

High-Dimensional Process Monitoring and Fault Isolation via Variable Selection

KAIBO WANG

Tsinghua University, Beijing 100084, P. R. China

WEI JIANG

The Hong Kong University of Science & Technology, Kowloon, Hong Kong

Both process monitoring and fault isolation are important and challenging tasks for quality control and improvement in high-dimensional processes. Under a practical assumption that not all variables would shift simultaneously, this paper proposes a variable-selection-based multivariate statistical process control (SPC) procedure for process monitoring and fault diagnosis. A forward-selection algorithm is first utilized to screen out potential out-of-control variables; a multivariate control chart is then set up to monitor suspicious variables. Therefore, detection of faulty conditions and isolation of faulty variables can be achieved in one step. Both simulation studies and a real example have shown the effectiveness of the proposed procedure.

Key Words: Forward Selection; Linear Regression; Multivariate Statistical Process Control; T^2 Chart.

STATISTICAL monitoring of high-dimensional processes and subsequent fault isolation are becoming increasingly important in modern manufacturing environments. For example, typical machining equipments may have multiple key variables to be measured continuously; a sensor network could centralize dozens or even more variables from multiple production lines or stations for decision making (Ding et al. (2006)); A profile that is useful for characterizing product quality in some applications (Woodall et al. (2004)) may consist of hundreds of gauge points, which form a dataset with an extremely high dimension. While monitoring dozens or even hundreds of variables simultaneously is crucial to process control and instrumentation, traditional statistical methods based on small- or medium-size samples may not be applicable due to the “curse-of-dimensionality” problem (Hastie et al. (2001)). Moreover, when a large number of variables are sampled simultaneously and fed to a central processor for decision making, impor-

tant signals could be easily concealed by noises, which make it difficult to detect potential process shifts and locate root causes of faults.

Interestingly, in high-dimensional applications, it is very rare to see all interested variables or quality characteristics change or shift all at the same time. Rather, a typical yet common phenomenon observed in practice is that a subset of variables, which is dominated by a common latent physical mechanism or component, deviate from their normal condition due to abnormal changes of the common mechanism or component (Mastrangelo et al. (1996), Choi et al. (2006), Li et al. (2008)). This is particularly true in distributed sensor systems where groups of sensors are responsible for detecting particular events. For this reason, if information collected from multiple streams of variables is treated equally, the effect of weak, or sometimes even strong, signals could become difficult to detect.

For example, consider Hotelling’s T^2 chart when the dimension p increases. Suppose that there is a shift in x_1 with magnitude 1.0 and no shifts in all other variables, i.e., we monitor x_1 , which is out of control, plus $(p-1)$ in-control variables. Without loss of generality, we assume an identity correlation matrix. Table 1 shows the probability for detecting the

Dr. Wang is an Assistant Professor in the Department of Industrial Engineering. His email address is kbwang@tsinghua.edu.cn.

Jiang is a Visiting Associate Professor in the Department of Industrial Engineering & Logistics Management. His email address is jiangwei08@gmail.com.

TABLE 1. Detection Probability of a Shift in x_1 with Magnitude 1 when p Varies (the false alarm probability is 0.002 for all cases)

p	Detection probability	p	Detection probability
1	0.982	11	0.354
2	0.964	12	0.281
3	0.937	13	0.216
4	0.897	14	0.163
5	0.843	15	0.120
6	0.776	16	0.086
7	0.697	17	0.060
8	0.612	18	0.041
9	0.524	19	0.028
10	0.436	20	0.018

shift in x_1 for achieving the same level of false-alarm probability when the dimension p increases from 1 to 20. The dimensions of the chart are listed in the first and the third columns. The false-alarm probability is set to be 0.002.

It is easy to see that a shift of size 1.0 can be detected with probability 0.982 if it is monitored individually ($p = 1$); however, if it is jointly monitored with another in-control variable ($p = 2$), the detection probability becomes 0.964. When $p = 10$, the detection probability decreases to 0.436, while when $p = 20$, it decreases to 0.018. This illustrates that power of detecting the signal of x_1 is diminished by other variables due to the inclusion of more noises. Runger and Alt (1996) found a similar phenomenon in their work. This example demonstrates the necessity to “reduce dimension” in multivariate statistical process control (SPC) because it will be hard to detect a weak signal in a high-dimensional scenario. In this research, we concentrate on high-dimensional cases where a few key variables can dominate the results. As we tend to believe that the number of such key variables that carry shift information is usually low, we propose using variable-selection techniques to locate potential out-of-control variables first, and then monitor these variables only. It is expected that monitoring only selected variables can be more efficient than monitoring all variables simultaneously.

Essentially there are two critical tasks for any SPC technique designed for high-dimensional processes to accomplish: the ability to trigger an alarm when a process deviates from its normal condition, and the ability to identify specific variables that are responsible for such alarms. The former task focuses on

shift detection, and the later one focuses on alarm diagnosis. Traditional SPC schemes usually focus on only one of the tasks. For example, Hotelling’s T^2 chart, multivariate versions of EWMA (MEWMA) and CUSUM (MCUSUM) charts are designed to detect process shifts but not to provide more insights regarding which variable(s) are responsible for the alarm (Woodall and Ncube (1985), Lowry and Montgomery (1995)). In addition, these charts treat all variables equally. In a process with dozens or even hundreds of variables, efficiency of such control charts could decrease quickly due to the foregoing reasons.

On the other hand, different fault diagnostic techniques have been developed separately. The regression-adjusted chart (Hawkins (1991), Hawkins (1993)) is one charting strategy that can provide diagnostic information. For a cascade process, this chart regresses downstream variables on upstream variables to identify root causes (Hawkins and Maboudou-Tchao (2008)). An adaptive regression-adjusted monitoring scheme was proposed by Liu et al. (2006); a procedure to search for root causes when an alarm is triggered was also presented. However, the authors indicated that the computational cost may increase significantly when the number of variables increases. The MYT-decomposition (Mason et al. (1995), Mason and Young (1999)) extends the idea to a more general situation by enumerating all possible combinations of variables to screen for out-of-control variables. The MYT decomposition can be optimized and utilized efficiently with large number of variables. However, MYT decomposition is designed for searching for root causes but not for detecting process shifts in nature. In this paper, we propose a variable-selection-based multivariate SPC (MSPC) control chart, which is denoted as the VS-MSPC chart hereafter. The VS-MSPC chart conducts an automated variable-selection procedure to screen out potential shift variables and estimate their shift magnitudes first, then monitor only the chosen variables. Because the number of simultaneously shift variables is usually rather small in a high-dimensional process, most in-control variables are removed from the monitoring procedure and only potential shift variables are examined. The treatment is expected to improve fault-detection performance and isolate root causes simultaneously.

