On-Line Multiscale Filtering of Random and Gross Errors without Process Models

Mohamed N. Nounou and Bhavik R. Bakshi
Dept. of Chemical Engineering, The Ohio State University, Columbus, OH 43210

Data Rectification by univariate filtering is popular for processes lacking an accurate model. Linear filters are most popular for online filtering; however, they are single-scale best suited for rectifying data containing features and noise that are at the same resolution in time and frequency. Consequently, for multiscale data, linear filters are forced to trade off the extent of noise removal with the accuracy of the features retained. In contrast, nonlinear filtering methods, such as FMH and wavelet thresholding, are multiscale, but they cannot be used for online rectification. A technique is presented for online nonlinear filtering based on wavelet thresholding. OLMS rectification applies wavelet thresholding to data in a moving window of dyadic length to remove random errors. Gross errors are removed by combining wavelet thresholding with multiscale median filtering. Theoretical analysis shows that OLMS rectification using Haar wavelets subsumes mean filters of dyadic length, while rectification with smoother boundary corrected wavelets is analogous to adaptive exponential smoothing. If the rectified measurements are not needed online, the quality of rectification can be further improved by averaging the rectified signals in each window, overcoming the boundary effects encountered in TI rectification. Synthetic and industrial data show the benefits of the online multiscale and boundary corrected translation invariant rectification methods.

Introduction

Measured process data are usually contaminated by random and gross errors due to sensor noise, disturbances, instrument degradation, and human errors. Since the performance of process operation tasks depends on the quality of information extracted from the measured data, the collected data need to be cleaned or ''rectified'' for efficient process operation. Data rectification is an ill-posed problem since given just the measured data, they cannot be rectified without additional information. Depending on the type of additional information used, techniques for data rectification may be categorized as: *rectification with fundamental process models, rectification with empirical process models, and rectification without process models.*

Rectification with fundamental process models minimizes the error between the measured and the rectified signals while requiring the rectified variables to satisfy a linear or nonlinear, steady-state or dynamic model. Methods in this category have received significant attention in academic research and often consist of the separate tasks of gross error detection and data reconciliation. Data reconciliation refers to the model-based removal of random errors, whereas data rectification is a more general term which includes the reduction of random as well as gross errors with or without process models (Kramer and Mah, 1994). For data from steady-state processes, data reconciliation may be performed as a weighted least-squares estimation where the sum of square errors normalized by the error covariance is minimized (Tamhane and Mah, 1985; Mah 1990). Crowe (1989) used a transformation called matrix projection in multivariate rectification problems with linear constraints, which exploits the redundancy among dependent process variables to eliminate Gaussian errors from data under steady-state conditions. Veverka (1992) extended the same approach to problems with nonlinear constraints. For dynamic process data, a wide range of approaches has been employed for data reconciliation. Kalman Filters are popular for the reconciliation of linear dynamic processes (Gelb, 1974; Sorenson, 1985). When a nonlinear process model is available, data may be rectified by an ex-

Correspondence concerning this article should be addressed to B. R. Bakshi.

tended Kalman filter or constrained nonlinear programming (Islam et al., 1994; Liebman and Edgar 1992; Tjao and Biegler 1991; Kim et al., 1990) or a sequential modular approach (Chiari et al., 1997). When fundamental process models are used for data rectification, the quality of the rectified data depends on the accuracy of the process models. Rectification with empirical process models overcomes the need for accurate fundamental process models by extracting the relationship between the variables by an empirical modeling method such as principal component analysis (PCA) (Kramer and Mah, 1994) or recurrent neural networks (Karjala and Himmelblau, 1994).

Data containing non-Gaussian or gross errors may be rectified by applying various statistical outlier detection techniques such as hypothesis testing and a generalized likelihood ratio approach (Mah, 1990). These tests may be applied to autocorrelated measurements by prewhitening the measurements using time-series modeling (Kao et al., 1990, 1992). Gross errors in multivariate data may also be detected by PCA (Tong and Crowe, 1995). For data from dynamic processes, robust reconciliation methods have been developed by Albuquerque and Biegler (1995) and Kim et al. (1997). A maximum likelihood approach has also been developed by Johnston and Kramer (1995) and Tjao and Biegler (1991) that removes random and gross errors simultaneously based on their probability distribution.

Since accurate process models relating the hundreds of measured variables are not easily obtained, the simplest and most widely used rectification methods do not rely on a fundamental or empirical process model. These methods rely on information about the nature of the errors or smoothness of the underlying signal and include various univariate filtering methods. Linear, univariate low pass filtering techniques such as mean filtering and exponential smoothing are the most commonly used methods in the chemical industry (Tham and Parr, 1994) as they are simple and can be easily used online. Another popular class of discrete linear filters can be designed using bilinear transformation of analog linear filters that have the desired frequency characteristics, and includes Butterworth and Chebyshev filters (Bose, 1985). Unfortunately, these linear filtering methods represent the data at a single scale or resolution in time and frequency which forces them to trade-off the extent of rectification with the quality of the rectified local features. Consequently, linear filters are not very effective in filtering signals containing features with different localizations in both time and frequency. The poor representation of the underlying localized features by linear filtering may be overcome by nonlinear filtering methods, such as finite impulse response (FIR) median hybrid filters (FMH) (Heinonen and Neuvo, 1987), and multiscale wavelet-based rectification (Donoho et al., 1995). FMH filters are most effective when applied to piecewise constant signals contaminated with white noise, require careful selection of the filter length, and are limited to off-line use. Multiscale filtering methods based on wavelet analysis represent the data as a weighted sum of orthonormal wavelets. The multiscale representation captures deterministic features in a few relatively large coefficients, while stationary autocorrelated errors are approximately decorrelated or whitened. Thus, stochastic stationary errors may be decreased with minimum distortion of the underlying deterministic signal by eliminating wavelet coefficients smaller than a threshold. The rectified signal is then recovered by reconstructing the thresholded coefficients to the time domain. This multiscale rectification technique has superior theoretical and practical properties (Donoho et al., 1995), but is restricted to off-line application and to data of dyadic length 2^J. Also, the wavelet representation lacks robustness to outliers or gross errors and has been extended for the rectification of data with such errors by multiscale median filtering (Bruce et al., 1994).

In this article, the existing multiscale approach for rectification of random and gross errors is extended for on-line rectification by processing data in a moving window of dyadic length. The theoretical properties of the resulting on-line multiscale (OLMS) rectification approach are studied to show that the OLMS approach subsumes existing methods such as mean filtering and exponential smoothing. Thus, OLMS rectification can automatically specialize to a linear filter by adapting the scale of the filter to that of each feature in the measured signal. Furthermore, if a time delay in the rectification is acceptable, then the quality of the rectified signal can be improved by averaging the rectified signals obtained in each moving window, resulting in an approach similar to the translation invariant (TI) rectification of Coifman and Donoho (1995), but without the boundary effects that occur in TI rectification. The OLMS and boundary corrected TI (BCTI) rectification are extended to removing non-Gaussian errors by combining them with multiscale median filtering (Bruce et al., 1994). The theoretical properties and application of the resulting method are also discussed and compared with FMH filtering. The performance of the multiscale filtering methods is compared with that of existing methods by Monte-Carlo simulation and application to industrial data. The terms rectification, filtering, and denoising are considered to be equivalent and are used interchangeably.

Data Rectification Without Process Models

Linear filtering

Linear filtering techniques rectify the signal by taking a weighted sum of previous measurements in a window of finite or infinite length. These techniques include finite impulse response (FIR) and infinite impulse response (IIR) filtering and are very popular in the chemical industry, because they are computationally efficient, easy to implement, and can be readily used for on-line rectification (Tham and Parr, 1994).

Linear filters are low pass filters which can be represented as

$$\hat{x}_t = \sum_{i=0}^{I-1} b_i x_{t-i} \tag{1}$$

where I is the filter length and $\{b_i\}$ is a sequence of weighting coefficients, which define the characteristics of the filter and satisfy the following condition

$$\sum_i b_i = 1. \tag{2}$$

The weighting sequence $\{b_i\}$ is the impulse response of the linear filter. Linear filters with finite window size are called

FIR filters. When all weighting coefficients $\{b_i\}$ are equal, the FIR filter reduces to a mean filter. Thus, for a filter of length I, any mean filtered data point can be represented mathematically in terms of the average of the last I measured data points as

$$\hat{x}_t = \frac{1}{I}(x_t + x_{t-1} + \ldots + x_{t-I+1}). \quad (3)$$

The mean filter can also be thought of as a convolution of the measured signal with a vector of I constant coefficients, each equals $\{1/I\}$. The weights used by a mean filter of length $I = 5$ are shown in Figure 1a.

Linear filters with an infinite filter length are called IIR filters. Any rectified data point using an IIR filter is represented as a weighted sum of all previous measurements. Exponentially weighted moving average (EWMA) is a popular IIR filter which smoothes a measured data point by exponentially averaging it with all the previous measurements. The EWMA filter is a recursive low pass filter that eliminates high frequency components from the measured signal. Computationally, it is implemented as

$$\hat{x}_t = \alpha x_t + (1-\alpha)\hat{x}_{t-1} \quad (4)$$

The parameter α is an adjustable smoothing parameter lying between 0 and 1, where a value of 1 corresponds to no smoothing and a value of zero corresponds to keeping only the first measured point. Expanding Eq. 4, $\hat{x}_t$ can be expressed as

$$\hat{x}_t = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \ldots + \alpha(1-\alpha)^{n-1} x_1$$

or

$$\hat{x}_t = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i x_{t-i}. \quad (5)$$

The weighting of the EWMA filter shows that the filter coefficients drop exponentially depending on the smoothing parameter α giving more importance to the more recent measurements, as shown in Figure 1b. IIR filters also include Butterworth filters which can be used to remove white, as well as correlated noise, from stationary stochastic signals. A more detailed discussion of different types of filters is presented by Strum and Kirk (1989).

Drawbacks of Linear Filtering Techniques. The basis functions representing raw measured data have a temporal localization equal to the sampling interval. Linear filters represent the measurements on basis functions with a broader temporal localization and narrower frequency localization, as shown in Figure 2a. These filters are single-scale in nature since all the basis functions have the same fixed time-frequency localization. Consequently, these methods face a tradeoff between the accurate representation of temporally localized changes and efficient removal of temporally global noise. Therefore, simultaneous noise removal and accurate feature representation of nonstationary measured signals cannot be effectively achieved by single-scale filtering methods.

Nonlinear filtering

Nonlinear filtering techniques are multiscale in nature and have been developed to overcome the inability of linear filters to capture features at different scales.

