# Data Analytics for Control Systems

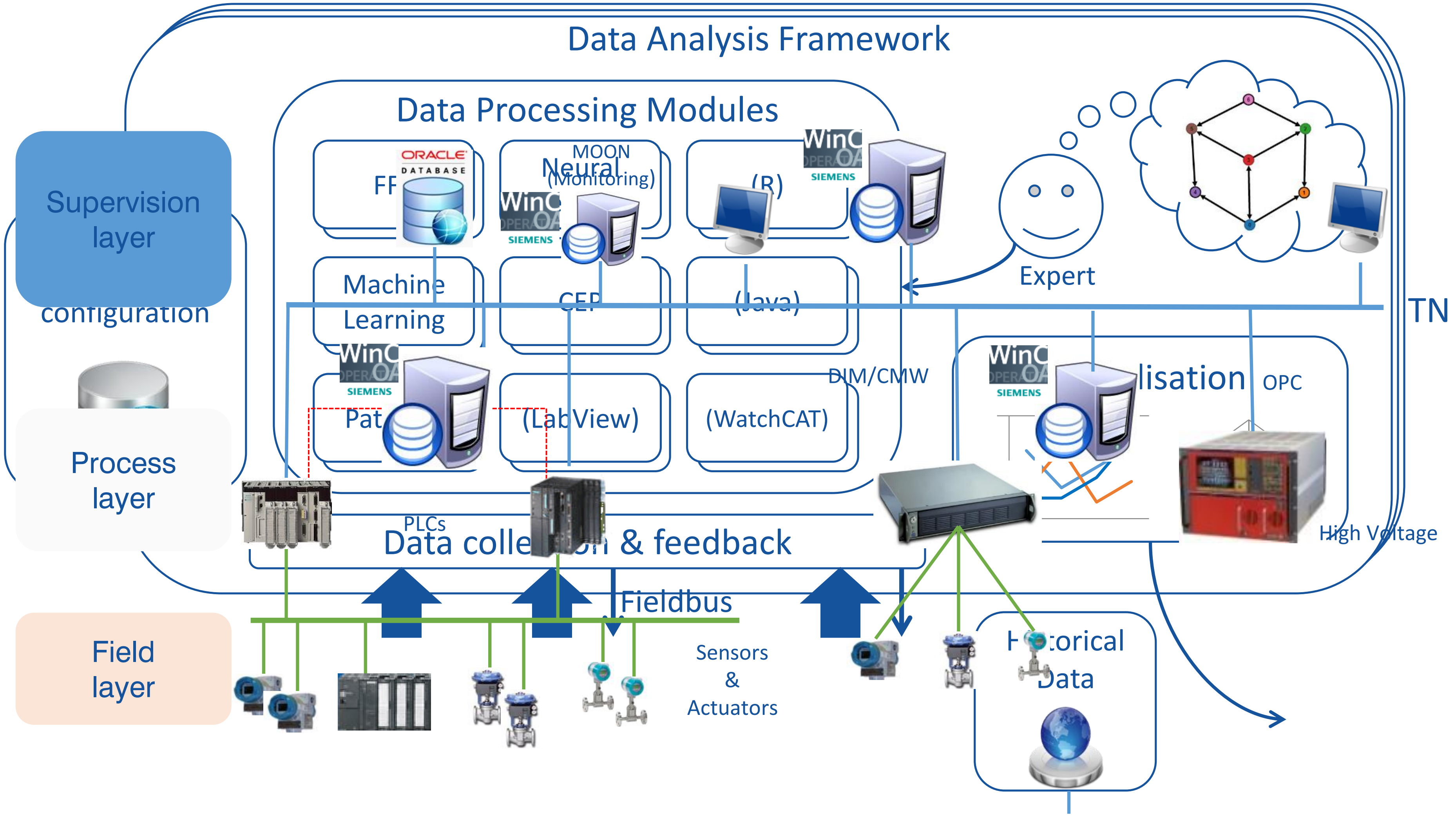
## DAQ Data Analytics Workshop

14 April 2016

*Axel Voitier*
*Filippo Tilaro*
*Manuel Gonzalez Berges*

# Our vision of the analysis framework

Scalable and fault-tolerant !!!



# Main expected features

- **Integration with CERN control system**
- **Scalability**
  - Scale the computation load across several hosts (OpenStack VMs)
  - Distributed storage for temporary results
- **Merging events and numerical data analysis**
  - Predictive trending
  - Temporal reasoning (CEP)
  - Statistical Analysis
- **Possibility to prototype additional plug-ins and algorithms**
  - Agree on a general API for new algorithm definition and integration
  - Integration with ‘R’
  - Data analysis flow definition in building blocks
- **Reporting**
  - Graphical visualization of huge list of signals/results
  - Interface to provide feedback to external systems (i.e.: WinCC OA)
- **Conversion into a Service**
  - On-line mode for continuous control system monitoring over custom time-windows
  - Support for historical analysis
- **Data management**
  - Different sampling rates / gap
  - Custom data model (i.e.: temperature in K/C)
  - Custom data access (i.e.: vector vs sequence)

# CERN control system use-cases

Based on real examples

# Use-cases classification

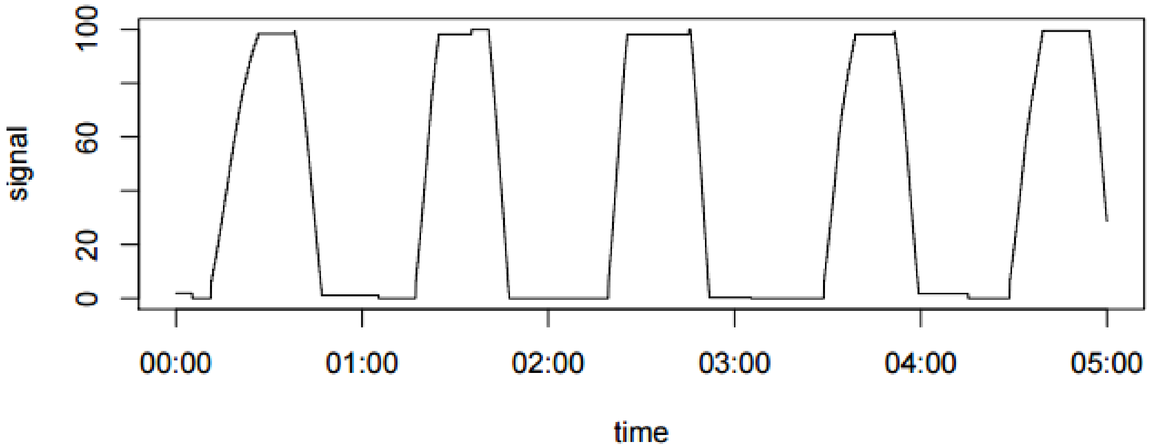
- **Online monitoring**
  - Continuous service to analyse the system status and inform operators in case of fault detection
- **Fault diagnosis**
  - “Forensics” analysis of system faults that have already happened in the past. In some cases root-cause analysis
- **Engineering design**
  - Analysis of historical data to draw conclusions about system behaviours which could be helpful to improve / optimize the system under analysis

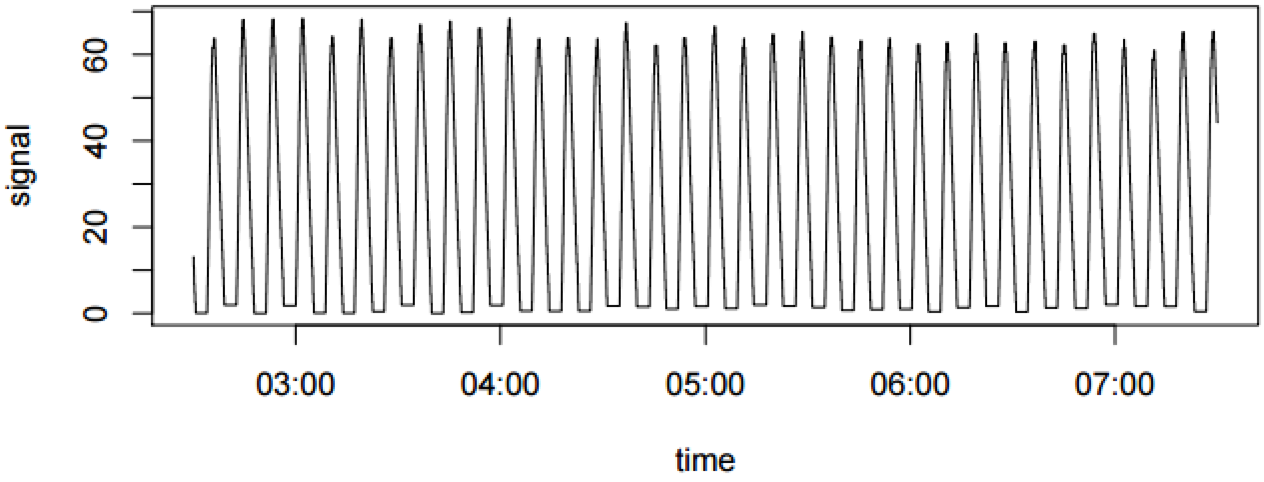
# Online monitoring

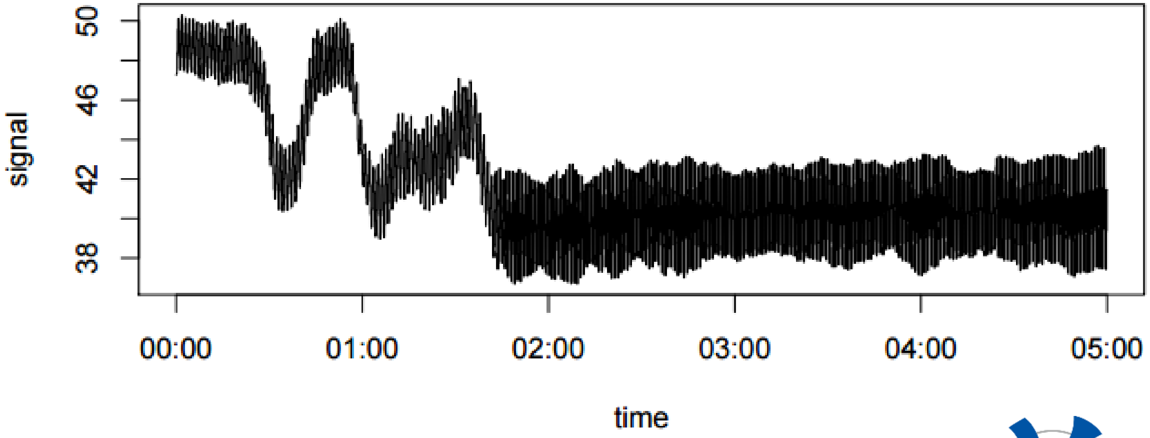
- Oscillation analysis in cryogenics valves (CRYO, CV)
- Online analysis of control alarms (MOON)
- Expert system on monitoring events (CMS)
- LHC dashboard (CRYO)

