The Compact Muon Solenoid Experiment

CMS Draft Note

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

2011/06/09
Head Id: 57905
Archive Id: 29932:59897MP
Archive Date: 2011/05/27
Archive Tag: trunk

Status of b-tagging tools for 2011 data analysis

The CMS Collaboration
University on the Moon

Abstract

The identification of jets containing the weak decay of a B-hadron is an essential tool for a wide range of analyses in the context of the Standard Model and beyond. A variety of algorithms exploit the long lifetime and the presence of soft leptons to discriminate these jets from those associated to light quarks. The status of the b-tagging tools and their commissioning with 2011 data is presented. New developments and improvements of the b-tagging algorithms are also documented.

1 Introduction

The identification of jets originating from b quarks is a crucial element for many physics analyses. In particular, high branching ratios to b quarks characterize a variety of Standard Model (SM) and discovery channels like the measurement of bottom or top pair production, the search for Higgs bosons and different other New Physics scenarios.

The hard fragmentation, the long lifetimes and high masses of B hadrons and the relatively high fraction of semileptonic decays distinguish these jets from those originating from gluons, light quarks and - to a lesser extent - from c quarks. Due to its precise inner tracking system and its lepton identification capabilities the CMS experiment is particularly suited for exploiting these features.

Detailed definitions of the studied quantities are given in [1], along with an explanation of the b-tagging algorithms. The b-tag commissioning with first collision data at 7 TeV is reported in [2]. The present note presents an update of the commissioning activities including data from the 2011 run. The validation of the main variables is discussed in Sections 4 to 8. One of the main differences to [2] is the presence of pileup, which is discussed in Section 10. The development and commissioning of additional higher-level b-tagging algorithms is documented in Section 11.

2 High Level Trigger

responsible editor: Jyothsna
Technical implementation to use b-tagging at High Level Trigger (HLT) enviromnent exists since few years **FIX ME add 2006 and 2007 references here**. These implementations exploit couple of properties of jets originating from b-quarks namely semi-leptonic decays and long lifetime. "BTagMu" implementation makes use of semi-leptonic b-decays and tags jets at HLT containing muons associated to them. Lifetime based b-tagging implementation referred to as "BTagIP" implementation relies mostly on offline equivalent of Track Counting High Efficiency algorithm for tagging the jets at HLT.

Implementation details of "BTagMu" and "BTagIP" along with the triggers developed for calibration and physics use cases is presented in the next sections.

2.1 BTagMu HLT implementation and use cases

Events containing a μ associated with a jet can be used to the measure the b-tag performance from data. Since only paths using BTagMu implementation are the b-tag calibration paths, these are explicitly mentioned in the below description of the implementation. The Level 1 seeds for these paths require a muon and a jet at Level 1 with E_T cuts specific to each designed HLT path and are listed in Table 1.

- Level-2: In 2011 Anti k_T corrected calo jets are used as Level-2 jets. In order to reduce the HLT rates dijets are required at Level-2 with varying jet E_T thresholds. These jet E_T cuts are tuned to have enough statistics available offline to derive the efficiency and scale factors in various offline jet E_T and η bins. In the recent data taking 20 GeV, 40 GeV, 70 GeV and 110 GeV are chosen as the thresholds.
- Level-2.5: The soft lepton b-tagging algorithm is run on the 4 highest E_T jets in the event with E_T threshold chosen same as the E_T cut on jets at Level-2. Level-2 muons, reconstructed using muon detector hits, are required to be near one of these 4 highest

E_T jets, $\Delta R(\mu, jet) < 0.4$, using the Soft Lepton b-tagging algorithm. Events pass if at least one jet is soft mu-tagged.

- Level-3: Refined tracking is used to reconstruct Level-3 muons. Only those Level-3 muons which pass $\mu\ p_T > 5\ GeV$ and have > 0 number of hits are used to select muon-in-jets with $\Delta R(\mu, jet) < 0.4$ requirement using the Soft Lepton b-tagging algorithm. Events pass if at least one jet is soft mu-tagged.

The Soft Lepton b-tagging algorithm used in both Level-2.5 and Level-3 make use of the same methods used in the offline reconstruction and they differ in muon and jet inputs at HLT wrt Offline.

The names of the "BTagMu" based triggers along with the Level-1 seeds are listed in Table 1.

Path name	**Level-1 seed**	**HLT prescale**	**Rate @1E33**
HLT_BTagMu_DiJet20_Mu5	L1_Mu3_Jet16	350	1.62 Hz
HLT_BTagMu_DiJet40_Mu5	L1_Mu3_Jet20	100	1.32 Hz
HLT_BTagMu_DiJet70_Mu5	L1_Mu3_Jet28	15	1.36 Hz
HLT_BTagMu_DiJet110_Mu5	L1_Mu3_Jet28	2	1.62 Hz

Table 1:

The offline selection required two jets of > 30 and $|\eta| < 2.4$ in order to stay within the tracker acceptance and in the range where b-tagging is typically applied. In order to increase the purity in B $\rightarrow \mu +$ X decays the event had to contain exactly one muon of > 6 and $|\eta| < 2.4$ with the following additional quality criteria:

- the muon was reconstructed as a "GlobalMuon" with at least one valid muon hit, more than one matching segment in the muon chambers and a $\chi^2/\mathrm{n}dof < 10$;
- the corresponding inner track had > 10 hits with at least one hit in the pixel system and a $\chi^2/\mathrm{n}dof < 10$;
- the z-distance between the reference point of the muon and the selected primary vertex was < 1cm.

The muon had to be in a cone defined by < 0.4 around the associated jet ("").

Turn on curves for the different BTagMu triggers have been computed by scaling the pT spectra for the data collected by each trigger by its integrated luminosity, and dividing the so obtained distribution by the one of the HLT_BTagMu_DiJet20_Mu5 trigger. The turn on curve for this last trigger has been computed by taking the pT spectrum of the muon-jetsin data collected by the HLT_Mu5 trigger as a reference. All these turn on curves are shown in Fig. **??**.

2.2 BTagIP HLT implementation and use cases

Many standard model and exotic physics channels contain b jets in the final state. By explicitly requiring b-tagged jets in the HLT paths for these physics channels, one can lower the jet E_T thresholds than cutting harder on the jets for rate reduction, increasing the purity as well as trigger efficiency for these channels.

A generic BTagIP implementation at HLT is described below. This implementation is adapted to suit the needs of physics channels by the trigger developers. The Level-1 seeds are also up to the trigger developers to choose depending on the physics channel and final state they are interested in.

- Level-2: The jet E_T thresholds and the number of jets vary depending on the needs

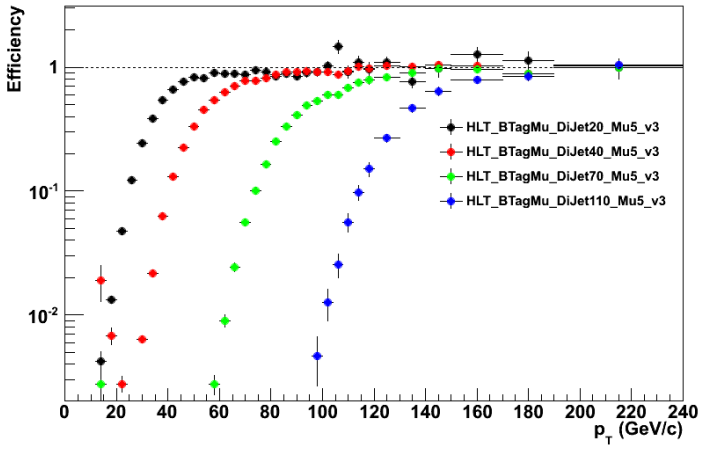

Figure 1: Turn on curve of all the above four paths using 2011 data

of the physics channel.

- Level-2.5: Tracks are reconstructed using the Pixel Tracker alone (each with at least 3 hits), and are then used to reconstruct the 1D primary vertex. The b-tag is run on 4-6 highest E_T jets in the event with E_T threshold chosen by the trigger developer, using the pixel tracks and the primary vertex as input. Most of the physics paths use Online Beam Spot in the transverse plane and fast estimation in Z as the primary vertex. 3D Primary vertex (PV) can also be used as reference. More details about the PV methods usage at HLT and the performnace are provide in the next subsection. Jets are tagged as b-jets if they have at least 1 or 2 tracks with 3D impact parameter significance greater than a threshold value tuned by the physics trigger developers. Events pass if at least one or two jets are b-tagged.
- Level-3: Tracks are reconstructed regionally in a cone size $\Delta R = 0.25$ either around the jets tagged as b-jets at Level-2.5 or around all the Level-2.5 jets. The choice of which jets are fed for the regional track reconstruction is also upto the developers. The track reconstruction is partial. stopping after 8 hits have been assigned to a track. The b-tag uses these tracks and the primary vertex reconstructed at Level-2.5. It selects jets having at least 2 tracks with 3D impact parameter significance greater than a threshold value tuned by the trigger developer. Events pass if at least one or two jets are b-tagged.

Trigger performance of couple of physics channels using BTagIP are presented in Fig **??** for QuadJet50_BTagIP trigger path and Fig **??** for HT300_CentralJet30_BTagIP_PFMHT55 trigger path as a function of offline b-tag algorithms.

QuadJet50_BTagIP trigger path is developed for all hadronic ttbar final state and to collect the events containing at least four jets at the HLT level with $E_T > 50$ GeV and amongst at least one of them is required to be b-tagged by TCHE algorithm at HLT. The cut TCHE cut chosen at HLT level is 2 for this path and as can be seen in bottom plot in Fig **??** the efficiency is 50% at offline TCHP value of 2.

SUSY analysis looking at missing E_T, jets and b-jet final state and developed trigger path HT300_CentralJet30_BTagIP_PFMHT55 to collect such events. Jet with $E_T > 30$ GeV is required to be b-tagged by TCHE algorithm at HLT and the cut applied is at 4.

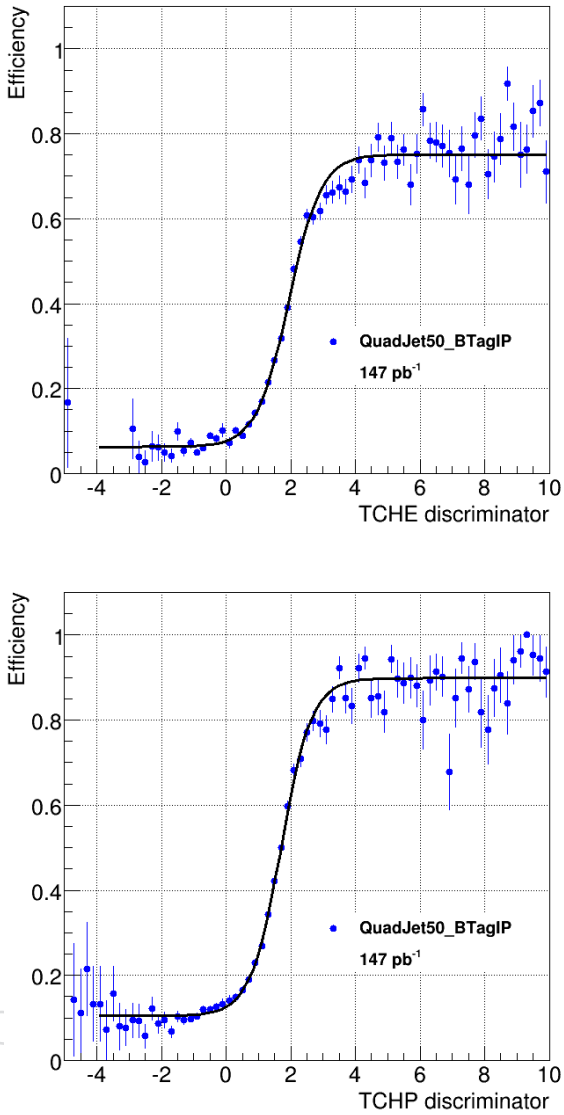

Figure 2: Turn on curve for QuadJet50_BTagIP trigger as a function of offline TCHE and TCHP discriminator obtained using 2011 data

2.3 3D Primary Vertex

responsible editor: Carlotta

To limit the CPU timing, the standard High Level Trigger reconstruction performs only a fast partial (1D) estimation of the primary vertex. Prompt pixel tracks are clustered in z and the longitudinal vertex position is determined by the Divisive Vertex Fitter algorithm. The online beamspot is used as rough estimation of the transverse vertex coordinate.

The full (3D) primary vertex reconstruction, implementing the Adaptive Vertex Fitter, could provide a reliable reference point for the track impact parameter calculation, input to the b-tagging algorithms, in case of movements of the beamspot. This will be crucial, at high luminosity, for those trigger paths that rely on b-tagging to reduce the rate by a factor greater than 10. To avoid any bias from an incorrect beamspot position, no beam constraint is applied in the vertex fitting procedure. The Gap clusterizer is used. To be included in the clusterization, pixel tracks are required to have at least 3 pixel hits, $\chi^2/ndof \leq 100.0$, and transverse impact

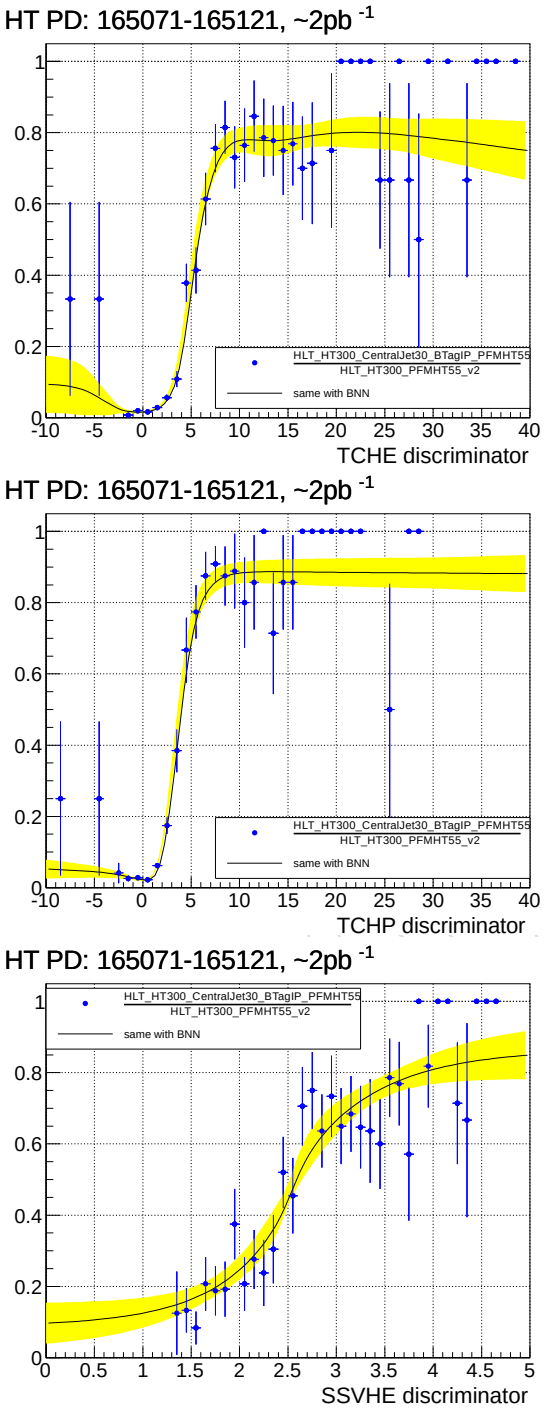

Figure 3: Turn on curve for HT300_CentralJet30_BTagIP_PFMHT55 as a function of offline TCHE, TCHP and SSVHE discriminator obtained using 2011 data

parameter significance ≤ 100.0. An additional cut on the track momentum can be applied and tuned in order to achieve reasonably good performance with a limited CPU consumption. Tracks that pass the selection are grouped into primary vertex candidates if their z distance to the nearest neighbor is smaller than 1 mm. A reconstructed primary vertex is finally rejected if the transverse distance to the beam is larger than 2 cm.

