J Intell Manuf (2009) 20:409–419
DOI 10.1007/s10845-008-0115-3

A hybrid approach for identification of concurrent control chart patterns

Chih-Hsuan Wang · Tse-Ping Dong · Way Kuo

Received: 5 February 2008 / Accepted: 13 May 2008 / Published online: 29 May 2008

Abstract Control chart patterns (CCPs) are widely used to identify the potential process problems in modern manufacturing industries. The earliest statistical techniques, including $\bar{X}$ chart and R chart, are respectively used for monitoring process mean and process variance. Recently, pattern recognition techniques based on artificial neural network (ANN) are very popular to be applied to recognize unnatural CCPs. However, most of them are limited to recognize simple CCPs arising from single type of unnatural variation. In other words, they are incapable to handle the problem of concurrent CCPs where two types of unnatural variation exist together within the manufacturing process. To facilitate the research gap, this paper presents a hybrid approach based on independent component analysis (ICA) and decision tree (DT) to identify concurrent CCPs. Without loss of generality, six types of concurrent CCPs are used to validate the proposed method. Experimental results show that the proposed approach is very successful to handle most of the concurrent CCPs. The proposed method has two limitations in real application: it needs at least two concurrent CCPs to reconstruct their source patterns and it may be incapable to handle the concurrent pattern incurred by two correlated process ("upward trend" and "upward shift").

Keywords Concurrent control chart · Pattern recognition · Independent component analysis · Decision tree

Introduction

SPC/EPC (statistical/ engineering process control) techniques play an important role to monitor the process situation and to maintain consistent product quality in modern manufacturing industry. Simply speaking, observed variation of quality characteristics results from either natural variation (common cause) or unnatural variation (assignable cause). Natural variation always exists in the manufacturing process regardless of the fact that how well the product is designed and how adequately the process is maintained. On the contrary, unnatural variation are often associated with specific assignable causes and lead to various types of anomaly patterns. These anomaly patterns often contain valuable information closely relevant to process parameters and process changes. Two quantities are commonly monitored in practice, including the mean and the range of the sample. The earliest techniques developed by Shewhart involve $\bar{X}$ chart (used for monitoring process mean) and R chart (used for monitoring process variance). Once the sources of assignable causes are correctly identified, quality practitioners can remove them and bring the abnormal process back to the normal condition (natural variation). Sometimes, control charts are inappropriately used without sufficient prior knowledge or historical data. Moreover, control charts also show poor performance to recognize different types of unnatural patterns (Guh and Tannock 1999a,b; Yang and Yang 2002; Wang and Kuo 2007).

Similar to most researchers (Cheng 1997; Guh and Tannock 1999a,b; Guh and Hsieh 1999; Guh and Shiue 2005; Pham and Oztemel 1994; Pham and Wani 1997; Yang and

C.-H. Wang (✉)
Department of Business Administration, Ming Chuan University, 250, Sec. 5, Chung Shan N. Rd., Taipei 11103, Taiwan
e-mail: jswang@mail2000.com.tw; chihwang@mcu.edu.tw

T.-P. Dong
Institute of Management, National Taiwan Normal University, Taipei, Taiwan

W. Kuo
Department of Industrial and Information Engineering, University of Tennessee, Knoxville, USA

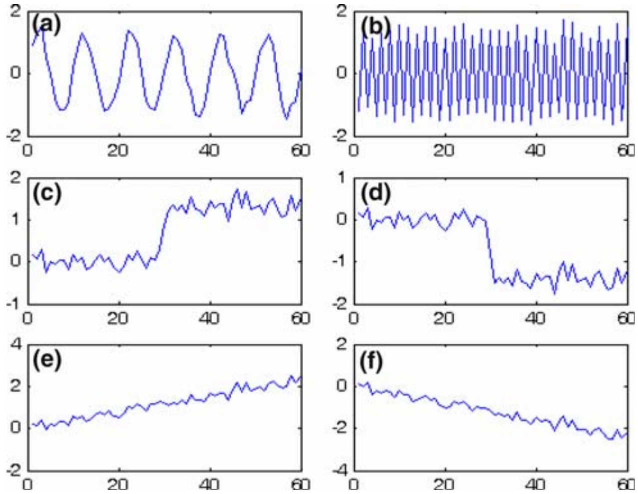


Fig. 1 Simple CCPs ((**a**) cyclic pattern, (**b**) systematic pattern, (**c**) upward shift, (**d**) downward shift, (**e**) upward trend, and (**f**) downward trend)

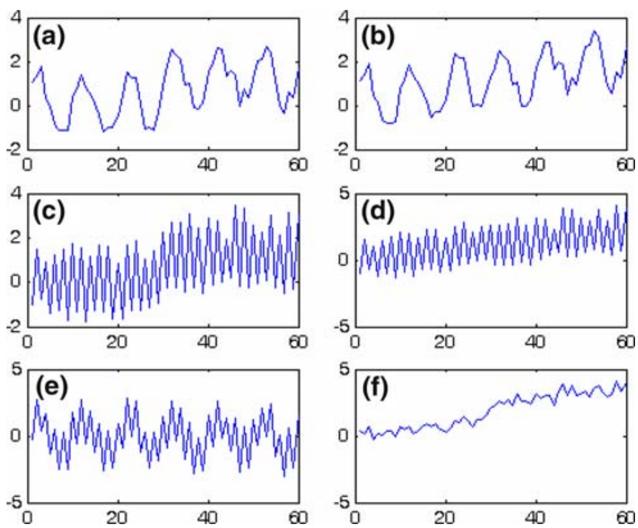


Fig. 2 Concurrent CCPs ((**a**) cyclic+upward shift, (**b**) cyclic+upward trend, (**c**) systematic+upward shift, (**d**) systematic+upward trend, (**e**) cyclic+systematic, (**f**) upward trend+upward shift)

Yang 2002, 2005; Yousef 2004; Wang and Kuo 2007), six common types of simple CCPs are illustrated in this research (see Fig. 1) and their assignable causes are listed below. Without loss of generality, six concurrent CCPs are also generated in Fig. 2.

