Anomaly detection for density data based on control charts

Alessandra Menafoglio[1*]

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

*alessandra.menafoglio@polimi.it

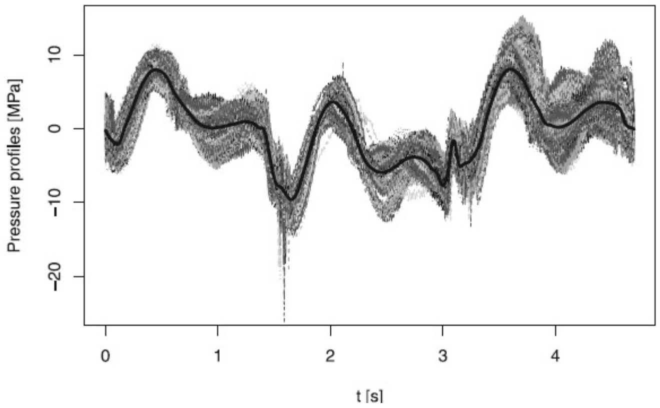
Table of contents

Table of contents

Introduction to profile monitoring

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

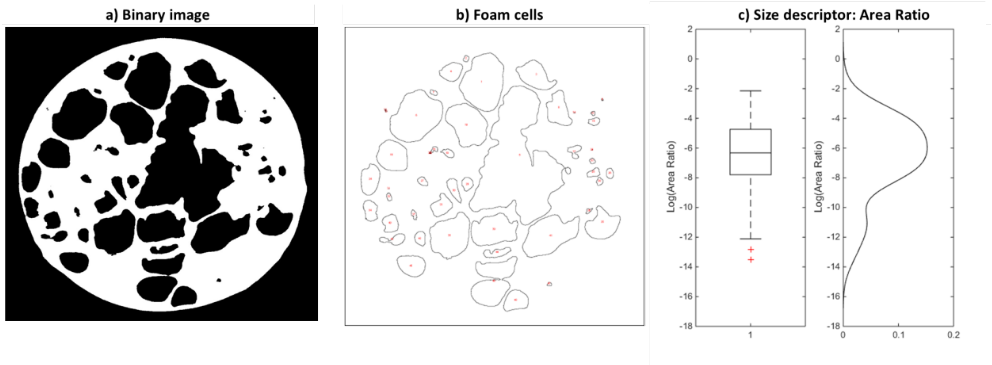
Waterjet cutting



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**

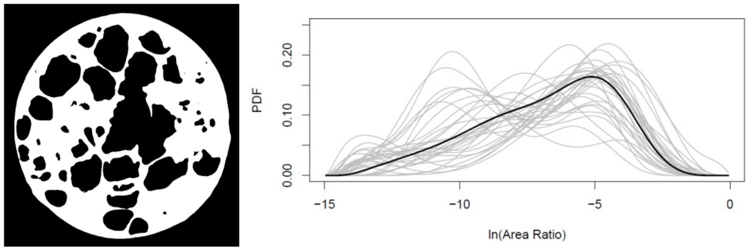
Other (more exploratory) methods for anomaly detection in density datasets (not covered here): methods based on depth measures
A. Menafoglio, L. Guadagnini, A. Guadagnini, P. Secchi (2021) "Object oriented spatial analysis of natural concentration levels of chemical species in large-scale groundwater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

Introduction to profile monitoring

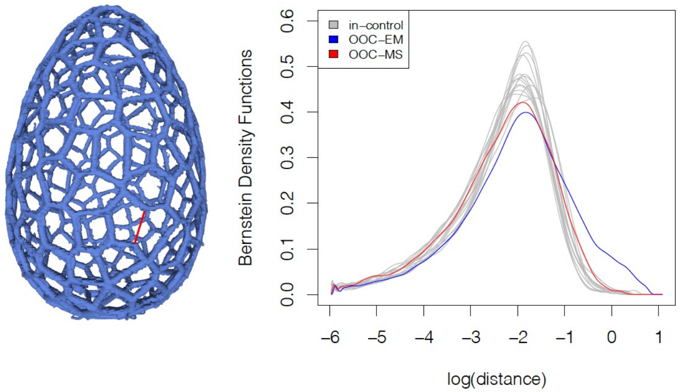
- Profile monitoring techniques have been based so far on (e.g., Colosimo, B.M., Pacella, M., 2007, 2010)
    1. Dimensionality reduction (e.g., via FPCA)
    2. Monitoring of the projection over a functional basis
    3. Monitoring of the residuals of the approximation

- 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

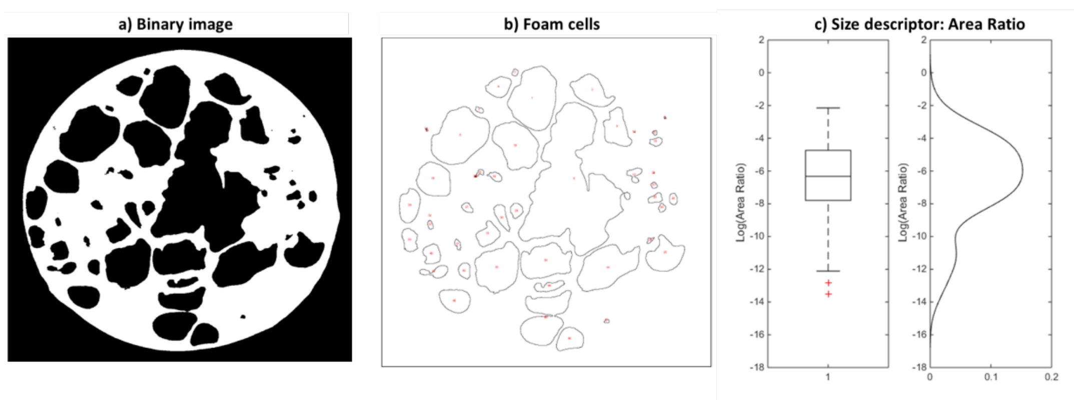


Distribution of point-to-point distances between the produced piece and the nominal object

Table of contents

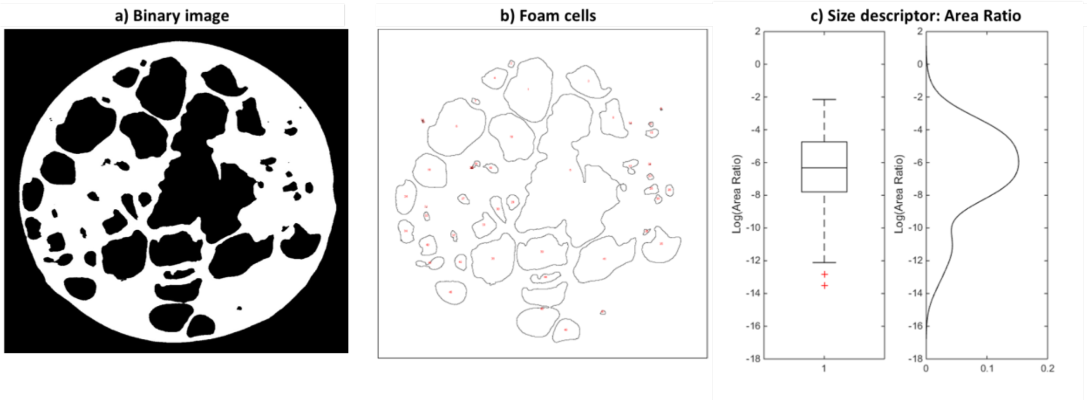
Quality control of foamed materials production

