Smart Structures and Systems, *Vol. 14, No. 3 (2014) 377-395*
DOI: http://dx.doi.org/10.12989/sss.2014.14.3.377

Bearing fault detection through multiscale wavelet scalogram-based SPC

Uk Jung[1] and Bong-Hwan Koh[*2]

[1]*Production and Operations Division, School of Business, Dongguk University-Seoul, Republic of Korea*
[2]*Department of Mechanical, Robotics, and Energy Engineering, Dongguk University-Seoul, 30 Pildong-ro, 1 gil Jung-gu, Seoul 100-715, Republic of Korea*

(Received July 18, 2013, Revised September 24, 2013, Accepted January 14, 2014)

Abstract. Vibration-based fault detection and condition monitoring of rotating machinery, using statistical process control (SPC) combined with statistical pattern recognition methodology, has been widely investigated by many researchers. In particular, the discrete wavelet transform (DWT) is considered as a powerful tool for feature extraction in detecting fault on rotating machinery. Although DWT significantly reduces the dimensionality of the data, the number of retained wavelet features can still be significantly large. Then, the use of standard multivariate SPC techniques is not advised, because the sample covariance matrix is likely to be singular, so that the common multivariate statistics cannot be calculated. Even though many feature-based SPC methods have been introduced to tackle this deficiency, most methods require a parametric distributional assumption that restricts their feasibility to specific problems of process control, and thus limit their application. This study proposes a nonparametric multivariate control chart method, based on multiscale wavelet scalogram (MWS) features, that overcomes the limitation posed by the parametric assumption in existing SPC methods. The presented approach takes advantage of multi-resolution analysis using DWT, and obtains MWS features with significantly low dimensionality. We calculate Hotelling's T^2-type monitoring statistic using MWS, which has enough damage-discrimination ability. A bootstrap approach is used to determine the upper control limit of the monitoring statistic, without any distributional assumption. Numerical simulations demonstrate the performance of the proposed control charting method, under various damage-level scenarios for a bearing system.

Keywords: statistical process control; fault detection; bootstrap; wavelet; scalogram

1. Introduction

As one of the most critical categories of industrial equipment, rotating machinery requires a condition monitoring technique for early warning of the occurrence of fault, before it develops to failure and breakdown. Reliable monitoring and evaluation techniques are responsible for providing decision-making to the operator regarding the operation of rotating machinery, based on a thorough evaluation of the available data in a timely fashion. Rolling element bearings are the most popular locations of potential defects in rotating machinery. Because the essence of bearing condition monitoring is a prognostic warning of incipient fault, advanced signal processing and

*Corresponding author, Associate Professor, E-mail: bkoh@dongguk.edu

http://www.techno-press.org/?journal=sss&subpage=8 ISSN: 1738-1584 (Print), 1738-1991 (Online)

fault diagnosis are very critical. In this context, signal processing of vibration measurement has been an active research area in recent years (Lin and Zhang 2005, Lei *et al*. 2013, Koh *et al*. 2008) Typically, a bearing system with a defect experiences periodic impulses in its vibration signal, due to repeated collisions between the defect and rolling components of the bearing.

The typical condition monitoring process is composed of two steps: first, acquisition of data from sensor measurements around the bearing system. Then, extracting damage-sensitive features from the measured data, and classifying the pattern for further diagnosis, such as allowing a decision for maintenance (Jardine *et al*. 2006). Over the last three decades, feature extraction from vibration signal has been widely investigated for monitoring bearing conditions, in conjunction with advanced signal processing techniques (Randall and Antoni 2011). Considering the non-stationary nature of bearing dynamics, many studies are devoted to the time-domain and/or time-frequency analysis. In the time domain, the simplest fault detection method is to evaluate the crest factor or kurtosis value from measurements (Dyer 1978). Also, empirical mode decomposition (Yu *et al*. 2006), wavelet transform (Prabhakar 2002), and autoregressive filter (Wang 2002) have been successfully employed for monitoring fault in rotating machinery.

Among many other fields, condition monitoring for rotating machinery using statistical process control (SPC) combined with statistical pattern recognition methodology, such as feature extraction, has become an area of promising research. Recent literature has presented several variants of a wavelet-decomposition-based multiscale data analysis method for vibration-based fault detection, which are, in concept, superior to those of the traditional single scale methods, such as the time series method (Jung and Koh 2009, Jung *et al*. 2006). A discrete wavelet transformation (DWT) uses appropriate low-pass and high-pass filters, to map the signal from the time domain into the time-scale domain. While DWT significantly reduces the dimensionality of the data, using denoising techniques, the number of retained wavelet coefficients can still be significantly large.

Hence, additional dimensionality reduction effort is necessary to summarize the wavelet coefficients. Consequently, a DWT is only a start, and further analysis of the coefficients is needed for multivariate monitoring process. Thus, in this present study, the multiscale wavelet scalogram (MWS) is selected as a set of defect-sensitive features, especially in very low dimension, for constructing a control chart, to avoid a high false alarm rate (FAR). MWS provides measures of signal energy at various frequency bands, and is commonly used in decision making tasks, including signal and image processing, astronomy and metrology (Rioul and Vetterli 1991, Scargle 1997, Peng *et al*. 2002, Jeong *et al*. 2003). Peng *et al*. (2002) used the wavelet scalogram to analyze vibration data of an industrial machine. The machine was 300MW pumped storage power generator unit (PSPGU) running at several operating conditions, and affected by many factors, such as hydraulic factors and electric factors. They showed the advantages of wavelet scalogram to extract the feature of defect. Jeong *et al*. (2003) applied the MWS concept to the problem of a rapid thermal chemical vapour deposition (RTCVD) process that deposits thin films on semiconductor wafers using a temperature-driven surface chemical reaction in order to detect process condition different from the nominal. They applied a commonly used data mining tool to the scalograms and showed the effectiveness of their techniques. MWS represents the scale-wise distribution of energies of the signal. These energies at different scales have distinct characteristics of different signals. Therefore, intuitively, it is possible that MWS will be useful in monitoring process changes, with data collected in time sequences.

Traditionally, a control chart assumes that the monitoring statistics follow a certain probability distribution. The most widely used multivariate control chart has been Hotelling's T^2 (T^2 chart)

(Hotelling 1947). The T^2 charts monitor T^2 statistics that measure the distance between an observation, and the scaled-mean, estimated from the in-control data. Assuming that the observed process data follow a multivariate normal distribution, the control limit of a T^2 control chart is proportional to the percentile of an β distribution in Phase I (retrospective analysis), and an F distribution in Phase II (future monitoring) (Mason and Young 2002). However, when the normality assumption of the data does not hold, a control limit based on the β and F distribution may be inaccurate, because the control limit determined this way may increase the false alarm rate (Chou *et al*. 2001). To tackle the limitation posed by the distributional assumption underpinning traditional control charts, nonparametric (or distribution-free) control charts have been developed. In particular, many studies have focused on the construction of nonparametric control charts by using a bootstrap procedure (Bajgier 1992, Seppala *et al*. 1995, Liu and Tang 1996, Lio and Park 2008, Polansky 2005, Phaladiganon *et al*. 2011). This procedure is favored over the traditional distribution-assuming procedures, because of its proven capabilities to effectively manage process data, without posing assumptions about their distribution.

