Measurements of $b$-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: The efficiency to identify jets containing $b$-hadrons ($b$-jets) is measured using a high purity sample of dileptonic top quark-antiquark pairs ($t\bar{t}$) selected from the $36.1\text{fb}^{-1}$ of data collected by the ATLAS detector in 2015 and 2016 from proton-proton collisions produced by the Large Hadron Collider at a centre-of-mass energy $\sqrt{s} = 13$ TeV. Two methods are used to extract the efficiency from $t\bar{t}$ events, a combinatorial likelihood approach and a tag-and-probe method. A boosted decision tree, not using $b$-tagging information, is used to select events in which two $b$-jets are present, which reduces the dominant uncertainty in the modelling of the flavour of the jets. The efficiency is extracted for jets in a transverse momentum range from 20 to 300 GeV, with data-to-simulation scale factors calculated by comparing the efficiency measured using collision data to that predicted by the simulation. The two methods give compatible results, and achieve a similar level of precision, measuring data-to-simulation scale factors close to unity with uncertainties ranging from 2% to 12% depending on the jet transverse momentum.

KEYWORDS: Hadron-Hadron scattering (experiments)

ArXiv ePrint: 1805.01845
1 Introduction

The identification of jets containing $b$-hadrons, referred to as $b$-jets, is vital for a large part of the physics programme of the ATLAS experiment [1] at the CERN Large Hadron Collider (LHC), including Standard Model (SM) precision measurements, studies of the Higgs boson’s properties and searches for new physics beyond the SM. The algorithms used to identify $b$-jets are referred to as $b$-tagging algorithms.

This paper describes a measurement of the $b$-jet tagging efficiency in proton-proton collision data recorded at $\sqrt{s} = 13$ TeV during Run 2 of the LHC. A very pure $t\bar{t}$ sample is selected, as these events have a high $b$-jets purity by virtue of the $t \rightarrow Wb$ branching fraction being close to 100% [2]. The number of additional non-$b$-jets in the sample is greatly reduced by requiring that both $W$ bosons decay leptonically. Two methods are used to measure the $b$-jet tagging efficiency: a new method which uses a tag-and-probe approach, referred to as the Tag-and-Probe method (T&P); and a combinatorial likelihood
approach, referred to as the Likelihood method (LH), which is based upon a method used during Run 1 ($\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV) of the LHC [3]. Having two methods enables reciprocal cross-checks to be made between them.

The $b$-jet tagging efficiency, $\varepsilon_b$, is measured for jets in the pseudorapidity range $|\eta| < 2.5$ and with transverse momentum $p_T > 20$ GeV for several operating points (OP). Operating points are defined by sets of selection criteria imposed upon the output of the $b$-tagging algorithm designed to provide a certain $b$-jet tagging efficiency. Four operating points are defined, corresponding to 60%, 70%, 77% and 85% $b$-jet tagging efficiencies in simulated $t\bar{t}$ events. Two sets of four operating points are implemented to provide a single-cut or a flat-efficiency operating point. The single-cut operating point provides the stated $b$-jet tagging efficiency when averaged over the transverse momentum distribution of $b$-jets in $t\bar{t}$ events, but the efficiency varies with jet $p_T$. On the other hand, the flat-efficiency operating point has a varying cut value, ensuring a constant $b$-jet tagging efficiency as a function of the jet $p_T$. Results are also presented in the form of data-to-simulation efficiency scale factors, defined as $\varepsilon_b^{\text{data}}/\varepsilon_b^{\text{sim}}$, where $\varepsilon_b^{\text{data}}$ is the efficiency measured in data, while $\varepsilon_b^{\text{sim}}$ represents the efficiency predicted by simulation using Monte Carlo (MC) generator-level information. In physics measurements, these scale factors can be applied jet by jet to correct the rate of events after applying a $b$-tagging requirement. The scale factors are measured for all operating points; however, this paper presents only results from a number of selected working points as examples. Separate measurements have also been made for the tagging efficiencies of jets containing $c$-hadrons, referred to as $c$-jets, and for jets containing neither a $b$-hadron nor a $c$-hadron, referred to as light-flavour jets, and are presented in ref. [3].

The paper is organised as follows. In section 2, the ATLAS detector and physics object reconstruction are described. Section 3 contains a description of the ATLAS $b$-tagging algorithms. In section 4, the data and simulated samples used in the $b$-jet tagging efficiency measurements are presented. Section 5 summarises the event selection criteria applied for both calibration methods, while in section 6 the T&P and LH methods are presented in detail. In section 7, the systematic uncertainties for each method are outlined, and results are presented in section 8. Finally, conclusions are given in section 9.

## 2 The ATLAS detector and object reconstruction

The ATLAS detector [1] at the LHC covers nearly the entire solid angle around the collision point. The detector comprises an inner tracking detector surrounded by a superconducting solenoid producing a 2 T axial magnetic field, a system of calorimeters, and a muon spectrometer (MS) incorporating three large toroid magnet assemblies. The inner detector (ID) consists of four layers of silicon pixel sensors and four layers of silicon microstrip sensors,
providing precision tracking in the pseudorapidity range $|\eta| < 2.5$. The innermost pixel layer, referred to as the insertable B-layer (IBL) [4, 5], was installed between Run 1 and Run 2 of the LHC. The IBL provides a hit measurement at an average radius of 33.3 mm, significantly closer to the interaction point than the closest pixel layer in Run 1 (radius of 50.5 mm). The additional pixel layer has a significant impact on the performance of both the tracking and vertexing algorithms, resulting in improved $b$-tagging performance. A straw-tube transition radiation tracker complements the measurements in the silicon layers by providing additional tracking and electron identification information for $|\eta| < 2.0$.

High-granularity electromagnetic (EM) and hadronic sampling calorimeters cover the region $|\eta| < 4.9$. All electromagnetic calorimeters, as well as the endcap and forward hadronic calorimeters, use liquid argon as the active medium and lead, copper, or tungsten absorber. The central hadronic calorimeter uses scintillator tiles as the active medium and steel absorber. The muon spectrometer measures the deflection of muons with $|\eta| < 2.7$ using multiple layers of high-precision tracking chambers located in a toroidal field of approximately 0.5 T or 1 T in the central and endcap regions of ATLAS, respectively.

The ATLAS detector incorporates a two-level trigger system, with the first level implemented in custom hardware and the second level implemented in software. This trigger system reduces the output from the detector electronics to about 1 kHz for offline storage.

Vertices are reconstructed using tracks measured by the inner detector [6]. Events are required to have at least one reconstructed vertex, with two or more associated tracks which have $p_T > 400$ MeV. The primary vertex is chosen as the vertex candidate with the largest sum of the squared transverse momenta of associated tracks.

Electrons are reconstructed from energy deposits (clusters) in the electromagnetic calorimeter matched to tracks reconstructed in the ID [7, 8]. Additionally, candidate clusters in the calorimeter barrel/endcap transition region, defined by $1.37 < |\eta_{\text{cluster}}| < 1.52$, as well as those of poor quality, are excluded. Muons are reconstructed from track segments in the MS that are matched to tracks in the ID [9, 10]. Combined muon tracks are then re-fit using information from both the ID and MS systems. The lepton tracks must be consistent with coming from the primary vertex of the event: the longitudinal impact parameter $z_0$ must satisfy $|z_0 \sin \theta| < 0.5$ mm, while the transverse impact parameter significance, $|d_0|/\sigma(d_0)$ must be less than 5 for electrons or less than 3 for muons. To reduce the contribution from hadronic decays (non-prompt leptons), photon conversions and hadrons misidentified as leptons, both the electrons and muons must also satisfy isolation and identification criteria. The loose, medium and tight working points of the isolation and identification algorithms are defined in ref. [8] for electrons, and in ref. [10] for muons. Two types of leptons are defined for the analyses presented in this paper. First, signal lepton candidates are required to have $p_T > 27$ GeV and $|\eta| < 2.5$, as well as to satisfy tight track- and calorimeter-based isolation criteria. Signal electrons (muons) are required to pass the medium electron (muon) identification criteria. Second, loose leptons are required to have $p_T > 7$ GeV and $|\eta| < 2.5$, as well as to satisfy loose identification and loose track-only isolation criteria.

