This article was downloaded by: *[Paynabar, Kamran]*
On: *13 February 2011*
Access details: *Access Details: [subscription number 933436582]*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

IIE Transactions
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713772245

Characterization of non-linear profiles variations using mixed-effect models and wavelets

Kamran Paynabar[a]; Jionghua (Judy) Jin[a]
[a] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA

Online publication date: 13 February 2011

To cite this Article Paynabar, Kamran and Jin (Judy) , Jionghua(2011) 'Characterization of non-linear profiles variations using mixed-effect models and wavelets', IIE Transactions, 43: 4, 275 — 290

To link to this Article: DOI: 10.1080/0740817X.2010.521807

URL: http://dx.doi.org/10.1080/0740817X.2010.521807

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

IIE Transactions (2011) **43**, 275–290
Copyright © "IIE"
ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2010.521807

Characterization of non-linear profiles variations using mixed-effect models and wavelets

KAMRAN PAYNABAR and JIONGHUA (JUDY) JIN*

Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109-2117, USA
E-mail: jhjin@umich.edu

Received November 2009 and accepted May 2010

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

There is an increasing research interest in the modeling and analysis of complex non-linear profiles using the wavelet transform. However, most existing modeling and analysis methods assume that the total inherent profile variations are mainly due to the noise within each profile. In many practical situations, however, the profile-to-profile variation is often too large to be neglected. In this article, a new method is proposed to model non-linear profile data variations using wavelets. For this purpose, a wavelet-based mixed-effect model is developed to consider both within- and between-profile variations. The utilization of wavelets not only simplifies the computational complexity of the mixed-effect model estimation but also facilitates the identification of the sources of the between-profile variations. In addition, a change-point model involving the likelihood ratio test is applied to ensure that the collected profiles used in the model estimation follow an identical distribution. Finally, the performance of the proposed model is evaluated using both Monte Carlo simulations and a case study.

Keywords: Change-point model, discrete wavelet analysis, functional data, likelihood ratio test, multiresolution analysis, process monitoring

1. Introduction

The rapid development of embedded sensing and computer technologies has led to online sensing and monitoring systems being increasingly used in manufacturing process control. In many practical situations, the sensor measurements are shown as time-dependent *functional data*, which are also called *profile data* or *waveform signals*. Some examples include the welding force responses recorded in resistance welding operations at uniform sampling time intervals (Chu *et al*., 2004), the tonnage signature signals measured in stamping processes, the equal crank angle sampling intervals (Jin and Shi, 1999), the vertical density profile of a particle board measured at fixed vertical depths (Walker and Wright, 2002), and the ram force signals used to press valve seats into engine heads in engine head assembly processes.

Most of the previous research reported in the literature has focused on linear profile monitoring; see, for example, Kang and Albin (2000), Kim *et al*. (2003), Mahmoud and Woodall (2004), Zou *et al*. (2006), Mahmoud *et al*. (2007), and Jensen *et al*. (2008). Non-linear profile modeling and monitoring has also generated increasing interest for the statistical process control of complicated profile data. For example, Gardner *et al*. (1997) utilized a smoothing spline to model non-linear profiles. Williams *et al*. (2007) used a four-parameter logistic regression and smoothing spline to model dose-response profiles to a drug. To monitor and distinguish out-of-control non-linear profiles in Phase I. Ding *et al*. (2006) considered each profile as a high-dimensional data set and applied Principal Component Analysis (PCA) and independent component analysis to reduce the dimension of the non-linear profile data while preserving the cluster structure of the profiles. Zou *et al*. (2008) used local linear smoothers to monitor non-linear profiles. Zou *et al*. (2009) applied the generalized likelihood ratio test to develop a monitoring procedure for non-linear profiles modeled by local linear kernel smoothing. To determine the control limit, they used the bootstrap method based on a few samples of in-control profiles.

In all of these studies it is assumed that the total variability of profiles can be modeled by random noises, which are typically assumed to be Normally Independently Distributed (NID). The random noises are mainly used to reflect the within-profile variation with a constant variance over all measurement points. In many practical situations, however, the variation among in-control profiles is too large to be solely handled by NID noises. Growth curves (Ramsay and Silverman, 1997), as shown in Fig. 1, are typical

*Corresponding author

0740-817X © 2011 "IIE"

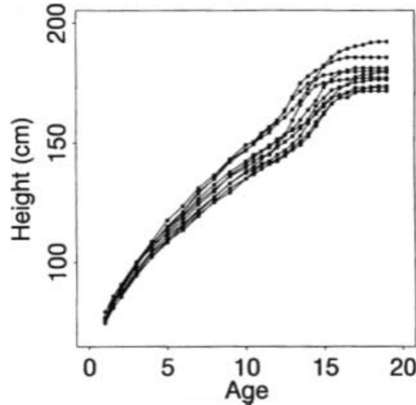

Fig. 1. Growth curves of 10 Swiss boys.

examples of such situations. Since the growth of one individual generally differs from that of others, a large amount of the profile variation is due to the person-to-person growth variability, which may not be fully explained by only the random noise within each person's growth curve.

As another example (to be discussed in detail in Section 7), the inserting forces of a pressing machine (as shown in Fig. 2) are used to press valve seat rings into an engine head. These force signals are continuously recorded during each cycle of repeated pressing operations. The overlapping multiple samples of signals collected at different cycles of in-control operations are shown in Fig. 2(a). Furthermore, to show the magnitude of random noises within each profile, an individual signal is also depicted in Fig. 2(b), in which the dotted points represent the actual measurements, and the solid line is profile fitted using the wavelet-based denoising method. As can be seen in Fig. 2(b), the within-profile variation obtained from the fitted model residuals e_t is much smaller than the part-to-part (i.e., curve-to-curve) variation shown in Fig. 2(a). In other words, a significant portion of the total inherent variation is reflected in the between-profile variation and is too large to be taken into account solely using random noises corresponding to the within-profile variation.

In practice, there are many causes for such inevitable between-profile variations, such as part-to-part variation, fixture or tooling tolerance, and/or process operation condition variations. The between-profile variation may affect the local profile shape differently at different segments of a profile. In contrast, the within-profile variation is mostly due to measurement errors and environmental disturbances, which independently and identically affect all observations of an entire profile. Therefore, characterization and estimation of between-profile and within-profile variations will not only help monitor the process more effectively but also provide us with a better understanding of the root causes of process variations, which can expedite further decision making for variation reduction and process improvement.

Recently, advanced modeling methods have been developed for variation modeling of non-linear profiles that consider both within-profile and between-profile variations via mixed-effect models (hereafter called *mixed models* for simplicity). For example, Mosesova *et al.* (2006) and Jensen and Birch (2009) developed a parametric mixed model by including a few model parameters as random effects to reflect the between-profile variation. Applying a parametric model, however, is not always achievable because it requires strong domain knowledge and major modeling

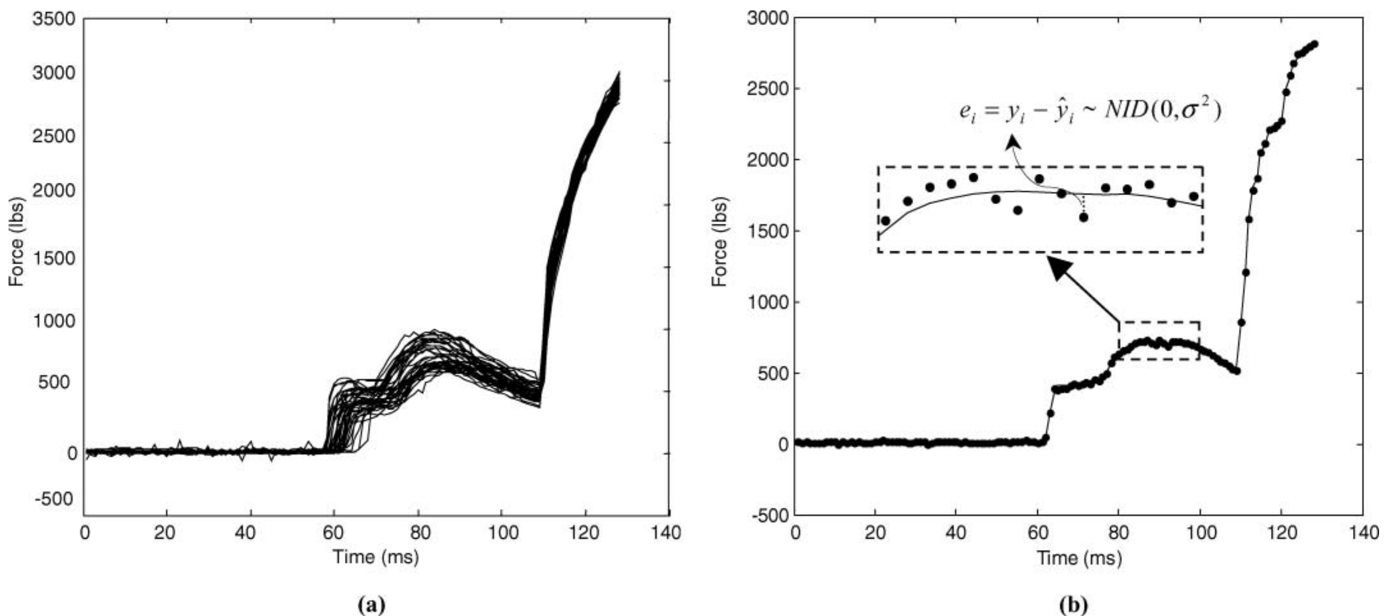

Fig. 2. Pressing force profile signals in a valve seat assembly operation: (a) overlapped multiple samples of profile signals and (b) one original profile (dot) and fitted profile (line).

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

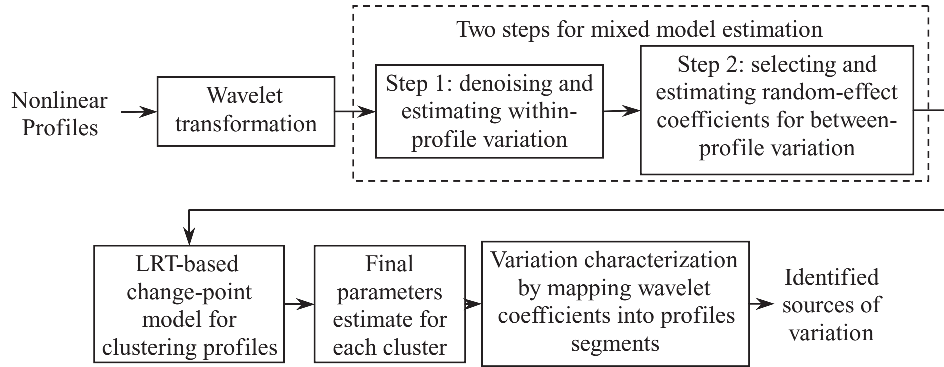

Fig. 3. Flow diagram of the proposed methodology.

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

efforts to identify an appropriate parametric model structure. To overcome this challenge, an alternative approach that uses non-parametric mixed models has attracted increasing attention. Mosesova *et al*. (2006) developed a mixed model by using a B-spline basis, whereas Shiau *et al*. (2009) used a random-effect B-spline model along with PCA to monitor non-linear profiles. Although splines and PCA are considered to be effective non-parametric approaches to the modeling and analysis of smooth non-linear profiles, they are inherently unable to model complicated non-linear profiles with local sharp jumps or non-differentiable points. The wavelet transform is a non-parametric alternative that can be effectively used for modeling non-linear profiles with sharp jumps.

