Model Learning Algorithms for Anomaly Detection in CERN Control Systems

ICALEPCS, Barcelona 2017

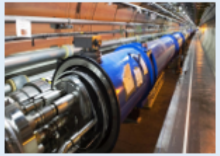
CERN, BE-Industrial Control & Safety Systems (ICS)

F. Tilaro, B. Bradu, M. Gonzalez-Berges, F. Varela, M. Roschchin

CERN: one of the world's largest automation systems

A multitude of Industrial Control Systems

Cooling & Ventilation



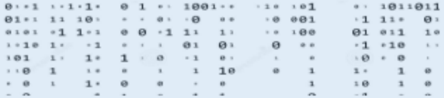
VACUUM

Cryogenics

GAS

Electric Grid

LHC Circuit, QPS, WIC,PIC, ...



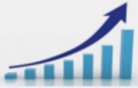
Storing +100 TB/year

Control Data analytics

- Specific industrial analytics algorithms
- Machine learning techniques to deal heterogeneous control systems

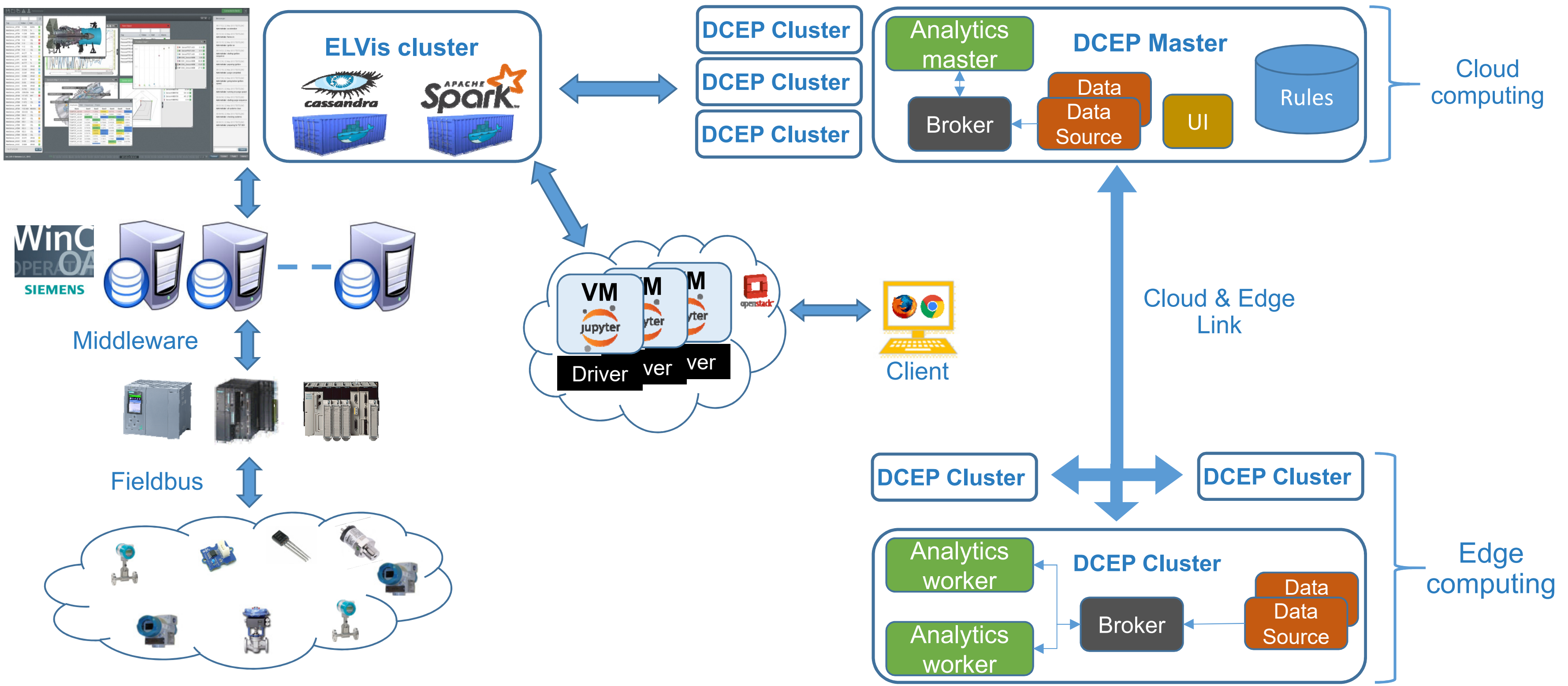
- Large-scale performance

- Take advantage of Big Data to improve:
  - ✓ Control system stability and efficiency
  - ✓ Reduce maintenance cost
  - ✓ Performances (even physic data quality)
  - ✓ Safety



Smart Data for Industrial Control Systems

Combining cloud and edge computing into a single framework



Anomaly detection in Cryogenics system

Main Goals:

