ATLAS NOTE

ATLAS-CONF-2011-102

July 20, 2011

Commissioning of the ATLAS high-performance *b*-tagging algorithms in the 7 TeV collision data

The ATLAS collaboration

Abstract

The ability to identify jets containing *b*-hadrons is important for the high-p_{T} physics program of a general-purpose experiment at the LHC such as ATLAS. Building on the successful commissioning of two robust algorithms used already for several analyses of the 2010 data, the commissioning of the high-performance *b*-tagging algorithms with the 2011 data is described, as well as the expected improvement in performance.

1 Introduction

The ability to identify jets containing b-hadrons is important for the high-p_{T} physics program of a general-purpose experiment at the LHC such as ATLAS [1, 2]. Two robust b-tagging algorithms taking advantage of the impact parameters of tracks (JetProb) or reconstructing secondary vertices (SV0) have been swiftly commissioned [3, 4] and used for several analyses of the 2010 and 2011 data [5–12]. Building on this success, more advanced b-tagging algorithms have been commissioned with the 2011 data. All these algorithms are based on Monte Carlo predictions for the signal (b-jet) or background (light- or in some cases c-jet) hypotheses. Assessing to which extent the simulation accurately reproduces the experimental data in the variables used for the high-performance tagging algorithms is therefore important for the performance of the algorithms and for their successful calibration for use in physics analyses. In this note, we make detailed comparisons between data and simulation of the tagger input variables and the resulting tagging rates in inclusive dijet event samples. We also present both Monte Carlo and data-driven studies that show the expected improvement in performance brought by the high-performance tagging algorithms.

2 Data samples

2.1 Experimental data sample

The analysis presented here is based on a sample of proton-proton collisions at $\sqrt{s} = 7$ TeV delivered by the LHC and collected by the ATLAS experiment between March 22nd and May 25th 2011. Only data collected during stable beam periods in which all sub-detectors were fully operational are used. The sample amounts to approximately 93M events and corresponds to about 330 pb^{-1} of integrated luminosity. Throughout the considered period, the LHC beam parameters were relatively stable. For instance, the average number of minimum-bias pile-up events, originating from the collisions of additional protons in the same bunch as the signal collision, was maintained at around six.

2.2 Simulated data samples

Experimental data are compared to a Monte Carlo simulation of QCD jet events. Simulated events were generated with the Pythia [13] 6.423 program using the MRST LO* [14] PDFs. This generator utilizes leading-order perturbative QCD matrix elements for $2 \to 2$ processes, along with a leading-logarithmic parton shower, an underlying event simulation with multiple parton interactions, and the Lund string model for hadronisation. Samples were generated using the ATLAS Minimum Bias Tune 1 (AMBT1) set of parameters [15], in which the non-diffractive model has been tuned to ATLAS measurements of charged particle production at $\sqrt{s} = 900$ GeV and $\sqrt{s} = 7$ TeV. In order to have sufficient statistics over the entire p_{T} spectrum, eight samples were generated for different exclusive ranges of the hard-scattering partonic transverse momentum $\hat{p_{\mathrm{T}}}$. The lowest-p_{T} sample consists of about 16M events, while the highest p_{T} ones have about 2M events each. The eight samples were mixed taking into account their respective production cross sections.

In addition, the estimation of the expected performance of the high-performance tagging algorithms described in Section 7.5 is based on a sample of simulated $t\bar{t}$ events in which both W bosons either decay leptonically or one decays leptonically and the other hadronically. The events were generated with the MC@NLO generator [16] v3.41 with PDF set CTEQ66 [17], assuming a top mass of 172.5 GeV.

On top of these signal events, simulated minimum-bias events were superimposed to simulate pile-up, obtained from Pythia non-diffractive, single and double diffractive events. The bunch train configuration of the LHC during this period was emulated.

The particle four-vectors from these generators were passed through a full simulation [18] of the ATLAS detector and trigger that is based on GEANT4 [19]. The simulated geometry corresponds to a perfectly aligned detector and most of the disabled detector elements for data-taking, notably the pixel modules, were masked in the simulation. Small run-to-run variations of the number of disabled channels are present in data but not corrected for in simulation. For instance, on the innermost pixel layer, six modules were permanently masked in simulation, while in data the number of disabled modules fluctuates around nine.

3 Reconstruction

Both the experimental data and simulated events were reconstructed using the same version of the ATLAS software. The reconstruction of the key objects for *b*-tagging purposes, namely the tracks, the primary vertex and the jets, is briefly described in the following.

The *b*-tagging performance relies critically on the accurate reconstruction of the charged tracks in the ATLAS Inner Detector. The innermost part, the pixel detector, has an intrinsic measurement accuracy of around 10 μm in the transverse plane ($r\phi$) and 115 μm along the beam axis (z) and its innermost layer of detection is located at a radius of 5 cm from the beam axis. This layer contributes the most to the accuracy of the measurement of the locations of tracks and displaced vertices, which are the main ingredients of *b*-tagging. The ATLAS tracking system allows the tracks to be measured efficiently and with good accuracy within a pseudorapidity range of $|\eta| < 2.5$ and down to $p_{\mathrm{T}} = 400$ MeV. For a central track with $p_{\mathrm{T}} = 5$ GeV, which is typical for *b*-tagging, the transverse momentum resolution is around 75 MeV and the transverse impact parameter resolution is about 35 μm.

The knowledge of the position of the primary interaction point (primary vertex) of the proton-proton collision is important for *b*-tagging since it defines the reference point with respect to which impact parameters and vertex displacements are measured. The reconstruction of primary vertices [20] relies on the reconstructed tracks and consists of two stages: first, the vertex finding, which associates reconstructed tracks to the vertex candidates, and second, the vertex fitting, which reconstructs the vertex position (and its error matrix). To ensure a good resolution on the vertex position, the primary vertex must be reconstructed from at least five tracks. For primary vertices with 70 tracks or $\sqrt{\sum_{\mathrm{tracks}} p_{\mathrm{T}}^2} \approx 12$ GeV, which is quite typical for this study, the resolution on the primary vertex position is about 23 μm in the transverse plane and 40 μm in the longitudinal direction. One important issue for *b*-tagging is the choice of the primary vertex, which is less trivial in the presence of minimum-bias events from pile-up: the primary vertex from a pile-up event may be mistakenly used as the signal vertex, or a fake primary vertex built from tracks from two different vertices may be reconstructed. The current strategy is to choose the primary vertex candidate that maximizes $\sum_{\mathrm{tracks}} p_{\mathrm{T}}^2$.

The jet direction is used to associate tracks to jets and subsequently to define the sign of the impact parameters of the tracks. Jet candidates are reconstructed by using the anti-k_t jet clustering algorithm [21] with a distance parameter $R = 0.4$. The inputs to this algorithm are three-dimensional topological calorimeter energy clusters [22]. The jet energies are corrected for inhomogeneities and for the non-compensating nature of the calorimeter by using p_{T}- and η-dependent calibration factors determined from Monte Carlo simulation [23]. The jet direction is corrected to take into account the position of the primary vertex along the z axis.

4 Trigger, event and jet selections

The ATLAS trigger system is based on a 3-level chain system. All the events analyzed were triggered by a jet trigger chain. At the Level 1 and Level 2, cluster-based jet triggers are used to select events. At

the last stage, the so-called Event Filter performs a full scan, reading out the complete calorimeter and running the offline anti-k_t jet finding algorithm with $R = 0.4$ on topological clusters. For this analysis, the events are required to come from single-jet triggers with different thresholds at the Event Filter, ranging from 10 GeV to 240 GeV. Most of these triggers, especially those with low-p_{T} thresholds, were increasingly heavily prescaled with the rapid increase in instantaneous luminosity over time. The triggers with the lowest p_{T} thresholds were prescaled by up to five orders of magnitude, and typically the same jet trigger is prescaled ten times more in the later data taking periods compared to the earlier ones. A logical OR of these triggers is used for selecting the events.

The offline event selection for this analysis consists of requiring the selected primary vertex candidate to have at least 5 tracks.

Different samples of jets are derived from the experimental data and from the simulated samples. The most commonly used sample is formed by the collection of the leading jet in all the events, provided they fulfill the requirements to potentially be tagged which are: $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. This sample is referred to as the inclusive leading jet sample and consists of about 10M jets in data. The average p_{T} of those jets is 145 GeV. In addition, another sample is defined in which the fraction of heavy-flavor jets is increased. The sample consists of 100k jets in data with a mean p_{T} of 180 GeV. To obtain this latter sample, dijet events were selected as follows: the leading jet (triggering jet) should fulfill $p_{\mathrm{T}1} > 40$ GeV and $|\eta_1| < 2.5$; the sub-leading jet is required to have $|\eta_2| < 2.5$; the two jets are required to be back-to-back and well-balanced in p_{T}: $\Delta\phi_{12} > 2.6$ rad, $|p_{\mathrm{T}1} - p_{\mathrm{T}2}| < 0.2(p_{\mathrm{T}1} + p_{\mathrm{T}2})$. In this sample, the probed jets consist of the collection of the sub-leading jet in the event, while the leading jet is tagged using the already commissioned early tagging algorithms. The heavy-flavor enrichment is obtained by requiring that the leading jet be identified as a b-jet by either the JetProb or the SV0 algorithms with high-purity cuts: a three-dimensional decay length significance L_{3D}/σ_L (cf. Section 7.2) greater than 8 for SV0 or, if no secondary vertex is found, a probability that the jet is compatible with the primary vertex lower than 10^{-6}, as computed by JetProb.

5 Simulation reweighting and corrections

In order to meaningfully compare data and simulation in the variables important for the high-performance taggers it must be ensured that the overall event kinematics are as well modelled in the simulation as possible.

