Statistical methods of data science

An introduction to Functional Data Analysis

Alessandra Menafoglio[1*]

[1]MOX, Department of Mathematics, Politecnico di Milano

*alessandra.menafoglio@polimi.it

5. Anomaly detection through control chart schemes

Agenda

Agenda

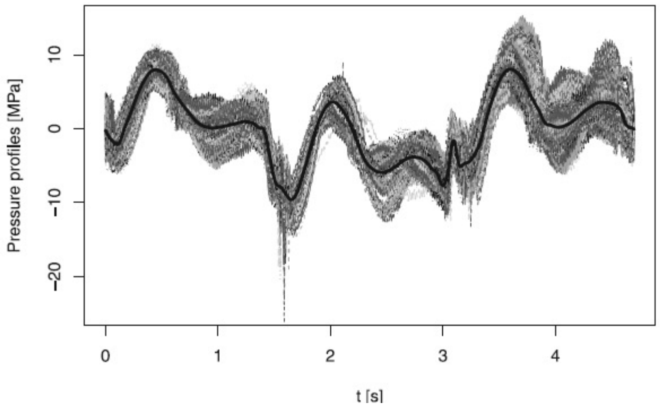
5. Functional control charts

- 5.1. Introduction & motivations: profile monitoring
- 5.2. Control charts for probability density functions
- 5.3. Anomaly detection in PDFs from image data
- 5.3. An application in Additive Manufacturing

Introduction to profile monitoring

- Monitoring the quality of a part or the stability of a process often relies on data that are functions

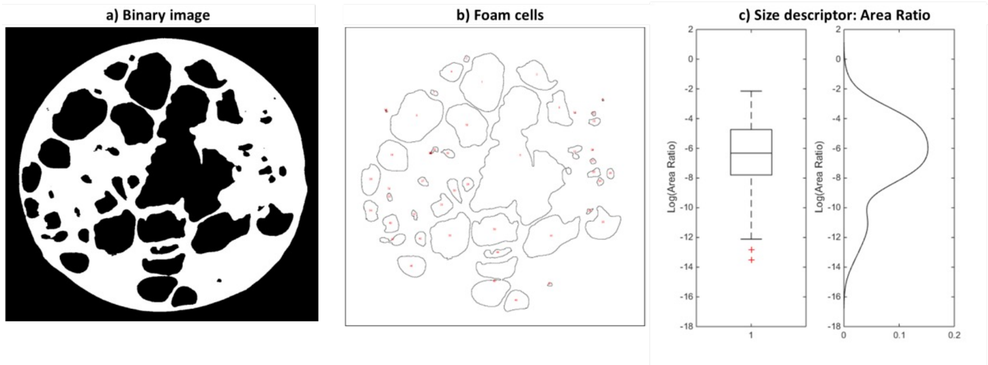
Waterjet cutting



Dynamic Pressure Profiles Under In-Control Conditions

From Grasso et al. (JQT, 2016)

Production of metal foams



Representation of the distribution of a quality indicator

From Menafoglio et al. (Technometrics, 2018)

Introduction to profile monitoring

- Monitoring the quality of a part or the stability of a process often relies on data that are functions
- When the observations are functional, typical multivariate methods for statistical process control (SPC) cannot be applied (curse of dimensionality)
- Methods for the SPC of functional data are typically called methods of **profile monitoring**

Introduction to profile monitoring

- Monitoring the quality of a part or the stability of a process often relies on data that are functions
- When the observations are functional, typical multivariate methods for statistical process control (SPC) cannot be applied (curse of dimensionality)
- Methods for the SPC of functional data are typically called methods of **profile monitoring**

- Most methods of profile monitoring assume that the data can be embedded in L^2
- When functional indicators represent aggregation of local indicators, PDF data arise
- SPC methods for data in general Hilbert spaces (particularly B^2) have been recently proposed

- In this lecture: **profile monitoring of PDF data**

Introduction to profile monitoring

- Monitoring the quality of a part or the stability of a process often relies on data that are functions
- When the observations are functional, typical multivariate methods for statistical process control (SPC) cannot be applied (curse of dimensionality)
- Methods for the SPC of functional data are typically called methods of **profile monitoring**

- Most methods of profile monitoring assume that the data can be embedded in L^2
- When functional indicators represent aggregation of local indicators, PDF data arise
- SPC methods for data in general Hilbert spaces (particularly B^2) have been recently proposed

- In this lecture: **profile monitoring of PDF data**

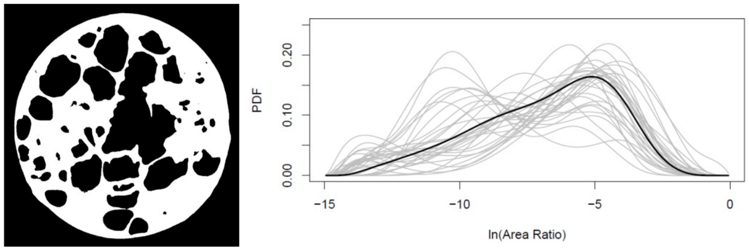
Other (more exploratory) methods for anomaly detection in density datasets (not covered here): methods based on depth measures
A. Menafoglio, L. Guadagnini, A. Guadagnini, P. Secchi (2021) "Object oriented spatial analysis of natural concentration levels of chemical species in large-scale groundwater bodies". Spatial Statistics, 43, 100494

Introduction to profile monitoring

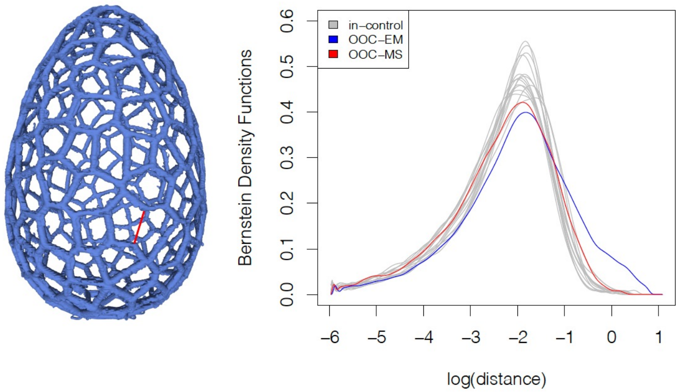
- Profile monitoring techniques have been based so far on (e.g., Colosimo, B.M., Pacella, M., 2007, 2010)
    1. Dimensionality reduction (e.g., via FPCA)
    2. Monitoring of the projection over a functional basis
    3. Monitoring of the residuals of the approximation

- This approach can be used in any Hilbert space

- We will consider this approach in the space B^2



Distribution of a quality indicator (Area Ratio) for a section of metal foam

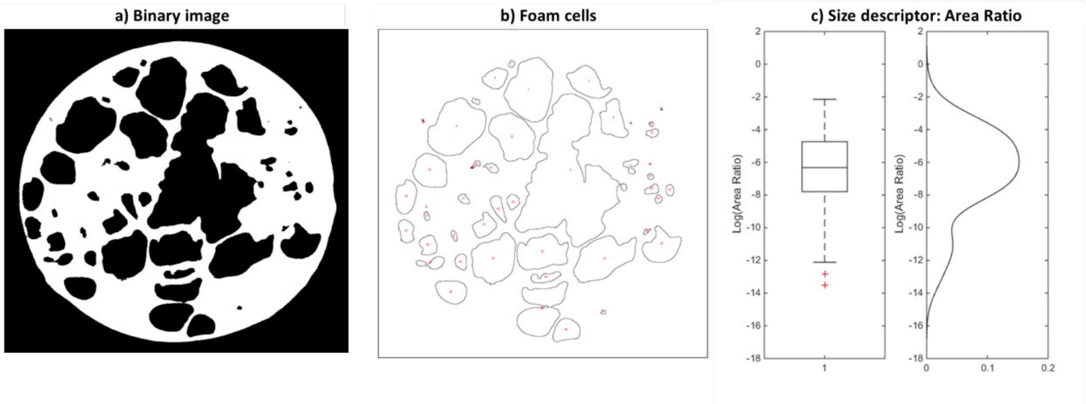


Distribution of point-to-point distances between the produced piece and the nominal object

Agenda

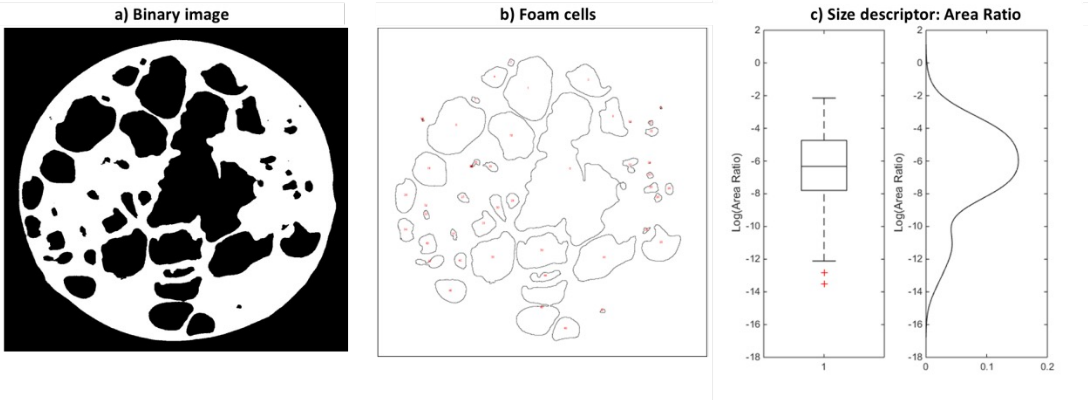
Quality control of foamed materials production

- **Metal foams**: porous metals with a cellular structure characterized by interesting combinations of physical and mechanical properties (e.g., high stiffness at low specific weight)
- **Key problem: monitoring the quality** of a part through **descriptors** (e.g., size or shape of random features), that can be summarized by
  - **few statistical moments (or QQ plot) → information loss**



Quality control of foamed materials production

- **Metal foams**: porous metals with a cellular structure characterized by interesting combinations of physical and mechanical properties (e.g., high stiffness at low specific weight)
- **Key problem: monitoring the quality** of a part through **descriptors** (e.g., size or shape of random features), that can be summarized by
  - **few statistical moments (or QQ plot) → information loss**
  - the **whole PDF** of the feature descriptors



Idea: the **shape of the PDF** can be used as a quality signature to determine both the quality of the part and the stability of the process

Statistical process control for PDFs

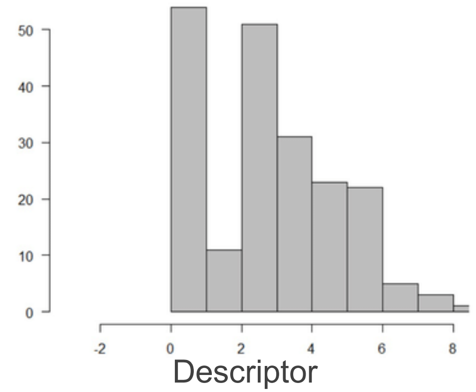
Available methods for statistical process control of distributional data still have limitations

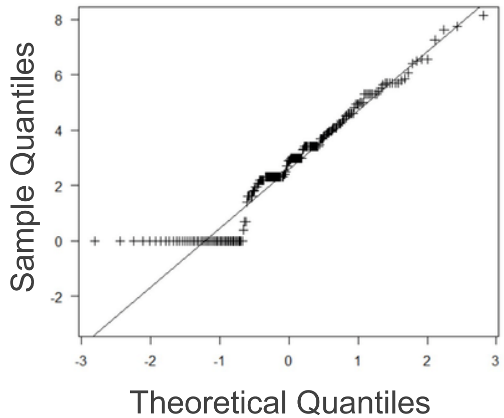
Monitoring of QQ plots

(Wang and Tsung, 2005; Wells et al., 2013)

1. Build QQ plot of descriptors
2. Monitoring QQ line (intercept and slope)

- **Pros**: accounts for data characteristics (distributional data)
- **Cons**: information loss





Profile monitoring

Idea: the data are functions, i.e., have infinite-dimensions

Profile monitoring: control charts for functional data

Statistical process control for PDFs

Profile monitoring

Idea: the data are functions, i.e., have infinite-dimensions

Profile monitoring: control charts for functional data

Profile monitoring in L^2 is not a good idea

The geometry of L^2 becomes meaningless with PDFs.

