Statistical Techniques for Monitoring Industrial Processes

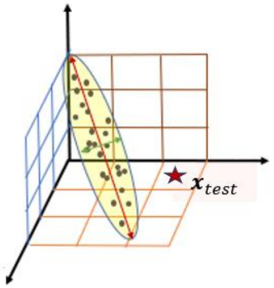
Lecture : PCA – Fault Detection

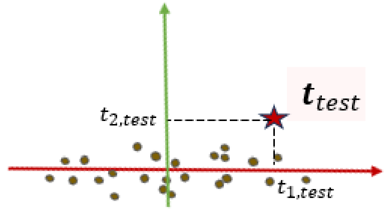
Module : PCA-based MSPM

Course TOC

- Introduction to Statistical Process Monitoring (SPM)
- Python Installation and basics (optional)
- Univariate SPM & Control Charts
  - Shewhart Charts
  - CUSUM Charts
  - EWMA Charts
- Multivariate SPM
  - Principal Component Analysis (PCA)-based MSPM
    - Dimensionality reduction
    - Fault detection & diagnosis (FDD) using PCA
    - Application to a Polymer Manufacturing process
  - Partial Least Squares (PLS) regression-based MSPM
  - Strategies for handling nonlinear, dynamic, multimode systems
- Deploying SPM solutions

PCA: Systematic Component and Noise





$\boldsymbol{x}_{test}$ → PCA Model → $\boldsymbol{t}_{test}$

Systematic component

$$T^2 = \frac{t_{1,test}^2}{\lambda_1} + \frac{t_{2,test}^2}{\lambda_2}$$

called Hotelling's T^2

quantifies where the test sample lies in the PC space

Noisy component

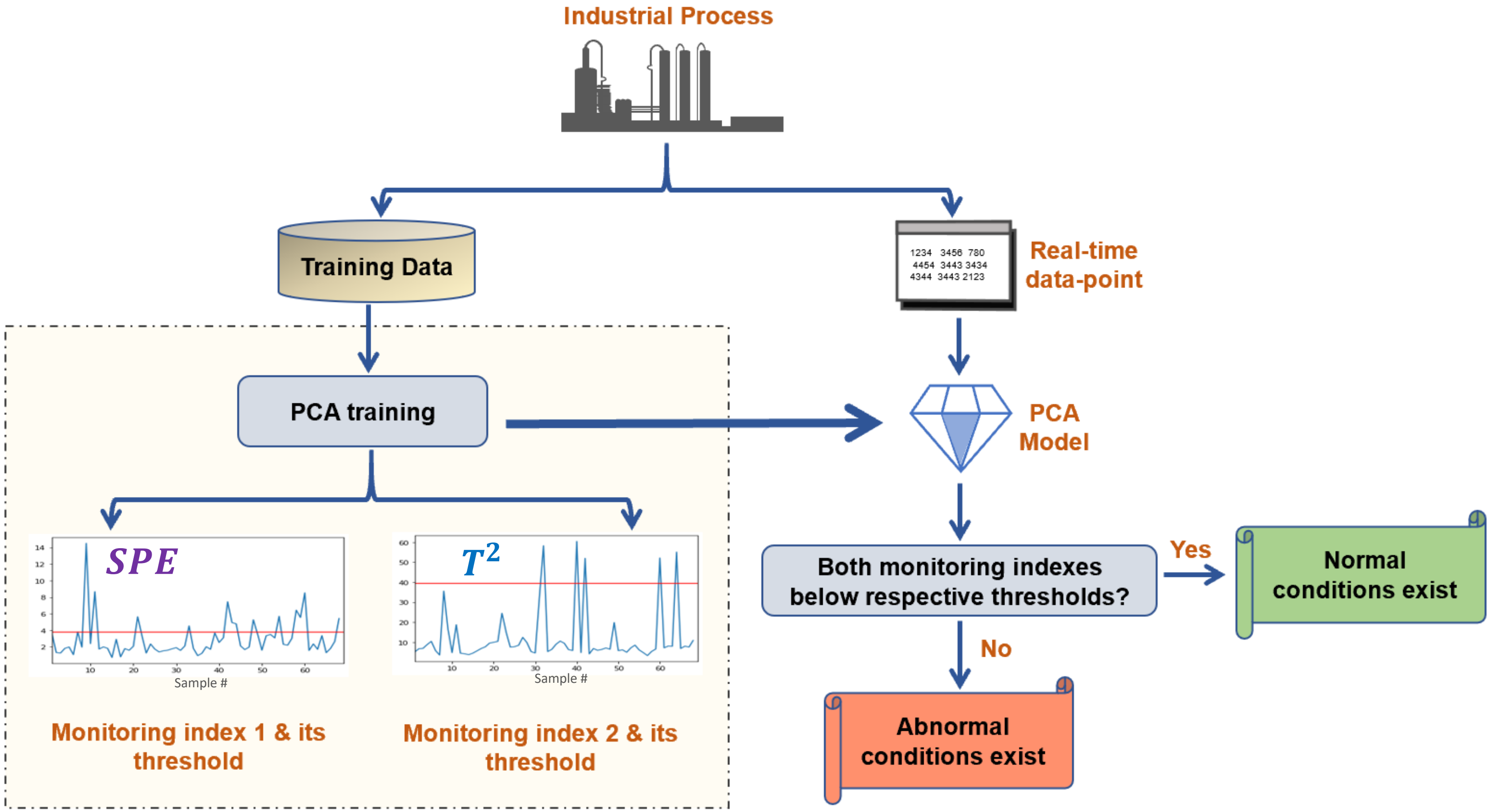
$$\boldsymbol{e}_{test} = \begin{bmatrix} e_{1,test} \\ e_{2,test} \\ e_{3,test} \end{bmatrix}$$

