Available online at www.sciencedirect.com
ScienceDirect
Nuclear Physics B Proceedings Supplement 00 (2014) 1–6
Nuclear Physics B Proceedings Supplement

CMS Alignement and Calibration workflows: lesson learned and future plans

Federico De Guio on behalf of the CMS Collaboration

federico.de.guio@cern.ch

Abstract

We review the online and offline workflows designed to align and calibrate the CMS detector. Starting from the gained experience during the first LHC run, we discuss the expected developments for Run II. In particular, we describe the envisioned different stages, from the alignment using cosmic rays data to the detector alignment and calibration using the first proton-proton collisions data (O(100 pb^{-1})) and a larger dataset (O(1 fb^{-1})) to reach the target precision. The automatisation of the workflow and the integration in the online and offline activity (dedicated triggers and datasets, data skims, workflows to compute the calibration and alignment constants) are discussed.

Keywords: CMS, Calibration, Alignment

1. Introduction: The Calibration Challenge

The design of the CMS detector [1] is based on a large super-conducting solenoid providing an intense (3.8 T) magnetic field, a high-precision silicon tracking system composed of about 76 millions channels including pixel and strips and hermetic calorimetry including a homogeneous Electromagnetic Calorimeter (ECAL) consisting of about 76000 $PbWO_4$ scintillating crystals. The return yoke of the solenoid houses a muon spectrometer used both for trigger and for tracking purposes.
The high level of complexity and the large number of detector channels reflect in an elaborated structure for the management and computation of the detector calibration and alignment. The present contribution reviews the workflows developed for this purpose focusing on a few selected examples.

2. Alignment and Calibration workflows

Most of the alignment and calibration workflows are fed with dedicated data samples, called AlCaReco, optimized both in terms of event selection and event content. Depending on the needs of the specific workflow, these samples can be selected offline, while performing the reconstruction, or directly online, at the High Level Trigger (HLT) level. The great flexibility of the HLT, which runs offline-quality software on a farm of commercial processors, is a key asset for this online selection since it guarantees an adequate rate of events that would not be selected by the standard trigger paths meant for physics analysis.

The CMS software and reconstruction framework accommodates time-dependent alignment and calibration conditions by "intervals of validity" (IOV), which are periods during which a specific set of constants retain the same values. Conditions changing on a short time scale require a special calibration workflow designed to allow updates with very short latency. The handling of the data streams in the first step of the offline processing on the Tier-0 farm at CERN reflects this need and is organized as follows:

- express processing: reconstruction of a limited selection of data in order to give prompt feedback about the detector status and physics performance and to provide data for calibration workflows. The results of the express reconstruction for a given run are usually available one or two hours after the raw data are collected;
- bulk processing: reconstruction of the main data stream for physics analysis. This reconstruction

step, also called *prompt reconstruction*, is delayed by 48 hours to allow for the computation of the fast-changing conditions. The output is divided in several Primary Datasets (PD) on the basis of the HLT decision;

- calibration streams: streams of events selected at the HLT level and processed at Tier-0 for calibration purposes.

2.1. The Prompt Calibration Loop

During normal operation of the CMS experiment in Run I about 300-400 Hz of data are processed in the bulk processing. Only a limited bandwidth, corresponding to about 10% of the bulk, is allocated for express processing in order to guarantee a fast reconstruction.
A selection of data from the express and calibration streams is used to compute the updated conditions for a given run while the bulk of the data is buffered on disk. In this way the *prompt reconstruction* can profit from the updated constants, reducing the need for offline reprocessing of the data. This workflow is called the prompt calibration loop (PCL) and is illustrated schematically in Fig. 1.
The conditions currently updated through this kind of workflow are:

- measurement of the beam-line parameters;
- monitoring and masking of problematic channels of the silicon strip tracker to respond to HV trips or noise;
- transparency corrections to compensate for the transient radiation damages of the $PbWO_4$ crystals of the ECAL calorimeter (see section 4);
- monitoring of possible movements of large structures of the silicon tracker mainly due to thermal stress;
- monitoring of problematic channels in the electromagnetic and hadronic calorimeters allowing for quick reaction time in case of "hot" regions identified in the express reconstruction.

2.2. Offline Alignment and Calibration

In order to reach the ultimate accuracy, more sophisticated alignment and calibration workflows are run offline. No time constraints are present in this case and the full treatment of the detector's alignment and calibration inter-dependencies can be studied and taken into account. The full statistics is exploited to provide the final set of condition which is then used in the reprocessing of the data. From the Run I experience, the full data reprocessing happens no more than once a year.
In the following we will focus on a selection of workflows:

- energy calibration of the electromagnetic calorimeter response (single channel and overall energy scale calibration);
- measure and correction of the tracker orientation with respect to the magnetic field;
- tracker module alignment.