# Oscillation analysis for cryogenics valves

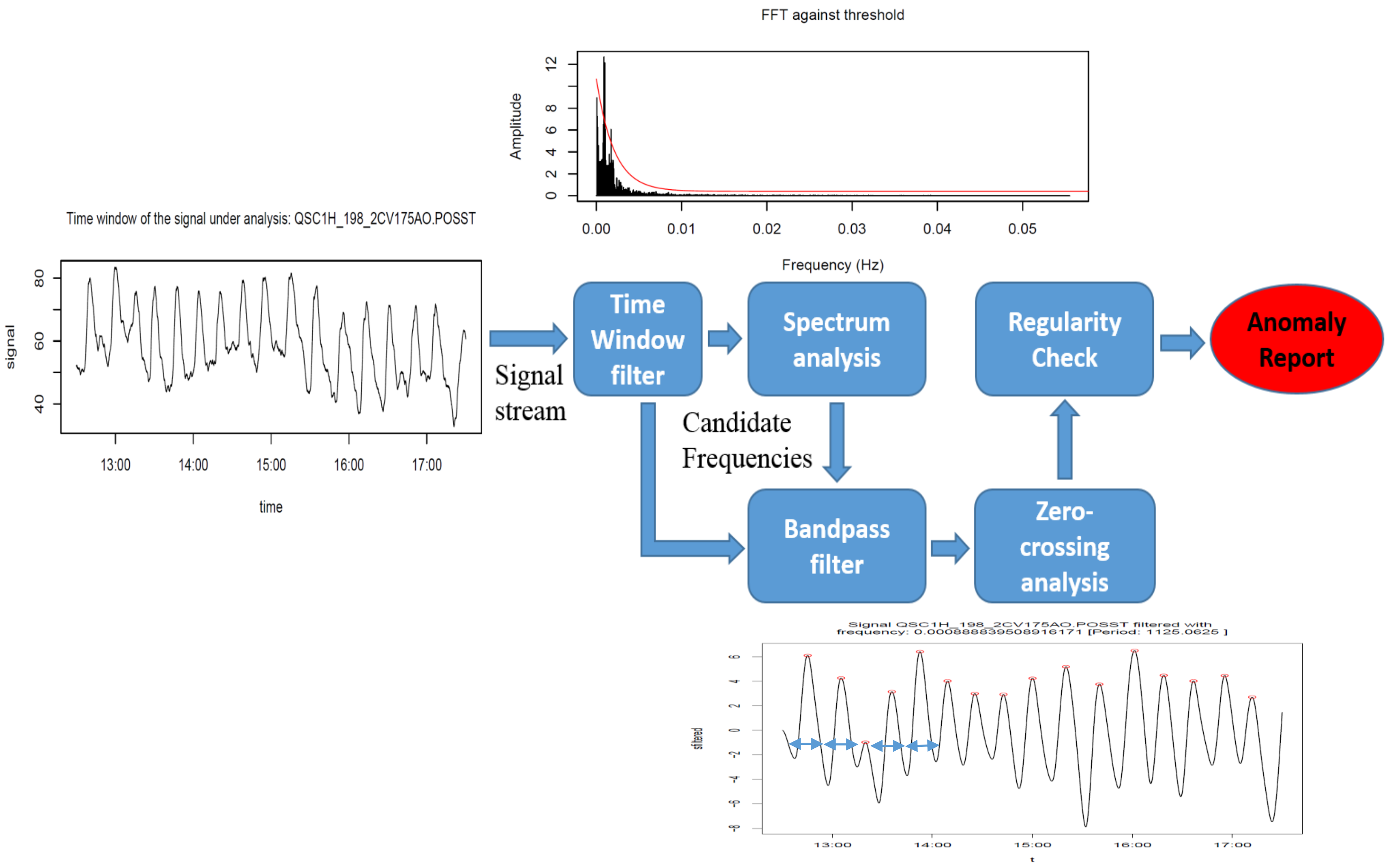
- Goal: detect whenever a signal is oscillating  in any anomalous way. Impact on:
  - Control system stability
  - Increased communication load
  - Maintenance (use of actuators)
  - Safety
  - Performances (Physic time)