The Monte Carlo samples used in this note were prepared to cover data-taking periods with higher bunch intensities than the ones under study: the luminosity-weighted average of the number of interactions per bunch-crossing, after trigger selections, is 10.4 in simulation and 5.7 in experimental data. The simulation events are therefore reweighted by using the distribution of the number of reconstructed primary vertices. This is a good estimator of the contribution of in-time pile-up events, namely additional pp collisions in the same bunch crossing as the signal event. Out-of-time pile-up events (pp collisions from neighboring bunches in the same train) also generate calorimeter activity and consequently extra jets. However, given the time resolution of the Inner Detector, and since the b-tagging algorithms reject jets with no track associated to them, the contribution of the out-of-time pile-up for this analysis is expected to be negligible.

Another important point is that the combination of the various jet triggers used and the fact that they were prescaled, which is not modelled in the simulation, results in rather different p_{T} and η spectra in data and simulation. To account for this, and also to correct for residual discrepancies between the experimental data and the modeling of the dijet event kinematics by the particle generator, a two-dimensional reweighting in (p_{T}, η) of the jets in simulation is performed. Since the most energetic jet in the event is more likely to fire the jet trigger, jets are classified in three categories according to their energy rank in the event: leading jet (1st rank), sub-leading jet (2nd rank) or softer jets (3rd rank and lower ranks).

For each jet rank, a 2D reweighting map is obtained by dividing the (p_{T}, η) distributions in data and in simulation. This procedure is applied separately for the inclusive jet sample and for the sample enriched in heavy-flavor jets. The same reweighting factor is applied to all jet flavors when reweighting the Monte Carlo, *i.e.* there is no attempt to model any possible flavor dependence. Note that the $\mathcal{O}(100)$ jets with a p_{T} above 800 GeV were ignored in the procedure and will not appear in the results discussed in this note.

Except where explicitly stated otherwise, all the distributions shown in the following have been obtained applying the reweightings and corrections described above.

6 Track selection and properties

In this section, some properties of the reconstructed charged tracks are examined, since tracks are a crucial ingredient of *b*-tagging. The tracks are required to fulfill the *b*-tagging track quality requirements and to be associated with a reconstructed jet, as explained in the following Sections 6.1 and 6.2. Prior to this, several relevant properties of the event were checked (primary vertex, environment of jets and effects of close-by jets, etc) and in all cases the simulation was found to adequately describe the experimental data after reweighting.

6.1 Track quality cuts

The track selection for *b*-tagging is designed to select well-measured tracks, rejecting fake tracks and tracks from long-lived particles (K_s, Λ and other hyperon decays, generically referred to as V^0 decays in the following) and material interactions (photon conversions or hadronic interactions).

The *b*-tagging quality selection requires at least seven precision hits (pixel or micro-strip hits) on the track. In addition, at least two hits in the pixel detector are required, of which one must be in the innermost pixel layer. Only tracks with $p_{\mathrm{T}} > 1$ GeV are considered. The transverse and longitudinal impact parameters (cf. Section 6.3) defined with respect to the primary vertex must fulfill $|d_0| < 1$ mm and $|z_0| \sin\theta < 1.5$ mm. This selection is used by all the tagging algorithms relying on the impact parameters of tracks.

6.2 Track association to jets

The actual tagging is performed on the sub-set of tracks in the event that are associated to jets. Tracks are associated to the jets with a spatial matching in $\Delta R_{(\text{jet, track})} \equiv \sqrt{\left(\eta_{\text{jet}} - \eta_{\text{track}}\right)^2 + \left(\phi_{\text{jet}} - \phi_{\text{track}}\right)^2}$. The association cut ΔR is varied as a function of the jet p_{T} in order to have a smaller cone for high-p_{T} jets which are more collimated. This reduces the number of spurious tracks from fragmentation, underlying event or pile-up events which would either dilute the discrimination (prompt tracks) or produce fake lifetime signatures. At 20 GeV, the ΔR cut is 0.45, while for more energetic jets with a p_{T} around 150 GeV the ΔR cut is 0.25. A given track can be associated to only one jet, the closest in ΔR. In both experimental data and simulation, the average number of tracks associated to a jet with $p_{\mathrm{T}} = 50$ GeV (resp. 200 GeV) and fulfilling the *b*-tagging quality criteria is 3.5 (resp. 7).

6.3 Impact parameters

The most critical track parameters for *b*-tagging are the transverse and longitudinal impact parameters. The transverse impact parameter d_0 is the distance of closest approach of the track to the primary vertex point in the $r\phi$ projection. The z coordinate of the track at this point of closest approach is referred to

as z_0. It is often called the longitudinal impact parameter[1]. On the basis that the decay point of the b-hadron must lie along its flight path, the impact parameter is signed to further discriminate the tracks from b-hadron decays from tracks originating from the primary vertex. The sign is positive if the track extrapolation crosses the jet direction in front of the primary vertex, and negative otherwise. Therefore, tracks from b/c hadron decays tend to have a positive sign. Figure 1 shows the distribution of the signed transverse impact parameter d_0 for b-tagging quality tracks in selected jets, for experimental data and simulation. The asymmetric shape of the distribution, with the enhanced positive part coming from heavy flavor decays, can be seen clearly. The simulation reproduces well the distribution in data, within 10% for most of the d_0 range. It can be shown that the disagreement is more pronounced at higher track momentum, suggesting that it may be an effect of residual misalignments. A smearing of the simulation is currently under investigation to correct the residual discrepancy, which originates in part from the fact that in simulation the measurement modules are perfectly aligned. The material appears to be well modelled by the simulation, since the agreement is better for low-momentum tracks, as seen previously [24]. For negative d_0, the agreement is slightly worse, as seen in previous studies [25] based on 2010 data. The error on d_0 as returned by the track fit[2] is shown in Figure 2 where the data/simulation agreement is also reasonable. Further studies in bins of $p^2 \sin^3\theta$ of the tracks show that the best description in simulation is for low-momentum tracks and/or tracks at large pseudorapidities. This indicates in particular that the description of the material in the simulation and its treatment in the track reconstruction are reasonable. The signed transverse impact parameter significance d_0/σ_{d_0} is shown on Figure 3. The significance, which gives more weight to tracks measured precisely, is the main ingredient of the tagging algorithms based on impact parameters. The moderate underestimation of the tails in the simulation is again due to high-momentum tracks: a further refined alignment could reduce the discrepancy. The significance z_0/σ_{z_0} is shown in Figure 4: in this case the resolution in data is better than in simulation. In both cases, the size of disagreements between the impact parameter significance distributions in data and Monte Carlo is minor and will be absorbed by the calibration procedures described in Ref. [27].

[1] Strictly speaking the impact parameter is $|z_0| \sin\theta$ where θ is the polar angle of the track.

[2] The error is actually convolved with the uncertainty on the primary vertex position.

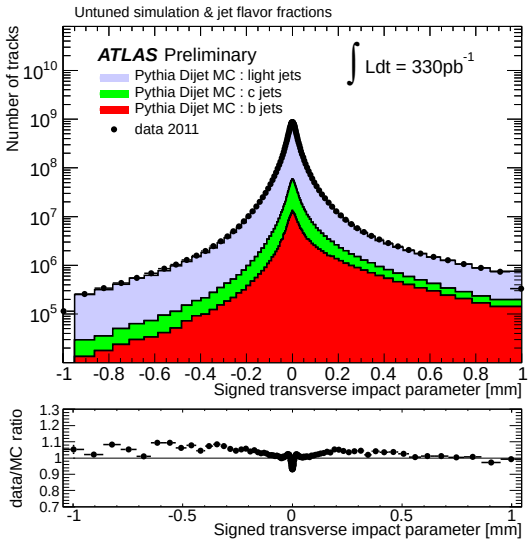


Figure 1: Distribution of the signed transverse impact parameter d_0 with respect to primary vertex for tracks of *b*-tagging quality associated to jets, for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). The ratio data/simulation is shown at the bottom of the plot.

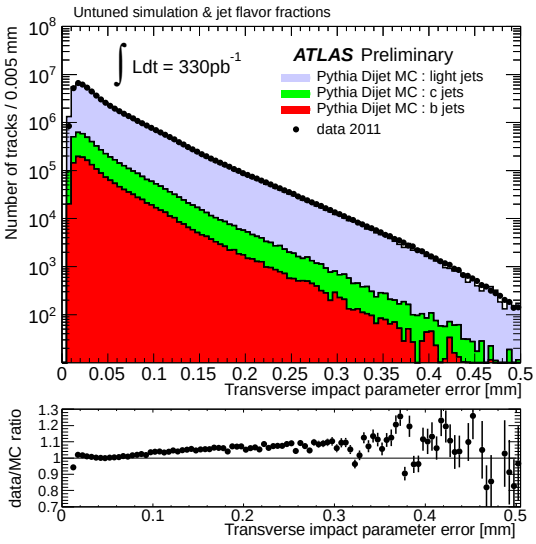


Figure 2: Distribution of the transverse impact parameter σ_{d_0} uncertainty with respect to primary vertex for tracks of *b*-tagging quality associated to jets, for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). The ratio data/simulation is shown at the bottom of the plot.

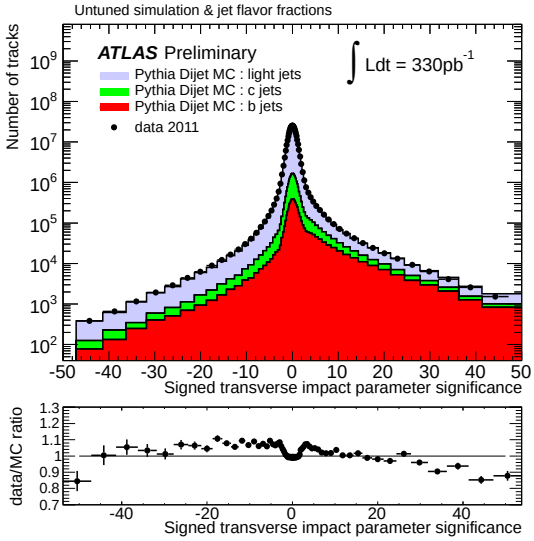


Figure 3: Distribution of the signed significance of the transverse impact parameter $d_0/\sigma(d_0)$ with respect to primary vertex for tracks of *b*-tagging quality associated to jets, for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). The ratio data/simulation is shown at the bottom of the plot.