2.3. Quasi Online Calibration

A third class of calibration workflows, the quasi online calibration for HLT consumption, is meant to update conditions while taking data and the measurement of the beam line parameters falls in this category as well. An application running in the Data Quality Monitor (DQM) framework [2] is responsible of deriving and storing conditions to a database which is then accessed at the High Level Trigger during the online event reconstruction. The feedback from the DQM to the HLT happens with a delay up to about 2 minutes.

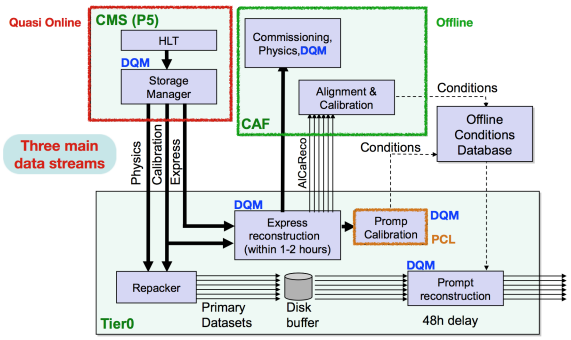

Figure 1: Schematic representation of the CMS prompt and offline calibration workflows.

3. Beam Spot Position Calibration

The measurement of the three-dimensional profile of the luminous region where the LHC beams collide at CMS is an important component of the event reconstruction being used as an estimate of the primary interaction point prior to the reconstruction of the primary vertex. The position of the center of the lumi-

nous region and its width are determined using two independent methods with complementary systematic uncertainties. The first uses the distribution of the reconstructed primary vertices to map the shape of the beam line; the mean three-dimensional position is determined with a 3D likelihood fit. The second method exploits the correlation between the transverse impact parameter and the azimuthal angle of tracks when the beam line is displaced from the expected position. With a sample of 1000 tracks, the position can be determined with a statistical precision of about 5 μm.
The fit is performed with a time granularity of one luminosity section (corresponding to 23 seconds of data taking) in the offline workflow and 5 luminosity sections in the PCL using tracks selected in the express stream. In a second step, ranges with stable parameters are "collapsed", increasing the statistical precision and reducing the database storage size. Finally, the calibration object is validated and uploaded to the database. This allows the best possible knowledge of the position of the luminous region within a few hours of data being collected, with a time granularity that allows possible movements during the fill to be followed, as shown in Fig. 2.

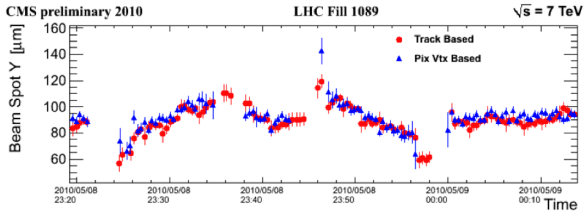

Figure 2: Fitted position of the beam line in the vertical plane as a function of time during a LHC fill where a luminosity scan was performed. The two sets of points correspond to independent methods used to determine the beam spot parameters as described in the text [3].

4. ECAL Calibration

Electromagnetic particles deposit their energy over several ECAL [4] crystals and the energy estimate implies a sum over the corresponding channels. Several factors contribute to the final reconstructed energy of electrons and photons. Among those a correction factor to each crystal signal amplitude is needed because of the radiation-induced response change with time.

4.1. Response variation with irradiation and monitoring of the performance stability

During LHC cycles, the ECAL response varies depending on irradiation conditions, which modify the transparency of each individual $PbWO_4$ crystal depending on its radiation hardness. The radiation damage, related to colour center creation during LHC fills, is recovered during inter-fills and technical stops. These effects take place on a time scale of hours and cause transparency changes of a few percent in the ECAL barrel. In the ECAL endcap at large $|\eta|$, where the dose rate is considerably higher, the observed loss reaches 50% at $|\eta| \sim 2.5$ (Fig. 3), the most forward electron acceptance in CMS. The transparency changes are monitored every 40 minutes by means of laser light injected into each crystal through optical fibres [5]. The corrections are computed online with crystal granularity and made available in time for the reconstruction of the bulk of the data exploiting the delay of the prompt calibration loop. Their availability before the *prompt reconstruction* processing starts is crucial to ensure a precise measurement of the energy of electrons and photons.

