Computing Method of Multivariate Process Capability Index Based on Normalized Pretreatment

Guangqi Ying¹, Yan Ran¹, Genbao Zhang, Yuxin Liu, Shengyong Zhang

State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing, China 400030

¹Corresponding Author: Yan Ran, Chongqing University, Chongqing, ranyan@cqu.edu.cn.

1. Introduction

Process capability is the consistency index of machining process. In order to measure, evaluate, analyze, and compare process capabilities, quantifiable indicators as Process Capability Index (PCI) are practically used. The concept of process capability ratio are firstly proposed by Juran[1], who compared process fluctuations with process specification as a quantitative index to evaluate process capability. Other quantifiable indicators that have emerged since then are based on process fluctuations and specifications, such as $C_p$, $C_{pk}$ and $C_{pm}$. With the improvement of processing level and increasingly complex parts, the processing parts generally have multiple quality characteristics, and the process capability evaluation needs to consider multiple quality characteristics. Since the concept of a multi-process capability index are proposed by Chan L K et al.[2], Process Capability Index multivariate analysis model was constructed as the following categories: the multi-process capability index (MPCI) was constructed by region ratio, such as the method for evaluating the process area that are presented by Shahriari et al.[3], and the spatial MPCI proposed by Wang S. X et al.[4] to solve the eccentricity of multivariate quality features. MPCI was constructed by a failure rate, such as Wierda et al.[5] Bothe et al.[6] and Chen K S et al.[7]. MPCI was constructed by principal component analysis, such as the process capability index of principal components proposed by WangKF and Chen JC[8], Wang FK and Du TC[9], Ma Yizhong[10] and Wang Rang CH[11] to build multi-quality characteristics index. And based on the correlation of the specification intervals, Shinde et al.[12] and Zhang M et al.[13] modified the specification interval. MPCI is constructed by vector method, such as a new multivariate capability vector that are proposed by Shahriari and Abdollahzadeh[14]. In addition to the methods above, there are another method. Pan[15] proposed two new multi-process capability indices: $NMC_p$ and $NMC_{pm}$.

In the traditional Principal Component Analysis (PCA) constructing the Multivariate Process Capability Index (MPCI), a few principal components mainly consider process variation but not process capability, which is unreasonable. Therefore, aiming at the rationality of principal components degree, an evaluation method for pre-processing data before constructing MPCI using PCA is proposed.

2. Principal Component Analysis theory

When the mass data of multi-quality feature parts obeys the multivariate normal distribution, the principal component analysis is proposed by WangKF and Chen JC[14] to construct the multivariate process capability index. They extended the one-unit capability index to pluralism and defined the multivariate process capability indices: $MC_p$, $MC_{pk}$, $MC_{pm}$ and $MC_{pmk}$.

When using the principal component analysis method, a small number of principal components can explain about 90% of the process variation. In terms of construction methods, the method of weighted averaging is used by Ma Yizhong[10]. Ac-
cording to the fluctuations, for each principal component index, different weights are assigned. Take \( C_p \) as an example:

\[ M_{Cp} = \sum_{i=1}^{r} C_p; PC_i \tag{1} \]

And Wang F K and Du T C\(^{[9]}\) and Wang Rang C H\(^{[1]}\) use geometric mean, namely:

\[ M_{Cp} = \left( \prod_{i=1}^{v} C_p; PC_i \right)^{1/v} \tag{2} \]

Where

\[ C_p; PC_i = \frac{USL_{PCi} - LSL_{PCi}}{6\sigma_{PCi}} \]

Where, \( r_i = \frac{1}{r} \cdot \text{tr}(\Sigma), \text{tr}(\Sigma) = \lambda_1 + \lambda_2 + \ldots + \lambda_r \), here, \( \lambda_1, \lambda_2, \ldots, \lambda_r \) is the feature vector of \( \Sigma \), \( C_p; PC_i \) is the single process capability index of the \( i \)-th principal component, and \( V \) represents the number of principal components which can explain about 90% of the process variation. The specification range and target value of the principal component are:

\[ LSL_{PCi} = u_i' LSL \quad USL_{PCi} = u_i' USL \tag{3} \]

Where, \( u_1, u_2, \ldots, u_p \) is the feature vector of \( \Sigma \).

