

RelMon: A General Approach to QA, Validation and Physics Analysis through Comparison of large Sets of Histograms

Danilo Piparo

CERN, CH-1211 Genve 23, Switzerland

E-mail: danilo.piparo@cern.ch

Abstract. The estimation of the compatibility of large amounts of histogram pairs is a recurrent problem in high energy physics. The issue is common to several different areas, from software quality monitoring to data certification, preservation and analysis. Given two sets of histograms, it is very important to be able to scrutinize the outcome of several goodness of fit tests, obtain a clear answer about the overall compatibility, easily spot the single anomalies and directly access the concerned histogram pairs. This procedure must be automated in order to reduce the human workload, therefore improving the process of identification of differences which is usually carried out by a trained human mind. Some solutions to this problem have been proposed, but they are experiment specific. RelMon depends only on ROOT and offers several goodness of fit tests (e.g. chi-squared or Kolmogorov-Smirnov). It produces highly readable web reports, in which aggregations of the comparisons rankings are available as well as all the plots of the single histogram overlays. The comparison procedure is fully automatic and scales smoothly towards ensembles of millions of histograms. Examples of RelMon utilisation within the regular workflows of the CMS collaboration and the advantages therewith obtained are described. Its interplay with the data quality monitoring infrastructure is illustrated as well as its role in the QA of the event reconstruction code, its integration in the CMS software release cycle process, CMS user data analysis and dataset validation.

1. Validation of the physics performance of high energy physics experiments

A fundamental aspect of the operations of a modern high energy physics (HEP) experiment is represented by the validation of the physics performance of its software. Several types of changes are regularly introduced in the chain that leads to the delivery of experiment's software deployed at the computing centres for data processing and analysis. They can consist for example in algorithmic improvements in the objects reconstruction, simulation or trigger, in code performance optimisations or upgrade of basic components involved in the software building cycle like the compiler. The effect of these variations on the characteristics of the physics objects used for data analysis must be continuously and carefully assessed. At least two products are necessary to achieve this goal: a high quality data quality monitoring (DQM) infrastructure allowing to collect in a collection of histograms the relevant properties of every dataset and a tool to compare different sets of histograms and estimate their level of compatibility. RelMon [1] is a possible experiment independent answer to this latter need.

2. The RelMon tool

Relmon is a general tool that allows to compare large sets of histogram pairs according to a statistical test defined by the user. It is implemented in about 4000 lines of Python and depends only on ROOT [2]. The tool provides handy command-line interfaces to manage the comparisons. The histograms that are analysed can be provided as ROOT histograms stored in ROOT files, also organised in a directory structure and the matching of corresponding histograms is done by name. The tests provided by RelMon to check the compatibility of two histograms are three: chi-squared, Kolmogorov-Smirnov and bin-to-bin. The first two are well known statistical procedures [3] and the tool relies on their implementation provided by ROOT. The bin-to-bin test is not a “traditional” statistical test, but is useful in case the identity of two sets of histograms is to be checked. The content of corresponding bins of each of the two histograms are checked to be identical and the outcome of the test is the fraction of bins which are different. The aforementioned tests are considered to fail if their p-values are greater than a value set by the user. In some real-life cases, it is impossible to apply these tests given a pair of histograms, for example when the number of bins or the axes’ limits are not the same. These exceptions are automatically handled by RelMon and labelled as “null” comparisons. All the information about the amount of succeeding, failing or null comparisons is aggregated per directory.

The result of the comparison of two sets of histograms can be displayed in two ways: either with a plain ascii report printed on screen or with an elegant web report (figure 1) where all the plots of the histograms overlays are displayed.

