Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

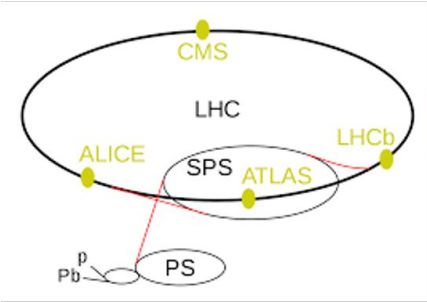
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

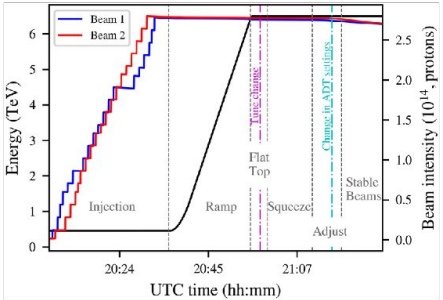
Anomaly detection for the online monitoring of the CMS Muon System (UC 2.2.1)

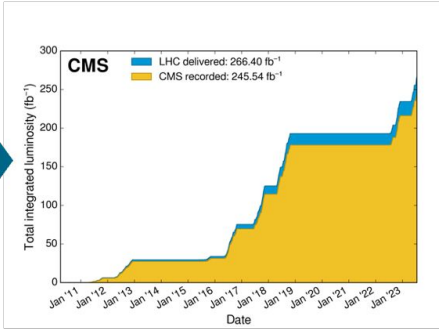
Marco Buonsante, **Federica Simone**, **Rosamaria Venditti** (INFN e Università/Politecnico di Bari)

Spoke2 WP2 Meeting, 21 Maggio 2024

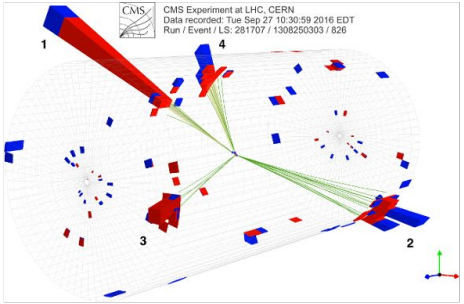
Data taking at CMS







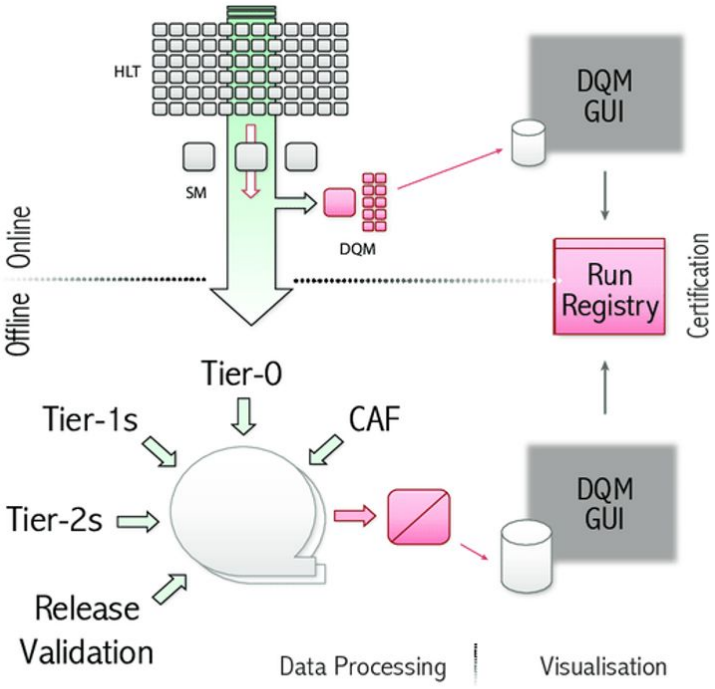
Recorded event



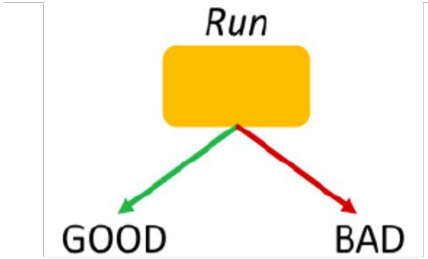
Lumisection (LS)