- **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

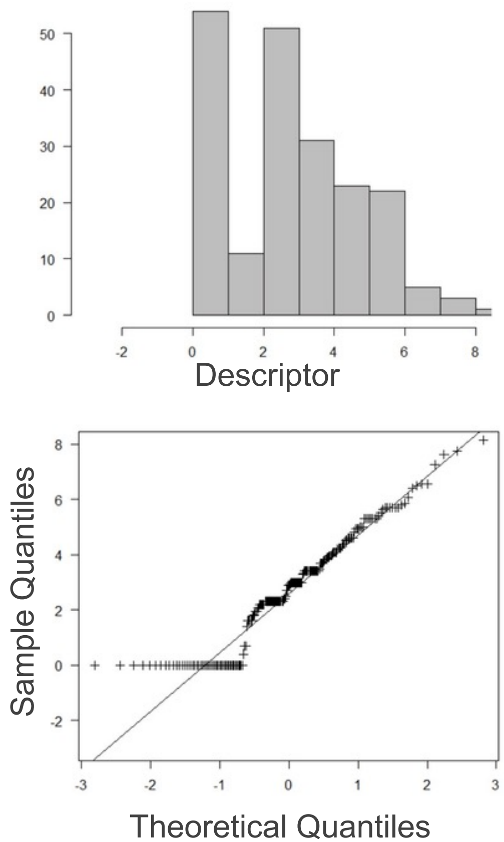
Available methods for statistical process control of distributional data still have limitations

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



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

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.

$\rightarrow$ poor monitoring performances with PDFs

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

Recall: Bayes spaces for PDF data

Bayes space geometry

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

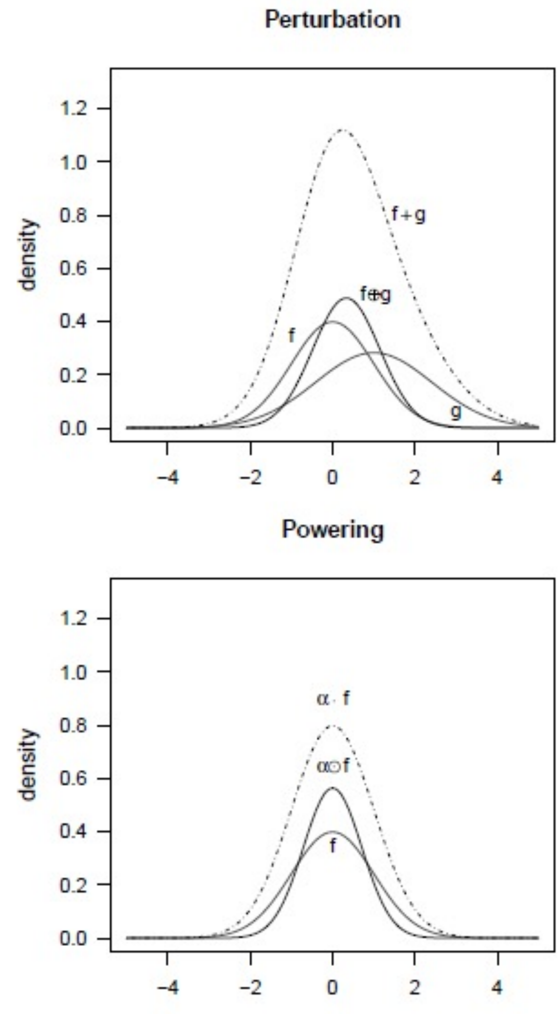
$$(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s)\,ds},$$

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

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

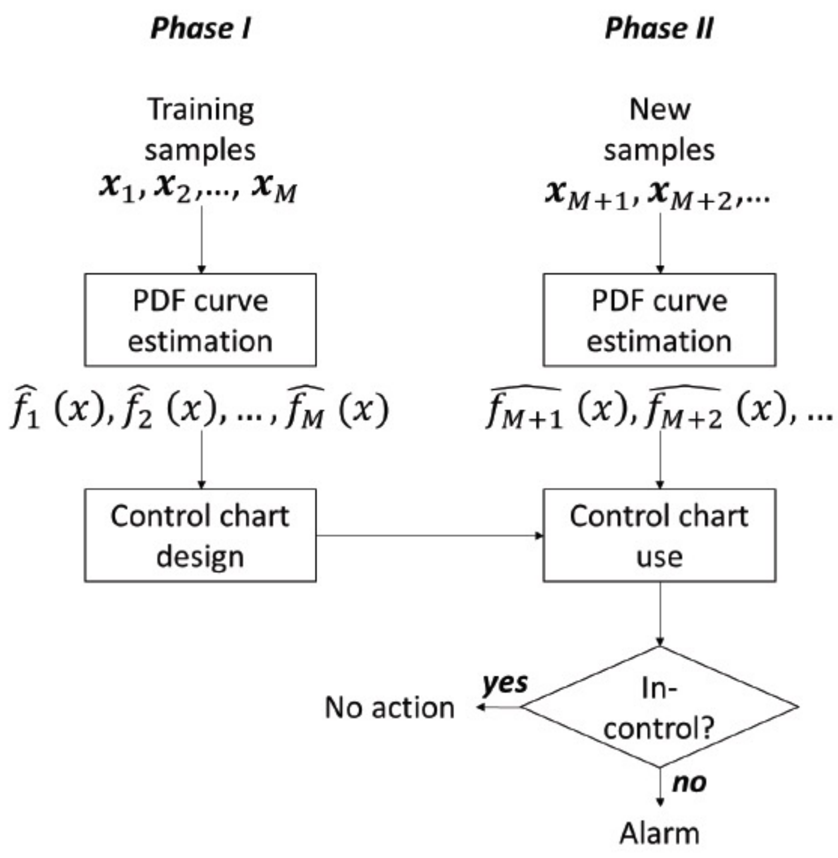
Only (log-)ratios between probabilities are meaningful (~ odds-ratio) as data represent the distribution of a *total* mass (=1) over a domain

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

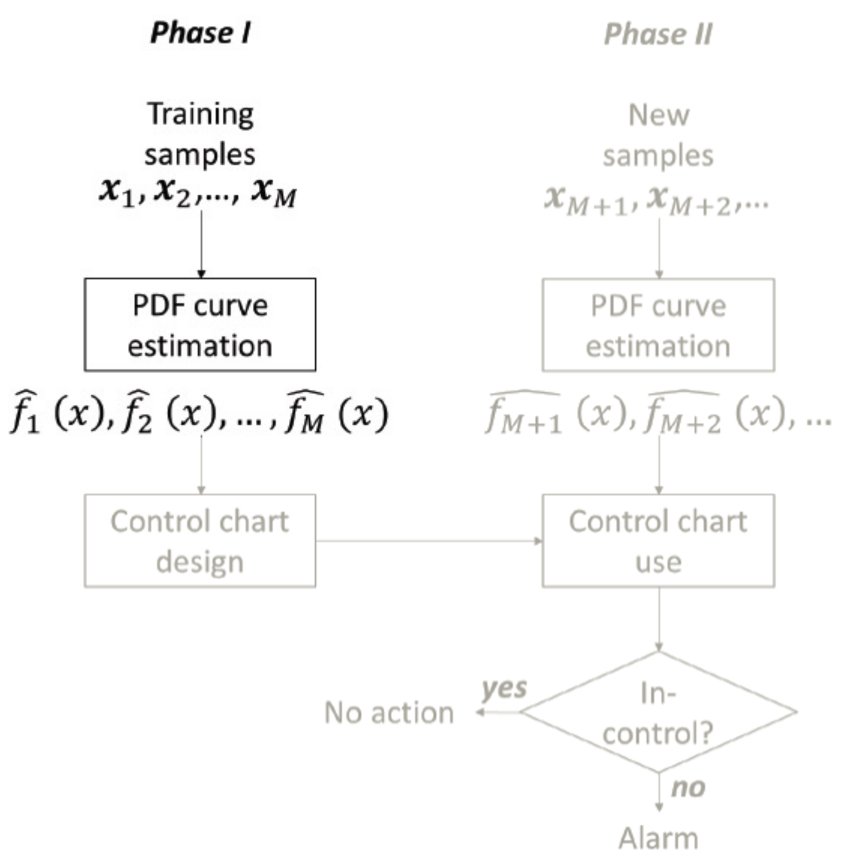


Hron et al (2016)

