ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in $pp$ collisions at $\sqrt{s} = 13$ TeV

ATLAS Collaboration

DOI
10.1140/epjc/s10052-019-7450-8

Publication date
2019

Document Version
Final published version

Published in
European Physical Journal C

License
CC BY

Citation for published version (APA):
ATLAS Collaboration (2019). ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in $pp$ collisions at $\sqrt{s} = 13$ TeV. European Physical Journal C, 79(11), [970]. https://doi.org/10.1140/epjc/s10052-019-7450-8

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
ATLAS Collaboration

CERN, 1211 Geneva 23, Switzerland

Received: 12 July 2019 / Accepted: 31 October 2019 / Published online: 25 November 2019
© CERN for the benefit of the ATLAS collaboration 2019

Abstract
The algorithms used by the ATLAS Collaboration during Run 2 of the Large Hadron Collider to identify jets containing $b$-hadrons are presented. The performance of the algorithms is evaluated in the simulation and the efficiency with which these algorithms identify jets containing $b$-hadrons is measured in collision data. The measurement uses a likelihood-based method in a sample highly enriched in $t\bar{t}$ events. The topology of the $t \rightarrow Wb$ decays is exploited to simultaneously measure both the jet flavour composition of the sample and the efficiency in a transverse momentum range from 20 to 600 GeV. The efficiency measurement is subsequently compared with that predicted by the simulation. The data used in this measurement, corresponding to a total integrated luminosity of 80.5 fb$^{-1}$, were collected in proton–proton collisions during the years 2015–2017 at a centre-of-mass energy $\sqrt{s} = 13$ TeV. By simultaneously extracting both the efficiency and jet flavour composition, this measurement significantly improves the precision compared to previous results, with uncertainties ranging from 1 to 8% depending on the jet transverse momentum.

Contents
1 Introduction .................................. 1
2 ATLAS detector ............................... 2
3 Key ingredients for $b$-jet identification ......... 3
4 Algorithms for $b$-jet identification .......... 4
  4.1 Training and tuning samples .......... 4
  4.2 Low-level $b$-tagging algorithms ...... 4
    4.2.1 Algorithms based on impact parameters 4
    4.2.2 Secondary vertex finding algorithm . 5
    4.2.3 Topological multi-vertex finding algo-
         rithm .................................. 5
  4.3 High-level tagging algorithms ........ 5
    4.3.1 MV2 .................................. 5
    4.3.2 DL1 .................................. 6
  4.4 Algorithm performance .............. 7
5 Data and simulated samples ............ 9
6 Event selection and classification .......... 10
  6.1 Lepton object definition ............ 10
  6.2 Event selection .......................... 10
  6.3 Event classification .................... 12
7 Extraction of $b$-jet tagging efficiency .... 12
8 Uncertainties ............................... 14
9 Results .................................. 16
10 Usage in ATLAS analysis ............... 17
  10.1 Smoothing ............................... 19
  10.2 Extrapolation to high-$p_T$ jets ...... 19
  10.3 Generator dependence ............... 20
  10.4 Reduction of systematic uncertainties .. 20
  10.5 Application to physics analyses ..... 20
11 Conclusion ................................ 21
References ................................ 22

1 Introduction

The identification of jets containing $b$-hadrons ($b$-jets) against the large jet background containing $c$-hadrons but no $b$-hadron ($c$-jets) or containing neither $b$- or $c$-hadrons (light-flavour jets) is of major importance in many areas of the physics programme of the ATLAS experiment [1] at the Large Hadron Collider (LHC) [2]. It has been decisive in the recent observations of the Higgs boson decay into bottom quarks [3] and of its production in association with a top-quark pair [4], and plays a crucial role in a large number of Standard Model (SM) precision measurements, studies of the Higgs boson properties, and searches for new phenomena.

The ATLAS Collaboration uses various algorithms to identify $b$-jets [5], referred to as $b$-tagging algorithms, when analysing data recorded during Run 2 of the LHC (2015–2018). These algorithms exploit the long lifetime, high mass and high decay multiplicity of $b$-hadrons as well as the properties of the $b$-quark fragmentation. Given a lifetime of the order of 1.5 ps ($\langle \tau \rangle \approx 450$ μm), measurable $b$-hadrons have a significant mean flight length, $\langle l \rangle = \beta \gamma c \tau$, in the detector before decaying, generally leading to at least one
vertex displaced from the hard-scatter collision point. The strategy developed by the ATLAS Collaboration is based on a two-stage approach. Firstly, low-level algorithms reconstruct the characteristic features of the $b$-jets via two complementary approaches, one that uses the individual properties of charged-particle tracks, later referred to as tracks, associated with a hadronic jet, and a second which combines the tracks to explicitly reconstruct displaced vertices. These algorithms, first introduced during Run 1 [5], have been improved and retuned for Run 2. Secondly, in order to maximise the $b$-tagging performance, the results of the low-level $b$-tagging algorithms are combined in high-level algorithms consisting of multivariate classifiers. The performance of tagging a $b$-jet (tagging efficiency, $\varepsilon_b$) and the probability of mistakenly identifying a $c$-jet or a light-flavour jet as a $b$-jet, labelled $e_c$ ($e_l$). In this paper, the performance of the algorithms is quantified in terms of $c$-jet and light-flavour jet rejections, defined as $1/e_c$ and $1/e_l$, respectively.

The imperfect description of the detector response and physics modelling effects in Monte Carlo (MC) simulations necessitates the measurement of the performance of the $b$-tagging algorithms with collision data [6–8]. In this paper, the measurement of the $b$-jet tagging efficiency of the high-level $b$-tagging algorithms used in proton–proton ($pp$) collision data recorded during Run 2 of the LHC at $\sqrt{s} = 13$ TeV is presented. The corresponding measurements for $c$-jets and light-flavour jets, used in the measurement of the $b$-jet tagging efficiency to correct the simulation such that the overall tagging efficiency of $c$-jets and light-flavour jets match that of the data, are described elsewhere [7,8]. The production of $t\bar{t}$ pairs at the LHC provides an abundant source of $b$-jets by virtue of the high cross-section and the $t \rightarrow Wb$ branching ratio being close to 100%. A very pure sample of $t\bar{t}$ events is selected by requiring that both $W$ bosons decay leptonically, referred to as dileptonic $t\bar{t}$ decays in the following. A combinatorial likelihood approach is used to measure the $b$-jet tagging efficiency of the high-level $b$-tagging algorithms as a function of the jet transverse momentum ($p_T$). This version of the analysis builds upon the approach used previously by the ATLAS Collaboration [6], extending the method to derive additional constraints on the flavour composition of the sample, which reduces the uncertainties by up to a factor of two relative to previous publication.

The paper is organised as follows. In Sect. 2, the ATLAS detector is described. Section 3 contains a description of the objects reconstructed in the detector which are key ingredients for $b$-tagging algorithms, while Sect. 4 describes the $b$-tagging algorithms and the evaluation of their performance in the simulation. The second part of the paper focuses on the $b$-jet tagging efficiency measurement carried out in collision data and the application of these results in ATLAS analyses. The data and simulated samples used in this work are described in Sect. 5. The event selection and classification performed for the measurement of the $b$-jet tagging efficiency are summarised in Sect. 6. The measurement technique is presented in Sect. 7 and the sources of uncertainties are described in Sect. 8. The results and their usage within the ATLAS Collaboration are discussed in Sects. 9 and 10, respectively.

2 ATLAS detector

The ATLAS detector [1] at the LHC covers nearly the entire solid angle around the collision point. It consists of an inner tracking detector (ID) surrounded by a superconducting solenoid, electromagnetic and hadronic calorimeters and a muon spectrometer incorporating three large superconducting toroid magnets.

The ID consists in a high-granularity silicon pixel detector which covers the vertex region and typically provides four measurements per track. The innermost layer, known as the insertable B-layer (IBL) [9], was added in 2014 and provides high-resolution hits at small radius to improve the tracking performance. For a fixed $b$-jet efficiency, the incorporation of the IBL improves the light-flavour jet rejection of the $b$-tagging algorithms by up to a factor of four [10]. The silicon pixel detector is followed by a silicon microstrip tracker (SCT) that typically provides eight measurements from four strip double layers. These silicon detectors are complemented by a transition radiation tracker (TRT), which enables radially extended track reconstruction up to the pseudorapidity $|\eta| = 2.0$. The TRT also provides electron identification information based on the fraction of hits (typically 33 in the barrel and up to an average of 38 in the endcaps) above a higher energy-deposit threshold corresponding to transition radiation. The ID is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the pseudorapidity range $|\eta| < 2.5$.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) sampling calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadronic calorimetry is provided by a steel/scintillator-tile calorimeter, segmented into three barrel structures within

---

1 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 $(r,\phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the $z$-axis. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. 

---

$\odot$ Springer
$|\eta| < 1.7$, and two copper/LAr hadronic endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic measurements, respectively.

The muon spectrometer comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by superconducting aircore toroids. The precision chamber system covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, complemented by cathode-strip chambers in the forward region. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel and thin-gap chambers in the endcap regions.

A two-level trigger system [11] is used to select interesting events. The first level of the trigger is implemented in hardware and uses a subset of detector information to reduce the event rate to a design value of at most 100 kHz. It is followed by a software-based trigger that reduces the event rate to a maximum of around 1 kHz for offline storage.

### 3 Key ingredients for $b$-jet identification

The identification of $b$-jets is based on several objects reconstructed in the ATLAS detector which are described here.

**Tracks** are reconstructed in the ID [12]. The $b$-tagging algorithms only consider tracks with $p_T$ larger than 500 MeV, with further selection criteria applied to reject fake and poorly measured tracks [13]. The combined efficiency of the track reconstruction and selection criteria, evaluated by using minimum-bias simulated events (in which more than 98% of charged-particles are pions), ranges from 91% in the central ($|\eta| < 0.1$) region to 73% in the forward ($2.3 \leq |\eta| < 2.5$) region of the detector. Additional selections on reconstructed tracks are applied in the low-level $b$-tagging algorithms described in Sect. 4.2 to specifically select $b$- and $c$-hadron decay track candidates and improve the rejection of tracks originating from pile-up.²

**Primary vertex (PV) reconstruction** [14,15] on an event-by-event basis is particularly important for $b$-tagging since it defines the reference point from which track and vertex displacements are computed. A longitudinal vertex position resolution of about 30 μm is achieved for events with a high multiplicity of reconstructed tracks, while the transverse resolution ranges from 10 to 12 μm, depending on the LHC running conditions that determine the beam spot size. At least one PV is required in each event, with the PV that has the highest sum of squared transverse momenta of contributing tracks selected as the primary interaction point. Displaced charged-particle tracks originating from $b$-hadron decays can then be selected by requiring large transverse and longitudinal impact parameters, $IP_{x\phi} = |d_0|$ and $IP_z = |z_0 \sin \theta|$, respectively, where $d_0 (z_0)$ represents the transverse (longitudinal) perigee parameter defined at the point of the closest approach of the trajectory to the $z$-axis [16]. Upper limits on the values of these parameters are used to reduce the contamination from pile-up, secondary and fake tracks.

