Available on CMS information server

CMS NOTE -2010/013

The Compact Muon Solenoid Experiment

CMS Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

11 February 2010 (v3, 24 August 2010)

Data Quality Monitoring for the CMS Resistive Plate Chamber Detector

A. Cimmino[a)], D. Lomidze, P. Paolucci

Istituto Nazionale di Fisica Nucleare - Sezione di Napoli

Abstract

Resistive Plate Chambers (RPCs), with their excellent time resolution ($\sim$ ns), were chosen as dedicated muon trigger detectors for the CMS experiment. RPCs fulfill the job of muon identification, estimate the momentum and unambiguously assign bunch crossing. The critical tasks of monitoring detector performance, debugging hardware, and certifying recorded data are carried out by the Data Quality Monitoring (DQM) system. The CMS DQM framework provides tools for creation, filling, storage, and visualization of histograms and scalar elements. It also offers standardized algorithms for performing statistical tests and automated data certification. Within this framework, the RPC DQM system was developed. The later is composed by a set of user defined algorithms and is intended to be used both online, during data taking, and offline, during the reconstruction stage at Tier-0 and re-reconstruction at the Tier-1s. Run by run, the system measures detector level and physics quantities which are subsequently stored in a dedicate database. We here describe the structure, functionalities, and performance of the DQM applications for the CMS RPC detector.

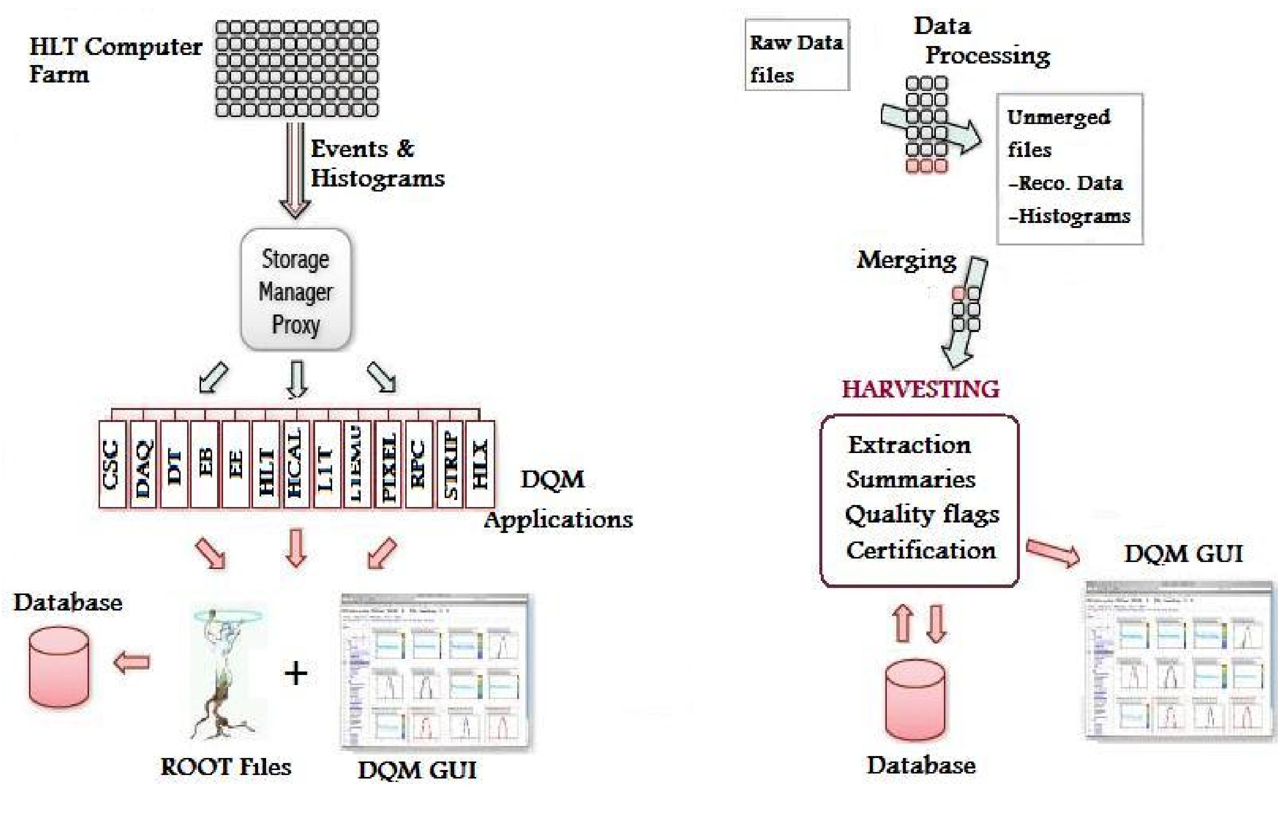
[a)] Also *Università degli Studi di Napoli Federico II*

1 Introduction

The Resistive Plate Chamber (RPC) [1] detector at the Compact Muon Solenoid (CMS) [2] experiment is a complex and massive system. A total of 912 double-gap chambers are installed in the CMS magnet return yoke, covering the pseudorapidity range $|\eta| \leq 1.6$.
Detector performance and quality of collected data are monitored by user defined algorithms based on the CMS Data Quality Monitoring (DQM) framework [3]. DQM has the objective of discovering recurring problems, both in hardware and reconstruction software, promptly and accurately enough to broadly pin-point its origin.

