Application of
Inductive Monitoring System (IMS)
for monitoring industrial processes

Enayet B. Halim, Harigopal Raghavan,
Sirish L. Shah and Frank Vanderham

Department of Chemical and Materials Engineering
University of Alberta

Outline

1. Motivation
2. Methodology
   a. Monitoring
   b. Diagnosis
3. Examples
   a. Rig Data
   b. Industrial Data
4. Advantages
5. Concluding Remarks

Motivation

Quote from a Process Engineer in industry:

"We neither have the time nor the resources (skills, tools and 'clean data') to obtain models of normal process operation and subsequently use these to detect abnormal events/situations."

Motivation

Problem with model-based system monitoring and diagnosis:

- Building models is often a difficult and time consuming process
- Systems may require models that are too complex to use for real time monitoring system

Solution is to use a data-based monitoring system:

- Clusters are defined from training data
- Each cluster represents a different mode of operation

Proposal: Inductive Monitoring System (IMS)

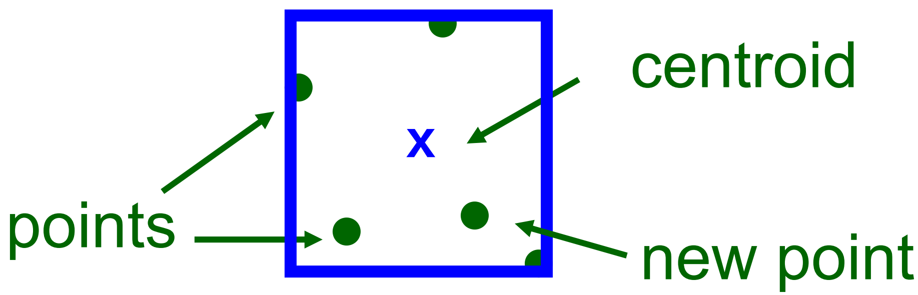
IMS automatically defines groups of consistent system parameter data by examining and generalizing from examples of nominal system data.

Methodology

Temp 1

All clusters are defined by their centroids and their boundaries

Boundaries are defined by the minimum and maximum values of the cluster points



If a new point falls within the boundaries of a cluster it is added as a member to the cluster

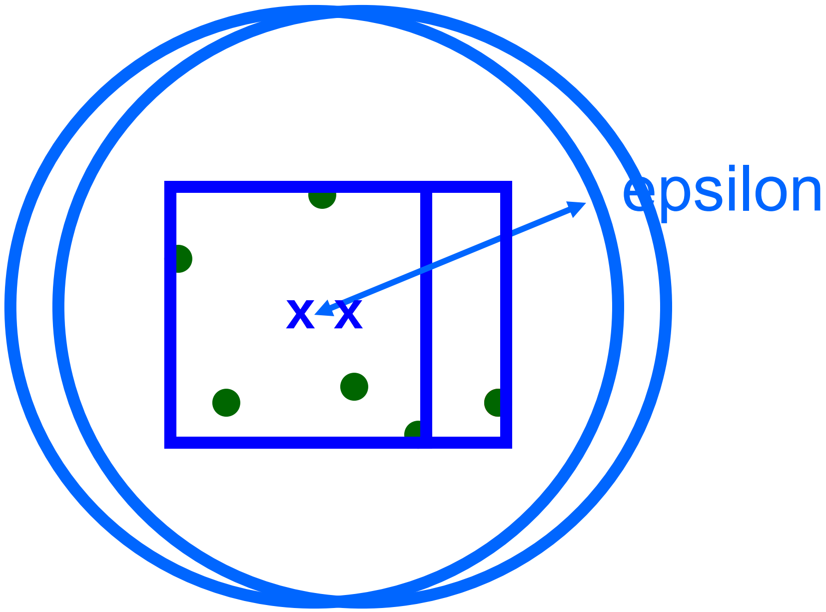
Temp 2

Methodology (cont.)

Temp 1

Also a new point is added to the cluster if it falls within a user defined (epsilon) radius

