Semi-local scaling exponent estimation with box-penalty constraints and total-variation regularisation

J. D. B. Nelson, C. Nafornita, and A. Isar

***Abstract*—We here establish and exploit the result that 2-D isotropic self-similar fields beget quasi-decorrelated wavelet coefficients and that the resulting localised log sample second moment statistic is asymptotically normal. This leads to the development of a semi-local scaling exponent estimation framework with optimally modified weights. Furthermore, recent interest in penalty methods for least squares problems and generalised Lasso for scaling exponent estimation inspires the simultaneous incorporation of both bounding box constraints and total variation smoothing into an iteratively reweighted least-squares estimator framework. Numerical results on fractional Brownian fields with global and piecewise constant, semi-local Hurst parameters illustrate the benefits of the new estimators.**

I. Introduction

COLLECTIVELY known as scaling processes, both long-range dependent and self-similar processes continue to offer rich and flexible models for a wide variety of signal and image processing problems including recent innovations in texture analysis for super-resolution [1], classification [2], segmentation [3], condition monitoring [4], decluttering [5], denoising [6], [7], [8], and general reconstruction [9].

In broad terms, scaling processes are characterised by the property of scaling behaviour whereby no one single scale is dominant. The most commonly studied subclasses of scaling processes, namely self-similar and long-range-dependent, possess two key features that manifest in the wavelet domain. The first is that the wavelet variance follows a power law with respect to scale; the energy decreases exponentially with finer scale levels. Secondly, the wavelet coefficients are quasi-decorrelated; i.e. the covariance between two wavelet coefficients decreases to zero rapidly with respect to lag and/or differences in scale.

The texture or roughness of a scaling process is determined by the decay, or scaling exponent, of the power law. Owing to the ubiquity of scaling processes, or at least of processes which comprise components that are scaling processes, it is perhaps not surprising that the scaling exponent has been exploited for a wide variety of signal and image processing applications.

Perhaps one of the most effective applications of scaling exponent estimation in image processing has been reconstruction. It is here that the wavelet domain provides an especially appealing setting in that it not only offers a convenient way to estimate scaling exponents but also provides a statistically well-principled means, via shrinkage, to perform denoising and reconstruction. Example applications include velocity measurements in the atmospheric boundary layer, [8]; synthetic aperture radar imagery, [10]; Weierstrass processes [6]; heart rate variability signals [7]; and natural imagery [11].

As Legrand and Lévy-Véhel intuitively explain, denoising can be performed by increasing the scaling exponent at each point [10]. This general idea is formalised into a Bayesian Energy Fraction Estimation approach by Vidakovic et al. [8]. Flandrin et al. used empirical mode decomposition, instead of fixed wavelet bases, to perform denoising of low frequency oscillation embedded in fractional Gaussian noise [7]. Furthermore, by interpreting the oscillatory component as signal, they also constructed a detrending method based on the (Hurst) scaling exponent.

Broadly put, these methods posit that either the signal or noise of interest satisfy a power-law relationship. Wavelet domain shrinkage or rescaling is then used to 'correct' any coefficients that significantly deviate from this prior model. Often, the scaling exponent is known or assumed, as in [8]. However, the work of Echelard and Lévy-Véhel [6], for example, estimates the scaling exponent using coefficients that are relatively uncorrupted by noise and uses this to adaptively correct the corrupted coefficients. This is also an idea that was explored, by Nafornita et al. [12], in the piecewise-constant Hurst scaling parameter and high-frequency noise setting; the authors noted there that improved scaling exponent estimation led directly to improved reconstruction results. This was also recently corroborated and extended by Nelson et al. [13] where a robust estimator of the Hurst parameter was used to subsequently suppress band-limited noise.

Intuitively, the ability of such a reconstruction paradigm is limited by the ability to segment the data into parts that are relatively noisy and those that are relatively noise-free. In turn, this ability is limited to the accuracy of the local scaling exponent estimates. It therefore follows that better scaling exponent estimates will deliver better solutions to tasks such as scaling exponent-based segmentation, reconstruction, and so on. In short, scaling parameter estimation is a relevant research goal in its own right.

Statistically rigorous wavelet-based approaches towards scaling parameter estimation were largely formalised by a series of research efforts by Abry, Veitch, and others, at the turn of the millennium, for processes on the real line [14],

J. D. B. Nelson is with the Department of Statistical Science at University College London and was partially supported by grants from Innovate UK and the Dstl

C. Nafornita and A. Isar are with the Communications Department at Politehnica University Timisoara

[15], [16]. That the scaling parameter can be estimated so conveniently and robustly via wavelets, and with such weak assumptions, is largely testament to a fundamental property of wavelet bases, namely their scale invariance— each wavelet member is merely a dilated (and translated) version of any other member.

We here revisit some of this earlier development and extend a key result to processes on $\mathbb{R}^2$, namely that (isotropic) self-similar processes (fields) give rise to quasi-decorrelated wavelet coefficients. Following and adapting Abry et al., this example leads us to propose an estimator for the semi-local scaling exponent parameter which, in turn, suggests a well-principled, modified, locally-weighted least-squares regression approach.

Since the power law and the quasi-decorrelation properties lead to a regressand which is asymptotically normal and since it might be natural for some problems such as texture segmentation, for example, to assume that the scaling exponent rate of change is limited, it is therefore tempting to extend the resulting weighted-least squares estimator by adding appropriate penalty terms to the (ℓ_2) log-likelihood and utilising some very recent developments in statistical regularisation. This is a highly active area of modern signal processing, computational statistics, and machine learning, which continues to exploit and develop a range of algorithms, many of which are inspired from the convex optimisation work of Boyd and Vandenberghe [17] and others, that perform estimation subject to prior knowledge, expressed in the form of constraints.

Against this backdrop, Nafornita et al. [12] and Pustelnik et al. [3] both very recently explored the use of total variation to spatially regularise the scaling exponent estimate. Inclusion of this penalty term is equivalent to assuming that the spatial differences of the scaling exponent are distributed as a Laplacian. It is therefore well-suited for estimation of piecewise-constant varying scaling exponents— a setting that has received relatively scant attention, especially for processes defined over $\mathbb{R}^2$.

Following this regularisation theme, we here introduce box-constraints into the estimation framework to reflect the fact that the scaling exponent is confined to a prescribed interval. As is especially evident for antipersistent processes with small scaling exponents, we will show with experiments that this furnishes more robust estimates.

Furthermore that the new developments are accommodated so readily into the regularisation framework leads us quite naturally to incorporate them into the generalised Lasso, total variation estimator devised by Nafornita et al. [12]. Our resulting approach, therefore, is able to permit any combinations of the proposed optimally-weighted semi-local estimator, the total-variation regulariser, and the new box-constraints. Remarkably, a very simple algorithm, based on the iteratively reweighted least-squares iterations is all that is required to perform the estimation for any such combination thereof.

Related work

With the advent of the multifractal formalism, focus in scaling processes on $\mathbb{R}^2$ has shifted away from multifractional models, i.e. the pointwise measurement of regularity, and instead towards a description of how regularity varies stochastically over space [18], [19], [20]. Whilst the multifractal formalism is a rich, fruitful, and current area of research that has led to many interesting results and useful applications, activity there can be contrasted with the relative apparent neglect of processes which have properties somewhere in-between the extremes of globally-fixed, and stochastically pointwise-varying, scaling exponents. Much of the interest in multifractional models [21], where the scaling parameter is allowed to vary smoothly or deterministically, is largely restricted to financial time-series models defined over the real line, such as in [22].

As such, there have only been a few efforts to estimate a spatially, smoothly varying scaling parameter. Extending Nelson and Kingsbury's local Hurst estimator [23], Nafornita et al. proposed a total variation estimator which permitted a generalised Lasso formulation [12]. Very recently, Nelson et al. extended this to a robust framework to accommodate heavy-tailed processes [13]. Pustelnik et al. used a local version of wavelet leaders— a maxima of wavelet coefficients across a local spatio-scale cone— together with a total variation penalty, specifically to perform texture segmentation [3], [24], [25]. Their (Hölder) scaling exponent estimation framework is based on the forward-backward, primal-dual algorithm. Instead of using local weights to accommodate the manner in which the variance increases over the larger scale levels, as we propose here, their algorithm attaches unknown weights to all the space-scale wavelet (leader) coefficients and optimises these jointly together with the local scaling exponent estimate. Furthermore, Pustelnik at al. do not constrain the scaling exponent parameter as we propose here. Also worth noting here is that piecewise constant Hurst estimation was very recently generalised by Regli and Nelson, via a Markov random field based approach, to realise a piecewise parameterised estimator [26].

This paper is structured as follows. We review some of the wavelet analysis of scaling processes in Section II and extend the result, that (isotropic) self-similar processes (fields) give rise to quasi-decorrelated wavelet coefficients, from the reals to $\mathbb{R}^2$. We propose a semi-local and optimally weighted scaling exponent in Section III. These estimators are embedded into a new box-constrained, total variation algorithmic framework in Section IV. Numerical experiments are performed in Section V for the case of self-similar processes which confirm the utility of the semi-local weighted estimator and the box-constraints. A summary is offered in Section VI where a number of ideas for further work are also outlined.

II. Scaling exponent estimation

Abry et al. [14] show that the spectral slope of self-similar processes, long-range dependent processes, second-order stationary $1/f$ processes, and fractal processes can all be treated and estimated in broadly equivalent ways. In particular, all of these so-termed scaling processes have second moments that, at least asymptotically, follow a simple power law. They also all have pseudo-decorrelated wavelet coefficients. Placing

the development of Abry et al. [14] into the local scaling exponent setting, we discuss below how these two properties— power law and decorrelation— naturally lend themselves to a simple least-squares estimation framework.

If a stochastic field $X\colon \mathcal{T} \mapsto \mathbb{R}$, defined over some spatial support $\mathcal{T} \subset \mathbb{R}^2$, has a second moment which follows a power law then

$$\mathbb{E}\big|(\mathcal{W}X)(\cdot\,; j, \theta)\big|^2 \propto 2^{2j(H+1)}, \tag{1}$$

where $\mathcal{W}$ is the wavelet operator defined by

$$(\mathcal{W}X)(t; j, \theta) = 2^{-j}\big\langle X, \psi_\theta(2^{-j}\cdot - t)\big\rangle_{L_2(\mathcal{T})},$$

with wavelet ψ defined over space t, orientation θ, and jth finest scale level. Sometimes, as is the case of long-range dependent processes for example, the proportionality in (1) is replaced with the weaker asymptotic equivalence but, in that case, the ensuing analysis is largely the same. Without loss of generality we refer to the H as the scaling exponent throughout this work. It determines the power-law rate of decay of the (wavelet) spectrum as the scale level becomes finer and finer.

For critically decimated wavelet coefficients, the transform is computed over $j \in \mathbb{Z}, k = (k_0, k_1) \in \mathbb{Z}^2$ via the fast wavelet pyramidal algorithm. For an image of size $\sqrt{n_0} \times \sqrt{n_0}$, this results in $n_j = 2^{-2j} n_0$ many wavelet detail coefficients, per directional subband, at the jth finest level, namely:

$$d_j[k] := 2^{-j}\langle X, \psi(2^{-j}\cdot - k)\rangle_{L_2(\mathcal{T})}, \quad k = 1, \ldots, n_j,$$

where we now drop the orientation parameter θ for convenience— although anisotropic estimation is possible (e.g. [23]), the analysis is broadly the same as the isotropic case but merely performed on a directional sub-band-by-sub-band basis. To estimate the scaling parameter H, it is therefore natural to consider the sample second moment of the wavelet coefficients:

$$\mu_j := \frac{1}{n_j}\sum_{k=1}^{n_j} |d_j[k]|^2 = \frac{1}{n_j}\big\|d_j[\cdot]\big\|^2_{\ell_2(\{1:n_j\})}, \tag{2}$$

as an approximation to the expectation in Equation (1). For then, H can be estimated by regressing $y_j = \log_2 \mu_j$ on j. To help fix some notation the explicit problem can be written as

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}_0\boldsymbol{\beta}\|_2^2, \tag{3}$$

where $\mathbf{y} = \big((y_j)_{j=j_-}^{j_+}\big)^\top \in \mathbb{R}^{J_0}$, $\boldsymbol{\beta} \in \mathbb{R}^2$, and

$$\mathbf{X}_0^\top = \begin{bmatrix} 1 & \cdots & 1 \\ j_- & \cdots & j_+ \end{bmatrix} \in \mathbb{R}^{2\times J_0}.$$

The fact that we can legitimately regress $\log_2 \mu_j$ onto j is not altogether trivial. We note that $\mathbb{E}\mu_j \propto 2^{2j(H+1)}$. However, in general $\mathbb{E}\log_2 \mu_j \neq \log_2 \mathbb{E}\mu_j = 2j(H+1) + C$. Instead, we appeal to the second property, namely that the wavelet coefficients of scaling processes are quasi-decorrelated. This is a well-reported result for 1-D processes. A similar result also holds in the isotropic 2-D case. To serve as a motivating example [14], it is instructive to consider self-similar processes. These will subsequently be studied further with numerical experiments in Section V. To this end, we note the following definition and result.

