ENGINEERING DRIVEN CAUSE-EFFECT MODELING AND STATISTICAL ANALYSIS FOR MULTI-OPERATIONAL MACHINING PROCESS DIAGNOSIS

Jian Liu, Jing Li, Jianjun Shi
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48105

ABSTRACT

Process fault identification for product quality improvement is a critical issue in both design and manufacturing, especially for multistage manufacturing processes. In this paper, an integrated approach is proposed to develop cause-effect models from engineering knowledge and to conduct associated statistical analysis of the measurement data. First, a cause-effect diagram and predicted symptom vectors (PSV) are formulated to recognize the cause-effect relationship between process variables and product qualities. Then factor analysis and factor rotating technique are employed to extract the relationships reflected from measurement data. Finally the potential process faults are identified by comparing predicted symptoms and extracted symptoms. A case study is conducted to demonstrate the effectiveness of the proposed methodology.

1. INTRODUCTION

Variation reduction is essential to ensure the competence of manufacturers, which demands quick detection and identification of process changes and their underlying fault sources. Recent years witnessed the marked advance in automated in-process sensing and data collecting technologies for measuring a large number of product/process features. The resulting volume of multivariate data creates tremendous potential for monitoring and diagnosing quality-related problems in the manufacturing process. Historically, Statistical process control (SPC) (Woodall and Montgomery, 1999) was adopted as a major technique to monitor quality and process. Recently, multivariate statistical tools, such as Principle Component analysis (PCA) (Jackson, 1980) were introduced in the quality control problems to extract the variation structure from measurement data. However, without interpretation with domain knowledge, the statistical methods provide little diagnostic capability.

Researches have been launched to address the knowledge-based root cause diagnosis issues. Ceglarek, Shi and Wu (1994) firstly attempted to provide a reasoning mechanism based on the hierarchical model of the auto-body assembly process. State space modeling methodology is also developed to represent the geometrical relation between variation sources and the key product characteristics (Zhou, *et. al.*, 2003). Although it successfully solves the diagnostic problem, the accompanied modeling efforts are tremendous. Li and Zuo (1999), following the Failure Modes and Effect Analysis (FMEA) procedure, proposed a hybrid approach to qualitatively diagnose the root cause of a semi-conductors manufacturing process. The qualitative analysis of this approach significantly

simplifies its implementation. However, the conceptual representation of the process knowledge provides less potential for itself to be linked with measurement data.

Above reviews identify the necessity of an approach to effectively integrate the engineering knowledge with existing statistical techniques. In this paper, a methodology is proposed to provide a platform for representing engineering knowledge in the format of cause-effect models. Multivariate statistical tools are employed to detect process changes and extract the variation patterns. The underlying process faults are identified through comparing statistical analysis results and the cause-effect relationships.

The remainder of this article is organized as follows. Section 2 gives an overview of the methodology. Section 3 introduces the formulation of the knowledge representation and the procedure of generating the cause-effect diagram and predicted fault symptom descriptions. Statistical tools used for detecting process changes and extracting variation patterns are introduced in section 4, together with the detail of fault diagnose procedure. Section 5 provides a set of case studies to demonstrate the proposed methodology and section 6 lists conclusions and some discussion.

2. OVERVIEW OF THE METHODOLOGY

Root cause diagnosis for manufacturing process is a procedure that consists of five major tasks. (i) The engineering knowledge is organized and appropriately represented. (ii) Data collected from in-process sensing are tested to detect process changes (iii) Statistical analysis is launched to extract the implicit data patterns. (iv)Through linking the extracted data patterns with the engineering knowledge, potential fault sources are identified. (v)The diagnosis results can be used for process control or system maintenance for taking corrective actions. Corresponding to these tasks, a diagnosis methodology is proposed in this research. It consists of three main parts: knowledge representation, data analysis and fault identification, as shown in figure 1.

Process planning information provides the basis for the engineering knowledge representation. The cause-effect diagram is formulated to recognize the cause-effect relationships between potential process faults and quality variables.

Fault predicted symptoms are pseudo-quantitative descriptions of these relationships. The purpose of the data analysis is to detect the errors, extract their patterns and represent them in an interpretable format. Multivariate statistical tools are employed to recognize the spatial nature of faults' impacts and provide adequate interpretability. By combining the engineering knowledge representation and statistical analysis of the measurement data, the process faults can be identified.
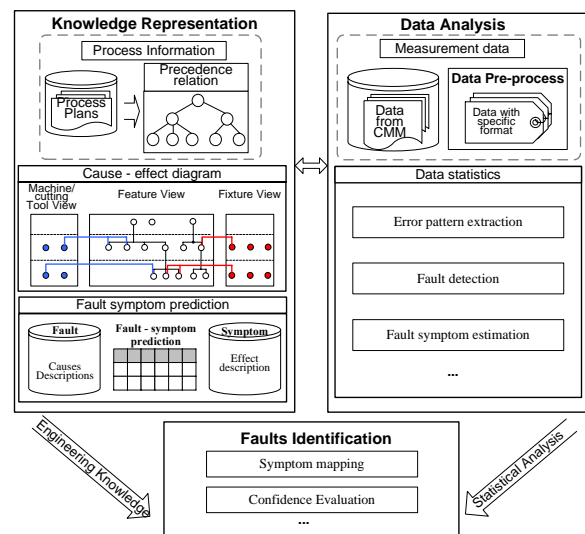

FIGURE 1. OVERVIEW OF THE METHODOLOGY

3. KNOWLEDGE REPRESENTATION

Often times, the lack of detailed description of manufacturing processes would hinder effective troubleshooting when problems occur. In this section, new concepts are proposed for representing the engineering knowledge and interfacing with statistical analysis results.