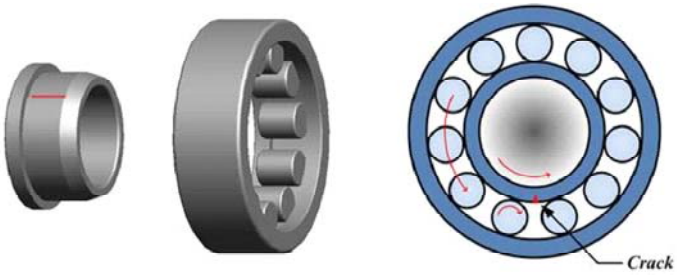
In this study, multivariate control chart analysis, which is the most commonly used SPC technique, and very suitable for automated condition monitoring, is applied to evaluate vibration signals, for further diagnosing the incipient fault in a bearing system. In our case, even though MWS composes a set of good features with high discriminating power in very low dimension, suitable for SPC application, the probability distributional information of MWS is theoretically intractable, due to the nature of the nonparametric approach, so that the traditional multivariate statistical control chart, such as Hotelling's T^2 control chart, is not a good choice. To overcome the limitation posed by the parametric assumption, this study proposes a nonparametric multivariate control chart method, using a bootstrap approach based on MWS-based monitoring statistic for fault detection. Before the proposed method is introduced, we first summarize a discrete wavelet transformation, and multivariate statistical process control techniques.

The organization of this paper is as follows. In Section 2, the motivating example of a bearing system and vibration signal acquisition is summarized. The theoretical background for discrete wavelet transformation and multivariate statistical process control is outlined in Section 3. In Section 4, the proposed method for monitoring a high-dimensional vibration signal is explained. In Section 5, the performance of the proposed method is evaluated, through simulation of the bearing signal. Finally, Section 6 contains the conclusion, and future research work.

2. Motivating example: bearing condition monitoring

2.1 Bearing system with localized defect

The bearing system in most rotating machinery needs a reliable measure for diagnosing its functionality, to avoid the risk of catastrophic accident, which may cause personal damage and economic loss. The most successful method for condition monitoring of a bearing system is using vibration measurements. The easiest way of measuring the vibration is mounting accelerometers on the surface of the outer frame of the bearing, and instrumenting them to the data acquisition system, for further signal processing.



(a) (b)

Fig. 1 Schematics of a roller-bearing system: (a) a defect on the surface of an inner race with rolling elements, and (b) rotating direction of the rolling elements

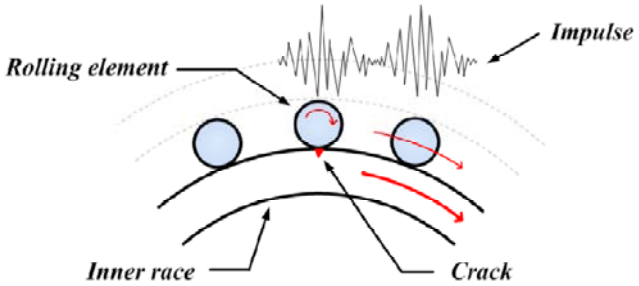


Fig. 2 Schematics of the impulsive stress wave generated by repeated collisions of crack and rolling elements

A roller bearing system includes three rolling components: the inner-race, outer-race and rollers, as shown in Fig. 1. The common defects of bearing systems are cracks, pits and spalls on the surface of the rolling elements. When a localized defect on the element hits its mating element, a sharp impulse inflicts the stress waves within a very short duration (see Fig. 2). And, this impulse penetrates the overall structure of the bearing system, including its mounting frame. Thus, the reflected stress wave is similar to the impulse response, incorporating the dynamics of the overall system. The pulse wave produces noise-corrupted vibration, which needs to be monitored to diagnosis the occurrence of fault in the bearing.

2.2 Simulation of the bearing vibration signal

Instead of using experimental data, many studies exploit bearing simulation schemes, in which we can arbitrary control the severity and collision frequency of defects. Previously, some researchers proposed bearing simulation methods, by producing series of repeated periodic oscillation with proportional damping terms (McFadden and Smith 1984, Cong *et al*. 2010). Again, the frequency of impulse is related to the collision frequency between the defect and rotating

elements. The model equation of the simulated bearing signal can be expressed as Eq. (1)

$$x(t) = \sum_{i=1}^{M} A_i w(t - iP - \tau_i) + n(t) \tag{1}$$

In the equation, P is the impulses period, $w(t)$ is the oscillating wave, and $n(t)$ is measurement noise. A_i is the periodic amplitude, after the T period. Note that the shaft speed for inner defects and cage speed for rolling elements becomes $1/T$. In the averaging period of impulses, minor and random fluctuation in oscillating waveform is denoted as τ_i. Considering natural frequency and damping of the bearing system, the equation of $w(t)$ becomes a cosine with a damping term, as $s(t) = e^{-Dt}cos(2\pi f_n t + \varphi_w)$. For exponential decaying to simulate damping, i.e., decrease of waveform to zero after a specific interval, D can be conveniently selected with appropriate value. Also, phase angle is denoted as φ_w. The amplitude of A_i can be presented as the amplitude of the impulse having cosine waveform $A_i = A_0 cos(2\pi f_n t + \varphi_A)$.

The limitation of the simulation approach using Eq. (1) is that it only represents a single frequency component. It is reasonable to believe that a bearing system has its own dynamics, and many modes of broad spectrum. Thus, we used a state-space model to represent the dynamics of the overall bearing system, instead of simply using Eq. (1). Specifically, the state-space model consists of two-degree-of-freedom system, i.e., two masses connected by springs and dampers. Each compoent of spring, mass, and damper simulates the overall dynamics of inner-race, outer-race, rolling elements, and housing structure. The effect of bearing fault is simulated by inflicting periodic impulses to the state-space system.

It is reasonable to consider that the severity of fault in a bearing is proportional to the power of periodic impulse caused by collision impact between the defect and rolling elements. We quantified the level of fault severity, denoted SL (Severity Level), as the relative intensity of impact force on the state-space model, having a scale from 0 to 1. Here, 0 represents no fault or pristine condition, while 1.0 indicates the condition from the most severe defect. In this paper, we used five levels of fault severity, SL =0.05, 0.1, 0.15, 0.2, and 0.25. Thus, the lower the SL value, the more difficult it is to differentiate the incipient defect from measurement noise or background resonances. The time-history data simulated from the model of the bearing system for the healthy case is shown in Fig. 3(a), and for the bearing having defect is depicted in Fig. 3(b), respectively. It is worth noting that the severity of a bearing defect reflected in terms of signal power can be easily masked by environmental noise, making its detection extremely difficult in practice. Thus, damage-sensitive features extracted by signal processing must be further interpreted, and evaluated for fault detection. Again, feature extraction is the process of identifying damage-sensitive properties derived from the measured vibration that allows one to differentiate the incipient fault from a healthy bearing. The number of data points has been fixed to 2^{12} =4096 points, which is large enough to include at least two defect-driven impulses during the rotation of the bearing.

