Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets

Jionghua JIN and Jianjun SHI

Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, MI 48109-2117
(*jhjin@engin.umich.edu*)
(*shihang@engin.umich.edu*)

Tonnage information is referred to as stamping force measurement in a complete forming cycle. Tonnage data contains rich information and features of stamping process failures. Due to its nonstationary nature and lack of physical engineering models, tonnage information cannot be effectively compressed using conventional data-compression techniques. This article presents a statistical method for "feature-preserving" data compression of tonnage information using wavelets. The technique provides more efficient data-compression results while maintaining key information and features for process monitoring and diagnosis. Detailed criteria, algorithms, and procedures are presented. A real case study is provided to illustrate the developed concepts and algorithms.

KEY WORDS: Data shrinkage; Feature-lossless; Multiresolution analysis; Process monitoring and diagnosis; Signal segmentation; Stamping tonnage information processing.

1. INTRODUCTION

Sheet-metal stamping is a very complex manufacturing process. In recent years, stamping tonnage sensors have been used widely to measure the stamping force for each stamped part for the sake of stamping process monitoring and fault diagnosis. This tonnage signal contains rich information and features of stamping process failures. An example of a stamping press, tonnage sensors, and a signal is shown in Figures 1 and 2. Figure 1 shows the tonnage sensors and some major press components related to press maintenance work. Figure 2 shows the total tonnage or stamping force, which is the sum of the outputs of all tonnage sensors mounted on the press. In Figure 2, the horizontal axis is the crank angle. A complete stamping cycle corresponds to the crank angle from 0° to 360°. Figure 2 shows only the relevant portion of a tonnage signal during the forming stage. The forming stage here refers to the period when the blank is pressed between the upper and the lower dies to produce the final part. The crank angle can be divided into several segments according to the different forming stages of a stamping process. Based on stamping engineering knowledge, different potential process/system failures may occur in different forming stages. Those potential failures and their corresponding segments of the tonnage signal in Figure 2 are listed in Table 1. More discussion on the segmentation and failures will be provided in Section 3. When different failures occur in production, the tonnage signal changes accordingly. Thus, features can be extracted from the tonnage signal for the purposes of process monitoring and diagnosis. A few examples of tonnage signatures/features and their corresponding process failures are shown in Figure 3. In the figure, excessive snap fault, caused by an extreme negative force, happens only at the specific forming stage when the upper die starts to rise and move away from the lower die. The gib-chatter problem shows jagged movement phenomena when the press goes down or up with no contact between the upper and lower dies. The problems of loose tie rods and worn bearings occur at the same stage of the press's bottom position, but they show different features of the peak tonnage. A flat peak shape is the feature of a loose tie rod, and an oscillation peak shape is the feature of a worn bearing. For the purpose of process monitoring and fault diagnosis, various tonnage signal-analysis techniques have been developed to perform feature extraction when the fault samples are available (Koh, Shi, and Williams 1995; Koh, Shi, and Black 1996; Jin and Shi in press).

Even though the techniques for tonnage signal analysis have been studied for stamping process monitoring and fault diagnosis, the question of how to compress the tonnage signal efficiently, while still retaining all of its potential useful features, has not been investigated thoroughly. This issue has become critical due to high stamping productivity and high demands on historical data storage. As an example, a typical progressive press can perform 200 strokes per minute. For each part, more than 1,500 data points will be measured using one tonnage sensor with a sampling interval of .1° of crank angle. If four sensors or channels are used, more than 6,000 data points will be collected in each stroke. In mass production, 2.88×10^6 data points will be sampled for 30 presses in 16 hours of production. Thus, performing data reduction for data storage is a critical problem. Furthermore, the advancement of teleservice and remote diagnosis (Lee 1995) requires in-process measurement data transfers through the Internet. In this

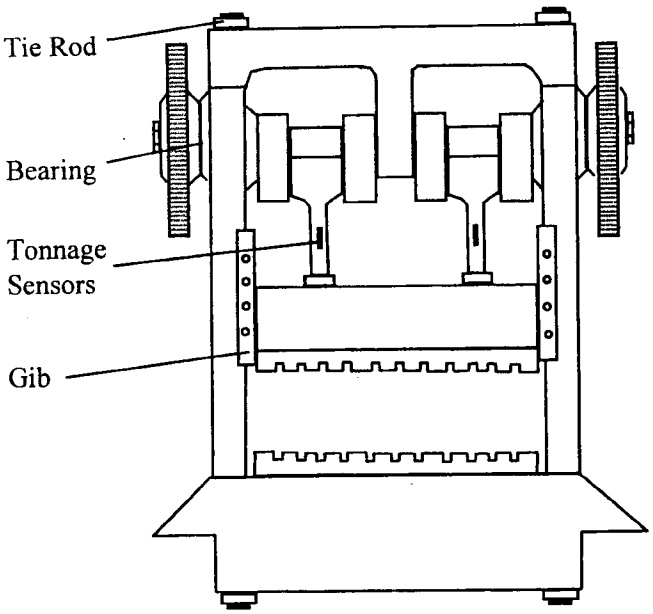

Figure 1. A Stamping Press and Tonnage Sensor Locations.

situation, development of an effective, efficient, and reliable data-compression technique has become an essential research challenge in remote diagnosis and teleserving applications (Shi, Ni, and Lee 1997).

This article studies the problem of how to apply wavelet analysis techniques in the process monitoring and diagnostic applications. More specifically, the problem is how to conduct the data compression for the purpose of process monitoring and diagnosis. It should be pointed out that data compression for the purposes of process monitoring and diagnosis is different from those data-compression methods used for a smooth estimate of the underlying function (Coifman and Wickerhauser 1994). In those data compressions, the analysis results in characterizing data as either "signal/estimated function" or "noise" (Coifman and Yale 1992; Donoho and Johnstone 1994, 1995; Donoho, Johnstone, Kerkyacharian, and Picard 1995; Johnstone and Silverman 1997); their objective is to obtain a smooth estimate of the function from the noisy data based on the compressed wavelet coefficients. The data compression should go beyond this step if it is for the purposes of process monitoring and diagnosis. In monitoring and diagnosis applications, various features are extracted from the measured signal and used for monitoring and diagnosis purposes. Under given diagnostic objectives, only certain features (not all signals) are of interest to the users (Jin and Shi in press). Thus, the irrelevant features, even those that belong to the true signals obtained by the approaches of the function estimation, may be removed from the data during data compression to accomplish higher data-compression efficiency. In this case, only the relevant features should be emphasized and maintained after the data compression. The aforementioned concept and technique are known as "feature-preserving" data compression.

The purpose of the proposed data compression, however, is not the same as that of the feature extraction as published in literature (Fukunaga 1990), where the features are extracted based on the training samples of the known faults to achieve the maximum separability among all trained classes. In our study, it is not necessary to know whether the measurement signal is a fault sample at the stage of data compression. Instead of extracting features to represent some specific faults, our efforts are made to maintain all potential fault features in the compressed data by using wavelet coefficients based on engineering knowledge. Thus, the remaining coefficients can be regarded as the "whole" potential feature space for process monitoring and diagnosis. This purpose is different from that of the feature-extraction method, which focuses on the selection of a feature space to maximize the separability of the trained samples.

"Feature-preserving" data compression for a tonnage signal is a very challenging problem due to its inherent nonstationary nature (Fig. 2). Various features of potential process failures are embedded in the signal, as shown in Figure 3. To perform the tonnage signal analysis, these critical features should be maintained after data compression. Examples of these features are shown in Figure 2 and include the following:

1. Transient jump edges, which correspond to the process transient time at the different working stages
2. Curvatures of low-frequency components, which indicate the change rate of the tonnage force
3. Peak tonnage, which shows the maximum force when the slide touches the bottom die position
4. Oscillation components, which represent the dynamic performance of a stamping process

Obviously, the data-compression methods that use a parametric model cannot be implemented in tonnage signal analysis due to the complexity and lack of appropriate physical models to describe the stamping process. When a nonparametric approach is considered for tonnage signal data compression, the wavelet analysis technique is superior due to its natural local properties in both time and frequency domains. Those favorable characteristics can be summarized as follows:

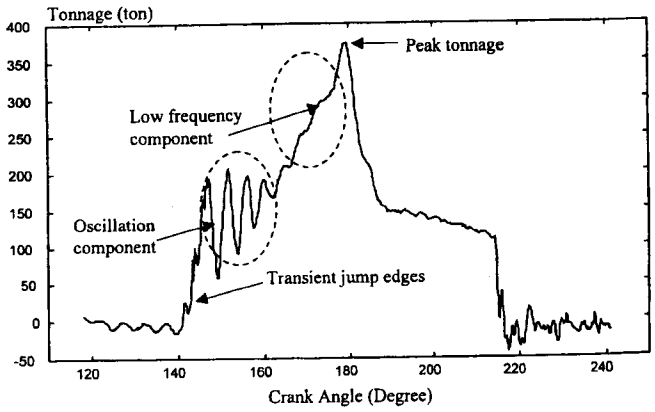

Figure 2. An Example of Tonnage Signals in a Forming Process.