Example 2.1: A stochastic field $X\colon \mathcal{T} \mapsto \mathbb{R}$, with finite variance and spatial support $\mathcal{T} \subset \mathbb{R}^2$, is weak statistically self-similar, or H-ss, if $\mathbb{E}X(\alpha\cdot) = \alpha^H \mathbb{E}X$ and $\mathbb{E}X(\alpha t)X(\alpha\cdot) = \alpha^{2H}\mathbb{E}X(t)X(\cdot)$, for $H, \alpha > 0$. In addition, if X also has stationary increments then it is called H-sssi and the scaling parameter $0 < H < 1$ is, in this case, also termed the Hurst parameter or exponent.

It is well-known and easy to show that self-similar processes satisfy the power-law property described by Equation (1). The following result establishes the second property of interest, namely pseudo-decorrelation.

Theorem 2.2: Let the compactly supported wavelet $\psi\colon \mathbb{R}^2 \mapsto \mathbb{R}$ possess M-many vanishing moments. Then the wavelet coefficients $d_j[k] = \langle B_H, \psi_{jk}\rangle$ of an isotropic H-sssi field B_H, defined over $\mathbb{R}^2$ with Hurst parameter H, satisfy the pseudo-decorrelation property, namely:

$$\mathbb{E}d_j[k]d_j[k'] = O\Big(\big\|2^j(k-k')\big\|^{2(H-M)}\Big). \tag{4}$$

The 1-D version of this result (i.e. for B_H defined over the reals) was established by, for example, Tewfik [27]. We present the 2-D version of the proof in Appendix A along with complementary results and note that this can be extended, quite readily, to higher dimensions. A similar result also holds for long-range dependent processes and fields as well as scaling processes/fields in general. It turns out that this result helps simplify a lot of the statistical analysis necessary to characterise the scaling exponent wavelet estimator. For Gaussian scaling processes, the wavelet coefficients are Gaussian and the quasi-decorrelation results in quasi-independence. As a result, we can venture the claim that μ_j is approximately Chi-squared. To make this more precise, we follow a similar argument to Abry et al. [14] and note that under the slightly stronger condition of decorrelation, together with the Gaussian assumption, we have that

$$\frac{n_j}{\mathbb{E}\mu_j}\mu_j \overset{\mathrm{d}}{\sim} \chi^2(n_j), \tag{5}$$

the Chi-squared distribution with n_j degrees of freedom. Hence

$$\begin{aligned} \log_2 \mu_j &\overset{\mathrm{d}}{\sim} \log_2 \mathbb{E}\mu_j - \log_2 n_j + \log_2 \chi^2(n_j) \\ &\overset{\mathrm{d}}{\sim} 2j(H+1) + \log_2 C - \log_2 n_j + \frac{\ln \chi^2(n_j)}{\ln 2}, \end{aligned}$$

and it follows (see Abry et al. [14]) that

$$\begin{aligned} \mathbb{E}\log_2 \mu_j &= 2j(H+1) + \log_2 C + g[j] \\ \operatorname{var}\log_2 \mu_j &= \frac{\zeta(2, n_j/2)}{\ln^2 2}, \end{aligned}$$

with $g[j] = \Psi(n_j/2)/\ln 2 - \log_2(n_j/2)$ and where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the Psi function and $\zeta(z,\nu)$ is the generalised Riemann Zeta function. Abry et al. [14] also note the asymptotic behaviour $g[j] \sim -(n_j \ln 2)^{-1}$ and $\operatorname{var}\log_2 \mu_j \sim 2(n_j \ln^2 2)^{-1}$, and Abry et al. [28] show that, asymptotically

$$\log_2 \mu_j \overset{\mathrm{d}}{\sim} \mathcal{N}\left(2j(H+1) + \log_2 C, \frac{2}{n_j \ln^2 2}\right), \tag{6}$$

which suggests, in the ideal case, a weighted least-squares regression of $\log_2 \mu_j$ onto j, namely

$$\min_{\boldsymbol{\beta}} \left\| \mathbf{W}_0^{1/2} \left(\mathbf{y} - \mathbf{X}_0 \boldsymbol{\beta} \right) \right\|_2^2, \tag{7}$$

with $\mathbf{W}_0 = \mathrm{diag}(w_j)_{j_-}^{j_+}$ and $w_j = 2^{-2j}$. In fact, as discussed in [14], [15], [16], the Gaussianity of $\log_2 \mu_j$ holds approximately quite well even for small data size and mild departures from decorrelation and distributional (Gaussian) assumptions. As such, Theorem 2.2 provides principled statistical justification to introduce a (weighted) least-squares approach to scaling exponent estimation of scaling processes over $\mathbb{R}^2$.

III. Processes with a piecewise-constant varying scaling exponent

That the asymptotic result (6) holds approximately, even for small sample size, suggests a more local analysis may be feasible. Indeed, this lends some theoretical strength to least-squares approaches that estimate scaling exponents as a piecewise constant [12], [13], [23], or piecewise parameterised [26], function. In the below, we first construct a semi-local scaling exponent estimator in Section III-A and then adapt and place this into an optimally weighted-least squares scheme in III-B.

A. Semi-local scaling exponent estimation

Localised estimation of the scaling exponent parameter requires a local estimate of the second moment in (1) about a point, t_i, say. Some previous work [23], [12] has reported some success by merely interpolating the decimated wavelet energies at each scale level up to the zeroth level. One then arrives at a point estimate of the second moment about location $t_i \in \mathcal{T}$ at the jth finest scale level, namely $\mu_j[i]$. Setting $y_j[i] = \log_2 \mu_j[i]$, the resulting regression problem then becomes a set of equations

$$y_j[i] = \beta_1[i] + j\beta_2[i] + \epsilon[i], \tag{8}$$

with $\beta_2[i] = 2(H_i + 1)$, and where $H_i = H(t_i)$ is the local scaling exponent which can now vary with location t_i. The (global) least-squares problem described by (3) now becomes a set of local least-squares problems, viz. $\min_{\boldsymbol{\beta}_i} \|\mathbf{y}_i - \mathbf{X}_0 \boldsymbol{\beta}_i\|_2^2$, with $\mathbf{y}_i = \left((y_j[i])_{j=j_-}^{j_+} \right)^\top \in \mathbb{R}^{J_0}$, $\boldsymbol{\beta}_i = (\beta_1[i], \beta_2[i])^\top \in \mathbb{R}^2$, and where $\mathbf{X}_0$ is as defined earlier. It will be convenient, for some of the extensions that follow, to write these equations in a single problem $\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$, where $\mathbf{X}$ is defined, and $\mathbf{y}$ and $\boldsymbol{\beta}$ are redefined, as follows

$$\begin{aligned} \mathbf{y} &= (\mathbf{y}_1^\top, \ldots, \mathbf{y}_{n_0}^\top)^\top \in \mathbb{R}^n, \\ \boldsymbol{\beta} &= (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_{n_0}^\top)^\top \in \mathbb{R}^p, \\ \mathbf{X} &= \mathbf{I}_{n_0} \otimes \mathbf{X}_0 \in \mathbb{R}^{n \times p}, \end{aligned}$$

with the $n_0 \times n_0$ identity $\mathbf{I}_{n_0}$, the Kronecker product $\otimes$, $n = J_0 n_0$, and $p = 2n_0$.

Performing a point estimate of the second moment at the zeroth level domain $\mathcal{T} = \{t_i\}$ is not necessarily optimal. Ideally, the second moment should be estimated by performing summations over wavelet coefficients which lie in those regions where the scaling exponent is constant. Unfortunately, these regions are not known a priori. Herein lies a trade-off. If the sizes of region over which the sample second moment is computed is too large then the estimator will fail to adequately capture the locally varying nature of the scaling exponent. In this case, coefficients sampled from regions with other scaling exponent values will contaminate the samples used in the second moment estimate. On the other hand, if the region is too small, then one may not be making full use of the data at hand; instead a suboptimal estimate of the second moment will be used. In this case, relevant coefficients will not be included in the summation. Clearly, the trade-off will largely depend on how quickly the scaling parameter varies over space. As such, the choice of region size is a parameter setting that must be decided or optimised based on some a priori knowledge or optimisation criterion.

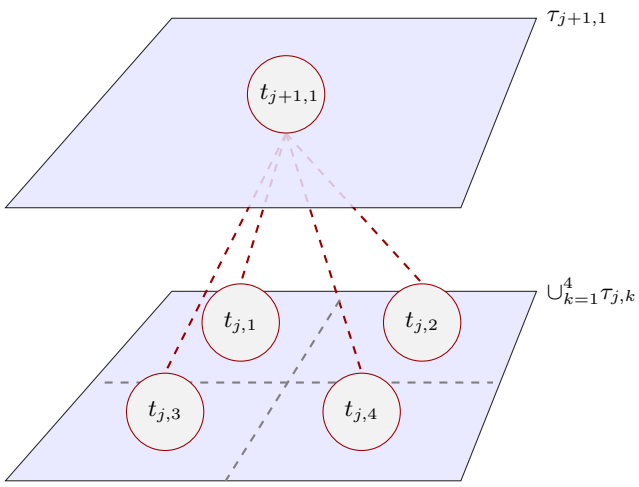


Fig. 1. The quad-tree structure of dyadically sampled wavelet coefficients over the two-dimensional lattice. The coefficient $t_{j+1,1}$ at the $(j+1)$th finest scale level is centred at a point in space which is equidistant from the four neighbouring 'children' coefficient locations $\{t_{j,i}\}_{i=1}^4$, say, at the jth finest scale level. Likewise the region $\tau_{j+1,1}$ centred on $t_{j+1,1}$ with width 2^{j+1} covers the same spatial support as the union of the four regions $\{\tau_{j,k}\}_{k=1}^4$, say, with widths 2^j, centred on the points $\{t_{j,i}\}_{i=1}^4$.

When the summation region size is greater than a single pixel, the numbers of decimated wavelet coefficients involved in the estimate will vary with respect to scale level. Similar to the global estimation case, the variance at each scale level will vary according to the numbers of coefficients summed. It is therefore important to understand how this relationship works in the semi-localised case where the moment is computed over a region with size somewhere in-between that of a pixel, at the one extreme, and the entire spatial support, on the other. To proceed, it is necessary to define what is meant by the magnitude of a decimated wavelet coefficient about an arbitrary location t_i in the spatial support $\mathcal{T}$.

The centre of mass of the wavelet at the $(j+1)$th finest scale level is centred at a point equidistant from four neighbours at the jth finest level. This quad-tree structure cascades down through all the scale levels. Let the set $\{t_{ji}\}_i \subset \mathcal{T}$ denote the spatial locations of the wavelet coefficients at the jth finest level. Define $\tau_{j_0 i} \subset \mathcal{T}$ to be the square region, centred on the point $t_{j_0 i}$, and without loss of generality, with width 2^{j_0}, cf.

Fig 1.

