Stat Comput manuscript No.
(will be inserted by the editor)

Spectral estimation for locally stationary time series with missing observations

Marina I. Knight · Matthew A. Nunes · Guy P. Nason

Received: date / Accepted: date

Abstract Time series arising in practice often have an inherently irregular sampling structure or missing values, that can arise for example due to a faulty measuring device or complex time-dependent nature. Spectral decomposition of time series is a traditionally useful tool for data variability analysis. However, existing methods for spectral estimation often assume a regularly-sampled time series, or require modifications to cope with irregular or 'gappy' data. Additionally, many techniques also assume that the time series are stationary, which in the majority of cases is demonstrably not appropriate. This article addresses the topic of spectral estimation of a non-stationary time series sampled with missing data. The time series is modelled as a locally stationary wavelet process in the sense introduced by Nason et al (2000) and its realization is assumed to feature missing observations. Our work proposes an estimator (the periodogram) for the process wavelet spectrum, which copes with the missing data whilst relaxing the strong assumption of stationarity. At the centre of our construction are second generation wavelets built by means of the lifting scheme (Sweldens, 1995), designed to cope with irregular data. We investigate the theoretical properties of our proposed periodogram, and show that it can be smoothed to produce a bias-corrected spectral estimate by adopting a penalized least squares criterion. We demonstrate our method with real data and simulated examples.

Keywords missing data · nondecimated transform · spectral estimation · wavelet lifting

Corresponding author: m.nunes@lancs.ac.uk

M. Knight
NHS Blood and Transplant, Fox Den Road, Stoke Gifford, Bristol, BS34 8RR, UK
Tel.: +44 (0)117 975 7588
Fax: +44 (0)117 975 7577
E-mail: marina.knight@nhsbt.nhs.uk

M. Nunes
Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, UK
Tel.: +44 (0)1524 593960
Fax: +44 (0)1524 592681
E-mail: m.nunes@lancs.ac.uk

G.P. Nason
School of Mathematics, University of Bristol, Bristol, BS8 1TW, UK
Tel.: +44 (0)117 928 8633
Fax: +44 (0)117 928 7999
E-mail: G.P.Nason@bristol.ac.uk

1 Introduction

The importance of spectral densities for stochastic processes and the usefulness of their estimation is well established in the time series analysis literature. In this article, we assume a basic knowledge of general time series concepts, but emphasize important results where needed.

For stationary processes, the question of spectral estimation has been studied extensively, see for example Priestley (1981). However, in various fields, such as finance (Mikosch and Starica, 2004; Fryźlewicz et al, 2006), and medicine (Nason et al, 2001; Cranstoun et al, 2002; Cazelles et al, 2007), modelling the observed data as stationary is not always appropriate. There have been many recent contributions to the literature dealing with non-stationary time series, such as the estimation of time-varying ARCH models (Dahlhaus and Subba Rao, 2006, 2007), spectral estimation methods based on SLEX bases (Ombao et al, 2002) or locally stationary wavelet-based approaches (Nason et al, 2000; Fryźlewicz and Nason, 2006; Van Bellegem and Von Sachs, 2008).

The second complexity is that time series with missing data frequently appear in practice. The 'patterns' of the missing observations and reasons behind them are varied:

for instance a whole sequence of data might be missing due to a malfunction of the machine recording the observations, the data may be censored, observations may be missing at random or follow a systematic pattern. The existence of missing observations induces irregularities in the time locations, while certain types of data naturally have irregularly-spaced observations, such as environmental time series (Witt and Schumann, 2005; Dilmaghani et al, 2007) or "high-frequency" financial data (Engle, 2000). In this context, data analysis cannot take place within the well-specified framework devoted to discrete time series measured at equal time intervals. Quite commonly, when missing observations are present in the data, they are imputed following various recommendations, for example by 'common sense', or some computations may be performed on the 'gappy' data (for more details, see e.g. Chatfield (2004)). Traditional spectral estimation methods can then be performed on the 'complete' time series.

Methods for autocovariance and spectral estimation for stochastic processes sampled at irregular locations have been developed (Hall et al, 1994; Bos et al, 2002). Some spectral estimation techniques involve mapping the inherent irregular structure of time series so that regularly-spaced spectral analysis can be performed, for example through sampling (Clinger and Van Ness, 1976; Broersen, 2008). For an overview of preprocessing methods for spectral estimation of irregular time series that parallel approaches built for regularly-spaced data, see Stoica and Sandgren (2006). The authors of this work underline the limited choice of spectral analysis techniques for irregularly-sampled data, and emphasize the large number of fields that could benefit from it, such as biomedicine, astronomy, seismology or engineering.

Models specifically developed for time series with missing data and their use for spectral estimation have been discussed in the literature (Broersen et al, 2004; Broersen, 2006). Mondal and Percival (2008) formulate unbiased spectral estimators assuming wavelet models of stationary time series and also investigate their asymptotic properties. If the missing observations occur with a periodic structure, Jones (1962) provides a development for spectral estimation of a stationary time series.

The current existing techniques in the literature of non-stationary time series do not easily extend to handle irregular data observations or missing data, as the constructions for missing data situations outlined above are only valid for *stationary* time series. Hence the focus in this article is to investigate the problem of spectral estimation for a non-stationary process with missing observations. In our approach, non-stationarity is defined in the sense introduced by Nason et al (2000) and our construction will make use of a 'nondecimated' wavelet algorithm introduced in Knight and Nason (2009). At the core of our construction are second generation wavelets built following the lifting scheme (Sweldens, 1995), that removes 'one coefficient at a time' (LOCAAT) by Jansen et al (2001, 2009) and used extensively in Nunes et al (2006) and Knight and Nason (2006).

This article is organized as follows. Section 2 briefly introduces (stationary) time series and outline ways of modelling time series data without imposing the strong assumption of stationarity. We discuss the concept of rescaled time introduced by Dahlhaus (1997), and then present the main results in the construction of a spectral estimator for locally stationary wavelet (LSW) processes of Nason et al (2000). Section 3 details our wavelet periodogram for a LSW process *with missing observations*. The missing data is handled by using a generalized wavelet transform, known as *lifting* and introduce the LOCAAT algorithm of Jansen et al (2001) and set out the 'nondecimated' lifting transform (NLT) of Knight and Nason (2009). We then provide a series of both actual and simulated data examples in Section 4. Section 5 investigates the raw periodogram and proposes a penalty criterion for removing its inherent bias and 'power leakage'. Section 6 concludes and outlines ideas for further work.

2 Spectral estimation for locally stationary time series

2.1 Locally stationary time series

In order to be able to make inferences on the characteristics of a time series (such as its variance or autocovariance function), certain assumptions must be imposed on its evolution. Most often, the process is assumed to be such that if we divide any of its realizations into smaller sections, then each section looks much like any other section of that realization, i.e. the statistical properties of the time series do not change with time. Such processes are called (strictly) stationary time series, and many excellent monographs are entirely devoted to studying them — see, for instance, Priestley (1981), Chatfield (2004) or Brockwell and Davis (2009).

We emphasize that, in practice, it is not always reasonable to assume that time series are stationary. However, once the stationarity assumption is dropped, other assumptions on the process, although less restrictive, still have to be imposed in order to be able to make inferences on the process characteristics, such as its evolving variance or autocovariance structure.

Throughout this article we shall concentrate on trend-free processes with a second order structure that varies slowly with time. Such time series appear to have a stationary behaviour over short stretches of time and so are called *locally stationary* (Dahlhaus, 1997; Nason and Von Sachs, 1999).

Dahlhaus (1997) introduced a new concept of rescaled time to provide a framework with which asymptotic process inference could be made: controlling the evolution of the individual amplitudes of the locally stationary process through

a function dependent on rescaled time ensures that its statistical characteristics, e.g. the autocovariance function or the process spectral density, can be (locally) estimated by pooling the observed data over the regions of local stationarity.

2.2 Locally stationary wavelet (LSW) processes

Wavelets have been so far used for a wide variety of problems that arise in time series analysis. For a review of the use of wavelets for time series analysis, see Nason and Von Sachs (1999) or the comprehensive monograph by Percival and Walden (2000).

Due to their nature, wavelets deliver a time–scale representation, complementary to the time–frequency interpretation that arises from a Fourier analysis and so the classical Fourier spectral analysis can be complemented by a wavelet spectral analysis.

This article builds upon the work of Nason et al (2000), who proposed a new way to model time series with a time-dependent second order structure, based on the concept of rescaled time of Dahlhaus (1997) and a family of discrete nondecimated wavelets $\{\psi_{j,k}(t)\}_{j,k}$, which replaces the set of sine and cosine waves in traditional Fourier analysis. Instead of assuming a stationary process behaviour, their process is assumed to have a stationary character *locally*, by constraining the model coefficients to change slowly within each scale. The authors refer to processes built as above under the name of locally stationary wavelet (LSW) processes.

In what follows we give the main points of the formal definition of a LSW process, and the interested reader can refer to Nason et al (2000) for the complete definition.

Definition 2.1 *A sequence of stochastic processes $\{X_{t,T}\}_{t\in\overline{0,T-1}}$, $T = 2^{J(T)}$ is a zero-mean LSW process if it admits the following representation*

$$X_{t,T} = \sum_{j=-J(T)}^{-1} \sum_{k\in\mathbb{Z}} w_{j,k;T}\psi_{j,k}(t)\xi_{j,k}, \tag{1}$$

where $\psi_{j,k}(t)$ is a nondecimated discrete wavelet at scale j and location k, $w_{j,k;T}$ is its corresponding amplitude and $\{\xi_{j,k}\}_{j,k}$ is a sequence of zero-mean, orthonormal random variables.

Within each scale j, the evolution of the amplitudes $\{w_{j,k;T}\}_{k\in\overline{0,T-1}}$ is regulated by the Lipschitz continuous function $W_j(\cdot)$, defined for rescaled time $z = \frac{k}{T}$.

Note that we (somewhat abusively) refer to the non-random component of the building block coefficients under the name of amplitudes. The functions $\{W_j(\cdot)\}_j$ control the degree of local stationarity of the process by forcing the amplitudes $\{w_{j,k;T}\}_k$ to vary slowly within each level.

The LSW process defined above has an associated *evolutionary wavelet spectrum* (EWS) $\{S_j(\cdot)\}_{j\in\overline{-J(T),-1}}$ that can be defined by

$$S_j(z) = |W_j(z)|^2 = \lim_{T\to\infty}|w_{j,\lfloor zT\rfloor;T}|^2, \tag{2}$$

where $z \in (0,1)$ and $\lfloor zT\rfloor$ denotes the largest integer not exceeding zT. The spectrum quantifies the contribution to the process variance made at location z and scale j.

For fixed T, the autocovariance of the process $(X_{t,T})_{t\in\overline{0,T-1}}$ depends both on the lag, τ and on the rescaled time location, z, and it is denoted by $c_T(z,\tau) = \mathrm{cov}(X_{\lfloor zT\rfloor}, X_{\lfloor zT\rfloor+\tau})$. Nason et al (2000) show that the autocovariance function $c_T(\cdot\,,\cdot)$ tends to an (asymptotic) local autocovariance $c(\cdot\,,\cdot)$: $|c_T(z,\tau) - c(z,\tau)| = O(T^{-1})$, where $c(z,\tau)$ is defined in the following.

Definition 2.2 *The local autocovariance (LACV) function of a LSW process defined in Definition 2.1 is given by*

$$c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau), \tag{3}$$

where $\Psi_j(\tau) = \sum_{k=max\{0,\tau\}}^{L_j-1+min\{0,\tau\}} \psi_{j,k}(0)\psi_{j,k}(\tau)$, $\tau \in \mathbb{Z}$ is the discrete autocorrelation wavelet at scale j.

Although representation (1) of a LSW process is not unique, the EWS is unique in terms of the local autocovariance, and vice versa (Nason et al (2000)).

The linear independence of the family $\{\Psi_j(\cdot)\}_{j\le -1}$ ensures the invertibility of the covariance–spectrum representation:

$$S_j(z) = \sum_{l=-\infty}^{-1} A^{-1}_{j,l}\left(\sum_{\tau} c(z,\tau)\Psi_l(\tau)\right), \tag{4}$$

where $A_J^{-1} = (A^{-1}_{j,l})_{j,l\in\overline{-J(T),-1}}$ is the inverse of the matrix A_J previously introduced (Nason et al, 2000). Formulae (3) and (4) are the analogues of the Fourier pair relationship between classical spectrum and autocovariance.

