New trends of SPC and Data Analysis in Industry

Marco S. Reis & Pedro Saraiva

October 27th, 2015

PSE group - CIEPQPF, Department of Chemical Engineering
University of Coimbra, Portugal

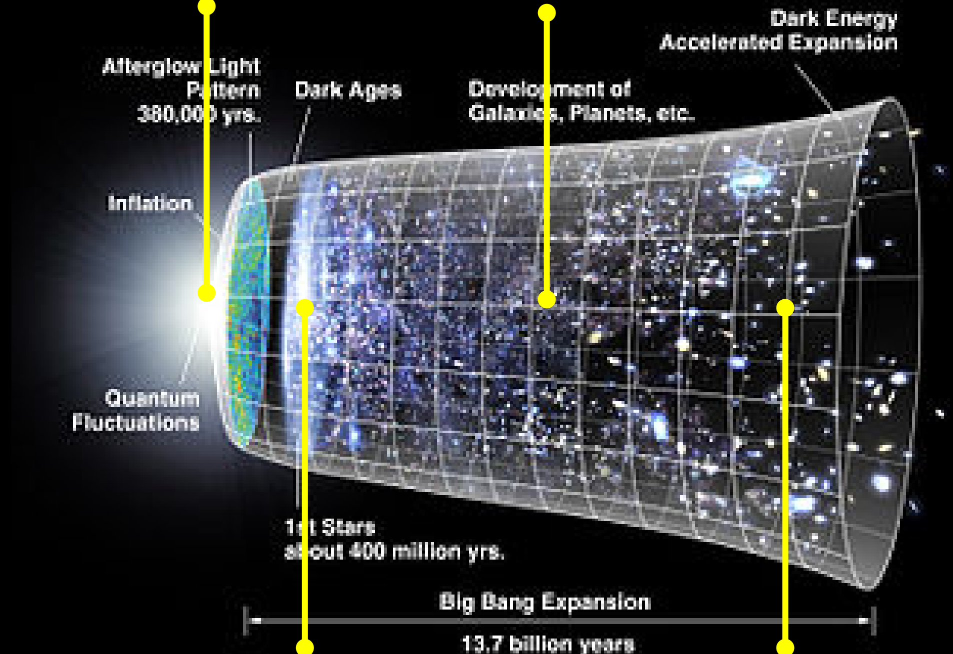
Outline

1. Introduction & Motivation
2. Examples & Current trends
   – Megavariate SPC in i.i.d. processes
   – Megavariate SPC in processes with autocorrelation
   – Monitoring of higher-order Profiles (1D, 2D, 3D, …)
   – Batch Processes Monitoring (BPM)
3. Conclusions

A brief story of … **SPC** for **large scale industrial processes**

1920's
W.A. Shewhart
Univariate methods
(Shewhart, EWMA, CUSUM, etc.)

1959
J.E. Jackson
Megavariate methods
(MSPC-PCA, -PLS)

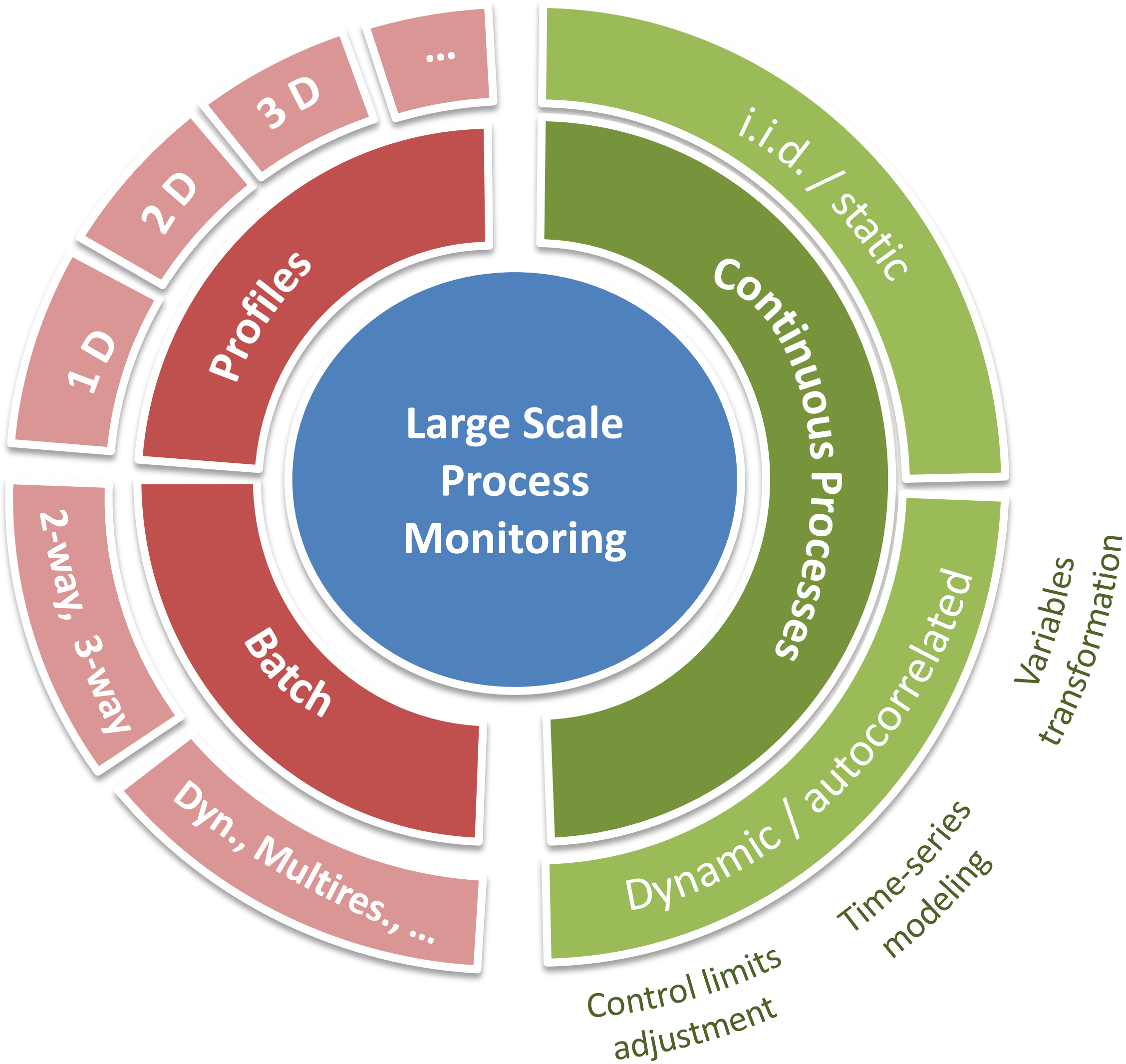


… end of story?

1931
H. Hotelling
Multivariate methods
(Hotelling's T^2, MEWMA, MCUSUM)

1995
Ku et al.
Megavariate dynamic methods
(DPCA)

Typology of Large Scale SPC applications



Goal

SPC for **Large Scale** Industrial Processes

- Examples
- Current trends

Topics

- Megavariate SPC in i.i.d. processes
- Megavariate SPC in processes with autocorrelation
- Monitoring of higher-order Profiles (1D, 2D, 3D, …)
- Batch Processes Monitoring (BPM)

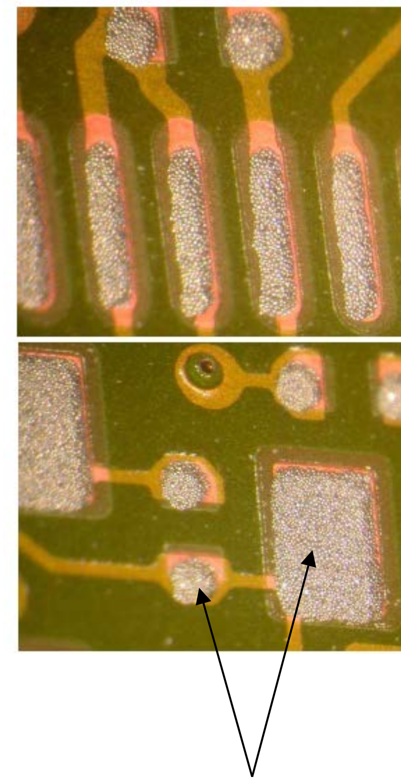
Megavariate SPC in i.i.d. processes

“from correlation-oriented to structured approaches”

- **Megavariate statistical process control** in electronic devices assembling
- Monitoring the **process correlated structure** incorporating the network structure of the system

Example:

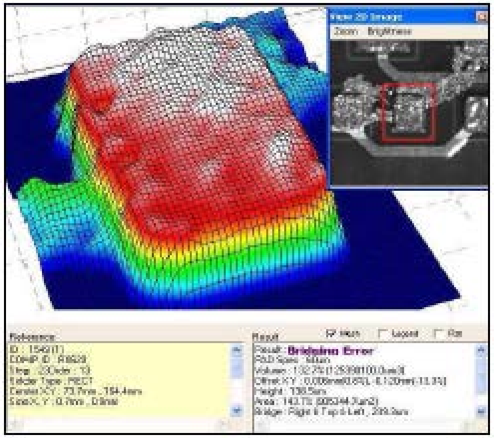
Megavariate statistical process control in electronic devices assembling



Solder Paste Deposits (SPD)

