

CHAPTER 1

Selective Review on Wavelets in Statistics

Yazhen Wang

*Department of Statistics, University of Connecticut
Storrs, CT 06269*

E-mail: yzwang@stat.uconn.edu

This article provides a brief review on main wavelet statistical methodologies developed in past fifteen years. The wavelet methods are used for estimating regression functions, detecting and estimating change-points, solving statistical inverse problems, and studying stochastic processes and self-similarity.

1. Introduction

Wavelets are viewed as a synthesis of ideas originated in engineering, physics, and pure mathematics. In past fifteen years, the field has grown at an explosive rate, and wavelet articles have appeared in a remarkable variety of publications. Wavelets are applied to a diverse set of problems and have made significant technology advancement.

Wavelets enjoy various properties including orthogonality, localization, and fast computational algorithms. Wavelets provide unconditional bases for many useful function spaces and have excellent compression capabilities for functions in the spaces. In other words, these classes of functions have economical descriptions (or sparse representations) in terms of wavelets: we do not need too many terms in the wavelet expansion to approximate a function closely. Successful applications of this property include wavelet shrinkage for signal and image compression and nonparametric regression.

Wavelets enable us to do time-frequency (or time-scale) analysis. Wavelet analysis can ‘unfold’ a signal of interest into a function over the time-frequency plane that tells us ‘when’ which ‘frequency’ occurs. This is similar to music notation, which tells the player which notes (frequency information) to play at any given moment (time information). It allows us to

simultaneously locate frequency content and time content of the signal. The time-frequency analysis is a suitable tool for analyzing non-stationary time series and signals (e.g. locally stationary time series and speech signals). Another application of this time-frequency localization is that wavelets can describe the local features (such as jumps) of a function and provide tools for studying change-points in statistics and edge detection in image processing.

Wavelets provide simultaneous quasi-diagonalization of a large class operators. A wavelet-vaguelette decomposition (WVD) can simultaneously quasi-diagonalize dilation-homogeneous operators (Calderon-Zygmund Operators) and a class of functions. The decomposition allows us to develop a new and better method for solving linear inverse problems and modeling non-stationary processes. Because of the localization property, the WVD provides an excellent tool for the study of change-points for indirect data.

Disordered structures and random processes that are self-similar on certain length and time scales are very common in nature. They are found on the largest and the smallest scales: in galaxies and landscapes, in earthquakes and fractures, in aggregates and colloids, in rough surfaces and interfaces, in glasses and polymers, in proteins and other large molecules. Self-similar structures in wavelets make them advantageous in handling self-similar phenomenon.

These properties of wavelets are utilized to develop wavelet based methodologies for solving statistical problems. The rest of the paper is organized as follows. Section 2 introduces wavelets. Sections 3 and 4 review wavelet shrinkage methods for function estimation in the nonparametric regression model and statistical inverse problems, respectively. Section 5 covers change-points for both direct and indirect data. Section 6 features local self-similarity and non-stationary stochastic processes. Section 7 provides a brief description of basis development in high dimensions.

2. Wavelets

Wavelet bases are derived from father wavelet and mother wavelet by a process of dyadic dilations and translations. We start by introducing Haar wavelet basis. Let

$$\phi(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\psi(x) = \phi(2x) - \phi(2x - 1) = \begin{cases} 1, & 0 \leq x \leq 1/2 \\ -1, & 1/2 < x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

In wavelet terminology, ϕ and ψ are called father and mother wavelets, respectively.

Define the continuous wavelet transformation of a function $f(x)$ by

$$Tf(s, u) = \frac{1}{\sqrt{s}} \int f(x) \psi\left(\frac{u-x}{s}\right) dx,$$

where s and u represent scale (or inverse of frequency) and spatial position (or time), respectively. The two dimensional space defined by the pair of variables (s, u) is called the scale-space (or time-frequency) plane, and the wavelet transformation $Tf(s, u)$ is a bivariate function of scale and position (or time and frequency).

Dyadically dilate and translate $\psi(x)$

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad j = 0, 1, \dots, \quad k = 0, 1, \dots, 2^j - 1.$$

It is easy to verify that $\phi(x), \psi_{j,k}(x)$ form an orthonormal basis in $L^2([0, 1])$. Then any integrable function $f(x)$ on $[0, 1]$ has an expansion

$$f(x) = \beta_{-1,0} \phi(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x),$$

where the wavelet coefficients

$$\beta_{-1,0} = \int f(x) \phi(x) dx, \quad \beta_{j,k} = \int f(x) \psi_{j,k}(x) dx.$$

Because $\psi_{j,k}$ has compact support of width 2^{-j} located at $k 2^{-j}$, $\beta_{j,k}$ corresponds to the continuous wavelet transformation of $f(x)$ at frequency 2^j and spatial position $k 2^{-j}$, and describes the information content of $f(x)$ at frequency 2^j and spatial location $k 2^{-j}$.

Although the Haar basis was discovered in 1910, it is not very useful in practice. Its basis functions are discontinuous, so it is not effective in approximating smooth functions or localizing frequency components. General wavelet bases are obtained as follows. Replacing Haar mother and father wavelets with a pair of very special functions (ϕ, ψ) , we dyadically dilate and translate ψ to obtain a wavelet basis $\phi, \psi_{j,k}$, $j = 0, 1, \dots$, $k = 0, 1, \dots, 2^j - 1$, which form an orthonormal basis in $L^2([0, 1])$ (boundary corrections are needed for some base functions whose supports stretch outside $[0, 1]$). Most useful wavelet bases require the pair (ϕ, ψ) to be smooth

and compactly supported. Unlike Haar basis or Fourier basis, such special ϕ and ψ are very complicated to construct and have no explicit forms. However, they can be numerically evaluated by the fast cascade algorithm. See Daubechies (1988, 1992).

