Università degli Studi di Padova

Department of Industrial Engineering

Machine Learning Laboratory #2

Prof. Pierantonio Facco

CAPE-Lab, Computer-Aided Process Engineering Laboratory

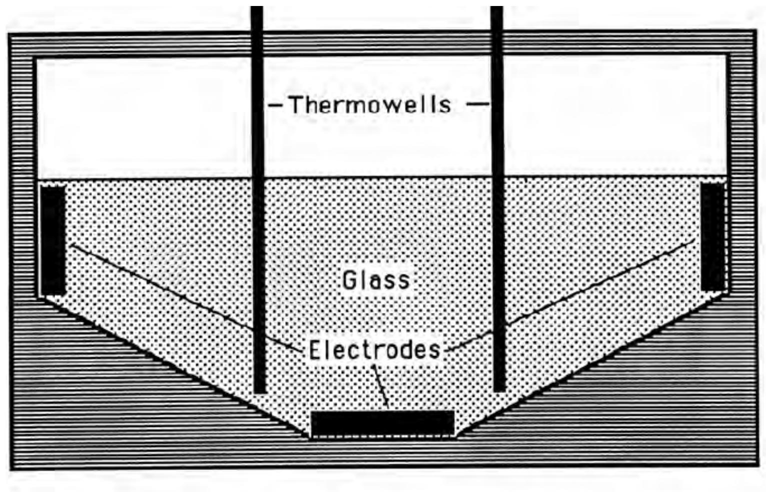
Email: pierantonio.facco@unipd.it

URL: https://research.dii.unipd.it/capelab/

Example #2: Slurry-fed ceramic melter monitoring

Slurry-fed glass melter

- High temperature glass fusion in a **slurry-fed glass melter***
  - nuclear waste stabilization in the glass
  - the product is stable for long time



- <u>Objective</u>: development of a **monitoring system** of the process

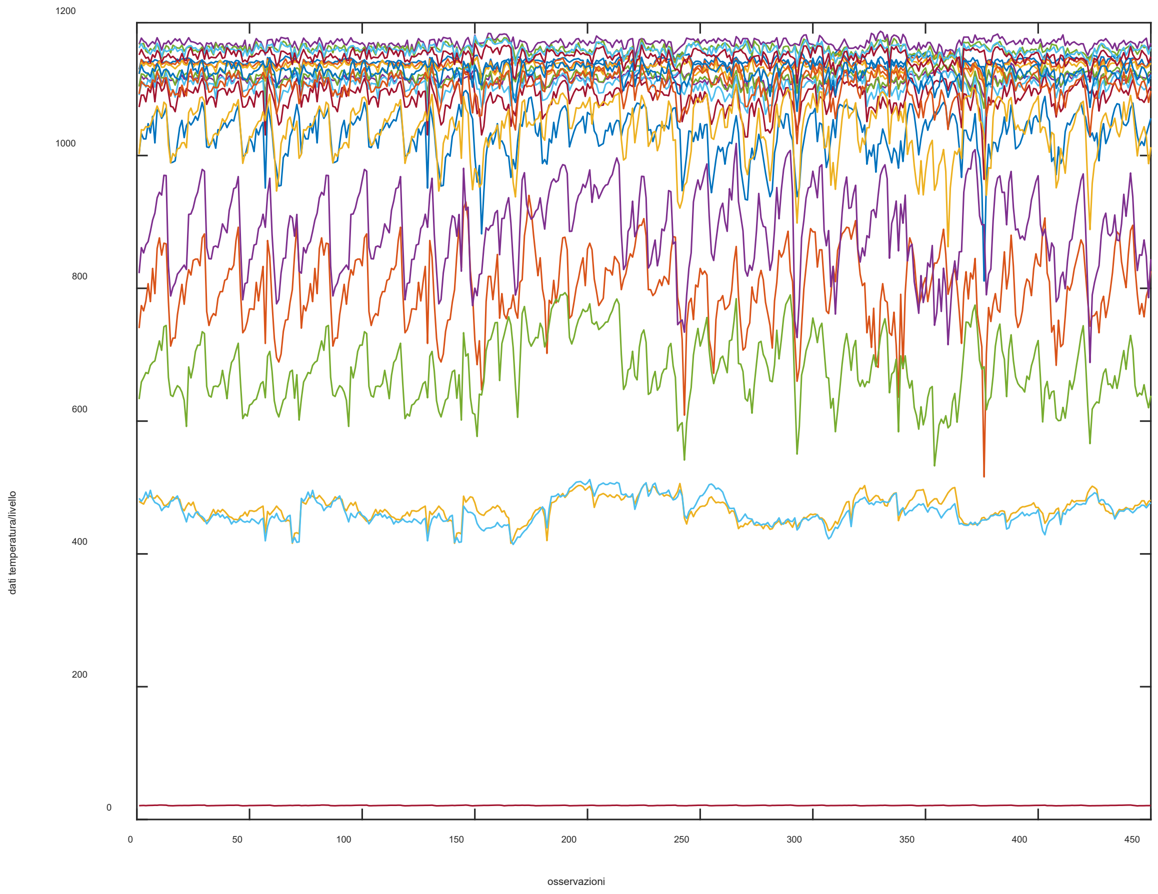
* Wise, B.M., D.J. Veltkamp, N.L. Ricker, B.R. Kowalski, S.M. Barnes and V. Arakali (1991). Application of Multivariate Statistical Process Control (MSPC) to the West Valley Slurry-Fed Ceramic Melter Process, *Waste Management '91 Proceedings*, Tucson (USA, AZ).

Data

- Data $\mathbf{X}$:
  - 450 observations
  - 21 variables:
    - temperatures in 20 points of the melter
    - glass level