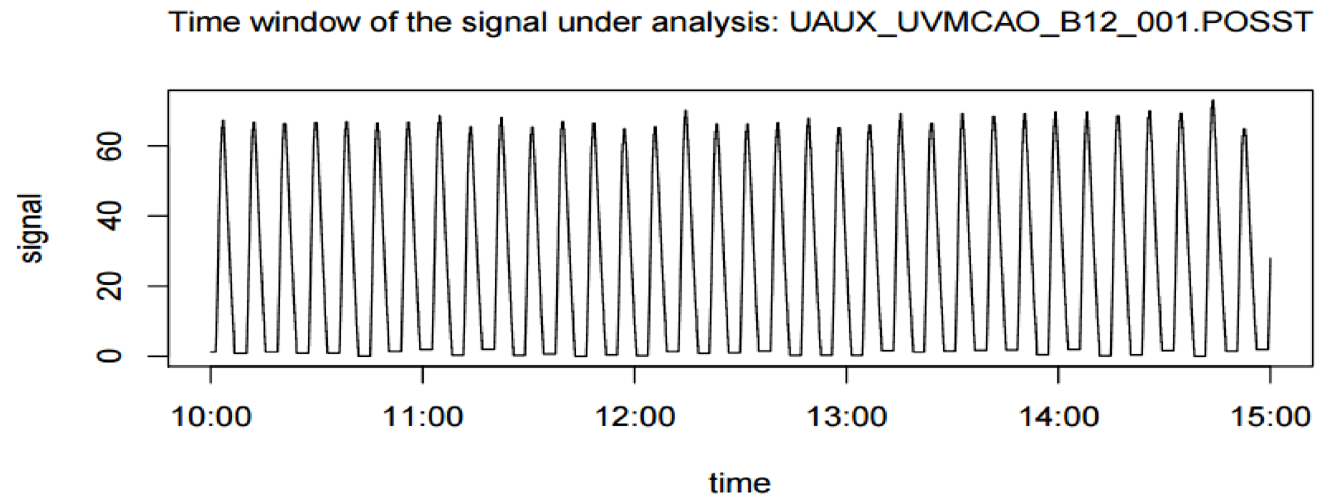
# Oscillation analysis flow

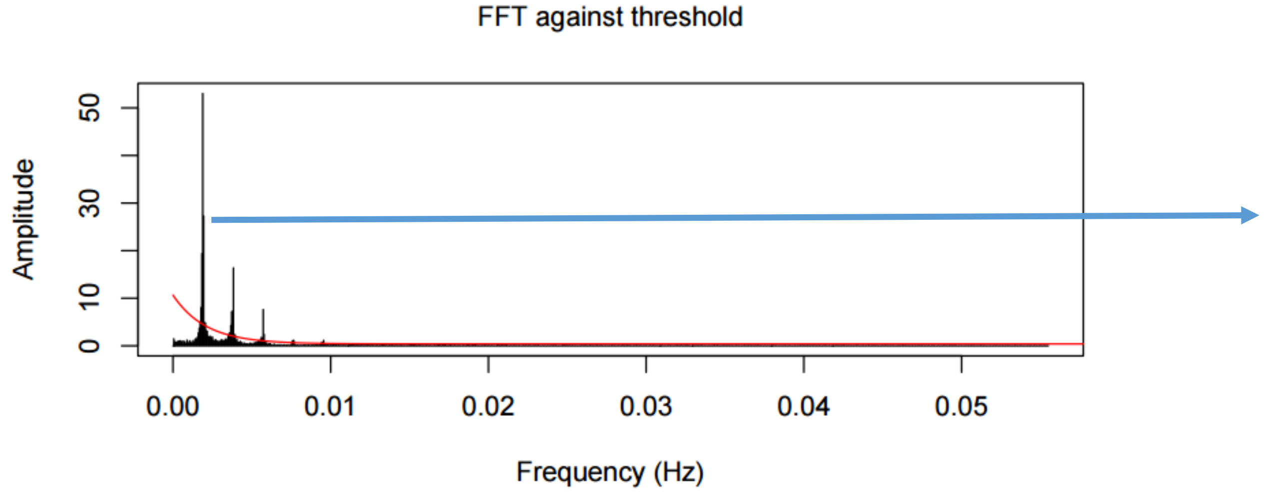


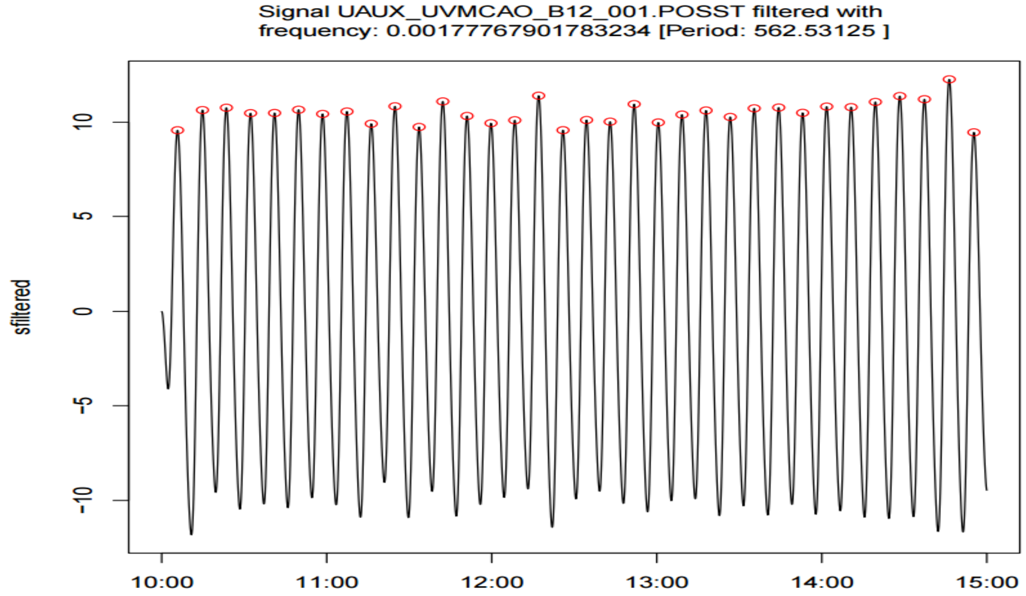

## On-line analysis:

- > 3000 sensors
- Continuous analysis
- Frequency: 24h

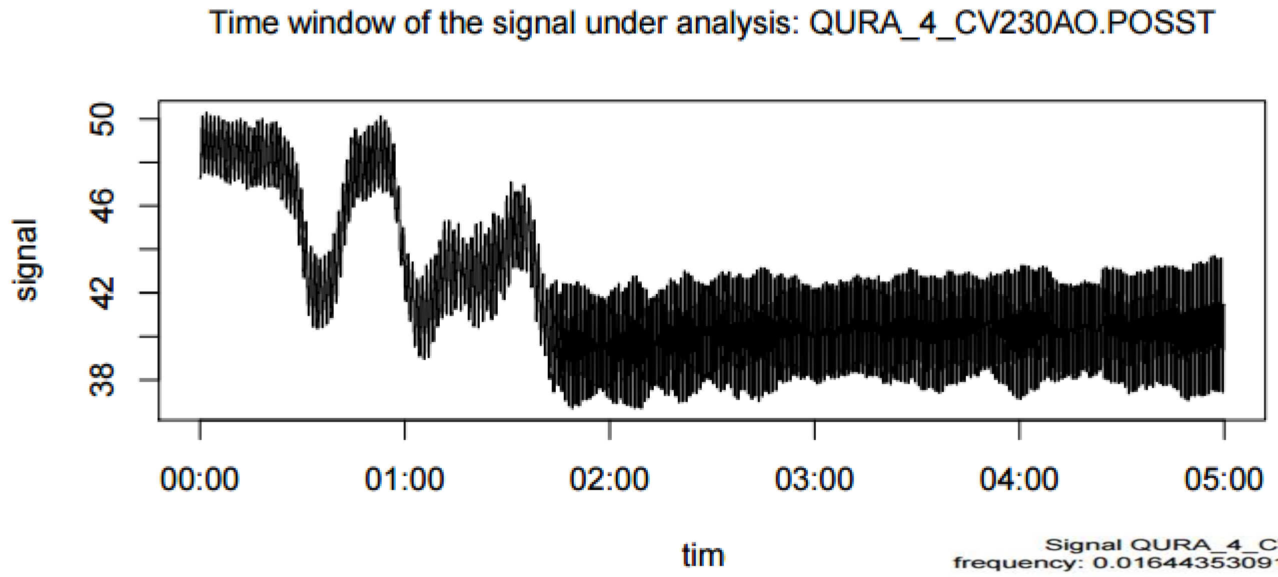
# Oscillation detection Ex#1

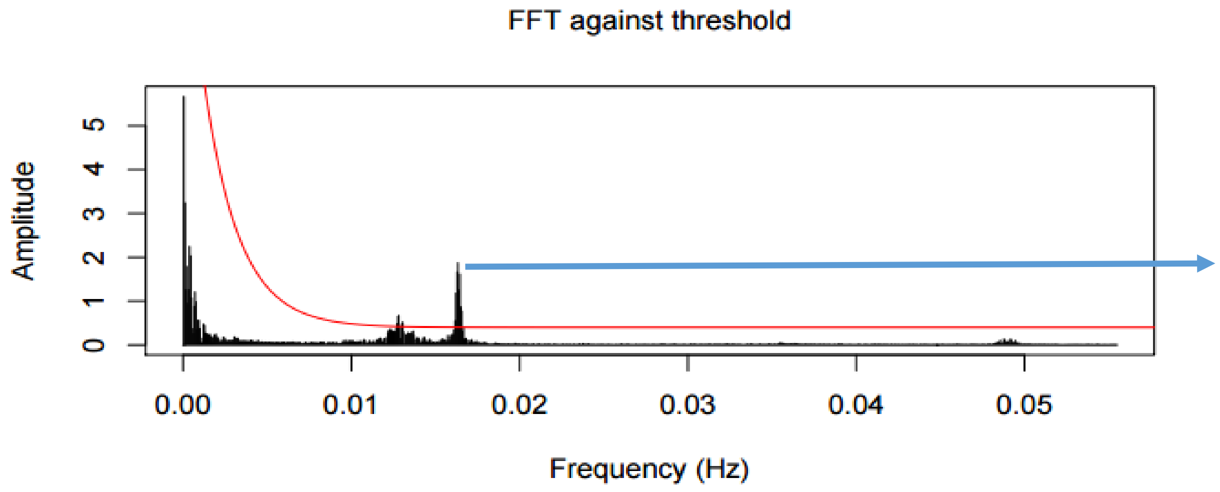


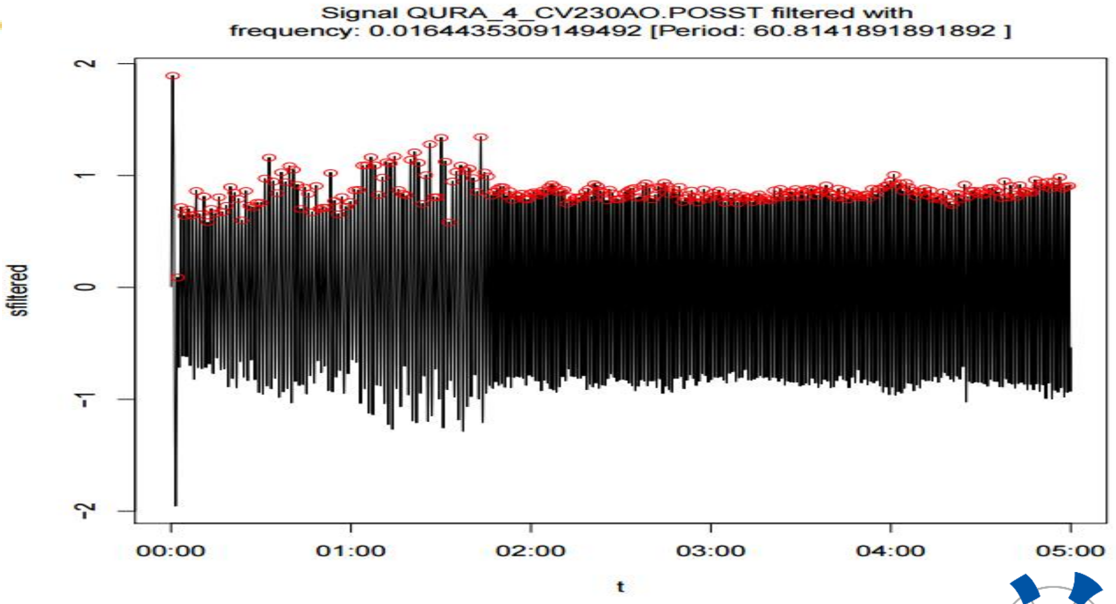



# Oscillation detection Ex#2

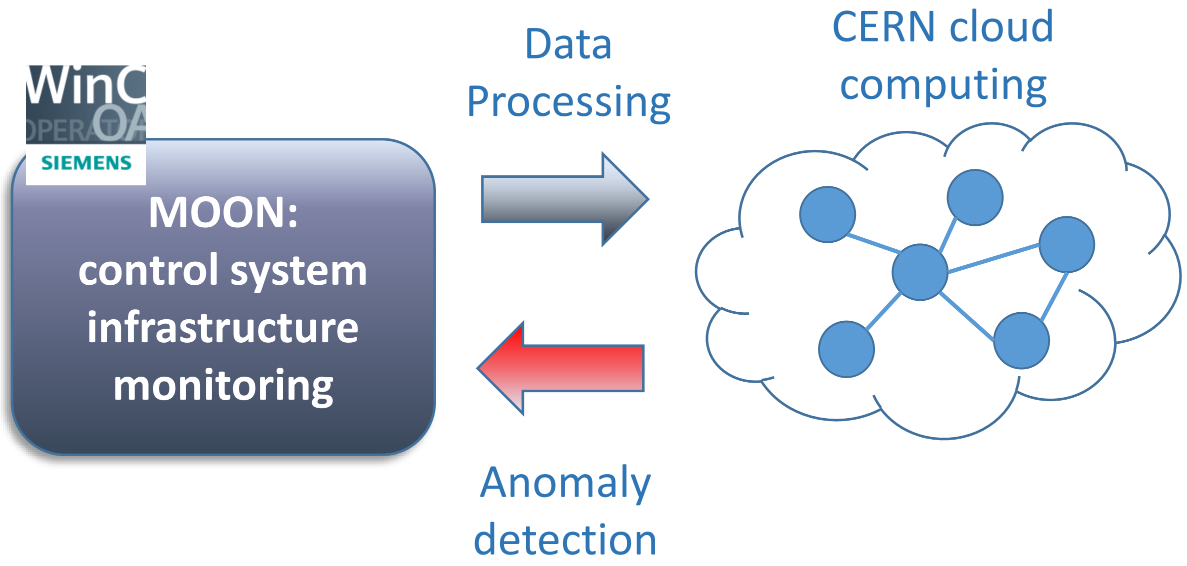


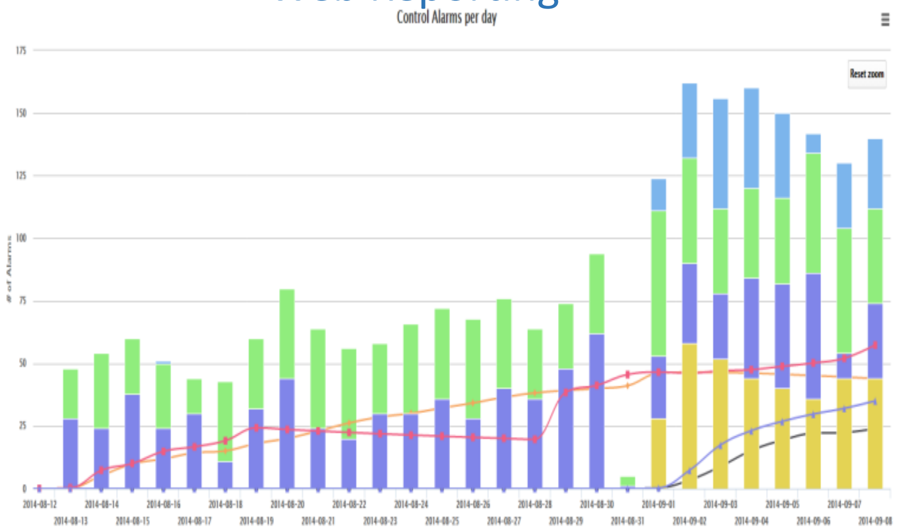



# Online analysis of control alarms

- Alarms analysis to detect anomalies or abnormal behaviors for thousands of devices
- Parallelization using the CERN OpenStack cluster
- Threshold learning algorithm and outliers detection techniques:
- Graphical visualization of the anomalies/outliers