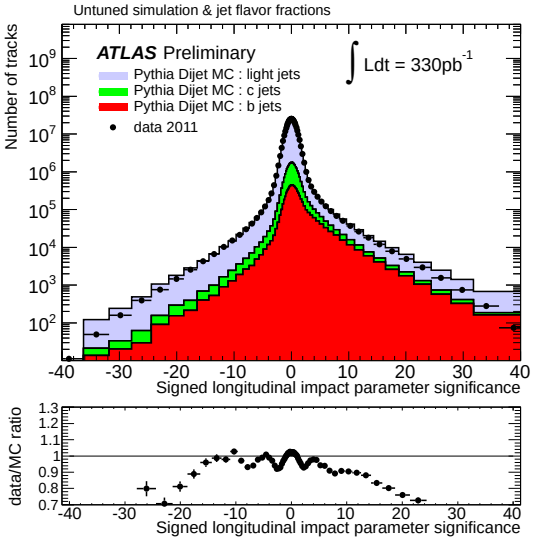


Figure 4: Distribution of the signed significance of the longitudinal impact parameter $z_0/\sigma(z_0)$ with respect to primary vertex for tracks of *b*-tagging quality associated to jets, for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). The ratio data/simulation is shown at the bottom of the plot.

7 Overview of the high-performance tagging algorithms

In this section, the *b*-tagging algorithms are described briefly. For more information, the reader is referred to Ref. [2].

A first stage, common to several algorithms, consists of attempting to reconstruct V^0 decays, in order to reject them since they can mimic tracks from *b*-hadron decays. The preselection cuts on track impact parameters reject a large fraction of long-lived particles and secondary interactions. The mass of the vertex formed by pairs of remaining tracks (Section 7.2) is used to reject the tracks that are likely to come from K_s, Λ decays and photon conversions. The radius of the vertex is compared to a crude description of the beam-pipe and the innermost pixel layers to reject secondary interactions in material.

7.1 Impact parameter-based algorithms

Combining the impact parameter significances of all the tracks in the jet is the basis of the first method to tag *b*-jets. A simple tagging algorithm doing this combination, JetProb, has been devised to be used for early data and has been extensively used in 2010 [3].

The IP3D high-performance tagging algorithm uses a likelihood ratio technique in which input variables are compared to pre-defined smoothed and normalized distributions for both the *b*- and light jet hypotheses, obtained from Monte Carlo simulation. The distributions in this case are two-dimensional histograms of the signed transverse impact parameter significance d_0/σ_{d_0} and longitudinal impact parameter significance z_0/σ_{z_0} of tracks, taking advantage of the correlations between the two variables.

7.2 Secondary vertex-based algorithms

To further increase the discrimination between *b*-jets and light jets, the inclusive vertex formed by the decay products of the *b*-hadron, including the products of the eventual subsequent charm hadron decay, can be sought. The search starts by building all two-track pairs that form a good vertex, using only tracks associated to the jet and far enough from the primary vertex. Vertices compatible with a V^0 or material interaction are rejected. All tracks from the remaining two-track vertices are combined into a single inclusive vertex, using an iterative procedure to remove the worst track until the χ^2 of the vertex fit is good.

The SV1 secondary vertex algorithm uses a looser track selection[3] than the one described in Section 6.1, primarily in order to maximize the efficiency to reconstruct V^0 decays and material interactions, whose corresponding tracks are subsequently removed for *b*-tagging purposes. The main differences in the selection cuts are the following: $p_{\mathrm{T}} > 400$ MeV, $|d_0| < 3.5$ mm (no cut on z_0), at least one hit in the pixel detector (no requirement on the innermost pixel layer), and no more than one hit on the track being shared with another track.

The decay length significance $L_{3D}/\sigma_{L_{3D}}$ measured in 3D and signed with respect to the jet direction can be used as a discriminating variable between *b*-jets and light jets: this is the principle of the SV0 tagger, extensively used as well with the 2010 data [4]. To increase the discriminating power, the high-performance tagging algorithm SV1 takes advantage of three of the vertex properties: the invariant mass of all tracks associated to the vertex, the ratio of the sum of the energies of the tracks in the vertex to the sum of the energies of all tracks in the jet, and the number of two-track vertices. These variables are combined using a likelihood ratio technique. SV1 relies on a 2D-distribution of the two first variables and a 1D-distribution of the number of two-track vertices. In addition the distance ΔR between the jet axis and the line joining the primary vertex to the secondary one is used.

[3] The track selection for the SV0 algorithm is even looser than the one of SV1, see [4].

7.3 Decay chain reconstruction with the JetFitter algorithm

A completely new algorithm, called JetFitter, is also available, which exploits the topology of weak *b*- and *c*-hadron decays inside the jet. A Kalman filter is used to find a common line on which the primary vertex and the *b*- and *c*-vertices lie, as well as their position on this line, giving an approximated flight path for the *b*-hadron. With this approach, the *b*- and *c*-hadron vertices are not necessarily merged, even when only a single track is attached to each of them. The discrimination between *b*-, *c*- and light jets is based on a likelihood using similar variables as in the SV1 tagging algorithm above, and additional variables such as the flight length significances of the vertices. This algorithm and its performance are also described in detail in Ref. [2, 26].

7.4 Combinations: IP3D+SV1 and IP3D+JetFitter

Thanks to the likelihood ratio method used for IP3D and SV1, the algorithms can be easily combined: the weights of the individual tagging algorithms are simply summed up. The combination JetFitter+IP3D is based on artificial neural network techniques with Monte Carlo simulated training samples and additional variables describing the topology of the decay chain.

7.5 Expected performance of the tagging algorithms

The expected improvement in *b*-tagging performance achieved by the high-performance tagging algorithms is discussed in this section, as determined using simulated $t\bar{t}$ events. Some comparisons of the algorithms using experimental data are discussed in Section 8.3. So far, the various techniques used to measure directly in data the *b*-tagging performance of the early tagging algorithms, JetProb and SV0, have shown [27] that the simulation properly describes both the *b*-jet tagging efficiency and the light jet rejection: the respective scale factors by which the *b*-tag efficiency and mistag rates in simulation have to be adjusted to be compared to data are both compatible with unity. While this may be slightly different with the high-performance algorithms, notably given their much higher light-jet rejection, preliminary results are very encouraging since scale factors are also very close to 1, and indicate that the following results based on simulation are giving an accurate picture of the performance to expect.

Figure 5 shows the light-jet rejection as a function of the *b*-jet tagging efficiency ε_b for the various ATLAS *b*-tagging algorithms. It is obtained by varying continuously the operating point of each tagger, *i.e.* the cut on its output discriminating variable. The jets are from $t\bar{t}$ simulated events, and must fulfill the following cuts: $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. The tagging efficiency is the fraction of jets labeled as *b*-jets that are properly tagged, while the rejection is the reciprocal of the fraction of jets that are labeled as light jets and are actually tagged incorrectly by the algorithm. The labeling procedure, which is not unambiguous, is based on the flavor of true quarks: a jet is labeled as a *b*-quark if a *b* quark with $p_{\mathrm{T}} > 5$ GeV is found in a cone of size $\Delta R = 0.3$ around the jet direction. The various labeling hypotheses are tried in this order: *b* quark, *c* quark and τ lepton. When none of these hypotheses are satisfied, the jet is labeled as a light jet. No attempt is made to distinguish light jets originating from gluons from those originating from quarks.

For the same tagging efficiency, the high-performance tagging algorithms are expected to have much lower mistagging rates than their counterpart algorithms designed for early data. In addition, by combining the vertexing techniques and the impact parameter information, the IP3D+SV1 and IP3D+JetFitter algorithms can reach very high tagging efficiencies. For a 50% (resp. 60%) *b*-tagging efficiency, these results show that a high-performance tagger like the IP3D+SV1 algorithm achieves a factor of five (resp. four) reduction in the mistag rate in events with $t\bar{t}$-like kinematics, compared to the early tagging algorithms used currently for physics analyses. At 70% *b*-tagging efficiency, the IP3D+JetFitter algorithm achieves a mistag rate lower than 1%.

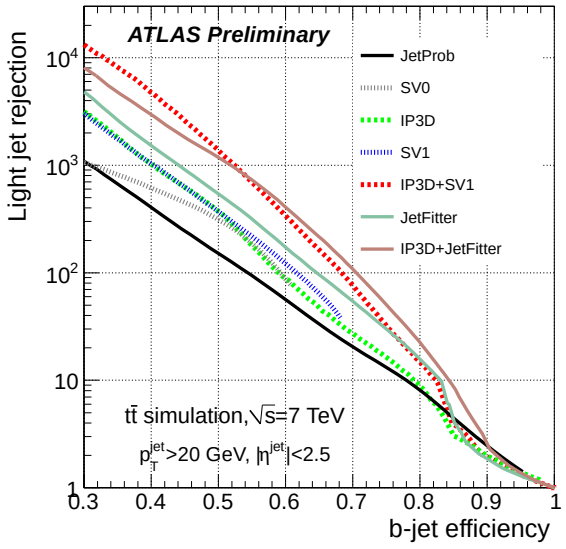


Figure 5: Light-jet rejection as a function of the *b*-jet tagging efficiency for the early tagging algorithms (JetProb and SV0) and for the high-performance algorithms, based on simulated $t\bar{t}$ events.