If $\hat{S}_j(z)$ denotes a spectrum estimator, then by taking $\hat{c}(z,\tau) = \sum_{j=-J(T)}^{-1}\hat{S}_j(z)\Psi_j(\tau)$ we obtain an estimator for $c(z,\tau)$. For certain choices of $\hat{S}_j(z)$, the estimator $\hat{c}(z,\tau)$ enjoys good properties, such as consistency (see Proposition 5 of Nason et al (2000)), which motivates obtaining a well-behaved estimator for the wavelet spectrum.

2.3 Spectral estimation for LSW processes

Nason et al (2000) introduced the *wavelet periodogram of a LSW process* $(X_{t,T})_{t\in\overline{0,T-1}}$ (constructed with respect to the nondecimated discrete wavelet family $\{\psi_{j,k}(t)\}_{j,k}$) given by:

$$I^j_{k,T} = d^2_{j,k;T}, \tag{5}$$

where $d_{j,k;T} = \sum_{t=0}^{T-1} X_{t,T}\psi_{j,k}(t)$ is the empirical wavelet coefficient at scale j and location k.

For $z \in (0,1)$, let us denote the (vector) wavelet periodogram by $\underline{I}_T(z) = (I^j_{\lfloor zT \rfloor,T})_{j\in\overline{-J(T),-1}}$. Similarly, the (vector) evolutionary wavelet spectrum is denoted by $\underline{S}(z) = (S_j(z))_{j\in\overline{-J(T),-1}}$.

Nason et al (2000) show that

$$\mathbb{E}(\underline{I}_T(z)) = A_J\underline{S}(z) + O(T^{-1}),\ z \in (0,1), \tag{6}$$

which implies for $z = \frac{k}{T}$,

$$\mathbb{E}(I^j_{k,T}) = \sum_{l=-J(T)}^{-1} A_{j,l}S_l\left(\frac{k}{T}\right) + O(T^{-1}). \tag{7}$$

Hence the expected value of the wavelet periodogram is (asymptotically) a linear combination of wavelet spectra, and Nason et al (2000) propose using a *corrected vector of periodograms*, $\underline{L}(z) = (L^j_{\lfloor zT \rfloor,T})_{j\in\overline{-J(T),-1}}$ for estimating $\underline{S}(z)$:

$$\underline{L}(z) = A_J^{-1}\underline{I}_T(z).$$

Relation (6) shows that $\underline{L}(z)$ is asymptotically an unbiased estimator for the evolutionary wavelet spectrum, $\underline{S}(z)$ for all $z \in (0,1)$. However, Nason et al (2000) show that $\underline{I}_T(z)$ has an asymptotically non-vanishing variance, so it is not a consistent estimator for the wavelet spectrum. To obtain consistency, $I^j_{\lfloor zT \rfloor,T}$ will be first smoothed as a function of z within each scale j. Then correction with A_J^{-1} of the smoothed $\underline{I}_T(z)$ will provide a wavelet spectrum estimator, $(\hat{S}_j(z))_{j\in\overline{-J(T),-1}}$. For properties of this estimator, the reader is referred to Nason et al (2000).

3 Spectral estimation for LSW processes with missing observations

We now derive an estimate for the evolutionary wavelet spectrum, when the observed LSW process features missing observations.

3.1 LSW processes with missing observations

Assume that for some T we observe $(X_{t,T})_{t\in\overline{0,T-1}}$, where $X_{t,T}$ admits the representation from Definition 2.1,

$$X_{t,T} = \sum_{j=-J(T)}^{-1}\sum_{k\in\mathbb{Z}} w_{j,k;T}\psi_{j,k}(t)\xi_{j,k},$$

but unlike before, we do not have an observed value $X_{t,T}$ for each $t \in \overline{0,T-1}$, i.e we start with a realization of a LSW process which features missing observations: at some time points we do not have the corresponding X values.

Let us denote the set of time points corresponding to observations on the process by $\mathscr{S} = \{t_1, t_2, \ldots, t_n\} \subseteq \{0,1,...,T-1\}$. We will use the notation $I_{\mathscr{S}}$ for the vector of $(I_{t_1}, I_{t_2}, \ldots, I_{t_n})$, and similarly $I_{\bar{\mathscr{S}}} = (I_t)_{t\in\bar{\mathscr{S}}}$ for the set of missing time points, where $\bar{\mathscr{S}} = \{0,1,...,T-1\}\setminus\mathscr{S}$.

For a future asymptotic theory to make sense, the elements of $\mathscr{S}$ cannot be constrained to belong to a fixed interval, and their number must increase with T, see Hall et al (1994). To reflect this we shall model $\mathscr{S}$ as follows. We define T independent identically distributed Bernoulli random variables which model the appearance of each time point for each $t \in \overline{0,T-1}$, say, $I_t \sim$ Bernoulli(p) by which we mean that each time point has probability $(1-p)$ of being missing. In this setting, the number of observations on the process, that is, the number of elements in $\mathscr{S}$, $|\mathscr{S}| = \sum_{t=0}^{T-1} I_t = n$, is a random variable $n \sim \text{Bin}(T,p)$. Therefore, the number of observations is in fact a function of T, $n(T)$, but to avoid notational clutter we denote it by n throughout the paper.

For a LSW process, defined as a sequence of stochastic processes (see Definition 2.1), there are two ways in which the locations of the missing values can arise for different values of T – we can either assume that the locations change with T, or that the missing time locations corresponding to the smaller T are fixed. These issues need to be further considered for an asymptotic development. Throughout the paper we will be working conditional on the time locations corresponding to observations on the process being fixed. In other words, we will assume that in practice we have available information at the n locations, $t_1, \ldots, t_n$ and we ignore the random character of these locations.

3.2 Wavelet periodogram in the missing data setting

The methodology of Nason et al (2000) for estimation of the process characteristics of interest (such as the EWS or the LACV) centres on obtaining the wavelet periodogram (defined by equation (5)) by computing the nondecimated empirical wavelet coefficients $d_{j,k;T}$ at each scale j and location k. It is clear that these classical wavelet formulae cannot be directly applied in the context of missing data. For this reason, we propose using a wavelet decomposition based on second generation wavelets, able to deal with irregularly-spaced data and hence able to produce empirical wavelet coefficients (details) at the observed time points.

3.2.1 The lifting scheme (LOCAAT)

Second generation wavelets are essentially a generalization of 'classical' wavelets, designed to cope with irregular settings or with data that is not of a dyadic length. Our approach uses wavelets constructed via the lifting scheme that

'removes one coefficient at a time' (LOCAAT) of Jansen et al (2001), and explored in Nunes et al (2006).

Briefly, the aim of the lifting scheme is to transform a function sampled at n irregularly-spaced locations (which we denote by $\{(x_i, f_i)\}_{i\in\overline{1,n}}$) into a set of say, L scaling and $(n-L)$ wavelet coefficients, where L is the desired primary resolution level. The algorithm is usually represented by recursively applying three steps: *split*, *predict* and *update*.

The *split* step consists in choosing a point to be removed, and essentially Jansen et al (2001) propose to remove points in an order dictated by the x-configuration: those points corresponding to denser areas are removed first, and further steps generate detail in progressively coarser areas. Each location is therefore associated with an interval which it intuitively 'spans': the shorter the interval, the more densely sampled the area around the location is.

The value of the function (f) is then *predicted* at the point selected for removal based on regression over its neighbourhood, and the prediction error will be the *detail coefficient* corresponding to that location.

In the *update* step, the f-values of the neighbouring points are updated by using a linear combination with the detail coefficient, such that the mean signal stays the same throughout the algorithm application. At this stage the lengths of the intervals associated to the neighbouring points also get updated in order to account for the decreasing number of scaling points that remain to 'span' the interval and accordingly, scale now has a continuous character. Jansen et al (2001, 2004) propose an artificial split into levels to mimick discrete scales from the classical wavelet setting, where each point uniquely corresponds to a scale.

In summary, the lifting scheme produces exactly one detail coefficient at each observed x-point, which is in turn associated to one (artificial) scale.

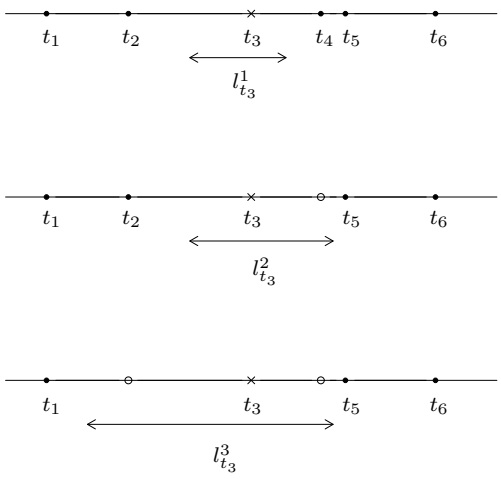

Fig. 1 Relationship between scale and order of point removal in the modified LOCAAT lifting algorithm. Top: t_3 (x) is the first point to be lifted. The scale associated to the detail coefficient $d^1_{t_3}$ is shown below it ($l^1_{t_3}$). Middle: t_3 is lifted after point t_4. Part of the interval corresponding to t_4 is redistributed to its neighbours and so the associated scale $l^2_{t_3}$ for $d^2_{t_3}$ increases. Bottom: t_3 is removed after both t_2 and t_4. In this situation, the detail coefficient $d^3_{t_3}$ has a large scale $l^3_{t_3}$ associated to it. By varying the order of point removal during the LOCAAT algorithm, the nondecimated lifting transform (NLT) is able to generate multiple detail coefficients for each location with different associated scales.

3.2.2 The nondecimated lifting transform (NLT)

A nondecimated transform in the lifting 'one coefficient at a time' context was introduced in Knight and Nason (2009), who proposed a technique that produces a set of wavelet coefficients associated to each x-location, at various artificial scales. This is known as the nondecimated lifting transform (NLT) for irregularly-spaced data and replaces the nondecimation in the classical, regularly spaced context.

Simply put, the application of the LOCAAT wavelet algorithm of Jansen et al (2001) transforms an irregular observed dataset into a set of wavelet (detail) coefficients, such that there is one detail coefficient corresponding to each 'lifted' point. In the approach introduced by Jansen et al (2001), the order of transforming scaling coefficients into detail coefficients is established by using the integral lengths of the scaling functions, which account for the 'span' of each point.

The NLT approach of Knight and Nason (2009) allows for full flexibility in choosing the order of obtaining the detail coefficients. The LOCAAT transform is modified to accommodate a *random* order of generating the wavelet coefficients, while the prediction and update steps are left unchanged. This transform is then repeatedly applied, every time following a different order of removing the points, and consequently generating the detail coefficients. A 'large enough' sample of these permutations ensures that a distribution of the empirical wavalet coefficients associated to each location is obtained.

Let us now formalize the NLT approach. The n observed time points $t_1, ..., t_n$ can be arranged in (ordered) vectors of length n in $n!$ ways. Out of this sample space, we randomly extract say m such orderings (trajectories), which will give the paths that the lifting algorithm will take.

For each selected trajectory, the modified lifting transform will generate a set of detail coefficients. Let us denote the n-dimensional (row) vector of detail coefficients by $\underline{d}^T = (d_{t_i,T})_{i\in\overline{1,n}}$. Using the matrix representation of the

wavelet transform, we can write

$$\begin{pmatrix} d_{t_1,T} \\ \cdots \\ d_{t_n,T} \end{pmatrix} = R \begin{pmatrix} X_{t_1,T} \\ \cdots \\ X_{t_n,T} \end{pmatrix},$$

where $R \in \mathscr{M}_{n,n}$ is the matrix built during the lifting transform.

From the above it follows that $d_{t_i,T} = \sum_{j=1}^{n} r_{i,j} X_{t_j,T}$, $\forall i \in \overline{1,n}$, and each detail is a linear combination of the observed $X_{t_i,T}$'s, $i \in \overline{1,n}$.