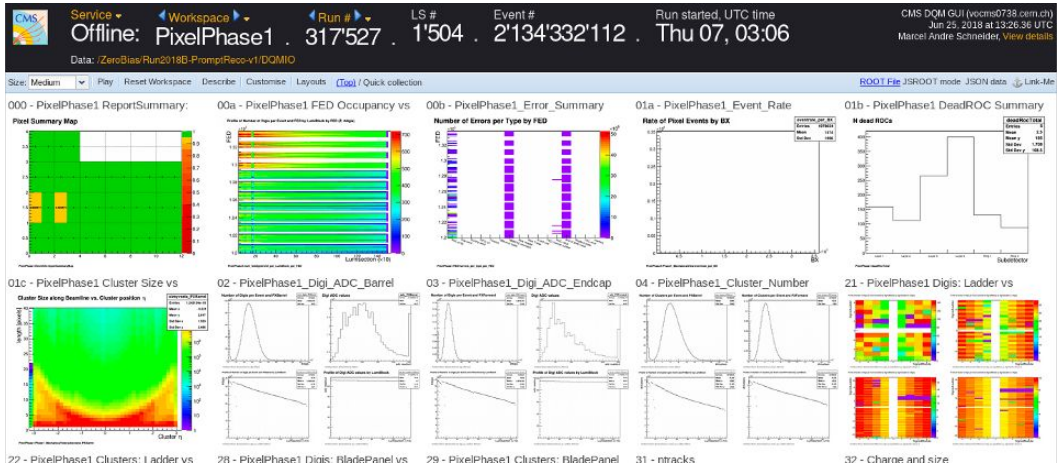
"Run" → thousands of LS

Data Quality Monitoring (DQM)

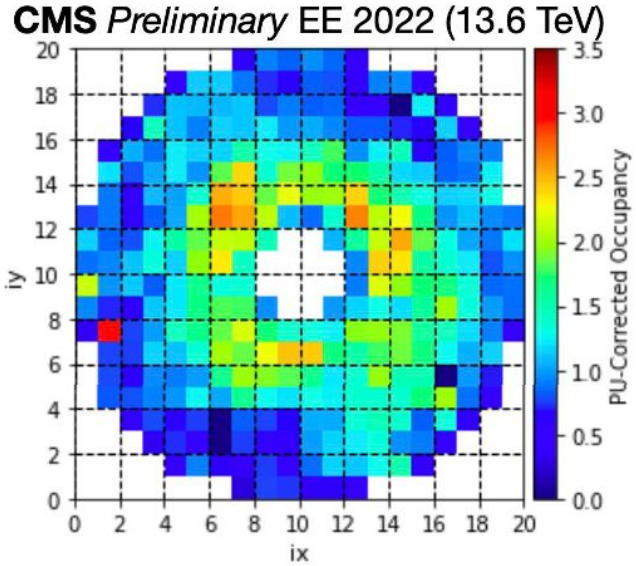


Online DQM: prontly raise alarms during data taking





Monitoring Elements (ME)



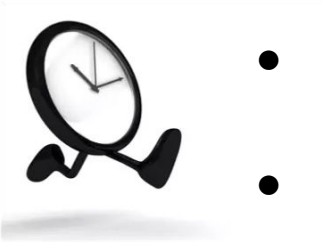
ECAL occupancy images

ME: quantities (i.e. histograms) which are inspected either automatically or by experts

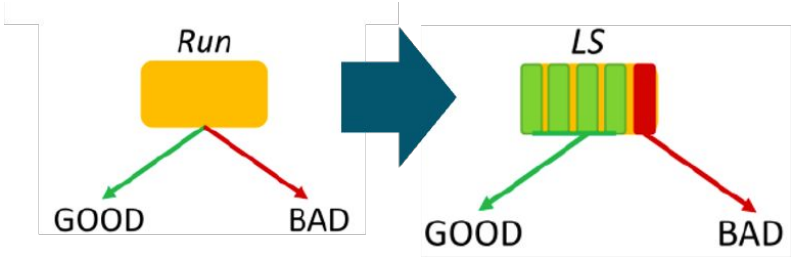
ME in the online DQM: histograms are filled in real time during the run.

