IL NUOVO CIMENTO Vol. 34 C, N. 2 Marzo-Aprile 2011
DOI 10.1393/ncc/i2011-10829-6

Colloquia: WISH2010

Underlying events and jet reconstruction in CMS

X. Janssen for the CMS Collaboration
Universiteit Antwerpen - Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

(ricevuto il 10 Gennaio 2011; approvato il 24 Gennaio 2011; pubblicato online il 16 Marzo 2011)

Summary. — In the following, we report on the techniques adopted by the CMS experiment to measure jets and on studies of the underlying event in pp interactions at $\sqrt{s} = 0.9$ and 7 TeV. Jets are reconstructed in the CMS detector on basis of the energy measured in the calorimeters ("Calorimeter jet") or tracks from charged particles ("Track-Jet"), or by combining both information ("Jet-Plus-Track" and "Particle Flow" algorithms). Results from detailed Monte Carlo studies are compared with the first pp data collected at the LHC and the jet energy response and resolutions are measured leading to a 10% (5%) jet energy scale uncertainty for Calorimeter jets (Jet-Plus-Track and Particle Flow jets), with an additional 2% uncertainty per unit of rapidity. The underlying event is studied by measuring the charged particle multiplicity and the energy density in the regions perpendicular to the plane of the hard 2-to-2 scattering which includes the beam and the jet directions. The underlying event activity increase with the leading jet transverse momentum ($p_{T,\mathrm{jet}}$) up to about 5–10 GeV/c and a plateau is observed at higher $p_{T,\mathrm{jet}}$ values. A factor two increase of the underlying event activity is observed at $\sqrt{s} = 7$ TeV with respect to 0.9 TeV. These studies allow to discriminate among several QCD Monte Carlo parametrisations which reproduce the Tevatron underlying event observations but diverge at higher energy as well as improve these models.

PACS 12.38.-t – Quantum chromodynamics.
PACS 12.38.Aw – General properties of QCD (dynamics, confinement, etc.).
PACS 12.38.Qk – Experimental tests.

1. – The LHC and the CMS experiment

In the fall of 2009, the Large Hadron Collider (LHC) started with proton-proton collisions at centre-of-mass energies ($\sqrt{s}$) equal to 900 GeV and 2.36 TeV. In March 2010 a running period at $\sqrt{s} = 7$ TeV started. The data presented in this report were collected by the Compact Muon Solenoid (CMS) experiment, one of the two general purpose detector at the LHC, either at 900 GeV or 7 TeV centre-of-mass energies.

The central feature of the CMS apparatus is a superconducting solenoid, 13 m in length and 6 m in diameter, which provides an axial magnetic field of 3.8 T. The bore of the solenoid is instrumented with various particle detection systems. The steel return

© Società Italiana di Fisica

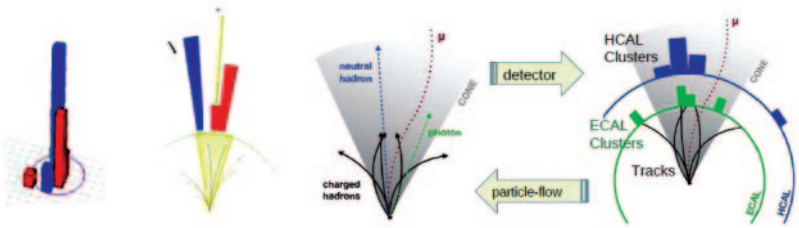

Fig. 1. – Schema of Calorimeter (left), Jet-Plus-Track (center), and Particle-Flow Jet reconstruction (right).

yoke outside the solenoid is in turn instrumented with gas detectors used to identify muons. Charged particle trajectories are measured by the silicon pixel and strip tracker, with full azimuthal coverage within $|\eta| < 2.5$, where the pseudorapidity η is defined as $\eta = -\ln\tan(\theta/2)$, with θ being the polar angle of the trajectory of the particle with respect to the counterclockwise beam direction. A lead-tungstate crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL) surround the tracking volume and cover the region $|\eta| < 3$. In addition, a steel/quartz-fibre forward calorimeter (HF) covers the region $2.9 < |\eta| < 5.2$. A more detailed description of the CMS detector can be found elsewhere [1].