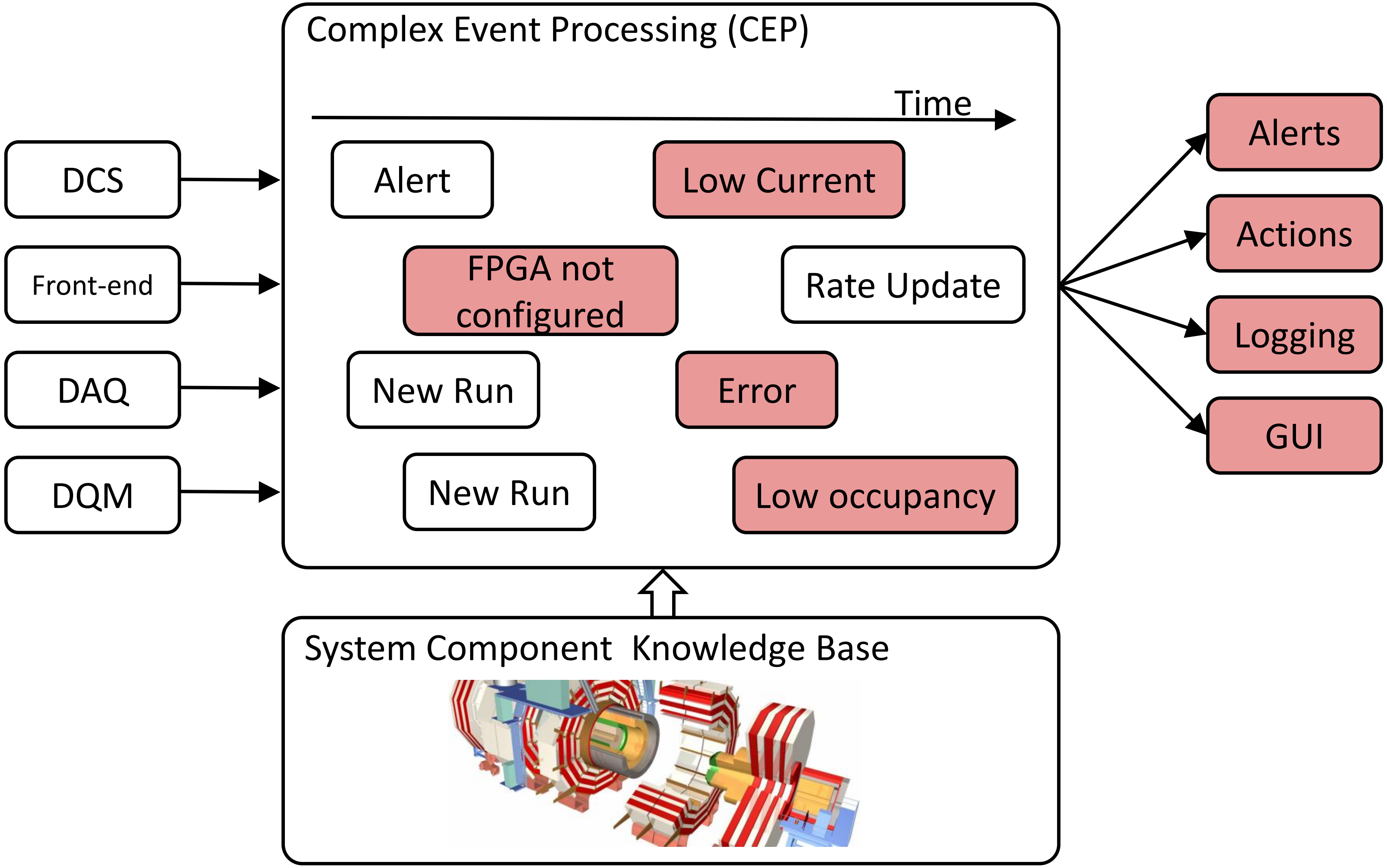

Web Reporting






CERN openlab

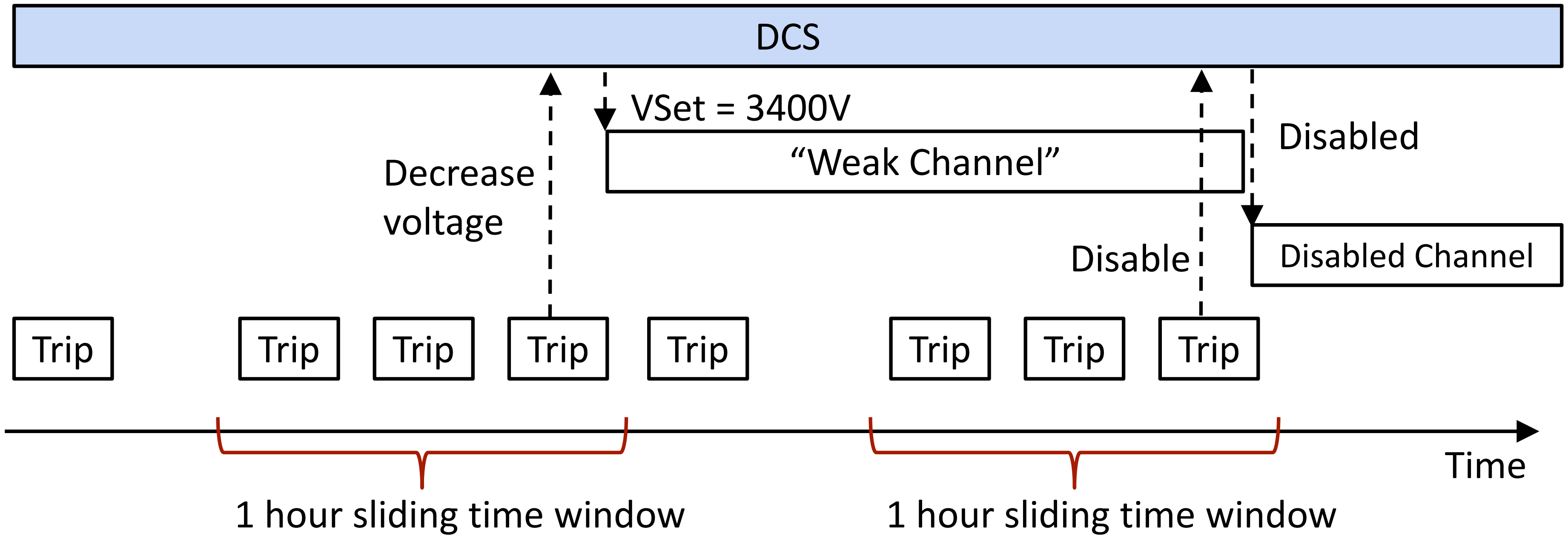
# CMS CSC Expert System
(Evaldas Juska )

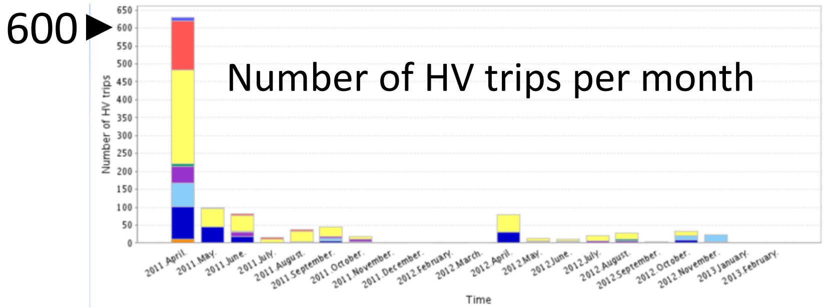


# CMS CSC Expert System Example

(Evaldas Juska )

HV trip recovery and channel management (10000 channels)

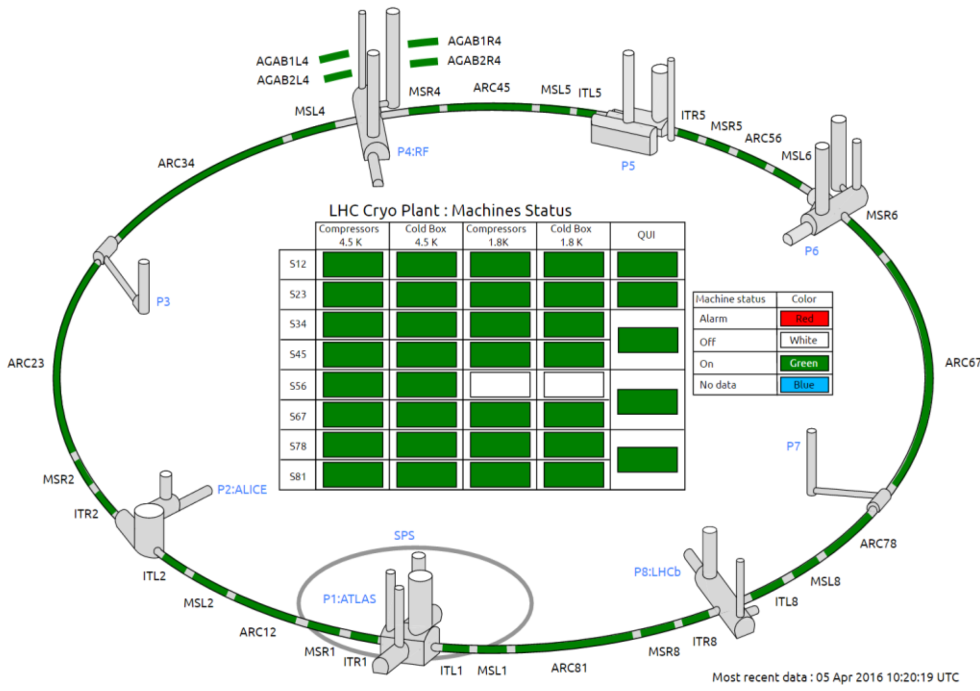



- Number of HV trips decreased dramatically in 1 month of operation
  - Minimal loss in efficiency
  - Only small fraction of channels were labeled "weak" or disabled