Once it is revealed that the extracted features are adequately sensitive to structural defects, the most direct way to deal with a vibration signal in SPC, is treating features extracted from the vibration signal as a realization of a multivariate process. However, the use of standard multivariate SPC charts is ill-advised, because when the number of monitored features exceeds the number of collected samples, the covariance matrix is singular, and the common multivariate

statistics cannot be calculated. Also, even though the number of monitored features is less than the number of collected samples, it is well known that the power of detecting process fault or structural damage will drop significantly, when the number of features becomes large (Fan 1996). This implies that the rate of false alarm will be significantly increased. Functional principal component analysis (FPCA; Ramsay and Silverman 1997) and related procedures (Hall *et al*. 2001) are useful as a dimension reduction tool, in modelling high dimensional data. However, there are some drawbacks in applying the FPCA approach to SPC procedure. For instance, the FPCA method presents difficulties of interpreting the relationship between changes in a few principal components, and the high-dimensional signal. To overcome the limitation posed by high-dimensionality, this study proposes a wavelet-based multivariate control chart method, for fault detection of a bearing system.

3. Theoretical background: DWT and SPC

3.1 Discrete wavelet transformation (DWT)

The feature extraction using vibration signal is mostly based on the Fourier transform, which breaks down a temporal signal into constituent sinusoids of different frequencies. The Fourier analysis transforms a time-based signal to a frequency-based one. Unfortunately, in transforming the domain, the time-based information is lost, and it is impossible to determine when or where the particular event took place. To a certain extent, this could be overcome by the application of different windowing techniques, leading to the short-time Fourier transforms (STFT). But these also have their own limitations, such as the fact that the information about time and frequency can be obtained with a limited precision that is determined by the size of the window. An alternative method is to use wavelets. This section provides a brief review of DWT (Mallat 1989, Vidakovic 1999).

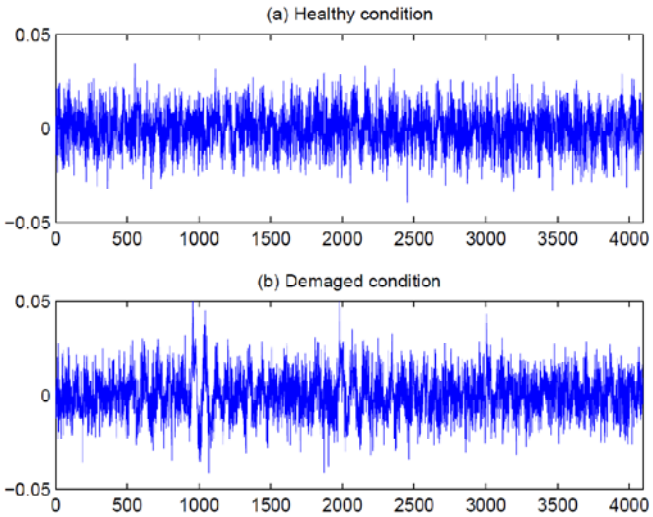


Fig. 3 Examples of time-history response from sensor for bearing: (a) healthy condition, and (b) damaged condition with SL =0.2

A wavelet is a function $\psi(t) \in L^2(R)$ with the following basic properties

$$\int_R \psi(t)dt = 0 \quad and \quad \int_R \psi^2(t)dt = 1, \tag{2}$$

where, $L^2(R)$ is the space of square integrable real functions defined on the real line R. Wavelets can be used to create a family of time-frequency atoms, $\psi_{s,u}(t) = s^{1/2}\psi(st-u)$, via the dilation factor s, and the translation u. We also require a scaling function $\phi(t) \in L^2(R)$ that satisfies

$$\int_R \phi(t)dt \neq 0 \quad and \quad \int_R \phi^2(t)dt = 1. \tag{3}$$

Selecting the scaling and wavelet functions as $\{\phi_{L,k}(t) = 2^{L/2}\phi(2^L t-k); k \in Z\}$, $\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t-k); j \geq L, k \in Z\}$, respectively, one can form an orthonormal basis to represent a signal function $f(t) \in L^2(R)$, as follows

$$f(t) = \sum_{k \in Z} c_{L,k}\phi_{L,k}(t) + \sum_{j \geq L}\sum_{k \in Z} d_{j,k}\psi_{j,k}(t), \tag{4}$$

where, Z denotes the set of all integers $\{0,\pm1,\pm2,\ldots\}$, and the coefficients $c_{L,k} = \int_R f(t)\phi_{L,k}(t)dt$ are considered to be the approximation level coefficients characterizing smoother data patterns, and $d_{j,k} = \int_R f(t)\psi_{j,k}(t)dt$ are viewed as the detail level coefficients describing (local) details of data patterns. In practice, the following finite version of the wavelet series approximation is used

$$\tilde{f}(t) = \sum_{k \in Z} c_{L,k}\phi_{L,k}(t) + \sum_{j=L}^{J}\sum_{k \in Z} d_{j,k}\psi_{j,k}(t), \tag{5}$$

where, $J > L$ and L correspond to the coarsest resolution level. Consider a sequence of data $\mathrm{y} = (y(t_1),\cdots,y(t_N))^T$ taken from $f(t)$. The discrete wavelet transform (DWT) of y is defined as

$$\mathrm{d} = (\mathrm{d}_{(1)},\mathrm{d}_{(2)},\ldots,\mathrm{d}_{(I)})^T \text{ where } \mathrm{I} = J - L + 2. \tag{6}$$

The main advantage of exploiting wavelets is performing local analysis of a highly localized signal, such as spikes. On the other hand, wavelets can be contained in a finite interval. Hence, they are well suited to represent or approximate signals containing discontinuities. This functionality is particularly important for fault detection using vibration signals. Moreover, due to the availability of a fast transform version, the computational effort to perform the signal processing is reduced. Because of these features, the wavelet transform has recently been considered as a powerful tool for fault detection and condition monitoring for rotating machinery (Prabhakar 2002). Wavelet-based examples showing the time-scale nature of damage characteristics in vibration responses can be found in Law (2005), Li (2006), and Li *et al*. (2007).

3.2 Statistical Process Control (SPC)

SPC is one of the most widely used techniques for quality control. The basic objective of SPC is to quickly detect the occurrence of special cause variation, so that the process can be investigated, and corrective action may be taken, before quality deteriorates and defective units are produced (Stoumbos *et al*. 2000). One important tool in SPC is the control chart, which is used to monitor the performance of a process over time, to keep the process in an in-control state. A broad spectrum of SPC methods have been developed, including methods for univariate SPC, such as Shewhart, moving-average (MA), exponentially weighted moving-average (EWMA), and cumulative-sum (CUSUM) charts. Methods for multivariate SPC include multivariate extensions of univariate methods, such as the Hotelling's T^2 statistic, and methods that monitor latent variables obtained by combining the measured variables with a lower dimension of space. Popular methods for reducing the dimensionality of the measured data include principal-component analysis (PCA), and partial least-square regression (PLS) (Ku *et al*. 1995). Many extensions and applications of these have been extensively investigated (Kresta *et al*. 1991, MacGregor and Kourti 1995).