$$Q_{test} = e_{1,test}^2 + e_{2,test}^2 + e_{3,test}^2$$

also called SPE (square prediction error)

quantifies the noisy component of $\boldsymbol{x}_{test}$ that could not be explained by the PCs

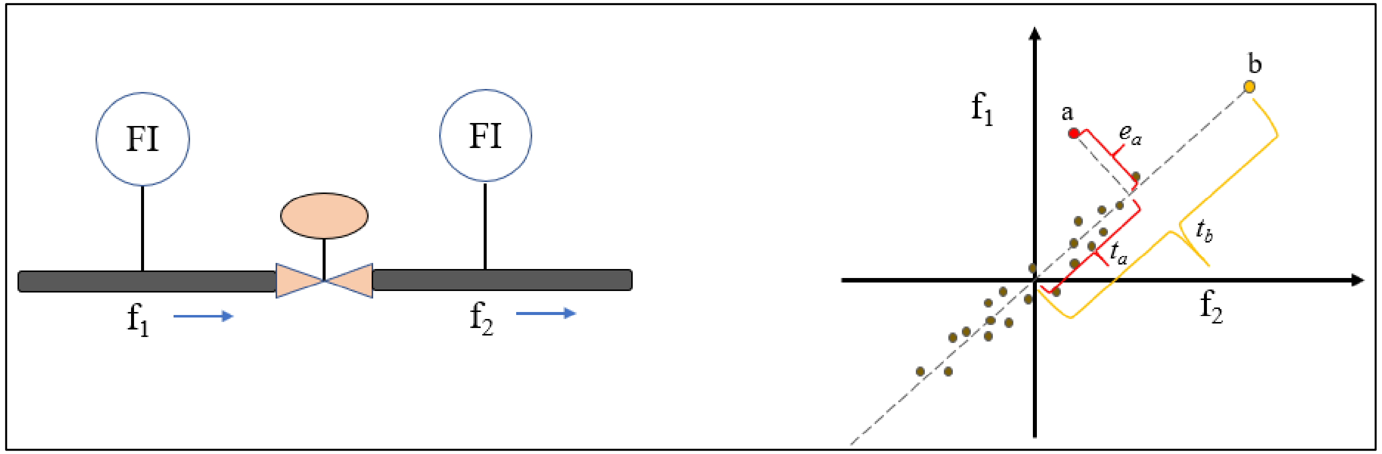
PCA-based Fault Detection Workflow



Importance of both T^2 and Q Statistics

- T^2 and Q are complementary metrics
- It is possible that an abnormal sample violates control limit of one metric, but not the other

