Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb$^{-1}$ of ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV

ATLAS Collaboration

DOI
10.1140/epjc/s10052-023-11736-z

Publication date
2023

Document Version
Final published version

Published in
European Physical Journal C

License
CC BY

Citation for published version (APA):
Calibration of the light-flavour jet mistagging efficiency of the $b$-tagging algorithms with $Z$+jets events using 139 fb$^{-1}$ of ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV

ATLAS Collaboration

CERN, 1211 Geneva 23, Switzerland

Received: 18 January 2023 / Accepted: 24 March 2023 / Published online: 14 August 2023
© CERN for the benefit of the ATLAS collaboration 2023

Abstract The identification of $b$-jets, referred to as $b$-tagging, is an important part of many physics analyses in the ATLAS experiment at the Large Hadron Collider and an accurate calibration of its performance is essential for high-quality physics results. This publication describes the calibration of the light-flavour jet mistagging efficiency in a data sample of proton–proton collision events at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 139 fb$^{-1}$. The calibration is performed in a sample of $Z$ bosons produced in association with jets. Due to the low mistagging efficiency for light-flavour jets, a method which uses modified versions of the $b$-tagging algorithms referred to as flip taggers is used in this work. A fit to the jet-flavour-sensitive secondary-vertex mass is performed to extract a scale factor from data, to correct the light-flavour jet mistagging efficiency in Monte Carlo simulations, while simultaneously correcting the $b$-jet efficiency. With this procedure, uncertainties coming from the modeling of jets from heavy-flavour hadrons are considerably lower than in previous calibrations of the mistagging scale factors, where they were dominant. The scale factors obtained in this calibration are consistent with unity within uncertainties.

1 Introduction

Many analyses in ATLAS [1], such as measurements or searches involving top quarks or Higgs bosons, rely on the identification of jets containing $b$-hadrons ($b$-jets) with high tagging efficiency and low mistagging efficiency for jets containing $c$-hadrons ($c$-jets) or containing neither $b$- nor $c$-hadrons (light-flavour jets). The relatively long lifetime and high mass of $b$-hadrons together with the large track multiplicity of their decay products is exploited by $b$-tagging algorithms to identify $b$-jets.

The $b$-tagging algorithms are trained using Monte Carlo (MC) simulated events and therefore need to be calibrated in order to correct for efficiency differences between data and simulation that may arise from an imperfect description of the data, e.g. in the parton shower and fragmentation modelling or in the detector and response simulation. The efficiency of identifying a $b$-jet ($\epsilon_b$) and the mistagging efficiencies ($\epsilon_c$ and $\epsilon_{\text{light}}$), which are the probabilities that other jets are wrongly identified by the algorithms as $b$-jets, are measured in data and compared with the predictions of the simulation. These tagging and mistagging efficiencies are defined as

$$\epsilon_f = \frac{N_{\text{pass}}}{N_{\text{all}}}$$

where $f \in \{b, c, \text{light}\}$, $N_{\text{pass}}$ is the number of jets of flavour $f$ selected by the $b$-tagging algorithm and $N_{\text{all}}$ is the number of all jets of flavour $f$. The flavour-tagging efficiency in data ($\epsilon_f^{\text{data}}$) is compared with the efficiency in MC simulation ($\epsilon_f^{\text{MC}}$) and a calibration factor, also called the scale factor (SF), is defined as

$$\text{SF}_f = \frac{\epsilon_f^{\text{data}}}{\epsilon_f^{\text{MC}}}$$

The calibration factors correct the efficiencies and mistagging efficiencies in simulation to the ones in data and are applied to all physics analyses in ATLAS that use $b$-tagging. The $b$-tagging efficiencies and the calibration SFs depend on the jet kinematics. The $b$-jet efficiency ($\epsilon_b$) is calibrated using the method described in Ref. [2], where the SFs $\text{SF}_b$ are extracted from a sample of events containing top-quark pairs decaying into a final state with two charged leptons and two $b$-jets. The $c$-jet mistagging efficiency ($\epsilon_c$) is calibrated via the method described in Ref. [3], where the SFs $\text{SF}_c$ are extracted from events containing top-quark pairs decaying into a final state with exactly one charged lepton and several jets. The events are reconstructed using a kinematic likelihood tech-
nique and include a hadronically decaying W boson, whose decay products are rich in $c$-jets.

This paper describes the measurement of the light-flavour jet mistagging efficiency ($\varepsilon_{\text{light}}$) of the DL1 $b$-tagger, which is widely used in ATLAS Run 2 physics analyses and is discussed in Sect. 5. The SFs$_{\text{light}}$ are extracted from particle-flow jets [4] using 139 fb$^{-1}$ of proton–proton ($pp$) collision data collected by the ATLAS detector during Run 2 of the LHC.

The mistagging efficiency $\varepsilon_{\text{light}}$ is difficult to calibrate because after applying a $b$-tagging requirement, the obtained sample of jets is strongly dominated by $b$-jets and the fraction of light-flavour jets passing a selection on the $b$-tagging score is too low to estimate $\varepsilon_{\text{data}}$. In order to extract an unbiased and precise SF$_{\text{light}}$, a sample enriched in mistagged light-flavour jets is required.

In this paper, the Negative Tag method [5,6], which relies on a modified tagger with a reduced $e^b$ and $e^c$ but similar $e^{\text{light}}$ with respect to the nominal tagger, is used. The method, described in detail in Sect. 6, has already been used to calibrate $\varepsilon_{\text{light}}$ by using dijet events in 2015–2016 data [7] from the early part of Run 2.

In this work, jets produced in association with Z bosons (Z+jets) are used instead, allowing the use of unprescaled lepton triggers instead of the prescaled single-jet triggers which were used in the previous calibration. The calibration precision is improved by extracting the SF$_{\text{light}}$ in a fit that simultaneously corrects $e^b$ using data. Previously, $e^b$ was estimated from simulation, resulting in large modelling uncertainties which impacted the precision of the result. An alternative approach [7] to the Negative Tag method exists in which $e^{\text{light}}_{\text{data}}$ is estimated from simulation by applying data-driven corrections to the input quantities of the low-level $b$-tagging algorithms. This method is used to assess the extrapolation uncertainty between the modified and nominal tagger as described in Sect. 7.

The paper is structured as follows. Section 2 describes the ATLAS detector. Section 3 presents the dataset and simulations used in this calibration. The reconstruction of jets, electrons and muons is summarised in Sect. 4, while the $b$-tagging algorithms are described in detail in Sect. 5. The calibration method is detailed in Sect. 6. Systematic uncertainties and results are presented, respectively, in Sects. 7 and 8 and conclusions are given in Sect. 9.

2 ATLAS detector

The ATLAS experiment [1] at the LHC is a multipurpose particle detector with a forward–backward symmetric cylindrical geometry and a near 4$\pi$ coverage in solid angle. It consists of an inner tracking detector (ID) surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer with a toroidal magnet system.

The inner tracking detector, which provides full coverage of a pseudorapidity range $|\eta| < 2.5$, consists of silicon pixel, silicon microstrip, and transition radiation tracking detectors. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first measurement normally being in the insertable B-layer (IBL) installed before Run 2, primarily to enhance the $b$-tagging performance [8,9]. Sampling calorimeters, made of lead and liquid Argon (LAr), provide electromagnetic (EM) energy measurements with high granularity in the pseudorapidity region $|\eta| < 3.2$. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$) and a copper/LAr hadron calorimeter covers the range $1.5 < |\eta| < 3.2$. The forward region is instrumented with LAr calorimeters in the range $3.1 < |\eta| < 4.9$, measuring both electromagnetic and hadronic energies in copper/LAr and tungsten/LAr modules. The muon spectrometer surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 Tm across most of the detector. The muon spectrometer includes a system of precision tracking chambers and fast detectors for triggering.