# Control data visualization

## LHC Dashboard (Brice Copy)






## Features:

- Faster data extraction
- Database query protection mechanism

- Data distribution
- User friendly navigation
- Easy to access

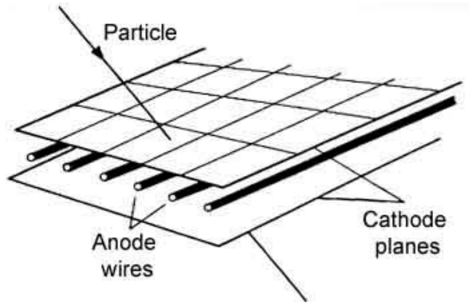
- Multiple sources
- Multiple output formats (Charts, Table, Text)
- Similar to the Atlas DDV (DCS Data Viewer)

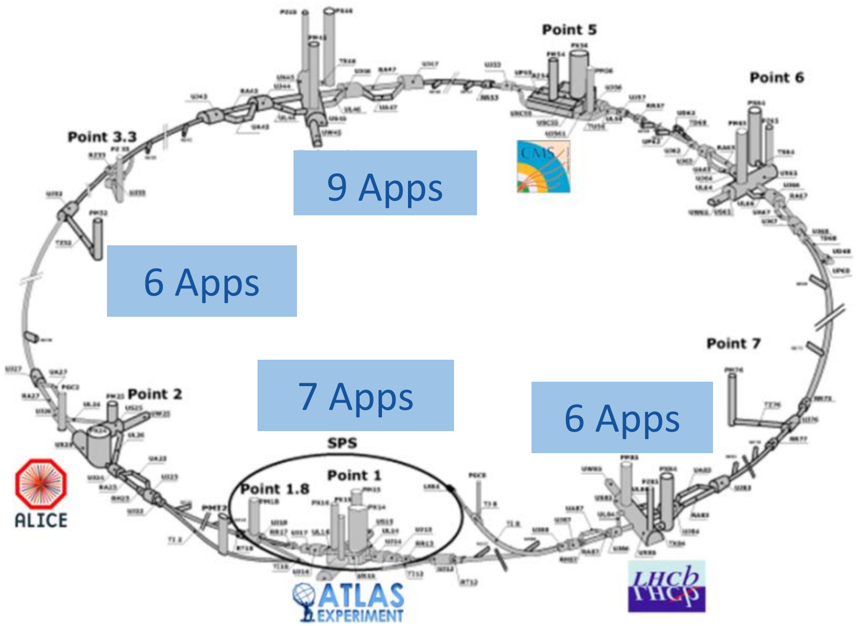
# Fault diagnosis (off-line)

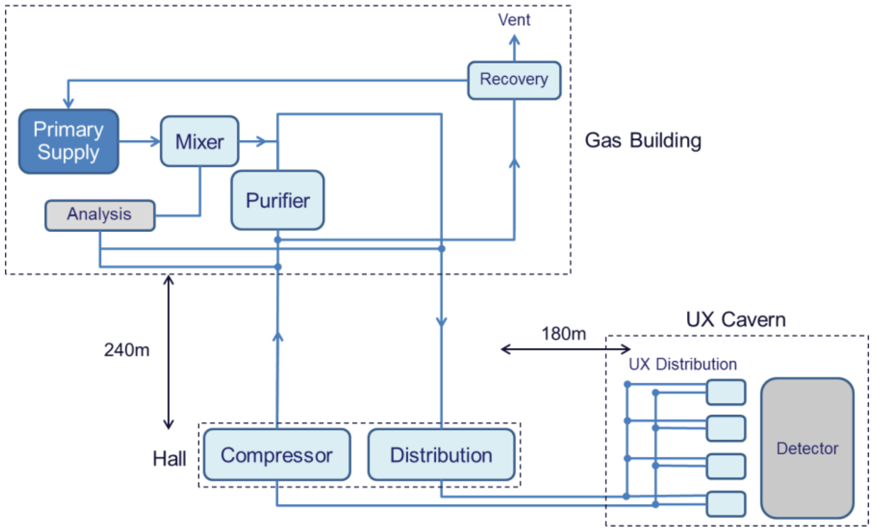
- Root cause analysis for control alarms avalanches (GAS system)

## An example:

# Gas control system @CERN

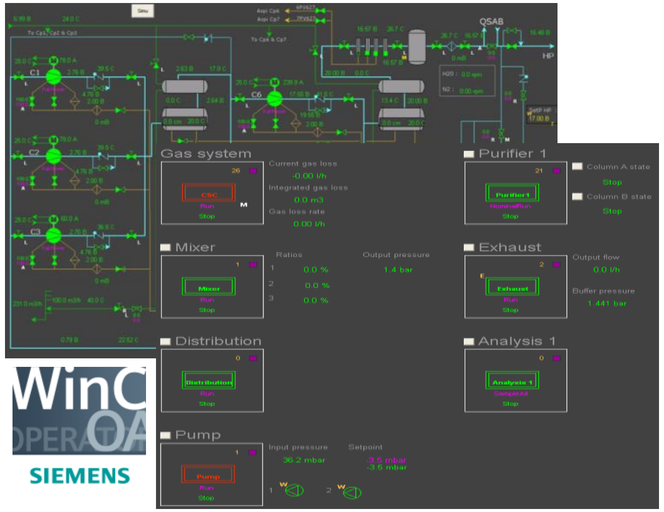


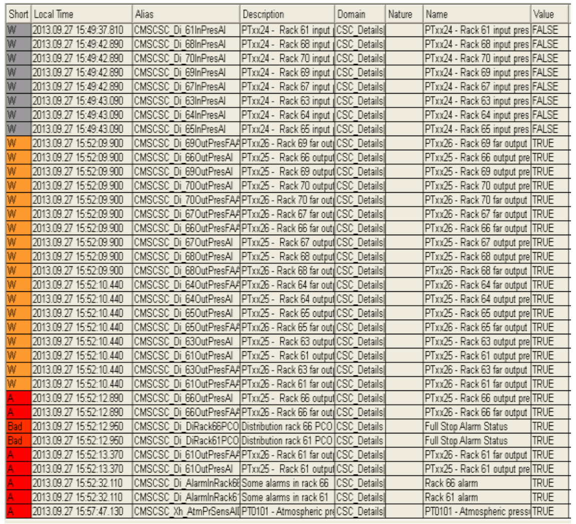




- 28 gas systems deployed around LHC
- 4 Data Server, 51 PLCs (29 for process control, 22 for flow-cells handling)
- Essential for particle detection
- Reliability and stability are critical
  - Any variation in the gas composition can affect the accuracy of the acquired data
- ~18 000 physical sensors / actuators