- Load:
  - `load replacedata`

- Data visualization:
  - no anomalies seem to be present



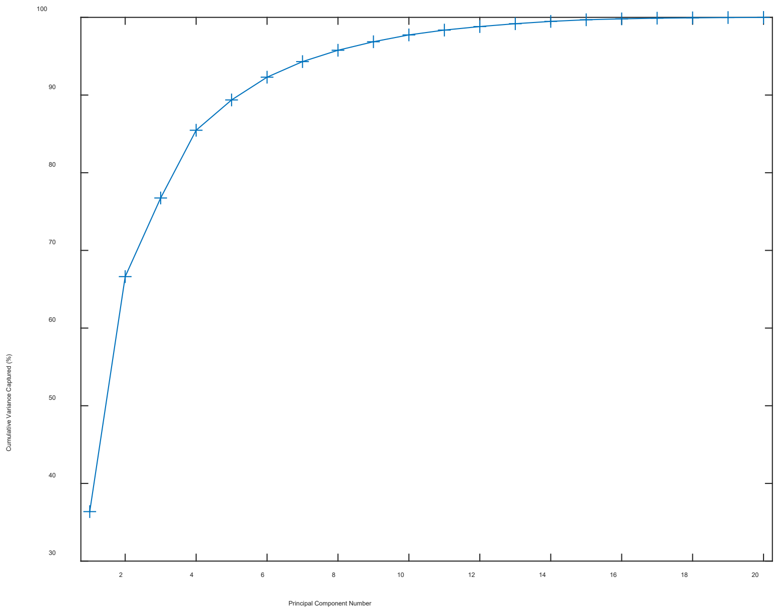
Exploratory and monitoring model

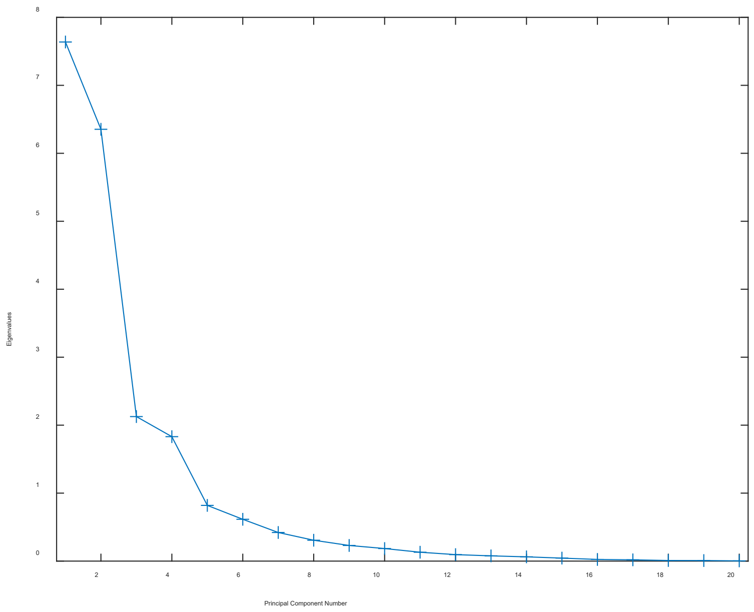
- **PCA model calibration** on «standard» process data (normal operating conditions NOC)
  - choice of the number A of PCs
  - loadings analysis to study variables correlations
  - scores analysis to study observation relations

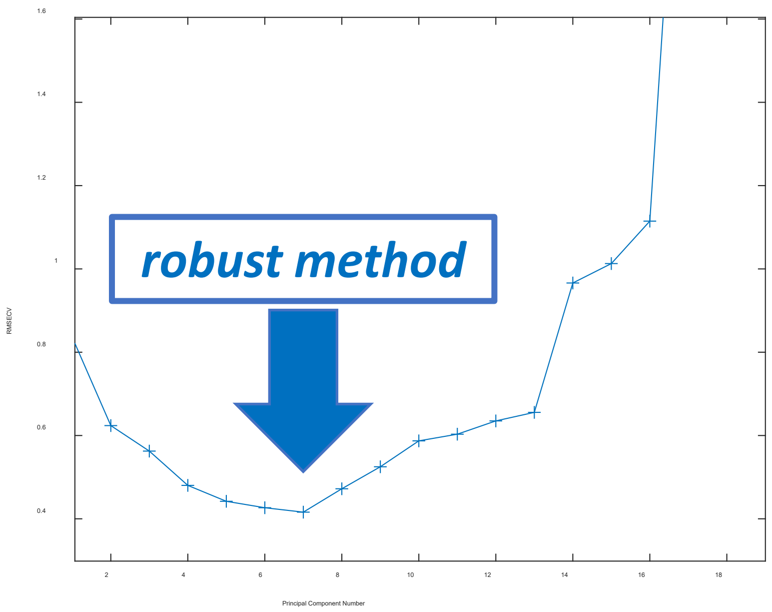
- **Multivariate monitoring of unknown observations**:
  - statistical monitoring chart
    - T^2 Hotelling chart
    - Q residuals chart
  - **anomalies detection**:
    - confidence limits
  - **anomalies diagnosis**
    - contribution plots

Choice of the PCA model dimension

- Autoscaling and PCA model calibration
- Choice of the PCs number:
  - cumulative variance: PCs $> 8/9$ do not add information
  - eigenvalues < 1 at PC5
  - cross validation: $\min(RMSECV)$ for **PC7**

