ATLAS Note

Draft version 0.1

Measuring the b-jet Identification Efficiency in a Sample of Jets Containing Muons in pp Collisions at $\sqrt{s}$ = 13 TeV recorded with the ATLAS Detector

Ingo Burmeister[a], Rachael Ann Creager[b], Joseph Kroll[b], Bingxuan Liu[c], Valentina Vecchio[d]

[a]*Technische Universität Dortmund,* [b]*University of Pennsylvania,* [c]*HEP Division, Argonne National Laboratory,* [d]*Universita e INFN Roma Tre*

20th July 2018

The identification of jets originating from b quarks, or "b-tagging", is crucial to many physics analyses conducted using data from the ATLAS experiment at the Large Hadron Collider. This note presents the measurement of the efficiency of b-tagging, ε_b, using a sample of jets containing a spatially-matched muon. These muons originate from the semileptonic decays of bottom and charm hadrons, and from hadrons in light quark and gluon jets that are incorrectly identified as muons.

Using the magnitude of the transverse momentum of the lepton relative to the jet+lepton axis, $p_{\mathrm{T}}^{\mathrm{rel}}$, the fraction of jets originating from b-quarks is measured before and after b-tagging, enabling the determination of ε_b. The efficiencies are reported as ratios (scale factors) between the efficiency measured in data and the efficiency in Monte Carlo simulation. The scale factors for the b-tagging algorithm known as MV2c10 are reported as a function of the transverse momentum of the jet for four different operating points, corresponding to 60%, 70%, 77% and 85% b-tagging efficiency in b-jets produced in top-quark pair production.

Contents

1 Introduction

The identification of jets containing b-hadrons, b-tagging, is an essential part of many different searches and measurements at the ATLAS experiment, including precision measurements in the top quark sector [1], Higgs couplings measurements [2] and a variety of searches for physics beyond the Standard Model. As b-tagging plays a critical role in these analyses, it is important to have a thorough understanding of b-tagging performance. An extensive summary of flavour-tagging efforts at the ATLAS experiment in Run 1 can be found in [3]. Taking advantages of the relatively larger semi-leptonic branching ratio of b-hadrons, this work calibrates the b-tagging efficiency in a sample of jets containing muons, using a kinematic variable that is independent from the b-tagging algorithms.

To evaluate b-tagging performance in data, this measurement relies on a discriminatory variable, $p_{\mathrm{T}}^{\mathrm{rel}}$, illustrated in Figure 1. In semileptonic b-hadron decays, the transverse momentum p_{T} of the lepton relative to the jet+lepton axis, $p_{\mathrm{T}}^{\mathrm{rel}}$, is larger than in light or charm hadron decays.

When a b-hadron decays semileptonically ($b \to \ell\nu_\ell q$), the maximum energy of the resulting lepton in the rest frame of the hadron is slightly less than half the b-quark mass. $p_{\mathrm{T}}^{\mathrm{rel}}$ is a direct measurement of this energy spectrum (ranging from the muon mass to half the b-quark mass). For semileptonic decays of non-b quarks, this distribution will be contained to a much smaller range (due to the smaller mass of the parent quark). In essense, the range and peak of the $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution is directly determined by the mass of the parent quark.

Because of this, $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions can be used to determine the fraction of jets originating from b-quark decays of a semileptonic sample. Calibration measurements performed with the $p_{\mathrm{T}}^{\mathrm{rel}}$ method using Run 1 ATLAS data are documented in [4].

The $p_{\mathrm{T}}^{\mathrm{rel}}$ of a muon in a jet, defined as its momentum component transverse to the jet+muon axis $\vec{p}_{jet}$, can be written as follows:

$$p_{\mathrm{T}}^{\mathrm{rel}} = \sqrt{\vec{p}_{\mu}^{\,2} - \left(\frac{\vec{p}_{\mathrm{jet}} \cdot \vec{p}_{\mu}}{|\vec{p}_{\mathrm{jet}}|}\right)^2}. \tag{1}$$

The $p_{\mathrm{T}}^{\mathrm{rel}}$ templates are produced for b, c, and light jets using Monte Carlo (MC) simulations (b and c) or by a data-driven approach (light). These templates are then fitted to the $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions of a sample of jets satisfying the b-tagging criterion and a sample of jets failing the b-tagging criterion. The fit determines the fraction of b-jets in each sample and hence the b-tagging efficiency as follows:

$$\varepsilon_b = \frac{N_b^{\mathrm{tagged}}}{N_b^{\mathrm{tagged}} + N_b^{\mathrm{untagged}}} \tag{2}$$

Here the b-tagging efficiency ε_b is defined using N_b^{tagged}, the number of b-jets tagged by the algorithm, and N_b^{untagged}, the number of b-jets that the algorithm rejected. The b-tagging efficiency in data can then be compared to the efficiency in MC calculated at truth level (by directly counting the number of tagged and untagged jets labeled with true b-hadron parents as defined in Section 2.4). The $p_{\text{T}}^{\text{rel}}$ method is further detailed in Section 3.

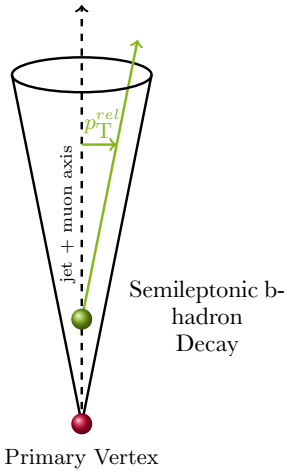


Figure 1: Schematic drawing showing a jet cone in black with a b-hadron in the jet, decaying semileptonically at a secondary vertex shown in green. The final state muon is shown in green as well as the projection of its momentum transverse to the jet+muon axis, the $p_{\text{T}}^{\text{rel}}$.

This measurement is performed on a sample of events containing jets with a reconstructed muon inside, collected with the ATLAS detector [5] in 2016 at a center of mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 24.5 fb^{-1}.

The MV2 classifier family [6] is a set of gradient-boosted decision tree algorithms used to identify jets as either originating from a b-quark or not. These algorithms use a set of 24 jet flavour-sensitive variables, constructed using basic b-tagging algorithms, as inputs. MV2c10, the algorithm studied here, was trained on a signal sample of b-jets with the background consisting of 7% c-jets and 93% light jets. MV2c10 is widely used in ATLAS during run 2 data taking[1]. The MV2c10 algorithm outputs a number between -1 and 1, with larger values corresponding to a higher probability of a jet to originate from a b-quark. For this b-tagging algorithm a set of “operating points" is defined, which are cut values placed on the output corresponding to a nominal b-tagging efficiency $\varepsilon_b^{\text{nom}}$ determind in a simulated $t\bar{t}$ sample. Four working points, 60%, 70%, 77% and 85% of the MV2c10 algorithm are calibrated in this note. Other algorithms such as SV1, a vertex-based algorithm, and IP3D, an impact parameter-based algorithm [3], are not calibrated here; these algorithms are used, however, in the event selections.

To calibrate the b-tagging efficiency ε_b of any classifier, two basic values need to be determined: $\varepsilon_b^{\text{MC}}$, the tagging efficiency in MC, and $\varepsilon_b^{\text{data}}$, the tagging efficiency on data, as well as the uncertainties on these values. The ratio of these two values is referred to as the *scale factor* κ_b (3):

$$\kappa_b = \frac{\varepsilon_b^{\text{data}}}{\varepsilon_b^{\text{MC}}} \tag{3}$$

[1] Run 2 is referring to the ATLAS data taking period starting in 2015 with a center of mass energy of 13 TeV

These scale factors are used to correct the b-tagging efficiency in the simulation. Because b-tagging performance depends strongly on the transverse momentum of a jet ($p_{\mathrm{T}}^{\mathrm{jet}}$), the scale factors are derived in bins of jet p_{T}. The bins used are: 20-30 GeV, 30-40 GeV, 40-50 GeV, 50-70 GeV, 70-90 GeV, 90-110 GeV, 110-140 GeV, 140-170 GeV, and 170-200 GeV. The choice of binning is further explained in Section 2.2.

2 Event and Object Selections

2.1 The ATLAS Detector

The ATLAS experiment [5] at the LHC is a multi-purpose particle detector with a forward-backward symmetric cylindrical geometry and a near 4π coverage in solid angle[2]. It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer. The inner tracking detector covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon micro-strip, and transition radiation tracking detectors. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A hadron (iron/scintillator-tile) calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The end-cap and forward regions are instrumented with LAr calorimeters for both EM and hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer surrounds the calorimeters and is based on three large air-core toroid superconducting magnets with eight coils each. Its bending power is in the range from 2.0 to 7.5 T m. It includes a system of precision tracking chambers and fast detectors for triggering. A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to reduce the accepted rate to at most 100 kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz on average depending on the data-taking conditions.

2.2 Trigger Selection

Muon-in-jet triggers are used to select multi-jet events enriched with semileptonic b-decays. These triggers collect jets passing a certain p_{T} threshold x with a spatially-matched muon passing a p_{T} threshold y. The spatial matching is made within a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$ around the jet axis, where η is pseudorapidity and ϕ is azimuthal angle in the plane transverse to the beam axis. A $\Delta z < 2$ mm is also required, where Δz is the distance of closest approach along the beam axis of the muon track to the primary vertex.

Due to the steeply falling jet p_{T} spectrum, different muon-in-jet triggers were used in different $p_{\mathrm{T}}^{\mathrm{jet}}$ regions. In each bin, one trigger is chosen by maximizing the yield while ensuring that the trigger efficiency shows no dependence on $p_{\mathrm{T}}^{\mathrm{jet}}$ across the entire bin.

[2] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

The triggers selected in this analysis are prescaled, i.e. only some fraction of events fulfilling the muon-in-jet trigger requirement are recorded. The effective luminosity collected by each trigger before selection is detailed in Table 1.

Table 1: Summary of $p_{\mathrm{T}}^{\mathrm{jet}}$ bins and the thresholds applied in the corresponding triggers.

Jet p_{T} Bin	Jet p_{T} Threshold	Muon p_{T} Threshold	Effective Luminosity
[20 GeV, 30 GeV], [30 GeV, 40 GeV]	15 GeV	4 GeV	6.41 pb^{-1}
[40 GeV, 50 GeV]	25 GeV	4 GeV	7.33 pb^{-1}
[50 GeV, 70 GeV]	35 GeV	4 GeV	9.26 pb^{-1}
[70 GeV, 90 GeV], [90 GeV, 110 GeV]	55 GeV	4 GeV	55.78 pb^{-1}
[110 GeV, 140 GeV]	85 GeV	6 GeV	376.00 pb^{-1}
[140 GeV, 170 GeV]	110 GeV	6 GeV	688.50 pb^{-1}
[170 GeV, 200 GeV]	150 GeV	6 GeV	2113.90 pb^{-1}

2.3 Simulation Samples

To produce the b- and c-jet $p_{\mathrm{T}}^{\mathrm{rel}}$ templates, a Pythia 8 dijet muon-filtered MC sample is used. This sample was produced using A14 Pythia8 tune with the NNPDF23LO set of PDFs[7] [8]. In total, 8 million events were used.

The muon filtering requirement is applied at generator level, so muons originating from processes simulated after event generation are not considered. This biases the relative amount of heavy and light flavour jets in the filtered sample. In addition to the muon-filtered MC, an unfiltered dijet Pythia 8 NNPDF23LO sample is also used. The unfiltered MC, which does not suffer from this bias, is used to determine the initial values for the b-fraction and light-to-charm ratio for the fitter and for calculating the true b-tagging efficiency. The unfiltered MC sample contains 8 million events.

The EvtGen simulation package [9] is used to improve the modelling of the decays of b- and c- hadrons. All MC samples are simulated with the full ATLAS detector simulation based on the Geant 4 toolkit [10]. The impacts from additional interactions (pileup) are taken into account by overlaying the minimum bias events simulated by Pythia 8.

2.4 Jet Flavour Labeling

Templates for the heavy flavour jets are generated from muon-filtered MC samples, as described in Section 2.3. The flavour of the jet is determined by a geometrical truth matching. The jet is iteratively checked for the existence of b-hadrons, c-hadrons or τ-leptons within a cone of $\Delta R < 0.3$. The b-hadrons and c-hadrons are required to have $p_{\mathrm{T}} > 5\,\mathrm{GeV}$. The truth record is first checked for any b-hadrons and labelled accordingly if one is found. If no b-hadrons are present, a similar check is done for c-hadrons. If neither b- or c-hadrons are identified in the truth record, the jet is labeled as light.

2.5 Event and Object Selection

To select semileptonic b-decays, the event is required to contain at least one muon with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$.

Jets are reconstructed from clusters of topologically connected calorimeter cells, which are then reconstructed into jets by the anti-k_t algorithm with $R = 0.4$ [11]. The jets must have $|\eta| < 2.5$ and satisfy a Jet Vertex Tagger (JVT) criterion which reduces the number of pileup jets by a factor of $\sim 10 - 100$ (depending on jet p_{T}) while retaining 92% of true hard scatter collisions [12]. In addition to the online trigger selection, further offline requirements are placed on the event topology. In all the p_{T} bins, at least two jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are selected to match a b-quark pair production topology.

Since muons inside more energetic jets have a harder p_{T} spectrum, the muon p_{T} threshold requirement depends on the jet p_{T} bin considered. In addition, the muons are required to satisfy $|\eta| < 2.5$, $d_0 < 2\,\mathrm{mm}$ and $|z_0 \cdot \sin\theta| < 4\,\mathrm{mm}$, where d_0 and z_0 and the transverse and longitudinal impact parameters, respectively[3]. This ensures that the muon is consistent with coming from the primary vertex of the event; the cut values were chosen for consistency with the $t\bar{t}$ based b-tagging calibrations. Only muons satisfying the “tight" selection criteria [13], i.e. muons measured in the inner tracking system as well as the outer muon spectrometer, passing the most stringent hit requirements, are used.

The event selection is summarized in Table 2.

Table 2: List of selection criteria for muon-in-jet samples

Jet Selections	$p_{\mathrm{T}} > 20$ GeV $\lvert\eta\rvert \leq 2.5$ JVT Cut
Muon Selections	$\lvert\eta\rvert \leq 2.5$ $\lvert d_0\rvert < 2$ mm $\lvert z_0 \cdot \sin\theta\rvert < 4$ mm Tight Quality
Muon p_T Requirement	5 GeV (if $p_T^{Jet} < 40$ GeV) 6 GeV (if $40 < p_T^{Jet} < 90$ GeV) 8 GeV (if $p_T^{Jet} > 90$ GeV)
Muon-Jet Matching	$\Delta R < 0.3$

2.6 Template Selections

The dominant production mechanisms of b-quarks in pp collisions produce $b\bar{b}$ pairs. If one jet in the event is identified as a b-jet, this enhances the probability that there is another b-jet in the event. To improve the b-jet purity of the jet sample used for the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration, the following procedure is applied:

- Require at least one jet to be identified as a b-jet by the MV2c10 algorithm at 85% working point

[3] Transverse impact parameter is defined as the distance of closest approach to the beam line, and longitudinal impact parameter is the value of z at the point of closest approach.

  - If multiple candidate b-jets are identified, preferentially choose the jet without a matched muon as the tag b-jet.
  - If multiple jets without a matched muon are tagged, or if all tagged jets contain muons, select the tag b-jet from the candidates randomly.
- Remove the tag b-jet to avoid bias and perform the analysis using the remaining jets.

This enhancement procedure is applied to the data selected for the measurement. Such an enhancement is not necessary in heavy flavour MC, because the sample purity is already quite high.

