IIE Transactions (2014) **46**, 497–511

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2013.814928

Statistical modal analysis for variation characterization and application in manufacturing quality control

WENZHEN HUANG[1,*], JINYA LIU[1], VIJYA CHALIVENDRA[1], DAREK CEGLAREK[2], ZHENYU KONG[3] and YINGQING ZHOU[4]

[1]*Department of Mechanical Engineering, University of Massachusetts Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747, USA*
E-mail: whuang@umassd.edu
[2]*International Digital Laboratory, WMG, University of Warwick, Coventry, CV4 7AL UK*
[3]*School of Industrial & Management Engineering, Oklahoma State University, Stillwater, OK 74078, USA*
[4]*Dimensional Control Systems Inc. Troy, MI 48084, USA*

Received June 2012 and accepted May 2013

A Statistical Modal Analysis (SMA) methodology is developed for geometric variation characterization, modeling, and applications in manufacturing quality monitoring and control. The SMA decomposes a variation (spatial) signal into modes, revealing the *fingerprints* engraved on the feature in manufacturing with a few truncated modes. A discrete cosine transformation approach is adopted for mode decomposition. Statistical methods are used for model estimation, mode truncation, and determining sample strategy. The emphasis is on implementation and application aspects, including quality monitoring, diagnosis, and process capability study in manufacturing. Case studies are conducted to demonstrate application examples in modeling, diagnosis, and process capability analysis.

Keywords: Manufacturing, quality, variation reduction, GD&T tolerancing

1. Introduction

With intense global competition, quality assurance becomes a critical challenge for design and manufacturing in industry. In a mass production environment, mechanical parts cannot be manufactured to completely conform to design specifications. There is a vital need to control manufacturing variations to optimize manufacturability and ensure product functionality (Ceglarek and Shi, 1995; Voelcker, 1998; Ceglarek *et al.*, 2004; Huang, Lin, Kong, and Ceglarek, 2007).

A major concern in manufacturing is to understand and control the effects of variations, in particular, that being made on critical geometric features. Three aspects in this regard involve (i) variation characterization and modeling; e.g., measuring and summarizing variation information on a manufactured part with a simplified math model that facilitates quality improvement in both design and manufacturing; (ii) identifying and extracting the most significant variation patterns for quality monitoring and diagnosis; and (iii) evaluating or predicting the capability of a process to produce quality products. The variation is also controlled in tolerance design. Geometric Dimensioning and Tolerancing (GD&T) describes all of the ranges and geometric ways that feature variations of size, position, orientation, form, and runout are allowed. To ensure quality, the variations must be simulated, and the impacts to manufacturability and product performance should be predicted via tolerance analyses. A model that can effectively characterize/summarize the variation and then convey the information to design for tolerance simulation is highly desirable but unfortunately unavailable. Huang and Ceglarek (2002) presented a Discrete Cosine Transformation (DCT)-based characterization technique that only partially solved the concern mentioned in (i) above. More methodological, application, and implementation issues still remain open, such as the concerns in (ii) and (iii) and the links between Statistical Modal Analysis (SMA) and tolerance design.

A unified variation model that bridges manufacturing and design is necessary but challenging. Currently available models are either good at characterizing signal variations (e.g., signal processing techniques for monitoring and diagnosis) or specialized in representing/manipulating curve/surface features (e.g., CAD parametric models). Neither serves both in a coherent manner.

1. The signal processing techniques in monitoring have accumulated and documented major amounts of information (process *fingerprints*) for characterization,

*Corresponding author
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uiie.

statistical modeling, diagnosis, and quality control (e.g., Statistical Process Control (SPC) and process capability). They are powerful in analyzing random signals with rich frequency components from measurements. Nevertheless, they have rarely been linked to or served as a GD&T model for replicating variations in tolerance design.

2. The modes identified by signal processing techniques have not been linked and interpreted in terms of GD&T tolerances, such as position, orientation, flatness and waviness, etc. This link is critical for root cause recognition in process control and in tolerance design simulation. For example, the first three rigid modes characterize kinematic variation of planar features; thus, by assigning these modes in a tolerance simulation model the size/position and flatness tolerance design can be evaluated through simulation.
3. To incorporate manufacturing uncertainty in GD&T design, the variations have been simulated by subjective statistical models (e.g., random CAD parametric models). These models have rarely been linked to actual variations in manufacturing because of their poor property in random signal processing and statistical modeling. The CAD parametric models, such as Bezier surface/curves, B-spline, NURBS, 16-point bicubic surface, etc., were initially designed for part curve/surface feature manipulation rather than the variations characterization. It is difficult to statistically model the random variations from measurement data. For example, a large number of parameters in NURBS (rational polynomials) on multiple surface patches must be estimated by nonlinear regression.

Huang and Ceglarek (2002) proposed the SMA concept and DCT algorithm for mode decomposition. However, it is an incomplete version of SMA methodology. It lacks some key components in theory and application, including statistical modeling, mode truncation criteria, sampling, the link between SMA and GD&T models as well as its applications in manufacturing. To address this problem, this article aims at comprehensively extending the work reported in Huang and Ceglarek (2002) by establishing (i) a complete SMA methodology with all of the above-mentioned components that were not covered or fully elaborated in the initial work and (ii) SMA application techniques in manufacturing; i.e., pattern extraction, monitoring. Statistical Process Control (SPC) and process capability with GD&T requirements. To keep a complete and smooth description of SMA, a certain level of overlap with the initial work remains, but it has been cut short as much as possible. Although the focus is on the manufacturing side, here the authors scratch only the surface of SMA for tolerance design; for more details please refer to Huang *et al.* (2010). In design, the SMA modes can be explained by GD&T tolerances such as size (e.g., first mode), orientation (e.g., second and third modes), and form (modes >3 in a certain frequency window). The statistical GD&T design calls for replicating the variations in the same way as they are being produced in the manufacturing process. Thus, it may provide a two-way traffic for design and manufacturing; i.e., utilizing process knowledge to compose the relevant modes (the *fingerprints*) into a SMA model for tolerance design simulation.

This article is organized as follows: a literature review is presented in Section 2, followed by assumptions and basic ideas of SMA in Section 3; Section 4 deals with a series of key issues in modeling such as statistical (spatial) modal analysis, mode truncation, and sampling; Section 5 emphasizes the strategy and implementation procedures for manufacturing application; i.e., identifying key variation patterns for diagnosis, process monitoring by mode-based control chart, and analyzing process capability with GD&T specification; case studies and a summary are included in Sections 6 and 7.

2. Related work

In the last several decades, tremendous efforts have been devoted to establishing variation and tolerance models for tolerance design and quality control in manufacturing.

Requicha (1983) pioneered the work on geometric tolerance representation theory. The concepts of Tolerance Zone (TZ) and variational class for representation of geometric tolerances were defined and studied (Requicha, 1983; Srinivasan and Jayaraman, 1989; Requicha and Chan, 1986). The TZ was defined as a region bounded by two virtual features, which are offsets from a design nominal feature. All actual features within the TZ belong to an acceptable variational class. Several techniques have been developed for computing offset feature surfaces in CAD models. Parametric semantics can also be directly extended to approximate TZ semantics to any accuracy by increasing the number of parameters (e.g., polynomial, Bezier, and bicubic surface patches). Turner and Wozny (1987) proposed a polynomial function to model form tolerance for nominal planar features. Alternative approaches based on Bezier's triangle fitting and triangle patches have also been proposed to represent planar surface (Gupta and Turner, 1993). Roy and Li (1998, 1999) (see also Li and Roy (2001)) developed a 16-point bicubic surface interpolation method. Most of these efforts focused on deterministic TZ construction in CAD systems rather than the random variability *per se*.

Random polynomial representation models have been developed in the last two decades. Roy and Li (1998, 1999) ensured that all of the control points on the interpolation surfaces were randomly selected in the TZ. In general, a combination of more than one surface patch is needed to represent the variational class of a complex feature with selected knots connecting them. Merkley (1998) also adopted random parameter Bezier functions for simulation of variations in assembly analysis. The random control points of the Bezier patch were used to simulate the patterns of

variations. They can be easily embedded into existing CAD systems. The parametric models used for statistical tolerance analysis rely on subjective assumptions on the distributions of model parameters. The estimation of these parameters and their distribution from measurement data is a cumbersome task because of high computation costs (e.g., in multi-dimensional spline surface fitting), nonlinear regression (e.g., NURBS), and problems associated with determiming the number and placement of knots, appropriate continuity constraints, model orders, etc.

In rigid assembly analysis, point-based variation models prevail (Ceglarek *et al.*, 2004; Huang, Lin, Kong, and Ceglarek, 2007). The covariance matrix method provides an alternative approach for variation analysis in compliant sheet metal assembly (Merkley, 1998; Camelio *et al.*, 2004). The covariance matrix, under a normality assumption, represents a point-based multivariate distribution (discretized geometric variations). Principal Component Analysis (PCA), a data-driven technique, can be used to eliminate correlations in the point-based data and thus provide a variation characterization model. The result of PCA is data or case dependent; thus, it may lead to inconsistent models from samples. Despite its wide applications in manufacturing for variation analysis and quality monitoring, the PCA model is not suitable for tolerance applications that require case-independent simulation models in their design. Therefore, there is no available PCA technique in the literature suitable for application to GD&T tolerance design.

Signal processing techniques have been used for many years in surface characterization stuies. Many orthogonal transformation techniques have been developed in image signal processing, such as the DCT (Ahmed *et al.*, 1974) and Karhunen–Loéve Transform (KLT; Srinivasan and Wood (1997)), etc. Huang *et al.* (2002) initiated DCT- based mode decomposition for stamping part variation analysis.

Data-driven methods, such as PCA and KLT, are sample specific and will produce inconsistent models. Srinivasan and Wood (1997) applied wavelets for feature extraction in manufacturing. The wavelet approach works excellently in characterizing highly localized and nonstationary signal patterns, but it is difficult to explain the obtained results, especially in terms of GD&T tolerances. These techniques have been applied almost exclusively in signal characterization for quality monitoring and diagnosis. There are only a few reports on GD&T design-related applications.