(1) Trend patterns. A trend can be defined as a continuous movement in either positive or negative direction. Possible causes are tool wear, operator fatigue, equipment deterioration, and so on.
(2) Shift patterns. A shift can be defined as a sudden change above or below the average of the process. This change may be caused by an alternation in process setting, replacement of raw materials, minor failure of machine parts, or introduction of new workers, and so forth.
(3) Cyclic patterns. Cyclic behaviors can be observed by a serial of peaks and troughs occurred in the process. Typical causes are the periodic rotation of operators, systematic environmental changes or fluctuation in the production equipment.
(4) Systematic patterns. The characteristic of systematic patterns is that a point-to-point fluctuation is systematically occurred. It means a low point is always followed by a high point and vice versa. Possible causes include difference between test sets and difference between production lines where product is sampled in rotation.
(5) Concurrent (mixture) patterns. A mixture is usually a combination of observations from separate disturbance or various types of anomalies. Possible causes include items from different suppliers, machines, or workers.

The above-mentioned studies are only capable to handle anomaly CCPs composed of single unnatural variation. Very limited work (Guh and Tannock 1999b; Yang and Yang 2005; Chen et al. 2007; Wang et al. 2007) has been reported on the identification of concurrent process patterns or signals. In practice, concurrent (mixture) process patterns composed of composite unnatural variation may exist and usually result in serious performance degradation for pattern classification. For instance, when the concurrent pattern occurs within the manufacturing process, most existing schemes will force the input to be classified into one of those predefined prototypes that is most similar to the input. Unfortunately, single anomaly type cannot represent the property of the mixture pattern and further tracking of its composite assignable causes also becomes impossible. Hence, the important clue to improve complex process problems via the concurrent CCP is easily missing or seriously biased.

Besides, all of above-mentioned papers do not consider the separation of concurrent CCPs and further reconstruction of their source patterns. In these cases, two concurrent CCPs composed of the same types of source patterns may be categorized into different results. In other words, the performance of pattern classifier is not robust with respect to different mixing coefficients of source patterns. To facilitate the research gap, this paper presents a hybrid approach that incorporates independent component analysis (ICA) and decision tree (DT) to solve the challenging problem. The remainder of this paper is structured as follows. Section "Previous work on CCP recognition" briefly reviews the previous work of CCP recognition and Section "The Proposed techniques" presents the proposed hybrid method. Computer synthetic results are conducted in Section "Computer synthetic results" and conclusions are drawn in Section "Conclusions".

Previous work on CCP recognition

Traditionally, there are two basic approaches developed in fault detection system: a model-based (so-called parametric) approach and a feature-based approach (Jin and Shi 2001). In the former approach, observations are assumed to follow a specific process distribution and hence sufficient data samples need to check the distribution characteristics in advance. Obviously, sufficient historical information of fault models is usually unavailable at the beginning of the manufacturing process. On the other hand, a feature-based approach is more flexible to deal with a complex process problem, especially when no prior information is available. Features could be obtained in various forms, including principal component analysis (Chen and Liu 2001; Yoon and MacGregor 2004), independent component analysis (Kano et al. 2003; Lee et al. 2004), and Fourier or wavelet transformation (Jin and Shi 2001; Yousef 2004; Chen et al. 2007). Transformed features cannot only reduce the size of input dimension but also decompose the complex functional relationship between time-series data and the associated manufacturing process. However, they are very difficult to be understood in practice and also need long computation.

Recently, fast computational developments in artificial intelligence or expert systems also motivate researchers to adopt the artificial neural network (ANN) based CCP classifier. ANN can be simply classified into two categories, including supervised ANN and unsupervised ANN. Most researchers (Cheng 1997; Chiu et al. 2003; Cook and Chiu 1998; Guh and Tannock 1999a,b; Guh and Hsieh 1999; Pham and Oztemel 1994; Yang and Yang 2002) use supervised paradigms, including multi-layer perceptron (MLP), radial basis function (RBF), and learning vector quantization (LVQ), to classify different types of CCPs. Other researchers (Guh and Shiue 2005; Pacella et al. 2004; Wang et al. 2007; Wong et al. 2006) use unsupervised neural networks, involving self organized maps (SOM) and adaptive resonance theory (ART), to fulfill the same objective. Due to its black-box property, the main criticism of using ANN is that its topology and structure cannot be systematically determined. In addition, the speed of adjusting network configuration for supervised ANN is quite slow and hence infeasible for on-line quality practitioners. Recently, a decision tree (DT) based classifier is also popular for the problem of CCP recognition (Pham and Wani 1997; Gauri and Chakraborty 2006, 2007; Guh 2005; Wang et al. 2008).

Furthermore, some hybrid methods are proposed to solve specific problems. Wong et al. (2006) use high-order statistical features with SOM network for the problem of multi-sensor condition monitoring. Lu et al. (2008) discuss the development of ANN with independent component analysis to identify the disturbance and recognize shifts in correlated process. Wang and Kuo (2007) incorporates wavelet filtering and robust fuzzy clustering to enhance the tolerance capability of process variance deviation. Most of the above-mentioned studies are capable to recognize single type of unnatural CCP. By contrast, the studies of concurrent CCPs incurred by two types of unnatural variation are relatively limited (Guh and Tannock 1999a,b; Yang and Yang 2005; Chen et al. 2007; Wang et al. 2007).

Guh and Tannock (1999b) uses a back-propagation ANN to recognize concurrent CCPs. Chen et al. (2007) also presents a hybrid approach by integrating wavelet method and back-propagation ANN for on-line recognition of concurrent CCPs. Wang et al. (2007) proposes a hybrid system that incorporates wavelet filtering and ART (adaptive resonance theory) to discover concurrent process signals. Instead of using ANN based classifier, Yang and Yang (2005) use a statistical correlation approach to speed up the recognition of concurrent CCPs. In this study, a hybrid approach (see Fig. 3) is proposed for the recognition of concurrent CCPs. For the purpose of fast computation, simple CCP shape features (see Section "Features of simple CCP") are used in this study for the input of DT.