To choose a region of size, say 2^{j_0}, over which to form a semi-local estimate of the sample second moment, the wavelet energies at scales coarser than j_0, i.e. for $j > j_0$, are simply upsampled (interpolated) to the points $\{t_{j_0 i}\}_i$. On the other hand, at scale levels finer than j_0 (namely $j < j_0$) the wavelet energies are averaged over the region $\tau_{j_0 i}$, i.e. $|d_j[i]|^2$ are averaged over the indexes $\kappa_{ji}^{j_0} := \{k\colon t_{jk} \in \tau_{j_0 i}\}$. In effect, all the coarser levels, greater than j_0, are upsampled to $\{t_{j_0 i}\}_i$ and all the finer levels, smaller than j_0, are 'downsampled' via summation to $\{t_{j_0 i}\}_i$. This leads to the semi-local sample second moment

$$\mu_j[i] = \begin{cases} \left(2^{(j-j_0)} \uparrow |d_j|^2\right)[i], & j \geq j_0 \\ \left|\kappa_{ji}^{j_0}\right|^{-1} \left\|d_j[\cdot]\right\|^2_{\ell_2(\kappa_{ji}^{j_0})}, & j < j_0 \end{cases} \tag{9}$$

with $\kappa_{ji}^{j_0} := \left\{k\colon t_{jk} \in \tau_{j_0 i}\right\}$ and where $a \uparrow x$ symbolises the operation of interpolating the sequence x by a factor of a. This could, for example, be via nearest neighbour or by bilinear interpolation.

As an example, one could consider the spatial support $\mathcal{T} = (0, 2^J)^2$. If this is sampled uniformly then the zeroth level over which the process is observed has, as its discrete domain, the 2-d lattice $\{t_{0i}\}_i = (k_1 - 1/2, k_2 - 1/2)_{k_1,k_2=1}^{2^J}$. The wavelet coefficients at the jth finest scale level will then be centred at the points $\{t_{ji}\}_i := \left(2^{j-1}(2k_1 - 1), 2^{j-1}(2k_2 - 1)\right)_{k_1,k_2=1}^{2^{J-j}}$ and the region centred on t_{ji} with width 2^j is

$$\tau_{ji} = \left[2^j(k_1 - 1), 2^j k_1\right) \times \left[2^j(k_2 - 1), 2^j k_2\right).$$

B. Locally weighted scaling exponent parameter estimation

When downsampling by a factor of $s = 2^{j_0}$, say, there will be a factor of $2^{2(j_0 - j)}$ many more coefficients involved in the estimate of the sample second moment at the jth finest level (when $j < j_0$) than the j_0th finest level. On the other hand, the jth finest scale levels for $j > j_0$ have just as many coefficients as the j_0th finest level, courtesy of the upsampling. Therefore, we propose that weights of $2^{2(j_0 - j)}$ be attached to the finer scale levels $j < j_0$ whereas weights of 1 are used for all the coarser scale levels $j \geqslant j_0$.

Furthermore, from Equation (9) it can be seen for scales finer than the j_0th finest level (i.e. for $j < j_0$) the μ_j statistic is derived from averaging over $2^{2(j_0 - j)}$ wavelet coefficients. Hence, using exactly the same arguments as were applied in the global case, we have that μ_j is Chi-squared, cf. Equation (5), only this time it is of order $2^{2(j_0 - j)}$ rather than order $N2^{-2j}$. On the other hand, for the coarser scale levels ($j \geqslant j_0$) only one (interpolated) coefficient is used— i.e. a point-estimate. If nearest neighbour interpolation is used then this point-estimate is distributed as Chi-squared, order one. This results in a slight modification of Equation (6), namely

$$\log_2 \mu_j \overset{\mathrm{d}}{\sim} \begin{cases} \mathcal{N}\left(2j(H+1) + \log_2 C, \dfrac{2s^2}{n_j \ln^2 2}\right), & j < j_0 \\ \mathcal{N}\left(2j(H+1) + \log_2 C, \dfrac{2}{\ln^2 2}\right), & j \geqslant j_0 \end{cases}$$

Hence, for a downsampling factor of s, we propose the optimal semi-local weighted-least squares (WLS) approach

$$\min_{\boldsymbol{\beta}} \left\| \mathbf{W}^{1/2} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) \right\|_2^2, \tag{10}$$

where, now

$$\begin{aligned} \mathbf{y} &= (\mathbf{y}_1^\top, \ldots, \mathbf{y}_{n_{j_0}}^\top)^\top \in \mathbb{R}^n, \\ \boldsymbol{\beta} &= (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_{n_{j_0}}^\top)^\top \in \mathbb{R}^p, \\ \mathbf{X} &= \mathbf{I}_{n_{j_0}} \otimes \mathbf{X}_0 \in \mathbb{R}^{n \times p}, \\ \mathbf{W} &= \mathbf{I}_{n_{j_0}} \otimes \mathbf{W}_0 \in \mathbb{R}^{n \times p}, \end{aligned}$$

with a slight abuse of the notation $n = n(s) = J_0 n_0 / s^2$ and $p = p(s) = 2n_0/s^2$; and with the $n_0 \times n_0$ identity $\mathbf{I}_{n_0}$, the Kronecker product $\otimes$, and weights $\mathbf{W}_0 = \mathrm{diag}(w_j)_{j_-}^{j_+}$, where

$$w_j = \max\left(1, s^2 2^{-2j}\right). \tag{11}$$

C. Simulation

The next section presents some estimation algorithms based on the preceding analysis. To study the efficacy, fractional Brownian fields are simulated according to the incremental Fourier synthesis method first proposed by Kaplan and Kuo [29]. Their method, which is also a key part of the Fraclab toolbox [30], provides a good balance between accuracy and computational complexity. This is important since, although on the one hand it is of course important to reproduce some of the key statistical properties of fractional Brownian fields, there is also often the requirement to perform many trials of many experiments over fields of various sizes, etc.

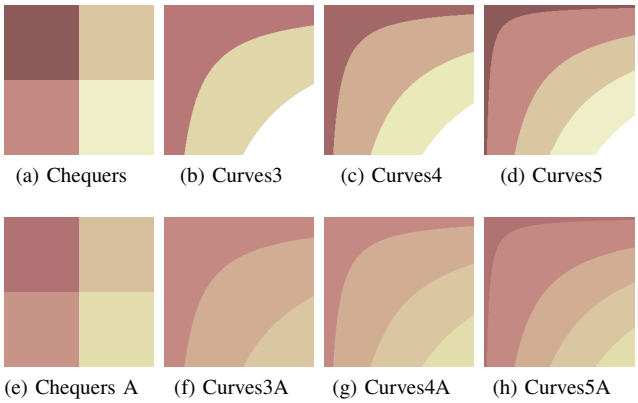
The incremental approach is based on the standard Fourier synthesis approach which, as the name suggests, simulates a field in the Fourier domain and then inverts it to arrive at B_H. The method proceeds by simply scaling a white noise process $\varepsilon(\cdot) \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, by the (generalised) power spectral density of a fractional Brownian field with Hurst parameter H, namely $\rho(\cdot) := \|\cdot\|^{-2H-2} \varepsilon(\cdot)$. The phase θ is drawn from the uniform distribution $\theta(\cdot) \overset{\text{iid}}{\sim} \text{Uniform}[0, 2\pi)$, to give the simulated spectrum $B_H^\wedge = \rho e^{i\theta}$. Finally, the fractional Brownian field is rendered via $B_H = \Re((\rho e^{i\theta})^\vee)$.

The key drawbacks to the standard synthesis method is that it results in a stationary process and fails to demonstrate the self-similarity property. Although a similar process is followed by the incremental Fourier synthesis approach, this time, the (stationary) first and second order increments of B_H are simulated first via the Fourier method. These are then carefully summed to produce the (non-stationary) fractional Brownian field. As Kaplan and Kuo illustrate, the incremental approach reproduces the self-similarity property more faithfully than the standard Fourier synthesis method with only double the computational cost [29].

Fields with piecewise, locally varying Hurst parameter can be constructed as follows. Consider a disjoint covering of the domain $\mathcal{T}$, namely $\bigcup_k \mathcal{T}_k = \mathcal{T}$, and $\mathcal{T}_k \cap \mathcal{T}_\ell = \emptyset$, for $k \neq \ell$. Construct the masks $S_k(t) := \mathbb{I}(t \in \mathcal{T}_k)$, with the indicator function $\mathbb{I}\colon \mathcal{T} \mapsto \{0,1\}$ and let $\{\eta(k)\}_k$ determine the codomain of the Hurst parameter $H = H(t), t \in \mathcal{T}$; i.e. the finite set of unique Hurst parameters present in the field.

TABLE I
CODOMAIN OF HURST VALUES IN SIMULATED DATA

Name	Codomain
chequers	$\{1/5, 2/5, 3/5, 4/5\}$
chequers A	$\{3/10, 13/30, 17/30, 7/10\}$
curves3	$\{1/3, 2/3, 1\}$
curves3A	$\{2/5, 1/2, 3/5\}$
curves4	$\{1/4, 1/2, 3/4, 1\}$
curves4A	$\{2/5, 1/2, 3/5, 7/10\}$
curves5	$\{1/5, 2/5, 3/5, 4/5, 1\}$
curves5A	$\{3/10, 2/5, 1/2, 3/5, 7/10\}$



(a) Chequers (b) Curves3 (c) Curves4 (d) Curves5

(e) Chequers A (f) Curves3A (g) Curves4A (h) Curves5A

Fig. 2. Piecewise-varying Hurst parameter for the chequers and curves simulated data.

First of all the set of surfaces $B_{\eta(k)} = B_{\eta(k)}(\cdot; \varepsilon_0)$ are simulated with the incremental Fourier synthesis method, using the same white noise process ε_0 for each surface. Then, a fractional Brownian field with locally varying H can be constructed with $B_H = \sum_k S_k B_{\eta(k)}(\cdot; \varepsilon_0)$. The resulting field will have Hurst parameter $\eta(k)$ at $t \in \mathcal{T}_k$.

The piecewise varying Hurst parameter function $H = H(t)$ of some of the simulated data used in the experiments in Section V is plotted in Figure 2. The codomains are tabulated in Table I. 'Chequers' simply comprises four smaller square regions where the Hurst parameter takes the values $1/5, 2/5, 3/5, 4/5$, top to bottom, left to right, starting in the top left corner. It therefore contains regions of both small transitions, between $1/5$ and $2/5$, say, as well as large jumps—for example between $1/5$ and $3/5$. The curves data have transitions along various curved edges. Note here that the alternative ('A') versions in the bottom panels of Figure 2 have smoother varying H. An instance of the 'Curves5A' data is depicted in Figure 3. Note how the local texture is rougher in the upper-left corner and becomes evermore smoother towards the lower-right corner as the Hurst parameter increases.

IV. ESTIMATION ALGORITHMS

We here extend the idea of exploiting prior information regarding: (i) how the scaling parameter varies over the spatial support and (ii) the fact that the scaling exponent often lies on some bounded, or semi-bounded interval, which is known apriori. Self-similar processes offer an example of a scaling exponent which lies on the unit interval.

More generally, the practitioner may have some prior knowledge about what interval(s) the scaling exponent may lie.

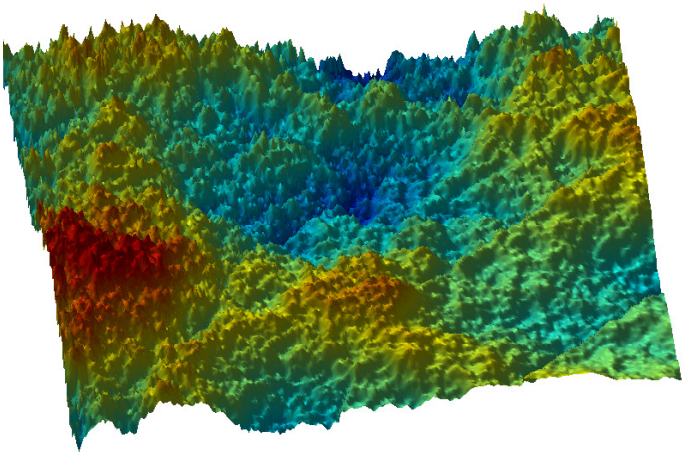


Fig. 3. An instantiation of the Fractional Brownian field 'Curves5A'.

This could come from training data, physical models, or other intuition about the data. Even quite conservative constraints on where the scaling exponent lies would offer extra information that one should ideally exploit.

We will see that it is possible to accommodate both the spatial regularisation and scaling exponent constraints with the flexible computational framework provided by the iteratively reweighted least-squares (IRLS) scheme.

Although not new, IRLS is a very flexible and extensible scheme has attracted some very recent attention. Most notable is that of Daubechies et al. [31] whose modifications ensure algorithm convergence. This is significant because the IRLS has exponentially fast convergence. Although each step in the algorithm can be expensive owing to a matrix inversion, Chen et al. recently formulated a preconditioner which greatly reduces this cost [32], [33] and Zhou et al. recently proposed the combination of IRLS with the popular alternating direction method of multipliers to avoid direct inversion [34].