ME in the offline DQM: histograms are available for inspection with per-run granularity

DQM challenges and limitations



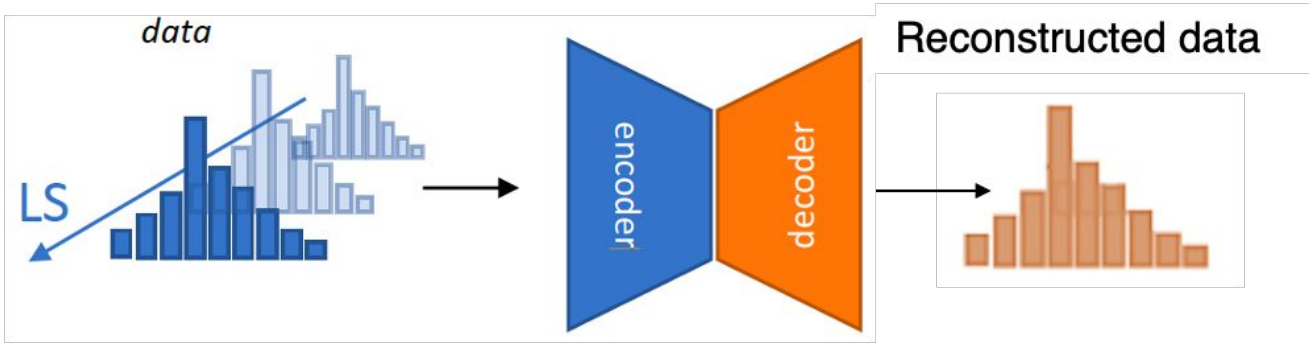
- Online monitoring is a highly time-sensitive operational task
- Impossible to foresee or simulate all potential failure scenarios
- Limited time granularity (run) can potentially hide transient issues only affecting few lumisections
  - Drawback: per-LS approach increases the number of MEs by a factor $\mathcal{O}$(10^3)



Machine Learning for the Online DQM

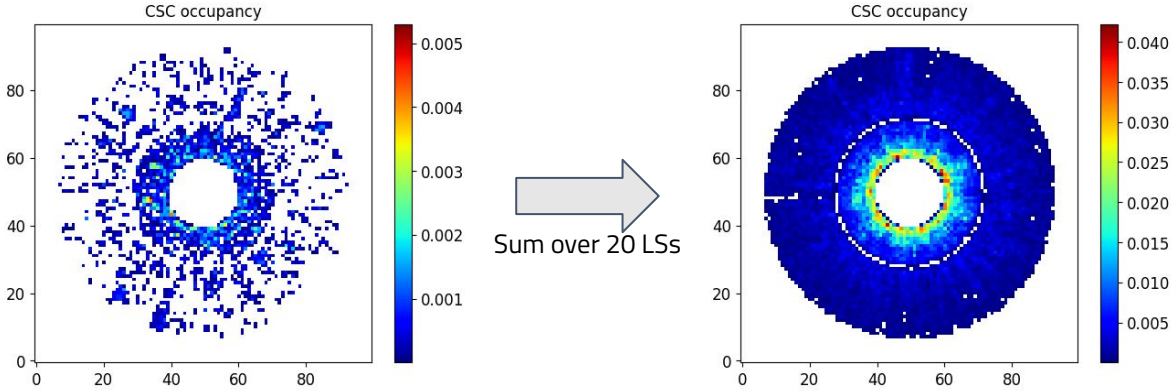
- High number of features
- Large class imbalance (most data is good)
- Non-exhaustive definition of failures

Unsupervised learning for anomaly detection



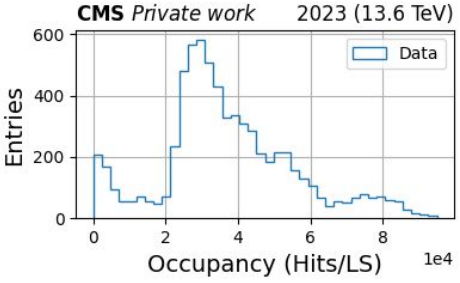
The Muon System use-case: data pre-processing