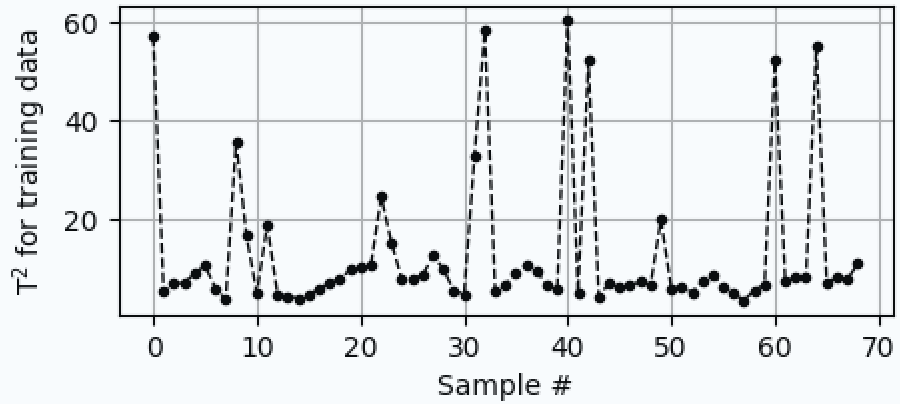
Both T^2 and Q are important and should be monitored



Flow measurements across a valve (left) and mean-centered flow readings with two abnormal instances (samples a and b)

Computing (Statistical) Control Limits for T^2 and Q Statistics

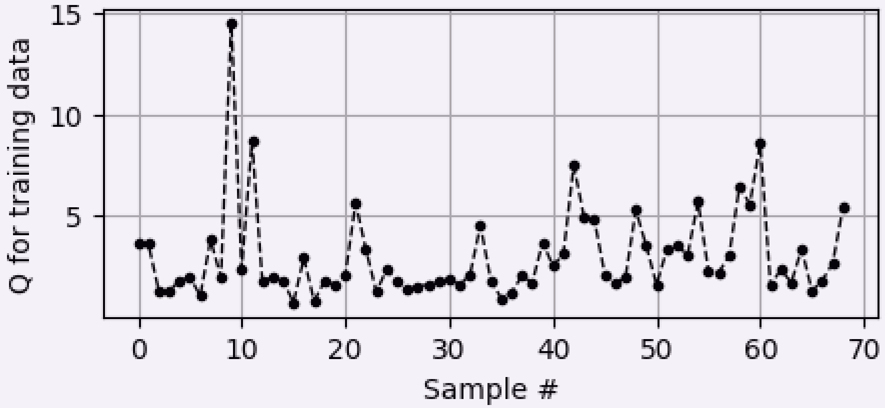
T^2



$$T^2_{CL} = \frac{k(N^2-1)}{N(N-k)} F_{k,N-k}(\alpha)$$

- α is false alert rate (usually 0.01 or 0.05)
- k = number of PCs retained
- N = number of training samples
- $F_{k,N-k}(\alpha)$ is (1- α) percentile of a F distribution with k and N – k degrees of freedom

Q



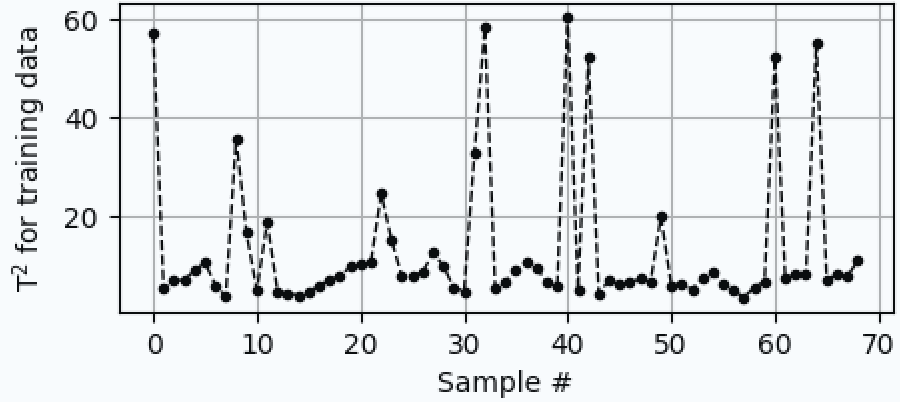
$$Q_{CL} = \theta_1 \left(\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (1-h_0)}{\theta_1^2} \right)^2$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}; \quad \theta_r = \sum_{j=k+1}^{m} \lambda_j^r \text{ ; } r\text{=1,2,3}$$