A. Total variation

The semi-local weighted scheme in the previous section explicitly exploited the assumed piecewise constant behaviour of the Hurst/scaling parameter. Another means to incorporate this prior knowledge is via statistical regularisation. To this end Nafornita et al. [12], proposed the total-variation regularised estimator

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{\nabla}\boldsymbol{\beta} \right\|_1, \tag{12}$$

where $\boldsymbol{\nabla} \in \mathbb{R}^{m \times p}$ performs a spatial regularisation on $\boldsymbol{\beta}$ (and therefore on H). The parameter λ controls the amount of spatial smoothness that is assumed in the solution. The probabilistic interpretation is that a negative-log-Laplacian prior is applied to the spatial differences and added to the negative-log-likelihood to establish the negative-log-posterior. In effect, Expression (12) describes a maximum a posteriori solution. Furthermore, Nafornita et al. [12] showed that this problem was an instance of a generalised Lasso, first considered by She [35]. Although She used a simulated annealing-based approach, the first path algorithm was found by Tibshirani and Taylor [36]. As the name suggests, this recovers all solutions

$\boldsymbol{\beta}(\lambda)$ over the values of λ. Recently, Nelson et al. [13] established an iteratively reweighted least-squares algorithm to solve (12). Their development proceeds by firstly noting that the usual Lasso objective can be rewritten as a weighted ridge regression one, namely:

$$\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_2^2+\lambda\left\|\boldsymbol{\beta}\right\|_1=\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_2^2+\lambda\left\|\boldsymbol{\Omega}^{1/2}\boldsymbol{\beta}\right\|_2^2,$$

where $\boldsymbol{\Omega}=\operatorname{diag}\left(|\beta_i|^{-1}\right)_{i=1}^{p}\in\mathbb{R}^{p\times p}$. This then suggests the following iterations which simply alternate between solving a weighted ridge problem and updating the weights:

$$\begin{aligned}\boldsymbol{\beta}^{(\ell+1)} &= (\mathbf{X}^\top\mathbf{X}+\lambda\boldsymbol{\Omega}^{(\ell)})^{-1}\mathbf{X}^\top\mathbf{y},\\ \boldsymbol{\Omega}^{(\ell+1)} &= \operatorname{diag}\left(|\beta_i^{(\ell+1)}+\delta|^{-1}\right)_{i=1}^{p},\end{aligned}\qquad(13)$$

where δ can be a small, fixed constant or updated as in [31].

However, as Nelson et al. [13] show, a very similar algorithm can also be realised for the generalised Lasso. In particular, we note that the generalised penalty in Equation (12) can be written as $\|\boldsymbol{\nabla}\boldsymbol{\beta}\|_1=\sum_i\left|\boldsymbol{\nabla}_i^\top\boldsymbol{\beta}\right|=\sum_i\left|\boldsymbol{\nabla}_i^\top\boldsymbol{\beta}\right|^{-1}(\boldsymbol{\nabla}_i^\top\boldsymbol{\beta})^2$, where $\boldsymbol{\nabla}_i\in\mathbb{R}^{p\times 1}$ such that $\boldsymbol{\nabla}=(\boldsymbol{\nabla}_1^\top,\ldots,\boldsymbol{\nabla}_m^\top)^\top$. The generalised Lasso objective can then be formulated in terms of a weighted ridge regression, viz.

$$\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_2^2+\lambda\left\|\boldsymbol{\nabla}\boldsymbol{\beta}\right\|_1=\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_2^2+\lambda\left\|\boldsymbol{\Omega}^{1/2}\boldsymbol{\nabla}\boldsymbol{\beta}\right\|_2^2,\quad(14)$$

where, now, $\boldsymbol{\Omega}=\operatorname{diag}\left(|\boldsymbol{\nabla}_i^\top\boldsymbol{\beta}|^{-1}\right)_{i=1}^{m}\in\mathbb{R}^{m\times m}$. In this way, their development leads to Algorithm 1 (for the case $\mathbf{W}=\mathbf{I}$). We note the generalisation $\mathbf{W}\neq\mathbf{I}$ is now possible as a consequence of the downsampled, weighted estimator presented in Section III-B. Again, this alternates between a ridge problem and an update. In practice, the algorithm is stopped if it either reaches some maximum number of iterations or if the distance between two successive estimates is smaller than some predefined threshold. It should be noted that, as is typical with IRLS-based algorithms, although convergence is reached very quickly (potentially exponentially fast [31] compared to the $O(1/k^2)$ complexity of gradient descent methods such as FISTA [37]) computational issues can arise when the data size, n_0 is large. For then, the inversion of the $p\times p$ matrix, required when estimating $\boldsymbol{\beta}^{\ell+1}$, can be problematic even when Cholesky factorisation and/or sparse matrix structures are exploited. Fortunately, one approach advocated here is the use of the downsampling factor. It is perfectly reasonable to combine this with the total-variation regularisation. Recall that downsampling by a factor of $s=2^{j_0}$ has the effect of shrinking $p=p(s)=2n_0/s^2$ by a factor of $s^2=2^{2j_0}$. Therefore, even a modest amount of downsampling, such as $s=4$, say, can reduce the problem size significantly.

Unlike many other methods, IRLS can be quite naturally extended to overlapping mixed, grouped-sparse, norms [32]. Indeed, very recent fast preconditioning IRLS-based methods have emerged [32], [33] which compare favourably to gradient descent methods and a variety of other state-of-the-art approaches. Furthermore, it should also be noted that, unlike ADMM, FISTA, and other convex optimisation approaches, IRLS can easily accommodate non-convex terms that, for example, could result from positing a hyper-Laplacian prior $\|\nabla\cdot\|_p$, with $0<p\leq 1$ on the spatial differences to induce a greater amount of sparsity than is available to the ℓ_1 penalty [31]. Moreover, it can also easily incorporate non-smooth fidelity terms which give rise to heavier-tailed likelihoods and, hence, can offer more robustness to non-Gaussian, noise [13].

In contrast to Tibshirani's generalised Lasso algorithm [36], IRLS (like other non-path algorithms such as FISTA [37] etc) does not yield the full solution path— for all λ. In practice, however, it is common and considered reasonable to select a suitable λ and solve over the $\boldsymbol{\beta}$ coefficients. The λ can be chosen with respect to some theoretical criterion such as AIC/BIC, etc, or alternatively is 'learnt' via cross-validation or otherwise. Further algorithmic considerations and comparisons to ADMM, gradient descent, and extensions thereof is beyond the scope of the paper and is left as further work.

Algorithm 1 Generalised Lasso (or total variation least squares) via iteratively reweighted least squares

Inputs:
design matrix $\mathbf{X}$, wavelet log-variances $\mathbf{y}$,
regularisation parameter λ
Initialize:
$\boldsymbol{\Omega}^{(0)}=\mathbf{I}$
for $\ell=0,1,\ldots,\ell_{\max}$ **do**
$\boldsymbol{\beta}^{(\ell+1)}=(\mathbf{X}^\top\mathbf{W}\mathbf{X}+\lambda\boldsymbol{\nabla}^\top\boldsymbol{\Omega}^{(\ell)}\boldsymbol{\nabla})^{-1}\mathbf{X}^\top\mathbf{W}\mathbf{y}$
$\boldsymbol{\Omega}^{(\ell+1)}=\operatorname{diag}\left(|\boldsymbol{\nabla}_i^\top\boldsymbol{\beta}^{(\ell+1)}+\delta|^{-1}\right)_{i=1}^{m}$
end for

B. Box-constraints

A common problem with scaling exponent estimation is that sometimes the estimates end up outside the defined, a priori, or reasonable, range of the given exponent. For example, it can be the case that Hurst parameter estimates land outside the valid interval $[0,1]$. In broader terms, the practitioner may be in one of the following scenarios: they have access to at least limited training data from which to estimate, learn, or tune sensible bounds; they have a physical model which suggests that the exponent to be estimated lies on some interval; and/or wish to constrain the solution to a known particular interval, as is sometimes the case in filtering or denoising tasks. It should also be worth mentioning that the bounds need not be optimally tight to the 'real' values of the scaling exponents in order to garner some benefits of introducing this prior knowledge. In this sense, even quite conservative bounds may be of some use to the extent that the incorporation of some prior information is better than none at all.

The occurrence of values outside the permitted range only increases when fewer samples are used— in particular, when a localised Hurst estimation strategy is followed. Although one could threshold values to force them to lie inside a reasonable or valid range, this is clearly not an ideal policy. Instead, we propose to incorporate constraints into the least-squares framework. Surprisingly, it appears that this has not been attempted before for scaling exponent estimation. Indeed it has not, in particular, been attempted for Hurst parameter estimation.

Happily, as with total-variation smoothing, we find that the constraints can be accommodated quite readily not only by the regularisation/penalty method but also by our iteratively

reweighted computational framework. We proceed by outlining the penalty approach to the ordinary least-squares problem, as outlined by Mead and Renaut [38]. They approach the constrained least-squares problem

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2, \\ \text{subject to} \quad & \boldsymbol{\beta}_- \leqslant \boldsymbol{\beta} \leqslant \boldsymbol{\beta}_+ \end{aligned} \tag{15}$$

Writing the box-constraint bounds as quadratic constraints

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2, \\ \text{subject to} \quad & \big(\boldsymbol{\beta}[i] - \overline{\boldsymbol{\beta}}[i]\big)^2 \leq \sigma_i^2, \end{aligned}$$

where $[i]$ indicates the ith element of the vector; and with $\overline{\boldsymbol{\beta}} = (\boldsymbol{\beta}_+ + \boldsymbol{\beta}_-)/2$, and $\boldsymbol{\sigma} = (\boldsymbol{\beta}_+ - \boldsymbol{\beta}_-)/2$, allows the problem to be expressed in penalty form, namely

$$\min_{\boldsymbol{\beta}} \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right\|_2^2 + \left\|\mathbf{C}_\epsilon^{-1/2}\left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}}\right)\right\|_2^2, \tag{16}$$

with $\mathbf{C}_\epsilon := \epsilon\mathbf{C}$ and ϵ is reduced, from unity, at each step as shown in line 8 of Algorithm 2. Note that the $\mathbb{I}$ symbol on line 6 denotes the indicator function which returns unity if the argument statement is true and otherwise returns zero. The condition $\mathbf{z} \neq 1$ on line 3 requires that all elements of $\mathbf{z}$ are unity or true before the loop stops. In practice, this could be relaxed to some maximum number of iterations or minimum difference in consecutive estimates of $\boldsymbol{\beta}$.

The adaptation to scaling exponent estimation can be achieved by letting $\mathbf{y}$ be the wavelet log-variances and then, assuming (cf. Eqn. 10) that the $\boldsymbol{\beta} = (\boldsymbol{\beta}[i])_{i=1}^{2n_0}$ are arranged as

$$\boldsymbol{\beta} = \big[\beta_1[1], \beta_2[1], \beta_1[2], \beta_2[2], \ldots, \beta_1[n_0], \beta_2[n_0]\big]^\top$$

gives the constraints as

$$\begin{aligned} \boldsymbol{\beta}_-[2i] &= B_-, & \boldsymbol{\beta}_-[2i+1] &= -L, \\ \boldsymbol{\beta}_+[2i] &= B_+, & \boldsymbol{\beta}_+[2i+1] &= L. \end{aligned}$$

Setting L to be large ensures that only the log-slope parameters $\beta_2[i]$ are constrained. The intercept is left unconstrained. Note that the unweighted case can be selected by simply choosing the weights $\mathbf{W}$ to be the identity matrix. A specific example, for Hurst exponent estimation for self-similar processes, say, would be to set $B_- = 2$ and $B_+ = 4$. This would ensure that the estimate for H lies in $[0, 1]$.

C. Box-constrained total variation

Since both total-variation and box-constraints can be managed under a similar penalised, iterative computational framework it seems quite natural to fuse them under our existing system. This hence leads us to propose the Bounded box constrained total variation (BBC-TV) least-squares scaling exponent estimation method as outlined in Algorithm 3.[1]

To help justify this, we note that since the total variation penalised least squares can be written as a weighted ridge regression, cf. (14), it can therefore be rewritten, in block-form, as a least-squares problem. Hence, it can be treated as a constrained least-squares problem in much the same way as the basic least-squares set-up, cf. (15). In summary (14) and (15) are combined to give

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \|\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}\|_2^2, \\ \text{subject to} \quad & \boldsymbol{\beta}_- \leqslant \boldsymbol{\beta} \leqslant \boldsymbol{\beta}_+ \end{aligned}$$

with

$$\mathbf{y}^* := \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{X}^* := \begin{bmatrix} \mathbf{X} \\ \sqrt{\lambda}\boldsymbol{\Omega}^{1/2}\nabla \end{bmatrix}, \tag{17}$$

Substituting $\mathbf{y}^*$ and $\mathbf{X}^*$ from (17) into Algorithm 2 gives Algorithm 3.