Jets are reconstructed from three-dimensional topological energy clusters [11] in the calorimeter using the anti-$k_t$ algorithm [12] with a radius parameter of $R = 0.4$. These
jets are referred to as calorimeter-jets. The clusters are calibrated to the electromagnetic energy response scale prior to jet reconstruction. The reconstructed jets are then calibrated to the jet energy scale (JES), corresponding to the particle scale, using corrections derived from simulation and in situ corrections based on 13 TeV data [13]. Jets are required to have calibrated \( p_T > 20 \text{ GeV} \) and to be within the acceptance of the inner detector, \( |\eta| < 2.5 \). Jet cleaning criteria are applied to identify jets arising from non-collision sources or noise in the calorimeter [14, 15]. Any event containing such a jet is removed. In order to reduce the contamination from jets arising from additional pp collisions in the same or nearby bunch crossings, called pile-up, a requirement on the Jet Vertex Tagger (JVT) [16] output is made. The JVT algorithm combines tracking information into a multivariate algorithm to reject jets which do not originate from the primary vertex, and is applied to jets with \( p_T < 60 \text{ GeV} \) and \( |\eta| < 2.4 \). Jets with \( p_T > 60 \text{ GeV} \) are assumed to have originated from the primary vertex.

Jets are also reconstructed from inner-detector tracks using the anti-\( k_t \) algorithm with a radius parameter of \( R = 0.2 \). These jets are referred to as track-jets. The tracks used in jet clustering are required to have \( p_T > 0.5 \text{ GeV} \) and to be matched to the primary vertex using impact parameter requirements on the tracks. Only track-jets with at least two tracks and with \( p_T > 10 \text{ GeV} \) and \( |\eta| < 2.5 \) are considered for the purposes of the \( b \)-jet tagging efficiency measurement. The results presented in this paper correspond to the jets reconstructed from the topological energy clusters in the calorimeter, which are referred to as jets throughout. Equivalent \( b \)-jet efficiency measurements are also performed for track-jets and the results made available to ATLAS analyses using those jets.

In order to avoid counting a single detector response as originating from two different objects, an overlap removal procedure is applied to the jet candidates and leptons passing the loose quality requirement. To prevent double-counting of electron energy deposits reconstructed as jets, the closest jet lying \( \Delta R < 0.2 \) from a selected electron is removed. Electron candidates that lie \( \Delta R < 0.4 \) from a jet surviving the selection are discarded to reduce the background from electrons that originate from heavy-flavour decays. Furthermore, to reduce the background from muons that originate from the decays of hadrons containing a heavy quark inside selected jets, muon candidates are removed if they are separated from the nearest selected jet by \( \Delta R < 0.4 \). However, if this jet has fewer than three associated tracks, the muon is kept and the jet is removed as it is likely that the energy is deposited in the calorimeter by the muon.

The missing transverse momentum (with magnitude \( E_T^{\text{miss}} \)) is defined as the negative vector sum of the transverse momenta of all selected and calibrated physics objects in the event, with an extra “soft” term added to account for low-momentum contributions from particles in the event that are not associated with any of the selected objects. This term is calculated using inner-detector tracks matched to the primary vertex to reduce the pile-up contamination [17].

\[ \text{The particle scale is defined as consisting of stable particles emerging from the p-p collision before interaction with the ATLAS detector.} \]
3 Definition of $b$-tagging algorithms

A new multivariate $b$-tagging algorithm, referred to as MV2c10, was developed for Run 2, and utilises a boosted decision tree (BDT). The algorithm is similar to the multivariate algorithms developed during Run 1 [3], but with a dedicated optimisation carried out for Run 2 to exploit the installation of the IBL and improved tracking software [18, 19]. The algorithms which provide the input variables for MV2c10 all exploit the relatively long $b$-hadron lifetime: a likelihood-based combination of the transverse and longitudinal impact parameter significances; the presence of a secondary vertex and related properties; and the reconstruction of the $b$-hadron decay chain using a Kalman filter to search for a common direction connecting the primary vertex to both the bottom and the tertiary charm decay vertices. Additionally, the jet $p_T$ and jet $\eta$ are included as BDT training variables to take advantage of correlations with other variables. In order to avoid any difference between the kinematic spectra of $b$-jets and background jets being used as a discriminating variable, the $b$-jet $p_T$ and $\eta$ spectra are reweighted to match the combined $c$-jet and light-flavour jet spectrum. The BDT was trained on a subset of events from a simulated $t\bar{t}$ sample, produced with POWHEG [20–23] interfaced with PYTHIA6 for the parton shower, hadronisation, and the underlying event [24] and using the CT10 [25] parton distribution function set, as described in more detail in section 4. The BDT training is performed by assigning $b$-jets as signal, and $c$-jets and light-flavour jets as background. In order to enhance the $c$-jet rejection, the $c$-jet fraction in the training is set to 7%, and the light-flavour jet background is set to 93%, as described in ref. [19].

The MV2c10 output for $b$-jets, $c$-jets and light-flavour jets in a $t\bar{t}$ sample, which is statistically independent from the training sample, is presented in figure 1(a). The rejection rates for light-flavour jets and $c$-jets are defined as the inverse of the efficiency for tagging a light-flavour jet or a $c$-jet as a $b$-jet, respectively. Figure 1(b) shows the corresponding light-flavour jet and $c$-jet rejection factors as a function of the $b$-jet tagging efficiency. The rejection rates for both the light-flavour jets and $c$-jets as a function of jet $p_T$ are given in figure 2(a) for the single-cut OP and figure 2(b) for the flat-efficiency OP, both for a 70% $b$-jet tagging efficiency.

4 Dataset and simulated event samples

The data used in these measurements were collected by the ATLAS detector from proton-proton collisions in 2015 and 2016 at a centre-of-mass energy $\sqrt{s} = 13$ TeV with 25 ns proton bunch spacing. The data correspond to a total integrated luminosity of $36.1 \pm 1.2$ fb$^{-1}$ after offline data quality selection, measured following ref. [26].

The dominant $t\bar{t}$ process was modelled using the matrix-element generator POWHEG-BOX v2 [20–23], which provides next-to-leading-order (NLO) accuracy in QCD. It used the CT10 parton distribution function (PDF) set [25] and it was interfaced to PYTHIA 6.428 [24] with the Perugia 2012 set [27] of tuned parameters (tune) for the modelling of the parton shower, fragmentation and the underlying event. The $h_{\text{damp}}$ parameter, which controls the $p_T$ of the first additional emission beyond the Born configuration, was set to
Figure 1. (a) The MV2c10 output for $b$-jets (solid line), $c$-jets (dashed line) and light-flavour jets (dotted line) in simulated $t\bar{t}$ events. (b) The light-flavour jet (dashed line) and $c$-jet rejection factors (solid line) as a function of the $b$-jet tagging efficiency of the MV2c10 $b$-tagging algorithm. The performance was evaluated on $t\bar{t}$ events simulated using POWHEG interfaced to PYTHIA6.