2 CMS DQM

The CMS DQM provides tools for creation, filling, storage, and visualization of histograms and scalar elements. 1D-, 2D-, and 3D-histograms, 1D- and 2D-profiles, integers, floats, and string messages can be booked, filled and updated anywhere in the analysis code. The infrastructure also offers users standardized algorithms for performing statistical tests and automated certification. DQM is intended to be used both online, during data taking, and offline, during the reconstruction stage at Tier-0 and re-reconstruction at the Tier-1s [4]. The online system monitors detector, trigger and DAQ hardware statuses. While offline applications certify the quality of reconstructed data and validate calibration results, software releases, and simulated data.
In the following subsections, the main features of the structure,visualization, and operation of the CMS DQM will be outlined.



(a) Online DQM system. (b) Offline DQM system.

Figure 1: CMS DQM workflows for online (a) and offline (b).

2.1 Online monitoring

Online DQM applications are an integral part of the event data processing, as illustrated in fig. 1(a). Each application, usually one per subsystem, receives event data through a dedicated storage manager event server. A special stream of events is used to perform DQM operations [5]. The stream contains detector and trigger raw data, Level1 and High Level Trigger (HLT) summary results, in addition to HLT by-products essential for monitoring trigger algorithms. Events are delivered to data quality applications at 10 - 15 Hz. There is no event sorting nor handling, and no guarantee parallel applications receive the same events.
Along with event data, the storage manager serves the DQM applications with a limited number of histograms filled in the HLT filter units (FU), i.e. the logical nodes that compose the HLT processer farm. In the FUs the event processing rate reaches 100kHz and all events can be accessed, even those that have not passed any HLT filter. Thus trigger rates and rejection factors can be computed. Identical histograms produced across different FUs are summed and sent to the storage manager proxy, which saves them to files and delivers them to the DQM consumers.
In the online environment, DQM output, which include histograms, alarm states and quality test results, is made available in real time to a central graphical user interface (GUI)[6], accessible form the web. Being web-based, this central GUI permits users all over the world to access the data and check results without installing experiment specific software. Monitoring data is also stored to ROOT files [9] periodically during the run. At the end run, the final result files are uploaded to a large disk pool on the central GUI. Subsequently, files are merged to larger size and backed up to tape. Recent monitoring data (several months worth) are cached on disk for easy access. The GUI was costum built to fulfill the need of shifters and experts for efficient visualization and navigation of DQM results, not as a physics analysis tool.
Starting and stopping DQM online applications, as well as the storage manager proxy is centrally managed by the CMS Run Control System [8], while the DQM GUI web server is independently managed.

2.2 Offline monitoring

Offline DQM, schematically represented in fig. 1(b), runs as part of the reconstruction process at Tier-0, of the re-reconstruction at the Tier-1s [4], and of the validation of software releases, simulated data, and alignment and calibration results. Despite the difference in location, data content and timing of these activities, offline monitoring is unique and formally divided into two steps. First, histograms are created and filled while data are processed event by event. Monitored information is stored along with normal event data files. The second step is called harvesting. During this step, histograms and monitoring information produced in step one are extracted and merged to yield full statistics. Efficiencies are calculated, summary plots are produced, and quality tests are performed. The automated data certification decision is taken here (see section 6 for more details).
The disadvantages of offline monitoring is the latency of reconstructed to raw data, which can be as long as a few days. Tier-0 has a time delay of one - two days, whereas the Tier-1 re-processing takes from days to weeks. Alignment and calibration quantities are validated with a latency of hours to few days. The validation cycle of simulated data depends entirely on sample production times, and varies anywhere from hours for release validation to weeks on large data samples. On the other hand, the advantages are substantial. All reconstructed events can be monitored and high level quantities are available. This allows rare or slowly developing problems to be identified.
At the end of the offline DQM process the output data file is uploaded to a large disk pool of the central DQM GUI server. There files are merged to larger size and backed up to tape; recent data is kept cached on disk for several months. Moreover, online and offline servers provide a common look and feel and are linked together as one entity. Thus, the entire CMS collaboration can access DQM data at one central location.

