Variational method and Wavelet Regression

Jianzhong Wang
Department of Mathematics and Statistics
Sam Houston State University
Huntsville, TX 77341
e-mail: mth_jxw@shsu.edu

Abstract

Wavelet regression is a new nonparametric regression approach. Compared with traditional methods, wavelet regression has the advantages of ideal minimax property, spatial inhomogeneous adaptivity, high dimension expansibility and fast algorithm. In this paper, we provide a brief review of wavelet shrinkage. We also apply the variational method for wavelet regression and show the relation between the variational method and wavelet shrinkage.

Keywords. Wavelets, Orthonormal wavelets, Wavelet regression, Variational Methods.

1 Introduction

Regression analysis is an important statistical technique for investigating and modeling the relationship between variables. The goal of regression is to get a model of the relationship between one variable $y$ and one or more variables $t$. Let $y = \hat{f}(t)$ be the observation (also called the observed function), which responds to the regressor $t$:

$$y_i := \hat{f}(t_i) = f(t_i) + \varepsilon_i \quad i = 1, 2, \ldots, N$$

where $t_i, i = 1, 2, \ldots, N$, are sampling points, $f(t_i)$ are values of an unknown function $f(t)$ (called the underlying function), and $\varepsilon_i$ are the statistical errors (called noise), which are often formulated as independent random variables. In this paper, we assume the sampling points are equal-spaced between 0 and 1.

The goal of a regression is to estimate the underlying function $f(t)$ from the sampling data $(t_i, y_i)_{i=1}^N$ (called the observed data), i.e., to establish a method to find the estimator $\hat{f}$, which achieves the minimal risk

$$R(\hat{f}, f) = N^{-1}E||\hat{f} - f||^2_{L_2,N}.$$  

The classical nonparametric regression methods mainly include orthogonal series estimators, kernel estimators and smoothing splines. Theoretical discussions about these methods can be found in [17] and [35]. Recently, a variety
of spatially adaptive methods has been developed in the statistical literature. Among them, wavelet adaptive methods have received great attention. References are given in Bock and Pliego [5], Vidakovic [36], Ogden [33], Donoho and Johnstone [13], [14], [15], [16], and their references. In wavelet regression, an important issue is to select a subset of wavelet coefficients, which well represent the underlying function \( f \) while removing most of noise.

Since 1980s, developing numerical methods for removing noise from an image while preserving edge also became an active research area. Mumford and Shah [28] proposed an energy functional for images so that the image processing such as noise removal and image segmentation can be formulated as a variational problem associated with the functional. Perona and Malik [34] proposed an anisotropic diffusion model for removing noise while enhancing edge. It is proven that the anisotropic diffusion equation in [34] was the steepest decent method for solving the variational problem associated with a certain energy functional [38]. References in this aspect can be found in the book [3].

The goal of this paper is to apply the variational method in the wavelet regression. The main idea is the following. Since the underlying function \( f \) is unknown, we cannot obtain the formula for \( R(\hat{f}, f) \). Hence to directly minimize \( R(\hat{f}, f) \) is impossible. We shall create a substitution of the risk functional, say \( GR(\hat{f}, \tilde{f}) \), in which the underlying function is not involved. Thus, we can use variational method to find \( \hat{f} \). The wavelet representation of data will enable us to establish the variational problems with a very simple structure. We outline the paper as follows. In Section 2, we briefly introduce orthonormal wavelet bases and adaptive wavelet regression (also called wavelet shrinkage). In Section 3, we describe the variational model for wavelet regression and establish the relation between it and wavelet shrinkages. In Section 4, we give some examples.

2 Preliminarily

2.1 Orthonormal Wavelet Bases

The regular orthonormal wavelet basis of \( L^2(R) \) is constructed from a multiresolution analysis (MRA) (see [24], [25], [26], [22], and [23]).

