Should we sample a time series more frequently?: decision support via multirate spectrum estimation

Guy P. Nason and Ben Powell,
University of Bristol, UK

Duncan Elliott
Office for National Statistics, Newport, UK

and Paul A. Smith
University of Southampton, UK

[Read before The Royal Statistical Society at a meeting organized by the General Applications Section, at the Society's 2016 annual conference in Manchester on Tuesday, September 6th, 2016, the President, Professor P. J. Diggle in the Chair]

Summary. Suppose that we have a historical time series with samples taken at a slow rate, e.g. quarterly. The paper proposes a new method to answer the question: is it worth sampling the series at a faster rate, e.g. monthly? Our contention is that classical time series methods are designed to analyse a series at a single and given sampling rate with the consequence that analysts are not often encouraged to think carefully about what an appropriate sampling rate might be. To answer the sampling rate question we propose a novel Bayesian method that incorporates the historical series, cost information and small amounts of pilot data sampled at the faster rate. The heart of our method is a new Bayesian spectral estimation technique that is capable of coherently using data sampled at multiple rates and is demonstrated to have superior practical performance compared with alternatives. Additionally, we introduce a method for hindcasting historical data at the faster rate. A freeware R package, regspec, is available that implements our methods. We illustrate our work by using official statistics time series including the UK consumer price index and counts of UK residents travelling abroad, but our methods are general and apply to any situation where time series data are collected.

Keywords: Aliasing; Bayesian statistics; Multirate; Spectrum estimation; Time series

1. Introduction

1.1. Practical context
In time series analysis it is well known that the sampling rate is an important consideration. If one samples infrequently, then information that changes rapidly (high frequencies) can be missed. In contrast, if one samples too frequently then one runs the risk of paying for sampling and storage of unnecessarily detailed information.

Often, one has access to a time series that has been sampled at a slow rate and a proposal is made to obtain more information by sampling at a faster rate. An equivalent situation arises when we consider rates and costs for publishing or communicating observations. Records of the number of UK residents travelling abroad (Fig. 1) provide one such example. One cannot

Address for correspondence: Guy P. Nason, School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK.
E-mail: G.P.Nason@bristol.ac.uk

© 2016 Royal Statistical Society 0964–1998/17/180353
Fig. 1. Quarterly and monthly (estimated) counts of trips abroad made by UK residents (O) predictions (hindcasts, explained later) for the monthly series given all the data and the estimated spectrum (O) (lines with lengths four times the prediction standard deviation; ———, guides for the eye; [-], indicator of when the sampling rate was increased) (source, Office for National Statistics, monthly overseas travel and tourism, May 2014, and overseas travel and tourism, quarter 1, 2014)

decrease the amount of information that is obtained by collecting samples more frequently. However, as increasing the sampling rate typically has real extra cost, the question ‘Is it worth sampling more frequently?’ is valid. Further analysis of the series of trips abroad by UK residents appears in example 3 below.

The concept of ‘worth’ is, in general, complex and situation dependent, depending not only on the measured variable in question but also the utility function of the observer and other stakeholders. Section 3 introduces a model that accounts for the various costs that are involved in sample rate switching and uses it to assess the resulting net benefit or loss of doing so, in a rational decision theoretic manner. This is achieved by concentrating on the second-order statistics of the process, with which we describe the joint population of observed and unobserved data. For second-order stationary time series these translate into the mean and the autocovariance, or its Fourier dual, the spectral density function, or spectrum. We focus, in particular, on the spectrum as its behaviour under different sampling rates is well understood.

For this paper we assume that the underlying quantity of interest is a (second-order) stationary stochastic process $X_t$ sampled on the integers $t$ with finite mean and variance. Now, purely for the exposition, since we shall eventually consider more general sampling schemes, suppose also that, maybe because of financial or technical constraints, we have only been able to collect a long series of observations at the even integers, $y_t = X_{2t}$, and there is a proposal to move to sampling $x_t = X_t$ on the integers. We would like to know the true spectrum $f_X(\omega)$ where $\omega$ is a frequency variable with domain $[0, \frac{1}{2})$.

Since $y_t$ is collected on the even integers, with sampling rate half that of the proposed $x_t$, an estimator of the spectrum based on $y_t$, $\hat{f}_Y(\omega)$ say, is typically considered to ‘see’ only frequencies in the range $[0, \frac{1}{4})$ because, owing to the well-known phenomenon of aliasing (which is discussed at length in Section 2), the top half of the frequencies in $X_t$ are folded onto and added to the bottom half. Clearly, the spectrum that we want, $f_X(\omega)$, cannot be uniquely identified from $f_Y(\omega)$. 
Should We Sample More Frequently?

However, if we use data sampled only at the faster rate, which we shall refer to as pilot or trial data, it could take some time before we can collect enough to obtain a reliable estimate of \( f_X(\omega) \). A key statistical innovation of this paper enables the incorporation of both the new (short) pilot data and the less frequently sampled established series together to obtain an estimate of \( f_X(\omega) \) better than those obtained with competing methods. Section 2 reviews aliasing and describes our new method for spectral estimation for multirate time series data.

There appears to be a relative paucity of classical statistical theory to support the many analysts who increasingly find themselves in receipt of time series at multiple rates. However, recently, the analysis of mixed frequency data has received much attention, mostly in econometrics. For example, mixed frequency data were used for forecasting in Ghysels et al. (2006), Clements and Galvão (2008), Armesto et al. (2010), Andreou et al. (2011) and Schorfheide and Song (2015), for vector auto-regressive modelling in Foroni et al. (2013, 2015), Cinadomo and D’Agostino (2015), Eraker et al. (2015) and Götz and Hauzenberger (2015), and for assessing Granger causality in Ghysels et al. (2015a, b). A recent overview of models for mixed frequency data is given by Foroni and Marcellino (2013). The goal of our paper is completely different in that we use a frequency domain approach to estimate spectra based on data taken at multiple rates, often with small amounts of faster data, and try to ascertain whether it is worth continuing to sample at the faster rate.

1.2. Modelling strategy

Our strategy for modelling time series in the type of situation that was sketched out in Section 1.1 centres on what we shall call linear Bayes (LB) spectrum estimation. This section briefly lays out our motivations for adopting this strategy, whereas the rest of the paper reinforces them.

Firstly, we approach the parameterization of second-order stationary time series in terms of a spectral density rather than, for example, a set of seasonal auto-regressive integrated moving average (SARIMA) parameters because the aliasing phenomenon, which will turn out to be so dominant in our problem, is well understood and easily described in the frequency domain. In the time domain, where the SARIMA parameters effectively live, it is not.

Secondly, we adopt a Bayesian mindset for the project as a whole because of its capacity to structure inferences combining heterogeneous sources of information coherently, and to make sense of inferences as components of decision problems.

Thirdly, we employ simple and fast LB methods, described in Goldstein and Wooff (2007), to structure our inferences because, in the absence of expedient conjugate relationships, implementation of an exact full Bayes inference for a high dimensional object, such as a spectrum within the span of a large number of basis functions, threatens to become computationally burdensome. Certainly, a Monte Carlo method could also be used for approximating full Bayes estimates. The scale and subtlety of the work required for implementation in such problems is discussed in Ceperley et al. (2012), who highlighted, in particular, the unresolved challenges that are faced when dealing with multimodal posterior density functions in high dimension. Example 2, which we present in Section 2.5, exhibits both features. As with the classical periodogram, we shall also want to refit our models with more degrees of freedom as more data become available. Incorporating this would necessitate a reversible-jump-like step, which further complicates issues of algorithm mixing and convergence, as discussed in Green and Hastie (2009).

One can identify many previous contributions to Bayesian spectrum estimation, ranging from the classical work of Whittle (1957) to contemporary Bayesian analyses including Mallick et al. (2002), Choudhuri et al. (2004), Pensky et al. (2007) and Rousseau et al. (2012). We would, however, place our work closest to the, chronologically intermediate and not expressly Bayesian, papers from Jones (1976) and Wahba (1980). Their application of simple, yet powerful,
smoothing formulae to the log-spectrum is built on here in the way that we apply similar formulae
to the assimilation of data at different sampling rates, and in our interpretation of mean-squared
error type statistics as posterior or adjusted variances, which allow us to quantify expected costs
associated with uncertainties for the spectrum.

1.3. Review of spectrum estimation for different types of missing data
Multirate data can be viewed as an example of single-rate data with systematically missing
observations. Good reviews of established methods for spectrum estimation with missing data
can be found in Broersen and Bos (2006) or section 10.4 of Broersen (2006) who divided the
field into three categories.

The first is least squares spectral analysis (LSSA), which was developed by Vaniček (1971),
and also known as the Lomb–Scargle periodogram of Lomb (1976) and Scargle (1982). Broersen
and Bos (2006) stated that this method works well for processes with strong sinusoidal behaviour
contaminated with additive noise, reproducing spectral peaks accurately but not spectral slopes.
Conventional implementations of LSSA ascertain an operational frequency range according to
an average spacing between all (fast and slow sampled) available data. They then proceeded
to compute periodogram-type statistics by projecting the data onto a set of sinusoids spread
over these frequencies. As a result, and by default, LSSA methods do begin to look for higher
frequencies as the fast sampled data are introduced by gradually moving the upper bound on
the spectrum’s effective support upwards. This is far too slow for us, however: as soon as we
know that the sampling rate may be doubled, regardless of the number of fast sampled data
we currently have, we know that we should be thinking about the spectrum at all frequencies
that are identifiable at the higher rate. To be clear, this is not a criticism of the Lomb–Scargle
method itself, but of its applicability to the type of problem that we are most interested in.

The second category centres on state space models, which encompass all SARIMA models.
These are naturally untroubled by missing data, which are effectively integrated out as and when
the filter-type algorithm, responsible for computing parameter likelihoods, reaches them. The
problem here, which we discuss at length in the on-line supporting materials, is that without
explicit formulation of a process’s spectrum we have no way to encode and prepare for the
known consequences of the aliasing phenomenon. Postponing a mathematical description, one
can probably recall, from watching car wheels turn on a television programme perhaps, how the
subsampling of a highly periodic signal gives rise to illusory oscillations far below the frequencies
of the original signal. It will turn out that the true and illusory frequencies correspond to
well-separated modes in the likelihood that state space calibration methods conventionally rely
on. Without explicit knowledge and consideration of this multimodality, all likelihood-based
SARIMA methods, from likelihood maximization to sophisticated Markov chain Monte Carlo
methods, run the risk of favouring one mode over another purely on the basis of arbitrary initial
conditions and/or random Markov chain Monte Carlo proposals. Data sampled at a higher rate
ought gradually to smooth out one of these modes but will only slowly alleviate this fundamental
problem.

The third group of methods relies on expectation–maximization algorithms to impute re-
peatedly missing data based on a current estimate for the spectrum, before re-estimating that
spectrum with conventional methods; see Wang et al. (2005) for an example. Such alternating
estimation procedures may also be adapted into Gibbs samplers, leading to the possibility of
spectrum variance estimation via a Markov chain Monte Carlo sample of simulated spectra.
By employing a spectral parameterization, the distinct modes of the SARIMA likelihood be-
come continuous ridges that reflect the idea that spectral mass can be transferred directly from
one aliased frequency to another. The ambiguity for the second-order structure is now better behaved, but it is still present and provides a large volume for stochastic inference algorithms to explore. For algorithms that are very fast or ingeniously adaptive such exploration need not be demanding, but we shall argue that the aliasing-induced ambiguity does not need thorough exploration because we already understand it perfectly.

The next section introduces our new method which is tailored for the slow–fast sampling scheme that we are considering. Its superiority over the first two established methods is explored below in example 2. We shall not, however, compare it with the third group of methods since we do not see them as competing models, but rather as a large class of alternative numerical methods for approximating the same Bayesian inference.

2. Infrequent sampling, aliasing and spectral estimation

2.1. Aliasing

Firstly, we establish notation for a long history of observations, with a subscript denoting low sampling rate, taken at every $K$th time point,

$$D_{\text{low}} = \{x_K, x_{2K}, \ldots, x_{N_{\text{low}}K}\},$$

and the subsequent series of observations sampled at integer time points,

$$D_{\text{high}} = \{x_S, x_{S+1}, x_{S+2}, \ldots, x_{S-1+N_{\text{high}}}\},$$

where $S$ is an integer used to locate $D_{\text{high}}$ in time, and $N_{\text{low}}$ and $N_{\text{high}}$ give the lengths of the series.

The key conceptual object for describing the variability within and between these data sets is the spectrum (scaled to exist on $0, \frac{1}{2}$ here, though $0, \pi, \pi$ or $-\pi, \pi$ are often used alternatives).

As is well known, the spectrum $f$ is related to the autocovariance function $\gamma$ for a stationary process according to

$$f(\omega) = 2 \gamma(0) + 4 \sum_{\tau=1}^{\infty} \gamma(\tau) \cos(2\pi\tau\omega), \quad \omega \in (0, \frac{1}{2}),$$

and variants of these formulae are found in the literature.

The effect of undersampling a stationary process is often interpreted in terms of the ‘aliasing phenomenon’, which causes components of a signal at certain frequencies to be confounded, or aliased, with others. The aliases of a particular frequency are pairs of reflections in a set of points given by integer multiples of the Nyquist frequency $\omega_N = \frac{\pi}{K}$, where $K$ is the time interval between observations. It follows that from $D_{\text{low}}$ we can learn about only a folded or aliased version of $f$,

$$s(K\omega) = K^{-1} \sum_{x \in A(\omega)} f(x)$$

where $A(\omega) = \{x = j/K \pm \omega | j \in \mathbb{N}, 0 \leq x \leq \frac{1}{2}\}$ is the set of frequencies that are aliased with $\omega$. It may be shown that equation (4) is the spectral density of the subsampled process for which time is rescaled so that observations appear to occur at integer time points. The extra information in data set $D_{\text{high}}$, concerning the full unfolded spectrum, can be appreciated as filling in the relative sizes of the summands in equation (4).
It may also be shown that observations of known linear combinations of the process at every \( K \)th time point are similarly only informative for the \( K \)-folded spectrum, and that only minor adjustments are needed below to apply it to the situation where a low sampling rate regime involves observing sums or averages of the process at intervals of \( K \) time units. This is so since the act of observing a particular known linear combination of values is equivalent to passing the signal through a filter, whose effect is to multiply the original spectrum by a second, known, spectrum; see, for example, Chatfield (2003), chapter 9.

Section 2.2, next, describes how we propose to estimate the spectrum by using single-rate data taken at every time point and then Section 2.3 adapts our methodology for every \( K \)th time point. Section 2.4 explains how we cope with subsampled and aggregated time series. Finally, Section 2.5 then explains how we can combine these techniques in a Bayesian learning framework to deal with multirate data and gives some examples.

### 2.2. Spectral estimation for a series observed at every time point

Our chosen method for estimating the spectrum of a stationary process treats the logged periodogram ordinates of a time series as noisy observations of the log-spectrum, with an additive bias and (almost) homogeneous noise variance following Wahba (1980). This leads to linear regression-type calculations for learning about the coefficients in a basis representation of the log-spectrum.

The regression parameters can be obtained from the \( D_{\text{high}} \) time series of length \( T = N_{\text{high}} \) by the raw periodogram, e.g. Brockwell and Davis (2009),

\[
I_T(\omega_j) = T^{-1} \left| \sum_{t=1}^{T} x_t \exp(2\pi i \omega_j t) \right|^2,
\]

where

\[
\omega_j = \frac{j}{2n_\omega}, \quad j = 0, \ldots, n_\omega, \quad n_\omega = \left\lceil \frac{T+1}{2} \right\rceil.
\]

The \( I_T(\omega_j) \) are asymptotically independently distributed as exponential or scaled \( \chi^2 \) random variables, depending on their index; see chapter 10 of Brockwell and Davis (2009) or chapter 4 of Shumway and Stoffer (2006), for example. Both distributions are special cases of gamma distributions whose log-moments are available in analytic form: for a general \( Y \) distributed according to a gamma distribution with shape parameter \( \alpha \) and rate parameter \( \beta \) we have

\[
\mathbb{E}\{\log(Y)\} = \psi(\alpha) - \log(\beta),
\]

\[
\text{var}\{\log(Y)\} = \psi_1(\alpha),
\]

where \( \psi \) is the digamma function and \( \psi_1 \) is the trigamma function; see identities 4.352(1) and 4.358(2) from Gradsteyn and Ryzhik (2007). This result suggests to us that we may usefully model the logged periodogram values as observations of their mean value, obscured by additive noise with variance independent of that mean value. The asymptotics thus lead us to the following model of the log-periodogram as proposed by Wahba (1980), page 123:

\[
\log\{I_T(\omega_j)\} = c_j + \log\{f(\omega_j)\} + e_j,
\]

with \( c_j = \psi(1) \approx -0.58, \mathbb{E}(e_j) = 0 \) and \( \text{var}(e_j) = \psi_1(1) \approx 1.64 \) for the interior frequencies that are asymptotically exponential distributed frequencies, and \( c_j = \psi(\frac{1}{2}) + \log(2) \approx -1.27, \mathbb{E}(e_j) = 0 \) and \( \text{var}(e_j) = \psi_1(\frac{1}{2}) \approx 4.93 \) for the ‘boundary frequencies’ that are asymptotically scaled \( \chi^2 \) distributed.
We then choose to model the log-spectrum as a superposition of $M$ basis functions,

$$
\log \{ f(\omega) \} = \sum_{k=0}^{M-1} \beta_k b_k(\omega),
$$

(8)

where the $\beta_k$ are scalar constants and the $b_k(\omega)$ are scalar functions that we shall later gather into (unsubscripted) $M$-dimensional vector quantities $\beta$ and $b(\omega)$. In the examples of Section 4 we employ a Fourier series representation for the log-spectrum, with the effect that the zeroth basis coefficient, which multiplies a unit constant basis function, attains special significance since it is equal to the logged forecast variance

$$
\text{var}_H(X_t) = \exp \left[ 2 \int_0^{1/2} \log \{ f(\omega) \} d\omega \right] = \exp(\beta_0),
$$

(9)

which is Kolmogorov’s formula. More precisely, equation (9) quantifies the mean-squared error of a best linear prediction for a process value given an infinite history of observations from it, which is alluded to by the subscript $H$, for history. We shall use Kolmogorov’s formula extensively in Section 3 where it is used to inform a cost analysis, enabling us to deduce whether switching to a faster sampling rate is expected to be worthwhile.

Choosing such a basis of sinusoids means that our parameterization coincides with a quantity that is known as the cepstrum (see, for example, Childers et al. (1977)) and renders the model particularly adept at describing spectra with regularly spaced peaks, which occur often in physical processes in the form of harmonics. Nevertheless, it would still be interesting, and unproblematic, to consider alternative bases that are more efficient at spanning other function spaces that are considered relevant for different problems.

