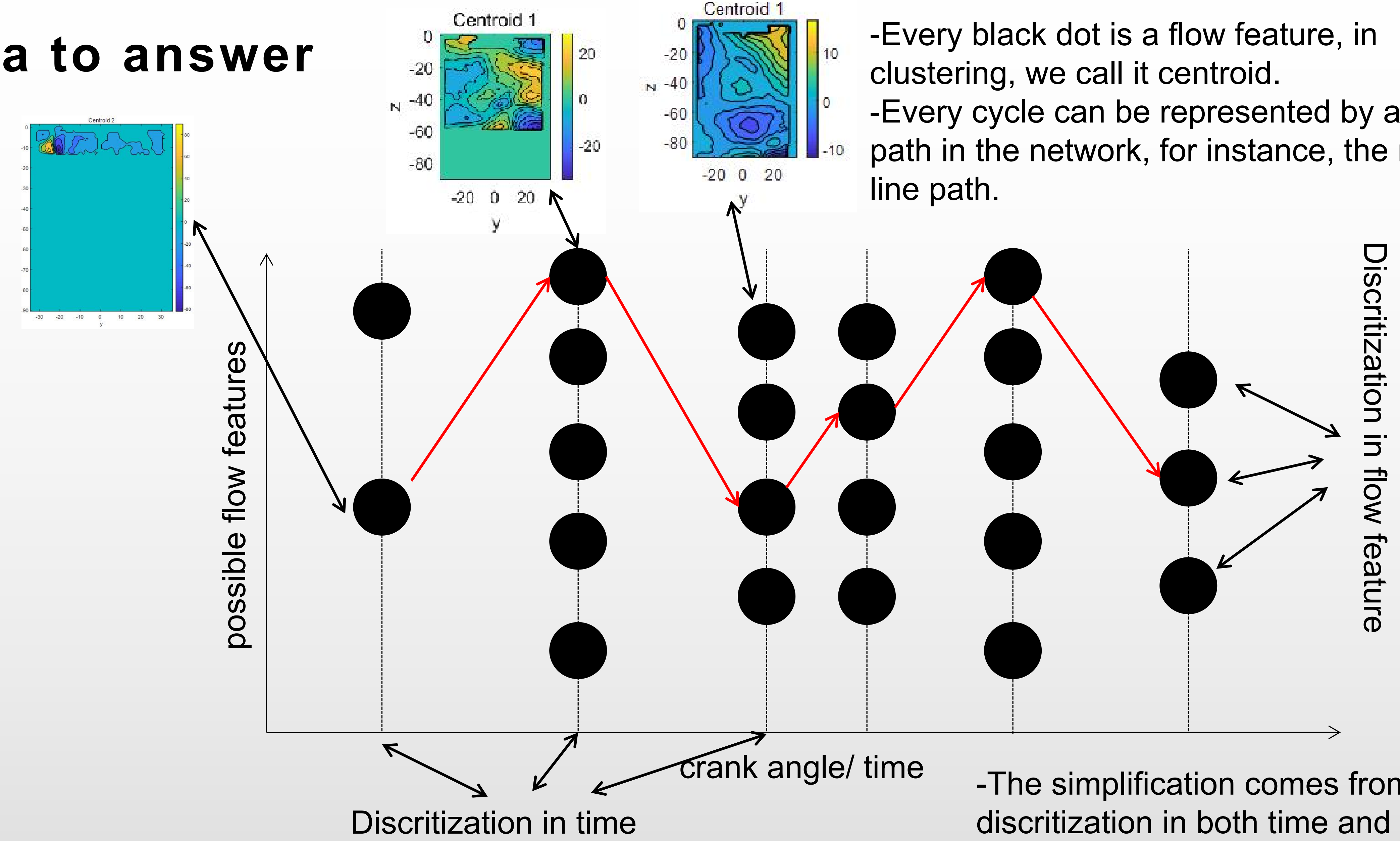
Cycle to Cycle Clustering preliminary report

Shuheng Liu

Motivation

- In TCCIII engine data set, hundreds of cycles are measured, including flow feature evolution and pressure signal.
- Is there a way to sort out flow features into groups so that we can simplify the description?
- How can we find out the most and the least deviated cycle? Can deviation of cycles be quantified?
- Can flow feature be predicted by just monitoring pressure signal?

Idea to answer

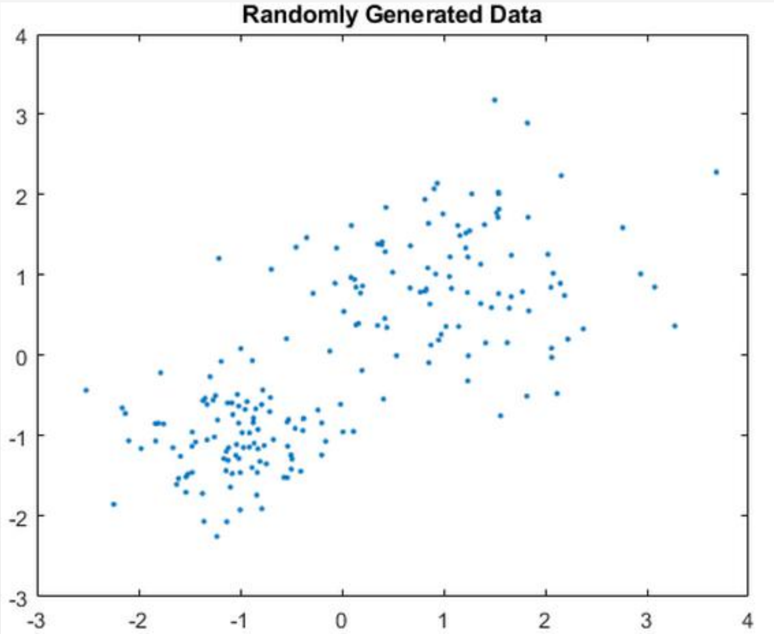


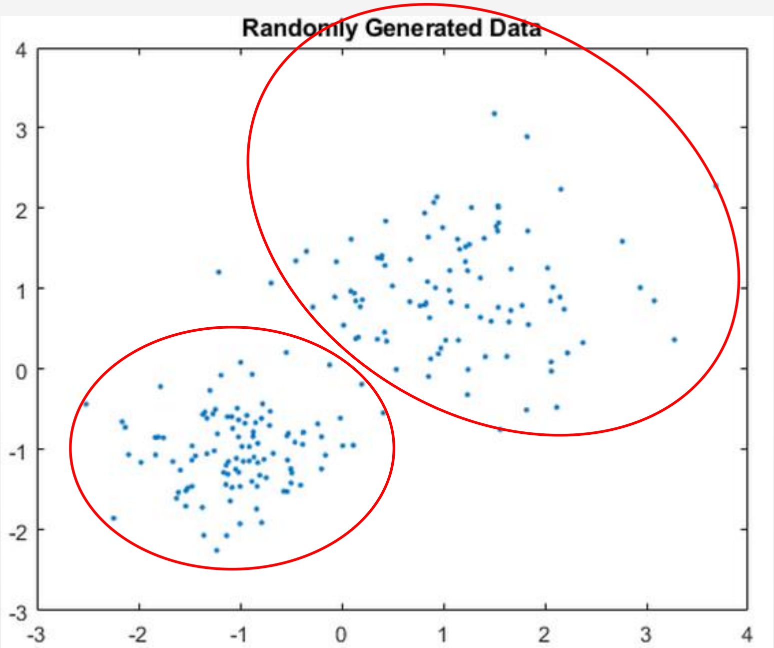
-Every black dot is a flow feature, in clustering, we call it centroid.
-Every cycle can be represented by a line path in the network, for instance, the red line path.

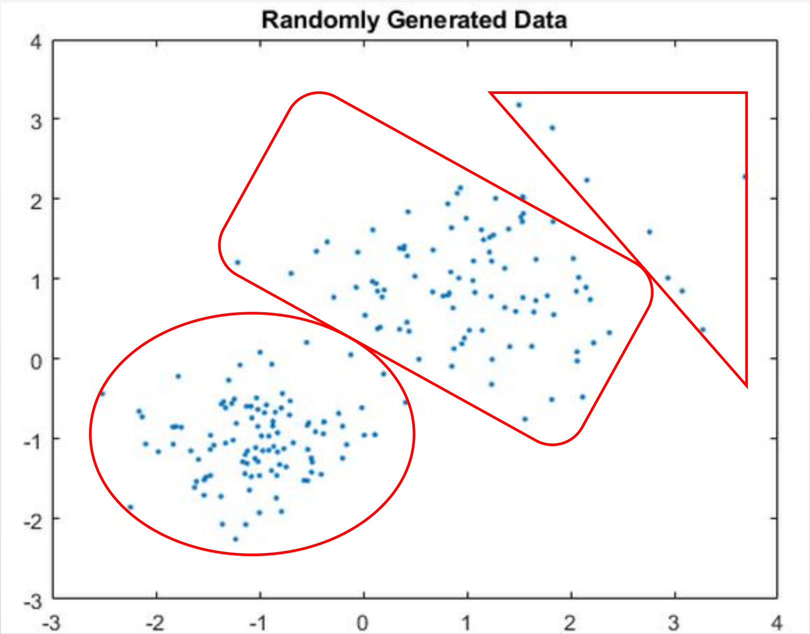
-The simplification comes from discritization in both time and flow feature

How to get the discritized flow features(black dots)

- Through unsupervised machine learning techniques: Clustering
- Widely used clustering techniques: kmeans, spectral clustering
- Function: Assign flow features to clusters and derive the centroid of each cluster.
- For example, in 2D space, there are a bunch of data points(Flow features are just data points in very high dimension), scattered in an xy coordinate system, these data can be put into 2 or 3 clusters. Number of clusters can be decided by users.

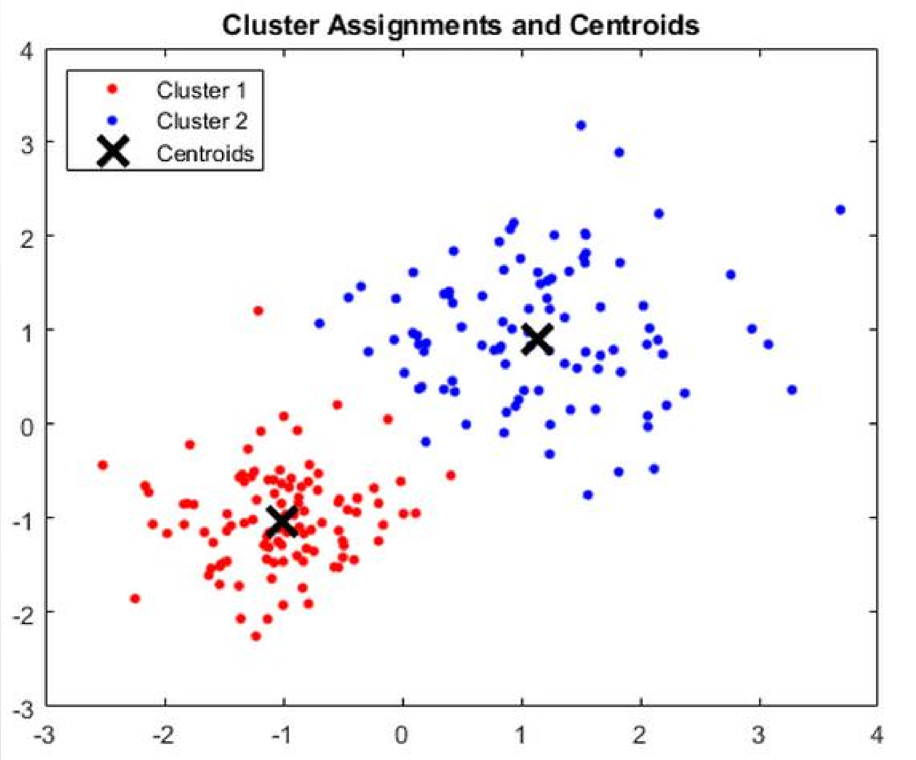






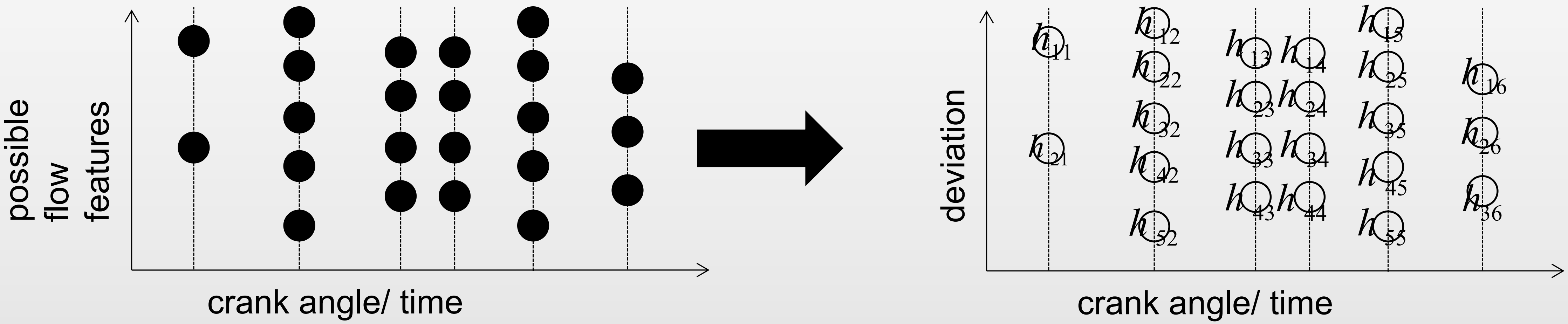
How to get the discritized flow features(black dots)

-After kmeans, resulting clusters and centroids are shown below:



How to quantify the deviation of centroids

- After clustering on all the flow features at interesting crank angles, there is a probability distribution for every interesting crank angle $\theta_1, \theta_2, \ldots, \theta_{N\theta}$, recorded by a probability distribution vector $\mathbf{y}_i = [y_{i1} \quad y_{i2} \quad \ldots \quad y_{iN_{ki}}], \mathrm{i} = 1, 2, \ldots, N_\theta$. For example, y_{i1} is number of flow features in cluster1, normalized by total number of flow features at *i* th interesting crank angle.
- To quantify how much centroid *j* deviates from mean of crank angle *i*, we compute distance h_{ji} between $\mathbf{y}_i$ and the exact distribution, for example $h_{2i} = \left\| [0 \quad 1 \quad 0 \quad \ldots \quad 0] - \mathbf{y}_i \right\|$.

