Calibration of light-flavour $b$-jet mistagging rates using ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV

The ATLAS Collaboration

The identification of jets containing $b$-hadrons is important for the physics programme of the ATLAS experiment at the Large Hadron Collider. Two evaluations of the misidentification rate of light-flavour jets for the $b$-tagging algorithm MV2c10, used in the LHC Run 2 ATLAS analyses, are described. The evaluations are performed in various ranges of jet transverse momenta and pseudorapidities with proton-proton collision data collected at a centre-of-mass energy of $\sqrt{s} = 13$ TeV during the years 2015 and 2016. The first evaluation is based on a data sample enriched in light-flavour jets thanks to the application of a dedicated algorithm with much reduced capabilities in tagging $b$-jets and similar performance in mistagging light-flavour jets when compared to the standard $b$-tagging algorithm. The second evaluation is based on a bottom-up approach where the underlying tracking variables in the simulation are adjusted to match performance observed in data; the $b$-tagging algorithm is then re-evaluated to assess the change in the light-flavour jet mistagging probability. The results of both methods are in good agreement and are used to calibrate the mistagging rate predicted by the nominal ATLAS simulation. Calibration factors are in the range of about 1.5 to 3 with uncertainties up to 50%.
1 Introduction

The identification of jets containing $b$-hadrons ($b$-tagging) is an important element of a number of prominent analyses performed with the ATLAS detector [1] at the Large Hadron Collider (LHC): for example measurements of standard model processes aiming to constrain the parton density functions of heavy-flavour (HF) quarks (i.e. $b$ and $c$ quarks) [2], studies of the top quark [3] and of the Higgs-boson [4, 5], and exploration of new physics scenarios [6, 7].

The $b$-tagging of a jet relies on the long lifetime of the $b$-hadrons ($\tau \sim 1.5$ ps, corresponding to a proper decay length of about $c\tau \sim 450 \text{ m}$) and large mass, resulting in the production of secondary decays displaced from the proton-proton interaction point and of a large multiplicity of decay products. These observables are reconstructed thanks to the charged-particle tracking capability of the ATLAS inner detector. The information is then combined using a multi-variate algorithm, MV2c10 [8], able to enhance the discrimination of a jet containing one or more $b$-hadrons ($b$-jet) with respect to a jet containing no $b$-hadrons but one or more $c$-hadrons ($c$-jet) or a jet containing neither $b$-hadrons nor $c$-hadrons (LF-jet, LF denoting light-flavour). Specific selections on the output value of a given $b$-tagging algorithm are called working points (WPs) [9, 10].

The performance of a $b$-tagging algorithm is characterized by the probability of tagging a $b$-jet ($\varepsilon_b$) and the probabilities of mistakenly tagging as a $b$-jet a $c$-jet ($\varepsilon_c$) or a LF-jet ($\varepsilon_{LF}$), referred to as “mistag rate” in the following. Ideally, Monte Carlo (MC) simulations including the various quark flavours could be used to evaluate the $b$-tagging performance. However, additional calibration is often needed to account for differences between data and simulation, originating for instance from an imperfect description of the detector response or from physics modelling effects. In practice, each working point of the algorithm is calibrated as a function of the jet transverse momentum ($p_T$) and, if relevant and statistics allows, absolute pseudorapidity ($|\eta|$).

This document presents the evaluations of the LF-jet mistag rate on ATLAS proton-proton collision data recorded at a center-of-mass energy of $\sqrt{s} = 13$ TeV for the MV2c10 algorithm, described briefly in Section 4. MV2c10 is the most commonly used $b$-tagging algorithm for ATLAS analyses during the years 2015-2016. Two methods are used: the negative tag method, described in Section 5, and the adjusted Monte Carlo (adjusted-MC) method, described in Section 6. The negative tag method consists of measuring the LF-jet mistag rate from a high statistics data sample enriched in LF-jets with the application of a modified algorithm which reverses some of the criteria used in the nominal identification algorithm. It has already been used by ATLAS during the LHC Run 1 (2010-2012) to calibrate the MV1 algorithm [8]. The analysis discussed here introduces a flipped tagger for the MV2c10 algorithm and extends the calibration to a wider $p_T$ range up to 3 TeV. The adjusted-MC method is a new calibration technique based on a bottom-up approach where the underlying tracking variables are adjusted to match the data. The effect is then propagated to the observables used by the $b$-tagging algorithm. The performance of the MV2c10 algorithm on $c$- and $b$-jets is described in separate works [8, 11].

