



MPCI: An R Package for Computing Multivariate Process Capability Indices

Edgar Santos-Fernández

Universidad Central “Marta Abreu”
de Las Villas

Michele Scagliarini

Università di Bologna

Abstract

Manufacturing processes are often based on more than one quality characteristic. When these variables are correlated the process capability analysis should be performed using multivariate statistical methodologies. Although there is a growing interest in methods for evaluating the capability of multivariate processes, little attention has been given to developing user friendly software for supporting multivariate capability analysis. In this work we introduce the package **MPCI** for R, which allows to compute multivariate process capability indices. **MPCI** aims to provide a useful tool for dealing with multivariate capability assessment problems. We illustrate the use of **MPCI** package through both simulated and real examples.

Keywords: multivariate process capability indices, multivariate normal distribution, principal component analysis, statistical quality control.

1. Introduction

Process capability indices such as, C_p , C_{pk} , C_{pm} and C_{pmk} are typically used as measures of process capability in the univariate domain (Kotz and Lovelace 1998). However, in the modern industry there are many manufacturing processes in which the quality is measured by the joint level of several correlated characteristics. Various process capability indices have been developed for assessing the performance of multivariate manufacturing processes. In general, multivariate process capability indices are constructed in one of the following ways: using the ratio of the volume of a tolerance region to the volume of a process region – see, for example, Shahriari, Hubele, and Lawrence (1995), Taam, Subbaiah, and Liddy (1993), and Pearn, Wang, and Yen (2007) and the references contained therein; using the proportion of nonconforming items – see Chen (1994) and Chen, Pearn, and Lin (2003); or using the

principal component analysis (PCA) – see Wang and Chen (1998), Xekalaki and Perakis (2002) and Wang (2005).

With the exception of Scagliarini and Vermiglio (2008), which developed a visual basic program to realize an Excel spreadsheet for computing two multivariate measures of capability, a current problem with multivariate measures of capability is the shortage of user-friendly software.

In this paper, with the purpose of providing a useful tool to help practitioners in performing multivariate capability analyses, we present the R (R Development Core Team 2012) package **MPCI** which is available from the Comprehensive R Archive Network at <http://www.cran.r-project.org/package=MPCI>. **MPCI** currently computes five multivariate process capability indices following the methods proposed by Shahriari *et al.* (1995), Taam *et al.* (1993) and the indices based on PCA proposed by Wang and Chen (1998), Xekalaki and Perakis (2002) and Wang (2005).

We structured this work by introducing each considered index together with its implementation in **MPCI**. The layout of the paper is as follows: Section 2 introduces the notation and the data sets that will be used throughout the article; Section 3 provides the theoretical background and the solutions offered by **MPCI** for performing multivariate capability analysis using the multivariate capability vector (Shahriari *et al.* 1995); Section 4 introduces the multivariate capability index MC_{pm} (Taam *et al.* 1993) and illustrates the use **MPCI**; Section 5 examines the statistical methodology and explains how to use **MPCI** to obtain multivariate process capability indices based on PCA (Wang and Chen 1998; Xekalaki and Perakis 2002; Wang 2005); Section 6 ends the paper with several concluding remarks.

2. Notation and data sets

Let \mathbf{X} represent the vector of the v quality characteristics of interest with mean vector μ and covariance matrix Σ . We assume that the joint probability distribution of the v quality characteristics is the multivariate normal distribution

$$\mathbf{X} \sim N(\mu, \Sigma) \quad (1)$$

Let us denote with $\mathbf{LSL}^\top = [LSL_1, LSL_2, \dots, LSL_v]$, $\mathbf{USL}^\top = [USL_1, USL_2, \dots, USL_v]$ and $\mathbf{T}^\top = [T_1, T_2, \dots, T_v]$ the v -vectors values of the lower specification limits, upper specification limits and target values respectively.