# Alarm flooding problem




Domino effect



| Short | Local Time | Alias | Description | Domain | Nature | Name | Value |
|---|---|---|---|---|---|---|---|
| W | 2013.09.27 15:49:37.810 | CMSCSC_Di_61InPresAl | PTxx24 - Rack 61 input | CSC_Details | | PTxx24 - Rack 61 input pres | FALSE |
| W | 2013.09.27 15:49:42.890 | CMSCSC_Di_68InPresAl | PTxx24 - Rack 68 input | CSC_Details | | PTxx24 - Rack 68 input pres | FALSE |
| W | 2013.09.27 15:49:42.890 | CMSCSC_Di_70InPresAl | PTxx24 - Rack 70 input | CSC_Details | | PTxx24 - Rack 70 input pres | FALSE |
| W | 2013.09.27 15:49:42.890 | CMSCSC_Di_69InPresAl | PTxx24 - Rack 69 input | CSC_Details | | PTxx24 - Rack 69 input pres | FALSE |
| W | 2013.09.27 15:49:42.890 | CMSCSC_Di_67InPresAl | PTxx24 - Rack 67 input | CSC_Details | | PTxx24 - Rack 67 input pres | FALSE |
| W | 2013.09.27 15:49:43.090 | CMSCSC_Di_63InPresAl | PTxx24 - Rack 63 input | CSC_Details | | PTxx24 - Rack 63 input pres | FALSE |
| W | 2013.09.27 15:49:43.090 | CMSCSC_Di_64InPresAl | PTxx24 - Rack 64 input | CSC_Details | | PTxx24 - Rack 64 input pres | FALSE |
| W | 2013.09.27 15:49:43.090 | CMSCSC_Di_65InPresAl | PTxx24 - Rack 65 input | CSC_Details | | PTxx24 - Rack 65 input pres | FALSE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_69OutPresFA | PTxx26 - Rack 69 far out | CSC_Details | | PTxx26 - Rack 69 far output | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_66OutPresAl | PTxx25 - Rack 66 output | CSC_Details | | PTxx25 - Rack 66 output pre | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_69OutPresAl | PTxx25 - Rack 69 output | CSC_Details | | PTxx25 - Rack 69 output pre | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_70OutPresAl | PTxx25 - Rack 70 output | CSC_Details | | PTxx25 - Rack 70 output pre | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_70OutPresFA | PTxx26 - Rack 70 far out | CSC_Details | | PTxx26 - Rack 70 far output | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_67OutPresFA | PTxx26 - Rack 67 far out | CSC_Details | | PTxx26 - Rack 67 far output | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_66OutPresFA | PTxx26 - Rack 66 far out | CSC_Details | | PTxx26 - Rack 66 far output | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_67OutPresAl | PTxx25 - Rack 67 output | CSC_Details | | PTxx25 - Rack 67 output pre | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_68OutPresAl | PTxx25 - Rack 68 output | CSC_Details | | PTxx25 - Rack 68 output pre | TRUE |
| W | 2013.09.27 15:52:09.900 | CMSCSC_Di_68OutPresFA | PTxx26 - Rack 68 far out | CSC_Details | | PTxx26 - Rack 68 far output | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_64OutPresFA | PTxx26 - Rack 64 far out | CSC_Details | | PTxx26 - Rack 64 far output | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_64OutPresAl | PTxx25 - Rack 64 output | CSC_Details | | PTxx25 - Rack 64 output pre | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_65OutPresAl | PTxx25 - Rack 65 output | CSC_Details | | PTxx25 - Rack 65 output pre | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_65OutPresFA | PTxx26 - Rack 65 far out | CSC_Details | | PTxx26 - Rack 65 far output | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_63OutPresAl | PTxx25 - Rack 63 output | CSC_Details | | PTxx25 - Rack 63 output pre | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_61OutPresAl | PTxx25 - Rack 61 output | CSC_Details | | PTxx25 - Rack 61 output pre | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_63OutPresFA | PTxx26 - Rack 63 far out | CSC_Details | | PTxx26 - Rack 63 far output | TRUE |
| W | 2013.09.27 15:52:10.440 | CMSCSC_Di_61OutPresFA | PTxx26 - Rack 61 far out | CSC_Details | | PTxx26 - Rack 61 far output | TRUE |
| A | 2013.09.27 15:52:12.890 | CMSCSC_Di_66OutPresAl | PTxx25 - Rack 66 output | CSC_Details | | PTxx25 - Rack 66 output pre | TRUE |
| A | 2013.09.27 15:52:12.890 | CMSCSC_Di_66OutPresFA | PTxx26 - Rack 66 far out | CSC_Details | | PTxx26 - Rack 66 far output | TRUE |
| Bad | 2013.09.27 15:52:12.950 | CMSCSC_Di_DiRack66PCO | Distribution rack 66 PCO | CSC_Details | | Full Stop Alarm Status | TRUE |
| Bad | 2013.09.27 15:52:12.950 | CMSCSC_Di_DiRack61PCO | Distribution rack 61 PCO | CSC_Details | | Full Stop Alarm Status | TRUE |
| A | 2013.09.27 15:52:13.370 | CMSCSC_Di_61OutPresFA | PTxx26 - Rack 61 far out | CSC_Details | | PTxx26 - Rack 61 far output | TRUE |
| A | 2013.09.27 15:52:13.370 | CMSCSC_Di_61OutPresAl | PTxx25 - Rack 61 output | CSC_Details | | PTxx25 - Rack 61 output pre | TRUE |
| A | 2013.09.27 15:52:32.110 | CMSCSC_Di_AlarmInRack6 | Some alarms in rack 66 | CSC_Details | | Rack 66 alarm | TRUE |
| A | 2013.09.27 15:52:32.110 | CMSCSC_Di_AlarmInRack61 | Some alarms in rack 61 | CSC_Details | | Rack 61 alarm | TRUE |
| A | 2013.09.27 15:57:47.130 | CMSCSC_Xh_AtmPrSensAl | PT0101 - Atmospheric pr | CSC_Details | | PT0101 - Atmospheric press | TRUE |

⊗ Fault in the distribution system

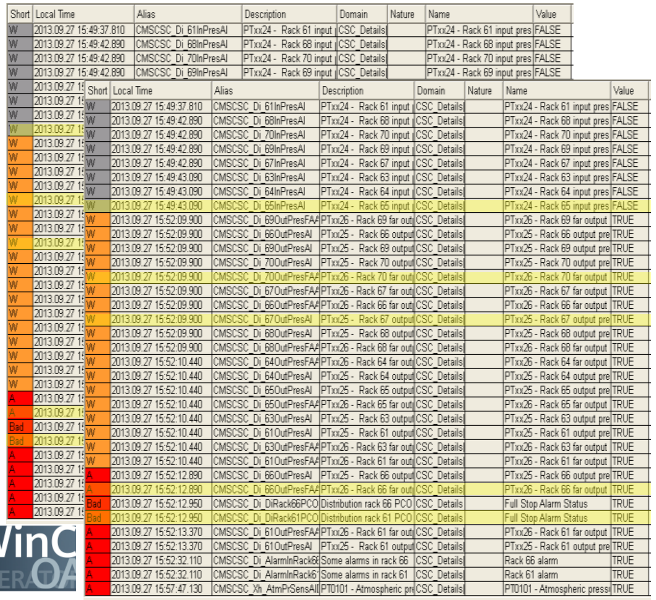
Alarms flooding

- **Diagnosing a fault is complex:** it may take weeks!
  - Alarms flooding: a single fault can generate up to a thousand of events
  - Number of different sequences:
    
    ~6x$10^{297}$ from: n!/(n-k)! , n=max seq. length, k=n/10
  - A single fault can stop the whole control process
  - The 1st alarm is not necessarily the most relevant for the diagnosis
  - Alarm generation depends on the system status



CERN openlab

# Events stream analysis



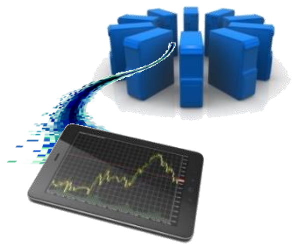

Event lists generated by the same fault

Data

## Analyze

Identify and detect fault / abnormal pattern for Diagnosis and Prognostics based on domain knowledge

## Learn



Provide experts with Root-cause and Gap Analysis using Rules and Patterns Mining

## Diagnose

Forecasts, Trends and Early-Warnings to increase Operating Hours

**Sequence:** Confidence: 83.3 % / Appearance count: 5

CMSCSC_Di_62PRegAI | Alarm Unacknowledged | Rising

**Sequence:** Confidence: 100.0 % / Appearance count: 6

CMSCSC_Di_62PRegAI | Alarm Unacknowledged | Rising
CMSCSC_Di_62PRegAI | Position Status (HH-LL) | Rising

CMSCSC_Di_YC60995 | Auto Off/Close Request Status | Falling
CMSCSC_Di_YC60995 | Auto On/Open Request Status | Rising
CMSCSC_Di_YC60995 | Off/Closed Status | Falling
CMSCSC_Di_YC60995 | On/Opened Status | Rising
CMSCSC_Di_YC60995 | Output Order Value Status | Rising

**Pattern** AA B A A B → **Alarm**

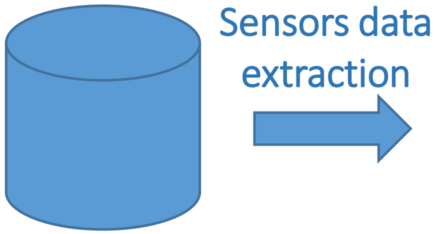
X T C D F A A E D N D B **K** D F A A B **K** D

# Anomaly detection by sensors data mining

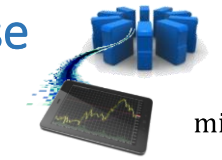
## Goal: detect abnormal/ unforeseen system behaviours

**LHC Logging Service**

Sensors data extraction

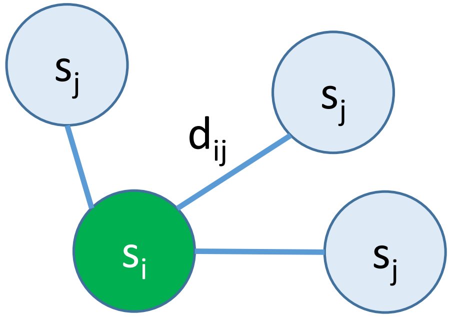


## Learning phase

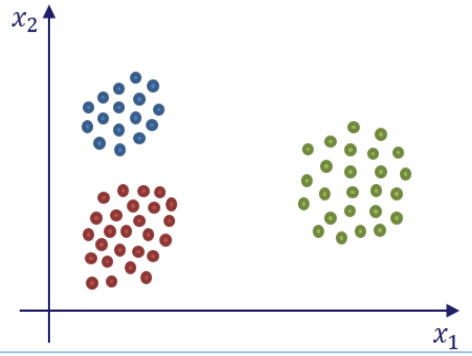


- Building a model based on historical data
- 3 different algorithms
  - Correlation index and KNN-graph
  - K-Mean clustering and probability model
  - Statistics expert-based model

$$E(d_i) = \sum_{j=1}^{k} d_{ij} * P(j|i)$$

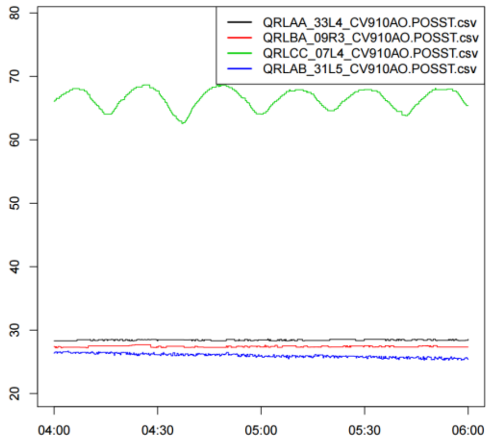


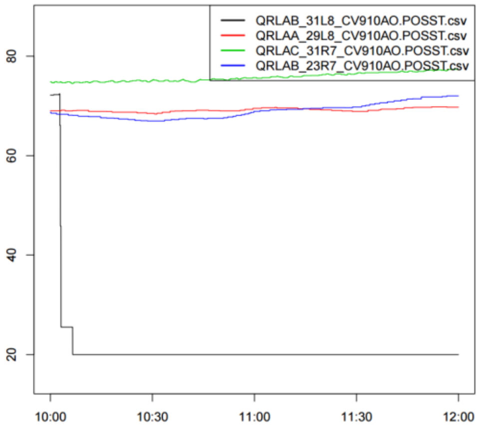

$$\min \frac{1}{C} \sum_{k=1}^{C} \max_{l \neq k} \left\{ \frac{\frac{\sum_i \| x_i - c_k \|}{N_k} + \frac{\sum_i \| x_i - c_l \|}{N_l}}{\| c_k - c_l \|} \right\}$$