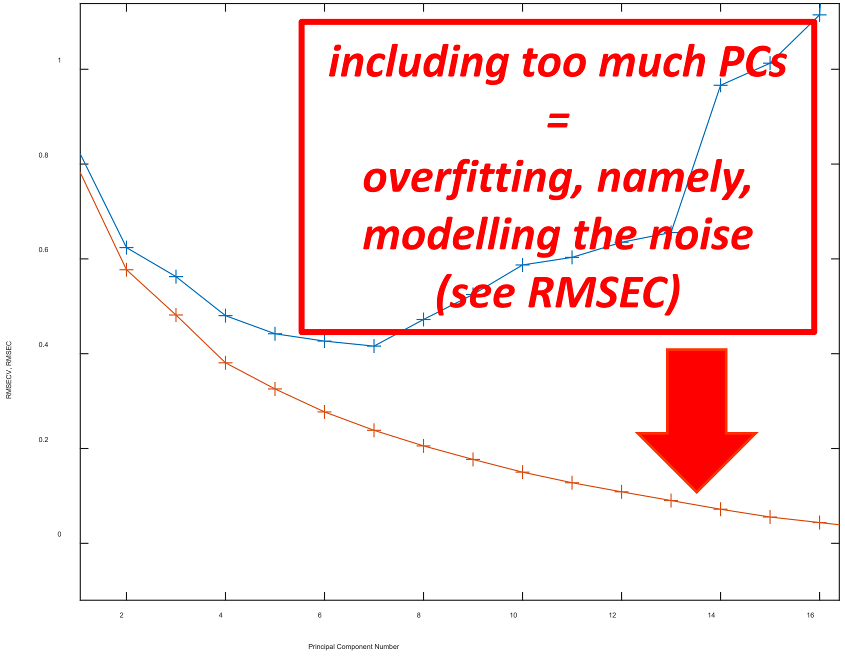






Cross validation

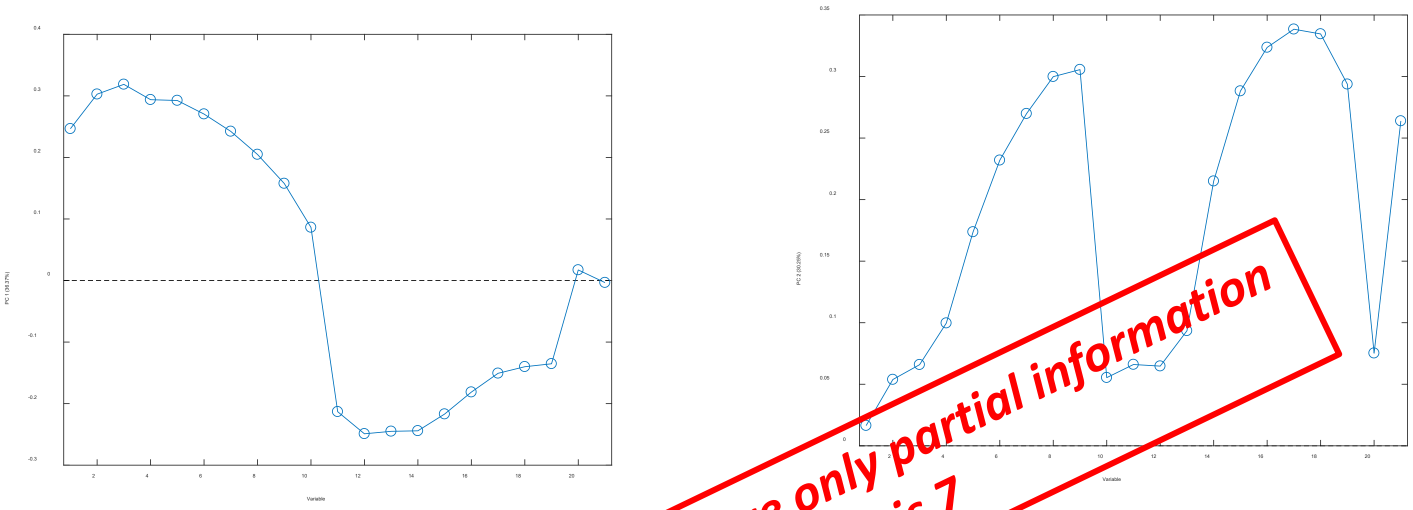
- Jackknife procedure for testing how the model works on new unknown data:
  - allows understanding the optimal model complexity
  - method:
    - science-based
    - robust
    - credible



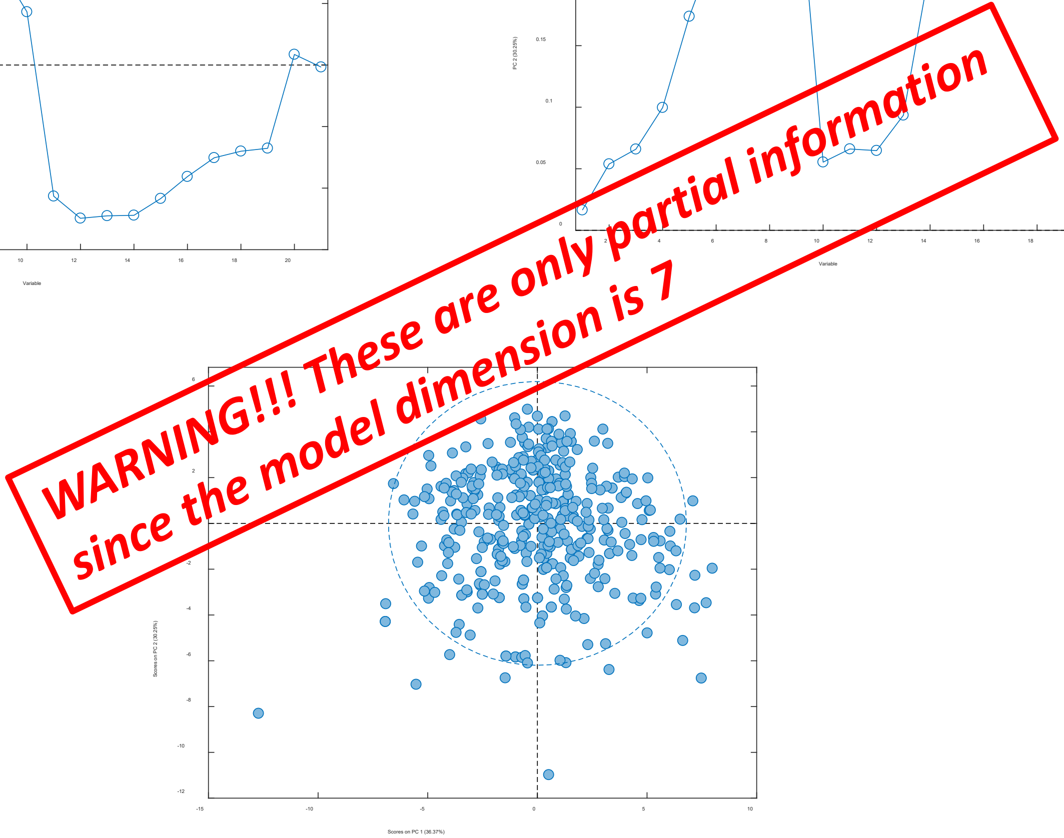
	Venetian Blinds	Contiguous Blocks	Random Subsets	Leave-One Out
Test sample selection scheme				

Exploratory analysis

- Variables correlation:
  - loading plot

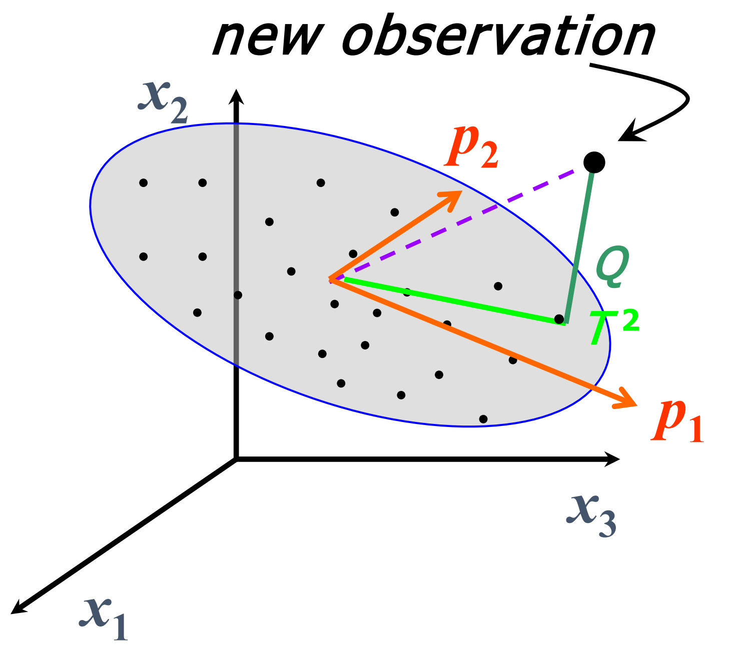


- Relation among observations:
  - score plot



Sample diagnostics

- Q **residuals** (squared prediction error SPE)
  - orthogonal distance between a new observation and the model
  - high Q = a new event happened (with a new correlation structure)
  - Q measures the variability that is not captured by the model