To illustrate the use of **MPCI** we make use of a simulated data set and real data taken from the literature. The `dataset1` in the package contain a sample of $n = 180$ observations of dimension $v = 5$ generated by assuming a multivariate normal distribution with mean vector and covariance matrix given by

$$\mu^\top = [30, 70, 15, 12, 120] \quad (2)$$

and

$$\Sigma = \begin{bmatrix} 2.100 & 0.178 & 0.318 & 0.579 & 0.136 \\ 0.178 & 2.400 & 0.099 & 0.245 & 0.207 \\ 0.318 & 0.099 & 1.700 & 0.086 & 0.057 \\ 0.579 & 0.245 & 0.086 & 0.800 & 0.331 \\ 0.136 & 0.207 & 0.057 & 0.331 & 4.100 \end{bmatrix} \quad (3)$$

respectively. The target values vector coincides with the mean vector $\mathbf{T} = \mu$, while the specification limits are

$$\mathbf{LSL}^\top = [24, 60, 10, 8, 100] \quad (4)$$

and

$$\mathbf{USL}^\top = [36, 80, 20, 16, 140] \quad (5)$$

The real data set, included as `dataset2` in the package, consists of the $n = 25$ sample observations of the bivariate example discussed by Wang and Chen (1998):

X_1 is the brinell hardness

143 200 160 181 148 178 162 215 161 141
 175 187 187 186 172 182 177 204 178 196
 160 183 179 194 181;

X_2 is the tensile strength brinell of a process

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 57 49.4 57.2 50.6 55.1 50.9 57.9
 45.5 53.9 51.2 57.5 55.6.

The specification limits for X_1 and X_2 are $[112.7, 241.3]$ and $[32.7, 73.3]$, respectively. The target vector is $\mathbf{T} = [177, 53]$.

3. The multivariate capability vector

The multivariate capability vector was proposed by Shahriari *et al.* (1995) based on the original work of Hubele, Shahriari, and Cheng (1991). The vector contains three components and appears as

$$(C_{pM}, PV, LI)$$

C_{pM} is the first component of the vector, it is a ratio of volumes. The numerator is the volume defined by the engineering tolerance region, while the denominator is the area or volume of a “modified process region”, defined as the smallest region similar in shape to the engineering tolerance region, circumscribed about a specified probability contour:

$$C_{pM} = \left[\frac{\text{Volume of engineering tolerance region}}{\text{Volume of modified process region}} \right]^{1/v} \quad (6)$$

The volume of the engineering tolerance region is

$$\prod_{i=1}^v (USL_i - LSL_i) \quad (7)$$

To compute the volume of the modified process region it is worth reminding that under the hypothesis of multivariate normality the statistic

$$(\mathbf{X} - \mu)^\top \Sigma (\mathbf{X} - \mu) = g(\mathbf{x}) \quad (8)$$

follows a χ^2 distribution with v degree of freedom. Therefore, the borders of the process region UPL_i , the upper process limit, and LPL_i , the lower process limit ($i = 1, 2, \dots, v$) are derived from the projection of the probability ellipsoid onto the respective axes. Specifically, they are determined by solving the systems of equations of first derivative, with respect to each X_i of the quadratic form (Nickerson 1994)

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma} (\mathbf{X} - \boldsymbol{\mu}) = \chi_{(v,\alpha)}^2 \quad (9)$$

where $\chi_{(v,\alpha)}^2$ is the $100(1 - \alpha)$ -th percentile of a χ^2 distribution with v degrees of freedom associated with the probability contour. Usually, in analogy with the “ 6σ ” in the denominator of the univariate indices, $\alpha = 0.0027$. The solutions (two for each dimension i) of Equation 9 are (Wang, Hubele, Lawrence, Miskulin, and Shahriari 2000):

$$UPL_i = \mu_i + \sqrt{\frac{\chi_{(v,\alpha)}^2 \det(\boldsymbol{\Sigma}_i^{-1})}{\det(\boldsymbol{\Sigma}^{-1})}} \quad (10)$$

$$LPL_i = \mu_i - \sqrt{\frac{\chi_{(v,\alpha)}^2 \det(\boldsymbol{\Sigma}_i^{-1})}{\det(\boldsymbol{\Sigma}^{-1})}} \quad (11)$$

where $i = 1, 2, \dots, v$ and $\det(\boldsymbol{\Sigma}_i^{-1})$ is the determinant of a matrix obtained from $\boldsymbol{\Sigma}^{-1}$ by deleting the i -th row and column. Thus, the volume of the modified process region is

$$\prod_{i=1}^v (UPL_i - LPL_i) \quad (12)$$

In practice $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. Given a random sample, $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, of size n from the process, to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ the sample mean

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \quad (13)$$

and the sample covariance matrix

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}) (\mathbf{X}_i - \bar{\mathbf{X}})^\top \quad (14)$$

can be used. Rencher (2002) provides a detailed discussion on the properties of the sampling distribution of $\bar{\mathbf{X}}$ and \mathbf{S} , while Vershynin (2012) investigates the optimal sample size that guarantees estimation of a covariance matrix by a sample covariance matrix with a fixed accuracy.

