Spectral correction for locally stationary Shannon wavelet processes[*]

Idris A. Eckley

Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK
e-mail: i.eckley@lancs.ac.uk

and

Guy P. Nason

School of Mathematics, University Walk, University of Bristol, BS8 1TW, UK
e-mail: g.p.nason@bris.ac.uk

Abstract: It is well-known that if a time series is not sampled at a fast enough rate to capture all the high frequencies then aliasing may occur. Aliasing is a distortion of the spectrum of a series which can cause severe problems for time series modelling and forecasting. The situation is more complex and more interesting for nonstationary series as aliasing can be intermittent. Recent work has shown that it is possible to test for the absence of aliasing in nonstationary time series and this article demonstrates that additional benefits can be obtained by modelling a series using a Shannon locally stationary wavelet (LSW) process. We show that for Shannon LSW processes the effects of dyadic-sampling-induced aliasing can be reversed. We illustrate our method by simulation on Shannon LSW processes and also a time-varying autoregressive process where aliasing is detected. We present an analysis of a wind power time series and show that it can be adequately modelled by a Shannon LSW process, the absence of aliasing can not be inferred and present a dealiased estimate of the series.

1. Introduction

This article considers the problem of spectral estimation in the presence of aliasing for a special class of discrete-time non-stationary time series: locally stationary Shannon processes. Aliasing is a well known though, arguably, under-appreciated phenomenon which can severely distort the estimates of the spectrum and autocovariance of a time series. Consider for a moment the situation in the (second-order) stationary setting. Given a discrete time series sampled at rate δt then, as is well known, the range of frequencies which can be observed undistorted in the spectrum is $[0, \pi/\delta t)$, where $\pi/\delta t$ is the Nyquist frequency. If,
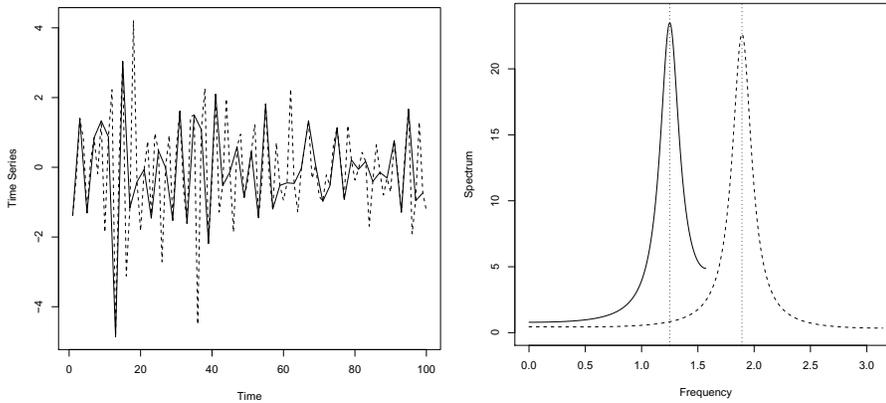

FIG 1. *Aliasing Example.* Both: *solid=subsampled series, dashed=original series.* Left: *realization of both series: X_t on the positive integers, $Y_t = X_{2t}$ on the odd positive integers.* Right: *spectra for both series: $f_X(\omega)$ on $[0, \pi)$, $f_Y(\omega)$ on $[0, \pi/2)$.*

given a fixed sampling rate for our discrete time series, there exist frequencies in the series which are in excess of the Nyquist, then the phenomenon of *aliasing* occurs. In particular, high frequency information reappears in the observed spectrum at lower frequencies. The following example demonstrates this effect more clearly.

Aliasing Example. Figure 1 demonstrates how aliasing can distort the truth. Suppose X_t is the stationary AR(3) model $X_t = -0.2X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} + Z_t$ where Z_t is independent standard Gaussian white noise. The spectrum of X_t on $[0, \pi)$ can be seen as the dashed line on the right plot of Figure 1 and the peak frequency is approximately $\omega = 1.89$.

Let $Y_t = X_{2t}$ be a subsampled series. After subsampling, the highest (Nyquist) frequency that can be observed in *any* series on the even integers is $\pi/2$. The spectrum of Y_t is displayed as the solid line on the right-hand plot in Figure 1 and now the peak frequency occurs at $\omega = 1.25$. Due to aliasing the true peak frequency of 1.89 has been reflected or folded to a peak at 1.25. The left-hand plot in Figure 1 shows the original series (dashed) and how it looks after subsampling (solid). Simply put, the higher frequencies in the original cannot be reproduced accurately in the subsample as the sampling rate is too slow. *The key point is that one often obtains series like Y_t with no knowledge of X_t. Hence, one can be deceived into thinking that there is genuine power at frequency $\omega = 1.25$, when the actual oscillation was at 1.89.* The practical question is: given a series Y_t is there any way of telling that aliasing occurred and the real peak is at 1.89?

We are by no means the first to consider the effect of aliasing in time series. Indeed many respected monographs describe the effect of aliasing, e.g. Hannan (1960) and Priestley (1983). Hinich and collaborators (Hinich and Wolinsky, 1988; Hinich and Messer, 1995) address the question of alias detection for *second-order stationary* series and introduce a hypothesis test for its detection. More

recently Eckley and Nason (2013) consider the more general problem of detection and location of aliasing within locally stationary wavelet time series, proposing a test for the absence of aliasing.

Our work. When dealing with non-stationary series it should be recognised that sometimes such a series can be aliased and sometimes not. Hence, our primary contribution is to propose an approach to estimate and remove the (potentially localised) effect of such time-dependent aliasing from wavelet-based spectral estimates for locally stationary Shannon processes. Specifically this article establishes that, for an (dyadically sampled) aliased Shannon locally stationary wavelet processes, one can identify both the amount of aliasing *and* obtain an (uncontaminated='dealiased') estimate of the evolutionary wavelet spectrum (EWS) for scales below the Nyquist limit.

Our article is arranged as follows: We begin in Section 2 with a brief review of LSW theory, together with a summary of the relevant results of Eckley and Nason (2013). Section 3 considers the behaviour of the Shannon-based EWS estimate under dyadic subsampling and describes an unbiased estimator for the 'below-Nyquist' scales. We then consider the behaviour of the 'dealiased' estimator for two special cases: with no aliasing (Section 3.3) and with white noise (Section 3.4). Finally in Section 4 we explore the performance of our approach on a simulated example and also consider its potential on an example arising from wind energy.

