Published in final edited form as:

Stat Comput. 2020 February ; 30(1): 167–185. doi:10.1007/s11222-019-09874-0.

Estimation of long-range dependence in gappy Gaussian time series

Peter F. Cragmile,

Department of Statistics, The Ohio State University, Columbus, OH 43210, USA.

Debashis Mondal

Department of Statistics, Oregon State University, Corvallis, OR 97331, USA.

Abstract

Knowledge of the long range dependence (LRD) parameter is critical to studies of self-similar behavior. However, statistical estimation of the LRD parameter becomes difficult when the observed data are masked by short range dependence and other noise, or are gappy in nature (i.e., some values are missing in an otherwise regular sampling). Currently there is a lack of theory for spectral- and wavelet-based estimators of the LRD parameter for gappy data. To address this, we estimate the LRD parameter for gappy Gaussian semiparametric time series based upon undecimated wavelet variances. We develop estimation methods by using novel estimators of the wavelet variances, providing asymptotic theory for the joint distribution of the wavelet variances and our estimator of the LRD parameter. We introduce sandwich estimators to compute standard errors for our estimates. We demonstrate the efficacy of our methods using Monte Carlo simulations, and provide guidance on practical issues such as how to select the range of wavelet scales. We demonstrate the methodology using two applications: one for gappy Arctic sea-ice draft data, and another for gap free and gappy daily average temperature data collected at 17 locations in south central Sweden.

Keywords

Between-wavelet-scale correlation; Hurst index; missing data; power law; sandwich estimator; self-similar noise

1 Introduction

Understanding the index or range of dependence in a time series has long been of interest to statisticians, hydrologist, economists, and other physical scientists (Beran (1994, Chapter 1) provides an excellent recount of this history). Long-range dependence (hereafter LRD) is often a norm in many physical phenomena, and refers to dependence structures that decay slowly as the lag distance between two observations grows. For Gaussian time series, LRD can be examined by looking at the rate of increase or decrease of its (generalized) spectral

density near zero frequency. A large value of this rate indicates a slow decay of statistical dependence that is typical of a process exhibiting smoother sample paths and complex dependencies. The knowledge of the rate parameter, also called the LRD parameter, has many applications. It provides information about the self-similar behavior of a time series as popularized by Hurst (1951) and Mandelbrot and van Ness (1968), and can be used to decide the order of differencing that is useful in constructing non-Markov models such as the fractional ARMA models made famous by Granger and Joyeux (1980) and Hosking (1981).

Over the decades, two approaches to the estimation of the LRD parameter have emerged; one based on spectral methods and the other using wavelets. (See Beran (1994) and Palma (2007) for a more complete review of time series methods for LRD.) The most used spectral method (e.g., Künsch, 1987; Robinson, 1995b) estimates the LRD parameter via log-periodogram regression, evaluated at the discrete Fourier frequencies. While this spectral method leads to interesting asymptotics, several adjustments such as finding the appropriate order of data differencing, optimal tapering, and optimally selecting an appropriate bandwidth of discrete frequencies for the estimator, often hamper the estimation procedures. This limits the ability to develop a single unified theory that includes these corrective steps; see, for example, Geweke and Porter-Hudak (1983), Velasco (1999), Hurvich and Chen (2000), Hurvich and Ray (2003), Moulines and Soulier (2003), Hurvich et al. (2005). Wavelet methods, on the other hand, focus on wavelet variances that define the variations in the differences in localized weighted averages over different time scales. For a LRD process, the log wavelet variance is linearly related to the LRD parameter multiplied by the log time scale, allowing us to estimate the LRD parameter by regressing the log wavelet variances (e.g. Abry and Veitch, 1998; Percival and Mondal, 2012). Unlike spectral-based methods, wavelet methods incorporate data differencing implicitly, removing polynomial trend automatically and, arguably, calling forth a more parsimonious analysis of LRD series (e.g., Craigmile et al., 2005). While wavelet methods have some limitations due to their restricted choices of time scales for which fast computations are permissible and problems surrounding plug-in estimation of uncertainty measures for the estimators, they are practically comparable to spectral-based ones: Jensen (1999), McCoy and Walden (1996), and more recently Moulines et al. (2008) and Fay et al. (2009) provide the revelation that wavelet-based methods have the potential to produce estimators of the LRD of a stationary time series with smaller root mean squared error, compared to existing spectral-based ones.

The primary focus of this paper is the estimation of the LRD parameter for time series with missing observations that can occur by accident (e.g., there are transmission problems in a satellite), by design (e.g., temperature data is not collected for a given period resulting in systematic gaps), or for other unforeseeable reasons. We call such a time series with missing observations *gappy*. Despite several works on spectral density function (SDF) estimation of short range dependent time series (e.g., Parzen, 1963; Scheinok, 1965; Bloomfield, 1970; Jones, 1971; Dunsmuir and Robinson, 1981; Dahlhaus, 1987) the theory for the estimation of the SDF for a gappy LRD time series, however, has remained unsolved, limiting the ability to further provide a useful spectral-based estimate of the LRD parameter. In this paper, we build upon the work of Mondal and Percival (2010) that derive consistent and asymptotically normal estimates of wavelet variances for gappy time series, including LRD ones. We develop theory for the joint estimation of wavelet variances over multiple time

scales, providing wavelet-based general least squares estimates of the LRD parameter for gappy Gaussian LRD series.

The main contributions of the paper are as follows. First, we develop an asymptotic theory for the joint distribution of the variogram type wavelet variance estimate of Mondal and Percival (2010) over multiple time scales, when the time scale approaches to infinity. This allows us to develop new estimates of the LRD parameter for gappy Gaussian LRD time series, and to provide corresponding theory to show that our method produces a consistent and asymptotic normal estimate in the presence of short-range dependence effects. Second, by embracing the correlations present between wavelet coefficients over different wavelet scales, we improve upon uncertainty estimates presented by others in the field. Typically, existing theory and methods (e.g. Abry and Veitch, 1998; Bardet et al., 2000; Moulines et al., 2007a,b; Roueff and Taqqu, 2009; Fay et al., 2009) for wavelet-based estimators of LRD are for gap free series, use a decimated wavelet transform, do not discuss uncertainty estimates, and the estimator (although not the asymptotic theory) assumes that coefficients over different wavelet scales are uncorrelated. Kouamo et al. (2013) proposes an interpolation-wavelet-based method for possibly gappy nonstationary Gaussian series and Knight et al. (2016) develop an estimation method based on wavelet lifting.

The rest the paper is laid out as follows. We introduce our gappy LRD process in Section 2, and provide theory for the estimation of wavelet variances of such series in Section 3. Our estimator of the LRD parameter and asymptotic theory is given in Section 4. We develop sandwich estimates for the standard error of the estimator which allow us to construct approximate confidence intervals for statistical inference. We also discuss the issues underlying the choice of wavelet scales that the estimator of the LRD parameter should be based upon. We investigate the performance of our estimators in both the gap free and gappy cases via simulation studies in Section 5. Through two carefully chosen applications we illustrate in Section 6 the versatility of settings in which these estimators can be used; for example, we can estimate the LRD parameter in the presence of nuisance trends, seasonality, and short range dependence. We close with a discussion in Section 7. Proofs are given in the appendix and extra tables and figures are presented in a supplement. R software (R Core Team, 2018) that accompanies this article is available from <https://github.com/petercraigmile/GappyLRD>.

2 Gaussian long range dependent time series with gaps

Assume that the time series $\{X_t, t \in \mathbb{Z}\}$ is Gaussian with a (generalized) SDF of the form

$$S_X(f) = \{2 \sin(\pi f)\}^{-2\delta} S_0(f), \quad f \in [-1/2, 1/2], \quad (1)$$

where the LRD parameter δ belongs to real numbers and the unknown even, non-negative, nuisance function $S_0(\cdot)$ represents the short memory (short range dependence) spectral density of $\{X_t\}$. Suppose that $S_0(\cdot)$ satisfies $S_0(0) > 0$ and a local Lipschitz condition at the zero frequency, namely,

$$|S_0(f) - S_0(0)| \leq C_1 S_0(0) |f|^\zeta, f \in (-\epsilon, \epsilon), \quad (2)$$

for a constant $C_1 > 0$, some ϵ in $(0, 1/2)$ and a fixed $0 < \zeta < 2$. The local Lipschitz condition, characterizing the regularity of $S_0(\cdot)$ near zero frequency, allow us to tease out the effect of the LRD parameter δ from the effect of $S_0(\cdot)$. Thus, near zero frequency the spectral density $S_X(\cdot)$ mimics a pure power law. This mild regularity condition emerged from the seminal work of Robinson (1994, 1995a,b) and is used extensively in the theory of estimators for the LRD parameter.

The above model for $\{X_t\}$ covers a large class of parametric models including the class of fractional autoregressive and moving averages, fractional exponential processes, as well as fractional Brownian motions and Gaussian noises observed on regular intervals. We usually encounter positive values of δ (we call $\{X_t\}$ anti-persistent when we observe negative values of δ , often as a result of differencing a time series). When $\delta < 1/2$, $\{X_t\}$ is covariance stationary, otherwise we view $\{X_t\}$ as a nonstationary intrinsic time series (e.g., Yaglom, 1987, Chapter 4) implying that its δ th order increments, for $\delta_t > \delta$, form a covariance stationary Gaussian time series whose SDF takes the form $\{2 \sin(\pi f)\}^{2(\delta_t - \delta)} S_0(f)$, for $f \in [-1/2, 1/2]$.

For the missing data mechanism, let $\{\eta_t\}$ denote a strictly stationary binary time series taking values zero or one according to whether or not X_t is missing. We assume that this gap process $\{\eta_t\}$ is independent of the time series $\{X_t\}$. Define

$$\pi_k^{-1} = P(\{\eta_t = 1\} \cap \{\eta_{t+k} = 1\})$$

to be the probability of observing two time series values at a distance of k lags apart. Necessarily $\pi_k^{-1} > 0$ for all values of k . In Section 3.3 we will assume that the gap process $\{\eta_t\}$ satisfies the mild mixing condition of Mondal and Percival (2010, Assumption 4.1) (The mixing condition ensures that the large sample distribution of $\hat{\pi}_{j,l,l'}^{-1}$ will be Gaussian). We allow for the possibility that there are no gaps in the series.

3 Wavelet analysis of gappy LRD data

Our estimation of the LRD parameter is based on an undecimated wavelet analysis of $\{X_t\}$ using the Daubechies class of wavelet filters (e.g., Daubechies, 1992, Section 6.2). This class of orthogonal filters, containing the Daubechies extremal phase (D) and least asymmetric (LA) sets of filters, are characterized by the filter width, L , or equivalently the maximum number of vanishing moments, $L/2$. We use an undecimated or maximum overlap wavelet transform for $\{X_t\}$ because it maximizes the amount of temporal information available at each wavelet level (e.g., Percival and Walden, 2000).