FMH Filtering. A FIR-Median Hybrid Filter (FMH) is a median filter which has a preprocessed input from M linear FIR filters (Heinonen and Neuvo, 1987, 1988). Thus, the

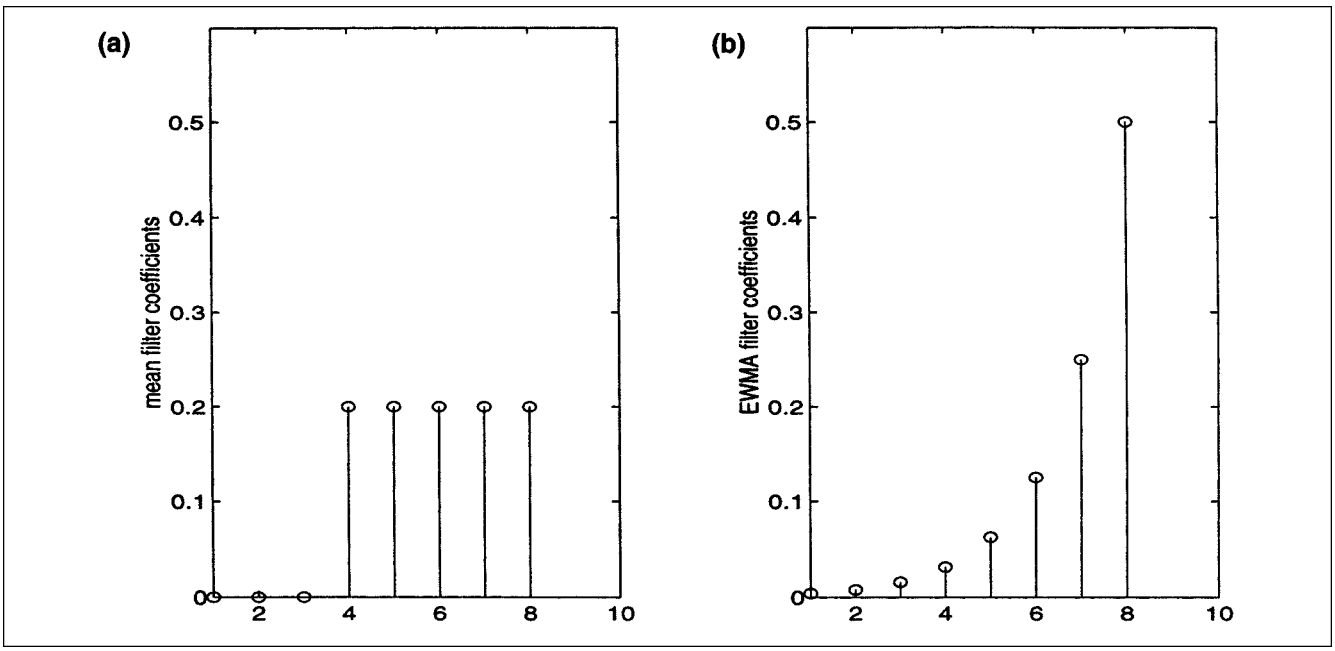

Figure 1. Filter used in (a) mean filtering and (b) exponential smoothing.

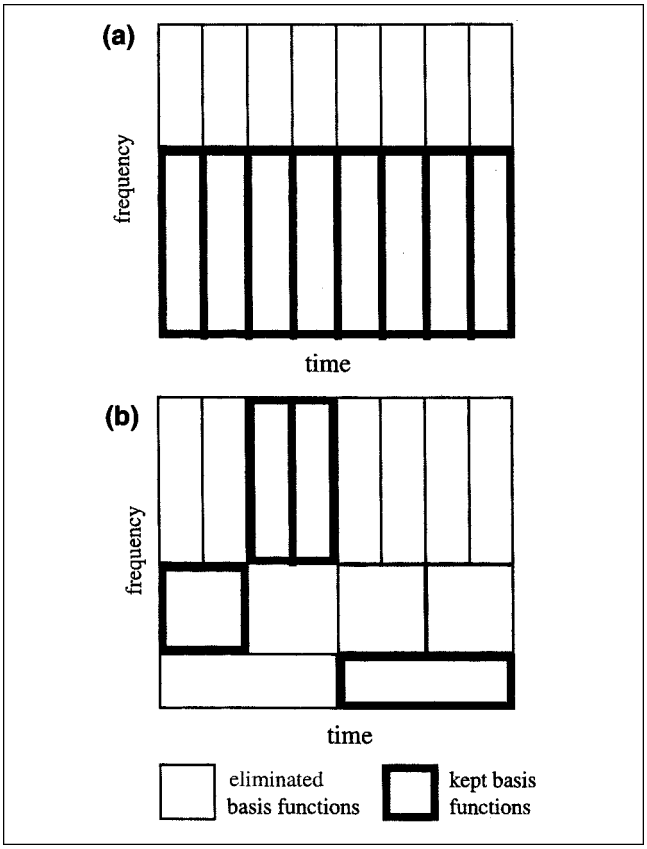

Figure 2. Time-frequency space decomposition by: (a) linear filtering techniques; (b) multiscale techniques.

FMH filter output is the median of only M values, which are the outputs of M FIR filters. An FMH filter of length $2I+1$ with 3 FIR substructures is shown in Figure 3. The three FIR filters used in Figure 3 are

$$y_1 = \left\{ \frac{x_{t-1} + x_{t-2} + \ldots + x_{t-I}}{I} \right\}$$
$$y_2 = x_t$$
$$y_3 = \left\{ \frac{x_{t+1} + x_{t+2} + \ldots + x_{t+I}}{I} \right\} \qquad (6)$$

where $\{x_{t-I}, x_{t-I+1}, \ldots, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \ldots, x_{t+I}\}$ are the $2I+1$ data points filtered by the FMH filter, and y_1, y_3, and y_2 are called the backward predictor, forward predictor, and the center value, respectively. FMH filtering is a batch filtering technique, which is most effective in capturing sharp changes in piece-wise constant signals. The lengths of the FIR filters are chosen to preserve the signal's features while eliminating the high frequency noise. Long FIR filters can oversmooth sharp edges, while short FIR filters may not eliminate enough noise. Since the central point in the FMH filter is the original noisy data, FMH filters tend to retain some noise. Better noise removal is possible by applying the FMH filter several times, which will result in a root signal that does not change with further filtering. Further extensions of FMH

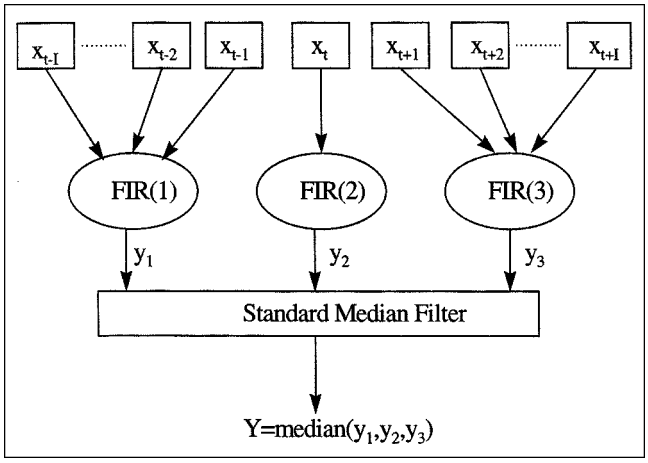

Figure 3. FIR median hybrid filter with three FIR substructures.

filters to using IIR and predictive FIR filters have also been developed (Heinonen and Neuvo, 1988).

FMH filters are also used to eliminate non-Gaussian errors from measured data (Heinonen, 1987, 1988). They are superior to the standard median filters due to their improved ability to preserve temporally localized features, while eliminating errors. However, proper selection of the FIR filters requires knowledge about the maximum duration of the gross errors. When such knowledge is available, the length of the FIR filters used can be determined so that a complete elimination of gross errors is achieved.

Multiscale Rectification. Signals with multiscale features can be analyzed by expressing the signal as a weighted sum of wavelet basis functions. Wavelets are a popular and computationally efficient family of multiscale basis functions. Any square-integrable signal may be represented at multiple scales by decomposition on a family of wavelets and scaling functions, as shown in Figure 4. The signals in Figures 4b, 4d, and 4f are at increasingly coarser scales as compared to the original signal shown in Figure 4a. These scaled signals are determined by filtering the data using a low pass filter H, which is equivalent to projecting the original signal on a set of orthonormal scaling functions of the form

$$\phi_{mk}(t) = \sqrt{2^{-m}}\,\phi(2^{-m}t - k). \qquad (7)$$

Each of the signals in Figures 4c, 4e, and 4g capture the details between the scaled signal at its scale and the scaled signal at the finer scale. These detail signals are computed by filtering the previous scaled signals using a high pass filter G, which is equivalent to projecting the original signal on a set of orthonormal basis functions called wavelets of the form

$$\psi_{mk}(t) = \sqrt{2^{-m}}\,\psi(2^{-m}t - k) \qquad (8)$$

where m, k, L, and n are the dilation parameter, translation parameter, maximum number of scales or decomposition depth, and the length of the original signal, respectively. The original signal can be represented as a sum of all the detail

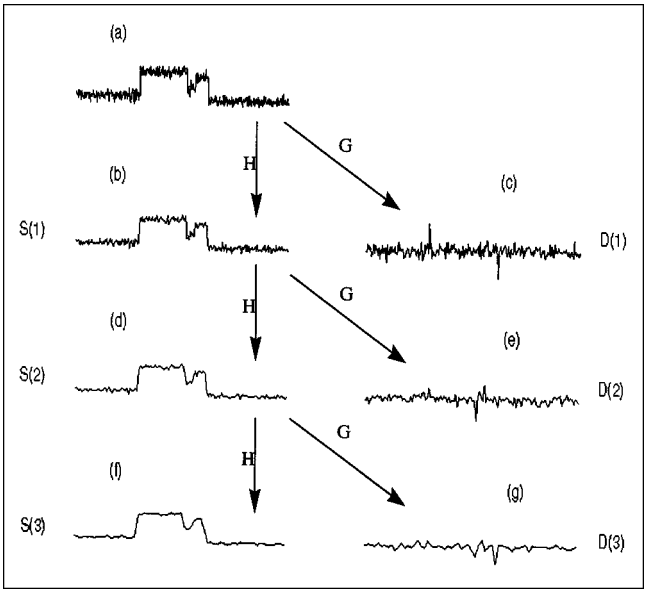

Figure 4. Multiscale decomposition of a stair-step signal using Haar.

signals and the last scaled signal as

$$x(t) = \sum_{k=1}^{n2^{-L}} a_{Lk}\phi_{Lk}(t) + \sum_{m=1}^{L}\sum_{k=1}^{n2^{-m}} d_{mk}\psi_{mk}(t) \qquad (9)$$

where a_{Lk} is the kth scaling function coefficient at the coarsest scale L, and d_{mk} is the kth wavelet coefficient at scale m. Fast wavelet decomposition and reconstruction algorithms of complexity $O(n)$ have been developed for a discrete signal of dyadic length (Mallat, 1989).

Multiscale rectification using wavelets is based on the observation that random errors in a signal are present over all the coefficients, while deterministic changes get captured in a small number of relatively large coefficients. Thus, stationary Gaussian noise may be removed by a three-step method (Donoho et al., 1995):

(1) Transform the noisy signal into the time-frequency domain by decomposing the signal on a selected set of orthonormal wavelet basis functions.

(2) Threshold the wavelet coefficients by suppressing coefficients smaller than a selected value.

(3) Transform the thresholded coefficients back into the original domain.

Donoho and coworkers have studied the statistical properties of wavelet thresholding and have shown that for a noisy signal of length n, the rectified signal will have an error within $O(\log n)$ of the error between the error-free signal and the signal rectified with *a-priori* knowledge about the smoothness of the underlying signal (Donoho and Johnstone, 1994).

Selecting the proper value of the threshold is a critical step in the rectification process, and several methods have been devised. For good visual quality of the rectified signal, the Visushrink method determines the threshold as

$$t_m = \sigma_m\sqrt{2\log n} \qquad (10)$$

where n is the signal length and σ_m is the standard deviation of the errors at scale m, which can be estimated from the wavelet coefficients at that scale by

$$\sigma_m = \frac{1}{0.6745}\,\text{median}\{|d_{mk}|\}. \qquad (11)$$

Other methods for determining the threshold are described by Nason (1996).