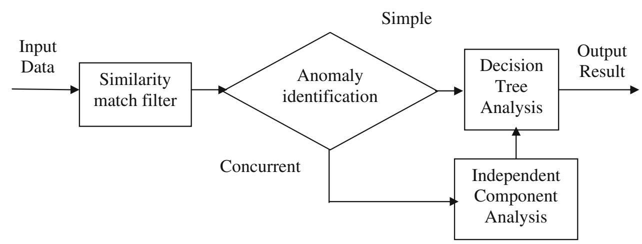


Fig. 3 The proposed hybrid approach

The proposed techniques

The proposed approach consists of three main steps: (1) the similarity measure between the input and the reference composed of simple CCPs is used to determine that the input belongs to "simple" anomaly or "concurrent" anomaly, (2) if the input belongs to the concurrent type, an ICA based separation will be applied and its source patterns can be reconstructed, (3) use of DT classifier with appropriate CCP features to identify its specific anomaly type. Based on the inner product, Al-Ghanim and Ludeman (1997) evaluated the correlation between the input and various reference vectors. Similarly, Yang and Yang (2005) use the statistical correlation coefficient (see Eq. 1) to determine the anomaly type of the input pattern.

$$corr = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}}, \tag{1}$$

where x_t, $\bar{x}(y_t, \bar{y})$ respectively denotes the input (reference) vector and its mean and n is the total length of the observing window. This study also utilizes the statistical correlation coefficient as the similarity measure between the input and various reference prototypes.

Features of simple CCP

Some statistical features such as mean, standard deviation, skewness, kurtosis, and autocorrelation, are adopted in Hassan et al. (2003) to improve the performance of ANN based classifier. Specifically, skewness provides the information regarding to the degree of asymmetry and kurtosis measures the relative peakness or flatness of its distribution. Their mathematical forms are respectively shown below.

$$mean = \frac{\sum_{t=1}^{n} x_t}{n}, \tag{2}$$

$$std = \sqrt{\frac{\sum_{t=1}^{n}(x_t - mean)^2}{n}}, \tag{3}$$

$$skew = \frac{\sum_{t=1}^{n}(x_t - mean)^3}{n(std)^3}, \tag{4}$$

$$kurt = \frac{\sum_{t=1}^{n}(x_t - mean)^4}{n(std)^4}, \tag{5}$$

where n is and x_t are similarly defined above.

Due to lack of intuitive meaning, statistical features are difficult to be applied for the input of DT to identify anomaly CCPs (Pham and Wani 1997). By observing their shape characteristics of CCPs, five CCP features are selected and used as our discriminators in this paper, including *LS-SLP*

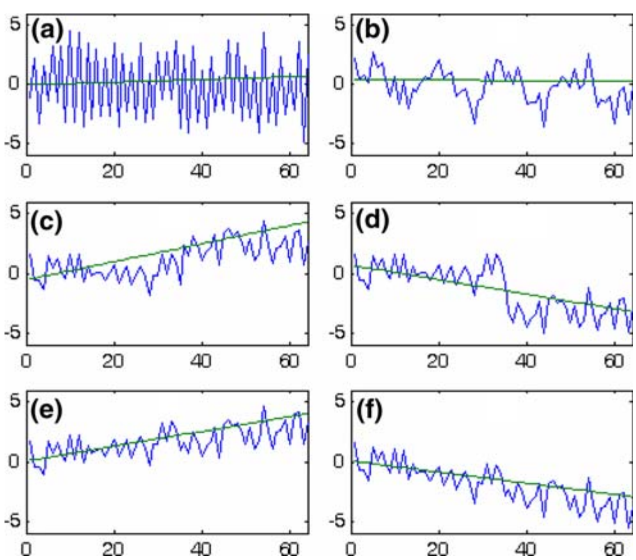


Fig. 4 The least-square regression of simple CCPs

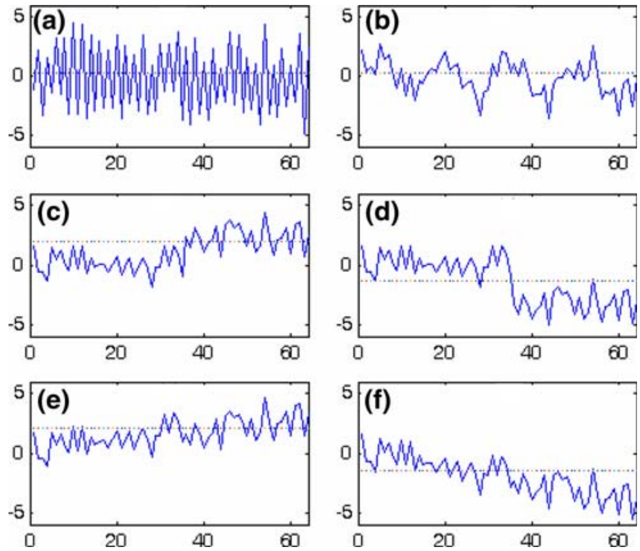


Fig. 5 The mean regression of simple CCPs

(the slope of the least-square regression), *LS-ERR* (the sum of the least-square error), *MN-ERR* (the sum of the mean error), *LS-CROS* (the number of crossings between the original pattern and its least-square regression), and *MN-CROS* (the number of crossings between the original pattern and its mean regression). Their mathematical forms are shown from Eq. 6 to 8 and detail explanation is supplied later. Figure 4 shows the generation of six simple CCPs and their least-square regression (see the solid lines) and Fig. 5 shows their mean regression (see the dashed lines). For simplicity, the labels of simple CCPs shown in two figures are denoted as: (a) systematic pattern, (b) cyclic pattern, (c) upward shift, (d) downward shift, (e) upward trend, and (f) downward trend.

$$LS - SLP = \frac{\sum_{t=1}^{n}(t - \bar{t})(x_t - \bar{x})}{\sum_{t=1}^{n}(t - \bar{t})^2}, \tag{6}$$

$$LS - ERR = \sum_{t=1}^{n}(x_t - \hat{x}_{LS}), \tag{7}$$

$$MN - ERR = \sum_{t=1}^{n}(x_t - \hat{x}_{MN}), \tag{8}$$

where $\hat{x}_{LS}$ and $\hat{x}_{MN}$ respectively denote the estimated signal based on least-square/mean regression.

(1) *LS-SLP*: the slope of the least-square line representing the pattern. The magnitude of the slope for both natural and cyclic patterns is approximately zero, while that for trend or shift patterns is obviously greater than zero. Therefore, *LS-SLP* may be a good candidate to differentiate natural and cyclic patterns from trend and shift patterns.