One of the research challenges in using wavelets for process monitoring is determining how to select a low dimension of monitoring features from the large dimension of the wavelet coefficients. Jin and Shi (1999) developed a feature-preserving wavelet-based thresholding method that extracts monitoring features from complicated tonnage waveform signals and then constructed a Hotelling's T^2 control chart based on the unthresholded wavelet coefficients for stamping process monitoring (Jin and Shi, 2001). However, their method is limited to detecting profile changes that are reflected by the selected unthresholded wavelet coefficients. To overcome this problem, Jeong *et al*. (2006) presented an adaptive thresholding procedure that thresholds the wavelet coefficients of each incoming profile and updates the selected coefficients based on those that are unthresholded.

Chicken *et al*. (2009) developed a change-point model based on the likelihood ratio test, in which all wavelet coefficients are taken into account. They showed that monitoring wavelet coefficients is equivalent to evaluating the hypothesis that the non-centrality parameter of a chi-square distribution is equal to zero. They estimated the non-centrality parameter based on the unthresholded coefficients, which can reduce the variance of the estimator and consequently improve the performance of monitoring methods. They also showed that the change-point model outperforms the methods proposed by Jin and Shi (2001) and Jeong *et al.* (2006).

All of the previously mentioned papers on wavelet-based monitoring approaches only consider the within-profile variation in modeling the total profile variability. Very little research has been done on wavelet-based profile modeling or on monitoring methods that can account for both within-profile and between-profile variations. Therefore, the objective of this article is to develop a mixed model based on wavelets for the following two purposes: (i) to characterize non-linear profile variations by considering both between-profile and within-profile variations, thus going beyond the existing wavelet-based non-parametric modeling methods that account for only the within-profile variation; and (ii) to characterize both global and local segmental variation patterns by mapping scale/detail wavelet coefficients into profile segments, which goes beyond the existing methods (such as PCA or splines) that mainly characterize global variations for smooth non-linear profiles.

In this article, the wavelet transform is selected by considering its following three unique merits over other non-parametric approaches: (i) wavelet-based modeling is capable of fitting complicated non-linear profiles with sharp jumps and non-differentiable points; (ii) the multi-scale wavelet coefficients have the unique capability to separate the within-profile noises (at the high-frequency range) from the true profile signal (mainly at the low-frequency range), which can significantly simplify the computation for estimating the mixed model parameters; and (iii) the mapping relationship between the multiresolution wavelet coefficients and the local profile segments can facilitate the identification of the sources of the between-profile variation.

Implementing the proposed mixed model involves two critical research issues. The first is to ensure that the collected profile samples used for model estimation have an identical mixed model distribution. It is well known that combining samples from different distributions can lead to a large estimation error for the model parameters, thus resulting in a misleading model. For this purpose, a

change-point model, which is used to group profiles based on their distributions, is developed based on a Likelihood Ratio Test (LRT). The other critical issue is knowing how to reduce the computational effort to implement a mixed model. It is well known that the computational effort required to estimate model parameters increases with the number of random parameters. Demidenko (2004) recommended a method for constructing a model starting with one random coefficient and then adding more random parameters one-by-one if needed. However, this step-by-step exploration approach may not be very effective considering the large number of wavelet coefficients transformed from profile signals. Therefore, although wavelet transform is an effective approach for modeling complex non-linear profiles, implementing a wavelets-based mixed model is still challenging. This article also discusses how to effectively select a low dimension for the wavelet is random effects in the construction of the proposed mixed model, which can be well suited for characterizing the between-profile variation.

The remainder of this article is organized as follows. Section 2 provides an overview of the proposed methodology, and Section 3 gives a brief review of the wavelet transform used for profile signals. The development of the proposed wavelet-based mixed model is discussed in detail in Section 4. In Section 5, an LRT-based change-point model is developed to check that the selected profile samples follow an identical distribution. The performance of the proposed approach is examined through both Monte Carlo simulations and a case study in Sections 6 and 7, respectively. Finally, conclusions are drawn in Section 8.

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

2. Overview of the proposed methodology

A general framework of the proposed methodology is shown in Fig. 3. First, the measured non-linear profile data are transformed into the wavelet domain by using a selected wavelet basis. Then, a mixed model is developed on the wavelet coefficients, in which a two-step modeling approach is developed that reduces the computational complexity in the mixed model estimation. At the first step, a wavelet denoising thresholding is performed on each profile in order to separate within-profile noise from between-profile variation. At the second step, in order to reduce the dimension of the parameters in the mixed model, a few wavelet coefficients are selected to act as random effects. Then, the LRT-based change-point model is applied to check the collected profile samples follow an identical distribution. This result is used to group profile samples based on their distributions for further estimation of mixed model parameters. Finally, a mapping between the wavelet coefficients selected to acts as random effects and the profile segments is conducted to facilitate the identification of variation sources. The detailed analysis of each step will be elaborated in subsequent sections.

3. Wavelet transformation for profile signals

Suppose there are m available profiles, each of which consists of n pairs of (t, y) discrete observations that can be generally described by

$$\mathbf{y}_i = f_i(\mathbf{t}) + \boldsymbol{\varepsilon}_i \quad \text{for } i = 1, \ldots, m, \tag{1}$$

where $\mathbf{y}_i$ is a vector of the discrete response measurements of profile i, $f_i(.)$ is an unknown non-linear function of profile i, $\mathbf{t}$ is a vector consisting of equally spaced sampling time or distance data, and $\boldsymbol{\varepsilon}_i$ is a vector of NID noises with $\boldsymbol{\varepsilon}_i \sim MVN(\mathbf{0}, \sigma^2\mathbf{I})$ to represent the within-profile variation, where I is an $n \times n$ identity matrix.

The first step of the proposed procedure, as shown in Fig. 3, is to transform each profile into the wavelet domain. It is well known that any function g in $L^2(\Re)$, the square-integrable functions space, can be expressed by a wavelet series of the form of (Daubechies, 1992):

$$g(t) = \sum_{k \in Z} c_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{\infty} \sum_{k \in Z} d_{jk} \psi_{jk}(t)$$

Functions $\phi(.)$ and $\psi(.)$ are known as the father and mother wavelet basis, respectively. They are used to decompose function g into two parts corresponding to low-frequency (coarse) and high-frequency (detail). The multiresolution decomposition of the wavelets is performed using a set of orthonormal wavelets $\phi_{j_0 k}(t) = 2^{j_0/2}\phi(2^{j_0}x - k)$ and $\psi_{jk}(t) = 2^{j/2}\psi(2^j x - k)$, for any non-negative integer $j \geq j_0$. The decomposed coefficients $c_{j_0 k}$ and d_{jk} are called the *approximate* and *detail wavelet coefficients*, and they are determined by the inner product of g and the corresponding wavelet functions; i.e., $c_{j_0 k} = \langle g, \phi_{j_0 k}\rangle$, and $d_{jk} = \langle g, \psi_{jk}\rangle$, where $\langle\rangle$ represents the inner product operator.

When the number of discrete measurements (n) in each profile is dyadic, i.e., $n = 2^J$, where J is a positive integer number, a fast numerical algorithm called the Discrete Wavelet Transform (DWT) algorithm can be used to determine the wavelet coefficients (Mallat, 1999). The matrix form of DWT is represented as $\mathbf{z} = \mathbf{W}\mathbf{y}$, where $\mathbf{W}_{n \times n}$ is an orthogonal real matrix that depends on the selected orthogonal wavelet basis, the vector $\mathbf{z} = [\mathbf{c}_{J-l_0}, \mathbf{d}_{J-l_0}, \mathbf{d}_{J-l_0+1}, \ldots, \mathbf{d}_{J-1}]^{\mathrm{T}}$ represents all decomposed wavelet coefficients, and the superscript T denotes the transpose operator. The elements of $\mathbf{c}_{J-l_0}$ denote the approximate coefficient vector at the decomposition level $l_0 (1 \leq l_0 \leq J)$, $\mathbf{y}$ can be represented by $\mathbf{c}_J$, and $\mathbf{d}_{J-l} (l = 1, 2, \ldots, l_0)$ denotes the detail coefficient vector at the decomposition level l. More details about the wavelets transform can be found in Daubechies (1992) and Mallat (1999).

In this article, an orthogonal Haar transform is used for the discretized profile data $\mathbf{y}_i = f_i(\mathbf{t}) + \boldsymbol{\varepsilon}_i$ and the resulting wavelet coefficients are represented as $\mathbf{z}_i = \boldsymbol{\theta}_i + \tilde{\boldsymbol{\varepsilon}}_i$, where $\boldsymbol{\theta}_i = \mathbf{W} f_i(\mathbf{t})$ is a vector of the true wavelet coefficients transformed from the true profile function $f_i(\mathbf{t})$, $\mathbf{z}_i = \mathbf{W}\mathbf{y}_i$ is a vector of the empirical wavelet coefficients transformed

from noisy profile $\mathbf{y}_i$, and $\tilde{\boldsymbol{\varepsilon}}_i = \mathbf{W}\boldsymbol{\varepsilon}_i$ is a random noise vector in the wavelet domain with $\tilde{\boldsymbol{\varepsilon}}_i \sim MVN(0, \sigma^2\mathbf{I})$.

4. Mixed model for the wavelet coefficients

To consider the between-profile variation, a mixed model, in which a few wavelet coefficients are selected to act as random effects, is utilized. Davidian and Giltinan (1995), Pinheiro and Bates (2000), and Demidenko (2004) have provided a comprehensive introduction to mixed models. Commonly, parametric non-linear mixed models are constructed and the corresponding parameters are estimated based on numerical methods that are computationally expensive and after do not converge when the number of random effects is large. In contrast, the wavelet transform approach provides a multi-scale linear transformation that is able to separate a the within-profile variation from the true between-profile variation. This ability of the wavelet transform approach allows us to simplify the estimation of the mixed model parameters.

To implement the mixed model based on wavelet coefficients, let $\boldsymbol{\theta}_i = \boldsymbol{\mu} + \mathbf{b}_i$, where $\boldsymbol{\mu}$ is the vector of fixed effects common to all profiles, $\mathbf{b}_i$ is the vector of random effects of profile i with $\mathbf{b}_i \sim MVN(\mathbf{0}, \boldsymbol{\Lambda})$, and $\boldsymbol{\Lambda}$ is a positive-definite matrix that represents the covariance structure of the random effects. In this article, $\boldsymbol{\Lambda}$ is assumed to be a diagonal matrix, which implies that the random effects are uncorrelated. The reason for this assumption is that the Maximum Likelihood (ML) estimate of the covariance matrix can become negative-definite if the sample size is less than the dimension of the covariance matrix, especially in the case of a wavelet-based mixed model where the number of wavelet coefficients is often large. However, if the sample size is large enough, the method presented in this article is applicable to cases in which no restriction on the covariance matrix structure (such as a diagonality assumption) exists. Furthermore, such a diagonality assumption only restricts the covariance among the random effects of between-profile variations, which implies neither the independency nor a constant variance across different data points within a profile. We also assume that in the equation $\mathbf{z}_i = \boldsymbol{\mu} + \mathbf{b}_i + \tilde{\boldsymbol{\varepsilon}}_i$, $\mathbf{b}_i$ is independent of $\tilde{\boldsymbol{\varepsilon}}_i$. Based on this mixed model, the parameters of $\boldsymbol{\mu}$ and $\mathbf{b}_i$ can be effectively used to represent the profile mean and between-profile variation, respectively.

