

# Hybrid Seismic Denoising Using Higher-Order Statistics and Improved Wavelet Block Thresholding

by S. Mostafa Mousavi and Charles A. Langston

**Abstract** We introduce a nondiagonal seismic denoising method based on the continuous wavelet transform with hybrid block thresholding (BT). Parameters for the BT step are adaptively adjusted to the inferred signal property by minimizing the unbiased risk estimate of Stein (1980). The efficiency of the denoising for seismic data has been improved by adapting the wavelet thresholding and adding a preprocessing step based on a higher-order statistical analysis and a postprocessing step based on Wiener filtering. Application of the proposed method on synthetic and real seismic data shows the effectiveness of the method for denoising and improving the signal-to-noise ratio of local microseismic, regional, and ocean bottom seismic data.

## Introduction

Over the last decade, the use of ambient seismic noise to deduce earth structure has changed the credibility and sensibility of the traditional definition of seismic noise as unwanted energy (e.g., Kumar and Ahmed, 2014). However, in many conventional seismic methods, it is still the case that the reliability of the seismic data and accuracy of parameter extraction, such as onset time, polarity, and amplitude, are directly affected by the noise level. As a result, the accuracy of event location and other attributes derived from seismic traces are also influenced by the noise content. Therefore, there is still a great need for developing suitable procedures that improve signal-to-noise ratios (SNRs) allowing for robust seismic processing. The development of methods for seismic noise attenuation continues to be a challenging problem in seismology because of the superposed nature of seismic signals. Good examples where SNRs are relatively small include marine experiments incorporating oceanic bottom seismometers (OBSs), seismological experiments in urban areas, and surface monitoring of microseismic events induced by hydraulic fracturing. One approach to improve the SNR in a seismic experiment is through the use of signal processing techniques.

Spectral filtering is a common practice in many seismic processing workflows. It attenuates some high- and/or low-frequency components to improve the SNR. However, filtering is not always effective because the signal and noise often share some frequency bands; part of the signal may be filtered out with the noise, whereas the noise might not be fully attenuated. Moreover, band-pass filtering inevitably distorts the signal. Although filtering might improve the SNR, it will not necessarily make phase onsets or polarities easier to observe (Douglas, 1997). Spectral filtering can also generate artifacts prior to impulsive arrivals that can be confused with seismic signal (Scherbaum, 1996).

Effective seismic denoising can be achieved through more complicated techniques, such as: principal component analysis (Hagen, 1982), f-x deconvolution (Canales, 1984), eigen image (Canales, 1984; Gulunay, 1986), singular value decomposition (Ursin and Zheng, 1985), Karhunen–Loève transform (Jones and Levy, 1987; Al-Yahya, 1991), time varying band-pass filters (Yilmaz, 1987), t-x prediction filtering (Abma and Claerbout, 1995), optimum (Wiener) filters (Douglas, 1997), band-pass filtering, f-k, and kx-ky filtering (Yilmaz, 2001), artificial neural networks (Djarfour *et al.*, 2008), Cadzow filtering (Trickett, 2008), matched filters (Gibbons and Ringdal, 2006; Eisner *et al.*, 2008), fuzzy methods (Hashemi *et al.*, 2008), singular spectrum analysis (Oropeza and Sacchi, 2011), nonlocal means algorithm (Bonar and Sacchi, 2012), empirical mode decompositions (Bekara and van der Baan, 2009; Han and van der Baan, 2015), band-variable filtering (Ditommaso *et al.*, 2012), S-transform (Pinnegar and Eaton, 2003; Schimmel and Gallart, 2007; Askari and Siahkoobi, 2008; Parolai, 2009; Ditommaso *et al.*, 2010; Tselentis *et al.*, 2012), 3C group sparsity constrained time–frequency transform (Rodriguez *et al.*, 2012), time–frequency reassignment (Han *et al.*, 2014), the wave packet transform (Galiana-Merino *et al.*, 2003; Shuchong and Xun, 2014), and methods based on the wavelet transform (Anestis and Oppenheim, 1995; Pazos *et al.*, 2003; Sobolev and Lyubushin, 2006; Chik *et al.*, 2009; To *et al.*, 2009; Ansari *et al.*, 2010; Beenamol *et al.*, 2012).

In this article, we introduce an adaptive algorithm for automatic noise reduction based on the continuous wavelet transform (CWT) incorporating higher-order statistics (HOS) and block thresholding (BT). BT is an adaptive, nondiagonal estimator introduced in mathematical statistics by Cai (1999) to obtain minimax signal estimators for nonparametric regression that uses information from neighboring time–frequency coefficients in the thresholding step.

In the following, we first introduce the theoretical background of CWT and wavelet denoising. Next, the proposed seismic denoising method will be described and the effectiveness of the method will be explored by applying it to synthetic and different types of real seismic data. Denoised results will be compared against results using standard spectral filtering and time–frequency thresholdings. The proposed method shows promise for increasing the SNR of seismic records of microseismic data, regional earthquake waveforms, and OBS data.

## Theoretical Background

### Time–Frequency Representation

Seismic signals are oscillatory signals with nonlinear and nonstationary patterns. In the time domain, a seismic signal  $f(t)$  can be modeled as a superposition of individual time-varying harmonic components  $f_k(t)$ :

$$f(t) = \sum_{k=1}^K f_k(t), \quad (1)$$

in which  $K$  is the maximum number of components in the signal. The observed time series  $y(t)$  will be contaminated by additive noise due to measurement errors and other processes  $\varepsilon(t)$  so that

$$y(t) = f(t) + \varepsilon(t). \quad (2)$$

The noise is often modeled as a zero-mean Gaussian process independent of the signal. Each component of the seismic signal can be modeled by harmonic signals of the form

$$f_k(t) = A_k(t) \cos(\theta_k(t)), \quad (3)$$

in which  $A_k(t)$  and  $\theta_k(t)$  are the instantaneous amplitude and phase of the  $k$ th signal component, respectively, found using the real part of the analytic signal. The analytic signal  $S(t)$  of a real-time series  $s(t)$  is given by

$$S(t) = s(t) - iH[s(t)], \quad (4)$$

in which  $H[\cdot]$  is the Hilbert transform. The instantaneous frequency  $\omega_k(t)$  of the  $k$ th component is derived from the instantaneous phase by

$$\omega_k(t) = \frac{\partial \theta_k(t)}{\partial t}. \quad (5)$$

This implies that components of the seismic signal  $f_k(t)$  can appear at different arrival times with different amplitudes and instantaneous frequencies. Hence, an analysis in the time or frequency domain alone may not provide adequate information. Time–frequency analysis of multilayered signals can give insights into the structure of the signal by localizing frequency so that noise can be distinguished from the signal more efficiently.

The CWT (Daubechies, 1992) is a popular time–frequency transform. It produces a time–frequency representation with better resolution compared with the short time Fourier trans-

form (Tary et al., 2014), because the signal is analyzed under different resolutions (or scales) at different frequencies. The idea behind the CWT is to apply a prototype analyzing function known as the mother wavelet over the time series. For a given mother wavelet  $\psi$ , the CWT of  $y(t)$  in relation (2) at scale  $a$  and time shift  $\tau$  is given by (Daubechies, 1992)

$$Wy(a, \tau) = \langle y, \psi_{a,\tau} \rangle = \int_{-\infty}^{+\infty} y(t) a^{-1/2} \psi^* \left( \frac{t-\tau}{a} \right) dt, \quad (6)$$

in which the  $*$  is the complex conjugate,  $\langle y, \psi_{a,\tau} \rangle$  is the inner product in  $L^2(R)$ , and  $Wy$  is the wavelet coefficient representing the observed signal  $y(t)$  at scale  $a$ ; assuming a wavelet  $\psi$ , with a Fourier transform that satisfies the admissibility condition (Daubechies, 1992; Farge, 1992)

$$C_\psi = \int_{-\infty}^{+\infty} |\omega|^{-1} |\hat{\psi}(\omega)|^2 d\omega < \infty. \quad (7)$$

The inversion of the CWT can be expressed as

$$y(t) = \frac{1}{C_\psi} \int_a \int_\tau Wy(a, \tau) d\tau \frac{da}{a^2}. \quad (8)$$

The CWT can be thought of as cross correlation of  $y(t)$  with a number of wavelets that are a stretched (or compressed) and shifted version of the original mother wavelet  $\psi$ . Depending on what signal features are of interest, a wavelet can be selected to facilitate detection of that feature. The advantage of using CWT for denoising over orthogonal wavelet transforms such as discrete wavelet transform or wavelet packet transform is in its translation-invariant nature. Nonsubsampled or redundant wavelet transforms like CWT can reduce pseudo-Gibbs artifacts (Coifman and Donoho, 1995) in the denoised signal (Elad and Aharon, 2006).

### Wavelet Estimation

The problem of seismic denoising can be restated as a nonparametric regression problem for a noisy, sampled version of  $y(t)$ :

$$y_i = f_i + \sigma \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (9)$$

in which  $\sigma$  is the known noise level and the  $\varepsilon_i$ s are independent standard random variates. The goal is to estimate the underlying regression function  $f$ , the seismic signal in equation (1), from the noisy data  $y_i$  without assuming any particular parametric structure for  $f$ . The estimated signal will be denoted by  $\hat{f}$ .

In signal processing, this nonparametric estimation process has long remained mostly Bayesian and linear (Mallat, 1999; Matz and Hlawatsch, 2000). Linear methods are not designed to handle spatially inhomogeneous functions with low regularity. For such functions, nonlinear methods are needed. Donoho and Johnston (Johnstone, 1992; Donoho and Johnstone, 1994, 1995; Donoho, 1995) in their influen-

tial works, proved that a simple thresholding for sparse representations can yield near-optimal nonlinear estimates in terms of the root mean square error (rmse).

The representation of a signal in a basis is termed sparse when the signal strength is concentrated in a few of the coefficients of the basis. The notion of sparsity of the signal representation is fundamental in wavelet estimation; wavelet bases provide such a sparsity. A wavelet transform can compact the energy of a normal function into very few large magnitude wavelet coefficients (DeVore *et al.*, 1992; Meyer, 1992).

Wavelet denoising methods are based on the use of the wavelet transform as a standard device for transforming a function estimation problem into a normal mean problem of estimating the wavelet coefficients in the sequence domain (e.g., Donoho and Johnstone, 1994; Cai, 1999). They generally consist of three steps. First, a noisy signal is decomposed into time–frequency atoms (e.g., equation 6) and then the time–frequency structure of the signal is discriminated from the noise. The time–frequency coefficients are then modified by a thresholding process to attenuate the noise and find an estimate of the signal as similar to  $f$  as possible. At the end, the estimated signal is reconstructed from modified coefficients using the inverse transform (equation 8).