Prior moments for the logged spectrum are induced by specifying prior moments for the coefficients $\beta_k$. In the examples to come, for example, we specify that the coefficients are all uncorrelated and have variances that decay as the frequency of the basis functions that they multiply increases. A background and further justification for this sort of specification, which favours smoother log-spectra, can be found in Wahba (1981), who reserved particular attention for variance specifications of the form

$$
\text{cov}(\beta_j, \beta_k) \propto \delta_{jk}(m + \lambda \nu_j)^{-m}
$$

(10)

for $j, k = 0, \ldots, M - 1$, where $\nu_j$ is the frequency of the $j$th basis function, $\lambda, m > 0$ act as roughness and shape (hyper)parameters and $\delta_{jk}$ is the Kronecker delta. By also specifying that only the expectation of $\beta_0$ is non-zero, we encode the belief that the log-spectrum will probably deviate smoothly from that of a white noise process. We are free to modify this specification, however, with the addition of log-spectra of other processes, to the right-hand side of equation (8) for example.

After calculation of the periodogram ordinates $I_T$, remembering that $T = N_{\text{high}}$, we adjust the basis coefficients $\beta$ by plugging them in, along with prior moments, $E(\beta)$ and $V$, for $\beta$ and those for the noise terms, namely the $e_j$ in equation (7), derived from expression (6), into LB adjustment equations to produce the adjusted expectation and variance quantities

$$
E_I(\beta) = E(\beta) + VB^T(BVB^T + \Sigma)^{-1}[\log(I_T) - E\{ \log(I_T) \}],
$$

(11)

$$
\text{var}_I(\beta) = V - VB^T(BVB^T + \Sigma)^{-1}BV,
$$

(12)
where superscript T signifies transpose and, using square brackets and subscripts to identify the scalar elements of an array,

\[
\begin{align*}
[\beta]_k &= \beta_k, \\
[I_T]_k &= I_T(\omega_k), \\
[V]_{j,k} &= \text{cov}(\beta_j, \beta_k), \\
[B]_{j,k} &= b_k(\omega_j).
\end{align*}
\]

(13)

When \( T = N_{\text{high}} \) is odd

\[
E[\log\{I_T(\omega_j)\}] = \begin{cases} 
 b(\omega_j)^T \mathbb{E}(\beta) + \psi(\frac{1}{2}) + \log(2), & j = 1, \\
 b(\omega_j)^T \mathbb{E}(\beta) + \psi(1), & j = 2, \ldots, T,
\end{cases}
\]

with \([\Sigma]_{j,k} = \pi^2/2\) if \( j = k = 1, \), \([\Sigma]_{j,k} = \pi^2/6\) if \( j = k = 2, \ldots, T \) and \([\Sigma]_{j,k} = 0\) otherwise. When \( T \) is even

\[
E[\log\{I_T(\omega_j)\}] = \begin{cases} 
 b(\omega_j)^T \mathbb{E}(\beta) + \psi(\frac{1}{2}) + \log(2), & j = 1, T, \\
 b(\omega_j)^T \mathbb{E}(\beta) + \psi(1), & j = 2, \ldots, T - 1,
\end{cases}
\]

with \([\Sigma]_{j,k} = \pi^2/2\) if \( j = k = 1, T, \), \([\Sigma]_{j,k} = \pi^2/6\) if \( j = k = 2, \ldots, T - 1 \) and \([\Sigma]_{j,k} = 0\) otherwise.

Equations (11) and (12) may be found in their more general form on pages 56–57 of Goldstein and Wooff (2007). Even without an appreciation of LB methodology, however, the general form of equations (11) and (12) will be familiar from their frequent recurrence as constrained estimators in the regression literature, in the form of Tikhonov regularization or ridge regression formulae for example.

To compute approximate credible intervals for the values of the log-spectrum, we shall later associate the adjusted expectations and variances for them, which are induced by the adjusted moments for the basis coefficients, with modes and expected squared deviations from the modes of probability distributions. Then, assuming unimodality of these hypothesized distributions, we use Gauss’s inequality (Pukelsheim, 1994) to compute conservative credible intervals for the variables that they describe. Using this inequality provides us with intervals that accommodate a wide range of posterior distributions while being considerably narrower than those derived for more general settings, such as Chebyshev’s inequality.

2.3. Spectral estimation with series of observations at every \( K \)th time point

The trick to extending our methodology so that we can make inferences from time series of observations at every \( K \)th time point is to employ a first-order Taylor expansion of the logarithm of the folded spectrum, \( s \) from equation (4), in terms of the logarithm of the unfolded spectrum, \( f \).

To use the trick in practice, we begin by taking the discrete Fourier transform of the subsampled data \( D_{\text{low}} \) of length \( R = N_{\text{low}} \), just as we did with \( D_{\text{high}} \) in the previous section:

\[
J_R(w_j) = R^{-1} \left| \sum_{r=1}^{R} x_{rK} \exp(2\pi iw_j r) \right|^2 ,
\]

(15)

with

\[
w_j = \frac{j}{2n_w}, \quad j = 0, \ldots, n_w, \quad n_w = \left[ \frac{R+1}{2} \right],
\]

and we understand these periodogram values as being informative for the folded spectrum.
The role of the Taylor series expansion is to provide an expression for the logarithm of the folded spectrum that is linear in the logarithm of the unfolded spectrum at the aliased frequencies, and so also linear in the basis coefficients $\beta$. This means keeping the expansion’s constant term and $K$ linear terms corresponding to each of the aliased frequencies, whereas the higher order terms are discarded on the assumption that they are negligibly small. So, using a subscript ‘Å’ to denote the central values (of the spectrum $f_\#(\omega)$ and folded spectrum $s_\#(\omega)$) around which we expand, we write the Taylor series expansion as

$$
\log\{s(K\omega)\} = -\log(K) + \log\left\{ \sum_{x \in A(\omega)} f(x) \right\}_{f_\#} + \sum_{x \in A(\omega)} \frac{\partial \log\{s(K\omega)\}}{\partial \log\{f(x)\}}_{f_\#} \left[ \log\{f(x)\} - \log\{f_\#(x)\} \right] + O\left( \sum_{x, x' \in A(\omega)} \left[ \log\{f(x)\} - \log\{f_\#(x)\} \right] \left[ \log\{f(x')\} - \log\{f_\#(x')\} \right] \right),
$$

and, filling in the derivatives by using the chain rule and truncating the series, we derive the expression

$$
\log\{s(K\omega)\} \approx -\log(K) + \log\left\{ \sum_{x \in A(\omega)} f_\#(x) \right\} + \sum_{x \in A(\omega)} \frac{f_\#(x)}{K s_\#(K\omega)} \left[ \log\{f(x)\} - \log\{f_\#(x)\} \right],
$$

where

$$
s_\#(K\omega_j) = K^{-1} \sum_{x \in A(\omega_j)} f_\#(x),
$$

$$
f_\#(\omega) = \exp \left\{ \sum_{k=0}^{M-1} \mathbb{E}(\beta_k) b_k(\omega) \right\}.
$$

The vector of log-periodogram values $J_R(w_j)$ with $j = 0, \ldots, n_w$, calculated from $D_{low}$ in equation (15), is understood to be a noisy observation of $\log(s)$. More precisely, the log-periodogram values $J_R(w_j)$ are modelled by exactly the same kind of model as in equation (7) except that $f$ is replaced by the folded spectrum $s$. Further, $\log(s)$ is approximately a biased weighted average of certain values of $\log(f)$, with the weights of the average determined by the values of the expansion’s centre, $f_\#$. Specifically, the ratio of the central spectrum at a particular frequency to that of the sum from all its aliases determines the degree to which $D_{low}$ modifies the estimate of the spectrum there. For example, when the central spectrum is constant over $(0, \frac{1}{2})$, all the $K$-coefficients in equation (16) are equal so observation of $J_R(w_j)$ affects our estimate for the spectrum equally at each of the aliased frequencies, resulting in the type of symmetry seen in Fig. 3(a) in Section 2.5. When the central spectrum puts the majority of its spectral mass on one aliased frequency, the effect of the periodogram ordinates $J_R(w_j)$ is to adjust this value while leaving the others relatively unchanged.

The expansion point $f_\#$ needs to be specified before we can construct the equations for adjusting our beliefs about $\beta$ given the subsampled data, and the natural candidate for this is the spectrum that is described by the prior expectation for $\beta$. Consequently, our prior expectation for $\beta$, informed by intuition, expert knowledge or other data, not only serves to inform estimates for the spectrum; it also directs the information in $D_{low}$ to certain parts of the spectral domain and so determines how we shall use it to adjust our beliefs.

In later versions of our code for spectral estimation we have also investigated recomputing the
linearization at the posterior expectation derived from the previous expansion point. With this strategy, each adjustment of the basis coefficients begins to look like a Newton optimization of an approximate posterior density. Although this development shows great promise, it is not integral to the principles underlying the method that is described here, and further exposition is reserved for future work.

The adjustment equations (11)–(12) for odd $R = N_{\text{low}}$, for the subsampled data, become

$$
E(J/\beta) = E(\beta) + VC^T (CVC^T + \Sigma)^{-1} \{ \log(J_{R}) - E\{\log(J_{R})\} \},
$$

(19)

$$
\text{var}_J(\beta) = V - VC^T (CVC^T + \Sigma)^{-1} CV,
$$

(20)

where

$$
[C]_{j,k} = \sum_{x \in A(w_j)} \frac{f_*(x)}{K} s_*(Kw_j) b_k(x),
$$

(21)

and

$$
E[\log\{J_R(w_j)\}] = \begin{cases} 
\log\{s_*(w_j)\} + \psi\left(\frac{1}{2}\right) + \log(2), & j = 1, \\
\log\{s_*(w_j)\} + \psi(1), & j = 2, \ldots, R.
\end{cases}
$$

(22)

The constants in the expectations (22) and $\Sigma$ in (equations 19) and (20) are derived from the same results for logged gamma variables as those in equations (11) and (12) but are necessarily of different dimension, depending on the lengths of the respective series. Analogous modifications are made for even $R$.

### 2.4. Spectral estimation with other derived series

Until now we have assumed that the historical data have arisen as the result of subsampling a finer resolution time series. In many situations, however, coarsely sampled data arise as the result of aggregation of finer series. For example, for the UK traveller data the quarterly number of trips by residents will be the sum taken over 3 months, and not the value in a given month.

We incorporate both subsampling and aggregation in the last stage of our method by considering the observations $z_t$ to be the filtered time series

$$
z_t = \theta_0 x_t + \theta_1 x_{t-1} + \ldots + \theta_U x_{t-U},
$$

where $x_t$ are the values of the process whose spectrum $f$ we are trying to estimate and the $\{\theta_u|u = 0, \ldots, U\}$ are known constants. To tackle this problem, we introduce additional notation for the spectrum of the filtered process and its folded version

$$
g(\omega) = \Theta(\omega) f(\omega),
$$

(23)

$$
t(K\omega) = K^{-1} \sum_{x \in A(\omega)} g(x)
$$

(24)

respectively, where $\Theta(\omega) = |\sum_{u=0}^U \theta_u \exp(-2\pi i u \omega)|^2$ is the squared gain of the filter $\{\theta_u\}_{u=0}^U$; see Chatfield (2003), section 9.3. We then denote the vector of periodogram values calculated from the series $\{z_K, z_{2K}, \ldots, z_{RK}\}$ by

$$
[L]_j = L_R(w_j) = R^{-1} \left| \sum_{r=1}^R z_{rK} \exp(-2\pi i w_j r) \right|^2,
$$

(25)

where $\{w_j\}$ is as in equation (15). The Bayes linear adjustment equations in this case are
\[ \mathbb{E}_L(\beta) = \mathbb{E}(\beta) + VF^T (FVF^T + \Sigma)^{-1} [\log(L_R) - \mathbb{E}\{\log(L_R)\}], \]  
(26)

\[ \text{var}_L(\beta) = V - VF^T (FVF^T + \Sigma)^{-1} FV, \]  
(27)

and include arrays which are populated analogously to those in Sections 2.2 and 2.3: \( \Sigma \) is a diagonal matrix with values \( \pi^2/2 \) or \( \pi^2/6 \); \( V \) and \( \beta \) have exactly the same interpretation. The objects

\[ [F]_{j,k} = \sum_{x \in A(w_j)} g_\star(x) \frac{g_\star(x)}{K_t\star(Kw_j)} b_k(x), \]  
(28)

where

\[ g_\star(\omega) = \Theta(\omega) f_\star(\omega) = \Theta(\omega) \exp\left\{ \sum_{k=0}^{M-1} \mathbb{E}(\beta_k) b_k(\omega) \right\}, \]  
(29)

\[ t_\star(K\omega) = K^{-1} \sum_{x \in A(\omega)} g_\star(x) \]  
(30)

denote the values of the aggregated and folded aggregated spectra corresponding to the Taylor series expansion’s centre, and

\[ \mathbb{E}[\log\{L_R(w_j)\}] = \begin{cases} 
\log\{t_\star(w_j)\} + \psi\left(\frac{1}{2}\right) + \log(2), & j = 1, \\
\log\{t_\star(w_j)\} + \psi(1), & j = 2, \ldots, R,
\end{cases} \]  
(31)

(for odd \( R \)) are defined analogously to equations (21) and (22).

### 2.5. Estimation with a mixture of types of series: sampling slow then fast

To begin with, before we have looked at any data, we specify a prior mean and variance for \( \beta \). Then, adjustment of these quantities given a time series requires that we pass them through equations (26) and (27), or the more specific alternatives (19) and (20) or (11) and (12), depending on whether we are dealing with the situations described in Sections 2.4, 2.3 or 2.2. The adjusted expectation and variance quantities that result from these equations are then carried forward to be plugged into those for the next adjustment as prior moments. In the present context, the ‘next adjustment’ means incorporating information that is obtained from the fast sampled data.

This simple picture is complicated slightly by the linearization of the folded spectrum, however. In a full Bayes analysis, the type in which there are no numerical constraints or approximations, the order in which adjustments are made ought not to affect their cumulative effect. This is not so here because the value of \( \mathbb{E}(\beta) \) influences adjustments via the value of \( f_\star \).

Although it is difficult to provide an authoritative answer to which order is best, we recommend, if possible, adjusting by the high frequency time series first so that the linearization for the folded spectrum, which tends to provide most of the spectrum shape information, provides a more faithful approximation to the true folded spectrum. In this way, we use \( D_{\text{high}} \) to tell us roughly where the spectral power is before using \( D_{\text{low}} \) to tell us more precisely what shape the spectrum takes there.

The methodology that was described in Sections 2.1–2.4 has been encapsulated in a freely available R (R Development Core Team, 2009) package called \texttt{regspec}. We now describe two examples showing \texttt{regspec} in action.
2.5.1. Example 1 (spectrum estimation for slow–fast sampled data)

This example illustrates the type of inference that is made possible by our new procedure’s ability to assimilate time series observations at two different sampling rates. Specifically, the task is to estimate an auto-regressive moving average ARMA(3,1) process’s spectrum from a thinned (slow sampled) series of 256 observations at even integer time points, followed by an unthinned (fast sampled) series of only 16 observations at integer time points. The process’s parameters are \( \phi = (-0.5, 0.4, 0.8) \) for the AR part and \( \theta = 0.2 \) for the MA part. Our prior for this example asserts that

\[
\begin{align*}
\mathbb{E}(\beta_j) &= 0, \\
\text{cov}(\beta_j, \beta_k) &= 9\delta_{jk} \{1 - \exp(-\lambda)\} \exp(-\lambda j),
\end{align*}
\]

where \( j, k = 0, \ldots, 99 \) and \( \exp(-\lambda) = 0.9 \). This covariance specification comes from taking the limit of expression (10) as \( m \to \infty \). Fig. 2 shows estimates for the spectrum based on just the 16 fast sampled observations for our method along with a standard out of the box AR and kernel-smoothed periodogram spectral estimates, giving an impression of their comparatively poor performance. The wide credible intervals are crucial in diagnosing the imprecision of all the estimates given so few data.

Estimates based on few data will clearly be poor. So, as we have only few fast sampled observations, can we improve our estimate by incorporating information from the ‘previously collected’ slow sampled data (sampled every second integer)? Fig. 3(a) shows a spectral estimate by using only the slow sampled data but with the knowledge that we shall progressively incorporate the fast sampled data. Hence, the horizontal axis, rather than being on \((0, 0.25)\) which would be the normal frequency range observable (for samples on the even integers), is plotted on \((0, 0.5)\), which is the range that is associated with the incoming fast sampled data. Fig. 3 has unfolded the spectral estimate from \((0, 0.25)\) in a symmetric way about 0.25: as we have not yet incor-

![Fig. 2. True spectrum for the ARMA process (— — —), spectral estimate produced by our new method (——), the AR estimate (— — —), kernel-smoothed periodogram estimate (— —), our approximate 50% and 90% credible intervals around (○), and periodogram values calculated from the short fast sampled series (○).](image-url)
Fig. 3. Same formatting as Fig. 2, showing equivalent estimates based on (a), (b) slow sampled series and (c), (d) the slow sampled and fast sampled series together in the cases when thinning leaves only (a), (c) every second observation and (b), (d) every third observation.
porated any higher frequency data, we have no reason to upweight contributions to the \((0, 0.25)\) spectrum from either the lower or higher halves of the new spectrum.

Fig. 3(c) shows the result of incorporating the additional 16 fast sampled observations. The spectral peak in Fig. 3(c) is estimated to a much higher degree of accuracy compared with that in Fig. 2 and the credible intervals are dramatically narrower.

Figs 3(b) and 3(d) show a similar situation to Figs 3(a) and 3(c) except that the thinning procedure takes only every third observation. The result is similar: incorporating additional less frequently sampled data can boost estimation performance for a small amount of more frequently sampled data.

Despite the usefulness of the plots in Fig. 3, there is much information that they do not communicate. Specifically, although the intervals here reflect the pointwise variance of the spectrum, they say nothing about the covariances between the values of the spectrum at different frequencies. These covariances quantify ideas along the lines that if the spectrum is smaller than expected at one place it ought to be correspondingly larger than expected in another. Utilizing this covariance information is important when aliasing is present; section 2 of the on-line supporting information provides more details of how this might be done.

### 2.5.2. Example 2 (comparison of spectrum estimators)

We now compare three methods in the example 1 scenario with \(N_{\text{low}} = 256\) observations sampled on the even integers followed by integer-sampled observations of varying lengths. The methods that are used are the kernel-smoothed Lomb–Scargle periodogram (LSSA), the spectrum corresponding to the ARMA model fitted by maximum likelihood (ML) and our LB adjusted expectation that is tailored for this scenario.

Our performance metric for each simulation \(m = 1, \ldots, 1000\) is the discrepancy statistic

\[
d_m = N^{-1} \sum_{i=1}^{N} \left[ \log \left\{ f(\omega_i) \right\} - \log \left\{ \hat{f}_m(\omega_i) \right\} \right]^2,
\]

where \(\hat{f}_m(\omega_i)\) is a point estimate of \(f(\omega_i)\) for simulation \(m\), and the \(\{\omega_i\}\) is an equally spaced set of \(N_\omega = 256\) points across \((0, \frac{1}{2})\).

