

Data preparation for the Compact Muon Solenoid experiment

Roberto Castello*

Centre for Cosmology, Particle Physics and Phenomenology

Université Catholique de Louvain

Louvain-la-Neuve, Belgium

E-mail: Roberto.Castello@cern.ch

During the first 3 years of operation at the Large Hadron Collider (LHC), the Compact Muon Solenoid (CMS) experiment has collected data across evolving conditions of center of mass energy, instantaneous luminosity and collisions pile up. Following this evolution, the CMS collaboration has constantly strived to guarantee the prompt availability of high quality reconstructed data, in order to ensure early and sound physics results. This has relied on a few key areas of constant attention covering: careful preparation and maintenance of the event simulation and reconstruction algorithms; efficient and robust strategies and algorithms for the calibration and the alignment of the detector elements; continuous scrutiny of the data quality and the validation of any changes to the software or calibrations which were introduced during the operations. This contribution covers the major development and operational aspects of the CMS offline workflows during the 2010-2013 data taking, underlying its essential role towards the main physics achievements and discoveries of the CMS experiment.

The European Physical Society Conference on High Energy Physics -EPS-HEP2013

18-24 July 2013

Stockholm, Sweden

*on behalf of CMS collaboration

1. The CMS experiment and the data preparation

The scientific program of CMS experiment [1] at the LHC covers a very broad spectrum of physics processes, focusing on the search for new phenomena in the TeV range. Excellent performance is crucial for reaching these scientific goals, which places very high demands on the level of precision of the calibration and alignment of the detector devices and on the systematic control and validation of the performance itself. One of the main challenges of CMS detector during the data taking consisted into promptly detecting the final decay products of known and new particles arising from the collision of protons at a centre of mass of 7 and 8 TeV (muons, electrons, hadrons, photons and taus), measuring their properties and reconstructing them as objects suitable for physics analysis. This data preparation process consists of several steps: based on events surviving the online High Level Trigger (HLT) selection, different reconstruction streams are activated, which are tightly coupled to the needs of workflows for data validation and detector monitoring and calibration. Beside this CMS target is also to provide a realistic simulation of the physics events, to be used for a comparison with the collected real data.

2. The data streams

Three main streams of data are processed on a dedicated CERN farm. An **express** stream (40 Hz bandwidth), shared among quasi-online calibration (50%), detector monitoring (25%) and physics monitoring (25%), is meant to be used for prompt feedback and beamspot measurement, providing reconstructed objects with a latency of 1-2 hours. A **calibration** stream is then entirely devoted to detector alignment and calibration purpose. Finally, a **physics** stream represents the bulk of data for physics analysis, the majority reconstructed promptly 48 h after, while others parked for a time-delayed reconstruction. Ten billion of events have been processed during 2011 and 2012 data-taking and distributed over 460 non-exclusive datasets (based on HLT filtering).

3. Data Quality Monitoring

The massive and continuous streaming of the data collected by such a complex detector requires a systematic control of the quality of the reconstructed data throughout the various steps. A robust Data Quality Monitoring (DQM) infrastructure allows for production of automated comparison of detector, object and analysis oriented distributions to be scrutinized. A dedicated **online** DQM is designed for a fast monitoring of detector performance during data-taking. Beside this, an **offline** DQM monitors the performance of reconstructed detector and physics quantities in the express and prompt streams, allowing for a fast feedback on the calibrations deployed on-line, providing a continuous monitoring (24 hours, 7days) of the performance. The same infrastructure is then also used offline for validation of software updates or for alignment and calibration updates derived with offline procedure aiming to provide better reconstruction of the data.

Finally, the sign-off of datasets valid for physics analysis is also based on the DQM monitoring of physics objects and sub-detectors status and it is performed at a level of granularity called Lumi Section (LS, corresponding to roughly 23 seconds of data-taking). The resulting certification efficiency for the data collected during pp collision was around 91% of CMS recorded luminosity and around (97) 94% during the (p)PbPb collisions.

4. Detector calibration, alignment and physics validation

According to the specific needs, several workflow for the calibration and alignment of CMS detector have been activated. A **quasi-online** one, for HLT and express stream calibration. Using a DQM-dedicated stream (sampling at 100 Hz) the beamspot measurement is provided, using a track based procedure together with a pixel-only vertexing. This calibration allows for a very fast turnaround, providing a value every 5 LS (about 2 minutes).

For the case of conditions changing on a short time scale, a special calibration workflow has been designed to allow updates with very short latency: the **prompt** calibration and alignment loop exploits the 48 hours delay between express and prompt reco at Tier-0, meaning that conditions are updated for a given run while the bulk of the data is buffered on disk. The most common workflow which are run at this stage are: the determination of the beam-spot position (measured every LS), the masking of Tracker problematic channels and the correction of sizeable movements of Pixel Barrel (BPIX) large structures (Fig. 1), potentially affecting vertexing and b-tagging.

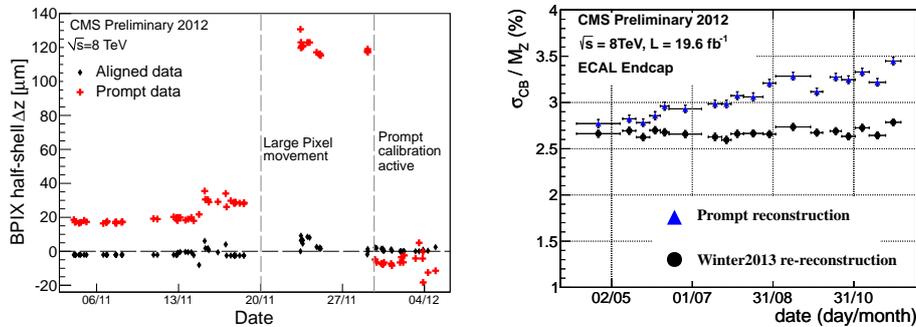


Figure 1: Left: Day-by-day value of the relative longitudinal shift between the two half-shells of the BPIX as measured with residuals between track and its primary vertex. The effect of correcting in the prompt (red marks) for the changes in the positions of the half-shells is visible after 29/11. Right: example of offline workflow for the calibration of electromagnetic calorimeter. Mass resolution of the Z peak is reconstructed from its di-electron decay mode, as a function of time.

Subsequently, a set of **offline** calibrations is run, aiming to provide more stable and improved conditions for collected data. The major updates concern alignment geometries (Tracker, Muon, ECAL) together with their inter-dependencies (beam-spot, b-tagging template probabilities) and calibration of calorimeter crystals, using different methods, like invariance of energy flow at fixed pseudo rapidity and corrections using Laser optical response (Fig. 1). Workflows are running taking calibration stream as input, while the output is stored on Oracle database. In general, selection of input events for the alignment and calibration purposes is tuned according to the needs, meaning that event content is reduced to optimize bandwidth/disk space usage (AICaReco format).

Finally a set of automated workflows and procedures for a robust validation of software and data-MonteCarlo reconstruction has been put in place. After a central re-reconstruction of datasets with improved conditions, an automated comparison and centralized scrutiny of physics performance represent the last steps towards the delivery of excellent data quality for the CMS physics analysis.

References

- [1] R. Adolphi et al. [CMS Collaboration], The CMS experiment at the CERN LHC, JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004
- [2] G. Cerminara, Alignment and Calibration of the CMS Detector, PoS EPS-HEP2011 (2011) 186
- [3] CMS collaboration, CMS-DP-2013/007
- [4] CMS collaboration, CMS-DP-2013/017