Compression of Chemical Process Data by Functional Approximation and Feature Extraction

Bhavik R. Bakshi
Dept. of Chemical Engineering, Ohio State University, Columbus, OH 43210

George Stephanopoulos
Laboratory for Intelligent Systems in Process Engineering, Dept. of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Effective utilization of measured process data requires efficient techniques for their compact storage and retrieval, as well as for extracting information on the process operation. Techniques for the on-line compression of process data were developed based on their contribution in time and in frequency using the theory of wavelets. Existing techniques for compression via wavelets and wavelet packets are inconvenient for on-line compression and are best suited for stationary signals. These methods were extended to the on-line decomposition and compression of nonstationary signals via time-varying wavelet packets. Various criteria for the selection of the best time-varying wavelet packet coefficients are derived. Explicit relationships among the compression ratio, local and global errors of approximation, and features in the signal were derived and used for efficient compression. Extensive case studies on industrial data demonstrate the superior performance of wavelet-based techniques as compared to existing piecewise linear techniques.

Introduction

Efficient management techniques of chemical process data are important, since measured data are crucial for performing several tasks in the chemical process industries. Process operation and control tasks are dependent on measured data. These tasks include feedback control, system identification, empirical process modeling and characterization, process monitoring, fault detection and diagnosis, supervisory and quality control, continuous process improvement, planning, and managerial decision-making. Measured data are a rich source of information, and if tapped can simplify and improve the operation of a chemical process. Improvements in sensors, measurement techniques and computer technology, combined with the increasing need for safe and efficient operation, have resulted in an explosion of measured process data. Furthermore, the trend towards computer integrated operation of chemical processes is expected to result in a data explosion, and extensive storage of process histories will be required for efficient and reliable decision-making (Venkatasubramanian and Stanley, 1993). With increasing development and use of the information superhighway and the need for remote and autonomous operation, process data are often transferred across networks to other operators, engineers, researchers and plant personnel. Recent environmental regulations and new quality management systems require collection and storage of real-time process data. In a typical large chemical process, it is estimated that 40–80 GB are required for storing a year's worth of measured data (Kennedy, 1993). Even with inexpensive storage devices, efficient data storage and retrieval are of critical importance because the complexity and speed of all the data-dependent tasks depends on the quantity and quality of stored data. Therefore, efficient techniques for the compaction and reconstruction of measured data are a crucial part of efficient data management, and are necessary for performing various data-dependent process tasks, with minimum utilization of computer and human resources.

Compact storage may be achieved by identifying and eliminating contributions to the signal that are irrelevant or redundant. For most process signals, the loss of the high frequency, low amplitude processes which are likely to be noise

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

is quite acceptable, particularly if the loss of accuracy can be controlled and monitored explicitly by the user. It is important that data compression techniques applied to chemical process data retain the transients and dynamics present in the measured variable, and provide acceptable accuracy at all the relevant frequencies. In case of multiple measured variables, it may also be important to retain the correlation structure between the signals.

Within chemical engineering, data compression has received surprisingly little attention. The advantages of real-time data compression were illustrated by Hale and Sellars (1981), and then by Bader and Tucker (1987a,b). They demonstrated the benefits of replacing the dedicated chart recorders with a more efficient and accurate data recording system based on the box car and backward slope methods. The swinging door method was proposed by Bristol (1990). Recently, Mah et al. (1995) have proposed a new method called piecewise linear on-line trending (PLOT). These data compression techniques produce piecewise linear interpolations of the data such that the local error between the actual and compressed signals is within given bounds. They are fast, but are unable to capture small process transients and upsets, and are best suited for steady-state analysis of the process. Other techniques like Vector Quantization (VQ) (Gray, 1984) have been applied to the compression of chemical process data (Feehs and Arce, 1988). Multivariate statistical methods for compression of signals from multiple variables by extracting the relationship between variables have also been suggested (MacGregor, 1994).

Our approach to process data compression (discussed in subsequent sections) is based on the observation that measured data consist of distinct features which may be extracted from a signal, by exploiting their distribution in the time-frequency space. Efficient techniques for time-frequency decomposition of a signal have been developed based on *wavelet* theory. If a measured signal does not have contributions from every frequency in every temporal range, then some of its wavelet coefficients will be zero, or negligibly small. These small coefficients may be neglected, and the signal may be reconstructed without significant loss of information. If the basis functions are local and orthonormal, then the square of the neglected coefficients is related to the global as well as the local errors of approximation of the reconstructed signal. Consequently, physically intuitive data compression techniques based explicitly on the global or local error incurred due to compression may be devised. Decomposition on *wavelet packets* introduces an additional degree of freedom in the wavelet decomposition, by exploring basis functions that cover the time-frequency range at different positions in time and frequency and at different scales. Wavelet packets provide a library of orthonormal basis functions, and may facilitate greater compression than the wavelet decomposition. For a nonstationary signal with changing time-frequency characteristics, the best wavelet packet representation may be time-varying. *Time-varying wavelet packets* involve adaptation of the wavelet packet basis with changing signal characteristics. The wavelet packet bases are evaluated over all time intervals, thus allowing selection of the best wavelet packet basis at all times. Decomposition of process signals on derivative wavelets may be used for data compression by extraction of relevant features from the signal.

In this article, we extend and adapt existing wavelet-based data compression methods for on-line application to chemical process data with explicit and physically intuitive criteria for specifying the quality of the reconstructed signal. A framework based on approximation theory is used to unify and compare the characteristics of various data compression techniques. A brief overview of signal decomposition in the time-frequency domain using wavelets and wavelet packets is presented in order to provide a physical interpretation of wavelet-based compression techniques. Techniques for the compression of chemical process data based on explicit control of the L^2 or L^∞ error of approximation or through feature extraction is described. Algorithms for the on-line decomposition of the measured signal, and criteria for the selection of the best basis and wavelet packet coefficients are also described. The practical issues in the implementation of wavelet-based techniques for the compression of chemical process data are discussed, as well as the issues underlying real-time compression, removing inaccuracies due to end effects, and selection of the mother wavelet. Finally, case studies on industrial data for an empirical comparison of the performance of traditional and wavelet-based techniques for data compression are presented. These case studies show that wavelet-based methods are efficient, intuitively appealing, and provide significantly greater compression than currently used techniques.

Techniques for Compression of Chemical Process Data

The primary objective of any data compression technique is to transform the data to a form that requires the smallest possible amount of storage space, while retaining all the relevant information. The desired qualities of a technique for efficient storage and retrieval of chemical process data are:

(1) Compacted data should require minimum storage space.

(2) Compaction and retrieval should be fast, often in real time.

(3) A clear and explicit measure of the quality of the signal obtained after retrieval should be available for use as a criterion for guiding the compression.

(4) The retrieved signal should have minimum distortion and should contain all the desired features.

(5) The compaction should be based on physically intuitive criteria and should require minimum *a priori* assumptions.

As will be shown in this article, techniques based on a time-frequency representation of measured data using wavelets satisfy the above requirements. Several techniques have been developed for data compaction, especially for the compaction of images, speech and seismic data. Techniques developed for these compaction tasks may not be directly applicable to the compaction of chemical process data for several reasons:

(1) Efficient execution of measured data dependent process tasks requires a consistent and unifying representation of measured data that facilitates the integration of various process tasks. The representation of the stored data should be consistent with that of other process tasks, and should facilitate integrated process operation.

(2) Compaction and retrieval of chemical process data are often necessary in real time. Data from dynamic processes are continuously collected, and the compaction and storage procedure should be able to keep up with the rate of data collection.

(3) Measures for the quality of the compressed process data may differ from those for data from other applications. For chemical process data, criteria based on the error of approximation and the retention of important features may be used.

Data compression techniques attempt to optimize the tradeoff between the compression ratio, and signal fidelity, or approximation error, and may be represented as solution techniques to an approximation problem defined as follows.

Definition: data compression problem

Determine the approximate representation $\hat{F}$ of a signal F so as to minimize the error of approximation given by:

$$\|F - \hat{F}\|_{L^p(I)} = \left[\int_I |F(x) - \hat{F}(x)|^p dx \right]^{1/p}, \quad 0 < p < \infty \tag{1a}$$

or

$$\|F - \hat{F}\|_{L^\infty(I)} = \sup_x |F(x) - \hat{F}(x)| \tag{1b}$$

The resulting compression ratio is given as, $N/\hat{N}$, where N is the number of coefficients in the original signal F, and $\hat{N}$ is the number of coefficients used to represent the approximated signal $\hat{F}$. The signal $\hat{F}$ that minimizes Eq. 1 is said to be the best approximation of the original signal F. Other criteria for quantifying the compression ratio may be defined based on the amount of memory required to store the information.

A popular technique for representing signals and for solving functional approximation problems is via expansion of the signal on a set of basis functions. The signal may be represented by Eq. 2, as a weighted sum of basis functions $\theta_i(x)$:

$$F(x) = \sum_{i=0}^{\infty} c_i \theta_i(x) \tag{2}$$

The approximation $\hat{F}$ may be achieved by ignoring the contribution of some of the basis functions. If the basis functions form an orthonormal set, then the best approximation of the original signal F is guaranteed, and the error from considering only the first K terms for the representation of $\hat{F}$ may be determined explicitly by the following formula:

$$\|F - \hat{F}\|^2_{L^2(I)} = E^2_{L^2(I)} = \sum_{k=K+1}^{\infty} c_k^2 \tag{3}$$

Since data compression may be posed as a functional approximation problem, all data compression techniques should be interpretable as approximations of the data through an expansion on a set of basis functions. This interpretation provides a consistent framework for comparing the properties of various data compression techniques, and for developing new ones. Various data compression techniques are described in the rest of this section, and are interpreted as expansion on a set of basis functions.

Piecewise linear approximation methods

The simplest of data compression methods that are being used extensively in chemical plants approximate the process signal by piecewise linear segments. The boxcar, backward slope (Hale and Sellars, 1981; Bader and Tucker, 1987a,b), swinging door (Bristol, 1990), and PLOT (Mah et al., 1995) represent the signal by a straight line until a user-specified error criterion is violated. Then, a new linear segment is initiated to continue the approximation. Data compression is achieved since only the end points of the linear segments need to be stored to reconstruct the compressed signal, resulting in a piecewise linear approximation. The error limit is usually defined to match the transducer's inherent accuracy. This avoids the measurement of any high frequency sensor noise. In general, these techniques approximate the process signal by representing it on a set of linear basis functions:

$$\hat{F}(x) = \sum_{i=1}^{\hat{N}-1} c_i \theta_i(x)$$

where

$$\theta_i(x) = m_i x + d_i, \quad x \in [x_i, x_{i+1}]$$

and

$$m_i = \frac{\hat{F}(x_{i+1}) - \hat{F}(x_i)}{x_{i+1} - x_i}, \quad d_i = \hat{F}(x_i) - m_i x_i$$

Piecewise linear compression methods are computationally fast with a computational complexity of $O(N)$, and perform well for process signals that have little noise and correspond to steady-state or slowly varying operations. The use of simple error bounds makes these methods inadequate for automatically recording moderate to high frequency, low amplitude transients, which may be very important for dynamic modeling and trend analysis. Also, the compressed signal does not provide the best approximation, since no error norm, as given by Eq. 1, is minimized. The compression parameters for piecewise linear methods are directly related to the maximum error of approximation, but significant experimentation and fine-tuning of the parameters are required for optimal performance.

