

UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

DEPARTMENT OF  
INDUSTRIAL ENGINEERING 

# Machine Learning Laboratory #2

Prof. Pierantonio Facco

CAPE-Lab, Computer-Aided Process Engineering Laboratory

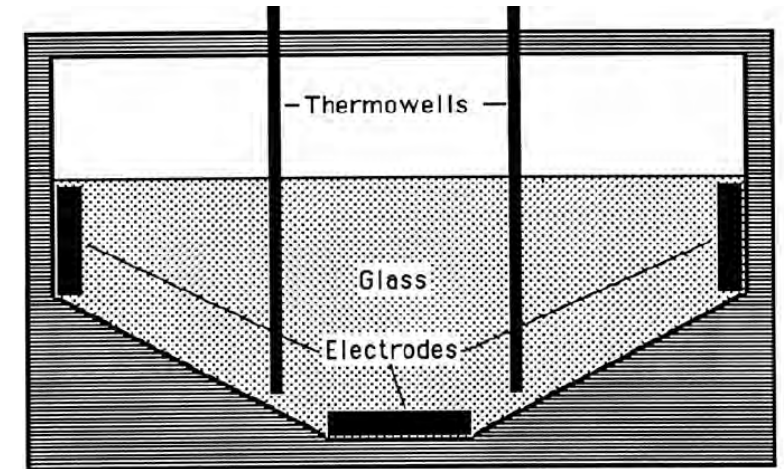
Email: [pierantonio.facco@unipd.it](mailto:pierantonio.facco@unipd.it)

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
- Objective: 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 X:

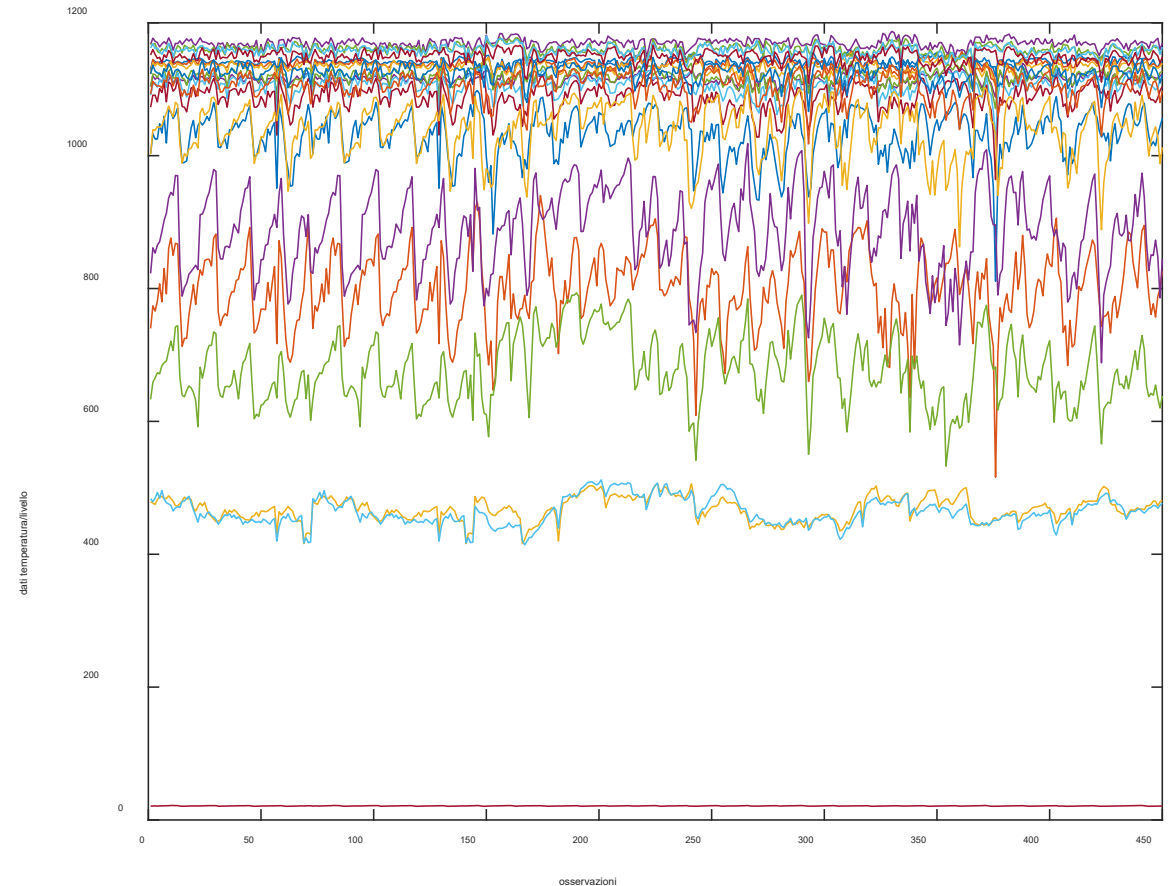
- 450 observations
- 21 variables:
  - temperatures in 20 points of the melter
  - glass level