In order to construct a wavelet-based mixed model, a two-step modeling approach is proposed (Fig. 3). At the first step, the within-profile variation is estimated and removed through wavelet denoising thresholding of each sample of profile signals. As a result, the sample-to-sample variability of the remaining wavelet coefficients mainly reflects the between-profile variation. Therefore, at the second step, the between-profile variation is estimated based on the remaining wavelet coefficients using all collected samples of profile signals if they follow an identical mixed model distribution. The investigation of the distributions among samples will be discussed in Section 5. In order to effectively estimate the mixed model parameters and identify the major sources of variations, a data dimension reduction approach is further explored by selecting a few significant wavelet coefficients, that are sufficient to characterize the majority of the between-profile variation. The following two subsections will discuss the details of the proposed two-step estimation of the mixed model parameters on the wavelet coefficients.

4.1. *Characterizing within-profile variation and denoising*

In this subsection, wavelet-based denoising thresholding is utilized to estimate and remove the within-profile variation of noises $\tilde{\boldsymbol{\varepsilon}}_i$. As Mallat (1989) indicated, only a few wavelet coefficients contribute to the original true function of profiles. Therefore, denoising thresholding can be effectively applied to wavelet coefficients to remove the within-profile variation.

Since the white noises equally contribute to the wavelet coefficients, the soft thresholding approach introduced by Donoho and Johnstone (1995) is applied with the following thresholding rule:

$$\eta(\mathbf{z}_i;\hat{\sigma}_i, n) = \text{sign}(\mathbf{z}_i)\big(|\mathbf{z}_i| - \hat{\sigma}\sqrt{2\log n}\big) \\ I\big(|\mathbf{z}_i| > \hat{\sigma}\sqrt{2\log n}\big), \quad (2)$$

where $\eta(\cdot)$ is the soft thresholding function, $I(\cdot)$ represents an indicator function, sign$(\cdot)$ is the sign function, and $\hat{\sigma}$ is the estimated standard deviation of $\tilde{\boldsymbol{\varepsilon}}_i$ and is calculated using $\hat{\sigma}^2 = \sum_{i=1}^{m} \hat{\sigma}_i^2/m$, with $\hat{\sigma}_i = \text{median}(|\mathbf{d}_{i,J-1} - \text{median}(\mathbf{d}_{i,J-1})|)/0.6745$, where $\mathbf{d}_{i,J-1}$ are the detail coefficients at the lowest decomposition level (Donoho and Johnstone, 1995). Based on Equation (2), if a coefficient is less than the threshold of $\hat{\sigma}\sqrt{2\log n}$, it will shrink to zero. The denoised profile data and wavelet coefficients are denoted as $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{z}}$, respectively. It is known that the maximum of n independent Gaussian white noises cannot exceed $\sigma\sqrt{2\log n}$ when n is large (Fan, 1996); i.e., $\Pr(\max|\tilde{\boldsymbol{\varepsilon}}| \le \sigma\sqrt{2\log n}) \to 1$ as $n \to \infty$ with $\tilde{\boldsymbol{\varepsilon}} \sim NID(0, \sigma^2\mathbf{I})$. This implies that with a high probability, all noises shrink towards zero when n is large. Thus, the remaining coefficients approximate the wavelet coefficients of the true profile signals.

As shown in Appendix A, for each profile, the conditional variance of the denoised wavelet coefficients, given $\mathbf{b}_i$ denoted by $\sigma^2_{\tilde{z}_{ir}}$ $(r = 1, 2, .., n)$, can be calculated as follows:

$$\sigma^2_{\tilde{z}_{ir}} = \left\{\left(\mu^r_{z_{ir}}(\zeta) - \zeta\right)^2 + \left(\sigma^r_{z_{ir}}(\zeta)\right)^2\right\}\Phi\left(\frac{\mu_{z_{ir}} - \zeta}{\sigma}\right) \\ + \left\{\left(\mu^l_{z_{ir}}(-\zeta) + \zeta\right)^2 + \left(\sigma^l_{z_{ir}}(-\zeta)\right)^2\right\}\Phi\left(\frac{-\mu_{z_{ir}} - \zeta}{\sigma}\right) - \mu^2_{\tilde{z}_{ir}}, \quad (3)$$

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

where ζ is the threshold value $\sigma\sqrt{2\log n}$; $\mu^r_{z_{ir}}(\zeta)$ and $\mu^l_{z_{ir}}(\zeta)$ are the right and left truncated means of z_{ir} with truncation point ζ, respectively; $\sigma^r_{z_{ir}}(\zeta)$ and $\sigma^l_{z_{ir}}(\zeta)$ are the right and left truncated standard deviation of z_{ir} with truncation point ζ, respectively; $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution; $\mu_{\tilde{z}_{ir}}$ is the conditional mean of the denoised wavelet coefficients; and $\mu_{z_{ir}} = \mu_r + b_{ir}$ with μ_r and b_{ir} being the rth element of vectors $\boldsymbol{\mu}$ and $\mathbf{b}_i$, respectively. The detailed derivations of $\mu_{\tilde{z}_{ir}}$, $\mu^r_{z_{ir}}(\zeta)$, $\mu^l_{z_{ir}}(\zeta)$, $\sigma^r_{z_{ir}}(\zeta)$, and $\sigma^l_{z_{ir}}(\zeta)$ are given in Appendix A.

4.2. *Characterizing between-profile variation*

After denoising thresholding (Section 4.1), the true wavelet coefficients $\boldsymbol{\theta}_i$ can be estimated by the remaining denoised wavelet coefficients $\tilde{\mathbf{z}}$; that is,

$$\hat{\boldsymbol{\theta}}_i = \tilde{\mathbf{z}}_i \text{ and } \boldsymbol{\theta}_i = \boldsymbol{\mu} + \mathbf{b}_i \tag{4}$$

As a result, the remaining unthresholded empirical wavelet coefficients can be used for modeling the between-profile variation and estimating the remaining unknown parameters of the mixed model. The ML estimate of $\boldsymbol{\mu}$ is obtained by

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{m} \tilde{\mathbf{z}}_i / m. \tag{5}$$

Let $\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}$ represent the covariance matrix of the denoised wavelet coefficients. The ML estimate of $\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}$ is $\hat{\boldsymbol{\Sigma}}_{\tilde{\mathbf{z}}} = \text{diag}[\hat{S}_r]$; $\hat{S}_r = \sum_{i=1}^{m}(\tilde{z}_{ir} - \bar{\tilde{z}}_r)^2/m$ for $r = 1, 2, \ldots, n$, where $\tilde{z}_{ir}$ is the rth denoised wavelet coefficient of the ith profile and $\bar{\tilde{z}}_r$ is the sample mean of the rth denoised wavelet coefficient among all profiles. Furthermore, it is well known that $\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}} = \text{var}\{E[\tilde{\mathbf{z}}|\mathbf{b}_i]\} + E\{\text{var}[\tilde{\mathbf{z}}|\mathbf{b}_i]\}$. In other words, in the wavelet domain, the variance obtained from the denoised coefficients of all profiles, $\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}$, can be decomposed into the between-profile variation, $\text{var}\{E[\tilde{\mathbf{z}}|\mathbf{b}_i]\}$, and the estimate's variation caused by denoising each profile $E\{\text{var}[\tilde{\mathbf{z}}|\mathbf{b}_i]\}$. Based on Equation (4), the term $\text{var}\{\text{E}[\tilde{\mathbf{z}}|\mathbf{b}_i]\}$ can be estimated by $\hat{\boldsymbol{\Lambda}}$, where $\hat{\boldsymbol{\Lambda}}$ is the estimate of the covariance matrix of $\mathbf{b}_i$. Also, the term $E\{\text{var}[\tilde{\mathbf{z}}|\mathbf{b}_i]\}$ can be estimated by Equation (3) provided that $\mu_{z_{ir}}$ and σ are replaced with $\hat{\mu}_r$ and $\hat{\sigma}$, respectively, where $\hat{\mu}_r$ represents the rth element of $\hat{\boldsymbol{\mu}}$. If the estimated $E\{\text{var}[\tilde{\mathbf{z}}|\mathbf{b}_i]\}$ is denoted by $\hat{\boldsymbol{\Sigma}} = \text{diag}[\hat{\upsilon}_r]$ for $r = 1, 2, \ldots, n$, with $\hat{\upsilon}_r$ as the estimate of $\sigma^2_{\tilde{z}_r}$, we can obtain $\hat{S}_r = \hat{\lambda}_r + \hat{\upsilon}_r$. Therefore, $\hat{\lambda}_r$ can be estimated by $\hat{\lambda}_r = (\hat{S}_r - \hat{\upsilon}_r)I(\hat{S}_r - \hat{\upsilon}_r > 0)$, with $I(\cdot)$ being an indicator function.

4.2.1. *Selecting wavelet coefficients with significant random effects*

To reduce the dimensionality of random effects in the mixed model, only a small number of wavelet coefficients that have large and significant random effects should be selected. For this purpose, two rules are suggested in this article for selecting appropriate wavelet coefficients in the mixed model. Rule 1 is used to select wavelet coefficients with larger contributions to the between-profile variance. Rule 2 is used to further check that the selected coefficients have a significant random effect.

Rule 1: Wavelet coefficients with a larger variance are chosen such that the cumulative variance contribution of the selected random effects exceeds a threshold of $Q(0 \leq Q \leq 1)$. The contribution of each wavelet coefficient as a random effect in the total between-profile variation is sorted by

$$q_r = \frac{\hat{\lambda}_r}{\text{trace}(\hat{\boldsymbol{\Lambda}})}, \hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_n. \tag{6}$$

Thus, the set of the selected wavelet coefficients can be represented by

$$A = \left\{ \tilde{z}_r | r \leq \arg\min_k \left\{ \sum_{d=1}^{k} q_d \geq Q \right\} \right\}.$$

Justification of Rule 1: The reason we use this criterion is because we are often interested in identifying the root sources for only the top $100Q$ percent of total variations. To show how q_r is related to the between-profile variation, let $\boldsymbol{\Sigma}_{f(t)}$ denote the between-profile covariance matrix. As proved in Appendix B, the total between-profile variance, calculated by $\text{trace}(\boldsymbol{\Sigma}_{f(t)})$, can be explained by the total variance of the random effects. That is, $\text{trace}(\boldsymbol{\Sigma}_{f(t)}) = \text{trace}(\boldsymbol{\Lambda})$. Therefore, the selected wavelet coefficients would be sufficiently described as more than $100Q$ percent of the total between-profile variation. There is no fixed value for the threshold Q. The choice of Q is subject to the specific application. Generally, a very small value of Q may result in information loss about between-profile variation. On the other hand, as will be explained in the next section, a very large value of Q may lead to too many selected wavelet coefficients that may affect the performance of the model used for grouping profile data.

Rule 2: Each wavelet coefficient in A is tested to check that it is a significant random effect.

For this purpose, the following hypothesis is formulated:

$$\begin{cases} H_0 : \lambda_r = 0 \\ H_a : \lambda_r \neq 0 \end{cases}, \ r \in A.$$

The statistic $F_r = \hat{\upsilon}_r / \hat{S}_r, r \in A$ is used to test the hypothesis. If the calculated F_r is larger than the critical value F_C, then it can be implied that the random effect corresponding to wavelet coefficient r is significant. The critical value F_C is obtained by the $100(1-\alpha)$th, $0 < \alpha < 1$ percentile of the empirical distribution of F_r when there is no random effect in the model that is obtained by the Monte Carlo simulation.

Justification of Rule 2: As stated earlier, in the wavelet domain, the estimated variance obtained from the denoised coefficients of all profiles can be expressed by the estimated

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

random effect plus the estimate's variation caused by denoising each profile; i.e., $\hat{S}_r = \hat{\lambda}_r + \hat{\nu}_r$. Therefore, when the value of $\hat{\nu}_r$ is close to the value of $\hat{S}_r$, it indicates that the corresponding random effect is not significant.

