A Wavelet-Based Procedure for Process Fault Detection

Emily K. Lada, Jye-Chyi Lu, and James R. Wilson

***Abstract*—To detect faults in a time-dependent process, we apply a discrete wavelet transform (DWT) to several independently replicated data sets generated by that process. The DWT can capture irregular data patterns such as sharp "jumps" better than the Fourier transform and standard statistical procedures without adding much computational complexity. Our wavelet coefficient selection method effectively balances model parsimony against data reconstruction error. The few selected wavelet coefficients serve as the "reduced-size" data set to facilitate an efficient decision-making method in situations with potentially large-volume data sets. We develop a general procedure to detect process faults based on differences between the reduced-size data sets obtained from the nominal (in-control) process and from a new instance of the target process that must be tested for an out-of-control condition. The distribution of the test statistic is constructed first using normal distribution theory and then with a new resampling procedure called "reversed jackknifing" that does not require any restrictive distributional assumptions. A Monte Carlo study demonstrates the effectiveness of these procedures. Our methods successfully detect process faults for quadrupole mass spectrometry samples collected from a rapid thermal chemical vapor deposition process.**

***Index Terms*—Monte Carlo, process fault detection, resampling procedures, reversed jackknifing, wavelets.**

I. Introduction

In MODERN manufacturing systems, various sensors are equipped to collect data for monitoring, controlling, and improving process performance; for examples, see [1] and [2]. Especially for costly products like semiconductors, such data are useful in improving production quality and efficiency as well as in detecting process faults and understanding process problems. One of the difficulties faced by many industries today, however, in fully utilizing the available data is an overabundance of such data. The objective of this paper is to present a case study in which wavelet transforms are used to reduce the size of large-volume, dynamic-trend data sets so that they can be used for (real-time) process fault detection, system behavior modeling, failure prediction, and other problem-solving and decision-making activities.

There are many different types of large-volume data sets in manufacturing processes. In this article, we focus on numerical data collected with various measurement tools, including different kinds of sensors and product-testing devices. Such data sets usually exhibit nonstationary dynamic trends with distinctive patterns (e.g., sharp jumps) caused by specific types of process faults (see the signals given in [3] for examples). Traditional statistical and signal-processing procedures, such as polynomial regression, time-series models, Fourier transforms, or neural networks, can be inappropriate to describe these data sets. As an alternative, wavelet transforms are being used by many researchers. For example, Jin and Shi [4] used tonnage signals to detect faults in a sheet-metal stamping process; and Wang *et al.* [5] used different catalyst recycling rates to diagnose failures in a residual fluid catalytic cracking process. In our study, we analyze quadrupole mass spectrometry (QMS) samples of a rapid thermal chemical vapor deposition (RTCVD) process (see Figs. 1 and 2) to detect significant deviations from the nominal (in-control) process.

Using an expert's knowledge of a particular process, we could derive a "feature-preserving" procedure to extract a particular data pattern and link it to a specific type of process fault; for example, see [4]. However, the purpose of our study is to develop and evaluate a generic "data-reduction" procedure that can handle large volumes of data without requiring an expert's knowledge or data visualization techniques to identify segments of fault signals.

Ideally, the uses of reduced-size data sets will not be limited to detecting specific types of known faults. Other possible uses include data reconstruction, fault classification, failure prediction, and other general purposes of data processing to improve manufacturing quality and efficiency. For example, wavelet transforms or locally focused wavelet neural networks [6] are useful in the following applications: 1) modeling equipment status and control signals that exhibit significant dependencies (correlations) over time and space as well as across equipment functions; and 2) predicting process failures in real time for in situ analysis or in postproduction mode for off-line analysis.

Beyond the generic approach to data reduction and process fault detection that we take in this paper, other techniques for trouble-shooting process problems include Isermann's survey [3] of model-based process fault detection methods and the neural network-based failure prediction procedures developed by Rietman and Beachy [7]. To extend the capabilities of these methods to handle large data sets with nonstationary patterns, one could create a wavelet-based reduced-size data set and feed it into their procedures instead of the original data.

Manuscript received June 19, 2001; revised September 17, 2001. The work of E. K. Lada and J. R. Wilson was supported in part by the National Science Foundation under Grant DMI-9900164. The work of J.-C. Lu was supported in part by the National Science Foundation under Grant DMS-0072960 and Grant EEC-0080315.

E. K. Lada is with the Graduate Program in Operations Research, North Carolina State University, Raleigh, NC 27695-7913 USA (e-mail: eklada@eos.ncsu.edu).

J.-C. Lu is with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 USA (e-mail: jclu@isye.gatech.edu).

J. R. Wilson is with the Department of Industrial Engineering, North Carolina State University, Raleigh, NC 27695-7906 USA (e-mail: jwilson@eos.ncsu.edu).

Publisher Item Identifier S 0894-6507(02)01030-8.

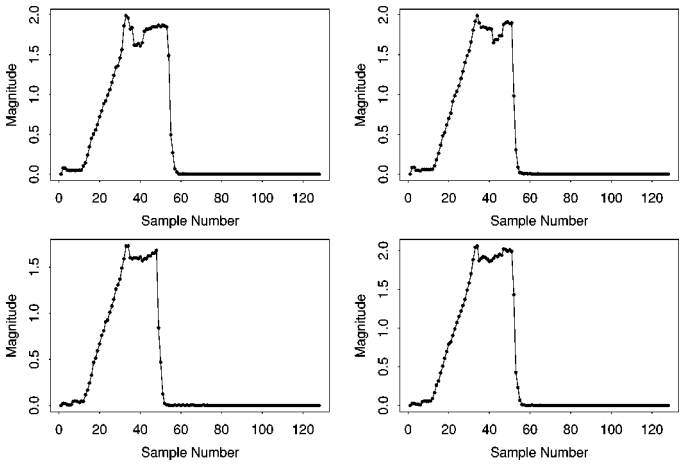

Fig. 1. QMS data for nominal (in-control) runs 1–4 of RTCVD process.

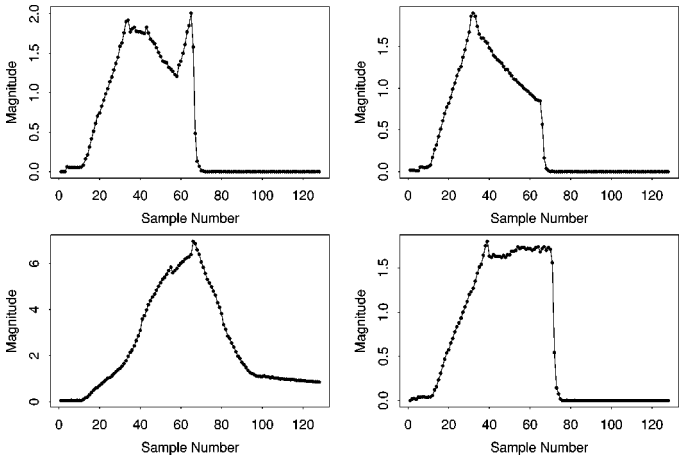

Fig. 2. QMS data for induced-fault runs 1–4 of RTCVD process.

Summary statistics, such as the mean and variance, are commonly used data reduction methods. However, this approach works only for data sampled from a population whose characteristics do not vary over time. For data with dynamic trends, this approach can be extended to the coefficients (and their "bases") of regression functions as well as Fourier and wavelet transforms. Our procedure selects a few important wavelet coefficients that represent key features of the data and discards those fine-scale wavelet coefficients that represent noise or secondary characteristics in the data. Thus, wavelet analysis can be used for both data reduction and data denoising. Because the commonly used wavelet model selection methods—for example, SURE by Donoho and Johnstone [8] and AMDL by Saito [9]—tend to overfit the data (that is, they use too many coefficients for effective data reduction), we propose a new method that balances a data-reduction ratio against a "normalized" version of the data reconstruction error. Based on several real-life case studies, our procedure compares favorably with the SURE and AMDL methods. Our model selection method also goes beyond the traditional procedures that are based on a single data set taken from the target process; thus our proposed techniques can handle multiple replications of time series generated by the target process. This type of study has never been explored previously.

To illustrate the use of reduced-size data sets for process fault detection, in this article, we calculate differences between 1) the wavelet coefficients estimated from several independent runs of the in-control process, and 2) the corresponding wavelet coefficients estimated from a single run of a process that must be tested for an out-of-control condition. We formulate a variant of Hotelling's T^2-statistic for two-sample problems to structure the sum of squares of the differences. The distribution of the T^2-statistic is explored by using a new "reversed jackknife" procedure. A Monte Carlo study characterizes the accuracy of approximating the distribution of the final test statistic with an appropriate F-distribution. This leads to a simple decision-making procedure suitable for detecting process faults. The proposed procedure is then tested on several data sets obtained by inducing faults in the RTCVD process. The results of this case study indicate that the proposed wavelet-based general data-reduction procedure is effective in detecting process faults.

