

Capability Analysis for Profiles

RUBÉN DARÍO GUEVARA GONZÁLEZ



NATIONAL UNIVERSITY OF COLOMBIA
FACULTY OF SCIENCE
DEPARTMENT OF STATISTICS
BOGOTÁ D.C., COLOMBIA
FEBRUARY 2014

Capability Analysis for Profiles

RUBÉN DARÍO GUEVARA GONZÁLEZ

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR IN SCIENCE - STATISTICS

ADVISOR
JOSÉ ALBERTO VARGAS NAVAS, PHD



NATIONAL UNIVERSITY OF COLOMBIA
FACULTY OF SCIENCE
DEPARTMENT OF STATISTICS
BOGOTÁ D.C., COLOMBIA
FEBRUARY 2014

Title in English

Process Capability Analysis for Profiles.

Título en español

Análisis de la Capacidad de Proceso para Perfiles.

Abstract: There are practical situations in which the quality of a process or product can be better characterized by a functional relationship between a response variable and one or more explanatory variables, this is called profile. Such profiles usually can be represented adequately using linear or nonlinear models. While there are several studies monitoring profiles, there are few studies to evaluate the capability of a process with profile quality characteristic, particularly if the process is characterized by a nonlinear functional relationship. This dissertation introduces methods to evaluate the capability of processes characterized by nonlinear profiles (univariate or multivariate), without distributional assumptions.

We propose two methods to measure the capability of processes characterized by univariate nonlinear profiles based on the concept of functional depth. These methods extend to functional data, the Process Capability Indexes proposed by Clements for measuring the capability of a process characterized by a random variable.

To evaluate the capability of processes characterized by multivariate nonlinear profiles, we consider each observation as a finite dimension vector whose elements are functions. Initially, we transform the original functional data into uncorrelated functions using a dimension reduction technique for multivariate functional data. Next, the capability for each functional component is evaluated. Two sets of process capability indices to measure the capability of these functional components are proposed, having into account if the random errors follow (or not) a multivariate normal distribution. Within case where the random errors do not follow a multivariate normal distribution, we use a method based on the concept of functional depth and apply the methods proposed for the case of univariate nonlinear profiles.

Performance of the methods proposed is evaluated through simulation studies. Examples illustrate the applicability of these methods. We offer conclusions and advice for future research at the end.

Resumen: Hay situaciones prácticas donde la calidad de un proceso o producto está mejor caracterizada por una relación funcional entre una variable de respuesta y una o más variables explicatorias, la cual es llamada perfil. Tales perfiles pueden ser representados usando modelos lineales o no lineales. Mientras que existen diferentes estudios en monitoreo de perfiles, hay pocos estudios para evaluar la capacidad de un proceso cuya característica de calidad es un perfil, particularmente de tipo no lineal. Esta disertación presenta métodos para evaluar la capacidad de estos procesos (univariados o multivariados), los cuales no emplean supuestos distribucionales.

Basados en el concepto de profundidad funcional, dos métodos para medir la capacidad de procesos caracterizados por perfiles no lineales univariados son propuestos. Estos métodos extienden al campo de los datos funcionales los índices propuestos por Clements para medir la capacidad de procesos caracterizados por una variable aleatoria.

Para evaluar la capacidad de procesos caracterizados por perfiles no lineales multivariados, cada observación es considerada como un vector de dimensión finita cuyos elementos son funciones. Inicialmente, los datos funcionales originales son transformados en funciones no correlacionadas usando una técnica de reducción para datos funcionales multivariados. A continuación, la capacidad para cada componente funcional es evaluada. Dos conjuntos de índices para medir la capacidad de estos componentes funcionales son propuestos, dependiendo si los errores aleatorios siguen o no una distribución normal multivariada. Para el caso donde los errores aleatorios

siguen una distribución multivariada no normal, un método basado en el concepto de profundidad funcional es propuesto.

El desempeño de los métodos propuestos es evaluado a través de estudios de simulación y su aplicabilidad es ilustrada en algunos ejemplos. Conclusiones y recomendaciones para futuras investigaciones son presentadas al final del documento.

Keywords: Functional Data, Functional Depth, Multivariate Functional Data, Multivariate Functional Principal Component Analysis, Nonlinear Profiles, Process Capability Analysis, Process Capability Indices.

Palabras clave: Datos Funcionales, Profundidad Funcional, Datos Funcionales Multivariados, Análisis de Componentes Principales para Funcionales Multivariados, Perfiles no Lineales, Análisis de Capacidad de Proceso, Índices de Capacidad de Proceso.

Acceptation Note

Thesis Work

“ mention”

Jury

Jury

Jury

Advisor
José Alberto Vargas N.

Bogotá, D.C., February/2014

Dedication

To my wife Olga, my son Darío, and my daughter Laura, and to all my family.

Acknowledgements

I would like to thank my wife Olga, my son Darío and, my daughter Laura for their encouragement, support and love.

I would like to specially thank my advisor, Dr. José Alberto Vargas Navas, for his help, guidance, commitment and encouragement. This dissertation could not have been completed without his insightful advice. I consider such an honor to have an advisor and friend like him.

I would also like to thank my committee members for their helpful suggestions and comments on the dissertation. Besides, I would like to thank Dr. Philippe Castagliola for his commentaries about the Hausdorff distance to measure the capability of processes characterized by nonlinear profiles.

I would also like to extend many thanks to the Statistic Department of the Universidad Nacional de Colombia for their great support and encouragement. I particularly want to express my gratitude to my professors for sharing their knowledge and friendship. I also commend the friendship of my partners. I am more than happy knowing I can count on them.

Contents

Contents	I
List of Tables	III
List of Figures	IV
Introduction	VII
1. Literature Review	1
1.1 Process capability	1
1.1.1 Process capability index	1
1.1.2 Multivariate process capability using principal components	3
1.1.3 Process capability analysis for profiles	4
1.2 Functional data	6
1.2.1 The concept of depth for functional data	6
1.2.2 Band depth for functional data	7
1.2.3 Functional outliers	9
1.2.4 Principal component for multivariate functional data	11
1.3 Hausdorff distance	14
2. Process Capability Analysis for NonLinear Profiles	16
2.1 Process capability of nonlinear profiles using depth functions	16
2.1.1 Simulations	20
2.1.2 Example	24
2.1.3 Conclusions	35
2.2 Evaluation of process capability in nonlinear profiles using Hausdorff distance . . .	35
2.2.1 Simulations	36

2.2.2	Example	37
2.2.3	Conclusions	40
3.	Process Capability Analysis for Multivariate NonLinear Profiles	41
3.1	Evaluation of Process Capability in Multivariate NonLinear Profiles	41
3.2	Simulations	44
3.3	Example	51
3.4	Conclusions	52
	Conclusions and future work	55
	Bibliography	57

List of Tables

2.1	Mean and standard deviation of $\widehat{C}_{pl(prf)}$ when the target PCI is 0,67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for Normal(0, 0.1) distribution.	25
2.2	Mean and standard deviation of $\widehat{C}_{pl(prf)}$ when the target PCI is 0,67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for LogNormal(0, 0.3) distribution.	26
2.3	Mean and standard deviation of $\widehat{C}_{pl(prf)}$ when the target PCI is 0,67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for Weibull(6,1) distribution.	27
2.4	Mean and standard deviation of $\widehat{C}_{pl(prf)}$ when the target PCI is 0,67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for Gamma(1.2, 5) distribution.	28
2.5	Mean and standard deviation of $\widehat{C}_{pl(prf)}$ when the target PCI is 0,67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for Beta(5, 5) distribution.	29
2.6	Relative bias of the several estimators of $C_{pl(prf)}$ when the target PCI is 0,67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for Normal(0, 0.1) distribution.	30
2.7	Relative bias of the several estimators of $C_{pl(prf)}$ when the target PCI is 0,67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for Gamma(1.2, 5) distribution.	31
2.8	Estimations and confidence intervals of $C_{pk(prf)}$ for the VDP data.	35
2.9	Execution time in seconds for the estimation of the median using seven estimators.	37
2.10	Relative bias and standard deviation (in gray) of seven estimators of $C_{pl(prf)}$ corresponding to the four-parameter logistic model, under five distributions of the error term, when the target PCI is 1 and $m = 30, 60, 90$ and 120 profiles.	38
2.11	Estimations of $C_{pk(prf)}$ for the VDP data.	40
3.1	Mean and standard deviation(gray colour) of two estimators of C_{pu} for the process generated by the equations (3.14), (3.15) and (3.16), with target $C_{pu} = 1$, $p = 2, 5, 10$ and 30 , and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9	46

List of Figures

1.1	Raw data (A), Functional bagplot with factor=2.58 (B), functional HDR boxplot for $\alpha = 0.01$ (C), and functional boxplot with factor=1.5 (D) of the sea surface temperatures measured monthly in degrees Celsius over the east-central tropical Pacific Ocean from 1951 to 2007.	12
	a	12
	b Functional bagplot	12
	c Functional HDR boxplot	12
	d Functional boxplot	12
2.1	Estimations of $Y_{0.5}$ (solid line), $Y_{0.99865}$ (dashed line) and $Y_{0.00135}$ (dotted line) using modified band depth.	18
2.2	Behaviour of several estimations of the median function. The red curve represents the true median curve, green curve is based on MBD, orange curve is supported by BD3, pink curve relies on subintervals with MBD, magenta curve uses the trimmed mean with deepest m curves, and black curve is based on pointwise estimations. Specification limits with blue curves.	19
2.3	30 Profiles from the 4-parameter logistic model under five distributions of the error term.	22
	a Normal(0,0.1) distribution	22
	b Beta(5,5) distribution	22
	c Gamma(1.2,5) distribution	22
	d Lognormal(0,0.3) distribution	22
	e Weibull(6,1) distribution	22
2.4	Median function (in black), $Y_{0.99865}$, $Y_{0.00135}$ (in blue), and specification functions (in red) for the 4-parameter logistic model using five distributions when $C_{pu} = 1.33$ and $C_{pl} = 1.33$	23
	a Normal(0,0.1) distribution	23
	b Beta(5,5) distribution	23
	c Gamma(1.2,5) distribution	23

d	Lognormal(0,0.3) distribution	23
e	Weibull(6,1) distribution	23
2.5	Vertical Density Profile of 24 particleboards.	32
2.6	Spline fits for the Vertical Density Profile data.	32
2.7	Three graphical methods for visualizing VDP data: a) Functional bagplot with factor=2.58. The red and turquoise curves are the outliers detected. b)The functional HDR boxplot for $\alpha = 0.01$. The red and turquoise curves are the outliers detected. c) The functional boxplot with factor=1.5. The red curve is the outlier detected.	33
a	Functional bagplot	33
b	Functional HDR boxplot	33
c	Functional boxplot	33
2.8	Spline fit for the Vertical Density Profile (in black), estimation of the median function using MBD (in green), estimations of the quantiles 0.99865 and 0.00135 using MBD (in orange), and specification functions (in blue).	34
2.9	Vertical Density Profile of 23 particleboards (black curves), with estimations of the median function (red curves) and estimations of the quantiles 0.99865 (green curves) and 0.00135 (blue curves) using maximum operator (on the top) and median (on the bottom) on the Hausdorff distance matrix between profiles.	39
3.1	Data generated from equations (3.14), (3.15) and (3.16) with $p = 5$, $\rho = 0.1$ (left) and $\rho = 0.9$ (right). The black curves are realizations of Y_1 , the red of Y_2 , the green of Y_3 , the blue of Y_4 , and the magenta of Y_5	45
3.2	Relative bias of $\widehat{MC}_{pl(prf)}$ (solid lines) and $\widehat{MC}_{pl(prf)}^\dagger$ (dotted lines) for the multivariate process generated by equations (3.14), (3.15) and (3.16), with target $C_{pl} = 1$, $p = 2$ (black lines), 5 (red lines), 10 (green lines) and 30 (blue lines), and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9	46
3.3	Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) o for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim N_p(0, \Sigma_{\varepsilon(t)})$, where $\Sigma_{\varepsilon(t)}$ is given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9	47
3.4	Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) o for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim t_p(3)$, $\Sigma_{\varepsilon(t)}$ given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9	48

3.5	Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) or for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim \text{gamma}_p(1.9, 4)$, $\Sigma_{\varepsilon(t)}$ given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9	49
3.6	Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) or for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim \text{beta}_p(1, 2)$, $\Sigma_{\varepsilon(t)}$ given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9	50
3.7	First principal component for $p = 10$ with $\rho = 0.9, 0.7, 0.5$ and 0.3 , and $\varepsilon \sim \text{beta}_p(1, 2)$	51
a	$\rho = 0.9$	51
b	$\rho = 0.7$	51
c	$\rho = 0.5$	51
d	$\rho = 0.3$	51
3.8	Raw fluorescence emission spectra sugar samples without outliers, sampled as a mean spanning eight hours equal to one shift during a three-month campaign (1995). Emission ranges were all 275 – 560 nm. The samples were measured at excitation wavelengths 305 and 325 nm, which constitute Y_5 (left) and Y_6 (right), respectively.	53
3.9	Upper specification function (red curve), estimation of the 0.99865 quantile function (green curve), estimation of the median function (orange curve) and fluorescence emission spectra (black curves) for Y_5 (left) and Y_6 (right), respectively.	53

Introduction

Process capability is the ability of a process to consistently meet specified customer-driven requirements. In general, a process is capable if the probability of obtaining nonconforming items (outside some specifications) is small, for example 0.0027. One metric commonly used to measure the process capability is the Process Capability Index (PCI). The PCIs are unitless and essentially measure the variability of a process relative to its specification limits. They have been widely used in different kind of industries such as the automotive, chemical, electronic or microelectronic and pharmaceutical. The PCIs have meaning if the process is in statistical control.

Traditional PCIs for processes with unilateral specifications are:

$$C_{pu} = \frac{USL - \mu}{3\sigma}, \quad (1)$$

and

$$C_{pl} = \frac{\mu - LSL}{3\sigma}, \quad (2)$$

where USL and LSL are the upper and lower specification limits, respectively; the parameters μ and σ are the mean and the standard deviation of the process, respectively. For processes with bilateral specifications, the traditional PCIs are:

$$C_p = \frac{USL - LSL}{6\sigma}, \quad (3)$$

and

$$C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right). \quad (4)$$

If all these indices have values larger than a specified threshold (1, 1.33 or 1.66 depending of the expected quality of the process), the process is said "capable".

These PCIs require that the distribution of the process to be normal. However, most of the processes that happen in the real world generate non-normal data, and the use of these PCIs based on the assumption of normality may yield misleading results. The first and simplest approach to deal with non-normal data is to transform this data into normally distributed data, and then use the classical indices to measure the capability. A second approach is to adjust a more appropriate parametric distribution. This can be attained approximating the distribution with some other parametric model. Within this approach the method of non-normal quantiles calculates the PCIs for a distribution of any shape, using the Pearson family of curves, proposed by [14]. The main advantage of this method is that no complicated distribution fitting is required and it is simple to use. Using Clements's proposal, the alternate definitions for the indices C_p and C_{pk} , are given by:

$$C_{p(Q)} = \frac{USL - LSL}{Y_{0.99865} - Y_{0.00135}}, \quad C_{pk(Q)} = \min(C_{pu(Q)}, C_{pl(Q)}) \quad (5)$$

where

$$C_{pu(Q)} = \frac{USL - Y_{0.5}}{Y_{0.99865} - Y_{0.5}}, \quad C_{pl(Q)} = \frac{Y_{0.5} - LSL}{Y_{0.5} - Y_{0.00135}}, \quad (6)$$

and Y_p is the p -quantile estimated from the process data.

When the process is described by two or more characteristics, several approaches of multivariate process capability indices (MPCIs) have been proposed to measure the capability of the process to meet the assigned specifications for each characteristic. In general, these MPCIs are constructed using principal component analysis (PCA), the ratio of the volume of a tolerance region to the volume of a process region or the proportion of nonconforming items.

Generally, the quality of a process or a product is represented by the distribution of an univariate or a multivariate quality characteristic. Anyhow, there are numerous processes where the quality is better characterized by a relationship between a response variable and one or more explanatory variables. This relationship usually is known as a profile. Such profiles can be represented using linear or nonlinear models. Some discussions concerning the general issues involving profile monitoring can be found in [110], [109] and [65].

The number of researches to measure the capability of processes characterized by profiles (linear or nonlinear) is scarce. For the case of nonlinear profiles, as far as we are concerned, the proposal presented by [103] is unique in this area. They evaluated the process yield of nonlinear profiles in manufacturing processes using the index S_{pk} proposed by [6], under the assumption that the process is normally distributed at the j th level of the independent variable, and that observations between levels are independent. For multivariate nonlinear profiles there is not research yet.

The goal of this dissertation is to evaluate the capability of processes characterized by univariate or multivariate nonlinear profiles, without distributional assumptions and taking into account the within-profile correlation. To this end, we focus on functional methods although most of the current work on profile in statistical process control relies fundamentally on procedures from classical regression and multivariate analysis.

Thanks to the quick progress in sensor and information technology and, the increase in storage capabilities and processing power of computers, functional data has become a topic of active statistical research. Functional data can be found in different disciplines such as medicine, biology, economics, chemometrics and engineering.

In the first part of this dissertation we extend to functional data the PCIs proposed by [14] (see equations (1.5) and (1.6)) that measure the univariate process capability. We assume that are working in the space of real continuous functions defined on a compact interval, where the curve of each profile is the realization of a univariate stochastic process in statistical control. With these assumptions, two methods to measure the capability of processes characterized by nonlinear profiles are proposed. In the first method, a discretized version of the process is considered and the PCIs are defined as the arithmetic mean of the one-dimensional PCIs evaluated for each point in the compact interval. In the second method, the PCIs are defined as the ratio of two integrals in view of the specification limits and the 0.99865, 0.5 and 0.00135 quantiles are continuous functions. In both methods, the proposed indices quantify the relationship between the actual performance of the nonlinear profiles and the specification limits.

The indices proposed can be estimated replacing the 0.99865, 0.5 and 0.00135 quantile functions by appropriate estimators that rely on sample observations. The estimation of these quantiles implies to solve the next question: How to order a sample of curves?. The notion of depth for functional data provides a center-outward ordering of the set of curves given.

Particularly, [57] proposed a notion of depth for functional data based on the graphic representation of the functions. This proposal, called band depth (BD), uses the bands defined by their graphs on the plane. A generalized version, called modified band depth (MBD), which consider the proportion of time that the curve of a function is inside the band, was proposed by the same authors. In this dissertation we use these two versions of depth for functional data to estimate the 0.99865, 0.5 and 0.00135 quantile functions.

Using the concepts of BD and MBD estimations of the 0.5 quantil function (sample median) are found. A sample median function is a curve from the sample with the highest depth value. If this is not unique, the median will be the average of the curves maximizing depth.

In multivariate data an ordinary q th quantile is a boundary point of a p th quantile inner region (p th central region) when q and p satisfy $p = \lfloor 2q - 1 \rfloor$. Extending this procedure to functional data, the estimations of 0.99865 and 0.00135 quantile functions are given by the functions that envelope the 0.9973 central region, which contains 99.73% of deepest curves from the sample.

In the first proposed method, a critical situation occurs when at a point, the estimated median and the estimated 0.99865 quantile function coincide or have similar values because the estimated C_{pu} in that point is infinite or it takes very large values. A similar situation occurs for C_{pl} . To solve this difficulty an estimator of the median function based on the functional trimmed mean is considered. Also, estimators based on BD and MBD obtained by subintervals are analyzed. The performance of the proposed estimators is evaluated through simulation studies for the two methods considered .

The estimators proposed based on BD or MBD require an intensive use of computational resource. For this reason, an alternative method based on the Hausdorff distance to evaluate process capability indices characterized by nonlinear profiles is analyzed. Using this distance, estimations of the 0.99865, 0.5 and 0.00135 quantiles functions are found and used to calculate the PCIs. Through simulation, we compare the accuracy, precision and execution time of the estimators based on the Hausdorff distance with the results obtained by the estimators based on the concept of band depth or modified band depth.

In the second part of this dissertation, we evaluate the capability of processes characterized by multivariate nonlinear profiles, where each observation is a finite dimension vector whose elements are functions. This proposal is an extension of the second method presented in the first part.

Initially, we transform the original functional data into uncorrelated functions using principal components for functional data on accord with the method presented by [4]. In this proposal, for each value of the domain on which the functions are defined, a classical multivariate PCA is carried out. Under this proposal, the principal components are linear combinations of the original functions in the data set. However, the weights for each value of the domain are unique only up to a change of sign; therefore, the principal components are not unique. To correct this problem, [4] proposed a criterion for choosing the sign so that the resulting components are smooth. The application of this procedure in processes with bilateral specifications, could do that for some values of the domain $USL > LSL$, while for the rest this relation change. A similar situation happens for processes with unilateral specification. This behaviour of the specification functions complicates the interpretation of the capability in the functional components. To overcome this drawback, we propose a method that preserves the relations between the specifications.

Once the functional components have been calculated, those that explain an appreciable percentage of the variance of the process, for example at least 90%, are identified and its capability is evaluated using the second method presented in the first part of this dissertation. Next, an overall PCI is obtained to measure the capability of the process. We consider several overall PCI. The performance of the proposed process capability indices is evaluated through simulation studies. An example of the sugar production which illustrates the applicability of the proposed method is showed here.

This dissertation is organized like this: Chapter 1 provides literature reviews on process capability, functional data, and the Hausdorff distance. First, we describe the main indices used to measure the capability of a process when it is characterized by a random variable, two or more characteristics, or a profile. The section associated to the functional data emphasizes about the depth for functional observations, functional outliers and a dimension reduction technique for multivariate functional data. Finally, we present the definition and properties of the Hausdorff distance, which constitutes a suitable distance for lines, curves or surfaces.

Chapter 2 provides a methodology to measure process capability characterized by nonlinear profiles using the concepts of band depth and modified band depth. This proposal does not have distributional assumptions about the data and has into account the correlation within the profile. Here we extend to functional data the Process Capability Indexes proposed by [14] to measure the capability of a process characterized by a random variable. The performance of the proposed estimators is evaluated through simulation studies. An example illustrates the applicability of these methods. This chapter is based on the paper "Process Capability Analysis for NonLinear Profiles Using Depth Functions" ([33]), published on-line by Quality and Reliability Engineering International.

An alternative method based on the Hausdorff distance to measure the process capability characterized by nonlinear profiles is presented in the section 2 of the chapter 2. The accuracy, precision and execution time of the estimators based on the Hausdorff distance is compared with the estimators based on the concept of band depth or modified band depth, presented in the section before. This section is based on the paper "Evaluation of Process Capability in Nonlinear Profiles Using Hausdorff Distance" ([35]), submitted for publication.

In Chapter 3 a method to analyze process capability characterized by multivariate nonlinear profiles is presented. This is based on a technique of principal component for multivariate functional data and the concept of functional depth. The performance of the proposed method is evaluated through simulation studies. An example about the production of sugar illustrates this method. This chapter is supported by the manuscript "Evaluation of Process Capability in Multivariate NonLinear Profiles" ([34]), submitted for publication.

Finally, we finish this dissertation with a summary of the contributions of this research and provides ideas for future works.

Literature Review

In this chapter the concepts used to evaluate the process capability characterized by nonlinear profiles are described. These are organized in three subsections: process capability, functional data and hausdorff distance.

1.1 Process capability

A process can be viewed as a series of actions or operations influenced by several elements or factors, all contributing to the eventual outcome. These elements or causes of variation can be generally broken into the following typical categories: Material, machine, method, manpower, environment and measurement, [76].

Process capability is the ability of a process to consistently meet specified customer-driven requirements. In general, a process is capable if the probability of obtaining nonconforming items (outside some specifications) is small, typically 0.0027. [70] emphasize the basic assumptions that a conventional capability analysis has:

- Process stability: The process is in a state of statistical control, no special causes are present, the process does no drift nor oscillate, and so forth.
- Representative samples: Obtained samples are representative of the population.
- Normality: The underlying process is normal.
- Independence: Observations are independent of each other.

1.1.1 Process capability index

A commonly used metric to measure the process capability is the Process Capability Index (PCI). Essentially a PCI measures the variability of a process relative to its specification limits. A PCI is an index of the quality of the process that measures the risk of producing defective articles due to the natural variability of the process, [89], [42], [1] and [32]. Capability indices are unitless and associate the process location and variance with one-sided or two-sided specifications, with or without a target value for the process, [90].

The first process capability index appearing in the literature was the precision index, C_p , [46], defined as:

$$C_p = \frac{USL - LSL}{6\sigma}, \quad (1.1)$$

where USL and LSL are the upper and lower specification limits and σ is the process standard deviation.

As C_p does not adequately deal with cases where process mean, μ , is not equal to the midpoint of the specification interval, [47] developed the index C_{pk} , defined as:

$$C_{pk} = \min \left\{ \frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma} \right\}. \quad (1.2)$$

For processes with bilateral specifications we only consider the indices C_p and C_{pk} , although in the literature there are more indices, for example the indices C_{pm} and C_{pmk} , which are defined as:

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} \quad \text{and} \quad C_{pmk} = \min \left\{ \frac{USL - \mu}{3\sqrt{\sigma^2 + (\mu - T)^2}}, \frac{\mu - LSL}{3\sqrt{\sigma^2 + (\mu - T)^2}} \right\},$$

where T is the target value.

When the process or product has a unique specification, we must consider the indices C_{pl} and C_{pu} . The process capability indices C_{pl} and C_{pu} , proposed by [47], measure larger-the-better and smaller-the-better process capabilities and are defined as:

$$C_{pl} = \frac{\mu - LSL}{3\sigma}, \quad C_{pu} = \frac{USL - \mu}{3\sigma}. \quad (1.3)$$

The indices C_p and C_{pk} can be expressed in function of the indices C_{pl} and C_{pu} through:

$$C_p = \frac{1}{2} (C_{pu} + C_{pl}), \quad C_{pk} = \min(C_{pu}, C_{pl}). \quad (1.4)$$

The indices C_{pu} , C_{pl} , C_p and C_{pk} , considered as the traditional PCIs, require that the distribution of the process to be normal and the process has symmetric tolerance, that is, the target value coincides with the midpoint of the specification interval. However, most of the processes that occur in the real world generate non-normal data, and the use of these PCIs based on the assumption of normality may yield misleading results. The simplest approach to deal with non-normal data is to transform the data into normally distributed data, and then use the classical indices to measure the capability. Following this approach, [45] built a system of distributions based on the moment method, [5] presented a family of power transformations, and [88] used a square-root transformation.

A second approach is to adjust a more appropriate parametric distribution. This can be attained by approximating the distribution with some other parametric model. Proposals presented by [9] and [14] are examples of this approach. [9] proposed estimating the proportion of non-conforming items, using a method based on Burr's distributions, and then converting this proportion to some traditional PCI. [14] proposed the method of non-normal quantiles to calculate the PCIs for a

distribution of any shape, using the Pearson family of curves. The main advantage of this method is that no complicated distribution fitting is required and it is also simple to use. By using Clements's proposal, the alternate definitions for the indices C_p and C_{pk} , are given by:

$$C_{p(Q)} = \frac{USL - LSL}{Y_{0.99865} - Y_{0.00135}}, \quad C_{pk(Q)} = \min(C_{pu(Q)}, C_{pl(Q)}) \quad (1.5)$$

where

$$C_{pu(Q)} = \frac{USL - Y_{0.5}}{Y_{0.99865} - Y_{0.5}}, \quad C_{pl(Q)} = \frac{Y_{0.5} - LSL}{Y_{0.5} - Y_{0.00135}}, \quad (1.6)$$

and Y_p is the p -quantile estimated from the process data.