Let $\{h_{1,l}: l = 0, \dots, L-1\}$ denote a wavelet filter of even width L with energy equal to a half, that is necessarily orthogonal to even shifts; i.e., $\sum h_{1,l}^2 = 1/2$; $\sum h_{1,l} h_{1,l+2n} = 0$, for all nonzero integers n with the convention that $h_{1,l} = 0$ for $l < 0$ and $l \geq L$ (e.g., Percival and

Walden, 2000, p.70). We construct higher level wavelet filters $\{h_{j,l}\}$ for $j > 1$ in terms of $\{h_{1,l}\}$. Let $H_1(f)$ denote the Fourier transform of the filter $\{h_{1,l}\}$, and take $G_1(f) = e^{-2\pi i(L-1)f} H_1(1/2 - f)$. The j th level wavelet filter, $\{h_{j,l}\}$, then arises from the convolution of j linear filters whose individual Fourier transforms take the form $G_1(f)$, $G_1(2f)$, ..., $G_1(2^{j-2}f)$ and $H_1(2^{j-1}f)$. The width of $\{h_{j,l}\}$ is equal to $L_j = (2^j - 1)(L - 1) + 1$.

We define the undecimated wavelet representation of $\{X_t\}$ by filtering this time series using $\{h_{j,l}\}$ for different values of positive integer j . The level j wavelet coefficient process, $\{W_{j,t}\}$, captures the changes in averages in the values of the time series $\{X_t\}$ over the (time) scale $\tau_j = 2^{j-1}$, and is defined by

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}. \quad (3)$$

The level j Daubechies wavelet filter has $L/2$ vanishing moments: i.e., $\sum_l l^k h_{j,l} = 0$ for $k = 0, \dots, L/2$. This means that this wavelet filter contains $L/2$ inherent differencing operations which allows us to decompose possibly nonstationary LRD series and to remove polynomial trends (Percival and Walden, 2000; Craigmile et al., 2005). Specifically, suppose that $\{X_t\}$ is a process of the form given by (1) with LRD parameter δ , but also assume an additive polynomial trend of degree r . Then as long as $\max(\delta, r+1) < L/2$, the level j wavelet coefficient process, $\{W_{j,t}\}$, will be a mean zero and stationary process. (Except in Section 6.2, we will assume no trend component for $\{X_t\}$.)

For our theoretical results we will also need the continuous wavelet transform. We define the wavelet function $\psi(\cdot)$ and scaling function $\phi(\cdot)$, both compactly supported on the interval $(-L+1, 0]$, whose Fourier transforms $\Psi(\cdot)$ and $\Phi(\cdot)$ are given respectively by

$$\Psi(f) = H_1(1/2f) \prod_{j=2}^{\infty} G_1(2^{-j}f), \quad \Phi(f) = \prod_{j=1}^{\infty} G_1(2^{-j}f).$$

The function $\psi(\cdot)$ corresponds to the continuous wavelet representation with its dyadic dilations and translations $\{\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)\}$ forming an orthogonal basis of all square integrable functions on the real line. It retains the vanishing moments property of the wavelet filters $\{h_{j,l}\}$, and at large values of j , $h_{j,l} \approx 2^{-j} \psi(-2^{-j}l)$, for each l . Furthermore, if $L > 2$, the wavelet function $\psi(\cdot)$ satisfies the regularity condition that for some $\alpha > 1$,

$$\sup_{u \in \mathbb{R}} |\Psi(u)| (1 + |u|)^{\alpha} < \infty. \quad (4)$$

This condition characterizes the decay of the function $\psi(\cdot)$ at infinity, which in turn corresponds to a smoothness property of the wavelet function $\psi(\cdot)$ at the origin. We will make this assumption from this point.

3.1 The wavelet variance

The level $j - 1$ wavelet variance (e.g., Percival and Walden, 2000, Chapter 8) is $v_j^2 = \text{var}\{W_{j,t}\}$; that is, the variance of level j wavelet coefficient process, as defined by (3). When $\{X_t\}$ is Gaussian with an SDF satisfying (1) and (2), the SDF mimics a pure power law at zero frequency. When the wavelet function $\psi(\cdot)$ satisfies (4) with a wavelet filter length of $L > \max\{2, 2\delta\}$, an immediate consequence of (4) is that the wavelet variances at larger levels become close to $S_0(0) K_\delta 2^{j(2\delta-1)}$, with

$$K_\delta = \int_{-\infty}^{\infty} |f|^{-2\delta} |\Psi(f)|^2 df.$$

Indeed, in this case, Moulines et al. (2007b, Theorem 1) show a more precise bound takes the form of

$$|v_j^2 - S_0(0) K_\delta 2^{j(2\delta-1)}| \leq C_1 C_2 S_0(0) 2^{j(2\delta-1-\zeta)}, \quad (5)$$

in which C_2 is an additional positive constant. Thus log of the wavelet variances at large scales fall on about a straight line, the slope of which is related to the LRD parameter δ : for large levels j ,

$$\log(v_j^2) \approx C + (\delta - 1/2)2j \log(2), \quad (6)$$

for some constant C . We use this property for the estimation of the LRD parameter.

3.2 Estimating the wavelet variance

Suppose that we observe a time series of length N that can be regarded as a realization of $\{X_t\}$, with gap process $\{\eta_t\}$. Suppose that the LRD parameter $\delta < L/2$ so that increments of $\{X_t\}$ has a square integrable SDF. Then the corresponding wavelet coefficients $\{W_{j,t}\}$, at a given level j , form realizations of a zero mean covariance stationary Gaussian time series and the wavelet variances satisfies (5). Then we can estimate the level j wavelet variance using the semi-variogram-type estimator (Mondal and Percival, 2010):

$$\hat{v}_j^2 = -\frac{1}{2N_{j,t}} \sum_{t=L_j-1}^{N-1} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \hat{\pi}_{j,l,l'} \times (X_{t-l} - X_{t-l'})^2 \eta_t - l \eta_{t-l'}, \quad (7)$$

where

$$\hat{\pi}_{j,l,l'}^{-1} = \frac{1}{N_{j,t}} \sum_{t=L_j-1}^{N-1} \eta_t - l \eta_{t-l'}, \quad 0 \leq l, l' \leq L_j - 1, \quad (8)$$

estimates the quantity $\pi_{l,l'}^{-1}$ at wavelet level j . In the above expression $N_j = N - L_j + 1 > 0$ denotes the number of wavelet coefficients on level j that are unaffected by filtering at the boundaries, and is the reason that the estimator of $\pi_{l,l'}$ becomes level dependent. While the wavelet filters $h_{j,l}$ can be calculated using a discrete wavelet transform (Percival and Walden,

2000, Section 4.12), except in non-gappy cases, the rest of the calculation is computationally intensive. Also note that (7) is invariant to affine shifts.

In (8) we must assume that each $\hat{\pi}_{j,l,l'}^{-1}$ is strictly positive, which holds asymptotically, but may fail for finite data with too many gaps (Mondal and Percival, 2010). It is unlikely that we will get some of these empirical probabilities to be zero unless we have 50% or more missing values. In such cases we recommend adjusting the estimator of the LRD parameter given by (15) to include only the wavelet levels with positive estimates of the wavelet variance. (Effectively we set a smaller maximum level, J .)

When there are no missing values $\hat{v}_j^2$ collapses to the usual unbiased estimator of the wavelet variance, $\hat{v}_j^2 = N_j^{-1} \sum_{t=L_j-1}^{N-1} W_{j,t}^2$. Conditioned on any gap pro-cess, the estimator given by (7) is unbiased.

3.3 Asymptotic properties of the estimated wavelet variance

Letting $V_{l,l',t} = \eta_{t-l} \eta_{t-l'}$, we assume that the covariances between $V_{l,l',t}$ and $V_{m,m',t'}$ are absolutely summable over t and t' for any indices l, l', m , and m' , and the higher order cumulants satisfy

$$\sum_{t_1=0}^{N-1} \dots \sum_{t_n=0}^{N-1} |\text{cum}(V_{l_1,l'_1,t_1}, \dots, V_{l_n,l'_n,t_n})| = o(N^{n/2})$$

for any indices $l_1, \dots, l_n$ and $l'_1, \dots, l'_n$ (see (Brillinger, 1981, Section 2.3) for a definition and properties of cumulants). Now define the process $\{Y_{j,t}\}$ given by

$$Y_{j,t} = -\frac{1}{2} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \pi_{l-l'} \times (X_{t-l} - X_{t-l'})^2 V_{l,l',t}. \quad (9)$$

This process has mean v_j^2 and collapses to $W_{j,t}^2$ in the gap free case. Then in terms of the $\{Y_{j,t}\}$ process the joint large sample properties of the semi-variogram type estimators of the wavelet variance can be summarized in the following theorem.

Theorem 1 Assume $\delta < 5/4$. Suppose we restrict to wavelet levels $j = J_0, J_0 + 1, \dots, J$, for fixed positive integers J_0 and J such that $J_0 < J$. For j in this range let $p_{j,J_0} = (N_j/N_{J_0})^{1/2}$ be the square root of the proportion of non-boundary wavelet coefficients on level j relative to level J_0 . As N approaches infinity,

$$N_{J_0}^{1/2} (p_{J_0,J_0}(\hat{v}_{J_0}^2 - v_{J_0}^2), p_{J_0+1,J_0}(\hat{v}_{J_0+1}^2 - v_{J_0+1}^2), \dots, p_{J,J_0}(\hat{v}_J^2 - v_J^2)),$$

is jointly asymptotically $(J - J_0 + 1)$ -variate normal w mean zero and dispersion matrix Σ , which is equal to the matrix-valued SDF of the vector-valued process

$$(Y_{J_0,t}, Y_{J_0+1,t}, \dots, Y_{J,t})^T, t = 0, 1, \dots$$

evaluated at zero frequency.

The proof of this result is given in the Appendix.

It is important note that that the regression relationship implied by Moulines et al. (2007b, Theorem 1) and thus (6) does not require $\delta < 5/4$ (or indeed Gaussianity) – we only need the assumption that $f^{L-2\delta}S_0(f)$ is square integrable and when $S_0(f)$ is square integrable it is enough to assume that $\delta < L/2$. However in the gappy case, to obtain a joint asymptotic distribution for the wavelet variances and thus the estimator of the LRD parameter δ introduced in the next section, Theorem 1 restricts the range of the LRD parameter to be $\delta < 5/4$. This range is larger than that typically assumed for Fourier methods (usually $\delta < 1/2$, but larger values can be entertained with specific choice of taper; e.g., Fay et al., 2009). We can obtain asymptotic normality results when analyzing time series with LRD parameters $\delta > 5/4$, by differencing the time series beforehand to drop the degree of LRD (e.g., Craigmile and Percival, 2005; Mondal and Percival, 2010).