**Hadronic jets** are built from topological clusters of energy in the calorimeter [17], calibrated to the electromagnetic scale, using the anti-$k_t$ algorithm with radius parameter $R = 0.4$ [18]. Jet transverse momenta are further corrected to the corresponding particle-level jet $p_T$, based on the simulation [19]. Remaining differences between data and simulated events are evaluated and corrected for using in situ techniques, which exploit the transverse momentum balance between a jet and a reference object such as a photon, $Z$ boson, or multi-jet system in data. After these calibrations, all jets in the event with $p_T \geq 20$ GeV and $|\eta| < 4.5$ must satisfy a set of loose jet-quality requirements [20], else the event is discarded. These requirements are designed to reject fake jets originating from sporadic bursts, large coherent noise or isolated pathological cells in the calorimeter system, hardware issues, beam-induced background or cosmic-ray muons. Jets with $p_T < 20$ GeV or $|\eta| \geq 2.5$ are then discarded. In order to reduce the number of jets with large energy fractions from pile-up collision vertices, the Jet Vertex Tagger (JVT) algorithm is used [21]. The JVT procedure builds a multivariate discriminant for each jet within $|\eta| < 2.5$ based on the ID tracks ghost-associated with the jet [22]; in particular, jets with a large fraction of high-momentum tracks from pile-up vertices are less likely to pass the JVT requirement. The rate of pile-up jets with $p_T \geq 120$ GeV is sufficiently small that the JVT requirement is removed above this threshold. The JVT efficiency for jets originating from the hard $pp$ scattering is 92% in the simulation. Scale factors of order unity are applied to account for efficiency differences relative to collision data.

**Track–jet matching** is performed using the angular separation $\Delta R$ between the track momenta, defined at the perigee, and the jet axis, defined as the vectorial sum of the four-momenta of the clusters associated with the jet. Given that the decay products from higher-$p_T$ $b$-hadrons are more collimated, the $\Delta R$ requirement varies as a function of jet $p_T$, being wider for low-$p_T$ jets (0.45 for jet $p_T = 20$ GeV) and narrower for high-$p_T$ jets (0.26 for jet $p_T = 150$ GeV). If more than one jet fulfills the matching criteria, the closest jet is preferred. The jet axis is also

² Pile-up interactions correspond to additional $pp$ collisions accompanying the hard-scatter $pp$ interaction in proton bunch collisions at the LHC.
used to assign signed impact parameters to tracks, where the sign is defined as positive if the track intersects the jet axis in the transverse plane in front of the primary vertex, and as negative if the intersection lies behind the primary vertex.

Jet flavour labels are attributed to the jets in the simulation.

Jets are labelled as $b$-jets if they are matched to at least one weakly decaying $b$-hadron having $p_T \geq 5$ GeV within a cone of size $\Delta R = 0.3$ around the jet axis. If no $b$-hadrons are found, $c$-hadrons and then $\tau$-leptons are searched for, based on the same selection criteria. The jets matched to a $c$-hadron ($\tau$-lepton) are labelled as $c$-jets ($\tau$-jets). The remaining jets are labelled as light-flavour jets.

4 Algorithms for $b$-jet identification

This section describes the different algorithms used for $b$-jet identification and the evaluation of their performance in simulation. Low-level $b$-tagging algorithms fall into two broad categories. A first approach, implemented in the IP2D and IP3D algorithms [23] and described in Sect. 4.2.1, is inclusive and based on exploiting the large impact parameters of the tracks originating from the $b$-hadron decay. The second approach explicitly reconstructs displaced vertices. The SV1 algorithm [24], discussed in Sect. 4.2.2, attempts to reconstruct an inclusive secondary vertex, while the JetFITTER algorithm [25], presented in Sect. 4.2.3, aims to reconstruct the full $b$- to $c$-hadron decay chain. These algorithms, first introduced during Run 1 [5], benefit from improvements and a new tuning for Run 2. To maximise the $b$-tagging performance, low-level algorithm results are combined using multivariate classifiers. To this end, two high-level tagging algorithms have been developed. The first one, MV2 [23], is based on a boosted decision tree (BDT) discriminant, while the second one, DL1 [23], is based on a deep feed-forward neural network (NN). These two algorithms are presented in Sects. 4.3.1 and 4.3.2, respectively.

4.1 Training and tuning samples

The new tuning and training strategies of the $b$-tagging algorithms for Run 2 are based on the use of a hybrid sample composed of $t\bar{t}$ and $Z'$ simulated events. Only $t\bar{t}$ decays with at least one lepton from a subsequent $W$-boson decay are considered in order to ensure a sufficiently large fraction of $c$-jets in the event whilst maintaining a jet multiplicity profile similar to that in most analyses. A dedicated sample of $Z'$ decaying into hadronic jet pairs is included to optimise the $b$-tagging performance at high jet $p_T$. The cross-section of the hard-scattering process is modified by applying an event-by-event weighting factor to broaden the natural width of the resonance and widen the transverse momentum distribution of the jets produced in the hadronic decays up to a jet $p_T$ of 1.5 TeV. The branching fractions of the decays are set to be one-third each for the $bb$, $cc$ and light-flavour quark pairs to give a $p_T$ spectrum uniformly populated by jets of all flavours. The hybrid sample is obtained by selecting $b$-jets from $t\bar{t}$ events if the corresponding $b$-hadron $p_T$ is below 250 GeV and from the $Z'$ sample if above, with a similar strategy applied for $c$-jets and light-flavour jets. More details about the production of the $t\bar{t}$ simulated sample, referred to as the baseline $t\bar{t}$ sample in the following, and the $Z'$ simulated sample are given in Sect. 5. Events with at least one jet are selected, excluding the jets overlapping with a generator-level electron originating from a $W$- or $Z$-boson decay.

4.2 Low-level $b$-tagging algorithms

4.2.1 Algorithms based on impact parameters

There are two complementary impact parameter-based algorithms, IP2D and IP3D [23]. The IP2D tagger makes use of the signed transverse impact parameter significance of tracks to construct a discriminating variable, whereas IP3D uses both the track signed transverse and the longitudinal impact parameter significance in a two-dimensional template to account for their correlation. Probability density functions (pdf) obtained from reference histograms of the signed transverse and longitudinal impact parameter significances of tracks associated with $b$-jets, $c$-jets and light-flavour jets are derived from MC simulation. The pdfs are computed in exclusive categories that depend on the hit pattern of the tracks to increase the discriminating power. The pdfs are used to calculate ratios of the $b$-jet, $c$-jet and light-flavour jet probabilities on a per-track basis. Log-likelihood ratio (LLR) discriminants are then defined as the sum of the per-track probability ratios for each jet-flavour hypothesis, e.g. $\sum_{i=1}^{N} \log \left( \frac{p_b}{p_c} \right)$ for the $b$-jet and light-flavour jet hypotheses, where $N$ is the number of tracks associated with the jet and $p_b$ ($p_c$) is the template pdf for the $b$-jet (light-flavour jet) hypothesis. The flavour probabilities of the different tracks contributing to the sum are assumed to be independent of each other. In addition to the LLR separating $b$-jets from light-flavour jets, two extra LLR functions are defined to separate $b$-jets from $c$-jets and $c$-jets from light-flavour jets, respectively. These three likelihood discriminants for both the IP2D and IP3D algorithms are used as inputs to the high-level taggers.

Both the IP2D and IP3D algorithms benefited from a complete retuning prior to the 2017–2018 ATLAS data taking period [23]. In particular, a reoptimisation of the track category definitions was performed, allowing the IBL hit pattern expectations and the next innermost layer information to be fully exploited. The rejection of tracks originating from photon conversions, long-lived particles decays ($K_s$, $\Lambda$) and interactions with detector material by the secondary vertex...
algorithms has also been improved. An additional set of new template pdfs was also produced using a 50%/50% mixture of $t\bar{t}$ and $Z'$ simulated events for extra track-categories with no hits in the first two layers, which are populated by long-lived $b$-hadrons traversing the first layers before they decay. The $t\bar{t}$ sample is used to populate all remaining categories.

4.2.2 Secondary vertex finding algorithm

The secondary vertex tagging algorithm, SV1 [24], reconstructs a single displaced secondary vertex in a jet. The reconstruction starts from identifying the possible two-track vertices built with all tracks associated with the jet, while rejecting tracks that are compatible with the decay of long-lived particles ($K_s$ or $\Lambda$), photon conversions or hadronic interactions with the detector material. The SV1 algorithm runs iteratively on all tracks contributing to the cleaned two-track vertices, trying to fit one secondary vertex. In each iteration, the track-to-vertex association is evaluated using a $\chi^2$-test. The track with the largest $\chi^2$ is removed and the vertex fit is repeated until an acceptable vertex $\chi^2$ and a vertex invariant mass less than 6 GeV are obtained. With this approach, the decay products from $b$- and $c$-hadrons are assigned to a single common secondary vertex.

Several refinements in the track and vertex selection were made prior to the 2016–2017 ATLAS data taking period to improve the performance of the algorithm, resulting in an increased pile-up rejection and an overall enhancement of the performance at high jet $p_T$ [24]. Among the various algorithm improvements, additional track-cleaning requirements are applied for jets in the high-pseudorapidity region ($|\eta| \geq 1.5$) to mitigate the negative influence of the increasing amount of detector material on the secondary vertex finding efficiency. The fake-vertex rate is also better controlled by limiting the algorithm to only consider the 25 highest-$p_T$ tracks in the jets, which preserves all reconstructed tracks from $b$-hadron decays, whilst limiting the influence of additional tracks in the jet. The selection of two-track vertex candidates prior to the $\chi^2$-fit was also reoptimised. Extra candidate-cleaning requirements were introduced to further reduce the number of fake vertices and material interactions, such as the rejection of two-track vertex candidates with an invariant mass greater than 6 GeV, which are not likely to originate from $b$- and $c$-hadron decays. Eight discriminating variables, including the number of tracks associated with the SV1 vertex, the invariant mass of the secondary vertex, its energy fraction (defined as the total energy of all the tracks associated with the secondary vertex divided by the energy of all the tracks associated with the jet), and the three-dimensional decay length significance, are used as inputs to the high-level taggers. The $b$-tagging performance of the algorithm is evaluated using a LLR discriminant based on likelihood functions combining pdfs extracted from some of the JetFitter output variables (vertex mass, energy fraction and decay length significance) and parameterised for each of the three jet flavours.