1.1.2 Multivariate process capability using principal components

When the process is described by two or more characteristics, several approaches of multivariate process capability indices (MPCIs) have been proposed to measure the capability of the process to meet the assigned specifications for each characteristic. In general, these MPCIs are constructed using: a) Principal component analysis (PCA), see [99], [100], [98], [80] and [69]; b) the ratio of the volume of a tolerance region to the volume of a process region such as the papers proposed by [93], [84], [101], [85] and [68]; and the proportion of nonconforming items, such as the works of [106], [12], [67], [71] and [10]. In this subsection some approaches based on PCA are explained.

[99] applied PCA to assess the capability of a multivariate normal process with ν quality characteristics. Using PCA, the original variables are projected onto new ν independent variables, called principal components PC , which are linear combinations of the original variables. The variance of each PC_j is equal to its eigenvalue λ_j , and together explain the variability in the data. The original multivariate specifications, **USL** and **LSL**, are projected into each PC_j . These projections constitute the specifications for each PC_j . [99] proposed an MPCI obtained as the geometric median of the capability indices evaluated on the principal components. This index is defined as:

$$MC_p = \left(\prod_{j=1}^q C_{p:PC_j} \right)^{1/q},$$

where $C_{p:PC_j} = \frac{USL_{PC_j} - LSL_{PC_j}}{6\sqrt{\lambda_j}}$ represents the univariate index C_p for the j th principal component, $USL_{PC_j} = \mathbf{u}'_j \mathbf{USL}$ and $LSL_{PC_j} = \mathbf{u}'_j \mathbf{LSL}$ are the specifications for PC_j , \mathbf{u}_j is the j th eigenvector of the covariance matrix Σ of the process, and q denotes the number of principal components used to assess the capability. If $C_{p:PC_j}$ is replaced by $C_{pk:PC_j}$, $C_{pm:PC_j}$ or $C_{pmk:PC_j}$ the indices MC_{pk} , MC_{pm} or MC_{pmk} are obtained, respectively.

[100] proposed an MPCI to evaluate the capability in multivariate non-normal processes. This index called MC_{pc} is calculated by the following equation:

$$MC_{pc} = \left(\prod_{j=1}^q C_{pc:PC_j} \right)^{1/q}$$

where $C_{pc:PC_j} = (USL_{PC_j} - LSL_{PC_j}) / 6\sqrt{\pi/2\bar{c}_j}$ and $\bar{c}_j = \frac{1}{n} \sum_{i=1}^n \left| PC_{ji} - \frac{USL_{PC_j} + LSL_{PC_j}}{2} \right|$.

The previous indices have the drawback that all the PCs have equal weight, but the variability of the data explained for each of them is different. [98] proposed an index using a weighted geometric mean, where the weights are the eigenvalues of each PC, λ_j . This index is defined as:

$$MC_{p(Q)} = \left(\prod_{j=1}^q [C_{p(Q);PC_j}]^{\lambda_j} \right)^{\frac{1}{\sum_{j=1}^q \lambda_j}},$$

where $C_{p(Q);PC_j}$ represents the index $C_{p(Q)}$ proposed by [14] to measure the capability of the j th component.

All these MPCIs are applicable to situations where the specifications of all the examined characteristics are bilateral. [69] extended the proposal of [99] to evaluate the multivariate process capability for processes with unilateral specifications. If the multivariate process has only lower specification limits, they proposed the index MC_{pl} defined as:

$$MC_{pl} = \left(\prod_{j=1}^q |C_{pl;PC_j}| \right)^{1/q},$$

where $|C_{pl;PC_j}|$ denotes the absolute value of the index C_{pl} for the j th principal component. They also suggested a new index considering the variance explained by each principal component, which is given by

$$MC'_{pl} = \frac{1}{\sum_{j=1}^q \lambda_j} \sum_{j=1}^q \lambda_j C_{pl;PC_j}$$

Similar indices were formulated for multivariate process with only upper specification limits, replacing $C_{pl;PC_j}$ by $C_{pu;PC_j}$

When these indices have values higher than 1, the multivariate process is capable.

1.1.3 Process capability analysis for profiles

Generally, the quality of a process or product is represented by the distribution of a univariate or multivariate quality characteristic. However, there are numerous processes where the quality is better characterized by a relationship between a response variable and one or more explanatory variables. This relationship is usually known as a profile. Such profiles can be represented using linear or nonlinear models.

A model is nonlinear if at least one derivative of the mean function with respect to the parameters depends on at least one parameter, [81]. Sometimes, the profile is modeled using nonlinear regression models given generally by

$$y_{ij} = f(\mathbf{t}_{ij}, \beta_i) + \varepsilon_{ij} \quad (1.7)$$

where \mathbf{t}_{ij} is a $k \times 1$ vector of regressors for the j th observation of the profile i , ε_{ij} is the random error, β_i is a $p \times 1$ vector of parameters for profile i , and f is nonlinear in the parameters. Different assumptions about of the random errors can be considered.

For simplicity of notation, the scalar model given in equation (1.7) can be expressed in matrix form as

$$\mathbf{y}_i = \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, m \quad (1.8)$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})'$, $\mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}_i) = (f(\mathbf{t}_{i1}, \boldsymbol{\beta}_i), f(\mathbf{t}_{i2}, \boldsymbol{\beta}_i), \dots, f(\mathbf{t}_{in}, \boldsymbol{\beta}_i))'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in})'$. For the i th random sample collected over time, estimates of $\boldsymbol{\beta}_i$ are obtained using an iterative method such as Gauss-Newton or Newton-Raphson.

The use of control charts for cases in which the quality of a process or product can be characterized by a functional relationship between a response variable and one or more explanatory variables is called profile monitoring, [109]. Some discussions of the general issues involving profile monitoring can be found in [110], [109] and [65]. Practical applications of profiles have been reported in several papers, such as [91], [48], [61], [105] and [51]. Some control charts approach for monitoring simple linear profiles have been developed by [48], [50], [114], [115] and [60]. Among the works on monitoring nonlinear profiles are [19], [107], [63], [13], [95], [23].

Process monitoring using control charts can be seen as a two stages process, Phase I and Phase II [108]. The goal in Phase I is to evaluate the stability of the process and, after dealing with some assignable cause, to estimate the in-control values of the process parameters. The goal of Phase II is to monitor the on-line data and quickly detect shifts in the process from the baseline established in Phase I. Different types of statistical methods are appropriate for the two phases with each type requiring different measures of statistical performance. In Phase I it is important to assess the probability of deciding that the process is unstable. In Phase II, the emphasis is on detecting process changes as quickly as possible. This is usually measured by parameters of the run length distribution, where the run length is the number of samples taken before an out-of-control signal. The average run length (ARL) is often used to compare the performance of competing control chart methods, [108] and [36]. Methods for monitoring linear profiles in Phase II have been proposed by [48], [50], [113], [112], and [79].

Recent studies have proposed some approaches to measure the capability of processes characterized by linear profiles. [40] employed the proportion of non-conformance to estimate Process Capability Index (PCI) of linear profiles in Phase I. [22] proposed two methods for measuring the process capability in simple linear profiles. The first method uses the percentage of nonconforming parts produced at each level of the independent variable to introduce a process capability index. The second method is a multivariate process capability approach where a vector of three components is introduced. The components of the vector assess the process dispersion, its centrality, and its location within the upper and the lower specification limits. [41] studied the process capability analysis for simple linear profiles under non-normality. [21] proposed three methods for measuring process capability in multivariate simple linear profiles. The first method is based on the percentage of nonconforming parts produced at each response variable. The second method is a multivariate capability vector and the third one applies the principal component analysis to measure process capability. [102] developed two indices to evaluate the process yield for simple linear profiles normally distributed with one-sided specification.

Recently, [49] proposed a functional method of the angle as explanatory variable to measure process capability index of circular profiles. [103] evaluated the process yield for nonlinear profiles in manufacturing processes using the index S_{pk} proposed by [6], under the assumption that the process is normally distributed at the j th level of the independent variable. To our knowledge, there is no research on the evaluation of capability of processes characterized by univariate or multivariate nonlinear profiles without distributional assumptions and taking into account the within-profile correlation.

This dissertation proposes two methods to evaluate the capability characterized by nonlinear profiles. The first is based on the concept of depth for functional data while the second uses the notion of the Hausdorff distance. For the case of multivariate nonlinear profiles, an approach based on a dimensional reduction method for functional data is presented. The next two sections introduce a review of these topics.

1.2 Functional data

Functional Data Analysis (FDA) is concerned with observations which are viewed as functions defined over some set T . If $T \subset \mathbb{R}$ the functional variable is a curve. The functional variable is a random surface when $T \subset \mathbb{R}^2$.

In theory the functional objects are defined in a continuous argument, however in practice, these are observed in a finite set of points. Then, a previous step in functional data analysis is to reconstruct the functional form of sample curves (or surfaces) from discrete observations. Thus, the typical functional objects, curves or surfaces, can be approximated by smooth functions. When the discretization is quite enough fine, the observations can be considered as functional objects. In this dissertation we consider $T \subset \mathbb{R}$.

Thanks to the fast progress in sensor and information technology and, the increase in storage capabilities and processing power of computers, functional data has become a topic of active statistical research. Functional data can be found in several fields as: biology, econometrics, environmetrics, the food industry, medical sciences, paper industry, speech recognition, etc. The books of [75], [28], [27], [26] and [39] offer a complete description about functional data.

In this section we review three very important problems in functional data: how to order a set of curves given, how to find outliers and how to reduce the dimension in multivariate functional data. The notion of depth for functional data provides a center-outward ordering of the set of curves given. The presence of outliers can lead to inaccurate conclusions on the modeling and forecasting of functional data. We introduce three graphical methods to detect functional outliers. The last problem belongs to the field of the multivariate functional data. In multivariate functional data the observations can be organized in a matrix, where each cell contains a function and the columns can be correlated. The dimensional reduction methods simplify the structure of the data set and retain the most of the information contained in the original structure.

1.2.1 The concept of depth for functional data

In one dimension, the definition of order statistics is straightforward and naturally arises from the intrinsic order on the real line; despite their simplicity, order statistics have a great importance in areas such as robust estimation and inference. In more than one dimension, the concept of order statistics and ranks is not clear and several definitions have been proposed. Proposed definitions are based on different notions of depth.

The concept of depth was introduced in multivariate data analysis for generalizing the notion of order to multivariate data. Depth measures the centrality of a point $\mathbf{t} \in \mathbb{R}^d$, with respect to a distribution function F or a given data cloud. Depth has not assumptions about the underlying distribution, it is based on the relative position of the data point (the shape of the data), [74]. Depth of a point decreases when the point moves away from the center of the data cloud. Thus, deeper points inside a data cloud have the higher depths, while those on the outskirts have lower depths, [118]. The “center”, given by the point of maximal depth if unique, otherwise by the average of

such points, represents a notion of multidimensional median on the given depth function, [82]. Some definitions of depth have been provided by [59], [94], [3], [66], [53], [86], [30], [96] and [117]. Based on this ordering, among others, [54], [116], [82], [83] and [104] have introduced quantitative and graphical methods for analyzing multivariate distributional characteristics such as location, scale, bias, skewness, kurtosis and quantiles.

In functional data analysis each observation is a real function $y_i(t)$, $i = 1, \dots, m$, $t \in T$, and $T \subset \mathbb{R}$. The notion of depth has been extended to functional data by [31], [17], [56], [55], [18], [25], [57] and [58]. The aim of functional depths is to measure the centrality of a given curve, y_i , within a group of trajectories or curves, y_1, \dots, y_m . The notion of depth for functional data provides a center-outward ordering of the set of curves. The curve with maximum depth may be defined as an estimate of the center of the functional distribution, [55] and [25]. Robust inference tools for functional data, based on the notion of depth for curves have been proposed by [55]. The ideas of trimmed regions, contours and central regions are extended from multivariate case to functions and their structural properties and asymptotic behavior are studied by these authors.

[55] and [57] proposed a notion of depth for functional data based on the graphic representation of the functions and makes use of the bands defined by their graphs on the plane. This depth provides a way to order a sample of curves from the center outward. These authors also proposed extending to functional data the concept of α -central region introduced by [54] for multivariate data. It can be defined as the band delimited by the α proportion of deepest curves from the sample. This concept leads to to define functional quantiles and the centrality or outlyingness of an observation. Having the ranks of curves, we propose a measure of capability for functional data, extending the nonparametric definition of the PCIs for random variable introduced by [14].

Graphical methods for visualizing functional data with outlier detection capability have been developed. [44] proposed functional versions of the bagplot and the functional highest density region boxplot, both of which are based on the first two robust principal component scores. [92] proposed an informative exploratory tool for visualizing functional data directly in the functional space, the functional boxplot, based on the center outward induced by band depth.

1.2.2 Band depth for functional data

[55] and [57] proposed a notion of depth for functional data based on the graphic representation of the functions and makes use of the bands defined by their graphs on the plane. As highlighted by [92], this depth allows for ordering a sample of curves from the center outward and, introduces a measure to define functional quantiles and the centrality or outlyingness of an observation. With the ranks of the curves, we measure the functional capability extending the nonparametric definition of the PCIs introduced by [14] and presented in equations (1.5) and (1.6)

Let $C(T)$ be the set of continuous functions defined on the compact interval T in \mathbb{R} . Let $y_1(t), \dots, y_m(t)$ be a collection of observations belonging to $C(T)$. The graph of a function y is the subset of \mathbb{R}^2 given by $G(y) = \{(t, y(t)) : t \in T\}$. The band in \mathbb{R}^2 delimited by the curves y_{i_1}, \dots, y_{i_k} is

$$\begin{aligned} B(y_{i_1}, y_{i_2}, \dots, y_{i_k}) &= \left\{ (t, y) : t \in T, \min_{r=1, \dots, k} y_{i_r}(t) \leq y \leq \max_{r=1, \dots, k} y_{i_r}(t) \right\} \\ &= \left\{ (t, y) : t \in T, y = \alpha \min_{r=1, \dots, k} y_{i_r}(t) + (1 - \alpha) \max_{r=1, \dots, k} y_{i_r}(t), \alpha \in [0, 1] \right\}. \end{aligned}$$

Let J be a fixed value with $2 \leq J \leq m$. For the functions y_1, \dots, y_m the band depth of any of these curves y is

$$BD_{m,J}(y) = \sum_{j=2}^J BD_m^{(j)}(y) \quad (1.9)$$

Where J is an exogenous parameter that indicates the maximum number of curves used to construct a band. [55] and [57] recommend using $J = 3$. $BD_m^{(j)}(y)$ expresses the proportion of bands $B(y_{i_1}, y_{i_2}, \dots, y_{i_j})$ determined by j different curves $y_{i_1}, y_{i_2}, \dots, y_{i_j}$ containing the whole graph of y . $BD_m^{(j)}(y)$ is defined for any function y in y_1, \dots, y_m and fixed j value with $2 \leq j \leq m$, as

$$BD_m^{(j)}(y) = \binom{m}{j}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq m} I\{G(y) \subseteq B(y_{i_1}, y_{i_2}, \dots, y_{i_j})\}, \quad (1.10)$$

where $I\{\cdot\}$ represents the indicator function.

If Y_1, Y_2, \dots, Y_J are independent copies of the stochastic process Y generating the observations y_1, \dots, y_m , the population version of the band depth for a given curve y with respect to the probability measure P is defined as

$$BD_J(y, P) = \sum_{j=2}^J BD^{(j)}(y, P) = \sum_{j=2}^J P\{G(y) \subset B(Y_1, Y_2, \dots, Y_j)\} \quad (1.11)$$

[57] also introduced a more flexible definition, the modified band depth (MBD). In this definition, instead of considering the indicator function they measured the proportion of times that a curve $y(t)$ is in the band. MBD is given by

$$MBD_m^{(j)}(y) = \binom{m}{j}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq m} \lambda_r(A(y_{i_1}, y_{i_2}, \dots, y_{i_j})), 2 \leq j \leq m,$$

where $A_j(y) \equiv A(y; y_{i_1}, y_{i_2}, \dots, y_{i_j}) \equiv \left\{ t \in T : \min_{r=i_1, \dots, i_j} y_r(t) \leq y(t) \leq \max_{r=i_1, \dots, i_j} y_r(t) \right\}$ and $\lambda_r(y) = \lambda(A_j(y)) / \lambda(T)$, if λ is the Lebesgue measure on T . When $y(t)$ is always inside the band, the value $\lambda_r(A_j(y))$ is one, and the modified band depth degenerates to the notion previous of band depth.

For simplicity we write $BD3$ to refer to the band depth method with $J = 3$, and MBD to the modified band depth method.

[57] defined a population median as a function in $C(T)$ maximizing the depths, $Y_{0.5}$, and a sample median function, $\hat{Y}_{0.5}$, as a curve from the sample with highest depth value: $\hat{Y}_{0.5} = \arg_{x \in \{y_1, \dots, y_m\}} \max D_m(\cdot)$, where $D_m(\cdot)$ indicates the depth of a curve, calculated by $BD3$ or MBD . If they are not unique, the median will be the average of the curves maximizing depth.

[53] introduced four desirable properties that an ideal depth function should possess. These properties were analyzed by [116] in a very general framework. These properties are:

- Affine invariance. The depth of a point $x \in R^d$ is independent of the coordinate system.
- Maximality at center. For a distribution having a uniquely defined "center", the depth function is maximal at this point.

- Monotonicity relative to deepest point. As a point $x \in \mathbb{R}^d$ moves away from the "deepest point" along any fixed ray through the center, the depth at x should decrease monotonically.
- Vanishing at infinity. The depth of a point x should approach zero as $\|x\| \rightarrow \infty$.

The finite-dimensional version of *BD3* and *MBD* satisfy these properties except affine invariance, which is not natural for functional data. The sample band depth present uniform consistency in the finite and functional case, see [57].

[54] introduced the concepts of trimmed region, central region and contour for multivariate data. [55] extended these concepts to functional observations.

Let y_1, \dots, y_m be a sample of continuous functions in $C(T)$. Let $y_{(1)}, \dots, y_{(m)}$ be the corresponding ordered functions according to the decreasing depth D_m (*BD3* or *MBD*). Let $y_{(1)}, \dots, y_{(r_\alpha)}$ be the observations with depth larger or equal than α . [55] estimate the α -trimmed region R_α as the band B delimited by the sample curves with depth larger than or equal to α :

$$\begin{aligned} \widehat{R}_{m,\alpha} &= B(y_{(1)}, \dots, y_{(r_\alpha)}) \\ &= \left\{ (t, y) \in T \times \mathbb{R} : \min_{i=1, \dots, r_\alpha} \{y_{(i)}(t)\} \leq y \leq \max_{i=1, \dots, r_\alpha} \{y_{(i)}(t)\} \right\}. \end{aligned}$$

They defined the p -central region, $C_{m,p}$, as the band defined by the fraction p of deepest sample curves:

$$C_{m,p} = B(y_{(1)}, \dots, y_{(\lceil mp \rceil)}), \quad (1.12)$$

where $\lceil mp \rceil$ is the smallest integer not less than $\lceil mp \rceil$.

[119] studied the structural properties of regions and contours for statistical depth functions in multivariate data. These properties are: affine equivariant, nested, connected and compact. [55] extended them to functional observations and showed that R_α and $R_{m,\alpha}$ are nested and affine equivariant.

In multivariate data the boundary of the p th central region is named the p th level contour and represents the p th quantile surface, [54] and [82]. A boundary point of the p th central region is the q th quantile when q and p satisfy $p = |2q - 1|$. Thus, for $p = 0.5$ the level contour represents an "interquantile region", and as $p \rightarrow 0$ it reduces to the median m , [82]. Based on this notion, [92] proposed a functional boxplot, where $C_{m,0.5}$ represents the inter-quartile range for functional data. The border of this region is the envelope representing the box in a classical boxplot. We have that the 99.73% central region represents the region delimited for $y_{0.99865}$ and $y_{0.00135}$.

1.2.3 Functional outliers

The presence of outliers can lead to inaccurate conclusions on the modeling and forecasting of functional data. A curve is an outlier if it has been generated by a stochastic process with a different distribution from the large majority of the curves. A curve is a functional outlier if it is very distant from the mean (magnitude outlier) or if it has a pattern different from the other curves (shape outlier).

For functional data, [44] and [92] have proposed graphical methods with outlier detection capability. The first developed the functional bagplot and the functional highest density region (HDR) boxplot, both of which are based on the first two robust principal component scores. The

functional bagplot is obtained by applying the bivariate bagplot [77] to the first two robust principal component scores, and then mapping the features of the bagplot into the functional space. The functional HDR boxplot is based on the bivariate HDR boxplot proposed by [43], which is applied to the first two robust principal component scores. [92] proposed an informative exploratory tool, the functional boxplot, for visualizing functional data, as well as its generalization, the enhanced functional boxplot. The functional boxplot works directly in the functional space rather than in the feature space that requires principal component analysis techniques.

The bivariate bagplot has a central point (the Tukey median), an inner region (the bag), and an outer region (the fence), beyond which outliers are shown as individual points. The bag is defined as the smallest depth region containing at least 50% of the total number of observations. The outer region of the bagplot is the convex hull of the points containing the region obtained by inflating the bag (relative to the Tukey median) by a factor ρ . [44] recommend to use $\rho = 2.58$ because it value allows that the fence contains the 99% of the observations when the projected bivariate scores follow a standard normal distribution.

The functional bagplot is a mapping of the bagplot of the first two robust principal component scores to the functional curves. The functional bagplot displays the median curve (the curve with the greatest depth), and the inner and outer regions. The inner region is defined as the region bounded by all curves corresponding to points in the bivariate bag. Thus, 50% of curves are in the inner region. The outer region is similarly defined as the region bounded by all curves corresponding to points within the bivariate fence region. The outer region (the fence) is obtained by inflating the inner region (the bag) by a constant factor 2.58 because this value allows that the fence contains the 99% of the observations when the projected bivariate scores follow a standard normal distribution.

[44] applied a robust principal component algorithm resistant to outliers, designed by [16], which uses a form of projection pursuit. Let $\{\phi_k(x)\}$ and $\{z_{i,k}\}$ be the principal components and the principal component scores from this algorithm. The first two score vectors, $(z_{1,1}, \dots, z_{n,1})$ and $(z_{1,2}, \dots, z_{n,2})$, are considered because they capture much of the information inherent in the original curves. Let $\mathbf{z}_i = (z_{i,1}, z_{i,2})$ be a bivariate score. Using halfspace location depth, a measure of multivariate depth proposed by [94], the bivariate scores are ordered. Then the curves can be ordered according to the increasing order of the bivariate scores using the halfspace depth.

The functional HDR boxplot is based on the bivariate HDR boxplot proposed by [43], which is applied to the first two robust principal component scores. The bivariate HDR boxplot is constructed using a bivariate kernel density estimate $\hat{f}(\mathbf{z})$, which is defined as

$$\hat{f}(\mathbf{z}) = \frac{1}{m} \sum_{i=1}^m K_{h_i}(\mathbf{z} - Z_i),$$

where Z_i represents a set of bivariate points, $K_{h_i}(\cdot) = K(\cdot/h_i)/h_i$, K is the kernel function, and h_i is the bandwidth for the i th dimension. The bandwidths are selected using smoothed cross-validation. $\hat{f}(\mathbf{z})$ is calculated from all of the bivariate robust principal component scores, obtained as in the functional bagplot.

In the HDR boxplot the functional data are ordered by values of o_i in a decreasing order, where $o_i = \hat{f}(\mathbf{z}_i)$. Thus, the first curve, curve with the highest density value, may be considered the "modal curve", whereas the last curve, curve with the lowest density value, may be considered the most unusual curve.

A HDR, "highest density region", is the region with coverage probability $1 - \alpha$ where all points within the region have a higher density estimate than any of the points outside the region.

HDR is defined as

$$H_\alpha = \{\mathbf{z} : \hat{f}(\mathbf{z}) \geq f_\alpha\},$$

where f_α is such that $\int_{H_\alpha} \hat{f}(\mathbf{z}) d\mathbf{z} = 1 - \alpha$. For a bivariate density, the HDR can be considered as contours, with an expanding coverage as α decreases.

The bivariate HDR boxplot displays the mode (the highest density point), defined as $\arg \sup \hat{f}(\mathbf{z})$, along with the 50% inner and (usually) 99% outer highest density regions. All points excluded from the outer HDR are outliers.

The functional HDR boxplot is a mapping of the bivariate HDR boxplot of the first two robust principal component scores to the functional curves. The functional HDR boxplot displays the modal curve (the curve with the highest density), and the inner and outer regions. The inner region is defined as the region bounded by all curves corresponding to points inside the 50% bivariate HDR. The outer region is similarly defined as the region bounded by all curves corresponding to the points within the outer bivariate HDR.

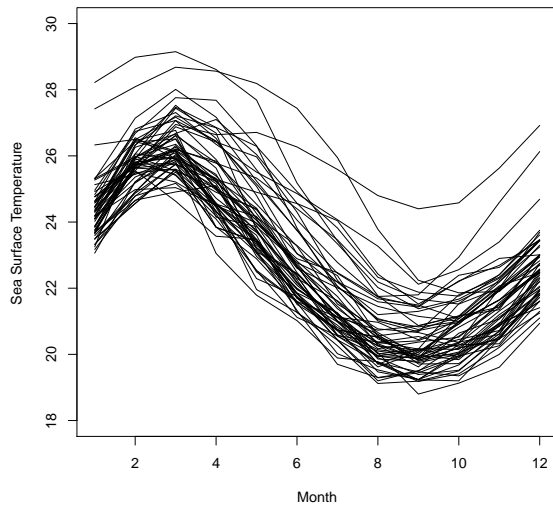
The functional boxplot is based on the center outward ordering induced by band depth for functional data. Here the users can find estimations of: the envelope of the 50% central region, the median curve, and the maximum non-outlying envelope. The functional boxplot is analog to the classical boxplot: the 50% central region is the analog to the inter-quartile range and the 1.5 times the 50% central region empirical rule (the fences) is employed to detect functional outliers. Any curve outside the fences is considered as outlier.

The functional bagplot, the functional HDR boxplot and the functional boxplot are illustrated on a series of sea surface temperatures related to the El Niño phenomenon, datased used by [44] and [92]. The data consist of monthly sea surface temperatures (SST) measured in degrees Celsius over the east-central tropical Pacific Ocean, see Figure 1.1a. Each curve represents one year of observed SST in degrees Celsius from January 1951 to December 2007. In the functional bagplot, Figure 1.1b, the dark and light gray regions show the bag and fence regions, respectively. The black line is the median curve and the curves outside the fence regions are shown as outliers of different colors. In the functional HDR boxplot, Figure 1.1c, the dark and light gray regions show the 50% HDR and outer HDR, respectively. The black line is the modal curve. The curves outside the outer HDR are shown as outliers of different colors. In the functional boxplot, Figure 1.1d, the blue curves denote envelopes, and the black curve represents the median curve. The red dashed curves are the outlier candidates detected by the 1.5 times the 50% central region rule. Simulation results on the performance of these outlier detection methods are reported in [44] and [92].