As far as the index is concerned, values of C_{pM} higher than 1 indicate that the modified process region is smaller than the engineering tolerance region, therefore we have high probability that the produced items will be classified as conform.

The second component of the vector (C_{pM}, PV, LI) is defined as the significance level of a Hotelling T^2 statistic computed under the assumption that the center of the engineering specifications is considered to be the true underlying mean of the process:

$$PV = \Pr \left(F_{(v, n-v)} > \frac{(n-v)}{v(n-1)} T^2 \right) \quad (15)$$

where

$$T^2 = n (\bar{\mathbf{X}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \quad (16)$$

and $F_{(v, n-v)}$ is the F distribution with v and $n - v$ degrees of freedom. Values of PV close to zero indicate that the center of the process is far from the engineering target value.

The third component of the vector summarizes a comparison of the location of the modified process region and the tolerance region. It indicates whether any part of the modified process region falls outside the engineering specifications. It has a value of 1 if the entire modified process region is contained within the tolerance region and, otherwise, a value of 0:

$$LI = \begin{cases} 1 & \text{if modified process region is contained within the tolerance region} \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

3.1. The multivariate capability vector using MPCVI

MPCVI consists of the core function `mpci` whose arguments are: `index` which specifies the index that will be computed, in this case `index = "shah"` (the other options are `"taam"`, `"wang"`, `"xeke"`, `"wangw"` and will be illustrated in the following sections); `x` the matrix of the data (the quality characteristics are the columns); `LSL` the vector of the Lower Specification Limits; `USL` the vector of the Upper Specification Limits; `Target` the vector of the target of the process; `alpha` which specifies the quadratic form (9) and conventionally $\alpha = 0.0027$. The function `mpci` uses also `npc`, `Method` and `graphic`. The arguments `npc` and `Method` refer to the indices based on PCA, therefore they will be discussed in Section 5. The logical argument `graphic` allows, in the bivariate case, for additional graphical representations concerning process and specifications and will be introduced in Section 4.

The function is thought to be easy to use, for this reason if `alpha` is missing, then the function assumes as default value $\alpha = 0.0027$ (in analogy with the with the “ 6σ ” in the denominator of the univariate indices). Similarly if the argument `Target`, the Target vector, is missing the program compute it as the midpoint of the specification interval (default value).

To illustrate how **MPCVI** works we use the `dataset1` included in the package.

```
R> library("MPCI")
R> data("dataset1")
R> x <- dataset1
R> Target <- c(30, 70, 15, 12, 120)
R> LSL <- c(24, 60, 10, 8, 100)
R> USL <- c(36, 80, 20, 16, 140)
R> alpha <- 0.0027
R> mpci(index = "shah", x, LSL, USL, Target, alpha)

[[1]]
[1] "Shahriari et al. (1995) Multivariate Capability Vector"
```

```
$CpM
```

```
[1] 1.265159
```

```
$PV
```

```
      [,1]
```

```
[1,] 0.770597
```

```
$LI
```

```
[1] 0
```

Now, the user can interpret the results: $C_{pM} = 1.265159$ indicates that the modified process region is smaller than the engineering tolerance region; $PV = 0.770597$ means that the center of the process is not far of the engineering target value; finally, $LI = 0$ implies that at least in one direction, the modified tolerance region exceeds the specification limits.

The function, without the specification of the arguments `Target` and `alpha` is

```
R> mpci(index = "shah", x, LSL, USL)
```

It works with the default values and, in this example, gives the same results.

4. The multivariate capability index MC_{pm}

The index MC_{pm} was proposed by Taam *et al.* (1993) and is defined as a ratio of two volumes. The numerator is the volume of the modified tolerance region R_1 and the denominator is the volume of the scaled 99.73 percent process region R_2 . Under the multivariate normality hypothesis R_2 is an ellipsoidal process region, while the modified tolerance region R_1 is the largest ellipsoid that is centered at the target completely within the original tolerance region. In the general case of v characteristics the region R_1 is an hyperellipsoid with a volume given by

$$\text{Vol.}(R_1) = \frac{2\pi^{v/2} \prod_{i=1}^v a_i}{v\Gamma\left(\frac{v}{2}\right)} \quad (18)$$

where a_i ($i = 1, 2, \dots, v$) are the lengths of the semi-axes.