Table 1. Tonnage Signal Segmentation and Fault Features

Segment S_r	Crank angle (data range)	Process failures	Fault signal features	Resolution of interest J_r (s = 3)	P_r %
S_1: slide going down	118.12°–141.88° [1, 198]	Gib chatter due to the excessive clearance	Dominant characteristic frequency component	J = 5, 6	5
S_2: blank holding & draw-bid forming	142.00°–162.52° [199, 370]	Excessive dynamic interaction between the press and the nitrogen cushion system	1. Transient rising edge dependent on the shut height and material thickness 2. Dominant frequency components dependent on the process dynamic characteristics 3. The mean of the signal segment dependent on the nitrogen cushion pressure	J = 3, 4, 5, 6, 7	10
S_3: part drawing	162.64°–176.32° [371, 485]	Large variation of the blank material	Signal slope and curvature	J = 3, 4, 5, 6	5
S_4: slide close to the press bottom	176.44°–182.32° [486, 535]	Loose tie rod, worn bearing, shut height change, die worn out	Profile shape and peak value	J = 3, 4, 5, 6	2
S_5: slide leaving away from the lower die	182.44°–188.32° [536, 585]	Larger variation of the blank material (spring back)	Signal slope and curvature	J = 3, 4, 5	5
S_6: slide only contacting with the lower binder	188.44°–213.52° [586, 795]	Change of the nitrogen cushion pressure	The mean of the signal segment dependent on the nitrogen cushion pressure	J = 3, 4	5
S_7: slide leaving away from the lower binder	213.64°–221.92° [796, 865]	Die bounce generated at the time of nitrogen cushion released	1. The transient dropping edge sensitive to the shut height and material thickness 2. The area of the excessive snap of the negative force sensitive to the die bounce	J = 3, 4, 5, 6, 7	10
S_8: slide continuously going up	222.04°–241.00° [866, 1024]	Gib chatter	Dominant characteristic frequency component	J = 5, 6,	5

1. The *orthogonality* of wavelets ensures the data-compression capability by expressing a function with a few coefficients and shrinks noises by thresholding wavelet coefficients.

2. The *compactness* of wavelets in the time domain ensures the detection of the transient jump edges.

3. The *multiresolution analysis* capability of wavelets decomposes a signal into different resolution levels, which can efficiently filter out the irrelevant signal from the relevant frequency bands.

Data compression using wavelets, called "data shrinkage," has been well developed since the significant achievements in denoising applications by Donoho and Johnstone (1994). A thresholding strategy was proposed for the data compression by setting an appropriate threshold of wavelet coefficients to filter out noises, thus reducing the number of nonzero data points in a signal. Much research has been done to determine an appropriate threshold. Donoho and Johnstone (1994) and Donoho et al. (1995) proposed a universal threshold that is incorporated into their VisuShrink procedure to get a smooth estimate of the underlying function of the signal. To estimate an unknown smoothness function, an adaptive threshold chooser, called SureShrink, was proposed later by Donoho and Johnstone (1995), based on Stein's unbiased risk estimation. The SureShrink chooser specifies a threshold value for each resolution level in a wavelet transform. These thresholds treat only signals with white noise. Further research has been done in dealing with correlated noise (Johnstone and Silverman 1997). Recently, many other thresholding rules have also been developed with different objectives, such as prediction error, false discovery rate, and the Bayesian perspective. To minimize prediction error, the cross-validation thresholding method was proposed and studied by Nason (1995, 1996) and Weyrich and Warhola (1995). The multiple-hypothesis testing procedure has been proposed to obtain an adaptive thresholding of wavelet coefficients under a satisfying false discovery rate (Benjamini and Hochberg 1995; Abramovich and Benjamini 1995, 1996). In terms of minimizing the predefined loss function, a point estimate of the wavelet coefficients can be obtained by using the Bayesian approach. It can be used to develop a statistical model for wavelet coefficients without assuming that the wavelet coefficients of

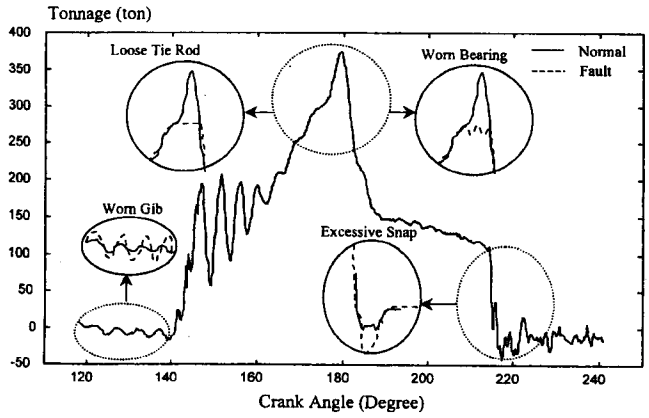

Figure 3. Examples of Fault Tonnage Signals.

an unknown signal are independent (Vannucci and Corradi 1997).

Even though much research has been done in data compression for image and audio signals, little research has been done in applying wavelets to data compression for process monitoring and diagnostic applications. In this article, the concept of feature-preserving data compression is developed for stamping process monitoring and diagnostic applications. As shown in Figure 4, instead of stopping at the denoising step, data are further compressed based on the frequency band of interest and the satisfactory approximation accuracy. Thus, the feature-preserving concept contains two essential aspects: (1) The signal features related to the purposes of process monitoring and diagnosis should be maintained—that is, feature preserving after data compression—and (2) other signal information beyond the interests of potential usage of signals should be removed to accomplish higher data-compression efficiency. In this article, tonnage signal data compression is studied to show the concept, general methodology, and results of feature-preserving data compression.

The article is organized as follows: After the introduction, a detailed study of feature-preserving criteria is proposed in Section 2. Those criteria are developed by combining the engineering knowledge of stamping processes and wavelet multiresolution analysis. Two indicators regarding the relevant resolution region and approximation accuracy are defined. Section 3 proposes a general data-compression procedure for stamping tonnage signals. A new segmental thresholding strategy is presented to satisfy the proposed feature-preserving criteria at each segment. A real example, based on the material thickness change, illustrates the feature-preserving data compression procedures and results. Finally, the impact of the data compression on the performance of monitoring and diagnosis is also investigated.

2. FEATURE–PRESERVING DATA COMPRESSION

2.1 Wavelet Basis and Notations

A few notations about wavelets used in this article are briefly introduced here. If $f(x) \in L^2(R)$, then $f(x)$ can be expressed as (Daubechies 1992)

$$f(x) = \sum_{k \in Z} c_{sk}\phi_{sk}(x) + \sum_{j=s,k \in Z}^{\infty} d_{jk}\psi_{jk}(x), \qquad (1)$$

where functions $\phi(x)$ and $\psi(x)$ are the two basic functions, known as a scaling function and a mother wavelet. The coefficient c_{sk} quantifies the basis function $\phi_{sk}(x)$. The d_{jk} quantifies the basis function $\psi_{jk}(x)$. The basis function $\psi_{jk}(x)$ has its center frequency at $2^j\omega_c$, where ω_c is the center frequency of the mother wavelet $\psi(x)$ (Cohen 1992).

When a discrete wavelet transform $\mathbf{W}$ is used for a dataset $\mathbf{Y}$ $(y_1, \ldots, y_N; N = 2^n)$, it can be expressed by

$$\mathbf{C} = \mathbf{WY}, \qquad (2)$$

where $\mathbf{W}$ is the wavelet transform matrix (Mallat 1989) and $\mathbf{C}$ is the wavelet coefficient. If $\mathbf{W}$ is an orthogonal wavelet, the original data $\mathbf{Y}$ can be recovered by the inverse discrete wavelet transform (without considering the boundary filter effect) as

$$\mathbf{Y} = \mathbf{W}^T\mathbf{C}. \qquad (3)$$

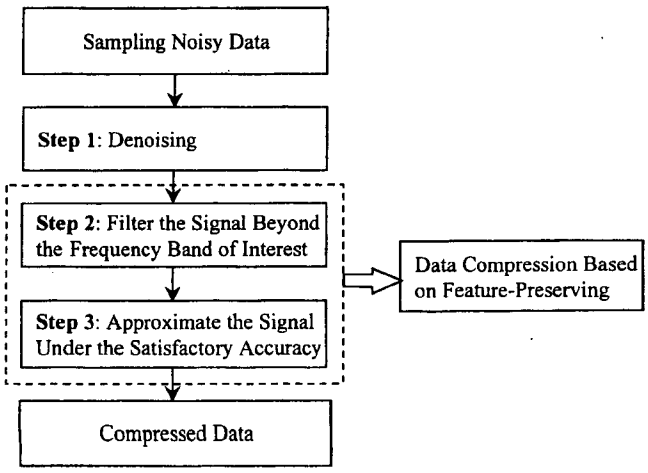

Figure 4. Analysis Procedures of Feature-Preserving Data Compression.

