Eindhoven – EURANDOM

November 2005

Benefits of Using the MEGA Statistical Process Control

Santiago Vidal Puig

Eindhoven – EURANDOM

November 2005

- Statistical Process Control: Univariate Charts USPC, MSPC and MegaSPC.

- Strategies for fault diagnosis during the monitoring of a multivariate process.

Eindhoven – EURANDOM

November 2005

Statistical Process Control:

Univariate Charts SPC, MSPC and MegaSPC

Control in a nonstationary world

The stable stationary state is the unnatural one

"If left to themselves machines do not stay adjusted, components wear out and managers and operators miscommunicate and change jobs"

Box G. and Luceño A.

Process Monitoring

Process Adjusting

Left to itself the entropy of any system can never decrease

The second law of thermodynamics

Statistical Process Control

- ***Establish a permanent and intelligent information system over the process evolution :***

Monitoring

  - ***Detect*** *the anomalies at an early stage (special causes)*

Detection

  - *Help to* ***identify*** *the causes of the anomalies*

Diagnosis

  - ***Eliminate the anomalies*** *and prevent their reappearance (or on the contrary incorporate them to the process if they improve its performance)*

Continuous improvement

Statistical Process Control

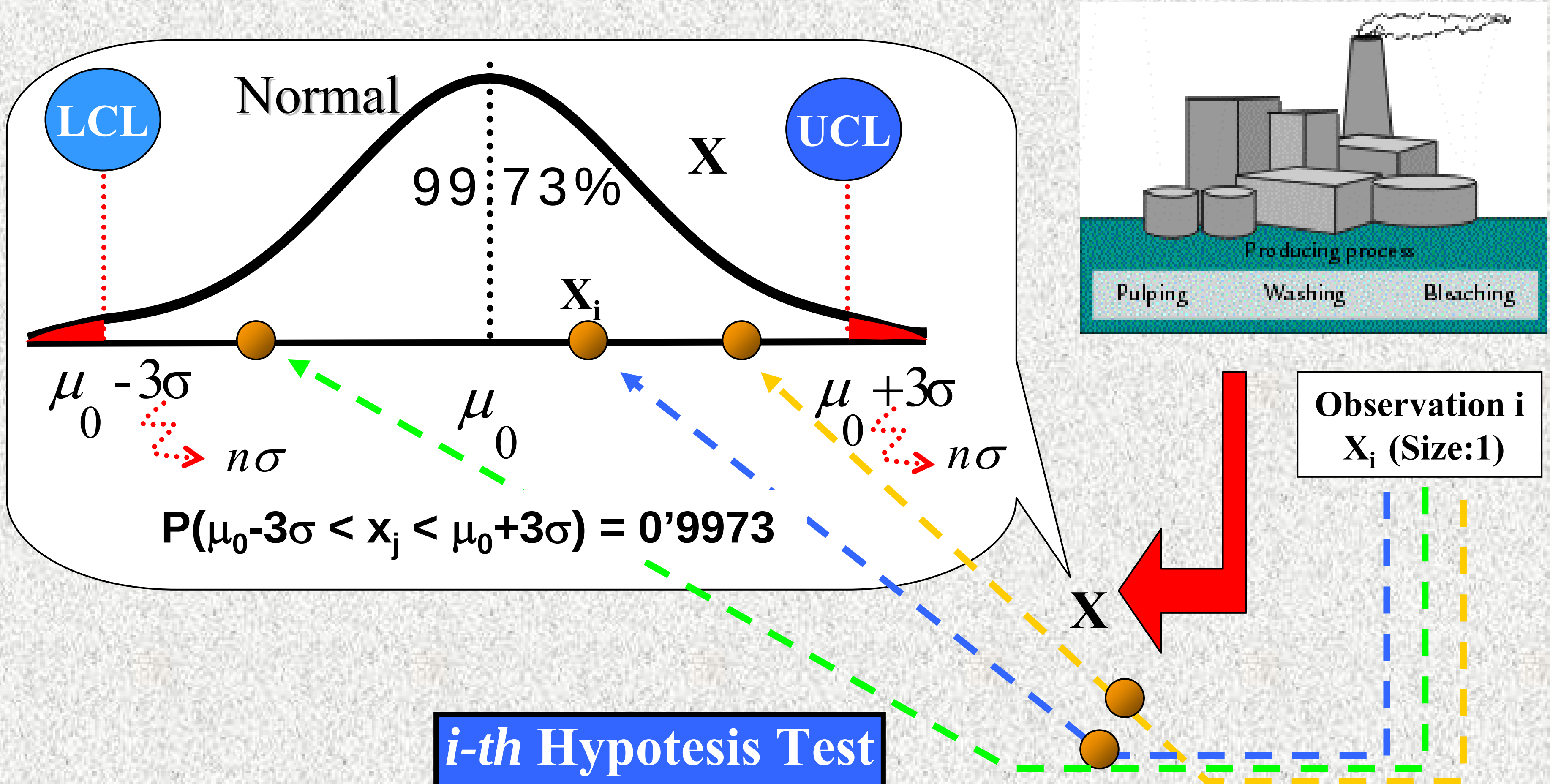
In industrial processes where process and quality variables are measured exist several strategies for statistical process control