In summary, the above research efforts have been largely devoted in two parallel directions. On one side, a variety of models have been developed for tolerance design. On the other, two-dimensional (2D) data-driven signal processing and spatial statistics techniques have been developed for spatial feature characterization and monitoring in manufacturing. The urgent desire from industry for a unified methodology to bridge the gap motivated the authors' research reported in this article.

3. Problem formulation

By projecting a surface to a plane (x, y), geometric variation on the surface feature is defined as the deviation of an actual feature from a nominal feature; i.e., $f(x, y) = F_{\text{actual}} - F_{\text{nominal}}$, where F_{actual} and F_{nominal} are the actual and nominal features, respectively. Two assumptions are introduced below.

Smoothness assumption: $f(x, y)$ is sufficiently smooth such that the high spatial frequency components (e.g., surface roughness) are small and can be ignored; i.e., $f(x, y)$ is band limited.

Height field assumption: $f(x, y)$ can be expressed as a height field (a single-valued function): $z = f(x, y)$ defined on a 2D domain and it is a stationary Gaussian random field process.

An actual feature variation can be measured by a Coordinate Measurement Machine (CMM); a similar case can be found in micro/nanoscale topography with scanning by atomic force microscopy. The probe of a CMM machine works as a low-pass filter that cuts off higher frequency components by bridging small grooves. Compared with global variation patterns, the microscale errors, such as roughness and micro-texture, can be ignored in many engineering applications. CMM results provide a depth map that consists of a bi-dimensional array of height z measurements at various discrete (x, y) locations. The normality assumption is justified by checking the CMM measurement data on sheet metal parts (Ceglarek and Shi, 1995).

A mode-based expression for geometric variations is proposed in this article. The idea is to decompose a variation signal into a linear combination of series of modes. The orthogonality of the modal shape functions ensures the mode independence. The approximation is achieved in a truncated modal space. An SMA model can approximate a variation to any accuracy by increasing the number of modes included in the model. The statistical property of the variation is captured by the distributions of the modal parameters. A correlation reduction technique, such as orthogonal transforms commonly used in digital image processing, can be applied for modal decomposition. Given measurement data $f(n\Delta x, m\Delta y) = f(n, m)$ at $x = n\Delta x$, $y = m\Delta y$, $n, m = 1, \ldots, N$, the variation is transformed from the spatial into the frequency (mode) domain. For a "smooth" variation a large proportion of "signal energy" is concentrated in a few modes (variation patterns). In general, for 2D signal arrays with a sample size of N^2 (or $M \times N$, if $M \neq N$), the forward and inverse transforms are given as

$$T(u, v) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} f(n, m)\, g(n, m, u, v) \qquad \text{(forward transform)} \quad (1)$$

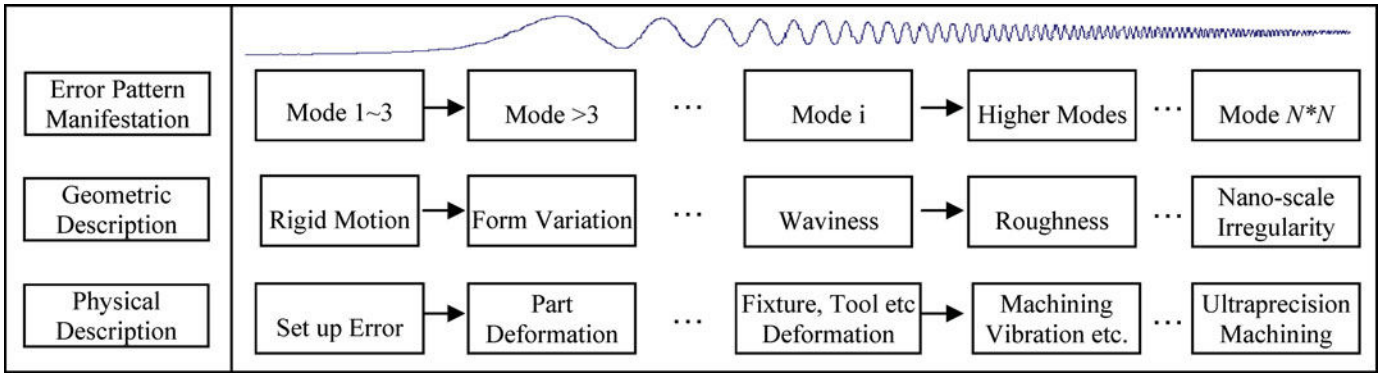

Fig. 1. Variation mode decomposition and interpretation.

$$f(n,m)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}T(u,v)\,h(n,m,u,v) \quad \text{(inverse transform)}, \tag{2}$$

where $T(u, v)$ represents the contribution of the mode indexed by mode numbers u, v in the x, y axes, respectively. $g(.)$ and $h(.)$ are called the forward and inverse transformation kernels (see also Section 4 for details). The transformation in Equation (2) is interpreted as a superposition of modes.

SMA modes (Equation (2)) reveal various variation patterns (Fig. 1) on an actual feature. For example, a kinetics mode (modes 1 to 3 or $h(n, m, 0, 0)$, $h(n, m, 1, 0)$, and $h(n, m, 0, 1)$) may represent fixture positioning/setup error (see Case I in Section 6); form or deformation modes ($u + v > 1$) represent part distortion or non-kinetics variation produced in manufacturing as summarized in Fig. 1 (also shown in Section 6). An SMA model carries rich information on variation patterns for manufacturing monitoring and diagnosis, enhancing signal interpretability, and quality control. It also provides a unified parametric framework for GD&T tolerance modeling in design. In comparison, existing tolerance models are largely parametric and thus lack the capability/flexibility of characterizing rich variation patterns.

4. SMA

The SMA methodology is developed in the following sections. In Section 4.1, the DCT method is initially introduced for variation modal analysis. Section 4.2 is devoted to statistical modeling. An inference method is applied for mode significance analysis in Section 4.3. Section 4.4 develops mode truncation criteria whereby the most significant modes are preserved in an SMA model.

4.1. *Modal decomposition via DCT transformation*

There are a number of discrete transformations that can be used for SMA modeling, such as Discrete Fourier Transform (DFT), discrete sine transform, DCT, slant transform, KLT, Wavelet Transform (WT), etc. DCT is selected here based on four criteria: (i) energy compaction (EC); (ii) analytical representation; (iii) frequency resolution; and (iv) interpretability.

1. EC: DCT has almost the highest EC; thus, it ensures the most compact model. DCT outperforms DFT by avoiding errors out of discontinuity from finite sampling. The DFT block or segment-based transforms leads to discontinuities at block boundaries (Gibbs phenomenon).
2. Analytical representation: Unlike the sample-dependent KLT approach, the cosine function-based approach provides a consistent analytical representation that is desirable in tolerance modeling and simulation.
3. High frequency resolution*:* Compared with WT, the superior frequency resolution of DCT is desirable in variation pattern identification, separation, representation, and interpretation.
4. Mode interpretation*:* Spatial frequency and mode are more interpretable than the results of WT in engineering (see Figs. 1 and 2 and Section 6).

The DCT transformation pair can be written as

$$\tilde{C}(u,v)=\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}f(n,m)g(n,m,u,v),$$
$$f(n,m)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\tilde{C}(u,v)g(n,m,u,v), \tag{3}$$

where $\{\tilde{C}(u, v)\}$ are the DCT coefficients, representing a *variation spectrum* on a feature (Fig. 2). The modes can be interpreted in terms of GD&T tolerances as illustrated in Figs. 1 and 2. The superposition of the first three modes in Equation (3) approximates the *rigid modes*. For example,

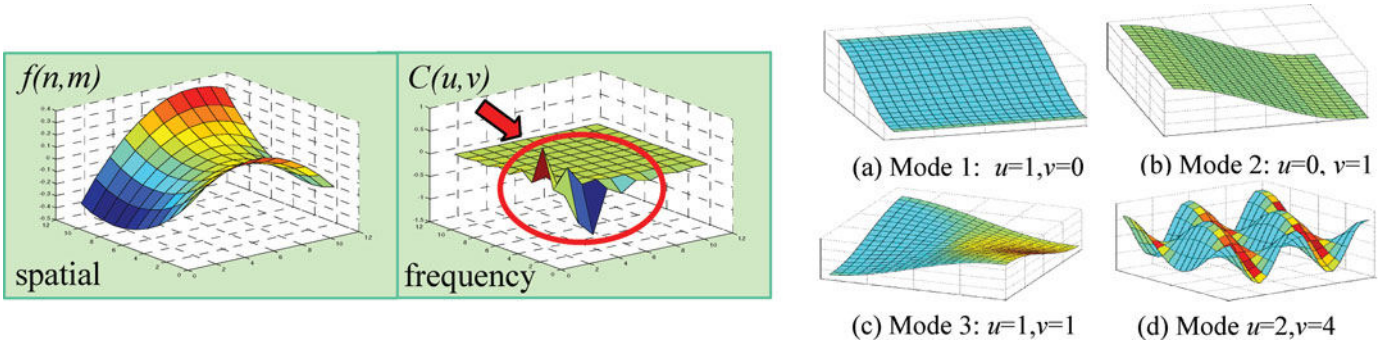

Fig. 2. The spatial frequency representation of geometric variation and its four modes.

when $u = v = 0$,

$$f(n, m)|_{1\text{st-mode}} = \frac{2}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} f(k, l),$$

represents a horizontal plane (variation controlled by size or location tolerances). When $u = 0$, $v = 1$, or $v = 0$, $u = 1$, the second and third modes ($C(1, 0)$ and $C(0,1)$ in Figs. 2(a) and 2(b)) are inclined quasi-rigid modes subject to orientation tolerance (e.g., flatness, straightness, etc.). Some higher modes are shown in Figs. 2(c) and 2(d). The combination of higher modes may represent form variations and even roughness according to their wavelengths.