Vector quantization

Vector quantization (VQ) is a technique derived from information theory, and is closely related to cluster analysis and unsupervised learning. It involves mapping parts of the signal onto a set of symbols or basis functions via an encoder. VQ relies on the fact that parts of a signal may be similar, and may be represented by the same encoder symbol. Instead of storing the signal, only the encoder symbols need to be stored, thus allowing data compression. For reconstructing the signal, the symbols are fed to a decoder which reproduces the signal. The reconstructed vector may be represented as a weighted sum of basis functions. The basis functions correspond to the codebook symbols or clusters and constitute a library of shapes which are present in the given signal. VQ minimizes the selected error criterion locally within each cluster for a range of signal values. This does not ensure the minimization of the overall least-squares error. Also, the selection of the basis functions requires significant trial and error. An overview of VQ is provided by Gray (1984), and an application to chemical process data is described by Feehs and Arce (1989). Due to its high computational expense, VQ has not been adopted by the chemical industry for on-line compression, but is a popular technique for final storage of the data.

Time-Frequency Representation of Process Data Using Wavelets

Representation of measured data in the time-frequency domain requires basis functions that are localized in time and frequency, and can cover the time-frequency space at various resolutions. In general, there are three adjustable parameters for decomposing the time-frequency space on a set of basis functions: the location of the basis function in the input or time domain u; the location in the frequency space ω; and the scale s, of the time-frequency window. The localization of the basis functions obtained by varying the time, scale and frequency parameters in a dyadic discretized manner is shown in Figure 1. The scale parameter represents the degree of localization of the basis function in the time or frequency domain. The discrete raw data may be represented at scale $s=1$ (or $m=0$), since each data point is local in time and global in frequency. For a given family of basis functions, the product $\Delta u \Delta \omega$ is a constant. Any combination of basis functions that spans the entire time-frequency domain forms a complete basis. The recent development of wavelets provides a formal framework for representing a signal in the time-frequency domain, and unifies several existing basis functions, as described in the rest of this section.

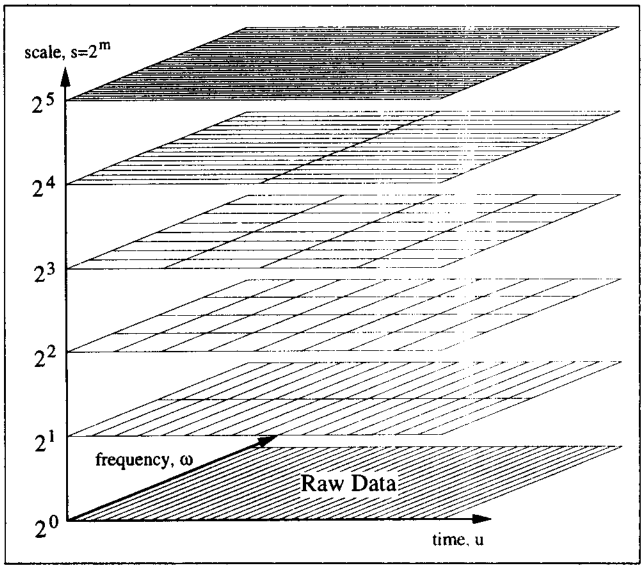

Figure 1. Basis functions in the time-scale-frequency space using dyadic discretization.

Each rectangle corresponds to the localization of a basis function.

Wavelet analysis

A family of wavelets $\psi_{su}(x)$ is derived from the translations and dilations of a mother wavelet $\psi(x)$, where the two labels s and u indicate the *dilation* and the *translation* of the mother wavelet. For practical applications to discrete data, the translation and dilation parameters may be discretized via dyadic sampling of both parameters with $s=2^m$ and $u=2^m k$ resulting in the representation of a signal of length N, by an equal number of coefficients. The wavelet coefficients d_{mk} capture the energy in the corresponding time-frequency window, and the lowest frequency scaling function coefficients a_{Lk} capture the energy at the lowest frequencies. Due to the orthonormality of the basis functions, the energy of the signal is equal to the sum of the squares of the coefficients:

$$\|F\|^2 = \sum_m \sum_k d_{mk}{}^2 + \sum_k a_{Lk}{}^2 \tag{4}$$

Since all wavelets belonging to a family are of constant shape, the dilation parameter s determines both the frequency location and the resolution of the wavelet in the time-frequency space. Therefore, a given frequency may be spanned by wavelets of a fixed resolution as shown in the decomposition of the time-frequency space by discrete dyadic wavelets in Figure 2c. For comparison, the decomposition of the time-frequency space by temporally localized delta functions and frequency localized trigonometric functions are shown in Figures 2a and 2b, respectively. More details may be found in Mallat (1989) and Daubechies (1992).

Detection of sharp changes in a signal is important in many applications like edge detection and pattern recognition. Special wavelets have been devised by Mallat and Zhong (1992) for detecting edges and inflexion points. The scaling function is chosen to be a smoothing function like a quadratic or cubic spline, or a Gaussian. The corresponding wavelet is the first derivative of the scaling function. Thus:

$$\psi(x) = \frac{d\phi(x)}{dx}$$

which implies that

$$W_{2^m u}F(x) = 2^m \frac{d}{dx} S_{2^m u}F(x)$$

Consequently, inflexion points in the scaled signal correspond to extrema in the detail signal. The translation parameter is sampled uniformly, that is, $u=n$, $n \in Z$ to enable accurate detection of the position and amplitude of the wavelet transform extrema. Mallat and Zhong (1992) have developed an algorithm for reconstructing a signal from its wavelet transform extrema.

The fast wavelet algorithm of Mallat (1989) may be applied to all the wavelets described above, with dyadic or uniform sampling of the translation parameter. An example of multiresolution analysis of a signal using dyadic and uniform

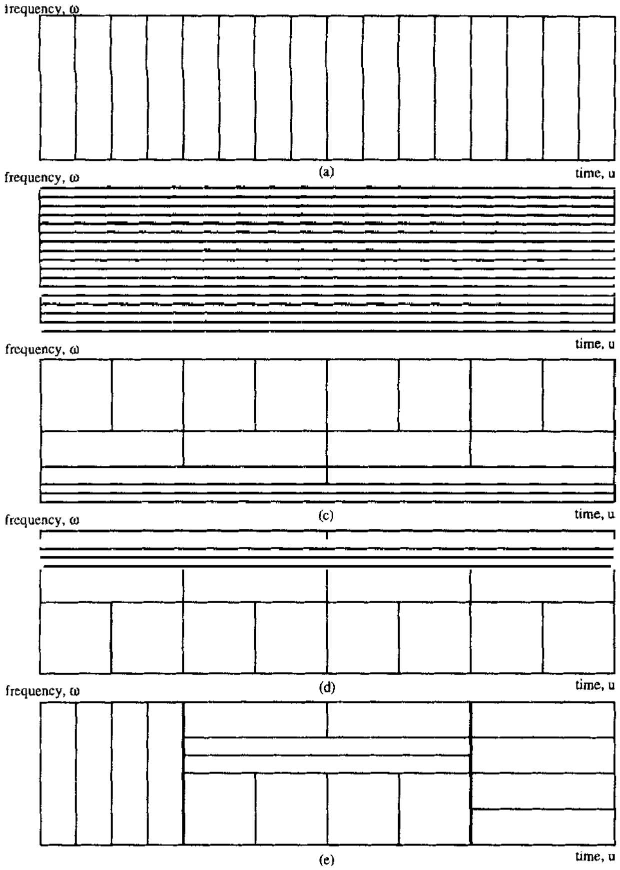

Figure 2. Decomposing the time-frequency space via (a) Dirac delta functions; (b) Fourier analysis; (c) wavelet analysis; (d) wavelet packets; (e) time-varying wavelet packets.

Decomposition provided by (a), (b) and (c) is fixed, while (d) and (e) adapt to the characteristics of the measured signal.

sampling is shown in Figure 3. All the detail signals and the last scaled signal constitute the wavelet decomposition of the original signal on an orthonormal basis with dyadic sampling in Figure 3a, and on a derivative wavelet basis with uniform sampling in Figure 3b.

Wavelet packets

The decomposition of the time-frequency space for different basis functions such as Dirac delta, trigonometric, windowed Fourier and wavelet basis are predetermined and fixed as shown in Figures 2a, 2b and 2c. The fixed, signal-independent decomposition of the time-frequency space by these basis functions makes them unsuitable for efficient representation of all types of signals. Ideally, the basis functions for decomposing a signal should be selected in a signal-dependent manner. Wavelet packets (Coifman and Wickerhauser, 1992) are an extension of wavelet analysis that subsume a variety of basis functions, such as those shown in Figures 2a–2d, and permits the selection of an adaptive basis from a library of orthonormal basis functions.

Consider the decomposition of the time-frequency space by wavelet analysis as shown in Figure 4a and the corresponding wavelet coefficients as shown in Figure 5a. A given

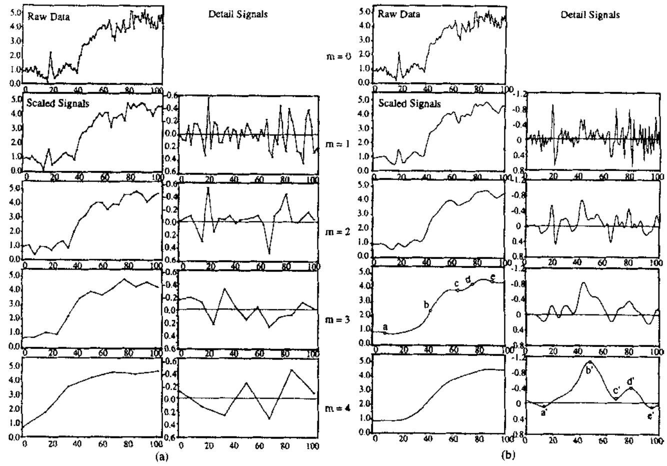

Figure 3. Wavelet decomposition.