In the gap free case, we can replace the assumption in Theorem 1 that $\delta < 5/4$ by $\delta < L/2$. This allows us to extend the range of LRD parameter that can be estimated by using wavelet filters of longer widths, L . In the gap free case, the dispersion matrix Σ becomes the matrix-valued SDF of the vector-valued process

$$(w_{J_0,t}^2, w_{J_0+1,t}^2, \dots, w_{J,t}^2)^T,$$

evaluated at zero frequency.

4 Estimation of the LRD parameter

Abry and Veitch (1998) proposed a weighted-least-squares estimate of the LRD parameter δ using the log wavelet variances, based on a decimated wavelet transform. The estimate of the LRD parameter δ is linearly related to the slope parameter. In that paper, the theory is based on the assumption that the data is gap free and that the Daubechies wavelet filter (in fact any finite length wavelet filter) is a perfect band pass filter, implying that the wavelet variances between scales are uncorrelated. In subsequent papers the estimator does not use between-scale information, but the theory does. The conservative rate of decorrelation between wavelet scales and the empirical illustrations provided in Craigmile and Percival (2005), provide evidence of correlations between wavelet coefficients over adjacent scales. Thus our estimation scheme and theory allows for gaps in the series and correlations between wavelet scales.

Accounting for the dependence between the different wavelet levels, let β_0 and β_1 be the intercept and slope in regressing the log wavelet variance upon the log time scale using the relationship implied by (6); the estimate of the LRD parameter is then $\beta_1 + 1/2$. More

specifically, suppose we choose a range of wavelet levels, $J_0, \dots, J$, to analyze over – a discussion of ways to select these levels are given in the Section 4.1. With $K = J - J_0 + 1$, let $\mathbf{z} = (\log \hat{v}_{J_0}^2, \dots, \log \hat{v}_J^2)^T$ denote the K -vector of estimated log wavelet variances. Let $\mathbf{A}$ be the $K \times 2$ design matrix with j th row ($j = 1, \dots, K$) given by $(1, 2(J_0 + j - 1)\log 2)^T$. (The definition of $\mathbf{A}$ follows from (6).) Using the delta method, and defining $\sigma_{j,j'}$ to be the (j, j') element of the dispersion matrix Σ of Theorem 1, the large sample dispersion matrix of $\mathbf{z}$, $\mathbf{D}$ say, has (j, j') element

$$d_{j,j'} = \frac{\sigma_{J_0-1+j, J_0-1+j'}}{N_{J_0-1+j}^{1/2} N_{J_0-1+j'}^{1/2} v_{J_0-1+j}^2 v_{J_0-1+j'}^2}. \quad (10)$$

Then, if the components of $\mathbf{D}$ were known, we could estimate the LRD parameter δ via $\hat{\delta} = \hat{\beta}_1 + 1/2$, where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$ is the general least squares estimator of $\beta = (\beta_0, \beta_1)^T$:

$$\hat{\beta} = (\mathbf{A}^T \mathbf{D}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{D}^{-1} \mathbf{z}.$$

The estimator of δ can be rewritten as,

$$\hat{\delta} = 1/2 + \mathbf{a}^T \mathbf{z}, \quad (11)$$

where letting $\mathbf{1}$ be a K -vector of ones, $\mathbf{w} = (2j \log 2 : j = J_0, \dots, J)$, $B_1 = (\mathbf{1}^T \mathbf{D}^{-1} \mathbf{1})$, $B_2 = (\mathbf{w}^T \mathbf{D}^{-1} \mathbf{1})$, and $B_3 = (\mathbf{w}^T \mathbf{D}^{-1} \mathbf{w})$,

$$\mathbf{a}^T = \frac{(\mathbf{B}_1 \mathbf{w}^T - \mathbf{B}_2 \mathbf{1}^T) \mathbf{D}^{-1}}{B_1 B_3 - \{B_2\}^2}. \quad (12)$$

In this setting, the following theorem summarizes the large sample properties of $\hat{\delta}$.

Theorem 2 Assume that the conditions underlying Theorem 1 hold with

$$\lim_{N \rightarrow \infty} N/2^J = \infty, \quad \lim_{N \rightarrow \infty} 2^{-J_0(1+2\zeta)} N = 0, \quad (13)$$

and

$$\sup_{j \geq 1} \sigma_{j,j} / (v_j^4 2^j) < \infty. \quad (14)$$

Then, as N approaches infinity,

$$N_J^{1/2} (\hat{\delta} - \delta) / \sqrt{\kappa}$$

is asymptotically normal with mean zero and dispersion 1. Here the term κ is equal to

$$\kappa = \frac{\mathbf{1}^T \boldsymbol{\Omega}^{-1} \mathbf{1}}{(\mathbf{1}^T \boldsymbol{\Omega}^{-1} \mathbf{1})(\mathbf{w}^T \boldsymbol{\Omega}^{-1} \mathbf{w}) - (\mathbf{w}^T \boldsymbol{\Omega}^{-1} \mathbf{1})^2},$$

where $\boldsymbol{\Omega}$ is a $K \times K$ matrix with (j, j') element $\sigma_{j, j'} / (v_j^2 v_{j'}^2)$.

Condition (14) ensures that the variance of the normalized LRD estimators stabilizes $N \rightarrow \infty$. It ensures the best possible rate of convergence of the LRD estimator (which is $N^{1/2}$). In the proof of the theorem we show that condition (14) always holds for gap-free processes.)

In practice $\mathbf{D}$ that appears in the estimator of the LRD parameter δ is unknown and we replace $\mathbf{D}$ by some $K \times K$ positive definite matrix $\tilde{\mathbf{D}}$, which may or may not be a function of the data; our estimator of δ then becomes

$$\hat{\delta} = 1/2 + \tilde{\mathbf{a}}^T \mathbf{z}, \quad (15)$$

where $\tilde{\mathbf{a}}$ is defined by substituting $\tilde{\mathbf{D}}$ for $\mathbf{D}$ in (12). For the estimator of the LRD parameter and its associated variance estimator, we consider two cases of $\tilde{\mathbf{D}}$ (each case relies on estimates of $\sigma_{j, j'}$, $\hat{\sigma}_{j, j'}$, that we discuss in detail below):

1. Choose $\tilde{\mathbf{D}}$ to have (j, j') element formed by plugging in $\hat{\sigma}_{j, j'}$ and $\hat{v}_j$ into (10):

$$\tilde{d}_{j, j'} = \frac{\hat{\sigma}_{j, j'}}{(N_j N_{j'})^{1/2} \hat{v}_j \hat{v}_{j'}}. \quad (16)$$

2. Define $\tilde{\mathbf{D}}$ via (16) for the diagonal elements only and zero elsewhere.

In the first case we call the estimator of the LRD parameter the full (dispersion) estimator, denoted $\hat{\delta}_F$. The second case, with a choice of diagonal $\mathbf{D}$ corresponds in the gap free case to the estimator of Abry and Veitch (1998). Thus we call this estimator of the δ the diagonal (dispersion) estimator, denoted $\hat{\delta}_D$. It can be simplified to

$$\hat{\delta}_D = \frac{1}{2} + \left[\sum_{j=J_0}^J (j B_4 - B_5) \tilde{d}_{j, j}^{-1} \log \hat{v}_j^2 \right] / (B_4 B_6 - \{B_5\}^2),$$

where $B_4 = \sum_{j=J_0}^J \tilde{d}_{j, j}^{-1}$, $B_5 = \sum_{j=J_0}^J j \tilde{d}_{j, j}^{-1}$, and $B_6 = \sum_{j=J_0}^J j^2 \tilde{d}_{j, j}^{-1}$.

A large sample to the variance of the LRD estimator is

$$\tilde{\kappa} = \tilde{\mathbf{a}}^T \mathbf{D} \tilde{\mathbf{a}}. \quad (17)$$

We obtain a sandwich estimate for this variance (e.g. White, 1980) by plugging in a estimate for $\mathbf{D}$ into (17) (not necessarily the same form used in (15)).

In practice we need to estimate Σ . In the gap free case for $j = j'$, define the sample cross-covariance at $\log k = -N_j + 1, -N_j + 2, \dots, N_j - 2, N_j - 1$ by

$$c_{j,j'}(k) = (N_j N_{j'})^{-1/2} \times \sum_{t=\max\{L_{j'}-1, L_j-k-1\}}^{\min\{N-1-k, N-1\}} W_{j,t+k} W_{j',t}, \quad (18)$$

and similarly define $c_{j,j'}(k)$ for $j > j'$. Then we estimate $\sigma_{j,j'}$ by $\hat{\sigma}_{j,j'} = \sum_k c_{j,j'}^2(k)$.

Because of the random nature of the gap process, we need to use a different estimator of Σ in the gappy case. We propose multitaper methods (e.g., Percival and Walden, 1993) to estimate $\hat{\sigma}_{j,j'}$. Let $R > (J - J_0 + 1)$ be some positive odd integer. At each wavelet level j we define a set of R orthonormal Slepian tapers $\{\lambda_{j,t}, t = 0, \dots, N_j - 1\}$ with a design bandwidth of $(R + 2)/N_j$. Then defining

$$T_{j,r} = \sum_t \lambda_{j,r,t} \hat{Y}_{j,t}, \quad r = 0, \dots, R - 1, \quad (19)$$

to be the tapered versions of $\{Y_{j,t}\}$ (here $\hat{Y}_{j,t}$ in equation (19) is equation (9) with $\pi_{l,l'}$ replaced by $\hat{\pi}_{j,l,l'}$), we estimate $\sigma_{j,j'}$ with the regression-based covariance

$$\hat{\sigma}_{j,j'} = \frac{1}{R} \sum_{r=0}^{R-1} (T_{j,r} - \xi_j \lambda_{j,r,+}) (T_{j',r} - \xi_j \lambda_{j',r,+}), \quad (20)$$

where $\lambda_{j,r,+} = \sum_t \lambda_{j,r,t}$ and

$$\xi_j = \frac{\sum_r T_{j,r} \lambda_{j,r,+}}{\sum_r \lambda_{j,r,+}^2}.$$

Consistency of the estimator (20) can be proven along the lines of Walden (2000).

4.1 Selection of wavelet scales

In practice, as long as we do not have a very long time series, the number of scales that we analyze, $J - J_0 + 1$, can be treated as a finite number. The value of the upper level, J , should be chosen to be a time scale at which there are a sufficient number of wavelet coefficients to produce an estimate of the wavelet variance with a certain confidence. Namely, we choose J to be the maximum level at which the estimated wavelet variance (7) has a sufficiently large effective degrees of freedom, calculated using the formula $2N_j(\hat{v}_j^2)/\hat{\sigma}_{j,j}$ (Percival and Walden, 2000, Chapter 8). An additional advantage of having a large effective degrees of freedom is that we ensure that the sampling distribution of $\hat{v}_j^2$ is approximately Gaussian. Referring again to the form of SDF of the time series, given by (1), J_0 marks the initial level at which the effect of short memory spectral density diminishes and the power law takes over. Typically in many applications, J_0 can be set to be two or three. However, larger values of J_0 may be required if S_0 is rich in structure and the time series follows a quasi-periodic behavior. Fay et al. (2009) make the point that the selection of J_0 is equivalent to the

selection of bandwidth in Fourier domain. This is because J_0 determines the effective number of samples to be used in the estimation of the LRD parameter.