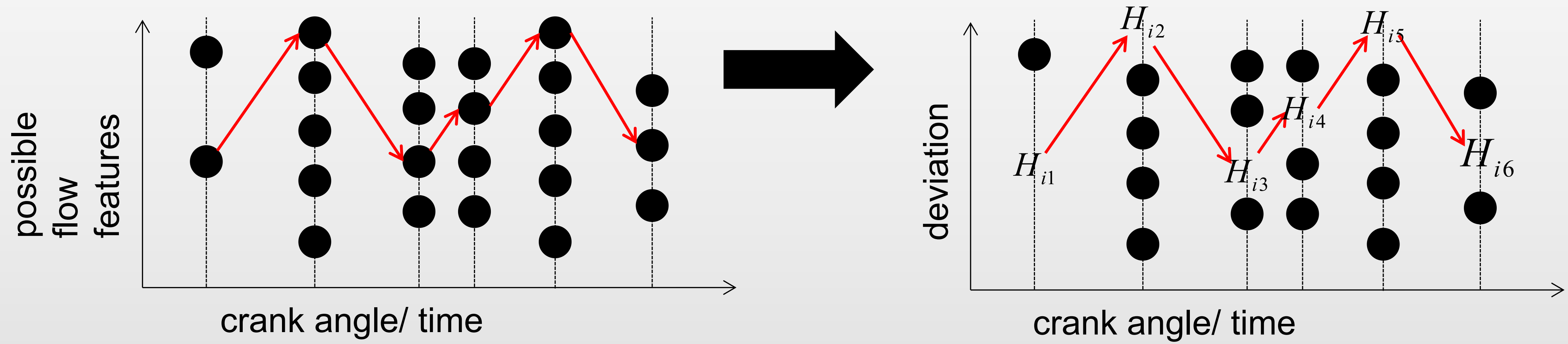


How to quantify deviation of cycles

-According to the line path of each cycle, a so-called distance evolving matrix **H** can be assembled:

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N_\theta} \\ H_{21} & \cdots & \cdots & H_{2N_\theta} \\ \cdots & \cdots & \cdots & \cdots \\ H_{N_{cyc}1} & H_{N_{cyc}2} & \cdots & H_{N_{cyc}N_\theta} \end{bmatrix}$$

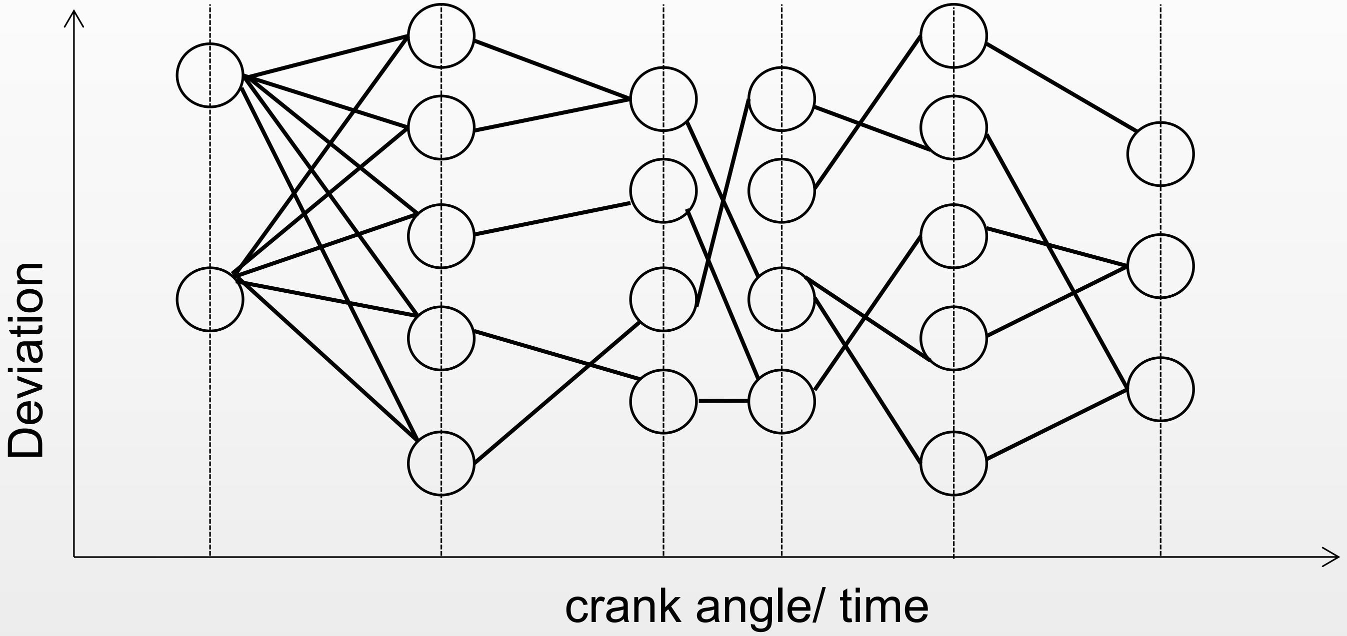
-The *i*th row of **H** contains the *h* sequence of the i th cycle.



-The significance of **H**: **H** contains the deviation information of all the cycles in a much reduced order

Possible ways to investigate H

- Just plot every or some of the row of **H**.
  - For example:



  - It may show the rough tendency of evolution, but may be too busy when number of cycles increase

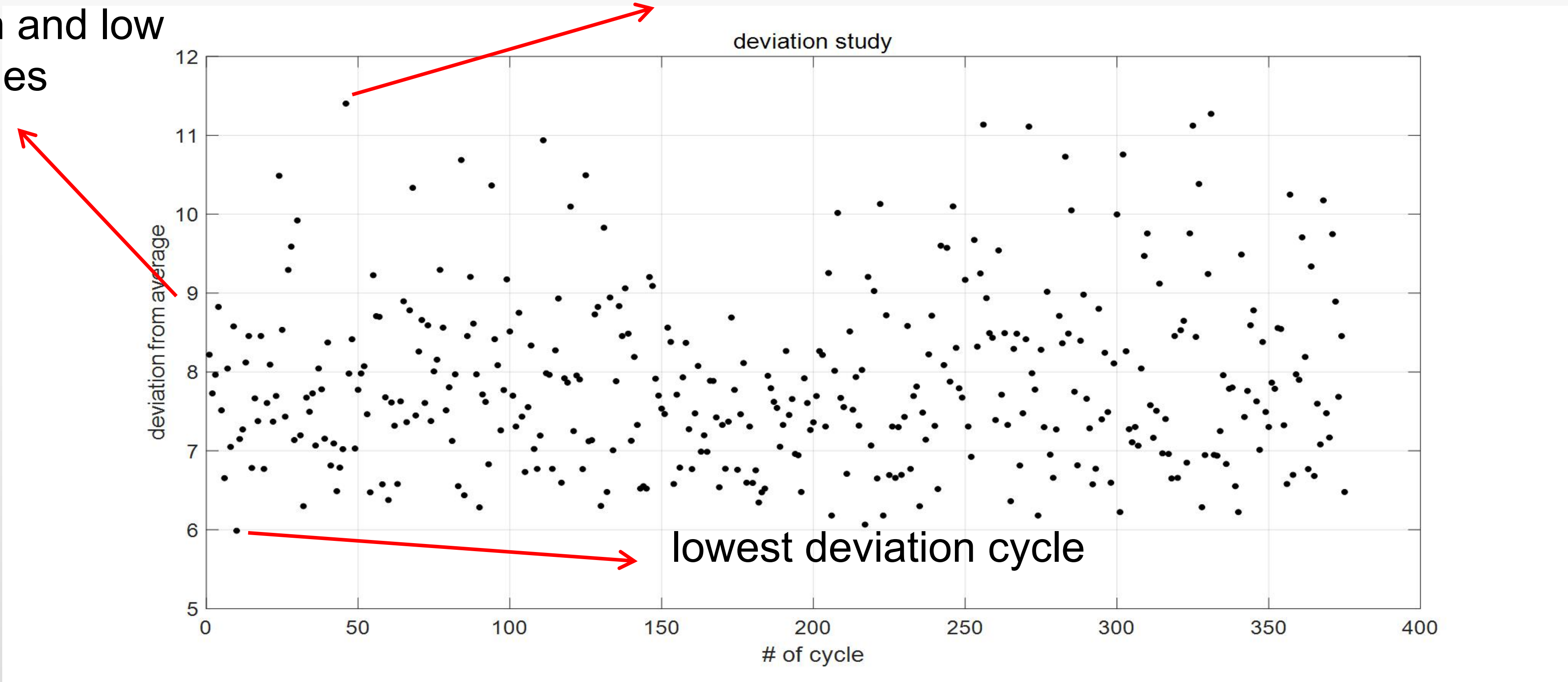
Possible ways to investigate H

-sum up every row of **H** to investigate the total deviation of every cycle. Associate every cycle with one value of deviation. The highest and lowest deviation cycles can be separated. High and low deviation cycles can be separated.

- An example is shown below:

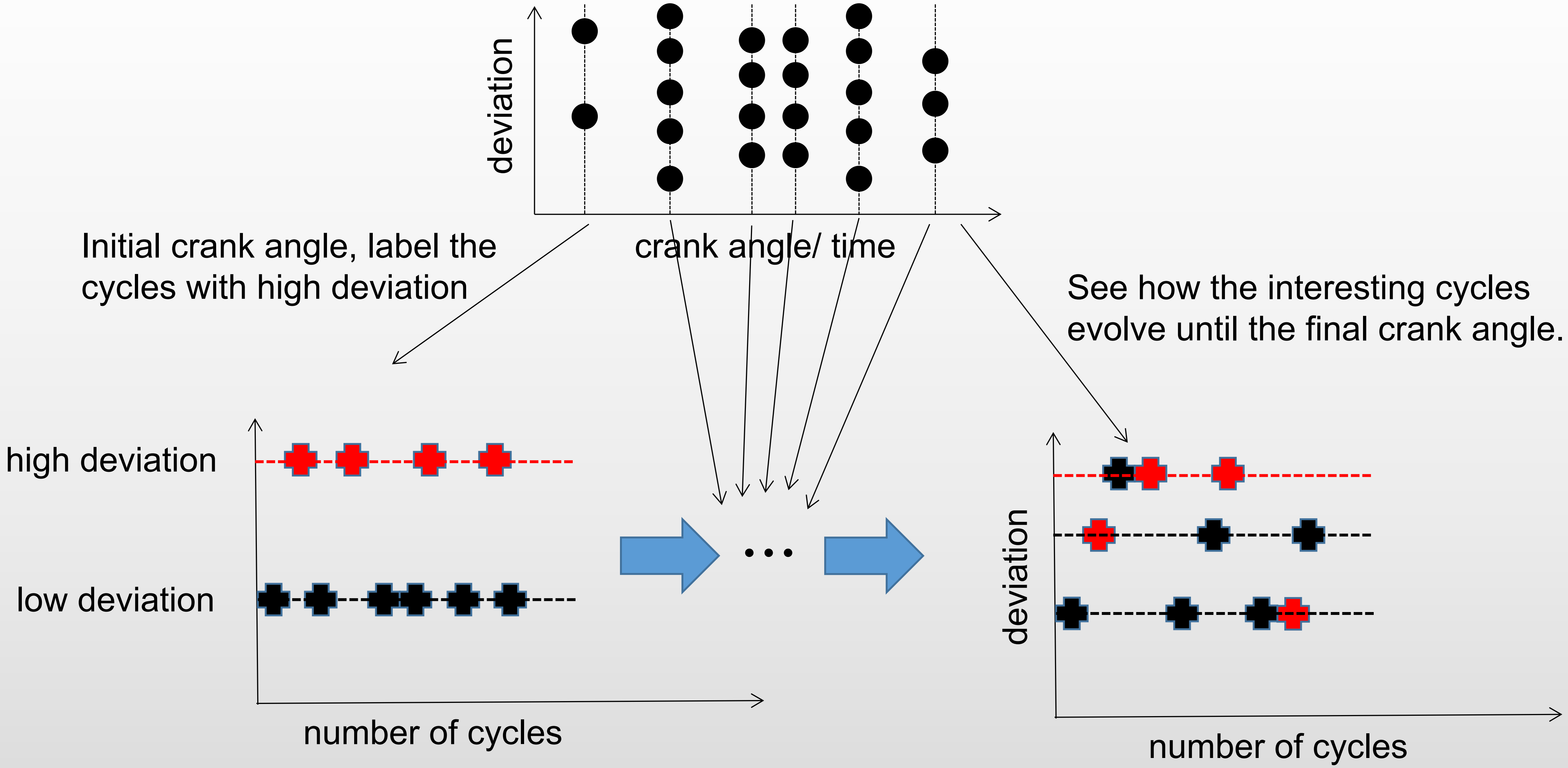
A self-decided boundary to separate high and low deviation cycles

highest deviation cycle



Possible ways to investigate H

- Track cycles with high deviation at the begining
- example:

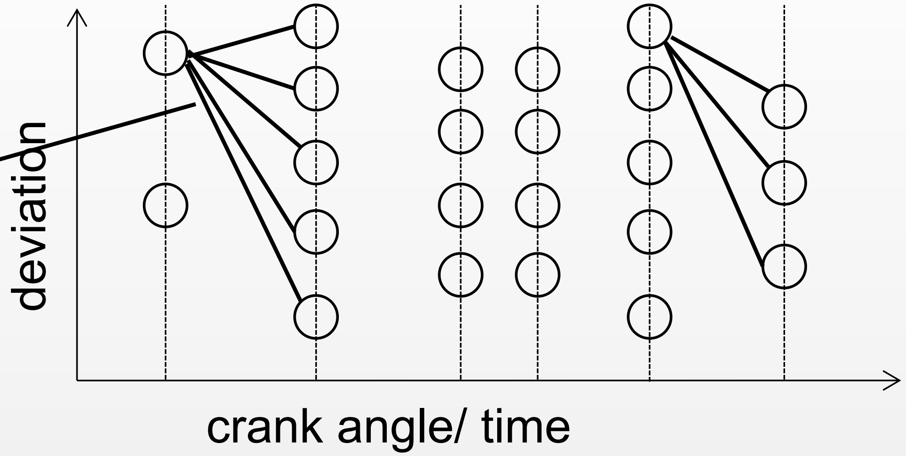


Possible ways to investigate H

- Compute transition probability

$$\mathbf{p}_1 = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet \end{bmatrix}$$

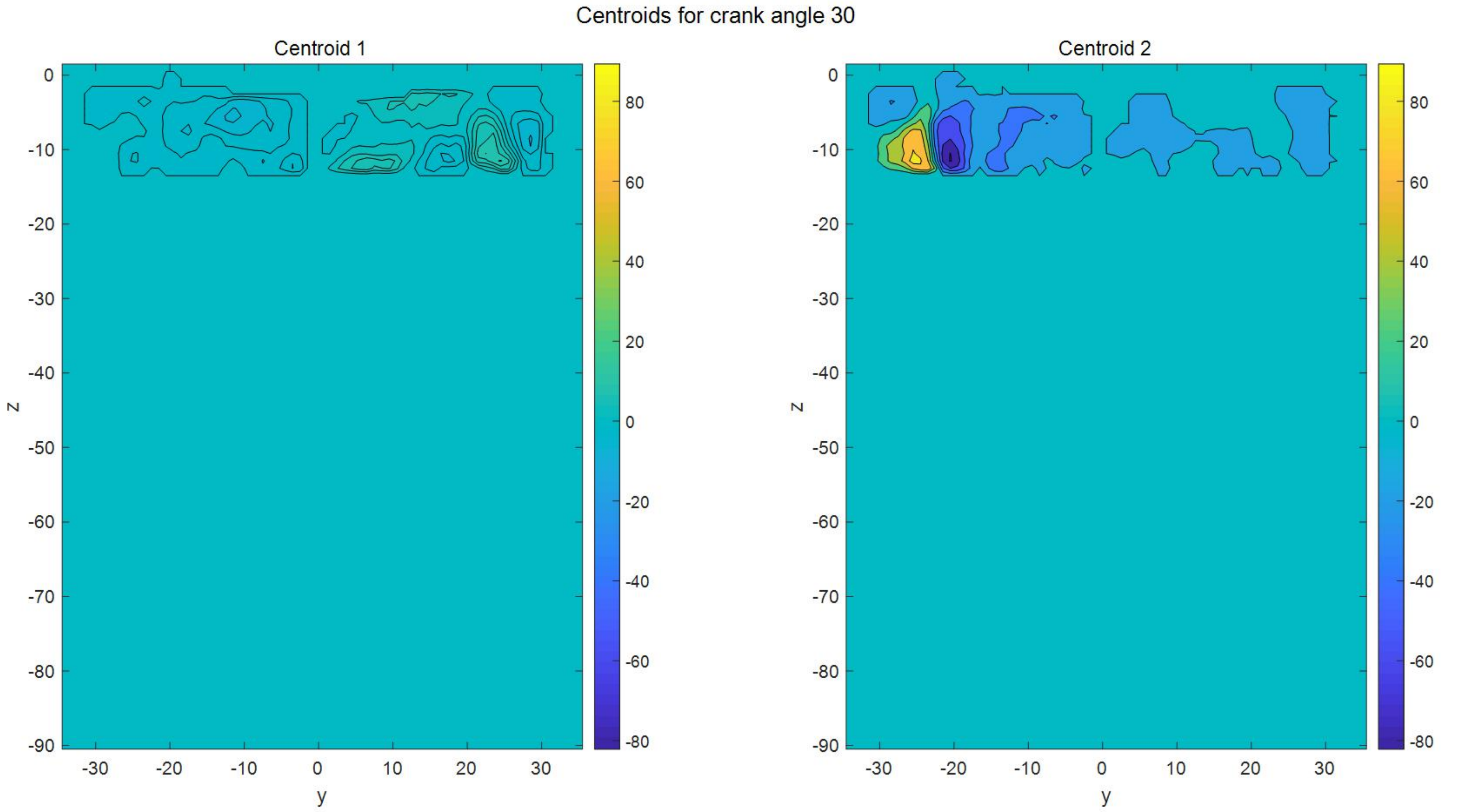
-Associate a transition probability distribution vector with every cluster except for the last crank angle



Implementation on real data

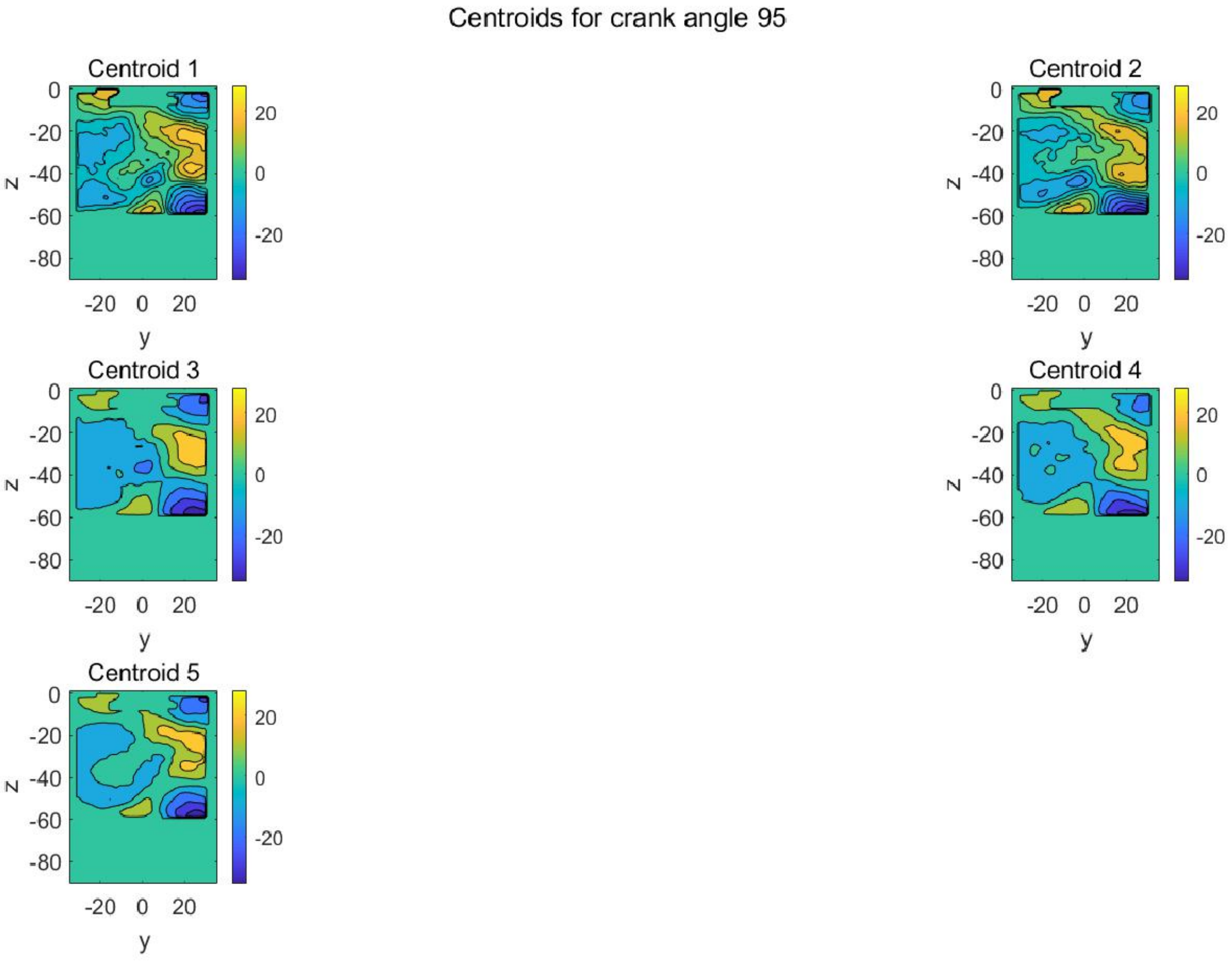
- Input data: TCCIII, Fired, Full View, image plane x=0, common grid id:S_2014_02_13_02
- Clustering algorithm: kmeans
- crank angles: $\mathbf{\Theta} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_{N_\theta=6} \end{bmatrix} = \begin{bmatrix} 30 & 95 & 165 & 230 & 310 & 360 \end{bmatrix}$
- Number of cluster for each crank angle: $\mathbf{N_k} = \begin{bmatrix} 2 & 5 & 5 & 5 & 5 & 5 \end{bmatrix}$

Clusters



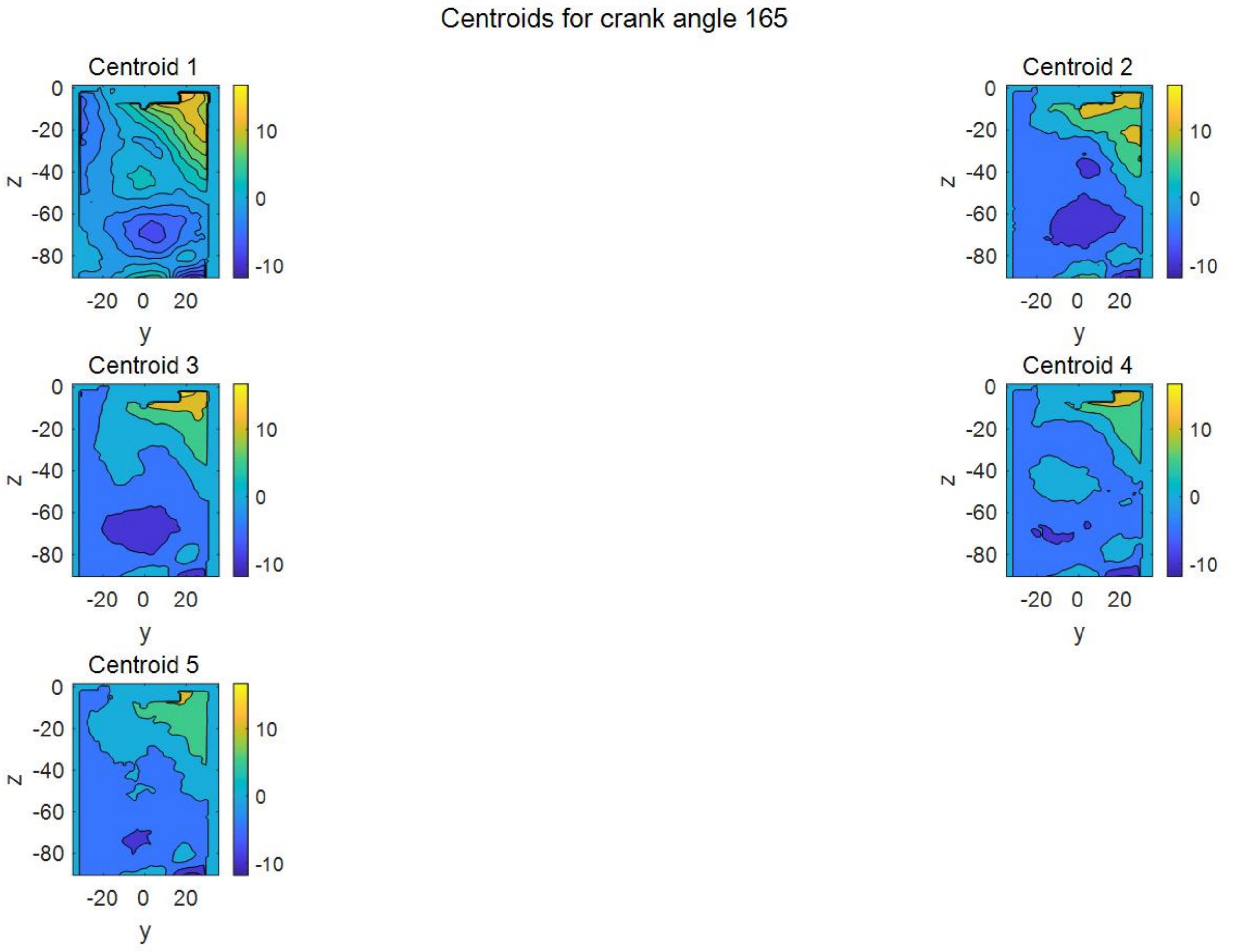
Centroids for crank angle 30
Centroid 1
Centroid 2
z
y
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-30
-20
-10
0
10
20
30
80
60
40
20
0
-20
-40
-60
-80