The advantage of denoising in the time–frequency domain over traditional spectral filtering is that it allows for noise in the same passband as the signal to be excluded from the signal so long as it is temporally separated from the arrivals. The major problems are in choice of a suitable thresholding value and a thresholding function for attenuating the wavelet coefficients associated with the inferred noise.

Thresholding uses the data itself to decide which coefficients are significant. Donoho and Johnstone (1994) showed that for the Gaussian white noise of variance  $\sigma^2$ , universal threshold is  $\lambda = \sigma\sqrt{2\log_e n}$  for a signal with length  $n$ . The main assumption behind the universal threshold method of Donoho and Johnstone (1994) is that the small wavelet coefficients are mostly due to the noise, whereas the signal is stored mainly in a few large coefficients. So wavelet coefficients can be set to zero if their magnitudes are less than the predetermined threshold  $\lambda$ . This keep-or-kill rule is called hard thresholding. Another popular scheme is known as soft thresholding (shrink-or-kill rule) where larger coefficients are smoothly reduced by the threshold value. Other methods like the customized thresholding of Yoon and Vaidyanathan (2004) or Shuchong and Xun (2014) have a behavior between hard and soft thresholding.

Although these algorithms give improved performance compared with previous methods, these approaches remain diagonal estimators, that is, the wavelet coefficients of the function are estimated based solely on their individual magnitudes and are kept or killed term by term. Term-by-term thresholding achieves a degree of trade-off between variance and bias contribution to the mean square error, but is not optimal. The resulting estimator is biased and has a suboptimal  $L^2$ -risk convergence rate (Antoniadis *et al.*, 2001). More-

over, in this scheme, isolated time–frequency coefficients can produce strong artifacts similar to highly annoying harmonic noise in audio signals (Mallat, 1999). This kind of noise can be removed by nondiagonal estimation schemes like BT, that regularize the geometry of the estimated signal and avoid leaving isolated points (Ephraim and Malah, 1984).

Hall *et al.* (1998, 1999a) and Cai (1999) studied local BT rules for function estimation using wavelet transforms to increase estimation precision. These methods group coefficients into nonoverlapping blocks and threshold them in groups rather than as individuals. The aim is to increase estimation accuracy by utilizing information about neighboring coefficients and improve convergence rates. The basic motivation of BT is that if neighboring coefficients contain some of the signal's energy then it is likely that the current coefficients do also, and so a lower threshold should be used, essentially yielding a different local trade-off between signal and noise. It has been shown that wavelet BT estimators have excellent mean square error performances relative to wavelet term-by-term thresholding estimators in finite sample situations (Antoniadis *et al.*, 2001). Moreover, BT estimators automatically adapt to the sparsity of the underlying wavelet coefficients and, hence, are completely data driven.

Hence, BT may be regarded as an automatic signal model selection procedure, which selects a set of important variables (wavelet coefficients) by omitting insignificant ones, then fits to the data a model consisting of only the important variables. Nondiagonal estimators such as BT (e.g., Cai, 1999; Hall *et al.*, 1999b; Cai and Silverman, 2001; Cai and Zhou, 2009) enjoy a number of advantages over the conventional, term-by-term thresholding methods. They increase estimation precision by utilizing information about neighboring wavelet coefficients and enjoy a high degree of adaptivity. Nonuniform or adaptive thresholding depends on the relationship between the energy distribution of the observed signal and that of the noise.

The use of different thresholds for wavelet coefficients seems reasonable since adaptive thresholding accounts for variation of the local statistics of wavelet coefficients. The degree of the adaptability, however, depends on the choice of the block size and threshold level, which influence the performance of the estimator. It has been shown that nondiagonal estimators clearly outperform diagonal estimators (e.g., Cai, 1999; Cai and Silverman, 2001; Cai and Zhou, 2009) and reduce the estimation risk (Ephraim and Malah, 1984; Kim and Chang, 2000; Wolf and Godsill, 2001; Cohen, 2002, 2005).

## Hybrid Seismic Denoising

Our hybrid method of seismic denoising consists of a preprocessing step using HOS criteria, a wavelet denoising based on the BT, and a postprocessing step based on the Wiener filtering. First, we decompose the signal into its time–frequency structure using CWT. Let us define coefficients of the wavelet expansion of the noisy observations,  $y_i$  in equation (9) by

$$Wy_i = Wf_i + \sigma W\varepsilon_i, \quad i = 1, 2, \dots, n, \quad (10)$$

in which  $Wy = \langle y, \psi_{a,\tau} \rangle$ ,  $Wf = \langle f, \psi_{a,\tau} \rangle$ , and  $W\varepsilon = \langle \varepsilon, \psi_{a,\tau} \rangle$ .

### Preprocessing

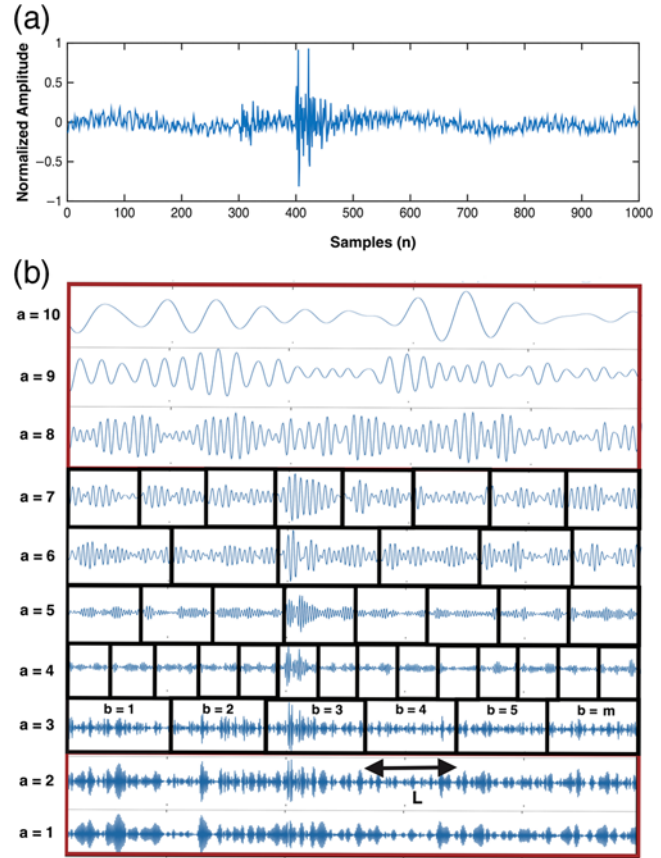
In wavelet denoising, it is assumed that the coefficients at the finest decomposition level (highest frequencies) are associated with pure noise. This assumption is usually used for the noise level estimation in wavelet thresholding. Frequency contents of seismic signals are concentrated in limited frequency bands that occur at middle decomposition levels in the wavelet domain. In addition to the highest frequencies (smallest scales or finest decomposition levels), most of the coefficients in the longer period section (largest scales) will be mainly associated with noise. High-power noises outside of the frequency range of the seismic signal can decrease efficiency of wavelet thresholding. Therefore, in our hybrid scheme of denoising, after the decomposition of the signal into its functional elements using CWT, we detect and remove elements that are mainly due to noise. This is done using the HOS to detect and remove any decomposition level with a Gaussian distribution from the time–frequency picture. HOS is referred to statistical descriptors of orders greater than two. HOS methods were used in several studies for denoising (Ravier and Amblard, 2001; Zhang and Zhao, 2001; Tsolis and Xenos, 2011). Computing the Kurtosis is a good choice for the Gaussianity measurement using HOS (DeCarlo, 1997). It is defined as the standardized fourth moment about the mean. When considering a numerical signal of  $N$  samples  $y_n = \{y_1, y_2, \dots, y_N\}$ , such as in real seismograms, the Kurtosis (kurt) is defined as (Bickel and Doksum, 1977)

$$\text{kurt}_y = \frac{\sum_{n=1}^N (y_n - \mu_y)^4}{N(\sigma_y^4)} - 3, \quad (11)$$

in which  $\mu_y$  is the mean and the  $\sigma_y$  is the estimated standard deviation of the signal  $y$ . Kurtosis can be used to classify signals as Gaussian or not. This is because the Kurtosis of Gaussian signals is equal to zero, which is not the case for non-Gaussian distributions. The bias and variance of the Kurtosis is estimated by  $B(\text{kurt}_y) = -6/N$  and  $\text{Var}(\text{kurt}_y) = 24/N$ , respectively (Tsolis and Xenos, 2011). A Kurtosis estimator can take values as follows:

$$|\text{kurt}_y| \leq \frac{\sqrt{24/N}}{\sqrt{1-a}}, \quad (12)$$

in which  $a$  is the level of confidence, with a numerically estimated optimum equal to 90% (Ravier and Amblard, 2001). In the preprocessing step, the Kurtosis is calculated for each decomposition level. The HOS criterion is applied to detect the scales capturing only Gaussian noise and removing them from the time–frequency picture. This step acts similar to automatic band-pass filtering and keeps just scales that are composed of coefficients associated with a noisy signal. Because of mode



**Figure 1.** (a) Noisy signal and (b) functional elements of the signal decomposed into different scales using continuous wavelet transform (CWT). Here, just the real part of the wavelet coefficients is plotted. Lower and higher frequencies (larger and smaller scales respectively) are mainly due to the noise and are detected and removed in the preprocessing step. The middle scales of the CWT contain both noise and signal energy. Block length  $L$  is selected for each scale by minimizing the Stein's unbiased risk estimator (SURE) statistic (see text). The color version of this figure is available only in the electronic edition.

mixing with the signal's components, the noise coefficients in remaining scales cannot be detected using HOS.

### Block Thresholding

To attenuate the remaining noise at other scales, we use wavelet thresholding. According to the data compression and localization properties of wavelets (Daubechies, 1992), it is reasonable to think of  $Wf$  as a high-dimensional, sparse, and normal mean vector (Cai, 1999). Hence, the mean  $Wf_i$  is approximately the true wavelet coefficients of  $f$ .