The light-jet template is generated with the events that pass the same trigger and selections but fail a high-efficiency MV2c10 cut (MV2c10 > -0.8). The true $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution shape is not correlated with MV2c10 score (see Figure 4(b)), so this requirement is not expected to bias the template shape. This creates a disjoint set of light-jet enriched events with low b-jet contamination. The heavy flavour (b- and c-jet) contamination is estimated using the unfiltered simulation sample. This contamination is corrected and assigned an uncertainty as discussed in Section 4.

3 The $p_{\mathrm{T}}^{\mathrm{rel}}$ Method

To calculate the b-tagging efficiency ε_b for a given tagging algorithm using Equation (2), two quantities need to be determined: N_b^{tagged}, the number of b-jets tagged by the algorithm, and N_b^{untagged}, the number of b-jets that the algorithm failed to tag. In MC, these quantities can be directly retrieved using truth level information (as described in Section 2.4). In data, these quantities need to be measured independently for each tagging criterion under study.

A template fit method is applied to measure the flavour fractions and hence the tagging efficiency in data. The discriminating variable used to perform the template fit is $p_{\mathrm{T}}^{\mathrm{rel}}$, as described below; hence this method is referred to as the $p_{\mathrm{T}}^{\mathrm{rel}}$ method.

3.1 Discriminating Variable: $p_{\mathrm{T}}^{\mathrm{rel}}$

Given the larger masses of b-hadrons relative to hadrons containing only charm or light quarks, their decay products are more energetic. Therefore, the final state particles have larger momentum in the rest frame of the decaying hadron, hereafter referred to as p^*, compared to the decays of less massive hadrons. Figure 2 shows the p^* of muons produced in the decays of B^0, B_s^0, D^0 mesons and Λ_b^0 baryons. Boosted into the laboratory frame, the larger p^* values results in a harder $p_{\mathrm{T}}^{\mathrm{rel}}$ spectrum for muons from b-hadrons compared to those from lighter hadrons.

The $p_{\mathrm{T}}^{\mathrm{rel}}$ of a muon in a jet is defined as its momentum component transverse to the jet+muon axis $\vec{p}_{jet}$ and is written in Equation 1.

Figure 1 illustrates the $p_{\mathrm{T}}^{\mathrm{rel}}$ of a muon within a jet cone. The jet+muon axis, depicted as a dashed line, is computed by first removing the energy corresponding to the minimal ionizing muon within the calorimeter and next adding the muon four-momentum to the jet. The $p_{\mathrm{T}}^{\mathrm{rel}}$ variable is not used as an input to MV2, so any correlation between $p_{\mathrm{T}}^{\mathrm{rel}}$ and the input variables to MV2 is not exploited by the algorithm.

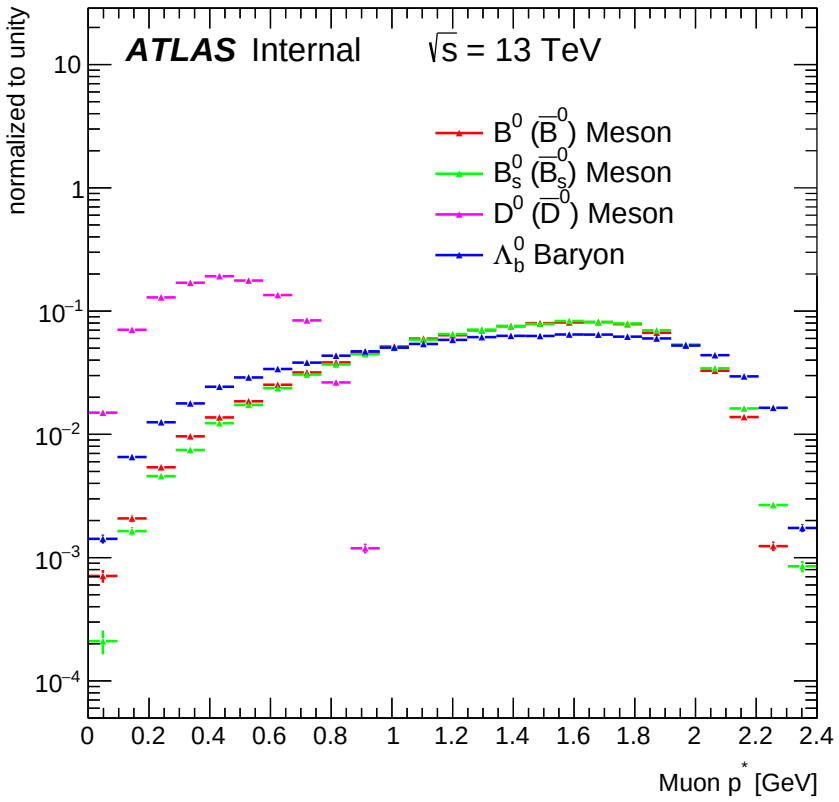


Figure 2: p^* distributions of final state muons from semileptonically decaying B^0, B_s^0, D^0 mesons and Λ_b^0 baryons in red, green, purple and blue, respectively. Due to the lower D^0-mass the respective final state muon has less energy and hence a smaller p^*. Only muons from direct b-hadron decays are included.

When studying semileptonic b-decays, this analysis focuses on final states containing a muon instead of an electron since muons can be reconstructed with higher efficiency and better resolutions.

3.2 Template Fit

Templates for the $p_{\mathrm{T}}^{\mathrm{rel}}$ spectrum of heavy flavour jets are generated from muon-filtered MC samples, described in Section 2.3. All the selection criteria discussed in Section 2 are applied. The procedure described in Section 2.4 is used to select b- and c-jets.

The light-jet template is constructed via a data-driven approach by selecting jets which fail a high-efficiency MV2c10 cut, as described in Section 2. After correcting for the expected heavy flavour contamination using MC prediction, ranging from 0-5% (15-30%) for $b(c)$-jet contaminations depending on jet p_{T} bin, the light-jet and c-jet templates are combined to a non-b-template. Figure 3 shows the normalized $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions of muons for the three jet flavours. The b-template shows a harder $p_{\mathrm{T}}^{\mathrm{rel}}$ spectrum compared to the c- and light-jet templates.

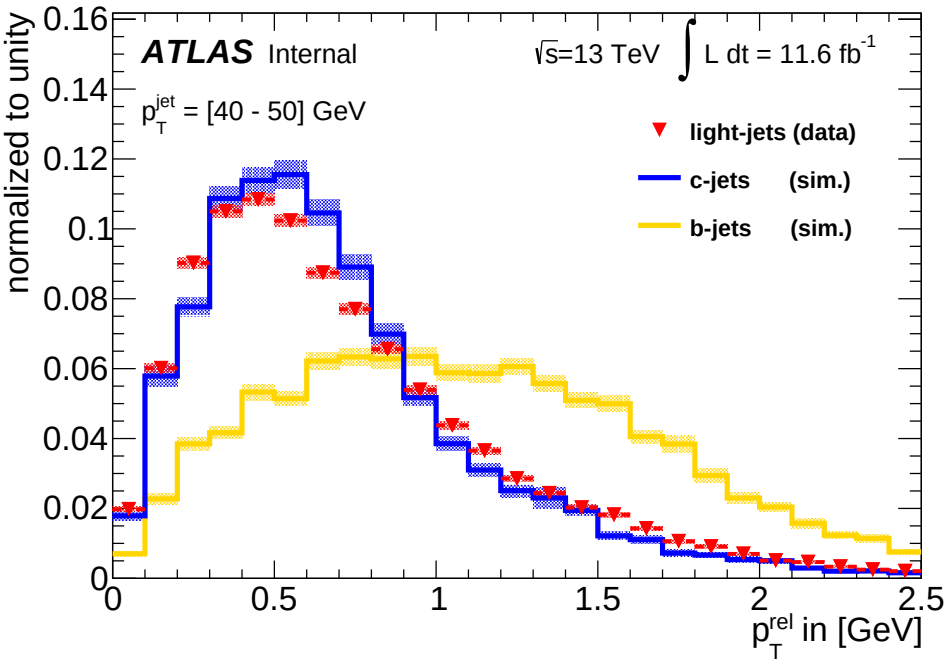


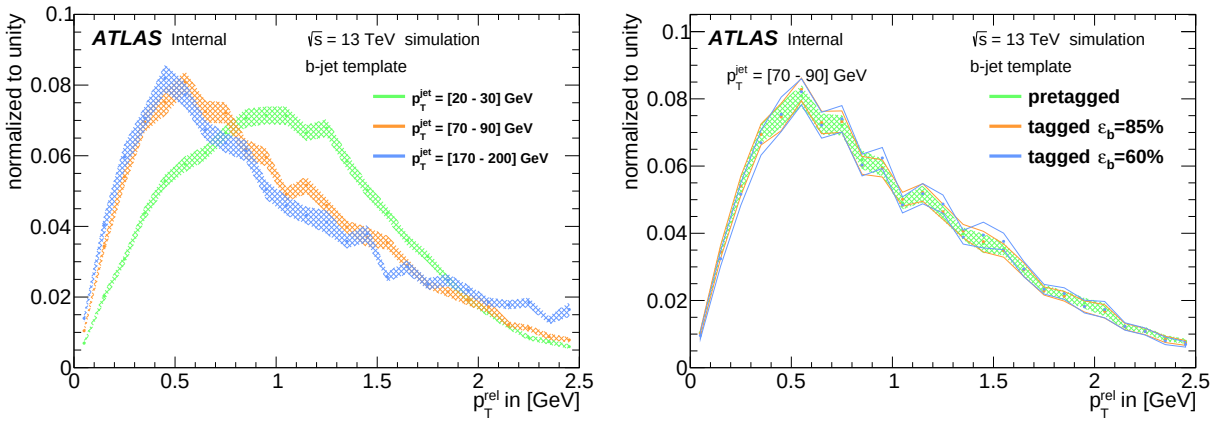
Figure 3: Muon $p_{\mathrm{T}}^{\mathrm{rel}}$spectra for the b-, c- and light-jet templates in yellow, blue and red, respectively. The heavy flavour templates are generated from simulation while the light-jet template is extracted in a data-driven approach.

The templates exhibit jet p_{T} dependence, as demonstrated in Figure 4(a). Here, the b-jet template is shown for three different $p_{\mathrm{T}}^{\mathrm{jet}}$ bins in different colors. A possible cause of the dependence seen here is the effect of cascade decays ($b \to c \to \mu + X$) and direct decays ($b \to \mu + X$) on $p_{\mathrm{T}}^{\mathrm{rel}}$ template shape. The muon p_{T} is softer in cascade decays because it originates from a c-quark instead of a b-quark. Hence, cascade decays will have a softer $p_{\mathrm{T}}^{\mathrm{rel}}$ template shape than direct decays. In Figure 4(a), the lowest $p_{\mathrm{T}}^{\mathrm{jet}}$ slice appears to have a harder $p_{\mathrm{T}}^{\mathrm{rel}}$ spectrum; this is because cascade decays in this slice are less likely to pass the muon p_{T} requirement. In the higher $p_{\mathrm{T}}^{\mathrm{jet}}$ slices, both direct and cascade decays pass the selection, so the $p_{\mathrm{T}}^{\mathrm{rel}}$ template is softer. Another contributing factor is that at high p_{T}, $p_{\mathrm{T}}^{\mathrm{rel}}$ becomes small relative to $p_{\mathrm{T}}^{\mathrm{jet}}$, making it more difficult to accurately reconstruct. This causes all three $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions (b, c, and light) to look more similar at high p_{T}, possibly resulting in the change in template shape observed.

Because of this $p_{\mathrm{T}}^{\mathrm{jet}}$ dependence, templates are generated with the same $p_{\mathrm{T}}^{\mathrm{jet}}$ binning as the data distributions when performing the fitting procedure. The template shapes show no dependences on the applied MV2 tagging requirement, as can be seen in Figure 4(b). The “pretagged" distribution refers to the sample before any tagging decision is applied. Since no dependence is observed, the templates are generated without applying any tagging requirement.

The overflow bin is not considered during fitting. The long tails in the $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution cause the overflow bin to become the most statistically significant bin. In this case, any attempt to fit templates to the data distribution is entirely driven by the overflow bin, ignoring differences in the distribution in the hadron mass range. Therefore, the $p_{\mathrm{T}}^{\mathrm{rel}}$ range of the templates is confined to 0-2.5 GeV.

Using the b-, c- and light-flavour templates, a log-likelihood fit is performed to the tagged and untagged



(a) Dependence on $p_{\mathrm{T}}^{\mathrm{jet}}$.

(b) Dependence on the tagger decision.

Figure 4: The dependence of the $p_{\mathrm{T}}^{\mathrm{rel}}$ template shape on the jet p_{T} and the decision of the tagging algorithm are shown on the left and right, respectively. On the left the b-jet template is shown for three different $p_{\mathrm{T}}^{\mathrm{jet}}$-bins in green, blue and orange with the statistical uncertainties depicted as filled error bands. A clear dependence on $p_{\mathrm{T}}^{\mathrm{jet}}$ is observed. On the right the b-jet template as used in the analysis is shown in green with the statistical uncertainty depicted by the filled error band. For comparison, the $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions for jets passing a loose or tight working point of the MV2c10 algorithm are shown in orange and blue. The spectra agree within the statistical precision, indicating no dependence on the tagging decision.

data distributions as shown in Figure 5. The fit to the tagged $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution is shown in Figure 5(a) while the untagged fit result is shown in Figure 5(b). The selected data distribution is shown as black dots. The b-fraction and light-to-charm ratio, both determined by the fitter, are used to determine the relative contribution of each flavour, visualized with the stacked templates.

3.3 Efficiency Extraction

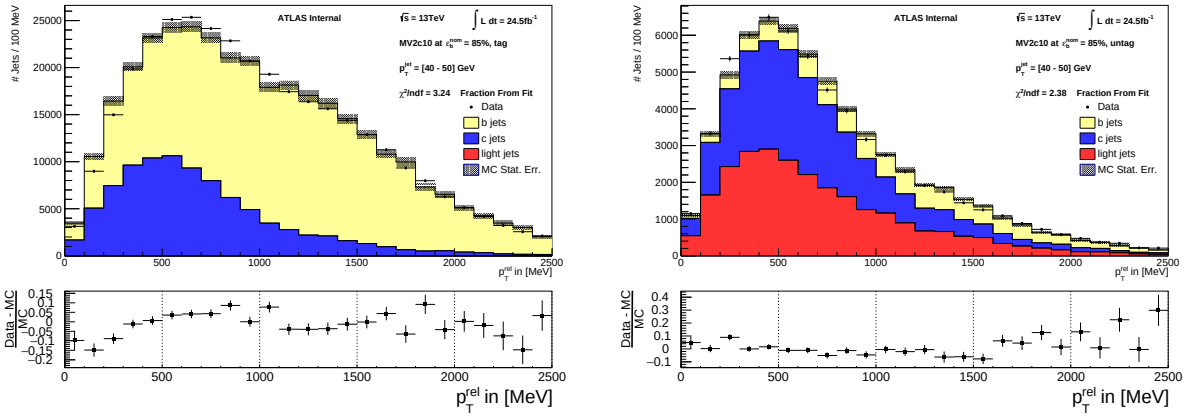
The $p_{\mathrm{T}}^{\mathrm{rel}}$ fit is used to determine the flavour composition of the QCD-dominated phase-space. With the resulting b-fractions f_b the tagging efficiency defined in Equation (2) can be calculated as

$$\varepsilon_b^{\mathrm{data}} = \frac{f_b^{\mathrm{tagged}} \cdot N_{\mathrm{data}}^{\mathrm{tagged}}}{f_b^{\mathrm{tagged}} \cdot N_{\mathrm{data}}^{\mathrm{tagged}} + f_b^{\mathrm{untagged}} \cdot N_{\mathrm{data}}^{\mathrm{untagged}}}, \tag{4}$$

where N_{data} is the number of jets in the respective data sample. This procedure is repeated for all the $p_{\mathrm{T}}^{\mathrm{jet}}$ regions defined before. Comparing the resulting efficiencies to the predictions from the unfiltered simulation, according to Equation (3), scale factors κ_b are calculated.