In practice we observe a function, a signal or a stochastic process only at finite number of points. The continuous wavelet transformation needs to be discretized as the discrete wavelet transformation which can be written as a linear transformation represented by an orthogonal matrix \mathcal{W} . Suppose we have observations z_1, \dots, z_n , $n = 2^J$. Let $\mathbf{z} = (z_1, \dots, z_n)$. The discrete wavelet transformation of \mathbf{z} is given by $\mathbf{w} = \mathcal{W}\mathbf{z}$. The elements of \mathbf{w} are called discrete wavelet coefficients, with the $n - 1$ elements of \mathbf{w} indexed dyadically by $w_{j,k}$ for $j = 0, 1, \dots, J - 1$, $k = 0, 1, \dots, 2^j - 1$, and the remaining element labelled as $w_{-1,0}$. Because \mathcal{W} is orthogonal, we can easily reconstruct \mathbf{z} from its discrete wavelet coefficients \mathbf{w} by the inverse discrete wavelet transformation $\mathbf{z} = \mathcal{W}^T \mathbf{w}$.

The rows of \mathcal{W} correspond to a discretized version of the wavelet basis $(\phi, \psi_{j,k})$, and $w_{j,k}$ relate to a discretized version of the continuous wavelet transformation $\beta_{j,k}$ as follows. If we dyadically index the first $n - 1$ rows of \mathcal{W} by $(2^j + k)$ for $j = 0, 1, \dots, J - 1$, $k = 0, 1, \dots, 2^j - 1$, and denote by $W_{j,k}(i)$ the i -th element of the $(2^j + k)$ -th row of \mathcal{W} , then $n^{1/2} W_{j,k}(i)$ is approximately equal to $\psi_{j,k}(i/n)$. If z_1, \dots, z_n are observations of $f(x)$ at $x = i/n$, $i = 1, \dots, n$, then $n^{1/2} w_{j,k} \approx \beta_{j,k}$.

Mallat's pyramidal algorithm requires only $O(n)$ operations for performing discrete wavelet transformation and inverse discrete wavelet transformation. Thus we can fast process wavelet analysis by computing the discrete wavelet coefficients of finite observations and reconstructing the observations from their corresponding discrete wavelet coefficients. See Mallat (1989, 1999).

3. Nonparametric regression

3.1. Wavelet shrinkage

Consider the nonparametric regression model

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n = 2^J, \quad (1)$$

where $x_i = i/n$, ε_i are i.i.d. normal random errors, and $f(x)$ is a function. The problem is to estimate $f(x)$ based on data y_i .

Because wavelets provide sparse representations for a wide class of functions, in a series of papers Donoho and Johnstone proposed wavelet shrink-

age to take advantage of wavelets' sparse representations and efficiently estimate regression functions. They established asymptotic minimax for estimating regression functions over a wide class of Besov spaces and showed that wavelet shrinkage estimators can be turned to achieve minimax rates over the entire scale of Besov spaces and significantly outperform linear estimators.

Wavelet shrinkage works as follows. First compute the discrete wavelet coefficients of data y_1, \dots, y_n ; second shrink the wavelet coefficients; third construct the estimator of $f(x)$ by using the shrunk wavelet coefficients. Specifically, let $(y_{j,k})$, $(\theta_{j,k})$ and $(\varepsilon_{j,k})$ be the discrete wavelet coefficients of (y_i) , $(f(x_i))$ and (ε_i) , respectively. Since discrete wavelet transformation is linear, from model (1) we yield

$$y_{j,k} = \theta_{j,k} + \varepsilon_{j,k}, \quad j = 0, 1, \dots, J-1, \quad k = 0, 1, \dots, 2^j - 1, \quad (2)$$

where $\varepsilon_{j,k}$ are i.i.d. normal, due to the orthogonality of discrete wavelet transformation.

Because of wavelet's sparse representations, there are relatively a small number of large $|\theta_{j,k}|$ and a large number of small $|\theta_{j,k}|$. Shrinking rules are used to select only those $y_{j,k}$ whose corresponding $\theta_{j,k}$ are of large magnitude. We use the selected $y_{j,k}$ to recover the large $\theta_{j,k}$ and reconstruct a function from the recovered wavelet coefficients as an estimator of f . Two shrinking rules are hard threshold rule

$$\hat{\theta}_{j,k} = \delta_h(y_{j,k}, \lambda) = y_{j,k} 1_{\{|y_{j,k}| > \lambda\}},$$

and soft threshold rule

$$\hat{\theta}_{j,k} = \delta_s(y_{j,k}, \lambda) = \text{sign}(y_{j,k}) (|y_{j,k}| - \lambda)_+,$$

where λ is threshold. The wavelet estimator \hat{f} is the function constructed using $\hat{\theta}_{j,k}$ as wavelet coefficients.

Various ways are proposed to select threshold λ . Universal threshold is defined by

$$\lambda = \sigma \sqrt{2 \log n},$$

which is used to shrink $y_{j,k}$ at all levels. It is derived from the fact that with probability tending to one, the maximum of $|\varepsilon_{j,k}|$ is bounded by $\sigma \sqrt{2 \log n}$.

SureShrink is to choose a threshold for $y_{j,k}$ at each level. SureShrink threshold λ_j for level j is the value that minimizes $SURE(y_{j,k}, \lambda_j)$ over all

λ_j , where

$$SURE(y_{j,k}, \lambda_j) = 2^j - 2 \sum_{k=0}^{2^j-1} 1_{\{|y_{j,k}| \leq \lambda_j\}} + \sum_{k=0}^{2^j-1} (|y_{j,k}| \wedge \lambda_j)^2.$$

It is based on the fact that $SURE(y_{j,k}, \lambda_j)$ is Stein's unbiased estimator of the ℓ_2 -risk of the wavelet estimator constructed using $\delta_s(y_{j,k}, \lambda_j)$, shrunk $y_{j,k}$ with soft threshold rule and threshold λ_j .

Cross-validation threshold is to treat threshold λ as a tuning parameter and apply classic cross-validation procedure to select λ . First divide data into two parts, one part with odd indices and one part with even indices; second use one part of the data to form a wavelet estimator with threshold λ and use the other part of the data to validate the prediction error of the wavelet estimator. The cross-validation threshold is defined to be the value that minimizes the prediction error with correction of multiplying a factor $(1 - \log 2 / \log n)^{-1/2}$. See Nason (1996).