4.3 High-level tagging algorithms

4.3.1 MV2

The MV2 algorithm [23] consists of a boosted decision tree (BDT) algorithm that combines the outputs of the low-level tagging algorithms described in Sect. 4.2 and listed in Table 1. The BDT algorithm is trained using the ROOT Toolkit for Multivariate Data Analysis (TMVA) [27] on the hybrid $t\bar{t} + Z'$ sample. The kinematic properties of the jets, namely $p_T$ and $|\eta|$, are included in the training in order to take advantage of the correlations with the other input variables. However, to avoid differences in the kinematic distributions of signal ($b$-jets) and background ($c$-jets and light-flavour jets) being used to discriminate between the different jet flavours, the $b$-jets and $c$-jets are reweighted in $p_T$ and $|\eta|$ to match the spectrum of the light-flavour jets. No kinematic reweighting is applied at the evaluation stage of the multivariate classifier. For training, the $c$-jet fraction in the background sample is set to 7%, with the remainder com-
Table 1 Input variables used by the MV2 and the DL1 algorithms. The JETFITTER c-tagging variables are only used by the DL1 algorithm

<table>
<thead>
<tr>
<th>Input</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinematics</td>
<td></td>
</tr>
<tr>
<td>$p_T$ Jet</td>
<td>Jet $p_T$</td>
</tr>
<tr>
<td>$\eta$ Jet</td>
<td>Jet $</td>
</tr>
<tr>
<td>IP2D/IP3D</td>
<td></td>
</tr>
<tr>
<td>log($P_b/P_{light}$)</td>
<td>Likelihood ratio between the $b$-jet and light-flavour jet hypotheses</td>
</tr>
<tr>
<td>log($P_b/P_c$)</td>
<td>Likelihood ratio between the $b$- and $c$-jet hypotheses</td>
</tr>
<tr>
<td>log($P_c/P_{light}$)</td>
<td>Likelihood ratio between the $c$-jet and light-flavour jet hypotheses</td>
</tr>
<tr>
<td>SV1</td>
<td></td>
</tr>
<tr>
<td>$m(SV)$</td>
<td>Invariant mass of tracks at the secondary vertex assuming pion mass</td>
</tr>
<tr>
<td>$f_E(SV)$</td>
<td>Energy fraction of the tracks associated with the secondary vertex</td>
</tr>
<tr>
<td>$N_{SV}$</td>
<td>Number of tracks used in the secondary vertex</td>
</tr>
<tr>
<td>$N_{SV}$</td>
<td>Number of two-track vertex candidates</td>
</tr>
<tr>
<td>$L_{SV}$</td>
<td>Transverse distance between the primary and secondary vertex</td>
</tr>
<tr>
<td>$S_{SV}$</td>
<td>Distance between the primary and the secondary vertex</td>
</tr>
<tr>
<td>$\Delta R(\vec{p}<em>{jet}, \vec{p}</em>{vtx})(SV)$</td>
<td>$\Delta R$ between the jet axis and the direction of the secondary vertex relative to the primary vertex</td>
</tr>
<tr>
<td>JETFITTER</td>
<td></td>
</tr>
<tr>
<td>$m(JF)$</td>
<td>Invariant mass of tracks from displaced vertices</td>
</tr>
<tr>
<td>$f_E(JF)$</td>
<td>Energy fraction of the tracks associated with the displaced vertices</td>
</tr>
<tr>
<td>$\Delta R(\vec{p}<em>{jet}, \vec{p}</em>{vtx})(JF)$</td>
<td>$\Delta R$ between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices</td>
</tr>
<tr>
<td>$S_{SV}$</td>
<td>Significance of the average distance between PV and displaced vertices</td>
</tr>
<tr>
<td>$N_{SV}$</td>
<td>Number of tracks from multi-prong displaced vertices</td>
</tr>
<tr>
<td>$N_{SV}$</td>
<td>Number of two-track vertex candidates (prior to decay chain fit)</td>
</tr>
<tr>
<td>$N_{SV}$</td>
<td>Number of single-prong displaced vertices</td>
</tr>
<tr>
<td>$N_{SV}$</td>
<td>Number of multi-prong displaced vertices</td>
</tr>
<tr>
<td>JETFITTER c-tagging</td>
<td></td>
</tr>
<tr>
<td>$L_{SV}$</td>
<td>Distance of 2nd or 3rd vertex from PV</td>
</tr>
<tr>
<td>$L_{SV}$</td>
<td>Transverse displacement of the 2nd or 3rd vertex</td>
</tr>
<tr>
<td>$m_{trk}(2nd/3rdvtx)(JF)$</td>
<td>Invariant mass of tracks associated with 2nd or 3rd vertex</td>
</tr>
<tr>
<td>$E_{trk}(2nd/3rdvtx)(JF)$</td>
<td>Energy fraction of the tracks associated with 2nd or 3rd vertex</td>
</tr>
<tr>
<td>$f_E(2nd/3rdvtx)(JF)$</td>
<td>Fraction of charged jet energy in 2nd or 3rd vertex</td>
</tr>
<tr>
<td>$N_{SV}$</td>
<td>Number of tracks associated with 2nd or 3rd vertex</td>
</tr>
</tbody>
</table>

Table 2 List of optimised hyperparameters used in the MV2 tagging algorithm

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trees</td>
<td>1000</td>
</tr>
<tr>
<td>Depth</td>
<td>30</td>
</tr>
<tr>
<td>Minimum node size</td>
<td>0.05%</td>
</tr>
<tr>
<td>Cuts</td>
<td>200</td>
</tr>
<tr>
<td>Boosting type</td>
<td>Gradient boost</td>
</tr>
<tr>
<td>Shrinkage</td>
<td>0.1</td>
</tr>
<tr>
<td>Bagged sample fraction</td>
<td>0.5</td>
</tr>
</tbody>
</table>
variables listed in Table 1. The latter relate to the dedicated properties of the secondary and tertiary vertices (distance to the primary vertex, invariant mass and number of tracks, energy, energy fraction, and rapidity of the tracks associated with the secondary and tertiary vertices). A jet $p_T$ and $|\eta|$ reweighting similar to the one used for MV2 is performed. The DL1 algorithm parameters, listed in Table 3, include the architecture of the NN, the number of training epochs, the learning rates and training batch size. All of these are optimised in order to maximise the likelihood when constructing the final output discriminant by solving the DL1 algorithm with a smaller memory footprint than BDT-based algorithms. The final DL1 $b$-tagging discriminant is defined as:

$$D_{DL1} = \ln \left( \frac{p_b}{p_c \cdot p_{\text{light}} + (1 - p_c) \cdot p_{\text{light}}} \right),$$

where $p_b$, $p_c$, $p_{\text{light}}$ and $f_c$ represent respectively the $b$-jet, $c$-jet and light-flavour jet probabilities, and the effective $c$-jet fraction in the background training sample. Using this approach, the $c$-jet fraction in the background can be chosen a posteriori in order to optimise the performance of the algorithm. An optimised $c$-jet fraction of 8% is used to evaluate the performance of the DL1 $b$-tagging algorithm in this paper.

The output discriminants of the DL1 $b$-tagging algorithms for $b$-jets, $c$-jets and light-flavour jets in the baseline $t\bar{t}$ simulated events are shown in Fig. 1b.

### 4.4 Algorithm performance

The evaluation of the performance of the algorithms is carried out using $b$-jet tagging single-cut operating points (OPs). These are based on a fixed selection requirement on the $b$-tagging algorithm discriminant distribution ensuring a specific $b$-jet tagging efficiency, $\epsilon_b$, for the $b$-jets present in the baseline $t\bar{t}$ simulated sample. The selections used to define the single-cut OPs of the MV2 and the DL1 algorithms, as well as the corresponding $c$-jet, $\tau$-jet and light-flavour jet rejections, are shown in Table 4. The MV2 and the DL1 discriminant distributions are also divided into five ‘pseudo-continuous’ bins, $(O^i)_{k=1\ldots5}$, delimited by the selections used to define the $b$-jet tagging single-cut OPs for 85%, 77%, 70% and 60% efficiency, and bounded...
Table 4 Selection and c-jet, τ-jet and light-flavour jet rejections corresponding to the different $b$-jet tagging efficiency single-cut operating points for the MV2 and the DL1 $b$-tagging algorithms, evaluated on the baseline $t\bar{t}$ events

<table>
<thead>
<tr>
<th>$\epsilon_b$</th>
<th>MV2 Selection</th>
<th>Rejection</th>
<th>DL1 Selection</th>
<th>Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection</td>
<td>Rejection</td>
<td></td>
<td>Rejection</td>
<td></td>
</tr>
<tr>
<td>c-jet</td>
<td>τ-jet</td>
<td>Light-flavour jet</td>
<td>c-jet</td>
<td>τ-jet</td>
</tr>
<tr>
<td>60%</td>
<td>&gt; 0.94</td>
<td>23</td>
<td>140</td>
<td>1200</td>
</tr>
<tr>
<td>70%</td>
<td>&gt; 0.83</td>
<td>8.9</td>
<td>36</td>
<td>300</td>
</tr>
<tr>
<td>77%</td>
<td>&gt; 0.64</td>
<td>4.9</td>
<td>15</td>
<td>110</td>
</tr>
<tr>
<td>85%</td>
<td>&gt; 0.11</td>
<td>2.7</td>
<td>6.1</td>
<td>25</td>
</tr>
</tbody>
</table>

Fig. 2 The (a) light-flavour jet and (b) c-jet rejections versus the $b$-jet tagging efficiency for the IP3D, SV1, JETFITTER, MV2 and DL1 $b$-tagging algorithms evaluated on the baseline $t\bar{t}$ events

by the trivial 100% and 0% selections. The value of the pdf in each bin is called the $b$-jet tagging probability and labelled $P_b$ in the following. The $b$-jet tagging efficiency of the $\epsilon_b = X\%$ single-cut OP can then be defined as the sum of the $b$-jet tagging probabilities in the range $[X\%,0\%]$. The light-flavour jet and c-jet rejections as a function of the $b$-jet tagging efficiency are shown in Fig. 2 for the various low- and high-level $b$-tagging algorithms. This demonstrates the advantage of combining the information provided by the low-level taggers, where improvements in the light-flavour jet and c-jet rejections by factors of around 10 and 2.5, respectively, are observed at the $\epsilon_b = 70\%$ single-cut OP of the high-level algorithms compared to low-level algorithms. This figure also illustrates the different $b$-jet tagging efficiency range accessible with each low-level algorithm and thereby their complementarity in the multivariate combinations, with the performance of the DL1 and MV2 discriminants found to be similar. The two algorithms tag a highly correlated sample of $b$-jets, where the relative fraction of jet exclusively tagged by each algorithm is around 3% at the $\epsilon_b = 70\%$ single-cut OP. The relative fractions of light-flavour jets exclusively mis-tagged by the MV2 or the DL1 algorithms at the $\epsilon_b = 70\%$ single-cut OP reach 0.2% and 0.1%, respectively.

