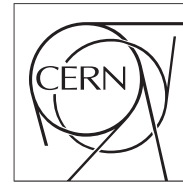


The Compact Muon Solenoid Experiment  
**Conference Report**

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



12 January 2011 (v2, 13 January 2011)

# Validation and Tuning of the CMS Simulation Software

Sunanda Banerjee, Mike Hildreth for the CMS Collaboration

## Abstract

The CMS simulation, based on the Geant4 toolkit, has been operational within the new CMS software framework for more than four years. The description of the detector, including the forward regions, has been completed. A detailed investigation of detector positioning and material budget is underway using collision data. Detailed modelling of detector noise has been performed and validated with the collision data.

Presented at *CHEP2010: International Conference on Computing in High Energy and Nuclear Physics 2010*

# Validation and Tuning of the CMS Full Simulation

S Banerjee<sup>1</sup>, M D Hildreth<sup>2</sup> (on behalf of the CMS Collaboration)

<sup>1</sup>Fermilab, Batavia, Illinois, 60510-5011 USA

<sup>2</sup>University of Notre Dame, Notre Dame, Indiana, 46556 USA

Email: Hildreth.2@nd.edu

**Abstract.** The CMS simulation, based on the Geant4 toolkit, has been operational within the new CMS software framework for more than four years. The description of the detector, including the forward regions, has been completed. A detailed investigation of detector positioning and material budget is underway using collision data. Detailed modelling of detector noise has been performed and validated with the collision data.

## 1. Introduction

The CMS Full Simulation is the primary tool for providing simulated physics events for the CMS experiment [1]. Based on the Geant4 toolkit [2] and interfaced to a diverse collection of Monte Carlo event generators, it has thus far provided the bulk of simulated data on which the first CMS physics results have been based. Here, we present the current status of the simulation validation and tuning, focusing on detailed agreement between simulation and data in terms of material description, electronics simulation, and the overall representation of particle interactions. We focus on various processes that give a qualitative assessment of the simulation performance, concentrating on derived or “physics” quantities that require many different aspects of the simulation to be correct in order to obtain good agreement between simulation and real data.

## 2. Overview of the CMS Full Simulation

A general overview of the CMS Full Simulation is given in Ref [3], which will be summarized here. Four-vectors representing the physics process of choice are passed to the Full Simulation framework in HepMC format. The Geant4 toolkit version 9.3.patch01 is used to propagate the particles through a detailed database-resident detector geometry, as described in Ref [4]. At this stage, one can optionally overlay the simulated Geant hits from additional low- $p_T$  hadronic events in order to simulate pileup interactions. The energy depositions in the sensitive detector volumes are then converted to electronic signals using algorithms based on the observed detector behavior, including the simulation of electronics noise and cross-talk. In many cases, the simulated electronic parameters are identical to those of the real electronics; the constants specifying performance, calibrations, and noise behavior can be read from the same database used for the reconstruction of the real collision data. The output of this stage is simulated data in a format identical to that of the real raw data read from the detector. Further processing uses this data to simulate the formation of the Level1 and High Level Trigger decisions using the same algorithms implemented online in the CMS Trigger system. The simulated raw data that is produced is processed in a manner identical to that followed by the real data from LHC collisions.

