Causation-Based T^2 Decomposition for Multivariate Process Monitoring and Diagnosis

Jing Li Jionghua Jin Jianjun Shi[1]
Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor MI 48109-2117

Abstract

Multivariate process monitoring and diagnosis is an important problem. Multivariate SPC using Hotelling T^2 statistic is widely adopted for change detection. However, T^2 control chart alone is not capable of identifying the root causes of the change. Thus, decomposition of T^2 is proposed (called MTY approach) which provides a way to identify the variables significantly contributing to an out-of-control T^2 signal. However, the MTY approach is computationally expensive and has a limited capability in root cause diagnosis for a large dimension of variables. This paper proposes a causation-based T^2 decomposition method by effectively integrating causal models with the MTY approach. Theoretical analysis and simulation studies demonstrate that the proposed method substantially reduces the computational complexity and enhances the diagnosability, compared with the MTY approach.

Keywords: Bayesian network, causal model, SPC, T^2 decomposition

Biography

Ms. Li is a research student in the Department of Industrial and Operations Engineering at the University of Michigan. Her email address is jinglz@umich.edu.

---

[1] Corresponding author: Dr. Jianjun Shi, phone 734-763-5321, fax: 734-764-3451, email: shihang@umich.edu

Dr. Jin is an associate professor in the Department of Industrial and Operations Engineering at the University of Michigan. She is a member of ASQ. Her email address is jhjin@umich.edu.

Dr. Shi is a professor in the Department of Industrial and Operations Engineering at the University of Michigan. He is a senior member of ASQ. His email address is shihang@umich.edu.

Introduction

With the wide adoption of sensors and sensor networks, multivariate process monitoring and diagnosis becomes more and more important. Multivariate SPC, typically the one based on Hotelling T^2 statistic (Alt (1985), and Tracy, Young and Mason (1992)), provides an effective way in mean shift detection by generating out-of-control signals. However, it does not provide any information on which variable(s) cause the change.

To improve the diagnosability of T^2 statistic, Mason, Tracy and Young (1995) proposed to decompose an out-of-control T^2 signal into independent terms and chart those terms individually to identify the variables that significantly contribute to the out-of-control signal. This approach is called the "MTY approach" in the paper. While the MTY approach is theoretically sound and appealing, it has some inherent deficiencies especially when the dimension of variables is high. For example, it requires $p!$ different decompositions of the T^2 statistic for a system of p variables, which results in $p \times 2^{p-1}$ distinct terms to be examined for identifying the out-of-control conditions. Although efforts were made to reduce the number of examined terms (Mason, Tracy and Young (1997)), this number still far exceeds p especially

when there are multiple "faults" in the system. In this paper, a fault is defined as the mean shift of a variable.

In addition to the computational issue, the diagnosability of the MTY approach is also a concern. Although the decomposition of T^2 provides a way for identifying the variables significantly contributing to the out-of-control T^2 signal, those variables may not be the root-cause variables of the fault in the system. For example, if a variable X has a mean shift, this shift will propagate to X's downstream variables that are affected by X. In this case, both X and the downstream variables may be identified as significant variables in the T^2 decomposition. However, as the fault only occurs in X, the diagnostic decision that concludes not only X but also the downstream variables are the root causes of the fault is misleading. This situation becomes more severe as the numbers of variables and faults increase.

This paper studies how to effectively utilize the causal relationships, represented by a Bayesian network (BN in short) (Korb and Nicholson (2003)), to reduce the computational complexity and improve the diagnosability of the MTY approach. The adoption of BN in this research is appealing mainly because it provides an effective model for describing causal influences. Although BN is mathematically defined in terms of probabilistic independence statements, a connection can be made between this characterization and the notion of causal influences (Heckman, Meek and Cooper (1997), pearl and Verma (1991), Spirtes, Glymour and Scheines (1993)). The causal interpretations facilitate the application of BN to solve many real-world problems (Mani and Cooper (1999), Friedman *et al* (2000), Li and Shi (2005)). Based on the BN model, a causation-based T^2 decomposition method is proposed in this paper. It will be shown that the number of terms to be examined in the decomposition of an out-of-control

T^2 statistic is reduced to p by this method. In addition to the enhancement of computational efficiency, the improvement in diagnosability is also significant.

The rest of the paper is organized as follows. First, the key notations and concepts of Bayesian network are briefly introduced. Next, the proposed causation-based T^2 decomposition method is illustrated. Then, an example is used to demonstrate the proposed method. Finally, the conclusion is given.

Key notations and concepts of Bayesian network

A BN is represented as a **directed acyclic graph** (DAG), i.e., a set of nodes $\{X_1,\ldots,X_p\}$ connected by directed arcs (see Figure 1 for an example). A directed graph is **acyclic** if there is no directed path $X_i \to \cdots \to X_j$ such that $X_i = X_j$. The nodes represent random variables. If there is a directed arc from X_i to X_j, i.e., $X_i \to X_j$, X_i is called a **parent** of X_j. If there is a directed path from X_j to X_k, i.e., $X_j \to \cdots \to X_k$, X_k is called a **descendant** of X_j. In this paper, let $\mathbf{PA}(X_j)$ denote the set containing all parents of X_j and let $\mathbf{DE}(X_j)$ denote the set containing all descendants of X_j. For example, in Figure 1, $\mathbf{PA}(X_5)=\{X_2, X_3\}$ and $\mathbf{DE}(X_4)=\{X_3, X_5\}$.

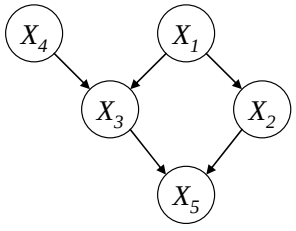


Fig. 1 The DAG of a Bayesian network

A DAG encodes the **Markov Condition**, i.e., a variable is independent of its non-descendants given its parents. If the directions of the arcs in a DAG have causal interpretations, i.e., the parents of a variable are its direct causes, then the BN becomes a causal model, or called as **causal network**.