3.1 Formulation of cause-effect diagram

The first step of cause-effect modeling is to define the process flow and machine datum strategy selected for real process. Traditionally, precedence diagram is one of the most powerful techniques to address this planning issue (Liu and Yuan, 1996). However, it describes only the precedence constraints among in-process features, not the quality variables and their relationships with process variables. Huang, *et. al.* (2000) proposed a similar strategy, D-graph, for the purpose of process diagnosability assessment. Although it is a good representation of datum strategy, D-graph has its limitation. It

records the precedence relations between datums and in-process features, but not the cause-effect relations in terms of impacts of a datum fault on the specific quality variables. Furthermore, besides the datums, error may also be introduced by fixtures, cutting tools and/or machine tools selected in each operation. If we adopt the assumption that the parts entering the system are perfect, these device-induced errors become the main source of potential root causes.

Based on the above analysis, the requirements for a knowledge representation include: (i) it should be able to describe the precedence information, (ii) potential process faults should be included as complete as possible, and (iii) the cause-effect relationships between process faults and affected quality variables should be expressed explicitly.
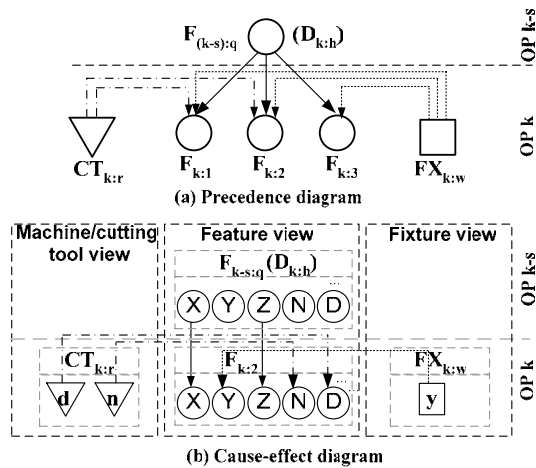

FIGURE 2. CAUSE-EFFECT DIAGRAM FORMULATION

Process FMEA is a group of activities intended to recognize and assess the potential failure modes associated with the manufacturing processes or quality concerns and to understand the effects of them (Prasad, 1991). Inspired by the FMEA technique and D-graph representation, this paper presents the cause-effect diagram, as shown in figure 2.

Process planning provides the information of devices (e.g. machine tool, cutting tools, fixtures, etc.), the precedence relations, datum strategies and in-process features formed in each operation. Figure 2 (a) is a partial description of them, where the following notations are adopted:
- The circle nodes, $\bigcirc$, denote the in-process features (e.g. $F_{(k-s):q}$, the qth feature generated in operation $k\text{-}s$.) and datum

features (e.g. $D_{k:h}$, the hth datum features used in operation k).
- The directed connections, $\rightarrow$, between them represent the precedence constraints.
- Triangle nodes, $\triangledown$, denote the cutting tools and machine tools (e.g. $CT_{k:r}$, the rth cutting tool used in operation k) selected to perform the manufacturing tasks.
- The directed connections, $\dashrightarrow$, from triangle nodes to circle nodes specify that the pointed features are generated by the pointing cutting tool and/or the pointing machine tool.
- The square nodes, $\square$, denote the fixture (e.g. $FX_{k:w}$, the wth fixture used in operation k) scheme selected to locate the work piece.
- The triangle nodes and square nodes are connected only to the in-process features in the same operation.

In this approach, all the quality variables of the in-process features are effect elements. For the causes, there are 3 types of faults considered.

(i) *Machine and cutting tool faults*, which are defined as the deviation of the cutting tools from their nominal path.
(ii) *Fixture faults*, which are defined as the imperfection of the locators.
(iii) *Datum faults*, which are defined as nonconformance of the in-process features used as datums in later operation(s).