The problem

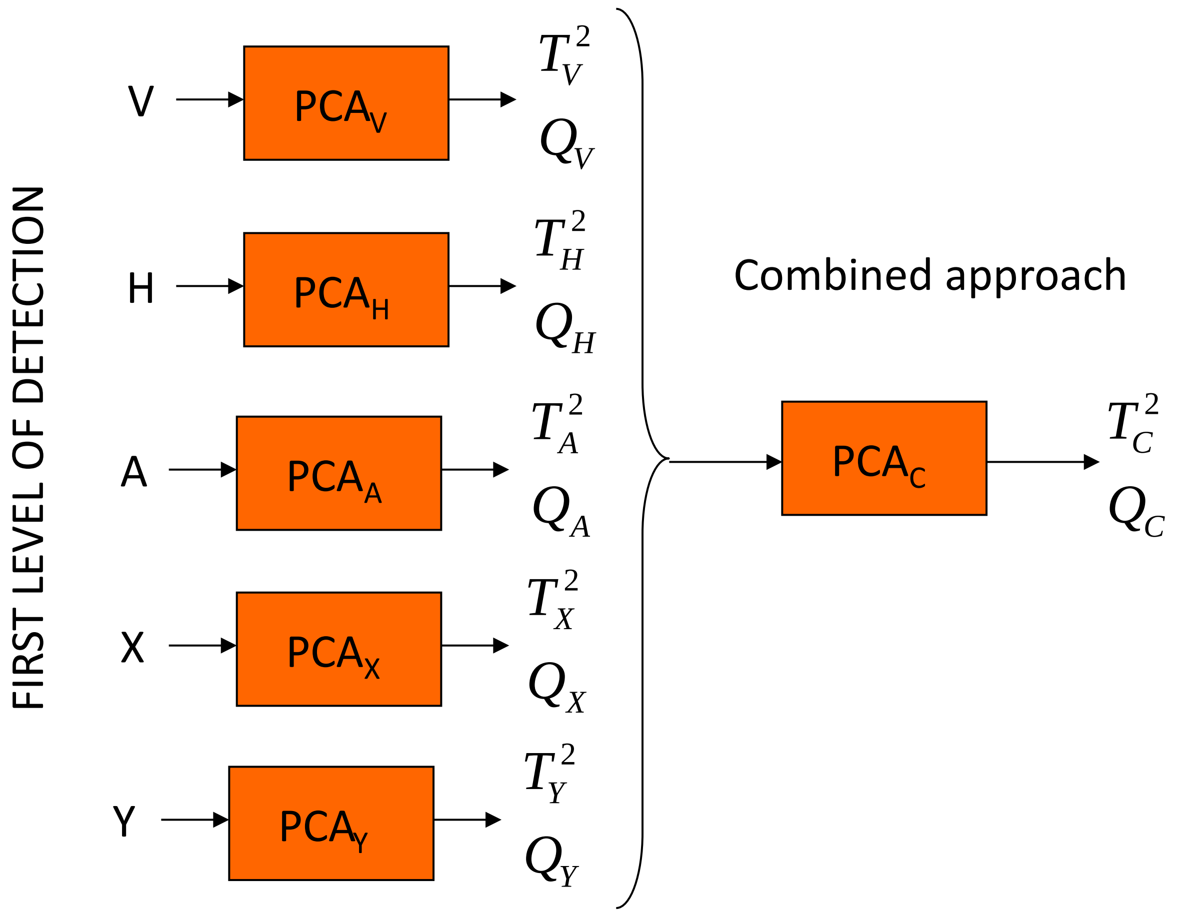
- 100% inspection of Printed Circuit Boards (PCB's).
- Each PCB has more than 3000 deposits (SPD's) of different shapes.
- Operators have less than 1 min to decide about the status of each PCB.
- Each solder deposit is evaluates according to 5 parameters obtained through Moiré interferometry
  - Volume (V)
  - Area (A)
  - Height (H)
  - Offset in the X coordinate (X)
  - Offset in the Y coordinate (Y)



> 15 000 measurements for each PCB!

- Multivariate Statistical Process Control using Principal Components Analysis (PCA-MSPC*)

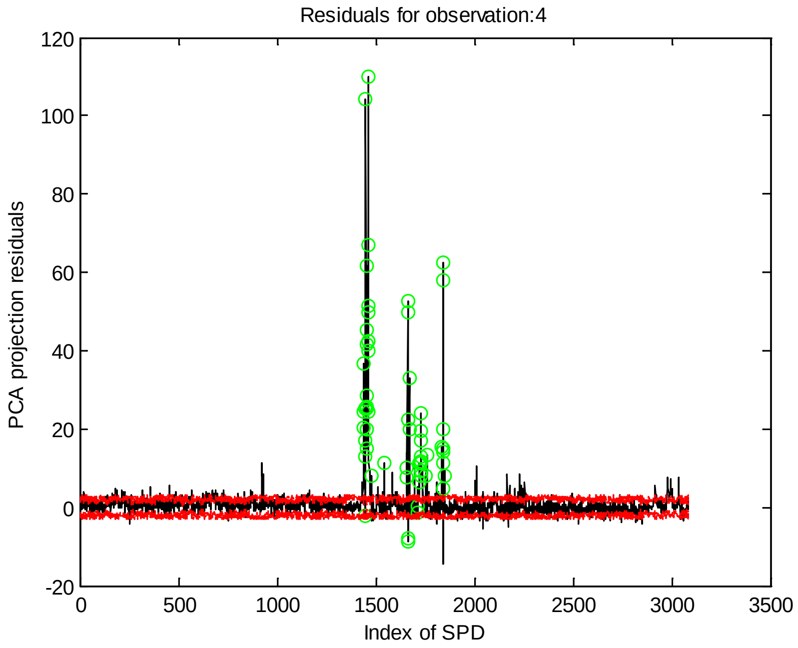
* J.E. Jackson, Technometrics, 1:4 (1959) 359-377.



· U C ·

SECOND LEVEL OF DETECTION

- Analysis of the residuals from the projection of each multivariate observation to the PCA subspace.



· U C ·

RESULTS FOR THE FIRST LEVEL OF DETECTION

Detection Statistics	Measurements used to compute the relative area under the **ROC curve** (values in %)					
	Height (H)	*Area (A)*	*Volume (V)*	*Offset X*	*Offset Y*	*Combined approach*
T^2	70.00	62.50	85.63	76.88	70.63	90.00
Q	93.13	93.75	91.88	85.00	83.13	90.63

10 PCB's classified as "good" were used to represent NOC data in SPC (estimate the PCA subspace, ...)
16 PCB's classified as "fail" (16) and "good" (5) were used to test the procedure

· U C ·

RESULTS FOR THE SECOND LEVEL OF DETECTION

Detection Statistics	Measurements used to identify abnormal SPD's (values in %)					
	Height (H)	*Area (A)*	*Volume (V)*	*Offset X*	*Offset Y*	*Combined approach*
Mean	80.33	65.82	76.04	60.23	54.38	72.47
Standard Deviation	20.07	29.97	21.02	21.23	29.93	17.31

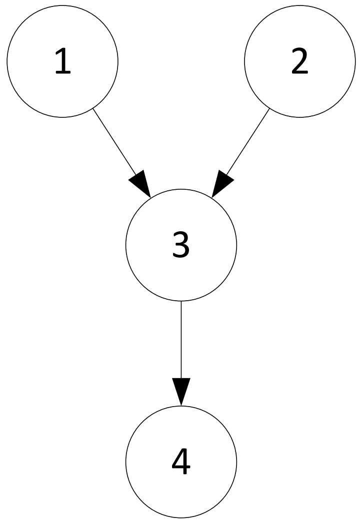
Reis, M.S. and P. Delgado, *A large-scale statistical process control approach for the monitoring of electronic devices assemblage.* Computers and Chemical Engineering, 2012. **39**: p. 163-169.

Emerging trend

- Most MSPC schemes are focused on **location** (mean level);
- Current MSPC procedures for multivariate dispersion are based on **marginal covariance**:
  - Likelihood ratio test;
  - Generalized variance;
  - PCA-MSPC.
- These methods are unable, by design, to discern **local changes in the process fine structure**;
- Current approaches **do not consider the underlying process causal structure**
  - Detection
  - Diagnosis

U C

Methods: Sensitivity enhancing transformation (SET)



1. Network Identification

$$x_1 \rightarrow x_3$$
$$x_2 \rightarrow x_3$$
$$x_3 \rightarrow x_4$$

2. Regress each variable onto its parents

$$y_1 = x_1$$
$$y_2 = x_2$$
$$y_3 = x_3 - x_1 b_{1,3} - x_2 b_{2,3}$$
$$y_4 = x_4 - x_3 b_{3,4}$$

3. Final model

$$\mathbf{Y} = \mathbf{XB}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & -b_{1,3} & 0 \\ 0 & 1 & -b_{2,3} & 0 \\ 0 & 0 & 1 & -b_{3,4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

4. Apply the Cholesky decomposition to the regression residuals thus obtained.

Methods

Process Data

↓

SET

↓

Uncorrelated data

↓

Monitoring statistics (PC, etc.)