Our approach of wavelet thresholding is BT. Let  $L \geq 1$  be the possible length of each block, and  $m = n/L$  be the number of blocks in each scale. We partition the wavelet coefficients at each resolution level into  $m$  nonoverlapping blocks of size  $L$  (length of  $L$  samples and width of one sample) (Fig. 1). Let  $Wy_b = (Wy_{(b-1)L+1}, \dots, Wy_{bL})$ ,  $Wf_b = (Wf_{(b-1)L+1}, \dots, Wf_{bL})$ , and  $W\varepsilon_b = (W\varepsilon_{(b-1)L+1}, \dots, W\varepsilon_{bL})$  represent coefficients of observa-



tions, signal, and noise in block  $b$ , respectively. The  $L^2$ -energy of the noisy signal in block  $b$  is denoted by the sum of squared wavelet coefficients in the block:

$$S_b^2 = \sum_{i \in b} W y_i^2, \quad \text{for } b = 1, 2, \dots, m. \quad (13)$$

The goal is to estimate  $Wf = (Wf_1, \dots, Wf_n)$  based on the observations  $Wy = (Wy_1, \dots, Wy_n)$  under the average mean square error:

$$R(W\tilde{f}, Wf) = \frac{1}{n} \sum_{i=1}^n E(W\tilde{f}_i - Wf_i)^2. \quad (14)$$

Cai (1999) considered a class of block-wise James–Stein estimators based on the James–Stein shrinkage rule (James and Stein, 1961) to estimate signal's coefficients on each block:

$$W\tilde{f}_b(\lambda, L) = \left( \frac{S_b^2 - \lambda L \sigma_N^2}{S_b^2} \right)_+ W y_b, \quad \text{for } i \in b, \quad (15)$$

in which  $\lambda \geq 0$  is the threshold level and  $\sigma_N$  is the noise level. A block  $b$  is deemed important if  $S_b^2$  is larger than the threshold  $\lambda$  so that all the coefficients in the block are retained, otherwise the block is considered negligible and all the observed coefficients of  $Wy_b$  in the block are shrunk to estimate the true wavelet coefficients,  $W\tilde{f}_b$  at each scale.

The true noise level is unknown and needs to be estimated. We use the robust median estimator of Donoho (1995) given by  $\sigma_N = \text{median}(W\epsilon - \text{median}(W\epsilon))/0.6745$ , in which the normalizing factor 0.6745 is the mean absolute deviation of a Gaussian distribution. However, our estimate is not based on the finest resolution level. The noise level is estimated on each decomposition level from coefficients of the observations prior to the approximate arrival of the signal. This is based on the assumption that in the seismic data wavelet coefficients prior to the high-energy arrivals are only associated with the noise. This level-dependent noise estimation and thresholding scheme improves denoising even when the added noise is no longer white noise, for example, colored Gaussian noise (Johnstone and Silverman, 1997).

The choice of block size  $L$  and threshold level  $\lambda$  largely determines the performance of the resulting estimator. Optimal block size and threshold are empirically derived by minimizing Stein's unbiased risk estimator (SURE) (Stein, 1980). If we write an estimator of signal coefficients as  $W\tilde{f} = Wy + g(Wy)$ , in which  $g$  is a function from  $\mathbb{R}^L$  to  $\mathbb{R}^L$ , based on Stein's theorem, when  $g$  is weakly differentiable, then the risk of estimator can be estimated by

$$E\{\|W\tilde{f}_b(\lambda, L) - Wf_b\|_2^2\} = E\{L\sigma^2 + \|g(Wy)\|_2^2 + 2\sigma^2 \nabla g(Wy)\}, \quad (16)$$

in which  $E\{\}$  is the expected value and  $\|\cdot\|_2^2$  is  $L^2$  norm. Similarly for the BT, we can write

$$W\tilde{f}_b(\lambda, L) = \left( \frac{S_b^2 - \lambda L \sigma^2}{S_b^2} \right)_+ W y_b = W y_b + g(W y_b). \quad (17)$$

$$g(W y_b) = \left( 1 - \frac{\lambda L \sigma^2}{S_b^2} \right)_+ W y_b - W y_b \quad (18)$$

converges to the same answer if taken in the limit from both sides, hence, it is weakly differentiable. Simple calculation shows that

$$E_{W\tilde{f}_b} \|W\tilde{f}_b(\lambda, L) - Wf_b\|_2^2 = E_{W\tilde{f}_b} (\text{SURE}(W y_b, \lambda, L, \sigma)), \quad (19)$$

in which

$$\begin{aligned} \text{SURE}(W y_b, \lambda, L, \sigma) &= \sigma^2 \left( L + \frac{\lambda^2 L^2 - 2\lambda L(L-2)}{\frac{S_b^2}{\sigma^2}} I\left(\frac{S_b^2}{\sigma^2} > \lambda L\right) \right. \\ &\quad \left. + \left(\frac{S_b^2}{\sigma^2} - 2L\right) I\left(\frac{S_b^2}{\sigma^2} \leq \lambda L\right) \right). \end{aligned} \quad (20)$$

This implies that the total risk is

$$E_{W\tilde{f}} \|W\tilde{f}(\lambda, L) - Wf\|_2^2 = E_{W\tilde{f}} [\text{SURE}(Wy, \lambda, L, \sigma)] \quad (21)$$

in which

$$\text{SURE}(Wy, \lambda, L, \sigma) = \sum_{b=1}^m \text{SURE}(W y_b, \lambda, L, \sigma), \quad (22)$$

is an unbiased estimate of the risk. Stein's unbiased risk estimator serves as a guide for choosing level-dependent block size and threshold levels. At each scale, we choose the block size  $L^s$  and threshold level  $\lambda^s$  to minimize  $\text{SURE}(Wy, \lambda, L, \sigma)$ :

$$(\lambda^s, L^s) = \arg \max_{\{L-2, 0\} \leq \lambda \leq \lambda^F, 1 \leq L \leq n^v} \min \text{SURE}(W y, \lambda, L, \sigma), \quad (23)$$

in which  $v(0 \leq v < 1)$  and  $\lambda^F = 2L \log_e n$  are additional restrictions on the search range proposed by Cai and Zhou (2009). Optimal block size and threshold level  $L^s$  and  $\lambda^s$ , respectively, are estimated using a fast grid search.

The estimator of  $Wf$  is then given by a hybrid scheme (Cai and Zhou, 2009) to overcome the drawbacks of BT in situations of extreme sparsity of wavelet coefficients as

$$\begin{cases} W\tilde{f}_b = W y_b \left( 1 - \frac{\lambda^s L^s \sigma_N^2}{S_b^2} \right)_+ & \text{for } i \in b \text{ if } T_n > \gamma_n \\ W\tilde{f}_i = W y_i \left( 1 - \frac{2 \log n}{W y_i} \right)_+ & \text{for } i \in b \text{ if } T_n \leq \gamma_n \end{cases}, \quad (24)$$

in which  $T_n = n^{-1} \sum (Wy_i^2 - 1)$ , and  $\gamma_n = n^{-1/2} \log_2^{3/2} n$ . When  $T_n \leq \gamma_n$ , the estimator is a degenerate block James–Stein estimator with block size  $L^s = 1$ . In this case, the estimator is also called the non-negative garrote estimator.

### Postprocessing

It is known that wavelet denoising can be improved via Wiener filtering (Ghael *et al.*, 1997; Yu *et al.*, 2008; Sieden-burg, 2012). Similarly, we use the initial estimate of the signal coefficients  $\tilde{W}\hat{f}$  from BT to design a wavelet-domain Wiener filter and apply it to the noisy coefficients as the final step for maximum suppression of noise (Yu *et al.*, 2008):

$$\begin{cases} \hat{W}\hat{f}_b = Wy_b \left( \frac{|\tilde{W}\hat{f}_b|^2}{|\tilde{W}\hat{f}_b|^2 + L^s \sigma^2} \right) & \text{for } T_n > \gamma_n \\ \hat{W}\hat{f}_i = Wy_i \left( \frac{|\tilde{W}\hat{f}_i|^2}{|\tilde{W}\hat{f}_i|^2 + L^s \sigma^2} \right) & \text{for } T_n \leq \gamma_n \end{cases} \quad (25)$$

By applying this Wiener filtering in the postprocessing step, a second estimate of the signal's coefficients  $\hat{W}\hat{f}$  will be obtained.

We now return to the nonparametric function estimation problem of equation (9). We have estimated the coordinates of the mean vector of wavelet coefficients in a level-dependent manner. Parameters are adjusted based on the nature of the recorded signal. Now, the estimate of regression function in equation (9), denoted by  $\hat{f}$ , is obtained by applying the inverse transform over  $\hat{W}\hat{f}$  coefficients.

### Summary of the Denoising Algorithm

Our algorithm is a new approach in signal denoising, primarily in its application to seismic data where noise and signal can often be examined separately in time. The outline of the proposed seismic denoising procedure can be summarized as follows:

1. Transform the time-series data into the time–frequency domain through the CWT (equation 6).
2. The HOS analysis is done for each scale of the wavelet representation of the signal using equations (11) and (12). The coefficients of scales with Gaussian distributions are removed and those with non-Gaussian distributions are kept.
3. The optimal block size  $L^s$  and threshold level  $\lambda^s$  for coefficients of each scale are estimated by minimizing Stein's unbiased risk estimate (equation 23).
4. Wavelet coefficients are attenuated using the hybrid BT scheme, equation (24) and an initial estimate of signal coefficients is made.
5. The attenuated coefficients are used in a postdenoising step to build a Wiener filter and obtain final time–frequency coefficients of the estimated signal (i.e., equation 25).
6. The final denoised signal is obtained via the inverse CWT of attenuated wavelet coefficients (equation 8).

Table 1

Velocity Model Used for Computing the Synthetic Seismogram

Thickness (km)	$V_s$ (km/s)	$V_p$ (km/s)	Density (gm/cm <sup>3</sup> )	$Q_s$	$Q_p$
0.65	0.72	1.8	1.2	100	200
2	3.48	6.02	1.83	500	1000
2	2.79	4.83	1.61	400	800
12.35	3.56	6.17	2.0	500	1000
23	3.8	6.6	2.2	500	1000
0	4.7	8.0	2.6	500	1000

