

The process-oriented multivariate capability index

E.J. FOSTER^{†*}, R.R. BARTON[‡], N. GAUTAM,
L.T. TRUSS[†] and J.D. TEW[†]

[†]General Motors Corporation, Mail Code 480-106-359, 30500 Mound Road, Warren, MI 48090-9055

[‡]The Pennsylvania State University, Business Administration Building,
University Park, PA 16802-3009

(Received September 2004)

Recent literature has proposed multivariate capability indices, but does not suggest a method for measuring quality characteristics in a way that links production irregularities directly to their causes. Our objective is to present a new approach to multivariate capability indices that uses process-oriented basis representation (POBREP) which allows the computing of cause-related index values. The proposed method focuses on independent process-oriented multivariate data by employing regression coefficients as data. These coefficients measure the amount of the characteristic patterns induced by particular problems or incidents that can occur in the system. Two examples from the electronics industry (the chip capacitor process and solder paste process) use simulated data and Monte Carlo integration to demonstrate the new process-oriented capability method. A reduction of estimation error was realized when using process-oriented capability. For the chip capacitor problem, capability error is 24–54% when using ordinary multivariate data. However, when using process-oriented data the error is less than 3%. Capability is difficult to compute from sample data in the solder paste example without the process-oriented approach. Future research should propose a multivariate capability measure for dependent process-oriented data.

Keywords: Capability; Process-oriented; Multivariate; Indices; Monte Carlo simulation

1. Introduction

Multivariate capability methods usually produce one number jointly representing capability for two or more quality attributes. Although a number of multivariate indices exist, a process-oriented approach allows a new index to be constructed that has a new interpretation. When process irregularities have known causes that influence two or more quality variables, the process-oriented viewpoint seeks to quantify those causes. In contrast, the ordinary capability viewpoint measures the deviations from target of two or more quality variables while ignoring measures of the causes; this is a data-oriented approach. Using a process-oriented multivariate capability approach requires process-oriented data. The interpretation of the new

*Corresponding author. Email: earnest.foster@gm.com

index is that non-conformance to specifications (when represented by a multivariate index) is thought of as resulting from the causes influencing the process. The actual causes connect with their non-conformance measures for an intuitive multivariate capability index that is useful for measuring a process-oriented performance. In addition, the process-oriented approach can allow for data reduction in the sense that the process variables identify the major sources of variation in a system. Ordinary variables may be more numerous and therefore require a more cumbersome and sometimes impractical multivariate capability computation. Some situations will not yield a multivariate capability measure using ordinary variables. Using process variables in many of these situations, however, would allow for computing multivariate capability.

Juran *et al.* (1974) first introduced the idea of capability ratios (now called indices). The first indices were univariate, measured process capability with regard to a single quality measure X and focused on the percentage non-conforming (Kotz and Lovelace 1998). Their purpose was simply to provide a ratio of the process specifications over process spread. Univariate indices are useful for identifying dimensional problems in parts. In recent years, multivariate capability indices were developed as a natural extension to the univariate concept.

A question that naturally arises is 'Why multivariate capability indices?'. Consider a part with independent quality characteristics X and Y . The part is not a good part if X or Y is beyond engineering specifications. If $C_p = 1.0$ (company standards are that 99.73% of parts must be good) is the value representing process potential for independent variables X or Y , then the parts are not capable when two variables are considered jointly (99.46% good parts overall = 0.9973×0.9973). Multivariate capability is useful here to represent the two quality variables as one number to reflect that the joint C_p (or multivariate C_p) of X and Y is less than 1.0. The practice of reporting two acceptable C_p numbers for X and Y is potentially misleading. Additionally, if there is correlation (when variables are not independent) between X and Y , a definite affect on multivariate proportion conforming results (in this example, the percentage conformance would be $> 99.46\%$ and $\leq 99.73\%$). Quantifying these joint probabilities allows for multivariate interpretations analogous to the univariate.

2. Literature

Multivariate capability indices appeared in the literature during the early 1990s. Most of them assumed multivariate normal data, a stable process, and were generalizations of their univariate counterparts. Wang *et al.* (2001) reviewed three multivariate methods (Taam *et al.* 1993, Chen 1994, and Shahriari *et al.* 1995) in detail and computed capability for four problems. In general, Taam *et al.* (1993) presented both a multivariate C_p and C_{pm} (MC_p and MC_{pm}). Given the elliptical equiprobability contours of the multivariate normal distribution, elliptical specifications were assumed. They addressed a hole-diameter application that required elliptical specifications. Shahriari *et al.* (1995) proposed a multivariate capability index using hyper-rectangular process boxes rather than ellipses as the specification area. Shahriari *et al.* (1995) recognized the need for multiple measures of multivariate capability and proposed a three-component multivariate capability vector. The first component of the vector is a multivariate capability

measure of process potential (C_{pM}) generalizing the univariate C_p . The second component of the Shahriari *et al.* (1995) vector is the PV (P -value). It is essentially the Hotelling T^2 significance level. If PV is close to 0, the observed process is 'not close' to centre. If PV is close to 1, the process is 'very close' to centre. The third vector component is LI , the location index. If $LI=1$, the sides of the process box fall within the specification box. These three components capture multivariate process knowledge and therefore require the interpretation of three separate quantities. Chen (1994) proposed a general multivariate capability index for process potential MC_p . The strength of this approach was the defining of multivariate capability in a general way that allows elliptical and rectangular specifications. Furthermore, Chen's approach does not require the multivariate normal assumption.