(a) Dyadic sampling using Daubechies-6 wavelet; (b) uniform sampling using cubic spline wavelet.

point in the time-frequency space is covered by wavelet basis functions at only a single scale. The detail signal at scale $m = 1$ in Figure 5a may be further decomposed on the selected orthonormal wavelet basis resulting in decomposition of the corresponding time-frequency space as shown in Figure 4b and giving the additional coefficients in Figure 5b. The de-

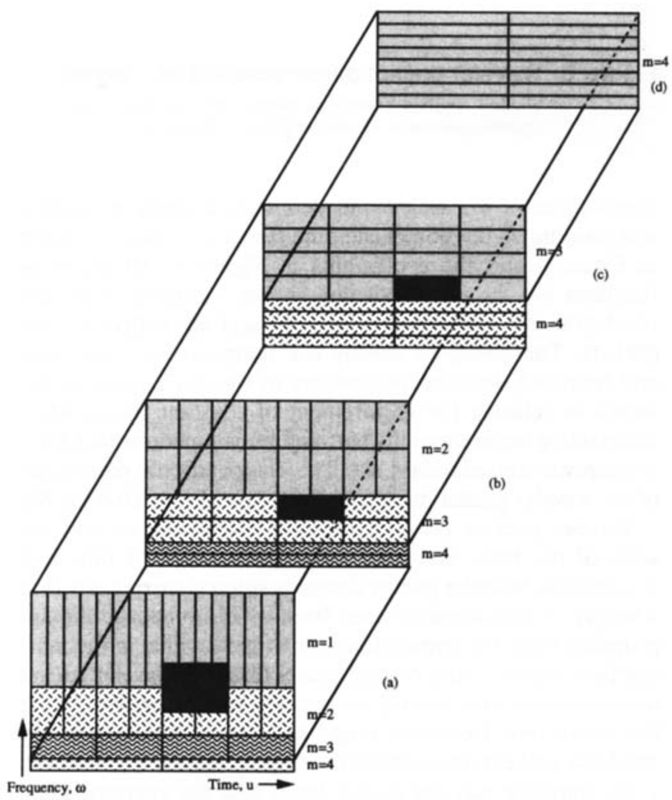

Figure 4. Decomposition of time-frequency space by wavelet packets.

Corresponding coefficients are calculated as shown in Figure 5. A signal localized in the dark region requires 6 coefficients for description by the wavelet decomposition in (a), and only 2 coefficients by wavelet packets in (b) and (c).

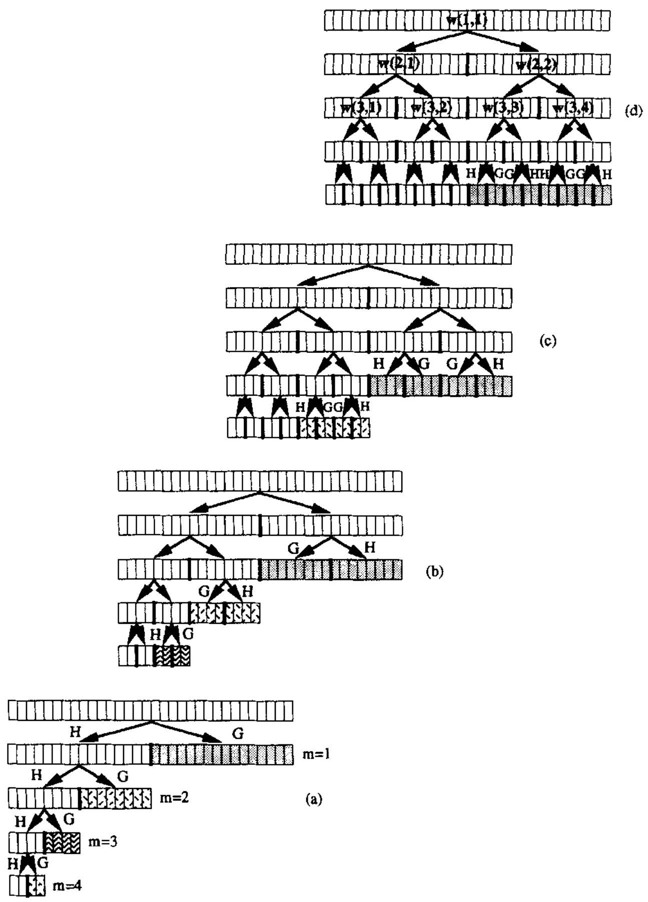

Figure 5. Wavelet packet decomposition of a signal.

Shaded regions represent coefficients for basis functions spanning similarly shaded regions in Figure 4.

composition of the scaled and detail signals may be continued resulting in division of the time-frequency space as shown in Figure 4 and the coefficients in Figure 5. All the basis functions for the decomposition shown in Figure 5 are derived from the same mother wavelet and are called wavelet packets. The ability to modify the temporal location, scale and frequency location parameters in wavelet packets is obtained by relaxing the requirement of constant shape, while maintaining orthonormality between basis functions that form a complete nonredundant set. The computational complexity of the wavelet packet decomposition algorithm is $O(N \log N)$.

Wavelet packets provide a formal framework to vary the scale of the basis functions for a given range of time and frequencies. Wavelet packet decomposition is carried out over a library of orthonormal bases from which to choose the appropriate basis for representing the original signal in the most compact manner. Any N coefficients from the wavelet packet decomposition that belong to packets which together cover the entire time-frequency range form a complete orthonormal basis and can reconstruct the original signal. Examples of some complete wavelet packet bases and the corresponding decomposition of the time-frequency space are shown in Figure 6. The library of orthonormal wavelet packets includes well-known and popular families of basis functions including trigonometric bases, windowed Fourier bases and wavelet bases. Thus, several time-frequency decomposition schemes are special cases of the library of orthonormal wavelet packets. Such a fine decomposition of the input-frequency space is likely to provide a more compact and efficient representation of input-frequency localized signals. Consider a signal covering the dark shaded region of the time-frequency domain shown in Figure 4. Six wavelet coefficients are required to represent this signal as indicated in Figure 4a, four at $m = 1$, and two at $m = 2$. On the other hand, the most compact representation of this time-frequency region requires only two wavelet packet coefficients, as shown in Figures 4b and 4c. Examples illustrating the superiority of wavelet packets for compression are given by Wickerhauser (1991). Any square integrable function may be represented as a weighted sum of orthonormal wavelet packets so that:

$$F(x) = \sum_{s} \sum_{u} \sum_{\omega} c_{su\omega}(x)\theta_{su\omega}(x) \tag{5}$$

where (u, ω) span the time-frequency space covered by the process signal, and the functions $\theta_{su\omega}(x)$ form a complete orthonormal basis. Consequently, an energy conservation equation similar to Eq. 4 may be easily derived. Efficient techniques for selecting the best bases for representing the time-frequency characteristics of the analyzed signal have been devised as described in the next section.

Compression of Chemical Process Data in Time-Frequency Domain

The basic premise behind compressing process data in the time-frequency domain is that a typical process signal may not have significant contributions in every region of the time-frequency space. The coefficients corresponding to basis functions that span regions of insignificant contribution from the signal may be neglected without losing much signal information and resulting in a compressed representation of the signal. In this section we adapt wavelet and wavelet packet-based data compression techniques for on-line compression of chemical process data through functional approximation. The cost criteria for best basis search given by Wickerhauser (1991) are modified for on-line application. We discuss the disadvantages of using wavelet packets for on-line compression, and show that time-varying wavelet packets are better suited for on-line compression with explicit control over the errors of approximation. Explicit relationships are derived between the L^2 and L^∞ errors of approximation and the neglected coefficients. Compression through feature extraction is based on using derivative wavelets and the identification and storage of only the wavelet transform extrema that correspond to significant features.

Compression via orthonormal wavelets

Signal decomposition of an orthonormal wavelet basis results in the same number of coefficients as the original signal. In order to permit on-line compression of the process data, the wavelet decomposition also needs to be performed on-line. Therefore, the wavelet decomposition methodology of Mallat (1989) is adapted for continuous signal decomposition.

Continuous Wavelet Decomposition. Due to the noncausal nature of the wavelet and scaling function filters, the wavelet

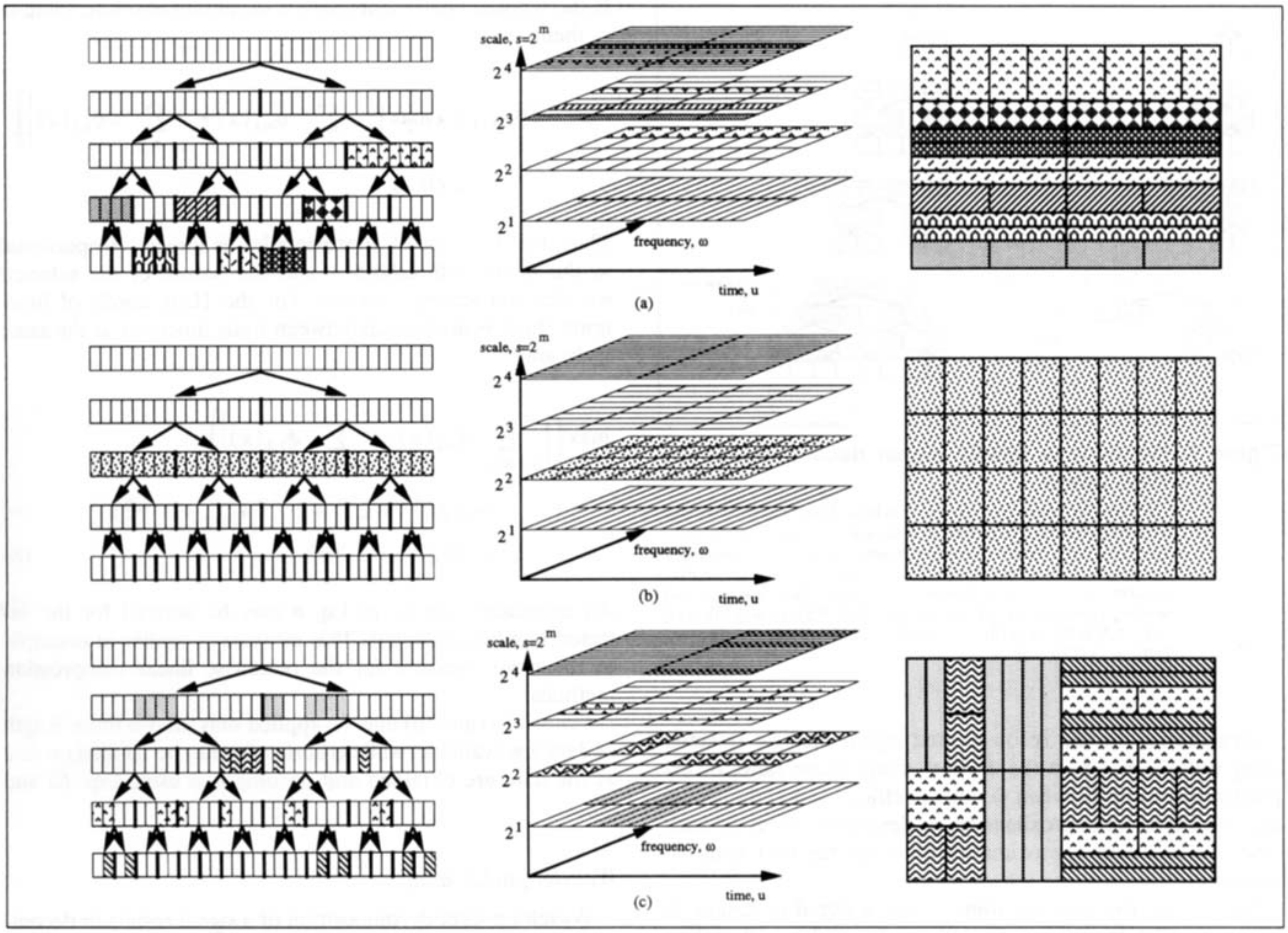

Figure 6. Some complete orthonormal wavelet packet bases.

