Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1260

Published online in Wiley Online Library

On the Implementation of the Principal Component Analysis–Based Approach in Measuring Process Capability

M. Perakis[a*†] and E. Xekalaki[b]

The use of principal component analysis in measuring the capability of a multivariate process is an issue initially considered by Wang and Chen (1998). In this article, we extend their initial idea by proposing new indices that can be used in situations where the specification limits of the multivariate process are unilateral. Moreover, some new indices for multivariate processes are suggested. These indices have been developed so as to take into account the proportion of variance explained by each principal component, thus making the measurement of process capability more effective. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: process capability indices; multivariate processes; principal component analysis

1. Introduction

Process capability indices aim at measuring the capability of a process to produce according to some assigned specifications related to some characteristics of the items produced. In most of the cases, it is assumed that the studied process is described through only one characteristic. However, sometimes the need arises for measuring the capability of processes described through two or more characteristics, that is, multivariate processes.

Among the suggested approaches for measuring the capability of multivariate processes one may note are those by Chan *et al.*[1], Pearn *et al.*[2], Taam *et al.*[3], Chen[4], Shahriari *et al.*[5], Bernardo and Irony[6], Wang and Chen[7], Wang and Du[8], Castagliola and Castellanos[9], and Shinde and Khadse[10]. The reviews of the abovementioned and some other proposals for capability indices for multivariate processes are provided by Kotz and Johnson[11,12], Kotz and Lovelace[13], Pearn and Kotz[14], and Wu *et al.*[15], whereas Wang *et al.*[16] and Zahid and Sultana[17] attempt a comparison of some of the suggested approaches. Furthermore, Spiring *et al.*[18] and Yum and Kim[19] provide an extensive coverage of the bibliography on process capability indices, inclusive of all the important contributions to the measurement of the capability of multivariate processes.

In this article, we consider the use of the principal component analysis in measuring the capability of a multivariate process. The use of this multivariate technique concerning process capability was initially considered by Wang and Chen[7]. The idea is to base the assessment of the capability of a process only on those of the first principal components that explain a large percentage of the total variability, as the remaining ones carry negligible information. In the sequel, on the basis of Xekalaki and Perakis[20], some alternative ways of implementing this technique are proposed through the definition of certain new indices.

In the remaining of this section a brief description of the idea of principal component analysis is given, mainly for the purpose of introducing notation and terminology that is used in the subsequent sections, and a presentation of the indices proposed by Wang and Chen[7] is made. In Section 2, extending Wang and Chen's[7] approach, some new indices are suggested for multivariate processes with unilateral specification limits. In Section 3, some new indices are proposed for multivariate processes with unilateral or bilateral specifications. These indices take into account the possibly unequal contributions of the individual principal components to the total variability. In Section 4, we present the results of a study that we performed, aiming at examining the properties of the suggested indices in various cases. Finally, some conclusions concerning the suggested indices are pointed out in Section 5.

A process that is described through r measurable characteristics is represented in the sequel by a random vector $\mathbf{X}=(X_1,\ldots,X_r)'$, with the ith element corresponding to the ith characteristic. The corresponding mean vector, variance–covariance matrix, and correlation matrix are denoted by $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\rho}$, respectively. When the elements of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\rho}$ are unknown, they can be estimated by $\overline{\mathbf{X}}$, $\boldsymbol{S}$,

[a]*General Accounting Office, Athens, Greece*

[b]*Department of Statistics, Athens University of Economics and Business, Athens, Greece*

**Correspondence to: Michael Perakis, General Accounting Office, 57 Panepistimiou Str., 105 64, Athens, Greece.*

[†]*E-mail: mperakis@gmail.com*

and **R**, respectively. The specifications of $\boldsymbol{X}$ are determined through the vector $\mathbf{l}=(L_1,\ldots,L_r)'$ of the lower specification limits for the r variables, the vector $\mathbf{u}=(U_1,\ldots,U_r)'$ of the upper specification limits, and the vector $\mathbf{t}=(T_1,\ldots,T_r)'$ of the target values.

The implementation of principal component analysis can be based either on $\boldsymbol{\Sigma}$ or on $\boldsymbol{\rho}$ (see, for example Mardia *et al.*[21] and Morisson[22]). If the r variables are expressed in similar measurement units, $\boldsymbol{\Sigma}$ has to be preferred, whereas if substantial differences in the measurement units are present, $\boldsymbol{\rho}$ is preferable.

If $\boldsymbol{\Sigma}$ is used, then denoting the r eigenvalues of $\boldsymbol{\Sigma}$ ordered in descending order by $\lambda_1 \geqslant \ldots \geqslant \lambda_r$ and the corresponding normalized eigenvectors by $\mathbf{u_1},\ldots,\mathbf{u_r}$, the r principal components are given by $Y_i = u_{1i}X_1 + \cdots + u_{ri}X_r$, $i = 1,\ldots,r$. When the matrix $\boldsymbol{\Sigma}$ is unknown, $\boldsymbol{S}$ is used whence the resulting estimates of the eigenvalues are denoted by $\hat{\lambda}_1 \geqslant \ldots \geqslant \hat{\lambda}_r$ and the estimates of the eigenvectors are denoted by $\hat{\mathbf{u}}_1,\ldots,\hat{\mathbf{u}}_r$. If the correlation matrix $\boldsymbol{\rho}$ is used, the r principal components obtained are $Y_i = u_{1i}\,Z_1 + \ldots + u_{ri}\,Z_r$, $i = 1\ldots\ r$. Here, $Z_i = (X_i - \mu_i)/\sigma_i$ and μ_i, σ_i denote the mean and standard deviation of the ith initial variable. Again, if μ_i, σ_i, λ_i, and $\mathbf{u_i}$ are unknown, they are replaced by their sample counterparts.

The choice of the number of principal components that should be included in the analysis can be determined either through some "rules of thumb" or via some tests. Jackson[23] and Wang and Chen[7] suggest the use of a test proposed by Anderson[24], which enables one to test the null hypothesis H_0: $\lambda_{q+1} = \ldots = \lambda_r$, against the alternative hypothesis that some of these eigenvalues are different. The test is sequentially applied for $q = 0,\ldots,r-2$, until the first time the null hypothesis is not rejected.

Wang and Chen[7] considered the capability assessment of a process using the principle components extracted from the r initial variables on the basis of the following indices

$$\mathrm{MC}_p = \left(\prod_{i=1}^{r} C_p^{Y_i}\right)^{\frac{1}{r}}, \mathrm{MC}_{pk} = \left(\prod_{i=1}^{r} C_{pk}^{Y_i}\right)^{\frac{1}{r}}, \mathrm{MC}_{pm} = \left(\prod_{i=1}^{r} C_{pm}^{Y_i}\right)^{\frac{1}{r}}, \textit{and}\ \ \mathrm{MC}_{pmk} = \left(\prod_{i=1}^{r} C_{pmk}^{Y_i}\right)^{\frac{1}{r}}$$

The use of these indices is meaningful when the distribution of the studied process is the multivariate normal, and thus the principal components are independent and normally distributed. Here, $C_p^{Y_i}$, $C_{pk}^{Y_i}$, $C_{pm}^{Y_i}$, and $C_{pmk}^{Y_i}$ denote the indices C_p, C_{pk}, C_{pm}, and C_{pmk} for the ith principal component, respectively. These indices have been developed for univariate normally distributed processes and are defined as

$$C_p = \frac{U-L}{6\sigma}, C_{pk} = \min\left\{\frac{\mu-L}{3\sigma}, \frac{U-\mu}{3\sigma}\right\}, C_{pm} = \frac{U-L}{6\sqrt{\sigma^2+(\mu-T)^2}}$$

and

$$C_{pmk} = \min\left\{\frac{\mu-L}{3\sqrt{\sigma^2+(\mu-T)^2}}, \frac{U-\mu}{3\sqrt{\sigma^2+(\mu-T)^2}}\right\}$$

respectively, where L and U denote the lower and the upper specification limit, T is the target value, and μ and σ are the mean and the standard deviation of the process.

In the assessment of the index values for the ith principal component, the specifications (L, U, and T) involved are linear combinations of the initial specifications with the elements of the corresponding eigenvector as coefficients. Obviously, Wang and Chen's[7] indices are geometric means of the values of the four basic process capability indices for the r principal components.

Wang and Chen[7] provide three examples and an actual data application that clarify the assessment of their indices. Note, however, that in these examples the number of principal components used for the assessment of the indices is not equal to r, as stated in their definitions, but is determined through the test procedure proposed by Anderson[24]. As the authors point out, the obtained values of the indices in all the examples are in accordance to the indices by Chan *et al.*[1], Taam *et al.*[3], Chen[4], and Shahriari *et al.*[5]. Moreover, they argue that an additional appealing feature of their technique is its simplicity in comparison to the other approaches.

2. New indices for multivariate processes with unilateral specifications

The four indices suggested by Wang and Chen[7] are applicable only to situations where the specifications of all the examined characteristics are bilateral. However, often one may be faced with a process whose characteristics have only a lower (upper) specification limit because their large (small) values may be desirable. In such cases, Wang and Chen's[7] approach can no longer be used. Moreover, the assessment of the most widely used process capability indices for multivariate processes, such as those suggested by Chan *et al.*[1], Taam *et al.*[3], Chen[4], and Shahriari *et al.*[5], becomes troublesome for processes connected with such specifications.