↓

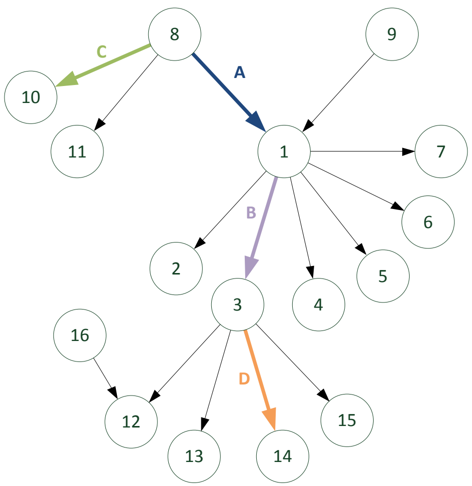
Statistics

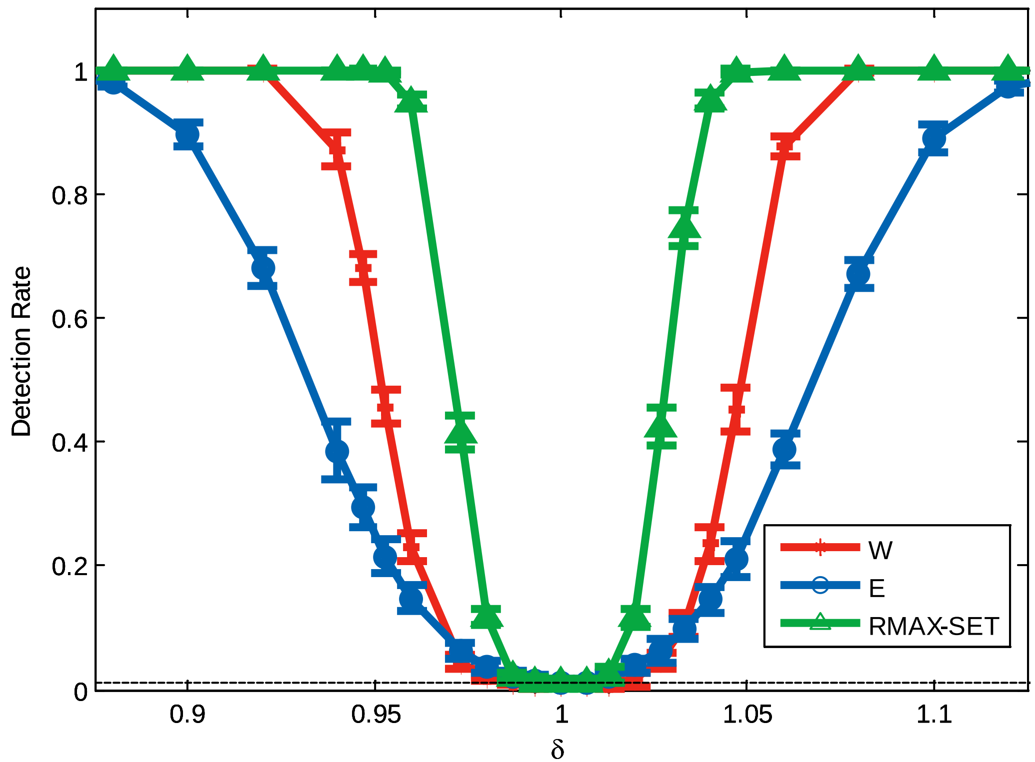
↓

MSPC

↓

Process State (Normal / Abnormal)





Rato, T.J. and M.S. Reis, *Sensitivity enhancing transformations for monitoring the process correlation structure.* Journal of Process Control, 2014. **24**: p. 905-915.

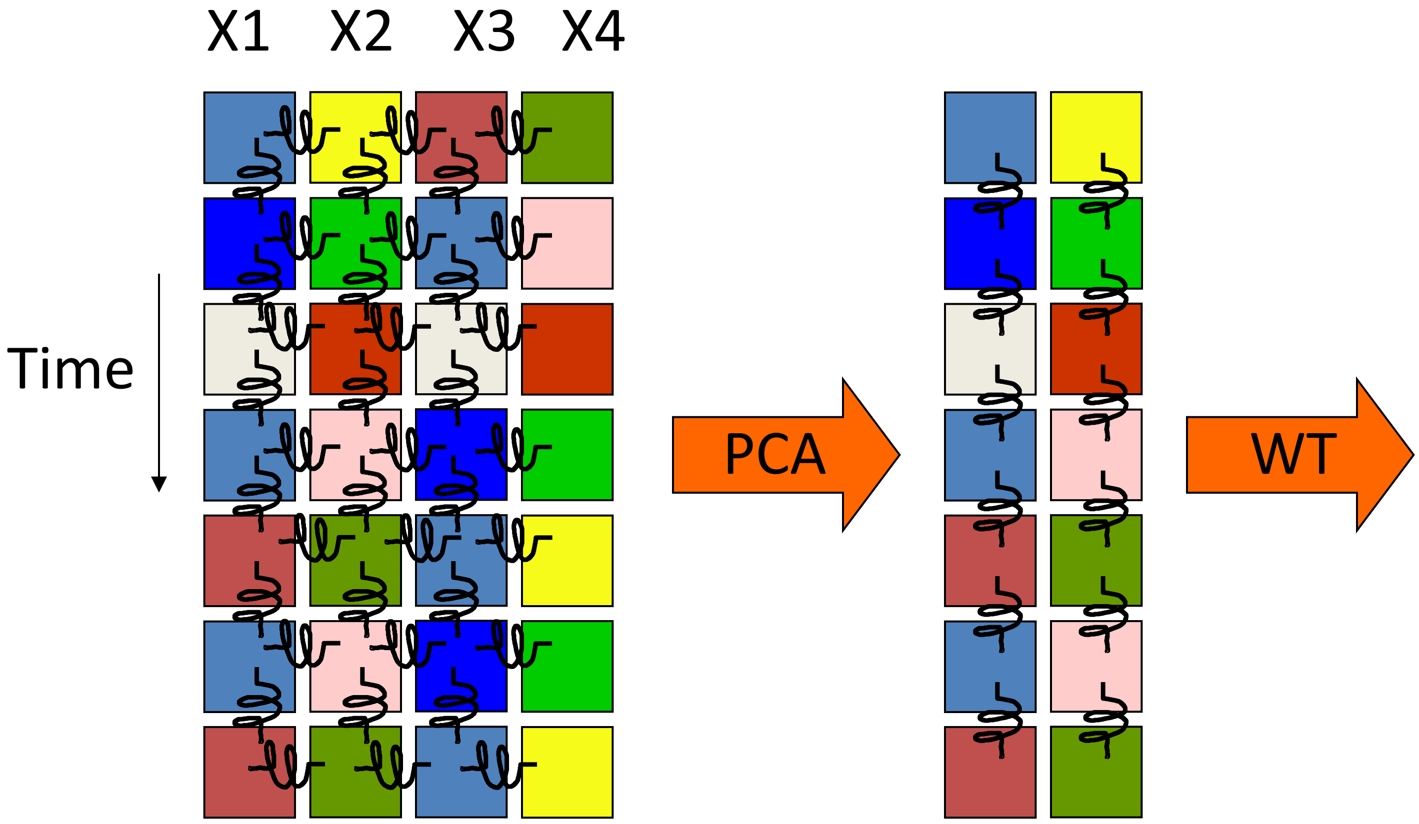
Rato, T.J. and M.S. Reis, *On-line process monitoring using local measures of association. Part I: Detection performance.* Chemometrics and Intelligent Laboratory Systems, 2015. **142**: p. 255-264.

Megavariate SPC in processes with autocorrelation

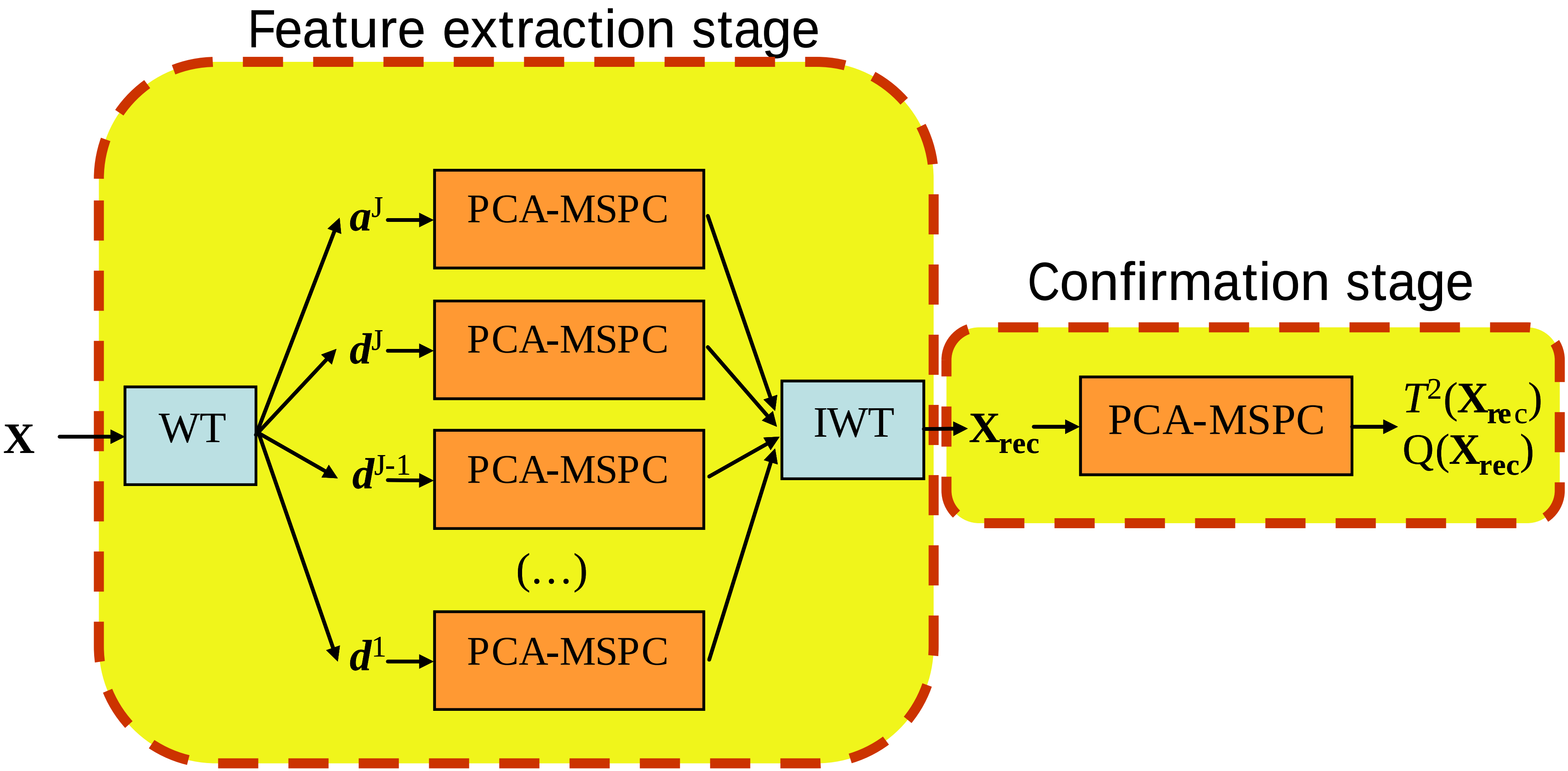
- Multiscale statistical process control (MS-SPC)
- Megavariate SPC in processes with autocorrelation