The rest of this paper is organized as follows. Section 2 derives a penalized likelihood function for variable selection, based on which the VS-MSPC chart

is proposed in Section 3. Section 4 investigates the performance of the proposed method and compares it with Hotelling’s T^2 chart. Section 5 concludes this paper with a summary of main features of the VS-MSPC chart.

Penalized Likelihood for Out-of-Control Variable Selection

In this section, we first derive a generalized likelihood ratio test (GLRT) statistic, which is useful for analyzing SPC algorithms. By considering a practical assumption that it is unusual for all variables in a high-dimensional process to shift simultaneously, a penalized likelihood is then considered to identify potential out-of-control variables. The solution of the penalized likelihood function is found by using existing variable-selection algorithms. The identified variables will be further utilized in Section 3 for process monitoring and root-cause discovery.

GLRT Statistic for Process Monitoring

Let \mathbf{y}_t be a p -dimensional observation collected at step t . Assume $\mathbf{y}_t \sim N_p(\mu, \Sigma)$. Checking the status of process means, which may be either in control or out of control, is equivalent to examining the following statistical hypothesis:

$$\begin{cases} H_0 : \mu \in \Omega_0 \\ H_1 : \mu \in \Omega_1, \end{cases} \quad (1)$$

where, without loss of generality, $\Omega_0 = \{0\}$ represents the parameter space when the process is in control; $\Omega_1 = \{\mu : \mu = \delta \mathbf{d}, \delta > 0\}$ is the parameter space when the process is out of control. Here, \mathbf{d} is a unit-length direction vector, $\|\mathbf{d}\|_{\Sigma^{-1}} = \sqrt{\mathbf{d}^T \Sigma^{-1} \mathbf{d}} = 1$; δ is a scalar that represents the length of a shift. Therefore, $\delta \mathbf{d}$ can represent mean shifts in any directions with any magnitudes.

To test the above hypothesis, we investigate the GLRT statistic with respect to it:

$$\lambda(\mathbf{y}_t) = \frac{\max_{\mu \in \Omega_0} L(\mathbf{y}_t, \mu)}{\max_{\mu \in \Omega_1} L(\mathbf{y}_t, \mu)},$$

where $L(\mathbf{y}_t, \mu)$ is the likelihood corresponding to \mathbf{y}_t . The null hypothesis is rejected and the alternative hypothesis is favored if $\lambda(\mathbf{y}_t) < c_1$, where $c_1 > 0$ is a constant that corresponds to a specific type-I error of the test.

Because $\Omega_0 = \{0\}$, which contains only a single point, the rejection region of the hypothesis can be

simplified as

$$\lambda(\mathbf{y}_t) = \frac{L(\mathbf{y}_t, 0)}{\max_{\mu \in \Omega_1} L(\mathbf{y}_t, \mu)} \leq c_1,$$

or equivalently,

$$\begin{aligned} \Lambda(\mathbf{y}_t) &= \log \lambda(\mathbf{y}_t) \\ &= \min_{\mu \in \Omega_1} (\log L(\mathbf{y}_t, 0) - \log L(\mathbf{y}_t, \mu)) < \log c_1. \end{aligned} \quad (2)$$

Under the assumption that $\mathbf{y}_t \sim N_p(\mu, \Sigma)$, the likelihood of \mathbf{y}_t is given by

$$L(\mathbf{y}_t, \mu) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-(\mathbf{y}_t - \mu)^T \Sigma^{-1} (\mathbf{y}_t - \mu) / 2}.$$

Therefore, the rejection region in Equation (2) is equivalent to

$$\Lambda(\mathbf{y}_t) = \min_{\mu \in \Omega_1} (-\mathbf{y}_t^T \Sigma^{-1} \mathbf{y}_t + (\mathbf{y}_t - \mu)^T \Sigma^{-1} (\mathbf{y}_t - \mu)) < \log c_1.$$

Let

$$S^2 = \min_{\mu \in \Omega_1} ((\mathbf{y}_t - \mu)^T \Sigma^{-1} (\mathbf{y}_t - \mu)). \quad (3)$$

We reject the null hypothesis in Equation (1) if

$$\Lambda(\mathbf{y}_t) = \mathbf{y}_t^T \Sigma^{-1} \mathbf{y}_t - S^2 > c \quad (4)$$

holds, where $C = -\log c_1$. The minimum of Equation (3) is attained when $\mu^* = \mathbf{y}_t$, which directly leads to $S^2 = 0$. The resulting inequality characterizes the rejection region of Hotelling’s T^2 chart, which in turn implies that Hotelling’s T^2 chart can be viewed as a GLRT procedure (Jiang and Tsui (2008)).

The Penalized Likelihood for Variable Selection

The solution of Equation (3), denoted by μ^* , can be viewed as an estimator of μ . However, the current mathematical solution, $\mu^* = \mathbf{y}_t$, has obviously ignored important engineering background. Without loss of generality, we assume $\mu = 0$ when the process is in control. As \mathbf{y}_t contains both useful and noise signals, setting $\mu^* = \mathbf{y}_t$ is obviously inaccurate because all elements of \mathbf{y}_t are contaminated by noises. Alternatively, it is more reasonable to assume that most elements of \mathbf{y}_t that have values close to zero are in control and their true mean is zero.

In order to force small coefficients to zero and increase model interpretability, we consider adding a penalty term to Equation (3). This leads to

$$S^2 = \min_{\mu \in \Omega_1} \left((\mathbf{y}_t - \mu)^T \Sigma^{-1} (\mathbf{y}_t - \mu) + \sum_{j=1}^p p_{\lambda_j}(|\mu_j|) \right), \quad (5)$$

where p is the dimension of \mathbf{y}_t and $|\mu_t|$ is the absolute value of the i th element of μ , i.e., the true mean of the i th variables. The newly added term, $p_{\lambda_j}(\cdot)$, is a penalty function that controls model complexity, i.e., the number of variables with nonzero mean value $|\mu_i| > 0$. By properly choosing $p_{\lambda_j}(\cdot)$, the solution of Equation (5) is expected to automatically set small means to zero and leave only potential out-of-control variables.

Different penalty functions have been proposed in the literature of variable selection. Fan and Li (2006) provided a review of these methods and suggested that, for an L_q -penalty function with $q \leq 1$, the penalized least square automatically performs variable selection by removing predictors with very small estimated coefficients. Two special forms of the penalized functions in Equation (5) are the L_1 penalty and L_0 penalty. With the L_1 penalty, $p_{\lambda_j}(|\mu_j|) = \lambda|\mu_j|$, minimizing the penalized likelihood function is equivalent to minimizing the original likelihood function in Equation (3) subject to the following constraint:

$$\sum_{i=1}^p |\mu_i| \leq c_2,$$

or simply,

$$S^2(\lambda) = \min_{\mu \in \Omega_1} \left((\mathbf{y}_t - \mu)^T \Sigma^{-1} (\mathbf{y}_t - \mu) + \lambda \sum_{j=1}^p |\mu_j| \right), \tag{6}$$

where λ is a penalty parameter that depends on c_2 . Equation (6) corresponds to the least absolute shrinkage and selection operator, or the LASSO, which was proposed by Tibshirani (1996).