To overcome this limitation, we propose, following Wang and Chen's[7] rationale, the indices

$$\mathrm{MCPL} = \left(\prod_{i=1}^{r} \left|\mathrm{CPL}^{Y_i}\right|\right)^{\frac{1}{r}} \textit{and}\ \ \mathrm{MCPU} = \left(\prod_{i=1}^{r} \left|\mathrm{CPU}^{Y_i}\right|\right)^{\frac{1}{r}}$$

where CPL^{Y_i} and CPU^{Y_i} denote the values of the well-known indices CPL and CPU for the ith principal component. Recall that the indices CPL and CPU are applicable to univariate processes with only a lower (CPL) or an upper (CPU) specification limit and are defined as $\mathrm{CPL} = (\mu - L)/3\sigma$ and $\mathrm{CPU} = (U - \mu)/3\sigma$, respectively.

The index MCPL has been developed for the case where all the characteristics of the studied process have only lower specification limits, whereas the index MCPU is applicable to situations where all the characteristics have only upper specification limits.

The reason why absolute values of the indices CPL and CPU are considered is connected to the fact that sometimes, because of the rotation of the axes that takes place in principal component analysis, the resulting linear combination of the initial lower (upper) specification limits for some principal components may take the place of an upper (lower) specification limit. Thus, if one does not take the absolute values of the univariate indices, he may end up to negative index values. This issue is clarified in the examples given in the following sections.

The introduction of the indices MCPL and MCPU is meaningful only if the *r* variables have unilateral specifications of the same type—either lower or upper. If some of the variables have only a lower specification limit and the remaining variables have only an upper specification limit, the assessment of the specifications of the principal components and consequently the measurement of the capability of the entire process following the suggested approach becomes troublesome.

The estimation of the indices MCPL and MCPU is illustrated in the sequel using some examples taken from the article of Wang and Chen[7]. These examples have been modified so as to become applicable to the case of processes with unilateral specifications.

2.1. Example 1

Wang and Chen[7] analyzed the data of Chan *et al.*[1] that refer to a bivariate process and represent measurements on the Brinell hardness (X_1) and tensile strength (X_2). The specifications are given by $\boldsymbol{l}$ = (112.7, 32.7)′, $\boldsymbol{u}$ = (241.3, 73.3)′, and $\boldsymbol{t}$ = (177, 53)′, and the sample size is 25. The sample data are given as follows:

X_1	143	200	160	181	148	178	162	215	161	141	175	187	187
X_2	34.2	57	47.5	53.4	47.8	51.5	45.9	59.1	48.4	47.3	57.3	58.5	58.2
X_1	186	172	182	177	204	178	196	160	183	179	194	181	
X_2	57	49.4	57.2	50.6	55.1	50.9	57.9	45.5	53.9	51.2	57.5	55.6	

From the data, one may easily obtain estimates of the mean vector and the variance–covariance matrix. In particular, $\bar{\boldsymbol{x}} = \left(\, 177.2 \quad 52.32 \,\right)'$ and $\boldsymbol{s} = \begin{bmatrix} 337.8 & 85.3308 \\ 85.3308 & 33.6247 \end{bmatrix}$, respectively. Moreover, the eigenvalues of $\boldsymbol{s}$ are $\hat{\lambda}_1 = 360.1269$ and $\hat{\lambda}_2 = 11.3331$, and their corresponding normalized eigenvectors are $\hat{\mathbf{u}}_1$ = (−0.9675, −0.2529)′ and $\hat{\mathbf{u}}_2$ = (0.2529, −0.9675)′. Wang and Chen[7] argue that according to Anderson's[24] test, only the first principal component should be used, and for this reason they assess the values of their indices without taking into account the second principal component (the first principal component explains 96.95% of the total variability).

Let us now assume that the two variables have only lower specification limits. In this case, one should proceed to the estimation of the value of MCPL. Actually, keeping only the first principal component, the index MCPL is given by $MCPL = \left|CPL^{Yi}\right| = \left|\frac{\mu_{Y_1} - L_{Y_1}}{3\sigma_{Y_1}}\right|$, with μ_{Y_1}, σ_{Y_1}, and L_{Y_1} denoting the mean, the standard deviation, and the lower specification limit for the first principal component. In this case, the estimated values of μ_{Y_1} and L_{Y_1} are $\bar{x}_{Y_1} = 177.2{\cdot}(-0.9675) + 52.32{\cdot}(-0.2529) = -184.67$ and $\hat{\ell}_{Y_1} = 112.7{\cdot}(-0.9675) + 32.7{\cdot}(-0.2529) = -117.31$, respectively, whereas that of σ_{Y_1} is $s_{Y_1} = 18.977$. Thus, the value of the index MCPL is estimated to be $\widehat{\text{MCPL}} = \left|\frac{-184.67-(-117.31)}{3{\cdot}18.977}\right| = 1.18$.

It should be remarked that the linear combination of the lower specification limits of the two initial variables becomes, after the rotation of the axes, an upper specification limit for the first principal component. Actually, the estimated mean of the first principal component is −184.67, whereas the specification limit is equal to −117.31. Thus, the use of MCPL without taking its absolute value would lead to a negative value.

2.2. Example 2

Let us now consider the second example given by Wang and Chen[7], in which the data of Example 1 are used, but the specifications now are $\boldsymbol{l}$ = (86.15, 24.75)′, $\boldsymbol{u}$ = (214.75, 65.35)′. Obviously, in this case the specification area, that is, the area containing all the acceptable process values, has been relocated, keeping, however, its area and its shape fixed. The initial (Example 1) and the relocated (Example 2) specifications are depicted in Figure 1. Because the mean vector is closer to the center of the initial position of the specification area, one would expect, in the bilateral case, to result in smaller index values. Actually, as shown by Wang and Chen[7], the indices, MC_{pk}, MC_{pm}, and MC_{pmk} result in smaller values after the change of the specification limits, whereas the value of MC_p remains the same because this index is not affected from changes of the location of the process.

On the other hand, when only the lower specification limits have been assigned, the value of MCPL is expected to become larger because the specification limits have become less strict and the process data remain unchangeable. Note that the initial (Example 1) and the relocated (Example 2) specification areas in the unilateral case are the areas above the bottom and the left segments of the two rectangles that determine the corresponding areas in the bilateral case (Figure 1). These areas are illustrated in Figure 2. Actually,

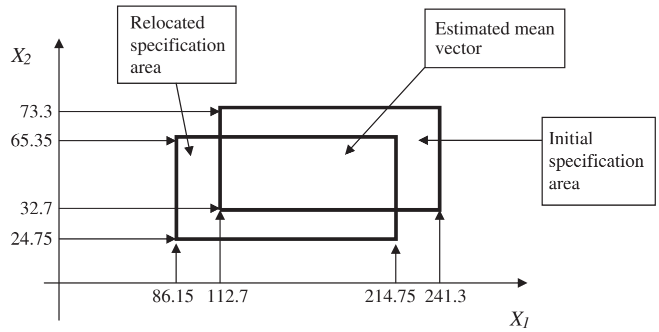

Figure 1. The specifications of Examples 1 and 2 in the bilateral specification case

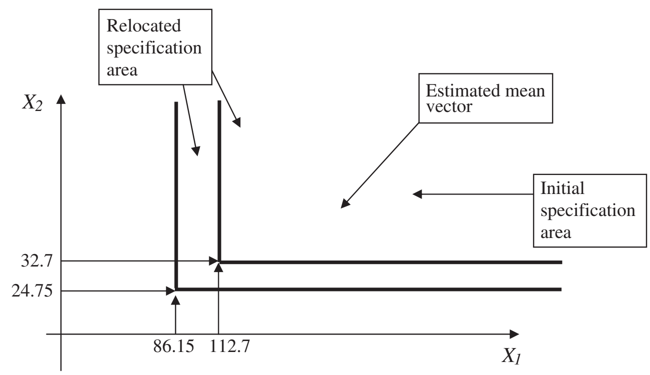

Figure 2. The specifications of Examples 1 and 2 in the unilateral specification case

in this case, the estimated values of μ_{Y_1} and σ_{Y_1} are, as previously, –184.67 and 18.977, respectively, whereas an estimate of the value of L_{Y_1} is given by

$$\hat{\ell}_{Y_1} = 86.15 \cdot (-0.9675) + 24.75 \cdot (-0.2529) = -89.61$$

Hence, an estimate of the value of the index MCPL is given by

$$\hat{MCPL} = \frac{|-184.67 - (-89.61)|}{3 \cdot 18.977} = 1.669$$

which is in accordance to our statements, that is, in this case the process is found to be more capable in comparison to Example 1.