2.2 Data Decomposition

To realize feature-preserving data compression, a three-layer data-decomposition structure is proposed, as shown in Figure 5. At Layer 1, the original sampling data $\mathbf{Y}$ is decomposed into two portions, the signal $\mathbf{F}$ and the noise $\mathbf{Z}$, which is the same as the data compression used for the smooth estimate of the underlying function approach (Donoho 1995). At Layers 2 and 3, based on process-engineering knowledge, the signal portion is further decomposed according to the relevant frequency band of a signal and a satisfactory accuracy requirement. These two layers of data decomposition form an important data structure for the following feature-preserving data compression. The detailed discussion of the data decomposition is given as follows:

At Layer 1, the original sampling data $\mathbf{Y}$ can be expressed by

$$\mathbf{Y} = \mathbf{F} + \mathbf{Z}, \qquad (4)$$

where $\mathbf{F}$ is a signal vector of samples $f(t_i), i = 1, \ldots, N$ with $N = 2^n$. $\mathbf{Z}$ is a noise vector drawn from a white-noise process $z_i, i = 1, \ldots, N$, with $z_i \sim$ iid $\mathrm{N}(0, \sigma^2)$.

At Layer 2, the signal frequency band of interest can be predetermined using process-engineering knowledge. The signal beyond the frequency band of interest should not be kept in the final compressed signal. In terms of the signal-frequency band of interest, the total signal $\mathbf{F}$ can be decomposed into two components:

$$\mathbf{F} = \mathbf{F}^I + \mathbf{F}^O, \qquad (5)$$

where $\mathbf{F}^I$ is the signal within the frequency/resolution band of interest and $\mathbf{F}^O$ is the signal out of the frequency/resolution band of interest. All frequency bands of

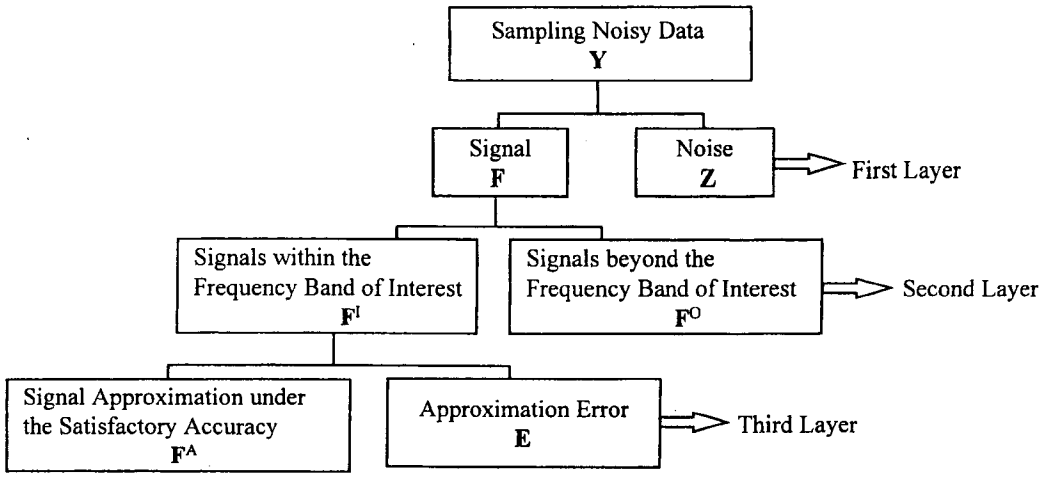

Figure 5. Data Decomposition for Feature-Preserving Data Compression.

interest form an interested frequency set, denoted as **J**. So, for any $j \in \mathbf{J}$, the decomposed signal with the frequency $2^j\omega_c$ or resolution 2^{-j} is within our interest.

At Layer 3, considering a satisfactory approximation accuracy, the signal $\mathbf{F}^I$ can be further expressed as

$$\mathbf{F}^I = \mathbf{F}^A + \mathbf{E}, \tag{6}$$

where $\mathbf{F}^A$ is the final compressed/approximated signal vector and **E** is the approximation error vector.

An index, P, is used to describe the satisfactory accuracy requirement, which is defined as

$$P = \frac{\|\mathbf{F}^I - \mathbf{F}^A\|^2}{\|\mathbf{F}^I\|^2} = \frac{\|\mathbf{E}\|^2}{\|\mathbf{F}^I\|^2} \times 100\%, \tag{7}$$

where $\|\cdot\|^2 = \sum_i [\cdot_i]^2$. The approximation error is relative to the signal $\mathbf{F}^I$ within the interested frequency set **J**.

2.3 Threshold Selection for Data Compression

To accomplish those three steps of the data compression shown in Figure 5, three different thresholds need to be selected.

2.3.1 Threshold Selection for Denoising. Following the general procedures of data compression, the first step is to conduct the discrete wavelet transform of noisy data:

$$\mathbf{WY} = \mathbf{WF} + \mathbf{WZ}. \tag{8}$$

The corresponding wavelet coefficients can be denoted by

$$\mathbf{C} = \mathbf{C}^F + \mathbf{C}^Z, \tag{9}$$

where **W** is a discrete wavelet transform, and $\mathbf{C}, \mathbf{C}^F$, and $\mathbf{C}^Z$ are the wavelet coefficients respective to the original noisy signal **Y**, the signal **F**, and the noise **Z**.

The second step is thresholding of wavelet coefficients. The threshold can be determined by using general denoising shrinkage methods. Because a white noise is being considered here, a universal threshold is used (Donoho and Johnstone 1994) as

$$\lambda^Z = \hat{\sigma}\sqrt{2\log N}, \tag{10}$$

where λ^Z is the threshold used for shrinking the noise coefficient $\mathbf{C}^Z$ and $\hat{\sigma}$ is an estimate of the noise level, which is determined by a scaled median absolute deviation of the empirical wavelet coefficients.

The shrinkage of wavelet coefficients is based on a hard thresholding rule, which is defined by

$$T^Z = \begin{cases} c_{jk}, & \text{if } |c_{jk}| > \lambda^Z \\ 0, & \text{if } |c_{jk}| \le \lambda^Z \end{cases}, \tag{11}$$

where T^Z is a hard thresholding rule for filtering out the noise **Z** based on the absolute value of the wavelet coefficient c_{jk}. If the smoothness of the estimated function is a critical concern in the application, some other thresholding methods, such as soft thresholding (Donoho 1995), garrot thresholding (Breiman 1995; Gao 1997), and Sureshrink (Donoho and Johnstone 1995) can be used.

2.3.2 Threshold Selection in the Frequency Domain. The objective of conducting data thresholding in the frequency domain is to eliminate all irrelevant frequency bands in the signal that are not related to the process faults. Based on the multiresolution analysis of the wavelets (Chui 1992), the signal **F** in Equation (5) can be rewritten as a wavelet decomposition form:

$$\mathbf{F} = \mathbf{V}_s^F + \sum_{j\in J} \mathbf{W}_j^F + \sum_{j\notin J} \mathbf{W}_j^F, \tag{12}$$

where $\mathbf{V}_s^F$ corresponds to the approximation of the signal **F** at the relevant coarsest level s; that is,

$$\mathbf{V}_s^F = \sum_{k\in[0,2^s-1]} \mathbf{c}_{sk}^F \phi_{sk}, \tag{13}$$

and $\mathbf{W}_j^F$ is the decomposition of the signal **F** at the single level j; that is,

$$\mathbf{W}_j^F = \sum_{k\in[0,2^j-1]} \mathbf{c}_{jk}^F \psi_{jk}. \tag{14}$$

Because $\mathbf{F}^O = \sum_{j\notin\mathbf{J}} \mathbf{W}_j^F$ is related to the signal beyond the relevant frequency band, the corresponding wavelet coefficients c_{jk}^F ($j \notin \mathbf{J}$) should be set to 0. This thresholding

rule is expressed as

$$T^O = \begin{cases} c_{jk}, & \text{if } j \in \mathbf{J} \\ 0, & \text{if } j \notin \mathbf{J} \end{cases}, \tag{15}$$

where T^O is the thresholding rule used for the shrinkage of the signal $\mathbf{F}^O$.