→ poor monitoring performances with PDFs

Better idea: profile monitoring by using a geometry appropriate for PDFs (Bayes space)

Bayes spaces for PDF data

Bayes space geometry

(Egozcue et al., 2006; van den Boogaart et al., 2014)

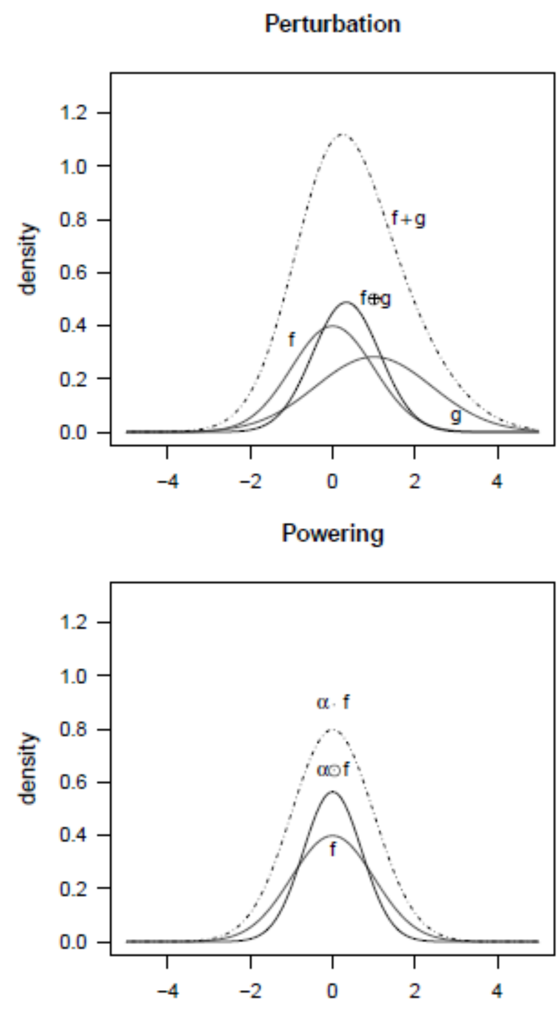
$$(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s)\,ds},$$

$$(\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \quad t \in I.$$

$$\langle f, g\rangle_{\mathcal{B}} = \frac{1}{2\eta}\int_I \int_I \ln\frac{f(t)}{f(s)} \ln\frac{g(t)}{g(s)}\,dt\,ds$$

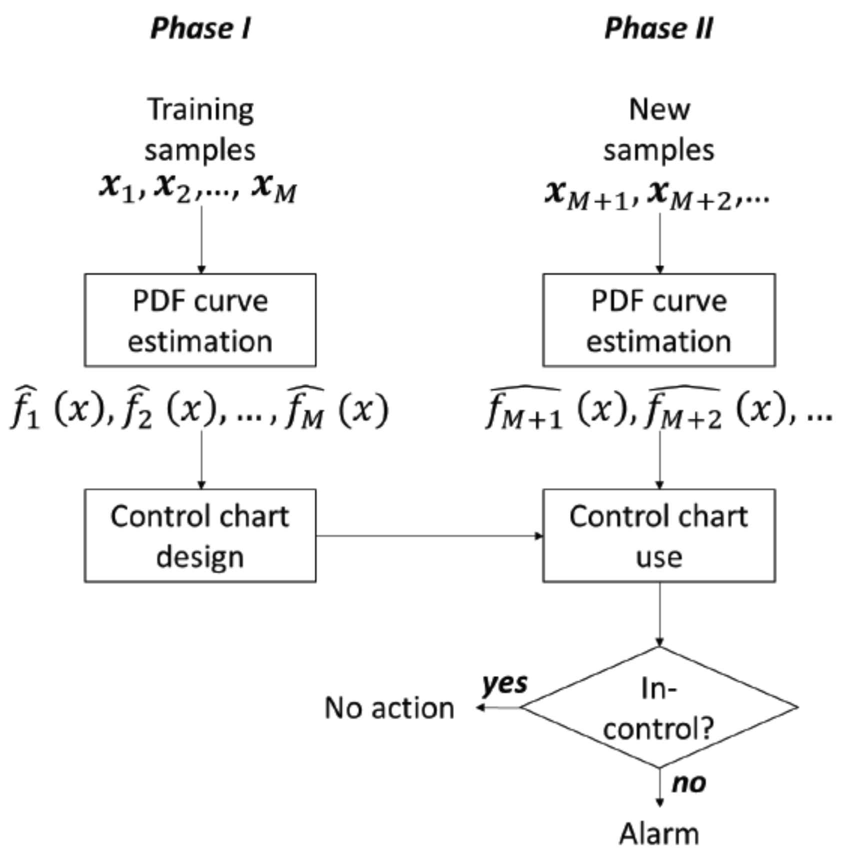
Only (log-)ratios between probabilities are meaningful (~ odds-ratio) as data represent the distribution of a *total* mass (=1) over a domain

Strategy: embedd the data in a Bayes space and here build a control chart

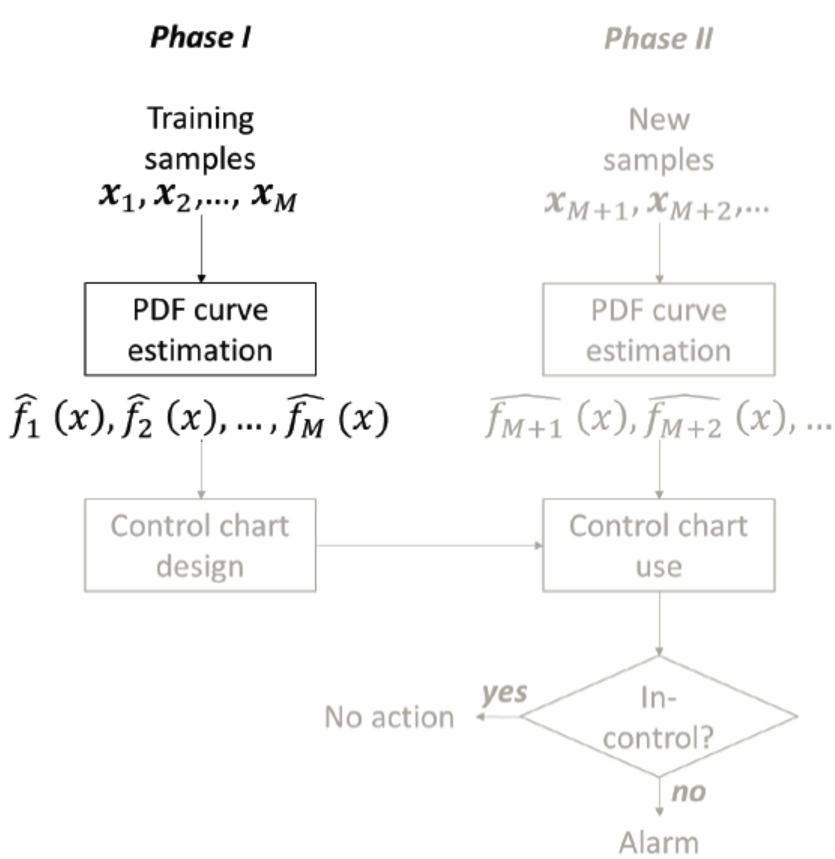


Hron et al (2016)

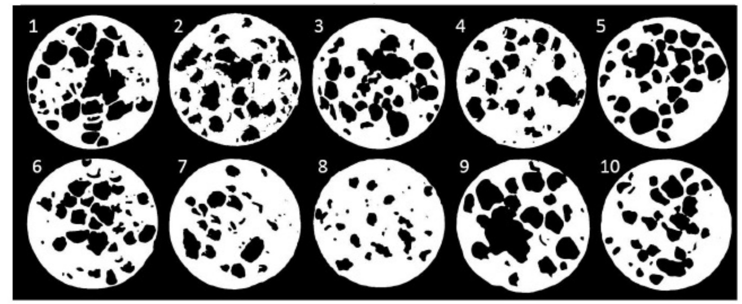
The problem of monitoring PDFs in Bayes spaces



The problem of monitoring PDFs in Bayes spaces



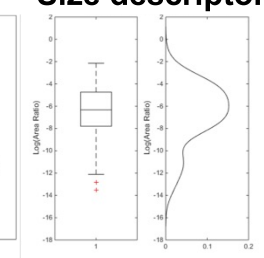
Training set of sections (Phase I)



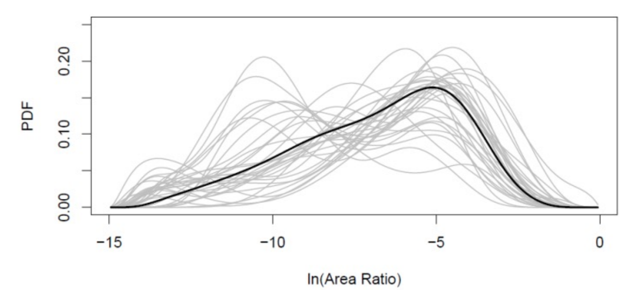
Binary Image

Foam cells

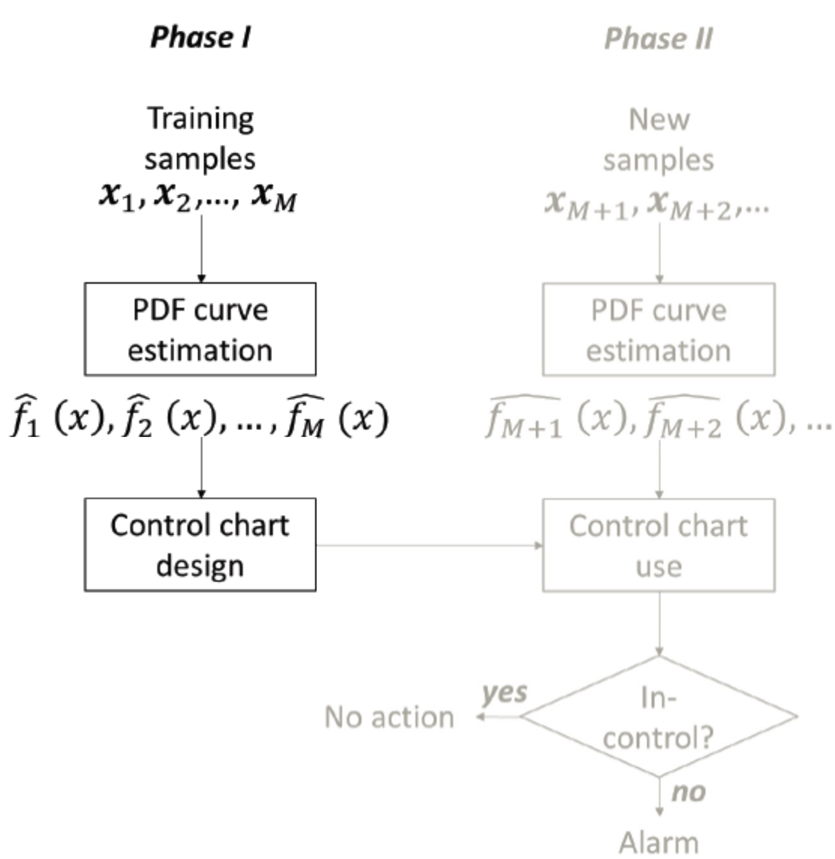
Size descriptor



PDFs of size descriptors in all the sections



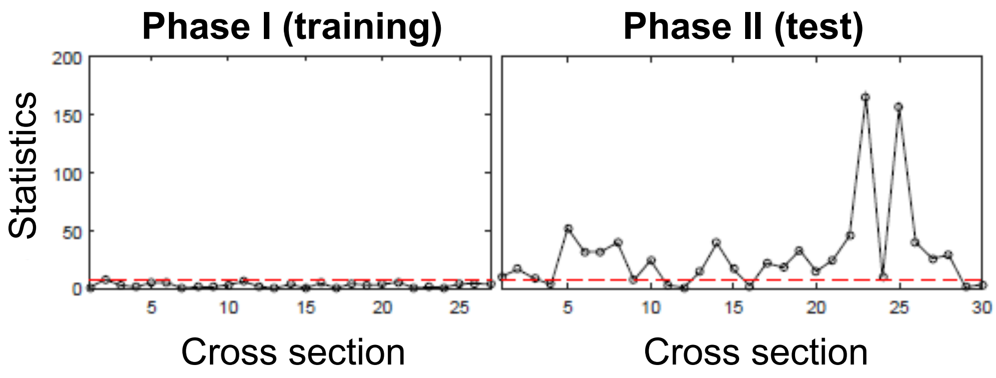
The problem of monitoring PDFs in Bayes spaces



Problem of building a control chart for PDF data in Bayes spaces

Profile monitoring of PDFs via Dimensionality Reduction in Bayes spaces

- **Problem**: building a **control chart in a Bayes space**
  - Build a statistics (functional constrained data)
  - Find a control limit (e.g., by empirical quantiles of the statistic on the training set)



Profile monitoring of PDFs via Dimensionality Reduction in Bayes spaces

- **Problem**: building a **control chart in a Bayes space**
  - Build a statistics (functional constrained data)
  - Find a control limit (e.g., by empirical quantiles of the statistic on the training set)

- **Strategy**:
  - **Reduce the dimensionality** of the problem in the Bayes space (PCA in Bayes spaces, Hron et al, 2016)
  - Build **multivariate control charts with probabilistic limit,** based on the dataset of reduced dimensionality

Note. This strategy extends to Bayes spaces the approach of Colosimo and Pacella (2007, 2010)

Dimensionality reduction in Bayes Spaces: SFPCA

- **Problem:** Given a dataset of *n* smoothed PDFs, $\widehat{f}_1, \ldots, \widehat{f}_M$ find the directions of maximum variability of the dataset, i.e., those solving

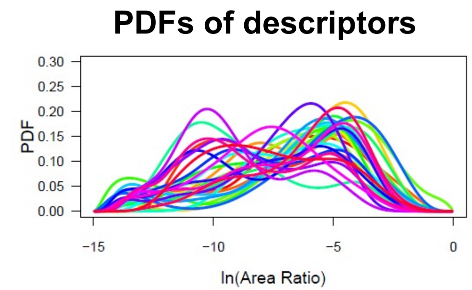
$$\sum_{j=1}^{M} \langle \widehat{f}_j \ominus \overline{f},\ \zeta \rangle^2 \quad \text{subject to} \quad \|\zeta\| = 1,\ \langle \zeta, \zeta_i \rangle = 0, i < j.$$

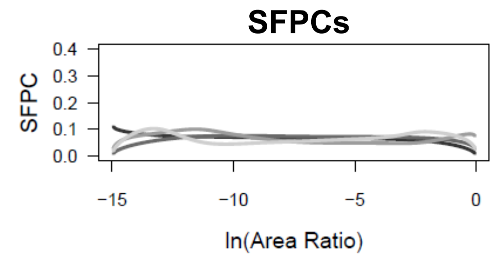
  Note. *Same problem as usual PCA, but in a different space*