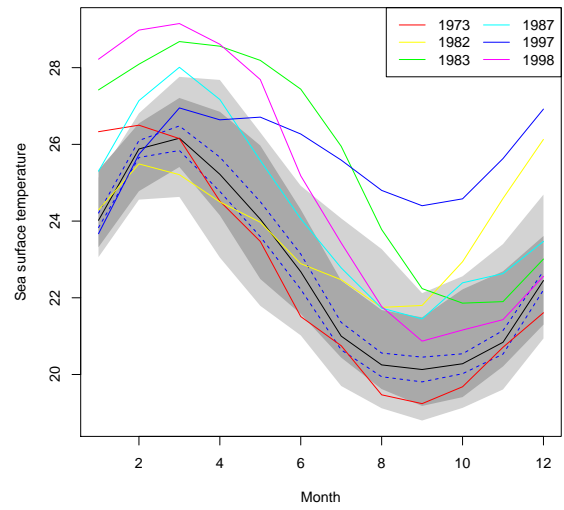
1.2.4 Principal component for multivariate functional data

In this subsection we review a principal component method for multivariate functional data proposed by [4]. In multivariate functional data each observation is a finite dimension vector whose elements are functions, which are realizations of a stochastic process. This method simplifies the structure of the data set by summarizing the vector of functions for each individual with a small set of functions that retains as much information as possible. In general, this method is a dimension reduction technique for multivariate functional.

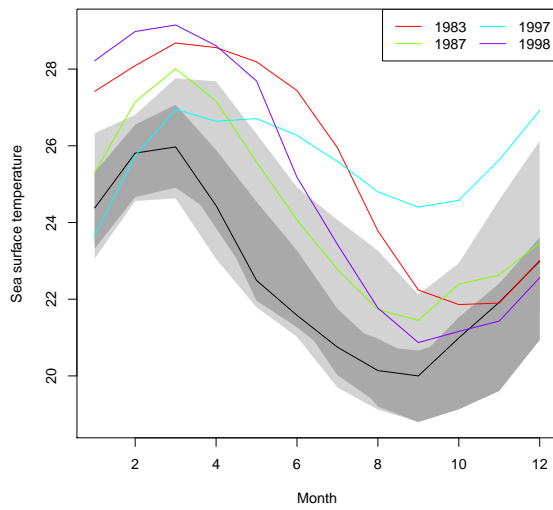
Let $\mathbf{Y} = (Y_1, \dots, Y_p)'$ an p dimensional stochastic process. Each random function Y_j , $j = 1, \dots, p$, is defined on the compact real interval $[c, d]$, with values in \mathbb{R} . For each $t \in [c, d]$, the random vector $\mathbf{Y}(t) = (Y_1(t), \dots, Y_p(t))'$ has a mean vector $\boldsymbol{\mu}(t) = \mathbf{0}$, and a positive-definite covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}(t)} = \mathbf{Y}(t)\mathbf{Y}(t)'$.



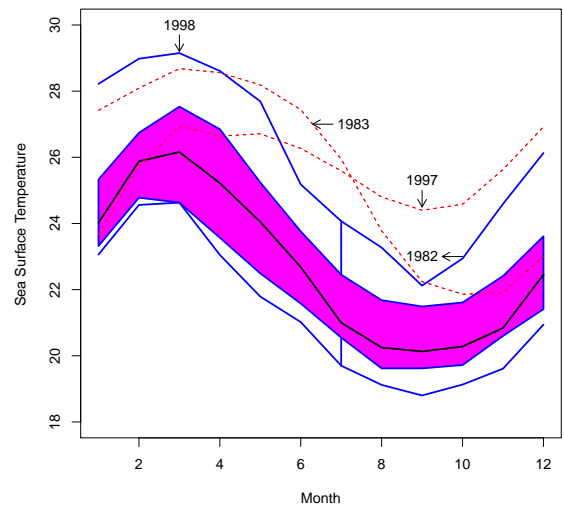
(A)



(B) Functional bagplot



(C) Functional HDR boxplot



(D) Functional boxplot

FIGURE 1.1. Raw data (A), Functional bagplot with factor=2.58 (B), functional HDR boxplot for $\alpha = 0.01$ (C), and functional boxplot with factor=1.5 (D) of the sea surface temperatures measured monthly in degrees Celsius over the east-central tropical Pacific Ocean from 1951 to 2007.

Let $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ be independent copies of \mathbf{Y} . The observations of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ can be organized in an $n \times p$ matrix, M , which provides a set of $n \times p$ -variate curves.

[4] carry out a classical multivariate PCA on $\mathbf{Y}(t)$ for each value t on which the functions are defined. Thus, the principal components are linear combinations of the original functions, defined as:

$$Z_r(t) = \mathbf{e}_r(t)' \mathbf{Y}(t), \quad r = 1, \dots, p, \quad (1.13)$$

where $\mathbf{e}_r(t)$ is a unit norm eigenvector corresponding to the r th eigenvalue of $\Sigma_{\mathbf{Y}(t)}$, with $\lambda_1(t) > \dots > \lambda_p(t) > 0$.

These functional principal components are functions uncorrelated, therefore the inner product between two different components with respect to the product measure $dt \times dP$ is zero, this is:

$$\langle Z_r, Z_s \rangle = E \left[\int_c^d Z_r(t) Z_s(t) dt \right] = 0.$$

To measure the information of \mathbf{Y} retained by each component, [4] considered the integrated variance, which is given by the following expression:

$$\int_c^d \mathbf{e}_r(t)' \Sigma_{\mathbf{Y}(t)} \mathbf{e}_r(t) dt.$$

In this proposal Z_r and λ_r are functions defined on the compact real interval $[c, d]$ with values in \mathbb{R} , while \mathbf{e}_r is a function defined on the same interval but with values in \mathbb{R}^p .

For a given value of t , $\mathbf{e}_r(t)$ and $-\mathbf{e}_r(t)$ can be considered as eigenvectors of $\Sigma_{\mathbf{Y}(t)}$. Therefore, the number of possible weighting functions of the linear combinations in each principal component is infinite. [4] proposed a criterion to select the sign of the weighting functions so that the resulting components are smooth and easier to interpret. The basic idea is to select the sign at t which is closer on average to the signs already determined for values in a neighborhood of t . This criterion allows that the principal components do not change abruptly with t . However, the application of this method can produce problems of interpretation with the specification functions of the components. For example, for processes with bilateral specifications, we have that $USL_{Z_r}(t) - LSL_{Z_r}(t) < 0$ for some values of t . To overcome this drawback and obtain smooth components, we have proposed a new method to select the sign of the weighting functions.

[4] presented two approaches to measure the proportion of variability explained by the r th principal component. The first approach is given by

$$\pi_r = \frac{1}{d-c} \int_c^d \frac{\lambda_r(t)}{V(t)} dt \quad (1.14)$$

where $V(t) = \sum_{j=1}^p \lambda_j(t)$ is the total variance at a given point $t \in [c, d]$. The second approach is a weighted version of π_r .

In practice $\Sigma_{\mathbf{Y}(t)}$ is unknown, and the functional principal components must be estimated using the matrix of sample functions. Let $\mathbf{y}_1, \dots, \mathbf{y}_n$, be realizations of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, with $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$, organized in the matrix M . Each row of M is a vector of p functions, and each ele-

ment (i, j) of M , denoted by y_{ij} , is the function j of the i th individual. Generally, $\Sigma_{\mathbf{Y}(t)}$ is replaced by $\widehat{\Sigma}_{\mathbf{Y}(t)}$, the sample covariance matrix of M evaluated at t .

The principal component method presented by [4] for multivariate functional data differs from the method proposed by [75]. [4] they carry out classical PCA for each value of the domain on which the functions are observed and suggest an interpolation method to build their functional principal components. While [75] propose to concatenate the observations of the functions on a fine grid of points (or the coefficients in a suitable basis expansion) into a single vector and then to perform a standard principal component analysis (PCA) on these concatenated vectors. When a basis expansion is used, this method forces to consider only orthonormal basis since the metric induced by the scalar product between the basis functions is not taken into account. [4] propose to simplify the structure of the data set by summarizing the vector of functions for each individual with a single function or a very small set of functions that retains as much information as possible from the original vector of functional observations. In [75] the final result is a vector of real numbers but not a function.

1.3 Hausdorff distance

The distance functions of metric spaces represent a way of quantifying the closeness of objects in a given domain. The distance most commonly used is the Euclidean distance, which represents the distance between two individual points. However it is not suitable for other type of objects as lines, curves or surfaces. For these objects, the Hausdorff distance has been used. Some applications of the Hausdorff distance are found in image processing, matching applications as face detection and tracking, medical image registration, etc.

Given two finite point sets, A and B , the Hausdorff distance between A and B is defined as

$$H(A, B) = \max(h(A, B), h(B, A)), \quad (1.15)$$

where

$$h(A, B) = \sup_{a \in A} \left\{ \inf_{b \in B} \|a - b\| \right\}, \quad h(B, A) = \sup_{b \in B} \left\{ \inf_{a \in A} \|a - b\| \right\}, \quad (1.16)$$

and $\sup\{\cdot\}$ represents the least upper bound of a set, $\inf\{\cdot\}$ the greatest lower bound of a set, and $\|\cdot\|$ some underlying metric define on the points of A and B , generally the Euclidean distance, which is assumed in this dissertation. The functions $h(A, B)$ and $h(B, A)$ are called the directed Hausdorff distance from A to B and from B to A , respectively. Not always $h(A, B) = h(B, A)$, therefore the directed Hausdorff distance functions are not really a distance function.

The Hausdorff distance is a metric, therefore satisfy the following properties:

- $H(A, B) \geq 0$ for all A and B ,
- $H(A, B) = 0$ iff $A = B$,
- $H(A, B) = H(B, A)$ and
- $H(A, B) \leq H(A, C) + H(B, C)$.

If A and B are non-empty close bounded sets (compact sets) the directed Hausdorff distance function from A to B given by (1.16) can be simplified into:

$$h(A, B) = \max_{a \in A} (d(a, B)) \quad (1.17)$$

where $d(a, B) = \min_{b \in B} (d(a, b))$ represents the distance from a to B , generally the Euclidean distance. $h(B, A)$ is defined in a way similar. For more details see [78] and [62].

The Hausdorff distance is sensitive to extreme values in the minimum distances, which may not be desirable, [29]. To solve this limitation [11] replaced the maximum operator in equation (1.17) by the median. This modification is denoted by

$$h^\dagger(A, B) = \text{med}_{a \in A} (d(a, B)). \quad (1.18)$$

By construction, this distance is symmetrical, so that $h^\dagger(A, B) = h^\dagger(B, A)$.

The computation of the Hausdorff distance can be classified into the following categories according to type of geometric objects dealt with: point sets, polygons and line segment sets, curves and surfaces and curve sets. See [2].

Process Capability Analysis for NonLinear Profiles

This chapter provides a methodology to measure process capability characterized by nonlinear profiles without distributional assumptions and having account the correlation within the profile. Here we extend to functional data the Process Capability Indexes proposed by [14] to measure the capability of a process characterized by a random variable. The performance of the estimators based on depth functions and Hausdorff distance is evaluated through simulation studies. An example illustrates the applicability of these methods.

2.1 Process capability of nonlinear profiles using depth functions

In this section, we propose two methods for measuring process capability characterized by nonlinear profiles without distributional assumptions. Let Y a one dimensional stochastic process defined on a probability space $(\Omega, \mathfrak{F}, P)$ according to equation (1.8). The distributional assumption of the random errors has been relaxed. The random errors are considered independent and identically distributed (iid), no necessarily with normal distribution. We restrict the methodology to functions in the space $C(T)$ of real continuous function on the compact interval T . Let $Y_{0.5}$, $Y_{0.99865}$ and $Y_{0.00135}$ be the population version of the median and the quantiles 0.99865 and 0.00135, respectively. Initially, we suppose that the process is evaluated in s points. For each t_i , $i = 1, \dots, s$, the one-dimensional PCIs can be calculated using the indices proposed by [14] and presented in equations (1.5) and (1.6). The PCIs that measure the process capability for the nonlinear profiles are defined as the arithmetic mean of the one-dimensional PCIs evaluated for each t_i , which are expressed by the following equations:

$$C_{pu(prf)}^A = \frac{1}{s} \sum_{i=1}^s \frac{USL(t_i) - Y_{0.5}(t_i)}{Y_{0.99865}(t_i) - Y_{0.5}(t_i)} \quad (2.1)$$

$$C_{pl(prf)}^A = \frac{1}{s} \sum_{i=1}^s \frac{Y_{0.5} - LSL(t_i)}{Y_{0.5}(t_i) - Y_{0.00135}(t_i)} \quad (2.2)$$

$$C_{p(prf)}^A = \frac{1}{s} \sum_{i=1}^s \frac{USL(t_i) - LSL(t_i)}{Y_{0.99865}(t_i) - Y_{0.00135}(t_i)} \quad (2.3)$$

$$C_{pk(prf)}^A = \frac{1}{s} \sum_{i=1}^s \min \left[\frac{USL(t_i) - Y_{0.5}(t_i)}{Y_{0.99865}(t_i) - Y_{0.5}(t_i)}, \frac{Y_{0.5}(t_i) - LSL(t_i)}{Y_{0.5}(t_i) - Y_{0.00135}(t_i)} \right]. \quad (2.4)$$

If $s = 1$, the proposed PCIs become the PCIs given by equations (1.5) and (1.6). This method is called A.

Based on the fact that USL , LSL , $Y_{0.00135}$, $Y_{0.5}$ and $Y_{0.99865}$ are continuous functions, we extend the indices proposed by [14] to functional data. The following indices constitute the second method called B:

$$C_{pu(prf)}^B = \frac{\int (USL(t) - Y_{0.5}(t)) dt}{\int_I (Y_{0.99865}(t) - Y_{0.5}(t)) dt} \quad (2.5)$$

$$C_{pl(prf)}^B = \frac{\int (Y_{0.5} - LSL(t)) dt}{\int_I (Y_{0.5}(t) - Y_{0.00135}(t)) dt} \quad (2.6)$$

$$C_{p(prf)}^B = \frac{\int (USL(t) - LSL(t)) dt}{\int_I (Y_{0.99865}(t) - Y_{0.00135}(t)) dt} \quad (2.7)$$

$$C_{pk(prf)}^B = \min \left(C_{pl(prf)}^B, C_{pu(prf)}^B \right) \quad (2.8)$$

where T is a compact interval in R .

In both methods, the proposed indices quantify the relationship between the actual performance of the nonlinear profiles and the specification limits.

These indices can be estimated replacing $Y_{0.5}(t)$, $Y_{0.99865}(t)$ and $Y_{0.00135}(t)$ by appropriate estimators that rely on sample observations. Let y_1, \dots, y_m be a sample of real continuous functions on the compact interval T . We initially consider the estimators defined by [55] and [57], which are based on the depths of the given curves, $D_m(y_1), \dots, D_m(y_m)$. $\hat{Y}_{0.5}$ is the curve from the sample with highest depth value: $\hat{Y}_{0.5} = \arg_{y \in \{y_1, \dots, y_m\}} \max D_m(y)$. If they are not unique, the median will be the average of the curves maximizing depth. $\hat{Y}_{0.99865}$ and $\hat{Y}_{0.00135}$ are the curves that envelope the 99.73% central region using equation (1.12) with $p = 0.9973$.

If the depths are calculated using the MBD method, we identify these estimators as $\hat{Y}_{0.5}^{MBD}$, $\hat{Y}_{0.99865}^{MBD}$ and $\hat{Y}_{0.00135}^{MBD}$, respectively. The indices obtained when these functional estimators are used in equations (2.1)-(2.4) are identified by $\hat{C}_{pu(prf)}^{MBDA}$, $\hat{C}_{pl(prf)}^{MBDA}$, $\hat{C}_{p(prf)}^{MBDA}$ and $\hat{C}_{pk(prf)}^{MBDA}$. We use the letter B instead of A when they are employed in equations (2.5)-(2.8). If the depths are calculated using the BD3 method, we identify the estimators using $BD3$ instead of MBD .

$\hat{Y}_{0.5}^{(MBD)}$ is the function with greater depth from a sample of functions using the MBD method, but this does not mean that at each point $t \in T$, $\hat{Y}_{0.5}^{MBD}(t)$ is the deepest point. A critical situation occurs when at a point t , $\hat{Y}_{0.5}^{MBD}(t) = \hat{Y}_{0.99865}^{MBD}(t)$ because $\hat{C}_{pu(prf)}^{MBDA} = \infty$, see Figure 2.1. When $\hat{Y}_{0.5}^{MBD}(t)$ is very close to $\hat{Y}_{0.99865}^{MBD}(t)$, $\hat{C}_{pu(prf)}^{MBDA}$ assumes values higher than the desired value. A similar situation occurs when $C_{pl(prf)}$ is estimated or if the depths are calculated using the BD3 method.

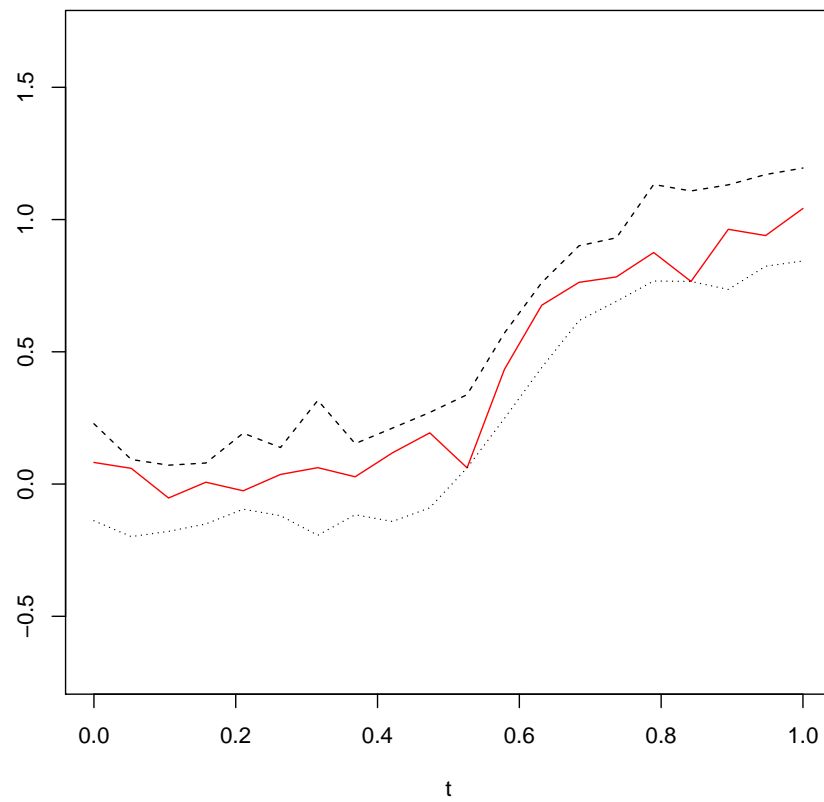


FIGURE 2.1. Estimations of $Y_{0.5}$ (solid line), $Y_{0.99865}$ (dashed line) and $Y_{0.00135}$ (dotted line) using modified band depth.

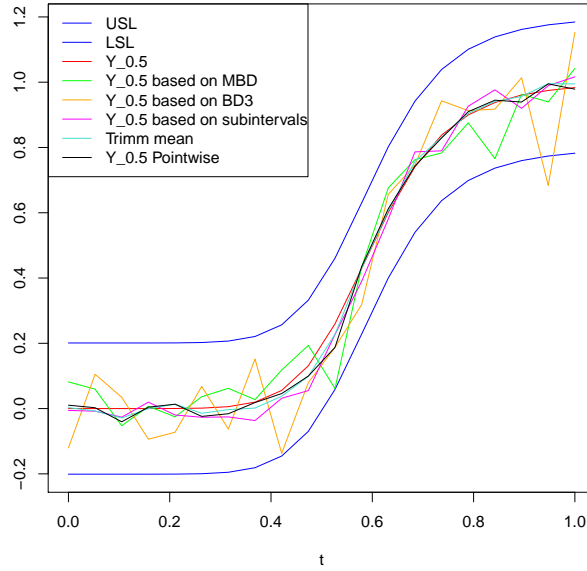


FIGURE 2.2. Behaviour of several estimations of the median function. The red curve represents the true median curve, green curve is based on MBD, orange curve is supported by BD3, pink curve relies on subintervals with MBD, magenta curve uses the trimmed mean with deepest m curves, and black curve is based on pointwise estimations. Specification limits with blue curves.

To solve this difficulty, we consider two additional estimators of $Y_{0.5}$: the functional α -trimmed mean and the median built by subintervals. α -trimmed mean is a robust location estimator of the center of the distribution, introduced by [31], which is defined as

$$\hat{\mu}_{TM\alpha} = \frac{1}{m - [\alpha m]} \sum_{i=1}^{m - [\alpha m]} y_i,$$

where $[\]$ indicates the integer part and $0 \leq \alpha \leq \frac{m-1}{m}$. α -trimmed mean is the mean of the most central $m - [\alpha m]$ curves. When $\alpha = 0$, the trimmed mean is the functional mean, and if $\alpha = \frac{m-1}{m}$ the trimmed mean is the functional median, [24]. The best estimator $\hat{\mu}_{TM\alpha}$ of the median depends on the α value, which can be found using simulation. When the depths are calculated using the MBD (or BD3) method, we note this estimator as $\hat{Y}_{0.5}^{TM_{MBD}}$ (or $\hat{Y}_{0.5}^{TM_{BD3}}$).

$Y_{0.5}$ also can be estimated by dividing the interval T in k subintervals. In each subinterval the deepest curve is found. The curve formed by the connection of these k curves is an estimator of $Y_{0.5}$, which is noted as $\hat{Y}_{0.5}^{Sub_{MBD}}$ when the depths are calculated by the MBD method, or $\hat{Y}_{0.5}^{Sub_{BD3}}$ if the BD3 method is used.

Figure 2.2 shows the true median function and the estimations obtained with the proposed estimators. Here we can observe that the estimations based on the depths MBD and BD3 are very irregular, while the estimations based on trimmed mean and subintervals are smoother and closer to the true median function.

To estimate the PCIs considering these new estimators of $Y_{0.5}$, we use the curves that envelope the central region $C_{m,0.9973}$ as estimators of $Y_{0.99865}$ and $Y_{0.00135}$. The PCIs estimated using these curves are noted as $C_{pu(prf)}^{SUB_{Depth}L}$, $C_{pl(prf)}^{SUB_{Depth}L}$, $C_{p(prf)}^{SUB_{Depth}L}$ and $C_{pk(prf)}^{SUB_{Depth}L}$, where $Depth$ indicates the

method used to calculate the depths of the curves, MBD or BD3; and L expresses the method employed to calculate the PCIs, A or B, with A associated to equations (2.1)-(2.4) and B to equations (2.5)-(2.8).

Finally, we estimate the PCIs using pointwise estimators for $Y_{0.5}$, $Y_{0.99865}$ and $Y_{0.00135}$. These pointwise estimators are defined by

$$\begin{aligned}\widehat{Y}_{0.5}^{PW}(t) &= \widehat{Q}_{0.5} \{y_1(t), y_2(t), \dots, y_m(t)\}, \\ \widehat{Y}_{0.99865}^{PW}(t) &= \widehat{Q}_{0.99865} \{y_1(t), y_2(t), \dots, y_m(t)\}, \\ \widehat{Y}_{0.00135}^{PW}(t) &= \widehat{Q}_{0.00135} \{y_1(t), y_2(t), \dots, y_m(t)\},\end{aligned}$$

for $t \in T$, where $\widehat{Q}_p\{\cdot\}$ is the sample p -quantile from the set given. The estimated PCIs using these pointwise estimators are noted with the expression PW as superindex.

We apply a bootstrap method to construct a confidence interval of the PCIs proposed. A bootstrap sample y_1^*, \dots, y_m^* is obtained by random sampling with replacement from the original functions y_1, \dots, y_m . The proposed PCIs are calculated using the bootstrap sample. This process is repeated B times so that a collection of B estimations of the PCIs is obtained. The $100(1 - \alpha)\%$ confidence interval of the PCIs is given by the percentiles $\alpha/2$ and $1 - \alpha/2$ calculated from B estimations of the PCIs.

This bootstrap method is different from the method proposed by [17]. They obtain a bootstrap sample $y_1^*(t_i), \dots, y_m^*(t_i)$ for a fixed t_i , with $i = 1, \dots, s$. Therefore, the obtained bootstrap sample is discretized and consists of s univariate bootstrap samples while in our proposal the selected elements are continuous functions.

2.1.1 Simulations

A series of simulations were conducted with a number of several profiles and distributions. We consider m profiles, with $m = 20, 30, 60$ and 90 . Normal(0,0.1), Lognormal(0, 0.3), Weibull(6,1), Gamma(1.2,5) and Beta(5,5) distributions are considered for the random terms ε_{ij} of equation (1.7). In this simulation study, we only consider the indices $C_{pu(prf)}$ and $C_{pl(prf)}$. Target values for these PCIs are set at 0.77, 1, 1.33, 1.66 and 2.0. For each target value of the PCI, the corresponding specification functions (USL or LSL) are obtained by solving the following equations:

$$USL(t) = C_{pu} [Y_{0.99865}(t) - Y_{0.5}(t)] + Y_{0.5}(t) \quad (2.9)$$

$$LSL(t) = Y_{0.5}(t) - C_{pl} [Y_{0.5}(t) - Y_{0.00135}(t)], \quad (2.10)$$

where $Y_{(p)}$ is the p th quantile of the considered distribution and $t \in T$.

Assuming the process is in statistical control, for each simulation run m profiles with s points are generated from the 4-parameter logistic model, employed in the pharmaceutical industry, which represents dose-response profiles of a drug. This model was considered by [107] and used after by [95], which is expressed by:

$$y_{it} = \beta_1 + \frac{\beta_2 - \beta_1}{1 + (t/\beta_3)^{\beta_4}} + \varepsilon_{it} \quad (2.11)$$

where y_{it} is the measured response of the subject i exposed to dose t , for $i = 1, \dots, m$, β_1 is the upper asymptote parameter, β_2 is the lower asymptote parameter, β_3 is the concentration corresponding to half distance between β_1 and β_2 , B_4 is the slope, and ε_{it} is the random error, [20]

Seven several estimations of the median function are obtained from the data: $\widehat{Y}_{0.5}^{MBD}$, $\widehat{Y}_{0.5}^{BD3}$, $\widehat{Y}_{0.5}^{TM_{MBD}}$, $\widehat{Y}_{0.5}^{TM_{BD3}}$, $\widehat{Y}_{0.5}^{Sub_{MBD}}$, $\widehat{Y}_{0.5}^{Sub_{BD3}}$, and $\widehat{Y}_{0.5}^{PW}$. The six first estimations are based on the functional depths, and for each of them, one estimation of $Y_{0.99865}$ and $Y_{0.00135}$ is obtained finding the curves that envelope the central region $C_{m,0.99763}$. When these estimations are employed in equations (2.1), (2.2), (2.5), and (2.6), estimations of the indices $C_{pu(prf)}^A$, $C_{pl(prf)}^A$, $C_{pu(prf)}^B$, and $C_{pl(prf)}^B$ are obtained, respectively.

Each run is replicated R times, for $R = 100$. To determine the best estimator of the indices $C_{pu(prf)}$ and $C_{pl(prf)}$, we compare the accuracy and the precision of the several estimators according to the amount of profiles and the given distribution. R program ([73]) is used to generate the simulations. In particular, we have used the function `fbplot` written by [92] to obtain the central regions $C_{m,p}$.