---

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 $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$. The transverse momentum is defined as $p_T = E/\cosh(\eta)$. 
2 ATLAS detector

The ATLAS experiment [1] at the LHC is a multi-purpose particle detector with a forward-backward symmetric cylindrical geometry and a near 4π coverage in solid angle. It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer.

The inner tracking detector (ID) covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon micro-strip, and transition radiation tracking detectors. Among them, the pixel detector is crucial for $b$-jet identification. A new inner pixel layer, the Insertable B-Layer (IBL) [12, 13], was added before the start of Run 2, at a mean sensor radius of 33.5 mm from the beam-line. The trajectory of charged particle tracks can be reconstructed with high-precision by this set of sub-detectors. In particular the reconstruction of the primary interaction vertex (PV) and of the impact parameter (IP) of tracks have a central role in the identification of $b$-jets. Vertices are reconstructed using at least two tracks with $p_T > 400$ MeV and the PV is identified as the vertex with the highest sum of the squared transverse momentum of the associated tracks [14]. The transverse and longitudinal impact parameters of a track, $d_0$ and $z_0$ respectively, are defined as the distance of closest approach of the track-trajectory to the PV in the transverse plane and as the distance in the $z$-direction between the PV and the track helix at the point of closest approach in $xy$-plane. The signed impact parameters are defined for a track reconstructed within a jet: it is positive if the angle between the jet direction and the line joining the primary vertex to the point of closest approach to the track is less than $\pi/2$ and negative otherwise. Many reconstructed tracks are used to reconstruct the PV, therefore the track IP measurement can be biased if the track is already used in the PV determination. The IP unbiased variables, $d_{0,\text{unbias}}$ and $z_{0,\text{unbias}}$, of a track are defined by recalculating the IP measurement after removing the track from the PV determination.

Outside the ID, the lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A hadron (iron/scintillator-tile) calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The end-cap and forward regions are instrumented with LAr calorimeters for both EM and hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer surrounds the calorimeters and is based on three large air-core toroid superconducting magnets with eight coils each. Its bending power is in the range from 2.0 to 7.5 Tm. It includes a system of precision tracking chambers and fast detectors for triggering.

A two-level trigger system, using custom hardware followed by a software-based level, is used to reduce the event storage rate to a maximum of around 1 kHz.

3 Data and Simulated Samples

3.1 Data Samples

The proton-proton collision data sample recorded by the ATLAS detector during the year 2015 and 2016 is used, corresponding to a time-integrated luminosity of 36.1 fb$^{-1}$. The uncertainty in the total integrated luminosity is 3.2%. It is derived following a methodology similar to that detailed in Ref. [15] from a preliminary calibration of the luminosity scale using $x$–$y$ beam-separation scans performed in August 2015 and May 2016. Only events recorded during stable beam conditions, with a reconstructed PV, and satisfying detector and data-quality requirements are considered.
The data used for the measurements described in this note were recorded using a suite of single jet triggers [16] requiring at least one hadronic jet with sufficient transverse momentum $p_T^{\text{jet}}$ and absolute pseudorapidity $|\eta^{\text{jet}}| < 3.2$ in the event. Given the very high rate of such events at the LHC, only a fraction of events satisfying the trigger requirement were recorded at low and medium $p_T^{\text{jet}}$ due to computational power and data storage limitations. The known fractional scaling of the trigger rate of a given class of events, named “prescale”, is optimized for the best bandwidth usage and data collection yield.