- **Hotelling** T^2
  - weighted distance from the average condition on calibration
    - weighted on the variance explained by each PC
  - high T^2 = too high variability with respect to the average calibration conditions
  - T^2 in a measure of the variability within the model

$$T^2_{N+1} = \hat{\mathbf{t}}^{\mathrm{T}}_{N+1}\boldsymbol{\Lambda}^{-1}\hat{\mathbf{t}}_{N+1} = \sum_{a=1}^{A}\frac{\hat{t}^2_{a,N+1}}{\lambda_a}$$

$$Q_{N+1} = \mathbf{e}^{\mathrm{T}}_{N+1}\mathbf{e}_{N+1}$$

Confidence limits

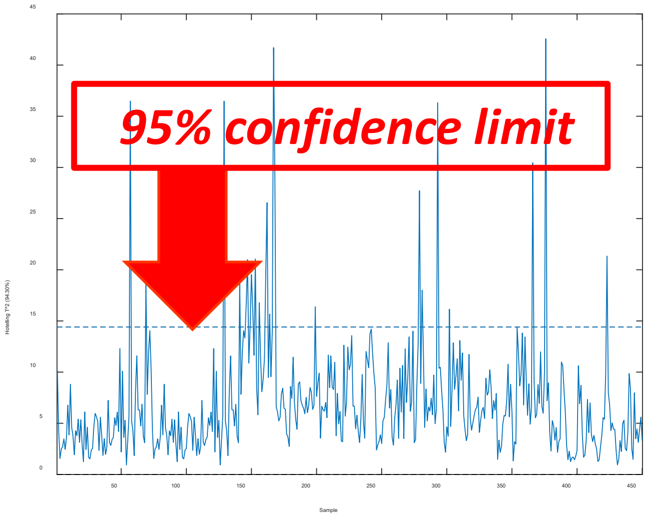
- Hotelling T^2
  - the limit comes from a F distribution of:

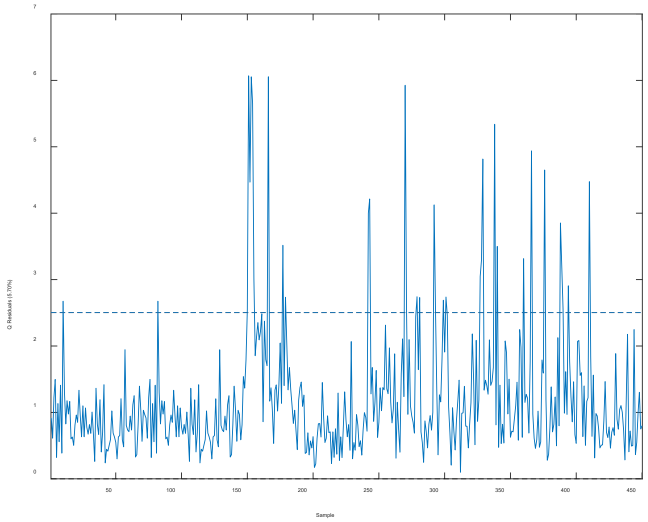
$$T^2_{\mathrm{lim}} = \frac{A(N-1)}{N-A} F_{A,N-A,1-\alpha}$$

- Q residuals
  - the limit comes from a χ^2 distribution of

$$Q_{\mathrm{lim}} = \frac{s_Q^2}{2\mu_Q} \chi^2_{\frac{2\mu_Q}{s_Q^2},1-\alpha}$$

- **5% of out of limits must be accepted**
  - **the outliers should be removed only if an anomaly is present**