4.2. *DCT and cosine-based regression*

The equivalency of DCT and cosine-based least squares regression is now demonstrated. Express Equation (3) as a discrete cosine-based regression model:

$$f(n, m) = \sum_{(u,v)\in\Omega} C(u, v) \cos(2un\pi/(2N - 1)) \times \cos(2vm\pi/(2N - 1)) + \varepsilon_t = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad (n, m = 1, 2, \ldots, N), \tag{4}$$

where $C(u, v) = \tilde{C}(u, v)\alpha(u)\alpha(v)$, $\boldsymbol{\beta} = \{C(u, v)\}^{\mathrm{T}}$, $\mathbf{X}$ is composed of mode shape vectors, Ω is a subset of $N \times N$ modes, and $\boldsymbol{\varepsilon}_t$ stands for a random error vector. Equation (4) includes the principal modes defined in Ω (see Sections 4.3 and 4.4) and the residual ε_t. ε_t is a linear combination of all minor mode components and is assumed to be normal. The least squares estimation for $f(n, m)$ is $\hat{\mathrm{f}} = \mathbf{Xb}$, $\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{f}$.

Theorem 1 shows that the DCT transform is equivalent to the least squares regression with 2D discrete cosine bases.

Theorem 1. *The least squares estimation of* $\boldsymbol{\beta}$*:* $\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{f}$ *is equivalent to modal parameter vector* $\{C(u, v)\}$*, with* $\mathbf{b} = \{C(u, v)\}$*, where* $\boldsymbol{\beta}$ *and* $\{C(u, v)\}$ *are defined in Equations* (3) *and* (4)*, respectively.*

Proof. The orthogonal cosine base yields

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = [\mathbf{I}] \quad \text{or} \quad \mathbf{b} = \mathbf{X}^{\mathrm{T}}\mathbf{f} \tag{5}$$

The least squares estimation of the error with DCT transformations presented in Equations (4) and (5) yields

$$\begin{aligned}\mathbf{b} = \mathbf{X}^{\mathrm{T}}\mathbf{f} &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} f(n, m)\alpha(u)\alpha(v) \cos(2nu\pi/(2N - 1)) \\ &\quad \times \cos(2mv\pi/(2N - 1)) \\ &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} f(n, m)g(n, m, u, v) = \{C(u, v)\}\end{aligned}$$

This completes the proof. ■

From Theorem 1 any results for $\mathbf{b}$ holds for $\{C(u, v)\}$; thus, in Sections 4.3 and 4.5 only vector $\mathbf{b}$ is used to describe DCT coefficients $\{C(u, v)\}$.

The DCT algorithm requires evenly spaced samples; however, unevenly scattered sampling can be more popular. For unevenly scattered sample points, the regression algorithm with the 2D cosine base function, similar to Vaníček's method in least squares spectral analysis, gives an alternative for DCT coefficients estimation (Wells *et al.*, 1985).

4.3. *Statistical significance test for SMA parameters*

Sections 4.1, 4.2 were devoted to point estimation of modal parameters. However, it is more desirable to have their statistical model for tolerance design application. In this section, the distribution is initially estimated and then the mode significance test is introduced for truncation criteria creation in Section 4.4.

Assuming $\boldsymbol{\varepsilon}_t(n, m) \sim N(0, \sigma^2\mathbf{I})$, it can be shown that $\mathbf{b}$ is an unbiased estimation of $\boldsymbol{\beta}$: $E(\mathbf{b}) = \boldsymbol{\beta}$, and $\mathrm{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \sigma^2[\mathbf{I}]$. $\mathbf{b}$ is a linear combination of $\mathbf{f}$ that is normally distributed; thus, normality of $\mathbf{b}$ is ensured. It is also easy to show for a portion of $\mathbf{b}_1$ out of $\mathbf{b}$ that $E(\mathbf{b}_1) = \boldsymbol{\beta}_1$.

Theorem 2. *Let* $\mathbf{f} = \mathbf{Xb}_1 + \boldsymbol{\varepsilon}_t$*,* $\mathbf{X}$ *be defined as in Equation* (7)*,* $\mathbf{b}_1 = \{b_1, b_2, \ldots b_r\}^{\mathrm{T}}$ *and* $\boldsymbol{\varepsilon}_t \sim N_N(0, \sigma^2\mathbf{I})$*. Let* $\bar{\mathbf{b}}_1$ *be the sample mean of vector* $\boldsymbol{\beta}_1$*. Then the* $100(1 - \alpha)\%$ *confidence*

region for $\boldsymbol{\beta}_1$ *is given by*

$$(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1)^{\mathrm{T}}(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1) \le r\, \mathrm{s}^2\, F_{r,N-r}(\alpha), \tag{6}$$

where $F_{r,N-r}(\alpha)$ *is the upper* (100%) *percentile of an F-distribution with* r *and* $N - r$ *degrees of freedom.* $s^2 = \bar{\boldsymbol{\varepsilon}}_t^{\mathrm{T}}\bar{\boldsymbol{\varepsilon}}_t/(N-r)$ *is the sample variance estimation of residual mean, and* $E(\mathrm{s}^2) = (\sigma/M)^2$. *The simultaneous* $100(1-\alpha)\%$ *confidence intervals for* $\beta_{1\mathrm{i}}$ *are given by* $\bar{b}_{1i} \pm s\sqrt{r\,F_{r,N-r}(\alpha)}$.

Proof. Sampling distribution of $\mathbf{b}_1 = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{f}$ is $N_r(\boldsymbol{\beta}_1, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$, so the distribution of $\bar{\mathbf{b}}_1$ with M samples is $N_r(\boldsymbol{\beta}_1, (\sigma/M)^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$. From the orthogonality of discrete cosine basis, $\mathbf{X}$ is a fixed r rank matrix. $\mathbf{X}(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1)$ is a linear combination of $(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1) \sim N_r(\mathbf{0},\ (\sigma^2/M^2)\mathbf{X}^{\mathrm{T}}\mathbf{X})$, so $(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1) \sim \chi_r^2$. Since $\mathbf{X}$ is a fixed matrix from discrete base function, the sample mean estimation of response is $\bar{\mathbf{f}}(n,m) = \mathbf{X}\boldsymbol{\beta}_1 + \bar{\boldsymbol{\varepsilon}}$, $\hat{\bar{\boldsymbol{\varepsilon}}}_t = [\mathbf{I} - \mathbf{X}\mathbf{X}^{\mathrm{T}}]\bar{\boldsymbol{\varepsilon}}_t$, where $\bar{\boldsymbol{\varepsilon}}_t \sim N(0, (\sigma/M)^2\mathbf{I})$. The covariance matrix becomes

$$\mathrm{Cov}\left(\begin{bmatrix}\bar{\mathbf{b}}_1\\ \hat{\bar{\boldsymbol{\varepsilon}}}_t\end{bmatrix}\right) = \frac{\sigma^2}{M^2}\begin{bmatrix}\mathbf{I} & | & 0\\ - - - & | & - - -\\ 0 & | & \mathbf{I} - \mathbf{X}\mathbf{X}^{\mathrm{T}}\end{bmatrix}.$$

Thus, $\hat{\bar{\boldsymbol{\varepsilon}}}_t = [\mathbf{I} - \mathbf{X}\mathbf{X}^{\mathrm{T}}]\bar{\boldsymbol{\varepsilon}}_t$ and $\bar{\mathbf{b}}_1$ are independent. The distribution of $\mathrm{s}^2 \sim \chi^2_{N-r}$ (Johnson and Wichern, 1998), so the $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}_1$ is $(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1)^{\mathrm{T}}(\boldsymbol{\beta}_1 - \bar{\mathbf{b}}_1) \le r\, \mathrm{s}^2\, F_{r,N-r}(\alpha)$. ■

Since the $\hat{\mathrm{Var}}(\bar{\mathbf{b}}_1) = (s/M)^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ and from the result of Equation (6) the simultaneous $100(1-\alpha)\%$ confidence intervals for the β_{1i} is $\bar{b}_{1i} \pm s\sqrt{r\,F_{r,N-r}(\alpha)}$, $i = 1, 2, \ldots, r$.

Replacing $F_{r,N-r}$ by the t-distribution, the one-at-a-time confidence interval for searching significant mode in part error signals is $\bar{b}_{1i} \pm t_{N-r}(\frac{\alpha}{2})s$, $i = 1, 2, \ldots, r$.

Only the coefficients with such confidence intervals that do not include zero are taken as statistically significant. However, a statistically significant mode can be meaningless if it has only minor engineering significance. More relevant significance indices are defined in Section 4.4 to account for both statistical and engineering significances, thereby allowing mode truncation criteria to be developed. The truncated model with all significant modes in Ω can be written as

$$f(x, y) = \sum_{u,v\in\Omega} \bar{b}_{u,v}\cos(ux\pi)\cos(vy\pi) = \mathbf{X}\bar{\mathbf{b}}_1. \tag{7}$$

As discussed in Section 4.1, it can be approximated by simplifying the quasi-rigid modes with linear terms:

$$f(x,\ y) = b_1 + b_2 x + b_3 y + \sum_{i=4}^{p} b_i\varphi_i(x, y) \tag{8}$$

4.4. *Mode significance and truncation criteria*

Three criteria are now proposed for significance checking. The importance of a mode is measured by significance indices defined in Section 4.4.1, including (i) statistical significance; (ii) energy significance; and (iii) Hausdorff distance between measured error surface and reconstructed surface from the model. Criteria (ii) and (iii) define the engineering significance of a mode. The mode truncation criteria are developed by using these indices in Section 4.4.2.

4.4.1. *Mode significance criteria*

Denote Ω as a significant mode index set with a specific subscript, and the following three criteria are defined for mode significance check.

1. Statistical Significance Criterion (SSC): Given confidence level α, denote the confidence interval in Section 4.3 as $I_{(u,v)}$, then the statistically significant mode set is defined as: $\Omega_{\mathrm{s}} = \cup\{(u, v)|\ 0 \notin I_{(u,v)}\}$.
2. Energy Significance Criterion (ESC): The contribution of each mode (u, v) can be characterized by the magnitude of the modal parameter $|C(u, v)|$. The signal energy in the mode is expressed as $C^2(u, v)$. Using Parseval's Theorem, it yields:

$$\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} C^2(u, v) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} f^2(n, m). \tag{9}$$

The ratio of the energy in a selected significant mode set to the total signal energy can be used to define the EC of the model. To achieve a given EC of $0 \le E \le 100\%$, the minimum number of significant modes that should be included in a coefficient index set Ω_i of the ith part is

$$\frac{\sum_{\Omega_i} C^2(u, v)}{\sum_{n,m=0}^{N-1} f^2(n, m)} \ge E \qquad (u, v) \in \Omega_i. \tag{10}$$

The above truncation criterion is based on coefficients of an individual part. With M samples, Ω_e is defined as the index set that satisfies $\Omega_e = \cup\Omega_i$, $i \le M$. Ω_e includes all of the mode indices that are energy significant in these M samples.