### Results

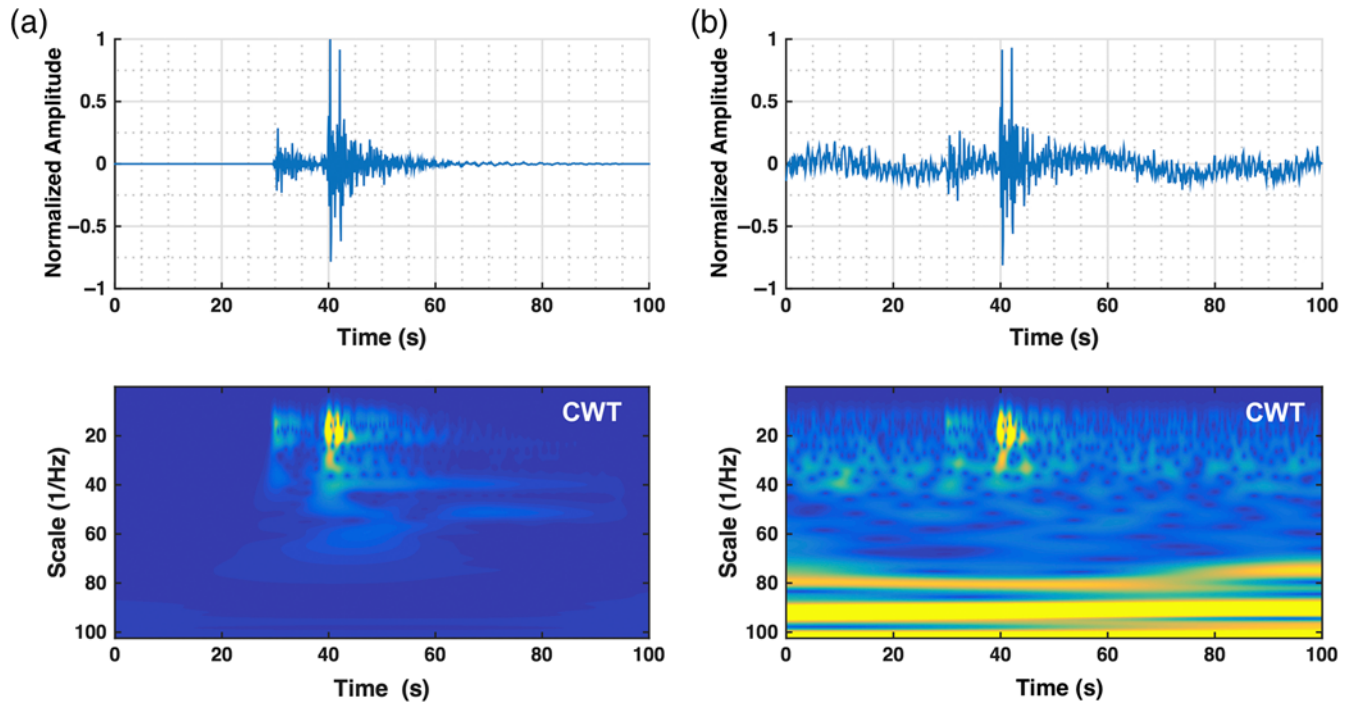
We first apply the algorithm to synthetic seismic signals. In this way, the shape of the whole output signal and the onset time of the first pulse can be compared with the ideal signal. Afterward, the denoising algorithm will be applied to a real induced microseismic signal, a regional earthquake seismogram, and then OBS data.

In our implementation of CWT, we use a Bump wavelet (Thakur *et al.*, 2013) as the mother wavelet with 100 scales. The noise level  $\sigma_N$  is estimated from the first 200 wavelet coefficients in each scale. In other words, we define the first 200 samples of data prior to the signal's arrival as the noise. However, this can be more precisely determined using any automatic procedure for onset picking or combining the denoising algorithm with an automatic onset picker in the wavelet domain, such as those proposed by Karamzadeh *et al.* (2013), Bogiatzis and Ishii (2015), or Mousavi *et al.* (2016).

### Synthetic Data

A synthetic seismogram is calculated using the frequency–wavenumber method (Zhu and Rivera, 2002) for the crustal model shown in Table 1. A point source was located at a depth of 12 km, and three component seismograms were computed for a receiver located on the surface at an epicentral distance of 80 km. Real seismic noise recorded by the New Madrid Cooperative Seismic Network was added to the synthetic seismogram in a way to yield to SNR of 2.5 for the resulting seismogram (Fig. 2).

To check the sensitivity of our method to the number of decomposition levels, we tested different numbers of scales for the wavelet transform. The denoising results were very similar. However, smaller scale numbers will speed up the denoising process. Optimal selection of the mother wavelet to produce the maximum number of wavelet coefficients close to zero while not reducing useful information is a crucial factor in wavelet thresholding. Different applications require different optimal wavelets. Luo and Zhang (2012) pointed out that the proper number of vanishing moments, size of support, regularity, and wavelet families are important parameters in selecting the optimal wavelet. Hence, we tried several mother wavelets for our hybrid denoising. However, selecting the optimal wavelet is a difficult task. For example,



**Figure 2.** (a) The synthetic signal and its CWT representation and (b) contaminated signal with real seismic noise. The color version of this figure is available only in the electronic edition.

selecting the Shannon wavelet increases the SNR and improves the rmse, but the amplitude of cross-correlation coefficients (CC) decreases. Using the Morlet wavelet attenuates more *P*-wave coda, but improved rmse and CC. However, other wavelets like the Mexican hat wavelet substantially degrade denoising performance. We achieved the best performance (Fig. 3) in terms of preserving the phase arrival shapes and highest SNR and CC with the Bump wavelet. Another parameter that we found important for the denoising is the length of presignal noise. The arrival time can be approximated but with longer presignal noise a more accurate noise level estimation can be made, and as a result, a more efficient denoising occurs.

The superior performance of the hybrid method compared with other methods is clearly shown in Figure 3. Although it was successful in estimating the *P*- and *S*-time-frequency coefficients and removing the noise energy, it also removed part of coda energy after the *P* and *S* waves. A detailed look at the first arriving pulse (Fig. 3c) clearly illustrates the good performance of filtering based on the adaptive BT method, particularly when compared with spectral filtering and hard and soft thresholding. However, the first cycle of the very small and emergent arrival has been slightly flattened although, in general, the polarity of arrival has been preserved (Fig. 3c). Band-pass filtering removes noise with frequencies higher and lower than the signal's frequency range but noise within the frequency range of the signal is untouched (Fig. 3b). Soft and hard thresholding preserved the structure of the signal; however, they were much less effective in noise removal. The qualitative assessment of good performance of our denoising method can be confirmed by

comparing the rmse, the value of the SNR after denoising, and cross-correlation results between the denoised and original waveforms presented in Table 2. In this table, performance of the hybrid BT method (including the pre- and postprocessing) on the synthetic signal is compared with the band-pass filtering and hard and soft thresholding (without pre- and postprocessing).

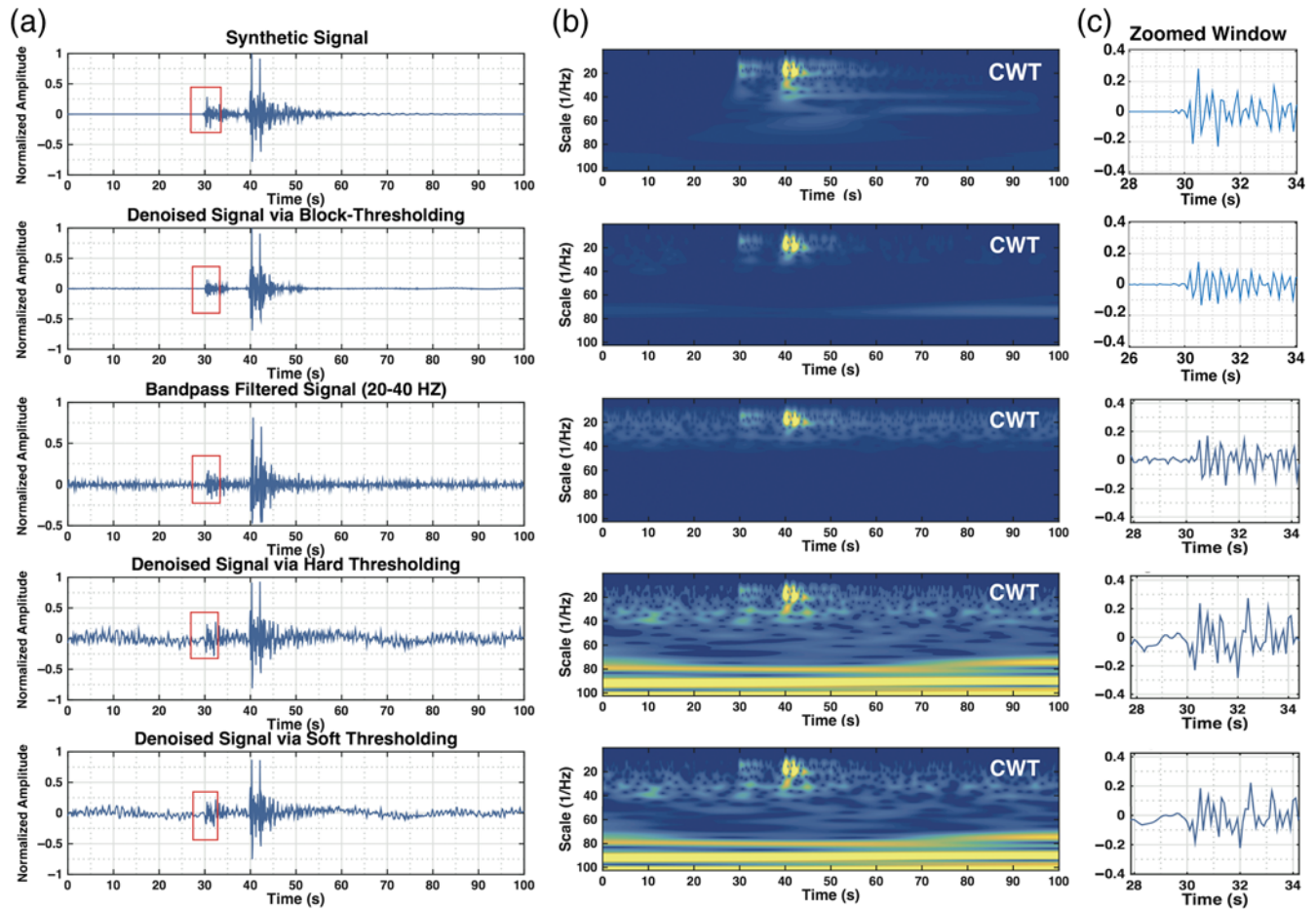
We compare the time-frequency structure of the denoised and original data by performing a cross-wavelet transform (XWT). The XWT is constructed from the wavelet coefficients of the original signal ( $CWT_o$ ) and the denoised signal ( $CWT_d$ ) and will expose their common power and relative phase in time-frequency space. The XWT measures the similarity of the wavelet representations of two signals and provides the ability to account for temporal (or spatial) variability in spectral character. Following Torrence and Compo (1998), the XWT is defined as

$$XWT_{o,d} = CWT_o \times CWT_d^*, \quad (26)$$

in which \* denotes complex conjugation. The cross-wavelet power is defined as  $|XWT_{o,d}|$ , and the complex argument  $\arg(XWT_{o,d})$  can be interpreted as the local relative phase between denoised and original seismograms (Grinsted *et al.*, 2004).