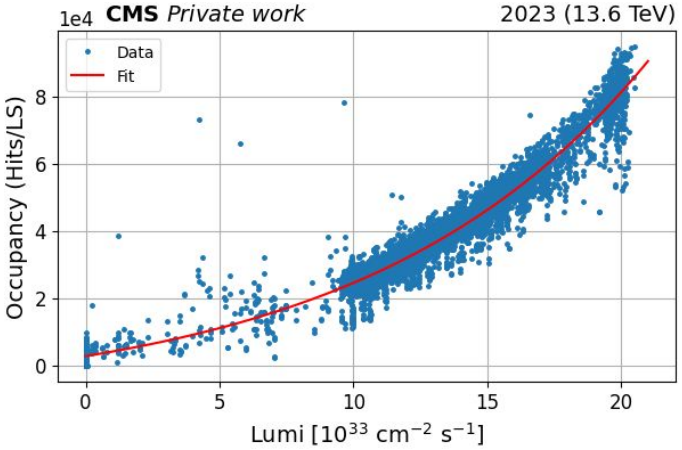
- Basic idea: train a model on occupancy plots from the different detector regions
- Need for preprocessing:
  - Muon System characterised by low occupancies (can't use a single lumisection for training)
  - Occupancy vs lumisection not uniform (depends on instantaneous luminosity)
  - Need to correct for the expected geometrical distribution of incoming particles (strong eta dependency)