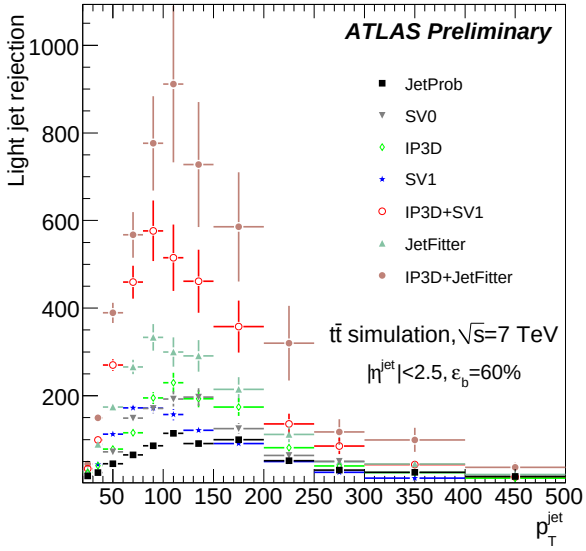


Figure 6: Light-jet rejection as a function of the jet transverse momentum p_{T}, for operating points of the various tagging algorithms leading to the same $\varepsilon_b^{t\bar{t}} = 60\%$, based on simulated $t\bar{t}$ events.

Note that the results depend strongly on the kinematics of the sample under study, and that this variation is specific to each tagging algorithm, as can be seen in Figure 6 which shows the light jet rejection as a function of the jet p_{T} for the various tagging algorithms operating in such a way that they all lead to the same $\varepsilon_b^{t\bar{t}} = 60\%$. One can notice that for all algorithms, the tagging performance is optimal for $p_{\mathrm{T}}^{\mathrm{jet}} \sim 100$ GeV: below this threshold, tracks in jets are relatively soft and therefore multiple scattering is compromising the resolution on the impact parameter, while above this p_{T} value several effects such as merged pixel clusters and pattern-recognition issues conspire to reduce the performance. Further information can be found in Ref. [2].

8 Commissioning studies

Two aspects of the Monte Carlo description of the data are studied during the commissioning of the algorithms: the level of agreement in the description of the tagging output discriminating variable and any associated auxiliary variables, and the description of the tagging rates in inclusive and heavy flavor-enhanced jet samples. The tagging rate is defined as the fraction of jets that are tagged out of those that could be tagged by a given algorithm, for a specific choice of its operating point. The fraction of tagged jets has some dependence on the jet p_{T} (and to a lesser extent on η) which is specific to each algorithm. For a typical sample with a mix of flavors and a fraction f_b of *b*-jets, the integrated tagging rate is the sum of the fraction of *b*-jets correctly tagged by the algorithm and the fraction of non-*b* jets tagged by mistake, *i.e.* $f_{\mathrm{tagged}} = f_b\varepsilon_b + (1 - f_b)\varepsilon_l$, where ε_l is the efficiency to tag a non-*b* jet by mistake and is algorithm- (and also p_{T}- and η-) dependent.

It is important to note that for the tagging rate histograms in the following, the data and simulation plots were not normalized to the same area: both the shape and normalization contain information. For the sake of comparison the operating points are not described by the actual cut value on the output of

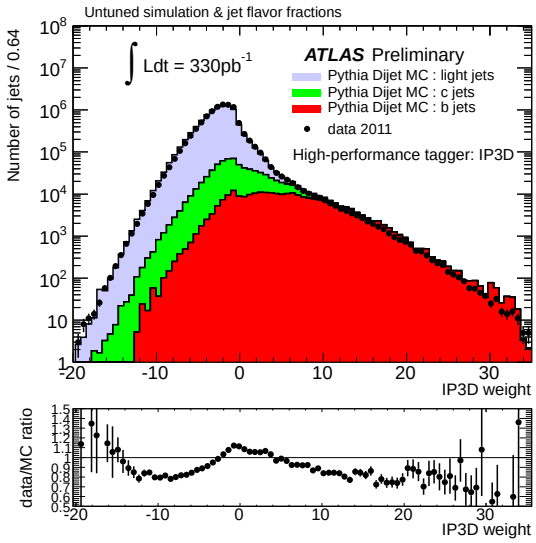


Figure 7: Distribution of the jet output of the IP3D tagging algorithm for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of the plot.

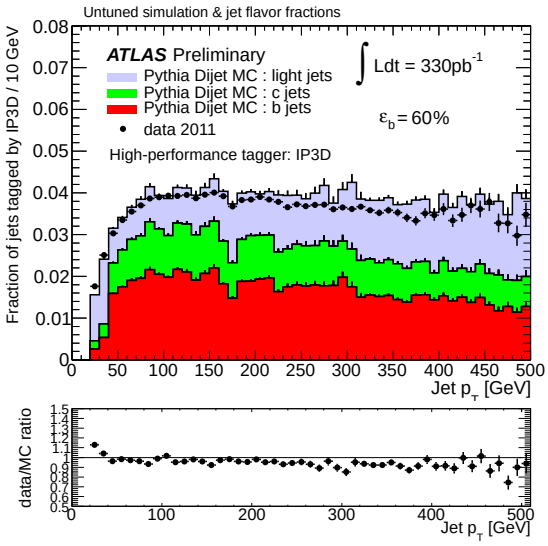


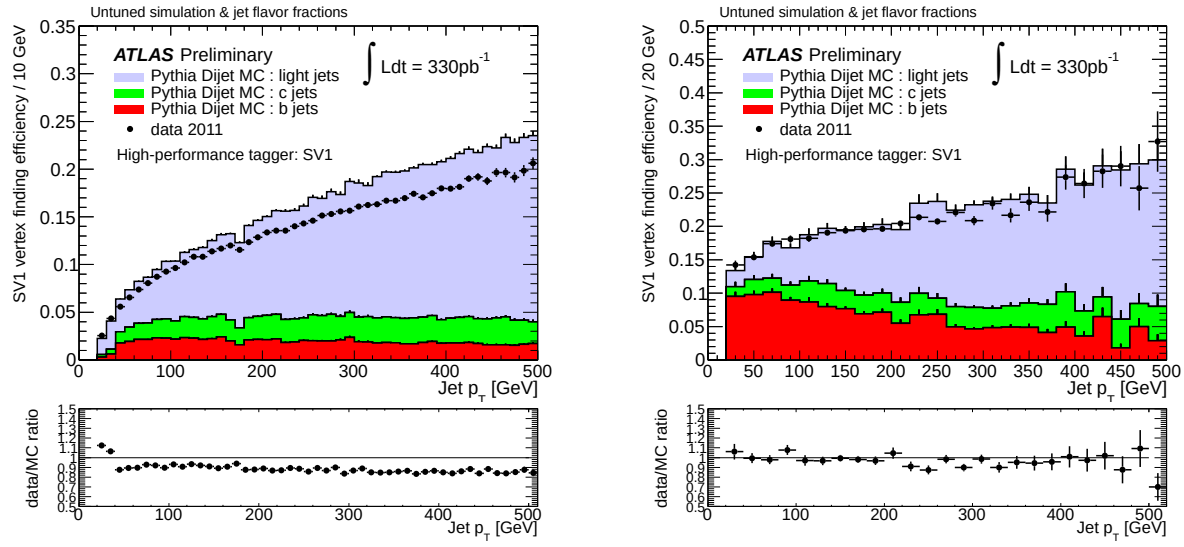
Figure 8: Distribution of the tagging rate for the IP3D tagging algorithm versus the jet p_{T} at an operating point $\varepsilon_b \approx 60\%$ for experimental data (solid black points) and for simulated data (filled histograms). Jets are from the inclusive leading jet sample.

each algorithm: instead the cuts are chosen to yield a fixed b-jet tagging efficiency (50% or 70% for instance) on a simulated $t\bar{t}$ sample. Therefore in the following, the operating points are referred to by their corresponding $\varepsilon_b^{t\bar{t}}$ efficiency.

It is also important to note that no systematic uncertainties are shown on the Monte Carlo predictions. Two important effects can influence the details of the agreement with data — the residual discrepancies in modeling the impact parameter resolutions, and the correct modeling of the b, c and light flavour composition of the jet samples by the Pythia Monte Carlo generator. Rather than evaluating these in detail (*e.g.* by using the impact parameter smearing procedures discussed in Section 6.3), these residual discrepancies will be addressed in practice by means of the tag calibration procedures [27]. This is discussed further in Section 8.3.

8.1 IP3D+SV1 algorithm

Since IP3D+SV1 is a combination of two algorithms, their individual behavior is discussed first. The output of the IP3D algorithm for taggable jets is shown in Figure 7, for experimental data and for simulated data. The simulated sample is broken into three components: b-jets, c-jets and light jets, according to the flavor composition predicted by the Pythia generator. The simulation gives a reasonably good description of the data, within 20%, with however a discrepancy in the region with negative weights, dominated by light jets. This discrepancy arises from the non-perfect modelling of the impact parameter discussed previously in Section 6.3, which can be accounted for with an *ad hoc* smearing of the simulated impact parameters. The positive tail is also not perfectly reproduced. Currently an operating point chosen in simulation would overestimate slightly the efficiency and would lead to fewer jets being tagged in data, as can be seen in Figure 8. This figure shows the tagging rate for the IP3D algorithm operating at a 60% b-jet efficiency. Despite the output weight discrepancy, the shape and the fraction of tagged jets are reasonably well reproduced by the simulation.