The purpose of a multivariate capability index is to summarize the multivariate process performance as an aid to process improvement initiatives. The multivariate methodologies should be analogous to univariate and provide a means to estimate multivariate proportion non-conforming. Since the likelihood of someone using a particular multivariate index decreases if it requires complex computations, we recommend considering computational ease when evaluating indices. There is a computational ease advantage using methodologies by Taam *et al.* (1993) and Shahriari *et al.* (1995). Wierda (1992) proposed a multivariate capability index analogous to C_{pk} that uses a p -dimensional rectangular specification area, and an expected proportion of non-conformance items approach. The success of this method depends on whether or not the data is normal. Let θ equal the percentage of good parts. The multivariate C_{pk} is defined as

$$MC_{pk} = \frac{1}{3} \Phi^{-1}(\theta). \quad (1)$$

Also, $\hat{\theta}$ (proportion conforming) is useful alone as a multivariate index. Since $\hat{\theta}$ is defined as proportion conforming, Wierda (1993) departed from the traditional definition of C_{pk} and estimated a multivariate C_{pk} that reflected the actual yield. Determine actual yield or the percentage conforming for the univariate case using:

$$\theta = \Phi\left(\frac{USL-\mu}{\sigma}\right) - \Phi\left(\frac{LSL-\mu}{\sigma}\right).$$

Capability, if computed using Wierda's (1992) approach, gives a more conservative value than the originally defined C_{pk} .

Capability indices, both univariate and multivariate, do not identify the causes associated with low (less than 99.73% conformance) index values. The difficult task of discovering causes of production irregularities is often left to process engineers. Identify the causes proactively before the capability study and link them to the patterns representing quality problems inherent in the process, and a more efficient method for using capability as statistical process control (SPC) tool results; this describes a process-oriented capability study.

3. An overview of the process-oriented basis method

Multivariate quality databases provide large amounts of information, but they present a challenge to the quality engineer who wants to use the data effectively. Multivariate SPC tools typically identify when irregularities in system behaviour

occur, and characterize, in a multivariate sense, the major components of this variation. The major directions of variation of the multivariate process measurement vector can be found by principal components analysis (PCA), which provides an alternate basis for the vector space of multivariate process measurements (Barton and Gonzalez-Barreto 1996).

A basis is a concept from linear algebra: a vector x can be expressed as a series of numbers (x_1, x_2, \dots, x_p) . This representation uses the standard basis $\{e_1, e_2, \dots, e_p\}$, which consists of the vectors $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, $e_p = (0, \dots, 0, 1)$. A basis can be thought of as a set of patterns, and a representation of the quality vector in that basis is a recipe, describing how much (either positive or negative) of each element is needed to construct the observed quality vector. To represent x using another basis, say using the pattern vectors (basis) $\{a_1, a_2, \dots, a_p\}$, one must solve the matrix equation

$$x = Az,$$

where A is a matrix whose columns are a_1, a_2, a_q , and $q \leq p$. The vector z is the representation of x in the basis $\{a_1 | a_2 | \dots | a_q\}$. The matrix-vector product Az describes x as a linear combination of the columns in A , and z describes how much of each column of A is needed to construct x . To find z given a full basis (p columns) A and a quality vector x , we find the inverse of the matrix A and compute

$$z = A^{-1}x.$$

Using an alternate basis can be a data reduction technique. If A has fewer than p columns, say q , then z will have only q components ($p-q$ variables will have no z components). In this case, the matrix A is not square and cannot be inverted. However, z can be found by solving the least-squares equation

$$z = (A'A)^{-1}A'x.$$

Data reduction is possible without using the process-oriented approach, by using only the first k principal components of variation for the basis of the quality vector. The principal components basis is data-oriented rather than process-oriented. Thus, the first basis element (pattern) will show the direction of the largest amount of variation in the quality vector. Since this direction is data-oriented, it may not provide a diagnosis of the causes of production irregularities. Therefore, it is up to the quality engineer to work with process experts to identify the cause or causes of these irregularities and to define the appropriate action.

A process-oriented basis, on the other hand, consists of basis elements or characteristic patterns induced by particular problems or incidents that can occur in the system. Describing the quality vector in terms of the process-oriented basis tells how much of each known problem type is present. The representation of the multivariate quality vector (either mean or variance) using this basis will identify a small set of potential causes, that is, those causes associated with basis elements which have large average coefficients or large variation in the coefficients in the process-oriented basis representation.

4. Multivariate capability using POBREP

The specific contribution of this paper is that it uses the coefficients generated in the process-oriented basis representation as data for computing a multivariate capability measure. This process-oriented approach for computing capability is unique in the literature; the literature uses product-oriented data to compute multivariate capability. From another perspective, this paper extends the POBREP SPC method of Barton and Gonzalez-Barreto (1996) by developing a multivariate process-oriented capability measure.