Multiscale SPC

- Process data usually presents cross-correlation and autocorrelation



MS-SPC

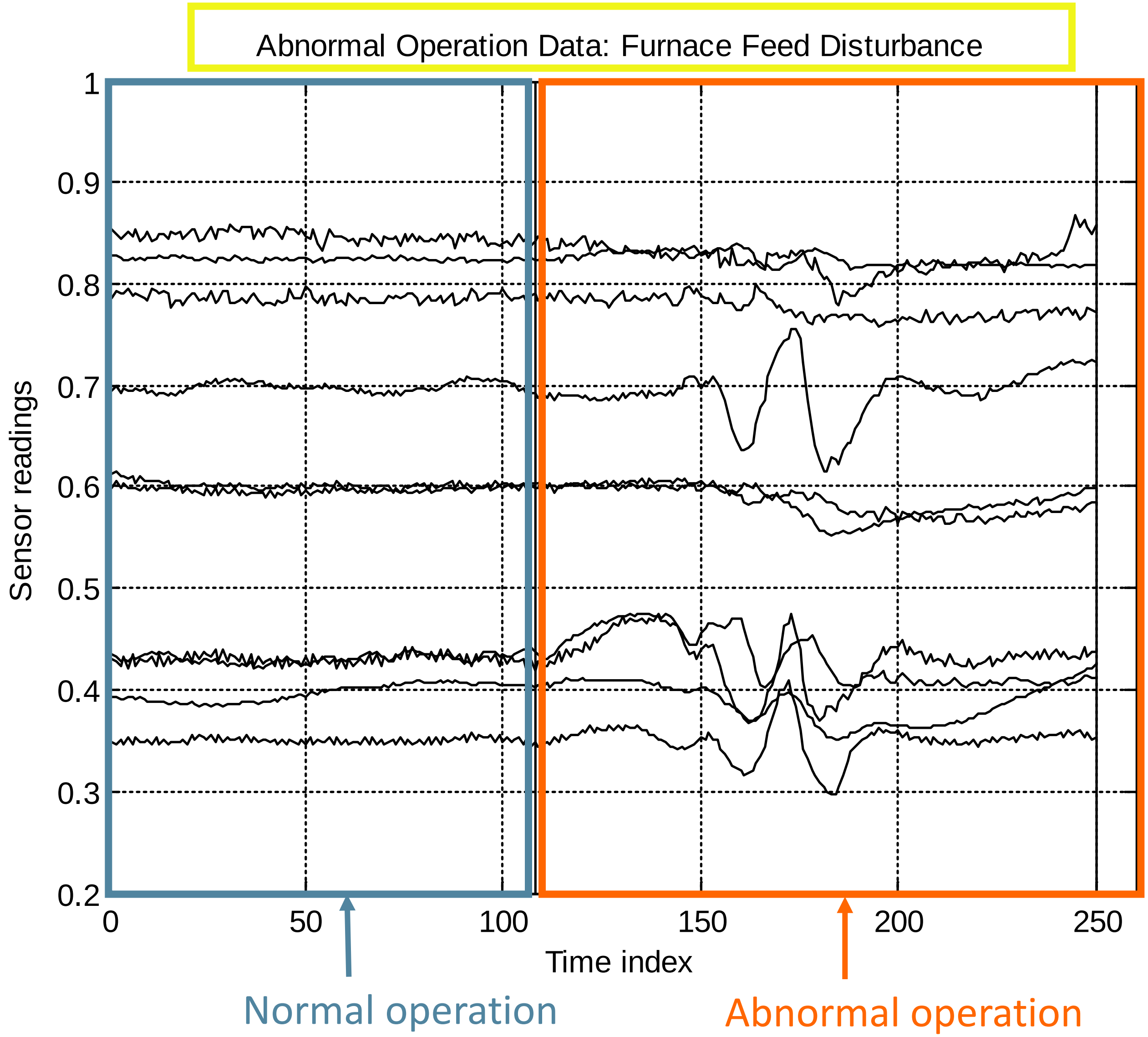


Bakshi BR. Multiscale PCA with Application to Multivariate Statistical Process Control. *AIChE Journal*. 1998;44(7):1596-1610

Some useful features of MSSPC

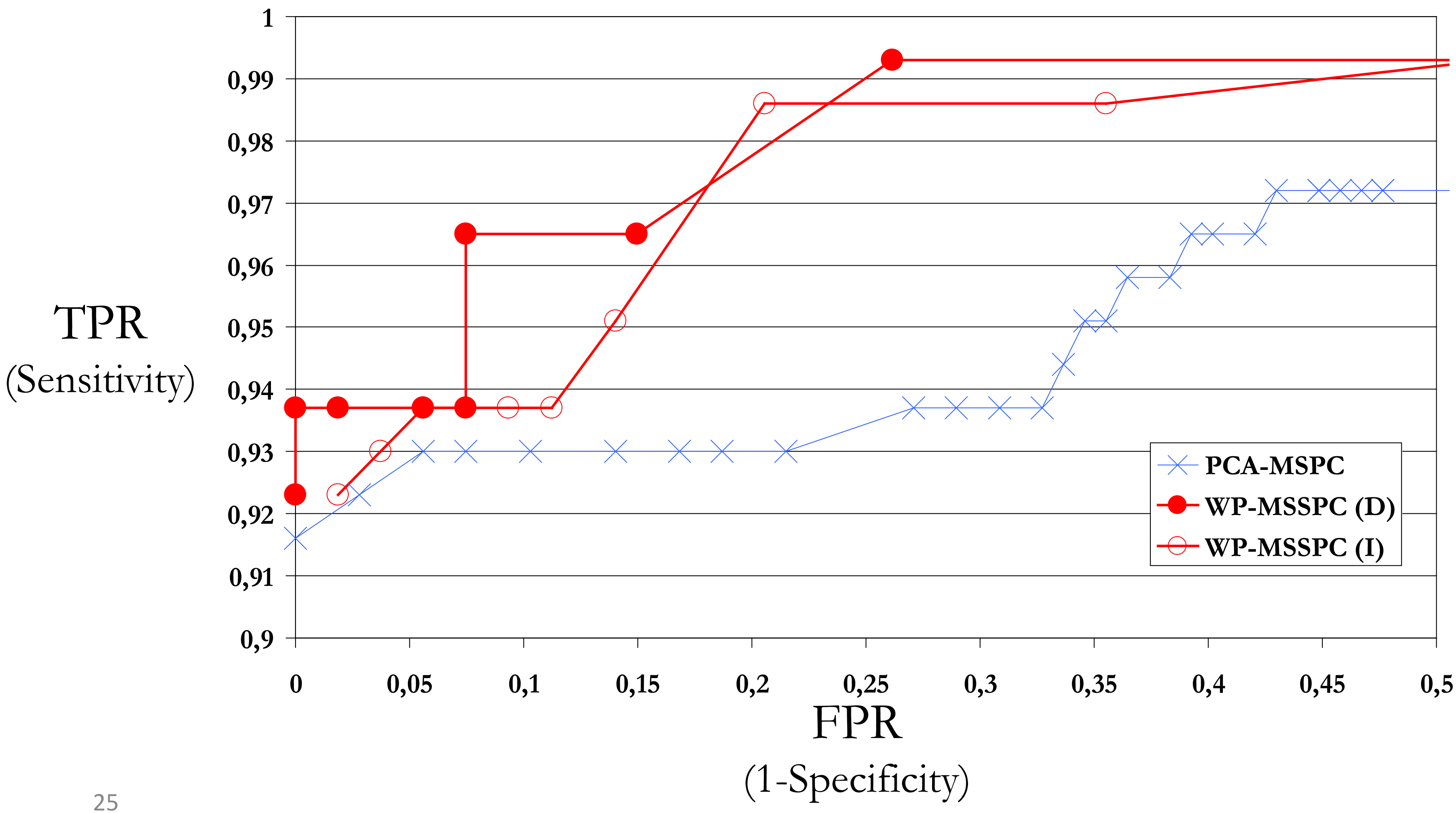
- Good performance for a wide variety of signals
  - Subsume other single-scale methods according to the scales selected!
- More adequate for detecting unknown disturbances, especially those with complex patterns
- Handles autocorrelation in a natural way
  - Wavelet coefficients are approximately uncorrelated and normal distributed.

Some Results: ROC studies



Some Results: ROC studies

- Perturbation: **Furnace Feed Disturbance**



Example:

Dynamic PCA with Decorrelated Residuals: DPCA-DR (Rato & Reis)