(2) *LS-CROS*: the number of the least-square crossings. Not surprisingly, *LS-CROS* is highest for natural and trend patterns, intermediate for shift patterns and lowest for cyclic patterns. Similarly, *LS-CROS* may be suitable to separate natural and trend patterns from other patterns.
(3) *MN-CROS*: the number of the mean crossings. *MN-CROS* is lowest for shift patterns, intermediate for cyclic and trend patterns and highest for natural patterns.
(4) *LS-ERR*: the sum of the least-square regression error. Natural and trend patterns have lowest least-square error, shift patterns have intermediate least-square error while cyclic patterns have highest least-square error.
(5) *MN-ERR*: the sum of the mean regression error. Natural patterns have lowest mean error; cyclic and trend patterns have intermediate mean error while shift patterns have highest mean error.

Very interestingly, the *LS-SLP* of both systematic pattern and cyclic pattern is approximately zero and much smaller than the absolute value of other four types of CCPs. Besides, the *LS-CROS* is the largest for the systematic pattern but the smallest for the cyclic pattern. The *MN-CROS* is the smallest for the shift pattern. And both *LS-ERR* and *MN-ERR* are the smallest for the trend pattern. Intuitively, these five selected features have more or less discriminative power for CCP classification.

Overview of independent component analysis (ICA)

A long-standing problem in statistics is how to find a good representation of multivariate data. Representation here means that some data transformation may be needed to make the dataset more visible. Principal component analysis (PCA) and its closely related Karhunen–Loève (KL) transform are two classic techniques in statistical data analysis, feature extraction, and data compression. PCA is also used for dimension reduction since it can effectively extract linear structure from high-dimensional data by projecting the data onto a lower dimensional subspace which contains most of the variance of the original data. However, PCA has serious drawbacks when the inherent data structure is nonlinear. To solve this problem, many methods have been developed for finding hidden/ latent factors that underlie sets of random variables or measurements, including nonlinear PCA or ICA. The connection between PCA and ICA involve decorrelation, whitening, and sphering (Hyvärinen et al. 2001).

ICA could be looked from a different viewpoint of finding a good representation of the dataset and this was closely related to the interesting "cocktail-party" problem or so-called "blind source separation" in the field of signal processing (Hyvärinen and Oja 2000). Its main idea is to utilize some information or statistical properties of the concurrent signals to estimate their mixing coefficients. The cock-tail party problem can be simply described as follows. Imagine that you are in a room where two people are speaking simultaneously. Two microphones held in different locations give you two recorded time signals denoted by $x_1(t)$ and $x_2(t)$. Each of these recorded signals is a weighted sum of the source signals emitted by two speakers, which can be denoted by $s_1(t)$ and $s_2(t)$. For convenience, we can express this scenario as the following linear equations:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t), \tag{9}$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t), \tag{10}$$

where $a_{11}, a_{12}, a_{21}, a_{22}$ are mixing parameters that depend on the distances of the microphones from the speakers. For a realistic industry application, two microphones can be replaced by two measuring sensors and different speakers represent various types of anomaly process pattern. It would be very nice if you could now estimate two original source signals using only the recorded mixture signals. If you know the mixing parameters, we can easily solve the problem. However, the problem is considerably more difficult in practice since you do not know the mixing parameters in advance.

For convenience, it is assumed that p measured variables $x_1, x_2, \ldots, x_p$ can be expressed as a linear combination of m unknown independent components $s_1, s_2, \ldots, s_m$ (where $m \leq p$). The relationship between them can be simplified as follows.

$$X = AS + E, \tag{11}$$

where $X = [x(1), x(2), \ldots x(n)] \in R^{p\times n}$ is the data matrix, $A = [a_1, a_2, \ldots, a_m] \in R^{p\times m}$ is the mixing matrix, $S = [s(1), s(2), \ldots, s(n)] \in R^{m\times n}$ the independent component matrix, $E \in R^{p\times n}$ is the residual or error matrix, and n is the dimension of measured data samples. When the property $(p = m)$ holds, the mixing matrix will become a square matrix. The basic idea of ICA is to estimate both the mixing matrix A and the independent components S from only the observed mixture data X. Alternatively, if we can find a demixing matrix $W \in R^{m\times p}$, the reconstructed source signals $\widehat{S} \in R^{m\times n}$ can be represented as:

$$\widehat{S} = WX, \tag{12}$$

However, to estimate the demixing matrix without any prior knowledge of the mixing process will be quite difficult. One approach to solve this problem is using some information on the statistical independence of the source signals to estimate the mixing parameters.

For simplicity, we assume p equals m from now on. The initial step in ICA is whitening, or so called sphering, which removes all cross-correlation between random variables. Consider a p dimensional random vector $x(t)$ at the kth sample, the eigen-decomposition of the covariance R_x could be given by:

$$R_x = E(x(k)x(k)^T) = UDU^T, \tag{13}$$

where E represents the expectation operator, D is a diagonal matrix composed of its eigenvalues, and U is an orthogonal matrix composed of its eigenvectors .

And the whitening transformation is expressed as

$$z(k) = Qx(k), \tag{14}$$

where $Q = D^{-1/2}U^T$. One can easily verify that $R_z = E(z(k)z(k)^T)$ becomes an identity matrix. After the whitening transformation, we have the following:

$$z(k) = Qx(k) = QAs(k) = Bs(k), \tag{15}$$

where B is an orthogonal matrix verified by the following:

$$E(z(k)z(k)^T) = BE(s(k)s(k)^T)B^T = BIB^T = I. \tag{16}$$

Currently, the problem of finding an arbitrary full-rank matrix A is successfully reduced to the simpler problem of finding an orthogonal matrix B since B has fewer parameters to estimate, which gives source signals as follows

$$\hat{s}(k) = B^T z(k) = B^T Qx(k), \tag{17}$$

From Eqs. 12 and 17, the relation between the demixing matrix W and the orthogonal matrix B can be expressed as:

$$W = B^T Q, \tag{18}$$

To calculate B matrix, it is initialized and gradually updated so that the reconstructed $\hat{s}(k)$ will have greater non-Gaussianity than the mixture pattern. Based on the central limit theorem, non-Gaussianity can also imply independence by Hyvärinen and Oja (2000). Specifically, there are various measures of non-Gaussianity, including the kurtosis, negentropy, and mutual information. In particular, Gaussian random variables will have zero kurtosis, largest entropy, and maximum mutual information. Hence, searching for a good W matrix can be alternatively achieved by maximizing kurtosis $K(y)$, negentropy $J(y)$ and mutual information $I(y)$.