Algorithm 2 Bounded box constrained (BBC) least-squares via the penalty method

1: **Inputs:**
design matrix $\mathbf{X}$, data $\mathbf{y}$, weights $\mathbf{W}$,
downsample factor $s = 2^{j_0}$,
constraints $\{\boldsymbol{\beta}_-, \boldsymbol{\beta}_+\}$
2: **Initialize:**
unweighted: $\mathbf{W} = \mathbf{I}$ or weighted:
$\mathbf{W} = \mathbf{I}_{n_{j_0}} \bigotimes \operatorname{diag}(\max(1, s^2 2^{-2j}))_{j_-}^{j_+}$
$\overline{\boldsymbol{\beta}} = (\boldsymbol{\beta}_+ + \boldsymbol{\beta}_-)/2$,
$\mathbf{C}_\epsilon^{-1} = \operatorname{diag}\left(\dfrac{4}{(\boldsymbol{\beta}_+[i] - \boldsymbol{\beta}_-[i])^2}\right)_{i=1}^{p}$
$\epsilon = 1$, $\ell = 0$, $\mathbf{z} = \mathbf{0} \in \mathbb{R}^p$
3: **while** $\mathbf{z} \neq \mathbf{1}$ **do**
4: $\quad\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{W}\mathbf{X} + \mathbf{C}_\epsilon^{-1})^{-1}(\mathbf{X}^\top\mathbf{W}\mathbf{y} + \mathbf{C}_\epsilon^{-1}\overline{\boldsymbol{\beta}})$
5: $\quad$**for** i **do**
6: $\quad\quad z[i] = \mathbb{I}(\boldsymbol{\beta}_-[i] \leqslant \widehat{\boldsymbol{\beta}}[i] \leqslant \boldsymbol{\beta}_+[i])$
7: $\quad$**end for**
8: $\quad\epsilon = \dfrac{\epsilon}{1 + \ell/10}$
9: $\quad\mathbf{C}_\epsilon^{-1} = \operatorname{diag}(1 + \epsilon^{-1}z[i])\mathbf{C}_\epsilon^{-1}$
10: $\quad\ell = \ell + 1$
11: **end while**

V. Experiments

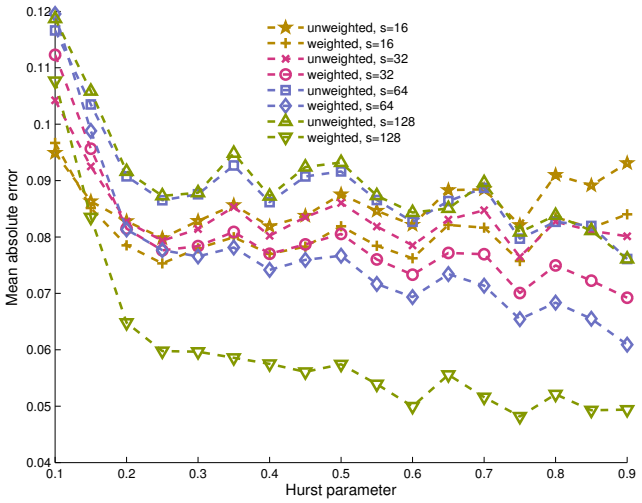
We here focus on Hurst estimation for self-similar processes but note that the ideas can be readily transferred to the more general case of scaling parameter estimation for any process which possesses a power-law and quasi-decorrelation properties.

A. Simulated data

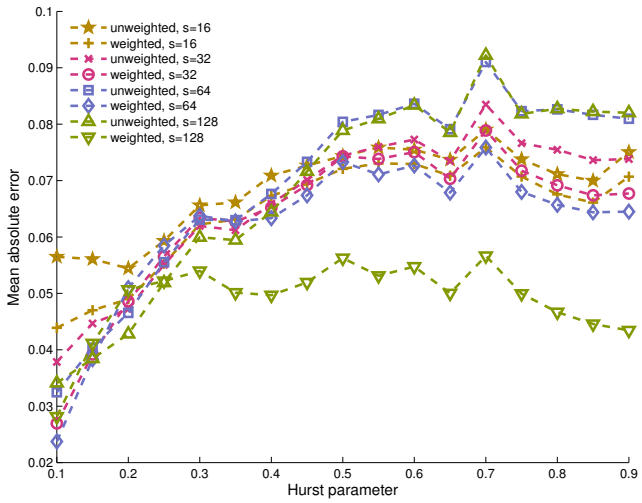
Figure 4 illustrates the absolute errors of the WLS Hurst estimates (cf. Eqn. (10)) for fractional Brownian fields of size 256×256 with globally constant Hurst with two different downsampling values, namely $s = 16$ and $s = 32$, and using three different weighting strategies. The unweighted scheme simply assigns $w_j \equiv 1$. The naive weighted scheme has $w_j = 2^{-2j}$; and, cf. Eqn. (11), the (proposed) weighted scheme chooses $w_j = \max\left(1, s^2 2^{-2j}\right)$. The proposed weights outperform both the naive weighted, and unweighted, schemes.

Figure 5 shows the absolute errors of Hurst estimates for fractional Brownian fields of size 256×256 with a globally constant Hurst parameter for the unconstrained and constrained, weighted and unweighted, methods. It can be

[1] Unlike Beck and Teboulle [37], who consider constrained TV regularisation for denoising and deblurring, the proposed algorithm is based on IRLS and is applied to the problem of scaling exponent estimation.



(a) Unconstrained weighted (WLS) and unweighted (OLS) least-squares.



(b) Constrained weighted (BBC-WLS) and unweighted (BBC-OLS).

Fig. 5. Mean absolute errors of local Hurst estimates obtained on fractional Brownian surfaces with globally constant Hurst parameter, for different values of the downsampling parameter s. Left: without constraints (OLS and WLS); right: with constraints (BBC-OLS and BBC-WLS).

Algorithm 3 Bounded box constrained total variation (BBC-TV) least-squares via the penalty method and iteratively reweighted least-squares

Inputs:
design matrix $\mathbf{X}$, data $\mathbf{y}$, weights $\mathbf{W}$,
downsample factor $s = 2^{j_0}$,
constraints $\{\boldsymbol{\beta}_-, \boldsymbol{\beta}_+\}$
regularisation parameter λ

Initialize:
unweighted: $\mathbf{W} = \mathbf{I}$ or weighted:
$\mathbf{W} = \mathbf{I}_{n_{j_0}} \bigotimes \mathrm{diag}(\max(1, s^2 2^{-2j}))_{j_-}^{j_+}$
$\overline{\boldsymbol{\beta}} = (\boldsymbol{\beta}_+ + \boldsymbol{\beta}_-)/2$,
$\mathbf{C}_\epsilon^{-1} = \mathrm{diag}\left(\dfrac{4}{(\boldsymbol{\beta}_+[i] - \boldsymbol{\beta}_-[i])^2}\right)_{i=1}^{p}$
$\epsilon = 1$, $\ell = 0$, $\mathbf{z} = \mathbf{0} \in \mathbb{R}^p$

while $\mathbf{z} \neq \mathbf{1}$ **do**
  $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{W} \mathbf{X} + \lambda \boldsymbol{\nabla}^\top \boldsymbol{\Omega}^{(\ell)} \boldsymbol{\nabla} + \mathbf{C}_\epsilon^{-1})^{-1} (\mathbf{X}^\top \mathbf{W} \mathbf{y} + \mathbf{C}_\epsilon^{-1} \overline{\boldsymbol{\beta}})$
  $\boldsymbol{\Omega}^{(\ell+1)} = \mathrm{diag}\left(|\boldsymbol{\nabla}_i^\top \widehat{\boldsymbol{\beta}} + \delta|^{-1}\right)_{i=1}^{m}$
  for i **do**
    $z[i] = \mathbb{I}(\boldsymbol{\beta}_-[i] \leqslant \widehat{\boldsymbol{\beta}}[i] \leqslant \boldsymbol{\beta}_+[i])$
  end for
  $\epsilon = \dfrac{\epsilon}{1 + \ell/10}$
  $\mathbf{C}_\epsilon^{-1} = \mathrm{diag}(1 + \epsilon^{-1} z[i]) \mathbf{C}_\epsilon^{-1}$
  $\ell = \ell + 1$
end while

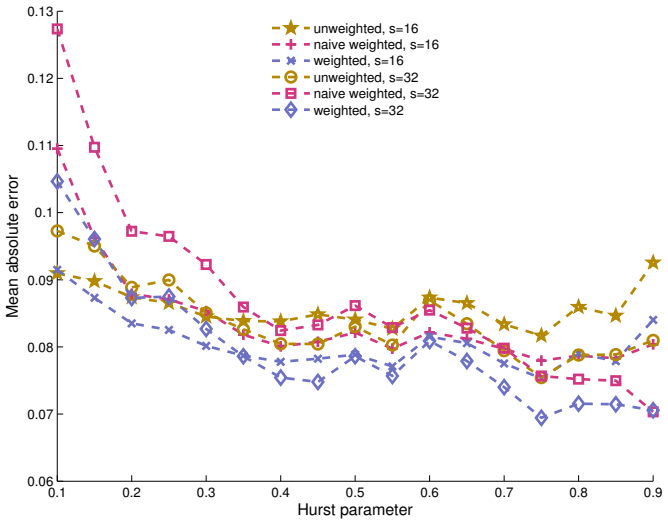


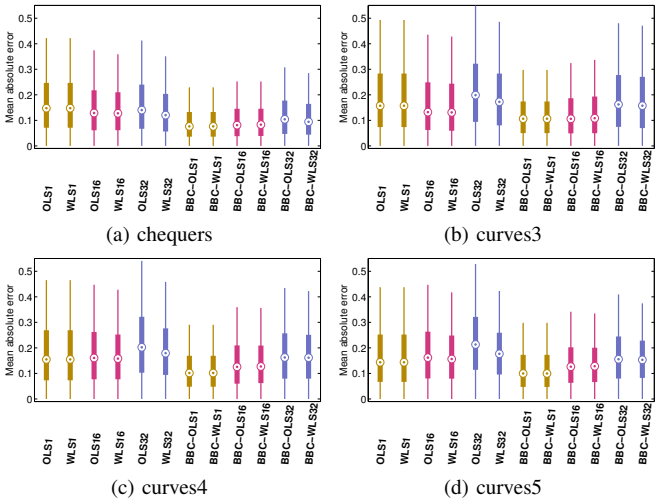
Fig. 4. Mean absolute errors of local Hurst estimates obtained on fractional Brownian surfaces with globally constant Hurst parameter, for different values of the downsampling parameter s and using: no weighting; naive weighting; and our proposed weighting across different scale levels.

seen that the weighted scheme holds an advantage over the unweighted scheme for all globally constant Hurst values and for both unconstrained and constrained.

The errors in both Figure 4 and Figure 5a are generally slightly greater for small Hurst values. At these parameter values, the spectral slope is quite shallow and small positive errors in the $\log_2 \mu_j$ statistic can lead to overly small, or indeed negative, Hurst estimates. This issue can be resolved somewhat by increasing the finest scale level j_- to be used in the regression. However, that may result in worse performance for larger values of the Hurst parameter. Since the Hurst parameter is not known a priori, practitioners must fix j_- (and j_+) based on cross-validation or other empirical testing. A comparison between the left and right panels of Figure 5 reveals that the constraints offer generally better estimates and that this advantage is especially evident, in this experiment, for the small values of the Hurst parameter. The right panel confirms that the proposed constraints can be successfully combined with the proposed weighted scheme.

Note that for both constrained or unconstrained the larger the downweighting factor s, the smaller the error. This is not surprising since the fractional Brownian field used to test the estimators in this experiment was endowed with a globally constant Hurst parameter. As such, the spatial locality sacrificed by choosing a large s has no detrimental effect.