The vector of details, $\underline{d}^T$ has a random character, inherited from the process $(X_{t_i,T})_{i\in\overline{1,n}}$. The elements of the matrix R depend on the prediction and update filters, which (for a linear transform) in turn depend only on the time locations and the regression order used in the prediction step (see Nunes et al (2006)). Therefore, since we work conditional on having fixed design points, $(t_k)_{k\in\overline{1,n}}$, the elements of the matrix R can be assumed fixed.

For each of the m 'paths', we apply the lifting algorithm and hence generate m different matrices $R^1, ..., R^m$. Correspondingly, we get m sets of wavelet vectors $\underline{d}^{1,T}, ..., \underline{d}^{m,T}$.

Each time location t_k is therefore associated to a set of detail coefficients, $\{d^{\alpha}_{t_k,T}\}_{\alpha\in\overline{1,m}}$. Each detail $d^{\alpha}_{t_k,T}$ is associated to an interval that intuitively accounts for the 'span' of time location t_k at the respective stage in the algorithm (see Figure 1). We shall denote the length of this interval by $l^{\alpha}_{t_k,T}$, and this will provide our measure of scale. Therefore, at *each* time location we obtain *a set* of details, which we can model as a function of their scale.

3.2.3 Proposed raw wavelet periodogram construction

In Section 2.3 we introduced the raw periodogram proposed by Nason et al (2000) for estimating the evolutionary wavelet spectrum (equation (5)). We saw that this was an array filled in with the values of the squared nondecimated detail coefficients corresponding to each level and time location, where the level had the usual multiresolution meaning on a $\log_2(T)$ scale. To obtain consistency, the raw periodogram was then first smoothed as a function of location within each scale and then was corrected by A_J^{-1} in order to provide an unbiased spectrum estimator.

In our context, the challenge comes from the irregularly-spaced data that hinders the application of classical wavelet techniques, such as nondecimation. Our aim is to propose a periodogram for estimating the wavelet spectrum of an LSW process despite dealing with a realization that features missing observations.

We now summarize the steps that we propose in order to obtain a raw periodogram for an LSW process sampled with missing data (see flowchart in Figure 2):

1. Apply the NLT of Knight and Nason (2009) introduced in the section above on the sampled process data, and obtain a set of detail coefficients at each observed time point.
2. The details associated to each location are in fact associated to various scales, as we explained in the previous section. Consequently we shall 'discretize' the scale in order to ensure comparability specifically with the construction from Nason et al (2000), and more generally with classical wavelet constructions. More exactly, we shall choose a set of 'evaluation' lengths, which we denote by $l^1, l^2, \ldots, l^{J^*}$ for some J^*, and through the choice of J^* we are in fact able to tune the proposed discreteness of the scale.
3. We want to estimate the function that links the squared detail coefficients with the scale at which they arise, in order to produce an estimate of the squared detail at each chosen scale l^i with $i \in \{1, \ldots, J^*\}$ and at each observed time location t_k with $k \in \{1, \ldots, n\}$.
   This can be achieved by taking a nonparametric regression approach in modelling the magnitude of the associated squared details $(d^1_{t_k,T})^2, \ldots, (d^m_{t_k,T})^2$ as a function of the corresponding interval lengths $l^1_{t_k,T}, \ldots, l^m_{t_k,T}$ for each fixed location t_k (refer to Figure 4). For each time location t_k, we denote by $f_{t_k,T}$ the function we want to estimate. In other words, for each t_k with $k \in \overline{1,n}$, we model the data as
   $$(d^{\alpha}_{t_k,T})^2 = f_{t_k,T}(l^{\alpha}_{t_k,T}) + \varepsilon_{\alpha}, \quad \alpha \in \overline{1,m}, \tag{8}$$
   and we want to obtain an estimate $\hat{f}_{t_k,T}(l^i)$ for each $i \in \overline{1,J^*}$. We estimate each $f_{t_k,T}(\cdot)$ by using a linear smoother, hence
   $$\hat{f}_{t_k,T}(l^i) = \sum_{\alpha=1}^{m} K_{\alpha}(l^i)(d^{\alpha}_{t_k,T})^2, \quad \forall i \in \overline{1,J^*}, \tag{9}$$
   where $K_{\alpha}(l^i)$ are weight functions that are non-zero only for those α values such that $l^{\alpha}_{t_k,T}$ is in a neighbourhood of l^i. We note that the weights $K_{\alpha}(\cdot)$ are different for each t_k, but we do not indicate it to avoid cluttering the notation. The above value of $\hat{f}_{t_k,T}(l^i)$ is an estimate of the magnitude of the squared detail $(d^{\alpha}_{t_k,T})^2$ at time t_k associated to the interval length l^i.
4. The matrix $(\hat{f}_{t_k,T}(l^i))_{i\in\overline{1,J^*},k\in\overline{1,n}}$ is our proposed raw periodogram, and corresponds to the raw periodogram $(d^2_{j,k;T})_{j\in\overline{-J(T),-1},k\in\overline{0,T-1}}$ introduced by Nason et al (2000).

Section 5 will show that our proposed raw periodogram is not an unbiased estimator for the EWS and will discuss the technical and computational challenge of correcting it.

3.3 Periodogram applicability and smoothing

Let us now make a few remarks on the periodogram construction.

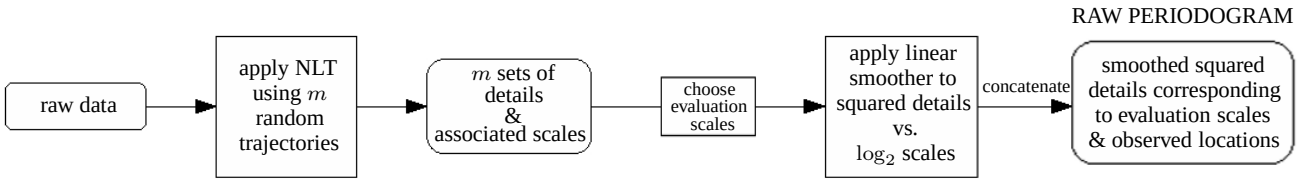

Fig. 2 Flowchart for constructing the NLT periodogram: (i) The modified nondecimated lifting algorithm (NLT) is applied to the raw observations using a fixed number of random lifting trajectories to form sets of detail coefficients and corresponding scales for each observation (ii) For each observation in turn, a linear smoother is applied to the NLToutput to predict the values of the squared details at a set of chosen evaluation scales (iii) The smoothed values for all observations are then concatenated (columnwise) to form a matrix which represents the raw periodogram.

Applicability. The construction is valid for irregularly-sampled locally stationary time series, not necessarily of a dyadic length, and in particular, is suitable for those time series whose irregular structure is induced from a regular time series with missing observations. Examples for both cases are shown in Section 4.

The periodogram construction also lends itself to extensions using data from repeated time series using the lifting transform modifications for multiple observations described in Nunes et al (2006). An alternative approach would be to simply average raw periodograms via merging time indices.

Scale interpretation. In classical wavelet theory, the notion of scale has a meaningful interpretation associated to the support of the wavelet, and, since the discrete wavelet transform (DWT) is defined on equispaced grids of dyadic length, the coefficients at a particular scale represent a dyadic powered number of the observations through decimation. This meaning transfers to the periodogram for the regular setting.

However, due to the lifting of one coefficient at each step, in the most general case our periodogram will not feature a dyadic structure in the scaling of wavelet support or have a dyadic number of coefficients represented from one scale to the next, although the scale is still connected to the wavelet support. In practice, to ensure that our final periodogram representation parallels the classical one, instead of the actual scale values, we use their $\log_2$ values. This has the interpretation that if the scale increases by one, the *"average"* wavelet function support doubles.

Smoothing. Nason et al (2000) approach the problem of the biasedness of their periodogram by first smoothing the periodogram via non-linear (translation-invariant) wavelet smoothing on the periodogram values, and then applying an inverse correction matrix. An alternative method using the Haar-Fisz transform has been proposed to smooth the periodogram (Fryźlewicz and Nason, 2006; Nason, 2008).

Ideally we would like to smooth our nondecimated lifting periodogram per level over time. However, due to the missing data in our framework, the data have an inherent irregularly-spaced structure and classical wavelet smoothing methods (such as those used for the usual regular time series setting) are not applicable here. Sanderson (2010) copes with this by averaging the periodogram values over a number of resolution bands to mirror the regular setting, and then smoothing their averaged periodogram over time by employing a simple moving average.

Since each periodogram value is produced from a specific run of a lifting transform, a natural approach to consider would be to smooth the squared coefficients across time by first denoising the detail coefficients within the run that produced them. This would not require pooling coefficient information from different NLT runs prior to smoothing as in the averaging approach of Sanderson (2010). For example, smoothing the periodogram values could be obtained by using lifting transforms to denoise the data (Nunes et al, 2006; Knight and Nason, 2009) and then pre- and post-transforming the values using the logarithm transformation. However, it remains unclear what strategy is best to consider when smoothing the periodogram over time, and is left as an area of future research.

4 Examples

We now give some illustrative examples of our periodogram for irregular time series. Our periodograms are produced in R by the lifting algorithm implementations *adlift* (Nunes and Knight, 2010) and *nlt* (Knight and Nunes, 2010).

4.1 Simulated example

Let us take the evolutionary wavelet spectrum $\{S_j(\cdot)\}_{j\leq -1}$, described in formula (10), which at the finest level (-1) exhibits a burst of activity, and at the coarser level -4 exhibits a squared sinusoidal behaviour (see Figure 3).

$$S_j(z) = \begin{cases} 1, & \text{for } j=-1,\ z\in(\frac{180}{256},\frac{209}{256}), \\ \sin^2(4\pi z), & \text{for } j=-4, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

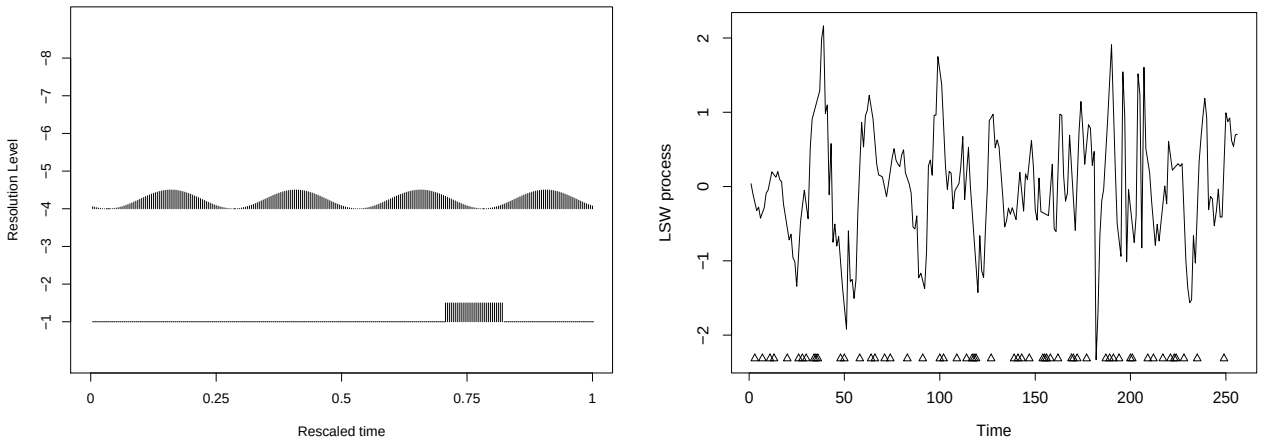

Fig. 3 Left: An evolutionary wavelet spectrum (EWS) with power at fine and mid-scales. The scale (y-axis) runs from finest (bottom) to coarsest (top); right: simulated LSW process corresponding to the spectrum featuring missing observations. Triangles indicate locations of missing time points.

Using the *LSWsim* function implemented in the R add-on package *WaveThresh* (Nason et al, 2008), we first simulate a LSW process of length $T = 256$ corresponding to the above spectrum. We take a random sample of $n = 200$ time points out of the 256, and then record their corresponding 'observed' process values in order to obtain an example of a LSW process with missing observations. An example of such a process (with missing data) appears in Figure 3 (right), and is represented on an irregular time grid. Note that the sinusoidal character might be guessed in the first half of the realization, but in the second half the burst masks it.