The performance of the 3D primary vertex reconstruction, compared to the standard 1D, is

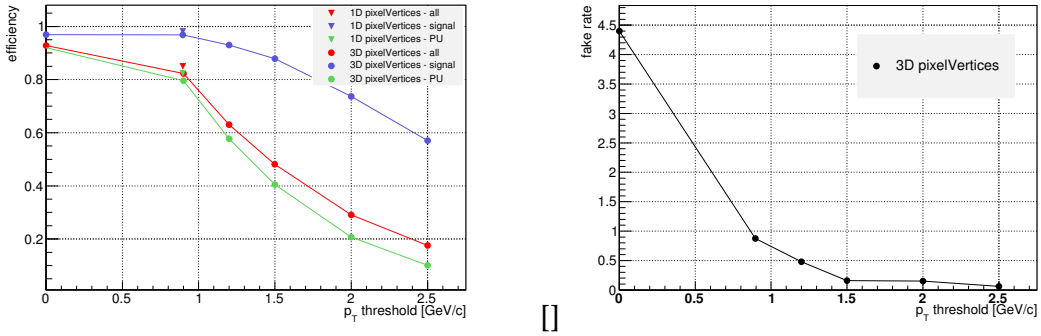

Figure 4: (a) primary vertex finding efficiency provided by the 3D reconstruction, for all vertices in the event, and separately for signal and pile up vertices, as a function of the p_T threshold applied in the pixel track selection. The result for the 1D pixel vertex reconstruction is also shown for a comparison. (b) rate of fake vertices in the event, as a function of the p_T threshold. The corresponding fake rate for the 1D reconstruction is 12.5% for a p_T cut of 0.9 GeV/c.

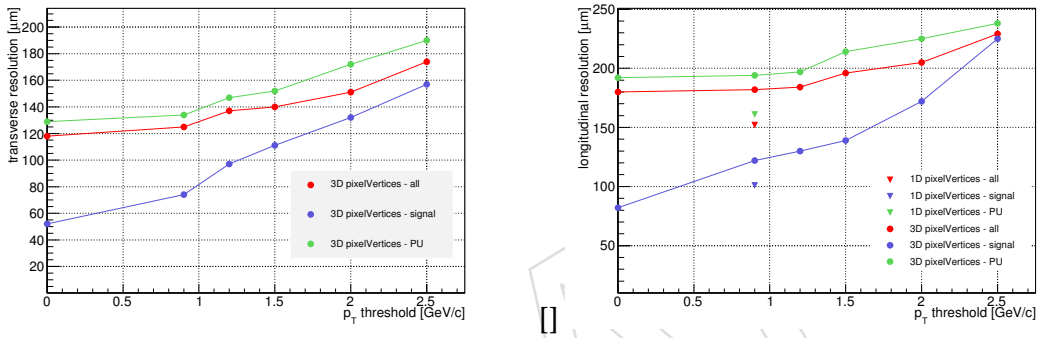

Figure 5: (a) transverse and (b) longitudinal 3D primary vertex position resolution as a function of the p_T threshold applied in the pixel track selection. In (b) the result for the 1D reconstruction is also shown for comparison.

evaluated using a Monte Carlo sample of QCD events with parton p_T greater than 15 GeV/c, with on average 10 pile up events. Fig.4 and 5 summarize the results for vertex finding efficiency, rate of fake vertices in the event, transverse and longitudinal position resolution, as a function of the p_T cut applied to the pixel tracks. From Fig.2.3 it is evident that a working point for this algorithm should be chosen in the region up to about 1.2 GeV/c, where the signal vertex efficiency is higher than 90%, and competitive with the 1D reconstruction. The fake rate in this range is between 4.5% and 0.5%, to be compared to 12.5% that we obtain for the 1D reconstruction. For the signal vertex, the longitudinal position resolution ranges between 80 and 120 μm. The transverse resolution is between two and four times worse than the beamspot width (about 25 μm), which is the limitation of the 1D algorithm. This should still assure sufficiently good b-tagging performance for the purpose of the online event selection.
A summary of the 3D vertex reconstruction performance is given in Table 2, together with an estimation of the running time (CPU) of the module.

p_T threshold (GeV/c)	running time (ms)
no threshold	32.6
0.9	15.5
1.2	10.0
1.5	7.1
2.0	4.8
2.5	3.9

Table 2: Running time (CPU) of the 3D primary vertex producer module as a function of the p_T threshold applied in the pixel track selection.

3 Data and Monte Carlo samples, reconstruction and selection

Monte Carlo Samples have been generated with PYTHIA6, tune Z2, in several bins of $\hat{p}_T$. Table 3 shows the names of the MC samples, and the number of processed events. The pileup conditions in data change rapidly so the Monte Carlo cannot be expected to match the data exactly. The Monte Carlo samples have therefore been gererated with a flat pileup distribution including a poissonian tail. This scenario covers roughtly the conditions of the 2011 run. The residual differences in pileup conditions are taken into account with a reweighting procedure as explained later in this section. The samples in Table 3 have been used for the main commissioning and validation studies in Sections 5 to 7. Another set of samples requiring the presence of a muon at generator level has been produced for the studies of muon jets in Section 8. The details of the muon enriched samples are given in Table 4.

Table 3: QCD Monte Carlo samples. All sample names have to be extended with the suffix /Spring11-PU_S1_START311_V1G1-v1/AODSIM.

sample	# events
/QCD_Pt_5to15_TuneZ2_7TeV_pythia6	3296192
/QCD_Pt_15to30_TuneZ2_7TeV_pythia6	8213600
/QCD_Pt_30to50_TuneZ2_7TeV_pythia6	6529320
/QCD_Pt_50to80_TuneZ2_7TeV_pythia6	4301392
/QCD_Pt_80to120_TuneZ2_7TeV_pythia6	6407732
/QCD_Pt_120to170_TuneZ2_7TeV_pythia6	6090400
/QCD_Pt_170to300_TuneZ2_7TeV_pythia6	5684160
/QCD_Pt_300to470_TuneZ2_7TeV_pythia6	6336960

Table 4: Muon Enriched Monte Carlo samples. All sample names have the suffix /Spring11-PU_S1_START311_V1G1-v1/AODSIM.

sample	# events
/QCD_Pt-15to20_MuPt5Enriched_TuneZ2_7TeV-pythia6	2884915
/QCD_Pt-20to30_MuPt5Enriched_TuneZ2_7TeV-pythia6	11352301
/QCD_Pt-30to50_MuPt5Enriched_TuneZ2_7TeV-pythia6	10909951
/QCD_Pt-50to80_MuPt5Enriched_TuneZ2_7TeV-pythia6	10686315
/QCD_Pt-80to120_MuPt5Enriched_TuneZ2_7TeV-pythia6	3183540
/QCD_Pt-120to150_MuPt5Enriched_TuneZ2_7TeV-pythia6	991024
/QCD_Pt-150_MuPt5Enriched_TuneZ2_7TeV-pythia6	1015900

We use "Particle Flow Jets" (PFjets) [3] using the anti-k_T jet clustering method [4] with cone radius parameter $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.5$. In addition, "L2L3" Jet energy corrections and the following jet ID cuts are applied:

```
pt > 10.0 && abs(eta) < 2.5 && neutralHadronEnergyFraction < 1.0 &&
neutralEmEnergyFraction < 1.0  && nConstituents > 1 &&
chargedHadronEnergyFraction > 0.0 && chargedMultiplicity > 0.0 &&
chargedEmEnergyFraction < 1.0
```

The jet ID cuts are introduced after the L2L3 energy corrections.

Tracks are reconstructed using the standard CMS iterative tracking procedure. The Combinatorial Kalman Filter algorithm [5–7] is applied in an iterative way with increasingly relaxed impact parameter constraint and different seeding layers.

The primary vertex (PV) is reconstructed from all tracks in the event which are compatible with the beam spot. The new "Deterministic Annealing Filter" (DA) algorithm is used for PV reconstruction (**FIXME: reference plus more details**)

The CMSSW software version 4.1.6 is used to analyze the Monte Carlo samples. This version has been patched with a backport of the DA primary vertex reconstructor and a fixed version of the secondary vertex producer. Data samples are analyzed with CMSSW 4.2.3 without any additional patches.

The commissioning studies in Sections 5 to 7 are using a single jet trigger `HLT_Jet60` for data, while the trigger `HLTJet30U` is applied in Monte Carlo samples. A jet p_t threshold of 60 GeV is applied both in data and Monte Carlo. It has been verified that the resulting jet p_t spectra are in reasonable agreement between data and Monte Carlo.

The studies requiring muons in jets have been made applying the jet+muon di-jet trigger `HLT_BTagMu_DiJet40_Mu5` in data. Both in data and Monte Carlo, the requirement of two jets with $p_t > 60$ and one jet with $p_t > 65$ GeV is applied. Muons are required to have transverse momentum $p_t > 7$ GeV.

In addition, another set of validation plots are produced with lower trigger threshold (`HLT_Jet30` in data and `HLTJet15U` in Monte Carlo) and lower jet momentum cut of $p_t > 30$ GeV in case of the single jet triggers. For the b-jet triggers another set of plots has been produced with `HLT_BTagMu_DiJet20_Mu5`, requiring two jets with $p_t > 40$ GeV, one of them with $p_t > 45$ GeV. Details are given in Appendix A.

An overview of the used data samples, triggers and effective integrated luminosity (including trigger prescales) is given in Table 3.

sample	trigger	effective lumi
`/Jet/Run2011A-PromptReco-v1(v2)`	`HLT_Jet30`	$3.06 \cdot 10^{-3}$ pb^{-1}
`/Jet/Run2011A-PromptReco-v1(v2)`	`HLT_Jet60`	0.056 pb^{-1}
`/METBTag/Run2011A-PromptReco-v1(v2)`	`HLT_BTagMu_DiJet20_Mu5`	1.62 pb^{-1}
`/METBTag/Run2011A-PromptReco-v1(v2)`	`HLT_BTagMu_DiJet40_Mu5`	3.12 pb^{-1}

The number of reconstructed primary vertices is shown in Section 10. It is visible that Monte Carlo does not describe the pileup multiplicity correctly. We therefore apply a vertex reweighting procedure in which all distributions obtained from Monte Carlo are reweighted so that the vertex multiplicity agrees between data and Monte Carlo. This reweighting procedure has been applied for all distributions in Sections 5 to 8.

4 Track selection

Tracks are associated to jets using a cone in $\eta - \phi$ space defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$. A new development which introduces a variable cone size is discussed in Section 11.

Furhter selection criteria for tracks are listed in the following. The corresponding distributions for all tracks within a distance $\Delta R < 0.5$ to the jet axis are displayed in Figures 6 to 8.

- number of pixel hits $>= 2$ (Figure 6)
- number of tracker hits (including pixel) $>= 8$ (Figure 6)
- transverse impact parameter $|IP_{2D}| < 0.2$ cm (Figure 6)
- transverse momentum $p_t > 1$ GeV$/c$ (Figure 7)

- normalized $\chi^2 < 5$ (Figure 7)
- longitudinal impact parameter $|IP_z| < 17$ cm (Figure 7)
- distance to jet axis < 0.07 cm (Figure 8)
- decay length < 5 (Figure 8)

Figures 6 to 8 show the track selection variables without any cuts, while Figures 9 to 11 show the same variables but with all cuts, except for the cut on the variable displayed (n-1 cuts).

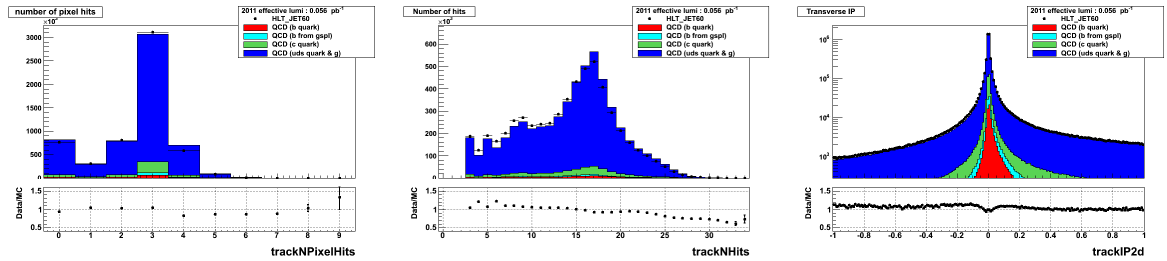

Figure 6: Left: number of hits in the pixel detector, middle: total number of hits in the tracker, right: transverse track impact parameter. No cuts on track variables were applied.