Cluster centroid and boundary changes accordingly



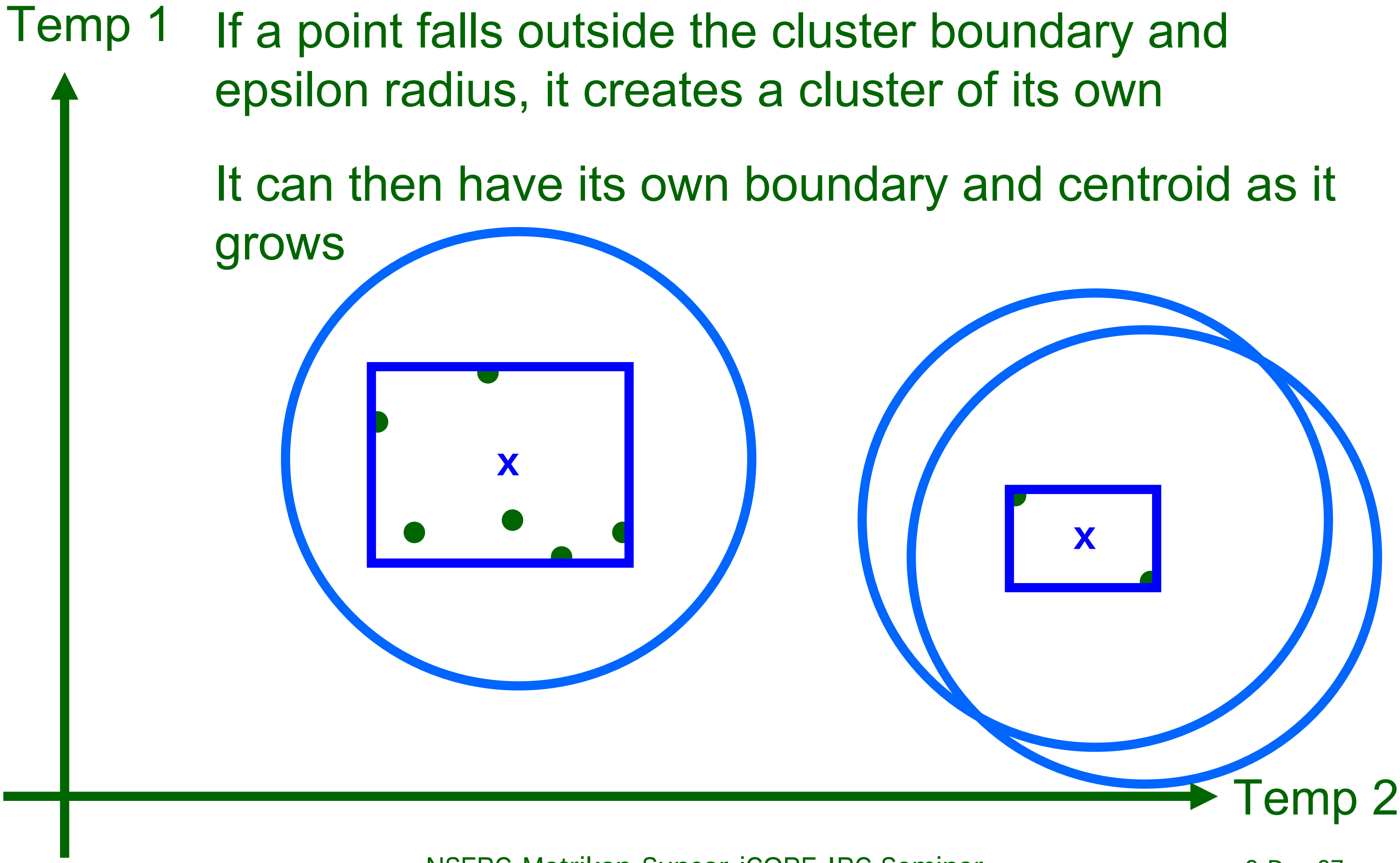
Epsilon radius relocates with the centroid

Temp 2

Methodology (cont.)

If a point falls outside the cluster boundary and epsilon radius, it creates a cluster of its own

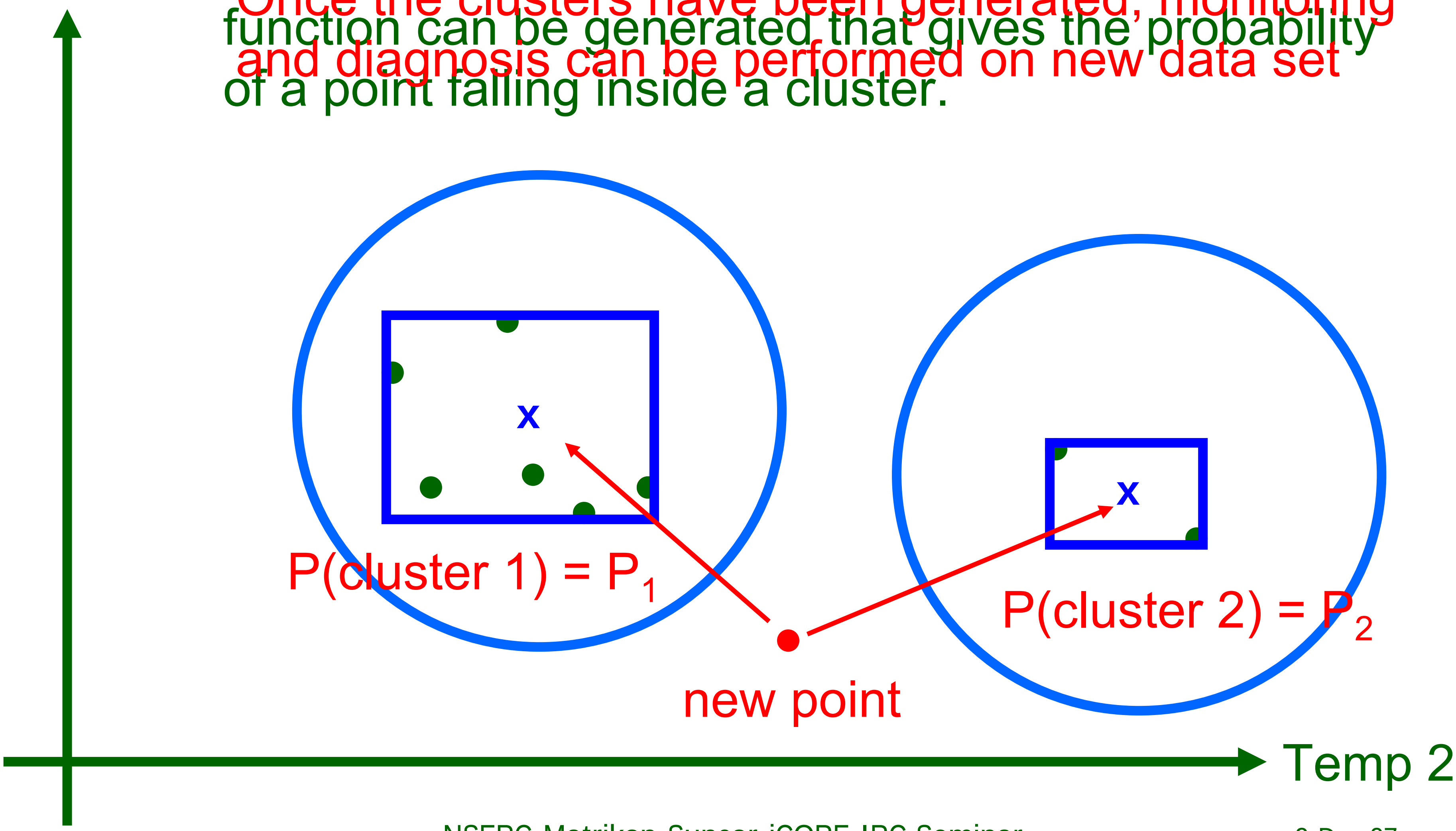
It can then have its own boundary and centroid as it grows



Monitoring

With enough training data, probability distribution function can be generated that gives the probability of a point falling inside a cluster.