The wavelet coefficients may be thresholded by hard or soft thresholding. Hard thresholding eliminates coefficients smaller than a threshold, whereas soft thresholding also shrinks the larger coefficients towards zero by the value of the threshold. Hard thresholding can lead to better reproduction of peak heights and discontinuities, but at the price of occasional artifacts that can roughen the appearance of the rectified signal, while soft thresholding usually gives better visual quality of rectification and less artifacts (Coifman and Donoho, 1995). An artifact, which is not present in the original signal, is created in the reconstructed signal when the wavelet function used to represent a feature in the signal and the feature itself do not align. Such artifacts are due to a localized Gibbs phenomenon which is caused by the lack of translational invariance in orthonormal wavelet decomposition. One approach proposed by Coifman and Donoho (1995) for suppressing artifacts is by shifting the signal several times, rectifying it, and then averaging out all translations to weaken the pseudo-Gibbs phenomena, as shown in Figure 5a. Hard thresholding and translation invariance combined give good visual and quantitative characteristics. This approach, called

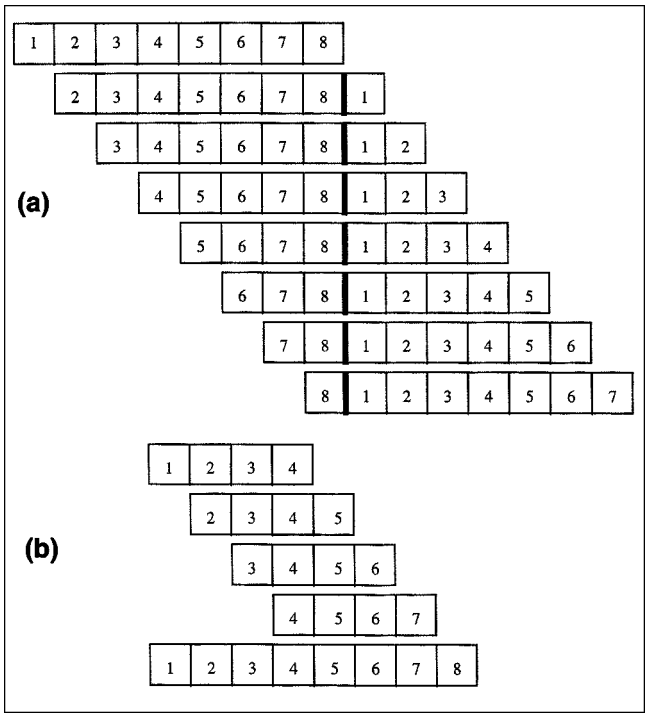

Figure 5. Translation mechanisms used in: (a) TI; (b) BCTI rectification.

Thick lines indicate augmentation of the signal ends.

translation invariant (TI) rectification, does improve the quality and smoothness of the rectified signal, but it suffers from the creation of end effects even when boundary corrected wavelets are used. These end effects are created because TI rectification considers the signal to be a cyclic list resulting in artificial discontinuities if the values of the signal end points do not match as indicated in Figure 5a. This disadvantage will be overcome by the boundary corrected TI (BCTI) rectification method developed in this article.

Wavelet-based multiscale rectification is a very effective approach for denoising signals contaminated by white, as well as correlated stationary Gaussian noise. If the traditional wavelet decomposition algorithm is applied to a signal with non-Gaussian errors, outliers will be present at multiple scales in both the scaled and detailed signals, and large coefficients corresponding to outliers get confused with those corresponding to important features. Thus, wavelet thresholding is not effective in eliminating non-Gaussian errors. This limitation may be overcome by combining wavelet thresholding with multiscale median filtering as in the robust multiscale rectification technique (Bruce et al., 1994). In this robust technique, outliers are eliminated at each scale using median filters as shown in Figure 6. A standard median filter or a FMH filter can be used in this approach, as they are both capable of eliminating gross errors. However, FMH filters usually result in better representation of the signal features. A more detailed discussion on the advantages and limitations of the standard median filter and the FMH filter in the robust multiscale technique is reported by Nounou (1997). The original signal, S(0) passes through a median filter at the finest scale, which results in the signal U(0). Then, the low and high pass wavelet filters H and G are applied on U(0) resulting in the scaled and detail signals at the finest resolution S(1) and D(1), respectively. The same process is repeated on the signal S(1), which passes through the median filter to get U(1) on which the low and high pass wavelet filters are applied to get S(2) and D(2). The process is then repeated to get scaled and detail signals at coarser scales. Due to the dyadic down-sampling used in wavelet transform, the effective median filter length increases at coarser scales. Therefore, outliers in short patches can be eliminated at finer scales while longer patches of outliers can be eliminated at coarser scales. When the effective median filter length at finer scales is shorter than the duration of scaling function coefficients corresponding to a particular outlier patch, outliers will leak into coarser scales. Such leakage may be eliminated by repeated robust multiscale filtering and by selecting a long enough median filter (Bruce et al., 1994). In theory, for a low pass wavelet filter of length M, leakage is prevented by using median filters of length $2M+1$ (Bruce and Gao, 1996). However, when long patches of outliers are present, a complete elimination of outliers can be accomplished by selecting the median filter long enough so that the effective median filter at the coarsest scale is longer than the longest patch of outliers, and by repeating multiscale median filtering. This robust multiscale technique is used in this article.

On-Line Multiscale Rectification

The existing nonlinear rectification techniques described in the previous section do perform better than linear filters for a broad variety of signals. However, a significant disadvantage of these nonlinear or multiscale methods is that they cannot be implemented online. In general, wavelet filters are noncausal in nature and require future measured data for calculating the current wavelet coefficient. This introduces a time delay in the computation that increases at coarser scales and smoother filters. This time delay may be overcome in a rigorous manner by using special wavelets at the edges that eliminate boundary errors while being orthonormal to the other wavelets (Cohen et al., 1993). These boundary corrected filters are causal and require no information about the future to compute wavelet coefficients at the signal end points. Another reason for restricting the wavelet-based methods to off-line use is the dyadic discretization of the wavelet parameters, which requires a signal of dyadic length for the wavelet decomposition. A signal containing a dyadic number of measurements can be decomposed as shown in Figure 7a. In contrast, if the number of measurements is odd, the last point cannot be decomposed without a time delay. For example, when three points are available, as shown in Figure 7b, only the first two data points can be decomposed, introducing a time delay of one data point. However, when seven data points are available, as shown in Figure 7c, only the first four data points can be decomposed to two scales and the next two data points can be decomposed to only one scale. Therefore, a time delay of one data point (point 7) at the finest scale and a time delay of three data points (points 5 through 7) at the coarser scale are introduced. In many applications such a time delay may be unacceptable. Consequently, this section describes an online method for multiscale rectification (OLMS), where absolutely no time delay is

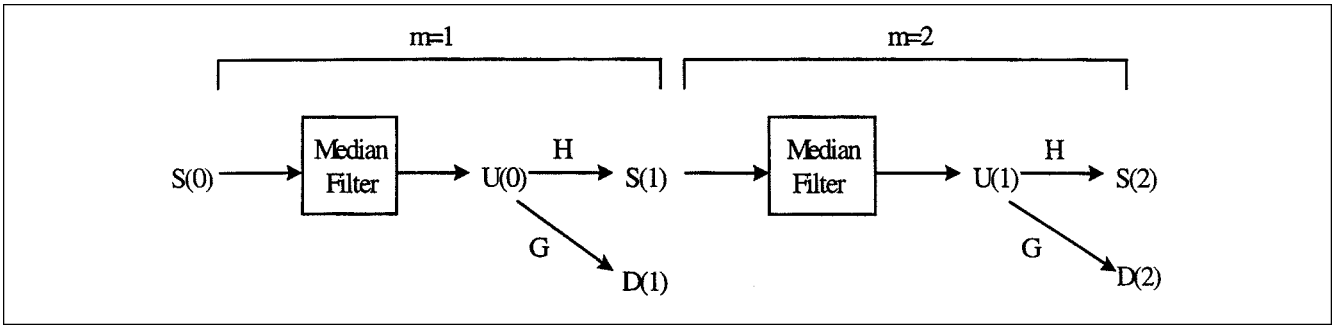

Figure 6. Robust multiscale wavelet decomposition method (Bruce et al., 1994).

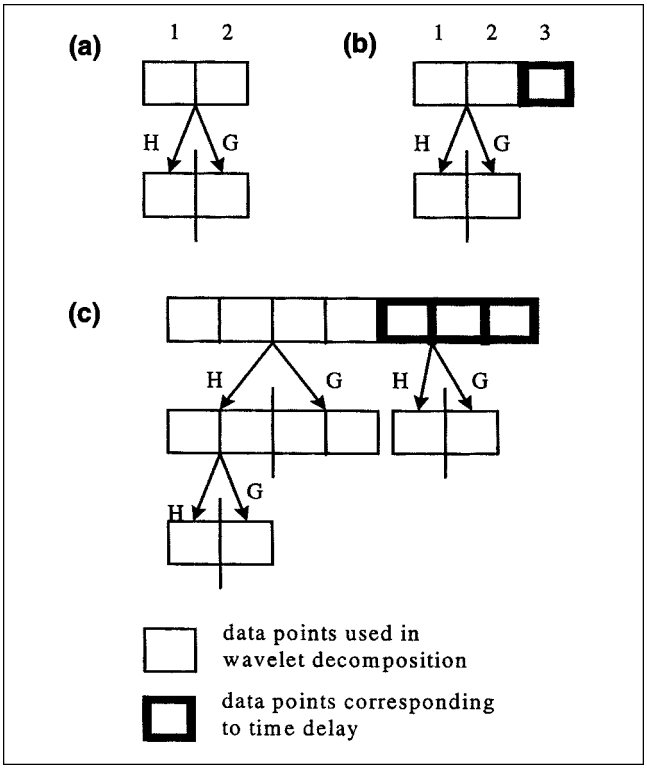

Figure 7. Time delay introduced due to the dyadic length requirement in wavelet decomposition.

allowed. When time delay in the rectification is acceptable, then OLMS rectification reduces to boundary corrected TI (BCTI) rectification. Both of these on-line rectification techniques are also extended to deal with measurements corrupted by non-Gaussian errors using the robust wavelet transform.

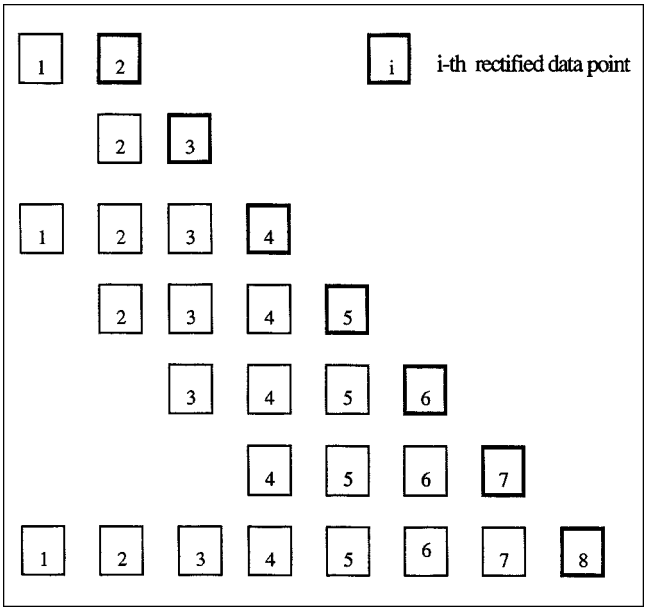

Figure 8. OLMS rectification.

