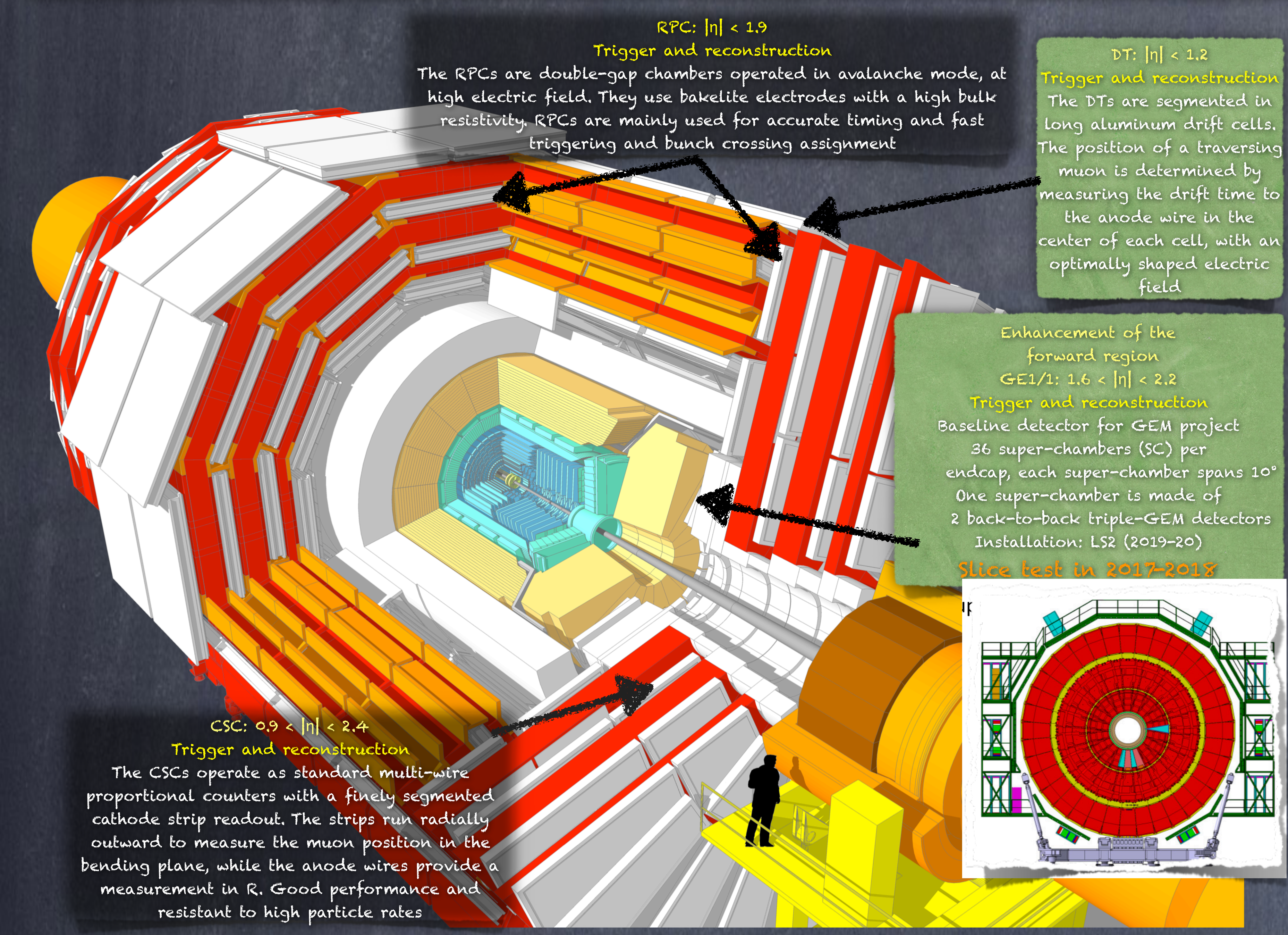
Monitoring tools for the CMS muon detector: present workflows and future automation