However, the additional JETFITTER c-tagging variables used by DL1 bring around 30% and 10% improvements in the light-flavour jet and c-jet rejections, respectively, at the $\epsilon_b = 70\%$ single-cut OP compared to MV2.
5 Data and simulated samples

In order to use data to evaluate the performance of the high-level \( b \)-tagging algorithms, a sample of events enriched in \( t \bar{t} \) dileptonic decays is selected.

The analysis is performed with a \( pp \) collision data sample collected at a centre-of-mass energy of \( \sqrt{s} = 13 \, \text{TeV} \) during the years 2015, 2016 and 2017, corresponding to an integrated luminosity of 80.5 fb\(^{-1} \) and a mean number of \( pp \) interactions per bunch crossing of 31.9. The uncertainty in the integrated luminosity is 2.0% [33], obtained using the LUCID-2 detector [34] for the primary luminosity measurements. All events used were recorded during periods when all relevant ATLAS detector components were functioning normally. The dataset was collected using triggers requiring the presence of a single, high-\( p_T \) electron or muon, with \( p_T \) thresholds that yield an approximately constant efficiency for leptons passing an offline selection of \( p_T \geq 28 \, \text{GeV} \).

The baseline \( t \bar{t} \) full simulation sample was produced using POWHEGBOX v2 [35–38] where the first-gluon-emission cut-off scale parameter \( h_{\text{damp}} \) is set to \( 1.5 m_t \), with \( m_{\text{top}} = 172.5 \, \text{GeV} \) used for the top-quark mass. POWHEGBOX was interfaced to PYTHIA 8.230 [39] with the A14 set of tuned parameters [40] and NNPDF30NNLO (NNPDF2.3LO) [41, 42] parton distribution functions in the matrix elements (parton shower). This set-up was found to produce the best modelling of the multiplicity of additional jets and both the individual top-quark and \( t \bar{t} \) system \( p_T \) [43].

Alternative \( t \bar{t} \) simulation samples were generated using POWHEGBOX v2 interfaced to HERWIG 7.0.4 [44] with the H7-UE-MMHT set of tuned parameters. The effects of initial- and final-state radiation (ISR, FSR) are explored by reweighting the baseline \( t \bar{t} \) events in a manner that reduces (reduces and increases) initial (final) parton shower radiation [43] and by using an alternative POWHEGBOX v2 + PYTHIA 8.230 sample with \( h_{\text{damp}} \) set to \( 3 m_{\text{top}} \) and parameter variation group VAR3 (described in Ref. [43]) increased, leading to increased ISR.

The majority of events with at least one ‘fake’ lepton in the selected sample arise from \( t \bar{t} \) production where only one of the \( W \) bosons, which originated from a top-quark decay, decays leptonically. These fake leptons come from several sources, including non-prompt leptons produced from bottom or charm hadron decays, electrons arising from a photon conversion, jets misidentified as electrons, or muons produced from in-flight pion or kaon decays. This background is also modelled using the \( t \bar{t} \) production described above. The rate of events with two fake leptons is found to be negligible.

Non-\( t \bar{t} \) processes, which are largely subdominant in this analysis, can be classified into two types: those with two real prompt leptons from \( W \) or \( Z \) decays (dominant) and those where at least one of the reconstructed lepton candidates is ‘fake’ (subdominant). Backgrounds containing two real prompt leptons include single top production in association with a \( W \) boson (\( Wt \)), diboson production (\( WW, WZ, ZZ \)) where at least two leptons are produced in the electroweak boson decays, and \( Z+jets \), with \( Z \) decaying into leptons. The \( Wt \) single top production was modelled using POWHEGBOX v2 interfaced to PYTHIA 8.230 using the ‘diagram removal’ scheme [45, 46] with the A14 set of tuned parameters and the NNPDF30NNLO (NNPDF2.3LO) [41, 42] parton distribution functions in the matrix elements (parton shower). Diboson production with additional jets was simulated using SHERPA [47, 48] v2.2.1 (for events where one boson decays hadronically) or SHERPA v2.2.2 (for events where no bosons decay hadronically), using the PDF set NNPDF30NNLO [41]. This includes the 4\( \ell \), \( \ell\ell\ell\ell, \ell\ell\ell\nu, \ell\ell\nu\nu, \ell\ell\tau\ell \) and \( \ell\ell\nu\nu \) final states, which cover \( WW, WZ \) and \( ZZ \) production including off-shell \( Z \) contributions. \( Z+jets \) production (including both \( Z \to \ell\ell \) and \( Z \to ee/\mu\mu \) was modelled using SHERPA v2.2.1 with PDF set NNPDF30NNLO. Processes with one real lepton include \( t \)–channel and \( s \)–channel single top production [49]. These processes were modelled with the same generator and parton shower combination as the \( Wt \) channel. \( W+jets \) production, with the \( W \) boson decaying into \( e\nu \) or \( \mu\nu \) with the \( t \)–lepton decaying leptonically, was modelled in a similar way to the \( Z+jets \) production described above.

Alternative samples of non-\( t \bar{t} \) processes include the \( Wt \) single top production using POWHEGBOX v2 interfaced to HERWIG 7.0.4. The effects of ISR and FSR are evaluated by reweighting the baseline single-top events in a manner that either reduces or increases parton shower radiation. An additional \( Wt \) sample using POWHEGBOX v2 interfaced to PYTHIA 8.230 with the alternative ‘diagram subtraction’ scheme [45, 46] is used to investigate the impact of the interference between \( t \bar{t} \) and \( Wt \) production. Uncertainties in di-boson and \( Z+jets \) production are estimated by reweighting the baseline samples, whereas uncertainties in processes with one real lepton are evaluated directly from data, as described later in Sect. 8.

As described in Sect. 4, the new Run 2 \( b \)-tagging algorithm training strategy is based on the use of a hybrid sample composed of both the baseline \( t \bar{t} \) event sample and a dedicated sample of \( Z' \) decaying into hadronic jet pairs. This \( Z' \) sample was generated using PYTHIA 8.2.12 with the A14 set of tuned parameters for the underlying event and the leading-order NNPDF2.3LO [42] parton distribution functions.

The EVTGEN [50] package was used to handle the decay of heavy-flavour hadrons for all samples except for those generated with the SHERPA generator, for which the default SHERPA configuration recommended by the SHERPA authors was used. All MC events have additional overlaid minimum-bias interactions generated with PYTHIA 8.160 with the A3 set of tunes parameters [51] and NNPDF2.3LO parton distribution functions to simulate pile-up background and are
weighted to reproduce the observed distribution of the average number of interactions per bunch crossing of the corresponding data sample. The nominal MC samples were processed through the full ATLAS detector simulation [52] based on GEANT4 [53], but most samples used for systematic uncertainty evaluation were processed with a faster simulation making use of parameterised showers in the calorimeters [54]. The simulated events were reconstructed using the same algorithms as the data.

6 Event selection and classification

A sample of events enriched in $t\bar{t}$ dileptonic decays is selected by requiring exactly two well-identified lepton candidates and two jets to be present in each event. Events are further classified on the basis of two topological variables to control processes including non-$b$-jets. The lepton definition, and the event selection and classification are described in this section.

6.1 Lepton object definition

In addition to the objects reconstructed for $b$-tagging, described in Sect. 3, the event selection for the efficiency measurement requires electron and muon candidates, defined as follows.

Electron candidates are reconstructed from an isolated energy deposit in the electromagnetic calorimeter matched to an ID track [55]. Electrons are selected for inclusion in the analysis within the fiducial region of transverse energy $E_T \geq 28$ GeV and $|\eta| < 2.47$. Candidates within the transition region between the barrel and end-cap electromagnetic calorimeters, $1.37 \leq |\eta| < 1.52$, are removed in order to avoid large trigger efficiency uncertainties in the turn-on region of the lowest $p_T$ trigger. A tight likelihood-based electron identification requirement is used to further suppress the background from multi-jet production. Isolation criteria are used to reject candidates coming from sources other than prompt decays from massive bosons (hadrons faking an electron signature, heavy-flavour decays or photon conversions). Scale factors (SFs), of order unity, derived in $Z \rightarrow e^+e^-$ events are applied to simulated events to account for differences in reconstruction, identification and isolation efficiencies between data and simulated events. Electron energies are calibrated using the $Z$ mass peak.

If electrons, muons or jets overlap with each other, all but one object must be removed from the event. The distance metric used to define overlapping objects is defined as $\Delta R' = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$ where $\Delta y$ represents the rapidity difference. To prevent double-counting of electron energy deposits as jets, jets within $\Delta R' = 0.2$ of a reconstructed electron candidate are removed. If the nearest remaining jet is within $\Delta R' = 0.4$ of the electron, the electron is discarded. To reduce the background from muons from heavy-flavour decays inside jets, muons are required to be separated by $\Delta R' \geq 0.4$ from the nearest jet. In cases where a muon and a jet are reconstructed within $\Delta R' < 0.4$, the muon is removed if the jet has at least three associated tracks; the jet is removed otherwise. This avoids an inefficiency for high-energy muons undergoing significant energy loss in the calorimeter.

6.2 Event selection

To be considered in this analysis, events must have at least one lepton identified in the trigger system. This triggered lepton must match an offline electron or muon candidate. For each applicable trigger, scale factors are applied to the simulation in order to correct for known differences in trigger efficiencies between the simulation and collision data [11].

In order to reject backgrounds with fewer than two prompt leptons, exactly two reconstructed leptons with opposite charges are required. Contributions from backgrounds with $Z$ bosons are reduced by requiring that one lepton is an electron and the other is a muon. The residual contribution from $Z \rightarrow \tau\tau$ events, which populate the low mass region, is further reduced by considering only events with $m_{e\mu} \geq 50$ GeV. The contribution from $t\bar{t}$ events with light-flavour jets from ISR or FSR or from $W$ bosons is reduced by requiring exactly two reconstructed jets.