On-line multiscale rectification of data with Gaussian errors

On-line multiscale rectification is based on multiscale rectification of data in a moving window of dyadic length, as shown in Figure 8. The OLMS methodology can be summarized as follows:

(1) Decompose the measured data within a window of dyadic length using a causal boundary corrected wavelet filter.

(2) Threshold the wavelet coefficients and reconstruct the rectified signal.

(3) Retain only the last data point of the reconstructed signal for on-line use.

(4) When new measured data are available, move the window in time to include the most recent measurement while maintaining the maximum dyadic window length.

The measurements in each window are rectified by the wavelet thresholding approach of Donoho et al. (1995) described in the previous section. This simple approach is very effective compared to the single-scale techniques, as shown by the theoretical analysis presented next and the illustrative examples. It retains the benefits of the wavelet decomposition in each moving window, while allowing each measurement to be rectified on-line.

Deeper insight into the properties and benefits of OLMS rectification may be obtained by relating it to the existing single-scale methods of mean filtering and exponential smoothing. The filters corresponding to the Haar scaling function at different scales are shown in Figure 9a. The shape of these filters is identical to that of a mean filter shown in Figure 1a. Due to its multiscale character, OLMS rectification using Haar wavelets is able to automatically select a mean filter of dyadic length that is best for representing each signal feature, as stated by the following theorem.

Theorem 1. For each measurement, OLMS Haar rectification at L scales can adaptively specialize to a mean filter of dyadic length ranging from 2^L to 1.

This theorem is proved in the Appendix based on the fact that a mean filtered signal using a filter of dyadic length 2^m may be generated by keeping the coefficients for the last scaled signal and the wavelet coefficients from the coarsest scale of decomposition L up to scale m. In practice, OLMS rectification using Haar wavelets subsumes a class larger than mean filters of dyadic lengths because if the last wavelet coefficient at some intermediate scale coarser than scale m is eliminated, then the effective filter will not be a mean filter of dyadic length. Instead, its coefficients will have different magnitudes to give more weighting to the more important features in the signal. For example, if a signal is rectified by decomposing it using Haar at two scales and by keeping the scaled signal at $m = 2$ and the last wavelet coefficient at scale $m = 1$, but eliminating the last wavelet coefficient at scale $m = 2$, the effective FIR filter will have the shape shown in Figure 10, which is not a mean filter. It gives more weighting to the most recent measurement to capture the important feature represented by the last retained wavelet coefficient at scale $m = 1$ and, at the same time, it gives a negative weighting to the previous measurement to account for the eliminated noise component represented by the eliminated wavelet coefficient at scale $m = 2$. The filter for OLMS rectification

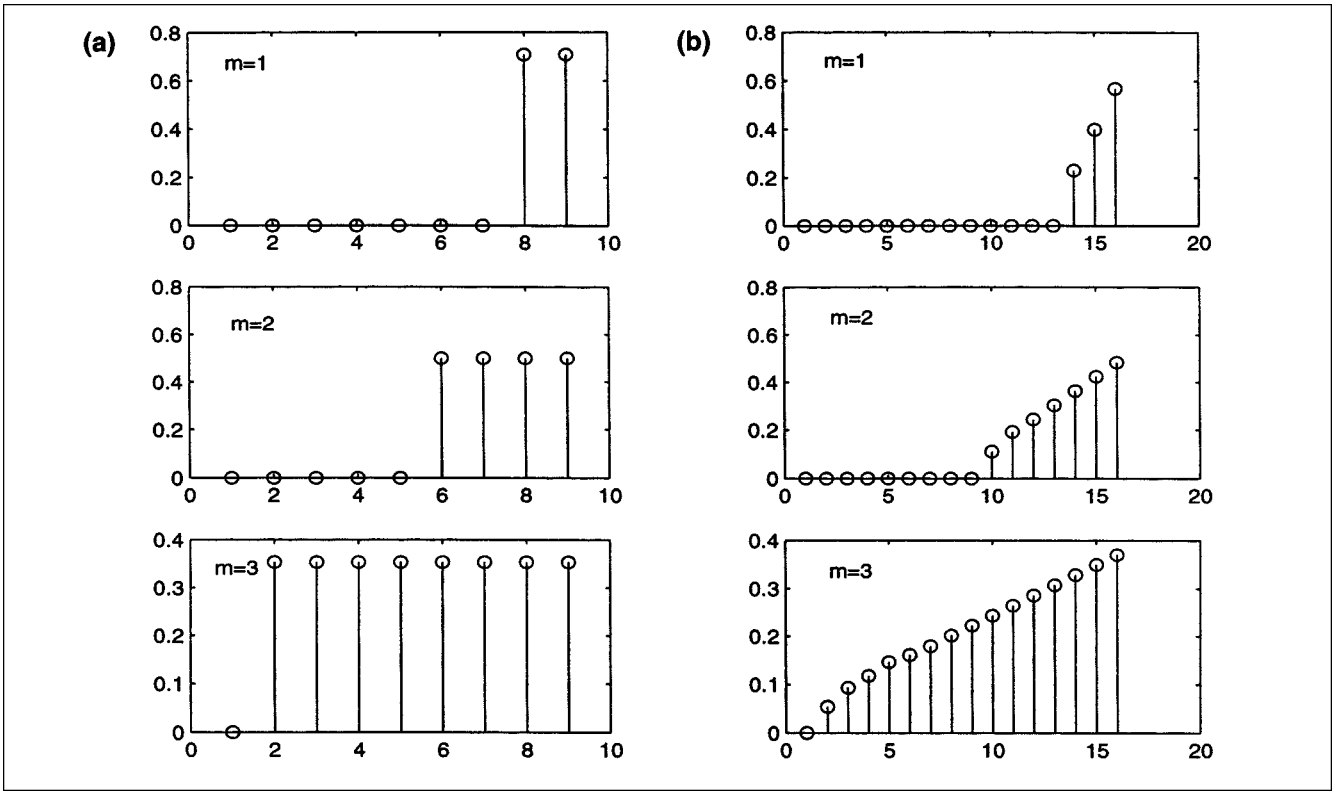

Figure 9. (a) Haar filter and (b) post-conditioned Daubechies D2 boundary corrected filter at multiple scales.

will also be able to adapt its length to best represent the important features in the measured signal. Thus, a short filter will be automatically selected for representing a fast change in the underlying signal, whereas a long filter will be selected for representing a slow change or constant segment. The nature of the filter selected for obtaining each rectified measurement is decided by the coefficients that are retained after thresholding.

This theoretical analysis shows that OLMS Haar rectification should be better than that obtained by a mean filter of dyadic length. In practice, due to its adaptive nature, OLMS Haar rectification performs better than mean filtering even for signals that are best for nondyadic mean filters, as shown by the illustrative examples. Haar is not the only wavelet filter that can be used for on-line rectification. It is used in this theoretical analysis because its simple form is convenient for

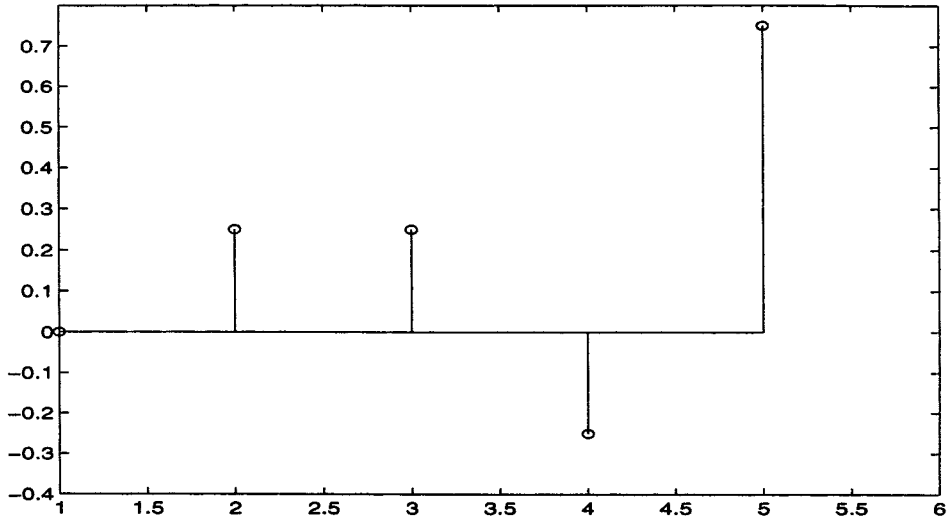

Figure 10. The effective FIR filter corresponding to OLMS Haar rectification at two scales ($L = 2$) by keeping all coefficients except the last wavelet coefficient at scale $m = 2$.

comparison with the single-scale method of mean filtering. OLMS rectification with wavelets other than Haar will require the use of boundary corrected filters. Figure 9b shows the filter applied to the right boundary by the boundary-corrected Daubechies second-order scaling functions at multiple scales. These filters are similar to that used for exponential smoothing shown in Figure 1b. Consequently, OLMS rectification using smoother wavelets will *approximately* subsume exponential smoothing to provide better rectification. The performance of OLMS rectification is compared with that of existing methods by illustrative examples in the next section.

Boundary corrected translation invariant (BCTI) rectification

If the rectified values are not required immediately, the quality of the rectified signal can be further improved by taking the mean of the OLMS rectified signals from different moving windows. For the rectification depicted in Figure 8, the BCTI rectified value for the first measurement will be the mean of three translations, whereas the rectified value for the fourth measurement will be the mean of five translations. Due to this averaging, the final rectified signal will have less spurious features than the result of on-line rectification. This approach is similar to the TI rectification (Coifman and Donoho, 1995), but since it does not assume the signal to be a cyclic list, it overcomes the problem of boundary effects encountered in TI rectification. This advantage is illustrated in Figure 5, which compares the translation mechanisms for TI and BCTI rectification. This comparison shows that in the elimination of errors at the boundaries, BCTI rectification will take the mean of fewer translations. Consequently, BCTI will require less computations than TI, but may be less smooth near the boundaries. The performance of this boundary corrected TI (BCTI) rectification technique is also illustrated in the next section.

Rectification of non-Gaussian errors

The capabilities of the OLMS and BCTI rectification techniques may be extended to simultaneous on-line elimination of Gaussian and non-Gaussian errors by combining these methods with multiscale median filtering. Similar to OLMS rectification, in the robust OLMS technique, data are rectified in a moving window of dyadic length that always includes the current measurement. The robust OLMS technique uses the robust wavelet transform algorithm illustrated in Figure 6, instead of the conventional wavelet transform to help suppress outliers at each scale. Application of the robust OLMS rectification to data containing sharp changes tends to smooth the change since the change is assumed to be an outlier until it persists beyond the coarsest scale of the median filter. Consequently, robust multiscale rectification is best suited for steady-state data as demonstrated by the illustrative examples.