The locations of the 56 missing time points are also represented in the figure. Note also that the region which features the activity burst has slightly more missing observations, which will probably influence the accuracy of the final spectrum estimator. Also, the estimation will be influenced by the overall proportion of missing observations.

For the lifting procedure there are 256! possible removal orderings of the time points that can be used in order to generate the detail coefficients. In what follows we take a simple random sample of $m = 1000$ trajectories out of the total 256!. Each trajectory gives the order in which the empirical lifted wavelet coefficients will be produced. For each case, we modify the lifting scheme such that it follows the corresponding random path, and use a prediction step that employs linear regression with an intercept and with 2 neighbours in a symmetrical configuration in order to generate the details at each step (see Nunes et al (2006)).

Let us now examine the relationship between details produced during the NLT transform and their associated scales. As an example we take the observation at time 15. Figure 4 shows the distribution of squared detail coefficients and

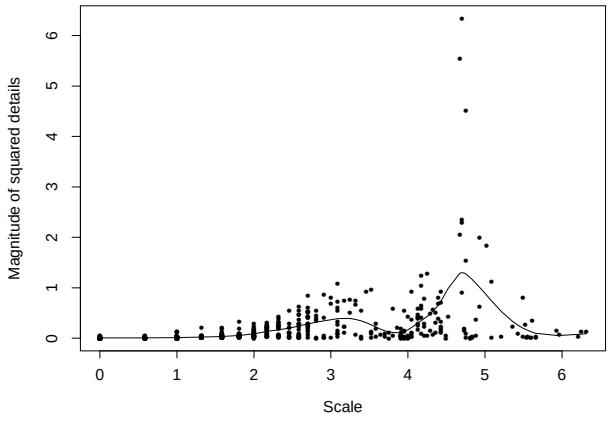

Fig. 4 Magnitude of the squared details associated to the observation at time 15, versus ($\log_2$ of) their associated integral lengths. The superimposed curve is obtained by smoothing using a cubic spline.

associated integral lengths for this point resulting from our 'nondecimated' lifting algorithm with $m = 1000$ trajectories. The cubic spline smoothing estimator is also shown.

Figure 5 gives two examples of our periodogram construction outlined in Section 3.2.3, each corresponding to a different 'evaluation' scale. The scale in the right picture is finer, and it essentially shows that we can tune the level of detail given by the periodogram. This zoom in – zoom out feature of wavelets is nicely reflected in our construction at the estimated spectrum level.

The range of these scales is roughly 0 to 8 (in a continuous manner, with smaller values corresponding to finer scales), and an approximate correspondence can be established with the initial discrete levels $-1, \ldots, -8$. Note though

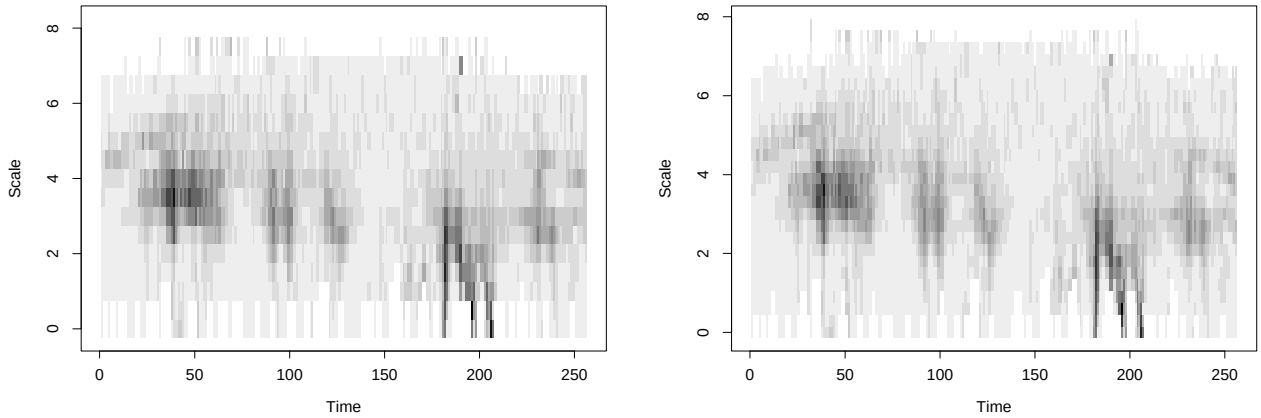

Fig. 5 Proposed 'raw' wavelet periodograms to estimate the wavelet spectrum of Figure 3. The smoothed squared details are represented on two different 'evaluation' scales (with 17, respectively 27 equally spaced divisions for the interval 0–8), with the picture on the right corresponding to the finer scale division. In each plot, the scale gets coarser from bottom upwards and darker pixels correspond to a higher estimated spectrum value.

that since not all time points are associated to integral lengths spanning the whole 0 to 8 range, missing values appear at the bottom and top rows of the matrices represented in Figure 5. However, the use of the finer scale diminishes this problem to some extent. Observe that both the burst and the four squared sinusoidal peaks are detected within the correct scales. The region approximately between times 150 to 175 does not contain much signal, which can be explained by the slightly higher proportion of missing observations from that area than from the rest.

4.2 Environmental time series

4.2.1 Orbital forcing data

In this example, we demonstrate our periodogram construction on environmental time series. A well-known school of thought in astronomy accepts that the positioning of the Earth as it moves around its orbit has an effect on climatic events over time, in particular the characteristics of ice ages, through changes in the Earth's insolation (received solar energy) measurement (Crucifix et al, 2007; Crucifix, 2008). This *orbital forcing* has been quantified via certain trigonometrical astronomical parameters so that time series representing the forcing can be generated (Berger, 1978). Since the time series produced from the mathematical framework of Berger (1978) are based on quantities derived from the physics of the Earth's orbit, the time series is unobserved (it is calculated), and thus contains no uncertainty or source of measurement error; this makes it a good candidate for testing whether the periodogram applied to these orbital forcing data can extract what is known about the cyclic variability components of the time series from orbital mechanics. It is also of interest to determine whether our spectral estimation method detects other frequency variance contributions that do not feature in other glacial characteristic investigations, namely energy at higher scales. The data was originally explored in Berger (1978).

The methodology as described in the flowchart (Figure 2) was applied to this orbital forcing data using $m = 5000$ trajectories in the nondecimated lifting algorithm. The resulting periodogram shows a clear periodic structure at mid-coarse scales (see Figure 6). This is the so-called *precession forcing* due to the movement of the periapsis between the Earth and the sun (Crucifix, 2008), and has an an average period of 21kyr (see Sanderson (2010)). This behaviour is thus evident at approximately at scale $14.4 = \log_2(21 \times 10^3)$. Note that the periodogram also exhibits characteristic 'troughs' representing the periodogram power leakage across the scales.

4.2.2 Trace carbon dioxide time series

Other paleoclimatic time series, for example, gas/isotope content in drilled ice-cores can also be used to map climatic events through history. Time series obtained from ice-cores are characterized by an uneven sampling rate, as deep within the ice-core, the snow/ice is under a strong mass pressure that results in depletion, pinching and swelling of layers (see Wolff (2005) for more details on the particular features of ice-core records). It is thus of interest to determine whether time series from these other sources can yield similar climatic information as orbital forcing time series. The comparison between orbital forcing and other ice-core time series has been investigated using a smoothed version of our lifting periodogram to some effect in Sanderson (2010).

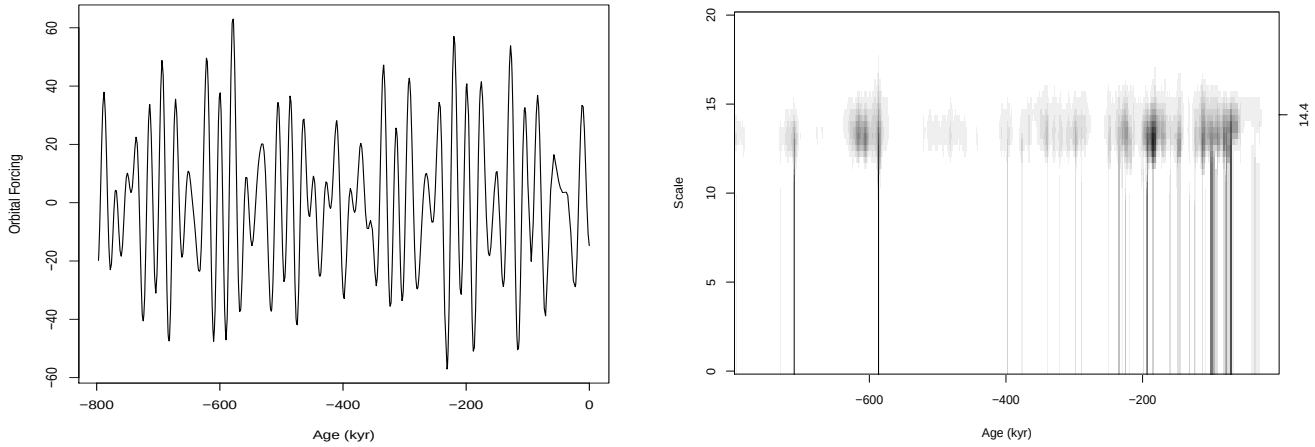

Fig. 6 Left: Plot of orbital forcing time series; right: periodogram of the orbital forcing time series using the algorithm in Section 3.2.3.

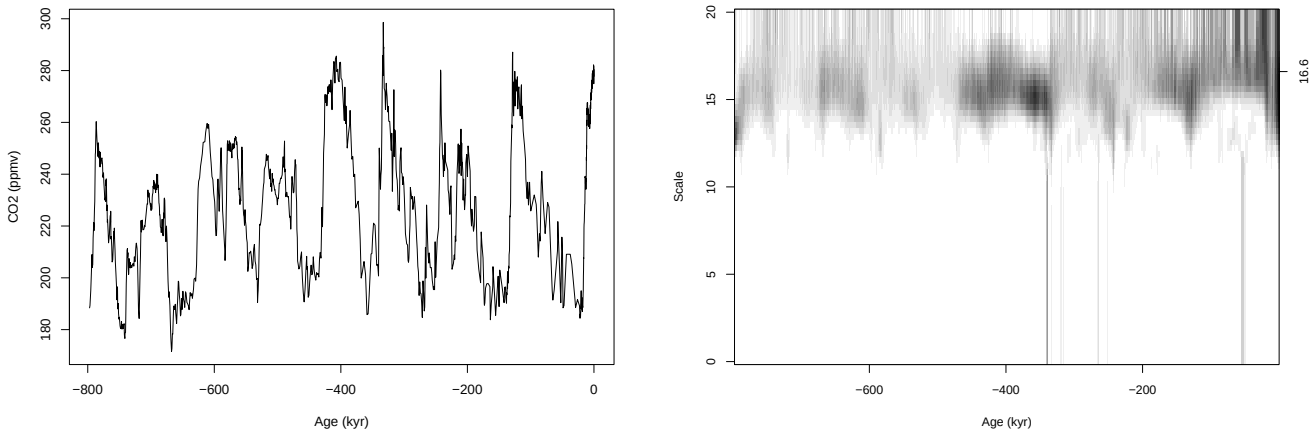

Fig. 7 Left: Plot of ice-core carbon dioxide content (parts per million by volume) versus associated historical age; right: NLT periodogram of the carbon dioxide time series.

The trace amount of trapped carbon dioxide in an ice-core can give indications of atmospheric changes during the time-span of the ice-core. We investigated a trace carbon dioxide time series with the NLT periodogram from Section 3.2.3 with $m = 5000$ paths. The time series has been analyzed by Lüthi et al (2008) was obtained from the World Data Center for Paleoclimatology in Boulder, USA.

Although some coarse-scale blurring is present, a periodic structure of roughly 100 kyr can be seen from the periodogram shown in Figure 7 (marked on the right axis at scale $\log_2(100 \times 10^3) = 16.6$)). The periodicity is clearest during the second half of the series; this spectral information agrees with evidence indicated by other studies of historical climatic changes, which acknowledges that there has been a climatic 100kyr cycle over the last 500 kyr (Crucifix and Rougier, 2009).