Figure 2. The light-flavour jet (squares) and $c$-jet rejection factors (triangles) at a $b$-tagging efficiency of 70% corresponding to (a) the single-cut OP and (b) the flat-efficiency OP as a function of the jet $p_T$ for the MV2c10 $b$-tagging algorithms in $t\bar{t}$ events simulated using POWHEG interfaced to PYTHIA6.

$m_t = 172.5$ GeV, a setting that was found to improve the description of the $p_T$ of the $t\bar{t}$ system when compared to data [28].

The dominant non-$t\bar{t}$ process is the associated production of a single top quark and a $W$ boson ($Wt$ process), which also contains a large fraction of $b$-jets. Other processes
Table 1. A summary of Monte Carlo generators used to simulate various physics processes, together with their basic parameter settings and corresponding cross-section order in pQCD at $\sqrt{s} = 13$ TeV. Whenever Pythia or Herwig++ is used for parton shower simulation, the parton shower PDFs are taken from CTEQ6L1.

such as $Z/\gamma^*$ + jets and diboson production constitute only a small fraction of the sample. Table 1 summarises the nominal MC generators used to simulate physics processes, along with the alternative samples used to estimate the systematic uncertainties related to the choice of MC generator and associated parameter settings. The contamination from $t\bar{t}$ produced in association with either a vector boson or a SM Higgs boson, and the gluon-gluon fusion or vector-boson fusion production of a Higgs boson, which subsequently decays into a pair of $W$ bosons, were found to be negligible. Therefore, these processes are not considered further.

Events in which one of the two selected lepton candidates is not a real prompt lepton (e.g. one coming from a $b/c$ hadron decay, a photon conversion, or a hadron misidentified as a lepton) are referred to as misidentified lepton events, and are estimated from data by changing the selection criterion from opposite-charge to same-charge leptons. In order to avoid double counting the contribution of misidentified leptons which is already reproduced by Monte Carlo, and to partially take into account a possible difference between opposite-charge and same-charge lepton events, the contribution from simulated same-charge events is subtracted from data for the estimate of the misidentified lepton background.
The EvtGen [41] package was used with all the hadronisation/fragmentation generators, except for Sherpa, to model the decays of $b$- and $c$-hadrons. All simulated samples were processed through the ATLAS detector simulation [42] based on GEANT4 [43]. Additional simulated $pp$ collisions generated with Pythia8 [35] were overlaid on all simulated samples to model the expected number of additional pile-up interactions in each event. Simulated events are corrected so that the lepton and jet identification efficiencies, energy scales and energy resolutions match those determined from data control samples at $\sqrt{s} = 13$ TeV.

In the simulated samples, a reconstructed jet is labelled as a $b$-jet if, within $\Delta R = 0.3$, there is a matching weakly decaying $b$-hadron with $p_T > 5$ GeV. The flavour labelling is exclusive, with the hadron matched to the closest jet in the $\Delta R$ phase space. If no $b$-hadron is found, but a $c$-hadron is matched to the jet, then it is labelled as a $c$-jet. If there is no $b$- or $c$-hadron, but a $\tau$-lepton is matched to the jet, then it is labelled as a $\tau$-jet, otherwise it is labelled as a light-flavour jet.

5 Event selection

Events were recorded using triggers requiring at least one lepton, with lepton isolation requirements and $p_T$ thresholds that vary depending on the data-taking conditions. In 2015 this threshold was 20 GeV for muons and 24 GeV for electrons, while in 2016 it was raised to 26 GeV for both the electrons and muons. These triggers are combined with higher-threshold triggers, of 50 GeV for muons and 60, 120 and 140 GeV for electrons, without isolation requirements, to improve the trigger efficiency for leptons with high transverse momentum.

Table 2 summarises the event selection criteria specific to the Likelihood and T&P methods. The events are selected by requiring two oppositely charged signal leptons ($e\mu$, $ee$ or $\mu\mu$) and two or three jets. One of the leptons must also be matched using a $\Delta R$ requirement to one of the objects that triggered the event. A veto is applied to events which contain one or more additional loose leptons. The T&P method uses only the $t\bar{t} \rightarrow e\mu\nu + 2$-jet category, with an additional requirement that at least one jet must be tagged by the MV2c10 algorithm at the 85% single-cut efficiency OP. The LH method also exploits events with exactly three jets, as well as events with same-flavour leptons in the final state. For events with same-flavour leptons in the final state, additional requirements of $E_T^{\text{miss}} > 60$ GeV and dilepton invariant mass $50 < m_{ll} < 80$ GeV or $m_{ll} > 100$ GeV are applied to suppress the contamination from on-shell $Z$ boson decays, multijet production and decays of $\gamma^*$, $\Upsilon$ and $J/\psi$ particles. In the $e\mu + 2$-jet channel of the LH (T&P) method, a $t\bar{t}$ purity of 82% (90%) is obtained, with sub-dominant contributions from single top, $Z/\gamma^* +$ jets and diboson processes and events containing a misidentified lepton. In the $e\mu + 3$-jet, $\ell\ell + 2$-jet and $\ell\ell + 3$-jet channels of the LH method, the $t\bar{t}$ contribution is 88%, 71% and 79%, respectively. Figure 3 shows the jet $p_T$ and $m_{ll}$ distributions for events passing the $e\mu\nu + 2$-jet selection before any requirement on the MV2c10 output is applied. Good agreement between the simulation and data is observed within the total statistical and systematic uncertainties.
Selection Requirement | Likelihood Method | T&P Method
--- | --- | ---
Leptons | 2 oppositely charged signal leptons ($e, \mu$) | 
Jets | 2 or 3 jets | 2 jets |
Region-specific cuts | $E_T^{\text{miss}}> 60$ GeV | At least 1 $b$-tagged jet \[(\text{at } 85\% \text{ efficiency OP})\] |
BDT cut | $D_{bb}^{\text{LH}}> 0.1$ | $D_{bb}^{\text{T&P}}> -0.02$ |

Table 2. The analysis regions and associated event selection criteria for the LH and T&P methods. The variables $D_{bb}^{\text{LH}}$ and $D_{bb}^{\text{T&P}}$ are BDT output discriminants trained to separate final states with at least two $b$-jets from all other events, in the LH and T&P methods, respectively. More details of the training and performance of the BDTs is presented in section 5.1.

Figure 3. The (a) jet $p_T$ distribution and (b) $m_{\ell\ell}$ distribution for the data (points) and simulated samples (stacked histograms) for the $e\mu+2$-jet selection in the LH method. The simulated samples are normalised to agree with a fit to the data as described in section 6.2. The bottom panels show the ratio of the data to the simulated samples, with the dotted uncertainty bands corresponding to the total MC statistical uncertainty and systematic uncertainty.

For the purposes of the calibration methods, simulated events are categorised using generator level information according to the flavours of the selected reconstructed jets. For the two-jet selection, three possible jet flavour combinations are considered: $bb$, $bj$ and $jj$, where $b$ represents a $b$-jet, and $j$ is defined as a non-$b$-jet. In the selected events, the $c$-jet contribution is sufficiently small that it can be considered together with the light-flavour jet component. In the same manner, for the three-jet selection, four possible jet flavour combinations are considered: $bbb$, $bbj$, $bjj$ and $jjj$. The dominant $t\bar{t}$ process contributes mostly to the $bb$ and $bj$ flavour combinations in the 2-jet channels, and to the $bbj$ and $bjj$ flavour combinations in the 3-jet channels. After adding all background processes, especially from $Z/\gamma^*+\text{jets}$ decays, the fraction of non-$b$-flavour jets increases. The overall
$b$-jet purity is calculated from simulation as the fraction of jets in the sample that are labelled as $b$-jets. In the selected samples of $e\mu$ events, it is found to be 72% (82%) in the two-jet case for the LH (T&P) method, while in the three-jet case it is 53%. For the events with $ee/\mu\mu$ final states, the overall $b$-jet purity is 62% and 48% in the two-jet and three-jet cases, respectively. As the number of background events containing at least one misidentified lepton is estimated from data, and therefore the jet flavour composition is unknown, it is assumed that only non-$b$-jets are produced in these events. A cross-check was performed assuming that only $b$-jets are produced in these events, and found to have a negligible effect on the calibration results.