The method for computing a multivariate POBREP capability uses coefficients as data in an $n \times q$ \mathbf{Z} matrix defined by $\mathbf{Z} = [z_1 | z_2 | \dots | z_q]$ where, for example, the z_1 through z_q are vectors (100 rows \times 1 column) of the respective z coefficients and q is $\leq p$. That is, assume the capability study will consist of 100 collected parts that individually generate q z coefficients for collection in matrix \mathbf{Z} . Since the columns of \mathbf{Z} are regression coefficients, the data contained in these columns are asymptotically normal. Therefore, assume the multivariate normal distribution for matrix \mathbf{Z} . The multivariate capability method using POBREP can potentially compute indices for the following cases:

1. $\mathbf{x} = \mathbf{Az}$ or $\mathbf{x} = \mathbf{Az} + \varepsilon$, \mathbf{z} independent, \mathbf{A} orthogonal, $\hat{\mathbf{z}} = \mathbf{A}^{-1}\mathbf{x}$, $\hat{\mathbf{z}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{x}$
2. $\mathbf{x} = \mathbf{Az}$ or $\mathbf{x} = \mathbf{Az} + \varepsilon$, \mathbf{z} independent, \mathbf{A} not orthogonal, $\hat{\mathbf{z}} = \mathbf{A}^{-1}\mathbf{x}$, $\hat{\mathbf{z}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{x}$

The focus of this paper is case 1, and the covariance matrix of \mathbf{Z} is represented as $\Sigma_{\mathbf{Z}}$ with diagonal elements $\sigma^2 z_1, \sigma^2 z_2, \dots, \sigma^2 z_q$. Off-diagonal elements of $\Sigma_{\mathbf{Z}}$ are set to zero and imply independence of causes. Let μ_i ($i = 1, \dots, q$) represent the means of the respective columns of \mathbf{Z} . Represent a vector of the column means as $\mu_{\mathbf{Z}}$. Input variables ($\mu_{\mathbf{Z}}, \Sigma_{\mathbf{Z}}$) and the z specifications are for computing $M\hat{C}pk$ analytically using

$$M\hat{C}pk = 1/3\Phi^{-1}(\hat{\theta}). \tag{2}$$

Our approach to capability is to rely on the original definition that capability is a measure to track the percentage of non-conformance. Wierda's (1992) approach directly incorporated this concept by using $\hat{\theta}$.

A major advantage of process-oriented multivariate capability is the direct link of cause to capability. The z_i 's represent the amount of pattern i (see figure 2) in the collected samples. Smaller capability values indicate that one or more causes, identified in advance by process experts, need investigation for off-target means. Multivariate capability indices should be used here as a diagnostic tool towards process improvement. They link patterns of joint variation in quality variables to their potential causes.

We can demonstrate computing multivariate capability using POBREP with the multivariate chip capacitor problem (see Barton and Gonzalez-Barreto 1996). Chip capacitors are manufactured by printing silver squares onto clay 'flats'. Figure 1 shows registration error measurements in both horizontal and vertical directions for the four squares near the flat perimeter. One of the key quality parameters is the offset of the printed silver square on the clay square. Registration errors occur when the silver squares are not properly located on the clay squares. The actual locations (solid-line squares) are shown in contrast to the

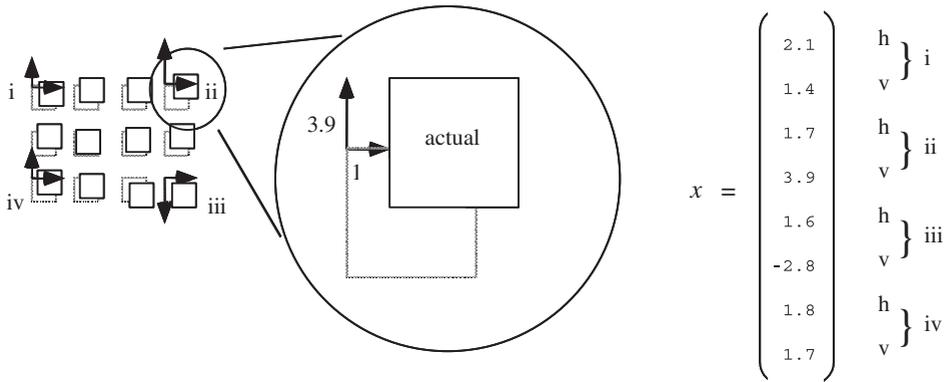


Figure 1. Registration error measurements for a multi-layer capacitor flat.

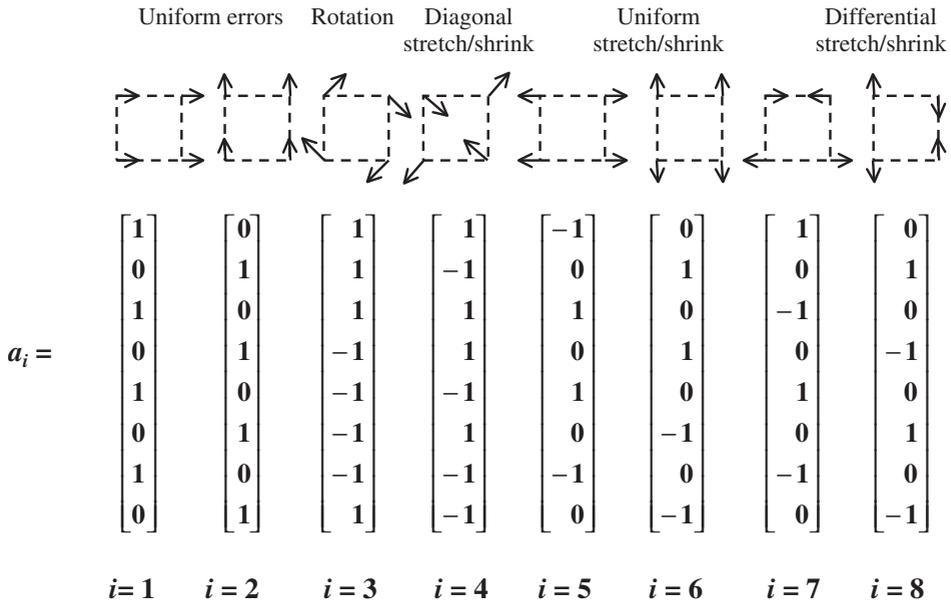


Figure 2. Process-oriented basis elements.

target (dotted-line squares). These registration errors produce a quality vector in eight-dimensional space.

