L'impatto dei metodi statistici per dati funzionali in ambito Industria 4.0

Christian Capezza
Dipartimento di Ingegneria Industriale
Università degli Studi di Napoli Federico II

Statistica e Data Science per il Business, 3 Maggio 2024, Padova - Palazzo del Bo - Aula Magna

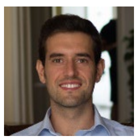
Agenda

- **The Statistics for Engineering Research (SFERe) group**
- Main industrial scenarios
- An example: our research in the automotive industry
- Technological transfer
- Conclusions

The SFERe group (sfere.unina.it)

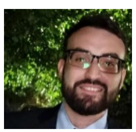
Biagio Palumbo: Associate Professor (Coordinator),
President of European Network for Business and Industrial Statistics (ENBIS)

Antonio Lepore: Associate Professor



Christian Capezza: Assistant Professor (RTD/A)

Fabio Centofanti: Assistant Professor (RTD/A)



Gianluca Sposito: PhD Candidate

Emanuele Rossi: PhD Student

Davide Forcina: PhD Student

The SFERe group: main research areas

We develop methods in these topics:

- Functional Data Analysis
- Interpretable Statistical Models
- Statistical Process Monitoring/Control

The main driver is… the industrial application!

Our role as Statisticians in a Dept. of Eng.

- **Gather problems** from industries
- **Select** those that may potentially involve impactful data analysis
- Identify the **broader industrial need** and the **statistical literature gap** to possibly fill in
- **Deal with real data**, big and time-consuming challenge
- **Develop new methods and models** from a general perspective to the extent of being inspired and not limited to the specific problem investigated
- Make the solution **accessible** to the industrial practitioner

Agenda

- The SFERe group
- **Main industrial scenarios**
- An example: our research in the automotive industry
- Technological transfer
- Conclusions

Industry 4.0: key enabling technologies

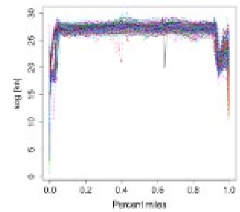
… advanced statistical methodologies for process monitoring and predictive maintenance in the context of large amounts of complex data acquired by multi-sensor systems …

Shipping: CO_2 emission monitoring

- Capezza, C., Centofanti, F., Lepore, A., Menafoglio, A., Palumbo, B., Vantini, S. (2022) Functional regression control chart for monitoring ship CO_2 emissions. ***Quality and Reliability Engineering International***, 38(3), 1519-1537.
- Capezza C., Lepore A., Menafoglio A., Palumbo B. Vantini S. (2020) Control charts for monitoring ship operating conditions and CO_2 emissions based on scalar-on-function regression. ***Applied Stochastic Models in Business and Industry***, 36(3):477–500.
- Reis, M. S., Rendall, R., Palumbo, B., Lepore, A., Capezza, C. (2020) Predicting ships' CO_2 emissions using feature-oriented methods. ***Applied Stochastic Models in Business and Industry***, 36(1):110-123.
- Capezza C., Coleman S., Lepore A., Palumbo B. Vitiello L. (2019) Ship fuel consumption monitoring and fault detection via partial least squares and control charts of navigation data. ***Transportation Research Part D: Transport and Environment***, 67:375-387.
- Lepore, A., Palumbo, B., Capezza, C. (2019) Orthogonal LS-PLS approach to ship fuel-speed curves for supporting decisions based on operational data. ***Quality Engineering***, 31(3):386–400.
- Lepore, A., Palumbo, B., Capezza, C. (2018) Analysis of profiles for monitoring of modern ship performance via partial least-squares methods. ***Quality and Reliability Engineering International***, 34:1424-1436.
- Lepore, A., Reis, M.S., Palumbo, B., Rendall, R., Capezza, C. (2017) A comparison of advanced regression techniques for predicting ship CO_2 emissions. ***Quality and Reliability Engineering International***, 33:1281–1292.

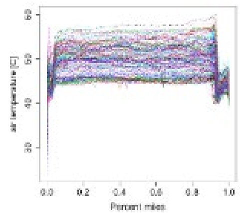
Shipping: CO_2 emission monitoring

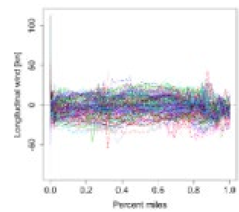
SCALAR RESPONSE: Y_i: total CO_2 emissions ($i = 1, \dots, N$)



FUNCTIONAL COVARIATES:

1. Speed over ground (SOG), [*kn*]
2. Acceleration (SOG derivative), [*kn/h*]
3. Power difference between propeller shafts, [*kW*]
4. Distance from the nominal route, [*NM*]
5. Longitudinal wind, [*kn*]
6. Side wind, [*kn*]
7. Air temperature, mean of four engines, [°*C*]
8. Cumulative sailing time, [*h*]
9. Trim, [*m*]





$$\widetilde{\boldsymbol{X}}(t) = \left(\tilde{X}_1(t), \dots, \tilde{X}_P(t)\right)$$

is a random element whose N realizations are
$\widetilde{\boldsymbol{X}}_i(t) = \left(\tilde{X}_{i1}(t), \dots, \tilde{X}_{iP}(t)\right)$ where $\tilde{X}_{ip} \in L^2(\mathcal{T})$, for all $p = 1, \dots, P$ and $i = 1, \dots, N$.

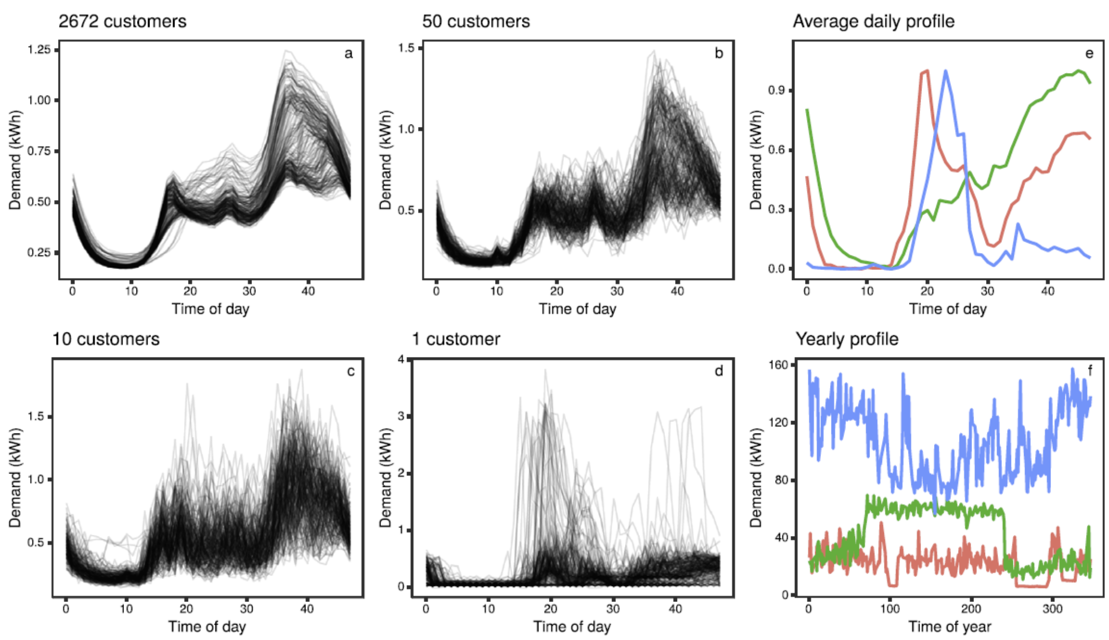
…

$\widetilde{\boldsymbol{X}}_i$ belongs to the Hilbert space $\mathbb{H}$ of the vectors whose elements are vectors of $L^2(\mathcal{T})$, with inner product $\left\langle \widetilde{\boldsymbol{X}}_{i_1}, \widetilde{\boldsymbol{X}}_{i_2} \right\rangle_{\mathbb{H}} = \sum_{p=1}^{P} \left\langle \tilde{X}_{i_1 p}, \tilde{X}_{i_2 p} \right\rangle = \sum_{p=1}^{P} \int \tilde{X}_{i_1 p}(t) \tilde{X}_{i_2 p}(t) dt$

