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

Cesare Calabria[1,2,*] on behalf of the CMS Collaboration

[1]INFN (Bari, Italy)
[2]CERN (Geneva, Switzerland)

Abstract. The CMS Muon System has been operated successfully during the two LHC runs allowing to collect a very high fraction of data with a quality that fulfils the requirements to be used for physics analysis. Nevertheless, the workflows used nowadays to operate and monitor the detector are rather expensive in terms of human resources. Focus is therefore being put on 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 state of the art of the developments toward more automatized monitoring and the roadmap for the future.

1 Introduction

1.1 The CMS experiment and its muon system

The Compact Muon Solenoid (CMS) experiment [1] is a general purpose particle physics detector operating at the CERN Large Hadron Collider (LHC) [2]. Data collected with the CMS detector have been used to produce many excellent scientific results in modern particle physics, for instance the discovery [3] and characterization[4] of the Standard Model Higgs boson. In CMS, muons are key ingredients and are measured with detection planes instrumented with four detector technologies: drift tubes (DT), cathode strip chambers (CSC), resistive plate chambers (RPC), and recently with gas electron multipliers (GEM) [6] for enhancing the forward region trigger and reconstruction [7]. A slice test consisting of five GEM chambers is being performed during the 2017-2018 data taking, while the full installation is foreseen for the LHC Second Long Shutdown (LS2) [7]. A detailed description of the CMS muon detectors and their performance can be found in [5].

1.2 The CMS data quality monitoring

Within the CMS Collaboration, physics analyses are performed on validated data, selected by imposing stringent quality criteria. During data taking, a subset of the collected data is processed in real time, to create a set of histograms filled with certain critical quantities. Statistical tests are performed to compare these histograms to a set of predefined references, representing the typical detector response during normal operating conditions. Using the

---

*e-mail: cesare.calabria@ba.infn.it

histogram comparison and the outcome of the tests, expert shifters acknowledge alarms and may decide to intervene (up to stopping the data taking), depending on the evaluation of the problem severity. The knowledge of the LHC running conditions and of the history of possible issues identified in the past are key ingredients in this decision process. Details about the infrastructure used for this Data Quality Monitoring (DQM) are given in [8]. The two monitoring chains are online monitoring, which provides live feedback on the quality of the data while they are being acquired, allowing the operator crew to react to unforeseen issues identified by the monitoring application; and offline monitoring, designed to certify the quality of the data collected and stored on disk using centralized processing. Despite their specific characteristics, these two steps rely on the same anomaly detection strategy: the scrutiny of a long list of predefined histograms, selected to detect a set of known failure modes. These histograms are monitored by detector experts, who compare each distribution to a corresponding reference, derived from good data in line with predetermined validation guidelines. This two-layer monitoring protocol was adopted by the CMS Collaboration for LHC Run I (2010-2012) and Run II (2015-2018).

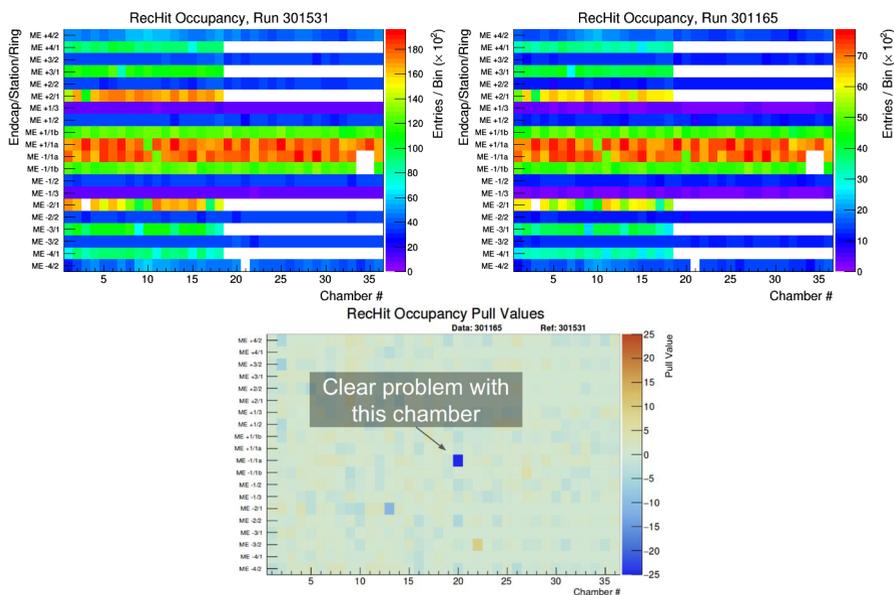### 1.3 Current limitations and future automation