5 Monte Carlo simulations

Here we employ three Monte Carlo simulation studies to address various statistical aspects that arise in the estimation of the LRD parameter. In Section 5.1 we compare the performance of different estimators of the LRD parameter of fractionally differenced processes for different choices of wavelet scales, sample lengths, and wavelet filters for the gap free case. We investigate sandwich estimators for the standard error of the LRD parameter in Section 5.2, and the performance of estimators of the LRD parameter for gappy series is studied in Section 5.3. In Section 5.4 we compare our estimator to a wavelet-lifting-based estimation method proposed by Knight et al. (2016).

5.1 Estimating gap free fractionally differenced processes

In this section we consider the finite sample performance of the full dispersion ($\hat{\delta}_F$) and diagonal ($\hat{\delta}_D$) estimators of the LRD parameter, δ , for gap free fractional differenced (FD) processes. A FD process, arguably the simplest LRD process, is defined by a SDF given by (1), assuming that $S_0(f)$ is a fixed constant for all $f \in [-1/2, 1/2]$. We fixed $S_0(f) = 1$.

Our first study compares the performance of the full and diagonal estimators, as we vary the LRD parameter δ ($\delta = 0.25, 0.40, 0.80$) and the sample length N ($N = 1024, 2048, 8192$). For each combination we compare the two estimators under three different choices of wavelet filter (D(4), D(6), LA(8)), based on 10,000 realizations of FD processes, simulated via the Davies-Harte algorithm (Davies and Harte, 1987; Craigmile, 2003). For this simulation, we fixed $J_0 = 3$ and $J = 7$ (we discuss the choice of J_0 and J in the next simulation). Comparisons of the bias, standard deviation (SD), and root mean square error (RMSE) for these estimators are summarized in Figure 1. In the figure the numbers denote the length of the wavelet filter, L . The shaded rectangles denote 95% bootstrap confidence intervals for each summary measure (gray: diagonal LRD estimator, $\hat{\delta}_D$; black: full LRD estimator, $\hat{\delta}_F$).

For a fixed wavelet filter, estimation method (diagonal or full), and LRD parameter δ , the bias in the estimating δ becomes smaller in magnitude (i.e., closer to zero) as the sample size, N , increases. For these fixed settings, the SD and RMSE decreases as the sample size increases. This is to be expected as we have more data to estimate the LRD parameter.

For a fixed wavelet filter, estimation method (diagonal or full), and sample size, the bias is smaller in magnitude for smaller values of the LRD parameter δ . Also, the SD and RMSE is smaller for larger values of δ , compared to smaller values of δ . Again this makes sense, as estimation should be more difficult when the data is more dependent.

Comparing the wavelet filters and estimation methods, we learn that the diagonal estimator of the LRD parameter with the D(4) wavelet filter has the smallest bias, across all values of δ and N . There is a suggestion that the full estimator $\hat{\delta}_F$ of the LRD parameter with the D(4)

filter may have the smallest SD, although there is some overlap in the confidence intervals between the SDs estimated for the full and diagonal estimators when we use the D(4) filter for smaller samples sizes, especially for smaller values of the LRD parameter δ . At smaller sample sizes the RMSE of the diagonal and full estimators are smallest when the D(4) filter is used, as compared to using the longer D(6) and LA(8) filters. As the sample size increases, there is a suggestion that the diagonal estimator with the D(4) filter has the smallest RMSE over all estimators and filters, especially for larger values of the LRD parameters δ .

Our second simulation fixes $N = 8192$, $\delta = 0.40$ with a D(4) wavelet filter, but varies the choice of the levels J_0 and J . Again based on 10,000 simulations of FD processes, we compare the bias, SD, and RMSE of the full and diagonal estimators. This comparison is summarized in Figure 2. (The supplement to this article shows figures for $N = 1024$ and $N = 2048$.) Each row shows a different value of J_0 ($J_0 = 1, 2, 3, 4$) and each panel plots the bias, SD, or RMSE versus J . The diagonal estimator is denoted by the gray circles, and the full estimator is denoted by the black triangles. 95% bootstrap confidence intervals for each statistic are smaller in width than the size of the symbols, and thus are omitted. These figures demonstrate consistently larger biases (relative to the SD) for small values of J_0 , indicating that the asymptotic expansion in (5) does not provide an accurate approximation in small time scales (This problem is exacerbated for the full estimator, as compared with the diagonal estimator). This suggests that there is a slight model misspecification – examination of the SDF at shorter frequencies indicates that the log wavelet variances do not exactly fall on a straight line at the smaller wavelet levels. Increasing J_0 reduces the bias, but increases the SD. In terms of the RMSE, a good compromise is to take $J_0 = 3$; with this choice, the performance of diagonal and full estimators are more similar (as confirmed by the first simulation with $J_0 = 3$ and $J = 7$). The curvature in the RMSE versus J indicates that, as discussed in Section 4.1, we should pick J not to be too large or small; taking uncertainty of the statistics into consideration, for $J_0 = 3$ a value of $J = 8$ looks to have the smallest RMSE. The figures provided in the supplement suggest that compared to $J_0 = 3$, a value of $J_0 = 2$ produces smaller RMSEs at shorter sample lengths, with a compromise on the bias. To ensure small magnitude biases, we choose $J_0 = 3$ and $J = 7$ for our remaining simulation studies when the sample length is $N = 1024$.

5.2 Comparisons of standard errors in the gap free case

In practice, we make inference based on observed data, and it is crucial that we compute accurate standard errors from just one observed time series. Our next simulation thus assesses these estimated standard errors. We fix $N = 1024$, $J_0 = 3$, and $J = 7$, but vary the LRD parameter ($\delta = 0.25, 0.40, 0.80$) and the choice of wavelet filter (given previous results, we restrict to the shorter D(4) and D(6) filters that are more efficient with respect to LRD parameter estimation). Our results are based on 10,000 realizations of FD processes for each value of δ , and are summarized in Table 1. The first two lines of the table display the estimated biases for the diagonal ($\hat{\delta}_D$) and full ($\hat{\delta}_F$) estimators of the LRD parameter, δ . The third and six lines (denoted “diagonal, SD” and “Full, SD”) calculates estimates of the SD of the two estimators based on these 10,000 simulations. Our interest is to compare these two sets of values with a summary (the average) of the estimated standard errors, calculated

using the square root of (17). Equation (17) depends on the type of estimator and the form of dispersion matrix $\mathbf{D}$, so we compare three different estimated standard errors, which we now describe. Let $\hat{\mathbf{D}}$ be the estimated dispersion matrix of log wavelet variance computed via (16). Next we compute the standard error of $\hat{\delta}_D$ in two ways; first by ignoring the between-wavelet-scale correlations (ese_D ; summarized in line four of the table), and second by using the sandwich estimator which includes the between-scale correlations (ese_S ; line five of the table). Line seven of Table 1 displays the estimated standard error for $\hat{\delta}_F$, ese_F which includes the full estimated dispersion matrix in (17).

We see that the Monte Carlo averages of ese_D (fourth row) are consistently much smaller than the standard deviations of $\hat{\delta}_D$ (third row), whereas averages of ese_S (fifth row) are closer, but still smaller, than the standard deviations of $\hat{\delta}_D$ (third row). Thus, the inclusion of between-wavelet-scale correlations with the sandwich estimator for the variance significantly improves the negative biases in the standard errors, but does not completely correct them. The SDs of $\hat{\delta}_F$ are lower than the SDs of $\hat{\delta}_D$ for this experimental setting. Again the average of the standard errors for the full estimator, ese_F are smaller than the actual SDs of $\hat{\delta}_F$. (We will discuss further the consistent bias in the SD estimates in Section 7.)

Next we briefly compare results for the two different wavelet filters. The Monte Carlo SDs $\hat{\delta}_D$ and $\hat{\delta}_F$ are about 10% – 13% larger when we use the $D(6)$ filter as compared to the $D(4)$ filter, which occurs largely due to the decrease in effective degrees of freedom with the increase in filter length. Second, SDs $\hat{\delta}_D$ and $\hat{\delta}_F$ are much closer to each other for $D(6)$ wavelet filter because the wavelet scales decorrelate better (e.g., Craigmile and Percival, 2005). Accordingly we see similar effects on the estimates of standard errors. In other words, estimates of standard errors are larger for $D(6)$, and estimates of standard errors for $\hat{\delta}_D$ and $\hat{\delta}_F$ are closer for $D(6)$. Furthermore, the decrease in effective degrees of freedom for $D(6)$ has some negative effect on the biases in standard errors; there is a little more bias in the averages of ese_S and ese_F compared to the $D(4)$ case.

5.3 Performance of the gappy LRD parameter estimates

We now assess the performance of the gappy estimator of the LRD parameter. We generate 10, 000 fractionally differenced time series of length 1024 each, with a LRD parameter $\delta = 0.40$. (The supplement to this article shows results for $\delta = 0.25$ and 0.80 .) Once we generate a time series, we considered three different gap patterns in our simulations:

1. No gaps.
2. Binomial gaps. We consider two cases with about 10% and 20% gaps at random.
3. Markov gaps. The gaps are dictated by a realization of a stationary Markov chain $\{\eta_t\}$ with $P(\eta_t = 0, | \eta_{t-1} = 0) = 0.05$, $P(\eta_t = 0, | \eta_{t-1} = 1) = 0.10$, and $P(\eta_t = 0) = 0.095$.

We then estimate the LRD parameter and obtain its standard errors. As in Section 5.2, we set $J_0 = 3$ and $J = 7$, but restrict ourselves only to the $D(4)$ wavelet filter. There are two sources

of variation in the gappy estimator: the variogram estimator of the wavelet variances, and the multitaper estimate of covariance. We distinguish between these effects by running the no gap case, where we estimate the LRD parameter using the gap free estimator based on sample covariances defined by equation (18) (“Cov”), and the gappy multitaper method with $R = 7$ and $R = 9$ Slepian tapers. We use the multitaper method with $R = 7$ and $R = 9$ for the gap patterns. Table 2 summarizes this experiment with the notation borrowed from Section 5.2.