(a) inclusive leading jet sample (b) heavy-flavor enriched jet sample

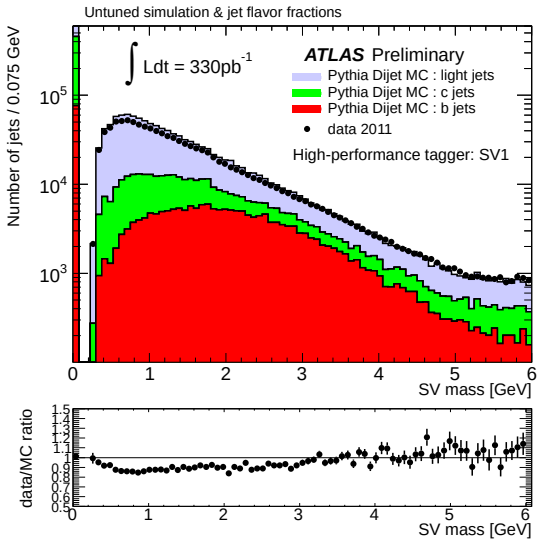
Figure 9: Distribution of the vertex finding efficiency by the SV1 tagging algorithm versus the jet transverse momentum for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). The plots are for two samples: (a) inclusive leading jets sample and (b) sample enriched in heavy-flavor jets. The ratio data/simulation is shown at the bottom of each plot.

The other ingredient is the SV1 algorithm. The reconstruction of an inclusive secondary vertex is not always successful. The fraction of jets in which such a vertex could be found is shown in Figure 9 for experimental and simulated data, as a function of the jet transverse momentum. Two samples with different flavor compositions are used, the inclusive leading jet sample and one enriched in heavy flavor jets. The simulated data agree reasonably well with the experimental data in the latter sample. In the sample dominated by light jets, the simulation slightly overestimates (by $\approx 10\%$) the secondary vertex reconstruction efficiency. The properties of the inclusive vertex found by SV1 are shown in Figure 10. The peaks at zero correspond to jets in which no secondary vertex could be reconstructed. All vertex properties shown are well modelled by the simulation within 20%.

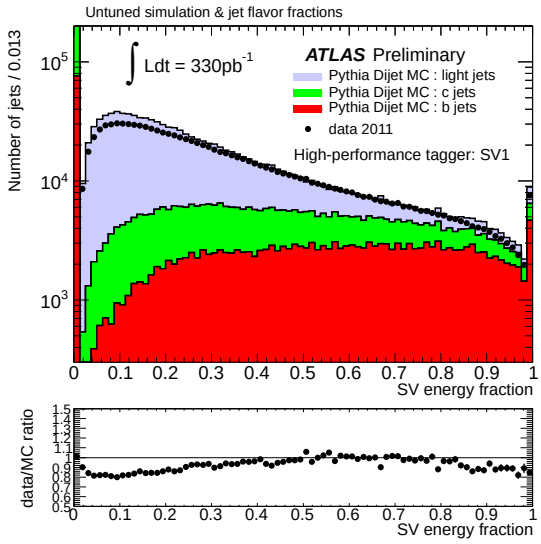
Finally, the resulting output weight of the IP3D+SV1 combined algorithm is shown in Figure 11. The behavior is very similar to the IP3D case, but the level of agreement between data and simulation is improved by the SV1 component which is better modeled than IP3D. The tagging rates as a function of the jet p_{T} and for an operating point leading to a b-jet efficiency of 60% in $t\bar{t}$ events are shown in Figure 12, for jets from the inclusive leading jet sample and for jets from the heavy-flavor enriched sample. The tagging rates on the inclusive leading jet sample are shown in Figure 13 as a function of the jet pseudorapidity and jet azimuthal angle. In the latter, one can see at $\phi \sim 2$ the impact of the two modules in the innermost pixel layer that were disabled for data-taking but not masked in the simulation. Apart from these well understood inconsistencies, shapes are well predicted by the simulation and the overall rates in simulation agree with the ones in data within 20%.

8.2 IP3D+JetFitter

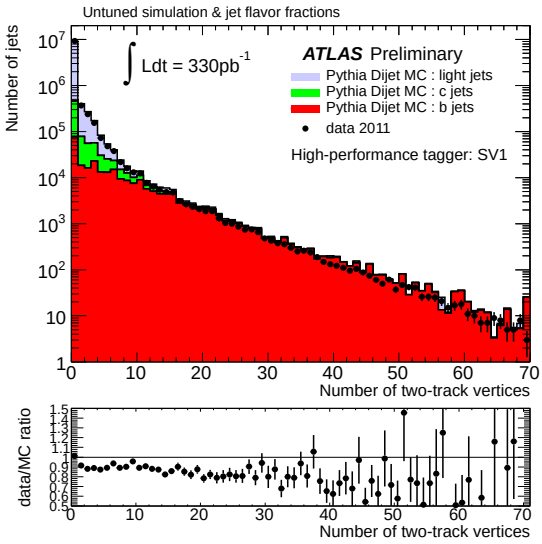
The same checks are now performed for the IP3D+JetFitter tagging algorithm. The output of the JetFitter algorithm (not combined with IP3D) is shown in Figure 14, while its tagging rate at an operating point $\varepsilon_b^{t\bar{t}} \approx 60\%$ is shown in Figure 15.



(a) vertex mass



(b) vertex energy fraction



(c) number of two-track vertices

Figure 10: Distribution of the properties of the vertex found by the SV1 tagging algorithm for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of each plot.

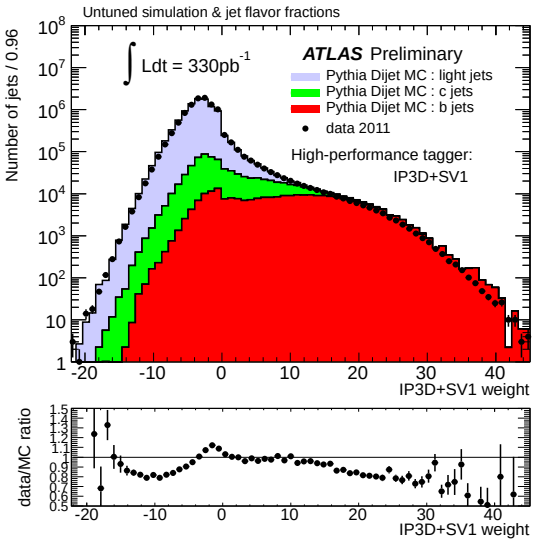
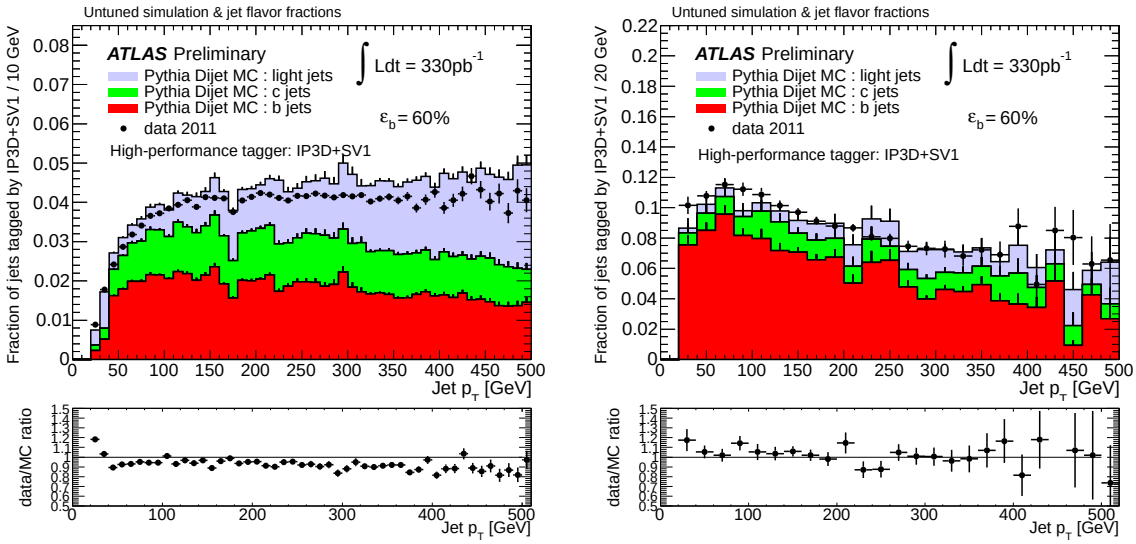


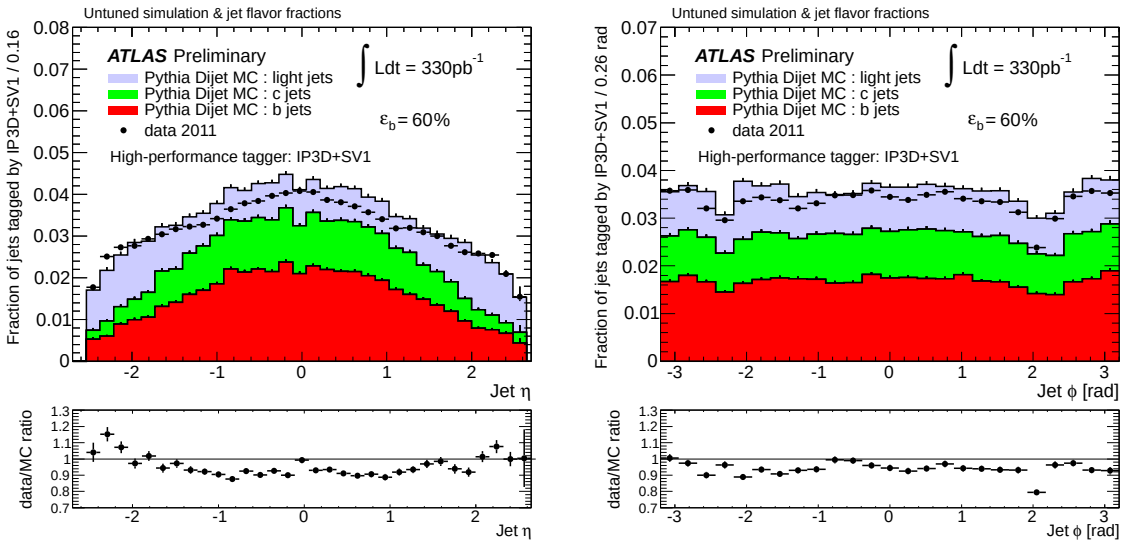
Figure 11: Distribution of the output of the IP3D+SV1 tagging algorithm for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of the plot.