The ever increasing detector complexity, and monitoring data volumes and the necessity to cope with different LHC running scenarios call for an increasing level of automation of the applications in the future. Already, the amount of histograms to monitor is challenging for a single shifter, while the number of histograms to monitor increases every time a new failure mode is identified and consequently added to the list of known potential problems. Furthermore, the human intervention and currently implemented tests require collecting a substantial amount of data, implying a detection delay. Finally, the cost in terms of human resources is substantial, i.e. the constant effort of the DQM shifters and the expert personnel responsible for updating the good data references and related instructions. For these reasons the CMS collaboration is moving towards a complete automation of the DQM procedure to solve those issues and to reduce first of all the human cost. The ongoing R&D within the CMS community is proving that automation is possible by using more advanced and modern approaches like artificial intelligence. In particular within the Muon community several tools are being developed that will be described in the next sections.

2 An automated statistical approach: AutoDQM

AutoDQM[9] is a standalone tool designed to offload the time consuming process of individually scanning plots looking for discrepancies. It is a semi-automated system based on established statistical tests that flags all outliers for further analysis/investigation aiming at reducing the amount of manual work of DQM shifters. The development started for the CSC subsystem, but the tool is flexible enough that it is being extended one-by-one to all the other CMS subsystems. AutoDQM works on histograms of the CMS DQM and it is designed to treat one- and two-dimensional plots differently.

A Kolmogorov-Smirnov (KS) test is used to judge the difference between the underlying distributions of the data/reference histograms. The empirical distribution functions (EDF) of each sample are calculated and then the maximal difference between the two EDFs is used as a metric to find the probability that the two samples are from the same distribution. AutoDQM uses the maximum Kolmogorov distance to decide whether or not to plot the histogram for shifter analysis. If the KS statistic is greater than a particular cut, then the plot is marked as anomalous and displayed on the AutoDQM GUI.
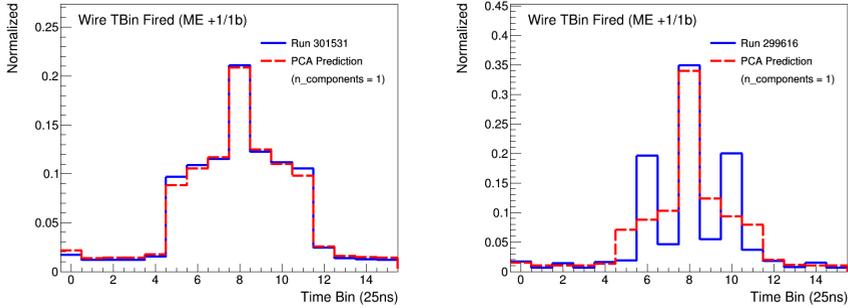
For the 2D histograms, AutoDQM measures the chi-squared values of the data and the reference graphs and plots the pull values of each bin onto a new identical graph. The upper row of Fig. 1 shows examples of typical 2D plots from the CSC DQM GUI: the plot on the left is a map of the system and represents its normal behaviour, while the right plot is obtained in a different moment of the data taking, and it hides an anomaly invisible to the human eye. The bottom plot of Fig. 1 is produced by AutoDQM (starting from the first two), and it is able to highlight clearly the problematic chamber in the system. The tool also stores the maximum pull value of each graph, and it uses the chi-squared value of each data-reference comparison as well as the maximum pull value of their pull histogram to decide whether or not to plot the histogram for the shifter analysis.