After using these two rules, the set of selected wavelet coefficients with a significant random effect is denoted as Ω_S, and Ω_U represents the remaining unselected wavelet coefficients. Furthermore, if the distribution of coefficients in Ω_S is identical, these coefficients can be used to estimate the between-profile variation in the wavelet domain. By applying the Inverse Discrete Wavelet Transformation (IDWT), we can map the coefficients in Ω_S to the corresponding local segments of profiles, which have significant between-profile variability. The mapped segments of profiles contributing to the between-profile variation about the process can further facilitate the identification of sources of variations along with engineering knowledge. A case study employing this mapping will be presented in Section 7.

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

5. LRT-based change-point model for clustering profiles

In estimating a mixed model, it is essential to ensure that all the selected profile samples follow an identical distribution. Thus, a change-point model involving LRT (LRT-CP) is used to check this condition. To develop the LRT-CP model, it is assumed that if there is a change in the process it only affects the mean of the profile distribution and the covariance would remains constant. If LRT-CP finds more than one group of profiles, the parameters of each group are be re-estimated.

5.1. *Construction of monitoring features*

If all coefficients are directly considered in the LRT-CP, the test power would decrease due to the "curse of dimensionality." To reduce the number of variables involved in LRT-CP, we use the taxonomy of coefficients presented in Section 4.2.1. It is widely known that the sample covariance matrix is sensitive to changes in the sample mean; i.e., if the mean of the process changes globally or locally, the estimated sample covariance matrix is inflated. Therefore, the larger the sample variance of a wavelet coefficient, the more likely it is that the coefficient mean has changed. Thus, coefficients with the random effect in Ω_S are chosen as the monitoring features to be directly used in LRT-CP. In order to further include the information of unselected coefficients in Ω_U, these coefficients are added together as a combined monitoring feature, which is defined as $\gamma_i = \sum_{r\in R} \tilde{z}_{ir}, i = 1, 2, \ldots, m$ with $R = \{r; \tilde{z}_{ir} \in \Omega_U\}$. Since the elements of vector $\tilde{\mathbf{z}}$ are NID, it yields:

$$\gamma_i \sim N\left(\sum_{r\in R} \mu_r, \sum_{r\in R} S_r\right).$$

Therefore, if there is a change in the coefficients in Ω_U, it can be detected by γ_i. If we use $\tilde{\boldsymbol{\gamma}}_i$ to denote all the selected coefficients in Ω_S, the monitoring feature vector can be formed as $\boldsymbol{\gamma}_i = [\tilde{\boldsymbol{\gamma}}_i^T \ \gamma_i]^T$ with $\boldsymbol{\gamma}_i \sim MVN(\boldsymbol{\mu}_\gamma, \boldsymbol{\Lambda}_\gamma)$, where $\boldsymbol{\mu}_\gamma$ and $\boldsymbol{\Lambda}_\gamma$ represent the mean vector and covariance matrix of the monitoring feature, respectively.

5.2. *LRT-based change-point model*

Various change-point models have been developed and applied with great successes to analyze and group collected observations. For example, an LRT-based change-point model has been developed for univariate normally distributed data (Sullivan and Woodall, 1996) and multivariate normally distributed data (Worsely, 1979; Sullivan and Woodall, 2000; Zamba and Hawkins, 2006). Sullivan (2002) developed a change-point model that was based on a clustering approach. Recent research on process monitoring using a change-point includes a method for monitoring linear profiles (Zou *et al.*, 2006; Mahmoud *et al.*, 2007) and for monitoring non-linear profiles using wavelet coefficients in Phase II control charts (Chicken *et al.*, 2009). In addition to the change-point models, there is another set of methods for grouping observations based on the clustering approach (see the work by Fraley and Raftery (1998), Kothari and Pitts (1999), Ertoz *et al.* (2003), and Zhang and Albin (2007)). The LRT-based change-point model, however, is preferred in this article for two reasons. First, for the purpose of grouping sequential observations, the change-point model directly utilizes information related to the data sequence order. Second, since the observations are assumed to follow a normal distribution in this article, a parametric model is preferred to a non-parametric model. The LRT-based change-point model is a commonly used parametric model-based approach. We extend the existing change-point model by including wavelet coefficients with random effects in order to account for the between-profile variability.

Suppose that the mean of the profiles changes at an unknown time τ. Assuming that the covariance matrix of non-linear profiles is constant, the distribution of $\boldsymbol{\gamma}_i$ can be written as

$$\boldsymbol{\gamma}_i \sim \begin{cases} MVN(\boldsymbol{\mu}_\gamma^0, \boldsymbol{\Lambda}_\gamma), & i \le \tau, \\ MVN(\boldsymbol{\mu}_\gamma^1, \boldsymbol{\Lambda}_\gamma), & i > \tau, \end{cases} \quad (7)$$

where $\boldsymbol{\mu}_\gamma^0$ and $\boldsymbol{\mu}_\gamma^1$ represent the mean vectors before and after the change, respectively. If all profiles are from an identical distribution, then τ does not exist and $\boldsymbol{\mu}_\gamma^0 = \boldsymbol{\mu}_\gamma^1$. Therefore, in order to check whether the collected profile data have an identical distribution, the following hypothesis test can be evaluated:

$$\begin{cases} H_0 : \underset{i\le\tau}{E}[\boldsymbol{\gamma}_i] = \underset{i>\tau}{E}[\boldsymbol{\gamma}_i] = \boldsymbol{\mu}_\gamma^0 \\ H_a : \underset{i\le\tau}{E}[\boldsymbol{\gamma}_i] \neq \underset{i>\tau}{E}[\boldsymbol{\gamma}_i] = \boldsymbol{\mu}_\gamma^1 \end{cases}, \quad \tau = 1, 2, \ldots, m, \quad (8)$$

where $E[\cdot]$ is the expectation. An LRT is utilized to evaluate the hypothesis test. As derived in Appendix C, the log likelihood ratio can be expressed as

$$\Gamma(\tau) = \frac{\tau(m-\tau)}{m} \sum_{i=\tau+1}^{m} \left(\hat{\mu}_{\gamma}^{1} - \hat{\mu}_{\gamma}^{0}\right)^{\mathrm{T}} \hat{\Lambda}_{\gamma}^{-1} \left(\hat{\mu}_{\gamma}^{1} - \hat{\mu}_{\gamma}^{0}\right), \quad (9)$$

where

$$\hat{\mu}_{\gamma}^{0} = \sum_{i=1}^{\tau} \gamma_i/\tau, \; \hat{\mu}_{\gamma}^{1} = \sum_{i=\tau+1}^{m} \gamma_i/(m-\tau),$$

and

$$\hat{\Lambda}_{\gamma} = \frac{\left\{\sum_{i=1}^{\tau}(\gamma_i - \hat{\mu}_{\gamma}^{0})(\gamma_i - \hat{\mu}_{\gamma}^{0})^{\mathrm{T}} + \sum_{i=\tau+1}^{m}(\gamma_i - \hat{\mu}_{\gamma}^{1})(\gamma_i - \hat{\mu}_{\gamma}^{1})^{T}\right\}}{(m-2)}.$$

The values of $\Gamma(\tau)$; $(\tau = 1, 2, \ldots, m)$ are compared with a limit of L. If $\Gamma(\tau) < L$ for $\tau = 1, 2, \ldots m$, all profiles are from an identical distribution. Otherwise, they can be categorized into different clusters and the value of τ maximizing $\Gamma(\tau)$ is an estimate of the change-point; i.e., $\hat{\tau} = \arg\max_{\tau=1,2,\ldots,m}\Gamma(\tau)$. If the LRT-CP detects a change, the profile samples can be separated into two groups based on the estimated change-point. Then, LRT-CP should be applied to each group to check whether more clusters exist in each group. Using this approach, one can categorize multiple change points and clusters. Additionally, knowing the estimated time at which the process changed could help process engineers effectively detect the root cause(s) of the change and identify the corresponding source(s) of variations. It should be noted that the LRT-CP model is often used to detect single and/or multiple sustained shifts in historical profile data. If one is interested in identifying outlier profiles, other methods such as multivariate T^2 control charts with a robust estimator of the covariance matrix (Vargas, 2003) can be utilized to examine the selected wavelet coefficients with random effect and detect outliers.

The limit L can be determined based on the desired Type-I error (α). In this article, L is determined via simulations since the values of $\Gamma(\tau)$ are not independent.

6. Performance evaluation using simulations

In this section, the performance of the proposed approach is evaluated using Monte Carlo simulations. This is accomplished in two stages. First, the performance of the proposed mixed LRT-CP model is assessed under different change scenarios and compared with another wavelet-based method recently proposed by Chicken *et al.* (2009), in which the between-profile variation is not considered. As mentioned earlier, if the collected profiles do not follow the same mixed model, the estimation results are misleading. In this article, two criteria are used for the performance evaluation: probability of detecting a change in the mean

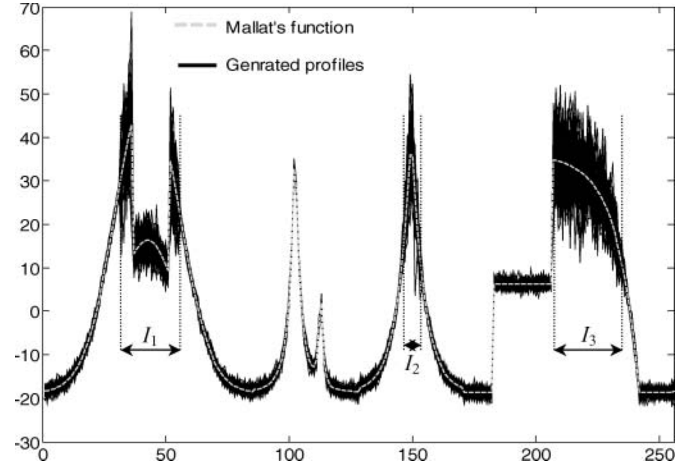

Fig. 4. Mallat's function and randomly generated profiles.

$(1-\beta)$ and the average estimation of change time ($\bar{\hat{\tau}}$). In the other stage, the accuracy of the proposed approach for estimating the between-profile variation is checked. This is evaluated based on the ratio of the standard deviations of the estimated to actual random effects, denoted by $\hat{\lambda}/\lambda$. Since the case of multiple changes can be boiled down to a single change case, only a single change is studied here.