FDR threshold (threshold based on false discovery rate) was investigated by Abramovich and Benjamini (1995) and Abramovich, Benjamini, Donoho and Johnstone (2005). It works as follows. Order $y_{j,k}$ in terms of their absolute values from the largest one to the smallest one and denote the ordered values by

$$|y|_{(1)} \geq |y|_{(2)} \geq \cdots \geq |y|_{(i)} \geq \cdots \geq |y|_{(n)},$$

and compute normal quantiles

$$\hat{t}_i = \sigma \Phi^{-1} \left(1 - \frac{iq}{2n} \right),$$

where q is a small positive number that controls the false discovery rate. Compare $|y|_{(i)}$ to \hat{t}_i and let κ be the largest index i for which $|y|_{(i)} \geq \hat{t}_i$. FDR threshold is taken to be \hat{t}_κ , and we use it to threshold $y_{j,k}$

$$\hat{\theta}_{j,k} = y_{j,k} 1_{\{|y_{j,k}| \geq \hat{t}_\kappa\}}.$$

Other thresholding methods include block thresholding rules, which shrink wavelet coefficients in groups rather than individually, and make simultaneous decisions to retain or to discard all coefficients within a block. See Cai (1999, 2002) and Hall, Kerkycharian and Picard (1998).

Above wavelet shrinkage methods are for normal random errors. It is important but not straightforward to extend the methods to non-normal data. See Antoniadis and Sapatinas (2001), Donoho and Jin (2005), and Kolaczyk and Nowak (2004, 2005) for some recent development on wavelet shrinkage for non-normal random errors.

3.2. Bayesian wavelet shrinkage

Because Bayes solutions for normal type problems have similar desirable property of wavelet shrinkage: heavily shrink small arguments and only slightly shrink the large arguments, Bayesian wavelet shrinkage was proposed as a competitor. By choosing appropriate prior and loss function, the resulting Bayesian wavelet shrinkage can be very close to thresholding, or even match hard- and soft- thresholding rules. Here we mainly review three approaches.

The first one is adaptive Bayesian wavelet shrinkage proposed by Chipman, Kolaczyk and McCulloch (1997). Assume the variance σ^2 of ε_i is known. From (2) the distribution of wavelet coefficients $y_{j,k}$ given $\theta_{j,k}$ is $N(\theta_{j,k}, \sigma^2)$. The prior on $\theta_{j,k}$ is a mixture of two normals

$$\theta_{j,k}|\gamma_j \sim \gamma_j N(0, \tau_j^2) + (1 - \gamma_j) N(0, \tau_j^2/c_j^2),$$

where γ_j are independent Bernoulli random variables with $P(\gamma_j = 1) = p_j$, and c_j, τ_j, p_j are hyperparameters. As p_j are taken to be very small, and c_j are chosen to be much bigger than 1, the two normals in the prior of $\theta_{j,k}$ are selected to capture the sparse representation of $f(x)$ by wavelets: a small number of large wavelet coefficients modeled by $\gamma_j N(0, \tau_j^2)$ and a large number of small coefficients described by $(1 - \gamma_j) N(0, \tau_j^2/c_j^2)$.

Under the squared error loss the posterior mean of $\theta_{j,k}$ is

$$\hat{\theta}_{j,k} = y_{j,k} \left(\frac{\tau_j^2}{\sigma^2 + \tau_j^2} P(\gamma_j = 1|y_{j,k}) + \frac{\tau_j^2}{c_j^2 \sigma^2 + \tau_j^2} P(\gamma_j = 0|y_{j,k}) \right),$$

where

$$P(\gamma_j = 1|y_{j,k}) = \frac{p_j \pi(y_{j,k}|\gamma_j = 1)}{p_j \pi(y_{j,k}|\gamma_j = 1) + (1 - p_j) \pi(y_{j,k}|\gamma_j = 0)},$$

$$\pi(y_{j,k}|\gamma_j = 1) \sim N(0, \sigma^2 + \tau_j^2), \quad \pi(y_{j,k}|\gamma_j = 0) \sim N(0, \sigma^2 + \tau_j^2/c_j^2).$$

An empirical Bayes method was proposed to tune the hyperparameters. Since the hyperparameters are level dependent, as a shrinker $\hat{\theta}_{j,k}$ is a smooth interpolation between two lines through origin with slopes $\tau_j^2/(\sigma^2 + \tau_j^2)$ and $\tau_j^2/(c_j^2 \sigma^2 + \tau_j^2)$, and it shrinks $y_{j,k}$ differently across levels. Thus, the Bayesian approach is an adaptive shrinker.

The second method is a full Bayesian approach proposed by Clyde, Parmigiani and Vidakovic (1998). They treat both $\theta_{j,k}$ and σ^2 unknown. Put an inverse chisquare distribution as a prior for σ^2 , and select the prior of $\theta_{j,k}$ as a mixture of a normal distribution and a point mass at zero, which

capture the strategy of keeping a small number of large wavelet coefficients and excluding a large number of small wavelet coefficients, respectively. That is,

$$\theta_{j,k}|\gamma_j, \sigma^2 \sim \gamma_j N(0, \tau_j^2) + (1 - \gamma_j) 1_{\{0\}},$$

$$\gamma_j \sim \text{Bin}(1, p_j), \quad \eta \nu / \sigma^2 \sim \chi_\nu^2,$$

where $c_j, \tau_j, p_j, \eta, \nu$ are hyperparameters. The standard Bayesian mechanism yields the posterior means of $\theta_{j,k}$ given $y_{j,k}$, which lack of explicit form, and Markov Chain Monte Carlo method can be used to numerically evaluate the posterior means.