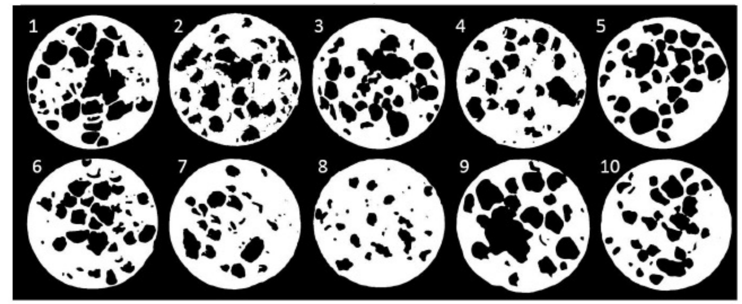
The problem of monitoring PDFs in Bayes spaces



The problem of monitoring PDFs in Bayes spaces



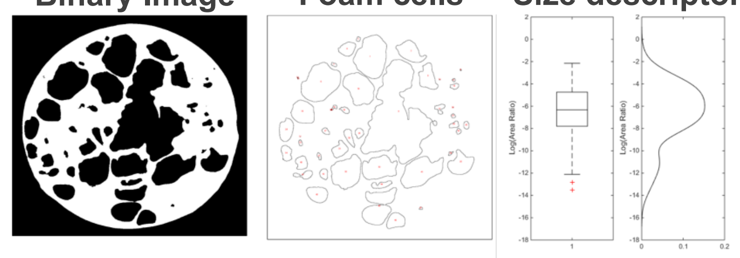
Training set of sections (Phase I)



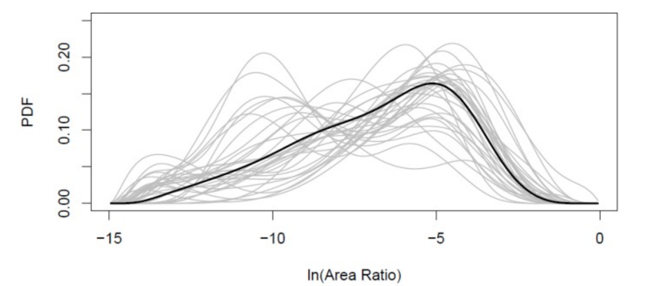
Binary Image

Foam cells

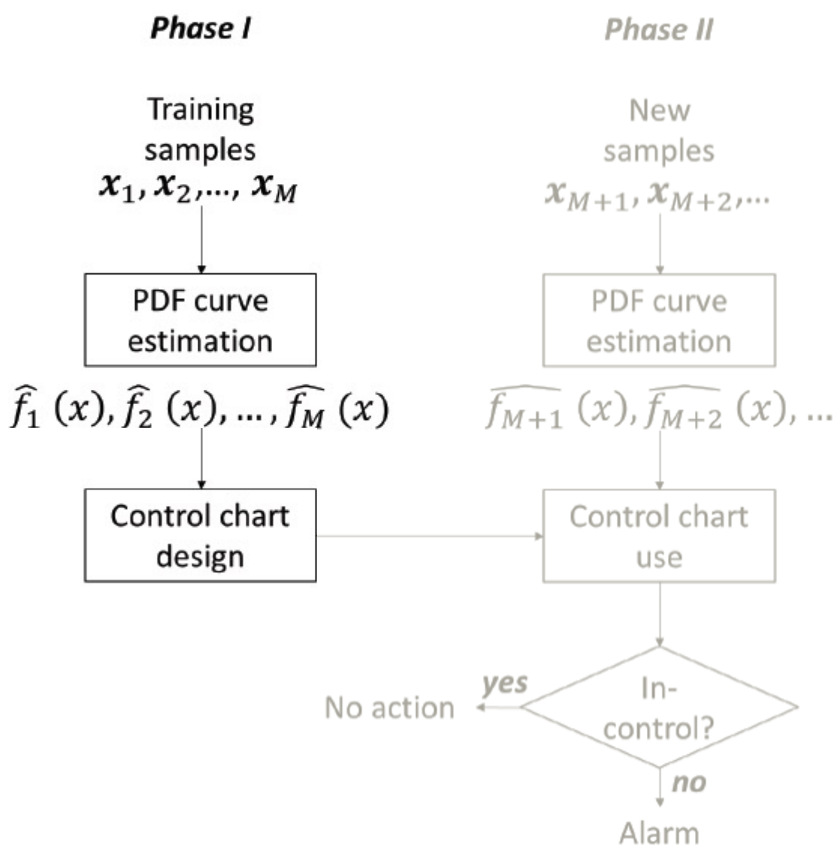
Size descriptor



PDFs of size descriptors in all the sections



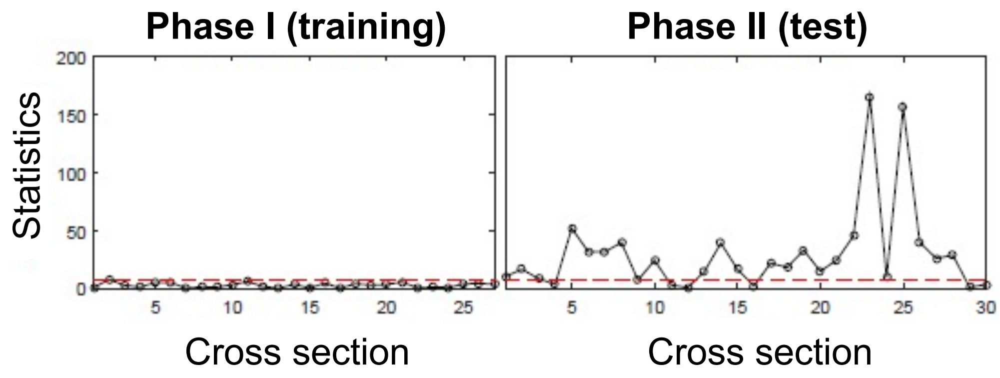
The problem of monitoring PDFs in Bayes spaces



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 (SFPCA in Bayes spaces)
    - Build **multivariate control charts with probabilistic limit,** based on the dataset of reduced dimensionality

Recall: 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

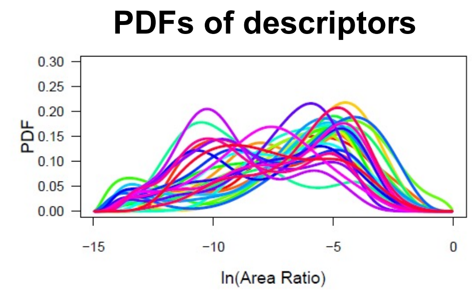
$$\sum_{j=1}^{M} \langle \widehat{f}_j \ominus \overline{f},\ \zeta \rangle^2 \quad \text{subject to} \quad \|\zeta\| = 1,\ \langle \zeta, \zeta_i \rangle = 0, i < j.$$