On the other hand, when the Hurst parameter varies over



(a) chequers (b) curves3

(c) curves4 (d) curves5

Fig. 6. Mean absolute errors of local Hurst parameter estimates obtained on fractional Brownian surfaces with spatially varying Hurst parameter for different values of the downsampling parameter s.

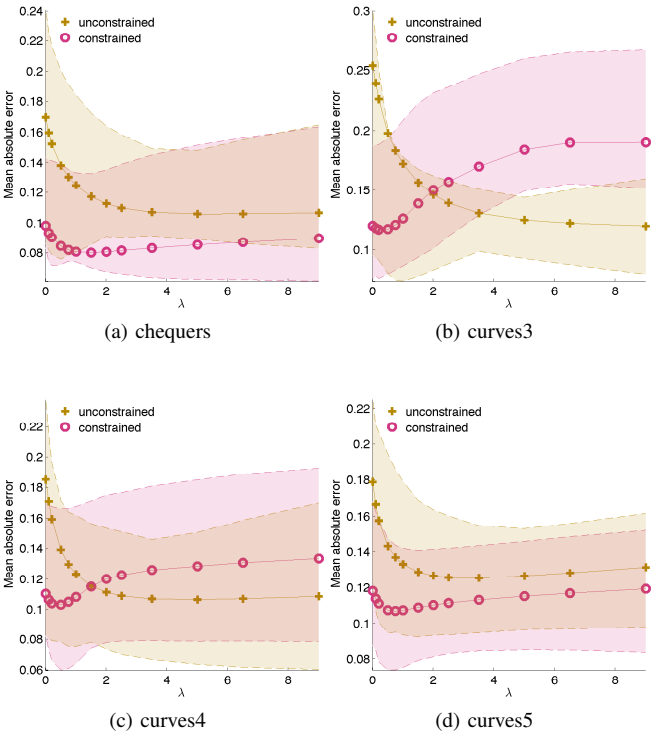
space, arbitrarily larger downsampling factors do not necessarily improve the estimates. Instead, a trade-off must be struck between the computational speed-up afforded by downsampling and the resulting loss in estimate accuracy relative to how locally one expects the Hurst parameter to vary.

Figure 6 uses box plots to illustrate the absolute errors of a variety of Hurst estimation methods using the data 'chequers', and 'curves' 3-5 (cf. Figs. 2a-2d) with size 64×64. The labels along the abscissa denote the unconstrained least-squares method with/without constraints (WLS/OLS), the constrained method with/without weights (BBC-WLS/BBC-OLS), using downsampling rates of 1, 16, and 32; e.g. the label 'BBC-OLS16' in the figure refers to the constrained, non-weighted least-squares scheme with a downsampling of $s = 16$.

From the figure, one can draw several conclusions, namely: the constraints generally improve accuracy (compare with/without BBC); some subsampling can be beneficial on data where the variation of the Hurst parameter is relatively small (compare OLS16 with OLS1 for 'chequers' and 'curves3'); however, when too much subsampling is used or when the Hurst parameter varies more locally the estimates lose local information and become poorer (compare OLS32 against OLS16 and OLS1)— hence here lies a trade-off between estimation error and computational speed; when too much subsampling is used, the weighting method appears to help with accuracy in this experiment (compare OLS32 with WLS32) although this benefit is mitigated somewhat when the constraints are added and is only small or negligible when the subsampling rate is small (compare other OLS against WLS).

Hence, overall, both the constrained method and the weighting method may have their use depending on the data, prior assumptions, other parameter settings such downsampling factor, and application considerations such as computational resources, time constraints, and required accuracy.

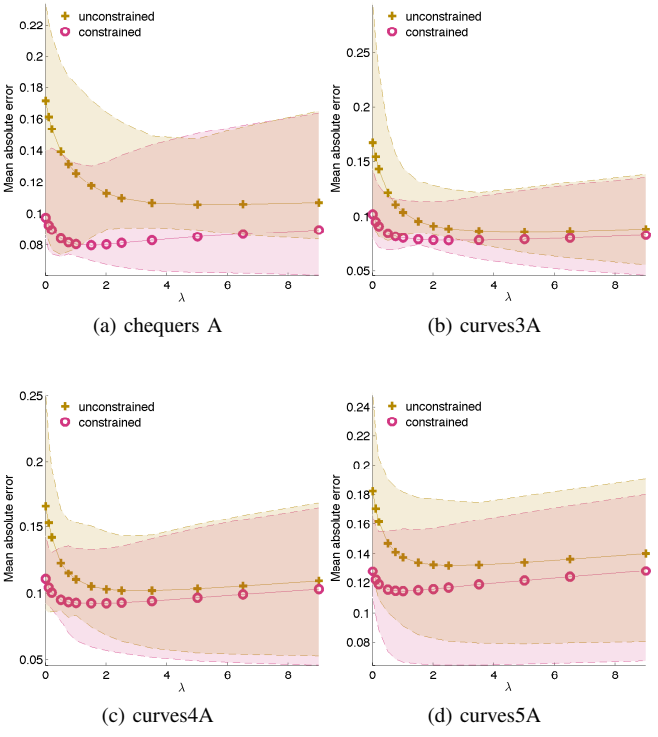
Figure 7 and 8 compare the absolute errors of the unconstrained and proposed constrained (total-variation regularised) Hurst estimation methods over different total-variation regularisation parameters using 64×64 versions of the data illustrated in Figure 2. Here the constrained version holds an



(a) chequers (b) curves3

(c) curves4 (d) curves5

Fig. 7. Mean absolute errors of local Hurst parameter estimates obtained on fractional Brownian surfaces (chequers and curves3, 4, 5) with spatially varying Hurst parameter using total variation regularisation with/without constraints, over different values of the regularisation parameter λ. Error bars represent upper-lower quartiles.



(a) chequers A (b) curves3A

(c) curves4A (d) curves5A

Fig. 8. Mean absolute errors of local Hurst parameter estimates obtained on fractional Brownian surfaces (chequers A and curves3A, 4A, 5A) with spatially varying Hurst parameter using total variation regularisation with/without constraints, over different values of the regularisation parameter λ. Error bars represent upper-lower quartiles.

advantage over the unconstrained one. In Figure 7, the optimal regularisation value of the constrained method is quite close

to zero. Whereas, when the alternative data is used (which contains less variation in the Hurst parameter) the optimal value is, as expected, larger. These results indicates two main behaviours. Firstly, since the constraints essentially restrict the interval over which the estimates can lie, less regularisation is required to optimise the resulting objective function; in a sense, the constraint performs some level of regularisation in addition to the total variation. Secondly, when less (Hurst) variation is present in the data, the optimal level of total variation will be larger so as to properly reflect the prior information that the Hurst parameter varies more smoothly.

B. Real data

As discussed in the introduction, as long as model assumptions hold, empirical results have corroborated the logic that improved Hurst estimation will give rise to better performance in scaling exponent-led image processing tasks such as segmentation or denoising. It is therefore anticipated that, with suitable adaptation, the superior performance of the weighted, TV-smoothing, and constrained strategies will further enhance many existing applications. Although a full exploration of these are beyond the present scope, we here offer a motivating example of segmentation for synthetic aperture sonar imagery. The upper panel of Figure 9 illustrates sonar imagery captured by the MUSCLE sensor. This imaging modality is used by environmental and defence scientists, amongst others, for a broad range of tasks including seabed classification and segmentation.

In defence applications, for example, one is typically concerned with the detection of mine-like objects. Unfortunately, as Groen et al. point out, computer-based mine-like detectors suffer from dramatic increases in false positives when confronted with sand-rippled seabeds [39]. This inspired Nelson and Kingsbury to propose a wavelet-based shrinkage method to adaptively suppress ripple energy and, hence, mitigate the deleterious effects of the seabed type on detection performance [5], [40], [41].

More precisely, the sonar imagery is modelled as a random field over the lattice $X\colon \mathcal{T} \mapsto \mathbb{R}$ with some underlying hidden state $Y\colon \mathcal{T} \mapsto \{0,1\}$ defined at each site, where 0 denotes non-ripple and 1 denotes ripple. The ripples are adaptively shrunk in the wavelet domain via a shrinkage operator defined as the multiplication of the wavelet coefficients with the marginal posterior probability of non-ripple, namely $(\mathcal{S}\cdot)_i := \cdot\,\mathbb{P}(Y_i = 0|X_i)$. The result is then transformed back into the original domain: $\widetilde{X} := \mathcal{W}^{-1}\mathcal{S}\mathcal{W}X$ thanks to the perfect reconstruction property of the dual-tree complex wavelet basis used. The adaptive (non-linear) filtered image $\widetilde{X}$ will contain significantly less energy than the original in the scale-orientation-spatial locations of the ripple. However, elsewhere, the filtered and original image should be very similar or exactly the same.

Happily, an effective choice of the likelihood function is the local and directional scaling exponent estimate. As such, this application represents a good, real-world application of a scaling exponent-based denoising task discussed, in generality, by [7] and, also [6], [10], [8]. In particular, here there is a need to adaptively separate the power-law signal (non-rippled seabed) and the pseudo-periodic noise (sand ripples).

The key to this task is performing an accurate and stable directional estimate of the scaling parameter. For that, the analysis in the preceding sections is applied to each wavelet sub-band direction separately. This results in a semi-local and directional estimate of the scaling parameter, say $H_m(t)$ as plotted in the bottom panel of Figure 9. For the dual-tree complex wavelets used here, there are six directions, namely $(30m - 15)^\circ$ for $m = 1, \ldots, 6$. Four combinations of methods are used for comparison, with/without constraints and with/without total variation smoothing. It can be seen that, in all methods, sub-bands two and three show a reasonably clear region of small scaling exponent in the ripple field and a relatively larger scaling exponent estimate outside the ripple field. This reflects both the location and general directionality of the sand ripples. On close inspection one may also be able to observe that the smoothing method offers more spatially coherent estimates, and that the constrained methods offer more homogeneous estimates, of the scaling exponent.

For illustrative clarity, Figure 10 shows thresholded versions of the estimates. Ideally, these should contain a contiguous black region in sub-bands two and three which reflects the position of the ripple field. It should arguably maybe contain some content inside the ripple region in the neighbouring bands one and four and be fairly clean in the other bands. The superiority of the TV smoother is evident. Rows two and four are both cleaner outside bands two and three and, at the same time, the dark region in bands two and three capture the true ripple field more faithfully than rows one and three. A modest improvement in the use of the constraints is also evident here. For example, the black regions in bands two and three of the fourth row cover more of the ripple field — there are fewer and smaller holes— than the second row. The segmentation results are, by no means, perfect. However, it should be noted that, (i) in practice one may either use the non-thresholded values to construct a likelihood of the ripple directly and/or combine this with a neighbourhood structure prior, such as a Markov random field to improve the segmentation; (ii) performing this segmentation in the wavelet domain affords the possibility to carry out shrinkage and suppress the ripple effects in a broadly similar way to [40]; (iii) the more accurate one can make the segmentation at this stage in the processing, the easier and more effective steps (i) or (ii) above will be.

VI. Conclusion

Theorem 2.2 and subsequent statistical analysis above established the fact that, since the wavelet coefficients of self-similar processes are quasi-decorrelated, the least-squares-based approach is properly motivated and can also be extended to the semi-local case. As a direct result, the following two main methodological developments have thus been added to the local scaling exponent estimation framework.

1) The first development, suggested by our semi-local version of Abry's asymptotic result, was the proposed weighting scheme. This provides a statistically well-principled means to conduct semi-local weighted-least squares estimation.