With the L_0 penalty, $p_{\lambda_j}(|\mu_j|) = \lambda I(|\mu_j| \neq 0)$, the penalized likelihood function becomes

$$S^2(\lambda) = \min_{\mu \in \Omega_1} ((\mathbf{y}_t - \mu)^T \Sigma^{-1} (\mathbf{y}_t - \mu) + \lambda M), \tag{7}$$

where $M = \sum_j I(|\mu_j| \neq 0)$ is the number of nonzero coefficients in the final model. Instead of limiting the sum of absolute values of the coefficients, the L_0 penalty constrains the number of nonzero coefficients in μ .

Both Equations (6) and (7) are capable of deleting variables that are close to zero in least square regression. In this research, we choose the L_0 penalty because it directly limits the number of nonzero coefficients in the model. The variables with coefficients being set to zero are treated as in control while other variables are potential out-of-control variables and

their shift magnitudes are estimated by the corresponding elements of μ^* . A solution of Equation (7) is derived in the following section.

Minimizing the Penalized Least Square

Because Σ is the variance-covariance matrix of \mathbf{y}_t in Equation (7), both Σ and Σ^{-1} are positive definite. Using the Cholesky decomposition of $\Sigma^{-1} = \mathbf{x}^T \mathbf{x}$, where \mathbf{x} is an upper triangular matrix, the penalized likelihood in Equation (7) can be written as as

$$\begin{aligned} S^2(\lambda) &= \min_{\mu \in \Omega_1} ((\mathbf{y}_t - \mu)^T \mathbf{x}^T \mathbf{x} (\mathbf{y}_t - \mu) + \lambda M) \\ &= \min_{\mu \in \Omega_1} ((\mathbf{x}\mathbf{y}_t - \mathbf{x}\mu)^T (\mathbf{x}\mathbf{y}_t - \mathbf{x}\mu) + \lambda M). \end{aligned}$$

Let $\mathbf{z}_t = \mathbf{x}\mathbf{y}_t$, then \mathbf{z}_t follows a multivariate normal distribution with mean $\mathbf{x}\mu$ and variance $\mathbf{x}\Sigma\mathbf{x}^T$. Because $\mathbf{x}\Sigma\mathbf{x}^T = \mathbf{I}$, $\mathbf{z}_t \sim N_p(\mathbf{x}\mu, \mathbf{I})$, i.e., \mathbf{z}_t is essentially a standardized vector of \mathbf{y}_t . Treating \mathbf{x} as the covariates or design points and \mathbf{z}_t as the response, the above optimization problem can be transformed to a penalized least square problem,

$$S^2(\lambda) = \min_{\mu \in \Omega_1} ((\mathbf{z}_t - \mathbf{x}\mu)^T (\mathbf{z}_t - \mathbf{x}\mu) + \lambda M), \tag{8}$$

in which μ is the coefficient vector that needs to be estimated.

Because the main purpose of solving Equation (8) is to identify potential out-of-control variables and estimate their shift magnitudes using μ^* , we can rewrite Equation (8) as

$$\begin{cases} \min_{\mu \in \Omega_1} ((\mathbf{z}_t - \mathbf{x}\mu)^T (\mathbf{z}_t - \mathbf{x}\mu)) \\ \text{s.t. } \sum_j I(|\mu_j| \neq 0) \leq s \end{cases} \tag{9}$$

That is, minimizing the least square with the number of nonzero predictors less than or equal to s . One straightforward way to find its solution is the best subset selection algorithm, which searches through all subsets of size k , $k \in \{0, 1, \dots, s\}$, to find the smallest sum of square of residuals. However, this searching procedure requires extensive computational resources. As indicated by Hastie et al. (2001) on page 55, even an efficient algorithm (the leaps and bounds procedure due to Furnival and Wilson (1974)) is feasible for dimensions as large as 30 or 40 only. Therefore, alternative solutions with approximately optimal performance have to be studied.

The forward-variable-selection procedure is a popular heuristic algorithm for fitting regression models that gives nearly the optimal performance. It starts from an empty set with no predictors. At each step,

the predictor that results in the largest decrease in sum of square errors or the largest increase in R^2 is added to the model. The usual selection criterion is the F -to-enter rule (Wilkinson and Dallal (1981)). Let

$$F = \frac{(R_{k+1}^2 - R_k^2)(n - k - 1)}{1 - R_{k+1}^2},$$

where R_k^2 is the R^2 statistic of the current model with k predictors and R_{k+1}^2 is the R^2 statistic when one more predictor is added to the model. The F -values of all variables are tested and the one with the highest value enters the model.

Other variable-selection algorithms, such as backward selection and stepwise, may also be considered. While in a high-dimensional process, the number of simultaneously shifted variables is usually much smaller than the process dimension p , it would be faster to add variables to a simple model than to remove variables from a full model. As the forward-variable-selection procedure is less demanding computationally when the number of candidate independent variables is large, we use this procedure to find the optimal μ^* to minimize Equation (9); specifically, the algorithm proposed by Efroymson (1960) is adopted for its high efficiency.

It should be noted that, different from the usual fully automated subset-selection procedures that stop searching based on statistical significance, such as p -values or F -values, the only stopping condition in solving Equation (9) is the constraint $\sum_j I(|\mu_j| \neq 0) \leq s$. Because increasing the number of nonzero coefficients can always decrease the sum of squares term, the boundary of the inequality in Equation (9) is deterministically achieved. Therefore, $M = \sum_j I(|\mu_j| \neq 0) = s$ always holds. That is, solving Equation (9) can help us find a fixed number of potential out-of-control variables. Such variables with nonzero coefficients will be charted against a control limit to check whether the process is in control or not. If an alarm is triggered during testing, it is easy to conclude that the variables with nonzero coefficients are responsible for such an alarm, which makes the diagnosis of alarms quite straightforward.

The Variable-Selection-Based Multivariate Control Chart

The solution to the penalized likelihood function locates potential out-of-control variables. In this section, we propose monitoring these variables against process shifts by a multivariate control chart. The

settings of unknown parameters in the proposed control chart are also discussed.