Then the multivariate capability index is written as

$$MC_{pm} = \frac{\text{Vol. (modified tolerance region)}}{\text{Vol. } \left((\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}_T^{-1} (\mathbf{X} - \boldsymbol{\mu}) \leq K(m) \right)} \quad (19)$$

where \mathbf{X} is the vector ($v \times 1$) of measurements from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_T = E \left[(\mathbf{X} - \mathbf{T})(\mathbf{X} - \mathbf{T})^\top \right]$ is the mean square error matrix from the process, \mathbf{T} is a vector of target values, and $K(m)$ is a 99.73-th percentile of a χ^2 with v degrees of freedom.

The denominator of MC_{pm} can be also expressed as a product of two terms:

$$\begin{aligned} \text{Vol.}(R_2) &= |\boldsymbol{\Sigma}|^{1/2} (\pi K)^{v/2} [\Gamma(v/2 + 1)]^{-1} \left[1 + (\boldsymbol{\mu} - \mathbf{T})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{T}) \right]^{1/2} \\ &= \text{Vol.}(R_3) \times \left[1 + (\boldsymbol{\mu} - \mathbf{T})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{T}) \right]^{1/2} \end{aligned} \quad (20)$$

where R_3 is the region in which 99.73% of the process values fall within.

Therefore MC_{pm} can be rewritten as:

$$MC_{pm} = \frac{\text{Vol.}(R_1)}{\text{Vol.}(R_3)} \times \frac{1}{\left[1 + (\boldsymbol{\mu} - \mathbf{T})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{T})\right]^{1/2}} = \frac{C_p}{D} \quad (21)$$

The MC_{pm} index is a function of two components: C_p which represents the process variability relative to the modified tolerance region; D which detects the process deviation from the target. Given a random sample of n measurements, $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, each of dimension v , the estimator for MC_{pm} is given by

$$\begin{aligned} \widehat{MC}_{pm} &= \frac{\text{Vol.}(R_1)}{|\mathbf{S}|^{1/2} (\pi K)^{v/2} [\Gamma(v/2 + 1)]^{-1}} \times \frac{1}{\left[1 + \frac{n}{n-1} (\bar{\mathbf{X}} - \mathbf{T})^\top \mathbf{S}^{-1} (\bar{\mathbf{X}} - \mathbf{T})\right]^{1/2}} = \\ &= \frac{\widehat{C}_p}{\widehat{D}} \end{aligned} \quad (22)$$

When the process mean vector equals the target vector, and the index has the value 1, then 99.73% of the process values lie within the modified tolerance region.

4.1. MC_{pm} using MPCl

MPCl computes the index MC_{pm} by setting `index = "taam"`. We illustrate this function using the `dataset2` and introducing `graphic`: an optional argument that, in the bivariate case, allows to plot the ellipse of the process region, the tolerance region, the modified process region and the modified tolerance region. The plot is obtained by setting `graphic = TRUE` (default, `graphic = FALSE`).

```
R> data("dataset2")
R> x <- dataset2
R> alpha <- 0.0027
R> Target <- c(177, 53)
R> LSL <- c(112.7, 32.7)
R> USL <- c(241.3, 73.3)
R> mpci(index = "taam", x, LSL, USL, graphic = TRUE)
```

The function produces Figure 1 and the output is:

```
[[1]]
[1] "Taam et al. (1993) Multivariate Capability Index (MCpm)"

$MCpm
      [,1]
[1,] 1.825283

[1] "CpM index of Shahriari et al. (1995) is the ratio of the
```