More complicated fault detection procedures could be developed based on the selected wavelet coefficients. For instance, if the process signals are collected at various time points, one could build time-dependent models of the wavelet coefficients and use those models to predict future behavior of the wavelet coefficients. Furthermore, for the purpose of process failure prediction, statistical process control limits could be built to signal the possibility of out-of-control events at future time points. These predictions of wavelet coefficient values could then be used to reconstruct an approximation to the data at selected future time points, thereby eliminating the need to reconstruct an original data set that may be very large. With various cross-functional, reduced-size data sets representing equipment status, control signals, or postproduction measurements, it is possible to predict failure events and identify their causes. Kim and May [2] explore the use of evidential reasoning [10] to build an expert system for diagnosis of equipment failures in the fabrication of integrated circuits. However, the scope of these extensions is beyond the purpose of this article on data reduction and should be addressed in future research.

The remainder of this paper is organized as follows. Section II contains a description of the RTCVD process and the data as well as a brief overview of wavelet basics. Section III presents a detailed fault detection procedure based on reduced-size data sets. Section IV gives concluding remarks and future research directions.

II. Process, Data, and Wavelet Transforms

RTCVD is a process capable of depositing chemical materials on wafers with great precision. It is used in many semiconductor manufacturing systems. Our RTCVD process is equipped with many in situ temperature sensors at various wafer locations that are useful in controlling deposition uniformity better than other processes. See Tedder *et al.* [11] for details of the RTCVD process used in this study. QMS data were collected from a series of process and diagnostic experiments involving polycrystalline Si deposition from 10% SiH_4/Ar in a single-wafer RTCVD tool ranging in temperature from 625°C to 725°C. The

data used in this study consists of QMS samples of 21 nominal (in-control) process runs and four induced-fault RTCVD process runs. Fig. 1 displays the data for the first four nominal runs, and Fig. 2 displays the data for the four induced-fault runs. Note that there is a certain amount of secondary "local variation" exhibited in these data sets.

Wavelet transforms are very popular in many engineering and computing fields for solving real-life application problems. Wavelets can model irregular data patterns, such as sharp "jumps" and "dips," better than the Fourier transform and other standard statistical procedures, such as splines and nonparametric regression. In the following overview of wavelet analysis, $\mathbb{R}$ denotes the real line and $L^2(\mathbb{R})$ denotes the space of square integrable real functions defined on $\mathbb{R}$. A wavelet is a function $\psi(t) \in L^2(\mathbb{R})$ with the following basic properties

$$\int_{\mathbb{R}} \psi(t)dt = 0 \text{ and } \int_{\mathbb{R}} \psi^2(t)dt = 1.$$

An example of the $\psi(t)$ function is the "sombrero" wavelet (see [12, p. 77]), which looks like a Mexican sombrero. Wavelets can be used to create a family of time-frequency atoms, $\psi_{\eta,\tau}(t) = \eta^{1/2}\psi(\eta t - \tau)$, via the dilation factor η and the translation τ. We also require a scaling function $\phi(t) \in L^2(\mathbb{R})$ that satisfies

$$\int_{\mathbb{R}} \phi(t)dt \neq 0 \text{ and } \int_{\mathbb{R}} \phi^2(t)dt = 1.$$

Let $\mathbb{Z}$ denote the set of all integers $\{0, \pm 1, \pm 2, \ldots\}$. Starting from suitable choices of the scaling function $\phi(t)$ and the wavelet function $\psi(t)$ and using dilations of the form $\eta = 2^{j_0}$ and $\eta = 2^j$ together with translations of the form $\tau = k$ for $j_0, j, k \in \mathbb{Z}$, we can construct an orthonormal basis for the space $L^2(\mathbb{R})$ consisting of the scaling functions $\left\{\phi_{j_0,k}(t) = 2^{j_0/2}\phi\left(2^{j_0}t - k\right) : k \in \mathbb{Z}\right\}$ and the wavelet functions $\left\{\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right) : j \geq j_0, k \in \mathbb{Z}\right\}$; and in this situation, any target function $f(t) \in L^2(\mathbb{R})$ can be expressed as

$$f(t) = \sum_{k\in\mathbb{Z}} c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(t) \qquad (1)$$

where the coefficients

$$c_{j_0,k} = \int_{\mathbb{R}} f(t)\phi_{j_0,k}(t)dt \text{ and } d_{j,k} = \int_{\mathbb{R}} f(t)\psi_{j,k}(t)dt$$

are defined as the inner product of $f(t)$ with the basis functions $\phi_{j_0,k}(t)$ and $\psi_{j,k}(t)$, respectively. The representation (1) is analogous to the Fourier series representation of a square integrable function on the interval $[0, 2\pi)$ in terms of trigonometric basis functions.

When a discrete wavelet transform (DWT) $\boldsymbol{W}$ is applied to a data set $\boldsymbol{Y} = (y_1, \ldots, y_N)^{\mathrm{T}}$ of size N (where the superscript $^{\mathrm{T}}$ represents the transpose operator and where in general we must have $N = 2^\nu$ for some positive integer ν), the vector of wavelet coefficients can be expressed by $\boldsymbol{\Theta} = \boldsymbol{WY}$. If $\boldsymbol{W}$ defines an orthogonal DWT, then the $N \times 1$ vector $\boldsymbol{Y}$ of original observations can be recovered using the inverse DWT, $\boldsymbol{Y} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Theta}$. The DWT transforms N data points into N wavelet coefficients. The original data can be expressed as a linear sum of products of wavelet coefficients and their corresponding basis functions, as in (1). Extending the analogy between wavelet analysis and Fourier analysis, we see that transforming a data set via the DWT closely resembles the process of computing the fast Fourier transformation (FFT) of that data set.

III. A Procedure for Detecting Process Faults

The procedure for detecting process faults based on a reduced-size data set can be broken down into the following steps: a) data reduction; b) construction of the nominal process data model; c) development of the process fault detection test statistic; and d) application of the test statistic to detect potential process faults.

A. Data Reduction

In this article, the selected wavelet coefficients are treated as the reduced-size data set. Our data-reduction step consists of two substeps: 1) selecting wavelet coefficients (and their corresponding basis functions) by working with a single data set; and 2) deciding on a data-reduction strategy for all replicates, where each replicate is a different set of signals collected from an independent, identically distributed instance of the same in-control process.

1) Selecting Wavelet Coefficients for a Single Data Set: When using wavelet transforms, one needs first to decide which family of wavelets to use. The default wavelet family used in the popular S-PLUS package for the DWT operation is the symmlet, s8; see Bruce and Gao [13, p. 17] for details. The s8 wavelet is an excellent overall choice for representing many functions since it is orthogonal, smooth, nearly symmetric, and nonzero on a relatively short interval. In a study not reported here, we experimented with other wavelet families, such as coiflets and daublets. The results were similar to the results obtained using s8 wavelets. In this article we limit our discussion to the results obtained with the s8 wavelet.

One method often used to fit data using wavelets is to compute a set of multiresolution approximations [12], [13]. This method involves first constructing an approximation to the data using the coarsest-scale signal and then adding increasingly finer levels of resolution. As more levels of resolution are used, the approximation to the target data set improves. Figs. 3–5 below depict multiresolution approximations to a data set that we obtained from an in-control run of the RTCVD process. Notice in particular the following aspects of Figs. 3–5:

i) In the top panel of Fig. 3, the solid curve represents the wavelet approximation to the data set based on taking the level index $j_0 = 2$ in (1) and then estimating only the coefficients $\{c_{2,k}: k = 0, 1, 2, 3\}$ of all relevant shifts (translations) of the j_0-level scaling functions $\{\phi_{2,k}(t): k \in \mathbb{Z}\}$ while setting all other coefficients on the right-hand side of (1) to zero. Thus the top panel of Fig. 3 depicts the accuracy of the wavelet approximation to the target data set based on four coefficients at the coarsest level.

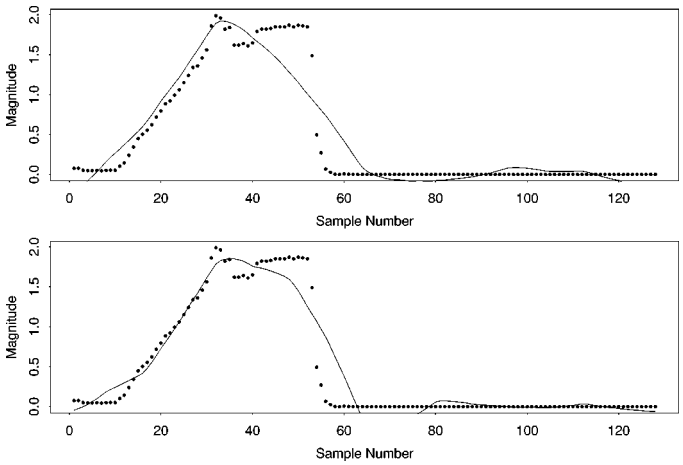

Fig. 3. Multiresolution approximation to an in-control run of the RTCVD process using four coefficients (top panel) and eight coefficients (bottom panel).