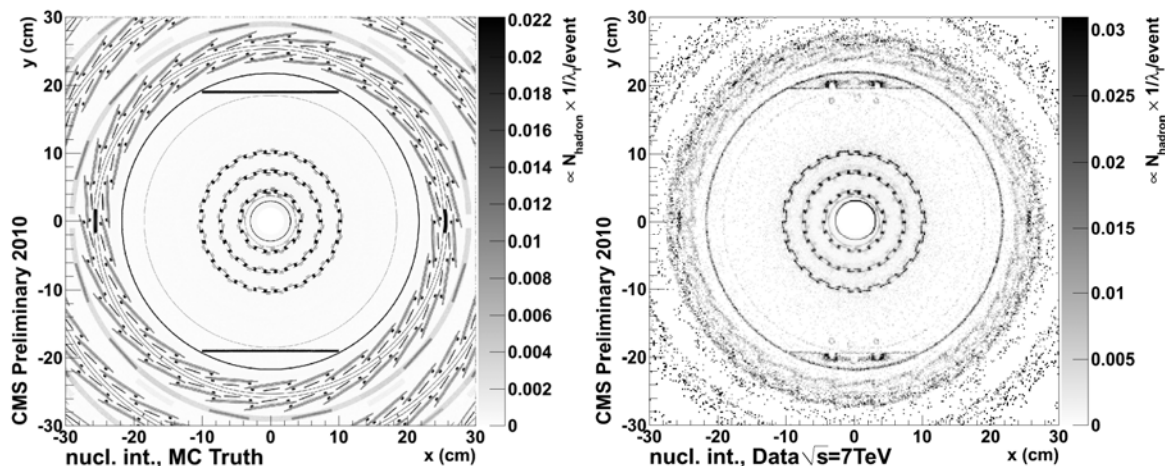
### 3. Comparison of Simulation with Results from Collision Data

In this section we present several comparisons of simulated events with those collected during the beginning of the 2010 physics run of the LHC. Except for some electroweak processes, the early data provides ample statistics for detailed comparisons between simulation and real data. As mentioned above, we will focus here on final-state or “physics” quantities used in physics analysis. For these quantities, the quality of description often depends on several aspects of the simulation. These will be highlighted in the discussion of each comparison.

#### 3.1. Status of the Central Tracker Simulation

Since it is by far the most segmented and complicated sub-detector, the tracker presents tremendous challenges to the simulation. Choices must be made in terms of the level of detail of the simulation geometry in order to optimize computation speed versus the correctness of the simulation. The CMS Central Tracker simulation geometry attempts to strike a balance between performance and detail by using a highly-symmetric and hierarchical detector description. A great deal of attention, however, has been paid to a correct representation of the amount, composition, and location of the material in the detector, since this is critical for a correct simulation of the resolutions of the tracking parameters. This is especially important in the tracking region of the CMS detector because the amount of material is quite large, with 60% of a radiation length at  $|\eta| = 1$ , and approaching 150% of a radiation length at  $|\eta| = 1.5$ . Energy loss and multiple-scattering effects dominate the tracker resolution.

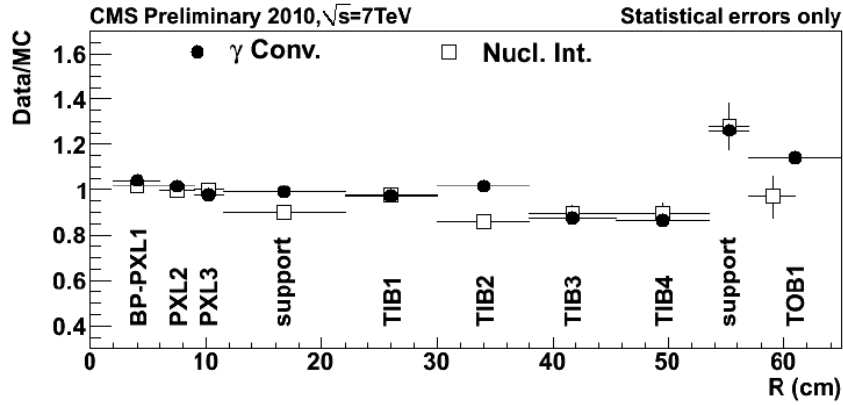
3.1.1. *Material Studies.* Recently, a series of studies [5] with collision data employing the characteristics of converted photons and nuclear interactions, occurring in the tracker, have been used to “map” the location and distribution of the tracker material. In general, the conversion or interaction probability per unit length is inversely proportional to the radiation length or the interaction length, respectively, of the material traversed. Figure 1 shows a comparison between the predicted number of nuclear interactions in the tracker volume and that reconstructed from collision data. The agreement is quite striking, as is the level of detail revealed. The reconstructed photon conversion distributions look similar, although the larger intrinsic uncertainty in the radial position of the conversion point leads to a “fuzzier” picture.



**Figure 1.** A comparison of the predicted location of nuclear interactions in the tracker volume (left) compared with those actually reconstructed in collision data. Note that the variation of the reconstruction efficiency with radius is not taken into account in the simulated distribution.