The third approach is to use weighted absolute error loss studied by Abramovich, Sapatinas and Silverman (1998). The prior of $\theta_{j,k}$ is a mixture of a normal and a point mass at zero,

$$\theta_{j,k} \sim \gamma_j N(0, \tau_j^2) + (1 - \gamma_j) 1_{\{0\}},$$

where

$$\tau_j^2 = C_1 2^{-\alpha j}, \quad \gamma_j = \min(1, C_2 2^{-\beta j}),$$

and C_1, C_2, α, β are hyperparameters. The parameters α and β govern the stochastically decaying speed of $\theta_{j,k}$ and thus control the smoothness of the underlying function $f(x)$. Abramovich, Sapatinas and Silverman (1998) showed that for known σ^2 , the Bayesian solution is the posterior median which has the form

$$\text{median}(\theta_{j,k}|y_{j,k}) = \text{sign}(y_{j,k}) \max(0, \zeta_{j,k}),$$

where

$$\zeta_{j,k} = |y_{j,k}| \frac{\tau_j^2}{\sigma^2 + \tau_j^2} - \frac{\tau_j \sigma}{\sqrt{\sigma^2 + \tau_j^2}} \Phi^{-1} \left(\frac{1 + \min(\omega_{j,k}, 1)}{2} \right),$$

$$\omega_{j,k} = \frac{1 - \gamma_j}{\gamma_j} \frac{\sqrt{\sigma^2 + \tau_j^2}}{\sigma} \exp \left\{ -\frac{\tau_j^2 y_{j,k}^2}{2 \sigma^2 (\sigma^2 + \tau_j^2)} \right\},$$

and Φ^{-1} is the inverse of the standard normal distribution function.

Empirical Bayesian approach was investigated by Clyde and George (2000) and Johnstone and Silverman (2005).

3.3. *Rough function estimation*

In scientific studies objects may be very far from smooth or piecewise smooth. Mathematically, fractal functions are used to model these rough objects, and fractal dimension is defined to quantitatively characterize these functions and measure their roughness. For the problem of estimating a fractal function based on noisy data, it is very critical to preserve its fractal dimension. Smooth or piecewise smooth functions can't well approximate a fractal function, and smoothing noisy data substantially reduces its fractal dimension and degrades its quality, so traditional smoothing methods may be inappropriate for fractal function estimation. Wang (1997a) studied estimation of fractal functions by wavelet shrinkage and showed that as sample size tends to infinity, the fractal dimension of the wavelet estimator over the observed resolution levels converges in probability to the fractal dimension of the underlying function. Therefore, for large samples, wavelet shrinkage can remove most of the noise and produce a wavelet estimate with fractal dimension well preserved. Moreover, the wavelet estimator can reveal that some parts of the function are more or less rougher than others and indicate varying local dimension for the underlying function.

3.4. *Dependent data*

So far we consider nonparametric regression with i.i.d. data. However, in many applications, data may be dependent. In fact, correlations between observations that are far apart may decay to zero at a slower rate than we would expect from independent data or short-range dependent data. Slow decay in correlation is often referred to as long-range dependence or long-memory. An alternative definition of long-range dependence for a stationary process is that its spectral density has a pole at zero.

Consider the problem of estimating $f(x)$ based on data from the nonparametric regression model (1), where random errors $\varepsilon_1, \dots, \varepsilon_n$ are zero-mean stationary process with correlation

$$\text{corr}(\varepsilon_i, \varepsilon_j) \asymp |i - j|^{-\alpha}, \quad |i - j| \rightarrow \infty, \quad \alpha \in (0, 1).$$

As functional Gaussian noise and fractional Brownian motion are often used to model phenomena exhibiting long-range dependence, Wang (1996) proposed the following fractional Gaussian noise model to approximate nonparametric regression with long-range dependent Gaussian errors.

Process $Y(x)$, $x \in [0, 1]$, is observed from the fractional Gaussian noise

model

$$Y(dx) = f(x) dx + \epsilon^{2-2H} B_H(dx), \quad (3)$$

where $f(x)$ is an unknown function, ϵ is the noise level, and $B_H(dx)$ is a fractional Gaussian noise, which is a formal derivative of a standard fractional Brownian motion $B_H(x)$, $H \in (0, 1)$.

Wang (1996) developed the decorrelation of a class of Gaussian processes including fractional Gaussian noise and fractional Brownian motion by using the idea of simultaneous diagonalization through WVD and the fact that the Gaussian processes are linked to dilation-homogeneous operators. The WVD simultaneously decorrelates fractional Gaussian noise and fractional Brownian motion and achieves a quasi-decorrelation of Brownian motion and white noise. Such simultaneous decorrelations are used to establish asymptotic minimax risks for the fractional Gaussian noise model and gave explicit forms for their convergence rates. In order to achieve minimax rates, the wavelet shrinkage estimators are required to use level dependent thresholds. Johnstone and Silverman (1997), Johnstone (1999) and von Sachs and Macgibbon (2000) investigated wavelet shrinkage for both short-range and long-range dependent data.

4. Inverse problems

Suppose we wish to recover a function $f(x)$ but observe only a transformation, $(Kf)(x)$, of the underlying function, where K is a linear transformation such as Abel Transform, Convolution transform, or Radon transform. In these problems, K is often non-invertible, in the sense that no inverse of K exists as a bounded linear operator, and we call them ill-posed inverse problems. Inverse problems are very challenging and difficult. The inverse problem in the presence of white noise is defined as follows,

$$y_i = (Kf)(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4)$$

where $x_i = i/n$, ε_i are i.i.d. normal errors, K is a transformation, and $f(x)$ is a function. Our goal is to estimate $f(x)$ based on data y_1, \dots, y_n .

The singular value decomposition (SVD) is widely used to solve inverse problems. As a wavelet analogue of SVD, Donoho (1995) created the WVD to solve inverse problems in the presence of white noise. WVD consists of three sets of basis functions, an orthogonal wavelet basis $(\psi_{j,k})$ and two near-orthogonal vaguelette bases $(u_{j,k})$ and $(v_{j,k})$. The three sets of basis functions are linked together as follows. The bases $(u_{j,k})$ and $(v_{j,k})$ are mutually biorthogonal, K transforms $\psi_{j,k}$ into $\kappa_j u_{j,k}$, and K^* transforms

$(v_{j,k})$ back to $\kappa_j \psi_{j,k}$, where κ_j are quasi-singular values. The WVD has the following reproducing formula

$$f(x) = \sum_{j,k} [Kf, u_{j,k}] \kappa_j^{-1} \psi_{j,k}(x).$$

We can see from above reproducing formula that like SVD, WVD can recover $f(x)$ from observations about $(Kf)(x)$. However, WVD is much better than SVD. It simultaneously represents the operator K in a quasi-diagonal form and effectively represents a large class of functions including functions with spatial variabilities such as jumps. Because of their localization property, the wavelets and vaguelettes in a WVD are much more effective in dealing with local behaviors of $f(x)$ than eigenfunctions in a SVD. The WVD based estimator of $f(x)$ is constructed as follows. First use data y_i to compute empirical vaguelette coefficients $[(y_i), u_{j,k}]$; second shrink the empirical vaguelette coefficients with level dependent threshold λ_j to obtain $\delta([(y_i), u_{j,k}], \lambda_j)$; third reconstruct $f(x)$ by the reproducing formula with $[Kf, u_{j,k}]$ replaced by $\delta([(y_i), u_{j,k}], \lambda_j)$. Donoho (1995) showed that the WVD estimator achieves minimax over a wide range of Besov spaces and outperforms linear estimators.