Since the aim of the study is to measure the $b$-jet tagging efficiency, it is useful to label simulated events according to the generator-level flavour of the two selected jets, fol-
Fig. 3  Distribution of the jet (a) $p_T$ and (b) $\eta$ in the events passing the selection. Simulated events are split into physics process. The ratio panels show the data-to-simulation ratio as well as the fraction of $t\bar{t}$ events among the simulated events.

Following the definitions introduced in Sect. 3, instead of the physics process they originate from. Events with two $b$-jets (non-$b$-jets) are labelled $bb$ ($ll$). Events with one selected $b$-jet and one non-$b$-jet are labelled $bl$ events if the $b$-jet $p_T$ is larger than the non-$b$-jet $p_T$ and $lb$ in the opposite case. According to the simulation, more than 90% of the non-$b$-jets are light-flavour jets, the rest being composed of $c$-jets, and more than 95% of the $b$-jets originate from a top-quark decay. The fraction of $\tau$-jets is predicted to be negligible.

In order to create $bb$, $bl$, $lb$ and $ll$-enriched regions in the selected sample, each of the two leptons is paired with a jet in an exclusive way to determine whether they originate from the same top-quark decay. The pairing is performed such that it minimises $(m_{j1,\ell_i}^2 + m_{j2,\ell_j}^2)$, where $j1$ ($j2$) is the highest-$p_T$ (second highest-$p_T$) $b$-jet, $\ell_i, j$ are the two leptons and $m_{j1,\ell_i}$ ($m_{j2,\ell_j}$) is the invariant mass of the system including the highest (second highest) $p_T$ $b$-jet and its associated lepton. Choosing the pairing that minimises this quantity relies on the fact that if the pairs of objects are from the same original particles then they are likely to have similar masses. Using the minimum of squared masses penalises asymmetric pairings with one high-mass lepton–jet pair, as well as combinations including two very high invariant masses, which are unlikely for those arising from top-quark decay. Events are required to have $m_{j1,\ell_i} \geq 20$ GeV and $m_{j2,\ell_j} \geq 20$ GeV in order to avoid configurations in which a soft jet and a soft lepton are close to each other, which are not well described by the simulation. The event classification based on these variables is described in more detail in the next section.

According to the simulation, about 85% of the events passing the selection are dileptonic $t\bar{t}$ events, about 65% of which are $bb$ events. Single top production in association with a $W$ boson accounts for 8% of the events, with about 30% of these events containing two selected $b$-jets. Diboson and $Z$+jet production represent respectively about 5% and 2% of the selected events, 85% of these events being $ll$ events. Events originating from $W$+jets production are negligible (< 0.1%). The main source of non-$b$-jets therefore originates from $t\bar{t}$ $bl$ or $lb$ events, i.e. dileptonic $t\bar{t}$ events with a high-$p_T$ light-flavour jet originating from ISR or FSR.

Figure 3 shows the level of agreement between data and simulation as a function of the $p_T$ and $\eta$ of the selected jets as well as the expected fraction of $t\bar{t}$ events. The overall level of agreement between data and simulation is fairly good, although some mismodelling is present at high jet $p_T$, possibly related to the modelling of the top-quark $p_T$ [57], which motivates the extraction of the $b$-jet tagging efficiency in jet $p_T$ bins. The distribution of the discriminant of the MV2 algorithm for events passing the selection is shown in Fig. 4. Generally, good modelling is observed, indicating similar $b$-jet tagging efficiencies in data and simulation.
6.3 Event classification

The distributions of the $m_{j1,\ell}$ and $m_{j2,\ell}$ observables are shown in Fig. 5. In the case of $t\bar{t}$ events with two $b$-jets, both $m_{j1,\ell}$ and $m_{j2,\ell}$ have an upper limit around $m_t = 172.5$ GeV and are usually significantly smaller due to the undetected neutrino. This is generally not the case for $bl$, $lb$ and $ll$ events, which result in high $m_{j1,\ell}$ and/or $m_{j2,\ell}$ values more often at high jet $p_T$. Therefore, the $m_{j1,\ell}$ and $m_{j2,\ell}$ observables discriminate between $bb$, $bl$, $lb$ and $ll$ events while being uncorrelated with the $b$-tagging discriminants, which do not make use of leptons outside jets.

The selected events are classified into 45 different bins according to the $p_T$ of the two jets, allowing the $b$-jet tagging efficiency to be measured as a function of the jet $p_T$. In addition, in each leading jet $p_T$, subleading jet $p_T$ bin ($p_{T,1}$, $p_{T,2}$), the events are further classified into four bins according to the $m_{j1,\ell}$ and $m_{j2,\ell}$ values:

- $m_{j1,\ell}, m_{j2,\ell} < 175$ GeV, signal region (SR): high $bb$ purity region used to measure the $b$-jet tagging efficiency,
- $m_{j1,\ell}, m_{j2,\ell} \geq 175$ GeV, $ll$ control region (CRLL): high $ll$ purity control region used to constrain the $bb$, $bl$, $lb$ and $ll$ fractions in the SR,
- $m_{j1,\ell} \geq 175$ GeV, $m_{j2,\ell} < 175$ GeV, $bl$ control region (CRBL): high $bl$ purity control region used to constrain the $bb$, $bl$, $lb$ and $ll$ fractions in the SR.

Finally, the events in the SR are further classified as a function of the pseudo-continuous binned $b$-tagging discriminant of the two jets, denoted $w_1$ and $w_2$, as defined in Sect. 4.4.

These classifications result in a total of 1260 orthogonal categories. A schematic diagram illustrating the event categorisation is shown in Fig. 6. The $bb$ event purity in the signal regions for the different $p_{T,1}$, $p_{T,2}$ bins is shown in Fig. 7. The lowest purity (19%) occurs when both jets have very low $p_T$; however, the majority of bins have a $bb$ event purity greater than 70% and the highest purity (where both jets have high $p_T$) reaches 93%. The CRLL, CRBL and CRLB control regions are enriched in their targeted backgrounds relative to the corresponding SR. Their purity in $ll$, $bl$ and $lb$ events varies across the $p_{T,1}$, $p_{T,2}$ plane and ranges in the simulation from 30% to 90% (CRLL), 32% to 79% (CRBL) and 20% to 74% (CRLB), respectively. The dominant background in each SR always benefits from a high-purity (i.e. $\geq 50\%$) control region.

7 Extraction of $b$-jet tagging efficiency

Once events have been selected and classified, the measurement of the $b$-jet tagging probabilities is performed. The precision of the previous ATLAS measurement [6] was limited by the uncertainty in the fractions of $bb$, $bl$, $lb$ and $ll$ events in the selected sample, which is driven by the modelling of top-quark pair production. The main novelty of this work in comparison to Ref. [6] lies in the measurement method, which uses both signal and control region data to define a joint log-likelihood function allowing the simultaneous estimate of the $b$-jet tagging probabilities and flavour compositions. This new technique leads to a reduction in the total uncertainties by up to a factor of two, as discussed in Sect. 8.

The general form of an extended binned log-likelihood function, after dropping terms that do not depend on the parameters to be estimated, is provided in Eq. (1):

$$\log L\left(v_{tot}, \hat{\Theta}\right) = -v_{tot} + \sum_{i}^{N} n_{i} \log v_{i}\left(v_{tot}, \hat{\Theta}\right), \quad (1)$$

where $v_{tot}$ is the total expected number of events, $\hat{\Theta} = (\Theta_{1}, \ldots, \Theta_{m})$ is the list of parameters to be estimated, including the parameters of interest (POI) and nuisance parameters, $v_{i} (n_{i})$ is the number of expected (observed) events in the bin $i$ and $N$ bins are considered in total.
Fig. 5 Distribution of the (a) $m_{j_1}$ and (b) $m_{j_2}$ observables in events passing the selection. Simulated events are classified in terms of the flavour composition of the two jets. The ratio panels show the data-to-simulation ratio as well as the fraction of $bb$ events among the simulated events.

Fig. 6 Schematic illustration of the event categorisation. Events are binned according to the $p_T$ of both the leading and subleading jet in the event. In each of these bins, events are further assigned to control (CR) and signal (SR) regions. The events in the SR are then classified according to the $b$-tagging discriminant of the two jets.

In this work, the POIs are the $b$-jet tagging probabilities, $P_b$, introduced in Sect. 4.4. They are defined in this measurement per $p_T$ bin, i.e., as the conditional probabilities for a $b$-jet with a transverse momentum falling in one of the nine $p_T$ bins ($T^m_{m=1...9}$) of the measurement to have a $b$-tagging discriminant $w$ falling in one of the five pseudo-continuous bins ($O^k_{k=1...5}$). The $b$-jet tagging efficiency of the single-cut OP $X$ in that jet $p_T$ bin, $\epsilon_b$, relates to the POIs as outlined below:

$$\epsilon_b(X|T^m) = \sum_{O^k \in X} P_b(O^k|T^m).$$

In each control region, the number of events in a given $p_T$, $p_T$ bin ($T^m$, $T^n$) is written as the sum of the $bb$, $bl$, $lb$ and $ll$ yields expected in that bin ($v_{bb}^{m,n}$, $v_{bl}^{m,n}$, $v_{lb}^{m,n}$, $v_{ll}^{m,n}$), corrected by $p_T$, $p_T$ dependent correction factors ($c_{bb}^{m,n}$, $c_{bl}^{m,n}$, $c_{lb}^{m,n}$, $c_{ll}^{m,n}$), forming the nuisance parameters:

$$v_{CR}(T^m, T^n) = c_{bb}^{m,n} v_{CR,bb}^{m,n} + c_{bl}^{m,n} v_{CR,bl}^{m,n} + c_{lb}^{m,n} v_{CR,lb}^{m,n} + c_{ll}^{m,n} v_{CR,ll}^{m,n}.$$
number of events expected in a given \( pT,1, pT,2, w_1, w_2 \) bin \((T^m, T^n, O^k, O^p)\) is thus written:

\[
v_{SR}(T^m, T^n, O^k, O^p) = c_{bb} v_{SR,bb} \cdot \mathcal{P}_b(O^k|T^m) \cdot \mathcal{P}_b(O^p|T^n) + c_{bl} v_{SR,bl} \cdot \mathcal{P}_l(O^k|T^m) \cdot \mathcal{P}_b(O^p|T^n) + c_{lb} v_{SR,lb} \cdot \mathcal{P}_b(O^k|T^m) \cdot \mathcal{P}_l(O^p|T^n) + c_{ll} v_{SR,ll} \cdot \mathcal{P}_l(O^k|T^m) \cdot \mathcal{P}_l(O^p|T^n),
\]

where \( \mathcal{P}_l \) is the effective \( b \)-jet tagging probability of the mix of \( c \)-jets and light-flavour jets predicted by the simulation in each \( pT,1, pT,2 \) bin. The POIs and correction factors are estimated by minimising the negative log-likelihood function defined above with the MINUIT algorithm [58]. Both the POIs and correction factors are free parameters during the minimisation procedure. Signal and control region data are provided as input as well as the \( \mathcal{P}_l \) conditional probabilities, which are estimated from the MC simulation corrected to match data (see Sect. 1). The simulation is also used to determine the \( bb, bl, lb \) and \( ll \) yield fractions according to the type of region (SR, CRs), as the correction factors are defined as a function of \( pT,1 \) and \( pT,2 \) only.