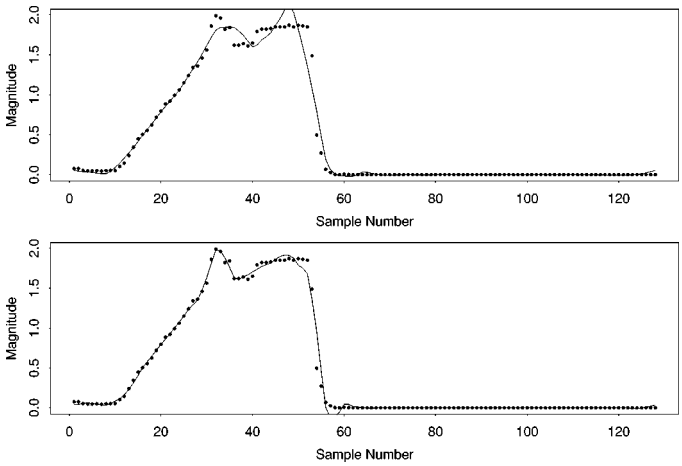

Fig. 4. Multiresolution approximation to an in-control run of the RTCVD process using 16 coefficients (top panel) and 32 coefficients (bottom panel).

ii) In the bottom panel of Fig. 3, the solid curve represents the wavelet approximation to the data set based on estimating the eight coefficients at the top two coarsest levels—namely, the coefficients $\{c_{2,k}: k = 0, 1, 2, 3\}$ of all relevant shifts of the j_0-level scaling functions $\{\phi_{2,k}(t): k \in \mathbb{Z}\}$ and the coefficients $\{d_{2,k}: k = 0, 1, 2, 3\}$ of the corresponding shifts of the j_0-level wavelet functions $\{\psi_{2,k}(t): k \in \mathbb{Z}\}$—while setting all other coefficients at the finer levels to zero.

iii) In the top panel of Fig. 4, the solid curve represents the wavelet approximation to the data set based on estimating 16 coefficients—namely, all the j_0-level coefficients mentioned in the previous item ii) and the coefficients $\{d_{3,k}: k = 0, 1, \ldots, 7\}$ of all relevant shifts of the (j_0+1)-level wavelet functions $\{\psi_{3,k}(t): k \in \mathbb{Z}\}$—while setting all other coefficients at the finer levels to zero.

iv) Similarly in the remaining panels of Figs. 4 and 5, the solid curve represents the wavelet approximation to the data set based on estimating the indicated number of coefficients up to and including all relevant shifts of the next-higher-level wavelet functions while setting all other coefficients at the finer levels to zero. At each level j in this example, there are 2^j nonzero coefficients $\{d_{j,k}: k = 0, 1, \ldots, 2^j - 1\}$ associated with the relevant shifts of the corresponding wavelet functions $\{\psi_{j,k}(t): k \in \mathbb{Z}\}$.

Notice the improvement in the approximation as more levels of detail are added by including coefficients at the finer levels hierarchically. At the finest level of resolution with index $j = j_0 + 4 = 6$, the total number of estimated wavelet coefficients equals the size of the data set $(N = 128)$ so that the data set is exactly reconstructed as shown in the bottom panel of Fig. 5.

While easy to use, this type of "linear" multiresolution approximation tends to oversmooth the data. For example, in the bottom panel of Fig. 3, the eight coefficients in the two coarsest-resolution levels are unable to represent the dip around sample number 40. Of course, if such a dip in the data is considered to be less important or data noise, then it is reasonable to filter out that dip. However, if that dip is considered to be an important characteristic of the target data set in representing process signatures, then we should use "nonlinear" approximation methods. In particular, nonlinear methods that select "important" wavelet coefficients (usually the largest in magnitude) and set to zero the "unimportant" coefficients (usually those representing noise) are effective in accurately representing small jumps or dips in the data with typically fewer coefficients than an approach based on a straightforward multiresolution approximation. Fig. 6 shows a reconstruction of the same data set depicted in Figs. 3–5 using the eight estimated wavelet coefficients having the largest magnitudes. Even though this reconstruction is not perfect, the dip around sample number 40 is represented. Fig. 7 shows the same data set reconstructed with the 19 estimated coefficients having the largest magnitudes. From this plot, it is clear that it is possible to achieve an excellent approximation to the data using as few as 19 coefficients.

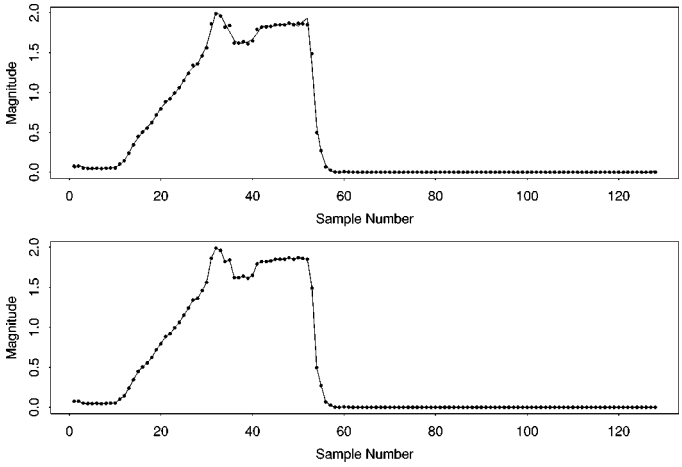

Fig. 5. Multiresolution approximation to an in-control run of the RTCVD process using 64 coefficients (top panel) and 128 coefficients (bottom panel).

There are many wavelet model selection procedures in the literature that are based on this idea of selecting "important" wavelet coefficients and setting to zero the "unimportant" coefficients. These methods attempt to find an optimal number of coefficients to accurately represent the data, thereby leading to

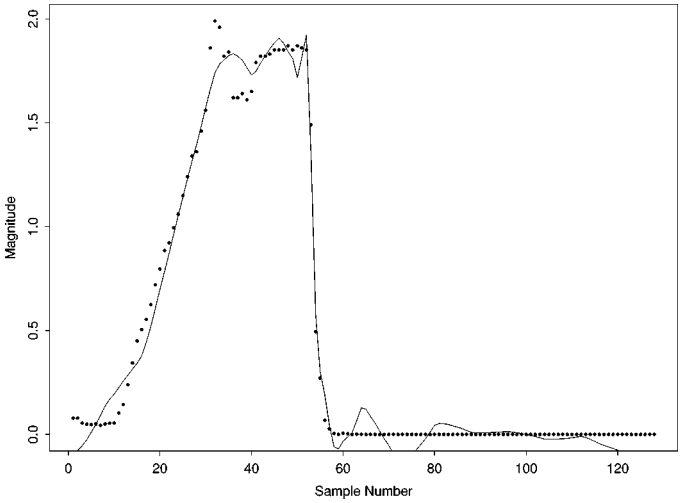

Fig. 6. Reconstruction of the in-control RTCVD data set using the eight largest-magnitude coefficients.

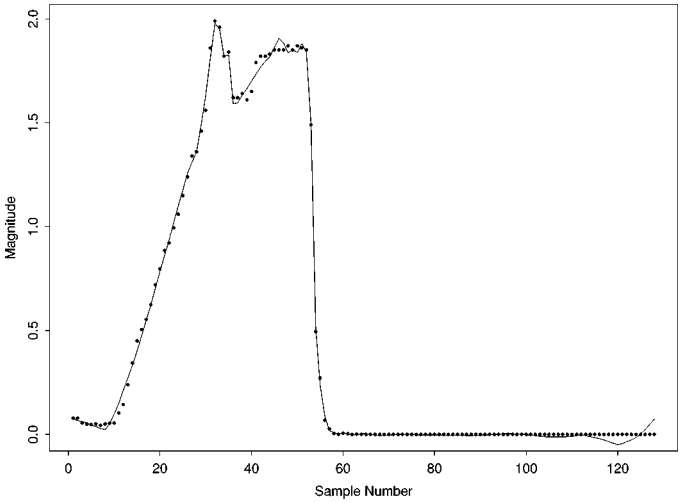

Fig. 7. Reconstruction of the in-control RTCVD data set using the 19 largest-magnitude coefficients.

a simplified and smoother (less noisy) data model. Two popular model selection examples are: the SURE (Stein's unbiased estimate of the risk function) method proposed by Donoho and Johnstone [8]; and the AMDL (approximate minimum description length) procedure proposed by Saito [9]. The AMDL method minimizes the cost function

$$\mathrm{AMDL}(C) = 1.5C\log_2 N + 0.5N\log_2\left[\sum_{i=1}^{N}(y_i - \widehat{y}_{i,C})^2\right] \tag{2}$$

where C is the number of wavelet coefficients selected to be nonzero and $\widehat{y}_{i,C}$ is the prediction of the ith observed response y_i based on the wavelet data model with the C largest-magnitude coefficients. As addressed in Antoniadis *et al.* [14], the AMDL function is similar to the Akaike information quantity commonly used in many statistical model selection procedures, including linear regression models.

While the existing wavelet model selection procedures, such as SURE and AMDL, are effective in denoising data, they tend to overfit the data and use an excessive number of coefficients. Because the goal of our wavelet coefficient selection method is data reduction, we would like to keep only a very small number of coefficients. However, it is also important to keep enough coefficients so that the data model represents the original data well. Below, we propose a new wavelet coefficient selection method to meet our goal of data reduction.