As shown in figure 2 (b), the cause-effect diagram contains three views: feature view, fixture view and machine/cutting tools view. The layers in each view represent the operation sequence. In feature view, in-process features (e.g. $F_{k:2}$) are decomposed and expressed in terms of quality variables, i.e. the position (X, Y, Z), orientation (N) and dimension (D). These quality variables are deployed in different layers according to the precedence relations, and the directed connections between them represent the cause-effect relationships. For instance the connection from the X of $F_{(k-s):q}$ to the X of $F_{k:2}$ means that the imperfection of X position of datum feature $F_{(k-s):q}$ will cause the error of X position of the generated feature $F_{k:2}$. In feature view, connections appear only between the quality variables of datums and that of the in-process features (generated based on those datums). In fixture view and machine/cutting tools view, nodes are also decomposed and expressed in terms of attributes of fixture and

machine/cutting tools. For instance, a locating pin $FX_{k:w}$ is described with its location (e.g. y) and size (e.g. d). Connections between these attributes and the quality variables also represent cause-effect relations. In the cause-effect diagram, all the nodes that have outgoing connections are cause nodes and those having incoming connections are effect nodes.

3.2 Predicted symptom representation

The cause-effect diagram cannot be directly matched with the statistical analysis results. However, it contains the information about causes, affected quality variables and effect patterns, which is called *symptom* in this paper. The description of the symptom of a potential fault source should include following aspects:

(i) What quality variables are affected by one possible fault source?
(ii) How are the effects exerted on those affected quality variables?
(iii) How do the effects of a specific fault source propagate?

The concept of predicted symptom vector (PSV) is proposed to capture those aspects. For a manufacturing process consisting of M operations, it can be known from the cause-effect diagram that there are totally n quality variables. For each operation k ($k = 1, 2, \dots, M$), there are m_k potential faults and the total number of potential process faults is p, where $p = \Sigma_{k=1}^{M} m_k$. Assuming that errors are introduced only by the devices (fixtures, tools and machines), therefore only device faults are included in m_k. For a process fault i_k and quality variable j, the impact of i_k on j is defined as $\theta_{i_k,j}$. It is time consuming, if not impossible, to derive the exact value of $\theta_{i_k,j}$'s. However, it is often comparatively easy to determine $\theta_{i_k,j}$'s sign and the interrelationship between $\theta_{i_k,j}$ and $\theta_{i_k,l}$, (i.e $\theta_{i_k,j} < \theta_{i_k,l}$ or $\theta_{i_k,j} > \theta_{i_k,l}$). For instance, if the positive deviation, in the fixture/machine coordinate system, of a process variable i_k causes the quality variable j deviate along the positive direction, in the part coordinate system (Zhou, *et. al*, 2003), then $\theta_{i_k,j} > 0$, and vice versa. Therefore, in this approach, $\theta_{i_k,j}$'s are only represented pseudo-quantitatively. Let $^0\phi_{i_k,j}$ be

the 0-order predicted cause-effect relation between the i_k^{th} potential fault in operation k and the jth quality variable.

$$^0\phi_{i_k,j} = \begin{cases} 1, & \text{if } \theta_{i_k,j} > 0 \\ 0, & \text{if } \theta_{i_k,j} = 0 \\ -1, & \text{if } \theta_{i_k,j} < 0 \end{cases} \tag{1}$$

Where $k = 1,2,\dots,M$; $i_k = 1,2,\dots,m_k$ and $j = 1,2,\dots,n$

By stacking up all the n $^0\phi_{i_k,j}$'s according to the operational sequence, the 0-order PSV of the i_k^{th} fault in operation k can be expressed as $_k^0 f_{i_k} = [\,^0\phi_{i_k,1},\,^0\phi_{i_k,2},\dots,\,^0\phi_{i_k,n}]^{\text{T}}$ Those non-zero $^0\phi_{i_k,j}$'s indicate the affected quality variables and the sign of non-zero $^0\phi_{i_k,j}$'s reflect how the effects are exerted (i.e. if a process variable changes in positive direction, how this change is reflected at the quality variables). Meanwhile, the interrelations between effects are also critical for identifying and differentiating the faults, especially for the case that different faults (non-orthogonal to each other) affecting the same set of quality variables present simultaneously. 1-order PSV describes these interrelations. Let

$$^1\phi_{i_k,j} = \begin{cases} 1 & \text{if } ^0\phi_{i_k,j} \neq 0,\ ^0\phi_{i_k,l} \neq 0,\ \theta_{i_k,j} > \theta_{i_k,l}, \\ & \text{and } ^0\phi_{i_k,t} = 0\,(t = l+1,l+2,\dots,j-1) \\ 0 & \text{if } ^0\phi_{i_k,j} \neq 0,\ ^0\phi_{i_k,l} \neq 0,\ \theta_{i_k,j} = \theta_{i_k,l}, \\ & \text{and } ^0\phi_{i_k,t} = 0\,(t = l+1,l+2,\dots,j-1) \\ -1 & \text{if } ^0\phi_{i_k,j} \neq 0,\ ^0\phi_{i_k,l} \neq 0,\ \theta_{i_k,j} < \theta_{i_k,l}, \\ & \text{and } ^0\phi_{i_k,t} = 0\,(t = l+1,l+2,\dots,j-1) \\ \infty & \text{if } ^0\phi_{i_k,j} = 0 \text{ or } ^0\phi_{i_k,v} = 0 \text{ for all } v < j \end{cases} \tag{2}$$

$^1\phi_{i_k,j}$ reflects the interrelation between magnitude of the jth effect and that of the previous non-zero effect. So, 1-order PSV $_k^1 f_{i_k} = [\,^1\phi_{i_k,1},\,^1\phi_{i_k,2},\dots,\,^1\phi_{i_k,n}]^{\text{T}}$.