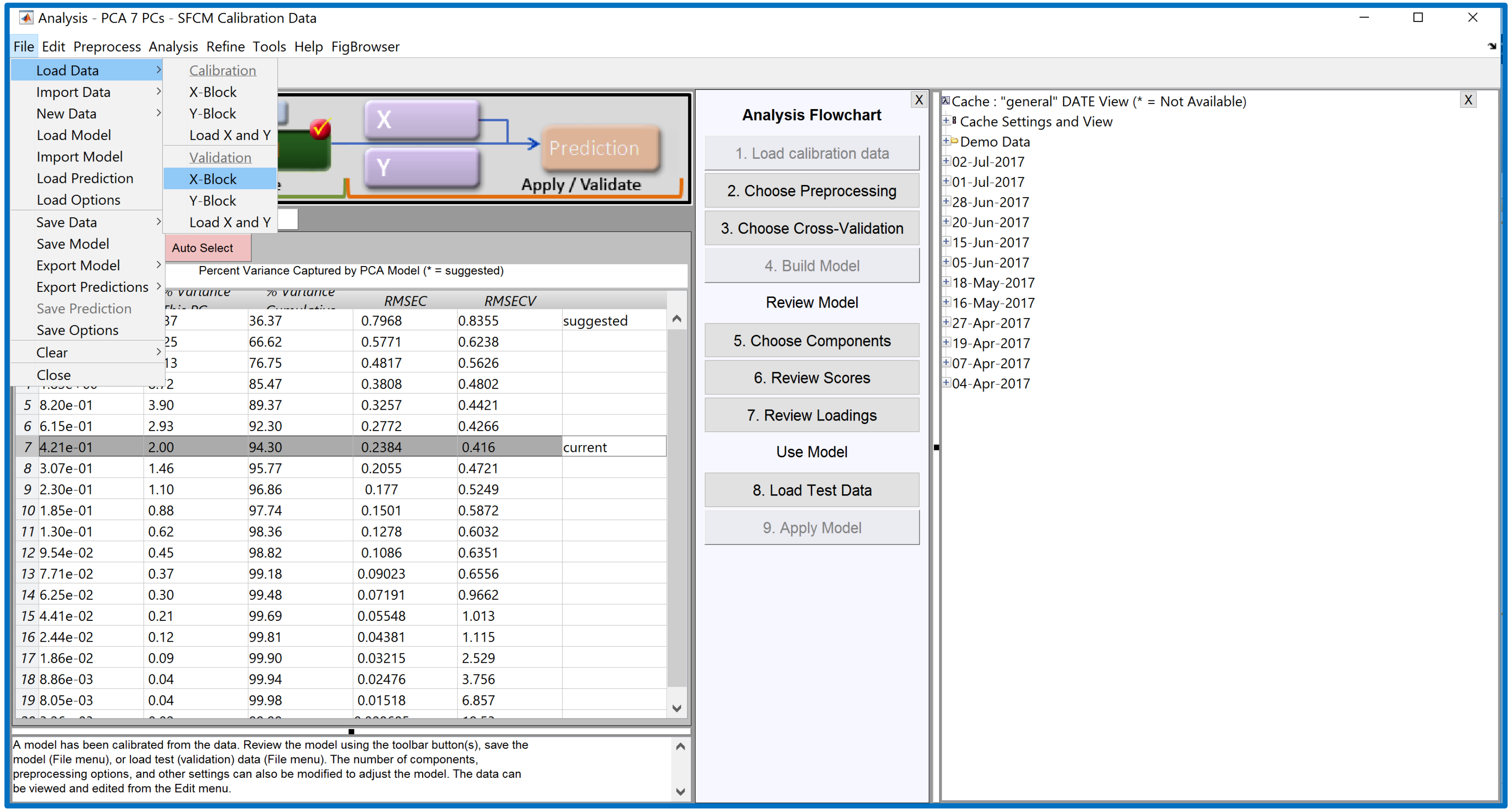




Monitoring system validation #1

New observation monitoring: validation #1

- Load validation dataset: **reptest1**



Validation observation #1

- New data $\mathbf{x}_{\mathrm{NEW}}$ is projected onto the hyperplane of the PCs:

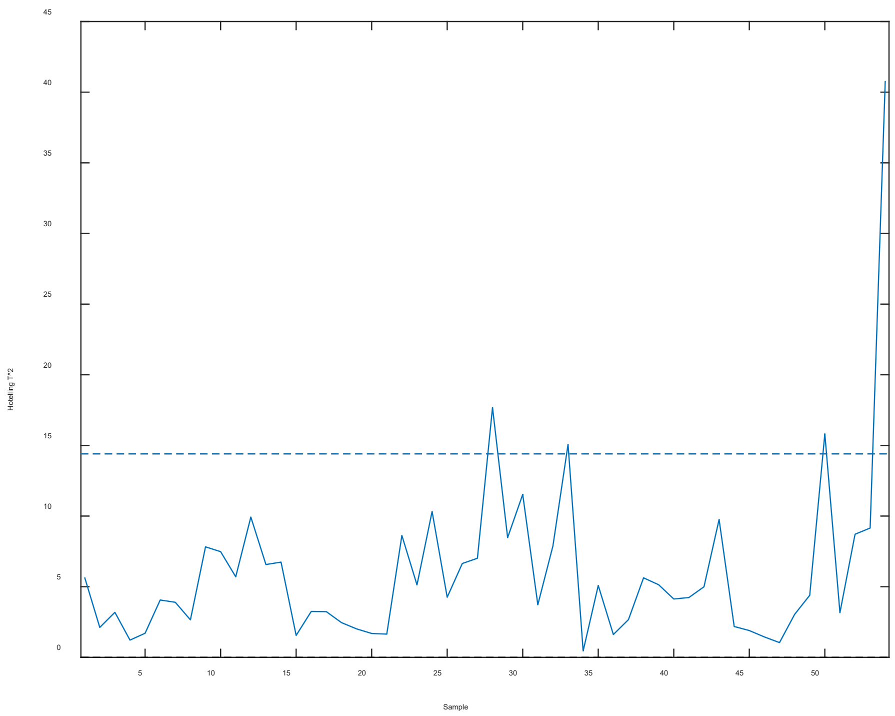
$$\mathbf{t}_{N+1} = \mathbf{x}_{N+1}\mathbf{P}$$

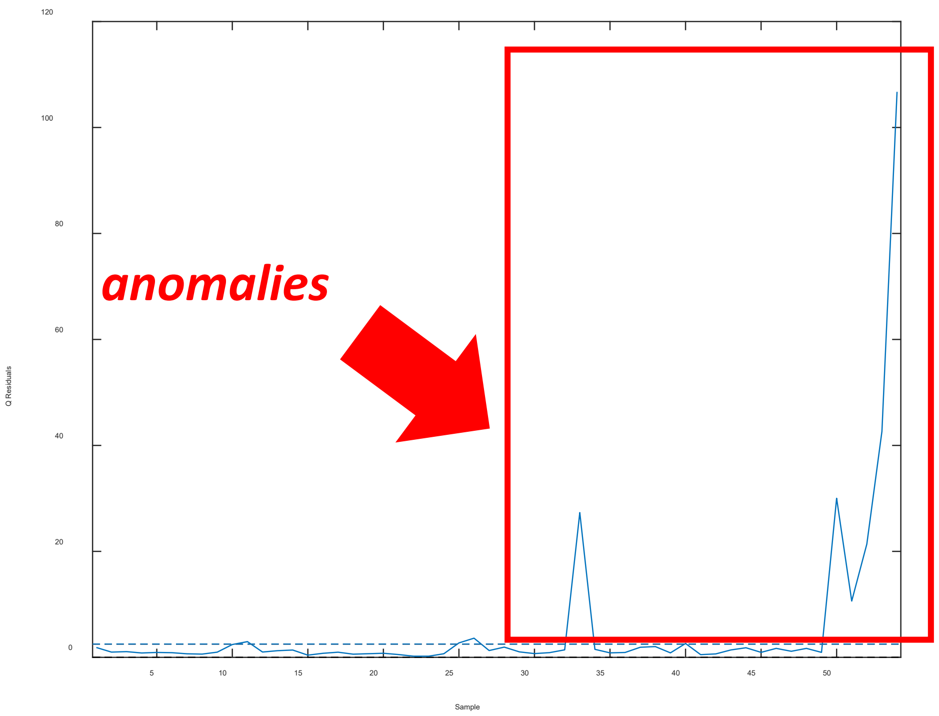
- The coordinates of the new observation in the PCs space are obtained
- Observe if the Hotelling and Q indices are well within the **confidence limits** obtained from the NOC calibration data