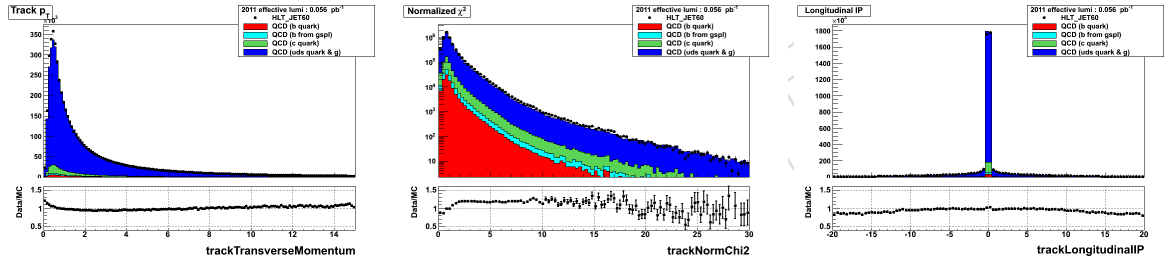

Figure 7: Left: transverse track momentum, middle: normalized track χ^2, right: longitudinal impact parameter. No cuts on track variables were applied.

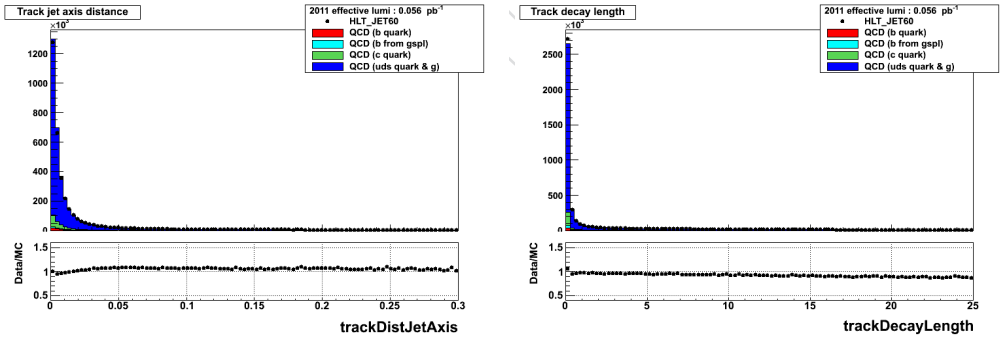

Figure 8: Left: distance to jet axis, right: track decay length. No cuts on track variables were applied.

This track selection is used for all impact parameter based algorithms, i.e. the track counting and track probability algorithms. The secondary vertex based algorithms apply slightly different selection criteria (only those which are different are listed in the following):

- jet-track association cone $\Delta R < 0.3$
- distance to jet axis < 0.2 cm
- no cut on decay length

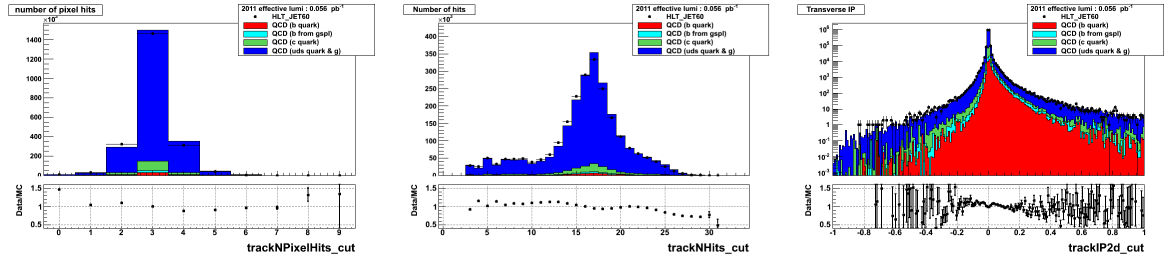

Figure 9: Left: number of hits in the pixel detector, middle: total number of hits in the tracker, right: transverse track impact parameter. All cuts on track selection variables were applied, except for the cut on the displayed quantity.

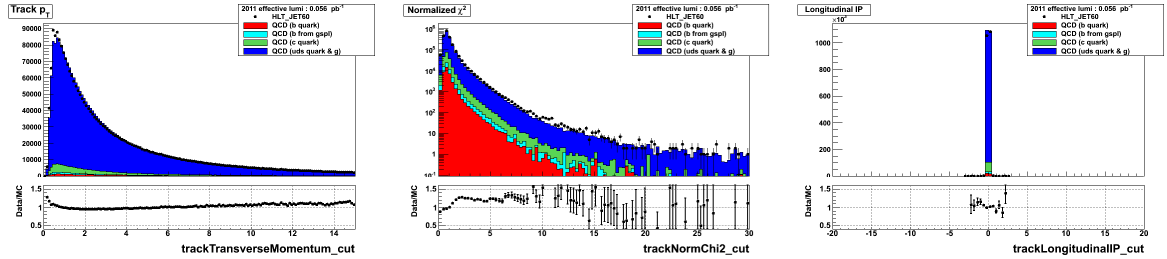

Figure 10: Left: transverse track momentum, middle: normalized track χ^2, right: longitudinal impact parameter. All cuts on track selection variables were applied, except for the cut on the displayed quantity.

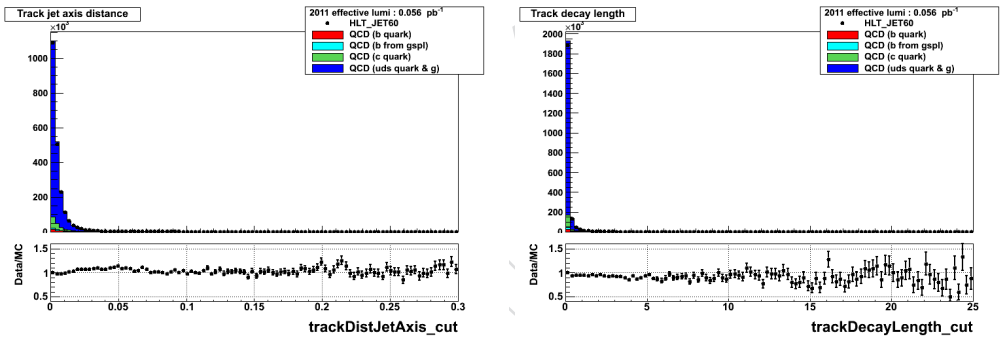

Figure 11: Left: distance to jet axis, right: track decay length. All cuts on track selection variables were applied, except for the cut on the displayed quantity.

- track quality class = "high purity"

The average number of tracks per jet is displayed in Figure 12 for the case with and without track selection cuts. The average number of tracks also depends on the jet energy which is shown in Figure 13. The discrepancy between data and simulation is attributed to the Monte Carlo event generator which is not reproducing the charged particle kinematics perfectly.

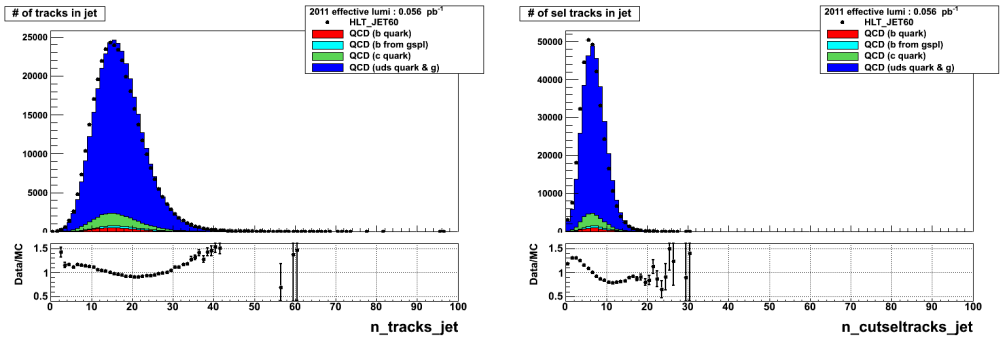

Figure 12: Left: number of tracks within $\Delta R < 0.5$ of the jet axis. Right: the same for tracks passing the selection criteria of the IP based algorithms as explained in the text.

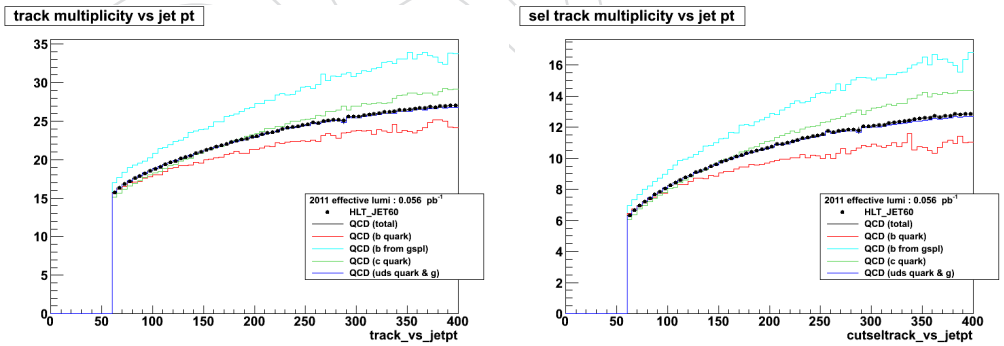

Figure 13: Left: average number of tracks associated to a jet depending on transverse jet momentum p_t. Right: average number of selected tracks associated to a jet depending on transverse jet momentum p_t.

5 Impact parameter

The impact parameter (IP) is defined as the minimum distance between the primary vertex and the trajectory of the track. Tracks produced by long lived particles such as B mesons are expected to have a sizable IP. In the ultra-relativistic limit the IP is Lorentz-invariant to a good approximation due to the cancellation of boost effects and the angle of the decay products with respect to the flight path. The precision of the IP measurement can be different from track to track and is between 30 μm and several hundreds μm. Given that the uncertainty can be of the same order as the IP value, the IP significance $S = IP/\sigma_{IP}$ is used for tagging b-jets.

The IP is "lifetime signed": tracks originating from the decay of particles traveling in the same direction of the jet are signed as positive, while those in opposite direction are tagged as negative. This is obtained by using the sign of the scalar product of the IP segment with the jet direction. It should be noted that a "sign flip" can happen to track produced in the region between the jet direction and the actual B-hadron flight direction.

The IP can be measured either in the transverse plane only or in three dimensions. The high resolution of the pixel detector also along the z coordinate allows the use of the 3D IP: despite the precision in z being slightly inferior with respect to the one in the transverse plane, by using the 3D significance the precision is not spoiled as the measurement errors are correctly taken into account.

3D Impact Parameter value, error and significance for first, second and third track in the jet (ordered by IP significance) are displayed in Figures 14 to 16. Figure 17 shows the same for all selected tracks in a jet (i.e. not ordered by IP significance).

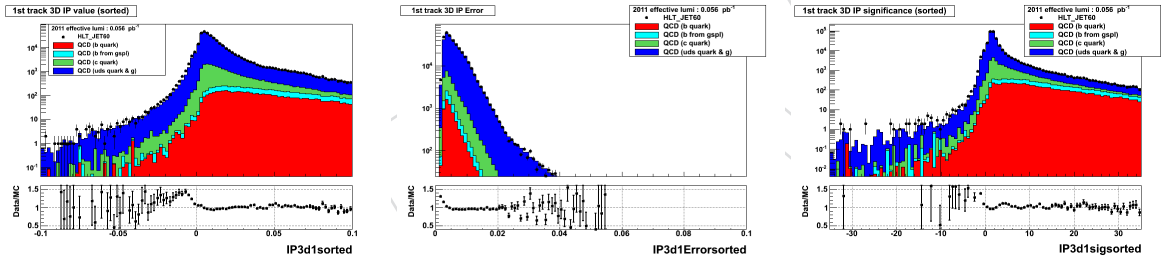

Figure 14: Left: IP value, middle: IP error, right: IP significance for the first track in the jet, ordered by IP significance.

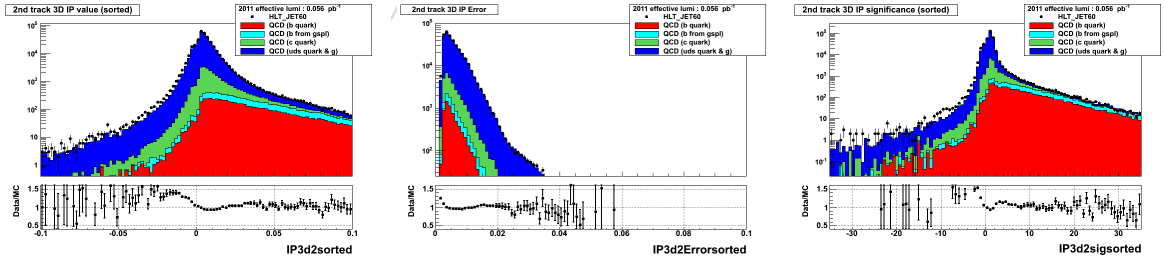

Figure 15: Left: IP value, middle: IP error, right: IP significance for the second track in the jet, ordered by IP significance.

The impact parameter has slightly different behaviour depending on the track momentum. This is shown in Figure 18 which displays the track IP values for six different track p_t bins.

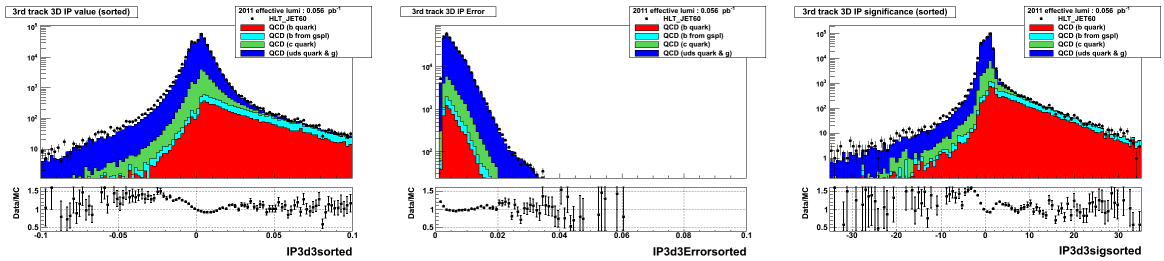

Figure 16: Left: IP value, middle: IP error, right: IP significance for the third track in the jet, ordered by IP significance.