The first two lines of the table show the bias for the diagonal ($\hat{\delta}_D$) and full ($\hat{\delta}_F$) estimators of δ . The standard deviations of these estimates are given on lines three and six of the table. Regardless of the estimator (diagonal or full), we see that the bias differs little in comparing the gappy methods to the gap free methods, but that the standard deviations are slightly higher. The increase in the standard deviations are both due to the use of the multitaper estimates of the co-variance as well as the increase in the uncertainty of the log wavelet variance estimates due to gaps. Typically in finite samples, multitaper estimators of the variances and covariances are slightly biased, and these biases in the between-wavelet-scale correlations are increasing the SD of the full dispersion estimator, making it a worse estimator of the LRD parameter δ . In other words, we are paying a price for not knowing the between-wavelet-scale correlations, and the RMSE for the diagonal estimator is smaller than that of the full estimator.

Next we focus on comparing the averages of the estimated of standard errors of the diagonal and full estimators. Let $\tilde{\mathbf{D}}$ be the estimated dispersion matrix of log wavelet variances computed via the Slepian multitaper method. Then, for each simulated gappy time series, we compute the standard error of the full estimator using equation (17); line seven of Table 2 gives the average of these values. We compute the standard error for the diagonal estimators in two ways: first by discarding all between-wavelet-scale correlation information (line four), and second by using the sandwich estimator (line five). We observe that the estimated standard error of the full estimator is heavily biased. It underestimates the SD by almost a factor of two or three. The estimated standard errors for the diagonal estimator is also biased, but to a lesser extent than the bias obtained for the full estimator. Using the sandwich estimator we get a slightly less biased estimate of the standard error.

Comparing the number of tapers for the no gaps case, we observe small differences in the results, with the exception of the standard deviation for the full estimator: in that case the standard deviation decreases by 16% when we move from $R = 7$ to $R = 9$ – also histograms of the estimates obtained indicate that the distributions are less heavy tailed for the larger value of R (we obtain more efficient estimates of the between-wavelet-scale covariances).

We see little difference in the bias when we move from 10% to 20% gaps in the binomial case, as well as for the Markov gaps case. The standard errors are slightly larger for the binomial 20% gaps case, as compared with the binomial 10% and Markov gaps case, indicating that our estimates of the LRD parameter are more uncertain as we increase the proportion of missing data. Across the different gappy cases, the results change little as we vary from $R = 7$ to $R = 9$. In the gappy cases, again we underestimate the standard error of

our estimate of the LRD parameter and the sandwich estimator is the least biased of all the estimates of the standard errors.

In conclusion, we find that our estimators perform well in terms of bias, SD, and RMSE, even in the gappy case. However, as was observed in the gap-free case, the estimated standard error of our estimator is underestimated. This level of underestimation is greater in the gappy case, compared to the gap-free case – using the sandwich estimator of the standard error helps reduce this bias. We found little difference in the effect of changing the number of tapers from $R = 7$ to $R = 9$, unless we want to reduce the standard deviation of the full estimator of the LRD parameter. We discuss the underestimation of variance further in Section 7.

5.4 Comparing to a wavelet-lifting-based-estimation scheme

As discussed in Section 1, the theory for the estimation of the SDF has remained unsolved, limiting the ability to further provide a useful spectral based estimate of the LRD parameter. We thus instead choose to compare our estimator to Knight et al. (2016), who develop an estimation method based on wavelet lifting.

In our limited study we simulate 1,000 realizations of $FD(d)$ processes, for $\delta = 0.25, 0.40$ and 0.80 , and simulate a binomial gappy scheme with about 10% gaps for each realization. We estimate the LRD parameter using our diagonal estimator with the D(4) wavelet filter and the Knight et al. estimator that is implemented in the liftLRD R package. We set $J_0 = 1$ to match with Knight et al. (2016) who use all the wavelet levels in their estimator. A comparison of the bias, SD, and RMSE is shown in Table 3. As can be seen, our diagonal estimator has a smaller bias and RMSE than the Knight et al. estimator across all values of δ . On the other hand, our estimator has a larger SD than the Knight et al. estimator. As discussed previously, it is possible to reduce the bias and RMSE of our estimator further by choosing $J_0 = 2$. This clearly demonstrates the efficacy of our estimation method for estimating LRD in gappy series.

6 Applications

In this section we apply our method of estimating LRD for gappy time series to the analysis of sea-ice drafts series in the Arctic Ocean and to modeling spatially located time series of daily average surface temperatures in south central Sweden.

6.1 Estimating LRD in sea-ice draft series

Percival et al. (2008) (hereafter PRTG) investigate whether sea-ice drafts series (a surrogate for ice thickness) exhibit LRD, when viewed as a function of distance. They consider data taken from two SCientific ICe Expedition (SCICEX) cruises within the Arctic Ocean (Original source: the Snow and Ice Data Center in Boulder, Colorado, USA)¹. A plot of the gappy average sea-ice draft (in meters) versus the distance in kilometers is shown in panel (a) of Figure 3. There are 21.4% of the values that are missing from this series. PRTG interpolate the data, and then investigate the statistical properties of the draft series,

¹The dataset we use in this article can be downloaded from <http://staff.washington.edu/dbp/DATA/draft-profile.txt>.

indicating that, after detrending, LRD is more plausible than short range dependence. With an assumption that the sea-ice draft series follows a $FD(\delta)$ model, they obtain a maximum likelihood estimate of $\hat{\delta}_{MLE} = 0.27$ with an approximate 95% confidence interval for δ of $[0.21, 0.33]$. Further, with a Haar wavelet-based weighted least squares on model-imputed-gap-filled data, they obtain an estimate of the LRD parameter of $\hat{\delta} = 0.26$ and a 95% confidence interval for δ nearly as narrow as the one based on maximizing the likelihood.

Our methods in Section 4 do not permit the use of the Haar wavelet filter with $L = 2$, and so here we use a Daubechies D(4) wavelet filter in combination with the semi-variogram type estimator $\hat{v}_j^2$ as defined by equation (7). The black circles in panel (b) of Figure 3 display $\log \hat{v}_j^2$ versus $\log(\tau_j)$ for $j = 1, \dots, 7$. In the panel, the black vertical lines show the approximate 95% confidence intervals based on multitaper estimates given in equation (20) with $R = 7$ tapers. The choice of a Daubechies D(4) filter allows us to accommodate a richer form of the short memory spectral density, $S_0(\cdot)$, in the LRD parameter estimation. To facilitate direct comparison with PRTG, we ignore any evidence of trends and periodicity in the data and use a minimum wavelet level of $J_0 = 1$. Unlike the maximum wavelet level of $J = 9$ used by PRTG, our choice of wavelet filter restricts us to go beyond $J = 7$ (our method requires that the number of wavelet coefficients on level J that are unaffected by filtering at the boundaries is positive; i.e., $N_j > 0$). Based on this choice of wavelet levels, the diagonal and the full dispersion estimates of the LRD parameter are $\hat{\delta}_D = 0.26$ and $\hat{\delta}_F = 0.25$, respectively. Next, we compute the estimated standard error of $\hat{\delta}_D$ in two ways: first by ignoring all between-wavelet-scale correlations, and then by applying the sandwich estimator using all the between-wavelet-scale correlations. As stated in Section 3.3, our estimators are robust to mild departures from Gaussianity. Also our estimators of δ , fit using least squares, indicate that $\delta < 5/4$. Thus using our asymptotic theory, approximate 95% confidence intervals for δ based on these two standard errors are, respectively, $[0.19, 0.33]$ and $[0.16, 0.36]$. Calculating a standard error for $\hat{\delta}_F$ using all the between-wavelet-scale correlations, an approximate 95% confidence interval for δ is $[0.23, 0.27]$. As evidenced by our simulation studies, we judge that the full dispersion estimate is slightly negatively biased, and that the approximate 95% confidence interval based on the sandwich standard error estimate and the diagonal estimate, i.e., $[0.16, 0.33]$, provides the most reliable conservative range of possible values for δ . The best fitted least square line drawn in panel (b) of Figure 3 with slope $\hat{\delta}_D$ clearly indicates that LRD prevails across all scales in this ice draft series.

Next, we take up the issue that linearly interpolating the gaps may introduce a bias in the LRD estimates, and underestimate the variability of the estimates. The log wavelet variances obtained after local linear interpolation of the draft series are indicated by the gray triangles in panel (b) of Figure 3. The corresponding gray vertical lines show the approximate 95% confidence intervals based on a Gaussian assumption for the data. These biased estimates are quite different than those derived from Section 4, and thus it is misguided to try to find information on the LRD parameter from the linearly interpolated data. This necessitates the use of model-based methods for missing data, as was suggested by PRTG, or the use of our gappy methodology, which requires less modeling assumptions. The very fact that our

semiparametric approach provides inference that is in good agreement with the parametric assumption of PRTG grants practitioners an advantage at the exploratory stage of the analysis.

6.2 LRD in daily temperature series in south central Sweden

Craigmile and Guttorp (2011) analyze daily time series of average temperatures collected at 17 sites in south central Sweden. The series were downloaded from the Swedish Meteorological and Hydrological Institute². The locations of these measurements are shown in the left hand panel of Figure 4. While all daily temperature series have an end date of June 30, 2008, the length of each series, and pattern of missingness vary by location (for a summary see the right hand panel of Figure 4).

After a naive interpolation of these missing values, Craigmile and Guttorp (2011) find that all temperature series in south central Sweden are marked by presence of time-trends, strong yearly seasonality, and complex stochasticity that include at least some seasonal volatility. These findings make the investigation of LRD in temperature series challenging, particularly when we need to discount the effect of annual cycles in series of a length between 12 and 48 years.

To elaborate on this point, the top three panels of Figure 5 displays the log wavelet variance profile of the raw daily average temperatures from three locations (Malung, Kilsbergen-Suttarboda, and Visby Flygplats), treating each series as realizations of an intrinsically stationary Gaussian process (nonstationary series that are stationary after differencing). Although the actual values differ, these plots show that the changes of temperature averages over successively longer time scales at the three different locations are remarkably similar. Furthermore, we see a peak at a time scale of $\tau_7 = 2^7 = 128$ days, spaced 256 days apart, which is due to the presence of a strong seasonal cycle. For the first and third locations, the wavelet variances subside after this time scale τ_7 , but it is not obvious if these decreases in the variances are linear on the log scale. This casts doubt that the effect of the seasonality is restricted to only one scale; it washes across a number of adjacent scales. At the second location, it is hard to explore the LRD because of the lack of longer time scale information. Overall, the effect of seasonality is so prominent that it is impossible to get a direct idea of LRD just from the raw series.