Process monitoring

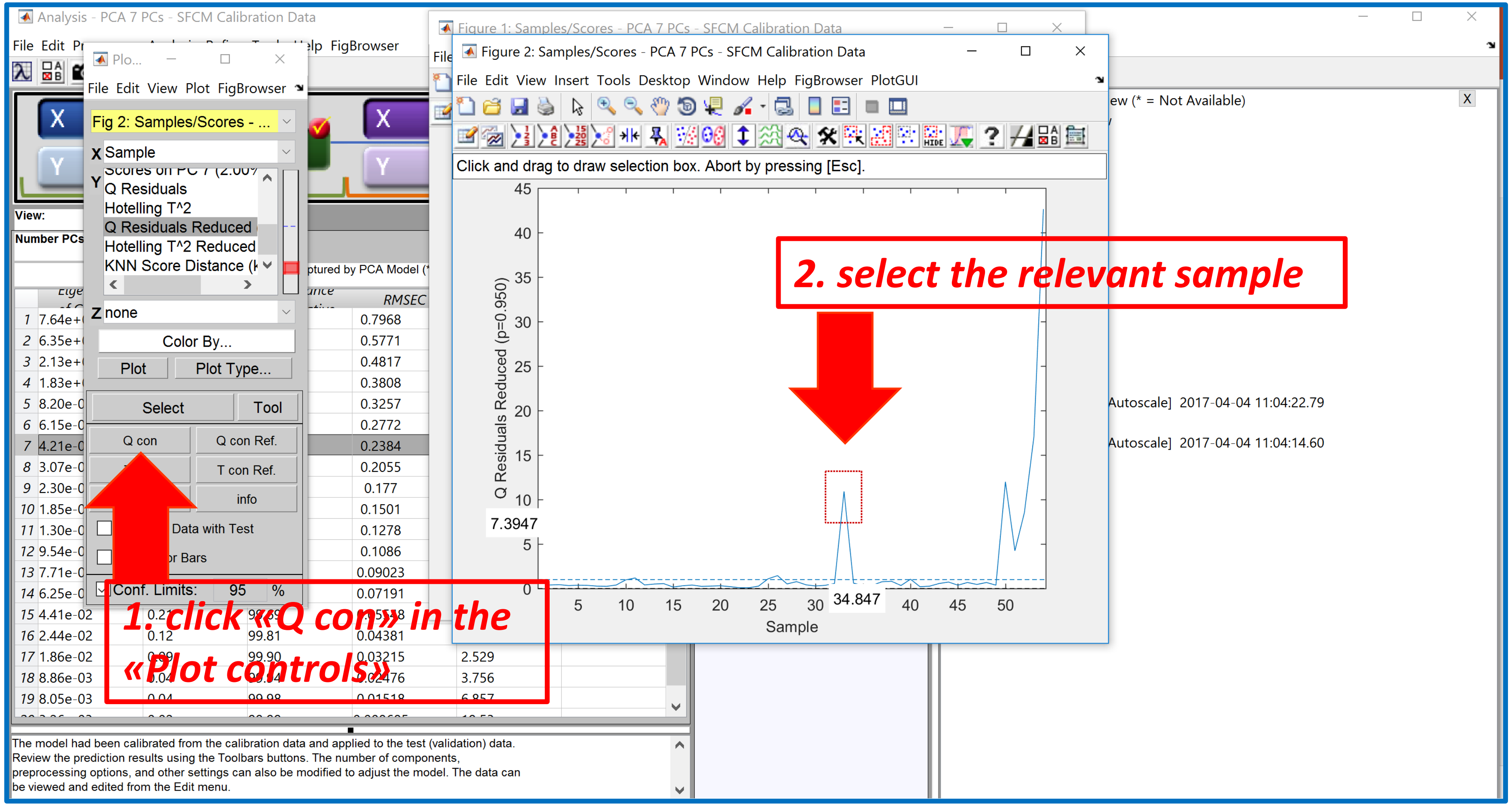
- Anomalies are detected for:
  - observation 33
  - observation 50-54

diagnosis is needed!



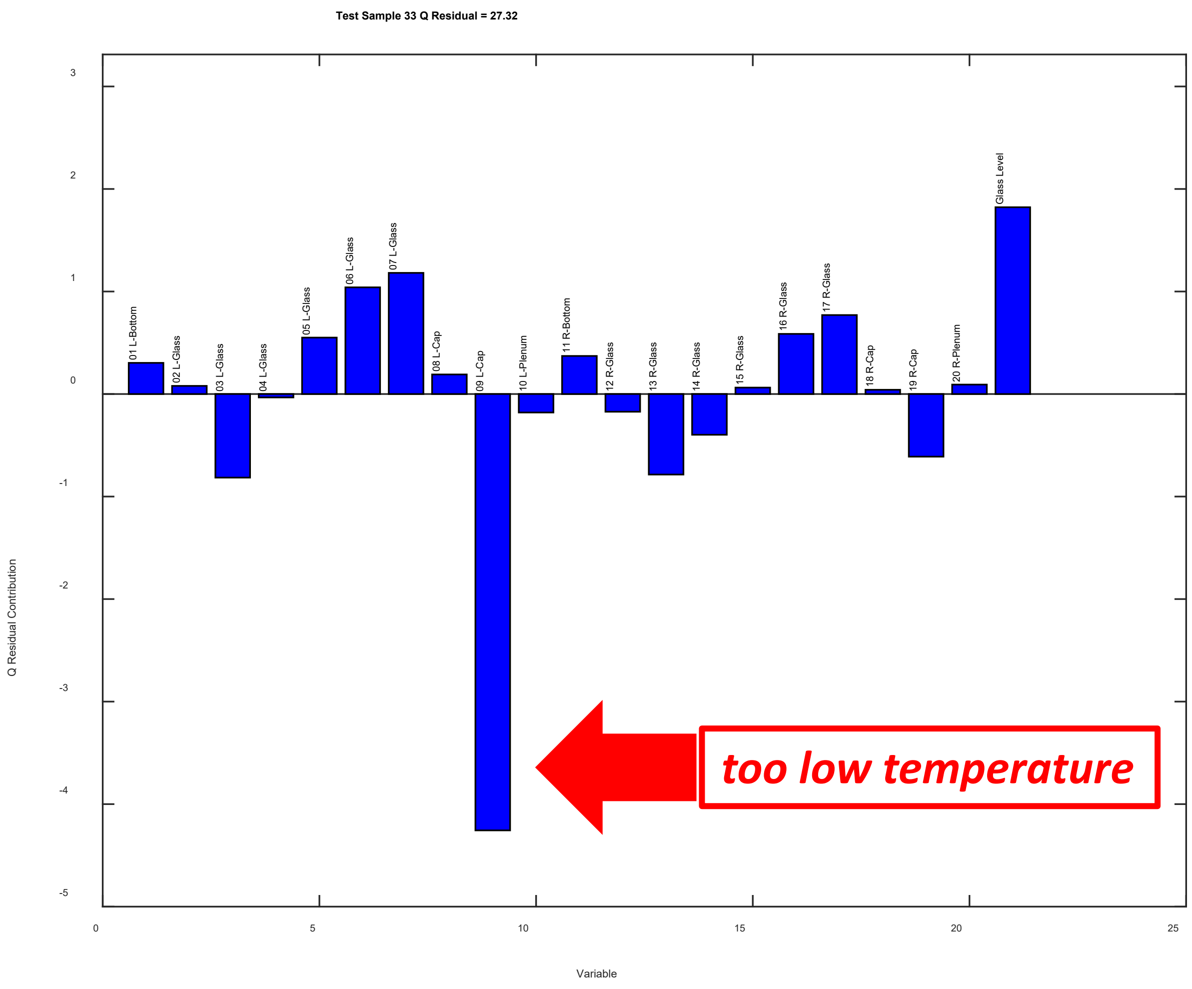


Diagnosis of anomalies: contribution plots for obs. 33



Diagnosis of the anomaly at obs. 33

- **Contribution plots** are used for diagnosis
  - **this diagnosis tool do not find the CAUSE of the anomaly, but the variables which are most correlated to the causes of the malfunction!!!**