4 Systemetic Uncertainties

Multiple sources of systematic uncertainties have been considered for this measurement. These systematics can be broadly categorized into detector and calibration uncertainties, physics modeling uncertainties, MC statistical uncertainties, and template selection uncertainties.



(a) Tagged fit result.

(b) Untagged fit result.

Figure 5: Showing the agreement of the data distribution after the $p_{\mathrm{T}}^{\mathrm{rel}}$ fit. The distribution on the left is tagged by MV2c10 at $\varepsilon_b^{\mathrm{nom}} = 85\%$ while the distribution on the right is untagged. The data is shown in black while the templates are shown as a stack.

In general, two kinds of systematics are used. Two-sided systematics are evaluated by doing an up- and downwards variation of the parameter under study and propagating the resulting effects to the scale factors. The systematic is then assumed to be symmetric and half the difference between the two variations is taken as a systematic. Single-sided systematics are evaluated by doing a variation and using the full difference to the nominal value as the uncertainty. The total systematic uncertainty is calculated from the sum in quadrature of all systematic uncertainties. Tables 4,5,6,7 summarize the systematic uncertainties for all the WPs.

4.1 Detector and Calibration Uncertainties

This analysis accounts for several uncertainties related to detector resolution and performance/calibration for jets and muons.

The jet energy scale calibration includes several MC-based, *in-situ*-validated corrections for pile-up effects: per-event area-based pile-up removal, calculated using median p_{T} density ρ, and residual p_{T} dependent N_{PV} and μ corrections, where N_{PV} is the number of reconstructed primary vertices and μ is the mean number of interactions per bunch crossing. The uncertainties on these corrections arise from potential MC mismodeling of N_{PV}, μ, and ρ topology and from p_{T} dependence of N_{PV} and μ terms used in the residual pile-up correction.

Jets originating from gluons, light quarks or heavy quarks differ in hadronization and jet structure and hence in calorimeter response. To cover uncertainty in calibrating these jets, a systematic variation is applied by comparing the calorimeter response for b-jets produced by multiple MC generators. The differences between quark- and gluon-initiated jets in jet composition and calorimeter response are also considered via a similar approach. Finally, differences in response between jets in forward and central η regions are handled similarly and validated in an *in-situ* tag-and-probe approach using dijet events [14].

Other jet uncertainties have been combined into nuisance parameters (NPs). 67 NPs are determined in *in-situ* Z/γ+jet and multi-jet p_{T} balance estimates and are then reduced to a set of six uncertainties via an eigen-decomposition. Five of these NPs constitute the greatest-magnitude principal components, and the remaining components are combined together quadratically into a single NP. Uncertainties related to jet energy resolution are combined into a single NP [14].

For the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration only combined (CB) reconstructed muon tracks were used. CB tracks are first reconstructed independently in the inner detector (ID) and muon spectrometer (MS), then combined together with a global refit using hits from both subdetectors. In this measurement, uncertainty in muon track reconstruction is included by applying a Gaussian smearing of muon tracks in both the ID and MS using the respective precision $\pm 1\sigma_{\mathrm{ID,MS}}$. In addition, the muon momentum scale in simulation is calibrated to data by studying $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ decays. The systematic and statistical uncertainties on this correction are propagated throughout the presented measurement and their impact on the SFs is calculated [13].

4.2 Modeling Uncertainties

The physics modeling uncertainties accounted for in this analysis have been grouped into those pertaining to muons and to jets.

4.2.1 Muon Uncertainties

The muons studied in this analysis originate both from direct $b \to \mu + X$ decays and from cascade $b \to c \to \mu + X$ decays, which result in a lower muon $p_{\mathrm{T}}^{\mathrm{rel}}$. The branching fractions of both decay modes have been measured and their ratio has been found to be $BR(b \to \ell X)/BR(b \to c \to \ell X) = 1.11 \pm 0.07$ [3, 15]. In the simulated samples, jets with muons originating from cascade decays have been scaled up or down corresponding to the variation of the branching ratios within the uncertainty. The difference in the results is taken as an up-down systematic.

The momentum of the muon in the b-hadron rest-frame p^* is the key distribution for this measurement. While its modeling is crucial, measurements of this distribution provided by the BaBar and Delphi collaborations [16, 17] do not fully agree within their uncertainties. The p^* spectrum of muons has been studied for the simulations discussed in Section 2.3. It has been found that the simulation agrees with the Delphi results. To estimate the impact of this limited knowledge, the simulation has been reweighted to match the BaBar distribution and the uncertainties have been propagated through the analysis.

In order to reduce contributions from fake muons, tight quality criteria are required in the analysis. Despite this, misidentified particles or decay-in-flight muons can still make the selection. Due to the muon filter on the MC simulation, the amount of these fake muons is underestimated in the simulation. To estimate the effect of additional fake muons, their amount has been increased in the simulation by reweighting jets with a fake muon by a factor of two and propagating the impact to the resulting scale factors.

4.2.2 Jet and Hadron Uncertainties

The muon $p_{\mathrm{T}}^{\mathrm{rel}}$ depends on the knowledge of the relative directions of the muon and the b-hadron. While the muon direction is measured with high precision, two aspects influence the knowledge of the b-hadron direction. First, the difference between the b-hadron flight direction and the reconstructed jet direction is studied in simulation. Second, the precision with which the jet direction is reconstructed is estimated by comparing calorimeter and track jets. These systematic effect have been found to be 0.004 and 0.008 in ϕ and η, respectively [3]. In order to estimate a jet angular resolution uncertainty, the jet direction has been varied within a Gaussian distribution with the width set to the respective values. These jets are then used to redo the measurement and the difference is taken as the systematic uncertainty.

Energetic gluon decays may result in jets with two or more heavy flavour hadrons inside the jet cone due to gluon-splitting. These jets have a different tagging efficiency, different probability to contain a muon and a different $p_{\mathrm{T}}^{\mathrm{rel}}$. Therefore, their modeling has a large impact on this measurement. The weight of events with two or more heavy flavour hadrons is scaled up or down by 50% to estimate conservatively the impact of this uncertainty.

The measurement depends on the modeling of the final state particles in the b-hadron rest-frame. Therefore, the fractions of hadrons (B^0, B^+, B_s, b-baryons) produced during the fragmentation process have to be well modeled. This modeling is limited by how precisely these fractions have been measured. For this measurement the fragmentation fractions have been set to the estimates by the HFAG [18]. To evaluate a systematic uncertainty the events have been reweighted to match flavour fractions measured by the CDF collaboration [19]. Table 3 compares the simulation and inclusive measurements.

Table 3: Fractions of b-hadrons as produced in simulated Z^0-decays, inclusive measurements performed at CDF [19], and compared to the average provided by the HFAG [18].

Hadron Fraction	Z^0-decays (MC)	Tevatron	HFAG-combination
B^+ fraction	0.410 ± 0.007	0.350 ± 0.020	0.406 ± 0.005
B^0 fraction	0.410 ± 0.007	0.350 ± 0.020	0.406 ± 0.005
B_s^+ fraction	0.100 ± 0.008	0.100 ± 0.010	0.105 ± 0.005
b-baryon fraction	0.080 ± 0.010	0.199 ± 0.044	0.083 ± 0.010

During the fragmentation process, the energy of the initial b-quark is passed on to the generated hadrons. The energy fraction X_b that is transferred to the b-hadron is altered by 5% in an up-down variation to estimate conservatively the systematic uncertainty due to the limited knowledge of the hadronization process.

For the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration it is a necessity to have a muon inside the jet cone, which is usually the case only for semileptonically decaying b-hadrons. The tagging efficiency on this sample, however, differs with respect to the inclusive b-tagging efficiency. This effect will cancel out in the calculation of the data-to-MC SFs as long as it is well modeled in the simulation. Possibile bias due to differences in the modeling quality between the semilpetonic and inclusive decays are found to be negligible in early ATLAS data. This measurement as function of the jet p_T was done on a set of $t\bar{t}$ events by comparing SFs determined in hadronic and leptonic top decays. A precision varying from 12% to 2% depending on the jet p_T and the b-tagging working point was reached. The full procedure is described in Appendix F.

4.3 MC Statistical Uncertainties

Two sources of MC statistical uncertainty have to be taken into account in this analysis. The simulation statistical uncertainty represents statistical limitations on the number of events in Equation (2) and is propagated to the simulation efficiency and the scale factors. The template statistical uncertainty covers changes in the template shapes due to statistical fluctuations. To evaluate this, ten thousand pseudo-experiments were performed to generate new templates according to a bin-by-bin Gaussian variation whose mean value and standard deviation are set to the numbers extracted from the nominal templates. The width of the resulting efficiency distribution is then used as the systematic uncertainty.

4.4 Template Selection Uncertainties

Since the light template is generated from data it is clear that there will still be some remaining heavy-flavour jets after the selection is applied. The amount of heavy flavour contamination is estimated from the unfiltered simulation and corrected for in the analysis. To account for the limited knowledge of this value, the flavour contamination fractions are varied up and down by one standard deviation and the scale factor is recalculated with the resulting numbers.

The Jet Vertex Tagger (JVT) [20] is used to reject pile-up jets. It has been checked that uncertainty on the estimated efficiency of the algorithm has no effect on the $p_{\mathrm{T}}^{\mathrm{rel}}$ analysis. Since this algorithm rejects jets with specific kinematic properties the choice of the JVT working point may have an impact on the measurement. To account for this, the JVT criterion has been exchanged for a loose and tight criterion when generating the templates. Upon alteration neither the selected data events nor the MC truth efficiencies are changed. Only the impact on the template shapes is used to re-evaluate the measured efficiencies.

Table 4: Systematic uncertainties for the MV2c10 tagging algorithm at 60% nominal *b*-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties	1.09	1.37	1.70	1.33	1.26	4.27	2.44	4.62	3.37
Modeling Uncertainties	2.92	3.35	3.40	3.79	6.79	5.86	5.54	5.94	11.98
MC Statistical Uncertainties	2.82	2.17	2.52	1.09	1.98	2.45	2.56	3.93	5.67
Template Selection Uncertainties	0.10	<0.01	<0.01	<0.01	0.02	0.01	0.11	0.08	0.33
Data Statistical Uncertainty	0.15	0.14	0.16	0.18	0.24	<0.01	0.36	0.56	0.73
Total Systematic Uncertainty	4.21	4.22	4.57	4.15	7.19	7.66	6.57	8.49	13.60

Table 5: Systematic uncertainties for the MV2c10 tagging algorithm at 70% nominal *b*-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties	0.86	0.87	0.57	0.93	1.04	3.04	1.87	2.88	2.36
Modeling Uncertainties	1.43	1.46	1.94	2.57	4.41	4.00	3.41	3.99	7.93
MC Statistical Uncertainties	1.91	1.38	0.71	0.78	1.45	1.92	1.90	2.90	4.16
Template Selection Uncertainties	0.17	<0.01	<0.01	<0.01	0.03	0.11	0.16	0.10	0.39
Data Statistical Uncertainty	0.17	0.15	0.17	0.19	0.25	0.60	0.39	0.67	0.78
Total Systematic Uncertainty	2.54	2.20	2.14	2.84	4.76	5.39	4.34	5.71	9.27

Table 6: Systematic uncertainties for the MV2c10 tagging algorithm at 77% nominal b-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[\,GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties	0.29	0.44	0.54	0.25	0.77	2.05	1.40	1.59	1.85
Modeling Uncertainties	0.93	1.02	1.05	1.26	2.42	2.28	1.82	2.65	4.92
MC Statistical Uncertainties	1.36	1.30	0.50	0.57	0.96	1.37	1.38	1.89	2.67
Template Selection Uncertainties	0.23	0.10	0.01	<0.01	0.04	0.17	0.19	0.11	0.41
Data Statistical Uncertainty	0.15	0.13	0.17	0.18	0.23	0.57	0.39	0.67	0.76
Total Systematic Uncertainty	1.69	1.72	1.29	1.41	2.72	3.36	2.69	3.62	5.91

Table 7: Systematic uncertainties for the MV2c10 tagging algorithm at 85% nominal b-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[\,GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties	0.06	0.06	0.04	0.05	0.09	0.25	0.07	0.09	0.27
Modeling Uncertainties	0.70	0.70	0.70	0.70	0.71	0.76	0.79	0.96	0.96
MC Statistical Uncertainties	1.34	0.57	0.63	0.58	1.38	1.33	0.21	0.33	0.57
Template Selection Uncertainties	0.08	0.14	<0.01	0.01	0.24	0.24	0.38	0.39	1.31
Data Statistical Uncertainty	0.10	0.09	0.11	0.12	0.10	0.40	0.11	0.16	0.16
Total Systematic Uncertainty	1.52	0.92	0.95	0.92	1.58	1.57	0.91	1.09	1.75

5 Results

As discussed in Section 3 the b-tagging efficiency is computed by performing a template fit to the tagged and untagged $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions. The procedure is repeated for all jet p_{T} bins and for four operating points for the algorithm MV2c10. The $p_{\mathrm{T}}^{\mathrm{jet}}$ dependent b-tagging efficiency ε_b in the selected data sample measured for the MV2c10 algorithm for the nominal efficiencies $\varepsilon_b^{\mathrm{nom}} = 60\%, 70\%, 77\%$, and 85% is shown in Figures 6. Data is shown in the black dots with uncertainty bands depicted in green, while the MC prediction is shown in gray. Dividing the two efficiencies on the left results in the scale factors shown on the right with the respective uncertainties. It can be seen that the SFs are generally consistent with unity, except for the 85% working point. However, a clear trend is visible showing higher SFs in the low $p_{\mathrm{T}}^{\mathrm{jet}}$regions and lower Sfs in the high $p_{\mathrm{T}}^{\mathrm{jet}}$regions, especially for $\varepsilon_b^{\mathrm{nom}} = 60\%$ and 70%.

6 Conclusion

In this note the first measurement of b-tagging efficiencies with the $p_{\mathrm{T}}^{\mathrm{rel}}$ method at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ was presented. Efficiencies have been measured for the MV2c10 for four working points. The results are presented as $p_{\mathrm{T}}^{\mathrm{jet}}$dependent scale factors.