High-power areas in Figure 4 indicate the correlation of high-magnitude coefficients in the denoised and original CWT spectrums. Concentrated energies in 30 and 40 s are associated with *P* and *S* respectively. The XWT shows that the two waveforms are in phase in all the sectors with





**Figure 3.** (a) Waveforms, (b) CWT spectrums, and (c) zoomed windows around the first arrival. The top row shows time series and the CWT for the synthetic signal without noise to compare with results from different denoising methods in the rows below. The method is annotated in the left panels. The right panel shows a zoomed time window of the first arrival within the box in the left panel for each row of figures. The color version of this figure is available only in the electronic edition.

significant common power (Fig. 4). Outside of the area with significant power the phase relationship is complicated but mostly antiphase. This shows that a strong link exists between the denoised and original data and that the denoising method preserves the time–frequency structure of *P* and *S*

Table 2

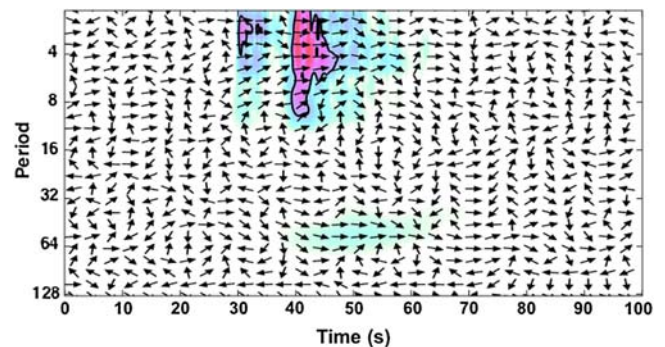
Root Mean Square Error (rmse), SNR, and Maximum CC between Denoised and Original Signal from the Synthetic Test Using Different Thresholding Methods

	rmse	SNR	CC
Block thresholding	0.027	42.831	0.935
Band-pass filtering	0.063	5.441	0.683
Hard thresholding	0.061	2.864	0.796
Soft thresholding	0.048	3.457	0.833

CC, correlation coefficients. In this table, performance of the hybrid block thresholding method (including the pre- and postprocessing) on the synthetic signal is compared to the band-pass filtering and hard and soft thresholding (without pre- and postprocessing). Signal-to-noise ratio (SNR) is measured as the root mean square amplitude in a time window around the signal to a same length window of preceding noise.

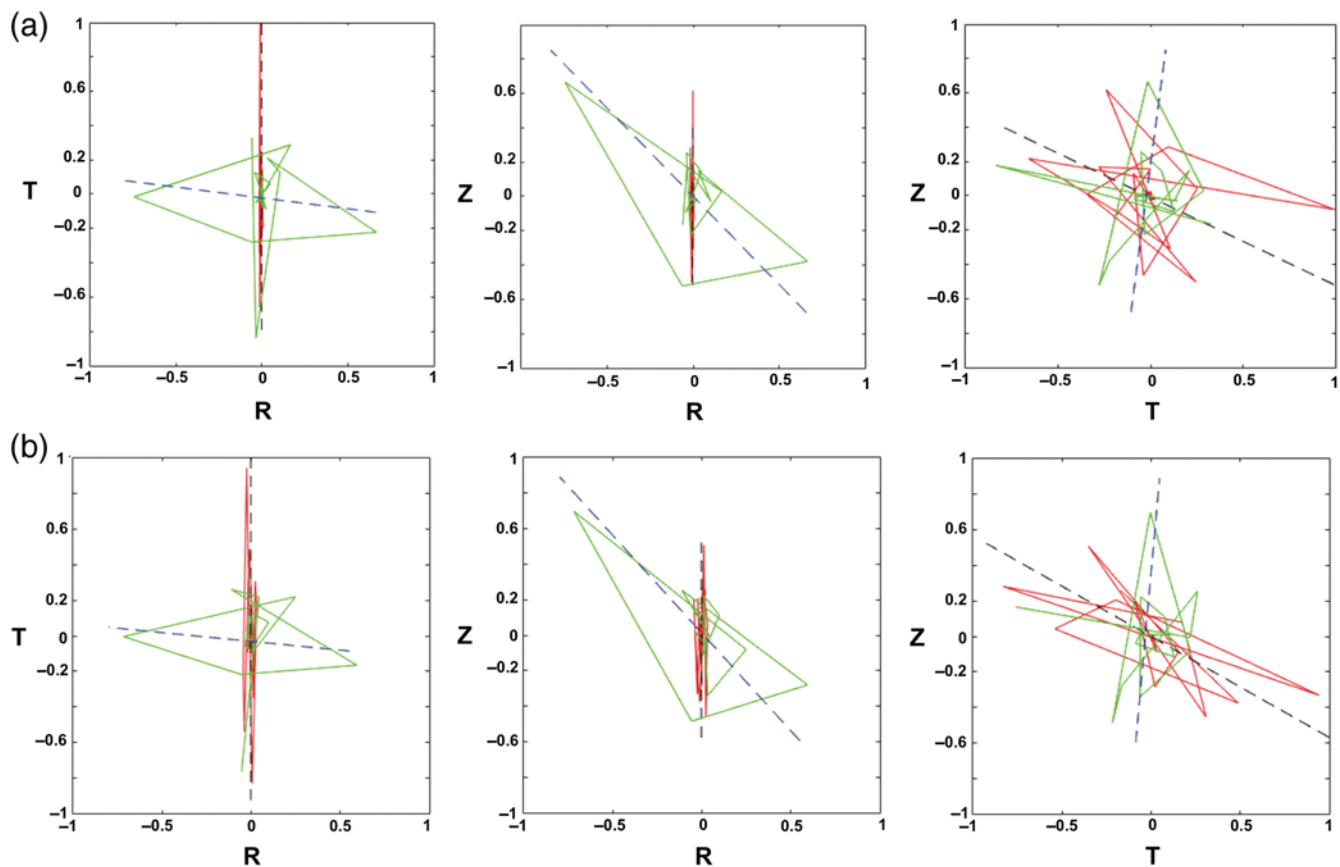
energy without affecting the phase information of the original data.

To check the effect of the denoising method on the polarization of the seismic data, we construct the hodogram of the



**Figure 4.** The cross-wavelet transform (XWT) of the original and denoised data. The phase arrows show the relative phasing of the two time series. Arrows pointing right: in-phase, left: antiphase, down: original phase leading the denoised phase by 90°, and up: denoised phase leading original phase by 90°. The color version of this figure is available only in the electronic edition.





**Figure 5.** (a) Original signal and (b) denoised one. The  $P$ - and  $S$ -wave particle motions are displayed (radial component [R] versus transverse [T] versus vertical [Z]). Dashed lines show the average particle motions as computed using principle component analysis. The  $P$ - and  $S$ -wave polarization azimuths, and the angle between  $P$ - and  $S$ -wave particle motions for the seismogram before the denoising are measured  $0^\circ$ ,  $-79^\circ$ , and  $94^\circ$ , respectively, and changes to  $0^\circ$ ,  $-86^\circ$ , and  $101^\circ$  for the denoised seismograms. The color version of this figure is available only in the electronic edition.

$P$ - and  $S$ -wave windows of three component data. The particle motions of  $P$  and  $S$  windows before and after the denoising are presented in Figure 5. These results show the overall polarization remained unchanged after the denoising.

### Real Seismic Data

We have applied the denoising method to real seismic data including a microseismic event induced during waste water injection in Arkansas recorded by a broadband seismometer at the surface, a regional earthquake ( $M_w$  3.9) in eastern Canada, and an  $M$  4.3 earthquake recorded on an OBS from the Cascadia experiment (Table 3). Applying the hybrid

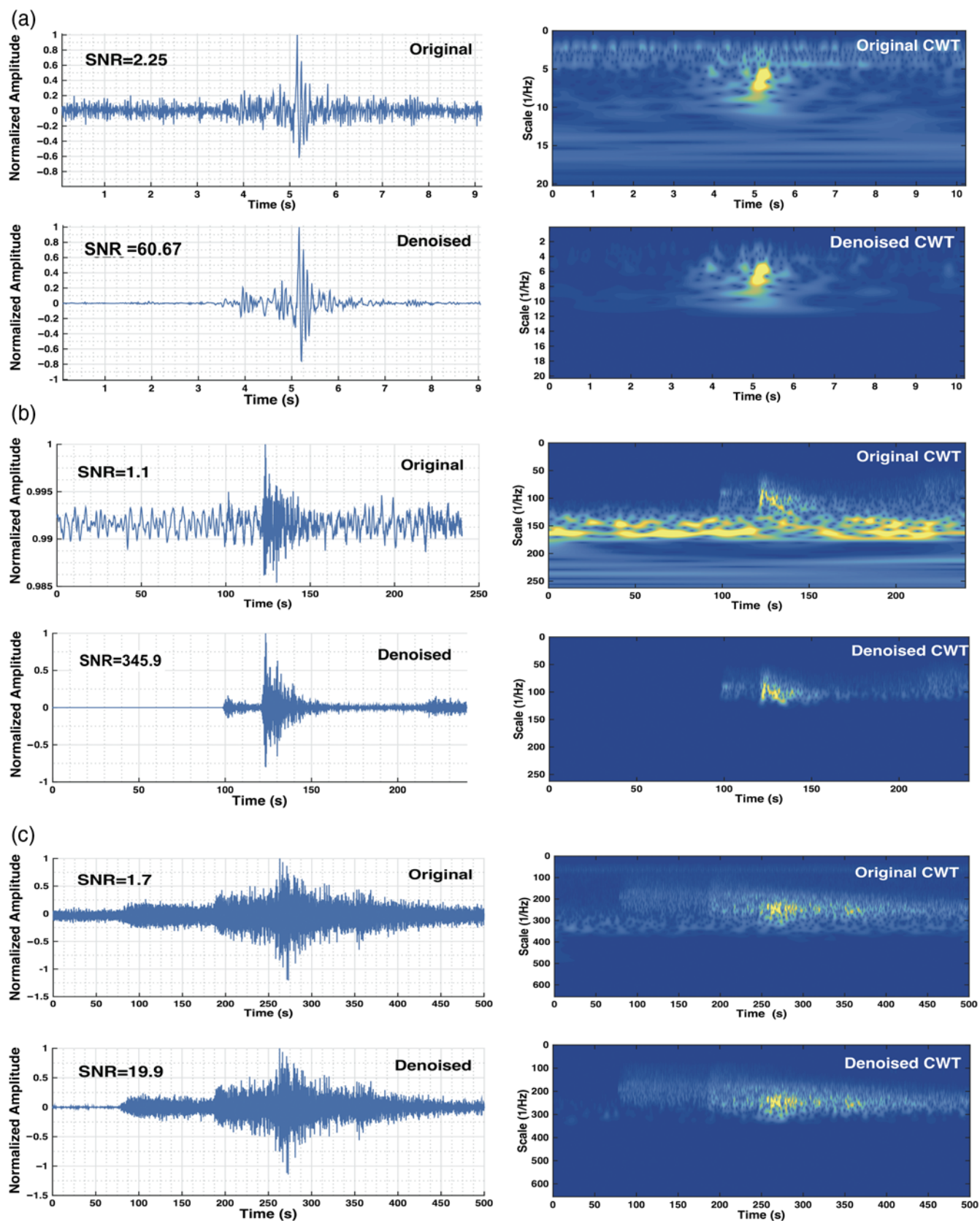
BT algorithm shows a significant improvement in the SNR in the order of several magnitudes for all cases (Fig. 6).