The values assigned to the dose-response model parameters are: $\beta_1 = 1$, $\beta_2 = 0$, $\beta_3 = 0.6$, and $\beta_4 = 8$. t is defined on the interval $[0, 1]$, and the random errors ε_{it} are considered iid. Figure 2.3 shows the curves of 30 profiles generated for each distribution. The functional specification limits are obtained using equations (2.9) and (2.10). Figure 2.4 shows $Y_{0.5}$, $Y_{0.99865}$, $Y_{0.00135}$, USL and LSL functions for each considered distribution, when $C_{pu(prf)}$ and $C_{pl(prf)}$ are set to 1.33.

To use the estimators of the median function based on the trimmed mean, $\widehat{Y}_{0.5}^{TM_{MBD}}$ or $\widehat{Y}_{0.5}^{TM_{BD3}}$, it is necessary to determine the quantity of curves j required to produce the best fit. We evaluate this fit using the mean integrated error, which is given by

$$E = \frac{1}{R} \sum_{r=1}^R EI(r)$$

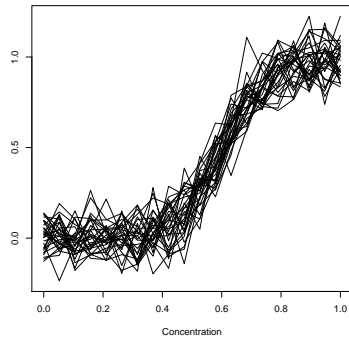
where $EI(r)$ is the integrated error for the r th replication, which is calculated at s points by the following expression:

$$EI(r) = \frac{1}{s} \sum_{i=1}^s (\widehat{\mu}_{TM_{\alpha}}(t_i) - Y_{0.5}(t_i))^2,$$

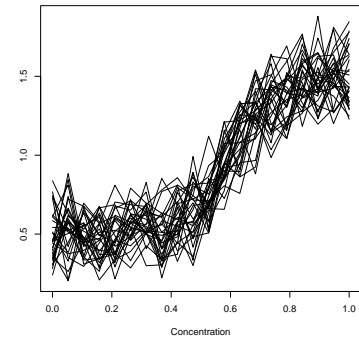
where $\alpha = \frac{m-j}{m}$, with $j = 1, \dots, m$.

For the Normal(0,0.1), Lognormal(0, 0.3), Weibull(6,1) and Beta(5,5) distributions, we found that the best fit is when $j = m$. For the Gamma(1.2,5) distribution, when the MBD method is used to calculate the depths, $j = 18, 20, 27$, and 37 for $m = 20, 30, 60$, and 90 , respectively. If the depths are calculated using the BD3 method, $j = m$.

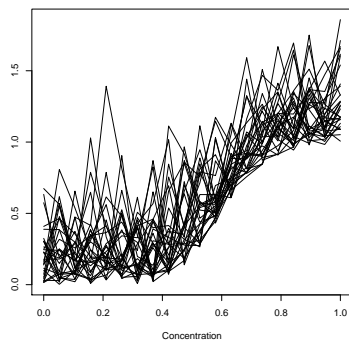
Tables 2.1, 2.2, 2.3, 2.4 and 2.5 show the mean and standard deviation of the estimators considered for $C_{pl(prf)}$, under five distributions, when the target values for the PCI are set at 0.67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 profiles, and $s = 20$ points. In general, the estimations based on method A (columns 3, 5, 7, 9, 11, 13 and 15) have greater overestimation than the based on method B (columns 4, 6, 8, 10, 12, 14 and 16). In particular, a high percentage of the estimations corresponding to the estimators $\widehat{C}_{pl(prf)}^{MBDA}$ and $\widehat{C}_{pl(prf)}^{BD3A}$ have infinite values or are very high, because in some points the estimated median is equal or close to $\widehat{Y}_{0.00135}$, see columns 3 and 5. Both methods are sensitive to the size of the sample. When m increases, the estimations of $C_{pl(prf)}$



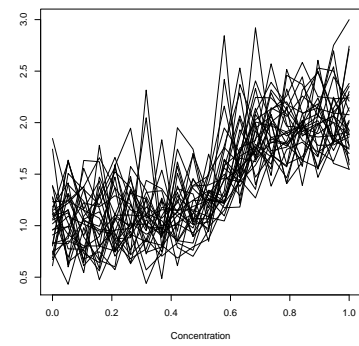
(A) Normal(0,0.1) distribution



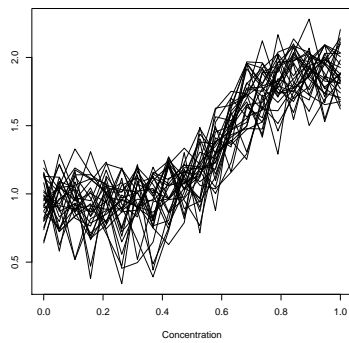
(B) Beta(5,5) distribution



(C) Gamma(1.2,5) distribution

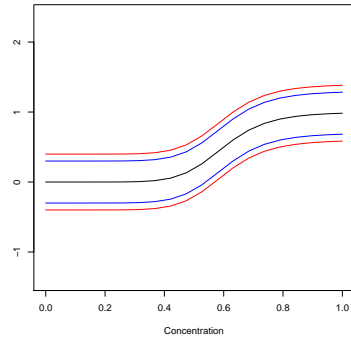


(D) Lognormal(0,0.3) distribution

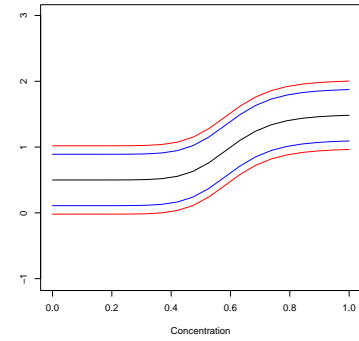


(E) Weibull(6,1) distribution

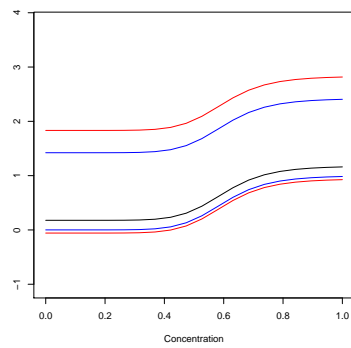
FIGURE 2.3. 30 Profiles from the 4-parameter logistic model under five distributions of the error term.



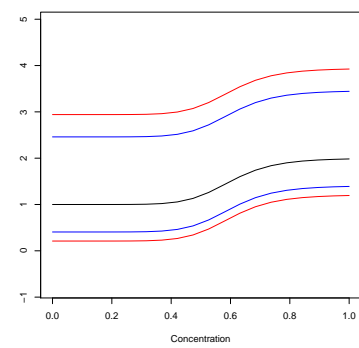
(A) Normal(0,0.1) distribution



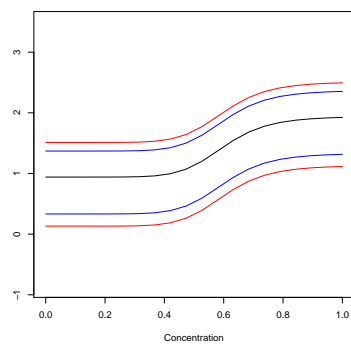
(B) Beta(5,5) distribution



(C) Gamma(1.2,5) distribution



(D) Lognormal(0,0.3) distribution



(E) Weibull(6,1) distribution

FIGURE 2.4. Median function (in black), $Y_{0.99865}$, $Y_{0.00135}$ (in blue), and specification functions (in red) for the 4-parameter logistic model using five distributions when $C_{pu} = 1.33$ and $C_{pl} = 1.33$.

are closer to the target PCI. For example in table 2.2, under Lognormal(0,0.3) distribution with target $C_{pl(prf)} = 1.33$, the mean of $\widehat{C}_{pl(prf)}^{TM_{MBDB}}$ is 1.787, 1.682, 1.548, and 1.477 when $m = 20, 30, 60$ and 90, respectively.

In general, the estimators based on the depth of the curves using method B have better performance than the pointwise estimators (last two columns). For example in table 2.4, under the Gamma(1.2, 5) distribution for the $C_{pl(prf)} = 1.66$, the mean of the pointwise estimator using method A, $\widehat{C}_{pl(prf)}^{PWA}$, is 1.887, 1.806, 1.734 and 1.711 when $m = 20, 30, 60$ and 90 respectively, while the mean of the estimator $\widehat{C}_{pl(prf)}^{BD3B}$ is 1.628, 1.593, 1.621 and 1.595.

To evaluate the accuracy of the estimators, we calculate their relative bias. Here we only report the simulation results for the Normal(0, 0.1) and Gamma (1.2, 5) distributions. The simulation results for other distributions are not reported here, but are available under request from the first author. Table 2.6 presents the relative bias of the estimators of $C_{pl(prf)}$ when the random terms have normal(0, 0.1) distribution. When $m = 20$ the overestimation of all the estimators is very high, greater to 60%. When $m = 90$, the estimators based on the functional depths using the method B have a relative bias between 0.214 and 0.222, while the best pointwise estimator has a relative bias of 0.236. For the Gamma(1.2, 5) distribution the estimators based on subintervals with method B present a relative bias of 0.028, 0.023, 0.020, 0.017 and 0.016 when the depths are calculated by MBD, $m = 90$ and the target values of the index $C_{pl(prf)}$ are 0.67, 1, 1.33, 1.66 and 2, respectively. For the same m and target PCI, the relative bias is 0.029, 0.023, 0.020, 0.015 and 0.017 when the depths are calculated using BD3, while the pointwise estimator has a relative bias of 0.027, 0.026, 0.025, 0.024 and 0.021, respectively. In this distribution some estimators present underestimation, see table 2.7.

The simulations show that the estimators of $C_{pl(prf)}$ employing the method B with subintervals or trimmed mean, based on the functional depths, have a better performance than the others estimators.

2.1.2 Example

We use the vertical density profile (VDP) data considered initially by [111] and after by [107]. In the manufacture of particleboard, the density is a very important quality characteristic. The density is measured using a profilometer which uses a laser device to take a series of measurements across the thickness of the board. The density depends on the depth where the measurement is performed. It is larger at the top and bottom faces of a board than the density near the core. The vertical density profile (VDP) of 24 particleboards were taken at 313 points, equally spaced (0.002 inches) with initial point at 0. The 24 profiles, y_i with $i = 1, \dots, 24$, are showed in the Figure 2.5.

[107] consider a nonparametric approach for modeling these profiles. They employed spline smoothing with 16 degrees of freedom, method initially proposed by [97]. Figure 2.6 shows the spline fits to each profile, \hat{y}_i , $i = 1, \dots, 24$. An average spline, \tilde{y} , is calculated as $\tilde{y} = \sum_{i=1}^{m=24} \hat{y}_i / m$.

In process capability analysis it is imperative that previously abnormal observations (outliers) are detected and excluded from analysis. Several informative exploratory tools have been developed to visualize functional outliers, among them are: the functional bagplot and the functional highest density region (HDR) boxplot presented by [44] and the functional boxplot formulated by [92]. The functional bagplot and the functional HDR boxplot show that the profiles corresponding to the boards A6 and B1 are outliers, while the functional boxplot detects only the board A6 as

TABLE 2.1. Mean and standard deviation of $\hat{C}_{PI}(prf)$ when the target PCI is 0.67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90 , $n = 20$ for Normal(0, 0.1) distribution.

$C_{PI}(prf)$	m	$MBDA$	$MBDB$	$BD3_A$	$BD3_B$	$SubMBDA$	$SubMBDB$	$SubBD3_A$	$SubBD3_B$	$TMMBDA$	$TMMBDB$	$TMBD3_A$	$TMBD3_B$	PW_A	PW_B		
0.67	20	-	1.095	-	1.182	1.094	1.181	1.094	1.181	1.157	1.093	1.157	1.093	1.177	1.101		
		-	0.078	-	0.083	0.076	0.082	0.075	0.082	0.075	0.072	0.074	0.072	0.074	0.074	0.075	
		-	0.986	-	0.985	0.986	1.045	0.987	1.045	0.987	1.035	0.986	1.035	0.986	1.049	0.995	
	30	-	0.054	-	0.055	0.064	0.053	0.066	0.053	0.066	0.060	0.053	0.060	0.053	0.061	0.052	
		-	0.874	-	0.874	0.900	0.874	0.900	0.874	0.900	0.899	0.874	0.899	0.874	0.911	0.885	
		-	0.045	-	0.044	0.042	0.045	0.042	0.045	0.042	0.041	0.045	0.041	0.045	0.041	0.044	
	60	-	0.811	-	0.813	0.811	0.811	0.832	0.811	0.832	0.832	0.811	0.832	0.811	0.845	0.826	
		-	0.043	-	0.040	0.036	0.039	0.036	0.039	0.036	0.036	0.039	0.036	0.039	0.035	0.038	
		-	1.616	-	1.616	1.760	1.607	1.761	1.607	1.761	1.705	1.605	1.705	1.605	1.736	1.615	
	1.00	20	-	0.124	-	0.131	0.129	0.103	0.126	0.103	0.101	0.094	0.101	0.094	0.110	0.099	
			-	1.490	-	1.498	1.575	1.490	1.577	1.491	1.577	1.550	1.486	1.550	1.486	1.572	1.499
			-	0.095	-	0.116	0.099	0.094	0.101	0.096	0.101	0.085	0.088	0.085	0.088	0.093	0.091
30		-	1.310	-	1.309	1.356	1.306	1.356	1.307	1.356	1.349	1.306	1.349	1.306	1.368	1.323	
		-	0.058	-	0.066	0.052	0.054	0.052	0.055	0.054	0.049	0.053	0.049	0.053	0.050	0.053	
		-	1.222	-	1.223	1.258	1.224	1.258	1.224	1.258	1.254	1.223	1.254	1.223	1.275	1.244	
60		-	0.050	-	0.053	0.044	0.050	0.044	0.050	0.044	0.043	0.049	0.043	0.049	0.043	0.048	
		-	2.171	-	2.164	2.365	2.147	2.366	2.149	2.366	2.261	2.147	2.261	2.147	2.305	2.160	
		-	0.184	-	0.188	0.213	0.148	0.218	0.148	0.218	0.141	0.137	0.141	0.137	0.162	0.144	
1.33		20	-	1.952	-	1.974	2.097	1.958	2.095	1.956	2.052	1.957	2.052	1.957	2.087	1.976	
			-	0.144	-	0.193	0.119	0.130	0.124	0.133	0.124	0.105	0.122	0.105	0.122	0.110	0.125
			-	1.719	-	1.732	1.789	1.721	1.789	1.721	1.789	1.775	1.722	1.775	1.722	1.802	1.745
	30	-	0.088	-	0.097	0.080	0.072	0.080	0.073	0.080	0.076	0.072	0.076	0.072	0.076	0.072	
		-	1.629	-	1.632	1.676	1.625	1.677	1.626	1.677	1.666	1.624	1.666	1.624	1.695	1.652	
		-	0.077	-	0.081	0.060	0.067	0.060	0.067	0.060	0.056	0.065	0.056	0.065	0.057	0.065	
	60	-	2.702	-	2.686	2.978	2.698	2.982	2.701	2.982	2.845	2.689	2.845	2.689	2.910	2.703	
		-	0.234	-	0.277	0.217	0.197	0.216	0.208	0.216	0.153	0.166	0.153	0.166	0.180	0.181	
		-	2.443	-	2.460	2.619	2.438	2.620	2.438	2.620	2.560	2.440	2.560	2.440	2.600	2.459	
	1.66	20	-	0.170	-	0.244	0.156	0.136	0.156	0.134	0.125	0.126	0.125	0.126	0.141	0.134	
			-	2.149	-	2.166	2.240	2.151	2.240	2.150	2.240	2.216	2.149	2.216	2.149	2.248	2.175
			-	0.141	-	0.156	0.098	0.107	0.099	0.107	0.099	0.090	0.105	0.090	0.105	0.093	0.104
30		-	2.030	-	2.027	2.087	2.025	2.088	2.025	2.088	2.075	2.022	2.075	2.022	2.114	2.060	
		-	0.106	-	0.114	0.078	0.079	0.080	0.080	0.080	0.074	0.077	0.074	0.077	0.074	0.077	
		-	3.210	-	3.301	3.568	3.207	3.582	3.215	3.582	3.408	3.215	3.408	3.215	3.493	3.241	
60		-	0.301	-	0.446	0.309	0.228	0.321	0.234	0.321	0.193	0.192	0.193	0.192	0.242	0.205	
		-	2.949	-	2.983	3.201	2.956	3.198	2.954	3.198	3.103	2.946	3.103	2.946	3.161	2.974	
		-	0.235	-	0.297	0.201	0.173	0.209	0.177	0.209	0.150	0.156	0.150	0.156	0.167	0.162	
90		-	2.619	-	2.608	2.719	2.605	2.714	2.601	2.601	2.685	2.601	2.685	2.601	2.731	2.637	
		-	0.146	-	0.184	0.119	0.126	0.117	0.122	0.117	0.102	0.109	0.102	0.109	0.110	0.113	
		-	2.442	-	2.703	2.444	2.431	2.503	2.430	2.490	2.490	2.429	2.490	2.429	2.533	2.472	
2.00	-	0.151	-	0.262	0.154	0.116	0.121	0.117	0.121	0.105	0.112	0.105	0.112	0.107	0.111		

TABLE 2.2. Mean and standard deviation of $\hat{C}_{pl(prf)}$ when the target PCI is 0.67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90, $n = 20$ for LogNormal(0, 0.3) distribution.

$C_{pl(prf)}$	m	$MBDA$	$MBDB$	$BD3_A$	$BD3_B$	$SubMBDA$	$SubMBDB$	$SubBD3_A$	$SubBD3_B$	$TMMBDA$	$TMMBDB$	$TMBD3_A$	$TMBD3_B$	PW_A	PW_B
0.67	20	-	0.945	-	0.948	0.985	0.943	0.985	0.943	0.976	0.948	0.976	0.948	0.986	0.948
	30	-	0.050	-	0.046	0.057	0.048	0.056	0.047	0.042	0.043	0.042	0.043	0.050	0.047
	60	-	0.886	-	0.886	0.900	0.882	0.900	0.882	0.908	0.892	0.908	0.892	0.907	0.887
	90	-	0.043	-	0.048	0.042	0.043	0.041	0.043	0.034	0.039	0.034	0.039	0.038	0.042
	20	-	0.810	-	0.813	0.812	0.804	0.812	0.804	0.828	0.820	0.828	0.820	0.821	0.811
	30	-	0.033	-	0.035	0.026	0.029	0.026	0.029	0.024	0.027	0.024	0.027	0.025	0.028
1.00	20	-	0.770	-	0.769	0.772	0.764	0.772	0.764	0.790	0.781	0.792	0.782	0.783	0.773
	30	-	0.027	-	0.033	0.023	0.023	0.023	0.023	0.020	0.021	0.020	0.020	0.021	0.021
	60	-	1.385	-	1.372	1.501	1.400	1.497	1.398	1.407	1.365	1.407	1.365	1.475	1.409
	90	-	0.085	-	0.089	0.101	0.072	0.099	0.072	0.062	0.060	0.062	0.060	0.077	0.069
	20	-	1.305	-	1.313	1.373	1.323	1.372	1.322	1.316	1.294	1.316	1.294	1.366	1.331
	30	-	0.063	-	0.080	0.069	0.060	0.070	0.059	0.049	0.051	0.049	0.051	0.061	0.057
1.33	20	-	1.187	-	1.185	1.216	1.194	1.216	1.194	1.190	1.178	1.190	1.178	1.222	1.205
	30	-	0.040	-	0.041	0.038	0.038	0.038	0.038	0.032	0.035	0.032	0.035	0.036	0.038
	60	-	1.144	-	1.195	1.142	1.164	1.148	1.165	1.149	1.137	1.147	1.136	1.175	1.162
	90	-	0.035	-	0.058	0.034	0.031	0.032	0.032	0.033	0.027	0.030	0.027	0.029	0.032
	20	-	1.847	-	1.811	2.025	1.870	2.022	1.868	1.837	1.787	1.837	1.787	1.960	1.874
	30	-	0.139	-	0.157	0.141	0.111	0.147	0.111	0.069	0.080	0.069	0.080	0.101	0.105
1.66	20	-	1.710	-	1.686	1.831	1.743	1.829	1.739	1.718	1.682	1.718	1.682	1.813	1.759
	30	-	0.106	-	0.116	0.077	0.067	0.080	0.073	0.056	0.058	0.056	0.058	0.071	0.071
	60	-	1.588	-	1.569	1.638	1.599	1.638	1.598	1.566	1.548	1.566	1.548	1.641	1.614
	90	-	0.081	-	0.082	0.049	0.053	0.049	0.053	0.041	0.046	0.041	0.046	0.048	0.053
	20	-	1.500	-	1.505	1.541	1.514	1.542	1.514	1.495	1.477	1.491	1.475	1.556	1.536
	30	-	0.065	-	0.069	0.045	0.044	0.045	0.043	0.037	0.037	0.035	0.036	0.040	0.041
2.00	20	-	2.250	-	2.262	2.532	2.319	2.517	2.314	2.271	2.205	2.271	2.205	2.458	2.344
	30	-	0.196	-	0.258	0.220	0.143	0.187	0.134	0.108	0.103	0.108	0.103	0.155	0.134
	60	-	2.132	-	2.085	2.292	2.158	2.290	2.160	2.119	2.064	2.119	2.064	2.267	2.181
	90	-	0.165	-	0.187	0.114	0.104	0.122	0.107	0.075	0.079	0.075	0.079	0.092	0.093
	20	-	1.960	-	1.942	2.025	1.971	2.027	1.970	1.924	1.895	1.924	1.895	2.034	1.996
	30	-	0.109	-	0.128	0.070	0.070	0.070	0.068	0.068	0.047	0.050	0.047	0.050	0.061
2.314	20	-	1.868	-	2.115	1.861	1.932	1.896	1.933	1.897	1.832	1.846	1.828	1.948	1.924
	30	-	0.089	-	0.332	0.092	0.056	0.057	0.061	0.041	0.047	0.041	0.046	0.046	0.051
	60	-	2.727	-	2.735	3.083	2.801	3.130	2.812	2.738	2.643	2.738	2.643	2.993	2.827
	90	-	0.267	-	0.418	0.228	0.204	0.345	0.209	0.113	0.125	0.113	0.125	0.172	0.169
	20	-	2.564	-	2.477	2.774	2.610	2.771	2.609	2.530	2.467	2.530	2.467	2.728	2.627
	30	-	0.204	-	0.224	0.124	0.112	0.123	0.115	0.072	0.085	0.072	0.085	0.092	0.098
2.458	20	-	2.370	-	2.332	2.451	2.384	2.450	2.385	2.299	2.273	2.299	2.273	2.450	2.410
	30	-	0.145	-	0.211	0.094	0.095	0.091	0.091	0.065	0.070	0.065	0.070	0.079	0.085
	60	-	2.268	-	2.619	2.238	2.336	2.285	2.338	2.216	2.189	2.216	2.189	2.347	2.314
	90	-	0.134	-	0.440	0.147	0.065	0.072	0.066	0.047	0.054	0.045	0.052	0.054	0.063

TABLE 2.3. Mean and standard deviation of $\hat{C}_{pl(prf)}$ when the target PCI is 0.67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90, $n = 20$ for Weibull(6,1) distribution.

$C_{pl(prf)}$	m	$MBDA$	$MBDB$	$BD3_A$	$BD3_B$	$SubMBDA$	$SubMBDB$	$SubBD3_A$	$SubBD3_B$	$TMMBDA$	$TMMBDB$	$TMBD3_A$	$TMBD3_B$	PWA	PWB	
0.67	20	-	1.071	-	1.073	1.173	1.072	1.176	1.072	1.156	1.073	1.156	1.073	1.170	1.078	
		-	0.081	-	0.084	0.104	0.081	0.113	0.081	0.087	0.082	0.087	0.082	0.088	0.079	
		-	0.973	-	0.972	1.035	0.974	1.035	0.974	0.974	1.031	0.973	1.031	0.973	1.041	0.982
	30	-	0.056	-	0.061	0.058	0.056	0.058	0.056	0.056	0.057	0.057	0.057	0.057	0.058	0.055
		-	0.861	-	0.861	0.894	0.865	0.894	0.865	0.865	0.890	0.861	0.890	0.861	0.905	0.877
		-	0.044	-	0.044	0.040	0.041	0.041	0.041	0.041	0.041	0.043	0.041	0.043	0.040	0.041
	60	-	0.806	-	0.806	0.830	0.809	0.830	0.809	0.809	0.827	0.805	0.827	0.805	0.844	0.824
		-	0.030	-	0.030	0.028	0.028	0.028	0.028	0.028	0.028	0.029	0.028	0.029	0.027	0.027
		-	1.629	-	1.659	1.787	1.621	1.621	1.793	1.623	1.763	1.639	1.763	1.639	1.773	1.631
	1.00	20	-	0.125	-	0.152	0.150	0.120	0.149	0.119	0.124	0.115	0.124	0.115	0.136	0.116
			-	1.471	-	1.481	1.566	1.465	1.565	1.464	1.558	1.476	1.558	1.476	1.562	1.474
			-	0.094	-	0.101	0.099	0.085	0.102	0.084	0.084	0.089	0.081	0.089	0.081	0.093
30		-	1.303	-	1.306	1.345	1.298	1.344	1.298	1.348	1.348	1.306	1.348	1.306	1.356	1.315
		-	0.063	-	0.064	0.056	0.057	0.057	0.057	0.057	0.055	0.058	0.055	0.058	0.054	0.055
		-	1.214	-	1.212	1.246	1.210	1.246	1.210	1.249	1.215	1.249	1.215	1.249	1.263	1.231
60		-	0.051	-	0.050	0.048	0.048	0.048	0.048	0.048	0.048	0.049	0.048	0.049	0.046	0.047
		-	2.195	-	2.209	2.410	2.159	2.415	2.163	2.415	2.361	2.196	2.361	2.196	2.366	2.174
		-	0.191	-	0.209	0.201	0.158	0.209	0.168	0.168	0.146	0.149	0.146	0.149	0.158	0.150
1.33		20	-	2.010	-	2.026	2.119	1.975	2.122	1.975	2.107	2.007	2.107	2.007	2.105	1.994
			-	0.160	-	0.180	0.156	0.130	0.159	0.131	0.132	0.129	0.132	0.129	0.131	0.127
			-	1.757	-	1.760	1.813	1.746	1.812	1.746	1.820	1.761	1.820	1.761	1.823	1.765
	30	-	0.103	-	0.121	0.083	0.091	0.081	0.081	0.090	0.082	0.092	0.082	0.092	0.080	0.089
		-	1.639	-	1.640	1.669	1.629	1.669	1.629	1.629	1.679	1.643	1.679	1.643	1.690	1.655
		-	0.079	-	0.084	0.074	0.073	0.074	0.073	0.073	0.072	0.073	0.072	0.073	0.069	0.070
	60	-	2.756	-	2.752	3.029	2.714	3.051	2.717	2.960	2.765	2.960	2.765	2.960	2.953	2.726
		-	0.255	-	0.290	0.290	0.194	0.299	0.192	0.192	0.201	0.180	0.201	0.180	0.219	0.180
		-	2.464	-	2.481	2.607	2.413	2.606	2.412	2.605	2.457	2.605	2.457	2.605	2.597	2.435
	1.66	20	-	0.192	-	0.205	0.168	0.137	0.173	0.143	0.144	0.137	0.144	0.137	0.154	0.139
			-	2.161	-	2.190	2.242	2.147	2.244	2.150	2.253	2.175	2.253	2.175	2.251	2.173
			-	0.124	-	0.145	0.108	0.099	0.109	0.099	0.102	0.102	0.102	0.102	0.102	0.101
30		-	2.039	-	2.038	2.084	2.017	2.083	2.017	2.083	2.045	2.045	2.103	2.045	2.109	2.052
		-	0.114	-	0.113	0.081	0.082	0.080	0.080	0.082	0.079	0.081	0.079	0.081	0.078	0.078
		-	3.294	-	3.373	3.599	3.236	3.613	3.243	3.552	3.306	3.552	3.306	3.552	3.555	3.260
60		-	0.333	-	0.431	0.320	0.239	0.326	0.248	0.248	0.254	0.228	0.254	0.228	0.291	0.237
		-	3.008	-	3.040	3.176	2.949	3.171	2.949	3.182	3.012	3.182	3.012	3.182	3.167	2.978
		-	0.250	-	0.232	0.197	0.169	0.197	0.170	0.170	0.179	0.176	0.179	0.176	0.190	0.176
90		-	2.636	-	2.641	2.710	2.603	2.712	2.604	2.729	2.643	2.729	2.643	2.729	2.724	2.634
		-	0.183	-	0.197	0.131	0.132	0.134	0.133	0.133	0.121	0.128	0.121	0.128	0.121	0.125
		-	2.436	-	2.456	2.492	2.416	2.495	2.419	2.520	2.455	2.520	2.455	2.520	2.523	2.459
90	-	0.134	-	0.133	0.106	0.104	0.107	0.107	0.107	0.101	0.104	0.101	0.104	0.098	0.098	

TABLE 2.4. Mean and standard deviation of $\hat{C}_{pl(prf)}$ when the target PCI is 0.67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90, $n = 20$ for Gamma(1.2, 5) distribution.

