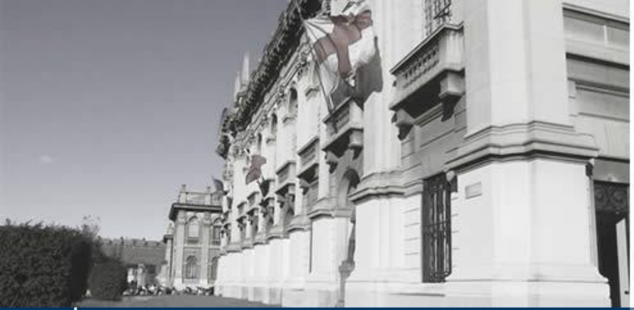


POLITECNICO DI MILANO

Exercises Session on Fault Detection

Stefano Marchetti
stefano.marchetti@polimi.it

Exercise 1

Method: AAKR (Signal Reconstruction)+*SPRT* (Decision)

Component: Gas Turbine

Exercise 2

Method: PCA

Component: Gas Turbine

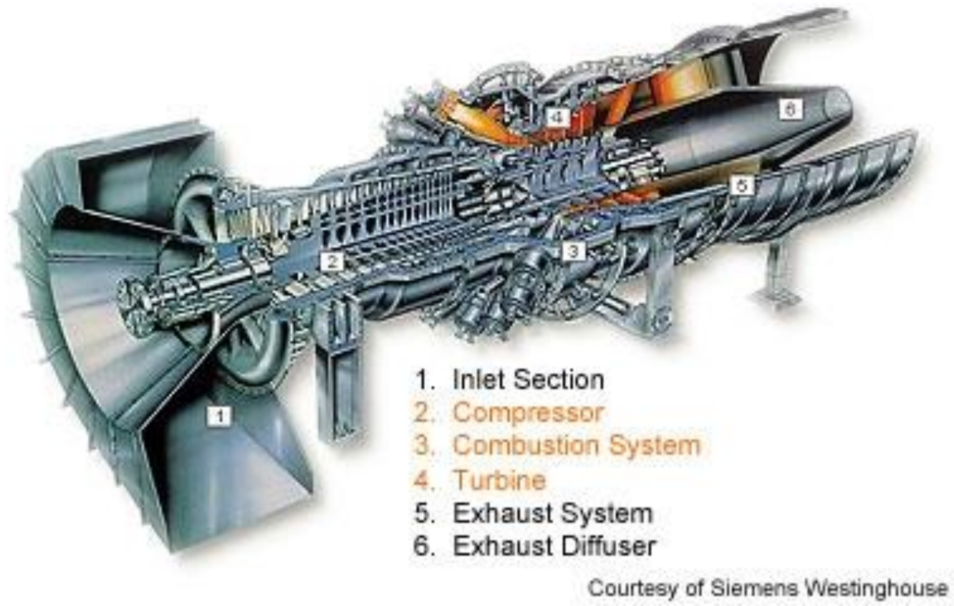
Exercise 3 (take home)

Method: you choose

Component: Wind Turbine

Exercise 1

Component: Gas Turbine

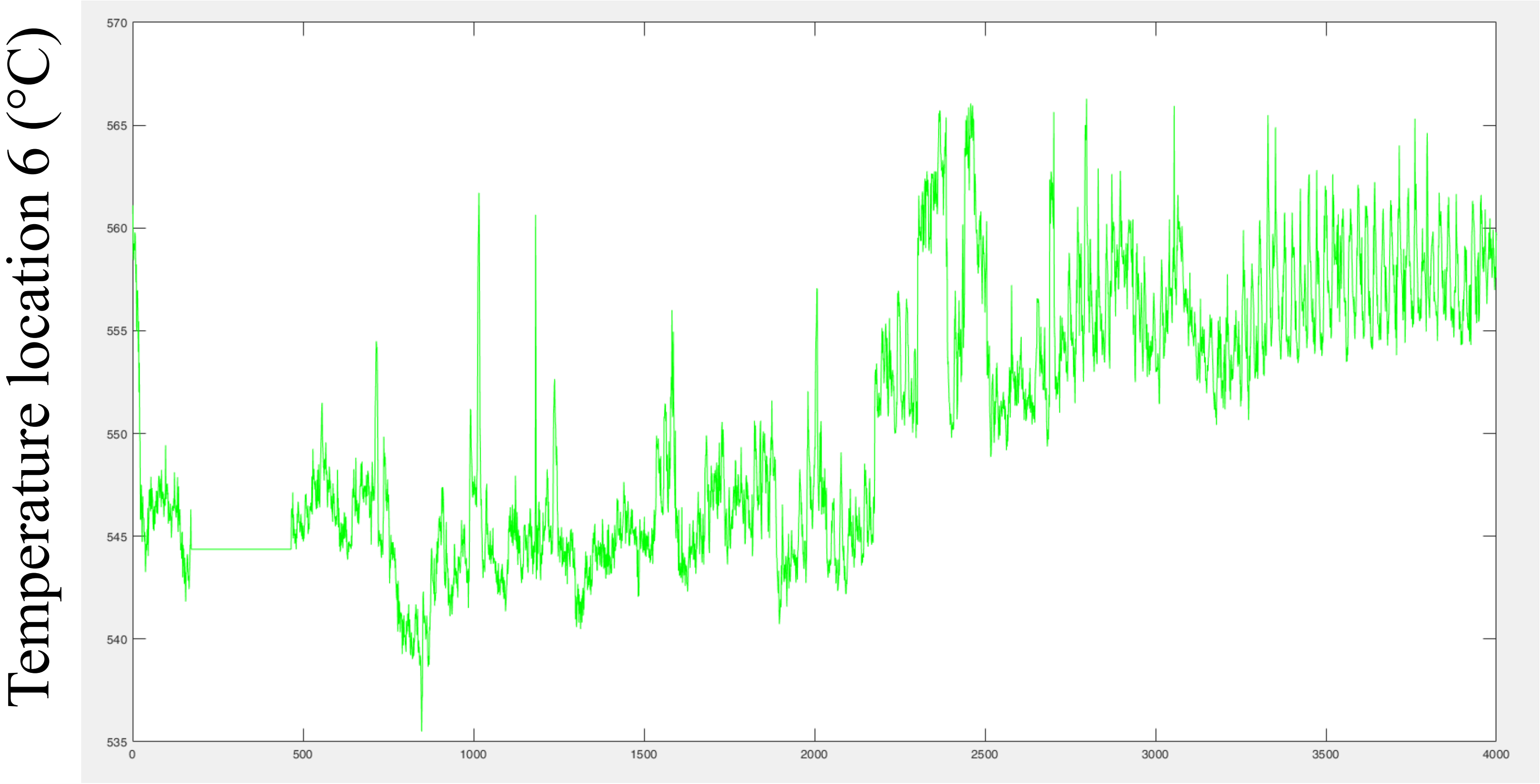


Measured Signals

Temperature location 1 (°C)
Temperature location 2 (°C)
Temperature location 3 (°C)
Temperature location 4 (°C)
Temperature location 5 (°C)
Temperature location 6 (°C)

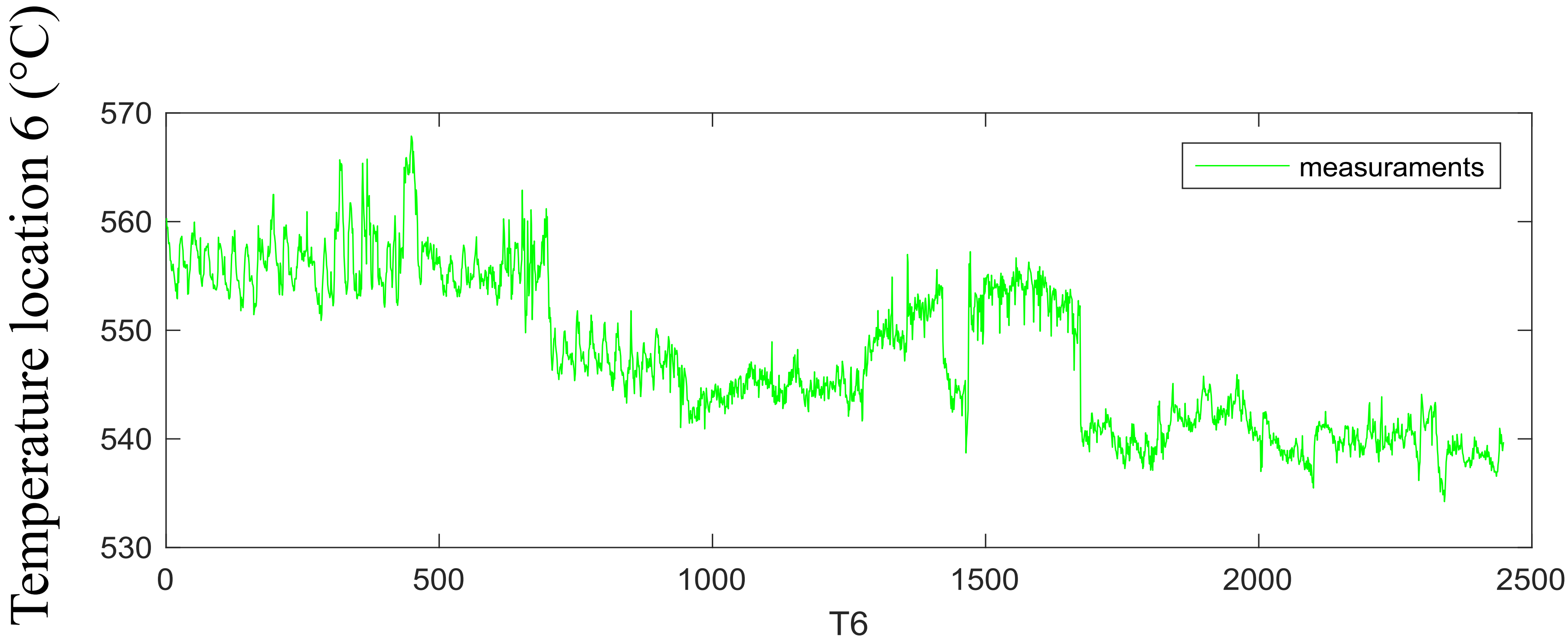
Available Data

- Train.dat → Normal condition data [6 signals, 4000 data points, frequency: 5200 measurements/year]