5.1 Multivariate event discriminant

In order to further enhance the $b$-jet purity of the selected samples in both methods, boosted decision trees (BDT) were trained using simulated events to separate final states with at least two $b$-jets, defined as the signal, from all other events, classified as a background. Each of the input variables is designed to select events with at least two $b$-jets based upon the topology and kinematics of the event, rather than exploiting any flavour-tagging-related properties of the jets, to ensure minimal bias in the MV2c10 discriminant. A dedicated optimisation was performed for each method, leading to a different choice of input variables, as shown in table 3. In both the LH and T&P methods, the BDTs are trained using the Toolkit for Multivariate Data Analysis, TMVA [44].

In the LH method, sample-selection BDTs are trained separately for the two- and three-jet categories. The training samples include not only the nominal $tt$ sample, but also the alternative $tt$ samples used for evaluating systematic uncertainties. The inclusion of the alternative $tt$ samples provides a larger training sample and allows the BDT to learn the topologies of events generated using alternative MC generators, helping to minimise uncertainties in the calibration due to the choice of generator.

For both the LH and T&P methods, all input variables, as well as $D_{bb}^{LH}$ and $D_{bb}^{T&P}$, are well modelled in simulation. The modelling of $D_{bb}^{LH}$ is shown in figure 4(a), combining all four event categories used in the LH method. Good agreement between the simulation and data is observed in all four channels. Only events with $D_{bb}^{LH} > 0.1$ are considered in the measurement. This threshold is found to be optimal, as it minimises the total uncertainty in the measurement of the calibration scale factor.

The modelling of $D_{bb}^{T&P}$ is shown in figure 4(b). The selection requirement on the BDT output was optimised such that it minimises the uncertainty associated with the data-to-simulation scale factors, and is found to be $D_{bb}^{T&P} > -0.02$.

Figure 5 shows the fraction of $b$-jets in bins of jet $p_T$ both before and after applying a requirement on $D_{bb}^{LH}$ ($D_{bb}^{T&P}$) for the nominal and alternative $tt$ generators in the $e\mu + 2$-jet category of the LH (T&P) method.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>T&amp;P</th>
<th>LH</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ell_1 p_T )</td>
<td>Leading lepton ( p_T )</td>
<td>( \times )</td>
<td>( \times )</td>
</tr>
<tr>
<td>Jet_1 ( p_T )</td>
<td>Leading jet ( p_T )</td>
<td>( \times )</td>
<td>( \times )</td>
</tr>
<tr>
<td>Jet_2 ( p_T )</td>
<td>Sub-leading jet ( p_T )</td>
<td>( \times )</td>
<td>( \times )</td>
</tr>
<tr>
<td>Jet_3 ( p_T )</td>
<td>Third-leading jet ( p_T ) (events with 3 jets only)</td>
<td>( \times )</td>
<td></td>
</tr>
<tr>
<td>( E_T^{\text{miss}} )</td>
<td>Missing transverse momentum</td>
<td>( \times )</td>
<td></td>
</tr>
<tr>
<td>( n_{\text{Fjets}} )</td>
<td>Number of jets with ( 2 &lt;</td>
<td>\eta</td>
<td>&lt; 4.5 )</td>
</tr>
<tr>
<td>( \Delta \phi(j_1,j_2) )</td>
<td>( \Delta \phi ) of leading 2 jets</td>
<td>( \times )</td>
<td></td>
</tr>
<tr>
<td>( \min \Delta R(j,j) )</td>
<td>Minimum ( \Delta R ) of all jet combinations</td>
<td>( \times )</td>
<td></td>
</tr>
<tr>
<td>Imbalance</td>
<td>((\text{jet}_1(p_T) - \text{jet}_2(p_T))/\text{jet}_1(p_T)) + \text{jet}_2(p_T))</td>
<td>( \times )</td>
<td></td>
</tr>
<tr>
<td>( m_{\text{avg}}(\ell_j) )</td>
<td>( \min((m(\ell_1,\text{jet}_i) + m(\ell_2,\text{jet}_j))/2) ), ( i,k = 1,2 ) (1, 2, 3) for events with 2 (3) jets</td>
<td>( \times )</td>
<td></td>
</tr>
<tr>
<td>( \min \Delta R(\ell_1,j) )</td>
<td>Minimum ( \Delta R ) separation of leading lepton from all jets</td>
<td>( \times )</td>
<td></td>
</tr>
<tr>
<td>( \min \Delta R(\ell_2,j) )</td>
<td>Minimum ( \Delta R ) separation of subleading lepton from all jets</td>
<td>( \times )</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Input variables used in the T&P and LH method BDT algorithms. The “\( \times \)” symbol in the T&P and LH columns indicates the BDT in which each variable is used. Quantities involving jets labelled \( j \) correspond to central jets, with \( |\eta| < 2.5 \).

**Figure 4.** The sample-selection BDT output distribution for the data (points) and simulated samples (stacked histograms) in (a) the combined \( ee/\mu\mu + 2/3 \)-jets sample used in the LH method and (b) for the \( e\mu + 2 \)-jets sample used in the T&P method. The simulated samples in the LH method are normalised to agree with a fit to the data as described in section 6.2, while in the T&P method, the normalisation is taken from the theoretical predictions. The cut applied to the \( D_{bb}^{\text{LH}} (D_{bb}^{\text{T&P}}) \) discriminant in the LH (T&P) method is indicated by a vertical dotted line. The bottom panels show the ratio of the data to the simulated samples, with the shaded uncertainty bands corresponding to the total MC statistical uncertainty and systematic uncertainty.
Figure 5. The fraction of $b$-jets in the selected sample as a function of the jet $p_T$, (a) before and (b) after the requirement on $D_{bb}^{\text{LH}}$, for the nominal and alternative $tt$ generators in the $e\mu + 2$-jet category of the LH method. The fraction of $b$-jets in the selected sample as a function of the jet $p_T$, (c) before and (d) after the requirement on $D_{bb}^{\text{T&P}}$, for the nominal and alternative $tt$ generators in the $e\mu + 2$-jet category of the T&P method.

6 Calibration methods

6.1 Tag-and-probe method

The sample used by the T&P method is 90% pure in $t\bar{t}$ events, providing a high-purity sample of $b$-jets. The $b$-jet tagging efficiency measurement is performed on a set of probe jets, where a jet is considered a probe jet if the other jet is $b$-tagged at the 85% efficiency OP.