Shinde and Khadse\(^{[12]}\) pointed out that the formulas calculated by Wang K F et al.\(^{[8-9]}\) are biased, because in principal component analysis, only the distribution of principal components is independent, the canonical intervals of different principal components are still interrelated, and their viewpoints are verified by examples. At the same time, new specification intervals and multi-process capability indices are defined.

\[ V = \{ (y_1, y_2, \ldots, y_r) | LSL_X \leq Uy \leq USL_X \}, \]

\[ y' = (y_1, y_2, \ldots, y_p), y_r = EY_r \]

Where, \( y \) is the observations of the principal component (PC), \( V \) is not a super-rectangular, it is composed of 2p linear inequalities (p is the number of quality features).

\[ M_{P1} = P \left\{ \left[ Y = (Y_1, Y_2, \ldots, Y_r) \right] \in V' \right\} \]

\[ y \sim N_r(\mu_y, \Sigma_y) = \text{diag}(\lambda_1, \ldots, \lambda_r) \tag{5} \]

\[ M_{P2} = P \left\{ \left[ Y = (Y_1, Y_2, \ldots, Y_r) \right] \in V' \right\} \]

\[ y \sim N_r(\mu_y, \Sigma_y) = \text{diag}(\lambda_1, \ldots, \lambda_r) \tag{6} \]

\( y \) represents the vector of PC, \( M_{P1} \) is similar to \( M_{Cp} \), \( M_{P2} \) is similar to \( M_{Cpk} \). If \( M_{P1} \geq 0.9973 \), it indicates that this process has potential capability. If \( M_{P2} \geq 0.9973 \), it indicates that this process has actual capability.

3. The principal component contribution rate

The contribution rate of the principal component was determined by the variance ratio of the quality feature and the correlation coefficient. For convenience of explanation, the following two quality characteristics are taken as examples to describe and calculate the principle component contribution rate.

3.1 Theoretical analysis of the influencing factors of the principal component contribution rate

It is known from Figure 1 that when the correlation coefficients of quality control features are same, the contribution rate of each principal component is only affected by the variance difference of each quality feature. Moreover, the higher the variance of quality characteristics, the higher the contribution rate of the first principal component. From the comparison of the three curves in Figure 1, it is found that with the increase of the variance ratio of the quality features, the correlation coefficient between quality characteristics has less influence on the contribution rate of the first principal component. When the standard deviation ratio of quality characteristics is more than 9 times, the contribution rate of the first principal component is independent from the correlation coefficient of the quality feature. What’s more, it is only determined by the ratio of variance between quality characteristics. At this time, the contribution rate of the first principal component is approximately equal to the proportion of the maximum variance to the total variance.

![Figure 1. Variance ratio of quality characteristic influence on principal component contribution degree](image)

As is shown in Figure 2, when the variances of the control quality characteristics are same, the contribution rate of each principal component is affected by the correlation coefficient between the various quality features, which indicates that the greater the correlation coefficient between the quality features, the greater the contribution rate of the first principal component.

From the theoretical analysis above, during analyzing the covariance matrix based on quality features, the contribution rate of principal components is determined by the variance ratio and the correlation coefficient between quality features. However, during calculating the multivariate process capability index, multivariate process capability index is not only related to the variance and the correlation coefficient of quality characteristics, but also to the quality characteristic interval. Therefore, it is not advisable and reasonable to use the contribution rate of principal components to represent and evaluate the contribution rate of multivariate process capability.