A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a rate below 100kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz on average depending on the data-taking conditions [10].

An extensive software suite [11] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Data and simulated samples

The presented results are based on data from $pp$ collisions with a centre-of-mass energy of $\sqrt{s} = 13$ TeV and a 25 ns proton bunch crossing interval. The data were collected between 2015 and 2018 with the ATLAS detector, corresponding to an integrated luminosity of 139 fb$^{-1}$ [12,13].

Footnote 1 continued

along the beam pipe. The $x$-axis points from the interaction point to the centre of the LHC ring, and the $y$-axis points upwards. Cylindrical coordinates ($\rho, \phi$) 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 $\eta = -\ln \tan(\theta/2)$. The angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$. Springer
Table 1 A summary of MC generators and their basic parameter settings used to simulate the various physics processes

<table>
<thead>
<tr>
<th>Process</th>
<th>Generator</th>
<th>PDF set</th>
<th>Hadronisation/fragmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z+jets</td>
<td>SHERPA 2.2.1</td>
<td>NNPDF3.0</td>
<td>SHERPA 2.2.1</td>
</tr>
<tr>
<td>$ZW$ and $q\bar{q}$-initiated $Z$</td>
<td>SHERPA 2.2.1</td>
<td>NNPDF3.0</td>
<td>SHERPA 2.2.1</td>
</tr>
<tr>
<td>$ZZ$ from gluon–gluon fusion</td>
<td>SHERPA 2.2.2</td>
<td>NNPDF3.0</td>
<td>SHERPA 2.2.2</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>POWHEG BOX v2</td>
<td>NNPDF3.0</td>
<td>PYTHIA 8.230</td>
</tr>
<tr>
<td>Z+jets ($Z \rightarrow ee$ and $Z \rightarrow \mu\mu$)</td>
<td>MADGRAPH 2.2.3</td>
<td>NNPDF3.0</td>
<td>PYTHIA 8.210</td>
</tr>
<tr>
<td>Z+jets ($Z \rightarrow \tau\tau$)</td>
<td>MADGRAPH 2.2.2</td>
<td>NNPDF2.3</td>
<td>PYTHIA 8.186</td>
</tr>
</tbody>
</table>

A set of single-electron [14] and single-muon triggers [15] were used, with transverse momentum ($p_T$) thresholds in the range 20–26 GeV and depending on the lepton flavour and data-taking period. All detector subsystems are required to have been operational during data taking and to fulfil data quality requirements [16].

Expected Standard Model (SM) processes were modelled using MC simulation. A summary of the simulated samples used in this calibration and their basic parameter settings is provided in Table 1. All samples were produced using the ATLAS simulation infrastructure [17] and GEANT4 [18]. The selection of additional $pp$ interactions per bunch crossing (pile-up) is accounted for by overlaying the hard-scattering process with minimum-bias events generated with PYTHIA 8.186 [19] using the NNPDF2.3LO set of parton distribution functions (PDF) [20] and the A3 set of tuned parameters [21]. Different pile-up conditions between data and simulation are taken into account by reweighting the mean number of interactions per bunch crossing in simulation to the number observed in data.

The events used in this study originate mostly from the production of a $Z$ boson in association with jets. This process is simulated with two different generators. Nominal predictions of $Z/\gamma^*$ production were generated using SHERPA 2.2.1 [22] with the NNPDF3.0NNLO PDF [23]. Diagrams with up to two additional parton emissions were simulated with next-to-leading-order (NLO) precision in QCD, and those with three or four additional parton emissions to leading-order accuracy. Matrix elements were merged with the SHERPA parton shower using the MEPS@NLO formalism [24–27] with a merging scale of 20 GeV and the five-flavour numbering scheme (5FNS). The sample used to evaluate the tagging efficiencies predicted by the MC simulation was produced with the MadGraph 2.2.3 [28] event generator with up to four additional partons at matrix-element level interfaced with PYTHIA 8.210 to model the parton shower, hadronisation, and underlying event with the NNPDF3.0NNLO PDF for the processes where the $Z$ boson decays into muons or electrons, and with MadGraph 2.2.2 interfaced with PYTHIA 8.186 with the NNPDF2.3LO PDF [20] for $Z$ boson decays into $\tau$-leptons. The predictions of the Z+jets MadGraph program are based on calculations using the 5FNS scheme. SHERPA 2.2.1 and MadGraph $Z$+jets simulations are normalised to the overall cross-section prediction at NNLO precision in QCD [29].

In addition to $Z$+jets production, some minor backgrounds contribute to the final event sample used for the calibration. These backgrounds consist mostly of diboson production ($ZZ$ and $ZW$) and top-quark pair production with subsequent decays into final states with two charged leptons. The $ZW$ production and the $ZZ$ production via quark–antiquark annihilation were simulated with SHERPA 2.2.1 with a MEPS@NLO configuration similar to what was used in the simulation of the $Z$+jets process described above, while loop-induced $ZZ$ processes initiated by gluon–gluon fusion were generated with SHERPA 2.2.2. Both the $ZZ$ and $ZW$ simulation samples used the NNPDF3.0NNLO PDF set. Top-quark pair production was simulated using the POWHEG BOX v2 generator at NLO precision in QCD [30] interfaced with PYTHIA 8.230 to model the parton shower, hadronisation, and underlying event, using the A14 set of tuned parameters [31]. This simulation was used to model top-quark pair production with subsequent decays into final states with two charged leptons and into final states with one charged lepton. The latter is used to evaluate the difference between the SFs obtained from the modified (DL1rFlip) and nominal (DL1r) $b$-tagging algorithms.

The decays of $b$- and $c$-hadrons were handled by EVTGEN 1.6.0 [32] in all simulations, except for those generated using SHERPA, for which the default configuration recommended by the SHERPA authors was used.

4 Object and event selection

Tracks of charged particles are reconstructed from energy deposits in the material of the ID [33]. Events are required to contain at least one vertex with two or more associated tracks that must have $p_T > 500$ MeV. Among all vertices, the vertex with the highest $p_T^2$ sum of the associated tracks is taken as the primary vertex (PV) [34]. The transverse ($d_0$) track impact parameter (IP) is defined as the distance of closest approach of the track-trajectory to the PV in the transverse plane. The
longitudinal track IP (z0) is defined as the distance in the z-direction between the PV and the track trajectory at the point of the closest approach in the x–y plane. Jets containing b- and c-hadrons are characterised by having in-flight decays due to the relatively long lifetime of the heavy-flavour hadrons and give rise to secondary vertices (SV) with associated tracks. To exploit the presence of in-flight decays in the jet direction, the signed IP is defined for each track within a jet. As shown in Fig. 1, a track has a positive IP if the angle between the jet direction and the line joining the PV to the point of closest approach to the track is less than π/2 and negative otherwise. The Single Secondary Vertex Finder (SSVF) algorithm [35] is used for the reconstruction of the SV.