Almost all model selection methods in the literature are linked to the mean-squared error (MSE)

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_{i,C})^2$$

which characterizes the accuracy of the approximation to the original data. Our method balances a dimensionless version of this error against a data reduction ratio, C/N. That is, similar to the AMDL method, we keep the C largest-magnitude wavelet coefficients that minimize the following (penalized) "relative reconstruction error":

$$\mathrm{RRE}(C) = \frac{\left[\sum_{i=1}^{N}(y_i - \widehat{y}_{i,C})^2\right]^{1/2}}{\left(\sum_{i=1}^{N} y_i^2\right)^{1/2}} + \frac{C}{N}. \tag{3}$$

The first component of (3) represents a "normalized" reconstruction error from the approximated wavelet model structured by a linear sum of products of selected wavelet coefficients and their corresponding bases, similar to (1). The second component is the normalized number of coefficients used. In addition, it is possible to add a constant multiplier λ to the second term in (3) to control the trade-off between the two terms. In this study, we decided to keep the method simple by setting $\lambda = 1$, thereby using equal weights for both terms in (3). More investigation is needed to explore the choice of λ and its impact.

Fig. 8 illustrates the results of applying the RRE and AMDL estimation methods to the data collected on in-control run 5 of the RTCVD process. The upper left-hand plot shows that the data approximation error becomes smaller when the number of coefficients included in the wavelet model increases. The lower left-hand plot shows that the RRE has a minimum value at $C = 10$. The lower right-hand plot shows that the approximation model with $C = 10$ wavelet coefficients captures almost all details of the data pattern. The upper right-hand plot shows that the AMDL criterion (2) has a minimum at $C = 77$, which is much larger than the value of $C = 10$ at which the RRE criterion (3) is minimized. Table I shows similar results obtained for six other nominal runs (the column headed "Run 5" in Table I corresponds to the nominal run illustrated in Fig. 8).

2) Deciding on a Data Reduction Strategy for All Replicates: As indicated in Table I, the number of wavelet coefficients used to represent a data set can be different for different sets of data from the same process. Moreover, even if the number of coefficients is the same, the coefficients (and the corresponding basis functions) can have different values. If the process has little noise, then the data across replications should be very similar, and thus the selected wavelet coefficients for each data set should be similar. Because there is a certain amount of noise in the data obtained from the nominal process used in this study, our challenge is to decide on a set of wavelet

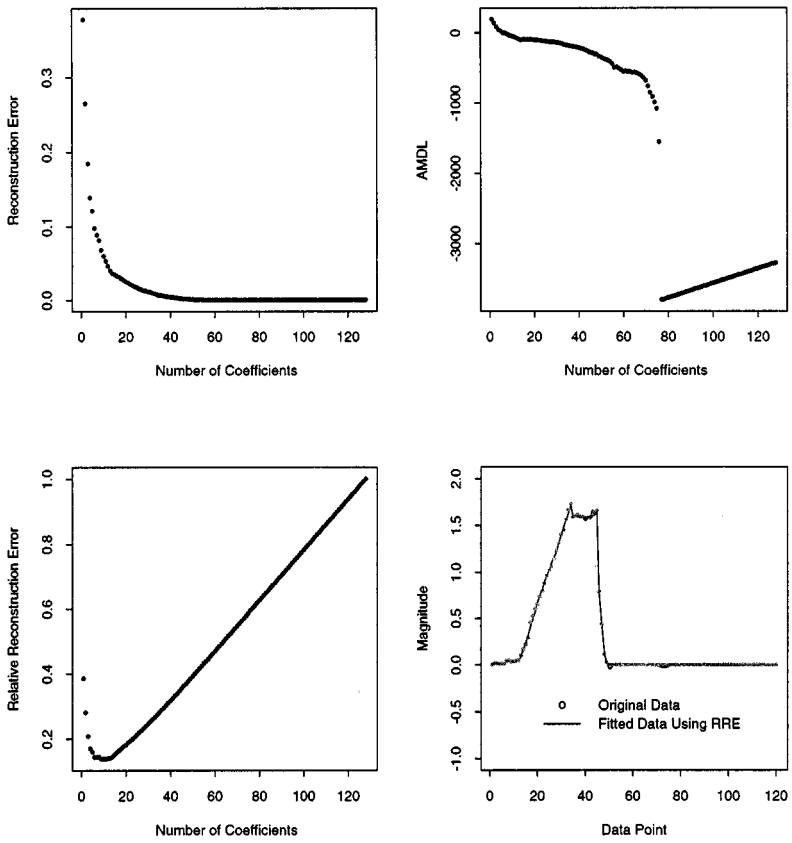

Fig. 8. Comparison of results of applying the RRE and AMDL estimation methods to run 5.

TABLE I
NUMBER OF COEFFICIENTS SELECTED FOR STANDARD RUNS

	Run						
Estimation Method	1	2	3	4	5	6	7
SURE	55	65	50	66	77	59	68
AMDL	86	86	93	86	77	90	84
RRE	11	14	10	11	10	12	10

coefficients to represent adequately the overall data structure of the nominal process.

First, we randomly selected seven data sets from the 21 signal data sets obtained from the nominal process. Using the RRE method, we selected wavelet coefficients for those seven data sets. To select representative coefficients for all replicates, there are a number of different strategies. For example, one can use the union set of all bases selected by the RRE method for the seven nominal data sets. This approach gives a comprehensive selection of the representative coefficients that covers many of the data fluctuations across replicates and captures the most important features of each set of data. However, the number of coefficients in the union set can be larger than the number of replicates of the in-control process. This will create problems later in computing the value of the overall performance statistic. An alternative approach is to select the most frequently occurring coefficients (that is, those coefficient positions that are selected by the RRE method on every replication), while at the same time keeping the number of selected coefficients smaller than the number of replicates. For our data, the most frequently occurring coefficients are $c_{2,0}$, $d_{2,0}$, $d_{2,1}$, $d_{2,2}$, $d_{3,0}$, and $d_{3,3}$, as shown in Table II below. The weakness of this method is that the more detailed data patterns described by the finer-level wavelet coefficients, such as $d_{6,26}$, are excluded. This will make the data-model approximation overly smoothed. A third alternative is to find a way to search through all coefficients to minimize the following overall relative reconstruction error (ORRE):

$$\mathrm{ORRE} = \sum_{m=1}^{M} \mathrm{RRE}(C_m)$$

where M is the number of replicates of the in-control process and $\mathrm{RRE}(C_m)$ is the relative reconstruction error for the mth data set. This approach involves a large number $(N \times M)$ of coefficients, however. Examining this many candidate models to find an appropriate data model can be very time consuming in practice. Finally, one could extend the procedures developed in the wavelet model selection literature to a multivariate case. Thus, the independent replicate data case becomes a special situation for applying the extended procedure. This idea could lead to a theoretically justified procedure. However, it is still under development and its performance remains to be studied [15].

To demonstrate the feasibility of using the wavelet coefficients estimated from a data set as the "reduced-size" data set for detecting process faults, we will use the 19 coefficients summarized in Table II as the selected coefficients in all subsequent discussion of the case study involving the RTCVD process. These 19 coefficients were obtained by taking the union set of the coefficients selected by the RRE method for each of the seven runs.

TABLE II
COEFFICIENTS SELECTED TO DESCRIBE THE NOMINAL PROCESS

Coefficient	Frequency	Coefficient	Frequency
$c_{2,0}$	7	$d_{4,5}$	4
$c_{2,1}$	4	$d_{4,6}$	4
$c_{2,3}$	1	$d_{4,7}$	2
$d_{2,0}$	7	$d_{5,12}$	5
$d_{2,1}$	7	$d_{5,13}$	3
$d_{2,2}$	7	$d_{6,22}$	1
$d_{3,0}$	7	$d_{6,25}$	1
$d_{3,2}$	4	$d_{6,26}$	2
$d_{3,3}$	7	$d_{6,27}$	1
$d_{3,4}$	4		

B. Construction of the Nominal Process Data Model

The estimated wavelet coefficients obtained from all replicates (at the bases shown in Table II) are treated as the reduced-size data set that will be used for making decisions concerning process performance. It is possible to utilize expert information to map the process fault patterns to some combination of these reduced-size data sets. Furthermore, artificial neural networks can be helpful in detecting and classifying fault types based on these wavelet coefficients. In this article, we use all coefficients listed in Table II to derive a hypothesis-testing procedure for detecting process faults. The nominal process data model will be approximated by using the coefficients listed in Table II, with their values computed according to the statistical-estimation procedure detailed in Section III-C below—in particular, see (6) below. If the values of the same 19 coefficients selected from a new data set are statistically different from the values of the coefficients obtained for the nominal process, then we conclude that the new process has a fault and is, therefore, out of control.