Univariate Charts: USPC

Multivariate Charts: MSPC

Megavariate Charts: Mega SPC

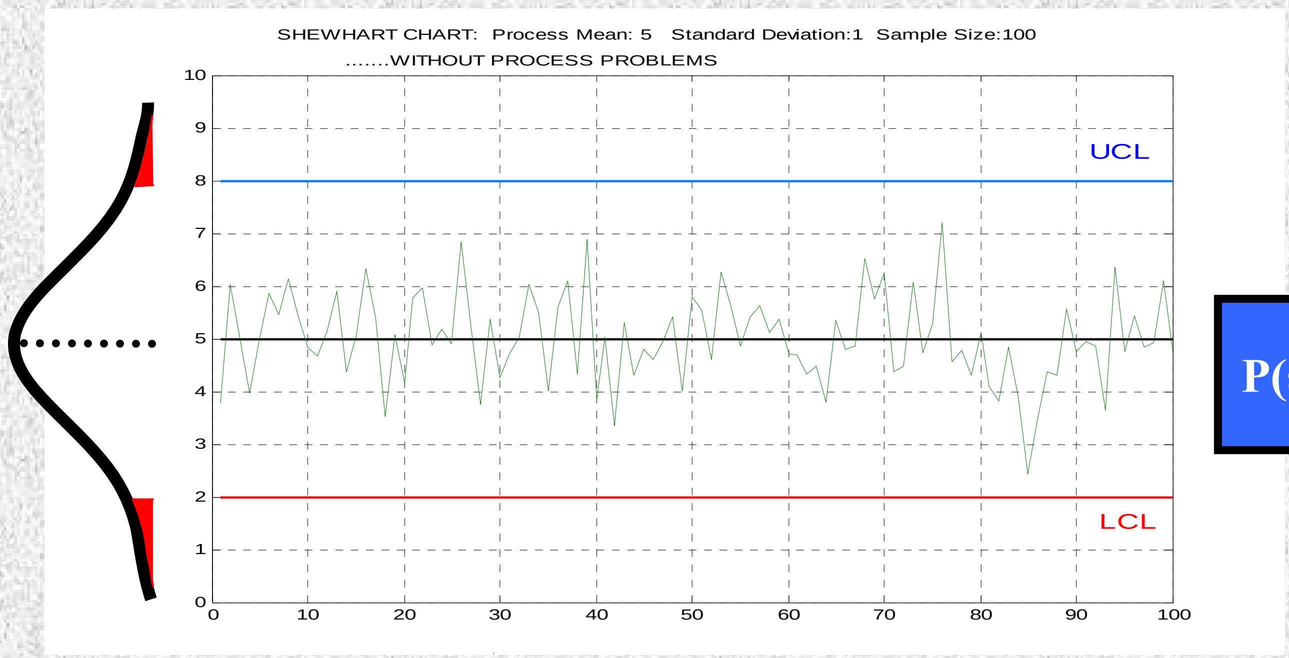
USPC: Shewhart Charts



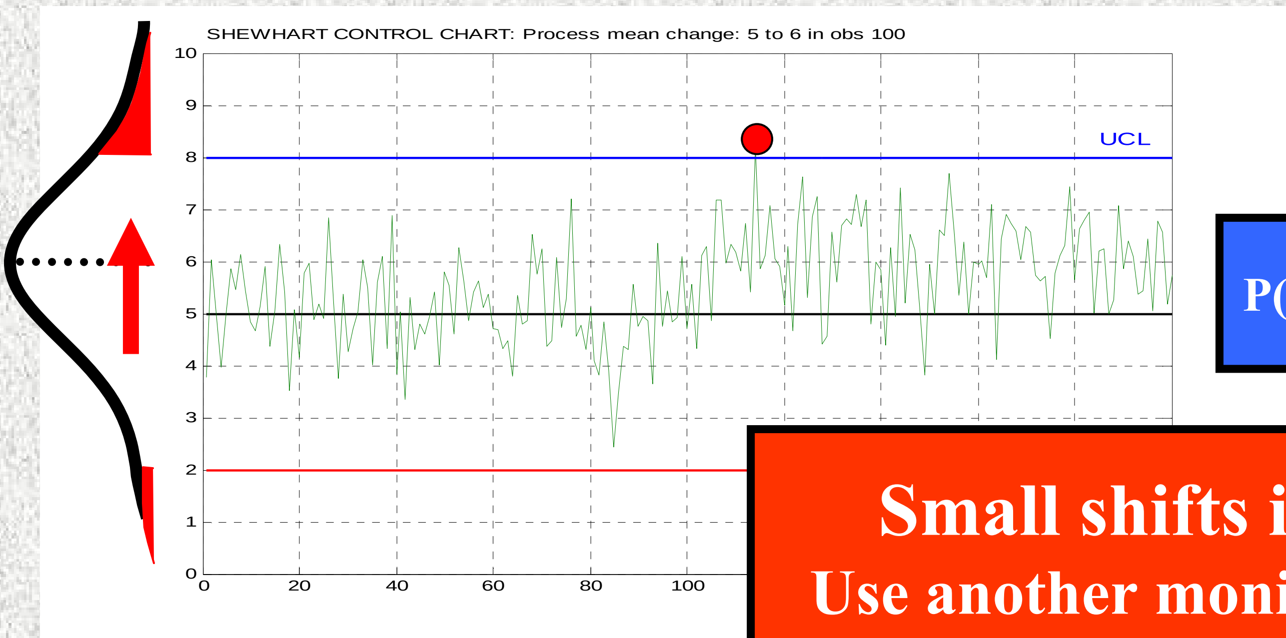
i-th Hypotesis Test

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$



P(signal in chart)= 0.0027



P(signal in chart)= 0.0229

Small shifts in the mean:
Use another monitoring strategies

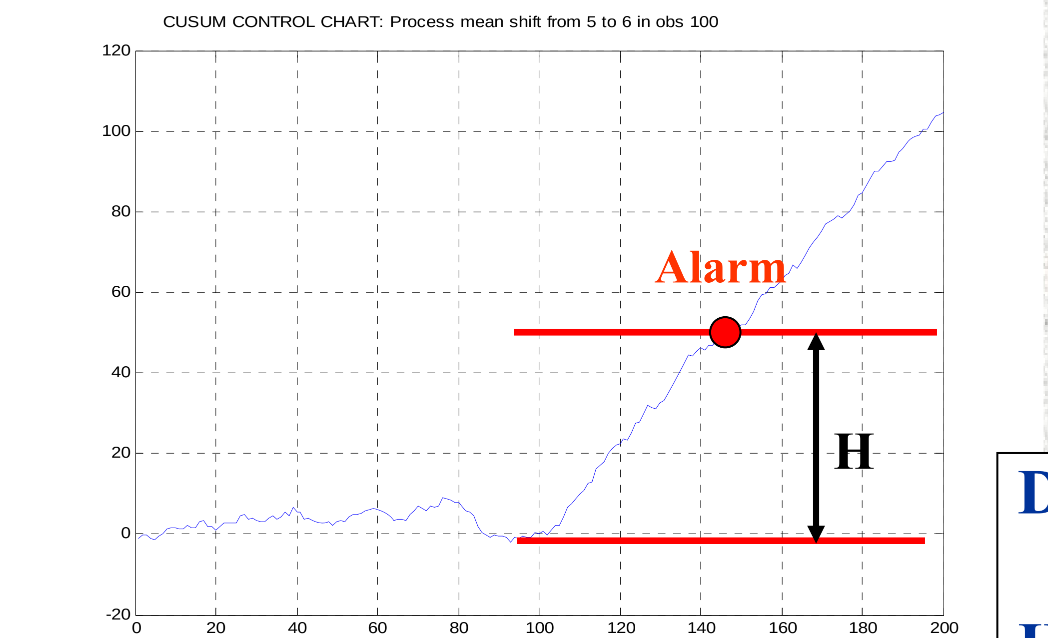
USPC: Cusum Charts

For the individual statistic of interest: x_i (if the sample size = 1): Summing deviations from the target value μ_0 in the sequence of observations

$$S_t = \sum_{i=1}^{t} (x_i - \mu_0)$$

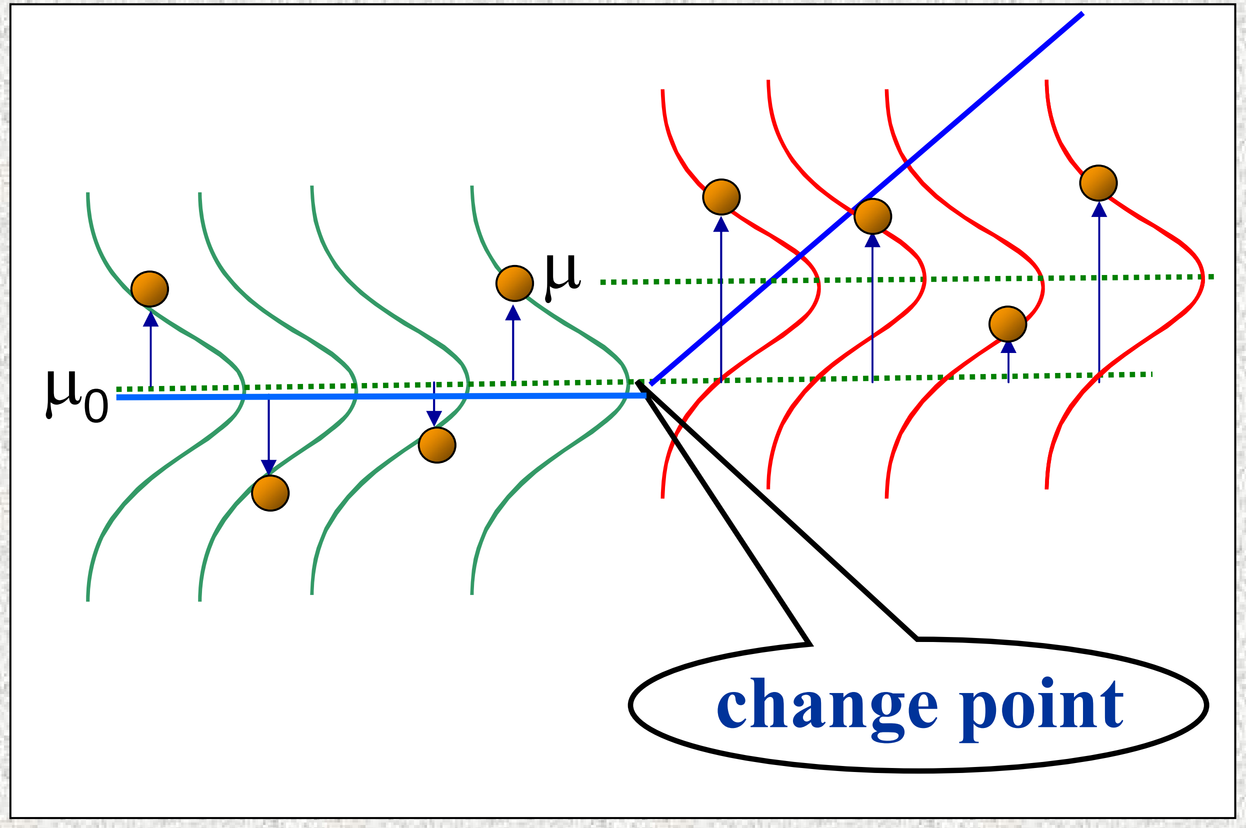
If there is a shift in the mean

Pronounced drift in S_t



D :Difference you want to detect.

H :Decisión interval



CUSUM VARIANTS:
- **Truncated Cusum Charts**
- **Centered Cusum**

USPC: EWMA Charts

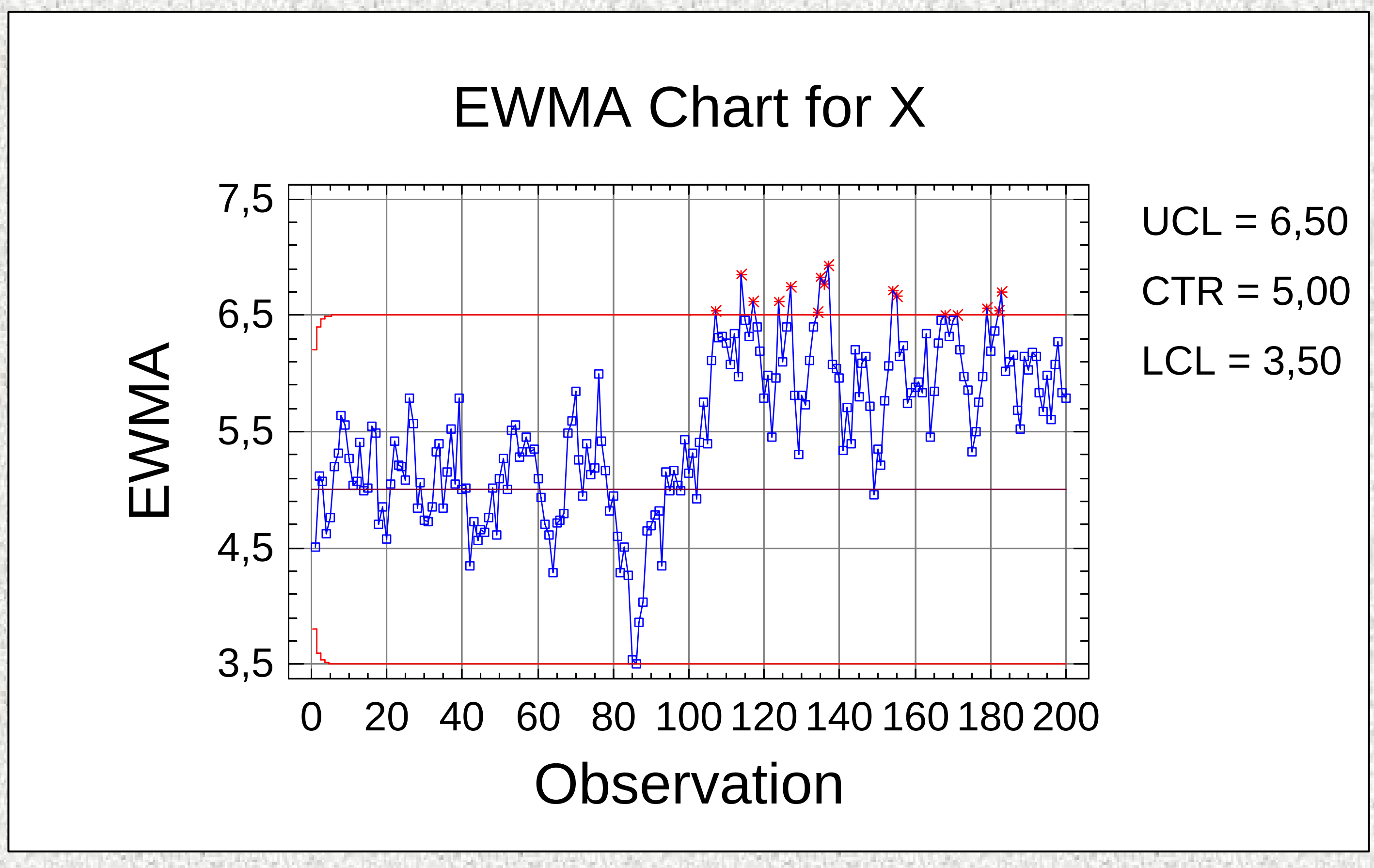
Exponentially weighted moving averages plot:
Past data values are remembered with geometrically decreasing weight

At time t plot the statistic: (sample size=1)

$$\hat{\mu}_t = (1-\lambda)\hat{\mu}_{t-1} + \lambda x_t$$

$$\hat{\mu}_t = (1-\lambda)(x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + ...)$$

$$Var(\hat{\mu}_t) = \sigma^2 \lambda^2 \frac{1-(1-\lambda)^{2t}}{1-(1-\lambda)^2} \ldots \xrightarrow[t\to\infty]{} \ldots Var(\hat{\mu}_t) = \sigma^2 \frac{\lambda}{2-\lambda}$$