The energy measured by ECAL for isolated electrons from W and Z decays is used to monitor the stability of the response and resolution. The stability of the ECAL resolution in 2011, monitored with the relative instrumental width of the $Z \rightarrow e^+e^-$ invariant mass peak (Fig. 4), is excellent in EB while a slight degradation is observed in EE partially due to the increased number of collisions per beam crossing (pileup interactions) during the year. The instrumental contribution to the Z width (σ_{CB}) is extracted from a fit to the invariant mass distribution of a Breit-Wigner convoluted with a Crystal-Ball response function.

4.2. Single channel inter-calibration

The main sources of channel-to-channel response variation are the crystal light yield variation in the ECAL barrel, about 15% at construction, and the gain spread of the photodetectors in the ECAL endcaps, about 25%. To reduce this spread and provide an acceptable performance at startup, calibration procedures with different levels of accuracy have been adopted during the construction and commissioning phase of ECAL. The inter-calibration precision has been improved with several techniques exploiting the properties of collision events. These include the invariance around the beam axis of the energy flow in minimum bias events, the π_0/η mass constraint on the energy of the two photons from the $\pi_0/\eta \rightarrow \gamma\gamma$ decays, and the momentum constraint on the energy of isolated electrons from W decays. The precision of each method has been estimated

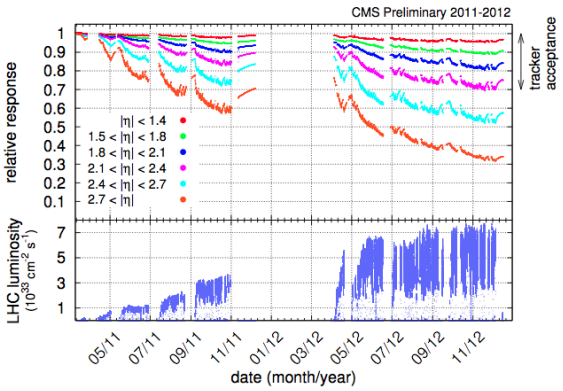

Figure 3: Relative response to laser light (440 *nm*) measured by the ECAL laser monitoring system, averaged over all crystals in bins of pseudorapidity, for the 2011 and 2012 data taking periods [6].

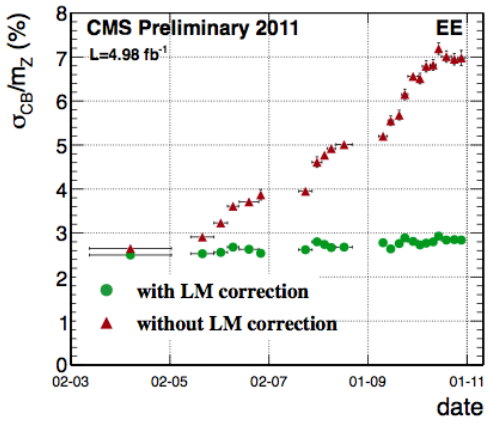

Figure 4: Stability of the ECAL resolution during 2011 with $Z \to ee$ events in the ECAL endcap. Red (green) points show data upon (after) the application of the laser corrections [6].

through the cross-comparison of the individual results, and cross-checked against precalibration constants. A weighted average of the calibration constants derived with different methods is performed a couple of times per year. Dedicated high rate trigger streams ensure that an high statistics for the mentioned physics processes is saved for calibration purposes.

The residual miscalibration of the channel response already ensures a contribution to the energy resolution below 0.5% in the central part of the barrel ($\eta < 1$), and below 2% in the endcaps. The final performance of the ECAL is reported in Fig. 5 where the energy resolution as a function of the pseudo-rapidity is shown.

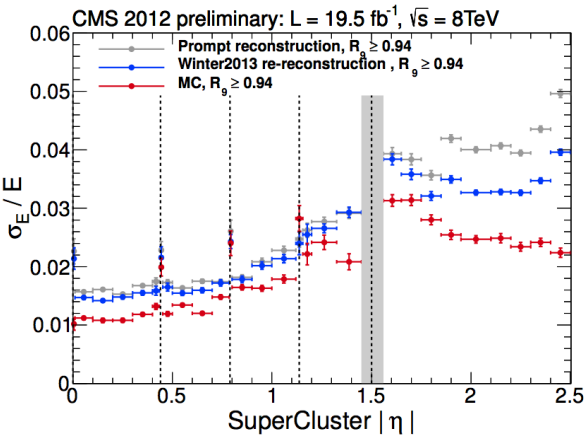

Figure 5: Relative energy resolution as a function of absolute pseudo-rapidity for $Z \to ee$ electrons from data and MC. The discrepancy observed in the simulation with respect to the data, is accommodated in CMS analyses by applying additional Gaussian smearing to the electron and photon energies in MC. The vertical dashed lines are the module boundaries in EB [6].