Hence, we adjust for seasonal effects and volatility in order to obtain some meaningful conclusions about LRD in the data. Following the procedures described in Craigmile and Guttorp (2011), first we remove a yearly seasonality from each series and then divide each deseasonalized series by estimates of the time-varying standard deviations to obtain gappy residual series that are nearly stationary and representative of LRD, short range dependencies (“weather”), as well as overall temperature trends. Normal quantile plots (omitted) indicate that normality for each gappy residual series is reasonable. Then, for each residual series, we apply our wavelet variance methodology to perform a site by site statistical inference. The bottom three panels of Figure 5 displays the log wavelet variance profile of the deseasonalized and standardized series for the same three locations (plots of

²<http://www.smhi.se/klimatdata/meteorologi/dataserierfor-observationsstationer-1961-2008-1.7375>

the log wavelet variance at all locations are shown in the supplement). We see that the structure of these wavelet variances of the deseasonalized and standardized series are markedly different from the raw series, a clear indication of the fact that seasonality was affecting the variability across different time scales. But once seasonality and volatility are removed from the data, the inherent differencing operations in the wavelet filters automatically accommodates for the presence of trends (see Section 3), and we see that significant LRD emerges from short range dependencies in each of the residual series (The supplement shows this to be the case at all locations). In particular, the log wavelet variances of the deseasonalized and standardized series at each location are diminishing linearly from time scales $\tau_4 = 2^{5-1} = 16$ onwards, permitting us, at each location, to estimate the LRD parameters and to compute their standard errors based on a value of $J_0 = 5$.

Panel (a) of Figure 6 displays the diagonal estimates of the LRD parameters for all 17 locations along with 95% confidence intervals for the LRD parameter based on sandwich estimates of the standard errors, which we found to be reliable through our simulation studies. As in the ice profile example, from the least square lines, there is clear evidence that $\delta < 5/4$, and thus our asymptotic theory applies. Furthermore, to better understand the geographic variations in these location-wise LRD estimates, panel (b) of Figure 6 shows these estimated LRD values on a map of south central Sweden. Typically, the diagonal estimates of the LRD parameters vary from 0.14 to 0.31, but taking uncertainty into consideration, most, but not all, 95% confidence intervals for the LRD parameters overlap. For example, the temperatures from location 14 in the south east (Harstena A) exhibit stronger LRD, as compared with the remaining locations. This analysis confirms that it is climatologically interesting to study the spatially-varying LRD in this region further, as studied by Craigmile and Guttorp (2011).

7 Discussion

We open our discussion by briefly commenting upon the estimation of LRD parameter for systematic gaps, which occurs, for example, in economic applications where trading is closed on weekends. Extending Mondal and Percival (2010), we can show that the estimators of wavelet variances in (7) are unbiased and jointly asymptotically normal. Furthermore, we can compute the asymptotic dispersion matrix via multitaper methods. This paves the way for statistical estimation and inference of the LRD parameter in the systematic gap case, in a way similar to what we have developed for random gaps.

In discussing the general evaluation of the LRD parameter estimation, we now interrogate some of the assumptions on which our method is based. For instance, one assumption that we have taken for granted is that of asymptotic normality of the estimators of the LRD parameter. This assumption can require reasonable lengths of time series. For small samples, when the normal asymptotics fails to yield accurate approximations to the small-sample distributions, an alternative is to base our inference of the LRD parameter on chi-square approximations, as shown in Percival and Walden (2000) and Percival and Mondal (2012). The chi-square approximation arises from the quadratic form of the estimator of a wavelet variance, which can be represented using a mixture of chi-square random variables with one degree of freedom. Typically, such approximations enjoy good small sample properties. In

the present context, we note that the chi-square approximations indicated in Percival and Mondal (2012) for the non-gappy case have the ability to improve biases in small samples. We are interested in extending this idea by developing Wishart approximations, both for non-gappy and gappy data, as the estimators of wavelet variances at several scales can be represented jointly as a mixture of Wishart random variables. Wishart approximations allow for the inclusion of information about the correlations between scales, while ensuring better small-sample bias and variance properties.

Another implicit assumption is that our time series are detrended or that we observe an additive polynomial of degree r , with $r + 1 \leq L/2$, where L is the length of the wavelet filter. If this is not the case, the presence of a long range trend, which can be difficult to detect and remove from the observed series, can affect the LRD parameter estimation (see, e.g., Künsch, 1986; Craigmile et al., 2004, 2005). Typically, small monotonic non-polynomial trends can have significant impact on the estimates of very large scale wavelet variances, and thus one needs to be careful about the selection of J . Rather than choosing J to be the maximum level at which the estimated wavelet variance is available, we should choose a smaller value of J to guard against possible trends. This necessitates further research on methods to choose J .

The value of J_0 can be chosen using some model selection procedure. Suppose, for $j = 1, 2, \dots, J - 1$, that $\mathcal{M}_j$ defines a model for which

$$\log v_j^2 - \log v_{j+1}^2 = \log v_{j+1}^2 - \log v_{j+2}^2 = \dots = \log v_{J-1}^2 - \log v_J^2 = 2\delta - 1,$$

and $\log v_1^2, \dots, \log v_J^2$ are arbitrary. We can then select J_0 by searching over models $\mathcal{M}_1, \dots, \mathcal{M}_{J-1}$ and identifying the model $\mathcal{M}_{J_0}$ which ‘best’ fits the estimates of log wavelet variances according to some suitable criteria. Thus, we can work with classical information criteria such as AIC (Akaike, 1973; Hurvich and Tsai, 1989) and BIC (Schwarz, 1978) or penalized likelihoods such as the one based on the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001) to obtain the best possible models. Since our estimators of the LRD parameter are based on an approximate regression relationship between the log wavelet variance and the log scale, the AIC index may be preferable, as it is perceived to have the ability to pick a slightly larger model. A larger model could guarantee a reasonable compromise between the bias and variance of the LRD parameter estimates. The other interesting choice is SCAD penalty, which gives rise to oracle properties, but its implementation requires sparse model specification in terms of second order differences of log wavelet variances. It is arguable that different information criteria or penalized likelihoods ascribe different significances to models, and as an informal diagnostics tools, it is best to summarize results for each of the best possible models. A Bayesian approach (possibly via reversible jump methods; Green, 1995) is also possible.

Our simulation studies indicate that our methods underestimate the standard errors of the LRD parameter estimates in both gap-free and gappy data scenarios. We learned that the problem was exacerbated when we used the multitaper spectral regression method to estimate the dispersion matrix Σ using (20). A sandwich estimator can decrease the bias in

estimating the SD of the estimate, but part of the underestimation of the SD is due to the fact that we do not account for the uncertainty in estimating the LRD parameter δ itself (e.g., see Craigmile et al., 2005, for an illustration of this problem in the parametric LRD setting). The use of a bootstrap (e.g., Sabatini, 1999) or the choice of a Bayesian estimator (e.g., Ko and Vannucci, 2006) can improve this situation. Also, improvements can be made to correct for this bias by making further parametric assumptions on S_0 , such as differentiability at or near zero frequency. We will investigate the advantages (and disadvantages) of making further assumptions in the future, by pursuing better nonparametric estimates of spectral densities at zero frequency (see also Heidelberger and Welch, 1981; Velasco and Robinson, 2001). We should also note that gappy methods are computationally intensive - further work will be needed to improve the computation time with longer series.

Finally, we note that this work can be extended to other time series models; e.g., locally stationary models (Roueff and von Sachs, 2010).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Prof. Noel Cressie and Prof. Mario Peruggia for helpful comments that improved this work.

Craigmile is supported in part by the US National Science Foundation (NSF) under grants NSF-DMS-1407604 and NSF-SES-1424481, and the National Cancer Institute of the National Institutes of Health under Award Number R21CA212308. Mondal is supported by National Science Foundation grant NSF-DMS-0906300 and NSF-DMS-1519890.

A Proofs

We first need the following lemmas.

Proposition 1 *Let $U_{l,l',t}$ and $V_{l,l',t}$ be stationary processes that are independent of each other for any choice of k, k', l and l' , and let*

$$U_{l,l',t} = \theta_{l,l'} + \int_{-1/2}^{1/2} e^{i2\pi ft} d\mathcal{U}_{l,l'}(f),$$

$$V_{l,l',t} = \omega_{l,l'} + \int_{-1/2}^{1/2} e^{i2\pi ft} d\mathcal{V}_{l,l'}(f)$$

be their respective spectral representations. For any k, k', l and l' , let $S_{k,k',l,l'}$ and $G_{k,k',l,l'}$ denote the respective cross spectrum between $U_{k,k',t}$ and $U_{l,l',t}$ and between $V_{k,k',t}$ and $V_{l,l',t}$. For fixed real numbers $\{a_{l,l'}\}$ define

$$Q_t = \sum_{l,l'} a_{l,l'} (U_{l,l',t} V_{l,l',t} - \theta_{l,l'} \omega_{l,l'}),$$

Then Q_t is a second order stationary process whose spectral density function is given by

$$S_Q(f) \equiv \sum_{k,k'} \sum_{l,l'} a_{k,k',l,l'} [\theta_{k,k'} \theta_{l,l'} G_{k,k',l,l'}(f) + -\omega_{k,k'} \omega_{l,l'} S_{k,k',l,l'}(f) + S * G_{k,k',l,l'}(f)], \quad (21)$$

where $S * G_{k,k',l,l'}(f) \equiv \int_{-1/2}^{1/2} G_{k,k',l,l'}(f - f') S_{k,k',l,l'}(f') df'$.

Proof The proof is given in the preprint that accompanies Mondal and Percival (2010).

Proposition 2 Let X_t be a real-valued Gaussian stationary process with zero mean, and SDF S_X that satisfies

$$\int_{-1/2}^{1/2} \sin^4(\pi f) S_X^2(f) df < \infty.$$

Let η_t be a binary-valued strictly stationary process that is independent of X_t and satisfies assumptions stated in Mondal and Percival (2010). Using $Y_{j,t}$ as defined by equation (9), then Q_b defined by

$$Q_t = \sum_{j=J_0}^J \gamma_j (Y_{j,t} - v_j^2), \quad (22)$$

is a second order stationary process whose SDF at zero is a strictly positive finite number.