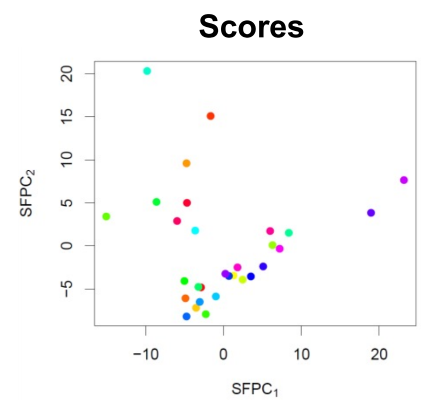
- **Scores:** Projection of the data along the SFPCs

$$z_{ji} = \langle \hat{f}_j \ominus \overline{f},\ \zeta_i \rangle$$

- **Dimensionality reduction**: keep the first *K* SFPCs that allows explaining a given amount of the variability (e.g., 95%, 98%)







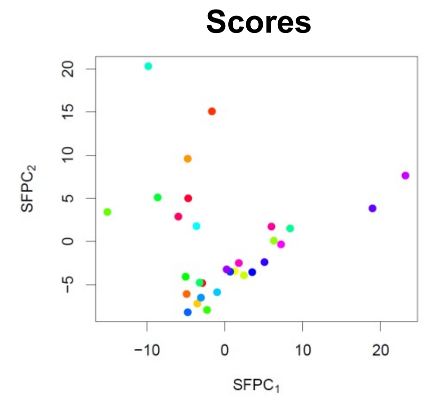
Hron, Menafoglio, Templ, Hruzová, Filzmoser (2016)

Control charts in Bayes spaces via SFPCA

Hotelling T² control chart on the scores

$$T_j^2(K) = \sum_{i=1}^{K} \frac{z_{ji}^2}{\rho_i}, \qquad j = 1, 2, \ldots,$$

to control deviations along the first *K* SFPCs

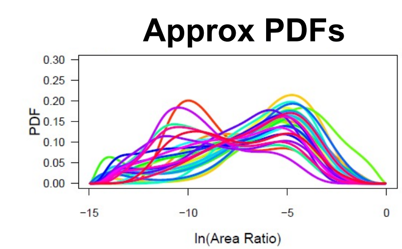


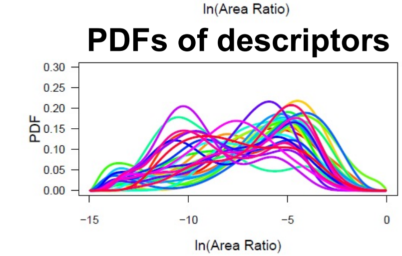
Residuals' control chart

$$SPE_j(K) = \| \hat{f}_j^* \ominus \hat{f}_j \|^2$$

$$j = 1, 2, \ldots,$$

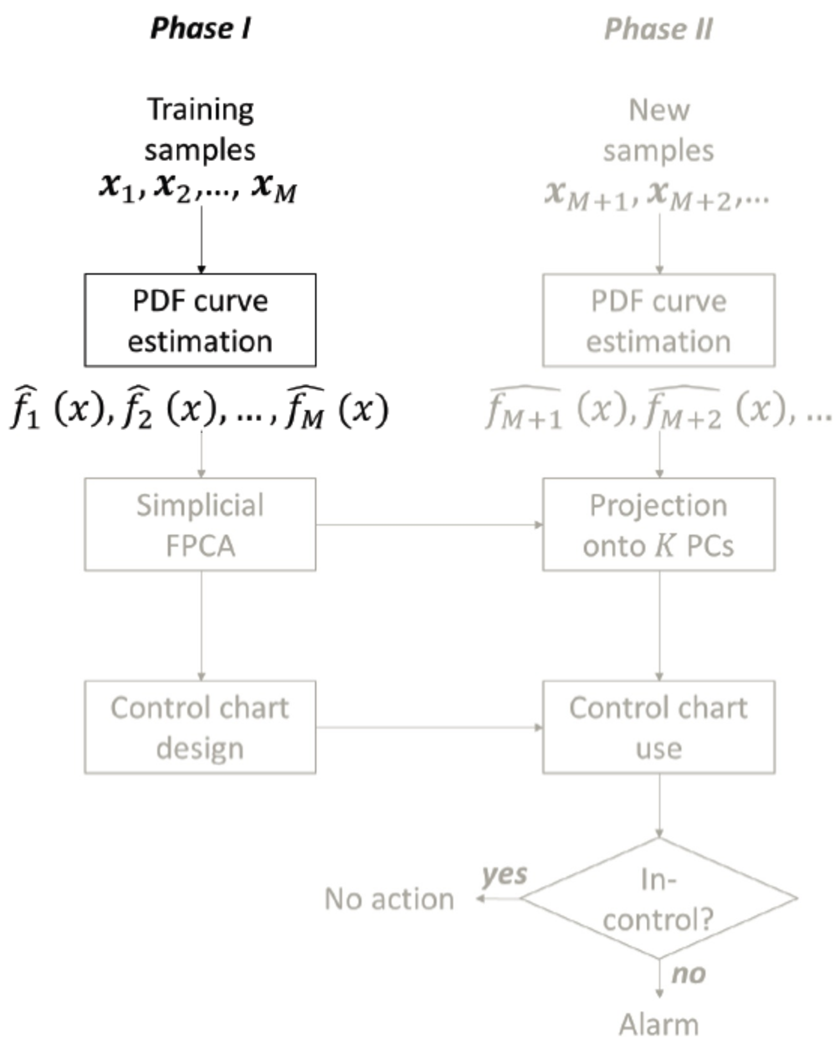
to detect shifts along directions orthogonal to the ones associated with the first *K* SFPCs



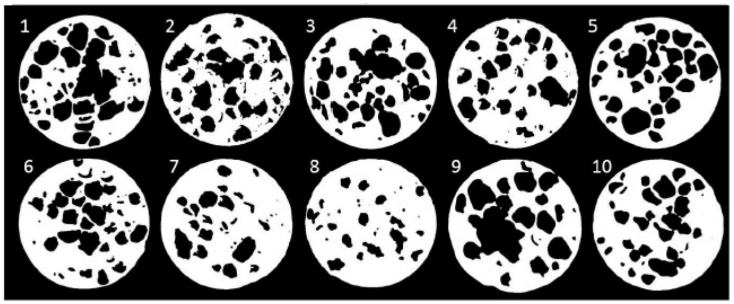


Control limits are found as empirical quantiles of the statistics on the training set

Profile monitoring for PDFs

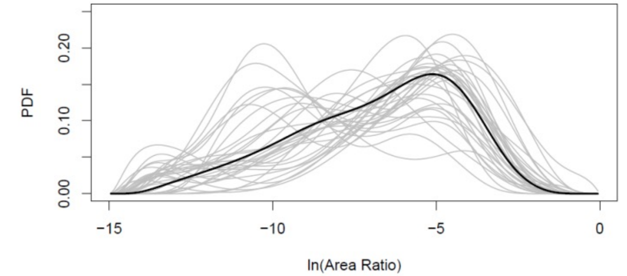


Training set of sections (Phase I)

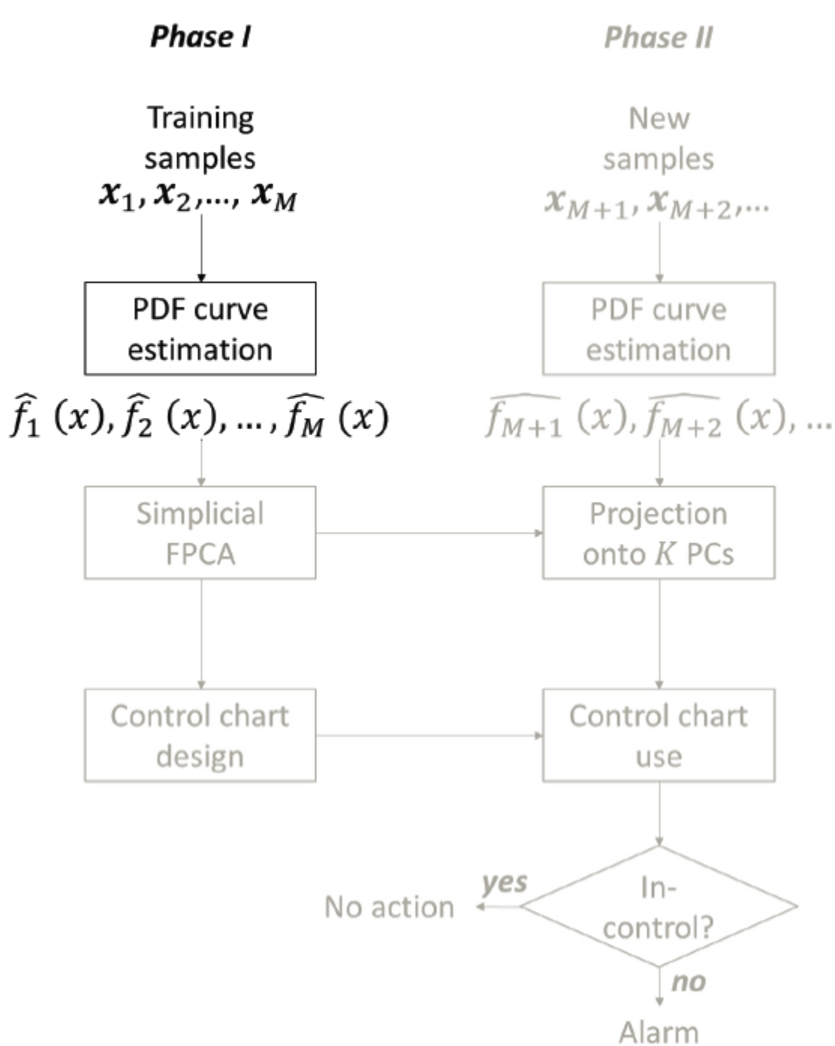


Binary Image **Foam cells** **Size descriptor**

PDFs of size descriptors in all the sections



Profile monitoring for PDFs



Bayes space geometry

(Egozcue et al., 2006; van den Boogaart et al., 2014)

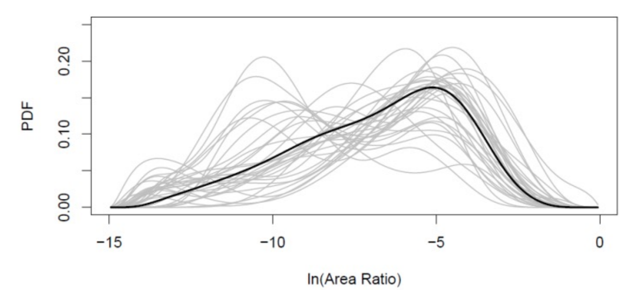
$$(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s)\, ds}$$

$$(\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_I f(s)^\alpha\, ds}, \quad t \in I.$$

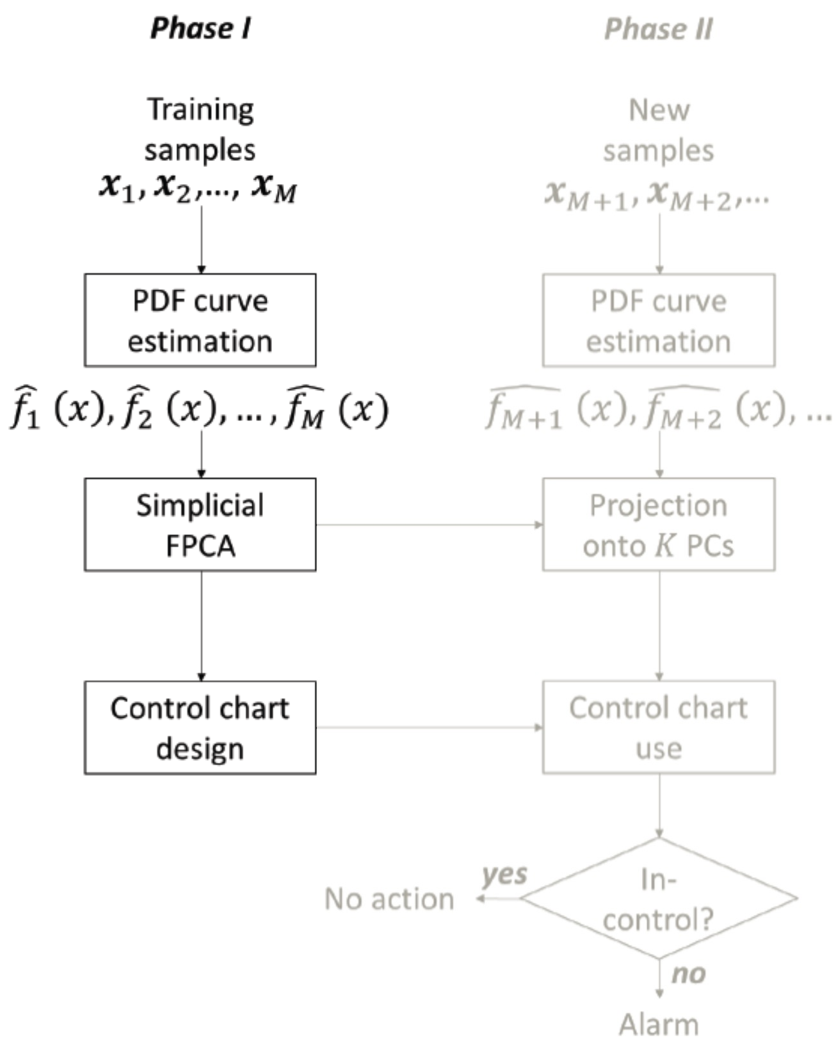
$$\langle f, g \rangle_{\mathcal{B}} = \frac{1}{2\eta} \int_I \int_I \ln \frac{f(t)}{f(s)} \ln \frac{g(t)}{g(s)}\, dt\, ds$$

Bayes space embedding

PDFs of size descriptors in all the sections



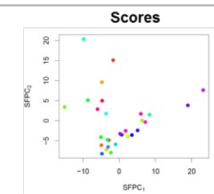
Profile monitoring for PDFs



Hotelling T^2 control chart on the scores

$$T_j^2(K) = \sum_{i=1}^{K} \frac{z_{ji}^2}{\rho_i}, \quad j = 1, 2, \dots,$$

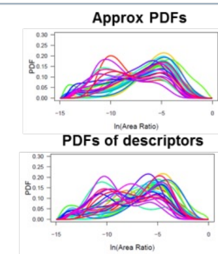
to control deviations along the first K SFPCs