3.2 Application of the principal component contribution rate

The gearbox base used in this paper has two quality characteristics, the width \( X_1 \) (60mm) and the height \( X_2 \) (90mm). \( S = \{ (X_1, X_2) | 57 \leq X_1 \leq 63, 89.4 \leq X_2 \leq 90.6 \} \) is the specification range. We collected 100 sets of sample data of this part. The sample mean vector and the covariance matrix are obtained as:
Fig 2. Correlation coefficient of quality characteristic influence on principal component contribution degree

Through principal component analysis based on covariance matrix, the principal component contribution rates of each quality characteristic are obtained as:

\[ PC_1 = u_1 X = X_1 - N(60.00, 0.09) \]
\[ PC_2 = u_2 X = X_2 \]

Figure 3 Principal component process capability

As shown in Figure 3, \( U = (u_1, u_2) \) is the eigenvector of \( S \). The ellipse represents the area of 95% process capability, and the rectangular refers to the specification interval.

Obviously, the process capability of the first principal component is greater than that of the second principal component, which indicated that using the first principal component to calculate the multivariate process capability will inevitably, overestimated the multivariate process capability. Because 91.7% of the first principal component contribution rate only explains 91.7% of the variation but not the multivariate process capability. It is ignoring the influence of the specification interval on the contribution rate of the multi-process capability, that leads to the over-estimation of the multi-process capability.

As illustrated by the above example, when calculating the multivariate process capability index, using the principal component contribution rate to evaluate the multivariate process capability directly is not precise. To solve this problem, we proposed a new method based on processing raw data.

4. MPCI with improved principal component contribution rate

Different from the traditional principal component analysis method, MPCI with improved principal component contribution rate needs to preprocess the quality features to standardize the specification intervals of each quality feature.

4.1 Data preprocessing

Interval standardization:

\[ H_i = \frac{(X_i - M_i)}{d_i} \]

where \( X_i \) is the original quality feature, \( H_i \) is the processed quality feature, \( M_i \) is the center of the \( X_i \) specification interval, and \( d_i \) is half of the interval between the upper and lower specification boundaries of \( X_i \):

\[ M_i = \frac{(USL_i + LSL_i)}{2} \]
\[ d_i = \frac{(USL_i - LSL_i)}{2} \]

Through processing the original data, the specification intervals of all quality features are normalized to [-1, 1]. Thus, in the process of calculating the multivariate process capability index with the new data meter, the multivariate process capability index is only determined by the variance and correlation coefficient of the quality features.

In the process of principal component analysis of new data, because the principal component process with a residual contribution rate of less than 10% is very strong, these principal components can be ignored, and the total component is explained by a small part of the principal component whose cumulative contribution rate reaches 90% or more.

4.2 Multivariate process capability index calculation

For a part with \( p \) feature qualities \( X_1, X_2, ..., X_p \), its feature quality vector is \( X' = [X_1, X_2, ..., X_p] \), each of which has \( n \) sample points or observation data, and obeys \( X \leq N_p(\mu, \Sigma) \). \( \Sigma \) is the covariance matrix. Symbols used are defined as:

\[ T = (T_1, T_2, ..., T_p) \]: \( X \) target value vector

\[ LSL_X = (LSL_{X_1}, LSL_{X_2}, ..., LSL_{X_p}) \]: Lower specification limit of \( X \)

\[ USL_X = (USL_{X_1}, USL_{X_2}, ..., USL_{X_p}) \]: Upper specification limit of \( X \)

\[ S = \{ x | LSL_X \leq x \leq USL_X \} \]: \( X \)’s super-rectangular specification interval

Quality characteristics \( H \) centralized treatment:

\[ Z_i = H_i \cdot \mu_{Z_i}, \quad i = 1, 2, ..., p \]
\[ \Sigma_z (X, X) = \text{Cov}(X, X) \quad \text{d.d.} \]
\[ \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_p \]: Characteristic value of \( \Sigma_z \)
\[ u_1, u_2, ..., u_p \]: Characteristic vector of \( \Sigma_z \)

Principal component vector:

\[ Y = U'Z \]

Where, \( U = (u_1, u_2, ..., u_p) \), \( E(Y) = 0 \).