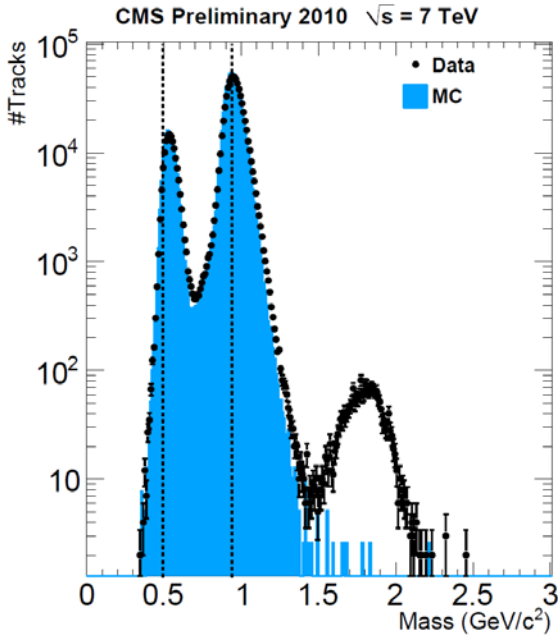
By integrating the conversion and interaction probabilities across different regions of the tracker volume, a quantitative comparison can be made between the material distributions represented in the simulation geometry and those measured in the collision data. This is shown in Figure 2, where

excellent agreement can be seen across most regions of the detector. The discrepancies are under investigation.



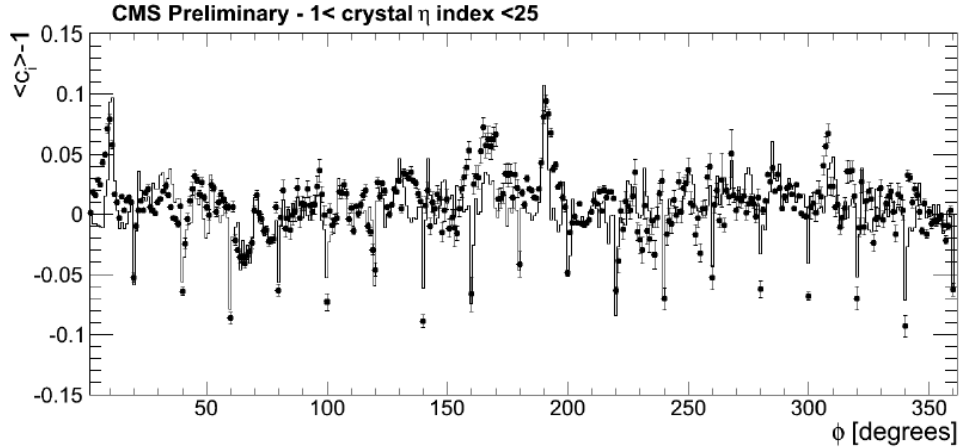
**Figure 2.** The ratio of the predicted rate of photon conversions (solid points) and nuclear interactions (open squares) to those measured in the data for various radial regions of the CMS tracker.

3.1.2. *dE/dx Measurement and Simulation.* Another indication of the quality of the tracker simulation is a comparison of the measured and simulated energy loss. In order for the  $dE/dx$  distributions to agree, the full tracker simulation chain must be correct, including the material thickness of a given tracker detector, the simulation of the charge transport and collection in the bulk silicon, and the amplifier characteristics, including noise and crosstalk. A comparison of the measured and simulated mass distributions (calculated from the momentum and  $dE/dx$  of each track) is shown in Figure 3 [6].