## ■ Load:

- **load replacedata**

## ■ Data visualization:

- no anomalies seem to be present

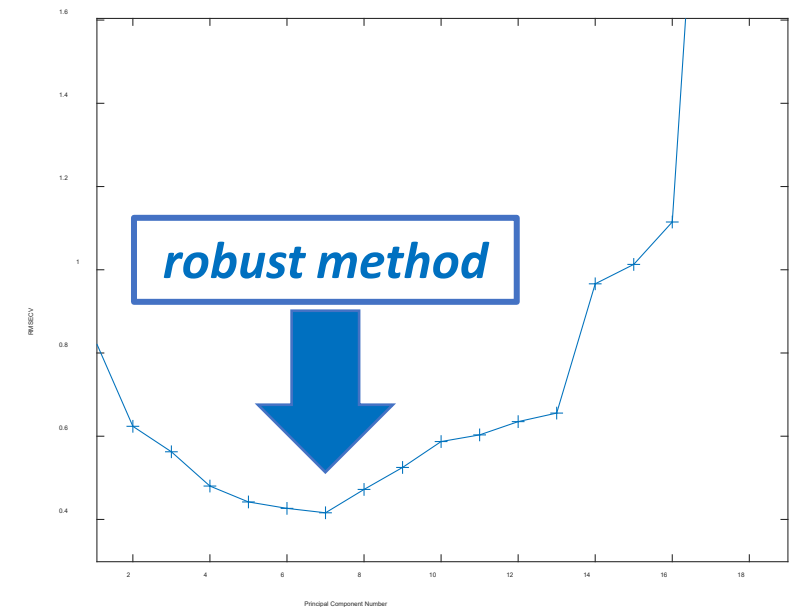
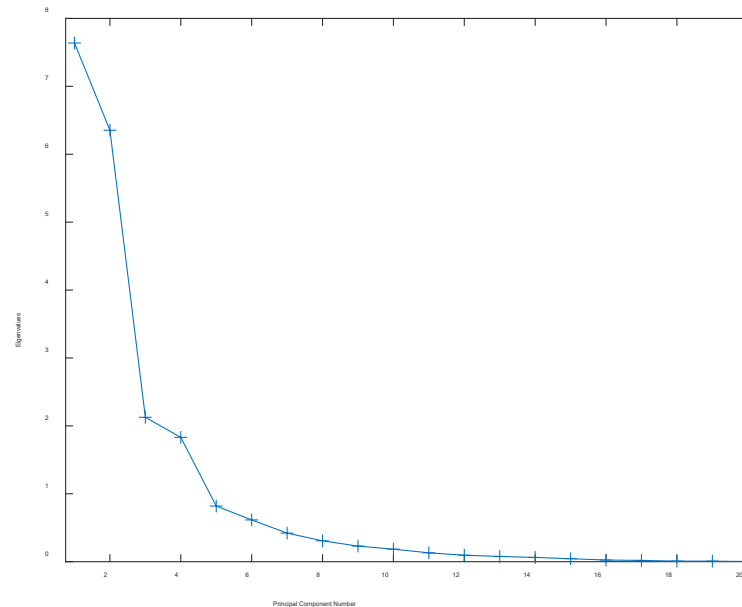
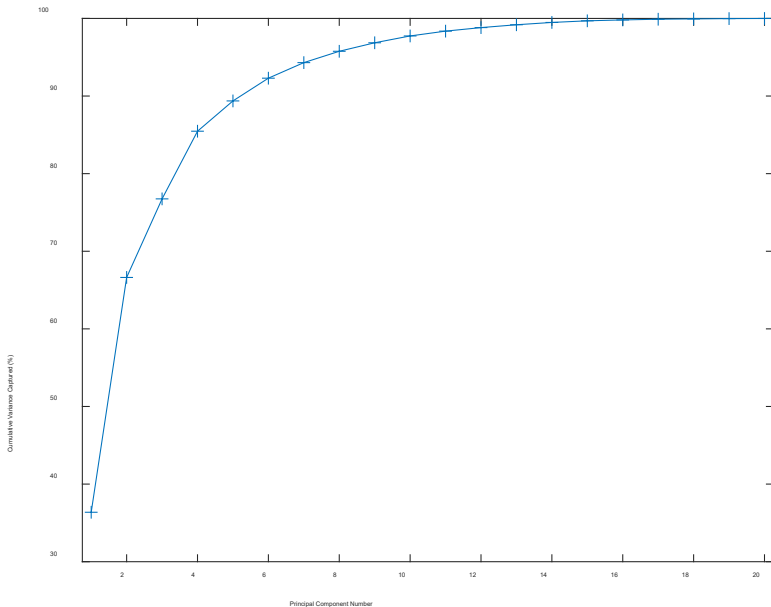