2.3.3 Threshold Selection in the Time Domain. The objective of conducting data thresholding in the time domain is to achieve a high data-compression rate with feature preserving by ensuring satisfactory approximation accuracy. Based on Equation (12), after filtering out the signal $\mathbf{F}^O$, the signal $\mathbf{F}^I$ can be expressed as

$$\mathbf{F}^I = \mathbf{V}_s^F + \sum_{j \in J} \mathbf{W}_j^F. \tag{16}$$

Based on Equation (6), it can be obtained:

$$\mathbf{WF}^I = \mathbf{WF}^A + \mathbf{WE}. \tag{17}$$

These corresponding wavelet coefficients can be denoted by

$$\mathbf{C}^I = \mathbf{C}^A + \mathbf{C}^E. \tag{18}$$

Considering a satisfactory approximation accuracy limit, P, for data compression of a signal, the shrinkage of the wavelet coefficients $\mathbf{C}^I$ ($c_{jk}^I, j \in \mathbf{J}$) can be further accomplished by using a hard thresholding rule T^E similar to Equation (11) with the threshold λ^E as follows:

$$\lambda^E = \sup\left\{\lambda \le \left[\frac{P \cdot \|\mathbf{C}^I\|^2}{n_z}\right]^{1/2}\right\} \tag{19}$$

and

$$T^E = \begin{cases} c_{jk}, & \text{if } |c_{jk}| > \lambda^E \\ 0, & \text{if } |c_{jk}| \le \lambda^E \end{cases}, \tag{20}$$

where n_z is the number of coefficients that are set to 0 by the threshold λ in this step. It excludes the initial number of zero coefficients due to the previous shrinkage of $\mathbf{Z}$ and $\mathbf{F}^O$. It can be proven that a satisfactory approximation accuracy limit P can be achieved by using a threshold defined in Equation (19). A short derivation is provided in the Appendix.

Based on Equations (19) and (20), the iterative calculation steps for λ^E are as follows:

1. Sort all coefficients C^I in absolute value from the largest to the smallest.
2. Set $n_z = 1$ and let λ be equal to the last nonzero coefficient.
3. Verify if Equation (21) is satisfied:

$$\lambda \le \left[\frac{P \cdot \|\mathbf{C}^I\|^2}{n_z}\right]^{1/2}. \tag{21}$$

4. If Equation (21) is satisfied in Step 3, set the last nonzero coefficient equal to 0 and go to Step 5; otherwise, stop the iteration and use the λ at the last iteration as λ^E.
5. Increase n_z by 1, and let λ^E be equal to the next nonzero coefficient; then go to Step 3.

Remark. The preceding iterative analysis procedure is similar to that of Ogden and Parzen (1996) in principle—that is, inclusion/exclusion of coefficients until a criterion is satisfied. There are, however, some differences in the thresholding criterion. The method of Ogden and Parzen (1996) uses a hypothesis test criterion. The level of likelihood ratio test is selected to ensure the smoothness of the estimate or the smoothness of the reconstruction of the underlying function or curve. In this article, the objective of data compression is to maintain all signal profile features for future process monitoring and fault diagnosis. Thus, a satisfactory approximation accuracy limit is defined as the criterion in the data compression.

3. DATA COMPRESSION FOR STAMPING TONNAGE SIGNALS

3.1 Tonnage-Signal Segmentation

3.1.1 The Concept of Segmentation. The concept of segmentation is used in this study. The word "blocking" (or "segmentation") has been used in image-analysis research. In the literature, there are two types of "blocks" used in image signal processing. The first type defines a block based on the texture of an image (Froment and Mallat 1992). In this method, the total energy of the wavelet coefficients in the block is compared with a predetermined threshold. The threshold is determined based on the criteria of texture distortion and human visual sensitivity. The second type of block is used in motion compensation (Rao and Bopardikar 1998). In this method, the motion compensation is performed on a block-by-block basis by dividing the images into blocks of equal size or segmenting based on criteria (Calzone, Chuang, and Divakaran 1997).

In this article, a different segmentation strategy is proposed for stamping process monitoring and diagnosis. In this segmentation strategy, the number of segments and the boundary partition of each segment are determined by stamping-engineering knowledge. The purpose of the segmentation is to partition the data into various segments so that each segment corresponds to fewer process faults. Based on the segmentation results, segmental thresholds are defined to achieve the feature-preserving objective—that is, to maintain all potential features for process monitoring and diagnosis. Detailed discussion is presented in the following subsections.

3.1.2 Segmentation of Tonnage Signals. The developed feature-preserving data-compression technique is used in tonnage signal study. As shown in Figure 2, the curvature changes of a tonnage signal are closely related to different working stages of a stamping process, which is usually referenced by a crank angle. To simplify the expression for the later wavelet analysis, however, the crank angle indices of a tonnage signal are now changed into sampling points. Generally, a crank-angle trigger is used in a data-acquisition system. The real crank angle can be easily recovered through the transformation of $\theta_0 + i^*\Delta\theta$, where θ_0 is the starting trigger position of the crank angle, i is the sampling point index, and $\Delta\theta$ is the sampling interval.

Because most process faults are associated with only some specific working stages, the partition of a tonnage signal into nearly disjointed segments can explicitly simplify the fault diagnostic problems (Koh et al. 1996). The boundaries of the partition are determined by the engineering knowledge of a stamping process, which defines the interactions between the press, the die, and the material (blank). In general, a total tonnage can be divided into eight segments as shown in Figure 6. The boundary of each segment corresponds to the change in a forming cycle. A complete stamping cycle can be described as follows: (1) The upper die (i.e., the outer binder and the inner punch) goes down from the top position until it touches the blank at data index 198. (2) The outer binder interacts with the lower die and generates binding force from index 199 to 370. (3) Then, the inner punch touches the blank and starts deep-drawing at index 371. (4) The deep draw continues, and the inner punch reaches the peak tonnage zone from index 486 to 535. (5) After that, the inner punch departs from the lower die and finally separates with the lower die at index 585. (6) After index 586, only the outer binder is in contact with the lower die. (7) The outer binder goes up at index 796, and the die cushion bounces back at the same time. The cushion's effects will disappear after index 865. And finally, (8) the upper die goes up to the initial position. The preceding eight steps correspond to a cycle of stamping one part. Each step corresponds to one segment in the tonnage-signal segmentation. Thus, the boundary and the size of each segment are determined purely by stamping-process knowledge. Meanwhile, the potential faults occurring in each segment can also be analyzed based on engineering knowledge because only certain components of the press, die, and blank interact in each segment. As an example, the gib-chatter fault in a press may occur only when the upper die goes down or up—that is, in segment S_1 or S_8. Similarly, a loose tie rod may be observed only when the maximum tonnage is required in the peak value—that is, segment S_4. Based on this knowledge, the potential faults for each segment can be defined.

Table 1 provides the summary of the process working stage, the potential faults, and their associated fault features at each segment S_r $(r = 1, \ldots, 8)$. In the table, $\theta_0 = 118°, i = 1, \ldots, N, N = 2^n = 1,024, n = 10$, and $\Delta\theta = .12°$. Due to the variability of the working stages, the frequency/resolution bands of interest and the satisfactory approximation accuracy P need to be defined individually at each segment and are denoted as $\mathbf{J}_r$ and P_r, respectively. The modified definition of P_r is given in Section 3.3, and the values of P_r and $\mathbf{J}_r$ are given in Table 1. In general, P_r is specified based on the magnitude range of each potential fault and the process noise variance in each segment, and $\mathbf{J}_r$ is specified based on the specific frequency ranges when potential faults occur in the segment. In practice, the frequency ranges corresponding to specific component failures are available from engineering design or can be analyzed for a given stamping system (ABC 1996).

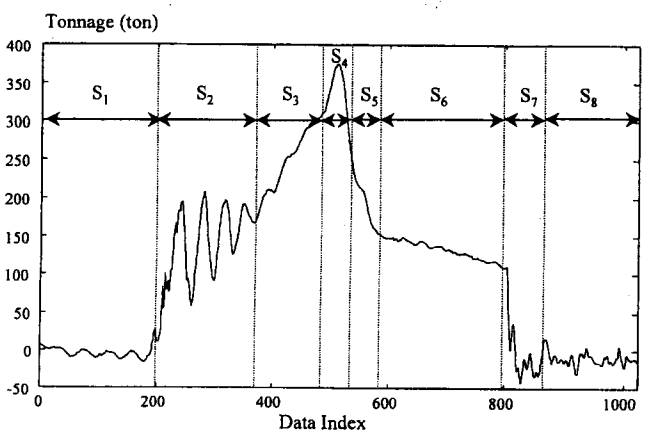

Figure 6. Segmentation of a Tonnage Signal.