We set \(N_{\text{low}} = 256\) and repeat the simulation 1000 times for each of a range of values for \(N_{\text{high}}\). Table 1 shows the superiority of our new LB estimate over the LSSA and ML ARMA methods.

| \(N_{\text{high}}\)   | Estimates for the following methods: |   |
|----------------------|--------------------------------------|---|
|                      | LSSA | ML ARMA | LB  |
| 8                    | 1.53 | 1.90    | 1.48|
| 16                   | 1.50 | 1.75    | 1.03|
| 32                   | 1.45 | 1.80    | 0.65|
| 64                   | 1.35 | 1.55    | 0.35|
| 128                  | 1.24 | 1.44    | 0.21|
| 256                  | 1.12 | 1.50    | 0.13|
| 512                  | 0.99 | 1.28    | 0.08|
| 1024                 | 0.92 | 1.07    | 0.05|
As alluded to in Section 1.3, and described extensively in the on-line supporting materials, we expect the ML ARMA method to do badly since it can easily be tricked into latching onto illusory frequencies when it follows its likelihood surface into a local mode. This is reflected nicely in Fig. 4 in the form of distinct modes in the discrepancy statistic that correspond to the differences between the modes of the likelihood.

The relatively poor performance of the LSSA method here is exacerbated by the true spectrum having a peak near the upper limit of the spectral domain. Only in the limit when the fast sampled data vastly outnumber the slow sampled data can the method even look at the highest frequencies.

### 2.5.3. Example 3 (UK residents’ trips abroad)

We revisit the number of UK residents travelling abroad series shown in Fig. 1 where \( N_{\text{low}} = 28 \) and \( N_{\text{high}} = 14 \). The plot is informal in so far as plotting data points with different meanings: the upper series consists of sums over quarters, whereas the lower values are sums over months.

For this example, we use our model to encode the difference of the log-spectrum of the observed process to that of a baseline SARIMA(1, 0, 0) × (1, 0, 0) model with both AR coefficients equal to 0.6 with seasonal period 12 months, constant intercept of 5200 and innovation variance of 3600. This prior baseline expectation recognizes the dominant annual periodicity that is expected and observed in these data. The priors for the difference are of the same form as expression (32) with \( \exp(-\lambda) = 0.8 \) and \( \delta_{j,k} \) instead of \( 9\delta_{j,k} \).

Fig. 5 shows estimates of the log-spectrum for the monthly series based on the \( D_{\text{low}} \) and \( D_{\text{high}} \) individually and combined \( D_{\text{low}} \cup D_{\text{high}} \). One should compare the improved estimate (Fig. 5(b)) which combines the slow and fast sampled data with the estimate constructed just from the fast sampled data (Fig. 5(a)). Incorporating the slow sampled data has tightened the annual peak at wavelength 12 months and appreciably improved the estimation of the lower frequencies in, approximately, the left-hand third of the plot. More precisely, the log-spectral estimate in Fig. 5(b) follows more closely the accurate low frequency information that is obtained from the quarterly data plot and, additionally, the credible intervals are noticeably tighter in the lower frequency region. At first glance it can be difficult to differentiate between the estimates in Fig. 5. Fig. 12 in the on-line supporting materials shows the differences between the adjusted means of the estimates, more clearly demonstrating the effects of combining the slow and fast sampled data.

Fig. 6 plots approximate means and 90% credible intervals for the auto-correlations of the trips abroad series. These are derived via a three-step approximation procedure where a multivariate normal distribution on the log-spectrum induces a log-normal distribution on the spectrum; a normal approximation of this log-normal spectrum induces a multivariate normal distribution on the autocovariances; finally, from the pragmatic moment approximations for ratios of correlated normal variables (Marsaglia, 2006), we derive means and variances for the auto-correlations, which inform the credible regions.

The most obvious change that is observed when comparing the auto-correlations estimated from just the monthly data (Fig. 6(a)) with those by using the combined series (Fig. 6(b)) is the tightening of the credible intervals, reflecting the increased information that is supplied by the quarterly data. We also note that Fig. 6(a) features estimated auto-correlations that are revised upwards in Fig. 6(b): this initial underestimation can be attributed to the inability of the short monthly series to contribute to estimates of the longer lag auto-correlations. Also, the shape of the peak in the auto-correlations around 12 months is noticeably sharper in Fig. 6(b) than that shown in Fig. 6(c).
Fig. 4. Selection of histograms of the discrepancy statistic for the estimated spectra over 1000 realizations of the time series $D_{\text{high}}$ and $D_{\text{low}}$: (a) LSSA, $N_{\text{high}} = 8$; (b) LSSA, $N_{\text{high}} = 32$; (c) LSSA, $N_{\text{high}} = 256$; (d) ML ARMA, $N_{\text{high}} = 8$; (e) ML ARMA, $N_{\text{high}} = 32$; (f) ML ARMA, $N_{\text{high}} = 256$; (g) LB, $N_{\text{high}} = 8$; (h) LB, $N_{\text{high}} = 32$; (i) LB, $N_{\text{high}} = 256$
Fig. 5. Estimates of the log-spectrum for the number of trips abroad made by UK residents each month given different training data: (a) informed only by 14 monthly data points (○, log-periodogram values calculated from the monthly data); (b) informed by both data sets; (c) informed by only the quarterly data
Fig. 6. Estimates of the auto-correlation values for the number of trips abroad made by UK residents each month (corresponding to the spectral estimates shown in Fig. 5) (the height of the shaded region at each lag is approximately 4 standard deviations of the auto-correlation value there): (a) informed only by the monthly data; (b) informed by both data sets; (c) informed by only the quarterly data (x, auto-correlations (obtained by interpolation of the log-spectrum and assuming smoothness) not at lags of multiples of 3 months)
Fig. 6(c) is interesting, as it shows how slow sampled data are leveraged against our prior for the log-spectrum (in this case, one that implies that the log-spectrum deviates smoothly from that of a seasonal auto-regressive process) to produce estimates for auto-correlations at completely unobserved lags (highlighted in Fig. 6 by using red crosses). Although the precise nature of this leveraging, whose legitimacy rests on the Bayes theorem and our approximations of its implications, is difficult to describe concisely in this instance, it is perhaps useful now to discuss it by providing a commentary on the flow of information through our model.

Our prior belief for a smooth log-spectrum translates to a prior for quickly decaying auto-correlations, which underlies the bounds of the auto-correlations tightening around zero in Fig. 6. The expected location of spectral power determines how the auto-correlations are effectively interpolated between values for which slow sampled data are immediately informative. Here, the filter gain takes the form of a spectral density with the greatest part of its mass over the first third of the spectral domain. With an implicit prior assumption the spectral mass of the monthly process is distributed evenly. Hence, periodogram ordinates from the aggregated series are predominantly informative for the spectrum over this third. When these ordinates are observed to be greater than expected, our power estimates over the lower frequencies are revised upwards. The accumulation of power at low frequencies then translates into the smooth (rather than oscillatory) interpolation of the auto-correlations observed in Fig. 6(c).

Our work with this example underlines the importance of combining information from lower and higher sampling rates after a change in sampling strategy, in particular for publication purposes. Doing so will facilitate the estimation of trends in the data, represented at lower frequencies and also of regular seasonal fluctuations that are often estimated and removed in official statistics publications. Discarding the quarterly information here can be seen to lead to wider credible intervals at seasonal and lower frequencies for example.

One particularly important use for estimates of the second-order quantities that we study here is the prediction of unobserved process values. To demonstrate our ability to do this, we derive monthly hindcasts for the ‘trips abroad series’ as best linear predictors by using all the available data (the quarterly pre-2011, and monthly post-2011) and autocovariance information derived from the spectral estimate in Fig. 5(b), which itself uses all the data. The hindcasts and error bars (of length four times the prediction standard deviation, for visibility) are plotted in blue in Fig. 1. Note that the direction of time here is unimportant, and the same method could be used to infer a process’s values given a densely sampled past and sparsely sampled future; a calculation relevant for scenarios in which the sampling rate is decreased.

Bayesian hindcasting methods have been developed in the fields of climatology and meteorology; see, for example, Solomon (2007), page 689, Werner and Holbrook (2011), Ortego et al. (2014) and references therein. However, hindcasting using multiple sample rate assimilation via a multirate spectral approach as developed here appears to be new and fully coherent in the Bayesian sense.

3. Pricing sampling strategies

Above, we established a strategy for learning about a spectrum in a multirate context. Now we seek to formulate and solve the decision problem of determining whether a switch to a higher rate is cost effective. To structure our understanding of possible realities, and so to provide a rationale for making decisions, we propose studying the cost function
Table 2. Cost function notation

| Notation | Meaning |
|----------|---------|
| $C_{\text{samp}}$ | Standard observation cost |
| $C_{\text{trial}}$ | Trial observation cost |
| $C_{\text{change}}$ | One-off change cost |
| $F_{\text{low}}$ | Forecast cost in the low frequency regime |
| $F_{\text{high}}$ | Forecast cost in the high frequency regime |
| $K$ | Slow sampling interval from Section 2.3 |
| $N_{\text{trial}}$ | Number of trial observations |
| $N_{\text{future}}$ | Number of future time intervals which contain either one or $K$ observations depending on whether we switch or not |
| $1_{\text{change}}$ | Sampling strategy change indicator function |

\[
C = N_{\text{trial}} C_{\text{trial}} + 1_{\text{change}} C_{\text{change}} + N_{\text{future}} [C_{\text{samp}} + F_{\text{low}} + 1_{\text{change}} \{(K - 1) C_{\text{samp}} + F_{\text{high}} - F_{\text{low}}\}],
\]

for which the notation is summarized in Table 2.

Although specification of the cost function serves to endow our uncertainties with meaning, in doing so, it necessarily forces us to entertain certain assumptions. Implicit to equation (33), for example, is the notion that our current default strategy is to collect data at the low sampling rate and that resources would need to be expended to switch to the higher sampling rate. Further implications are discussed in section 1.2 of the on-line supplementary material.

Returning to our motivating problem, we would like to be able to advise a client considering a change to the measurement of a time series on whether they ought to sample more often, and on how easy it would be for them to collect evidence informing this decision. In terms of the cost function, we interpret these goals as providing recommendations for an optimal value for $1_{\text{change}} \in \{0, 1\}$ and for $N_{\text{trial}} \in \mathbb{N}$. For now, we assume that the costs $C_{\text{samp}}$, $C_{\text{trial}}$ and $C_{\text{change}}$, and the number $N_{\text{future}}$, are fixed constants that are given to us. We shall return to the plausibility of this assumption in the fifth paragraph of Section 5.2. We also anticipate that specifications for the forecast costs $F_{\text{low}}$ and $F_{\text{high}}$ are likely to be highly context specific. Certainly, in real applications, one can easily imagine discussing, and attempting to elicit, complex relationships between forecasting skill and actual costs. Although it will be interesting to investigate more realistic relationships in the future, we concentrate now on a widely applicable and, crucially, tractable relationship whereby forecast costs are considered to be proportional to the logarithm of the forecast variance, which is derived from the log-spectrum via Kolmogorov’s formula:

\[
F_{\text{low}} = C_u \log \{\text{var}_{H_{\text{low}}}(x_t)\} = 2 C_u \int_0^{1/2} \log \{s(\omega)\} d\omega,
\]

\[
F_{\text{high}} = C_u \log \{\text{var}_{H_{\text{high}}}(x_t)\} = 2 C_u \int_0^{1/2} \log \{f(\omega)\} d\omega,
\]

where $f(\omega)$ and $s(\omega)$ are the spectrum and folded spectrum as defined in equations (2) and (4) respectively, and $C_u$ is a calibration constant that converts prediction success into monetary terms based on one’s utility. The intuition behind this selection is that the ability to forecast the future provides a tangible benefit. These cost specifications coincide up to proportionality with
Should We Sample More Frequently?

the expected log-likelihoods for process values given infinite histories of observations from the process, which we have denoted here with $H_{low}$ and $H_{high}$.

To make a specification for $I_{change}$, and so for the future sampling strategy that we currently favour, we must ask whether the cost of switching relative to not switching,

$$Q := C_{change} + N_{future}(K - 1)C_{samp} + F_{high} - F_{low},$$

is greater than 0, in which case we stick with the low frequency sampling regime, or less than 0, in which case we switch to the high frequency sampling regime. In practice, however, since $F_{high} - F_{low}$ is unknown, we can only estimate the cost and base our decision for which sampling strategy to choose on this estimate. This decision process gives rise to the additional random variable that defines a future cost reduction based on our current beliefs for $Q$:

$$RD := \begin{cases} Q, & \mathbb{E}_D(Q) \leq 0, \\ 0, & \mathbb{E}_D(Q) > 0, \end{cases}$$

where $\mathbb{E}_D(Q)$ denotes our expectation for $Q$ given a generic data set $D$. We also note that $Q$ is bounded above since $F_{high} < F_{low}$, reflecting the fact that better informed forecasts should always lead to smaller costs:

$$Q < Q_{max} = C_{change} + (K - 1)N_{future}C_{samp}.$$  

The same Taylor series expansion that enabled us to incorporate undersampled time series in our inferences in Section 2.3 now allows us to produce an approximate mean and variance for $F_{high} - F_{low}$, and thus for $Q$, since we approximate it with the following linear combination of the elements of $\beta$:

$$F_{high} - F_{low} \approx C_u \left( \beta_0 - 2 \int_0^{1/2} \log \left\{ \frac{s_*(\omega)}{\omega} \right\} d\omega \\ - 2K \int_0^{1/(2K)} \frac{f_*(x)}{Ks_*(K\omega)} \left[ \log \left\{ f(x) \right\} - \log \left\{ f_*(x) \right\} \right] d\omega \right) \\ = C_u \left[ \beta_0 - 2 \int_0^{1/2} \log \left\{ \frac{s_*(\omega)}{\omega} \right\} d\omega - 2K \left\{ \int_0^{1/(2K)} \frac{f_*(x)}{Ks_*(K\omega)} b(x) d\omega \right\}^T \\ \times (\beta - \beta_0) \right],$$

where $b(x)$ is the same vector of basis function values introduced in equation (8), and superscript $T$ signifies the matrix transpose. To formulate the expected or anticipated cost reduction that is afforded by a trial, we need only to consider the events which will cause us to adjust our expectation for $Q$ from one side of zero to the other. These are the instances in which the information that is gathered in the trial causes us to change our mind for the long-term sampling strategy to employ. To proceed, we must recognize that our adjusted expectation for $Q$ given the trial data is a random variable, since we have not yet performed the trial. We then need to calculate the expected cost reduction as the probability that our adjusted expectation for $Q$ crosses the decision threshold, multiplied by the average expected cost reduction given that this happens.

To gather information for the behaviour of our future adjusted expectation we return to the variance estimate (12). This formula tells us how our prior variance for the coefficients $\beta$ ought to be reduced after observing some data, or, seen in another way, how our prior variance for $\beta$
can be partitioned into a portion that disappears once data have been observed and a portion that remains. The part that disappears is the variance for what our adjusted expectation will be after observing the data. Additionally, we also know our current expectation for this adjusted expectation: it must be the same as our current expectation for $\beta$ for our beliefs to be coherent.

Since we have specified covariance properties for $\log(f)$ a priori through the coefficient vector $\beta$, and since estimate (12) does not depend on the periodogram values, we can anticipate how the variance for our beliefs for $\log(f)$ and the linearized version of $F_{\text{high}} - F_{\text{low}}$ (37) will evolve. Consequently, given the expectation and variance of $Q$ informed by some (possibly empty) historical data set $D_0$, both the expectation and the variance of our adjusted expectation given the additional series $D_{\text{trial}}$ (with length $N_{\text{trial}}$) are known in advance of performing the trial, taking us a step closer to calculating the expected benefit of trials of various length.

To quantify explicitly the expected cost reduction that is afforded by a trial, we load the moments of the adjusted expectation into a particular distribution and calculate

$$
\mathbb{E}(R_{D_0 \cup D_{\text{trial}}} - R_{D_0}) = \begin{cases} 
\int_0^{Q_{\text{max}}} \mathbb{E}_{D_0 \cup D_{\text{trial}}}(Q) \phi \{ \mathbb{E}_{D_0 \cup D_{\text{trial}}}(Q) \}, & \mathbb{E}_{D_0}(Q) \leq 0, \\
-\int_{-\infty}^{0} \mathbb{E}_{D_0 \cup D_{\text{trial}}}(Q) \phi \{ \mathbb{E}_{D_0 \cup D_{\text{trial}}}(Q) \}, & \mathbb{E}_{D_0}(Q) > 0,
\end{cases}
$$

where $\phi$ is the probability density function assigned to $\mathbb{E}_{D_0 \cup D_{\text{trial}}}(Q)$.

In our work so far with this methodology, we have found it effective to plug the moments for the future expectation for $Q$ into a reflected and translated gamma distribution with support $(-\infty, Q_{\text{max}}]$. We have also experimented with using a normal distribution here, ignoring the upper bound $Q_{\text{max}}$. Our findings appear reasonably robust to this choice, in so far as producing trial length recommendations that still lead to average reduced costs in simulations, but, since calculations are straightforward in both cases, we favour the gamma distribution to account for the upper bound (36).

Identifying an appropriate value for $N_{\text{trial}}$ requires that we compute approximate values (38) for all feasible values of $N_{\text{trial}}$, and adding to them the costs that are required to pay for the trials. The trial length corresponding to the minimal net expected cost then constitutes our recommendation for $N_{\text{trial}}$. To clarify, according to our reasoning framework we always have an expectation for $Q$ and, consequently, always have a preference between high and low sampling rates. By identifying a non-zero value for $N_{\text{trial}}$ at which the expectation for equation (33) is minimized, however, we are identifying that the risk that is associated with acting on our current preference equates to a cost that is greater than that of a trial. If the minimizing trial length is zero, then we are ready to present our recommendation for the switching decision variable, $\mathbf{1}_{\text{change}} = \mathbf{1}\{\mathbb{E}(Q) \leq 0\}$, immediately without collecting any trial observations.

4. Examples

4.1. Example 4 (estimating forecast variances)

In this example we test the accuracy of our log-spectrum-derived estimate for the statistic (37), which quantifies the potential for increasing the precision of forecasts. The investigation takes the form of a Monte Carlo simulation experiment in which we draw a large ensemble of log-spectra from a multivariate normal prior with moments of the form (10), and simulating two time series of normally distributed values from each. We then use these time series to recover the log-spectra. The first series here is of length 90 and consists of observations at every other time point, whereas the second is of length 48 and consists of observations at every time point.
Our estimates for statistic (37) are their adjusted expectations, which follows from that of $\beta$. They are accompanied by approximate conservative 90% credible intervals calculated by using Gauss’s inequality.