EWMA Chart for X
EWMA
7,5
6,5
5,5
4,5
3,5
0
20
40
60
80
100
120
140
160
180
200
Observation
UCL = 6,50
CTR = 5,00
LCL = 3,50

Measuring the Charts Performance

Average Run Length: Nº observations expected before to get a signal
ARL

$ARL_{in\ control\ -\ \mu_0}$

$ARL_{out\ control\ -\ \mu}$

▲ **Expected nºobservations for a false alarm**

▼ **Expected nº observations to detect the problem**

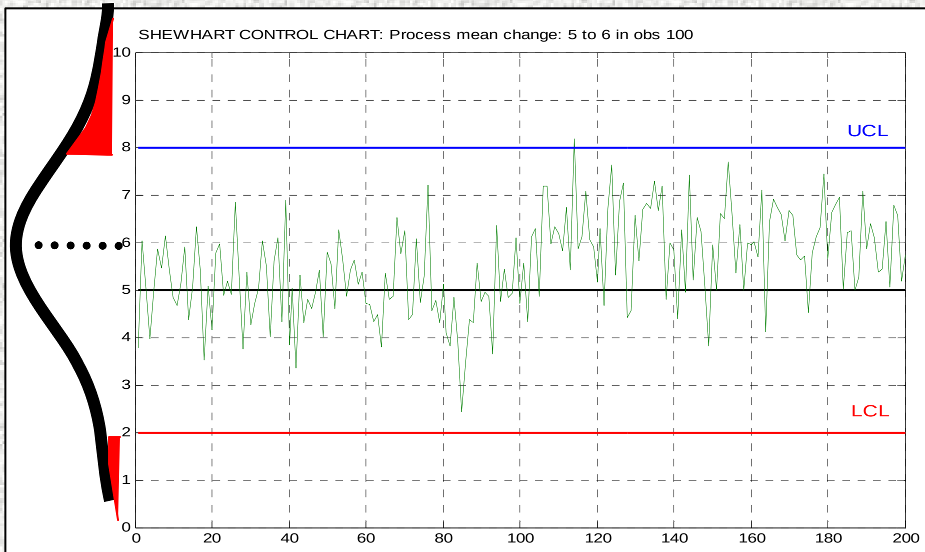
Run Length

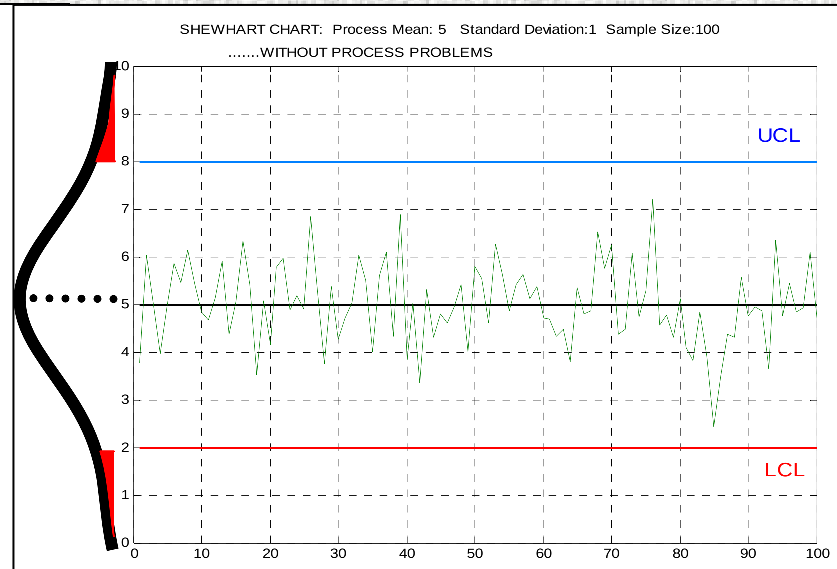
Geometric distribution (P)

P= Probability of a chart signal

ARL = E(RL) = 1/P

Example





ARL **out control - μ** = 6

$$P = P(N(\mu\ ;\sigma) \notin [\mu_0 - 3\sigma \quad \mu_0 + 3\sigma])$$

P=0.0229 **ARL = 1/P= 44**

ARL **in control - μ_0** = 5

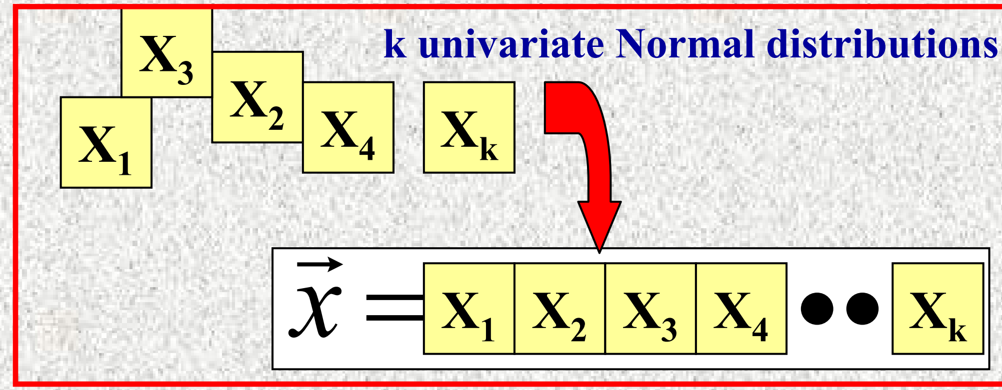
$$P = P(N(\mu_0;\sigma) \notin [\mu_0 - 3\sigma \quad \mu_0 + 3\sigma])$$

***P*=0.0027** **ARL=1/*P*=370**

Multivariate Charts: MSPC

There are different kind of charts that may be used:

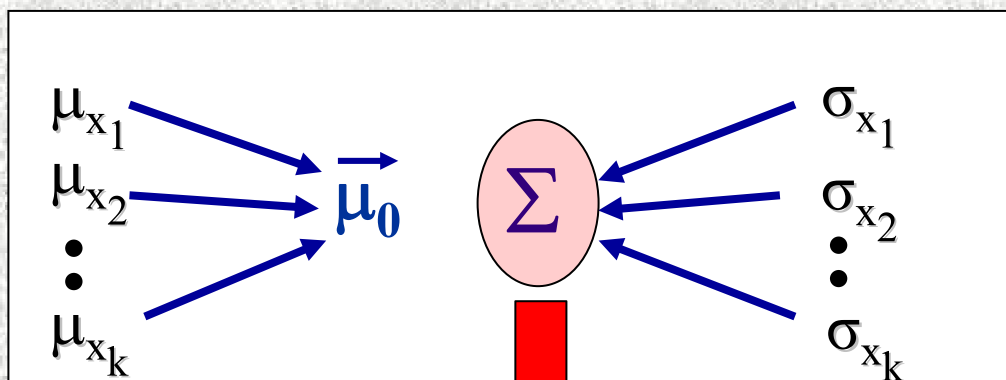
- T^2 Hotelling
- MCusum
- MEWMA



Multivariate Normal Density Function

$$f(\vec{x}) = k e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_0)'\overline{\overline{\Sigma}}^{-1}(\vec{x}-\vec{\mu}_0)}$$

k th-dimensional Multivariate Normal



Likelihood: "a prior" probability of the observation

$$D_{Mahalanobis} = (\vec{x}-\vec{\mu}_0)'\overline{\overline{\Sigma}}^{-1}(\vec{x}-\vec{\mu}_0)$$

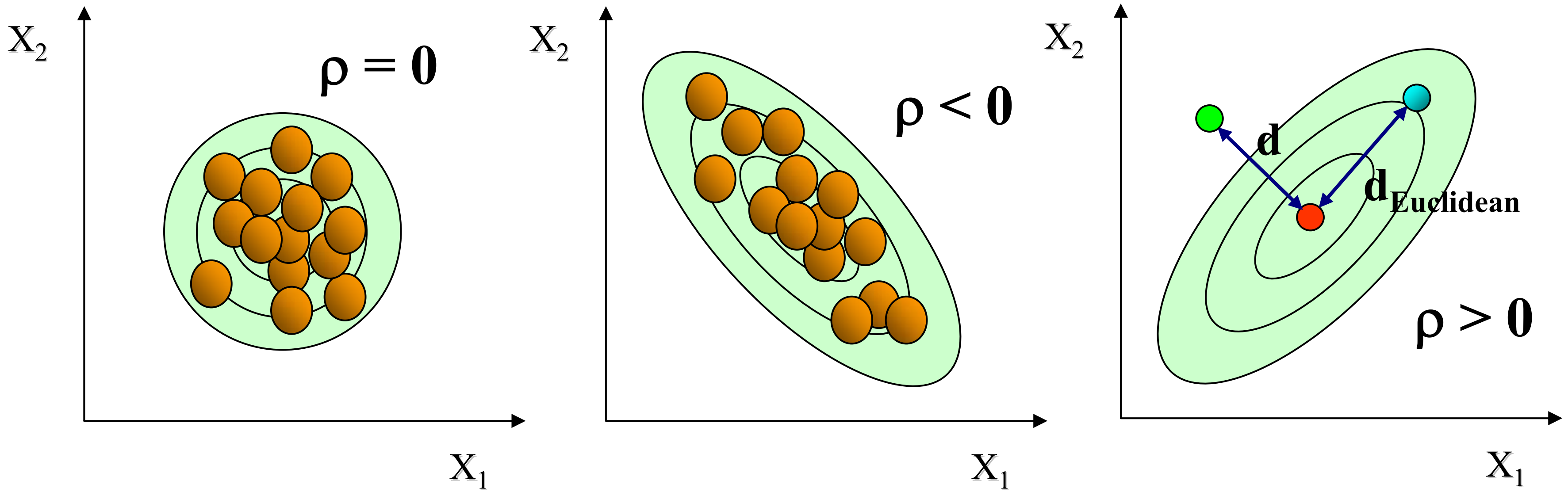
Covariance Matrix

σ_{x_i} $\sigma_{x_i x_j}$

D. Mahalanobis

Stastistical distance: ***Probability that the new observation would differ*** **a certain** **<u>euclidean distance</u> from the process mean in a certain <u>direction</u> of the space**