Hypothesized causes of production irregularities are linked to patterns inherent in the eight-component quality vector. Patterns in this example can be classified in five groups: displacement, rotation, diagonal stretch/shrink, uniform stretch/shrink, and differential stretch/shrink. These eight process-oriented basis elements are the columns of the \mathbf{A} matrix: $\mathbf{A}=[a_1|a_2| \dots |a_8]$. Figure 2 shows the full set of basis elements in graphical form. They were chosen because they are independent, and engineers could proactively link process irregularities to them.

The eight basis coefficients (z_i) are computed by $\mathbf{z} = \mathbf{A}^{-1}\mathbf{x}$ in this full basis example. Coefficients are non-zero in practice, and the coefficients with the largest magnitudes are used for diagnosis.

For the chip capacitor problem, assume 100 capacitor flats are used for a capability study, and assume specifications (± 2 units) are available for the quality vector \mathbf{x} . The \mathbf{Z} matrix contains 100×8 z_i coefficients, and all z_i are generated using the orthogonal basis \mathbf{A} . We compared capability results using \mathbf{Z} as data (the process-oriented capability method) to results using \mathbf{X} (Wierda’s multivariate capability method). When using Wierda’s (1992) capability method, computing capability for \mathbf{X} is difficult to perform accurately due to the requirement of multivariate integration. The covariance matrix $\Sigma_{\mathbf{X}}$ may have significant correlations and thus create a difficult multivariate normal integration problem. To compute POBREP and ordinary multivariate indices for this example problem, random data sets are computed. Let a_{ij} represent individual matrix elements of basis matrix \mathbf{A} . Generate random basis coefficients \mathbf{z} , and let k represent sample number. Thus, a random eight component quality vector \mathbf{x} is computed from

$$\begin{aligned}
 x_{k1} &= \sum z_{kj} a_{1j} + \varepsilon_{k1} \\
 x_{k2} &= \sum z_{kj} a_{2j} + \varepsilon_{k2} \\
 &\vdots \\
 x_{k8} &= \sum z_{kj} a_{8j} + \varepsilon_{k8}
 \end{aligned}
 \tag{3}$$

where \mathbf{z} are simulated data from a normal distribution with variances and means according to table 1. Error ε is random normal (0, 0.2). The variable j is incremented from 1 to 8 corresponding to the number of columns of \mathbf{A} . The first three cases in table 1 consider at least a single active process cause (corresponding to basis element No. 1). The next three consider at least three active process causes.

Since $\mathbf{z} = \mathbf{A}^{-1}\mathbf{x}$ gives eight estimated values for each sample, $\mathbf{Z} = [z_1 | z_2 | \dots | z_m]$, capability can be computed using \mathbf{Z} and the specification limits. Rectangular \mathbf{x} specifications also apply to $\mathbf{A}\mathbf{z}$ (since $\mathbf{x} = \mathbf{A}\mathbf{z}$). Since specifications for the x_i ’s are assumed equal to ± 2 units, the components of $\mathbf{A}\mathbf{z}$ and \mathbf{x} can be checked for conformance to specifications using a simple Monte Carlo approach. Using the

Table 1. How sample z data is generated by modifying $\mu_{\mathbf{Z}}$ and $\Sigma_{\mathbf{Z}}$ (datasets 1–6).

Dataset cases for computing \mathbf{Z} matrix capability	Diagonal variances for $\Sigma_{\mathbf{Z}}$
1. Base 1 ($\mu_{\mathbf{Z}} = \mathbf{0}$)	(1 ² , 0.05 ²)
2. Base 1 with z_1 mean shift = 0.5	(1 ² , 0.05 ²)
3. Base 1 with z_1 variance increase ($\mu_{\mathbf{Z}} = \mathbf{0}$)	(1.5 ² , 0.05 ²)
4. Base 2 ($\mu_{\mathbf{Z}} = \mathbf{0}$)	(1 ² , 1 ² , 1 ² , 0.05 ²)
5. Base 2 with z_1 mean shift = 0.5	(1 ² , 1 ² , 1 ² , 0.05 ²)
6. Base 2 with z_1 variance increase ($\mu_{\mathbf{Z}} = \mathbf{0}$)	(1.5 ² , 1 ² , 1 ² , 0.05 ²)

Table 2. Wierda's (1992) multivariate $\hat{\theta}$ as a capability measure: A comparison using capacitor chip simulated data. Bias corrected bootstrap confidence intervals are included for the estimated yield using the \mathbf{Z} matrix ($\hat{\theta}$ = estimated multivariate yield).