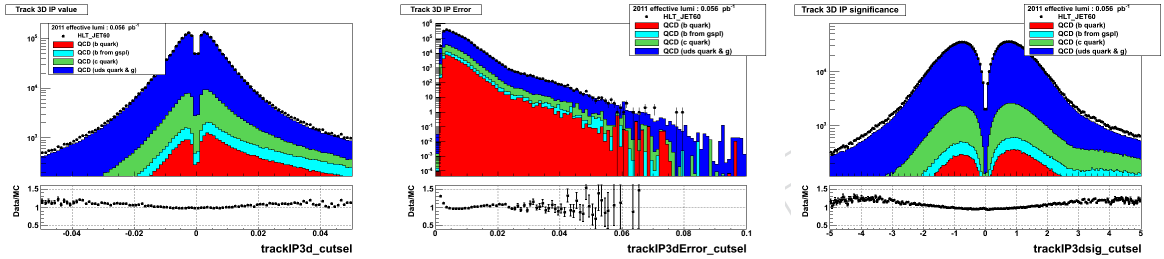

Figure 17: Left: IP value, middle: IP error, right: IP significance for all selected tracks in the jet. The track selection as defined in Section 4 has been applied.

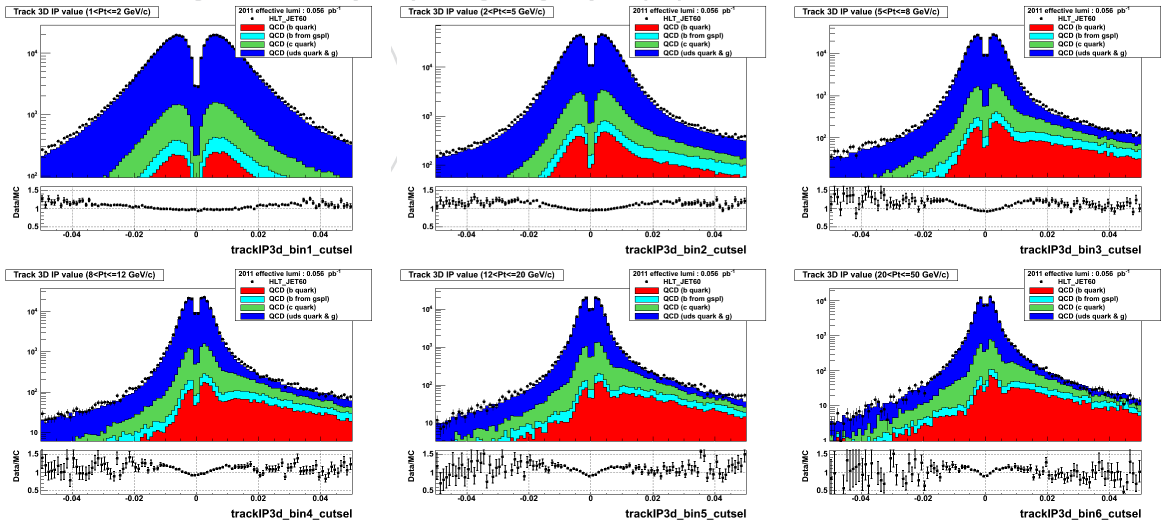

Figure 18: The IP value in six bins of track p_t. From top left to bottom right (in units of GeV/c): $1 < p_t < 2$; $2 < p_t < 5$; $5 < p_t < 8$; $8 < p_t < 12$; $12 < p_t < 20$; $20 < p_t < 50$.

6 Secondary vertices

Secondary vertex reconstruction is performed using the Adaptive Vertex Finder [8], which performs a fully inclusive vertex search in a list of given tracks. The approach is to fit a vertex from all tracks and iteratively repeat the fit with tracks that were not compatible with the vertices obtained in previous iterations. This procedure is repeated until the list of tracks is exhausted or the vertex fit fails. The Adaptive Vertex Fitter is used for the actual vertex fit performed at each iteration. Since all candidate tracks are passed to it at once, it is able to intrinsically identify and deal with outliers to allow for the fit to converge. It therefore applies an iterative procedure, by which outlier tracks are increasingly downweighted. This is done until the fit converges and thus only compatible tracks, which have sizable weights, remain.

The parameters used are primcut = 1.8 and seccut = 6.0. Both denote the track-vertex compatibility cutoff parameter used for the first and all subsequent fits, respectively. The first fit attempt is additionally constrained to the beam spot in order to avoid a successful fit of a secondary vertex with the small cutoff parameter designed for identification of tracks from the primary vertex. The cutoff parameter of 6.0 is deliberately chosen this large in order to increase the vertexing efficiency for cases of a b-c decay chain where the two individual secondary and tertiary vertices cannot be resolved, but both decays yield enough tracks to form a common vertex. While this vertex definition is slightly unphysical, it increases the b-tagging performance of the combined secondary vertex algorithm. The vertex finder considers a track to be an outlier if it has been assigned a fit weight of less than 0.5.

The resulting list of vertices is then subject to a cleaning procedure which applies the following selection criteria:

- fraction of tracks shared with primary vertex < 0.65
- distance from beam spot in transverse plane < 2.5 cm
- DR of the flight direction with respect to the jet axis < 0.5
- $D_{xy}/\sigma_{D_{xy}}$ (2D flight distance significance with respect to reconstructed primary vertex) > 3
- $D_{xy} > 0.1$ mm

In addition, a rejection of vertices due to K_s mesons is applied by rejecting vertices with an invariant mass in the K_s mass window of 0.5 ± 0.05 GeV$/c^2$.

The average charged track multiplicity of a B hadron decay is about five and despite the tight track quality cuts the efficiency of being able to reconstruct respective decay vertices is very high. Efficiency limiting factors in reconstruction arise from tracking inefficiencies, tracks lost due to quality or acceptance cuts or tracks that are also compatible with the primary vertex and hence excluded from the secondary vertex fit. The number of reconstructed vertices per jet is shown on the left in Figure 19, while the number of tracks at the reconstructed secondary vertex is shown in the middle. The dependence of the track multiplicity on the jet momentum is shown on the right. It is visible that the fraction ob b-jets is significantly enhanced for vertices with three or more tracks. This is also visible in Figure 20 which shows the reconstructed vertex mass with a minimum of two (left plot) or three (middle plot) tracks at the Secondary Vertex. The right plot in Figure 20 shows the transverse momentum of the secondary vertex (with two tracks) which is determined using the sum of the momentum vectors of the tracks at the vertex.

An important quantity which is sensitive to the lifetime of B hadron decays is the distance between primary and secondary vertex. As for the impact paramter, the significance of this quantity is used in b-tagging algorithms. Figure 21 shows the flight distance significance, the

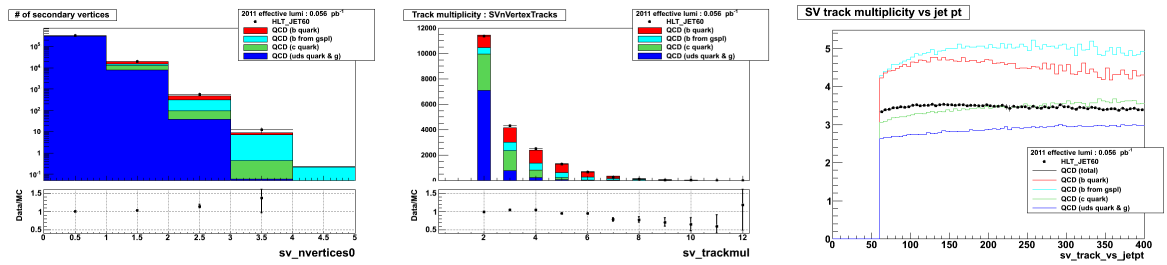

Figure 19: Left: number of reconstructed secondary vertices per jet, middle: number of tracks at the reconstructed secondary vertex, right: average number of tracks at the secondary vertex versus jet p_t.

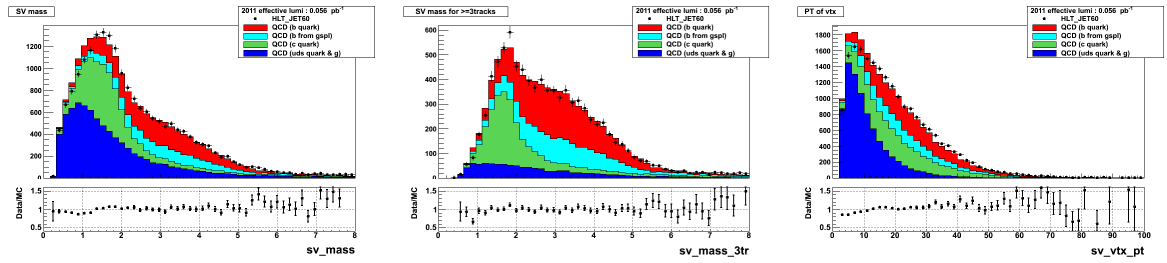

Figure 20: Left: vertex mass with two or more reconstructed tracks at the vertex. Middle: vertex mass with three or more tracks at the vertex. Right: transverse momentum of the secondary vertex (with two or more tracks).

normalized χ^2 of the vertex fit and the energy ratio of tracks at the secondary vertex with respect to all tracks in the jet.

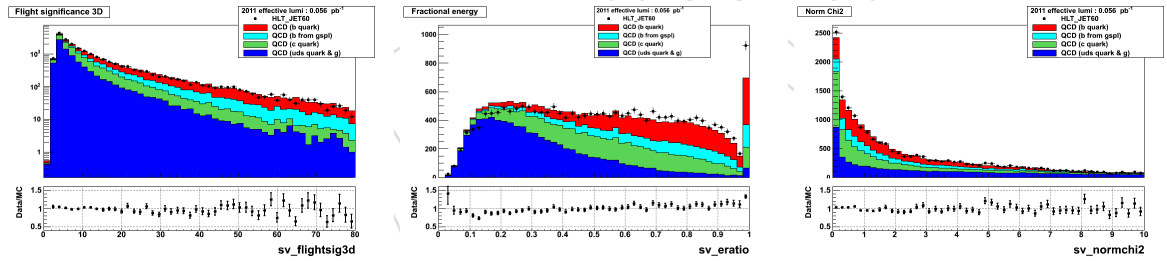

Figure 21: Left: vertex flight distance significance. Middle: ratio of track energy at the secondary vertex with respect to all selected tracks in the jet. Right: vertex fit normalized χ^2.

Two different directions can be defined at the secondary vertex: the flight direction, which points from the primary vertex to the secondary vertex and the direction of the vertex momentum which is the sum of all vertex track momenta. The angle between these two directions measured in ΔR as well as the angle with the jet axis are shown in Figure 22.

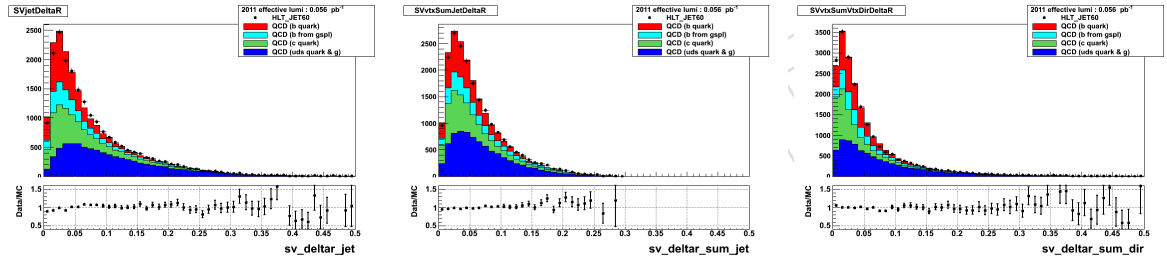

Figure 22: Left: angular distance in ΔR between jet axis and vertex direction. Middle: angular distance in ΔR between jet axis and the sum of track momenta at the vertex, Right: angular distance in ΔR between vertex direction and the sum of track momenta at the vertex.

7 Discriminators and tagging rates

Details about the calculation of the discriminators are given in [1]. The distributions of track counting discriminators are shown in Figure 23, jet probability discriminators are shown in Figure 24 and simple secondary vertex discriminators are shown in Figure 25.

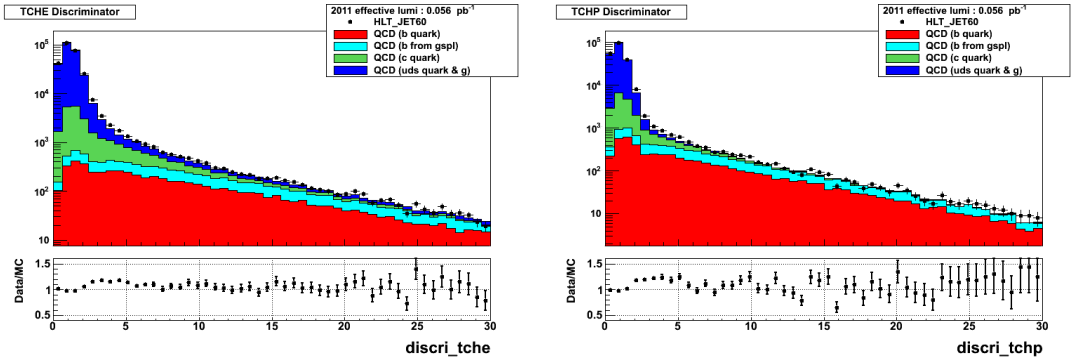

Figure 23: Left: track counting high efficiency, Right: track counting high purity discriminators.

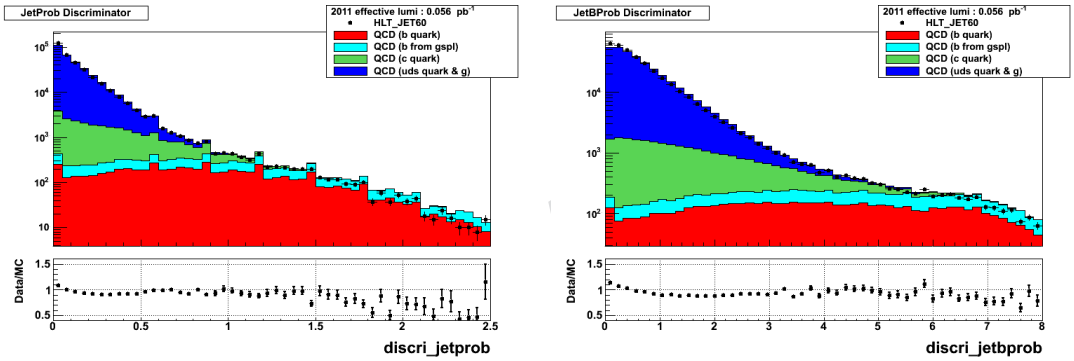

Figure 24: Left: jet probability , Right: jet B probability discriminators.