In general, control chart problems in SPC can be divided into two phases (Woodall and Montgomery 1999, Woodall 2000). Phase I analysis tries to isolate the in-control (baseline) data from an unknown historical data set, and establish the control limits for future monitoring (Zhang and Albin 2007). Phase II analysis monitors the process using control charts derived from the "cleaned" in-contol data set from the Phase I analysis. With a simple plot of the set of monitoring statistics derived from the original samples, the control chart can effectively determine whether or not a process is in an in-control state. In addition to the monitoring statistics, another important component of control charts is determining the control limits, which are often calculated based on the probabilistic distribution of the monitoring statistics (Sukchotrat *et al*. 2010).

Common characteristics used for comparing the performance of control charts are first an in-control Average Run Length (ARL0), and secondly an out-of-control Average Run Length (ARL1). ARL0 is defined as the expected number of observations taken from an in-control process, until the control chart falsely signals out-of-control. ARL0 is regarded as acceptable, if it is large enough to keep the level of false alarms at an acceptable level. ARL1 is defined as the expected number of observations taken from an out-of-control process, until the control chart correctly signals that the process is out-of-control. Ideally, the ARL1 should be as small as possible.

3.2.1 Parametric Hotelling's T^2 statistic and limitation

Hotelling extended the univariate control chart to handle multivariate problems (Hotelling 1947). Hotelling's T^2 chart is a multivariate control chart that can monitor a multivariate process efficiently. Mason and Young give the basic steps for the implementation of multivariate statistical process control using the T^2 statistic, and they published a textbook on the practical development and application of multivariate control techniques using the T^2 statistic (Mason and Young 2002). There is an crucial matter of the sample size n of each sample statistic in a control chart. If $n=1$, then special care must be taken. As Lowry and Montgomery suggest, the appropriate use of a sample statistic can be divided into four categories (i.e., (1) Phase I and $n=1$, (2) Phase I and $n>1$, (3) Phase II and $n=1$, (4) Phase II and $n>1$).

For charts constructed using individual observations ($n=1$), the test statistic for the ith

individual observation has the form

$$T_i^2 = (\mathrm{x}_i - \mu_0)^T \Sigma_0^{-1} (\mathrm{x}_i - \mu_0) \tag{7}$$

where, x_i is the i th, $i = 1,2,\cdots,m$ observation following $N_p(\mu_0, \Sigma_0)$, where μ_0 and Σ_0 are the known vector of means and the known variance-covariance matrix, respectively. The T_i^2 statistic follows a χ^2-distribution with p degrees of freedom (Seber 1984).

In Eq. (7), if μ_0 is replaced by $\overline{\mathrm{x}}_0$, Σ_0 is replaced by S_0, and x_f is a future individual observation, which is independent of the estimators $\overline{\mathrm{x}}_0$ and S_0, then the $T_i^2 / p(m+1)(m-1)(m(m-p))^{-1}$ statistic follows an F-distribution with p and $(m-p)$ degrees of freedom. Thus, a multivariate Shewhart control chart for the process mean, with unknown parameters, has the upper control limit (Tracy *et al*. 1992) in Eq. (8) below, and this control chart is called a Phase II T^2-chart.

$$UCL = \frac{p(m+1)(m-1)}{m(m-p)} F_{1-\alpha, p, (m-p)} \tag{8}$$

Although Hotelling's T^2 chart is one of most popular multivariate control charts, it is difficult to apply this chart to the non-normal data that is often found in many applications, due to the distributional assumption of the T^2 chart. A number of non-parametric control charts have been developed to address the limitation of the distributional assumptions (Chakraborty *et al*. 2001, Liu *et al*. 2004, Bakir 2006, Kim *et al*. 2007, Qiu 2008). However, no consensus exists about which of them best satisfies all the conditions encountered in modern process systems. A detailed review of non-parametric control charts is beyond the scope of this study.

3.2.2 Nonparametric approach for non-normality situations: bootstrapping

In traditional control charts, the control limits are determined based on the underlying distribution of the monitoring statistic, with the user-specified value of Type I error rate (α). Here, Type I error rate represents the probability of a false alarm. In contrast, the distribution of some non-parametric monitoring statistics of new control charts is unknown, due to its non-parametric nature. This provides the motivation to develop an appropriate non-parametric procedure to establish the control limits (Chou *et al*. 2001, Sukchotrat *et al*. 2010). In particular, many studies have focused on the construction of non-parametric control charts, by using a bootstrap procedure. The bootstrap, originally proposed and named by Efron (1979), is a computational technique that can be used to effectively estimate the sampling distribution of a statistic. One can use the non-parametric bootstrap to estimate the sampling distribution of a statistic, while the only assumptions are that the sample is representative of the population from which it is drawn, and that the observations are independent and identically distributed.

Bajgier (1992) introduced a univariate control chart whose lower and upper control limits were estimated by using the bootstrap technique, and Seppala *et al*. (1995) proposed a subgroup bootstrap chart, to compensate for the limitations of Bajgier's control charts. Liu and Tang (1996) proposed a bootstrap control chart that can monitor both independent and dependent observations. Note that several control charts based on the bootstrap are compared in the study from Jones and Woodall (1998). We apply the concept of the bootstrap procedure in Section 4.2 to construct a control limit for the our test statistic, with a more explicit description.

4. The proposed control chart: MWS-based bootstrap control chart

4.1 Multiscale Wavelet Scalogram (MWS)

When the ith component of d in Eq. (6) is defined as $\mathrm{d}_{(i)} = (d_{(i)1}, d_{(i)2}, \cdots, d_{(i)m_i})^T$, where m_i is the number of wavelet coefficients at the (i)th resolution level, the element of multiscale wavelet scalogram (MWS) at the (i)th resolution level is defined as (Vidakovic 1999, Jeong *et al*. 2003)

$$s_{(i)} = \sum_{k=1}^{m_i} d_{(i)k}^2, \quad i = 1,2,\cdots,I \tag{9}$$

which is the sum of the squares of all wavelet coefficients at the (i)th resolution level, where $I = J - L + 2$. Then, a vector of MWS that represents a single vibration signal, is defined as

