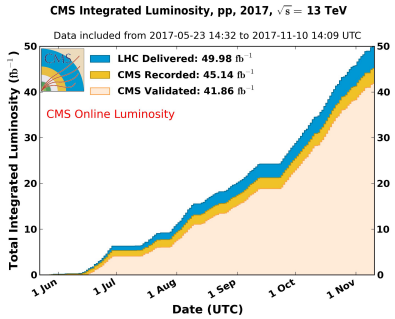
Prospects for automatic data quality monitoring at the CMS pixel detector using machine learning

Luka Lambrecht, on behalf of the CMS Collaboration
e-mail: luka.lambrecht@cern.ch

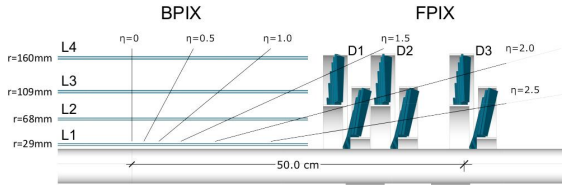
Research Foundation Flanders
Opening new horizons

GHENT UNIVERSITY

Problem

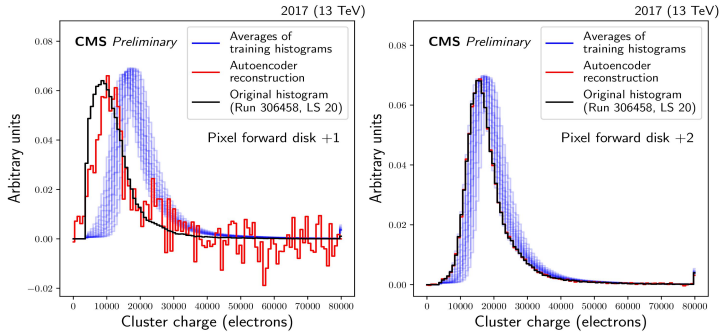


- Goals:
  - optimal usage of the LHC delivered luminosity.
  - filter compromised data from the "good for physics analyses" set.
- Need accurate flagging of detector issues.
- Investigate automatic methods
  - to assist human shifters and experts,
  - to access finer time granularity.



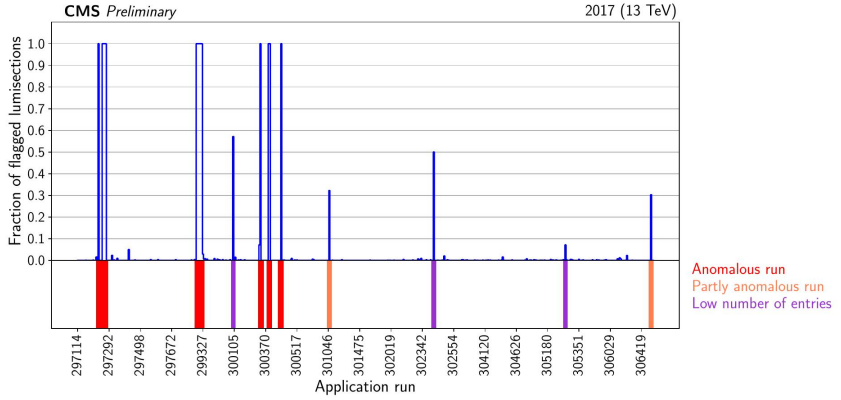
- Case study using monitoring elements from the CMS pixel tracker.
- Look at distribution of electric charge per cluster for the different regions in the tracker.

Method



- Challenges:
  - no reliably labeled data.
  - large class imbalance (few examples of anomalies).
- Problem formulation:
  - unsupervised anomaly detection.
  - given a large set of histograms, find the anomalous ones.
- Approach:
  - train autoencoders on large data volume.
  - good histograms are accurately reconstructed, while anomalies are not.
  - use the reconstruction error as anomality measure.

Results



- Accurate flagging of anomalous luminosity sections.
- Both in 'global training' (e.g. for legacy reprocessing) and 'local training' (e.g. for ongoing data taking).
- Some more work needed to reduce the sensitivity to discrete detector condition changes between runs.

Future developments

- Optimize choice of reference histograms for local training.
- Extend to other monitoring elements.
- Further validation and commissionning in Run-3 data.
- Implement in online DQM software for live data taking.