(a) inclusive leading jet sample

(b) heavy flavor enriched jet sample

Figure 12: Distribution of the tagging rate for the IP3D+SV1 tagging algorithm at an operating point $\varepsilon_b \approx 60\%$ for experimental data (solid black points) and for simulated data (filled histograms for the various flavors) versus the jet transverse momentum, for two jet samples: (a) the inclusive jet sample and (b) the sample enriched in heavy-flavor jets. The ratio data/simulation is shown at the bottom of each plot.



(a) versus jet η (b) versus jet ϕ

Figure 13: Distribution of the tagging rate for the IP3D+SV1 tagging algorithm at an operating point $\varepsilon_b \approx 60\%$ for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of each plot.

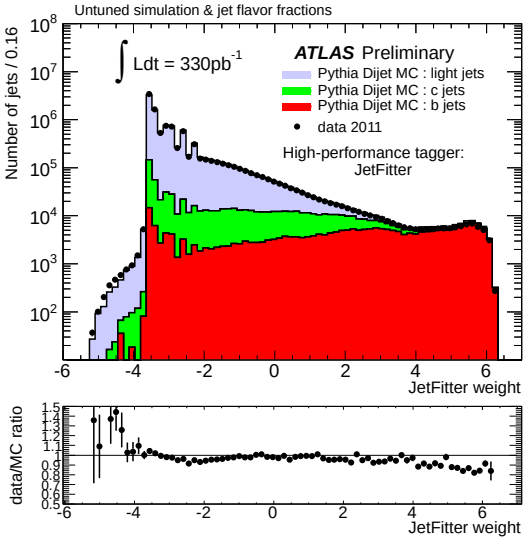


Figure 14: Distribution of the jet output of the JetFitter tagging algorithm for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of the plot.

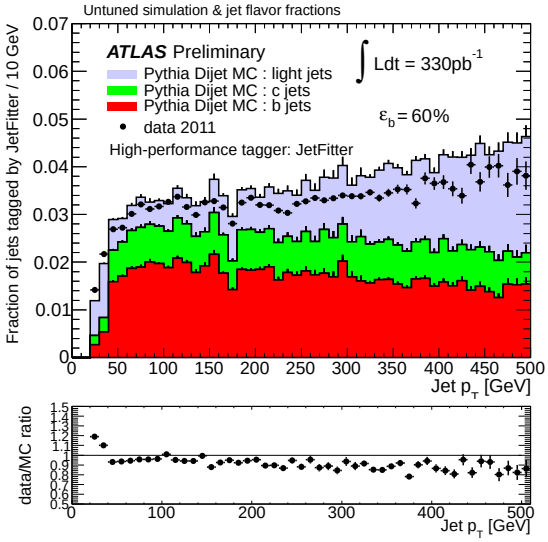


Figure 15: Distribution of the tagging rate for the JetFitter tagging algorithm versus the jet p_{T} at an operating point $\varepsilon_b \approx 60\%$ for experimental data (solid black points) and for simulated data (filled histograms). Jets are from the inclusive leading jet sample.

Some of the properties of the decay chain found by JetFitter are shown in Figure 16: the number of vertices with two or more tracks, the number of single-track vertices, the invariant mass of the tracks used anywhere along the decay chain and the fraction of their energy compared to the jet energy. The simulation reproduces these distributions fairly well, which is remarkable given these are the results of very complex algorithms.

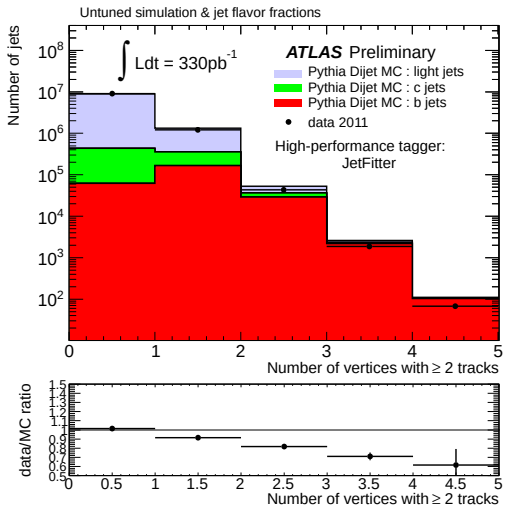
The jet weight resulting from the combination of JetFitter with IP3D can be seen in Figure 17. The shape in experimental data is closely reproduced by the simulation, except in the negative weight region, dominated by light jets, where the impact parameter resolution issue already discussed (Section 6.3) manifests itself through the combination with IP3D. The tagging rates for the inclusive jet sample and for the one enriched in heavy-flavor jets are shown on Figure 18, this time with a looser configuration of the algorithm ($\varepsilon_b^{t\bar{t}} \approx 70\%$). The tagging rates as a function of the jet pseudorapididty and azimuthal angle are shown in Figure 19. In all cases, the tagging rate predicted by the simulation agrees with experimental data to within 20%.

8.3 Comparison of the algorithms

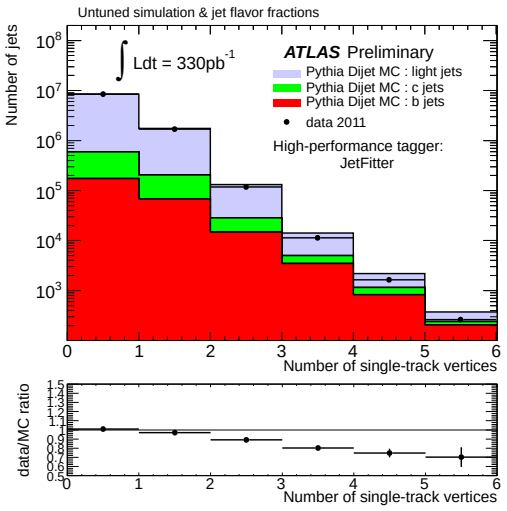
In this section, we discuss briefly how the performance of the high-performance tagging algorithms compares with the one of the early-data algorithms, using experimental data. As seen in Section 7.5, the performance of the tagging algorithms is characterized, for a given operating point, by the efficiency to tag real b-jets $\varepsilon_b(p_{\mathrm{T}}, \eta)$ and by the rejection of non-b jets, which is the inverse of the fraction of mistagged jets $\varepsilon_l(p_{\mathrm{T}}, \eta)$. Since it is not possible to select data samples with a high-enough purity in b-jets, and even less so with a high purity of light jets, a detailed comparison would in principle require to disentangle the behavior of the taggers and the actual flavor composition of the sample. Calibration techniques to do so are described in Ref. [27] and will be used in the future, but are beyond the scope of this note. Instead a qualitative approach is used in the following to give an indication of the improved performance to be expected from the high-performance b-tagging algorithms.

The SV0 algorithm, which has been used so far in many physics analysis at an operating point of $\varepsilon_b^{t\bar{t}} = 50\%$ in simulation, is compared to the new IP3D+SV1 algorithm at an operating point chosen to give the same efficiency $\varepsilon_b^{t\bar{t}}$. Figure 20(a) shows the SV0 tagging rate in data as a function of the jet p_{T} in the inclusive jet sample. For illustration, the tagging rates in the simulated samples are also shown. In Figure 20(b), analogous tagging rates are shown for the IP3D+SV1 algorithm. The IP3D+SV1 algorithm appears to be less prone than the SV0 algorithm to mistakenly tag light jets as b-jets. It is important to note however that the comparison is qualitative only and that it is not possible to infer for instance the ratio of the mistag rates of the two algorithms from these plots. This is because, as can be seen from Figure 5, a change in the operating point leading to a minor change in the b-tag efficiency will have a large impact on the mistag rate. Moreover, even if the operating points of the two algorithms were chosen to give the same inclusive b-tag efficiency in a simulated $t\bar{t}$ sample, there is no guarantee that the algorithms would operate at exactly the same tagging efficiency in data in the dijet sample. On the contrary, this is quite unlikely since the kinematics of the jet sample considered here are different from $t\bar{t}$ and the tagging efficiency depends strongly on the p_{T} of the jet. Slight differences between the b-tag efficiencies in data and simulation and the associated change in the light jet rejection also allow for the observed differences in the tagging rates when comparing data and Monte-Carlo without posing a major problem for the commissioning of these algorithms.