To simulate profiles, the popularly used piecewise smooth function of Mallat (1999) is utilized to generate simulated profiles, as shown in Fig. 4. This simulated function is a complicated function with several non-differentiable points that cannot be easily modeled by parametric models or other non-parametric models such as splines. Additionally, Chicken *et al.* (2009) used this function to evaluate the performance of their monitoring procedure. Also, it is assumed that the between-profile variations only occur at three segments: $I_1 = [32,55]$, $I_2 = [146,153]$, and $I_3 = [207,236]$ as shown in Fig. 4, which cover 62 data points, or 24% of the entire profile. In those segments, each response $y_{ir}, i = 1, 2, \ldots m, r = 1, 2, \ldots, n$ is generated based on $y_{ir} = f(t_r) + b_{ir}^{f}$, where $f(t_r)$ is the value of Mallat's function at t_r, and $b_{ir}^{f} \sim N(0, s^2 f^2(t_r))$ is the random effect with a coefficient of variation s, which sets the standard deviation of each response y_{ir} to be proportional to the value of its mean $f(t_r)$. Finally, to include the within-profile variation, NID noises with a mean of zero and variance $\sigma^2 = 1$ are added to y_{ir}. Figure 4 shows the 50 simulated profiles with $n = 256$, $t \in [1, 256]$, and $s = 0.2$. An alternative procedure that can be used to simulate random profiles is to generate both within- and between-profile variations on the wavelet coefficients and then transform the coefficients back to the original domain using IDWT. In this article, we prefer to use the first procedure since we are interested in modeling and characterizing variations of original profiles. For the comparison purpose, similar to Chicken *et al.* (2009), the Haar basis with the complete

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

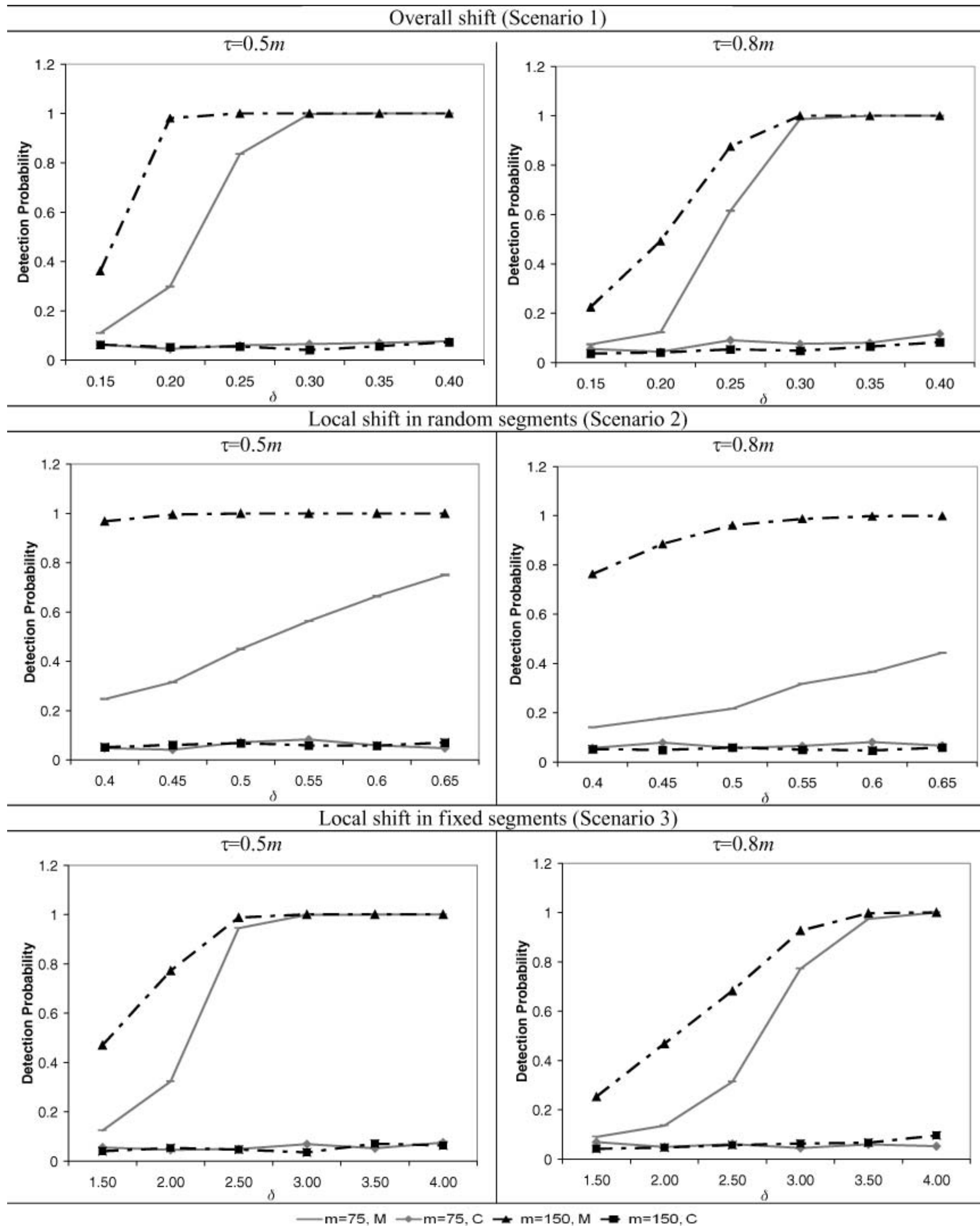

Fig. 5. Detection probability of the proposed LRT-CP model (M) and (C) that of the method proposed by Chicken *et al.* (2009).

decomposition (i.e., eight levels of decomposition) is used to develop the wavelet-based mixed model.

In order to assess the capability of the methods in detecting different types of profile changes, three change scenarios with different magnitudes are examined.

Scenario 1: Overall mean change, where the whole profile is shifted vertically; that is, $\boldsymbol{\mu}_\gamma^1 = \boldsymbol{\mu}_\gamma^0 + \delta(s\boldsymbol{\mu}_\gamma^0 + \sigma)$.

Scenario 2: Local mean change in the segments of [41, 46] and [208, 215] with $\boldsymbol{\mu}_\gamma^1 = \boldsymbol{\mu}_\gamma^0 + \delta(s\boldsymbol{\mu}_\gamma^0 + \sigma)$. These two segments contain the between-profile variation.

Scenario 3: Local mean change in the segments of [6,22], [89,106], and [129,145] with $\boldsymbol{\mu}_\gamma^1 = \boldsymbol{\mu}_\gamma^0 + \delta\sigma$. These three segments are not comprised of between-profile variations.

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

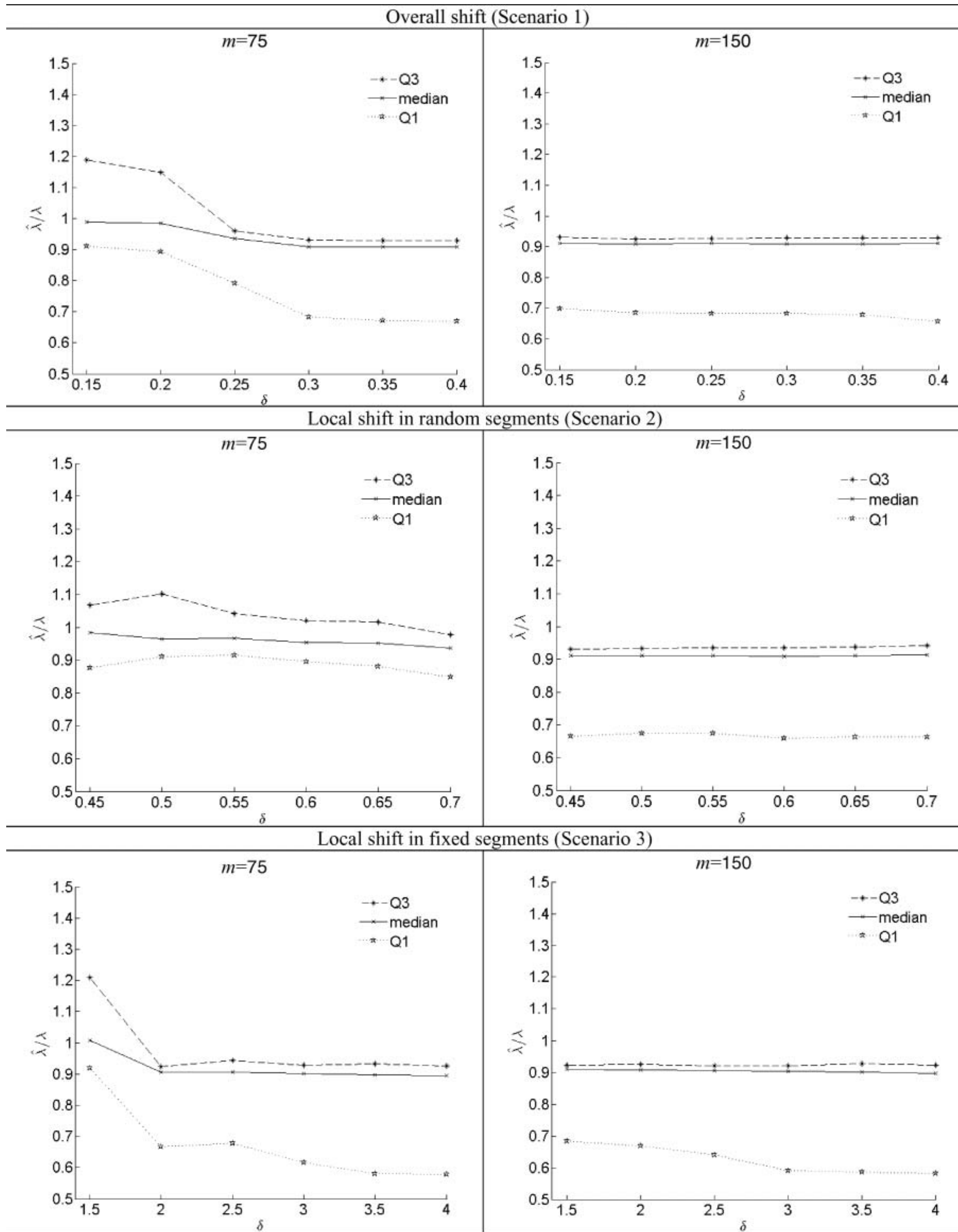

Fig. 6. Q_1, median, and Q_3 of $\hat{\lambda}_i/\lambda_i$ under different shift scenarios.

Furthermore, to investigate the performance sensitivity to the parameters of m and τ, the simulations are carried out for $m = 75$ and 150; $\tau = 0.5m$ and $0.8m$ with 1000 replications. Also, $Q = 0.80$ is used for all simulation scenarios.

To compare the performance of the methods, the limit L is chosen so that the estimated probability of false signal (α) is approximately equal to 0.05. In the proposed method, the probability of the signal is estimated by the proportion of simulation runs where at least one of $\Gamma(\tau)$, $\tau = 1, 2, \ldots, m$ is plotted beyond L. Thus, the 95th percentile of $\max \Gamma(\tau)$, $\tau = 1, 2, \ldots, m$ obtained from 1000 simulation runs is chosen as the limit L. It is clear that

Table 1. L values for different p and m with $\alpha = 0.05$

	p									
m	*5*	*10*	*15*	*20*	*25*	*30*	*35*	*40*	*45*	*50*
75	25.81	41.38	58.70	80.12	108.22	140.97	187.64	254.12	351.40	529.93
100	24.73	38.77	52.08	67.62	86.72	106.90	130.81	164.97	206.79	256.82
125	24.40	36.27	49.31	62.39	78.12	93.92	111.80	133.16	159.12	186.31
150	23.84	35.55	47.99	58.90	73.84	86.49	102.60	118.03	139.07	157.97
175	23.16	34.94	45.50	56.54	70.51	82.54	96.46	110.71	125.19	140.76
200	23.30	34.64	45.24	55.81	66.10	78.13	89.66	104.15	118.21	132.76
250	23.70	33.56	42.55	53.42	63.68	73.70	86.45	98.34	108.15	120.35
300	22.16	33.58	43.73	52.53	62.32	72.05	82.87	94.42	105.10	115.19
350	22.96	32.86	43.64	52.12	61.99	70.89	79.92	90.93	101.09	112.71
400	23.67	33.30	42.65	51.74	60.76	70.47	78.77	89.99	97.83	106.43
450	23.21	33.94	41.87	50.67	59.76	69.56	77.73	88.48	97.94	105.58
500	23.03	32.99	42.43	51.49	59.77	67.41	76.83	87.27	96.02	104.79

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

for a specific α, the value of L depends on the number of collected profiles m and the dimension of $\boldsymbol{\gamma}_i$, denoted as p. The estimated values of L based on 1000 simulations for $\alpha = 0.05$ under different m and p are provided in Table 1.