Clusters



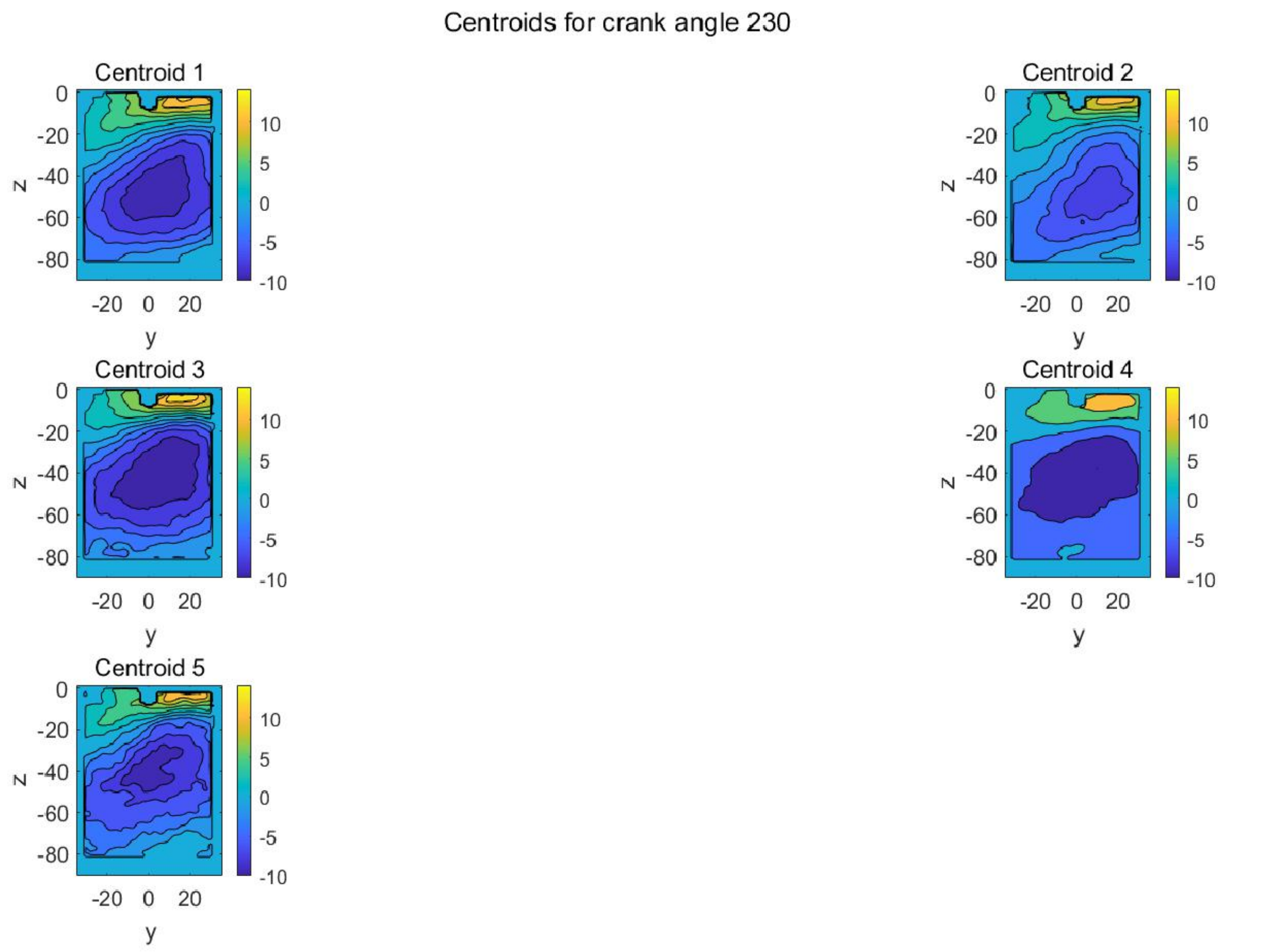
Centroids for crank angle 95
Centroid 1
Centroid 2
Centroid 3
Centroid 4
Centroid 5
z
y
0
-20
-40
-60
-80
-20
0
20

Clusters

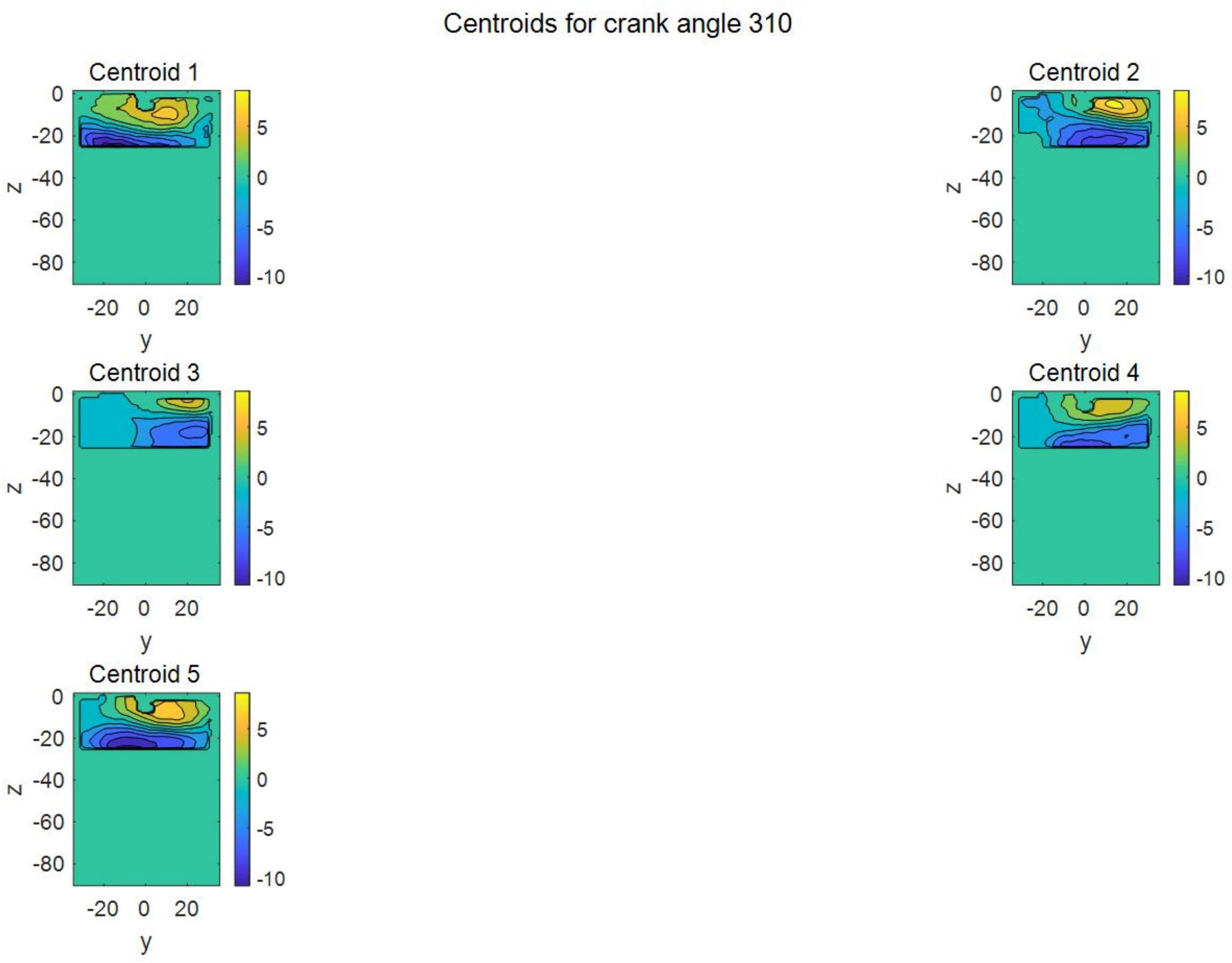


Centroids for crank angle 165
Centroid 1
Centroid 2
Centroid 3
Centroid 4
Centroid 5
z
y
0
-20
-40
-60
-80
-20 0 20
10
0
-10

Clusters

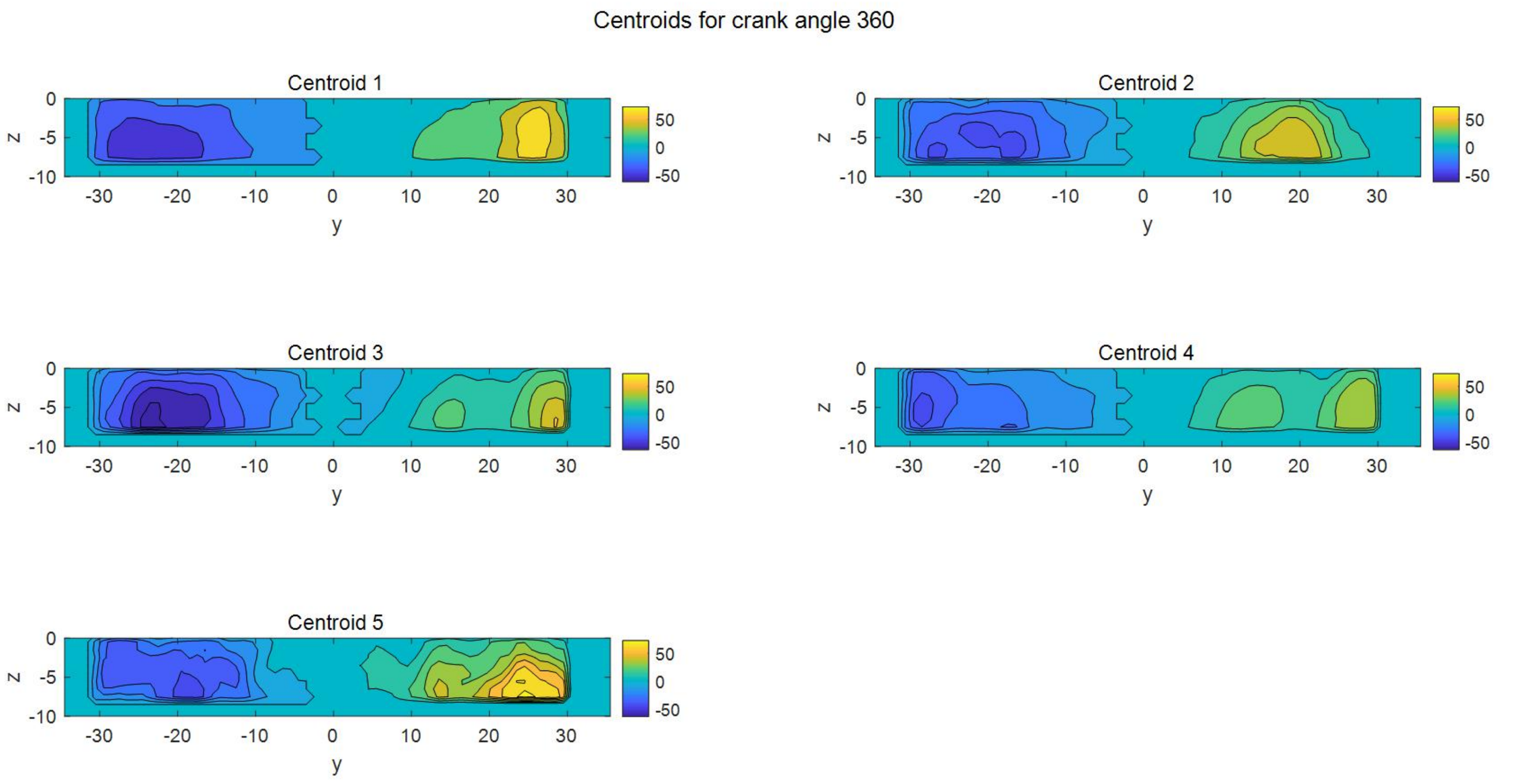

Centroids for crank angle 230
Centroid 1
Centroid 2
Centroid 3
Centroid 4
Centroid 5
z
y
0
-20
-40
-60
-80
-20 0 20
10
5
0
-5
-10

Clusters



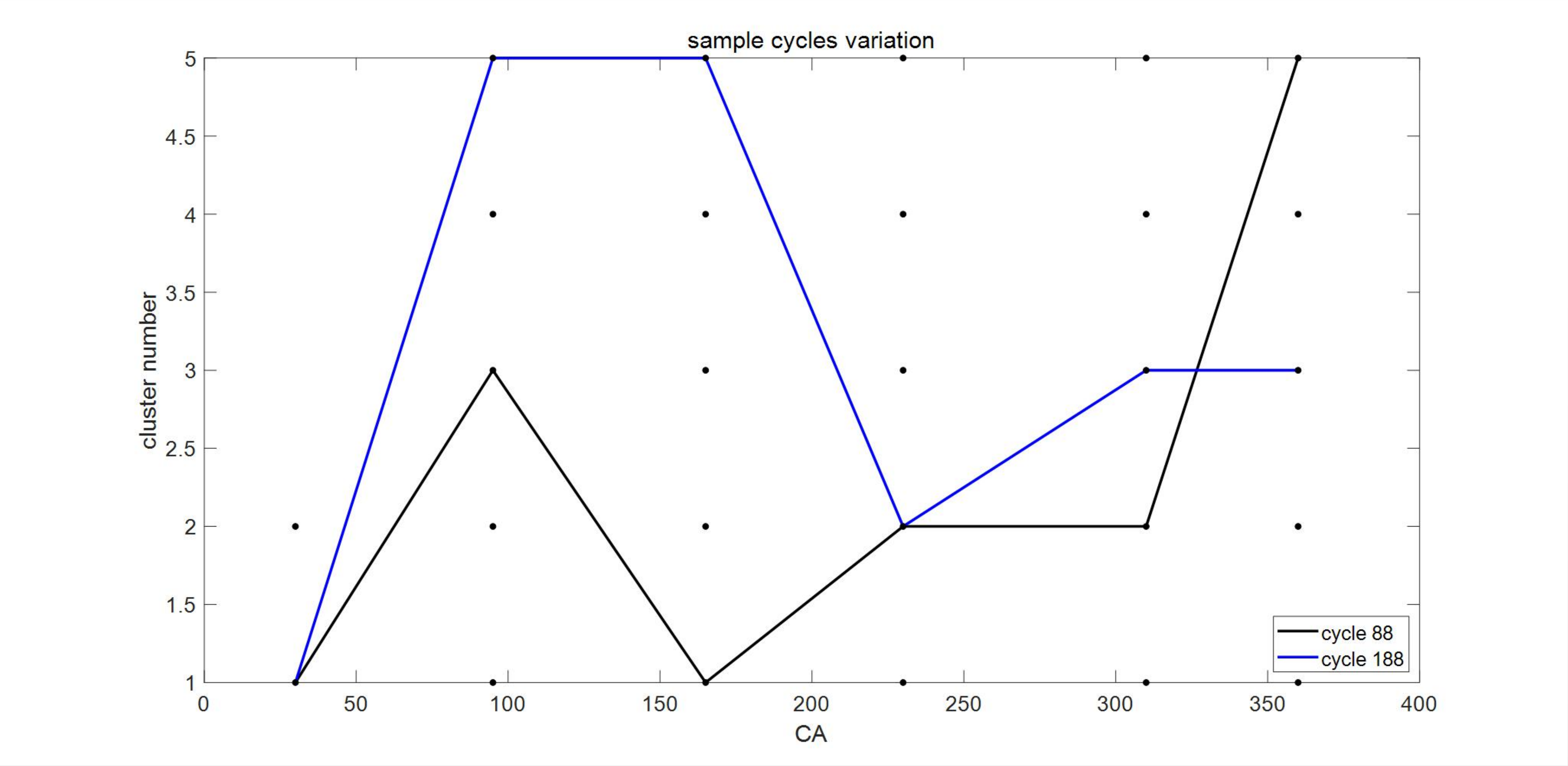
Centroids for crank angle 310
Centroid 1
Centroid 2
Centroid 3
Centroid 4
Centroid 5
z
y
0
-20
-40
-60
-80
-20
0
20
5
0
-5
-10