Kolaczyk (1996) investigated the WVD shrinking method and applied it to tomographic image reconstruction. Wang (1997b) investigated function estimation for inverse problems in the presence of long-memory noise by employing two WVDs - one for the inverse problem which simultaneously quasi-diagonalizes both the operator and the prior information and one for long-range dependence which decorrelates long-memory noise - to convert the linear inverse problem into a new inverse problem with white noise. Their minimax risks converge to zero at rates that depend on both transformation and long-range dependence, which differ from those for problems with either direct observations or indirect observations with independence or short-range dependence. It is also of interest to point out that by the use of WVD diagonalization, the two problems, a linear inverse problem and estimation with long-range dependence, can be formally converted from one to another. Abramovich and Silverman (1998) considered a vaguelette-wavelet decomposition for solving inverse problems. However, WVD can't be used to solve inverse problems with scale preferred transformation like boxcar deconvolution. Johnstone, Kerkycharian, Picard and Raimondo (2004) and Johnstone and Raimondo (2004) proposed wavelet methods to solve deconvolution problems.

5. Change-points

5.1. Direct data

Wavelets have a property to “zoom in” on very short lived frequency phenomena, such as transients in signals and singularities in functions, and hence provide a tool to study localized changes. Wang (1995, 1998) proposed wavelet methods to study change-points of functions in one and two dimensions. The methods are first to compute wavelet transformation of the noisy data and then to compare the wavelet coefficients with the estimated threshold. It uses the spatial positions at which the wavelet transformation across fine scale levels exceeds the threshold to detect and locate change-points such as jumps and sharp cusps.

A function $f(x)$ in one dimension is said to have an α -cusp at x_0 for $\alpha \in [0, 1)$ if there exists a positive constant C such that, as h tends to zero from left or right,

$$|f(x_0 + h) - f(x_0)| \geq C|h|^\alpha.$$

$\alpha = 0$ corresponds to that $f(x)$ has a jump at x_0 . If f is smooth at u , $Tf(s, u)$ has the order $s^{3/2}$ as s tends to zero, and if f has a sharp-cusp at u , the maximum of $|T(s, u)|$ over a neighborhood of u with size proportional to the scale s converges to zero at a rate no faster than $s^{\alpha+1/2}$ as s tends to zero.

Suppose we have observations y_1, \dots, y_n from model (1), and $f(x)$ may have sharp cusps and is smooth otherwise. Our goal is to detect if $f(x)$ has sharp cusps, and if it has, estimate the locations of the sharp cusps.

From (2) and the property of the wavelet transformation for a function with sharp cusps, we can see that while at time point $t = k2^{-j}$ where $f(x)$ is smooth, $\theta_{j,k}$ is of order $2^{-3j/2}$, and nearby a sharp cusp of $f(x)$, $2^{(\alpha+j)/2} \theta_{j,k}$ has an absolute value bounded below from zero. Thus, at high resolution levels, $\theta_{j,k}$ dominate $\varepsilon_{j,k}$ nearby where $f(x)$ has sharp cusps, and is negligible in comparison with $\varepsilon_{j,k}$ where $f(x)$ is smooth. This fact implies that, at some high resolution levels j_n with $2^{j_n} \sim n^{1/(2\alpha+1)}/\log^2 n$, nearby sharp cusps of $f(x)$, $y_{j_n,k}$ are dominated by $\theta_{j_n,k}$ whose absolute value is significantly large, and hence significantly larger than the others. Define

$$T_n = \max\{|y_{j_n,k}| : k = 0, 1, \dots, 2^{j_n} - 1\}.$$

Then T_n is of much larger order under $f(x)$ with sharp cusps than under smooth $f(x)$, and thus can be served as a statistic to detect sharp cusps in $f(x)$. To carry out the detection, we need a threshold value for T_n so that if

T_n exceeds the threshold value, we can claim that $f(x)$ has sharp cusps. One choice of threshold is the universal threshold $D_n = \hat{\sigma} \sqrt{2 \log n/n}$, where $\hat{\sigma}$ is median of $|y_{j_n,k}|$ divided by 0.6745. It is based on asymptotic theory of the maximum of Gaussian processes and works well when noise ε_i is Gaussian or has light-tailed.

We estimate the sharp cusps of $f(x)$ by the locations of $|y_{j_n,k}|$ that exceed threshold D_n , that is, if $|y_{j_n,k}| > D_n$ for some k , the corresponding jump location is estimated by $\hat{\tau} = k 2^{-j_n}$.

Wang (1995) and Raimondo (1998) established asymptotic theory for the wavelet method and developed fast algorithms to practically implement the method. They showed that the wavelet estimators of the change-points are optimal or nearly optimal. Wang (1998) proposed the wavelet method for change curves in two dimensions and applied the method to edge detection in image processing. Raimondo and Tajvidi (2003) considered the change-points problem when the random errors ε_i follow a heavy-tail distribution. They designed sharp cusp detection by employing a peak-over-threshold method to model the wavelet coefficients of the heavy-tail noise and using the generalized Pareto distribution to approximate the exceedances of the wavelet coefficients.