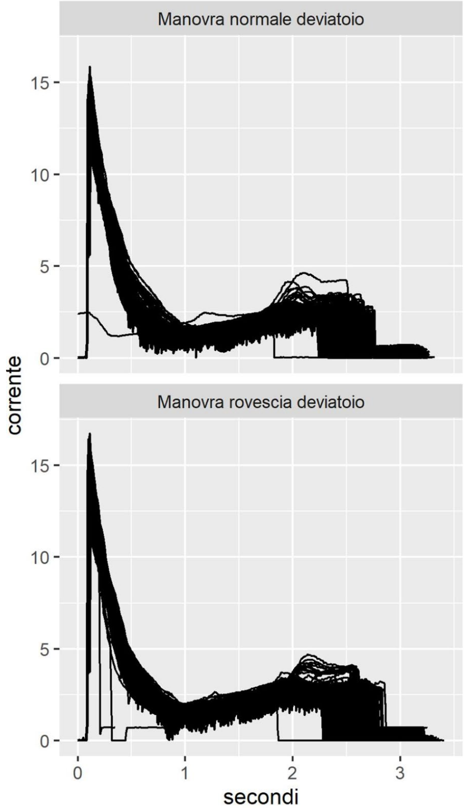
Capezza, C., Lepore, A., Menafoglio, A., Palumbo, B. & Vantini, S. (2020+), 'Control charts for monitoring ship operating conditions and CO_2 emissions based on scalar-on-function regression'. *Applied Stochastic Models in Business and Industry*.

Electricity Demand Forecasting

- Ongoing research activity with Matteo Fasiolo and Euan Enticott (PhD student at the School of Mathematics of the *University of Bristol*)
- Capezza, C., Palumbo, B., Goude Y., Wood, S.N., Fasiolo, M. (2021) Additive Stacking for Disaggregate Electricity Demand Forecasting. ***The Annals of Applied Statistics***, 15(2):727-746



Railway: Railroad switch monitoring



Manufacturing: automotive industry

- Ongoing research activity with Kamran Paynabar, *Georgia Tech*, Stream-Based Active Learning for Anomaly Detection.
- Capezza, C., Capizzi, G., Centofanti, F., Lepore, A., Palumbo, B. (2024+) An Adaptive Multivariate Functional EWMA Control Chart. **arXiv:2403.03837**.
- Capezza, C., Centofanti, F., Lepore, A., Palumbo, B. (2024) Robust Multivariate Functional Control Chart. ***Technometrics***.
- Capezza, C., Centofanti, F., Lepore, A., Palumbo, B. (2021) Functional clustering methods for resistance spot welding process data in the automotive industry. ***Applied Stochastic Models in Business and Industry***, 37(5):908-925.

Many problems have similarities…

- In all these problems, the statistical unit of interest is one or multiple profiles.
- One of the most common questions in industrial applications regard the prompt identification of anomalies.
- The generalizability of a solution to several industrial problems becomes the main driver of our research.

Data complexity

Statistical methods deal with the increasing complexity of the available data:

- **univariate, scalar**
- **multivariate, scalar**
- **multivariate, profiles**
  at each voyage and for each variable, multiple measurements are acquired over time and profiles are available (functional data)

In this context, two approaches are possible:

→ Simplify by extracting features from each profile

→ Deal with the complexity of the data and treat curves as single objects to analyze (functional data analysis)

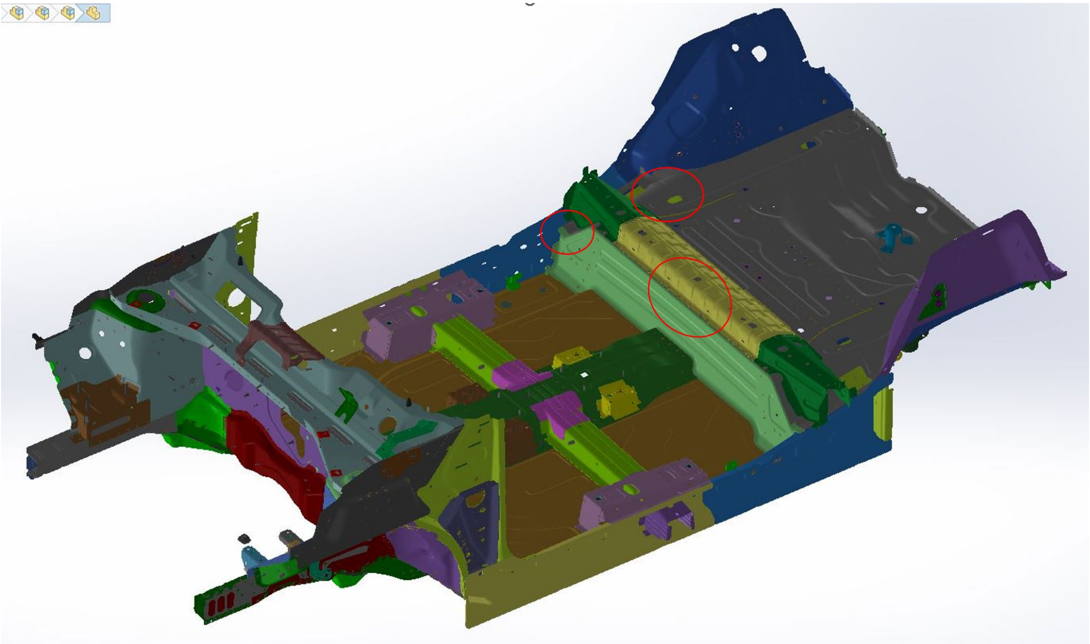
INCREASING DATA COMPLEXITY

Agenda

- The Statistics for Engineering Research (SFERe) group
- Main industrial scenarios
- **An example: our research in the automotive industry**
- Technological transfer
- Conclusions

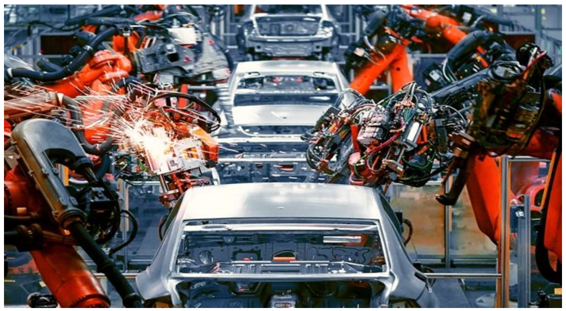
Automotive Industry

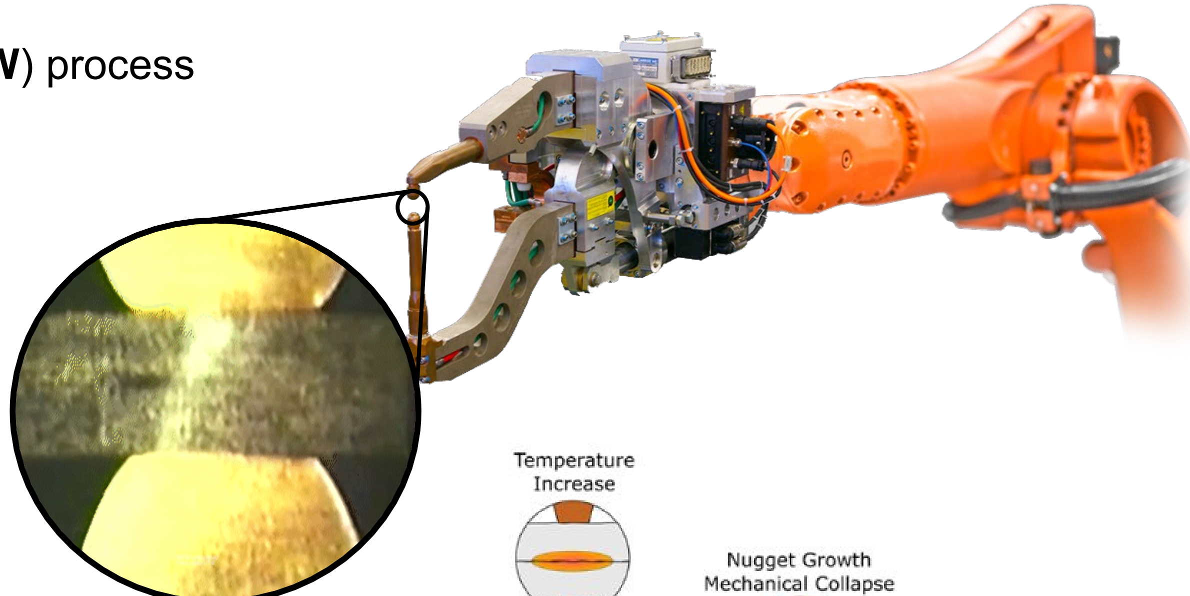
- Assembly of car bodies in the automotive industry
- Resistance spot welding (**RSW**) process
- **Many points** are welded for each car body