- z_α is (1- α) percentile of a standard Gaussian distribution

Computing (Statistical) Control Limits for T^2 and Q Statistics

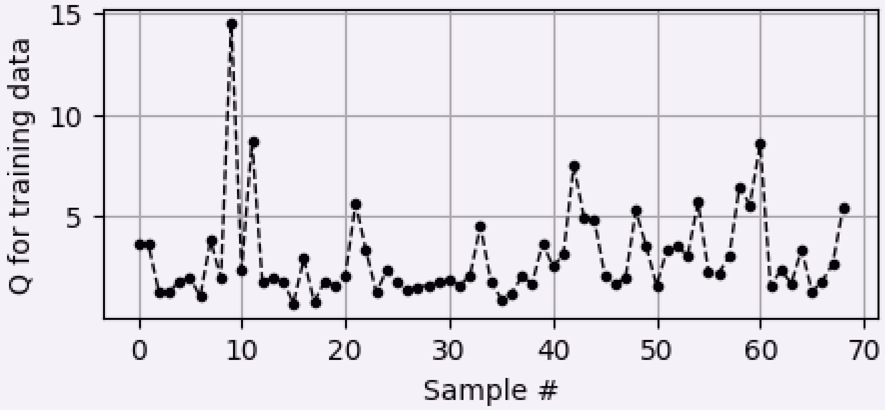
T^2



$$T^2_{CL} = \frac{k(N^2-1}{N(N-k}$$

- α is false aler
- k = number o
- N = number of training samples
- $F_{k,N-k}(\alpha)$ is (1- α) percentile of a F distribution with k and N – k degrees of freedom

Q



$$\frac{\theta_2 h_0(1-h_0)}{\theta_1^2}\Bigg)^2$$

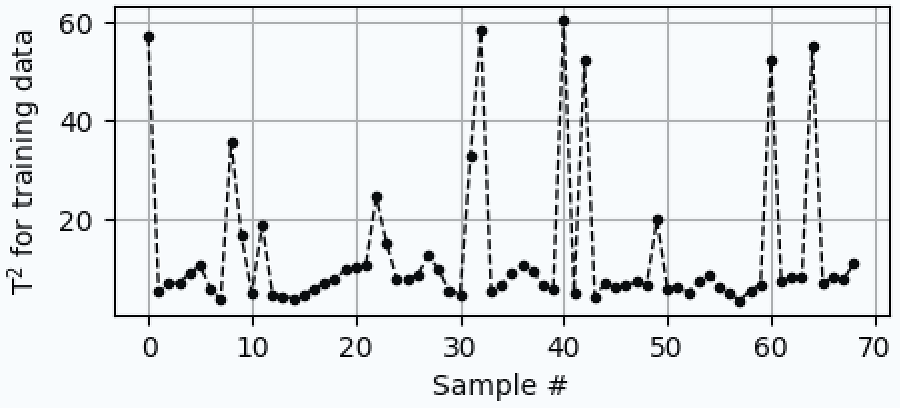
$k+1\ \lambda_j^r$; r=1,2,3

- z_α is (1- α) percentile of a standard Gaussian distribution

These statistical expressions for control limits are valid only if the training data follow a multivariate Gaussian distribution

Computing (Empirical) Control Limits for T^2 and Q Statistics

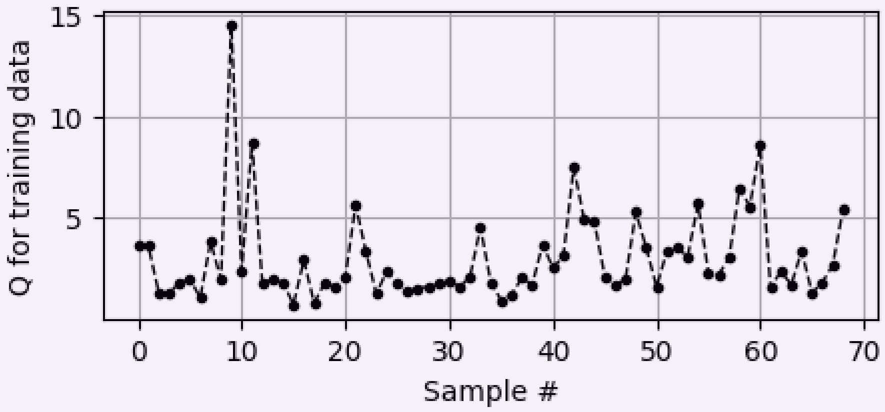
T^2



T^2_{CL} = np.percentile(T2_values, 99) # alpha=0.01

- Control limit is simply the 99th percentile of the T^2 values for the training samples

Q



Q_{CL} = np.percentile(Q_values, 99) # alpha=0.01

- Control limit is simply the 99th percentile of the Q values for the training samples

Statistical Techniques for Monitoring Industrial Processes

Next Lecture : PCA – Fault Detection Implementation

Module : PCA-based MSPM