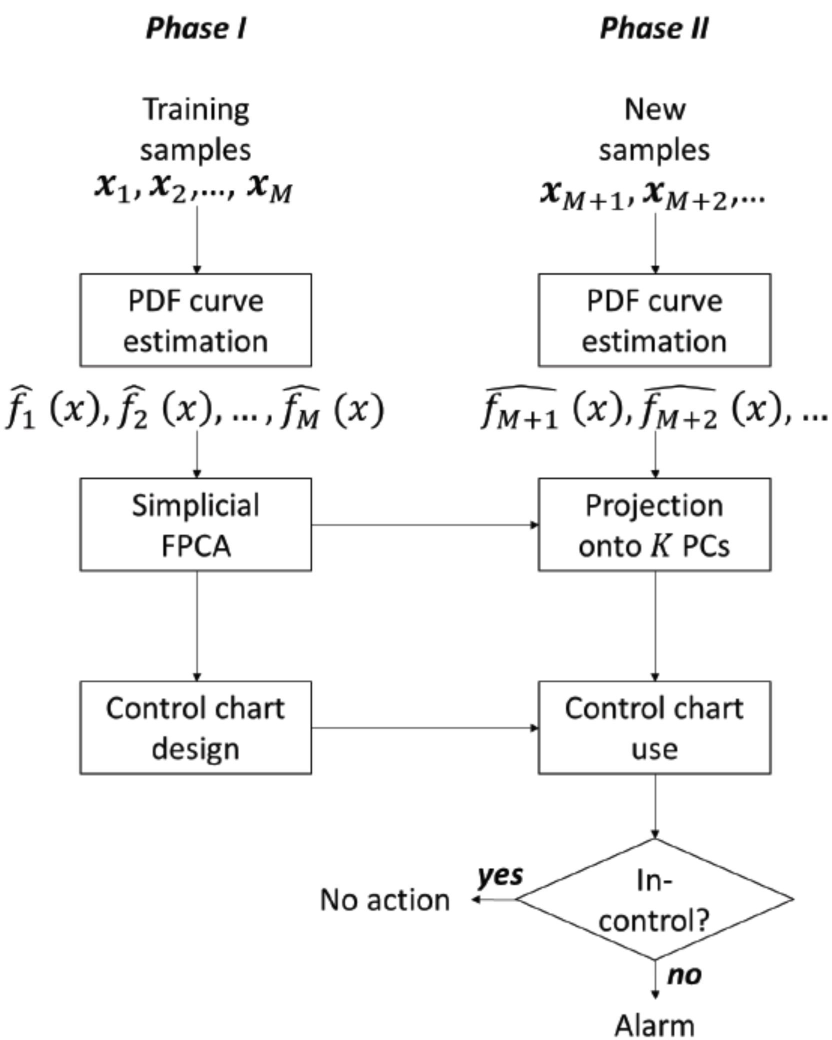
Residuals' control chart

$$SPE_j(K) = \|\hat{f}_j^* \ominus \hat{f}_j\|^2 \quad j = 1, 2, \dots,$$

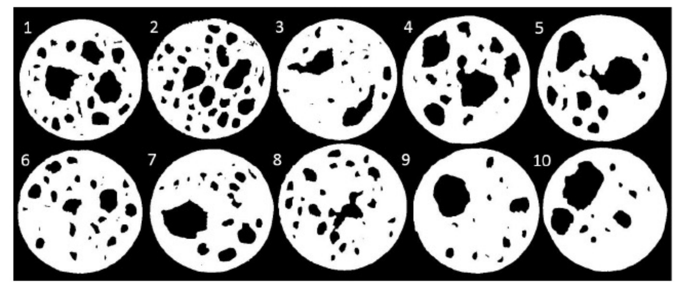
to detect shifts along directions orthogonal to the ones associated with the first K SFPCs



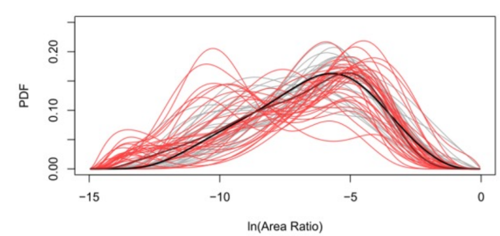
Profile monitoring for PDFs



Application to Phase II sections



PDFs of size descriptors in Phase II sections



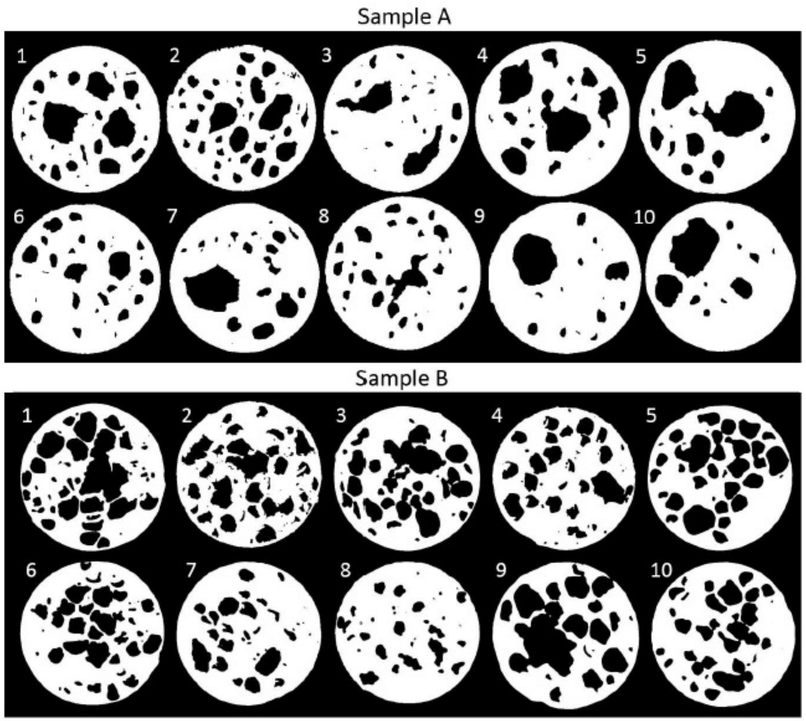
Agenda

Quality control of foamed materials production

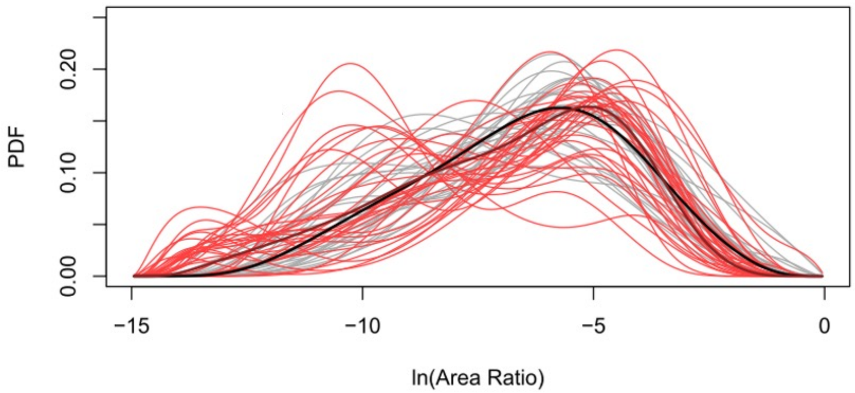
- Two samples of diameter $D = 35\ mm$, produced under the same process conditions, but different methods for slice polishing on the preparation plane.
- We expect to detect a shift in the pore-size distribution, caused by the change of the polishing treatment.
- We employ as indicator the area ratio

$$A_r(i,j) = \frac{A_i(j)}{A_{tot}(j)}, \quad i = 1, ..., N_j;\ j = 1, 2, ...$$

with $A_i(j)$ the area of the i-th pore in the j-th section



Results on metal foam data

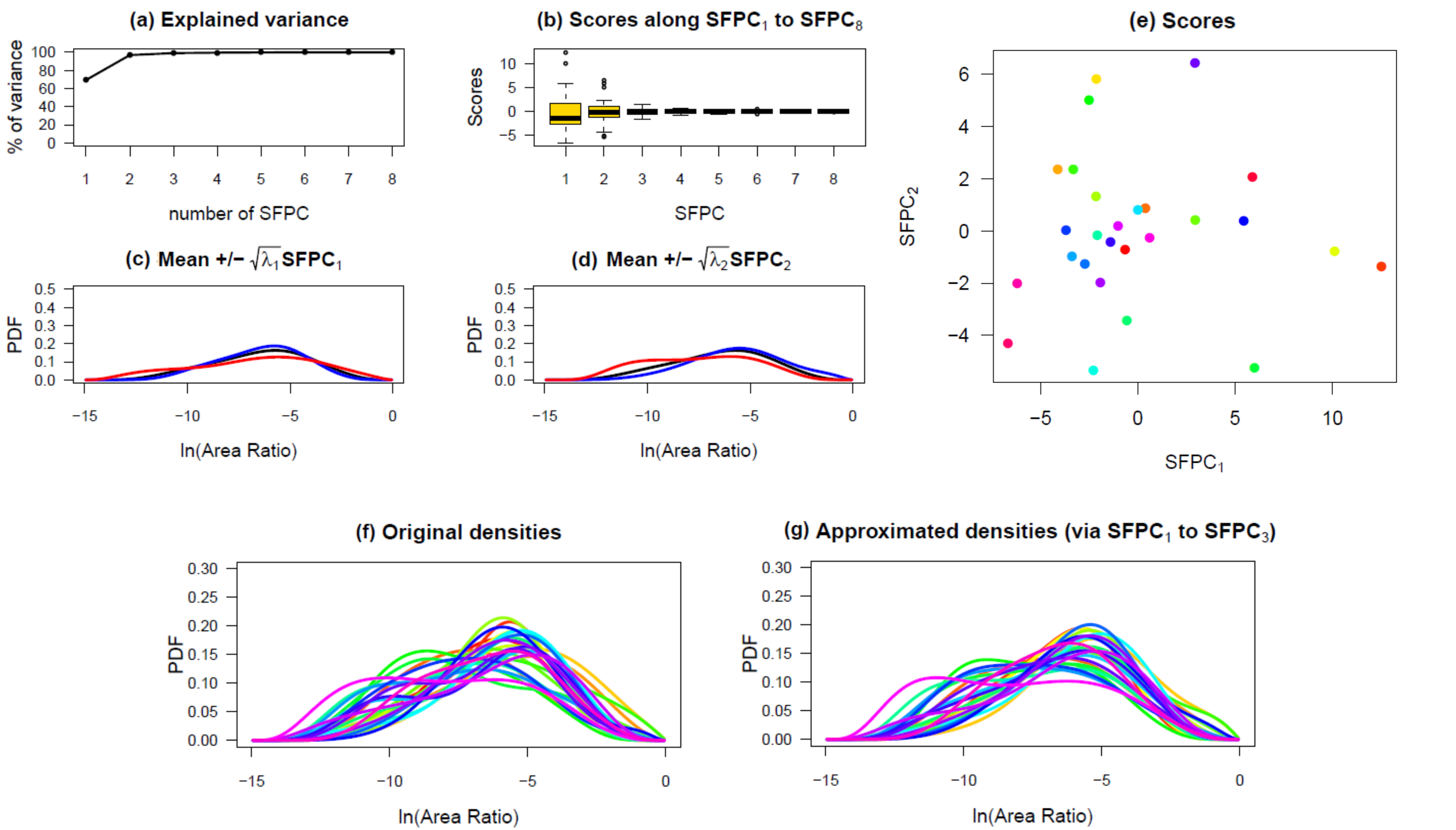


- Grey curves: Phase I dataset
- Red curves: Phase II dataset

- We performed SFPCA on the Phase I dataset and retained *K=3* SFPCs (98% of the variability)