3 RPC DQM Requirements

The RPC DQM system is composed of a set of dedicated tasks which debug hardware, monitor detector performance and assess data quality by monitoring detector level and physics quantities. The parameters the RPC performance group has chosen to monitor are various. Among them there is chamber occupancy, i.e. the distribution of single hits per channel. Other monitored quantities are multiplicity, i.e. the number of hits per event, and number of clusters, i.e the number of consecutive strips fired per event, as well as the number of clusters per event. The fraction of clusters with size equal to 1 is also a parameter of interest, as explained in section 6. Moreover, data integrity is constantly controlled. Inconsistencies in detector, readout channel, or electronic channel identification numbers are checked as well as data size and data format. Particular attention is given to synchronization parameters. Delay of RPC signals with respect to trigger signals is monitored by DQM applications. The delay is given in units of bunch crossings (25 ns). In addition, time alignment of hits due to a crossing muon is calculated. The

spread in time is also given in units of bunch crossings. Chamber efficiencies is monitored only in the offline environment, where reconstructed objects are available. Efficiency, defined as the number of reconstructed hits over the number of expected ones, is measured for each roll. It is evaluated by extrapolating muon segments reconstructed in the DTs or CSCs to the closest RPC plane and looking for matching hits. Additionally, to allow comparisons among different detector parts and subsequent runs, normalization criteria, using the number of channels and total number of events are applied where needed.
Fig. 2 shows an example of how a problem in the integrity of collected data can be promptly spotted using DQM tools. A general 2D-histogram (fig. 2(a)) shows the shifter the RPC Front End Drivers (FED) [10] are in warning state. The expert, once alerted, checks more detailed histograms. He or she confirms that one of the RPC FEDs, namely FED 790, is sending error messages at almost every event (non-zero entries in the first bin of fig. 2(b)). Accessing more refined information allows to spot the nature of the problem and intervene accordingly. In this particular example, data contained invalid IDs from the RPC data acquisition system and the problem was solved resetting that particular FED. Fig. 3(a) shows the occupancy of a given roll in the Barrel region as computed by

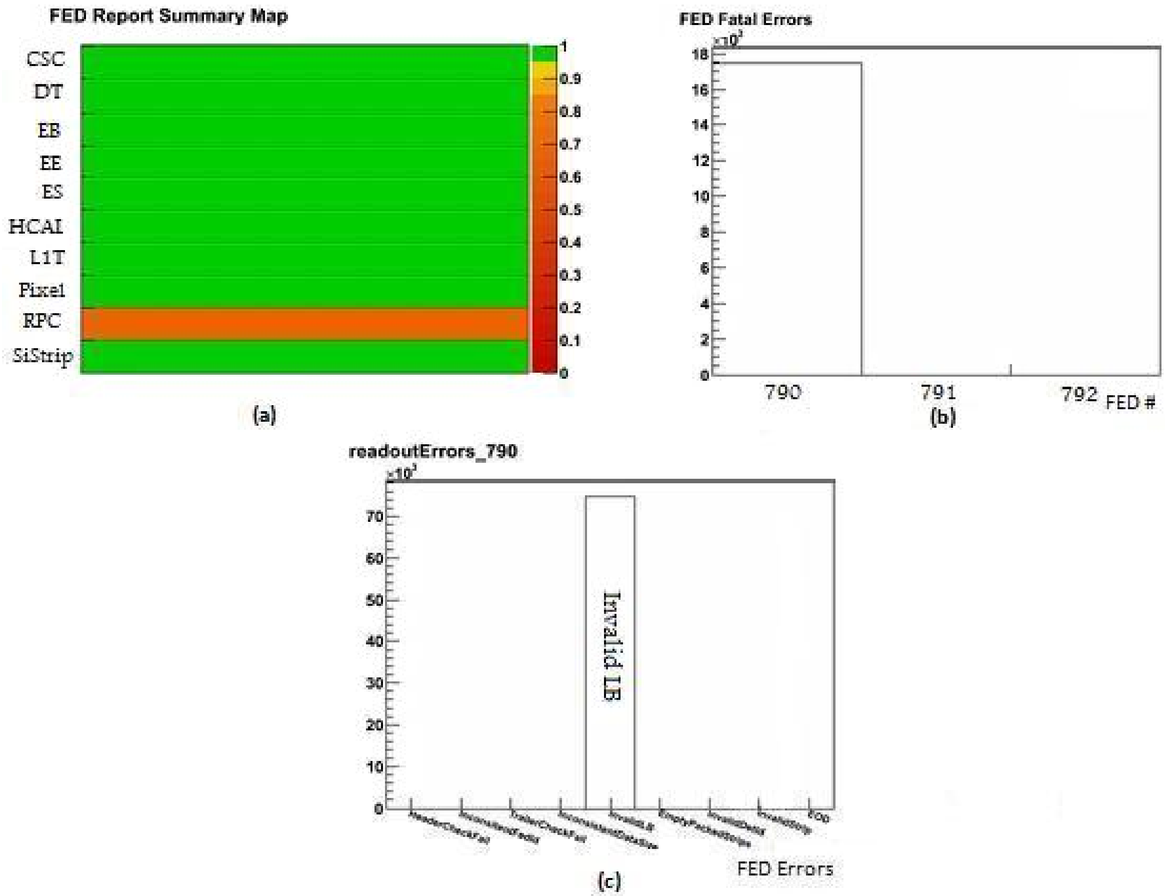


Figure 2: Examples of a FED error messages correctly detected by the DQM. The warning state is clearly visualized in the shifter summary plot (a). Non-zero entries in first bin of (b) pin point the problem to FED 790. Figure (c) shows the nature of the problem.