Example: Bivariate case

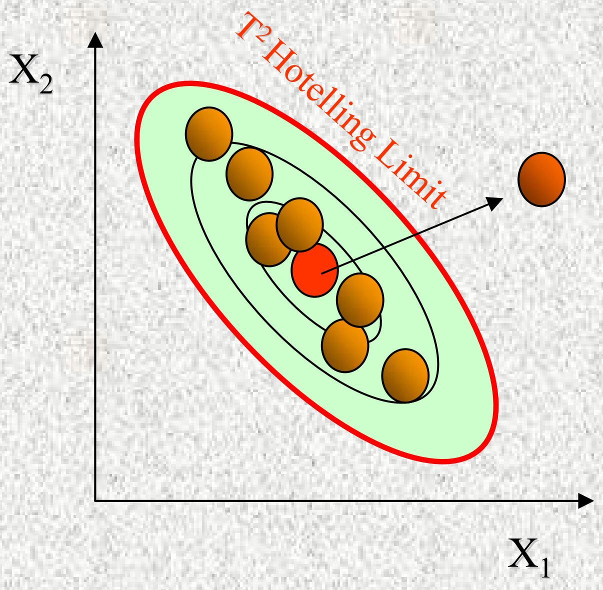


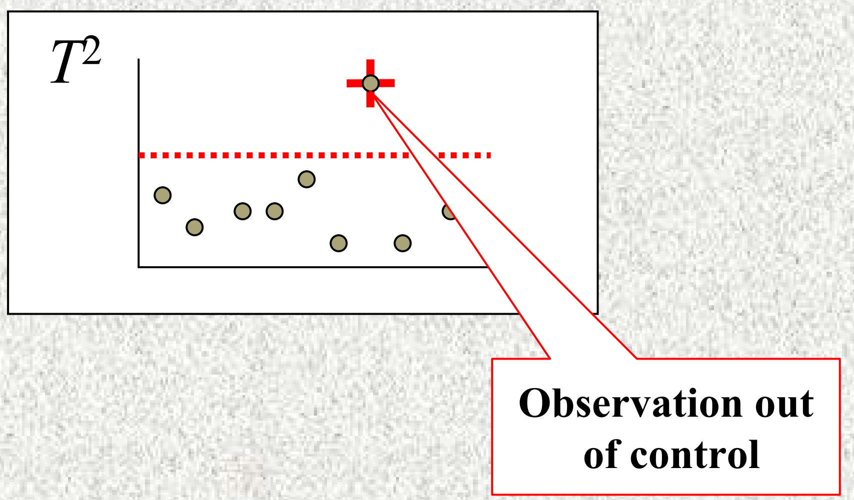
MSPC: T^2 Hotelling

$$D_{Mahalanobis} = (\vec{x} - \vec{\mu}_0)' \overline{\overline{\Sigma}}^{-1} (\vec{x} - \vec{\mu}_0)$$

$$T^2_{Hotelling} = (\vec{x} - \vec{\mu}_0)' S^{-1} (\vec{x} - \vec{\mu}_0)$$

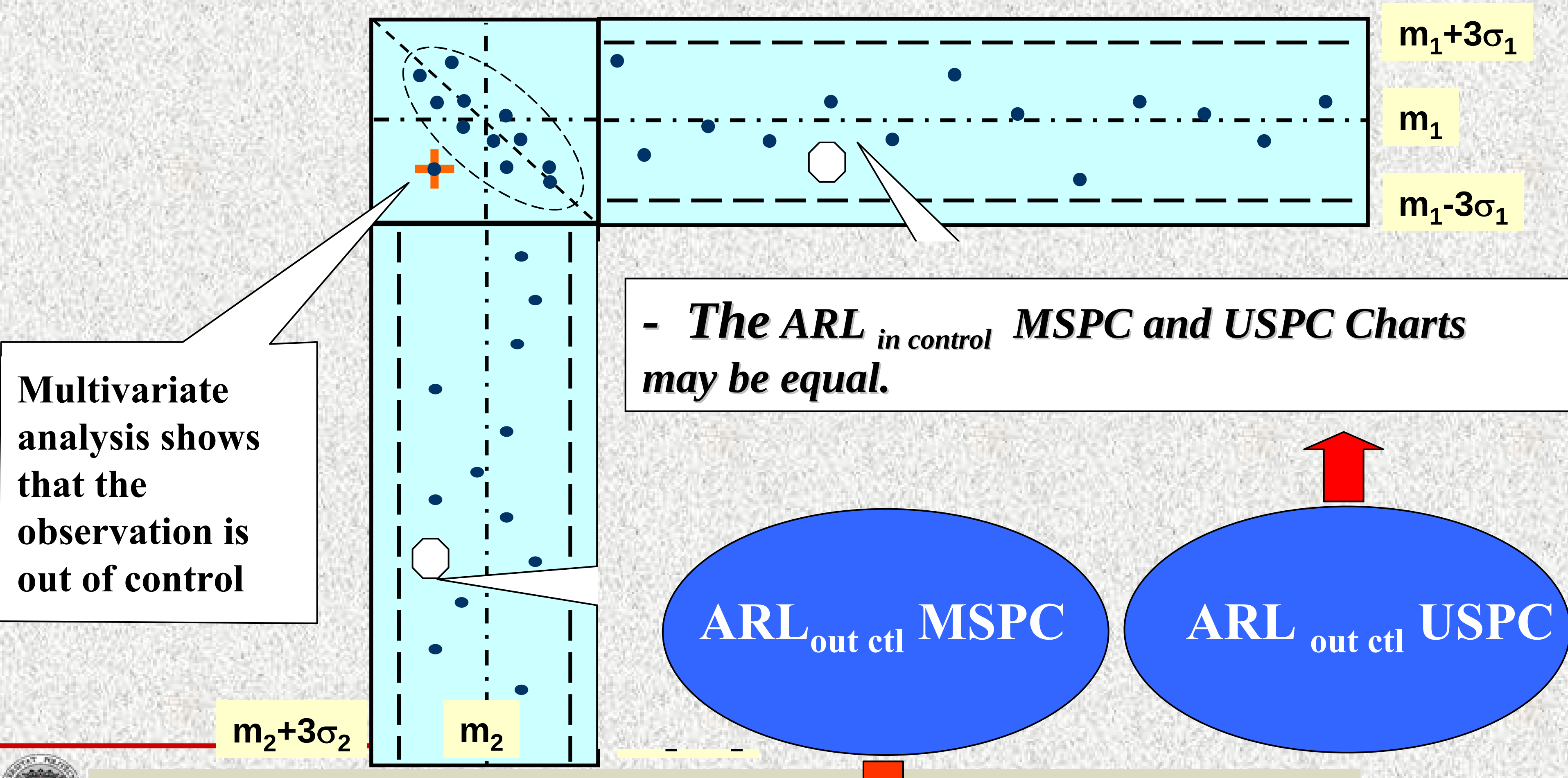
T^2-Hotelling: *Estimated* D-Mahalanobis



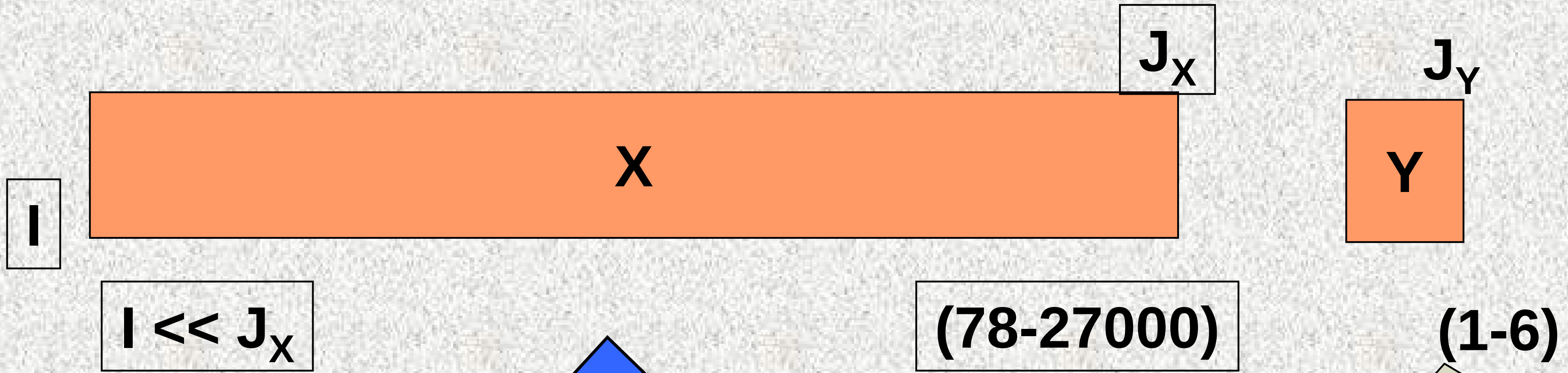


SPC on univariate charts: limitations

- ***Quality is often a multivariate property***
- ***Univariate control charts (ignores correlation)***



Nature of data: evolution



Data-rich environments

- *Hundreds of process variables*
- *Measured on-line (sensors)*
- *High sampling rate (seconds-hours)*
- *High-dimensional data*
- *Highly collinear data*
- *Missing data problems*
- *Limiting cost: sampling*

Data-poor systems

- *Few quality properties*
- *Measured off-line (Lab or manually)*
- *Low sampling rate (hours-days)*
- *Low-dimensional data*
- *Slightly correlated data*
- *Limiting cost: analysis*

Standard SPC

How to adapt SPC for data-rich environments?