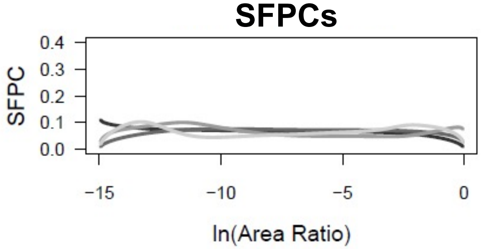
  Note. *Same problem as usual PCA, but in a different space*

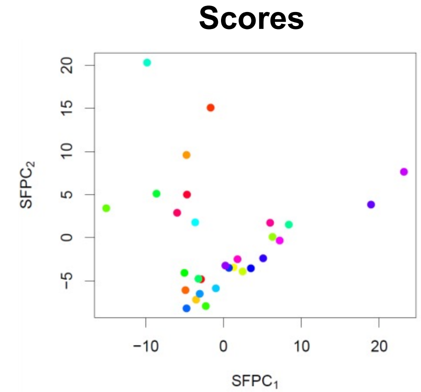
- **Scores:** Projection of the data along the SFPCs

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

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





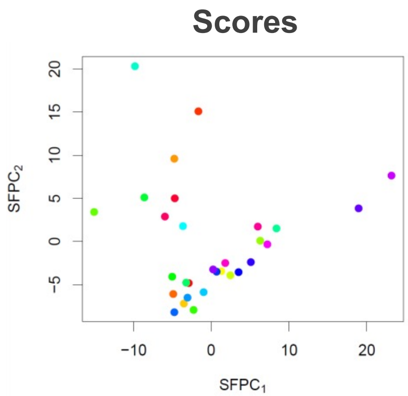


Control charts in Bayes spaces via SFPCA

Hotelling T^2 control chart on the scores

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

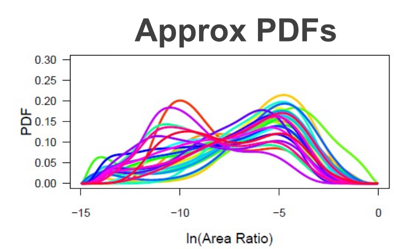
to control deviations along the first *K* SFPCs

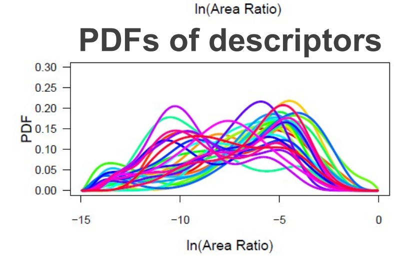


Residuals' control chart

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

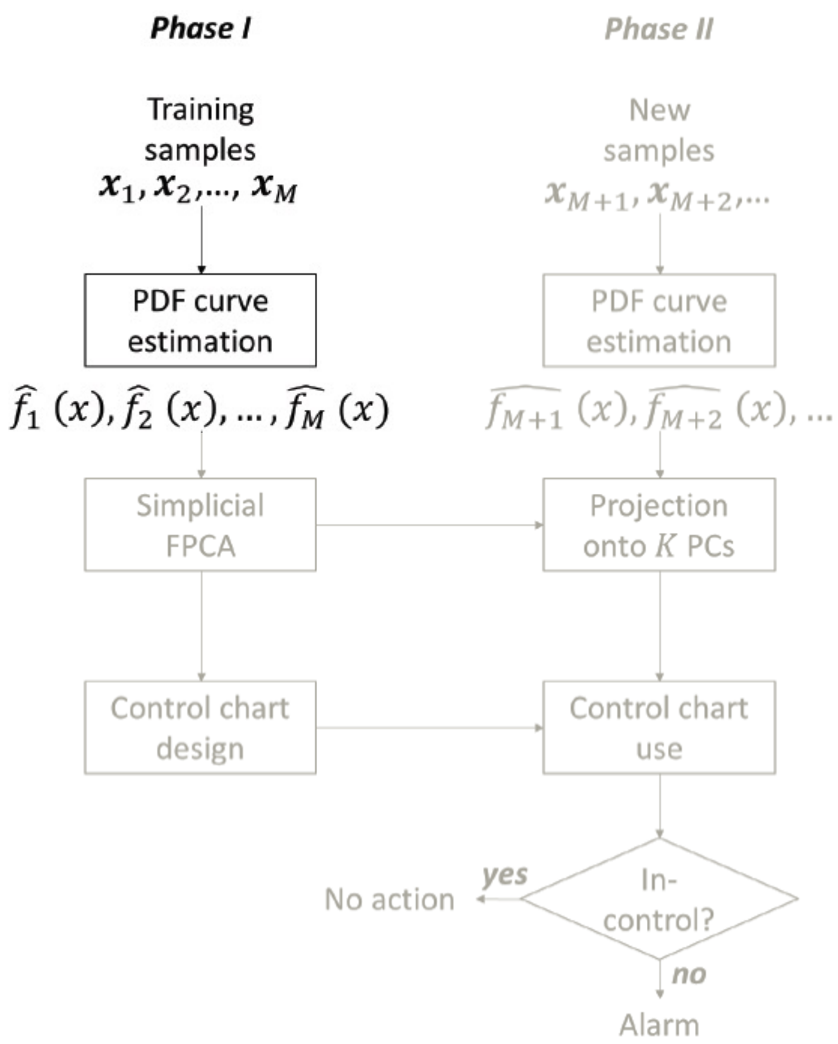
to detect shifts along directions orthogonal to the ones associated with the first *K* SFPCs