For the error propagation consideration, since the in-process features generated in one stage may be selected as the datums, their imperfections will bring errors to the features generated based on them even when there is no process faults present in later operations. Due to their pseudo-quantitative nature, PSV's are not capable to describe the propagated effects. However, a process fault affects only specific quality variables of the features generated in the same operation. For the quality variables of features generated in other operations, the

corresponding $^{0}\phi_{i_k,j}$'s equal to 0 and $^{1}\phi_{i_k,j}$'s equal to ∞. Also because the statistical analysis tool introduced in section 4.2 has the capability to identify this difference, it is rational to dismiss from PSV's the description of propagated effects.

4. STATISTICAL DATA ANALYSIS

To implement the root cause diagnosis, PSV's should be compared with the symptoms extracted from in-process measurements. In this paper, the linear cause-effect relationships between dimensional quality and process variables are adopted based on the assumption that process faults are often small in magnitudes. Thus, a general nonlinear relationship can be approximated through Taylor series expansion around its nominal value.

Let $\mathbf{y} = [y_1 \quad y_2 \quad \ldots \quad y_n]^T$ be an $n \times 1$ random vector formed by stacking up quality variables of the in-process features generated in different operations. Those variables are defined as the deviation from nominal value. Let $\mathbf{y}_i$ $(i=1,2,...,N)$ be a random sample of N observations of $\mathbf{y}$. The cause-effect relationship between potential process faults and the quality variables can be represented as

$$\mathbf{y} = \mathbf{\Gamma}\mathbf{u} + \mathbf{v} \tag{3}$$

Where $\mathbf{u} = [u_1 \quad u_2 \quad \ldots \quad u_p]^T$ is a $p \times 1$ random vector that represents the p ($p \leq n$) potential faults (i.e. causes). $\mathbf{\Gamma} = [\mathbf{\phi}_1 \quad \mathbf{\phi}_2 \quad \ldots \quad \mathbf{\phi}_p]$ is an n by p constant matrix determined by product design and process plan. $\mathbf{\phi}_j$ $(j = 1, 2, \ldots, n)$ is an $n \times 1$ vector. $\mathbf{v} = [v_1 \quad v_2 \quad \ldots \quad v_n]^T$ is an $n \times 1$ zero-mean random vector with covariance matrix $\mathbf{\Sigma}_{\mathbf{v}} = \sigma^2 \mathbf{I}_{n \times n}$, where $\mathbf{I}_{n \times n}$ denotes an n by n identity matrix.

The process faults defined in $\mathbf{u}$ often manifest themselves as mean shifts and/or increase of variances. As a result, u is split into two parts: a deterministic part $\overline{\mathbf{u}}$ and a zero-mean random part $\widetilde{\mathbf{u}}$. So, Eq. (3) can be written as:

$$\mathbf{y} = \mathbf{\Gamma}(\overline{\mathbf{u}} + \widetilde{\mathbf{u}}) + \mathbf{v} = \mathbf{\Gamma}\overline{\mathbf{u}} + \mathbf{\Gamma}\widetilde{\mathbf{u}} + \mathbf{v} \tag{4}$$

Where $\overline{\mathbf{u}} = \mathrm{E}(\mathbf{u})$ is a deterministic vector, $\widetilde{\mathbf{u}} = \mathbf{u} - \overline{\mathbf{u}}$ and $\mathbf{v}$ are independent zero-mean

random vectors. When process faults present, the measurement data will correspondingly demonstrate shift of mean and/or increase of variance. Let $\mathbf{y} = \overline{\mathbf{y}} + \widetilde{\mathbf{y}}$,

$$\overline{\mathbf{y}} = \mathbf{\Gamma}\overline{\mathbf{u}} \tag{5}$$

$$\widetilde{\mathbf{y}} = \mathbf{\Gamma}\widetilde{\mathbf{u}} + \mathbf{v} \tag{6}$$

The objectives of the statistical analysis include: (i) detecting the process changes from measurement data and (ii) extracting the spatial nature of the process faults, i.e. the fault directions $\mathbf{\phi}_k$'s, and (iii) matching them with PSV's to identify the potential process faults. According to Eq. (5) and (6), the process faults manifest as mean shifts have fixed effects on quality variables and those manifest as variance increase cause random effects. In this paper, the identification of the fault sources which cause random effects is addressed.