Figure 1. Upper row: The left plot shows a typical and normal 2D-histogram selected from the CSC DQM for a certain reference run. The right plot is obtained for another run and it hides an anomaly invisible to the human eye. Bottom plot: The plot is a final plot produced by AutoDQM filled with the pull values for each bin after the data/reference comparison. The blue bins clearly indicate an abnormal deviation from the reference plot, warning about a possible anomaly in the system.

The AutoDQM implementation, as described previously, requires that a good reference run is always provided to evaluate the quality of a data run. In order to overcome this problem, a machine learning approach was considered with success. The Principal Component Analysis (PCA) was used to analyse the quality of 1D DQM histograms. PCA reduces a distribution to some number of principal "components" that describe its "normal" variation (since anomalous runs are very rare). What was found is that 95% of the variance in all the histograms is typically described by the first few components belonging to physical properties of the runs like luminosity. From those components it is possible for any data histogram to "reconstruct" the expected distribution. The quality of the reconstruction is evaluated by summing the squared errors between the input data and its reconstruction. Since the first few principal components describe the "normal" variation of the histograms, normal runs are reconstructed quite well from just these components. On the other hand, "bad" or "outlier" runs cannot be well described from these components, so the reconstructed histogram will not

match the original well, as it can be seen in Fig. 2. This approach brings several advantages. In fact, it is a generic method that can be used on arbitrary plots. Moreover, it is unsupervised, so no manual labelling of thousands of plots is needed, it is able to handle a wide variety of running conditions, and it can identify not-before-seen issues.

Figure 2. Example 1D distributions for a specific CSC chamber reconstructed starting from the principal components for a "good" (left) and a "bad" run (right). Blue is the original distribution, red is the PCA prediction.
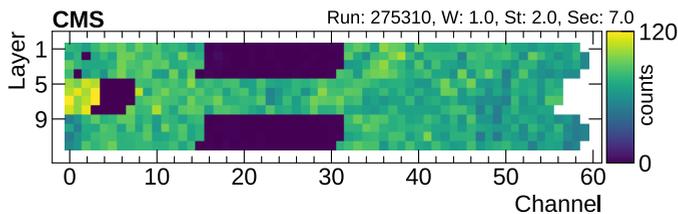
3 A Machine Learning approach for the Muon DQM

3.1 Neural networks for detecting DT anomalies

The DT system consists of five wheels, each one divided into 12 azimuthal sectors. Chambers are arranged in four stations at different radii, named MB1, MB2, MB3 and MB4. Each station consists of 12 chambers (one per sector) except for MB4 (which contains 14 chambers). The total number of chambers is then 250. Each DT chamber consists of 12 layers of drift tubes, each containing a variable number of tubes, up to 96. Each tube corresponds to one readout channel. Particles carrying an electromagnetic charge and traversing a tube release an electronic signal by ionizing the gas in the tube giving rise to a so called "hit". By combining the information provided by all the channels, one can determine the trajectory of the particle crossing the chamber. For each chamber in a given run, the current DQM infrastructure records an occupancy matrix, which contains the total number of particle hits at each channel for a given time interval.