When time delay is acceptable, improved elimination of Gaussian and non-Gaussian errors is possible by averaging the rectified signals from different time shifts. The luxury of having extra time to make a decision about the nature of a particular change allows better judgment resulting in a robust BCTI rectification method. The type of median filter, standard or FMH filter, used in the robust BCTI technique is not a major concern, because spikes and edge outliers are greatly reduced by averaging different translations, which is already achieved by BCTI. The advantages of combining BCTI and the robust wavelet transform are illustrated through synthetic and industrial examples.

Practical issues in OLMS and BCTI rectification

Any filtering method requires typical data or information about the underlying signal and noise for selecting the filter parameters. In OLMS rectification, the filter tuning parameters are the value of the threshold and the maximum depth of the wavelet decomposition. For the robust OLMS rectification, the length of the median filter at the finest scale also needs to be decided. Other practical issues include selecting a wavelet and the maximum length of the moving window. These issues are discussed in the rest of this section.

Value of Threshold. The threshold may be estimated by applying the Visushrink method to the available measurements. For data corrupted by stationary errors, the threshold value stops changing much after an adequate number of measurements are available. Consequently, for stationary noise, the threshold may be estimated from the measurements until the change is below a user-specified value. This approach for estimating the threshold cannot be performed recursively due to the median operator used in Eq. 11 and will require storage of a large number of measurements. Modifications for determining the threshold for time-dependent errors have been suggested (Neumann and Von Sachs, 1995).

Depth of Decomposition. Thresholding wavelet coefficients at very coarse scales may result in the elimination of important features, whereas thresholding only at very fine scales may not eliminate enough noise. Therefore, the depth of wavelet decomposition needs to be selected to optimize the quality of the rectified signal. Empirical evidence suggests that a good initial guess for the decomposition depth is about half of the maximum possible depth, that is $\{[\log_2(n)]/2\}$, where n is the moving window length. However, a smaller depth might be more appropriate in OLMS rectification if a long boundary corrected filter with a large support is used in the decomposition since the filters at the two edges might overlap at very coarse scales. In all synthetic examples used in this work, the decomposition depth is selected to minimize the mean-square error between the underlying and the rectified signals. For signals with non-Gaussian errors, the decomposition depth also needs to be chosen so that the effective median filter at the coarsest scale is longer than the longest patch of outliers.

Selected Wavelet Filter. The type, length, and nature of the wavelet filter used in OLMS and BCTI affect the quality of the rectification. Since the OLMS rectification uses only the last rectified data point from each translated signal, it is crucial that only boundary corrected or causal wavelet filters are used. If boundary corrected filters are not used, then the last point is among the least accurate ones due to end effect errors. This is not a major concern in BCTI rectification, because inaccuracies due to end effects are greatly reduced by averaging different translations.

In off-line rectification, the similarity between the shape of the wavelet filter and the shape of the signal enhances the

quality of the rectified signal. For example, the Haar filter is a better choice for a stair-step signal than a smoother filter, such as Daubechies; however, a smooth Daubechies filter is a better choice than Haar for a smooth signal. Since the rectified signal using BCTI is just the average of different off-line rectified signals from different windows, the quality of BCTI rectification is improved when the wavelet filter resembles the signal's features. For OLMS rectification, however, this advantage does not necessarily hold. Just as EWMA often gives a smaller mean-square error than mean filtering and preserves sudden changes more accurately, OLMS rectification using Daubechies second-order boundary corrected filter often results in smaller mean-square error than OLMS rectification using Haar as illustrated in the next section.

Illustrative Examples

In this section, the advantages and limitations of OLMS and BCTI rectification of Gaussian and non-Gaussian errors are illustrated through examples based on synthetic and industrial data. The synthetic signals used in this analysis are generated as

$$x_t = y_t + e_t + \delta_t \tag{14}$$

where x_t is the noisy signal, y_t is the noise-free deterministic signal, e_t is the stationary error, which can be white or correlated, and δ_t is the non-Gaussian error component containing outliers of different patch lengths. Two types of synthetic noise-free signals are used, namely, stair-step and HeaviSine, to better illustrate the performance of the developed techniques on data with different degrees of smoothness. The industrial data are measurements of a tray temperature of a distillation column. The following set of examples compares the performance of the OLMS and BCTI techniques with those of the single-scale and the nonlinear techniques, respectively. For the synthetic data, since the underlying signal is known and, for a fair comparison of the different methods, the tuning parameters for each method are selected to minimize the mean-square error between the rectified and the noise-free signals using trial and error. Furthermore, the VisuShrink method of Donoho et al. (1995) is used to calculate the threshold in all examples, soft thresholding is used in all OLMS examples, and hard thresholding is used in all BCTI examples.

Example 1. Stair-Step Signal with White Noise. The noisy step-stair signal contaminated with white noise of variance 0.5 and the noise-free signal are shown in Figure 11a. The filter parameters used to minimize the mean-square errors are a scale depth of 5 for OLMS using Haar, a mean filter length of 2, and an EWMA smoothing parameter value of 0.6. OLMS rectification gives the smallest mean-square error followed by exponential smoothing and mean filtering. The superior performance of OLMS is confirmed by averaging the results of rectifying hundred of realizations of the signal. The mean-square error results of this Monte Carlo simulation are listed in column 2 of Table 1. For this example, an even smaller mean-square error of 0.2592 is obtained using OLMS with Daubechies second-order boundary corrected filter through a similar Monte Carlo simulation. This advantage of OLMS using Daubechies over Haar is analogous to the advantage of EWMA over mean filtering in capturing sharp changes, as explained in the practical issues section. Since Haar and Daubechies OLMS are approximately multiscale generalizations of mean filtering and exponential smoothing, respectively, more oversmoothing of edges is observed in OLMS rectification using Haar than OLMS rectification using Daubechies, which results in a higher mean-square error.

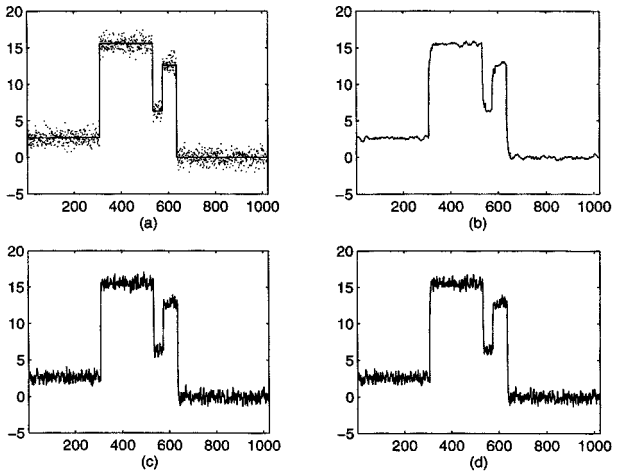

Figure 11. (a) Stair-step signal with white noise of variance 0.5; (b) OLMS Haar rectification (MSE = 0.2865); (c) mean filtering (MSE = 0.3805); (d) exponential smoothing (MSE = 0.3232).

Example 2. HeaviSine Signal with White Noise. The signal used in this example is a HeaviSine function with a white noise of variance 0.5. Daubechies second-order boundary corrected filter is used in OLMS rectification. The scale depth in OLMS, the mean filter length, and the exponential smoothing parameter that minimize the mean-square errors are 5, 4, and 0.35, respectively. Once again, OLMS results in the smallest mean-square error for the data in Figure 12 and for the Monte Carlo simulation summarized in column 3 of Table 1.

Example 3. Stair-Step Signal with Colored Noise. A similar comparison is performed for the case of correlated measurement noise. An autoregressive AR(1) noise model of the form, $e_t = \sqrt{0.5}\{e_{t-1} + a_t\}$ is used in this analysis. Figure 13 compares the performance of OLMS, mean filtering, and exponential smoothing. For OLMS, the Haar filter is used with scale depth of 4. The mean filter length and the exponential smoothing parameter used are 1 and 0.7, respectively. A mean

Table 1. Monte Carlo Simulation Comparing Various Filtering Techniques

	Stair-Step White Noise Using Haar	HeaviSine White Noise Using BC D2	Stair-Step Correlated Noise Using Haar	HeaviSine Correlated Noise Using BC D2
OLMS	0.2739	0.1177	0.3737	0.3147
Mean Filtering	0.3606	0.1796	0.5130	0.3917
EWMA	0.2970	0.1605	0.4533	0.3365

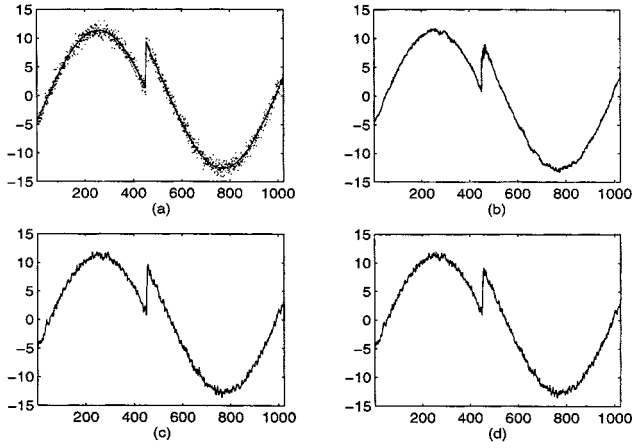

Figure 12. (a) HeaviSine signal with white noise of variance 0.5; (b) OLMS rectification using D2 boundary corrected filter (MSE = 0.1027); (c) mean filtering (MSE = 0.2085); (d) exponential smoothing (MSE = 0.1788).

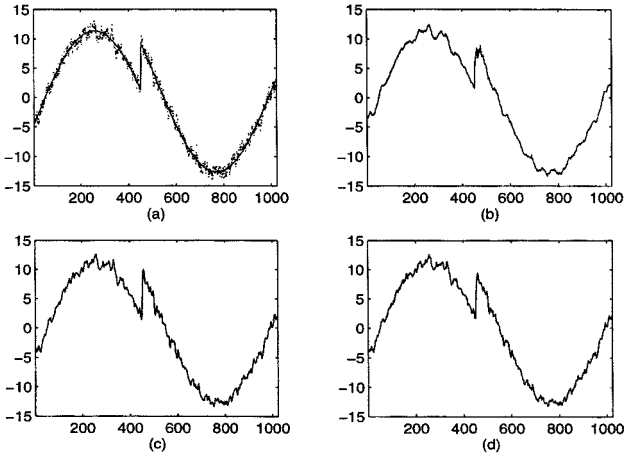

Figure 14. (a) HeaviSine signal with correlated noise of variance 0.5; (b) OLMS rectification using D2 boundary corrected filter (MSE = 0.3059); (c) mean filtering (MSE = 0.4375); (d) exponential smoothing (MSE = 0.3688).

filter of length one implies no filtering. It resulted in the least mean-square error because longer mean filters oversmooth the sharp changes in the stair-step signal increasing the mean-square error. As shown in Figure 13 and column 4 of Table 1, once again, OLMS outperforms linear filtering.

Example 4. HeaviSine Signal with Colored Noise. This example confirms the trend in the relative performance of OLMS, mean filtering, and exponential smoothing. The contaminating AR(1) noise model is the same as that used in Example 3. A Daubechies boundary corrected filter with a scale depth of 5, a mean filter of length 5, and an exponential smoothing parameter of 0.28 are used for the smallest mean-square errors. The results are shown in Figure 14, and the Monte Carlo simulation results are summarized in column 5 of Table 1. The results in these examples demonstrate the benefits of OLMS rectification over other linear filtering techniques. The rectified signals using OLMS rectification are visually more appealing and have smaller mean-square errors.