Electron candidates are reconstructed from energy clusters in the EM calorimeter matched to an ID track [36]. Candidates must satisfy pT > 18 GeV and |η| < 2.47, excluding the barrel–endcap transition region of 1.37 < |η| < 1.52, and their associated tracks must fulfil |d0|/σd0 < 5 and |z0| sin θ < 0.5 mm, where σd0 is the uncertainty in d0. Electrons are identified using a likelihood-based discriminant that combines information about shower shapes in the EM calorimeter, track properties and the quality of the track-to-cluster matching [36]. Electrons must fulfill a ‘Tight’ identification selection criterion [38] and satisfy |d0|/σd0 < 3 and |z0| sin θ < 0.5 mm. The isolation requirement reduces the contamination from non-prompt muons by placing an upper bound on the amount of energy measured in the tracking detectors and the calorimeter (combined using the particle-flow algorithm [4]) within a cone of variable size (for pTμ < 50 GeV) or fixed size (for pTμ > 50 GeV) around the muon. Efficiency scale factors are used to correct for differences in muon reconstruction, identification, vertex association, isolation and trigger efficiencies between simulation and data [38].

Jets are reconstructed from particle-flow objects combining information from the tracker and calorimeter [4], using the anti-kT algorithm [39,40] with a radius parameter of R = 0.4. The jet energy is calibrated to the particle scale by using a sequence of corrections, including simulation-based corrections and in situ calibrations [41]. The jet-vertex tagging technique (JVT) [42], which uses a multivariate likelihood approach, is applied to jets with |η| < 2.4 and pT < 60 GeV to suppress jets from pile-up activity. The ‘Tight’ selection criterion, corresponding to a JVT score > 0.5, is used. Scale factors are applied to match the JVT MC efficiencies to those in data. All selected jets must have a pT > 20 GeV and |η| < 2.5.

The b-tagging algorithms use tracks matched to jets as input. This matching uses the angular separation between the track momenta, defined at the point of closest approach to the PV, and the jet axis, ΔR(track, jet axis). The selection requirement on ΔR(track, jet axis) varies as a function of the jet pT because the b-hadron decay products are more collimated at larger hadron pT [2].

The jet-flavour labelling in simulation is based on an angular matching of reconstructed jets to generator-level b-hadrons, c-hadrons and τ-leptons with pT > 5 GeV. If a b-hadron is found within a cone of size ΔR = 0.3 around the jet axis, the jet is labelled as a b-jet. If no matching to any b-hadron is possible, the matching procedure is repeated sequentially for c-hadrons and τ-leptons, and the matched jets are called c-jets and τ-jets, respectively. A jet is labelled as a light-flavour jet by default if no matching to any of these particles was successful.

A series of requirements on the angular separation ΔR between muons, electrons and jets are applied to remove overlaps between objects. If an electron candidate shares an ID track with a muon candidate, the electron candidate is rejected. Jets within a cone of ΔR = 0.2 around a lepton are rejected, unless the lepton is a muon and the jet has more than three associated tracks, in which case the muon is rejected. Finally, lepton candidates that are found to be 0.2 < ΔR < 0.4 from any remaining jet are discarded.

A jet sample enriched in light-flavour jets is needed for the calibration. A sample constructed using the leading jet in pT in a Z+jets selection is expected to contain a 7%–8%
Table 2: Expected contributions of processes in the calibration region, compared with the selected data obtained with the full Run 2 dataset. The quoted uncertainties are statistical uncertainties only. Normalisation differences between data and MC simulation are taken into account in the calibration strategy and do not affect the calibration results.

<table>
<thead>
<tr>
<th>Process</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z+jets (SHERPA)</td>
<td>(8.766 ± 0.015)·10^6</td>
</tr>
<tr>
<td>tt</td>
<td>(7.673 ± 0.006)·10^4</td>
</tr>
<tr>
<td>WZ</td>
<td>(4.188 ± 0.005)·10^4</td>
</tr>
<tr>
<td>ZZ</td>
<td>(2.526 ± 0.004)·10^4</td>
</tr>
<tr>
<td>Predictions</td>
<td>(8.910 ± 0.015)·10^6</td>
</tr>
<tr>
<td>Data</td>
<td>9.560·10^6</td>
</tr>
</tbody>
</table>

The fraction of e-jets and a 4%–6% fraction of b-jets, depending on the jet pT. The mistagging efficiency calibration is performed on the leading jet in a sample of Z+jets candidate events. The distinct signature of a Z boson decaying into either two electrons or two muons allows a clean sample of Z+jets events to be selected. Events are selected for further analysis using single-lepton triggers, where one of the leptons must be matched to an object that triggered the recording of the event. The event must contain exactly two prompt leptons with opposite charges and the same flavour (i.e. either exactly two electrons or exactly two muons). The leading lepton must have pT > 28 GeV and the invariant mass of the dilepton system, mℓℓ, must satisfy 81 < mℓℓ < 101 GeV. Only events with a reconstructed Z boson with pT > 50 GeV are considered because the overall modelling is better in this range [43]. The event must also contain at least one jet with pT > 20 GeV and |η| < 2.5.

The event yields obtained after applying all selection requirements are listed in Table 2, while comparisons between the predictions by the simulation and the data are shown in Fig. 2 for selected variables. The predictions of the MC simulation are normalised to data for these comparisons. The shapes of the distributions predicted by the simulation are consistent with those observed within statistical uncertainties for most of the considered kinematic range. The discrepancy observed in the SV mass (mSV) distribution is expected because it is affected by a known mismodelling of the jet-flavour fractions [44,45]; this is mitigated by the fit described in Sect. 6. The mSV observable is calculated using the SSVF algorithm. Negative values of mSV indicate that no SV was found in the jet. The mismodelling in jet pT does not affect the calibration.

5 The DL1r b-tagging algorithm

Properties of b-hadrons, such as their relatively high mass, long lifetime and large multiplicity of reconstructed tracks from their decay products, are exploited to distinguish b-jets from c-jets and light-flavour jets. Individual low-level taggers [46,47] are designed to target the specific properties of b-jets. The IP2D, IP3D and RNNIP algorithms exploit the individual properties of the tracks from charged particles matched to the jet, especially the track d_0 and z_0. RNNIP is based on a recurrent neural network and it explores the correlations between the IPs of different selected tracks in the jet. The SV1 algorithm exploits the output of the SSVF algorithm and JetFitter [48] reconstructs secondary and tertiary vertices, following the expected topology of b-hadron decays. The DL1r algorithm is a deep neural network that takes the output of the low-level taggers and jet kinematics as input and builds a single discriminant [49,50].

ATLAS analyses use selection requirements defining a lower bound on the DL1r discriminant to select b-jets with a certain efficiency. Four of these so-called single-cut operating points (OPs) are defined, corresponding to b-jet selection efficiencies of 85%, 77%, 70% and 60%. The OPs are evaluated in a sample of b-jets from simulated tt events. These selection requirements divide the DL1r score into five intervals to form the pseudo-continuous OPs, where the lower edge of the lowest interval corresponds to 100% DL1r b-tagging efficiency, and the upper edge of the highest interval corresponds to 0% efficiency.