2. Background: LSW processes

We start by recalling the key LSW process definitions from Nason, von Sachs and Kroisandt (2000), henceforth referred to as NvSK00. The key building block in the NvSK00 approach is to construct LSW processes using discrete wavelets. These wavelets are usually defined via the low and high pass filters associated with Daubechies' compactly supported wavelets. More formally:

Definition 2.1. Let $\{h_k\}$ and $\{g_k\}$ be the usual low and high pass quadrature mirror filters that are associated with the construction of the Daubechies (1992) compactly supported orthogonal continuous time wavelets. Then the associated **discrete wavelets** $\psi_j = (\psi_{j,0}, \psi_{j,1}, \ldots, \psi_{j,N_j-1})$ are vectors of length N_j for scales $j \in \mathbb{N}$ and can be calculated as follows:

$$\psi_{1,n} = \sum_k g_{n-2k}\delta_{0,k} = g_n, \text{ for } n = 0, \ldots, N_1 - 1, \tag{1}$$

$$\psi_{j+1,n} = \sum_k h_{n-2k}\psi_{j,k}, \text{ for } n = 0, \ldots, N_{j+1} - 1. \tag{2}$$

Here $N_j = (2^j - 1)(N_h - 1) + 1$ where N_h is the number of non-zero elements of $\{h_k\}$ and $\delta_{0,k}$ is the Kronecker delta. The number of vanishing moments of the wavelet is $N = N_h/2$, i.e. $\int x^m \psi(x)\, dx = 0$ for $m \in \mathbb{N}$ such that $0 \leq m < N$. Finally **nondecimated** discrete wavelets are a particular indexing of discrete wavelets given by $\psi_{j,k}(t) = \psi_{j,k-t}$ for $k, t \in \mathbb{Z}$, $j \in \mathbb{N}$.

In the limit, when the number of vanishing moments $N \to \infty$, we obtain the *Shannon wavelets*. The Shannon discrete wavelets, $\psi_{j,n}$ are no longer compactly supported, do decay rapidly over n and are more efficiently defined in the Fourier domain. However, our work, up until Section 3, is valid for Daubechies' compactly supported wavelets and Shannon wavelets, so we will defer a precise definition of Shannon wavelets until Section 3.

One of the key innovations proposed by NvSK00 is to embed discrete wavelets within the LSW process model as follows:

Definition 2.2. Let $\{X_{t,T}\}_{t=0,\ldots,T-1}$, $T = 2^J$, $J \in \mathbb{N}$, be a sequence of doubly-indexed stochastic processes. Then such a process is said to be a **locally stationary wavelet (LSW)** process if it has the following representation in the mean-square sense:

$$X_{t,T} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k;T} \psi_{j,k}(t) \xi_{j,k}. \tag{3}$$

In this representation the $\{\xi_{j,k}\}$ is a collection of uncorrelated random variables with mean zero and variance one, $\{\psi_{j,k}\}_{j \in \mathbb{N}, k \in \mathbb{Z}}$ form a set of nondecimated discrete wavelets, whilst $\{w_{j,k;T}\}_{j \in \mathbb{N}, k \in \mathbb{Z}}$ is a set of amplitudes satisfying the following conditions:

There exists, for each $j \in \mathbb{N}$, a Lipschitz continuous function $W_j : (0,1) \to \mathbb{R}$, fulfilling the following properties:

   i) $\sum_{j=1}^{\infty} |W_j(z)|^2 < \infty$, uniformly in $z \in (0,1)$;
  ii) the Lipschitz constants, L_j, are uniformly bounded in j and $\sum_{j=1}^{\infty} 2^j L_j < \infty$;
 iii) $\exists \{C_j\}_{j \in \mathbb{N}}$, such that for each T $\sup_k |w_{j,k;T} - W_j(k/T)| \leq C_j/T$, where for each j the supremum is over $k = 0, \ldots, T-1$ and $\{C_j\}$ is such that $\sum_{j=1}^{\infty} C_j < \infty$

Due to the various assumptions made in the above definition, the $\{w_{j,k}\}$ are a collection of amplitudes that are 'smooth' in a particular way as a function of k and $w_{j,k}^2 \approx S_j(k/T)$. Specifically the rate of evolution of the second-order properties of the series $X_{t,T}$ is controlled by smoothness constraints on $S_j(z)$ as a function of z via those imposed on $W_j(z)$ in (i)–(iii) in Definition 2.2. Henceforth, for brevity, we drop the second T subscript in $X_{t,T}$.

Spectral power for a LSW time series is quantified by the evolutionary wavelet spectrum, the time-scale analogue of the usual stationary spectrum, $f(\omega)$.

Definition 2.3. The LSW process $\{X_t\}_{t=0,\ldots,T-1}$, for the infinite sequence $T \geq 1$ has **evolutionary wavelet spectrum (EWS)** defined by $S_j(z) := |W_j(z)|^2$ for $j \in \mathbb{N}$ and $z \in (0,1)$ with respect to $\{\psi_{j,k}\}$.

The spectrum, $S_j(z)$, controls how much variance there is in the process at different scales or frequency bands at time z. Roughly speaking, the quantity $S_j(z)$ corresponds to the variance of the process integrated over the approximate frequency band $[2^{-j}\pi, 2^{1-j}\pi]$.

Finally, we recall the wavelet periodogram used for EWS estimation. The **nondecimated wavelet coefficients**, $\{d_{\ell,m}\}$, of a time series, Y_t, are given

by $d_{\ell,m} = \sum_t Y_t \psi_{\ell,m}(t)$ for $\ell \in \mathbb{N}, m \in \mathbb{Z}$. The **raw wavelet periodogram** of Y_t is defined to be $I_{\ell,m} = d_{\ell,m}^2$ for $\ell \in \mathbb{N}$ and $m \in \mathbb{Z}$ where $\{d_{\ell,m}\}$ are the nondecimated wavelet coefficients of Y_t. The **autocorrelation wavelet**, $\Psi_j(\tau)$, is defined by $\Psi_j(\tau) := \sum_k \psi_{j,k} \psi_{j,k-\tau}$ for $j \in \mathbb{N}, \tau \in \mathbb{Z}$ and the inner product operator of the autocorrelation wavelets is given by: $A_{j,\ell} = <\Psi_j, \Psi_\ell> = \sum_\tau \Psi_j(\tau) \Psi_\ell(\tau)$ for $j, \ell \in \mathbb{N}$.