The MV2c10 distributions for probe jets in data and simulation are presented in figure 6, broken down by both process and probe jet flavour. Good agreement between
simulation and data is observed in the $b$-jet dominated MV2c10 output region. In the light-flavour jet dominated MV2c10 output region, some disagreement between simulation and data is observed, however, this is accounted for by the light-flavour jet SFs and associated uncertainties [3]. The $b$-tagging efficiency is measured by first determining the fraction of the probe jets that satisfy a given $b$-tagging criterion, $f_{\text{tagged}} = N_{\text{pass}} / N$, where $N$ is the total number of probe jets and $N_{\text{pass}}$ is the number of probe jets satisfying the criterion. This fraction is measured in data by subtracting the contribution from non-$t\bar{t}$ processes, as predicted by simulation.

$$f_{\text{tagged}} = \frac{N_{\text{pass}}}{N_{\text{data}}} - \frac{N_{\text{pass}}}{N_{\text{non}-t\bar{t}, \text{sim}}}$$

where $N_{\text{data}}$ is the number of probe jets in data and $N_{\text{non}-t\bar{t}, \text{sim}}$ is the number of probe jets from non-$t\bar{t}$ events predicted by simulation. It should be noted that non-$t\bar{t}$ processes can contribute $b$-jets as well as light-flavoured jets.

The determination of the $b$-jet tagging efficiency relies on the assumption that the fraction of probe jets containing a tagged jet in data, $f_{\text{tagged}}$, is given by

$$f_{\text{tagged}} = f_b \varepsilon_b + (1 - f_b) \varepsilon_j,$$

where $f_b$ and $(1 - f_b)$ are the fractions of $b$-jets and non-$b$-jets in $t\bar{t}$ events, and $\varepsilon_b$ and $\varepsilon_j$ are the $b$-jet and non-$b$-jet tagging efficiencies. The $b$-jet tagging efficiency can be determined by measuring $f_{\text{tagged}}$ in data, and estimating the other parameters from simulation. In this approach, jets are considered as uncorrelated objects within an event. The $b$-tagging efficiency is extracted from $f_{\text{tagged}}$ in bins of the jet $p_T$

$$\varepsilon_b = \frac{f_{\text{tagged}} - (1 - f_b) \varepsilon_j}{f_b}.$$  

The main sources of systematic uncertainty arise from the determination of $f_b$ and $\varepsilon_j$, as both of these parameters are taken from simulation.

### 6.2 Combinatorial likelihood method

The LH method is performed separately for the $e\mu$ and combined $ee/\mu\mu$ final states in the two- and three-jet bin categories due to differences in the background composition. The yields of $t\bar{t}$ and $Z/\gamma^* + \text{jets}$ events are normalised using dedicated control samples in data. The selection criteria applied to define these control regions are given in table 4. Normalisation factors are determined simultaneously with the maximum-likelihood fit to the observed number of events in all regions, and used to scale the overall normalisation of the $t\bar{t}$ and $Z/\gamma^* + \text{jets}$ processes. The values of the fitted normalisation factors, along with their associated statistical uncertainties, are also presented in table 4.

Unlike the T&P method, the LH method exploits the per-event jet correlations. For example, in the case of events with two jets, the fraction of events with one $b$-tagged jet, $f_{1\text{-tag}}$, or two $b$-tagged jets, $f_{2\text{-tag}}$, can be measured in data using

$$f_{1\text{-tag}} = 2f_{bb}\varepsilon_b (1 - \varepsilon_b) + f_{bj} \left[ \varepsilon_j (1 - \varepsilon_b) + (1 - \varepsilon_j) \varepsilon_b \right] + (1 - f_{bb} - f_{bj}) 2\varepsilon_j (1 - \varepsilon_j)$$

$$f_{2\text{-tag}} = f_{bb}\varepsilon_b^2 + f_{bj} \varepsilon_j \varepsilon_b + (1 - f_{bb} - f_{bj}) \varepsilon_j^2.$$
Figure 6. The MV2c10 distribution of the probe jets used for the calibration in the T&P method, (a) broken down by process, and (b) probe jet flavour.

<table>
<thead>
<tr>
<th>Control sample</th>
<th>$e\mu$</th>
<th>$ee/\mu\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$tt$</td>
<td>$e\mu$ selections before $D_{bb}^{LH}$ requirement</td>
<td>$ee/\mu\mu$ selections before $D_{bb}^{LH}$ requirement</td>
</tr>
<tr>
<td>NF (2-jet)</td>
<td>$0.924 \pm 0.004$</td>
<td>$0.906 \pm 0.007$</td>
</tr>
<tr>
<td>NF (3-jet)</td>
<td>$1.010 \pm 0.004$</td>
<td>$0.986 \pm 0.008$</td>
</tr>
<tr>
<td>$Z/\gamma^* +$ jets</td>
<td>Same lepton flavour, $80 &lt; m_{\ell\ell} &lt; 100$ GeV</td>
<td>Same lepton flavour, $E_T^{miss} &gt; 60$ GeV, $80 &lt; m_{\ell\ell} &lt; 100$ GeV</td>
</tr>
<tr>
<td>NF (2-jet)</td>
<td>$1.104 \pm 0.001$</td>
<td>$1.096 \pm 0.005$</td>
</tr>
<tr>
<td>NF (3-jet)</td>
<td>$1.258 \pm 0.002$</td>
<td>$1.281 \pm 0.008$</td>
</tr>
</tbody>
</table>

Table 4. Definition of the control samples in data used for the determination of the normalisation factors for $tt$ and $Z/\gamma^* +$ jets processes in the LH method. The obtained normalisation factors (NFs) with their associated statistical uncertainties in each category are also presented.

where $f_{bj}$ and $f_{bb}$ are the fraction of events with one $b$-jet and two $b$-jets, respectively, and $\varepsilon_b$ ($\varepsilon_j$) is the $b$- (light-flavour) jet tagging efficiency. Using these equations, $\varepsilon_b$ is determined by measuring $f_{1\text{-tags}}$ and $f_{2\text{-tags}}$ from data, with $f_{bj}$, $f_{bb}$ and $\varepsilon_j$ determined from MC. In this way, the correlation of the jet flavour information in the 1-tag and 2-tag regions is added, resulting in a more precise efficiency measurement. A measurement in $N$ kinematic bins results in $2 \times N^2$ coupled equations. It is possible to solve such a system of non-linear equations, but in practice it is much simpler to model the same system by using a more flexible and powerful likelihood function and solve the system numerically by maximising the likelihood.

Using the probability density functions, $P$, the per-event likelihood term for the two jets in the event to have transverse momenta $p_{T,1}$ and $p_{T,2}$ and MV2c10 weight outputs
where $P_{f_1 f_2}(p_{T,1}, p_{T,2})$ is a two-dimensional probability density function for jets of flavour $f_1$ and $f_2$ to have transverse momenta $p_{T,1}$ and $p_{T,2}$, and $P_f(w; p_T)$ is a probability density function of the $b$-jet tagging weight for a jet of flavour $f$ at a given $p_T$. The factors $f_{bb}$, $f_{bj}$, and $f_{jj} = 1 - f_{bb} - f_{bj}$ are the overall flavour fractions in the two-jet case. All probability density functions are determined from simulation, except the one for the $b$-jet weight, which contains the information to be extracted from data.

The $b$-jet tagging efficiency corresponding to the MV2c10 weight cut of $w_{\text{cut}}$ is given by

$$
\varepsilon_b(p_T) = \int_{w_{\text{cut}}}^{\infty} dw' P_b(w', p_T).
$$

For the extraction of the $b$-jet tagging efficiency for a single OP, $P_b(w', p_T)$ corresponds to a histogram with two $w'$ bins for each jet $p_T$ bin. The bin above $w_{\text{cut}}$ corresponds to the $b$-jet tagging efficiency.

For events containing three jets, the likelihood is constructed in a way analogous to the two-jet case, resulting in six equivalent likelihood terms instead of four.