$$S = (s_{(1)}, s_{(2)}, \cdots, s_{(I)}) \tag{10}$$

where, $s_{(1)}$ is for the approximation level, and $s_{(i)}, 2 \le i \le I$ is for the detail levels. MWS provides measures of signal energy at various frequency bands (Rioul and Vetterli 1991). MWS represents the scale-wise distribution of the energies of a signal. Thus, a vector of MWS in Eq. (10) is in very low dimension, depending on the predetermined depth of decomposition (decomposition level= $I-1$). An adequate depth of decomposition is required, to avoid the risk of eliminating important features, and over-smoothing of the signal. For a signal of length N in an off-line mode, or a window of length N in an on-line mode, empirical evidence suggests that the depth of decomposition should be half the maximum possible depth. According to Ganesan et al. (2004), for a dyadic data length $N = 2^j$, the recommended depth of decomposition is $j/2$ (i.e., $I = j/2 + 1$). Fig. 3 shows the result of boxplot analysis based on MWS, for the bearing signal having N=4096 datapoints and 6 decomposition levels ($S = (s_{(1)}, s_{(2)}, \cdots, s_{(7)})$), towards the healthy and damaged cases. Here, a thousand sets of bearing signals for both healthy (SL =0) and damaged (SL =0.2) cases are generated for demonstration purpose (each set has N=4096 datapoints). As shown in the figure, the approximation level $s_{(1)}$, and relatively high frequency scales, such as $s_{(4)}$, $s_{(5)}$, $s_{(6)}$, $s_{(7)}$, are less sensitive to the bearing defect, while low frequency scales, such as $s_{(2)}$ and $s_{(3)}$, are quite sensitive to the defect. For example, the highest frequency scale $s_{(7)}$ in Fig. 3(g) has almost the same interquartile ranges between the health and damaged case, while the lowest frequency scale $s_{(2)}$ in Fig. 3(b) has no overlapping interquartile ranges. Fig. 4 represents a 2-D visualization of the discriminating ability of MWS. The goal of MSW-based clustering is to discriminate the existence of damage in a bearing. Here, the scales $s_{(2)}$ and $s_{(3)}$ are used for extracting damage-sensitive features for clustering. Obviously, all 1000 sets of signals generated from the damaged bearing with (SL =0.2) have been sufficiently separated from the 1000 sets of signals from the healthy bearing (SL =0). Thus, in this present

study, MWS is regarded as the set of extracted features that are damage-sensitive, in terms of scale-wise energy representation of vibration signals. MWS is also, in very low dimension, suitable for constructing a multivariate control chart.

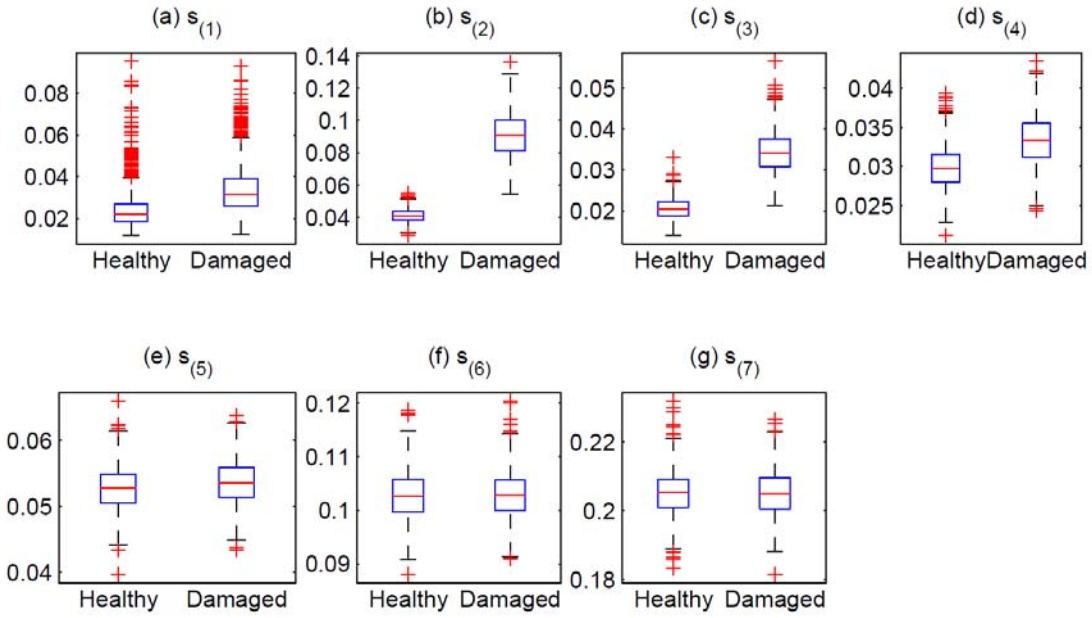


Fig. 4 Boxplot analysis based on MWS at different scales with two different classes (healthy and damaged condition)

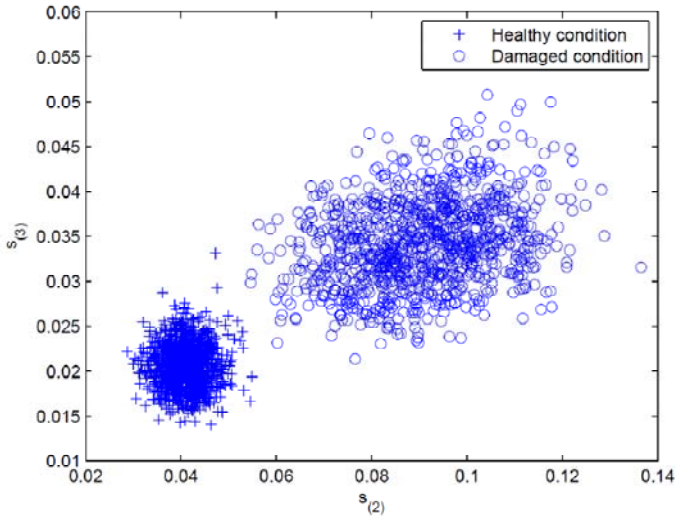


Fig. 5 Two-dimensional visualization of clusters at bi-scale scalograms ($s_{(2)}$ and $s_{(3)}$)

4.2 MWS-based Bootstrap Control Chart (MWSB chart)

The proposed monitoring statistic using MWS, $T^2_{(S)}$, measures the distance between an observed MWS vector of the vibration signal, and the scaled-mean MWS vector, estimated from the in-control data. The monitoring statistic for the i th individual vibration signal observation can be calculated by the following equation, which is similar to Eq. (7)

$$T^2_{(S)i} = (S^{(i)} - \bar{S})^T \Sigma_S^{-1} (S^{(i)} - \bar{S}) \tag{11}$$

where, $S^{(i)}$ is a vector of MWS for the i th observation, $\bar{S}$ and Σ_S are a sample mean vector and a sample covariance matrix of those vectors of MWS for vibration signals from an in-control process, respectively. Then, the procedure for constructing our multiscale wavelet scalogram-based control chart via bootstrap (MWSB chart) is summarized as follows, in a similar manner used in Phaladiganon *et al*. (2011).

1. Compute the $T^2_{(S)i}$ statistic for $i = 1, \cdots, m$ with m observations from an in-control dataset, using Eq. (11).
2. Let $T^{2(b)}_{(S)1}$, $T^{2(b)}_{(S)2}$, $\cdots$, $T^{2(b)}_{(S)m}$ be a set of m $T^2_{(S)i}$ values from the b th bootstrap sample ($b = 1, \cdots, B$) randomly drawn from the initial $T^2_{(S)i}$ statistics, with replacement. In general, B is a large number (e.g. $B > 5{,}000$).
3. In each of B bootstrap samples, determine the $100 \cdot (1-\alpha)$th percentile value, given a user-specified value α, with a range between 0 and 1.
4. Determine the control limit, by taking the average of B $100 \cdot (1-\alpha)$th percentile value ($\bar{T}^2_{(S),100\cdot(1-\alpha)}$). Note that statistics other than the average can be used (e.g., the median).
5. Use the established control limit to monitor a new observation. That is, if the monitoring statistic of a new observation exceeds ($\bar{T}^2_{(S),100\cdot(1-\alpha)}$), we declare that specific observation as out of control.