$C_{pl(prf)}$	m	$MBDA$	$MBDB$	$BD3_A$	$BD3_B$	$SubMBDA$	$SubMBDB$	$SubBD3_A$	$SubBD3_B$	$TMMBDA$	$TMMBDB$	$TMBD3_A$	$TMBD3_B$	PW_A	PW_B		
0.67	20	-	0.779	-	0.793	0.694	0.754	0.694	0.755	0.800	0.810	0.804	0.812	0.725	0.747		
		-	0.046	-	0.053	0.040	0.035	0.044	0.034	0.034	0.018	0.018	0.017	0.018	0.027	0.028	
		-	0.756	-	0.789	0.688	0.727	0.689	0.729	0.776	0.776	0.785	0.787	0.792	0.707	0.721	
	30	-	0.051	-	0.051	0.031	0.029	0.031	0.028	0.028	0.017	0.017	0.014	0.014	0.022	0.023	
		-	0.725	-	0.740	0.674	0.698	0.675	0.699	0.752	0.752	0.759	0.771	0.774	0.688	0.696	
		-	0.049	-	0.053	0.023	0.024	0.022	0.024	0.013	0.013	0.013	0.009	0.010	0.014	0.015	
	60	-	0.726	-	0.471	0.725	0.672	0.689	0.672	0.689	0.748	0.753	0.767	0.770	0.682	0.688	
		-	0.044	-	0.234	0.054	0.019	0.019	0.019	0.019	0.009	0.009	0.006	0.006	0.011	0.011	
		-	1.083	-	1.081	1.121	1.095	1.119	1.094	1.078	1.078	1.073	1.076	1.072	1.111	1.099	
	1.00	20	-	0.026	-	0.029	0.031	0.025	0.031	0.025	0.018	0.018	0.017	0.018	0.026	0.025	
			-	1.059	-	1.055	1.078	1.066	1.079	1.066	1.056	1.052	1.053	1.050	1.077	1.070	
			-	0.017	-	0.018	0.019	0.016	0.020	0.016	0.016	0.013	0.012	0.012	0.012	0.018	0.016
30		-	1.030	-	1.029	1.039	1.035	1.035	1.038	1.035	1.029	1.028	1.027	1.026	1.040	1.038	
		-	0.009	-	0.009	0.010	0.009	0.010	0.009	0.007	0.007	0.007	0.006	0.007	0.009	0.010	
		-	1.021	-	1.041	1.021	1.024	1.023	1.024	1.023	1.019	1.019	1.017	1.017	1.026	1.026	
60		-	0.007	-	0.027	0.006	0.006	0.006	0.006	0.006	0.004	0.005	0.004	0.005	0.006	0.006	
		-	1.391	-	1.359	1.557	1.441	1.557	1.440	1.440	1.359	1.341	1.352	1.336	1.501	1.457	
		-	0.075	-	0.088	0.072	0.051	0.084	0.049	0.049	0.028	0.028	0.025	0.026	0.047	0.043	
1.33		20	-	1.358	-	1.317	1.464	1.401	1.467	1.402	1.333	1.316	1.313	1.304	1.441	1.414	
			-	0.070	-	0.076	0.052	0.041	0.054	0.040	0.040	0.022	0.021	0.017	0.018	0.029	0.028
			-	1.341	-	1.311	1.399	1.369	1.398	1.369	1.304	1.295	1.281	1.278	1.389	1.378	
	30	-	0.054	-	0.069	0.027	0.028	0.028	0.028	0.029	0.015	0.014	0.010	0.010	0.019	0.019	
		-	1.317	-	1.318	1.375	1.356	1.376	1.357	1.357	1.290	1.283	1.268	1.265	1.370	1.363	
		-	0.055	-	0.051	0.020	0.020	0.022	0.022	0.011	0.011	0.011	0.007	0.007	0.013	0.014	
	60	-	1.711	-	1.628	1.988	1.784	1.784	1.983	1.781	1.639	1.608	1.625	1.599	1.887	1.808	
		-	0.172	-	0.145	0.137	0.091	0.133	0.094	0.094	0.040	0.039	0.034	0.035	0.073	0.066	
		-	1.667	-	1.593	1.844	1.734	1.848	1.734	1.608	1.580	1.576	1.576	1.559	1.806	1.761	
	1.66	20	-	0.112	-	0.135	0.077	0.067	0.084	0.073	0.031	0.030	0.023	0.024	0.045	0.046	
			-	1.632	-	1.621	1.752	1.699	1.753	1.701	1.580	1.563	1.533	1.527	1.734	1.717	
			-	0.110	-	0.128	0.055	0.051	0.058	0.053	0.028	0.028	0.029	0.015	0.016	0.030	0.030
30		-	1.617	-	2.202	1.595	1.725	1.688	1.723	1.685	1.562	1.552	1.517	1.513	1.711	1.700	
		-	0.108	-	1.023	0.113	0.041	0.040	0.039	0.041	0.020	0.022	0.012	0.013	0.024	0.026	
		-	2.008	-	1.925	2.432	2.150	2.418	2.138	1.934	1.886	1.913	1.872	2.290	2.181		
60		-	0.219	-	0.251	0.190	0.136	0.182	0.126	0.126	0.053	0.052	0.050	0.049	0.104	0.087	
		-	1.957	-	1.891	2.273	2.109	2.275	2.104	1.906	1.870	1.855	1.834	2.197	2.136		
		-	0.167	-	0.198	0.123	0.101	0.129	0.104	0.045	0.045	0.048	0.031	0.035	0.066	0.069	
90		-	1.910	-	1.871	2.131	2.051	2.131	2.052	1.862	1.833	1.797	1.786	2.102	2.073		
		-	0.156	-	0.178	0.068	0.070	0.067	0.067	0.034	0.033	0.020	0.021	0.037	0.040		
		-	1.960	-	1.920	2.088	2.031	2.091	2.034	1.840	1.820	1.776	1.769	2.061	2.042		
2.00	-	0.184	-	0.169	0.066	0.065	0.068	0.066	0.027	0.027	0.016	0.017	0.034	0.035			

TABLE 2.5. Mean and standard deviation of $\hat{C}_{pl(prf)}$ when the target PCI is 0.67, 1, 1.33, 1.66 and 2, $m = 20, 30, 60$ and 90, $n = 20$ for Beta(5, 5) distribution.

$C_{pl(prf)}$	m	$MBDA$	$MBDB$	$BD3_A$	$BD3_B$	Sub_{MBDA}	Sub_{MBDB}	Sub_{BD3_A}	Sub_{BD3_B}	TM_{MBDA}	TM_{MBDB}	TM_{BD3_A}	TM_{BD3_B}	PW_A	PW_B
0.67	20	-	0.969	-	0.968	1.023	0.969	1.023	0.969	1.007	0.969	1.007	0.969	1.018	0.975
	-	0.058	-	0.059	0.072	0.058	0.058	0.074	0.058	0.054	0.057	0.054	0.057	0.057	0.057
	30	-	0.895	-	0.893	0.922	0.896	0.923	0.896	0.920	0.897	0.920	0.897	0.928	0.903
	-	0.042	-	0.047	0.044	0.041	0.044	0.041	0.041	0.043	0.042	0.043	0.042	0.043	0.041
	60	-	0.804	-	0.801	0.818	0.806	0.818	0.806	0.821	0.806	0.821	0.806	0.829	0.814
	-	0.035	-	0.038	0.027	0.029	0.029	0.027	0.029	0.027	0.029	0.027	0.029	0.027	0.028
1.00	20	-	0.770	0.746	0.770	0.779	0.771	0.779	0.771	0.781	0.772	0.781	0.772	0.791	0.782
	-	0.028	0.046	0.046	0.032	0.024	0.025	0.024	0.025	0.023	0.024	0.023	0.024	0.023	0.024
	30	-	1.444	-	1.439	1.548	1.441	1.552	1.442	1.504	1.438	1.504	1.438	1.530	1.447
	-	0.100	-	0.086	0.102	0.087	0.108	0.090	0.108	0.082	0.083	0.082	0.083	0.094	0.088
	60	-	1.346	-	1.346	1.401	1.345	1.402	1.345	1.379	1.343	1.379	1.343	1.398	1.353
	-	0.074	-	0.080	0.075	0.070	0.075	0.070	0.075	0.063	0.066	0.063	0.066	0.068	0.068
1.33	20	-	1.210	-	1.211	1.236	1.209	1.236	1.209	1.231	1.209	1.231	1.209	1.245	1.222
	-	0.043	-	0.046	0.041	0.041	0.041	0.042	0.041	0.039	0.040	0.039	0.040	0.039	0.041
	30	-	1.159	-	1.159	1.174	1.157	1.174	1.158	1.172	1.158	1.172	1.158	1.186	1.173
	-	0.034	-	0.033	0.030	0.033	0.031	0.033	0.031	0.030	0.033	0.030	0.033	0.029	0.033
	60	-	1.924	-	1.953	2.071	1.911	2.077	1.912	1.996	1.913	1.996	1.913	2.039	1.929
	-	0.132	-	0.186	0.132	0.111	0.126	0.108	0.126	0.095	0.097	0.095	0.097	0.112	0.106
1.66	20	-	1.779	-	1.780	1.866	1.774	1.867	1.776	1.830	1.775	1.830	1.775	1.854	1.786
	-	0.108	-	0.134	0.095	0.088	0.096	0.096	0.089	0.075	0.081	0.075	0.081	0.084	0.085
	30	-	1.602	-	1.609	1.644	1.604	1.644	1.604	1.629	1.603	1.629	1.603	1.651	1.621
	-	0.071	-	0.089	0.057	0.058	0.059	0.058	0.058	0.052	0.057	0.052	0.057	0.054	0.057
	60	-	1.526	1.690	1.529	1.550	1.523	1.549	1.523	1.542	1.523	1.542	1.523	1.562	1.542
	-	0.059	0.348	0.057	0.045	0.044	0.044	0.046	0.045	0.040	0.043	0.040	0.043	0.041	0.043
2.00	20	-	2.396	-	2.427	2.615	2.402	2.619	2.399	2.500	2.393	2.500	2.393	2.555	2.407
	-	0.173	-	0.226	0.191	0.153	0.207	0.149	0.149	0.120	0.116	0.120	0.116	0.149	0.128
	30	-	2.211	-	2.223	2.332	2.206	2.328	2.204	2.274	2.204	2.274	2.204	2.308	2.220
	-	0.152	-	0.173	0.114	0.100	0.100	0.114	0.100	0.087	0.090	0.087	0.090	0.098	0.096
	60	-	2.024	-	2.047	2.063	2.010	2.064	2.010	2.046	2.013	2.046	2.013	2.072	2.033
	-	0.102	-	0.152	0.073	0.069	0.075	0.072	0.072	0.066	0.066	0.066	0.066	0.068	0.068
36.945	20	5.781	1.922	5.843	1.927	1.951	1.914	1.950	1.913	1.939	1.913	1.939	1.913	1.965	1.937
	-	0.090	0.090	36.942	0.091	0.062	0.062	0.061	0.061	0.053	0.055	0.053	0.055	0.053	0.054
	30	-	2.895	-	2.872	3.187	2.904	3.184	2.892	3.022	2.891	3.022	2.891	3.095	2.914
	-	0.251	-	0.340	0.247	0.183	0.257	0.191	0.191	0.150	0.146	0.150	0.146	0.178	0.169
	60	-	2.703	-	2.693	2.818	2.661	2.821	2.662	2.742	2.657	2.742	2.657	2.787	2.677
	-	0.227	-	0.285	0.155	0.146	0.158	0.147	0.147	0.117	0.118	0.117	0.118	0.129	0.129
36.945	20	-	2.434	-	2.439	2.488	2.415	2.487	2.414	2.458	2.413	2.458	2.413	2.488	2.437
	-	0.152	-	0.172	0.095	0.091	0.093	0.088	0.088	0.083	0.082	0.083	0.082	0.089	0.088
	30	-	2.300	-	2.305	2.334	2.297	2.333	2.295	2.319	2.294	2.319	2.294	2.353	2.326
	-	0.111	-	0.111	0.068	0.069	0.068	0.069	0.069	0.055	0.057	0.055	0.057	0.058	0.059
	60	-	2.300	-	2.305	2.334	2.297	2.333	2.295	2.319	2.294	2.319	2.294	2.353	2.326
	-	0.111	-	0.111	0.068	0.069	0.068	0.069	0.069	0.055	0.057	0.055	0.057	0.058	0.059

TABLE 2.6. Relative bias of the several estimators of $C_{pl}(prf)$ when the target PCI is 0.67, 1, 1.33, 1.66 and 90, $m = 20, 30, 60$ and 90, $n = 20$ for Normal(0, 0.1) distribution.

$C_{pl}(prf)$	m	MBD_A	MBD_B	$BD3_A$	$BD3_B$	Sub_{MBDA}	Sub_{MBDB}	Sub_{BD3A}	Sub_{BD3B}	TM_{MBDA}	TM_{MBDB}	TM_{BD3A}	TM_{BD3B}	PW_A	PW_B
0.67	20	-	0.635	-	0.634	0.764	0.634	0.762	0.633	0.726	0.632	0.726	0.632	0.757	0.643
	30	-	0.472	-	0.470	0.559	0.472	0.560	0.473	0.545	0.472	0.545	0.472	0.566	0.485
	60	-	0.305	-	0.304	0.344	0.304	0.343	0.304	0.341	0.304	0.341	0.304	0.359	0.321
	90	-	0.210	-	0.213	0.241	0.211	0.241	0.211	0.241	0.211	0.241	0.211	0.261	0.232
1.00	20	-	0.616	-	0.616	0.760	0.607	0.761	0.607	0.705	0.605	0.705	0.605	0.736	0.615
	30	-	0.490	-	0.498	0.575	0.490	0.577	0.491	0.550	0.486	0.550	0.486	0.572	0.499
	60	-	0.310	-	0.309	0.356	0.306	0.356	0.307	0.349	0.306	0.349	0.306	0.368	0.323
	90	-	0.222	-	0.223	0.258	0.224	0.258	0.224	0.254	0.223	0.254	0.223	0.275	0.244
1.33	20	-	0.633	-	0.627	0.779	0.614	0.779	0.616	0.700	0.614	0.700	0.614	0.733	0.624
	30	-	0.467	-	0.485	0.577	0.473	0.575	0.471	0.543	0.471	0.543	0.471	0.569	0.485
	60	-	0.293	-	0.302	0.345	0.294	0.345	0.294	0.335	0.294	0.335	0.294	0.355	0.312
	90	-	0.225	-	0.227	0.260	0.222	0.261	0.223	0.253	0.221	0.253	0.221	0.275	0.242
1.66	20	-	0.627	-	0.618	0.794	0.625	0.796	0.627	0.714	0.620	0.714	0.620	0.753	0.628
	30	-	0.471	-	0.482	0.578	0.469	0.578	0.469	0.542	0.470	0.542	0.470	0.566	0.481
	60	-	0.295	-	0.305	0.349	0.296	0.349	0.295	0.335	0.295	0.335	0.295	0.354	0.310
	90	-	0.223	-	0.221	0.257	0.220	0.258	0.220	0.250	0.218	0.250	0.218	0.274	0.241
2.00	20	-	0.605	-	0.650	0.784	0.603	0.791	0.608	0.704	0.608	0.704	0.608	0.746	0.620
	30	-	0.474	-	0.492	0.600	0.478	0.599	0.477	0.551	0.473	0.551	0.473	0.581	0.487
	60	-	0.310	-	0.304	0.359	0.303	0.357	0.301	0.343	0.301	0.343	0.301	0.365	0.319
	90	-	0.221	0.352	0.222	0.252	0.216	0.251	0.215	0.245	0.214	0.245	0.214	0.266	0.236

TABLE 2.7. Relative bias of the several estimators of $C_{pl}(prf)$ when the target PCI is 0.67, 1, 1.33, 1.66 and 90, $n = 20$ for Gamma(1.2, 5) distribution.

$C_{pl}(prf)$	m	MBD_A	MBD_B	$BD3_A$	$BD3_B$	Sub_{MBDA}	Sub_{MBDB}	Sub_{BD3A}	Sub_{BD3B}	TM_{MBDA}	TM_{MBDB}	TM_{BD3A}	TM_{BD3B}	PW_A	PW_B
0.67	20	-	0.163	-	0.184	0.036	0.125	0.036	0.126	0.194	0.208	0.199	0.212	0.082	0.115
	30	-	0.128	-	0.177	0.027	0.086	0.028	0.088	0.158	0.172	0.175	0.183	0.055	0.077
	60	-	0.082	-	0.105	0.006	0.041	0.008	0.043	0.122	0.133	0.151	0.155	0.027	0.039
1.00	20	-	0.083	-	0.081	0.003	0.028	0.003	0.029	0.117	0.124	0.145	0.149	0.019	0.027
	30	-	0.059	-	0.055	0.078	0.066	0.079	0.066	0.056	0.052	0.053	0.050	0.077	0.070
	60	-	0.030	-	0.029	0.039	0.035	0.038	0.035	0.029	0.028	0.027	0.026	0.040	0.038
1.33	20	-	0.021	0.041	0.021	0.024	0.023	0.024	0.023	0.019	0.019	0.017	0.017	0.026	0.026
	30	-	0.046	-	0.022	0.170	0.084	0.171	0.083	0.022	0.008	0.017	0.004	0.128	0.095
	60	-	0.021	-	-0.009	0.101	0.054	0.103	0.054	0.002	-0.010	-0.012	-0.019	0.083	0.064
1.66	20	-	0.008	-	-0.014	0.052	0.030	0.051	0.029	-0.020	-0.026	-0.037	-0.039	0.044	0.036
	30	-	-0.009	-	-0.009	0.034	0.020	0.034	0.020	-0.030	-0.035	-0.047	-0.049	0.030	0.025
	60	-	0.031	-	-0.020	0.197	0.075	0.194	0.073	-0.012	-0.032	-0.021	-0.036	0.137	0.089
2.00	20	-	0.004	-	-0.040	0.111	0.044	0.113	0.045	-0.032	-0.048	-0.050	-0.061	0.088	0.061
	30	-	-0.017	-	-0.024	0.055	0.024	0.056	0.025	-0.048	-0.058	-0.076	-0.080	0.045	0.034
	60	-	-0.026	0.327	-0.039	0.039	0.017	0.038	0.015	-0.059	-0.065	-0.086	-0.089	0.031	0.024
90	20	-	0.004	-	-0.037	0.216	0.075	0.209	0.069	-0.033	-0.057	-0.044	-0.064	0.145	0.090
	30	-	-0.022	-	-0.054	0.136	0.055	0.137	0.052	-0.047	-0.065	-0.072	-0.083	0.099	0.068
	60	-	-0.045	-	-0.064	0.065	0.026	0.065	0.026	-0.069	-0.084	-0.101	-0.107	0.051	0.037
90	20	-	-0.020	-	-0.040	0.044	0.016	0.046	0.017	-0.080	-0.090	-0.112	-0.116	0.031	0.021

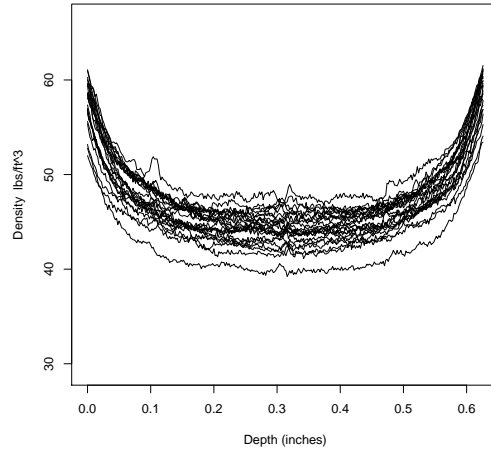


FIGURE 2.5. Vertical Density Profile of 24 particleboards.

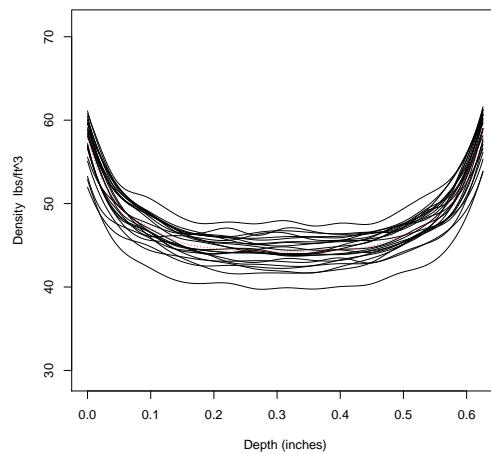
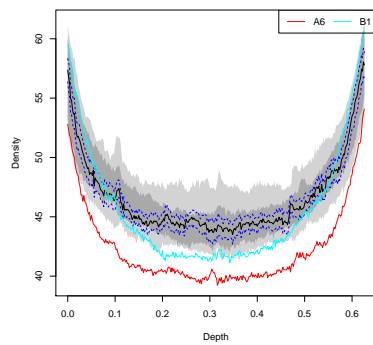
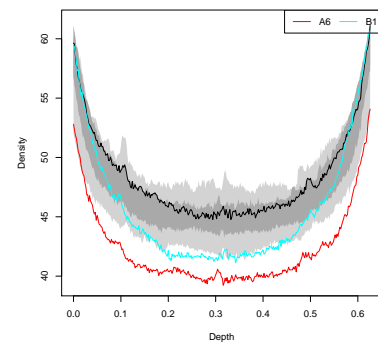


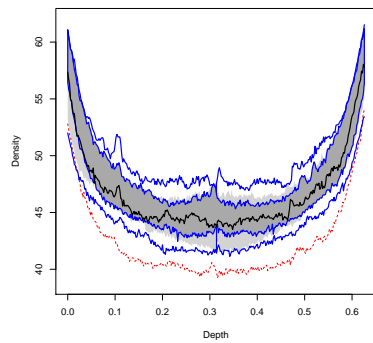
FIGURE 2.6. Spline fits for the Vertical Density Profile data.



(A) Functional bagplot



(B) Functional HDR boxplot



(C) Functional boxplot

FIGURE 2.7. Three graphical methods for visualizing VDP data: a) Functional bagplot with factor=2.58. The red and turquoise curves are the outliers detected. b) The functional HDR boxplot for $\alpha = 0.01$. The red and turquoise curves are the outliers detected. c) The functional boxplot with factor=1.5. The red curve is the outlier detected.

outlier, see Figure 2.7. This last result coincides with the outlier obtained via some non-parametric methods employed by [107]. The profile of the board A6 is excluded from the analysis.

For the panel types recognized in ANSI Standard 208.1, three density classes are identified: high, medium and low. With wood at 7 percent moisture content, high class approximates density values of over 50 pounds per cubic foot (lb/ft^3), while medium class includes density values between 40 and 50 lb/ft^3 , and less than 40 lb/ft^3 is covered by low class, see [15]. We consider that the process is of medium density class, thus $USL(t) = 50$ and $LSL(t) = 40$.

The median function and the quantile functions, $Y_{0.99865}$ and $Y_{0.00135}$, are estimated out of spline fits for the VDP data using the concept of band depth. Figure 2.8 shows the specification functions and the quantiles 0.99865, 0.5 and 0.00135 using MBD.

We propose the following procedure to construct a confidence interval for the PCIs:

- A bootstrap sample $y_1^*(t), \dots, y_m^*(t)$ is obtained by random sampling with replacement from the original functions $y_1(t), \dots, y_m(t)$. Each selected element is a curve.
- The proposed PCIs are calculated using the bootstrap sample.
- This process is repeated B times, so that a collection of B estimations of the PCIs is obtained. The $100(1 - \alpha)\%$ confidence interval of the PCIs is given by the percentiles $\alpha/2$ and $1 - \alpha/2$ calculated from B estimations of the PCIs.

Table 2.8 shows the estimations and the confidence intervals obtained for the index $C_{pk(prf)}$ using the estimators based on depth functions with method B with $alpha = 0.05$. The confidence intervals are calculated from 1000 bootstrap samples and indicate that the process is capable. However, for some subintervals of depth the functions do not meet the specifications, see Figure 2.8. The percentile intervals are asymmetrical to the right side. In general, the bootstrap percentile interval is asymmetrical with asymmetry depending on the sample, [38].

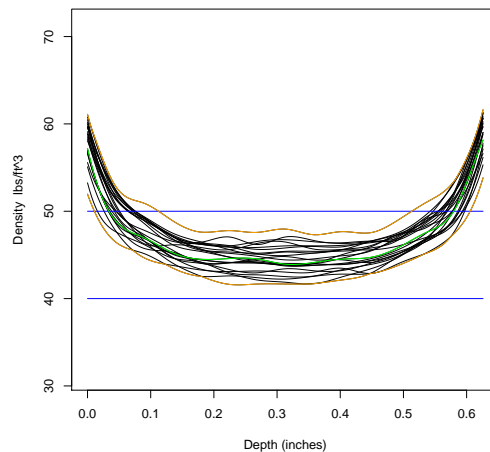


FIGURE 2.8. Spline fit for the Vertical Density Profile (in black), estimation of the median function using MBD (in green), estimations of the quantiles 0.99865 and 0.00135 using MBD (in orange), and specification functions (in blue).

The bootstrap percentile interval is inaccurate when bias or skewness is present in the bootstrap distribution, unless the sample sizes are very large, [37]. Besides, the ordinary bootstrap does not

TABLE 2.8. Estimations and confidence intervals of $C_{pk(prf)}$ for the VDP data.