A closure test was performed by applying the full method to the simulated events. This sample of simulated events is normalised to an integrated luminosity of 36.1 fb$^{-1}$, and is treated as “pseudo data” with the expected number of events in each bin taken as the mean of a Poisson distribution to estimate the statistical uncertainty. The resulting scale factors are close to unity within the statistical uncertainty of the pseudo data sample for all bins except the lowest jet-$p_T$ bin in the 3-jet sample, verifying that the method has no significant bias. An additional 3% uncertainty is added to cover the observed non-closure in the lowest $p_T$ bin for the three-jet sample.

### 7 Systematic uncertainties

Three categories of systematic uncertainty are considered in the measurements presented. First, MC generator modelling uncertainties that affect the modelling of kinematic distributions and the jet flavour composition in simulated events. Second, normalisation uncertainties that account for uncertainties in the cross-section of simulated samples. Third, experimental uncertainties, which are related to detector effects and the reconstruction of the physics objects in the simulated samples.

Uncertainties from the MC generator modelling are evaluated in the simulated $t\bar{t}$, single-top, and $Z/\gamma^* +$ jets samples, by comparing the nominal samples to ones created with alternative generators and settings. The alternative generators induce changes in the event kinematics and flavour composition, thereby affecting the extracted scale factors. The difference between the scale factors is taken as the systematic uncertainty. Uncertainties are estimated for the predictions from the nominal
The $t\bar{t}$ sample (POWHEG+PYTHIA6), by altering the choice of parton shower and hadronisation generator (POWHEG+HERWIG++) or by altering the matrix element generator (MG5_aMC@NLO+HERWIG++). Changing the settings of the nominal generator to increase or decrease the amount of parton radiation and varying the choice of parton distribution function (PDF) set (MG5_aMC@NLO+HERWIG++ with CT10 PDFs or PDF4LHC15_NLO PDFs) gives additional sources of uncertainty. Further uncertainties due to the observed mismodelling of the top quark and \textit{t}\textit{t}-system $p_T$ are evaluated by taking the difference between the default \textit{t}\textit{t} prediction and a sample in which the top quark and \textit{t}\textit{t}-system $p_T$ distributions are reweighted to match predictions at NNLO accuracy in QCD [45, 46]. For $Wt$ single-top production, uncertainties are estimated by varying the parton shower and hadronisation modelling (POWHEG+HERWIG++) and varying settings of the nominal generator to increase or decrease the amount of parton radiation. The uncertainty in the treatment of the interference between $Wt$ and \textit{t}\textit{t}$ is assessed by replacing the nominal diagram removal (DR) scheme with a diagram subtraction (DS) scheme [47] in POWHEG+PYTHIA6. An additional 6% uncertainty is applied to the normalisation of the single-top samples to account for the uncertainty in the predicted cross-section [30]. For the $Z/\gamma^* + \text{jets}$ process, the nominal samples (MG5_aMC@NLO) are compared to the alternative POWHEG+PYTHIA8 sample. In addition an uncertainty is estimated for the modelling of the jet $p_T$ spectrum in the $Z/\gamma^* + \text{jets}$ events with the same-flavour leptons in the final state by reweighting the spectrum to match the data in the $Z/\gamma^* + \text{jets}$ control sample.

To account for the extrapolation of the $Z/\gamma^* + \text{jets}$ normalisation from the control sample to the sample used in the $b$-tagging efficiency measurement, a 20% uncertainty in the $Z/\gamma^* + \text{jets}$ estimate is applied. The size of this uncertainty is determined by comparing the data to MC simulation in the relevant kinematic distributions. An additional 50% uncertainty is applied to the events with at least one $b$- or $c$-jet, as observed in $Z + b$ measurements [48]. Due to the small contribution from the diboson backgrounds, only a normalisation uncertainty is assigned to this sample. This uncertainty is assumed to be 50% in the two-jet channel, and 70% in the three-jet channel, as determined from MC studies. Likewise for the backgrounds with misidentified leptons, only a normalisation uncertainty of 50% is considered. Identical normalisation uncertainties are applied in the LH and T&P methods.

In addition, in the T&P method, which has a tighter event selection than the LH method, the uncertainties arising from the limited size of the simulated samples have a non-negligible effect on the order of 1% on the total scale factor uncertainty. They are evaluated using 10,000 pseudo experiments. In each pseudo experiment the efficiency and the data-to-simulation scale factor in each jet $p_T$ bin is computed. The standard deviation of the scale factor in all of the pseudo experiments is taken as the systematic uncertainty due to limited MC sample size. The impact of MC statistical uncertainties is significantly lower in the more inclusive LH method, and therefore the impact of MC statistical uncertainties in the likelihood model itself is not considered.

The experimental uncertainties include those related to the reconstruction of electrons, muons, jets and $E_T^{miss}$, uncertainties in the mis-tagging of $c$- and light-flavour jets as $b$-jets,
uncertainties in the modelling of pile-up and in the integrated luminosity. For both the electrons [8] and muons [10], uncertainties are estimated for the energy scale and resolution, as well as the reconstruction, identification, and trigger efficiencies using 13 TeV data. The uncertainties in the jet energy scale and resolution are evaluated using 13 TeV data [13], and so is an uncertainty in the efficiency of the JVT selection [16]. The uncertainties in the energy scale and resolution of the jets and leptons are propagated to the calculation of the $E_{\text{T}}^{\text{miss}}$, which also has additional dedicated uncertainties from the momentum scale, resolution and efficiency of the tracks not associated with any of the reconstructed objects, along with the modelling of the underlying event. The predicted rate of mistakenly tagging non-$b$-jets is corrected using data-to-simulation efficiency scale factors measured separately for $c$-jets and light-flavour jets [3]. The uncertainty in this prediction is estimated by varying these scale factors within their associated uncertainties. For Run 2 data, the $c$-jet efficiency scale factor uncertainty varies from $\sim 15\%$ for a jet $p_T$ of 100 GeV, to $\sim 30\%$ for a jet $p_T$ of 300 GeV. For light-flavour jets, the uncertainty varies from $\sim 40\%$ for a jet $p_T$ of 100 GeV, to $\sim 30\%$ for a jet $p_T$ of 300 GeV. The uncertainty due to the reweighting of the distribution of the expected average number of interactions per bunch crossing, $\langle \mu \rangle$, from the simulation to the one measured in data is estimated by varying the nominal reweighting scale factor by the size of the nominal correction. The uncertainty in the combined 2015 and 2016 integrated luminosity is 3.2%. It is derived, following a methodology similar to that detailed in ref. [26], from a preliminary calibration of the luminosity scale using $x$-$y$ beam-separation scans performed in August 2015 and May 2016.

The effect of each source of systematic uncertainty on the $b$-jet tagging efficiency data-to-simulation scale factors is computed by replacing the nominal simulated sample with the sample affected by the systematic variation, and rerunning the fit to data. The uncertainty is taken as a difference relative to the scale factor measured in the nominal case. When combining all four channels in the LH method, all single systematic variations are treated as fully correlated, except for the background modelling uncertainties, for which a 50% correlation is assumed. This partial correlation is applied, as each modelling variation is expected to account for more than one effect.