A BN can also be represented as the conditional distribution of each variable given its parents, i.e., $P\left(X_j \mid \mathbf{PA}\left(X_j\right)\right)$. By applying the chain rule of probabilities and utilizing the conditional independencies implied by the Markov condition, the joint distribution of all variables in a BN can be written as

$$P\left(X_1,\ldots,X_p\right)=\prod_{j=1}^{p} P\left(X_j \mid \mathbf{PA}\left(X_j\right)\right). \tag{1}$$

Therefore, a BN representation allows the joint distribution to be decomposed into a set of independent conditional probabilities.

In this paper, it is assumed that all variables in a BN are continuous and a variable X_j, given its parents of m_j variables denoted by $\mathbf{PA}\left(X_j\right)=\left\{PA_1\left(X_j\right),\ldots,PA_{m_j}\left(X_j\right)\right\}$, follows a Gaussian distribution, i.e.,

$$P\left(X_j \mid pa_1\left(X_j\right),\ldots,pa_{m_j}\left(X_j\right)\right)\sim N\left(\alpha_{0j}+\sum_{k=1}^{m_j}\alpha_{kj}\,pa_k\left(X_j\right),\sigma_j^2\right), \tag{2}$$

where $pa_k\left(X_j\right)$ denotes the value of $PA_k\left(X_j\right)$; α_{0j}, α_{kj} are the coefficients that linearly link the values of X_j's parents with the mean of X_j; σ_j^2 is the variance of X_j and σ_j^2 is not influenced by the values of X_j's parents. This representation is called **Linear Gaussian** in BN. If all

variables in a network have linear Gaussian conditional distributions, then the joint distribution is a multivariate Gaussian (Lauritzen and Wermuth (1989)).

If the linear Gaussian network has causal interpretations, an alternative representation to (2) can be given as (Korb and Nicholson, (2003))

$$X_j = \alpha_{0j} + \sum_{k=1}^{m_j} \alpha_{kj} PA_k\left(X_j\right) + V_j, \tag{3}$$

where V_j is the disturbance, $V_j \sim N\left(0, \sigma_j^2\right)$. In order to directly compare the relative strength of the causal influences by $PA_k\left(X_j\right)$, it is customary to use standardized variables (i.e., the variables having means equal to 0 and standard deviations equal to 1) in a linear Gaussian BN, denoted as Z_j. Then (3) can be written as

$$Z_j = \sum_{k=1}^{m_j} p\left(PA_k\left(Z_j\right), Z_j\right) PA_k\left(Z_j\right) + E_j, \tag{4}$$

where $p(\bullet,\bullet)$ are called the **path coefficient;** $PA_k\left(Z_j\right)$ is the $PA_k\left(X_j\right)$ after standardization; $E_j \sim N\left(0, \tilde{\sigma}_j^2\right)$ and $\tilde{\sigma}_j^2$ is given by (Weisberg (1985))

$$\tilde{\sigma}_j^2 = 1 - \sum_{k=1}^{m_j} p\left(PA_k\left(Z_j\right), Z_j\right) \rho\left(PA_k\left(Z_j\right), Z_j\right), \tag{5}$$

where $\rho(\bullet,\bullet)$ denotes the correlation coefficient between two variables**.**

Causation-based T^2 decomposition

The MTY approach generates $p!$ distinct decompositions. With the aid of the causal relationships implied by a BN, it will be shown in this section that those decompositions can be classified into two categories (i.e., Type-A and Type-B) according to their diagnosability. The

characteristics of each type of decompositions are discussed. Then, it is proven that all Type-B decompositions will converge to one decomposition, called "causation-based T^2 decomposition". Finally, the analytic results are summarized and the procedure of applying causation-based T^2 decomposition in monitoring and diagnosis is illustrated.

Brief review of the MTY approach

Let $\mathbf{X}$ represent a $p-$dimensional random vector and $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\mathbf{X}})$ with known mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$. The Hotelling's (1947) T^2 statistic is

$$T^2 = (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2 \tag{6}$$

where $(\mathbf{X} - \boldsymbol{\mu})^T$ is the transpose of $(\mathbf{X} - \boldsymbol{\mu})$ and χ_p^2 represents a χ^2 distribution with p degrees of freedom. Without loss of generality, it is assumed in this paper that all variables are standardized, i.e., $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}$ is used to denote the correlation matrix of the standardized variables $\mathbf{Z}$. Then (6) can be written as:

$$T^2 = \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} \tag{7}$$

Mason, Tracy and Young (1995) proposed a procedure to decompose the T^2 into independent components. This procedure starts with specifying an ordering of the p variables and then partitions the T^2 as

$$T^2 = {}^rT^2 = \sum_{j=1}^{p} {}^rT^2_{j \bullet 1,\ldots,j-1} \,. \tag{8}$$

In (8), r serves as an indicator of the specific ordering. Since p variables have $p!$ different orderings, there are $p!$ distinct decompositions, denoted by ${}^rT^2$ ($r = 1, \ldots, p!$), which yield the

same T^2. ${}^rT_{j\bullet 1,\ldots,j-1}$ is the j^{th} variable in ordering r, denoted by ${}^rZ_j$, adjusted by the mean and standard deviation of the conditional distribution of ${}^rZ_j$ given ${}^rZ_1,\ldots,{}^rZ_{j-1}$, i.e.,

$$ {}^rT_{j\bullet 1,\ldots,j-1} = \frac{{}^rZ_j - {}^r\mu_{j\bullet 1,\ldots,j-1}}{{}^r\sigma_{j\bullet 1,\ldots,j-1}} \tag{9} $$

where

$$ {}^r\mu_{j\bullet 1,\ldots,j-1} = \sum_{i=1}^{j-1} {}^r\beta_i \; {}^rZ_i, \tag{10} $$

${}^r\beta_i$ are the regression coefficients of ${}^rZ_j$ regressed on ${}^rZ_i$, $i = 1,\ldots, j-1$, and according to Weisberg (1985),

$$ {}^r\sigma^2_{j\bullet 1,\ldots,j-1} = 1 - \sum_{i=1}^{j-1} {}^r\beta_i \; \rho\left({}^rZ_j, {}^rZ_i\right). \tag{11} $$