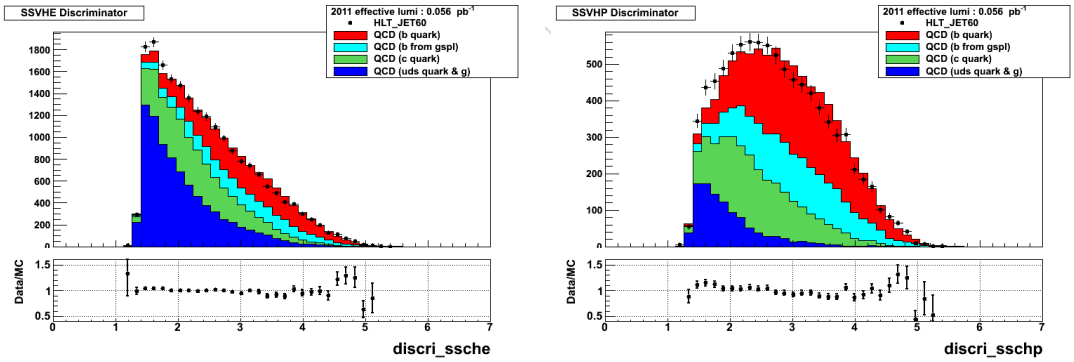

Figure 25: Left: simple secondary vertex high efficiency , Right: simple secondary vertex high purity discriminators. The underflow bin for jets which do not contain a reconstructed secondary vertex is not displayed.

The tagging rates can be calculated by integrating the discriminator distributions (Figures 23 to 25) from a given discriminator cut to infinity, divided by the total integral. The tagging rates are displayed in Figures 26 to 28.

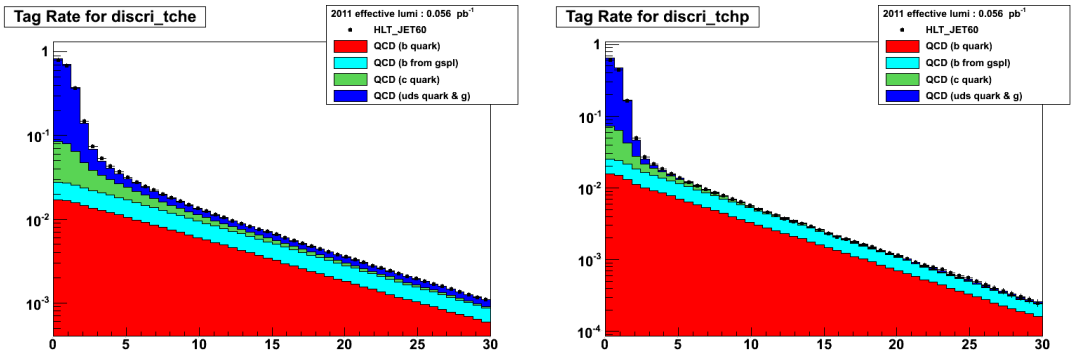

Figure 26: Left: track counting high efficiency tagging rate, Right: track counting high purity tagging rate.

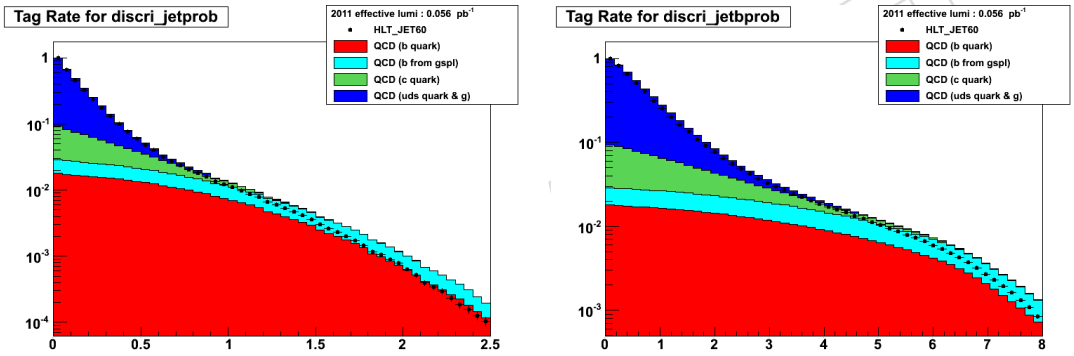

Figure 27: Left: jet probability tagging rate, Right: jet B probability tagging rate.

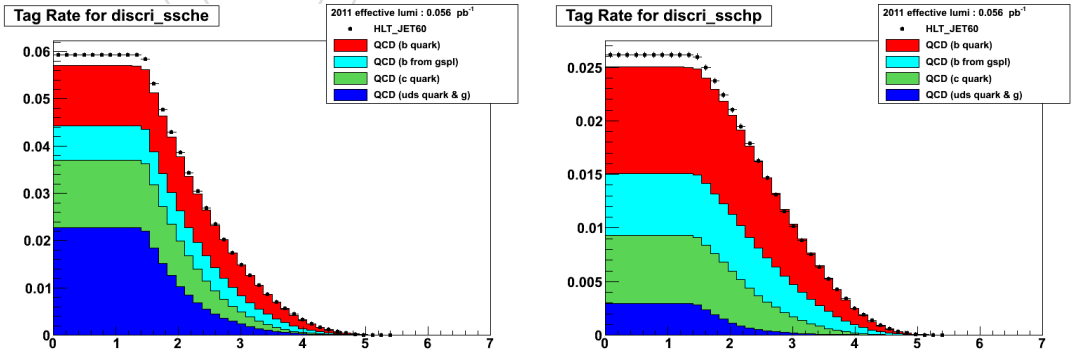

Figure 28: Left: simple secondary vertex high efficiency tagging rate, Right: simple secondary vertex high purity tagging rate.

8 Muons in jets

This section describes the special properties of jets containing muons. As mentioned in Section 3, a dedicated di-jet plus muon trigger has been used in data. The Monte Carlo samples are similar to the QCD samples used in the previous section except for the requirement of the presence of a muon at generator level.

Muons are seeded from the CMS muon chambers, and are then linked to tracks found in the tracking system to form global muons [9, 10].

The muon selection requirements are the following:

- transverse momentum $p_t > 7$ GeV
- number of valid inner hits > 10
- number of pixel hits > 1
- number of missed outer hits < 3
- number of matches > 0
- inner $\chi^2/ndof < 10$
- global $\chi^2/ndof < 10$
- longitudinal distance to primary vertex < 2 cm
- ΔR to jet axis < 0.4

Figure 29 shows the transverse momentum, the impact parameter significance and the reconstructed number of muons per jet.

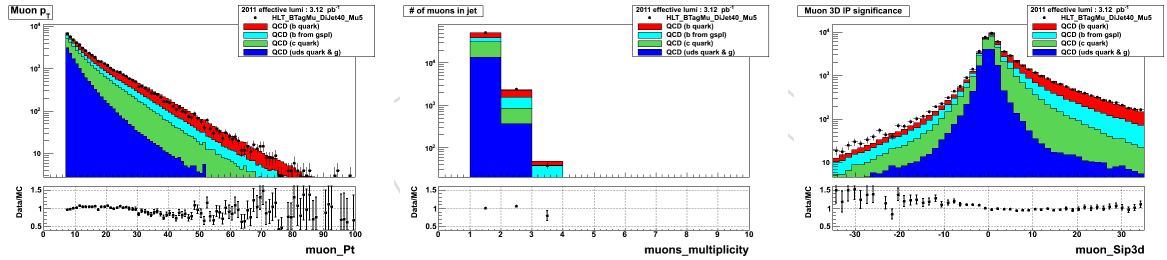

Figure 29: Left: transverse momentum p_t of muons in jets. Middle: number of reconstructed muons per jet. Right: 3D impact parameter significance of muons in jets.

An important quantity for performance measurements is the relative transverse momentum with respect to the jet, p_t^{rel}. This is displayed in Figure 30 (left). There is an apparent discrepancy in this distribution which is also visible in the angle between the muon and the jet axis in units of ΔR (Figure 30 right).

The angular discrepancies in muon-jets are also visible in the secondary vertex angles. Figure 31 shows the same distributions as Figure 22, but for jets with muons. The trend towards larger angles in data is enhanced in muon jets.

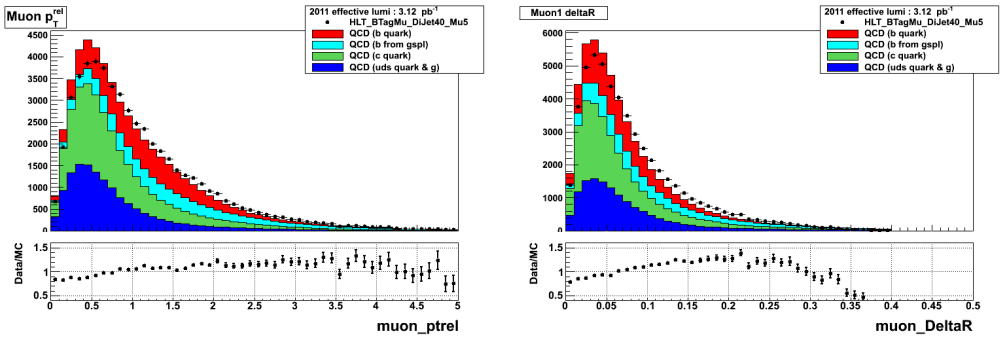

Figure 30: Left: transverse momentum of the muon with respect to the jet axis p_t^{rel}. Right: angle (in units of ΔR) between the muon and the jet axis.

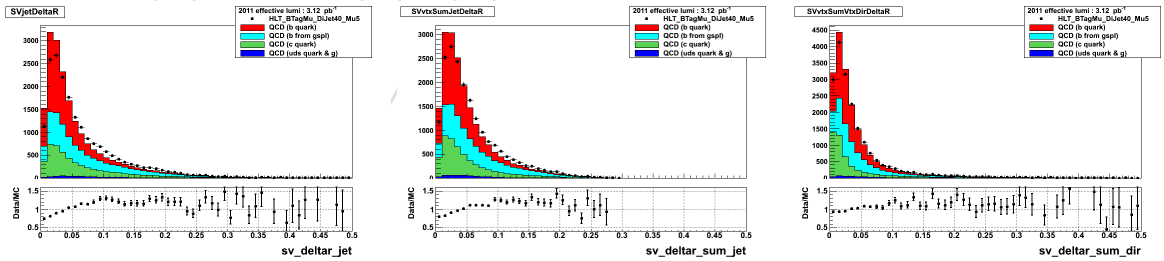

Figure 31: Left: angular distance in ΔR between jet axis and vertex direction. Middle: angular distance in ΔR between jet axis and the sum of track momenta at the vertex, Right: angular distance in ΔR between vertex direction and the sum of track momenta at the vertex.

9 Alignment

9.1 Alignment of the Inner Silicon Tracker

The determination of the alignment corrections for the Inner Silicon Tracker to be used for the reprocessing of the 2010 data was done with the global method [11, 12] using a mixture of tracks coming from atmospheric cosmic rays and minimum bias collisions. The global method aligns the highest level structures (half-barrels, endcaps) with all the six degrees of freedom together with all module units with the most sensitive degrees of freedom each: u, w and γ for the strip modules (also α and β in TIB) and u, v, w and γ for the pixel modules [1].

During the 2010 LHC run, the geometry of the pixels was monitored on a daily basis using the so-called unbiased track-to-primary vertex residuals. This method consists in selectively removing a track from the event, determine a primary vertex with the others, compute the transverse and longitudinal projections of impact parameter of the probe track with respect the primary vertex and finally studying the mean value of the distribution of the residuals as a function of ϕ, η of the track.

Despite starting from a pre-calibrated detector, a discontinuity in the distribution of the mean longitudinal impact parameter as a function of azimuthal angle was observed, with the height of the step changing few times during the year. This behavior was interpreted as relative movements, up to 90 μm in magnitude, along the z-direction of the two BPIX half-barrels, movements which are allowed since the two halves are mechanically independent.

To cope with this, in a common fit separate sets of alignment corrections for each of the three BPIX layers in each half-barrel and each of the four FPIX half-disks in each endcap were provided for seven different periods of the data-taking. The relative position of the modules with respect the supporting structure was instead assumed to be the same along all the 2010 data-taking. Figure 32 shows the distribution of longitudinal separation of the BPIX half-barrels as estimated from the unbiased track-to-primary vertex residuals as a function of time before and after the alignment procedure.

The statistical precision reached by the alignment was checked looking at the distribution of the median of the unbiased track-to-hit residuals computed for each module. The RMS of these distributions amount to 2 μm (4 μm) for the r-ϕ (z) measurement coordinate in the BPIX and to 6 μm (11 μm) for the r-ϕ (r) measurement in the FPIX, for all the data-taking intervals described above.

A set of alignment parameter errors, calibrated to provide pulls of the track-to-hit residuals with gaussian standard deviation close to unity, was provided together with the alignment constants.

Finally, prior to the restart of the 2011 LHC operations, the alignment corrections for the two BPIX half-barrels (6 dofs) and for the four FPIX half-disks (3 translational dofs) were recomputed using a sample of about 15k cosmic ray tracks. The analysis of the track-to-primary vertex residuals on events from 2011 collisions indicates that a potentially uncorrected separation of the two BPIX half-barrels is at most 10 μm.

[1] A local right-handed coordinate system is defined for each module, with u being the more precisley measured coordinate, v orhogonal to the u-axis and in the module plane, and the w-axis normal to the module plane. Angles α, β and γ are the rotations around the u, v and w-axes respectively.

9.2 Simulation-based study of the impact of misalignment on b-tagging performances

The performance of the alignment procedure described above is essentially not limited by the statistical precision. To properly describe in the simulation the alignment accuracy reached at the end of 2010, a misalignment scenario was prepared following the same approach described in [13] and using a sample of simulated events with approximately the same composition of the sample used for the alignment in the data. The same alignment parameter errors determined in the data complemented the misalignment scenario, hereafter referred to as MC2010.

The performances of the different b-tagging algorithms obtained with the MC2010 misalignment scenario were compared with respect those obtained with a perfectly aligned detector and with a misalignment scenario, prepared before the installation of the Inner Silicon Tracker in CMS, supposed at that epoch to describe the uncertainty on the alignment parameters expected after 10/pb of collected data [14]. The results, obtained from a sample of about 1.5 millions simulated $t\bar{t}$ events are shown in Figure 33. For all the taggers the performance with the MC2010 scenario are equal to those obtained with a perfectly aligned detector.

The same sample of simulated events was used to evaluate the deterioration of the performances of the b-tagging algorithms in case shifts along the z-direction of the two BPIX half-barrels, similar to those observed in 2010, were present but not corrected by the alignment procedure. For this study, the positions and orientations of all the other components of the Tracker were supposed to be perfectly known.