3. Hausdorff Distance Criterion (HDC): The Hausdorff distance is a generic distance between two non-empty sets. Let $\mathrm{d}(p, S')$ be the Euclidean distance between a point $p \in$ surface S and a surface S', or $\mathrm{d}(p, S') = \min_{p\in S}||\ p - p'\ ||_2$. Hausdorff distance is defined as $d(S, S') = \max_{p'\in S'}\{d(p, S')\}$. Given a threshold δ the Hausdorff distance criterion is defined as

$$\begin{aligned} d(\tilde{f}(n,m),\ f(n,m)) &= \max_{(u,v)\in\Omega_j}^{p\in\tilde{f}(n,m)}\{d(p,\ f(n,m))\} \le \delta\\ \tilde{f}(n,m) &= \sum_{\Omega_j} C(u, v)\cos(2un\pi/(2N-1))\\ &\quad \times \cos(2vm\pi/(2N-1)),\\ f(n,m) &= \tilde{f}(n,m) + \varepsilon_t(n,m). \end{aligned} \tag{11}$$

The Hausdorff distance can be simplified as a point-to-point distance between sampled error surface and reconstructed surface. It provides an intuitive geometric

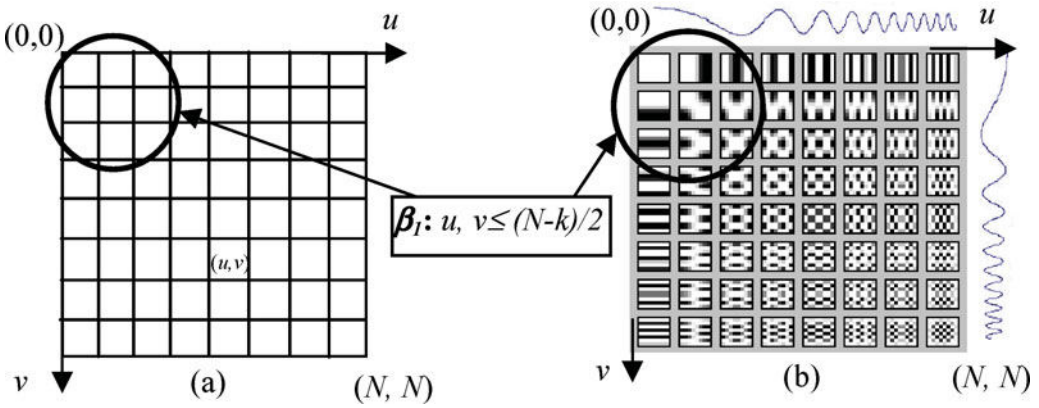

Fig. 3. (a) Frequency domain and (b) mode shapes ($N = 8$, $k = 2$).

measure of the maximum distance between the truncated model and the measured model. Analogous to the definition of Ω_e for M part samples, Ω_h is defined as the index set that satisfies $\Omega_h = \cup\Omega_j$, $j \leq M$. Ω_h includes all of the mode indices that satisfy the HDC.

4.4.2. *Mode truncation criteria*

The modes to be preserved must be statistically significant and meaningful in terms of engineering. For example, the modes for texture irregularity (e.g., roughness) can be statistically significant, but they usually only have an insignificant effect on assembly stackup analysis. The presented significance criteria are combined to define the rules for mode truncation. The union and intersection of the sets of indices are introduced to exclude insignicant modes and three truncation criteria are thus defined.

1. Mode Selection Criterion 1: Given confidence level α and an EC E the truncated SMA model must include all modes $(u, v) \in \Omega$: $\Omega = \Omega_e \cap \Omega_s$. Ω represents a set that includes those modes that are significant both statistically and in EC.
2. Mode Selection Criterion 2: Given confidence level α and a maximum Hausdorff distance δ, the truncated SMA model must include all modes $(u, v) \in \Omega$: $\Omega = \Omega_h \cap \Omega_s$. Ω represents all the modes satisfying the HDC and are statistically significant.
3. Mode Selection Criterion 3: Given confidence level α, a maximum Hausdorff distance δ, and an EC E, the truncated SMA model must include all modes $(u, v) \in \Omega$: $\Omega = \Omega_s \cap (\Omega_e \cup \Omega_h)$.

The truncated model with the coefficient index set Ω is

$$\tilde{f}(n, m) = \sum_{\Omega} C(u, v) \cos(2un\pi/(2N-1)) \times \cos(2vm\pi/(2N-1)).$$

4.5. *Sampling and aliasing*

An implicit assumption in the SMA model is that a feature is sampled with sampling rate that ensures all the variation patterns of interest are included. However, a CMM cannot automatically ensure a sufficient sampling rate. A validity check method is now introduced for a predefined sampling rate (Fig. 3).

Before a part is sampled an appropriate sampling rate has to be decided to avoid aliasing error. Sampling theorem ensures that aliasing errors can be avoided if the sampling points are taken with periods of $1/\Delta x$ and $1/\Delta y$ greater than or equal to the Nyquist frequencies 2F and 2G, respectively. Otherwise, the aliasing error has to be taken into account.

Although the assumption was made in Section 3 that the form error $f(x, y)$ is band limited, the cut-off frequencies F and G are actually unknown. The cut-off frequencies can be estimated using process knowledge. For example, tool and tool path parameters can be used to estimate periodic variations on a feature in machining processes. A sampling rate can be initially determined from experience, and then the validity of the rate must be checked. A likelihood ratio test is introduced below for validity checking the predefined sampling rate. It is assumed that the modal magnitudes are monotonically decaying with increasing frequencies u and v.

Denote $\boldsymbol{\beta}_s$ as a subset of coefficients $\boldsymbol{\beta}$ of a complete regression model. $\boldsymbol{\beta}_s$ is defined as $\{(u, v) \mid u \leq N - k, v \leq N - k\}$, $k \geq 1$. The integer $k \geq 1$ is selected to provide the degrees of freedom for s^2 estimation. $\boldsymbol{\beta}_s$ is split into two groups: $\boldsymbol{\beta}_1$ includes all of the modes with index $\{(u, v) \mid u \leq (N-k)/2, v \leq (N-k)/2\}$, and $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_s - \boldsymbol{\beta}_1$ includes all of the modes with index $\{(u, v) \mid (N-k)/2 < u \leq N - k, (N-k)/2 < v \leq N - k\}$. The statement that $\boldsymbol{\beta}_2$ does not affect $\bar{\mathbf{f}}(u, v)$ can be expressed as H_0 hypothesis:

$$H_0\text{: } \boldsymbol{\beta}_2 = 0; \quad \text{i.e., } \bar{\mathbf{f}} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}; \quad \text{and}$$
$$H_1\text{: } \boldsymbol{\beta}_2 \neq 0; \quad \text{i.e., } \bar{\mathbf{f}} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

Under hypothesis H_0, $\mathbf{X}$ and $\boldsymbol{\beta}_s$ are

$$\mathbf{X} = \left[[\mathbf{X}_1]_{N\times\frac{(N-k)^2}{4}}, \quad [\mathbf{X}_2]_{N\times\frac{3(N-k)^2}{4}} \right],$$

$$\boldsymbol{\beta}_\mathbf{s} = \begin{bmatrix} [\boldsymbol{\beta}_1]_{\frac{(N-k)^2}{4}\times 1} \\ [\boldsymbol{\beta}_2]_{\frac{3(N-k)^2}{4}\times 1} \end{bmatrix}$$

The likelihood ratio test rejects H_0 if

$$\frac{4\times(SS_{\text{res}}(\mathbf{X}_1) - SS_{\text{res}}(\mathbf{X}))}{3(N-k)^2 s^2} > F_{3(N-k)^2/4,\ k(2N-k)}(\alpha) \quad (12)$$

Here $s^2 = (\bar{\mathbf{f}} - \mathbf{X}\bar{\mathbf{b}}_s)^{\mathrm{T}}(\bar{\mathbf{f}} - \mathbf{X}\bar{\mathbf{b}}_s)/k(2N-k))$, $SS_{\text{res}}(\mathbf{X}) = (\bar{\mathbf{f}} - \mathbf{X}\bar{\mathbf{b}}_s)^{\mathrm{T}}(\bar{\mathbf{f}} - \mathbf{X}\bar{\mathbf{b}}_s)$ and $SS_{\text{res}}(\mathbf{X}_1) = (\bar{\mathbf{f}} - \mathbf{X}_1\bar{\mathbf{b}}_1)^{\mathrm{T}}(\bar{\mathbf{f}} - \mathbf{X}_1\bar{\mathbf{b}}_1)$. If the H_0 cannot be rejected, the contribution of modes with $\boldsymbol{\beta}_2$ as coefficients can be taken as being insignificant; thus, conservative upper bound estimates of the cut-off frequencies are $F \le (N-k)/2$ and $G \le (N-k)/2$. The sampling frequencies satisfy sampling theorem: $2F \le N = 1/\Delta x$, $2G \le N = 1/\Delta y$, so the sampling rate is valid. If H_0 is rejected, larger sampling frequencies must be used.

5. SMA model for quality monitoring and evaluation

5.1. *SMA for detection and diagnosis*

The SMA model provides a tool for variation pattern isolation and identification. These patterns are *fingerprints* of a process carrying the manufacturing process information and can be used for monitoring and diagnosis (they also provide a guide for mode composition in tolerance simulation). A domain expert can easily attribute a major pattern to the cause of a problem. For example, in a subassembly or a final product, a rigid mode error represents datum- or locator-induced part position/orientation deviations, whereas a higher mode pattern reveals errors caused by a manufacturing process, such as tool (e.g., die or fixtures) misalignment, spring back, fixture deformation, the machine's structural compliances (stiffness and directional orientation), thermal effects, etc. (Dornfeld and Lee, 2008).