Clusters



Centroids for crank angle 360
Centroid 1
Centroid 2
Centroid 3
Centroid 4
Centroid 5
z
y
0
-5
-10
-30
-20
-10
0
10
20
30
50
0
-50

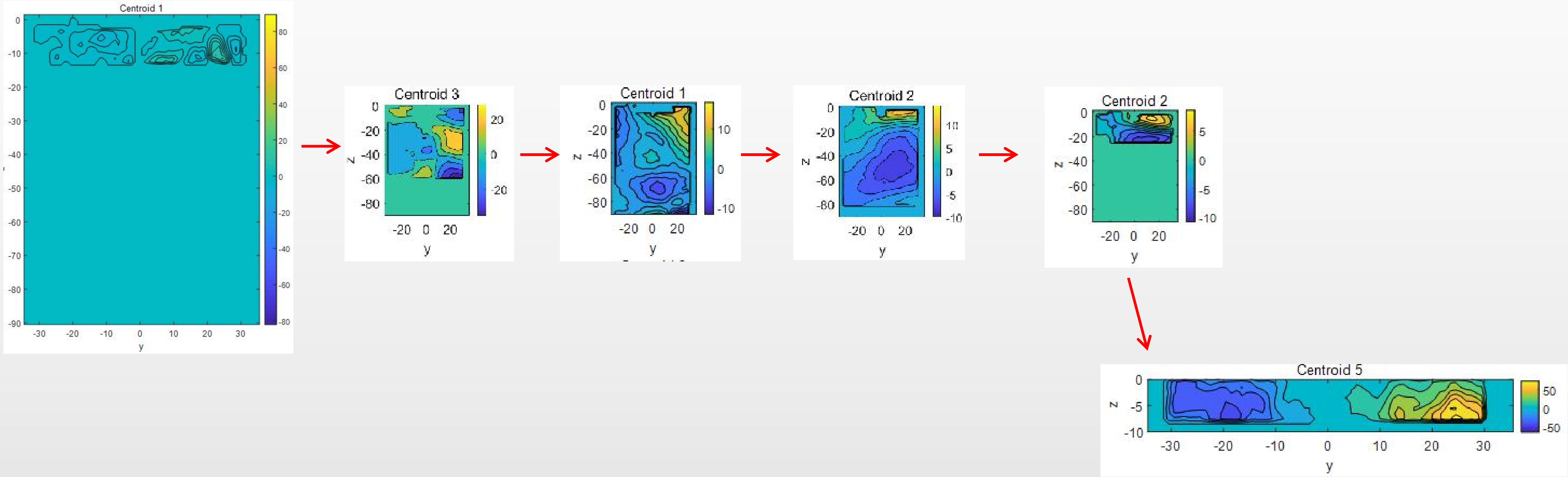
Diagrams



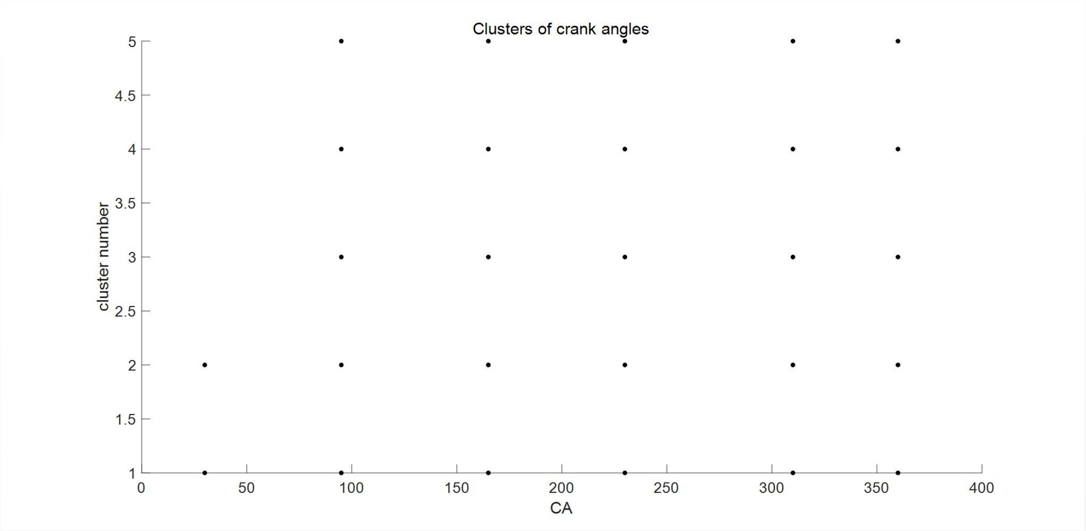
Cycle 88 and 188 are randomly chosen. All the cycles can be visualized by a line pattern.

Further interpretation

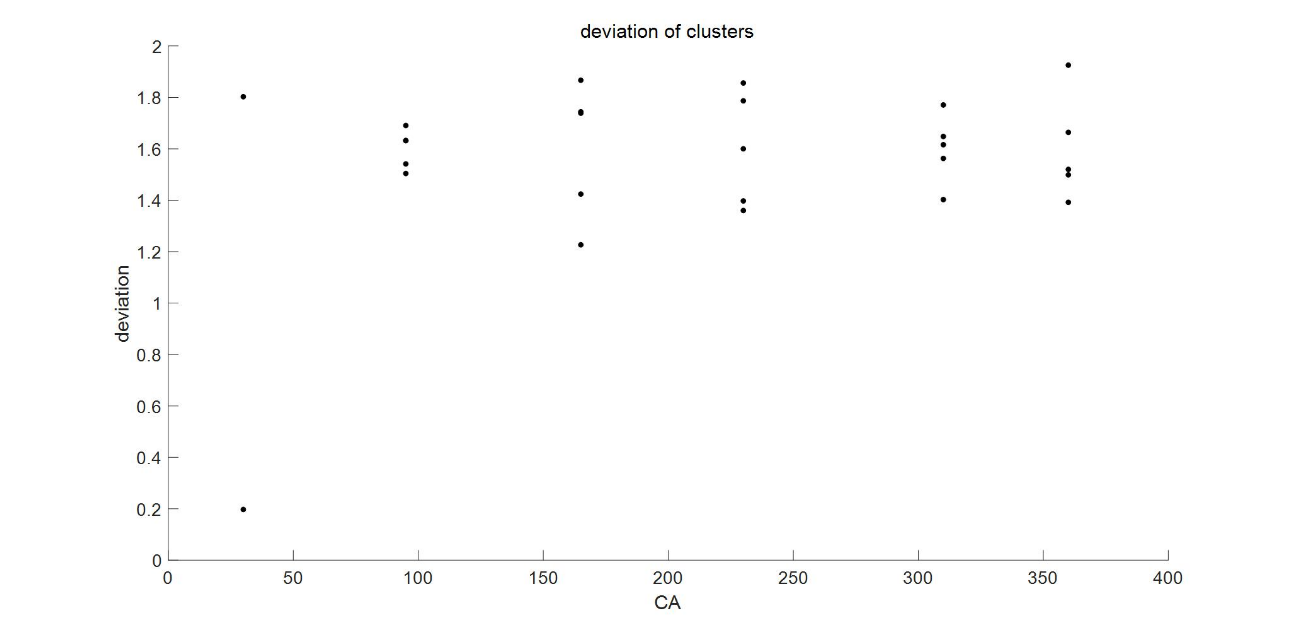
- We will know cycle 88, for example, go through clusters 1-3-1-2-2-5



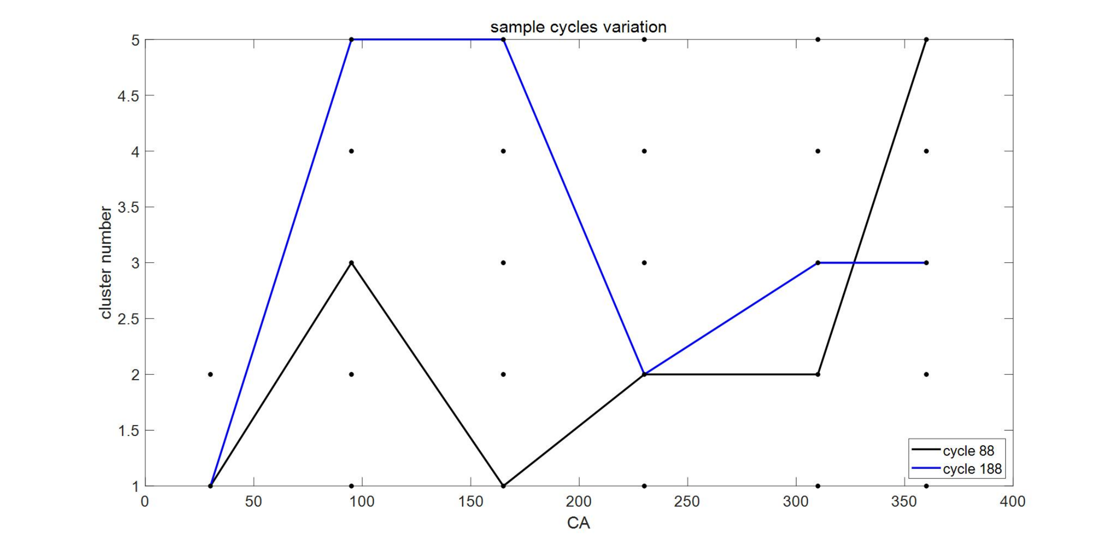
Quantify deviation of centroids

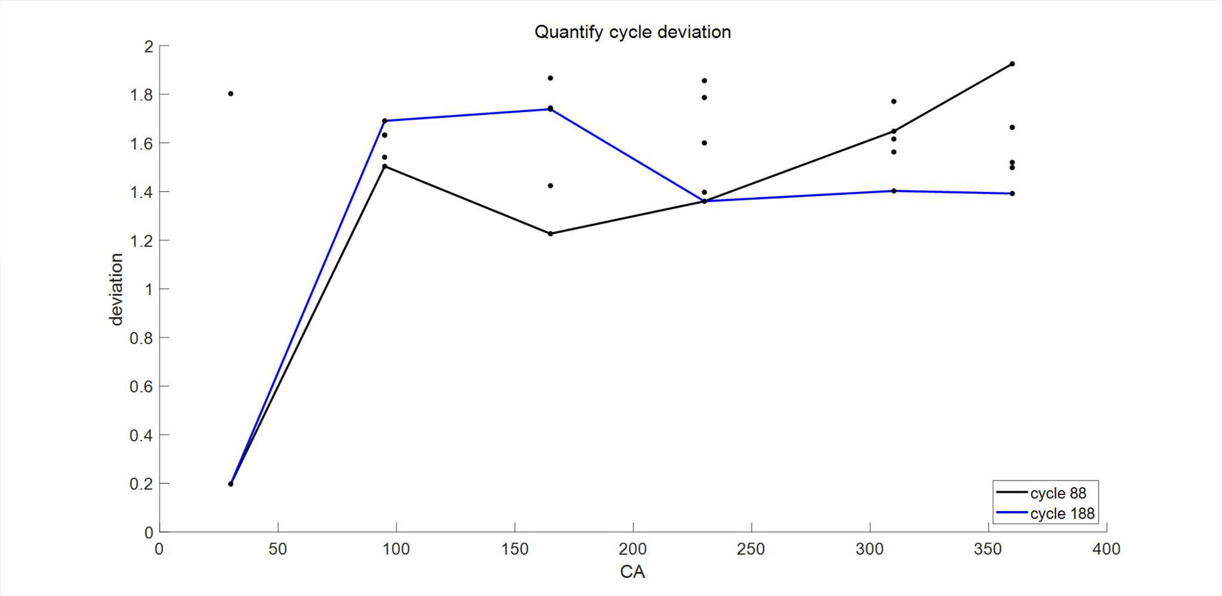


-Every cluster is associated with a deviation value. Note the vertical axis change from cluster number to deviation.