In the case of the microearthquake (Fig. 6a), noises with higher- and lower-frequency contents, compared with the signal frequency, were attenuated from time–frequency representation of data. High-energy noise components at 19 and 27 Hz (Fig. 7a) can be associated with strong electronic noise. Removing these high-magnitude noises from the spectra indicates that the proposed method is not just limited to random noise. Single channel application of the method makes it possible to detect and remove this type of noise that might not be coherent across an array. SNR was increased from 2.25 to 60.67. This can improve the detection of small

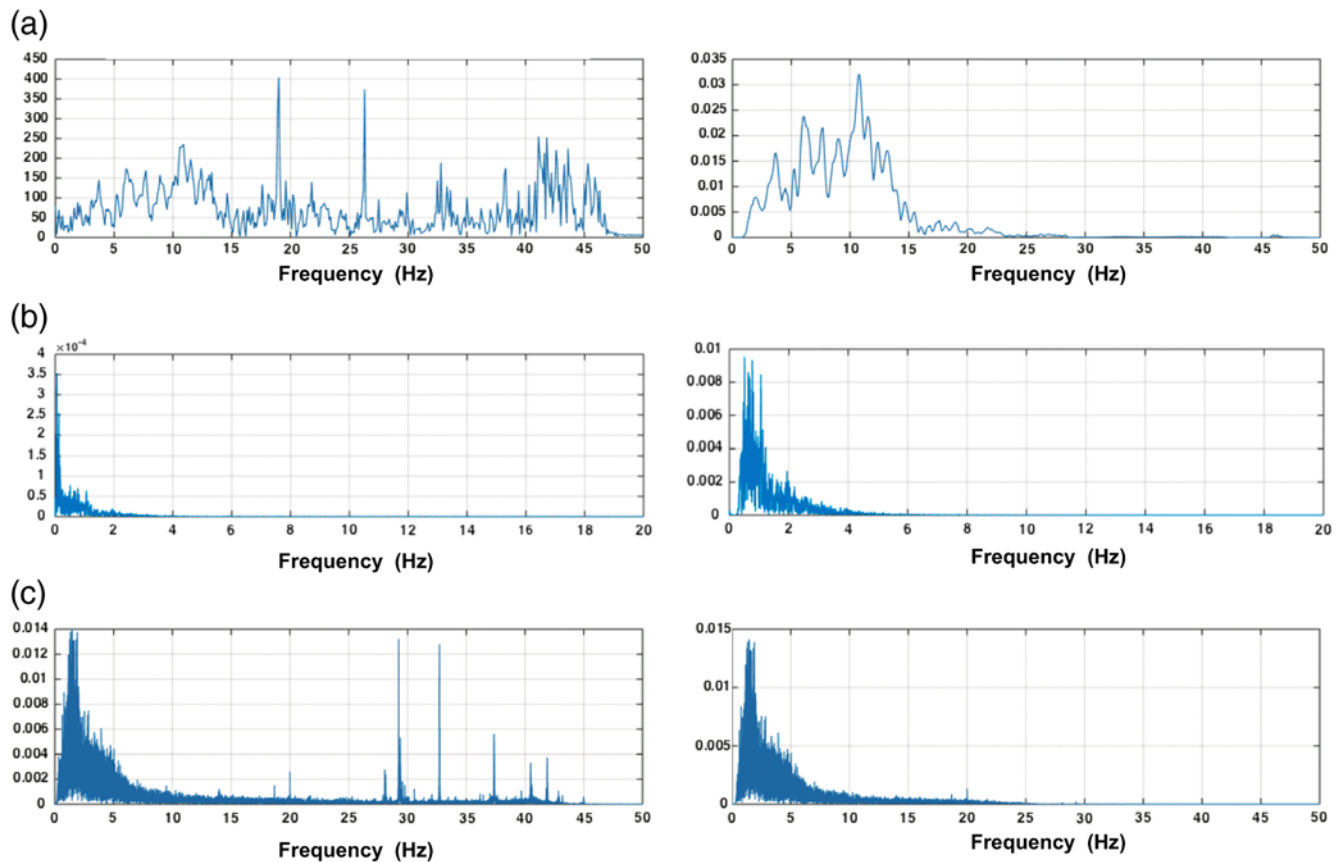
Table 3  
Information for Events Used in This Study

Event	Date	Time (hh:mm:ss)	Latitude ( $^\circ$ )	Longitude ( $^\circ$ )	Depth (km)	Magnitude	Station
Regional	8 July 2012	11:13:36	52.04	$-60.863$	11	$3.9 M_w$	MOQ
Local OBS	20 February 2013	12:17:30	40.41	$-127.035$	10	$4.3 M$	7D.FS02B
Microearthquake	25 October 2010	05:39:14	35.29	$-92.3245$	4.4	$0.5 M$	ARK1

OBS, oceanic bottom seismometer.



**Figure 6.** (a) Induced microearthquake, (b) local earthquake recorded by oceanic bottom seismometer (OBS), and (c) regional earthquake. Event data are listed in Table 3 and vertical components of motion displayed. Each major panel (a through c) shows the original time-series data in the upper left panel and its CWT to the right. Below are the denoised seismogram and its CWT for comparison. The color version of this figure is available only in the electronic edition.



**Figure 7.** (a) Single side amplitude spectra for microearthquake, (b) OBS data, and (c) regional earthquake (left) for raw data and (right) after denoising using proposed method. There are high-power bursts at 19 and 27 Hz in the spectra of raw microseismic data associated with electronic noise. The proposed method removed these bursts and was also successful in removing high-power low-frequency noises in OBS data and high-power high-frequency noises in the regional data. The color version of this figure is available only in the electronic edition.

events, which is a challenging problem in surface monitoring of microseismic events (Eisner *et al.*, 2010), using single channel data. On the other hand, in addition to preserving main features of the seismic signal, such as the waveform and polarity of the surface wave, a more clear *P* and *S* wave provided after the denoising makes phase arrival-time estimations more robust. Moreover, some small amplitude features such as the emergent arrival at 3.5 s (denoised waveform in Fig. 6a) that can be associated to a headwave, were totally buried under the background noise before the denoising. This can facilitate picking of the first arrival times and their polarity which has special importance for the source location and fracture imaging. Thus, denoising microearthquakes can help improve the detection, location, and understanding of the wave propagation, which are crucial steps in microseismic processing.

In OBS data (Fig. 6b), noise was mainly coherent, long-period, and high-amplitude (ground-roll noise). This high-power noise (Fig. 7b) at the higher scales is not attenuated using traditional wavelet thresholding. On the other hand, if we use a simple high-pass filter for removing this noise, it is not easy to choose the corner frequency because higher corners will damage some important features of the signal,

whereas selecting a lower corner frequency will fail to remove enough noise. However, the proposed hybrid scheme was successful in attenuating these high-power noises in addition to the finer high-frequency noises that have overlap with the frequencies of the seismic event and increasing the SNR from 1.1 to 345.92.

High-energy spikes have been removed from spectra after the denoising (Fig. 7c). More clear *P*- and *S*-wave arrivals were obtained after the denoising of regional waveform (Fig. 6c) and SNR was increased from 1.7 to 19.9. A relatively low-frequency band of noise can be identified from a spectrum of original data. This noise gradually merges into the time-frequency structure of signal (spectrum in Fig. 6c). Denoising algorithm identified and removed this noise from the data. This can assemble the spectral content of dominant energy in the signal (between 250 and 300 s, associated with *Lg*), which has special importance for regional attenuation studies.

## Discussion

Development and application of the denoising technique as proposed here will ultimately improve signal detection,

seismogram composition studies, and source discrimination for small local/regional seismic sources. Denoising the signal in this manner will significantly lower detection thresholds and allow small seismic events to be detected more easily, either through conventional detection algorithms such as short-term average/long-term average or more advance schemes in the time–frequency domain.

It has been well known that the typical background seismic noise level is much higher at the seafloor than on land, especially at frequencies below 1 Hz (e.g., Sutton and Barstow, 1990; Romanwicz *et al.*, 1998; Webb, 1998; Webb and Crawford, 1999). Efforts have been concentrated on attenuating some specific types of noises that have dominant effects on the SNR of data in seafloor seismic measurements. Some examples include the proposed method of Crawford and Webb (2000) for identifying and removing tilt noises with frequencies less than 0.1 Hz from vertical data or the proposed method of Chen *et al.* (2015) for attenuating of ground-roll noise. The effectiveness of our hybrid algorithm for denoising of OBS data has been shown in the previous section. However, the proposed method of this study is not limited to any specific type of noise and effectively can identify and remove noises from data.

Another application of the proposed method can be in local or regional attenuation studies. Denoising the signal can assemble the spectral content of the phase of interest more precisely and decrease the uncertainties in Q estimation (McNamara *et al.*, 2012; Mousavi *et al.*, 2014), improving the ability of models to match observed amplitudes. In ambient noise studies, earthquake signal energy needs to be taken out of the noise in a preprocessing step. However, this can be done using a similar thresholding method in a reverse manner to kill the buried signals and keep the noise. After ambient noise cross correlation, our denoising approach can be used again to further improve the SNR of obtained Green's functions. An interesting extension of this technique could be in application to phased seismic array data. Data from regional seismic array stations add an important dimension to wave analysis because wave apparent velocity can theoretically be determined, helping in phase identification. Our denoising technique can be used directly on each element of an array and may significantly improve normal beam forming. However, wave apparent velocity (or slowness) represents an additional transform (into the slowness domain) that can be added to the time–frequency formulation.

## Conclusions

We have introduced a powerful adaptive seismic denoising method based on the CWT, HOS, BT, and Wiener filtering. Our approach of seismic denoising is a hybrid method that integrates the power of these techniques to effectively suppress noise in the signal and increase the output SNR. Our approach takes advantage of a preprocessing step based on the HOS. Kurtosis operates as a Gaussianity estimator and the signal's Gaussian components are detected and removed.

This step improves performance of wavelet thresholding by removing high-power noise components. BT improves the efficiency of wavelet thresholding by considering information of neighboring coefficients on the time–frequency plane. These steps make the thresholding scheme an adaptive, data-driven approach in the sense that all the filter parameters for denoising are dynamically adjusted to the characteristics of the signal. This denoising method is adaptive and robust to signal type variations and can improve the SNR. In a postprocessing step, a Wiener filter is designed based on the thresholded coefficients and applied to the time–frequency coefficients to improve denoising efficiency. Our seismic denoising method has been applied to both synthetic and real seismic data. Results show that the new method outperforms spectral filtering and classical wavelet thresholding methods and considerably improves the SNR of the waveforms without affecting the phase or polarization information of the signal.