the online DQM during a run of noise. The plot clearly shows a dead region in the roll, which corresponds to a non functioning chip in a front end board of the RPC chamber. This is more clear, if compared to the occupancy distribution of a fully working neighboring roll (see fig. 3(b)). Reasons for dead regions in a roll are various: malfunctioning supply of readout electronics, faulty or disconnected cables, wrong channel mapping, or simply the region was intentionally masked because affected by other problems. In all cases, investigation by a detector expert is required.
The examples described above are only two of the possible scenarios in which prompt feedback from the DQM can be used to give fast and accurate feedback of possible problems or discrepancies.

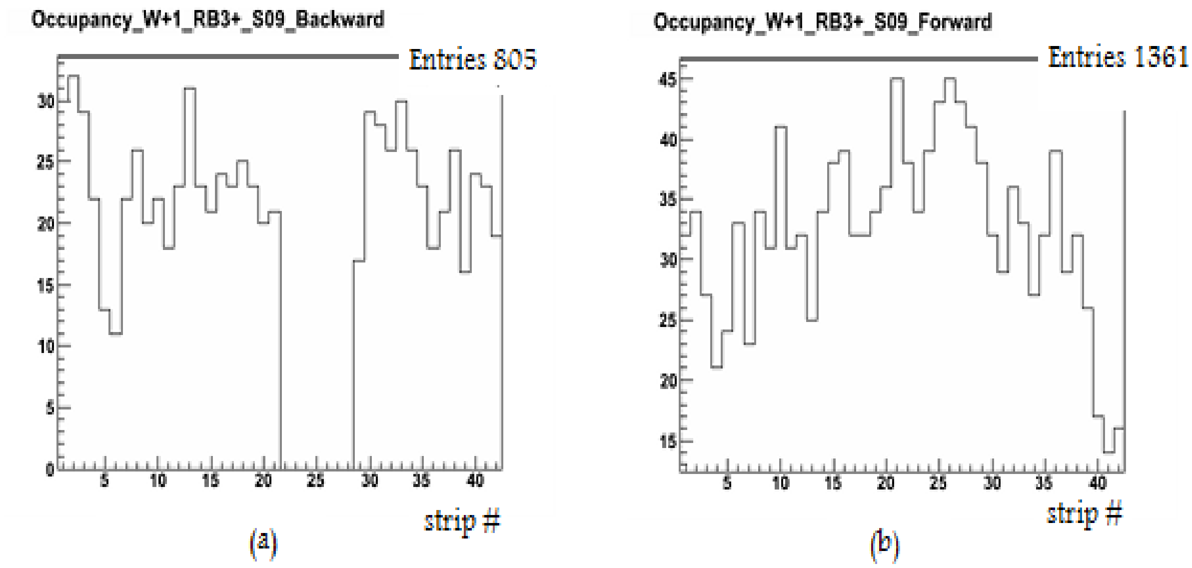


Figure 3: The figure on the left (a) shows the occupancy plot for the given roll in the Barrel region. A dead region is clearly present. The plot on the right (b) shows the occupancy of a fully functioning neighboring roll.

4 RPC DQM Structure

The RPC DQM system has a source/client structure, as shown in fig 4. The *source* modules access information from the events, define the quantities to be monitored, and fill histograms. Histograms are defined for each roll or for larger detector segments, such as the the sectors (60) and wheels (5) in the barrel and rings (12) and disks (6) in the forward regions (see section 1). Event selection is performed at this level. In the online environment, the RPC performance group decided to accept all single and double muon trigger path and, for early data only, minimum bias events too. The selection is done upstream of all source modules and is externally configurable. There is no need to change monitoring code. The offline applications instead run on the full statistics of every dataset sample processed in Tier-0 and in the Tier-1s.
The *client* modules access the histograms periodically and perform analyses. The frequency of the access depends on the monitored quantity, varying from every luminosity section (23s) to once a run. Clients have the tasks of: creating summary histograms, performing quality tests, calculating alarm levels, saving the output in ROOT files, and taking a preliminary data certification decision. In addition, summary plots are created by the clients by combining information from individual low level histograms. Summaries give a general and fast overview on the entire RPC system. This two-step process design is used both in online and offline DQM environments. Commissioning of the system during cosmic data taking and early collision data has shown that the online algorithms could process data up to a rate of 40 Hz. The offline modules were successfully integrated in the standard CMS software releases, proving to be stable and robust.
A total of $\sim 1.7 \cdot 10^3$ histograms are produced in the online environment and $\sim 3 \cdot 10^4$ in the offline. All information is stored to a ROOT file, intended for expert use, and uploaded to the central GUI. A screen shot of the web GUI with a selection of RPC DQM plots is shown in fig.5.