Multivariate Statistical Process Monitoring

Once the solution μ^* is obtained via the variable-selection procedure, potential out-of-control variables are identified and their shift magnitudes are obtained at the same time. Substituting μ^* for μ in Equation (8) yields

$$S^2 = (\mathbf{z}_t - \mathbf{x}\mu^*)^T(\mathbf{z}_t - \mathbf{x}\mu^*) + \lambda M.$$

Because $\Sigma^{-1} = \mathbf{x}^T\mathbf{x}$ and $\mathbf{z}_t = \mathbf{x}\mathbf{y}_t$, it follows that

$$\begin{aligned} S^2 &= \mathbf{z}_t^T\mathbf{z}_t - \mathbf{z}_t^T\mathbf{x} - \mu^{*T}\mathbf{x}^T\mathbf{z}_t + \mu^{*T}\mathbf{x}^T\mathbf{x}\mu^* + \lambda M \\ &= \mathbf{y}_t^T\mathbf{x}^T\mathbf{x}\mathbf{y}_t - \mathbf{y}_t^T\mathbf{x}^T\mathbf{x}\mu^* - \mu^{*T}\mathbf{x}^T\mathbf{x}\mathbf{y}_t \\ &\quad + \mu^{*T}\mathbf{x}^T\mathbf{x}\mu^* + \lambda M \\ &= \mathbf{y}_t^T\Sigma^{-1}\mathbf{y}_t - \mathbf{y}_t^T\Sigma^{-1}\mu^* - \mu^{*T}\Sigma^{-1}\mathbf{y}_t \\ &\quad + \mu^{*T}\Sigma^{-1}\mu^* + \lambda M \\ &= \mathbf{y}_t^T\Sigma^{-1}\mathbf{y}_t - 2\mathbf{y}_t^T\Sigma^{-1}\mu^* - \mu^{*T}\Sigma^{-1}\mu^* + \lambda M. \end{aligned}$$

Now we substitute the above result for S^2 in Equation (4) and obtain

$$\Lambda(\mathbf{y}_t) = 2\mathbf{y}_t^T\Sigma^{-1}\mu^* - \mu^{*T}\Sigma^{-1}\mu^* - \lambda M > c.$$

As previously discussed, we choose a fixed number of nonzero coefficients via the variable-selection procedure. Therefore, M will reach the boundary condition s and become a constant. An equivalent charting statistic is obtained by removing the constant term and modifying the threshold value,

$$\Lambda(\mathbf{y}_t) = 2\mathbf{y}_t^T\Sigma^{-1}\mu^* - \mu^{*T}\Sigma^{-1}\mu^* > c'. \quad (10)$$

Equation (10) defines a new control chart for monitoring process status. Because a variable-selection procedure is first conducted to select potential out-of-control variables and estimate their shift magnitudes, we name the new scheme the VS-MSPC chart. Steps to implement the VS-MSPC chart are summarized as follows:

- (a) Variable selection. At each step, once a new observation is collected, variable selection is conducted to identify potential out-of-control variables. The resulting nonzero coefficients estimate shift magnitudes of corresponding variables. Those variables with zero-value coefficients are treated as in control.
- (b) Process monitoring. Variables with nonzero coefficients are charted using Equation (10). If the charting statistic exceeds a preset control limit, an out-of-control alarm will be triggered.

- (c) Signal diagnosis. Following an out-of-control alarm, all variables with nonzero coefficients identified in Step (a), or equally, all the variables monitored in Step (b), are the variables that should be responsible for this alarm.

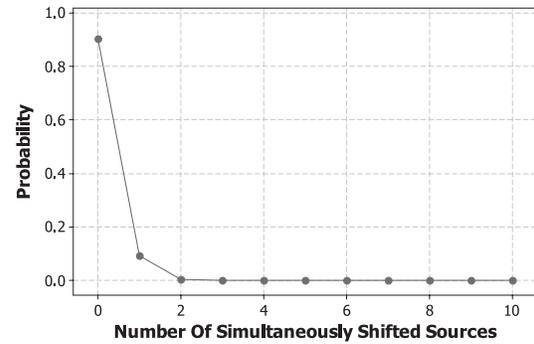
The above procedure shows that the advantages of the VS-MSPC procedure lie not only on the capability to identify out-of-control variables that are responsible for an alarm but also on the capability to estimate the corresponding shift magnitudes. The tasks of monitoring and diagnosis in SPC are naturally integrated and conveniently solved simultaneously.

Parameter Determination for the VS-MSPC Chart

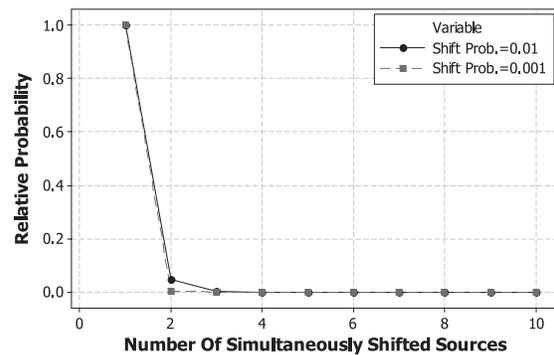
The maximum number of nonzero coefficients in Equation (9), s , or equivalent M in Equation (8), is an arbitrary parameter that should be determined based on knowledge and experience of specific processes. In practice, shifts in process variables could be caused by physical failures, component malfunction, etc. The probability that multiple variables shift simultaneously is generally low.

Suppose there are 10 independent sources of failures in a process and the probability that any failure occurs at each step is 1%. Figure 1(a) shows the probability that multiple variables shift simultaneously. It is seen from Figure 1 that the probability that no variables shift is 90.44%, one source shifts with a probability of 9.14%, while the probability that two sources fail simultaneously drops sharply to 0.4%. The probability for more than two sources shifting is very close to zero. Figure 1(b) shows the relative probability that multiple sources shift simultaneously. The scale of the vertical axis is divided by the probability that only one source shifts. It is found that the likelihood that one source shifts is dominating. This dominance becomes even stronger if the probability that any individual source shifts becomes smaller. Therefore, it is practical and meaningful to assume that, at any single step, only one source would shift in a process.

In industrial processes, multiple sensors are sometimes placed to monitor a common physical mechanism. As a result, changes in one source will be reflected in multiple channels. Under such circumstances, the number of simultaneously shifted process variables can be determined by domain knowledge and experience about specific processes (Li et al. (2008)). When such domain knowledge is not avail-



(a)



(b)

FIGURE 1. The Probability That Multiple Variables Shift Simultaneously.

able, in general, one may consider using “penalized principal component analysis (PCA)” for determining s . Principal component analysis is a way to find hidden structures of systems. Usually, the resulting principal components are linear combinations of original variables. Penalized PCA generates principal components that are linear combinations of partial variables but not all variables (so that the loadings of some variables are set to zero). Based on the penalized PCA, we can identify the variables that are influenced by a common source of variation. In this way, the number of simultaneously shifted variables can be determined. Jolliffe et al. (2003) proposed a modified PCA, which is equivalent to the concept of “penalized PCA” we illustrated above. The use of the penalized PCA and its integration with the VS-MSPC chart is a topic that should be further studied based on real-world examples.