	$\widehat{C}_{pk(prf)}$	Confidence Interval
<i>MBD</i>	1.027	(1.025, 1.943)
<i>BD3</i>	1.027	(1.024, 1.525)
<i>Sub_{MBD}</i>	1.032	(1.027, 1.815)
<i>Sub_{BD3}</i>	1.032	(1.027, 1.799)
<i>TM_{MBD}</i>	1.031	(1.029, 1.677)
<i>TM_{BD3}</i>	1.031	(1.029, 1.677)

work very well for statistics such as the median or other quantiles that depend heavily on a small number of observations out of a larger sample, [38]. For these reasons a more appropriate bootstrap interval to estimate the PCIs is a topic of future research.

The ANSI standard for panels was established to processes represented by the distribution of a univariate characteristic and not for a functional relationship. This is the reason why the specifications are parallel lines to the horizontal axis. This standard should be revised so that in the future the specifications will be defined as functions of depth. Probably, these specifications will exhibit a bathtub-like curve.

2.1.3 Conclusions

In this section, we propose two methods to measure the process capability of nonlinear profiles without distributional assumptions. Both methods are based on the depth for functional data and constitute an extension of the PCIs proposed by [14] to functional data. The first method calculates the PCIs for the nonlinear profiles as the mean of pointwise PCIs, while the second method considers the PCIs as the ratio of two expressions, where each represent an approach to the distance between two functions.

The results of the simulations show that the estimations of the PCIs for nonlinear profiles based on these two methods present overestimation. However, the first method present greater overestimation than the second method. Both methods are sensitive to the size of the sample. The relative bias decreases when the amount of profiles considered in the sample increases. Should be ideal to use 90 or more profiles.

In general, the estimations of the PCIs for nonlinear profiles using the second method with estimations of the median function and the quantiles 0.99865 and 0.00135 based on subintervals or trimmed means have better performance than the estimations based on pointwise estimations.

Since the band depth and modify band depth incorporate the within-profile correlation, the proposed PCIs takes into account this correlation.

2.2 Evaluation of process capability in nonlinear profiles using Hausdorff distance

In the previous section the indices proposed to measure the process capability characterized by nonlinear profiles were estimated replacing the 0.99865, 0.5 and 0.00135 quantile functions by appropriate estimators based on the concept of band depth or modified band depth. An alternative method based on the Hausdorff distance to estimate the capability of these processes is presented in

this section. The accuracy, precision and execution time of the estimators based on the Hausdorff distance is compared with the estimators based on the concept of band depth or modified band depth.

We consider three sets of estimators for $Y_{0.5}$, $Y_{0.99865}$ and $Y_{0.00135}$ based on the Hausdorff distance proposed by [11], see equation (1.18). Initially, we consider that the distance from the profile i to the rest of profiles is given by:

$$z_i = z(y_i) = \max_{\substack{j=1,\dots,m \\ j \neq i}} h^\dagger(y_i, y_j). \quad (2.12)$$

[11] proposed that "the most representative profile" among the m profiles is the profile R that satisfies:

$$\min_{i=1,\dots,m} (z(y_i)) = z_R. \quad (2.13)$$

The ordered values of z_i provide a center-outward ordering of the profiles, where profiles near the center have smaller z values. The profile with smaller z value is the "center", so that a natural estimator of the median profile is given by $\hat{Y}_{0.5}^{HD+} = \arg_{y \in \{y_1, \dots, y_m\}} \min z(y)$, which coincides with "the most representative profile" proposed by [11]. $\hat{Y}_{0.99865}$ and $\hat{Y}_{0.00135}$ are the curves that envelope the 99.73% of the sample curves according to the ordered z values. If the maximum operator in equation (2.12) is changed by the median or by the minimum operator we obtain two new sets of estimators, denoted by $\hat{Y}_{0.5}^{HD\pm}$ and $\hat{Y}_{0.5}^{HD-}$, respectively.

2.2.1 Simulations

Assuming that the process is in statistical control, for each simulation run, m profiles with s points are generated from the 4-parameter logistic model (see equation (2.11)), with $s = 120$ and $m = 30, 60, 90$ and 120 . Normal(0,0.1), Lognormal(0, 0.3), Weibull(6,1), Gamma(1.2,5) and Beta(5,5) distributions are considered for the random terms ε_{ij} . Each run is replicated $R = 100$ times. In this simulation study, we only consider the index $C_{pl(prf)}$ with target value set at 1. The specification function (LSL) is obtained by solving the equation (2.10).

Initially nine estimators of the median function are considered. First, we take the estimators based on BD_m and MBD , which are denoted as $\hat{Y}_{0.5}^{BD}$ and $\hat{Y}_{0.5}^{MBD}$, respectively. The concepts of BD and MBD can be applied by subintervals, so that the corresponding estimators of the median function are noted as $\hat{Y}_{0.5}^{SubBD}$ and $\hat{Y}_{0.5}^{SubMBD}$. The median function can be estimated using the notion of trimmed mean. These estimators are identified by $\hat{Y}_{0.5}^{TmBD}$ and $\hat{Y}_{0.5}^{TmMBD}$. Finally, three estimators based on the Hausdorff distance are considered: $\hat{Y}_{0.5}^+$, $\hat{Y}_{0.5}^\pm$ and $\hat{Y}_{0.5}^-$. The estimators of $Y_{0.99865}$ and $Y_{0.00135}$ are the curves that envelope the central region $C_{m,0.99763}$.

The execution time, accuracy and precision of the several estimators are compared to determine the best estimator of the index $C_{pl(prf)}$, according to the amount of profiles and the given distribution. Programs developed in R ([73]) are used to generate the simulations. In particular, we have used the function `fbplot` written by [92] to obtain the central regions and the estimations of the median based on BD or MBD . The simulations have been realized on a computer with 4 GB of RAM memory and processor Intel Core i7-2600S of 2.80 GHz.

TABLE 2.9. Execution time in seconds for the estimation of the median using seven estimators.

Number of Profiles (m)	MBD	BD	SubMBD	SubBD	TmMBD	TmBD	Hausdorff Distance
30	0.75	5.46	8.65	84.47	0.75	5.46	0.96
60	2.68	45.77	30.83	576.42	2.68	45.77	2.25
90	6.57	178.76	70.03	1911.57	6.57	178.76	5.77
120	8.83	352.89	89.74	3526.53	8.83	352.89	7.83
180	21.17	1278.04	186.00	11364.59	21.17	1278.04	16.94

The execution time of the estimators is presented in Table 2.9. For example, when the number of profiles is $m = 180$, the median function is estimated in 21.17 and 1278.04 seconds using MBD and BD, respectively; while the Hausdorff distance method uses just 16.94 seconds. If the bands are applied for subintervals, the execution time is 186 and 11364.59 seconds using MBD and BD, respectively. In general, the estimations based on BD require more time than the estimations based on MBD . The execution time of the estimations, obtained when the bands are applied by subintervals, is much larger than when these are applied to the entire interval. Except for $m = 30$, the execution time of the estimators based on the Hausdorff distance is less than the time employed by MBD . When the number of profiles increases, the execution time using the Hausdorff distance is much better.

In our simulations the relative bias of the estimators based on BD is similar to the relative bias of the estimators based on MBD . For this reason and, as the execution time of the estimators that employ BD is very high, we do not present the relative bias of these estimators. Table 2.10 shows the relative bias of the estimators of $C_{pl(prf)}$ based on MBD and Hausdorff distance, when the number of profiles is $m = 30, 60, 90$ and 120 . The number of points considered in each profile is $s = 120$, under the five distributions considered. In all the investigated cases, the relative bias decreases when the number of profiles increases, and their values are very similar. A relative large set of profiles is recommended. The standard deviation indicates a high precision of the all estimators.

2.2.2 Example

We consider the example presented in the previous section about the vertical density profile (VDP). The profile of the board A6 has been removed from the dataset, because it was detected as an outlier by the functional boxplot. Therefore, this process will become in statistical control. Figure 2.9 shows the profiles of the particleboards without outliers and the estimations of the quantiles 0.99865, 0.5 and 0.00135 using the maximum operator and the median on the Hausdorff distance matrix between profiles.

TABLE 2.10. Relative bias and standard deviation (in gray) of seven estimators of $C_{pl(prf)}$ corresponding to the four-parameter logistic model, under five distributions of the error term, when the target PCI is 1 and $m = 30, 60, 90$ and 120 profiles.

Distribution	Number of Profiles (m)	MBD	SubMBD	TmMBD	$HD+$	$HD\pm$	$HD-$
Normal(0,0.1)	30	0.473	0.471	0.471	0.472	0.472	0.470
		0.037	0.034	0.033	0.038	0.036	0.033
	60	0.292	0.292	0.292	0.293	0.293	0.292
		0.024	0.022	0.022	0.024	0.024	0.022
	90	0.214	0.214	0.214	0.214	0.215	0.214
		0.021	0.021	0.020	0.020	0.021	0.021
	120	0.167	0.165	0.166	0.166	0.166	0.165
		0.022	0.021	0.021	0.021	0.022	0.021
Lognormal(0,0.3)	30	0.290	0.309	0.283	0.296	0.296	0.287
		0.025	0.023	0.019	0.026	0.027	0.019
	60	0.183	0.194	0.179	0.187	0.190	0.180
		0.016	0.014	0.013	0.016	0.017	0.014
	90	0.134	0.142	0.131	0.137	0.139	0.132
		0.014	0.013	0.012	0.015	0.015	0.012
	120	0.107	0.113	0.104	0.109	0.110	0.105
		0.009	0.010	0.009	0.010	0.011	0.008
Weibull(6,1)	30	0.484	0.470	0.484	0.480	0.479	0.481
		0.042	0.035	0.035	0.038	0.041	0.036
	60	0.296	0.291	0.299	0.294	0.294	0.299
		0.028	0.026	0.026	0.028	0.028	0.027
	90	0.214	0.209	0.215	0.213	0.212	0.214
		0.022	0.021	0.022	0.021	0.022	0.022
	120	0.167	0.164	0.168	0.167	0.166	0.167
		0.018	0.016	0.017	0.018	0.018	0.017
Gamma(1.2,5)	30	0.054	0.066	0.051	0.056	0.057	0.052
		0.007	0.006	0.004	0.006	0.006	0.005
	60	0.027	0.034	0.026	0.028	0.029	0.026
		0.003	0.003	0.003	0.003	0.004	0.003
	90	0.018	0.023	0.017	0.019	0.019	0.017
		0.002	0.002	0.002	0.002	0.003	0.002
	120	0.013	0.017	0.012	0.014	0.014	0.013
		0.002	0.002	0.001	0.002	0.002	0.001
Beta(5,5)	30	0.332	0.332	0.332	0.332	0.332	0.332
		0.019	0.019	0.019	0.019	0.019	0.019
	60	0.205	0.205	0.205	0.205	0.205	0.205
		0.012	0.012	0.012	0.012	0.012	0.012
	90	0.149	0.149	0.149	0.149	0.149	0.149
		0.011	0.011	0.011	0.011	0.011	0.011
	120	0.116	0.116	0.116	0.116	0.116	0.116
		0.008	0.008	0.008	0.008	0.008	0.008

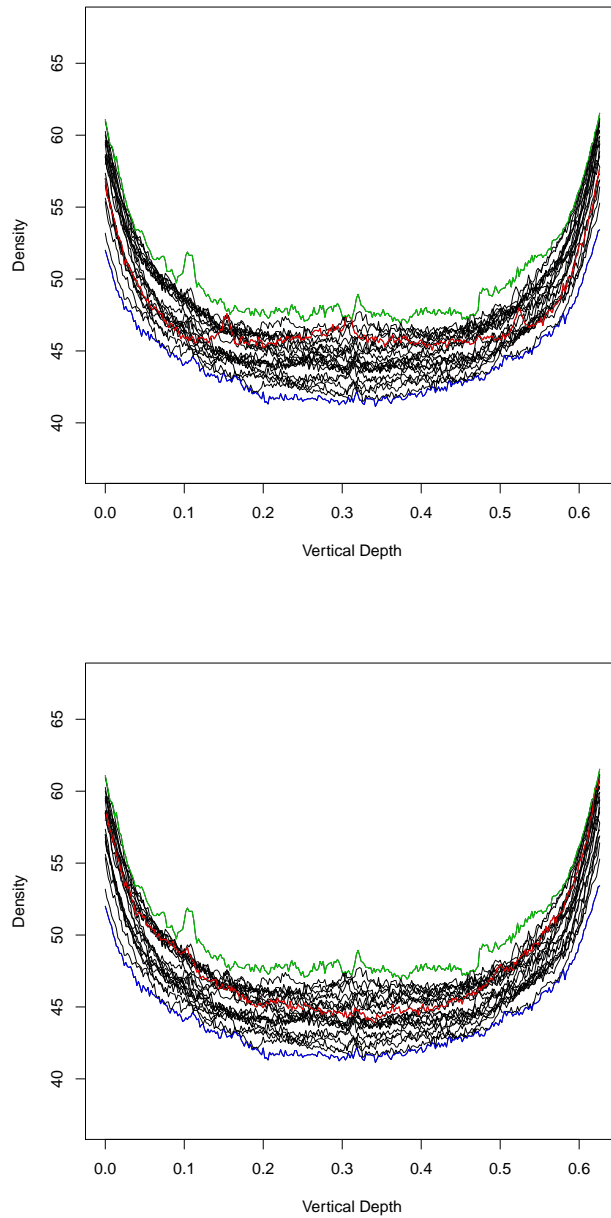


FIGURE 2.9. Vertical Density Profile of 23 particleboards (black curves), with estimations of the median function (red curves) and estimations of the quantiles 0.99865 (green curves) and 0.00135 (blue curves) using maximum operator (on the top) and median (on the bottom) on the Hausdorff distance matrix between profiles.

TABLE 2.11. Estimations of $C_{pk(prf)}$ for the VDP data.

Estimator	$\widehat{C}_{pk(prf)}$
<i>MBD</i>	1.027
<i>SubMBD</i>	1.032
<i>TmMBD</i>	1.031
<i>HD+</i>	1.033
<i>HD±</i>	1.041
<i>HD-</i>	1.026

The specification functions are established for a process of medium density class, so that $USL(t) = 50$ and $LSL(t) = 40$, [15]. Table 2.11 shows the estimations obtained for the index $C_{pk(prf)}$ employing estimators based on MBD and Hausdorff distance, which oscillate between 1.026 and 1.041, indicating that the process is capable.

2.2.3 Conclusions

In this section we propose an approach based on the Hausdorff distance to measure the process capability characterized by nonlinear profiles. This approach uses the functional version of PCIs of Clements proposed by [33], but the estimations of the 0.99865, 0.5 and 0.00135 quantiles functions are obtained using the Hausdorff distance.

The simulation study shows that although the relative bias and the precision of the estimators based on the Hausdorff distance are very similar to the obtained by the estimators based on band depth, the execution time is significantly lower, and their performance improves when the number of profiles increases.

Further developments can extend the method proposed to other PCIs.

Process Capability Analysis for Multivariate NonLinear Profiles

In this chapter we propose a method to measure the process capability characterized by multivariate nonlinear profiles based on a principal component method for multivariate functional data and the concept of functional depth. Performance of the proposed method is evaluated through simulation studies. An example from the production of sugar illustrates this method.

3.1 Evaluation of Process Capability in Multivariate NonLinear Profiles

Let $\mathbf{Y} = (Y_1, \dots, Y_p)'$ be a p dimensional stochastic process defined on a probability space $(\Omega, \mathfrak{F}, P)$. Each random function Y_j , $j = 1, \dots, p$, is defined on the compact real interval $[c, d]$, with values in \mathbb{R} , which is modeled by a nonlinear regression model given by

$$Y_j(t) = f_j(\mathbf{X}_j, \beta_j, t) + \varepsilon_j(t), \quad (3.1)$$

where f_j is a nonlinear function, \mathbf{X}_j is a vector of explanatory variables, β_j is a vector of parameters, ε_j is the random error and $t \in [c, d]$. Without loss of generality we assume that for each $t \in [c, d]$, the random vector $\mathbf{Y}(t) = (Y_1(t), \dots, Y_p(t))'$ has a mean vector $\mu(t) = \mathbf{0}$, and a positive-definite covariance matrix $\Sigma_{\mathbf{Y}(t)} = \mathbf{Y}(t)\mathbf{Y}(t)'$.

The specification functions of the process is given by the vector $\mathbf{USL} = (USL_1, \dots, USL_p)$ and $\mathbf{LSL} = (LSL_1, \dots, LSL_p)$, where USL_j and LSL_j are the upper and lower specification functions of the random function Y_j .

Let us assume that there are m random samples available from a historical data set in control. For the k th random sample collected over time, we have a vector of p functions in a set of n_k observations, which are organized in an $n_k \times p$ matrix M_k whose (i, j) entry is the random function Y_j corresponding to the i th observation. We assume that the observations within each of the function Y_j are independent.

We initially pool all the m samples into one sample of size $N = \sum_{k=1}^m n_k$, which is organized in a $N \times p$ matrix. Next, we find the p functional principal components according with the proposal of [4].

Let Z_r be the r th functional principal component defined as in equation (1.13), λ_r the r th eigenvalue function and e_r the r th eigenvector function corresponding to λ_r .

If for all $t \in [c, d]$ the random errors follow a multivariate normal distribution, the distribution of the $Z_r(t)$ is normal. The capability on the r th functional component is measured through out the following indices:

$$C_{pu;Z_r} = \frac{\int_c^d USL_{Z_r}(t) - \mu_{Z_r}(t) dt}{3 \int_c^d \sqrt{\lambda_r(t)} dt}, \quad (3.2)$$

$$C_{pl;Z_r} = \frac{\int_c^d \mu_{Z_r}(t) - LSL_{Z_r}(t) dt}{3 \int_c^d \sqrt{\lambda_r(t)} dt}, \quad (3.3)$$

$$C_{p;Z_r} = \frac{\int_c^d USL_{Z_r}(t) - LSL_{Z_r}(t) dt}{6 \int_c^d \sqrt{\lambda_r(t)} dt}, \quad (3.4)$$

$$C_{pk;Z_r} = \min \{C_{pu;Z_r}, C_{pl;Z_r}\}, \quad (3.5)$$

where $USL_{Z_r}(t) = e_r(t)'USL_{Y_r}(t)$ and $LSL_{Z_r}(t) = e_r(t)'LSL_{Y_r}(t)$ represent the upper and lower specification functions of $Z_r(t)$, respectively; and $\mu_{Z_r}(t)$ is the population mean function of the r th principal component.

Based on the fact that the components are independent functions, we propose the overall capability index $MC_{pu(prf)}$ to measure the capability of the process with only upper specification functions, which is given by:

$$MC_{pu(prf)} = \frac{1}{q} \sum_{r=1}^q C_{pu;Z_r}, \quad (3.6)$$

where q is the number of functional principal components comprising at least the 90% of the process variability. Similarly, we can define $MC_{pl(prf)}$, $MC_{p(prf)}$ and $MC_{pk(prf)}$ by replacing $C_{pu;Z_r}$ with $C_{pl;Z_r}$, $C_{p;Z_r}$ and $C_{pk;Z_r}$, respectively.

These indices assign the same importance to all the q functional principal components considered. To overcome this deficiency, we use the measure π_r , see equation (1.14), which expresses the proportion of variability explained by the r th component. So the new index, suggested for processes with only upper specification functions, is defined by:

$$MC_{pu(prf)}^\dagger = \sum_{r=1}^q \pi_r C_{pu;Z_r}. \quad (3.7)$$

The indices $MC_{pl(prf)}^\dagger$, $MC_{p(prf)}^\dagger$ and $MC_{pk(prf)}^\dagger$ can be obtained by replacing $C_{pu;Z_r}$ with $C_{pl;Z_r}$, $C_{p;Z_r}$ and $C_{pk;Z_r}$, respectively.

If for any t , the distribution of the random errors is not known or it is a multivariate nonnormal, we calculate the capability of the functional components using indices proposed by [33], expressed in equations (2.5) - (2.8). So, the capability in each functional component is calculated through the following indices:

$$C_{pu(Q);Z_r} = \frac{\int_c^d USL_{Z_r}(t) - Z_{r,0.5}(t) dt}{\int_c^d Z_{r,0.99865}(t) - Z_{r,0.5}(t) dt}, \quad (3.8)$$

$$C_{pl(Q);Z_r} = \frac{\int_c^d Z_{r,0.5}(t) - LSL_{Z_r}(t) dt}{\int_c^d Z_{r,0.5}(t) - Z_{r,0.00135}(t) dt}, \quad (3.9)$$

$$C_p(Q);Z_r = \frac{\int_c^d USL_{Z_r}(t) - LSL_{Z_r}(t) dt}{\int_c^d Z_{r,0.99865}(t) - Z_{r,0.00135}(t) dt}, \quad (3.10)$$

$$C_{pk(Q);Z_r} = \min \{C_{pu;Z_r}, C_{pl;Z_r}\}, \quad (3.11)$$

where $Z_{r,0.5}(t)$ is the population median function of the Z_r functional component, and $Z_{r,0.99865}(t)$ and $Z_{r,0.00135}(t)$ are their functional quantiles 0.99865 and 0.00135, respectively.

For processes characterized by multivariate nonlinear profiles which have only the upper specification function, the capability without distributional assumptions is calculated using $C_{pu(Q);Z_r}$ instead of $C_{pu;Z_r}$ in the equations (3.6) and (3.7). These new indices are defined by:

$$MC_{pu(Q);prf} = \frac{1}{q} \sum_{r=1}^q C_{pu(Q);Z_r}, \quad (3.12)$$

and

$$MC_{pu(Q);prf}^\dagger = \sum_{r=1}^q \pi_r C_{pu(Q);Z_r}. \quad (3.13)$$

Similar versions can be obtained replacing $C_{pu(Q)}$ by $C_{pl(Q)}$, $C_p(Q)$ or $C_{pk(Q)}$.

The weights at t are unique except for the sign. Therefore, the number of possible weighting functions, \mathbf{e} , defining each principal component is non-finite. [4] proposed a criterion for choosing the sign of the weighting functions so that the resulting components are smooth and easier to interpret. The basic idea consists selecting the sign at t which is closer on average to the signs already determined for values in a neighborhood of t . However, using this method for processes with bilateral specifications, sometimes $USL_{Z_r}(t) - LSL_{Z_r}(t) < 0$ for some $t \in [c, d]$. In processes with unilateral specifications, these could be positive for some values of t and negative for the others. This behaviour complicates the interpretation of the capability in the functional components. To overcome this drawback and ensure that the PCIs evaluated in the principal components have the same sign of their corresponding PCIs calculated on the original functions, we select the sign at t for the weighting functions $\mathbf{e}(t)$ such that $USL_{Z_r}(t) - LSL_{Z_r}(t) \geq 0, \forall t \in [c, d]$, when the processes are associated to bilateral specifications. For processes with only upper specification functions we set the criterion as: $sign(USL_{Z_r}(t) - \widehat{Z}_{r,0.5}(t)) = sign(USL_{Y_r}(t) - \widehat{Y}_{r,0.5}(t))$, $r = 1, \dots, p$. The estimations of the median can be replaced by estimations of the mean if the indices expressed by equations (3.2)-(3.5)

are employed. A similar condition for processes with only lower specification functions can be established. Using this method, we have additionally that the first components are smooth functions.

In practice, we do not know $\lambda_r(t)$ (eigenvalue function), $\mu_{Z_r}(t)$, $Z_{r,0.5}(t)$, $Z_{r,0.99865}(t)$ and $Z_{r,0.00135}(t)$. $\lambda_r(t)$ is estimated from the sample covariance matrix, $\mu_{Z_r}(t)$ is estimated using the sample mean function defined as $\bar{Z}_r(t) = \frac{1}{N} \sum_{i=1}^N z_{ir}(t)$. and the functional quantiles of $Z_r(t)$ are estimated using the concept of band depth for functional data, which is based on the bands defined by the graph of the functions on the plane, see [57]. In particular, $\hat{Z}_{r,0.5}$ is the sample median function, a curve from the sample with highest depth value, $\hat{Z}_{r,0.5} = \arg \max_{\{z \in z_{1r}, \dots, z_{Nr}\}} BD_N(z)$, where z_{1r}, \dots, z_{Nr} are the observations of the r th functional principal component, and $BD_N(z)$ is the band depth of any of these curves. $\hat{Z}_{r,0.99865}(t)$ and $\hat{Z}_{r,0.00135}(t)$ are the curves that envelope the estimated 0.9973 central region, [33]. The estimated α -central region is the band delimited by the α proportion of deepest curves from the sample, [55] and [57].

Generally the data are discretized, therefore a smoothing of the data is needed, for example a spline smoothing.

3.2 Simulations

In this section, the proposed indices to measure the capability of processes characterized by multivariate nonlinear profiles with only lower specification functions, are evaluated via simulation studies, under the assumption that the process is in statistical control. The simulations are performed using the language R ([72]).

100 Multivariate functional data sets of size $N \times p$, with $N = 100$ and $p = 2, 5, 10$ and 30 are considered. These data sets are realizations of the vector of functions $\mathbf{Y}(t) = (Y_1(t), \dots, Y_p(t))$, where

$$Y_1(t) = \sin(t) + 0.5\epsilon_1(t) \quad (3.14)$$

and

$$Y_i(t) = 3(i-1)\sin(t) + 0.5\epsilon_i(t), \quad (3.15)$$

for $t \in [0, 2\pi]$ and $i = 2, \dots, p$. The multivariate normal, multivariate t , multivariate gamma and multivariate beta distributions are considered for the random vectors $\epsilon(t)$. The observations within each curve are independent and the structure of the covariance matrix for the random vectors $\epsilon(t)$ is given by:

$$\Sigma_{\epsilon(t)} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}, t \in [0, 2\pi], \quad (3.16)$$

with $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9 . Figure 3.1 shows the simulated functions for $p = 5$ and $\rho = 0.1$ and 0.9 when the random vectors $\epsilon(t) \sim N_p(\mathbf{0}, \Sigma_{\epsilon(t)})$.

For each random function Y_r , $r = 1, \dots, p$, the target value of the $C_{pl;Y_r}$ index is set at 1 and the corresponding specification function, $LSL_{Y_r}(t)$, is found. If the distribution of Y_r is normal,

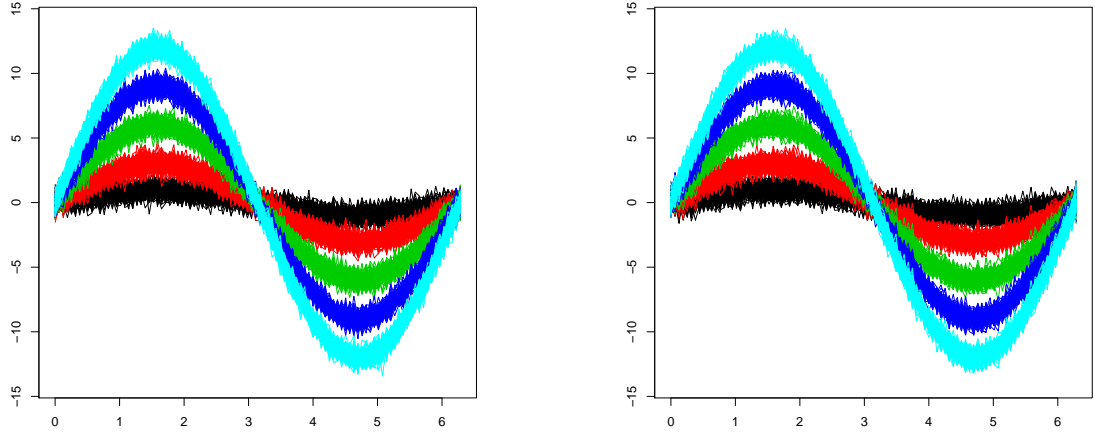


FIGURE 3.1. Data generated from equations (3.14), (3.15) and (3.16) with $p = 5$, $\rho = 0.1$ (left) and $\rho = 0.9$ (right). The black curves are realizations of Y_1 , the red of Y_2 , the green of Y_3 , the blue of Y_4 , and the magenta of Y_5 .