Results on metal foam data

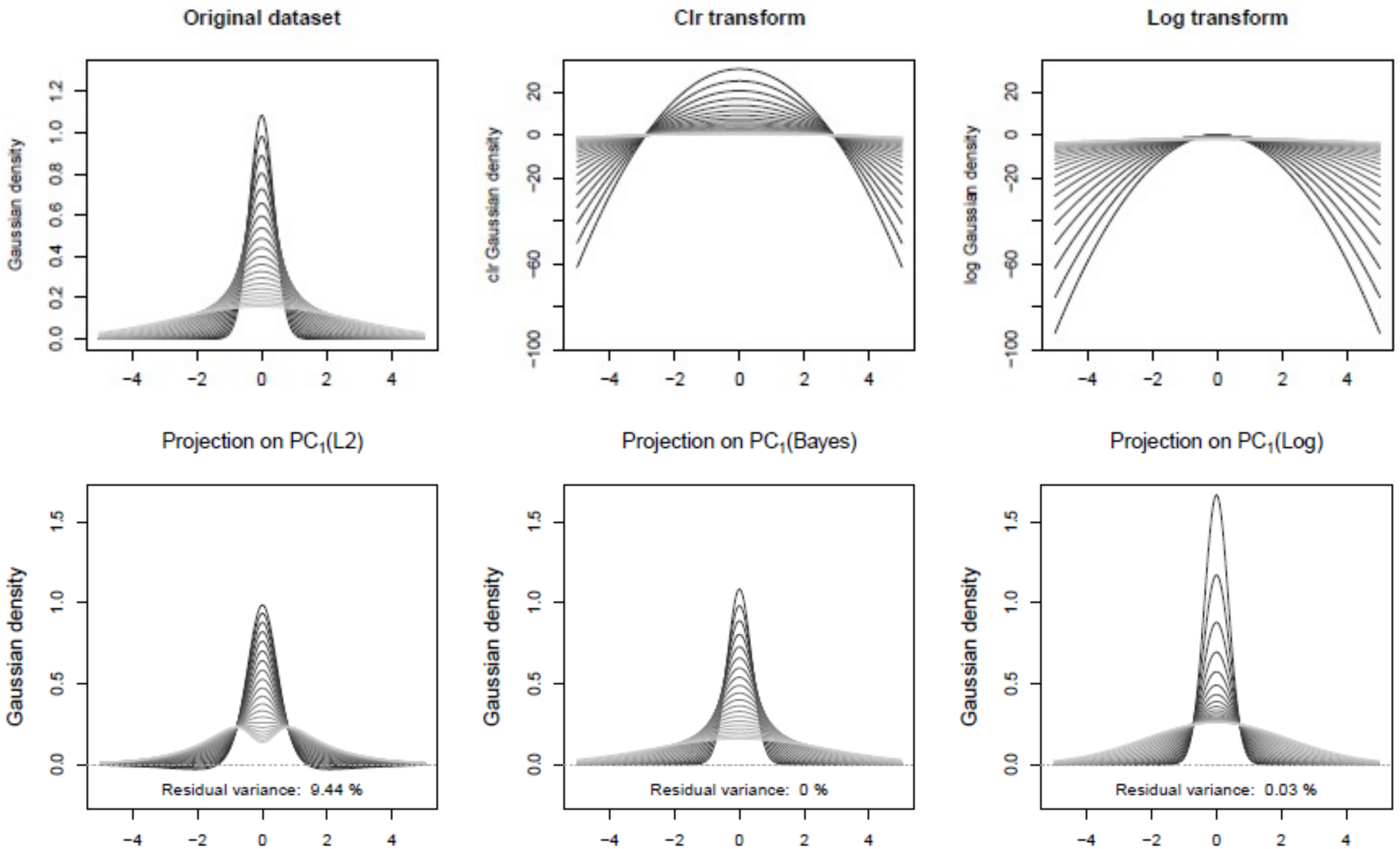
Menafoglio, Grasso, Secchi, Colosimo (Technometrics, 2018)



Recall: L^2 versus Bayes space

Menafoglio, Grasso, Secchi, Colosimo (Technometrics, 2018)

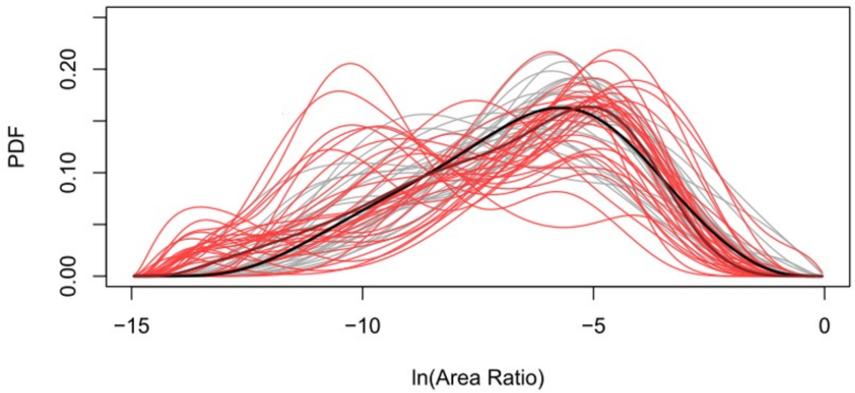
FPCA of Gaussian densities in L^2, in Bayes spaces or based on log-transform



$$[\mathrm{clr}(f)](t) = \log(f(t)) - \int_0^1 \log(f(\tau))\, d\tau, \quad t \in [0, 1].$$

Results on metal foam data

Menafoglio, Grasso, Secchi, Colosimo (Technometrics, 2018)

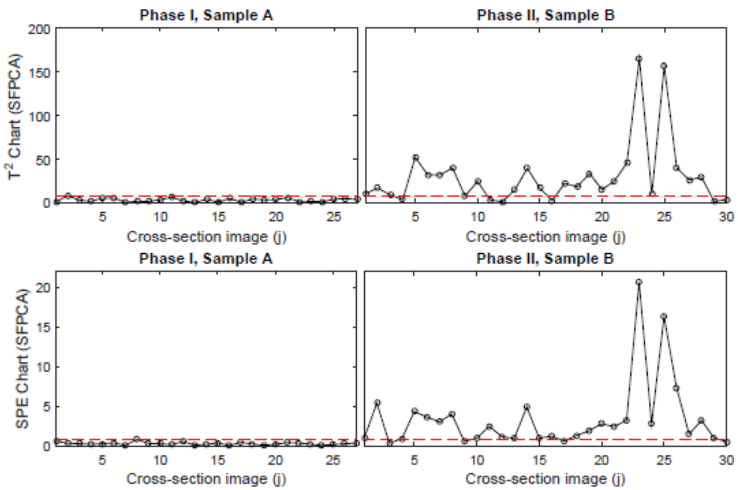


- Grey curves: Phase I dataset
- Red curves: Phase II dataset

- We performed SFPCA on the Phase I dataset and retained *K=3* SFPCs (98% of the variability)
- We built the control charts for
  - the scores along the SFPCs
  - the residuals
- We applied the control chart to the Phase II dataset

Results on metal foam data

Menafoglio, Grasso, Secchi, Colosimo (Technometrics, 2018)

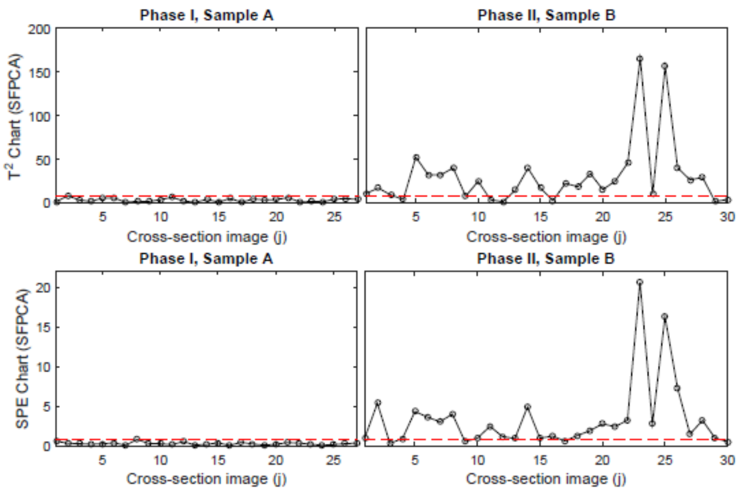


PDF monitoring in Bayes space

- We performed SFPCA on the Phase I dataset and retained *K=3* SFPCs (98% of the variability)
- We built the control charts for
  - the scores along the SFPCs
  - the residuals
- We applied the control chart to the Phase II dataset

Results on metal foam data

Menafoglio, Grasso, Secchi, Colosimo (Technometrics, 2018)



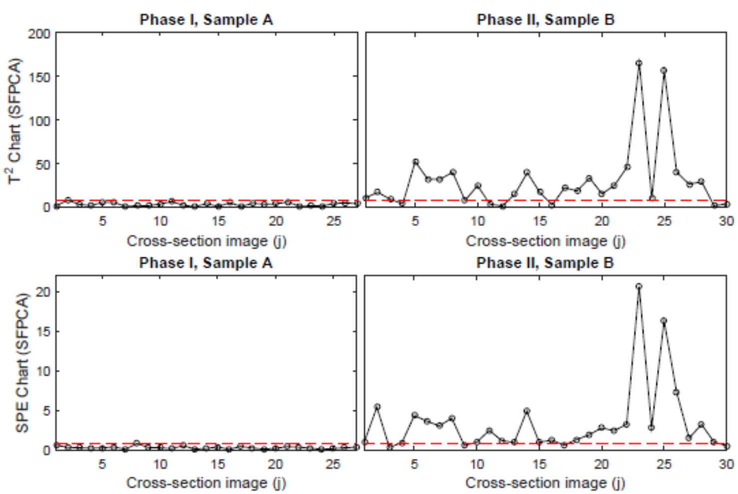
PDF monitoring in Bayes space

Method	Type II error
SFPCA	0.1
FPCA	0.1333
Q-Q plot	0.5333
Shewhart	0.4

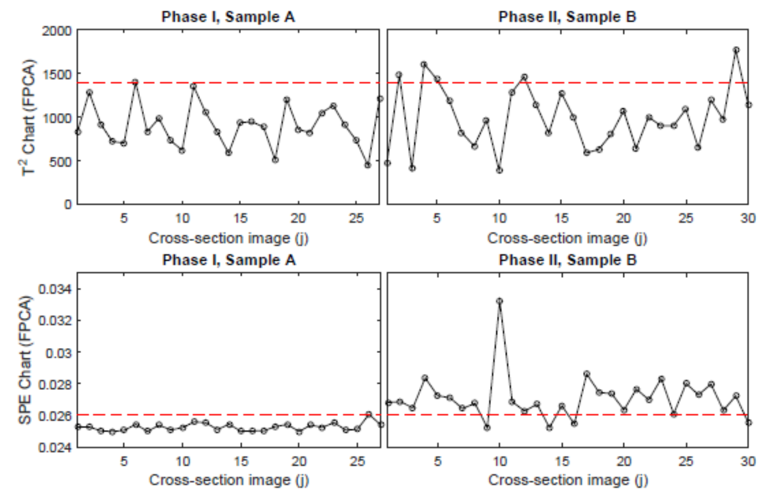
Simulations confirmed that PDF monitoring in Bayes spaces outperforms competitor methods in terms of type II error and ARL

Results on metal foam data

Menafoglio, Grasso, Secchi, Colosimo (Technometrics, 2018)



PDF monitoring in Bayes space

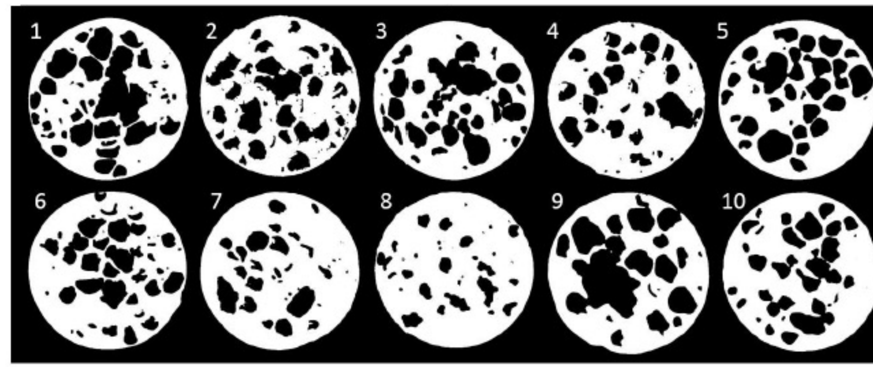


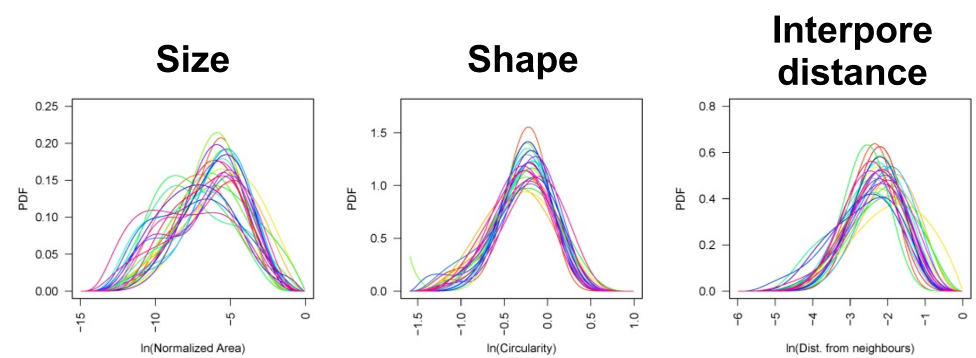
PDF monitoring in L^2

Simulations confirmed that PDF monitoring in Bayes spaces outperforms competitor methods in terms of type II error and ARL

Final remarks

- Monitoring the **whole PDF** allows to capture more precisely deviations from in-control conditions
- **Multivariate extensions** are possible, e.g., building multiple charts for a number of interesting indicators and control the global type I error via Bonferroni corrections





Agenda

5. Functional control charts

5.1. Introduction & motivations: profile monitoring

5.2. Control charts for probability density functions

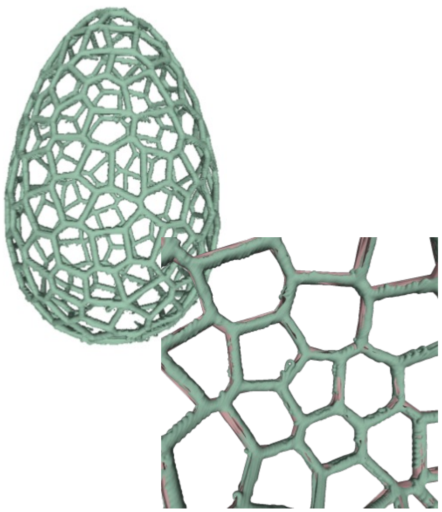
5.3. Anomaly detection in PDFs from image data

5.3. An application in Additive Manufacturing

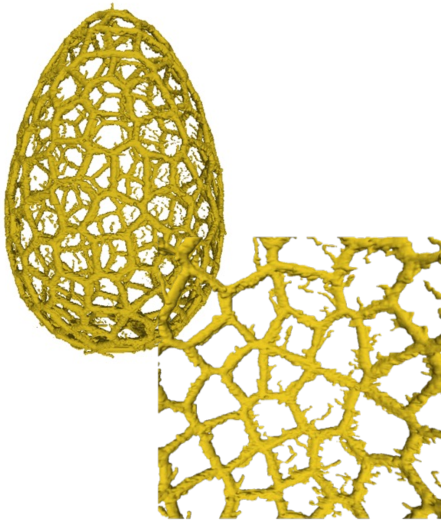
The applicative context

- Additive manufacturing: 3D printing of varied types of complex shapes
- How to perform Statistical Process Control (SPC) in the presence of complex shapes?
- Real produced objects are not ideal: in-control production cannot be expected to coincide with nominal eggs.