Let us now assume that only upper specification limits have been assigned. In Example 1, the specifications are $\boldsymbol{u} = (241.3, 73.3)'$, whereas in Example 2, the specifications are $\boldsymbol{u} = (214.75, 65.35)'$. Obviously, in this case the specification areas are the areas below the top and the right segments of the two rectangles in Figure 1. In this case, one would expect to result in smaller index value in the second case because the estimate of the mean vector $\bar{\boldsymbol{x}} = (177.2 \quad 52.32)'$ is located closer to the corresponding specification limits. Actually, for $\boldsymbol{u} = (241.3, 73.3)'$, it follows that $\hat{u}_{Y_1} = 241.3 \cdot (-0.9675) + 73.3 \cdot (-0.2529) = -252,00$, whereas for $\boldsymbol{u} = (214.75, 65.35)'$, it follows that $\hat{u}_{Y_1} = 214.75 \cdot (-0.9675) + 65.35 \cdot (-0.2529) = -224.30$ and the corresponding estimates of the index MCPU are

$$\hat{MCPU} = \frac{|-184.67 - (-252.00)|}{3 \cdot 18.977} = 1.18$$

and

$$\hat{MCPU} = \frac{|-184.67 - (-224.30)|}{3 \cdot 18.977} = 0.70$$

Figures 3 and 4 provide some insight into the properties of the suggested indices for various values of the specification limits and not only for a specific change. These depict contour plots of MCPL (Figure 3) and MCPU (Figure 4) for the process defined in Examples 1 and 2. These plots cover a wide range of specification limits and illustrate the fact that the values of the new indices are appreciably sensitive to changes in the specification area. Note that such changes in the specification area do affect the capability of the process because they induce changes to the corresponding proportion of conforming items.

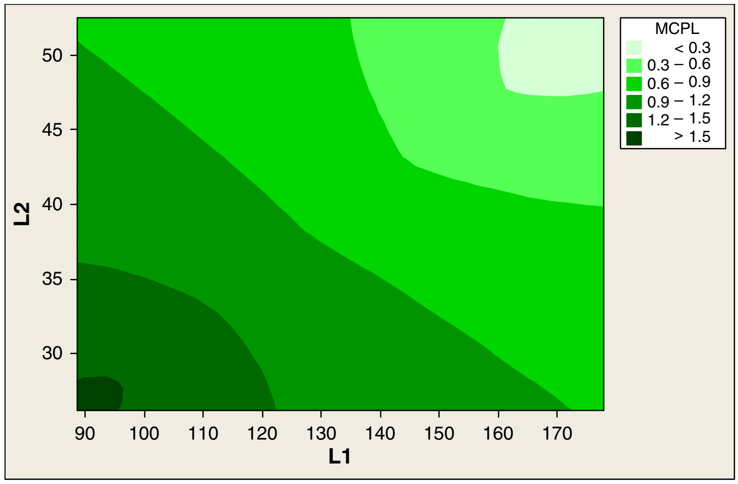

Figure 3. Contour plot of MCPL for various values of L_1 and L_2 for the process of Examples 1 and 2

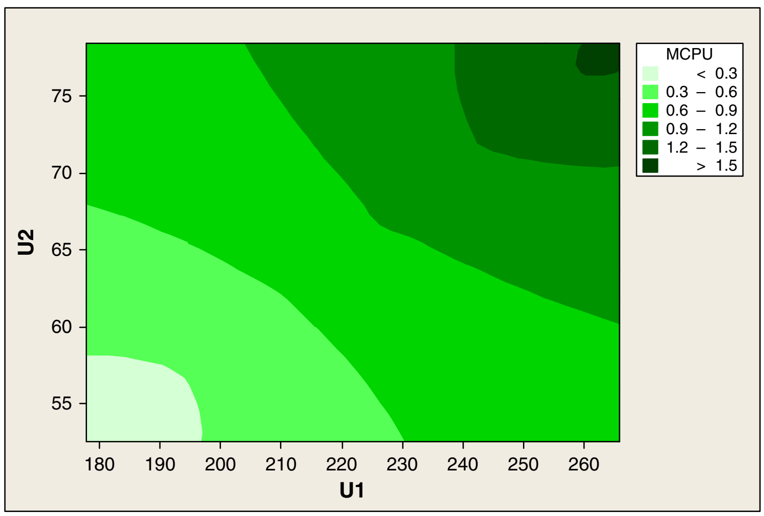

Figure 4. Contour plot of MCPU for various values of U_1 and U_2 for the process of Examples 1 and 2

In particular, from Figure 3, we observed that the value of the index MCPL is high when both specifications are small, that is, when they are far from the mean of the process ($\bar{\boldsymbol{x}} = (\,177.2 \quad 52.32\,)'$). On the other hand, when the specification limits move toward the mean, the value of the index becomes close to zero. Therefore, the suggested index seems to perform rather satisfactorily.

Similar conclusions may be drawn for the index MCPU. Actually, from Figure 4, one may observe that the value of this index is high when both specifications are large, whereas its value decreases as the specification limits move toward the mean.

Finally, some insight into the behavior of the index MCPL is provided by the following example referring to an actual data set considered by Wang and Chen[7].

2.3. *Example 3*

The data in question come from a plastics manufacturer in Taiwan. The variables considered are the depth (X_1), the length (X_2), and the width (X_3) of a plastic product. The vectors of the specification limits are given by $\boldsymbol{l} = (2.1, 304.5, 304.5)'$ and $\boldsymbol{u} = (2.3, 305.1, 305.1)'$, respectively. On the basis of a random sample of 50 pieces of the product, the mean vector and the variance–covariance matrix are estimated to be $\bar{\boldsymbol{x}} = (\,2.16, \quad 304.72, \quad 304.77\,)'$ and $\boldsymbol{s} = \begin{bmatrix} 0.0021 & 0.0008 & 0.0007 \\ 0.0008 & 0.0017 & 0.0012 \\ 0.0007 & 0.0012 & 0.0020 \end{bmatrix}$, respectively. Wang and Chen[7] have implemented Anderson's[24] test and concluded that only the first two principal components should be used (they explain 89.04% of the total variability). The estimated eigenvalues and the corresponding eigenvectors of the first two principal components are $\hat{\lambda}_1 = 0.0037$, $\hat{\lambda}_2 = 0.0015$, $\hat{\mathbf{u}}_1 = (0.5222, 0.5824, 0.6230)'$, and $\hat{\mathbf{u}}_2 = (0.8385, -0.2172, -0.4998)'$.

Assuming that only the lower specification limits for the three variables have been assigned, we obtain $\bar{x}_{Y_1} = 2.16{\cdot}0.5222 + 304.72{\cdot}0.5824 + 304.77{\cdot}0.6230 = 368.4686$, $\hat{\ell}_{Y_1} = 2.1{\cdot}0.5222 + 304.5{\cdot}0.5824 + 304.5{\cdot}0.6230 = 368.1409$, and $s_{Y_1} = 0.0608$ for the first principal component and $\bar{x}_{Y_2} = 2.16{\cdot}0.8385 + 304.72{\cdot}(-0.2172) + 304.77{\cdot}(-0.4998) = -216.698$, $\hat{\ell}_{Y_2} = 2.1{\cdot}0.8385 + 304.5{\cdot}(-0.2172) + 304.5{\cdot}(-0.4998) = -216.5657$, and $s_{Y_2} = 0.03873$ for the second principal component. Therefore, an estimate of MCPL is given by

$$\hat{MCPL} = \left(\prod_{i=1}^{2}\left|\hat{CPL}^{Y_i}\right|\right)^{\frac{1}{2}}$$
$$= \sqrt{\frac{|368.4686 - 368.1409|}{3{\cdot}0.0608}\cdot\frac{|-216.698-(-216.5657)|}{3{\cdot}0.03873}}$$
$$= \sqrt{(1.7966)(1.1386)}$$
$$= 1.43.$$

Let us now assume that the (unilateral) specifications for the three variables become stricter and determined by $\boldsymbol{l} = (2.15, 304.6, 304.6)'$. In this case, we would expect to find a smaller (estimated) value of MCPL because the probability of producing outside the specification area becomes now higher (the mean is now closer to the specification limits and the variability, as expressed by $\boldsymbol{s}$ remains unchanged). In this case, the values of $\bar{x}_{Y_1}$, s_{Y_1}, $\bar{x}_{Y_2}$, and s_{Y_2} remain unchanged because they are affected only from the data and not from the specifications. On the other hand, the lower specification limit for the first principal component becomes $\hat{\ell}_{Y_1} = 2.15{\cdot}0.5222 + 304.6{\cdot}0.5824 + 304.6{\cdot}0.6230 = 368.288$, and the lower specification limit for the second principal component becomes $\hat{\ell}_{Y_2} = 2.15{\cdot}0.8385 + 304.6{\cdot}(-0.2172) + 304.6{\cdot}(-0.4998) = -216.595$.

Therefore, an estimate of MCPL is given by

$$\hat{MCPL} = \sqrt{\frac{|368.4686 - 368.288|}{3{\cdot}0.0608}\cdot\frac{|-216.698-(-216.595)|}{3{\cdot}0.03873}}$$
$$= \sqrt{(0.9925)(0.8828)}$$
$$= 0.936$$

and thus the change of the specification limits is reflected by the decrease of the value of the index.