3.2 Monte Carlo Simulated Samples

Samples of simulated multijet events from strong interaction processes are generated with the Pythia 8.186 [17] MC generator, referred to as Pythia 8 in the following, with the NNPDF 2.3 Leading Order (LO) parton distribution functions (PDFs) [18]. This generator utilizes leading-order perturbative quantum chromodynamics (pQCD) matrix elements, along with a leading-logarithmic parton shower [19], an underlying event (UE) simulation with multiple parton interactions, and the Lund string model for hadronisation [20].

The parameters for the modelling of the interaction features not represented by the matrix element are provided by the A14 tune [21]. Alternative simulated samples of the same class of events are generated with the HERWIG++ 2.7.1 MC generator [22], referred to as HERWIG++ in the following, with the CTEQ6L1 LO PDFs [23] and the UEEE5 [24] tune for the modelling of the interaction features not represented by the matrix elements (parton shower, hadronization, underlying event).

Generated events are propagated through a full simulation of the ATLAS detector [25] based on Geant4 [26] that simulates the particle interactions with the detector material. All of the generated events were processed with the nominal ATLAS software version used for physics analysis during the year 2016.

The effect of additional proton-proton interactions per bunch crossing (pileup) is accounted for by overlaying onto the hard-scattering process minimum-bias events generated with Pythia 8. Different pileup conditions between data and simulation are taken into account by reweighting the mean number of interactions per bunch crossing in simulation to the one observed in data.

The difference, at the few percent level, between data and simulation regarding the jets association to the primary hard-interaction vertex [27] (discussed in the next section) is corrected by means of event-wise scale factors applied to the simulation [28]. No corrections related to $b$-tagging performance are applied.

4 Jet reconstruction and $b$-tagging algorithms

Hadronic jets are reconstructed from clustered energy deposits [29] in the ATLAS calorimeter with the anti-$k_t$ algorithm [30] and a parameter $R = 0.4$. Jets considered in this analysis are required to be within the ID pseudorapidity acceptance ($|\eta^{\text{jet}}| < 2.5$), satisfy cleaning criteria [31], originate from the primary hard-interaction vertex [27], and have $p_T^{\text{jet}} > 20$ GeV after final jet energy scale calibration [32]. For the purpose of $b$-tagging, tracks are associated to jets on the basis of a shrinking cone cut on the angular separation between the track and the jet axis directions, as detailed in Ref. [8]. Simulated jets are labelled according to their flavour by spatially matching the jet with generator level hadrons: if a $b$-hadron with $p_T > 5$ GeV is found within $\Delta R = 0.3$ of the jet direction, the jet is labelled as a $b$-jet, if no $b$-hadrons are found the procedure is repeated for $c$-hadrons and $\tau$ leptons, and a jet is labelled as light-flavour if none of the preceding associations can be made.
The $b$-tagging algorithm most commonly used in the ATLAS analyses of data collected in the years 2015-2016 is called MV2c10 [9, 10], a multi-variate method trained to discriminate $b$-jets against LF-jets. The MV2c10 algorithm combines in a single variable the information coming from several lower-level algorithms briefly described in the following.

- **Impact Parameter based Algorithms (IP2D and IP3D):** due to the long lifetime of $b$-hadrons, tracks generated from $b$-hadron decay products tend to have larger impact parameters than tracks associated with LF-jets. The IP2D and IP3D algorithms make use of the signed impact parameter significance of the tracks associated with the jet, introduced previously in Section 2. IP3D uses both the transverse ($d_0$) and longitudinal ($z_0$) impact parameter significances taking into account their correlations, while IP2D uses the $d_0$ significance only. The probability density functions (pdfs) for the signed impact parameter significance of these tracks are used to define ratios of the $b$, $c$ and light-flavour jet hypotheses, and these are then combined in three log likelihood ratio discriminants (LLR): $b$/LF, $b/c$ and $c$/LF.