Dataset from table 1	Average actual yield $\hat{\theta}$ values for the \mathbf{X} matrix	Average estimated yield $\hat{\theta}$ values for the \mathbf{X} matrix	Average estimated yield \mathbf{Z} matrix values
1	0.92	0.69	0.92 (0.91, 0.93)
2	0.88	0.67	0.88 (0.87, 0.92)
3	0.76	0.50	0.76 (0.75, 0.78)
4	0.51	0.25	0.50 (0.45, 0.60)
5	0.48	0.25	0.47 (0.42, 0.56)
6	0.39	0.18	0.39 (0.35, 0.49)

data sets labelled 1–6, table 2 presents overall ‘percentage conforming’ capability values for the \mathbf{X} and \mathbf{Z} matrices.

Estimated capability values (in table 2) can be obtained using the process-oriented data or the ordinary data. Using a sample of 100 capacitor chip parts that is collected for a process-oriented multivariate capability study, we wish to generate multivariate normal random numbers having a means vector and variance-covariance structure similar to that of the sample. This is accomplished using the Cholesky decomposition method of multivariate number generation. Estimated results in this paper rely on estimating parameters (variance-covariance matrix and the means vector) from a sample of 100 to emulate an actual capability study. Next, the parameters are used to generate 10 000 multivariate random vectors that represent simulated parts that are compared to multivariate specifications for computing the estimated capability. The data needed for the estimated yield computations is computed using $\hat{\mathbf{Z}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{x}$ where a realistic error ε for each component of the \mathbf{x} vector is included. Actual yield is computed if the z 's are known, while error ε is zero. In practice, however, the z 's are unknown and therefore estimated. Estimated and actual results reported are average values from 1000 simulations.

The bootstrap is a re-sampling method that can be used to obtain confidence intervals for the estimated multivariate capability measure. The percentile method is a particular bootstrap method used to obtain the results in the last column of tables 2 and 4. See Efron (1982) for more information on the bootstrap method. Precedents for using bootstrap confidence intervals for multivariate capability include Littig *et al.* (1992) and Polansky (2001). Using bias corrected bootstrap intervals (also used in tables 2 and 4) is a method to fairly centre the bootstrap interval around the average (Gunter 1992) if its initial values are biased.

Supposing that a POBREP capability was not available, multivariate capability for \mathbf{X} would be computed. Table 2 shows that estimated $\hat{\theta}$ computed using \mathbf{Z} is almost identical to actual $\hat{\theta}$ computed when using \mathbf{X} . Furthermore, estimated yield for \mathbf{X} differs greatly from actual yield for \mathbf{X} (0.26 is the largest difference). Given that a capability study using 100 parts is assumed economically feasible, good estimates of multivariate yield may be attainable from the estimated \mathbf{Z} matrix results. When using Monte Carlo estimation, estimated multivariate capability based on original (\mathbf{X}) measurements is inaccurate for these examples (24–54% error).

Thus, a potential advantage of POBREP capability is that it accurately estimates capability using the same number of samples that would be used in an

ordinary capability study. Further, for many electronics industry parts, POBREP capability is feasible because the process-oriented causes can be assumed to operate independently. This results in a diagonal covariance matrix for the z 's, further reducing the number of parameters that must be estimated.

Capability can also be computed using less than a full basis. In the chip capacitor example, we can assume only two process-oriented variables. The intention is to show that capability continues to be effectively computed due to the assumption that the remaining basis elements capture the structure of the data while leaving no patterns in the data (Gonzalez-Barreto 1996). Before applying the capability method when the basis is less than full, we propose a few guidelines. Check the required condition of 'no patterns' left in the quality vectors statistically by plotting control charts for all the possible patterns (Barton and Gonzalez-Barreto 1996). When the plotted coefficients associated with potentially significant patterns show little variation or no trends relative to the control charts' centrelines, the pattern is not included since it does not account for much variability in the quality vectors. Additionally, as proposed by Barton and Gonzalez-Barreto (1996), use graphical displays such as the position dimension diagram (Seder 1950) to identify the significant POBREP patterns using. Finally, use process experts' input for deciding what basis elements would capture the significant variation in the data structure.

Tables 3 and 4 show how data was generated and the resulting multivariate capability if only two process-oriented variables capture variation in the quality

Table 3. Chip capacitor problem: using two basis elements (\mathbf{a}_1 and \mathbf{a}_8). How sample z data is generated by modifying μ_Z and Σ_Z (datasets 1–6).

Dataset cases for computing \mathbf{Z} matrix capability	Diagonal variances for Σ_Z
1. ($\mu_Z = \mathbf{0}$)	(1 ² , 0.05 ²)
2. z_1 mean shift = 0.5	(1 ² , 0.05 ²)
3. z_1 variance increase ($\mu_Z = \mathbf{0}$)	(1.5 ² , 0.05 ²)
4. ($\mu_Z = \mathbf{0}$)	(1 ² , 1 ²)
5. z_1 mean shift = 0.5	(1 ² , 1 ²)
6. z_1 variance increase ($\mu_Z = \mathbf{0}$)	(1.5 ² , 1 ²)

Table 4. Wierda's (1992) multivariate $\hat{\theta}$ as a capability measure: a comparison using simulated capacitor chip data and two basis elements (\mathbf{a}_1 and \mathbf{a}_8), ($\hat{\theta}$ = estimated multivariate yield).