Fig. 7 compares our estimates against the truth but also against alternative estimates, computed by using more conventional time series methods, that do not take the structural relationship between the forecast variances in the two sampling regimes into account. We shall refer to these as AR estimates. These are derived by, firstly, fitting AR models by ML to the pairs...
of time series, and calculating corresponding pairs of AR residuals. By conditioning on the ML estimates for the AR coefficients and entertaining improper uniform priors for the innovation variances for the two processes, we induce inverse gamma posterior distributions for them and a scaled $F$-distribution posterior for their ratio. We proceed to truncate the $F$-distribution in recognition of the idea that the forecast variance in the high sampling frequency regime cannot exceed that in the low sampling frequency regime. Finally, the median, which we take as a point estimator, and a central credible interval for this approximate posterior are computed from particular quantiles of the $F$-distribution, which are easily evaluated with standard statistical software.

Straightaway, we can see that the AR estimates perform relatively badly, achieving a mean absolute error over the simulations of 0.24 against an equivalent figure of 0.11 for the LB estimates. We also note that, over the ensemble of 1000 90%-approximate credible intervals calculated with the LB and AR approaches, the former are on average 0.58 as wide as the latter, and they capture the true value of the log-variance ratio in 86% and 81% of cases respectively.

4.2. Example 5 (UK consumer price index)

Here, we analyse a monthly consumer price index CPI series produced by the Office for National Statistics which measures the change in prices by using a chain-linked fixed basket approach. Informally, we understand CPI as being informative for typical consumer day-to-day spending as a percentage of a 2005 baseline. We preprocess the data firstly by passing them through the log-transform and then by removing a linear trend. Following this, we model the difference from the spectrum of a white noise process with variance $0.03^2$ and prior specification for the $\beta$s as in equation (32) but with $\exp(-\lambda) = 0.95$ and $4\delta_{j,k}$ instead of $9\delta_{j,k}$.

Our next step is to modify the data set to recreate the type of scenario that we anticipated in Section 2. This involves separating the first 270 data points and thinning them to every third value so that our historical low frequency data set $D_{low}$ consists of quarterly observations. For $D_{high}$ we take the subsequent 4 years of data at the monthly rate. From these two time series (plotted in Fig. 8) we produce a spectral estimate that encodes an expectation and variance for the linearized version of $F_{high} - F_{low}$, which is interpretable as a log-ratio of the forecast variance in the high sampling frequency regime to that in the low sampling frequency regime. Using the same strategy as used to produce plots for the log-spectrum, we associate these two moments with a mode and expected squared deviation from the mode before employing Gauss’s inequality to compute bounds for a conservative credible interval. By exponentiating the expectation and the bounds we produce forecast variance ratio values $0.34$, $0.47$, $0.65$, meaning that our analysis implies increasing the sampling rate will lead to a forecast variance of between 34% and 65% of that achieved otherwise, with probability at least 0.9. By looking at residuals from AR models fitted to the data sets $D_{low}$ and $D_{high}$, as we did in example 4, we compute a point estimate for $F_{high} - F_{low}$ of $-1.48$, which, on exponentiation, suggests a forecast variance reduction to 21% of the low sampling rate variance. The accompanying approximate credible interval, derived by using the $F$-distribution, suggests that the forecast variance can be reduced to between 15% and 35% with probability 0.9.

Both these estimates may be compared with the exponentiated bounds of the 90% conservative credible interval calculated by using the LB model for the log-spectrum given the whole, fast sampled time series, which we refer to as the oracle estimate. Informally, we understand the oracle estimate as being as close as we can reach to the sort of hidden true values available in our simulation experiments. The three variance ratio estimates and their associated prediction intervals are compared graphically in Fig. 9. We can see here that the AR estimate appears to
Fig. 8. Thinned (slow sampled) detrended log-CPI time series
overestimate the extra precision that is made possible by high frequency sampling. The LB point estimate for the forecast variance ratio is closer to that of the oracle and is accompanied by an interval which overlaps with the greater part of the oracle’s interval.

So we suggest here, without proposing specific but essentially arbitrary cost parameters that would serve to quantify the argument, that in the scenario that is described in this example, we would have been able to specify a more appropriate strategy in a variety of situations using the LB model rather than the calculations based on the AR fits. Firstly, we would argue that the evidence in favour of sampling at the monthly rate is not as strong as the AR estimate suggests. Thus the LB inference may have prevented us from incurring sampling costs which were not recouped by more precise forecasts in the future. Secondly, the wider credible interval for the LB inference encodes the idea that we would have been able to make the switching decision with less confidence than indicated by the AR calculations. The implication this time is that additional data and a postponement of a sampling strategy decision would have been of greater value than suggested by the AR analysis.

5. Discussion and conclusions

5.1. Methodology overview

We have shown the power and simplicity of log-spectrum smoothing techniques and how they can be extended to conduct inference using time series data sampled at multiple rates in a Bayesian framework. Our methods permit us to anticipate the set of plausible spectra confounded by the aliasing phenomenon in a way that existing procedures fail to do. Further, we have developed a method for quantifying our uncertainty for the spectrum in terms of its implications for forecasting and hindcasting, and proceeded to translate this uncertainty into a cost that may be weighed against the costs of obtaining further information, thus informing a decision problem that is genuinely relevant to applied statisticians and official statistics in particular.

An R Development Core Team (2009) package, titled re specs, for implementing the calculations is available from the authors and is available from the Comprehensive R Archive Network.

5.2. Further avenues and applications

One area where our central problem is very real is in official statistics, where surveys are often
Should We Sample More Frequently?

relatively large and costly, but there is constant pressure for more timely information to support evidence-based decision making in government, balanced against the need for efficiency in the use of public funds. In 1994–1995, for example, there was a review of the options for turning the Labour Force Survey into a true monthly survey by Steel (1997), but the subsequent public consultation on the options determined that there was insufficient benefit for the cost of £7 million–£8 million (in 1996 prices) as discussed by Werner (2006). However, a reduced cost solution of producing rolling 3-monthly estimates each month was introduced from April 1998, costing around £0.2 million (much of the cost was to produce an integrated labour market statistics release, not all for calculation of the rolling estimates). An experimental monthly series has been calculated from the existing (quarterly) design since 2004 as an aid to interpretation but is not considered to be of sufficient quality to stand in its own right.

All of these steps in the evolution of the Labour Force Survey could be assessed retrospectively by using the methods developed here; many of the data are available, and some of the cost estimates. The issue of whether a true monthly design would be beneficial remains topical, but the high cost of data collection for such a design has tended to make it an unlikely development. However, the topic is often revisited whenever labour market statistics receive heightened attention, such as when the Bank of England considered linking interest rate change decisions to the rate of unemployment during 2013. So there is a possibility of prospective analysis for future change.

Section 3 discussed the importance of the smoothness of the log-spectrum for the value of pilot observations to the decision. This is also related to the cost and design of a faster rate survey (whether the pilot or a full survey). It is unusual for a survey design to be carried forward wholesale for implementation at a faster rate; usually a design change is involved, often a change to the sample size, which affects the quality of the new series. For example, the construction output statistics produced by the Office for National Statistics changed from a sample of 12000 businesses per quarter to 8000 businesses per month in 2010 as described by Crook and Sharp (2010). This change would be expected to make the sampling variance of the high frequency (monthly) series higher than that of the low frequency series.

A change such as that in the construction output statistics suggests that other forms of the cost model (33) might be worthy of study. We could begin, for example, by separating out the costs for high frequency and low frequency observations to allow for differences in design. This would mean replacing the third term in model (33) to give

\[
C = N_{\text{trial}} C_{\text{trial}} + 1_{\text{change}} C_{\text{change}} \\
+ N_{\text{future}} \{1_{\text{change}} K(C_{\text{high}} + F_{\text{high}}) + (1 - 1_{\text{change}})(C_{\text{low}} + F_{\text{low}})\}. \tag{39}
\]

The cost function (39) then has some interesting special cases:

(a) \(C_{\text{trial}} = C_{\text{high}}\)—when the trials use the proposed new method, rather than being a smaller pilot; this allows straightforward \textit{post hoc} evaluation of the net benefit of a change already made to a statistical collection;

(b) when there is no change in data collection cost, but there is a change in publication frequency, i.e. when \(KC_{\text{high}} = C_{\text{low}}\).

This is like the Labour Force Survey example, where the outputs change periodicity without a change in the on-going collection costs (in practice there is usually a smaller cost to changing the estimation and compilation processes, and the publication schedule). There are doubtless many alternative ways in which the costs of changes could be measured.

In both model (33) and model (39), \(N_{\text{future}}\) codifies the period over which the cost–benefit
trade-off is to be evaluated for the decision process. It is only possible to recommend change when the third term in the cost function is negative and, given that it is indeed negative, the larger $N_{\text{future}}$, the more often we shall conclude that change is worthwhile. Some general guidance on appraisal of decisions like this is available from Her Majesty’s Treasury (2003) and is based on the service life of the asset under consideration. There is no general guidance on the service life of a statistical survey, but using information on service lives of research and development assets in Ker (2013) for information and communication services industries we can use 8–20 years as a rule of thumb. In fact a period of 20 years corresponds to the current life of some redesigns (such as the Labour Force Survey) and therefore feels appropriate, though this is not based on a comprehensive analysis of survey changes. The trade-off between cost and benefit is also affected by the relative size of the $C$- and $F$-parameters in the cost function; in general there may be a challenge in translating the variances (the $F$-parameters) into suitable monetary values.

Further motivation for the synthesis of data at multiple sampling rates can be identified in recent reports from the European Statistical System network. Among the network members’ documented concerns are issues relating to company value-added tax (VAT) registrations, which provide administrative data for monthly and quarterly turnover estimates. Vlag et al. (2010), in particular, discussed the difficulties that were faced by Statistics Netherlands in 2009 after abolition of legislation that obliged companies to produce monthly VAT declarations. Although the precise details are too complex to describe fully here, it is clear that the informativeness of their monthly VAT series was severely degraded, the annual series was not, since annual reports remained compulsory, and monthly turnover estimates were still required to fulfil commitments to national and European authorities. We contend that our methodology would have been of value here for its ability to propagate knowledge for a process’s variability from one sampling regime to another. Specifically, we would have been able to estimate both the spectral properties and the actual values of the process underlying the VAT figures from the historic monthly series and contemporary annual series. The deliverable of this procedure would have been a ‘nowcast’ accompanied by coherent credible intervals, resulting in a series of estimates qualitatively equivalent to the hindcast of Fig. 1 but with the time axis reversed. Squeezing information from the degraded contemporary monthly series would be challenging and, together with analysis of the UK construction output survey, represents an important direction for further work.

With an eye to the future, the consideration of changes to the way that the population census is conducted in England and Wales also fits within this framework. Office for National Statistics (2013) sets out options for a census using existing government data and compulsory annual surveys, which has considerable transition and on-going costs but has an advantage in allowing higher frequency (possibly annual) updates to some statistics on the population relative to the low frequency (decennial) population census. The cost estimates for the various approaches are also set out in the consultation document, so an appropriate cost function could be used to bring this information together to aid decision making.

It appears that the applications of this approach are many in topical areas of official statistics, and there are other areas of statistics where similar applications would also be valuable in evidence-based decision making.

Acknowledgements

We thank the Joint and Associate Editors and three reviewers who provided a wealth of helpful comments and suggestions that improved the paper. This work was supported by Engineering and Physical Sciences Research Council grant ER/K020951/1.
Kostas Triantafyllopoulos (University of Sheffield)

I warmly congratulate Nason and his colleagues for a clearly written and thought-provoking paper. They tackle an important, but not much discussed, issue in time series analysis, i.e. sampling frequency. They motivate their methodology and its performance and usefulness by considering official statistics time series, namely the consumer price index and counts of UK residents travelling abroad, alongside several simulated data sets. In the core of the methodology there is a proposal for spectrum estimation, adopting a linear model for the log-periodogram, which is estimated via Bayes linear methods. This is initially proposed for the unfolded version of the data (time series sampled at high frequency). Consequently, they propose a way to obtain the spectrum of the folded data (low frequency) from that of the unfolded version. Various extensions are provided to accommodate for linear functions of the unfolded data. A decision support mechanism, which considers costs, is proposed and a relatively simple solution in closed form is provided. The paper is illustrated by using simulated and real data and the discussion is thorough with detailed comments. These examples and the associated R package help to disseminate the methodology and to highlight its usefulness.

I thank the authors for bringing design into time series. This is a topic that is rarely discussed in time series and more should be done in this direction. The authors illustrate clearly that in the context they discuss there is significant benefit from sampling a time series at multirate frequencies. The methodology proposed is quite general and, although the authors have done a good job illustrating it in the context of their examples, the methodology may be applied more widely to other time series data. One particular area that I am interested in is ‘big data’ and data mining where massive data sets are obtained. Such data frequently emerge in manufacturing, biology and finance, as data extraction mechanisms have been recently developed with the advance of computing. Perhaps the approach of this paper could be used to mine the data at very high frequencies and to obtain a sparse time series, which holds much of the information, but is easier to handle. A similar interest of mine is sampling continuous time time series, in which again it may be advantageous to consider sampling at lower frequencies and evaluating the loss of information as we move from high to low frequency. An extension of the methodology could be to

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary Material for “Should we sample a time series more frequently?: decision support via multirate spectrum estimation (with discussion)”’.

Discussion on the paper by Nason, Powell, Elliott and Smith
consider multivariate time series. Indeed, many of the macroeconomic time series are observed in multiple variables which are cross-correlated over time. In my opinion the authors’ approach could extend to such processes, but I do not know how the cross-spectrum would be estimated.

One of the disadvantages of the methodology comes from the limitation of Bayes linear methods used to estimate the coefficients of the log-periodogram model. The authors have argued convincingly why they avoid Markov chain Monte Carlo or other Monte Carlo estimation procedures, but I still believe that uncertainty analysis around the regression coefficients is limited to the Bayes linear optimality. For example, it is well known that Bayes linear estimation minimizes the expected risk as well as least squares, but it is less clear how this choice may affect spectrum estimation in the context of this paper and the data considered. One of the examples that the authors discuss involves a count time series (counts of UK residents travelling abroad). Much of the recent literature is focused on developing models for dealing specifically with count time series and not relying on transformations and stationarity. For example, for low values of the count (including 0s) the seasonal auto-regressive integrated moving average–stationarity transformation could be unreliable. Can the authors elaborate how such counts could be incorporated in their model and what challenges such data could pose? Similar comments apply to other discrete values time series, such as binary or binomial time series. Finally, I extend my congratulations to the authors for a thought-provoking paper, which I have no doubt will open up discussion on time series design and is likely to lead to further research in the near future.

Sofia Olhede (University College London)
I congratulate Nason and his colleagues on their very thought-provoking contributions to the estimation of spectra with different time series sampling mechanisms. Sampling time series is an important and currently much neglected area of statistics that they are to be commended on for having both recognized and contributed to. The main contribution of this paper is to question how time series should be sampled: rather than with a fixed sampling period, that has been preselected to avoid aliasing, the paper instead is considering the scenario when sampling can be altered or modified.

The paper starts by addressing how to quantify and alleviate aliasing. This is done by focusing on modelling the log-spectrum, using a non-parametric series expansion of this function—work that can be traced back to Jones (1976) and Wahba (1980). Using penalization (which can be interpreted in a Bayesian posterior inference context, as is done in this paper) is not a new concept, and related ideas for model choice have been explored for parametric time series analysis; see Hurvich and Tsai (1989, 1991).

The non-parametric representation of Nason and his colleagues can also be interpreted as the cepstrum (Bogart et al., 1963; Moir and Barrell, 2003), which is used in speech processing for analysing the pitch of human speech. Cosine series of the log-spectrum have been used to create new parametric model families (see for example Bloomfield (1973) and Beran (1993)), and these models are popular.

Finite sample effects also cause the observed variance of the discrete Fourier transform not to equate to the spectrum. Such effects have been shown to have a significant effect in finite sample estimation (Thomson, 1982; Riedel and Sidorenko, 1995; Sykulski et al., 2016). Perhaps such blurring effects are more dominant than aliasing in realistic sampling scenarios. Depending on what frequencies are more important to the estimation of a given parameter the trade-off between sampling period and number of samples can be made formally, and prediction of future values may not in all contexts be as important as small mean-square error for parameter estimation. If we remain on a coarser grid, for predictive purposes, better spectral estimates may not ensure better predictions. In some instances more complex procedures partially reaching a higher Nyquist frequency, or a finer frequency grid, may be the answer, adaptively changing both the sample duration and sampling period. Such thinking is complementary to the notion of costs and predictive performance that is given in the paper.

Estimation based on basis expansions of the log-spectrum, such as implemented by Nason and his colleagues, can lead to averaging of the log-periodogram rather than using least squares of the periodogram for the exponential of the log-periodogram model; this leads to delicate variance trade-offs (Robinson, 1994, 1995). Finite sample size effects are difficult to quantify, and even identifiability of continuous time models may be more difficult to treat from discrete observations than might be envisioned (Jones, 1981).

Any stochastic process that is sampled and inferred starts life in continuous time. Aliasing arises because variation cannot be disambiguated. Models that are discussed in the paper look at integer multiples of a set sampling period. Naturally, we may start directly from a continuous model of the process $X(t)$ allowing for any type of sampling. We can easily, in more general settings than econometrics, imagine that the process
could be sampled by using any increasing sequence \( \{t_n\} \). It is possible with the right choice of sequence to remove all aliasing completely (Masry, 1978). For some econometrics applications implementing non-regular sampling may be impossible; but equally for many applications removing preprocessing steps may just recover data at a finer irregular resolution, containing important high frequency information.

The choice of priors, where Nason and his colleagues have been inspired by Wahba, is also important in this setting. Most smoothing priors will encourage log-spectral smoothness, and a rapid decay of the correlation structure in time. This may not be appropriate for realistic signals with strong forms of periodicities. Some care must be adopted in their selection, and especially if families of Fourier coefficients should be kept or killed together, which would encourage usage of the group penalties and forms of structured sparsity (Wolfe et al., 2004; Yuan and Lin, 2006).

In conclusion, the authors are to be congratulated for their very thought-provoking contributions to the analysis of time series. Sampling is often treated as problem provided, and set in stone, when actually this is not so. Throwing off the yoke of unrealistic discrete modelling, and reverting to the true continuous time model, yields many opportunities to uncover more exact inference from our data, and the authors are to be commended for recognizing this fact.

The vote of thanks was passed by acclamation.

**Yi Yu (University of Cambridge) and Ivor Cribben (University of Alberta)**

We congratulate Nason and his colleagues on their paper. Motivated by a survey data set, their paper proposes a new Bayesian spectral estimation method that allows for the coherent fusion of data sampled at different rates.