What is in-control?



Characterization of in control variability

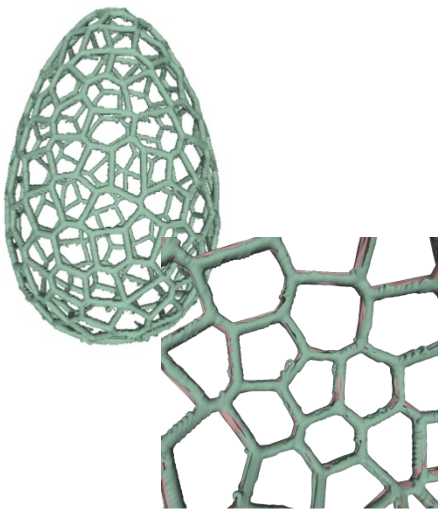


Identification of out-of-control

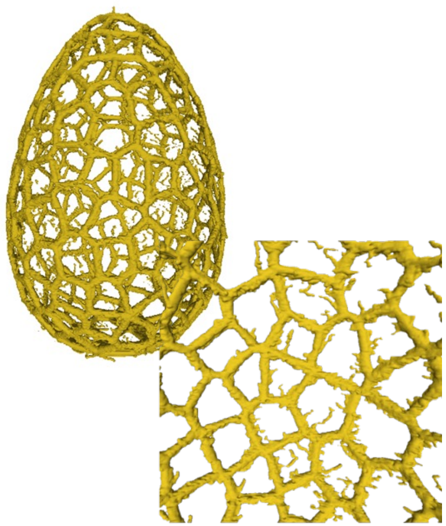
The applicative context

General objectives:

- Characterize the in-control variability (within- and between-part)
- Design a control chart to identify out-of-control conditions in the presence of a series of products



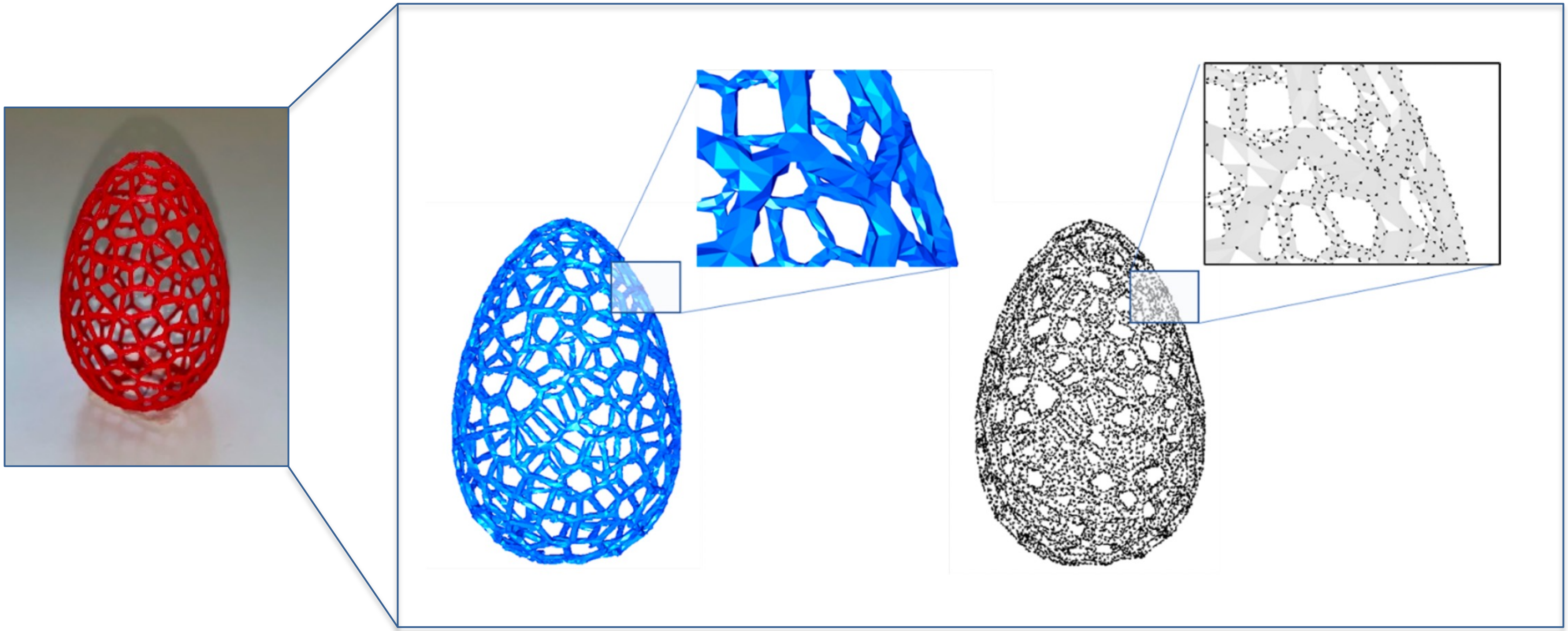
Characterization of in control variability



Identification of out-of-control

From produced shapes to data

- Data are collected via tomography and represented as (very fine) meshes
- Meshes are graphs, represented as vertices, edges, and faces
- Problem reduced to characterize the variability of a point-cloud (that of the vertices) w.r.t. a nominal



Deviation between two objects

- To understand the variability of the phenomenon, we need to define how dissimilar the produced shape Y (point clouds) is with respect to the nominal shape X
- In practice, we define two maps:

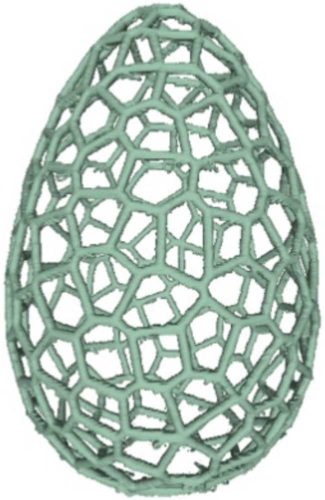
$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$$

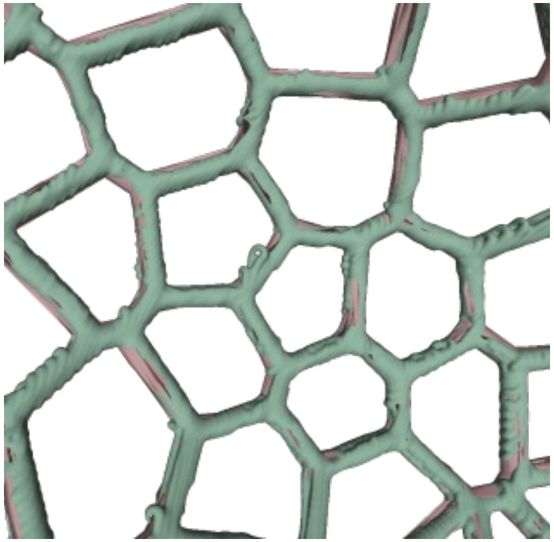
How far the closest point of Y is from x

$$d_X : Y \to \mathbb{R}^+,\ d_X(y) := \inf_{x \in X} d(x, y)$$

How far the closest point of X is from y

Produced and nominal shapes, after alignment





High values of the maps are informative of local defects

Deviation between two objects

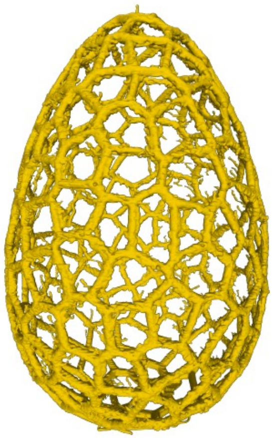
- To understand the variability of the phenomenon, we need to define how dissimilar the produced shape Y (point clouds) is with respect to the nominal shape X
- In practice, we define two maps:

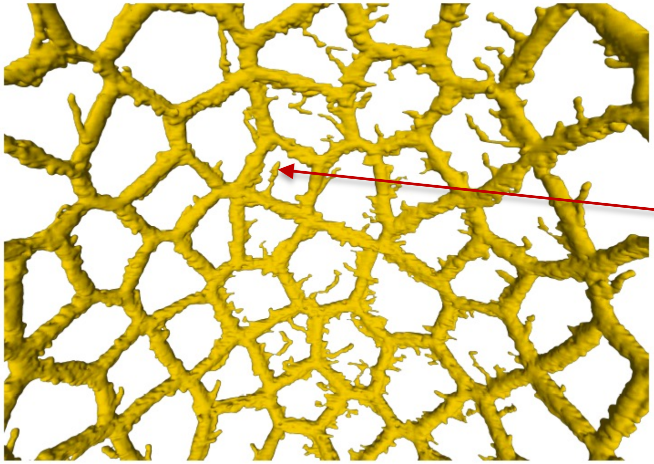
$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$$

How far the closest point of Y is from x

$$d_X : Y \to \mathbb{R}^+,\ d_X(y) := \inf_{x \in X} d(x, y)$$

How far the closest point of X is from y





Addition of material will be associated with locally high values of $d_X(y)$

Deviation between two objects

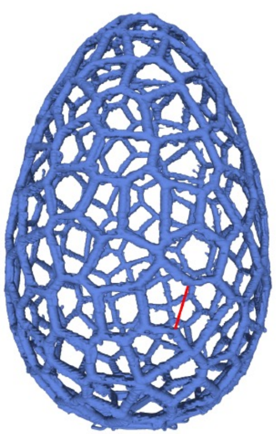
- To understand the variability of the phenomenon, we need to define how dissimilar the produced shape Y (point clouds) is with respect to the nominal shape X
- In practice, we define two maps:

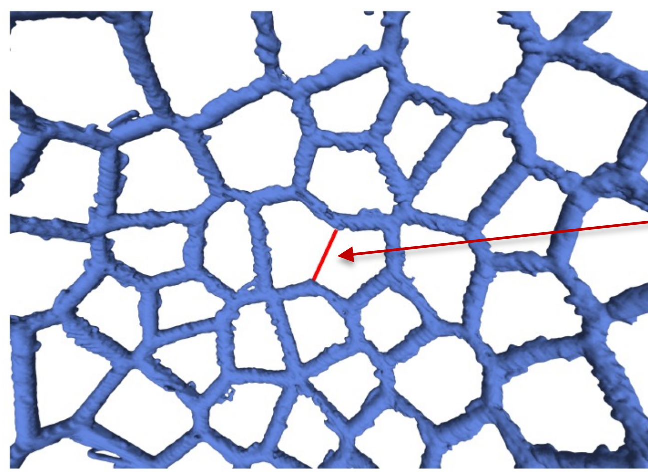
$$d_Y : X \to \mathbb{R}^+, \ d_Y(x) := \inf_{y \in Y} d(x, y)$$

How far the closest point of Y is from x

$$d_X : Y \to \mathbb{R}^+, \ d_X(y) := \inf_{x \in X} d(x, y)$$

How far the closest point of X is from y





Missing material will be associated with locally high values of $d_Y(x)$

Deviation between two objects

- To understand the variability of the phenomenon, we need to define how dissimilar the produced shape Y (point clouds) is with respect to the nominal shape X
- In practice, we define two maps:

$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$$

How far the closest point of Y is from x

$$d_X : Y \to \mathbb{R}^+,\ d_X(y) := \inf_{x \in X} d(x, y)$$

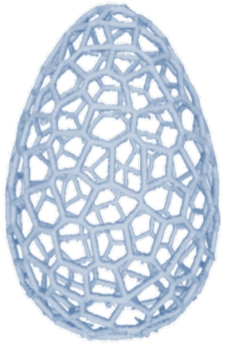
How far the closest point of X is from y

Both maps are informative, and each informs on a different type of deviation!