5.2. *Multivariate process monitoring by Hotelling T^2 control charts*

5.2.1. *Mode-based T^2 control chart*

Feature variability in an SMA model is expressed as $f(x, y) = \mathbf{X}\mathbf{b}_1 + \boldsymbol{\varepsilon}$ with all of the reserved significant modes in $\mathbf{X}\mathbf{b}_1$ and residuals $\boldsymbol{\varepsilon}$. In manufacturing $\mathbf{b}_1$ is determined by measurements on manufactured feature(s). Thus, $\mathbf{b}_1$ carries process information for monitoring.

Suppose $f(x, y)$ is measured on m parts, p modes are preserved and estimated by DCT, $\mathbf{b}_1 \sim \boldsymbol{N}_p(\boldsymbol{\beta}_1, \boldsymbol{\Sigma})$. $\boldsymbol{\beta}_1$, $\boldsymbol{\Sigma}$ can be estimated by sample mean $\hat{\boldsymbol{\beta}}_1$ and sample covariance matrix $\hat{\boldsymbol{\Sigma}}$, respectively. $\hat{\boldsymbol{\beta}}_1$ is the sample mean vector, $\hat{\boldsymbol{\beta}}_1 = [\hat{\boldsymbol{\beta}}_{1_1}, \hat{\boldsymbol{\beta}}_{1_2}, \ldots \hat{\boldsymbol{\beta}}_{1p}]^{\mathrm{T}}$, and $\hat{\boldsymbol{\beta}}_{1k} = (1/m)\sum_{i=1}^{m} b_{ki}\ k = 1, 2, \ldots,$ p is the estimated mean of the kth modal parameter. The covariance matrix is estimated by

$$> \hat{\boldsymbol{\Sigma}} = \frac{1}{m-1}\sum_{i=1}^{m}(b_{1i} - \bar{\hat{\beta}}_1)(b_{1i} - \bar{\hat{\beta}}_1)^{\mathrm{T}} = \hat{\sigma}^2\mathbf{I}.$$

A mode-based T^2 control chart is introduced below for the variability monitoring. For a single observation, this control chart signals a shift in the mean $\hat{\boldsymbol{\beta}}_1$,

$$T_i^2 = (b_1 - \hat{\beta}_1)^{\mathrm{T}}\hat{\Sigma}^{-1}(b_1 - \hat{\beta}_1) = \frac{(b_1 - \hat{\beta}_1)^{\mathrm{T}}(b_i - \hat{\beta}_1)}{\hat{\sigma}^2} > h_1, \quad (13)$$

where h_1 is a specified control limit. In Phase I, a preliminary charting procedure should be constructed to purge the outliers in observations. The control limit h_1 for Phase I is determined by (Lowry and Montgomery, 1995): $h_1 = \{(m-1)^2/m\}B_{[\alpha;p/2,(m-1-p)/2]}$; $B_{[\alpha;p/2,(m-1-\mathrm{p})/2]}$ is the upper αth quartile of the beta distribution with parameters $p/2$ and $(m-1-p)/2$. For the observation with subgroup of size n, out-of-control is detected by

$$T_i^2 = n(\bar{b}_{1i} - \hat{\beta}_1)^{\mathrm{T}}\bar{\hat{\Sigma}}^{-1}(\bar{b}_{1i} - \hat{\beta}_1) = \frac{n(\bar{b}_{1i} - \hat{\beta}_1)^{\mathrm{T}}(\bar{b}_{1i} - \hat{\beta}_1)}{\hat{\sigma}^2} > h_2, \quad (14)$$

where $\bar{b}_{1i}$ is the sample mean within ith subgroup and the ground mean $\hat{\beta}_1 = (1/m)\sum_{i=1}^{m}\bar{b}_{1i}$; $\bar{\hat{\boldsymbol{\Sigma}}} = (1/m)\sum_{i=1}^{m}\hat{\boldsymbol{\Sigma}}_i$. $\hat{\boldsymbol{\Sigma}}_i$ is the within ith sample group covariance matrix. The control limit h_2 is determined by (Mason and Young, 2002)

$$h_2 = \frac{p(m-1)(n-1)}{mn-m-p+1}F_{\alpha,p,mn-m-p+1}.$$

In Phase II, the control limits for individual observation ($n = 1$) are (Mason and Young, 2002)

$$UCL_1 = \frac{p(m+1)(m-1)}{m^2-mp}F_{\alpha,p,m-p}, \quad LCL_1 = 0. \quad (15)$$

For the subgroup of size $n > 1$, the corresponding upper and lower control limits are

$$UCL_2 = \frac{p(m+1)(n-1)}{mn-m-p+1}F_{\alpha,p,mn-m-p+1}, \quad LCL_2 = 0. \quad (16)$$

5.2.2. *Residual T^2 control chart*

The residuals in the SMA model are $\boldsymbol{\varepsilon} = f(x, y) - \mathbf{X}\hat{\boldsymbol{\beta}}_1$. $\boldsymbol{\varepsilon}$ characterizes the feature deviation from the average variation patterns. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}^{-1} = 1/\hat{\sigma}^2$. If the process is stable and in control the residuals are assumed to be trivial unless some special causes drive some of these modes to a significant level (mean shift or variance change). As a necessary supplement to the above three schemes, a method is proposed in this section based on the Hotelling T^2 chart. It

is suggested that it is used together with any of the other schemes.

For individual observation let us take the residual of sample i as an observed random variables, since it is usual to appeal to large sample approximation, so we have

$$T_\varepsilon^2 = \boldsymbol{\varepsilon}\,\hat{\boldsymbol{\Sigma}}_\varepsilon^{-1}\,\boldsymbol{\varepsilon} \sim \chi^2_{N^2-p} \quad UCL = \chi^2_{N^2-p}(\alpha) \quad (17)$$

$$E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = \sigma^2(N^2 - p). \qquad \hat{\sigma}^2 = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{(N^2 - p)}$$

For $n > 1$ the control chart is based on a group mean of residuals:

$$\bar{\boldsymbol{\varepsilon}} = \sum_{i=1}^{m} \boldsymbol{\varepsilon}_i \sim N(0, \sigma^2/m), \quad T_\varepsilon^2 = \bar{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\Sigma}}_\varepsilon^{-1}\bar{\boldsymbol{\varepsilon}} \sim \chi^2_{N^2-p} \text{ and}$$

$$UCL = \chi^2_{N^2-p}(\alpha) \quad (18)$$

5.3. *Process capability analysis*

The multivariate process capability indices in the literature exclusively rely on a normality assumption and a specification domain with a simple regular shape (e.g., rectangle, cube, ellipsoid, etc.). However, the GD&T tolerances specified on a feature are usually interrelated and form an irregular specification domain (e.g., Maximum Material Condition (MMC), containment/floating relations among position/size, orientation, form tolerances, etc.). Yield, a conformity-based process capability index, provides a more versatile and generic alternative (Liu *et al.*, 2013). Yield is defined as the probability of a feature conforming to a specification domain Ω; i.e., $Y = \text{Prob}[f(x, y) \in \Omega]$.

The SMA mode-based model conforms to GD&T tolerances: the rigid modes characterize variations in size/position/orientation of a perfect feature, whereas the higher modes represent form variations. Thus, it allows for a yield estimation process in a capability study.

With an SMA model, $f(x, y) = \mathbf{X}\mathbf{b}_1 + \boldsymbol{\varepsilon}$ with p reserved mode $\mathbf{b}_1 = \{b_i\}_{p\times 1}, i = 1, \ldots, p$, can be randomly simulated with the known distributions of $\mathbf{b}_1$ that are characterized by sample means and variances. The statistical model of $\mathbf{b}_1$ is estimated with the data from a number of sampled parts. With a multivariate normality assumption, the probability density distribution (p.d.f.) of the key mode coefficients $\mathbf{b}_1$ can be conveniently estimated. Kernel p.d.f. estimation gives an alternative method for more generic cases when the normality is not warranted (Huang *et al.*, 2012; Liu *et al.* 2013).

The yield-based process capability study involves two major steps: (i) generating a number of random feature samples via Monte Carlo simulation from estimated p.d.f. and (ii) conformity checks. When a kernel p.d.f. is used, Metropolis–Hasting sampling or Monte Carlo integration algorithms should be utilized for random sample generation or conformity check (Huang *et al.*, 2012).

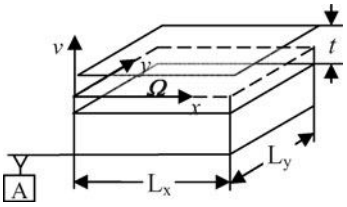

Fig. 4. A TZ.

To illustrate the capability analysis with interrelated GD&T tolerances, Fig. 4 shows a TZ t for a position on the top face of a rectangular planer feature. Other tolerances such as orientation t' and form t'' are not shown. The interrelationships among them will be described later. The feature variation is composed of rigid modes and form variation modes. The linear terms $v_S = b_1$, $v_O = b_2x + b_3y$ below are *rigid modes* if a variation is considered as a spatial signal. These terms approximate the first three modes in the SMA model (Equation 4 and Fig. 2). Thus, $\{b_1, b_2, b_3\}$ characterizes the rigid-body motion of a perfect feature.

The modes in $v_F = \sum_{i=4}^{k}\sum \gamma_i\varphi_i(x, y)$ characterize a *form variation* (warping, bending, twisting, etc.):

$$f(x,\ y) = v_S + v_O + v_F = b_1 + b_2x + b_3y + \sum_{i=4}^{k} b_i\varphi_i(x, y). \quad (19)$$

The variations in Equation (19) are decomposed into modes as illustrated in Fig. 5. On the left side of Fig. 5 there are examples of feature variations on the top face in Fig. 4.