4.3 Infant ECG data

In this section, we compare the lifting periodogram construction introduced in Section 3.2.3 with that of Nason et al (2000) for the regular data setting (see Section 2.3). The data to which we have applied both methods is a time series recording the heart rate (ECG) during sleep of a young infant. The data consist of 2048 regularly-spaced observations sampled at $\frac{1}{16}$Hz (see Nason et al (2000)). The dataset has been made available in the R add-on package *WaveThresh* (Nason et al, 2008).

To compare our periodogram with one applicable for regularly-spaced data, we randomly selected 10% of the data to be treated as missing. After removing the selected points, the ECG time series had $n = 1843$ irregularly-spaced observations (205 were treated as missing). The irregular time series can be seen in the top-right of Figure 8; the triangles

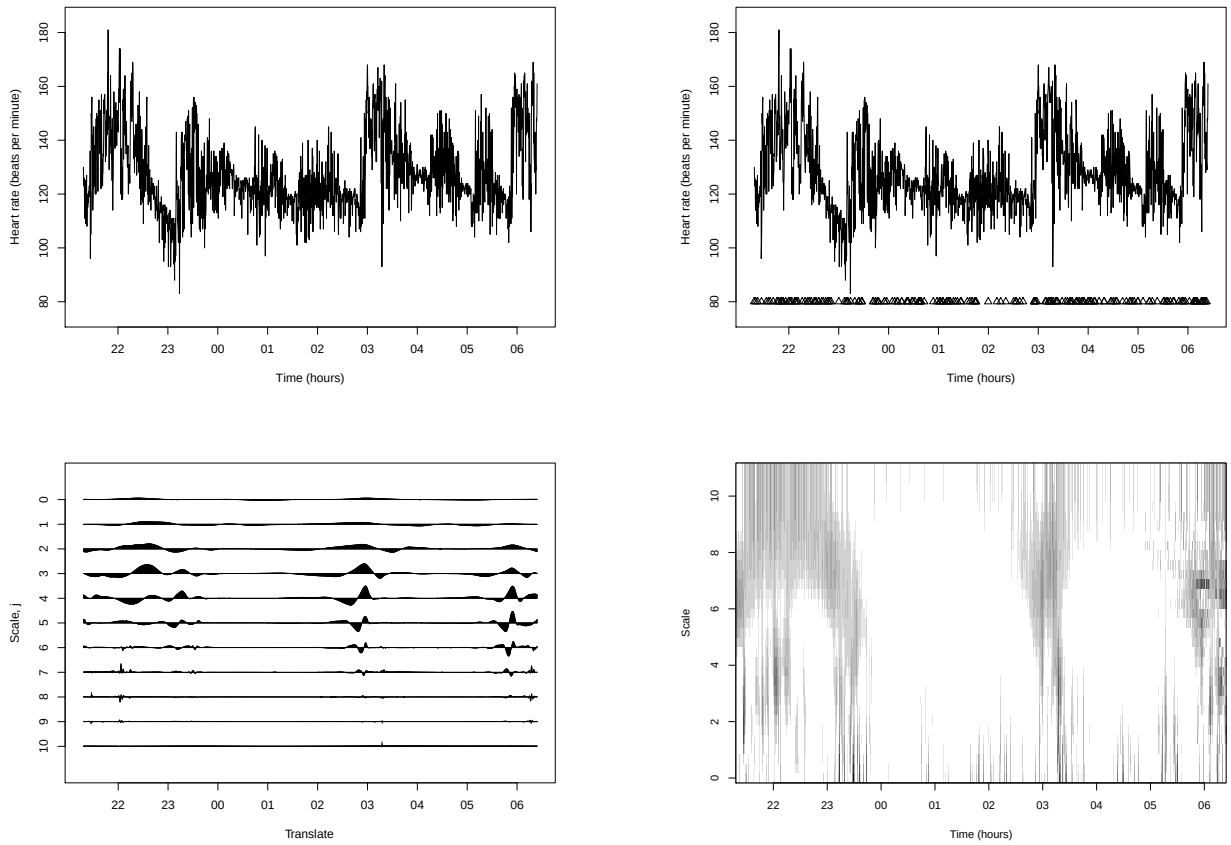

Fig. 8 Top left: original infant ECG time series; top right: infant ECG data with 205 artificially created missing values (shown below the time series); bottom left: the spectral estimate using the NvSK method; bottom right: NLT periodogram construction. For both periodograms the scale progress from fine at the bottom to coarse scales at the top.

below the time series represent the locations of the missing observations. We applied the lifting periodogram to the irregular ECG observations ($m = 5000$ lifting paths). This is shown in Figure 8 (bottom-right), together with the corresponding regular data setting spectral estimate of Nason et al (2000) from Section 2.3 (using all 2048 observations). We denote this spectral estimate by NvSK.

We consider that the periodogram displays similar spectral characteristics as that of Nason et al (2000) using the full dataset (Figure 8, bottom-left). In particular, the main features of the NvSK periodogram are evident in the NLT periodogram, namely at the mid-coarse scales (approximately at times 22:00, 23:15, 03:00 and 06:00). Furthermore, the fine scale behaviour is also present (at 22:00, 03:00 and 06:00). However, it is debatable whether the spectral information in our periodogram at fine scales at time 23:15 is a real feature of the data: it is noticeable that the periodogram shows some spurious spectral information across observations at the extremes of the scale range (below scale 1 and near scale 10).

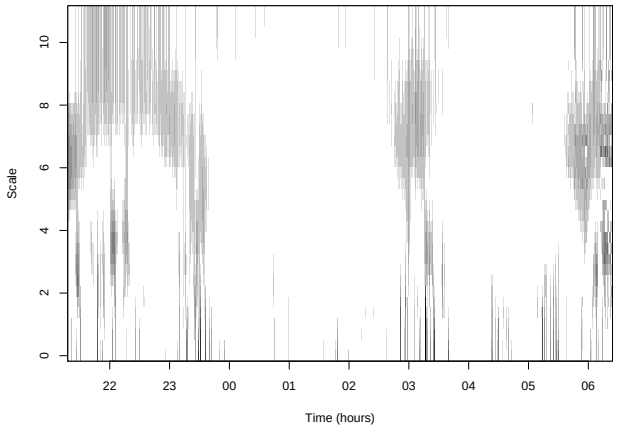

Fig. 9 Periodogram for infant ECG data with 25% of the observations deemed missing.

We have repeated this example with increased proportions of missingness up to 25% without any significant degradation in the spectral information shown in the resulting

NLT periodograms. The periodogram for the ECG data with 25% missing is shown in Figure 9.

To investigate how the NLT spectral estimates change with the amount of data considered missing from the infant ECG dataset, we performed the following study. In the first instance, the method described in Figure 2 was used to create a periodogram for the ECG time series with a single randomly chosen datapoint. We then calculated the (squared) error between this and the periodogram produced with an increased degree of missingness, treating the periodogram with one missing datapoint as 'the truth' (only the common subset of timepoints were considered in the error calculation). The motivation behind this comparison is that the calculation gives an overall (rather than pointwise) measure of how the periodograms change as the proportion of missing datapoints increases. The error calculation was then repeated $K = 100$ times (each repetition corresponding to independently resampling the datapoints to be considered missing) and then averaged over the K sampling runs. Note that in the calculations described, the periodograms were scaled prior to performing the error calculation so that they represented the same variance as the infant ECG data.

Table 2 shows the average mean squared error (MSE) of the degradation for different chosen amounts of missingness. The table reveals that, as expected, the fidelity to the 'true' periodogram decreases as the amout of missingness becomes more severe, with the degradation increasing slightly more significantly for higher percentages of missing data.

Missingness (%)	Error (MSE)
5	0.144
10	0.156
15	0.178
20	0.214
25	0.252

Table 1 Mean square error (MSE) results indicating degradation of infant ECG NLT spectral estimates for different degrees of missingness.

At this point we should stress that a periodogram for nonstationary processes with missing and/or irregular data is so far unavailable using the spectral estimation methods curently available in the literature. Indeed, the NvSK technique for classical LSW processes cannot be used even with only a single missing observation. As noted in Section 1, spectral estimation techniques often rely on imputation in order to handle missing observations, or in the simplest case, the data is treated as equispaced. This less principled way of obtaining periodograms for irregular data can introduce errors in estimation, especially with significant missingness (for example greater than 5%). It is thus reassuring that our raw periodogram correctly detects the spectral structure in this difficult situation, albeit with obvious power leakage across the scales.

Note that in the examples above the data has been randomly selected and treated as missing. If the missing points are clustered such that some feature of the data is lost, then the spectral estimate will not be able to capture the power activity in that region of the time series. However, this observation would be true of all spectral estimation methods for irregular data in this case.

5 Correction of raw periodogram (theory)

In this section, we establish the relationship between the initial (unknown) evolutionary wavelet spectrum and our proposed 'raw' periodogram, in order to make a step towards a bias-corrected periodogram. We then discuss the spectral correction of the periodogram and highlight its technical difficulty, reflected in its computational complexity. Future avenues for development of this periodogram correction are also considered.

5.1 Relationship between the proposed periodogram and the evolutionary wavelet spectrum

In what follows, we aim to obtain $\mathbb{E}(\hat{f}_{t_k,T}(l^i)|I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0})$. This can be viewed as establishing an equivalent formula in our setting to that from the LSW approach in Nason et al (2000) (see equation (7)). The following treatment is only meant to parallel such a formula and is not a rigorous asymptotic development. It is clear from the illustrations shown thus far (e.g. Figure 5) that some kind of blurring is present in our proposed periodogram, and the formula we derive in this section suggests that the blurring can, in principle, be corrected.

We shall first obtain the covariance structure of the wavelet coefficients as a function of the initial spectrum, both within the lifting scheme corresponding to each trajectory, and also between different trajectories. In other words, we are interested in $\mathrm{cov}(d^{\alpha}_{t_i,T}, d^{\beta}_{t_{i'},T}|I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0})$, $\forall i, i' \in \overline{1,n}, \forall \alpha, \beta \in \overline{1,m}$. Note that we are in fact also conditioning on the trajectories being fixed, rather than take into account their randomness, as this would in turn induce randomness in the matrices $R^1, ..., R^m$. This conditioning will be assumed for all results in this section, and so we omit noting it explicitly throughout the paper.

As a first step, the following lemma will establish a link between the variance–covariance matrix of the detail coefficients and the (sample) autocovariance matrix of the initial LSW process at the observed time points. The proof of this lemma and subsequent results in this section can be found in Appendix A.

Lemma 5.1 *Under the previous notation, the following relation holds: for any* $\alpha, \beta \in \overline{1,m},\ i,i' \in \overline{1,n}$ *,*

$$\operatorname{cov}(d^{\alpha}_{t_i,T}, d^{\beta}_{t_{i'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = \sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} \operatorname{cov}(X_{t_j,T}, X_{t_{j'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) r^{\beta}_{i',j'}. \tag{11}$$

We have expressed the detail coefficient covariance as a linear combination of *sample* autocovariances of the initial process $(X_{t,T})_{t\in\overline{0,T-1}}$ involving *only the observed locations*. We now extend this relation to express it in terms of the *local* autocovariance of the process.

Proposition 5.2 *For any* $\alpha,\ \beta \in \overline{1,m},\ i,\ i' \in \overline{1,n}$ *, we have*

$$\operatorname{cov}(d^{\alpha}_{t_i,T}, d^{\beta}_{t_{i'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = \sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} c\left(\frac{t_j}{T}, t_{j'} - t_j\right) r^{\beta}_{i',j'} + \tilde{R}_T, \tag{12}$$

where $\tilde{R}_T$ *is a term of order* $O(T^{-1})$.

Using the definition of the local autocovariance and equation (12), we can obtain an expression of the detail coefficient covariance in terms of the EWS of the LSW process, $\{S_l(\cdot)\}_l$, and the discrete autocorrelation wavelets $\{\Psi_l(\cdot)\}_l$ of Nason et al (2000) introduced in Section 2.2.