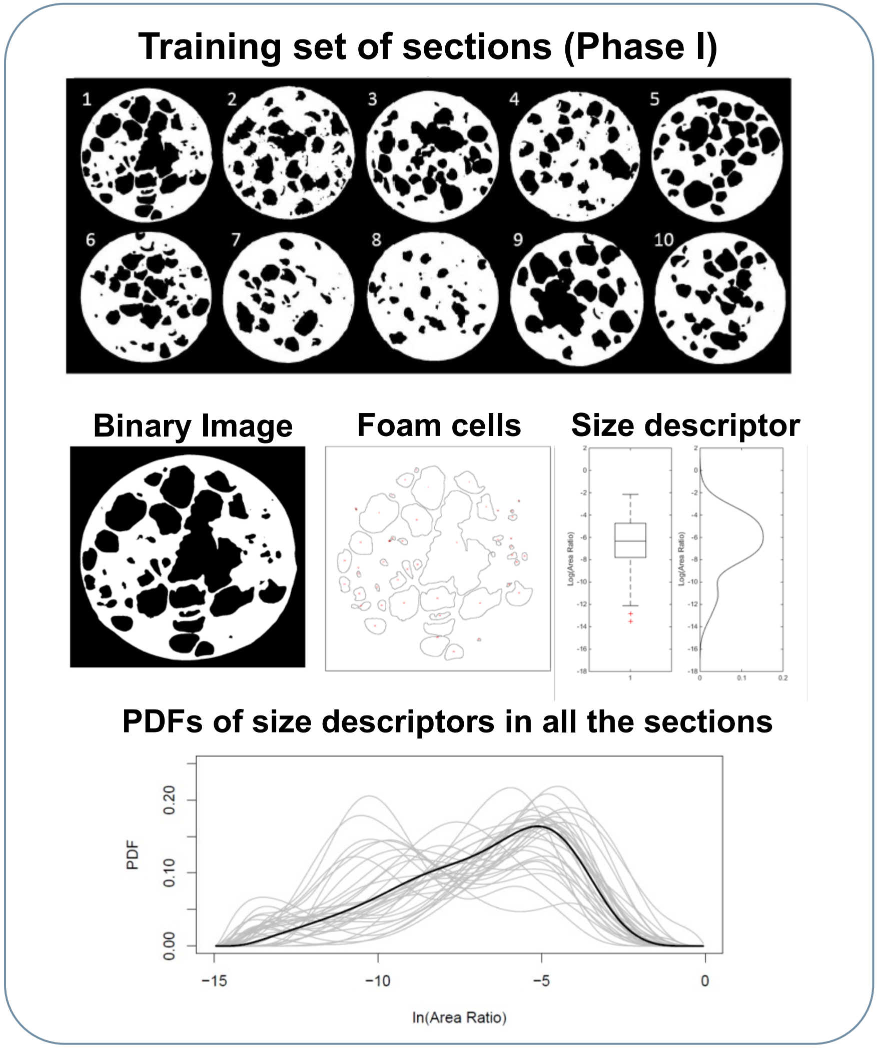




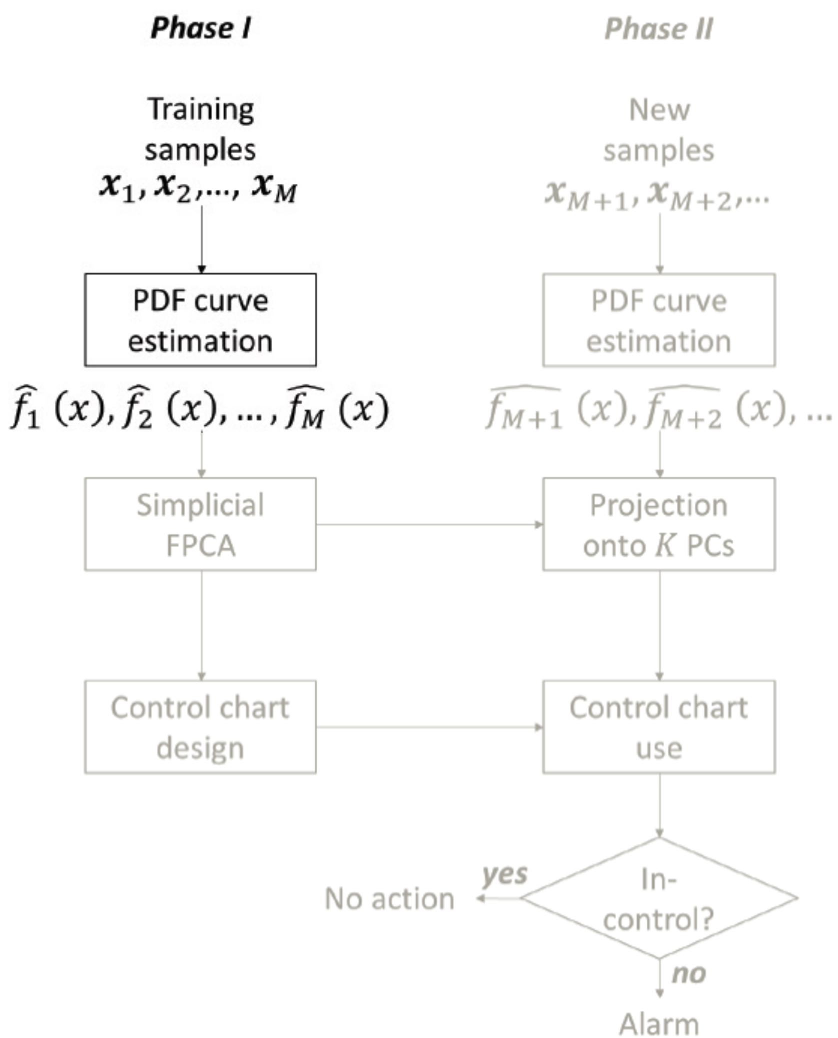
Control limits are found as empirical quantiles of the statistics on the training set

Profile monitoring for PDFs





Profile monitoring for PDFs



Bayes space geometry

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

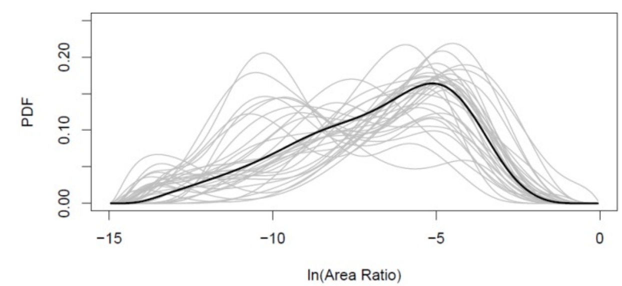
$$(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s)\,ds}$$

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

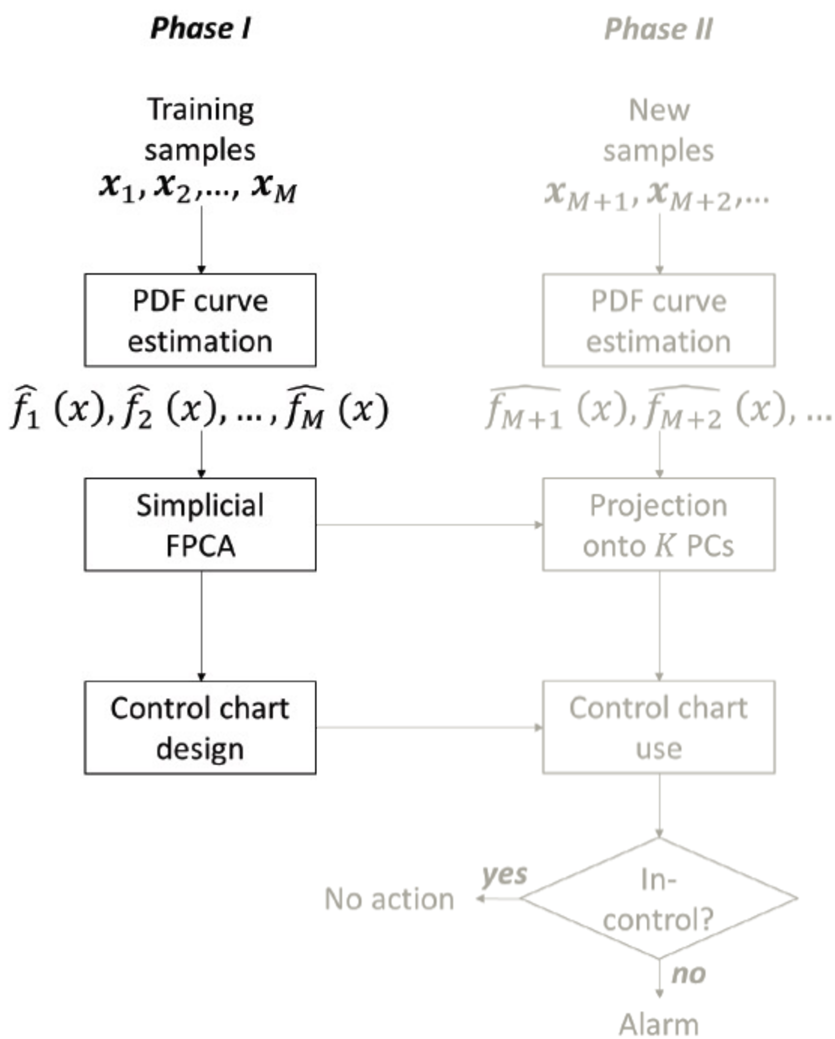
$$\langle f, g \rangle_{\mathcal{B}} = \frac{1}{2\eta} \int_I \int_I \ln\frac{f(t)}{f(s)} \ln\frac{g(t)}{g(s)}\,dt\,ds$$