The contour plot of Figure 5 reflects a behavior by the index MCPL consistent with that revealed in Example 2 for various values of the specification limits. Note that without loss of generality, it was assumed in this case that $L_2 = L_3$ for ease in the construction of a contour plot. One may observe that, as in the case of Example 2, the value of the index is high when all the specifications are low and decreases as the specification limits move toward the mean.

3. New indices that take into account the variance explained by the principal components

A drawback of the indices defined by Wang and Chen[7] is the fact that they assign the same importance to all the principal components that are found to be significant and are thus used in their assessment. To illustrate this, consider again Wang and Chen's[7] actual data used in Example 3 of the previous section.

To decide how many principal components to include in their analysis, they implemented Anderson's[24] test. They initially tested the hypothesis H_0: $\lambda_1 = \lambda_2 = \lambda_3$. The value of the test statistic (37.96) exceeded $\chi^2_{5,0.95} = 11.07$; hence, they proceeded with testing the hypothesis H_0: $\lambda_2 = \lambda_3$. Again, the value of the test statistic (9.94) was in the 5% critical region ($\chi^2_{2,0.95} = 5.99$). Hence, taking into

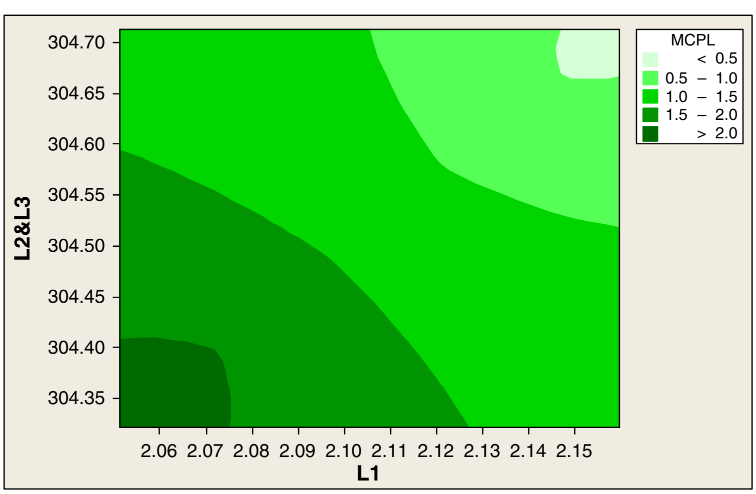

Figure 5. Contour plot of MCPL for various values of L_1 and L_2 and L_3 for the process of Example 3

account the fact that the first two principal components explained the 89.04% of the total variability, they based the assessment of their indices solely on the first two principal components.

It is obvious, however, that the first two principal components are treated as having the same importance in terms of the indices (they are geometric means of the corresponding indices for univariate processes). However, the first principal component explains 63.93% of the total variability, whereas the second principal component explains only 25.11% of it. Furthermore, Anderson's[24] test of the hypothesis H_0: $\lambda_1 = \lambda_2$ yields $T = -(50-1)\sum_{j=1}^{2} \ln\hat{\lambda}_j + (2)(50-1)\ln\left(\sum_{j=1}^{2}\left(\hat{\lambda}_j/2\right)\right) = 9.66$.

This is in the 5% critical region ($\chi^2_{2,0.95} = 5.99$). Therefore, the fact that the first two principal components differ significantly in their associated eigenvalues, and hence in the percentages of the variance explained by them, is completely disregarded by the indices MC_p, MC_{pk}, MC_{pm}, and MC_{pmk}.

To overcome this deficiency, we suggest some new indices that allow for potential differences in the portions of the variance explained by the principal components considered. These differences are taken into account by assigning unequal weights to the index values corresponding to the principal components used and, in particular, proportional to the portions of the variance explained by them, as determined by their respective eigenvalues. Specifically, if the studied process has bilateral tolerances, we suggest the indices $MC'_p = \frac{1}{V}\sum_{i=1}^{r'}\lambda_i C_p^{Y_i}, MC'_{pk} = \frac{1}{V}\sum_{i=1}^{r'}\lambda_i C_{pk}^{Y_i}, MC'_{pm} = \frac{1}{V}\sum_{i=1}^{r'}\lambda_i C_{pm}^{Y_i}$ and $MC'_{pmk} = \frac{1}{V}\sum_{i=1}^{r'}\lambda_i C_{pmk}^{Y_i}$ whereas in the case of unilateral specifications, we propose the index $MCPL' = \frac{1}{V}\sum_{i=1}^{r'}\lambda_i\left|CPL^{Y_i}\right|$ if all of them are lower specification limits and the index $MCPU' = \frac{1}{V}\sum_{i=1}^{r'}\lambda_i\left|CPU^{Y_i}\right|$ if all of them are upper specification limits. Here, r' denotes the number of the principal components used, as determined through Anderson's[24] test or any other criterion, and V denotes the sum of the eigenvalues associated with the selected principal components, that is, $V = \sum_{i=1}^{r'}\lambda_i$.

Obviously, the proposed indices are weighted averages of the indices for univariate processes for the selected principal components, with weights on the basis of the proportion of variability explained by each of them. Alternatively, if one wishes to keep all the principal components in the analysis, the formulae for the suggested indices can be modified by replacing r' and V by r and $\sum_{i=1}^{r}\lambda_i$, respectively. If the variance–covariance matrix $\boldsymbol{\Sigma}$ is unknown, the values of the indices can be estimated using the eigenvalues and the eigenvectors of the sample variance–covariance matrix $\boldsymbol{S}$. Besides, in the case where the r variables that describe the process are expressed in measurement units with substantial differences, one may alternatively assess the values of the indices using the eigenvalues and the eigenvectors of the correlation matrix.

For the data considered previously in Example 3, $\hat{\lambda}_1 = 0.0037$ and $\hat{\lambda}_2 = 0.0015$. Therefore, $\hat{V} = 0.0052$, and thus the corresponding estimates of the suggested indices are $\hat{MC}'_p = 1.94, \hat{MC}'_{pk} = 1.60, \hat{MC}'_{pm} = 1.26$ *and* $\hat{MC}'_{pmk} = 1.06$.

The corresponding estimates of the indices by Wang and Chen[7] are $\hat{MC}_p = 1.60$, $\hat{MC}_{pk} = 1.42$, $\hat{MC}_{pm} = 1.22$, and $\hat{MC}_{pmk} = 1.08$. As one may observe, the estimates of the new indices vary from those of the corresponding indices by Wang and Chen[7], especially for the first two indices, that is, $\hat{MC}'_p$ and $\hat{MC}'_{pk}$. The cause of this discrepancy is the fact that the univariate indices C_p and C_{pk} result in distant values for the first two principal components. Specifically, the estimated index C_p is 2.27 for the first principal component and 1.13 for the second. The corresponding estimates of C_{pk} are 1.80 and 1.12, respectively. However, the first principal component explains 63.93% of the total variability whereas the second only 25.11% of it. Thus, it is intuitively reasonable to give more importance to the index values that have been assessed for the first principal component because they carry a larger amount of information for the entire process examined. This issue is reflected by the values of the suggested indices, which are closer to the corresponding value for the first principal component in comparison to the indices by Wang and Chen[7]. The properties of the suggested indices and their advantages in comparison to the indices by Wang and Chen[7] are discussed in the next section.

4. Properties of the new indices

To investigate the performance of the indices proposed in Section 3 and to compare them to the indices suggested by Wang and Chen[7], we carried out a simulation study about bivariate processes for ease of illustration. In particular, our aim in this study is to examine the behavior of the suggested indices, in the bivariate case, in situations where the values of the parameters of the process change and to see if their behavior is in accordance to the capability of the process. Without loss of generality, it was assumed that the specifications of the two processes are those given in Table I.

For the specification limits given in Table I, a wide range of process parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) are considered. Thus, the values of the mean vector considered, for both processes, range from 0 to 1 with increments of 0.1, that is, 0, 0.1, 0.2,. . ., 0.9, 1.0. It should be noted that we considered only mean values greater than the target value, because the selected specifications are symmetric (i.e. each target value coincides with the mean of the corresponding lower and upper specification limits), and the results for mean values smaller than the target value are similar. The values of the parameters σ_1^2, σ_2^2, and $\sigma_{1,2} = Cov(X_1, X_2)$ considered and the corresponding eigenvalues are given in Table II.

For the selected combinations of these parameter values, the indices MC_p, MC_{pk}, MC_{pm}, and MC_{pmk} by Wang and Chen[7] and the new indices MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} proposed in Section 3 have been assessed. Figures 6–10 depict plots of the eight indices for the four variance combinations and for 11 different (common) mean values corresponding to $\sigma_{1,2}$ = 0.1, 0.2, 0.3, 0.4, and 0.5.