For example, if $p = 3$, 3!=6 decompositions are possible, as shown in (12):

$$ \begin{aligned} T^2 = {}^1T^2 &= T_1^2 + T_{2\bullet 1}^2 + T_{3\bullet 1,2}^2 \\ = {}^2T^2 &= T_1^2 + T_{3\bullet 1}^2 + T_{2\bullet 1,3}^2 \\ = {}^3T^2 &= T_2^2 + T_{3\bullet 2}^2 + T_{1\bullet 2,3}^2 \\ = {}^4T^2 &= T_2^2 + T_{1\bullet 2}^2 + T_{3\bullet 1,2}^2 \\ = {}^5T^2 &= T_3^2 + T_{1\bullet 3}^2 + T_{2\bullet 1,3}^2 \\ = {}^6T^2 &= T_3^2 + T_{2\bullet 3}^2 + T_{1\bullet 2,3}^2 \end{aligned} . \tag{12} $$

With the aid of the decompositions, it is possible to identify which variables significantly contribute to an out-of-control T^2 signal.

Classification of the decompositions

BN with causal interpretations provides an effective tool to identify how the fault in one variable propagates to other variables. Specifically, if a fault occurs in Z_j, it will affect all Z_j's

descendents but leave the non-descendents unaffected. This information can be utilized in diagnosing a fault, i.e., not only identifying the variables that significantly contribute an out-of-control T^2 signal, but also exactly locating "the" variables that are the root causes of the fault.

However, in the MTY approach, not all $p!$ decompositions facilitate the fault diagnosis. Hawkins (1993) points out that if a variable is regression-adjusted by its downstream variables, the mean shift in this variable will be diluted. Here downstream variables are conceptually equivalent to the descendents in a BN. Because the ${}^{r}T_{j\bullet 1,\ldots,j-1}$ defined in (9) is obtained by applying regression adjustment to ${}^{r}Z_j$ for ${}^{r}Z_1,\ldots,{}^{r}Z_{j-1}$, Hawkins' statement implies that if $\{{}^{r}Z_1,\ldots,{}^{r}Z_{j-1}\}$ contains the descendents of ${}^{r}Z_j$, then the mean shift in ${}^{r}Z_j$ will be diluted in ${}^{r}T_{j\bullet 1,\ldots,j-1}$, i.e., using ${}^{r}T_{j\bullet 1,\ldots,j-1}$ may not be able to precisely identify the mean shift in ${}^{r}Z_j$. Based on this consideration, the $p!$ decompositions in the MTY approach can be divided into two categories in terms of whether there exits a ${}^{r}T_{j\bullet 1,\ldots,j-1}$ in the decomposition that dilutes the mean shift in ${}^{r}Z_j$. The definition of those two categories is given bellow, where $\cap$ denotes the intersection of two sets and Φ denotes an empty set.

Definition 1: For a decomposition ${}^{r}T^2$, if there exists a term ${}^{r}T_{j\bullet 1,\ldots,j-1}$ such that $\mathbf{DE}({}^{r}Z_j)\cap\{{}^{r}Z_1,\ldots,{}^{r}Z_{j-1}\}\neq\Phi$, i.e., $\{{}^{r}Z_1,\ldots,{}^{r}Z_{j-1}\}$ contains at least one descendent of ${}^{r}Z_j$, ${}^{r}T^2$ is called Type-A decomposition. Otherwise, if $\{{}^{r}Z_1,\ldots,{}^{r}Z_{j-1}\}$ contains no descendent of ${}^{r}Z_j$, ${}^{r}T^2$ is called Type-B decomposition.

According to Definition 1, a Type-A decomposition must contain a term ${}^{r}T_{j\bullet 1,\ldots,j-1}$ in which $\{{}^{r}Z_1,\ldots,{}^{r}Z_{j-1}\}$ includes some descendents of ${}^{r}Z_j$. If ${}^{r}Z_j$ has a mean shift, using ${}^{r}T_{j\bullet 1,\ldots,j-1}$

for diagnosis has a risk of miss-detecting the shift as the shift is diluted in ${}^{r}T_{j\bullet 1,\ldots,j-1}$. Thus, compared with Type-B decomposition, Type-A decomposition has lower diagnosability. This point is further illustrated through the following example.

An illustrative example:

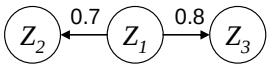


Fig. 2 The linear Gaussian BN of a three-variable system

Consider a simple system of three variables Z_1, Z_2 and Z_3, whose causal relationships are represented by the linear Gaussian BN in Figure 2. (The number on top of each arc is the path coefficient.) The 3!=6 possible decompositions of a T^2 statistic by the MTY approach are given in (12). Among the six decompositions, ${}^{1}T^2$ and ${}^{2}T^2$ are Type-B decompositions while ${}^{3}T^2$, ${}^{4}T^2$, ${}^{5}T^2$ and ${}^{6}T^2$ are Type-A decompositions, according to Definition 1. In ${}^{3}T^2$, the term $T_{1\bullet 2,3}$ regresses Z_1 on its descendents Z_2 and Z_3. If Z_1 has a mean shift δ, the detected shift using $T_{1\bullet 2,3}$ is $E\left(T_{1\bullet 2,3}\right)=0.52\delta$ (derivation given in the appendix II). It can be seen that the mean shift in Z_1 is diluted in $T_{1\bullet 2,3}$. Similar phenomena can be found in other Type-A decompositions ${}^{4}T^2$, ${}^{5}T^2$ and ${}^{6}T^2$. However, in the Type-B decomposition ${}^{1}T^2$, the mean shift in Z_j will not be diluted in $T_{j\bullet 1,\ldots,j-1}$, $j=1,2,3$. The computations that support this conclusion are summarized in Table 1(a) (derivation given in the appendix II). For example, if Z_2 has a mean shift δ, the detected shift using $T_{2\bullet 1}$ is $E\left(T_{2\bullet 1}\right)=1.4\delta$. Thus, the mean shift in Z_2 is not diluted but enlarged in $T_{2\bullet 1}$. This enlargement increases the sensitivity in detecting the mean shift. Similar results can be obtained for the other Type-B decomposition ${}^{2}T^2$, as shown in Table 1(b).