The containment and floating relations of GD&T tolerances is defined by $t'' \subseteq t' \subseteq t$, where t, t', t'' are position, orientation, and form tolerances. Their relations can be expressed as

$$|v_0| = |v_S + v_O + v_F| = |b_1 + b_2x + b_3y + \sum_{i=4}^{k} b_i\phi_i(x, y)| \le t/2 \quad (t = \text{width of size tolerance}), \quad (20a)$$

$$|v_1| = |v_O| = |b_2x + b_3y + \sum_{i=4}^{k} b_i\phi_i(x, y)| \le t'/2 \quad (t' = \text{width of orientation tolerance}), \quad (20b)$$

$$|v_2| = |v_F| = \left|\sum_{i=4}^{k} b_i\phi_i(x, y)\right| \le t''/2 \quad (t'' = \text{width of form tolerance}), \quad (20c)$$

where the three constraints must be subjected to the Maximum/Minimum Material Condition (MMC) specified in GD&T (Rule #1). The containment and floating relations of GD&T tolerances allow wobbling and shrinking of t' inside t and t'' inside t'. These relationships lead to irregularity of the specification domain. Figure 6 illustrates an example of this relation between orientation and

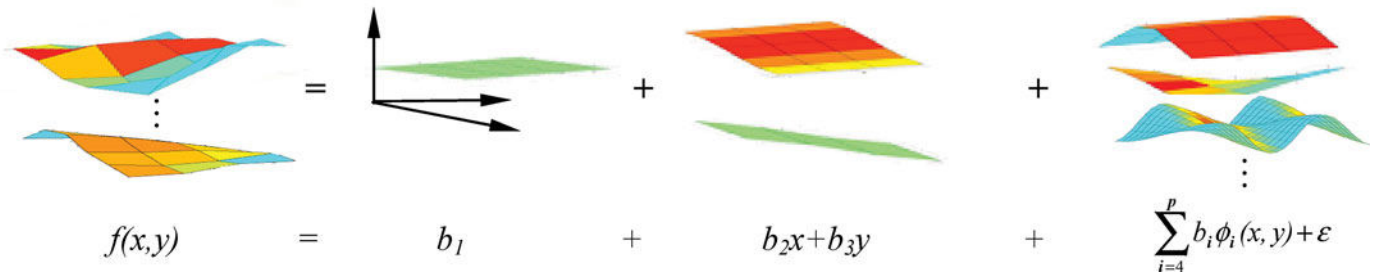

Fig. 5. Modal decomposition of a 2D planar surface feature.

position tolerances; i.e., t' and t. As shown in the figure, when a planar feature position b_1 approaches its tolerance boundary $t/2$, the orientation tolerance window t' must be adjusted (shrunk) accordingly, allowing a smaller allowance for b_2 and b_3.

The conformity check procedure is as follows: (i) estimating p.d.f. of mode coefficients $\mathbf{b}_1$; (ii) generating random feature sample $\{b_1\ b_2, \ldots, b_n\}$ from the p.d.f. by Monte Carlo or other sampling algorithms; and (iii) checking conformity of all sample points to the GD&T specification region.

6. Case studies

Case study I: One typical ship panel used in shipbuilding is shown in Fig. 7. Measurement data of 40 panel samples are utilized; 21 × 17 nodes evenly scatter in the x and y directions on each panel. A signal energy of 98%, confidence level $1 - \alpha = 99\%$, and maximum Hausdorff distance $\delta = 0.1$ were set as the criteria for significant mode selection. All three mode truncation criteria produced identical results. Five significant modes are $\Omega = \{(0, 0), (2, 0), (4, 0), (0, 2), (0, 4)\}$. One of the initial panel variation and the recovered variation (from SMA model with the five modes) are shown in Fig. 8. The five significant modes are shown in Fig. 9.

There may be many contributors to the variations in panel fabrication such as material handling, plasma cutting, geometrical design of the panel, assembly sequence, welding process parameters, etc. Experienced process engineers may easily link the mode patterns to the causes by looking at the mode shapes. The interpretations of the significant modes in the case study are given in Table 1. If the five modes are selected as key quality characteristics, the multivariate control chart can be established as illustrated in Case III.

The SMA analysis can also be extended to those non-planar features that can be developed onto a 2D plane (e.g., $x-\varphi$ or $x-y$ plane), such as cylindrical, conic, non-flat free surface, spherical surfaces, etc. The measured variations on these features are actually treated as images on the projected 2D planes.

Case study II: An application on a MEMS surface for image feature recognition and extraction is illustrated below (Fig. 10). An RF-switch MEMS has two surfaces that are in an on–off contact working status with a high frequency. The flatness of the contacting surfaces affects the component's performance and reliability. Contact analysis requires the multi-scale simulation of surface topography in a range of nanometers up to 10^2 μm. An RF-switch MEMS bottom layer surface was scanned by AFM on a 10×10 μm^2 area as shown in Fig. 10(a). The topography is represented by a 256 × 256 pixel (in the vertical or z direction) image matrix. A regular fractal structure was found in the signal. The regular hexagonal patterns match the evenly distributed small holes on the top metal layer. The holes are designed to allow chemicals to flow down through the top layer in an etching process. These patterns are composed of the remnants of undesired material, that deteriorating the flatness of the surface. One of the authors' ongoing research topics is to identify the regular hexagonal pattern and characterize both regular and fractal components with appropriate models; i.e., deterministic and fractal models. SMA and the ESC were adopted for extracting

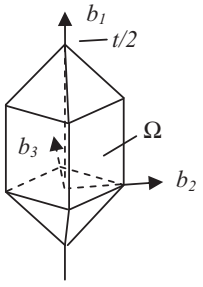

Fig. 6. The three-dimensional tolerance region.

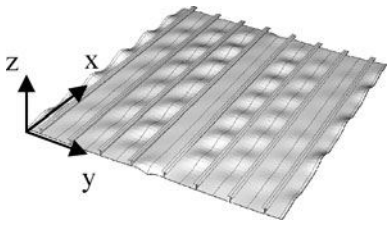

Fig. 7. Variation patterns on a planar part.

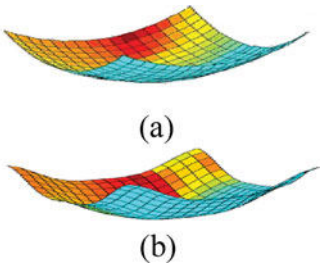

Fig. 8. (a) Measured surface and (b) recovered surface.

and modeling the hexagonal patterns. The original image signal was contaminated by MEMS sample orientations (a gradual color change from the left to the right in Fig. 10(a)) about the x and y directions, attributable to setup errors (e.g., by uneven glue below the component for holding the sample during AFM measurements). In addition, the zero mode, representing overall signal level in the z direction, is irrelevant to the flatness of the surface and needs to be eliminated. These undesirable signal components must be removed before further processing.

To characterize and model the undesirable hexagonal patterns for quality monitoring and further contact analysis, a DCT transformation was conducted on the topography image. The zero and orientations modes ($C(1, 1), C(1, 2)$, $C(2, 1)$) were removed by nullifying these mode coefficients (Fig. 10(b)) from the DCT result. The key modes were selected and reserved using ESC by checking and comparing the energy contribution ratio of each mode against a selected threshold δ. Over 97% of signal energy (when $\delta = 0.000\ 01$) was reserved in these modes and they satisfactorily extract and approximate these patterns as shown in Fig. 10(c). The data compression rate is about 98%. A tradeoff must be made between accuracy and dimension reduction (compression rate) in selecting δ. The smaller the δ, the higher the model dimension.

This topography SMA model serves two purposes: (i) characterizing the hexagonal patterns for contract and performance analysis in MEMS design and itching process improvement and (ii) providing a compressed variable space for quality monitoring (e.g., multivariate SPC). The complete topography model for contact analysis may require modeling both the hexagonal pattern and fractal irregularity; the latter is beyond the scope of this article and will be presented elsewhere.

Case study III: **A** mode-based Hotelling T^2 control chart was constructed using simulated modes of planar features. Five modes ($p = 5$) were selected to characterize manufactured planar features. Mode coefficients $C(0, 0)$, $C(1, 0)$, $C(0, 1)$, $C(1, 1)$, and $C(2,1)$ were simulated using a known multi-normal distribution that was hypothetically estimated from an in-control process. To construct the T^2 control chart we assumed $p = 5, m = 35, n = 1$; i.e., using 35 initial samples in Phase I. These parameters were selected only for validation and illustration purposes. With a Type I error $\alpha = 0.005$ the upper/lower control limits were obtained by Equation (15) as $UCL = 24.655$ and $LCL = 0$. The hypothetically observed T^2 were calculated from simulated mode coefficients. The initial 200 T^2 values (denoted as $T_{\text{-sq}}$) were simulated from the in-control process with all points being within the control limits (Fig. 11). Then, the combined mean shifts in $C(0, 0)$ and $C(1, 0)$ modes by 2σ and 1σ were intentionally created, respectively. A rising trend can be seen starting from the 200th point. The shifted process was simulated for subsequent T^2 point calculations. The first two out-of-control points are at the 233th and 268th samples ($T^2 = 25.664$ and $T^2 = 26.035$).

Reducing Type I error α by lifting UCL (e.g., $\alpha = 0.0027$ or $ARL_0 = 370$ as for a Shewhart chart; thus $UCL = 26.076$ and the first out-of-control point is at the 282th sample with $T^2 = 27.345$) increases the Type II error, deteriorating the ability of the chart to detect anomolies and resulting in a higher average run length ARL_1. This chart, like Shewhart charts, seems to be insensitive to small or moderate changes. The authors therefore suggest using multivariate CUSUM or EWMA techniques to improve the chart's sensitivity. If an out-of-control signal is triggered on a T^2 chart checking individual mode charts (using a univariate Shewhart chart) may help to identify major contributors to the process change. The mode interpretations as shown in Table 1 shows the possible root causes (e.g., misalignment, residual stress, spring back, etc.), rendering the SMA T^2 control chart a unique diagnostic advantage.

Case study IV: It is impossible for conventional univariate methods and indices (e.g., C_p, C_{pk}) to evaluate process capability with GD&T specifications. GD&T specified on a planar feature, for example, usually produces an irregularly-shaped specification region and involves multiple quality characteristics (Liu *et al.*, 2013). For example, GD&T on a planar feature may allow the feature variations in position/size, orientations around two axes, waviness, etc., within their respective tolerances. Their interrelated relationships define an irregular region as shown in Fig. 6. Liu *et al.* (2013) illustrate more irregular tolerance regions

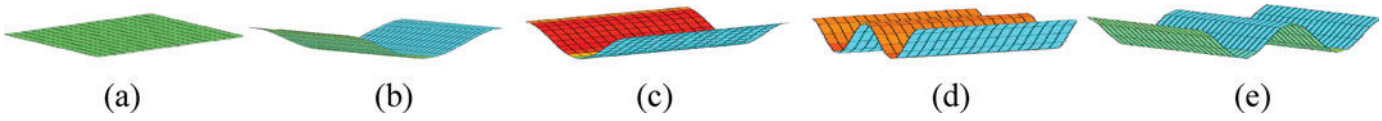

Fig. 9. Significant modes selected from 40 panels.