Dataset from table 3	Average actual yield values for the \mathbf{X} matrix	Average estimated yield values for the \mathbf{X} matrix	Average estimated yield \mathbf{Z} matrix values (bias corrected bootstrap confidence intervals constructed from 1000 bootstraps using the percentile method)
1	0.95	0.69	0.95 (0.93, 0.96)
2	0.92	0.67	0.92 (0.88, 0.95)
3	0.81	0.50	0.82 (0.75, 0.85)
4	0.89	0.48	0.91 (0.87, 0.95)
5	0.86	0.47	0.88 (0.84, 0.93)
6	0.76	0.34	0.78 (0.72, 0.84)

vectors. Here, estimated capability using the process-oriented approach continues to yield values that are close to actual.

5. The solder paste POBREP problem

We further demonstrate the POBREP method for capability using the solder paste problem (see Gonzalez-Barreto and Barton 1995) from the electronics industry. Square fine pitch components have 208 leads total (52 per side) with corresponding solder paste deposit volumes. The multivariate measurement vector \mathbf{x} contains the errors from nominal for 208 deposit volumes measured by a vision system. Here, we also assume specifications are ± 2 units. Basis elements can be constructed that represent process failures (irregularities in volume deposits), and patterns can be drawn to visually represent the basis elements. Table 5 (Gonzalez-Barreto and Barton 1995) represents four basis elements, and figure 3 represents their basis element patterns. In figure 3 the dotted lines represent the desired level for all solder paste volumes, and the solid lines represent the actual levels of solder paste being below or above the desired levels for the 208 leads. Causes are proactively linked to each pattern allowing diagnosis of the root causes of irregular solder paste volumes.

The four basis elements in table 5 represent process variables that are monitored (rather than naively monitoring 208 leads individually).

To compute multivariate capability, we first simulate data using equation 5.1, where $j = 1$ to 4 patterns, and k is 1 to 100 samples. The error ε is simulated normal $(0, 0.2^2)$, and the \mathbf{a}_{1j} to \mathbf{a}_{208j} represent locations within the \mathbf{A} matrix.

$$\begin{aligned}
 x_{k1} &= \sum z_{kj} a_{1j} + \varepsilon_{k1} \\
 x_{k2} &= \sum z_{kj} a_{2j} + \varepsilon_{k2} \\
 &\vdots \\
 x_{k208} &= \sum z_{kj} a_{208j} + \varepsilon_{k208}
 \end{aligned}
 \tag{4}$$

Table 6 shows how the \mathbf{z} data are simulated from a multivariate normal distribution. Cases 1 and 2 consider a variance increase for z_1 , and the cases 3 and 4 consider variance increases for z_1 and z_2 , and a mean shift of the z_1 variable is considered.

Using the datasets labelled 1 to 6 from table 6, table 7 presents overall capability values (estimated from a study of 100 parts) for the \mathbf{X} matrix using 208 ordinary

Table 5. Vector representation of basis elements.

Position (lead)	Basis element			
	\mathbf{a}_1	\mathbf{a}_2	\mathbf{a}_3	\mathbf{a}_4
(1–52)	+1	+1	0	-1
(53–104)	+1	-1	+1	0
(105–156)	+1	+1	0	+1
(157–208)	+1	-1	-1	0

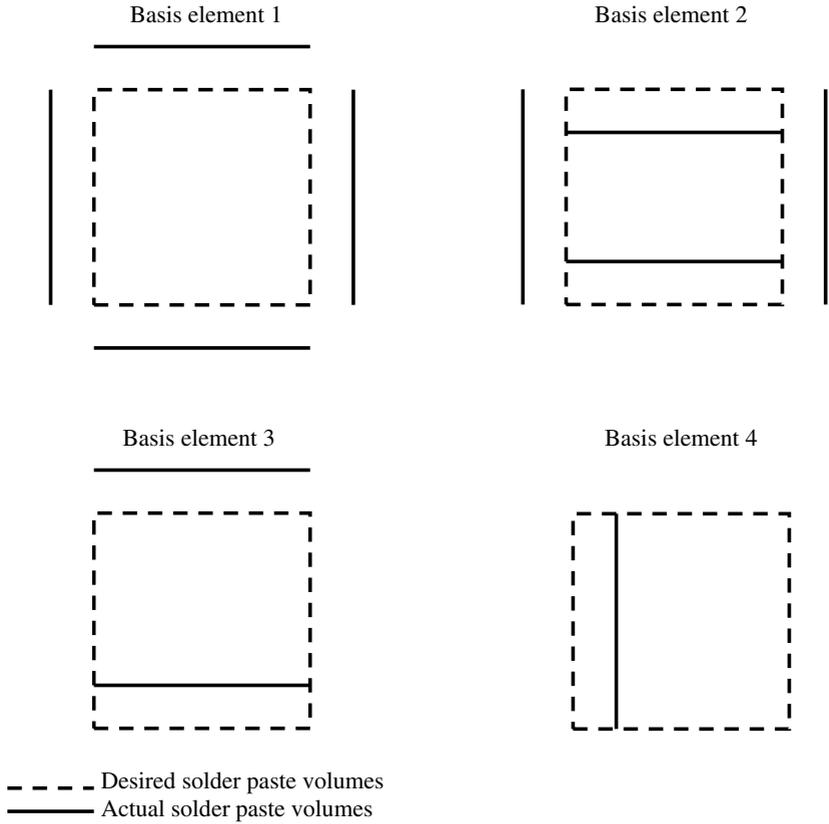


Figure 3. Four basis elements for fine pitch component.

Table 6. Solder paste example: How sample z data are generated.