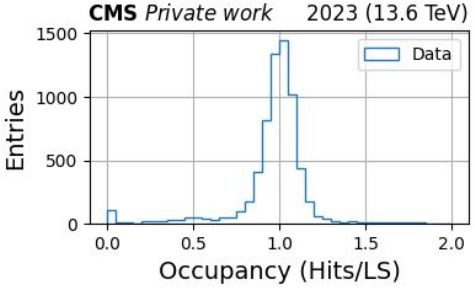


The Muon System use-case: data pre-processing

Entries normalised to instantaneous luminosity



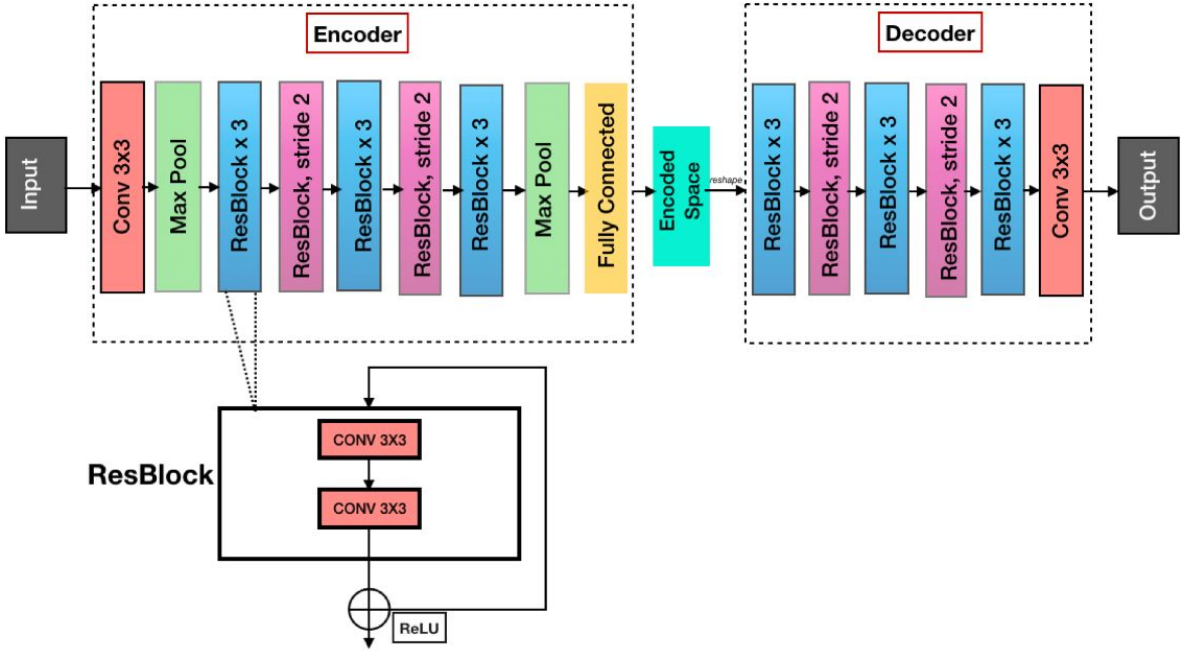




Model architecture

ResNet for unsupervised reconstruction of images

Input: 100x100 pre-processed images (small dataset, 300 examples)

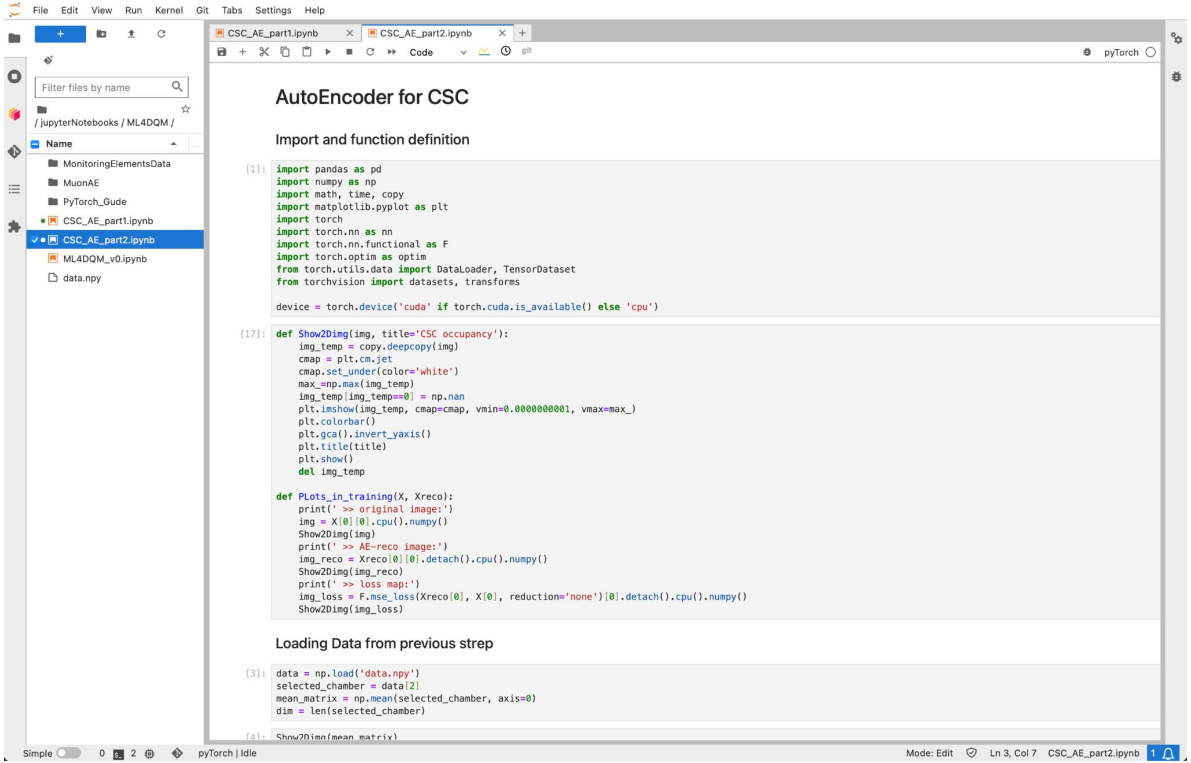


https://doi.org/10.48550/arXiv.2309.10157

Computing resources for training

Leveraging the ReCas JupyterHub service with GPU

https://jvino.github.io/cluster-hpc-gpu-guides/interactive_services/jupyter-lab-for-recas-users/