J

I

I >> J

Standard SPC
uni&multivariate

Multivariate projection methods: PCA, PLS

I

I << J

Data structure

A

Compression

I

I > A

Discovering
latent structures

Process is driven by a few underlying common cause events

Latent variables space strategies

Principal Component Analysis (PCA)

Compresses the information

↓

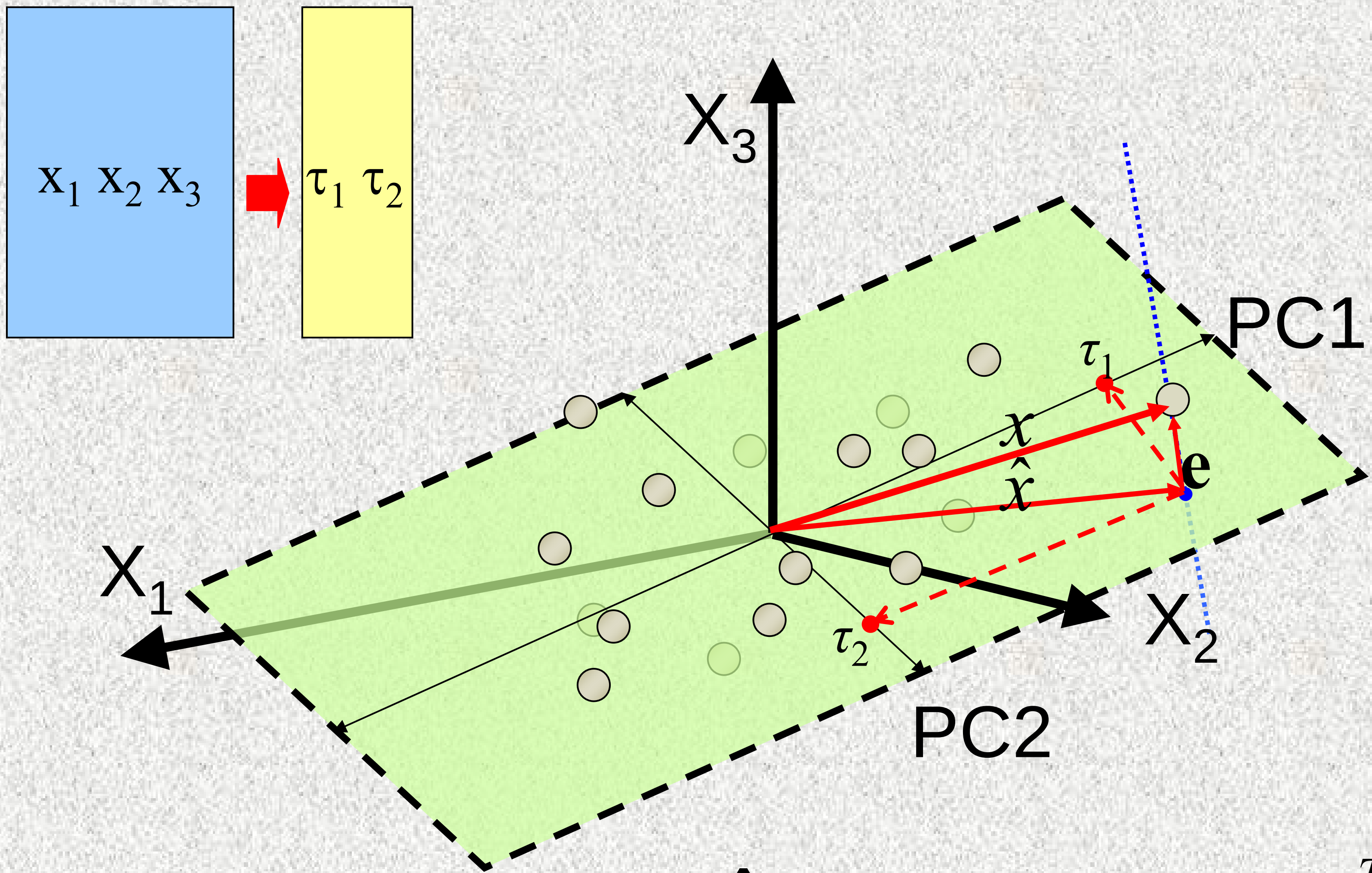
Reduced number of independent latent variables

The observations are projected into a space of A dimensions.

$$\mathbf{X} = \sum_{a=1}^{A} t_a \mathbf{p}_\mathbf{a}^\mathbf{T} + \mathbf{E}$$

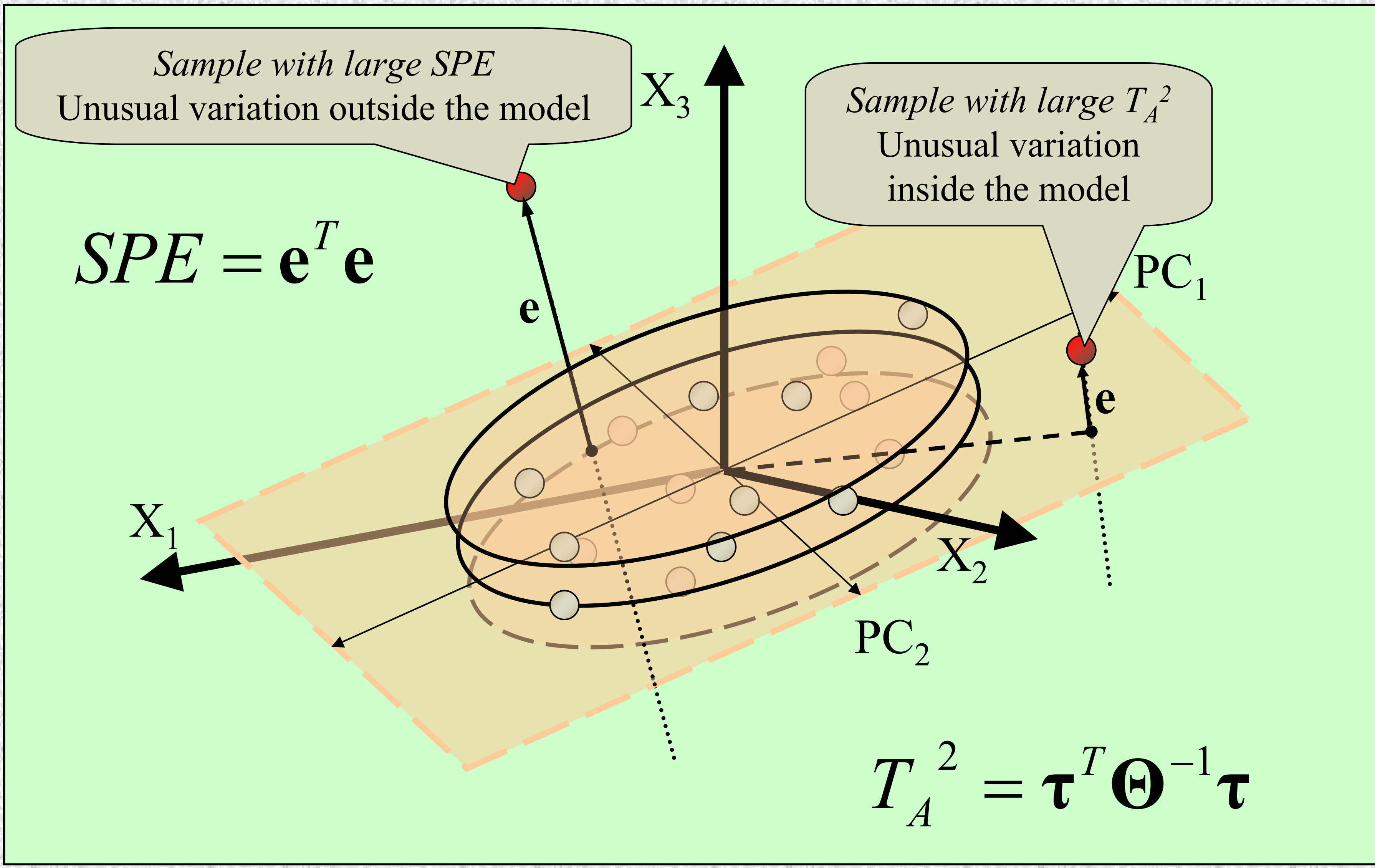
- These new variables explain most of the process variability.
- Residual information not captured by the model (E).