The estimated detection probabilities for different change scenarios are shown in Fig. 5. As can be seen from these figures, under all scenarios and different values for the parameters m and τ, the proposed mixed LRT-CP method outperforms The method of Chicken *et al.* (2009). This is because the LRT-CP model accounts for the between-profile variation, while the other does not do so. Moreover, the detection probability of LRT-CP is improved as the number of sampled profiles (m) increases. Clearly, by increasing m, the estimation of parameters in the mixed model becomes more precise, thus resulting in this improvement. Also, the performance of LRT-CP is better when $\tau = 0.5m$ than when $\tau = 0.8m$. This is because in the case of $\tau = 0.5m$, there are equal samples available in each group to estimate the parameters, which leads to a better estimation for the two groups on average.

Furthermore, the performance of change-point estimation was also studied. The mean value and the standard error (given in the parentheses) for each estimated τ by using the LRT-CP method are shown in Tables 2 to 4. As an example, under Scenario 2 with $m = 150$, $\tau = 75$ and $\delta = 0.60$, a mean of the estimated change-point is 74.97 with a standard error of 0.07. Similar to the effect of τ and m on the detection performance of LRT-CP, the performance of the change-point estimator also improves when m increases and/or when $\tau = 0.5m$. Generally, the higher the detection probability of LRT-CP, the more accurate and precise is the change-point estimator. Based on Tables 2 to 4, it can also be seen that the absolute biases of the estimated change points are all less than five, except for in a few cases (where values are underlined). Therefore, the estimation performance of change points by using LRT-CP is quite reasonable.

Furthermore, in order to investigate the performance of the proposed approach in estimating the between-profile variation, the ratio of $\hat{\lambda}_r/\lambda_r$, $r = 1, 2, \ldots, 62$ was calculated for every point within all three random-effect

Table 2. Overall shift (Scenario 1)

		δ					
m	τ	*0.15*	*0.20*	*0.25*	*0.30*	*0.35*	*0.40*
75	$0.5m = 38$	40.32	39.03	38.43	38.24	38.10	38.02
		(0.59)	(0.36)	(0.17)	(0.06)	(0.02)	(0.01)
	$0.8m = 60$	43.86	52.50	58.98	60.18	60.07	60.02
		(0.72)	(0.61)	(0.25)	(0.03)	(0.01)	(0.01)
150	$0.5m = 75$	76.43	76.35	75.61	75.19	75.04	75.00
		(0.79)	(0.21)	(0.06)	(0.02)	(0.01)	(0.00)
	$0.8m = 120$	104.79	111.83	118.87	120.12	120.05	120.00
		(1.18)	(0.85)	(0.35)	(0.01)	(0.01)	(0.01)

Table 3. Local shift in random segments (Scenario 2)

		δ					
m	τ	*0.40*	*0.45*	*0.50*	*0.55*	*0.60*	*0.65*
75	$0.5m = 38$	37.73	37.41	38.34	37.46	37.75	37.86
		(0.44)	(0.35)	(0.31)	(0.27)	(0.24)	(0.21)
	$0.8m = 60$	52.21	53.11	56.11	55.81	56.33	57.13
		(0.56)	(0.52)	(0.41)	(0.42)	(0.4)	(0.32)
150	$0.5m = 75$	74.77	75.00	75.25	74.97	75.04	75.06
		(0.25)	(0.17)	(0.12)	(0.08)	(0.07)	(0.06)
	$0.8m = 120$	117.19	118.60	119.25	119.48	119.86	120.03
		(0.51)	(0.35)	(0.25)	(0.23)	(0.10)	(0.08)

Table 4. Local shift in fixed segments (Scenario 3)

m	τ	δ 1.50	2.00	2.50	3.00	3. 50	4.00
75	$0.5m = 38$	37.58	39.11	38.17	38.03	38.00	38.00
		(0.57)	(0.39)	(0.1)	(0.02)	(0.01)	(0.002)
	$0.8m = 60$	40.03	51.63	57.92	59.19	59.85	59.99
		(0.79)	(0.55)	(0.31)	(0.19)	(0.10)	(0.01)
150	$0.5m = 75$	74.39	75.27	75.20	75.03	75.00	75.00
		(0.70)	(0.46)	(0.11)	(0.01)	(0.00)	(0.00)
	$0.8m = 120$	106.46	114.85	116.68	119.48	120.01	120.00
		(1.15)	(0.80)	(0.58)	(0.21)	(0.03)	(0.004)

segments $[I_1, I_2, I_3]$. In order to show the estimation performance, we use the median of all $\hat{\lambda}_r/\lambda_r$ ratios to assess the average estimation performance and use the third and first quartile values of $\hat{\lambda}_r/\lambda_r$, respectively denoted by Q_3 and Q_1, to show the estimation uncertainty. The results under different change scenarios with $\tau = 0.5m$ are presented in Fig. 6. Values of $\hat{\lambda}_r/\lambda_r$ greater and less than one imply overestimation and underestimation, respectively, while values close to one indicate the unbiased estimates. From Fig. 6, it is clear that the medians of $\hat{\lambda}_r/\lambda_r$ are close to one, which shows that they have a very good average estimation performance. In the case of $m = 150$, the estimates are more stable than those for $m = 75$ across all δ values. Therefore, the standard deviation estimates become steadier when m is large. Moreover, when $m = 75$, the accuracy of the estimates becomes more stable as δ increases. The reason for this is that the detection probabilities of change points increase as δ increases, thus resulting in the better change-point estimates.

From Fig. 6, it can be observed that the variances of random effects can sometimes be overestimated under a scenario with a small shift, which is due to the poor performance in detecting and estimating a change-point with a small shift. An increase in the shift magnitude δ results in an improvement in the estimation performance and the estimated variance approaches the true variance (i.e., the ratio of $\hat{\lambda}_i/\lambda_i$ is closer to one, as shown in Fig. 6). However, the stable value of $\hat{\lambda}_i/\lambda_i$ is generally less than one because in our mixed model, only a subset of random effect coefficients is selected to explain the $100Q$ percent the total between-profile variation. Therefore, without the effect of the estimation error of change points, the estimated variance should be less than the true total variance.

In short, the simulation results show that the proposed methodology has a reasonable performance in classifying different groups of profiles as well as in characterizing the variance of each group of profiles.

7. Case study

In this section, the proposed methodology is applied to real-world profile data, which were collected for a valve seat pressing operation that is performed as part of an engine head assembly process. At every cycle of the pressing operation, a valve seat is pressed into a seat counterbore pocket of a cylinder head, which generates one cycle of press force signals, which becomes a sample of profile data. Pictures of the engine head (upper left), valve seat pocket (lower left), and pressing machine (right) are shown in Fig. 7. In this process, one of the important quality characteristics is the gap between the seat bottom and the pocket. However, there is no automatic sensing technology for directly measuring this gap during production. Another aspect to take into account is that the product quality is very sensitive to the pressing force on the ram, which can be measured online by load sensors installed on the pressing machine. Therefore, pressing force signals are often used for process

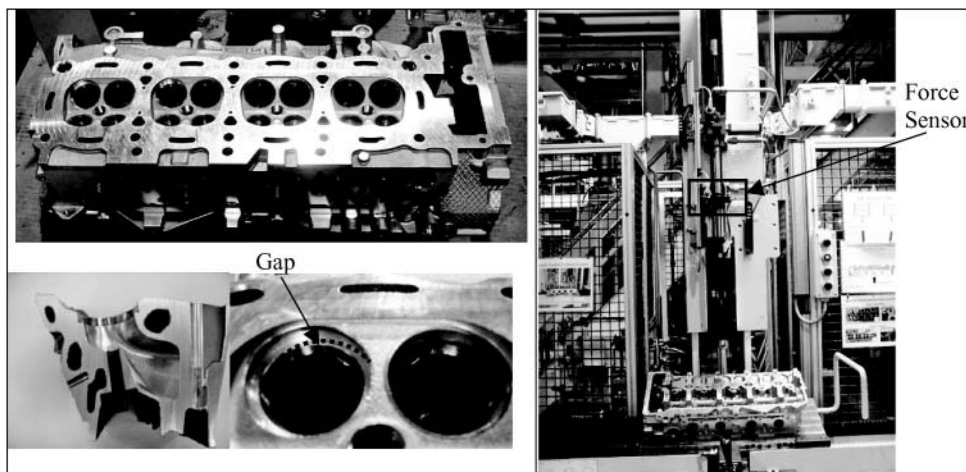

Fig. 7. Engine head (upper left panel), cross-section view of valve seat pocket and gap between valve seat and pocket (lower left panel), valve seat assembly process (right panel).

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

Table 5. Summary information of the fitted mixed model

Segment	*Corresponding profile observations*	*Corresponding wavelet coefficients*	*Wavelet coefficients mean (fixed effect)*	*Wavelet random effect variances*	*Average of segmental between-profile variance*
a	$y_{57} - y_{73}$	$c_{5,16}$	5.736×10^2	1.1163×10^5	1.7026×10^4
		$c_{5,19}$	1.076×10^3	5.9810×10^4	
		$c_{5,17}$	8.129×10^2	4.8550×10^4	
		$c_{5,15}$	1.238×10^2	4.1900×10^4	
		$c_{5,18}$	8.951×10^2	2.4110×10^4	
		$d_{5,19}$	-8.410×10^1	2.0460×10^4	
b	$y_{74} - y_{92}$	$c_{5,20}$	1.362×10^3	6.8650×10^4	1.1852×10^4
		$c_{5,21}$	1.525×10^3	5.0160×10^4	
		$c_{5,22}$	1.538×10^3	3.4070×10^4	
		$c_{5,23}$	1.464×10^3	2.4350×10^4	
		$c_{5,19}$	1.076×10^3	5.9810×10^4	
c	$y_{109} - y_{116}$	$c_{5,28}$	1.998×10^3	3.8790×10^4	7.7238×10^3
		$c_{5,29}$	3.651×10^3	2.3000×10^4	

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

control (i.e., reduction in the variation in pressing force signals will lead to the improvement of product quality). In this case study, 50 force profiles were collected for process variation evaluation by the following analysis (as shown in Fig. 8).

A mixed model was developed to characterize the process variation using the methodology given in Fig. 3. Based on Section 4.2.1, $Q = 0.80$ was used as the threshold for selecting the wavelet coefficients to create random effects in Ω_S. The LRT-CP model presented in Section 5.2 was also used to check whether all 50 profiles followed an identical distribution. The LRT-CP result indicated a change-point of $\hat{\tau} = 42$; thus these 50 profiles are clustered into two groups. In Fig. 8, the profiles corresponding to clusters 1 and 2 are plotted with a solid line and a dashed line, respectively. LRT-CP was also applied to the profiles within each group, but it did not find a new change-point within each group of profiles. Since the number of profiles in cluster 2 was not large enough, only cluster 1 was used for further identification of the critical segments with a large variability. A similar method could be applied to cluster 2 once more profiles are collected.

In order to identify sources in variations in cluster 1, a mixed model was constructed. Based on the fitted mixed model, the estimated within-profile variance ($\hat{\sigma}^2$) was equal to 82.28. The wavelet coefficients used to create random effects in Ω_S were $[c_{5,16}, c_{5,20}, c_{5,19}, c_{5,21}, c_{5,17}, c_{5,15}, c_{5,28}, c_{5,22}, c_{5,23}, c_{5,18}, c_{5,29}, d_{5,15}]^T$ in descending order of their variances. Then, the mapping of these coefficients in Ω_S to the associated profile segments was conducted using IDWT.