The single-cut and pseudo-continuous OPs are calibrated in order to correct for efficiency differences between data and simulation. The calibration SF, defined in Eq. (1), is measured relative to a reference efficiency εMC. The performance of the b-tagging algorithm in simulation is affected by the hadronisation and fragmentation model used in the parton shower simulation [43]. To account for their differences, simulation-to-simulation SFs are applied to those simulated samples that have a fragmentation model different from the default. The calculation of these simulation-to-simulation correction factors is described in Ref. [51]. The top-quark pair production sample produced with POWHEG BOX v2 + PYTHIA 8.230 is used to define the reference MC efficiencies in the b- and c-jet calibrations. As no equivalent Z+jets simulation produced with the POWHEG BOX v2 generator is available, the Z+jets MadGraph + PYTHIA 8 simulation is used for the definition of the reference efficiency in the calculation of SFlight for light-flavour jets. Since the Z+jets MadGraph + PYTHIA 8 simulation uses the same parton shower model and set of tuned parameters as the top-quark pair production simulation, the light-flavour jet results and the b- and c-jet calibrations are obtained relative to similar reference models.

6 The negative tag method

The DL1r algorithm rejects 97.52% and 99.96% of the light-flavour jets, respectively, at the 85% and 60% OPs, according
Fig. 2 Comparisons between data and simulation for the most relevant kinematic variables after event selection. The MC expectations are normalised to data. The shaded area corresponds to the statistical uncertainties of the MC simulations. The $Z+\text{jets}$ events are simulated using SHERPA 2.2.1. Negative values of $m_{SV}$ indicate that no SV was found in the jet.

Tracks matched to $b$-jets have relatively large and positively signed IPs due to the long lifetime of the $b$-hadrons and the presence of displaced decay vertices. In contrast, tracks matched to light-flavour jets typically have IP values consistent with zero within the IP resolution such that a more symmetric IP distribution is expected. The expected IP distributions of the tracks associated with $b$-jets, $c$-jets and simulated $t\bar{t}$ events. Therefore, the fraction of light-flavour jets passing the threshold is too low to estimate $\varepsilon^\text{light}$ in data. The Negative Tag method [5,6] calibrates $\varepsilon^\text{light}$ data using a modified version of the DL1r algorithm (called DL1rFlip) that achieves lower $\varepsilon^b$ to $\varepsilon^\text{light}$ ratios without changing $\varepsilon^\text{light}$ significantly.

1 The tracks matched to light-flavour jets have a slight bias towards positive-sign values due to the presence of some long-lived particles (e.g., $K_S$, $\Lambda_1$). The contribution from the mismodelling of long-lived particles is expected to be negligible relative to the mismodelling of the $d_0$ and $z_0$ resolutions [7].
or light-flavour jets are shown in Fig. 3. The Negative Tag method assumes that the probability for a light-flavour jet to be mistagged remains almost the same when inverting the IP signs of all tracks and displaced vertices. This is based on the assumption that light-flavour jets are misidentified as $b$-jets mainly due to resolution effects in the track reconstruction which result in tracks and vertices from tracks with positive IPs inside the jet. Given the symmetric IP distributions, the fractions of tracks and vertices from tracks with positive IPs remain stable after inverting the IP signs of all tracks and vertices. The presence of the positive tail in the IP distribution challenges this assumption and its impact is taken into account by a dedicated uncertainty, which is called the DL1rFlip-to-DL1r extrapolation uncertainty in the following.

The DL1rFlip algorithm inverts the signs of the track IPs and the decay length, while using the same algorithm training and OP-threshold definitions as the DL1r algorithm. The $\varepsilon_{\text{light}}$ values obtained by the two algorithms are approximately the same, while the DL1rFlip algorithm selects a smaller fraction of $b$-jets than the nominal version. The discriminants of the DL1r and DL1rFlip algorithms are compared between data and simulation in Fig. 4. These comparisons show that the heavy-flavour fractions are much lower for high DL1rFlip values than for high DL1r values. At the 85% (60%) single-cut OP, the DL1rFlip algorithm selects a sample of jets in which 25% (1%) are light-flavour jets, while the DL1rFlip algorithm selects a sample in which a higher fraction, 38% (6%), are light-flavour jets at the same OP.

The calibration is performed independently in jet $p_T$ intervals in order to account for the $p_T$ dependence of $\varepsilon_{\text{light}}$. A simultaneous binned fit to the $m_{SV}$ distribution in each pseudo-continuous interval of the DL1rFlip discriminant is performed in order to simultaneously determine $\varepsilon_{\text{data}}$ and $\varepsilon_{\text{light}}$ in the DL1rFlip discriminant intervals. The sensitivity of the fit does not allow the SFs of all three jet flavours to be derived simultaneously. Therefore, $\varepsilon_{\text{data}}$ is constrained to the MC predictions and $\varepsilon_{\text{light}}$ is fixed to unity within an uncertainty of 30%, as suggested by studies of the $c$-jet mistagging efficiency calibration [3].

For a given interval of jet $p_T$, the expected number of jets for a defined discriminant interval $i$ is given by

$$N_i(m_{SV}) = N_{i,MC} \times C \times \left[ \sum_{f=\text{light},c,b} F^f \times SF^f_i \times \varepsilon_{i,MC}^f \times P_i^f(m_{SV}) \right],$$

where $N_{i,MC}$ is the predicted flavour-inclusive event yield for each discriminant interval; $C$ is a global normalisation factor and $F^f$ are the jet-flavour fractions; $P_i^f(m_{SV})$ is the probability density function of $m_{SV}$ for jet flavour $f$ in the $i$-th DL1rFlip discriminant interval, taken from simulation. The $P_i^f(m_{SV})$ is defined in such a way to integrate an additional bin ($m_{SV} < 0$ GeV) representing the number of events where no secondary vertex is found. The $m_{SV}$ has been obtained with tracks with nominal sign as input to the SSVF algorithm.

The $C$, $F^f$ and $SF_i^{b,\text{light}}$ parameters are allowed to float in the fit, while $N_{i,MC}$ and $\varepsilon_{i,MC}^f$ are fixed to the predictions from simulated events and $SF_i$ is set to 1.0 ± 0.3. The constraints $\sum_{i=1}^5 \varepsilon_{i,MC}^f \times SF^f_i = 1$, where the contributions to the first bin (corresponding to the 100%–85% OP interval) are allowed to vary such that unitarity is preserved, and $F_{\text{light}} = F_b - F_c$ are applied, reducing the number of free parameters to 11. Figure 5 shows the post-fit $m_{SV}$ and DL1rFlip discriminant distributions for the 50–100 GeV $p_T$ interval. The parameters of interest, i.e. the SFs, are obtained from the DL1rFlip algorithm. In order to apply this calibration to the DL1r algorithm, an extrapolation uncertainty is added to account for the differences between these two tagging algorithms. Details of this DL1rFlip-to-DL1r extrapolation uncertainty are presented in Sect. 7.

### 7 Systematic uncertainties

The measurement of $S_{\text{light}}$ is affected by four types of uncertainties, including those due to experimental effects, the modelling of the $Z$+jets and background processes, and the limited number of events in data and simulation. For each source of uncertainty, one parameter of the fit model is varied at a time, and the effect of this variation on $S_{\text{light}}$ is evaluated. This approach is chosen to prevent the fit from using the data to constrain the impact of individual systematic uncertainties. The probability density functions in $m_{SV}$ are derived from the SHERPA 2.2.1 $Z$+jets simulation sample. The probability density function for each flavour component is separately normalised to unity, so only the shape of the $m_{SV}$ distributions is estimated using the SHERPA 2.2.1 $Z$+jets simulation. The simulated efficiencies from the MADGRAPH $Z$+jets simulation listed in Table 1 are used in the likelihood definition to predict the expected number of events for each flavour category. Given that the flavour fractions and most SFs are extracted from the data, the systematic uncertainty calculation considers mainly effects on the template shapes. An additional uncertainty accounts for differences between $S_{\text{light}}$ values from the DL1rFlip and DL1r algorithms to ensure the applicability of this calibration to the DL1r algorithm.
Fig. 3 Signed-$d_0$ (left) and signed-$z_0$ (right) distributions for tracks matched to $b$-jets, $c$-jets and light-flavour jets in simulated $t\bar{t}$ events. The selected tracks are matched to particle-flow jets with $p_T > 20$ GeV and $|\eta| < 2.5$, and pass the jet-vertex tagger selection. The distributions are normalised to unity. Statistical uncertainties are also shown.