Histograms are organized in a hierarchical tree-like folder structure reproducing detector geometry, starting from the overall RPC detector down to the single electronic channel. Since navigating through folders and several thousand histograms is complicated for non-experts, special layouts containing only summary histograms are prepared for both RPC and central DQM shifters. Purpose of shifter histograms is to allow the shift crew to quickly identify problems and take action. The histograms are meaningful, not overwhelmed with information and equipped with a clear set of instructions. Reference histograms may be superimposed and quality tests are applied. Checks are performed on the fraction of dead channels, the average cluster size for each chamber η-partition, synchronization, and eventual presence of Front End Driver (FED) errors[10]. The tests are repeated every 5 luminosity sections ($\sim$ 115 seconds) and alarm states are displayed on the GUI. Thus the operator is promptly informed about potential inconsistencies.

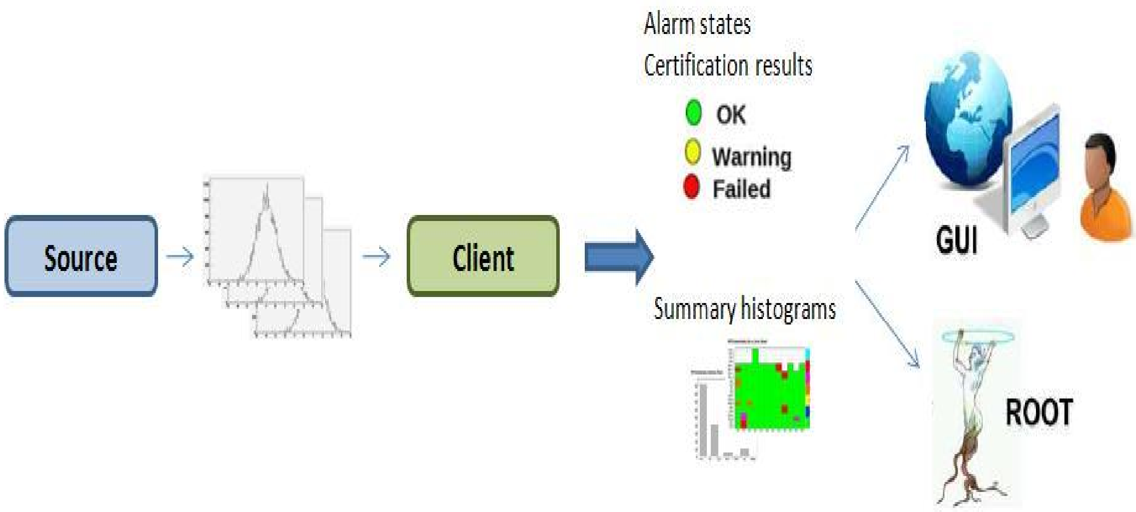


Figure 4: Schema of the source/client structure of the RPC DQM.

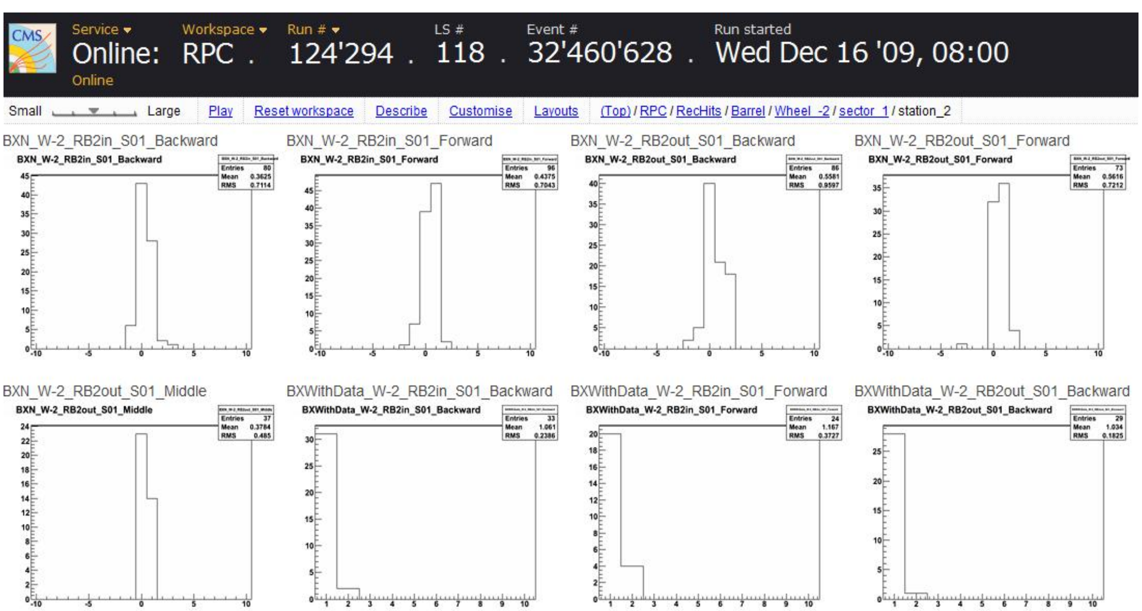


Figure 5: Screen shot of the DQM GUI showing a selection of RPC histograms.

