IL NUOVO CIMENTO Vol. 33 C, N. 4 Luglio-Agosto 2010
DOI 10.1393/ncc/i2010-10654-5

Colloquia: TOP2010

Leptons, b-tagging and MET reconstruction at CMS after the first data

J. Maes on behalf of the CMS Collaboration

Interuniversity Institute for High Energies, Vrije Universiteit Brussel - Brussels, Belgium

(ricevuto l' 8 Agosto 2010; approvato l' 8 Agosto 2010; pubblicato online il 28 Settembre 2010)

Summary. — The performance of reconstruction and analysis tools is an essential ingredient for top quark physics analyses. This presentation gives an overview of the current tools in the CMS experiment, their expected performance obtained with simulations and the performance studies using the first data. The presented data consists of several data-taking campaigns. The Cosmic Run at Four Tesla (CRAFT08/09) provided millions of cosmic muons and was primarily used to study the tracker and the muon systems. In the first weeks after the start of the LHC experiment at the end of 2009, proton collision data at a center-of-mass energy of 900 GeV and 2360 GeV were collected. During the first months of 2010 data at a center-of-mass energy of 7 TeV was collected. During the redaction of this paper data taking at 7 TeV is ongoing.

PACS 14.65.Ha – Top quarks.

1. – Leptons

1·1. *Muons*. – The muon reconstruction in CMS is performed in three stages with the *Combinatorial Track Finder* algorithm [1]. In the first stage, the local reconstruction, information of the muon subsystems is combined into track segments which serve as regional seeds for further trajectory building. These regional seeds are combined in the standalone reconstruction step to build the muon trajectory in the muon system only. In the third and final step, the global muon reconstruction, the trajectories of the local muons are extrapolated towards the interaction point to search for compatible tracks in the tracking system. After a compatible track is identified a last refit of the trajectory is performed to determine the muon four-momentum and its trajectory.

Five track parameters, resulting from the track reconstruction, describe the helical trajectory of a track at the point closest to the nominal interaction point. The performance of the muon track reconstruction strongly depends on the alignment of the tracker, the alignment of individual muon chambers and the alignment of global muon system with respect to the tracker.

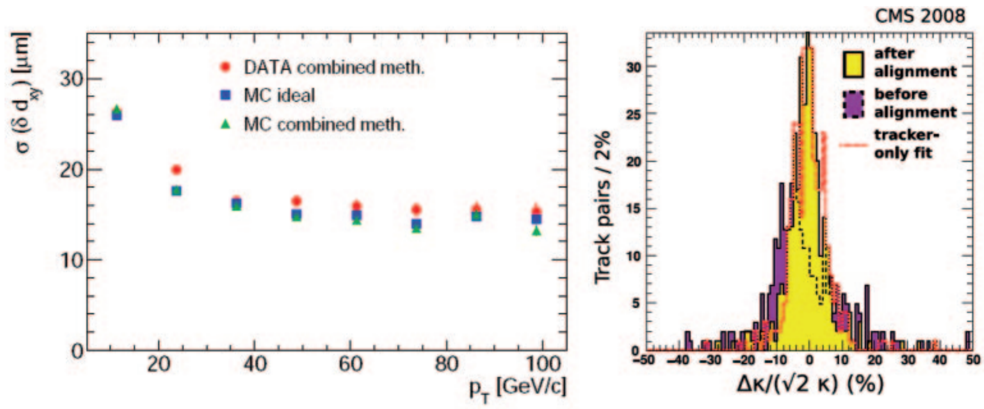

Fig. 1. – Left: difference between upper and lower track segment transverse impact parameters, *vs.* the p_T of the track. Right: fractional curvature difference between top and bottom parts of CMS muons before and after alignment.