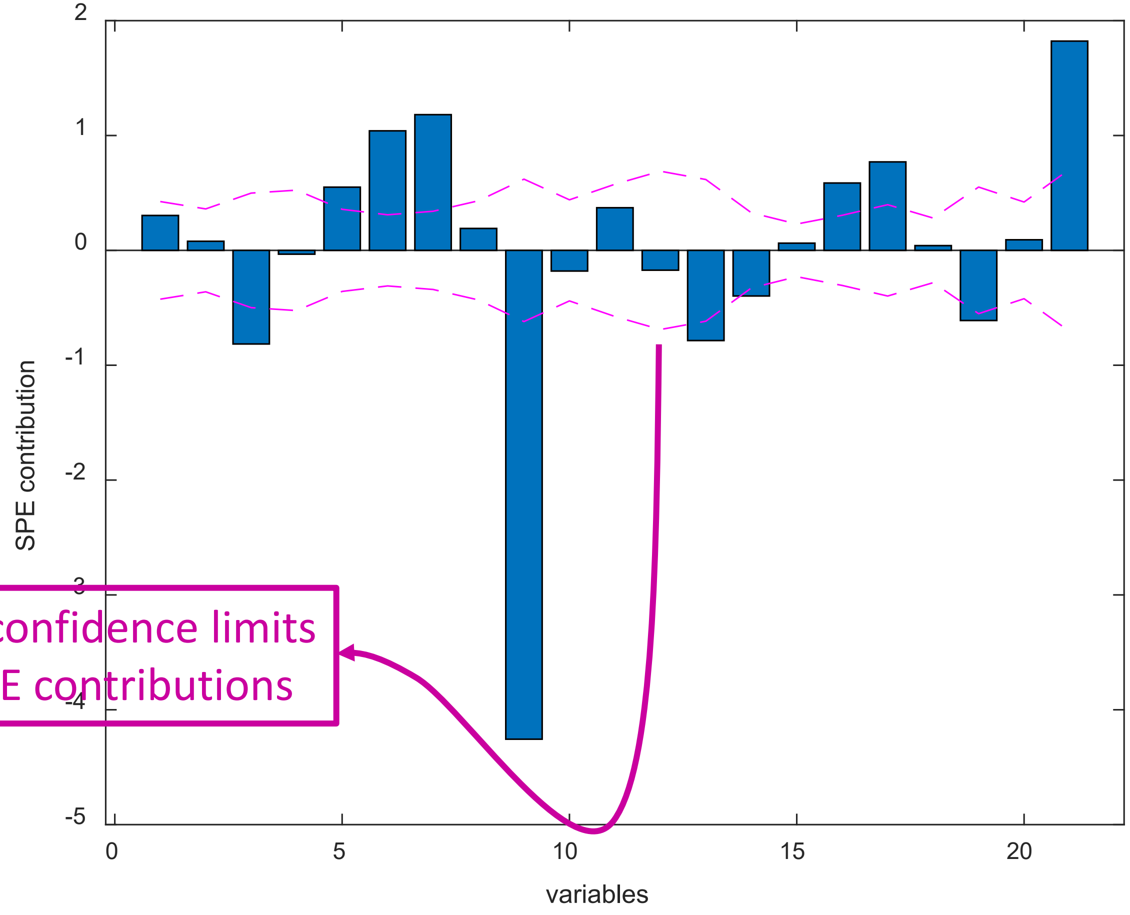
Improved diagnosis of the anomalies

- **Contribution plots** are used for diagnosis in an improved way if you build a contribution plot chart with the respective confidence limits:
  - the confidence limits on the contributions on T^2 or SPE are based on the assumption that scores and residuals are normally distributed
  - under this hypothesis it is easy to build a sort of Shewhart chart on each variable with the respective 95% confidence limits

1. calculate the mean and standard deviation of each contribution (either of T^2 or SPE) for the calibration data
2. verify if the mean are sufficiently close to zero
3. calculate the 95% (i.e., $\pm 1.96\sigma$) confidence limits
4. compare the contribution of each variable for new projected observations to the respective limit

Improved diagnosis of the anomaly at obs. 33

- The improved diagnosis tell us much more: the issue here is related to more than one variable!
  - variable 9 is highly deviating
  - variables 5, 6, 7, 16, 17 and 21 are too high
  - variables 3 and 13 are too low