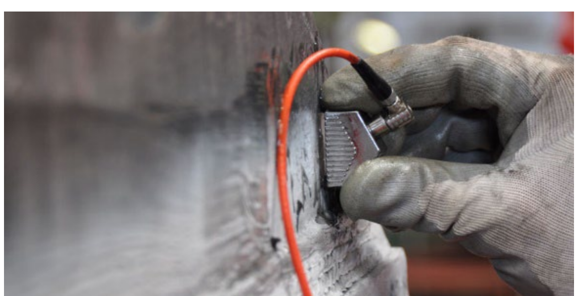
Automotive Industry

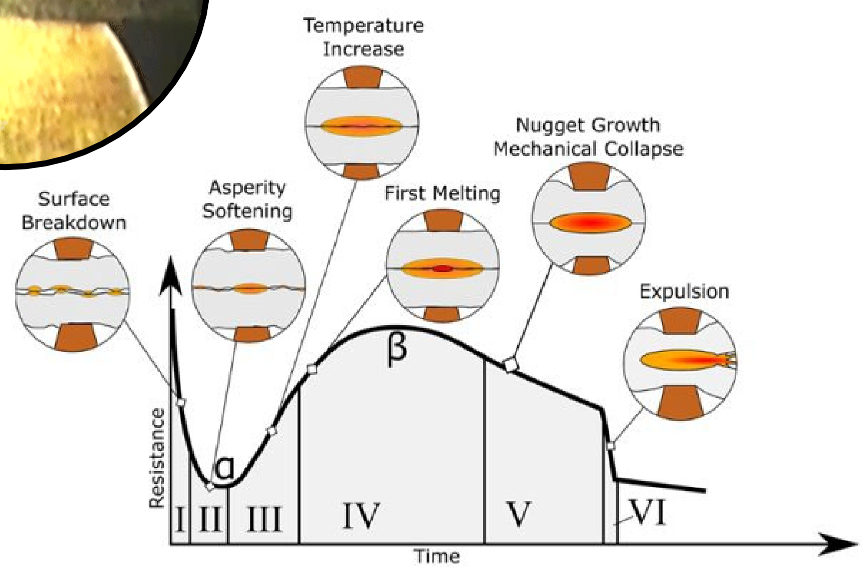
- Resistance spot welding (**RSW**) process





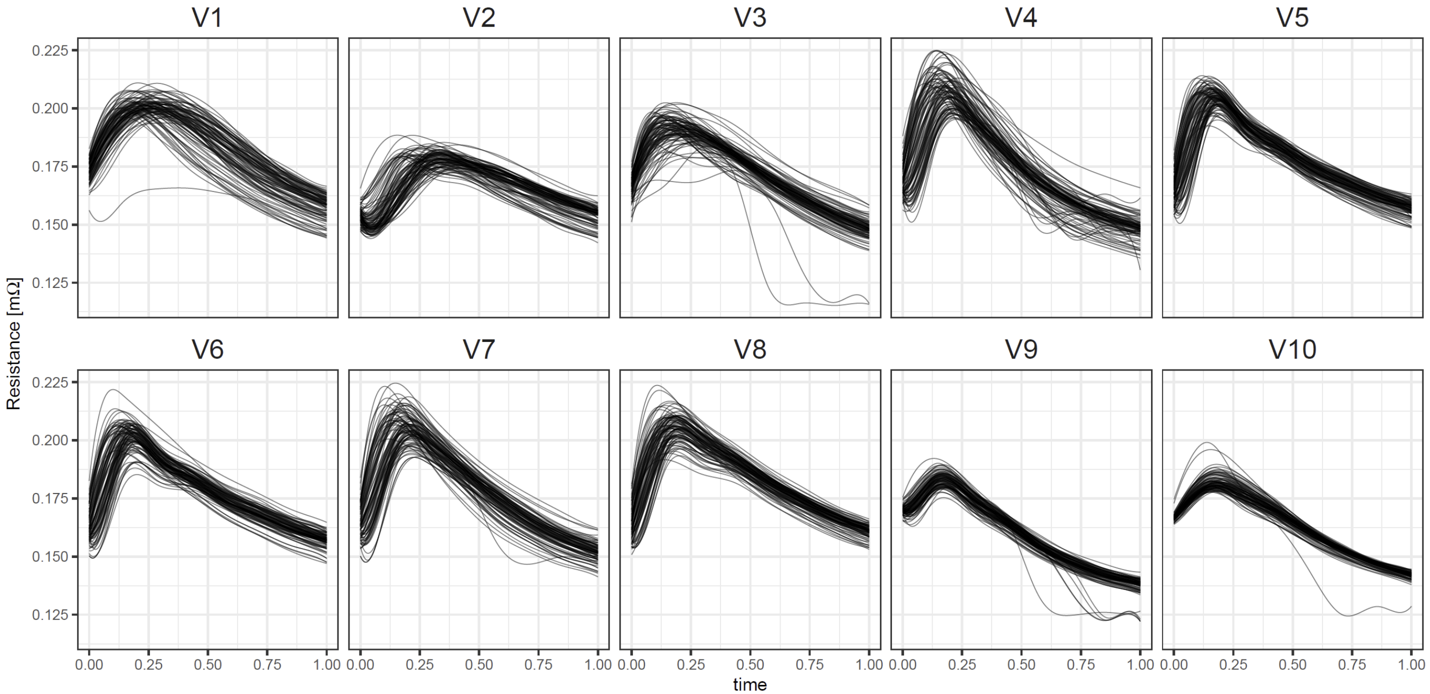
- Technological signature of the process: the dynamic resistance curve (**DRC**)





Automotive Industry

- Focus on the set of 10 points made by one robot arm on one car body.
- Profiles corresponding to spot welds made immediately after electrode renewal are used to form the Phase I sample.
- The remaining observations are used in Phase II to evaluate the proposed chart performance.