Once the clusters have been generated, monitoring and diagnosis can be performed on new data set

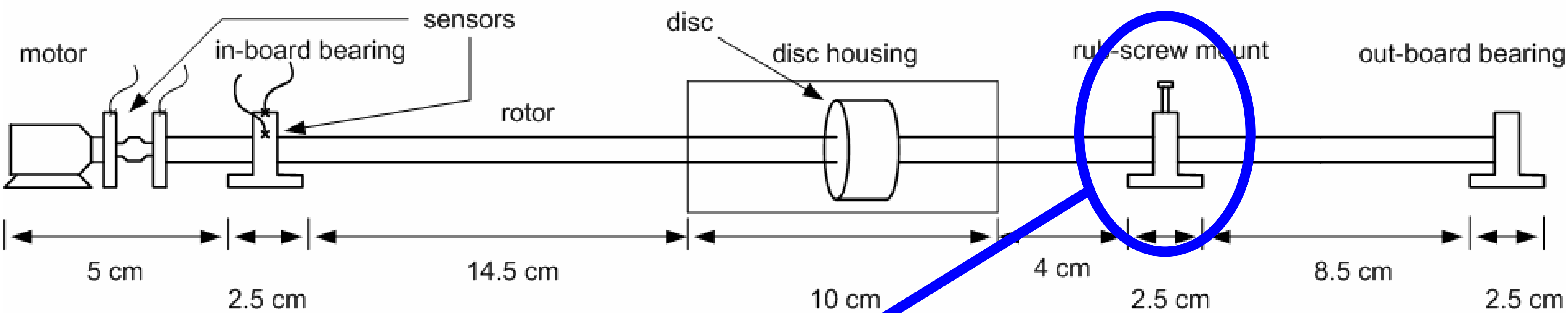


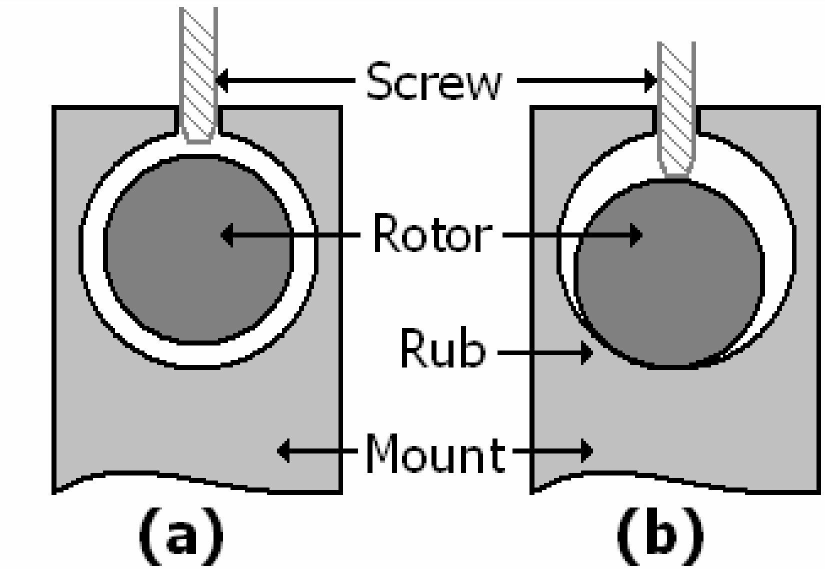
Diagnosis

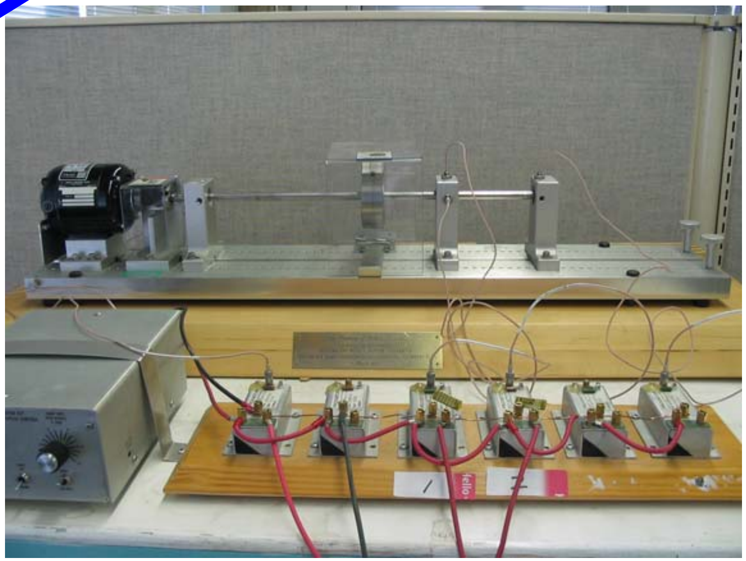
1. IMS fuzzy scheme gives the probability (in numerical value) of a sample being classified as a fault.
2. IMS not only detects fault samples but also groups fault samples into fault clusters. Each of the fault clusters represent different modes of failure.
3. Contribution from different variables to the fault sample can help specify the variable(s) most responsible for the mode of failure.
4. IMS does not require any supervised learning. The only user defined input is the Epsilon, but it can be optimized too using the Fukuyama-Sugeno Index, which in turn optimizes the number and size of the clusters.

Example – Rig data

The rig is located in the Reliability Lab in the
Mechanical Engineering Building at the University of Alberta, Canada







Example – Rig data (cont.)