Bayes space embedding

PDFs of size descriptors in all the sections



Profile monitoring for PDFs

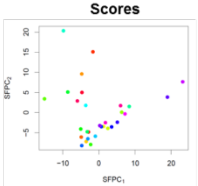


Hotelling T^2 control chart on the scores

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

to control deviations along the first K SFPCs

Scores

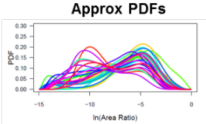


Residuals' control chart

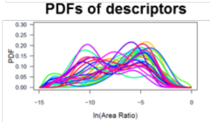
$$SPE_j(K) = \|\hat{f}_j^* \ominus \hat{f}_j\|^2 \quad j = 1, 2, \dots,$$

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

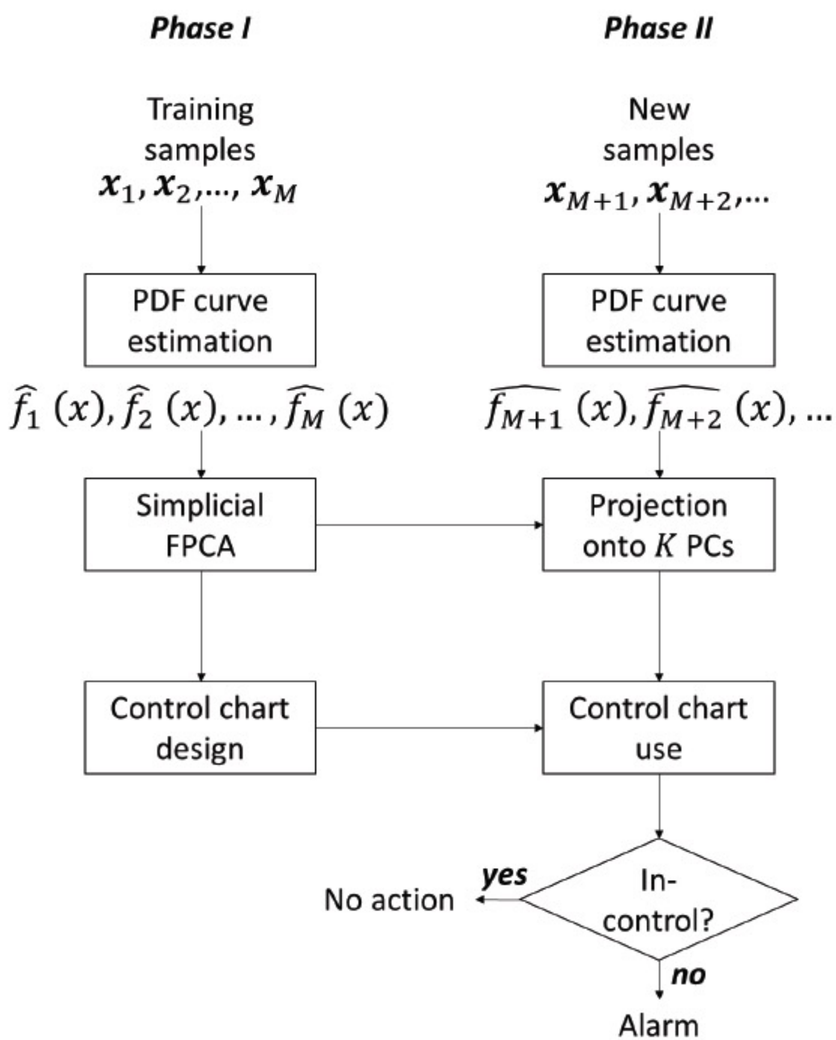
Approx PDFs



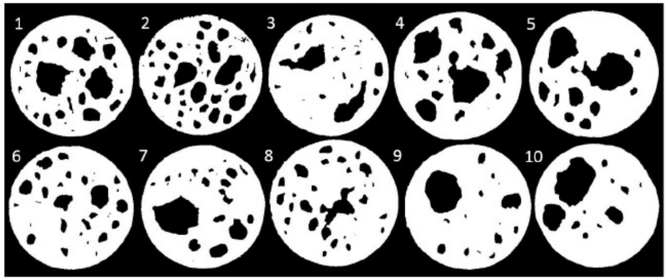
PDFs of descriptors



Profile monitoring for PDFs



Application to Phase II sections



PDFs of size descriptors in Phase II sections

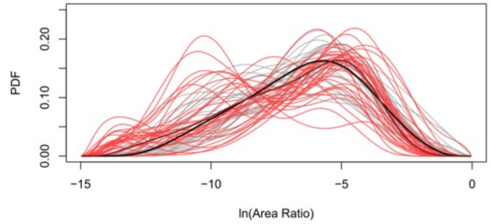


Table of contents

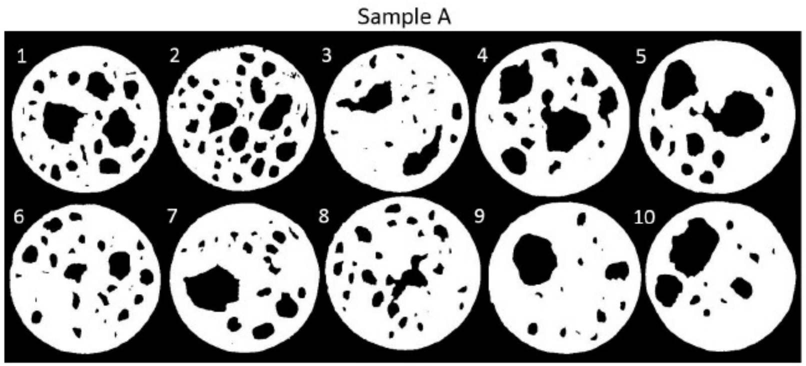
ANALYSIS OF METAL FOAM DATA

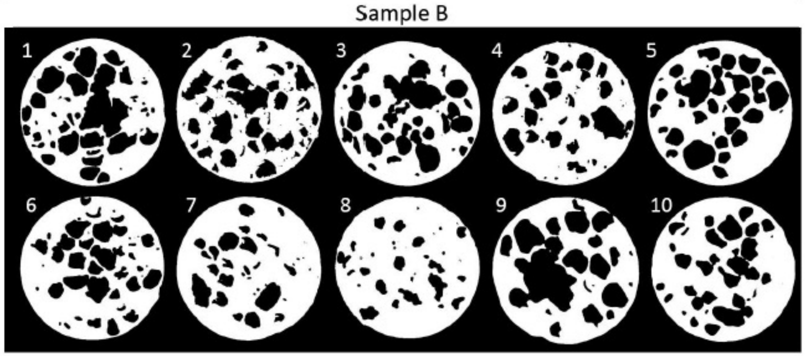
Quality control of foamed materials production