Proof Note that Q_t in (22) can be written as

$$\sum_{l,l'=0}^{L_J-1} a_{l,l'} (U_{l,l',t} V_{l,l',t} - \theta_{l,l'} \omega_{l,l'}).$$

where $U_{l,l',t} = 1/2 (X_{t-l} - X_{t-l'})^2$, $V_{l,l',t} = \eta_{t-l} \eta_{t-l'}$, $a_{l,l'} = \sum_{j=J_0}^J \gamma_j h_{j,l} h_{j,l'} \pi_{l-l'}$, $\theta_{l,l'} = \gamma_{X,l-l'}$, and $\omega_{l,l'} = \pi_{l-l'}^{-1}$. Next, following the argument given in Mondal and Percival (2010), we can show that the bivariate process $\mathbf{U}_t = [1/2(X_{t-k} - X_{t-k'})^2, 1/2(X_{t-l} - X_{t-l'})^2]^T$, for any choice of k, k', l and l' , has a spectral matrix $\mathcal{S}_{\mathbf{U}}$ that is continuous. In addition, if $S_{k,k',l,l'}$ is the (k, k', l, l') component of $\mathcal{S}_{\mathbf{U}}$, then

$$\sum_{k,k'} \sum_{l,l'} h_{j,k} h_{j,k'} h_{j,l} h_{j,l'} S_{k,k',l,l'}(0) > 0.$$

Now by Lemma 1, the SDF of Q_t is given by the right-hand side of equation (21), where $G_{k,k',l,l'}$ is the cross spectrum between $\eta_{t-k} \eta_{t-k'}$ and $\eta_{t-l} \eta_{t-l'}$. Since $a_{l,l'} \omega_{l,l'} = \sum_{j=J_0}^J \gamma_j h_{j,l} h_{j,l'}$,

$$\begin{aligned}
& \sum_{k,k'} \sum_{l,l'} a_{k,k'} a_{l,l'} \omega_{k,k'} \omega_{l,l'} S_{k,k',l,l'}(0) \\
&= \sum_{j=J_0}^J \sum_{j'=J_0}^J \sum_{k,k'} \sum_{l,l'} \gamma_j \gamma_{j'} h_{j,k} h_{j,k'} h_{j',l} h_{j',l'} S_{k,k',l,l'}(0)
\end{aligned} \tag{23}$$

and after a few lines of algebra, reduces to

$$2 \int_{-1/2}^{1/2} \left(\sum_{j=J_0}^J \gamma_j |H_j(f)|^2 \right)^2 S_X^2(f) df$$

where $|H_j(f)|^2$ is the squared gain function associated with the level j wavelet filter $h_{j,f}$. This integral is a strictly positive finite number. Now $a_{k,k'}, a_{l,l'}, \theta_{k,k'}, \theta_{l,l'}, G_{k,k',l,l'}(0)$ and $a_{k,k'}, a_{l,l'}, S * G_{k,k',l,l'}(0)$ are nonnegative and finite because $G_{k,k',l,l'}$ and $S_{k,k',l,l'}$ are entries of spectral density matrices. Finally, $a_{k,k'}, a_{l,l'}, \theta_{k,k'}, \theta_{l,l'}, G_{k,k',l,l'}(0)$ is also nonnegative and finite. This completes the proof.

Before we prove Theorem 1, the following lemmas are consequences of Brillinger (1981, p.21), and are proved in Mondal and Percival (2010).

Proposition 3 Assume that X_t satisfies the conditions stated in Theorem 1. Let $U_{p,t} = 1/2(X_t - X_{t-1})^2$ and $EU_{p,t} = \theta_p$, in which $p = (l, l')$. Then for $n \geq 3$ and fixed $p_1, \dots, p_n$,

$$\sum_{t_1, \dots, t_n} |\text{cum}(U_{p_1, t_1} - \theta_{p_1}, \dots, U_{p_n, t_n} - \theta_{p_n})| = o(M^{n/2}),$$

where each t_i ranges from 0 to $M-1$.

Proposition 4 Let $U_{p,t}$ be either as in Lemma 3, and assume $\kappa_n(p_1, \dots, p_n, t_1, \dots, t_n) = \text{cum}(U_{p_1, t_1} - \theta_{p_1}, \dots, U_{p_n, t_n} - \theta_{p_n})$. Define for $i = 1, 2, \dots, n-1$

$$\kappa_n(p_1, \dots, p_n, t_1, \dots, t_i) = \sum_{t_{i+1}, \dots, t_n} M^{-1/2(n-i-1)} \kappa_n(p_1, \dots, p_n, t_1, \dots, t_n),$$

where the summation in t_j ranges from 0 to $M-1$. Then $\kappa_n(p_1, \dots, p_n, t_1, \dots, t_i)$ is bounded and satisfies

$$\sum_{t_1, \dots, t_i} \kappa_n(p_1, \dots, p_n, t_1, \dots, t_i) = o(M^{1/2(i+1)}) \tag{24}$$

for $i = 1, 2, \dots, n$.

Proof (Proof of Theorem 1)

First we will apply the Cramer Wold theorem to derive the asymptotic normality of the time averages of the multivariate process $(Y_{J_0,t}, Y_{J_0+1,t}, \dots, Y_{J,t})^T$. Thus, as in Lemma 2, let γ_j , $j = j_0, \dots, J$ be real constants and take the univariate time series formed by the linear combination, i.e.,

$$Q_t = \sum_{j=J_0}^J \gamma_j (Y_{j,t} - v_j^2) = \sum_{l,l'=0}^{L_J-1} a_{l,l'} (U_{l,l',t} V_{l,l',t} - \theta_{l,l'} \omega_{l,l'}),$$

where $U_{l,l',t} = 1/2(X_{t-l} - X_{t-l'})^2$, $V_{l,l',t} = \eta_{t-l} \eta_{t-l'}$, $a_{l,l'} = \sum_{j=J_0}^J \gamma_j h_{j,l} h_{j,l'} \pi_{l-l'}$, $\theta_{l,l'} = \gamma_{X,l-l'}$, and $\omega_{l,l'} = \pi_{l-l'}^{-1}$. We first prove a central limit theorem for $R = M^{-1/2} \sum_{t=0}^{M-1} Q_t$. Using Zurbenko (1986, p.2), to write the log of the characteristic function of R as

$$\log F(z) = \sum_{n=1}^{\infty} \frac{i^n z^n}{n!} \sum_{t_1, \dots, t_n} \frac{B_n(t_1, \dots, t_n)}{M^{n/2}},$$

where B_n is the n th order cumulant of Q_t , and each t_i ranges from 0 to $M-1$. Since Q_t is centered, $B_1(t_1) = 0$.

By Lemma 2, Q_t is stationary, the autocovariances $s_{Q,\tau}$ of Q_t are absolutely summable and $M^{-1} \sum_{t_1} \sum_{t_2} B_2(t_1, t_2) \rightarrow \sum_{\tau} s_{Q,\tau} = S_Q(0) > 0$. In order to prove the CLT for R , it suffices to show that $\sum_{t_1, \dots, t_n} M^{-n/2} B_n(t_1, \dots, t_n) \rightarrow 0$ for $n = 3, 4, \dots$

Now using Brillinger (1981, p.19), we break up the n th order cumulant as follows:

$$B_n(t_1, \dots, t_n) = \sum_{p_1} \dots \sum_{p_n} a_{p_1} \dots a_{p_n} \text{cum}(U_{p_1,t_1} V_{p_1,t_1} - \theta_{p_1} \omega_{p_1}, \dots, U_{p_n,t_n} V_{p_n,t_n} - \theta_{p_n} \omega_{p_n}).$$

But, applying the argument given in Mondal and Percival (2010), it follows that

$$\text{cum}(U_{p_1,t_1} V_{p_1,t_1} - \theta_{p_1} \omega_{p_1}, \dots, U_{p_n,t_n} V_{p_n,t_n} - \theta_{p_n} \omega_{p_n}) = o(M^{n/2}).$$

Note that $\hat{v}_j^2 - v_j^2$ is the average of Q_t over $L_j - 1 \leq t \leq N - 1$ with $n_{l,l'}$ replaced by its consistent estimate $\hat{\eta}_{l,l'}$. Then invoke Slutsky's theorem to complete the proof that $\hat{v}_j^2$ is jointly asymptotically normal.

Proof (Proof of Theorem 2) Note that $\mathbf{a}$ depends on J . We have $\mathbf{a}^T \mathbf{1} = 0$ and $\mathbf{a}^T \mathbf{w} = 1$. Then since $K = J - J_0 + 1$ is finite and $\mathbf{a}$ is normalized, as $J \rightarrow \infty$, $\mathbf{a}^T = O(1^T)$; that is, for each i , $a_i = O(1)$.

Letting $\xi = (\log v_j^2 : j = J_0, \dots, J)$, we have

$$\begin{aligned}
& N_j^{1/2}(\hat{\delta} - \delta)/\sqrt{\kappa} \\
&= N_j^{1/2}(1/2 + \mathbf{a}^T \mathbf{z} - \delta)/\sqrt{\kappa} \\
&= N_j^{1/2} \mathbf{a}^T (\mathbf{z} - \xi)/\sqrt{\kappa} + N_j^{1/2}(\mathbf{a}^T \xi - (\delta - 1/2))/\sqrt{\kappa} \\
&= N_j^{1/2} \mathbf{a}^T (\mathbf{z} - \xi)/\sqrt{\kappa} + N_j^{1/2} \mathbf{a}^T \{\xi - \mathbf{w}(\delta - 1/2) - \mathbf{1} \log(S_0(0)K_\delta)\}/\sqrt{\kappa}.
\end{aligned}$$

An immediate consequence of equations (2) and (13) is that the second term in the above equation goes to zero. The asymptotic normality of the first term holds by Theorem 1 and (14) in conjunction with the applications of the delta method and Slutsky's theorem.

To see that condition (14) holds in the gap-free case, note that $\sigma_{j,j}$ is equal to

$$2 \int_{-1/2}^{1/2} |H_j(f)|^4 S_X^2(f) df = 2S_0^2(0)K_{1,\delta} 2^{j(4\delta-1)} + \text{a small term},$$

where $K_{1,\delta} = \int |f|^{-4\delta} |\Psi(f)|^4 df$. This implies that

$$\sigma_{j,j}/\sqrt{\kappa} = c_1 2^j + \text{a small term},$$

for some constant c_1

References

- Abry P, Veitch D (1998) Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory* 44:2–15
- Akaike H (1973) Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60:255–265
- Bardet J, Lang G, Moulines E, Soulier P (2000) Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes* 3:85–99
- Beran J (1994) *Statistics for Long Memory Processes*. Chapman and Hall, New York, NY
- Bloomfield P (1970) Spectral analysis with randomly missing observations. *Journal of the Royal Statistical Society Series B* 32:369–380
- Brillinger DR (1981) *Time Series: Data Analysis and Theory*. Holden-Day Inc.
- Craigmile PF (2003) Simulating a class of stationary Gaussian processes using the Davies-Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis* 24:505–511
- Craigmile PF, Guttorp P (2011) Space-time modeling of trends in temperature series. *Journal of Time Series Analysis* 32:378–395
- Craigmile PF, Percival DB (2005) Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Transactions on Information Theory* 51:1039–1048
- Craigmile PF, Guttorp P, Percival DB (2004) Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics* 15:313–335
- Craigmile PF, Guttorp P, Percival DB (2005) Wavelet based estimation for polynomial contaminated fractionally differenced processes. *IEEE Transactions on Signal Processing* 53:3151–3161
- Dahlhaus R (1987) Nonparametric spectral analysis with missing observations. *Sankhyā The Indian Journal of Statistics Series A* 49:347–367
- Daubechies I (1992) *Ten Lectures on Wavelets CBMS-NSF Series in Applied Mathematics*, SIAM, Philadelphia, PA
- Davies RB, Harte DS (1987) Tests for Hurst effect. *Biometrika* 74:95–101