Statistical Performance Analysis

To investigate the performance of the proposed VS-MSPC chart, we compare it with Hotelling’s T^2

TABLE 2. ARL Comparison, $p = 10$

δ	T^2	VS-MSPC					
		$M = 1$	$M = 2$	$M = 3$	$M = 4$	$M = 5$	$M = 6$
0.00	200.00	199.68	200.99	200.96	200.36	199.78	200.05
0.50	132.28	145.70	138.67	134.39	133.10	132.73	133.09
1.00	50.78	59.23	52.14	50.45	49.74	49.75	50.35
1.50	17.14	19.82	16.48	15.92	15.92	16.18	16.49
2.00	6.42	7.37	5.83	5.74	5.88	6.01	6.16
2.50	2.96	3.32	2.65	2.64	2.71	2.78	2.85
3.00	1.72	1.87	1.56	1.56	1.60	1.63	1.66
3.50	1.25	1.32	1.18	1.18	1.20	1.21	1.23
4.00	1.07	1.10	1.04	1.04	1.05	1.06	1.06
4.50	1.02	1.02	1.01	1.01	1.01	1.01	1.01
5.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

chart, which is the most widely used chart for multivariate process monitoring. Processes with dimensions $p = 10, 20, 50,$ and 100 are considered. Without loss of generality, we assume $\mathbf{y}_t \sim N_p(\mathbf{0}, \mathbf{I})$ when the process is in control and $\mathbf{y}_t \sim N_p([\delta, \delta, 0, \dots, 0]^T, \mathbf{I})$ when the process is out of control. That is, two out of the p variables are assumed to have shifts with magnitudes $\delta \in (0, 5]$. Because the exact number of shifted variables is usually unknown in practice, fixing M in the VS-MSCP chart may under- or overestimate this parameter. Therefore, we choose $M = 1-6$ for $p = 10$ and $M = 1, 2, 3$ for $p = 20, 50,$ and 100 to simulate different scenarios. The in-control average run length (ARL) of all control charts are set to

200. Each ARL is obtained using at least 10,000 replicates. ARL has been popularly used for control-chart analysis. The larger the in-control ARL, the more robust the control chart; the smaller the out-of-control ARL, the more efficient the control chart.

Tables 2-5 show the ARL performance of the VS-MSPC chart and Hotelling's T^2 chart. It is seen that, even though all charts have an equal in-control ARL, most variable-selection-based MSPC charts (except the one with $M = 1$ when $p = 10$) outperform the T^2 chart for $\delta \geq 1.0$. Simulation results also suggest that the proposed method is quite robust even when M is misspecified. In monitoring the process with

TABLE 3. ARL Comparison, $p = 20$

δ	T^2	VS-MSPC		
		$M = 1$	$M = 2$	$M = 3$
0.00	200.00	200.14	200.43	200.13
0.50	151.29	164.05	158.70	156.43
1.00	73.60	82.40	74.88	72.70
1.50	29.04	28.87	24.90	24.57
2.00	11.29	10.26	8.66	8.48
2.50	4.90	4.38	3.57	3.58
3.00	2.54	2.28	1.90	1.90
3.50	1.60	1.48	1.29	1.30
4.00	1.22	1.16	1.08	1.08
4.50	1.07	1.04	1.02	1.02
5.00	1.02	1.01	1.00	1.00

TABLE 4. ARL Comparison, $p = 50$

δ	T^2	VS-MSPC		
		$M = 1$	$M = 2$	$M = 3$
0.00	200.00	200.22	199.55	199.55
0.50	169.33	182.02	177.59	176.24
1.00	106.37	114.75	108.62	107.74
1.50	54.07	46.87	42.32	42.02
2.00	24.69	16.09	14.00	14.12
2.50	11.15	6.21	5.23	5.33
3.00	5.38	2.95	2.47	2.50
3.50	2.92	1.73	1.51	1.53
4.00	1.84	1.26	1.16	1.16
4.50	1.35	1.08	1.04	1.04
5.00	1.13	1.02	1.01	1.01

TABLE 5. ARL Comparison, $p = 100$

δ	T^2	VS-MSPC		
		$M = 1$	$M = 2$	$M = 3$
0.00	200.00	199.45	199.22	199.12
0.50	178.54	187.44	185.62	184.31
1.00	128.79	137.94	132.71	131.37
1.50	77.99	64.38	60.41	61.03
2.00	41.84	22.46	20.55	21.07
2.50	21.10	8.27	7.29	7.49
3.00	10.60	3.65	3.16	3.22
3.50	5.57	2.00	1.76	1.79
4.00	3.19	1.37	1.25	1.26
4.50	2.04	1.12	1.07	1.07
5.00	1.48	1.03	1.01	1.01

$p = 20, 50,$ or $100,$ the VS-MSPC chart is superior to Hotelling's T^2 chart in detecting medium and large shifts for most cases. For some cases, $M = 2$ may not necessarily give the lowest out-of-control ARL due to the existence of noise. However, its performance is always close to the optimal one.

It is interesting to note that the performance of the VS-MSPC chart varies with process dimension $p.$ We now define an index, relative efficiency (RE), as the ratios between the ARLs of Hotelling's T^2 chart and the VS-MSPC chart. Obviously, the higher this index is, the better is the VS-MSPC chart. Based on the ARLs shown in Tables 2–5, we choose the VS-MSPC chart with $M = 2$ to compute their relative efficiency. As is clearly seen from Figure 2, the relative efficiency of the VS-MSPC chart becomes higher when p becomes larger, which implies that the benefits of using the VS-MSPC chart for high-dimensional processes are more prominent. This property is well explained by the fact that larger p values correspond to more observations in the penalized likelihood estimation in Equation (5) or the restricted least square estimator in Equation (9). It implies that applying variable selection is an effective way to filter out potentially out-of-control variables so that the control chart can focus on suspects and make more accurate decisions about process status.