- 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

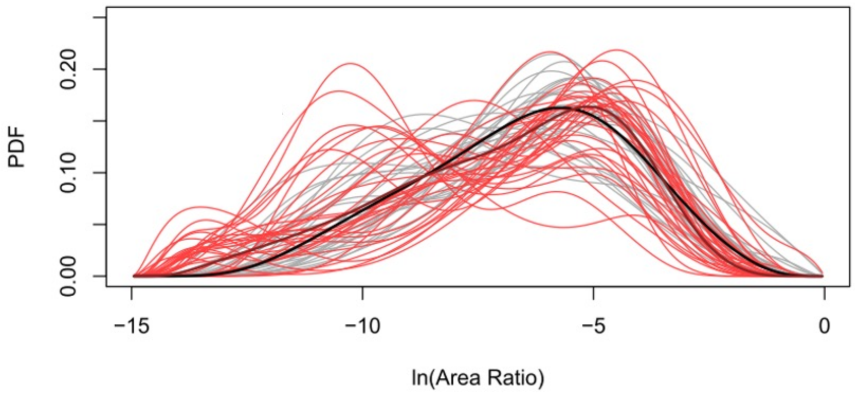
$$A_r(i,j) = \frac{A_i(j)}{A_{tot}(j)}, \quad i = 1, ..., N_j;\ j = 1, 2, ...$$

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





Results on metal foam data

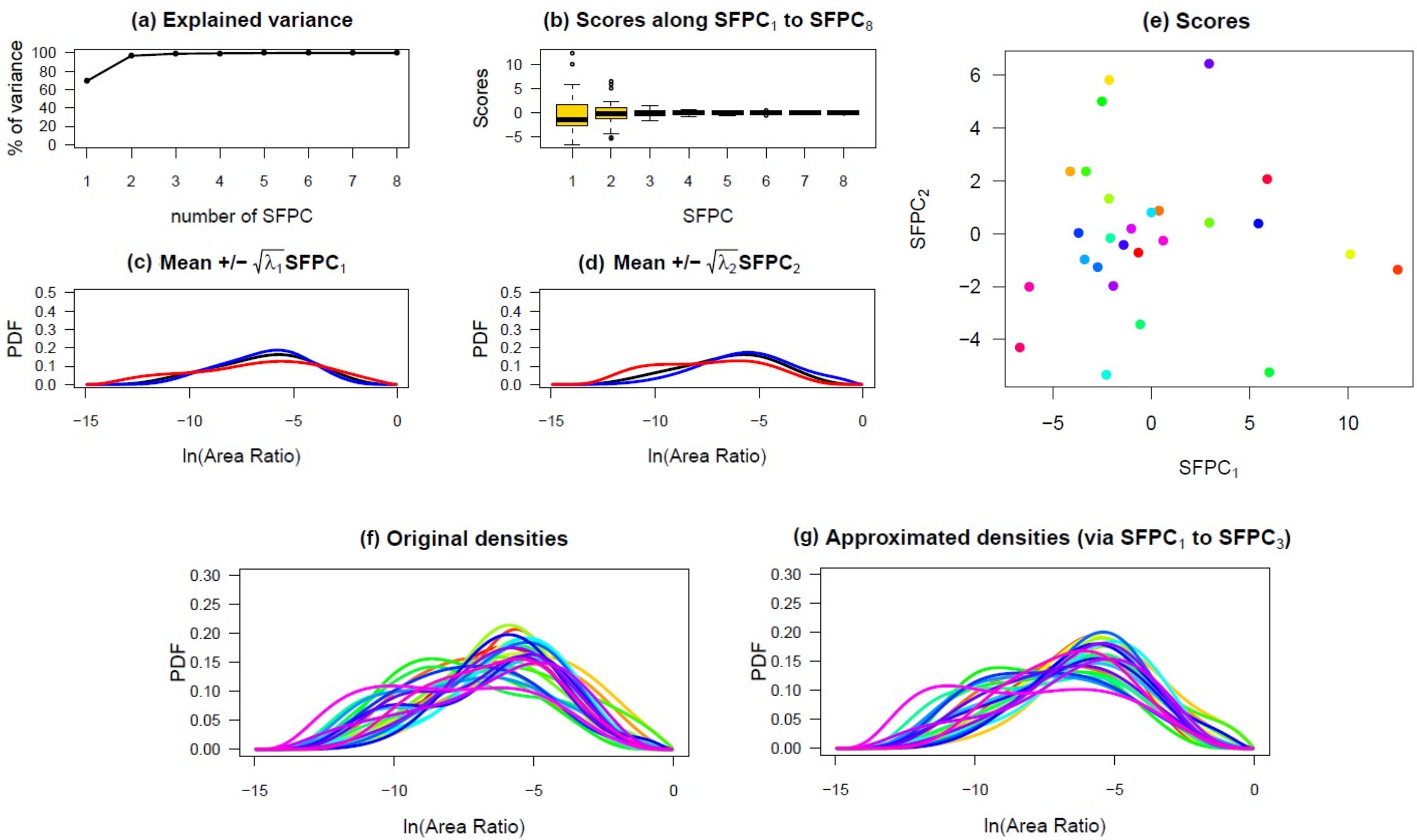


- SFPCA on the Phase I dataset; retained $K=3$ SFPCs (98% of the variability)

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

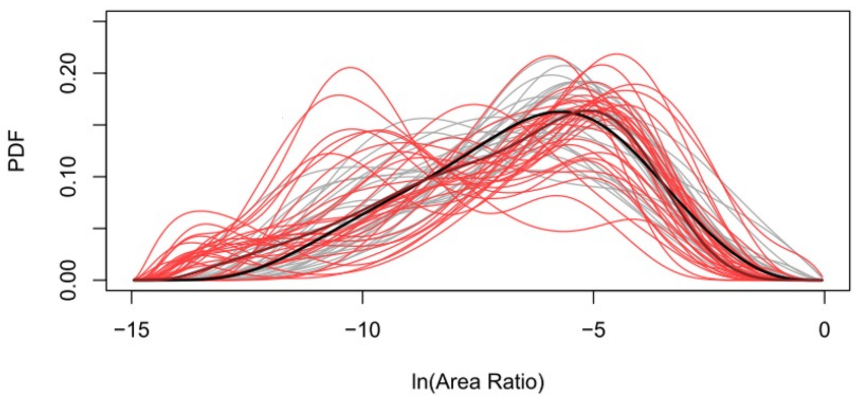
Results on metal foam data

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



Results on metal foam data

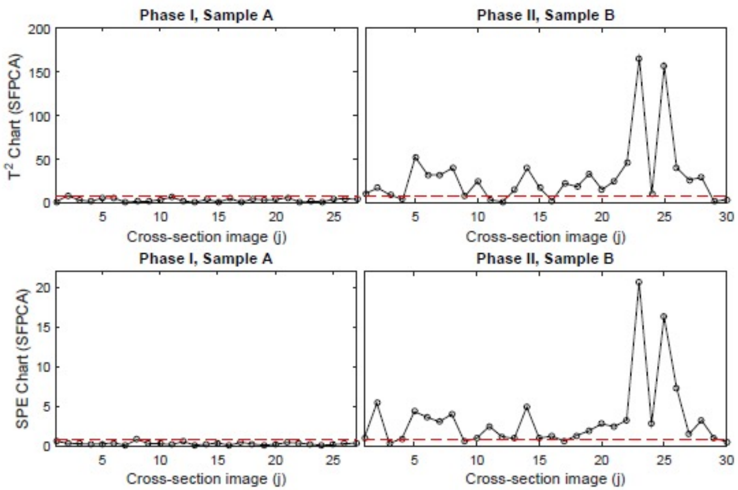
Menafoglio, Grasso, Secchi, Colosimo (Technometrics, 2018)