## Data and Resources

All seismic data used in this study can be obtained from the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC) at [www.iris.edu](http://www.iris.edu) (last accessed August 2015). Synthetic seismograms were computed using frequency–wavenumber ( $f$ - $k$ ) synthetic seismogram package, provided by Lupei Zhu, Department of Earth and Atmospheric Sciences, St. Louis University, <http://www.eas.slu.edu/People/LZhu/home.html> (last accessed August 2015).

## Acknowledgments

We thank Paul Ogwari for providing the information about microseismic events. This study was supported by the Air Force Research Laboratory under Contract #FA9453-16-C-0015. We thank two anonymous reviewers for their insightful remarks.

## References

- Abma, R., and J. Claerbout (1995). Lateral prediction for noise attenuation by  $t$ - $x$  and  $f$ - $x$  techniques, *Geophysics* **60**, 1887–1896.
- Al-Yahya, K. M. (1991). Application of the partial Karhunen–Loève transform to suppress random noise in seismic sections, *Geophys. Prospect.* **39**, 77–93, doi: [10.1111/j.1365-2478.1991.tb00302.x](https://doi.org/10.1111/j.1365-2478.1991.tb00302.x).
- Anestis, A. A., and G. Oppenheim (1995). *Wavelet Methods in Statistics*, Vol. 103, Springer-Verlag, New York, New York, 5–14.
- Ansari, A., A. Noorzad, H. Zafarani, and H. Vahidifard (2010). Correction of highly noisy strong motion records using a modified wavelet denoising method, *Soil Dynam. Earthq. Eng.* **30**, 1168–1181.
- Antoniadis, A., J. Bigot, and T. Sapatinas (2001). Wavelet estimators in nonparametric regression, a comparative simulation, study, *J. Stat. Software* **6**, 1–83.
- Askari, R., and H. R. Siahkoobi (2008). Ground roll attenuation using the  $S$  and  $x$ - $f$ - $k$  transforms, *Geophys. Prospect.* **56**, no. 1, 105–114, doi: [10.1111/j.1365-2478.2007.00659.x](https://doi.org/10.1111/j.1365-2478.2007.00659.x).
- Beenamol, M., S. Prabavathy, and J. Mohanalin (2012). Wavelet based seismic signal de-noising using Shannon and Tsallis entropy, *Comput. Math. Appl.* **64**, 3580–3593.
- Bekara, M., and M. van der Baan (2009). Random and coherent noise attenuation by empirical mode decomposition, *Geophysics* **74**, V89–V98, doi: [10.1190/1.3157244](https://doi.org/10.1190/1.3157244).



- Bickel, P. J., and K. A. Doksum (1977). *Mathematical Statistics*, Holden-Day, San Francisco, California.
- Bogiatzis, P., and M. Ishii (2015). Continuous wavelet decomposition algorithms for automatic detection of compressional- and shear-wave arrival times, *Bull. Seismol. Soc. Am.* **105**, no. 3, 1628–1641, doi: [10.1785/0120140267](https://doi.org/10.1785/0120140267).
- Bonar, D., and M. Sacchi (2012). Denoising seismic data using the nonlocal means algorithm, *Geophysics* **77**, A5–A8.
- Cai, T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach, *Ann. Stat.* **27**, 898–924.
- Cai, T., and B. W. Silverman (2001). Incorporation information on neighboring coefficients into wavelet estimation, *Sankhya* **63**, 127–148.
- Cai, T., and H. Zhou (2009). A data-driven block thresholding approach to wavelet estimation, *Ann. Stat.* **37**, no. 2, 569–595, doi: [10.1214/07-AOS538](https://doi.org/10.1214/07-AOS538).
- Canales, L. L. (1984). Random noise reduction, *SEG Tech. Program Expanded Abstr.* **3**, 525–527.
- Chen, Y., S. Jiao, J. Ma, H. Chen, Y. Zhou, and S. Gan (2015). Ground-roll noise attenuation using a simple and effective approach based on local band-limited orthogonalization, *Geosci. Rem. Sens. Lett. IEEE* **12**, no. 11, 2316–2320, doi: [10.1109/LGRS.2015.2475280](https://doi.org/10.1109/LGRS.2015.2475280).
- Chik, Z., T. Islam, S. A. Rosyidi, H. Sanusi, M. R. Taha, and M. M. Mustafa (2009). Comparing the performance of Fourier decomposition and wavelet decomposition for seismic signal analysis, *Eur. J. Scientific Res.* **32**, 314–328.
- Cohen, I. (2002). Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator, *IEEE Signal Process. Lett.* **9**, 113–116 (April 2002).
- Cohen, I. (2005). Relaxed statistical model for speech enhancement and a priori SNR estimation, *IEEE Trans. Speech Audio Process.* **13**, 870–881 (September 2005).
- Coifman, R., and D. L. Donoho (1995). Translation invariant de-noising, in *Wavelets and Statistics (Lecture Notes in Statistics)*, A. Antoniadis and G. Oppenheim (Editors), Springer-Verlag, New York, New York, 125–150, doi: [10.1007/978-1-4612-2544-7\\_9](https://doi.org/10.1007/978-1-4612-2544-7_9).
- Crawford, W. C., and S. C. Webb (2000). Identifying and removing tilt noise from low-frequency (<0.1 Hz) seafloor vertical seismic data, *Bull. Seismol. Soc. Am.* **90**, no. 4, 952–963.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, Vol. 61, SIAM, Philadelphia, Pennsylvania, 357 pp., doi: [10.1137/1.9781611970104](https://doi.org/10.1137/1.9781611970104).
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis, *Psychol. Meth.* **2**, 292–307.
- DeVore, R., B. Jawerth, and V. Popov (1992). Compression of wavelet decompositions, *Am. J. Math.* **114**, 737–785.
- Ditomaso, R., M. Mucciarelli, M. R. Gallipoli, and F. C. Ponso (2010). Effect of a single vibrating building on free-field ground motion: Numerical and experimental evidences, *Bull. Earthq. Eng.* **8**, no. 3, 693–703, doi: [10.1007/s10518-009-9134-5](https://doi.org/10.1007/s10518-009-9134-5).
- Ditomaso, R., M. Mucciarelli, and F. C. Ponso (2012). Analysis of non-stationary structural systems by using a band-variable filter, *Bull. Earthq. Eng.* **10**, no. 3, 895–911, doi: [10.1007/s10518-012-9338-y](https://doi.org/10.1007/s10518-012-9338-y).
- Djarfour, N., T. Aifa, K. Baddari, A. Mihoubi, and J. Ferahtia (2008). Application of feedback connection artificial neural network to seismic data filtering, *Compt. Rendus Geosci.* **340**, no. 6, 335–344, doi: [10.1016/j.crte.2008.03.003](https://doi.org/10.1016/j.crte.2008.03.003).
- Donoho, D. (1995). De-noising by soft-thresholding, *IEEE Trans. Inform. Theor.* **41**, 613–627.
- Donoho, D., and I. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**, 425–455.
- Donoho, D. L., and I. M. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage, *J. Am. Stat. Assoc.* **90**, 1200–1224.
- Douglas, A. (1997). Bandpass filtering to reduce noise on seismograms: Is there a better way? *Bull. Seismol. Soc. Am.* **87**, 770–777.
- Eisner, L., D. Abbott, W. Barker, J. Lakings, and M. Thornton (2008). Noise suppression for detection and location of microseismic events using a matched filter, *78th Annual Meeting of the Society of Exploration Geophysicists*, Las Vegas, Nevada, 9–14 November 2008, 1431–1435, Expanded Abstracts.
- Eisner, L., V. Grechka, and S. Williams-Stroud (2010). Future of microseismic analysis: Integration of monitoring and reservoir simulation, *AAPG Hedberg Conference*, Austin, Texas, 5–10 December 2010.
- Elad, M., and M. Aharon (2006). Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Process.* **15**, 3736–3745.
- Ephraim, Y., and D. Malah (1984). Speech enhancement using a minimum mean square error short-time spectral amplitude estimator, *IEEE Trans. Acoust. Speech Signal Process.* **32**, 1109–1121.
- Farge, M. (1992). Wavelet transforms and their applications to turbulence, *Annu. Rev. Fluid Mech.* **24**, 395–458, doi: [10.1146/annurev.fl.24.010192.002143](https://doi.org/10.1146/annurev.fl.24.010192.002143).
- Galiana-Merino, J. J., J. Rosa-Herranz, J. Giner, S. Molina, and F. Rotella (2003). De-noising of short period seismograms by wavelet packet transform, *Bull. Seismol. Soc. Am.* **93**, 2554–2562.
- Ghael, S., A. Sayeed, and R. Baraniuk (1997). Improved wavelet denoising via empirical Wiener filtering, *Proc. of SPIE, Math. Imag.*, San Diego, California, July 1997.
- Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* **165**, 149–166.
- Grinsted, A., J. C. Moore, and S. Jevrejeva (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series, *Nonlinear Process. Geophys.* **11**, 561–566.
- Gulunay, N. (1986). FXDECON and complex Wiener prediction filter, *56th Annual Meeting of the Society of Exploration Geophysicists*, Houston, Texas, 2–6 November 1986, 279–281, Expanded Abstracts.
- Hagen, D. C. (1982). The application of principal components analysis to seismic data sets, *Geoexploration* **20**, 93–111, doi: [10.1016/0016-7142\(82\)90009-6](https://doi.org/10.1016/0016-7142(82)90009-6).
- Hall, P., G. Kerkycharian, and D. Picard (1998). Block threshold rules for curve estimation using kernel and wavelet methods, *Ann. Stat.* **26**, 922–942.
- Hall, P., G. Kerkycharian, and D. Picard (1999a). On the minimax optimality of block thresholded wavelet estimators, *Statistica Sinica* **9**, 33–50.
- Hall, P., G. Kerkycharian, and D. Picard (1999b). A note on the wavelet oracle, *Stat. Probab. Lett.* **43**, 415–420.
- Han, J., and M. van der Baan (2015). Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding, *Geophysics* **80**, no. 6, 69–80, doi: [10.1190/GEO2014-0423.1](https://doi.org/10.1190/GEO2014-0423.1).
- Han, L., M. D. Sacchi, and L. Han (2014). Spectral decomposition and de-noising via time-frequency and space-wavenumber reassignment, *Geophys. Prospect.* **62**, no. 2, 244–257, doi: [10.1111/1365-2478.12088](https://doi.org/10.1111/1365-2478.12088).
- Hashemi, H., A. Javaherian, and R. Babuska (2008). A semi-supervised method to detect seismic random noise with fuzzy GK clustering, *J. Geophys. Eng.* **5**, no. 4, 457–468, doi: [10.1088/1742-2132/5/4/009](https://doi.org/10.1088/1742-2132/5/4/009).
- James, W., and C. Stein (1961). Estimation with quadratic loss, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, University of California Press, Berkeley, California, 20 June–30 July 1961, 361–379, <http://projecteuclid.org/euclid.bsm/1200512173> (last accessed September 2015).
- Johnstone, I. M. (1992). Wavelets and the theory of non-parametric function estimation, *Phil. Trans. Roy. Soc. Lond.* **357**, 2475–2493.
- Johnstone, I. M., and B. W. Silverman (1997). Wavelet threshold estimators for data with correlated noise, *J. Roy. Stat. Soc.* **59**, 319–351, ISSN 13697412.
- Jones, I. F., and S. Levy (1987). Signal-to-noise ratio enhancement in multi-channel seismic data via the Karhunen–Loève transform, *Geophys. Prospect.* **35**, no. 1, 12–32, doi: [10.1111/j.1365-2478.1987.tb00800.x](https://doi.org/10.1111/j.1365-2478.1987.tb00800.x).
- Karamzadeh, N., G. Javan Doloei, and A. Reza (2013). Automatic earthquake signal onset picking based on the continuous wavelet transform, *IEEE Trans. Geosci. Remote Sens.* **51**, 2666–2674, doi: [10.1109/TGRS.2012.2213824](https://doi.org/10.1109/TGRS.2012.2213824).
- Kim, N. S., and J. H. Chang (2000). Spectral enhancement based on global soft decision, *IEEE Signal Process. Lett.* **7**, 108–110 (May 2000).
- Kumar, D., and I. Ahmed (2014). Seismic noise, in *Encyclopedia of Solid Earth Geophysics*, H. K. Gupta (Editor), Springer, The Netherlands, 1157–1161, ISBN 978-90-481-8701-0, doi: [10.1007/978-90-481-8702-7\\_146](https://doi.org/10.1007/978-90-481-8702-7_146).