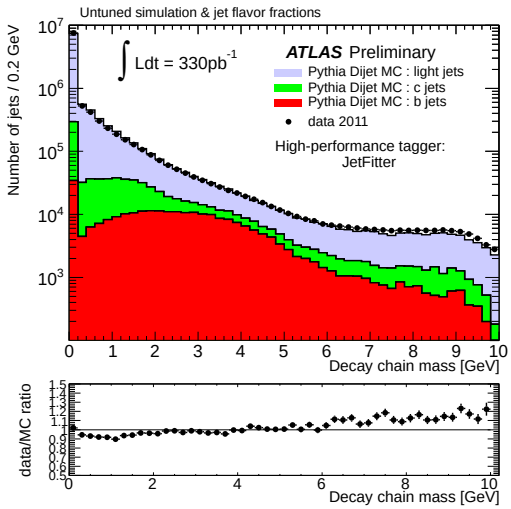
As mentioned already, one feature of the new tagging algorithms is their ability to operate at a high b-jet tagging efficiency, *e.g.* 70%, while still achieving a high rejection of light jets. Of the early algorithms, only the JetProb algorithm could reach such high efficiencies, but with a poor light jet rejection. The SV0 algorithm efficiency is intrisically limited to 50%-60%, which is the efficiency to reconstruct a secondary vertex. Therefore the new IP3D+JetFitter algorithm at $\varepsilon_b^{t\bar{t}} = 70\%$ is compared to JetProb operating at the same $\varepsilon_b^{t\bar{t}}$. Figure 21(a) shows the tagging rate as a function of the jet p_{T} for the JetProb algorithm in the



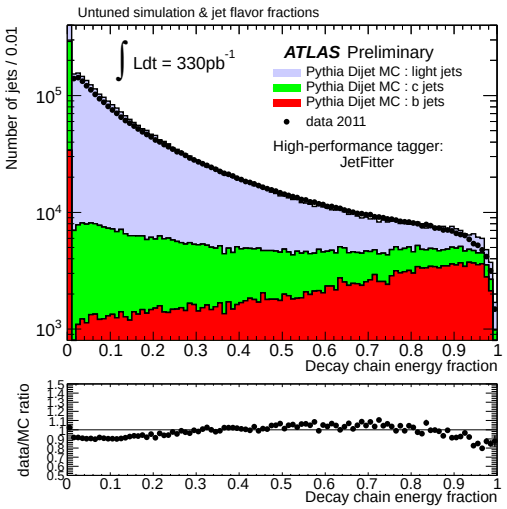
(a) number of vertices with ≥ 2 tracks



(b) number of single-track vertices



(c) decay chain mass



(d) decay chain energy fraction

Figure 16: Distribution of some of the properties of the vertex found by the JetFitter tagging algorithm for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of each plot.

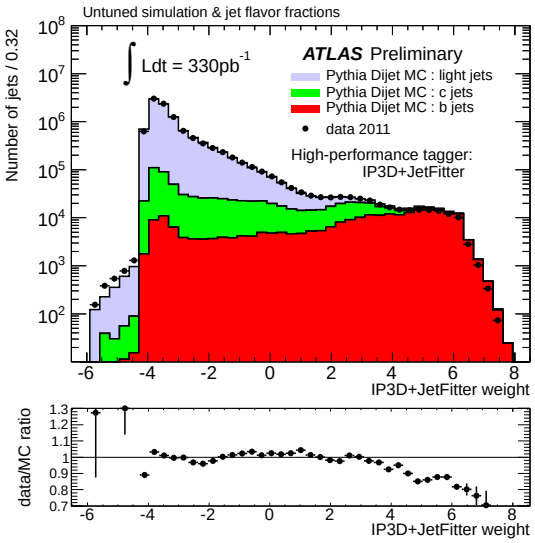
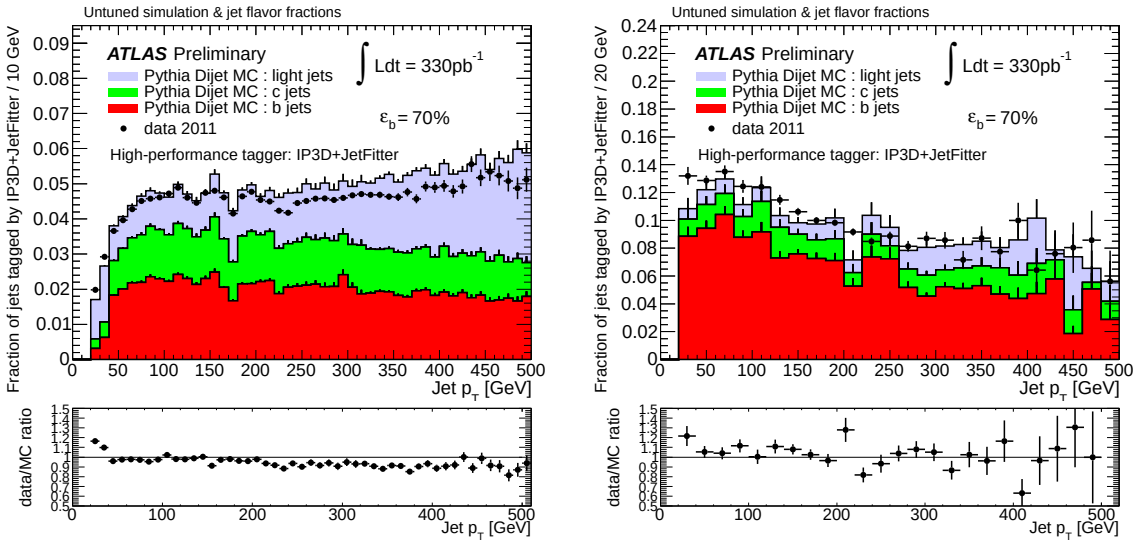


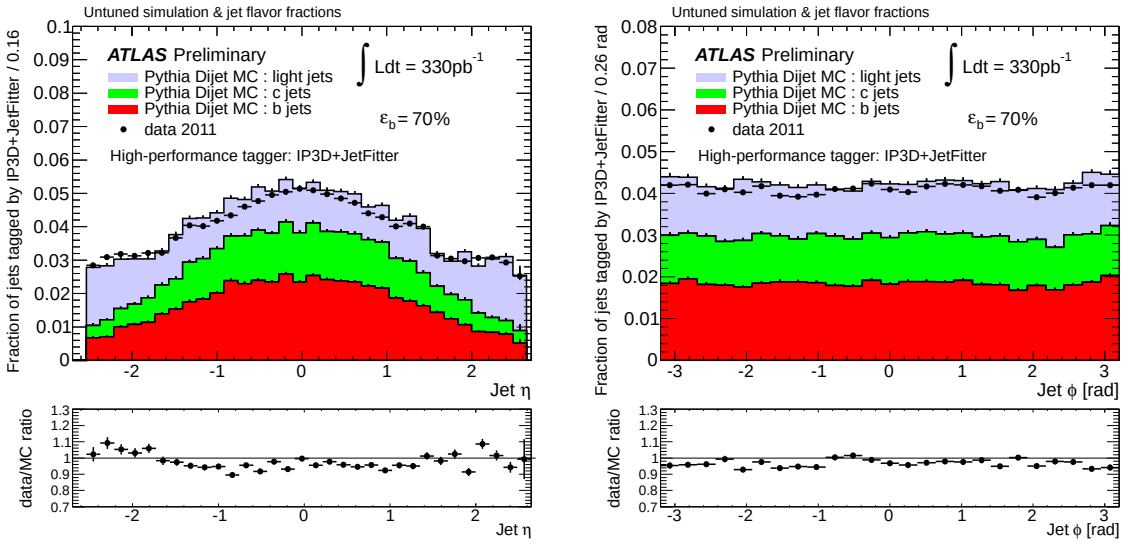
Figure 17: Distribution of the output of the IP3D+JetFitter tagging algorithm for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of the plot.



(a) inclusive leading jet sample

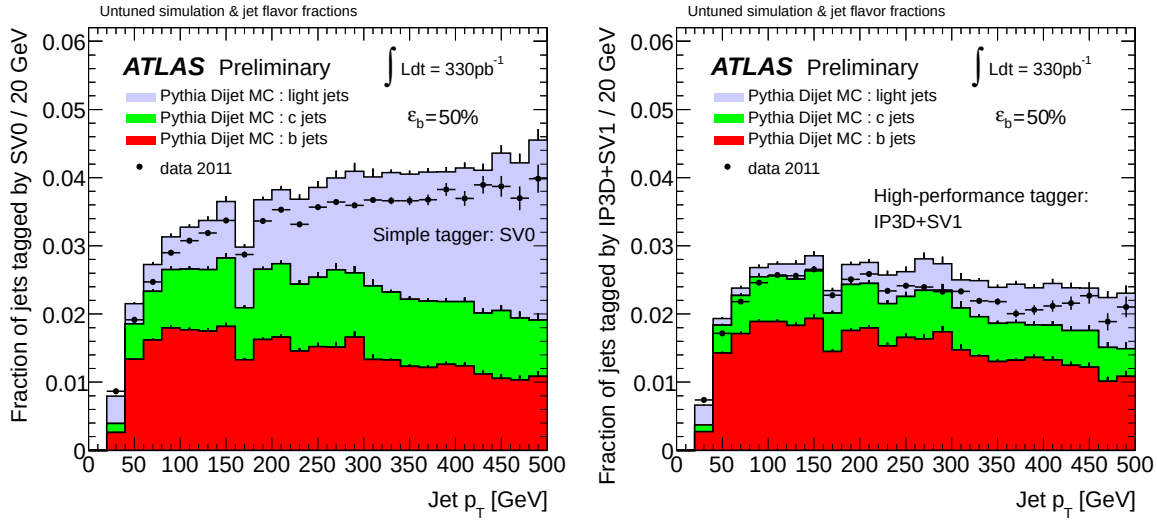
(b) heavy-flavor enriched jet sample

Figure 18: Distribution of the tagging rate for the IP3D+JetFitter tagging algorithm at an operating point $\varepsilon_b \approx 70\%$ for experimental data (solid black points) and for simulated data (filled histograms for the various flavors) versus the jet transverse momentum, for two jet samples: (a) inclusive light jet sample and (b) sample enriched in heavy-flavor jets. The ratio data/simulation is shown at the bottom of each plot.