$LSL_{Y_r}(t)$ is calculated by solving the following equation:

$$LSL_{Y_r}(t) = \mu_{Y_r}(t) - 3\sigma_{Y_r}(t)C_{pl;Y_r}, \quad (3.17)$$

otherwise

$$LSL_{Y_r}(t) = Y_{r0.5}(t) - (Y_{r0.99865}(t) - Y_{r0.5}(t))C_{pl;Y_r}. \quad (3.18)$$

To determine the best estimator of the process capability, we compare the relative bias and the precision from the 100 replications.

Initially, we assume that the distribution of the error vectors $(\varepsilon_1(t), \dots, \varepsilon_p(t))$ is $N_p(0, \Sigma_{\varepsilon}(t))$. Figure 3.2 illustrates the relative bias of $\widehat{MC}_{pl(prf)}$ and $\widehat{MC}_{pl(prf)}^\dagger$ under different values of p and ρ . The accuracy of $\widehat{MC}_{pl(prf)}^\dagger$ improves when the correlation among the random functions increases regardless of the process dimension. For multivariate process where the correlation between the random functions is small, the accuracy of $\widehat{MC}_{pl(prf)}$ decreases when the dimension increases. The accuracy of $\widehat{MC}_{pl(prf)}$ deteriorates when the correlation among the random functions or the dimension of the process increases. Table 3.1 shows the mean and standard deviation of $\widehat{MC}_{pu(prf)}$ and $\widehat{MC}_{pu(prf)}^\dagger$ for several values of p and ρ . The standard deviation for $\widehat{MC}_{pu(prf)}$ oscillates between 0.0021 and 0.0129; while that for $\widehat{MC}_{pu(prf)}^\dagger$ this oscillates between 0.0027 and 0.0095.

When the multivariate distribution of the random vectors is not known or nonnormal, the capability of the multivariate process is estimated using the equations (3.12) and (3.13), where the $Z_{r,0.5}$ function is estimated using the concept of MBD applied on the entire interval or by subintervals, and the $Z_{r,0.99865}$ and $Z_{r,0.00135}$ functions are estimated employing the concept of α central region, with $\alpha = 0.99723$, see [57] and [33].

Figures 3.3, 3.4, 3.5 and 3.6 show the behavior of the relative bias as function of the correlation when the distribution of the random errors is $N_p(\mathbf{0}, \Sigma_{\varepsilon})$, $t_p(3)$, $gamma_p(1.9, 4)$ and $beta_p(1, 2)$, respectively, for $p = 2, 5$ and 10.

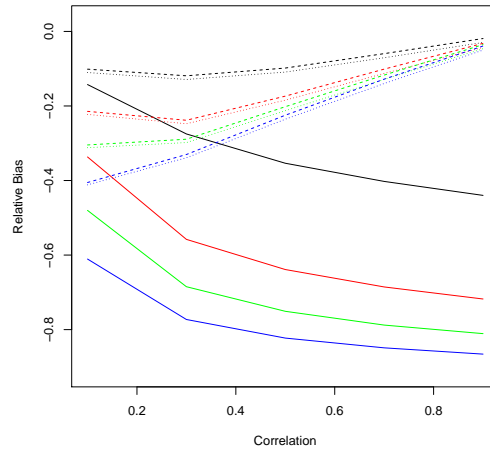


FIGURE 3.2. Relative bias of $\widehat{MC}_{pl(prf)}$ (solid lines) and $\widehat{MC}_{pl(prf)}^\dagger$ (dotted lines) for the multivariate process generated by equations (3.14), (3.15) and (3.16), with target $C_{pl} = 1$, $p = 2$ (black lines), 5 (red lines), 10 (green lines) and 30 (blue lines), and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9 .

TABLE 3.1. Mean and standard deviation (gray colour) of two estimators of C_{pu} for the process generated by the equations (3.14), (3.15) and (3.16), with target $C_{pu} = 1$, $p = 2, 5, 10$ and 30 , and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9 .

p	Estimator	ρ				
		0.1	0.3	0.5	0.7	0.9
2	$\widehat{MC}_{pu(prf)}$	0.8578	0.7239	0.6460	0.5975	0.5598
		0.0129	0.0096	0.0082	0.0066	0.0057
	$\widehat{MC}_{pu(prf)}^\dagger$	0.8901	0.8702	0.8907	0.9292	0.9712
		0.0095	0.0074	0.0067	0.0064	0.0075
5	$\widehat{MC}_{pu(prf)}$	0.6633	0.4419	0.3609	0.3146	0.2830
		0.0125	0.0075	0.0049	0.0046	0.0038
	$\widehat{MC}_{pu(prf)}^\dagger$	0.7775	0.7520	0.8151	0.8872	0.9579
		0.0089	0.0050	0.0047	0.0055	0.0064
10	$\widehat{MC}_{pu(prf)}$	0.5199	0.3150	0.2486	0.2124	0.1890
		.0090	0.0057	0.0033	0.00358	0.0031
	$\widehat{MC}_{pu(prf)}^\dagger$	0.6879	0.7015	0.7868	0.8713	0.9532
		0.0065	0.0046	0.0035	0.00545	0.0078
30	$\widehat{MC}_{pu(prf)}$	0.3894	0.2270	0.1773	0.1512	0.1343
		0.0062	0.0033	0.0025	0.0021	0.0018
	$\widehat{MC}_{pu(prf)}^\dagger$	0.5883	0.6609	0.7644	0.8606	0.9500
		0.0041	0.0027	0.0034	0.0040	0.0068

For $p = 5$ if the random errors follow a multivariate normal distribution, the best estimator is $\widehat{MC}_{pl(Q);prf}^\dagger$ when the median is estimated by MBD on the entire interval. For multivariate gamma the best accuracy occurs using this same estimator when $\rho > 0.6$, for the other values of ρ the best accuracy occurs using $\widehat{MC}_{pl(Q);prf}^\dagger$ with the median estimated by MBD on subintervals. This last estimator presents the better performance for multivariate beta distribution.

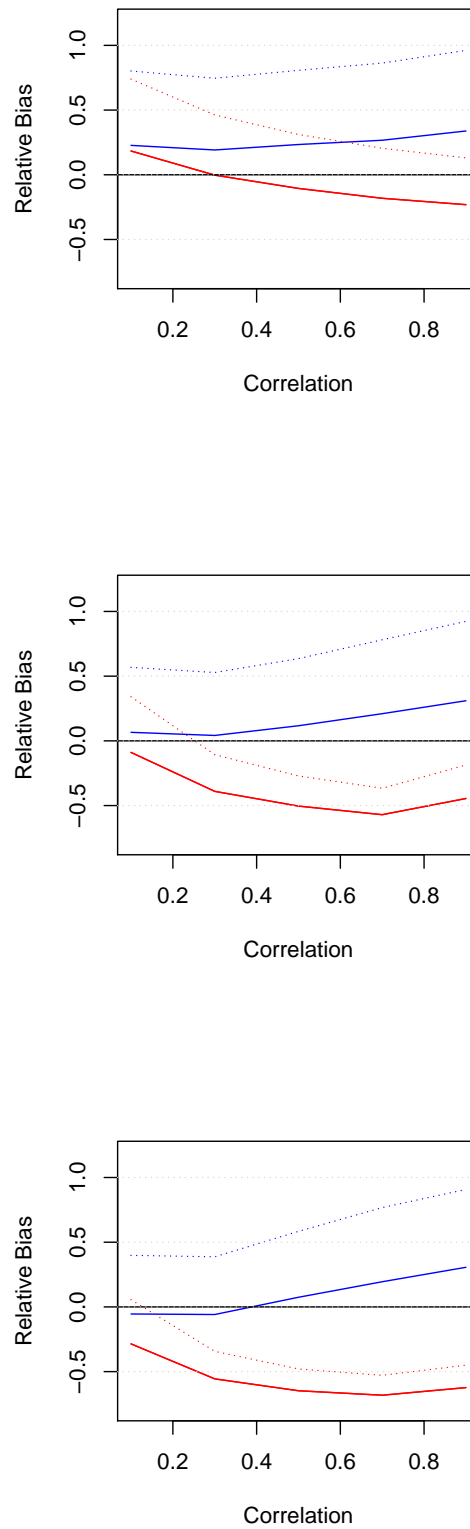


FIGURE 3.3. Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) or for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim N_p(0, \Sigma_{\varepsilon(t)})$, where $\Sigma_{\varepsilon(t)}$ is given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9

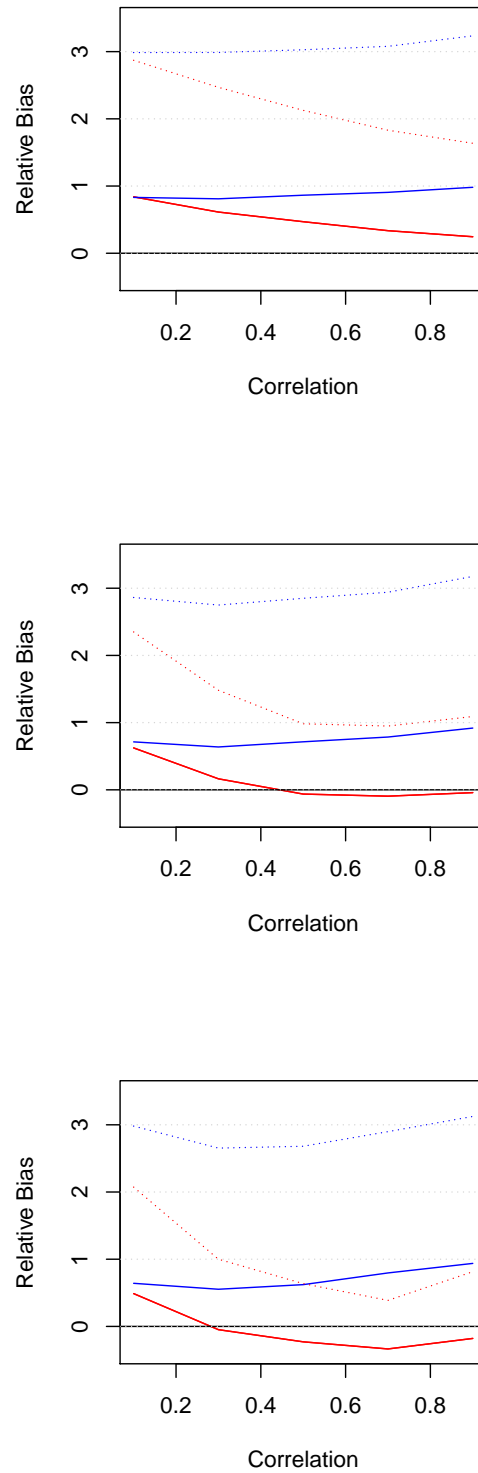


FIGURE 3.4. Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) or for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim t_p(3)$, $\Sigma_{\varepsilon(t)}$ given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9 .

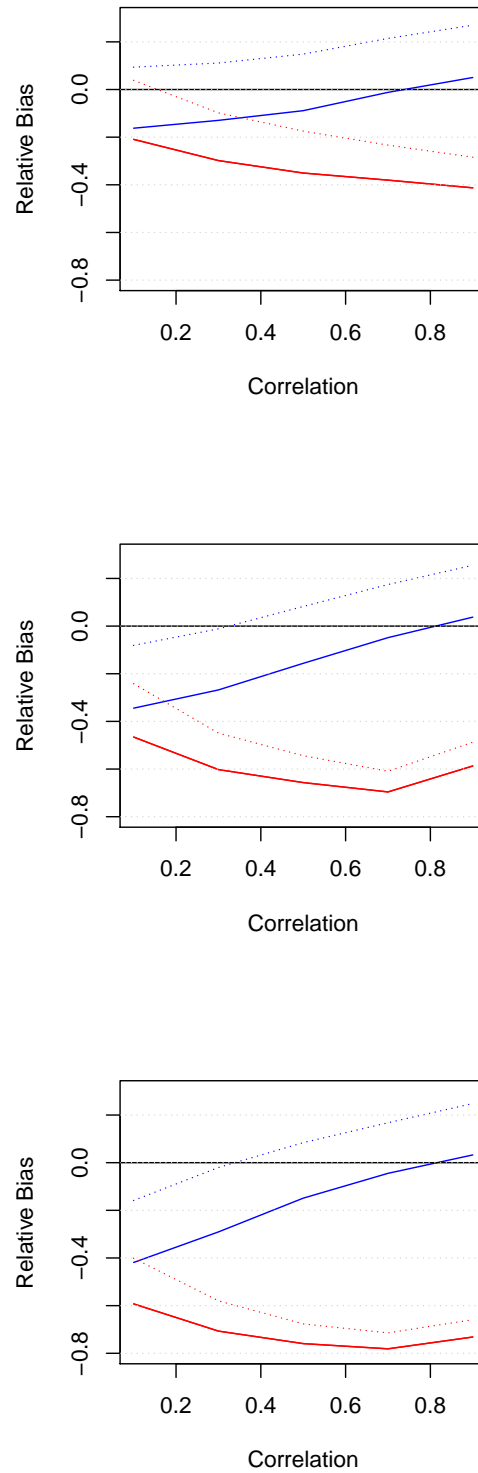


FIGURE 3.5. Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) or for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim \text{gamma}_p(1.9, 4)$, $\Sigma_{\varepsilon(t)}$ given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9 .

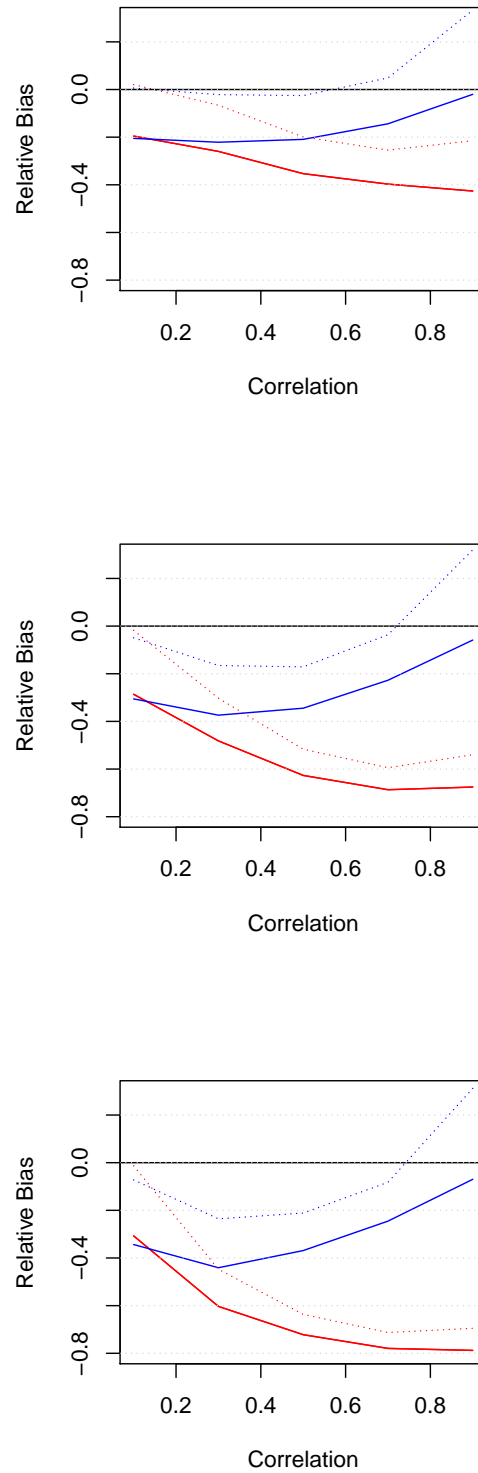


FIGURE 3.6. Relative bias of $\widehat{MC}_{pl(Q);prf}$ (in red) and $\widehat{MC}_{pl(Q);prf}^\dagger$ (in blue) associated to the multivariate process given by equations (3.14) and (3.15) with target $C_{pl} = 1$, when the medians are estimated using MBD on the entire interval (solid line) or for subintervals (dotted line), for $p = 2$ (top), $p = 5$ (middle) and $p = 10$ (bottom), with $\varepsilon \sim \text{beta}_p(1, 2)$, $\Sigma_{\varepsilon(t)}$ given by the equation (3.16) and $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9 .

In general, the estimator $\widehat{MC}_{pl(Q);prf}^\dagger$ presents better performance than the estimator $\widehat{MC}_{pl(Q);prf}$. For multivariate normal distribution this estimator presents the best accuracy when the median has been estimated by MBD on the entire interval. For multivariate gamma and beta distributions this same estimator presents the best accuracy for high values of ρ . For small or moderate values of ρ the best accuracy occurs when this estimator employs the median estimated by MBD on subintervals.

Figure 3.7 shows the first principal component for $p = 10$ with $\rho = 0.9, 0.7, 0.5$ and 0.3 , and $\varepsilon \sim \text{beta}_p(1, 2)$. Here, we can observe that these curves are steadier when ρ increases. A similar behaviour exists for the other simulated cases.

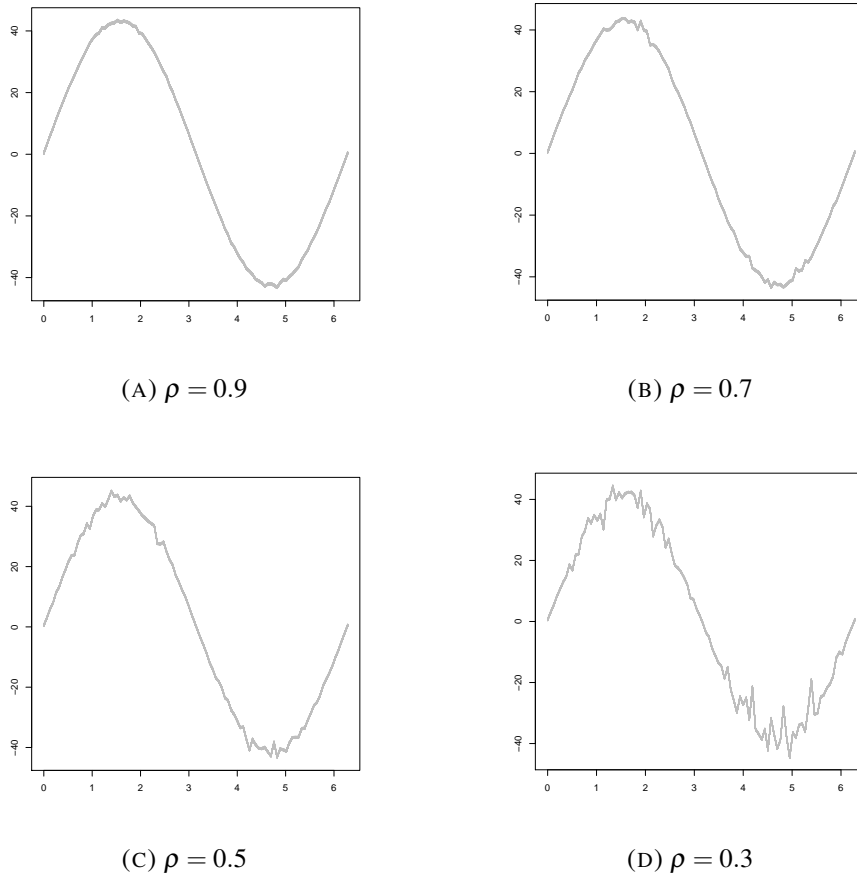


FIGURE 3.7. First principal component for $p = 10$ with $\rho = 0.9, 0.7, 0.5$ and 0.3 , and $\varepsilon \sim \text{beta}_p(1, 2)$.

3.3 Example

We considered the dataset presented by [64] and used after by [8]. This dataset contains the fluorescence information about the sugar production using fluorescence spectroscopy. Fluorescence is the property of some atoms and molecules to absorb light at a particular wavelength (excitation) and to subsequently emit light of longer wavelength (emission) after a brief interval. Fluorescence spectroscopy is often used in analytical chemistry, food analysis, environmental analysis etc. It is a very sensitive technique that can be performed nondestructively and provides qualitative and

quantitative information of diverse types of chemical analytes,[87]. The data obtained using this technique is generally presented as emission spectra, which is visualized as a plot of the fluorescence intensity versus wavelength (nanometers) for an excitation wavelength, [52]. The spectra are a series of peaks, or lines, superimposed upon noise, where each peak arises from either a characteristic absorption or a characteristic compound, [7].

In the research of [64], sugar was sampled continuously during eight hours to make a mean sample representative for one 'shift' (eight hour period). Samples were taken during the three months of operation (the so-called campaign) in late autumn from a sugar plant in Scandinavia giving a total of 268 samples. The sugar was sampled directly from the final unit operation (centrifuge) of the process. The sugar was dissolved in un-buffered water (2.25g/15mL) and the solution was measured spectrofluorometrically in a 10 by 10 mm cuvette on a PE LS50B spectrofluorometer. Raw non-smoothed data was output from the fluorometer. For every sample the emission spectra from 275-560 nm were measured in 0.5 nm intervals (571 wavelengths) at seven excitation wavelengths (230, 240, 255, 290, 305, 325, 340 nm).

We consider that the data are observations of $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7)$, a p dimensional stochastic process with $p = 7$, where Y_j indicates the emission spectra at one specific excitation wavelength. For the excitation wavelengths 230, 240, 255, 290, 305, 325, 340 nm, the emission spectra are noted Y_1, Y_2, \dots, Y_7 , respectively. These data are organized in a 268×7 matrix M , whose row i is the emission spectra of the sample i measured at emission wavelength from 275 – 560 nm, when the sample has been excited with light at wavelengths 230, 240, 255, 290, 305, 325, 340 nm. Each column contains the profiles corresponding to one excitation wavelength, (Y_j). M provides a set of 268×7 -variate curves.

Specification functions for each Y_j do not exist. For this reason, we obtain these functions from a subset of the data. Let M_1 and M_2 two submatrices extracted from M of dimensions 100×7 and 168×7 , respectively, . The specification functions of Y_j , $j = 1, \dots, 7$, are calculated from M_1 and correspond to the curves that envelope the α central region, with $\alpha = 0.90$. M_2 is employed to measure the capability of \mathbf{Y} .

The functional boxplot proposed by [92] is used to detect any outliers of M_1 and M_2 . For M_1 the profiles 10, 14, 16, 17, 38 and 71 are detected as outliers, and for M_2 13 outliers are found, corresponding to the profiles 112, 129, 130, 131, 158, 162, 164, 166, 197, 198, 199, 200 and 268. These outliers are excluded from the analysis and therefore the process will become in statistical control. Figure 3.8 shows the observations corresponding to Y_5 and Y_6 from M_2 .

Figure 3.9 shows for Y_5 and Y_6 the upper specification function, estimations of the 0.99865 and 0.5 quantile functions using MBD, and the fluorescence emission from M_2 . Replacing the 0.99865 and 0.5 quantile functions by these estimations in the equations (3.12) and (3.13), using estimations of the median based on MBD and employing $q = 1$ because the proportion of variance explained by the first component is $\pi_1 = 0.87$, the estimations of the capability of this process are $\widehat{MC}_{pu(Q);prf} = 1.08$ and $\widehat{MC}_{pu(Q);prf}^\dagger = 1.23$; therefore the process is capable.

3.4 Conclusions

In this chapter we propose a method to measure the capability of processes characterized by multivariate nonlinear profiles, based on a principal component method for multivariate functional data. In this method, each profile is considered as the realization of a p -dimensional stochastic process defined on a compact interval, where the observations are curves and are organized in a matrix.

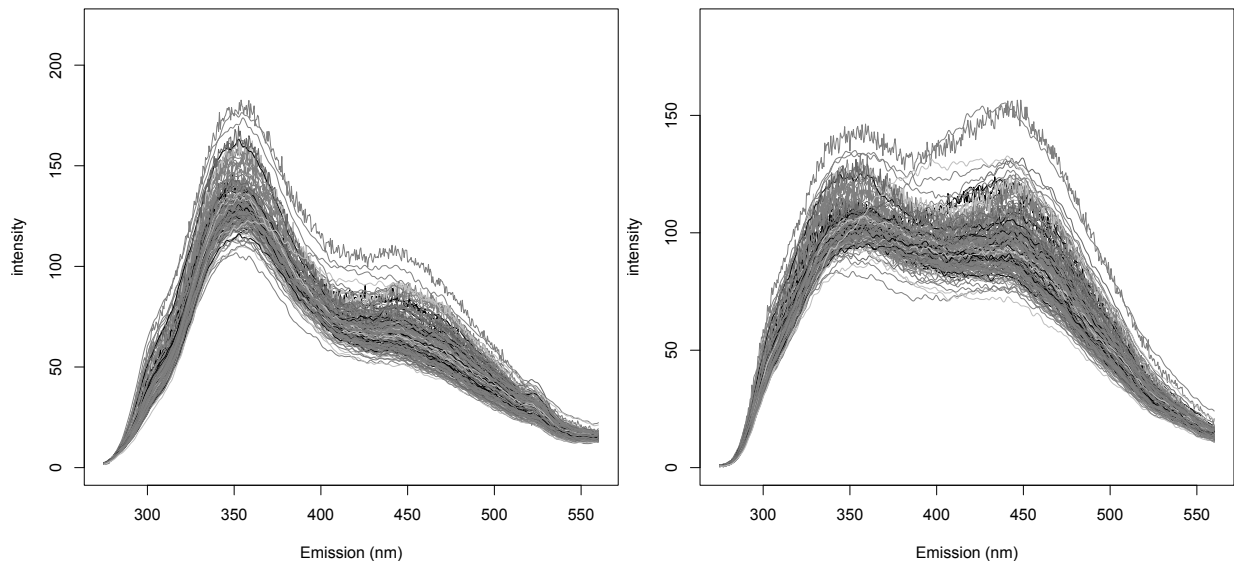


FIGURE 3.8. Raw fluorescence emission spectra sugar samples without outliers, sampled as a mean spanning eight hours equal to one shift during a three-month campaign (1995). Emission ranges were all 275 – 560 nm. The samples were measured at excitation wavelengths 305 and 325 nm, which constitute Y_5 (left) and Y_6 (right), respectively.