- Luo, G., and D. Zhang (2012). Wavelet denoising, in *Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology*, D. Baleanu (Editor), InTech, doi: [10.5772/37424](https://doi.org/10.5772/37424).
- Mallat, S. (1999). A wavelet tour of signal processing, in *Wavelet Analysis & Its Applications*, Second Ed., Academic Press, San Diego, California, 620 pp.
- Matz, G., and F. Hlawatsch (2000). Minimax robust nonstationary signal estimation based on a p-point uncertainty model, *J. Franklin Inst.* **337**, 403–419.
- McNamara, D. E., M. Meremonte, J. Z. Maharrey, S.-L. Mildore, J. R. Altidore, D. Anglade, S. E. Hough, D. Given, H. Benz, L. Gee, *et al.* (2012). Frequency-dependent seismic attenuation within the Hispaniola Island region of the Caribbean sea, *Bull. Seismol. Soc. Am.* **102**, 773–782, doi: [10.1785/0120110137](https://doi.org/10.1785/0120110137).
- Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, New York, New York, doi: [10.1017/CBO9780511623820](https://doi.org/10.1017/CBO9780511623820).
- Mousavi, S. M., C. H. Cramer, and C. A. Langston (2014). Average  $Q_{Lg}$ ,  $Q_{Sn}$ , and observation of  $Lg$  blockage in the continental Margin of Nova Scotia, *J. Geophys. Res.* **119**, 7722–7744, doi: [10.1002/2014JB011237](https://doi.org/10.1002/2014JB011237).
- Mousavi, S. M., C. A. Langston, and S. P. Horton (2016). Automatic microseismic denoising and onset detection using the synchrosqueezed-continuous wavelet transform, *Geophysics* **81**, no. 4, 1–15, doi: [10.1190/GEO2015-0598.1](https://doi.org/10.1190/GEO2015-0598.1).
- Oropeza, V., and M. Sacchi (2011). Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis, *Geophysics* **76**, V25–V32.
- Parolai, S. (2009). Denoising of seismograms using the S transform, *Bull. Seismol. Soc. Am.* **99**, 226–234, doi: [10.1785/0120080001](https://doi.org/10.1785/0120080001).
- Pazos, A., M. J. Gonzalez, and G. Alguacil (2003). Non linear filter using the wavelet transform applied to seismological records, *J. Seismol.* **7**, 413–429.
- Pinnegar, C. R., and D. W. Eaton (2003). Application of the S transform to prestack noise attenuation filtering, *J. Geophys. Res.* **108**, 2422, doi: [10.1029/2002JB002258](https://doi.org/10.1029/2002JB002258).
- Ravier, P., and P. O. Amblard (2001). Wavelet packets and denoising based on higher-order-statistics for transient detection, *Signal Process.* **81**, 1909–1926.
- Rodriguez, I. V., D. Bonar, and M. Sacchi (2012). Microseismic data denoising using a 3C group sparsity constrained time–frequency transform, *Geophysics* **77**, no. 2, V21–V29, doi: [10.1190/GEO2011-0260.1](https://doi.org/10.1190/GEO2011-0260.1).
- Romanowicz, B., D. Stakes, J. P. Montagner, P. Tarits, R. Urhammer, M. Begnaud, E. Stutzmann, M. Pasyanos, J.-F. Karczewski, S. Etchemendy, *et al.* (1998). MOISE: A pilot experiment towards long term sea-floor geophysical observatories, *Earth Planets Space* **50**, 927–937.
- Scherbaum, F. (1996). Of poles and zeros: Fundamentals in digital seismology, *Modern Approaches in Geophysics*, G. Nolet (Editor), Vol. 15, Kluwer Academic Press, Dordrecht, The Netherlands, 257 pp.
- Schimmel, M., and J. Gallart (2007). Frequency-dependent phase coherence for noise suppression in seismic array data, *J. Geophys. Res.* **112**, no. B04303, doi: [10.1029/2006JB004680](https://doi.org/10.1029/2006JB004680).
- Shuchong, L., and C. Xun (2014). Seismic signals wavelet packet de-noising method based on improved threshold function and adaptive threshold, *Comput. Model. New Tech.* **18**, 1291–1296.
- Siedenburg, K. (2012). Persistent empirical Wiener estimation with adaptive threshold selection for audio denoising, *Proc. of the 9th Sound and Music Comput. Conf.*, Copenhagen, Denmark, 11–14 July 2012, 426–433.
- Sobolev, G., and A. Lyubushin (2006). Microseismic impulses as earthquake precursors: Izvestiya, *Phys. Solid Earth* **42**, 721–733, doi: [10.1134/S1069351306090023](https://doi.org/10.1134/S1069351306090023).
- Stein, C. (1980). Estimation of the mean of a multivariate normal distribution, *Ann. Stat.* **9**, 1135–1151.
- Sutton, G. H., and N. Barstow (1990). Ocean bottom ultra-low frequency (ULF) seismo-acoustic ambient noise: 0.002–0.4 Hz, *J. Acoust. Soc. Am.* **87**, 2005–2012.
- Tary, J. B., R. H. Herrea, J. Han, and M. V. D. Baan (2014). Spectral estimation—What is new? What is next? *Rev. Geophys.* **52**, no. 4, 723–749, doi: [10.1002/2014RG000461](https://doi.org/10.1002/2014RG000461).
- Thakur, G., E. Brevdo, N. S. Fuckar, and H. T. Wu (2013). The synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications, *Signal Process.* **93**, 1079–1094.
- To, C. A., J. R. Moore, and S. D. Glaser (2009). Wavelet denoising techniques with applications to experimental geophysical data, *Signal Process.* **89**, 144–160, doi: [10.1016/j.sigpro.2008.07.023](https://doi.org/10.1016/j.sigpro.2008.07.023).
- Torrence, C., and G. P. Compo (1998). A practical guide to wavelet analysis, *Bull. Am. Meteorol. Soc.* **79**, 61–78.
- Trickett, S. (2008). F-xy cadzow noise suppression, *SEG Tech. Program Expanded Abstr.* **27**, 2586–2590.
- Tselentis, G. A., N. Martakis, P. Paraskevopoulos, A. Lois, and E. Sokos (2012). Strategy for automated analysis of passive microseismic data based on S-transform, Otsu's thresholding, and higher order statistics, *Geophysics* **77**, KS43–KS54, doi: [10.1190/GEO2011-0301.1](https://doi.org/10.1190/GEO2011-0301.1).
- Tsolis, G., and T. Xenos (2011). Signal denoising using empirical mode decomposition and higher order statistics, *Int. J. Signal Process. Image Process. Pattern Recogn.* **4**, 91–106.
- Ursin, B., and Y. Zheng (1985). Identification of seismic reflections using singular value decomposition, *Geophys. Prospect.* **33**, 773–799, doi: [10.1111/j.1365-2478.1985.tb00778.x](https://doi.org/10.1111/j.1365-2478.1985.tb00778.x).
- Webb, S. C. (1998). Broadband seismology and noise under the ocean, *Rev. Geophys.* **36**, no. 1, 105–142.
- Webb, S. C., and W. C. Crawford (1999). Long period seafloor seismology and deformation under ocean waves, *Bull. Seismol. Soc. Am.* **89**, no. 6, 1535–1542.
- Wolfe, P. J., and S. J. Godsill (2001). Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement, *Proc. IEEE Workshop Stat. Signal Process.*, Orchid Country Club, Singapore, 6–8 August 2001, 496–499.
- Yilmaz, Ö. (1987). *Seismic Data Processing*, First Ed., Society of Exploration Geophysics, Tulsa, Oklahoma.
- Yilmaz, Ö. (2001). *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*, Second Ed., Society of Exploration Geophysicists, Tulsa, Oklahoma.
- Yoon, B. J., and P. P. Vaidyanathan (2004). Wavelet-based denoising by customized thresholding, *IEEE Int. Conf. On Acoustic Speech Signal Process. (ICASSP)*, Montreal, Québec, 17–21 May 2004.
- Yu, G., S. Mallat, and E. Bacry (2008). Audio denoising by time–frequency block thresholding, *IEEE Trans. Signal Process.* **56**, 1830–1839.
- Zhang, W., and X. H. Zhao (2001). Wavelet thresholding using higher order statistics for signal denoising, *Proceedings of the International Conference on Info-Tech and Info-Net (ICII '01)*, Beijing, China, 29 October–1 November, Vol. 1, 363–368.
- Zhu, L., and L. A. Rivera (2002). A note on the dynamic and static displacements from a point source in multi-layered media, *Geophys. J. Int.* **148**, 619–627.

University of Memphis  
Center for Earthquake Research and Information (CERI)  
3876 Central Avenue, Suite 1  
Memphis, Tennessee 38152  
smousavi@memphis.edu

Manuscript received 29 November 2015;  
Published Online 7 June 2016