Fig. 4 Comparisons between data and simulation for the distributions of the DL1r discriminant (left) and DL1rFlip discriminant (right). Statistical uncertainties of the yields are shown.

Uncertainties in the modelling of the detector response have a negligible impact on the calibration results. The impact of the jet energy scale uncertainties was estimated by globally shifting the jet energy scale (JES) by 5%, in accord with the most conservative estimate of the JES uncertainty in $p_T$ derived in Ref. [41]. This conservative assumption about the magnitude of the jet energy scale uncertainties leads to an estimated impact of less than 1% on the calibration results, owing to the low $p_T$ dependence of the scale factors. The impact of $d_0$ and $z_0$ IP resolution modelling uncertainties on the $m_{SV}$ templates was estimated by applying data-driven corrections to the simulated $d_0$ and $z_0$ in a sample of simulated top-quark pairs with a method similar to the one in Ref. [7]. The impact of these corrections is transferred to the fit templates and is symmetrised around the central value. This approach was chosen in order to mitigate the effect of statistical fluctuations in the available $Z$+jets samples. Other experimental uncertainties were found to be negligible because they are not correlated with the discriminant used in the fit.

Simulations of $Z$+jets via SHERPA 2.2.1 are used as the nominal model to derive the fit templates for $b$-jets, $c$-jets and light-flavour jets in the $m_{SV}$ distribution. In order to assess the impact of the resulting choice of parton shower model, shower matching scheme and order of the perturbative QCD calculations on the template shapes, SFs$^{\text{light}}$ are derived using an alternative model, provided by the $Z$+jets MADGRAPH + PYTHIA 8 simulation. The difference between the obtained SFs$^{\text{light}}$ is taken as the estimate of the MC modelling uncertainty. The effect of QCD scale uncertainties is estimated by independently doubling or halving the renormalisation and factorisation scales. The impact of choosing
Fig. 5 Post-fit distribution of the DL1rFlip discriminant (left) and the $m_{SV}$ distribution in the 77–70% OP interval (right). Both distributions are shown for the 50–100 GeV $p_T$ interval. The shaded area corresponds to the total statistical uncertainty. The binning in $m_{SV}$ allows good separation between the jet flavours and also uses the minimum number of bins necessary to constrain the free parameters of the fit. Choosing the coarsest binning possible allows the statistical uncertainties to be minimised.

a particular PDF set is estimated by propagating its uncertainties. However, the impacts of the QCD scale and PDF set uncertainties are found to be covered by the statistical uncertainties of the MC simulation and these uncertainties are therefore not included.

The effect of the 30% uncertainty in $S F_{c}$, which was fixed to a value of 1.0 in the fit, is estimated by repeating the fit with $S F_{c}$ set to 1.3 and 0.7 instead. The mean impact of the two variations on $SF_{\text{light}}$ is taken as the impact of the charm calibration uncertainty on the light-flavour jet calibration.

The $S F_{\text{light}}$ value is assumed to be the same for DL1rFlip and DL1r to first order, as described in Sect. 6. However, mismodelling of the IP resolution, the fake-track rate or the parton shower can have different effects on the tagging performance and $SF_{\text{light}}$ values of the DL1r and DL1rFlip algorithms. A DL1rFlip-to-DL1r extrapolation uncertainty is added to account for residual differences in $SF_{\text{light}}$ between the two algorithms. $SF_{\text{light}}$ cannot be measured in data for the DL1r algorithm. Therefore, a method similar to the one in Ref. [7] is used to derive $SF_{\text{light}}$ for DL1rFlip and DL1r from MC simulation in order to compare the two calibrations. The DL1rFlip-to-DL1r extrapolation uncertainty consists of two components. One component estimates the impact of the mismodelling of tracking variables on the DL1rFlip and DL1r tagging performance, evaluated in a bottom-up approach correcting the simulation, and the second component estimates the impact of the shower and hadronisation model on the tagging performance difference between DL1rFlip and DL1r. The impact of the mismodelling of tracking variables is evaluated by applying data-driven corrections to underlying tracking variables affecting the $b$-tagging performance in a sample of simulated $t\bar{t}$ events generated with Powheg Box v2 + Pythia 8.230. The impact on $S F_{\text{light}}$ of correcting these observables is evaluated with respect to the uncorrected simulation. Only corrections which have been shown to have a significant impact are considered [7]. These include the IP resolution and fake-track rate modelling [53]. The total impact on the mistagging efficiency is obtained by multiplying the effects of all corrections. The total impact on $S F_{\text{light}}$ is compared between DL1r and DL1rFlip and the relative difference is assigned as the first component of the DL1rFlip-to-DL1r extrapolation uncertainty. A comparison of these simulation-based calibration factors is shown in Fig. 6. The impact of the parton shower modelling on the difference between $S F_{\text{light}}$ in DL1r and DL1rFlip is estimated by multiplying the $S F_{\text{light}}$ for DL1r (DL1rFlip), derived using the corrected Powheg Box v2 + Pythia 8.230 simulation, by the simulation-to-simulation $S F$s defined as the ratio of the DL1r (DL1rFlip) $S F_{\text{light}}$ for the Sherpa 2.2.1 and the MadGraph + Pythia 8 simulations. The simulation-to-simulation $S F$s have been derived using the $Z$+jets samples listed in Table 1. This gives an estimate of the DL1rFlip-to-DL1r extrapolation uncertainty for an alternative shower model. The difference between this estimate and the estimate obtained using the corrected Powheg Box v2 + Pythia 8.230 simulation is used as the component assessing the impact of the shower and hadronisation model on the tagging performance difference between DL1rFlip and DL1r. The two components are added in quadrature to obtain the DL1rFlip-to-DL1r extrapolation uncertainty.
Fig. 6  Comparison of the light-flavour jet efficiency scale factor results for the DL1r and DL1rFlip tagger obtained with the MC-based calibration on a POWHEG BOX v2 + PYTHIA 8.230 t¯t MC simulation, considering the 85% and 70% operating points, as a function of the jet $p_T$. The error bars correspond to MC statistical uncertainties. The lower panel shows the ratio of DL1rFlip to DL1r tagger MC-based calibration efficiency scale factors. The envelope of this ratio across the considered jet $p_T$ interval is used as an estimate for the impact of the tracking variable mismodelling on the difference between the DL1rFlip and DL1r calibrations, which is a part of the DL1rFlip-to-DL1r extrapolation uncertainty, estimating the impact of applying the calibrations derived for the DL1rFlip algorithm to the DL1r algorithm.

Fig. 7  SFs to calibrate the $\varepsilon$ light of the pseudo-continuous OPs. The size of the uncertainty bands in the direction of the jet $p_T$ axis is arbitrary and corresponds to the choice of the calibration intervals in $p_T$. 