4.1 Hypothesis testing for fault detection

Assume that the measurement data $\mathbf{y}$ follows multivariate normal distribution. Suppose totally N random samples are measured and denoted by $\mathbf{Y} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \ldots \quad \mathbf{y}_N]^T$. From (3), under normal manufacturing condition, $\mathbf{\mu}_{\mathbf{Y}} = 0$ and $\mathbf{\Sigma}_{\mathbf{Y}} = diag(\sigma_1^2, \sigma_2^2, \ldots \sigma_n^2)$, where $\mathbf{\mu}_{\mathbf{Y}}$ and $\mathbf{\Sigma}_{\mathbf{Y}}$ represent the mean vector and covariance matrix of respectively. Apley and Shi (2001) discussed the applicability to assume that the covariance matrix of $\mathbf{Y}$ is a diagonal matrix with equal diagonal elements, i.e. $\mathbf{\Sigma}_{\mathbf{Y}} = \sigma^2 \mathbf{I}$. Process fault will result in a different variation pattern. Hence, statistical hypotheses testing on sample covariance can be conducted to detect the process changes.

$$H_0 : \mathbf{\Sigma}_{\mathbf{Y}} = \sigma^2 \mathbf{I} \; vs. \, H_a : \mathbf{\Sigma}_{\mathbf{Y}} \neq \sigma^2 \mathbf{I} \tag{7}$$

Where, σ^2 is an unknown positive value. The test is based on the unbiased likelihood ratio statistic

$$\lambda = |\mathbf{S}| / (n^{-1} tr(\mathbf{S}))^n \tag{8}$$

Where, $\mathbf{S} = (N-1)^{-1} \Sigma_{i=1}^N (\mathbf{y}_i - \overline{\mathbf{Y}})(\mathbf{y}_i - \overline{\mathbf{Y}})^T$. At the αth level of significance, it rejects H_0 in favor of H_a if the observed statistic

$$-[(N-1) - (2n^2 + n + 2)/6n] \log \lambda > \chi^2_{g,\alpha} \tag{9}$$

Where, $\chi^2_{g,\alpha}$ is an asymptotical chi-square distribution with $g = n(n+1)/2 - 1$ degrees of freedom. If H_0 is rejected at significance level α, it means that process faults, which manifest themselves as variance increase, are present.

4.2 Symptom extraction for fault identification

Assuming $E(\tilde{\mathbf{u}}) = \mathbf{0}$ and $Cov(\tilde{\mathbf{u}}) = \mathbf{I}$, Eq. (6) can be treated as an orthogonal factor model with p common factors (Richard and Dean, 1998). Therefore, factor analysis is used in this paper to identify the linear cause-effect relationship between quality variables and process faults. The p common factors are the process faults needed to be figured out. Under the assumption of linear relationship, the covariance between measurement data $\tilde{\mathbf{Y}}$ and the random process faults $\tilde{\mathbf{u}}$ is defined by $Cov(\tilde{\mathbf{Y}}, \tilde{\mathbf{u}}) = \mathbf{\Gamma}$. Principal component method is used to estimate the factor loading matrix $\mathbf{\Gamma}$. Let the covariance of $\mathbf{Y}$,

$$\mathbf{\Sigma_Y} = Cov(\mathbf{Y}) = \mathrm{E}(\mathbf{Y} - \mathbf{\mu_Y})(\mathbf{Y} - \mathbf{\mu_Y})^{\mathrm{T}} = \mathbf{\Gamma\Gamma}^{\mathrm{T}} + \mathbf{\Psi} \qquad (10)$$

have eigenvalue-eigenvector pairs (λ_i, e_i) with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ and $span\{e_i\}_{i=1}^p = span\{\mathbf{\phi}_i\}_{i=1}^p$, and assume that $E(\tilde{\mathbf{u}}) = 0$ and $Cov(\tilde{\mathbf{u}}) = \mathrm{I}$, $\mathbf{\Sigma_Y}$ can be decomposed as:

$$\mathbf{\Sigma_Y} = \mathbf{Z}_\mathrm{p}\mathbf{\Lambda_p}\mathbf{Z_p} + diag\{\psi_1, \psi_2, \ldots, \psi_n\} \qquad (11)$$

Where $\mathbf{Z}_p = [\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_p]$, $\mathbf{\Lambda}_p = diag\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$, $\psi_i = \sigma_{ii} - \sum_{j=1}^p \varphi_{ij}^2$ for $i=1,2,\ldots,n$. Matrix $\mathbf{\Gamma}$ can be estimated by $\mathbf{Z}_p\mathbf{\Lambda}_p^{-1/2}$. The number of process faults equals to the number of retained principle components p ($p \leq n$). Let $\mathbf{T}$ be a $p{\times}p$ orthogonal matrix, so that $\mathbf{TT}^{\mathrm{T}} = \mathbf{T}^{\mathrm{T}}\mathbf{T} = \mathbf{I}$. Eq. (6) can be written as

$$\tilde{\mathbf{y}} = \mathbf{\Gamma}\tilde{\mathbf{u}} + \mathbf{v} = \mathbf{\Gamma TT}^{\mathrm{T}}\tilde{\mathbf{u}} + \mathbf{v} = \mathbf{\Gamma}^*\tilde{\mathbf{u}}^* + \mathbf{v} \qquad (12)$$