Available Data

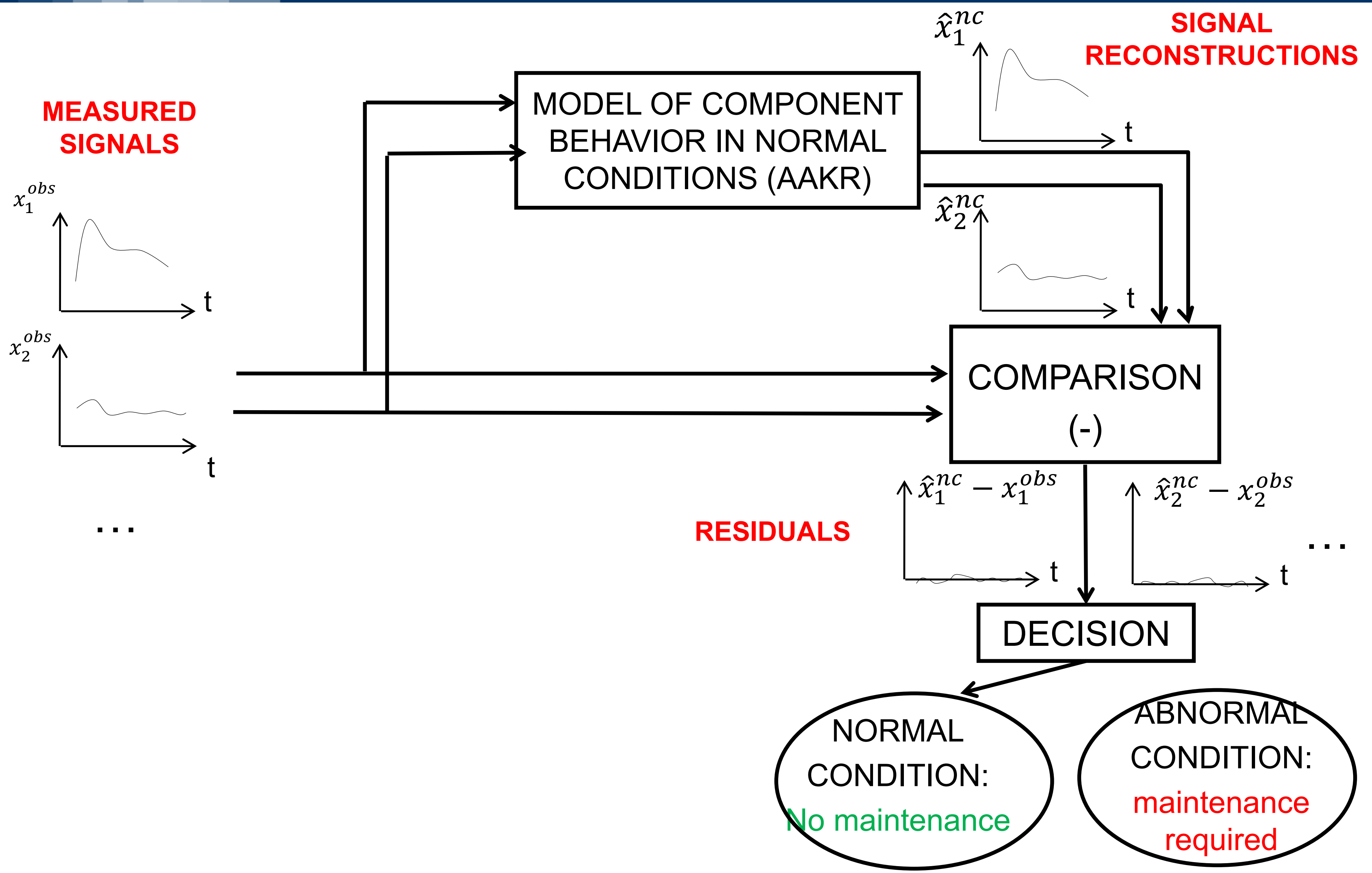
- Validation.dat → Normal condition data [6 signals, 2500 data points, frequency: 5200 measurements/year]



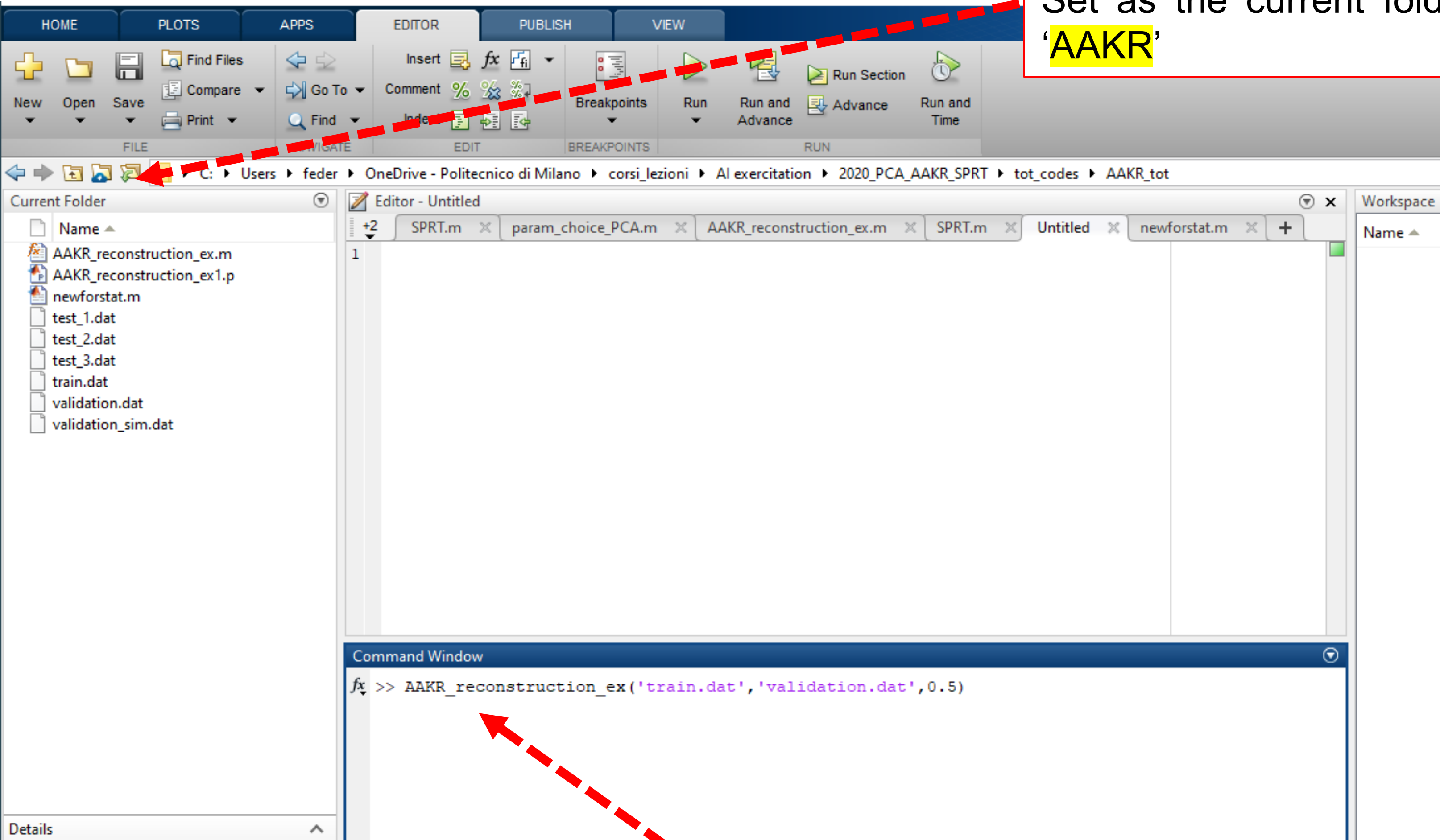
Exercise 1

Method: AAKR (Signal Reconstruction)+*SPRT* (Decision)

Fault Detection using AAKR for signal reconstruction



The AAKR code in Matlab: how to run the code?



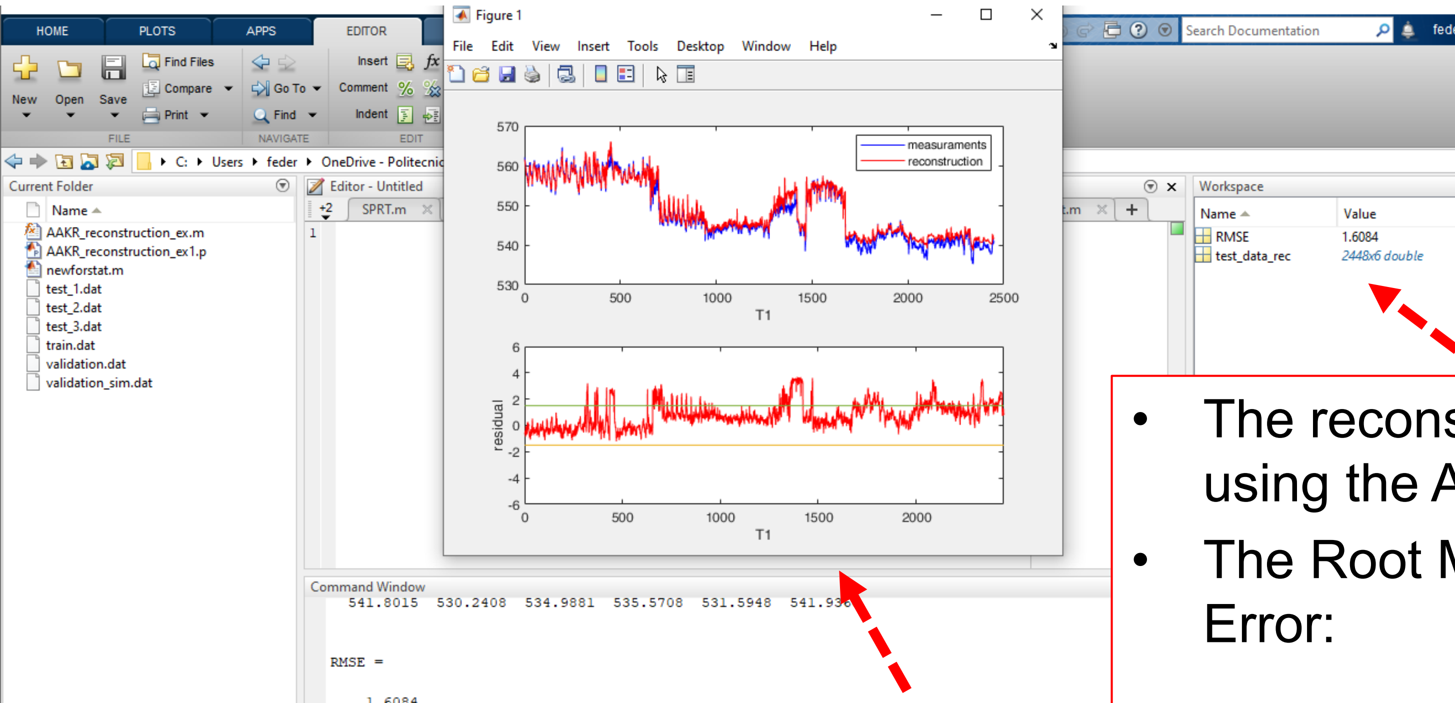
Set as the current folder the folder 'AAKR'

Run the code in the command window:

$[test\_data\_rec, RMSE]=AAKR\_reconstruction\_ex('\mathrm{train.dat}','\mathrm{validation.dat}', \mathrm{h})$

h: bandwidth parameter

The AAKR code in Matlab: Output



One figure for each signal representing:

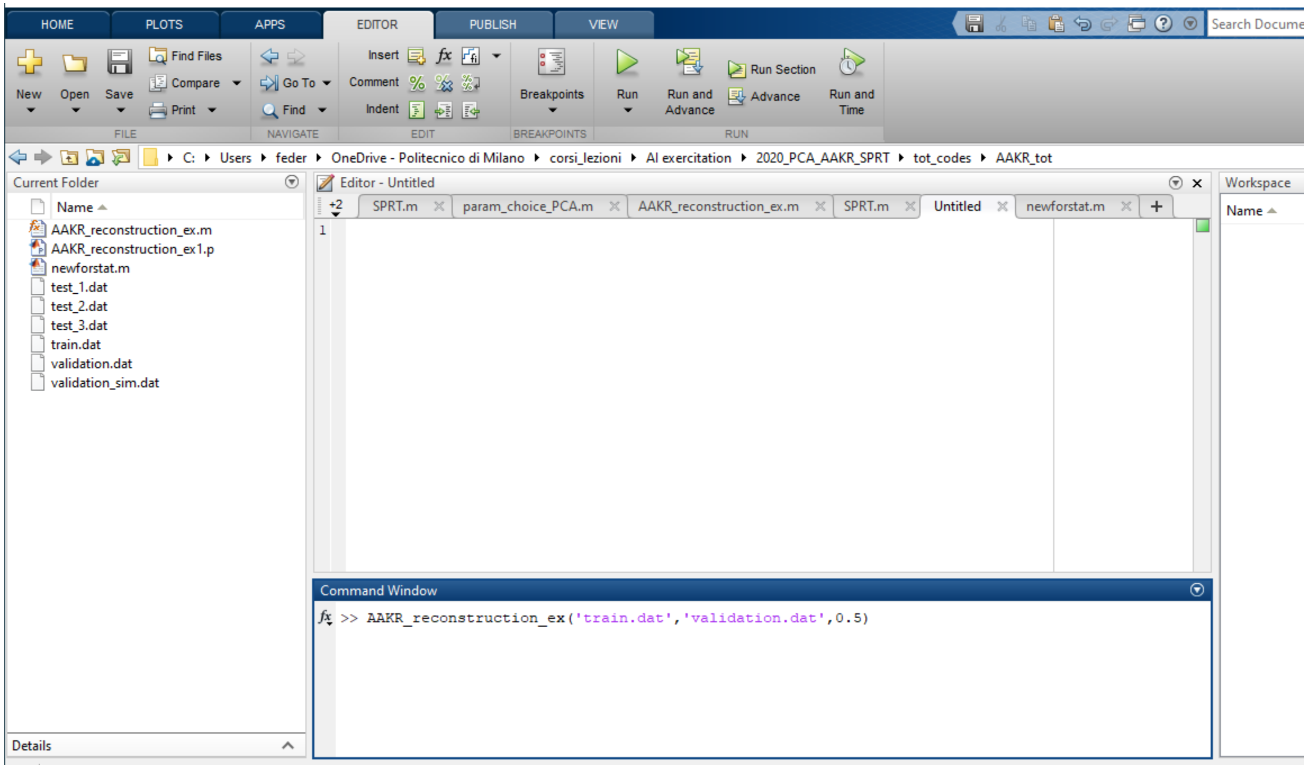
- The original data and the reconstruction obtained using the AAKR algorithm
- The residual for each data point

- The reconstruction obtained using the AAKR algorithm
- The Root Mean Square Error:

$$RMSE = \frac{\sum_{j=1}^{n} \frac{\sum_{t=1}^{N_{valid}} \left|\vec{\hat{x}}^{nc}(t,j) - \vec{x}^{obs}(t,j)\right|^2}{N_{valid}}}{n}$$

The AAKR code in Matlab: Example

5 minutes



Run the algorithm considering the bandwidth parameter *h*=0.1

Execute the code in the command window:

[test_data_rec, RMSE]=$AAKR\_reconstruction\_ex(\text{'train.dat'},\text{'validation.dat'},\text{h})$

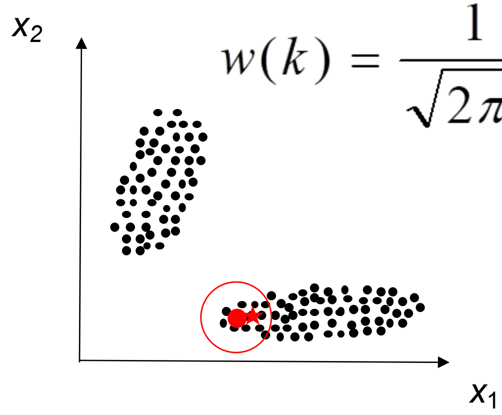
Exercise 1

Method: AAKR

Setting the model parameter:

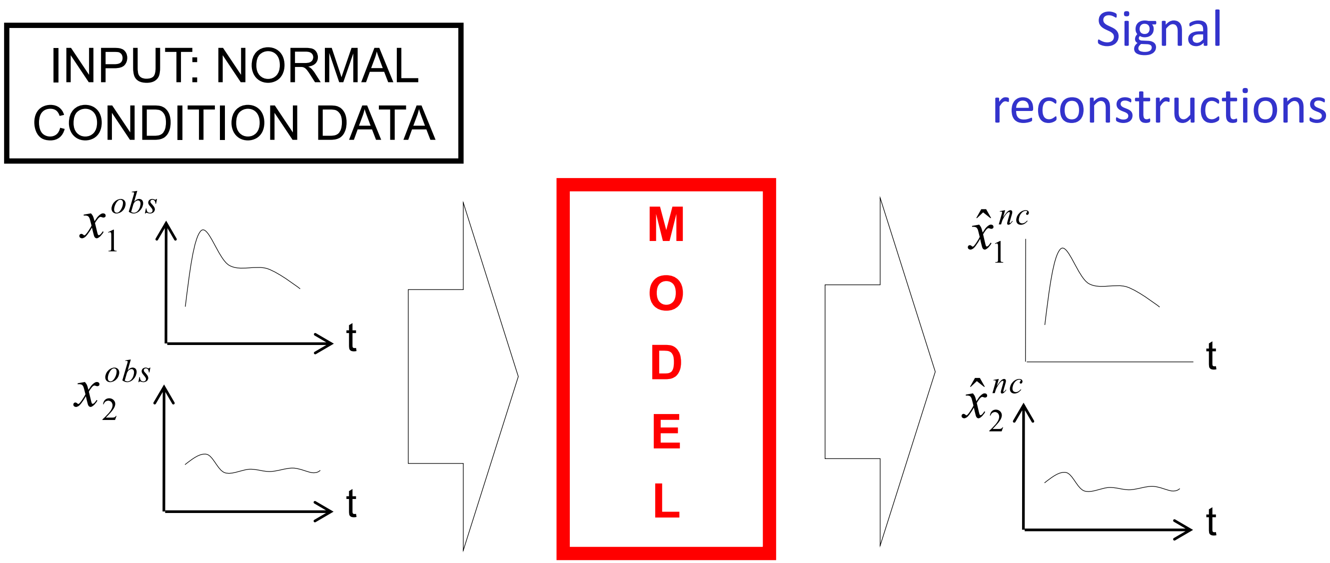
- *h =bandwidth parameter*

$$w(k) = \frac{1}{\sqrt{2\pi h}} e^{-\frac{d^2(k)}{2h^2}}$$



- Objective: ACCURATE and ROBUST reconstruction model

Accuracy



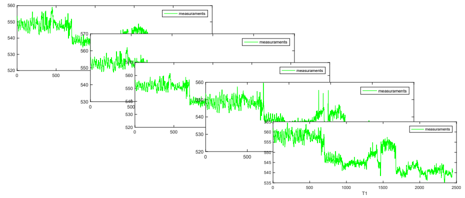
- Accurate reconstruction model if:

$$\vec{\hat{x}}^{nc} \cong \vec{x}^{obs}$$

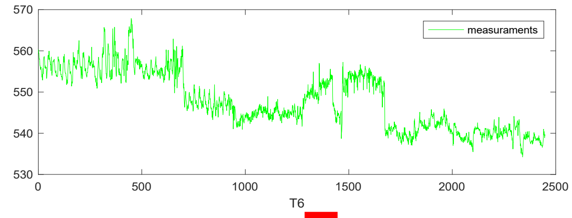
- Metric to measure accuracy: Root Mean Square Error ($RMSE$):

$$RMSE = \frac{\sum_{j=1}^{n} \frac{\sum_{t=1}^{N_{valid}} \left| \vec{\hat{x}}^{nc}(t,j) - \vec{x}^{obs}(t,j) \right|^2}{N_{valid}}}{n}$$

Accuracy – Results wit *h*=0.05 and *h*=0.4



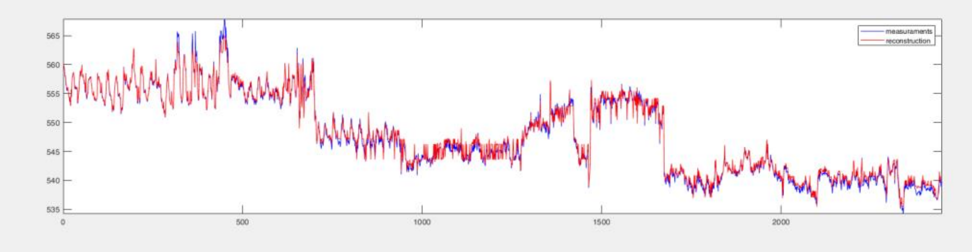
Signal 6 normal condition

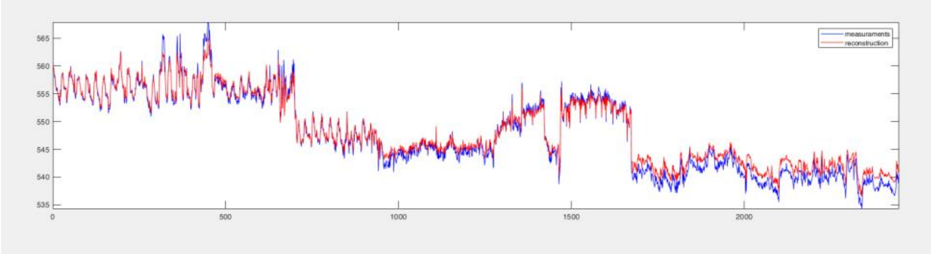


AAKR

h=0.05

h=0.4



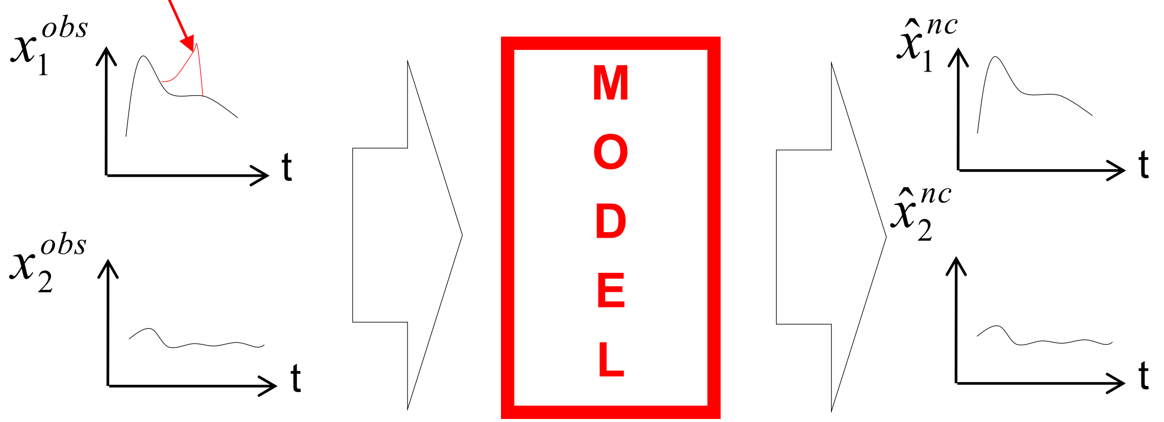


h=0.05 results in a more accurate model

Robustness

INPUT: ABNORMAL CONDITION DATA (Simulated)