PCA Model

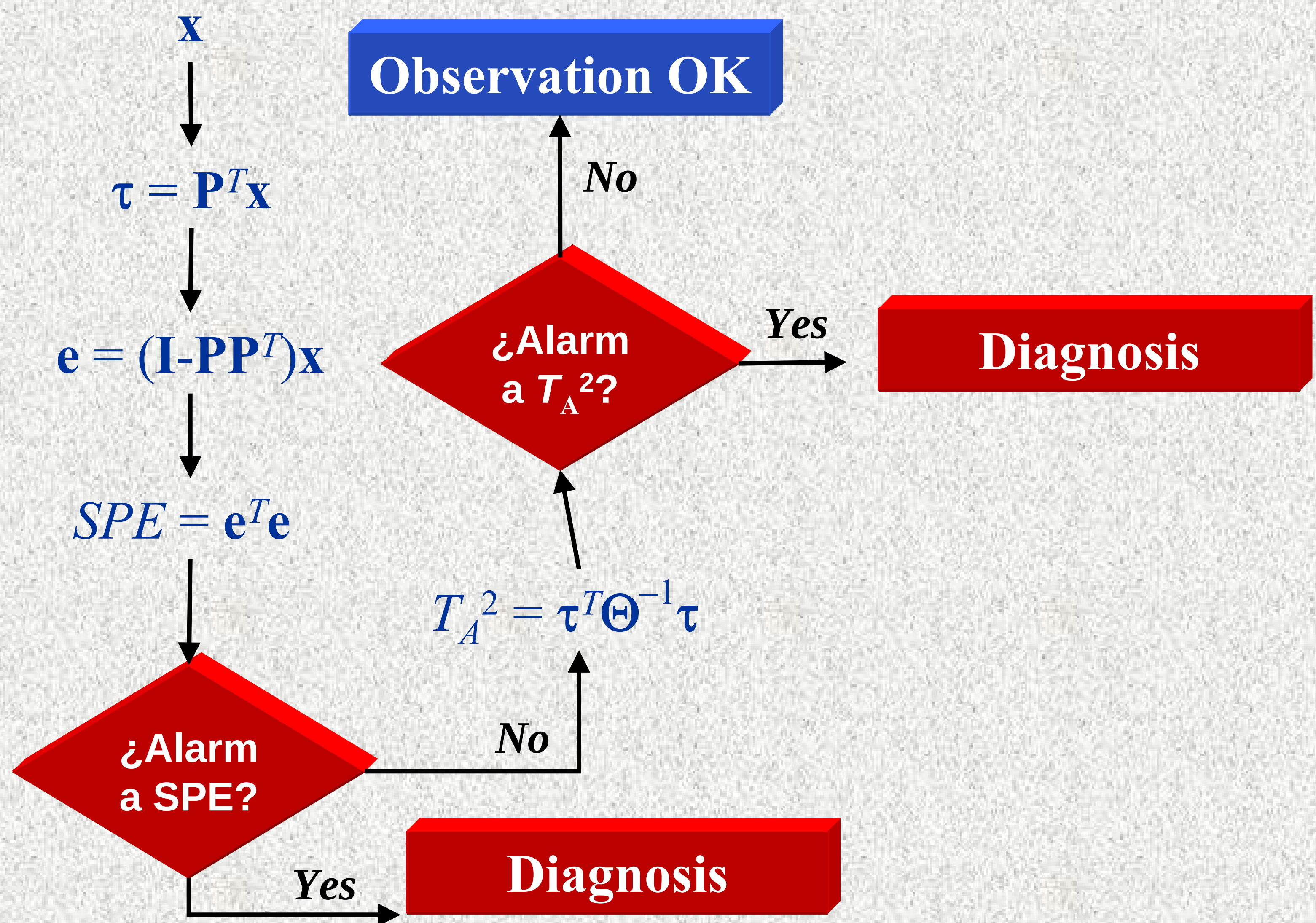


$$x = \hat{x} + e = \mathbf{P}\boldsymbol{\tau} + \mathbf{e} = PP^T x + e$$

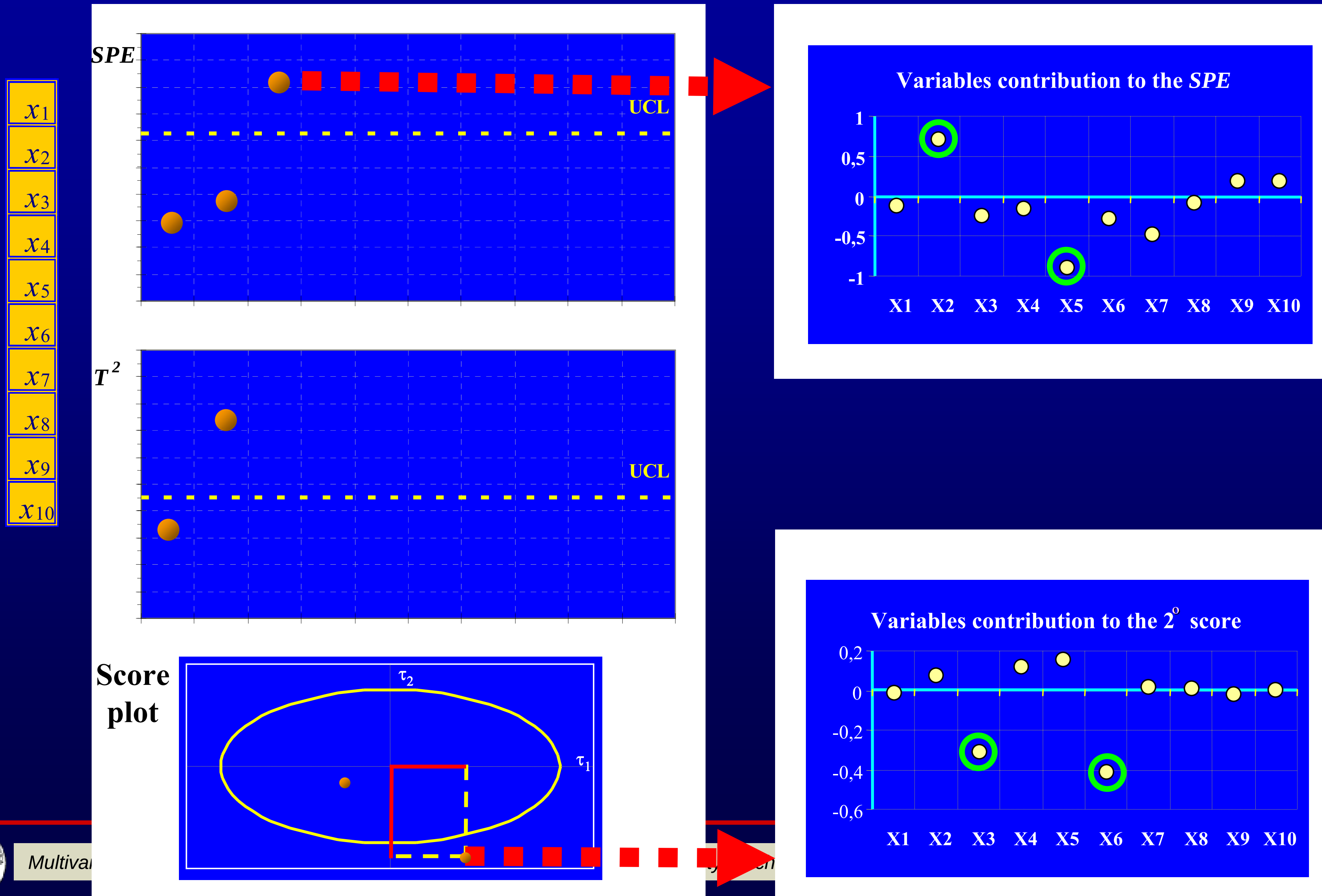
Out of Control Observations



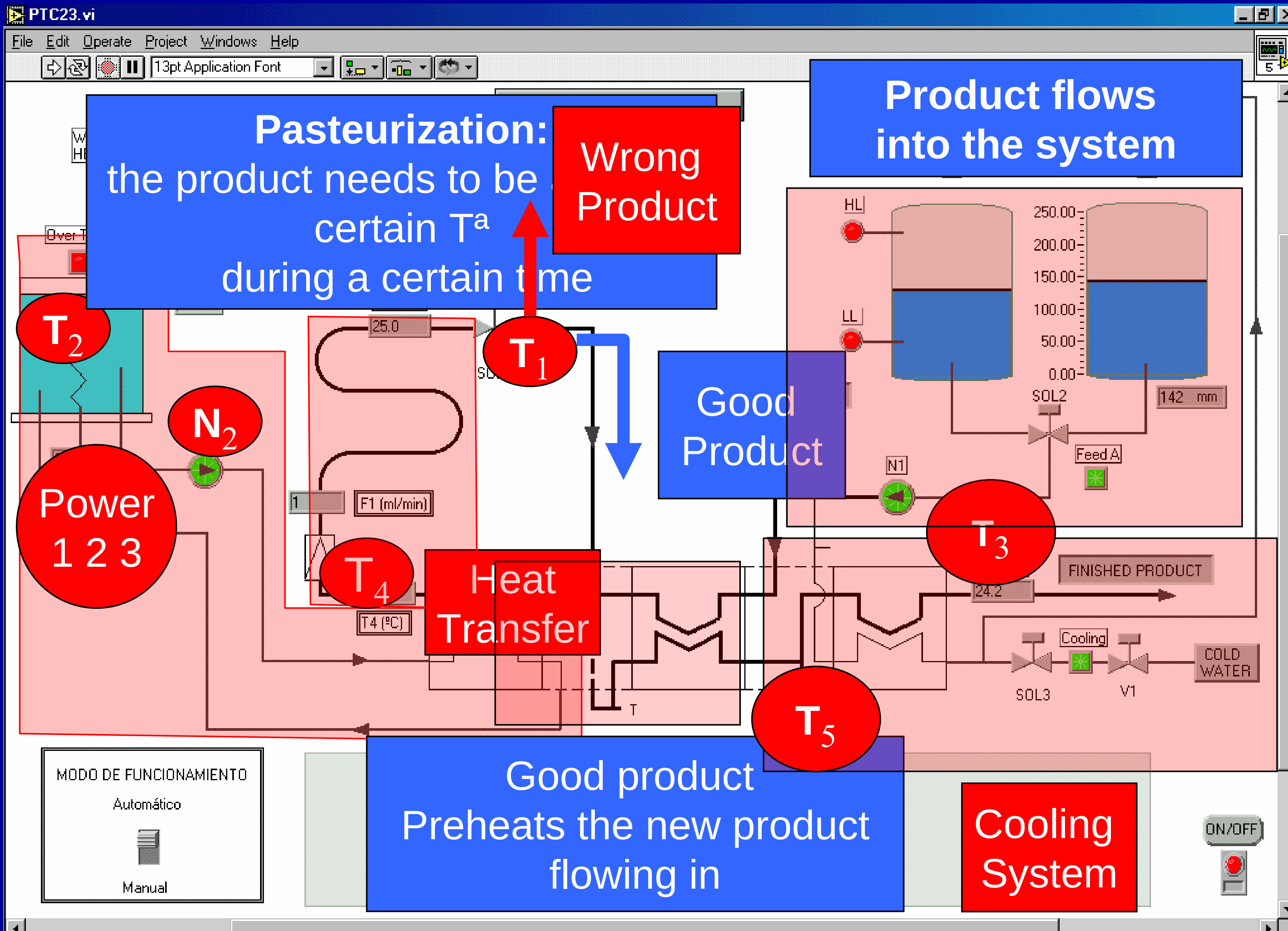
Monitoring with PCA



New Observations Monitoring Schedule



PASTEURIZATION PROCESS



PROCESS DESCRIPTION

Variables

1. **Tank Level**
2. **T_1 :** Temperature after the curved pipe
3. **T_2 :** Temperature of the heated water
4. **T_3 :** Temperature of the final product
5. **T_4 :** Temperature before the curved pipe
6. **T_5 :** Temperature of final product used to preheat
7. **Flow**
8. **SP Flow**
9. **Power 1**
10. **Power 2**
11. **Power 3**

These three variables measure some aspect of the power used to heat the water

12. **% Pump1:** control the flow speed of product
13. **% Pump2:** control the flow speed of the heating water

Loop Controls

1. SetPoint $T_2 \Rightarrow$ Power
2. SetPoint $T_1 \Rightarrow$ Pump 2
3. SetPoint Flow $\Rightarrow$ Pump 1

$T^2 \Longleftrightarrow T^2_A$, SPE

$$T^2 = T_A^2 + \sum_{k=A+1}^{K} \frac{t_i^2}{\lambda_i}$$

$$SPE = \sum_{k=A+1}^{K} t_i^2$$

Data Set	1	2
Total Number of observations	13249	7505
Deleted observations original (T^{2})	6026	1084
Deleted observations latent (T_A^2 , SPE)	4215	18
Deleted by T^2_A	3417	0
Deleted by T^2 y T^2_A	3383	0

Hotelling T^2: noisy & ↑false alarms
T^2_A : stable

CONCLUSIONS

- In a multivariate process with correlations between the measured variables, MSPC performs better than USPC
- In a multivariate process with colinearity problems or a bad conditioned sample covariance matrix. MegaSPC performs better than MSPC (T^2_A is less noisy than T^2)
- In a multivariate process with missing data or where the number of variables is larger than the number of observations we have only a choice: the MegaSPC

Eindhoven – Eurandom

October 2005

Strategies for fault diagnosis during the monitoring of a multivariate process

Santiago Vidal Puig

UNIVERSIDAD POLITÉCNICA DE VALENCIA

Outline

Fault Diagnosis

- Introduction

- Strategies based on the original variables space.

- Strategies based on the latent variable space

- Future Objectives

Fault Diagnosis: Introduction

Process Fault: Special causes affecting the process being monitored.

Sensor Fault: Special causes affecting the measurement of any process variable being monitored.

Process Fault → Statistical diagnosis of the fault

↓

Aim to the suspected responsible variables

Process Engineers → Assign the Fault

Fault Diagnosis: MSPC Approach

Based on non-causal empirical correlation models

Models built from Normal plant operating data (NOC) when only commom causes of variation are present.

Fault Diagnosis: Strategies

Several strategies have been proposed in the last 15 years.