Table 1. Case study I: Mode interpretation

Modes	*Coefficients*	*Error interpretation*
Fig. 9(a)	$C(0,0)$	Mean shift caused by fixture calibration
Figs. 9(b)–(e)	$C(2,0)$, $C(0,2)$, $C(4,0)$, $C(0,4)$	Material handling-induced modes: residual stress in milling process and in material handling on shop floor (moving) as well as uneven support in conveyance systems

*Error interpretation was based on the information provided by DCS, Inc. Troy, MI.

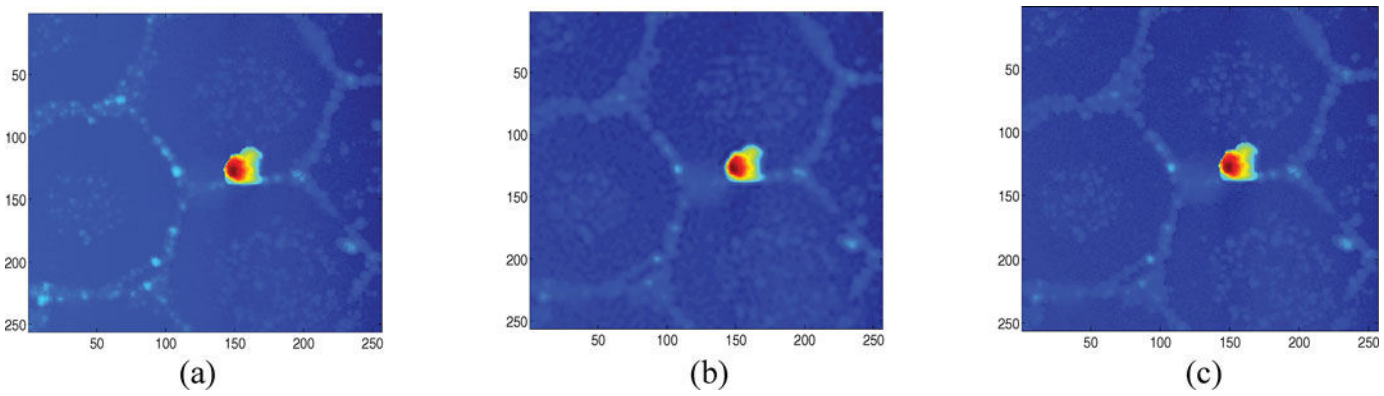

Fig. 10. (a) Original bottom layer topography by AFM (10×10 μm^2), (b) the topography after the zero mode and orientation components are removed, and (c), image recovered by the reserved modes.

with various GD&T designs. In addition, the multivariate normality may not be warranted.

Unlike conventional indices C_p and C_{pk}, the conforming probability p to a GD&T specification is introduced here as an index for the evaluation. A rectangular planar feature is used for illustrating process capability analysis with GD&T tolerances. As shown in Fig. 4, the feature has the dimensions of $L_x = 60$, $L_y = 40$, size tolerance $t = 0.5$, orientation tolerance $t' = 0.2$, and we ignore t'' to simplify the visual illustration. With Rule 1 in the GD&T standard

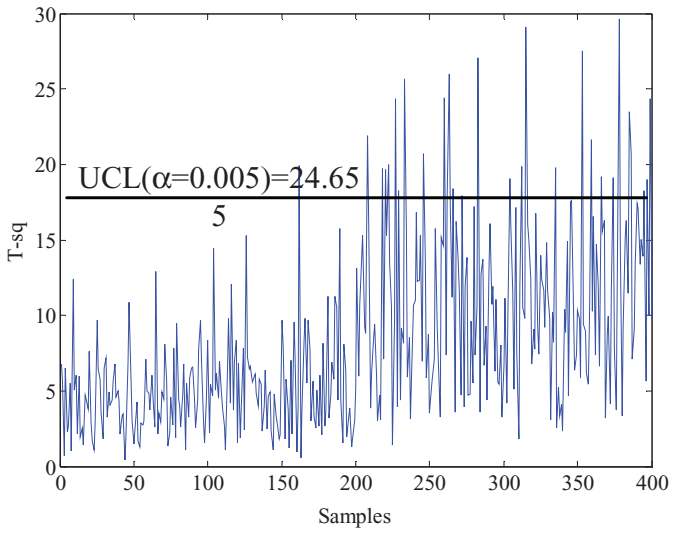

Fig. 11. Mode-based T^2 control chart UCL($\alpha = 0.005$) = 24.655.

the two tolerances have a containment relation; i.e., the orientation tolerance must float within the size tolerance. To evaluate process capability, a sufficient number of manufactured part features must be initially measured from an in-control process in order to estimate a statistical model. In the case study we assume that (i) the first three mode coefficients (characterizing size and orientation variations) are normally distributed and (ii) their p.d.f. is estimated from measurement data and DCT results. A kernel method can also be adopted for p.d.f. assessment when normality is not warranted, as shown in Huang *et al.* (2012) and Liu *et al.* (2013). With the p.d.f. there are two methods for conformity p estimation; i.e., random simulation and Monte Carlo integration. The former is adopted in the case study; the latter can be found in Liu *et al.* (2013).

An orientation tolerance zone Ω_o is shown in Fig. 12. A simulation algorithm was applied in the case study to generate samples. Sample size $N = 100\ 000$ was chosen to ensure accuracy. The overall conformity p with both position and orientation is about 0.9952. Marginal conformities $p_1 = 99.73\%$ and $p_2 = 99.78\%$ were estimated by similar analysis algorithm that counts only the conformity to each individual tolerance; i.e., size and orientation. p_1, p_2 explain the probabilities (= 0.9973 and 0.9978) of manufactured features falling in the size and orientation tolerance ranges, respectively. When p is not satisfactory, marginal conformities help to identify the contributions from different sources, prioritizing process improvement actions toward the source(s) with lower

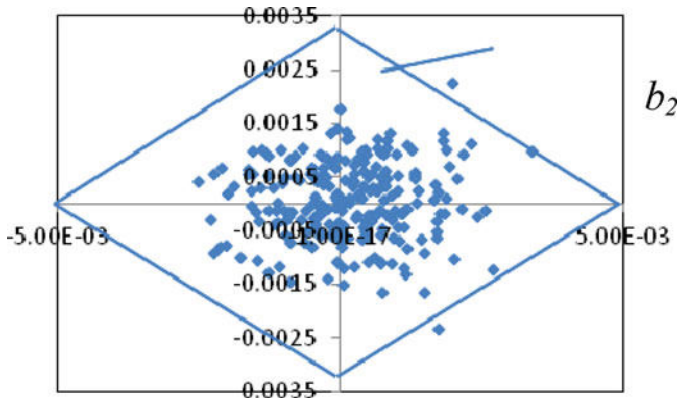

Fig. 12. Tolerance conformity of a rectangular planar feature.

marginal conformities. For example, if $p_1 < p_2$ increasing p_1 is more effective for improving overall conformity p.

7. Summary and discussion

The information carried in variations on part features can be utilized for quality control in two ways: (i) in manufacturing: pattern identification for monitoring, predicting changes via statistical models, evaluating process capability, and supporting decision making; and (ii) in design: replicating variation patterns and predicting/optimizing manufacturability and product performance in tolerance design. The proposed SMA methodology is an approach to create a unified and coherent model to meet both requirements in design and manufacturing. SMA has many facets in application as shown in Sections 5 and 6. The emphasis here is on the manufacturing side, covering the topics of quality monitoring, diagnosis, and process capability study.

A mode-based model provides a platform for planar geometric feature (surface) variation characterization. It can find applications in both manufacturing (for quality monitoring and control) and design (for GD&T tolerance simulation). The model can also be easily extended to those features (in the authors' ongoing work) that can be developed on a plane (e.g., the features with axes of rotational symmetry, such as cylindrical, conic, spherical surfaces, etc.). This property generalizes its applicability to a variety of features. A systematic procedure is designed for statistical modal analysis and modeling, including mode decomposition via DCT or regression, mode truncation criteria, and sampling. DCT is highly efficient for the SMA analysis on evenly spaced samples; regression with the same cosine base functions gives an alternative method for uneven samples. The significance in both engineering and statistics is considered in mode truncation.

The DCT algorithm uses orthogonal expansion to approximate feature variations. The resultant mode-based model has a much smaller variable size (i.e., coefficients of a few key modes) that alleviates the difficulty for control chart construction that may otherwise involve tens of thousands of variables. The interpretations of the modes based on process knowledge and experiences can be linked to the generating mechanisms such as die misalignment, setup errors, force-induced deformation, vibration, etc. It benefits root cause identification and diagnosis for quality improvement.

The modes can also be interpreted in the GD&T context (e.g., position/size, orientation, form, etc.). The GD&T interpretation makes tolerance simulation possible by mode composition (e.g., inverse DCT as shown in the recent literature (Huang *et al.*, 2010; Liu *et al.*, 2013). The tolerance interpretation also allows mapping mode coefficients directly into a tolerance region for conformity checks, especially for the lowest three modes that describe kinematic movements of a planar feature. A numerical simulation method for multivariate process capability (MPC) analysis is created using conformity as an index. The interactions among GD&T tolerances often lead to an irregular tolerance region, preventing conventional MPC analysis from application in this case.

Funding

The authors gratefully acknowledge the financial support from National Science Foundation awards (NSF-CMMI: 0928609, 1129741, and 0927557).

References

Ahmed, N., Natarajan, T. and Rao, K.R. (1974) Discrete cosine transform. *IEEE Transactions on Computers*, **23**, 90–94.