Example 5. Comparison of TI, BCTI, and FMH Rectification. When time delay is acceptable, the quality of rectification can be further improved by BCTI rectification. BCTI results in smoother rectified signals as in TI rectification, but provides better estimates of the signal boundaries than TI rectification. This advantage is illustrated through the following example, in which the BCTI rectification of a stair-step and HeaviSine signals are compared with those of TI and FMH rectification, as shown in Figures 15 and 16, respectively. For

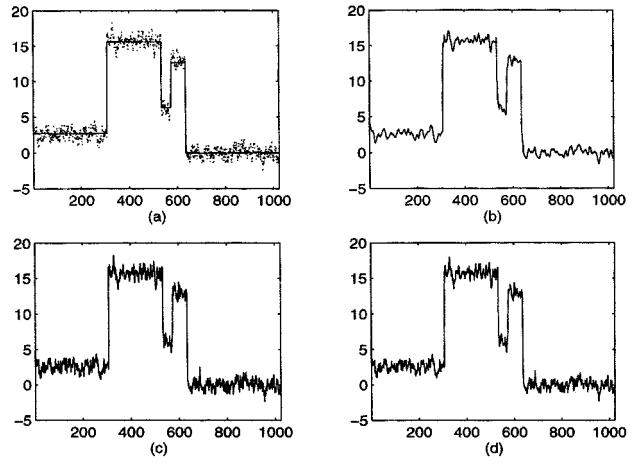

Figure 13. (a) Stair-step signal with correlated noise of variance 0.5; (b) OLMS rectification using Haar (MSE = 0.3747); (c) mean filtering (MSE = 0.5372); (d) exponential smoothing (MSE = 0.4963).

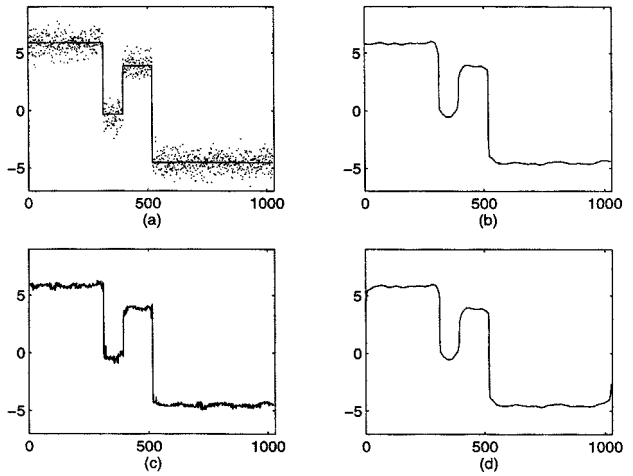

Figure 15. (a) Stair-step signal with white noise; (b) BCTI rectification using Haar (MSE = 0.0439); (c) FMH filtering (MSE = 0.0317); (d) TI rectification using Haar (MSE = 0.0644).

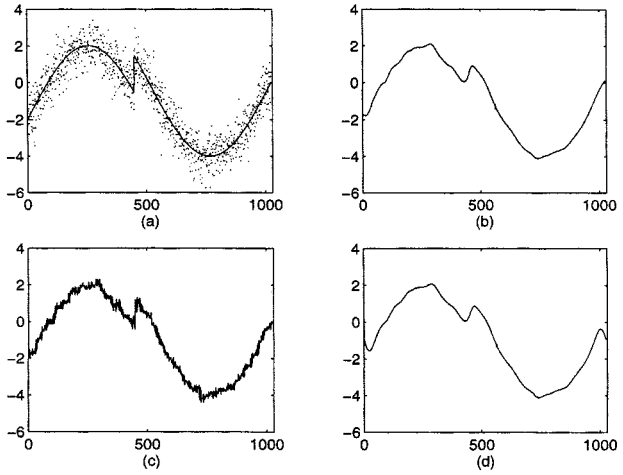

Figure 16. (a) HeaviSine signal with white noise; (b) BCTI rectification using D2 boundary corrected filter (MSE = 0.0227); (c) FMH filtering (MSE = 0.0383); (d) TI rectification using D2 boundary corrected filter (MSE = 0.0341).

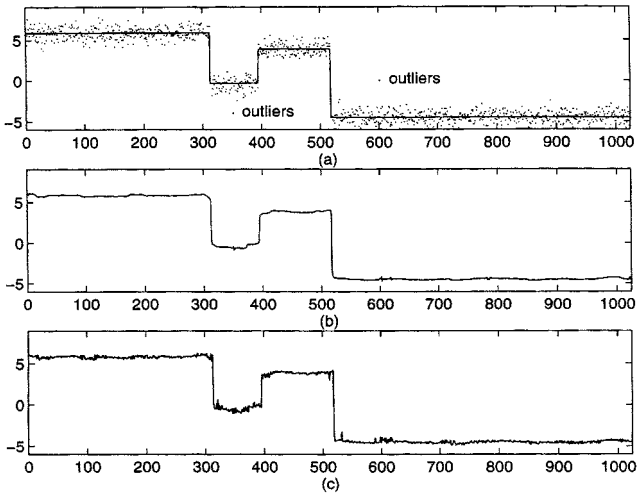

Figure 18. (a) Noisy and noise-free stair-step signal with outlier patches of length 2; (b) robust BCTI Haar and scale depth of 6 (MSE = 0.01830); (c) FMH filtering with a filter of length 13 (MSE = 0.01686).

the stair-step signal contaminated by white noise of variance 0.5, BCTI rectification with a scale depth of 6, a FMH filter length of 29, and TI Haar rectification with scale depth of 6 are used. In this example, FMH filtering gives a smaller mean-square error since a piece-wise constant signal is ideally suited for this method (Heinonen and Neuvo, 1988). The results in Figure 15 indicate that FMH filtering retains more noise than BCTI rectification but does not round off the edges as much. The advantage of BCTI rectification over TI rectification at the signal boundaries is also clearly demonstrated in Figures 15b and 15d.

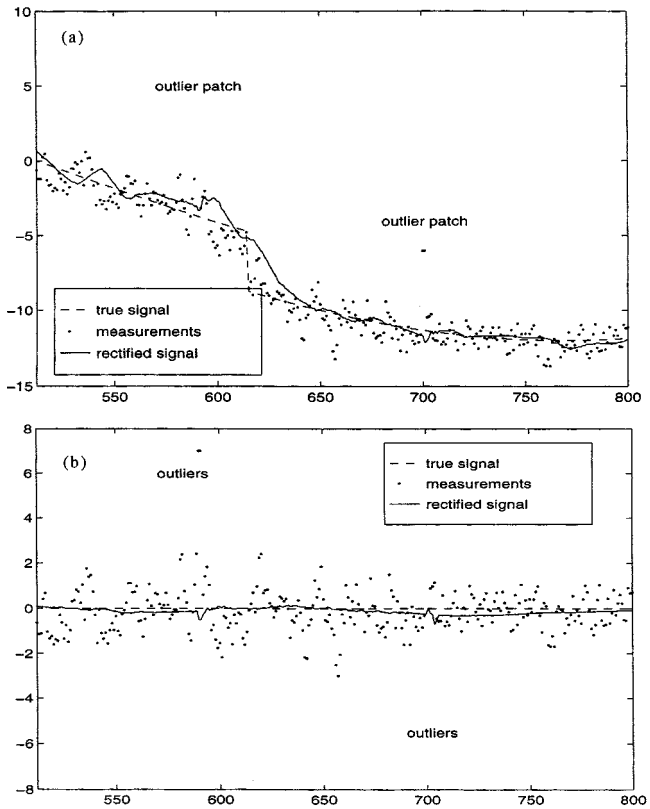

Figure 17. Robust Haar OLMS rectification.

(a) dynamic signal with AR(1) noise and outlier patches of length 2, scale depth = 5, FMH filter length = 7, (MSE = 0.347); (b) constant signal with AR(1) noise and outlier patches of length 2, FMH filter length = 9, scale depth = 6, (MSE = 0.031).

For the HeaviSine example shown in Figure 16, BCTI rectification with a D2 boundary corrected filter and a scale depth of 6, a FMH filter length of 25, and TI rectification with D2 boundary corrected filter and scale depth of 5 are used. In this example, BCTI gives a smaller error than FMH. BCTI does also better than TI due to the absence of errors at the signal ends.

Example 6. Robust OLMS Rectification. The performance of the robust OLMS rectification is illustrated for a dynamic and steady-state signal. The OLMS rectification in Figure 17a shows that robust OLMS rectification tends to oversmooth sharp changes in the data. As discussed in an earlier section, this smoothing is due to the fact that a significant change is initially considered to be an outlier until the change persists for a duration longer than the coarsest scale of the multiscale median filter. However, steady-state data with non-Gaussian errors can be rectified without distortion by the robust OLMS technique as shown in Figure 17b.

Example 7. Robust BCTI Rectification. When off-line rectification is acceptable, the classification of sharp changes becomes easier since measurements before and after the change are usually available. The performance of BCTI rectification is shown and compared with that of FMH filtering in Figure 18. Robust BCTI rectification with Haar and a scale depth of 6, a median filter of length 13, and a FMH filter of length 31 are used since they minimize the mean-square error. Both techniques identify and eliminate all outliers. However, FMH rectification retains more noise than the robust BCTI but

captures the edges more accurately leading to a smaller mean-square error.

Example 8. Industrial Data. The performance of OLMS, BCTI and their robust counterparts are illustrated through application to set temperature measurements from an industrial distillation column tray. In OLMS rectification, shown in Figure 19a, the Haar wavelet and an initial window length of 128 are used. Note that this initial window length is not a restriction. Theoretically, OLMS rectification can start with any dyadic set of measurements, starting at two. However,