# Exploratory and monitoring model

- **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

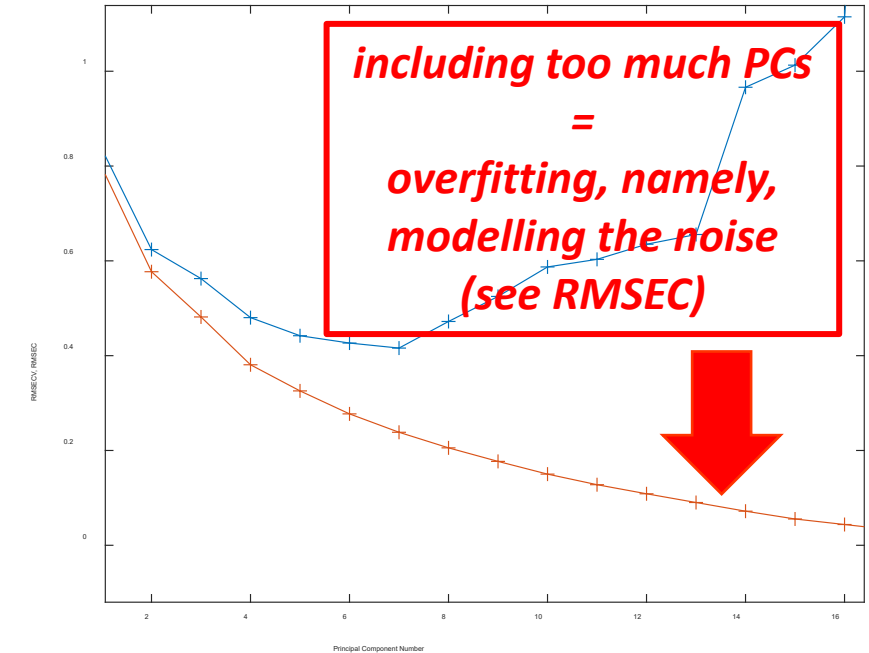
# Choice of the PCA model dimension

- 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

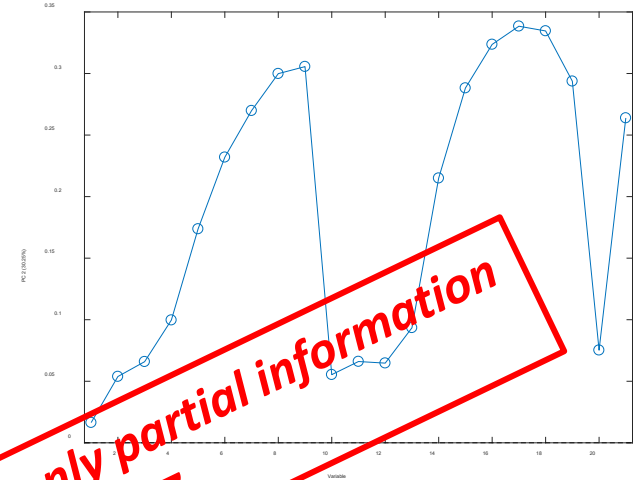
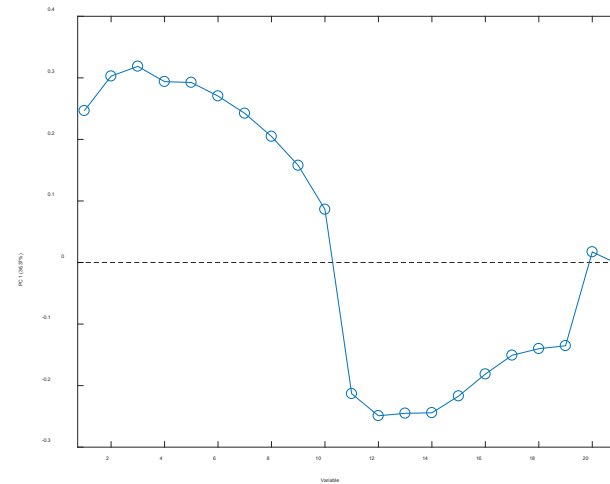


	Venetian Blinds	Contiguous Blocks	Random Subsets	Leave-One Out
Test sample selection scheme				

# Exploratory analysis

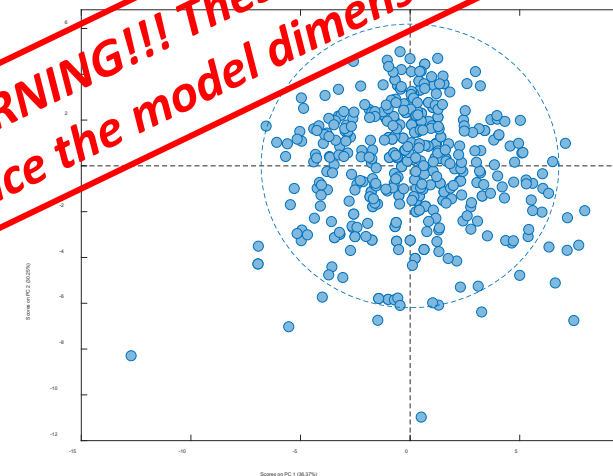
- Variables correlation:

- loading plot



- Relation among observations:

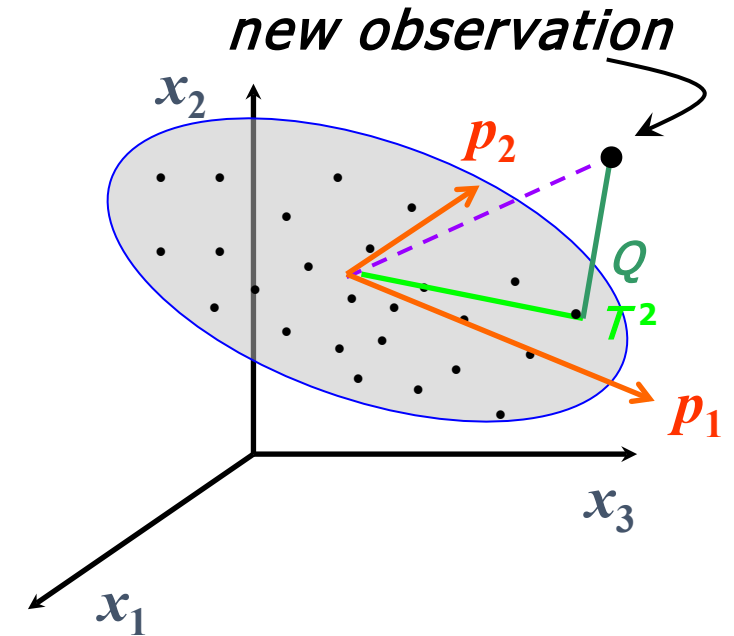
- score plot



**WARNING!!! These are only partial information since the model dimension is 7**

# 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
  - $T^2$  is a measure of the variability within the model



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

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

# Confidence limits

- Hotelling  $T^2$

- the limit comes from a F distribution of:

$$T_{\text{lim}}^2 = \frac{A(N-1)}{N-A} F_{A, N-A, 1-\alpha}$$

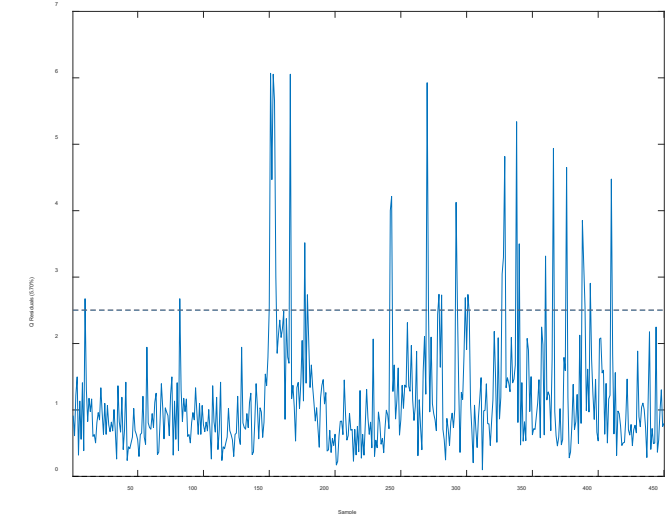
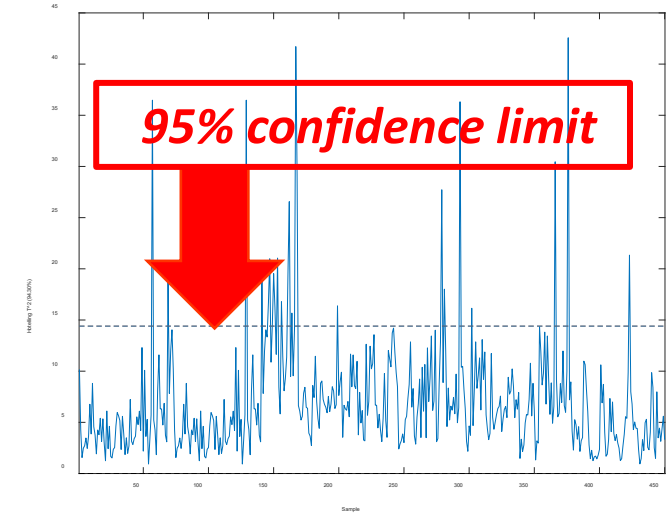
- Q residuals

- the limit comes from a  $\chi^2$  distribution of

$$Q_{\text{lim}} = \frac{s_Q^2}{2\mu_Q} \chi_{\frac{2\mu_Q}{s_Q^2}, 1-\alpha}^2$$

- 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**

Analysis - PCA 7 PCs - SFCM Calibration Data

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Load Data > Calibration  
Import Data > X-Block  
New Data > Y-Block  
Load Model > Load X and Y  
Import Model > Validation  
Load Prediction > X-Block  
Load Options > Y-Block  
Save Data > Load X and Y  
Save Model > Auto Select  
Export Model > Percent Variance Captured by PCA Model (\* = suggested)

	% variance	% variance	RMSEC	RMSECV		
	Component	Component				
1	37	36.37	0.7968	0.8355	suggested	
2	25	66.62	0.5771	0.6238		
3	13	76.75	0.4817	0.5626		
4	12	85.47	0.3808	0.4802		
5	8.20e-01	3.90	89.37	0.3257	0.4421	
6	6.15e-01	2.93	92.30	0.2772	0.4266	
7	4.21e-01	2.00	94.30	0.2384	0.416	current
8	3.07e-01	1.46	95.77	0.2055	0.4721	
9	2.30e-01	1.10	96.86	0.177	0.5249	
10	1.85e-01	0.88	97.74	0.1501	0.5872	
11	1.30e-01	0.62	98.36	0.1278	0.6032	
12	9.54e-02	0.45	98.82	0.1086	0.6351	
13	7.71e-02	0.37	99.18	0.09023	0.6556	
14	6.25e-02	0.30	99.48	0.07191	0.9662	
15	4.41e-02	0.21	99.69	0.05548	1.013	
16	2.44e-02	0.12	99.81	0.04381	1.115	
17	1.86e-02	0.09	99.90	0.03215	2.529	
18	8.86e-03	0.04	99.94	0.02476	3.756	
19	8.05e-03	0.04	99.98	0.01518	6.857	

A model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing options, and other settings can also be modified to adjust the model. The data can be viewed and edited from the Edit menu.