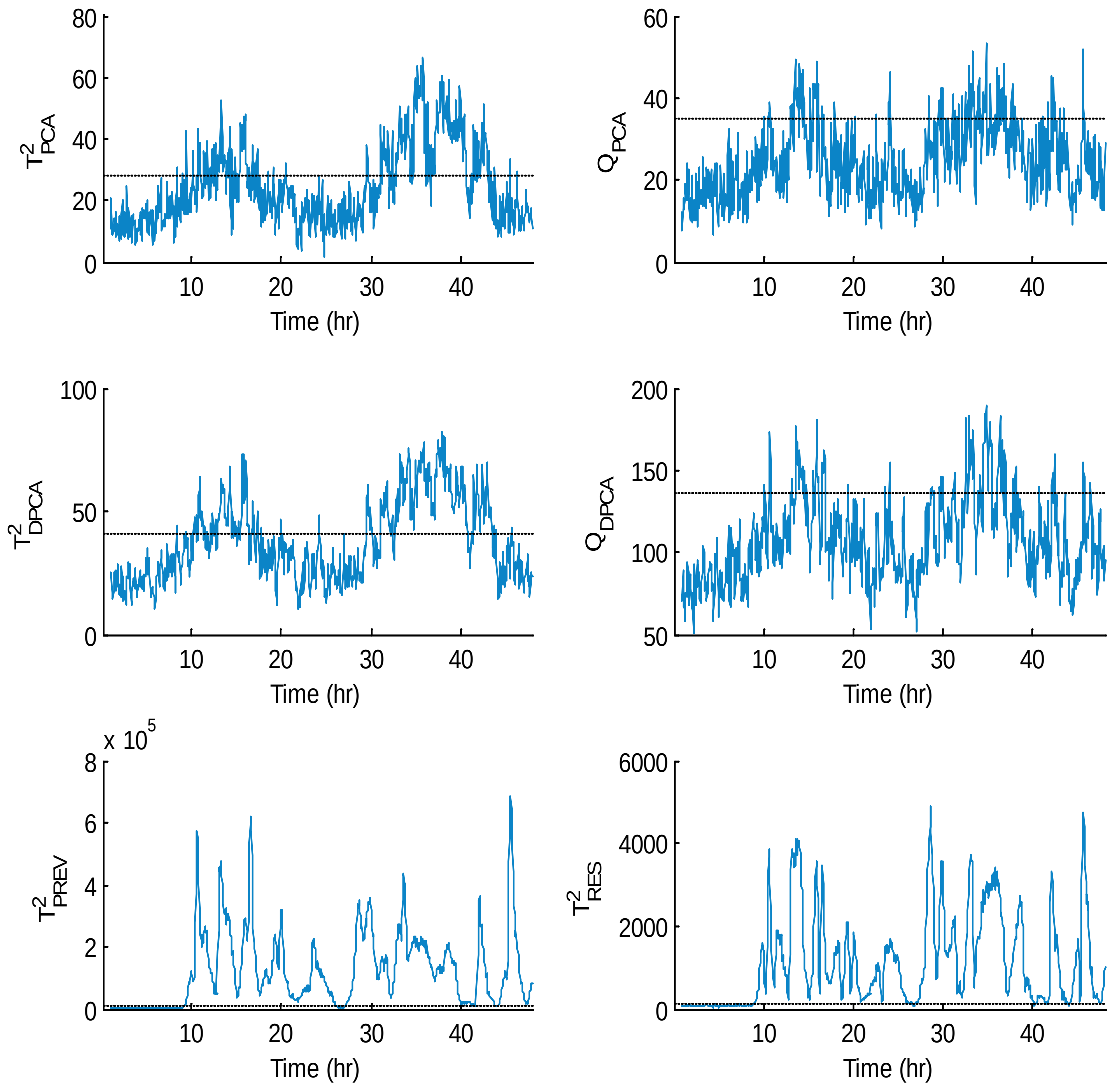
$$\mathbf{X} = [\mathbf{X}(k)\ \mathbf{X}(k-1)\ \ldots\ \mathbf{X}(k-l)];$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{\#} \\ \mathbf{x}^{*} \end{bmatrix}, \quad \mathbf{S} = \mathbf{P}\Lambda\mathbf{P}^{\mathrm{T}}, \quad \mathbf{P} = \begin{bmatrix} \mathbf{P}^{\#} \\ \mathbf{P}^{*} \end{bmatrix}$$

$$\hat{\mathbf{t}}_{1:k} = \Lambda \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k\times(n-k)} \end{bmatrix} \mathbf{P}^{*\mathrm{T}} \left(\mathbf{P}^{*} \Lambda \mathbf{P}^{*\mathrm{T}} \right)^{-1} \mathbf{x}^{*}$$

Scores	$T^2_{PREV} = \left(\mathbf{t} - \hat{\mathbf{t}}\right)^{\mathrm{T}} \mathbf{S}^{-1}_{\mathbf{t}-\hat{\mathbf{t}}} \left(\mathbf{t} - \hat{\mathbf{t}}\right)$
Residuals	$T^2_{RES} = \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right)^{\mathrm{T}} \mathbf{S}^{-1}_{\hat{\mathbf{t}}} \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right)$

Tennessee Eastman process



Tennessee Eastman process

A \ B		PCA		DPCA		DPCA-MD	
		T^2	Q	T^2	Q	T^2_{PREV}	T^2_{RES}
PCA	T^2		0.388 (-)	0.414 (-)	0.152 (-)	**0.002 (-)**	**0.003 (-)**
	Q	0.388 (+)		0.540 (+)	**0,046 (-)**	**0,004 (-)**	**0,008 (-)**
DPCA	T^2	0,414 (+)	0,540 (-)		0,257 (-)	**0,002 (-)**	**0,004 (-)**
	Q	0,152 (+)	**0,046 (+)**	0,257 (+)		**0,007 (-)**	**0,013 (-)**
DPCA-MD	T^2_{PREV}	**0,002 (+)**	**0,004 (+)**	**0,002 (+)**	**0,007 (+)**		0,115 (+)
	T^2_{RES}	**0,003 (+)**	**0,008 (+)**	**0,004 (+)**	**0,013 (+)**	0,115 (-)	

Rato, T.J. and M.S. Reis, *Fault detection in the Tennessee Eastman process using dynamic principal components analysis with decorrelated residuals (DPCA-DR).* Chemometrics and Intelligent Laboratory Systems, 2013. **125**: p. 101-108.

Monitoring Profiles (1D, 2D, 3D, …)

"(…) We view the monitoring of process and product profiles as the most promising area of research in statistical process control. (…)"

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004). Using Control Charts to Monitor Process and Product Quality Profiles. *Journal of Quality Technology*, 36(3), 309-320.

- Definition [Profile, ***P***]:

An array of data, indexed by time and/or space, that characterizes a given entity (product, process).

$$P:\left\{\mathbf{Y}\left(ix,iy,iz,it\right)\right\}_{\underbrace{ix,iy,iz}_{\text{Spatial indices}},\underbrace{it}_{\text{Time index}}\in\Omega_x\times\Omega_y\times\Omega_z\times\Omega_t}$$

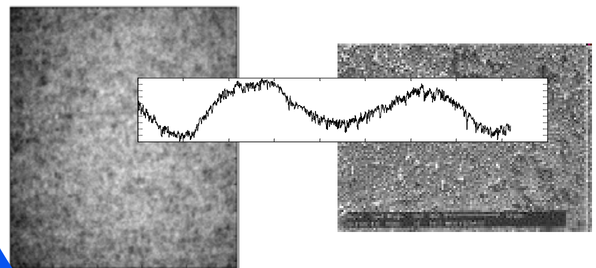
$$\left\{\mathbf{Y}\right\}_{ix,iy,yz,it}\in\Re^n$$

“Dual domain” classification

Frequency domain localization

High

Frequency Localized Random Profiles



Fully Localized Profiles

More:
Acoustic signals
Seismic signals

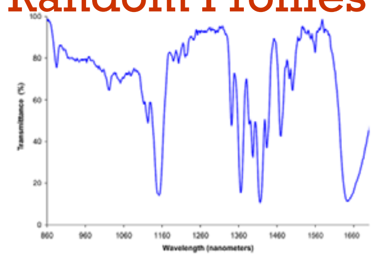
“Hard” constraint:

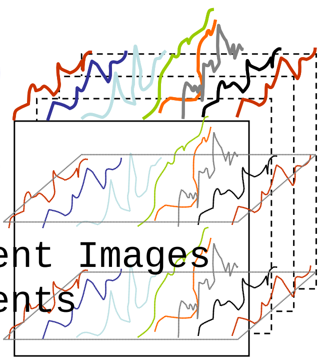
$$\sigma(g)\cdot\sigma(\hat{g}) \geq \frac{1}{2}$$

Low

Delocalized Random Profiles

Time-Space Localized Random Profiles





More:
Analysis of Coherent Images
Perfusion experiments
fMRI, …

Low | High

Time/Space localization

Monitoring paper surface: Formation* (Reis, MS & Bauer, A.)

- Currently is evaluated off-line: few times per day (e.g. after each paper reel production)
  - Very high delay, regarding the production speed of current paper machines (**~100 Km/h!)**

* Level of uniformity in the way fibres are distributed across the paper surface.