Cesare Calabria on behalf of the CMS Collaboration

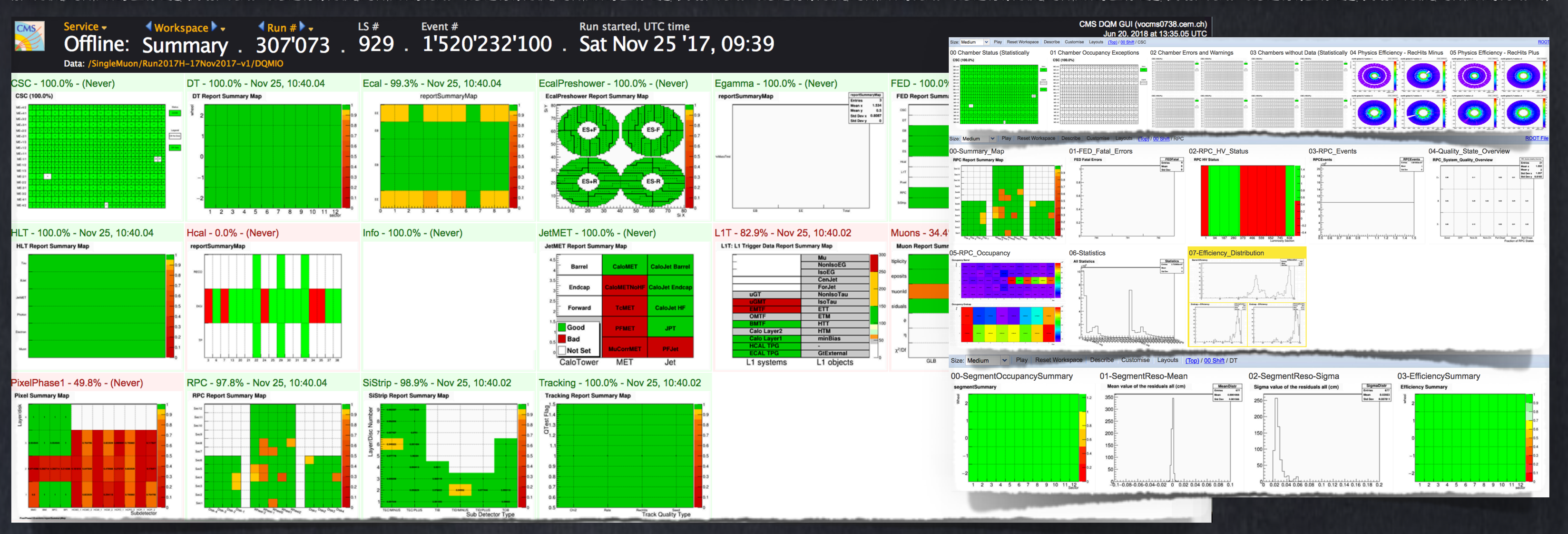
CHEP 2018 Conference, 9-13 July 2018, Sofia, Bulgaria

The CMS muon system presently consists of three detector technologies equipping different regions of the spectrometer. Drift Tube chambers (DT) are installed in the muon system barrel, while Cathode Strip Chambers (CSC) cover the end-caps; both serve as tracking and triggering detectors. Moreover, Resistive Plate Chambers (RPC) complement DT and CSC in barrel and end-caps respectively and are mostly used in the trigger. Finally, Gas Electron Multiplier (GEM) chambers are getting installed in the muon spectrometer end-caps at different stages of the CMS upgrade programme. The CMS muon system has been operated successfully during the two LHC runs allowing to collect a very high fraction of data whose quality fulfills the requirements to be used for physics analysis. Nevertheless the workflows used nowadays to run and monitor the detector are rather expensive in term of human resources. Focus is therefore being put in improving such workflows, both by applying automated statistical tests and exploiting modern machine learning algorithms, in view of the future LHC runs. The ecosystem of tools presently in use will be presented, together with the status of the art of the developments toward more automatized monitoring and the roadmap for the future.