Table I. Specification limits for the studied processes

	L	T	U
X_1	−2	0	2
X_2	−2.2	0	2.2

Table II. Values of the parameters σ_1^2, σ_2^2, and $\sigma_{1,2}$ and the corresponding eigenvalues

σ_1^2	σ_2^2	$\sigma_{1,2}$	λ_1	λ_2
0.5	0.6	0.1	0.66	0.44
		0.2	0.76	0.34
		0.3	0.85[a]	0.25[a]
		0.4	0.95[a]	0.15[a]
		0.5	1.05[a]	0.05[a]
0.75	1.0	0.1	1.04	0.71
		0.2	1.11	0.64
		0.3	1.20	0.55
		0.4	1.29	0.46
		0.5	1.39	0.36
1.0	1.5	0.1	1.52	0.98
		0.2	1.57	0.93
		0.3	1.64	0.86
		0.4	1.72	0.78
		0.5	1.81	0.69
1.2	1.7	0.1	1.72	1.18
		0.2	1.77	1.13
		0.3	1.84	1.06
		0.4	1.92	0.98
		0.5	2.01	0.89

[a]Cases where the first principal component explains at least 75% of the total variability. In these cases, the values of the indices suggested in Section 3 are identical to those by Wang and Chen[7] because only the first principal component is used.

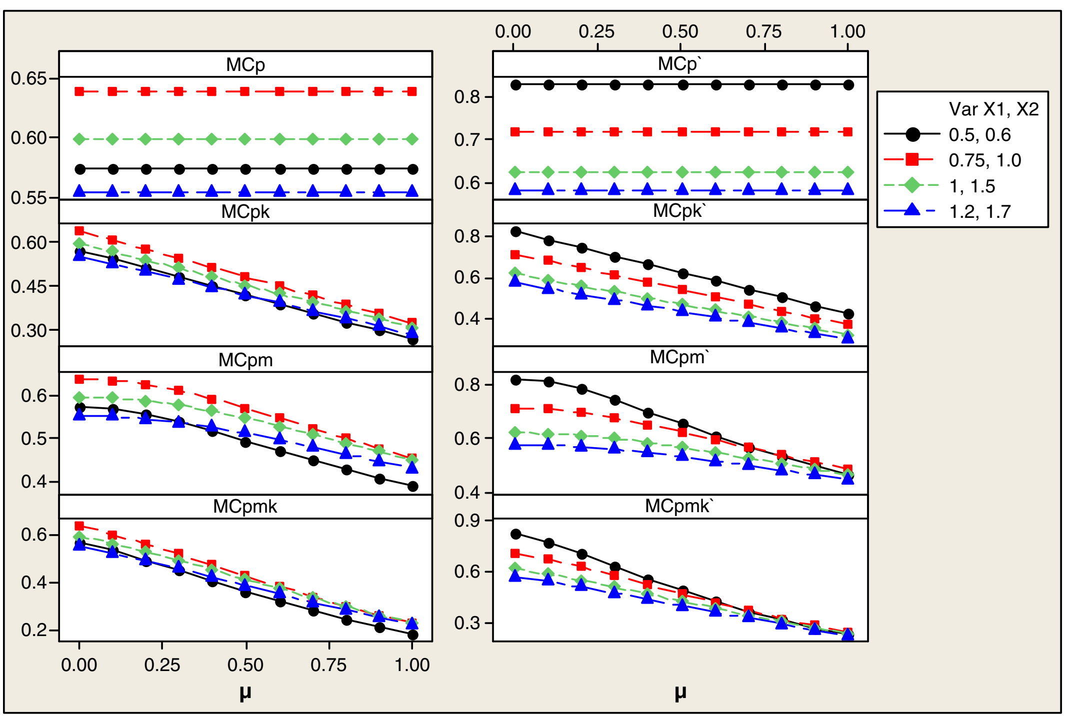

Figure 6. Plot of the indices MC_p, MC_{pk}, MC_{pm}, MC_{pmk}, MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for various common values μ for μ_1 and μ_2 and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2} = 0.1$

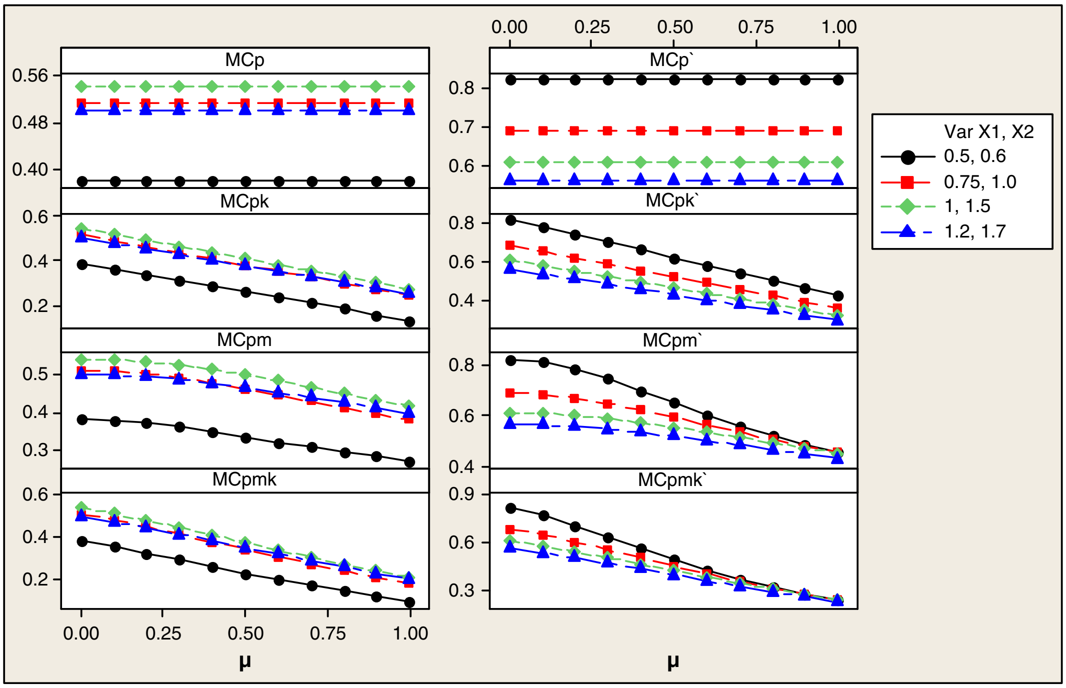

Figure 7. Plot of the indices MC_p, MC_{pk}, MC_{pm}, MC_{pmk}, MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for various common values μ for μ_1 and μ_2 and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2} = 0.2$

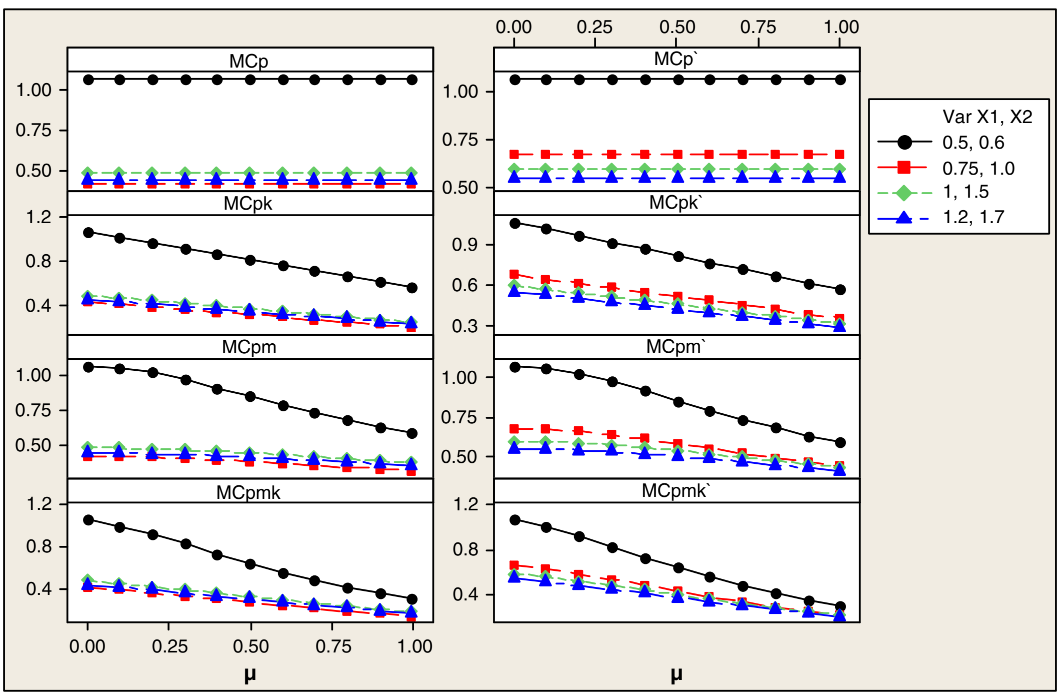

Figure 8. Plot of the indices MC_p, MC_{pk}, MC_{pm}, MC_{pmk}, MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for various common values μ for μ_1 and μ_2 and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2} = 0.3$

In each case, for fixed values of $\sigma_{1,2}$, μ_1, and μ_2, one would expect to end up with higher index values when the variances of the two processes are small, that is, in the case where $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 0.6$, and to lower values as the variances increase. As one may observe, the four new indices possess this property in all the cases examined because the line corresponding to the case $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 0.6$ is always above the other lines and the three other lines are positioned according to the variances of the respective processes. Moreover, the differences in the four variance combinations examined, between the values of each of the indices MC'_{pk}, MC'_{pm}, and MC'_{pmk}, are greater for values of μ_1 and μ_2 that are close to the target value and become smaller as the two processes