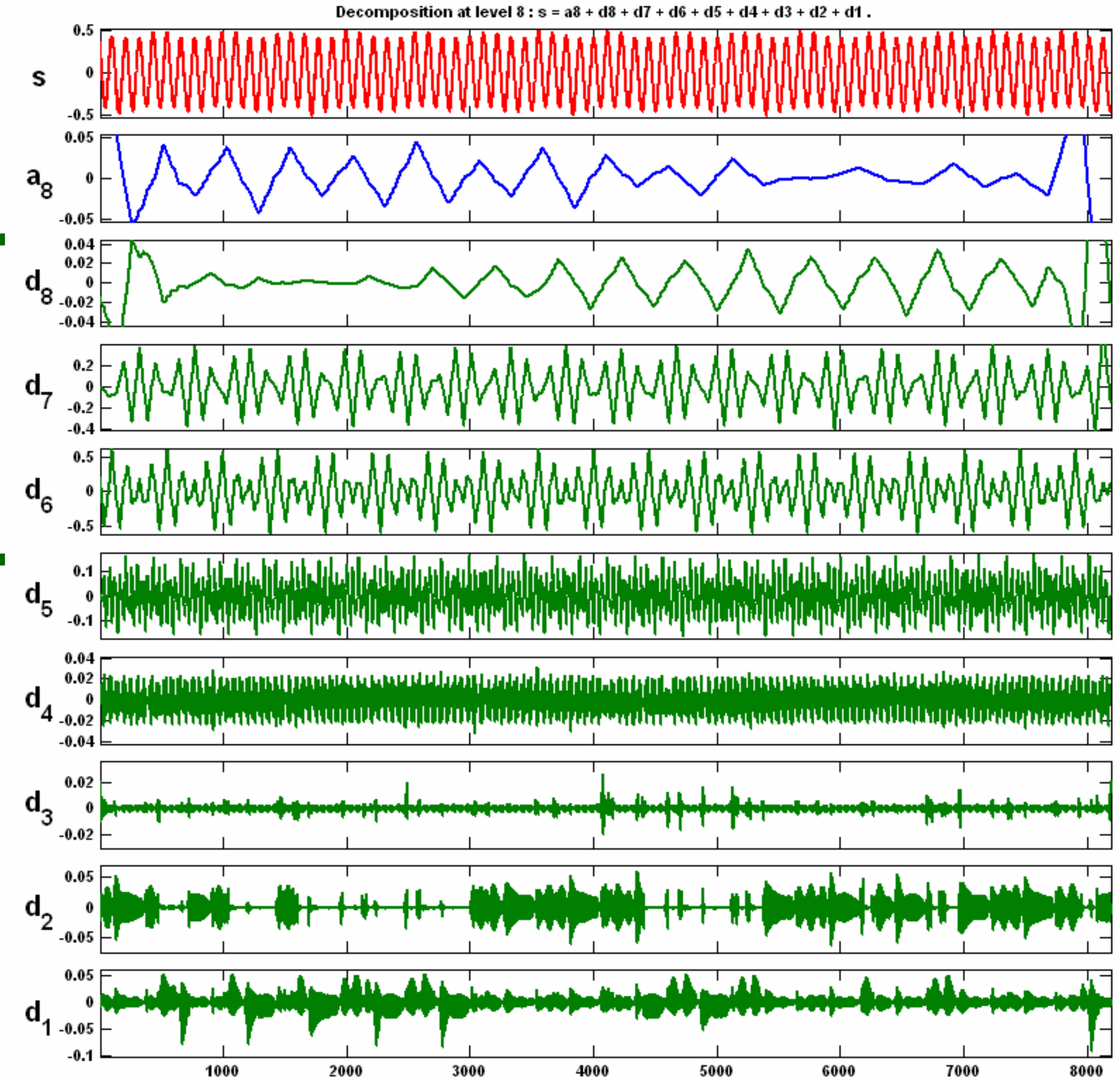
Data has been collected under 4 modes of operation:

1. Normal Operation
2. Mild Rub introduced
3. Severe Rub introduced
4. Normal Operation + white noise with SNR = 2.5

Vibration data in univariate. It is transformed to multivariate data by dividing it into different frequency bands using Discrete Wavelet Transform (DWT).

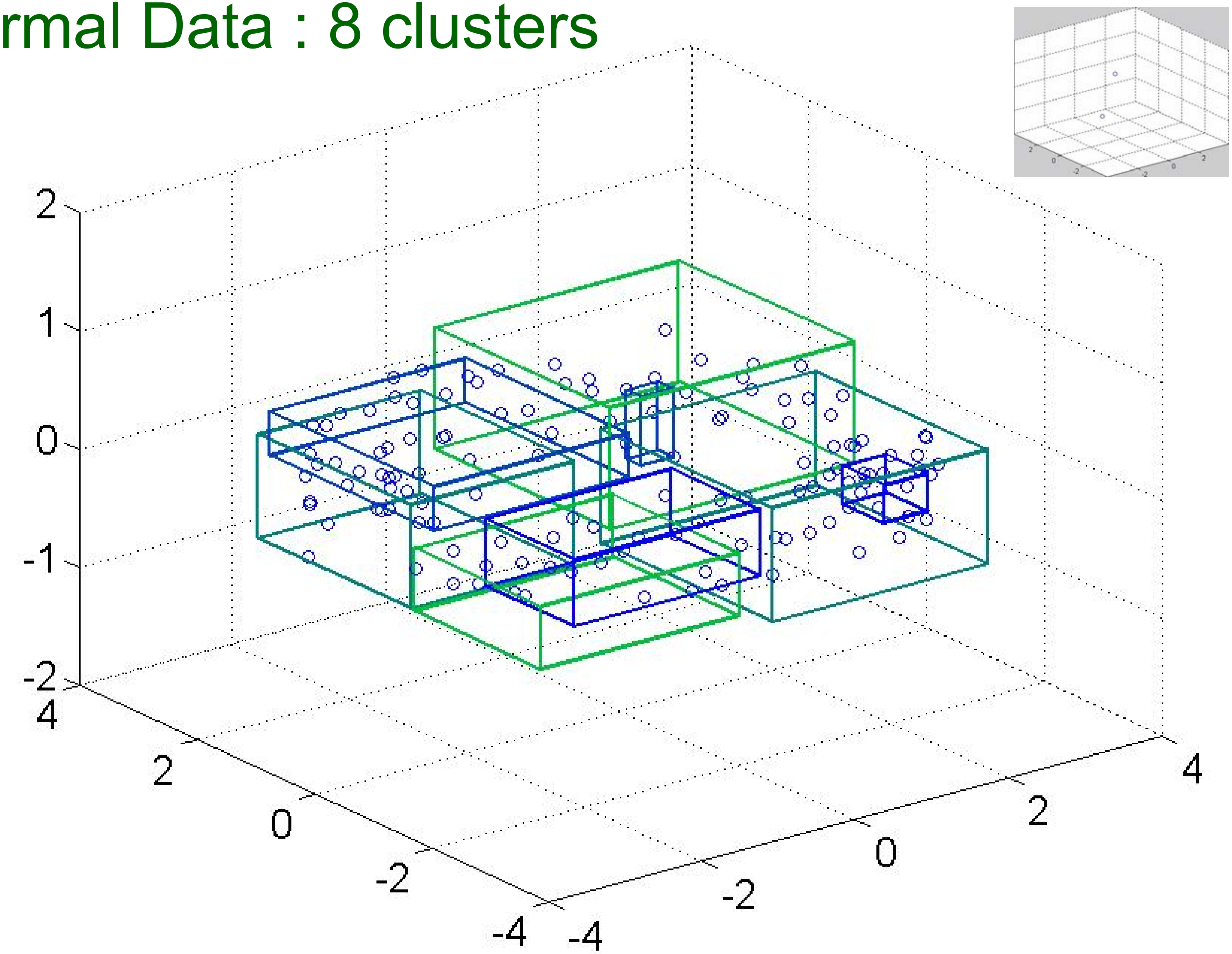
Example – Rig data (cont.)