The multirate sampled phenomenon also appears in human neuroscience. Here, a great challenge is determining causal pathways in which different brain areas interact to support cognition and behaviour. We hope that their proposed method can be used to find a sensitive and stable estimate of Granger causality pathways on functional magnetic resonance imaging data in combination with data collected by using dynamic functional magnetic resonance inverse imaging, which achieves a very fast sampling rate. In addition, their proposed method could also play a role in the simultaneous recording and analysis of electroencephalography and functional magnetic resonance imaging data. This combination allows researchers to achieve both high temporal and high spatial resolution in the recording of human brain function, but the combination remains an on-going challenge. Another area facing a similar challenge is econometrics where models incorporate variables sampled at different frequencies (Ghysels et al., 2004; Clements and Galvão, 2008). For example, the gross domestic product figure is released quarterly, whereas a range of leading and coincident indicators are available more frequently. We hope to see an extension of the proposed method to these topical areas, although we are aware that the method does not assume overlapping sampling periods.

The paper concentrates on data that are collected at a slow-then-fast rate. However, we are interested in more general settings and the flexibility of their method. We use the `Dfexample` example from `regspec` to investigate the setting in which the first and the last 51 time points are sampled at every integer with the data in between sampled at every three time points. To obtain an estimator of the spectrum, we compare six different settings for updating the slow sampled region.

(a) Settings 1 and 2: use the estimators from only the first and second fast sampled regions.
(b) Setting 3: use the average of the estimators from the two fast sampled regions.
(c) Settings 4 and 5: update respectively the first and second fast sampled region with the estimators from respectively the second and first fast sampled region. Use the result as the estimator.
(d) Setting 6: use the estimator from splitting the second fast sampled region, and use one region to update the other.

Fig. 10 and Table 3 show the plots and the discrepancy criterion,

\[
d = N_w^{-1} \sum_{w=1}^{N} \left[ \log \left( \hat{f}(w) \right) - \log \left( \hat{f}(w) \right) \right]^2,
\]

used in the paper. Interestingly, in terms of the discrepancy values, the best results are obtained by setting 6.

| Setting | 1   | 2   | 3   | 4   | 5   | 6   |
|---------|-----|-----|-----|-----|-----|-----|
| \( d \) | 0.433 | 0.268 | 0.268 | 0.221 | 0.221 | 0.165 |
Fig. 10. (a) Setting 1; (b) setting 2; (c) setting 3; (d) setting 4; (e) setting 5; (f) setting 6
We also attained similar results when we increased the fast sample size and when the slow rate data were sampled at every four, five and six time points. Hence, the method proposed can handle more general data settings, and iterative updates may improve the final estimation.

Granville Tunnicliffe Wilson (Lancaster University)
I suggest that insight into the nature of the aliasing problem can be gained by fitting spectral components directly to the multirate sampled series. Using a fine grid of frequencies $f_k = k/N$ consistent with the highest sampling rate we represent the series $y_t$, sampled at any regular or irregular times, by

![Figure 11. A multirate sample from example 1](image1)

![Figure 12. Sample spectrum of the multirate series assuming different model spectra: (a) uniform spectrum assumed; (b) higher frequencies weighted; (c) lower frequencies weighted](image2)
\[ y_i = \sum_{k=0}^{\infty} a_k \cos(2\pi f_k) + b_k \sin(2\pi f_k) + \epsilon_i. \]

Initially, we need to assume a model spectral density \( S(f) \) for \( y_t \), so that the coefficients \( \psi = \{a_k, b_k\} \) are taken to be uncorrelated with variances \( 0.5 S(f_k)/N \). The observation error \( \epsilon_t \) is white noise with negligibly small variance. This random-effects model is fitted by the best linear unbiased prediction equations

\[ \hat{\psi} = (Z'Z + D^{-1}\sigma^2_e)^{-1}Z'y, \]

where the diagonal elements of \( D \) are the coefficient variances. For a sample series from example 1 shown in Fig. 11, with \( N = 1000 \), we obtain a sample spectrum proportional to \( \hat{a}_k^2 + \hat{b}_k^2 \) as shown in Fig. 12(a).

For regularly sampled series such sample spectra are very insensitive to the assumed model spectrum. The fitted coefficients are then the Fourier transform of the series extrapolated by using the model spectrum. For irregularly sampled series they are the transform of the series both extrapolated and interpolated, and become much more sensitive. Figs 12(b) and 12(c) show how the sample spectrum varies when the assumed model spectrum has greater magnitude at higher or lower frequencies. This reveals the ambiguity of representation which can be resolved by a suitable parsimonious model, as proposed in the paper. See Tunnicliffe Wilson et al. (2015), chapter 8, for an application to a series with completely irregular sampling.

A mixed model is sensible for the series of example 3, with components at the seasonal frequencies \( j/12 \) being taken as fixed effects. The design is modified to represent the quarterly totals. Assuming a uniform model spectrum, the mixed model equations

\[ \left( \begin{array}{cc} X'X & X'Z \\ Z'X & Z'Z + D^{-1}\sigma^2_e \end{array} \right) \left( \begin{array}{c} \hat{\theta} \\ \hat{\psi} \end{array} \right) = \left( \begin{array}{c} X'y \\ Z'y \end{array} \right), \]

give the sample spectra in Fig. 13. These plots shows no sign of possible difficulty with aliasing, nor any indication of time varying seasonality.

For the maximum likelihood auto-regressive moving average model simulation experiment better soft-
ware should be used. The results in Table 1 by using Genstat are 20% of those in the column headed LB.

Further, fitting the series of UK trips abroad with fixed monthly effects and an AR(2) error, monthly values during the period of quarterly data are well estimated as shown in Fig. 14. Auto-regressive moving average modelling works well for this example but does have limitations and this paper is to be commended as opening up new methodology for multirate series.

Marina I. Knight (University of York) and Matthew A. Nunes (Lancaster University)

We congratulate Nason and his colleagues for a stimulating paper which challenges current statistical thinking and practice to consider time series sampled at different sampling rates—unfortunately, an often overlooked question. This work has the potential to help with principled decision making in a range of

![Fig. 15. Spectral estimation of a fractional Gaussian noise series of length $n = 1024$, with Hurst exponent $H = 0.8$: (a) spectral estimate by using the original series; (b) distortion of the spectral estimate induced by dyadic subsampling; (c) spectral estimate corrected by using the first 30 unit-sampled observations via regspec](image)
scientific areas and, in this respect, we also commend the release of the \texttt{regspec} software accompanying the paper to aid practitioners.

Whereas the examples in the paper focus on arguably ‘well-behaved’ stationary processes with short memory, one could also consider the benefits of the proposed methodology in more complex settings. Specifically, consider processes with different spectral characteristics, such as spikes or exponential decay, present for long memory time series such as fractionally integrated processes or fractional Gaussian noise (see for example Beran et al. (2013)). In such contexts, the focus is often on the estimation of long-range dependence intensity in the series, quantified through a single quantity: the Hurst exponent $H$. Although several long memory estimation techniques exist, both in the time and in the frequency domains, we are not aware of any that could indeed handle combining information captured through the process sampling at more than just one sampling rate.

Interestingly, this work could pave the way to achieving just that: more reliable long memory estimation by means of corroborating information from subsampled series. As is intuitive, process subsampling results in poorer Hurst exponent estimation, for both time- and spectral-domain-based methods. Classical spectral estimation techniques such as Lomb–Scargle or least squares spectral analysis estimation (Vaníček, 1971; Lomb, 1976; Scargle, 1982) have been shown to be unreliable for long memory processes, and particularly so if the series has been subsampled (Broersen et al., 2000; Broersen, 2007). We conjecture that the by-product of this work, i.e. the spectral estimate that incorporates all process information, could be conducive to a new, more reliable long memory estimator constructed in the frequency domain.

An example of the proposed methodology on a long memory process is shown in Fig. 15. Although only illustrative, this example shows that the \texttt{regspec} correction seems promising here also. Could the authors comment on the wider influence of their work on estimation of secondary quantities often derived from such spectra, e.g. the Hurst exponent, and the potential of any theoretical development in this direction?

We would also like to hear the authors’ insights on the possibility of extending the work to the analysis of multivariate (multirate) series for, for example, cross-spectral or polyspectral estimation.

**Xiao-Li Meng (Harvard University, Cambridge)**

A toast and congratulations to the author(s) should be in order whenever a discussant enjoys reading a paper. But often this enjoyment leads to a self-condolence for the discussant: ‘Well, there goes another paper that I should have written’. The grapes are particularly sour when the discussant has been planting similar varieties in a neighbouring vineyard. In Meng and Xie (2014), who demonstrated that more data do not guarantee better results, we studied the effect of the sampling frequency, and the related interplay between data patterns and model assumptions, on estimating the auto-correlation in the simple autoregressive AR(1) model. The findings were somewhat intriguing, and I have been wondering what they would look like for more complex time series models.

Specifically, suppose in principle we can observe any part of an AR(1) series of indefinite length into the future

$$Y_t = \rho Y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad t = 0, 1, \ldots, \tag{40}$$

but in reality we can only afford to take $n$ observations. What is then the sampling frequency for optimally estimating $\rho$? Assuming that $n$ is sufficiently large to render the adequacy of the Fisher information approximation, we proved that the optimal spacing (between consecutive sampling times) is 1 when $\rho^2 \leq 1/3$, and it is 2 when $\frac{1}{3} < \rho^2 \leq \sqrt{(21/20) - 0.5}$. In general, the optimal spacing goes up with $\rho^2$ at the rate of $\left\{-\log(\rho^2)\right\}^{-1}$ as $\rho^2 \uparrow 1$; so does the maximal relative gain in efficiency for estimating $\rho$ compared with using the single spacing.

We also investigated the efficiency gain for estimating $\rho$ from knowing the value of $\sigma^2$. When the spacing $s = 1$, the relative gain in Fisher information is bounded above by $1/(n-1)$. However, it can be as high as 50% once $s = 2$ (achieved when $\rho = 0$), and it approaches $\infty$ as $\rho^2 \downarrow 0$ when $s = 3$. From a time domain perspective, this is because, once $s > 1$, estimating $\rho$ and estimating $\sigma^2$ become tangled because

$$Y_{t+|t} \sim N(\rho Y_t, k_s(\rho^2)\sigma^2), \quad k_s(x) = \sum_{j=0}^{s-1} x^j, \tag{41}$$

and hence $\text{var}(Y_{t+|t}|Y_t)$ depends on both $\rho$ and $\sigma^2$ when $s > 1$. In other words, the Fisher information matrix for $\{\rho, \sigma^2\}$ is diagonal if and only if $s = 1$.

The authors emphasize the gain of insights from a frequency domain perspective about missing observations in a time series. Since every time series can be represented equivalently in the frequency domain and in the time domain, I would be very interested in gaining additional statistical insights from the authors’ perspective about these mathematical findings.
I congratulate the authors on an excellent and thought-provoking paper. It follows in the very best traditions of papers read before the Society, developing novel methods motivated by a very real and important problem, while also opening up many interesting questions which I feel sure will give rise to several new research directions. You have also left me wanting to ask more of the methodology than one could ever reasonably hope to capture in a single paper.

If I follow the paper correctly, it assumes that the observed time series is real valued. However, several official statistics measures, including the Office for National Statistics’s international passenger survey data referred to in the paper, are integer valued. I wonder therefore how readily this methodology might be adapted to this special class of time series. If not what is the consequence of approximating such a series by the current approach?

The paper also sets out some very compelling scenarios by which data arising from different sampling rates might be utilized. Typically the scenario proposed is one of an existing series that might be replaced, or supplemented, by a more frequently sampled series. In practice I wonder whether the data available might actually be even richer than this, with the analyst having access to a hierarchy of data at different sampling rates, each representing a different level of granularity. For example, quarterly reported data might indeed be accessed monthly, but conceivably weekly or even daily data might be readily accessible in the future. Assuming that this hierarchy of data is present, how straightforward might it be to extend the approach to allow for data sampled at these different rates and could the approach be used to provide insight on which levels are worth including?

Finally, as the authors are aware, in recent years there has been a significant focus on developing theory and methods for non-stationary time series within the literature. Indeed, the question of alias detection has already received some attention in this setting (see, for example, Eckley and Nason (2014) or Eckley and Nason (2016)). I therefore wonder whether the authors have considered the adaptation of the new Bayesian spectral estimation technique to non-stationary time series. If so, how feasible might this approach be, computationally, in a potentially ‘big data’ setting?

This is a very interesting and impressively wide-ranging paper.

Frequentist linear analysis of log-periodogram ordinates has asymptotic efficiency just over 60% (Bartlett and Kendall, 1946; Cox and Hinkley, 1968): not disastrous but sufficiently low to signal a need for more elaborate methods of analysis sometimes. The result presumably applies also to linear Bayes analysis with flat priors although a Bayesian analysis with informative priors would fare better.

For series with a strong signal at a specified frequency, complex demodulation in which the imaginary part is plotted against the real part of a moving Fourier transform could presumably be adapted to data with unequal spacing. The technique, due, I think, to J. W. Tukey, was very carefully explained by Bloomfield (2000).

In the discussion of sampling intervals there are two different situations. In one the values observed are point values of a notional underlying process whereas in the other the values observed are essentially cumulative totals since the last sampling point. I was not always clear in the paper which was being addressed.

Given the number of observations allowed in a specified interval and given objectives requiring information over a very broad frequency band optimal sampling schemes might formally be regarded as within the province of the Kiefer–Wolfowitz theory of optimal design, although I imagine it would be a formidable task to implement that.

Many congratulations go to Nason and his colleagues for this important contribution on a very timely topic. Its influence may be beyond what has been explicitly stated in the paper: it is commonplace in this information age that data are collected irregularly and/or almost continuously over time. Analysis is often done with the accumulated data over regularly spaced time intervals. This study may provoke different workings in cleansing and in analysing those irregularly sampled time series data.

The estimated spectrum depicted in Fig. 3(c) is a significant improvement from that in Fig. 3(a), by using 16 extra ‘fast’ sampled data points. In fact the improved estimator is remarkably accurate and is free from the so-called ‘aliasing’, which seems too good to be true, as, for example, only those 16 added data points carry the information on the autocovariance at lag 1. One possible reason for this radical improvement is smoothing which, however, is a double-edged sword. Some empirical guidelines on how to choose the
amount of smoothness, especially in high dimensional cases, would be helpful in applying the method proposed.

The authors advocate a non-parametric approach, which is natural given the availability of multirate data. Nevertheless if our interest is solely on the first two moments properties of a second-order stationary process (such as its spectrum), an auto-regressive moving average (ARMA) specification provides a natural parametric framework with which the spectral domain inference tools such as the Whittle likelihood are well developed and understood. Does it help to know that the underlying process is ARMA? Are there any advantages to using multirate data in estimating the spectrum of an ARMA process?

The following contributions were received in writing after the meeting.

Amir Ahmad (United Arab Emirates University, Al Ain)
I congratulate Nason and his colleagues for writing this interesting paper. The paper discusses the question: is it worth sampling the series at a faster rate?

The cost function given in equation (33) is an important part of the paper. The cost function has many terms defined in Table 2. It would be interesting if the authors could give some values of these terms in a given domain. It would help readers to understand in a better way the importance of different terms.

Adam Butler (Biomathematics and Statistics Scotland, Edinburgh), Kate R. Searle and Francis Daunt (Centre for Ecology and Hydrology, Edinburgh)
Nason and his colleagues have highlighted an important applied problem that arises in the context of official statistics and have proposed an innovative methodology for addressing this problem. We note that the issue of determining an appropriate sampling rate is also of crucial importance in the field of movement ecology, which is concerned with describing, understanding and predicting the movement of wild animals (Hansson and Åkesson, 2014). Data on animal movement are usually collected by using electronic tags (such as Global Positioning System tags and accelerometers), and these tags will typically record data at a sampling rate that is specified by the user. The selection of this rate is therefore a key issue in the design of studies that involve electronic tagging data.

This problem is clearly similar to that addressed by the authors in the present paper, but there are two substantive differences.

(a) Within the context of electronic tagging data typically no explicit financial cost is involved in collecting data at higher frequency, because the financial costs relate to the purchasing of the tags and to the costs of capturing the animal. The choice of sampling rate is, however, non-trivial, because it will typically involve a trade-off—the battery life, for example, will often constrain the total number of records that can be collected from a single tag deployment.

(b) Although the design problem for electronic tags relates to the selection of the temporal sampling rate (as in the paper) the data that are associated with each record will often be either spatial (e.g. for Global Positioning System tags) or high dimensional (e.g. for accelerometers), and the approaches to analysis will therefore differ from those used in the analysis of time series data—within movement ecology they frequently include hidden Markov models and state space models (Langrock et al., 2012), continuous time models (Niu et al., 2016) and semiparametric approaches (e.g. kernel density estimation; Fleming et al. (2015)).

We currently adopt a simulation-based approach to the selection of the optimal sampling rate within our work, because an analytic approach does not appear to be tractable, but it would be interesting to know whether the approaches that have been developed within this paper can either be extended to deal with electronic tag data or else can give qualitative insights into the design issues that are associated with these types of data.

P. F. Craigmile (Ohio State University, Columbus) and D. B. Percival (University of Washington, Seattle)
Statistical analyses of time series typically assume that data are recorded at times separated by a fixed interval $\Delta$. Many phenomena in the physical sciences can be recorded at different intervals, making $\Delta$ a design parameter. Setting $\Delta$ is important because it impacts the degree to which aliasing can adversely alter potentially interesting portions of the spectrum. How to fix $\Delta$ is covered in classic works such as Blackman and Tukey (1958) and Brillinger (1981) and presumes some knowledge (from physical theory, previous analyses or a pilot study) about the spectrum for different $\Delta$. This interesting paper questions the assumption of fixing $\Delta$. It does so in the context of a time series (UK residents’ trips abroad) for
Discussion on the Paper by Nason, Powell, Elliott and Smith

which the spectrum may not be of much interest and for which the choice of $\Delta$ would seem to be dictated
more by reporting considerations (monthly versus seasonal) than by aliasing. What is innovative here is the
development of a quasi-empirical-Bayes sequential method that uses all available data, while minimizing
some cost function.

In any design problem, choice of the cost function is critical. Here the cost function is related to minimizing
the mean-squared error of estimating the log-spectrum (a proxy for minimizing forecast variance). This
cost function, although ideally adapted to the estimation method employed, is somewhat disconnected
from the tourism application if we take the goal to be identification of key seasonal features rather than forecasting. Is there a cost function that is cognisant of this goal?

Physical applications suggest two adaptations. First, the periodogram can be substantially biased. A
better starting point is a multitaper estimator (Percival and Walden, 1993). This approach, taken by Walden
et al. (1998), leads to log-multitaper models closely paralleling log-periodogram models. Second, physical applications often involve complex-valued series, so an extension involving two-sided spectra is of interest.