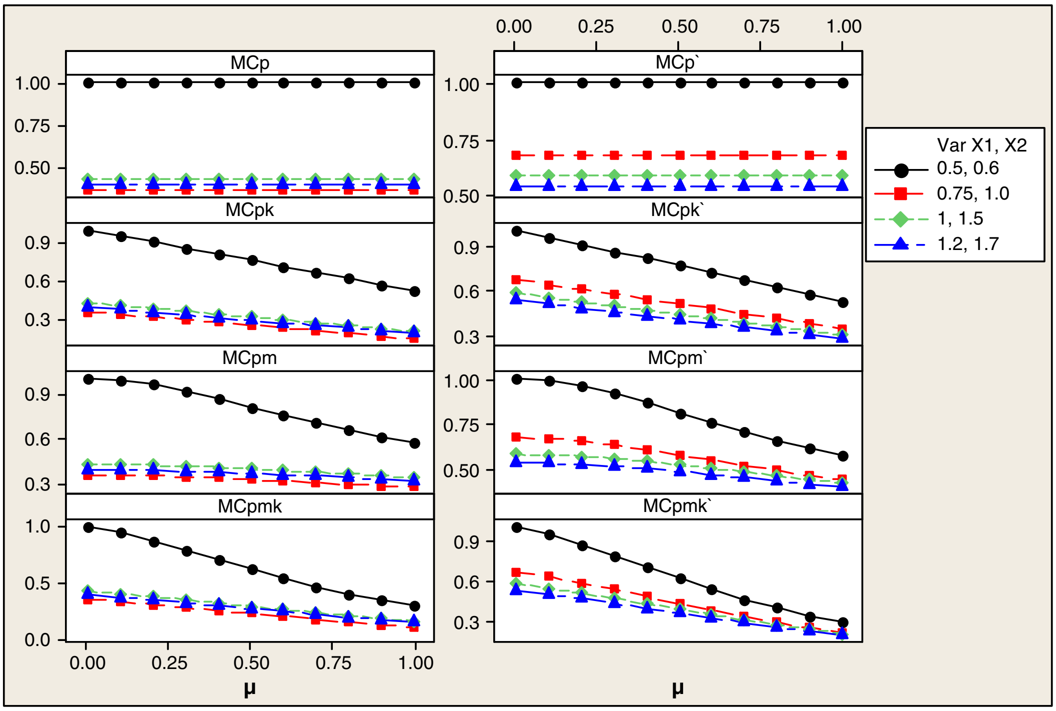

Figure 9. Plot of the indices MC_p, MC_{pk}, MC_{pm}, MC_{pmk}, MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for various common values μ for μ_1 and μ_2 and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2}=0.4$

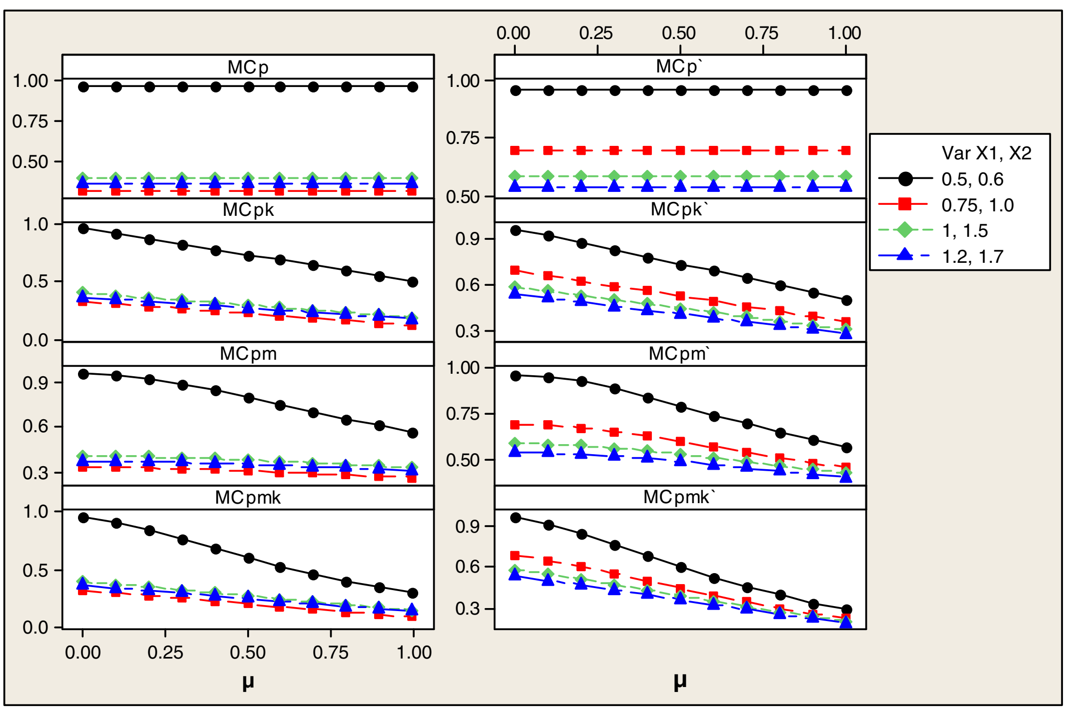

Figure 10. Plot of the indices MC_p, MC_{pk}, MC_{pm}, MC_{pmk}, MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for various common values μ for μ_1 and μ_2 and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2}=0.5$

become off target. This is connected to the fact that for an on-target process, it is very important to have small variability because in such a case the probability of producing near the target is very high. On the other hand, for an off-target process, it is not always preferred to have small variability because in such a case the process produces within the specification area but the probability of producing near the target is negligible. On the contrary, an off-target process with relatively high variability may produce outside the specification area, but some of the produced items are near the target. This property of the suggested indices is quite appealing

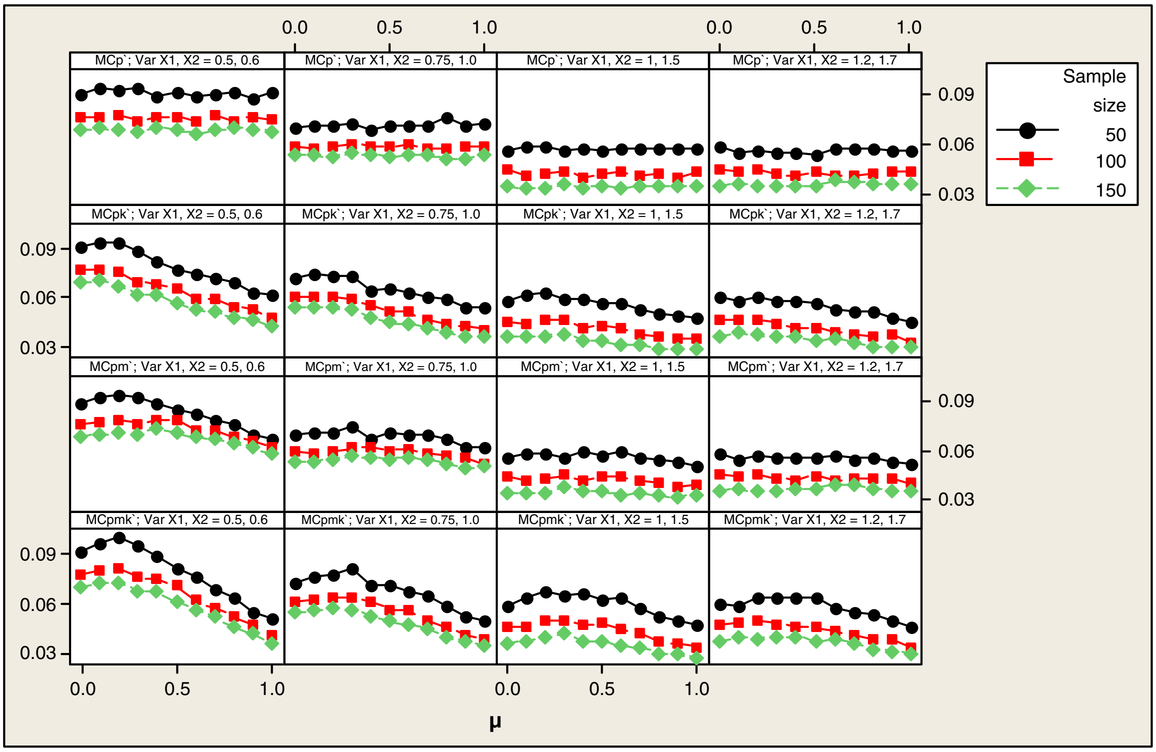

Figure 11. Plot of the standard errors of the estimators of the indices MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for three sample sizes, various common values μ for μ_1 and μ_2, and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2}=0.1$