**Definition 1** A multiresolution analysis of \( L^2(R) \) is a sequence \( (V_j)_{j \in \mathbb{Z}} \) of closed subspaces of \( L^2(R) \) such that the followings hold:

1. \( V_j \subset V_{j+1} \quad j \in \mathbb{Z} \)
2. \( \bigcup_{j \in \mathbb{Z}} V_j = L^2(R) \) and \( \bigcap_{j \in \mathbb{Z}} V_j = \{0\} \)
3. \( f(x) \in V_j \iff f(2x) \in V_{j+1} \)
4. \( f(x) \in V_j \iff f(x - 2^{-j}k) \in V_j \quad j, k \in \mathbb{Z} \)
5. There exists a function \( \phi \in V_0 \) such that \( \{\phi(x-n)\}_{n \in \mathbb{Z}} \) forms an unconditional basis of \( V_0 \), i.e., \( \{\phi(x-n)\}_{n \in \mathbb{Z}} \) is a basis of \( V_0 \) and there exist two constants, \( A, B > 0 \) such that, \( \forall (c_n) \in l^2 \), the following inequality holds

\[
A \sum |c_n|^2 \leq \| \sum c_n \phi(\cdot-n) \|^2 \leq B \sum |c_n|^2.
\]
The function $\phi$ in Definition 1 is called an MRA generator. Furthermore, if \(\{\phi(x-n)\}_{n \in \mathbb{Z}}\) is an orthonormal basis of \(V_0\), then $\phi$ is called an orthonormal MRA generator (also called an orthonormal scaling function). Assume $\phi$ satisfies the two-scale equation

$$
\phi(x) = 2 \sum_{k \in \mathbb{Z}} h(k) \phi(2x - k), \quad (h(k))_{k \in \mathbb{Z}} \in l^2.
$$

Define the wavelet function $\psi$ by

$$
\psi(x) = 2 \sum_{k \in \mathbb{Z}} (-1)^k h(1 - k) \phi(2x - k).
$$

For a function $f$, we write

$$
f_{j,k}(x) = 2^j f(2^j x - k), \quad j, k \in \mathbb{Z}.
$$

Then \(\{\phi_{j,k}\}_{k \in \mathbb{Z}}\) forms an orthonormal basis of \(V_j\). Let \(W_j \subset L^2(\mathbb{R})\) be the subspace spanned by \(\{\psi_{j,k}\}_{k \in \mathbb{Z}}\). Then \(W_j\) is called the wavelet subspace at level \(j\). It is clear that \(W_j \perp V_j\) and \(W_j \oplus V_j = V_{j+1}\). Hence, \(\bigoplus_{j=-\infty}^{\infty} W_j = L^2(\mathbb{R})\) and \(\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}\) forms an orthonormal basis of \(L^2(\mathbb{R})\), called a standard orthonormal wavelet basis. Recall that we also have

$$
L^2(\mathbb{R}) = V_{J_0} \oplus \bigoplus_{j \geq J_0} W_j.
$$

Hence, \(\{\phi_{J_0,k}\}_{k \in \mathbb{Z}} \cup \{\psi_{J_0,k}\}_{j \geq J_0, k \in \mathbb{Z}}\) also forms an orthonormal basis of \(L^2(\mathbb{R})\), which is called a hybrid orthonormal wavelet basis. Using this basis we can decompose a function $f \in L^2(\mathbb{R})$ into the following form

$$
f = f_{J_0} + \sum_{j \geq J_0} g_j, \quad f_{J_0} \in V_{J_0}, g_j \in W_j,
$$

$$
f_{J_0} = \sum b_{J_0,k} \phi_{J_0,k}, \quad g_j = \sum w_{j,k} \psi_{j,k}.
$$

Particularly, a function $f \in V_j$ has the decomposition

$$
f = f_{J_0} + \sum_{j = J_0}^{J-1} g_j, \quad f_{J_0} \in V_{J_0}, g_j \in W_j.
$$

Hybrid orthonormal wavelet bases are very useful in many applications because in the decomposition above $f_{J_0}$ provides an estimator of $f$ while $g_j, j = J_0, \ldots, J - 1$, preserve the details of $f$ at different levels (see Figures 1-3). For more discussions of wavelet bases, we refer to [12] and [24].