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model

Review Model

5. Choose Components
6. Review Scores
7. Review Loadings

Use Model

8. Load Test Data
9. Apply Model

Cache: "general" DATE View (\* = Not Available)

- Cache Settings and View
- Demo Data
- 02-Jul-2017
- 01-Jul-2017
- 28-Jun-2017
- 20-Jun-2017
- 15-Jun-2017
- 05-Jun-2017
- 18-May-2017
- 16-May-2017
- 27-Apr-2017
- 19-Apr-2017
- 07-Apr-2017
- 04-Apr-2017

# Validation observation #1

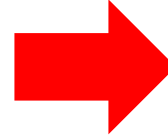
- New data  $\mathbf{x}_{\text{NEW}}$  is projected onto the hyperplane of the PCs:

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

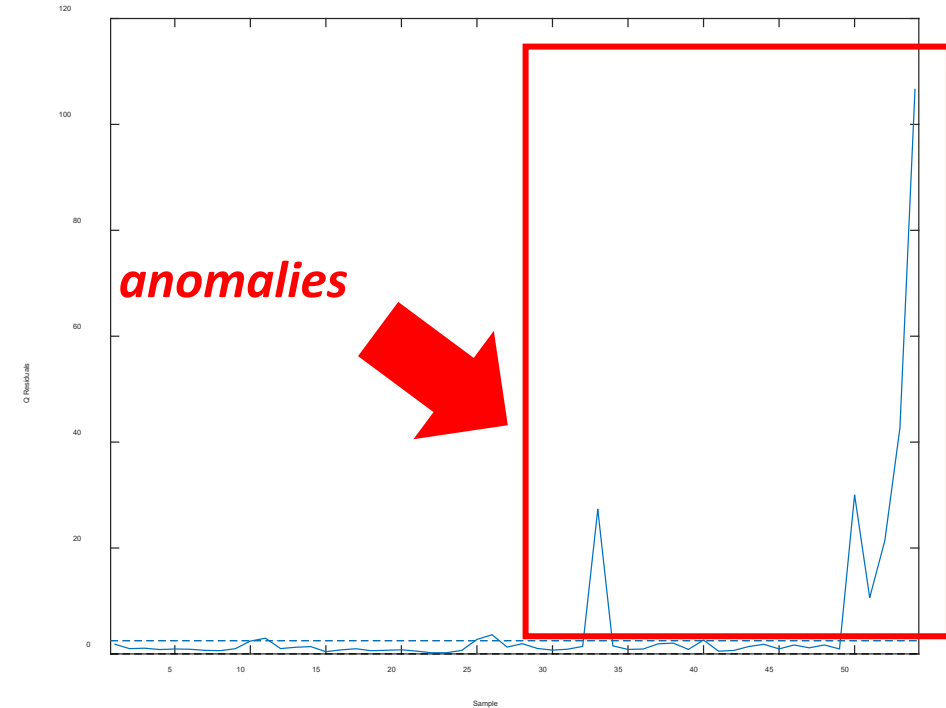
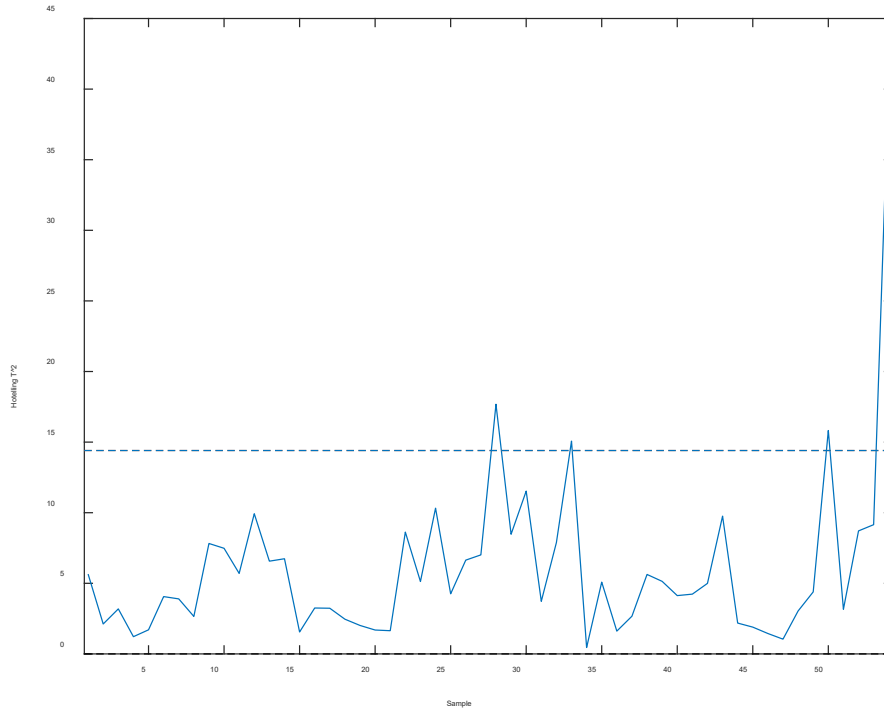
- The coordinates of the new observation in the PCs space are obtained
- Observe if the Hotelling and  $Q$  indices are well within the **confidence limits** obtained from the NOC calibration data