1˙1.1. Alignment of the tracker using cosmic rays. The alignment procedure for the tracker [2] is a complex optimization problem to determine corrections for the module positions or alternatively the alignment parameters. Two statistical methods were employed at CMS to solve the alignment problem, the global alignment algorithm, *Millepede II* and the local iterative algorithm, the *Hits and Impact Points* algorithm. Both alignment algorithms were used to obtain module position corrections independently using more than three million cosmic ray charged particles collected during the CRAFT data-taking period. After verifying the consistency between the two methods, the final results were obtained by applying the two algorithms in sequence to take advantage of their complementary strengths.

Track parameter resolutions were validated with an independent reconstruction of upper and lower *legs* of cosmic ray tracks. The performance of the track parameters is found to be already very close to the ideal performance from Monte Carlo simulated samples with ideal detector geometry, see fig. 1 (left). The positions of the modules were determined to an average precision of 34 microns RMS in the barrel and 314 microns RMS in the endcaps in the most sensitive coordinate.

1˙1.2. Alignment of the muon system using cosmic rays. The alignment of the layers within the drift tube chambers and the alignment of the cathode strip chambers relative to one another was performed using locally fitted track segments. Additional to this internal alignment each muon chamber was aligned relative to the tracker [3] using the tracks from the tracker propagated to the muon system with a detailed map of the magnetic field and material distribution of CMS. The *reference-target* algorithm divides the tracking volume into two regions: a reference region (the tracker), in which normal track-fitting is performed, and a target region (the muon chambers), in which unbiased residuals are computed from the propagated tracks. Residuals are the differences between the predicted particle trajectories and the muon chamber data.

The alignment procedure is validated by studying the improvement in the resolutions of the muons after re-fitting with the new geometry. To study the sensitivity of the effect of misalignment of the muon system, energetic cosmic rays are selected with $p_T > 200\,\mathrm{GeV}/c$, a sample which is independent of the $100\,\mathrm{GeV}/c < p_T < 200\,\mathrm{GeV}/c$

tracks used to perform the alignment. The top half and bottom half of the cosmic ray trajectory are reconstructed separately, split at the point of closest approach to the LHC beamline. Any difference in track parameters between the top and bottom fits is purely instrumental. A significant improvement of the resolution of the muon is observed, see fig. 1 (right).

1˙1.3. The performance of the muon reconstruction. The performance of muon reconstruction in CMS is evaluated using cosmic-ray muons during the CRAFT08 data-taking campaign [4]. The efficiency of the muon reconstruction algorithm was measured by selecting events with a good-quality global muon reconstructed in one hemisphere of the detector and examining the presence of a corresponding track in the opposite hemisphere. The efficiency to reconstruct the global muon is found to be $(99.7 \pm 0.1)\%$. For a wide range of transverse momentum of the reconstructed muon essentially no p_T and η dependence has been observed. The relative momentum resolution for muons crossing the barrel part of the detector is found to be better than 1% at $10\,\mathrm{GeV}/c$ and is about 8% at $500\,\mathrm{GeV}/c$, the latter being only a factor of two worse than expected with ideal alignment conditions. Muon charge misassignment is found to be ranging from less than 0.01% at $10\,\mathrm{GeV}/c$ to about 1% at $500\,\mathrm{GeV}/c$.

1˙2. *Electrons.* – Electron candidates are reconstructed by associating a supercluster in the electromagnetic calorimeter (ECAL) to a track reconstructed in the tracker. Two complementary approaches to reconstruct electrons in the CMS detector have been developed [5]. In the first approach, the track search is seeded by the supercluster: the initial track segments formed with two hits in the pixel detector layers and/or inner silicon strip detector layers are considered only if they match the supercluster in η and ϕ and in p_T, accounting for the track curvature expected from the supercluster E_T and the four-Tesla magnetic field. Both charge hypotheses are considered, and the electron is assumed to originate from the measured beam spot. This approach is well suited for high-p_T electrons ($p_T > 5$–$10\,\mathrm{GeV}/c$), where the supercluster energy and position estimations are reliable and kinks due to bremsstrahlung do not impair track reconstruction much. In the second approach, short track segments with as little as 3 hits reconstructed with the standard tracking algorithm are loosely matched to ECAL deposits in η, ϕ and E_T. To reduce the number of charged pions faking electrons, a pre-identification is performed using a boosted decision tree. The input variables are the ECAL matching variables, preshower information in the endcaps, the number of hits along the trajectory, the track fit χ^2 and an estimation of the bremsstrahlung energy loss from the difference in the track momentum estimated at the vertex and at the ECAL surface. This approach is expected to be well suited for electrons with $p_T < 5$–$10\,\mathrm{GeV}$ and for non-isolated electrons.