© Springer
The envelope over $p_T$ of the $\text{DL1rFlip}$-to-$\text{DL1r}$ extrapolation uncertainty is used for the final uncertainty estimate. The shower modelling makes the largest contribution to the $\text{DL1rFlip}$-to-$\text{DL1r}$ extrapolation uncertainty and ranges up to 10%. The largest uncertainty contribution in correcting the MC simulation is obtained from the track IP resolution modelling corrections, ranging up to 10% as well. The total $\text{DL1rFlip}$-to-$\text{DL1r}$ extrapolation uncertainty is 10%–12% depending on the $b$-tagging OP and is the dominant systematic uncertainty overall.

Statistical uncertainties of the data and the $m_{SV}$ templates from simulated MC events are taken into account. The statistical uncertainties of the MC-based fit templates are implemented following the light-weight Beeston–Barlow method [54]. As their effect is non-negligible, cross-check studies were performed using a ‘toy’ approach. The fit template bin entries were fluctuated randomly around the nominal value according to a Gaussian probability distribution with a standard deviation equal to the MC statistical uncertainty in the template bins. SF$_{\text{light}}$ was extracted for each of these new fit templates and the standard deviation of the resulting SF$_{\text{light}}$ distribution was taken as the statistical uncertainty. The statistical uncertainty estimates via the Beeston–Barlow model and the toy approach were found to be consistent.

8 Results

The calibration SFs$_{\text{light}}$ of the $\text{DL1r}$ algorithm are presented. The 85%, 77% and 70% OPs were successfully calibrated in data by using the Negative Tag method. However, it is not feasible to calibrate the 60% OP in data because of insufficient statistics and the relatively large contamination by heavy-flavour jets.

Due to these limitations and since the measured SFs for the closest 77–70% and 70–60% pseudo-continuous OPs do not show any significant deviation from unity compared to their respective measurement uncertainties, the assumption is made that the same must hold for the 60% OP. The SF for this OP is thus set to 1. Its dominant uncertainty, the $\text{DL1rFlip}$-to-$\text{DL1r}$ extrapolation uncertainty, is derived specifically for the 60% OP, and found to be slightly larger than for the 70% OP. The other (sub-leading) uncertainties are assumed to be identical to the 70% OP, since their evaluation suffers
The breakdown of the different uncertainties is shown for the different OPs.

**Table 3** Uncertainties on the mistagging efficiency scale factors for the DL1r tagger at the 85% efficiency OP presented as a function of the jet $p_T$.

<table>
<thead>
<tr>
<th>$p_T$ bin [GeV]</th>
<th>20 &lt; $p_T$ &lt; 50</th>
<th>50 &lt; $p_T$ &lt; 100</th>
<th>100 &lt; $p_T$ &lt; 150</th>
<th>150 &lt; $p_T$ &lt; 300</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-jet SF</td>
<td>6.1</td>
<td>5.9</td>
<td>6.2</td>
<td>5.3</td>
</tr>
<tr>
<td>‘Flip’ extrapolation</td>
<td>9.7</td>
<td>9.7</td>
<td>9.7</td>
<td>9.7</td>
</tr>
<tr>
<td>MC modelling</td>
<td>0.38</td>
<td>2.4</td>
<td>4.4</td>
<td>1.0</td>
</tr>
<tr>
<td>Impact parameter ($d_0, z_0$)</td>
<td>1.0</td>
<td>1.3</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Data statistical</td>
<td>0.52</td>
<td>0.56</td>
<td>1.1</td>
<td>1.4</td>
</tr>
<tr>
<td>MC statistical</td>
<td>0.88</td>
<td>0.84</td>
<td>0.72</td>
<td>0.4</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>1.1</td>
<td>1.1</td>
<td>1.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
</tr>
</tbody>
</table>

The results for the mistagging efficiency calibration of light-flavour jets from the DL1r tagger, using particle-flow jets in Run 2 data, are shown in Fig. 7 for the pseudo-continuous OPs of the DL1r algorithm. The calibration of the DL1r single-cut OPs, which is derived from the calibration results of the pseudo-continuous OPs, is shown in Fig. 8. Furthermore, Fig. 9 shows a breakdown of the uncertainties in the SFs$^{\text{light}}$ for each jet $p_T$ interval and single-cut OP.

The measured SFs$^{\text{light}}$ are consistent with unity within uncertainties, except for the 85–77% pseudo-continuous OP and the 85% single-cut OP, where the data prefer SFs$^{\text{light}}$ which differ from one by about one standard deviation. Therefore, the results indicate the MC simulation predicts $\epsilon^{\text{light}}$ to be similar to that in data.

---

Fig. 9 DL1r mistagging efficiency calibration uncertainties. The breakdown of the different uncertainties is shown for the different OPs.
Table 4 Uncertainties on the mistagging efficiency scale factors for the DL1r tagger at the 77% efficiency OP presented as a function of the jet $p_T$

<table>
<thead>
<tr>
<th>$p_T$ bin [GeV]</th>
<th>20 &lt; $p_T$ &lt; 50</th>
<th>50 &lt; $p_T$ &lt; 100</th>
<th>100 &lt; $p_T$ &lt; 150</th>
<th>150 &lt; $p_T$ &lt; 300</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-jet SF</td>
<td>6.4</td>
<td>6.6</td>
<td>8.2</td>
<td>7.8</td>
</tr>
<tr>
<td>‘Flip’ extrapolation</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>MC modelling</td>
<td>3.1</td>
<td>3.9</td>
<td>1.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Impact parameter ($d_0$, $z_0$)</td>
<td>4.8</td>
<td>6.5</td>
<td>2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Data statistical</td>
<td>1.3</td>
<td>1.5</td>
<td>3.1</td>
<td>4.1</td>
</tr>
<tr>
<td>MC statistical</td>
<td>5.1</td>
<td>5.2</td>
<td>4.5</td>
<td>2.7</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>5.2</td>
<td>5.4</td>
<td>5.5</td>
<td>5.1</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>15</td>
<td>16</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 5 Uncertainties on the mistagging efficiency scale factors for the DL1r tagger at the 70% efficiency OP presented as a function of the jet $p_T$

<table>
<thead>
<tr>
<th>$p_T$ bin [GeV]</th>
<th>20 &lt; $p_T$ &lt; 50</th>
<th>50 &lt; $p_T$ &lt; 100</th>
<th>100 &lt; $p_T$ &lt; 150</th>
<th>150 &lt; $p_T$ &lt; 300</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-jet SF</td>
<td>7.8</td>
<td>6.9</td>
<td>8.2</td>
<td>8.8</td>
</tr>
<tr>
<td>‘Flip’ extrapolation</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>MC modelling</td>
<td>11</td>
<td>5.3</td>
<td>4</td>
<td>2.6</td>
</tr>
<tr>
<td>Impact parameter ($d_0$, $z_0$)</td>
<td>9.3</td>
<td>14</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Data statistical</td>
<td>2.6</td>
<td>3</td>
<td>6</td>
<td>8.8</td>
</tr>
<tr>
<td>MC statistical</td>
<td>9.3</td>
<td>12</td>
<td>10</td>
<td>6.3</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>9.7</td>
<td>12</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>22</td>
<td>23</td>
<td>19</td>
<td>18</td>
</tr>
</tbody>
</table>

Mistags of light-flavour jets are mostly caused by the presence of fake tracks and the limited impact parameter resolution [7]. The fake-track rate is underestimated by the MC simulation and the impact parameter resolution is overestimated [7,53]. An increase in the fake-track rate and a decrease in impact parameter resolution both give rise to larger mistagging efficiencies [7,55]. Observed SFs slightly larger than unity are therefore expected.