Camelio, J., Hu, S.J. and Marin, S. (2004) Compliant assembly variation analysis using component geometric covariance. *Transactions of the ASME, Journal of Manufacturing Science & Engineering*, **126**, 355–360.

Ceglarek, D., Huang, W., Zhou, S., Ding, Y., Ramesh, K. and Zhou, Y. (2004) Time-based competition in manufacturing: stream-of-variation analysis (SOVA) methodology—review. *International Journal of Flexible Manufacturing Systems*, **16**(1),11–44.

Ceglarek, D. and Shi, J. (1995) Dimensional variation reduction for automotive body assembly. *Manufacturing Review*, **8**(2), 139–154.

Dornfeld, D. and Lee, D. (2008) *Precision Manufacturing*, Springer, New York, NY.

Gupta, S. and Turner, J.U. (1993) Variational solid modeling for tolerance analysis. *IEEE Computer Graphics and Applications*, **13**, 64–74.

Huang, W. and Ceglarek, D. (2002) Mode-based decomposition of part form error by discrete-cosine-transform with implementation to assembly and stamping system with compliant parts. *Annals of the CIRP*, **51**(1), 21–26.

Huang, W., Konda, B.R. and Kong, Z. (2010) Geometric tolerance simulation model for rectangular and circular planar features. *Transactions NAMRI / SME*, **38**, 363–370.

Huang, W., Lin, J., Bezdecny, M.R., Kong, Z. and Ceglarek, D. (2007) Stream-of-variation modeling I: a generic 3D variation model for rigid body assembly in single station assembly processes. *Transactions of the ASME, Journal of Manufacturing Science & Engineering*, **129**(4), 821–831.

Huang, W., Lin, J., Kong, Z. and Ceglarek, D. (2007) Stream-of-variation (SOVA) modeling II: a generic 3D variation model for rigid body assembly in multi station assembly processes. *Transactions of the ASME, Journal of Manufacturing Science & Engineering*, **129**(4), 832–842.

Huang, W., Pahwa, A. and Kong, Z. (2012) Kernel density & Metropolis–Hastings sampling in process capability analysis, in *Proceedings of the ASME International Manufacturing Science & Engineering Conference*, pp. 651–660.

Johnson, R.A. and Wichern, D.W. (1998) *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Li, B. and Roy, U. (2001) Relative positioning of toleranced polyhedral parts in an assembly. *IIE Transactions*, **33**, 323–336.

Liu, J., Kong, Z., Zhou, Y. and Huang, W. (2013) Process capability analysis with GD&T specifications, in *ASME 2013 International Mechanical Engineering Congress & Exposition (IMECE2013)*.

Lowry, C.A. and Montgomery, D.C. (1995) Review of multivariate control charts. *IIE Transactions,* **27**, 800–810.

Mason, R.L. and Young, J.C. (2002) *Multivariate Statistical Process Control with Industrial Applications*, SIAM, Philadelphia, PA.

Merkley, K.G. (1998) Tolerance analysis of compliant assemblies. Ph.D. thesis, Brigham Young University, Provo, UT.

Requicha, A.A.G. (1983) Toward a theory of geometric tolerancing. *International Journal of Robotics Research*, **2**(4), 45–60.

Requicha, A.A.G. and Chan, S.C. (1986) Representation of geometric features, tolerances and attributes in solid modelers based constructive geometry. *IEEE Journal of Robotics & Automation*, **2**(3), 156–186.

Roy, U. and Li, B. (1998) Representation and interpretation of geometric tolerances for polyhedral objects—I. Form tolerances. *Computer-Aided Design*, **30**, 151–161.

Roy, U. and Li, B. (1999) Representation and interpretation of geometric tolerances for polyhedral objects II. Size, orientation and position tolerances. *Computer-Aided Design*, **31**, 273–285.

Srinivasan, V. and Jayaraman, R. (1989) Geometric tolerancing: II. Conditional tolerances. *IBM Journal of Research and Development*, **33**(2), 105–124.

Srinvasan, R.S. and Wood, K.L. (1997) A form tolerancing theory using fractals and wavelets. *Journal of Mechanical Design*, **119**, 185–193.

Turner, J.U. and Wozny, M.J. (1987) Tolerances in computer-aided geometric design. *The Visual Computer*, **3**, 214–216.

Voelcker, H.B. (1998) The current state of affairs in dimensional tolerancing: 1997. *Integrated Manufacturing Systems*, **9**(4), 205–217.

Wells, D., Vanicek, P. and Pagiatakis, S. (1985) Least squares spectral analysis revisited. Technical Report no. 84, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, NB, Canada.

Appendix

Notation

$\mathbf{b}$, $\mathbf{b}_1$ = estimated model coefficient vector and its sub-vector

$\bar{\mathbf{b}}_1$ = a sample mean of $\mathbf{b}_1$ estimated from multiple parts

$\{\tilde{C}(u, v)\}$ = Discrete-cosine-transformation coefficient vector

$C(u, v)$ = $\tilde{C}(u, v)\alpha(u)\alpha(v)$

$\mathrm{d}(S, S')$ = Housdorff distance of surfaces S and S'

$f(x, y)$ = variation at a coordinate of (x, y)

$\mathbf{f}$,$\hat{\mathrm{f}}$, = observed and estimated variation vectors

$\hat{f}_0$ = predicted variation at a new coordinate (x_0, y_0)

$F_{r, N-r}$ = F distribution with degrees of freedom of r and N-r

$h(n, m, u, v)$, $g(n, m, u, v)$ = inverse transformation kernel, and forward transformation kernel

$\mathbf{H}^{\mathrm{T}}_{u, v}$ = a discretized cosine base vector

$I_{(u, v)}$ = a confidence interval of $C(u, v)$

M, N = sample sizes

r = dimension of vector $\mathbf{b}_1$

t_{N-r} = t distribution with degrees of freedom $N-r$

$T(u, v)$ = $\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} f(n, m)g(n, m, u, v)$ a transformation coefficient

u, v = mode indices in x and y directions

$\mathbf{X}$ = a matrix composed of $\bar{I}$and all the vectors $\boldsymbol{H}_{u, v}$

$\alpha(u)$, $\alpha(v)$ = constant factors

α = confidence level

$\boldsymbol{\beta}$, $\boldsymbol{\beta}_1$ = regression coefficient vector and its sub-vector

χ^2_r = chi-square distribution with a degrees of freedom r

δ = a specified Housdorff distance threshold

$\boldsymbol{\varepsilon}_t$, $\bar{\boldsymbol{\varepsilon}}_t$ = a residual vector and its sample mean

$\hat{\bar{\boldsymbol{\varepsilon}}}_t$ = $[\mathbf{I} - \mathbf{X}\mathbf{X}^{\mathrm{T}}]\bar{\boldsymbol{\varepsilon}}_t$

σ = standard deviation of $\boldsymbol{\varepsilon}_t$

Ω = a mode index set

Biographies

Wenzhen Huang is an Associate Professor in Mechanical Engineering at the University of Massachusetts, Dartmouth. He earned B.S. (1982) and M.S. (1984) degrees in Mechanical Engineering from Jilin University and Shanghai Jiaotong University, China, respectively. He earned a Ph.D. in Industrial and System Engineering from the University of Wisconsin–Madison in 2004. His recent research interests are GD&T tolerance modeling, statistical tolerance design, surface characterization of nano/microstructures, and multivariate statistics in quality control. His recent research has been supported by ATP/NIST and the NSF. He has published over 45 peer-reviewed journal articles and a number of conference papers.

Jinya Liu is a Ph.D. candidate at the University of Massachusetts Dartmouth (UMASSD). Her present work focuses on modeling for multi-scale MEMS surface, contact analysis, as well as performance and reliability improvement. She earned an M.S. degree in Mechanical Engineering at UMASSD in 2012. The work performed for her Master's degree emphasized multivariate process capability analysis with GD&T specifications.

Vijaya Chalivendra is an Associate Professor in Mechanical Engineering at the University of Massachusetts Dartmouth. He earned a Ph.D. in Mechanical Engineering and Applied Mechanics from the University of Rhode Island. His recent research interests are nano-mechanical

characterization, surface characterization of MEMS structures, biological materials, and biomimicking. He has published over 35 peer-reviewed journal articles and 40 conference proceedings. He served as the Editor for a special issue of *Experimental Mechanics* in 2012.

Dariusz (Darek) Ceglarek earned a Ph.D. degree in Mechanical Engineering from the University of Michigan–Ann Arbor in 1994. He was on the research faculty at the University of Michigan–Ann Arbor from 1995 to 2000. He was an Assistant Professor, Associate Professor, and Professor in the Department of Industrial and Systems Engineering in 2000, 2003, and 2005, respectively. Since 2006 he has been Professor and EPSRC Research Chair at the University of Warwick, UK. His research focuses on product life cycle modeling and analysis with an emphasis on production and healthcare systems convertibility, scalability, and Six Sigma quality and root cause analysis during design, manufacturing, and service phases. He has received numerous awards including Best Paper Awards by ASME, a UK EPSRC STAR Award (granted to an "exceptional senior faculty, recognized international leader in his/her research field"), U.S. National Science Foundation 2003 CAREER Award, etc.

Zhenyu (James) Kong received B.S. (1993) and M.S. (1995) degrees in Mechanical Engineering from the Harbin Institute of Technology, Harbin, China, and a Ph.D. (2004) degree from the Department of Industrial and System Engineering, University of Wisconsin–Madison. He is an Associate Professor with the School of Industrial Engineering and Management at Oklahoma State University. He has authored or co-authored a number of articles in various journals. His research has been sponsored by the National Science Foundation, the Department of Transportation, and Dimensional Control Systems Inc. His research focuses on automatic quality control for complex manufacturing processes/systems. He is a member of IIE, INFORMS, ASME, and SME.

Yingqing Zhou earned his Ph.D. in Systems Engineering from the Case Western Reserve University, Cleveland, Ohio, in 1992. After graduation, he was a senior developer in Trikon Design from 1992 to 1995. Since 1995, he has been the Director of Research and Development with Dimensional Control Systems, Inc. His current researches includes dimensional variation simulation and optimization in kinematic motion systems and deformable components. Dr. Zhou has been the chief editor for *Dimensional Engineering News*, a monthly newsletter, since May 2003.