Note: The domain of $d_Y(x)$ are the points in the nominal shape X; ***the domain of $d_X(y)$ are the points in the produces shape Y***

Statistical process control

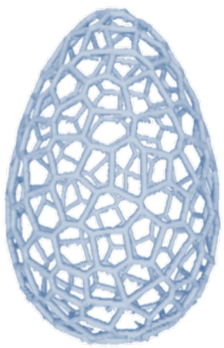
- **Objective**: Characterize the in-control variability (within- and between-part) from a sample of N produced objects



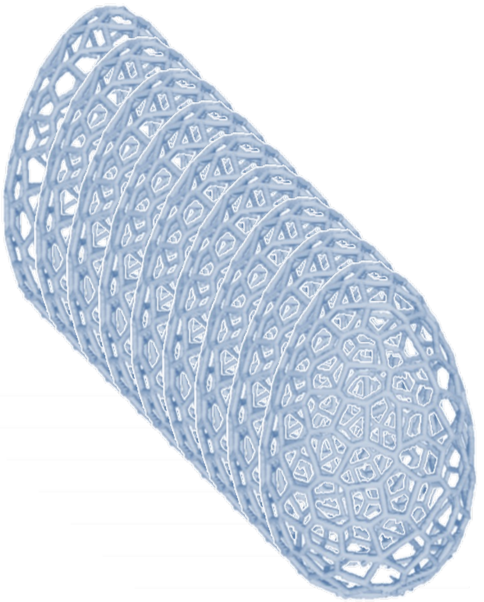
Nominal shape

Statistical process control

- **Objective**: Characterize the in-control variability (within- and between-part) from a sample of N produced objects



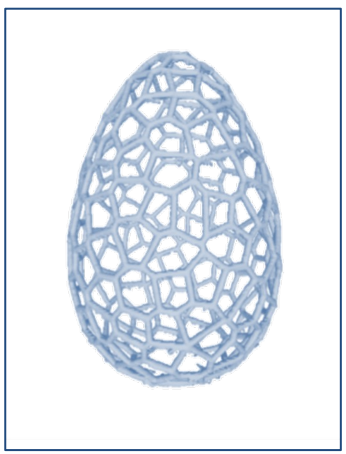
Nominal shape

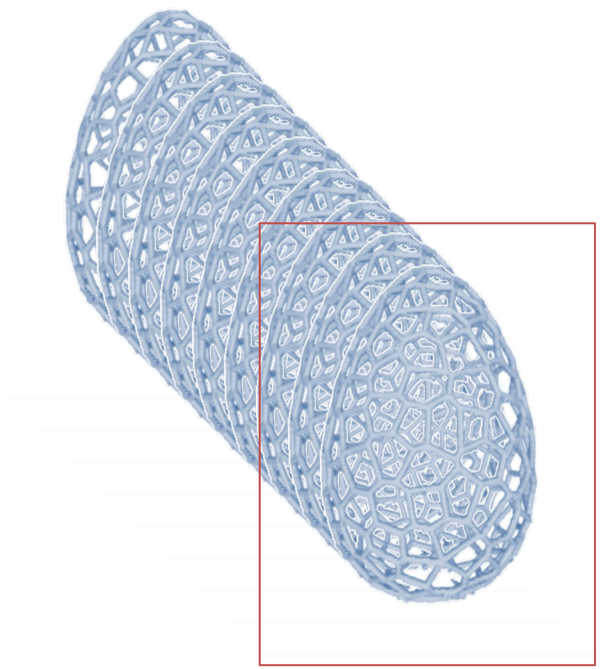


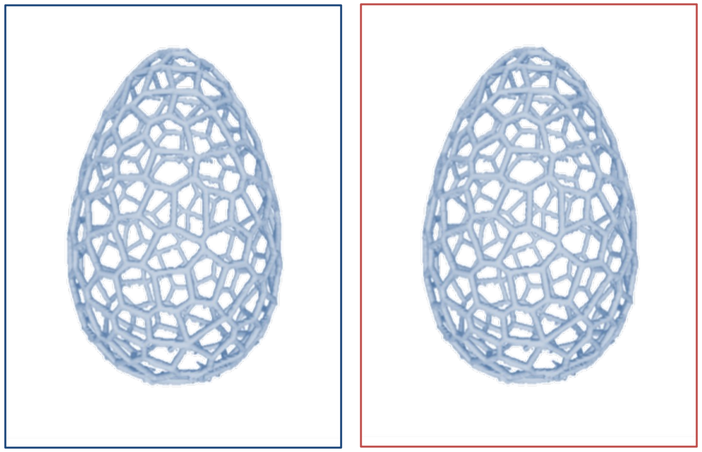
Produced shapes

Statistical process control

- **Objective**: Characterize the in-control variability (within- and between-part) from a sample of N produced objects







$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$$

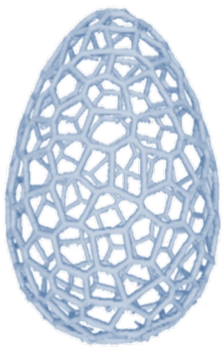
$$d_X : Y \to \mathbb{R}^+,\ d_X(y) := \inf_{x \in X} d(x, y)$$

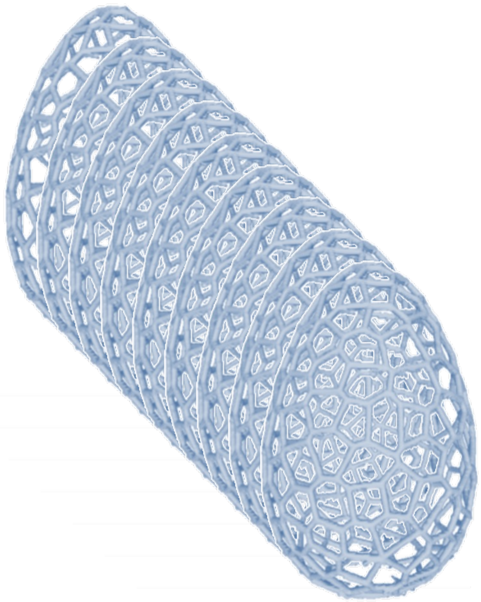
Nominal shape

Produced shapes

Comparison of each produced shape with nominal through the pair $d_{Y,i}(x)$, $d_X(y)$

- **Objective**: Characterize the in-control variability (within- and between-part) from a sample of N produced objects





$d_{Y,1}(x),\ d_X(y)$

$d_{Y,2}(x),\ d_X(y)$

…

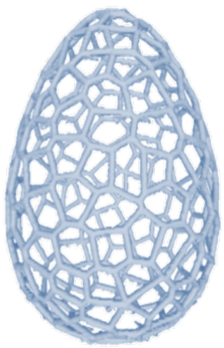
$d_{Y,i}(x),\ d_X(y)$

…

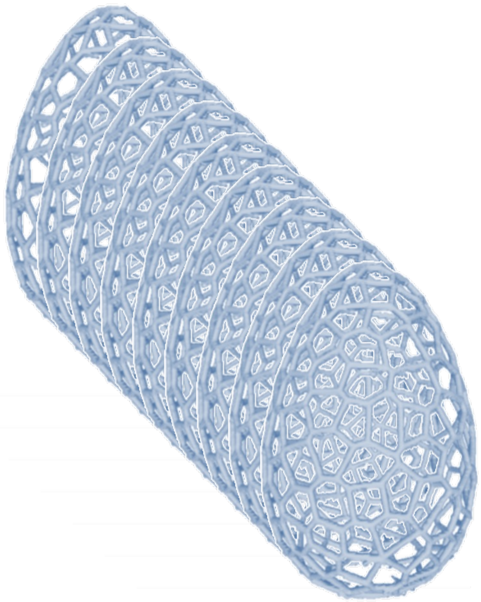
$d_{Y,N}(x),\ d_X(y)$

Nominal shape → Produced shapes → Profile monitoring of $\{d_{Y,i}(x),\ d_X(y)\}$?

- **Objective**: Characterize the in-control variability (within- and between-part) from a sample of N produced objects



Nominal shape



Produced shapes

$d_{Y,1}(x)$, $d_X(y)$

$d_{Y,2}(x)$, $d_X(y)$

…

$d_{Y,i}(x)$, $d_X(y)$

…

$d_{Y,N}(x)$, $d_X(y)$

Profile monitoring of $\{d_{Y,i}(x), d_X(y)\}$?

- **Objective**: Characterize the in-control variability (within- and between-part) from a sample of N produced objects

Recall:

$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$$

How far the closest point of Y is from x: defined over the nominal X

$$d_X : Y \to \mathbb{R}^+,\ d_X(y) := \inf_{x \in X} d(x, y)$$

How far the closest point of X is from y: defined over the produced shape Y

$d_{Y,1}(x)$, $d_X(y)$

$d_{Y,2}(x)$, $d_X(y)$

…

$d_{Y,i}(x)$, $d_X(y)$

…

$d_{Y,N}(x)$, $d_X(y)$

Profile monitoring of $\{d_{Y,i}(x), d_X(y)\}$?

Statistical process control

- **Objective**: Characterize the in-control variability (within- and between-part) from a sample of N produced objects

Recall:

$$d_Y : X \to \mathbb{R}^+,\; d_Y(x) := \inf_{y \in Y} d(x,y)$$

How far the closest point of Y is from x: defined over the nominal X

$$d_X : Y \to \mathbb{R}^+,\; d_X(y) := \inf_{x \in X} d(x,y)$$

How far the closest point of X is from y: defined over the produced shape Y

$d_{Y,1}(x)$, $d_X(y)$

$d_{Y,2}(x)$, $d_X(y)$

…

$d_{Y,i}(x)$, $d_X(y)$

…

$d_{Y,N}(x)$, $d_X(y)$

Cannot be directly compared!

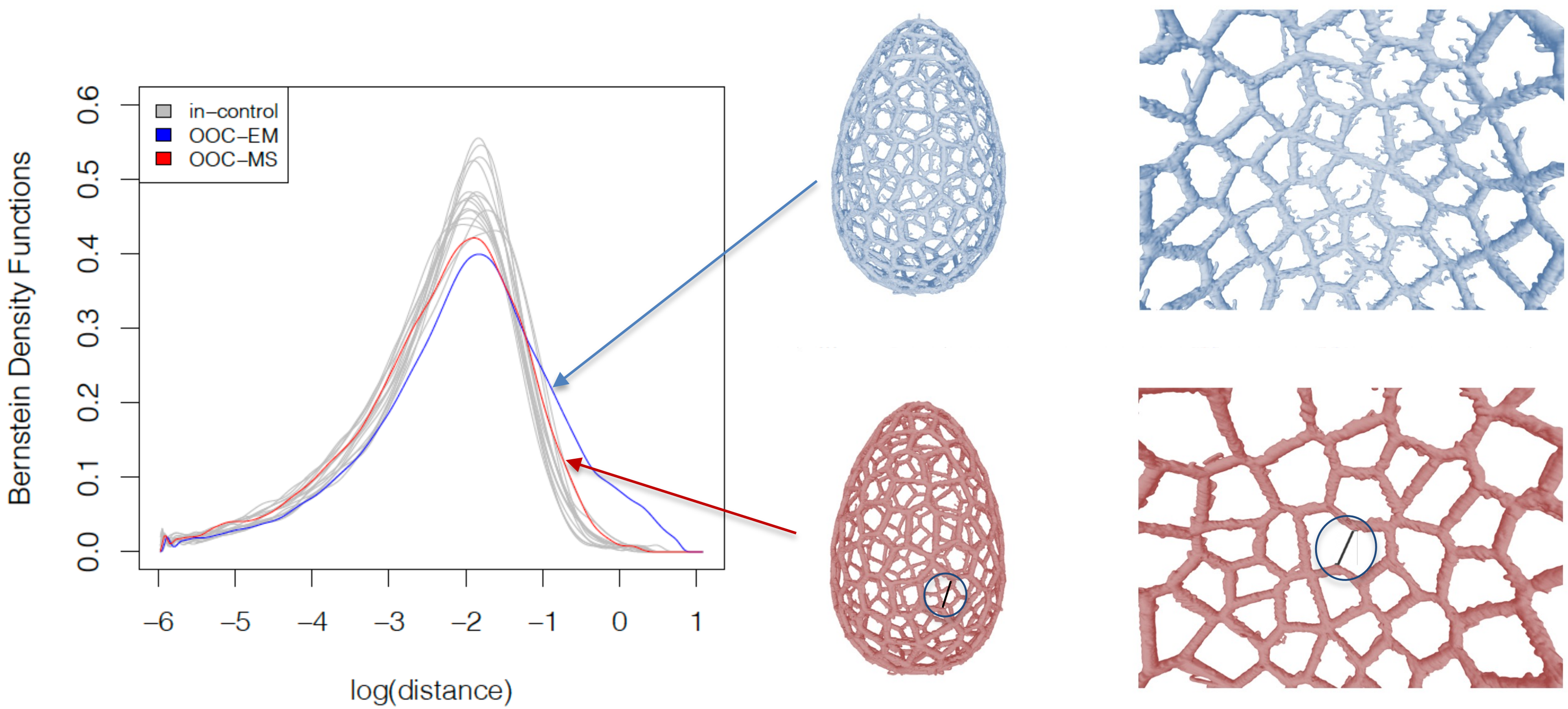
Profile monitoring of $\{d_{Y,i}(x), d_X(y)\}$?

Representing deviation maps through densities

- Each produced object can however be represented through a pair of **exhaustive summaries** of the $d_{Y,i}(x)$, $d_X(y)$, i.e., the **PDFs of the deviation maps**

Representing deviation maps through densities

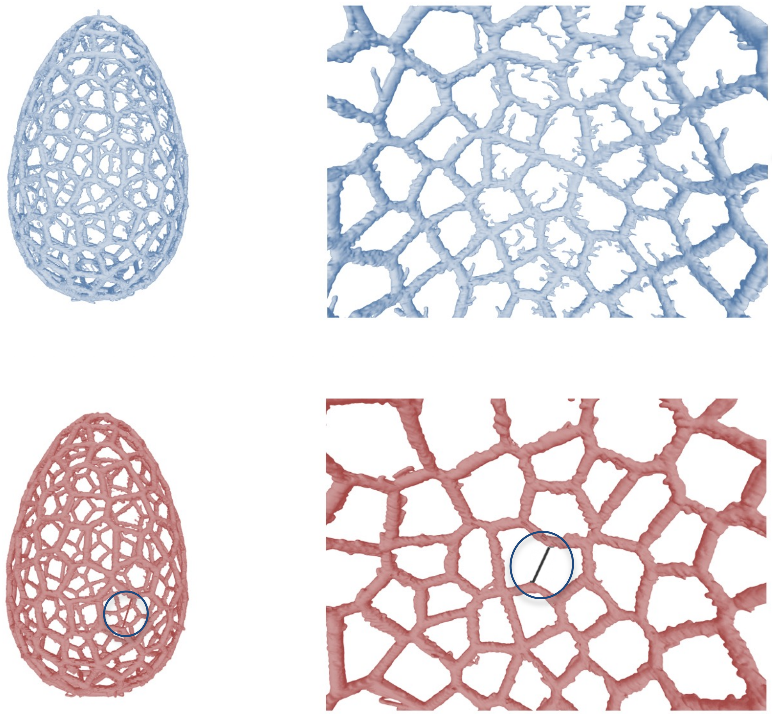
- Each produced object can however be represented through a pair of **exhaustive summaries** of the $d_{Y,i}(x)$, $d_X(y)$, i.e., the **PDFs of the deviation maps**