Three different scenarios were investigated corresponding to 40 μm, 80 μm and 160 μm absolute separation of the BPIX half-barrels. No change in the b-tagging performance is observed for the SSVHE and SSVHP taggers. For the other taggers (Figure 34), a significant decrease of the b-tagging efficiency is observed only for the 160 μm separation. For the track-counting taggers the decrease is more pronounced for TCHP, about -5.5% decrease, already setting-in at the medium contamination working point, while for the TCHE the decrease is visible only at the loose contamination working point (about -2.5% reduction). In case of the CSVB tagger the decrease is about -2.5% at 160 μm separation. The largest drop in b-tagging efficiency is observed for the JP and JBP taggers, about -7% at all the tight, medium and loose working points for the 160 μm separation scenario.

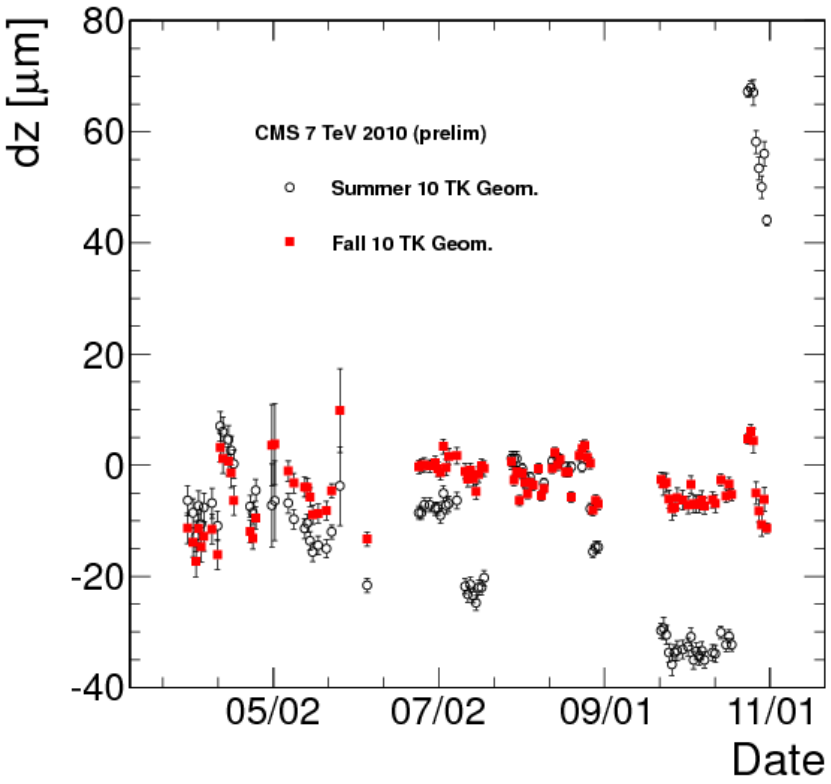

Figure 32: Longitudinal separation of the BPIX half-barrels as estimated from the unbiased track-to-primary vertex residual method as a function of time for the 2010 LHC proton-proton run. Empty (filled) dots are the pre-(post-)alignment values.

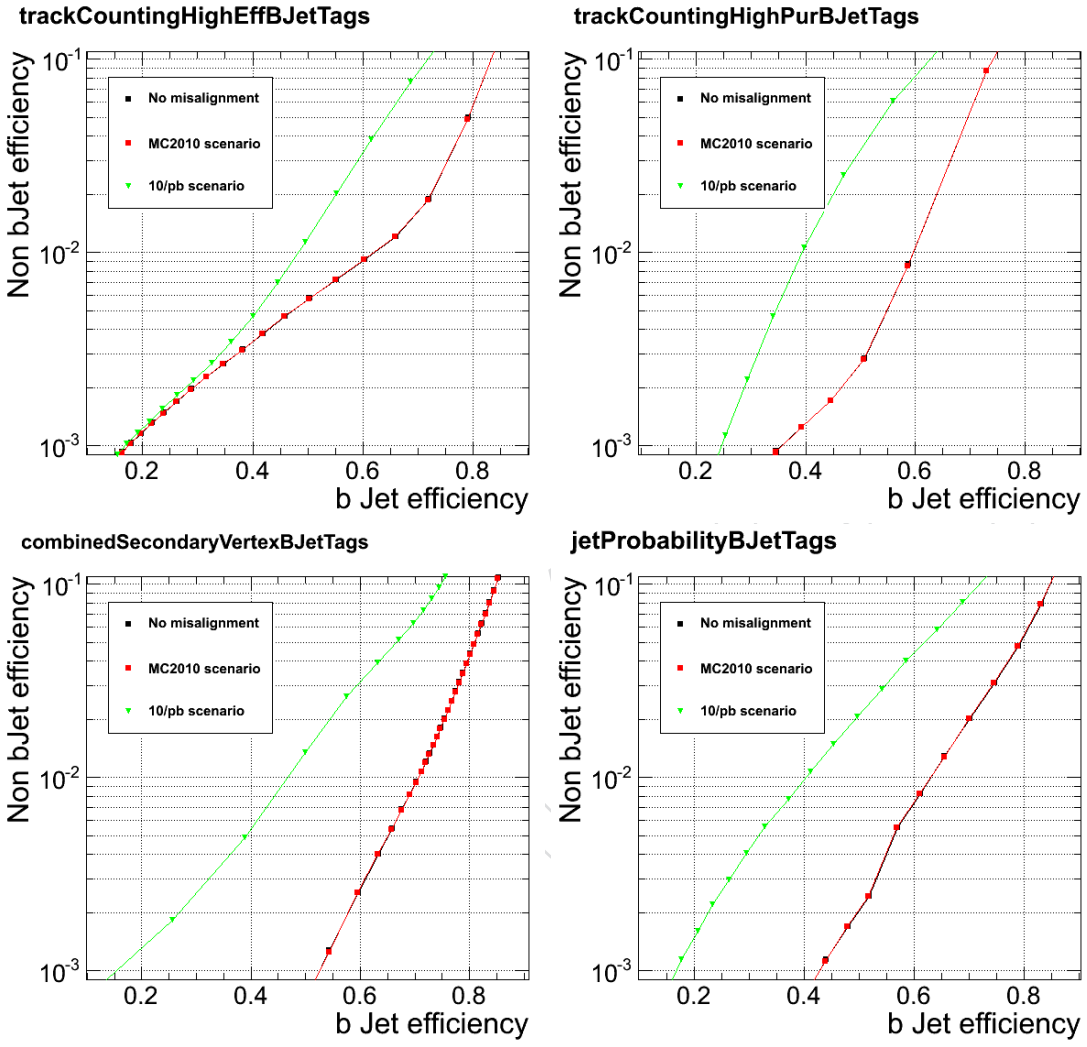

Figure 33: Contamination vs. efficiency for the TCHE, TCHP, CSV and JP b-tagging algorithms for a scenario describing the estimated current accuracy in alignment compared to a perfectly aligned detector. For reference the performance expected after 10/pb of collected lumi, based on a previuos study, are also shown.

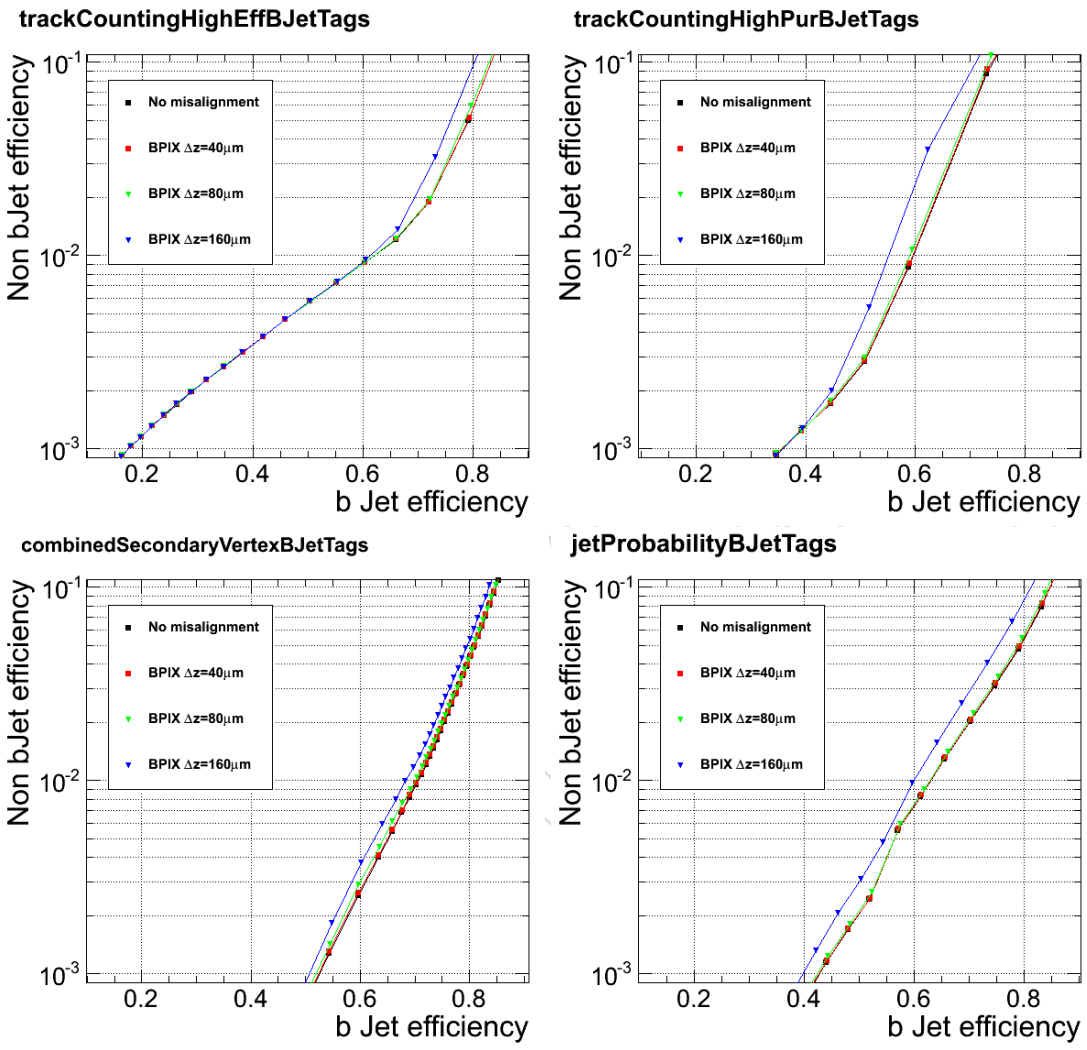

Figure 34: Contamination vs. efficiency for the TCHE, TCHP, CSV and JP b-tagging algorithms for scenarios with an artificial separation of the two BPIX half-barrels of 40, 80, 160 μm.

10 Pileup

Due to the high LHC luminosity in the 2011, several proton collisions are taking place simultaneously in one bunch crossing. The number of reconstructed primary vertices per event is shown in Figure 35. The constant change of the luminosity conditions makes a correct simulation in the Monte Carlo samples difficult. We therefore use Figure 35 to obtain reweighting factors to equalize the vertex multiplicity in Monte Carlo. The reweighting procedure has been applied for the results shown in Section 5 to 8.

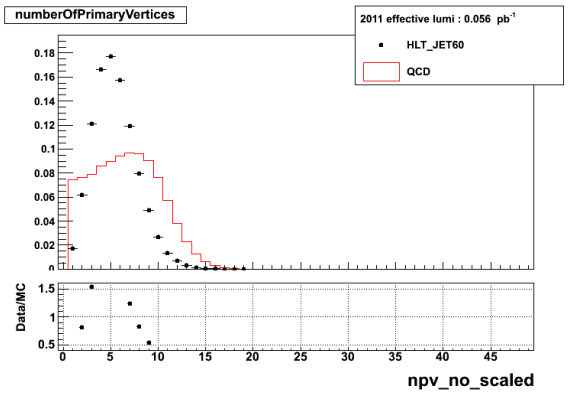

Figure 35: Number of reconstructed primary vertices per event.

The signal primary vertex is chosen to be the one with the largest sum of track p_t^2 which is currently the standard CMS convention. Tracks from pileup vertices can be wrongly associated to a jet from the signal vertex. This effect is visible in Figure 36 (left) which shows the number of tracks associated to a jet for three different primary vertex multiplicities. The right plot in Figure 36 shows the number of selected tracks. The track selection is clearly rejecting those additional tracks from nearby primary vertices. The most efficient observable to reject pileup tracks is the distance of the track to the jet axis (compare Section 4). In the ideal case, tracks from B decays are tangential to the jet axis and therefore have zero distance to the jet axis. Tracks from pileup events are well separated from the jet axis in longitudinal direction.

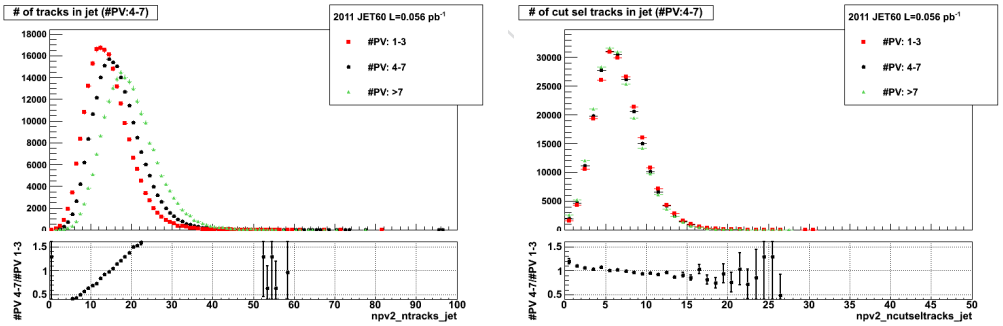

Figure 36: Left: number of tracks associated to a jet without any selection cuts. Right: number of tracks associated to a jet passing the selection cuts.

Some additional validation plots, for three different vertex multiplicities are shown in Figure 37. The most important observables, such as impact parameter and vertex flight distance seem to be in good agreement among the different pileup multiplicities. The number of reconstructed secondary vertices per jet for the three pileup cases is shown in Figure 38.