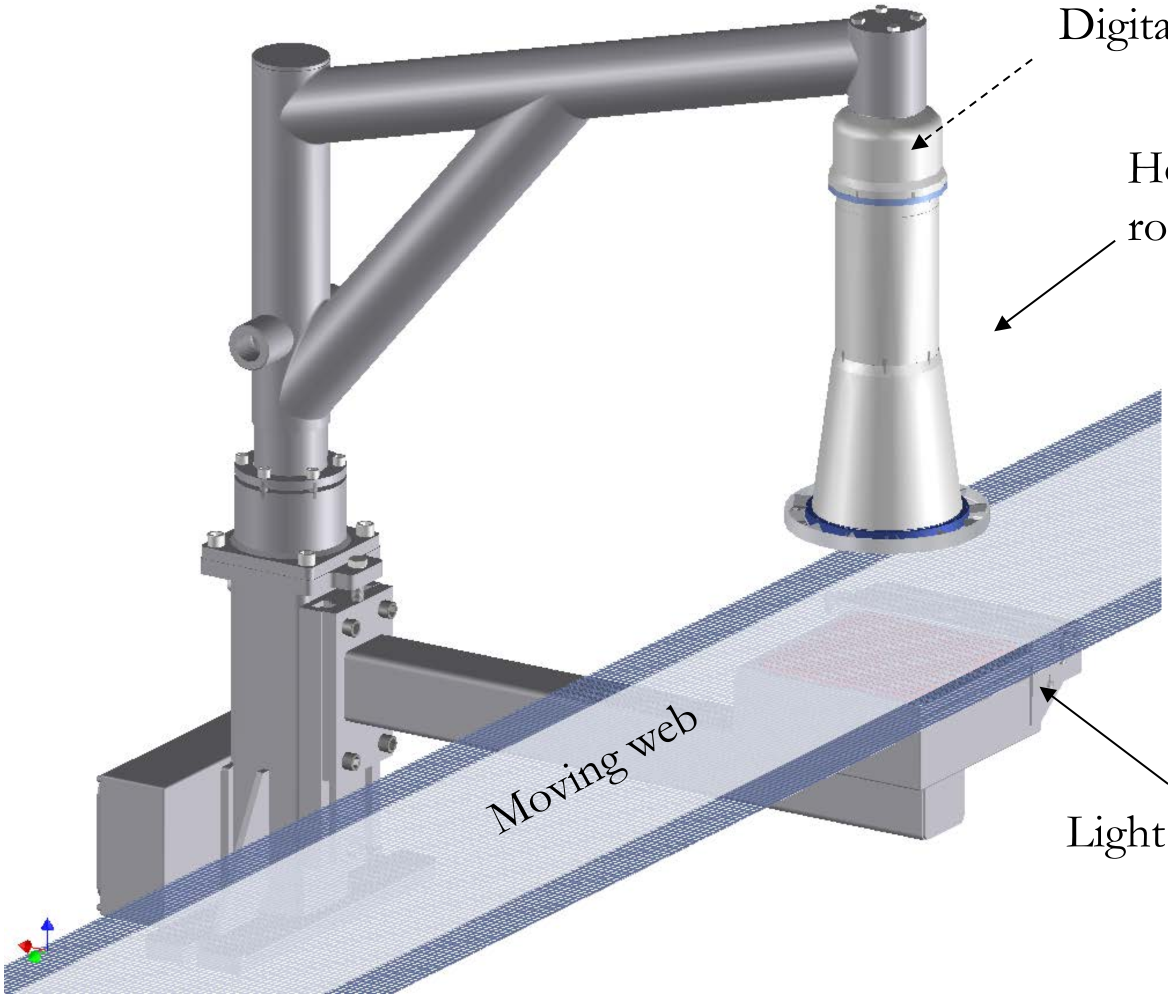
Goal

Develop a technology for

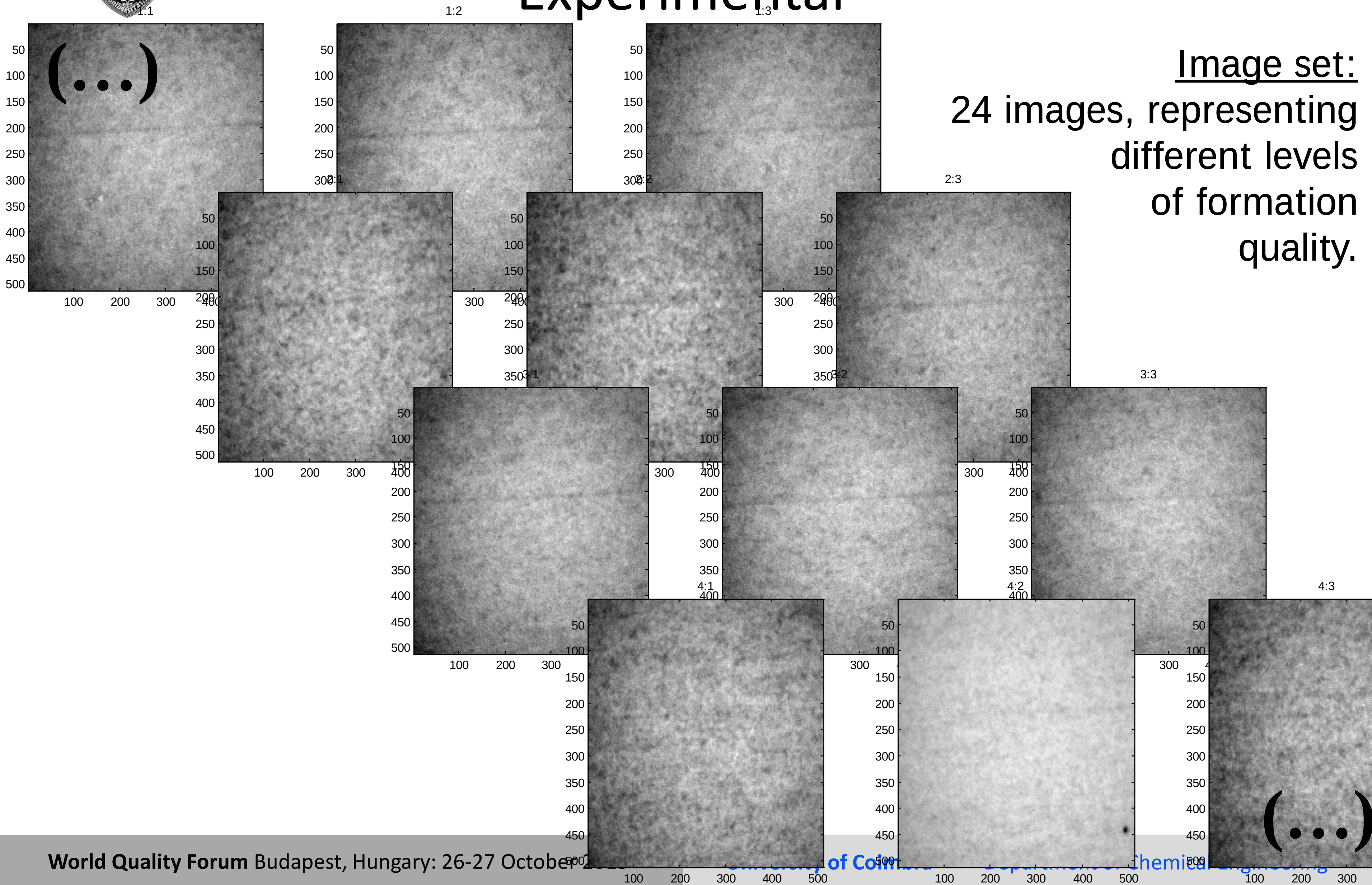
on-line monitoring of the paper formation.