Data analysed were recorded either on basis of the inclusive jet trigger requesting a minimal total transverse momentum being observed in the calorimeters or through the so-called MinBias trigger. The latter is based on two of the CMS subdetectors acting as LHC beam monitors, the Beam Scintillation Counters (BSC) and the Beam Pick-up Timing for the eXperiments (BPTX) devices. The BSCs are located along the beam line on each side of the interaction point (IP) at a distance of 10.86 m. The two BPTX devices, which are located inside the beam pipe at distances of 175 m from the IP, are designed to provide precise information on the bunch structure and timing of the incoming beams.

2. – Jet reconstruction methods

Four types of jets are reconstructed at CMS, which combine differently individual contributions from subdetectors to form the inputs to the jet clustering algorithm: calorimeter jets, Jet-Plus-Track (JPT) jets, Particle-Flow (PFlow or PF) jets, and track jets. Jets in the studies presented here are reconstructed using the Anti-k_T [2] clustering algorithm with the size parameter $R = 0.5$.

Calorimeter jets are reconstructed using energy deposits in the electromagnetic (ECAL) and hadronic (HCAL) calorimeter cells, combined into calorimeter towers, see fig. 1(left). In order to suppress the contribution from calorimeter readout electronics noise, thresholds are applied on energies of individual cells when building towers for jets reconstruction [3]. In addition, to suppress the contribution from event pile-up (additional proton-proton interactions in the same bunch crossing), calorimeter towers with transverse energy of $E_T^{towers} < 0.3\,\text{GeV}$ are not used in jet reconstruction.

In the *Jet-Plus-Tracks* algorithm (fig. 1(center)) [4], calorimeter jets are reconstructed first as described above, then charged particle tracks are associated with each jet based on spatial separation in $\eta - \phi$ between the jet axis and the track momentum measured at the interaction vertex. The associated tracks are projected onto the surface of the calorimeter and classified as in-cone tracks if they point to within the jet cone around

the jet axis on the calorimeter surface, or out-of-cone otherwise. The momenta of both in-cone and out-cone tracks are then added to the energy of the associated calorimeter jet. For in-cone tracks the expected average energy deposition in the calorimeters is subtracted based on the momentum of the track. The direction of the axis of the original calorimeter jet is also corrected by the algorithm.

The *Particle Flow* algorithm combines the information from all CMS sub-detectors to identify and reconstruct all particles in the event (see fig. 1(right)), namely muons, electrons, photons, charged hadrons and neutral hadrons [5, 6]. PFlow jets are then reconstructed from the resulting list of particles. The jet momentum and spatial resolutions are expected to be improved with respect to calorimeter jets as the use of the tracking detectors and of the excellent granularity of the ECAL allows to resolve and precisely measure charged hadrons and photons inside jets, which constitute $\sim 90\%$ of the jet energy.

Track jets [7] are reconstructed from tracks of charged particles measured in the central tracker. Only well-measured tracks, based on their association with the primary vertex and their quality, are used by the algorithm. The method is completely independent of the calorimetric measurements, allowing for cross-checks.

3. – Jet energy scale calibration

Data used for the studies presented here were recorded by the CMS detector during March–July 2010 in proton-proton collisions at LHC at $\sqrt{s} = 7\,\text{TeV}$ and correspond to the integrated luminosity up to $73\,\text{nb}^{-1}$ [3]. Data filtering both on-line (trigger) and off-line was applied to reject beam background events. Further jet quality criteria ("Jet ID") were applied to retain the vast majority of real jets while rejecting most fake jets arising from calorimeter and/or readout electronics noise. The current JEC status in CMS, which was obtained after this conference, is described in [8] but is not discussed in this report.