3.2 Wavelet Decomposition of Tonnage Signals

Based on the characteristics of tonnage signals, Daubechies wavelet DB3 is selected as the basis of the wavelet transform. The selection of DB3 in the analysis is based on the following two major considerations.

The first consideration is that energy loss has been used as one of the criteria in the thresholding, as defined in Equation (7). In the stamping signal, the integral of the tonnage-signal square is considered to be the forming energy. Thus, a relationship between the tonnage signal and wavelet coefficients should be developed. According to Parseval's theorem, the norm of the wavelet coefficients is the same as the norm of the function being spanned (Burrus, Gopinath, and Guo 1998). Thus, the energy loss due to the shrinkage of the wavelet coefficients is described by the sum of squares of shrinkage wavelet coefficients. Based on this, the threshold can be simply determined easily using a given satisfactory accuracy requirement as defined in Equation (19). As a result, the major reason to choose the Daubechies orthogonal wavelet is because Parseval's theorem holds whereas for biorthogonal wavelets, it does not (Burrus et al. 1998).

The second consideration is the smoothness or the number of vanishing moments required in the stamping signal. For Daubechies wavelets DBp with the support length $2p - 1$, the number of vanishing moments of ψ equals p, and the regularity increases with p (Misiti, Misiti, Oppenheim, and Poggi 1996). In our application, the changes of stamping tonnage signals due to process faults are reflected mostly on the waveform profile change, which can be seen from examples such as loose tie rod and excessive snap shown in Figure 3, and material thickness change shown in Figure 8, Subsection 3.4.1. In general, those waveform profile changes can be described by a second-order polynomial approximation, which supports the selection of DB3 in our application.

Using DB3, Figure 7 shows the wavelet decomposition results at each level, where the approximation signals $\mathbf{V}_j$ (j is the decomposed level, $j = 3, 4, \ldots, 9, s = 3$) are shown on the left panels and the related decomposed detail signal $\mathbf{W}_j$ is shown on the right panels. Based on the multiresolution analysis (Chui 1992), it can be seen that

$$\mathbf{V}_{j+1} = \mathbf{V}_j + \mathbf{W}_j \tag{22}$$

and

$$\mathbf{Y} = \mathbf{V}_j + \sum_{k=j}^{n-1} \mathbf{W}_k, \tag{23}$$

where $j = s, \ldots, n-1$ $(s = 3, n = 10)$.

3.3 Segmental Thresholding

In different signal segments S_r $(r = 1, \ldots, 8)$, there are different relevant frequency sets $\mathbf{J}_r$ and satisfactory accuracy requirements P_r, as shown in Table 1. Thus, the segmental thresholds need to be used at each segment for data compression in Layers 2 and 3, as shown in Figure 5. For this purpose, the threshold selection discussed in Section 2.3 must be modified. Before the detailed discussion of the modified thresholding procedures, two important terms will be introduced, "overlap coefficients" and "nonoverlap coefficients."

Overlap coefficients are coefficients corresponding to wavelets that overlap the boundary between segments. The calculation of these wavelet coefficients is related to multiple segment data.

Nonoverlap coefficients are coefficients corresponding to wavelets that are included within one segment. The calculation of these wavelet coefficients is related only to one segment's data.

3.3.1 Identifying Wavelet Coefficients at Each Segment. Because the wavelet transform is conducted for the total tonnage signal with all segments, this subsection discusses how to identify the wavelet coefficients for thresholding in a given segment. Those wavelet coefficients include nonoverlap coefficients but may, or may not, include overlap coefficients. Consequently, in terms of whether to include the overlap coefficients in a given segment, two definitions are given as follows:

1. *Type 1 coefficients:* In a given data segment, Type 1 coefficients include only the nonoverlap coefficients.
2. *Type 2 coefficients:* In a given data segment, Type 2 coefficients include both the nonoverlap coefficients and the overlap coefficients.

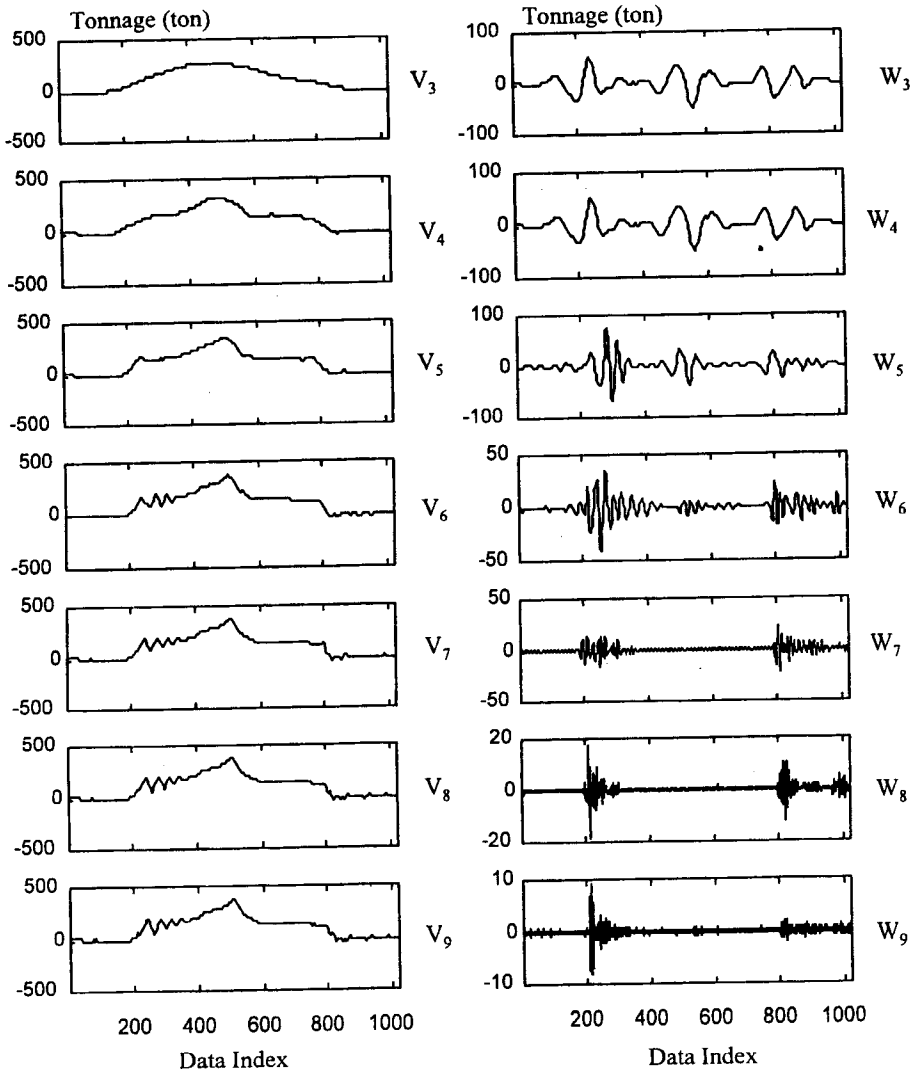

Figure 7. Wavelet Decomposition of a Tonnage Signal.

Based on the preceding definitions, these two types of oefficients can be selected from all wavelet coefficients sing the following algorithms:

Type 1 coefficients selection. For a given data segment $_r$ with data points $y_i(i \in [a_r, b_r]), a_r$ and b_r are the start-ng data index and the ending data index of this segment. he starting wavelet coefficient index ca_r^{j-1} and the end-ng wavelet coefficient index cb_r^{j-1} at the decomposed level -1 of segment S_r can be obtained by using the respective ndices at level j:

$$\text{if} \quad cb_r^j - m > ca_r^j \quad \text{then}$$

$$ca_r^{j-1} = \text{int}\left(\frac{ca_r^j + 1}{2}\right) \quad \text{and}$$

$$cb_r^{j-1} = \text{int}\left(\frac{cb_r^j - m}{2}\right); \tag{24}$$

$$\text{otherwise } ca_r^{j-1} = cb_r^{j-1} = 0. \tag{25}$$

At the finest level—that is, $j = n = 10$ for $N = 1{,}024$ ampling data—ca_r^n and cb_r^n are equal to the starting data ndex a_r and the ending data index b_r of segment S_r. For Daubechies wavelets DBp, the support length of ψ is $m = 2p - 1$ (Misiti et al. 1996).