Signal reconstructions

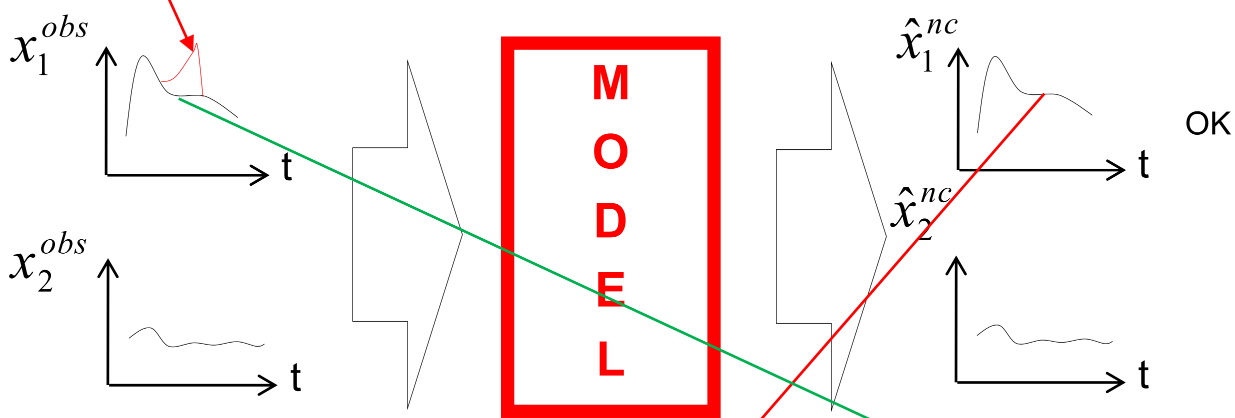


- Robust reconstruction if: $\vec{\hat{x}}^{nc} \cong \vec{x}^{obs-nc}$

Robustness

INPUT: ABNORMAL CONDITION DATA (Simulated)

Signal reconstructions



- Robust reconstruction if: $\hat{\vec{x}}^{nc} \cong \vec{x}^{obs-nc}$

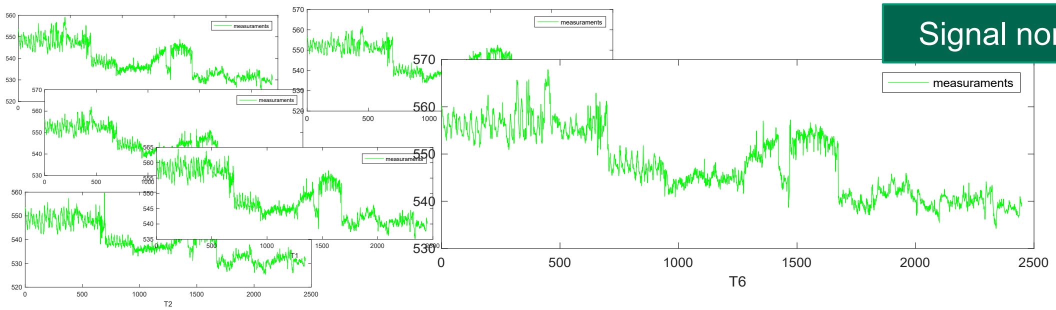
- Metric to measure Robustness:

Reconstruction of the anomaly

Normal condition data (before anomaly simulation)

$$\text{Robustness} = \frac{\sum_{t=1}^{N_{valid}} \left| \hat{\vec{x}}_1^{nc}(t) - \vec{x}_1^{obs-nc}(t) \right|^2}{N_{valid}}$$

How to Verify Robustness in Practice?

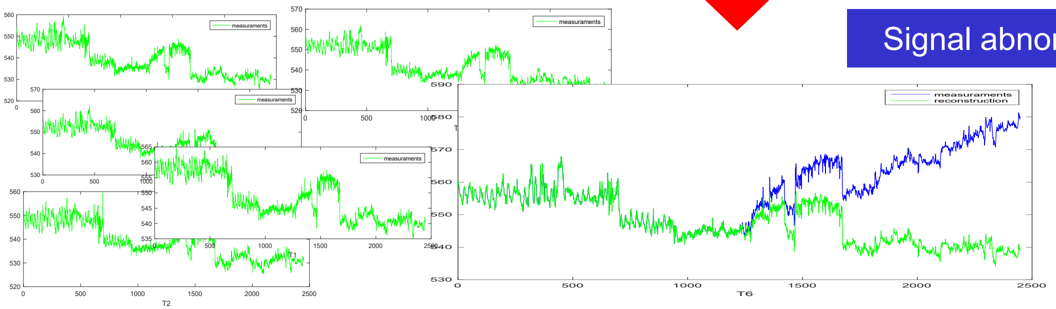


Signal normal condition

Artificial simulated abnormal condition on signal 6

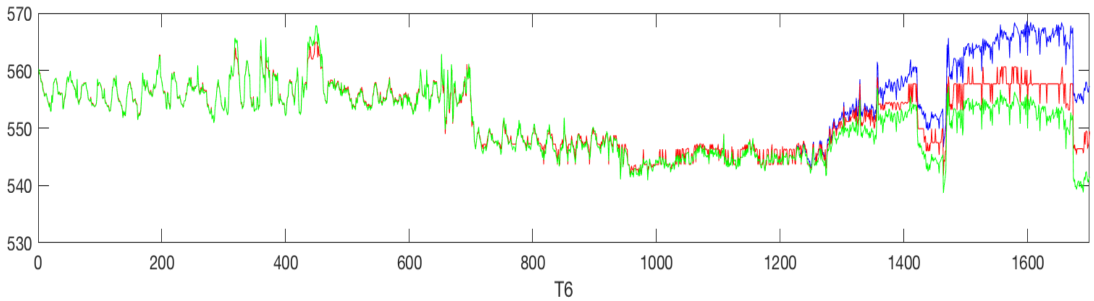
The other signals remain in normal condition

Signal abnormal condition



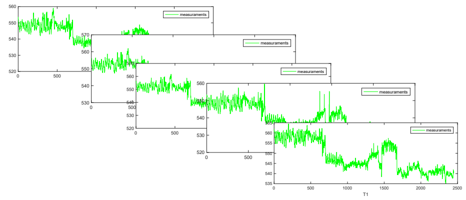
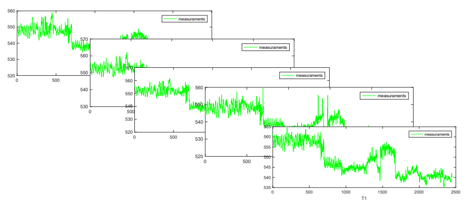
AAKR

-- Signal normal condition
-- Signal abnormal condition
-- AAKR signal reconstruction

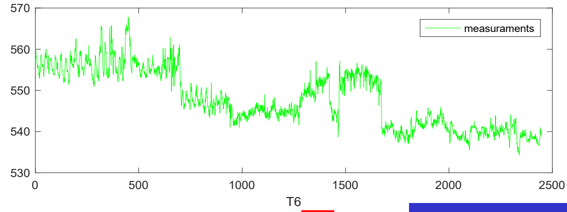


The method is robust if AAKR reconstruction is equal to the signal in normal condition

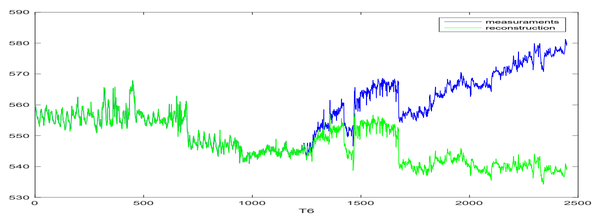
Robustness – Results wit *h*=0.05 and *h*=0.4

Signal normal condition



Signal abnormal condition



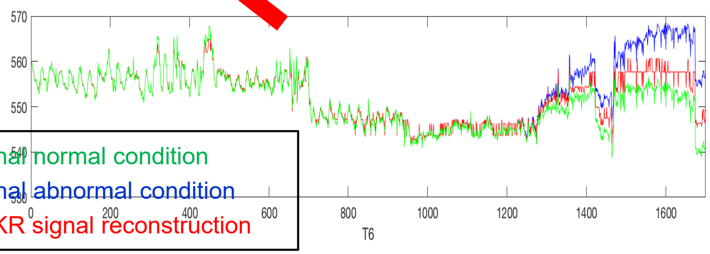
not robust
but accurate

AAKR

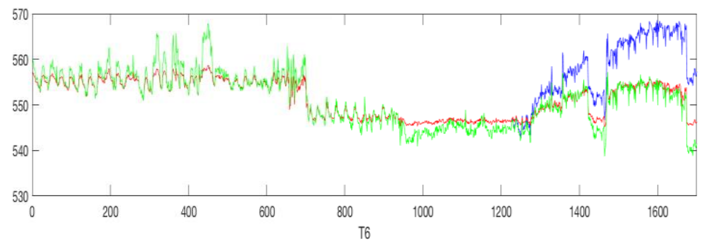
h=0.05

h=0.4

More robust
but less
accurate



-- Signal normal condition
-- Signal abnormal condition
-- AAKR signal reconstruction



Optimal value of *h* is a trade off between accuracy and robustness

Code to measure the method robustness


```
val = load('validation.dat');
[npat,nsig]=size(val);
anomaly=load('validation_sim.dat');

[test_reconstruction,rmse]=AAKR_reconstruction('train.dat','validation_sim.dat',0.05);

robustness=sum(((test_reconstruction(:,1)-val(:,1)).^2)/npat)^0.5
```

Exercise 1.1: Assignment

- Set the bandwidth parameter (h) of the AAKR-based reconstruction model

You can use the following files:

- Validation.dat → data in normal condition
- Validation_sim.m→ data containing a simulated abnormal condition on **signal 1**

30-40 minutes

When you don't need the plots of the signals, you can add one more parameter to the function and set it to 0:

[test_data_rec,rmse]=AAKR_reconstruction_ex(train_data_name,test_data_name,h,**0**)

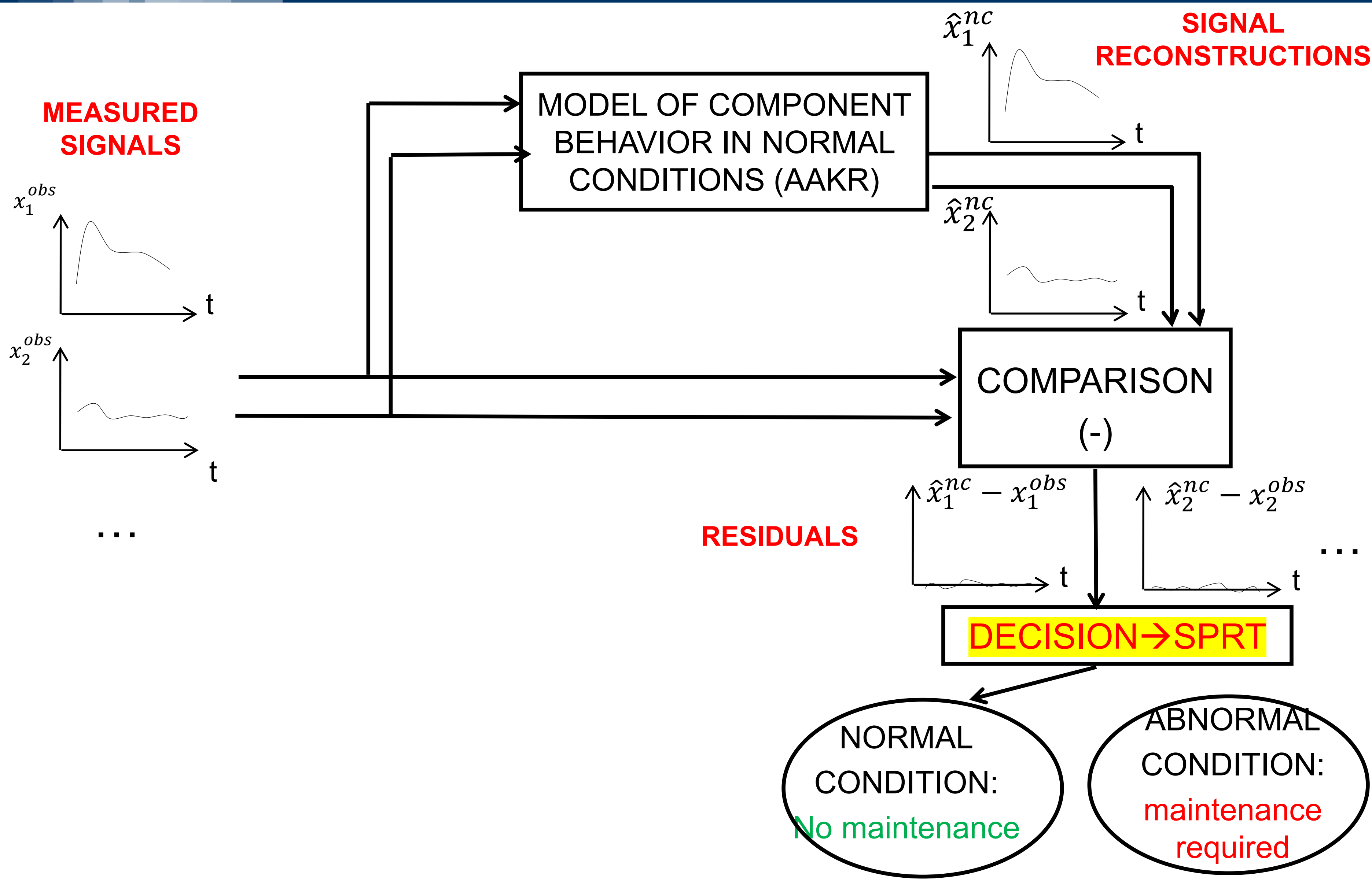
Parameter to avoid plotting

Exercise 1.2 - Assignment

- Perform the reconstruction of the signal measurements in the 4 files test_1.dat, test_2.dat, test_3.dat and test_4.dat.
- In which files can you detect abnormal conditions? Do you have any hypothesis on the type of abnormal condition?
- Draw your conclusions on the possibility of using the developed model for fault detection.