The jet energy measured in the detector is typically different from the corresponding particle jet energy. The main cause for this energy mismatch is the non-uniform and non-linear response of the CMS calorimeters. Furthermore, electronics noise and additional pp interactions in the same bunch crossing (event pile-up) can lead to extra unwanted energy. The purpose of the jet energy correction is to relate, on average, the energy measured in the detector to the energy of the corresponding particle jet. CMS utilize a factorized multi-step procedure, offset, relative and absolute corrections, for the jet energy calibration (JEC):

$$E_{Corrected} = (E_{Uncorrected} - E_{Offset}) \times C_{Rel}(\eta, p''_T) \times C_{Abs}(p'_T)$$

where p''_T is the transverse momentum of the jet corrected for offset and $p'_T = p''_T \times C_{Rel}(\eta, p''_T)$ is the transverse momentum of the jet corrected for offset and pseudorapidity dependence. The offset correction aims to correct the jet energy for the excess unwanted energy due to electronics noise and pile-up. The relative correction removes variations in jet response versus jet η relative to a central control region. The absolute correction removes variations in jet response *versus* jet p_T.

3˙1. *Monte Carlo truth based JEC.* – In this approach, the JEC is estimated from the Monte Carlo jet response which is measured from QCD events generated with PYTHIA6 [9] and processed through the full, GEANT4 [10] based, CMS detector simulation. The same jet clustering algorithm is applied to all stable generated particles and

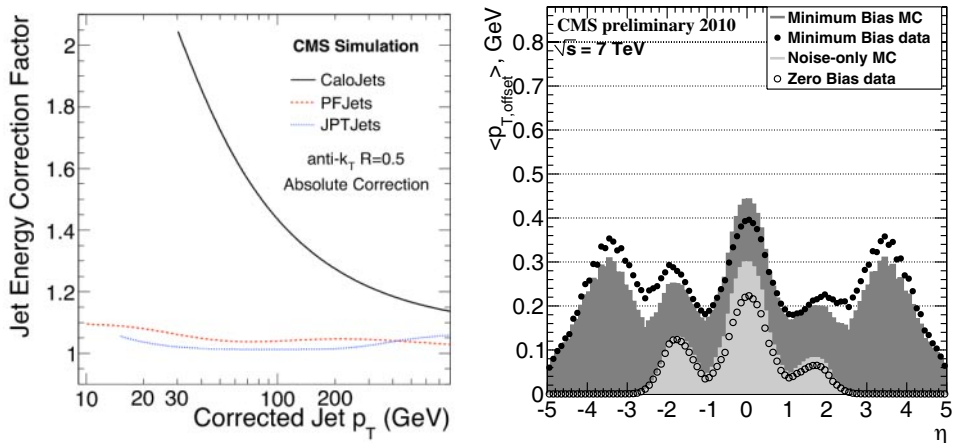

Fig. 2. – Left: absolute jet energy correction factors C_{Abs} derived from simulation for calorimeter, JPT, and PFlow jets at $\sqrt{s} = 7\,\text{TeV}$ as a function of corrected jet transverse momentum. Right: average offset p_T as a function of η from noise-only and from noise+one pile-up.

these generator jets are then associated to reconstructed ones, by requiring a small distance between the jet axes. For the matched jets, we study the quantity p_T^{Jet}/p_T^{GenJet} to extract jet calibration factors as a function of uncalibrated jet p_T and η. The truth-based JEC Monte Carlo do not factorize out the offset correction which is lumped together with the relative and absolute corrections. Figure 2(left) shows the absolute correction factors C_{Abs} as a function of corrected jet transverse momentum for the three jet types. At low p_T, calorimeter jets need to be corrected by a large multiplicative factor (up to 2) due to the non-compensating nature of the CMS calorimeters. JPT and PFlow jets require much smaller corrections as these jets rely heavily on the tracking information. Details on the η dependence of the Monte Carlo truth-based JEC can be found in [3]. By re-deriving the JEC for corrected jets, a 2% uncertainty on Monte Carlo truth-based JEC for jets in simulation is derived.

3˙2. *Data driven JEC estimation.* – At the current initial stage of LHC running, Monte Carlo truth-based JEC is used to correct jets in both data and Monte Carlo simulation, since the limited statistics do not allow yet for a complete data-driven estimation of the JEC. The uncertainty for jet in data are obtained from real collision events, as discussed below, based on the factorisation approach.