To sum up, there are two principles for ICA estimation: nonlinear decorrelation and maximally non-Gaussian components. The independence property used in ICA is stronger than the uncorrelatedness used in PCA since "independence" also implies "nonlinear uncorrelatedness". Another very intuitive point of finding independent components is maximum non-Gaussianity. Now, ICA has been widely applied in many areas, such as denoising, feature extraction or data compression, signal or image separation, and time-series prediction. In brain imaging, the sensors outside the head can measure mixed signals that are emitted by different brain areas and ICA can be applied to separate their source signals. In econometrics or financial area, parallel time series may reveal different components hidden in the dataset and ICA can be used to obtain a good insight. For further details, interested readers can refer to Hyvärinen et al. (2001).

Principle of the decision tree (DT)

DT is a simple yet powerful model since her hierarchical structure is easy to be understood and incorporated with human knowledge. Today, DT is widely used for supervised classification in practice, ranging from medical diagnosis to credit evaluation. Generally speaking, tree models are usually divided into regression trees (when the response variable is quantitative continuous) and classification trees (when the response variable is qualitative categorical or quantitative discrete). When DT comes to classification trees, there are three major algorithms used in practice, including CART ("Classification and Regression Trees"), C4.5, and CHAID ("Chi-square Automatic Interaction Detection"). The famous CART algorithm was developed by the statistical community (Breiman et al. 1984). Another famous CHAID algorithm was also developed by the statistical community (Kass 1980). By contrast, C4.5 and its later version, C5.0 (Quinlan 1993), were very popular among computer scientists.

All three algorithms start from the root node, create classification rules by constructing a tree-like structure in a top-down and divide-and-conquer manner, and also restrict the size of the resulting tree to avoid the problem of over-fitting. The main difference can be shortly described as follows. CHAID achieve its goal by testing a statistical stopping rule that prevents tree growth. In contrast, both CART and C4.5 first grow the full tree and then prune it back to a reasonable size. Another difference between CART and the other two algorithms is that CART only allows "binary" splitting whereas the other two allow "multiple" splitting. Moreover, CART decides the split by calculating the amount of homogeneity within class but CHAID tests a hypothesis regarding the dependence between the splitting variable and the response variable. In brief, CART tends to be more suitable for prediction whereas CHAID seems to more powerful for data segmentation. Owing to its fast computation, C4.5 is adopted in this study to classify the specific type of simple CCP.

C4.5 is an extended form of Iterative Dichotomizer 3 (ID3) with additional characteristics such as the ability to handle continuous attribute, noisy data, and alternative measures for selecting attributes and pruning decision trees (Quinlan 1983, 1986). Guh and Shiue (2005) indicated the three phases to conduct classification rules for C4.5. For simplicity, we assume that the dataset S consists of s data samples to introduce the construction process of DT. Let the class label C_i has m distinct values denoting m distinct classes and let s_i denote the number of samples in class C_i. Using the information theory, the expected information (so-called entropy) needed to classify a given sample can be defined by

$$I(S) = -\sum_{i=1}^{m} p_i \log_2 p_i, \tag{19}$$

where $p_i = s_i/s$ is the probability that an sample belongs to class C_i. Suppose attribute A has n distinct values, $\{a_1, \ldots, a_n\}$, and it can be used to partition the dataset S into n subsets, $\{S_1, \ldots, S_n\}$, where S_j represents those samples in S that have the attribute value a_j of A. Let s_{ij} denote the number of samples from class C_i in subset S_j. If attribute A is selected as the best attribute for splitting the current node, this attribute should have the largest information gain or the greatest entropy reduction. The expected information based on the partitioning into attribute A's subsets is given by

$$I(A) = \sum_{j=1}^{n} \frac{s_j}{s} I(S_j), \tag{20}$$

where the term s_j/s calculated by the number of samples in S_j divided by the total number of samples in S acts as the weight of the *jth* subset. For any given subset S_j,

$$I(S_j) = -\sum_{i=1}^{m} p_{ij} \log_2 p_{ij}, \tag{21}$$

where $p_{ij} = s_{ij}/s_j$ is the probability that a sample in S_j belongs to the class C_i. Therefore, the information gain due to branching on attribute A can be described as: $G(A) = I(S) - I(A)$. In other words, $G(A)$ is the expected entropy reduction caused by selecting attribute A. Basically, the attribute with the highest information gain will be selected as the best discriminator for the current node and the recursive process will be continued until all samples in each leaf node belong to the same class.

Note, splitting at each internal node represents a test on an attribute and this process will be terminated until all samples in the leaf node belong to the same class. When DT is initially built in this manner, many branches inside the tree also reflect a lot of noisy data or outliers originated from the training dataset. This phenomenon is so-called overfitting and hence tree pruning approaches are needed to remove unnecessary branches. There are two common methods proposed for tree pruning: "pre-pruning" is achieved by halting the tree construction early and "post-pruning" removes branches from a fully grown tree. In this paper, pre-pruning is used to save the computation time and its details will be described in Section "DT based CCP identification".

Computer synthetic results

Pattern generation

Six simple CCPs are generated in Fig. 1, including cyclic pattern, systematic pattern, upward shift, downward shift, upward trend and downward trend. Without loss of generality, six concurrent CCPs originated from simple CCPs are also generated in Fig. 2. Specifically, the amplitude of cyclic patterns, the magnitude of shift patterns and the slope of trend patterns are randomly determined within a specific range. The random setting in shift quantity of those anomaly CCPs is for the purpose of testing the adaptive capability of the proposed method. All details to generate simple and concurrent CCPs are shown below.

(a) Natural pattern:

$$x(t) = n(t) \sim N(0, 1), \tag{22}$$

where $x(t)$ is a sample at time t (from a standard Gaussian distribution).