Determination of the appropriate number of bootstrap samples to use is not obvious. However, with reasonably large numbers of bootstrap samples, the variation of results is insignificant. The computational time required has been perceived as one of the disadvantages of the bootstrap technique, but this is no longer a critical issue, because of the computing power currently available.

5. Simulation study for performance evaluation

5.1 Comparison of control limits

In order to use MWS features for bearing condition monitoring, we need to first decide which wavelet basis function type should be used. Several wavelet basis function types are available in

the literature. Some of these are the Haar's, Daubechies', coiflets, symlets, and bi-orthogonal wavelets. For a detailed description on the mathematical properties of the wavelet basis function, refer to Goswami and Chan (1999). In this study, we used Daubechies' wavelet type (db5), with depth of decomposition 6 (i.e. N =4096, $\frac{log_2(N)}{2}$ =6).

We generated two groups of 1,000 sets of signals from healthy condition (in-control observations). The first group of 1,000 in-control sets of signal was used to determine the control limits of the $T^2_{(s)}$ statistic, from the F -distribution and bootstrap approach. The second group of 1,000 in-control sets of signal was monitored on the control chart, which were based on the control limits established by the first group of in-control sets of signal. The control chart that produces a similar value for both the actual FAR and the nominal FAR would be considered the better one. To check multivariate normality of the in-control observations, we conducted a Royston's H test (Royston 1983). The p -value, which measures the plausibility that the dataset follows a multivariate normal distribution, was almost zero (i.e., Royston's statistic: 188.067, degrees of freedom: 7.000048, P-value: 0.0000). This strongly indicates that this dataset does not come from a multivariate normal distribution.

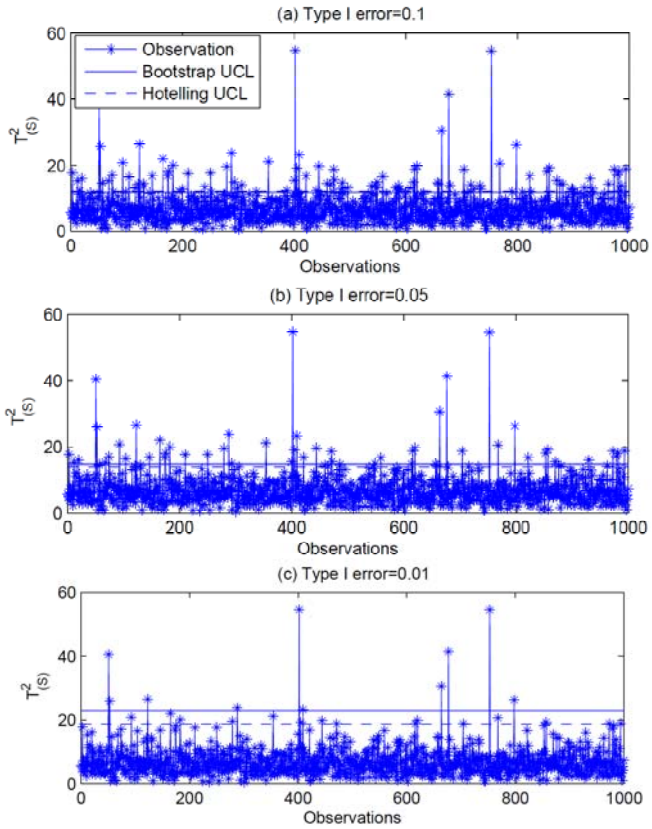


Fig. 6 Control limits established by the Hotelling's F-distribution, and the proposed bootstrap percentile on the monitoring statistic $T^2_{(S)i}$

Fig. 5 shows the proposed $T^2_{(s)}$ -based control charts from the second group of 1,000 in-control observations. The control limits were computed using the F -distribution and bootstrap approach, respectively. We specified the nominal FARs from 0.01, 0.05, and 0.1. As illustrated in Fig. 5(a) and (b) for the nominal FARs 0.1 and 0.05 respectively, the control limits from the F -distribution and bootstrap approach generate similar actual FARs. However, in Fig. 5(c) for the nominal FAR 0.01, the control limit from the F -distribution tended to generate higher FAR than the bootstrap approach (FAR with F -distribution: 0.025 and FAR with bootstrap: 0.01). The control limit determined by the bootstrap approach produced similar FAR to the nominal one.

5.2 ARL0 performance evaluation

As mentioned in earlier chapters, ARL0 is defined as the average number of observations required for the control chart to detect a change under the in-control process of Phase I. To compare the competing control charts, we need to determine the better one, corresponding to the higher value of ARL0. Given a Type I error rate (α), which is assumed as equal to 0.01, 0.05, and 0.1, the two competing approaches (using F -distribution and the bootstrap approach) can be compared in terms of ARL0 (in-control ARL). In this study, the ARL0 was calculated using bootstrap techniques. Starting from a group of 1,000 in-control sets of signal as a sample, we simulated 5,000 bootstrap samples, where each sample is selected randomly, with replacement, and with the same size 1,000 as the original sample. For each bootstrap sample, the run length (RL) to detect a change under the healthy condition was calculated.

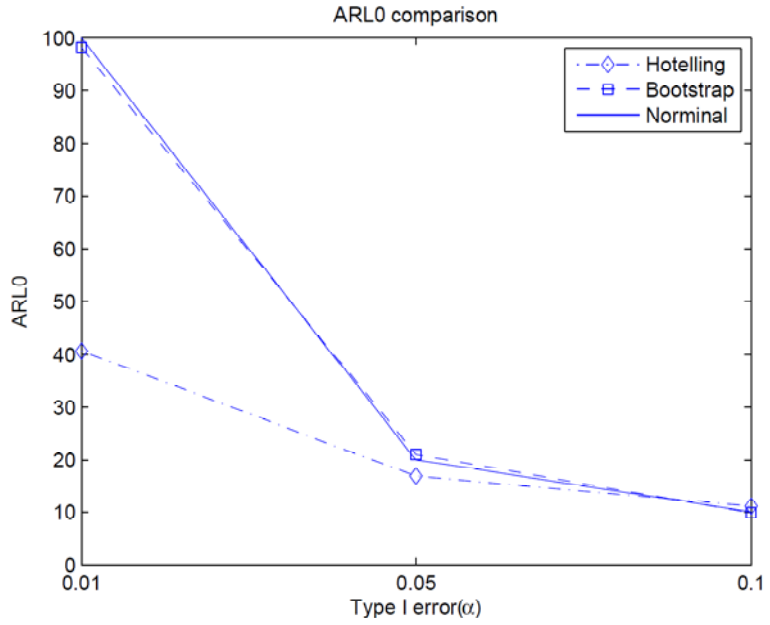


Fig. 7 ARL0 comparison between F-distribution and the bootstrap approach, for different type I error rates

Fig. 6 shows the nominal ARL0 values and the actual ARL0 values as obtained by F -distribution and the bootstrap approach, for different Type I error rate α. This figure shows that across the different approaches, the actual ARL0 values are close to the nominal ARL0 values, when the Type I error rate α is equal to 0.05 and 0.1. However, for the Type I error rate $\alpha = 0.01$, the difference between the actual ARL0, as determined by the F -distribution and the nominal ARL0, increases. In contrast, the bootstrap approach generated similar actual and nominal ARL0 results. This suggests that choosing the bootstrap approach, instead of F-distribution, will reduce the false alarm rate.