1˙2.1. The performance of the electron reconstruction. The commissioning of the electron and photon reconstruction was studied with approximately 200000 minimum bias events at a center-of-mass energy of 900 GeV collected at the end of 2009 [6]. Electrons are reconstructed using both reconstruction algorithms. In the collected minimum bias sample, only very low p_T electron candidates are expected. The sample of electron candidates is predicted to be dominated by charged hadrons or electrons coming from photon conversions. In the simulation, a total of 33.9% of the reconstructed candidates are found to be matched to a generator level electron (4.6%) or photon (29.3%) in a cone of $R = 0.15$. Most of the reconstructed electrons and photons in the data sample are thus coming from fakes. Figure 2 shows the transverse momentum (left) and the pseudorapidity (right) of the electron and photon candidates. No discrepancies between

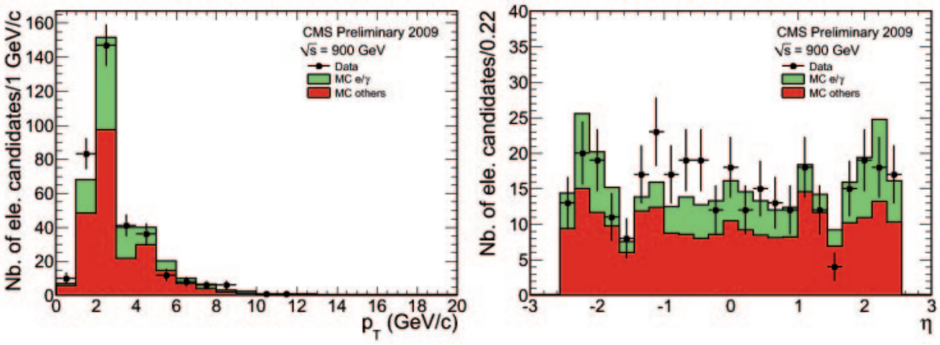

Fig. 2. – Electron and photon candidate transverse momentum (left) and pseudorapidity (right) compared to simulation (filled histogram). The simulation is normalized to the total number of electron and photon candidates in the data. The simulation expectation for electron and photon candidates, matched to a generated electron or photon, is also shown.

data and simulation are found for the kinematic properties nor for the electron isolation variables.

In the CMS experiment several methods, such as *tag and probe* methods have been developed to measure the electron and muon properties in data. They have been validated with simulations and in general a good agreement is found between the predictions of the method in the simulation and the expectations in simulated data [7, 8]. During the data-taking periods at a center-of-mass energy of 7 TeV some electron pair and muon pair candidates have been identified [9]. With the currently increasing integrated luminosity *tag and probe* methods will soon deliver data-driven estimations of reconstruction properties of the electrons and muons.

2. – b-tagging

The identification of b-jets is a key point in studies involving top quark production. The hard fragmentation, long lifetimes and high masses of B hadrons, or alternatively the relatively high fraction of semileptonic decays of the B hadron, distinguish these jets from those originating from gluons, light quarks and, to a lesser extent, c quarks. Due to the precise inner tracking system of CMS and its lepton identification capabilities, the CMS experiment is well positioned to exploit these features, using the so-called b-tag algorithms [10]. Depending on the exploited property, two main classes of b-tag algorithms can be distinguished.