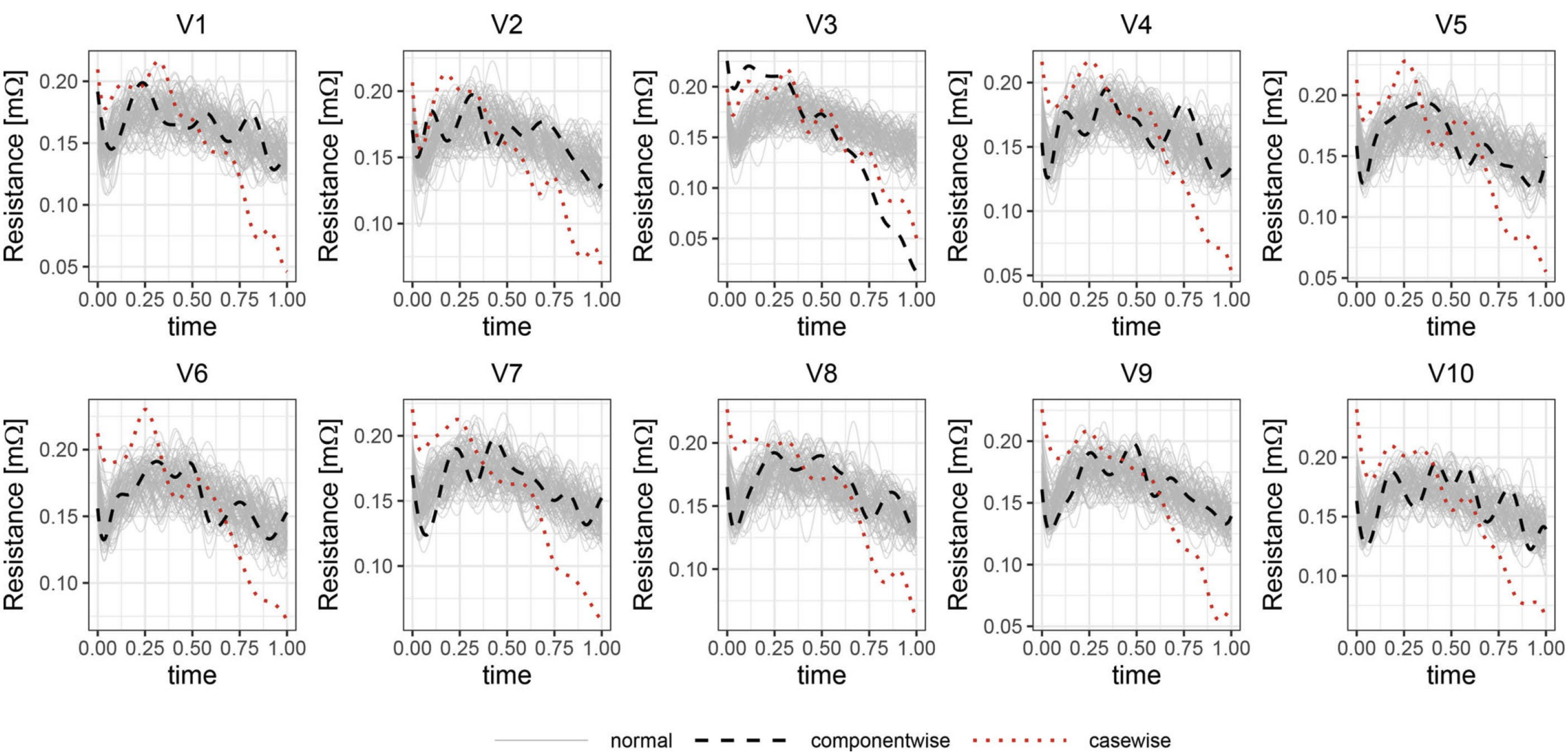


Robust Multivariate Functional Control Chart

- SPM + functional data = **profile monitoring**
- Usually, SPM is divided into
  - **Phase I**: build reference data set from in-control (IC) process and estimate control chart limits;
  - **Phase II**: check process stability on new data.
- However, the Phase I sample is often contaminated by the presence of **outliers.**
- To deal with outliers (in Phase I) in the statistical literature we roughly have
  - **diagnostic** approaches: may fail to detect more moderate outliers;
  - **robust** approaches: keep all data points and reduce the impact of outliers.
- Functional casewise vs componentwise outliers

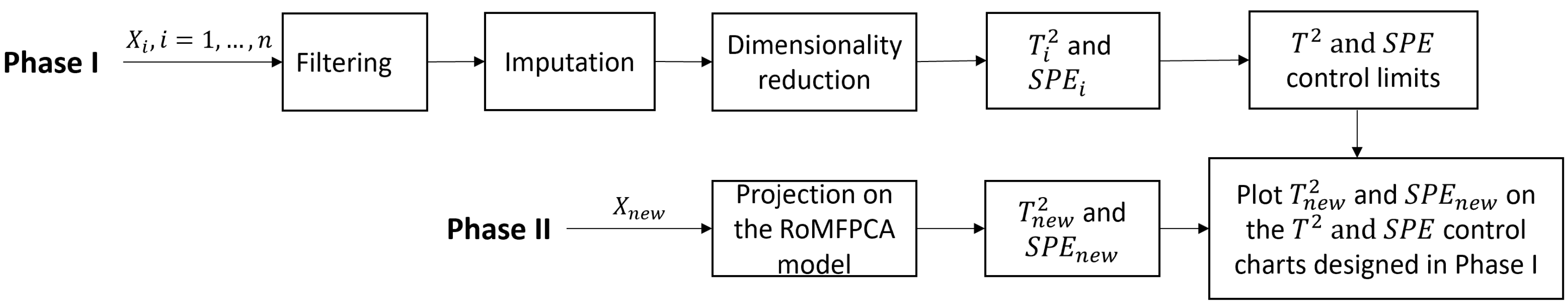
Capezza, C., Centofanti, F., Lepore, A., & Palumbo, B. (2024). Robust Multivariate Functional Control Chart. *Technometrics*

Functional Casewise and Componentwise Outliers



Capezza, C., Centofanti, F., Lepore, A., & Palumbo, B. (2024). Robust Multivariate Functional Control Chart. *Technometrics*

The Control Charting Implementation

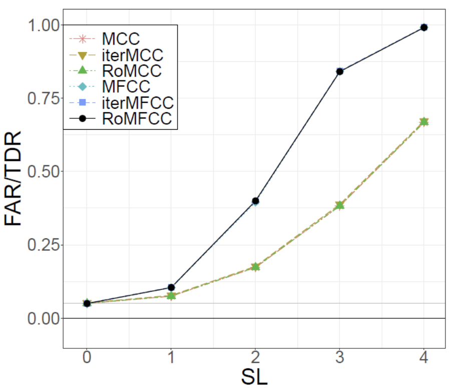


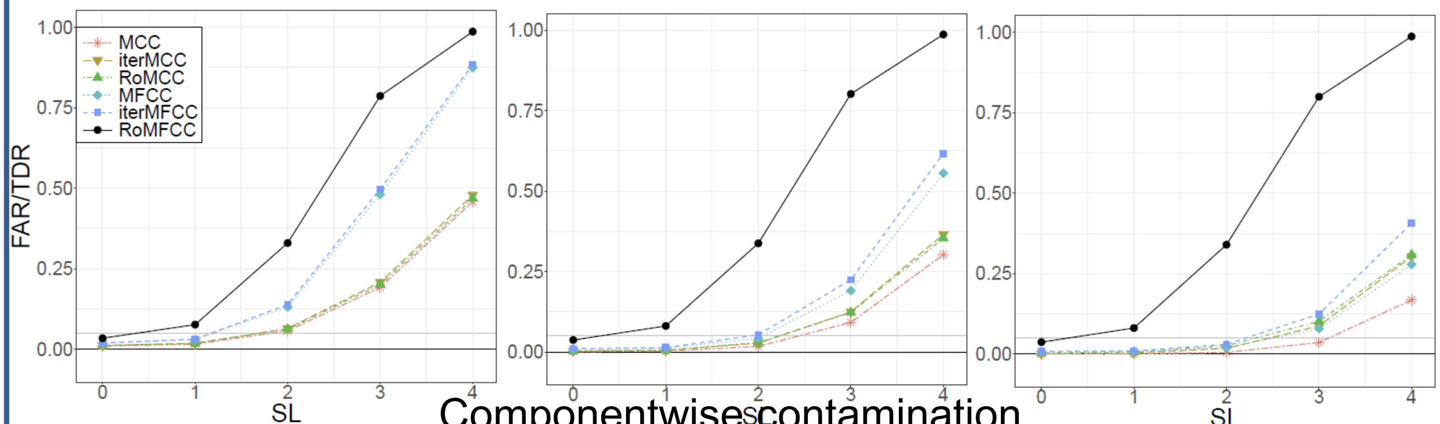
Capezza, C., Centofanti, F., Lepore, A., & Palumbo, B. (2024). Robust Multivariate Functional Control Chart. *Technometrics*