- **Secondary Vertex Finding Algorithm (SV1):** SV1 [33] aims to explicitly reconstruct an inclusive displaced secondary vertex within the jet. The first step consists of reconstructing two-track vertices using the candidate tracks. Tracks are rejected if they form a secondary vertex which can be identified as likely originating from the decay of a long-lived particle (e.g. $K_s$ or $\Lambda$), photon conversions or hadronic interactions with the detector material. A new vertex is then fitted with all tracks surviving this selection, outlier tracks being iteratively removed from this set of tracks. The vertex information is condensed in the eight following observables: the invariant mass of the tracks at the secondary vertex assuming pion masses, the ratio of the sum of the energies of these tracks to the sum of the energies of all tracks associated with the jet, the number of tracks used in the secondary vertex, the number of two-track vertex candidates, the transverse distance between the primary and secondary vertices, the total distance between the primary and secondary vertices, the distance between the primary and secondary vertices divided by its uncertainty (i.e. the impact parameter significance), and the $\Delta R$ between the jet axis and the direction of the secondary vertex relative to the primary vertex.

- **Decay Chain Multi-Vertex Algorithm (JetFitter):** JetFitter exploits the topological structure of weak $b$- and $c$-hadron decays inside the jet and tries to reconstruct the full $b$-hadron decay chain. A Kalman filter is used to find a common line on which the primary vertex and the HF vertices lie, approximating the $b$-hadron flight path as well as their positions. Hence, HF vertices can be resolved even when only a single track is attached to them whenever the resolution allows. The vertex information is condensed in the eight following observables: the number of two-track vertex candidates prior to the decay chain fit, the invariant mass of the tracks from displaced vertices assuming pion masses, the ratio of the sum of the energies of these tracks to the sum of the energies of all tracks associated with the jet, the number of displaced vertices with one track, the number of displaced vertices with more than one track, the number of tracks from displaced vertices with at least two tracks and the $\Delta R$ between the jet axis and the vector-sum of the momenta of all tracks attached to displaced vertices.

The MV2c10 multivariate algorithm is trained using a background sample including 7% of $c$-jets and 93% of LF-jets. The algorithm assigns to each jet a $b$-tagging output, $w$, ranging from -1 to 1. The distribution for LF jets peaks towards -1 and that for $b$-jets towards +1, and MV2c10 values for $c$-jets tend to lie between the two. Different WPs can be defined according to the desired tagging efficiency required for
b-jets, evaluated on simulated $t\bar{t}$ events, by requiring a minimum threshold on the MV2c10 output \cite{ATL-PHYS-PUB-2016-008}. The thresholds commonly used in ATLAS 2015-2016 analyses are listed in Table 1.

<table>
<thead>
<tr>
<th>WP</th>
<th>Cut value $X$</th>
<th>$b$-jet efficiency ($\varepsilon_b$)</th>
<th>$c$-jet mistag rate ($\varepsilon_c$)</th>
<th>LF-jet mistag rate ($\varepsilon_{LF}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>85%</td>
<td>0.1758</td>
<td>85%</td>
<td>32%</td>
<td>2.9%</td>
</tr>
<tr>
<td>77%</td>
<td>0.6459</td>
<td>77%</td>
<td>16%</td>
<td>0.77%</td>
</tr>
<tr>
<td>70%</td>
<td>0.8244</td>
<td>70%</td>
<td>8.3%</td>
<td>0.26%</td>
</tr>
<tr>
<td>60%</td>
<td>0.9349</td>
<td>60%</td>
<td>2.9%</td>
<td>0.065%</td>
</tr>
<tr>
<td>50%</td>
<td>0.9769</td>
<td>50%</td>
<td>0.94%</td>
<td>0.017%</td>
</tr>
</tbody>
</table>

Table 1: $b$-tagging working points considered in this document. Each WP is defined by a cut value $X$ on the MV2c10 output distribution ($w > X$, $w$ ranging in $[-1, 1]$). The resulting $b$-tagging efficiency ($\varepsilon_b$) and $c$- and LF-jet mistag rates ($\varepsilon_c$, $\varepsilon_{LF}$) as measured in a $t\bar{t}$ simulated sample are also shown.