# Process monitoring

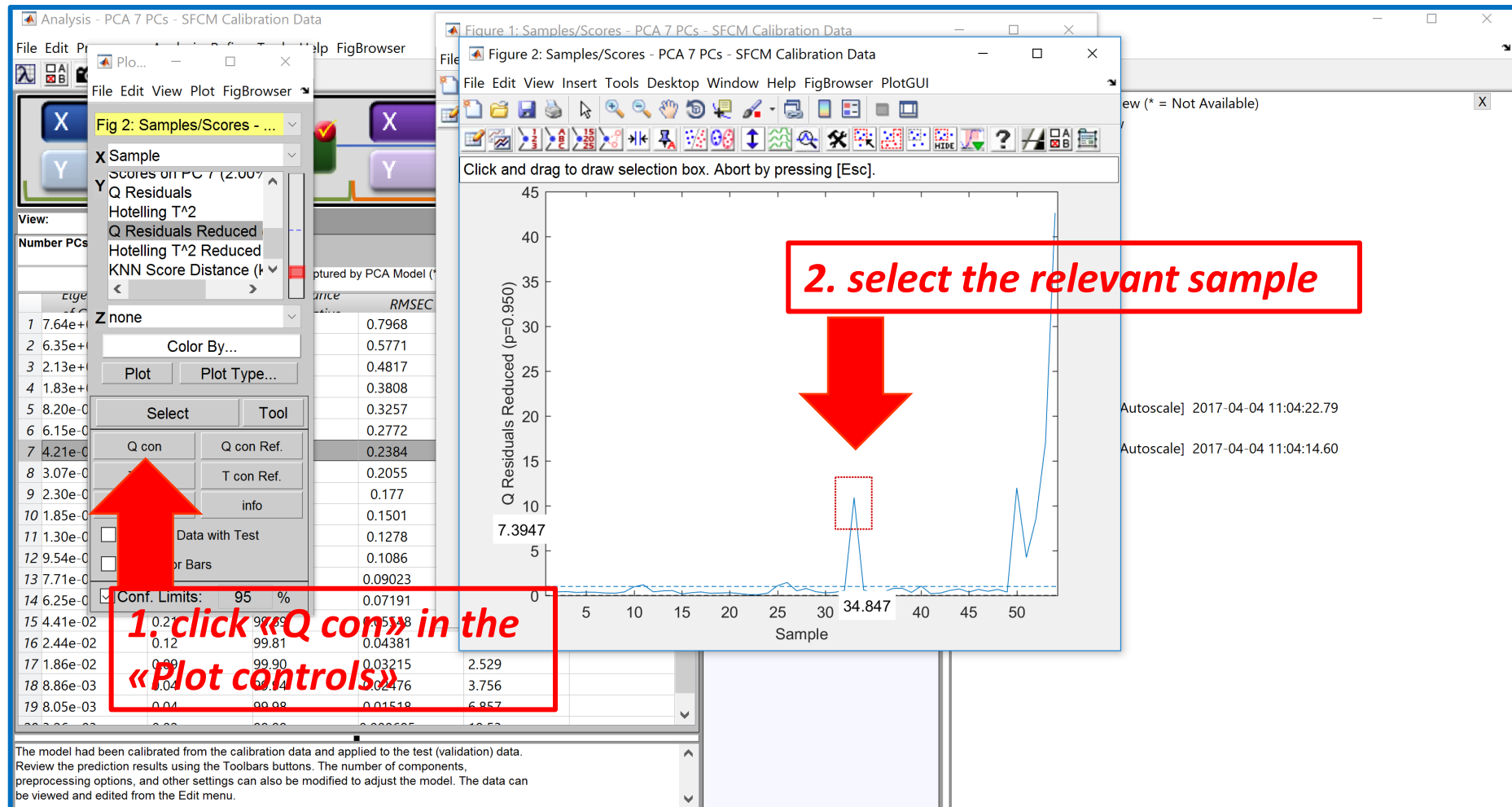
- Anomalies are detected for:
  - observation 33
  - observation 50-54