Table 1 The detected mean shifts by Type-B decompositions

(a) ${}^{1}T^2 = T_1^2 + T_{2\bullet 1}^2 + T_{3\bullet 1,2}^2$ (b) ${}^{2}T^2 = T_1^2 + T_{3\bullet 1}^2 + T_{2\bullet 1,3}^2$

The variable with a mean shift δ	The detected mean shift by the terms in ${}^{1}T^2$
Z_1	$E(T_1) = \delta$
Z_2	$E(T_{2\bullet 1}) = 1.4\delta$
Z_3	$E(T_{3\bullet 1,2}) = 1.7\delta$

The variable with a mean shift δ	The detected mean shift by the terms in ${}^{2}T^2$
Z_1	$E(T_1) = \delta$
Z_2	$E(T_{2\bullet 1,3}) = 1.4\delta$
Z_3	$E(T_{3\bullet 1}) = 1.7\delta$

The above discussion and example illustrate that Type-A decompositions have lower sensitivity in detecting the mean shifts than Type-B decompositions. Thus, to facilitate the root cause diagnosis, Type-B decompositions should be used.

Causation-based T^2 decomposition

This section proves that all Type-B decompositions converge to a unique decomposition, called causation-based T^2 decomposition.

Proposition 1: If ${}^{r}T^2$ is a Type-B decomposition, then ${}^{r}T^2 = \sum_{j=1}^{p} {}^{r}T_{j\bullet \mathbf{PA}({}^{r}Z_j)}^2$.

This proposition implies that if ${}^{r}T^2$ is a Type-B decomposition, then each individual term in the decomposition using the MTY approach, i.e., ${}^{r}T_{j\bullet 1,\ldots,j-1}^2$, becomes ${}^{r}T_{j\bullet \mathbf{PA}({}^{r}Z_j)}^2$. Detailed proof can be found in the appendix.

Once a BN is built, the parents of a variable can be identified from the DAG. In other words, the parents $\mathbf{PA}(Z_j)$ of Z_j do not change with the different orderings of the variables. Therefore, the left superscript r in Proposition 1 can be removed, indicating that all Type-B decompositions converge to a unique decomposition, which is the causation-based T^2 decomposition defined as follows.

Definition 2: The causation-based T^2 decomposition of a T^2 statistic is

$$T^2 = \sum_{j=1}^{p} T^2_{j\bullet \mathbf{PA}(Z_j)} \ . \tag{13}$$

To illustrate, the example in Figure 2 is revisited. The causation-based T^2 decomposition by Definition 2 is $T^2 = T_1^2 + T_{2\bullet 1}^2 + T_{3\bullet 1}^2$. According to Proposition 1, the two Type-B decompositions ${}^1T^2$ and ${}^2T^2$ converge to the causation-based T^2 decomposition, i.e., $T^2 = {}^1T^2 = {}^2T^2 = T_1^2 + T_{2\bullet 1}^2 + T_{3\bullet 1}^2$.

The proof of Proposition 1 includes two necessary steps. First, it needs to be proven that in a Type-B decomposition ${}^rT^2 = \sum_{j=1}^{p} {}^rT^2_{j\bullet 1,\ldots,j-1}$, the conditional set $\{{}^rZ_1,\ldots,{}^rZ_{j-1}\}$ of the variable ${}^rZ_j$ include all the parents of ${}^rZ_j$, $j = 1,\ldots,p$. This is Lemma 1 given bellow. Second, it needs to be shown that the regression coefficients of ${}^rZ_j$ regressed on its non-parents in $\{{}^rZ_1,\ldots,{}^rZ_{j-1}\}$ are equal to zero, $j = 1,\ldots,p$. Proving this utilizes the conclusion in Lemma 2 and the definition of ${}^rT_{j\bullet 1,\ldots,j-1}$ in (9), (10) and (11).

Lemma 1: If ${}^rT^2$ is a Type-B decomposition, then for any ${}^rT^2_{j\bullet 1,\ldots,j-1}$ $(j = 1,\ldots,p)$, $\mathbf{PA}({}^rZ_j) \subseteq \{{}^rZ_1,\ldots,{}^rZ_{j-1}\}$.

Proof by contradiction: If there exists a ${}^rT^2_{j\bullet 1,\ldots,j-1}$ such that at least one parent of ${}^rZ_j$, denoted by ${}^rZ_l$, does not belong to $\{{}^rZ_1,\ldots,{}^rZ_{j-1}\}$, i.e., $l \notin \{1,\ldots,j-1\}$ in ordering r, then $l \in \{j+1,\ldots,p\}$. Thus, according to (8), ${}^rZ_j \in \{{}^rZ_1,\ldots,{}^rZ_{l-1}\}$ for ${}^rT^2_{l\bullet 1,\ldots,l-1}$, i.e., $\{{}^rZ_1,\ldots,{}^rZ_{l-1}\}$

contains a descendent of ${}^{r}Z_l$, implying that ${}^{r}T^2$ is a Type-A decomposition, which conflicts with the given condition that "${}^{r}T^2$ is a Type-B decomposition". Δ

Lemma 2: if Z_i is a non-descendent of Z_j, then the correlation between Z_i and Z_j, $\rho(Z_i, Z_j)$, can be written as $\rho(Z_i, Z_j) = \sum_{k=1}^{m_j} \rho(Z_i, PA_k(Z_j)) p(PA_k(Z_j), Z_j)$.

Proof:

$\rho(Z_i, Z_j) = E(Z_i Z_j) = E\left(Z_i \left(\sum_{k=1}^{m_j} p(PA_k(Z_j), Z_j) PA_k(Z_j) + E_j \right) \right) = \sum_{k=1}^{m_j} \rho(Z_i, PA_k(Z_j)) p(PA_k(Z_j), Z_j) + E(Z_i E_j)$, where the 2^{nd} equality holds by replacing Z_j with the right-hand side of (4). Because E_j is the disturbance to Z_j, it will impact Z_j and also Z_j's descendents through Z_j. But E_j cannot affect Z_i which is a non-descendent of Z_j. Thus, $E(Z_i E_j) = 0$. Δ

Graphically, Lemma 2 can be depicted by Figure 3, where dash lines represent correlations and solid lines with arrows are the arcs in a BN.