Daubechies ([11], [12]) created compactly supported orthonormal wavelets bases, which are very useful in wavelet regression. In regression the sample data are finite. In order to deal with finite data, we need the multiresolution analysis, scaling functions, wavelets, and wavelet bases on intervals. The details of these modifications can be found in [10]. In wavelet regression, the wavelet bases on
final intervals are used in the following way. Let \( \{V_j\}_{j=0}^\infty \) be the MRA defined on the interval \([0, 1]\) and \( \{W_j\}_{j=0}^\infty \) be the corresponding wavelet subspaces. For a fixed \( J \), we assume \( (\phi_{j,k}) \) forms an orthonormal basis of \( V_j \), where \( \phi_{j,k} \) has been modified from its original version \( 2^{-j/2}\phi(2^j x - k) \) as in [10],[13]. After the modification, the subscript \((J, k)\) indicates the spatial location of \( \phi_{j,k} \). A function \( f \in V_J \) is now decomposed to

\[
    f_J(t) = \sum a_{J,k} \phi_{J,k}(t), \quad a_{J,k} = \langle f, \phi_{J,k} \rangle.
\]

(5)

where the sum in (5) is a finite one. Similarly, we have \( V_J = V_{J_0} \bigoplus_{j=J_0}^{J-1} W_j \) and

\[
    f_J(t) = f_{J_0}(t) + \sum_{j=J_0}^{J-1} g_j(t) = \sum a_{J_0,k} \phi_{J_0,k}(t) + \sum_{j=J_0}^{J-1} \sum w_{j,k} \psi_{j,k}(t),
\]

(6)

where \( a_{J_0,k} = \langle f, \phi_{J_0,k} \rangle \) and \( w_{j,k} = \langle f, \psi_{j,k} \rangle \).

For convenience, we write \( a_J = (a_{J,k}) \), \( a_{J_0} = (a_{J_0,k}) \), \( w_j = (w_{j,k}) \), \( v_a = [a_{J_0}, w_{J_0}, \ldots, w_{J,J-1}] \), and denote the transform from \( a_J \) to \( v_a \) by \( W \):

\[
    v_a = Wa_J.
\]

We call \( W \) the \textit{discrete wavelet transform} (DWT) and call its inverse \( W^{-1} \) the \textit{inverse discrete wavelet transform} (IDWT). It is easy to see that when \( \phi \) is an orthonormal MRA generator and \( \psi \) is the corresponding orthonormal wavelet, \( W \) is an orthonormal matrix. It follows that \( W^{-1} = W^T \). Without loss of generality, later we shall always assume \( J_0 = 0 \).

Based on pyramidal algorithms, Mallat in [18] developed fast algorithms for computing DWT and IDWT, which are also called Mallat’s algorithms. They have the computational complexity of \( O(n \log n) \), where \( n = \text{length}(a_J) \).

### 2.2 Adaptive Wavelet Regression

We now return to the model (1). Since the sample data are finite, without loss of generality, we assume \( N \) in (1) is equal to the dimension of the space \( V_J \) for a certain level \( J \). (If it is not, we can extend the sample data to let it be.) Then we can identify the data in (1) to a function of the space \( V_J \). Wavelet regression adopts a wavelet estimator to estimate the underlying function. The estimator is select from a subspace of \( V_J \) and the selection is dependent on the sample data. Hence, wavelet regression usually is a nonlinear regression. For the given sample data in (1), the wavelet regression can be outlined in the following diagram.

Let $W$ be the DWT. Let $T_a : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the adaptive operator, which changes wavelet coefficients for a certain purpose. Then the algorithm that performs wavelet regression can be written as follows.

$$v_y = Wy,$$
$$\hat{v}_y = T_a v_y,$$
$$\hat{y} = W^{-1} \hat{v}_y.$$  

We write

$$W = W^T T_a W.$$  

Then $W$ denotes the wavelet regression operator. If the adaptive operator $T_a$ is a compression one, then the adaptive wavelet regression is a wavelet shrinkage since the number of non-vanished entries of $\hat{v}_y$ is smaller than that one of $v_y$.

The wavelet regression is based on the following facts.

- The orthonormal transform of white noise $\sim N(0, \sigma^2)$ is still a white noise. $\sim N(0, \sigma^2)$. (See [13].)
- If $f$ is a noise-free function, then most of its wavelet coefficients vanish, only a very few of them have non-neglected values, which represent the details of the function.
- If a function $f$ carries noise, then its smooth component $f_{J_0}(t)$ in (6) is not influenced by noise very much, while the wavelet components keep noise.

Figures 1 and 2 show the wavelet decompositions of noise-free functions, and Figures 3 and 4 show them of noisy functions.