$\tilde{\mathbf{u}}^* = \mathbf{T}^{\mathrm{T}}\tilde{\mathbf{u}}$ has the same statistical properties as $\tilde{\mathbf{u}}$. The loading $\mathbf{\Gamma}^* = \mathbf{\Gamma T} = \mathbf{Z}_p\mathbf{\Lambda}_p^{-1/2}\mathbf{T}$ can be distinguished from $\mathbf{\Gamma}$. This is very much akin to clarifying the cause-effect relations between a specific process fault and the quality variables. So in this research, the rotated factor loadings $\mathbf{\phi}_i^*$'s ($i = 1, 2, \ldots, p$) are "extracted

symptom vectors (ESV)", which can be compared with the PSV's to identify the faults. To implement it, 0-order ($^0\mathbf{\phi}_i^* = \begin{bmatrix} ^0\varphi_{i,1}^* & ^0\varphi_{i,2}^* & \ldots & ^0\varphi_{i,n}^* \end{bmatrix}^{\mathrm{T}}$) and 1-order ($^1\mathbf{\phi}_i^* = \begin{bmatrix} ^1\varphi_{i,1}^* & ^1\varphi_{i,2}^* & \ldots & ^1\varphi_{i,n}^* \end{bmatrix}^{\mathrm{T}}$) ESV's are formulated for each φ_i^* ($i = 1, 2, \ldots, p$). Let

$$^0\varphi^*_{i,j} = \begin{cases} 1, & \text{if } \varphi_{i,j} > 0 \text{ and } |\varphi_{i,j}| > h \\ 0, & \text{if } |\varphi_{i,j}| < h \\ -1, & \text{if } \varphi_{i,j} < 0 \text{ and } |\varphi_{i,j}| > h \end{cases} \qquad (13)$$

$$^1\varphi^*_{i,j} = \begin{cases} 1 & \text{if } ^0\varphi^*_{i,j} \neq 0,\ ^0\varphi^*_{i,k} \neq 0,\ \varphi^*_{i,j} > \varphi^*_{i,l}, \\ & \text{and } \varphi^*_{i,t} = 0\, (t = l+1, l+2, \ldots, j-1) \\ 0 & \text{if } ^0\varphi^*_{i,j} \neq 0,\ ^0\varphi^*_{i,k} \neq 0,\ \varphi^*_{i,j} = \varphi^*_{i,k}, \\ & \text{and } \varphi^*_{i,t} = 0\, (t = l+1, l+2, \ldots, j-1) \\ -1 & \text{if } ^0\varphi^*_{i,j} \neq 0,\ ^0\varphi^*_{i,k} \neq 0,\ \varphi^*_{i,j} < \varphi^*_{i,k}, \\ & \text{and } \varphi^*_{i,t} = 0\, (t = l+1, l+2, \ldots, j-1) \\ \infty & \text{if } ^0\varphi^*_{i,j} = 0 \text{ or } ^0\varphi^*_{i,v} = 0 \text{ for all } v < j \end{cases} \qquad (14)$$

Where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ and h is a positive threshold. As discussed in section 3, process faults that are present in different operations cause different effects. The factor analysis (Richard and Dean, 1998) and factor rotation techniques tend to find the loading structure that each factor (process fault) has high contribution to some variables and small contribution to the others. Therefore, factor analysis provides the capability to identify this difference and to distinguish process fault present in different operations.

The extracted symptom vectors for each individual operation $^0_k\mathbf{\phi}_i^*$ and $^1_k\mathbf{\phi}_i^*$ ($k = 1, 2, \ldots, M$) are derived from $^0\mathbf{\phi}_i^*$ and $^1\mathbf{\phi}_i^*$, as illustrated in figure 3. Each $^0_k\mathbf{\phi}_i^*$ or $^1_k\mathbf{\phi}_i^*$ contains n (total number of quality variables) elements. The elements corresponding to manufacturing features generated in operation k are copied from $^0\mathbf{\phi}_i^*$
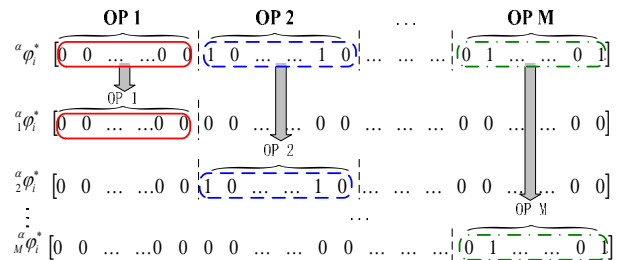