The anomaly detection method currently used in the online monitoring production system is able to identify regions of cells not providing any electronic signal, large enough to affect the track reconstruction in the chamber. This is the most frequent issue, usually related to transient problems in the readout electronics. An example of this kind of failure is shown in Fig. 3. Although it quantifies the fault severity on the basis of the fraction of affected channels, the current detection method does not identify specific faulty layers. A novel approach has been proposed to go beyond the functionalities of the current production algorithm[10]. Starting from the identification of layers with under-performing cells, it provides effective identification of faulty chambers. In particular, two complementary approaches are considered, in order to spot local problems (at chamber level) and even intra-chamber issues (at layer level).
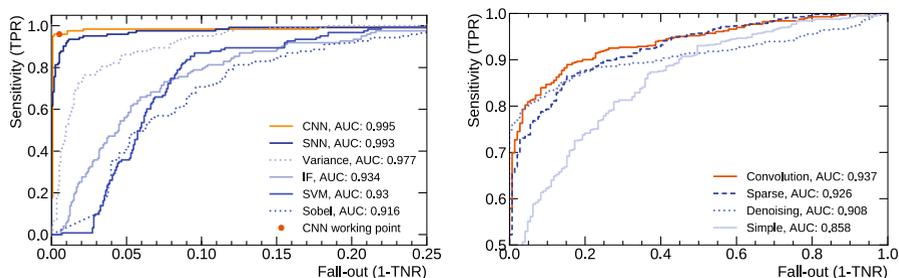
For the identification of local problems a classifier is trained considering each layer independently from the others. The goal is to identify regions of channels not registering any hits

Figure 3. Example of visualization of input data for a DT chambers. The chamber shows regions of low occupancy across the 12 layers and should be classified as faulty.

(dead channels), or having lower detection efficiency (hence lower hit counts with respect to the neighbouring ones in the same layer) or being dominated by electronic noise (called noisy channels). These are by far the most frequent failure modes. Several supervised and unsupervised benchmark algorithms (simple variance, Sober filter, Isolation Forest (IF), μ-SVM) have been considered and compared to supervised learning algorithms, i.e. a fully connected shallow neural network (SNN), and a convolutional neural network (CNN). The ground truth to train and test the algorithms is established by field experts by visually inspecting the input sample before any preprocessing. The performance of the different algorithms is shown in the right plot of Fig. 4. Compared to statistical, image processing, or other machine learning based solutions, supervised deep learning clearly outperforms all the other algorithms[10].

In order to identify intra-chamber problems, the relative occupancy patterns of the layers within a chamber are used. The aim is to detect failure modes where the occupancy of hits decreases uniformly in a specific layer or set of layers. Typical examples of these kinds of failures are problems related with the high-voltage bias of the drift cells, but they are quite rare. For this reason a supervised approach is not feasible and a semi-supervised one has been preferred, training only against labelled good data. In order to test the model, a subset of the data, during which a specific layer (layer 9) of some chambers was operating at a voltage lower than the nominal one, is used. The following semi-supervised approaches have been considered: simple bottleneck autoencoder, convolutional autoencoder, denoising autoencoder, autoencoder with sparsity regularization in the hidden layers (details can be found in [10]). All models are instructed to minimize the mean squared error (MSE) between original and reconstructed samples. The right plot of Fig. 4 shows that the performance of all models is good, especially of the convolutional autoencoder[10].

Figure 4. Receiver operating characteristic (ROC) curves for different models used in the local (left) and regional (right) approach. The Area Under the Curve (AUC) is quoted to compare the performance.

3.2 CMS L1 Muon trigger monitoring with ML

The CMS Trigger System

The LHC provides proton-proton (p–p) and heavy-ion collisions at high interaction rates. For protons the beam crossing interval is 25 ns. Depending on luminosity, several collisions occur at each crossing of the proton bunches. Since it is impossible to store and process the large amount of data associated with the resulting high number of events, a drastic rate reduction has to be achieved. This task is performed by the trigger system. The rate is reduced in two steps called Level-1 (L1) Trigger and High-Level Trigger (HLT), respectively. The L1 trigger reduces the event rate from the LHC bunch crossing frequency (40 MHz) to the maximum sustainable readout rate, i.e. 100 kHz. The Level-1 Trigger consists of custom-designed, largely programmable electronics, whereas the HLT is a software system implemented in a filter farm of about one thousand commercial processors. The L1 Trigger uses coarsely segmented data from the calorimeters and the Muon System, while holding the high-resolution data in pipelined memories in the front-end electronics. The HLT stage makes use of the whole detector with full granularity and the output rate is limited to an average of ~1 kHz by the offline resources. All three muon systems take part in the trigger. Thanks to a regional segmentation muon tracks are treated separately depending on the pseudorapdity η. It distinguishes a barrel region (low η), an endcap region (high η), and a transition region between them ($|\eta|$ ~1) called overlap. Such regions will result in different triggering algorithms but also in different deployments of hardware processors. More details about the CMS trigger system can be found in [11].