- Detect whenever a sensor or actuator (e.g. valve) is behaving in any anomalous way
- Inform engineers and operators in order to take proper actions

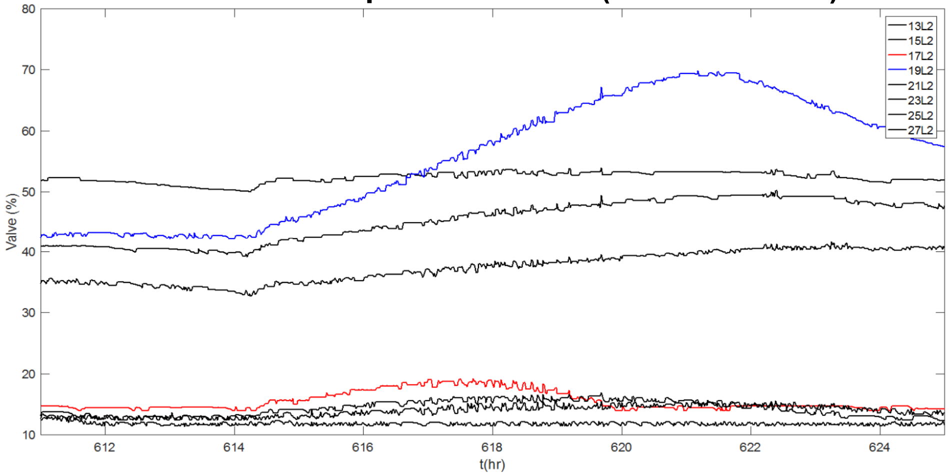
Multiple issues:

- sensors faults/glitches
- hardware failures/degradations,
  false measurements, wrong tuning/structure, …

Impacts:

- Control system stability
- Increased communication load
- Maintenance (use of actuators)
- Performances (even physic data quality)
- Safety

Valves CV910 positions in L2 (26th June 2017)

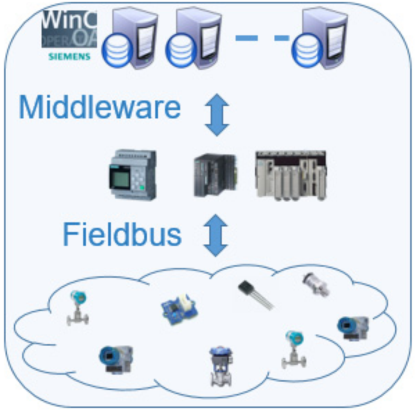


Anomaly detection by sensors data mining

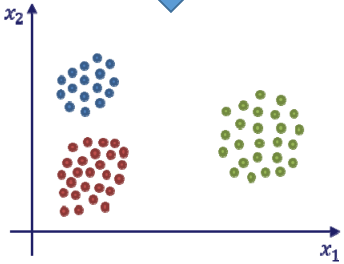
Build a model to detect abnormal or unforeseen system behaviours:
- Exploit historical data (~4GB/day for Cryo)
- Combine Machine Learning techniques with Experts' knowledge

Sensors/actuators measurements mining to learn:
- Logical relations
- Physical relations

Challenges:
- Different application domains/systems
- Different types of anomaly
- Noise and duration of an anomaly
- Not precise boundaries between normal/anomalous
- Mostly unsupervised training
- Dynamic system => dynamic model