By these properties, in wavelet regression, the adaptive operator $T_a$ is selected so that it does not change $a_0$, i.e.,

$$T_a = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{T} \end{bmatrix},$$

where $\mathbf{T}$ is the operator on $W := \bigoplus_{j=1}^{J_{\min}} W_j$. Let $\mathbf{w} = [w_0, \ldots, w_{J-1}]$ and assume the length of $\mathbf{w}$ is $n$. We simply denote $\mathbf{w} = (w_i)_{i=1}^n$. Thus,

$$\mathbf{w} = \theta + \mathbf{z}, \quad w_i = \theta_i + z_i,$$  

(7)

where $\theta_i$ is the wavelet coefficient of the underlying function $f$ and $z_i$ is the wavelet transform of $\epsilon_i$. If $\epsilon_i$ is assumed to be the white noise $\sim N(0, \sigma^2)$, so is $z_i$. Then the wavelet regression is essentially to find a $\mathbf{w}$ that minimizes the risk

$$R(\mathbf{w}, \theta) = \frac{1}{n} E(||\mathbf{w} - \theta||^2).$$  

(8)

Since small wavelet coefficients mostly contribute to noise while large ones to signal, wavelet regression removes the small wavelet coefficients from $\mathbf{w}$. 
Donoho and Johnstone in [13] used two different shrinkages to design $T$. For a given threshold $\lambda > 0$, the hard threshold function is

$$\eta_h(x; \lambda) = \begin{cases} 
0, & |x| \leq \lambda \\
x, & |x| > \lambda 
\end{cases},$$

(9)

and the soft threshold function is

$$\eta_s(x; \lambda) = \begin{cases} 
0, & |x| \leq \lambda \\
sign(x) (|x| - \lambda), & |x| > \lambda 
\end{cases}.$$  

(10)

For a vector $w$, we write $\eta_h(w; \lambda) = (\eta_h(w_i; \lambda))$ and $\eta_s(w; \lambda) = (\eta_s(w_i; \lambda))$ respectively. Then the adaptive operators corresponding to $\eta_h(w; \lambda)$ and $\eta_s(w; \lambda)$ are $T^h w = \eta_h(w; \lambda)$ and $T^s w = \eta_s(w; \lambda)$ respectively.

The choice of the threshold $\lambda$ is a fundamental issue in wavelet shrinkage. A large threshold cuts too many coefficients and will result in an oversmoothing estimator. Conversely, a small threshold does not remove noise well and will produce a wiggly, undersmoothing estimator. The proper threshold ought to take a careful balance. A lot of work have been done in this aspect. (See [1], [2], [7], [12], [13], [14], [15], [16], [29], [30], [31], and [32].)

Figure 1. Wavelet decomposition of Block. Up to Level 4.
Figure 2. The wavelet decomposition of Bumps.

Figure 3. Wavelet decomposition of Blocks with noise.
3 Variational Method for Wavelet Regression

Wavelet shrinkage is based on the principle that each wavelet coefficient whose absolute value is smaller than the threshold contributes to noise; otherwise it contributes to the signal. The threshold in wavelet shrinkage is used to balance the oversmoothing and undersmoothing. The value of the threshold is dependent of the noise level, the size of the sampling data, and the smoothness required by the regression. The authors of [13] and [14] established several rules for determining thresholds, including Universal thresholding, MiniMax thresholding, SureShrink, Heursure and so on. (See also [15], [16], [29], [30].) These methods are very effective when noise is white and its level is known. However, in many applications, the noise type is unknown and the noise level is not uniform over all sampling areas. It is necessary to study wavelet regression in a general framework. By definition, regression is the minimization of the risk function, where the key issue is to remove noise while keeping the features of the data. Hence, we discuss the variational method for wavelet regression and reveal the relation between the variational method and the wavelet shrinkage.

Following the idea of [28], to balance the smoothness and the sharpness of an estimator, we introduce two energy functions. The first energy measures the distance between the estimator and the sampling data. It is clear that the estimator should be close to the sample data. Recall that in wavelet regression, we do not change the function \( f_{J_0} \in V_{J_0} \). Hence, the distance can be measured
by the following wavelet deviation energy

$$A(u) = \|u - w^0\|^2,$$

(11)

where $w^0$ represents the vector of wavelet coefficients of the observation. By the properties of wavelets, $A(u)$ controls the undersmoothing. Since $w^0$ carries noise, the estimator will be undersmoothing if $A(u)$ is close to 0. On the other hand, the wavelet components of a function represent its wiggle and noise. Therefore, we can create a weighted wavelet energy to measure the wiggle together with the noise for data $u$:

$$F(u) = u^T G u,$$

(12)

where $G(u)$ is a positive definite (or semi-definite) matrix. Under the assumption that noises are independent random variables, we may assume $G$ is a diagonal matrix, in which each entry on the diagonal is a weight function of $u_i$. Recall that small wavelet coefficients contribute to noise, and large ones contribute to signal. For the purpose of regression, large weights ought to assign to small coefficients and vice versa. Hence, we have the following.