Remark: It transpires that the raw wavelet periodogram is a biased estimator of the EWS. Specifically NvSK00 established that in the unaliased setting

$$\lim_{T \to \infty} \mathbb{E}(d_{\ell,m}^2) = \sum_j A_{\ell,j} S_j(m/T) + \mathcal{O}(T^{-1}). \tag{4}$$

Fortunately the inner product matrix, A, is invertible and can be constructed efficiently and so an asymptotically unbiased estimate can be obtained (see NvSK00 or Eckley and Nason (2005) for further details).

2.1. Aliasing within LSW processes

Since aliasing within a discrete time process is related to the rate at which the process is observed, it is apparent that by subsampling an unaliased series one can eventually induce aliasing. Recent work by Eckley and Nason (2013) has focussed on precisely this problem, considering the effect of subsampling on LSW series for general Daubechies wavelets. Specifically they consider a setting where an observed LSW process Y_t is obtained from dyadic samples of an (unaliased) series, X_t. In other words $Y_t = X_{2^r t}$ for $t \in \mathbb{Z}$ and $r = 1, \ldots, J-1$. As the Y_t are sampled at a lower frequency than in the original unaliased series, there is potential for high frequency structure in X_t to become aliased in Y_t.

A number of interesting results flow from this setting. For example, Eckley and Nason (2013) establish that $\{Y_t\}$ admits the following decomposition: $Y_t = F_t + L_t$ where $\{L_t\}$ is LSW with the same underlying wavelet family and F_t is a process with zero mean and autocovariance function of known form which, under certain conditions can be shown to be stationary white noise. Of particular interest for the work presented here, Eckley and Nason (2013) establish the following result which describes the form of the raw wavelet periodogram under r-levels of dyadic subsampling:

$$D_{l,m}^{(r)} \equiv \lim_{T \to \infty} \mathbb{E}(d_{l,m}^2) = \sum_{j=1}^{r} S_j(2^r m/T) + \sum_{j=r+1}^{\infty} A_{j-r,\ell} S_j(2^r m/T) + \mathcal{O}(T^{-1}). \tag{5}$$

Comparing this form with the expression for the expectation of the raw wavelet periodogram in the unaliased setting (equation 4) we note that any structure which is sampled always appears as a 'contaminant' $\sum_{j=1}^{r} S_j(2^r m/T)$ at *every* scale of analysis and that this is observed at a grid spacing of $(2^r m)/T$ instead of m/T. This results holds for all Daubechies' wavelets and also Shannon wavelets. However to establish the result for the Shannon setting, we also require that the EWS, $S_j(z)$, has continuous first derivative for each $j > 0$.

Given the above, it is natural to consider whether some form of hypothesis might be formed relating to the presence or absence of aliasing at a given time point $z \in (0, 1)$. Such a test is proposed by Eckley and Nason (2013), although due to the fact that white noise causes similar consequences as aliasing, the test which can be proposed has a very specific form: H_0: no aliased component *and* no white noise component at a specific time point z, against an alternative H_A: aliased component *or* white noise component present at z. We refer interested readers to Eckley and Nason (2013) for further details of this test.

3. The Shannon succesive scale method

Can we do more than the hypothesis test described in the previous section? In particular can we go further than identify periods of aliasing/white noise contamination? This section shows that for Shannon LSW processes one can identify both the contamination, $\sum_{j=1}^{r} S_j(z)$, *and* obtain an (uncontaminated = 'dealiased') estimate of the EWS for scales below the Nyquist limit. In regular EWS estimation such scales become contaminated according to the spectral modification shown in equation (5).

Such 'dealiasing' is known to be possible for certain restricted bandpass signals (those whose nonzero spectrum is contained in a compact interval of frequencies, $\omega \in \mathbb{R}$ not containing zero), see Harris (2004). However, as we explain below, Shannon LSW seems to be special as 'dealiasing' can be achieved for processes that are not bandpass or bandlimited. Shannon's advantage stems from the orthogonality of its autocorrelation wavelets at different scales, mathematically expressed by $A_{j,\ell} = 0$ for $j \neq \ell$ and $A_{j,j} = 2^j$, see NvSK00 A.5. In particular, the orthogonality means that we do not need to concern ourselves with the effects of 'hidden' scales (i.e. the very coarsest scales).

Section 3.1, next, provides a brief review of Shannon wavelets. Section 3.2 considers the behaviour of the Shannon-based EWS estimate under dyadic subsampling and describes an unbiased estimator for the 'below-Nyquist' scales. We then consider the behaviour of the 'dealiased' estimator for two special cases: with no aliasing in Section 3.3 and with white noise in Section 3.4.

3.1. Brief review of Shannon wavelets

The Shannon wavelet is most efficiently defined in the Fourier domain, see, for example, Chui (1997), pages 46 and 47 or Appendix A.5 of NvSK00. Define the set $C_j = [-2^{-j+1}\pi, -2^{-j}\pi] \cup [2^{-j}\pi, 2^{-j+1}\pi]$. The discrete Fourier transform of the Shannon wavelet is given by $\hat{\psi}_j(\omega) = -2^{j/2} \exp(-2^{j-1}i\omega)\chi_{C_j}(\omega)$, where $\chi_A(\omega)$ is the indicator function of the set A. The underlying mother wavelet is given by $\psi(t) = \pi^{-1}(t - \frac{1}{2})^{-1}(\sin 2\pi t - \cos \pi t)$. Figure 2 shows a picture of a Shannon wavelet which is not compactly supported but has decay like $\mathcal{O}(t^{-1})$ in the time domain. As mentioned above, the A matrix for Shannon wavelets is diagonal with $A_{j,j} = 2^j$, for $j \in \mathbb{N}$.
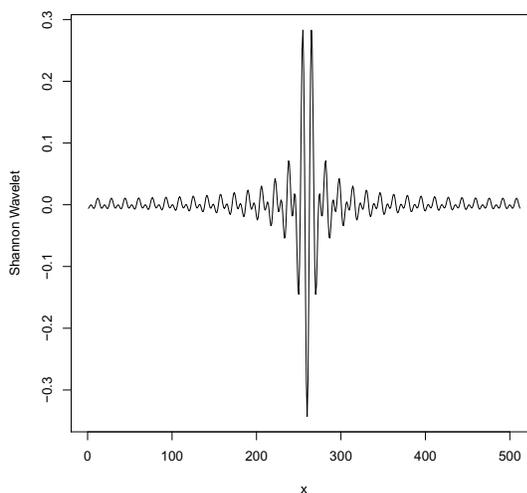
FIG 2. *A Shannon wavelet.*