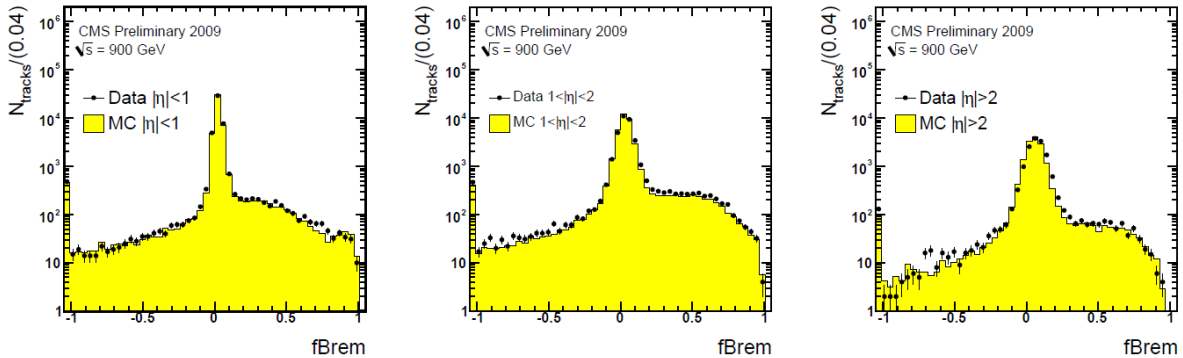
**Figure 3.** The particle mass, calculated from the energy loss and momentum measurements, of charged particles with  $p_T < 2.0$  GeV/c. Note that deuterons (the peak seen at  $\sim 1.8$  GeV/c<sup>2</sup>) are not simulated in the generator used to produce this simulated data.

3.2. Simulation of the Electromagnetic Calorimeter  
 Appropriate modelling of the Electromagnetic Calorimeter (Ecal) necessarily includes a description of the material particles will traverse between the interaction point and the calorimeter, a simulation of the energy deposition, light generation and collection, an emulation of the electronics response, and a proper calorimeter geometry description including the dead material of the calorimeter support structures. Many studies have been performed comparing the simulated and real Ecal. One of the most revealing is an investigation into the corrections that must be applied during the calibration procedure for each calorimeter cell to ameliorate the effects of dead material and interstitial gaps. Figure 4 shows the corrections necessary to achieve a uniform Ecal response as a function of azimuthal angle for one half of the barrel Ecal in rapidity [7]. The corrections are computed separately for the data and a simulation of the calibration procedure. The detailed agreement demonstrates that the ingredients of the Ecal simulation have been properly rendered.



**Figure 4.** Average difference from unity of the inter-calibration constants derived with the  $\phi$  symmetry method for data (solid circles) and simulation (histogram). In the absence of the systematic effects described in the text, a flat distribution with statistical fluctuations around zero is expected.

Additional verification of the geometry, material description, and the physics models in the simulation comes from a study of charged track energy loss as particles traverse the material in front of the Ecal. The quantity  $f_{\text{Brem}} = (p_{\text{in}} - p_{\text{out}})/p_{\text{in}}$  [8] is a measure of the fractional energy loss due to interactions in the material. The momentum and the beginning ( $p_{\text{in}}$ ) and end ( $p_{\text{out}}$ ) of the track is determined using a Gaussian Sum Filter tracking algorithm [9] that can follow the local curvature of a charged track in the track fitting. Figure 5 shows a study [10] of the energy loss in three different regions of the track rapidity, comparing simulation and data. In each plot, the peak corresponding to  $f_{\text{Brem}} = 0$  is predominantly due to pions. The region of the plot with  $f_{\text{Brem}} > 0$  is populated by electrons, while that where  $f_{\text{Brem}} < 0$  is mainly due to resolution effects in the track fit.