*diagnosis is needed!*

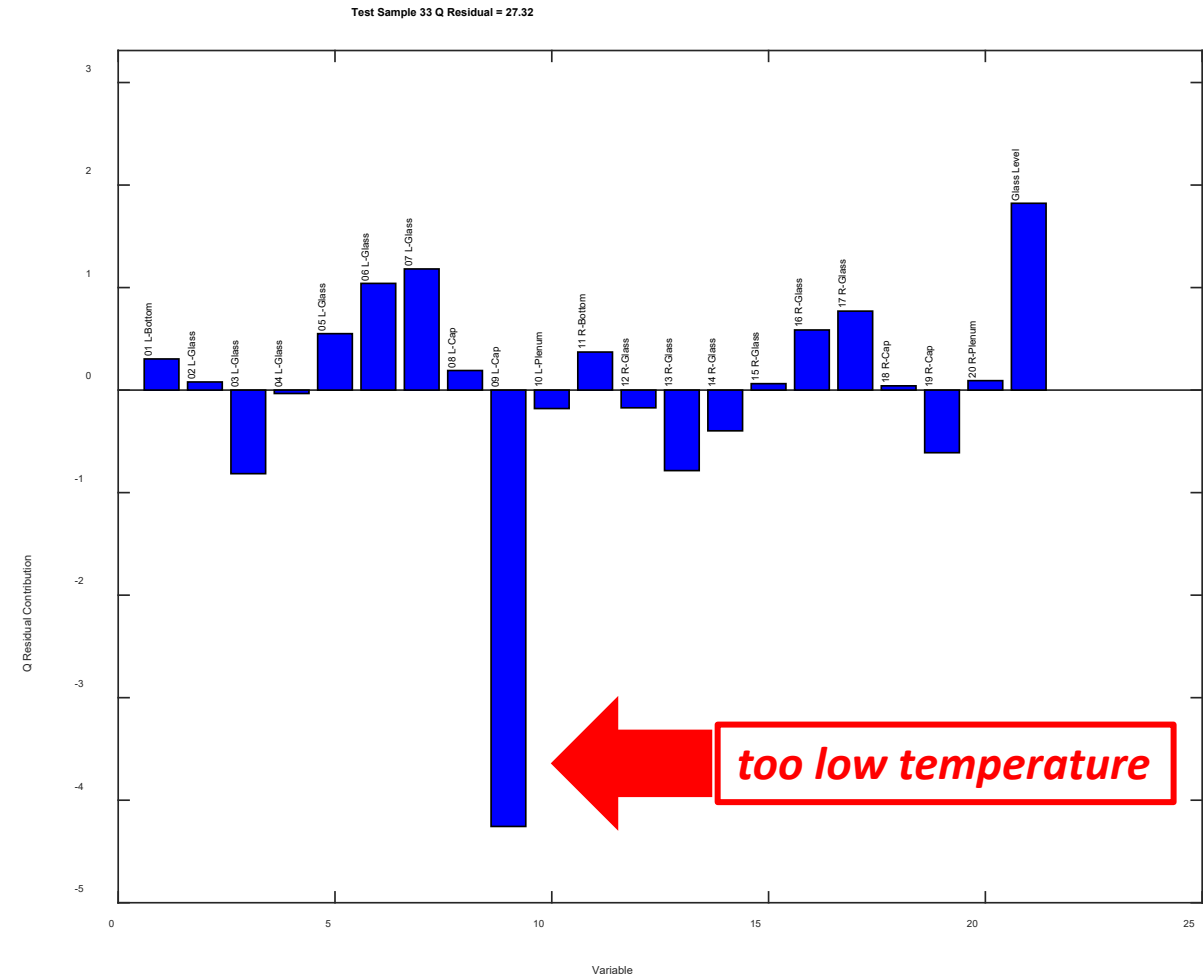


# Diagnosis of anomalies: contribution plots for obs. 33



# 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!!!

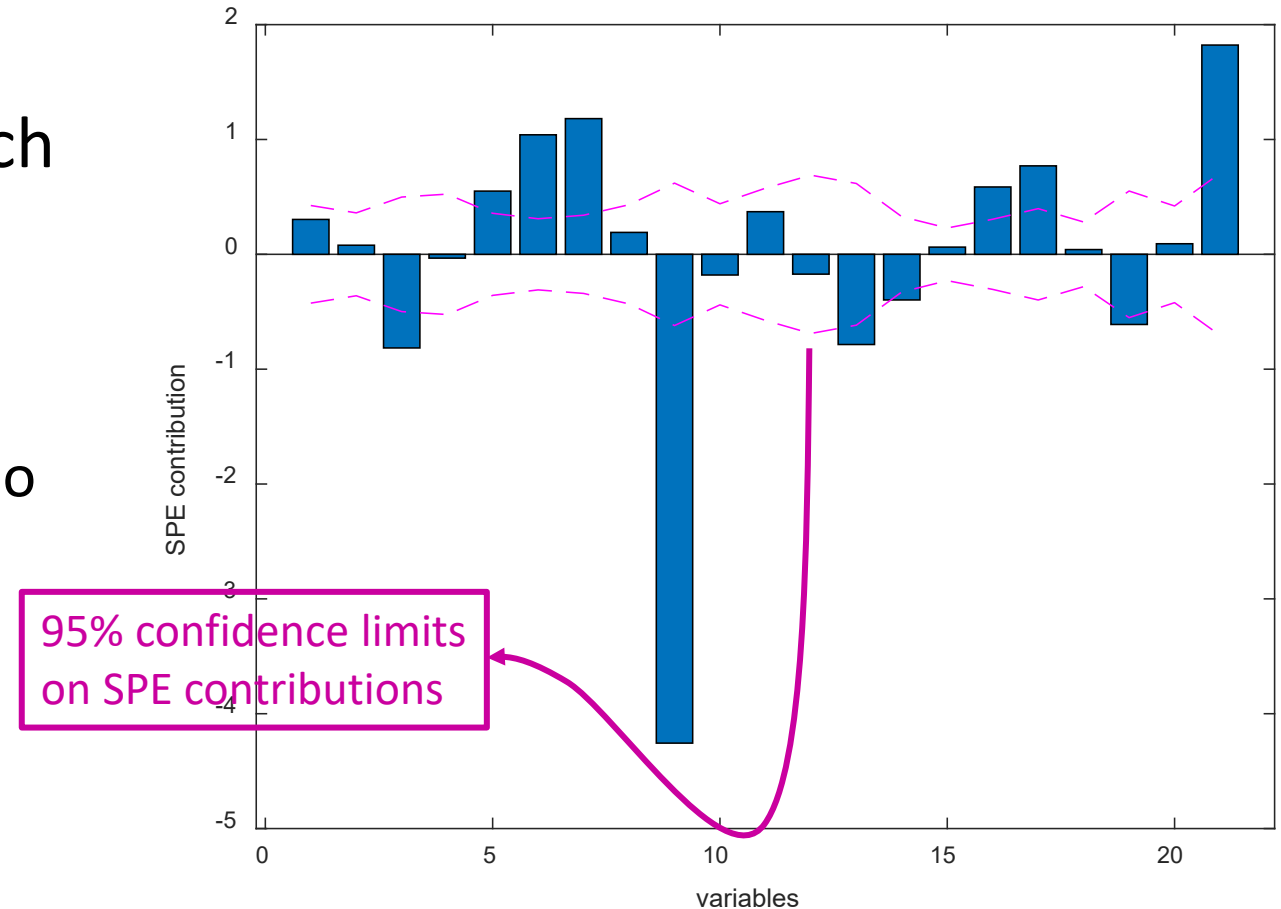


# Improved diagnosis of the anomalies

- **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

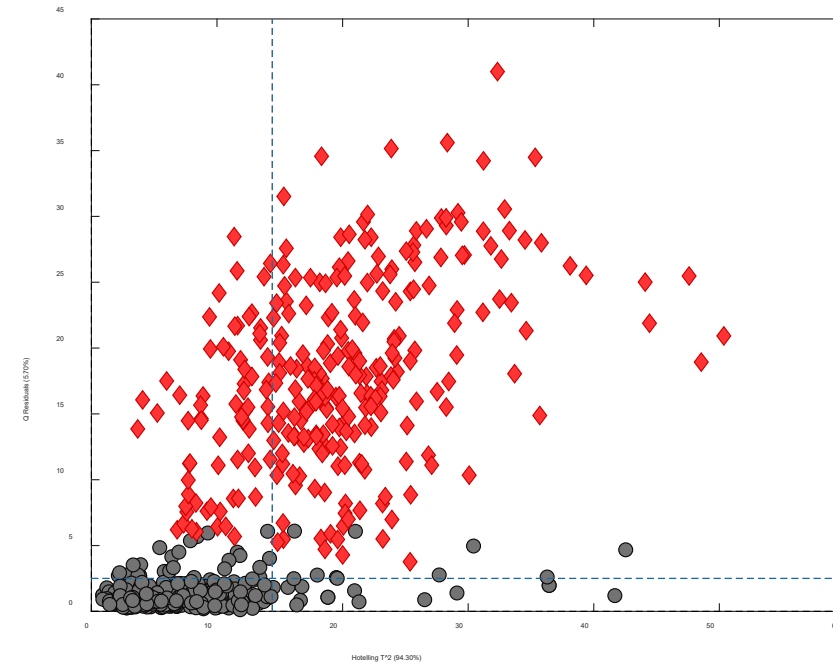
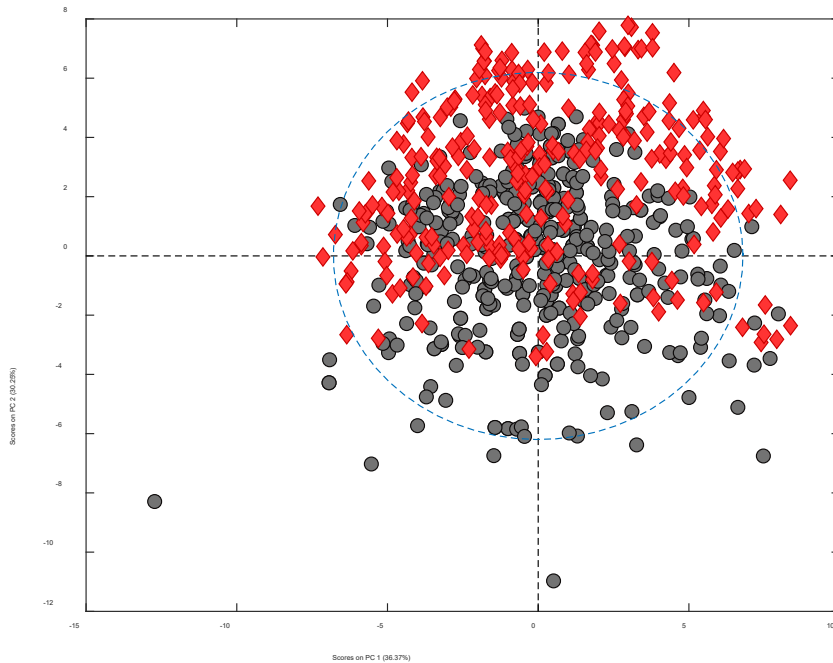
- The improved diagnosis tell us much more: the issue here is related to more than one variable!
  - variable 9 is highly deviating
  - variables 5, 6, 7, 16, 17 and 21 are too high
  - variables 3 and 13 are too low



# 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

```
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');
```



% 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))
```



```
% 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))
```



```

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!