Proof. The wavelet coefficients at level $j-1, C_{(j-1)k}$ $k = 1, \ldots, N_{j-1}; N_{j-1}$ is the number of wavelet coeffi-cients at level $j-1$) are obtained through the highpass filter Strang and Nguyen 1996) on the coefficients C_{jk} at level . The decimation principle is that only half of the high-pass filter outputs corresponding to the odd index at level are kept to form coefficients at level $j-1$. Based on this principle, the number of wavelet coefficients at level $j-1$ N_{j-1}) can be obtained as

$$N_{j-1} = \text{int}\left(\frac{N_j - m - 1}{2}\right) + m. \tag{26}$$

Here $j = 1, \ldots, n$ ($n = 10$), and N_n equals the length of he original data series ($N_n = N = 1{,}024$).

The starting coefficient index ca_r^{j-1} of Type 1 contribu-ion coefficients at level $j-1$ in segment S_r can be de-ermined by the starting coefficient index ca_r^j at level j. f ca_r^j is an odd index, the highpass filter output at this dd index is kept according to the preceding decimation principle. Because only half outputs at the odd indices are kept, the starting index ca_r^{j-1} at level $j-1$ is $ca_r^{j-1} = (ca_r^j + 1)/2$; otherwise, if the starting coefficient index ca_r^j is an even index, the highpass filter output at this even index is gnored. The actual starting index at level $j-1$ is shifted to he next point; thus, $ca_r^{j-1} = ca_r^j/2 + 1$. In both cases, one has

$$ca_r^{j-1} = \text{int}\left(\frac{ca_r^j + 1}{2}\right), \tag{27}$$

where the int operation obtains the nearest integer and nt(.5) = 1.

For the ending coefficient index at segment S_r, because Type 1 contribution coefficients do not include overlap co-efficients, the actual ending index cb_r^j should be shifted back m points at level j. Otherwise, the ending index will contribute to the excessive region beyond cb_r^j. Similar to the preceding discussion on the starting coefficient index, the ending coefficient index can be obtained as follows based on the decimation principle:

$$cb_r^{j-1} = \text{int}\left(\frac{cb_r^j - m}{2}\right) \tag{28}$$

If $cb_r^j - m < ca_j^r$, it means that the ending index is located before the starting index, and thus no relevant Type 1 contribution coefficient is included; that is, $ca_r^{j-1} = cb_r^{j-1} = 0$.

Remark. Type 1 coefficients are used in the segmental thresholding of coefficients $\mathbf{C}^I$ under the satisfactory accuracy limit P_r at segment S_r. The approximation error $\mathbf{E}_r$, which is defined at segment S_r, is caused by shrinkage of the nonoverlap coefficients in only one segment. As a result, it guarantees the satisfactory approximation accuracy that will be defined in Equation (31).

Type 2 coefficients selection. For a given data segment S_r including data points $y_i(i \in [a_r, b_r]), a_r$ and b_r are the starting data index and the ending data index of this segment. The starting wavelet coefficient index ca_r^{j-1} and ending wavelet coefficient index cb_r^{j-1} at level $j-1$ can be obtained by using the respective indices at level j,

$$ca_r^{j-1} = \text{int}\left(\frac{ca_r^j - m + 1}{2}\right) \quad \text{if} \quad ca_r^j - m \geq 1$$

$$= 1 \quad \text{if} \quad ca_r^j - m < 1 \tag{29}$$

and

$$cb_r^{j-1} = \text{int}\left(\frac{cb_r^j}{2}\right) \quad \text{if} \quad cb_r^j + m \leq N_j + 1$$

$$= N_{j-1} \quad \text{if} \quad cb_r^j + m > N_j + 1. \tag{30}$$

Because Type 2 contribution coefficients may overlap more than one signal segment, the actual starting index must be shifted back m points at level j unless it reaches the first point. Therefore, similar to the discussion on the starting index of Type 1 contribution coefficients, Equation (29) holds. For the ending coefficient index of Type 2 contribution coefficients, there is no need to shift back m points at level j. The ending index cb_r^{j-1} is simply equal to half of cb_r^j unless $cb_r^j + m$ is beyond the ending point N_j. In that case, N_{j-1} is used as the ending index instead, as in Equation (30).

Remark. Type 2 coefficients in the segmental thresholding are used to filter out signals beyond the relevant frequency set $\mathbf{J}_r$ at segment S_r. Because the overlap coefficients are also involved, they are thresholded here to achieve more efficient data compression. Because the overlap coefficients are compressed based on different thresholds at different segments, however, a conservative coordination between the segmental thresholding is required to ensure the

feature-preserving requirement. It is suggested that the coefficients can be set to 0 if and only if they have been set to 0 in all segments. The detailed thresholding procedures will be provided in the next subsection.

3.3.2 Coordinate the Segmental Thresholds in the Frequency Domain. As mentioned previously, there are several threshold rules applied on the overlap coefficients, which can lead to different thresholding results. Thus, coordination is required for these overlap coefficients. The basic idea in the coordination is to keep as few coefficients as possible under the constraint of feature preserving. In other words, a coefficient cannot be set to 0 at this stage unless it is set to 0 in all segments. The detailed procedures include the following:

1. For segment S_r $(r = 1, \ldots, 8)$, determine the coefficient indices of Type 2 contribution coefficients.
2. For each coefficient of $\mathbf{C}^F$, identify whether it belongs to the preceding set of selected coefficients. If not, set it to 0; otherwise, further check to see whether it is within the relevant frequency band $\mathbf{J}_r$. If so, keep it unchanged; if not, set it to 0. As a result, when Step 2 is finished, a new compressed coefficient vector $\mathbf{C}_r$ is obtained for each segment S_r. The dimension of $\mathbf{C}_r$ is equal to that of $\mathbf{C}^F$.
3. The final coefficient vector $\mathbf{C}^I$ is obtained by a conservative coordination between the coefficients $\mathbf{C}_r$ at each segment, using $c^I_{jk} = c^F_{jk}$ when $\sum_{r=1}^{8} c_{r(jk)} \neq 0$. Here, c^I_{jk}, c^F_{jk}, and $c_{r(jk)}$ are coefficient elements in vectors $\mathbf{C}^I, \mathbf{C}^F$, and $\mathbf{C}_r$, respectively.

3.3.3 Coordinate the Segmental Thresholds in the Time Domain. Because the satisfactory accuracy limit P_r has to be associated with each segment S_r $(r = 1, \ldots, 8)$, the definition of segment P of Equation (7) is modified as follows:

$$P_r = \frac{\|\mathbf{F}^I_r - \mathbf{F}^A_r\|^2}{\|\mathbf{F}^I_r\|^2} \times 100\%, \tag{31}$$

where r is the segmentation index. Based on this definition, the threshold λ^E of Equation (19) is also modified as follows:

$$\lambda^E_r = \sup\left\{\lambda^E_r \leq \left[\frac{P_r \cdot \|\mathbf{C}^I_r\|^2}{n_r}\right]^{1/2}\right\}, \tag{32}$$

where each parameter is limited to segment S_r instead of the whole data range. The detailed iteration steps are similar to the steps given in Section 2.3.3. The only difference is that the data are separately processed for each segment rather than for the whole data range at one time.

3.4 An Example of Stamping Tonnage-Signal Compression

3.4.1 Feature-Preserving Data-Compression Results. As shown in Figure 8, two sets of tonnage signals with different material thickness are used to validate the feature-preserving criteria in the data compression of tonnage signals. The solid curve represents the tonnage signal using the blank with 10% greater than the normal thickness,

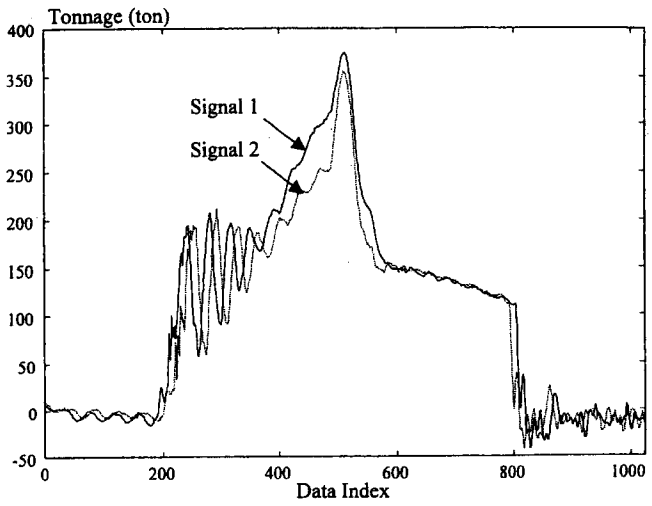

Figure 8. Tonnage Signals With Different Material Thickness (1,024 Data Points).

and the dashed curve is related to the thinner blank with 10% less than the normal thickness. The curvature difference of these two tonnage signals is clearly shown in Segment 3.