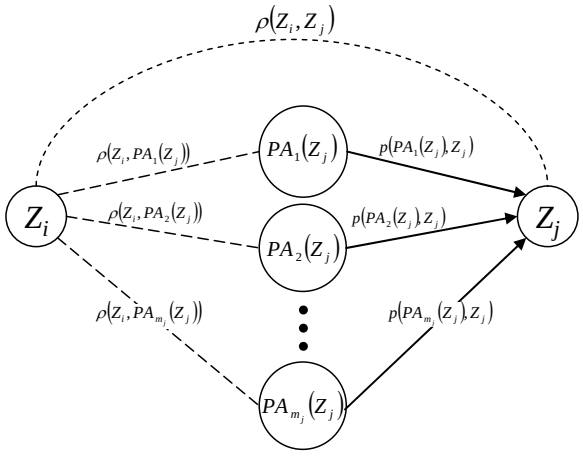


Fig. 3 Graphical illustration of Lemma 2

Monitoring and diagnosis with causation-based T^2 decomposition

Figure 4 shows the flow chart of applying causation-based T^2 decomposition to the monitoring and diagnosis in a system of p variables $\{Z_1,\ldots,Z_p\}$. A linear Gaussian BN is first constructed to describe the causal relationships among the variables. Then, a T^2 control chart (Alt (1985)) is established on the samples with the T^2 statistic calculated by (7). Once an out-of control T^2 is detected, this T^2 is decomposed by applying the causation-based T^2 decomposition defined in (13). In (13), each $T^2_{j\bullet\mathbf{PA}(Z_j)}$ is independent of one another. If the $\mathbf{\Sigma}$ of $\{Z_1,\ldots,Z_p\}$ is known, $T^2_{j\bullet\mathbf{PA}(Z_j)}$ follows a χ^2_1 distribution. Thus, one can compare each $T^2_{j\bullet\mathbf{PA}(Z_j)}$ with $\chi^2_{1,\alpha}$ (the upper α percentage point of χ^2_1) to determine which $T^2_{j\bullet\mathbf{PA}(Z_j)}$ significantly contributes to the out-of-control T^2. A significant $T^2_{j\bullet\mathbf{PA}(Z_j)}$ implies that Z_j is high likely to have experienced a mean shift.

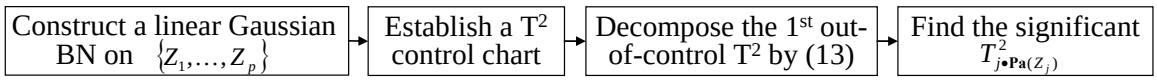


Fig. 4 Flow chart of applying causation-based T^2 decomposition to monitoring and diagnosis

Example

The diagnosability of the proposed causation-based T^2 decomposition method is demonstrated through an example of a manufacturing system. The system consists of five variables: $Z_1 \sim Z_4$ are process variables and Z_5 is the product quality variable. A linear Gaussian BN is constructed on those variables and represented as the DAG in Figure 5. The number beside each arc is the path coefficient.

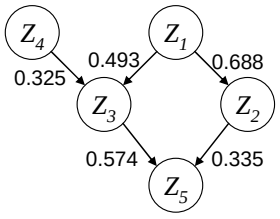


Fig. 5 The linear Gaussian BN of a manufacturing system

Because the manufacturing system has five variables, there are five potential single-fault scenarios, in each of which only one variable has a mean shift, and 26 potential multiple-fault scenarios, in each of which more than one variable have mean shifts. A total of 31 fault scenarios are shown in the first six columns of Table 1, where δ denotes the magnitude of the mean shift in the unit of standard deviation.

Table 1 Fault scenarios and the associated $T^2_{j\bullet \mathbf{PA}(Z_j)}$, $j = 1,\ldots,5$ ($\delta = 3$)

s	Z_1	Z_2	Z_3	Z_4	Z_5	$T^2_{1\cdot \mathbf{Pa}(Z1)}$	$T^2_{2\cdot \mathbf{Pa}(Z2)}$	$T^2_{3\cdot \mathbf{Pa}(Z3)}$	$T^2_{4\cdot \mathbf{Pa}(Z4)}$	$T^2_{5\cdot \mathbf{Pa}(Z5)}$	d_s
1	δ					22.78*	0.6	0.78	0.56	0.03	1
2		δ				0.29	23.84*	0.03	3.68	0.12	1
3			δ			1.6	2.1	5.89*	3.58	0.13	1
4				δ		0.04	0.07	0.03	13.71*	1.69	1
5					δ	0.9	1.05	1.46	2.1	17.44*	1
6	δ	δ				11.14*	20.22*	0.01	1.29	0.48	1
7	δ		δ			8.7*	0.65	9.26*	0.79	0	1
8	δ			δ		8.62*	1.3	0.04	5.29*	0.84	1
9	δ				δ	12.65*	0.14	0.95	0.2	14.4*	1
10		δ	δ			0.32	20.77*	15*	0.01	3.28	1
11		δ		δ		0.61	17.69*	0.71	10*	0.45	1
12		δ			δ	0.11	32.55*	0.02	0.15	18.31*	1
13			δ	δ		0.11	0.91	13.59*	4.75*	0.8	1
14			δ		δ	1.04	0.37	14.12*	1.79	19.75*	1
15				δ	δ	0.09	0.06	0.32	7.19*	20.65*	1
16	δ	δ	δ			6.04*	14.71*	6.76*	0.06	2.17	1
17	δ	δ		δ		14.6*	10.63*	0.03	6.5*	0.86	1
18	δ	δ			δ	9.69*	14.7*	2.58	1.13	27.68*	1
19	δ		δ	δ		9.45*	0	14.59*	2.94	1.13	0
20	δ		δ		δ	5.23*	2.94	11.06*	0.01	23.28*	1
21	δ			δ	δ	15.51*	0.05	0.17	12.36*	22.82*	1
22		δ	δ	δ		3.1	33.17*	6.67*	6.49*	0.09	1
23		δ	δ		δ	0.09	9.9*	13.37*	2.25	16.15*	1
24		δ		δ	δ	0.34	19.53*	0.2	25.67*	18.89*	1
25			δ	δ	δ	0.02	0.39	16.86*	7.05*	14.37*	1
26	δ	δ	δ	δ		2.84	26.82*	13.97*	4.46*	4.92*	0
27	δ	δ	δ		δ	3.87*	11.61*	21.86*	2.86	23.73*	1
28	δ	δ		δ	δ	13.84*	13.06*	0.04	9.78*	38.11*	1
29	δ		δ	δ	δ	20.83*	1.29	22.25*	9.7*	25.77*	1
30		δ	δ	δ	δ	1.24	13.73*	18.23*	16.29*	13.6*	1
31	δ	δ	δ	δ	δ	9.96*	29.71*	8.49*	10.64*	19.08*	1

The simulation study includes the following steps.