8 Results

Figure 7 shows the measured efficiency in data and simulation and the data-to-simulation scale factors as a function of the jet $p_T$ for both the T&P and LH methods, corresponding to the 70% $b$-jet tagging efficiency single-cut OP, for $R = 0.4$ calorimeter-jets. The efficiencies determined in simulation and data agree within their uncertainties, resulting in scale factors close to unity. It can be seen that the resulting data-to-simulation scale factors are in agreement between the two methods, with similar central values and uncertainties. Scale factors were also measured as a function of the average number of interactions per bunch crossing, in selected $p_T$ bins, and the jet $\eta$, using both the LH and T&P methods, and are shown in figures 8 and 9, respectively, for the single-cut OP. The data-to-simulation scale factors are observed not to have a strong dependence on either variable. The $b$-jet tagging
efficiency in simulation varies by less than 1% over the range $0 < \langle \mu \rangle < 50$, and by up to 5% of the range $0 < |\eta| < 2.5$.

Tables 5 and 6 show the data-to-simulation scale factors, and the statistical, systematic and total uncertainties separately for each $p_T$ bin. Depending on the $p_T$ bin, the total uncertainties range between 2% and 12% for the LH method and 2% and 9% for the T&P method, with the statistical uncertainty component ranging between 0.3% and 1.8% for the LH method and 0.5% and 2.8% for the T&P method. A reduction in the statistical uncertainty is achieved in the LH method by combining measurements from four channels,
as well as exploiting the correlations in the events, while in the case of the T&P method only the $e\mu + 2$-jet channel is used. The systematic uncertainty component varies from 1.5% to 8.6% depending on the jet $p_T$ for the T&P method, while in the case of the LH method, the effect of systematic uncertainties ranges between 1.8% and 12%. However, in the LH method, the total uncertainty is smaller for a larger jet-$p_T$ range, and this method is therefore used as the default $b$-jet calibration. The dominant sources of uncertainty in both methods relate to the modelling of the $t\bar{t}$ sample and alter the predicted flavour
Figure 9. Data-to-simulation scale factors, corresponding to the 70% $b$-jet tagging efficiency single-cut OP, as a function of the jet $|\eta|$, in (a) the LH method, and (b) the T&P method, for $R = 0.4$ calorimeter-jets. Both the statistical uncertainties (error bars) and total uncertainties (shaded region) are shown.

composition, to which both methods are particularly sensitive. The application of the sample-selection BDT reduces the impact of these uncertainties by up to 50% due to the increase of the $b$-jet purity and the removal of regions of phase space which have large modelling uncertainties. At very low and high jet $p_T$, the uncertainties related to the measurement of the jet energy scale and resolution also become significant. Normalisation and modelling of the $Z/\gamma^* +$ jets background, as well as the normalisation of the diboson backgrounds have a larger effect in the LH method than in the T&P method. This is due to the inclusion of the events with three jets and events with the same-flavour leptons in the final state, as these regions have a larger contribution from the $Z/\gamma^* +$ jets and diboson backgrounds.

An additional uncertainty is included to extrapolate the measured uncertainties to higher jet $p_T$, which is not measured here but is of interest in some physics analyses. This term is calculated from simulated events by considering variations of the quantities affecting the $b$-tagging performance such as the impact parameter resolution, percentage of poorly measured tracks, description of the detector material, and the track multiplicity per jet. The dominant effect on the uncertainty when extrapolating at high $p_T$ is related to the different tagging efficiencies after smearing the tracks’ impact parameters according to the resolutions measured in data and simulation. The difference in the impact parameter resolution is due to effects from alignment, dead modules and additional material not properly modelled in the simulation. The impact of the $b$-tagging efficiency uncertainty increases with jet $p_T$ and reaches 15% above 1.5 TeV.

Similar measurements were conducted for other types of reconstructed jets and other $b$-tagging OPs used within the ATLAS physics program, and are used accordingly. As an additional example, figure 10 presents the measured data-to-simulation scale factors as a
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale factor</td>
<td>1.013</td>
<td>1.035</td>
<td>1.029</td>
<td>1.019</td>
<td>0.984</td>
<td>0.964</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>0.123</td>
<td>0.030</td>
<td>0.018</td>
<td>0.022</td>
<td>0.026</td>
<td>0.037</td>
</tr>
<tr>
<td>Statistical uncertainty</td>
<td>0.012</td>
<td>0.003</td>
<td>0.004</td>
<td>0.004</td>
<td>0.010</td>
<td>0.018</td>
</tr>
<tr>
<td>Systematic uncertainty</td>
<td>0.123</td>
<td>0.030</td>
<td>0.018</td>
<td>0.021</td>
<td>0.024</td>
<td>0.032</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Systematic Uncertainties [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix element modelling (tt)</td>
</tr>
<tr>
<td>Parton shower / Hadronisation (tt)</td>
</tr>
<tr>
<td>NNLO top $p_T$, tt $p_T$ reweighting (tt)</td>
</tr>
<tr>
<td>PDF reweighting (tt)</td>
</tr>
<tr>
<td>More / less parton radiation (tt)</td>
</tr>
<tr>
<td>Matrix element modelling (single top)</td>
</tr>
<tr>
<td>Parton shower / Hadronisation (single top)</td>
</tr>
<tr>
<td>More / less parton radiation (single top)</td>
</tr>
<tr>
<td>DR vs. DS (single top)</td>
</tr>
<tr>
<td>Modelling (Z+jets)</td>
</tr>
<tr>
<td>$p_T$ reweighting (Z+jets)</td>
</tr>
<tr>
<td>MC non-closure</td>
</tr>
<tr>
<td>Normalisation single top</td>
</tr>
<tr>
<td>Normalisation Z+jets</td>
</tr>
<tr>
<td>Normalisation Z + b/c</td>
</tr>
<tr>
<td>Normalisation diboson</td>
</tr>
<tr>
<td>Normalisation misid. leptons</td>
</tr>
<tr>
<td>Pile-up reweighting</td>
</tr>
<tr>
<td>Electron efficiency/resolution/scale/trigger</td>
</tr>
<tr>
<td>Muon efficiency/resolution/scale/trigger</td>
</tr>
<tr>
<td>$E_T^{\text{miss}}$</td>
</tr>
<tr>
<td>JVT</td>
</tr>
<tr>
<td>Jet energy scale (JES)</td>
</tr>
<tr>
<td>Jet energy resolution (JER)</td>
</tr>
<tr>
<td>Light-flavour jet mis-tag rate</td>
</tr>
<tr>
<td>c-jet mis-tag rate</td>
</tr>
<tr>
<td>Luminosity</td>
</tr>
</tbody>
</table>

**Table 5.** Data-to-simulation scale factors and associated uncertainties for the 70% $b$-jet tagging efficiency single-cut operating point of the MV2c10 $b$-jet tagging algorithm using the LH method, for $R = 0.4$ calorimeter-jets, as a function of the jet $p_T$. 

function of the jet $p_T$ for both the T&P and LH methods, corresponding to the 85% $b$-jet tagging efficiency single-cut OP for $R = 0.4$ calorimeter-jets.

Scale factors for the anti-$k_t$ $R = 0.2$ track-jets at the 70% single-cut OP are presented in figure 11. The efficiencies determined in simulation and data agree within their uncertainties, resulting in scale factors close to unity.
Figure 10. Data-to-simulation scale factors, corresponding to the 85% $b$-jet tagging efficiency single-cut OP, as a function of the jet $p_T$, in (a) the LH method, and (b) the T&P method, for $R = 0.4$ calorimeter-jets. Both the statistical uncertainties (error bars) and total uncertainties (shaded region) are shown.

Figure 11. Data-to-simulation scale factors as a function of the jet $p_T$ using (a) the LH method and (b) the T&P method for $R = 0.2$ track-jets. Both the statistical uncertainties (error bars) and total uncertainties (shaded region) are shown. The results correspond to the 70% $b$-jet tagging efficiency single-cut operating point of the MV2c10 $b$-tagging algorithm.
Table 6. Data-to-simulation scale factors and associated uncertainties for the 70% \( b \)-jet tagging efficiency single-cut operating point of the \( \text{MV2c10} \) \( b \)-jet tagging algorithm using the T&P method, for \( R = 0.4 \) calorimeter-jets, as a function of the jet \( p_T \).