3.2. *Shannon wavelet periodogram under dyadic subsampling*

We now turn to consider the behaviour of the Shannon-based EWS estimate under dyadic subsampling and describes an unbiased estimator for the 'below-Nyquist' scales.

Setup. Shannon's autocorrelation wavelet orthogonality is extremely useful as it means (5) can be considerably simplified to:

$$D_{j,m}^{(r)} = \sum_{\ell=1}^{r} S_\ell(z_m) + A_{j,j}S_{r+j}(z_m), \quad \text{for } r, j \in \mathbb{N}. \tag{6}$$

The key simplification in (6) is that it involves (i) the aliased content $\sum_{\ell=1}^{r} S_\ell(z_m)$ and (ii) *only* the spectral level S_{r+j}. In particular, spectral information from coarser levels, e.g. the infinity of levels in the right-hand sum of (5) do not appear in (6). Next we turn to consider how one might estimate the alias and spectral components within this particular setting.

Estimation. Define $I_m := (I_{1,m}, \ldots, I_{J,m})^T$ to be the vector of raw wavelet periodograms at time m. Estimates of

$$\tilde{S}_m := \left(\sum_{\ell=1}^{r} S_\ell(z_m), S_{r+1}(z_m), \ldots, S_{r+J}(z_m) \right),$$

can be obtained by inverting the series of equations in (6) and substituting the empirical quantity $d_{j,m}^2$ for its asymptotic expectation $D_{j,m}^{(r)}$. The astute reader will have noted that we have J equations for $J+1$ unknowns. A further equation can be obtained by defining the local *feasible* variance of the series

$\sigma_m^2 = \sum_{\ell=1}^{r+J} S_\ell(z_m)$. Then by substituting a suitable estimator $\hat{\sigma}^2$ for σ^2 we have a system with $J + 1$ equations and unknowns which can be solved.

At this point it is worth explaining further the difference between feasible and total variance. The total variance, $\sum_{j=1}^{\infty} S_j(z)$, specifically includes contributions from increasing scales without bound. In other words, the total variance is the variance of the underlying stochastic processes at time z. The feasible variance is computed from the sample where we only ever observe a finite but potentially large, number of scales. Property (i) of Definition 2.2 implies that $S_j(z) \to 0$ as $j \to \infty$, hence power has to decay eventually.

With the additional equation, the matrix representation of (6) is:

$$\tilde{D}_m^{(r)} := \begin{pmatrix} \sigma_m^2 \\ D_{1,m}^{(r)} \\ \vdots \\ D_{J,m}^{(r)} \end{pmatrix} = \tilde{A}_J \tilde{S}_m = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 0 & \cdots & 0 \\ 1 & 0 & 4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 2^J \end{pmatrix} \begin{pmatrix} \sum_{\ell=1}^{r} S_\ell(z_m) \\ S_{r+1}(z_m) \\ S_{r+2}(z_m) \\ \vdots \\ S_{r+J}(z_m) \end{pmatrix}, \quad (7)$$

where $\tilde{A}_J$ is a $(J + 1)$-dimensional square matrix which is invertible by the next Proposition.

Proposition 3.1. *$\tilde{A}_J$ is invertible with* $\det \tilde{A}_J = 2^{J(J-1)/2}$ *for all integers* $J > 0$.

Proof. In the appendix, along with an explicit formula for $\tilde{A}_J^{-1}$.

The Estimator. To estimate $\tilde{S}_m$ we introduce the following estimator $\hat{\tilde{S}}_m = \tilde{A}_J^{-1} \hat{\tilde{D}}_m^{(r)}$, where $\hat{\tilde{D}}_m^{(r)} = (\hat{\sigma}_m^2, I_m)^T$. The following restult establishes that this estimator is unbiased.

Proposition 3.2. *Let X_t be a Shannon LSW process, $Y_t = X_{2^r t}$. Then $\hat{\tilde{S}}_m$ is an unbiased estimator of $\tilde{S}_m$.*

Proof. The proof is simply a rearrangement of (7).

In summary, Proposition 3.2 shows that, for Shannon LSW processes, under aliasing we can recover the spectral components that would be contaminated using the usual EWS estimation process. In the language of stationary time series: we can successfully estimate below-Nyquist frequency components that would have previously contaminated by the higher frequencies wrapping round due to aliasing. Note, that since r is arbitrary this also includes the aliasing of continuous time processes by the sampling process. This new method works for signals that are not bandpass or bandlimited: however, they have to be of the Shannon LSW form. Proposition 3.2 also permits us to recover $\sum_{\ell=1}^{r} S_\ell(z_m)$.

Smoothing for consistency. NvSK00 note that the raw wavelet periodogram is not consistent and, like the classical case, needs to be smoothed to obtain a consistency. Such smoothing has been carried out using a variety of methods in the literature, for example, second-stage wavelet shrinkage in NvSK00 or Haar-Fisz estimation for piecewise constant spectra in Fryzlewicz and Nason (2006).

Any of these methods could be used here to obtain a consistent estimator of $\tilde{A}_J \tilde{S}_m$ with the same rates of convergence. Estimates of $\tilde{S}_m$ (without the pre-multiplier of $\tilde{A}_J$) can be carried out in precisely the same manner as in NvSK00 with the same rates of convergence as in their Theorem 4. Such theory is based on existing results on quadratic forms of Gaussian variables which are χ^2, see Neumann and von Sachs (1995).