More exactly, by substituting the local autocovariance $c(z,\tau) = \sum_{l=-\infty}^{-1} S_l(z)\Psi_l(\tau)$ from equation (3) in equation (12), we obtain, for any $\alpha,\ \beta \in \overline{1,m},\ i,\ i' \in \overline{1,n}$,

$$\operatorname{cov}(d^{\alpha}_{t_i,T}, d^{\beta}_{t_{i'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = \sum_{l=-\infty}^{-1}\sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} r^{\beta}_{i',j'} \Psi_l(t_j - t_{j'}) S_l\left(\frac{t_j}{T}\right) + \tilde{R}_T. \tag{13}$$

In the above formula we used the symmetry around 0 of $\{\Psi_l(\cdot)\}_l$.

Fryźlewicz (2003) observes that in order to achieve the convergence of the autocovariance $c_T(\cdot,\cdot)$ (established in proposition 1 of Nason et al (2000)), the 'tail' of the sequence $\{S_j(\cdot)\}_{j\le -1}$ needs to be controlled. An approach to this would be to allow non-zero contributions to $\{S_j(\cdot)\}_{j\le -1}$ only from levels say $j \in \{-J', \ldots, -1\}$ for a large enough J', which would in turn mean that l would have a finite range in the above formula (and therefore also in the subsequent ones).

Equation (13) links the detail coefficient variances, and implicitly, $\mathbb{E}(d^{\alpha}_{t_i,T} d^{\beta}_{t_{i'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0})$, to the (unknown) wavelet spectrum at the observed time points, $\{S_l(\frac{t_j}{T})\}_{l,j}$, involving only tractable coefficients. We shall now re-write the previous expression in (13) in terms of this expectation.

Proposition 5.3 *For any* $\alpha,\ \beta \in \overline{1,m},\ i,\ i' \in \overline{1,n}$ *, we have*

$$\mathbb{E}(d^{\alpha}_{t_i,T} d^{\beta}_{t_{i'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = \sum_{l=-\infty}^{-1}\sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} r^{\beta}_{i',j'} \Psi_l(t_j - t_{j'}) S_l\left(\frac{t_j}{T}\right) + \tilde{R}_T, \tag{14}$$

where $\tilde{R}_T$ *is a term of magnitude* $O(T^{-1})$.

We can now obtain the expectation of the smoothed squared details, $E(\hat{f}_{t_k,T}(l^i) | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0})$, which will give us an insight into the relationship between our proposed periodogram and the wavelet spectrum of the process.

Theorem 5.4 *For the wavelet periodogram estimators* $\hat{f}_{t_k,T}(\cdot)$ *constructed in (9), and for all* $i \in \overline{1,J^*}$, $k \in \overline{1,n}$ *the following relation holds:*

$$\mathbb{E}(\hat{f}_{t_k,T}(l^i) | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = \mathit{Trace}(A^{l^i,k} S^T) + \tilde{R}^*_T, \tag{15}$$

where

$\tilde{R}^*_T = O(T^{-1})$,
$S = (S_{l,j})_{l\le -1, j\in\overline{1,n}}$ *with* $S_{l,j} = S_l(\frac{t_j}{T})$,
$A^{l^i,k} = (a^{l^i,k}_{l,j})_{l\le -1, j\in\overline{1,n}}$ *with*
$a^{l^i,k}_{l,j} = \sum_{j'=1}^{n}\left\{\sum_{\alpha=1}^{m} K_\alpha(l^i) r^{\alpha}_{k,j} r^{\alpha}_{k,j'}\right\}\Psi_l(t_j - t_{j'})$ *and*
$\{K_\alpha(\cdot)\}_\alpha$ *are as defined in equation (9).*

The result in the previous theorem corresponds to (7) in the development of Nason et al (2000). However, recall that our result is conditional on the time locations corresponding to the observations on the process being fixed, and on ignoring the randomness in the lifting trajectories. For further work, it would be interesting to try and eliminate these restrictions, as well as rigorously set a framework in which to investigate the asymptotic behaviour of our estimator.

Equation (15) shows that our proposed 'raw' periodogram is not an unbiased estimator for the wavelet spectrum, and it therefore needs correction. This does not come as a surprise, given the similar result that follows from (7) for the simpler case of observing a LSW process with no missing observations. Formula (15) also highlights that the used smoother will influence the amount of bias.

5.2 Periodogram correction

In this section we discuss the computational issues and implications arising from missing data when estimating the wavelet spectrum of a LSW process, and the challenge is to work towards correcting the proposed raw periodogram in the missing data setting.

Relation (15) indicates a way for proposing a better estimator (than the raw periodogram) for the spectrum matrix S

curtailed down to $J(T)$ rows, $S = (S_l(\frac{t_j}{T}))_{l\in\overline{-J(T),-1},j\in\overline{1,n}}$. To achieve this, we shall first re-write the unknown spectrum values S into vector format,

$$\underline{\tilde{s}} = ((S_{-1,j})_{j\in\overline{1,n}} \mid \cdots \mid (S_{-J(T),j})_{j\in\overline{1,n}}) \in \mathscr{M}_{1,J(T)\times n}.$$

On the same principle as above, let us put each matrix $A^{l^i,k}$ into vector format, as follows

$$\underline{\tilde{a}}^{l^i,k} = ((a^{l^i,k}_{-1,j})_{j\in\overline{1,n}} \mid (a^{l^i,k}_{-2,j})_{j\in\overline{1,n}} \mid \cdots \mid (a^{l^i,k}_{-J(T),j})_{j\in\overline{1,n}}),$$

where for each $l \in \overline{-J(T),-1}$, $(a^{l^i,k}_{l,j})_{j\in\overline{1,n}}$ is an n-dimensional row vector, and hence $\underline{\tilde{a}}^{l^i,k} \in \mathscr{M}_{1,J(T)\times n}$, $\forall k \in \overline{1,n}$, $\forall i \in \overline{1,J^*}$. For each observed point t_k, i.e. for each $k \in \overline{1,n}$, define the associated matrix

$$\tilde{A}^k = \begin{pmatrix} \underline{\tilde{a}}^{l^1,k} \\ \underline{\tilde{a}}^{l^2,k} \\ \cdots \\ \underline{\tilde{a}}^{l^{J^*},k} \end{pmatrix} \in \mathscr{M}_{J^*,J(T)\times n}.$$

Similarly, also define for each $k \in \overline{1,n}$,

$$\underline{\hat{f}}^k = (\hat{f}_{t_k,T}(l^1), \hat{f}_{t_k,T}(l^2), \ldots, \hat{f}_{t_k,T}(l^{J^*})) \in \mathscr{M}_{1,J^*}.$$

In this notation, an estimator for the vector of wavelet spectrum values corresponding to the observed locations, $\underline{\tilde{s}}$, can be obtained by solving the following system with $J(T)\times n$ unknowns and $J^*\times n$ equations

$$\begin{pmatrix} \tilde{A}^1 \\ \tilde{A}^2 \\ \cdots \\ \tilde{A}^n \end{pmatrix} \underline{\tilde{s}}^T = \begin{pmatrix} (\underline{\hat{f}}^1)^T \\ (\underline{\hat{f}}^2)^T \\ \cdots \\ (\underline{\hat{f}}^n)^T \end{pmatrix}. \tag{16}$$

Solving this large system of equations in equation (16) can obviously be computationally intensive. Upon investigation of the A-matrices structure, they exhibit a sparse structure: for $A^{l^i,k}$, only those columns corresponding to neighbouring time points of t_k are non-zero (see Figure 10).

In order to take advantage of this sparsity, we rearrange the A-matrices into matrices $B_{j,k} = (b^{l^i,l}_{j,k} = a^{l^i,k}_{l,j})_{i\in\overline{1,J^*},l\in\overline{1,J}}$, to obtain

$$E(\underline{\hat{f}}^k) \sim \sum_{j \text{ s.t. } t_j\in V_{t_k}} B_{j,k} S_{\cdot,j}, \tag{17}$$

where V_{t_k} is a neighbourhood of the time point t_k. For fixed t_k, the $B_{j,k}$-matrices are non-zero only for those time points t_j that are around t_k. Due to the reduction in the size of the system, we also reduce computational costs.

Nason et al (2000) note that wavelet periodograms exhibit '*power leakage*' from fine to coarser scales, also exhibited in our raw periodograms (see Figure 5). Hence we formulate a set of *penalized* linear least squares problems

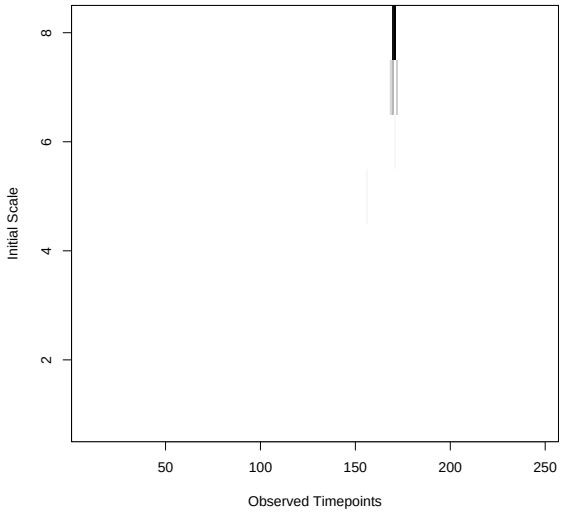

Fig. 10 Example of an A matrix. Darker pixels correspond to higher values.

(with inequality constraints) in order to estimate S using the formulation (17).

More exactly, for each time point t_k we solve

$$\min\left\{ \left\| \underline{\hat{f}}^k - \sum_{j \text{ s.t. } t_j\in V_{t_k}} B_{j,k} S_{\cdot,j} \right\|^2 + \sum_{l=-J(T)}^{-1} 2^{-(l+1)} (\lambda \|S_l''\|^2 + \mu \|S_l\|^2) \right\}$$

to estimate the unknown spectrum values, subject to $S[,t_j] \geq 0$, where $\lambda, \mu \in \mathbb{R}$ are unknown constants. This penalization criterion incorporates a cost for high power at coarse scales as well as a smoothness constraint for spectral content over time at particular scales.

For computational reasons, we take the neighbourhood V_{t_k} to be small in practice, e.g. $j \in \{k-1, k, k+1\}$. This represents a narrow (vertical) strip around the point of interest.

The solution to the penalized linear least squares problem above is our proposed *corrected* periodogram that estimates the evolutionary wavelet spectrum, S.

Since the penalized least squares correction employs a search for values over the time series, it is essentially looking for J x n best fitting (real) spectrum values subject to a convergence tolerance. This means that the correction could potentially find spectral solutions which look quite different on separate implementations of the search. However, we have found that the spectral estimates resulting from the penalized least squares correction seem to be more stable as long as the weight μ is chosen to be non-zero. The norm weights λ and μ can also be incorporated in the optimization search.

For time series of increasing length, the number of scales also increases, and so the number of real values to find grows

across time as well as scale as n increases. Hence the corrected spectrum search increases dramatically in computational cost and time.

We also note that the penalized correction used above can be applicable to irregular time series generally, by mapping the irregular time structure to the regular time series with missing observations framework.

5.3 Example

As a demonstration of this method of correction for the proposed periodogram from Section 3.2.3, let us consider a wavelet spectrum characterized by a fine-level burst (Figure 11). As with the example in Section 4, we simulate a LSW process from the spectrum, and then remove a number of observations, forming a time series from which we hope to estimate the original spectrum (shown in the top-right of Figure 11).

The raw periodogram of Section 3.2.3 was obtained by using the algorithm from the flowchart (Figure 2) on the time series with missing observations, by using $m = 1000$ trajectories in the NLT. The resulting periodogram (bottom-left) shows a well-defined power burst (in terms of time) at the location of the burst in the original spectrum, but also exhibits apparent power leakage across the finer scales.

The penalized spectral search algorithm of Section 5.2 was then applied to the raw periodogram to form a corrected spectrum estimate (bottom-right of Figure 11). The edges of the burst are still well-defined, but the penalized criterion has successfully removed a lot of bias and power leakage from our raw periodogram.