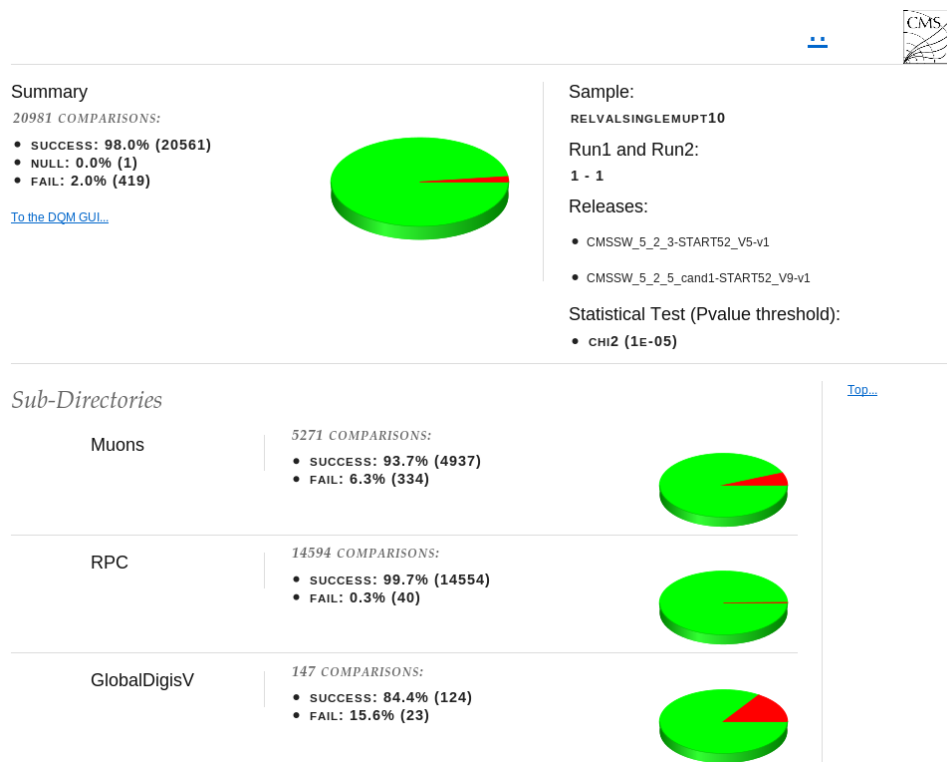


Figure 1. An example of web report generated with RelMon. Pie-charts are used to aggregate the overall rate of failure and success within a directory. With a click on the piechart is possible to inspect the contents of the directory. Useful meta data are shown on each page of the report to give a context to the displayed overlays.

3. RelMon in CMS

In this section, after a brief outline of the CMS Data Quality Monitoring (DQM), different usages of RelMon within the CMS collaboration are characterised.

3.1. Data quality monitoring

During the years, the CMS collaboration developed a high quality DQM infrastructure for the certification of the acquired and Monte Carlo generated data [4]. The underlying principle is to append monitoring modules to the algorithms chain used for regular data processing. These modules allow to collect in histograms all the relevant quantities both for the Monte Carlo and real data samples. The CMS framework then takes care of writing such histograms in dedicated ROOT files.

In order to efficiently make available to the physicists the enormous amount of histograms deriving from the operations of the CMS detector, DQM GUI servers are provided [5]. The ROOT files obtained running monitoring modules are automatically uploaded to the DQM servers which further process their content and store it in an internal database for performance reasons. A web interface can be accessed in order to browse all the histograms contained in a given DQM server: the DQM GUI (figure 2).

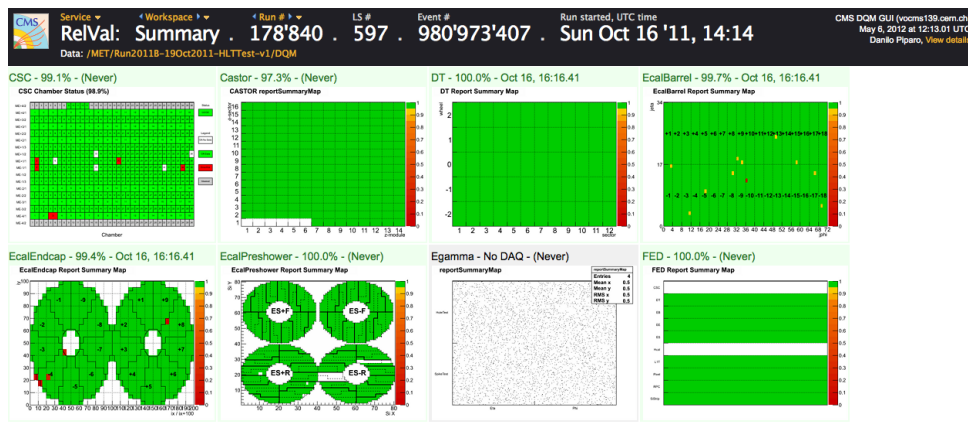


Figure 2. A snapshot of the CMS DQM GUI. All the histograms indexed in the internal DQM server database can be accessed through it.

Advanced manipulations of the stored histograms can be performed directly within the browsers like axes zooming, logarithmic scaling of axes and histograms overlays to ease datasets comparison. These features are offered by a component called “plot-fairy”. RelMon can produce plots of histograms overlays as images on disk taking advantage of the well-known ROOT graphics routines or directly interfacing the web-reports to the DQM GUI and exploit the plot-fairy to obtain the comparison plots as shown in figure 3.