5 Calibration with the negative tag method

5.1 Method description

The negative-tag method relies to a large extent on the assumption that LF-jets are mistagged as $b$-jets mainly because of the finite resolution of the reconstructed inner detector track trajectories and impact parameters. Under this assumption, the signed impact parameter distribution of the tracks associated with LF-jets is expected to be symmetric around zero. This property is approximately true in simulated events, as shown in Figure 1. On the contrary, the tracks originating from $b$- and $c$-hadrons (and so associated with $b$- and $c$-jets) exhibit a signed impact parameter distribution with a higher tail at large positive values due to the long lifetime \cite{ATL-PHYS-PUB-2016-008}. This difference is one of the main features used by the IP2D and IP3D algorithms to discriminate between $b$-, $c$- and LF-jets. Therefore, the tag rate of LF-jets obtained by running the IP2D and IP3D algorithm after flipping the impact parameter sign of the jet tracks is expected to be a good approximation of the LF-jet mistag rate for these algorithms. Moreover, one expects such a tag rate to be comparable for the three flavors because the $b$-, $c$- and LF-jets impact parameter distribution tails are much more similar on the negative side than on the positive side. A similar feature is expected for distributions related to secondary vertices with signed decay length significance, which are seeded from tracks, such that this reasoning can be extended to the SV1 and JetFitter algorithm.

A dedicated algorithm denoted MV2c10Flip is defined in the following. MV2c10Flip aims to tag jets that contain a significant amount of tracks with negative impact parameters and secondary vertices with negative lifetime, i.e. with an angle between the jet direction and the line joining the primary vertex to the point of closest approach of the track or to the secondary vertex less than $\pi/2$. MV2c10Flip uses the same input variables as MV2c10 and was not retrained. However the input variables are determined from modified, “flipped”, versions of the IP2D, IP3D, SV1 and JetFitter algorithms, denoted respectively IP2DNeg, IP3DNeg, SV1Flip and JetFitterFlip. These are defined as follows:

- **Negative Impact Parameter based Algorithms (IP2DNeg, IP3DNeg):** only tracks with negative $d_0$ are selected as input. The $d_0$ sign of the selected tracks is flipped before running the IP2D/IP3D
Figure 1: Normalised signed-\(d_0\) (left) and signed-\(z_0\) (right) distributions for tracks associated to LF-, \(c\)- and \(b\)-jets in multijet events generated with the Pythia 8 event generator.

algorithms. In the standard versions of IP2D/IP3D, both positive and negative \(d_0\) tracks are used as input and no flipping of the \(d_0\) sign is performed.

• **Flipped Secondary Vertex Finding Algorithm (SV1Flip):** the signed impact parameter selection applied to the candidate tracks used to reconstruct two-track vertices is inverted. Inclusive secondary vertices reconstructed behind the PV with respect to the jet direction (i.e. negative lifetime vertices) are identified by the algorithm. The negative direction of the jet axis is used to compute observables of interest, as the \(\Delta R\) between jet and secondary vertex directions, and, where needed, the absolute value of quantities is used in order to compute positive-definite MV2c10Flip input variables.

• **Flipped Decay Chain Multi-Vertex Algorithm (JetFitterFlip):** the \(d_0\) sign of the tracks used as input for two-track vertex reconstruction is flipped before running the JetFitter algorithm. Once defined, the sign of the lifetime of the two-track vertex candidates used to seed tracks for JetFitter is also flipped. Finally, only the final vertices with negative lifetime are considered to compute the MV2c10Flip input variables.

Figure 2 shows one example (the appendix reports additional variables) of the effect of using flipped versus standard algorithms on one of the MV2c10 inputs: the response to LF-jets of the secondary vertex mass of SV1Flip and SV1 algorithms remains about the same while the \(b\)-jet discrimination power is drastically reduced in SV1Flip.

A jet is considered negatively tagged if the MV2c10Flip output, ranging between -1 and 1, satisfies the nominal WP threshold value defined earlier in Table 1. Following the reasoning explained above, MV2c10Flip is expected to have a LF-jet mistag rate similar to MV2c10 but much lower \(b\)-, \(c\)-jet selection efficiency. This reduces the impact of the \(c\)- and \(b\)-jets, present in the data, on the mistag rate measurement.