· U C ·

Experimental: Measurement Sensor

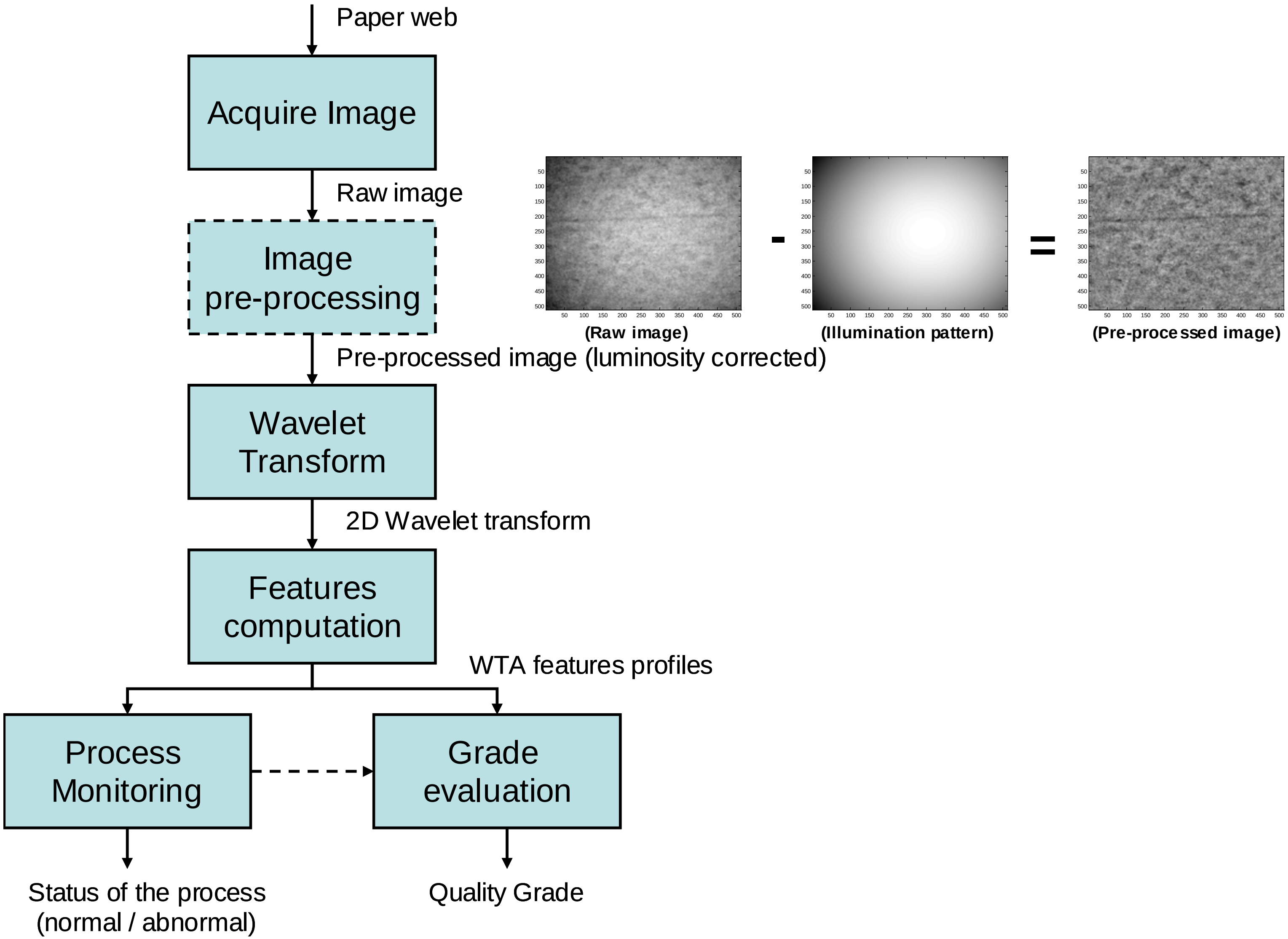


Experimental



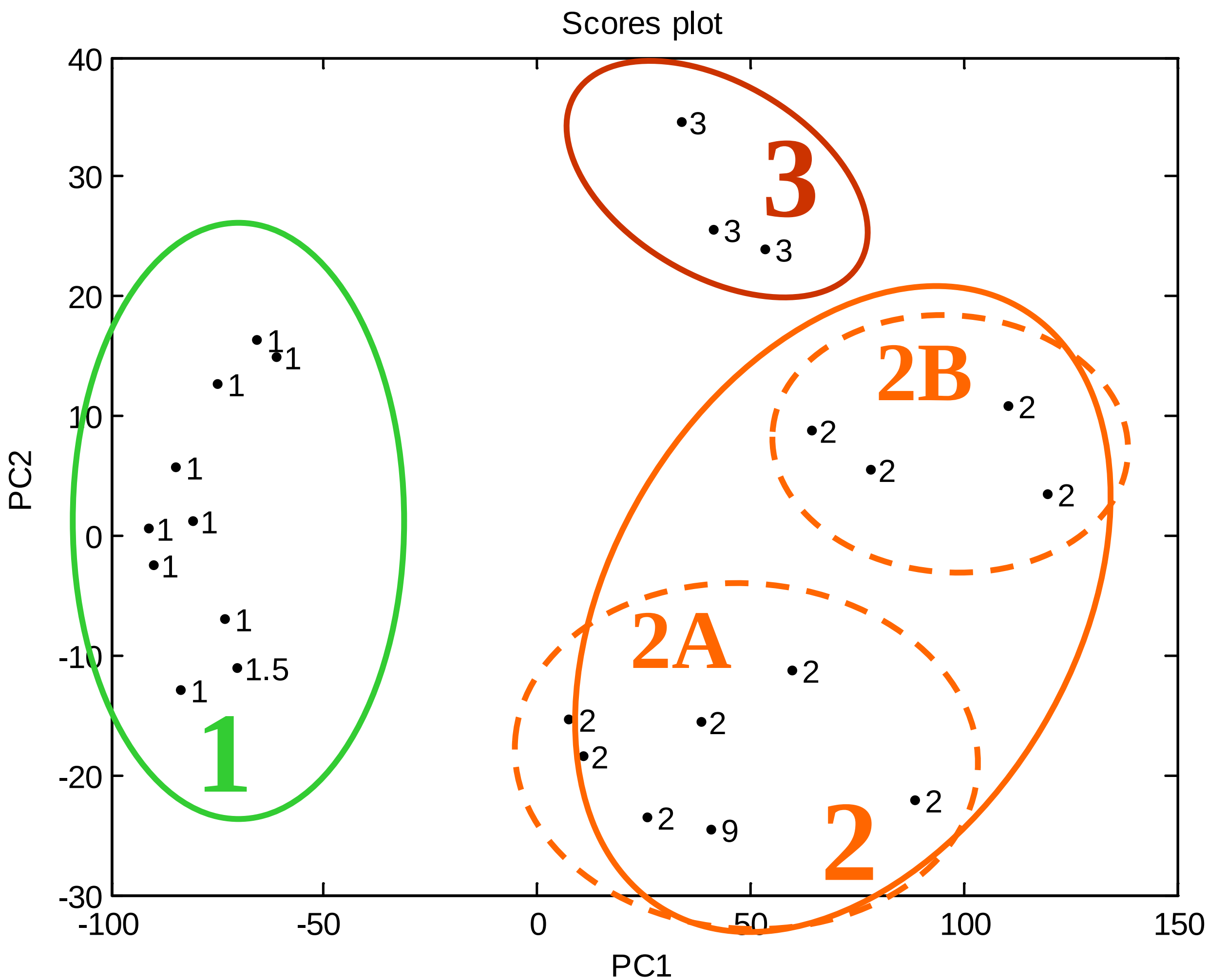
<u>Image set:</u>
24 images, representing different levels of formation quality.

Methods


Paper web
Acquire Image
Raw image
Image pre-processing
Pre-processed image (luminosity corrected)
Wavelet Transform
2D Wavelet transform
Features computation
WTA features profiles
Process Monitoring
Grade evaluation
Status of the process (normal / abnormal)
Quality Grade
(Raw image)
-
(Illumination pattern)
=
(Pre-processed image)

Results (RQ1)

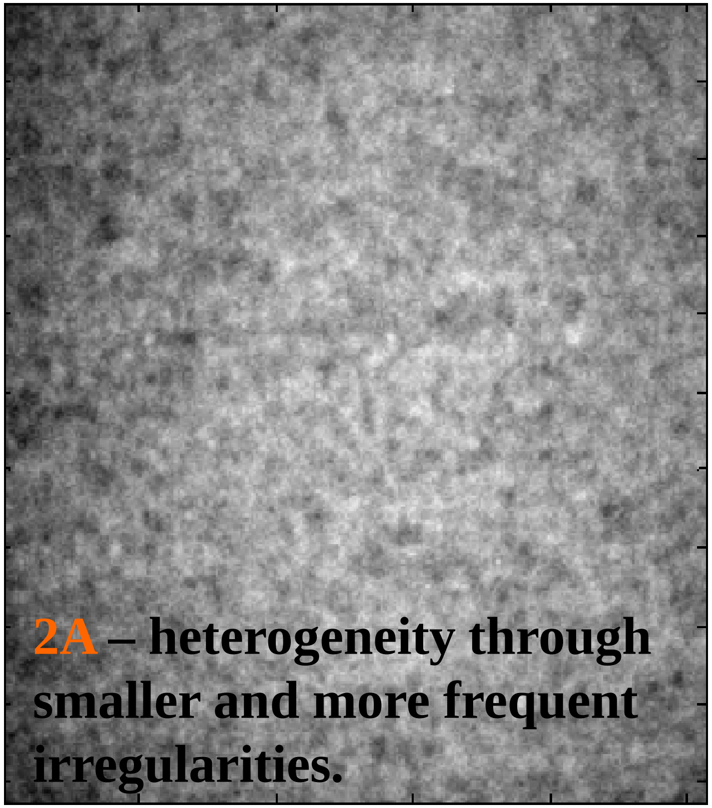
- PCA analysis of wavelet signatures

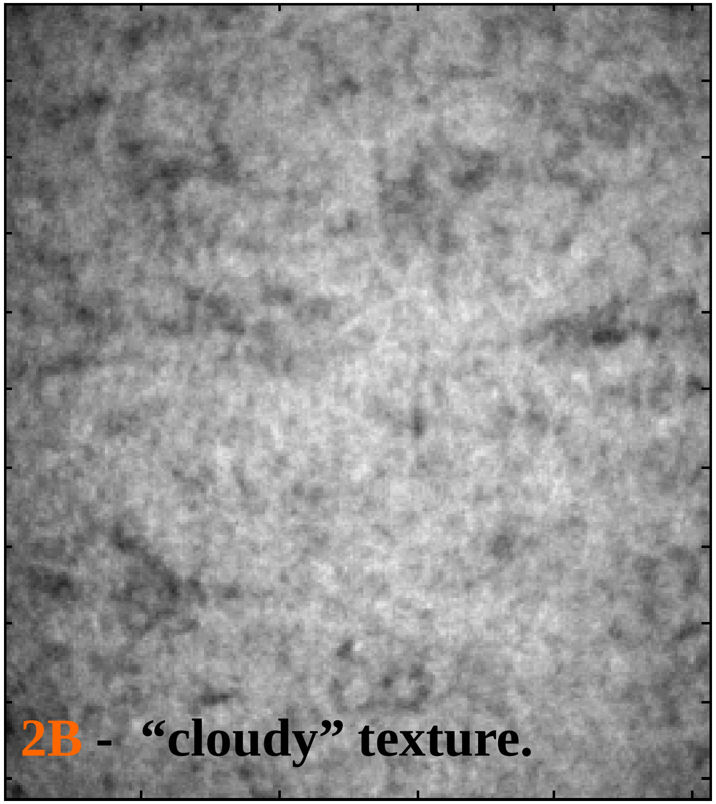


2PC's – 97.96% of the overall variability

Results (RQ1)

- Prototypes of clusters 2A and 2B

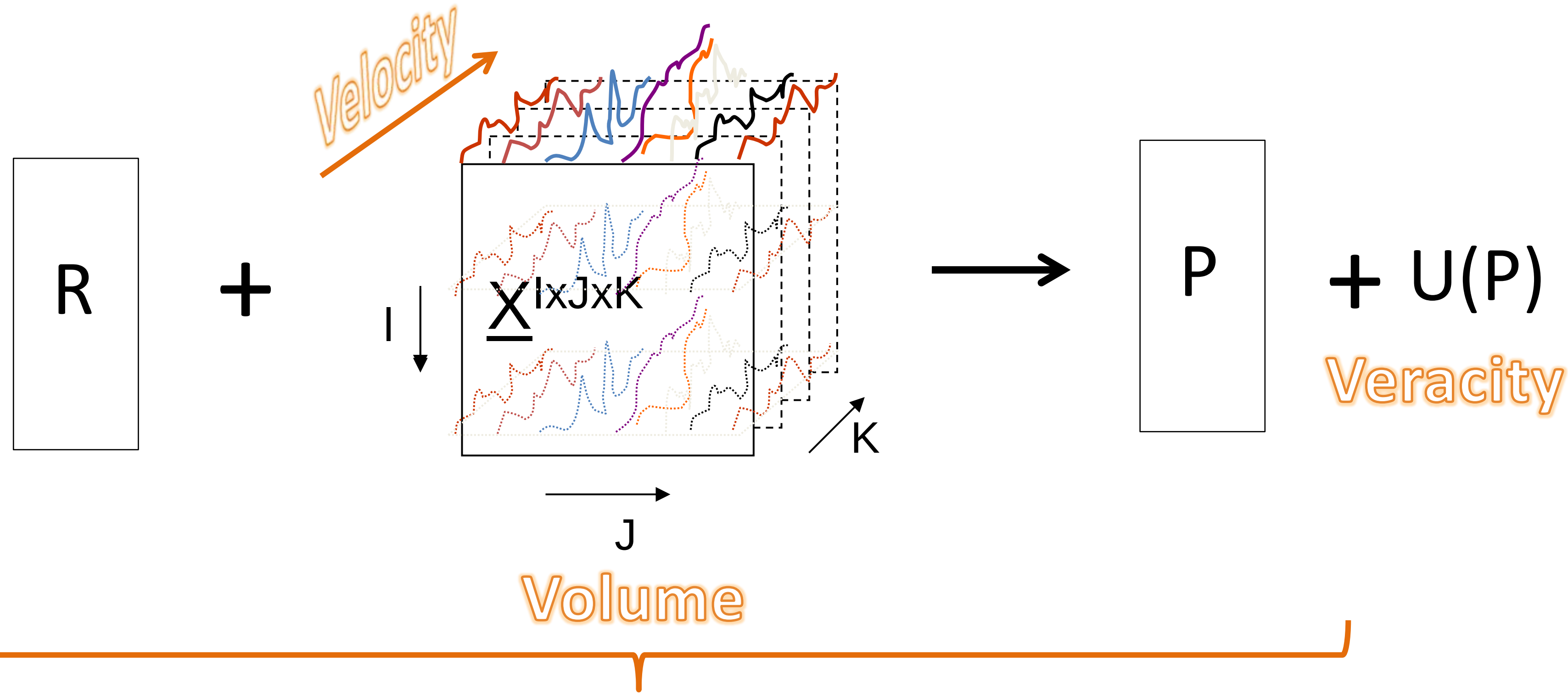




Batch Process Monitoring (BPM)