It should be noted that the major advantage of the VS-MSPC chart is not only its capability in detecting process shifts, but also in searching for potential out-of-control variables. That is, the chart is capable of suggesting responsible variables when an alarm is triggered. To investigate the diagnostic capability of

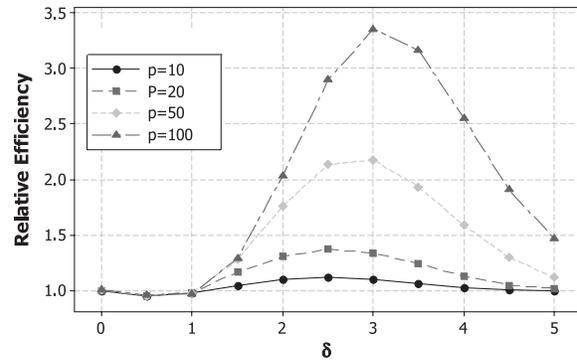


FIGURE 2. Relative Efficiency Between the VS-MSPC Chart and Hotelling's T^2 Chart.

the proposed scheme, a process with $p = 10$ is simulated and shifts of size δ are added to the first two variables (y_1 and y_2). A VS-MSPC chart with $M = 2$ is set up and 200 out-of-control alarms are collected from it. Corresponding variables that are responsible for such alarms are also recorded and shown in Table 6. Each column with titles from y_1 to y_2 shows the number of times that the corresponding variables are chosen to be responsible for out-of-control alarms. The last column, titled *Correctness*, is obtained by dividing the added number of times that y_1 and y_2 are chosen by the sum of each row. Correctness can be used as an indicator to check how often correct variables are identified in the study. Table 6 clearly shows that, although diagnostic capability of the VS-MSPC chart is low for weak process shifts, when the shift sizes increase, the probability that out-of-control variables are correctly identified is increasing significantly. Even for a shift size of 1, the correctness probability is higher than 50%. For shifts size of three or larger, more than 90% of the cases, out-of-control alarms are correctly tracked down.

The above simulations are conducted assuming an identity covariance matrix for the process. While when the original variables are correlated or one works on transformed independent components first and needs to transform back to original variables, the above findings may not hold accurately. In the following section, a real example with a general covariance matrix is presented and used to demonstrate the effectiveness of the VS-SPC chart for high-dimensional process monitoring.

A Real Example

Vertical density profiles (VDP) of engineered wooden boards is a measure of board quality (Young

TABLE 6. Diagnostic Capability Study of the VS-MSPC Chart. $p = 10$, $M = 2$

δ	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	Correctness (%)
0.5	58	58	38	34	34	31	36	37	38	36	29.0
1.0	110	109	24	18	18	25	25	23	23	25	54.8
1.5	132	141	12	19	9	17	14	24	15	17	68.3
2.0	152	157	10	17	6	11	8	15	10	14	77.3
3.0	183	187	5	1	4	6	4	6	1	3	92.5
4.0	193	192	0	1	1	5	5	1	0	2	96.3
5.0	199	199	0	0	0	1	1	0	0	0	99.5

et al. (1999)). A VDP is formed by measuring density of a board at different depths. A dataset of 24 VDPs is presented by Walker and Wright (2002) and reviewed by Woodall et al. (2004). Several typical profiles and the sample average profile are shown in Figure 3. Each profile consists of 314 points, correspond to 314 random variables. Because there are only 24 samples available, which are not sufficient for estimating a process variance-covariance matrix, we resample each profile by taking one point out of

every 16 points to reduce the dimension of the curves to 20. Then each curve has 20 variables left. In addition, all profile variables are standardized to have a mean of zero and variances of one.

In this section, we plot these profiles on a VS-MSPC chart and compare it with Hotelling's T^2 chart. During variable selection, five variables are retained to have nonzero coefficients. The sample variance-covariance matrix is calculated and utilized

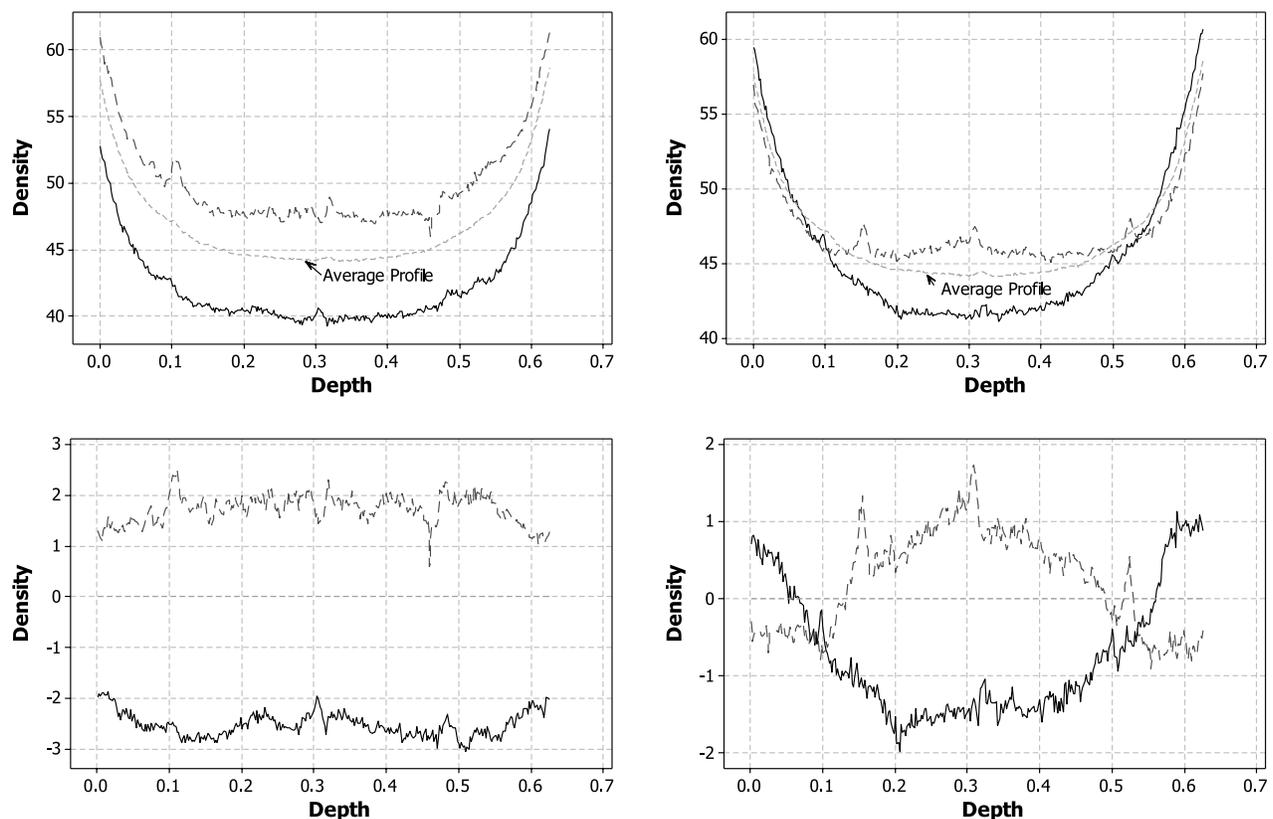


FIGURE 3. Typical Vertical Density Profiles and the Average Profile. Top row, raw profiles; bottom row, standardized profiles.