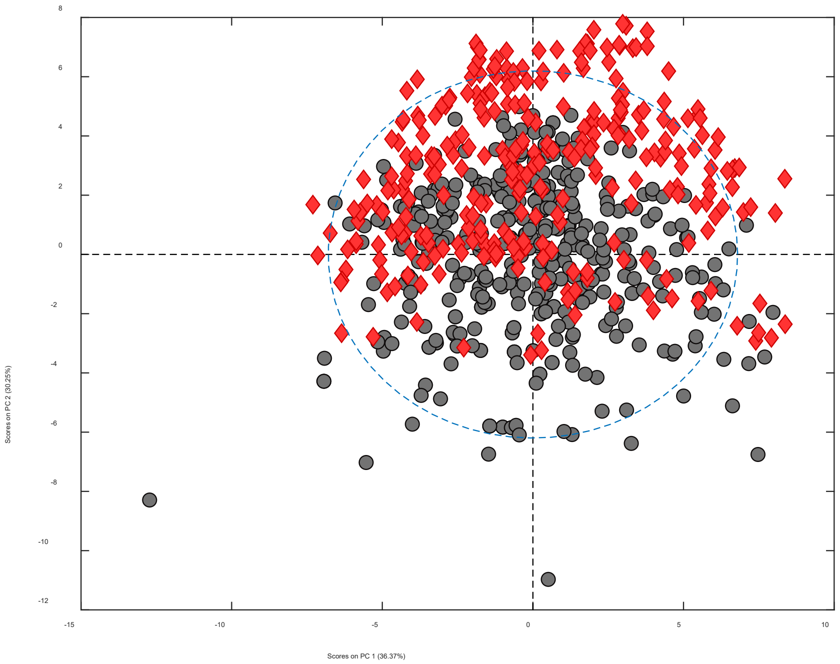


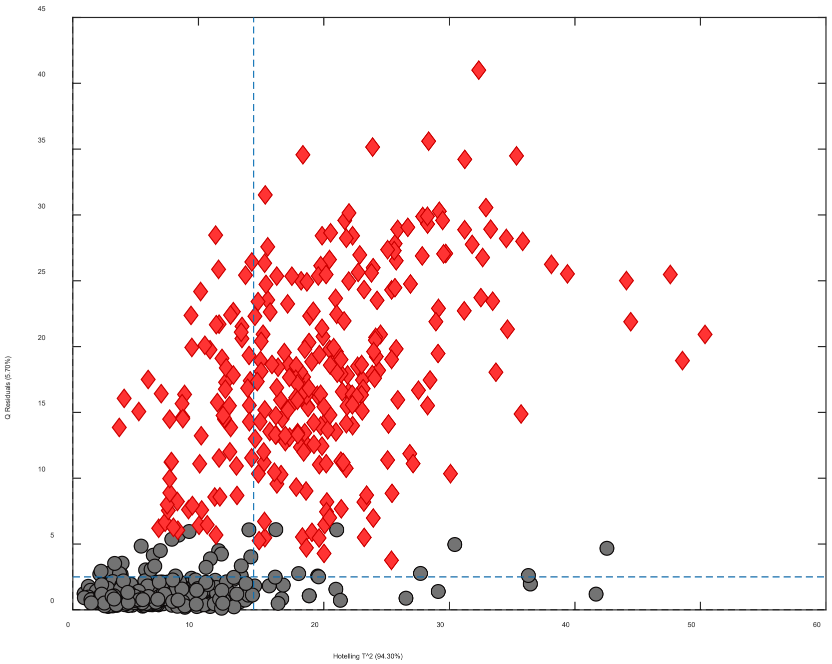
95% confidence limits on SPE contributions

Monitoring system validation #2

New observation monitoring: validation #2

- Load dataset: `reptest2`
- Score plot: calibration + validation:
  - validation set shows different mean with respect to calibration
  - new process conditions are found





Take home message

- Multivariate statistic process monitoring utilizes **complementary monitoring charts** that must be observed together
  - Hotelling T^2
  - squared prediction error Q
- Both of them must be overviewed at the same time!
- Control charts **detect malfunctions**
  - and subtle phenomena as well
    - drifts
    - seasonal effects
- The malfunctions causes can be **diagnosed** studying the variables which are most correlated to the anomaly

Matlab® code (1/4)

```
clear all
close all
clc

load replacedata

Xc=double(repcal);
Xv1=double(reptest1);
Xv2=double(reptest2);

% Data visualization
figure;plot(Xc(:,1:2),'DisplayName','X(:,1:2)');xlabel('sample');ylabel('temperature [°C]');
figure;plot(Xc(:,1:3),'DisplayName','X(:,1:3)');xlabel('sample');ylabel('temperature [°C]');
figure;plot(Xc(:,4:6),'DisplayName','X(:,4:6)');xlabel('sample');ylabel('temperature [°C]');
figure;scatter(Xc(:,1),Xc(:,2));xlabel('temperature 1 [°C]');ylabel('temperature 2 [°C]');box on;
figure;scatter(Xc(:,1),Xc(:,3));xlabel('temperature 1 [°C]');ylabel('temperature 3 [°C]');box on;
figure;scatter(Xc(:,1),Xc(:,11));xlabel('temperature 1 [°C]');ylabel('temperature 11 [°C]');box on;

Cc7_8=corrcoef(Xc(:,7:8));
Cm=corrcoef(Xc);
corrmap(Xc);
figure;imagesc(corrcoef(Xc));xlabel('variables');ylabel('variables’);
```

Matlab® code (2/4)

```
% PCA model building
opca.display='off';
opca.plots='none';
opca.preprocessing='autoscale';
pcam=pca(Xc,7,opca);

% Verification of the monitoring charts performances
T2=pcam.tsqs{1,1};
Q=pcam.ssqresiduals{1,1};
T2lim=pcam.tsqlim{1,1}(1,1);
Qlim=pcam.reslim{1,1}(1,1);
sum(T2>T2lim)/length(Xc)
sum(Q>Qlim)/length(Xc)

% Verification of the residuals
T=pcam.loads{1,1};
P=pcam.loads{2,1};
E=auto(Xc)-T*P';
histogram(E(:,1))
histogram(E)
normplot(E(:,1))
```

Matlab® code (3/4)

```
% Verification of the control limits
T2lim2=7*449/(450-7)*finv(0.95,7,451-7);
Qlim_chi=chilimit(Q);
Qlim_jm=jmlimit(7,pcam.detail.ssq(:,2),0.95)
sQ=std(Q);
mQ=mean(Q);
Qlim_chi2=sQ^2/(2*mQ)*chi2inv(0.95,round(2*mQ/sQ^2))

% Projection of the validation data
[Tv1,Qv1,T2v1]=pcapro(Xv1,pcam,0);
[Tv2,Qv2,T2v2]=pcapro(Xv2,pcam,0);


% Contribution plots for diagnosis
T2conC=tconcalc(Xc,pcam);
QconC=qconcalc(Xc,pcam);

figure;hist(T2conC(:,1))
figure;normplot(T2conC(:,1))
figure;hist(T2conC(:,3))
figure;hist(QconC(:,1))
```

Matlab® code (4/4)

```
mT2c=mean(T2conC);
sT2c=std(T2conC);
mQc=mean(QconC);
sQc=std(QconC);

T2conV1=tconcalc(Xv1,pcam);
figure;bar(T2conV1(1,:));hold on;plot(mT2c-2*sT2c,'--m');plot(mT2c+2*sT2c,'--m');hold off;xlabel('variable');ylabel('T^2 contribuution');title('validation 1 - observation 1')
figure;bar(T2conV1(28,:));hold on;plot(mT2c-2*sT2c,'--m');plot(mT2c+2*sT2c,'--m');hold off;xlabel('variable');ylabel('T^2 contribuution');title('validation 1 - observation 28')

QconV1=qconcalc(Xv1,pcam);
figure;bar(QconV1(1,:));hold on;plot(mQc-2*sQc,'--m');plot(mQc+2*sQc,'--m');hold off;xlabel('variable');ylabel('Q contribuution');title('validation 1 - observation 1')
figure;bar(QconV1(50,:));hold on;plot(mQc-2*sQc,'--m');plot(mQc+2*sQc,'--m');hold off;xlabel('variable');ylabel('Q contribuution');title('validation 1 - observation 50')
figure;bar(QconV1(51,:));hold on;plot(mQc-2*sQc,'--m');plot(mQc+2*sQc,'--m');hold off;xlabel('variable');ylabel('Q contribuution');title('validation 1 - observation 51')
figure;bar(QconV1(54,:));hold on;plot(mQc-2*sQc,'--m');plot(mQc+2*sQc,'--m');hold off;xlabel('variable');ylabel('Q contribuution');title('validation 1 - observation 54')
```

… per sempre a fianco a me!