Preliminary results

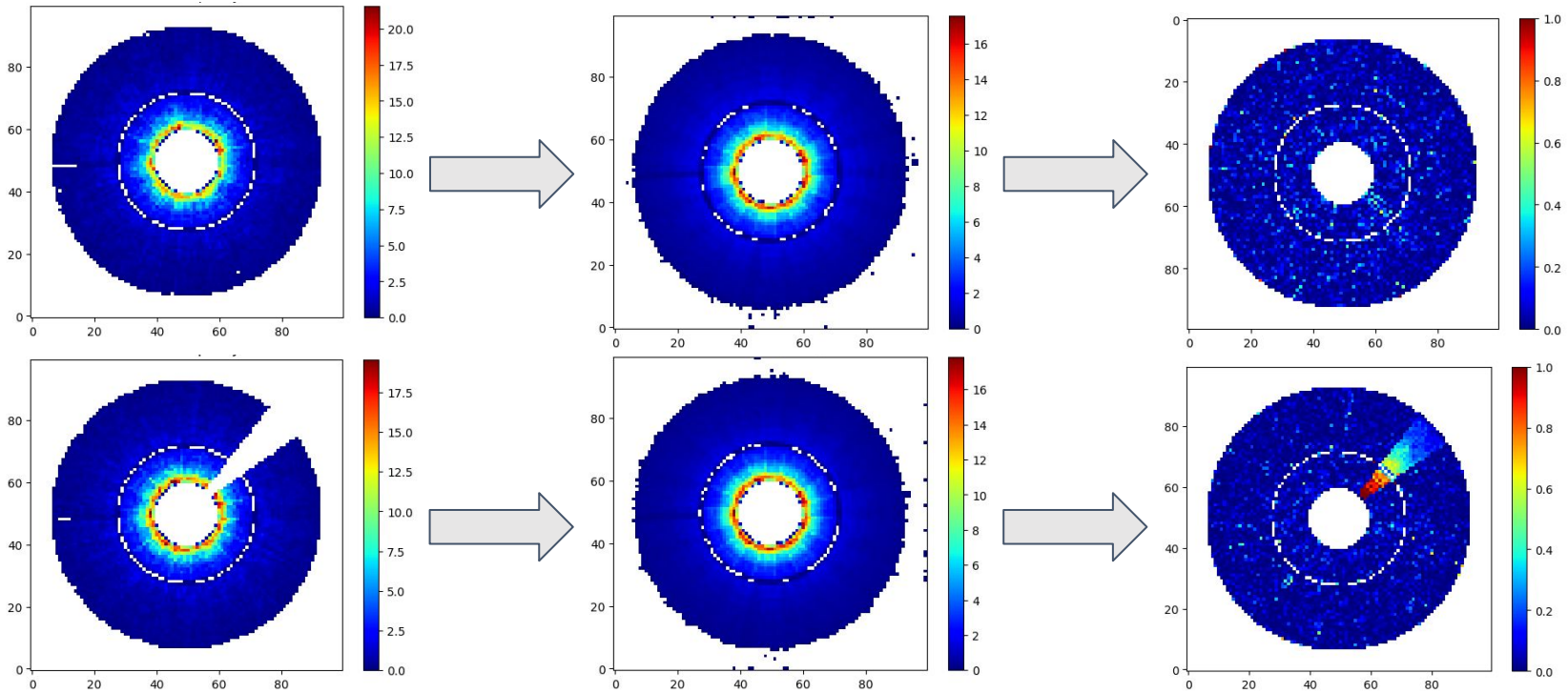
Original image:

AE Reco-image:

Normalized loss map:

Predictions on a good example:

Predictions on a trivial example with an "artificial hole":



To-dos and summary

- Ongoing development for automatised anomaly detection for the Muon System https://github.com/fsimone91/MuonAE
- Requested some muon-related ME to be included in the DQM file production → larger statistics for training will be available soon https://github.com/cms-sw/cmssw/pull/44610
- Preliminary results on a specific detector region (CSC endcap)
  - To-do: extend to other regions/subdetectors
  - To-do: simulate anomalies (e.g. dead channels) to access performance

- This development will be included in the next CMS ML Hackathon dedicated to DQM (July 2024)
- Abstract submitted to CHEP2024 https://indico.cern.ch/event/1338689/abstracts/175693/