Covariance of principal components:

\[ COV(Y) = \text{Diag}(\lambda_1, \lambda_2, ..., \lambda_p) \]

Covariance of principal components:
Preprocess $X$ according to equations (7) and (8):

$$H_t = \{X_i - M_t\} / d_t, \quad Z_t = H_t - \mu_{Z_t}, \quad i = 1, 2, ..., p$$

Calculate the $Z$-means vector $\overline{Z}$ and the covariance matrix $S_z$ according to equation (9):

$$\overline{Z} = \begin{bmatrix} 0.0424133 \\ 0.044994 \\ 0.026843 \\ 0.086942 \\ -0.002168 \\ 0.002034 \\ -0.001018 \\ 0.110592 \end{bmatrix}$$

$$S_z = \begin{bmatrix} 0.0424133 & 0.044994 & 0.026843 & 0.086942 \\ 0.044994 & 0.026843 & 0.086942 & -0.002168 \\ 0.026843 & 0.086942 & -0.002168 & 0.110592 \end{bmatrix}$$

5. Application

Take the gearbox housing produced by a domestic manufacturer as an example. The proposed multi-process capability index (MPCI) is used for the capability analysis of the gearbox housing. In the manufacturing process, the specification intervals of the quality characteristics $X_1$, $X_2$, $X_3$, and $X_4$ respectively are [-0.1, 0.1], [-0.035, 0.035], [-0.018, 0.004], and [-0.021, 0.004], and the target values are 0, 0.0104, and -0.0085, respectively.

From the production process, 50 observation samples were randomly selected and according to the method of PAN et al.

The Shapiro-Wilk statistic was calculated to check whether the sampled data follows the multivariate normal distribution. The Shapiro-Wilk statistic was calculated to check whether the sampled data is considered to follow a multivariate normal distribution. Therefore, the sampled data is considered to follow a multivariate normal distribution.

Calculate the sample mean vector $\overline{X}$ and the covariance matrix $S$:

$$\overline{X} = \begin{bmatrix} 0.0084141 & 0.000281 \\ -0.007473 & -0.00821 \end{bmatrix}$$

$$S = \begin{bmatrix} 0.0004213 & 0.00014698 & 0.00008388 \\ 0.00005425 & 0.00002518 & 0.00001052 \\ -0.00000271 & 0.00000089 & -0.00000014 \\ 0.00001728 \end{bmatrix}$$

5.1 MPCI with improved principal component contribution rate

Preprocess $X$ according to equations (7) and (8):

$$H_t = \{X_t - M_t\} / d_t, \quad Z_t = H_t - \mu_{Z_t}, \quad i = 1, 2, ..., p$$

Calculate the $Z$-means vector $\overline{Z}$ and the covariance matrix $S_z$ according to equation (9):

$$\overline{Z} = \begin{bmatrix} -0.1 \pm 0.00342 \\ 0.1 \\ -0.035 \pm 0.00298 \\ 0.035 \\ -0.018 \pm 0.00826 \\ 0.011 \\ -0.021 \pm 0.00889 \\ 0.0125 \end{bmatrix}$$

As shown in Figure 4, the common area enclosed by 8 lines is the revised specification area:

$$R = \begin{bmatrix} 0.5883Y_1 + 0.01627Y_2 \leq 0.1 + 0.00342 \\ 1 \\ -0.035 - 0.00298 \leq 0.5883Y_1 + 0.01627Y_2 \leq 0.035 + 0.00298 \\ -0.018 + 0.00826 \leq 0.6800Y_1 + 0.01627Y_2 \leq 0.011 + 0.00826 \\ 0.011 \pm 0.000889 \\ -0.0125 \pm 0.000889 \end{bmatrix}$$

According to equation (11), (12) and (13), the target values of principal component covariance and $Y$ are calculated as follow:

$$\Sigma_y = \begin{bmatrix} 0.01753 \\ 0.035 - 0.00298 \\ 0.011 \\ 0.0125 \end{bmatrix}$$

$$\Sigma_y = \begin{bmatrix} 0.0422 \\ -0.0283 \end{bmatrix}$$

Calculate the probability that the principal component satisfies the specification region according to equations (15) and (17)}
P1 = 0.9965  P2 = 0.9962  

MCP is calculated according to equation (17) and (18):  