Hence, the negative-tag method consists of measuring the negative-tag rate, \(\varepsilon_{\text{neg}}^{\text{data}}\), in a data sample enriched in LF-jets and extracting from it the calibration of the mistagging rate of LF-jets in the standard \(b\)-tagging configuration. However, several aspects of such LF-jet calibration extraction rely on simulation. In particular the HF and LF-jets negative-tag rates are not expected to be identical because of the presence of
mis-signed tracks from HF-hadron decays in HF-jets due to resolution effects and misalignment between the jet axis and the HF-hadron flight path. Furthermore LF-jets can be mistagged because of the presence of tracks from displaced vertices of long-lived particles (e.g. $K_S$ or $\Lambda$) and material interactions (hadronic interactions and photon conversions), which introduce a bias towards a positive value in the impact parameter distribution of LF-jet tracks. Correction factors accounting for true $b$- and $c$-jets in the sample and true displaced vertices due to long-lived particle decays and material interactions in LF-jets are derived from MC as discussed in the next paragraphs.

The following formula is used to extract the LF-jet mistagging rate calibration:

$$\epsilon_{LF} = \epsilon_{LF, neg}^{MC} \cdot K_{LL}^{MC}, \quad (1)$$

where $\epsilon_{LF}$ is the LF-jet mistag rate, $\epsilon_{LF, neg}^{data}$ is the inclusive negative-tag rate measured in data and $K_{LL}^{MC}$ and $K_{HF}^{MC}$ are correction factors extracted from MC simulation and defined as follows:

$$K_{LL}^{MC} = \left( \frac{\epsilon_{LF}}{\epsilon_{LF, neg}} \right)^{MC}, \quad K_{HF}^{MC} = \left( \frac{\epsilon_{LF, neg}}{\epsilon_{all, neg}} \right)^{MC}, \quad (2)$$

where $\epsilon_{LF}$ and $\epsilon_{LF, neg}$ are respectively the LF-jet mistag rate and negative LF-jet mistag rate for LF-jets and $\epsilon_{all, neg}$ is the negative-tag rate in the sample, inclusive in flavor. The use of Equation 1 requires reliable MC predictions concerning:

1. the fraction $f_{HF}$ of HF-jets before any tag,
2. the ratio of negative-tag efficiencies for LF- and HF-jets, $K_{HF}^{LF} = \frac{\epsilon_{HF, neg}}{\epsilon_{HF, neg}}$, since $K_{HF}^{MC} = \frac{1}{f_{HF} \times K_{HF}^{LF} + (1-f_{HF})}$,
3. the ratio of normal and negative-tag efficiencies for LF-jets, $K_{LL}^{MC}$.

In the extreme case of HF-jets dominating the negative-tag sample, practically only $\epsilon_{HF, neg}$ is measured in data and, because of condition 2, which constrains data/MC differences to be the same for $\epsilon_{LF, neg}$ and $\epsilon_{HF, neg}$, the measurement of the negative-tag rate of HF-jets would heavily influence the extrapolation to the LF-jet mistag rate, $\epsilon_{LF}$. These conditions are met for the very tight selection WPs analysed (i.e. 60%...
and 70%) and at high $p_T^{\text{jet}}$ where the predicted fraction of HF-jets after negative tag exceeds 40% in most of the $p_T^{\text{jet}},|\eta^{\text{jet}}|$ bins, see Table 2. Conservative systematic uncertainties are adopted in order to account for this extrapolation, as discussed in Section 5.4.

5.2 Event selection

Events are required to have at least one reconstructed vertex with at least two associated well-reconstructed tracks. At least two jets selected according to the Section 4 criteria must be present to retain the event. If more than two jets satisfy these criteria, the two jets with the highest transverse momenta are selected.