In the next section, we describe some results which validate the behaviour of our estimator for some key cases of interest.

3.3. No aliasing and no white noise

What about the case where there is no aliasing or white noise? In other words, the raw periodogram vector I_m can be, and is, computed in the regular case where it is a precursor to spectral estimation. From NvSK00, in the non-aliased case we know that

$$\tilde{D}_m^{(0)} = \begin{pmatrix} \sum_{j=1}^{J} S_j(z_m) \\ AS_m \end{pmatrix}, \tag{8}$$

where A is the inner product matrix from NvSK00 and $S_m = (S_1(z_m), \ldots, S_J(z_m))^T$.

However, suppose we just compute the raw wavelet periodogram, I_m and apply our 'dealiasing' matrix, $\tilde{A}_J^{-1}$, but the underlying situation is that of *no* aliasing. Then, the following happens:

$$\tilde{A}_J^{-1} \tilde{D}_m^{(0)} = \tilde{A}_J^{-1} \begin{pmatrix} \sum_{j=1}^{J} S_j(z_m) \\ AS_m \end{pmatrix} = \tilde{A}_J^{-1} \begin{pmatrix} \sum_{j=1}^{J} S_j(z_m) \\ 2S_1(z_m) \\ \vdots \\ 2^J S_J(z_m) \end{pmatrix}, \tag{9}$$

the latter equality because A is diagonal for Shannon wavelets with $A_{j,j} = 2^j$.

Proposition 3.3. *With no aliasing the top row of the 'dealiased' vector is zero, and the remaining elements are an unbiased estimate of the true EWS. Mathematically,*

$$\tilde{A}_J^{-1} \tilde{D}_m^{(0)} = (0, S_1(z_m), \ldots, S_J(z_m))^T. \tag{10}$$

Proof. In the appendix.

This means that if there is no aliasing then application of our 'dealiasing' matrix $\tilde{A}_J^{-1}$ behaves as we would expect, telling us that the top row is zero and the remainder of the rows consist of the true EWS.

3.4. White noise

As mentioned in Section 2.1, if Y_t is white noise with variance σ^2 then the expectation of the wavelet periodogram is given by $D_{\ell,m} = \sigma^2$ for all ℓ, m.

What happens if we subject white noise to our dealiasing method, i.e. applying $\tilde{A}_J^{-1}$ to $\sigma^2 1_{J+1}$, the $(J+1)$-dimensional vector of σ^2s?

Proposition 3.4.

$$\tilde{A}_J^{-1}\sigma^2 1_{J+1} = \sigma^2(1, 0, \ldots, 0)^T. \tag{11}$$

Proof. This is easy to see as $\tilde{A}_J$ is invertible and its first column is 1_{J+1}.

Propositions 3.3 and 3.4 together mean that if you have any (a) white noise component or (b) aliased power in your series then it migrates to the top row of the 'dealiased vector'. Sections 3.2 and 3.3 show that the other rows in the 'dealiased vector' contain *uncontaminated* estimates of the remainder of the spectrum: something that classical spectral analysis does not do.

4. Examples

In this section we explore the application of our approach to simulated and real data. In Section 4.1 we consider the performance of our estimation scheme in a simulation study based on stationary time series, paying particular attention to how the accuracy of our estimates behaves for different lengths of time series. Then, in Section 4.3 we apply the proposed spectral estimation scheme to an example arising from the wind energy sector.

4.1. Four-scale LSW process: Shannon 'dealiasing'

We begin by exploring the performance of our approach in a simulated setting. Here we let X_t be a stationary Shannon LSW process with $S_1 = 1$, $S_2 = 2$, $S_3 = 3$ and $S_4 = 0.5$ and Gaussian innovations (since the process is stationary, $S_j(z) = S_j$ for all $z \in (0,1)$). We also set $Y_t = X_{2t}$ – in other words on sampling, all the finest scale structure (i.e. scale $j = 1$) becomes aliased. Table 1 shows the results of estimating $\tilde{S}$ using the methods of Section 3, taken over many simulations and shows that the sample size has to be moderately large (e.g., $T = 1024$, say) before the dealiasing becomes accurate. However, the results are not unreasonable for even quite small sample sizes ($T = 256$) bearing in mind that the methodology is designed to work with nonstationary series.

TABLE 1
Estimated mean values of $\tilde{S}_j$ with their standard deviations in parentheses for the Four-scale LSW process of different lengths T as defined in the text with aliasing. *The means/sds are taken over 200 simulations to 2 d.p.*

	Est. Aliased	Corrected Spectrum			
T	j=1	j=2	j=3	j=4	j=5
Truth	1	2	3	0.5	0
128	2.65 (0.98)	1.26 (0.67)	2.54 (0.96)	0.32 (0.33)	0.00 (0.07)
256	1.80 (0.61)	1.59 (0.53)	2.75 (0.76)	0.38 (0.24)	0.00 (0.05)
512	1.39 (0.36)	1.81 (0.39)	2.86 (0.60)	0.45 (0.17)	0.00 (0.03)
1024	1.18 (0.25)	1.91 (0.25)	2.97 (0.39)	0.48 (0.12)	0.00 (0.02)
2048	1.10 (0.18)	1.93 (0.18)	2.95 (0.29)	0.49 (0.08)	0.00 (0.01)

TABLE 2
Estimated mean values of $\tilde{S}_j$ with their standard deviations in parentheses for the Four-scale LSW process of different lengths T as defined in the text with no aliasing. *The means/sds are taken over 200 simulations to 2 d.p.*

	Est. Aliased	Corrected Spectrum			
T	j=1	j=2	j=3	j=4	j=5
Truth	0	2	3	0.5	0
128	1.41 (0.33)	1.33 (0.51)	2.65 (1.01)	0.33 (0.25)	0.00 (0.02)
256	0.68 (0.10)	1.64 (0.31)	2.79 (0.70)	0.41 (0.17)	0.00 (0.01)
512	0.34 (0.04)	1.84 (0.21)	2.89 (0.53)	0.46 (0.12)	0.00 (0.00)
1024	0.17 (0.01)	1.91 (0.19)	2.94 (0.37)	0.47 (0.09)	0.00 (0.00)
2048	0.09 (0.00)	1.96 (0.12)	2.97 (0.27)	0.49 (0.06)	0.00 (0.00)

Rarely, our estimator is slightly negative and so, as is common in the literature, when this occurs we truncate it to zero. The simulation was repeated for $S_1 = 0$, i.e. no aliasing, and these results are shown in Table 2.