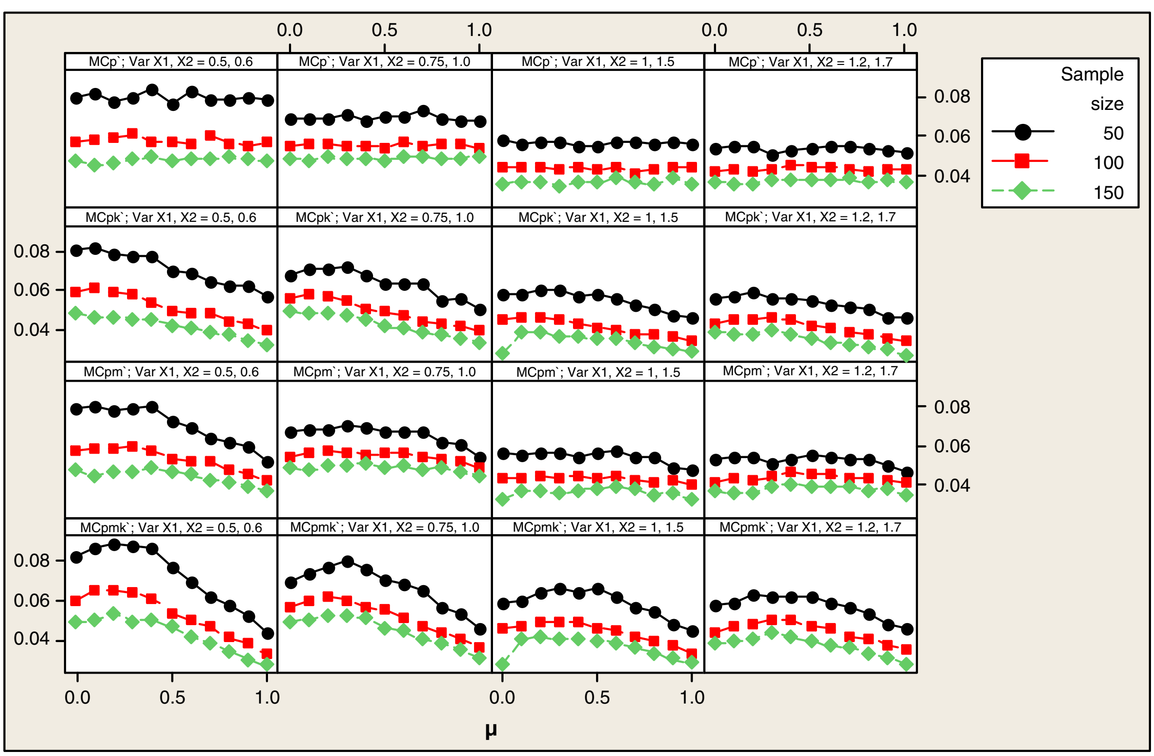

Figure 12. Plot of the standard errors of the estimators of the indices MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for three sample sizes, various common values μ for μ_1 and μ_2, and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2}=0.2$

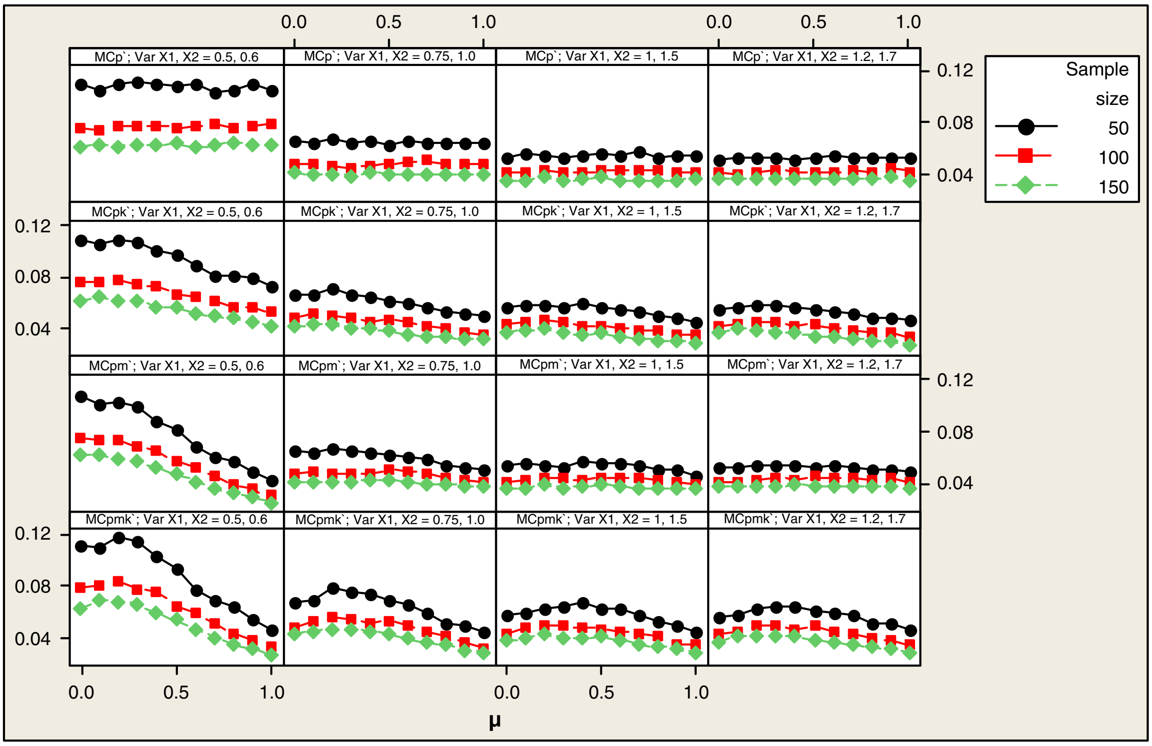

Figure 13. Plot of the standard errors of the estimators of the indices MC'_{p}, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for three sample sizes, various common values μ for μ_1 and μ_2, and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2}=0.3$

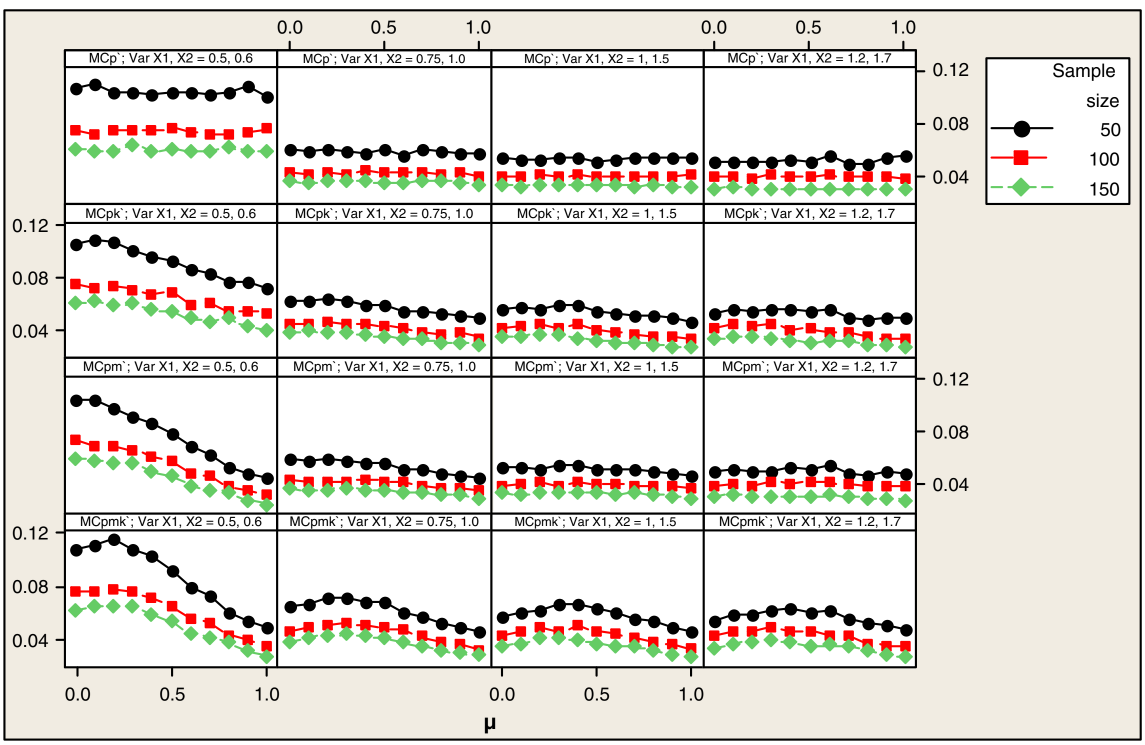

Figure 14. Plot of the standard errors of the estimators of the indices MC'_{p}, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for three sample sizes, various common values μ for μ_1 and μ_2, and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2}=0.4$

and is reflected mainly by the two indices that take into account the target value, that is, by the indices MC'_{pm} and MC'_{pmk}. It should be noted that the index MC'_{p} does not possess this property because it is not affected by the means of the processes. Therefore, the suggested indices perform rather satisfactorily in all the cases considered.