The extraction method is validated using pseudo-data generated with a known flavour composition. This is created by combining events from either nominal or alternative MC simulation fluctuated according to the statistical uncertainty expected from the actual dataset. The input parameters of the minimisation procedure are taken from the nominal MC simulation in all cases. The size of the non-closure effects observed when using pseudo-data based on the nominal and alternative MC simulation are compared respectively with the expected data statistical uncertainty (0.6–3.7%), and to the sum in quadrature of the expected data statistical uncertainty, the MC statistical uncertainty and the physics modelling uncertainties quoted for the final measurement (0.9–5.4%). The non-closure effects are found to be within uncertainties in each \( pT \) bin such that no additional uncertainty related to the signal extraction method is considered.

8 Uncertainties

Uncertainties affecting the measurement which originate from statistical sources are considered together with systematic uncertainties related to the detector calibration and physics modelling.

The data statistical uncertainty in the \( b \)-jet tagging probabilities, and their bin-to-bin correlations, are obtained from the error matrix returned by MINUIT [58] and propagated to the \( b \)-jet tagging efficiencies via a basis transformation. The data statistical uncertainty reaches about 4% (2%) for jets within \( 20 \leq p_T < 30 \text{ GeV} \) (30 \( \leq p_T < 40 \text{ GeV} \)), ranges from 1% to 3% for jet \( p_T \geq 140 \text{ GeV} \) and is below 1% elsewhere.

The bootstrap resampling technique [59] is used to assess the MC statistical uncertainty by creating an ensemble of statistically equivalent measurements in which the weight of each simulated event used in the nominal measurement is multiplied by an additional term, randomly chosen for each event from a Poisson distribution with a mean of one. The standard deviation of the distribution of these measurements is taken as the MC statistical uncertainty. This method allows all correlations to be preserved and the uncertainty in the value of any parameter to be extracted. One hundred
bootstrap replicas of each simulated sample are used for this evaluation. The MC statistical uncertainty in the b-jet tagging efficiencies is found to be non-negligible only for jet $p_T \leq 40 \text{ GeV}$, where it reaches about 2% and 1% for jets within $20 \text{ GeV} \leq p_T < 30 \text{ GeV}$ and $30 \text{ GeV} \leq p_T < 40 \text{ GeV}$, respectively.

The systematic uncertainties are derived by varying a parameter in the simulated events, repeating the complete analysis with this varied parameter and taking the difference between the updated measurement of the b-jet tagging efficiency or probability and the nominal measurement as the (bin-wise correlated) uncertainty. For b-jet tagging efficiencies, the bootstrap replicas of simulated events are then used to evaluate the MC statistical uncertainty in each systematic variation. Variations of the b-jet tagging efficiency that are not statistically significant undergo a bin-merging procedure over an increasing number of $p_T$ bins to improve their significance. Following this procedure, only statistically significant variations are considered as systematic uncertainties.

Uncertainty sources related to the energy scale and resolution of hadronic jets [19] encompass both the modelling of the detector response and the analysis techniques used to derive the calibration. The impact of the jet energy scale uncertainty reaches from 4 to 5% for jet $p_T \leq 30 \text{ GeV}$, 1% for $30 \text{ GeV} \leq p_T < 40 \text{ GeV}$, and is negligible elsewhere. It is dominated by the prediction of the quark/gluon origin of the light-flavour jets and by the difference in their energy response, as well as the difference in the calorimeter energy response as a function of $\eta$. The uncertainty originating from the jet energy resolution is negligible. Uncertainty sources related to the performance of the JVT algorithm [21], the b-tagging performance for light-flavour jets [8] and c-jets [7] as well as the modelling of pile-up interactions were investigated and found to be negligible, as were lepton-related uncertainties, including energy/momentum scale and resolution, identification, isolation, trigger and track–vertex association efficiency.

The uncertainty in the physics modelling of top-quark events is evaluated by changing the parton shower and hadronisation model from PYTHIA 8 to HERWIG 7 and increasing or decreasing the amount of ISR and FSR within PYTHIA 8 [43]. The uncertainties originating from parton distribution functions (PDF) are quantified following the PDF4LHC recommendations [60]. An additional source of uncertainty originates from the mismodelling of the interference between single top $Wt$ and $t\bar{t}$ production. It is evaluated by switching the nominal single-top simulation sample, based on the ‘diagram removal’ scheme, to the one based on the ‘diagram subtraction’ scheme [45]. The final $t\bar{t}$ modelling uncertainty reaches 3% (2%) for jet $p_T < 30 \text{ GeV}$ ($30 \leq p_T < 40 \text{ GeV}$) and about 1% for $p_T \geq 40 \text{ GeV}$. It is dominated at low $p_T$ by PDF and ISR/FSR variations whereas at higher $p_T$ the choice of parton shower and hadronisation model is the dominant contribution. The single-top uncertainty reaches about 3% for jet $p_T < 30 \text{ GeV}$ due to the parton shower and hadronisation model variation, and 1% for jet $p_T \geq 250 \text{ GeV}$, where the uncertainty in the interference with $t\bar{t}$ events is the dominant contribution. It is below 1% elsewhere. The uncertainties associated with the modelling of top-quark events are reduced by up to a factor of two relative to the previous ATLAS analysis [6] due to the new b-jet tagging efficiency extraction method, which allows the bb event yield to be determined at a precision of a few percent in each $p_T,1$, $p_T,2$ bin.

The uncertainty in the modelling of diboson and Z+jets production [61,62] is evaluated by varying the total cross-section and the factorisation and renormalisation scales for these processes, as well as propagating the uncertainty from the PDF. The total cross-section is kept constant when performing the scale and PDF variations such that only the shapes of the kinematic distributions are impacted. The total cross-section is varied by $\pm 6\% (\pm 5\%)$ for Z+jets (diboson) production. The scale uncertainties are estimated simultaneously by varying the nominal values by a factor of two up and down and taking the largest deviations from the nominal predictions in each direction as uncertainties. PDF uncertainties are evaluated using the 100 bootstrap replicas provided by the NNPDF30NNLO [41] using the same method as outlined for the MC statistical uncertainty earlier in this section. The final diboson and Z+jets uncertainties are found to be negligible in the entire range covered by the analysis.

The number of events with a selected muon not originating from a Z- or W-boson decay is predicted by the simulation to be negligible after the event selection. This is due to the tight muon identification and isolation criteria applied. The number of events with a selected electron not originating from a Z- or W-boson decay passing the event selection (1NPel, for 1 non-prompt electron) is also predicted by the simulation to be very small but one order of magnitude higher, reaching about 0.3\% of the total event yield after selection. An uncertainty in this yield is derived by comparing the number of data and MC events in an alternative region defined by requiring two same-sign (SS) leptons instead of opposite-sign (OS). The SS region is predicted by the simulation to have a composition that is 12% 1NPel events, with the remaining 88\% of the sample coming from non-1NPel events, which is dominated by diboson production. This is estimated from simulation and subtracted from the data. The remaining data events are then compared with MC predictions in bins of electron $p_T$. The data-to-simulation ratio ranges from values close to 3 for $p_T < 120 \text{ GeV}$ to values close to 1 for $p_T \geq 300 \text{ GeV}$. These values are used as simulation-to-data scale factors to correct the yield of simulated 1NPel events in the OS region in order to estimate an uncertainty in the fake-lepton modelling. The b-jet tagging efficiency measurement is then repeated with these scale factors applied and com-
pared with the nominal measurement. Differences of about 1% to 2% for jet $p_T < 40$ GeV and negligible elsewhere are observed and accounted for as an additional systematic uncertainty.

9 Results

The goodness-of-fit is evaluated by computing a Pearson’s $\chi^2$ and comparing it with the number of degrees of freedom ($ndf$) of the fit [63]. This procedure tests the hypothesis that the remaining differences between observed and expected yields post-fit originate only from the limited size of the dataset. The $\chi^2/ndf$ value obtained for the nominal measurement is 0.98, corresponding to a $p$-value of about 0.65. This result illustrates the high goodness-of-fit already observed before accounting for the other sources of uncertainty discussed in Sect. 8.

The $bb$, $bl$, $lb$ and $ll$ yield post-fit correction factors are of order unity, compatible with unity within uncertainties, and typically constrained within 2–5% for $bb$, 5–10% for $bl$, and 7–20% for $lb$ and $ll$. The central values of the $bb$ yield correction factors tend to be a few percent below unity, pointing to a slight underestimate of the number of light-flavour jets in the nominal simulation. The yield correction factors deviate more strongly from unity when running on the alternative simulated samples.

The $b$-jet tagging efficiency measurement for the $\epsilon_b = 70\%$ single-cut OP of the MV2 algorithm is presented in Fig. 8a as a function of jet $p_T$ together with the efficiency derived from $t\bar{t}$ simulated events passing the signal region selection. The corresponding $b$-jet tagging efficiency simulation-to-data scale factors, defined as the ratio between the measured $b$-jet tagging efficiency to the $b$-jet tagging efficiency derived from the simulation, are shown in Fig. 8b. Scale factors are derived for all single-cut OPs and for the DL1 tagger using the same technique, resulting in similar results, as illustrated in Fig. 9. The scale factors have values very close to one and are approximately constant throughout the entire $p_T$ range, illustrating the good modelling of the $b$-jet tagging performance. The $b$-jet tagging efficiency measurement for the $\epsilon_b = 70\%$ single-cut OP of the MV2 algorithm as a function of jet $|\eta|$ and the corresponding simulation-to-data scale factors are presented in Fig. 10. The $b$-jet tagging probability and efficiency measurement was also repeated considering only data and simulated events with either less or more than 28 additional $pp$ interactions per bunch crossing, and separately for 2015–2016 and 2017 data. In all cases, consistent results were observed.