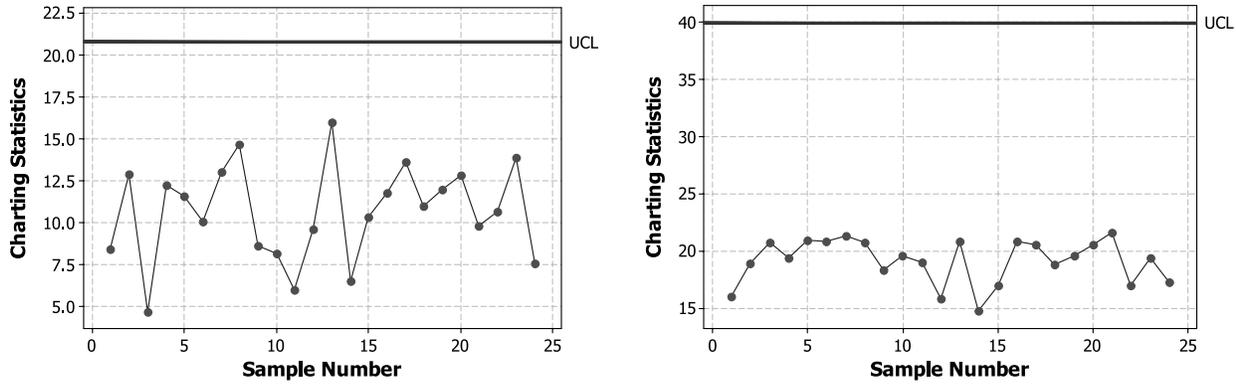


FIGURE 4. Statistical Monitoring of VDPs. Left, VS-MSPC chart; right, Hotelling's T^2 chart.

for control-chart design. The resulting VS-MSPC chart is shown in Figure 4. The UCL corresponding to an in-control ARL of 200 is shown. The control limit of Hotelling's T^2 chart is set to $\text{Inv-}\chi_{0.995,20}^2 = 39.9968$, with a corresponding in-control ARL of 200.

As seen in Figure 4, both charts show no out-of-control points. However, the relative shift magnitudes of samples are seen to be different. For example, VS-MSPC reports that sample 13 is farther away from the baseline than sample 21, while Hotelling's T^2 chart reports the opposite situation. If we check the standardized samples 13 and 21, as shown in Figure 5, we can find that most points of sample 13 stay far above zero. Sample 21, on the other hand, is far deviated from zero only near its two ends. Because the VS-MSPC chart has taken the correlation structure of the variables into consideration and chosen most representative variables to characterize the profiles, it can differentiate these profiles in a more easily understandable way.

It can be seen from Figure 4 that the points on the VS-MSPC charts are closer to the control limit than those on the T^2 chart. To assess the closeness to the decision boundary, p -values of the observed T^2 points can be evaluated. The largest point on the VS-MSPC chart has a p -value of 0.139, while the largest point on the T^2 chart has a p -value of 0.360, which is much higher. This shows that the VS-MSPC chart is expected to be more sensitive than the T^2 chart if any abnormal conditions happen.

Williams et al. (2007) analyzed the same dataset and concluded that with a type-I error of $\alpha = 0.05$, sample 15 is the most profound outlier, but sample 13 is not suspected by any charts the authors tried. However, in their analysis, the correlation structure among the variables was ignored. In fact, all variables are strongly correlated in this example. Most correlation coefficients are larger than 0.7 or even 0.9. Figure 5 shows the plots of samples 13 and 15. It is seen that, even though sample 15 deviates far from zero near its high end, it stays closer to zero than sample

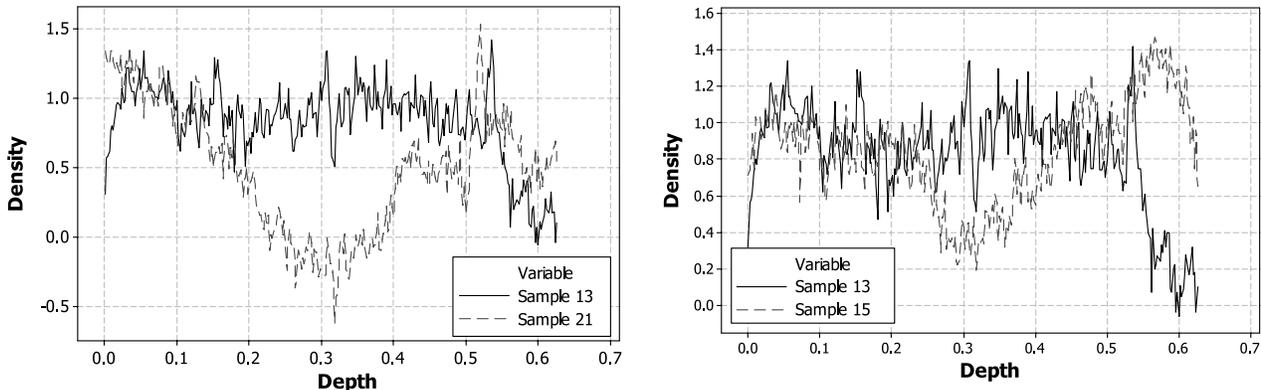


FIGURE 5. Plots of Samples. Left, samples 13 and 21; right, samples 13 and 15.

13 at most of the rest of the locations. Again, the VS-MSPC gives results that are consistent with our graphical observations.

Conclusions

The statistical monitoring of high-dimensional applications, such as sensor networks and complicated manufacturing systems, usually exhibit poor performance due to the fact that shifts in partial variables are averaged out by the huge amount of in-control variables. When out-of-control alarms are received, it is also difficult to diagnose such alarms and locate responsible variables. In this research, a variable-selection-based MSPC control chart, namely the VS-MSPC chart, is proposed to tackle the aforementioned challenges. An automated forward-selection procedure is utilized to screen out potential out-of-control variables. Shift magnitudes of these variables are also estimated. A multivariate SPC chart is then set up to monitor only the suspicious variables.

Simulation results show that the proposed scheme is superior to the Hotelling's T^2 chart for high-dimensional processes in detecting moderate and large shifts. The proposed scheme is also robust to parameter misspecification. Even if the exact number of out-of-control variables is not accurately set in the control chart, the resulting performance is still promising. Most important, the VS-MSPC chart integrates process monitoring and diagnosis in a coherent step and provides a complete solution for SPC of high-dimensional systems.