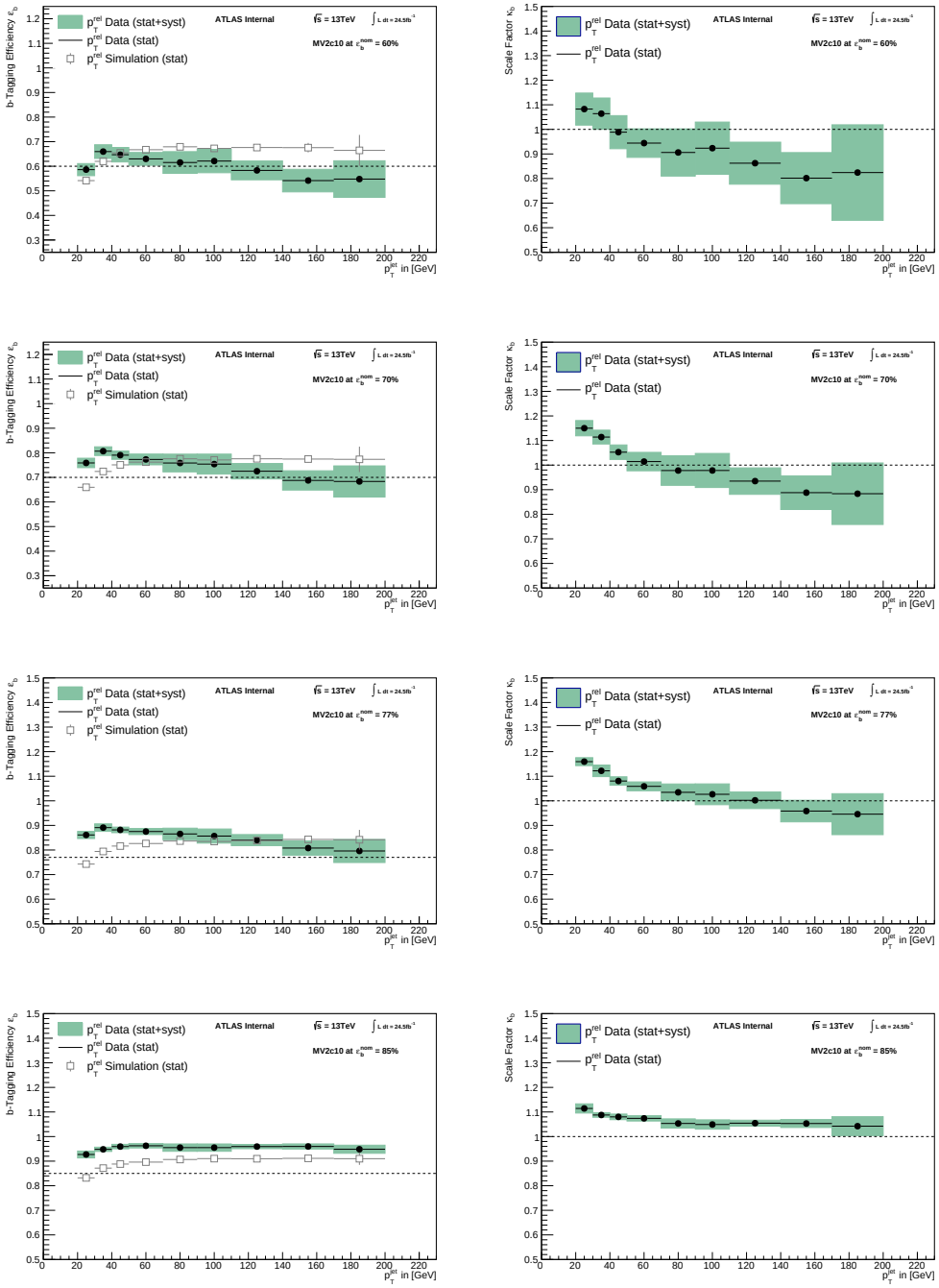


Figure 6: Jet p_{T} dependent efficiencies (left) and scale factors (right) measured by the $p_{\mathrm{T}}^{\mathrm{rel}}$ method for the MV2c10 algorithm for the nominal efficiency of $\varepsilon_b^{\mathrm{nom}}$ = 60%, 70%, 77%, and 85% (top to bottom). MC predictions are shown in gray while the data is depicted in black dots. The statistic and systematic uncertainties are shown in the green error bands.

Appendix

A Data and Simulation Samples Details

Table 8: Unfiltered MC Samples Summary

Slice	Jet p_{T} Range	QCD Process Type	Dataset ID	AMI tags	Number of Events
JZ1W	20 – 60 GeV	Soft (inelastic)	361021	e3569 s2576_s2132 r8459_r7676 p3228	1999000
JZ2W	60 – 160 GeV	Hard	361022	e3668 s2576_s2132 r8459_r7676 p3228	1994600
JZ3W	160 – 400 GeV	Hard	361023	e3668 s2576_s2132 r8459_r7676 p3228	7884500
JZ4W	400 – 800 GeV	Hard	361024	e3668 s2576_s2132 r8459_r7676 p3228	7979800

Table 9: Muon-Filtered MC Samples Summary

Slice	Jet p_{T} Range	QCD Process Type	Dataset ID	AMI tags	Number of Events
JZ0W	0 – 20 GeV	Soft (inelastic)	427000	e3968 s2608_s2183 r7725_r7676 p3228	1993600
JZ1WA	20 – 40 GeV	Soft (inelastic)	427030	e3968 s2608_s2183 r7725_r7676 p3228	996700
JZRW1B	40 – 60 GeV	Hard	427031	e3968 s2608_s2183 r7725_r7676 p3228	997799
JZRW2	60 – 160 GeV	Hard	427032	e3968 s2608_s2183 r7725_r7676 p3228	1989600
JZRW3	160 – 400 GeV	Hard	427033	e3968 s2608_s2183 r7725_r7676 p3228	1991600
JZRW4	400 – 800 GeV	Hard	427034	e3968 s2608_s2183 r7725_r7676 p3228	1995200

The muon-filtered and the unfiltered MC samples are generated in slices of jet momentum[4] to provide the same statistical uncertainty across the whole $p_{\mathrm{T}}^{\mathrm{jet}}$ range. Starting with JZRW1B, hard QCD slices in the muon-filtered sample were generated by taking advantage of internal Pythia8 reweighting. Further details on Pythia8 configuration and reweighting options can be found at [21]. For the unfiltered MC additional truth filtering techniques were used to improve data-MC agreement in the low $p_{\mathrm{T}}^{\mathrm{jet}}$ region.

B Event and Object Selection Details

The trigger efficiency turn-on ratios were calculated using the bootstrap method [22], with the lower p_{T} muon-in-jet triggers as support. For each bin, the trigger with the lowest threshold is picked from

[4] The jet slices are determined on jets clustered on stable particles in the truth record with the anti-k_t algorithm with R=0.6

the available triggers that have reached their full efficiency in that bin. The $p_{\mathrm{T}}^{\mathrm{jet}}$ bin boundaries and the corresponding trigger thresholds are listed in Table 1. Note that the effective luminosities are calculated before selection criteria are applied.

The trigger turn-on behavior has been studied by selecting events based on a lower threshold, fully efficient trigger and compute the $p_{\mathrm{T}}^{\mathrm{jet}}$ dependence efficiency of the trigger under consideration as a function of jet p_{T}. This procedure is not practical for the lowest threshold trigger mu4j15. A p_{T} reweighting scheme has been used to study the impact of p_{T} modeling effects in low $p_{\mathrm{T}}^{\mathrm{jet}}$ regions. This is further discussed in Section C.2.

For the template selection, a selection to optimize the *b*-purity is applied as described in Section 2.6. This selection was studied in order to update the cut values from their Run I values. The most important metric for evaluating this selection is *b*-efficiency. The data sample used for calibration should have increased *b*-efficiency (to provide a purer sample of *b*-jets for calibrating), whereas the data sample used to produce the light template should have low *b*-efficiency (to provide a light template with low *b*-contamination). Similarly, the heavy-flavour contamination is determined from a MC sample using the same cut as the light template sample. Finally, the muon-filtered MC sample does not need a *b*-purity enhancement cut, since the light contamination in the *b*-template is already low without any additional cuts.

In this study, we compared the effect of four different cut values These cut values were chosen to cover a large range of potential cuts. The loosest cut (MV2c10 > -0.8) was chosen to include the peak in number of light jets between -1.0 and -0.8 MV2c10 (see Fig. 7). The tightest cut (MV2c10 > 0.1758) corresponds to the 85% working point. Two other cut values (MV2c10 > -0.3 and 0.0) were chosen to provide coverage in between these extremes.

Naively, it would be “best" to use the tightest cut on the calibration data and unfiltered MC samples and the loosest cut on the light template and heavy-flavour contamination samples. We were concerned, however, that these cuts might hurt our statistics or increase our systematic uncertainties. After careful study, however, we found that these worries were unfounded. Therefore, we chose this “best" combination of *b*-purity optimization tags.

C Low Jet p_{T} Region Studies

As shown in Table 1, each jet p_{T} bin applies a trigger whose threshold is well below the lower edge of that p_{T} bin so that the trigger is fully efficient. Ensuring that the event selections operate on the trigger efficiency plateau is even more crucial to this analysis as the MC samples used do not have triggers simulated (due to a production-level trigger calculation issue).

However, the trigger applied in the two lowest jet p_{T} bins, HLT_mu4j15, has a very slow turn-on so that requiring $p_{\mathrm{T}}^{\mathrm{jet}} \geq 20$ GeV is not optimal. Unfortunately HLT_mu4j15 is the trigger with the lowest threshold that can collect enough data for this analysis. Due to the above facts, the jet p_{T} spectra in the $[\,20, 30]$ p_{T} bin have significant discrepancies between data and MC as show in Figure 8. In addition, a thereshold of 15 GeV can be triggerd by the sub-leading jet in the event as well. As a result, the inclusive jet p_{T} spectrum in the $[\,30, 40]$ p_{T} bin shows a secondary turn-on effect in data while the MC does not have this feature as one can see in Figure 8.

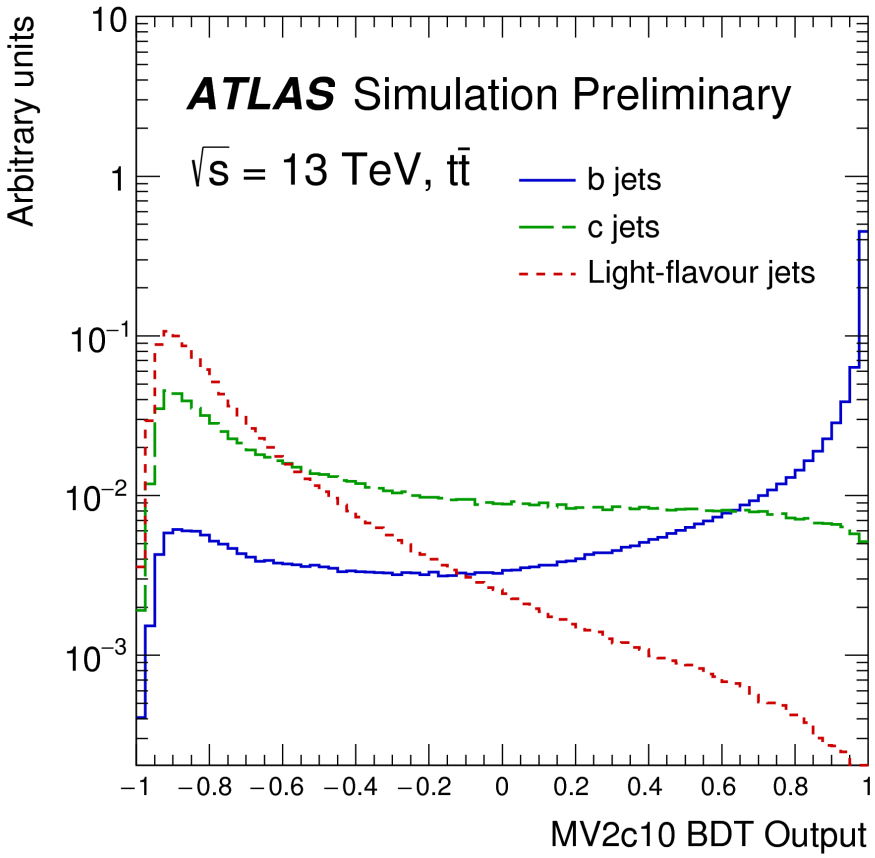


Figure 7: MV2c10 BDT output for b- (solid blue), c- (dashed green) and light-flavour (dotted red) jets evaluated with $t\bar{t}$ events [6]

Given the features discussed above, it is important to make sure that the discrepancies in jet p_{T} do not bias the template shapes as the fitting results are affected if it is the case. Therefore two different cross-checks are performed as will be illustrated below.

C.1 Jet p_{T} Reweighting

A simple reweighting technique is applied to the MC samples based on the inclusive jet p_{T} spectrum in data. The same event selections are applied, except the b-tagging criteria as listed in Table 2 are applied to both data and MC. In each jet p_{T} region as summarized in Table 1, the MC is normalized to data and then a scaling factor is obtained for each jet p_{T} bin with a step of 2 GeV to make MC match the data. Figure 9 shows the scaling factors. The scale factors in the high p_{T} region are close to unity while the ones in the low p_{T} region show large fluctuations, deviated more from unity.

However, the fitting results are not improved systematically applying the jet p_{T} scaling factors.

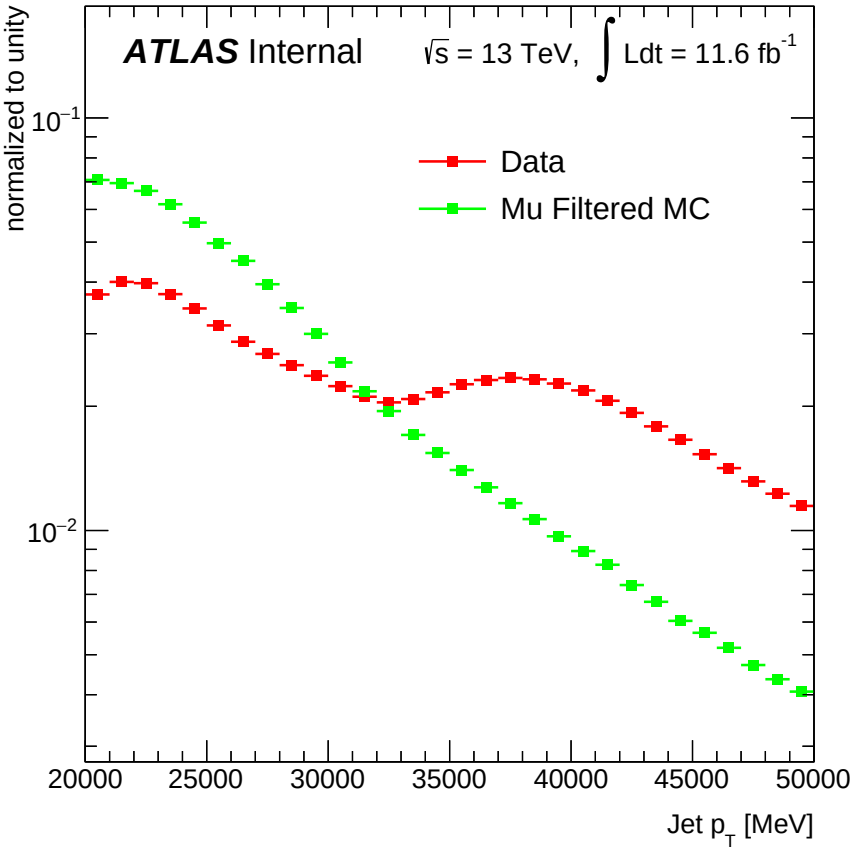


Figure 8: Comparison of the jet p_{T} spectrum in $[\,20, 60]$ p_{T} bin between data and MC.

C.2 Trigger Emulation

As reweighting the jet p_{T} does not change the fitting results systematically, a trigger emulation method is applied to further check the possible bias (again, note that the MC samples used do not have triggers simulated due to a production-level trigger calculation issue). The offline quantities are smeared according to a gaussian distribution and then taken as the online quantities. A jet with such low p_{T} around 15 GeV may not have its full energy recorded at the trigger level as only trigger towers with the energy deposited exceeding a given threshold are taken into account. The p_{T} of the jet is divided into three energy deposits equally with each deposit being smeared. The deposits that exceed 5 GeV are summed as the emulated online jet p_{T} which is then compared with the jet p_{T} threshold of the trigger. Muon is a single object so that smearing the offline muon p_{T} directy is enough.

As a sanity check, various values for the mean and standard deviations are applied in each step and no significant biases are observed in the final results.