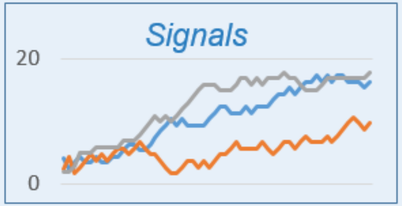


Model building



Different algorithms for anomaly detection

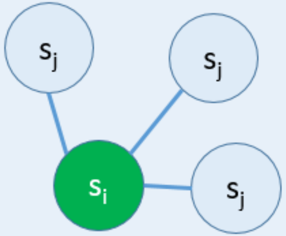
Correlation and K-NN



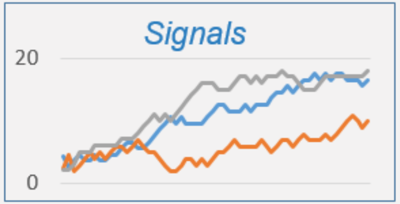
Pearson Correlation Index

$$d_{ij} = -\ln|a_{ij}|$$

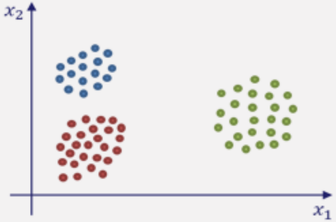
$$E(d_i) = \sum_{j=1}^{k} d_{ij} * P(j|i)$$



Stochastic clustering

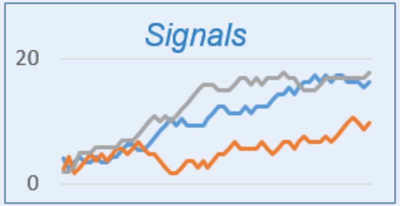


K-Mean & Davies-Bouldin index



$$\mathrm{P}(X_j = k) = \binom{n}{k} * P(g_j)^k * \left(1 - P(g_j)\right)^{n-k}$$

Experts' knowledge



- Statistical indexes:
  - Cluster and single signals derivatives
  - β = tolerance factor.

$$\begin{cases} x_i = x_{i-1} + \dfrac{dt}{\nabla Tf + dt}(x_i - x_{i-1}) \\ tw^{-1}\left(\sum_{tw} \nabla xi \ - \ \nabla X\right) > \beta, \\ X = \{xi \in Cluster\}, tw = time\ window \end{cases}$$

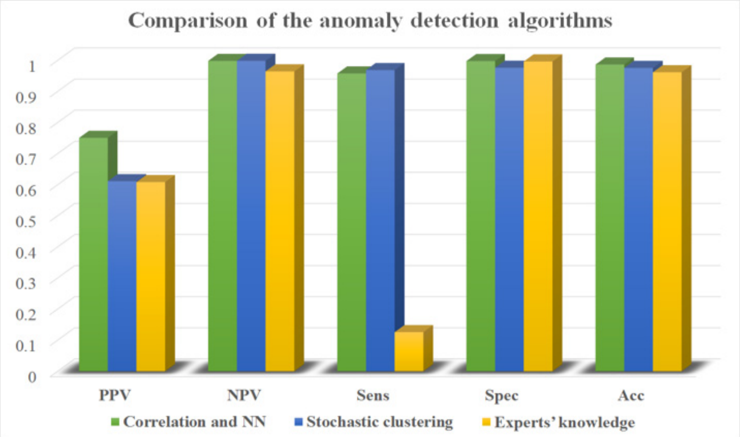
Algorithms comparison

Different fault types detected with a single configuration:

- Spike / Noise / Flipping / Offset / Drifting
- Wide range of fault frequency and amplitude

5 performance indexes:

- Positive/Negative predictive value (PPV/NPV), sensitivity (Sens), specificity (Spec) and accuracy (Acc)



Least robust: Stochastic clustering due to high false positives

Low Sensitivity and high Specificity: Experts' knowledge

Most robust and general algorithm: Correlation and K-NN

$$PPV = \frac{\sum TP}{\sum TP + \sum FP},\ NPV = \frac{\sum TN}{\sum FN + \sum TN},$$