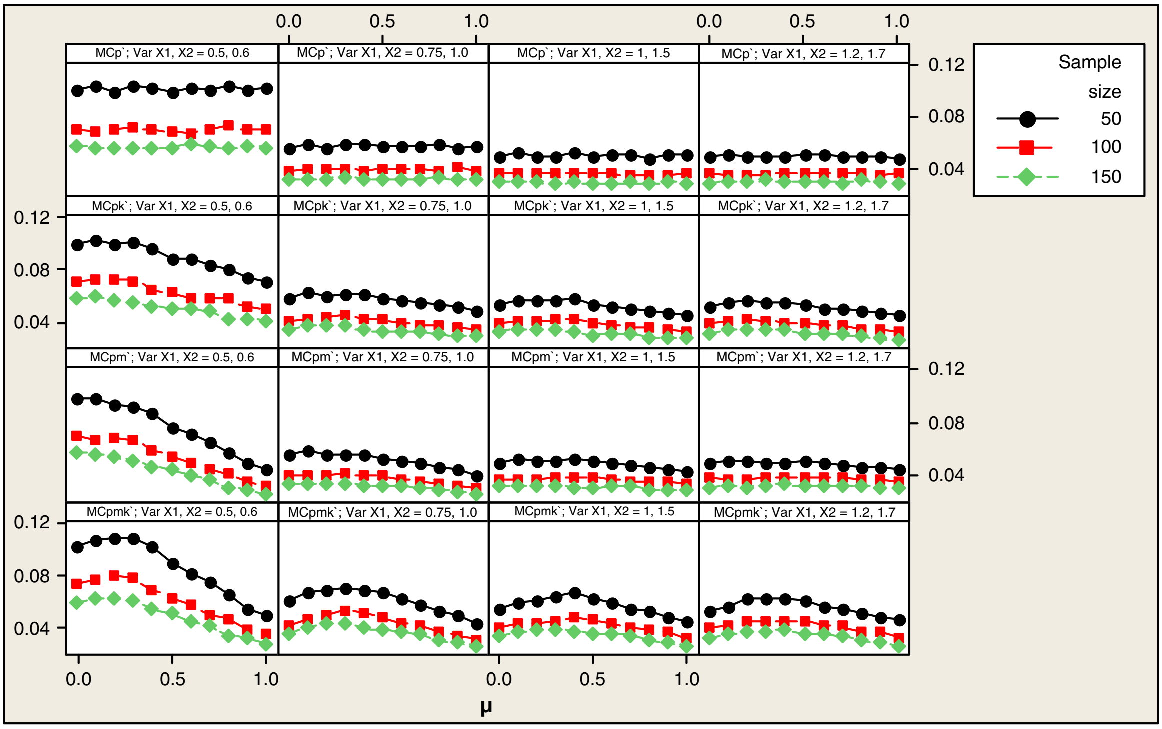

Figure 15. Plot of the standard errors of the estimators of the indices MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} for three sample sizes, various common values μ for μ_1 and μ_2, and various values of σ_1^2 and σ_2^2 with $\sigma_{1,2}=0.5$

On the other hand, the indices by Wang and Chen[7] do not possess the appealing features of the new indices introduced in this article. For instance, from Figure 6, we observed that the value of MC_p is 0.64 when $\sigma_1^2=0.75$ and $\sigma_2^2=1.0$, whereas it is 0.57 when $\sigma_1^2=0.5$ and $\sigma_2^2=0.6$, although in the second case the variability is smaller. Similar conclusions may be drawn from Figures 7–10.

Except for the properties of the new indices and their ability to measure process capability satisfactorily, we examined the accuracy of their estimators. This can be achieved by assessing their standard errors. As already stated earlier, their estimators are obtained by substituting $\bar{\mathbf{X}}$, $\boldsymbol{S}$, and $\mathbf{R}$ for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\rho}$, respectively.

The standard errors of the estimators of the indices MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} are estimated, via simulation, using the so-called *parametric bootstrap technique* (see, e.g. Efron and Tibshirani[25]). The cases examined in this simulation study are those used earlier for examining the properties of the indices, that is, the bivariate normal distribution with values of σ_1^2, σ_2^2, and $\sigma_{1,2}$ as given in Table II and mean vector, for both processes, from 0 to 1 with increments of 0.1. For each of these cases, the parametric bootstrap technique is implemented by generating 1000 random samples of sizes 50, 100, and 150. For each of the generated samples, the values of MC'_p, MC'_{pk}, MC'_{pm}, and MC'_{pmk} are estimated. The standard deviation of the 1000 estimates, which has been assessed for a specific index, can be used for estimating the standard error of its corresponding estimator.

The estimated standard errors are depicted in Figures 11–15. The basic findings are listed as follows:

- In all the cases, the value of standard error decreases as the sample size increases.
- The standard error of the estimator of MC'_p is not affected by the mean of the process because, by its definition, this index is affected only by the variability of the process.
- The standard errors of the estimators of MC'_{pk}, MC'_{pm}, and MC'_{pmk} seem to become smaller as the mean moves further from the target value.

5. Conclusions

In this article, extending the initial idea of Wang and Chen's[7], some alternative approaches have been proposed for using principal component analysis in the measurement of the capability of multivariate processes. Some new indices have been suggested that make possible the assessment of process capability for processes with unilateral tolerances. Moreover, new indices have been introduced, which take into account the proportion of the variance explained by each principal component. For all the suggested indices, illustrative examples have been provided.

The suggested indices constitute a useful tool for the practitioners in the case of multivariate processes because they perform quite satisfactorily, as illustrated via the given examples and the study of their properties provided in Section 4. They can be used for either unilateral or bilateral specifications, and their assessment is rather simple.

Acknowledgement

The authors thank the referee for comments that helped improve an earlier version of the article.

References

1. Chan LK, Cheng SW, Spiring FA. A multivariate measure of process capability. *Journal of Modelling and Simulation* 1991; **11**(1):1–6.
2. Pearn WL, Kotz S, Johnson NL. Distributional and inferential properties of process capability indices. *Journal of Quality Technology* 1992; **24**(4): 216–231.
3. Taam W, Subbaiah P, Liddy JW. A note on multivariate capability indices. *Journal of Applied Statistics* 1993; **20**(3):339–351.
4. Chen H. A multivariate process capability index over a rectangular solid tolerance zone. *Statistica Sinica* 1994; **4**:749–758.
5. Shahriari H, Hubele NF, Lawrence FP. A multivariate process capability vector. *Proceedings of the 4th Industrial Engineering Research Conference*, *Institute of Industrial Engineers* 1995; 304–309.
6. Bernardo JM, Irony TZ. A general multivariate Bayesian process capability index. *The Statistician* 1996; **45**(3):487–502.
7. Wang FK, Chen JC. Capability index using principal component analysis. *Quality Engineering* 1998; **11**(1):21–27.
8. Wang FK, Du TCT. Using principal component analysis in process performance for multivariate data. *Omega* 2000; **28**(2):185–194.
9. Castagliola P, Castellanos JG. Capability indices dedicated to the two quality characteristics case. *Quality Technology and Quantitative Management* 2005; **2**(2):201–220.
10. Shinde RL, Khadse KG. Multivariate process capability using principal component analysis. *Quality and Reliability Engineering International* 2009; **25**(1):69–77.
11. Kotz S, Johnson NL. Process Capability Indices. Chapman and Hall: London, 1993.
12. Kotz S, Johnson NL. Process capability indices - a review, 1992–2000. *Journal of Quality Technology* 2002; **34**(1):2–19.
13. Kotz S, Lovelace CR. Process Capability Indices in Theory and Practice, Arnold A: London, 1998.
14. Pearn WL, Kotz S. Encyclopedia and Handbook of Process Capability Indices: A Comprehensive Exposition of Quality Control Measures. World Scientific Publishing: Singapore, 2006.
15. Wu CW, Pearn WL, Kotz S. An overview of theory and practice on process capability indices for quality assurance. *International Journal of Production Economics* 2009; **117**(2):338–359.
16. Wang FK, Hubele NF, Lawrence FP, Miskulin JD, Shahriari H. Comparison of three multivariate process capability indices. *Journal of Quality Technology* 2000; **32**(3):263–275.
17. Zahid MA, Sultana A. Assessment and comparison of multivariate process capability indices in ceramic industry. *Journal of Mechanical Engineering* 2008; **39**(1):18–25.
18. Spiring F, Leung B, Cheng S, Yeung A. A bibliography of process capability papers. *Quality and Reliability Engineering International* 2003; **19**(5): 445–460.
19. Yum BJ, Kim KW. A bibliography of the literature on process Capability Indices: 2000–2009. *Quality and Reliability Engineering International* 2011; (to appear).
20. Xekalaki E, Perakis M. The Use of Principal Components Analysis in the Assessment of Process Capability Indices. *Proceedings of the Joint Statistical Meetings of the American Statistical Association - Section on Physical & Engineering Sciences (SPES)*, New York, August 2002; 3819–3823.
21. Mardia KV, Kent JT, Bibby JM. Multivariate Analysis. Academic Press: London, 1979.
22. Morrison DF. Multivariate Statistical Methods. McGraw-Hill: New York, 1990.
23. Jackson JE. Principal components and factor analysis: Part I – principal components. *Journal of Quality Technology* 1980; **12**:201–213.
24. Anderson TW. Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics* 1963; **34**:122–148.
25. Efron B, Tibshirani RJ.. An introduction to the bootstrap. Chapman and Hall: New York, 1993.