5.3 ARL1 performance evaluation

The objective in Phase II of control charting is to quickly detect any change in the process from its in-control state. In this phase of control charting, the competing methods are compared in terms of ARL1, where the run length is defined as the number of samples taken, until an out-of-control signal is issued. In order to evaluate the performance of control charts in this operating phase, the occurrence of assignable causes (faults) were simulated by a total of five out-of-control conditions. Each condition is then characterized by a damage severity level parameter (SL), directly proportional to the damage severity of the out-of-control (i.e., $SL = 0.05,0.1,0.15,0.2,0.25$) introduced in the baseline model (i.e., $SL = 0$ for the in-control state).

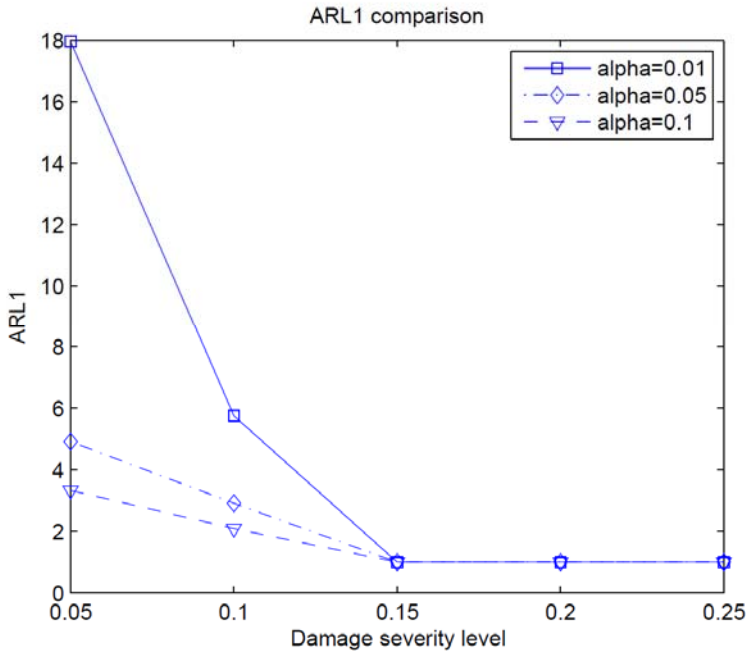


Fig. 8 ARL1 comparison among several type I error rates, for different damage severity levels

Fig. 7 summarizes the simulation results under study. This figure reports the ARL1 performance estimated by computing a set of 5,000 bootstrap run lengths for α=0.01, 0.05, and 0.1, in a similar manner to the case of ARL0 calculation, given that the new sets of signal are simulated according to a specific out-of-control model (i.e., SL = 0.05,0.1,0.15,0.2,0.25). In order to select the better approach in each different study (where each case study is characterized by a specific out-of-control), we need to determine the better method, corresponding to the lower value of ARL1. Fig. 7 shows that the actual ARL1 values with larger α are smaller, as one would expect. More importantly, as SL increases, the actual ARL1 decreases, until SL becomes 0.15. This means that as damage becomes more severe, it is more quickly detected. If the damage severity level is greater than or equal to 0.15, the damages are all detected in average at the first observation, across all different Type I errors. This results indicate that the proposed control chart is very sensitive to the severity of incipient fault in a bearing system.

6. Conclusions

This study introduced a new feature extraction and dimension reduction technique, to alleviate the high dimensionality problem of implementing multivariate statistical process control, when the quality characteristic is a vibration signal from a bearing system. A set of multiscale wavelet scalogram features was generated to reduce the dimensionality of data, and the bootstrapping technique applied as nonparametric density estimation, to set up an upper control limit of the control chart. Our example and numerical simulation of a bearing system demonstrated that the proposed method has satisfactory fault-discriminating ability, without any distributional assumption.

The proposed bearing condition monitoring approach has several desirable attributes. First, the fault detection is conducted in an unsupervised learning mode. That is, data from the damaged system are not needed. The ability to perform the fault detection in an unsupervised learning mode is very important, because data from damaged bearing are typically not available for most real-world practices. Second, the approach presented herein is very attractive for the development of an automated continuous monitoring system because of its simplicity, and because it requires minimal interaction with users. However, it should be pointed out that the procedure developed has only been verified on relatively simple numerical simulations, using a bearing model. To verify that the proposed method is truly robust, it will be necessary to examine many time records, corresponding to a wide range of operational and environmental cases, a wide range of damaged and pristine condition, as well as different damage scenarios.

In this research, we emphasized a damage severity-change problem, but this can be easily extended to damage classification problems. The other direction that this research could be extended to, is to estimate the damage severity level from vibration signals using nonparametric regression techniques, such as support vector regression. Finally, there are several parameters still to be determined, such as the wavelet type, the level of decomposition, and the number of bootstrap samples for density estimation, before the proposed approach can be used effectively. Some theoretical guidelines are needed in future research, to select ‘optimal’ sets of parameters.

Acknowledgements

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A2003787).

References

Bakir, S. (2006), "Distribution-free quality control charts based on signed rank-like statistics", *Commun. Stat. - Theor. M.*, **35**, 743-757.

Bajgier, S.M. (1992), "The use of bootstrapping to construct limits on control charts", *Proceedings of the Decision Science Institute*, San Diego, CA.

Chakraborti, S., Van der Laan, P. and Bakir, S.T. (2001), "Nonparametric control chart: an overview and some results", *J. Quality Technol.*, **33**(3), 304-315.

Chou, Y.M., Mason, R.L. and Young, J.C. (2001), "The control chart for individual observations from a multivariate non-normal distribution", *Commun. Stat. – Simul. C.*, **30**(8-9), 1937-1949.

Cong, F., Chen, J., and Dong, G. (2010), "Research on the order selection of the autoregressive modelling for rolling bearing diagnosis", *Proceedings of the Institution of Mechanical Engineers*, Part C: *J. Mech. Eng. Sci.*, **224**(10), 2289-2297.

Dyer, D. and Stewart, R.M., (1978), "Detection of rolling element bearing damage by statistical vibration analysis", *J. Vib. Acoust.*, **100**(2), 229-235.

Efron, B. (1979), "Bootstrap method: another look at jackknife", *Ann. Stat.*, **7**(1), 1-26.

Fan, J. (1996), "Test of significance based on wavelet thresholding and Neyman's truncation", *J. Am. Stat. Assoc.*, **91**, 674-688.

Ganesan, R., Das, T.K. and Venkataraman, V. (2004), "Wavelet-based multiscale statistical process monitoring: A literature review", *IIE Trans.*, **36**, 787-806.

Goswami, J.C. and Chan, A.K. (1999), *Fundamentals of wavelets: theory, algorithms, and applications*, Wiley, New York, NY.

Hall, P., Poskitt, D.S. and Presnell, D. (2001), "Functional data-analytic approach to signal discrimination", *Technometrics*, **43**(1), 1-9.

Hotelling, H. (1947), *Multivariate quality control, in techniques of statistical analysis*, (Eds. Eisenhart, C., Hastay, M.W. and Wills, W.A.), McGraw-Hill, New York, NY.