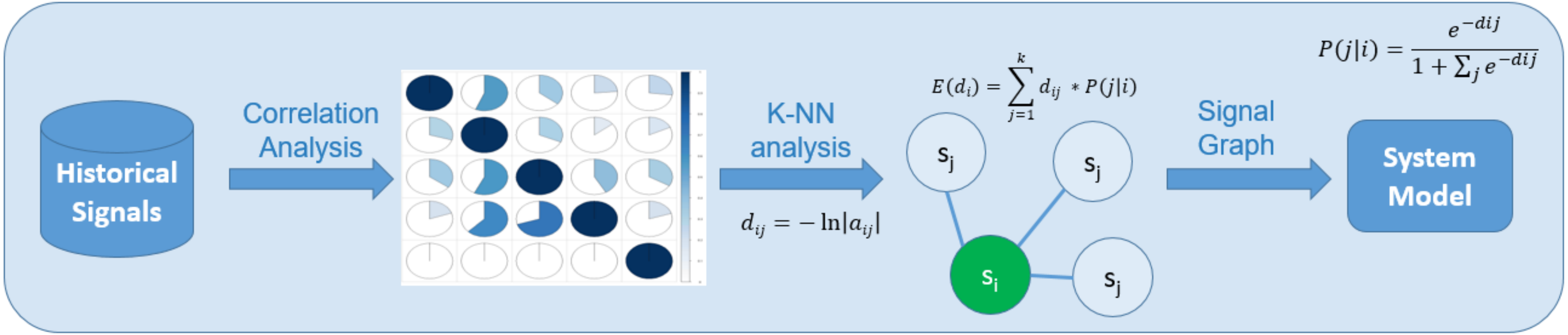
$$Sens = \frac{\sum TP}{\sum TP + \sum FN},\ Spec = \frac{\sum TN}{\sum FP + \sum TN},$$

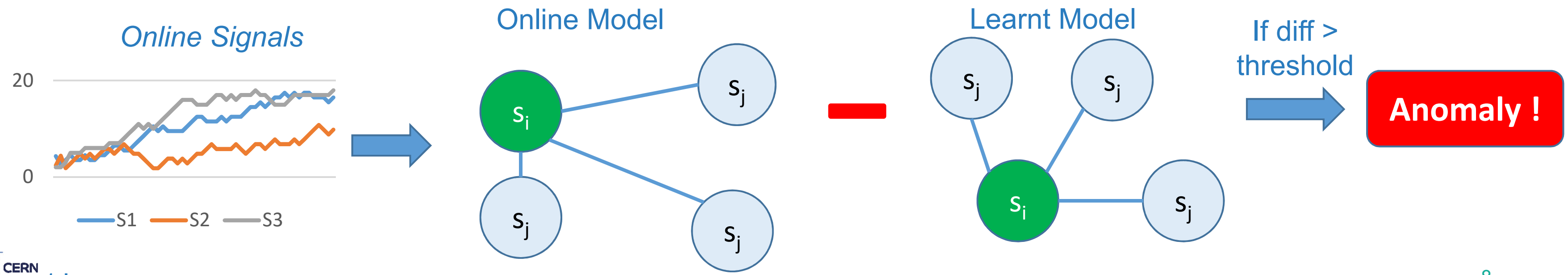
$$Acc = \frac{\sum TP + \sum TN}{\sum TP + \sum FP + \sum FN + \sum TN}$$