1. For every fault scenario s $(s = 1,\ldots,31)$, the steps 1.1~1.5 bellow are performed.

   1.1 A mean shift of three standard deviations (i.e., $\delta = 3$) is introduced to a variable Z_j, if Z_j has a mean shift in this fault scenario.

   1.2 One dataset is generated with 100 samples, in which the introduced mean shift occurs at the 1st sample.

   1.3 A T^2 chart is constructed. $\chi^2_{5,0.05} = 11.07$ is adopted as the control limit.

   1.4 Once an out-of-control T^2 is detected, it is further decomposed into five independent terms $T^2_{j\bullet\mathbf{PA}(Z_j)}$ $(j = 1,\ldots,5)$ based on (13). Each $T^2_{j\bullet\mathbf{PA}(X_j)}$ is compared with $\chi^2_{1,0.05} = 3.84$. A significant $T^2_{j\bullet\mathbf{PA}(X_j)}$ indicates that Z_j has a mean shift.

   1.5 The diagnostic result in step 1.4 is compared with the truth of the fault scenario in step 1.1. If they are consistent, $d_s = 1$; otherwise, $d_s = 0$.

2. The diagnostic accuracy is computed as $\gamma_{f_\text{causation}} = \left(\sum_{s=1}^{31} d_s \Big/ 31 \right)\%$.

3. Steps 1 and 2 are repeated for 1000 times and the mean and standard deviation of $\gamma_{f_\text{causation}}$ are computed.

For better understanding of the simulation procedure, one table from steps 1~2 is shown in Table 1, where $\gamma_{f_\text{causation}} = 29/31 = 93.5\%$. 1000 such tables are generated in step 3. The mean and standard deviation of $\gamma_{f_\text{causation}}$ are recorded in the second row of Table 2. The diagnostic accuracy of causation-based T^2 decomposition is $73.6\% \pm 7.7\%$.

Table 2 Comparison of diagnostic accuracy between causation-based T^2 decomposition method and the MTY approach ($\delta = 3$)

	Mean	Standard deviation
$\gamma_{f_\text{causation}}$	73.6%	7.7%
γ_{f_MTY}	37.1%	7.4%

The proposed causation-based T^2 decomposition is also compared with the MTY approach. Because the MTY approach requires examining an excessive number of terms from the $p!$ decompositions, which is computationally involved and needs tremendous interpretation efforts, Mason, Tracy and Young (1997) proposes a sequential scheme to reduce the computations to a reasonable number. Briefly, this scheme starts with computing the unconditional terms T_j^2's. If any T_j^2's are significant, Z_j's are removed and T^2 statistic is calculated based on the rest variables. If this T^2 is insignificant (i.e., in control), the computation ends; otherwise, $T_{j\bullet i}^2$'s are computed for the remaining variables. The Z_j and Z_i corresponding to a significant $T_{j\bullet i}^2$ are removed and T^2 statistic is re-calculated. If this T^2 is significant (i.e., out of control), higher-order conditional terms are computed in a similar fashion until no variables remain for calculating T^2. The diagnostic decision will be made based on all the removed variables after the entire computation is complete. Specifically, if a variable Z_j is removed when checking the unconditional terms, it indicates that Z_j has a mean shift. Otherwise, it indicates that the relationship between Z_j and other variables has changed.

Following this scheme, the 31 fault scenarios are revisited. The simulation uses the same datasets as those generated in steps 1~3. The simulation procedure is similar to that in causation-

based T^2 decomposition except that step 1.4 is replaced with the step 1.4' bellow and the diagnostic accuracy is denoted by γ_{f_MTY}.

1.4' The out-of-control T^2 statistic is investigated according to the scheme. Specifically, each $T_j^2 (j = 1,\ldots,5)$ is compared with $\chi^2_{1,0.05} = 3.84$. A significant T_j^2 indicates that Z_j has a mean shift.

It should be pointed out that in applying the scheme to the simulation datasets, the computation ends after significant T_j^2 are identified, i.e., no conditional terms need to be calculated. This is because only mean shifts are introduced to the variables in this study and there are no changes in the variable relationships.

The mean and standard deviation of γ_{f_MTY} are recorded in the third row of Table 2. The diagnostic accuracy of the MTY approach is $37.1\% \pm 7.4\%$. Considering that the standard deviations of $\gamma_{f_causation}$ and γ_{f_MTY} are not significantly different, it can be concluded that the diagnostic accuracy of causation-based T^2 decomposition is almost twice as high as that of the MTY approach.

Conclusion

Multivariate process monitoring and diagnosis is an important and challenging issue. Multivariate SPC based on T^2 statistic can effectively detect system change but is not capable of diagnosing the root causes of the change. The MTY approach makes efforts to improve the diagnosability by decomposing the T^2 statistic. However, there are two major deficiencies of this

approach. First, it is computationally expensive as $p!$ different decompositions of the T^2 are needed. Second, the improvement in diagnosability is still limited.

This paper proposes a causation-based T^2 decomposition method, which integrates the causal relationships, represented by a BN, with the T^2 decomposition. The development of the method includes several steps. First, based on the BN, the $p!$ decompositions were classified into two categories, Type-A and Type-B. It was found that diagnosis with Type-A decompositions has a risk of miss-detecting the faults. Thus, Type-B decompositions should be used to ensure the diagnosability. Second, it was proved that all Type-B decompositions converge to a unique decomposition, called causation-based T^2 decomposition. Finally, the procedure of applying causation-based T^2 decomposition to monitoring and diagnosis was given.