Tables 3, 4, 5 and 6 feature a detailed breakdown of the uncertainties in the results. Overall, the largest systematic uncertainty is the $\text{DL1rFlip} \rightarrow \text{DL1r}$ extrapolation uncertainty, with values of 10–12%, depending on the $b$-tagging OP. The impact of the charm calibration uncertainty amounts to 5–10% depending on the jet $p_T$ and the $b$-tagging OP, and is therefore one of the larger uncertainties. The MC modelling uncertainty is a few percent in most cases, but reaches its maximum of 11% for the 70% single-cut OP in the $20 \leq \text{jet } p_T \leq 50$ GeV interval.

The effect of the $d_0$ and $z_0$ impact parameter resolution modelling uncertainty is generally of the order of a few percent and therefore mostly subdominant. However, the effect can range up to 14% for low jet $p_T$ and tight OPs. The data statistical uncertainties give subdominant contributions except for high jet $p_T$ and tight OPs, where contributions range from 0.5 to 9.5%. The MC statistical uncertainty is one of the dominant contributions, ranging from 0.4 to 13%, depending on the OP and the jet $p_T$. The total uncertainty in $S_{\text{Flight}}$ is between 11 and 23%. In general, higher precision is obtained for the looser OPs and the size of the dominant uncertainties does not depend significantly on the jet $p_T$.

In the previous calibration with the Negative Tag method using dijet events in 2015–2016 data, the total uncertainty in $S_{\text{Flight}}$ ranged from 14% to 76% for jets with $p_T$ between 20 and 300 GeV. Dominant uncertainties arose from the modelling of the track impact parameter resolution and the uncertainty in the fraction of heavy-flavour jets which was estimated using simulation. The reduced uncertainty in the present calibration stems from both improved modelling of the track impact parameters in simulation and the $m_{SV}$ fit, which allows the extraction of $S_{\text{Flight}}$ while simultaneously correcting $\varepsilon^b$ and the jet-flavour fractions using data. Previously, $\varepsilon^b$ and the heavy-flavour fractions were estimated from simulation, resulting in large modelling uncertainties, which were a limiting factor in the precision of the result.
Negative Tag method using 2015–2016 data. These calibration uncertainties are considerably lower than the previous 14% to 76% uncertainties from the momentum. The scale factors typically exceed unity by around 10–20%, with total uncertainties ranging from 11 to 23%, correcting the light-flavour jet mistagging efficiency, ε, is used. Data-to-simulation scale factors for the application of an alternative b-tagging algorithm, designed to facilitate the measurement of the light-flavour jet mistagging efficiency, is used. Data-to-simulation scale factors for correcting ε in simulation are measured in four different jet transverse momentum intervals, ranging from 20 to 300 GeV, for four separate quantiles of the b-tagging discriminant. The scale factors typically exceed unity by around 10–20%, with total uncertainties ranging from 11 to 23%, and do not exhibit any strong dependence on jet transverse momentum. These calibration uncertainties are considerably lower than the previous 14% to 76% uncertainties from the Negative Tag method using 2015–2016 data.

### 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; BMIWF and FWF, Austria; ANAS, Azerbaijan; CNPq and FAPESP, Brazil; NSEA, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Mincienciencias, Colombia; MEYS CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/Irfu, France; SRNSFG, Georgia; BMDF, HGF and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCUK, Norway; MEInP, Poland; FCT, Portugal; MNE/IFA, Romania; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TENMAK, Turkey; STFC, UK; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; PRIMUS 21/SCI/017 and UNCE SCI/013, Czech Republic; COST, ERC, ERDF, Horizon 2020 and Marie Sklodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir IDEX and 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 MINERVA, Israel; Norwegian Financial Mechanism 2014-2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenS Programmes Generalitat Valenciana, Spain; Giron Gustafssonfons Stiftelse, Sweden; 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 (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. [56].

**Table 6** Uncertainties on the mistagging efficiency scale factors for the DL1r tagger at the 60% efficiency OP presented as a function of the jet p_T

<table>
<thead>
<tr>
<th>p_T bin [GeV]</th>
<th>20 &lt; p_T &lt; 50</th>
<th>50 &lt; p_T &lt; 100</th>
<th>100 &lt; p_T &lt; 150</th>
<th>150 &lt; p_T &lt; 300</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-jet SF</td>
<td>7.6</td>
<td>6.8</td>
<td>10</td>
<td>9.5</td>
</tr>
<tr>
<td>‘Flip’ extrapolation</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>MC modelling</td>
<td>11</td>
<td>5.2</td>
<td>5</td>
<td>2.8</td>
</tr>
<tr>
<td>Impact parameter (d₀, z₀)</td>
<td>9.1</td>
<td>13</td>
<td>3.8</td>
<td>2.2</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>20</td>
<td>20</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Data statistical</td>
<td>2.6</td>
<td>3</td>
<td>7.6</td>
<td>9.5</td>
</tr>
<tr>
<td>MC statistical</td>
<td>9</td>
<td>12</td>
<td>13</td>
<td>6.8</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>9.4</td>
<td>12</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>22</td>
<td>23</td>
<td>23</td>
<td>20</td>
</tr>
</tbody>
</table>

9 Conclusions

The light-flavour jet mistagging efficiency ε_ligh of the DL1r b-tagging algorithm has been measured with a 139 fb⁻¹ sample of $\sqrt{s} = 13$ TeV pp collision events recorded during 2015–2018 by the ATLAS detector at the LHC. The measurement is based on an improved method applied to a sample of Z+jets events. The Negative Tag method, based on the application of an alternative b-tagging algorithm, designed to facilitate the measurement of the light-flavour jet mistagging efficiency, is used. Data-to-simulation scale factors for correcting ε_ligh in simulation are measured in four different jet transverse momentum intervals, ranging from 20 to 300 GeV, for four separate quantiles of the b-tagging discriminant. The scale factors typically exceed unity by around 10–20%, with total uncertainties ranging from 11 to 23%, and do not exhibit any strong dependence on jet transverse momentum. These calibration uncertainties are considerably lower than the previous 14% to 76% uncertainties from the Negative Tag method using 2015–2016 data.

**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 licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. Material that is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copy-
right holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
Funded by SCOAP3. SCOAP3 supports the goals of the International Year of Basic Sciences for Sustainable Development.