The uncertainty in the efficiency measurement for the $\epsilon_b = 70\%$ single-cut OP of the MV2 tagger is summarised in Table 5. The total uncertainty reaches about 1% for $40 \text{ GeV} \leq p_T < 250 \text{ GeV}$, where it is dominated by the uncertainty in the physics modelling of $t\bar{t}$ events and the data statistical uncertainty. At lower $p_T$ values ($20 \text{ GeV} \leq p_T < 40 \text{ GeV}$), the total uncertainty increases to 8% due to higher uncertainties in the jet energy scale, the modelling of $t\bar{t}$ and single-top-quark events, the limited number of data and MC events and the modelling of fake leptons. For jet $p_T \geq 250$ GeV, the uncertainty increases to about 3% due to the limited number of data events. These observations are consistent across single-cut OPs and taggers.

The measurement of the $b$-jet tagging probabilities in the MV2 and DL1 algorithm output bins is presented in Fig. 11a, c together with the $b$-jet tagging probabilities derived from $t\bar{t}$ simulated events passing the signal region selection. The
Fig. 9  $b$-jet tagging efficiency simulation-to-data scale factors as a function of jet $p_T$. The 60%, 70%, 77% and 85% single-cut OP of the (a) MV2 and (b) DL1 taggers are shown. The various groups of points are offset for visual effects but computed in the same jet $p_T$ range. Vertical error bars represent the total uncertainty.

Fig. 10  The (a) $b$-jet tagging efficiency and (b) $b$-jet tagging efficiency simulation-to-data scale factor for $\epsilon_b = 70\%$ single-cut OP of the MV2 tagger as a function of jet $|\eta|$. Vertical error bars include data statistical uncertainties only while the green bands correspond to the sum in quadrature of statistical and systematic uncertainties. The dots are located at the mean of the $b$-jet $|\eta|$ distribution in each $|\eta|$ bin.

10 Usage in ATLAS analysis

This section details how the simulation-to-data scale factors are incorporated into ATLAS physics analyses. Scale factors are smoothed, extrapolated beyond the jet $p_T$ range of the data measurement and corrected taking into account the generator dependence in the simulation. The number of systematic uncertainties is reduced while preserving the bin-by-bin correlations. The scale factors are then applied to ATLAS physics analyses by correcting the $b$-jet tagging response in simulation and by applying related uncertainties to the correction.

probabilities are shown for jets with $110 \text{ GeV} \leq p_T < 140 \text{ GeV}$, which is located close to the $b$-jet tagging efficiency maximum. The corresponding $b$-jet tagging probability scale factors are shown in Fig. 11b, d. Scale factors are derived for all single-cut OPs using the same technique, resulting in similar results. The scale factors have values close to one and are about constant throughout the pseudo-continuous bins.

The uncertainty in this measurement is summarised in Table 6. The total uncertainty varies from about 9% in the 100–85% bin to about 1% in the 60–0% bin. It is driven by the $t\bar{t}$ modelling uncertainties and data statistics, which is consistent with the result reported for the 70% single-cut OP in this $p_T$ range.
Table 5 Breakdown of the systematic uncertainties in the $b$-jet tagging efficiency measurement for the 70% single-cut OP of the MV2 tagger as a function of the jet $p_T$ bin. The ‘$t\bar{t}$ modelling’ and the ‘Single top modelling’ uncertainties correspond to the sum in quadrature of the uncertainty in the parton shower, hadronisation model, initial-state and final-state radiation and PDF for $t\bar{t}$ and single top-quark production, respectively. The ‘Single top modelling’ uncertainties include an additional source originating from the interference between single top and $t\bar{t}$ production. ‘Other sources’ corresponds to the sum in quadrature of the uncertainties related to jet energy resolution, electron and muon performance, $b$-tagging performance for light-flavour jets and $c$-jets, JVT performance, diboson and $Z$+jet modelling (including normalisation and shape uncertainties) and pile-up modelling. All systematic uncertainties are fully correlated bin-by-bin whereas the statistical uncertainty correlations are evaluated following the procedures described in Sect. 8. In the case of correlated systematic uncertainties, the relative sign of the uncertainty in each bin is taken into account, even if not shown here.

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>Relative uncertainty on $\varepsilon_b$ [%] per jet $p_T$ bin [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data statistics</td>
<td>3.7</td>
</tr>
<tr>
<td>MC statistics</td>
<td>2.2</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>4.5</td>
</tr>
<tr>
<td>$t\bar{t}$ modelling</td>
<td>3.2</td>
</tr>
<tr>
<td>Single top modelling</td>
<td>2.5</td>
</tr>
<tr>
<td>Fake leptons modelling</td>
<td>1.8</td>
</tr>
<tr>
<td>Other sources</td>
<td>1.4</td>
</tr>
<tr>
<td>Total</td>
<td>7.7</td>
</tr>
</tbody>
</table>

Fig. 11 The (a, c) $b$-jet tagging probability and (b, d) $b$-jet tagging probability simulation-to-data scale factors for the (a, b) MV2 and (c, d) DL1 tagger for jets with $110 < p_T < 140$ GeV in the various pseudo-continuous bins. The probability measurement is shown together with the probabilities derived from $t\bar{t}$ simulated events passing the signal region selection. Vertical error bars include data statistical uncertainties only while the green bands correspond to the sum in quadrature of statistical and systematic uncertainties. The dots are located at the bin centres.

 Springer
Table 6 Breakdown of the systematic uncertainties in the b-jet tagging probability measurement of the MV2 tagger as a function of the ‘pseudo-continuous’ bins for jets satisfying $110 \leq p_T \leq 140$ GeV. The ‘$t\bar{t}$ modelling’ and the ‘Single top modelling’ uncertainties correspond to the sum in quadrature of the uncertainty in the parton shower, hadronisation model, initial-state and final-state radiation and PDF for $t\bar{t}$ and single-top-quark production, respectively. The ‘Single top modelling’ uncertainties include an additional source originating from the interference between single top and $t\bar{t}$ production. ‘Other sources’ corresponds to the sum in quadrature of the uncertainties related to jet energy resolution, electron and muon performance, b-tagging performance for light-flavour jets and c-jets, JVT performance, diboson and Z+jet modelling (including normalisation and shape uncertainties) and pile-up modelling. All systematic uncertainties are fully correlated bin-by-bin whereas the statistical uncertainty correlations are evaluated following the procedures described in Sect. 8. In the case of correlated systematic uncertainties, the relative sign of the uncertainty in each bin is taken into account, even if not shown here.

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>Impact on $\mathcal{P}_b$ [%] per pseudo-continuous OP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100–85</td>
</tr>
<tr>
<td>Data statistics</td>
<td>4.2</td>
</tr>
<tr>
<td>MC statistics</td>
<td>1.4</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>1.0</td>
</tr>
<tr>
<td>$t\bar{t}$ modelling</td>
<td>7.2</td>
</tr>
<tr>
<td>Single top modelling</td>
<td>2.4</td>
</tr>
<tr>
<td>Fake leptons modelling</td>
<td>0.1</td>
</tr>
<tr>
<td>Other sources combined</td>
<td>2.1</td>
</tr>
<tr>
<td>Total</td>
<td>9.1</td>
</tr>
</tbody>
</table>

### 10.1 Smoothing

The simulation-to-data scale factors for single cut OPs are smoothed in jet $p_T$ using a local polynomial kernel estimator with a bandwidth parameter of 0.2 following the procedure described in Ref. [6]. This procedure prevents distortions in the variables of interest induced by the application of the scale factors.

### 10.2 Extrapolation to high-$p_T$ jets

The analysis described in this paper provides a precise measurement of the b-jet tagging efficiency in data and compares it with the one obtained from MC simulation. Since there are currently not many b-jets in data for jet $p_T$ above 400 GeV in di-lepton $t\bar{t}$ events, an alternative assessment of the uncertainty in the b-jet tagging efficiency for jet $p_T$ in this range is developed to extend the single cut OP calibration to the entire jet $p_T$ range inspected by physics analyses in ATLAS. Underlying quantities that are known to affect the b-tagging performance are varied in the simulation one by one and the b-jet tagging efficiency is recomputed in each case. The difference from the b-jet tagging efficiency obtained in the nominal simulation is then taken as an additional systematic uncertainty.

Four distinct sets of variables, related to the reconstruction of tracks, of jets, the modelling of the b-hadrons and the interaction of long-lived b-hadrons with the detector material, are considered. Among the uncertainties related to the reconstruction of tracks, the ones that are found to most affect the b-tagging performance are those related to the track impact-parameter resolution, the fraction of fake tracks, the description of the detector material, and the track multiplicity per jet. The uncertainty in the impact-parameter resolution includes the effects of alignment, dead modules and additional material not accurately modelled in the simulation. The uncertainty is derived from several event topologies, including dijet events where effects due to tracking in dense environments, such as in the cores of high-energy jets, are included [12]. No dedicated studies of samples enriched in high-energy b-jets, where collimated tracks from displaced decay vertices conspire to create a challenging environment for the track reconstruction algorithm, are included at this stage. The effect of the parton shower simulation and b-quark fragmentation function is evaluated by comparing the b-jet tagging efficiency with the one obtained from the alternative $t\bar{t}$ event simulations described in Sect. 5. In standard ATLAS MC simulations, interactions with detector material are simulated only for the decay products of the b-hadron and not for the b-hadron itself. Given that about 5% of the b-hadrons within b-jets with jet $p_T = 150$ GeV decay after the innermost pixel detector layer, differences in the b-jet tagging efficiency at high $p_T$ are expected. In order to evaluate the size of the effect, the $Z'$ sample described in Sect. 5 was enhanced to include the interaction of b-hadrons with the detector material, and the b-jet tagging efficiency derived from this sample is compared with the one obtained from the nominal $Z'$ sample.