However, Figure 36 (right) indicates a slightly reduced track selection efficiency for high pileup multiplicities, because the average number of tracks decreases with more pileup. This results

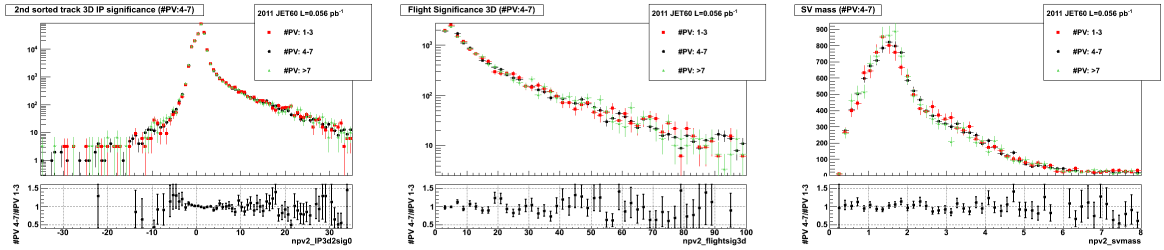

Figure 37: Left: impact parameter significance of the second track, ordered by IP significance. Middle: secondary vertex flight distance significance. Right: secondary vertex mass. All distributions are split into three different primary vertex multiplicities.

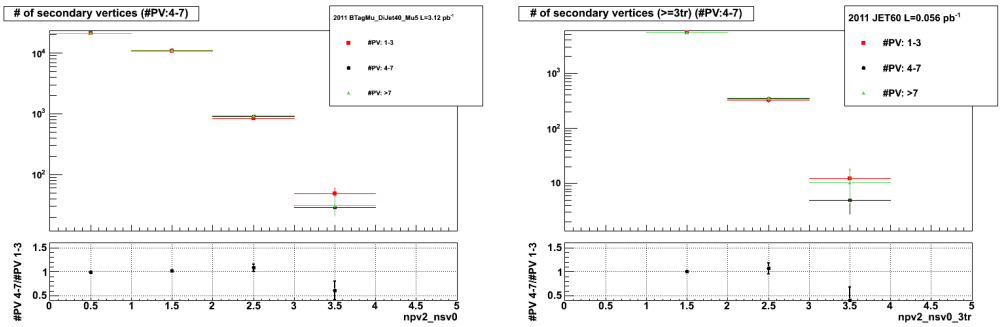

Figure 38: Left (right): number of reconstructed secondary vertices with three (two) tracks per jet. The distributions are split into three different primary vertex multiplicities.

in a reduced b-tagging efficiency in pileup events. A Monte Carlo based study has been done which shows a clear degradation of the b-tagging performance for events with pileup. This is illustrated in Figure 39 which shows the light flavor mistag efficiency versus b-tagging efficiency for several b-tagging algorithms.

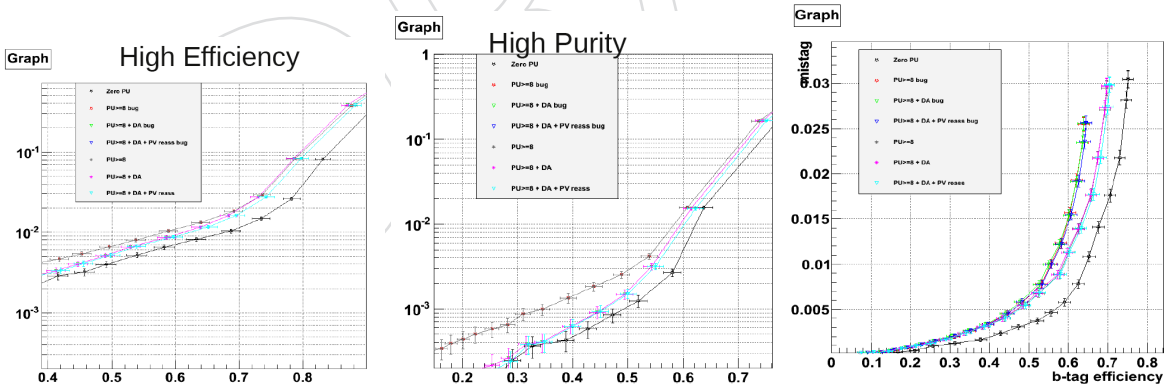

Figure 39: Light flavor mistag efficiency versus b-tagging efficiency for different piluep scenarios. (**FIXME: produce plots properly**). Left: track counting high efficiency algorithm. Middle: track counting high purity algorithm. Right: simple secondary vertex high efficiency algorithm.

(**FIXME:**) discussion on PV-jet association

11 Development of new algorithms

11.1 Improved jet-track association

11.2 Combined secondary vertex algorithm

The discrimination power of all secondary vertex b-taggers described in [1] can be further improved by combining them in a multivariate analysis (MVA) [15]. A new implementation of this technique is in development, using a general MVA framework in CMSSW.

A simple classifier was chosen as a first step, to make sure that the workflow is well-understood. Input variables are combined using a Naive Bayes likelihood ratio (which assumes that they are uncorrelated), and are normalized and rotated to remove linear correlations.

The discriminator is separately optimized in three broad classes: RecoVertex (a secondary vertex is fully reconstructed), PseudoVertex (no secondary vertex is reconstructed, but at least two tracks are inconsistent with the primary vertex with a transverse impact parameter significance greater than 2.0), and NoVertex (neither of the above). These classes determine the set of input variables available for optimization.

The work on this project is ongoing. The MVA framework is modular enough to replace classification engines with minimal impact on the workflow; thus, the project has a natural upgrade path to more sophisticated MVA techniques.

References

[1] CMS Collaboration, "Algorithms for b Jet identification in CMS", *CMS PAS* **BTV-09-001** (2009).

[2] CMS Collaboration, "Commissioning of b-jet identification with pp collisions at $\sqrt{s}$ = 7 TeV", *CMS PAS* **BTV-10-001** (2010).

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

[4] G. S. Matteo Cacciari, Gavin P. Salam, "The anti-k_t jet clustering algorithm", *JHEP* **0804** (2008) 063. arXiv:0802.1189 [hep-ph].

[5] CMS Collaboration, "CMS Physics Technical Design Report, Volume I: Detector performance and software", *CMS PTDR* **1** (2008).

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

[7] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", *CMS PAS* **TRK-10-005** (2010).

[8] W. Waltenberger, "Adaptive Vertex Reconstruction", *CMS Note* **2008/33** (2008).

[9] **CMS** Collaboration, "The CMS experiment at the CERN LHC", *JINST* **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[10] **CMS** Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV", *CMS PAS* **MUO-10-002** (2010).

[11] V. Blobel, "Software alignment for tracking detectors", *Nucl. Instrum. Meth.* **A566** (2006) 5–13. doi:10.1016/j.nima.2006.05.157.

[12] G. Flucke, P. Schleper, G. Steinbruck et al., "CMS silicon tracker alignment strategy with the Millepede II algorithm", *JINST* **3** (2008) P09002. doi:10.1088/1748-0221/3/09/P09002.

[13] **CMS** Collaboration, CMS Collaboration, "Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays", *JINST* **5** (2010) T03009. doi:10.1088/1748-0221/5/03/T03009.

[14] T. Lampén, N. De Filippis, F. Schilling et al., "Comprehensive Set of Misalignment Scenarios for the CMS Tracker", *CMS Note* **2008/29** (2008).

[15] C. Weiser, "A combined secondary vertex based B-tagging algorithm in CMS",. CERN-CMS-NOTE-2006-014.

A Appendix

Figures 40 to 56 show a selection of the plots from Sections 5 to 8 but with a lower trigger threshold (`HLT_Jet30` in data and `HLTJet15U` in Monte Carlo) and lower jet momentum cut of $p_t > 30$ GeV in case of the single jet triggers. For the b-jet triggers another set of plots has been produced with `HLT_BTagMu_DiJet20_Mu5`, requiring two jets with $p_t > 40$ GeV, one of them with $p_t > 45$ GeV.

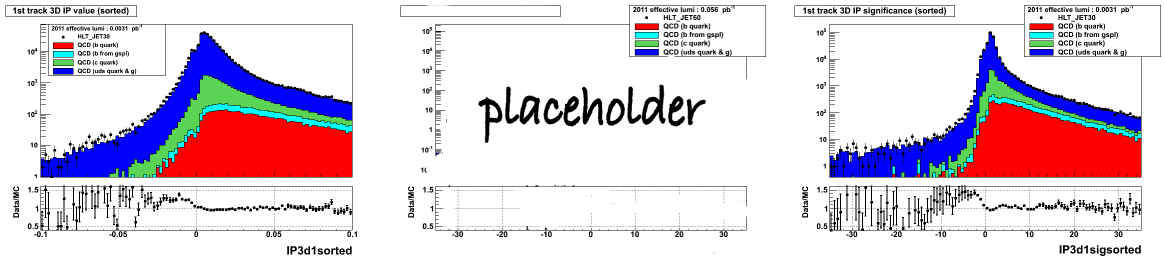

Figure 40: Left: IP value, middle: IP error, right: IP significance for the first track in the jet, ordered by IP significance.

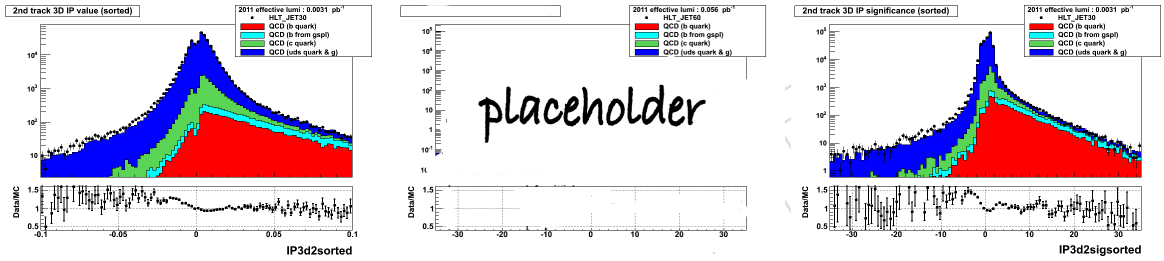

Figure 41: Left: IP value, middle: IP error, right: IP significance for the second track in the jet, ordered by IP significance.

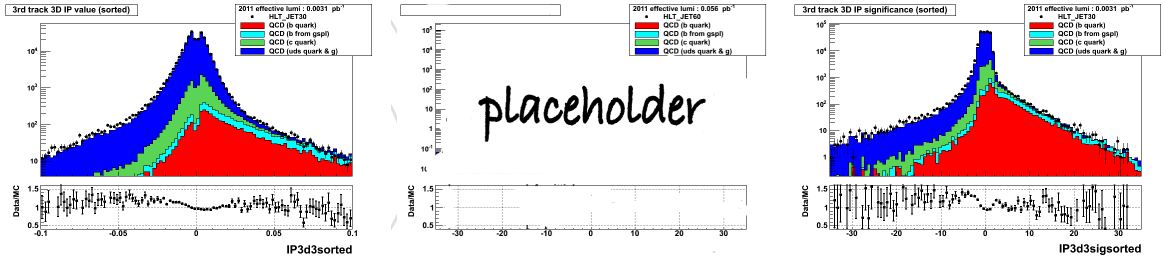

Figure 42: Left: IP value, middle: IP error, right: IP significance for the third track in the jet, ordered by IP significance.

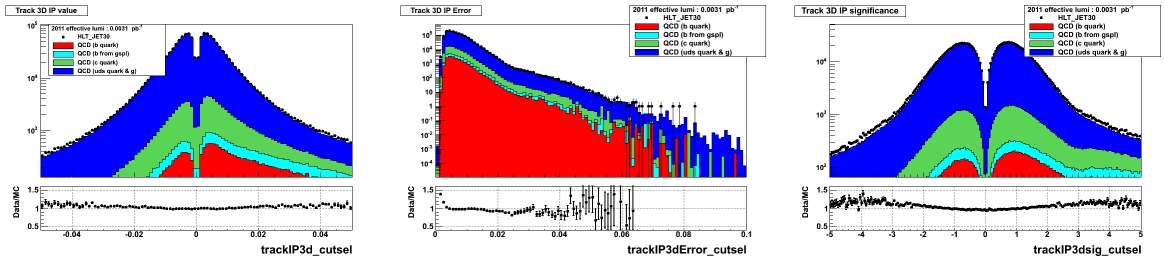

Figure 43: Left: IP value, middle: IP error, right: IP significance for all selected tracks in the jet. The track selection as defined in Section 4 has been applied.

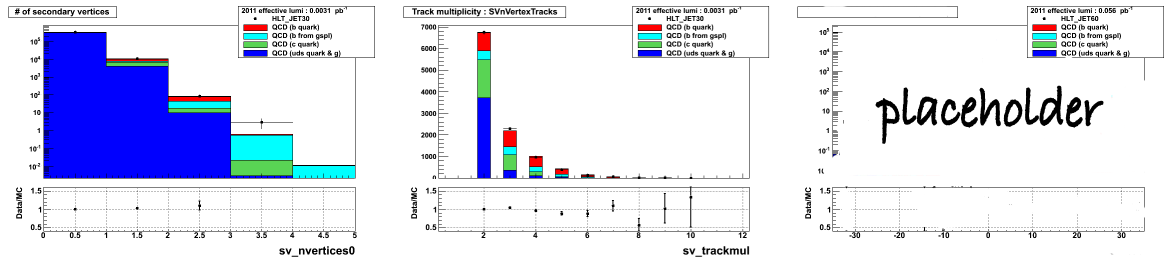

Figure 44: Left: number of reconstructed secondary vertices per jet, middle: number of tracks at the reconstructed secondary vertex, right: average number of tracks at the secondary vertex versus jet p_t.

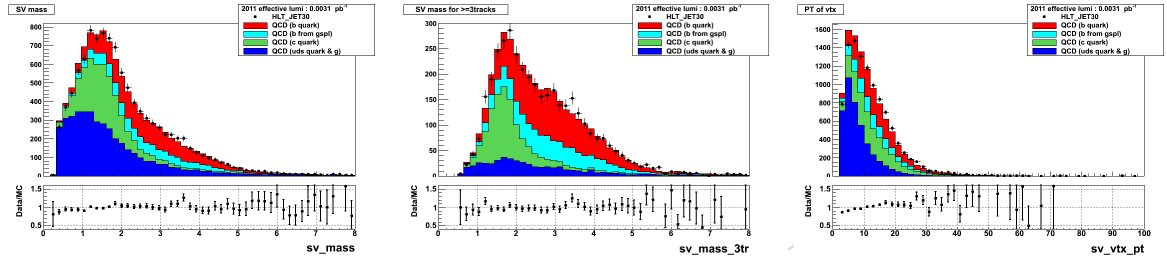

Figure 45: Left: vertex mass with two or more reconstructed tracks at the vertex. Middle: vertex mass with three or more tracks at the vertex. Right: transverse momentum of the secondary vertex (with two or more tracks).