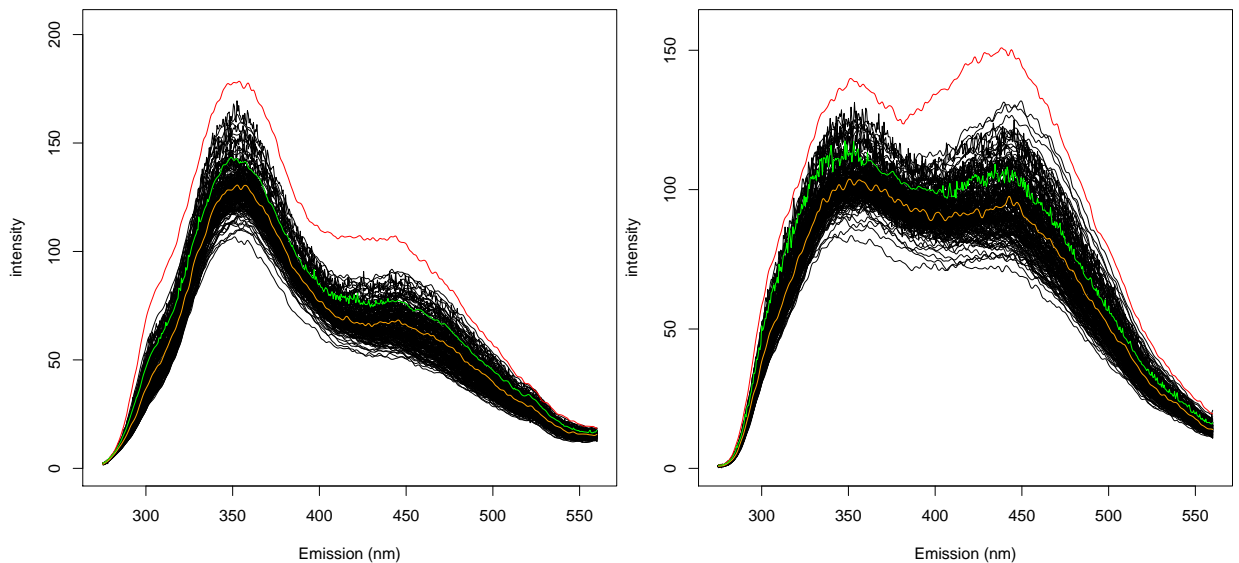


FIGURE 3.9. Upper specification function (red curve), estimation of the 0.99865 quantile function (green curve), estimation of the median function (orange curve) and fluorescence emission spectra (black curves) for Y_5 (left) and Y_6 (right), respectively.

The method initially creates uncorrelated new functions that are linear combinations of the original ones. A reduction of the dimension of the process is done on accord with the proportion of explained variance. Then, for each uncorrelated new function the capability is measured. Finally, an overall index is calculated.

Two cases are considered, depending on whether or not the distribution of the observations for each point that belongs to the compact interval is multivariate normal. For each case, two estimators have been evaluated. One is defined as the arithmetic mean of the indices associated to the components, and the other is the weighted mean for the proportion of variance explained by the components.

A simulation study considering a process with only lower specification functions was performed. If for each point of the compact interval, the distribution of the observations is multivariate normal, the accuracy of the estimator that employs the proportion of explained variance for the components improves when the correlation among the random functions increases regardless of the dimension of the process. For multivariate process where the correlation between the random functions is small, the accuracy of this estimator decreases when the dimension increases. The accuracy of the other estimator is affected by the correlation among the random functions and the dimension of the process. For multivariate non-normal distributions, in general, the estimator that employs the proportion of explained variance by each component presents better performance.

Conclusions and future work

In this dissertation we propose methodologies to evaluate the capability of processes characterized by univariate or multivariate nonlinear profiles, without distributional assumptions in the data and taking into account the within-profile correlation. To this end, we focus on functional methods.

In the first part of this dissertation we extend to functional data the PCIs proposed by Clements ([14]) that measure the univariate process capability. We assume that are working in the space of real continuous functions defined on a compact interval, where the curve of each profile is the realization of a univariate stochastic process in statistical control. With these assumptions, two methods to measure the capability of processes characterized by nonlinear profiles are proposed. In the first method, a discretized version of the process is considered and the PCIs are defined as the arithmetic mean of the one-dimensional PCIs evaluated for each point in the compact interval. In the second method, the PCIs are defined as the ratio of two integrals in view of the specification limits and the 0.99865, 0.5 and 0.00135 quantiles are continuous functions. In both methods, the proposed indices quantify the relationship between the actual performance of the nonlinear profiles and the specification limits.

The indices proposed can be estimated replacing the 0.99865, 0.5 and 0.00135 quantile functions by appropriate estimators that rely on sample observations. We use the concepts of band depth, modified band depth and Hausdorff distance to found estimations of these functions. The results of the simulations show that the estimations of the PCIs for nonlinear profiles based on these two methods present overestimation. However, the first method present greater overestimation than the second method. Both methods are sensitive to the size of the sample. The relative bias decreases when the amount of profiles considered in the sample increases. Should be ideal to use 90 or more profiles. The simulation study shows that although the relative bias and the precision of the estimators based on the Hausdorff distance are very similar to the obtained by the estimators based on band depth, the execution time is significantly lower, and their performance improves when the number of profiles increases.

In the second part of this dissertation we propose a method to measure the capability of processes characterized by multivariate nonlinear profiles, based on a principal component method for multivariate functional data. In this method, each profile is considered as the realization of a p -dimensional stochastic process defined on a compact interval, where the observations are curves and are organized in a matrix. The method initially creates uncorrelated new functions that are linear combinations of the original ones. A reduction of the dimension of the process is done on accord with the proportion of explained variance. Then, for each uncorrelated new function the capability is measured. Finally, an overall index is calculated.

Two cases are considered, depending on whether or not the distribution of the observations for each point that belongs to the compact interval is multivariate normal. For each case, two

estimators have been evaluated. One is defined as the arithmetic mean of the indices associated to the components, and the other is the weighted mean for the proportion of variance explained by the components.

A simulation study considering a process with only lower specification functions was performed. If for each point of the compact interval, the distribution of the observations is multivariate normal, the accuracy of the estimator that employs the proportion of explained variance for the components improves when the correlation among the random functions increases regardless of the dimension of the process. For multivariate process where the correlation between the random functions is small, the accuracy of this estimator decreases when the dimension increases. The accuracy of the other estimator is affected by the correlation among the random functions and the dimension of the process. For multivariate non-normal distributions, in general, the estimator that employs the proportion of explained variance by each component presents better performance.

Since the band depth, modify band depth and the Hausdorff distance incorporate the within-profile correlation, the proposed PCIs takes into account this correlation. An application is given by the vertical density profile presented in the example. An study that measures and describes the effect of the within-profile correlation could be interesting.

In some industrial applications the response variable of interest is discrete, for example proportion of conformance or number of defective items, where the specifications are given in terms of the minimum acceptable proportion or number of minimum defective items. In other applications, the response could consist of more than two categorical outcomes as in the case of failure mode study where a failed product can be classified as no damaged, slightly damaged, somewhat damaged, moderately damaged, somewhat severely damaged, and severely damaged. It should be interesting to study how to measure the capability in these processes.

In this dissertation a principal component method for multivariate functional data is used as dimensional reduction technique. Other dimension reduction methods for multivariate functional data can be used, for example multivariate functional partial least squares. It would be worthwhile to study and compare the performance of the PCIs resulted from these techniques.

Several types of bootstrap confidence intervals have been developed, for example the standard bootstrap confidence interval, the percentile bootstrap confidence interval, the biased corrected percentile bootstrap confidence interval, the biased corrected and accelerated percentile bootstrap confidence interval, etc. It should be interesting to study the behavior of several bootstrap confidence interval for estimating the process capability indices for functional data.

The band depth, modified band depth and the Hausdorff distance can be used to develop control charts for nonlinear profile monitoring. These control charts would not have distributional assumptions and would have account the within-profile correlation. It should be interesting to study how these control charts can be developed and compare their performance with the existing control charts for monitoring nonlinear profiles.

In some situations the number of profiles is small, so it is necessary to develop methods to measure the capability with an acceptable bias.

Further developments can extend the methods proposed to other PCIs. An implementation to multivariate functional data could be interesting.

Bibliography

- [1] Mohammed Z. Anis, *Basic process capability indices: An expository review*, International Statistical Review **76** (2008), no. 3, 347–367.
- [2] Yan Bing Bai, Jun Hai Yong, Chang Yuan Liu, Xiao Ming Liu, and Yu Meng, *Polyline approach for approximation hausdorff distance between planar free-form curves*, Computer-Aided Design **43** (2011), no. 6, 687–698.
- [3] V. Barnett, *The ordering of multivariate data.*, Journal of the Royal Statistical Society. Series A **139** (1976), 319–354.
- [4] J. R. Berrendero, A. Justel, and M. Svarc, *Principal components for multivariate functional data*, Computational Statistics and Data Analysis **55** (2011), no. 9, 2619–2634.
- [5] G. E. P. Box and D. R. Cox, *An analysis of transformation*, Journal Royal Statistical Society B **26** (1964), no. 2, 211–243.
- [6] R. A. Boyles, *Process capability with asymmetric tolerances*, Communications in Statistics - Simulation and Computation **23** (1994), no. 3, 615–643.
- [7] Richard G. Brereton, *Chemometrics: Data analysis for the laboratory and chemical plant*, John Wiley & Sons, Chichester, England, 2003.
- [8] Rasmus Bro, *Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis*, Chemometrics and Intelligent Laboratory Systems **46** (1999), no. 2, 133–147.
- [9] P. Castagliola, *Evaluation of non-normal process capability indices using Burr's distributions*, Quality Engineering **8** (1996), no. 4, 587–593.
- [10] P. Castagliola and J. V. Garcia Castellanos, *Capability indices dedicated to the two quality characteristics case*, Quality Technology & Quantitative Management **2** (2005), no. 2, 201–220.
- [11] Philippe Castagliola and Ariane Ferreira Porto Rosa, *Monitoring of batch processes with varying durations based on the hausdorff distance*, International Journal of Reliability and Safety Engineering **13** (2006), no. 3, 1–24.
- [12] H. F. Chen, *A multivariate process capability index over a rectangular solid tolerance zone*, Statistica Sinica **4** (1994), 749–758.

-
- [13] Eric Chicken, Joseph J Pignatiello, and James R Simpson, *Statistical process monitoring of nonlinear profiles using wavelets*, Journal of Quality Technology **41** (2009), no. 2, 198–212.
- [14] J.A. Clements, *Process capability calculations for non-normal distributions*, Quality Progress **22** (1989), no. 9, 95–97.
- [15] CPA, *Ansi particleboard standard a208.1-1999*, Composite Panel Association, 1999.
- [16] Christophe Croux and Anne Ruiz-Gazen, *High breakdown estimators for principal components: the projection-pursuit approach revisited*, Journal of Multivariate Analysis **95** (2005), no. 1, 206–226.
- [17] A. Cuevas, M. Febrero, and Fraiman R., *On the use of bootstrap for estimating functions with functional data*, Computational Statistics and Data Analysis **51** (2006), 1063–1074.
- [18] A. Cuevas, Febrero M., and R. Fraiman, *Robust estimation and classification for functional data via projection-based depth notions*, Computational Statistics **22** (2007), 481–496.
- [19] Yu Ding, Li Zeng, and Shiyu. Zhou, *Phase I analysis for monitoring nonlinear profiles in manufacturing processes*, Journal of Quality Technology **38** (2006), no. 3, 199–216.
- [20] Alex Dmitrienko, Christy Chuang-Stein, and Ralph DAgostino, *Pharmaceutical statistics using sas: A practical guide*, SAS Institute Inc, Cary, North Carolina, 2007.
- [21] M. Ebadi and A. Amiri, *Evaluation of process capability in multivariate simple linear profiles*, Scientia Iranica **19** (2012), no. 6, 1960–1968.
- [22] Mohsen Ebadi and Hamid Shahriari, *A process capability index for simple linear profile*, International Journal of Advanced Manufacturing Technology **64** (2013), no. 5-8, 857–865.
- [23] S.-K. S. Fan, Y.-J. Chang, and N. Aidara, *Nonlinear profile monitoring of reflow process data based on the sum of sine functions*, Quality and Reliability Engineering International **29** (2013), no. 5, 743–758.
- [24] M. Febrero, P. Galeano, and W. González-Manteiga, *A functional analysis of NOx levels: Location and scale estimation and outlier detection*, Computational Statistics **22** (2007), 411–427.
- [25] Manuel Febrero, Pedro Galeano, and W. González-Manteiga, *Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels*, Environmetrics **19** (2008), 331–345.
- [26] Frédéric Ferraty (ed.), *Recent advances in functional data analysis and related topics*, Physica-Verlag, Berlin Heidelberg, 2011.
- [27] Frédéric Ferraty and Yves Romain (eds.), *The oxford handbook of functional data analysis*, Oxford University Press, 2011.
- [28] Frédéric Ferraty and Philippe Vieu, *Nonparametric functional data analysis*, Springer, New York, 2006.
- [29] Flavio S. Fogliatto, *Multiresponse optimization of products with functional quality characteristics*, Quality and Reliability Engineering International **24** (2008), no. 8, 927–929.
- [30] R. Fraiman and J. Meloche, *Multivariate L-estimation.*, Sociedad de Estadística e Investigación Operativa **8** (1999).

-
- [31] R. Fraiman and G. Muniz, *Trimmed means for functional data*, *Test* **10** (2001), 419–440.
- [32] Isabel González and Ismael. Sánchez, *Capability indices and nonconforming proportion in univariate and multivariate processes*, *International Journal of Advanced Manufacturing Technology* **44** (2009), 1036–1050.
- [33] Rubén Darío Guevara and José Alberto Vargas, *Process capability analysis for nonlinear profiles using depth functions*, *Quality and Reliability Engineering International* **in Press** (2013).
- [34] Rubén Darío. Guevara and José Alberto Vargas, *Evaluation of process capability in multivariate nonlinear profiles*, *International Journal of Advanced Manufacturing Technology*. Submitted for publication (2014).
- [35] Rubén Darío Guevara, José Alberto Vargas, and Philippe Castagliola, *Evaluation of process capability in nonlinear profiles using hausdorff distance*, Submitted to *Quality Technology and Quantitative Management*. Tentative acceptance subject to minor corrections (2014).
- [36] S. Gupta, D.C. Montgomery, and W.H. Woodall, *Performance evaluation of two methods for online monitoring of linear calibration profiles*, *International Journal of Production Research* **44** (2006), no. 10, 1927–1942.
- [37] T. Hesterberg, D.S. Moore, S. Monaghan, A. Clipson, and R. Epstein, *Bootstrap methods and permutation tests: Companion chapter 18 to the practice of business statistics*, W.H. Freeman and Company, 2003.
- [38] Tim Hesterberg, *Bootstrap*, *Wiley Interdisciplinary Reviews: Computational Statistics* **3** (2011), 497–526.
- [39] Lajos Horváth and Piotr Kokoszka, *Inference for functional data with applications*, Springer, New York, 2012.
- [40] S. Hosseini-fard and B. Abbasi, *Evaluation of process capability indices of linear profiles*, *International Journal of Quality & Reliability Management* **29** (2012), no. 2, 162 – 176.
- [41] S. Z. Hosseini-fard and B. Abbasi, *Process capability analysis in non normal linear regression profiles*, *Communications in Statistics - Simulation and Computation* **41** (2012), no. 10, 1761–1784.
- [42] Susan K. Humphrey and Timothy C. Krehbiel, *Managing process capability*, *Mid-American Journal of Business* **14** (1999), no. 2, 7–12.
- [43] R. J. Hyndman, *Computing and graphing highest density regions*, *The American Statistician* **50** (1996), no. 2, 120–126.
- [44] Rob J. Hyndman and Han Lin Shang, *Rainbow plots, bagplots, and boxplots for functional data*, *Journal of Computational and Graphical Statistics* **19** (2010), no. 1, 29–45.
- [45] N.L. Johnson, *System of frequency curves generated by methods of translation*, *Biometrika* **36** (1949), no. 1-2, 149–176.
- [46] J. M. Juran, *Juran’s quality control handbook*, 3rd ed., McGraw-Hill, 1974.
- [47] Victor E. Kane, *Process capability indices*, *Journal of Quality Technology* **18** (1986), 41–52.

-
- [48] Lan Kang and Susan L. Albin, *On-line monitoring when the process yields a linear profile*, Journal of Quality Technology **32** (2000), no. 4, 418–426.
- [49] Ramezan Nemati Keshteli, Reza Baradaran Kazemzadeh, Amirhossein Amiri, and Rassoul Noorossana, *Functional process capability indices for circular profile*, Quality and Reliability Engineering International (2013), To appear.
- [50] Keunpyo Kim, Mahmoud A. Mahmoud, and William H. Woodall, *On the monitoring of linear profiles*, Journal of Quality Technology **35** (2003), no. 3, 317–328.
- [51] Andrew Kusiak, Haiyang Zheng, and Zhe. Song, *Models for monitoring wind farm power*, Renewable Energy **34** (2009), no. 3, 583–590.
- [52] Joseph R. Lakowicz, *Principles of fluorescence spectroscopy, 3rd ed*, Springer, Singapore, KYO, 2006.
- [53] R. Y. Liu, *On a notion of data depth based on random simplices.*, Annals of Statistics **18** (1990), 405–414.
- [54] R. Y. Liu, J. M. Parelius, and K. Singh, *Multivariate analysis by data depth: Descriptive statistics, graphics and inference*, Annals of Statistics **27** (1999), 783–858.
- [55] Sara López-Pintado and J. Romo, *Depth-based inference for functional data*, Computational Statistics & Data Analysis **51** (2007), no. 10, 4957–4968.
- [56] Sara López-Pintado and Juan Romo, *Depth-based inference for functional data*, Tech. report, Universidad Carlos III de Madrid, Madrid, 2006.
- [57] Sara López-Pintado and Juan Romo, *On the concept of depth for functional data*, Journal of the American Statistical Association **104** (2009), no. 486, 718–734.
- [58] Sara López-Pintado and Juan Romo, *A half-region depth for functional data*, Computational Statistics and Data Analysis **55** (2011), 1679–1695.
- [59] P. Mahalanobis, *On the generalized distance in statistics.*, National Institute of Sciences of India, vol. 12, 1936, pp. 49–55.
- [60] Mahmoud A. Mahmoud, Peter A. Parker, William H. Woodall, and Douglas M. Hawkins, *A change point method for linear profile data*, Quality and Reliability Engineering International **23** (2007), no. 2, 247–268.
- [61] Mahmoud A. Mahmoud and William H. Woodall, *Phase I analysis of linear profiles with calibration applications*, Technometrics **46** (2004.), no. 4, 380–391.
- [62] D. Min, L. Zhilin, and C. Xiaoyong, *Extended hausdorff distance for spatial objects in gis*, International Journal of Geographical Information Science **21** (2007), no. 4, 459–475.
- [63] J. M. Moguerza, A. Muñoz, and S. Psarakis, *Monitoring nonlinear profiles using support vector machines*, Lecture Notes in Computer Science **4756** (2007), 574–583.
- [64] L. Munck, L. Norgaard, S.B. Engelsen, R. Bro, and C.A. Andersson, *Chemometrics in food science—a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance*, Chemometrics and Intelligent Laboratory Systems **44** (1998), no. 1-2, 31–60.

-
- [65] Rassoul Noorossana, Abbas Saghaei, and Amirhossein Amiri, *Statistical analysis of profile monitoring*, John Wiley & Sons, Hoboken, New Jersey., 2011.
- [66] Hannu Oja, *Descriptive statistics for multivariate distributions.*, Statistics & Probability Letters **1** (1983), 327–332.
- [67] S. Pal, *Performance evaluation of a bivariate normal process*, Quality Engineering **11** (1999), no. 3, 379–386.
- [68] Jeh-Nan Pan and Chun-Yi Lee, *New capability indices for evaluating the performance of multivariate manufacturing processes*, Quality and Reliability Engineering International **26** (2010), no. 1, 3–15.
- [69] M. Perakis and E. Xekalaki, *On the implementation of the principal component analysis-based approach in measuring process capability*, Quality and Reliability Engineering International **28** (2012), no. 4, 467–480.
- [70] J. J. Pignatiello and J.S. Ramberg, *Process capability indices: Just say no.*, Transactions of ASQC 47th Annual Quality Congress, 1993, pp. 92–104.
- [71] A. M Polansky, *A smooth nonparametric approach to multivariate process capability*, Technometrics **43** (2001), no. 2, 199–211.
- [72] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013, ISBN 3-900051-07-0.
- [73] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0.
- [74] Eynat Rafalin and Diane Souvaine, *Data depth contours-a computational geometry perspective*, Tech. report, Tufts University, 2004.
- [75] J. O. Ramsay and B. W. Silverman, *Functional data analysis*, Springer, New York, 2005.
- [76] K. Rezaie, B. Ostadi, and M.R. Taghizadeh, *Applications of process capability and process performance indices.*, Journal of Applied Sciences **6** (2006), no. 5, 1186–1191.
- [77] P. Rousseeuw, I. Ruts, and J. W Tukey, *The bagplot: A bivariate boxplot*, The American Statistician **53** (1999), no. 4, 382–387.
- [78] William Rucklidge, *Efficient visual recognition using the hausdorff distance*, Springer, 1996.
- [79] Abbas Saghaei, Marzieh Mehrjoo, and Amirhossein. Amiri, *A CUSUM-based method for monitoring simple linear profiles*, International Journal of Advanced Manufacturing Technology **45** (2009), 1252–1260.
- [80] Michele Scagliarini, *Multivariate process capability using principal component analysis in the presence of measurement errors*, Advances in Statistical Analysis **95** (2011), no. 2, 113–128.
- [81] Oliver Schabenberger and Francis J. Pierce, *Contemporary statistical models for the plant and soil sciences*, CRC Press, Boca Raton, FL, 2002.
- [82] R. Serfling, *Quantile functions for multivariate analysis: Approaches and applications*, Statistics Neerlandica **56** (2002), no. 2, 214–232.

-
- [83] ———, *Nonparametric multivariate descriptive measures based on spatial quantiles*, Journal of Statistical Planning and Inference **123** (2004), 259–278.
- [84] H Shahriari, N. Hubele, and F. Lawrence, *A multivariate process capability vector*, 4th Industrial Engineering Research Conference (Nashville), May 1995, pp. 303–308.
- [85] Hamid Shahriari and Mohammadreza Abdollahzadeh, *A new multivariate process capability vector*, Quality Engineering **21** (2009), no. 3, 290–299.
- [86] K Singh, *A notion of majority depth*, 1991.
- [87] Age Smilde, Rasmus Bro, and Paul Geladi, *Multi-way analysis with applications in the chemical sciences*, John Wiley & Sons, Chichester, England, 2004.
- [88] S. Somerville and D. Montgomery, *Process capability indices and non-normal distributions*, Quality Engineering **19** (1996), no. 2, 305–316.
- [89] F.A. Spiring, *Process capability: A total quality management tool*, Total Quality Management **6** (1995), no. 1, 21–33.
- [90] Zachary G. Stoumbos, *Process capability indices: Overview and extensions*, Nonlinear Analysis: Real World Applications **3** (2002), 191–210.
- [91] F.S. Stover and R.V. Brill, *Statistical quality control applied to ion chromatography calibrations*, Journal of Chromatography **804** (1998), no. 1-2, 37–43.
- [92] Ying Sun and Marc G. Genton, *Functional boxplots*, Journal of Computational and Graphical Statistics **20** (2011), no. 2, 316–334.
- [93] W. Taam, P. Subbaiah, and J. Liddy, *A note on multivariate capability indices*, Journal of Applied Statistics **20** (1993), no. 3, 339–351.
- [94] J. W. Tukey, *Mathematics and the picturing of data*, International Congress of Mathematicians, vol. 2, Canadian Mathematical Congress, 1975, pp. 523–531.
- [95] A. Vaghefi, Sam D. Tajbakhsh, and R. Noorossana, *Phase II monitoring of nonlinear profiles*, Communications in Statistics - Theory and Methods **38** (2009), no. 11, 1834–1851.
- [96] Yehuda Vardi and Cun-Hui Zhang, *The multivariate l_1 -median and associated data depth*, Proceedings of the National Academy of Sciences of the United States of America **97** (2000), no. 4, 1423–1426.
- [97] E. Walker and S. Wright, *Comparing curves using additive models*, Journal of Quality Technology **34** (2002), 118–129.
- [98] Chung Ho Wang, *Constructing multivariate process capability indices for short-run production*, International Journal of Advanced Manufacturing Technology **26** (2005), no. 11-12, 1306–1311.
- [99] F. K. Wang and James. C. Chen, *Capability index using principal component analysis*, Quality Engineering **11** (1998), no. 1, 21–27.
- [100] F. K. Wang and T. C. T. Du, *Using principal component analysis in process performance for multivariate data*, Omega **28** (2000), no. 2, 185–194.

-
- [101] F. K. Wang, N. F. Hubele, F. P. Lawrence, J. D. Miskulin, and H. Shahriari, *Comparison of three multivariate process capability indices*, Journal of Quality Technology **32** (2000), no. 3, 263–275.
- [102] Fu Kwun Wang, *Measuring the process yield for simple linear profiles with one-sided specification*, Quality and Reliability Engineering International (2013), To appear.
- [103] Fu Kwun Wang and Yi Cyuan Guo, *Measuring process yield for nonlinear profiles*, Quality and Reliability Engineering International (2013), To appear.
- [104] Jin Wang and R. Serfling, *Data depth: Robust multivariate analysis, computational geometry, and applications*, vol. 72, ch. On Scale Curves for Nonparametric Description of Dispersion, pp. 37–48, American Mathematical Society, Providence, Rhode Island, USA, 2006.
- [105] Kaibo Wang and Fugee. Tsung, *Using profile monitoring techniques for a data-rich environment with huge sample size*, Quality and Reliability Engineering International **21** (2005), no. 7, 677–688.
- [106] Siebrand J. Wierda, *A multivariate process capability index*, Proceedings of the Annual Quality Congress of the American Society for Quality **47** (1993), no. 0, 62–67.
- [107] James D. Williams, William H. Woodall, and Jeffrey B. Birch, *Statistical monitoring of nonlinear product and process quality profiles*, Quality and Reliability Engineering International **23** (2007), no. 8, 925–941.
- [108] W.H. Woodall, *Controversies and contradictions in statistical process control*, Journal of Quality Technology **32** (2000), 341–378.
- [109] William H. Woodall, *Current research in profile monitoring*, Producao **17** (2007), no. 3, 420–425.
- [110] William H. Woodall, Dan J. Spitzner, Douglas C. Montgomery, and Shilpa Gupta, *Using control charts to monitor process and product quality profiles*, Journal of Quality Technology **36** (2004), no. 3, 309–320.
- [111] T.M. Young, P.M. Winistorfer, and S. Wang, *Multivariate control charts of mdf and osb vertical density profile attributes*, Forest Products Journal **49** (1999), no. 5, 79–86.
- [112] Jiujun Zhang, Zhonghua Li, and Zhaojun. Wang, *Control chart based on likelihood ratio for monitoring linear profiles*, Computational Statistics and Data Analysis **53** (2009), 1440–1448.
- [113] Changliang Zou, Fugee Tsung, and Zhaojun. Wang, *Monitoring general linear profiles using multivariate exponentially weighted moving average schemes*, Technometrics **49** (2007), no. 4, 395–408.
- [114] Changliang Zou, Yujuan Zhang, and Zhaojun. Wang, *A control chart based on a change-point model for monitoring linear profiles*, IIE Transactions **38** (2006), no. 12, 1093–1103.
- [115] Changliang Zou, Chunguang Zhou, Zhaojun Wang, and Fugee. Tsung, *A self-starting control chart for linear profiles*, Journal of Quality Technology **39** (2007), no. 4, 364–375.
- [116] Y. Zuo and R. Serfling, *General notions of statistical depth function*, Annals of Statistics **28** (2000), 461–482.

-
- [117] Yijun Zuo, *Projection-based depth functions and associated medians*, *Annals of Statistics* **31** (2003), no. 5, 1460–1490.
- [118] Yijun Zuo and Xuming He, *On the limiting distributions of multivariate depth-based rank sum statistics and related tests*, *Annals of Statistics* **34** (2006), no. 6, 2879–2896.
- [119] Yijun Zuo and R. Serfling, *Structural properties and convergence results for contours of sample statistical depth functions*, *Annals of Statistics* **28** (2000), no. 2, 483–499.