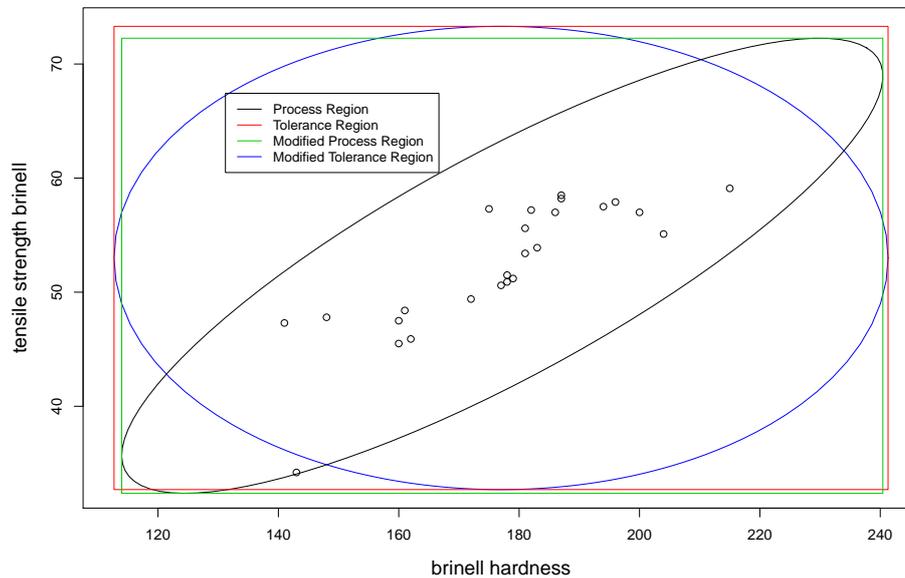


Figure 1: Process region, tolerance region, modified process region, modified tolerance region using `graphic = TRUE`.

Tolerance Region and the Modified Process Region"

[1] "MC_{pm} index of Taam et al. (1993) is the ratio of the ellipsoids: Modified Tolerance Region and the Process Region "

Examining the value of the computed index, $MC_{pm} = 1.825283$, the user can assert that the process is capable since the process has a smaller variation than allowed by the specification limits. Furthermore, by means of the visualization offered by Figure 1, the user has an immediate idea of the process status with respect to the specifications.

The same example without using `graphic` is:

```
R> data("dataset2")
R> x <- dataset2
R> alpha <- 0.0027
R> Target <- c(177,53)
R> LSL <- c(112.7, 32.7)
R> USL <- c(241.3, 73.3)
R> mpci(index = "taam", x, LSL, USL)

[[1]]
[1] "Taam et al. (1993) Multivariate Capability Index (MCpm)"

$MCpm
      [,1]
[1,] 1.825283
```

5. Multivariate capability indices based on PCA

Wang and Chen (1998) proposed a method for constructing multivariate capability indices using PCA. PCA uses the spectral decomposition of the covariance matrix Σ

$$\Sigma = \mathbf{U}\mathbf{D}\mathbf{U}^\top \quad (23)$$

where $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_v)$ is the matrix of eigenvectors of Σ with columns \mathbf{u}_i ($i = 1, 2, \dots, v$) and $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_v)$ is the diagonal matrix of the eigenvalues. In this framework, the engineering specifications and the target values of the i -th principal component (PC _{i}) are

$$LSL_{PC_i} = \mathbf{u}_i^\top \mathbf{L}\mathbf{S}\mathbf{L} \quad (24)$$

$$USL_{PC_i} = \mathbf{u}_i^\top \mathbf{U}\mathbf{S}\mathbf{L} \quad (25)$$

$$T_{PC_i} = \mathbf{u}_i^\top \mathbf{T} \quad (26)$$

for $i = 1, 2, \dots, v$, respectively.

The proposal by Wang and Chen (1998) evaluates the capability of a multivariate process considering a subset m ($m \leq v$) of PCs. They defined MPCIs MC_p , MC_{pk} , MC_{pm} and MC_{pmk} using the univariate process capability indices of the PCs. The multivariate capability index MC_p defined in Wang and Chen (1998) is

$$MC_p = \left(\prod_{i=1}^m C_{p;PC_i} \right)^{1/m} \quad (27)$$

where

$$C_{p;PC_i} = (USL_{PC_i} - LSL_{PC_i}) / 6\sigma_{PC_i} \quad (28)$$

is the univariate measure of capability for the i -th PC, $\sigma_{PC_i} = \sqrt{\lambda_i}$ and m denotes the number of PCs used for assessing the capability. Similarly, they have defined MC_{pk} , MC_{pm} and MC_{pmk} by replacing $C_{p;PC_i}$ with $C_{pk;PC_i}$, $C_{pm;PC_i}$ and $C_{pmk;PC_i}$, respectively, for $i = 1, 2, \dots, m$. Values of the indices greater than 1 indicate that the process is capable.