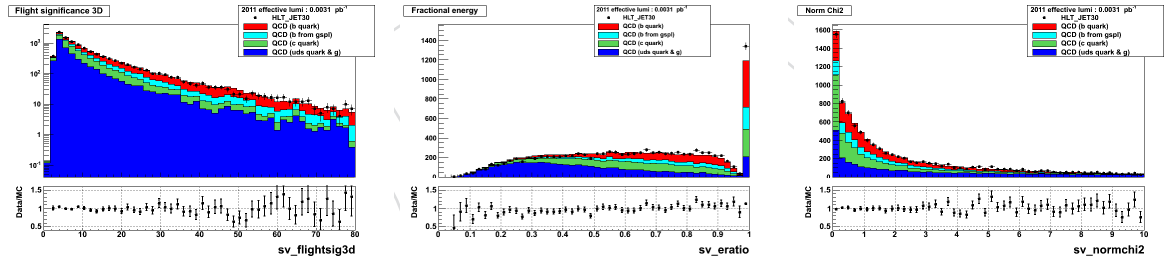

Figure 46: Left: vertex flight distance significance. Middle: ratio of track energy at the secondary vertex with respect to all selected tracks in the jet. Right: vertex fit normalized χ^2.

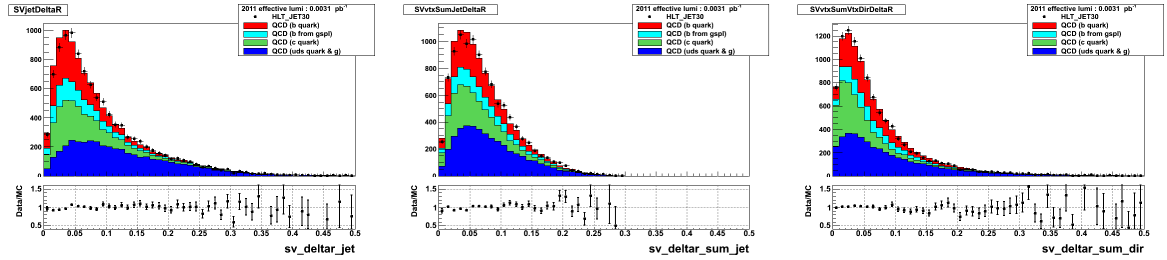

Figure 47: Left: angular distance in ΔR between jet axis and vertex direction. Middle: angular distance in ΔR between jet axis and the sum of track momenta at the vertex, Right: angular distance in ΔR between vertex direction and the sum of track momenta at the vertex.

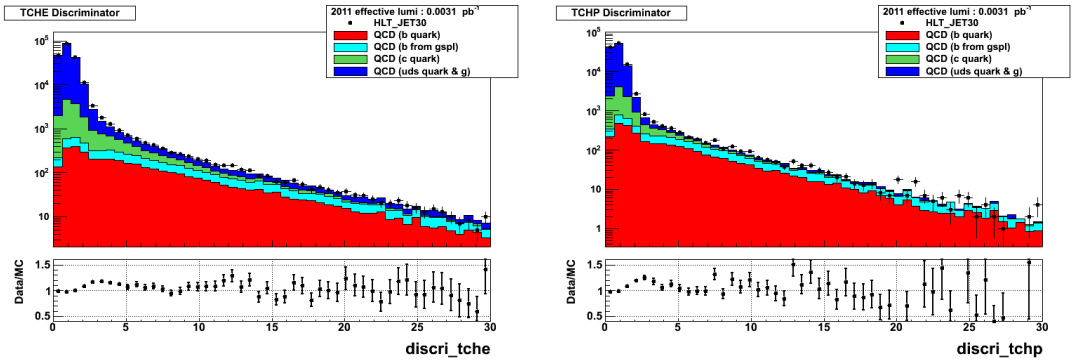

Figure 48: Left: track counting high efficiency, Right: track counting high purity discriminators.

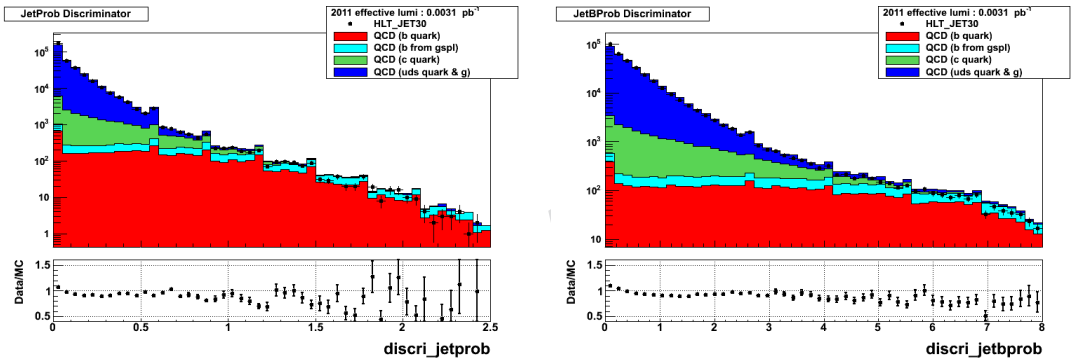

Figure 49: Left: jet probability , Right: jet B probability discriminators.

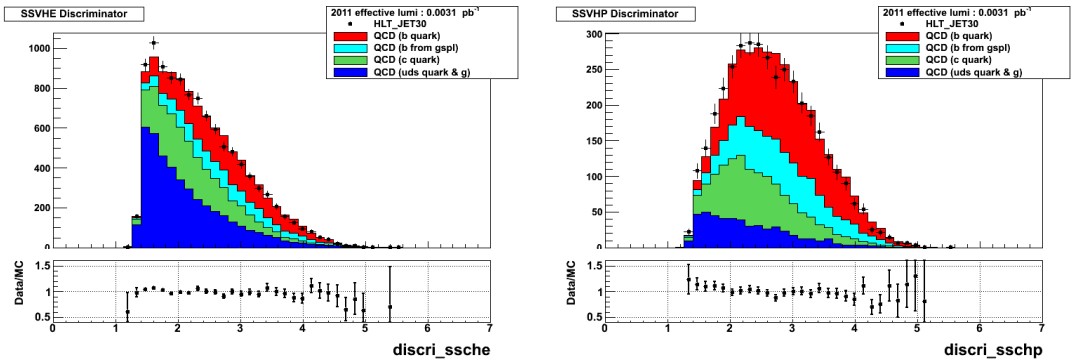

Figure 50: Left: simple secondary vertex high efficiency , Right: simple secondary vertex high purity discriminators. The underflow bin for jets which do not contain a reconstructed secondary vertex is not displayed.

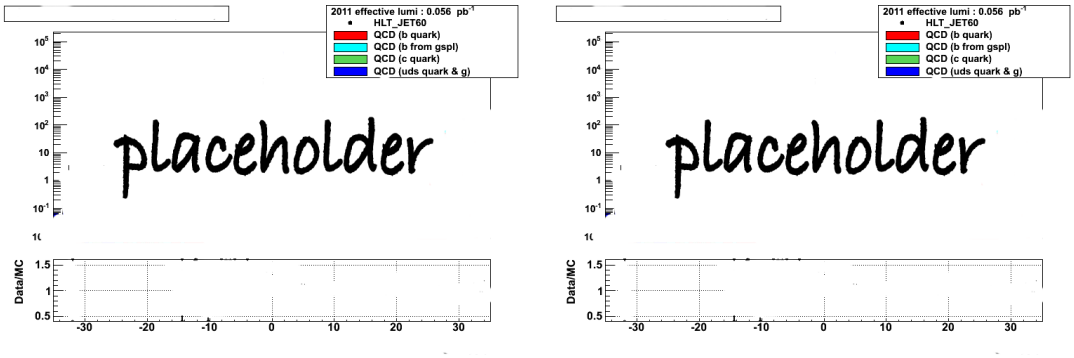

Figure 51: Left: track counting high efficiency tagging rate, Right: track counting high purity tagging rate.

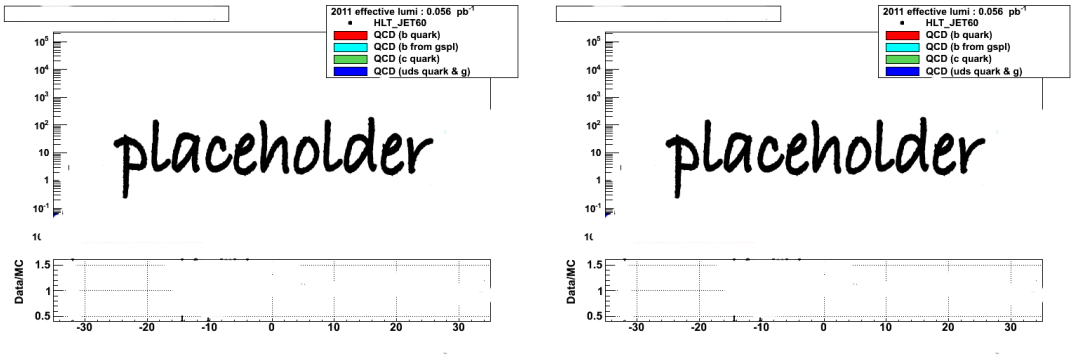

Figure 52: Left: jet probability tagging rate, Right: jet B probability tagging rate.

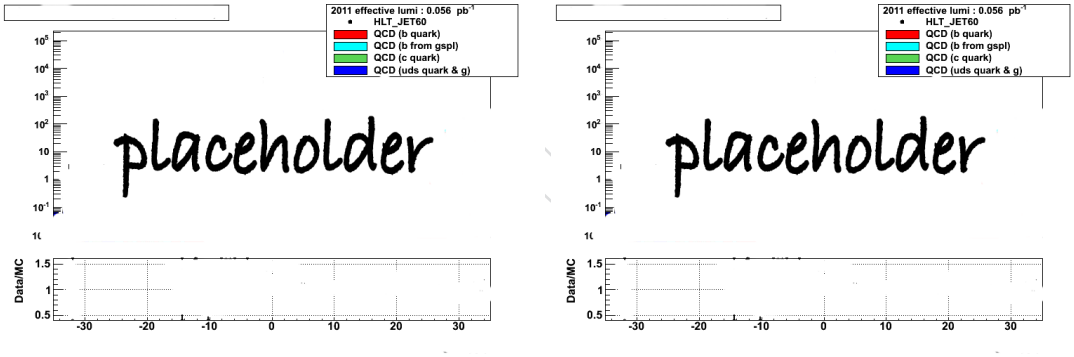

Figure 53: Left: simple secondary vertex high efficiency tagging rate, Right: simple secondary vertex high purity tagging rate.

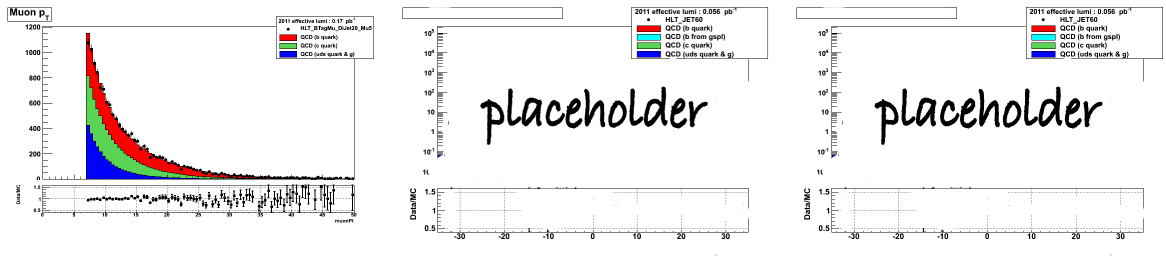

Figure 54: Left: transverse momentum p_t of muons in jets. Middle: number of reconstructed muons per jet. Right: 3D impact parameter significance of muons in jets.

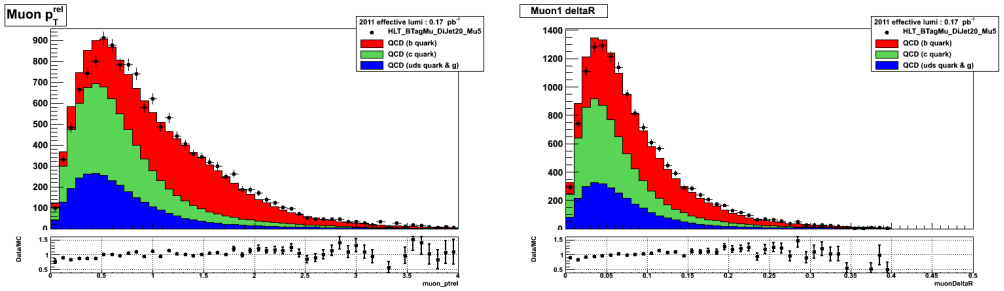

Figure 55: Left: transverse momentum of the muon with respect to the jet axis p_t^{rel}. Right: angle (in units of ΔR) between the muon and the jet axis.

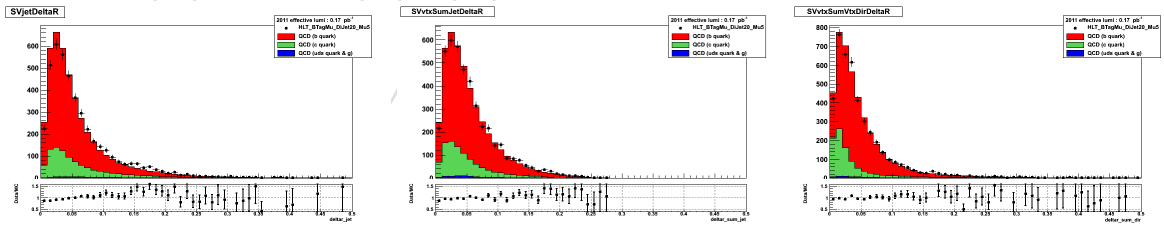

Figure 56: Left: angular distance in ΔR between jet axis and vertex direction. Middle: angular distance in ΔR between jet axis and the sum of track momenta at the vertex, Right: angular distance in ΔR between vertex direction and the sum of track momenta at the vertex.