Dataset cases for computing \mathbf{Z} matrix capability	Variances for $\Sigma_{\mathbf{Z}}$
1. z_1 mean shift = 1	$(0.9^2, 0.3^2, 0.3^2, 0.3^2)$
2. $(\mu_{\mathbf{Z}} = \mathbf{0})$	$(0.9^2, 0.3^2, 0.3^2, 0.3^2)$
3. z_1 mean shift = 1	$(0.9^2, 0.9^2, 0.3^2, 0.3^2)$
4. $(\mu_{\mathbf{Z}} = \mathbf{0})$	$(0.9^2, 0.9^2, 0.3^2, 0.3^2)$
5. z_1 mean shift = 1	$(0.3^2, 0.3^2, 0.3^2, 0.3^2)$
6. $(\mu_{\mathbf{Z}} = \mathbf{0})$	$(0.3^2, 0.3^2, 0.3^2, 0.3^2)$

variables and the \mathbf{Z} matrix using four process-oriented variables ($z_1, z_2, z_3,$ and z_4). It shows that Monte Carlo simulation cannot estimate capability using ordinary variables. In contrast, using the four process-oriented variables allows for estimating capability values that are close to actual capability.

We can repeat the capability study for the solder paste problem using only two process variables (3rd and 4th basis element in table 5) rather than four. Table 9 shows that the actual capability, when using table 8 datasets, is also close to the estimated capability when only two process variables are used.

Table 7. Capability for the solder paste example ($\hat{\theta}$ = estimated multivariate yield).

Dataset from table 6	Average actual yield $\hat{\theta}$ values for the X matrix	Average estimated yield $\hat{\theta}$ values for the X matrix	Average estimated yield Z matrix values (bootstrap confidence intervals constructed from 1000 bootstraps using the percentile method)
1	0.70	—	0.70 (0.62, 0.74)
2	0.89	—	0.89 (0.85, 0.93)
3	0.50	—	0.50 (0.42, 0.56)
4	0.69	—	0.69 (0.61, 0.74)
5	0.90	—	0.90 (0.84, 0.93)
6	1.00	—	1.00 (1.00, 1.00)

Table 8. Solder paste example datasets using basis elements 3 and 4.

Dataset cases for computing Z matrix capability	Variances for Σ_Z
1. z_1 mean shift = 1	$(0.9^2, 0.3^2)$
2. $(\mu_Z = \mathbf{0})$	$(0.9^2, 0.3^2)$
3. z_1 mean shift = 1	$(0.9^2, 0.9^2)$
4. $(\mu_Z = \mathbf{0})$	$(0.9^2, 0.9^2)$
5. z_1 mean shift = 1	$(0.3^2, 0.3^2)$

Table 9. Capability for the solder paste example with two process variables (a_3 and a_4): ($\hat{\theta}$ = estimated multivariate yield).

Dataset from table 8	Average actual yield $\hat{\theta}$ values for the X matrix	Average estimated yield $\hat{\theta}$ values for the X matrix	Average estimated yield Z matrix $\hat{\theta}$ values (bootstrap confidence intervals constructed from 1000 bootstraps using the percentile method)
1	0.87	—	0.87 (0.81, 0.92)
2	0.97	—	0.97 (0.94, 0.99)
3	0.84	—	0.84 (0.80, 0.90)
4	0.95	—	0.95 (0.91, 0.97)
5	1.00	—	1.00 (1.00, 1.00)

Using the ordinary variables to determine capability using Wierda’s (1992) percentage conforming approach is not feasible due to the ‘curse of dimensionality’. To obtain accurate estimates for higher dimensional multivariate problems, a larger sample size is required that increases exponentially with the number of variables. However, using the smaller process variable set rather than ordinary variables allows an accurate POBREP capability estimation.

Meeting the assumption of multivariate normal distribution is even more important for the multivariate indices relative to univariate indices. The z coefficients

will tend to collectively form normal distributions due to the central limit theorem (CLT) effect under certain circumstances. That is, when the columns of the basis matrix have four or more non-zero numbers, the z coefficients, which would then be linear combinations of four numbers, are approximately normal. This principle is based on studies done for X-bar charts (Rigdon *et al.* 1994) that showed that X-bar charts require normal distributions for the proper percentage of parts to fall outside of their control limits. An even stronger CLT effect results when greater numbers of non-zero elements are in the basis matrix (solder paste example). There is, however, no CLT effect for non-POBREP multivariate or univariate capability. Many POBREP applications will fit the orthogonal (or independent) case because deliberate steps are taken to construct an orthogonal basis matrix, and we recommend the use of process-oriented multivariate capability for these cases.

When a POBREP application does not fit the orthogonal case, POBREP coefficients could be unstable, have the wrong size, or even the wrong sign. This would create an inaccurate multivariate process-oriented capability measure. Hoerl and Kennard (1970) proposed a remedy termed ridge regression for when the prediction vectors (equivalent to the $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ vectors) are not orthogonal. The method adds small positive quantities to the diagonal of $\mathbf{A}'\mathbf{A}$, and the ridge trace (a two-dimensional plot) would show the resulting effects on non-orthogonality. Let \mathbf{z}_h^* equal the ridge regression estimated z parameters. Then

$$\mathbf{z}_h^* = [\mathbf{A}'\mathbf{A} + c\mathbf{I}]^{-1}\mathbf{A}'x.$$

When $c=0$, the component values of the vector \mathbf{z}_h^* are equal to the ordinary least squares regression z coefficients. The issue of what c to choose is a judgmental one (Neter *et al.* 1996) and could be based on when the variance inflation factor (VIF), a measure of the dependence of the columns, first becomes acceptable. The $\text{VIF}_n = (1 - R_n^2)^{-1}$ where $n=1, 2, \dots, p$, and R_n^2 is the coefficient of multiple determination when \mathbf{a}_n is regressed on the other $p-1$ \mathbf{a} variables. VIF values over 5 indicate a situation of moderate to extreme nonorthogonality.