The selected coefficients and corresponding regions in time and frequency are indicated by similar shading.

decomposition algorithm can be applied only after collecting a finite length of data. In a chemical process, since measured process data are continuously obtained, we present a scheme for computing the wavelet coefficients continuously, as soon as adequate data are available.

Consider a scaling function and wavelet with q_h and q_g denoting the number of coefficients of the corresponding filters. In general:

$$q_g = q_{g-} + q_{g+} + 1 \quad q_h = q_{h-} + q_{h+} + 1$$

where q_{g-} is the number of wavelet filter coefficients with negative locations in time, and q_{g+} is the number of wavelet filter coefficients with positive locations in time. For the well-known Haar wavelet, we have, $q_{h-} = q_{\tilde{h}+} = q_{g-} = q_{\tilde{g}+} = 0$, $q_{h+} = q_{\tilde{h}-} = q_{g+} = q_{\tilde{g}-} = 1$. In general, the H and G filters may be written as:

$$H = [h(-q_{h-})h(-q_{h-}+1)\ldots h(0)h(1)\ldots h(q_{h+}-1)h(q_{h+})]$$

$$G = [g(-q_{g-})g(-q_{g-}+1)\ldots g(0)g(1)\ldots g(q_{g+}-1)g(q_{g+})]$$

The decomposition and reconstruction filters are noncausal in nature, and require samples from the future to compute the coefficients at the present time. Consequently, the computation of the wavelet and scaling function coefficients at a scale m is delayed by $2^m q_{g-}$, and $2^m q_{h+}$ time units, respectively, and is given by:

$$d_{2^{m+1},k2^{m+1}} = [a_{2^m,(k-q_{g-})2^m} \ldots a_{2^m,k2^m} \ldots a_{2^m,(k+q_{g+})2^m}]^* G \tag{6a}$$

$$a_{2^{m+1},k2^{m+1}} = [a_{2^m,(k-q_{h-})2^m} \ldots a_{2^m,k2^m} \ldots a_{2^m,(k+q_{h+})2^m}]^* H \tag{6b}$$

On-line dyadic wavelet decomposition for the Haar bases is shown in Figure 7a and may be easily extended for using other filters. The coefficients at scale $m = 1$ may be calculated as soon as $2^m q_{g+} = 2$ points are obtained. Similarly, computation of the coefficients at higher scales is delayed by 4 measured samples at $m = 2$, 8 measured samples at $m = 3$, and so on.

Criteria for Selection of Wavelet Coefficients. The set of wavelet coefficients that provide the most compact signal

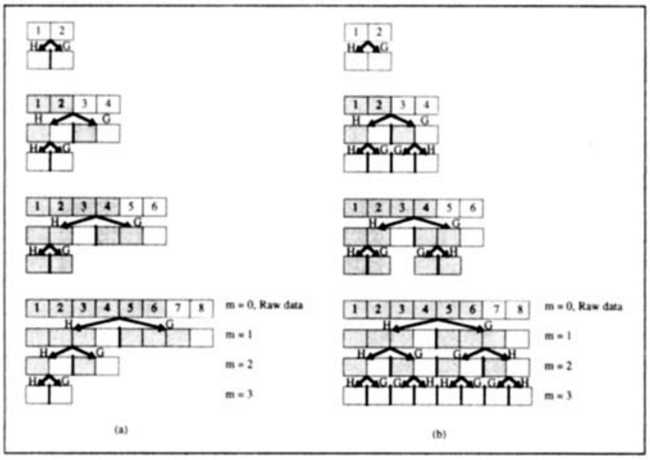

Figure 7. Continuous dyadic signal decomposition on Haar bases.

(a) Wavelet basis; (b) wavelet packets. Each coefficient can be computed by the H and G filters after $2^m q_{h+}$ and $2^m q_{g+}$ samples have been obtained respectively. Each rectangle corresponds to a coefficient. Shaded rectangles indicate previously computed coefficients. For the Haar wavelet and scaling function: $H = [h(0)\ h(1)] = [0.5\ 0.5]$, $G = [g(0)\ g(1)] = [-0.5\ 0.5]$; $H = [\bar{h}(-1)\ \bar{h}(0)] = [h(1)\ h(0)]$; $G = [\bar{g}(-a)\ \bar{g}(0)] = [g(1)\ g(0)]$.

representation in the reconstructed signal may be selected using criteria based on the desired compression ratio, or explicit relationship between wavelet coefficients and the global and local error of approximation, as described below. In each case, the resulting approximation provides the best approximation.

(a) *Desired Compression Ratio.* Since a signal of length N is represented by an equal number of wavelet and scaling function coefficients, for a desired compression ratio (CR) of C, the largest $\hat{N} = N/C$ decomposition coefficients are selected.

(b) *Acceptable Mean-Squares Error of Approximation.* The relationship between the neglected wavelet and scaling function coefficients and the mean-squares or L^2 error of approximation may be written as:

$$\left(\|F - \hat{F}\|_{L^2(I)}\right)^2 = \sum_{d_{mk} \in S_-} d_{mk}{}^2 + \sum_{a_{Lk} \in S_-} a_{Lk}{}^2 \qquad (7)$$

where S_- is the set of wavelet and scaling function coefficients that are ignored in the compression. For a given error of approximation over a predetermined length of data, the largest decomposition coefficients are retained, until the sum of their squares is greater than or equal to the specified global mean-squares error of approximation

(c) *Acceptable Local Point-Wise Error of Approximation.* Explicit control over the local error of approximation may be exercised by eliminating wavelet decomposition coefficients that are smaller than a selected threshold. The maximum local or L^∞ error of approximation is a function of the selected threshold value and the selected basis functions, as derived below:

$$\|F - \hat{F}\|_{L^\infty(I)} = \max_x \left(\left| \sum_{d_{mk} \in S_-} d_{mk} \psi_{mk}(x) + \sum_{a_{Lk} \in S_-} a_{Lk} \phi_{Lk}(x) \right| \right)$$

If the ignored coefficients have a maximum absolute value of ϵ, then:

$$\|F - \hat{F}\|_{L^\infty(I)} \leq \epsilon \max_x \left(\left| \sum_{mk \in S_-} \psi_{mk}(x) + \sum_{Lk \in S_-} a\phi_{Lk}(x) \right| \right)$$
$$= O(\epsilon)$$

Thus, the L^∞ error of approximation is directly proportional to the selected threshold ϵ and the nature of the selected wavelet and scaling function. For the Haar family of functions, there is no overlap between basis functions at the same scale and:

$$\max_x \left(\left| \sum_{mk \in S_-} \psi_{mk}(x) + \sum_{Lk \in S_-} \phi_{Lk}(x) \right| \right)$$
$$= 2(2^{-1/2} + 2^{-1} + 2^{-3/2} + \ldots + 2^{-L/2})$$
$$= 4.828(1 - 2^{-L/2}) \qquad (8)$$

An expression similar to Eq. 8 may be derived for the selected family of wavelets. This measure is similar in principle to the error measure for the piecewise linear compression methods.

Criteria (a) and (b) may be applied only after a finite length of data are available, while criterion (c) may be applied on-line as the data are obtained and decomposed using Eqs. 6a and 6b.

Wavelet packet analysis

Wavelet packet decomposition of a signal results in decomposition over several complete orthonormal bases for different combinations of the value of the translation (u), frequency (ω) and scale (s) parameters. Subsequently, it is necessary to select the basis that provides the most compact representation of the process signal.

Continuous Wavelet Packet Decomposition. A signal may be decomposed continuously on a library of orthonormal wavelet packets by simply extending the continuous wavelet decomposition procedure given in the previous section, and applying Eqs. 6a and 6b as measured data are obtained. The continuous computation of wavelet packet coefficients for Haar bases is illustrated in Figure 7b. The coefficients computed by applying the H and G filters at each scale are delayed by $2^m q_{h+}$ and $2^m q_{g+}$ sampling intervals respectively. For most signals, the best basis may not be known *a priori*, and the selection of the best basis and compression has to be performed in a batch manner after collecting a predetermined window of measured data. The requirements of continuous wavelet packet decomposition, best basis selection and compression are satisfied much more effectively within the framework of time-varying wavelet packets, described later in this section.

Best Basis Selection. Since wavelet packets consist of a library of orthonormal basis functions, it is necessary to select the set of complete orthonormal basis functions that represent the signal in the best manner (Coifman and Wickerhauser, 1992; Wickerhauser, 1991). As shown by Wickerhauser (1991), the wavelet packet tree may be searched via branch and bound in $O(N \log N)$ time by using additive measures of the information cost of a wavelet packet. Several measures of information cost have been suggested by Wickerhauser (1991). Since most of these cost measures are not particularly well-suited for the compression of chemical process data with explicit control over the error of approximation, some modified cost measures are described below:

(a) Number of Coefficients Necessary for an Acceptable Mean-Squares Error of Approximation. This cost measure is the number of coefficients with the largest absolute value that will represent the signal with an acceptable mean-squares error of approximation. These $\hat{N}_{wp}$ coefficients represent the signal with the specified L^2 error in the most compact form. The cost function is computed by eliminating the smallest coefficients for the selected wavelet packet basis until the condition:

$$\sum_{c_{su\omega} \in S_-} c_{su\omega}^2 < E_{L^2(I)}^2 \tag{9}$$

is violated, where S_- is the set of wavelet packet coefficients that are eliminated, and is of length $(N_{wp} - \hat{N}_{wp})$, where N_{wp} is the total number of coefficients in the corresponding wavelet packet.

(b) Number of Coefficients Necessary for an Acceptable Maximum Pointwise Error of Approximation. This information cost measure is given by the number of coefficients whose absolute value is greater than a selected threshold ϵ:

$$\hat{N}_{wp} = \text{length}(|c_{su\omega}| > \epsilon)$$

A relationship similar to Eq. 8 may be easily derived, indicating that the L^∞ error is $O(\epsilon)$. This criterion enables selection of basis functions that provide the most compact representation while satisfying the L^∞ error.

On-Line Compression. Performing on-line wavelet packet decomposition, selection of the best basis and data compression simultaneously is not very convenient, since it is necessary to make assumptions about the best basis or select a fixed window length of data. If the best basis for representing a signal is known *a priori*, then on-line compression may be accomplished in conjunction with decomposition on the selected basis, as described earlier. For most signals, the best basis may not be known *a priori*, and a predetermined number of samples have to be collected before selecting the best basis. If the wavelet packet cost is measured by the threshold criterion, the elimination of small coefficients to obtain the desired compression ratio may be combined with the selection of the best basis. Nevertheless, it is necessary to select a finite window of data, introducing some arbitrariness in the compression procedure and lower compression ratio, as illustrated by the case studies. A modification of the best basis selection procedure results in time-varying or adaptive wavelet packets, which overcome the need to select an arbitrary window size, and is well-suited for on-line compression.