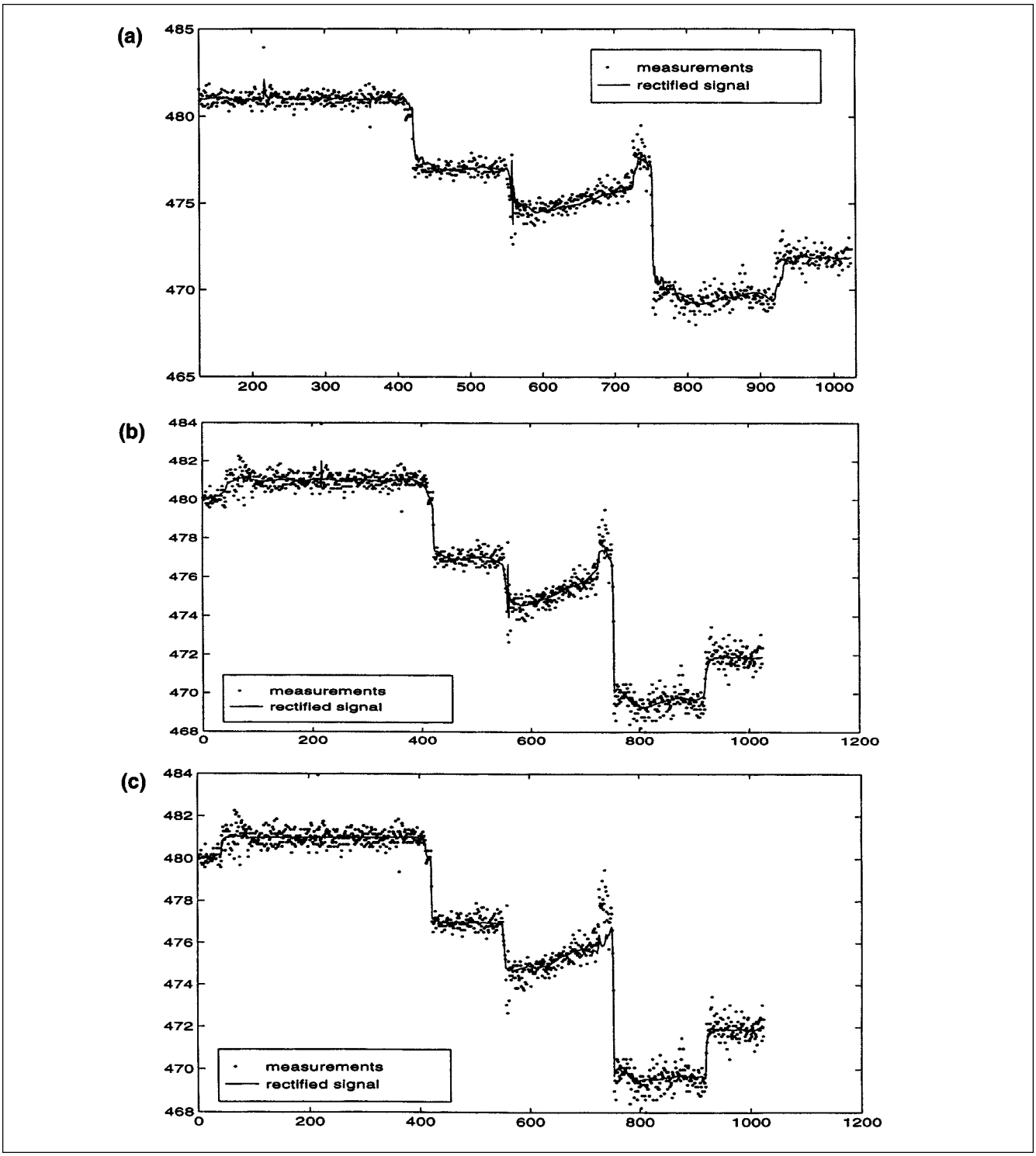

Figure 19. Rectification of industrial data.

This uses (a) OLMS and Haar using an initial window of 128 points; (b) BCTI and Haar using initial window of 128 points; (c) robust BCTI and Haar, FMH filter length = 7.

since the threshold is estimated from the data in the moving window, the threshold estimate improves as the moving window length increases. When the noise is assumed to be stationary, the threshold stops changing after a large set of measurements are collected, and, thus, the moving window length can be held constant. Also, the decomposition depth used in this example is two scales deeper than half of the maximum possible depth. These parameters are chosen such that a good visual quality of rectification is achieved. Figure 19a shows that the OLMS rectified signal follows the trend of the measurements while capturing the changes quite well.

In BCTI rectification, shown in Figure 19b, the Haar wavelet, a decomposition depth of two scales deeper than half of the maximum possible depth, and an initial window of length 512 are used. This initial window length is chosen to achieve simultaneous reduced boundary effects and smoothness by averaging rectified signals at different time shifts. If all available measurements, in this case 1,024 data points, are used as the initial window, smoothness will deteriorate as no averaging is done unless the data are assumed to be a cyclic list where boundary effects will be observed as in TI rectification. The resulting BCTI rectified signal, shown in Figure 19b, is smoother than the one obtained using OLMS rectification. However, both OLMS and BCTI rectification seem to be affected by the apparent outliers present in the data. Rectification by robust BCTI decreases the sensitivity to the suspected outliers, as shown in Figure 19c, in which a FMH filter of length 7 is used. The result also shows that some suspected features between the 700th and 750th data points are eliminated. In this example, the true nature of these features is unknown. If they are not outliers, then a smaller decomposition depth and a shorter median filter are recommended for this example.

Conclusions

This article extends the benefits of multiscale analysis using wavelets to on-line rectification of measured data without fundamental or empirical process models. The method for on-line multiscale (OLMS) rectification is based on rectifying data in a moving window of dyadic length by wavelet thresholding (Donoho et al., 1995). The measurements in each window are decomposed using the selected wavelet, the wavelet coefficients smaller than a threshold are eliminated, and the signal is reconstructed. For on-line rectification, only the most current rectified measurement is retained.

If the rectified signal is not required on-line, the quality of rectification may be improved by taking the average of the rectified signals in each moving window. This approach is equivalent to translation invariant rectification (Coifman and Donoho, 1995), but without the boundary errors, and is called boundary corrected translation invariant (BCTI) rectification. Both OLMS and BCTI rectification are extended to rectify signals corrupted by random and gross errors by combining wavelet thresholding in each moving window with multiscale median filtering (Bruce et al., 1994).

Theoretical analysis shows that OLMS rectification using Haar wavelets subsumes mean filtering by automatically selecting the best filter at each measurement from a library of mean filters of dyadic length. OLMS rectification using smoother wavelets is analogous to a multiscale generalization of exponential smoothing. Similarly, the robust OLMS rectification automatically selects the best mean and median filters from a dyadic library. The theoretical properties and the illustrative examples demonstrate the superiority of OLMS rectification over linear filtering, and the advantage of BCTI over TI at the boundaries and over FMH in terms of smoothness.

This article has focused on multiscale rectification of a single variable contaminated by Gaussian stationary noise and outliers when accurate process models are not available. Multiscale techniques have been developed for the off-line removal of other types of errors such as nonstationary and heteroscedastic noise (Neumann and Von Sachs, 1995). These methods may be extended for on-line rectification by incorporating them in the methodology developed in this article. Linear filters are popular for a variety of applications such as statistical process control (Montgomery, 1996) and adaptive modeling (Wold, 1994). The multiscale approach developed in this article may also be used to improve upon these methods (Top and Bakshi, 1998). Most practical data rectification problems are multivariate in nature, and may have fundamental process models available. A multiscale approach for the rectification of such data is described by Bakshi et al. (1998).

Acknowledgments

Partial financial support from the TAPPI foundation (PE-357-95), American Chemical Society Petroleum Research Fund (30523-G9), and National Science Foundation (CTS 9733627) are gratefully acknowledged.

Literature Cited

Albuquerque, J., and L. Biegler, "Data Reconciliation and Gross-Error Detection for Dynamic Systems," *AIChE J.*, **42**(10), 2841 (1995).

Bakshi, B. R., M. N. Nounou, P. K. Goel, and X. Shen, "Multiscale Data Reconciliation with Fundamental Process Models," AIChE Meeting, Miami Beach (1998).

Bose, N. K., *Digital Filters Theory and Applications*, Elsevier Science, New York (1985).

Bruce, A., and H.-Y. Gao, *Applied Wavelet Analysis with S-Plus*, Springer, New York (1996).

Bruce, A. G., D. L. Donoho, H.-Y. Gao, and R. D. Martin, "Denoising and Robust Non-Linear Wavelet Analysis," *Proc SPIE*, Orlando, FL, **2242**, 325 (1994).

Chiari, M., G. Bussani, M. G. Grottoli, and S. Pierucci, "On-Line Data Reconciliation and Optimization: Refinery Applications," *Comput. Chem. Eng.*, **21**, suppl., s1185 (1997).

Cohen, A., I. Daubechies, and V. Pierre, "Wavelets on the Interval and Fast Wavelet Transforms," *Appl. Comput. Harmonic Anal.*, **1**, 54 (1993).

Coifman, R. R., and D. L. Donoho, "Translation-Invariant De-Noising," *Lecture Notes in Statistics*, Springer-Verlag, **103**, 125 (1995).

Crowe, C. M., "Observability and Redundancy of Process Data for Steady State Reconciliation," *Chem. Eng. Sci.*, **44**, 2909 (1989).

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptotia?," *J. Roy. Stat. Soc. B*, **57**, 301 (1995).

Donoho, D. L., and I. M. Johnstone, "Ideal De-noising in an Orthonormal Basis Chosen from a Library of Bases," Technical Report, Dept. of Statistics, Stanford Univ., Stanford, CA (1994).

Gallagher, N. C., Jr., and G. Wise, "Theoretical Analysis of the Properties of the Median Filters," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-**29**, 1136 (1981).

Gelb, A., *Applied Optimal Estimation*, MIT Press, Cambridge, MA (1974).
Heinonen, P., and Y. Neuvo, "FIR-Median Hybrid Filters with Predictive FIR Substructures," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **36**, 892 (1988).
Heinonen, P., and Y. Neuvo, "FIR-Median Hybrid Filters," *IEEE Trans. on Acoustics, Speech, and Signal Process*., ASSP-**35**, 832 (1987).
Islam, K. A., G. H. Weiss, and J. A. Romagnoli, "Non-Linear Data Reconciliation for an Industrial Pyrolysis Reactor," *Comput. Chem. Eng*., **18**, suppl., s218 (1994).
Johnston, L. P. M., and M. A. Kramer, "Maximum Likelihood Data Rectification: Steady-State Systems," *AIChE J.*, **41**, 2415 (1995).
Kao, C.-S., A. C. Tamhane, and R. H. Mah, "Gross Error Detection in Serially Correlated Data. 2. Dynamic Systems," *Ind. Eng. Chem. Res*., **31**, 254 (1992).
Kao, C.-S., A. C. Tamhane, and R. H. Mah, "Gross Error Detection in Serially Correlated Data," *Ind. Eng. Chem. Res*., **29**, 1004 (1990).
Karjala, T. W., and D. M. Himmelblau, "Dynamic Data Rectification by Recurrent Neural Networks vs. Traditional Methods," *AIChE J*., **40**, 1865 (1994).
Kim, I.-W., M. J. Liebman, and T. F. Edgar, "A Sequential Error-In-Variables Methods for Nonlinear Dynamic Systems," *Comput. Chem. Eng*., **15**, 663 (1990).
Kim, I.-W., M. S. Kang, S. Park, and T. F. Edgar, "Robust Data Reconciliation and Gross Error Detection: The Modified MIMT using NLP," *Comput. Chem. Eng*., **21**, 775 (1997).
Kramer, M. A., and R. S. H. Mah, "Model Based Monitoring," *Proc. Int. Conf. on Foundations of Comput. Aided Process Oper*., D. Rippen, J. Hale, and J. Davis, eds., CACHE, Austin, TX (1994).
Liebman, M. J., and T. F. Edgar, "Efficient Data Reconciliation and Estimation for Dynamic Processes using Nonlinear Programming Techniques," *Comput. Chem. Eng*., **16**, 963 (1992).
Mah, R. S. H., *Chemical Process Structure and Information Flows*, Butterworth, Boston (1990).
Mallat, S. G., "A Theory of Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. on Pattern Analysis and Machine Intellig*., **11**, 764 (1989).
Montgomery, D. C., *Introduction to Statistical Quality Control*, Wiley, New York (1996).
Nason, G. P., "Wavelet Shrinkage using Cross-Validation," *J. of Roy. Stat. Soc*., **58**, 463 (1996).
Neumann, M., and R. Von Sachs, "Wavelet Thresholding: Beyond the Gaussian I.I.D. Situation," *Lecture Notes in Statistics*, Springer-Verlag, **103**, 103 (1995).
Nounou, M. N., "On-Line Multiscale Rectification of Random and Gross Errors without Process Models," MS Thesis, The Ohio State Univ., Columbus (1997).
Sorenson, H. W., ed., *Kalman Filtering: Theory and Applications*, IEEE Press, New York (1985).
Strum, R. D., and D. E. Kirk, *First Principles of Discrete Systems and Digital Signal Processing*, Addison-Wesley, Reading, MA (1989).
Tamhane, A. C., and R. S. H Mah, "Data Reconciliation and Gross Error Detection in Chemical Process Networks," *Technometrics*, **27**, 409 (1985).
Tham, M. T., and A. Parr, "Succeed at On-Line Validation and Reconstruction of Data," *Chem. Eng. Prog.*, **90**, 46 (1994).
Tjao, I. B., and L. T. Biegler, "Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems," *Comput. Chem. Eng*., **15**, 679 (1991).
Tong, H., and C. M. Crowe, "Detection of Gross Errors in Data Reconciliation by Principal Component Analysis," *AIChE J.*, **41**, 1712 (1995).
Top, S., and B. R. Bakshi, "Improved Statistical Process Control using Wavelets," *Foundations of Computer Aided Process Operations*, J. Pekny and G. Blau, eds., Snowbird, UT, in press (1999).
Veverka, V., "A Method of Reconciliation of Measured Data with Nonlinear Constraints," *Appl. Math. Comput*., **49**, 141 (1992).
Wold, S., "Exponentially Weighted Moving Principal Components Analysis and Projections to Latent Structures," *Chemom. and Intell. Lab. Sys*., **23**, 149 (1994).