- Challenges in Batch Processes Monitoring (BPM)

Example: **B**ig Data
Batch Processes



Quantitative and
Qualitative Variables

Sensor data

Spectra
Images
Quality parameters

Variety

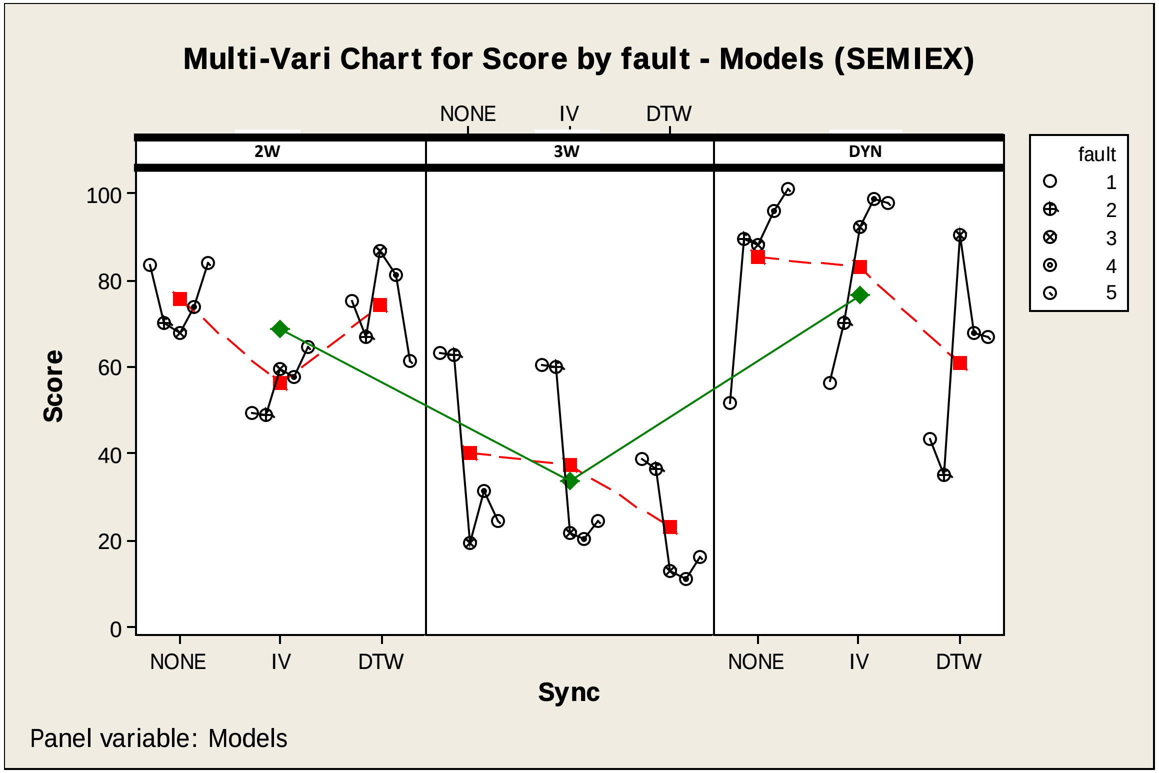
Scope & Motivation

- Batch processes
  - Widely used in industry (high added-value specialties, but also commodities)
    - Semiconductor (~s, min)
    - Chemical and Petrochemical (~hr)
    - Pharmaceutical (~days)
    - Food & Drinks (~hr, weeks, years)
    - (…)
  - Flexible (multipurpose, many degrees of freedom for intervention, scalable to different production ranges)

Scope & Motivation

- Many Batch Process Monitoring (BPM) methods and variants have been proposed:
  - 2-Way
    - Batch-Wise unfolding (Nomikos & MacGregor, 1994, 1995)
    - Variable-Wise unfolding (Wold et al., 1987, 1998)
  - 3-Way
    - PARAFAC (Bro, 1997; Westerhuis et al., 1999)
    - TUCKER3 (Geladi, 1989; Louwerse & Smilde, 2000)
  - Dynamic
    - ARPCA (Choi et al., 2008)
    - BDPCA (Chen & Liu, 2002)
  - Hierarchical (Rännar, MacGregor & Wold, 1998)
  - Local, Evolving (Ramaker et al., 2005)
  - Kernel methods (Lee, J.-M. et al., 2004; Jia et al., 2010)
  - Multiscale (Rato et al., 2015)
  - (...)

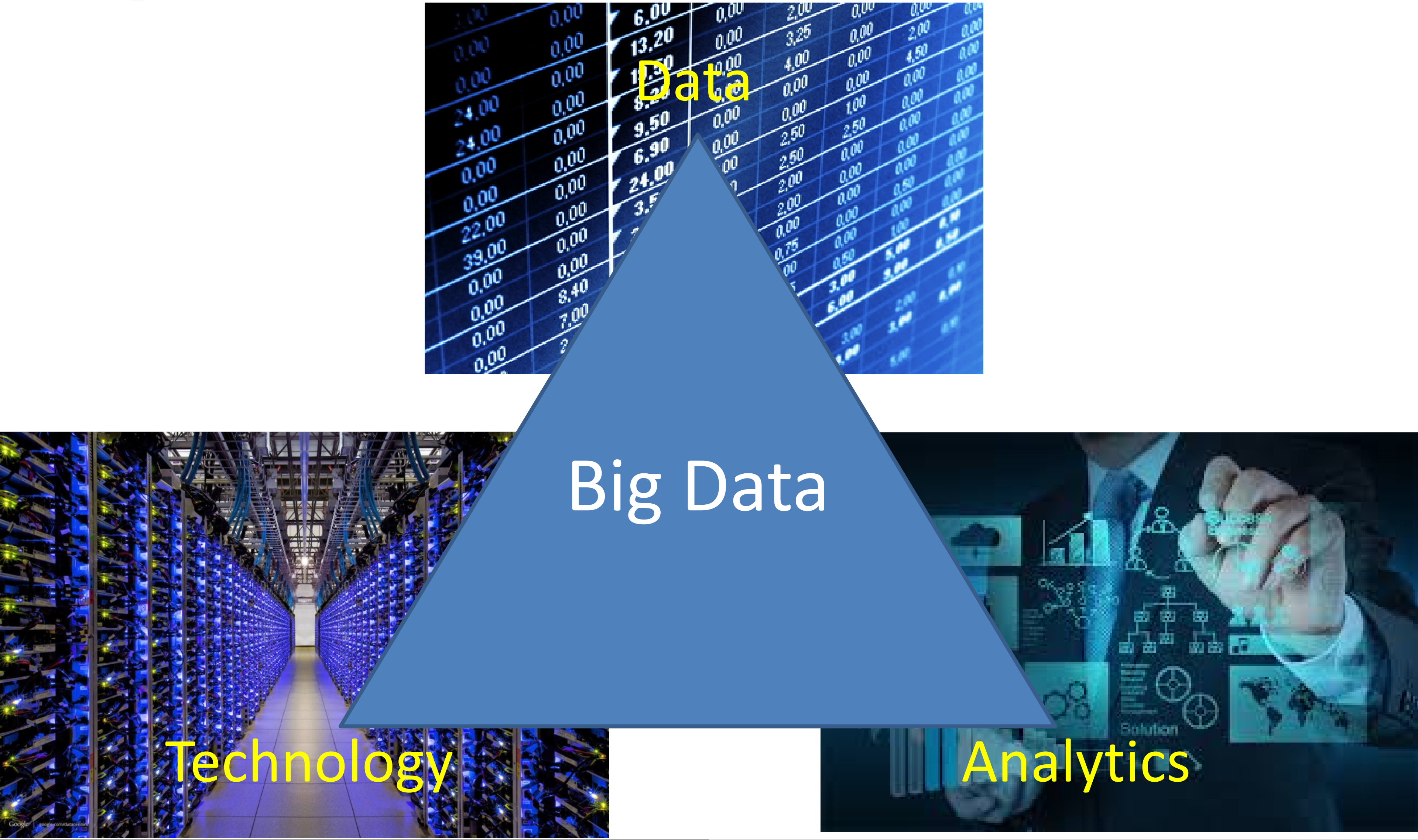
Case study: SEMIEX



SEMIEX / Crossed effects: Methods x Synchronization x Faults

· U C ·

SPC in the big data era



Conclusions

- 90+ years after its introduction, SPC is still an exciting and evolving field!
- SPC should be complemented with effective Diagnosis tools
- New challenges include
  - Handling complex dynamics: multiscale methods
  - Integrating the structure of the system and existing domain knowledge: SET, Bayesian methods
  - Handling multiple data structures (profiles): multi-block methods

and … making everything **simple to use** and **robust**!

Acknowledgements

Tiago Rato
Ricardo Rendall
Veronique Medeiros
Pedro Delgado

PC LAB Process Chemometrics Laboratory

http://www.eq.uc.pt/~marco/research/pclab/