Time-varying wavelet packet analysis

The time-frequency characteristics of a measured process signal can change with time. In order to represent such a nonstationary signal efficiently, the basis function should also

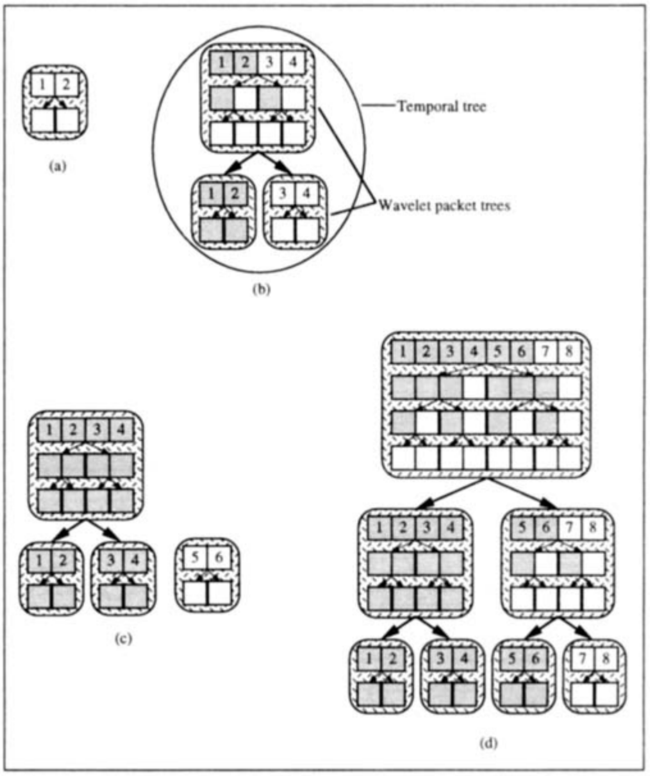

Figure 8. Continuous computation and construction of double tree on time-varying Haar wavelet packets.

Each node of the temporal tree contains a wavelet packet decomposition tree. Shaded rectangles indicate previously computed coefficients.

change with time. The best wavelet packet basis selected by the method described above can result in a time frequency decomposition, as shown in Figures 2d, 6a and 6b, which is best suited for representing signals with stationary time-frequency character. A signal with time-frequency content of the type shown in Figures 2e and 6c cannot be efficiently represented by the wavelet packets selected by Coifman and Wickerhauser's best basis search procedure. Time-varying wavelet packet analysis, developed in parallel with Herley et al. (1993), is a modification of wavelet packet analysis that is well-suited for representing stationary as well as nonstationary signals. The time-frequency decompositions possible via time-varying wavelet packets subsume all other wavelet-based techniques, and can result in any of the decompositions shown in Figures 2 and 6, while enabling continuous best basis selection and compression as described below.

Continuous Decomposition. As soon as the number of samples obtained are equal to $2q_{h+}$ and $2q_{g+}$, adequate for applying the H and G filters respectively of the selected mother wavelet, the measured signal is decomposed on a wavelet packet basis and the best basis selected using the cost criterion. As more samples are obtained, the signal may be decomposed to more levels of scale. For example, using Haar wavelet packets, every set of two data points may be decomposed to one level of scale, as shown in Figure 8a. After two more samples are obtained, the four data points taken together may be decomposed to two levels, as shown in Figure

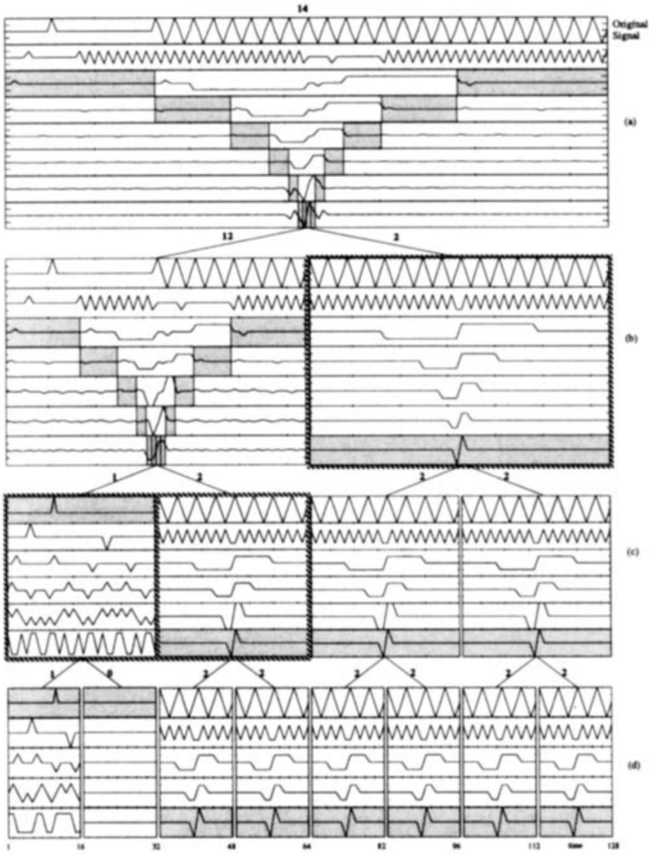

Figure 9. Double tree for selection of best basis and compression via time-varying wavelet packets.

Shaded region on each node of temporal tree shows the best wavelet packet basis for the corresponding signal selected using cost criterion (b). Minimum cost for each node of the temporal tree is written at the binary splits. The best time-varying wavelet packets are indicated by thick borders. The corresponding time-frequency representation is shown in Figure 10d.

8b. The decompositions themselves may be arranged in the form of a "double tree" with each node of the temporal tree containing a wavelet packet decomposition tree, as indicated in Figure 8b. The double tree may be constructed as more samples are obtained, as shown in Figure 8c and 8d. The best basis may now be selected by searching over this double tree.

On-Line Best Basis Selection and Compression. Due to their additive nature, the selection criteria described previously may be used for selecting the best time-varying wavelet packet bases as well as compressing the signal, via a branch and bound search of the double tree. At each node in the temporal tree the best wavelet packet basis may be searched by using the methodology described above, resulting in a cost measure for the nodes in the temporal tree. Now, the temporal tree may be searched via branch and bound to select the nodes that result in minimization of the cost criterion.

The best basis selection and compression are best illustrated by considering a simple signal consisting of a pulse followed by a sine wave. The signal is nonstationary in character since the pulse is localized in time, whereas the sinusoid is localized in frequency. A portion of the double tree for this signal is shown in Figure 9. Each node contains the wavelet packet decomposition of a portion of the signal being analyzed. Using the threshold cost criterion, Criterion (b) with $\epsilon = 0$ each node in the double tree may be searched for the best basis that captures the entire signal. At the level of the temporal tree shown in Figure 9d, each node contains the wavelet packet decomposition of segments of the signal of length 16. The best bases are selected for each segment, and denoted by the shaded regions in each node. The first 16 samples of the signal consist of the pulse at the 11th sample, and the best basis is the original segment itself. Since there is only one coefficient of value greater than the selected threshold ($\epsilon = 0$), the cost of the selected best basis is 1, as shown on the edge connecting the nodes in the temporal tree. The next segment in Figure 9d is constant at zero, and the best basis is the original segment itself with a cost of 0. For the segment from samples 33 through 48, the time-frequency character of the signal changes, and the best basis is the last level with a cost of 2 due to the two nonzero coefficients. Similarly, the nodes in Figure 9c contain the wavelet packet decomposition of segments of length 32, and the best wavelet packet bases for each segment are shown shaded. In this manner, the entire double tree may be constructed and searched for the best basis. The best time-varying wavelet packet bases selected by searching the double tree in Figure 9 are indicated by thick borders, and result in dividing the signal into three segments: the first two segments with 32 samples each containing the pulse and a portion of the sine wave respectively; and the third segment with 64 samples containing the rest of the sine wave. The entire signal may be represented by coefficients for only 5 basis functions, resulting in a compression ratio of 25.6 without any approximation error. On the other hand, compression via wavelet decomposition requires 80 coefficients, or a compression ratio of 1.6, and compression via wavelet packets using the best basis search results in 14 coefficients, or a compression ratio of 9.1. Notice that the signal decomposition in Figure 9a is the wavelet packet decomposition of the entire signal, and the shaded region corresponds to the best wavelet packet basis. The time frequency representation obtained by wavelet decomposition, wavelet packets and time-varying wavelet packets is shown in Figure 10. As expected, the temporal localization of the pulse and the frequency localization of the sine wave get diffused over several basis functions for wavelets and wavelet packets, as seen in Figures 10b and 10c, respectively. Time-varying wavelet packets, on the other hand, result in the most accurate and compact time-frequency representation, as shown in Figure 10d.

There is a clear tradeoff between the compression ratio and the computational complexity for each of the orthonormal wavelet-based methods described so far in this section. The compression ratio improves in going from wavelets to wavelet packets to time-varying wavelet packets, but the computational complexity increases from $O(N)$ for wavelets to $O(N\log N)$ for wavelet packets and $O(N\log^2 N)$ for time-varying wavelet packets. The performance of the time-varying wavelet packet method as described may be further improved by relaxing the constraint of dyadic discretization of the time-frequency domain, and incorporating techniques for selection of the best mother wavelet for a signal. In addition to providing explicit control over the global and local errors of approximation, the compression methods described above ex-

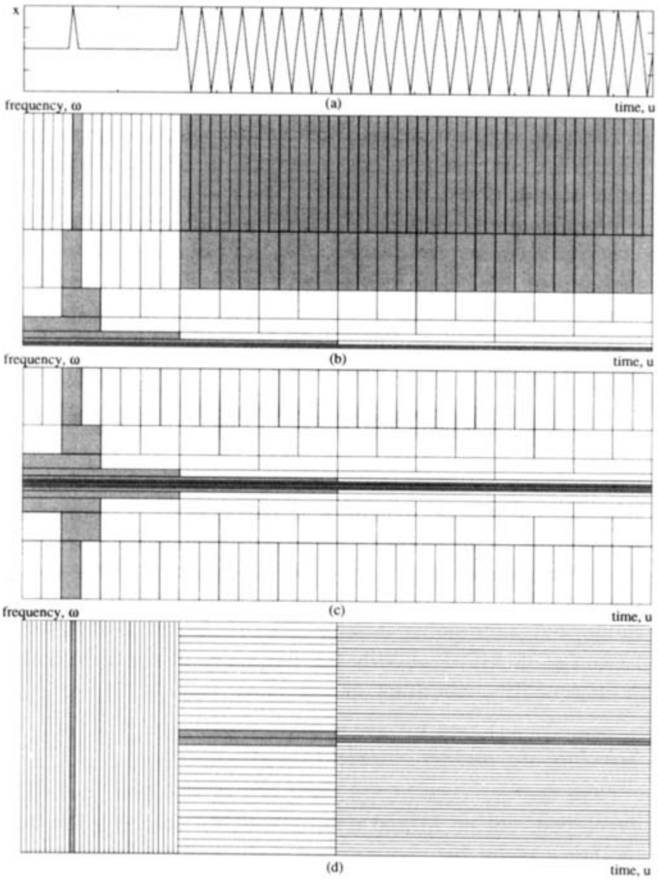

Figure 10. Time-frequency decomposition of signal in Figure 9.