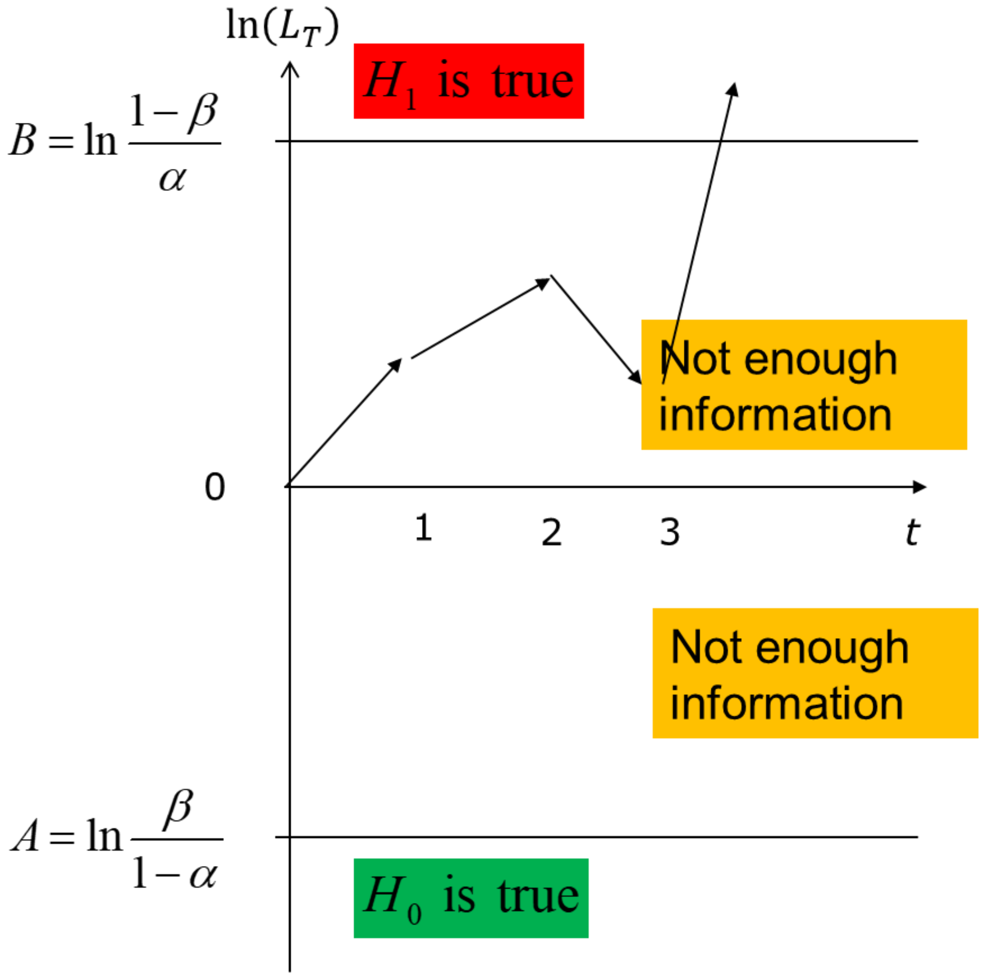
30 minutes

Sequential Probability Ratio Test



SPRT

$$L_0 = 1 \rightarrow \ln(L_0) = 0$$
$$\ln(L_1) = \ln(L_0) + \frac{\mu_1}{\sigma^2}\left(r^{(1)} - \frac{\mu_1}{2}\right)$$
$$\ln(L_2) = \ln(L_1) + \frac{\mu_1}{\sigma^2}\left(r^{(2)} - \frac{\mu_1}{2}\right)$$



MATLAB FUNCTION

$$[SPRT_{Index}] = SPRT(test_{name}, test\ rec_{name}, \alpha, \beta, \mu_{H_0}, \sigma_{H_0}, \mu_{H_1}, \sigma_{H_1})$$

OUTPUT:

- $SPRT_{Index}$ = N-by-1 vector containing the $SPRT$ index

INPUTS:

- $test_{name}$ = N-by-1 data matrix containing *N* sequential measurements of one signal
- $test\ rec_{name}$ = N-by-1 data matrix containing the *N* reconstructions of the data in $test_{name}$.
- α = False alarm rate
- β = Missed alarm rate.
- μ_{H_0}= the mean of the Gaussian distribution of H_0 hypothesis (the component is working in normal conditions).
- σ_{H_0}= variance of Gaussian distribution H_0
- μ_{H_1}= the mean of the Gaussian distribution of H_1 hypothesis (the component is working in abnormal conditions).
- σ_{H_1}= variance of Gaussian distribution H_1

Example:
[$SPRT_{Index}$] = $SPRT$('test_4A_meas_S5.dat','test_4A_rec_S5.dat',0.01,0.01,0,0.15,1,0.15)

Exercise 1.3: Assignment

- Apply the SPRT to detect abnormal conditions to the measurements of signal 5 in files "test_4A.dat" and "test_4B.dat". The following setting of the SPRT parameters is suggested:

  - $\alpha = 0.01$
  - $\beta = 0.01$

  - and Gaussian distributions for the two hypothesis H_0 and H_1 with parameters:

    - $\mu_{H_0} \approx 0,\ \sigma_{H_0} = 0.15$
    - $\mu_{H_1} = 1,\ \sigma_{H_1} = 0.15$

- Do you detect the occurrence of abnormal conditions? When?
- Compare the results with those of a threshold-based method with threshold = 1.
- Draw your conclusions on the possibility of using the SPRT in decision-making for fault detection.

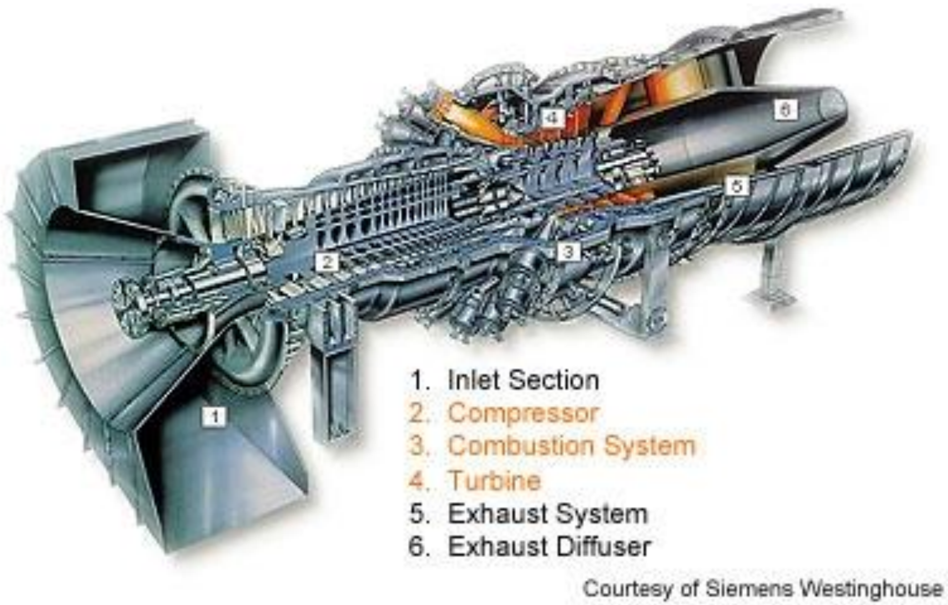
Exercise 2

Fault detection with PCA considering turbine data

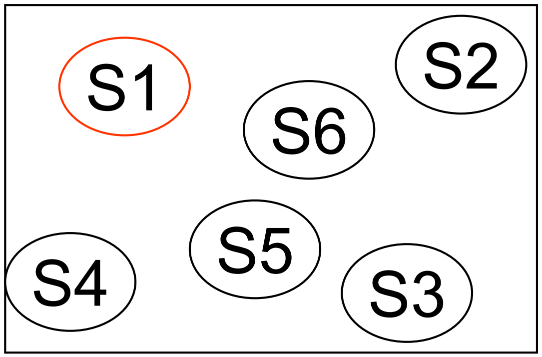
1. *Optimize the PCA to achieve:*
    a. *Accuracy*
    b. *Robustness*
2. *Apply the PCA model on test datasets*
3. *Compare AAKR and PCA results*