We close with three remarks. First, modelling the log-spectrum by using cosine bases is equivalent to
fitting an exponential process (Bloomfield, 1973). The number of bases included can be critical in
estimating the spectrum consistently; typically the number should increase as the sample size increases
(see, for example, Hurvich and Brodsky (2001) and Holan et al. (2009) in the context of long memory
processes). Second, Anderson's discussion of Diggle and al Wasel (1997) elucidates problems in regarding
the periodogram evaluated at the Fourier frequencies as being independent (the stringent conditions for
independence of Fuller (1996) may not apply here). Finally, might this problem be best expressed as a
state space problem, thus minimizing reliance on approximations and allowing for the formal updating of
information via filtering (e.g. Harrison and West (1999) and Hyndman et al. (2008))?

Peter J. Diggle (Lancaster University)
I very much welcome an all-too-rare contribution to design methodology for time series and would like to
comment on possible connections with the related area of geostatistical design.

In this two-dimensional setting, space is recognized as being continuous, and the issue of sampling
frequency can be hedged to some extent by using irregularly spaced sampling locations. If, for example,
we want to understand the behaviour of a spatial phenomenon, $S(x)$ say, over a region $A$, a good general
design strategy is to combine a spatially regular set of locations to cover $A$ with a sprinkling of closely
spaced pairs of locations; the former lead to precise predictive inferences given the model for $S(x)$, whereas
the latter enable estimation of high frequency effects that play an important role in defining the optimal
amount of smoothing to apply to the data. See, for example, Chipeta et al. (2016).

The time series analogue could present administrative difficulties in the context of official statistics, but
might it be worth considering in other areas of application?

Amira Elayouty, Marian Scott and Claire Miller (University of Glasgow)
The paper presents a very interesting novel Bayesian (frequency domain) method to determine whether it
is worth sampling a time series more frequently to gain a better understanding of the process under study.
According to the methodology proposed, the information obtained from the historical data sampled at the
slow rate and the pilot data sampled at the faster rate are summarized by their spectral densities which are
then mixed to obtain a better estimate of the spectral density. Nason and his colleagues have illustrated
the methodology by using the records of UK residents travelling abroad.

In the environmental context, the tremendous technical improvements in monitoring equipment allow
data to be recorded at high frequencies (faster rate) to gain a better understanding and to form a more complete picture of the phenomenon under study. However, in many cases, the benefits of the high frequency sampling are not always clear in terms of process understanding and present greater analysis challenges. It may often be that temporally aggregated data are more informative, being less impacted by very short and transient events—it depends on the scientific question of interest! Being able to determine an appropriate sampling rate is therefore highly appealing in environmental applications.

Environmental systems typically are highly dynamic and responsive to external events (such as intense rainfall). This could argue that really what is needed is adaptive sampling. It would then be practically important to have ‘realtime’ diagnostics to identify when the sampling equipment should switch from one sampling rate to another.

The methodology that is proposed in the paper relies on the definition of a cost function to determine whether collecting data at the faster rate is cost effective. However, quantifying the cost function may not be a straightforward task, involving both the monetary cost and the potential information loss (e.g. if predicting risk of flooding).
Nonetheless we welcome this paper and the methodology proposed; we would suggest an alternative question related to the data analysis, which would be can this tool inform about the most appropriate temporal aggregation for a given data series? In its final form, the question might be could this approach extend to determine what the faster sampling should be (so an estimation and inference question)?

Konstantinos Fokianos (University of Cyprus, Nicosia)

It gives me a great pleasure to congratulate Nason and his colleagues on a very nice and important contribution. The problem of obtaining a suitable sampling rate is of central importance to time series analysis. They propose a sound approach which is supported by theory and examples.

My discussion is focused on their basic model, given by equation (8) of the paper. The notion of cepstrum, i.e. the logarithm of the spectral density function, was introduced by Bogart \textit{et al.} (1963). The cepstral coefficients (or correlations) are the coefficients obtained by a standard Fourier expansion of the log-spectrum (see Pourahmadi (1984), for more). Because the logarithm of the spectral density function is a well-behaved function, Bloomfield (1973) introduced the exponential model of order \( p \) for the spectrum by

\[
f(\omega) = \frac{\tau^2}{2\pi} \exp \left\{ 2 \sum_{r=1}^{p} \theta_r \cos(\omega r) \right\}, \quad 0 < \omega < \pi,
\]

where \( \tau^2 \) and \( \theta = (\theta_1, \ldots, \theta_p) \) are unknown parameters. Obviously, model (8) in the paper belongs to this framework. In fact, any auto-regressive moving average process can be approximated by expression (42) (see Fokianos and Savvides (2008), page 319). Some further work on model (42) can be found in Parzen (1993), Holan (2004) and Holan \textit{et al.} (2016), among others.

Related work by Fokianos and Savvides (2008) shows that the log-ratio of the periodogram ordinates depends asymptotically on the log-ratio of their respective spectral densities. Motivated by this fact, they proposed modelling the log-ratio of two or more unknown spectral density functions for comparison and clustering; see also Savvides \textit{et al.} (2008). Earlier analogous references include the works by Coates and Diggle (1986) and Diggle (1990) who modelled the log-ratio of two spectral densities by a quadratic function. Given \( G \) unknown spectral density functions, say \( f_j(\cdot) \), let

\[
\log \left\{ \frac{f_j(\omega)}{f_G(\omega)} \right\} = a_j^T Z(\omega) - \pi < \omega < \pi,
\]

for \( j = 1, 2, \ldots, G - 1 \) where \( a_j = (a_{j0}, a_{j1}, \ldots, a_{jp})^T \) is a \((p + 1)\)-dimensional vector of unknown parameters and \( Z(\omega) = (1, 2 \cos(\omega), 2 \cos(2\omega), \ldots, 2 \cos(p\omega))^T \). The order \( p \) is assumed to be known and fixed. When \( a_j = 0 \), for all \( j \), then all \( G \) spectral densities are equal, i.e. the time series have identical second-order properties.

A useful idea is to extend the author's methodology to the case of \( G \) independent time series. Suppose that \( G = 2 \). Then equation (8) in the paper can be considered afresh in connection with expression (43); in this case, the constants \( \beta_k \) of equation (8) are interpreted as the 'generalized' difference of the cepstrum correlations between the two series. Then, it might be interesting to develop a methodology along the lines of Nason and his colleagues for combining data on different sampling rates. If, in addition, the newly defined parameters \( \beta_k \) fluctuate around 0 then it might be worth investigating clustering procedures which discover similar series (in the sense of second-order properties).

Subhashis Ghosal (North Carolina State University, Raleigh) and Anindya Roy (University of Maryland, Baltimore)

Nason and his colleagues study an important practical problem with implications in various fields including official statistics. They provide a thought-provoking solution for the problem whether it is worth increasing the sampling rate in a stationary time series for better prediction given a cost constraint. In their setting, the problem essentially reduces to estimating \( F_{\text{high}}(\cdot) \) on the basis of infrequently sampled data. The solution proposed by the authors through Taylor’s expansion is a nice way to avoid the non-linearity of the functional and to estimate this quantity. Nevertheless some aspects may require further attention. In model (8), \( M \) is a smoothing parameter which in principle should be made larger with increasing data, or be given a prior so that it can be automatically determined by the data. A deeper issue is estimation of \( f \) (and its functional \( F_{\text{high}}(\cdot) \)) based on infrequently sampled data. Since the folded spectrum \( s \) does not uniquely identify \( f \), the posterior of the coefficient vector (or \( f \)), given the low frequency data, can be multimodal. This may make the remainder in expression (16) substantial. The reference point \( f_{\text{ref}} \) if chosen as the posterior mean may not
be one of the points of concentration, and hence the remainder may be large. The problem may be alleviated by using all possible modes as references points, estimating $F_{\text{high}} - F_{\text{low}}$ by using the representation (37) based on posterior samples near each mode and merging the estimated values according to the weights that the posterior attributes to these modes. This can be done by using the posterior samples obtained using the infrequently sampled data. Nevertheless, the estimate of $F_{\text{high}}$ will be generally inconsistent except possibly in some low dimensional parametric families where all parameters are uniquely determined by $s$. Thus, in general, one can only estimate the projection of $F_{\text{high}}$ on the space of functional of $s$.

Finally, instead of using a log-periodogram regression (7) with model (8) to drive posterior moments, the exponential model

$$I_T(\omega) \sim \text{Exp}\{1/f(\omega)\}$$

based on the Whittle likelihood may be used and a random series prior based on splines with positive coefficients (such as inverse gamma for ‘conjugacy-like’ structure) may be used to model $f$ directly; see Shen and Ghosal (2015). This approach will lead to the same basis coefficients for $f$ and $s$ except that the basis for $s$ will be the folded version of that of $f$.

**Eric Ghysels** (University of North Carolina, Chapel Hill, and Centre for Economic and Policy Research, London), **Reza Solgi** (Harvard University, Cambridge) and **Antonietta Mira** (Università della Svizzera Italiana, Lugano, and University of Insubria, Como)

Nason and his colleagues acknowledge that, whereas the literature on the topic is sparse, there is a growing literature on so-called mixed data sampling (MIDAS) regression models and several of its contributions, including Andreou et al. (2011), Armesto et al. (2010), Clements and Galvão (2008) and Ghysels et al. (2006). Using the setting of MIDAS regressions, we would like to raise a few issues.

Consider two of the examples in the paper; the UK price index and UK travel abroad. Various high frequency series are available in both cases that help to predict the (low frequency) series of interest. For inflation, one might think of oil prices or other financial variables—available daily; see for example Monte-forte and Moretti (2013). For tourism, one can use Google search data (see for example Bangwayo-Skeete and Skeete (2015)), which are in principle available at any sampling frequency. As sampling frequency is part of the model choice in a MIDAS regression, we think that the reverse question, namely ‘Should we sample a time series less frequently?’ is as relevant as the question posed by the authors. In a data rich environment we can run regressions with daily predictors, or weekly, or monthly, etc. Storing high frequency data is more costly and cumbersome to handle. It is therefore useful to ask the question whether we should keep data at their original ultrahigh frequency or whether to predict low frequency objects of interest, such as macroeconomic entities, we should settle for some level of temporal aggregation. Ghysels et al. (2006) addressed this issue in the context of monthly volatility prediction. Ghysels et al. (2016) approached the same problem in a Bayesian setting. They constructed a set of candidate models, $\{M_1, \ldots, M_n\}$, with the explanatory variables sampled at different frequencies. In this framework, one can estimate the ratio of the marginal likelihoods of each pair of models (the so-called Bayes factors) $BF_{ij} = m(D | M_j) / m(D | M_i)$, $D$ being the data, and perform model selection based on the posterior probability of the models, obtained by using Bayes theorem from the marginal likelihood and the model prior probabilities. Note that we are also indirectly relying on the nowcasting literature; i.e. MIDAS regressions using high frequency data can be used for realtime updates of low frequency series—which might be a less costly solution compared with the increased univariate scheme considered by the authors (see Andreou et al. (2013)).

**Rebecca Killick and Ben Norwood** (Lancaster University)

Nason and his colleagues should be commended on an interesting paper exploring an area that is often overlooked.

They have applied their techniques to consumer price index and international passenger data available from the Office for National Statistics. An interesting extension of their work would be to consider a series such as gross domestic product that is made up from several underlying series potentially collected at different sampling rates. Under this example one may wish to consider increasing the sampling rate of one or more of the underlying series and the effects this would have on the estimation of gross domestic product. This is a specific example of a multivariate extension to their work where not all components have the same existing sampling rate and not all may benefit from moving to an increased sampling rate. Could the authors comment on whether their methodology could be generalized to this scenario and the potential challenges which might be encountered?

Underlying the authors’ decision rule is spectral estimation. The authors have described an approach for spectral estimation from multirate data but have considered all of this in the stationary context. It is
widely acknowledged within the Office for National Statistics that many of the time series they encounter have seasonal breaks or evolving seasonal structures and hence piecewise constant or evolving spectra. The estimation of the evolving multirate spectra is one challenge and the extension of the decision rule to non-stationary spectra is a second.

For estimating the spectrum, Nettheim (1965) suggested various methods to account for evolving seasonality, the chief contribution being a time varying Fourier spectrum estimated across a number of moving windows. These, or more recent locally stationary Fourier (Dahlhaus, 1997), basis functions could be used in place of the time invariant functions in the paper. Could the authors comment on the potential effect of using such a process on the overall variance of the forecasts, and similarly the effect of using the stationary assumption when in the presence of evolving seasonality.

Finally, the results surrounding the decision rule appear to hinge on the fact that the new information reduces the variance of the estimated coefficients (equation (12)). In the evolving spectra scenario this description is no longer valid as the new information does not give you more information about what you have already seen. Could the authors comment on whether they envisage that a decision rule taking into account evolving information would be achievable?

A. J. Lawrance (University of Warwick, Coventry)

My view is that this paper delivers original and useful research on a relevant methodology in time series analysis, and I congratulate Nason and his colleagues. The need to consider the effect of increasing or reducing the frequency of government or economic time series statistics is cogently made and can easily be appreciated from the authors’ interesting data sets. The authors’ methodology is soundly based on existing foundations of spectral analysis in time series, which are then attractively built on to allow mixed frequency data sets. The original work in the paper contains many useful technical insights and employs practically useful theoretical ideas, such as modifications of linear Bayes spectrum analysis, in judicious ways. Some of the technical aspects around the key estimation methodology may be beyond the likes of me to understand fully and to self-implement without a large amount of study. In mitigation, the announced provision of R code, hopefully with example runs, should make the work practically accessible, I hope so and would stress the importance of easily implemented and well-documented code. I particularly like the graphics, emphasizing statistical gains with their uncertainty shading, from using the methodology. The presentation of both spectral and auto-correlation analysis is appreciated. If there is to be a criticism of the methodology, its second-order nature might be mentioned. I make this from the point of view that somewhere, I am not sure where, the theory may rest in part on linear Gaussian assumptions. Particularly in prediction, this can be unsatisfactory. The authors do refer towards the end of Section 2 to the prediction of unobserved process values, both hindcasts and forecasts. I can see that these will be the same when based on second-order quantities, as the authors note by saying that the direction of time is unimportant. However, certainly for non-Gaussian non-linear time series models and many observed series, their stochastic parts are directional and this is important. The attempt in Section 3 at obtaining sampling strategies is laudable, but my own feeling is that practitioners will deal with the problem in more pragmatic ways. Nevertheless, the contribution is thought provoking, as is the rest of the paper.

Jorge Mateu (University Jaume I, Castellón) and Guillermo Ferreira (University of Concepción)

Nason and his colleagues are to be congratulated on a valuable contribution and a thought-provoking paper on the topic of sampling in the time series context, focusing their attention on the sampling rate implemented sequentially in time. The sampling rate on time series is an extremely interesting topic transversely involved in different areas of science, in particular, econometrics, social sciences and political science. The authors propose a methodology to decide whether it is suitable to move from a slow to a faster sampling rate, taking into account the costs that can be incurred to make this switch. This is carried out in the frequency domain of the time series, and thus spectrum estimation is essential. Our discussion is focused on spectral analysis and a possible extension of their spectral Bayesian estimator.

The spectrum of a time series is responsible for describing the distribution of the variance as a function of the frequencies. In the context of multivariate data, frequencies in different samples can be mistaken with each other; this fact is known as ‘aliasing’. To deal with this phenomenon, the authors propose a ‘folded spectrum’ at low sampling rates, which is defined as a function of the unfolded spectrum (or original spectrum) and frequencies that are aliased with frequencies in the whole spectral domain. To estimate the spectrum both at low and high sampling rates, the authors recommend first using high frequency followed by low frequency time series, employing a linearization of the folded spectrum through a first-order Taylor series expansion of the logarithm of the unfolded spectrum.
Discussion on the Paper by Nason, Powell, Elliott and Smith

However, the authors are missing a nice (and widely encountered) case in which the time series exhibit both long-range dependence and a cyclical behaviour. This phenomenon occurs in revenue series, quarterly UK inflation rates (as nicely described in Franses and Ooms (1997)), monetary aggregates and gross national product series, and in monthly river flow, among many other practical econometrical and environmental problems. A particular example of this type of processes is the seasonal auto-regressive fractional integrated moving average model, whose spectral density has one singularity at the origin and with possible poles at frequencies different from 0. This behaviour of the spectral density should be considered in the linear Bayesian estimator proposed by the authors for completeness of their approach.

Craig H. McLaren and Nick Vaughan (Office for National Statistics, Newport) Within an economic statistics context, there is a strong user demand to obtain estimates of the economy at an increased pace. The frequency and compilation of time series estimates are heavily influenced by cost but also on user needs and legislation. To meet user and policy needs the Office for National Statistics already publishes quarterly gross domestic product estimates on a month 1, month 2 and month 3 basis as additional data become available (Lee et al., 2015).

Published estimates typically involve the combination of complex statistical processes, e.g. treating missing observations; conceptual, coverage and alignment coverage adjustments, benchmarking, deflation, unchaining, chain linking, seasonal adjustment and aggregation. Quarterly data are often benchmarked and aligned to annual estimates which can impact the levels of the quarterly estimates. Analysis of quarterly estimates, as in this paper, would need to consider the effect of these processes, perhaps through the correlation model in this framework. For monthly estimates, trading day is often an important component. It was not clear how this component, and the other processing steps like chain linking, would be considered in this framework when moving from quarterly to monthly. The example given for the UK consumer price index is interesting but the conclusions unsurprising as this is typically stable, with no large movements. Some countries publish monthly and quarterly balance of payments. Comparing this method where published consistent monthly and quarterly data are available would be a useful illustration.

We have interpreted the process as similar to the application of splining. It would be good to expand on the effect of this sinusoid approach to ensure that it captures characteristics of real world estimates. For higher frequency outputs, the additional noise can mask underlying movements or turning points. Trend estimates for higher frequency indicators may be relevant. This framework could be expanded to consider whether quarterly outputs in a seasonally adjusted form should be reconsidered as monthly outputs in a trend form. Related to Steel (1997), McLaren and Steel (2000) considered alternatives to sample surveys, where significant savings are made by changing the rotation pattern but focusing on trend estimates.

The Office for National Statistics has a clear agenda to increase the use of administrative sources. Issues will arise in linking and aligning this to survey returns. Applications of this method could include issues such as assessment of subsampling of large administrative data sets, or as a diagnostic tool to verify the validity of administrative data before efforts are made to link or reduce survey collections.

Antonis A. Michis (Central Bank of Cyprus, Nicosia) First I begin by congratulating Nason and his colleagues for their extremely interesting paper. The subject of the paper is very important for organizations engaged in the compilation and analysis of time series data like national statistical institutes and central banks.

It is also very important for market research firms that provide retail and consumer measurement services to producers and manufacturers. Market research firms usually collect data with weekly, monthly or bimonthly frequency and the costs and benefits associated with each sampling rate are frequently under consideration. Similar considerations also exist for retail trade census surveys that are usually repeated every 5 years and are associated with large budgets.