5 Quality State Machine

The CMS RPC system is complex and massive. These characteristics require a high level of automation in monitoring processes, to reduce human errors and optimize recovery procedures. But automation comes with the need to describe the behavior of the system in an accurate way. For these reasons, the RPC performance group developed, within the compass of the DQM framework, a fully automated procedure for debugging hardware and assessing the state of the detector. The algorithm borrows many features from finite state automata. The first step of the implementation was defining seven detector states: Good, Off, Dead, Partially Dead, Noisy Strips, Noisy Roll, Bad Occupancy Shape. A costum analysis algorithm performs a series of subsequent tests (see fig. 6) to

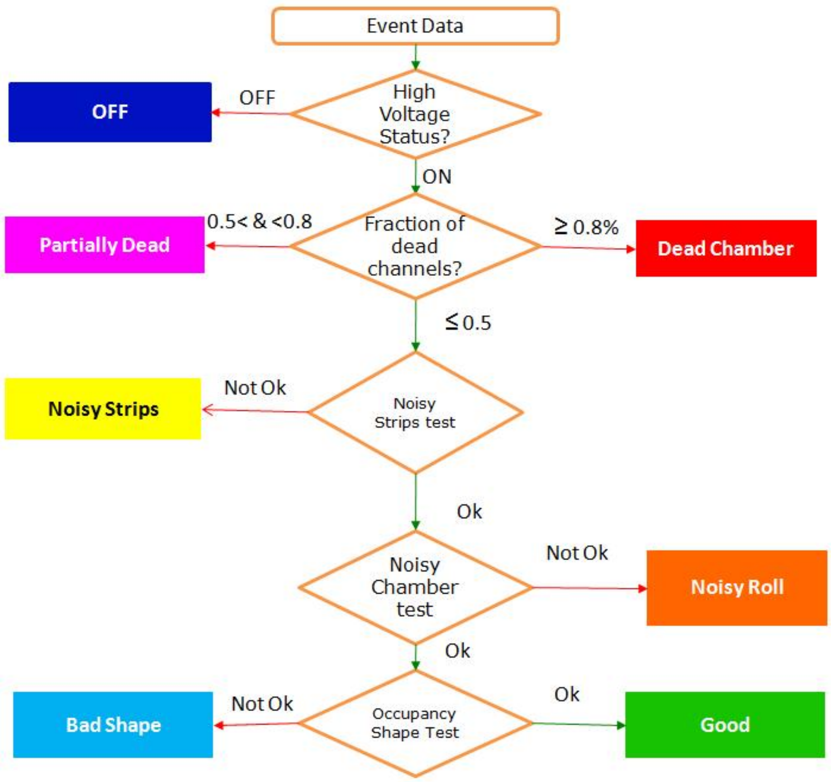


Figure 6: Scheme of the RPC certification workflow.

determine each roll's status. First of all it checks if the system is at its working point by reading Detector Control System (DCS) [7] information in the event data. Subsequently it checks if the fraction of non responding channels in a roll is under a given threshold (between 0.5 and 0.8 the roll is declared Partially Dead, while anything higher is considered Dead). The algorithm then tries to spot the possible presence of noisy strips. Careful studies have shown that if 85% or more of the clusters in a roll have size 1, noisy strips are surely affecting detector behavior. In addition, a strip is declared noisy if its occupancy exceeds by a given factor the average occupancy per strip in that rolls. This factor is currently set to 3.5. The next step is to evaluate if the roll, as a whole, is noisy. A roll is said to be noisy when its average multiplicity is ≥ 6. The last step checks the shape of occupancy distributions. An asymmetric distribution fails the test. The percentage of asymmetry currently admitted is less than 30%. If all tests are passed successfully, the roll is declared "Good". When a roll is found to be in a given state, the algorithm exits and moves to the evaluation of the next roll. Thus, this analysis is light weight and respects time limits set for processing. The algorithm was extensively tested during early commissioning phases. Using cosmic data we were able to determine the range of validity of each test and the value the parameters used as cuts. All parameters are externally configurable. Only runs where detector conditions were known and the status of each chamber had been verified were used. The algorithm results are displayed in eleven 2-D histograms representing major detector elements, as shown in the fig. 7. Distributions of the number of rolls that fall in each of the seven categories are also produced. Fig. 8 gives at example of such distributions

Figure 7: This figure shows the results of the RPC certification algorithm for Wheel 0 during a collision run taken in 2010. The x-axis represents the 12 sectors in which the wheel is divided in the azimuthal angle ϕ, while rolls are displayed on the y-axis. The numeric/color code is: 1 = Good, 2 = Off, 3 = Noisy Strips, 4 = Noisy Roll, 5 = Partially Dead Roll, 6 = Dead Roll, and 7 = Bad Occupancy Shape.