## Anomaly detection

- Use the previous model to detect anomalies
- On-line analysis over a time window of 1 day
- Continuous analysis against thousands of sensors



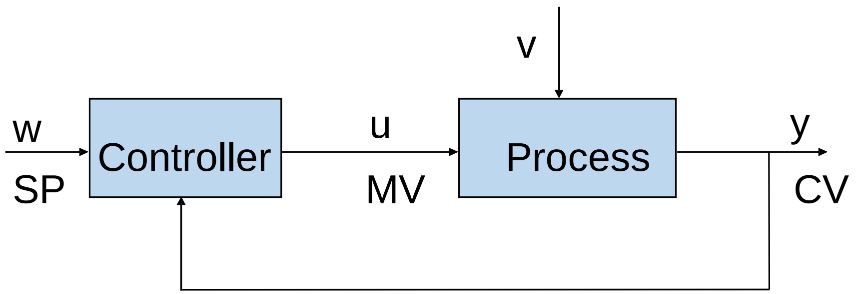



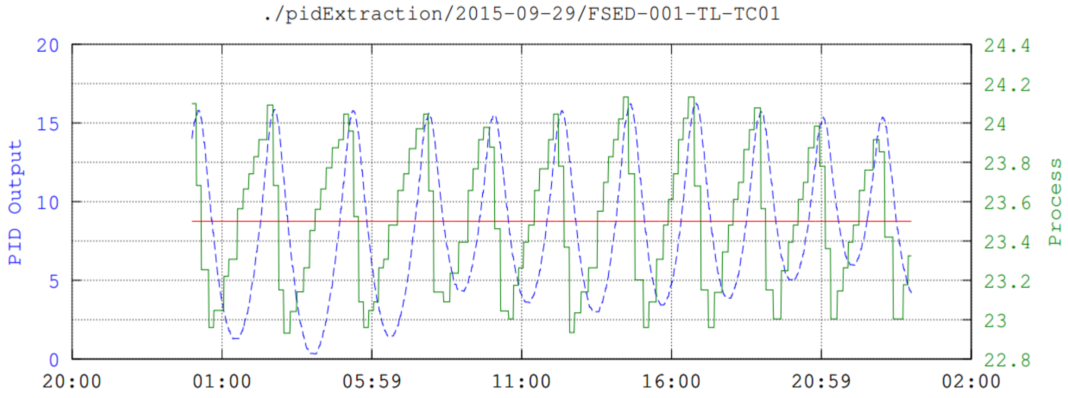
# Engineering design

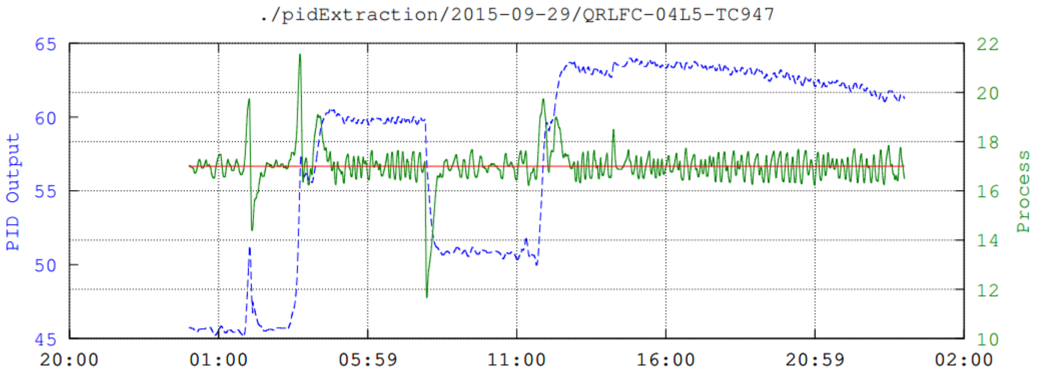
- PID supervision (CRYO, CV)
- Recommendation system for WinCC OA users (PSEN)

# Evaluation of PID supervision

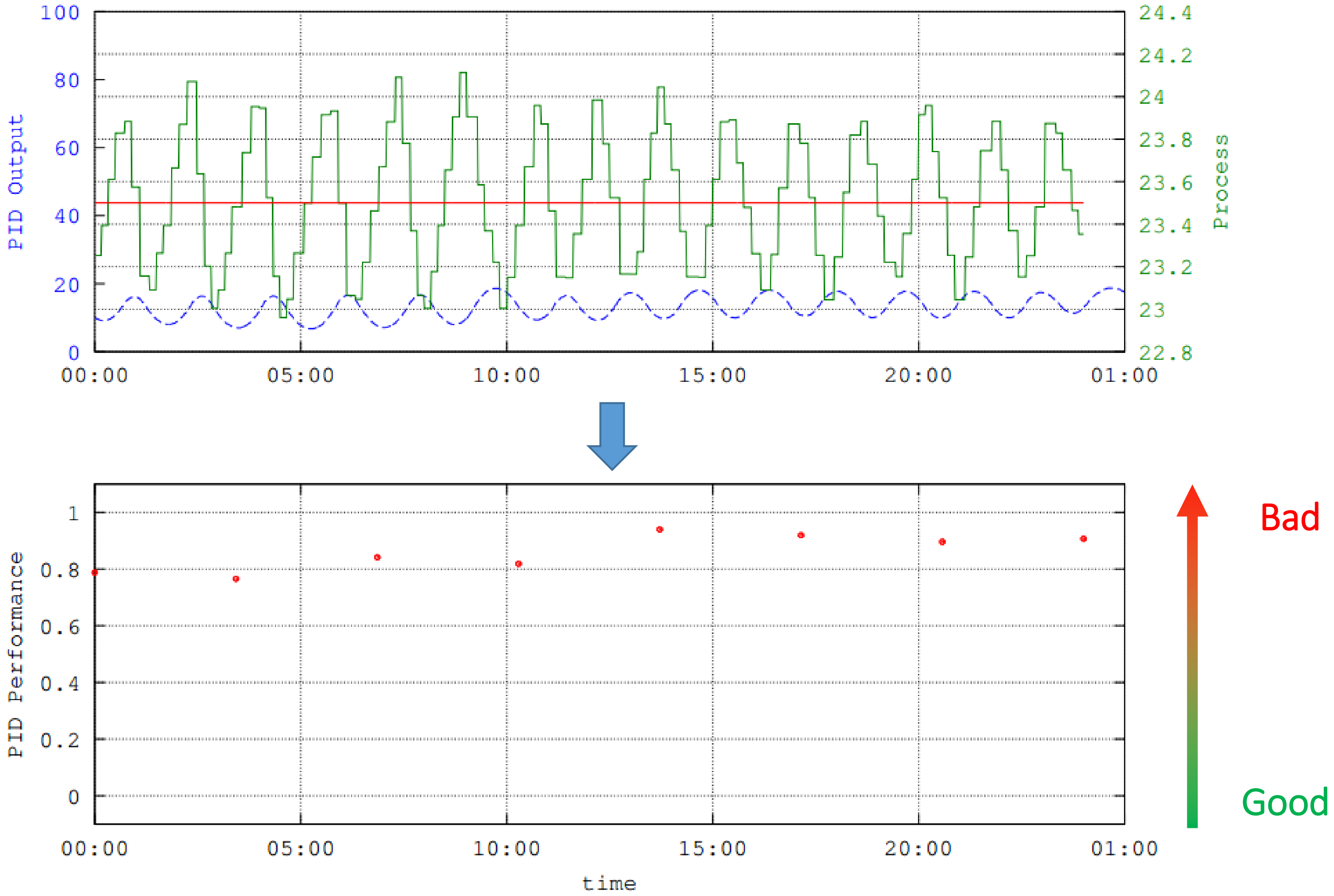
- In collaboration with the University of Valladolid
  *Based on: "Performance monitoring of industrial controllers based on the predictability of controller behaviour", R. Ghraizi, E. Martinez, C. de Prada*
- PID performance has an impact on:
  - Process security
  - Quality of physics
  - Maintenance (stress on the equipment)
- Issues:
  - Many sources of faults/malfunctions
  - System status dependency
  - External disturbances/factors
  - Bad tuning
  - Wrong controller type/structure
  - Slow degradation