These sources of uncertainties are found to have a similar impact on the b-jet tagging efficiency of the MV2 and DL1 taggers in the jet $p_T$ range 400 GeV to 1 TeV. In this jet $p_T$ regime, the modelling uncertainties are dominant, reaching 2% for $p_T \sim 400$ GeV and growing linearly to ~4% at the TeV scale. The uncertainty due to the interaction with the detector material is also important and found to be ~1% at $p_T \sim 700$ GeV, growing to ~2% at ~1 TeV. Other leading uncertainties include the jet energy scale and track impact-parameter resolution uncertainties, reaching about 2.5% and 1% at ~1 TeV, respectively. At the TeV scale, the impact of the extrapolation uncertainty is different for MV2 and DL1, due to the differing efficiency profile of the two b-taggers. The b-jet tagging efficiency of DL1 falls more steeply at high $p_T$ compared to that of MV2, which is approximately constant. This results in the jet energy scale uncertainty having a much larger impact for the DL1 tagger, due to the increased impact of the migration of jets between the $p_T$ bins.
The simulation-to-data scale factor measured in the highest $p_T$ bin considered in the collision data analysis is extrapolated for $p_T \geq 400$ GeV. The mean value and uncertainties after smoothing at $p_T = 400$ GeV are assumed to stay valid for higher jet $p_T$. An extrapolation uncertainty is then constructed as the sum in quadrature of all the uncertainties described above, rescaled in proportion to their respective values in the highest $p_T$ bin of the data measurement, and added in quadrature to the pre-existing uncertainties. The result of the procedures of smoothing and extrapolating the single cut OP scale factors are shown in Fig. 12, where both the $b$-jet tagging efficiency as directly measured in data and its extrapolation are shown for the $\varepsilon_b = 70\%$ single cut OP of the MV2 and DL1 taggers.

10.4 Reduction of systematic uncertainties

The individual application in a physics analysis of each independent systematic uncertainty included in Fig. 12 would lead to a large number of variations. A method for reducing the total number of uncertainties while preserving the bin-by-bin correlations is provided for use in ATLAS physics analyses and is described in Ref. [6]. This is achieved by constructing the covariance matrix for each source of uncertainty and by summing these matrices together. Bin-by-bin correlations are kept as non-zero off-diagonal elements. As this equates to the total covariance matrix, which is symmetric and positive-definite, an eigenvector decomposition is performed. The resulting number of variations equals the number of jet $p_T$ bins and is further reduced where eigenvalue variations are shown to have a negligible impact on a result.

10.5 Application to physics analyses

For each jet where $b$-jet tagging is applied in ATLAS physics analyses, a weight is applied in simulation to match the tagging rate as measured in data by the calibration analyses. The weight is jet-flavour dependent. The calibration analysis described in this paper is the baseline correction for jets labelled as $b$-jets. If the jet is tagged using a single cut OP in MC simulation the weight is simply the smoothed simulation-to-data scale factor itself:

$$w_{\text{jet}} = SF(p_T),$$

where $SF(p_T)$ is the smoothed $b$-jet tagging efficiency scale factor evaluated at a given $p_T$. If the jet is not tagged the weight becomes:

$$w_{\text{jet}} = \frac{1 - \varepsilon_b^{\text{data}}(p_T)}{1 - \varepsilon_b^{\text{MC}}(p_T)} = \frac{1 - SF(p_T)\varepsilon_b^{\text{MC}}(p_T)}{1 - \varepsilon_b^{\text{MC}}(p_T)}.$$

![Fig. 12 b-jet tagging efficiency simulation-to-data scale factors for the $\varepsilon_b = 70\%$ single cut OP of the (a) MV2 and (b) DL1 taggers, including the smoothed and extrapolated results. The bin centres are used for the smoothing whereas the dots are located at the mean of the b-jet $p_T$ distribution in each $p_T$ bin.](image)
to 8% for single cut OPs. The precision of the measurement is limited at low \( p_T \) by the uncertainty in the jet energy scale, which is detector-related, and at high \( p_T \) by the size of the dataset, which will grow in the future. It was previously limited by the modelling of top-quark pair production. Further procedures relating to the simulation-to-data scale factors are undertaken to correct for generator dependences, smooth the shape, extrapolate beyond the range of the data measurement, reduce the number of nuisance parameters and apply them in ATLAS analyses.

This result demonstrates a significant improvement in the precision of the \( b \)-jet tagging efficiency measurement for the ATLAS experiment, which is typically limited by systematic uncertainties. The total uncertainty is reduced by up to a factor of two relative to the previous ATLAS results. This is achieved by the improvement in the measurement method, which simultaneously extracts the \( b \)-jet tagging efficiency and jet flavour composition.

Acknowledgements 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; BMWFW and FWF, Austria; ANAS, Azerbaijan; STFC, Belarus; CNPq and FAPESP, Brazil; NSEC, RRC 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 and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, The 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, USA. In addition, individual groups and members have received support from BCKDF, CANARIE, CRC and Compute Canada, Canada; COST, ERC, ERDF, Horizon 2020, and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya, Spain; The Royal Society and Leverhulme Trust, UK. 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 (The Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [64].

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: All ATLAS scientific output is published in journals, and preliminary results are made available in Conference Notes. All are openly available, without restriction on use by external parties beyond copyright law and the standard conditions agreed by CERN. Data associated with journal publications are also made available: tables and data from plots (e.g. cross section...
values, likelihood profiles, selection efficiencies, cross section limits, ...) are stored in appropriate repositories such as HEPDATA (http://hepdata.cedar.ac.uk/). ATLAS also strives to make additional material related to the paper available that allows a reinterpretation of the data in the context of new theoretical models. For example, an extended encapsulation of the analysis is often provided for measurements in the framework of RIVET (http://rivet.hepforge.org/). This information is taken from the ATLAS Data Access Policy, which is a public document that can be downloaded from http://opendata.cern.ch/record/413 [opendata.cern.ch].

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Funded by SCOAP³.

References

28. F. Chollet et al., Keras. https://keras.io (2015)
34. G. Avoni et al., The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS. JINST 13, P07017 (2018)
28 (a) Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia; (b) Department of
Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice, Slovak Republic
29 Physics Department, Brookhaven National Laboratory, Upton, NY, USA
30 Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
31 California State University, CA, USA
32 Cavendish Laboratory, University of Cambridge, Cambridge, UK
33 (a) Department of Physics, University of Cape Town, Cape Town, South Africa; (b) Department of Mechanical
Engineering Science, University of Johannesburg, Johannesburg, South Africa; (c) School of Physics, University of the
Witwatersrand, Johannesburg, South Africa
34 Department of Physics, Carleton University, Ottawa, ON, Canada
35 (a) Faculté des Sciences Ain Chock, Réseau Universitaire de Physique des Hautes Energies-Université Hassan II,
Casablanca, Morocco; (b) Faculté des Sciences, Université Ibn-Tofail, Kénitra, Morocco; (c) Faculté des Sciences
Semlalia, LPHEA-Marrakech, Université Cadi Ayyad, Marrakesh, Morocco; (d) Faculté des Sciences, Université
Mohamed Premier and LPTPM, Oujda, Morocco; (e) Faculté des sciences, Université Mohammed V, Rabat, Morocco
36 CERN, Geneva, Switzerland
37 Enrico Fermi Institute, University of Chicago, Chicago, IL, USA
38 LPC, Université Clermont Auvergne, CNRS/IN2P3, Clermont-Ferrand, France
39 Nevis Laboratory, Columbia University, Irvington, NY, USA
40 Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
41 (a) Dipartimento di Fisica, Università della Calabria, Rende, Italy; (b) Laboratori Nazionali di Frascati, INFN Gruppo
Collegato di Cosenza, Cosenza, Italy
42 Physics Department, Southern Methodist University, Dallas, TX, USA
43 Physics Department, University of Texas at Dallas, Richardson, TX, USA
44 National Centre for Scientific Research “Demokritos”, Agia Paraskevi, Greece
45 (a) Department of Physics, Stockholm University, Stockholm, Sweden; (b) Oskar Klein Centre, Stockholm, Sweden
46 Deutsches Elektronen-Synchrotron DESY, Hamburg and Zeuthen, Germany
47 Lehrstuhl für Experimentelle Physik IV, Technische Universität Dortmund, Dortmund, Germany
48 Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden, Germany
49 Department of Physics, Duke University, Durham, NC, USA
50 SUPA-School of Physics and Astronomy, University of Edinburgh, Edinburgh, UK
51 INFN e Laboratori Nazionali di Frascati, Frascati, Italy
52 Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
53 II. Physikalisches Institut, Georg-August-Universität Göttingen, Göttingen, Germany
54 Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève, Switzerland
55 (a) Dipartimento di Fisica, Università di Genova, Genoa, Italy; (b) INFN Sezione di Genova, Genoa, Italy
56 II. Physikalisches Institut, Justus-Liebig-Universität Giessen, Giessen, Germany
57 SUPA-School of Physics and Astronomy, University of Glasgow, Glasgow, UK
58 LPSC, Université Grenoble Alpes, CNRS/IN2P3, Grenoble INP, Grenoble, France
59 Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge, MA, USA
60 (a) Department of Modern Physics and State Key Laboratory of Particle Detection and Electronics, University of Science
and Technology of China, Hefei, China; (b) Institute of Frontier and Interdisciplinary Science and Key Laboratory of
Particle Physics and Particle Irradiation (MOE), Shandong University, Qingdao, China; (c) School of Physics and
Astronomy, Shanghai Jiao Tong University, KLPPAC-MoE, SKLPPC, Shanghai, China; (d) Tsung-Dao Lee Institute,
Shanghai, China
61 (a) Kirchhoff-Institut für Physik, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; (b) Physikalisches Institut,
Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
62 Laboratory for Applied Information Science, Hiroshima Institute of Technology, Hiroshima, Japan
63 (a) Department of Physics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China; (b) Department of Physics,
University of Hong Kong, Hong Kong, China; (c) Department of Physics and Institute for Advanced Study, Hong Kong
University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
64 Department of Physics, National Tsing Hua University, Hsinchu, Taiwan
65 Department of Physics, Indiana University, Bloomington, IN, USA
Also at Dipartimento di Matematica, Informatica e Fisica, Università di Udine, Udine, Italy
Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow, Russia
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 Instituto Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain
Also at Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, The Netherlands
Also at Institute for Nuclear Research and Nuclear Energy (INRNE) of the Bulgarian Academy of Sciences, Sofia, Bulgaria
Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary
Also at Institute of Particle Physics (IPP), Vancouver, 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 Instituto de Fisica Teorica, IFT-UAM/CSIC, Madrid, Spain
Also at Istanbul University, Dept. of Physics, Istanbul, Turkey
Also at Joint Institute for Nuclear Research, Dubna, Russia
Also at LAL, Université Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
Also at Louisiana Tech University, Ruston LA, USA
Also at LPNHE, Sorbonne Université, Université de Paris, CNRS/IN2P3, Paris, France
Also at Manhattan College, New York NY, USA
Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
Also at National Research Nuclear University MEPhI, Moscow, Russia
Also at Physics Department, An-Najah National University, Nablus, Palestine
Also at Physics Dept, University of South Africa, Pretoria, South Africa
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, USA
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