Shaded region denotes nonzero coefficients. (a) original signal, 128 points; (b) dyadic wavelet decomposition, 80 points; (c) wavelet packets, 14 points; (d) time-varying wavelet packets, 5 points.

tract prominent features and minimize signal distortion. A compression technique based on derivative wavelets for compression via explicit feature extraction is described next.

Feature extraction via derivative wavelets

The compression methods described in previous sections implicitly retain conspicuous features with minimum distortion. Compression of process data through explicit feature extraction requires (1) techniques for representing features in the process data in an explicit manner to allow selection of the relevant features while satisfying the desiderata in the second section; (2) criteria for determining what features in a process signal are relevant and worth storing, and what features may be lost due to compression.

Decomposition of a signal using derivative wavelets and uniform sampling of the translation parameter allows satisfaction of both these requirements. The conspicuousness of features in the process signal may be determined based on their persistence over multiple scales. These properties of derivatives wavelets have been exploited for extracting features from process data by Bakshi and Stephanopoulos (1994). In this section, we briefly revisit some of these techniques and show their application to the compaction of process data.

Derivative wavelets are useful for feature extraction because they provide (1) a translationally invariant representation of the signal at multiple scales; (2) an unambiguous representation of the signal features, and their evolution over frequencies. As described in the third section, and portrayed in Figure 3b, the inflexion points in the scaled signal obtained via a derivative wavelet at a scale m appear as extrema in the wavelet transform at the scale $(m+1)$. Criteria for determining the most relevant or conspicuous features in a process signal may be developed by connecting the wavelet transform extrema at all scales to generate a wavelet interval tree for the decomposition in Figure 3b. Such a tree representation is useful for storing the process signal containing information about only those features selected by the user or automatically using specific criteria given below.

Selection of Wavelet Transform Extrema. The automatic selection of conspicuous features is based on the stability of episodes and extrema. The region between adjacent extrema is called an *episode*. Each rectangle in the wavelet interval tree represents an episode over its frequency range. The *stability of an episode* is defined as the number of scales over which the episode persists without the appearance of new extrema in the corresponding episode at lower scales. Similarly, the *stability of an extremum* is defined as the number of scales over which it persists, without new extrema appearing in its range. The range of an extremum is equal to the region covered by the G filter at the corresponding position and scale, and is equal to $2^{m+1}q+1$, where the filter is of length $2q+1$, and there are 2^m-1 zeros between consecutive filter coefficients.

Conspicuous temporally localized features can be represented by selecting those episodes and extrema that are most stable. Some criteria for selecting the wavelet transform extrema for representing the compressed signal are illustrated below and described in greater detail in Bakshi and Stephanopoulos (1994, 1995).

(1) *Stable Episodes at a Given Scale.* The trend reconstructed by selecting stable extrema at scale 3 is shown in Figure 11g. It is qualitatively identical to the scaled signal at scale 2 in Figure 3b, but is quantitatively more accurate.

(2) *Witkin's Stability Criterion.* The reconstructed trend obtained by selecting the episodes and extrema based on Witkin's stability criterion (Witkin, 1983) is shown in Figure 11f.

(3) *Threshold Criterion.* Extrema of magnitude larger than a user specified threshold are retained. This criterion is similar to the threshold criterion for selecting wavelet packet coefficients, and is related to the L^∞ error.

The signal may be reconstructed from the selected wavelet transform extrema by an iterative process described by Mallat and Zhong (1992). The performance of this technique is illustrated in Figures 11f and 11g.

Practical Issues in Compression of Chemical Process Data

Several practical and implementation-related issues arise in the utilization of the techniques described in this article for compression of chemical process data. Some of these issues are related to the application of the specific methodology to real-time process data, while others are related to the familiarity of the plant personnel with currently used process

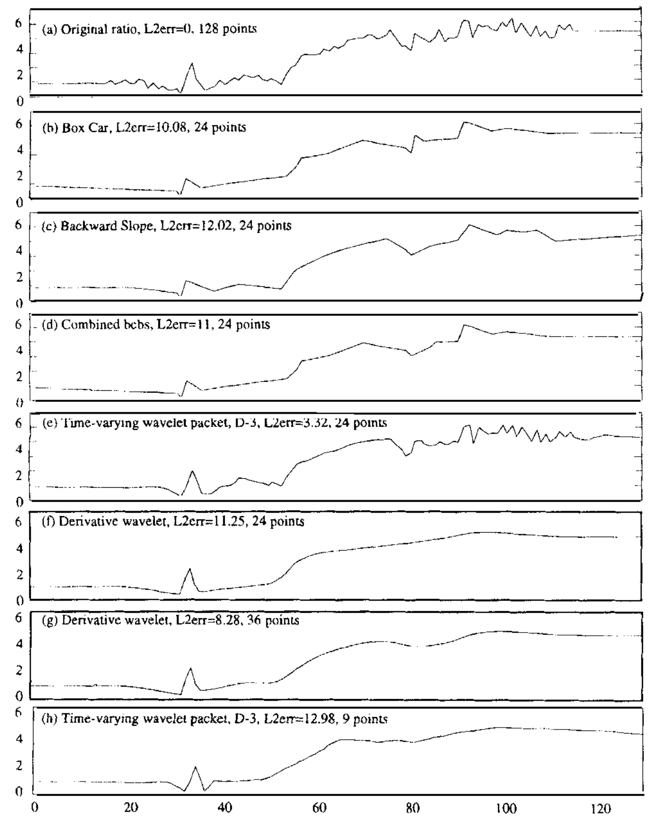

Figure 11. Comparison of compression by various techniques.

historians based on the piecewise linear approximation techniques.

Inaccuracies due to end effects

A practical problem in decomposing and reconstructing signals using wavelets involves the errors introduced due to boundary effects caused by the finite length of the measured signal. Direct application of the filters to the data at the boundaries is not possible due to inadequate data. Traditionally, the data at the boundaries are filtered by assuming a mirror image of the signal beyond its end points. Depending on the filters used, the mirror image assumption results in errors in the wavelet coefficients, as well as the reconstructed signal near the end points. Such errors are unacceptable, particularly for data that are inherently finite in nature, such as the measured signal obtained from batch processes. Techniques for eliminating end effects have been devised based on the modifying the signal or the filter coefficients.

Signal Augmentation by Constant Segments. The finite signal is augmented by padding it with constant segments at both ends with value equal to the respective end point. The length of the fictitious padded segment is determined based on the nature of the decomposition and reconstruction filters. In general, if the signal is augmented at both ends by a pad of length q_g^m, then boundary effects may be eliminated. The augmented signal results in an increase in the computation time, which increases exponentially with the number of scale levels m. The effect on the compression ratio is minimal since a constant segment results only in zero wavelet coefficients which are not stored in the compressed representation. A similar approach may be utilized to eliminate errors and signal discontinuities due to windowing. The appropriate length of the measured signal may be used on the lefthand side, and a constant segment on the right of the window. Such techniques are also used for computing the Fourier transform of finite signals (Press et al., 1992).

Modification of Boundary Filters. An alternate, and more efficient approach for dealing with end effects involves modification of the filter coefficients applied near the signal boundaries. The application of a wavelet and scaling function filters with dyadic sampling may be represented in matrix form as:

$$y = Tx$$

where x is a vector of the scaled signal, and y is a vector of the scaled and detail signals at a coarser scale. The matrix T is made up of filter coefficients, and ideally, all rows of T should be orthonormal. Due to the finite length of the signal, the rows of T that operate near the boundaries are not orthonormal. Techniques developed by Herley et al. (1993) and Cohen et al. (1993) modify the boundary filters to maintain orthogonality of the filter matrix T.

Selecting the mother wavelet

Over the last decade, several types of wavelets have been developed, but no formal criteria exist for selecting the best mother wavelet for compressing a given signal. Usually, the mother wavelet is selected based on qualitative criteria and experience-based heuristics, some of which are presented here. The accuracy of the error relationships such as Eqs. 7 and 8 also depends on the selected wavelet. Orthonormal wavelets that are compactly supported and corrected for boundary effects provide the most accurate satisfaction of the error relations. For wavelets that are not compactly supported, such as the Battle-Lemarie family of wavelets, the truncation of the filters contributes to the error of approximation in the reconstructed signal, resulting in a lower compression ratio for the same approximation error. The smoothness of the reconstructed signal depends directly on the nature of the basis functions. Several orthonormal wavelets with different degrees of smoothness have been designed. The orthonormal wavelets with compact support designed by Daubechies (1988) are reasonably smooth for orders greater than 6. The Haar wavelets provide piecewise constant approximation which may be adequate for some process signals, such as those of manipulated variables. Among derivative wavelets, quadratic and cubic wavelets are described by Mallat and Zhong (1992). The first derivative of a Gaussian is an infinitely differentiable wavelet. *A priori* knowledge of the nature of the signal may be used to select the mother wavelet based on its smoothness. For applications requiring a large compression ratio, wavelets with high degrees of smoothness are preferred since the wavelets may become visible in the reconstructed signal. The effect of the selected mother wavelet on the compression of a process signal is illustrated by Example 2 in the next section. The cost criterion and method used for selection of the time-varying best basis described in the previous section may also be used for selecting the mother wavelet. Wavelet packets based on several mother wavelets may be evaluated, and the one that minimizes the cost criterion may be selected. The additional freedom provided by choosing the best mother wavelet increases the computational complexity by a constant factor of the number of wavelets considered N_w, resulting in a complexity of $O(N_w N \log^2 N)$. If multiple mother wavelets are used for different portions of a signal, discontinuities and inaccuracies may be introduced near the boundary where the new wavelet is introduced. Once again, the boundary filters may be orthogonalized as suggested above.

On-line vs. archival compression

The practical issues as well as the speed of data compression vary depending on the stages of data acquisition at which compression is performed. Every measured variable may be compressed as the data are collected on-line. This type of compression is most common in the chemical process industry today through the use of process historians (Kennedy, 1993). In intelligent sensors it may even be necessary to do some preliminary data compression as the data are collected. The techniques described in this article are best suited for such on-line application to a single measured variable.

Usually data for several measured variables are collected for several days or weeks, and then stored into the company data archives. These data may be retrieved at a later stage for studying various aspects of the process operation. If online compression techniques were applied, the archived data are compressed to some degree. Additional compression may be achieved by eliminating the redundancy between multiple measured signals. An integrated method for on-line as well as archival compression is presently being developed (Bakshi, 1994).