MCP1 = 0.9748  
MCpk1 = 0.9651  

The target values of principal component covariance and y':  

\[ \Sigma_y' = \begin{pmatrix} 0.0004839 & 0 \\ 0 & 0.0000299 \end{pmatrix} \]  

\[ T_y' = (-0.0023, 0.0037) \]  

The probability that the principal component satisfies the specification region:  

\[ P_1' = 0.99939 \]  
\[ P_2' = 0.99936 \]  

The MCPI:  

MCPc = 1.1428  
MCpk = 1.139  

5.2 Traditional principal component method of MPCI  

Calculate the principal component based on the covariance:  

\[ \lambda_1 = 0.0004839 \]  
\[ PC_1 = Y_1' = -0.9287Z_1' - 0.3491Z_2' - 0.1250Z_3' + 0.0048Z_4' \]  
\[ \lambda_2 = 0.0000299 \]  
\[ PC_2 = Y_2' = 0.3622Z_1' - 0.9068Z_2' - 0.1637Z_3' - 0.1404Z_4' \]  
\[ \lambda_3 = 0.0000170 \]  
\[ PC_3 = Y_3' = 0.0564Z_1' - 0.1247Z_2' - 0.0333Z_3' + 0.9900Z_4' \]  
\[ \lambda_4 = 0.0000022 \]  
\[ PC_4 = Y_4' = 0.0562Z_1' + 0.2006Z_2' - 0.9780Z_3' - 0.0109Z_4' \]  

Principal components \( Y_p, Y_p', Y_p, \) and \( Y_p, \) are independent from each other and subject to normal distribution, and their parameters are \( \mu, \sigma^2 = (0, 0.00004839), (0, 0.00000299, 0.000000170), \) and \( (0, 0.0000022, 0.0000002) \), respectively. The target vector of the principal component is \((-0.0023, 0.0037, 0.0009, -0.0016)\). The first two principal components explain the overall difference of 96.4%, so the first two principal components were used to analyze the multivariate process capability.

As shown in Figure 5, the common area surrounded by 8 lines is the revised specification area:

\[ R' = \begin{pmatrix} 0.09658 & -0.9287 \leq Y_1' \leq 0.3622Y_2' \leq 0.10342 \\ -0.03798 \leq -0.3491Y_1' \leq -0.9068Y_2' \leq 0.03202 \\ -0.00974 \leq -0.1250Y_1' - 0.1637Y_2' \leq 0.0048 \end{pmatrix} \]  

The multi-process capability index should be less than the process capability index of any single quality feature. The potential process capability indices of the individual quality features \( X_1, X_2, X_3, \) and \( X_4 \) are calculated to be 1.62, 1.27, 1.13, and 1.002, so the multivariate process capability index should be less than 1.002.

In the example above, the potential MPCI with improved principal component contribution rate is 0.9748, and the potential MPCI of the traditional principal component method is 1.1428. The MPCI with improved principal component contribution rate is more realistic than the traditional principal component method. The analysis found that in the MPCI of the traditional principal component method, the variance of the quality feature \( X_4 \) is smaller and the correlation is weaker than other quality features, so the contribution rate of the quality feature \( X_4 \) is small and ignored (in the second section, it has been derived that the smaller the variance ratio and the weaker the contribution of the quality features, the smaller the contribution rate), resulting in MPCI being overestimated.

6. Conclusion  

1) Through the theoretical analysis of the influence factors of the contribution rate of principal components, we found that the principal component contribution rate is determined only by the variance ratio and correlation coefficient between the quality features. Moreover, when one of them is constant, the contribution rate of the first principal component is positively linearly relative to another.  

2) Based on the principal component method, we found.
that the principal component contribution rate is not equivalent to the contribution rate of multivariate process capability practically. So, using the principal component contribution rate to directly evaluate the multivariate process capability is not reasonable or precise.

3) To solve the problem mentioned above, through normalizing the specification range of the processed quality features to [-1,1], eliminating the principal components with smaller variance to simplify the data. We proposed a method based on processing original data. Moreover, the rationality of the method is proved by an example.

References