Strategies based on the *original variables* space

Monitor the T^2 Hotelling on the original variables

$$T^2 = \mathbf{x}^T_{\mathbf{new}} \mathbf{S}^{-1} \mathbf{x}_{\mathbf{new}}$$

Strategies based on the *latent variables* space "projection methods"

Monitor the T_A^2 on A latent variables

$$T_A^2 = \boldsymbol{\tau}^T \boldsymbol{\Theta}^{-1} \boldsymbol{\tau}$$

Monitor the SPE(square prediction error)

$$SPE = \sum_{j=1}^{K} e_{ij}^2 = \sum_{j=1}^{K} (x_{ij} - \hat{x}_{ij})^2$$

Strategies based on the original variables space

Doganaksoy, Faltin and Tucker Method (1991)

Murphy
Hawkins
Hayter
.
.
Montgomery

Mason, Tracy and Young Method (1995)

Doganaksoy, Faltin and Tucker Method

Creates a ***Ranking*** of the variables according to their probability of participation in the detected change.

$$t = \frac{\overline{X}_{i,new} - \overline{X}_{i,ref}}{\left[S_{ii} \left(\frac{1}{n_{new}} + \frac{1}{n_{ref}} \right) \right]^{1/2}}$$

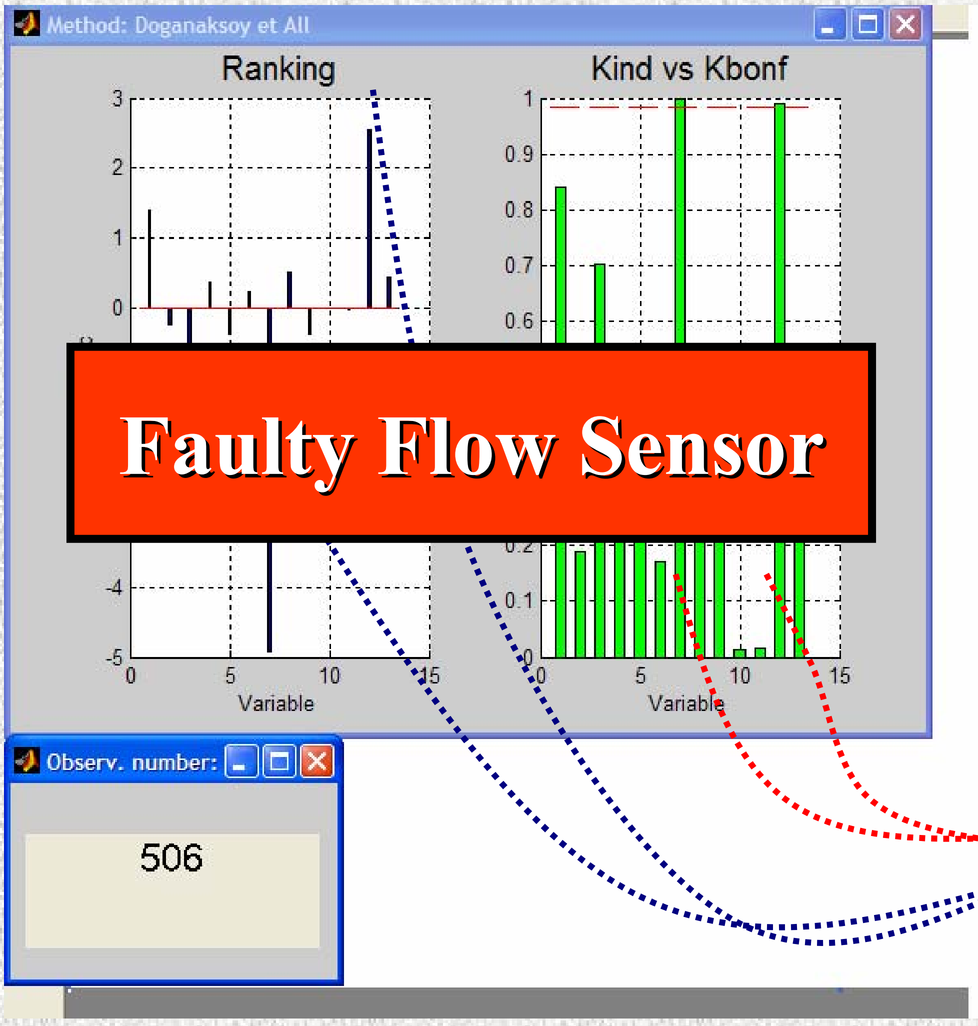
Use **univariate t** statistics for the difference of means

Guide of highly suspected variables

Weak points:

- Does not use the information provided by the **correlation structure** among the variables in the diagnosis.

Doganaksoy method in a simple fault



$T^2_{Hotelling}$ = 4199
UCL = 4.2841

ALARM!!

Variable 7: Flow
Variable 12: Pump 1

Fault in the flow sensor in the main system (the value fall down)
Signal in pump 1 signal due to the feed back control (increasing the pump1 %)

Mason, Tracy and Young Method

Orthogonal descomposition of the T^2 Hotelling

- Terms with a T^2 structure
- Straightforward interpretation
- Known distribution (F Snedecor)

Unconditional Terms

T_i^2

- Marginal contribution of each variable to the T^2
- Detects changes in the operational values of the variable without considering the correlation structure.

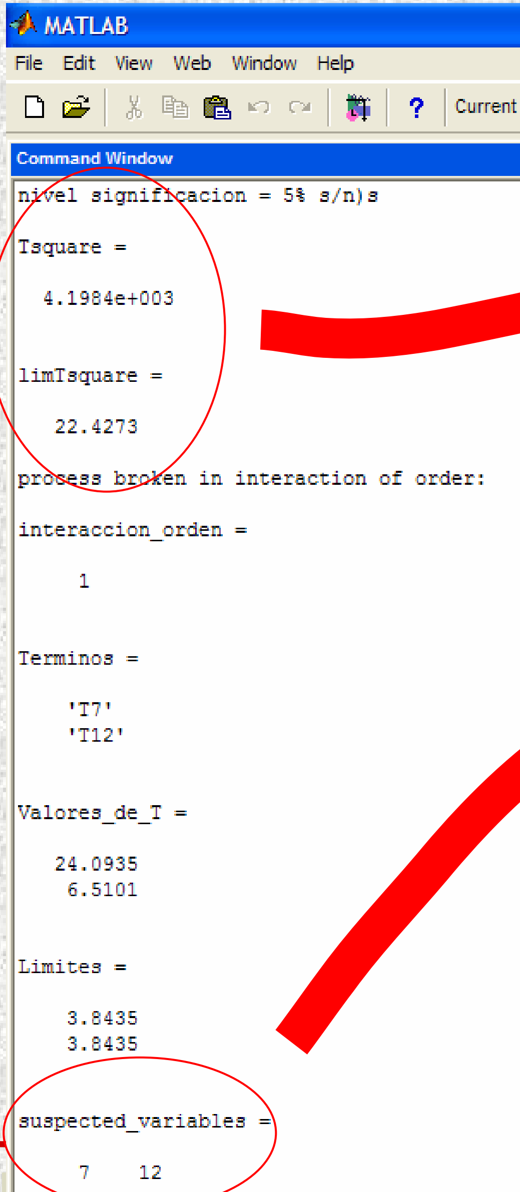
Conditional Terms

$T_i^2|_{1.2..i-1.i+1..p}$

- Contribution of each variable to the T^2 after being adjusted by regression.
- Detects changes that break the correlation structure

Weak Point: The number of terms grows very fast with K

Mason *et al* method in a simple fault



ALARM!!

Variable 7: Flow
Variable 12: Pump 1

Faulty Flow Sensor or Pump 1

Fault Diagnosis in the latent variable space

When a ***a problem is detected*** we must find the cause identifying some original variables as suspicious or indicating the fault.

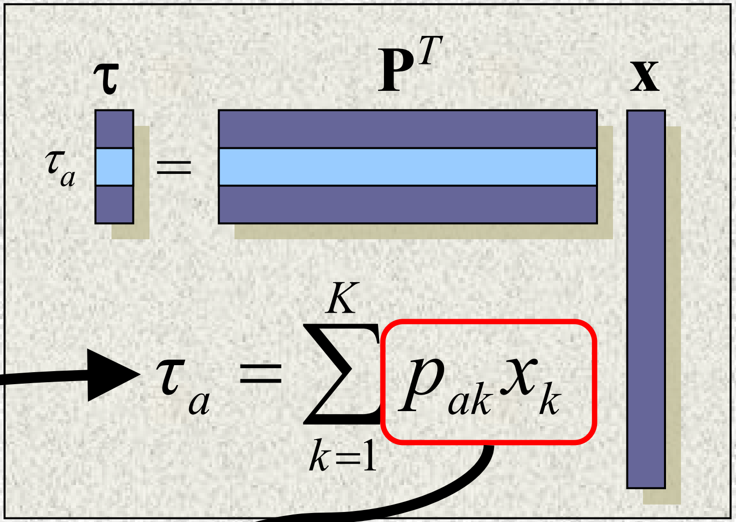
Contributions
Kourti, T. ; MacGregor, J.F. (1996)

Fault reconstrution
Qin. J (2003)

Fault Signatures
Yoon S, MacGregor J.F (2001)

Contribution to the Scores

The score vector is obtained by projection of an observation onto the model space : $\boldsymbol{\tau} = \mathbf{P}^T \mathbf{x}$