Monitoring the status of the L1 trigger is vital for the CMS operations. Also in this case the DQM and its histograms are the primary source of information that human shifters check continuously during the data taking facing the same issues already highlighted in sections 1.2 and 1.3. The CMS Muon community decided to start an R&D activity to exploit machine learning and deep learning techniques for developing an innovative tool for monitoring the L1 muon trigger rate in the CMS barrel region. Firstly the algorithm must correlate trigger rates and online instantaneous luminosities coming from the CMS online database and identify chamber(s) with rate problem(s). Eventually the algorithm must correlate more informations coming from further DQM histograms to make a more precise prediction of the issue and to lighten the shifter work load. Though the dependency of trigger rate on the instantaneous luminosity is expected to be linear for almost all the chambers, an approach based on artificial intelligence can be advantageous especially for those chambers exhibiting a non-linear behaviour or affected by a high level of noise. Once the development is completed, the natural consumers will be the online shifter and the Muon System online operation teams. The R&D activity to produce a proof-of-concept started with the deployment of a monitoring tool for the DT local trigger rates, with the plan of extending it to the RPC local trigger rates and other stages of the barrel muon trigger chain.

DT local trigger monitoring with Machine Learning

The minimal set of features needed to automatize the trigger anomaly detection in the DT system is composed of the position of the chamber in the system itself (expressed in terms of wheel, sector, and station number), the local trigger rate, and the instantaneous luminosity. Since anomalies are quite rare and the tool must be sensitive also to unforeseen issues, a supervised approach is not feasible. Semi-supervised algorithms trained on data coming from healthy chambers seemed to be the most reasonable choice. The price to pay is the search for a good reference run, but this is not a difficult task for the DT system. In fact, from the study of the data it turned out the DT system is very stable and almost all the chambers exhibit

mostly a linear behaviour. This resulted in the possibility to use another derived quantity as a feature that is able to fully characterize every chamber in the system. The rate over instantaneous luminosity ratio, which will be called the cross-section (CS) in the rest of this report, is more or less flat for every chamber during entire runs and it is quite sensitive to any emerging problem, see for example the left plot in Fig. 5.
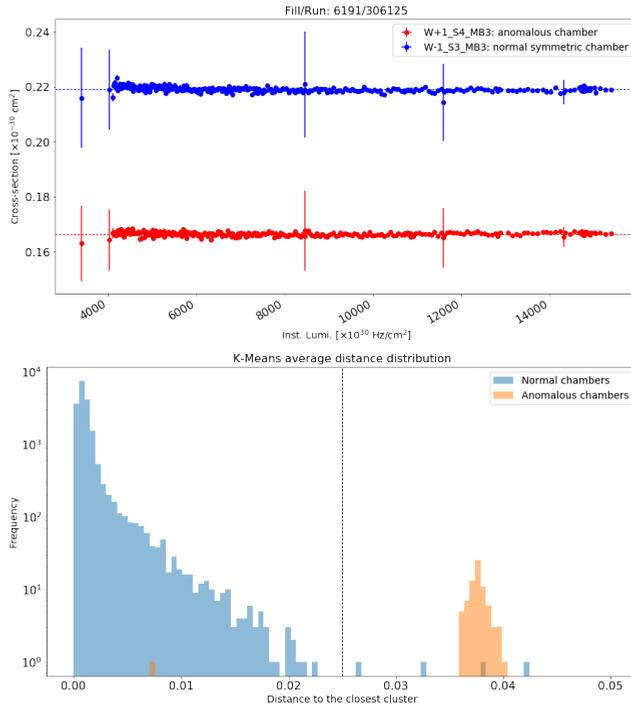
During the R&D activity several algorithms (supervised, unsupervised and semi-supervised) have been investigated by testing the performance on a known problematic chamber that for this reason provides a lower rate with respect to its symmetric chamber in the system. As already anticipated, supervised models are not a viable choice, since it is impossible to provide examples for all the possible issues in the system, and this would result in a model not able to identify efficiently issues different from the one used for the training phase. Unsupervised models need to be further investigated, but they might need to collect a large amount of data to make a reliable inference online. The most promising results in terms of efficiency (~100%) and false positive rate (<1%) came from the use of an autoencoder model and from K-Means clustering. Both the algorithms are trained towards the good data of the chosen reference run and tested on an independent test sample where data from the anomalous chamber are included. For the autoencoder the score is calculated as the mean squared error between original and reconstructed samples. For K-Means clustering the distance from the centroid of the closest good cluster is used as score. Given the dimensionality of the problem, and considering its high convergence speed, K-Means clustering seems to be the most suitable choice for an online monitoring. Fig. 5 shows the distribution of scores obtained by using the K-Means clustering model to make some inference on an independent test sample. As expected, normal chambers are close to the centroids of good clusters built by training the algorithm towards the good data of the chosen reference run, while anomalous chambers are placed quite far away. This gives the possibility to set (and adjust properly) a threshold on the distance in order to efficiently recognize anomalies in the DT local trigger rates, providing also geographical information to the experts. The method has been successfully tested also on other transient issues that happened during the 2017 and 2018 data taking.