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

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

Results on metal foam data

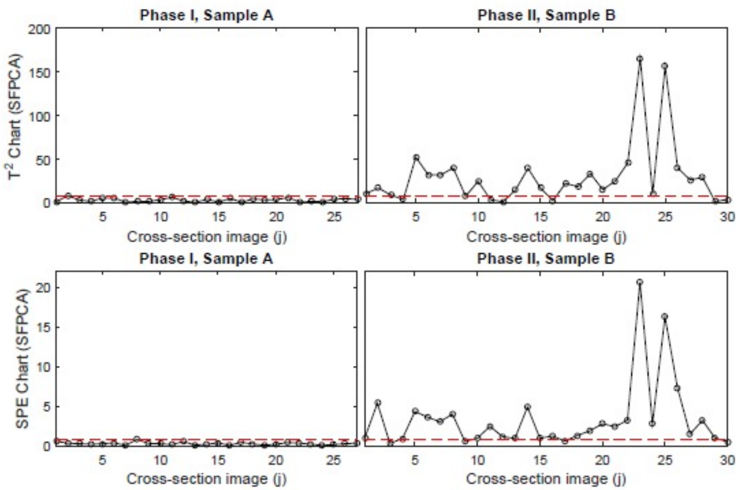


PDF monitoring in Bayes space

- SFPCA on the Phase I dataset; retained K=3 SFPCs (98% of the variability)
- Control charts for
  - the scores along the SFPCs
  - the residuals
- Control chart applied to the Phase II dataset

Results on metal foam data

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



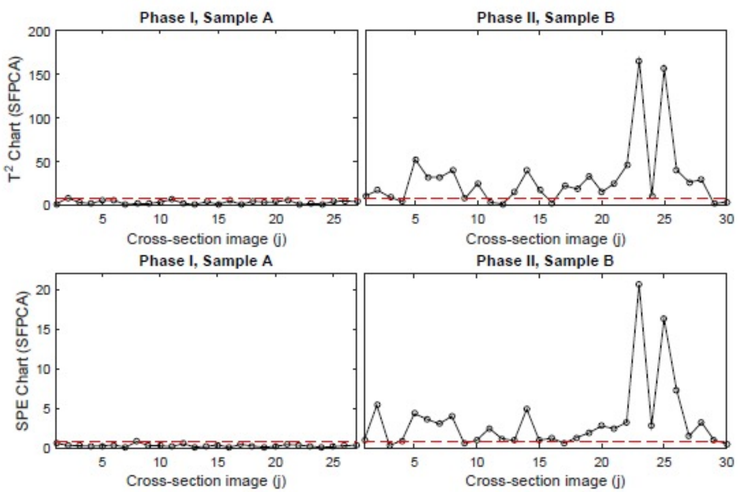
Method	Type II error
SFPCA	0.1
FPCA	0.1333
Q-Q plot	0.5333
Shewhart	0.4

PDF monitoring in Bayes space

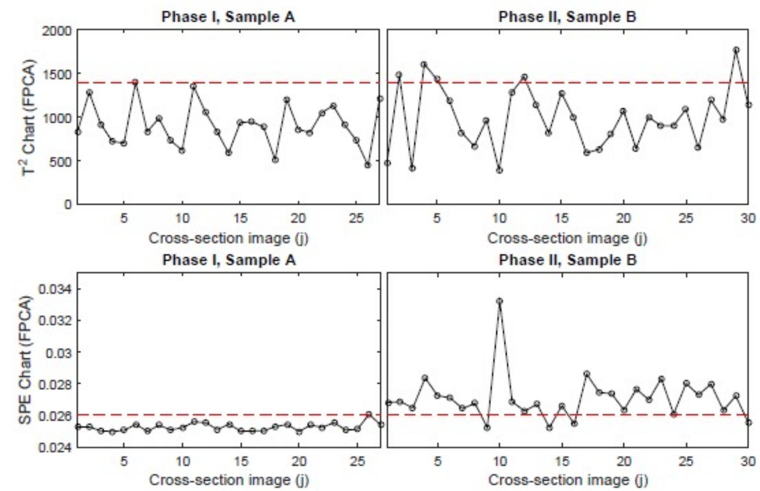
Simulations confirmed that PDF monitoring in Bayes spaces outperforms competitor methods in terms of type II error and ARL

Results on metal foam data

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



PDF monitoring in Bayes space

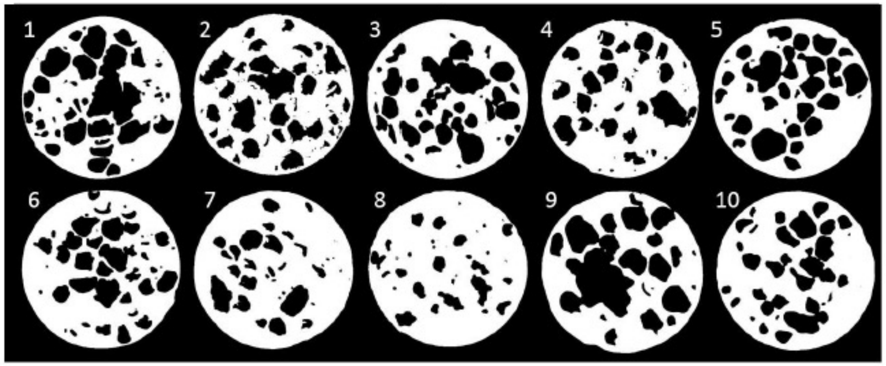


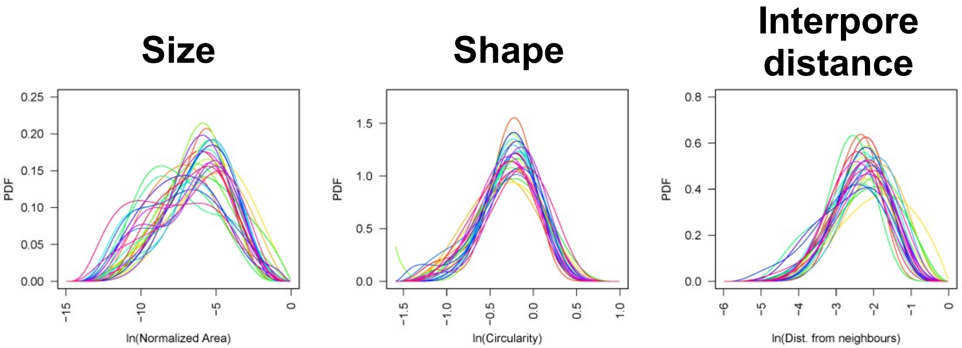
PDF monitoring in L^2

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

Take home messages

- 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





CONCLUSIONS

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.

Codes available on GitHub: github.com/AMenafoglio/BayesSpaces-codes

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**.
- **Scimone, R., Taormina, T., Colosimo, B.M., Grasso, M., Menafoglio, A., Secchi P. (2021). Statistical modeling and monitoring of geometrical deviations in complex shapes with application to Additive Manufacturing. MOX-report 23/2021.**
- 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.

Anomaly detection for density data based on control charts

Alessandra Menafoglio[1*]

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

*alessandra.menafoglio@polimi.it