The original signal is divided into 8 details using db-3 wavelet.

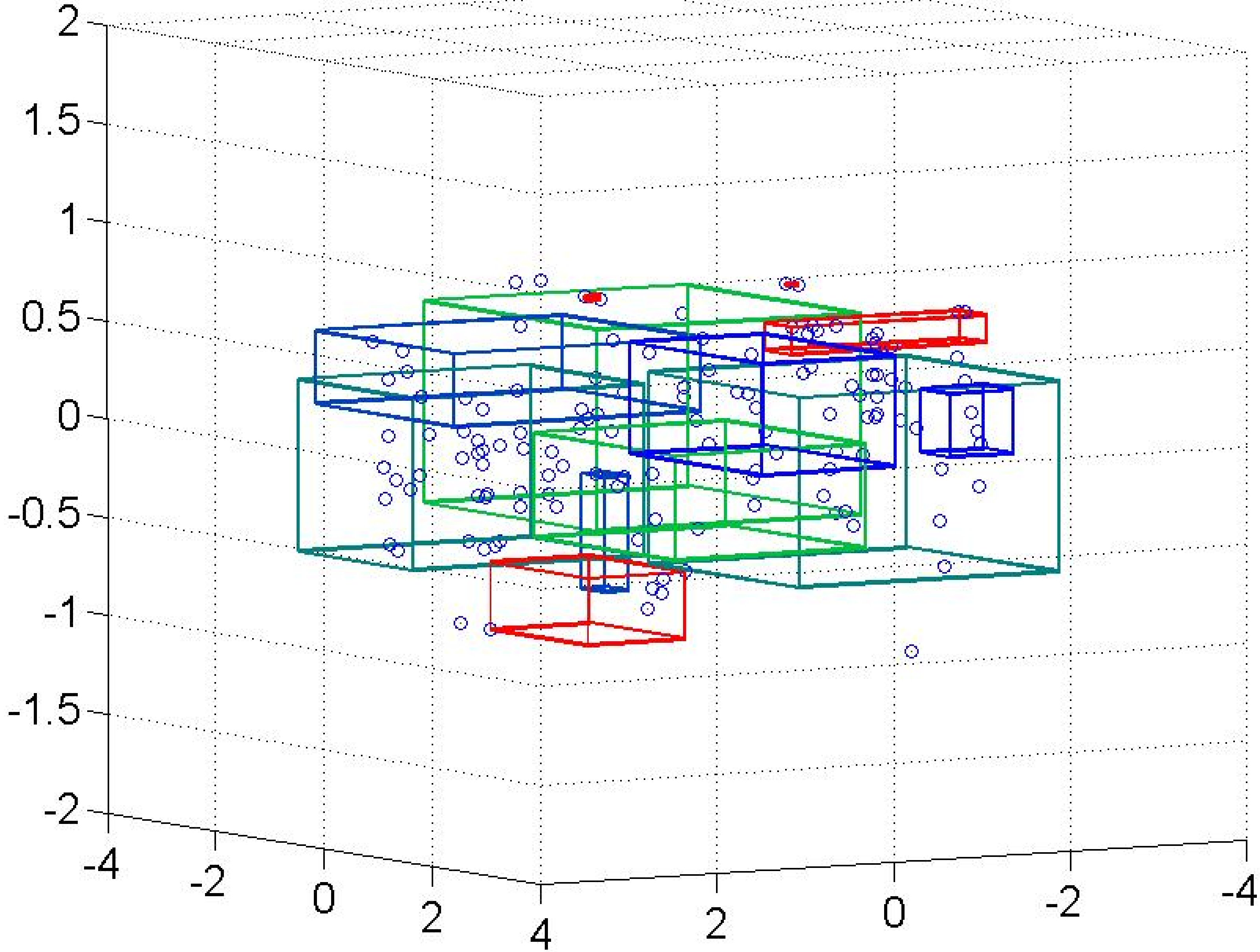


For the sake of visualization, details 6, 7 and 8 were taken as separate variables and analyzed.