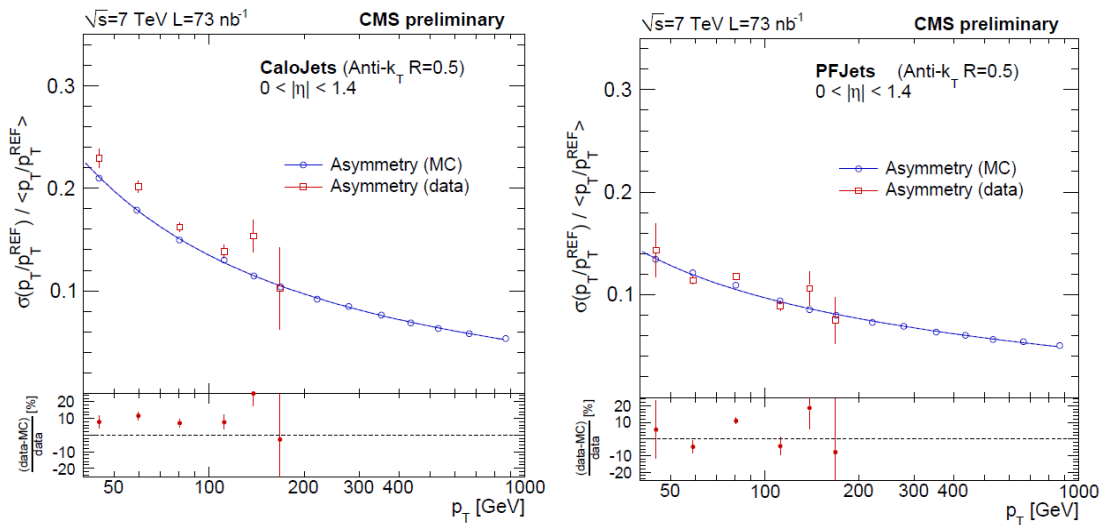
**Figure 5.** A comparison between simulation (solid yellow histogram) and data (black points) of the fractional energy loss for high-quality charged tracks reconstructed with a Gaussian Sum Filter tracking fit. The central peak corresponds to particles (mostly pions) with no bremsstrahlung.

### 3.3. Simulation of the CMS Calorimeter System

One can also study the simulation of the combined Hadronic calorimeter (Hcal) and Ecal. Investigations range from the performance of the Geant4 physics models on single particles [11] to studies of overall agreement of higher-level quantities such as particle jet energy and missing energy resolutions. To obtain good agreement between data and simulation for the more global quantities, many elements of the simulation must be correct, including the physics models, the detector material description, sampling simulation, electronics and electronic noise simulation, and the physics process generators. These represent the most complicated objects that one describes using the simulation.

3.3.1 *Jet Energy Resolution.* There are three main algorithms for jet reconstruction in CMS [12]. Calorimeter jets use only reconstructed energies in the calorimeter for jet formation. In Jet-Plus-Track (JPT) jets, an energy equivalent to the average calorimeter response for a given reconstructed charged particle is subtracted from the Calorimeter jet and replaced with the charged particle momentum. The Particle Flow (PF) technique attempts to reconstruct all final-state single particles (electron, muons, photons, charged and neutral hadrons) from the tracks and calorimeter energy depositions. Jets are then formed from clusters of these reconstructed particles.

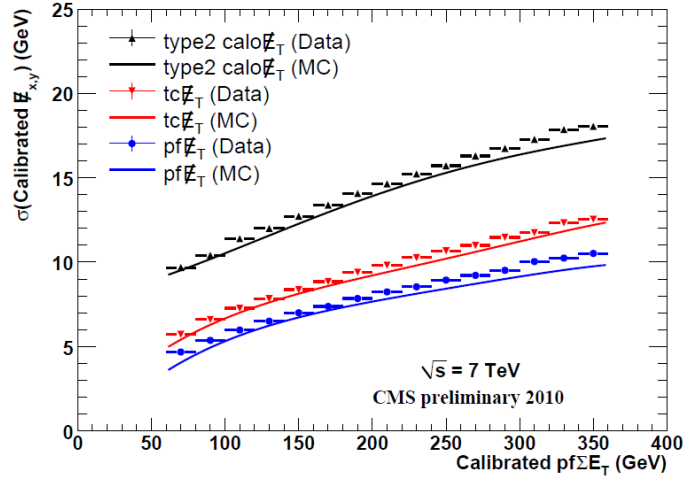
Figure 6 presents a study of the jet energy resolution using a di-jet asymmetry method [12], where the jet energy imbalance in di-jet events is used as a measure of the resolution. The same method is applied to data and simulated di-jet samples, and good agreement is seen. Note that the Particle Flow method has intrinsically better resolution at lower transverse momentum.