6 Conclusions and further work

In this article we have addressed the problem of spectral estimation for a non-stationary process that exhibits missing observations, a problem which so far has not been addressed in the literature. Non-stationarity was understood here as local stationarity, and the wavelet model introduced by Nason et al (2000) was adopted. Second generation wavelet methods constructed by means of the lifting scheme that 'removes one coefficient at a time' (Jansen et al, 2001) were employed, due to their flexibility of working with irregularly-spaced data not necessarily of a dyadic length. In this context a 'nondecimated' lifting transform (Knight and Nason, 2009) was used to ensure that a set of empirical wavelet coefficients is available at each (observed) time location throughout a continuous distribution of scales. Exploiting the flexibility behind the continuous nature of scale in second generation wavelet approaches, we proposed a 'raw' periodogram for estimating the wavelet spectrum at the (rescaled) observed locations. We presented two sets of examples from environmental and medical time series that show the utility of our approach in obtaining local spectral information from locally stationary time series which suffer from missing data. In both cases we were able to elicit frequencies at which significant power existed and, in the case of the environmental series, agree with the current accepted knowledge in the field.

Theoretically, we showed that the periodogram is not an unbiased estimator for the evolutionary wavelet spectrum, and have also explored initial work towards a corrected periodogram through using a penalized criterion on the spectrum. However, our approach is highly computationally intensive.

For the future, it would be interesting to further investigate the properties of the corresponding estimator, as well as its asymptotic behaviour. Also, to this moment we have not explored the possible advantages of using the *adaptive* lifting of Nunes et al (2006) in our development, which would give our method the potential of not having to choose the wavelet basis a priori.

An existing challenge is to set up a locally stationary wavelet type model that would directly handle the problem of correcting the periodogram for irregular data.

Finally we note that the methods and ideas presented in this paper of using nondecimated lifting for spectral analysis can be readily generalized to multidimensional settings, for example, by modifying the Voronoi polygon- or tree-based lifting transforms introduced in Jansen et al (2009). The use of nondecimated lifting techniques similar to those in this article for multivariate time series is an interesting avenue of research and initial work in this area seems promising (see Sanderson (2010)).

7 Acknowledgments

The authors gratefully acknowledge financial support from the University of Bristol Applied Research project funded by Her Majesty's Government. The authors would also like to thank Jean Sanderson for many interesting discussions. The authors would like to thank P. Fleming, A. Sawczenko and J. Young of the Institute of Child Health, Royal Hospital for Sick Children, Bristol for supplying the ECG data.

A Proofs

This appendix gives the proofs of the results from Section 5.1, following the notation outlined in the text.

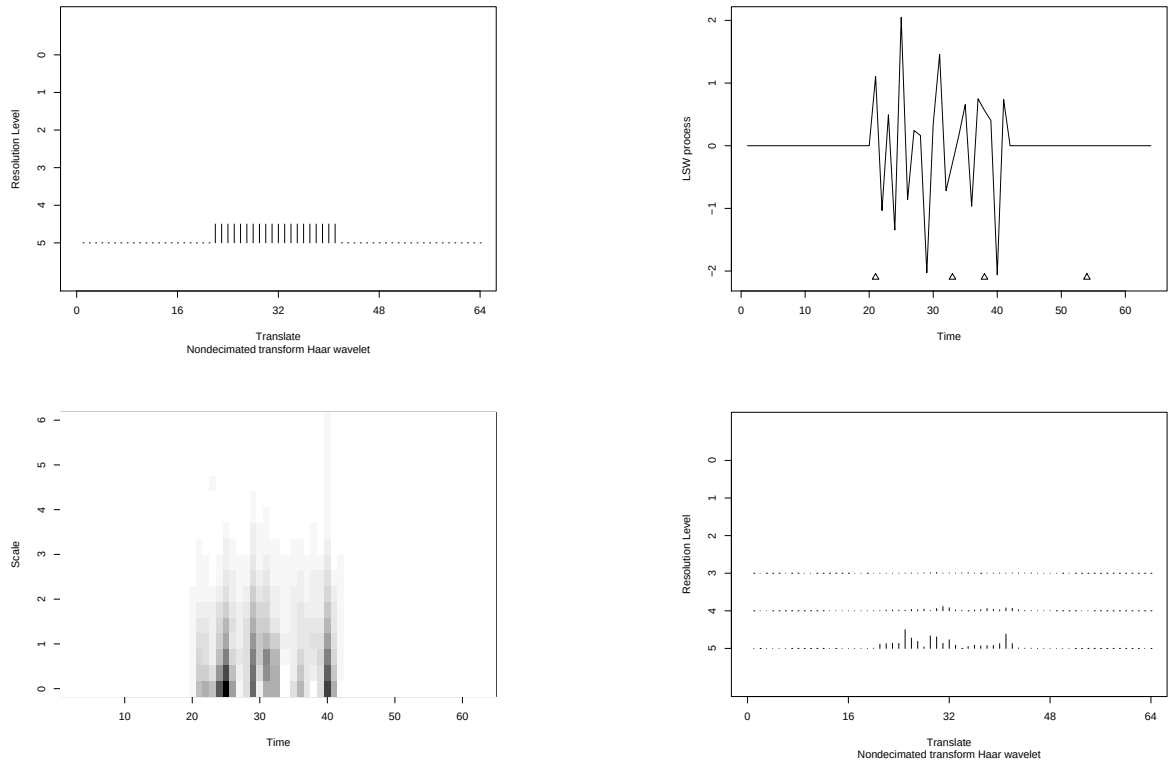

Fig. 11 Top left: Wavelet spectrum – finest scale on bottom. Top right: Simulated LSW process of length 64 with 4 observations deemed missing. Bottom left: (Uncorrected) proposed raw periodogram. Bottom right: penalized linear least squares estimate of S (with $\lambda = \mu = 35$).

A.1 Proof of Lemma 5.1

For any $\alpha,\ \beta \in \overline{1,m}$, let us denote

$$\Sigma^{\alpha,T} = \mathrm{var}((\underline{d}^{\alpha,T})^T | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}, \text{fixed paths}) \in \mathscr{M}_{n,n}.$$
$$A^{\alpha,\beta,T} = \mathrm{cov}((\underline{d}^{\alpha,T})^T, (\underline{d}^{\beta,T})^T | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}, \text{fixed paths}) \in \mathscr{M}_{n,n}.$$

The variance–covariance matrix of the vector $(\underline{d}^{\alpha,T}, \underline{d}^{\beta,T})^T \in \mathscr{M}_{2n,1}$ thus takes the form

$$\mathrm{var}\left(\begin{pmatrix}(\underline{d}^{\alpha,T})^T\\(\underline{d}^{\beta,T})^T\end{pmatrix} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}, \text{fixed paths}\right) = \begin{pmatrix}\Sigma^{\alpha,T} & A^{\alpha,\beta,T}\\(A^{\alpha,\beta,T})^T & \Sigma^{\beta,T}\end{pmatrix}. \quad (18)$$

Since $(\underline{d}^{\alpha,T})^T = R^\alpha((X_{t_i,T})_{i\in\overline{1,n}})^T$ for $\alpha \in \overline{1,m}$, it follows that for any $\alpha, \beta \in \overline{1,m}$ we have

$$\begin{pmatrix}(\underline{d}^{\alpha,T})^T\\(\underline{d}^{\beta,T})^T\end{pmatrix} = \begin{pmatrix}R^\alpha\\R^\beta\end{pmatrix}((X_{t_i,T})_{i\in\overline{1,n}})^T.$$

Hence

$$\mathrm{var}\left(\begin{pmatrix}(\underline{d}^{\alpha,T})^T\\(\underline{d}^{\beta,T})^T\end{pmatrix} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}, \text{fixed paths}\right) = \begin{pmatrix}R^\alpha\Sigma^{(T)}(R^\alpha)^T & R^\alpha\Sigma^{(T)}(R^\beta)^T\\R^\beta\Sigma^{(T)}(R^\alpha)^T & R^\beta\Sigma^{(T)}(R^\beta)^T\end{pmatrix}, \quad (19)$$

where

$$\Sigma^{(T)} = \mathrm{var}(((X_{t_i,T})_{i\in\overline{1,n}})^T | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = (\sigma_{j,k;T})_{j,k\in\overline{1,n}}$$

is the (symmetric) variance-covariance matrix of the observed signal (with missing observations), having assumed that the missing points are deterministic rather than random quantities.

Using relation (18), we obtain

$$\Sigma^{\alpha,T} = R^\alpha\Sigma^{(T)}(R^\alpha)^T, \quad \forall\alpha \in \overline{1,m}, \quad (20)$$
$$A^{\alpha,\beta,T} = R^\alpha\Sigma^{(T)}(R^\beta)^T, \quad \forall\alpha,\beta \in \overline{1,m}. \quad (21)$$

Written explicitly, equation (21) takes the form

$$\mathrm{cov}(d^\alpha_{t_i,T}, d^\beta_{t_{i'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}, \text{fixed paths}) = \sum_{j=1}^n\sum_{j'=1}^n r^\alpha_{i,j}\,\mathrm{cov}(X_{t_j,T}, X_{t_{j'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) r^\beta_{i',j'},$$

for any $\alpha,\beta \in \overline{1,m}$, $i,i' \in \overline{1,n}$.

A.2 Proof of Proposition 5.2

If we let $z_j = \frac{t_j}{T} \in (0,1)$, then the process autocovariance can be written as

$$\mathrm{cov}(X_{t_j,T}, X_{t_{j'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = \mathrm{cov}(X_{\lfloor z_j T\rfloor}, X_{\lfloor z_j T\rfloor + (t_{j'} - t_j)} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}).$$

Therefore, we can write

$$\mathrm{cov}(X_{t_j,T}, X_{t_{j'},T} | I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}) = c_T(\frac{t_j}{T}, t_{j'} - t_j), \quad (22)$$

where $c_T(\cdot\,,\cdot\,)$ is the autocovariance of the LSW process $(X_{t,T})_{t\in\overline{0,T-1}}$, introduced in Section 2.2, and we assume the conditioning still holds.

From the result in Lemma 5.1 and equation (22), we obtain

$$\mathrm{cov}(d^{\alpha}_{t_i,T}, d^{\beta}_{t_{i'},T}|I_{\mathscr{S}}=\underline{1}, I_{\bar{\mathscr{S}}}=\underline{0}, \text{fixed paths}) = \sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} c_T(\frac{t_j}{T}, t_{j'}-t_j) r^{\beta}_{i',j'}.$$

Nason et al (2000) proved that the process autocovariance and local autocovariance functions are linked through the relation $c_T(z,\tau) = c(z,\tau) + R_T$, for any rescaled time location z and lag τ, where R_T is a term of magnitude $O(T^{-1})$. From the above relation, the following becomes apparent:

$$\mathrm{cov}(d^{\alpha}_{t_i,T}, d^{\beta}_{t_{i'},T}|I_{\mathscr{S}}=\underline{1}, I_{\bar{\mathscr{S}}}=\underline{0}, \text{fixed paths}) = \sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} c(\frac{t_j}{T}, t_{j'}-t_j) r^{\beta}_{i',j'} + \sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} R_T r^{\beta}_{i',j'},$$

for any $\alpha,\ \beta \in \overline{1,m},\ i,\ i' \in \overline{1,n}$.

Let us denote

$$\tilde{R}_T = R_T \sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{i,j} r^{\beta}_{i',j'} = R_T \sum_{j=1}^{n} r^{\alpha}_{i,j} \sum_{j'=1}^{n} r^{\beta}_{i',j'}.$$

Since matrices associated to a LOCAAT lifting transform have a sparse character, for a fixed i the sums of the type $\sum_{j=1}^{n} r^{\alpha}_{i,j}$ only involve a finite number of elements, independent of the magnitude of n. If more data is collected, then there is a chance that the new observations will be involved in $\sum_{j=1}^{n} r^{\alpha}_{i,j}$ for a fixed i, but the combination will still be sparse, and so $\sum_{j=1}^{n} r^{\alpha}_{i,j} := C^{\alpha} < \infty$. As $R_T = O(T^{-1})$, it follows that $\exists\, k < \infty$ such that $|\tilde{R}_T| \le kT^{-1}C^{\alpha}C^{\beta} < \infty$, so $\tilde{R}_T$ also has magnitude $O(T^{-1})$.