FIGURE 3. DECOMPOSITION OF ESV

and $^1\varphi_i^*$, respectively. Other elements, which represent the quality variables of manufacturing features generated in other operations, are set to zeros. As a result, every $^0\varphi_i^*$ (or $^1\varphi_i^*$) has M $^0_k\varphi_i^*$ (or $^1_k\varphi_i^*$), $k = 1,2,\ldots,M$. The smallest index k^* of the $^0_k\varphi_i^*$'s that contain non-zero elements indicates the stage where the fault introduced. In Figure 3, $k^* = 2$. And, $^0_{k^*}\varphi_i^*$'s and $^1_{k^*}\varphi_i^*$'s with non-zero elements will be compared with $^0_k f_{j_k}$'s and $^1_k f_{j_k}$'s ($j_k = 1,2,\ldots,m_k$), respectively. Let $^0_k\kappa_{i:j_k}$ and $^1_k\kappa_{i:j_k}$ be the 0-order and 1-order matching vectors (MV), respectively,

$$^0_k\kappa_{i:j_k} = {}^0_{k^*}\varphi_i^* - {}^0_k f_{j_k} \tag{15}$$

$$^1_k\kappa_{i:j_k} = {}^1_{k^*}\varphi_i^* - {}^1_k f_{j_k} \tag{16}$$

Also let 0-order matching index (MI) $^0_k\rho_{i:j_k}$ = number of non-zero elements in $^0_k\kappa_{i:j_k}/n$ and 1-order MI $^1_k\rho_{i:j_k}$ = number of non-zero elements in $^1_k\kappa_{i:j_k}/n$. These two indices evaluate the confidence of the matching conclusion. Therefore, fault identification is a procedure to find the j that maximize $^0_k\rho_{i:j_k}$ and $^1_k\rho_{i:j_k}$.

5. CASE STUDY

A case study was conducted by applying the proposed methodology in a real engine head machining process. The detailed process description and state space (SS) model introduced by Zhou, *et. al.* (2003) were used in this paper to derive cause-effect diagram and PSV's of potential faults, simulate process faults and generate corresponding measurement data. Totally 168 quality variables of 28 features generated in 3 different operations were considered in y. Two faults were purposely introduced to operation 3 by:

(i) Adding a variance source (fault 1) to the process variable corresponding to the X position of the 4-way pin, and
(ii) Adding a variance source (fault 2) to the process variable corresponding to the X position of the 2-way pin.

Their corresponding PSV's are listed in table 1 and table 2. 1,000 samples were simulated with the SS model and the corresponding measurements were generated.
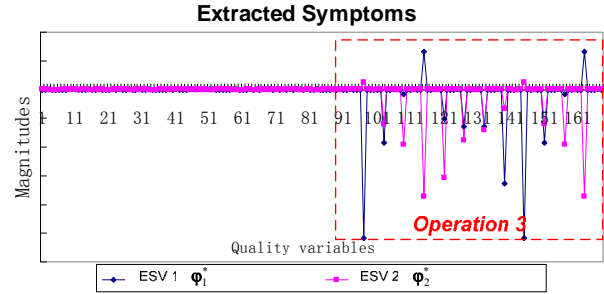

FIGURE 4. EXTRACTED SYMPTOMS

Hypotheses testing results indicated the process changes. For the purpose of root cause diagnosis, two ESV's were achieved by applying the FA on the simulated measurement data. These two ESV's, as shown in figure 4, were compared with all the PSV's of potential faults. The detailed 0-order and 1-order comparisons between the two ESV's and the PSV's of the two faults are listed in table 1 and table 2, respectively. Resultant MI's indicated that ESV 1 was identified as fault 2 and ESV 2 was identified as fault 1.

6. CONCLUSION

In this paper, a methodology for detecting and identifying process faults of multi-operational manufacturing system is proposed. It avoids time-consuming quantitative modeling and provides an effective representation of the engineering knowledge. The limitations of the current works include: (i) Different statistical analysis tools are required to extract the different variation patterns. (ii) Sensitivity analysis is another critical issue to be addressed, especially for the cases that the fault directions of two or more potential faults are not orthogonal to each other. (iii) A systematic mechanism to assess that identification result, especially for the non-orthogonal faults identification, is desirable. These limitations define the scope and directions of the future work.

TABLE 1. 0-ORDER MATCHING FOR FAULT IDENTIFICATION

	Identification of fault 1	Identification of fault 2
0-order PSV	$_3^0 f_1 = [0_{1\times48} \quad | 0_{1\times48} | \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_1$ $\Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1}$ $\Xi_{-1} \quad \Xi_1]^T$	$_3^0 f_2 = [0_{1\times48} \quad | 0_{1\times48} | \quad \Xi_1 \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1}$ $\Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_1 \quad \Xi_{-1}$ $\Xi_{-1} \quad \Xi_{-1}]^T$
0-order ESV's	$^0\varphi_2^* = {}_3^0\varphi_2^* = [0_{1\times48} \quad | 0_{1\times48} | \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1}$ $\Xi_1 \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1}$ $\Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_1]^T$	$^0\varphi_1^* = {}_3^0\varphi_1^* = [0_{1\times48} \quad | 0_{1\times48} | \quad \Xi_1 \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1}$ $\Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_{-1} \quad \Xi_1 \quad \Xi_{-1}$ $\Xi_{-1} \quad \Xi_{-1}]^T$
0-order MV	$_3^0\kappa_{2:1} = {}_3^0\varphi_2^* - {}_3^0 f_1 = 0_{168\times1}$	$_3^0\kappa_{1:2} = {}_3^0\varphi_2^* - {}_3^0 f_2 = 0_{168\times1}$
0-order MI	$_3^0\rho_{2:1} = 1$	$_3^0\rho_{1:2} = 1$