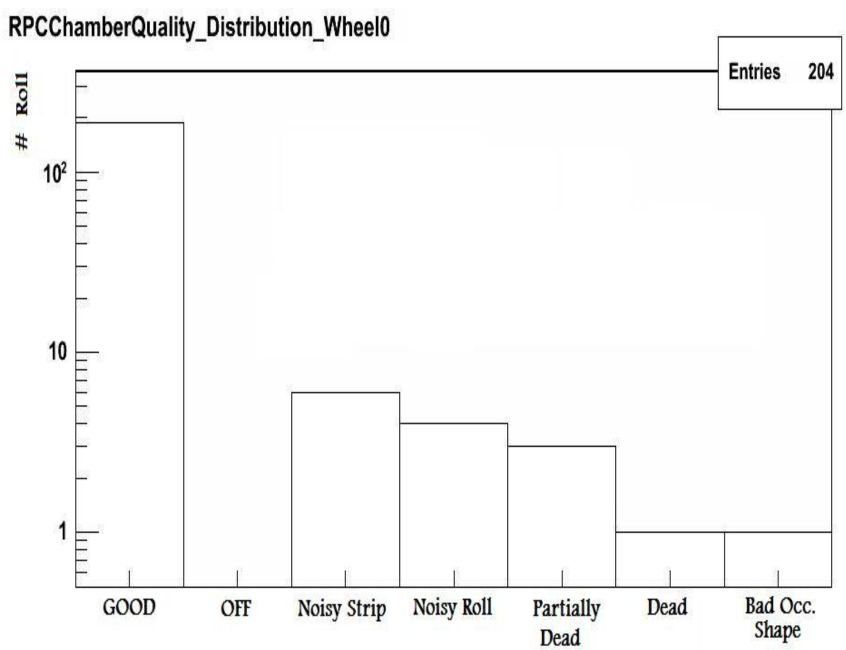


Figure 8: This figure is the 1D representation of the 2D-histogram in fig. 7.

6 Data Certification

Data certification [11] for RPCs is mainly based on the information coming from DQM online and offline but uses DAQ data too. The procedure can be divided in automatic and manual certification.
Automatic certification is based on the results of standard quality tests applied to the occupancy distributions of each roll. The fraction of alive channels is computed and weighed by geometrical considerations. The ultimate result is a float number between 0 and 1 reflecting detector performance and a quality flag, i.e. “good”, “bad”, and in case no quality calculations were performed “unknown”. Following CMS specification, the quality flag is set to “bad” and expert intervention is required when the floating point value is beneath 0.95. Both flags are assigned to various detector segments. The RPC community chose to assess quality results at the granularity of the sector. There are 60 sectors in the barrel region and 36 in the endcaps. The quality information is displayed on the GUI as a list of floats and summarized in user friendly 2-D histograms (see fig. 9). In addition, a DAQ quality flag is calculated. This flag represents the percentage of allocated Front End Drivers (FEDs).
Manual certification is performed both online and offline by central DQM shifters. The shifters are asked to

Subsystem	Summary	Run	Lumi section	Event	Last update	Last event	Processed events	Event rate	Monitor elements
RPC	96.7%	124230	68	24500042	Dec 15, 01:55.35	Dec 15, 01:55.34	178474	13.34	16807
Disk-1	100.0%								
Disk-2	100.0%								
Disk-3	100.0%								
Disk1	100.0%								
Disk2	100.0%								
Disk3	100.0%								
Wheel-1	95.9%								
Wheel-2	97.1%								
Wheel0	97.5%								
Wheel1	96.9%								
Wheel2	96.3%								

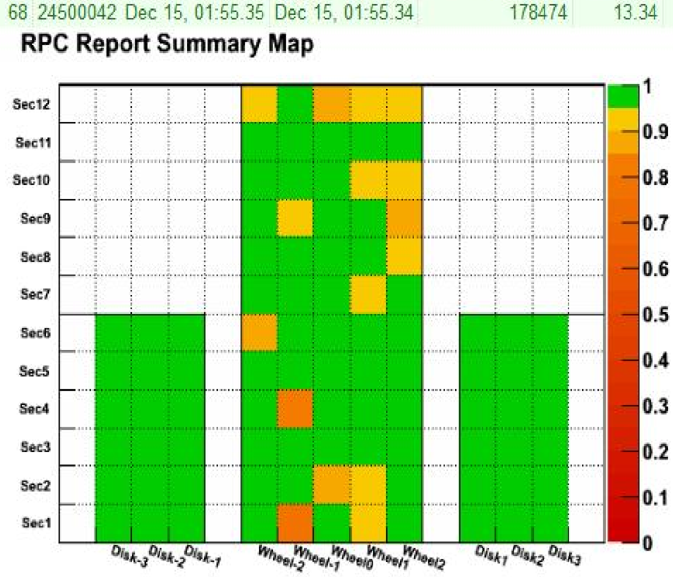


Figure 9: The automated certification results as displayed by the DQM GUI.