3.2. Software integration builds

One of the fundamental building blocks that leads to the delivery of high quality releases of the CMS software (CMSSW) is the integration build process adopted by the collaboration. Twice a day, the code candidate to be part of an official CMSSW release is compiled on different platforms. This allows to perform important sanity checks of the code but also to process small datasets and check the overall physics performance of the potential releases. Also in this case the DQM modules are ran and RelMon is executed to check the compatibility of the histograms therewith obtained with a known reference [6].

TrackerMuonsNotGlobalMuons



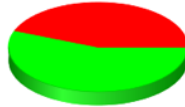
Muons/MuonIdentificationV

Summary

200 COMPARISONS:

- SUCCESS: 56.0% (112)
- FAIL: 44.0% (88)

[To the DQM GUI...](#)



Sample:

RELVALTTBAR

Run1 and Run2:

1 - 1

Releases:

- CMSSW_5_0_0_pre4-START50_V3-v1
- CMSSW_5_0_0_pre4-START50_V3_FastSim-v1

Statistical Test (Pvalue threshold):

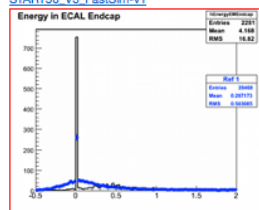
- CHI2 (1E-05)

Failing Comparisons

hEnergyEMEndcap

Chi2: 0.00E+00

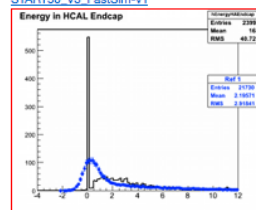
[CMSSW_5_0_0_pre4-START50_V3-v1](#)
[CMSSW_5_0_0_pre4-START50_V3_FastSim-v1](#)



hEnergyHAEndcap

Chi2: 0.00E+00

[CMSSW_5_0_0_pre4-START50_V3-v1](#)
[CMSSW_5_0_0_pre4-START50_V3_FastSim-v1](#)



hCaloCompat

Chi2: 1.80E+283

[CMSSW_5_0_0_pre4-START50_V3-v1](#)
[CMSSW_5_0_0_pre4-START50_V3_FastSim-v1](#)

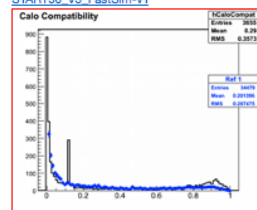


Figure 3. An example of histogram overlay accessible with RelMon.

This procedure represents a first line of defence against unexpected behaviours of the CMS software or bugs introduced by the developers.

3.3. Reconstruction, simulation, calibration constants and trigger

Starting from successful integration builds, CMSSW pre-releases are provided once a week in order to consolidate the state of the code, test interdependencies among software components while releases are cut approximately once per month. They close a development cycle and then are used for central processing, Monte Carlo production and analysis. For every (pre)release, a complete set of release validation datasets is provided which includes Monte Carlo simulated and re-processed proton-proton collisions events [7]. Experts of all CMS sub-detectors, physics objects and physics analysis groups scrutinise these datasets against the ones of the preceding (pre)release in order to obtain a solid incremental validation of the features introduced in the CMS software for simulation, reconstruction and trigger or the calibration constants. The main tool used to perform these regressions is RelMon.

A centralised regression is provided, which considers all the datasets of the release validation collection. Not only comparisons from successive releases are provided, but also a comparison of fast and full simulation [8]. The latency between the arrival of the ROOT files containing the DQM histograms and the appearance of the RelMon report on the web can be of ~ 30 minutes, even though more than one million histograms are compared (see figure 4). This strategy exposes many potentially critical aspects of the physics output produced by the two releases: anomalies can be immediately pin-pointed.