Selecting the compression criterion

Several different criteria for selection of the wavelet packets and decomposition coefficients are given throughout the article. The threshold criterion is best suited for on-line compression using time-varying wavelet packets or derivative wavelets. The explicit relationship between the local L^∞ error of approximation and the threshold value provides a useful guide for selecting this adjustable parameter.

Retaining actual data points in the compressed signal

Currently used data compression techniques such as box car, backward slope and swinging door methods retain actual measured values of some of the points, while sacrificing the best approximation of the signal, whereas wavelet-based methods retain an approximation of the measured values, while providing the best approximation of the signal. Due to the familiarity of plant operators and process engineers with piecewise linear compression techniques, there may be some resistance to a method that does not store the actual measured values. As can be seen from all the examples solved in this article, all wavelet-based methods retain the most conspicuous features in the signal with minimum distortion depending on the compression ratio. In fact, compression via feature extraction explicitly ensures the accuracy of the most conspicuous features.

Case Studies

Several case studies of compression of chemical process data are described in this section. The performance of the wavelet-based data compression techniques is compared with currently used conventional techniques. The properties of data compression via feature extraction and functional approximation are also illustrated. All the data compression algorithms are implemented in MatLab on a Macintosh personal computer.

Example 1

The raw data in Figure 11a consists of 128 points. The plots in Figures 11b through 11f show the reconstructed data from only 24 of the original 128 points, (CR = 5.33). Figures 11b, 11c, and 11d are the result of applying the box car, backward slope and combined box car and backward slope methods, respectively. Each of these methods results in a piecewise linear approximation of the signal. The data reconstructed after time-varying wavelet packet compression using the Daubechies third-order wavelet (Daubechies, 1988) for the same compression ratio are shown in Figure 11e. Finally, the result of compression via feature extraction using derivative wavelets is shown in Figure 11f. Notice that the signal in Figure 11e obtained from time-varying wavelet packet compression performs best by giving the smallest least-squares error of approximation and retaining the most information for the given compression ratio. The signals obtained via the piecewise linear methods and via derivative wavelets result in similar least-squared errors, but the piecewise linear methods cause significant distortion of features in the signal, such as the spike at $t = 34$. The reconstructed signal in Figure 11f consists of only the most conspicuous features determined by applying Witkin's stability criterion. A signal similar to that in Figure 11f may also be obtained by compression with Daubechies-3 time-varying wavelet packets, as shown in Figure 11h, if only 9 coefficients are retained (CR = 14.22). Figures 11f and 11h indicate that compression with time-varying wavelet packets gives a greater degree of compression than compression with derivative wavelets, but derivative wavelets provide a translationally invariant representation and explicit feature extraction.

Comparison of the performance of piecewise linear and wavelet-based compression methods indicates the superiority of wavelet-based methods. The L^2 and L^∞ errors of approximation are plotted vs. a range of compression ratios in Figure 12. Among piecewise linear methods, the box car gives the best performance which is compared with the performance of the Daubechies-3 time-varying wavelet packets. Time-varying wavelet packets, as well as wavelet packets, perform significantly better than piecewise constant methods while minimizing the distortion of conspicuous features in the signal.

Example 2

The selection of the mother wavelet for decomposing a signal on a wavelet packet basis is often based on qualitative or heuristic arguments since no formal criteria are available. We have studied the compression of several sets of industrial process data using a variety of wavelets, and the results of wavelet packet compression for the raw data shown in Figure

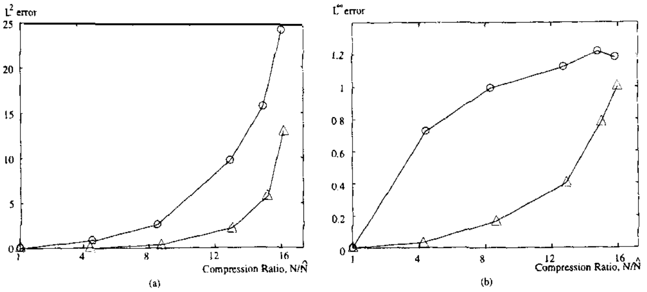

Figure 12. Relative performance of piecewise linear and orthonormal wavelet packet compression for the raw data shown in Figure 11a.

(a) Variation of L^2 error with compression ratio; (b) variation of L^∞ error with compression ratio. Original signal has 128 points. (○) Piecewise constant compression methods, box car; (Δ) Time-varying wavelet packet compression, Daubechies-3.

11a are depicted in Figure 13 for several orthonormal wavelets and a compression ratio of 4. The filter coefficients for the Daubechies wavelets are available in Daubechies (1988). The relationship between properties of the mother wavelet such as its smoothness, length of compact support, and oscillatory nature, and the nature of the reconstructed signal and the error of approximation for a given compression ratio are illustrated by the reconstructed signals in Figure 13. The reconstructed signals have the same degree of smoothness or regularity as the underlying mother wavelet. For example, the

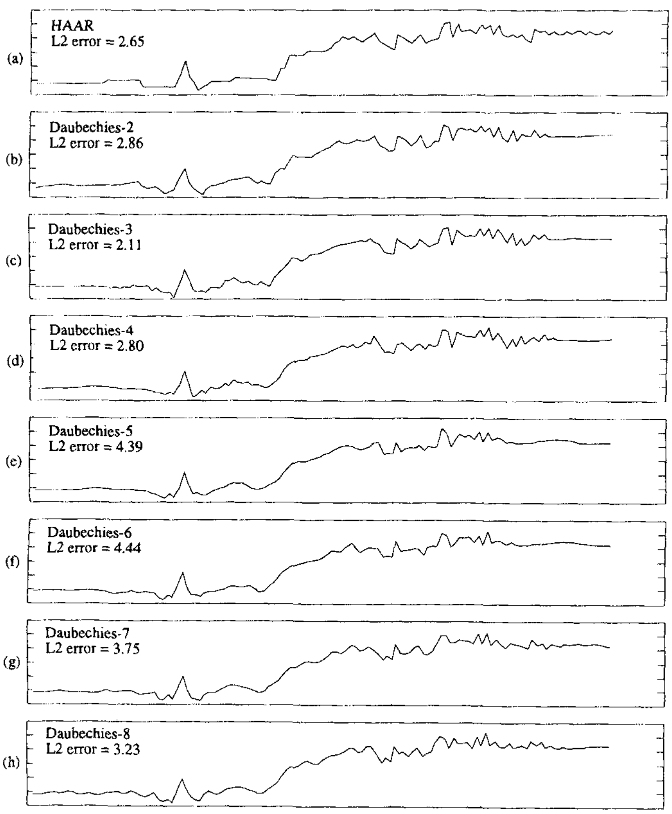

Figure 13. Reconstructed signal for raw data in Figure 11a using various orthonormal wavelets and a compression ratio of 128/32 = 4.

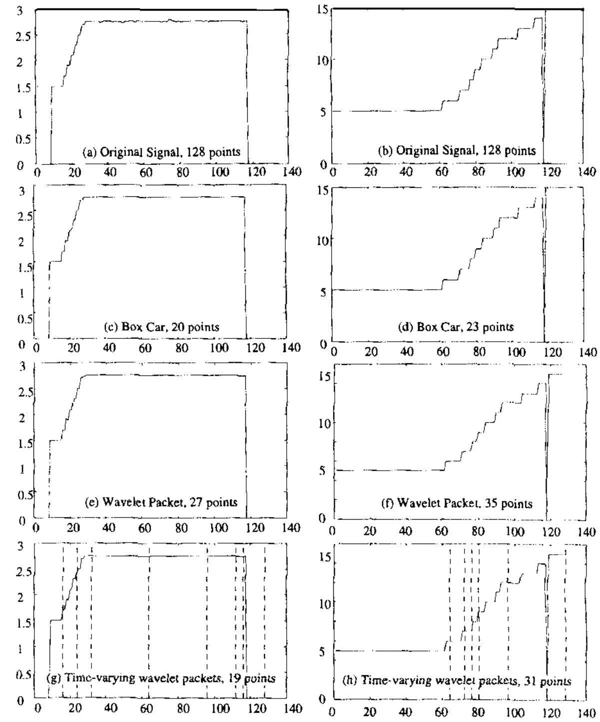

Figure 14. Compression of piecewise linear signals.

signals reconstructed from Haar wavelet packets are piecewise constant, as shown in Figure 13a. The absence of high degrees of smoothness of the low-order Daubechies wavelets is seen in the irregular nature of the approximations in Figures 13a, 13b and 13c, as opposed to the increasingly smooth nature of the curves in going from in Figures 13d to 13h.

Example 3

Manipulated variables in chemical processes often have regions of constant behavior and a few sharp changes, reflecting operator action. Typical examples of the signal generated by a manipulated variable for a batch process are shown in Figures 14a and 14b. These signals are taken from measured data for an industrial fed-batch fermentor. Such piecewise constant signals should be ideally suited for piecewise linear approximation techniques. The results of applying the box car method, Haar wavelet packets and time-varying Haar wavelet packets for compressing the signal are shown in Figures 14c through 14h. Haar wavelets are selected since their piecewise constant nature matches with that of the signals, and should result in better approximation than that obtained by other smoother wavelets. The box car method does outperform Haar wavelet packets by providing higher degrees of compression for an acceptable signal. As described in the second section, the box car method was designed for compressing exactly such signals that have stretches of constant or linear operation. A larger number of Haar wavelet packet coefficients are required for comparable errors of approximation because of the dyadic nature of the wavelet packet decomposition. Time-varying wavelet packets perform better than the box car method for the signal in Figure 14a, and worse for the signal in Figure 14b. Current research in the use of time-varying wavelet packets for compression is directed towards relaxing the dyadic nature of the discretization, and is ex-

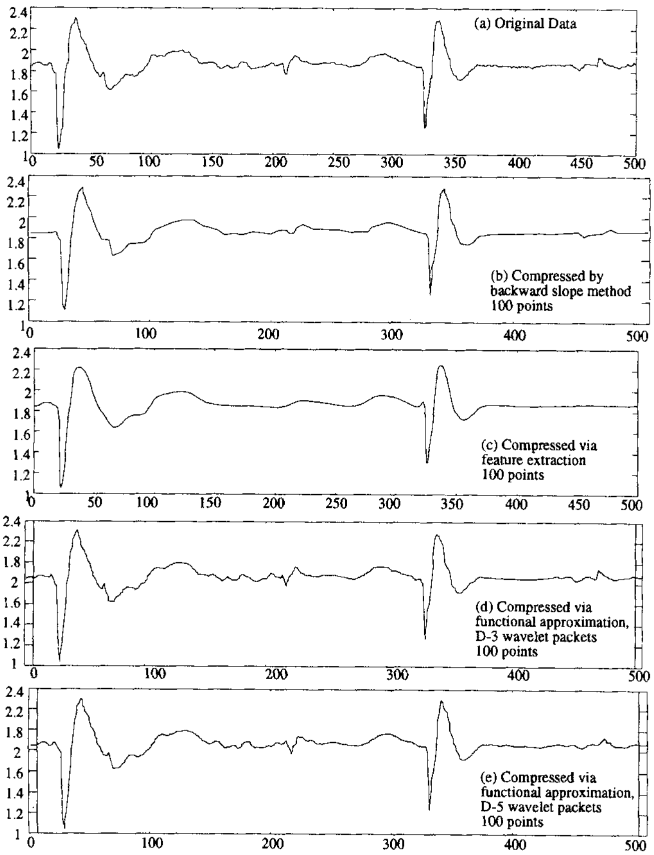

Figure 15. Compression results for Example 4.