**Definition 2** Let $G$ be a nonnegative diagonal matrix defined by

$$G = \text{diag}(c_i(u_i))_{i=1}^n,$$

(13)

where each $c_i$ is an even function satisfying the conditions: (1) $c_i(s) \geq 0$ and $\lim_{s \to \infty} c_i(s) = 0$; (2) $c_i(|s|)$ is decreasing on $[0, \infty)$. Then

$$F(u) = u^T G u$$

is called the weighted wavelet energy and $G$ is called the weight matrix.

Since $G$ is a diagonal matrix, $F(u)$ can be simplified to the quadratic form $\sum_{i=1}^n s^2 c_i(s)$. Write $g_i(s) = s^2 c_i(s)$. Then $(g_i(s))$ is the energy density. The weighted wavelet energy controls oversmoothing. If $F(u) = 0$, then all wavelet components vanish, which indicates that the estimator is in $V_{J_0}$, i.e., the regression will cause oversmoothing.

We now combine $G(u)$ and $A(u)$ to make an energy function

$$\Gamma(u) = G(u) + \lambda A(u),$$

(14)

where $\lambda$ is the parameter used to balance oversmoothing and undersmoothing. Thus, wavelet regression can be formulated to the following variational problem:

Find $w$ such that

$$w = \arg(\min \Gamma(u)).$$

(15)

Assume each $c_i(s)$ is differentiable. Then the solution of (15) satisfies the Euler-Lagrangian equation

$$g'(w) + 2\lambda(w - w^0) = 0,$$
where \( g'(w) = (g'_i(w_i)) \). It follows that
\[
2\lambda w + g'(w) = 2\lambda w^0,
\]
i.e.,
\[
2\lambda w_i + g'(w_i, w^0_i) = 2\lambda w^0_i, \quad i = 1, \ldots, n.
\]
If the matrix \( G \) is not a diagonal one, then a nonlinear matrix equation will replace (17).

We now establish the relation between the variational method and the wavelet shrinkage. A general linear wavelet shrinkage is formulated as
\[
\Gamma DS(w_i, \delta_i) = (\delta_i w_i)_{i=1}^n, \quad \delta_i \in [0, 1].
\]
We show that if in \( F(u) = \sum g_i(u) \), each density \( g_i \) is linear on \([0, \infty)\), then the solution of the variational problem (15) leads to a linear shrinkage.

**Theorem 3** If a variational problem (15) satisfies
\[
g_i(s) = \mu_i |s|, \quad \mu_i \geq 0
\]
then it yields a linear wavelet shrinkage.

**Proof.** Without loss of generality, we can assume the balance parameter \( \lambda \) in \( \Gamma(u) \) is 1/2. Otherwise, we use \( \lambda \mu_i/2 \) to replace \( \mu_i \). We have
\[
\rho'_i(s) = \mu_i \text{sgn}(s),
\]
which leads to the Euler-Lagrangian equation
\[
s + \mu_i \text{sgn}(s) = w^0_i, \quad i = 1, \ldots, n.
\]
Since \( \rho_i'(0) \) does not exist, 0 is also a critical point. When \( |w^0_i| < \mu_i \), the equation (19) has no solution; and when \( |w^0_i| \geq \mu_i \), the equation has the unique solution
\[
s = \text{sgn}(w^0_i) (|w^0_i| - \mu_i).
\]
Then it is easy to verify that the vector \( w = (w_i) \) with
\[
 w_i = \text{sgn}(w^0_i) (|w^0_i| - \mu_i)_+ = \frac{|w^0_i| - \mu_i}{|w^0_i|} w^0_i, \quad i = 1, \ldots, n,
\]
minimizes \( \Gamma(u) \), where \( x_+ = \max(x, 0) \). Let \( \delta_i = (|w^0_i| - \mu_i)/|w^0_i| \). Then \( \delta_i \in [0, 1] \).

As applications of Theorem 3, we discuss the wavelet shrinkages by using hard threshold and soft threshold respectively.