Our limited simulation experiments have shown that the VS-MSPC chart is not so sensitive in detecting small shifts compared with Hotelling's T^2 chart. This is mainly due to the random noises that conceal the weak signals. Similar to the ideas behind the CUSUM- and EWMA-type of control charts, cumulating historical observations is expected to help detect small shifts in the VS-MSPC chart. This will be further pursued in another paper. Moreover, when latent variables are available in a high-dimensional system, Runger et al. (2007) proposed a systematic way for process monitoring and fault isolation by using weighted least squares methods to regress observations on latent faults/sources. It is expected that their process-modeling strategies are helpful to the VS-MSPC chart in achieving better process understanding and determining optimal charting parameters.

In this research, we treat all mean vectors and co-

variance matrices as unknown. The estimate of these unknown parameters and determination of sample size for parameter estimation are also important topics that deserve future research efforts. Finally, the forward-variable-selection procedure is considered in this research to find potential out-of-control variables. Other variable-selection algorithms may be considered in future research.

Acknowledgments

The authors would like to thank the editor and two anonymous referees for their valuable suggestions, which have helped improve this manuscript greatly. Dr. Kaibo Wang's work was partially supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (SRFDP) under grant 200800031056 and the Beijing Natural Science Foundation under grant 9092005. Dr. Wei Jiang's work was supported by HKUST-DAG08/09.EG10.

References

- CHOI, S. W.; MARTIN, E. B.; MORRIS, A. J.; and LEE, I. B. (2006). "Adaptive Multivariate Statistical Process Control for Monitoring Time-Varying Processes". *Industrial & Engineering Chemistry Research* 45, pp. 3108–3118.
- DING, Y.; ELSAYED, E. A.; KUMARA, S.; LU, J. C.; NIU, F.; and SHI, J. (2006). "Distributed Sensing for Quality and Productivity Improvements". *IEEE Transactions on Automation Science and Engineering* [see also *IEEE Transactions on Robotics and Automation*] 3, pp. 344–359.
- EFROYMSON, M. A. (1960). "Multiple Regression Analysis". In *Mathematical Methods for Digital Computers*, A. Ralston and H. S. Wilf, eds., Ch. 17. New York, NY: Wiley.
- FAN, J. and LI, R. (2006). "Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery". *Proceedings of the International Congress of Mathematicians*, pp. 595–622.
- FURNIVAL, G. and WILSON, R. (1974). "Regression by Leaps and Bounds". *Technometrics* 16, pp. 499–511.
- HASTIE, T.; TIBSHIRANI, R.; and FRIEDMAN, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.
- HAWKINS, D. M. (1991). "Multivariate Quality-Control Based on Regression-Adjusted Variables". *Technometrics* 33, pp. 61–75.
- HAWKINS, D. M. (1993). "Regression Adjustment for Variables in Multivariate Quality Control". *Journal of Quality Technology* 25, pp. 170–182.
- HAWKINS, D. M. and MABOUDOU-TCHAO, E. M. (2008). "Multivariate Exponentially Weighted Moving Covariance Matrix". *Technometrics* 50, pp. 155–166.
- JIANG, W. and TSUI, K.-L. (2008). "A Theoretical Framework and Efficiency Study of Multivariate Control Charts". *IIE Transactions on Quality and Reliability* 40, pp. 650–663.
- JOLLIFFE, I. T.; TRENDAFILOV, N. T.; and UDDIN, M. (2003). "A Modified Principal Component Technique Based on the

- Lasso". *Journal of Computational & Graphical Statistics* 12, pp. 531–547.
- LI, J.; JIN, J. H.; and SHI, J. J. (2008). "Causation-Based T-2 Decomposition for Multivariate Process Monitoring and Diagnosis". *Journal of Quality Technology* 40, pp. 46–58.
- LIU, H. C.; JIANG, W.; TANGIRALA, A.; and SHAH, S. (2006). "An Adaptive Regression Adjusted Monitoring and Fault Isolation Scheme". *Journal of Chemometrics* 20, pp. 280–293.
- LOWRY, C. A. and MONTGOMERY, D. C. (1995). "A Review of Multivariate Control Charts". *IIE Transactions* 27, pp. 800–810.
- MASON, R. L.; TRACY, N. D.; and YOUNG, J. C. (1995). "Decomposition of T-2 for Multivariate Control Chart Interpretation". *Journal of Quality Technology* 27, pp. 99–108.
- MASON, R. L. and YOUNG, J. C. (1999). "Improving the Sensitivity of the T-2 Statistic in Multivariate Process Control". *Journal of Quality Technology* 31, pp. 155–165.
- MASTRANGELO, C. M.; RUNGER, G. C.; and MONTGOMERY, D. C. (1996). "Statistical Process Monitoring with Principal Components". *Quality and Reliability Engineering International* 12, pp. 203–210.
- RUNGER, G. C. and ALT, F. B. (1996). "Choosing Principal Components for Multivariate Statistical Process Control". *Communications in Statistics—Theory and Methods* 25, pp. 909–922.
- RUNGER, G. C.; BARTON, R. R.; DEL CASTILLO, E.; and WOODALL, W. H. (2007). "Optimal Monitoring of Multivariate Data for Fault Patterns". *Journal of Quality Technology* 39, pp. 159–172.
- TIBSHIRANI, R. (1996). "Regression Shrinkage and Selection Via the Lasso". *Journal of the Royal Statistical Society Series B—Methodological* 58, pp. 267–288.
- WALKER, E. and WRIGHT, S. P. (2002). "Comparing Curves Using Additive Models". *Journal of Quality Technology* 34, pp. 118–129.
- WILKINSON, L. and DALLAL, G. E. (1981). "Tests of Significance in Forward Selection Regression with an F-to-Enter Stopping Rule". *Technometrics* 23, pp. 377–380.
- WILLIAMS, J. D.; WOODALL, W. H.; and BIRCH, J. B. (2007). "Statistical Monitoring of Nonlinear Product and Process Quality Profiles". *Quality and Reliability Engineering International* 23, pp. 925–941.
- WOODALL, W. H. and NCUBE, M. M. (1985). "Multivariate CUSUM Quality-Control Procedures". *Technometrics* 27, pp. 285–292.
- WOODALL, W. H.; SPITZNER, D. J.; MONTGOMERY, D. C.; and GUPTA, S. (2004). "Using Control Charts to Monitor Process and Product Quality Profiles". *Journal of Quality Technology* 36, pp. 309–320.
- YOUNG, T. M.; WINISTORFER, P. M.; and WANG, S. Q. (1999). "Multivariate Control Charts of MDF and OSB Vertical Density Profile Attributes". *Forest Products Journal* 49, pp. 79–86.