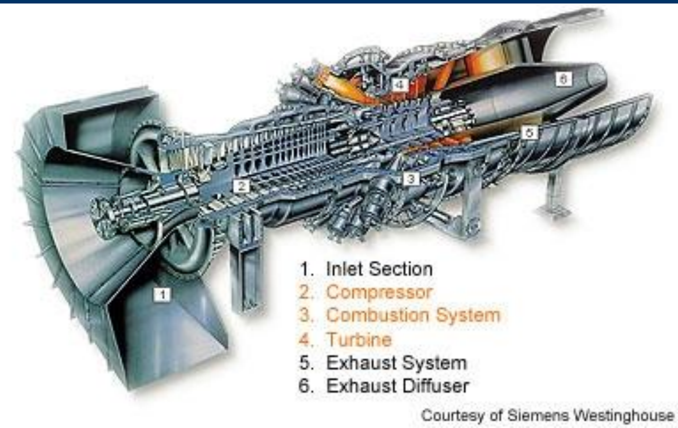
Exercise 2

Component: Gas Turbine



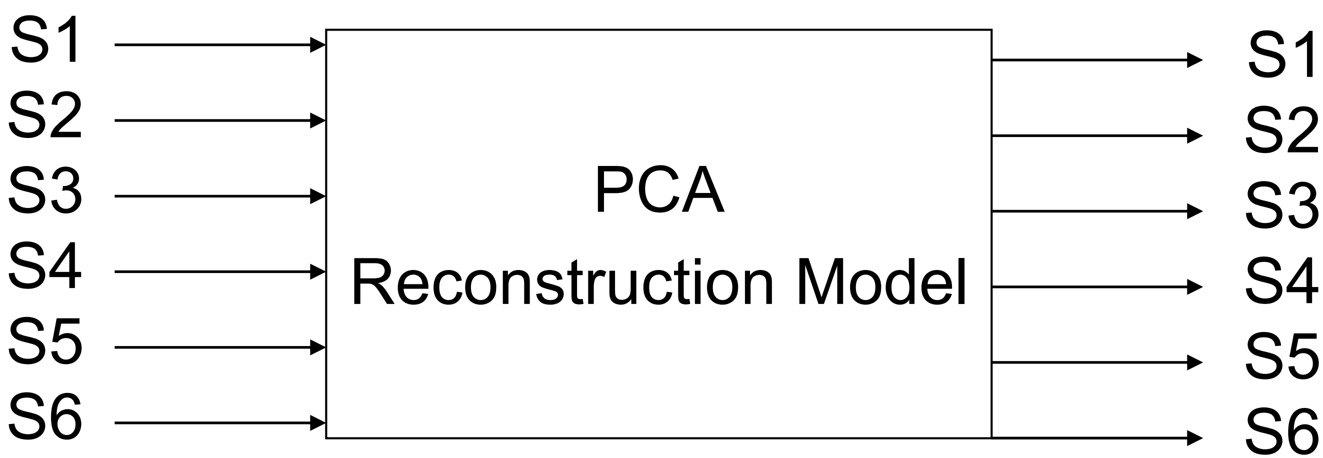
Component



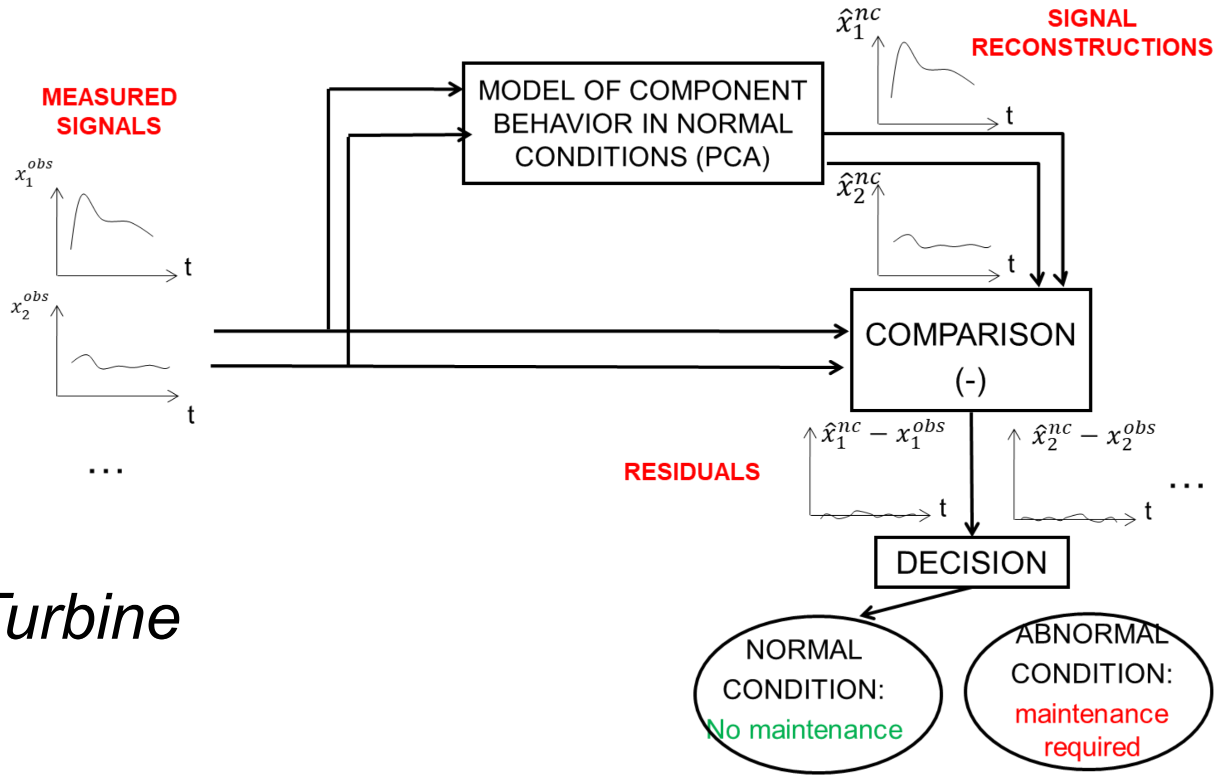


Signal Reconstructions = Expected signal values in normal conditions

Real measurements



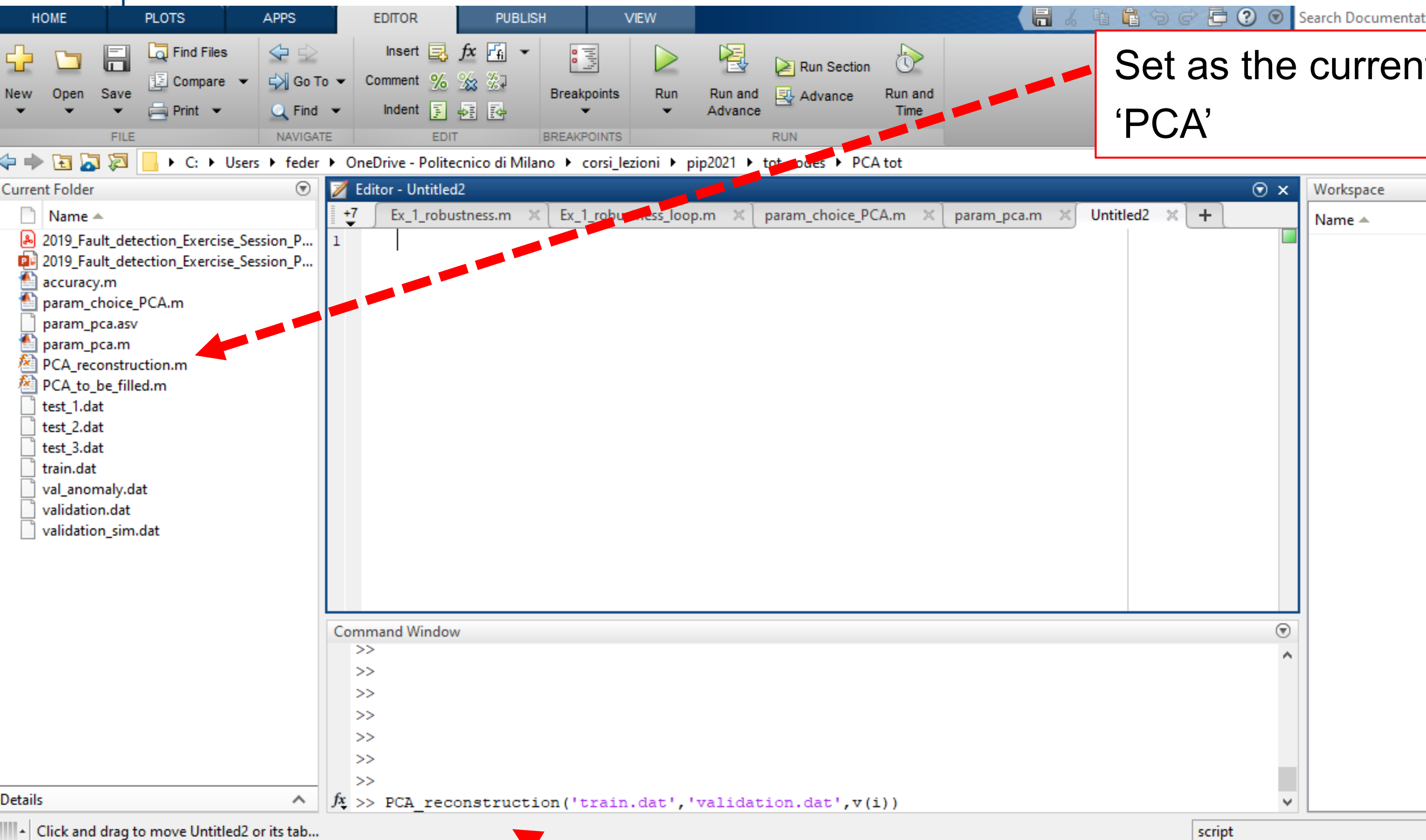
Exercise 2



Method: PCA

Component: Gas Turbine

The PCA code in Matlab: how to run the code?



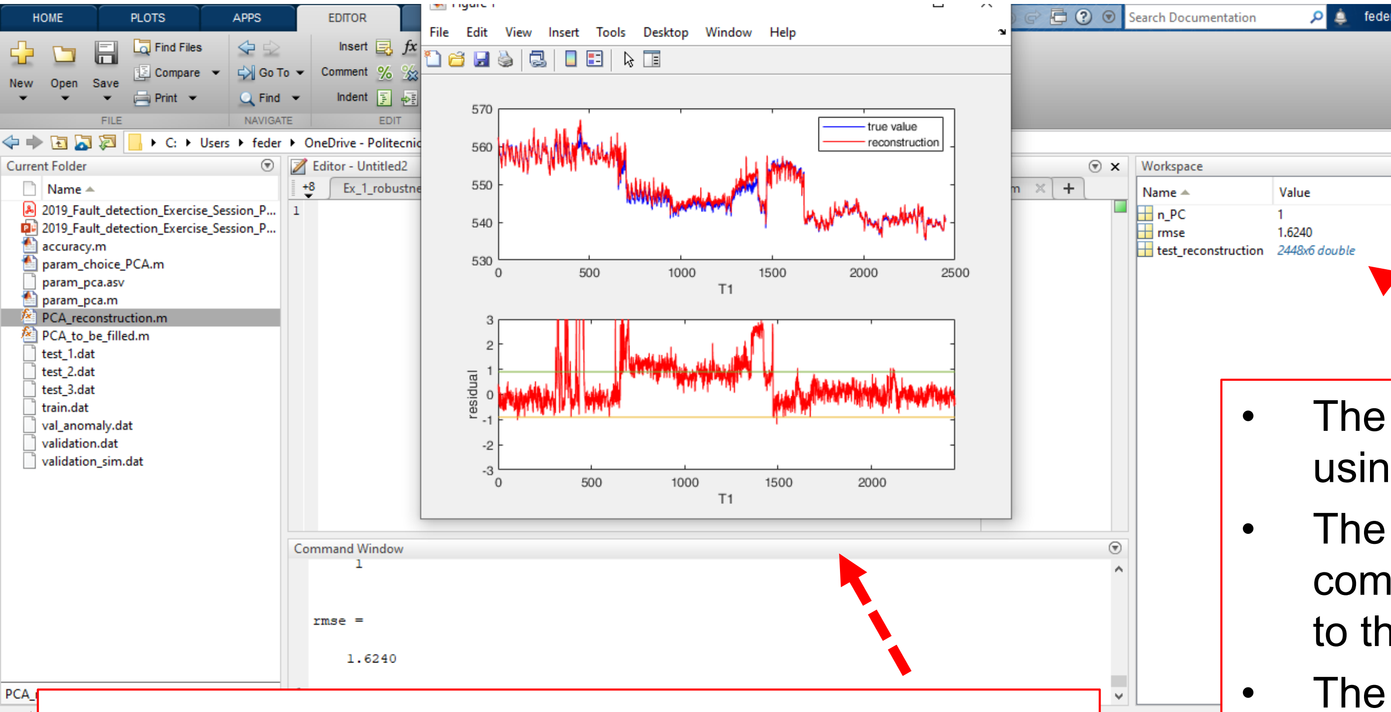
Set as the current folder the folder: 'PCA'

Execute the code in the command window:

[test_data_rec, RMSE]= $PCA\_reconstruction('\text{train.dat}', '\text{validation.dat}', \text{v})$

v=minimum percentage of variance to be represented in the PCA space [e.g., 0.95]