4.2. Time-varying AR(1) process

In this example we investigate what happens when an aliased time-varying autoregressive process (TVAR) of order one is subjected to our methodology. We simulated a single realization from the TVAR(1) model $X_t = \alpha_t X_{t-1} + Z_t$, for $t = 1, \ldots, 512$, α_t varied linearly from $\alpha_1 = 0.9$ to $\alpha_{512} = -0.9$ over the extent of the series and Z_t was Gaussian white noise with mean zero and unit variance. Simulations from this model can be obtained using the `tvar1sim()` function from the `locits` package developed by Nason (2013). We then subsampled the series to obtain $Y_t = X_{2t}$ and the subsampled realization is shown in the left hand plot of Figure 3. Figure 3 also shows the estimate $\hat{\tilde{S}}_m$ for this realization after smoothing (right hand plot). The highest frequencies in the realization of
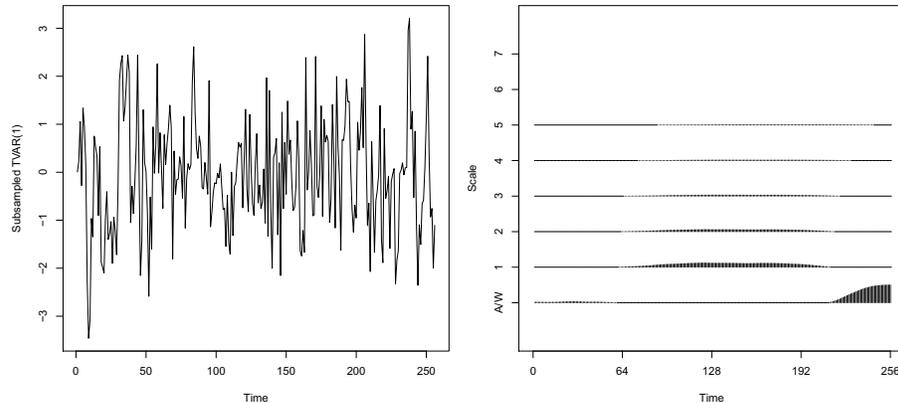

FIG 3. Left: *Subsampled TVAR(1) realization.* Right: *Smoothed estimate of $\tilde{S}$ for TVAR(1) realization. The lowest scale labelled A/W contains the aliased or white noise content, the other scales, j correspond to S_{j+r} in $\tilde{S}$.*

X_t are on the far right-hand side of the series and indeed our method presents power in the 'aliasing/white noise' (lowest) band on the far right-hand side of the right plot in Figure 3. This example is particularly interesting as aliasing (or white noise) has clearly been detected even though the underlying process is not a Shannon LSW process and indeed not a LSW process. However, there are undoubtedly other TVARMA processes that are 'further away' from LSW processes and further research would be required to see whether aliasing could be as well detected as in this case.

4.3. Aberporth wind power series

The problem of wind speed and power forecasting has been a topic of increasing focus in recent years, see for example Landberg et al. (2003) or Genton and Hering (2007). This has been, to a greater extent, motivated by the need to develop reliable forecast tools to enable effective integration of wind farm output into electricity grids. Several authors have considered the applicability of ARIMA models to such data, see, for example Brown, Katz and Murphy (1984), Huang and Chalabi (1995), Sfetsos (2002)), where, naturally the fitting of such models would require the assumption of no aliasing within the time series.

IDA and Shannon LSW Model Feasibility. This example considers hourly wind speed data measured at the Meteorological Office station in Aberporth, Wales, previously studied by Hunt and Nason (2001), Nason and Sapatinas (2002) and Cardinali and Nason (2010). The first differences (to remove trend) of the wind speed series are shown in Figure 4.

To verify the feasibility of representating this data by a Shannon LSW model, we consider the regular spectrum, its confidence intervals and the estimated Shannon spectrum. These are all shown in Figure 4. In a nutshell, this permits
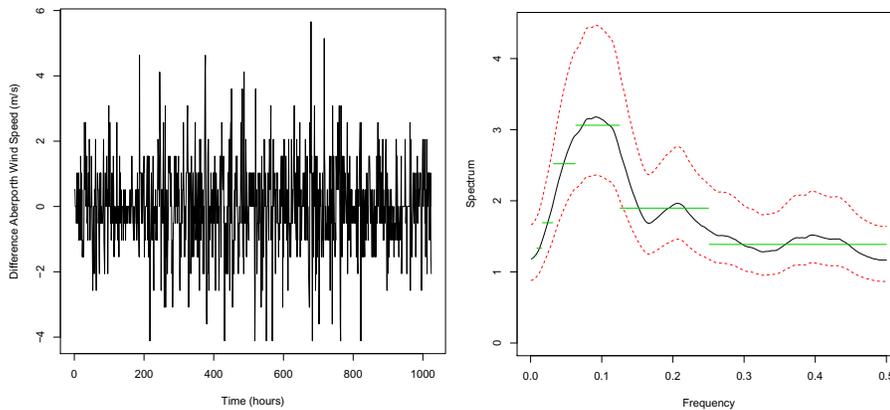