- Dunsmuir W, Robinson PM (1981) Asymptotic theory for time series containing missing and amplitude modulated observations. *Sankhyā The Indian Journal of Statistics Series A* 43:260–281
- Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96:1348–1360
- Fay G, Moulines E, Roueff F, Taqqu MS (2009) Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics* 151:159–177
- Geweke J, Porter-Hudak S (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4:221–238
- Granger CWJ, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1:15–29
- Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732
- Heidelberger P, Welch PD (1981) A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM* 24:233–245
- Hosking JRM (1981) Fractional differencing. *Biometrika* 68:165–176
- Hurst HE (1951) The problem of long-term storage in reservoirs. *Transactions of the American Society of Civil Engineers* 116:776–808
- Hurvich C, Lang G, Soulier P (2005) Estimation of long memory in the presence of a smooth nonparametric trend. *Journal of the American Statistical Association* 100:853–871
- Hurvich CM, Chen WW (2000) An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis* 21:155–180
- Hurvich CM, Ray BK (2003) The local Whittle estimator of long-memory stochastic volatility. *Journal of Financial Econometrics* 1:445–470
- Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. *Biometrika* 76:297–307
- Jensen M (1999) Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting* 18:17–32
- Jones RH (1971) Spectrum estimation with missing observations. *Annals of the Institute of Statistical Mathematics* 23:387–398
- Knight MI, Nason GP, Nunes MA (2016) A wavelet lifting approach to long-memory estimation. *Statistics and Computing* 27:1453–1471 [PubMed: 32025109]
- Ko K, Vannucci M (2006) Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes. *Journal of Statistical Planning and Inference* 136:3415–3434
- Kouamo O, Levy-Leduc C, Moulines E (2013) Central limit theorem for the robust log-regression wavelet estimation of the memory parameter in the gaussian semi-parametric context. *Bernoulli* 19:172–204
- Künsch H (1986) Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability* 23:1025–1030
- Künsch H (1987) Statistical aspects of self-similar processes. In: *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, Utrecht, Netherlands, vol 1, pp 67–74
- Mandelbrot BB, van Ness JW (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10:422–436
- McCoy EJ, Walden AT (1996) Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics* 5:26–56
- Mondal D, Percival D (2010) Wavelet variance analysis for gappy time series. *Annals of the Institute of Statistical Mathematics* 62:943–966
- Moulines E, Soulier P (2003) Semiparametric spectral estimation for fractional processes In: Doukhan P, Oppenheim G, Taqqu MS (eds) *Theory and Applications of Long-range Dependence*, Birkhäuser, Boston, MA, pp 251–301
- Moulines E, Roueff F, Taqqu MS (2007a) Central limit theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context. *Fractals* 15:301–313

- Moulines E, Roueff F, Taqqu MS (2007b) On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter. *Journal of Time Series Analysis* 28:155–187
- Moulines E, Roueff F, Taqqu MS (2008) A wavelet Whittle estimator of the memory parameter of a non-stationary gaussian time series. *The Annals of Statistics* 36:1925–1956
- Palma W (2007) *Long-Memory Time Series: Theory and Methods*. Wiley-Interscience, Hoboken, NJ
- Parzen E (1963) On spectral analysis with missing observations and amplitude modulation. *Sankhyā The Indian Journal of Statistics Series A* 25:383–392
- Percival DB, Mondal D (2012) A wavelet variance primer In: Rao TS, Rao CR (eds) *Time Series, Handbook of Statistics*, vol 30, Elsevier, Chennai, India, chap 22
- Percival DB, Walden AT (1993) *Spectral Analysis for Physical Applications Multitaper and Conventional Univariate Techniques*. Cambridge University Press, Cambridge, England
- Percival DB, Walden AT (2000) *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, England
- Percival DB, Rothrock DA, Thorndike AS, Gneiting T (2008) The variance of mean sea-ice thickness: Effect of long-range dependence. *Journal of Geophysical Research* 113:1958–1970
- R Core Team (2018) R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>
- Robinson PM (1994) Semiparametric analysis of long-memory time series. *Annals of Statistics* 22:515–539
- Robinson PM (1995a) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23:1630–1661
- Robinson PM (1995b) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23:1048–1072
- Roueff F, von Sachs R (2010) Locally stationary long memory estimation. *Stochastic Processes and their Applications* 121:813–844
- Roueff F, Taqqu MS (2009) Asymptotic normality of wavelet estimators of the memory parameter for linear processes. *Journal of Time Series Analysis* 30:534–558
- Sabatini AM (1999) Wavelet-based estimation of 1/f-type signal parameters: Confidence intervals using the bootstrap. *IEEE Transactions on Signal Processing* 47:3406–3409
- Scheinok PA (1965) Spectral analysis with randomly missed observations: The binomial case. *Annals of Mathematical Statistics* 36:971–977
- Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics* 6:461–464
- Velasco C (1999) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20:87–127
- Velasco C, Robinson PM (2001) Edgeworth expansions for spectral density estimates and studentized sample mean. *Econometric Theory* 17:497–539
- Walden AT (2000) A unified view of multitaper multivariate spectral estimation. *Biometrika* 87:767–788
- White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48:817–838
- Yaglom AM (1987) *Correlation Theory of Stationary and Related Random Functions Volume I: Basic results*. Springer, NY, New York
- Zurbenko IG (1986) *The spectral analysis of time series*. Elsevier Science, Amsterdam

Fig. 1.

A comparison of the bias, SD, and RMSE for the diagonal (gray rectangles) and full (black rectangles) estimators of the LRD parameter δ , for various wavelet filter lengths (the numbers 4, 6, and 8), sample lengths N , and LRD parameter. The rectangles denote 95% bootstrap confidence interval for the summary measure.

Fig. 2.

When $N = 8192$ and $\delta = 0.40$, a comparison of the bias, SD, and RMSE for the diagonal (gray circles) and full (black triangles) estimator versus J . Each row shows a different value for J_0 .

Fig. 3.

(a) A plot of the gappy average sea-ice draft (in meters) versus the distance in kilometers;
 (b) A plot of the gappy log wavelet variances (black symbols), and the log wavelet variances after linear interpolation (gray symbols) versus the log wavelet scale. The vertical lines in each case denote pointwise 95% confidence intervals for each log wavelet variance. The regression line fit is used to estimate δ using the gappy diagonal method.

Fig. 4.
 (a) A map of south central Sweden indicating the 17 locations of the temperature series; (b)
 The light gray regions indicate the time series measurement dates by location. The solid
 black vertical lines denote periods of missingness. The number of days missing for each
 series is listed on the right hand side of the panel.

Fig. 5.

The top panels display the log wavelet variance of the raw daily average temperatures versus the log wavelet scale, at three different locations. The horizontal line indicates the estimates, and the vertical bars give their 95% confidence intervals. The straight line across time scales is based on diagonal estimator of LRD parameter. The bottom panels display the wavelet variances of the deseasonalized and standardized temperature series at the same three locations.

Fig. 6.

(a) By the location ID, the circles denote diagonal estimates of the LRD parameter and the vertical lines denote the associated pointwise 95% CIs for the LRD parameter. Gray denotes locations with missing data, black denotes locations without missing data; (b) The diagonal estimates of the LRD parameter are shown spatially over south-central Sweden. The small numbers denote the location ID.

Comparisons of the bias and estimated standard errors, for the diagonal and full estimators, as a function of the wavelet filter and LRD parameter. The bootstrap SE for the bias is no larger than 0.0008; the bootstrap SE for the SD is no larger than 0.0005; the SE for the averaged ESE is no larger than 3×10^{-5} .

Table 1

	$\delta = 0.25$		$\delta = 0.40$		$\delta = 0.80$	
	$D(4)$	$D(6)$	$D(4)$	$D(6)$	$D(4)$	$D(6)$
Diagonal, Bias	-0.019	-0.024	-0.018	-0.023	-0.019	-0.026
Full, Bias	-0.023	-0.028	-0.023	-0.028	-0.025	-0.032
Diagonal, SD	0.058	0.065	0.058	0.066	0.060	0.068
Diagonal, Avg. ESE	0.044	0.049	0.044	0.049	0.045	0.050
Diagonal, Avg. Sandwich ESE	0.052	0.056	0.052	0.057	0.054	0.059
Full, SD	0.057	0.064	0.057	0.065	0.058	0.067
Full, Avg. ESE	0.050	0.055	0.050	0.055	0.051	0.057

Table 2

Comparisons of the bias and estimated standard errors, for the diagonal and full estimators of the LRD parameter for an FD(0.40) process, under a different number of gap patterns. In each case the D(4) wavelet filter is used with $J_0 = 3$, and $J = 7$. The bootstrap SE for the bias is no larger than 0.0008; the bootstrap SE for the SD is no larger than 0.0005; the SE for the averaged ESE is no larger than 3×10^{-5} .

	Cov	No gaps			Binomial, 10% gaps			Binomial, 20% gaps			Markov		
		$R = 7$	$R = 9$	$R = 7$	$R = 9$	$R = 7$	$R = 9$	$R = 7$	$R = 9$	$R = 7$	$R = 9$	$R = 7$	$R = 9$
Diagonal, Bias	-0.019	-0.017	-0.017	-0.018	-0.018	-0.017	-0.018	-0.017	-0.016	-0.017	-0.016	-0.017	-0.016
Full, Bias	-0.023	-0.020	-0.019	-0.020	-0.019	-0.018	-0.019	-0.018	-0.016	-0.018	-0.017	-0.018	-0.017
Diagonal, SD	0.058	0.062	0.063	0.065	0.065	0.067	0.065	0.067	0.067	0.064	0.064	0.064	0.064
Diagonal, Avg. ESE	0.044	0.035	0.035	0.036	0.037	0.038	0.037	0.038	0.039	0.036	0.037	0.036	0.037
Diagonal, Avg. Sandwich ESE	0.052	0.038	0.038	0.040	0.040	0.042	0.040	0.042	0.042	0.040	0.040	0.040	0.040
Full, SD	0.057	0.088	0.073	0.091	0.076	0.094	0.076	0.094	0.079	0.092	0.076	0.092	0.076
Full, Avg. ESE	0.050	0.027	0.030	0.036	0.037	0.038	0.037	0.038	0.039	0.036	0.037	0.036	0.037

Table 3

Comparisons of the bias, SD, and RMSE for the diagonal and Knight et al. estimators of the LRD parameter δ for an FD(δ) process and a binomial gap pattern with about 10% missing cases, as the LRD parameter is varied. The bootstrap SE for the bias, SD, and RMSE is no larger than 0.001.

	$\delta = 0.25$			$\delta = 0.40$			$\delta = 0.80$		
	Diagonal	Knight, et al.		Diagonal	Knight, et al.		Diagonal	Knight, et al.	
Bias	-0.034	-0.051		-0.050	-0.072		-0.080	-0.103	
SD	0.031	0.026		0.031	0.026		0.033	0.030	
RMSE	0.046	0.058		0.059	0.077		0.086	0.107	