Simulation Study

Contamination level 1 | Contamination level 2 | Contamination level 3

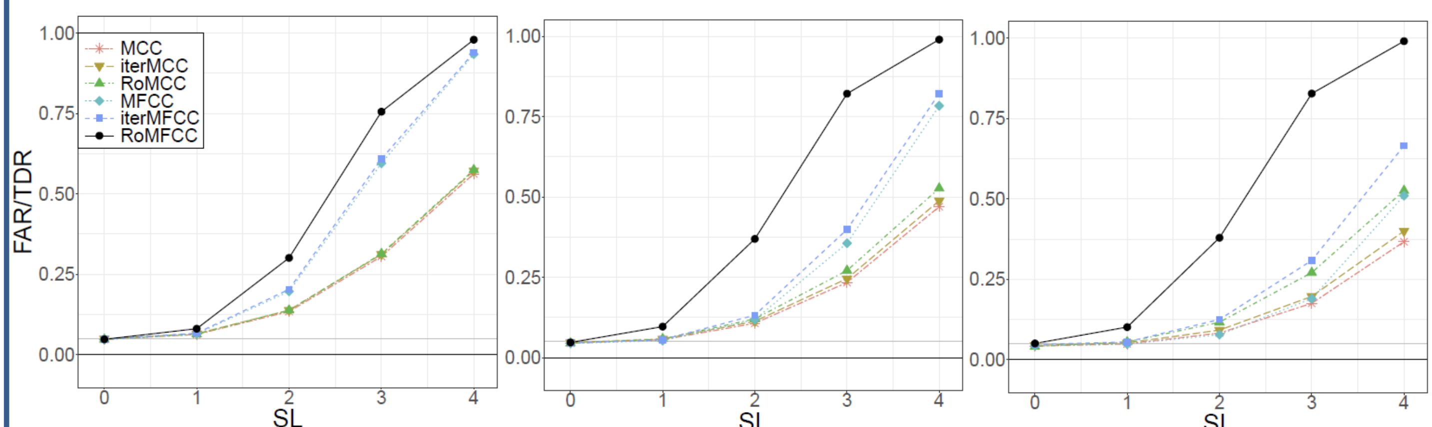
No contamination





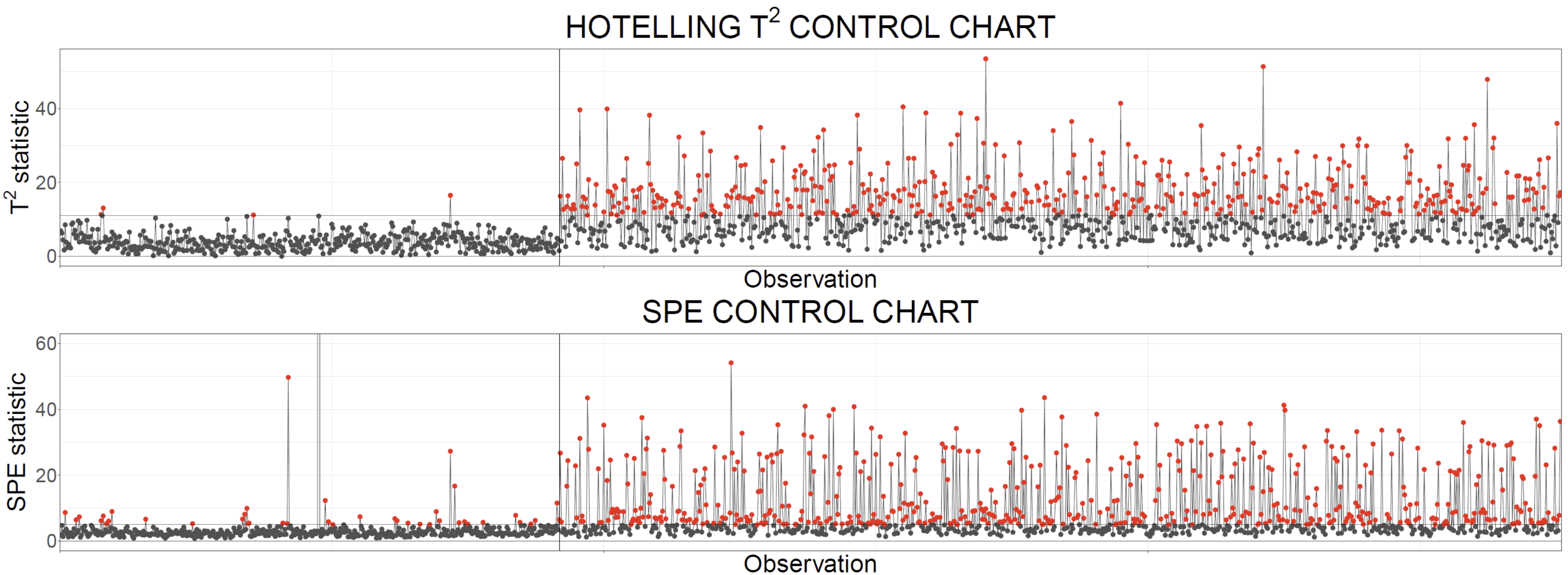
Componentwise contamination

Casewise contamination



Real-case Study

- A significant number of tuning set observations (left of vertical line) are signaled as OC, highlighted in red (expected because these points may include functional casewise outliers not filtered out by the FUF).
- In Phase II, **72.3%** of the points are signaled as OCs by the RoMFCC.



Capezza, C., Centofanti, F., Lepore, A., & Palumbo, B. (2024). Robust Multivariate Functional Control Chart. *Technometrics*

Real-case Study

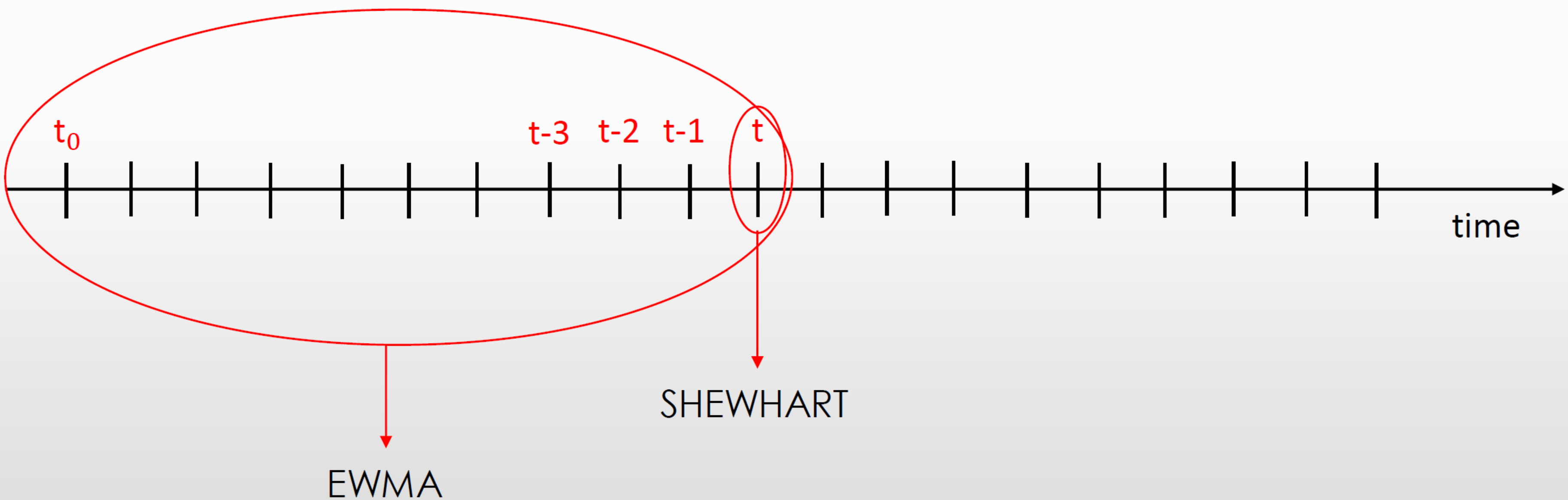
- Finally, the proposed method is compared with the competing methods presented in the simulation study (with **uncertainty** evaluation via **bootstrap**)
- RoMFCC **outperforms** the competing control charts and stands out as the best method to promptly identify OC conditions in the RWS process caused by an increased electrode wear with a Phase I sample contaminated by functional outliers.

	$\widehat{TDR}$	$\overline{TDR}$	CI
M	0.336	0.335	[0.305,0.368]
Miter	0.462	0.461	[0.428,0.496]
MRo	0.513	0.512	[0.481,0.547]
MFCC	0.541	0.541	[0.511,0.574]
MFCCiter	0.632	0.632	[0.595,0.664]
RoMFCC	0.723	0.723	[0.695,0.753]

Capezza, C., Centofanti, F., Lepore, A., & Palumbo, B. (2024). Robust Multivariate Functional Control Chart. *Technometrics*

Adaptive multivariate functional EWMA chart

- Difference between Shewhart and EWMA



Capezza, C., Capizzi, G., Centofanti, F., Lepore, A., & Palumbo, B. (2024). An Adaptive Multivariate Functional EWMA Control Chart. arXiv:2403.03837

Adaptive multivariate functional EWMA chart

$$Y_n = \lambda X_n + (1-\lambda) Y_{n-1}$$

- The EWMA statistic Y_n is a ***weighted average*** of the observation X_n and Y_{n-1} .