In order to allow for potential differences in the portion of variance explained by the principal components, Xekalaki and Perakis (2002) suggested a series of new indices that assign unequal weights to the univariate index values corresponding to the principal components employed, in particular in proportion to the percentages of variance explained by them, as determined by their respective eigenvalues. They specifically proposed the following index

$$MXC_p = \frac{\sum_{i=1}^m \lambda_i C_{p;PC_i}}{\sum_{i=1}^m \lambda_i} \quad (29)$$

and offered similar definitions of MXC_{pk} , MXC_{pm} and MXC_{pmk} . Values of the indices greater than 1 indicate a capable process.

Within the framework of a short-run process capability assessment, Wang (2005) proposed the use of the weighted geometric mean. The weights are, once again, the eigenvalues λ_i of each principal component. The index is

$$MWC_p = \left(\prod_{i=1}^m C_{p;PC_i}^{\lambda_i} \right)^{\frac{1}{\sum_{i=1}^m \lambda_i}} \quad (30)$$

The author proposed similar definitions for MWC_{pk} , MWC_{pm} and MWC_{pmk} .

5.1. The indices based on PCA using MPCI

MPCI computes multivariate capability indices according to Wang and Chen (1998), Xekalaki and Perakis (2002) and Wang (2005), by setting `index = "wang"`, `index = "xeke"` and `index = "wangw"` respectively. In this framework a key role is played by the number of principal components used for computing the multivariate indices. Therefore, in order to provide a wide range of options to the users, the package allows different solutions. The user can choose the number of principal components “a priori” by specifying the parameter `npc`. Note that, as guide for this choice, the function `summary(princomp(x))` can be employed to obtain a summary on the importance of the components. Otherwise, by specifying the input argument `Method`, the user can choose one of the following five methods (see e.g., Rencher 2002):

1. Method 1 (Percentage). This method selects the number of principal components that explain at least 80% percent of the total variability.
2. Method 2 (Average). This method works with the principal components whose eigenvalues are greater than the average of the eigenvalues.
3. Method 3 (Scree). This option shows the Scree Diagram and the user can choose the number principal components (Rencher 2002).
4. Method 4 (Bartlett’s test). The number of principal components used for the assessment of the indices is determined through the Bartlett’s test (Bartlett 1950).
5. Method 5 (Anderson’s test). The number of principal components used for the assessment of the indices is determined through the Anderson’s test (Anderson 1933).

Note that, when the function `mpci` computes one of the three indices based on PCA, if `npc` or `Method` are not specified, then `mpci` computes the indices using the Method 1. Furthermore, when the user chooses for `Method = 4` or `Method = 5` the parameter `alpha` corresponds to the significance level used by the corresponding test. If `alpha` is missing, then $\alpha = 0.05$ (default value).

Example with `index = "wang"` (Wang and Chen 1998) using the default `Method = 1`:

```
R> data("dataset2")
R> x <- dataset2
R> Target <- c(177,53)
R> LSL <- c(112.7,32.7)
R> USL <- c(241.3, 73.3)
R> mpci(index = "wang", x, LSL, USL)

[[1]]
[1] "Wang and Chen (1998) Multivariate Process Capability
    Indices(PCI) based on PCA"

$`number of principal components`
[1] 1

$MCp
```

```
[1] 1.180205
```

```
$MCpk
```

```
[1] 1.179954
```

```
$MCpm
```

```
[1] 1.180205
```

```
$MCpmk
```

```
[1] 1.179954
```

In the output the user can find the number principal components selected by the default Method 1 and the values of the indices MC_p , MC_{pk} , MC_{pm} and MC_{pmk} . The interpretation of these indices is analogous to the univariate case. Values higher than 1, as in this case, are interpreted as that the process is capable, whereas values smaller than 1 suggest that some quality improvement activities were required.