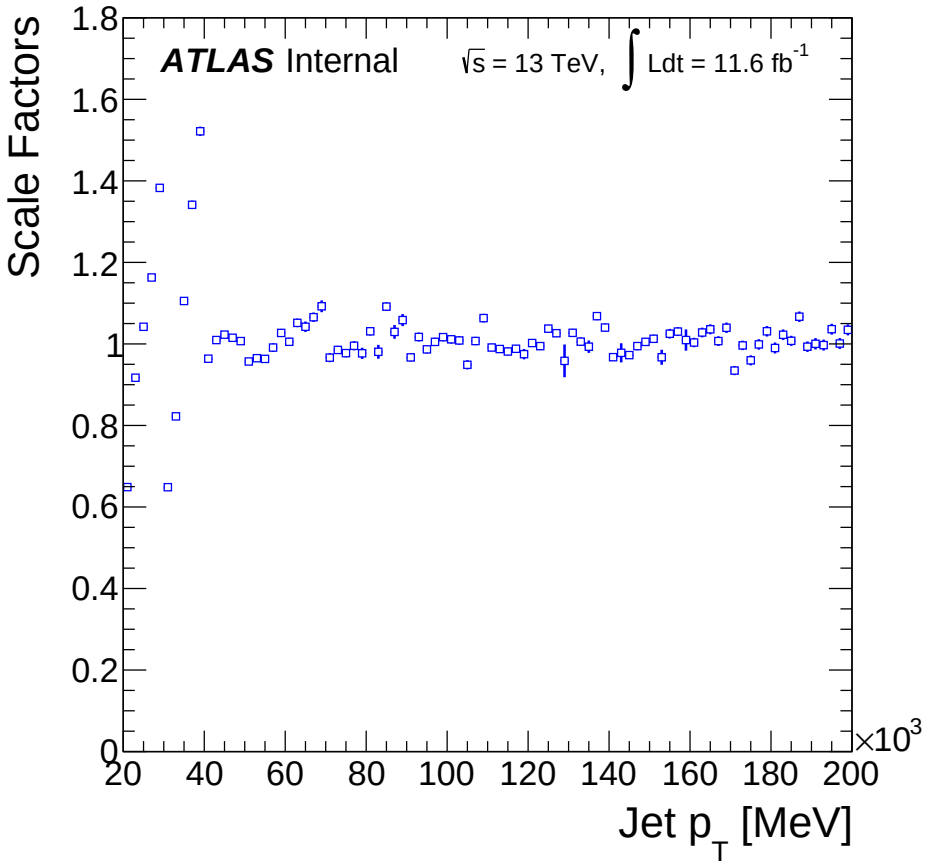


Figure 9: Scale factors applied to correct jet p_{T} in MC.

D Likelihood Definition for the $p_{\mathrm{T}}^{\mathrm{rel}}$ Fit

Assuming Poisson statistics the likelihood $\mathcal{L}$ for the expectation value λ given k observed events is

$$\mathcal{L}(\lambda|k) = \frac{\lambda^k}{k!}\exp(-\lambda). \tag{5}$$

Using the natural logarithm the likelihood of multiple Poisson observations becomes a sum over $p_{\mathrm{T}}^{\mathrm{rel}}$ bins i with the log-likelihood

$$\log \mathcal{L}_i(\lambda_i(f_b)|N_i^{\mathrm{data}}) = N_i^{\mathrm{data}} \log \lambda_i(f_b) - \lambda_i(f_b), \tag{6}$$

$$\text{with} \quad \lambda_i(f_b) = f_b \cdot N_i^b + (1 - f_b) \cdot N_i^{\text{non-}b}. \tag{7}$$

Where N_i^{data} is the number of observed events, N_i^b is the number of events in the b template and $N_i^{\text{non-}b}$ is the number of events in the combined c- and light-templates. The b-fraction f_b is the free parameter being optimized during the fitting procedure allowing to determine the flavour fractions in the tagged and untagged samples.

E Muon p^* spectrum

The modeling of the muon momentum in the b-hadron rest frame p^* is crucial for the p_T^{rel} method. Therefore, the agreement of the MC simulation with current measurements has been tested as shown in Figure 10. B-mesons are the most commonly produced b-hadrons for the considered selection. A set of B-mesons has been simulated and the p^* spectrum for the final state muons is shown in blue in Figure 10. This is compared to measurements of the final state lepton p^* measured by Delphi in red and BaBar in green [16, 17]. While the measurement by BaBar only includes B^0 mesons and B^+ mesons the Delphi measurement includes additional contributions from heavier b-hadrons. However the collaboration provides results which are corrected for those contributions (B_s, b-Baryons) using MC predictions [17]. Hence both measurements provide a good estimate for the expected B-meson p^* spectrum. It can be observed that the simulation matches the Delphi measurement a bit better, while both measurements show some disagreement. In order to account for this limited knowledge an uncertainty gets assigned, as discussed in Section 4.2, where the simulation gets reweighted to match the BaBar results. In the low p^* region no values for the BaBar measurement are available. Therefore, the weight calculated for the lowest p^* value available in both measurements is used for all lower p^* values.

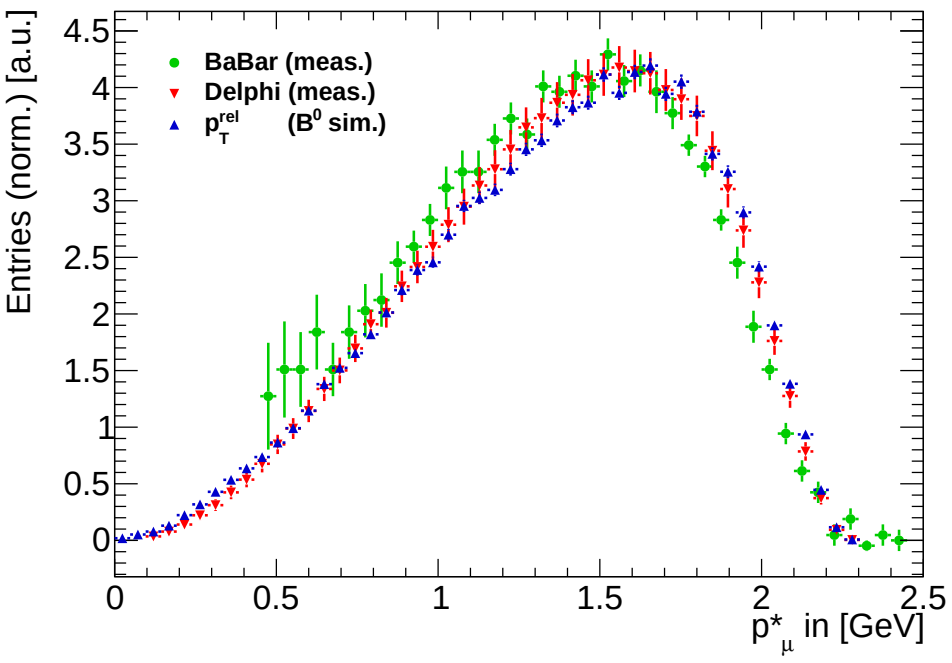


Figure 10: Final state lepton p^* from B-meson decays generated in simulation as used for the p_T^{rel} measurement shown in blue. In comparison measurements from Delphi and BaBar are shown in red and green, respectively [16, 17].

F Extrapolation to inclusive b-jets

The p_T^{rel} method provides the b-tagging efficiency calibration of jets initiated by semileptonic decays of b-hadrons. The scale factor derived with the p_T^{rel} method may be not safely usable in a sample of inclusive b-hadron decays: the extrapolation to inclusive b-jets needs to be studied in order to eventually assign a systematic uncertainty. This study has been performed also in the Run I analysis, giving a flat 4% value as systematics uncertainty [3].

The differences between the semileptonic and hadronic cases stand in the intrinsic properties of the decay and in the way the jets are experimentally selected. Concerning the intrinsic properties, they differ in the number of charged particles generated in the secondary vertex. Then semileptonic jets have to satisfy the request of containing an high momentum and well-reconstructed muon. It is observed (left plot in Figure 11) that low momentum hadronic b-jets have 15% loss of b-tagging efficiency if compared to the semileptonic case; the two efficiencies smoothly became the same after jet $p_T > 60$ GeV. The motivation

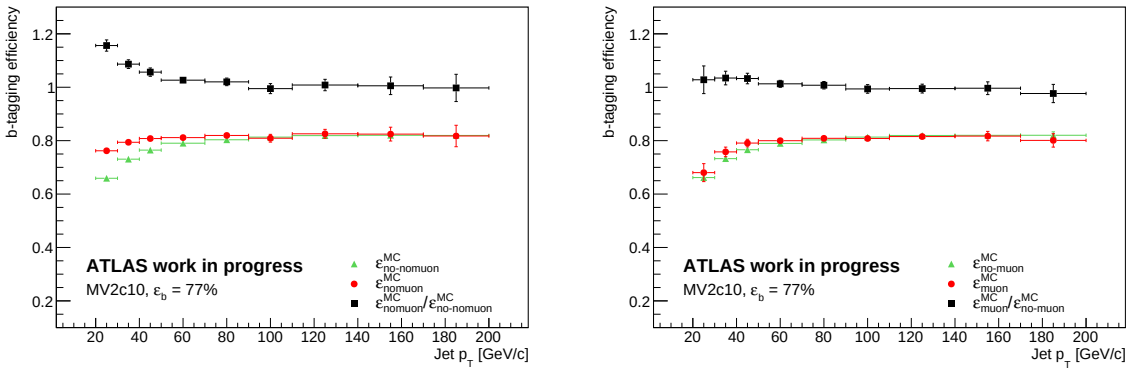


Figure 11: Simulated b-tagging efficiency as function of the jet transverse momentum of jets containing muons (red curve), jet not containing muons (green curve) and their ratio (black curve). The semileptonic correction to the jet momentum is applied in the right plot.

stand on the mis-reconstruction of low momentum particles from the secondary vertex that causes the b-tagging algorithm to fail the identification. It is shown in the right plot of Figure 11 how the trend of the ratio between the two efficiencies is partially recovered by applying a correction on the momentum of semileptonic jets. The correction consists in adding the muon momentum to the jet momentum and also removing the muon calorimeter energy deposits. The corrected jet represents in this way a better proxy for the b-hadron, which is not yet perfect due to the impossibility of re-summing the neutrinos contribution. However, if the simulation adequately reproduces the differences among the two cases, the scale factor evaluated by the p_T^{rel} method can be used also in the inclusive case.

In order to estimate the semileptonic-to-inclusive extrapolation uncertainty, b-tagging efficiency scale factors have been evaluated for jets containing a muon in a $\Delta R < 0.4$ and for jets failing this requirement. The estimation of these scale factors has been performed using the tag-and-probe technique in a high purity sample of b-jets of dileptonic $t\bar{t}$ events. A total integrated luminosity of $36.1\,fb^{-1}$ of data collected during 2015 and 2016 has been used for this purpose.

The following selections are applied to define the signal region:

- One prompt electron of $p_T > 25$ GeV and $|\eta| < 2.47$, excluding the $1.37 < |\eta| < 1.57$ region

- One prompt muon of $p_T > 25$ GeV and $|\eta| < 2.4$
- Opposite electric charge between the prompt electron and the prompt muon
- Two jets of $p_T > 20$ GeV and $|\eta| < 2.5$
- At least one b-tagged jet at the 85% working point of the MV2c10 tagger

It is estimated from the Montecarlo simulation that these selections ensure a 91% of $t\bar{t} \to e\mu + 2j$ in the signal region. The main background processes are the single-top and diboson productions, contributing in the signal region with the 8% of events. The production of W and Z bosons in association to jets and to top-quark pairs represent a 1% contribution. A good agreement between data and simulation is found in all the kinematical distributions.

In the tag-and-probe method the 85% b-tagged jet is fixed as tag while the other jet is used as probe to measure the b-tagging efficiency. In the case in which the two jets in the event are b-tagged at the 85%, then both of them are used as probe. The simulated b-tagging efficiency is computed using the information from the Montecarlo truth as follows.

$$\varepsilon_b^{sim} = \frac{N_{truth-b}^{pass}}{N_{truth-b}} \tag{8}$$

The same quantity is calculated in data events Equation (10). As a first step subtracting the contribution of the background Equation (9) and then correcting by the number of probe jets coming from b-quarks and the mis-tag rate estimated in the $t\bar{t}$ simulation.

$$\varepsilon_b^{uncorr} = \frac{N_{data}^{pass} - N_{bkg}^{pass}}{N_{data} - N_{bkg}} \tag{9}$$

$$\varepsilon_b^{data} = \frac{\varepsilon_b^{uncorr} - (1 - f_b^{t\bar{t}}) \cdot \varepsilon_{non-b}^{t\bar{t}}}{f_b^{t\bar{t}}} \tag{10}$$

The resulting scale factors in both the semileptonic and hadronic case are found to be consistent with unity in all the jet p_T bins and so is their ratio. Figure 12 shows the comparison of b-tagging efficiency for data and simulated events.

Figure 13 shows the ratio between the scale factors for jets with or without muons for the 77% working point: its flatness can ben interpreted as the goodness of the simulation to reproduce the different topologies of semileptonic and hadron decays of all b-hadrons. The distribution of the scale factors ratio as function of the jet p_T has been fitted using a constant function. The result of the fit is compatible with the unity for all the b-tagging working points: the uncertainty on the constant value from the fit is taken as the semileptonic-to-inclusive extrapolation uncertainty for the p_T^{rel} calibration.

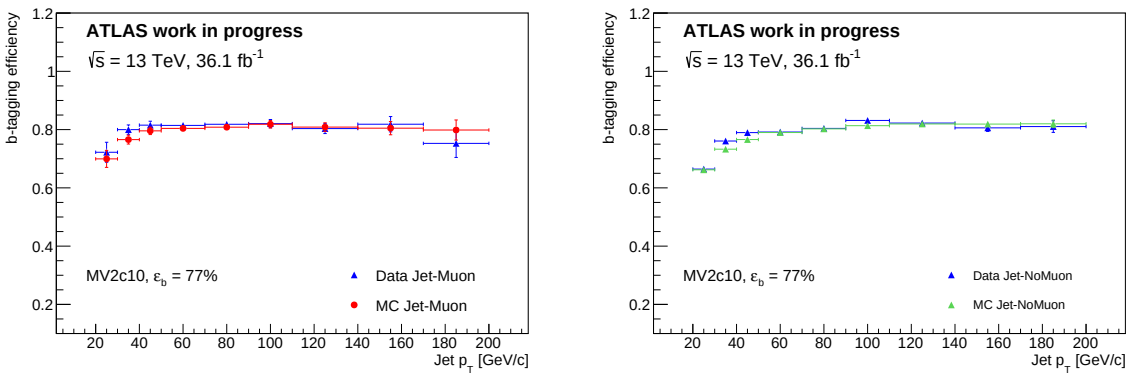


Figure 12: b-tagging efficiency in data and simulation as function of the jet transverse momentum, for jet containing muons (left) and jet without muons (right)

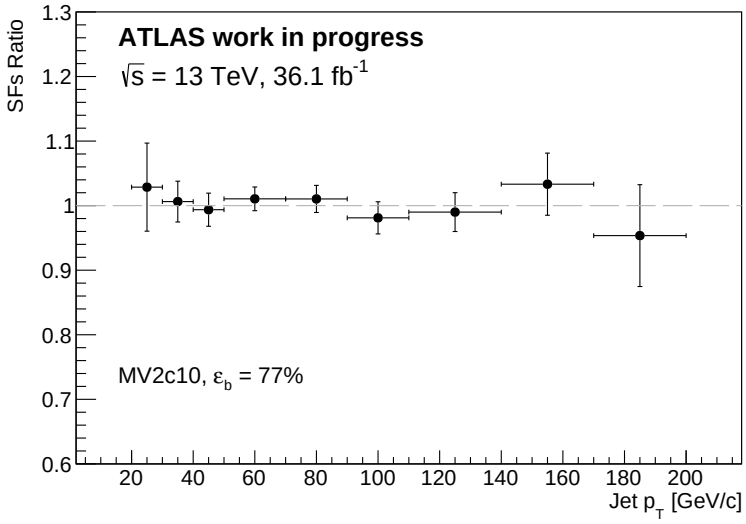


Figure 13: Ratio between the b-tagging efficiency scale factor of jet with and without muons for the 77% working point as function of jet p_T bins.