(b) Cyclic pattern:

$$x(t) = n(t) + a \sin(2\pi t/T), \tag{23}$$

where $1.5 \leq a \leq 3$ and $T = 15$ respectively denote the amplitude and the period of cyclic patterns.

(c) Systematic pattern:

$$x(t) = n(t) + (-1)^t s, \tag{24}$$

where $1.5 \leq s \leq 3$ denotes the magnitude of shift.

(d) Upward shift/downward shift:

$$x(t) = n(t) \pm su(t - t_h), \tag{25}$$

where $u(t)$ stands for a unit step function shown below.

$$u(t - t_h) = \begin{cases} 0, & t < t_h \\ 1, & t \geq t_h, \end{cases} \tag{26}$$

(e) Upward trend/downward trend:

$$x(t) = n(t) \pm dt, \tag{27}$$

where $0.05 \leq d \leq 0.1$ stands for the slope of trend patterns.

(f) Concurrent/Mixture patterns:

$$x(t) = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} S_1(t) \\ S_2(t) \end{bmatrix}, \tag{28}$$

Table 1 The kurtosis measure w.r.t. simple CCPs

	Cyclic	Systematic	Upshift	Downshift	Uptrend	Downtrend
Kurtosis	1.5943	1.0408	1.1877	1.1832	1.7479	1.8049

where $S_1(t), S_2(t)$ represents two source patterns composed of simple CCPs.

ICA based separation

PCA is based on second-order statistics but ICA uses high-order statistical cumulants (i.e. kurtosis) to analyze the dataset. Two zero-mean random variables are said to be uncorrelated if their covariance is zero.

$$COV(X, Y) = E[(X - \mu_X)(Y - \mu_Y)^T] = 0, \quad (29)$$

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)], \quad (30)$$

where $COV(X, Y)$ is the covariance and μ_X, μ_Y, respectively denote their expectation values. By taking the form of $g(X) = X$ and $h(Y) = Y$, "independence" will automatically imply "uncorrelatedness". According to the central limit theorem, non-Gaussianity is a good indicator of independence for two mixed signals. Therefore, concurrent CCPs originated from mixing simple CCPs are closer to Gaussian than their original simple patterns. In this study, a kurtosis based fixed-point algorithm (Hyvärinen and Oja 1997) is used for the reason of fast computation. Note, kurtosis is zero for a Gaussian random variable but kurtosis is positive (supergaussian) or negative (subgaussian) for most non-Gaussian random variables.

Based on Table 1, six simple CCPs have nonzero kurtosis and this implies that they are nearly non-Gaussian by nature. A further independence test of the concurrent CCPs is conducted in Table 2. If we compare the 3rd row and the 4th row of Table 2, the mixture kurtosis after mixing is less than the sum of the original kurtosis before mixing. This phenomenon meets the central limit theorem. By observing the second row of Table 2, the covariance of the mixture composed of "upshift" and "uptrend" is much larger than the other entries. This also implies that both "upshift" and "uptrend" are two correlated process and hence ICA based separation may be ineffective on this case. Not surprisingly, except in the category of "upward trend + upward shift", other categories of concurrent CCPs have been successfully separated and reconstructed by ICA (see Figs. 6 and 7).

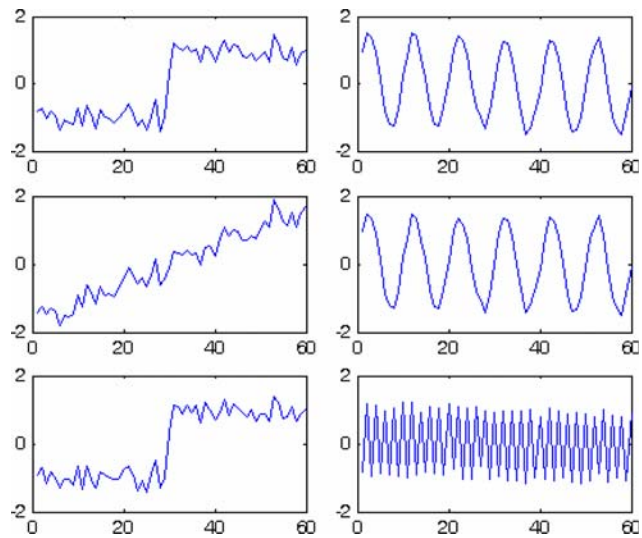


Fig. 6 Separation of concurrent CCPs (upshift + cyclic, uptrend + cyclic, upshift + systematic)

DT based CCP identification

Five CCP features involving *LS-SLP*, *LS-CROS*, *MN-CROS*, *LS-ERR* and *MN-ERR* are selected for identification of CCPs and their typical values are conducted in Table 3. Since each feature has more or less discriminative power to separate various types of CCPs, the optimal priority of five features along the DT could be achieved by comparing their information entropy. In simple words, the best feature at each node of the DT for branch splitting is selected by searching for the feature with the maximal information gain (see the underline marked in each entry of Table 4). For convenience, the number denoted in the first column of Table 4 represents the level of the decision tree (the root node is denoted by level 0).

Owing to its maximal information gain (1.59), *LS-SLP* is first selected as the discriminative feature at the root node (level 0) and then the DT is accordingly separated into three subgroups: zero-slope (comprising systematic & cyclic patterns), positive-slope (comprising positive trend & shift patterns), and negative-slope (comprising negative trend & shift

Table 2 The independence test w.r.t. concurrent CCPs

	Cyclic + Upshift	Cyclic + Uptrend	Systematic + Upshift	Systematic + Uptrend	Cyclic + Systematic	Upshift + Uptrend
Covariance	0.025954	0.067299	0.092926	0.075434	0.067299	0.42232
Mixture kurtosis	2.2921	2.4051	1.652	1.5398	1.9512	1.3064
Sum of original kurtosis	2.782	3.3422	2.2285	2.7887	2.635	2.9356