Offset correction: The contributions from calorimeter electronic noise and extra pp interactions within the same bunch crossing (pile-up) to the offset corrections are estimated separately. To estimate the noise-only contribution, data from random trigger together with a veto on real collision events leading to activity within the CMS detector are used. In this sample, the average calorimeter p_T summed up inside a cone of radius $R = 0.5$ at a given η ($p_{T,Offset}(\eta)$) is measured. To estimate the offset from one additional interaction event, we select Minimum Bias trigger events in early runs, (where the fraction of events with more than one interaction per bunch crossing is small) and study the same quantity. Both results are shown in fig. 2(right) for data and Monte Carlo. The maximum offset in p_T being $400\,\text{MeV}/c$, this contribution to the JEC is neglected.

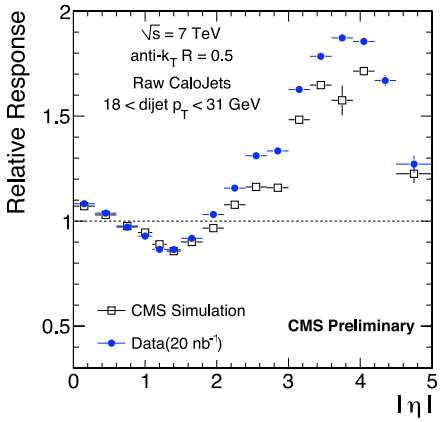

Fig. 3. – Relative jet energy response for Calorimeter jets as a function of $|\eta|$ for $18 < p_T^{dijet} < 31\,\mathrm{GeV}/c$.

Relative response from dijet p_T balance: This correction is estimated using the p_T balance in back-to-back dijet events with one jet in the central control region of the calorimeter, $|\eta| < 1.3$ (barrel jet), and the other jet at arbitrary η (probe jet). Events collected on the basis of the jet trigger are requested offline to contain at least one good primary vertex and two jets, with one of them in the barrel region of $|\eta| < 1.3$. The two leading jets must be azimuthally separated by $\Delta\phi > 2.7$ rad and no additional jets with $p_T^{3rdJet}/p_T^{dijet} > 0.2$ are allowed, where $p_T^{dijet} = (p_T^{probe} + p_T^{barrel})/2$. The relative response is estimated as $R(\eta_{probe}, p_T^{dijet}) = (2+\langle B\rangle)/(2-\langle B\rangle)$, where $B = (p_T^{probe} - p_T^{barrel})/p_T^{dijet}$ is the dijet balance, and is shown for one bin in p_T^{dijet} in fig. 3 for calorimeter jets as a function of $|\eta|$. An additional correction factor to take into account in the remaining discrepancy at large $|\eta|$ between data and Monte Carlo is hence applied to data in top of the Monte Carlo truth-based JEC and a systematic of $2\%|\eta|$ is associated to it. Similar results are obtained for other jet types [3].

Absolute response measurements from photon+jet events: This method uses the p_T balance in events with a jet back-to-back to a photon. Photon candidates with $p_T^{\gamma} > 15\,\mathrm{GeV}/c$ within $|\eta| < 1.3$ are required to be well isolated from any other activity. In the selected photon sample, we require the presence of a barrel jet ($|\eta| < 1.3$) recoiling against the photon candidate in azimuth by $\Delta\phi > 2\pi/3$. Events containing a second jet with $p_T^{2ndjet} > 0.5 p_T^{\gamma}$ are discarded. The absolute response is measured by studying the average response $< p_T^{jet}/p_T^{\gamma}$ as a function of p_T^{γ} and is of the order of 50% (70%) for calorimeter (JPT and PF) jets [3]. A global 10% (5%) systematic on the Monte Carlo truth-based JEC is estimated from the discrepancy between data and Monte Carlo for calorimeter jets (JPT and PF jets). This is however dominated by low statistics in the current event sample and a conservative 10% uncertainty is presently used.