2˙1. *Lifetime based b-tag algorithms.* – The first class of b-tag algorithms relies on the long lifetime of the B hadron and the consequent occurrence of displaced tracks with substantial transverse impact parameter and/or the potential presence of a reconstructed secondary vertex.

2˙1.1. Impact parameter based b-tag algorithms. Tracks are associated to a jet based on the angular distance to the jets. To minimize fake and badly reconstructed tracks, basic track quality requirements are imposed. The most powerful single-track observable is the impact parameter, IP, and is defined as the distance between the track and the primary vertex at the point of closest approach. For B hadrons with finite lifetime, the

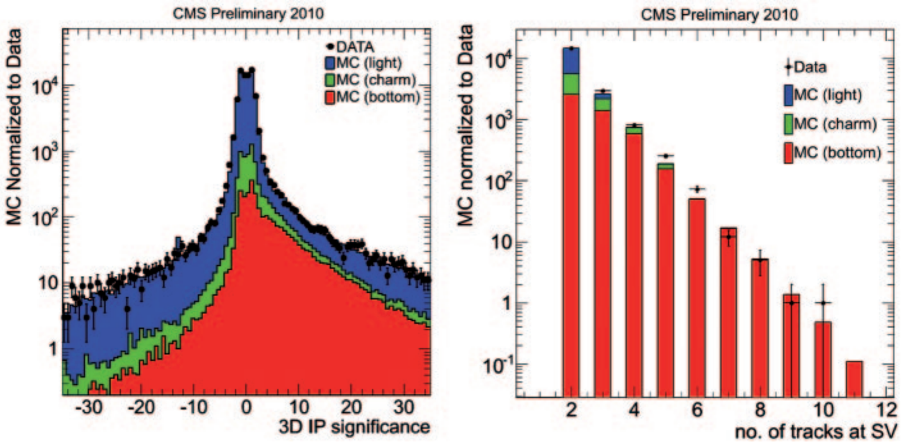

Fig. 3. – Comparison between simulation and 7 TeV proton collision data. Left: signed three-dimensional impact parameter distribution. Right: the distribution of the number of tracks associated to the secondary vertex.

impact parameter is Lorentz invariant and the typical scale is set by $c\tau \approx 0.5\,\text{cm}$. The impact parameter can be calculated either in the transverse plane or in three dimensions. In CMS, the good z resolution provided by the pixel system allows the use of the three-dimensional IP. Given that the uncertainty $\sigma(IP)$ can be of the same order of magnitude as the impact parameter itself, a better observable for b-tagging is the impact parameter significance defined as $IP/\sigma(IP)$. The impact parameter is lifetime signed and is obtained from the sign of the scalar product of the IP segment with the jet direction. A sign flip can happen due to differences between the reconstructed jet axis and the true B hadron flight direction.

Figure 3 (left) shows the signed three-dimensional impact parameter distribution for all selected tracks for anti-k_T particle flow jets with $p_T > 40\,\text{GeV}/c$ and $|\eta| < 1.5$ reconstructed in 7 TeV data [11]. The simulation is normalized to the data and shows to be in very good agreement with data.

The simplest way to produce a discriminator based on the impact parameters of tracks associated to jets is the so-called track counting algorithm. The track counting algorithm identifies a jet as a b-jet if there are at least N tracks associated to that jet with a significance of the impact parameter exceeding S. To produce a continuous discriminator for this algorithm one needs to fix the value N, and consider as discriminating variable the impact parameter significance of the N-th track (ordered in decreasing significance). If one is interested in a high efficiency for b-jets, the second track can be used, while for higher purity selections the third track is a better choice.