TABLE 2. 1-ORDER MATCHING FOR FAULT IDENTIFICATION

	Identification of fault 5_3	Identification of fault 4_3
1-order PSV	$_3^1 f_1 = [\infty_{1\times48} \quad | \infty_{1\times48} | \quad \infty_{1\times6} \quad \Omega_1 \quad \Omega_1 \quad \Omega_1$ $\Omega_{-1} \quad \Omega_{-1} \quad \Omega_{-1} \quad \Omega_{-1} \quad \Omega_{-1} \quad \Omega_1$ $\Omega_1 \quad \Omega_1]^T$	$_3^1 f_2 = [\infty_{1\times48} \quad | \infty_{1\times48} | \quad \infty_{1\times6} \quad \Omega_{-1} \quad \Omega_{-1} \quad \Omega_{-1}$ $\Omega_1 \quad \Omega_1 \quad \Omega_1 \quad \Omega_1 \quad \Omega_1 \quad \Omega_{-1}$ $\Omega_{-1} \quad \Omega_{-1}]^T$
1-order ESV	$_3^1\varphi_2^* = [\infty_{1\times48} \quad | \infty_{1\times48} | \quad \infty_{1\times6} \quad \Omega_1 \quad \Omega_1 \quad \Omega_1$ $\Omega_{-1} \quad \Omega_{-1} \quad \Omega_0 \quad \Omega_{-1} \quad \Omega_{-1} \quad \Omega_1$ $\Omega_1 \quad \Omega_1]^T$	$_3^1\varphi_1^* = [\infty_{1\times48} \quad | \infty_{1\times48} | \quad \infty_{1\times6} \quad \Omega_{-1} \quad \Omega_{-1} \quad \Omega_{-1}$ $\Omega_1 \quad \Omega_1 \quad \Omega_1 \quad \Omega_1 \quad \Omega_1 \quad \Omega_{-1}$ $\Omega_{-1} \quad \Omega_{-1}]^T$
1-order MV	$_3^1\kappa_{2:1} = {}_3^1\varphi_2^* - {}_3^1 f_1 = [0_{1\times132} \quad -1 \quad 0_{1\times35}]$	$_3^1\kappa_{1:2} = {}_3^1\varphi_1^* - {}_3^1 f_2 = 0_{168\times1}$
1-order MI	$_3^1\rho_{2:1} = 167/168 = 0.994$	$_3^1\rho_{1:2} = 1$

Where $\Xi_t = [t \ 0 \ 0 \ 0 \ 0 \ 0]$, $\Omega_t = [t \ \infty \ \infty \ \infty \ \infty \ \infty]$, $\infty_{1\times n} = [\infty \ \infty \ ... \ \infty]_{1\times n}$ and $0_{1\times n} = [0 \ 0 \ ... \ 0]_{1\times n}$

REFERENCES

Apley, D. and Shi, J., (2001), "A Factor-Analysis Method for Diagnosing Variability in Multivariate Manufacturing Processes", *Technometrics*. Vol. 43, No.1 pp84 – 95

Ceglarek, D. and Shi, J., (1995), "Dimensional Variation Reduction for Automotive Body Assembly", *J. of Mfg. Review*, Vol. 8, pp139-154.

Huang, Q., Zhou, N. and Shi, J., (2000), "Stream of Variation Modeling and Diagnosis of Multi-station Machining Processes", *Proc. 2000 ASME IMECE*, MED-Vol. 11, pp.81-88, November, Orlando, FL.

Jackson, J. E. (1980), "Principal Components and Factor Analysis: part I – Principal Components," *J. of Quality Tech.*, Vol. 12, pp201-213

Li, H. X., Zuo, M. J., (1999), "A hybrid approach for identification of root causes and reliability improvement of a die bonding process- a case study", *Reliability Eng. & Sys. Safety*, Vol. 64, pp43-48.

Liu, T. I. and Yuan, G., (1996), "PDES - an expert system for constructing the precedence diagram", *J. of the Franklin Inst.*, v 333B, n 6, p 975-990.

Prasad, S., (1991), "Improving manufacturing reliability in IC package assembly using the FMEA technique"**,** *IEEE Trans. on Components, Hybrids and Mfg. Tech.*, v 14, n 3, Sep, p 452-456.

Richard A. J., Dean W. W., (1998) "Applied Multivariate Statistical Analysis", *Prentice Hall.*

Woodall, W. H. and Montgomery, D. C., (1999), "Research Issues and Ideas in Statistical Process Control", *J. of Quality Tech.*, Vol.31, pp376-386.

Zhou, S., Huang Q. and Shi J., (2003), "State Space Modeling of Dimensional Variation Propagation in Multistage Machining Process Using Differential Motion Vector", *IEEE Trans. on Robotics and Automation*, Vol. 19, No. 2, pp296-309.