Summary Table

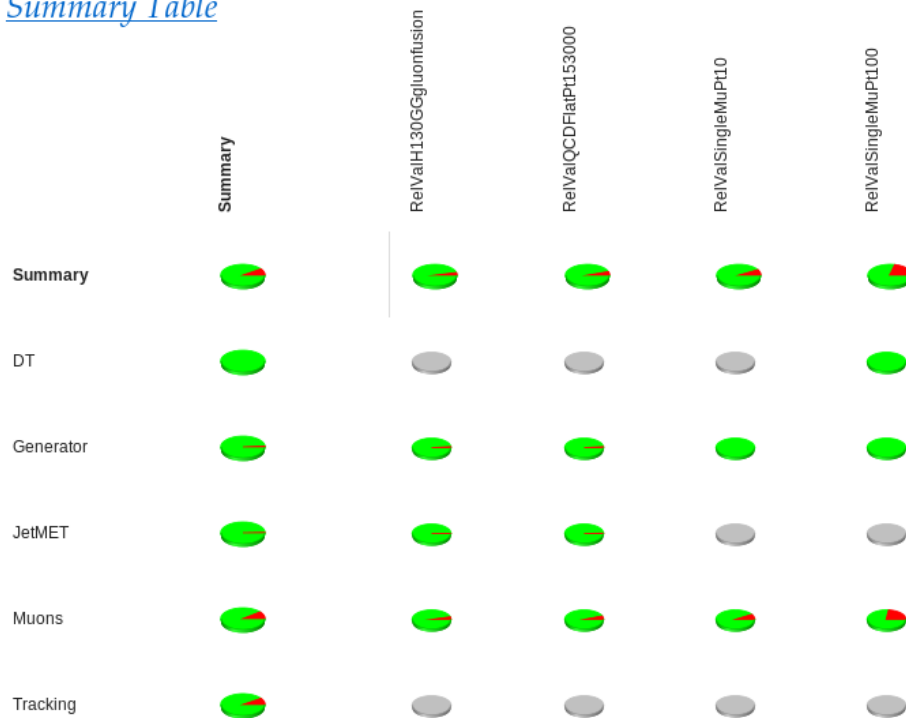


Figure 4. With RelMon it is possible to perform regressions of several pairs of files simultaneously and aggregate the information in a single report. CMS exploits this possibility to compare the DQM histograms sets coming from release validation samples created with successive releases to perform incremental validation. The picture shows one part of the full “validation matrix”, where all the CMS subsystems are compared using all the release validation datasets, pie-charts are used to summarise the information about the histograms comparisons. The gray pie-charts represent comparisons not executed according to a blacklist compiled by subsystem experts.

3.4. Software performance optimisations

The processing of the complex data acquired by large HEP experiments requires high performance scientific software. For this reason, the CMS collaboration puts an enormous effort in the continuous improvement of the CPU performance of its software without affecting the quality of the detector’s measurements. This process implies a precise scrutiny of the results obtained with the improved software: the quality of the measured data cannot be jeopardised in any way.

RelMon can help in this process, with the bin-to-bin test. CMS took advantage of this feature in several occasions, for example during the transition from the GCC compiler collection [9] from the 4.3 to the 4.6 version or when the jemalloc [10] allocator was adopted.

3.5. User analysis

A possible use case in the field of data analysis where RelMon is of great help is the one of the assessment of the quality of the description of the data by the Monte Carlo simulation. Two sets of histograms must be provided, one obtained from data and one from the simulation.

One of the main weaknesses of data analysis code is the fact that its robustness and sanity is sometimes not thoroughly checked, principally because of the tight schedules involved by analyses that can lead to a discovery and because it is not always included in the official software release

of the experiment. An analysis group can take advantage of RelMon even in this context. For example, checking the compatibility of the results produced by an analysis on a known dataset before and after new developments, as the CMS $Z(\rightarrow \mu\mu) + jet$ calibration workgroup [11].

4. Conclusions

RelMon is a fast and easy to use tool depending only on ROOT, which allows to assess the compatibility of large sets of histogram pairs producing easily accessible reports.

In combination with its DQM infrastructure, CMS adopted RelMon as a part of the CMS software integration build cycle, to implement automatic quality assurance and testing of detector conditions and code for the simulation, reconstruction and trigger. In addition, RelMon proved to be an extremely powerful tool to track down unexpected features introduced in the code during campaigns of software performance optimisation. Furthermore, the assessment of the Monte Carlo simulation description of the acquired data and the test of compatibility between different kinds of simulations can rely on RelMon.

References

- [1] <https://twiki.cern.ch/twiki/bin/view/CMSPublic/RelMon>.
- [2] Rene Brun et al. Root - an object oriented data analysis framework. *Nucl. Inst. Meth. In Phys.*, 389:81–86, 1997.
- [3] Glen Cowan. *Statistical data analysis*. Oxford Oxford Univ. Press, 1998.
- [4] The CMS Collaboration. CMS data quality monitoring: Systems and experiences. *J. Phys. Conf. Ser.*, 219, 2010.
- [5] The CMS Collaboration. CMS data quality monitoring web service. *J. Phys. Conf. Ser.*, 219, 2010.
- [6] <http://cmssdt.cern.ch/SDT/html/showIB.html>.
- [7] The CMS Collaboration. Validation of software releases for CMS. *J. Phys. Conf. Ser.*, 219, 2010.
- [8] The CMS Collaboration. Comparison of the fast simulation of CMS with the first lhc data. (CERN-CMS-DP-2010-039), Oct 2010.
- [9] <http://gcc.gnu.org>.
- [10] <http://www.canonware.com/jemalloc>.
- [11] Thomas Hauth Joram Berger. A procedure for authomatic q/a of jet calibration code. private communication, 2011.