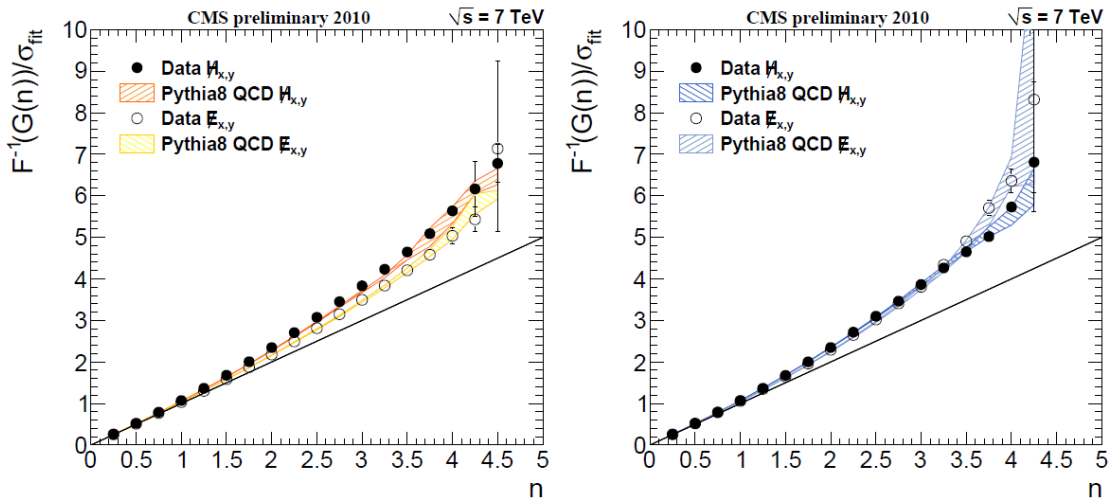
**Figure 6.** Plots comparing the jet energy resolution obtained in di-jet events in data (red points) and simulation (blue points). The left plot shows the quantities as determined from jets formed using purely calorimeter information; the right plot shows the results for particle flow jets.

3.3.2 *Missing Energy and Total Energy Resolution.* In a sense, the resolution of the missing transverse energy is the most stringent test of the overall calorimeter simulation. Physics models, noise simulation, and response simulation all contribute to the total resolution. Figure 7 shows a comparison of the resolution of the missing transverse energy versus the total calorimeter energy measured in di-jet events [13]. Results are shown for data and simulation for Calorimeter jets, JPT jets, and PF jets, again with very good agreement.

An additional study [13] compares the shapes of the missing transverse energy and the total transverse jet energy ( $H_T$ ). Here, the non-Gaussian nature of the tails of the distributions is compared with simulation. Figure 8 shows how many “sigma” of the measured distribution are required in order to contain  $n$  sigma of a Gaussian. A perfectly-Gaussian distribution would thus result in a straight line of slope one. As non-Gaussian tails become larger, more of the central width is required to contain the same fraction of events, hence the deviation upward from the straight line. The agreement between data and simulation for both missing and total energy is quite striking, especially for the tails of such a complicated distribution.



**Figure 7.** A measure of the missing energy resolution vs. the total transverse energy where three different jet algorithms (black: CaloJets, red: JPT jets, blue: PF jets) have been used to reconstruct the jets. Results from the simulation (solid lines) and real data (points) show good agreement.