5.2. Indirect data

SVD and traditional change-point methods have great difficulty in dealing with change-points for indirect observations. Tools to detect and locate change-points for indirect observations require two properties. They must extract information about $f(x)$ from indirect observations, and they must characterize the local features of $f(x)$. Eigenfunctions in a SVD, like the Fourier basis, often have trouble in focusing on the local behavior of $f(x)$, while conventional detection techniques based on smoothing can't recover information about $f(x)$ from indirect observation. In contrast to the huge amount of literature on change-points for direct data, there has been little study of change-points for indirect data.

WVD can not only extract information about $f(x)$ from its indirect observations but also characterize localized features of $f(x)$ near a point. It is very suitable for studying change-points for indirect data. Utilizing this special property of WVD, Wang (1999) proposed a method to detect and estimate change-points of the underlying functions $f(x)$ based on observations about $(Kf)(x)$ from model (4). The method uses a WVD to extract the information about the wavelet transformation of $f(x)$ from in-

direct data. Once we have the wavelet transformation of $f(x)$, similar to wavelet methods for direct data, we can detect and estimate change-points by the wavelet transformation across fine scale levels. Asymptotic theory for the detection and estimation was established.

6. Local self-similarity and non-stationary stochastic process

Many natural phenomena exhibit some sort of self-similarity, and scientists have applied self-similarity models to many areas including image processing (fractal image compression and segmentation), dynamic systems (turbulence), and biology and medicine (physiological time series). Although self-similar stochastic processes were first introduced in a theoretical context by Kolmogorov in 1941, statisticians were made aware of the practical applicability of such processes through the work of B.B. Mandelbrot (Mandelbrot and van Ness, 1968)

We say a stochastic process $Y(t)$ is a self-similar process with self-similarity parameter H if for any positive stretching factor c , the distribution of the rescaled and reindexed process $c^{-H} Y(ct)$ is the same as that of the original process $Y(t)$. That is, for any sequence of time points t_1, \dots, t_n and any positive constant c , the collections $\{c^{-H} Y(ct_1), \dots, c^{-H} Y(ct_n)\}$ and $\{Y(t_1), \dots, Y(t_n)\}$ have the same probability distribution. As a consequence, the sample path of $Y(t)$ has the qualitative features that are invariant to magnification or shrinkage, so that stochastically any sections of the sample path have the same general appearance regardless of the locations and length of the sections.

The value of the self-similarity parameter or scaling exponent H dictates the dynamic behavior of a self-similar process $Y(t)$. If $Y(t)$ has finite second moments, and its associated increment process $X(t) = Y(t) - Y(t-1)$ is stationary, then H is assumed to be between zero and one, and the value of H is used to describe the autocorrelation of $X(t)$. For $H > 1/2$, $X(t)$ has slowly decaying autocorrelation which is often referred to as long-range dependence or long memory, and $X(t)$ is widely used to model phenomena with power-law correlation and long-memory. For $H < 1/2$, $X(t)$ is characterized by serial correlations that decay rapidly and sum to zero. $H = 1/2$ corresponds to that $X(t)$ is serially uncorrelated. The estimation of H as a constant has been extensively studied, predominantly in the context of long memory where it is assumed that $H \in (1/2, 1)$.

Self-similarity models with constant self-similarity parameter H assume

that the self-similar features of the underlying phenomenon persist over time. However, many phenomena often display self-similarity only on certain time scale, and/or have self-similarity in different time periods with different self-similarity parameters. That is, self-similar patterns are local and change as the phenomenon itself evolves. To adequately model such phenomena, we need to introduce non-stationary self-similar processes by allowing the scaling exponent to vary as a function of time, and develop a statistical procedure to characterize the exponent's progression. Mallat, Papanicolaou and Zhang (1998) used local cosine bases to model a class of locally stationary processes with approximately convolution covariance operators. With the efficient modeling of covariance operators by local cosine bases, Donoho, Mallat and von Sachs (1998) investigated estimation of the covariance of locally stationary processes. Neuman and von Sachs (1997) and Nason, von Sachs and Kroisandt (2000) considered adaptive wavelet estimation of the evolutionary spectra. Nason and Sapatinas (2002) used wavelet packets to model non-stationary time series. Gançlaves and Flandrin (1993) and Flandrin and Gançlaves (1994) proposed locally self-similar processes whose scaling exponents vary with time and discussed their applications.

Cavanaugh, Wang and Davis (2003) introduced the following two locally similar processes. The first example is defined by the stochastic integral

$$Y(t) = \int_{-\infty}^0 \left\{ (t-u)^{H(t)-1/2} - (-u)^{H(t)-1/2} \right\} dB_u + \int_0^t (t-u)^{H(t)-1/2} dB_u, \quad (5)$$

where B_u is standard Brownian motion, and $H(t) \in (0, 1)$ represents the scaling function. As $Y(t)$ is an extension of fractional Brownian motion and allows its self-similarity parameter to vary over time, we call it *generalized fractional Brownian motion*, gfBm. For gfBm, Cavanaugh, Wang and Davis (2003) provided an explicit forms for its covariance function and showed that gfBm can be locally approximated by fBm, and its sample path has fractal dimension $2 - \min\{H(t) : 0 \leq t \leq 1\}$ and local dimension $2 - H(t_0)$ at a given time point t_0 . The increment process of fBm is stationary, while gfBm has a non-stationary increment (unless $H(t)$ is constant) and if $H(t)$ is smooth, the increment process of gfBm is locally stationary.

The second example is defined through difference equation

$$\Phi(B)(1-B)^{H(t)-1/2}X(t) = \Theta(B)\epsilon(t), \quad (6)$$

where B is a backshift operator given by $BX(t) = X(t-1)$, $\Phi(B)$ and $\Theta(B)$ are polynomials in B with roots outside the unit circle, $\epsilon(t)$ is Gaussian

white noise, and $H(t) \in (0, 1)$ is the scaling function. Similar to gFBM, $X(t)$ is an extension of a fractional autoregressive integrated moving-average process with self-similarity parameter allowing to evolve over time, we thus refer it to as a *generalized fARIMA* or gFARIMA process. A special case of (6) is defined by

$$(1 - B)^{H(t)-1/2} X(t) = \epsilon(t). \quad (7)$$

We refer it to as *generalized fractionally integrated noise*, since it is an extension of fractionally integrated noise.