C. Development of the Process Fault Detection Statistic

Let $\left\{\widehat{\boldsymbol{\beta}}_i^{\circ}: i = 1, \ldots, M\right\}$ denote the $1 \times p$ dimensional vectors of estimated wavelet coefficients selected from the M replicates of the in-control process, where p is the number of wavelet coefficients used to describe the nominal process. (In our case study, recall that $p = 19$ and $M = 21$; and thus on the ith run of the in-control RTCVD process, the elements of $\widehat{\boldsymbol{\beta}}_i^{\circ}$ are the 19 corresponding estimates of the wavelet coefficients listed in Table II and computed via the DWT of the data set of size $N = 128$ recorded during that run, where $i = 1, \ldots, 21$.) Because these replicates are independent and identically distributed (i.i.d.) data-signal streams, the selected wavelet coefficients will naturally be i.i.d. across replications. However, the coefficients in a single data set are not i.i.d. since part of the same data set is used in determining the values of different coefficients. Let $\boldsymbol{\beta}^{\circ}$ and $\boldsymbol{\Sigma}^{\circ}$ respectively denote the mean vector and covariance matrix of the multivariate distribution of each in-control random vector $\widehat{\boldsymbol{\beta}}_i^{\circ}$. Note that we do not specify the distribution of each $\widehat{\boldsymbol{\beta}}_i^{\circ}$. If the original data are samples from a normal distribution, then each $\widehat{\boldsymbol{\beta}}_i^{\circ}$ will have a multivariate normal distribution since the wavelet coefficients obtained from the DWT are linear combinations of the original data. However, if the original data are not normal, then in general the wavelet coefficient vector $\widehat{\boldsymbol{\beta}}_i^{\circ}$ will not be multivariate normal. Our study will cover both the normal and nonnormal cases.

If we obtain a new set of data signals from a process whose in-control status is to be tested, then the new data set is transformed using the DWT in the same manner as for the original M runs that are known to be in control. We let $\widehat{\boldsymbol{\beta}}_{\text{new}}$ denote the vector of selected wavelet coefficients estimated from the new data set. We assume that the random vector $\widehat{\boldsymbol{\beta}}_{\text{new}}$ has expected value $\boldsymbol{\beta}_{\text{new}}$, which is not necessarily equal to $\boldsymbol{\beta}^{\circ}$, and covariance matrix $\boldsymbol{\Sigma}^{\circ}$, the same covariance matrix as for the $\left\{\widehat{\boldsymbol{\beta}}_i^{\circ}\right\}$.

To see if the new process is also in control, one can select a few (or all) wavelet coefficients and formulate a test statistic. The following hypothesis-testing procedure includes all p coefficients. The same idea works for a few targeted coefficients. The null hypothesis that the new process is in control can be written as

$$H_0: \boldsymbol{\beta}_{\text{new}} = \boldsymbol{\beta}^{\circ}. \tag{4}$$

The first step in testing this null hypothesis is to take independent samples of size s without replacement from the set of estimated wavelet coefficient vectors corresponding to the independent replicates of the in-control process,

$$\left\{\widehat{\boldsymbol{\beta}}_i^{\circ}: i = 1, \ldots, M\right\}. \tag{5}$$

This sampling operation is independently replicated K times. For the kth random sample of size s selected without replacement from (5), we let $\overline{\widehat{\boldsymbol{\beta}}}_k$ denote the corresponding sample mean. For reasons explained in Appendix A, we refer to this sampling scheme as a "reversed jackknife" method.

We observe that the $\left\{\overline{\widehat{\boldsymbol{\beta}}}_k: k = 1, \ldots, K\right\}$ are independent and identically distributed since they are obtained via K independent replications of the basic operation of taking a sample of size s without replacement from (5). From this random sample we calculate the grand mean

$$\overline{\overline{\widehat{\boldsymbol{\beta}}}} = \frac{1}{K}\sum_{k=1}^{K}\overline{\widehat{\boldsymbol{\beta}}}_k \tag{6}$$

and the sample covariance matrix

$$\boldsymbol{S} = \frac{1}{K-1}\sum_{k=1}^{K}\left(\overline{\widehat{\boldsymbol{\beta}}}_k - \overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)^{\mathrm{T}}\left(\overline{\widehat{\boldsymbol{\beta}}}_k - \overline{\overline{\widehat{\boldsymbol{\beta}}}}\right). \tag{7}$$

We combine these in-control sample statistics with the estimated wavelet coefficient vector $\widehat{\boldsymbol{\beta}}_{\text{new}}$ for the new process to obtain a variant of the classical two-sample Hotelling's T^2-statistic that has been adapted to the reversed jackknife sampling scheme:

$$T^2_{\text{new}} = \left[\frac{KM\left(1 - \frac{s}{M}\right)}{sKM + M - s}\right]\left(\widehat{\boldsymbol{\beta}}_{\text{new}} - \overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)\boldsymbol{S}^{-1} \times \left(\widehat{\boldsymbol{\beta}}_{\text{new}} - \overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)^{\mathrm{T}} \tag{8}$$

where T^2_{new} has the T^2-distribution with $K - 1$ degrees of freedom, provided that the $\left\{\widehat{\boldsymbol{\beta}}_i^{\circ}: i = 1, \ldots, M\right\}$ and $\widehat{\boldsymbol{\beta}}_{\text{new}}$ are

i.i.d. normal with mean vector $\boldsymbol{\beta}^{\circ}$ and covariance matrix $\boldsymbol{\Sigma}^{\circ}$. Furthermore, under this normality assumption, we see that

$$F_{\text{new}} = \frac{T^2_{\text{new}}(K-p)}{(K-1)p} \tag{9}$$

has an F-distribution with p and $K-p$ degrees of freedom. Appendix A contains a derivation of (8) and (9) under the normality assumption and the null hypothesis (4). If the value of the test statistic (9) is larger than the F critical value for a prespecified significance level (for example, $\alpha = 0.05$ or $\alpha = 0.10$), then we conclude that there are faults in the process from which $\widehat{\boldsymbol{\beta}}_{\text{new}}$ was estimated.

1) Formulation of Nonparametric Fault-Detection Procedure for Nonnormal Data: If one does not assume that $\left\{\widehat{\boldsymbol{\beta}}_i^{\circ}: i = 1, \ldots, M\right\}$ and $\widehat{\boldsymbol{\beta}}_{\text{new}}$ are normal, then the F-distribution with p and $K-p$ degrees of freedom is only an approximation to the distribution of the test statistic (9). In the case where the data are not normal, resampling procedures such as bootstrapping [16] are typically used to develop an approximation to the distribution of (9). However, standard resampling procedures can yield an estimate of the covariance matrix $\boldsymbol{\Sigma}^{\circ}$ that is singular; and in such situations, a test statistic like (9) cannot be computed. In our case study with $M = 21$ and $p = 19$, the probability of generating a conventional bootstrap sample of size M from (5) (i.e., a random sample selected with replacement from (5)) that possesses a nonsingular sample covariance matrix is less than 3.526×10^{-4}; see Appendix A of Lada [17]. Clearly the traditional approach to bootstrapping will not work in this situation; and this is the fundamental reason for the reversed jackknife sampling method that we proposed in (5)–(7) to compute the final test statistic (9).

To handle the case where the data are not necessarily normal, we develop an alternative to traditional bootstrapping techniques. In our procedure, the empirical distribution of the test statistic (9) under the null hypothesis (4) is used to determine if a new data set, represented by $\widehat{\boldsymbol{\beta}}_{\text{new}}$, is in control or not. Fig. 9 contains a formal statement of the procedure to detect process faults in the nonnormal case. Notice that steps **1(a)–1(c)** of Fig. 9 generate an empirical distribution $\{F_r\}$ of the test statistic (9); and then the given vector $\widehat{\boldsymbol{\beta}}_{\text{new}}$ is tested for the in-control condition in steps **3(a)–3(c)** of Fig. 9 by referring the associated value of F_{new} to the empirical distribution $\{F_r\}$. See Appendix B for more details on how this procedure was derived.

2) Setting s, K, and M for the Fault-Detection Procedure: Some care is required in selecting appropriate values of s and K so that the reversed jackknife sampling scheme can be used effectively in a process fault detection procedure. Given the size M of the original set of in-control estimated wavelet coefficient vectors (5), we see that there are in general $\binom{M}{s} = M!/[s!(M-s)!]$ ways to choose s distinct vectors that will constitute a sample of size s selected without replacement from (5). In view of the discussion in Section III-C1 and in Appendix A, we see that to ensure invertibility of the sample covariance matrix $\boldsymbol{S}$ computed from the sample means $\left\{\overline{\widehat{\boldsymbol{\beta}}}_k: k = 1, \ldots, K\right\}$ based on the reversed jackknife method, the value of s must be chosen so as to guarantee that $\binom{M}{s}$ is sufficiently large. More precisely, $\binom{M}{s}$ should be much larger than K to ensure that with probability close to one, the set $\left\{\overline{\widehat{\boldsymbol{\beta}}}_k: k = 1, \ldots, K\right\}$ will include p linearly independent vectors so that $\boldsymbol{S}^{-1}$ exists and the test statistic (9) can actually be calculated. We observe that if the $\left\{\overline{\widehat{\boldsymbol{\beta}}}_k\right\}$ were continuous random vectors, then it would follow from Proposition 2 of [18] that $\boldsymbol{S}$ would be invertible with probability one. Moreover as s increases, a central limit-type effect ensures that the $\left\{\overline{\widehat{\boldsymbol{\beta}}}_k\right\}$ will converge in distribution to multivariate normality and thus will tend to behave like continuous random vectors. This observation coupled with extensive computational experience using the process fault detection procedure of Fig. 9 has led us to the conclusion that in addition to requiring $\binom{M}{s} \gg K$, we should take s just large enough so that the set of sample mean vectors $\left\{\overline{\widehat{\boldsymbol{\beta}}}_k: k = 1, \ldots, K\right\}$ will pass the Shapiro–Wilk test for multivariate normality at a standard level of significance (say, $\alpha = 0.05$ or $\alpha = 0.10$). In addition to ensuring that $\boldsymbol{S}^{-1}$ exists with probability close to one, we have found that this approach to setting s improves the adequacy of approximating the null distribution of (9) with either the F-distribution having p and $K-p$ degrees of freedom or with the empirical distribution of the $\{F_r\}$ generated in step **1** of Fig. 9.