Example with `index = "wangw"` (Wang 2005) and `Method = 2`:

```
R> data("dataset1")
R> x <- dataset1
R> Target <- c(30, 70, 15, 12, 120)
R> LSL <- c(24, 60, 10, 8, 100)
R> USL <- c(36, 80, 20, 16, 140)
R> mpci(index = "wangw", x, LSL, USL, Method = 2)

[[1]]
[1] "Wang (2005) Multivariate Process Capability Indices(PCI)
    based on PCA"

$`number of principal components`
[1] 2

$MCp
[1] 2.71438

$MCpk
[1] 2.697944

$MCpm
[1] 2.709994

$MCpmk
[1] 2.693585
```

The interpretation is similar to example illustrated above. Example with `index = "xeke"` (Xekalaki and Perakis 2002) and `Method = 3`:

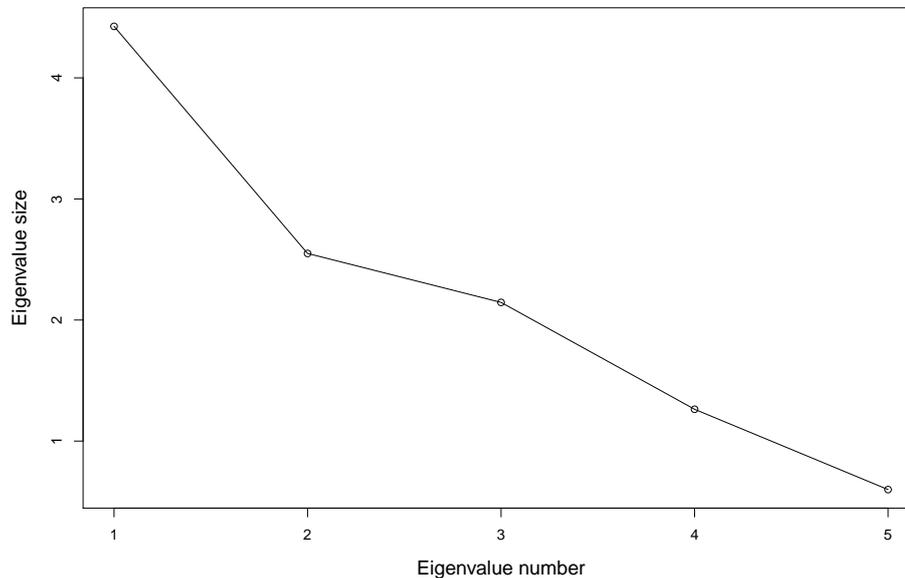


Figure 2: Scree diagram produced by function `xekeMPCI`.

```
R> data("dataset1")
R> x <- dataset1
R> Target <- c(30, 70, 15, 12, 120)
R> LSL <- c(24, 60, 10, 8, 100)
R> USL <- c(36, 80, 20, 16, 140)
R> mpci(index = "xeke", x, LSL, USL, Method = 3)
```

The function produces Figure 2 and asks to enter the number of principal components. Entering 2 the output is

Enter the number of principal component(npc) according to the scree graph:

1: 2

Read 1 item

```
[[1]]
```

```
[1] "Xekalaki and Perakis (2002)
```

```
      Multivariate Process Capability Indices(PCI) based on PCA"
```

```
$`number of principal components`
```

```
[1] 2
```

```
$MCp
```

```
[1] 2.868629
```

```
$MCpk
```

```
[1] 2.850806
```

```
$MCpm
```

```
[1] 2.863180
```

```
$MCpmk
[1] 2.845392
```

Using the same data and specifications but choosing `Method = 4`:

```
R> mpci(index = "xeke", x, LSL, USL, Method = 4)
```

we obtain:

```
[[1]]
[1] "Xekalaki and Perakis (2002)
     Multivariate Process Capability Indices(PCI) based on PCA"

$`number of principal components`
[1] 4
$MCp
[1] 2.119719
$MCpk
[1] 2.102786
$MCpm
[1] 2.115772
$MCpmk
[1] 2.09887
```

6. Concluding remarks

Processes with multiple quality characteristics often occur in manufacturing industries, consequently quantitative measures of process performance for multivariate quality characteristics are of great interest to quality control practitioners.

The R package **MPCI** provides an easy-to-use tool for computing multivariate process capability indices. To our knowledge, **MPCI** is the first available R software for computing multivariate process capability indices. Our aim is to provide useful tools for performing multivariate process capability analysis. In such a way we hope to contribute to the diffusion of multivariate process capability analysis. **MPCI** is still undergoing active development. Planned extensions include the multivariate indices based on the proportion of nonconforming items – see [Chen \(1994\)](#) and [Chen *et al.* \(2003\)](#) – and several new capability indices [Pan and Lee \(2010\)](#) recently proposed.