The track counting algorithm only combines the impact parameter of one single track, a natural extension is to use the information of all tracks. One can use the impact parameter of a track to define a track-by-track probability by extracting the probability density function for tracks not coming from b-jets. The jet probability b-tag algorithm combines the track probability previously defined. Two discriminators are provided; one named jet probability is strictly related to the combined probability that all the tracks in the jet come from the primary vertex. Alternatively the jet B probability estimates how likely it is that the four most displaced tracks are compatible with the primary vertex;

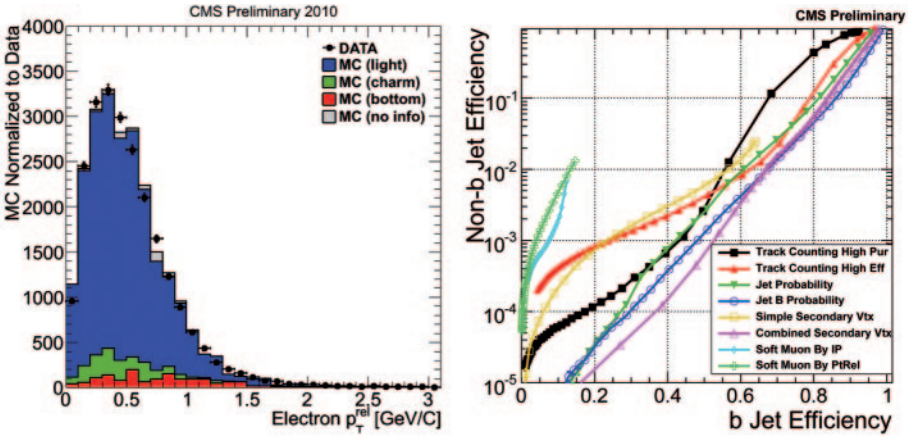

Fig. 4. – Left: distribution of p_T^{rel} compared between data and simulations. Right: performance of b-tag algorithms in simulation.

the selection comes from the fact that the average charged track multiplicity in weak b hadron decay is ≈ 5, and from the average track reconstruction efficiency, about 80% for tracks in jets.

2˙1.2. Secondary vertex based b-tag algorithms. The reconstruction of secondary vertex candidates is performed with the Adaptive Vertex Fitter algorithm. A simple b-tag algorithm based on the presence and the properties of at least one secondary vertex is called the simple secondary vertex b-tag algorithm. If no such vertex is found, the algorithm returns no discriminator, limiting its maximum b-jet efficiency to around 60–70%, the probability of finding a vertex in the presence of weak B hadron decays. The significance of the three-dimensional flight distance is used as a discriminating variable for this tagger. A more complex approach involves the use of various properties of the secondary vertices, together with other lifetime-based information, like, *e.g.*, the impact parameter significance of tracks. By using additional variables, the combined secondary vertex algorithm provides discrimination even when no secondary vertex is found. Some of the variables used in this algorithm include the transverse flight distance significance, vertex mass, number of tracks at the vertex, number of tracks in the jet, three-dimensional signed impact parameter significances for all tracks in the jet, etc. These variables are used as input to a likelihood ratio. In fig. 3 the distribution of the number of tracks associated to the secondary vertex is compared between simulations and 7 TeV data [12].

2˙2. *Soft-lepton–based b-tag algorithms.* – The second class of b-tag algorithms relies on the leptonic decay of B hadrons which delivers in 20% of the b-jets and electron or a muon with a high transverse momentum with respect to the jet axis. The presence of a lepton close to the jet is already a hint of a weak decay of a B hadron. This is complemented with some additional quantity, in order to build a discriminator. In the *soft lepton by* p_T^{rel} algorithm the p_T of the lepton with respect to the jet axis is used. In the *soft muon by IP significance* the impact parameter significance of the muon is used instead, but only when found to be positive. In all the cases, when more than one lepton is reconstructed, the one with the highest discriminator value is used. Figure 4 shows the distribution of p_T^{rel} compared between data and simulations.