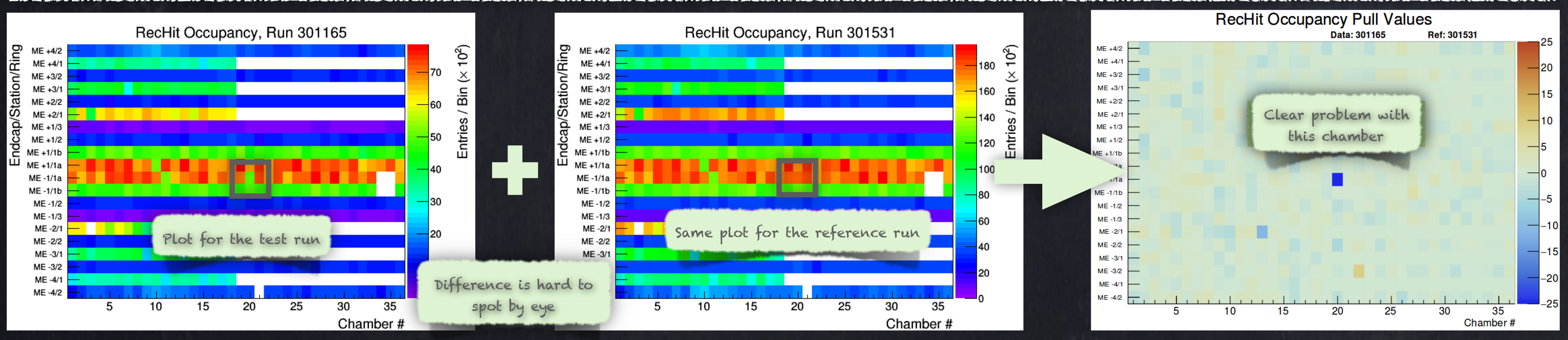
CMS Data Quality Monitoring system

- A critical asset to guarantee a high-quality data for physics analyses (online and offline)
- Online DQM assess data goodness and identifies emerging problems in the detector
  - Data with poor quality are flagged by eyeballing DQM GUI and comparing a set of histograms to a reference good sample
- Problems with current strategy:
  - Delay: human intervention and tests require collecting sufficient statistics
  - Volume budget: amount of quantities a human can process in a finite time period
  - Human driven decision process: alarms based on shifter judgment
  - Changing running conditions: reference samples change over time
  - Manpower: the effort to train a shifter and maintain instructions
- R&D is ongoing to automate the CMS DQM procedure by looking also at innovative machine and deep learning techniques



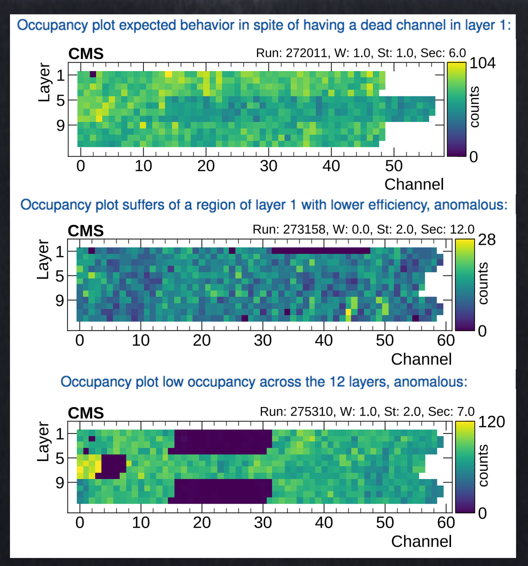
AutoDQM

- Developed a tool called AutoDQM to assist DQM shifters in looking for hard to spot problems
- User specifies reference run and the tool flags plots using statistical tests (bin-by-bin pull values, Kolmogorov-Smirnov test, etc.)
- It is configurable for all the subsystems, no need to manually scan hundreds of plots
- User still needs to specify reference run: time-consuming and not always an obvious choice