For locally self-similar process, its scaling function often carries important, even decisive, information about the behavior of the process, so it is desirable to develop statistical inference based on sample observations of the process. Wang, Cavanaugh and Song (2001) and Cavanaugh, Wang and Davis (2003) proposed a procedure based on wavelets for constructing an estimator of the time-varying scaling exponent of a locally self-similar process. They established an approximate local log-linear relationship between the square of the wavelet transformation of $Y(t)$ and the scale for the transformation and used local least-squares regression to estimate $H(t)$. Specifically, let ψ be a mother wavelet and

$$TY(a, t) = a^{-1/2} \int \psi\left(\frac{u-t}{a}\right) Y(u) du = a^{1/2} \int \psi(x) Y(t+ax) dx$$

be wavelet transformation of $Y(t)$. Then

$$E \{|TY(a, t)|^2\} \asymp C_1 a^{1+2H(t)}, \quad \text{as } a \rightarrow 0.$$

Set

$$y_t(a) = \log \{|TY(a, t)|^2\},$$

$$C_2 = E \left(\log \left[|TY(a, t)|^2 / E \{|TY(a, t)|^2\} \right] \right),$$

$$\varepsilon_t(a) = \log \left[|TY(a, t)|^2 / E \{|TY(a, t)|^2\} \right] - C_2.$$

Then we have an approximate regression model

$$y_t(a) \approx c + \{2H(t) + 1\} \log a + \varepsilon_t(a), \quad \text{for small scale } a, \quad (8)$$

with $c = \log C_1 + C_2$. The least squares estimator of $H(t)$ is given as follows.

1. Select a sequence of small scales $a_1 > \dots > a_k$, say $a_j = 2^{-j}$ where $j = 1, \dots, k$.

2. Define a set of bivariate data (x_j, y_j) , $j = 1, \dots, k$, by setting

$$x_j = \log a_j \quad \text{and} \quad y_j = y_t(a_j) \quad \text{for each } j.$$

3. Evaluate the least-squares estimate of $H(t)$ in (8) via

$$\hat{H}(t) = \left\{ \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{\sum (x_j - \bar{x})^2} - 1 \right\} / 2,$$

$$\text{where } \bar{x} = \sum x_j / k, \bar{y} = \sum y_j / k.$$

Consistency of $\hat{H}(t)$ was established and extensive simulations were conducted to check the effectiveness of the procedure. The method was applied to several practical examples. However, asymptotic distribution is not available for $\hat{H}(t)$, although simulations indicate that $\hat{H}(t)$ is asymptotically normal. See Craigmile and Percival (2005), Katul, Vidakovic and Albertson (2001) and Vidakovic, Katul and Albertson (2000) for related work.

7. Beyond wavelets

Wavelets are designed to handle point singularities. Functions in one dimension have only point singularities but complex singularity structures may occur in high dimensional functions. For example, a function in three dimensions may be singular along curves or surfaces. Donoho and his collaborators have created high dimensional bases like ridgelets, curvelets and beamlets to better represent complex functions in high dimensions and use them to find some hidden structures from noisy high dimensional data. See Donoho (2000) and other related papers listed at <http://www-stat.stanford.edu/~donoho>.

Acknowledgments

This work was supported by the NSF grant DMS-0504323.

References

1. F. Abramovich and Y. Benjamini (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In *Wavelets and Statistics* (eds. Antoniadis and Oppenheim), 5-14. New York: Springer-Verlag.
2. F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone (2005). Adapting to Unknown Sparsity by Controlling the False Discovery Rate. To appear in *Ann. Statist.*
3. F. Abramovich, T. Sapatinas, and B. W. Silverman (1998). Wavelet thresholding via a Bayesian approach. *J. Roy. Statist. Soc. B* **60**, 725-749.

4. F. Abramovich and B. W. Silverman (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85**, 115-129.
5. C. Angelini and B. Vidakovic (2004). Gamma-minimax wavelet shrinkage: a robust incorporation of information about energy of a signal in denoising applications. *Statistica Sinica* **14**, 103-125.
6. A. Antoniadis and P. Fryzlewicz (2005). Parametric modeling of threshold across scales in wavelet regression. Manuscript.
7. A. Antoniadis and T. Sapatinas (2001). Wavelet shrinkage for natural exponential families with quadratic variance functions. *Biometrika* **88**, 805-820.
8. S. Barber and G. P. Nason (2004). Real nonparametric regression using complex wavelets. *J. R. Statist. Soc. Ser. B* **66**, 927-939.
9. S. Barber, G. P. Nason and B. W. Silverman (2002). Posterior probability intervals for wavelet thresholding. *J. R. Statist. Soc. Ser. B* **64**, 189-206.
10. T. Cai (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 898-924.
11. T. Cai (2002). On block thresholding in wavelet regression: Adaptivity, block size, and threshold level. *Statistica Sinica* **12**, 1241-1273.
12. J. E. Cavanaugh, Y. Wang and W. Davis (2003). Local self-similar processes and their wavelet analysis, in *Handbook of Statistics, Volume 21: Stochastic Processes: Modeling and Simulation* (eds. Shanbhag and Rao), Elsevier Science, 93-135.
13. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch (1997). Adaptive Bayesian wavelet shrinkage. *J. Amer. Statist. Assoc.* **92**, 1413-1421.
14. M. Clyde and E. George (2000). Flexible empirical Bayes estimation for wavelets. *J. Roy. Statist. Soc. B* **62**, 681-698.
15. M. Clyde, G. Parmigiani, and B. Vidakovic (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika* **85**, 391-401.
16. P. F. Craigmile and D. B. Percival (2005). Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Transactions on Information Theory* **51**, 1039-1048.
17. I. Daubechies (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41**, 909-996.
18. I. Daubechies (1992). Ten Lectures on Wavelets. CBMS-NSF series in applied mathematics, no. 61, SIAM, Philadelphia.
19. D. L. Donoho (1995). Nonlinear solutions of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harm. Anal.* **2**, 101-126.
20. D. L. Donoho (2000). Ten Lectures on Beyond Wavelets. Available from <http://www-stat.standord.edu/~donoho>
21. D. L. Donoho and J. Jin (2005). Asymptotic minimaxity of FDR for sparse exponential model. To appear in *Ann. Statist.*
22. D. L. Donoho and I. M. Johnstone (1994). Ideal spatial adaptation by wavelets shrinkage. *Biometrika* **81**, 425-455.
23. D. L. Donoho and I. M. Johnstone (1995). Adapting to unknown smoothness via wavelets shrinking. *J. Amer. Statist. Assoc.* **90**, 1200-1224.
24. D. L. Donoho and I. M. Johnstone (1998). Minimax estimation via Wavelets shrinkage. *Ann. Statist.* **26**, 879-921.