4. – Underlying event studies in the transverse region to the leading jet

The underlying event in a hard scattering process is everything accompanying an event but the hard scattering component of the collision. Results at 900 GeV correspond

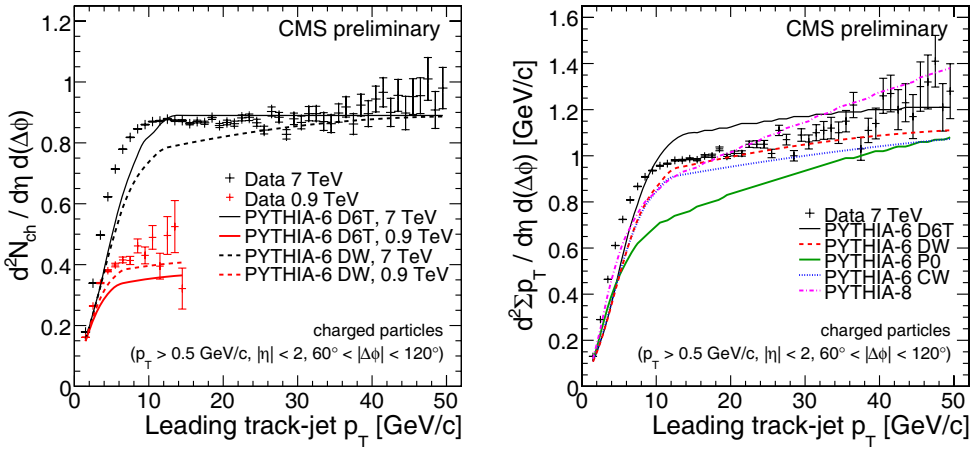

Fig. 4. – The average charge multiplicity (left) and scalar sum p_T (right) for charged particles with $p_T > 0.5\,\mathrm{GeV}/c$ and $|\eta| < 2$ in the transverse region to the leading jet as a function of the leading jet p_T, for data at $\sqrt{s} = 0.9$ and 7 TeV, compared with predictions of several PYTHIA6 Tunes and of PYTHIA8.

to an integrated luminosity of $1\,\mathrm{mb}^{-1}$ [11] while those at 7 TeV amounts to $1\,\mathrm{nb}^{-1}$ [12] and the data are collected on basis of the MinBias trigger.

The analysis uses the "jet" structure of the hard proton-proton collision to experimentally study the underlying event activity. Track-jets are reconstructed from tracks using a cone algorithm. The direction of the hard scattering is assumed correlated with the leading track-jet direction and is used to isolate regions of $\eta - \phi$ space that are maximally sensitive to the underlying event. Tracks are used to account the activity in the transverse region to the track-jet direction, defined by $\pi/3 < |\phi_{track} - \phi_{jet}| < 2\pi/3$. The p_T of the leading track-jet defines the energy scale of the event. The underlying event activity is studied looking at the charged track multiplicity and scalar sum of the p_T as a function of leading track-jet in the transverse region. The data are compared with various tunes of PYTHIA6 (D6T [13], DW [14], P0 [15] and CW which is a modified version of DW) and with PYTHIA8 [16].

4˙1. *Event and track selection.* – The tracks with $p_T > 0.5\,\mathrm{GeV}/c$ and $|\eta| < 2$ are kept for the underlying event studies if they originate from the primary vertex. In order to suppress contamination by secondary tracks from decays of long-lived particles and photon conversion, a small distance of closest approach between track and primary vertex is required. The leading track-jet is reconstructed with the tracks selected using the SISCone algorithm [17]. The data are not corrected for detector effects but are rather compared to the reconstructed level Monte Carlo predictions. An average systematic relative error of 2% has been computed by taking into account uncertainties such as tracks selection criteria, tracker alignment, tracker material content, background contamination, trigger condition and variation of the beam spot condition.