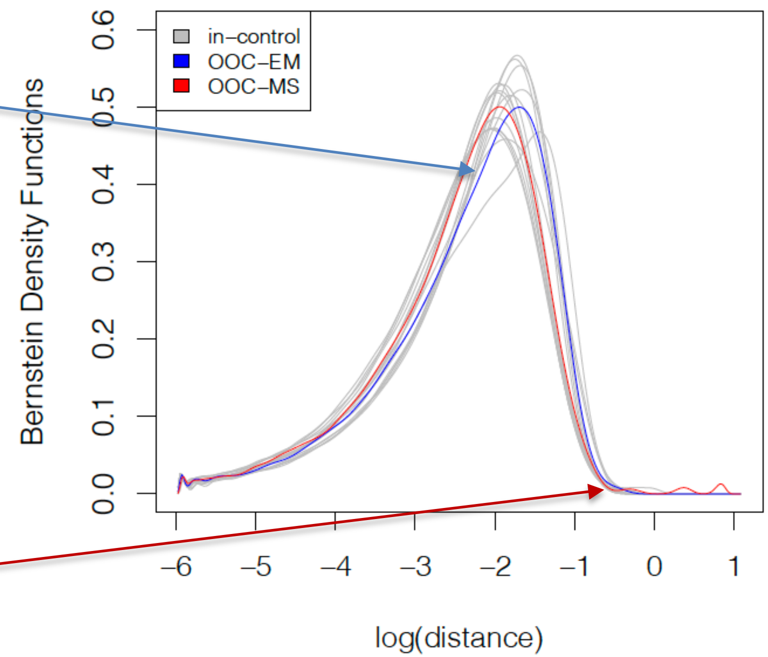


Densities of map $d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x,y)$ *(How far the closest point of Y is from x, X being the nominal)*

Representing deviation maps through densities

- Each produced object can however be represented through a pair of **exhaustive summaries** of the $d_{Y,i}(x)$, $d_X(y)$, i.e., the **PDFs of the deviation maps**

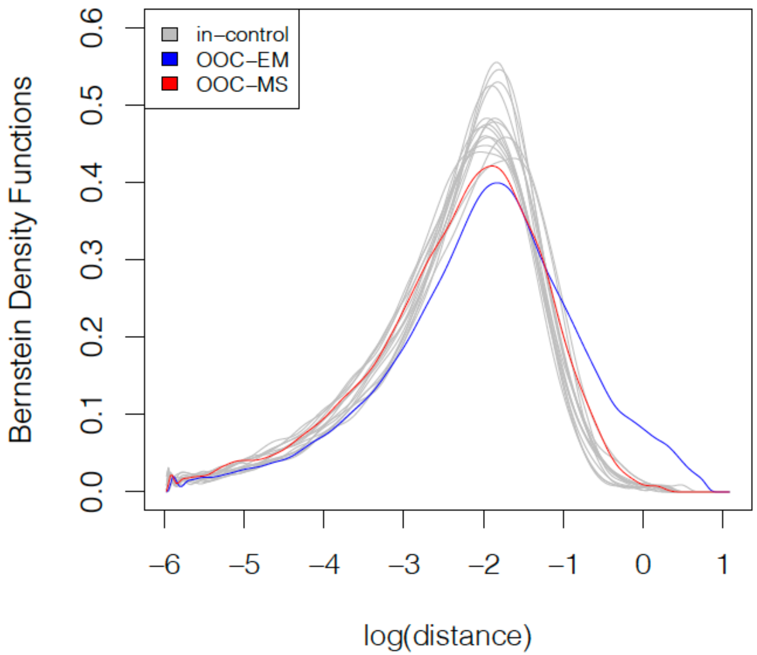




Densities of map $d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$ *(How far the closest point of Y is from x, X being the nominal)*

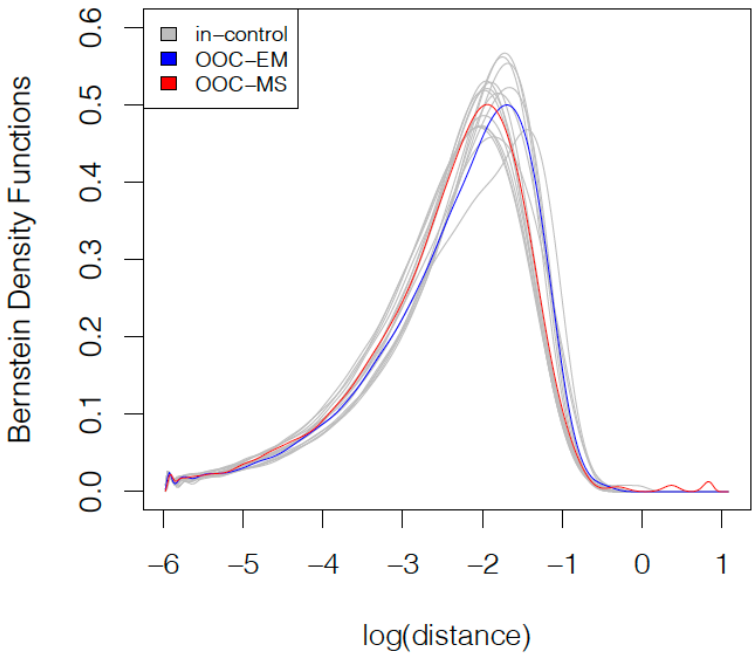
Representing deviation maps through densities

- Each produced object can however be represented through a pair of **exhaustive summaries** of the $d_{Y,i}(x)$, $d_X(y)$, i.e., the **PDFs of the deviation maps**



Densities of map

$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$$



Densities of maps

$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x, y)$$

Representing deviation maps through densities

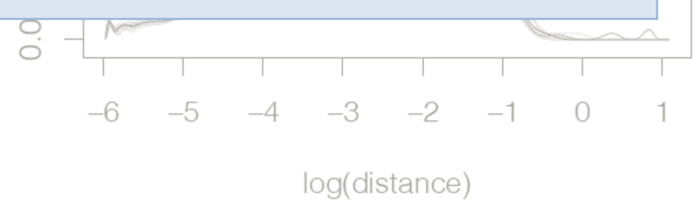
- Each produced object can however be represented through a pair of **exhaustive summaries** of the $d_{Y,i}(x)$, $d_X(y)$, i.e., the **PDFs of the deviation maps**

Having represented the data through densities, an object-oriented control chart scheme can be built, e.g. by:

- Dimensionality reduction (SFPCA)
- Control charts on the scores (T2 statistic) and on residuals from the SFPCA representation (Q statistic)

Densities of map

$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x,y)$$

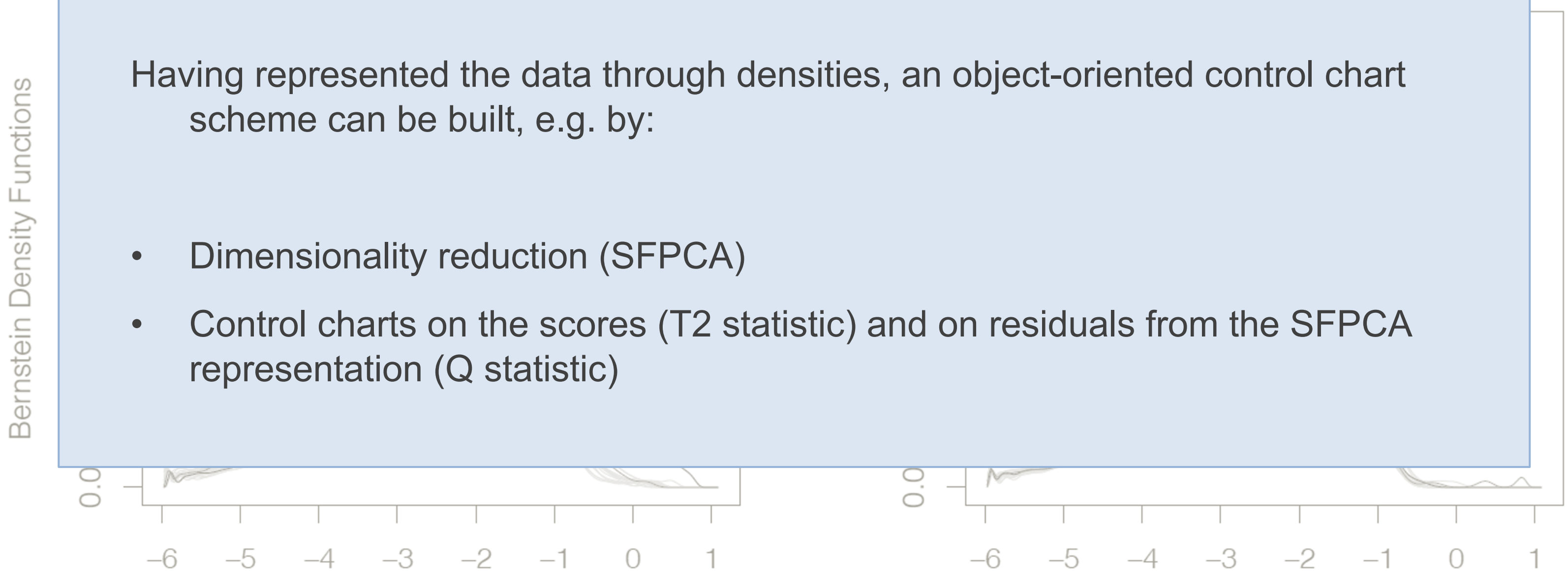


Densities of maps

$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x,y)$$

Representing deviation maps through densities

- Each produced object can however be represented through a pair of **exhaustive summaries** of the $d_{Y,i}(x)$, $d_X(y)$, i.e., the **PDFs of the deviation maps**



Having represented the data through densities, an object-oriented control chart scheme can be built, e.g. by:

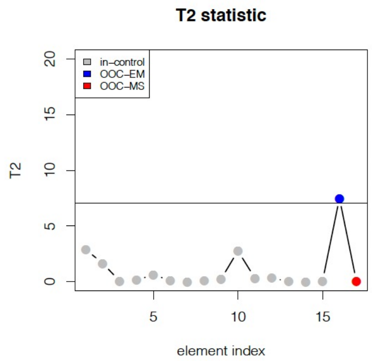
- Dimensionality reduction (SFPCA)
- Control charts on the scores (T2 statistic) and on residuals from the SFPCA representation (Q statistic)

A relevant part of an FDA analysis could be to find the right representation for the data object!

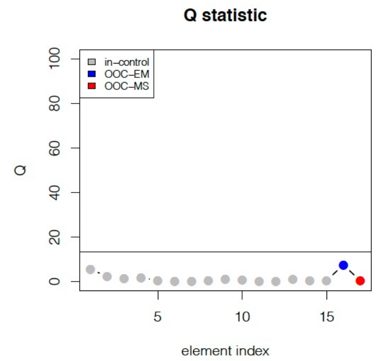
$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x,y)$$

$$d_Y : X \to \mathbb{R}^+,\ d_Y(x) := \inf_{y \in Y} d(x,y)$$

Control charts

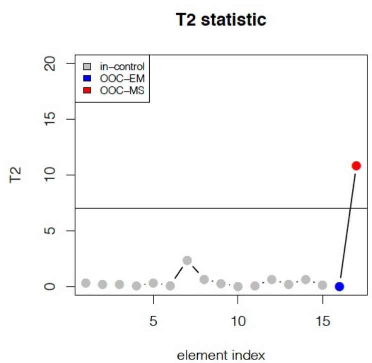


(a) $T^{2(P)}$ chart

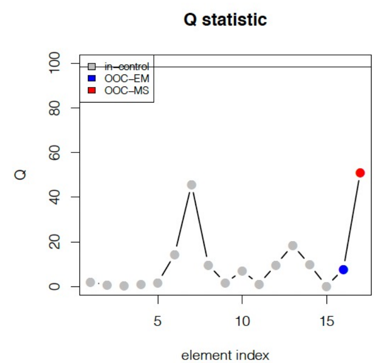


(b) Q^P chart

Control chart built on the first map: it identifies excess of material



(c) $T^{2(S)}$ chart



(d) Q^S chart

Control chart built on the second map: it identifies lack of material

The two control charts can be combined using Bonferroni correction

Conclusions and take home messages

- Anomaly detection is a crucial problem in industrial settings, when anomalies represent out-of-control behavior in the production process
- Control chart schemes can be build in the Bayes space, to control process or product quality from density data
- Density data may not only arise as the «raw data» but from aggregation of scalar (local) indicators
- Monitoring the entire distribution (through the Bayes space theory) instead of some summary indices (e.g., mean and variance) allows one to significantly gain in power of the statistical process control.

Selected references

- Colosimo, B.M., Pacella, M. (2007). On the Use of Principal Component Analysis to Identify Systematic Patterns in Roundness Proles. Quality and Reliability Engineering International, 23(6), 707-725.
- Colosimo, B.M., Pacella, M. (2010). A Comparison Study of Control Charts for Statistical Monitoring of Functional Data. International Journal of Production Research, 48(6), 1575-1601.
- Egozcue, J.J., Dìaz-Barrero, J.L., Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. Acta Mathematica Sinica (English Series) 22(4), 1175-1182.
- Hron, K., Menafoglio, A., Templ, M., Hruzova, K., Filzmoser, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. Computational Statistics & Data Analysis, 94, 330-350.
- **Menafoglio, A., Grasso, M., Secchi, P., Colosimo, B.M. (2018). Profile Monitoring of Probability Density Functions via Simplicial Functional PCA with application to Image Data. Technometrics, 60(4), 497-510**.
- **R. Scimone, T. Taormina, B. M. Colosimo, M. Grasso, A. Menafoglio, P. Secchi (2021) "Statistical Modeling and Monitoring of Geometrical Deviations in Complex Shapes With Application to Additive Manufacturing", Technometrics. ISSN:0040-1706. DOI: 10.1080/00401706.2021.1961870.**
- Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V. (2014). Bayes Hilbert spaces. Aust. N. Z. J. Stat., 56(2), 171-194.
- Wang, K., & Tsung, F. (2005). Using prole monitoring techniques for a data-rich environment with huge sample size. Quality and Reliability Engineering International 21(7). 677-688.