4 Conclusions and outlook

The CMS Muon community is actively pursuing a variety of modern solutions to push the DQM process towards higher levels of automation. Work is ongoing with room for improvements within all the subsystems, but many promising results have been achieved so far in a variety of data quality monitoring applications by using both traditional statistical methods and modern neural network-based approaches. This document has shown how detector malfunctions can be identified with high accuracy by a set of automatic procedures. Currently, the algorithms for the DT occupancy monitoring are being integrated into the CMS online DQM infrastructure in order to be commissioned with the data of the 2018 Run. AutoDQM is already accessible via a web interface and is being extended to other subdetectors. The R&D activity concerning the monitoring of the L1 muon trigger in the CMS barrel region will continue by refining the models and extending to the RPC system. Since many monitored quantities in typical high-energy physics experiments are based on histograms, the approach proposed in this document could be extended beyond the presented use cases.

References

[1] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST **3** S08004 (2008).
[2] The LHC Study Group, "The Large Hadron Collider: conceptual design", Technical report, CERN/AC/95-05 (LHC) Geneva (1995).

Figure 5. Left: cross-section vs. instantaneous luminosity for W+1_S4_MB3 (red) affected by a known issue that reduces its rate and its normal symmetric chamber W-1_S3_MB3 (blue). Right: distribution of scores predicted by K-Means clustering for W+1_S4_MB3 (yellow) and all the other normal chambers (light blue) in the DT system during run 306125.

[3] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Physics Letters B **716(1)**:30–61 (2012).

[4] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV," Eur. Phys. J. C **75(5)** (2015), 212.

[5] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV," JINST **13** P06015 (2018).

[6] F. Sauli, "GEM: A new concept for electron amplification in gas detectors", Nucl. Instrum. Meth. A **386** (1997) 531–534.

[7] A. Colaleo, A. Safonov, A. Sharma and M. Tytgat, "CMS Technical Design Report for the Muon Endcap GEM Upgrade", CERN-LHCC-2015-012. CMS-TDR-013 (2015).

[8] F. De Guio, "The data quality monitoring challenge at CMS: experience from first collisions and future plans", Technical report (2015).

[9] AutoDQM GITHub webpage: https://github.com/jkguiang/AutoDQM/wiki.

[10] A. A. Pol, G. Cerminara, C. Germain, M. Pierini and A. Seth, "Detector monitoring with artificial neural networks at the CMS experiment at the CERN Large Hadron Collider," arXiv:1808.00911 [physics.data-an].

[11] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade," CERN-LHCC-2013-011, CMS-TDR-12, CMS-TDR-012.