As an economist working for a central bank, I am frequently interested in causal relationships between economic time series. However, causal relationships tend to be negatively affected (weakened) when data collection is not performed with the appropriate frequency. For this reason, the methodology that is proposed by the authors can contribute towards improving the accuracy and usefulness of statistical analysis with time series data.

A worthwhile topic for future research would be to extend the methodology that is described in the paper to temporal disaggregation problems. In these problems, statistical methods are used to disaggregate low frequency time series into compatible high frequency time series. The hindcasting application that is described by the authors provides a useful starting point for examining temporal disaggregation problems.
In applied work, temporal disaggregation is frequently achieved by incorporating in the analysis high frequency indicator variables that are known to exhibit sufficient correlation with the time series that are observed at lower frequencies. For example, one application can concern the use of monthly data on exports and imports to disaggregate quarterly sales data.

A variety of temporal disaggregation methods have been proposed in the literature. Most of these methods are based either on regression models that make use of indicator variables observed at higher frequencies (see for example Chow and Lin (1971)), or on purely time series models that do not make use of any indicator variables (see for example Wei and Stram (1990)). It would be interesting to analyse the temporal disaggregation problem by using high frequency indicator variables in the context of the linear Bayes spectral method that is described in the paper. It can potentially convey superior frequency domain information in the high frequency time series generated.

Anastasia Papavasiliou (University of Warwick, Coventry) and Hao Ni (University College London)

We thank Nason and his colleagues for bringing to the discussion such an important question and we congratulate them on their innovative approach.

In this note, we focus not on the methodology but on the way that the question is posed. Sampling at very fine timescales is just one way of describing complex data streams and can be very inefficient, in particular if we are interested in its effects when driving a process. An alternative non-linear, higher order description for data streams is the so-called signature of a stream, which originated in the theory of rough paths. In that context, one would ask how often and, most importantly, up to what level should one sample the signature of the data stream?

For example, Clark and Cameron (1980) studied the numerical approximation of a solution to a multi-dimensional stochastic differential equation and provided a lower bound for the approximate solution at time $T = 1$ (an effect) via sampling the driving signal (stream) at $n$ deterministic locations in $[0,1]$. As a generalization of this result, Dickinson (2007) (proposition 17) showed that any $n$-dimensional linear functional of the stream is a similarly poor predictor. By combining linear regression with the signature feature set, we can achieve much better prediction results (see Levin et al. (2016)).

Here is a simple example, aiming to demonstrate why the signature can play an important role in the question of sampling. The second level of the log-signature of a one-dimensional stream $X_t$ and its delay, namely the area, is defined as

$$\int_t^{t+T} \int_t^u dX_t dX_{u-} - \int_t^{t+T} \int_t^u dX_{u-} dX_u.$$ 

A simple computation shows that high frequencies, even at low amplitudes, can have a significant contribution to the area. In other words, a fast growing area suggests the need to sample more often. Constructing an adaptive sampling scheme based on the area leads to a description of the data that implicitly contains information about the first two levels of the signature. Alternatively, one can keep information about both increments and area at a coarser homogeneous grid, as a way to describe the frequency content between sampling points.

To summarize, we believe that in some cases the concept of the signature of a data stream can be a powerful tool for describing the data and is worth further investigation within the context of statistics.

Dongho Song (Boston College)

Nason and his colleagues have provided us with an insightful and stimulating paper. The paper has an extensive discussion on how to use data sampled at multiple rates and shows through various exercises that there are gains to combining information from higher and lower sampling rates. Several empirical exercises presented in this paper are revealing in this regard.

I would like to complement the analysis by observing that there are greater gains to using mixed frequency time series data if one is concerned with volatility estimation. Schorfheide et al. (2016) have shown in a time domain approach that the estimation with mixed frequency data, e.g. monthly and annual data, leads to improvements in the conditional mean or trend estimates (consistent with this paper) and also in the conditional volatility estimates (as shown by tightening the credible intervals). Nason and his colleagues motivate the use of spectral density and I am left wondering how much of the current method can be extended to volatility estimation if the true data-generating process has stochastic volatility.

The other comment is that some data, e.g. consumption data in the USA, are singular although they are available at monthly, quarterly and annual frequency. It is costly for the statistical agency to conduct a comprehensive survey at higher frequency so they often interpolate the lower frequency data, e.g. annual
data, to quarterly or monthly by using (potentially ill-measured) proxy series. Thus, higher frequency series aggregate perfectly to lower frequency by definition, which makes the joint use of data measured at different frequencies impossible. The common trade-off that researchers face is either to use more frequent but noisy information or to use less frequent but cleaner information. Example 5 (UK consumer price index) in Section 4.2 does not apply the temporal aggregation restriction. What is the authors’ stance on this?

Milan Stehlik and Jean Paul Maidana (Johannes Kepler University in Linz and University of Valparaiso)

We congratulate Nason and his colleagues for introducing readers to the challenges of time series sampling frequency.

The application of discrepancy statistics $d_m$ is the main point addressed in this discussion. In Table 1, the superiority of the new linear Bayes estimate is shown. However, one should point out that these kinds of result depend on the choice of the proper distance measure (see Kanamori and Sugiyama (2014)), and the

![Fig. 16. For each value of $N_{\text{high}}$, 5000 independent simulations from an ARMA(3,1) model were generated with the TSA package of the Comprehensive R Archive Network; each of these simulations were used to estimate the spectral density with the three proposed methods (a) least squares spectral analysis, (b) maximum likelihood ARMA and (c) linear Bayes; a comparison with the theoretical spectral density with four discrepancy measures is given](image-url)
### Table 4. Mean values of each measure of discrepancy for 5000 independent simulations

| $N_{\text{high}}$ | $d_N$  | $d_{S_1}$ | $d_{S_2}$ | $d_{S_3}$ |
|-------------------|--------|-----------|-----------|-----------|
|                   | Least squares spectral analysis | Maximum likelihood | Linear Bayes | Least squares spectral analysis | Maximum likelihood | Linear Bayes | Least squares spectral analysis | Maximum likelihood | Linear Bayes |
| 8                 | 0.569  | 0.576     | 0.047     | 121.587   |
| 16                | 0.566  | 0.573     | 0.047     | 121.032   |
| 32                | 0.547  | 0.562     | 0.046     | 118.589   |
| 64                | 0.525  | 0.550     | 0.045     | 115.855   |
| 128               | 0.491  | 0.529     | 0.044     | 111.388   |
| 256               | 0.450  | 0.504     | 0.042     | 105.908   |

Discussion on the Paper by Nason, Powell, Elliott and Smith
choice of distance in the paper has the form of φ-divergence. For some remarks on the relationship between
φ-divergences and statistical information (e.g. expressed by Bayes risks) see Stehlik (2012). However, an
informative choice of proper distance measure can change the position of illustrated methods.

Several methods for spectral density estimation have been proposed; in this comparison we use least
squares spectral analysis, maximum likelihood from an auto-regressive moving average (ARMA) model
and the linear Bayes method proposed by Nason and his colleagues.

To compare these three methods, we simulate time series from an ARMA(3,1) model with auto-regressive
parameters \( \phi = (-0.5, 0.4, 0.8) \) and \( \theta = 0.2 \) for the moving average part with the TSA package from the
Comprehensive R Archive Network; this time series is composed of 256 observations at even integer
points (\( N_{\text{slow}} \)) followed by an unthinned part (\( N_{\text{high}} = 8, 16, 32, 64, 128, 256 \)). For each of these \( N_{\text{high}} \)-values
1000 independent samples from the ARMA(3, 1) model were simulated with the \texttt{fArima} package of the
Comprehensive R Archive Network. Finally to compare the different spectral density estimates from the
three methods proposed we use four measures for the discrepancy between the theoretical spectral density
and proposed density. The distances are

\[
d_{N,m} = N^{-1} \sum_{j=1}^{Nw} \left[ \log \{ f(w_i) \} - \log \{ \hat{f}(w_i) \} \right]^2
\]

and

\[
d_{S,m} = N^{-1} \left( \sum_{j=1}^{Nw} \left[ \log \{ f(w_i) \} - \log \{ \hat{f}_m(w_i) \} \right] \right)^{1/8}, \quad i = 1, 2, 0.5, \quad m = 1, \ldots, 5000,
\]

where \( \hat{f}_m(w_i) \) is the point estimate of \( f_m(w_i) \) for simulation \( m \) and \( \{w_i\} \) are 256 equally spaced points
between 0 and 0.5.

In Table 4 and Fig. 16 it can be seen that least squares spectral analysis surprisingly is the better ap-
proximation for the spectral density for each of the four measures of discrepancy and smaller values of
\( N_{\text{high}} \); conversely the linear Bayes and maximum likelihood ARMA models are better for long values of
\( N_{\text{high}} \).

A. Stein (University of Twente, Enschede)
I read the paper with much pleasure. Sampling is important and sometimes costly, and any step forward in
optimizing the sample size is to be welcomed. For the time series data that are considered in the
paper, working with spectra and periodograms makes sense. However, many times series come at irregular
intervals, sometimes even continuously, and other time series have serious gaps for sometimes unclear
reasons. It may be a challenge for the methodology presented to provide a clear solution for such cases. In
fact, in the data that I often use, the problem of the gaps is the most serious. A strong point of the research
is that there is attention to trends. As indicated at the start of the paper, we should sample more intensively
where changes are to be expected. The emphasis on seasonality is thus somewhat specific.

Of some interest would be whether the methods presented can be easily extended towards two-dimensional
spatial sampling, or even towards combined spatiotemporal sampling. In spatial sampling the issue of sea-
sonality or other forms of periodicity is not so strong. Maybe for that reason, spectral solutions are not
so prominent in spatial sampling, despite good and serious attempts. There is more attention to statistical
(design-based) sampling schemes versus non-statistical optimal (model-based) sampling schemes where
optimality is defined on the basis of stakeholder interests. If the two dimensions are seemingly indepen-
dent, like hourly temperatures and monthly travel trips abroad, then I could imagine that relatively simple
two-dimensional multirate sampling can be defined. However, when the two are closer related, like two
co-ordinates, or minimum and maximum average monthly temperatures, then the challenge may be bigger.
Further, spatiotemporal sampling is becoming quite prominent in the abundance of satellite images that
are collected at regular intervals from the same area. Those may contain clouds, and in the evening not
so much is seen: hence the problem of dealing with gaps is rather critical. Possibly, the authors have some
clear ideas on how to proceed in such circumstances or may find it promising to elaborate on it in their
future research.

Adam M. Sykulski (University College London)
The method proposed for time series sampled at variable rates is an important innovation in spectral
analysis, for which I congratulate Nason and his colleagues. I can see immediate application benefit in
the field of oceanography with data collected from freely drifting satellite-tracked instruments. Recently
Discussion on the Paper by Nason, Powell, Elliott and Smith

Fig. 17. Spectral density estimates generated from an AR(4) process given by $X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9238X_{t-4} + \epsilon_t$ where $\{\epsilon_t\}$ is a Gaussian white noise process with mean 0 and unit variance, as studied in Percival and Walden (1993), page 46 (the Slepian taper is applied using dpss from the R package multitaper, with bandwidth parameter set to 4, and the time series used is perci- valAR4 from the same package): (a) estimates from 1024 observations sampled every integer point (——, true AR(4); ——, regspec; ———, periodogram; ————, regspec (tapered); ————–, tapered spectrum); (b) estimates from combining 256 observations sampled every three integer points, and 32 observations sampled every integer point (——, true AR(4); ——, regspec; ———, regspec (tapered))

1-hourly interpolated data sets since 2005 were made available by the Global Drifter Program (Elipot et al., 2016), which can now be combined naturally with 6-hourly data going back to 1979.

I would like to discuss the paper’s use of the periodogram in forming spectral density estimates. It is well known that the periodogram is often a badly biased estimate of the process’s true spectral density; see for example Dahlhaus (1988) and Sykulski et al. (2016). This is due to the leakage caused by the edge effects of truncating a sequence to finite length. A commonly proposed solution is to apply a data taper to ameliorate edge effects (Percival and Walden, 1993). Here, I shall investigate whether tapering can be effectively combined with the method of Nason and his colleagues.

In the analysis of Fig. 17, I have applied a Slepian taper (Slepian, 1978) to the time series; before then applying the authors’ R package regspec to obtain spectral estimates. I have used the classical AR(4) process studied in Percival and Walden (1993). This process has a larger dynamic range in the spectrum than the auto-regressive moving average ARMA(3,1) process studied by Nason and his colleagues, and thus has greater bias from leakage. In Fig. 17(a) I examine various spectral density estimates of a regularly sampled series of length 1024. The basis of the periodogram at high frequencies is apparent, and this has been carried over to the regspec estimate. However, after applying a Slepian taper (of bandwidth 4) to the time series, then the direct tapered spectral estimate—as well as the regspec estimate—is no longer badly biased at high frequencies.

It is now interesting to see whether tapering can be combined with regspec with variable sampling rates. In Fig. 17(b), I apply the Slepian taper to 256 observations from the AR(4) process sampled every three integer points, followed by 32 observations sampled every integer point. The high frequencies are now significantly more difficult to resolve; but regspec when combined with tapering is still considerably less biased than if no taper were applied. The ability of the method to benefit naturally from data tapers for such processes is a highly desirable and useful feature.

Alexander Vandenberg-Rodes and Hernando Ombao (University of California at Irvine)

This paper highlights an important point: statisticians often treat the sampling rate as given, leaving us vulnerable to aliasing artefacts if the underlying signal has significant power at frequencies above the sampling frequency. We share two comments. First, we draw the connection between the proposed linear Bayes methodology and classical filtering theory. Second, sampling at higher rates that are not integer multiples of the slow sampling rate can have advantages.

(a) Treat the vector $\beta$ (of basis coefficients $\beta_k$) as the hidden state and the log-periodograms $I_T$ and $J_R$ as the noisy observations. Under ‘fast sampling’ rewrite equation (7)
Discussion on the Paper by Nason, Powell, Elliott and Smith

\[ I_T = c_{\text{fast}} + B\beta + e_{\text{fast}}, \]

where \( B \) is the matrix of basis functions. The ‘adjusted’ expectation and variance of \( \beta \), in equations (11) and (12), are then one step of the classical (linear Gaussian) Kalman filter when \( \beta \) is constant.

Note that aliasing results in a non-linear observation equation

\[ J_R = c_{\text{slow}} + h(\beta) + e_{\text{slow}}, \]

where \( h(\cdot) \) is a log-sum-exponential function. Classical Kalman filtering does not apply and hence the authors propose a linearization. The Bayesian update for \( \beta \) is one step of the extended Kalman filter. Furthermore, recomputing the linearization around the new posterior expectation for \( \beta \) is the iterated extended Kalman filter. Viewed in this new light, other methods for non-linear state space model, such as unscented filtering, assumed density filtering and particle filtering (Särkkä, 2013; Douc et al., 2014) might produce better approximations to the full posterior of \( \beta \), when compared with the extended Kalman filter.

(b) It might be advantageous to choose a ‘high’ sampling rate that is a non-integer multiple of the ‘low’. When sampling a continuous time signal at 1 Hz and 0.5 Hz, any frequencies in the band from 0.75 to 1.25 Hz will have identical aliasing artefacts in the band from 0 to 0.25 Hz under both sampling rates. When sampling at 0.6 Hz (high) and 0.5 Hz (low), then the only frequencies \( \omega \) that alias to the same frequency in the band from 0 to 0.25 Hz under both high and low sampling schemes solve \( \omega - 0.5k = (\omega - 0.6l) \) for positive integers \( k \) and \( l \). This is related to multi-coset sampling (Mishali and Eldar, 2009), which allows exact reconstruction of deterministic signals with support on (unknown) frequency bands, despite sampling significantly below the Nyquist rate.

The authors replied later, in writing, as follows.

We thank all discussants for their incisive and well-considered comments, it was particularly pleasing to see that the sampling rate question is one that provokes such wide-ranging interest from theory into applications. For brevity we have categorized our response and we apologize in advance for not being able to respond directly to all discussants.

The wide range of concepts that are relevant to the main paper’s content (aliasing, spectral estimation, sample rate, cost function and fusion) meant that many potential avenues of investigation had to be ignored. Hence, we focused closely on univariate time series using a popular spectral estimate, an efficient inference engine and streamlined cost assessment. We are grateful to the discussants, who have highlighted particularly important avenues and suggested extensions or modifications to the choices that we made, which we now summarize.

Applications to other areas
It was heartening to learn that many of the ideas in the main paper would resonate with so many interesting application areas: big data and data mining (Triantafyllopoulos), functional magnetic resonance imaging and electroencephalography, econometrics (Yu and Cribben), long memory time series (Knight and Nunes), market research (Michis), ecology (Butler, Searle and Daunt), spatial statistics (Stein and Diggle) and oceanography (Sykulski) and we look forward to developments in these areas. Song raises the possibility of developing the methodology to estimate volatility, which we would be interested to see.

For big data it seems natural to select an appropriate sampling interval to reduce the size of the data to something more manageable. In the official statistics context, this is likely to be an important component of dealing with large administrative data sets, as suggested by McLaren and Vaughan. Such large data sets are becoming increasingly important in official statistics and working out how best to gather the information in them at low cost is critical.

Different kinds of data
Many discussants suggested scenarios where our fusion methodology might find application after substantial customization. Perhaps the most important scenario to consider, as identified by Knight and Nunes, and Killick and Norwood, is the analysis of multivariate series. With the latter’s allusion to gross domestic product statistics, which are composed from a range of inputs, they hint at the considerable contextual complexities that sit on top of the technical challenges that are involved here. McLaren and Vaughan also address the challenge that component time series in gross domestic product may not contribute their properties directly but may be modified by post-processing, such as balancing (which may depend on many
series and have no analytical representation). Without doubt, however, the potential benefits of developing multivariate methodology are huge.

Separately, Yu and Cribben, and Michis mention the estimation of causal relationships from time series data, and how this might be improved with higher sampling rates. This is a fascinating direction in which to take this work, which we had not anticipated.

Both Triantafyllopoulos and Eckley separately raise the question of whether our methodology could be used for data living in more exotic locations than the real line, such as count data. We believe that the answer is ‘yes’ in principle but substantial work would be required to develop the details. In the international migration example the treatment of the data as continuous has little practical import since the values are large. However, as the sampling rate increases the counts will inevitably reduce and more specialized methods might be required.

**Different sampling schemes**

Our proposed sampling scheme was the scheme that seemed most natural and fitted well with current practice in official statistics. However, several discussants queried whether, or suggested that, our methodology could be used for different sampling schemes. For example, Olhede mentions sampling taken over a period of time rather than instantaneously and presumably this also extends to many variants, such as non-linear sampling functions. Olhede mentions non-regular sampling as a way of bypassing the aliasing problem, but in many applied situations this might not be completely practicable, e.g. persuading a government department to produce another value for international passenger numbers minutes after the last one! However, for some data sets, such as Web scraping of price data, irregular sampling might be possible and include high frequency sampling as necessary. Diggle also advocates mixed regular and irregular sampling in the spatial context which should be further explored for the time series case. The multi-coset sampling (Vandenberg-Rodes and Ombao) and the alternative of the stream signature (Papavasilou and Ni) both sound fascinating and it would be good to learn more. In particular, how does the stream signature relate to the gamut of existing statistical and signal processing sampling theory, including compressed sensing ideas, and would it also be useful for analysing small data sets?