5. Pixel and Tracker Alignment

The task of the CMS tracker [7, 8] is to measure the trajectories of charged particles (tracks) with very high momentum, angle, and position resolutions, in combination with high reconstruction efficiency. The complete set of parameters describing the geometrical properties of the modules composing the tracker is called the tracker geometry and is one of the most important inputs used in the reconstruction of tracks.

5.1. Pixel Barrel Alignment

Misalignment of the tracker geometry is a potentially limiting factor for its performance therefore it is important to provide to the physics analyses the best possible

geometry for use in the *prompt reconstruction*, immediately correcting any possible time-dependent large misalignment. Specifically, the position of the large structures (beyond the module level) in the pixel detector is relevant for the performance of b-tagging algorithms.
While the alignment at the level of the single modules needs data accumulated over substantial periods of time, the stability of the position of the large structures can be controlled with relatively small amounts of data or via a system of infrared lasers. A system of laser beams is able to monitor the positions of a restricted number of modules in the silicon strip tracker. Movements of large structures in the pixel tracker can be detected with high precision with collision tracks by a statistical study of the primary-vertex residuals, defined as the distance between the tracks and the primary vertex at the point of closest approach of the tracks to the vertex. All these techniques allow the monitoring of the position of the large structures on a daily basis. In Fig. 6 the daily evolution of the relative longitudinal shift between the two half-shells of the pixel barrel as measured with the primary-vertex residuals method.

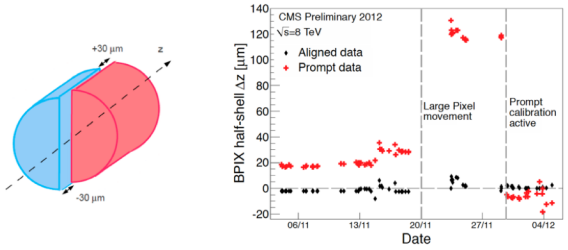

Figure 6: Daily evolution of the relative longitudinal shift between the two half-shells of the pixel detector as measured with the primary-vertex residuals [9].

5.2. Tracker Orientation with respect to the Magnetic Field

The magnetic field of the CMS solenoid is to good approximation parallel to the z-axis. The orientation of the tracker relative to the magnetic field is of special importance, since the correct parameterization of the trajectory in the reconstruction depends on it. This global orientation is described by the angles θ_x and θ_y. Uncorrected overall tilts of the tracker relative to the magnetic field could result in biases of the reconstructed parameters of the tracks and the measured masses of resonances inferred from their charged daughter particles.
The measurement of the tilt angles is based on the study of overall track quality as a function of the θ_x and θ_y angles. The tilt angles θ_x and θ_y are scanned in appropriate intervals centred around zero. For each set of values, the standard CMS track fit is applied to the whole set of tracks, and an overall track quality estimator is determined. The track quality is estimated by the total χ^2 of all the fitted tracks, $\sum\chi^2$.
It is not expected that tilt angles will change significantly with time, hence their measurement are performed offline few times per year. The plot in Fig. 7 shows the result obtained from data; the θ_x values are systematically shifted by 0.3 *mrad*, while the θ_y values are close to zero.

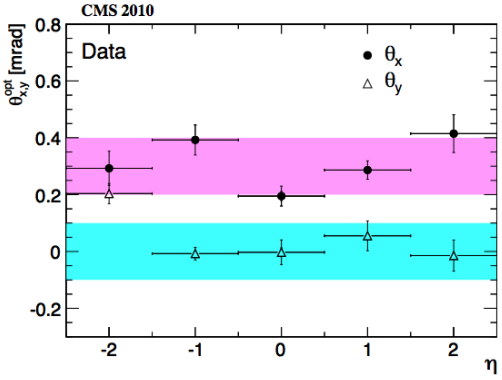

Figure 7: Tracker tilt angles θ_x (filled circles) and θ_y (hollow triangles) as a function of track pseudorapidity [9].