A.3 Proof of Proposition 5.3

In the LSW model, the sequence of processes $(X_{t,T})_{t\in\overline{0,T-1}}$, is assumed to have zero mean, i.e. $\mathbb{E}(X_{t,T}) = 0,\ \forall t \in \overline{0, T-1},\ \forall T$.

Since

$$\mathbb{E}(d^{\alpha}_{t_i,T} d^{\beta}_{t_{i'},T}) = \mathrm{cov}(d^{\alpha}_{t_i,T}, d^{\beta}_{t_{i'},T}) + \mathbb{E}(d^{\alpha}_{t_i,T})\mathbb{E}(d^{\beta}_{t_{i'},T})$$

and from formula (13)

$$\mathbb{E}(d^{\alpha}_{t_i,T}|I_{\mathscr{S}}=\underline{1}, I_{\bar{\mathscr{S}}}=\underline{0}, \text{fixed paths}) = \sum_{j=1}^{n} r^{\alpha}_{i,j}\mathbb{E}(X_{t_j,T}),$$

we obtain the desired equation.

A.4 Proof of Theorem 5.4

Since $\hat{f}_{t_k,T}(l^i) = \sum_{\alpha=1}^{m} K_{\alpha}(l^i)(d^{\alpha}_{t_k,T})^2,\ \forall i \in \overline{1,J^*},\ \forall k \in \overline{1,n}$, it follows that

$$\mathbb{E}(\hat{f}_{t_k,T}(l^i)|I_{\mathscr{S}}=\underline{1}, I_{\bar{\mathscr{S}}}=\underline{0}, \text{fixed paths}) = \sum_{\alpha=1}^{m} K_{\alpha}(l^i)\mathbb{E}((d^{\alpha}_{t_k,T}{}^2)|I_{\mathscr{S}}=\underline{1}, I_{\bar{\mathscr{S}}}=\underline{0}, \text{fixed paths}).$$

By taking $\alpha = \beta$ and $i = i' := k$ in (14), $\mathbb{E}(\hat{f}_{t_k,T}(l^i)|I_{\mathscr{S}} = \underline{1}, I_{\bar{\mathscr{S}}} = \underline{0}, \text{fixed paths})$ can be expressed as

$$\begin{aligned}
&\sum_{\alpha=1}^{m} K_{\alpha}(l^i)\left\{\sum_{l=-\infty}^{-1}\sum_{j=1}^{n}\sum_{j'=1}^{n} r^{\alpha}_{k,j} r^{\alpha}_{k,j'} \Psi_l(t_j - t_{j'}) S_l(\frac{t_j}{T}) + \tilde{R}_T\right\} \\
&= \sum_{l=-\infty}^{-1}\sum_{j=1}^{n}\left[\sum_{j'=1}^{n}\left\{\sum_{\alpha=1}^{m} K_{\alpha}(l^i) r^{\alpha}_{k,j} r^{\alpha}_{k,j'}\right\}\Psi_l(t_j - t_{j'})\right] S_l(\frac{t_j}{T}) \\
&\qquad + \tilde{R}_T \sum_{\alpha=1}^{m} K_{\alpha}(l^i), \quad (23)
\end{aligned}$$

$\forall i \in \overline{1,J^*},\ \forall k \in \overline{1,n}$.

As $a^{l^i,k}_{l,j} = \sum_{j'=1}^{n}\left\{\sum_{\alpha=1}^{m} K_{\alpha}(l^i) r^{\alpha}_{k,j} r^{\alpha}_{k,j'}\right\}\Psi_l(t_j - t_{j'})$, the above equation can be equivalently written as

$$\mathbb{E}(\hat{f}_{t_k,T}(l^i)|I_{\mathscr{S}}=\underline{1}, I_{\bar{\mathscr{S}}}=\underline{0}, \text{fixed paths}) = \sum_{l=-\infty}^{-1}\sum_{j=1}^{n} a^{l^i,k}_{l,j} S_l(\frac{t_j}{T}) + \tilde{R}_T \sum_{\alpha=1}^{m} K_{\alpha}(l^i). \quad (24)$$

Therefore

$$\begin{aligned}
&\mathbb{E}(\hat{f}_{t_k,T}(l^i)|I_{\mathscr{S}}=\underline{1}, I_{\bar{\mathscr{S}}}=\underline{0}, \text{fixed paths}) \\
&\qquad = \sum_{l=-\infty}^{-1} (A^{l^i,k} S^T)_{l,l} + \tilde{R}_T \sum_{\alpha=1}^{m} K_{\alpha}(l^i) \\
&\qquad = \mathrm{Trace}(A^{l^i,k} S^T) + \tilde{R}_T \textstyle\sum_{\alpha=1}^{m} K_{\alpha}(l^i).
\end{aligned}$$

Observe that in order to obtain $(A^{l^i,k}S^T)_{l,l}$ for a fixed time t_k and scale l^i, only the terms corresponding to time locations $t_j,\ t_{j'}$ such that $(t_j - t_{j'})$ does not exceed the support of the autocorrelation wavelet $\Psi_l(\cdot\,)$ are contributing to the sum.

Let us denote $\tilde{R}^*_T = \tilde{R}_T \sum_{\alpha=1}^{m} K_{\alpha}(l^i)$. For finite m, $\tilde{R}^*_T$ has magnitude $O(T^{-1})$ as $\tilde{R}_T$ has magnitude $O(T^{-1})$ from the previous proposition.

References

Berger AL (1978) Long-term variations of daily insolation and quaternary climatic changes. J Atmos Sci 35:2362–2367

Bos R, de Waele S, Broersen PMT (2002) Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data. IEEE Trans Instrum Meas 51(6):1289–1294

Brockwell PJ, Davis RA (2009) Time series: theory and methods, 2nd edn. Springer Verlag

Broersen PMT (2006) Automatic spectral analysis with missing data. Digit Sig Proc 16(6):754–766

Broersen PMT (2008) Time series models for spectral analysis of irregular data far beyond the mean data rate. Meas Sci Tech 19(1):015,103, DOI http://stacks.iop.org/0957-0233/19/i=1/a=015103

Broersen PMT, de Waele S, Bos R (2004) Autoregressive spectral analysis when observations are missing. Automatica 40(9):1495–1504

Cazelles B, Chavez M, Magny GC, Guégan J, Hales S (2007) Time-dependent spectral analysis of epidemiological time-series with wavelets. J Roy Soc Interface 4(15):625–636

Chatfield C (2004) The analysis of time series: an introduction. Chapman & Hall/CRC Pr I Llc

Clinger W, Van Ness JW (1976) On unequally spaced time points in time series. Ann Stat 4(4):736–745

Cranstoun SD, Ombao HC, von Sachs R, Guo W, Litt B (2002) Time-frequency spectral estimation of multichannel EEG using the Auto-SLEX method. IEEE Trans Biomed Eng 49(9):988–996

Crucifix M (2008) Global change: Climate's astronomical sensors. Nature 456(7218):47–48

Crucifix M, Rougier J (2009) On the use of simple dynamical systems for climate predictions. The Euro Phys Journal 174(1):11–31

Crucifix M, Loutre MF, Berger A (2007) The climate response to the astronomical forcing. In: Calisesi Y, Bonnet RM, Gray L, Langen J, Lockwood M (eds) Solar Variability and Planetary Climates, Space Sciences Series of ISSI, vol 23, Springer New York, pp 213–226

Dahlhaus R (1997) Fitting time series models to nonstationary processes. Ann Stat 25(1):1–37

Dahlhaus R, Subba Rao S (2006) Statistical inference for time-varying ARCH processes. Ann Stat 34(3):1075–1114

Dahlhaus R, Subba Rao S (2007) A recursive online algorithm for the estimation of time-varying ARCH parameters. Bernoulli 13(2):389–422

Dilmaghani S, Henry IC, Soonthornnonda P, Christensen ER, Henry RC (2007) Harmonic analysis of environmental time series with missing data or irregular sample spacing. Environ Sci Tech 41(20):7030–7038

Engle RF (2000) The econometrics of ultra-high-frequency data. Econometrica 68(1):1–22

Fryźlewicz P (2003) Wavelet techniques for time series and poisson data. PhD thesis, University of Bristol, UK

Fryźlewicz P, Nason GP (2006) Haar-Fisz estimation of evolutionary wavelet spectra. J Roy Stat Soc B 68:611–634

Fryźlewicz P, Sapatinas T, Rao S (2006) A Haar-Fisz technique for locally stationary volatility estimation. Biometrika 93(3):687

Hall P, Fisher NI, Hoffmann B (1994) On the nonparametric estimation of covariance functions. Ann Stat 22(4):2115–2134

Jansen M, Nason GP, Silverman BW (2001) Scattered data smoothing by empirical Bayesian shrinkage of second generation wavelet coefficients. In: Unser M, Aldroubi A (eds) Wavelet Applications in Signal and Image Processing IX, SPIE, vol 4478, pp 87–97

Jansen M, Nason GP, Silverman BW (2004) Multidimensional nonparametric regression using lifting. Tech. Rep. 04:17, Statistics Group, Department of Mathematics, University of Bristol, UK

Jansen M, Nason GP, Silverman BW (2009) Multiscale methods for data on graphs and irregular multidimensional situations. J Roy Stat B 71(1):97–125

Jones RH (1962) Spectral analysis with regularly missed observations. Ann Math Stat 33(2):455–461

Knight MI, Nason GP (2006) Improving prediction of hydrophobic segments along a transmembrane protein sequence using adaptive multiscale lifting. SIAM J Multiscale Mod and Sim 5:115–129

Knight MI, Nason GP (2009) A nondecimated lifting transform. Stat Comput 19(1):1–16

Knight MI, Nunes MA (2010) nlt: a nondecimated lifting scheme algorithm. R package version 2.1-1

Lüthi D, Le Floch M, Bereiter B, Blunier T, Barnola JM, Siegenthaler U, Raynaud D, Jouzel J, Fischer H, Kawamura K, et al (2008) High-resolution carbon dioxide concentration record 650,000–800,000 years before present. Nature 453(7193):379–382

Mikosch T, Starica C (2004) Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Rev Econom Stat 86(1):378–390

Mondal D, Percival DB (2008) Wavelet variance analysis for gappy time series. Ann Inst Stat Math pp 1–24

Nason GP (2008) Wavelet methods in statistics with R. Springer Verlag

Nason GP, Von Sachs R (1999) Wavelets in time-series analysis. Phil Trans Roy Soc London A 357(1760):2511–2526

Nason GP, Von Sachs R, Kroisandt G (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. J Roy Stat Soc B 62(2):271–292

Nason GP, Sapatinas T, Sawczenko A (2001) Wavelet packet modelling of infant sleep state using heart rate data. Sankhy$\bar{a}$ B pp 199–217

Nason GP, Kovac A, Maechler M (2008) Wavethresh: Software to perform wavelet statistics and transforms. R package version 4.2-1

Nunes MA, Knight MI (2010) Adlift: an adaptive lifting scheme algorithm. R package version 1.2-3

Nunes MA, Knight MI, Nason GP (2006) Adaptive lifting for nonparametric regression. Stat and Comput 16(2):143–159

Ombao H, Raz J, Von Sachs R, Guo W (2002) The SLEX model of a non-stationary random process. Ann Inst Stat Math 54(1):171–200

Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press: Cambridge

Priestley MB (1981) Spectral analysis and Time Series. Academic Press

Sanderson J (2010) Wavelet methods for time series with bivariate observations and irregular sampling grids. PhD thesis, University of Bristol, UK

Stoica P, Sandgren N (2006) Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches. Digit Sig Proc 16(6):712–734

Sweldens W (1995) The lifting scheme: A new philosophy in biorthogonal wavelet construction. In: Laine A, Unser M (eds) Wavelet Applications in Signal and Image Processing III, Proc. SPIE 2569, pp 68–79

Van Bellegem S, Von Sachs R (2008) Locally adaptive estimation of evolutionary wavelet spectra. Ann Stat 36:1879–1924

Witt A, Schumann AY (2005) Holocene climate variability on millennial scales recorded in Greenland ice cores. Nonlin Proc Geophys 12(3):345–352

Wolff EW (2005) Understanding the past-climate history from Antarctica. Antarctic Sci 17(04):487–495