### 8.1 Generator dependence of the efficiency scale factors

The use of EvtGen ensures that Pythia and Herwig use a consistent lifetime and decay model for all \( b \)-hadron species (e.g. \( B^+, B^0, B^0_s \)), which reduces the differences between the \( b \)-jet tagging efficiencies predicted by the two generators. Nevertheless, the intrinsic tagging efficiency of a \( b \)-jet still depends on several aspects which are not harmonised between the different generators, such as: the initial production fractions of the different \( b \)-
hadron species, the fragmentation function, the number of additional charged particles not from the $b$-hadron in the jet and the relative topology of the $b$-hadron and the jet. These differences cause the intrinsic $b$-jet tagging efficiency of a sample to vary depending on the hadronisation/fragmentation generator. Therefore, when using a simulated sample with a different fragmentation model to that used to derive the data-to-simulation scale factors (i.e. POWHEG+PYTHIA6), it is necessary to include additional generator-dependent scale factors. Generator-dependent data-to-simulation scale factors are determined as the ratio of the predicted $b$-jet tagging efficiencies in each jet for the generator in question and the reference of POWHEG+PYTHIA6, with the scale differing from 1 by less than 5% for $b$-jets.

8.2 Smoothing of the efficiency scale factors

For use in physics measurements, the data-to-simulation efficiency scale factors are smoothed from the initial six bins in jet $p_T$ using a local polynomial kernel estimator [49]. This procedure is performed in order to avoid any boundary effects, and to prevent distortions in the distributions of interest in analyses when applying the scale factors.

The result of the smoothing of the $b$-jet scale factor for the 70% operating point in the LH method is shown in figure 12. All the per-bin systematic uncertainties are added in quadrature and shown together with the statistical uncertainties for the calibrated bins of jet $p_T$.

8.3 Reduction of the nuisance parameters

The total uncertainties in the data-to-simulation efficiency scale factors presented in section 8 are calculated as a sum in quadrature of the statistical uncertainty and individual
components of the systematic uncertainty. However, for the application to physics analyses, a statistically more correct approach based on varying each source of uncertainty by $\pm 1\sigma$, independently, and considering its effect on the data-to-simulation efficiency scale factors in each bin, gives a more accurate estimate of the effect of the $b$-tagging uncertainty on the result. If done in this way, a large number of uncertainties (one per source) would need to be taken into account. Thus, reducing the number of systematic uncertainties that need to be considered, while still conserving the correct dependence on the jet $p_T$ and jet $\eta$, is beneficial.

A method for reducing the number of systematic uncertainties while preserving the bin-to-bin correlations was developed, and is based on an eigenvalue decomposition of the covariance matrix of systematic and statistical variations. It starts from the construction of the $6 \times 6$ covariance matrix corresponding to each source of uncertainty in the six bins of jet $p_T$ used for the calibration. Since bin-to-bin correlations are assumed, these matrices have non-zero off-diagonal elements. The total covariance matrix is constructed by summing these covariance matrices corresponding to different sources of uncertainty. As the total covariance matrix is a symmetric, positive-definite matrix, an eigenvector decomposition can be performed. Such a procedure provides orthogonal variations whose size is given by the square root of the corresponding eigenvalues. The resulting number of variations is six, corresponding to the number of bins used for the calibration, and is an important simplification in the implementation of systematic uncertainties in physics analyses.

Finally, most of the eigenvalue variations are very small and can be neglected without impacting the correlations or total uncertainty. The remaining eigenvalue variations can be further reduced by removing eigenvalue variations below a chosen threshold. However, preservation of the correlations comes at a cost, with some of the total uncertainty incorrectly removed. Thus, a tradeoff is made as to how much of the total uncertainty is preserved versus the correlations. Three different schemes of eigen-variation reduction are implemented: ‘loose’ provides a complete description of the total uncertainty and correlations, ‘medium’ has a small amount of loss in the total uncertainty or correlation loss (of the order of 3% relative difference), and ‘tight’ has a more aggressive reduction, where more loss in the total uncertainty or correlation is tolerated (of the order of 10–50% relative difference).

9 Conclusion

The $b$-jet tagging efficiency of the ATLAS $b$-tagging algorithm has been measured using a high-purity sample of dileptonic $t\bar{t}$ events selected from the $36.1\,\text{fb}^{-1}$ of data collected by the ATLAS detector in 2015 and 2016 from proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ at the LHC. A boosted decision tree, based on event topology only, is used to select events in which two $b$-jets are present, reducing the contamination from events in which only one $b$-jet is reconstructed in the detector acceptance. The implementation of a boosted decision tree in the event selection reduces the dominant uncertainty in the modelling of the flavour of the jets in the $t\bar{t}$ events by up to 50%. Two methods are used to extract the efficiency from the $t\bar{t}$ events, a tag-and-probe method and a combinatorial...
likelihood approach. The efficiency is extracted for $R = 0.4$ calorimeter-jets in a transverse momentum range from 20 to 300 GeV, with data-to-simulation scale factors calculated by comparing the efficiency extracted from collision data to that obtained from simulation. The two methods produce consistent results with similar precision. The measured data-to-simulation scale factors are close to unity with a total uncertainty ranging from 2% to 12%. In addition, the data-to-simulation scale factors are measured as a function of the jet $\eta$ and the average number of interactions per bunch crossing, in selected bins of the jet $p_T$, and are found not to have a significant dependence on either of these variables. The generator dependence of the data-to-simulation scale factors is assessed, along with procedures for smoothing the scale factors, and reducing the number of nuisance parameters arising from the data-to-simulation scale factors.

Acknowledgments

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, ERDF, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [50].
Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


Also at Department of Physics, University of Fribourg, Fribourg; Switzerland.

Also at Department of Physics, University of Michigan, Ann Arbor MI; United States of America.

Also at Dipartimento di Fisica E. Fermi, Università di Pisa, Pisa; Italy.

Also at Giresun University, Faculty of Engineering, Giresun; Turkey.

Also at Graduate School of Science, Osaka University, Osaka; Japan.

Also at Hellenic Open University, Patras; Greece.

Also at Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest; Romania.

Also at II. Physikalisches Institut, Georg-August-Universität Göttingen, Göttingen; Germany.

Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona; Spain.

Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg; Germany.

Also at Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen; Netherlands.

Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest; Hungary.

Also at Institute of Particle Physics (IPP); Canada.

Also at Institute of Physics, Academia Sinica, Taipei; Taiwan.

Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku; Azerbaijan.

Also at Institute of Theoretical Physics, Ilia State University, Tbilisi; Georgia.

Also at Istanbul University, Dept. of Physics, Istanbul; Turkey.

Also at LAL, Université Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay; France.

Also at Louisiana Tech University, Ruston LA; United States of America.

Also at Manhattan College, New York NY; United States of America.

Also at Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia.

Also at National Research Nuclear University MEPhI, Moscow; Russia.

Also at Near East University, Nicosia, North Cyprus, Mersin; Turkey.

Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg; Germany.

Also at School of Physics, Sun Yat-sen University, Guangzhou; China.

Also at The City College of New York, New York NY; United States of America.

Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing; China.

Also at Tomsk State University, Tomsk, and Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia.

Also at TRIUMF, Vancouver BC; Canada.

Also at Universita di Napoli Parthenope, Napoli; Italy.

* Deceased