**Example 1. [Hard thresholding]** Let \( \delta \) be a given threshold. We assume in \( w^0 \) each sample value whose absolute value is less than \( \delta \) represents noise.
Otherwise it does not carry on noise. The assumption suggests the following weight:

$$c_i(s) = \begin{cases} \delta/|s|, & |w_i^0| \leq \delta, \\ 0, & |w_i^0| > \delta, \end{cases}$$

which leads

$$\rho_i(s) = \begin{cases} \delta |s|, & |w_i^0| \leq \delta, \\ 0, & |w_i^0| > \delta. \end{cases}$$

and

$$\rho_i'(s) = \begin{cases} \delta \text{sgn}|s|, & |w_i^0| \leq \delta, \\ 0, & |w_i^0| > \delta. \end{cases}$$

By (20), the solution $w$ of the variational problem (15) is

$$w_i = \begin{cases} \frac{1}{2\lambda} \cdot w_i^0, & |w_i^0| \leq \delta, \\ |w_i^0| > \delta. \end{cases}$$

Hence, when $\lambda \leq \frac{1}{2}$, the solution provides the wavelet shrinkage using the hard threshold $\delta$.

**Example 2. [Soft thresholding]** We now assume in $w^0$ each sample value carries on noise~$N(0, \sigma^2)$. Hence, we adopt the following weight

$$c_i(s) = \delta/|s|,$$

which leads $\rho_i(s) = \delta|s|$ and therefore

$$\rho_i'(s) = \delta \text{sgn}(s).$$

By (21), the solution $w$ of the variational problem (15) is

$$w_i = \text{sgn}(w_i^0) \left(|w_i^0| - \frac{1}{2\lambda} \delta\right)_+,$$

which provides the wavelet shrinkage using the soft threshold $\delta/(2\lambda)$.

### 4 Examples

In this section, we give several examples to illustrate the function of $\lambda$ in the wavelet regression. We select four functions to test the regression: “Blocks” is a step function, which has several step jumps. According to [20] and [21], Lipschitz order of a step jump is 0. “Heavy Sine” is a broken sine wave, which has two jumps. “Bumps” is a continuous function but has a lot of non-differentiable points, which are classified to be with Lipschitz orders in $(0, 1)$. “Doppler” is smooth everywhere except at the origin, where the function has infinite oscillations. The white noise is added to the sample data of them. Let $\sigma(f)$ and $\sigma(n)$
denote the standard deviation of the function and the noise respectively. In Figure 5 – Figure 8, we set \( \sigma(f)/\sigma(n) = 5 \). In the regression, we set \( \delta = \sqrt{2\sigma(n)} \ln N \), where \( N \) is the size of the data. We also set \( \lambda = 2, 1, 1/2, 1/4, 1/10 \), which yield the thresholds \( \delta/2, \delta, 2\delta, 4\delta, 10\delta \) respectively. In Figure 9, we choose \( \lambda = 2 \), which yield the threshold \( 4\delta \). The results show that the regression is not very sensitive to small \( \lambda \), say \( \lambda \leq 1/2 \), but sensitive to \( \lambda > 1 \). To regress the functions with singularity as in “Doppler”, a larger \( \lambda \) should be chosen. Other tests show that the regression does not very much rely on the choices of particular wavelets. (The figures are not included in this paper.) Using different wavelets in a regression causes little differences, provided that the size of sampling data is large (say \( \geq 2^{10} \) when wavelet level is \( \leq 4 \)).

When the sampling data carry on the noise with uniform distribution, the wavelet regression still works well. Figure 10 – Figure 13 show the wavelet regression of the four functions with uniform noise. Again, the estimator of “Doppler” produces a larger error.

Acknowledgment. The author thanks the referees for their valuable comments.
Figure 6. Wavelet regression for Heavy Sine.

Figure 7. Wavelet regression for Bumps.
Figure 8. Wavelet regression for Doppler.

Figure 9. Wavelet regression for $\lambda = 2$, with $\frac{\sigma(f)}{\sigma(n)} = 5$. 
Figure 10. Wavelet regression of Block with uniform noise.

Figure 11. Wavelet regression of Heavy Sine with uniform noise.
Figure 12. Wavelet regression of Bumps with uniform noise.

Figure 13. Wavelet regression of Doppler with uniform noise.

Acknowledgments: This work was supported in part by ER-Grant 2003 of Sam Houston State University.
References