G Dependence on Run Conditions

The analysis presented in this note extends throughout multiple data taking periods which have different conditions like e.g. the amount of pile-up or the pre-scales for the utilised triggers. In order to check for systematic effects the analysis data has been split into multiple data taking periods and the same analysis has been conducted for all of them independently. Figure 14 shows the measurement of the b-tagging efficiency for MV2c10 at $\varepsilon_b^{nom} = 70\%$ split up into multiple data taking periods. Error bars only include statistical uncertainties. No systematic effect in dependence of run conditions is observed.

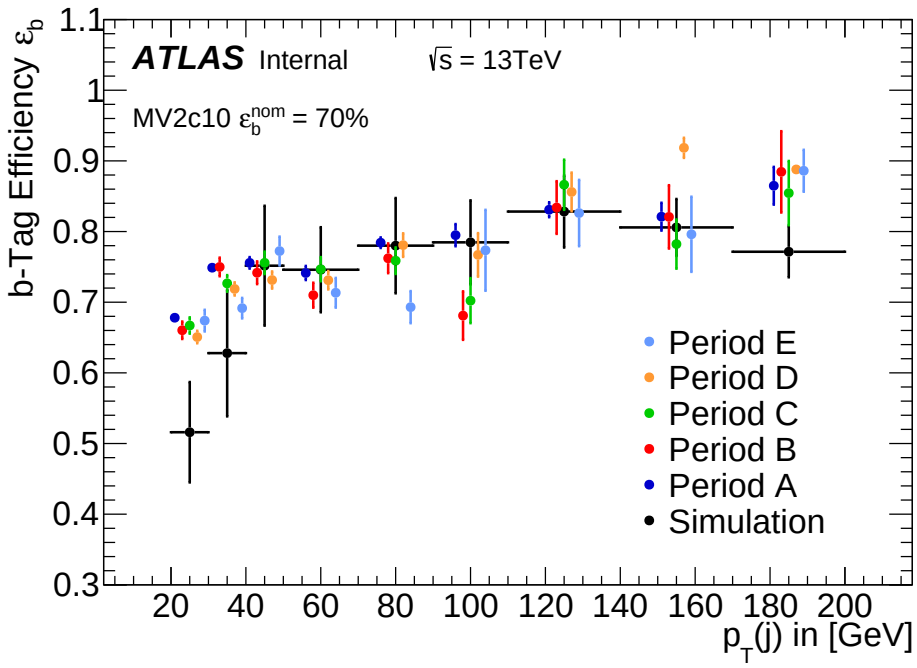


Figure 14: Efficiency ε_b measured for MV2c10 at $\varepsilon_b^{nom} = 70\%$ in dependence of jet p_{T}. The total data sample is split up into multiple data taking periods shown in different colors. The $p_{\mathrm{T}}^{\mathrm{jet}}$ binning for all periods is identical to the binning used for the prediction from simulation shown in black.

H Distributions of $p_{\mathrm{T}}^{\mathrm{rel}}$ using fit results

Using the templates generated as described in Section 2 the fit to the $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution is performed. The results for the $\varepsilon_b^{\mathrm{nom}} = 70\%$ WP of the MV2c10 algorithm and all $p_{\mathrm{T}}^{\mathrm{jet}}$ bins are shown in figures 15-17. The fit is performed for the tagged and the untagged distributions independently as described in Section 3.

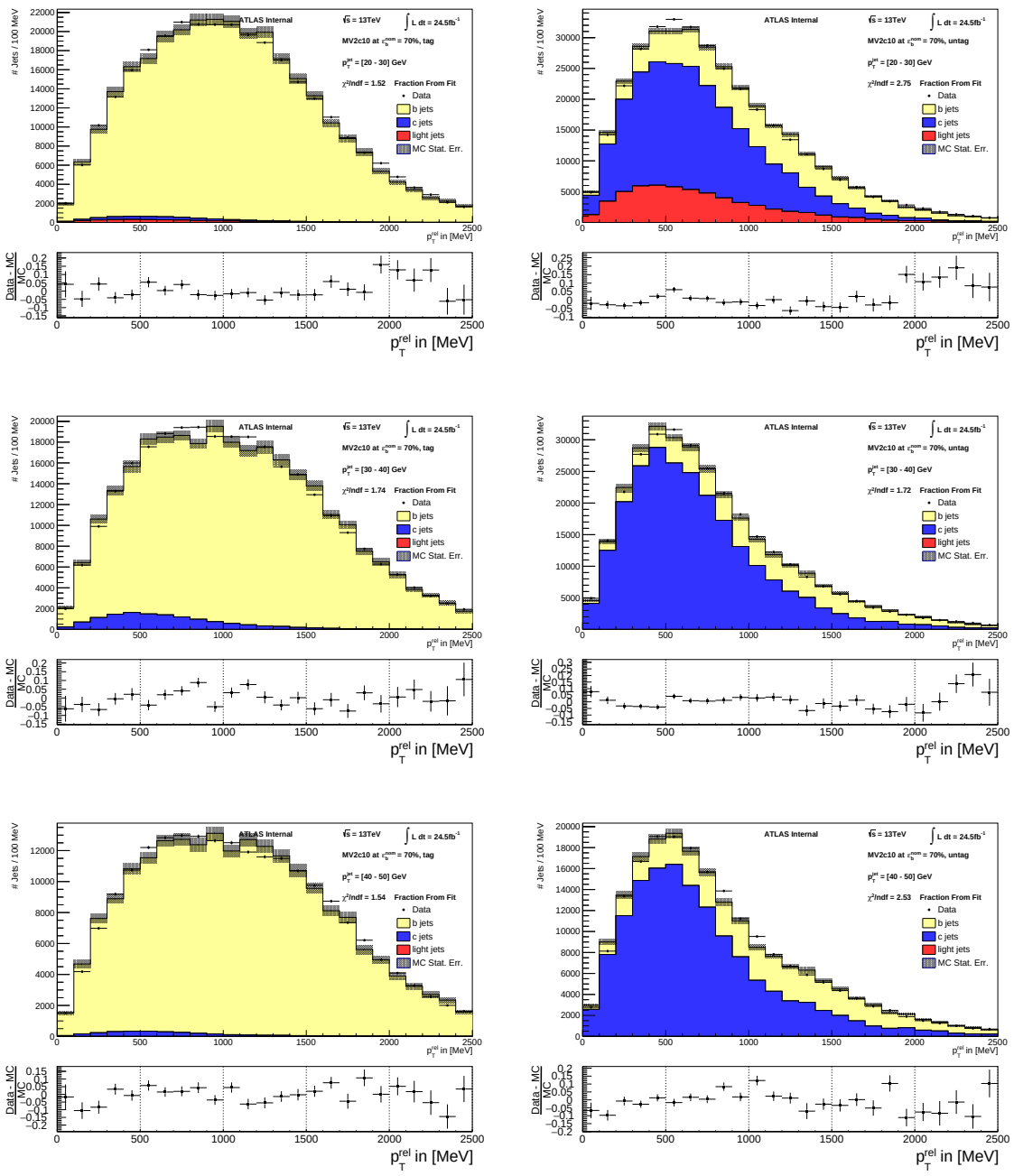


Figure 15: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions obtained by the log-likelihood fit.

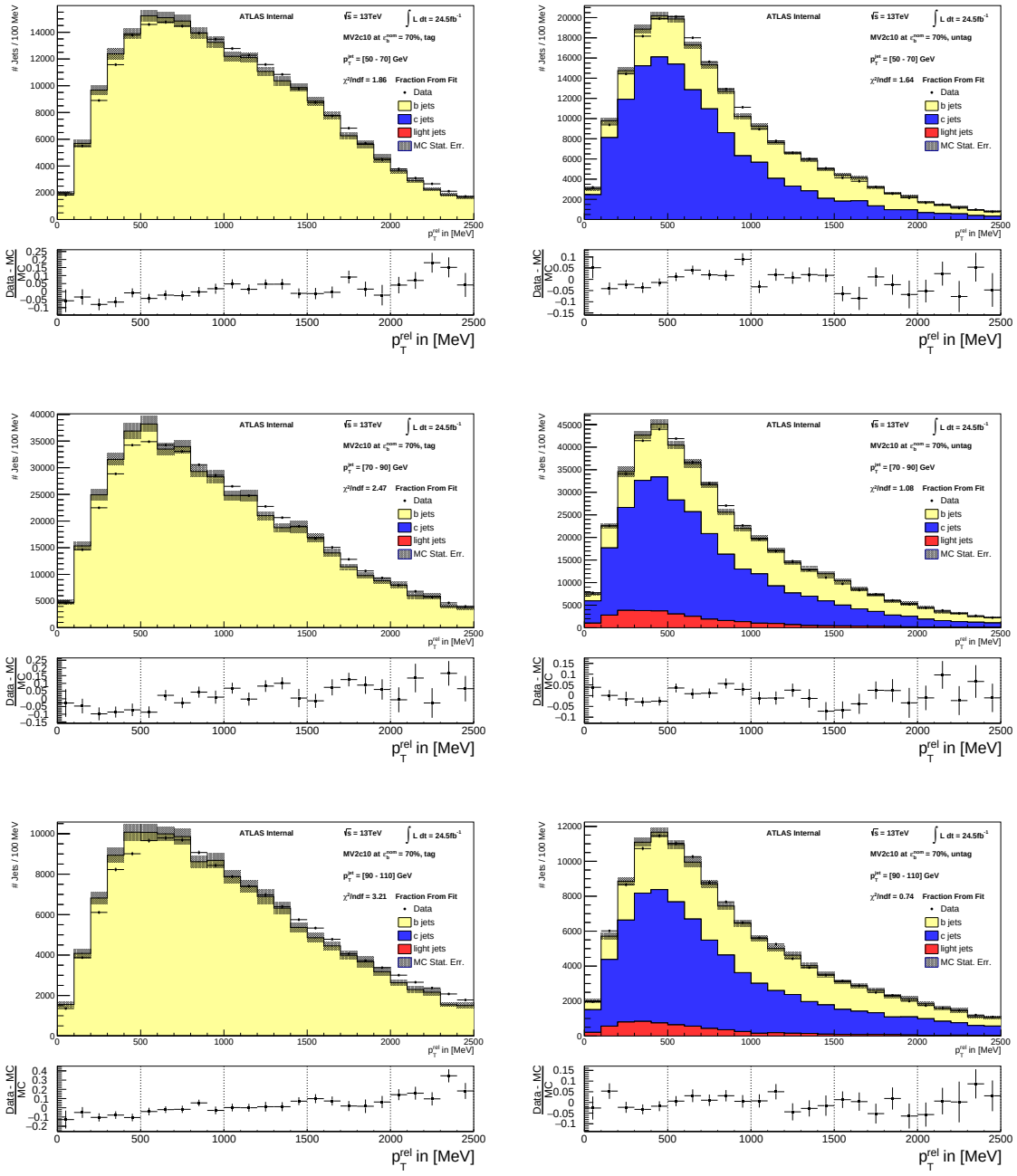


Figure 16: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions obtained by the log-likelihood fit.

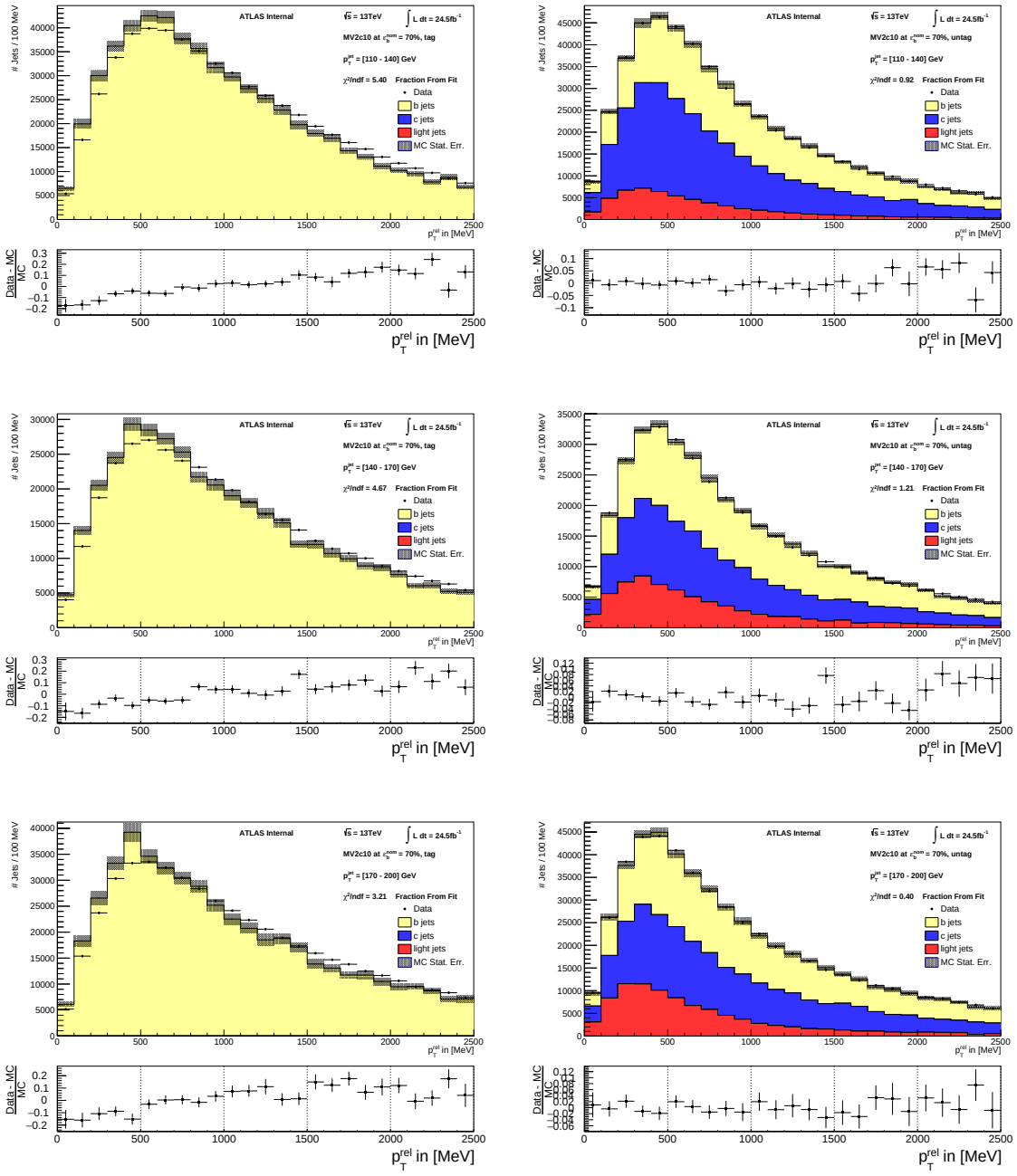


Figure 17: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions obtained by the log-likelihood fit.