Signals correlation & conditional K-Nearest Neighbour

A two phase analytical process

Offline model learning analysis

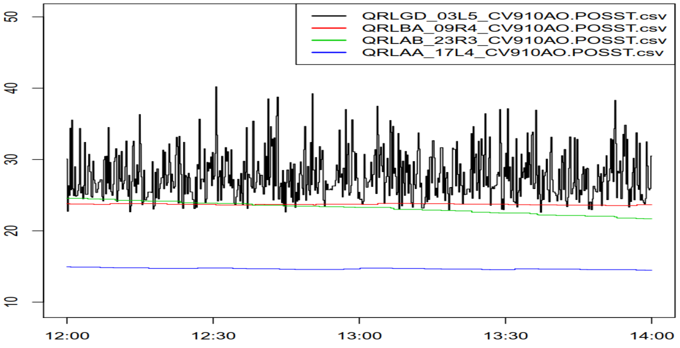




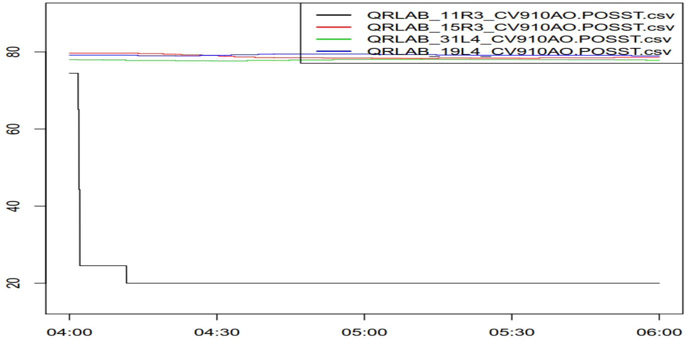
Signals Correlation and K-NN in action!

Detection of different anomalies in Cryogenics

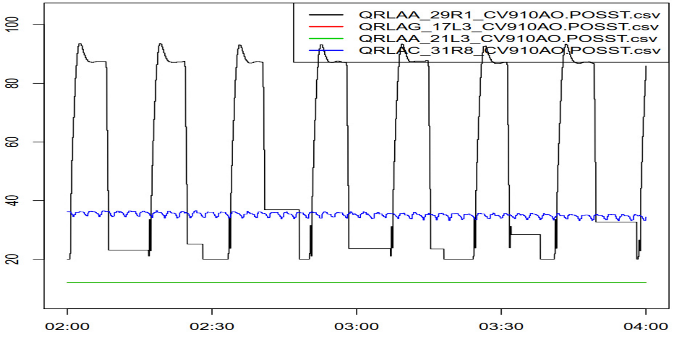
Flipping fault detection



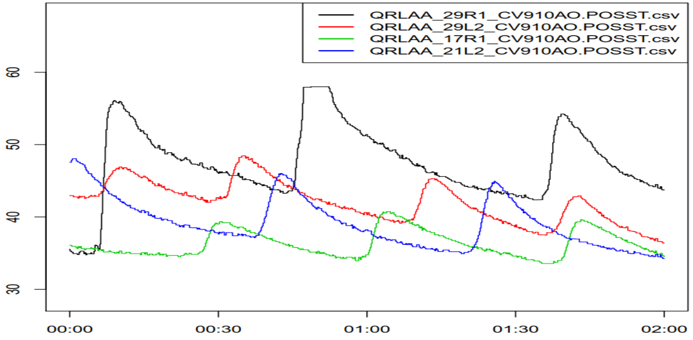
Signal offset detection



Oscillation detection



Faulty amplitude detection

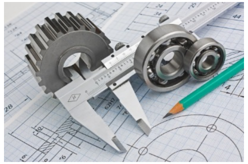


Conclusions and future work

Deployment of different analyses for anomaly detection for control systems

Benefits:

- Operation support
  - Online anomaly detection to take proper actions in time
    "Data Analytics Reporting Tool for CERN SCADA Systems",
    P. Seweryn, M. Gonzalez-Berges, B. Schofield, F. Tilaro, ICALPECS 2017
  - Prevent possible faults and system downtime
- Diagnosis support
  - Automatic monitoring of a multitude of sensors/actuators to identify anomalous patterns
  - Scale and accelerate analysis
- Engineering support
  - Evaluate and improve operational performance (PID tuning)
    "Automatic PID performance monitoring applied to LHC Cryogenics",
    B. Bradu, E. Blanco Vinuela, R. Marti, F. Tilaro, ICALPECS 2017
  - Increase reliability and efficiency by design



Possible future extensions:

- Threshold model and reinforcement learning
- Root-cause analysis

Thank You!

CERN BE-ICS

https://be-dep-ics.web.cern.ch/