FIG 4. Left: *First differences of wind speed at Aberporth* (ms^{-1}). Right: *Spectral analysis of* `das` *series. The dashed lines are 99% confidence intervals. The solid horizontal lines are projections of the spectrum onto the Shannon LSW process.*
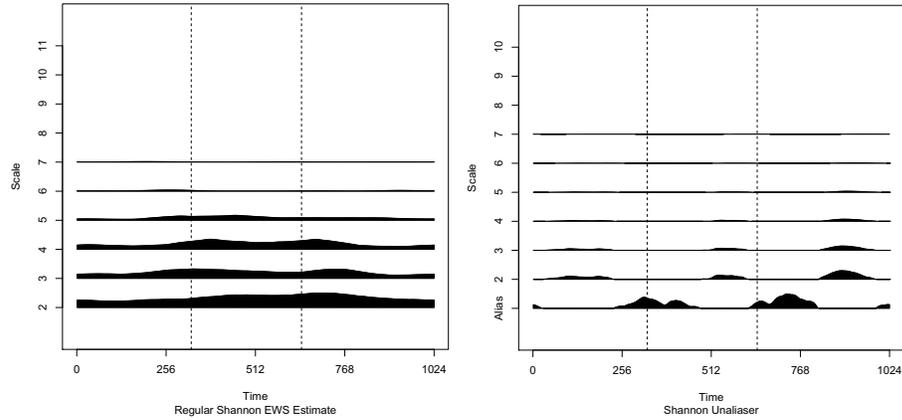
FIG 5. Left: *Regular Shannon wavelet EWS estimate of* das *series.* Right: *Dealiased Shannon wavelet EWS estimate of* das *series. The levels at 2 and coarser are again present. An extra level is present containing the sum of all potential aliases. The vertical lines in both estimates at $t = 328, 644$ are explained in the text.*

us to verify whether the Shannon LSW spectrum lies within the confidence intervals of a nonparametric spectrum estimate from the data. If it does we deem the Shannon LSW process to be a feasible model for the series. Note that splitting the series into smaller parts improves the fit to the Shannon basis, hence providing ample justification for this model for this series (for the whole series we model a *locally* stationary Shannon series).

Aliasing? The standard Shannon wavelet EWS is shown in the left hand plot of Figure 5. Here the finest scale referred to is scale $j = 2$, as we are investigating whether there is power at finer scales. Now focus on time $t = 328$ indicated by a vertical line. There is clearly oscillatory power at scale $j = 3$ at this time which slowly tails off. This 'tailing off' coincides with an increase of power at the finer scale $j = 2$ immediately afterwards. Indeed, careful scrutiny shows some power starting earlier still at scale $j = 5$. This pattern of power is indicative of a pulse increasing in frequency over time. Something similar happens at time $t = 644$, power increasing in successively finer scales at slightly later times. The key thing is that the left hand plot in Figure 5 suggests a possibility of two pulses of power increasing in frequency over time. So, could this apparent frequency increase have increased enough to cause aliasing?

Figure 5 (right hand plot) shows the result of applying our dealiasing methodology. The vertical lines at $t = 328, 644$ are again shown. What is interesting about this figure is that our method indeed suggests significant power is present in the 'aliased' band, and looking at the coarser scales the 'frequency increase' from 2 to 'aliased' is present in both cases (i.e. power at scale 2 just before the aliasing occurred).

From this figure it might be thought that the presence of power is either present in the aliased band or the 'non-aliased bands' (two to ten) but not

both, i.e. they are mutually exclusive at any given point in time. This is actually illusory, there is actually power operating at some of the coarser scales simultaneously with that in the 'aliased' band, but it is of a much smaller order than that in the 'aliased' band.

In conclusion it is important to realize that we are not necessarily detecting aliasing: the power in the 'aliased' band could be due to a white noise component (and our test cannot detect the difference between the two). However, we might be suspicious that aliasing is occurring due to the suggestion that there is power of increasing frequency over time in the left hand plot of Figure 5. More formal clarification of this could be obtained by applying the Hinich and Wolinsky (1988) on assumed stationary segments of the relevant parts of the series.

5. Concluding remarks

This article introduces new theory on what happens to a Shannon LSW process, and its spectrum, under dyadic subsampling. For this special case we can estimate below-Nyquist spectral components, even for time series that are not bandlimited because of the orthogonality of Shannon autocorrelation wavelets. Additional simulation studies have also demonstrated that the Shannon analysis wavelet works well even if the process was synthesized using other Daubechies compactly supported wavelets.

Improved knowledge of aliasing, white noise component, or neither clearly has important implications for the modelling, estimation and forecasting of time series. For example, if there is strong evidence for aliasing then forecasting might be inaccurate as forecasting techniques often rely heavily on spectral estimates. On the other hand, knowledge that aliasing is unlikely to be affecting a time series means that the sampling rate need not be increased.

Other directions for future research Clearly there is much scope for future work in this area. For example it would also be interesting to explore whether some of these ideas can also be extended to the locally stationary Fourier or other settings, such as the SLEX framework of Ombao et al. (2002). Finally, in all that we have considered here we have focussed on those situations where the process generating wavelet is known *a priori*. However as early work by Gott and Eckley (2013) reports, in practice this will be unknown and the effect of analysing the time series by another wavelet can be significant. What this means in the aliased world still remains to be investigated.

Acknowledgements

Appendix: Proofs

Proof of Proposition 3.1 (Invertibility of $\tilde{A}_J$)

Write $\tilde{A}_J$ in block matrix form as follows:

$$\tilde{A}_J = \begin{pmatrix} I & B \\ C & A \end{pmatrix}, \tag{12}$$

where $I = (1)$, $B = C^T = (1, \ldots, 1)$ and $A = \text{diag}(2, 2^2, \ldots, 2^J)$ is just the inner product matrix from NvSK00. Clearly, A is invertible and hence, using basic properties of determinants, we have

$$\det \tilde{A}_J = \det(A) \det(1 - BA^{-1}C). \tag{13}$$

Here $A^{-1} = \text{diag}(2^{-1}, 2^{-2}, \ldots, 2^{-J})$, and $BA^{-1}C = \sum_{j=1}^{J} 2^{-j} = 1 - 2^{-J}$. Hence:

$$\det \tilde{A}_J = 2^{1+2+\cdots+J-1} = 2^{J(J-1)/2} > 0 \quad \text{for } J \in \mathbb{N}. \tag{14}$$