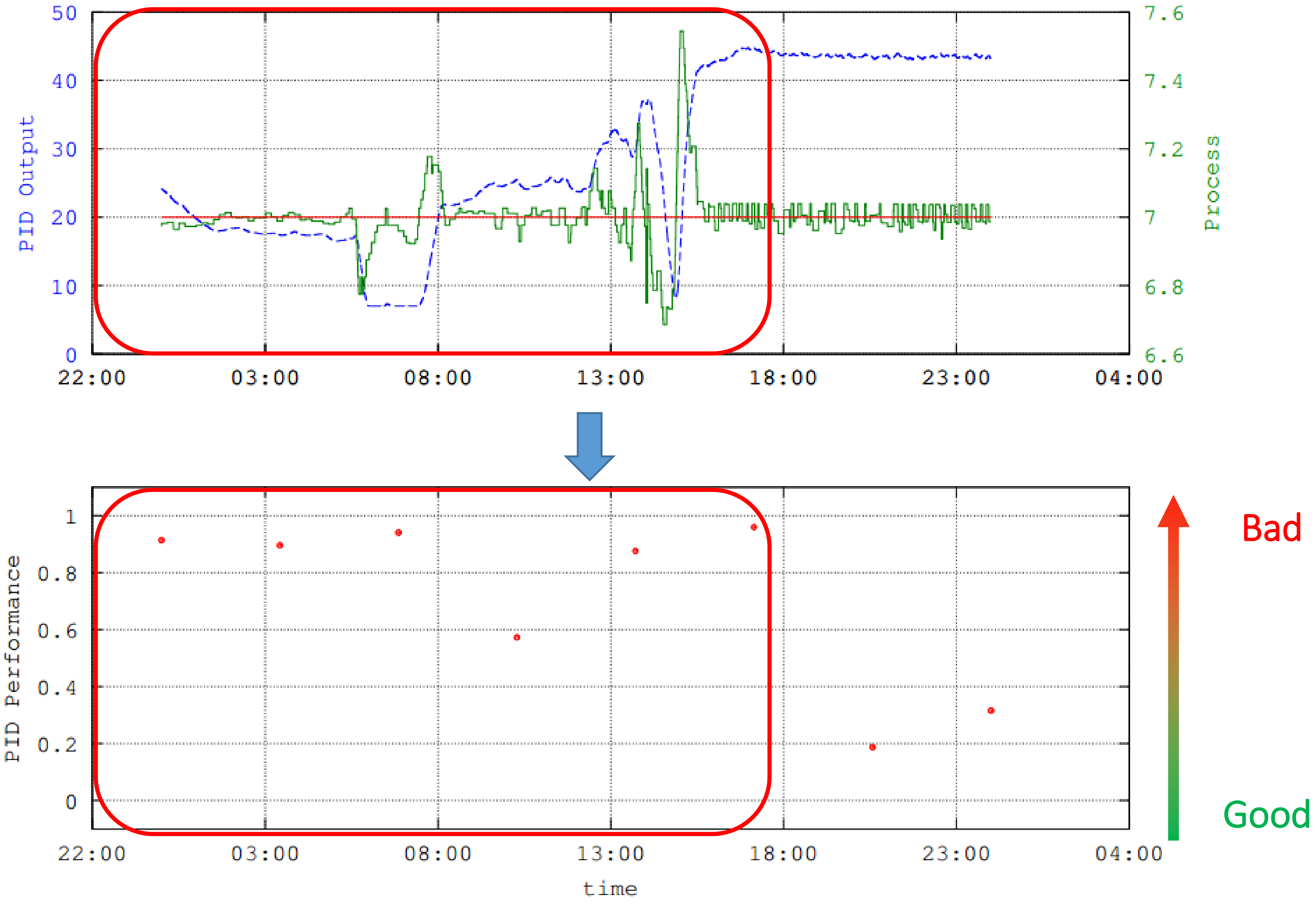



# PID supervision Ex#1



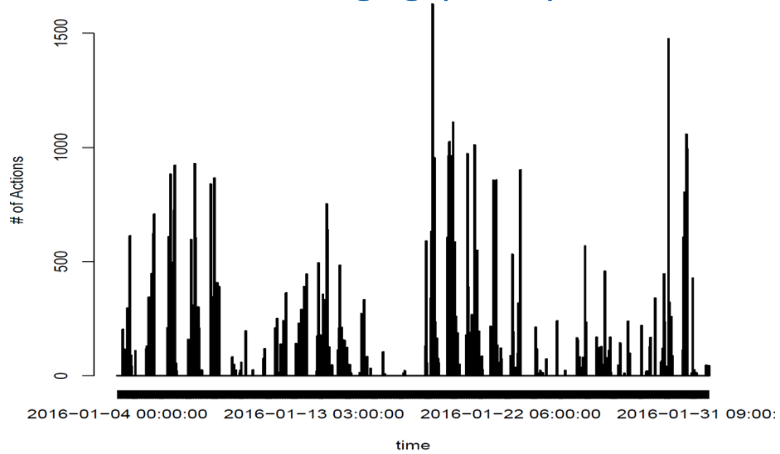
# PID supervision Ex#2



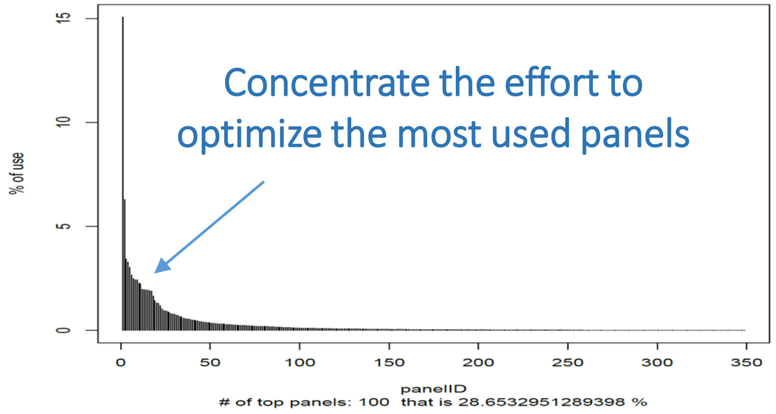
# Recommendation system for WinCC OA users

Users' actions extraction

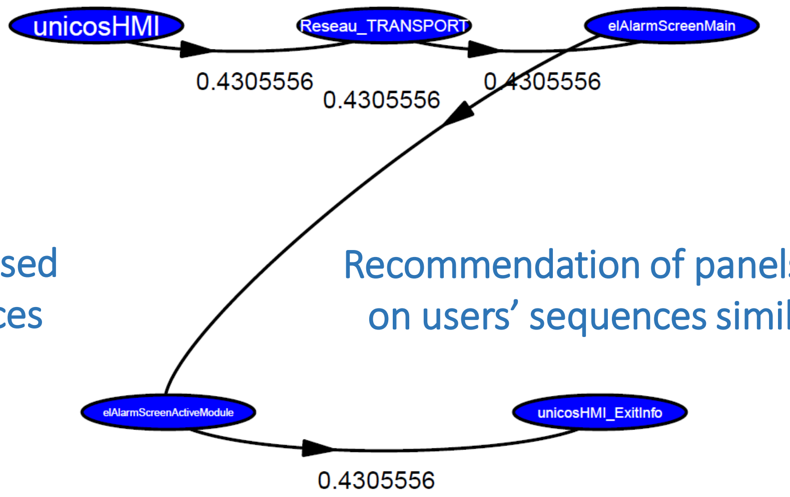
## Users' usage gap analysis




## Normalized distribution of panels usage



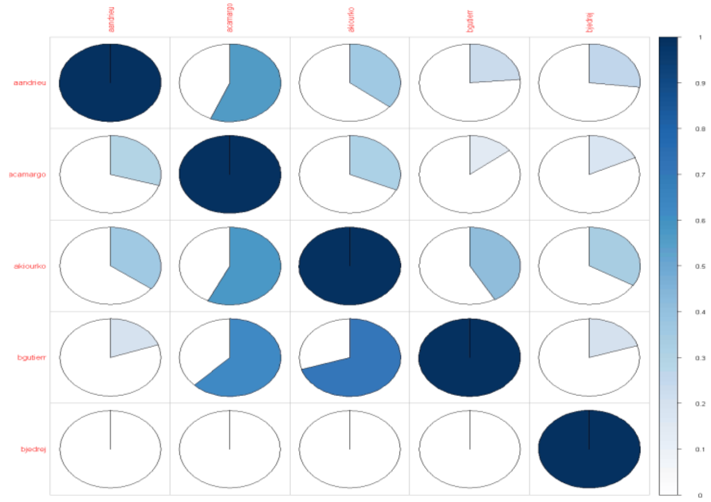

## Users' frequent sequences



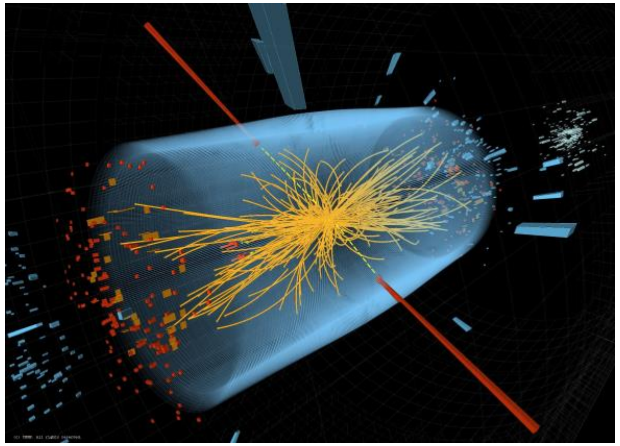

Recommendation of panels based on the specific users' sequences

Recommendation of panels based on users' sequences similarities

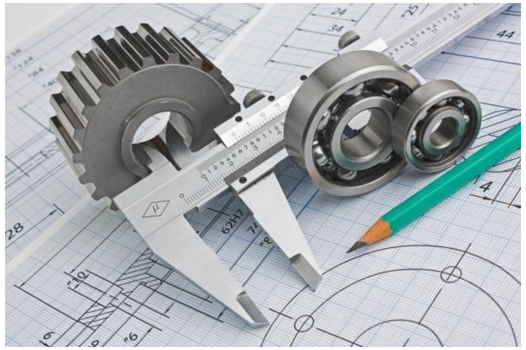
## Jaccard Sequences Similarity



# Data Analytics Benefits



- **Increased System Reliability**
  - Minimized forced outages
- **Complete data analysis**
  - Reduced service effort: weeks → hours
- **24/7 Expert Knowledge Availability**
  - One central knowledge base



### Operation support

- Big data visualization
- Forecast system status and take proper actions in time
- Prevent possible faults and system downtime

### Diagnosis support

- Identify root causes
- More accurate analysis
- Accelerate analysis From weeks to hours
- Identify hidden patterns

### Engineering support

- Evaluate and improve operational performance
- Increase reliability and efficiency by design
- Lead control system decisions

# Conclusions

- **Multiple data analytical activities in the experiments**

- **Data analytics brings an important added value to control systems**
  - A price to pay to integrate it into DCS

- **Many data analytical activities started**
  - in an uncoordinated manner
  - different technologies (ES, Storm, Esper, DroolsFusion…)
  - Effort to homogenize all the activities under a common analytic platform

# Use-cases: a partial list

- **Online monitoring**
  - Control System Health
  - Electrical power quality of service
  - Looking for heat in superconducting magnets
  - Oscillation in cryogenics valves
  - Discharge of superconducting magnets heaters
- **Faults diagnosis**
  - Anomalies in the process regulation
  - PLC anomalies
  - Data loss detection
  - Root-cause analysis for complex WinCC OA installations
  - Analysis of sensors functioning and data quality
  - Analysis of LHCb configuration management system
  - Analysis of OPC-CAN middleware
  - Data loss in LHCb DAQ
  - Analysis of electrical power cuts
  - Cryogenic system breakdowns
- **Engineering design**
  - Electrical consumption forecast
  - Efficiency of electric network
  - Predictive maintenance of control systems elements
  - Predictive maintenance for control disks storage
  - Vibration analysis
  - Efficiency of control process