References


43. ATLAS Collaboration, Measurements of b-tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. JHEP 08, 089 (2018), arXiv:1805.01854 [hep-ex]


66 IJCLab, Université Paris-Saclay, CNRS/IN2P3, 91405, Orsay, France
67 Centro Nacional de Microelectrónica (IMB-CNM-CSIC), Barcelona, Spain
68 Department of Physics, Indiana University, Bloomington, IN, USA
69 (a) INFN Gruppo Collegato di Udine, Sezione di Trieste, Udine, Italy; (b) ICTP, Trieste, Italy; (c) Dipartimento Politecnico di Ingegneria e Architettura, Università di Udine, Udine, Italy
70 (a) INFN Sezione di Lecce, Lecce, Italy; (b) Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy
71 (a) INFN Sezione di Milano, Milan, Italy; (b) Dipartimento di Fisica, Università di Milano, Milan, Italy
72 (a) INFN Sezione di Napoli, Naples, Italy; (b) Dipartimento di Fisica, Università di Napoli, Naples, Italy
73 (a) INFN Sezione di Pavia, Pavia, Italy; (b) Dipartimento di Fisica, Università di Pavia, Pavia, Italy
74 (a) INFN Sezione di Pisa, Pisa, Italy; (b) Dipartimento di Fisica E. Fermi, Università di Pisa, Pisa, Italy
75 (a) INFN Sezione di Roma, Rome, Italy; (b) Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy
76 (a) INFN Sezione di Roma Tor Vergata, Rome, Italy; (b) Dipartimento di Fisica, Università di Roma Tor Vergata, Rome, Italy
77 (a) INFN Sezione di Roma Tre, Rome, Italy; (b) Dipartimento di Matematica e Fisica, Università Roma Tre, Rome, Italy
78 (a) INFN-TIFPA, Trento, Italy; (b) Università degli Studi di Trento, Trento, Italy
79 Universität Innsbruck, Department of Astro and Particle Physics, Innsbruck, Austria
80 University of Iowa, Iowa City, IA, USA
81 Department of Physics and Astronomy, Iowa State University, Ames, IA, USA
82 (a) Departamento de Engenharia Elétrica, Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, Brazil; (b) Universidade Federal do Rio De Janeiro COPPE/EE/IF, Rio de Janeiro, Brazil; (c) Instituto de Física, Universidade de São Paulo, São Paulo, Brazil; (d) Rio de Janeiro State University, Rio de Janeiro, Brazil
83 KEK, High Energy Accelerator Research Organization, Tsukuba, Japan
84 Graduate School of Science, Kobe University, Kobe, Japan
85 (a) AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Kraków, Poland; (b) Marian Smoluchowski Institute of Physics, Jagiellonian University, Kraków, Poland
86 Institute of Nuclear Physics Polish Academy of Sciences, Kraków, Poland
87 Faculty of Science, Kyoto University, Kyoto, Japan
88 Kyoto University of Education, Kyoto, Japan
89 Research Center for Advanced Particle Physics and Department of Physics, Kyushu University, Fukuoka, Japan
90 Instituto de Física La Plata, Universidad Nacional de La Plata and CONICET, La Plata, Argentina
91 Physics Department, Lancaster University, Lancaster, UK
92 Oliver Lodge Laboratory, University of Liverpool, Liverpool, UK
93 Department of Experimental Particle Physics, Jožef Stefan Institute and Department of Physics, University of Ljubljana, Ljubljana, Slovenia
94 School of Physics and Astronomy, Queen Mary University of London, London, UK
95 Department of Physics, Royal Holloway University of London, Egham, UK
96 Department of Physics and Astronomy, University College London, London, UK
97 Louisiana Tech University, Ruston, LA, USA
98 Fysiska institutionen, Lunds universitet, Lund, Sweden
99 Departamento de Física Teorica C-15 and CIAFF, Universidad Autónoma de Madrid, Madrid, Spain
100 Institut für Physik, Universität Mainz, Mainz, Germany
101 School of Physics and Astronomy, University of Manchester, Manchester, UK
102 CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France
103 Department of Physics, University of Massachusetts, Amherst, MA, USA
104 Department of Physics, McGill University, Montreal, QC, Canada
105 School of Physics, University of Melbourne, Melbourne, VIC, Australia
106 Department of Physics, University of Michigan, Ann Arbor, MI, USA
107 Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA
108 Group of Particle Physics, University of Montreal, Montreal, QC, Canada
109 Fakultät für Physik, Ludwig-Maximilians-Universität München, Munich, Germany
110 Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), Munich, Germany
111 Graduate School of Science and Kobayashi-Maskawa Institute, Nagoya University, Nagoya, Japan
112 Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA
TRIUMF, Vancouver, BC, Canada; Department of Physics and Astronomy, York University, Toronto, ON, Canada
Division of Physics and Tomonaga Center for the History of the Universe, Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Japan
Department of Physics and Astronomy, Tufts University, Medford, MA, USA
United Arab Emirates University, Al Ain, United Arab Emirates
Department of Physics and Astronomy, University of California Irvine, Irvine, CA, USA
Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden
Department of Physics, University of Illinois, Urbana, IL, USA
Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia-CSIC, Valencia, Spain
Department of Physics, University of British Columbia, Vancouver, BC, Canada
Department of Physics and Astronomy, University of Victoria, Victoria, BC, Canada
Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, Würzburg, Germany
Department of Physics, University of Warwick, Coventry, UK
Waseda University, Tokyo, Japan
Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot, Israel
Department of Physics, University of Wisconsin, Madison, WI, USA
Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany
Department of Physics, Yale University, New Haven, CT, USA

a Also Affiliated with an institute covered by a cooperation agreement with CERN, Geneva, Switzerland
b Also at An-Najah National University, Nablus, Palestine
c Also at Borough of Manhattan Community College, City University of New York, New York, NY, USA
d Also at Bruno Kessler Foundation, Trento, Italy
e Also at Center for High Energy Physics, Peking University, Beijing, China
f Also at Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Thessaloniki, Greece
g Also at Centro Studi e Ricerche Enrico Fermi, Rome, Italy
h Also at CERN, Geneva, Switzerland
i Also at Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève, Switzerland
j Also at Departament de Fisica de la Universitat Autonoma de Barcelona, Barcelona, Spain
k Also at Department of Financial and Management Engineering, University of the Aegean, Chios, Greece
l Also at Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA
m Also at Department of Physics, Ben Gurion University of the Negev, Beer Sheva, Israel
n Also at Department of Physics, California State University, East Bay, USA
o Also at Department of Physics, California State University, Sacramento, USA
p Also at Department of Physics, King’s College London, London, UK
q Also at Department of Physics, Stanford University, Stanford, CA, USA
r Also at Department of Physics, University of Fribourg, Fribourg, Switzerland
s Also at Department of Physics, University of Thessaly, Volos, Greece
i Also at Department of Physics, Westmont College, Santa Barbara, USA
u Also at Hellenic Open University, Patras, Greece
v Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain
w Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany
x Also at Institute of Applied Physics, Mohammed VI Polytechnic University, Ben Guerir, Morocco
y Also at Institute of Particle Physics (IPP), Canada
z Also at Institute of Physics and Technology, Ulaanbaatar, Mongolia
aa Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
ab Also at Institute of Theoretical Physics, Ilia State University, Tbilisi, Georgia
ac Also at L2IT, Université de Toulouse, CNRS/IN2P3, UPS, Toulouse, France
ad Also at Lawrence Livermore National Laboratory, Livermore, USA
ae Also at National Institute of Physics, University of the Philippines Diliman (Philippines), Quezon City, Philippines
af Also at Technical University of Munich, Munich, Germany
ag Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing, China
ah Also at TRIUMF, Vancouver, BC, Canada
ai Also at Università di Napoli Parthenope, Naples, Italy
aj Also at University of Chinese Academy of Sciences (UCAS), Beijing, China
ak Also at Department of Physics, University of Colorado Boulder, Colorado, USA
al Also at Washington College, Chestertown, MD, USA
am Also at Yeditepe University, Physics Department, Istanbul, Türkiye
an Also at Institute for Nuclear Research and Nuclear Energy (INRNE) of the Bulgarian Academy of Sciences, Sofia, Bulgaria