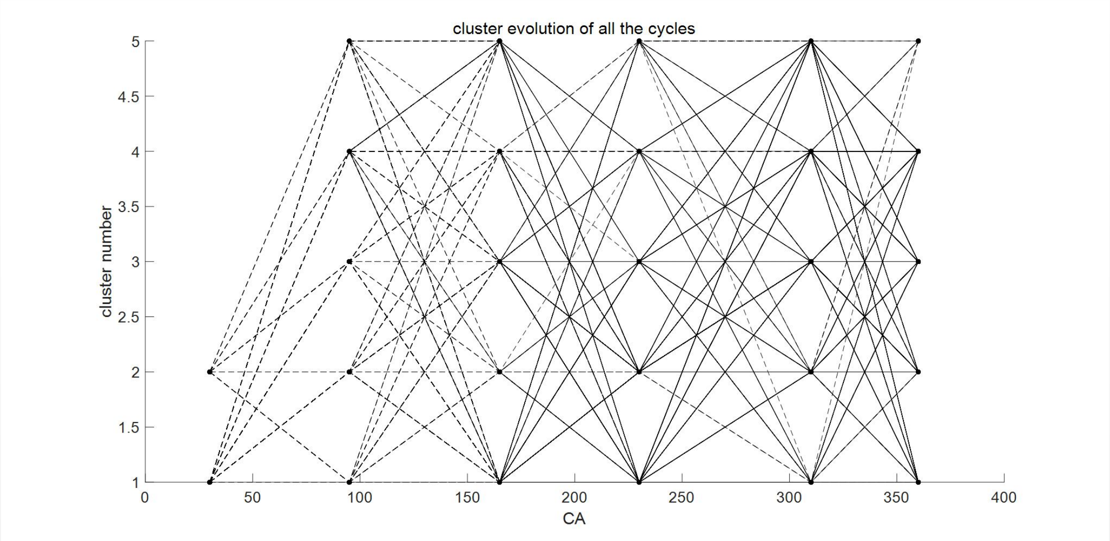
Quantify deviation of cycles





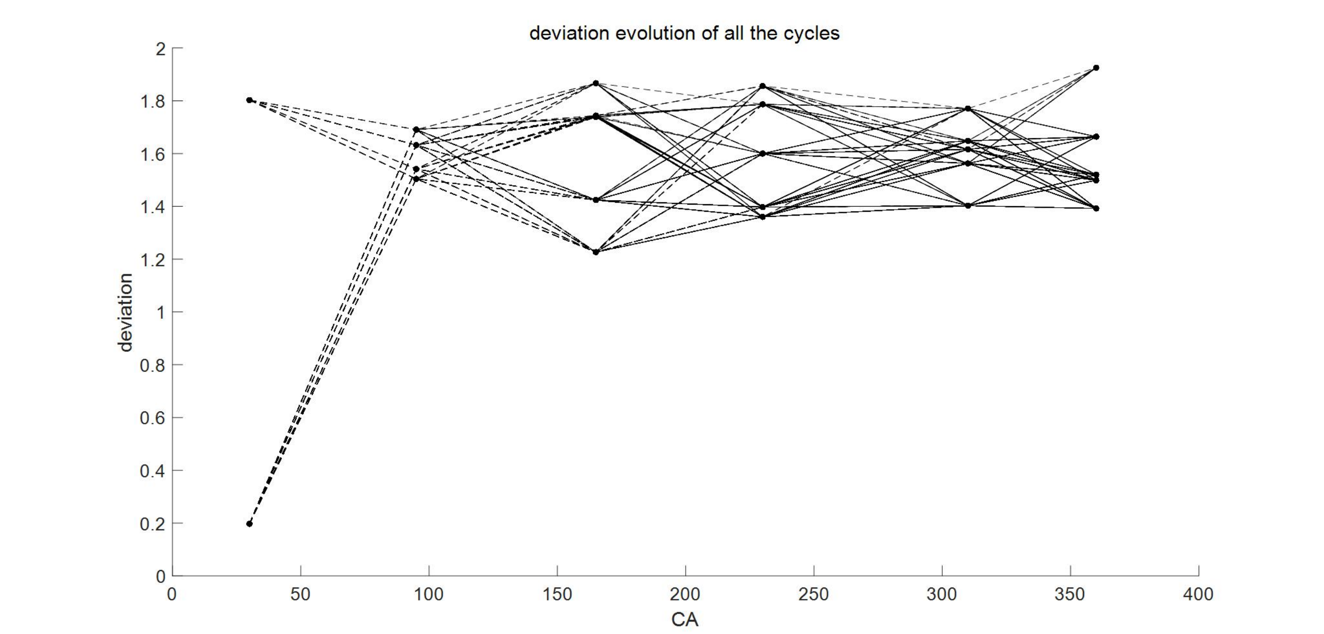
-Cluster transition process of a cycle is replaced by a deviation sequence.

Overall view of quantification(Method 1 to investigate H)



-Both plots show the evolution of all cycles, the first plot shows cluster transition, while the second shows deviation sequence.

-In fact, the second plot contains all the information about **H** (distance evolving matrix)



Investigate H (Method 2)