Clearly the selection of appropriate values for s and K depends critically on the selection of an appropriate value for M; and as in all resampling schemes (including conventional bootstrapping schemes), the validity of our nonparametric process fault detection procedure depends on the selection of a sufficiently large value of M at the outset. We performed an extensive set of Monte Carlo experiments to formulate effective rules of thumb for setting s, K, and M in general applications of our fault detection procedure. Based on $R = 1000$ replications of steps **1** and **3** of the procedure, Fig. 10 displays the following c.d.f.'s for a typical Monte Carlo experiment in which we took $p = 20$, $M = 200$, $s = 5$, and $K = 60$ so that the original set of wavelet coefficient vectors $\left\{\widehat{\boldsymbol{\beta}}_i^{\circ}: i = 1, \ldots, M\right\}$ was sampled from a 20-dimensional multivariate uniform distribution, with covariance structure typical of estimated wavelet coefficients (specifically for this example, the lag-1 correlations were set to 0.4 while all other correlations were set to zero):

i) the empirical c.d.f. of the $\{F_r: r = 1, \ldots, R\}$ generated by Step **1**;
ii) the empirical c.d.f. of F_{new} based on $R = 1000$ replications of Step **3** in which $\widehat{\boldsymbol{\beta}}_{\text{new}}$ was sampled from the same underlying multivariate uniform distribution from which we sampled the original set (5) of in-control estimated wavelet coefficient vectors; and
iii) the c.d.f. of the corresponding F-distribution with p and $K-p$ degrees of freedom.

It is clear from Fig. 10 that in general the distribution of the final test statistic F_{new} in item ii) above differs not only from the conventional normal-theory approximation in item iii) above but also from our reversed jackknife approximation in item i) above. Based on extensive computational experience with the process fault detection procedure of Fig. 9, we have found that the approximations i) and iii) to the null distribution of the final

Nonparametric Procedure for Detecting Process Faults

1. For $r = 1, \ldots, R$ ($R \geq 1000$), perform the rth replication of steps **1(a)–1(c)** below.

(a) Independently generate K samples of size s, where the s units in each sample are selected without replacement from (5). Let $\overline{\overline{\beta}}_{k,r}$ denote the sample mean of the kth sample for $k = 1, \ldots, K$.

(b) Calculate the grand mean vector and the sample covariance matrix of the $\left\{\overline{\overline{\beta}}_{k,r} : k = 1, \ldots, K\right\}$,

$$\overline{\overline{\overline{\beta}}}_r = \frac{1}{K}\sum_{k=1}^{K}\overline{\overline{\beta}}_{k,r} \quad \text{and} \quad \boldsymbol{S}_r = \frac{1}{K-1}\sum_{k=1}^{K}\left(\overline{\overline{\beta}}_{k,r} - \overline{\overline{\overline{\beta}}}_r\right)^{\mathrm{T}}\left(\overline{\overline{\beta}}_{k,r} - \overline{\overline{\overline{\beta}}}_r\right). \tag{10}$$

(c) Randomly sample $\widehat{\boldsymbol{\beta}}_r$ from (5) and compute the test statistic

$$F_r = \frac{T_r^2(K-p)}{(K-1)p}, \tag{11}$$

where

$$T_r^2 = \left[\frac{\left(1-\frac{s}{M}\right)/s}{\left(1-\frac{1}{M}\right)+\left(1-\frac{s}{M}\right)/(sK)}\right]\left(\widehat{\boldsymbol{\beta}}_r - \overline{\overline{\overline{\beta}}}_r\right)\boldsymbol{S}_r^{-1}\left(\widehat{\boldsymbol{\beta}}_r - \overline{\overline{\overline{\beta}}}_r\right)^{\mathrm{T}}.$$

2. Sort the test statistics $\{F_r : r = 1, \ldots R\}$ in ascending order. Let F_L and F_U be the $\alpha/2$ and $1-\alpha/2$ percentiles, respectively, of the $\{F_r\}$.

3. To test a new vector of estimated wavelet coefficients $\widehat{\boldsymbol{\beta}}_{\text{new}}$ for the in-control condition, perform steps **3(a)–3(c)** below.

(a) As in step **1(a)** above, independently generate K samples of size s, where the s units in each sample are selected without replacement from (5); then compute the corresponding sample means $\left\{\overline{\overline{\beta}}_{k,0} : k = 1, \ldots, K\right\}$; and finally compute the corresponding overall statistics $\overline{\overline{\overline{\beta}}}_0$ and $\boldsymbol{S}_0$ as in (10) with $r = 0$.

(b) Compute the T^2-value and the corresponding F-value

$$T_{\text{new}}^2 = \left[\frac{KM\left(1-\frac{s}{M}\right)}{sKM+M-s}\right]\left(\widehat{\boldsymbol{\beta}}_{\text{new}} - \overline{\overline{\overline{\beta}}}_0\right)\boldsymbol{S}_0^{-1}\left(\widehat{\boldsymbol{\beta}}_{\text{new}} - \overline{\overline{\overline{\beta}}}_0\right)^{\mathrm{T}} \quad \text{and} \quad F_{\text{new}} = \frac{T_{\text{new}}^2(K-p)}{(K-1)p}.$$

(c) If F_{new} falls outside the $100(1-\alpha)\%$ tolerance interval (F_L, F_U), then the process yielding $\widehat{\boldsymbol{\beta}}_{\text{new}}$ is out of control; otherwise that process is considered to be in-control.

Fig. 9. Procedure for detecting process faults when the original data sets are nonnormal.

test statistic F_{new} will be adequate in practice if we select M, K, and s according to the following rules of thumb:

$$M \geq 15p, \quad K \approx \max\left\{p+20, \frac{M}{s}\right\}, \text{ and } \binom{M}{s} \gg K. \tag{12}$$

D. Process Fault Detection Applications

As mentioned previously, we observed $M = 21$ in-control runs of the RTCVD process to form the set (5) of in-control estimated wavelet coefficient vectors. We then took $K = 40$ samples of size $s = 5$ without replacement from the set (5) to compute the sample mean vectors $\left\{\overline{\overline{\beta}}_k : k = 1, \ldots, K\right\}$ and the overall sample statistics $\overline{\overline{\overline{\beta}}}$ and $\boldsymbol{S}$. The four induced-fault runs of the RTCVD process shown in Fig. 2 were used to test the effectiveness of the proposed process fault detection procedure. Notice that two of the faults are significantly different from the in-control process while the other two faults, representing a more difficult case to detect, are similar to the in-control process. For each of the four induced-fault runs, we computed the corresponding wavelet coefficient vector $\widehat{\boldsymbol{\beta}}_{\text{new}}$; and finally we computed the test statistic F_{new} corresponding to $\widehat{\boldsymbol{\beta}}_{\text{new}}$.

To ensure adequate performance of the process fault detection statistic (9), we attempted to apply the rules of thumb (12) while taking s large enough to ensure that the set of p-dimensional sample means $\left\{\overline{\overline{\beta}}_k : k = 1, \ldots, K\right\}$ would pass the Shapiro–Wilk test for multivariate normality. Using Royston's algorithm [19] for the extended univariate Shapiro–Wilk test, we developed an S-PLUS program to compute the multivariate Shapiro–Wilk test statistic W^*. This program is available from the authors on request. Corresponding to $K = 40$ samples

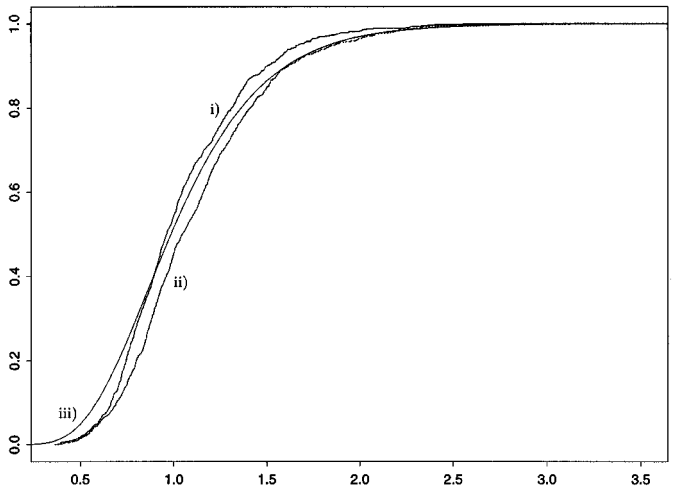

Fig. 10. c.d.f.'s for the nonparametric procedure of Fig. 9: i) c.d.f. of the $\{F_r: r = 1, \ldots, R\}$ (solid step function) generated by Step **1**; ii) empirical c.d.f. of F_{new} based on $R = 1000$ replications of Step **3** with $\widehat{\boldsymbol{\beta}}_{\text{new}}$ sampled from the same in-control distribution as for (5) (dashed step function); and iii) the corresponding F c.d.f. (solid curve) with p and $K - p$ degrees of freedom.