25. D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard (1995). Wavelet shrinkage: Asymptopia ? (with discussion). *J. Roy. Statist. Soc. B* **57** 301-369.
26. D. L. Donoho, S. G. Mallat and R. von Sachs (1998). Estimating covariances of locally stationary processes: convergence of best-basis methods. Technical report, Stanford University.
27. J. Fan (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91**, 674-688.
28. P. Flandrin and P. Gançlaves (1994). From wavelets to time-scale energy distributions. In *Advances in Wavelet Analysis* (eds. Schumaker and Webb), Academic Press, New York, 309-334.
29. P. Gançlaves and P. Flandrin, P. (1993). Bilinear time-scale analysis applied to local scaling exponents estimation. In *Progress in Wavelet Analysis and Applications* (eds. Meyer and Roques), Frontieres, Paris, 271-276.
30. P. Hall and P. Patil (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non-linear wavelet methods. *J. Roy. Statist. Soc. B* **58**, 361-378.
31. P. Hall, G. Kerkycharian and D. Picard. (1998). Block threshold rules for curve estimation using kernel and wavelet method. *Ann. Statist.* **26**, 922-942.
32. I. M. Johnstone (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statistica Sinica* **9**, 51-83.
33. I. M. Johnstone and B. W. Silverman (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. B* **59**, 319-351.
34. I. M. Johnstone and B. W. Silverman (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.* **33**, 1700-1752.
35. I. M. Johnstone, G. Kerkycharian, D. Picard and M. Raimondo (2004). Wavelet deconvolution in a periodic setting (with discussion). *J. Roy. Statist. Soc. B* **66**, 547-573.
36. I. M. Johnstone and M. Raimondo (2004). Periodic boxcar deconvolution and diophantine approximation. *Ann. Statist.* **32**, 1781-1804.
37. G. Katul, B. Vidakovic and J. Albertson (2001). Estimating global and local scaling exponents in turbulent flows using wavelet transformations. *Physics of Fluids* **13**, 241-250.
38. E. D. Kolaczyk (1996). A wavelet shrinkage approach to tomographic image reconstruction. *J. Amer. Statist. Assoc.* **91**, 1079-1090.
39. E. D. Kolaczyk, J. Ju and S. Gopal (2005). Multiscale, multigranular statistical image segmentation. To appear in *J. Amer. Statist. Assoc.*
40. E. D. Kolaczyk and R. D. Nowak (2004). Multiscale likelihood analysis and complexity penalized estimation. *Ann. Statist.* **32**, 500-527.
41. E. D. Kolaczyk and R. D. Nowak (2005). Multiscale generalized linear models for nonparametric function estimation. To appear in *Biometrika*.
42. S. G. Mallat (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intel.* **11**, 674-693.
43. S. G. Mallat (1999). A Wavelet Tour of Signal Processing. Academic Press, San Diego, CA. Second Edition.

44. S. G. Mallat, G. Papanicolaou and S. Zhong, S. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26**, 1-47.
45. B. B. Mandelbrot and J. W. van Ness (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422-437.
46. G. P. Nason (1996). Wavelet shrinkage using cross-validation. *J. Roy. Statist. Soc. B* **58**, 463-479.
47. G. P. Nason and T. Sapatinas (2002). Wavelet packet transfer function modelling of non-stationary time series. *Statist. Comput.* **12**, 45-56.
48. G. P. Nason, R. von Sachs and G. Kroisandt (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc. Ser. B* **62**, 271-292.
49. M. Neumann and R. von Sachs (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. To appear in *Ann. Statist.* **25**, 38-76.
50. D. P. Percival and P. Guttorp (1994). Long-memory processes, the Allen variance and wavelets. In *Wavelets in Geophysics*(eds. Foufoula-Georgiou and Kumar), New York, Academic Press, 325-357.
51. M. Raimondo (1998). Minimax estimation of sharp change points. *Ann. Statist.* **26**, 1379-1397.
52. M. Raimondo and N. Tajvidi (2003). A peak over threshold model for change-point detection by wavelets. *Statistica Sinica* **14**, 395-412.
53. M. G. Tadesse, J. G. Ibrahim, M. Vannucci and R. Gentleman (2005). Wavelet thresholding with Bayesian false discovery rate control. *Biometrics* **61**, 25-35.
54. B. Vidakovic (1999). Statistical Modeling by Wavelets. Wiley, New York.
55. B. Vidakovic, G. Katul and J. Albertson (2000). Multiscale denoising of self-similar process. *J. Geophys. Res.-Atmos.* **105**, 27049-27058.
56. von Sachs, R. and Macgibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist.* **27**, 475-499.
57. Y. Wang (1995). Jump and sharp cusp detection via wavelets. *Biometrika* **82**, 385-397.
58. Y. Wang (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24**, 466-484.
59. Y. Wang (1997a). Fractal function estimation via wavelet shrinkage. *J. Roy. Statist. Soc. B* **59**, 603-613.
60. Y. Wang (1997b). Minimax estimation via wavelets for indirect long-memory data. *J. Statist. Plan. Infer.* **64**, 45-55.
61. Y. Wang (1998). Change curve estimation via wavelets. *J. Amer. Statist. Assoc.* **93**, 163-172.
62. Y. Wang (1999). Change-point analysis via wavelets for indirect data. *Statistica Sinica* **9**, 103-117.
63. Y. Wang, J. E. Cavanaugh and C. Song (2001). Self-similarity index estimation via wavelets for locally self-similar processes. *J. Statist. Plan. Infer.* **99**, 91-110.