Acknowledgments

We are grateful to the Associate Editor and two reviewers for their thoughtful and constructive comments. This research was partially supported by a 2008 grant (project no. 2008CEFF37 001, sector: Economics and Statistics) for research projects of national interest, provided by the Italian Ministry of Universities and Scientific and Technological Research.

References

- Anderson TW (1933). “Asymptotic Theory for Principal Component Analysis.” *The Annals of Mathematical Statistics*, **34**, 122–148.
- Bartlett MS (1950). “Tests of Significance in Factor Analysis.” *The British Journal of Psychology*, **3**, 77–85.
- Chen H (1994). “A Multivariate Process Capability Index over a Rectangular Solid Zone.” *Statistica Sinica*, **4**, 749–758.
- Chen KS, Pearn WL, Lin PC (2003). “Capability Measures for Processes with Multivariate Characteristics.” *Quality and Reliability Engineering International*, **19**, 101–110.
- Hubele NF, Shahriari H, Cheng CS (1991). “A Bivariate Process Capability Vector.” In JB Keats, DC Montgomery (eds.), *Statistical Process Control in Manufacturing*, pp. 299–310. Marcel Dekker, New York.
- Kotz S, Lovelace CB (1998). *Process Capability Indices in Theory and Practice*. Arnold, London.
- Nickerson D (1994). “Construction of a Conservative Confidence Region from Projections of an Exact Confidence Region in Multiple Linear Regression.” *The American Statistician*, **48**(2), 120–124.
- Pan JN, Lee CY (2010). “New Capability Indices for Evaluating the Performance of Multivariate Manufacturing Processes.” *Quality and Reliability Engineering International*, **26**, 3–15.
- Pearn WL, Wang FK, Yen CH (2007). “Multivariate Capability Indices: Distributional and Inferential Properties.” *Journal of Applied Statistics*, **34**, 941–962.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rencher AC (2002). *Methods of Multivariate Analysis*. 2nd edition. John Wiley & Sons, New York.
- Scagliarini M, Vermiglio R (2008). “Computing Multivariate Process Capability Indices (Excel).” *Journal of Modern Applied Statistical Methods*, **7**, 346–357.
- Shahriari H, Hubele NF, Lawrence FP (1995). “A Multivariate Process Capability Vector.” *Proceedings of the 4th Industrial Engineering Research Conference*, **1**, 304–309.
- Taam W, Subbaiah P, Liddy WJ (1993). “A Note on Multivariate Capability Indices.” *Journal of Applied Statistics*, **20**, 339–351.
- Vershynin R (2012). “How Close is the Sample Covariance Matrix to the Actual Covariance Matrix?” *Journal of Theoretical Probability*. doi:10.1007/s10959-010-0338-z. Forthcoming.

- Wang CH (2005). “Constructing Multivariate Process Capability Indices for Short-Run Production.” *The International Journal of Advanced Manufacturing Technology*, **26**, 1306–1311.
- Wang FK, Chen JC (1998). “Capability Index Using Principal Components Analysis.” *Quality Engineering*, **11**, 21–27.
- Wang FK, Hubele NF, Lawrence FP, Miskulin JD, Shahriari H (2000). “Comparison of Three Multivariate Process Capability Indices.” *Journal of Quality Technology*, **32**, 263–275.
- Xekalaki E, Perakis M (2002). “The Use of Principal Component Analysis in the Assessment of Process Capability Indices.” In *Proceedings of the Joint Statistical Meetings of the American Statistical Association, The Institute of Mathematical Statistics, The Canadian Statistical Society. New York*, pp. 3819–3823.

Affiliation:

Edgar Santos-Fernández
Universidad Central “Marta Abreu” de Las Villas
D St. 17. Santa Catalina. Santa Clara
Villa Clara 50300, Cuba
E-mail: edgar.santos@etecca.cu

Michele Scagliarini
Dipartimento di Scienze Statistiche
Università di Bologna
Via Belle Arti 41
40126 Bologna, Italy
E-mail: michele.scagliarini@unibo.it
URL: <http://www2.stat.unibo.it/scagliarini/>