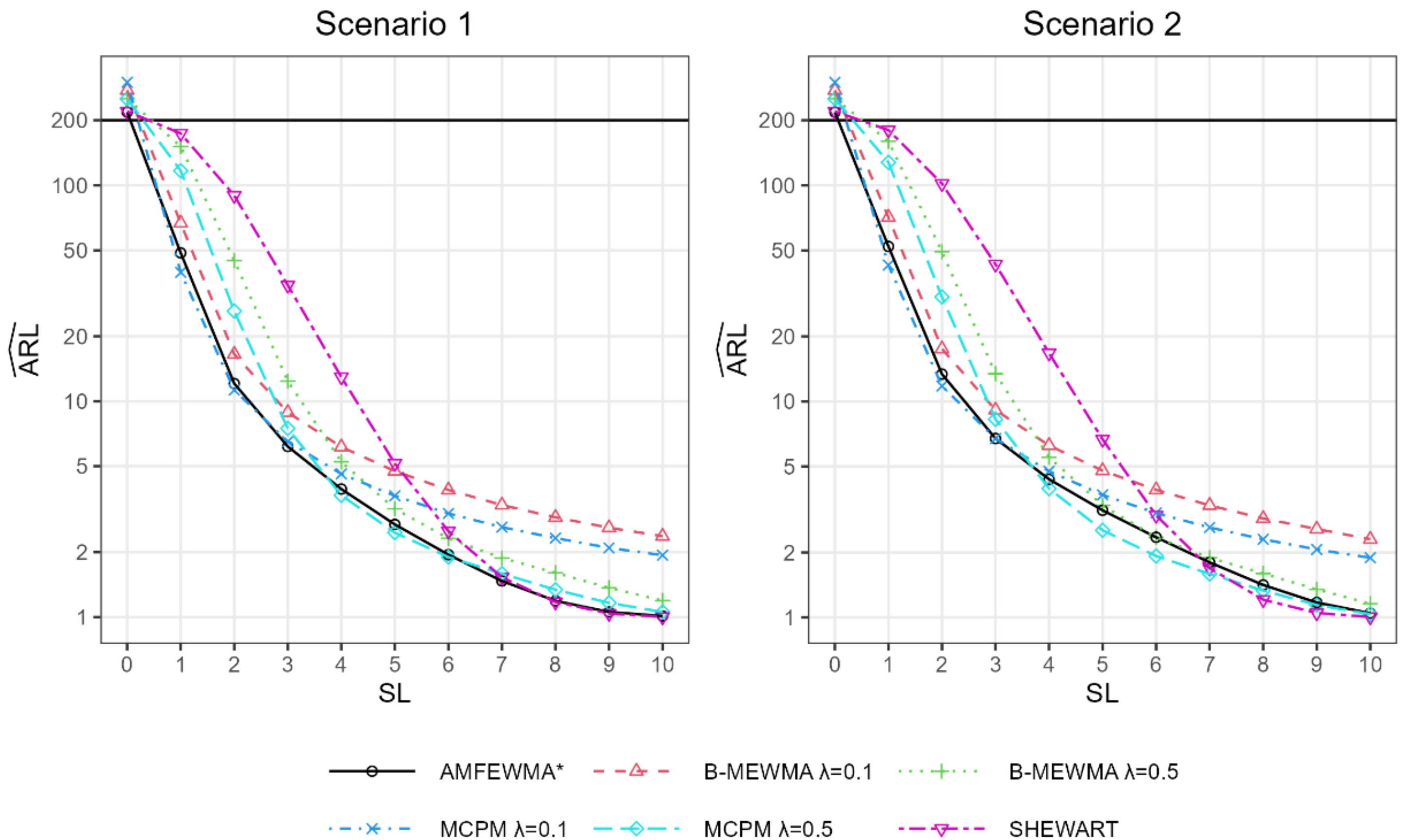
- Capizzi and Masarotto (2003) provided a solution to ***the problem of choosing λ adaptively*** :

$$Y_n = W(e_n) X_n + (1-W(e_n)) Y_{n-1}$$

$$\mathrm{e_n = X_n - Y_{n-1}}$$

Capezza, C., Capizzi, G., Centofanti, F., Lepore, A., & Palumbo, B. (2024). An Adaptive Multivariate Functional EWMA Control Chart. arXiv:2403.03837

Simulation study



Capezza, C., Capizzi, G., Centofanti, F., Lepore, A., & Palumbo, B. (2024). An Adaptive Multivariate Functional EWMA Control Chart. arXiv:2403.03837

Case study

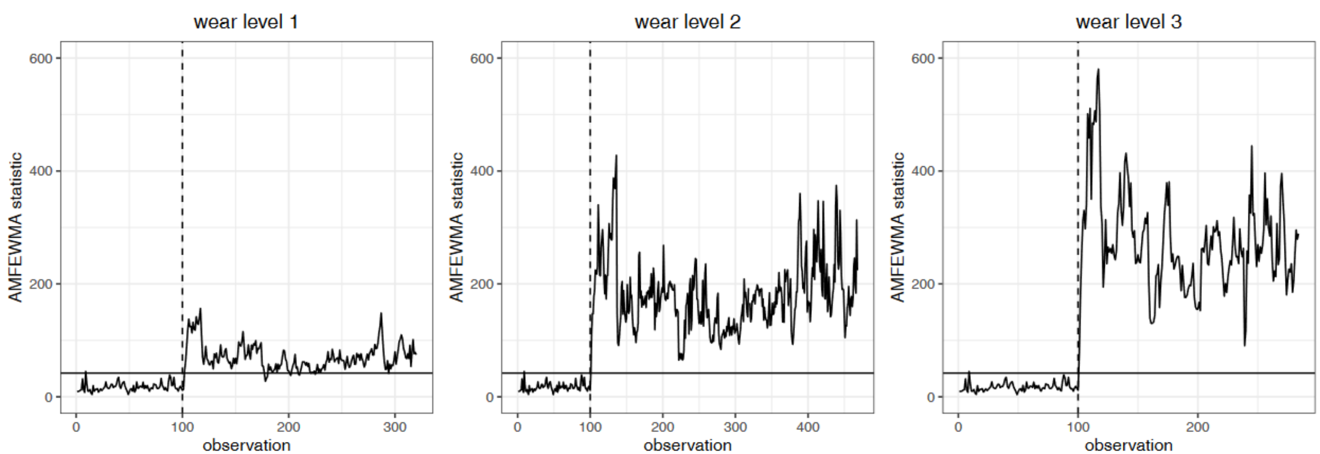


Figure 4: AMFEWMA* control chart under wear levels 1, 2, and 3. In each plot, the first 100 observations are randomly sampled from the tuning set, while the Phase II observations are reported after the dashed vertical line.

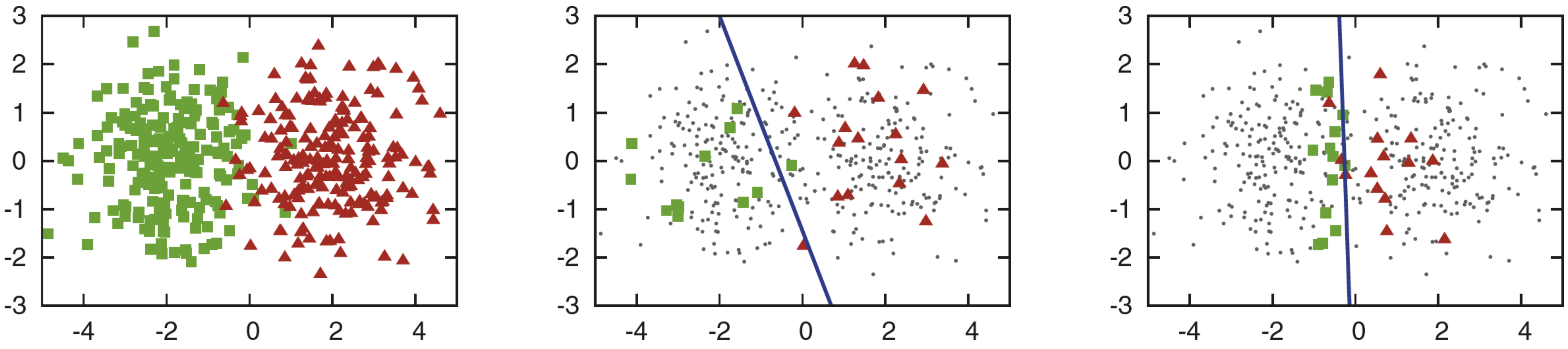
Table 5: $\widehat{\text{ARL}}$ calculated on the three Phase II datasets (corresponding to wear level 1, 2, and 3, respectively) in the case study, for the proposed AMFEWMA and the competing control charts. In bold, the lowest $\widehat{\text{ARL}}$ value for each row is reported.