4˙2. *Results.* – Figure 4 shows the average charge multiplicity and scalar sum p_T in the transverse region to the leading jet as a function of its p_T. Both quantities show the same features in data and Monte Carlo: a fast rise due to multiple parton interactions

(MPI) in the event, the main component of the underlying event activity, at low p_T followed by a slower increase due to radiation above $5\,\mathrm{GeV}/c$ and $8\,\mathrm{GeV}/c$ for 0.9 and 7 TeV respectively. The rise of the transverse charge multiplicity is underestimated by the Monte Carlo predictions, while the observed plateau is best reproduced by the PYTHIA6 D6T tune. For the scalar sum p_t in the rising region, however, the Monte Carlo predict too little activity, PYTHIA8 being more successful than the other tunes at the lowest p_T values. The flattening of the distributions is described by CW and DW, whereas the increase of activity with increasing leading track-jet p_T observed for D6T and for PYTHIA8 is significantly too large, with P0 predictions being systematically below the data. Within CMS, a new PYTHIA6 tune [18] which better describes the main feature of the underlying event has been recently developed on basis of these results.

5. – The jet-median approach to the underlying event at $\sqrt{s} = 900\,\mathrm{GeV}$

Recently, a new method to study the underlying event has been proposed to take into account all jets in the event and evaluates, on an event-by-event basis, the median of the ratio of the transverse momenta and the areas of all jets in the event [19]. The areas of the jets are determined with the active area clustering method, provided by the FastJet package [20].

The median is expected to be robust against outliers such as hard di-jets and is therefore a suitable measure for underlying event activity. However, due to the low activity in minimum bias events, the observable has been adjusted to take into only physical jets in the event, while pure ghost jets are ignored. The adjusted observable is:

$$\rho' = \underset{j \in \text{physical jets}}{\text{median}} \left[\left\{ \frac{p_{Tj}}{A_j} \right\} \right] \cdot C$$

where p_{Tj} is transverse momentum of jet j, $A_j = N_j^{ghosts}/\rho^{ghosts}$ is the active area defined by overlaying in an event a uniform grid of extremely soft pseudo-particles (ghosts) with a density ρ^{ghosts} and counting the number of ghosts clustered into a jet j (N_j^{ghosts}) and C is the occupancy of the event, which is the summed area $\sum_j A_j$ covered by all physical jets divided by the considered detector region A_{tot}.

For the present analysis [21], only the early 2009 LHC data at $\sqrt{s} = 900\,\mathrm{GeV}$ have been analysed. In order to stay consistent with the traditional underlying event measurement above performed on the same dataset, the event and track selection were closely matched but with a track transverse momentum of $0.3\,\mathrm{GeV}/c$. The data were compared with the same Monte Carlo as in the previous section. The observed distribution of the jet median ρ' is compared with these predictions in fig. 5 and the ratio between them is extracted. Similarly to the traditional underlying event measurement, none of the PYTHIA6 tunes nor PYTHIA8 are able to describe all the data features which demonstrates the sensitivity of the modified jet median variable ρ'. Among all tunes, DW is however the closest to the data.

6. – Conclusions

A first study of jet energy calibration from early pp collisions recorded by CMS at $\sqrt{s} = 7\,\mathrm{TeV}$ has been presented. Four different techniques to reconstruct jets in CMS are used: calorimeter jets, Jet-Plus-Track jets, Particle-Flow and track jets for which the Anti-k_T clustering algorithm with cone size of $R = 0.5$ is used. Following the approach