look at a limited number of summary histograms, produced by the quality state machine (section 5) and report on possible inconsistencies or problems. Instructions are available to facilitate this task. The shifter also monitors FED errors. In case of fatal errors, such as wrong FED IDs or inconsistent data size, the system is immediately flagged as bad. Offline shifters are also asked to monitor chamber efficiencies. An efficiency higher than 90% is considered “good”.
Both during commissioning phases of the CMS experiment and during early LHC operations, the emphasis was on the manual run-by-run certification, which relies on visual inspection of histograms. Presently, strong efforts are being made to move to the automatic certification with higher time resolution keeping the manual one as a control instance.
The end result is a list of “good” and “bad” runs, which is reviewed by detector experts. Weekly meetings are held to collect the final decisions and communicate them to the entire CMS collaboration. At this stage quality flags are copied to the offline condition database and to the Dataset Bookkeeping System (DBS) [12]. During the year 2009, RPC DQM certified 1979 runs, 547 of which are in presence of beams. The detector was flagged as ”good” in 89% of the runs, “bad” in 8% of them. Certification procedures were not applied in the remaining 60 runs mainly due to lack of statistics or to the failure of DQM applications. All “bad” runs have been reviewed by an RPC expert. Five of them showed synchronization or trigger configuration problems, 50 presented low chamber efficiencies ($< 90\%$), 5 had FED errors, and 10 presented high noise levels. Thirty runs exhibited occupancy distributions different from expected. This was mainly due to experts working on system configurations. In 4 runs high voltage

was off, while in 20% of the “bad” runs the system was in safe mode (low voltage on, but high voltage in standby) because of instable beams. Therefore, the detector was operating as expected but not in data taking conditions. Finally, in 28 runs DQM applications problems caused automatic certification to fail and the quality flag was set erroneously to “bad” by the central DQM shifter. A clearer set of shifter instruction have been produced to avoid this mistake in the future.

7 Conclusions

The RPC DQM system has been heavily tested during detector conditions and data taking. The system was used by both experts and shifters. Problems in the detector, FEDs and trigger were promptly spotted. Experts were able to use DQM information to find the origin of such problems. The shifter instructions and layout proved to be clear and user friendly. Finally the data certification procedure successfully certified the quality of more than $1.9 \cdot 10^3$ runs. Results, confirmed by experts were written to the offline database and DBS.
Development of the system will continue. Presently we are introducing conditions data in the certification algorithm, we are working on the implementation of trend plots, i.e. plots of the evolution in time of the average significant parameters, and we are studying new criteria for event selection.

References

[1] M. Abbrescia et al., “The RPC system for the CMS experiment at the LHC”, Nucl. Instrum. Methods A 508 (2003) 137-141.

[2] The CMS Collaboration, “The Compact Muon Solenoid Technical Proposal” (CERN/LHCC-94-38), CERN, 1994.

[3] L. Tuura, A.B. Meyer, I. Segoni and G. Della Ricca “CMS data quality monitoring: systems and experiences” Proc. CHEP09, Computing in High Energy Physics, (Prague, Czech Republic) 2009.

[4] CMS Collaboration, “CMS: The computing project. Technical design report”,. CERN-LHCC-2005-023.

[5] CMS Collaboration, “Commissioning of the CMS High-Level Trigger with cosmic rays, arXiv:0911.4889 [physics.ins-det]

[6] L.Tuura, L. Eulisse, A.Meyer, “CMS Data Quality monitoring web service”, Proc. CHEP09, Computing in High Energy

[7] P. Paolucci, G. Polese, R. Gomez-Reino, C. Viviani, H. Shahzad, T. Khurshid “The Detector Control Systems for the CMS Resistive Plate Chamber”, CMS CR -2009/136, 2009.

[8] V. Briglijevic et al., “Run Control and Monitor System for the CMS Experiment, CERN,Geneve,Svizzera.

[9] ROOT - A data analysis framework, 2009, http://root.cern.ch.

[10] The CMS Collaboration, “The Trigger and Data Acquisition project, Volume2 - Technical Design Report”, CERN/LHCC 2002-26, CMS TDR 6.2 (2000).

[11] CMS Collaboration “CMS Data Processing Workflows during an Extended Cosmic Ray Run”, arXiv:0911.4842v2 [physics.ins-det].

[12] A. Afaq et al., “The CMS Dataset Bookkeeping Service”, Journal of Physics: Conference Series 119 (2008) 072001.