	MFEWMA					AMFEWMA*
Wear level	λ=0.1	λ=0.2	λ=0.3	λ=0.5	$\lambda = 1$	$\lambda = 0.5, k = 4$
1	5.52	4.48	**4.14**	5.46	19.36	4.96
2	2.98	2.32	1.96	1.60	1.70	**1.56**
3	2.36	2.00	1.62	1.22	**1.12**	1.18

Capezza, C., Capizzi, G., Centofanti, F., Lepore, A., & Palumbo, B. (2024). An Adaptive Multivariate Functional EWMA Control Chart. arXiv:2403.03837

Stream-Based Active Learning for Anomaly Detection

What is Active Learning?



(a) a 2D toy data set (b) random sampling (c) uncertainty sampling

Image from: Settles, B. (2012). Active Learning. Synthesis Lectures On Artificial Intelligence And Machine Learning.

Joint work with Kamran Paynabar, Georgia Tech

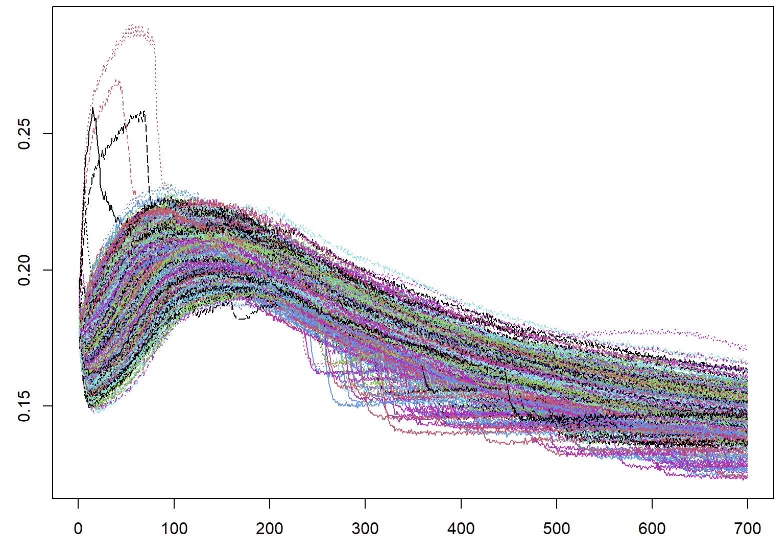
Stream-Based Active Learning for Anomaly Detection

Plot of dynamic resistance curves related to one specific point

Ultrasonic inspection give labels, however only for very few items

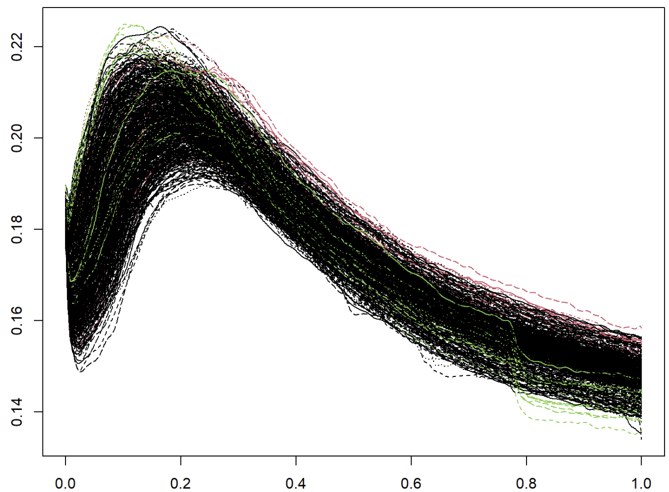
From the curves, some patterns seem to indicate a limited number of more common anomalies:

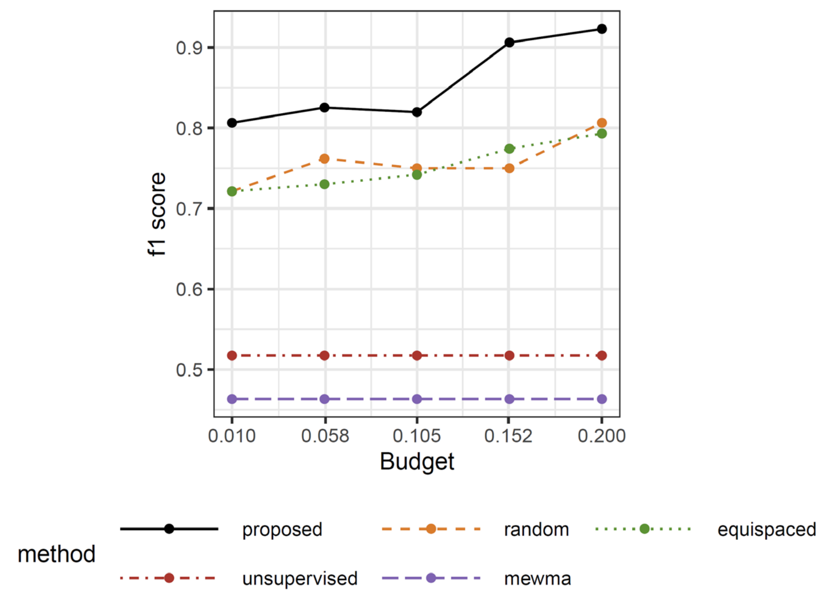
- **High initial** resistance suggests poor contact between the welding electrodes and the workpieces, possibly due to surface contaminants or misaligned electrodes
- **Abrupt changes** in resistance: sudden drops in resistance during the welding process indicates expulsion (molten material is expelled from the weld)
- **Cold weld**: inadequately sized weld nugget



Joint work with Kamran Paynabar, Georgia Tech

Case study





Joint work with Kamran Paynabar, Georgia Tech

Agenda

- The Statistics for Engineering Research (SFERe) group
- Main industrial scenarios
- An example: our research in the automotive industry
- **Technological transfer**
- Conclusions

Technological transfer

Challenge in partnership with industries:

- Make new methods accessible!
- Free software solutions
- R/Python packages


SFERe

Statistics For Engineering Research

Home People Publications Software Seminars

Software

R packages

- funcharts: Functional Control Charts.
- slasso: S-LASSO Estimator for the Function-on-Function Linear Regression.
- sasfunclust: Sparse and Smooth Functional Clustering.
- adass: Adaptive Smoothing Spline (AdaSS) Estimator for the Function-on-Function Linear Regression
- rofanova: Robust Functional Analysis of Variance

Python packages

- NN4MSP: Neural network for multiple stream processes
- NN4OCMSP: Neural network for out-of-control signal interpretation in multiple stream processes

The funcharts R package

The newest version 1.4.1 is available on CRAN:

funcharts: Functional Control Charts

Provides functional control charts for statistical process monitoring of functional data, using the methods of Capezza et al. (2020) <doi:10.1002/asmb.2507> and Centofanti et al. (2021) <doi:10.1080/00401706.2020.1753581>. The package is thoroughly illustrated in the paper of Capezza et al (2023) <doi:10.1080/00224065.2023.2219012>.