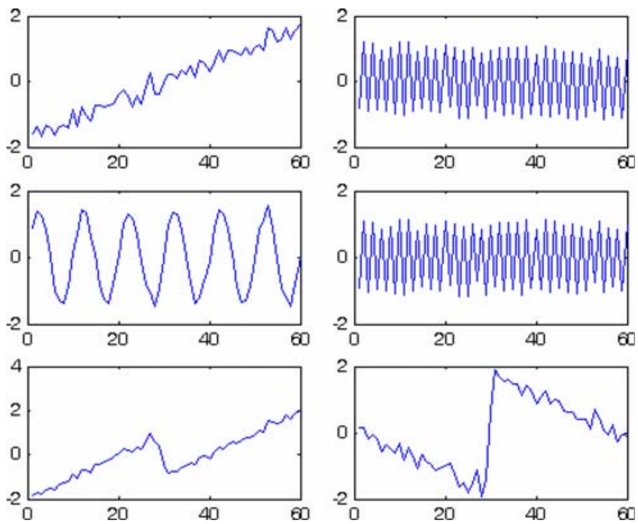


Fig. 7 Separation of concurrent CCPs (uptrend + systematic, cyclic + systematic, uptrend + upshift)

Table 3 Typical values of five CCP features

	LS-SLP	*LS-CROS*	*MN-CROS*	*LS-ERR*	*MN-ERR*
Cyclic	−0.0111	18.82	19.32	116.94	120.24
Systematic	0.0004	62.59	63.25	583.65	584.78
Up shift	0.0708	23.63	9.38	90.627	181.94
Down shift	−0.0711	24.03	9.095	88.404	180.43
Up trend	0.055	30.15	18.755	58.931	114.68
Down trend	−0.055	30.15	19.185	58.931	115.26

Table 4 Information gain of five CCP features

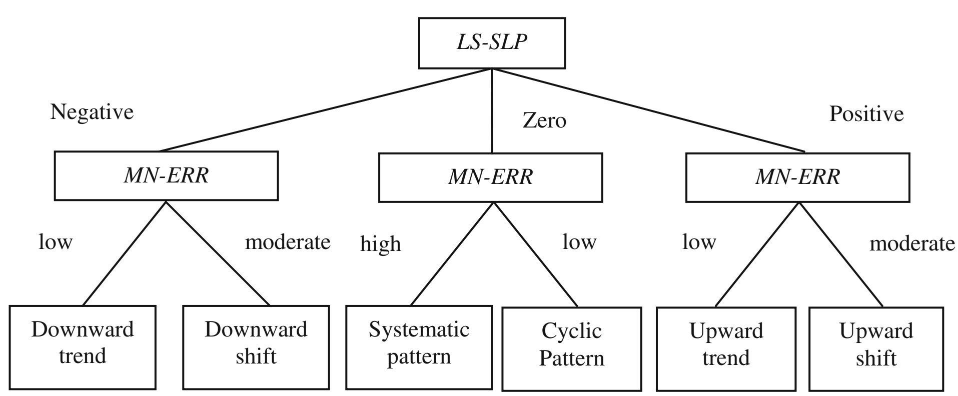
	LS-SLP	*LS-CROS*	*MN-CROS*	*LS-ERR*	*MN-ERR*
Level 0: Root node	1.59	0.98	0.99	1.16	1.12
Level 1-1: Zero-slope	N/A	1	1	1	1
Level 1-2: Positive-slope	N/A	0.23	0.45	0.48	0.61
Level 1-3: Negative-slope	N/A	0.38	0.53	0.42	0.7

patterns). Note, *LS-SLP* selected at level 0 could not be reselected at the next level (N/A denotes "not acceptable"). Similarly, *MN-ERR* is selected to discriminate the shift pattern from the trend pattern at both level 1–2 and level 1–3. Because all features of the systematic pattern are much different from the cyclic pattern, four features have the same discriminative power at level 1–1 except *LS-SLP*. For convenience, *MN-ERR* is also used within the zero-slope group to distinguish the systematic pattern from the cyclic pattern. Finally, the construction of the DT for CCP identification can be conducted in Fig. 8. In fact, a couple of "*If-Then*" rules embedded in the DT could be easily read. For instance, if the negative *LS-SLP* and the low *MN-ERR* hold simultaneously, the input sample will be classified into the category of "downward trend" (also see Fig. 8).

Conclusions

Concurrent (mixture) CCPs incurred by two types of unnatural variation together usually exist within the manufacturing process. However, most existing schemes will force the concurrent pattern to be classified into one of those predefined categories that represent single type of unnatural variation. This action will cause great difficulties in tracking the assignable causes hidden in the observed mixtures. Moreover, they do not consider the separation of concurrent CCPs or further reconstruct their corresponding source patterns. As a consequence, the performance of pattern classifier is not robust with respect to different mixing coefficients and the results may be inconsistent. In other words, two concurrent CCPs that comprise the same types of source patterns may be categorized into different results. In this research, a hybrid approach that incorporates ICA and DT is proposed to separate concurrent CCPs, to reconstruct their source patterns, and to identify the specific anomaly types. The proposed method has two limitations in real application: it needs at least two mixed CCPs to reconstruct their source patterns and it may be incapable to handle the concurrent patterns

Fig. 8 DT construction to identify anomaly CCPs



arising from two correlated process (i.e. "upward trend" and "upward shift"). Experimental results show that the proposed approach is very successful in most concurrent CCPs. More importantly, the presented method is very promising to be applied to other temporal data, such as the separation of composite biomedical signals (Graupe et al. 2007) or composite financial time-series (Kiviluoto and Oja 1998).

Acknowledgements The authors would thank two anonymous referees for their helpful comments and suggestions. This paper is financially supported by Taiwan National Science Council.

References

AI-Ghanim, A. M., & Ludeman, L. C. (1997). Automated unnatural pattern recognition on control charts using correlation analysis techniques. *Computers & Industrial Engineering, 32*(3), 679–690. doi:10.1016/S0360-8352(96)00214-8.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth.

Chen, J., & Liu, J. (2001). Derivation of function space analysis based PCA control charts for batch process monitoring. *Chemical Engineering Science, 56*, 3289–3304. doi:10.1016/S0009-2509(01)00004-5.

Chen, Z., Lu, S., & Lam, S. (2007). A hybrid system for SPC concurrent pattern recognition. *Advanced Engineering Informatics, 21*, 303–310. doi:10.1016/j.aei.2007.03.002.

Cheng, C. S. (1997). A neural network approach for the analysis of control chart patterns. *International Journal of Production Research, 35*(3), 667–697. doi:10.1080/002075497195650.