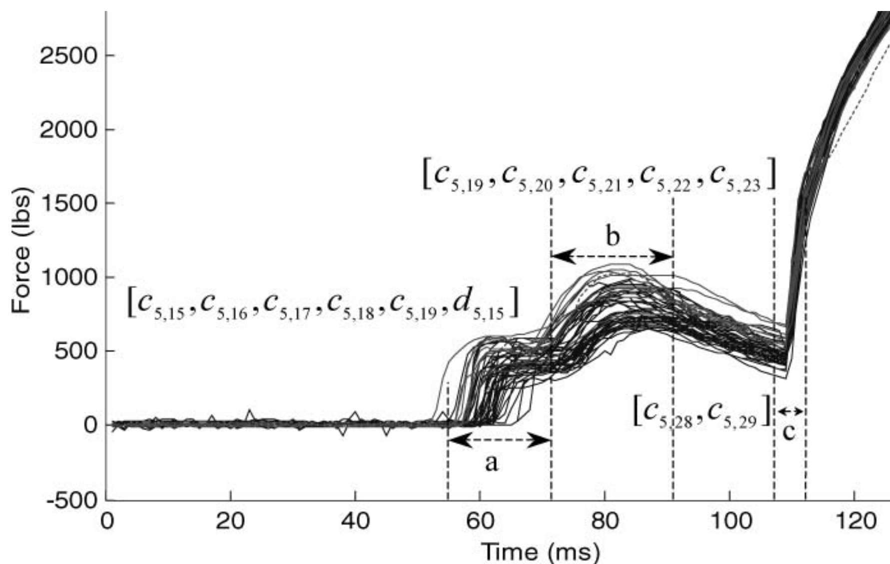

Fig. 8. Force against time profiles.

These mapped segments, along with their corresponding coefficients, are shown in Fig. 8. There are three segments that contribut 80% of the total between-profile variation. Table 5 summarizes the obtained information for each segment and the corresponding between-profile variances. In Table 5, we see that the between-profile variance is much larger than the within-profile variance, which implies that the sources of variations causing between-profile variations are more important for process improvement. The wavelet coefficients chosen to create random effects along with the estimated mean and variance (reported in Table 5) can further serve as a basis to implement control charts for process monitoring.

Furthermore, based on the segments obtained from IDWT and engineering knowledge, the sources contributing to the between-profile variations can be identified. The variation in "segment a" is due to the position variations of engine head surfaces. This source is mainly related to the variation of initial contacting points induced by the variation of an engine head's pocket depth due to previous manufacturing stages. The clearance tolerance between the valve seat and the seat packet is the major source of variations for "segment b." The pressure variation of the assembly machine could be causing the force signal variation in "segment c." The first two sources of the variations are considered part-to-part variations, but the source for "segment a" is related to the process variation at previous manufacturing stages, while the source for "segment b" is related to the current assembly process variation. The average between-profile variance for each segment can also be obtained by adding up the variance of wavelet coefficients in each segment and dividing the sum by the length of the segment. These values, reported in Table 5, can be used to prioritize further actions for variation reduction and process improvement.

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

8. Conclusions and future research

In most real applications, the total inherent variation of profiles often consists of both within-profile and between-profile variations. Characterizing both types of variations and identifying their variation sources can be used to guide proactive action to improve processes. In this article, a mixed model was developed that is based on wavelet coefficients and is capable of modeling both types of variations for complex non-linear profiles. In constructing the mixed model, an LRT-based change-point model was utilized in order to check that the collected profiles used into model estimation followed the same distribution. To reduce the computation complexity, a two-step estimation procedure was developed for mixed model estimation. Furthermore, a method for effective selection of monitoring features was proposed in order to improve LRT-CP performance. Monte Carlo simulations and a case study were conducted to demonstrate the effectiveness of the proposed approach.

In this article, it was assumed that within-profile noises are IND. If the independency assumption does not hold, the denosing procedure used for removing the within-profile variation may not perform properly since the estimated within-profile variance would be highly biased. Developing wavelet-based mixed models for modeling non-linear profile data in the presence of autocorrelated noises will be an interesting topic for future research. Furthermore, the extension of the developed mixed model to online process monitoring and diagnosis would be an interesting development.

Acknowledgements

The authors gratefully acknowledge the financial support of the NSF under grant DMI F0541750 the Engineering Research Center for Reconfigurable Manufacturing Systems (NSF grant EEC-9529125) at the University of Michigan and General Motors Corporation. The authors would also like to thank the editor and referees for their insightful comments.

References

Chicken, E., Pignatiello, J.J. and Simpson, J.R. (2009) Statistical process monitoring of nonlinear profiles using wavelets. *Journal of Quality Technology,* **41**, 198–212.

Chu, Y.X., Hu, S.J., Hou, W.K., Wang, P.C. and Marin, S.P. (2004) Signature analysis for quality monitoring in short circuit GMAW. *Welding Journal*, **83**, 336s–343s.

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

Davidian, M. and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurements Data*, Chapman and Hall, London, UK.

Demidenko, E. (2004) *Mixed Models: Theory and Applications*, Wiley, New York, NY.

Ding, Y., Zeng, L. and Zhou, S. (2006) Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology*, **38**, 199–216.

Donoho, D.L. and Johnstone, I.M. (1995) Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, **90**, 1200–1224.

Ertoz, L., Steinbach, M. and Kumar, V. (2003) Finding clusters of different sizes, shapes and densities in noisy high dimensional data. Presented at the SIAM International Conference on Data Mining, San Francisco, CA, May 2003.

Fan, J. (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, **91**, 674–688.

Fraley, C. and Raftery, A. (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, **41**, 578–588.

Gardner, M., Lu, J., Gyurcsik, R., Wortman, J., Hornung, B., Heinisch, H., Rying, E., Rao, S., Davis, J. and Mozumder, P. (1997) Equipment fault detection using spatial signatures. *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C (Manufacturing)*, **20**, 294–303.

Jensen, W.A. and Birch, J.B. (2009) Profile monitoring via nonlinear mixed models. *Journal of Quality Technology*, **41**, 18–34.

Jensen, W.A., Birch, J.B. and Woodall, W.H. (2008) Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology*, **40**, 167–183.

Jeong, M.K., Lu, J.C. and Wang, N. (2006) Wavelet-based SPC procedure for complicated functional data. *International Journal of Production Research*, **44**, 729–744.

Jin, J. and Shi, J. (1999) Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*, **41**, 327–339.

Jin, J. and Shi, J. (2001) Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. *Journal of Intelligent Manufacturing*, **12**, 140–145.

Johnson, N.L. and Kotz, S. (1970) *Distributions in Statistics. Continuous Univariate Distributions—Volume 1*. Wiley, New York, NY.

Kang, L. and Albin, S.L. (2000) On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, **32**, 418–426.

Kim, K., Mahmoud, M.A. and Woodall, W.H. (2003) On the monitoring of linear profiles. *Journal of Quality Technology*, **35**, 317–328.

Kothari, R. and Pitts, D. (1999) On finding the number of clusters. *Pattern Recognition Letters*, **20**, 405–416.

Mahmoud, M.A., Parker, P.A., Woodall, W.H. and Hawkins, D.M. (2007) A change point method for linear profile data. *Quality and Reliability Engineering International*, **23**, 247–268.

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011

Mahmoud, M.A. and Woodall, W.H. (2004) Phase I analysis of linear profiles with calibration applications. *Technometrics*, **46**, 380–391.

Mallat, S. (1999) *A Wavelet Tour of Signal Processing*, Academic Press, Burlington, MA.

Mallat, S.G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

Mosesova, S.A., Chipman, H.A., MacKay, R.J. and Steiner, S.H. (2006) Profile monitoring using mixed-effects models. Technical report RR-06-06, University of Waterloo, Ontario, Canada.

Pinheiro, J.C. and Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York, NY.

Ramsay, J.O. and Silverman, B.W. (1997) *Functional Data Analysis*, Springer-Verlag, New York, NY.

Shiau, J.H., Huang, H., Lin, S. and Tsai, M. (2009) Monitoring nonlinear profiles with random effects by non-parametric regression. *Communications in Statistics-Theory and Methods*, **38**, 1664–1679.

Sullivan, J.H. (2002) Estimating the locations of multiple change points in the mean. *Computational Statistics*, **17**, 289–296.

Sullivan, J.H. and Woodall, W.H. (1996) A control chart for preliminary analysis of individual observations. *Journal of Quality Technology*, **28**, 265–278.

Sullivan, J.H. and Woodall, W.H. (2000) Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations. *IIE Transactions*, **32**, 537–549.

Vargas, J.A. (2003) Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*, **35**, 367–376.

Walker, E. and Wright, S.P. (2002) Comparing curves using additive models. *Journal of Quality Technology*, **34**, 118–129.

Williams, J.D., Woodall, W.H. and Birch, J.B. (2007) Phase I analysis of nonlinear product and process quality profiles. *Quality and Reliability Engineering International*, **23**, 925–941.

Worsley, K.J. (1979) On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association*, **74**, 365–367.

Zamba, K.D. and Hawking, D.M. (2006) A multivariate change-point model for statistical process control. *Technometrics*, **48**, 539–549.

Zhang, H. and Albin, S. (2007) Determining the number of operational modes in baseline multivariate SPC data. *IIE Transactions*, **39**, 1103–1110.

Zou, C., Qiu, P. and Hawkins, D. (2009) Non-parametric control chart for monitoring profiles using change point formulation and adaptive smoothing. *Statistica Sinica*, **19**, 1337–1357.

Zou, C., Tsung, F. and Wang, Z. (2008) Monitoring profiles based on non-parametric regression methods. *Technometrics*, **50**, 512–526.

Zou, C., Zhang, Y., and Wang, Z. (2006) A control chart based on a change-point model for monitoring profiles. *IIE Transactions*, **38**, 1093–1103.

Appendices

Appendix A: Derivation of $\mu_{\tilde{z}_{ir}}$ and $\sigma^2_{\tilde{z}_{ir}}$

First, we derive the conditional mean of denoised coefficients $\mu_{\tilde{z}_{ir}}$

$$\begin{aligned}
\mu_{\tilde{z}_{ir}} &= E[\tilde{z}_{ir}\,|b_{ir}] = E[\text{sign}(z_{ir})(|z_{ir}|-\zeta)\,I(|z_{ir}|>\zeta)\,|b_{ir}] \\
&= E[\text{sign}(z_{ir})(|z_{ir}|-\zeta)\,I(|z_{ir}|>\zeta)\,|(|z_{ir}|<\zeta),b_{ir}] \\
&\qquad \Pr(|z_{ir}|<\zeta\,|b_{ir}) \\
&\quad + E[\text{sign}(z_{ir})(|z_{ir}|-\zeta)\,I(|z_{ir}|>\zeta)\,|(|z_{ir}|>\zeta), \\
&\qquad b_{ir}]\Pr(|z_{ir}|>\zeta\,|b_{ir}) \\
&= E[(z_{ir}-\zeta)\,|z_{ir}>\zeta,b_{ir}]\Pr(z_{ir}>\zeta\,|b_{ir}) \\
&\quad + E[(z_{ir}+\zeta)\,|z_{ir}<-\zeta,b_{ir}]\Pr(z_{ir}<-\zeta\,|b_{ir}). \qquad \text{(A1)}
\end{aligned}$$