Version:	1.4.1
Depends:	R (≥ 3.6.0), robustbase
Imports:	fda, ggplot2, rlang, parallel, tidyr, patchwork, RSpectra, matrixStats, roahd, dplyr, stringr, fda.usc, rrcov, rofanova, Matrix, MASS, mvtnorm
Suggests:	covr, knitr, rmarkdown, testthat
Published:	2024-02-22
Author:	Christian Capezza [cre, aut], Fabio Centofanti [aut], Antonio Lepore [aut], Alessandra Menafoglio [aut], Biagio Palumbo [aut], Simone Vantini [aut]
Maintainer:	Christian Capezza <christian.capezza at unina.it>
BugReports:	https://github.com/unina-sfere/funcharts/issues
License:	GPL-3
URL:	https://github.com/unina-sfere/funcharts
NeedsCompilation:	no
Citation:	funcharts citation info
Materials:	README NEWS
CRAN checks:	funcharts results

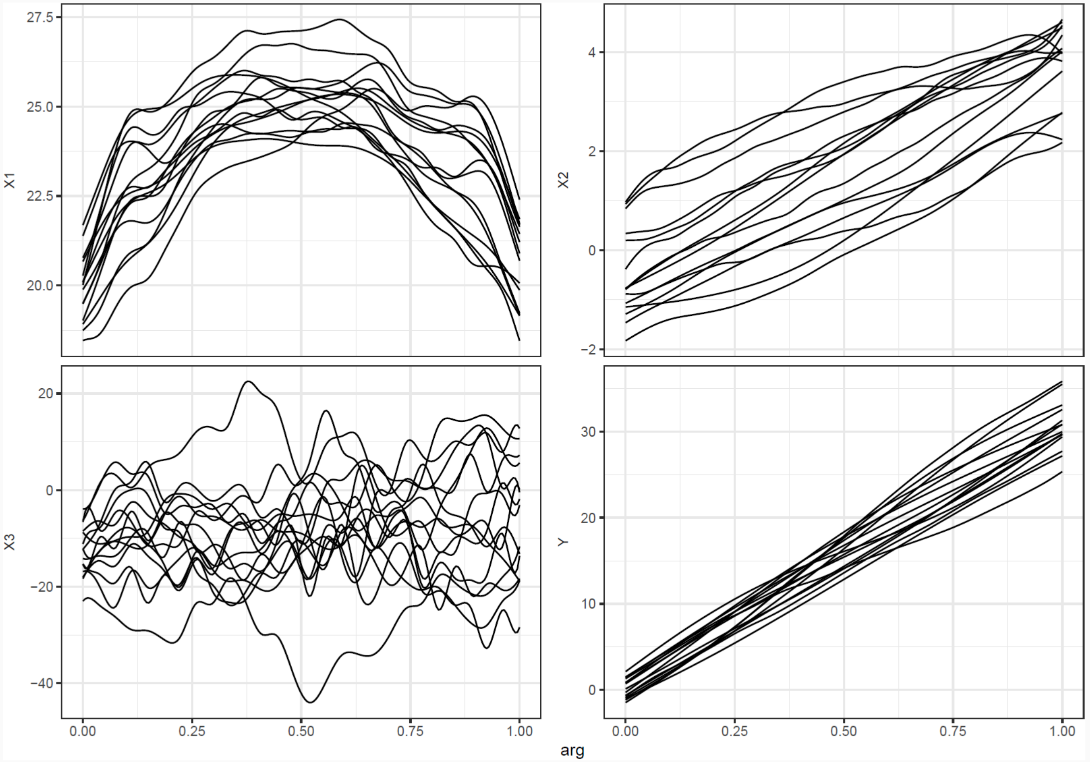
The paper is thoroughly illustrated in:

Capezza, C., Centofanti, F., Lepore, A., Menafoglio, A., Palumbo, B., & Vantini, S. (2023). Funcharts: Control charts for multivariate functional data in R. *Journal of Quality Technology*, 1-18.

The funcharts R package

It allows to simulate multivariate functional data

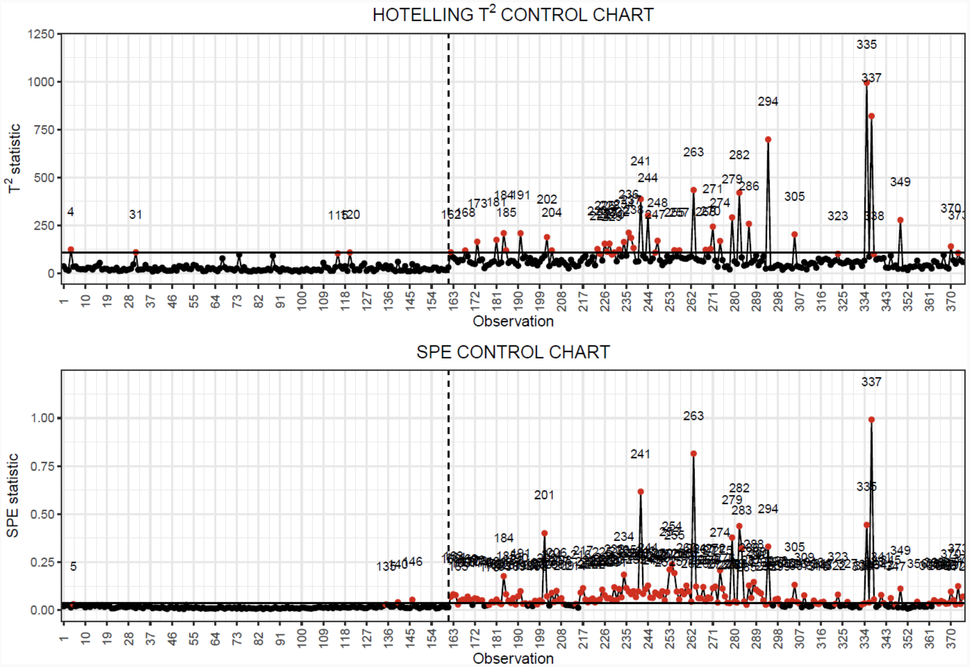
```
mfdobj <- get_mfd_list(dat[1:4])
plot_mfd(mfdobj)
```



The funcharts R package

It allows to build control charts for multivariate functional data

```
cclist_frcc <- regr_cc_fof(mod_fof_pc, mfdobj_y_new = y2, mfdobj_x_new = x2)
plot_control_charts(cclist_frcc)
```

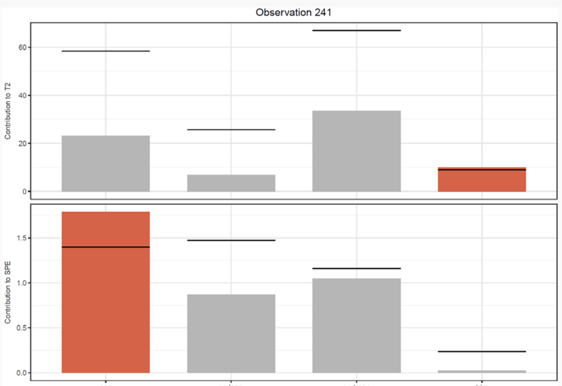


The funcharts R package

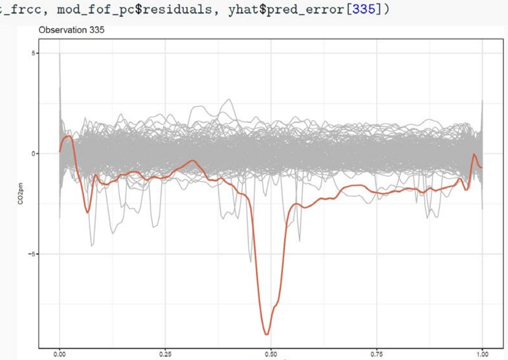
It allows to perform fault detection to identify anomalous functional variables

Contribution plots → Plot of the anomalous variables

```r
cont_plot(cclist_pca, id_num = 241)
```




```r
yhat <- predict_fof_pc(object = mod_fof_pc, mfdobj_y_new = y2, mfdobj_x_new = x2)
plot_mon(cclist_frcc, mod_fof_pc$residuals, yhat$pred_error[335])
```



Agenda

- The Statistics for Engineering Research (SFERe) group
- Main industrial scenarios
- An example: our research in the automotive industry
- Technological transfer
- **Conclusions**

Conclusions

- It is possible to compete in the industrial sector through interpretable statistical methods.
- Statistics catalyzes the knowledge process in the industrial field.
- Addressing broader issues is more important rather than solving individual problems.
- Companies are called to invest in human resources able to interpret complex industrial scenarios via a statistical approach.
- Science before data.
- Critical analysis is needed of the management costs of a large and undifferentiated quantity of data compared to the increase in knowledge derived from their use.

Thanks for the attention! Q&A