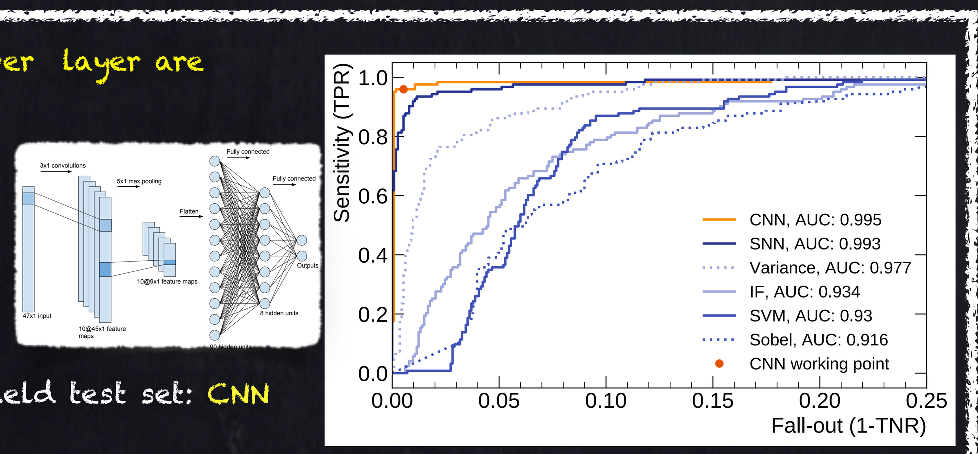


Machine Learning for DT DQM

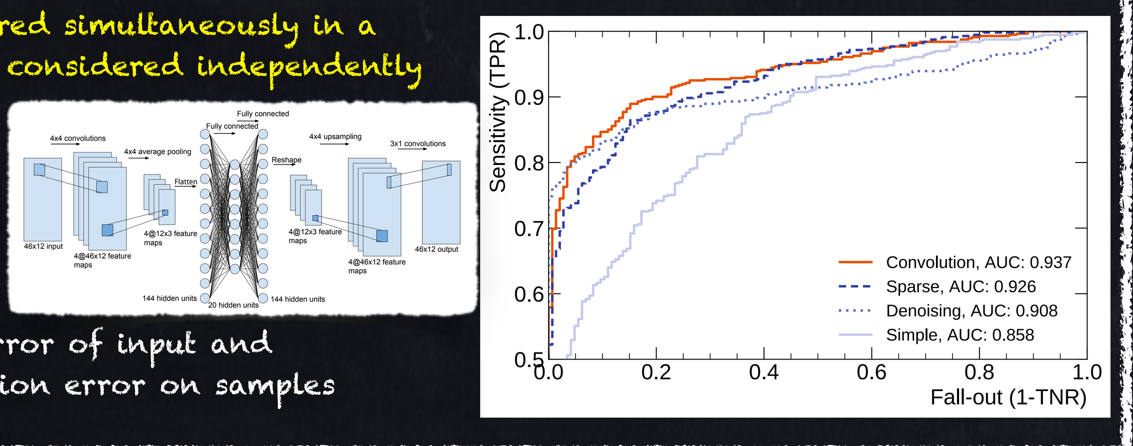
- Occupancy plots are among the most important DQM plots. They show the frequency of hits in given detector channels and are used to quickly identify and diagnose problems
- Very high-dimensional data → standard outlier detection techniques give poor performance. Utilize both supervised and semi-supervised neural-network methods based on the use-case
- More details in the talk by Adrian Alan Pol on monday (July 9): "Online detector monitoring using AI: challenges, prototypes and performance evaluation for automation of online quality monitoring of the CMS experiment exploiting machine learning algorithms"



- Local approach: data collected in each muon chamber layer are treated independently from the other layers
- Preprocess and standardize: interpolate to make the same size, median filter to remove isolated faulty channels, max scaling
- Train a 1D 3x1 CNN using each 47-channel muon chamber layer as a sample
- Compared to numerous benchmark models on withheld test set: CNN gives best performance