Since $z_{ir}|b_{ir} \sim N(\mu_{z_{ir}},\sigma^2)$, the random variable $(z_{ir}|z_{ir}>\zeta,b_{ir})$ follows a right truncated normal distribution with parameters $(\mu_{z_{ir}},\sigma^2,\zeta)$. Similarly, $(z_{ir}|z_{ir}<-\zeta,b_{ir})$ follows a left truncated normal distribution with parameters $(\mu_{z_{ir}},\sigma^2,-\zeta)$. Therefore, Equation (A1) can be written as

$$\begin{aligned}
\mu_{\tilde{z}_{ir}} &= \left(\mu^r_{z_{ir}}(\zeta)-\zeta\right)\Phi\left(\frac{\mu_{z_{ir}}-\zeta}{\sigma}\right)+\left(\mu^l_{z_{ir}}(-\zeta)+\zeta\right)\Phi \\
&\quad \times\left(\frac{-\mu_{z_{ir}}-\zeta}{\sigma}\right), \qquad \text{(A2)}
\end{aligned}$$

where $\mu^r_{z_{ir}}(\cdot)$ and $\mu^l_{z_{ir}}(\cdot)$, respectively, are the right and left truncated means of z_{ir} with truncation point $(\cdot)$ and can be calculated by

$$\begin{aligned}
\mu^r_{z_{ir}}(\zeta) &= \mu^+_{z_{ir}}\frac{\phi((\mu_{z_{ir}}-\zeta)/\sigma)}{\Phi((\mu_{z_{ir}}-\zeta)/\sigma)}\sigma \text{ and} \\
\mu^l_{z_{ir}}(-\zeta) &= \mu^+_{z_{ir}}\frac{-\phi((-\mu_{z_{ir}}-\zeta)/\sigma)}{\Phi((-\mu_{z_{ir}}-\zeta)/\sigma)}\sigma, \qquad \text{(A3)}
\end{aligned}$$

where $\phi(\cdot)$ is the probability distribution function of a normal standard random variable (Johnson and Kotz, 1970). The conditional variance $\sigma^2_{\tilde{z}_{ir}}$ is also obtained based on the derived $\mu_{\tilde{z}_{ir}}$:

$$\begin{aligned}
\sigma^2_{\tilde{z}_{ir}} &= E\left[\tilde{z}^2_{ir}\,|b_{ir}\right]-\mu^2_{\tilde{z}_{ir}} = E[(z_{ir}-\zeta)^2\,|z_{ir}>\zeta,b_{ir}] \\
&\qquad \times \Pr(z_{ir}>\zeta\,|b_{ir})+E[(z_{ir}+\zeta)^2\,|z_{ir}<-\zeta,b_{ir}] \\
&\qquad \times \Pr(z_{ir}<-\zeta\,|b_{ir})-\mu^2_{\tilde{z}_{ir}} \\
&= \left\{E^2[(z_{ir}-\zeta)\,|z_{ir}>\zeta,b_{ir}]+\text{var}[(z_{ir}-\zeta)\right. \\
&\qquad \left.|z_{ir}>\zeta,b_{ir}]\right\}\Pr(z_{ir}>\zeta\,|b_{ir}) \\
&\quad +\left\{E^2[(z_{ir}+\zeta)\,|z_{ir}<-\zeta,b_{ir}]+\text{var}[(z_{ir}+\zeta)\right. \\
&\qquad \left.|z_{ir}<-\zeta,b_{ir}]\right\}\Pr(z_{ir}<-\zeta\,|b_{ir})-\mu^2_{\tilde{z}_{ir}}
\end{aligned}$$

$$= \left\{\left(\mu^r_{z_{ir}}(\zeta) - \zeta\right)^2 + \left(\sigma^r_{z_{ir}}(\zeta)\right)^2\right\} \Phi\left(\frac{\mu_{z_{ir}} - \zeta}{\sigma}\right) + \left\{\left(\mu^l_{z_{ir}}(-\zeta) + \zeta\right)^2 + \left(\sigma^l_{z_{ir}}(-\zeta)\right)^2\right\} \Phi \times \left(\frac{-\mu_{z_{ir}} - \zeta}{\sigma}\right) - \mu^2_{\tilde{z}_{ir}},$$

where $(\sigma^r_{z_{ir}}(\cdot))^2$ and $(\sigma^l_{z_{ir}}(\cdot))^2$, respectively, are the right and left truncated varainces of z_{ir} with truncation point $(\cdot)$ and can be calculated by

$$\left(\sigma^r_{z_{ir}}(\zeta)\right)^2 = \sigma^2\left[1 - \frac{((\mu_{z_{ir}} - \zeta)/\sigma)\phi((\mu_{z_{ir}} - \zeta)/\sigma)}{\Phi((\mu_{z_{ir}} - \zeta)/\sigma)} - \left(\frac{\phi((\mu_{z_{ir}} - \zeta)/\sigma)}{\Phi((\mu_{z_{ir}} - \zeta)/\sigma)}\right)^2\right],$$

and

$$\left(\sigma^l_{z_{ir}}(\zeta)\right)^2 = \sigma^2\left[1 - \frac{((-\mu_{z_{ir}} - \zeta)/\sigma)\,\phi((-\mu_{z_{ir}} - \zeta)/\sigma)}{\Phi((-\mu_{z_{ir}} - \zeta)/\sigma)} - \left(\frac{\phi((-\mu_{z_{ir}} - \zeta)/\sigma)}{\Phi((-\mu_{z_{ir}} - \zeta)/\sigma)}\right)^2\right]. \quad \text{(A4)}$$

Appendix B: Proof of trace($\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}$) ≈ trace($\boldsymbol{\Sigma}_{f(\mathbf{t})}$).

It is known that $f(\mathbf{t})$ can be obtained by applying IDWT to true wavelet coefficients, that is

$$f(\mathbf{t}) = \mathbf{W}^{-1}\boldsymbol{\theta}. \quad \text{(A5)}$$

Therefore, the covariance matrix of $f(\mathbf{t})$ can be expressed by

$$\boldsymbol{\Sigma}_{f(t)} = \mathbf{W}^{-1}\text{var}(\boldsymbol{\theta})(\mathbf{W}^{-1})^{\mathrm{T}} = \mathbf{W}^{-1}\text{var}(\boldsymbol{\mu} + \mathbf{b})(\mathbf{W}^{-1})^{\mathrm{T}} = \mathbf{W}^{-1}\boldsymbol{\Lambda}(\mathbf{W}^{-1})^{\mathrm{T}}. \quad \text{(A6)}$$

This is true because $\boldsymbol{\mu}$ is deterministic. Also, since $\mathbf{W}$ is an orthogonal wavelet basis, it can be implied $(\mathbf{W}^{-1})^{\mathrm{T}} = \mathbf{W}$.

Taking the trace(.) yields

$$\text{trace}(\boldsymbol{\Sigma}_{f(t)}) = \text{trace}(\mathbf{W}^{-1}\Lambda\mathbf{W}) = \text{trace}(\mathbf{W}^{-1}\mathbf{W}\boldsymbol{\Lambda}) = \text{trace}(\boldsymbol{\Lambda}). \quad \text{(A7)}$$

Appendix C: Derivation of the LRT statistic

The log likelihood function under the alternative hypothesis in Equation (8) can be written as

$$l_1 = \log\left\{\prod_{i=1}^{\tau} h\left(\boldsymbol{\gamma}_i; \boldsymbol{\mu}^0_\gamma, \boldsymbol{\Lambda}_\gamma\right) \prod_{i=\tau+1}^{m} h\left(\boldsymbol{\gamma}_i; \boldsymbol{\mu}^1_\gamma, \boldsymbol{\Lambda}_\gamma\right)\right\} = -mc/2\log(2\pi) - m/2\log(|\boldsymbol{\Lambda}_\gamma|) - 1/2 \times \sum_{i=1}^{\tau}\left(\boldsymbol{\gamma}_i - \boldsymbol{\mu}^0_\gamma\right)^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}_\gamma(\boldsymbol{\gamma}_i - \boldsymbol{\mu}^0_\gamma) - 1/2 \sum_{i=\tau+1}^{m}\left(\boldsymbol{\gamma}_i - \boldsymbol{\mu}^1_\gamma\right)^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}_\gamma\left(\boldsymbol{\gamma}_i - \boldsymbol{\mu}^1_\gamma\right), \quad \text{(A8)}$$

where $h(.)$ is the multivariate normal probability distribution function, and c is equal to the cardinality of $\boldsymbol{\Omega}_s$.

Under H_0, the corresponding log likelihood function would be

$$l_2 = \log\left\{\prod_{i=1}^{m} h\left(\boldsymbol{\gamma}_i; \boldsymbol{\mu}^0_\gamma, \boldsymbol{\Lambda}_\gamma\right)\right\} = -mc/2\log(2\pi) - m/2\log(|\boldsymbol{\Lambda}_\gamma|) - 1/2 \sum_{i=1}^{m}\left(\boldsymbol{\gamma}_i - \boldsymbol{\mu}^0_\gamma\right)^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}_\gamma\left(\boldsymbol{\gamma}_i - \boldsymbol{\mu}^0_\gamma\right). \quad \text{(A9)}$$

The maximum likelihood estimators for mean parameters are $\hat{\boldsymbol{\mu}}^0_\gamma = \sum_{i=1}^{\tau}\boldsymbol{\gamma}_i/\tau$ and $\hat{\boldsymbol{\mu}}^1_\gamma = \sum_{i=\tau+1}^{m}\boldsymbol{\gamma}_i/(m-\tau)$, respectively. The estimate of $\boldsymbol{\Lambda}_\gamma$ is the pooled-sample covariance matrix; i.e., $\hat{\Lambda}_\gamma = \{\sum_{i=1}^{\tau}(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\mu}}^0_\gamma)(\varphi_i - \hat{\boldsymbol{\mu}}^0_\gamma)^{\mathrm{T}} + \sum_{i=\tau+1}^{m}(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\mu}}^1_\gamma)(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\mu}}^1_\gamma)^{\mathrm{T}}\}/(m-2)$. Replacing $\boldsymbol{\mu}^1_\gamma$, $\boldsymbol{\mu}^0_\gamma$, and $\boldsymbol{\Lambda}_\gamma$ in Equation (C1) by their ML estimators, after simplification, the log likelihood ratio can be expressed as

$$\Gamma(\tau) = l_1 - l_2 = \frac{\tau(m-\tau)}{m}\sum_{i=\tau+1}^{m}(\hat{\boldsymbol{\mu}}^1_\gamma - \hat{\boldsymbol{\mu}}^0_\gamma)^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}_\gamma(\hat{\boldsymbol{\mu}}^1_\gamma - \hat{\boldsymbol{\mu}}^0_\gamma) \quad \text{(A10)}$$

Biographies

Kamran Paynabar is a Ph.D. candidate in the Department of Industrial and Operations Engineering at the University of Michigan. He received his B.Sc. and M.Sc. in Industrial Engineering at Iran University of Science and Technology and Azad University in 2002 and 2004, respectively, and his M.A. in Statistics at the University of Michigan in 2010. His research interest include developing new approaches for system modeling, process monitoring, and fault diagnosis of manufacturing and health care systems. His research expertise is in the area of variation modeling and fault diagnosis of high-dimensional functional data and waveform signals by integrating statistical learning techniques and enigneering models. He is a member of ASQ, IIE, and INFORMS.

Jionghua (Judy) Jin is an Associate Professor in the Department of Industrial and Operations Engineering at the University of Michigan. She received her Ph.D. degree from the University of Michigan in 1999. Her recent research focuses on data fusion for complex system improvement with the goal of developing novel methodologies for variation reduction, condition monitoring and fault diagnosis, process control, knowledge discovery, and decision making. Her research emphasizes a multidisciplinary approach by integrating applied statistics, signal processing, reliability engineering, system control, and decision-making theory. She has received a number of awards including a NSF CAREER Award in 2002 and a PECASE Award in 2004 and six Best Paper Awards between 2005 and 2010. She is a member of ASME, ASQC, IEEE, IIE, INFORMS, and SME.

Downloaded By: [Paynabar, Kamran] At: 17:21 13 February 2011