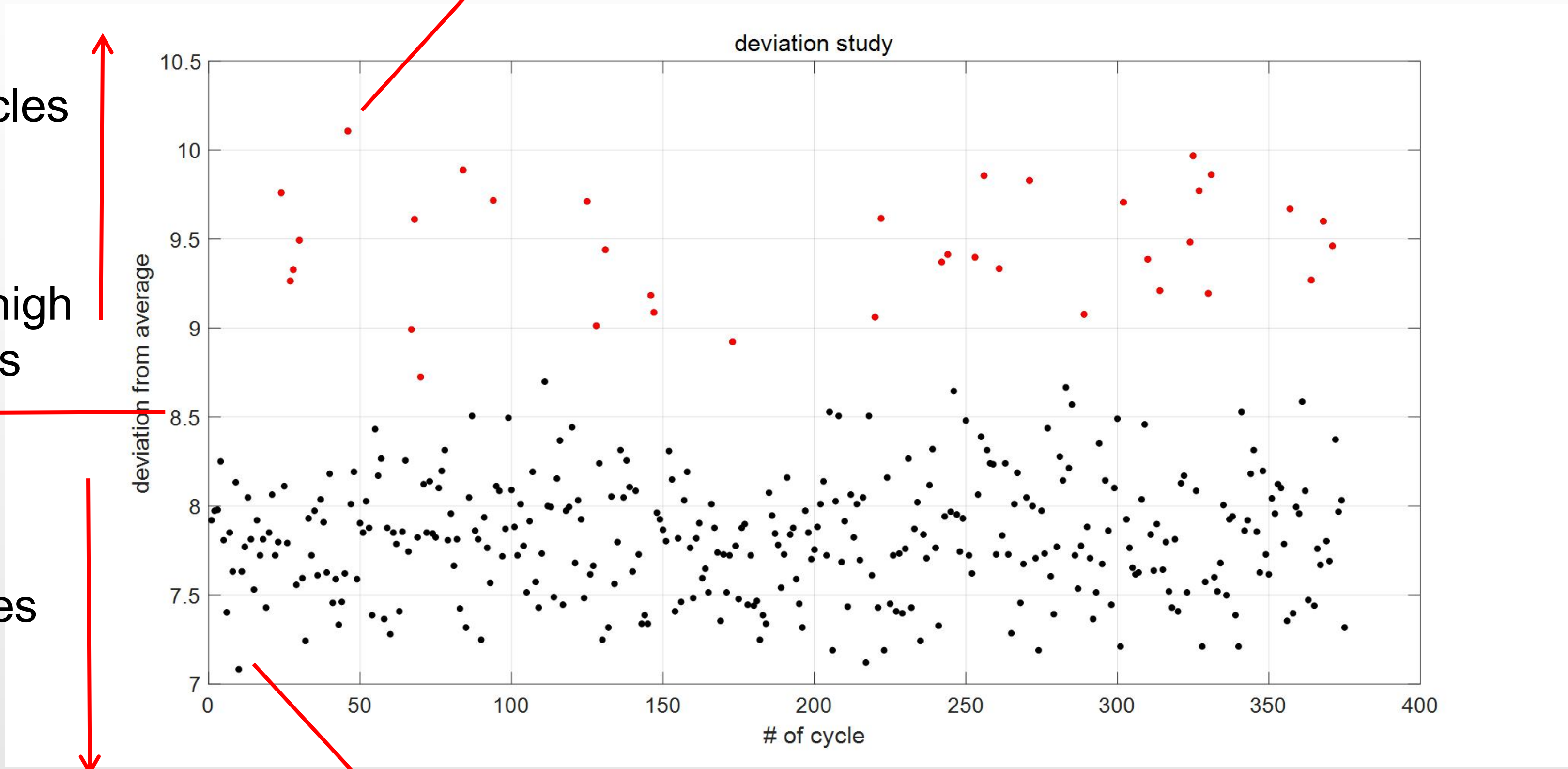
- Sum up every row of **H** and plot

highest deviation cycle #46

High deviation cycles

Threshold to separate high and low deviation cycles

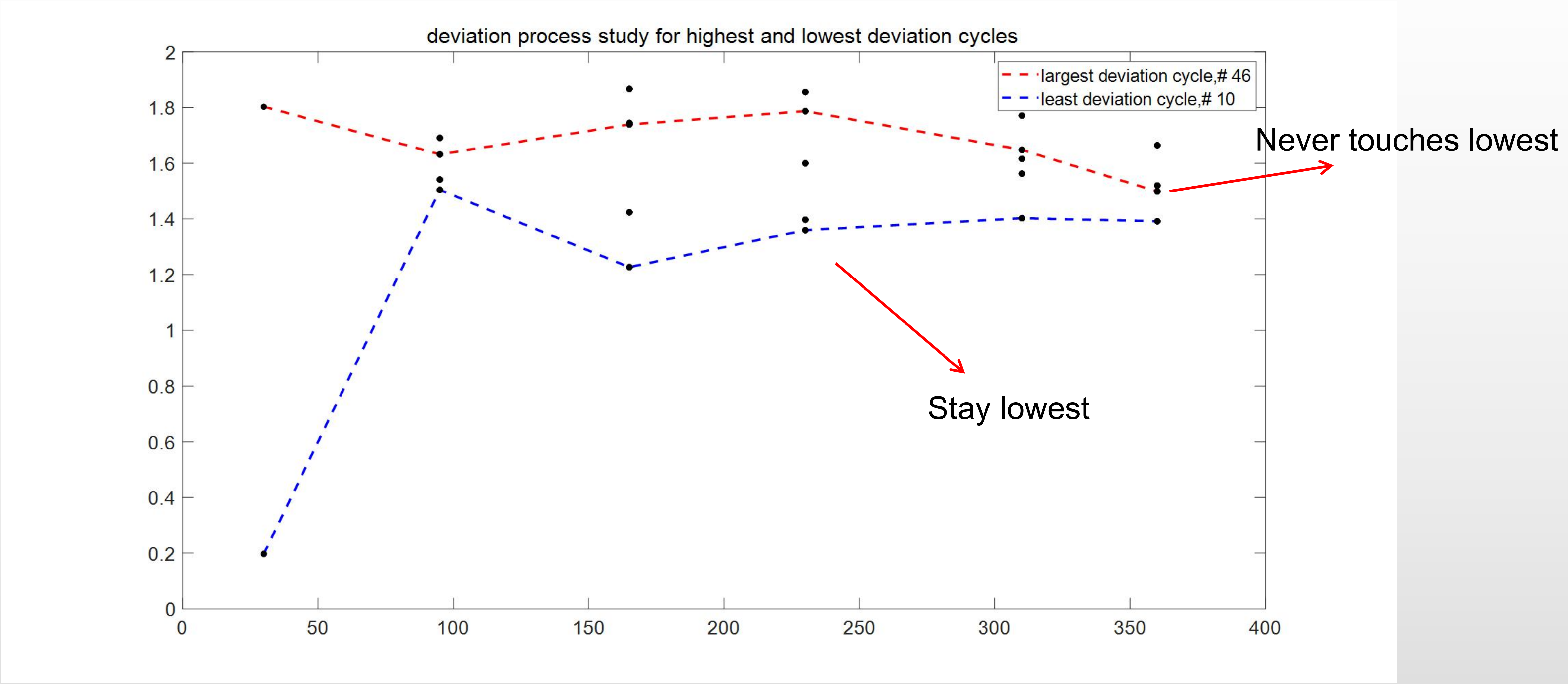
Low deviation cycles



lowest deviation cycle #10

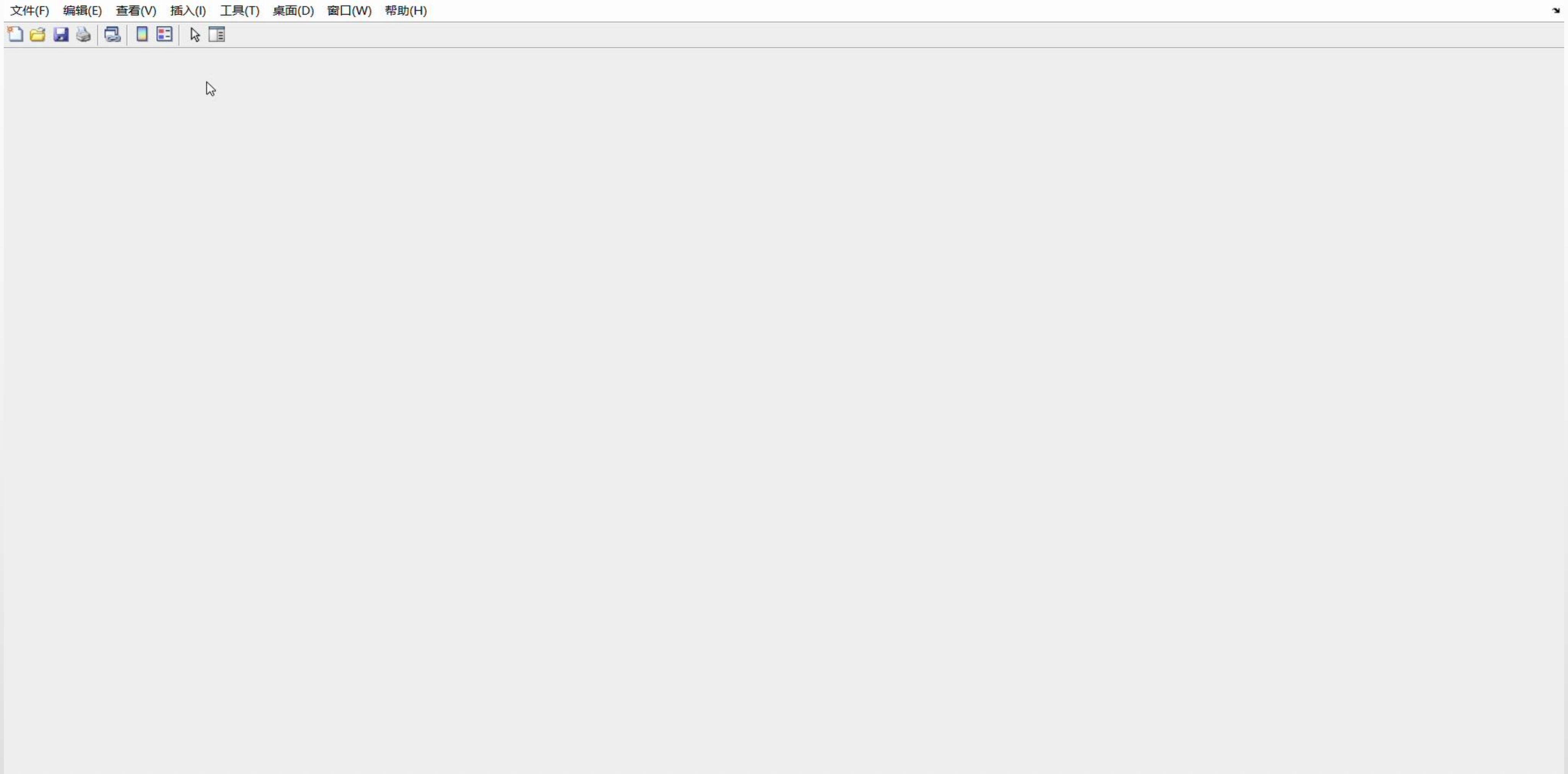
Investigate H (Method 2)

- See the deviation sequences of the highest and lowest deviation cycles.



Investigate H (Method 2)

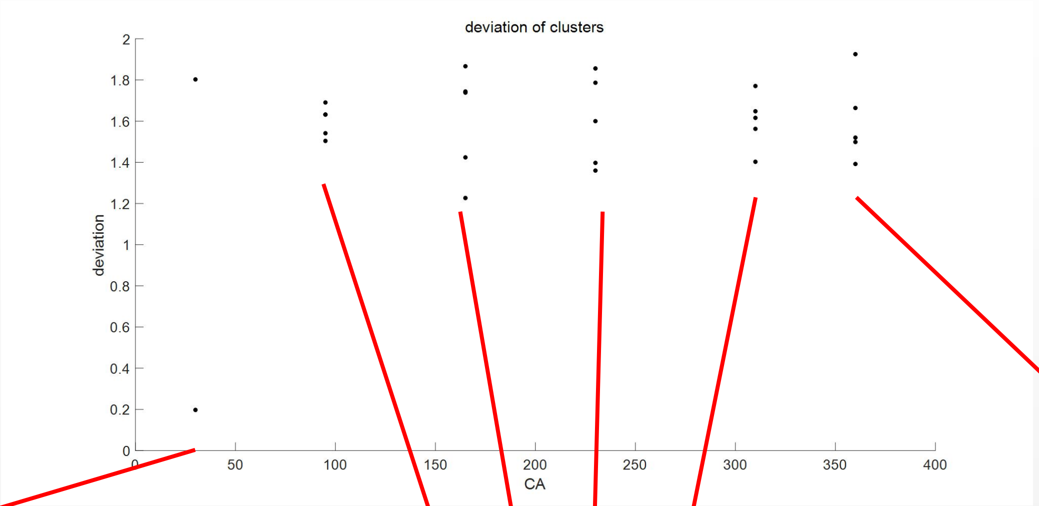
-We can visualize highest and lowest deviation cycles and compare.

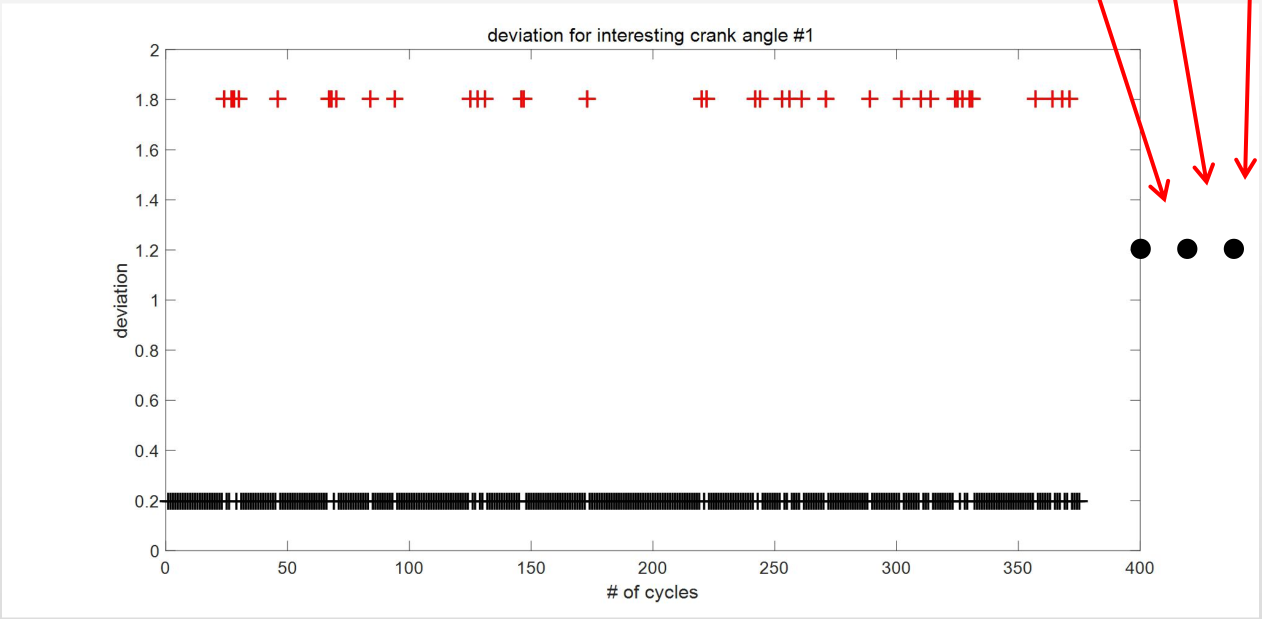


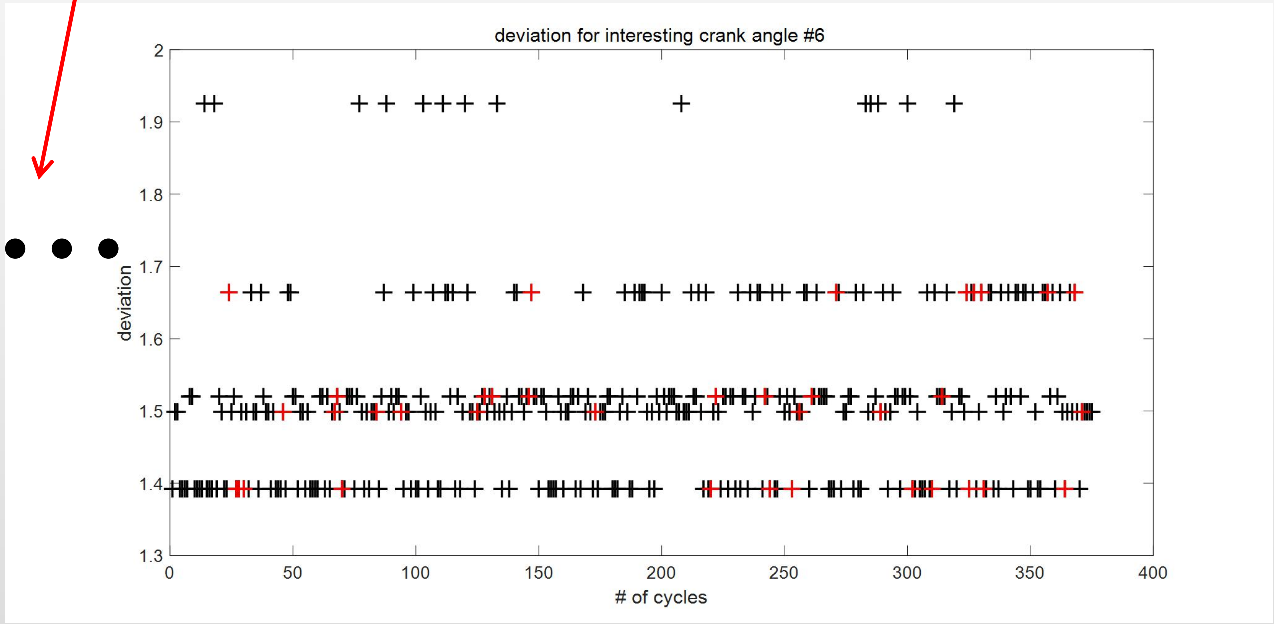
Investigate H (Method 3)

- Let's track cycles with high deviation at the beginning

-To find out whther high deviation at the beginning ends up with high deviation at the end
-If that is the case, the system could be unstable
-For this input data, that's not the case. None of the interesting cycles ends up with high deviation