5.3. Tracker Modules Alignment

Track-hit residual distributions are generally broadened if the assumed positions of the silicon modules used in track reconstruction differ from the true positions. Therefore standard alignment algorithms follow the least squares approach and minimise the sum of squares of normalised residuals from many tracks. The ambitious task is to perform a unique fit for 24000 sensors resulting in a number of free parameters of the order of 200000. The target precision is 10 μm and it has been proved that 100 pb^{-1} are sufficient to reach such level of precision.

The muonic decays of Z bosons provide a standard reference that can be used for validating the aligned geometry. The agreement between the data and the simulation is good and suggests that in terms of performance the aligned geometry in data is very close to a perfectly aligned tracker, with a beneficial impact on the physics measurements of CMS (Fig. 8). The modules alignment procedure is performed approximately once per year.

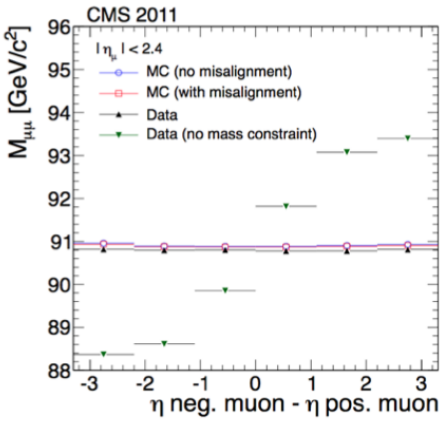

Figure 8: Invariant mass of $Z \rightarrow \mu^+\mu^-$ candidates as a function of the η separation of the two muons. Distributions from aligned data are shown as black upward-pointing triangles. Distributions from a simulation without misalignment and with realistic misalignment are presented as blue hollow circles and red hollow markers, respectively. The same distribution from the data but with a geometry produced without using the Z-boson mass information is presented with green downward-pointing triangles [9].

6. Summary and Outlook

During Run I the alignment and calibration workflows proven to be well designed. In Run II no fundamental changes are foreseen in the data flow. The increased level of automatisation in extracting calibration and alignment conditions will minimise the manual intervention improving the stability and the reliability of the procedures.
In preparation of the Run II startup, the described infrastructures will be exercised during the planned cosmic runs where cosmic data will be collected and used to evaluate the initial alignment conditions after the detector recommissioning. Meanwhile, the ECAL crystals transparency recovery will be monitored with lasers.
The potential impact of the startup calibration and alignment accuracy on the analyses performance is assessed with a dedicated simulation campaign: the so called Computing Software Analysis Challenge. Two main scenarios are simulated:

- startup like scenario: simulating the expected accuracy of the alignment and calibration conditions after the cosmics calibration campaign and a few 100 pb^{-1} of collision data have been acquired. The reference instantaneous luminosity for this sample is $7.5 \times 10^{33} cm^{-2} s^{-1}$ with 50 ns bunch spacing implying an average of 40 pile up collisions per event;
- asymptotic scenario: simulating the expected accuracy of the alignment and calibration conditions after 1 fb^{-1} of collision data has been acquired. The reference instantaneous luminosity for this sample is $7.0 \times 10^{33} cm^{-2} s^{-1}$ (20 pile up) + $1.4 \times 10^{34} cm^{-2} s^{-1}$ (40 pile up) at 25 ns.

The goal is to have the full calibration chain up and running right at the beginning of Run II and to be able to deliver high quality data to be used for publications few hours after the data taking.

References

[1] CMScollaboration,The CMS experiment at the CERN LHC, 2008 JINST 3 S08004.
[2] Federico De Guio, The CMS data quality monitoring software: experience and future prospects, 2014 J. Phys.: Conf. Ser. 513 032024
[3] The CMS Collaboration, Tracking and Primary Vertex Results in First 7 TeV Collisions, CMS-PAS-TRK-10-005.
[4] The CMS Collaboration, The Electromagnetic Calorimeter Project: Technical Design Report, CERN/LHCC 1997, CMS TDR 4.
[5] Anfreville M et al., Laser monitoring system for the CMS lead tungstate crystal calorimeter, Nucl. Instr. and Meth. A 594 (2008) 292-320.
[6] CMS Collaboration, Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s}$ = 7 TeV, 2013 JINST 8 P09009.
[7] CMS Collaboration,The CMS tracker system project: Technical Design Report, CMS TDR CERN-LHCC-98-006 (1998).
[8] CMS Collaboration,The CMS tracker: addendum to the Technical Design Report, CMS TDR CERN-LHCC-2000-016 (2000).
[9] CMS Collaboration, Alignment of the CMS tracker with LHC and cosmic ray data, 2014 JINST 9 P06009