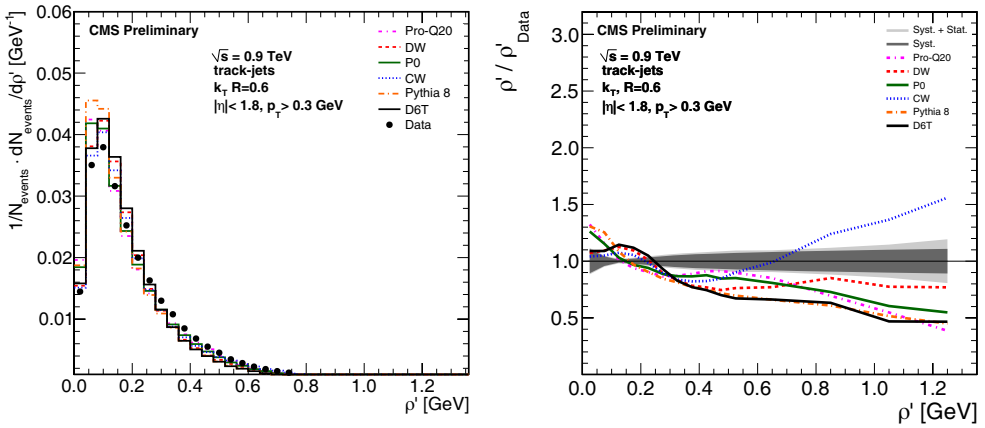

Fig. 5. – Left: median of p_T over jet area for track-jets reconstructed from collision data (black circles) and for the different PYTHIA6 tunes and PYTHIA8 default tune. Right: ratio of the median of p_T over jet area for the same Monte Carlo predictions with respect to data. The dark-gray shaded band corresponds to the systematic uncertainty and the light-gray shaded band to the total uncertainty.

of multi-step factorized jet energy calibration adopted by CMS, offset, relative and absolute jet energy corrections have been studied separately from various data samples. Significantly better performance for the jet types employing the tracking information have been observed compared to the jets using calorimeter-only information. Current physics analyses in CMS use 10% (5%) JEC uncertainties for calorimeter jets (JPT and PFlowjets), with the additional 2% uncertainty per unit rapidity. Observations from the current limited statistics datasets support these numbers as conservative estimates. The systematic uncertainty of the jet energy resolution is estimated to be 10% for all jet types. Observations from the data support this number as a reasonable estimate.

A study of the underlying event activity in proton-proton interactions at both 900 GeV and 7 TeV has been presented. A significant activity increase with the jet transverse momentum is reported, confirming the so-called "Pedestal Effect". A factor two increase of the underlying event activity is observed at 7 TeV with respect to 900 GeV. The comparison between data collected and the predictions from different PYTHIA6 tunes and PYTHIA8 show the best agreement with the DW tune and confirm the expected sensitivity to Monte Carlo models. The CMS collaboration is hence presently working on obtaining a new tune which describes these data.

REFERENCES

[1] CMS Collaboration, *JINST*, **3** (2008) S08004.
[2] Cacciari M., Salam G. P. and Soyez G., *JHEP*, **0804** (2008) 063.
[3] CMS Collaboration, CMS-PAS-JME-10-003 (2010); http://cdsweb.cern.ch/record/1279362.
[4] CMS Collaboration, CMS-PAS-JME-09-002 (2009); http://cdsweb.cern.ch/record/1190234.
[5] CMS Collaboration, CMS-PAS-PFT-09-001 (2009); http://cdsweb.cern.ch/record/1194487.

[6] CMS Collaboration, CMS-PAS-PFT-10-002 (2010); http://cdsweb.cern.ch/record/1279341.
[7] CMS Collaboration, CMS-PAS-JME-10-006 (2010); http://cdsweb.cern.ch/record/1275133.
[8] CMS Collaboration, CMS-PAS-JME-10-010 (2010); http://cdsweb.cern.ch/record/1308178.
[9] Sjostrand T., Mrenna S. and Skands P. Z., *JHEP*, **05** (2006) 026.
[10] GEANT4 Collaboration, *Nucl. Instrum. Methods A*, **506** (2003) 250.
[11] CMS Collaboration, *Eur. Phys. J. C*, **70** (2010) 555.
[12] CMS Collaboration, CMS-PAS-QCD-10-010 (2010); http://cdsweb.cern.ch/record/1279345.
[13] Field R., *Acta Phys. Polon. B*, **39** (2008) 2611.
[14] Field R., arXiv:1003.4220.
[15] Skands P. Z., arXiv:0905.3418.
[16] Sjostrand T., Mrenna S. and Skands P. Z., *Phys. Commun.*, **178** (2008) 852.
[17] Cacciari M. and Salam G. P., *Phys. Lett. B*, **641** (2006) 57.
[18] Field R., arXiv:1010.3558.
[19] Cacciari M. *et al.*, *JHEP*, **1004** (2010) 065.
[20] Cacciari M. and Salam G. P., *Phys. Lett. B*, **641** (2006) 57.
[21] CMS Collaboration, CMS-PAS-QCD-10-005; http://cdsweb.cern.ch/record/1278706.