6. Conclusions

Multivariate capability using process-oriented basis representations is a new method for measuring process performance. For many POBREP applications, we can assume independent data by using the process-oriented variables' coefficients when the original data exhibits dependence. The current multivariate capability methods do not provide a direct link to the causes of poor process performance and are not always practical for processes that collect large amounts of multivariate data. More research is needed to determine POBREP capability when non-orthogonal basis elements yield significantly correlated basis coefficients.

Acknowledgements

This work was supported in part by NSF GOALI grant DMI 0084909 and by a grant from the General Motors Corporation. Figures 1 and 2 are courtesy of Quality Engineering.

References

- Albin, S.L., King, L. and Gerald, S., An X and EWMA chart for individual observations. *J. Qual. Technol.*, 1997, **29**, 41–48.
- Barton, R.R. and Gonzalez-Barreto, D.R., Process-oriented basis representation for multivariate process diagnostics. *Qual. Eng.*, 1996, **9**, 107–118.
- Barton, R.R. and Gonzalez-Barreto, D.R., Process-oriented basis representations: linking manufacturing process design and diagnosis. *Proc. Euro. Conf. Con. Eng.*, 1999, **9**, 109–114.
- Birgoren, B., Multivariate statistical process control for quality diagnostics and applications to process-oriented basis representations. Ph.D. Dissertation, 1998 (Pennsylvania State University).
- Chen, H., A multivariate process capability index over a rectangular solid zone. *Statistica Sinica*, 1994, **4**, 749–758.
- Devor, R.E., Chang, T. and Sutherland, J.W., *Statistical Design and Control*, 1992 (Macmillan Publishing Company: New York).
- Efron, B., The jack knife, the bootstrap and other resampling plans, 1982, CBMS-NSF Regional Conference Series in Applied Mathematics.
- Gonzalez-Barreto, D.R., Process-oriented basis representation for multivariate process diagnostics and control. Ph.D. Dissertation, 1996 (Pennsylvania State University).
- Gonzalez-Barreto, D.R. and Barton, R.R., 1995, Process-oriented basis representations for statistical process control (SPC). In *Proceedings of the 4th Industrial Engineering Research Conference*, Institute of Industrial Engineers, Nashville, TN, pp. 759–768.
- Gunter, B.H., How to make something from almost nothing and get statistically valid answers—Part 4: Bias-corrected intervals. *Qual. Prog.*, 1992, **25**, 79–83.
- Hotelling, H., Multivariate quality control. In *Techniques of Statistical Analysis*, edited by C. Eisenhart, M.W. Hastny, and W.A. Wallis, 1947 (McGraw-Hill: New York).
- Juran, J.M., Gryna, F.M. and Bingham, R.S. Jr, *Quality Control Handbook*, 1974 (McGraw-Hill: New York).
- Kane, V.E., Process capability indices. *J. Qual. Tech.*, 1986, **18**, 41–52.
- Kotz, S. and Johnson, N.L., *Process Capability Indices*, 1993 (Chapman & Hall: London).
- Kotz, S. and Lovelace, C., *Process Capability Indices in Theory and Practice*, 1998 (Arnold: London).
- Littig, S.J., Lam, C.T. and Pollock, S.M., Capability measurements for multivariate processes: definition and an example for a gear carrier. Technical Report, 92-42, 1992, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI.
- Mitra, A., *Quality Control and Improvement*, 1998 (Prentice-Hall: New Jersey).
- Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W., *Applied Linear Statistical Models* (4th edition), 1996 (Richard D. Irwin: Burr Ridge, IL).
- Polansky, A.M., A smooth nonparametric approach to multivariate process capability. *Technometrics*, 2001, **43**, 199–211.
- Rigdon, S.E., Cruthis, E.N. and Champ, C.W., Design strategies for individuals and moving range control charts. *J. Qual. Tech.*, 1994, **26**, 274–289.
- Seder, L.A., Diagnosis with diagrams. Part II, *Indust. Qual. Cont.*, 1950, **6**, 7–10.
- Shahriari, H., Hubele, N.F. and Lawrence, F.P., A multivariate process capability vector. In *Proceedings IIE 4th Industrial Engineering Research Conference*, 1995, pp. 303–308.
- Taam, W., Subbaiah, P. and Liddy, J.W., A note on multivariate capability indices. *J. Appl. Stat.*, 1993, **20**, 339–351.
- Wang, F.K., Hubele, N.F., Lawrence, F.P., Miskulin, J.D. and Shahriari, H., Comparison of three multivariate process capability indices, *J. Qual. Tech.*, 2000, **32**, 263–275.
- Wierda, S.J., Multivariate quality control—estimation of the percentage of good products. Technical Report, 1992, Department of Econometrics, University of Groningen, The Netherlands.
- Wierda, S.J., A multivariate process capability index, in *ASQC Quality Congress Transactions, Boston*, 1993, pp. 343–348.
- Wood, M., Three suggestions for improving control charting procedures. *Int. J. Qual. Reliab. Manage.*, 1995, **12**, 61–74.