of size $s = 5$, the resulting multivariate Shapiro–Wilk test statistic for the $\left\{\widehat{\overline{\boldsymbol{\beta}}}_k\right\}$ is $W^* = 0.670$, and the associated P-value is 0.521. Note moreover that in this application, we have $\binom{M}{s} = 20{,}349 \gg 40 = K$, which clearly satisfies the third condition in (12). In all applications of our process fault detection procedure to this case study and to other problems in which we have taken s just large enough to ensure that the $\left\{\widehat{\overline{\boldsymbol{\beta}}}_k\right\}$ pass the multivariate Shapiro–Wilk test and $\binom{M}{s} \gg K$, we have never generated a sample $\left\{\widehat{\overline{\boldsymbol{\beta}}}_k\right\}$ for which the corresponding sample covariance matrix $\boldsymbol{S}$ failed to be invertible. We believe that this observation provides good evidence of the effectiveness of our rules of thumb in (12) for setting s and K in practical applications of the process fault detection procedure. On the other hand, it should be noted that since $p = 19$ in the RTCVD case study, the first two conditions in (12) would require $M \geq 285$ and $K \approx 57$. Since we were constrained to taking $M = 21$ with the corresponding value of $K = 40$, we do not have an adequate overall approximation to the distribution of F_{new} under the null hypothesis (4) based on either the $\{F_r\}$ in our nonparametric resampling scheme or the F-distribution with p and $K - p$ degrees of freedom.

Some care is required in interpreting the results that we obtained from applying the process fault detection procedure to the induced-fault runs of the RTCVD process. We obtained the 95% tolerance interval (0.54, 2.00) for an in-control value of the test statistic (9). Table III displays the computed value of (9) for each of the four induced-fault runs, as well as their P-values computed from the empirical distribution $\{F_r: r = 1, \ldots, R\}$. Clearly, all four induced-fault test statistic values fall well outside the given 95% tolerance interval. Furthermore, note that the estimated P-value based on the $\{F_r\}$ is zero for all four runs. Table III also shows the P-value when the test statistic (9) is compared to the F-distribution with $p = 19$ and $K - p = 21$ degrees of freedom. Notice that these P-values are also very close to 0. Since we do not have an adequate overall approximation to the distribution of F_{new} under the null hypothesis (4), the 95% tolerance interval (0.54, 2.00) is of questionable validity. Nevertheless, we see that in the extreme tails of the distribution of F_{new}, both of our approximations are sufficiently accurate to justify the conclusion that all four induced-fault runs are out of control. Therefore, we concluded that all four runs have potential process faults, and steps should be taken to identify and remove these faults.

TABLE III
TEST STATISTIC VALUES FOR THE FOUR FAULT RUNS

		Calculated P-value	
Fault	F_{new}	From $\{F_r\}$	From $F(p, K-p)$
1	8.48	0	4.73×10^{-6}
2	94.7	0	3.33×10^{-16}
3	5190.	0	$< 10^{-38}$
4	205.2	0	$< 10^{-38}$

IV. CONCLUSIONS AND FUTURE RESEARCH

We believe that the results presented in this paper provide good evidence of the effectiveness of the proposed method for detecting process faults from a reduced-size data set consisting of a selected set of estimated wavelet coefficients. If the original data are normal, or if one simply wants a quick test for process faults with a minimum amount of computation, then the value of the proposed test statistic (9) for a new set of data can be compared to the F-distribution with p and $K - p$ degrees of freedom. Otherwise, to obtain the most accurate results, we recommend that the test statistic value for a new data set be compared to the empirical null distribution generated by the algorithm of Fig. 9—provided that the conditions of (12) are satisfied.

The next logical step after detecting possible faults in a process is to determine the cause of the fault. When there is enough information to map the wavelet coefficients to known faults, artificial neural networks or other pattern recognition procedures can be employed to classify faults and suggest remedial procedures. Further research on building fault classification signatures from the reduced-size data set (that is, the wavelet coefficients) is needed. By using process fault detection and fault classification techniques, manufacturing systems can improve their quality and operational efficiency.

APPENDIX A
"REVERSED JACKKNIFE" METHOD

Typical jackknife procedures [20], such as the "delete-1 jackknife" and the "delete-2 jackknife," delete one or two data points from the original data set that contains M replications to construct a new sample for calculating the test statistic value. In the delete-1 jackknife procedure, this process is repeated M times (since there are M different ways to generate samples of size $M - 1$ from the original sample of size M). Similarly, in the delete-2 jackknife procedure, the process is repeated $\binom{M}{2}$ times. In our case study, we take random samples of size s, which means that our procedure randomly deletes $M - s$ of the data points. Unlike the jackknife procedure, our procedure repeats the resampling process K times (it does not go through all the

cases of deleting $M-s$ data points). Therefore, in some sense, our procedure can be considered a "reversed jackknife" procedure.

Since each sample of size s is selected without replacement from the original data set of estimated wavelet coefficients for the M in-control processes, $\overline{\widehat{\boldsymbol{\beta}}}_k$ is the average of s quantities that in general are assumed to be identically distributed random vectors with mean $\boldsymbol{\beta}^\circ$ and covariance matrix $\boldsymbol{\Sigma}^\circ$. Note, however, that the vectors going into the computation of $\overline{\widehat{\boldsymbol{\beta}}}_k$ are not independent because they are sampled without replacement from the finite set (5); and we have

$$\operatorname{Cov}\left(\overline{\widehat{\boldsymbol{\beta}}}_k\right)=\frac{\operatorname{Cov}\left(\widehat{\boldsymbol{\beta}}_i^\circ\right)}{s}\left(1-\frac{s}{M}\right)=\frac{\boldsymbol{\Sigma}^\circ}{s}\left(1-\frac{s}{M}\right) \quad (13)$$

where the quantity $(1-(s/M))$ is the "finite population correction" for sample size s and population size M [21]. If the sampling fraction s/M is small, then this correction is close to unity and therefore it has little effect on the standard error of the sample mean. Otherwise, if the correction is ignored, the covariance matrix of the sample mean will be overestimated. Finally invoking the Central Limit Theorem, we see that provided s is sufficiently large, $\overline{\widehat{\boldsymbol{\beta}}}_k$ is approximately normal with mean vector $\boldsymbol{\beta}^\circ$ and covariance matrix (13).

To define the fault detection test statistic, we compute the overall sample statistics, $\overline{\overline{\widehat{\boldsymbol{\beta}}}}$ and $\boldsymbol{S}$, as given by (6) and (7), respectively. Notice that

$$\begin{aligned}\mathrm{E}\left[\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right]&=\mathrm{E}\left[\overline{\widehat{\boldsymbol{\beta}}}\right]=\boldsymbol{\beta}^\circ \text{ and } \mathrm{E}\left[\boldsymbol{S}\right]=\operatorname{Cov}\left(\overline{\widehat{\boldsymbol{\beta}}}_k\right)\\&=\frac{\boldsymbol{\Sigma}^\circ}{s}\left(1-\frac{s}{M}\right).\end{aligned} \quad (14)$$

Hence, $\boldsymbol{S}$ is an unbiased estimator of $\operatorname{Cov}\left(\overline{\widehat{\boldsymbol{\beta}}}_k\right)$, and $\overline{\overline{\widehat{\boldsymbol{\beta}}}}$ is approximately normally distributed with mean $\boldsymbol{\beta}^\circ$ (the same mean as for the original M nominal runs) and covariance matrix

$$\operatorname{Cov}\left(\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)=\frac{\operatorname{Cov}\left(\overline{\widehat{\boldsymbol{\beta}}}_k\right)}{K}=\frac{\boldsymbol{\Sigma}^\circ}{sK}\left(1-\frac{s}{M}\right). \quad (15)$$

Now if the $\left\{\widehat{\boldsymbol{\beta}}_i^\circ: i=1,\ldots,M\right\}$ and $\widehat{\boldsymbol{\beta}}_{\text{new}}$ are i.i.d. normal with mean vector $\boldsymbol{\beta}^\circ$ and covariance matrix $\boldsymbol{\Sigma}^\circ$, then we are able to construct the following test statistic for determining if the wavelet coefficients (the reduced-size data sets) obtained from the new and the nominal processes are equal statistically:

$$\boldsymbol{D}=\sqrt{\frac{sKM}{sKM+M-s}}\left(\widehat{\boldsymbol{\beta}}_{\text{new}}-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right) \quad (16)$$

where $\widehat{\boldsymbol{\beta}}_{\text{new}}$ is a vector of wavelet coefficient values obtained from a new process. Note that under the null hypothesis (4), we have $\mathrm{E}[\boldsymbol{D}]=0$, and

$$\begin{aligned}\operatorname{Cov}(\boldsymbol{D})&=\frac{sKM}{sKM+M-s}\left[\operatorname{Cov}\left(\widehat{\boldsymbol{\beta}}_{\text{new}}\right)+\operatorname{Cov}\left(\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)\right]\\&=\frac{sKM}{sKM+M-s}\left[\boldsymbol{\Sigma}^\circ+\frac{\boldsymbol{\Sigma}^\circ}{sK}\left(1-\frac{s}{M}\right)\right]=\boldsymbol{\Sigma}^\circ.\end{aligned} \quad (17)$$

By multiplying the difference $\left(\widehat{\boldsymbol{\beta}}_{\text{new}}-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)$ by $\sqrt{sKM/(sKM+M-s)}$, we obtain a test statistic $\boldsymbol{D}$ that has the same covariance matrix as $\left\{\widehat{\boldsymbol{\beta}}_i^\circ\right\}$ and $\widehat{\boldsymbol{\beta}}_{\text{new}}$.