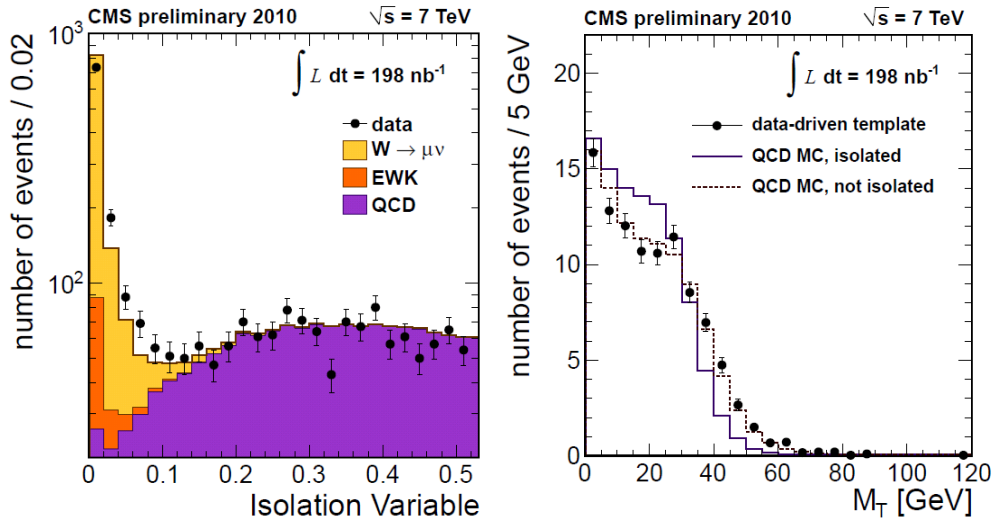


**Figure 8.** A study quantifying the non-Gaussian nature of the tails on missing energy and  $H_T$ . Di-jet events reconstructed with the Calorimeter jet (left) and PF jet (right) algorithms are shown. See text for a description of the axes.

### 3.4. Simulation of the CMS Muon System

The Muon System simulation relies on a combination of physics models (for punch-through), material description, and electronics and noise simulation. Future simulation will also have to contend with backgrounds from neutrons which are expected to fill the CMS cavern during high-luminosity running. Figure 9 shows two examples [14] of distributions related to muon identification in leptonic  $W$  boson decays, where modeling of interactions and backgrounds play important roles. The first plot (left) compares simulation and data for the amount of energy in a  $\Delta R = 0.3$  cone around the muon direction, normalized to the muon momentum. Good agreement is seen in both the signal and background regions. The second plot (right) compares a background template for the transverse mass distribution derived from data events with non-isolated muons with two versions of the mass distribution from QCD Monte Carlo events. These background templates are an essential ingredient for the extraction of the  $W$  production cross section. The simulated distribution derived in the same

manner as the data is in excellent agreement, demonstrating that the simulation of jets, punch-through, and fake muons is close to what is found in the real data.



**Figure 9.** Two distributions demonstrating excellent agreement between simulation and real data for quantities associated with muon identification. See text for a description.

#### 4. Conclusions

We have presented a small fraction of the studies showing excellent agreement between the CMS Full Simulation and the 2010 LHC collision data. These results represent the work of many people over many years. The Full Simulation is ready to serve as the basis for CMS physics program.

#### References

- [1] CMS Collaboration, *The CMS experiment at the CERN LHC*, *J. Instrum.* **3** (2008) S08004
- [2] S. Agostinelli et al., *Geant4 - A Simulation Toolkit*, *Nucl. Instrum. and Methods* **A506** (2003) 250-303
- [3] S. Banerjee, *J.Phys.Conf.Ser.* 119:032006 (2008)
- [4] F. Cossutti, *J.Phys.Conf.Ser.* 219:032005 (2010)
- [5] CMS Collaboration, *CMS Tracking Performance Results from Early LHC Operation*, *Eur. Phys. J. C* **70** (2010) 1165
- [6] CMS Collaboration, *Search for Heavy Stable Charged Particles in pp collisions at √s = 7 TeV*, CMS PAS EXO-10-004
- [7] CMS Collaboration, *Electromagnetic Calorimeter Calibration with 7 TeV Data*, CMS PAS EGM-10-003
- [8] S. Baffioni et al., *Eur. Phys. J. C* **49** (2007), no. 3, 1099
- [9] W. Adam, et al., *J. Phys. G: Nucl. Part. Phys.* **31** (2005) N9–N20
- [10] CMS Collaboration, *Electromagnetic Physics Objects Commissioning with First LHC Data*, CMS PAS EGM-10-001
- [11] See S. Banerjee, *Validation of Geant4 Physics Models with LHC Collision Data*, these proceedings
- [12] See the CMS Collaboration, *Jet Performance in pp Collisions at √s = 7 TeV*, CMS PAS JME-10-003, and references therein
- [13] CMS Collaboration, *Missing Transverse Energy Performance in Minimum-Bias and Jet events from proton-proton Collisions at √s = 7 TeV*, CMS PAS JME-10-004
- [14] CMS Collaboration, *Measurements of Inclusive W and Z Cross Sections in pp Collisions at √s = 7 TeV*, CMS PAS EWK-10-002