(a) Original signal; (b) by backward slope method; (c) using wavelet transform extrema; (d) using Daubechies-3 wavelet packets; (e) using Daubechies-5 wavelet packets.

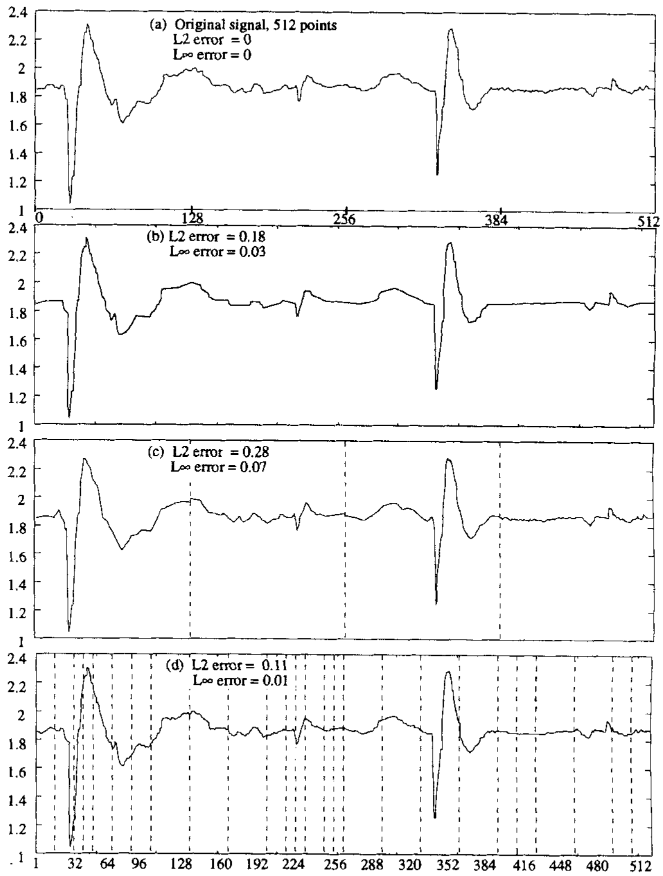

Figure 16. Results of on-line compression for compression ratio of 512/166 = 3.1.

(a) Original signal; (b) compression via box car method; (c) compression via wavelet packets, window length selected arbitrarily as 128 samples; (d) compressed signal using time-varying Haar wavelet packets, window length selected adaptively for best compression.

pected to result in better performance with signals such as that in Figure 14b. Nevertheless, piecewise linear methods have a smaller computational complexity of $O(N)$, as compared to the complexity of time-varying wavelet packets of $O(N\log^2 N)$, and may be better for piecewise linear signals.

Example 4

This example illustrates the behavior of the compression methods for a signal consisting of sharp changes. A portion of the signal representing pressure variation in an industrial flare tower consists of 512 points, as shown in Figure 15a. The reconstructed signal obtained by retaining only 100 points is shown in Figures 15b through 15e for various compression methods. Once again, compression with wavelet packets using the Daubechies-3 wavelet gives the best quality of the compressed signal. Due to the presence of sharp changes, small, localized oscillations, or a Gibbs phenomenon may be exhibited near the sharp change for derivative wavelets, as shown in Figure 15c at about 330 s.

Example 5

On-line compression using the box car method, wavelet packets, and time-varying wavelet packets are compared for a compression ratio of 3.1 (= 512/166). On-line compression using wavelet packets is implemented in a batch manner by applying wavelet packet compression after an arbitrarily determined finite window of 128 samples is collected. The box car method and time-varying wavelet packets do not require assumptions of a finite window of data before attempting compression. The raw data are shown in Figure 16a, which is the same as that in Figure 15a. The selection of wavelet packet compression is based on achieving a maximum L^∞ error of 0.01 in each window of data. The compressed signal and the least-squares error in each window are shown in Figure 16c. Comparing the results in Figure 16c with those of the box car method in Figure 16b shows that on-line compression using wavelet packets using finite and fixed windows of data gives quite poor compression. Time-varying wavelet packets adapt the window size for wavelet packet decomposition based on the time-frequency character of the signal, and result in the best performance, as shown in Figure 16d.

Conclusions

The process data compression techniques presented in this article are based on decomposing the signal on basis functions that are localized in the time-frequency domain. Physically meaningful criteria such as the acceptable loss of accuracy or relevance of features are used for selecting the "best" bases and coefficients for representing the compressed signal. Each compression method is based on multiresolution analysis using techniques derived from wavelet theory. Existing wavelet and wavelet packet decomposition algorithms are modified for on-line implementation. Combined on-line decomposition, selection of the best orthonormal bases, as well as compression for satisfaction of a user-specified maximum pointwise error of approximation is achieved via the technique of adaptive, or time-varying wavelet packets.

Unlike currently used techniques for compression of chemical process data such as the box car, backward slope, and swinging door methods, and other techniques such as vector quantization, data compression in the time-frequency domain possesses several advantages and satisfies the desiderata for an efficient process data storage and retrieval technique. Thus, wavelet-based data compression:

- Stores the data in terms of contributions in the time-frequency domain, providing a consistent representation that may be exploited effectively by other process operation tasks (Bakshi, 1994) and may propitiate integrated process operation
- Provides greater compression for the same degree of approximation, or signal fidelity
- Is physically intuitive due to the explicit mathematical relationship between the orthonormal wavelet and wavelet packet coefficients, and the local and global errors of approximation. For derivative wavelets, compression may be achieved by extracting the most conspicuous or relevant features in the process signal
- Orthonormal wavelets and wavelet packet bases guarantee the best approximation of the original signal at all compression ratios, with minimum distortion of the conspicuous features.

Wavelet theory is a fast developing field, and new developments are likely to further improve the data compaction ability of wavelet-based techniques. The superior performance and characteristics of wavelet-based data compression techniques, as well as the practical issues in their implementation, are illustrated via case studies using industrial data.

Acknowledgments

We thank Stephane Mallat and coworkers for the use of their "xwavel" program. The early part of this research was done while the first author was at Aware, Inc., Cambridge, MA. Their support and computer resources are gratefully acknowledged.

Literature Cited

Bader, F. P., and T. W. Tucker, "Data Compression Applied to a Chemical Plant Using a Distributed Historian Station," *ISA Trans.*, **26**(4), 9 (1987a).

Bader, F. P., and T. W. Tucker, "Real-Time Data Compression Improves Plant Performance Assessment," *InTech*, 53 (1987b).

Bakshi, B. R., "Efficient Storage and Retrieval of Process Data for Improved Process Monitoring and Control," *AIChE Meeting*, San Francisco, (Nov. 1994).

Bakshi, B. R., and G. Stephanopoulos, "Representation of Process Trends: III. Multi-Scale Extraction of Trends from Process Data," *Comp. Chem. Eng.*, **18**(4), 267 (1994).

Bakshi, B. R., and G. Stephanopoulos, "Reasoning in Time; Modeling, Analysis and Pattern Recognition of Temporal Process Trends," *Paradigms of Intelligent Systems in Process Engineering*, Chap. 8, G. Stephanopoulos and C. Han, eds., Academic Press (1995).

Bristol, E. H., "Swinging Door Trending: Adaptive Trend Recording?" *ISA Conf. Proc.*, p. 749 (1990).

Cheung, J. T-Y., and G. Stephanopoulos, "Representation of Process Trends: II. The Problem of Scale and Qualitative Scaling," *Comp. Chem. Eng.*, **14**, 511 (1990).

Coifman, R. R., and M. V. Wickerhauser, "Entropy-Based Algorithms for Best Basis Selection," *IEEE Trans. on Inform. Theory*, **38** (2), 713 (1992).

Daubechies, I., "Orthonormal Bases of Compactly Supported Wavelets," *Comm. Pure and Applied Math.*, **XLI**, 909 (1988).

Daubechies, I., *Ten Lectures on Wavelets*, SIAM Press, Philadelphia (1992).

Feehs, R. J., and G. R. Arce, "Vector Quantization for Data Compression of Trend Recordings," Univ. of Delaware, Dept. of Electrical Engineering, *Technical Report*, 88-11-1 (1988).

Gray, R. M., "Vector Quantization," *IEEE ASSP Magazine*, 4-29, (Apr. 1984).

Hale, J. C., and H. L. Sellars, "Historical Data Recording for Process Computers," *Chem. Eng. Prog.*, 38 (Nov., 1981).

Herley, C., J. Kovacevic, K. Ramchandran, and M. Vetterli, "Tilings of the Time-Frequency Plane: Construction of Arbitrary Orthogonal Bases and Fast Tiling Algorithms," *IEEE Trans. Sig. Proc.*, **41**,12 (1993).

Kennedy, J. P., "Data Treatment and Applications: Future of the Desktop," *Proc. Int. Conf. on Foundations of Computer Aided Process Operations*, D. Rippin, J. Hale, and J. Davis, eds., CACHE, Austin, TX (1993).

MacGregor, J. F., "Statistical Process Control of Multivariate Processes, *IFAC ADCHEM, Kyoto, Japan* (1994).

Mah, R. S. H., A. C. Tamhane, S. H. Tung, and A. N. Patel, "Process Trending with Piecewise Linear Smoothing," *Comp. Chem. Eng.*, **19**(2), 129 (1995).

Mallat, S. G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pat. Anal. and Mach. Intel*, **11**(7), 674 (1989).

Mallat, S. G., Zero Crossings of a Wavelet Transform, "IEEE Trans. on Inform. Theory, **37**(4), 1019 (1991).

Mallat, S., and S. Zhong, "Characterization of Signals from Multiscale Edges," *IEEE Trans. Pat. Anal. and Mach. Intel.*, **14**(7) 710 (1992).

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York (1992).

Venkatasubramanian, V., and G. Stanley, "Integration of Process Monitoring, Diagnosis and Control: Issues and Emerging Trends," *Proc. Second Int. Conf. on Foundations of Computer Aided Process Operations*, D. Rippin, J. Hale, J. Davis, eds., CACHE (1993).

Wickerhauser, M. V., *INRIA Lectures on Wavelet Packet Algorithms*, Yale Univ., New Haven, CT (1991).

Witkin, A. P., "Scale Space Filtering: A New Approach to Multi-Scale Description," *Image Understanding*, S. Ullman and W. Richards, eds., Ablex, Norwood, NJ, p. 79 (1983).

Manuscript received Dec. 2, 1994, and revision received Mar. 27, 1995.