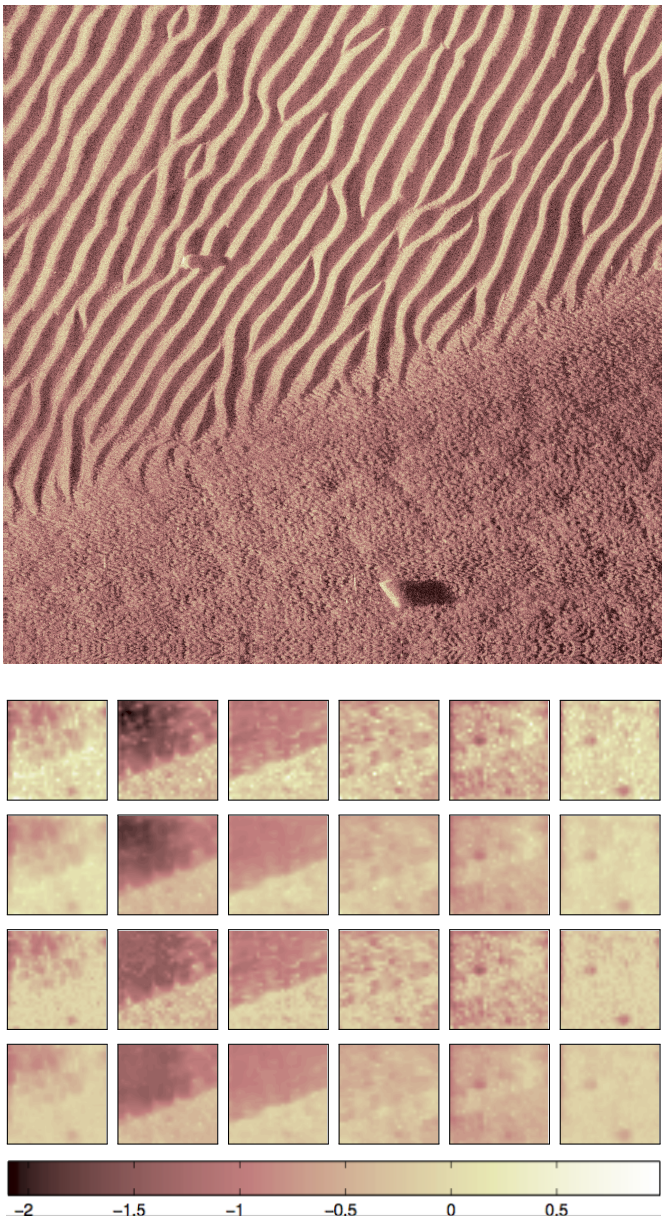


Fig. 9. Upper panel: MUSCLE sonar imagery, courtesy of the Centre for Maritime Research and Experimentation. Note the sand rippled region in the upper-left part of the image. Also present are two mine-like objects. Bottom panel— estimated scaling exponent of the MUSCLE data using: unconstrained without TV smoothing [row 1]; unconstrained with TV smoothing [row 2]; constrained without TV smoothing [row 3]; constrained with TV smoothing [row 4]. The columns of the estimate sub-figures the directional response $(30m - 15)^\circ$, for $m = 1, \ldots, 6$.

2) The second development was to pull through work on the use of the penalty method to incorporate box-constraints into the least-squares framework.

We have shown that, thanks to the least-squares framework and iteratively reweighted least squares algorithm, these two innovations can be successfully combined and, indeed, can be incorporated into the recent total variation regularisation framework to achieve better estimates. Table II summarises the various combinations of methods introduced here.

There remain many open questions for further study, including the following.

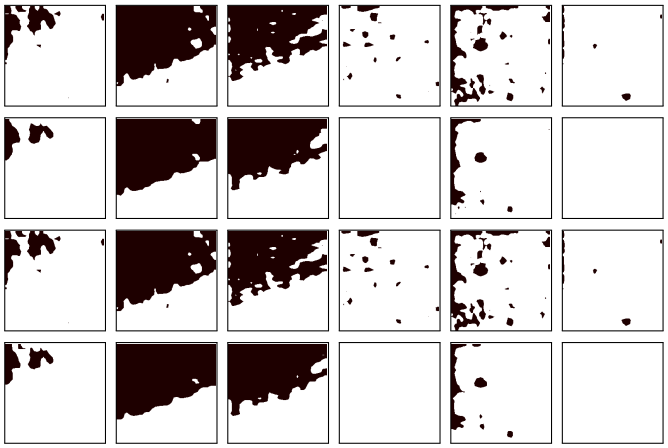


Fig. 10. Thresholded versions of the lower panel of Figure 9, namely the estimated scaling exponent of the MUSCLE data using: unconstrained without TV smoothing [row 1]; unconstrained with TV smoothing [row 2]; constrained without TV smoothing [row 3]; constrained with TV smoothing [row 4].The columns of the estimate sub-figures the directional response $(30m - 15)^\circ$, for $m = 1, \ldots, 6$.

(F1) Although no obvious convergence problems were suffered during our simulations, a convergence proof for the proposed BBC-TV algorithm is outstanding and would provide a natural research target.

(F2) The quasi-decorrelation result, along with the weighted, constrained, and TV-regularised estimation, could be extended to the anisotropic case. This would provide an updated framework of the anisotropic OLS work in [23] and/or a non-stationary extension of [42].

(F3) An alternative direction would be to consider estimator robustness for the non-Gaussian case. Abry et al. [14] studied the robustness of the Hurst estimator as the Gaussianity assumptions were relaxed somewhat. A theoretical and numerical study of such robustness for the model discussed here would be of interest. Recent work by Nelson et al. [13] has generalised the Gaussian likelihood to the exponential family in the regularised Hurst estimation framework to accommodate outliers. Their analysis was by no means complete in that the source of the outliers was not rigorously described. It would be of interest to incorporate such analysis into the framework presented here and to further strengthen the theory.

(F4) The estimation framework here is based purely on a maximum a posteriori approach. This does not immediately offer confidence intervals on the estimates. Any estimation variance reported here is arrived at empirically via repeated experiment with full knowledge of ground truth. Generating uncertainty bounds without recourse to ground truth would, of course, be preferable.

(F5) The related work of Pustelnik et al. [3], [24], [25] comprises some noteworthy features which could offer interesting ways forward when combined with the work herein. The use of wavelet leaders for overcomplete bases such as the dual-tree complex wavelets used here is yet to be investigated, for example.

(F6) A comparison of scaling exponent based methods with

TABLE II
THE FAMILY OF THE PROPOSED CONSTRAINED, WEIGHTED LEAST-SQUARES AND TOTAL VARIATION LOCAL HURST ESTIMATION METHODS: $f(\boldsymbol{\beta};\mathbf{W},\lambda)=\left\|\mathbf{W}^{1/2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\|_2^2+\lambda\left\|\boldsymbol{\beta}\right\|_1$, WITH/WITHOUT CONSTRAINTS $\boldsymbol{\beta}_-\leqslant\boldsymbol{\beta}\leqslant\boldsymbol{\beta}_+$

Form	Name	Proposed (as Hurst estimator)
Unconstrained		
$f(\cdot;\mathbf{I},0)$	OLS	(e.g. Nelson and Kinsgbury [23])
$f(\cdot;\mathbf{W},0)$	WLS	here: Eqn (10) (cf. [14], [15], [16])
$f(\cdot;\mathbf{I},\lambda)$	TV (gen. Lasso)	Nafornita et al. [12], [13] (cf. [36])
$f(\cdot;\mathbf{W},\lambda)$	TV (weighted)	here: Algorithm 1
Constrained		
$f(\cdot;\mathbf{I},0)$	BBC-OLS	here: Algorithm 2 (cf. [38])
$f(\cdot;\mathbf{W},0)$	BBC-WLS	here: Algorithm 2
$f(\cdot;\mathbf{I},\lambda)$	BBC-TV	here: Algorithm 3
$f(\cdot;\mathbf{W},\lambda)$	BBC-WTV	here: Algorithm 3

other approaches to common image processing tasks such as denoising, deblurring, segmentation, etc.

VII. ACKNOWLEDGEMENTS

The authors thank the reviewers for some discerning comments which greatly improved the presentation of the paper.

APPENDIX A
PROOF OF 2D WAVELET COVARIANCE DECAY

To prove Theorem 2.2 we first establish a couple of results as follows.

Lemma A.1: Let $\psi\colon\mathbb{R}^2\mapsto\mathbb{R}$ be a separable (2-d) wavelet which is constructed from a 1-d wavelet with M-many vanishing moments. Then, the function

$$\Lambda(t):=\int_{\mathbb{R}^2}\psi(s)\psi(s-t)\mathrm{d}s$$

has $2M$ vanishing moments in the non-diagonal directions and $4M$ vanishing moments in the diagonal directions in that

$$\int_{\mathbb{R}^2}\|t\|^m\,\Lambda_\theta(t)\mathrm{d}t=0,\quad\text{for } m=0,\ldots,M_\theta-1,$$

with $M_\theta=2M$ for $\theta\in\{0,1\}$, say, and $M_2=4M$.

Proof With some abuse of notation, separable two-dimensional wavelets take the form $\psi_0(t_0,t_1)=\phi(t_0)\psi(t_1)$, $\psi_1(t_0,t_1)=\psi(t_0)\phi(t_1)$, and $\psi_2(t_0,t_1)=\psi(t_0)\psi(t_1)$, say, where the subscript on the ψ denotes orientation. We note that

$$\int_{\mathbb{R}^2}\|t\|^m\,\Lambda(t)\mathrm{d}t=-\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\|s-t\|^m\,\psi(s)\psi(t)\mathrm{d}s\mathrm{d}t$$

$$=-\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\sum_{n=0}^{m/2}\binom{m/2}{n}(s_0-t_0)^{m-2n}(s_1-t_1)^{2n}\psi(s)\psi(t)\mathrm{d}s\mathrm{d}t$$

Taking the summation outside the integrals reveals that, for orientation $\theta=0$, say, we have a weighted sum of terms with a factor of the form

$$\int\int(s_1-t_1)^{2n}\psi(s_1)\psi(t_1)\mathrm{d}s_1\mathrm{d}t_1,$$

where $n=0,\ldots,m/2$. The term $(s_1-t_1)^{2n}$, itself, admits a binomial expansion, the terms of which are all be annihilated by either the $\psi(s_1)$ or $\psi(t_1)$ functions courtesy of the vanishing moments of ψ unless $m\geqslant 2M$. A similar argument for $\theta\in\{1,2\}$ completes the proof. ■

Theorem A.2: The wavelet coefficients of an isotropic self-similar process with stationary increments defined over $\mathbb{R}^2$, are stationary.

Proof First we note that the mean $\mathbb{E}d_j[k]=0$ does not vary with k. For isotropic H-sssi fields it is well known (e.g. [43]) that

$$\gamma_X(s,t):=\mathbb{E}X(s)X(t)\propto\|s\|^{2H}+\|t\|^{2H}+\|s-t\|^{2H}. \quad (18)$$

For $k,k'\in\mathbb{Z}^2$, and $\psi_{jk}:=2^{-j}\psi(2^{-j}\cdot-k)$, we have

$$\mathbb{E}d_j[k]d_j[k']=\mathbb{E}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}X(s)X(t)\psi_{j,k}(s)\psi_{j,k}(t)\mathrm{d}s\mathrm{d}t.$$

The expectation can be taken inside the integral and, from (18) and some simple substitutions, we have

$$\mathbb{E}d_j[k]d_j[k']=C\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\gamma_X(s+k',t+k)\psi_{j,k}(s)\psi_{j,k}(t)\mathrm{d}s\mathrm{d}t.$$

Since $\int\psi(t)\mathrm{d}t=0$, this becomes

$$\mathbb{E}d_j[k]d_j[k']=C2^{2j(H+1)}\int_{\mathbb{R}^2}\|t+k'-k\|^{2H}\,\Lambda(t)\mathrm{d}t, \quad (19)$$

with

$$\Lambda(t):=\int_{\mathbb{R}^2}\psi(s)\psi(s-t)\mathrm{d}s,$$

as in Lemma A.1. The result is then obtained by noting that the right-hand-side is a function of lag $(k-k')$ only. ■

The main theoretical result is here established as follows.