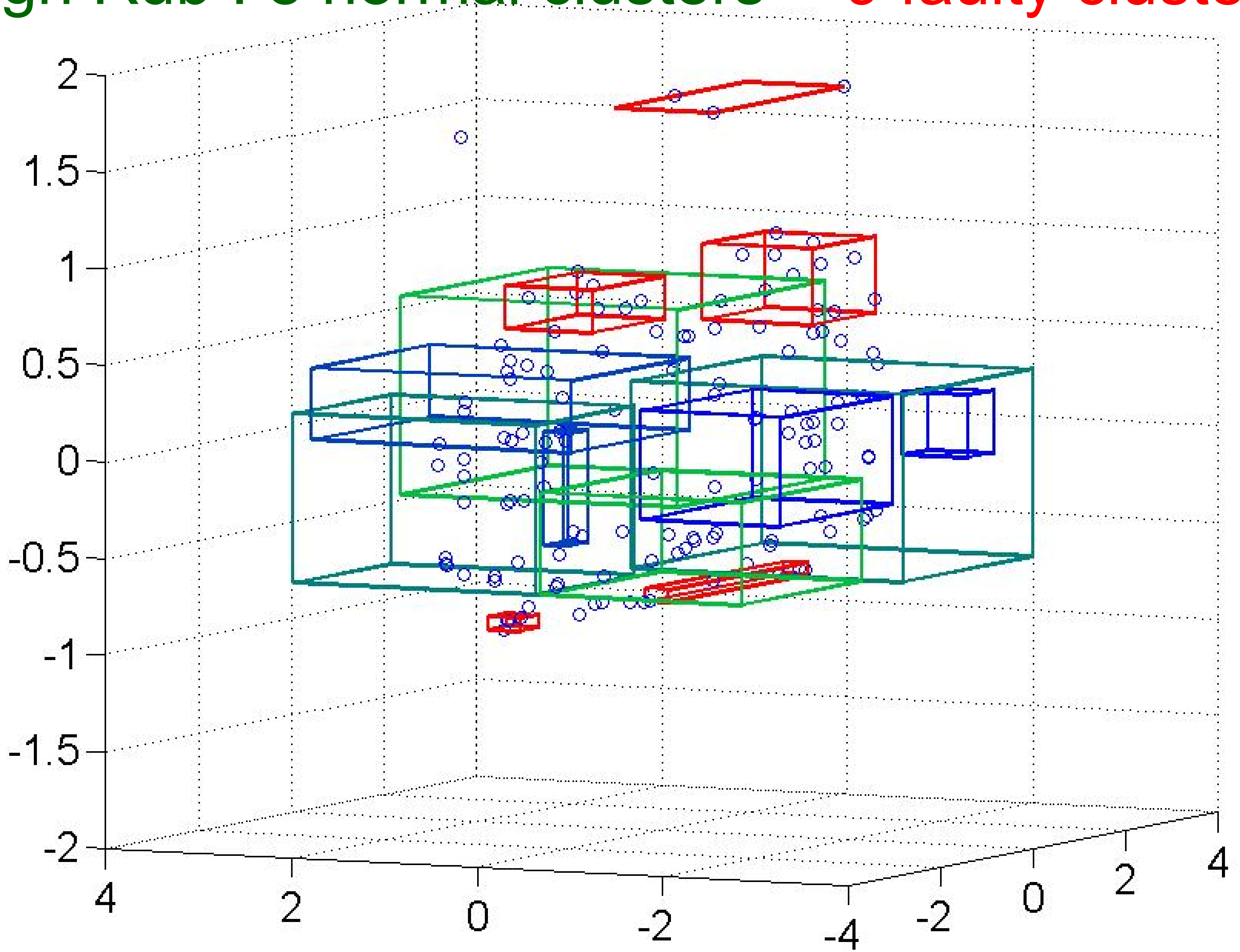
Normal Data : 8 clusters



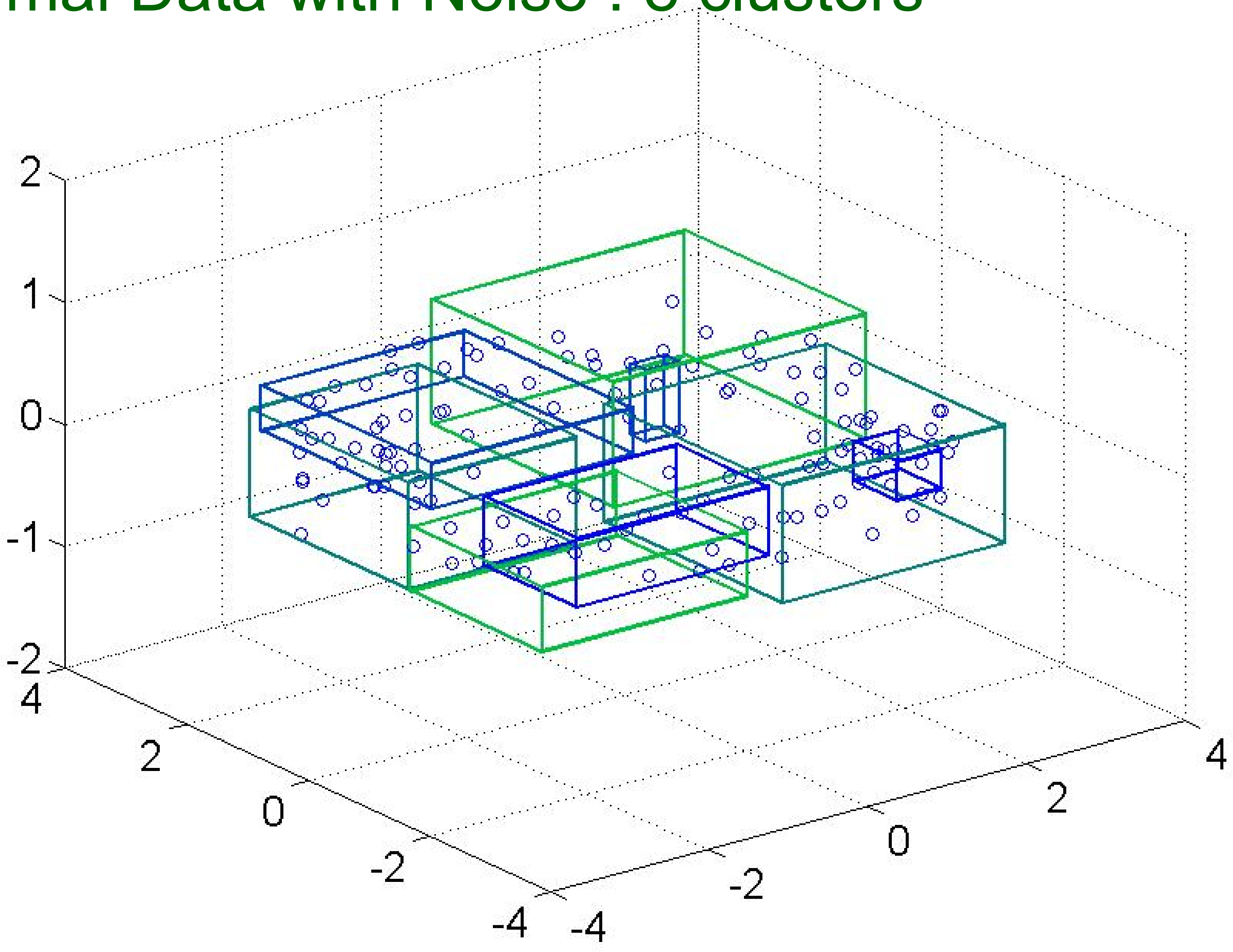
Low Rub : 8 normal clusters + 7 faulty clusters