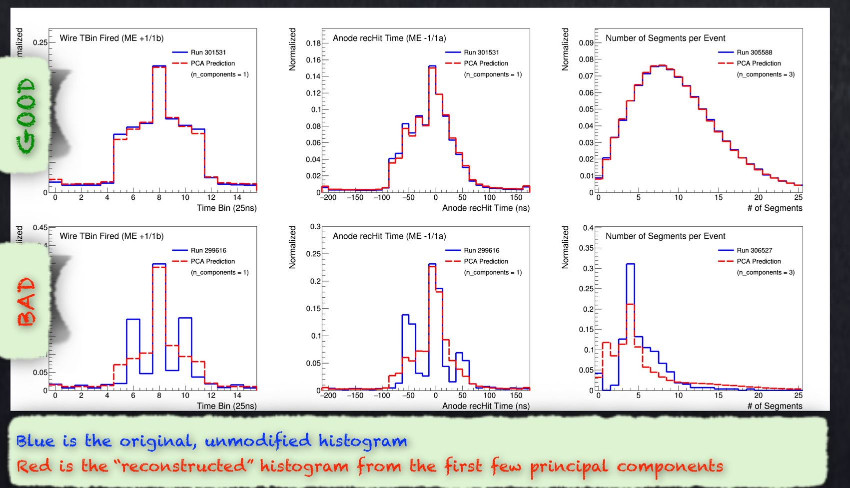


- Regional approach: all layers are considered simultaneously in a muon chamber, but each muon chamber is considered independently from the others
- Few examples of this mode of failure → fully supervised approach will not work, try a semi-supervised AutoEncoders (AE)
- All models minimize the mean squared error of input and reconstructed samples → high reconstruction error on samples that are problematic



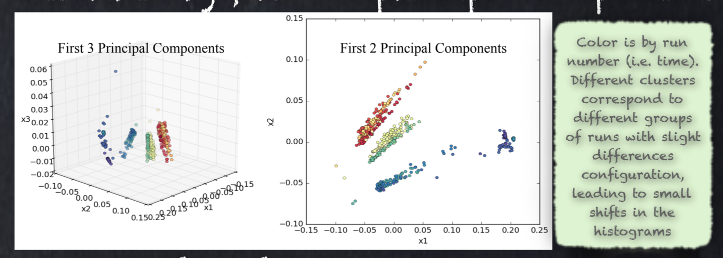
Machine Learning for CSC DQM

- Problem: hundreds of unique 1D DQM histograms. How to detect anomalous ones?
- Want an algorithm that simply flags a DQM histogram as potentially problematic, given nothing but the histogram itself
- Shouldnt require manual labeling of thousands of DQM plots to train algorithm
- Should be easily adaptable to new subsystems/plots
- Should be able to handle a wide variety of running conditions
- Shouldnt depend on a large number of example bad plots (these are quite rare), and should be sensitive to never-before-seen anomalies



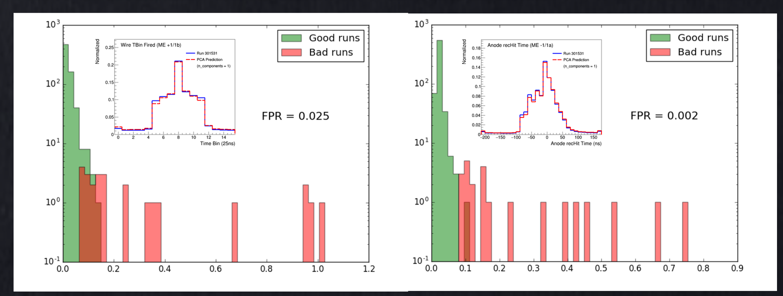
Blue is the original, unmodified histogram
Red is the "reconstructed" histogram from the first few principal components

- Idea: Treat input histograms as (nbins)-dimensional points and reduce to just a few dimensions using Principal Component Analysis (PCA)
- Find that all DQM plots almost completely described by first 1-3 principal components



Color is by run number (i.e. time). Different clusters correspond to different groups of runs with slight differences configuration, leading to small shifts in the histograms

- Since the first few principal components describe the "normal" variation of the histograms, normal runs are reconstructed quite well from just these components
- "Bad" or "outlier" runs cannot be described well from these components, so the reconstructed histogram will not match the original well
- Use sum-of-squared-errors as a measure of similarity between histograms



Machine Learning for Barrel L1Muon Trigger Monitoring

- L1 and L1Muon trigger are implemented in custom electronics: many input components and multiple channels of electronic communication that can fail, leading to "anomalous" trigger rates
  - Important to monitor the system in real time and quickly identify and fix any problems
- GOAL: exploit ML/DL techniques to develop an innovative tool for the L1Muon Barrel Trigger rate monitoring in CMS
- The algorithm must:
  - correlate trigger rates and instantaneous luminosities coming from different Muon detectors and electronics components
  - identify chamber(s) with rate problem(s)
  - correlate different sources of information to make a diagnosis of the issue
- Work is ongoing with promising results especially from the semi-supervised (AutoEncoder) and completely unsupervised approaches (LOF, K-Means clustering)
  - Ultimately, unsupervised approach is probably best as it can in theory generalize to never-before-seen problems