Theorem A.3: Let the compactly supported wavelet $\psi\colon\mathbb{R}^2\mapsto\mathbb{R}$ possess M-many vanishing moments. Then the wavelet coefficients $d_j[k]=\langle B_H,\psi_{jk}\rangle$ of an isotropic H-sssi field B_H, defined over $\mathbb{R}^2$ with Hurst parameter H, satisfy

$$\mathbb{E}d_j[k]d_j[k']=O\Big(\left\|2^j(k-k')\right\|^{2(H-M)}\Big).$$

Proof From Theorem A.2 and Equation (19) we have

$$\gamma_{d_j}(\alpha):=\mathbb{E}d_j[\cdot]d_j[\cdot+\alpha]\propto\int_{\mathbb{R}^2}\|t+\alpha\|^{2H}\,\Lambda(t)\mathrm{d}t.$$

We note that, since ψ has compact support so too does Λ. Hence, the integral has finite support $\mathrm{mes}(\mathcal{S}^2)<\infty$. Furthermore, since $\|t+\alpha\|^{2H}\geq 0$ and $\Lambda(t)$ are bounded and integrable over $\mathcal{S}^2$ it follows, from the Mean-Value Theorem (for double integrals), that there exists $t_0\in\mathcal{S}^2$ such that

$$\left|\int_{\mathcal{S}^2}\|t+\alpha\|^{2H}\,\Lambda(t)\mathrm{d}t\right|=|\Lambda(t_0)|\int_{\mathcal{S}^2}\|t+\alpha\|^{2H}\,\mathrm{d}t.$$

Similarly, there exists a $t_1\in\mathcal{S}^2$ such that

$$\left|\int_{\mathcal{S}^2}\left(\|t\|+\|\alpha\|\right)^{2H}\Lambda(t)\mathrm{d}t\right|=|\Lambda(t_1)|\int_{\mathcal{S}^2}\left(\|t\|+\|\alpha\|\right)^{2H}\mathrm{d}t.$$

Now, since $\|t+\alpha\|^{2H}\leqslant\left(\|t\|+\|\alpha\|\right)^{2H}$, it therefore follows that there exists a $\rho_\Lambda=|\Lambda(t_0)/\Lambda(t_1)|$ such that

$$\left|\int_{\mathcal{S}^2}\|t+\alpha\|^{2H}\,\Lambda(t)\mathrm{d}t\right|\leqslant\rho_\Lambda\left|\int_{\mathcal{S}^2}\left(\|t\|+\|\alpha\|\right)^{2H}\Lambda(t)\mathrm{d}t\right|.$$

To this end, it is now possible to proceed, in spirit, in a similar fashion to Tewfik et al. [27] by working with the right-hand-side expression to establish a bound for the covariance. We begin by reformulating this as

$$\left|\gamma_{d_j}(\alpha)\right| \leqslant c_1 \left\|\alpha\right\|^{2H} \left| \int_{\mathcal{S}^2} \left(\left\|t\right\| / \left\|\alpha\right\| + 1\right)^{2H} \Lambda(t) \mathrm{d}t \right|.$$

For α sufficiently large (outside the smallest disc that encloses $\mathcal{S}^2$, the $(\cdot)^{2H}$ term can be expanded as an infinite binomial series, viz.

$$\left(\left\|t\right\| / \left\|\alpha\right\| + 1\right)^{2H} = \sum_{m=0}^{\infty} \eta_m \left(\frac{\left\|t\right\|}{\left\|\alpha\right\|}\right)^m,$$

with

$$\eta_m = \begin{cases} 1, & \text{for } m = 0 \\ \dfrac{2H(2H-1)\cdots(2H-m+1)}{m!}, & \text{otherwise} \end{cases}$$

Now, since Λ has $2M$ vanishing moments and since η is a monotonically decreasing sequence, it follows that

$$\begin{aligned} \left|\gamma_{d_j}(\alpha)\right| &\leqslant c_2 \left\|\alpha\right\|^{2H} \sum_{m=1}^{\infty} \left|\eta_m\right| \left| \int_{\mathcal{S}^2} \left(\frac{\left\|t\right\|}{\left\|\alpha\right\|}\right)^m \Lambda(t) \mathrm{d}t \right| \\ &\leqslant c_3 \left\|\alpha\right\|^{2H} \left|\eta_{2M}\right| \sum_{m=2M}^{\infty} \left| \int_{\mathcal{S}^2} \left(\frac{\left\|t\right\|}{\left\|\alpha\right\|}\right)^m \Lambda(t) \mathrm{d}t \right|. \end{aligned}$$

By Cauchy-Schwartz $\left|\Lambda(t)\right| = \left|\langle \psi(s-t), \psi(s) \rangle\right| \leqslant \left\|\psi\right\|^2$ and

$$\begin{aligned} \left|\gamma_{d_j}(\alpha)\right| &\leqslant c_4 \left\|\alpha\right\|^{2H} \int \left|\psi(s)\right|^2 \mathrm{d}s \sum_{m=2M}^{\infty} \left| \int_{\mathcal{S}^2} \left(\frac{\left\|t\right\|}{\left\|\alpha\right\|}\right)^m \mathrm{d}t \right| \\ &= c_5 \left\|\alpha\right\|^{2H} \sum_{m=0}^{\infty} \int_{\mathcal{S}^2} \left(\frac{\left\|t\right\|}{\left\|\alpha\right\|}\right)^{m+2M} \mathrm{d}t. \end{aligned}$$

Hence, by another application of the Mean Value Theorem, there exists a $t_0 \in \mathcal{S}^2$ such that

$$\begin{aligned} \left|\gamma_{d_j}(\alpha)\right| &\leqslant c_6 \left\|\alpha\right\|^{2H} \sum_{m=0}^{\infty} \left(\frac{\left\|t_0\right\|}{\left\|\alpha\right\|}\right)^{m+2M} \\ &= c_7 \left\|\alpha\right\|^{2(H-M)} \sum_{m=0}^{\infty} \left(\frac{\left\|t_0\right\|}{\left\|\alpha\right\|}\right)^m. \end{aligned}$$

Since α is sufficiently large, the summands are less than one and the series converges. Pulling in the 2^{2j} factor from Equation (19) into the norm and carrying this through all the above arguments completes the proof. ∎

REFERENCES

[1] I. Zachevsky and Y. Y. Zeevi, "Single-image superresolution of self-similar textures," *IEEE International Conference on Image Processing*, 2013.

[2] P. Abry, H. Wendt, and S. Jaffard, "When Van Gogh meets Mandelbrot: Multifractal classification of painting's texture," *Signal Processing*, vol. 93, no. 3, pp. 554–572, 2013.

[3] N. Pustelnik, H. Wendt, P. Abry, and N. Dobigeon, "Local regularity, wavelet leaders and total variation based procedures for texture segmentation," *ArXiv e-print arXiv:1504.05776*, 2015.

[4] J. Blackledge and D. Dubovitski, "A surface inspection machine vision system that includes fractal analysis," *International Society for Advanced Science and Technology, Journal of Electronics and Signal Processing*, vol. 3, no. 2, pp. 76–89, 2008.

[5] J. D. B. Nelson and N. G. Kingsbury, "Fractal dimension based sand ripple suppression for mine hunting with sidescan sonar," *Int. Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, 2010.

[6] A. Echelard and J. L. Véhel, "Wavelet denoising based on local regularity information," *Proceedings of the European Signal Processing Program*, 2008.

[7] P. Flandrin, P. Gonçalvès, and G. Rilling, "Detrending and denoising with empirical mode decompositions," in *Proceedings of the European Signal Processing Program*, 2004, pp. 1581–1584.

[8] B. D. Vidakovic, G. G. Katul, and J. D. Albertson, "Multiscale denoising of self-similar processes," *Journal of Geophysical Research*, vol. 105, no. D22, pp. 27 049–27 058, 2000.

[9] O. Pont, A. Turiel, and H. Yahia, "An optimized algorithm for the evaluation of local singularity exponents in digital signals," *Combinatorial Image Analysis*, vol. 6636, pp. 346–357, 2011.

[10] P. Legrand and J. L. Véhel, "Local regularity-based image denoising," *IEEE International Conference on Image Processing*, vol. 2, pp. 377–380, 2003.

[11] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for bayesian wavelet based image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.

[12] C. Nafornita, A. Isar, and J. D. B. Nelson, "Regularised, semi-local Hurst estimation via generalised lasso and dual-tree complex wavelets," *IEEE International Conference on Image Processing*, pp. 2689–2693, 2014.

[13] J. D. B. Nelson, C. Nafornita, and A. Isar, "Generalised M-lasso for robust, spatially regularised Hurst estimation," *IEEE Global Conference on Signal and Image Processing*, 2015.

[14] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation, and synthesis of scaling data," in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, Eds. New York: John Wiley & Sons, 2000.

[15] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, Jan 1998.

[16] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 878–897, Apr 1999.

[17] S. Boyd and L. Vandenberghe, "Convex optimization," 2004.

[18] P. Abry, P. Gonçalvès, and J. L. Véhel, *Scaling Fractals and wavelets*. Wiley, 2009.

[19] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, *Self-similarity and long-range dependence through the wavelet lens*. Birkhäuser, 2002, pp. 527–556.

[20] S. Jaffard, B. Lashermes, and P. Abry, "Wavelet leaders in multifractal analysis," in *Wavelet Analysis and Applications*, ser. Applied and Numerical Harmonic Analysis, M. I. V. T. Qian and X. Yuesheng, Eds. Birkhäuser, 2007, pp. 201–246.

[21] A. Ayache, S. Cohen, and J. L. Véhel, "The covariance structure of multifractional Brownian motion," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3810–3813.

[22] S. Corlay, J. Lebovits, and J. L. Véhel, "Multifractional stochastic volatility models," *Mathematical Finance*, vol. 24, pp. 364–402, 2014.

[23] J. D. B. Nelson and N. G. Kingsbury, "Dual-tree wavelets for estimation of locally varying and anisotropic fractal dimension," *IEEE International Conference on Image Processing*, pp. 341–344, 2010.

[24] N. Pustelnik, H. Wendt, and P. Abry, "Local regularity for texture segmentation: Combining wavelet leaders and proximal minimization," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5348–5352, 2013.

[25] N. Pustelnik, P. Abry, H. Wendt, and N. Dobigeon, "Inverse problem formulation for regularity estimation in images," *IEEE International Conference on Image Processing*, pp. 6081 – 6085, 2014.

[26] J.-B. Regli and J. D. B. Nelson, "Piecewise parameterised Markov random fields for semi-local Hurst estimation," *Proceedings of the European Signal Processing Conference*, 2015.

[27] A. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 904–909, 1992.

[28] P. Abry, P. Gonçalvès, and P. Flandrin, "Wavelets, spectrum analysis and $1/f$ processes," *Lecture Notes in Statistics*, vol. 103, pp. 15–29.

[29] L. M. Kaplan and C. C. J. Kuo, "An improved method for 2-d self-similar image synthesis," *IEEE Transactions on Image Processing*, vol. 5, no. 5, pp. 754–761, 1996.

[30] I. S. C. d. P. Regularity Team, "Fraclab Toolbox," http://fraclab.saclay.inria.fr, 2014.

[31] I. Daubechies, R. DeVore, M. Fornasier, and S. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[32] C. Chen, J. Huang, L. He, and H. Li, "Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2713–2720, 2014.
[33] C. Chen, J. Huang, L. He, and H. Li, "Fast iteratively reweighted least squares algorithms for analysis-based sparsity reconstruction," *ArXiv e-print arXiv:1411.5057*, 2014.
[34] X. Zhou, R. Molina, F. Zhou, and A. K. Katsaggelos, "Fast iteratively reweighted least squares for lp regularized image deconvolution and reconstruction," *IEEE International Conference on Image Processing*, pp. 1783–1787, 2014.
[35] Y. She, "Sparse regression with exact clustering," *Electronic Journal of Statistics*, vol. 4, pp. 1055–1096, 2010.
[36] R. J. Tibshirani and J. Taylor, "The solution path of the generalised Lasso," *Annals of Statistics*, vol. 39, no. 3, pp. 1335–1371, 2010.
[37] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.
[38] J. Mead and R. A. Renaut, "Least squares problems with inequality constraints as quadratic constraints," *Linear Algebra and its Applications*, vol. 432, no. 8, pp. 1936–1949, 2010.
[39] J. Groen, E. Coiras, and D. Williams, "Detection rate statistics in synthetic aperture sonar images," *Proceedings of the 3rd International Conference & Exhibition on Underwater Acoustic Measurements (UAM)*, pp. 367–374, 2009.
[40] J. D. B. Nelson and N. G. Kingsbury, "Fractal dimension, wavelet shrinkage, and anomaly detection for mine hunting," *IET Signal Processing Journal*, vol. 6, no. 5, pp. 484–493, 2012.
[41] ——, "Multiresolution markov random field wavelet shrinkage for ripple suppression in sonar imagery," *IMA International Conference on Mathematics in Signal Processing*, 2012.
[42] A. Isar and C. Nafornita, "On the statistical decorrelation of the 2d discrete wavelet transform coefficients of a wide sense stationary bivariate random process," *Digital Signal Processing*, vol. 24, pp. 95–105, 2014.
[43] P. Flandrin, "Wavelets analysis and synthesis of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, 1992.