As the first step (the first layer shown in Fig. 5), the denoising threshold of wavelet coefficients is used. The data-compression results are shown in Figure 9. It can be seen that the number of nonzero wavelet coefficients is reduced from the original 1,024 data points to 621/640 respective to the thicker/thinner material thickness. In this step, some irrelevant signals with the higher resolutions beyond our relevant frequency band are still kept in the compressed signals. To achieve more efficient data compression, these data are further compressed by using the proposed segmental thresholding strategy. After removing irrelevant frequency signals (the second layer in Fig. 5), the remaining nonzero coefficients are dramatically reduced from 621/640 to 82/95, corresponding to the thicker/thinner material thickness. Further data compression is conducted under the

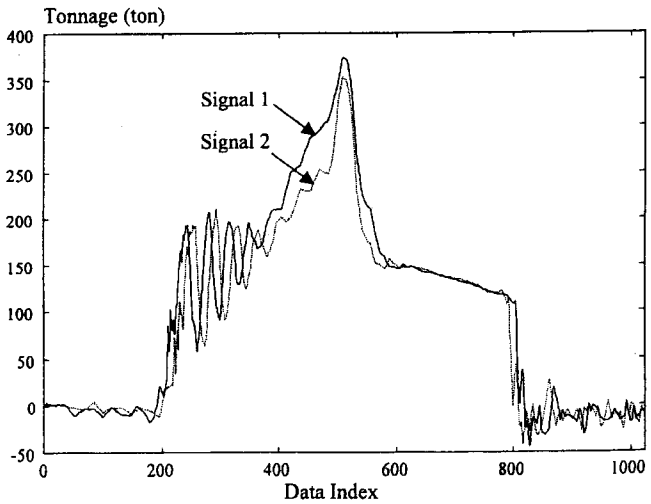

Figure 9. Denoised Tonnage Signals (the number of nonzero wavelet coefficients of signal 1 is 621; the number of nonzero wavelet coefficients of signal 2 is 640).

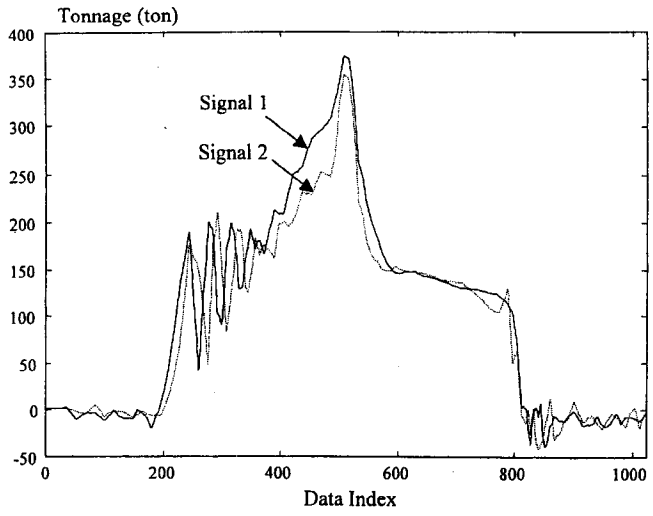

Figure 10. Compressed Tonnage Signals Using the Feature-Preserving Approach (the number of nonzero wavelet coefficients of signal 1 is 66; the number of nonzero wavelet coefficients of signal 2 is 73).

satisfactory accuracy index (the third layer in Fig. 5). As a result, the nonzero coefficients are reduced to 66/73. The final compressed results are shown in Figure 10. From this figure, it can be seen that the irrelevant signals are efficiently removed while the fault features are kept very well. This case study indicates that data compression using feature-preserving criteria is much more efficient than conventional data compression. About 92% data reduction has been accomplished using the presented method, as compared to 38% data reduction using denoising data compression.

The effectiveness of the developed feature-preserving data compression can be further demonstrated with the fault-detection power in process monitoring and fault diagnosis. Some discussion on this topic is presented in the next subsection.

3.4.2 Assessment of Fault-Detection Performance. Feature-preserving data compression may have an impact on the performance of process fault detection. As an example, the material-thickness change is reflected by a mean shift of the wavelet coefficient $c_{3(6,25)}$ at the decomposition level 6 (a nonoverlap coefficient in segment S_3; i.e., $r = 3, j = 6, k = 25$). Because the detection power is associated with a specific detection algorithm, an X-bar chart is used to monitor the mean shift of the coefficient $c_{3(6,25)}$ in the article.

It is assumed that a mean shift μ_1 is generated due to the material thickness change. The detection power D respective to this mean shift can be calculated by (Montgomery 1996)

$$D = 1 - \Phi(L - K\sqrt{n_0}) + \Phi(-L - K\sqrt{n_0}), \quad (33)$$

where $\Phi(x)$ is the cumulative probability function of a standard normal distribution. L is associated with the control limits under a given Type I error α, n_0 is the sample size of the detected subgroup, and K represents the mean shift μ_1 relative to the process standard deviation σ ($K = \mu_1/\sigma$).

After data compression, some coefficients will be set to 0 if their absolute values of the monitored coefficients are less than the threshold λ. If each wavelet coefficient before data compression follows a distribution with the probability density function $g(x)$, after data compression, the estimate of mean value can be obtained by

$$\mu_1 + \Delta\mu = \int_{-\infty}^{-\lambda} xg(x)\,dx + \int_{-\lambda}^{\lambda} x \cdot 0 \cdot dx + \int_{\lambda}^{\infty} xg(x)\,dx, \quad (34)$$

where μ_1 is the original mean shift before data compression, $\Delta\mu$ is the additional mean shift due to the data compression, λ is the threshold, and here $\lambda = \lambda_3^E$. The magnitude of $\Delta\mu$ will affect the power of the process change detection. From Equation (34), $\Delta\mu$ can be calculated by

$$\Delta\mu = -\int_{-\lambda}^{\lambda} xg(x)\,dx. \quad (35)$$

If assuming that $g(x)$ is for a normal distribution, $\Delta\mu$ can be obtained by solving Equation (35):

$$\Delta\mu = \frac{\sigma}{\sqrt{2\pi}}\left(e^{-(\lambda-\mu_1)^2/2\sigma^2} - e^{-(\lambda+\mu_1)^2/2\sigma^2}\right] - \mu_1 \cdot \left[\Phi\left(\frac{\lambda-\mu_1}{\sigma}\right) - \Phi\left(\frac{-\lambda-\mu_1}{\sigma}\right)\right]. \quad (36)$$

Similar to the definition of K, a relative mean shift can be defined as $\Delta K = \Delta\mu/\sigma$; that is,

$$\Delta K = \frac{1}{\sqrt{2\pi}}\left[e^{-(\lambda-\mu_1)^2/2\sigma^2} - e^{-(\lambda+\mu_1)^2/2\sigma^2}\right] - K \cdot \left[\Phi\left(\frac{\lambda-\mu_1}{\sigma}\right) - \Phi\left(\frac{-\lambda-\mu_1}{\sigma}\right)\right]. \quad (37)$$

The effect of ΔK on the detection power can be analyzed by differentiating Equation (33):

$$\Delta D = \sqrt{n_0}(g(L - K\sqrt{n_0}) - g(-L - K\sqrt{n_0})) \cdot \Delta K. \quad (38)$$

Equations (37) and (38) provide a quantitative relationship between the threshold value and the detection power loss due to the data compression. Thus, if threshold λ is selected in the data compression, the process detection power will be reduced from its original D (in the absence of data compression) to $D + \Delta D$ ($\Delta D < 0$) (with data compression). In our example, $L = 3$ and $n_0 = 5$ are used. The detection power of the mean shift due to the material thickness change as shown in Figure 8 is $D = .9356$ before data compression. After data compression using the proposed method, the loss of the detection power is equal to $\Delta D = -.0085$. Thus, the relative loss of the detection power is .91%, which is acceptable in real applications.

ACKNOWLEDGMENTS

This research was partially supported by the NSF CAREER award DMI 9624402. We also gratefully acknowledge the referees for their many insightful comments.