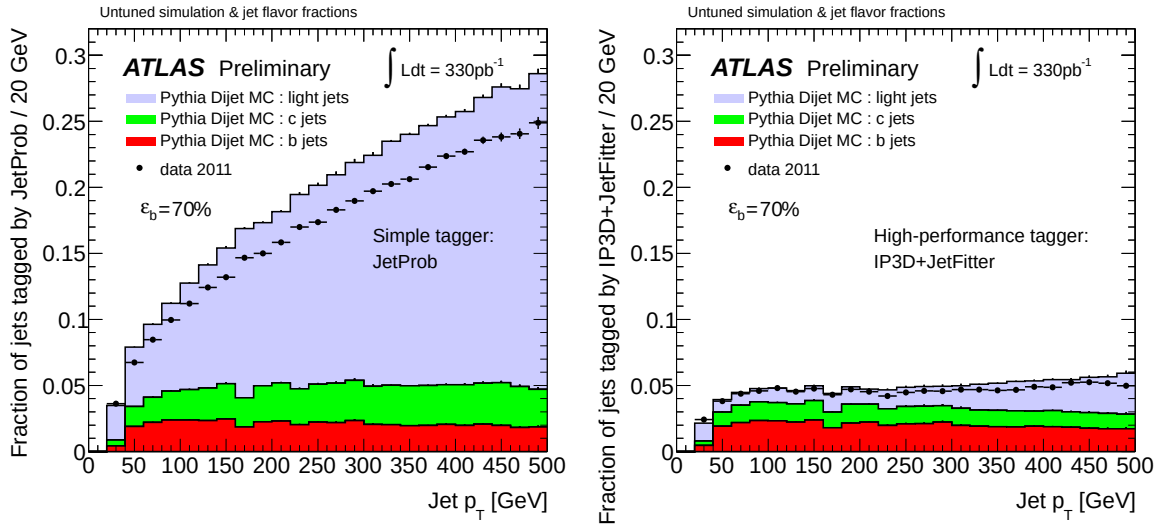
(a) versus jet η (b) versus jet ϕ

Figure 19: Distribution of the tagging rate for the IP3D+JetFitter tagging algorithm at an operating point $\varepsilon_b \approx 70\%$ for experimental data (solid black points) and for simulated data (filled histograms for the various flavors). Jets are from the inclusive leading jet sample. The ratio data/simulation is shown at the bottom of each plot.



(a) SV0 (b) IP3D+SV1

Figure 20: Distribution of the tagging rates on the inclusive leading jet sample, for the SV0 and IP3D+SV1 tagging algorithms with their respective operating points tuned to a priori give the same 50% b-jet efficiency on a $t\bar{t}$ simulated sample. Note that since there is no guarantee that the two algorithms actually tag exactly the same fraction of b-jets in the data sample considered (cf. text), no quantitative conclusion on the relative mistag rates of the two algorithms should be drawn from this comparison.



(a) JetProb

(b) IP3D+JetFitter

Figure 21: Distribution of the tagging rates on the inclusive leading jet sample, for the JetProb and IP3D+JetFitter tagging algorithms with their respective operating points tuned to a priori give the same 70% *b*-jet efficiency on a $t\bar{t}$ simulated sample. Note that since there is no guarantee that the two algorithms actually tag exactly the same fraction of *b*-jets in the data sample considered (cf. text), no quantitative conclusion on the relative mistag rates of the two algorithms should be drawn from this comparison.

inclusive jet sample, while Figure 21(b) shows the same information for the IP3D+JetFitter algorithm. The same conclusion holds, namely that the fraction of mistagged non-*b* jets appears to be substantially higher for JetProb than for the more powerful IP3D+JetFitter algorithm.

9 Conclusion

In this note, high-performance *b*-tagging algorithms have been studied with a sample of 330 pb^{-1} of pp collision data collected by ATLAS in 2011, and compared with simulated events.

Most of the distributions of their input variables as well as their output variables are well reproduced by the simulation, typically at the 10% level. One of the main reasons for residual differences between data and simulation is the slightly better impact parameter resolution in simulation compared to data. This can be adressed *e.g.* by smearing the impact parameter distribution in simulation to agree with that in data — work that is ongoing but beyond the scope of this note. However the magnitude of the problem is not such that it should delay the commissioning of the high-performance algorithms.

It has been shown that, once an operating point is selected for an algorithm, the fraction of tagged jets in data and in simulation agree well within 20%. This has been tested on the leading jet of the events, on sub-leading jets, and also on events with a different flavor composition. It is known from past experience that a description of the data by simulation at the level of 20%, which is similar to what was achieved for the early-data algorithms, is sufficient to allow for a successful calibration of the tagging algorithms. Indeed, such a level of agreement is remarkable given the complexity of the taggers and the number of effects that have to be correctly modelled in the Monte Carlo simulation to get to this level.

For the same operating point, the gain to be expected with the high-performance taggers in comparison with the early-data taggers has also been estimated qualitatively directly in data. The material

presented in this note gives confidence that the high-performance tagging algorithms are ready to be used in physics analyses as soon as their performance has been calibrated in data. They are expected to bring a substantial gain for most analyses, since for the same tagging efficiency as the algorithms used in 2010 they are projected to yield a mistag rate between two and five times lower. In addition, since these algorithms can be operated at a higher b-tagging efficiency ($\geq 70\%$) with still an acceptable light jet rejection, they are also very promising for searches of New Physics with low production cross sections.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** S08003 (2008).

[2] ATLAS Collaboration, *Expected performance of the ATLAS Detector, Trigger and Physics*, CERN-OPEN-2008-020, arXiv:0901.0512 (2009).

[3] ATLAS Collaboration, *Performance of impact parameter-based b-tagging algorithms with the ATLAS detector using pp collisions at $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2010-091 (2010).

[4] ATLAS Collaboration, *Performance of the ATLAS secondary vertex b-tagging algorithm in 7 TeV collision data*, ATLAS-CONF-2010-042 (2010).

[5] ATLAS Collaboration, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV*, EPJC 71, 1577 (2011).

[6] ATLAS Collaboration, *Search for supersymmetry using final states with one lepton, jets, and missing transverse momentum with the ATLAS detector in $\sqrt{s} = 7$ TeV pp collisions*, Phys. Rev. Lett. 106, 131802 (2011).

[7] ATLAS Collaboration, *Search for supersymmetry in pp collisions at $\sqrt{s} = 7$ TeV in final states with missing transverse momentum and b-jets*, Phys. Lett. B701, 398 (2011).

[8] ATLAS Collaboration, *Determination of the Top-Quark Mass from the $t\bar{t}$ Cross Section Measurement in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector*, ATLAS-CONF-2011-054 (2011).

[9] ATLAS Collaboration, *Measurement of the inclusive and dijet cross sections of b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, ATLAS-CONF-2011-056 (2011).

[10] ATLAS Collaboration, *Search for FCNC Top Quark Processes at $\sqrt{s} = 7$ TeV with the ATLAS Detector*, ATLAS-CONF-2011-061 (2011).

[11] ATLAS Collaboration, *Search for $t\bar{t}$ production in the all-hadronic channel in ATLAS with $\sqrt{s} = 7$ TeV data*, ATLAS-CONF-2011-066 (2011).

[12] ATLAS Collaboration, *Observation of t-channel Single Top-Quark Production in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, ATLAS-CONF-2011-088 (2011).

[13] *PYTHIA 6.4 physics and manual*, T. Sjöstrand, S. Mrenna and P. Skands, JHEP 0605 (2006) 026.

[14] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. C63 (2009) 189, arXiv:hep-ph/0901.0002;
A. Sherstnev and R. S. Thorne, *Different PDF approximations useful for LO Monte Carlo generators*, arXiv:hep-ph/0807.2132.

[15] ATLAS Collaboration, *Charged particle multiplicities in pp interactions at $\sqrt{s} = 900$ GeV and 7 TeV in a diffractive limited phase-space measured with the ATLAS detector at the LHC and new PYTHIA6 tune*, ATLAS-CONF-2010-031 (2010).

[16] S. Frixione, P. Nason and B.R. Webber. *Matching NLO QCD and parton showers in heavy flavour production*, JHEP 0308 (2003) 007, arXiv:hep-ph/0305252;
S. Frixione, E. Laenen and P. Motylinski, *Single-top production in MC@NLO*, JHEP 0603 (2006) 092, arXiv:hep-ph/0512250.

[17] P.M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev, D78 (2008) 013004, arXiv:hep-ph/0802.0007.

[18] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C70 (2010) 823874, arXiv:physics.ins-det/1005.4568.

[19] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. A506 (2003) 250.

[20] ATLAS Collaboration, *Performance of primary vertex reconstruction in pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*, ATLAS-CONF-2010-069 (2010).

[21] M. Cacciari, G. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP 0804 (2008) 063, arXiv:arXiv:0802.1189. M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B 641 (2006) 5761, hep-ph/0512210. G. P. Salam, M. Cacciari, and G. Soyez, *FastJet package*, http://www.lpthe.jussieu.fr/ salam/fastjet/.

[22] ATLAS Collaboration, *Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using pp Collisions at $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2010-053 (2010).

[23] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty for jets produced in pp collisions at $\sqrt{s} = 7$ TeV and measured with the ATLAS detector*, ATLAS-CONF-2010-056 (2010).

[24] ATLAS Collaboration, *Charged-particle multiplicities in pp interactions at $\sqrt{s} = 900$ GeV measured with the ATLAS detector at the LHC*, Phys. Lett. B 688 (2010) 21-42.

[25] ATLAS Collaboration, *Tracking studies for b-tagging in 7 TeV collision data with the ATLAS detector*, ATLAS-CONF-2010-040 and ATLAS-CONF-2010-070 (2010).

[26] G. Piacquadio, C. Weiser, *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*, J.Phys.Conf.Ser. 119 032032 (2008).

[27] ATLAS Collaboration, *Calibrating the b-tag Efficiency and Mistag Rate in 35/pb of Data with the ATLAS Detector*, ATLAS-CONF-2011-089 (2011).