I Distributions of $p_{\mathrm{T}}^{\mathrm{rel}}$ using MC predictions

The templates can also be combined using the flavour fractions as predicted in MC. Using the fractions from simulation the resulting $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions for the $\varepsilon_b^{\mathrm{nom}}$ = 70% WP of the MV2c10 algorithm and all $p_{\mathrm{T}}^{\mathrm{jet}}$ bins are shown in figures 18-20.

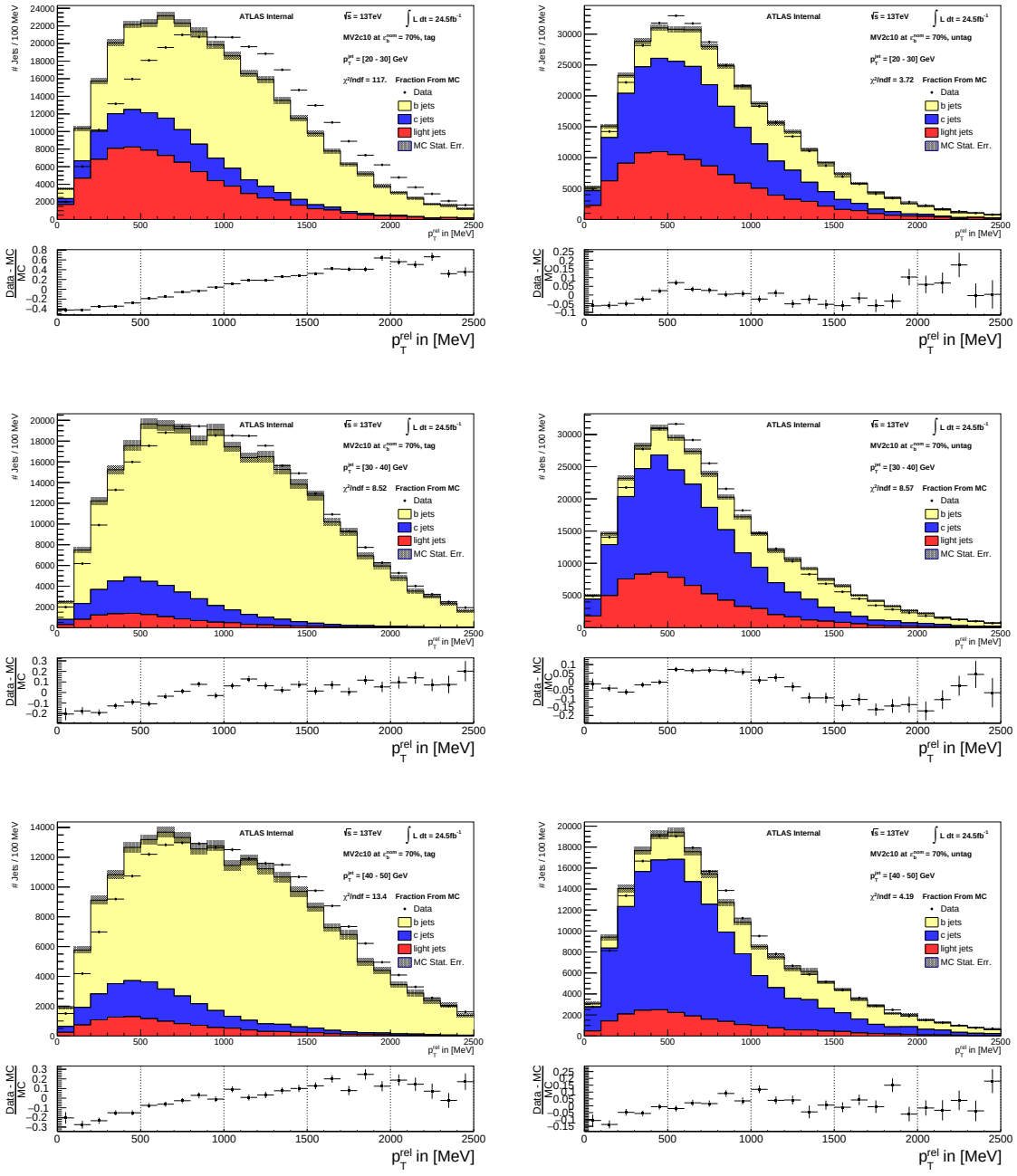


Figure 18: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions as predicted in simulation.

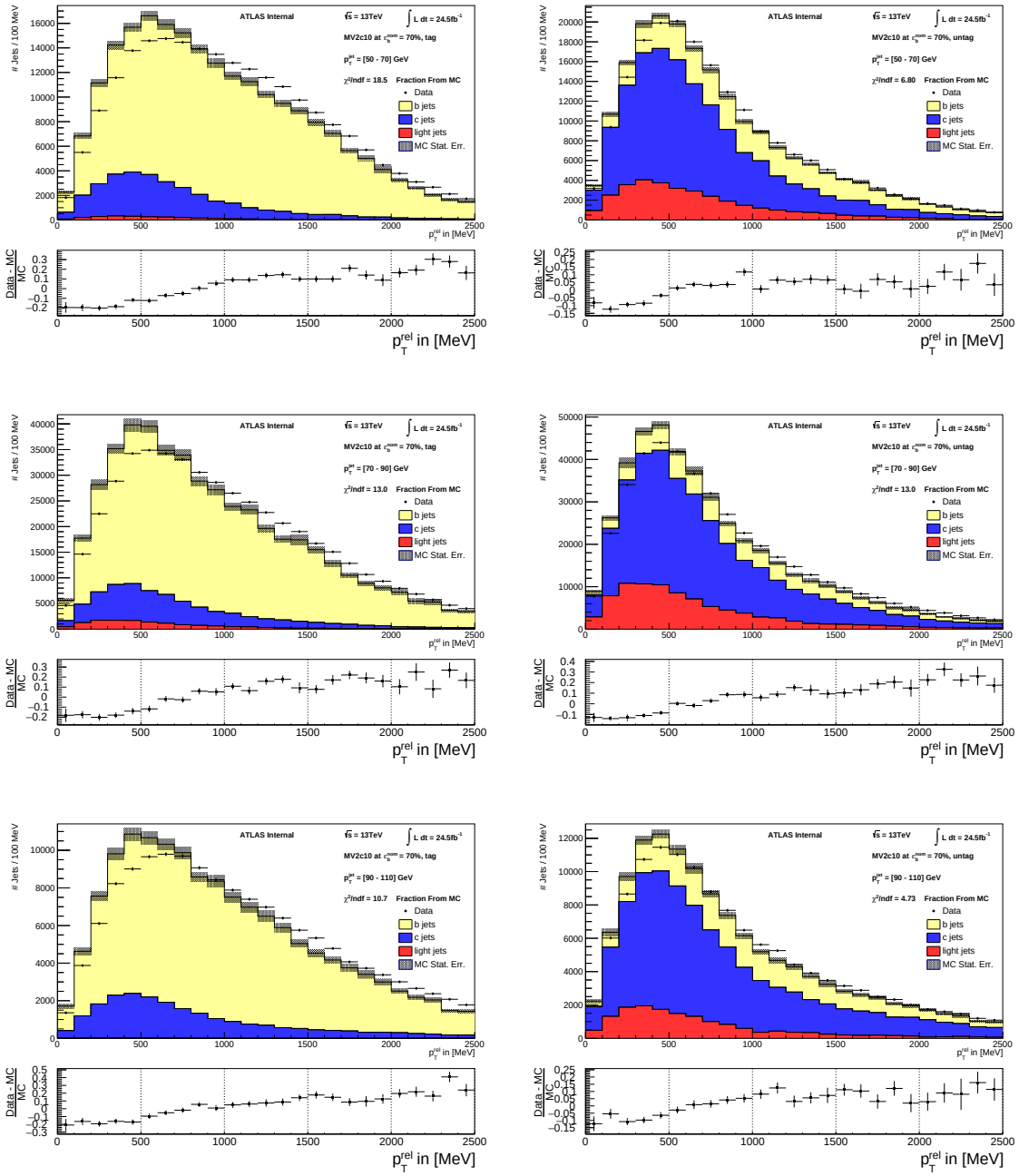


Figure 19: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions as predicted in simulation.

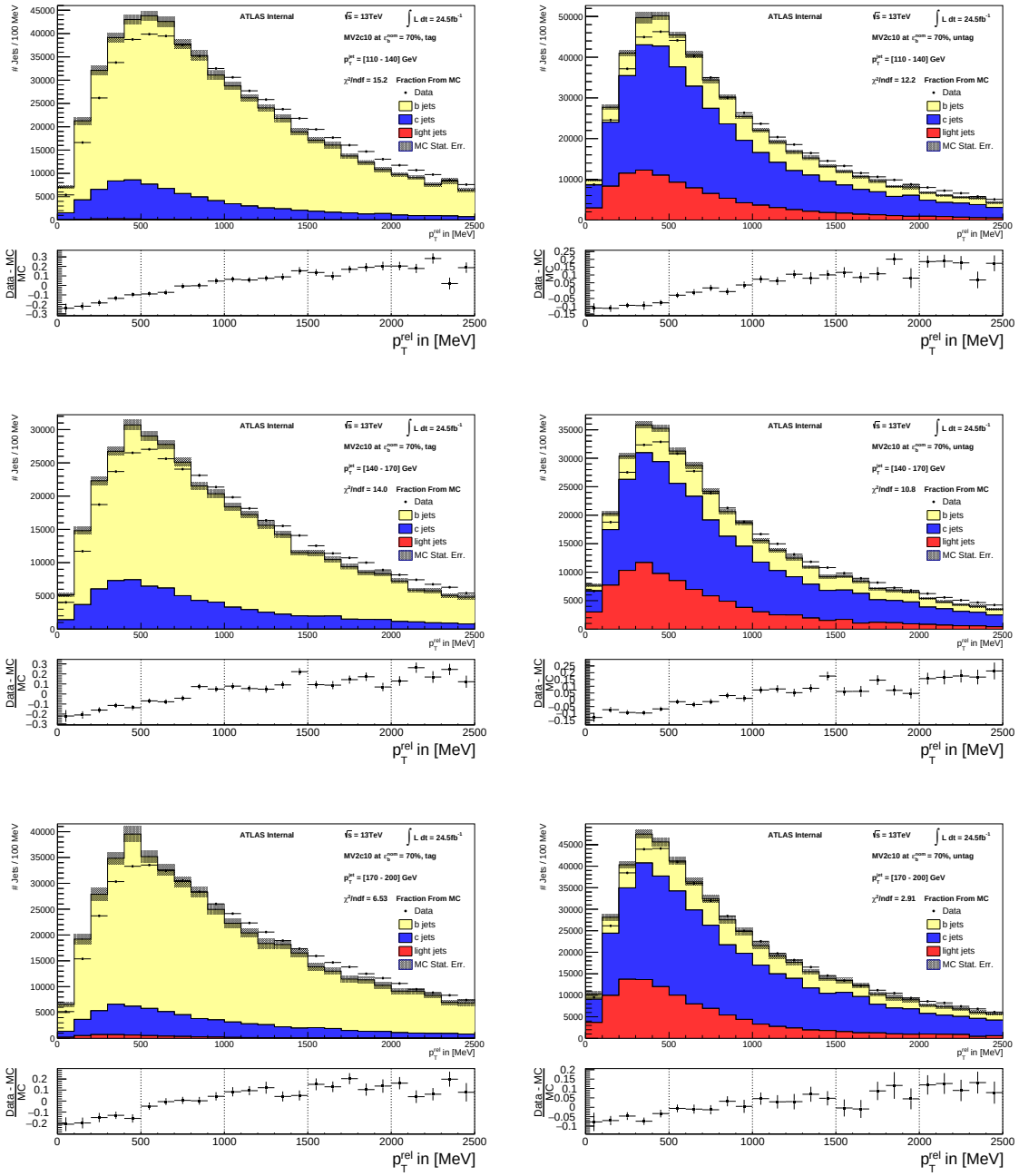


Figure 20: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions as predicted in simulation.

J Overview of Systematics Tables

Table 10: Systematic uncertainties for the MV2c10 tagging algorithm at 60% nominal *b*-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[GeV]$ Bins							
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties									
Pileup Offset Mu	0.08	0.02	0.02	0.06	0.09	0.24	0.02	0.14	0.27
Pileup Offset NPV	0.04	0.19	<0.01	0.08	0.17	0.25	0.55	0.32	0.04
Pileup Pt Term	0.04	0.07	0.18	0.25	0.67	0.02	0.54	0.66	1.15
Pileup Rho Topology	0.66	0.73	0.62	0.51	0.63	0.07	1.01	0.38	0.33
B JES Response	0.20	0.40	0.62	0.53	0.58	0.04	0.60	0.40	0.69
Flavour Composition	0.03	0.21	0.14	0.07	0.04	0.04	0.01	0.14	0.18
Flavour Response	0.13	0.06	0.01	0.01	0.04	0.02	0.01	<0.01	0.25
Eta Intercalibration Modeling	0.24	0.18	0.20	0.23	0.26	0.19	0.22	0.25	1.88
Eta Intercalibration Total Statistics	0.09	0.07	0.21	0.10	0.27	0.39	0.47	0.05	0.36
Effective NP 1	0.51	0.31	0.41	0.22	0.25	0.04	0.57	0.71	0.55
Effective NP 2	0.01	<0.01	0.12	0.05	0.19	0.34	0.40	0.22	0.54
Effective NP 3	0.02	0.02	0.11	0.06	0.21	0.36	0.41	0.25	0.76
Effective NP 4	0.03	0.01	0.16	0.06	0.22	0.44	0.50	0.19	0.09
Effective NP 5	0.06	<0.01	0.18	0.01	0.09	<0.01	0.08	0.01	0.05
Effective NP 6	<0.01	0.01	0.14	0.01	<0.01	0.21	0.22	0.05	0.23
Effective NP 7	0.03	0.01	0.15	0.01	0.03	0.16	0.22	0.08	0.08
Effective NP 8 Rest Term	0.02	0.01	0.11	0.01	0.08	0.18	0.20	0.03	0.24
JER Single NP	0.60	0.98	1.29	1.01	0.08	4.16	1.62	4.36	1.89
Muon ID	0.01	<0.01	0.04	0.01	0.12	<0.01	0.17	0.23	0.42
Muon MS	0.03	0.01	0.03	0.01	0.05	0.09	0.01	0.69	0.64
Muon Scale	0.05	0.06	0.05	0.04	0.03	0.25	0.07	0.47	0.40
Modeling Uncertainties									
Axis Smearing	0.36	0.64	0.76	1.02	1.06	2.88	0.64	2.73	10.62
B Decay Fractions	0.43	0.59	0.67	0.96	1.48	1.48	1.68	1.52	0.57
B Decay p* Spectrum	2.46	2.89	2.81	2.86	4.75	2.75	2.20	2.34	1.21
Fake Muons	0.42	0.28	0.20	0.28	0.51	1.18	0.25	1.43	1.11
Gluon Splitting B	0.07	0.17	0.60	1.25	4.17	3.51	4.31	3.66	2.39
Gluon Splitting C	0.03	0.06	0.06	0.15	0.42	0.26	0.83	0.10	0.53
B Fragmentation Fractions	0.37	0.44	0.52	0.80	0.02	0.58	0.63	0.96	3.44
B Fragmentation Function	0.13	0.08	0.38	0.26	0.76	0.62	1.03	1.34	2.42
Extrapolation to Inclusive	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
MC Statistical Uncertainties									
Simulation Statistics	2.67	2.03	1.04	0.69	0.84	1.12	0.67	0.76	0.89
Template Statistics	0.91	0.78	2.30	0.85	1.79	2.18	2.47	3.86	5.60
Template Selection Uncertainties									
Light Contamination	0.10	<0.01	<0.01	<0.01	0.02	0.01	0.11	0.08	0.33
JVT Efficiency	0.03	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Data Statistical Uncertainty	0.15	0.14	0.16	0.18	0.24	<0.01	0.36	0.56	0.73
Total Systematic Uncertainty	4.21	4.22	4.57	4.15	7.19	7.66	6.57	8.49	13.60