APPENDIX: THRESHOLDING UNDER SATISFACTORY APPROXIMATION ACCURACY CRITERIA

Because Parseval's theorem holds for the orthogonal wavelets (Chui 1992), the approximation losses/errors can be obtained by using Equations (6) and (18):

$$\|\mathbf{F}^I - \mathbf{F}^A\|^2 = \|\mathbf{C}^I - \mathbf{C}^A\|^2. \tag{A.1}$$

The approximation at the coarsest level $\mathbf{V}_s^F$ in Equation (16) is kept unchanged during the data compression (Donoho and Johnstone 1994). Thus, the approximation loss is considered only on the detailed components. Thus, Equation (A.1) can be written as

$$\begin{aligned}\|\mathbf{F}^I - \mathbf{F}^A\|^2 &= \sum_{j\in\mathbf{J},k\in[0,2^j-1]} (c_{jk}^I - c_{jk}^A)^2 \\ &= \sum_{j\in\mathbf{J},k\in[0,2^j-1]} (\Delta c_{jk})^2 \end{aligned} \tag{A.2}$$

where c_{jk}^A is the coefficient after thresholding using Equations (19) and (20) and $\Delta c_{jk} = c_{jk}^I - c_{jk}^A$. So, Δc_{jk} can be obtained by

$$\Delta c_{jk} = \begin{cases} c_{jk}^I & \text{if} \quad c_{jk}^I \le \lambda^E \\ 0 & \text{if} \quad c_{jk}^I > \lambda^E \end{cases}. \tag{A.3}$$

From Equations (A.2) and (A.3), it can be seen that

$$\|\mathbf{F}^I - \mathbf{F}^A\|^2 = \|\mathbf{C}^I - \mathbf{C}^A\|^2 \le n_z \cdot (\lambda^E)^2. \tag{A.4}$$

Because $\|\mathbf{F}^I\|^2 = \|\mathbf{C}^I\|^2$, Equation (A.4) can be rewritten as

$$\frac{\|\mathbf{F}^I - \mathbf{F}^A\|^2}{\|\mathbf{F}^I\|^2} = \frac{\|\mathbf{C}^I - \mathbf{C}^A\|^2}{\|\mathbf{C}^I\|^2} \le \frac{n_z \cdot (\lambda^E)^2}{\|\mathbf{C}^I\|^2}. \tag{A.5}$$

By substituting Equation (19) into Equation (A.5), a satisfactory approximation accuracy limit can be achieved:

$$\frac{\|\mathbf{F}^I - \mathbf{F}^A\|^2}{\|\mathbf{F}^I\|^2} \le P. \tag{A.6}$$

[Received November 1997. Revised June 1999.]

REFERENCES

ABC (1996), "Press Components' Characteristic Frequencies," Technical Report, NZS-102, National Institute of Standards and Technology, Advanced Manufacturing Program.

Abramovich, F., and Benjamini, Y. (1995), "Thresholding of Wavelet Coefficients as Multiple Hypotheses Testing Procedure," in *Wavelets and Statistics, Lecture Notes in Statistics*, eds. A. Antoniadis and G. Oppenheim, New York: Springer-Verlag, pp. 103, 5–14.

——— (1996), "Adaptive Thresholding of Wavelet Coefficients," *Computational Statistics & Data Analysis*, 22, 351–361.

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Ser. B, 57, 289–300.

Breiman, L. (1995), "Better Subset Regression Using the Nonnegative Garrote," *Technometrics*, 37, 373–384.

Burrus, C., Gopinath, R., and Guo, H. (1998), *Introduction to Wavelets and Wavelet Transforms: A Primer*, Englewood Cliffs, NJ: Prentice-Hall.

Calzone, S., Chuang, C. C., and Divakaran, A. (1997), "Video Compression by Mean-Corrected Motion Compensation of Partial Quadtrees," *IEEE Transactions on Circuits and System for Video*, 7, 86–96.

Chui, C. K. (1992), *An Introduction to Wavelets*, San Diego, CA: Academic Press.

Cohen, A. (1992), "Biorthogonal Wavelets," in *Wavelet: A Tutorial in Theory and Applications*, ed. C. K. Chui, San Diego, CA: Academic Press, pp. 123–152.

Coifman, R. R., and Wickerhauser, M. V. (1994), "Adapted Waveform Analysis as a Tool for Modeling, Feature Extraction, and Denoising," *Optical Engineering*, 33, 2170–2174.

Coifman, R. R., and Yale, F. M. (1992), "Adapted Waveform Analysis and Denoising," in *Progress in Wavelet Analysis and Applications*, eds. Y. Meyer and S. Roques, Proceedings of the International Conference "Wavelets and Applications," Toulouse, France, Gif-sur-Yvette, France: Editions Frontieres, pp. 63–93.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: Society for Industrial and Applied Mathematics.

Donoho, D. L. (1995), "De-Noising by Soft-Thresholding," *IEEE Transactions on Information Theory*, 41, 613–627.

Donoho, D. L., and Johnstone, I. M. (1994), "Ideal Spatial Adaptation via Wavelet Shrinkage," *Biometrika*, 81, 425–455.

——— (1995), "Adapting to Unknown Smoothness via Wavelet Shrinkage," *Journal of the American Statistical Association*, 90, 1200–1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995), "Wavelet Shrinkage: Asymptopia?" *Journal of the Royal Statistical Society*, Ser. B, 57, 301–369.

Froment, J., and Mallat, S. (1992), "Second Generation Compact Image Coding With Wavelets," in *Wavelet: A Tutorial in Theory and Applications*, ed. C. K. Chui, San Diego, CA: Academic Press, pp. 655–679.

Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition*, New York: Academic Press.

Gao, H. Y. (1997), "Wavelet Shrinkage Denoising Using Non-Negative Garrote," *Wavelet Digest*, 6, 3, p. 2, preprints, available electronically (*http://www.wavelet.org/wavelet/digest_06/digest_06.03html*).

Jin, J., and Shi, J. (in press), "Diagnostic Feature Extraction From Stamping Tonnage Signal Based on Design of Experiment," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 122.

Johnstone, I. M., and Silverman, B. W. (1997), "Wavelet Threshold Estimators for Data Correlated Noise," *Journal of the Royal Statistical Society*, Ser. B, 59, 319–351.

Koh, C. K. H., Shi, J., and Black, J. (1996), "Tonnage Signature Attribute Analysis for Stamping Process," *NAMRI/SME Transactions*, 23, 193–198.

Koh, C. K. H., Shi, J., and Williams, W. (1995), "Tonnage Signature Analysis Using the Orthogonal (Harr) Transforms," *NAMRI/SME Transactions*, 23, 229–234.

Lee, J. (1995), "Modern Computer-Aided Maintenance of Manufacturing Equipment and System: Review and Perspective," *Computers & Industrial Engineering*, 28, 793–811.

Mallat, S. (1989), "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Pattern Analysis and Machine Intelligence*, 11, 674–693.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. M. (1996), *Wavelet Toolbox: For Use with MATLAB*, Natick, MA: The MATH Works Inc.

Montgomery, D. (1996), *Introduction to Statistical Quality Control*, New York: Wiley.

Nason, G. P. (1995), "Choice of the Threshold Parameter in Wavelet Function Estimation," in *Wavelets and Statistics* (Lecture Notes in Statistics), eds. A. Antoniadis and G. Oppenheim, New York: Springer-Verlag, pp. 103, 261–280.

——— (1996), "Wavelet Shrinkage Using Cross-Validation," *Journal of the Royal Statistical Society*, Ser. B, 58, 463–479.

Ogden, R. T., and Parzen, E. (1996), "Change-Point Approach to Data Analytic Wavelet Thresholding," *Statistics and Computing*, 6, 93–99.

Rao, R., and Bopardikar, A. (1998), *Wavelet Transforms: Introduction to Theory and Applications*, Reading, MA: Addison-Wesley Longman.

Shi, J., Ni, J., and Lee, J. (1997), "Research Challenges and Opportunities in Remote Diagnosis and System Performance Assessment," in *Proceedings of* 4^{th} *IFAC, IMS '97, Seoul, South Korea*, Oxford, U.K.: Elsevier, pp. 213–218.

Strang, G., and Nguyen, T. (1996), *Wavelets and Filter Banks*, Wellesley, MA: Wellesley-Cambridge Press.

Vannucci, M., and Corradi, F. (1997), "Some Findings on the Covariance Structure of Wavelet Coefficients: Theory and Models in a Bayesian Perspective," *Wavelet Digest*, 6, 2, p. 10, preprints, available electronically (*http://www.wavelet.org/wavelet/digest_06/digest_06.02.html*).

Weyrich, N., and Warhola, G. T. (1995), "De-Noising Using Wavelet and Cross-Validation," in *Approximation Theory, Wavelets and Applications* (NATO ASI Ser. C), ed. S. P. Singh, Dordrecht: Kluwer, pp. 454, 523–532.