High Rub : 8 normal clusters + 9 faulty clusters



Normal Data with Noise : 8 clusters



Example – Industrial data

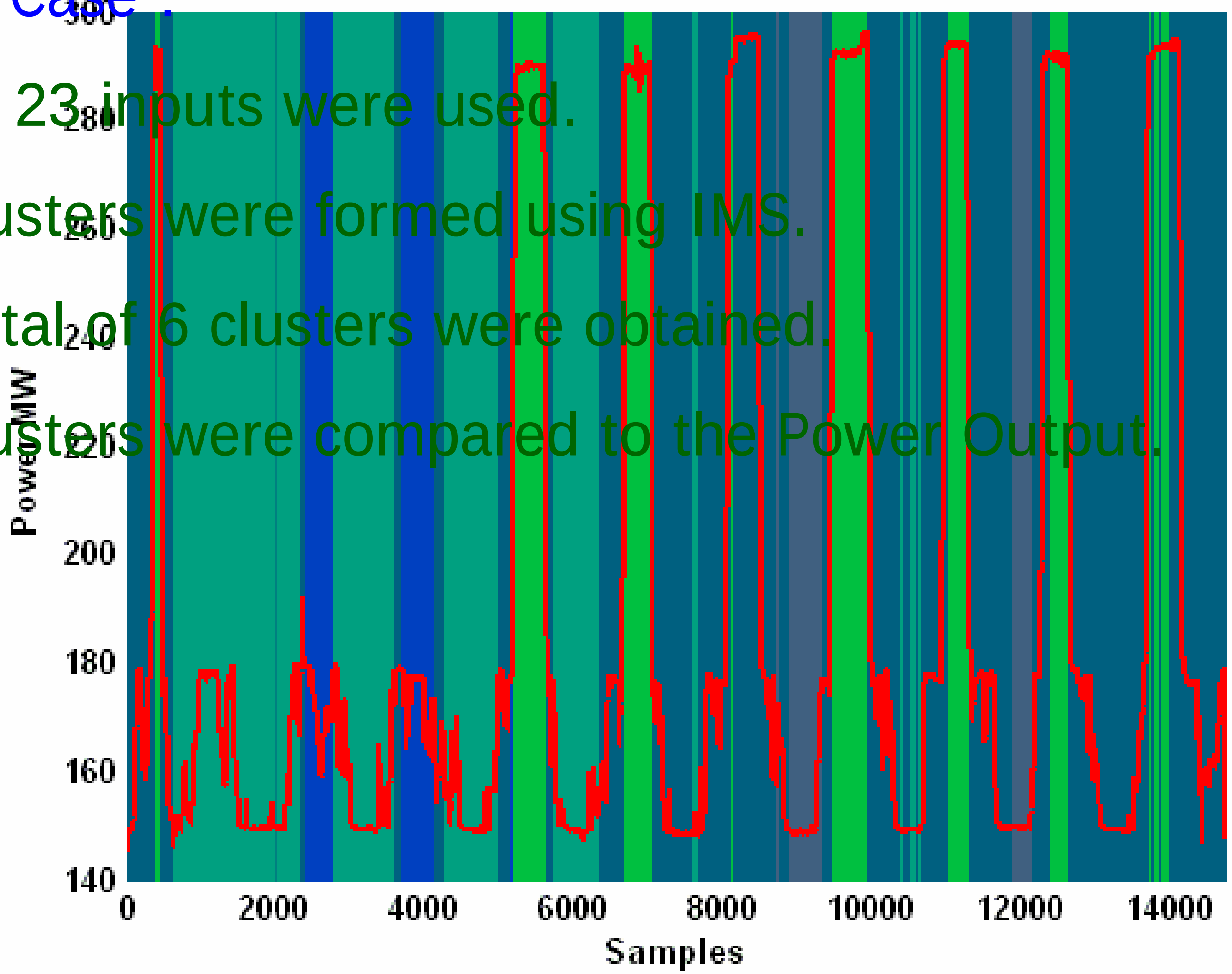
Data was taken from a power plant.

23 input variables
2 output variables including Total Power Output in MW.

Example – Industrial data (cont.)