The PCA code in Matlab: Output



- The reconstruction obtained using the PCA algorithm
- The number of principal components corresponding to the selected variance
- The RMSE

One figure for each signal representing:

- The original data and the reconstruction obtained using the PCA algorithm
- The residual for each data point

Exercise 2: Assignment

Consider the same data used in Exercise 1

1. Optimize the PCA reconstruction model to achieve:
    a. Accuracy
    b. Robustness   Validation_sim.m→data containing a simulated abnormal condition on **signal 6**

2. Apply the PCA model on test datasets: test_1.dat, test_2.dat, test_3.dat and test_4.dat.

3. Compare AAKR and PCA results on test_3.dat

Hints:

- Consider different numbers of Principal Components by trying different values for the parameter *v=minimal percentage of variance considered in the PCA space*
- Consider the root mean square error as a performance measure and compute it on test data under normal condition and simulated abnormal condition

[test_reconstruction,n_PC, RMSE]=PCA_reconstruction('train.dat','validation.dat',0.95)

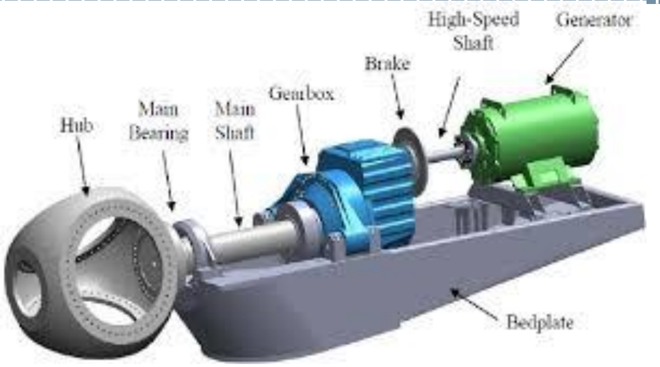
Exercise 3 (take home)

Method: you choose

Component: Wind Turbine

The Monitoring System

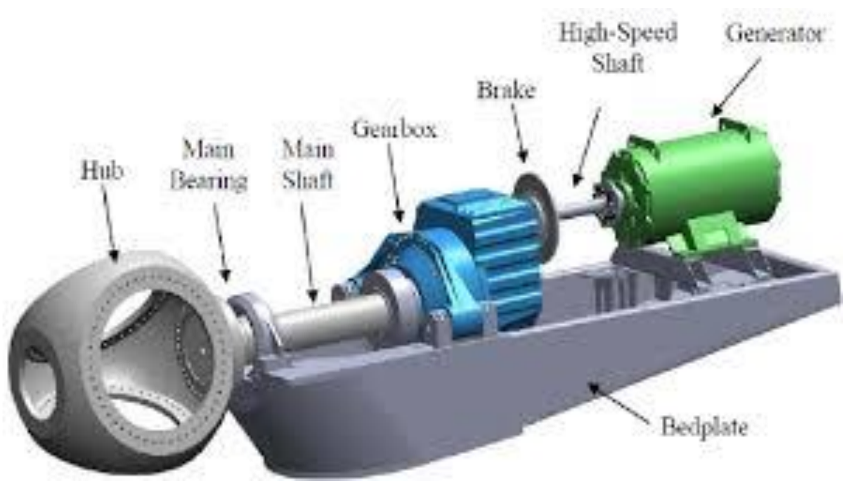
Wind turbine farm



- Main Bearing+ Planetary Gear box + Gearbox + Generator
- Monitoring system: 6 accelerometers and 1 sensor measuring the rotating speed.

Some possible failures modes

a) Crashing of the highspeed shaft of the gear box
b) Breaking of a teeth in the planetary stage
c) Misalignment between the generator and the gearbox shafts
d) Wearing of the rear bearing housing of the generator

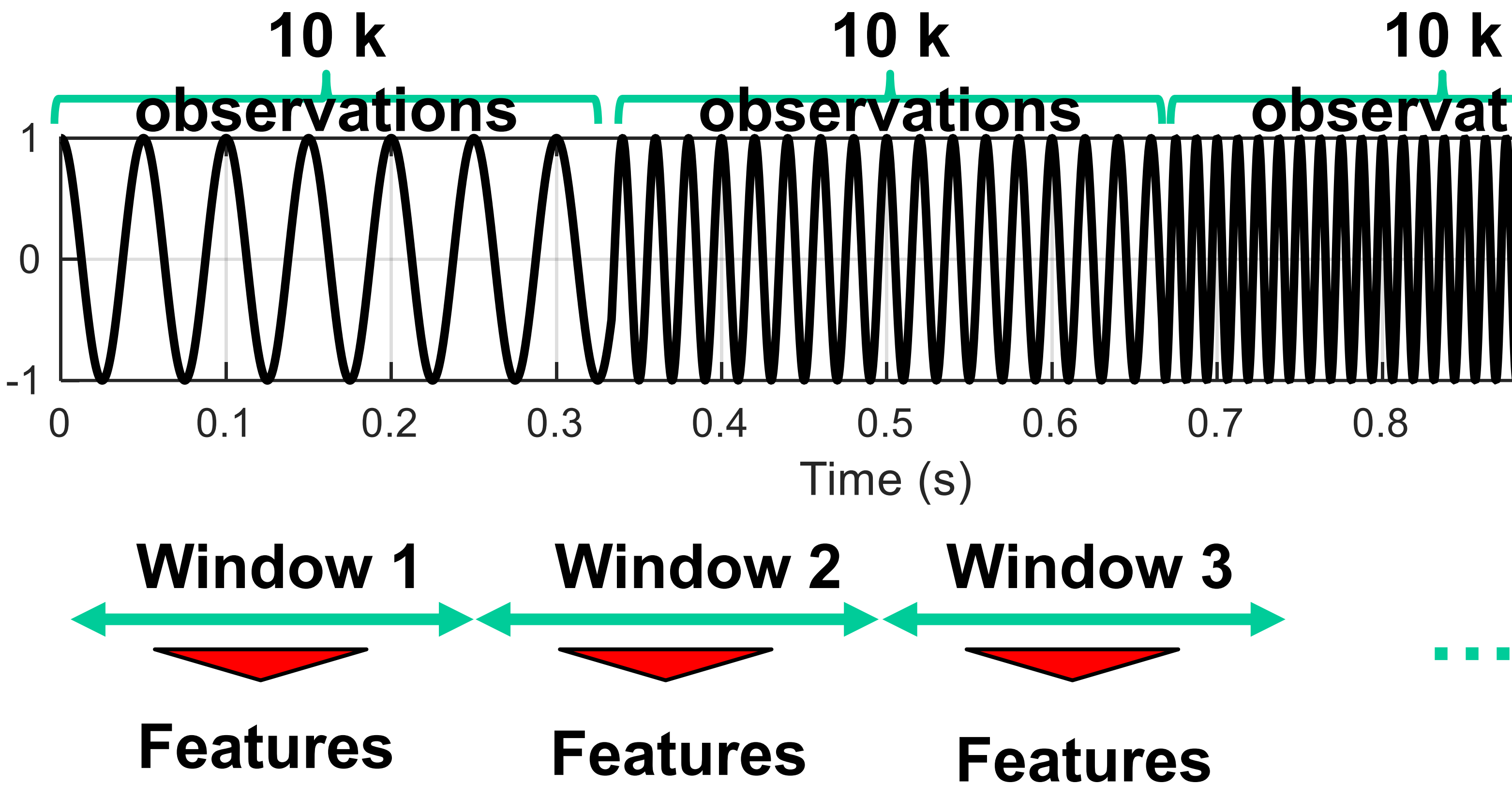


Gearbox-Generator Failure

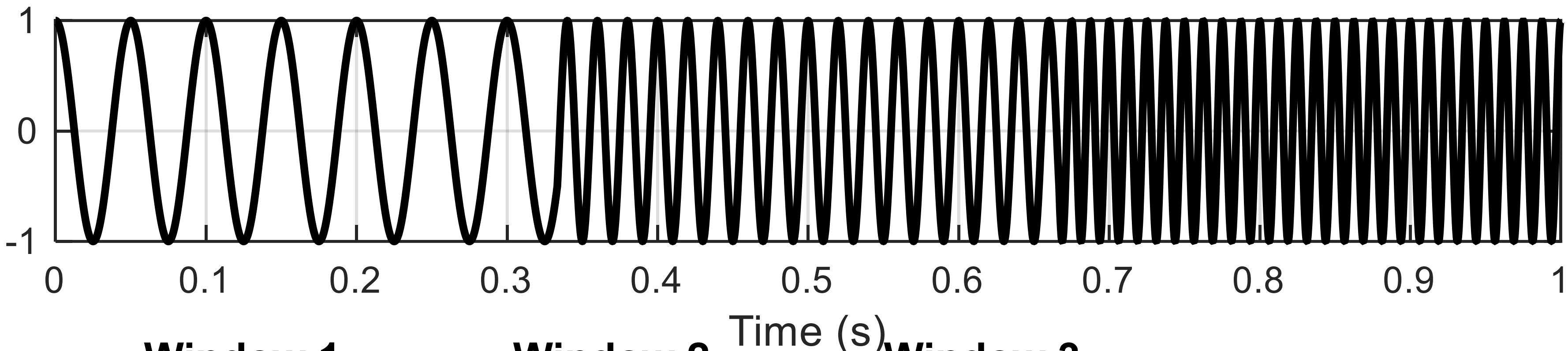
↓

- **Long turbine downtime [up to some months]**
- **Intervention of expensive tools**

High frequency sensor:



Feature Extraction

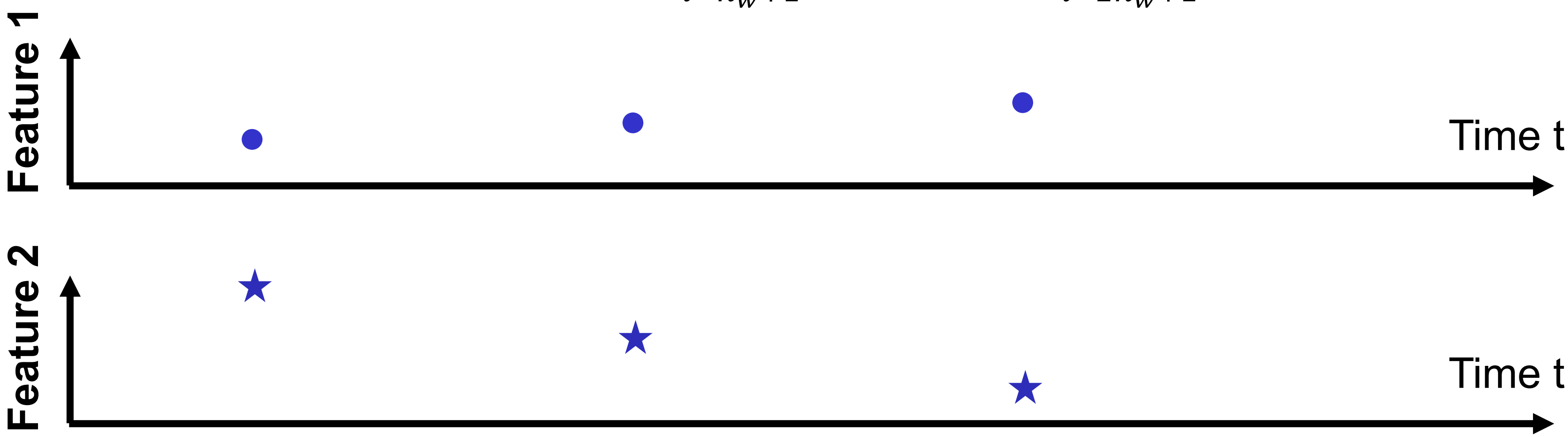


Window 1 — $W_1 = \{x_i\}_{i=1}^{n_w}$

Window 2 — $W_2 = \{x_i\}_{i=n_w+1}^{2n_w}$

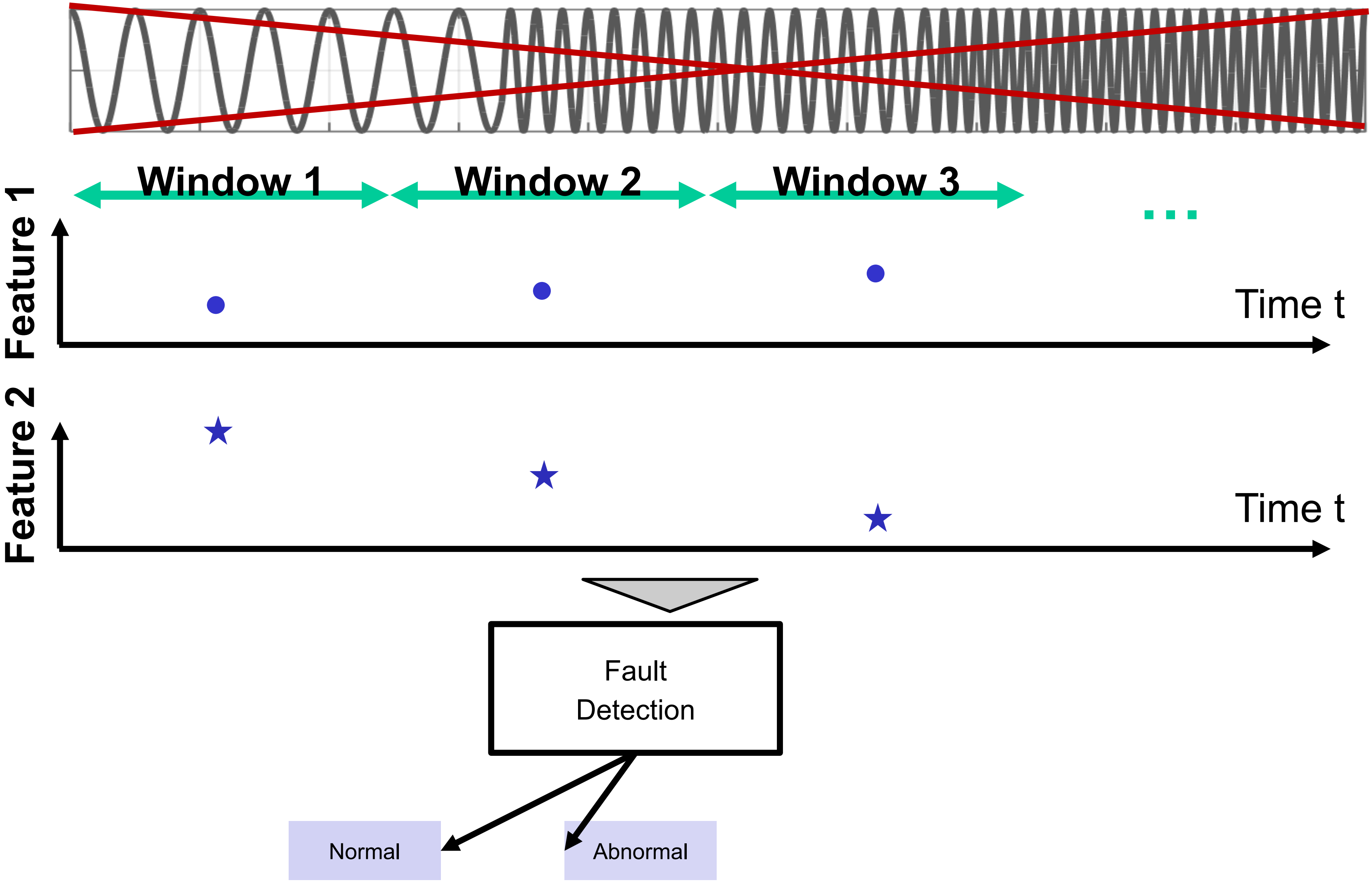
Window 3 — $W_3 = \{x_i\}_{i=2n_w+1}^{3n_w}$

...



...

Feature Extraction



FEATURE EXTRACTION

- Each 5 sec signal data are divided in 40 segments(windows) in order to extract features

- The following statistical features are considered:

Mean, standard deviation, Kurtosis, Skewnesss, min, max, 2nd moment, 3rd moment

- A window contains N values W $= \{x_i\}_{i=1}^{N}$, the corresponding Statistical Features are:

Sample Mean:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Sample Variance:

$$\sigma^2 \approx s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Standard Deviation: σ

k-th Moments:

$$\mu_k \approx \frac{1}{N}\sum_{i=1}^{N} x_i^k$$

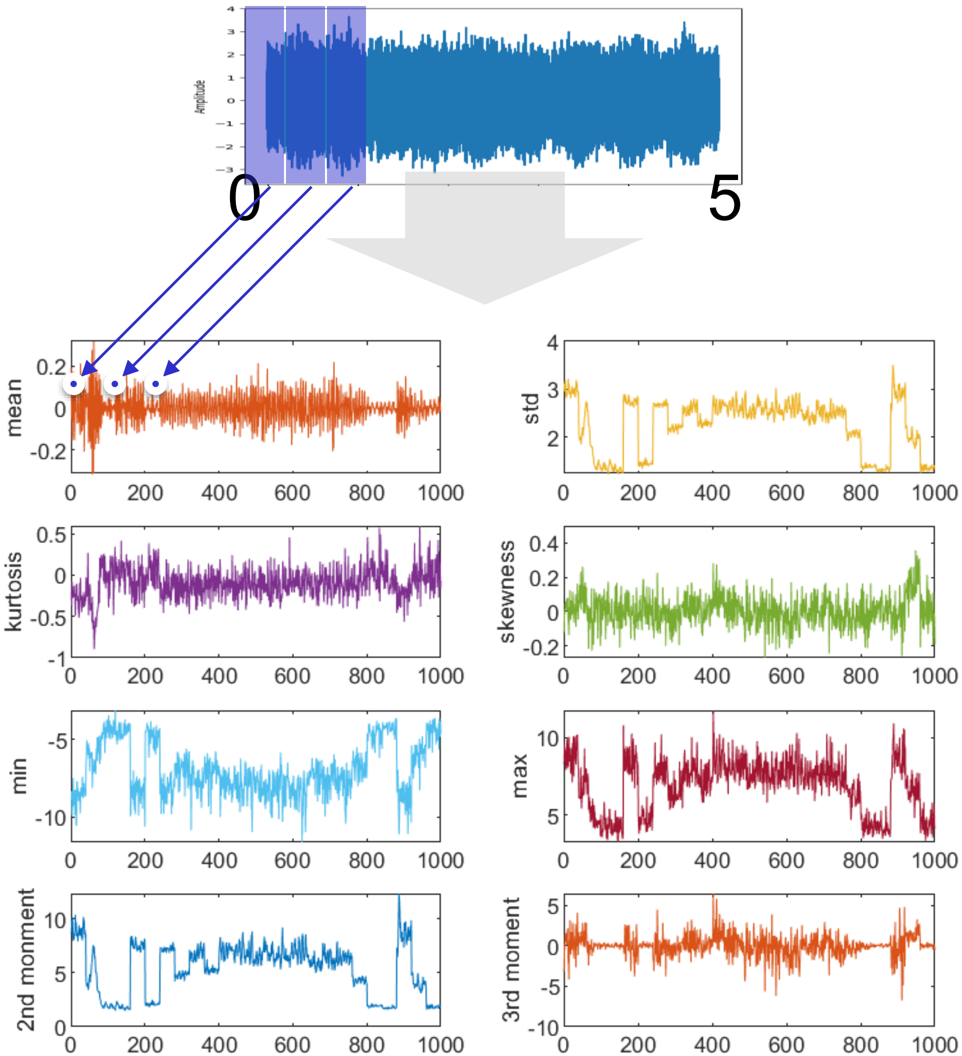
Skewness➔ “asymmetry”
(the third standardized moment)

$$\frac{\mu_3}{\sigma^3}$$

Kurtosis ➔ “peak”
(the fourth standardized moment)

$$\frac{\mu_4}{\sigma^4}$$

FEATURE EXTRACTION



Exercise 3: Assignment

<u>Given</u>:

1) The file <u>'train.dat' which </u>contains the features already extracted in normal condition
2) The file <u>'validation.dat' which </u>contains the features already extracted in normal condition
3) The file <u>'validation_sim.dat' which </u>contains the features in a simulated abnormal condition

train

800x8 double

	1	2	3	4	5	6	7	8
1	0.1918	2.9473	-0.2172	-0.1264	-8.9417	8.3333	8.6760	-3.2251
2	-0.0553	3.1255	-0.3734	0.0047	-8.3208	9.5974	9.7568	0.1444
3	-0.0273	3.0340	-0.0800	-0.0165	-9.9243	8.4782	9.1938	-0.4595
4	-0.1549	3.1874	-0.2684	0.0424	-8.4574	9.1952	10.1471	1.3688
5	0.0066	3.0410	-0.2734	0.0520	-9.3214	8.4448	9.2366	1.4564
6	0.1720	2.8165	-0.1769	0.0581	-7.4707	7.9281	7.9228	1.2936
7	-0.0533	2.9428	-0.1788	0.0175	-9.2303	9.3448	8.6492	0.4444
8	0.1164	2.8412	-0.1314	0.0722	-8.2652	9.2579	8.0627	1.6507
9	-0.0958	2.8827	-0.2671	0.0034	-8.6233	9.9347	8.2996	0.0815
10	-0.1537	2.9199	-0.2698	0.0793	-8.7042	8.0308	8.5154	1.9674
11	0.0615	2.9858	-0.2282	0.0806	-9.8658	9.6261	8.9039	2.1387
12	-0.0193	3.2212	-0.4878	-0.0667	-8.2399	8.1581	10.3633	-2.2198
13	0.0821	3.0265	-0.2508	-0.0710	-8.5418	9.9236	9.1485	-1.9599
14	0.0416	3.1149	-0.1363	-0.1381	-10.4691	10.0679	9.6909	-4.1581
15	0.0044	2.9303	-0.1060	0.0274	-10.5794	8.6044	8.5764	0.6874
16	-0.0010	2.8905	-0.1156	0.0113	-7.7779	8.4754	8.3445	0.2727
17	-0.0743	2.9348	-0.3117	0.1239	-7.6838	9.3669	8.6025	3.1201
18	-0.0671	2.9080	-0.2487	-0.0200	-8.6593	8.2900	8.4459	-0.4903
19	0.0190	2.8127	-0.2573	-0.0495	-8.1409	7.6882	7.9017	-1.0966
20	0.0753	2.9773	-0.3975	0.0638	-8.9418	8.6402	8.8533	1.6776

mean	std	kurtosiss	skewness	min	max	2^{nd} Moment	3^{rd} Moment

<u>You are required to:</u>

1) **<u>Develop</u>** a fault detection tool for the turbine generator [you can use AAKR or PCA, as you prefer];
2) **<u>Apply the developed tool </u>**to the data in the file "`test.mat`" and Identify possible abnormal conditions period.

Exercise 3.1

Fault detection with AAKR on new data

1. *Optimize the AAKR to achieve Accuracy and Robustness*
2. *Apply the AAKR model on test data*

Use data in folder ‘second case study’

Exercise 3.2

Fault detection with PCA

1. *Optimize the PCA to achieve Accuracy and Robustness*
2. *Apply the PCA model on test data*

Use data in folder ‘second case study’