Compared with the MTY approach, the proposed method has the advantages of significantly reduced computational complexity and enhanced diagnosability. Besides the analytical study, an example of a manufacturing system was also provided. Results showed that the diagnostic accuracy of causation-based T^2 decomposition is almost twice as high as that of the MTY approach.

Appendices

A.I Proof of Proposition 1

By Lemma 1, $\mathbf{PA}\left({}^{r}Z_j\right) \subseteq \left\{{}^{r}Z_1,\ldots,{}^{r}Z_{j-1}\right\}$. Let $\overline{\mathbf{PA}}\left({}^{r}Z_j\right)$ denote the complement of $\mathbf{PA}\left({}^{r}Z_j\right)$ with respect to $\left\{{}^{r}Z_1,\ldots,{}^{r}Z_{j-1}\right\}$. Based on the definition of ${}^{r}T_{j\bullet 1,\ldots,j-1}$ in (9), (10) and (11),

it is known that to prove this proposition is to prove $\boldsymbol{\beta}_{\overline{Pa}} = \mathbf{0}$ in the regression bellow, where $\boldsymbol{\beta}_{Pa}$ and $\boldsymbol{\beta}_{\overline{Pa}}$ are the regression coefficients of the predictors $\mathbf{PA}\left({}^{r}Z_j\right)$ and $\overline{\mathbf{PA}}\left({}^{r}Z_j\right)$, respectively.

$$ {}^{r}Z_j = \boldsymbol{\beta}_{Pa}^{T}\mathbf{PA}\left({}^{r}Z_j\right) + \boldsymbol{\beta}_{\overline{Pa}}^{T}\overline{\mathbf{PA}}\left({}^{r}Z_j\right) + {}^{r}e_j . \tag{A-1} $$

For a general regression model $Y = \boldsymbol{\beta}^{T}\mathbf{X} + \varepsilon$ (assuming all variables are standardized), least square estimation (LSE) gives $\mathbf{S}_{XX}\hat{\boldsymbol{\beta}} = \mathbf{S}_{XY}$, where $\mathbf{S}_{XX}$ is the sample correlation matrix of predictors $\mathbf{X}$, $\mathbf{S}_{XY}$ is a vector of sample correlations between predictors $\mathbf{X}$ and response Y, and $\hat{\boldsymbol{\beta}}$ is the LSE estimate of regression coefficients $\boldsymbol{\beta}$. If sample errors are not considered, $\mathbf{S}_{XX}$ and $\mathbf{S}_{XY}$ are replaced by their population counterparts $\boldsymbol{\Sigma}_{XX}$ and $\boldsymbol{\Sigma}_{XY}$ respectively, and $\boldsymbol{\beta}$ can be solved by $\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta} = \boldsymbol{\Sigma}_{XY}$. Substituting this general LSE notation with the notation in the specific regression model of (A-1) gives

$$ \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{T} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{Pa} \\ \boldsymbol{\beta}_{\overline{Pa}} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \\ \mathbf{F} \end{bmatrix}, \tag{A-2} $$

where $\mathbf{A} = cor\left(\mathbf{PA}\left({}^{r}Z_j\right), \mathbf{PA}\left({}^{r}Z_j\right)\right)$, $\mathbf{D} = cor\left(\overline{\mathbf{PA}}\left({}^{r}Z_j\right), \overline{\mathbf{PA}}\left({}^{r}Z_j\right)\right)$, $\mathbf{B} = cor\left(\mathbf{PA}\left({}^{r}Z_j\right), \overline{\mathbf{PA}}\left({}^{r}Z_j\right)\right)$, $\mathbf{G} = cor\left(\mathbf{PA}\left({}^{r}Z_j\right), {}^{r}Z_j\right)$, $\mathbf{F} = cor\left(\overline{\mathbf{PA}}\left({}^{r}Z_j\right), {}^{r}Z_j\right)$, and $cor(\bullet,\bullet)$ denotes the correlation matrix between two sets of random variables.

Let $f_h = \rho\left(\overline{PA}_h\left({}^{r}Z_j\right), {}^{r}Z_j\right)$ where $\overline{PA}_h\left({}^{r}Z_j\right) \in \overline{\mathbf{PA}}\left({}^{r}Z_j\right)$. Thus, $f_h \in \mathbf{F}$. Because ${}^{r}T^2$ is a Type-B decomposition, $\overline{PA}_h\left({}^{r}Z_j\right)$ must be a non-descendent of ${}^{r}Z_j$. According to Lemma 2,

$$f_h = \sum_l \rho\left(\overline{PA}_h\left({}^rZ_j\right), PA_l\left({}^rZ_j\right)\right) p\left(PA_l\left({}^rZ_j\right), {}^rZ_j\right), \tag{A-3}$$

where $PA_l\left({}^rZ_j\right) \in \mathbf{PA}\left({}^rZ_j\right)$.

By (A-2),

$$f_h = \sum_l \rho\left(\overline{PA}_h\left({}^rZ_j\right), PA_l\left({}^rZ_j\right)\right)\beta_{pa_l} + \sum_{h'} \rho\left(\overline{PA}_h\left({}^rZ_j\right), \overline{PA}_{h'}\left({}^rZ_j\right)\right)\beta_{\overline{pa}_{h'}}, \tag{A-4}$$

where β_{pa_l} is regression coefficient for the l^{th} parent in $\mathbf{PA}\left({}^rZ_j\right)$, and $\beta_{\overline{pa}_{h'}}$ is regression coefficient for the h'^{th} non-parent in $\overline{\mathbf{PA}}\left({}^rZ_j\right)$.