Chiu, C. C., Shao, Y. E., Lee, T. S., & Lee, K.M. (2003). Identification of process disturbance using SPC/EPC and neural networks. *Journal of Intelligent Manufacturing, 14*, 379–388. doi:10.1023/A:1024657911399.

Cook, D. F., & Chiu, C. C. (1998) Using radial basis function neural networks to recognize shifts in correlated manufacturing process parameters. *IIE Transactions, 30*, 227–234.

Gauri, S. K., & Chakraborty, S. (2006). Feature based recognition of control chart patterns. *Computers & Industrial Engineering, 51*, 726–742. doi:10.1016/j.cie.2006.07.013.

Gauri, S. K., & Chakraborty, S. (2007). A study on the various features for effective control chart pattern recognition. *International Journal of Advanced Manufacturing Technology, 34*(3), 385–398. doi:10.1007/s00170-006-0591-6.

Graupe, D., Zhong, Y., & Graupe, M. H. (2007). Extracting fetal from maternal ECG for early diagnosis: Theoretical problems and solutions—BAF and ICA. In *Proceeding of the International Conference on Biomedical Engineering* (pp. 352–356). Innsbruck, Austria.

Guh, R. S. (2005). A hybrid learning-based model for on-line detection and analysis of control chart patterns. *Computers & Industrial Engineering, 49*, 35–62. doi:10.1016/j.cie.2005.03.002.

Guh, R. S., & Hsieh, Y. C. (1999). A neural network based model for abnormal pattern recognition of control charts. *Computers & Industrial Engineering, 36*, 97–108. doi:10.1016/S0360-8352(99)00004-2.

Guh, R. S., & Shiue, Y. R. (2005). On-line identification of control chart patterns using self-organized approaches. *International Journal of Production Research, 43*(5), 1225–1254. doi:10.1080/00207540420002688884.

Guh, R. S., & Tannock, J. D. T. (1999a). A neural network approach to characterize pattern parameters in process control charts. *Journal of Intelligent Manufacturing, 10*, 449–462. doi:10.1023/A:1008975131304.

Guh, R.S., & Tannock, J.D.T. (1999b). Recognition of control chart concurrent patterns using a neural network approach. *International Journal of Production Research*, *37*(8), 1743–1765.

Hassan, A., Shariff, M. N. B., Shaharoun, A. M., & Jamaludin, H. (2003). Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research, 41*(7), 1587–1603. doi:10.1080/0020754021000049844.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: Wiley.

Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation, 9*(7), 1483–1492. doi:10.1162/neco.1997.9.7.1483.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*, 411–430. doi:10.1016/S0893-6080(00)00026-5.

Jin, J., & Shi, J. (2001). Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. *Journal of Intelligent Manufacturing, 12*, 257–268. doi:10.1023/A:1011248925750.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., & Ohno, H. (2003). Monitoring independent components for fault detection. *Journal of American Institute Chemical Engineering, 49*(4), 969–976.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*, 119–127. doi:10.2307/2986296.

Kiviluoto, K., & Oja, E. (1998). Independent component analysis for parallel financial time series. In *Proceeding of the International Conference on Neural Information Processing* (Vol. 2, pp. 895–898). Tokyo, Japan.

Lee, J. M., Yoo, C. K., & Lee, I. B. (2004). Statistical process monitoring with independent component analysis. *Journal of Process Control, 14*, 467–485. doi:10.1016/j.jprocont.2003.09.004.

Lu, C. J., Wu, C. M., Keng, C. J., & Chiu, C. C. (2008). Integrated application of SPC/EPC/ICA and neural networks. *International Journal of Production Research, 46*(4), 873–893.

Pacella, M., Semeraro, Q., & Anglani, A. (2004). Adaptive resonance theory-based neural algorithms for manufacturing process quality control. *International Journal of Production Research, 42*(21), 4581–4607. doi:10.1080/00207540410001715706.

Pham, D. T., & Oztemel, E. (1994). Control chart pattern recognition using learning vector quantization networks. *International Journal of Production Research, 32*(3), 721–729. doi:10.1080/00207549408956963.

Pham, D. T., & Wani, M. A. (1997). Feature-based control chart recognition. *International Journal of Production Research, 35*(7), 1875–1890. doi:10.1080/002075497194967.

Quinlan, J. R. (1983). *Learning efficient classification procedures and their application to chess endgames*. Machine Learning.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Los Altos, CA: Morgan Kaufmann.

Wang, C. H., Kuo, W., & Qi, H. (2007). An integrated approach for process monitoring using wavelet analysis and competitive neural network. *International Journal of Production Research, 45*(1), 227–244. doi:10.1080/00207540500442393.

Wang, C. H., & Kuo, W. (2007). Identification of control chart patterns using wavelet filtering and robust fuzzy clustering. *Journal of Intelligent Manufacturing, 18*, 343–350. doi:10.1007/s10845-007-0028-6.

Wang, C. H., Guo, R. S., Chiang, M. H., & Wong, J. Y. (2008). Decision tree based control chart pattern recognition. *International Journal of Production Research* (in press).

Wong, M. L. D., Jack, L. B., & Nandi, A. K. (2006). Modified self-organizing map for automated novelty detection applied to vibration signal monitoring. *Mechanical Systems and Signal Processing, 20*, 593–610. doi:10.1016/j.ymssp.2005.01.008.

Yang, M. S., & Yang, J. H. (2002). A fuzzy-soft learning vector quantization for control chart pattern recognition. *International Journal of Production Research, 40*(12), 2721–2731. doi:10.1080/00207540210137639.

Yang, J. H., & Yang, M. S. (2005). A control chart pattern recognition scheme using a statistical correlation coefficient method. *Computers & Industrial Engineering, 48*, 205–221. doi:10.1016/j.cie.2005.01.008.

Yoon, S., & MacGregor, J. F. (2004). Principal component analysis of multivariate data for process monitoring and fault diagnosis. *Journal of American Institute Chemical Engineering, 50*(11), 2891–2903.

Yousef, A. A. (2004). Recognition of control chart patterns using multi-resolution wavelets analysis and neural networks. *Computers & Industrial Engineering, 47*, 17–29. doi:10.1016/j.cie.2004.02.007.