2˙3. *Performance of b-tag algorithms.* – The performance of b-tag algorithms is expressed by the b-jet identification efficiency *vs.* c- or light jet rejection efficiency, additionally dividing light jets into uds- and gluon jets. Figure 4 shows an overview of the non-b jet efficiency *vs.* the b-jet efficiency for the b-tag algorithms described above for simulated multi-jet events. In [13] the effect of misalignment and miscalibration of the CMS detector on the performance of b-tag algorithms is studied. Although the assumptions in the presented study, concerning the alignment of the tracker, found to be too pessimistic with the current knowledge about the alignment performance, the conclusion that the simplest b-tag algorithms are the most robust still holds. The recommended b-tag algorithms to be used for the first data in CMS are the track counting high efficiency and the simple secondary vertex algorithm.

2˙4. *b-Tag performance measurement in data.* – In CMS a calibration of the b-tag identification efficiency based on data is foreseen. Several methods to fulfill this task have been studied over the past years. The system 8 method [14] solves a system of eight equations constructed from the total number of events in two samples with different b jet content, before and after tagging with two b-tagging algorithms. The $p_T^{\rm rel}$ method [14] relies directly on a fit of the $p_T^{\rm rel}$ distribution of the muon before and after tagging the muon-jet. An alternative method based as well on $p_T^{\rm rel}$ properties of the soft lepton is the counting method and makes no explicit use of the $p_T^{\rm rel}$ templates. A method to measure the mistag rate from negative tags is presented in [15]. Based on the constraint $|V_{tb}| = 1$, top quark events contain a large sample of b-jets. A method based on top quark events is presented in [16]. All methods expect roughly a relative uncertainty on the b-tag efficiency of 15% for 10/pb, down to 5% for 1000/pb, the top quark based measurements is only feasible with an integrated luminosity of ≈ 100/pb.

3. – Missing E_T

Neutrinos and other hypothetical weakly interacting particles pass through CMS without detection. The presence of such particles in a collision can be inferred from the imbalance of the total transverse energy of the event. Several methods exist to obtain a measure for the missing transverse energy of an event. Three methods, ordered in increasing complexity, are presented here.

3˙1. *CaloMET.* – The traditional method for missing-transverse-momentum determination at hadron colliders is based on the calorimeter information only [17]. In CMS, it is calculated as the negative value of the vector sum of the transverse energies deposited in the calorimeter towers. The latter is corrected for the presence of muons and the under-estimation of the hadronic energy in the calorimeters. First, identified muons are corrected for by replacing the minimum ionizing transverse energy expected in the calorimeters by the transverse momentum of the associated track reconstructed in the central tracker. Second, the transverse energies of the reconstructed jets are replaced by those of the jet-energy-scale corrected jets. The sequential application of the muon and the jet-energy-scale corrections defines the current standard missing transverse energy, called CaloMET.

3˙2. *tcMET.* – Track-corrected missing E_T is calculated using the uncorrected CaloMET, muons, electrons and tracks [17]. The missing E_T is corrected for muons if they pass user-defined selection criteria by subtracting the p_T of the muon and adding a calorimeter deposit, $E_T^{\rm MIP}$, consistent with the particle being approximately minimum

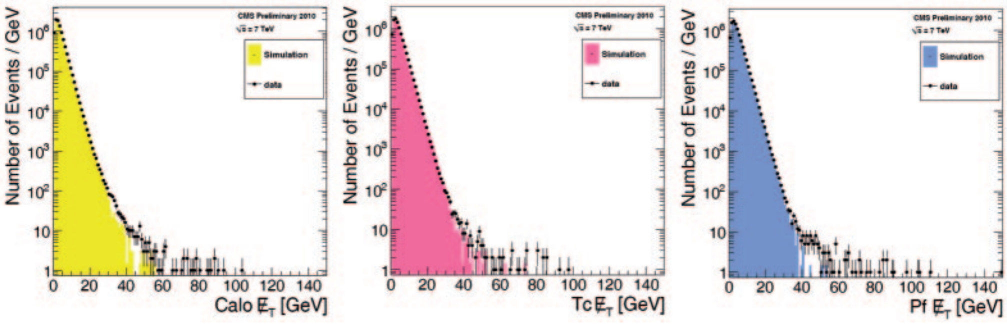

Fig. 5. – Comparison between simulated and measured missing transverse energy in 7 TeV data for caloMET (left), tcMET (middle) and pfMET (right).

ionizing. Tracks not matched to electrons or muons and passing a loose set of selections are used to correct the E_T further. The actual correction is implemented by removing the expected energy E_T deposited by each good track in the calorimeter, determined using the response function, and replacing it with the track momentum at the vertex.