Equalizing (A-3) and (A-4) gives one solution to β_{pa_l} and $\beta_{\overline{pa}_{h'}}$, i.e.,

$$\beta_{pa_l} = p\left(PA_l\left({}^rZ_j\right), {}^rZ_j\right) \text{ and } \beta_{\overline{pa}_{h'}} = 0. \tag{A-5}$$

Similar to f_h, let $g_l = \rho\left(PA_l\left({}^rZ_j\right), {}^rZ_j\right)$. Thus $g_l \in \mathbf{G}$. According to Lemma 2,

$$g_l = \sum_{l'} \rho\left(PA_l\left({}^rZ_j\right), PA_{l'}\left({}^rZ_j\right)\right) p\left(PA_{l'}\left({}^rZ_j\right), {}^rZ_j\right). \tag{A-6}$$

By (A-2),

$$g_l = \sum_{l'} \rho\left(PA_l\left({}^rZ_j\right), PA_{l'}\left({}^rZ_j\right)\right)\beta_{pa_{l'}} + \sum_h \rho\left(\overline{PA}_l\left({}^rZ_j\right), \overline{PA}_h\left({}^rZ_j\right)\right)\beta_{\overline{pa}_h} \tag{A-7}$$

Equalizing (A-6) and (A-7) also gives one solution to $\beta_{pa_{l'}}$ and $\beta_{\overline{pa}_h}$, i.e.,

$$\beta_{pa_{l'}} = p\left(PA_{l'}\left({}^rZ_j\right), {}^rZ_j\right) \text{ and } \beta_{\overline{pa}_h} = 0, \tag{A-8}$$

which is the same as the solution in (A-5), implying that (A-5) is a solution to (A-2). Assuming that there is no multicollinearity in the variables, $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{bmatrix}$ is non-singular and (A-5) is the unique solution to (A-2). Thus, $\boldsymbol{\beta}_{\overline{Pa}} = \mathbf{0}$. Δ

A.II Calculation of $E\left(T_{j\bullet 1,\ldots j-1}\right)$ when Z_j has a mean shift δ

This section shows how to compute $E\left(T_{1\bullet 2,3}\right)$ when Z_1 has a mean shift δ. In a similar way, the $E\left(T_{j\bullet 1,\ldots j-1}\right)$ in Table 1 can be obtained.

Based on the definition in (9), $T_{1\bullet 2,3} = \left(Z_1 - \mu_{1\bullet 2,3}\right)/\sigma_{1\bullet 2,3}$, where $\mu_{1\bullet 2,3} = \beta_2 Z_2 + \beta_3 Z_3$ by (10) and $\sigma^2_{1\bullet 2,3} = 1 - \left(\beta_2 \rho\left(Z_1, Z_2\right) + \beta_3 \rho\left(Z_1, Z_3\right)\right)$ by (11). Thus,

$$E\left(T_{1\bullet 2,3}\right) = \frac{E\left(Z_1\right) - \left(\beta_2 E\left(Z_2\right) + \beta_3 E\left(Z_3\right)\right)}{\sqrt{1 - \left(\beta_2 \rho\left(Z_1, Z_2\right) + \beta_3 \rho\left(Z_1, Z_3\right)\right)}} \quad . \tag{A-9}$$

In (A-9), it is given that $E\left(Z_1\right) = \delta$. Then $E\left(Z_2\right) = 0.7\delta, E\left(Z_3\right) = 0.8\delta$ according to the BN in Figure 2. The path coefficients in the BN indicate that $\rho\left(Z_1, Z_2\right) = 0.7$, $\rho\left(Z_1, Z_3\right) = 0.8$ and $\rho\left(Z_2, Z_3\right) = 0.7 \times 0.8 = 0.56$. β_2 and β_3 can be solved by least square estimation, i.e.,

$$\begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 1 & \rho\left(Z_2, Z_3\right) \\ \rho\left(Z_2, Z_3\right) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho\left(Z_1, Z_2\right) \\ \rho\left(Z_1, Z_3\right) \end{bmatrix} = \begin{bmatrix} 1 & 0.56 \\ 0.56 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.7 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 0.3671 \\ 0.5944 \end{bmatrix}.$$

Inserting these numerical values into (A-9) gives $E\left(T_{1\bullet 2,3}\right) = 0.52\delta$. Δ

Reference

Alt, F. B. (1985). "Multivariate Quality Control" in *Encyclopedia of Statistical Sciences*, (Kotz, S.; Johnson, N.L.; and Read. C. R. eds.) Vol 6, pp. 111-122.

Friedman, N., *et al*, (2000). "Using Bayesian Networks to Analyze Expression Data", *Journal of Computational Biology*, 7(3-4), pp. 601-620.

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.

Hecherman, D., Meek, C. and Cooper, G. (1997). "A Bayesian Approach to Causal Discovery", Technical Report MSR-TR-97-05, Microsoft Research.

Hotelling, H. (1947). "Multivariate Quality Control", *Techniques of Statistical Analysis*, Eisenhart, Hastay and Wallis (eds.), McGraw-Hill, New York.

Korb, K.B. and Nicholson, A.E., (2003). *Bayesian Artificial Intelligence*, London, UK: Chapman & Hall/CRC.

Lauritzen, S. L. & Wermuth, N. (1989), "Graphical Models for Associations between Variables, Some of Which are Qualitative and Some Quantitative", *Annals of Statistics* 17, pp.31–57.

Li, J., and Shi, J (2005). "Knowledge Discovery from Observational Data for Process Control using Causal Bayesian", accepted by *IIE Transactions.*

Mani, S., and Cooper, G.F. (1999). "A study in Casual Discovery from Population-based Infant Birth and Death Records", *Proceedings of the AMIA Annual Fall Symposium*, Philadelphia, PA, pp. 315-319.

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., Tracy, N. D., and Young, J. C. (1997). "A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals". *Journal of Quality Technology* 29, pp. 396-406.

Pearl, J. and Verma, T. S. (1991). "A Theory of Inferred Causation", in *Principles of Knowledge Representation and Reasoning*: Proceedings of 2nd International Conference (KR'91'), pp. 441-452.

Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction and Search*, Springer-Verlag.

Tracy, N. D., Young, J. C., and Mason, R. L. (1992). "Multivariate Control Charts for Individual Observations". *Journal of Quality Technology* 24, pp. 88-95.

Weisberg, S. (1985). *Applied Linear Regression*, New York: John Wiley.