Table 11: Systematic uncertainties for the MV2c10 tagging algorithm at 70% nominal *b*-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}$[*GeV*] Bins							
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties									
Pileup Offset Mu	0.06	0.01	0.07	0.05	0.08	0.15	0.02	0.10	0.18
Pileup Offset NPV	0.04	0.16	<0.01	0.06	0.17	0.13	0.41	0.20	0.03
Pileup Pt Term	0.04	0.03	0.07	0.17	0.51	0.02	0.40	0.59	0.85
Pileup Rho Topology	0.66	0.30	0.23	0.32	0.48	0.03	0.71	0.16	0.31
B JES Response	0.20	0.16	0.24	0.35	0.45	0.04	0.43	0.22	0.52
Flavour Composition	0.06	0.10	0.05	0.06	0.04	0.04	0.01	0.11	0.17
Flavour Response	0.07	0.04	0.02	0.01	0.03	0.02	0.01	0.01	0.18
Eta Intercalibration Modeling	0.15	0.07	0.06	0.15	0.20	0.18	0.07	0.29	1.47
Eta Intercalibration Total Statistics	0.05	0.06	0.08	0.06	0.23	0.26	0.32	0.01	0.28
Effective NP 1	0.43	0.12	0.17	0.12	0.22	0.07	0.41	0.43	0.46
Effective NP 2	0.03	0.02	0.05	0.04	0.17	0.21	0.30	0.15	0.42
Effective NP 3	0.02	0.01	0.06	0.04	0.19	0.23	0.31	0.16	0.53
Effective NP 4	0.02	0.01	0.08	0.04	0.19	0.29	0.35	0.17	0.12
Effective NP 5	0.03	0.03	0.08	<0.01	0.08	<0.01	0.06	<0.01	0.06
Effective NP 6	0.02	0.01	0.06	0.01	0.01	0.13	0.16	0.02	0.15
Effective NP 7	0.03	0.01	0.06	0.01	0.03	0.11	0.16	0.06	0.12
Effective NP 8 Rest Term	0.02	0.01	0.05	0.01	0.08	0.12	0.15	0.05	0.17
JER Single NP	0.20	0.76	0.36	0.74	0.29	2.97	1.34	2.66	1.00
Muon ID	<0.01	0.02	0.04	0.01	0.08	0.11	0.12	0.13	0.40
Muon MS	0.01	<0.01	0.02	<0.01	0.03	0.02	0.03	0.42	0.40
Muon Scale	0.03	0.02	0.04	0.03	0.02	0.24	0.02	0.40	0.26
Modeling Uncertainties									
Axis Smearing	0.25	0.44	0.57	0.82	0.73	2.21	0.25	1.86	7.00
B Decay Fractions	0.01	0.03	0.08	0.59	0.94	0.97	1.12	1.06	0.46
B Decay p* Spectrum	0.87	0.85	1.48	1.84	3.07	1.89	1.36	1.94	0.73
Fake Muons	0.11	0.01	0.07	0.27	0.39	0.96	0.18	1.35	0.83
Gluon Splitting B	0.01	0.01	0.08	0.78	2.63	2.04	2.49	1.91	1.65
Gluon Splitting C	0.02	0.05	0.05	0.13	0.34	0.23	0.67	0.05	0.21
B Fragmentation Fractions	0.01	0.11	0.05	0.50	0.01	0.36	0.44	0.52	2.41
B Fragmentation Function	0.03	0.04	0.02	0.16	0.47	0.40	0.64	0.78	1.62
Extrapolation to Inclusive	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
MC Statistical Uncertainties									
Simulation Statistics	1.83	1.23	0.42	0.38	0.53	0.76	0.44	0.45	0.72
Template Statistics	0.56	0.63	0.57	0.68	1.35	1.76	1.85	2.86	4.10
Template Selection Uncertainties									
Light Contamination	0.17	<0.01	<0.01	<0.01	0.03	0.11	0.16	0.10	0.39
JVT Efficiency	0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Data Statistical Uncertainty	0.17	0.15	0.17	0.19	0.25	0.60	0.39	0.67	0.78
Total Systematic Uncertainty	2.54	2.20	2.14	2.84	4.76	5.39	4.34	5.71	9.27

Table 12: Systematic uncertainties for the MV2c10 tagging algorithm at 77% nominal *b*-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}$[*GeV*] Bins							
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties									
Pileup Offset Mu	0.01	0.02	0.07	0.02	0.06	0.09	0.03	0.08	0.14
Pileup Offset NPV	0.01	0.08	0.01	0.02	0.10	0.04	0.30	0.14	0.04
Pileup Pt Term	0.01	0.02	0.04	0.05	0.33	<0.01	0.29	0.48	0.66
Pileup Rho Topology	0.22	0.14	0.20	0.03	0.29	0.13	0.47	0.08	0.41
B JES Response	0.06	0.05	0.21	0.06	0.29	0.08	0.28	0.13	0.47
Flavour Composition	0.02	0.04	0.05	0.04	0.03	0.02	<0.01	0.04	0.11
Flavour Response	0.02	0.02	0.01	0.01	0.01	0.01	<0.01	0.02	0.11
Eta Intercalibration Modeling	0.05	0.01	0.08	0.02	0.14	0.13	<0.01	0.30	1.08
Eta Intercalibration Total Statistics	0.01	0.04	0.05	<0.01	0.16	0.15	0.19	0.01	0.17
Effective NP 1	0.13	0.04	0.12	0.02	0.15	0.12	0.29	0.29	0.48
Effective NP 2	0.01	0.02	0.04	0.02	0.12	0.11	0.19	0.09	0.39
Effective NP 3	0.01	0.01	0.04	0.01	0.13	0.12	0.21	0.10	0.44
Effective NP 4	0.01	0.01	0.06	<0.01	0.14	0.16	0.23	0.14	0.12
Effective NP 5	0.01	0.01	0.05	0.02	0.06	<0.01	0.04	0.02	0.03
Effective NP 6	<0.01	<0.01	0.04	0.01	0.01	0.07	0.09	0.02	0.09
Effective NP 7	0.01	<0.01	0.04	0.01	0.03	0.06	0.09	0.03	0.10
Effective NP 8 Rest Term	0.01	0.01	0.03	0.01	0.05	0.06	0.09	0.03	0.09
JER Single NP	0.09	0.40	0.40	0.23	0.41	2.01	1.10	1.38	0.77
Muon ID	0.01	0.02	0.01	<0.01	0.04	0.06	0.07	0.05	0.30
Muon MS	<0.01	<0.01	0.03	<0.01	0.03	0.03	<0.01	0.21	0.21
Muon Scale	0.01	<0.01	0.03	0.02	0.02	0.15	0.01	0.27	0.14
Modeling Uncertainties									
Axis Smearing	0.06	0.28	0.42	0.49	0.29	1.29	0.04	1.27	4.37
B Decay Fractions	<0.01	0.01	0.01	0.03	0.53	0.55	0.66	0.64	0.32
B Decay p* Spectrum	0.15	0.31	0.25	0.66	1.68	0.96	0.61	1.67	0.90
Fake Muons	0.03	0.08	0.14	0.22	0.31	0.67	0.18	0.98	0.93
Gluon Splitting B	<0.01	0.01	0.03	0.07	1.26	0.95	1.12	0.49	0.11
Gluon Splitting C	0.01	0.03	0.05	0.09	0.22	0.17	0.48	<0.01	0.04
B Fragmentation Fractions	<0.01	0.06	0.01	0.06	0.01	0.19	0.21	0.14	1.34
B Fragmentation Function	<0.01	0.02	0.01	0.01	0.26	0.19	0.29	0.37	0.85
Extrapolation to Inclusive	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
MC Statistical Uncertainties									
Simulation Statistics	1.35	0.43	0.03	0.13	0.32	0.55	0.27	0.29	0.45
Template Statistics	0.16	1.23	0.50	0.55	0.91	1.25	1.35	1.87	2.63
Template Selection Uncertainties									
Light Contamination	0.23	0.10	0.01	<0.01	0.04	0.17	0.19	0.11	0.41
JVT Efficiency	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Data Statistical Uncertainty	0.15	0.13	0.17	0.18	0.23	0.57	0.39	0.67	0.76
Total Systematic Uncertainty	1.69	1.72	1.29	1.41	2.72	3.36	2.69	3.62	5.91

Table 13: Systematic uncertainties for the MV2c10 tagging algorithm at 85% nominal *b*-tagging efficiency

Systematic Uncertainty Source	Systematic Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}$[*GeV*] Bins							
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Detector and Calibration Uncertainties									
Pileup Offset Mu	<0.01	<0.01	0.01	<0.01	0.01	0.01	0.01	<0.01	0.01
Pileup Offset NPV	<0.01	0.01	<0.01	<0.01	0.02	0.02	0.02	0.01	0.01
Pileup Pt Term	<0.01	<0.01	<0.01	<0.01	0.03	0.01	0.02	0.03	0.07
Pileup Rho Topology	0.05	0.01	0.01	0.01	0.04	0.01	<0.01	0.02	0.01
B JES Response	0.01	0.01	0.01	<0.01	0.04	0.02	0.01	0.01	0.01
Flavour Composition	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	0.02
Flavour Response	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01
Eta Intercalibration Modeling	0.01	<0.01	<0.01	<0.01	0.01	0.03	0.03	0.04	0.04
Eta Intercalibration Total Statistics	<0.01	0.01	<0.01	<0.01	0.01	0.02	0.01	0.01	<0.01
Effective NP 1	0.03	0.01	0.01	0.01	0.03	0.04	<0.01	0.01	<0.01
Effective NP 2	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	0.01	<0.01	0.01
Effective NP 3	<0.01	<0.01	<0.01	<0.01	0.01	0.01	0.01	0.01	0.01
Effective NP 4	<0.01	<0.01	<0.01	<0.01	0.01	0.02	<0.01	<0.01	<0.01
Effective NP 5	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.01	0.01	0.01
Effective NP 6	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	0.01	0.01
Effective NP 7	<0.01	<0.01	<0.01	<0.01	<0.01	0.03	0.01	0.01	<0.01
Effective NP 8 Rest Term	<0.01	<0.01	<0.01	<0.01	0.01	0.01	0.01	0.01	<0.01
JER Single NP	0.02	0.06	0.04	0.05	0.04	0.22	0.04	0.06	0.25
Muon ID	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	<0.01
Muon MS	<0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.01	0.01	<0.01
Muon Scale	<0.01	<0.01	<0.01	<0.01	<0.01	0.09	0.01	0.01	<0.01
Modeling Uncertainties									
Axis Smearing	0.02	0.05	0.05	0.07	0.10	0.05	0.33	0.54	0.53
B Decay Fractions	<0.01	0.01	<0.01	<0.01	0.02	0.01	0.05	0.10	0.13
B Decay p* Spectrum	0.08	0.02	0.01	0.01	0.04	0.03	0.13	0.05	0.10
Fake Muons	<0.01	0.01	0.01	0.03	0.04	0.21	0.02	0.03	0.01
Gluon Splitting B	<0.01	<0.01	<0.01	<0.01	0.05	0.18	0.02	0.34	0.33
Gluon Splitting C	<0.01	<0.01	0.01	0.01	0.04	0.06	0.09	0.08	0.08
B Fragmentation Fractions	<0.01	0.01	<0.01	0.01	0.01	0.06	0.06	0.01	0.01
B Fragmentation Function	0.01	<0.01	<0.01	<0.01	0.01	0.08	0.01	0.06	0.12
Extrapolation to Inclusive	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
MC Statistical Uncertainties									
Simulation Statistics	0.59	0.57	0.16	0.18	0.10	0.20	0.08	0.12	0.13
Template Statistics	1.20	0.03	0.61	0.55	1.38	1.31	0.19	0.31	0.55
Template Selection Uncertainties									
Light Contamination	0.08	0.14	<0.01	0.01	0.24	0.24	0.38	0.39	1.31
JVT Efficiency	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Data Statistical Uncertainty	0.10	0.09	0.11	0.12	0.10	0.40	0.11	0.16	0.16
Total Systematic Uncertainty	1.52	0.92	0.95	0.92	1.58	1.57	0.91	1.09	1.75

References

[1] ATLAS Collaboration,
Measurement of the top quark mass in the $t\bar{t} \to$ dilepton channel from $\sqrt{s} = 8$ TeV ATLAS data,
Phys. Lett. **B761** (2016) 350, arXiv: 1606.02179 [hep-ex].

[2] ATLAS Collaboration, 'Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a $b\bar{b}$ pair in pp collisions at 13 TeV using the ATLAS detector',
tech. rep. ATLAS-CONF-2016-091, CERN, 2016,
URL: http://cds.cern.ch/record/2206813.

[3] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*,
JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex].

[4] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector*, ATLAS-CONF-2012-043 (2012),
URL: https://cds.cern.ch/record/1435197.

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

[6] ATLAS Collaboration,
'Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run',
tech. rep. ATL-PHYS-PUB-2016-012, CERN, 2016,
URL: https://cds.cern.ch/record/2160731.

[7] A. Buckley, 'ATLAS Pythia 8 tunes to 7 TeV data', tech. rep. ATL-PHYS-PROC-2014-273, CERN, 2014, URL: https://cds.cern.ch/record/1974411.

[8] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244,
arXiv: 1207.1303 [hep-ph].

[9] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. **A462** (2001) 152.

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

[11] ATLAS Collaboration, 'Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s} = 13$ TeV',
tech. rep. ATL-PHYS-PUB-2015-036, CERN, 2015,
URL: https://cds.cern.ch/record/2044564.

[12] ATLAS Collaboration,
Performance of pile-up mitigation techniques for jets in pp collisions with the ATLAS detector,
(2015), arXiv: 1510.03823 [hep-ex].

[13] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. **C76** (2016) 292, arXiv: 1603.05598 [hep-ex].

[14] ATLAS Collaboration, 'Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector',
tech. rep. ATL-COM-PHYS-2016-213, CERN, 2016,
URL: https://cds.cern.ch/record/2136864.

[15] K. Nakamura et al. (Particle Data Group), *Review of Particle Physics*,
J. Phys. G **37** (2010), 075021.

[16] BABAR Collaboration,
Measurement of the electron energy spectrum and its moments in inclusive $B \to X e \nu$ decays,
Phys. Rev. D **69.111104** (11 2004).

[17] DELPHI Collaboration, *Determination of heavy quark non-perturbative parameters from spectral moments in semileptonic B decays*, Eur. Phys. J. **C45** (2006) 35,
arXiv: hep-ex/0510024 [hep-ex].

[18] Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2014*, (2014), arXiv: 1412.7515 [hep-ex].

[19] CDF Collaboration, *Measurement of Ratios of Fragmentation Fractions for Bottom Hadrons in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$-TeV*, Phys. Rev. **D77.072003** (2008), arXiv: 0801.4375 [hep-ex].

[20] ATLAS Collaboration, 'Tagging and suppression of pileup jets with the ATLAS detector',
tech. rep. ATLAS-CONF-2014-018, CERN, 2014,
URL: https://cds.cern.ch/record/1700870.

[21] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159,
arXiv: 1410.3012 [hep-ph].

[22] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*,
Eur. Phys. J. **C77** (2017) 317, arXiv: 1611.09661 [hep-ex].