Initial Case :

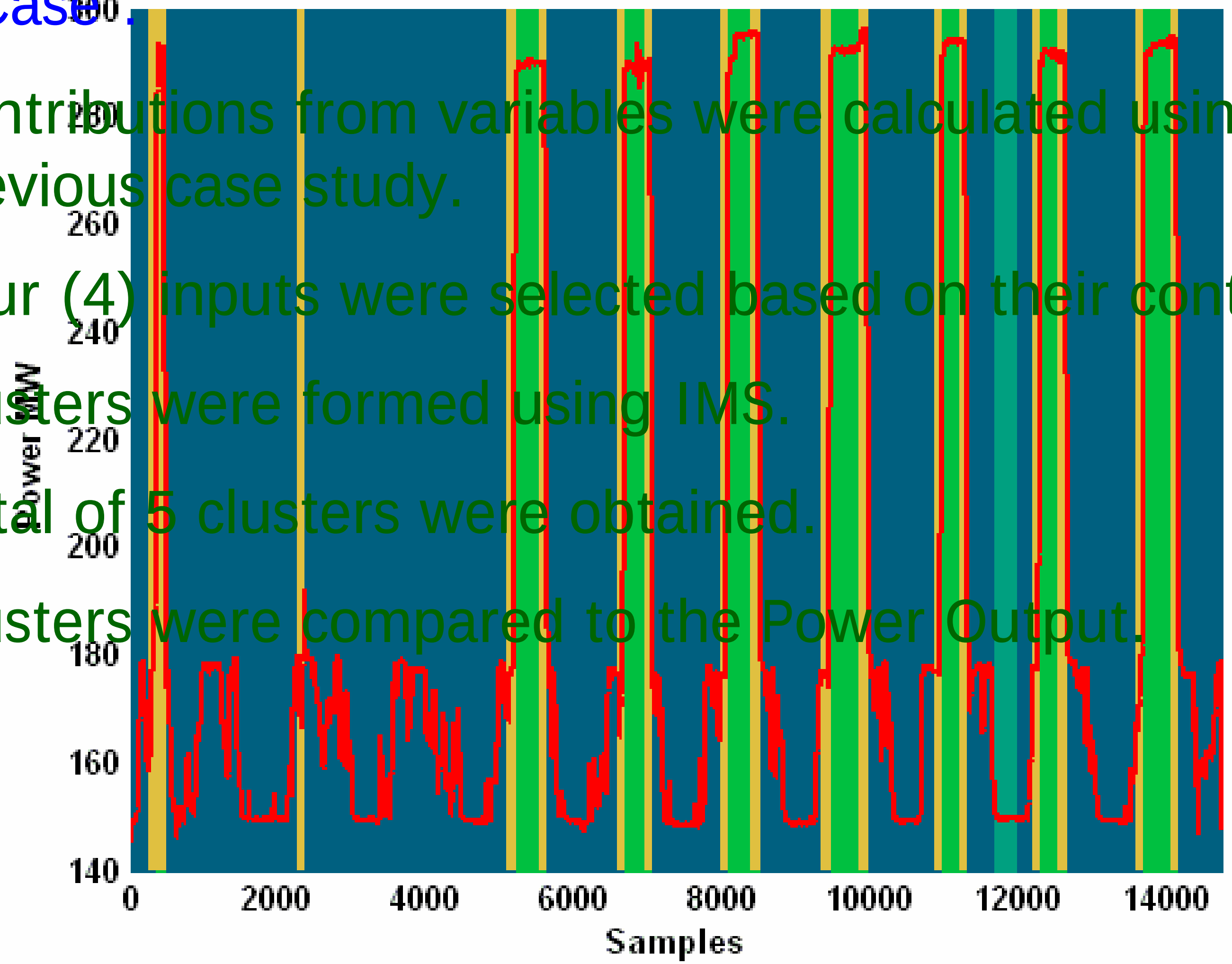
1. All 23 inputs were used.
2. Clusters were formed using IMS.
3. Total of 6 clusters were obtained.
4. Clusters were compared to the Power Output.



Example – Industrial data (cont.)

Final Case:

1. Contributions from variables were calculated using previous case study.
2. Four (4) inputs were selected based on their contribution.
3. Clusters were formed using IMS.
4. Total of 5 clusters were obtained.
5. Clusters were compared to the Power Output.



Advantages

1. IMS is much faster than most other clustering algorithms.

   It does not require distance calculation for most of the sample points.

2. IMS is robust and requires minimal supervision.

   It decides on the size and number of clusters on its own and trains itself automatically on nominal operation data.

3. IMS is informative. The output can be easily interpreted.

   It can indicate the contribution of each of the variables to the clusters.
   The output clusters can be attributed to different modes of operation.

4. IMS is simple and can be tweaked according to need.

   The concept is easy to understand and can be easily related to.
   Variables can be weighted and the cluster size or number can be changed easily according to the need of the operation.

Concluding Remarks

1. A fully automated robust clustering method has been proposed.

2. Inductive Monitoring System has been applied successfully on both Rig and Industrial data sets.

3. Future work includes development of better visualization tools and interfaces.

QUESTIONS?

Optimizing parameters

IMS does not require any supervised learning. The only user defined input is the Epsilon, but it can be optimized too.

Fukuyama-Sugeno Index is used to find the optimum number of clusters.

$$FS = \sum_{i=1}^{C} \sum_{j=1}^{N} \left[\left\| x_{ij} - v_i \right\|^2 - \left\| v_i - v \right\|^2 \right]$$

x_{ij} = member #j of cluster #i
v_i = centroid of cluster #i
v = centroid of all v_i's

FS = Fukuyama-Sugeno Index
C = number of clusters
N = number of members in a cluster

When, FS≈0, the number of clusters is optimum.

Optimizing parameters (cont.)

Minimum Epsilon
FS calculated

Maximum Epsilon
FS calculated

Epsilon is interpolated so that FS→0
Larger Epsilon means bigger clusters

Clusters are generated based on Epsilon
Larger Epsilon means less number of clusters

FS Index is calculated
If FS≈0 then the Epsilon value
and number of clusters is optimum

Optimum size and
number of clusters
are obtained