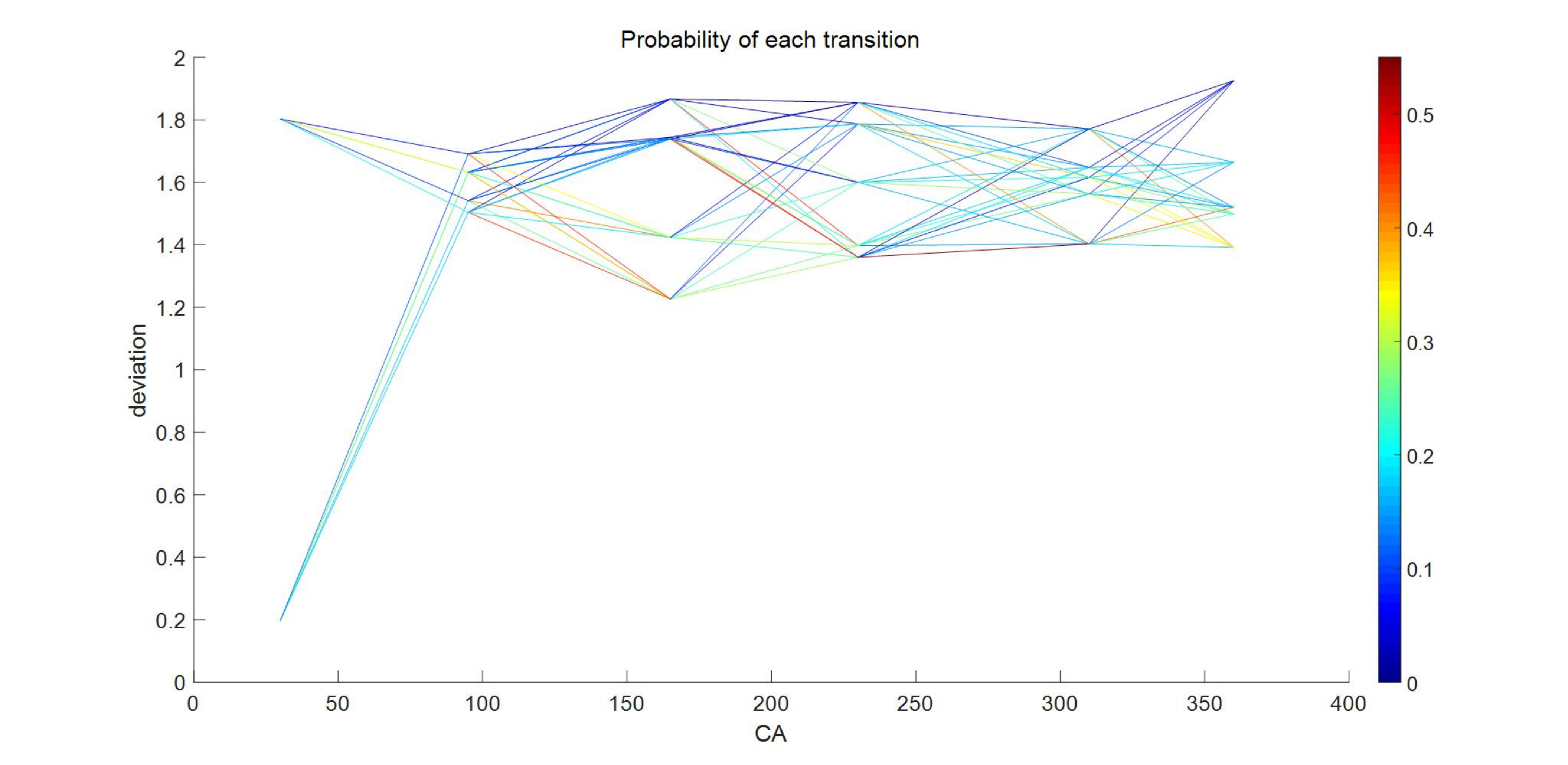






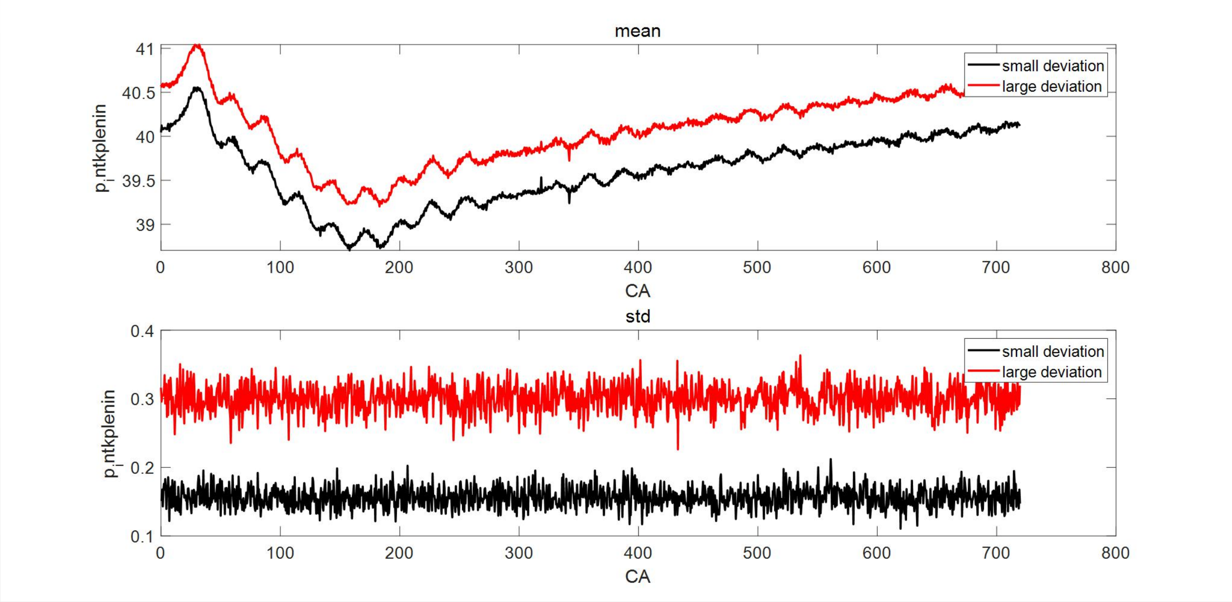
Investigate H (Method 4)

- Let's compute every probability of transition between clusters
  - In these cycles, they tend to transit to lower deviation. They may tend to go stable.



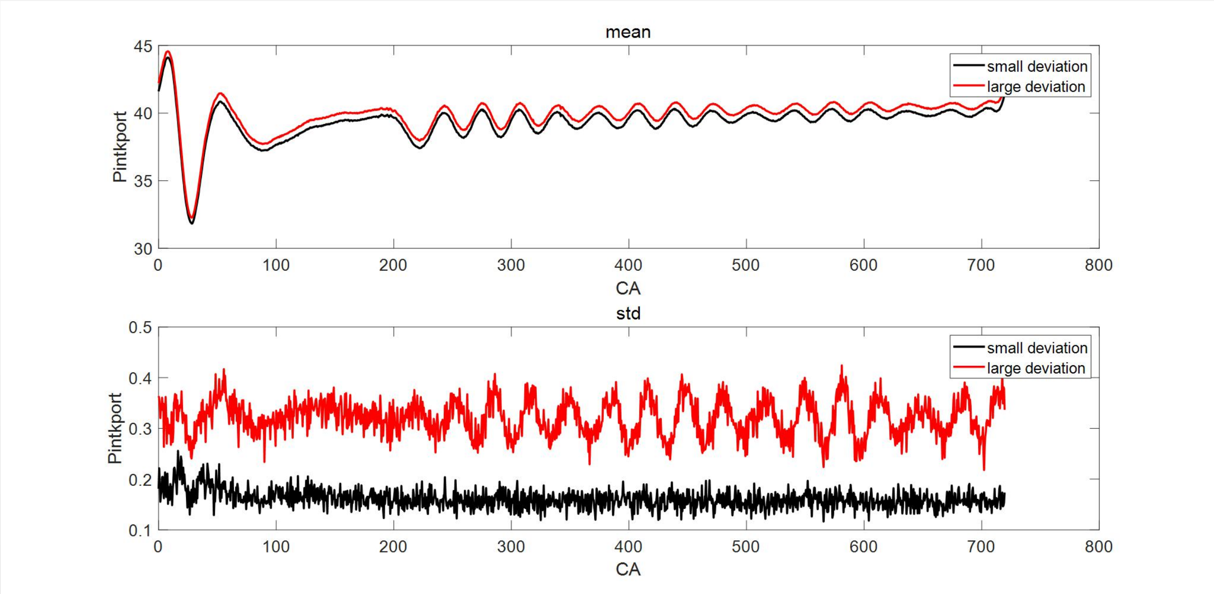
Relate to pressure signal

- Since high and low deviation cycles are separated (Method 2 to investigate **H**), it is possible to find out whether their corresponding pressure signals are also separated.

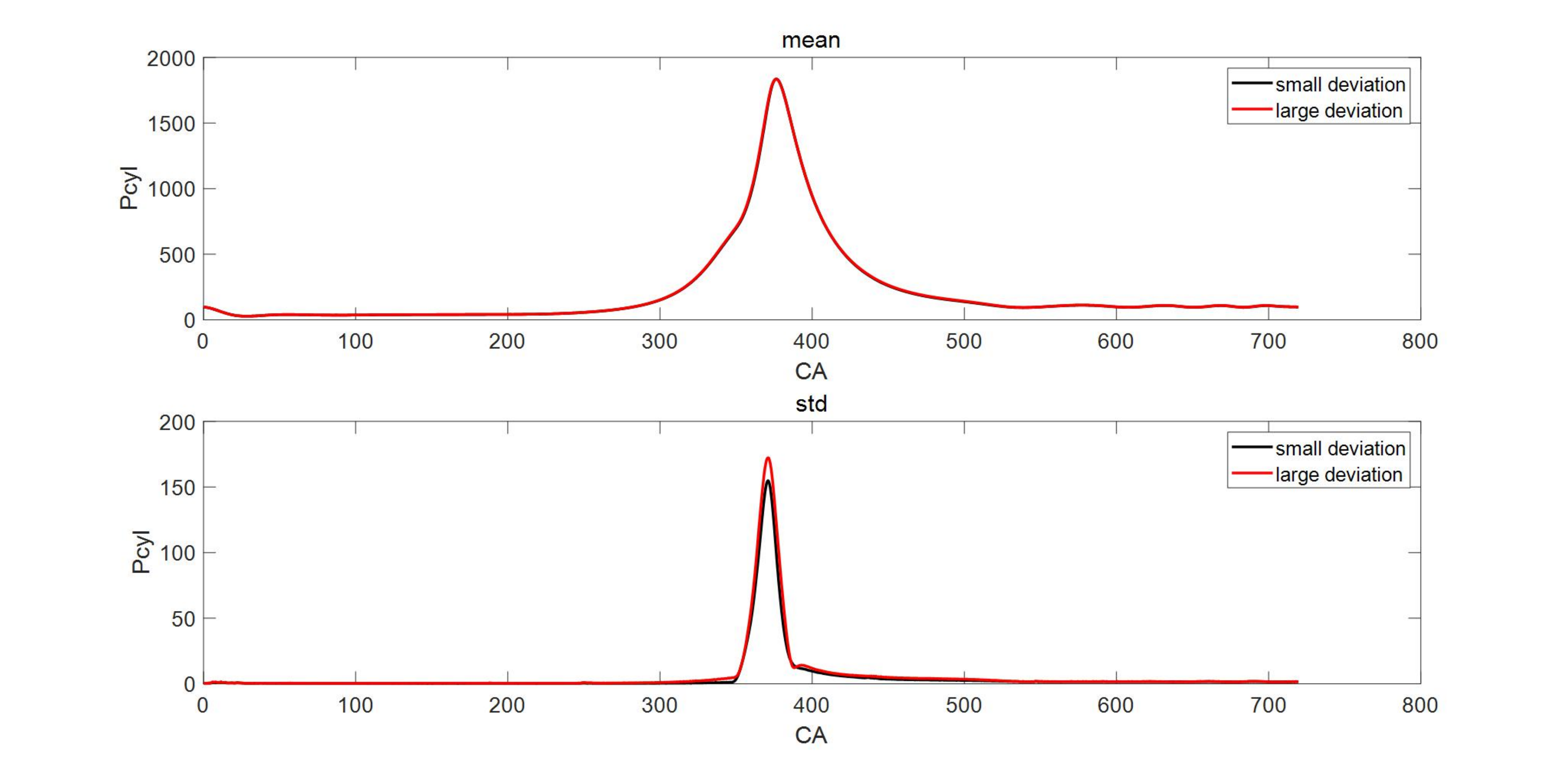


- From these two sensors, we see the separation

of pressure signal is obvious

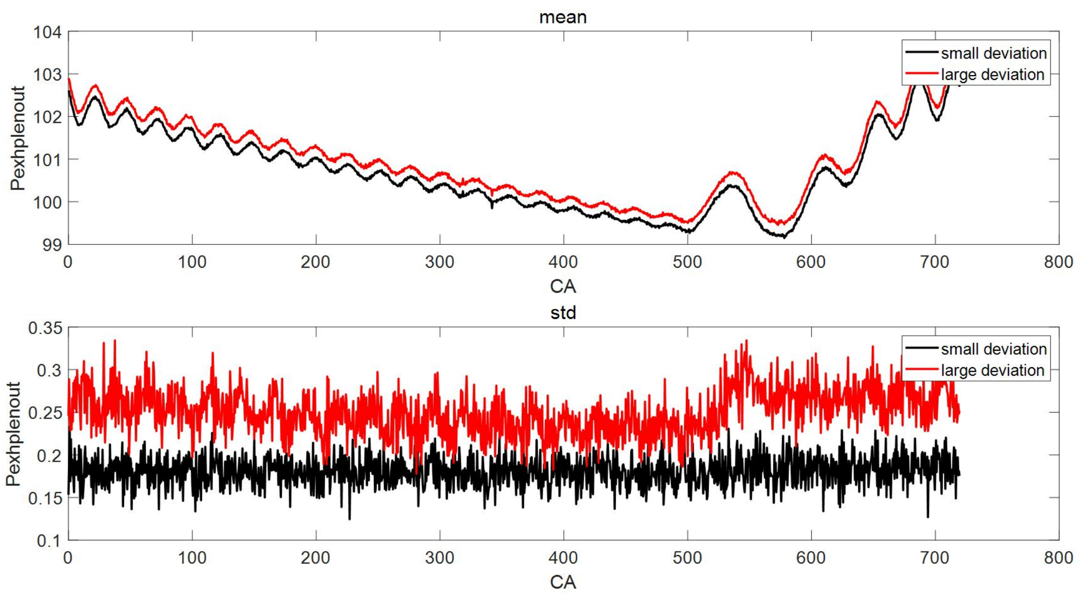


Relate to pressure signal

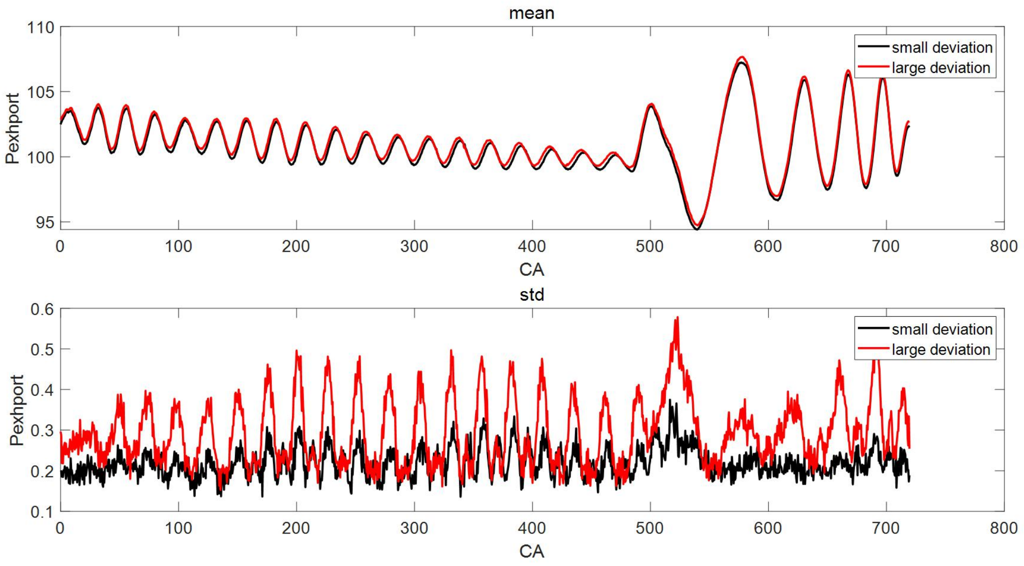


-This sensor does not convey obvious separation between two groups of cycles.

Relate to pressure signal



-From thse two sensors, separation of pressure signals is also obvious.
-It may be true that for this system, large deviation in pressure indicates large deviation in flow feature.



Conclusion

- An approach to simplify the description of similar cycles is proposed. It is realized by discretising time and clustering through flow features.
- An idea to quantify deviation of flow features and cycles is derived afterwards. They are computed as distance between probability distributions.
- Several methods to investigate the deviations are implemented, which may distinguish between high and low deviation cycles, reflect the tendency of variation and verify the stability of a system.
- Pressure signals may indicate the deviation of flow features by studying regulation beforehand.