Each individual score can be expressed as:

$$\tau_a = \mathbf{p}_a^T \mathbf{x}$$

***k*-th variable contribution on the *a*-th score:**

$$c_k^{\tau_a} = p_{ak} x_k$$

Assumptions: Data are multivariate normal distributed (centered and scaled)

$$c_k^{\tau_a} \sim N\left(0; p_{ak}^2\right)$$

Contribution to the SPE

$$SPE = \sum_{j=1}^{K} e_{ij}^2 = \sum_{j=1}^{K} (x_{ij} - \hat{x}_{ij})^2$$

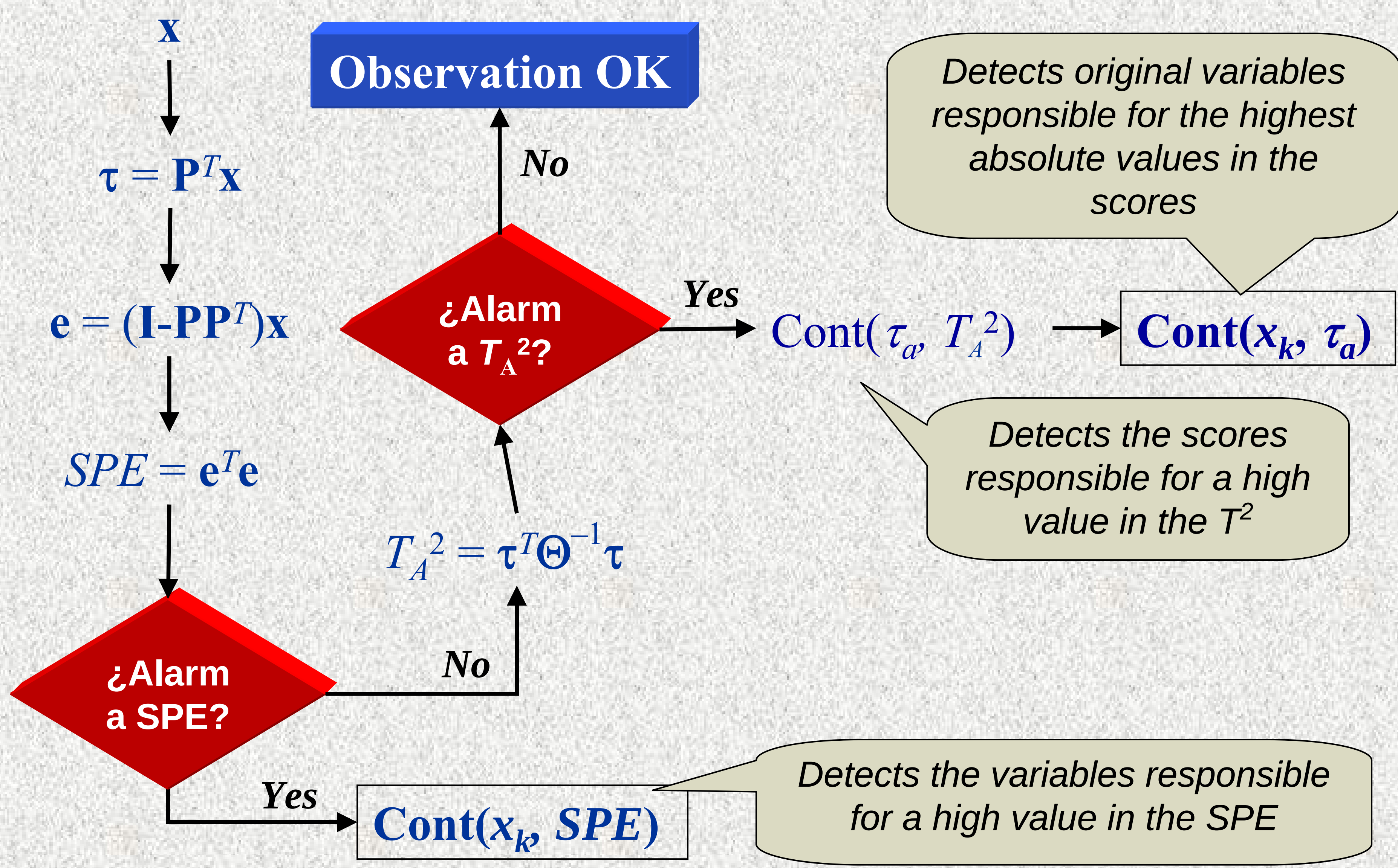
$$c_k = (x_k - \hat{x}_k)^2 = e_k^2$$

$$c_k = (x_k - \hat{x}_k) = e_k$$

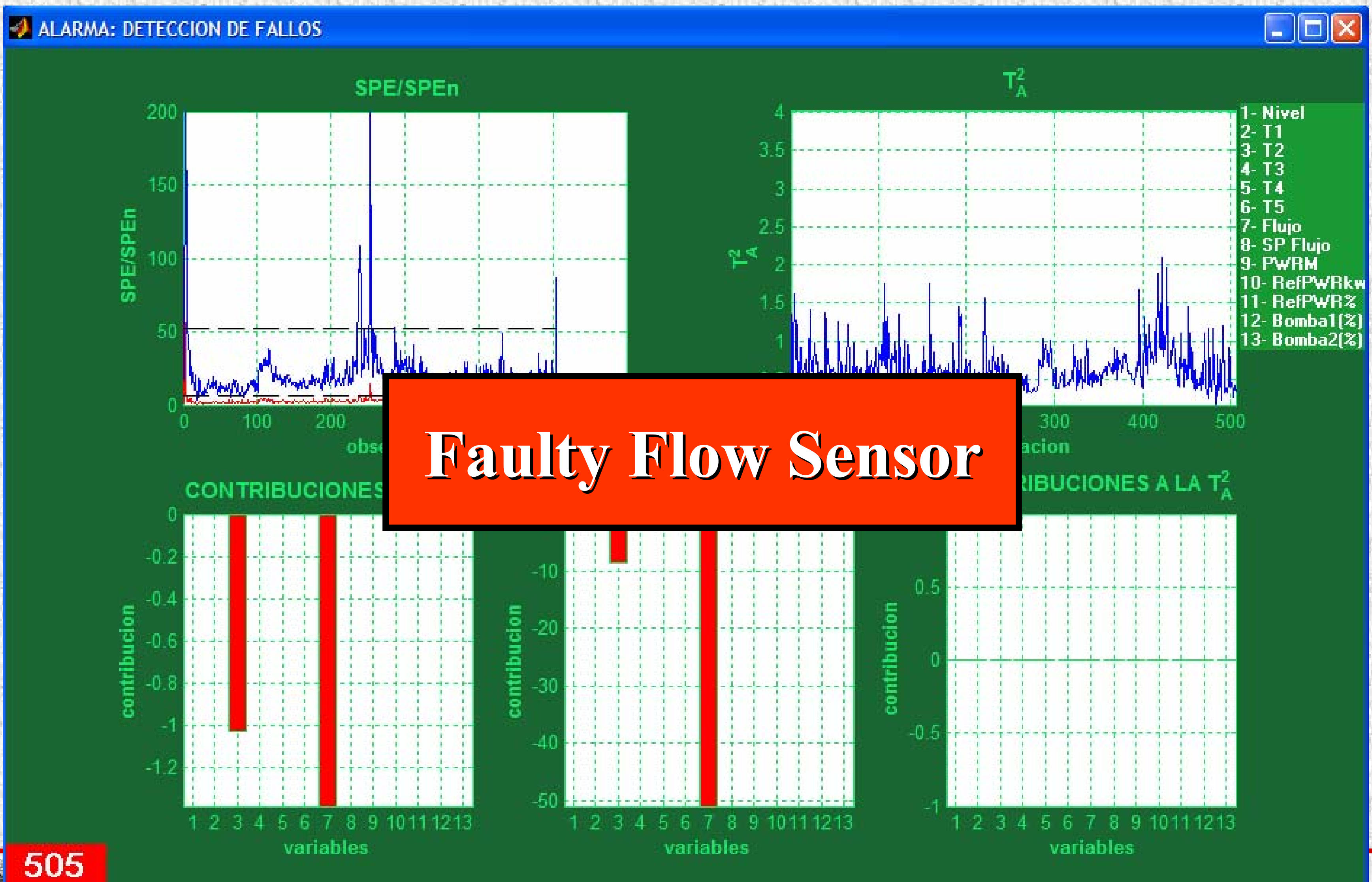
Standardized SPE : Some Authors propose this statistic approach to improve the sensitivity of the diagnosis (J. Westerhuis)

$$c_k = \frac{(x_k - \hat{x}_k)^2}{s_{k,res}^2} = \frac{e_k^2}{s_{k,res}^2}$$

Monitoring with PCA



Contribution plots method in a simple fault

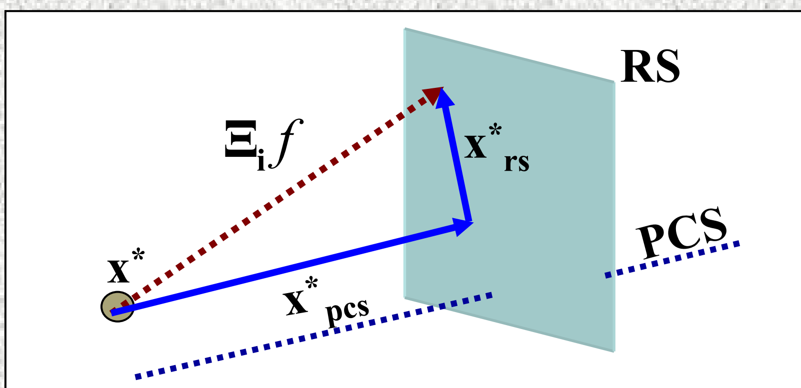


Fault Reconstruction

Uses the estimation and reconstruction of the observations registered under fault conditions to identify the variables responsible of the fault

a **reconstructed observation vector $\mathbf{x}_i$** is obtained correcting the observation $\mathbf{x}$:

$$\mathbf{x}_i = \mathbf{x} - \Xi_i f_i$$



x^*	**Observation free of fault**
x^*_{pcs}	**Projection of the fault in PCS**
x^*_{rs}	**Projection of the fault in RS**

Ξ_i is the "direction of faults matrix"

f is the estimation of the fault magnitude

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$$

Fault sensors 1 y 3

Orthonormal

Fault Reconstruction

For each kind of fault, we will search for the value f_i which minimises the SPE following the reconstruction.

Identification of the fault: the one which leads to a minimun SPE

Detectability, Identifiability and Reconstructability conditions

If there are several faults which reconstructed SPE near to the minimun

Fault can not be univocally identified

Contributions Method

Does not delete the efect of the fault

Requires no previous knowledge about the nature of the faults that affect the process

Signal: suspected responsible variables for the fault

Fault Reconstruction

Deletes the effect of the fault

Leads to correcting the measures of the faulty sensors and in that way anticipate real process faults in regulated processes

Requires previous knowledge about the nature of the faults which affect the process (fault directions matrix)

Signal: suspected responsible variables and suspected fault

Objectives

- Investigate existence of relationships among different methodologies

- Study the performance of these methodologies when applied to real processes and simulations.

- Study the effect of these strategies on phase I
  (Enbis Conference 2005 Vidal-Puig S.; Janssen P.M.A; Sanchis,J; Ferrer A.)

- Study the effect of autocorrelation in the monitored statistics
- Improvement of these methodologies

Acknowledgements

- *Multivariate Statistical Engineering Group; Dept. Applied Statistics, O.R. & Quality; Technical University of Valencia (SPAIN)*

  Alberto Ferrer Riquelme
- *Dept. of Systems Engineering and Control: DISA Technical University of Valencia (SPAIN)*

  Javier Sanchís Saez

Research Project: *Development of new strategies for the integration of the Automatic Process Control (APC) and the Statistical Process Control (SPC) in the regulation and monitoring of continuous multivariate industrial processes*

Santiago Vidal Puig