The statistic $\boldsymbol{D}$ can now be used to calculate a two-sample version of Hotelling's T^2-statistic. In view of (13)–(17), we have

$$\begin{aligned}T_{\text{new}}^2=&\sqrt{\frac{sKM}{sKM+M-s}}\left(\widehat{\boldsymbol{\beta}}_{\text{new}}-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)\left[\frac{s}{\left(1-\frac{s}{M}\right)}\boldsymbol{S}\right]^{-1}\\&\times\sqrt{\frac{sKM}{sKM+M-s}}\left(\widehat{\boldsymbol{\beta}}_{\text{new}}-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)^{\mathrm{T}}\\=&\left[\frac{KM\left(1-\frac{s}{M}\right)}{sKM+M-s}\right]\left(\widehat{\boldsymbol{\beta}}_{\text{new}}-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)\boldsymbol{S}^{-1}\\&\times\left(\widehat{\boldsymbol{\beta}}_{\text{new}}-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)^{\mathrm{T}}\end{aligned} \quad (18)$$

where T_{new}^2 has the T^2-distribution with $K-1$ degrees of freedom. Note that in order for the matrix $\boldsymbol{S}$ to be nonsingular so that (16) can be computed, a necessary and sufficient condition is that the corresponding random sample $\left\{\overline{\widehat{\boldsymbol{\beta}}}_k: k=1,\ldots,K\right\}$ must contain p linearly independent vectors and must satisfy $K\geq p+1$; see Theorem 7.5.2 of [22]. Finally, Theorem 5.2.2 of [22] ensures that if the $\left\{\widehat{\boldsymbol{\beta}}_i^\circ: 1\leq i\leq M\right\}$ and $\widehat{\boldsymbol{\beta}}_{\text{new}}$ are i.i.d. normal with mean vector $\boldsymbol{\beta}^\circ$ and covariance matrix $\boldsymbol{\Sigma}^\circ$, then (9) has an F-distribution with p and $K-p$ degrees of freedom.

APPENDIX B
EMPIRICAL DISTRIBUTION OF TEST STATISTIC FOR NONNORMAL DATA

We can compute the empirical distribution of the test statistic (11) (see Fig. 9) under the null hypothesis (4) in the following way. To generate the rth random sample from this distribution for $r=1,\ldots,R$, we calculate $\overline{\overline{\widehat{\boldsymbol{\beta}}}}_r$ and $\boldsymbol{S}_r$ according to (6) and (7), respectively, based on a new sample $\left\{\overline{\widehat{\boldsymbol{\beta}}}_{k,r}: k=1,\ldots,K\right\}$. Finally to compute the rth observation of (11), we randomly sample $\widehat{\boldsymbol{\beta}}_r$ from the original set of M wavelet coefficient vectors representing the nominal process. Taking into account the finite population correction, we see that $\operatorname{Cov}\left(\widehat{\boldsymbol{\beta}}_r\right)=\boldsymbol{\Sigma}^\circ(1-1/M)$. This results in the slightly altered test statistic

$$\begin{aligned}T_r^2=&\left[\frac{1}{\left(1-\frac{1}{M}\right)+\frac{\left(1-\frac{s}{M}\right)}{sK}}\right]\left(\frac{1-\frac{s}{M}}{s}\right)\left(\widehat{\boldsymbol{\beta}}_r-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)\boldsymbol{S}^{-1}\\&\times\left(\widehat{\boldsymbol{\beta}}_r-\overline{\overline{\widehat{\boldsymbol{\beta}}}}\right)^{\mathrm{T}}\end{aligned} \quad (19)$$

which is easily seen to be equivalent to the expression for T_r^2 in step **1(c)** of Fig. 9.

REFERENCES

[1] O. A. Basir and H. C. Shen, "Modeling and fusing uncertain multi-sensory data," *J. Robotic Syst.*, vol. 13, no. 2, pp. 95–109, 1996.

[2] B. Kim and G. S. May, "Real-time diagnosis of semiconductor manufacturing equipment using a hybrid neural network expert system," *IEEE Trans. Comp., Packag., Manufact. Technol.—Part C*, vol. 20, pp. 39–47, Mar. 1997.

[3] R. Isermann, "Process fault detection based on modeling and estimation methods—A survey," *Automatica*, vol. 20, no. 4, pp. 387–404, 1984.
[4] J. Jin and J. Shi, "Feature-preserving data compression of stamping tonnage information using wavelets," *Technometrics*, vol. 41, no. 4, pp. 327–339, 1999.
[5] X. Z. Wang, B. H. Chen, S. H. Yang, and C. McGreavy, "Application of wavelets and neural networks to diagnostic system development, 2, an integrated framework and its application," *Comput. Chem. Eng.*, vol. 23, pp. 945–954, 1999.
[6] E. A. Rying, G. L. Bilbro, and J.-C. Lu, "Focused local learning with wavelet neural networks," *IEEE Trans. Neural Networks*, to be published.
[7] E. A. Rietman and M. Beachy, "A study on failure prediction in a plasma reactor," *IEEE Trans. Semiconduct. Manufact.*, vol. 11, pp. 670–680, Nov. 1998.
[8] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.
[9] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelets in Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds. New York: Academic, 1994, pp. 299–324.
[10] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[11] L. L. Tedder, G. W. Rubloff, B. F. Cohaghan, and G. N. Parsons, "Dynamic rate and thickness metrology during poly-Si rapid thermal chemical vapor deposition from SiH_4 using real time in situ mass spectrometry," *J. Vacuum Sci. Technol.*, ser. A, vol. 14, no. 2, pp. 267–270, 1996.
[12] S. G. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.
[13] A. Bruce and H.-Y. Gao, *Applied Wavelet Analysis With S-PLUS*. New York: Springer-Verlag, 1996.
[14] A. Antoniadis, I. Gijbels, and G. Grégoire, "Model selection using wavelet decomposition and applications," *Biometrika*, vol. 84, no. 4, pp. 751–763, 1997.
[15] B. Vidakovic, Private Communication, 2001.
[16] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
[17] E. K. Lada, "A Wavelet-Based Procedure for Process Fault Detection," Master's thesis, Graduate Program in Operations Research, North Carolina State Univ., Raleigh, NC, 2000.
[18] A. M. Porta Nova and J. R. Wilson, "Estimation of multiresponse simulation metamodels using control variates," *Manag. Sci.*, vol. 35, no. 11, pp. 1316–1333, 1989.
[19] J. P. Royston, "An extension of Shapiro and Wilk's *W* test for normality to large samples," *Appl. Statist.*, vol. 31, no. 2, pp. 115–124, 1982.
[20] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: Soc. for Industrial and Applied Mathematics, 1982.
[21] W. G. Cochran, *Sampling Techniques*, 3rd ed. New York: Wiley, 1977.
[22] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.

Emily K. Lada received the B.A. degree in mathematics from the University of North Carolina at Chapel Hill and the M.S. degree in operations research from North Carolina State University, Raleigh. She is working toward Ph.D. degreee in the Operations Research Graduate Program at North Carolina State University.

She is a member of Omega Rho and INFORMS.

Jye-Chyi Lu received the Ph.D. degree in statistics from the University of Wisconsin at Madison in 1988.

In 1988, he joined the faculty of North Carolina State University, where he remained until 1999. Since 1999, he is a professor in the School of Industrial and Systems Engineering (ISyE) at the Georgia Institute of Technology. He is very active in promoting research, education, and extension-service programs with a focus on information systems engineering, e-business, e-design, and industrial statistics areas. He has published about 40 journal papers in these areas. He serves as an associate editor of IEEE TRANSACTIONS ON RELIABILITY.

James R. Wilson is Professor and Head of the Department of Industrial Engineering at North Carolina State University. From 1988 to 1992, he served as a departmental editor of *Management Science*; and from 1997 to 2001, he served as an area editor of *ACM Transactions on Modeling and Computer Simulation*. He has also held various offices in INFORMS-College on Simulation; and he currently serves as the corepresentative of that organization to the Board of Directors of the Winter Simulation Conference.

Dr. Wilson is a member of ASA, ACM, IIE, and INFORMS.