We now find the inverse formula. We have already mentioned that A is invertible and it can be seen from above that $1 - BA^{-1}C = 2^{-J}$. Hence the quantity $(1 - BA^{-1}C)$ is invertible and its inverse equal to 2^J. Hence, using well-known results on the inverse of block matrices we obtain:

$$\tilde{A}_J^{-1} = \begin{pmatrix} (1 - BA^{-1}C)^{-1} & Q \\ Q^T & P \end{pmatrix} \tag{15}$$

where

$$Q = -(1 - BA^{-1}C)^{-1}BA^{-1} = -2^J BA^{-1}, \tag{16}$$

and

$$P = A^{-1} + A^{-1}C(1 - BA^{-1}C)^{-1}BA^{-1} = A^{-1} + 2^J A^{-1}CBA^{-1}. \tag{17}$$

After some algebra it can be shown that $Q = -(2^{J-1}, 2^{J-2}, \ldots, 1)$ and $P = (p_{i,j})$ where

$$p_{i,j} = \begin{cases} 2^{-j}(1 + 2^{J-j}) & \text{for } i = j, \\ 2^{J-(i+j)} & \text{for } i \neq j. \end{cases} \tag{18}$$

Hence:

$$\tilde{A}_J^{-1} = \begin{pmatrix} 2^J & -2^{J-1} & -2^{J-2} & \cdots & -1 \\ -2^{J-1} & & & & \\ \vdots & & & P & \\ -2^{-1} & & & & \end{pmatrix} \tag{19}$$

Proof of Proposition 3.3: $\tilde{A}_J^{-1}\tilde{D}_m^{(0)}$

From (9) we write

$$\tilde{D}_m^{(0)} = \begin{pmatrix} v \\ AS_m \end{pmatrix}, \tag{20}$$

where $v = \sum_{j=1}^J S_j(z_m)$. Then

$$\tilde{A}_J^{-1}\tilde{D}_m^{(0)} = \begin{pmatrix} 2^J & Q \\ Q^T & P \end{pmatrix}\begin{pmatrix} v \\ AS_m \end{pmatrix} = \begin{pmatrix} 2^J v + QAS_m \\ Q^T v + PAS_m \end{pmatrix}, \tag{21}$$

where Q, P were defined in the proof of Proposition 3.1 above.

The top row of the right-hand side of (21) is

$$2^J v + QAS_m = 2^J \sum_{j=1}^J S_j(z_m) - 2^{J-1} \cdot 2S_1(z_m)$$
$$- 2^{J-2} \cdot 2^2 S_2(z_m) - \cdots - 1 \cdot 2^J S_J(z_m) = 0. \tag{22}$$

For the second row, the quantity $Q^T v = -\sum_{j=1}^J S_j(z_m)(2^{J-1}, \ldots, 1)$ and from (17) we have:

$$
\begin{aligned}
PAS_m &= (A^{-1} + 2^J A^{-1} CBA^{-1})AS_m = (I + 2^J A^{-1} CB)S_m \quad (23) \\
&= (I - Q^T v). \quad (24)
\end{aligned}
$$

Hence, rows 2 to $J+1$ of (21) are the vector S_m and we have the remarkable result that

$$\tilde{A}_J^{-1}\tilde{D}_m^{(0)} = (0, S_1(z_m), \ldots, S_J(z_m))^T. \tag{25}$$

References

BROWN, B. G., KATZ, R. W. and MURPHY, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *J. Clim. Appl. Meteorol.* **23** 1184–1195.

CARDINALI, A. and NASON, G. P. (2010). Costationarity of locally stationary time series. *J. Time Ser. Econom.* **2** Article 1. MR2915633

CHUI, C. K. (1997). *Wavelets: A Mathematical Tool for Signal Analysis.* SIAM, Philadelphia. MR1443204

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia. MR1162107

ECKLEY, I. A. and NASON, G. P. (2005). Efficient computation of the discrete autocorrelation wavelet inner product matrix. *Stat. Comput.* **15** 83–92. MR2137272

ECKLEY, I. A. and NASON, G. P. (2013). A test for the absence of aliasing for locally stationary wavelet time series (submitted for publication).

FRYZLEWICZ, P. and NASON, G. P. (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *J. R. Statist. Soc.* B **68** 611–634. MR2301011

Genton, M. and Hering, A. (2007). Blowing in the wind. *Significance* **4** 11–14. MR2359227

Gott, A. N. and Eckley, I. A. (2013). A note on the effect of wavelet choice on the estimation of the evolutionary wavelet spectrum. *Commun. Stat. Simulat.* **42** 393–406. MR3017003

Hannan, E. J. (1960). *Time Series Analysis*. Chapman and Hall, London. MR0114281

Harris, F. J. (2004). *Multirate Signal Processing for Communication Systems*. Prentice Hall, Upper Saddle River, NJ, USA.

Hinich, M. J. and Messer, H. (1995). On the principal domain of the discrete bispectrum of a stationary signal. *IEEE Trans. Sig. Proc.* **43** 2130–2134.

Hinich, M. J. and Wolinsky, M. A. (1988). A test for aliasing using bispectral analysis. *J. Am. Statist. Ass.* **83** 499–502. MR0971378

Huang, M. and Chalabi, Z. S. (1995). Use of time-series analysis to model and forecast wind speed. *J. Wind Eng. Ind. Aerod.* **56** 311–322.

Hunt, K. and Nason, G. P. (2001). Wind speed modelling and short-term prediction using wavelets. *Wind Eng.* **25** 55-61.

Landberg, L., Giebel, G., Nielsen, H. A., Nielsen, T. and Madsen, H. (2003). Short-term prediction – An overview. *Wind Energy* **6** 273–280.

Nason, G. P. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Statist. Soc.* B **75** 879–904. MR3124795

Nason, G. P. and Sapatinas, T. (2002). Wavelet packet transfer function modelling of nonstationary time series. *Stat. Comput.* **12** 45–56. MR1877579

Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc.* B **62** 271-292. MR1749539

Neumann, M. and von Sachs, R. (1995). Wavelet thresholding: Beyond the Gaussian iid situation. In *Wavelets and Statistics*, (A. Antoniadis and G. Oppenheim, eds.). *Lecture Notes in Statistics* **103** Springer-Verlag, New York.

Ombao, H. C., Raz, J., von Sachs, R. and Guo, W. (2002). The SLEX model of non-stationary random processes. *Ann. Inst. Statist. Math.* **54** 171–200. MR1893549

Priestley, M. B. (1983). *Spectral Analysis and Time Series*. Academic Press, London.

Sfetsos, A. (2002). A novel approach for the forecasting of mean hourly wind speed time series. *Renew. Energ* **27** 163–174.