**Improved estimators of the spectrum and related models**

Many contributors rightly commented on the fact that several estimators exist that are better than the simple smoothed periodogram (Olhede, Craigmile, Tunnicliffe Wilson and Cox). Sykulski and Tunnicliffe Wilson’s contributions both provide detailed examples of the kind of improvement that can be achieved with multitapering and Fourier methods with Hannan–Quinn order selection respectively. Our defence here, though, is pedagogical in that we wished for the description of our methodology to be widely accessible and some of these other estimators, although demonstrably better, would have added an extra layer of complexity. However, we think that many of the improved estimators could be ‘plugged into’ our scheme without too much effort and we should have at least admitted the possibility in the paper.

Olhede, Craigmile and Fokianos highlight the connection between our work with the cepstrum and exponential models and provide a valuable list of references to that body of work. Ghosal and Roy suggest using a different, but related, model which would benefit from further investigation in this context.

**Priors**

Of course, priors are important and ours were selected to be of the auto-regressive integrated moving average (ARIMA) or seasonal ARIMA type due to the context of the problem at hand. However, we welcome the suggestion of Mateu and Ferreira who suggest the seasonal auto-regressive fractional integrated moving average model which would also be useful in many real circumstances. In this area Yao wonders whether our choice of prior means that we smoothed too much. We should say that the prior was selected to be in tune with our prior beliefs, although Yao’s point is well made and further exploration of prior sensitivity should be carried out. Olhede mentions that the plain Wahba priors might not work well for signals with strong periodicities and it is for that reason that we used the seasonal ARIMA priors.

**Long memory, non-stationary and non-Gaussian series**

Knight and Nunes provide a lovely example of our methodology applied to a long memory series; clearly if the spectral estimation works then the associated machinery should be able to do a reasonable job. Of course, long memory estimation focuses on the low frequencies whereas our initial story looks to improve estimation of the higher frequencies and so their example is highly apposite. It would be interesting to consider a theoretical analysis to investigate whether and how a combination of sampling rates can
be utilized effectively for long memory estimation, which could be of great utility, for example, in risk assessment problems where one has little access to samples across large periods of time.

Eckley, and Killick and Norwood separately wonder how our methods could apply to the tricky, but practically important, case of non-stationary (including locally stationary) time series. Killick and Norwood suggest looking at windowed approaches which might be a feasible way forward. However, there is, maybe, a conceptual difficulty with non-stationary series; our methodology is concerned with spectral fusion from different time points and if the spectra are different in the two locations then fusion does not make sense. Worse, you might not have enough information in either of the two locations to discern whether the two spectra are different (although the methods that are suggested by Fokianos might help here). Of course, if the spectrum was evolving sufficiently slowly it might work (but again, how to tell?): recent work on measuring non-stationarity by Das and Nason (2016) might help here) or if fusion was achieved contemporaneously by fusing multivariate series. Also, estimating non-stationarity often requires much more data than we are used to in stationary series problems and, for the fast sampled data envisaged in the main paper, there are not usually enough observations.

Lawrance’s point about the second-order nature of our work is well made. In particular, consideration would have to be given to situations outside the linear Gaussian framework although, again, much rests on the success of the second-order spectrum estimation and, for more general data, it might be possible, in theory at least, to consider the behaviour of higher order spectra.

Other improvements
We use linear Bayes methods as our inference engine but other, maybe full Bayesian, methods would improve performance but, we suspect, be much slower (Triantafyllopoulos, and Vandenberg-Rodes and Ombao) and/or more difficult to manage as the paper mentions. Yu and Cribben investigate the ordering of the slow–fast sampling rates and we are grateful for their extended analysis which sheds light on good orderings. This idea might be taken further theoretically to suggest ‘optimal’ combinations of sampling schemes in different scenarios. We agree with Ghosal and Roy that the $M$-parameter, the number of basis functions, is a key parameter and more thought should be given to its selection, ensuring that it increases with increasing sample size. Another possibility would be to incorporate it into the Bayesian scheme more formally as they suggest. Likewise we are keen on their idea of not relying on one mode but averaging results over a set of alternative possible modes. Stehlik and Maidana’s assessment of spectral estimates using different measures of spectral discrepancy is another valuable contribution to this topic.

We were also interested by the suggestion of Michis who proposes disaggregating quarterly sales data by using related monthly exports data, particularly as a way to aid hindcasting. Song also mentions that this technique is used by statistical agencies guided though by potentially ill-measured proxy series. The success of these procedures will depend on the quality and relevance of the proxy, but a high quality rapidly sampled proxy could be used in our work and indeed this is related to the value-added tax example that is mentioned in Section 5.2 of the paper.

Sampling less frequently
Elayouty, Scott and Miller advocate aggregation when the time series might be subject to short-term transient events; the aggregation locally smooths out the effects of such transients. Our view of this depends on what the transients are. If they are infrequent then maybe they are outliers and may be removed from the analysis. If they are frequent then they are part of the normal stochastic structure of the signal and classed as signal information and a proper component of the spectrum. In the latter case aggregation might result in an unwarranted loss of information. However, Elayouty, Scott and Miller are clear that the action to take depends on the purpose of the analysis: we fully agree with this. Ghysels, Solgi and Mira also point to the usefulness of high sampled frequency signals to assist in the analysis of low sampled frequency signals. This view reflects the very fundamentals of statistics in providing a low dimensional and useful summary of information from a signal and, of course, we are in complete agreement with this philosophy.

The above discussion promotes the idea of adaptive sampling, i.e. adapt the sampling rate to match the information content in the signal. Our contribution here would be to assess and control the costs that are associated with that scheme. Such ideas, based on the information content only, were proposed by, for example, Hall and Penev (2004).

Cost function
Craigmile makes the good suggestion to use the cost function to take account of seasonal effects: our cost function is clearly presented as a suggestion as one of the many contextually sensible things to do. Ahmad asks whether we can provide more detailed information on the costs involved. In general, this is difficult
to do. Partly this is because not all of the costs are easily identifiable but also, with the exception of rare detailed studies, because of the difficulty that is faced by national statistical offices in delineating the costs to themselves and to the users of their statistics in a way which would facilitate the estimation of the cost of policy change.

The main paper discussed the Labour Force Survey where some of this detail was estimated in Steel (1996), including the one-off cost of the change; Penneck et al. (1993) estimated some of the costs of individual elements of business surveys, but not the one-off cost of a change in periodicity. Elayouty, Scott and Miller confirm that cost function identification is often not straightforward although proxies and approximations are often useful as the paper shows. Butler, Searle and Daunt remind us that different cost functions, e.g. relating to the total number of observations rather than their frequency, are often merited.

For the longer term
Several ideas in the paper and, particularly, the discussion are open for future development. It would be interesting to see how our methodology might

(a) open the door to understand Granger causality in functional magnetic resonance imaging data (Yu and Cribben),
(b) be utilized to identify different spectral densities (Fokianos) and
(c) show how Kiefer–Wolfowitz optimal design theory (Cox) could be utilized in our situation.

Meng, in his vineyard, presents an intriguing different slant on a cognate problem: what sampling rate is best for certain parameters? The analysis is compelling but is it easily generalizable to more complex models, and how well can it do when parameters are also estimated? In any case, more viticulture is probably required to turn all these methods into celebrated vintages.

References in the discussion

Andreu, E., Ghysels, E. and Kourtellos, A. (2011) Forecasting with mixed-frequency data. In Oxford Handbook of Economic Forecasting (eds M. P. Clements and D. F. Hendry), pp. 225–245. Oxford: Oxford University Press.
Andreu, E., Ghysels, E. and Kourtellos, A. (2013) Should macroeconomic forecasters use daily financial data and how? J Bus. Econ. Statist., 31, 240–251.
Armesto, M. T., Engemann, K. M. and Owyang, M. T. (2010) Forecasting with mixed frequencies. Fed. Resrv. Bnk St Louis Rev., 92, 521–536.
Bangwayo-Skeete, P. F. and Skeete, R. W. (2015) Can Google data improve the forecasting performance of tourist arrivals?: Mixed-data sampling approach. Toursm Mangmnt, 46, 454–464.
Bartlett, M. S. and Kendall, D. G. (1946) The statistical analysis of variance-heterogeneity and the logarithmic transformation. J. R. Statist. Soc., suppl., 8, 128–138.
Beran, J. (1993) Fitting long-memory models by generalized linear regression. Biometrika, 80, 817–822.
Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013) Long-memory Processes. New York: Springer.
Blackman, R. and Tukey, J. (1958) The Measurement of Power Spectra. New York: Dover Publications.
Bloomfield, P. (1973) An exponential model for the spectrum of a scalar times series. Biometrika, 60, 217–226.
Bloomfield, P. (2000) The Fourier Analysis of Time Series, 2nd edn. New York: Wiley.
Bogart, B. P., Healy, M. J. R. and Tukey, J. W. (1963) The frequency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and safe-cracking. In Proc. Symp. Time Series Analysis (ed. M. Rosenblatt), pp. 209–243. New York: Wiley.
Brillinger, D. R. (1981) Time Series: Data Analysis and Theory. Holt: Wiley.
Broersen, P. M. T. (2007) Time series models for spectral analysis of irregular data far beyond the mean data rate. Measurn Sci. Technol., 19, 1–13.
Broersen, P. M. T., De Waele, S. and Bos, R. (2000) The accuracy of time series analysis for laser-Doppler velocimetry. 10th Int. Symp. Application of Laser Techniques to Fluid Mechanics.
Chipeta, M. G., Terlouw, D. J., Phiri, K. S. and Diggle, P. J. (2016) Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. Environmetrics, to be published.
Chow, G. C. and Lin, A.-L. (1971) Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. Rev. Econ. Statist., 53, 372–375.
Clark, J. and Cameron, R. (1980) The maximum rate of convergence of discrete approximations for stochastic differential equations. In Stochastic Differential Systems Filtering and Control, pp. 162–171. Berlin: Springer.
Clements, M. P. and Galvño, A. B. (2008) Macroeconomic forecasting with mixed-frequency data: forecasting output growth in the United States. J. Bus. Econ. Statist., 26, 546–554.
Coates, D. S. and Diggle, P. J. (1986) Tests for comparing two estimated spectral densities. J. Time Ser. Anal., 7, 7–20.
Cox, D. R. and Hinkley, D. V. (1968) A note on the efficiency of least squares estimates. J. R. Statist. Soc. B, 30, 284–289.
Discussion on the Paper by Nason, Powell, Elliott and Smith

Dahlhaus, R. (1988) Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Ann. Statist.*, 16, 808–841.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Ann. Statist.*, 25, 1–37.

Das, S. and Nason, G. P. (2016) Measuring the degree of nonstationarity of a time series. *Stat*, 5, in the press.

Dickinson, A. S. (2007) Optimal approximation of the second iterated integral of Brownian motion. *Stoch. Anal. Appl.*, 25, 1109–1128.

Diggle, P. J. (1990) *Time Series*. New York: Oxford University Press.

Diggle, P. J. and al Wasel, I. (1997) Spectral analysis of replicated biomedical time series (with discussion). *Appl. Statist.*, 46, 31–71.

Douc, R., Moulines, E. and Stoffer, D. (2014) *Nonlinear Time Series: Theory, Methods and Applications with R Examples*. Boca Raton: CRC Press.

Eckley, I. A. and Nason, G. P. (2014) Spectral correction for locally stationary Shannon wavelet processes. *Electron. J. Statist.*, 8, 184–200.

Eckley, I. A. and Nason, G. P. (2016) A test for the absence of aliasing or local white noise in locally stationary wavelet time series. To be published.

Elipot, S., Lumpkin, R., Perez, R. C., Lilly, J. M., Early, J. J. and Sykulski, A. M. (2016) A global surface drifter data set at hourly resolution. *J. Geophys. Res. Oceans*, 121, 2937–2966.

Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P. and Calabrese, J. M. (2015) Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. *Ecology*, 96, 1182–1188.

Fokianos, K. and Savvides, A. (2008) On comparing several spectral densities. *Techometrics*, 50, 317–331.

Frances, P. and Ooms, M. (1997) A periodic long-memory model for quarterly UK inflation. *Int. J. Forecast.*, 13, 117–126.

Fuller, W. A. (1996) *Introduction to Statistical Time Series*, 2nd edn. Hoboken: Wiley.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004) The MIDAS touch: mixed data sampling regression models. *Finance*.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006) Predicting volatility: getting the most out of return data sampled at different frequencies. *J. Econometr.*, 131, 59–95.

Ghysels, E., Solgi, R. and Mira, A. (2016) Bayesian frequency selection for MIDAS regression. *Working Paper*. Hall, P. and Penev, S. (2004) Wavelet-based estimation with multiple sampling rates. *Ann. Statist.*, 32, 1933–1956.

Hannan, E. J. and Quinn, B. G. (1979) The determination of the order of an autoregression. *J. R. Statist. Soc. B*, 41, 190–195.

Hansson, L.-A. and Åkesson, S. (eds) (2014) *Animal Movement across Scales*. Oxford: Oxford University Press.

Hyndman, R., Koehler, A. B., Ord, J. K. and Snyder, R. D. (2008) *Forecasting with Exponential Smoothing: the State Space Approach*. Berlin: Springer Science and Business Media.

Jones, R. H. (1981) Fitting a continuous time autoregression to discrete data. In *Applied Time Series Analysis*, vol. II (ed. D. F. Findley), pp. 651–682. New York: Elsevier.

Kanamori, T. and Sugiyama, M. (2014) Statistical analysis of distance estimators with density differences and density ratios. *Entropy*, 16, 921–942.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. and Morales, J. M. (2012) Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93, 2336–2342.

Lee, P., McCrae, A., Denley, H. and Osborn, E. (2015) *A Short Guide to the UK National Accounts*. Newport: Office for National Statistics.

Levin, D., Lyons, T. and Ni, H. (2016) Learning from the past, predicting the statistics for the future, learning an evolving system. *Preprint arXiv:1309.0260*.

Lomb, N. (1976) Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.*, 39, 447–462.

Masry, E. (1978) Alias-free sampling: an alternative conceptualization and its applications. *IEEE Trans. Inform. Theory*, 24, 317–324.

McLaren, C. H. and Steel, D. G. (2000) The impact of different rotation patterns on the sampling variance of trend estimates. *Surv. Methodol.*, 26, 163–172.
Meng, X.-L. and Xie, X. (2014) I got more data, my model is more refined, but my estimator is getting worse: Am I just dumb? *Econometrica*, **33**, 218–250.

Mishali, M. and Elder, C. (2009) Blind multiband signal reconstruction: compressed sensing for analog signals. *IEEE Trans. Signal Process.*, **57**, 993–1009.

Moir, T. J. and Barrett, J. F. (2003) A kepsrum approach to filtering, smoothing and prediction with application to speech enhancement. *Proc. R. Soc. Lond. A*, **459**, 2957–2976.

Monteforte, L. and Moretti, G. (2013) Real-time forecasts of inflation: the role of financial variables. *J. Forecast.*, **32**, 51–61.

Nettheim, N. F. (1965) Fourier methods for evolving seasonal patterns. *J. Am. Statist. Ass.*, **60**, 492–502.

Niu, M., Blackwell, P. G. and Skarin, A. (2016) Modelling interdependent animal movement in continuous time. *Biometrics*, **72**, 315–324.

Parzen, E. (1993) Stationary time series analysis using information and spectral analysis. In *Developments in Time Series Analysis: in Honour of M. B. Priestley* (ed. T. S. Rao), pp. 139–148. London: Chapman and Hall.

Pourahmadi, M. (1984) Taylor expansion of $\exp(\sum_{k=0}^{\infty} a_k z^k)$ and some applications. *Am. Math. Mnthly*, **91**, 303–307.

Robinson, P. M. (1994) Semiparametric analysis of long-memory time series. *Ann. Statist.*, **22**, 515–539.

Robinson, P. M. (1995) Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, **23**, 1630–1661.

Särkkä, S. (2013) *Bayesian Filtering and Smoothing*. Cambridge: Cambridge University Press.

Savvides, A., Promponas, V. J. and Fokianos, K. (2008) Clustering of biological time series by cepstral coefficients based distances. *Pattern Recogn.*, **41**, 2398–2412.

Scargle, J. (1982) Studies in astronomical time series analysis: II—Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.*, **263**, 835–853.

Schorfeide, F., Song, D. and Yaron, A. (2016) Identifying long-run risks: a Bayesian mixed-frequency approach. *Manuscript*. Boston College, Boston.

Shen, W. and Ghosal, S. (2015) Adaptive Bayesian procedures using random series priors. *Scand. J. Statist.*, **42**, 1194–1213.

Slepian, D. (1978) Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case. *Bell Syst. Tech. J.*, **57**, 1371–1430.

Steel, D. (1996) *Options for Producing Monthly Estimates of Unemployment According to the ILO Definition*. London: Central Statistical Office.

Steel, D. (1997) Producing monthly estimates of unemployment and employment according to the International Labour Office definition (with discussion). *J. R. Statist. Soc. A*, **160**, 5–46.

Stehlik, M. (2012) Decompositions of information divergences: recent development, open problems and applications. *AIP Conf. Proc.*, **1493**, 972–976.

Sykulski, A. M., Olhede, S. C., Lilly, J. M. and Early, J. J. (2016) The de-biased Whittle likelihood for second-order stationary stochastic processes. *Preprint arXiv:1605.06718*.

Thomas, D. J. (1982) Spectrum estimation and harmonic analysis. *Proc. IEEE*, **70**, 1055–1096.

Tunnliciffe Wilson, G., Reale, M. and Haywood, J. (2015) *Models for Dependent Time Series*. New York: CRC Press.

Vaníček, P. (1971) Further development and properties of the spectral analysis by least-squares. *Astrophys. Space Sci.*, **12**, 10–33.

Wahba, G. (1980) Automatic smoothing of the log periodogram. *J. Am. Statist. Ass.*, **75**, 122–132.

Walden, A. T., Percival, D. B. and McCoy, E. J. (1998) Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Trans. Signal Process.*, **46**, 3153–3165.

Wei, W. W. S. and Stram, D. O. (1990) Disaggregation of time series models. *J. R. Statist. Soc. B*, **52**, 453–467.

Wolfe, P. J., Godsill, S. J. and Ng, W.-J. (2004) Bayesian variable selection and regularization for time-frequency surface estimation. *J. R. Statist. Soc. B*, **66**, 575–589.

Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. B*, **68**, 49–67.