3˙3. *pfMET*. – The particle-flow event reconstruction aims at reconstructing and identifying all stable particles in the event, *i.e.* electrons, muons, photons, charged hadrons and neutral hadrons, with a thorough combination of all CMS sub-detectors towards an optimal determination of their direction, energy and type. The principle to determine missing E_T after the particle-flow event reconstruction is rather simple. It consists of forming the transverse-momentum-vector sum over all reconstructed particles in the event and then taking the opposite of this azimuthal, momentum two-vector [18].

Figure 5 shows the distributions of the various missing E_T definitions for 7 TeV collision data compared to simulations [19].

REFERENCES

[1] Adam W., Mangano B., Speer T. and Todorov T., [CERN-CMS-NOTE-2006-041].
[2] Chatrchyan S. *et al.* (CMS Collaboration), *JINST*, **5** (2010) T03020 [arXiv: 0911.4022].
[3] Chatrchyan S. *et al.* (CMS Collaboration), *JINST*, **5** (2010) T03009 [arXiv: 0910.2505].
[4] Chatrchyan S. *et al.* (CMS Collaboration), *JINST*, **5** (2010) T03022 [arXiv: 0911.4994].
[5] Vanlaer P. (CMS Collaboration), *Electron and photon reconstruction in CMS* (July 2009) [CMS CM-2009/201].
[6] CMS Collaboration, *Electromagnetic physics objects commissioning with first LHC data* (2010) [CMS-PAS-EGM-10-001].
[7] CMS Collaboration, *Measuring Electron Efficiencies at CMS with Early Data* (Dec 2008) [CMS-PAS-EGM-07-001].
[8] CMS Collaboration, *Towards a measurement of W and Z cross sections into muons in pp collisions at 14 TeV* [CMS-PAS-EWK-07-002].
[9] CMS Collaboration, *W and Z candidates recorded with the CMS detector during the first LHC runs in 2010* (May 2010) [CERN-CMS-DP-2010-012].
[10] CMS Collaboration, *Algorithms for b Jet identification in CMS* [CMS-PAS-BTV-09-001].

[11] CMS Collaboration, *B-tagging observables and event display from first 7 TeV data* (May 2010) [CERN-CMS-DP-2010-015].

[12] CMS Collaboration, *Tracking and Vertexing Results from First Collisions* (2010) [CMS-PAS-TRK-10-001].

[13] CMS Collaboration, *Impact of Tracker Misalignment on the CMS b-Tagging Performance* (Aug 2007) [CMS-PAS-BTV-07-003].

[14] CMS Collaboration, *Performance Measurement of b tagging Algorithms Using Data containing Muons within Jets* (July 2007) [CMS-PAS-BTV-07-001].

[15] CMS Collaboration, *Evaluation of udsg Mistags for b-tagging using Negative Tags* (July 2007) [CMS-PAS-BTV-07-002].

[16] Lowette S., D'Hondt J., Heyninck J. and Vanlaer P., *Offline Calibration of b-Jet Identification Efficiencies* (Jan 2006) [CERN-CMS-NOTE-2006-013].

[17] CMS Collaboration, *Performance of Track-Corrected Missing Transverse Energy in CMS* (Dec 2009) [CMS-PAS-JME-09-010].

[18] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET* (Apr 2009) [CMS-PAS-PFT-09-001].

[19] CMS Collaboration, *Jet and MET Commissioning Results from 7 TeV Collision Data* (May 2010) [CERN-CMS-DP-2010-015].