Jardine, A.K.S., Lin D. and Banjevic, D. (2006), "A review on machinery diagnostics and prognostics implementing condition-based maintenance", *Mech. Syst. Signal Pr.*, **20**(7), 1483-1510.

Jeong, M.K., Chen, D. and Lu, J.C. (2003), "Thresholoded scalogram and its applications in process fault detection", *Appl. Stoch. Model. Bus.*, **19**, 231-244.

Jones, L.A. and Woodall, W.H. (1998), "The performance of bootstrap control charts", *J. Quality Technol.*, **30**(4), 362-375.

Jung, U., Jeong, M.K. and Lu, J.C. (2006), "A vertical-energy-thresholding procedure for data reduction with multiple complex curves", *IEEE T. Syst. Man. Cy. - B*, **36**(5), 1128-1138.

Jung, U. and Koh, B.H. (2009), "Structural damage localization using wavelet-based silhouette statistics", *J. Sound Vib.*, **321**, 590-604.

Kim, S.H., Alexopoulos, C., Tsui, K.L. and Wilson, J.R. (2007), "A distribution-free tabular CUSUM chart for autocorrelated data", *IIE Trans.*, **39**(3), 317-330.

Koh, B.H., Nagarajaiah, S. and Phan, M.Q. (2008), "Reconstructing structural changes in a dynamic system from experimentally identified state-space models", *J. Mech. Sci. Technol.*, **22**(1), 103-112.

Kresta, J., MacGregor, J.F. and Marlin, T.E. (1991), "Multivariate statistical monitoring of process operating performance", *Can J. Chem. En.*, **69**(1), 35-47.

Ku, W., Storer, R.H. and Georgakis, C. (1995), "Disturbance detection and isolation by dynamic principal component analysis", *Chemomet. Intell. Lab.*, **30**(1), 179-196.
Law, S.S., Li, X.Y., Zhu, X.Q. and Chan, S.L. (2005), "Structural damage detection from wavelet packet sensitivity", *Eng. Struct.*, **27**, 1339-1348.
Lei, Y., He, Z. and Zi, Y. (2011), "EEMD method and WNN for fault diagnosis of locomotive roller bearings", *Expert Syst. Appl.*, **38**(6), 7334-7341.
Li, H., Deng, X. and Dai, H. (2007), "Structural damage detection using the combination method of EMD and wavelet analysis", *Mech. Syst. Signal Pr.*, **21**, 298-306.
Li, Z., Xia, S., Wang, J. and Su, X. (2006), "Damage detection of cracked beams based on wavelet transform", *Int. J. Impact Eng.*, **32**, 1190-1200.
Lin, J. and Zhang, A. (2005), "Fault feature separation using wavelet-ICA filter", *NDT&E Int.*, **38**(6), 421-427.
Lio, Y.L. and Park, C. (2008), "A bootstrap control chart for Birnbaum-Saunders percentiles", *Qual. Reliab. Eng. Int.*, **24**, 585-600.
Liu, R.Y. and Tang, J. (1996), "Control charts for dependent and independent measurements based on bootstrap methods", *J. Am. Stat. Assoc.*, **91**, 1694-1700.
Liu, R.Y., Singh, K. and Teng, J.H. (2004), "DDMA-charts: nonparametric multivariate moving average control charts based on data depth", *Allgemeines Stat. Archiv.*, **88**(2), 235-258.
MacGregor, J.F. and Kourti, T. (1995), "Statistical process control of multivariate processes", *Control Eng. Pract.*, **3**(3), 403-414.
Mallat, S.G. (1989), *A wavelet tour of signal processing*, Academic Press, San Diego.
McFadden, P.D. and Smith, J.D. (1984), "Model for the vibration produced by a single point defect in a rolling element bearing", *J. Sound Vib.*, **96**(1), 69-82.
Mason, R.L. and Young, J.C. (2002), *Multivariate statistical process control with industrial applications*, ASA/SIAM: Philadelphia, PA.
Peng, Z., Chu, F. and He, Y. (2002), "Vibration signal analysis and feature extraction based on reassigned wavelet scalogram", *J. Sound Vib.*, **253**(5), 1087-1100.
Phaladiganon, P., Kim, S.B., Chen, V.C.P., Baek, J.G. and Park, S.K. (2011), "Bootstrap-based T^2 multivariate control charts", *Commun. Stat. – Simul. C.*, **40**, 645-662.
Polansky, A.M. (2005), "A general framework for constructing control charts", *Qual. Reliab. Eng. Int.*, **21**, 633-653.
Prabhakar, S., Mohanty, A.R. and Sekhar, A.S. (2002), "Application of discrete wavelet transform for detection of ball bearing race faults", *Tribol. Int.*, **35**(12), 793-800.
Qiu, P. (2008), "Distribution-free multivariate process control based on log-linear modeling", *IIE Trans.*, **40**(7), 664-677.
Randall, R.B. and Antoni, J. (2011), "Rolling element bearing diagnostics – a tutorial", *Mech. Syst. Signal Pr.*, **25**(2), 485-520.
Rioul, O. and Vetterli, M. (1991), "Wavelets and signal processing", *IEEE Signal Proc. Mag.*, 14-38.
Royston, J.P. (1983), "Sone techniques for assessing multivariate normality based on the Shapiro-Wilk W", *Appl. Stat.*, **32**(2), 121-133.
Scargle, J.D. (1997), *Wavelet methods in astronomical time series analysis, Application of Time Series Analysis in Astronomy and Meteorology*, Chapman & Hall, New York.
Seber, G.A.F. (1984), *Multivariate observations*, Wiley, New York.
Seppala, T., Moskowitz, H., Plante, R. and Tang, J. (1995), "Statistical process control via the subgroup bootstrap", *J. Qual. Technol.*, **27**, 139-153.
Sohn, H., Czarnecki, J.J. and Farrar, C.R. (2000), "Structural health monitoring using statistical process control", *J. Struct. Eng. – ASCE*, **126**, 1356-1363.
Stoumbos, Z.G., Reynolds, M.R., Ryan, T.P. and Woodall, W.H. (2000), "The state of statistical process control as we proceed into the 21st century", *J. Am. Stat. Assoc.*, **95**, 992-998.
Sukchotrat, T., Kim, S.B. and Tsung, F. (2010), "One-class classification-based control chart for

multivariate process monitoring", *IIE Trans.*, **42**, 107-120.
Vidakovic, B. (1999), *Statistical modeling by wavelets*, John Wiley & Sons.
Wang, W. and Wong A.K. (2002), "Autoregressive model-based gear fault diagnosis", *J. Vib. Acoust.*, **124**, 172-179.
Woodall, W.H. (2000), "Controversies and contradictions in statistical process control", *J. Qual. Technol.*, **32**(4), 341-350.
Woodall, W.H. and Montgomery, D.C. (1999), "Research issues and ideas in statistical process control", *J. Qual. Technol.*, **31**(4), 376-386.
Yu, Y., Yu, D. and Junsheng, C. (2006), "A roller bearing fault diagnosis method based on EMD energy entropy and ANN", *J. Sound Vib.*, **294**(1-2), 269-277.

CY