Appendix: Proof of Theorem 1

The decomposition of the signal $x(t)$ at L scales can be written as

$$x(t) = \sum_{k=1}^{n2^{-L}} a_{Lk}\phi_{Lk}(t) + \sum_{m=1}^{L}\sum_{k=1}^{n2^{-m}} d_{mk}\psi_{mk}(t). \quad \text{(A1)}$$

The decomposition vector, containing the coarsest resolution coefficients and all detail coefficients, can be determined by multiplying the signal $x(t)$ by the decomposition matrix W

$$\begin{bmatrix} a_{L,1} \\ a_{L,2} \\ \cdot \\ a_{L,n2^{-L}} \\ d_{L,1} \\ \cdot \\ d_{L,n2^{-L}} \\ d_{L-1,1} \\ \cdot \\ d_{1,n2^{-1}} \end{bmatrix} = W * x(t) = \begin{bmatrix} h_{1,L} & h_{2,L} & \cdot & h_{2^L,L} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & h_{1,L} & \cdot & h_{2^L,L} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & h_{1,L} & \cdot & h_{2^L,L} \\ g_{1,L} & g_{2,L} & \cdot & g_{2^L,L} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & g_{1,L} & \cdot & g_{2^L,L} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & g_{1,L} & \cdot & g_{2^L,L} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ g_{11} & g_{12} & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & g_{11} & g_{12} & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & g_{11} & g_{12} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \quad \text{(A2)}$$

when all wavelet coefficients are eliminated, the decomposition vector becomes

$$[a_{L,1} \quad \cdot \quad \cdot \quad a_{L,n2^{-L}} \quad 0 \quad \cdot \quad \cdot \quad 0]^T (n \times 1). \quad \text{(A3)}$$

which is equivalent to multiplying $x(t)$ by the matrix

$$W' = \begin{bmatrix} h_{1,L} & h_{2,L} & \cdot & h_{2^L,L} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & h_{1,L} & \cdot & h_{2^L,L} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \cdot & 0 & h_{1,L} & \cdot & h_{2^L,L} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}. \quad \text{(A4)}$$

For the Haar wavelet, the matrix W' is

$$W' = \begin{bmatrix} \frac{1}{\sqrt{2^L}} & \frac{1}{\sqrt{2^L}} & \cdot & \frac{1}{\sqrt{2^L}} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & 0 & \frac{1}{\sqrt{2^L}} & \cdot & \frac{1}{\sqrt{2^L}} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \frac{1}{\sqrt{2^L}} & \cdot & \frac{1}{\sqrt{2^L}} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}. \quad \text{(A5)}$$

Therefore, the final rectified signal is

$$\hat{\underline{x}}(t) = W^T W' x(t) = \begin{bmatrix} \frac{1}{2^L} & \cdot & \frac{1}{2^L} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \frac{1}{2^L} & \cdot & \frac{1}{2^L} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & \frac{1}{2^L} & \cdot & \frac{1}{2^L} & 0 & \cdot & 0 \\ 0 & \cdot & 0 & \cdot & \cdot & \cdot & 0 & \cdot & 0 \\ 0 & \cdot & 0 & \frac{1}{2^L} & \cdot & \frac{1}{2^L} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \frac{1}{2^L} & \cdot & \frac{1}{2^L} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \frac{1}{2^L} & \cdot & \frac{1}{2^L} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} \frac{1}{2^L}(x_1 + \ldots + x_{2^L}) \\ \cdot \\ \frac{1}{2^L}(x_1 + \ldots + x_{2^L}) \\ \cdot \\ \cdot \\ \frac{1}{2^L}(x_{n-2^L+1} + \ldots + x_n) \\ \cdot \\ \frac{1}{2^L}(x_{n-2^L+1} + \ldots + x_n) \end{bmatrix} \quad \text{(A6)}$$

Equation A6 can be written in terms of the Haar scaling function as follows

$$\hat{x}(t) = \left\{ \sum_{k=1}^{n2^{-L}} a_{Lk} \phi_{Lk} \right\}. \quad \text{(A7)}$$

However, in OLMS rectification, only the last point of each rectified translated signal matters. The last point in the rectified signal in Eq. A6 is

$$\hat{x}(n) = \frac{1}{2^L}(x_{n-2^L+1} + \ldots + x_n) = \overline{MF\{2^L\}} \quad \text{(A8)}$$

where $MF\{2^L\}$ means a mean filter of length 2^L.

Since the Haar filter of length 2, only the last wavelet coefficient at each scale can affect the last rectified data point, which can affect OLMS rectification. This can be shown by eliminating only the last coefficient from every scale, keeping all other interior wavelet coefficients, which is equivalent to using the following matrix for wavelet decomposition

$$\begin{bmatrix} \frac{1}{\sqrt{2^L}} & \cdot & \cdot & \frac{1}{\sqrt{2^L}} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \frac{1}{\sqrt{2^L}} & \cdot & \cdot & \frac{1}{\sqrt{2^L}} \\ \frac{1}{\sqrt{2^L}} & \cdot & \cdot & -\frac{1}{\sqrt{2^L}} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{2^{L-1}}} & \cdot & -\frac{1}{\sqrt{2^{L-1}}} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad \text{(A9)}$$

Therefore, the rectified translated signal is

$$\begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & \frac{1}{2^L} & \cdot & \cdot & \frac{1}{2^L} \\ 0 & \cdot & 0 & \frac{1}{2^L} & \cdot & \cdot & \frac{1}{2^L} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix}$$

$$= \left\{ \sum_{k=1}^{n2^{-L}} a_{Lk}\phi_{Lk} \right\}. \quad \text{(A10)}$$

$$\begin{bmatrix} \frac{1}{\sqrt{2^L}} & \cdot & \frac{1}{\sqrt{2^L}} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & \frac{1}{\sqrt{2^L}} & \cdot & \cdot & \frac{1}{\sqrt{2^L}} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & \frac{1}{\sqrt{2^L}} & \cdot & \cdot & -\frac{1}{\sqrt{2^L}} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}. \quad \text{(A12)}$$

Therefore, the rectified translated signal is

$$\begin{bmatrix} \frac{1}{2^L} & \cdot & \cdot & \frac{1}{2^L} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \frac{1}{2^L} & \cdot & \cdot & \frac{1}{2^L} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & \frac{1}{2^{L-1}} & \cdot & \frac{1}{2^{L-1}} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & \frac{1}{2^{L-1}} & \cdot & \frac{1}{2^{L-1}} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \frac{1}{2^{L-1}} & \cdot & \frac{1}{2^{L-1}} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \frac{1}{2^{L-1}} & \cdot & \frac{1}{2^{L-1}} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix}. \quad \text{(A13)}$$

Again, the last point in the rectified translated signal is

$$\hat{x}(n) = \frac{1}{2^L}(x_{n-2^L+1} + \ldots + x_n) = MF\{2^L\} \quad \text{(A11)}$$

which is a mean filter of length 2^L. This means that all wavelet coefficients other than the last one in each scale are irrelevant to the last point in any translated signal and, thus, the OLMS rectification using Haar. This is a great simplification in this theoretical analysis because now we are only concerned about how retaining the last wavelet coefficient at a particular scale relates to mean filtering.

The decomposition matrix corresponding to the decomposition at L scales and eliminating all wavelet coefficients except the last one at scale L is

Equation A13 can be written in terms of the last scale scaling function coefficients and last wavelet coefficient as

$$\hat{x}(t) = \left\{ \sum_{k=1}^{n2^{-L}} a_{Lk}\phi_{Lk} + d_{L,n2^{-L}}\psi_{L,n2^{-L}} \right\}. \quad \text{(A14)}$$

Therefore, the last rectified data point in the translated signal is

$$\hat{x}(n) = \frac{1}{2^{L-1}}(x_{n-2^{L-1}+1} + \ldots + x_n) = MF\{2^{L-1}\} \quad \text{(A15)}$$

which is equivalent to a mean filter of length 2^{L-1}.

Furthermore, it can easily show the keeping the last wavelet coefficients at the last two scales L and $L-1$ is equivalent to applying a mean filter of length 2^{L-2}, which can be written as

$$\hat{x}(t) = \left\{ \sum_{k=1}^{n2^{-L}} a_{Lk}\phi_{Lk} + d_{L,n2^{-L}}\psi_{L,n2^{-L}} + d_{L-1,n2^{-(L-1)}}\psi_{L-1,n2^{-(L-1)}} \right\}. \quad \text{(A16)}$$

Again, keeping the last point

$$\hat{x}(n) = \frac{1}{2^{L-2}}(x_{n-2^{L-2}+1} + \ldots + x_n) = MF\{2^{L-2}\} \quad \text{(A17)}$$

is equivalent to a mean filter of length 2^{L-2}.

Extending the relations shown in Eqs. A11, A15, and A17, by keeping the wavelet coefficients at scales L through 1, the rectified signal can be represented as

$$\hat{x}(t) = \left\{ \sum_{k=1}^{n2^{-L}} a_{Lk}\phi_{Lk} + d_{L,n2^{-L}}\psi_{L,n2^{-L}} + d_{L-1,n2^{-(L-1)}}\psi_{L-1,n2^{-(L-1)}} + \ldots + d_{1,n2^{-1}}\psi_{1,n2^{-1}} \right\} \quad \text{(A18)}$$

and the last rectified data point

$$\hat{x}(n) = x(n) = MF\{2^0 = 1\}$$

which is equivalent to a mean filter of length 1.

Therefore, OLMS rectification using Haar is equivalent to an adaptive mean filter of dyadic lengths, ranging from 2^L to 1. It is called adaptive, because the equivalent mean filter length is dependent on the wavelet coefficients kept, which depend on the signal's features. Furthermore, if some Haar wavelet coefficients at intermediate scales are included, other types of FIR filters are created to adapt the localized changes in the underlying signal.

Manuscript received Sept. 8,1998, and revision received Feb. 22, 1999.