In order to optimise the statistical power of the measurements, each $p_T^{\text{jet}}$ bin considered in the measurements makes use of the single-jet trigger with lowest prescale (i.e. with highest recorded integrated luminosity) that is more than 99.9% efficient in that range. This condition cannot be achieved in the lowest $p_T^{\text{jet}}$ bin (20 GeV < $p_T^{\text{jet}}$ < 60 GeV) and therefore a kinematic bias from trigger inefficiency affects the highest transverse momentum jet (leading jet) distribution in data in this range. This issue is overcome by basing the nominal measurements in 20 GeV < $p_T^{\text{jet}}$ < 60 GeV in subleading jets. The other measurements ($p_T^{\text{jet}}$ > 60 GeV) are based on the leading jets. The integrated luminosity of the data sample depends on $p_T^{\text{jet}}$ and ranges from 0.02 pb$^{-1}$ (20 GeV < $p_T^{\text{jet}}$ < 60 GeV, highly prescaled) to 36.1 fb$^{-1}$ ($p_T^{\text{jet}}$ > 500 GeV, unprescaled).

A good angular separation between the two jets in the transverse plane ($\Delta\phi_{jj}$ > 2 rad.) is also required in order to reject events with high transverse momentum jets originating from the hadronization of a gluon which split into two heavy-flavour quarks or originating from beam-induced background [34].

The final data and simulated samples are split in two $|\eta^{\text{jet}}|$ intervals (\(|\eta^{\text{jet}}|<1.2, 1.2<|\eta^{\text{jet}}|<2.5\)) called central and forward in the following, which are further divided in eight $p_T^{\text{jet}}$ intervals (20-60, 60-100, 100-200, 200-300, 300-500, 500-750, 750-1000, 1000-3000 GeV). A mistag rate measurement is performed in each interval. The number of selected jets used in the calibration before any $b$-tagging requirement and prescale corrections ranges from about $6 \times 10^4$ to $2 \times 10^7$. The highest (lowest) statistics categories correspond to central (forward) jets in the mid- (low- and high-) $p_T^{\text{jet}}$ range(s). The amount of jets in data after a negative-tag selection corresponding to the 85% WP (see Section 5.3) ranges from a few thousands of events to a few hundreds of thousands of events. The number of jets in data after the negative-tag selection corresponding to the 60% WP (see Section 5.3) ranges from a few tens of events to a few thousands of events. The highest and lowest statistics categories stay the same before and after negative-tag requirement.

5.3 Expected performance of MV2c10 and MV2c10Flip

The MV2c10 and MV2c10Flip output distributions for $b$-, $c$- and LF-jets in the Pythia 8 multijet sample are shown in Figure 3 for jets with $p_T^{\text{jet}}$ > 60 GeV and $|\eta^{\text{jet}}| < 2.5$. All distributions are normalised to unity. As expected, the LF-jet output shape is comparable between the two algorithms and MV2c10Flip discriminates much less between the three types of jets than MV2c10. The first property allows a good estimate of the LF-jet mistag rate to be obtained using the negative tag. The second property reduces the impact of the $b$- and $c$-jet contamination on the measurement.

However, due to the effects described in Section 5.1, the normalised HF-jet MV2c10Flip distributions is higher than the normalised LF-jet MV2c10Flip distribution at values close to 1. The HF-jet contamination
Table 2: Fraction of heavy flavour jets for the 85% and 60% selection WPs for the MV2c10 and MV2c10Flip algorithms for different intervals of $p_T^{\text{jet}}$ and $|\eta^{\text{jet}}|$. The fractions are extracted from multijet events generated with the Pythia 8 event generator. The events are required to pass the event selection described in the Sections 3 and 5 and the leading jet is considered.

<table>
<thead>
<tr>
<th>Algorithm WP</th>
<th>$p_T^{\text{jet}}$ [GeV]</th>
<th>20-60</th>
<th>60-100</th>
<th>100-200</th>
<th>200-300</th>
<th>300-500</th>
<th>500-750</th>
<th>750-1000</th>
<th>1000-3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV2c10 85%, HF frac. (%)</td>
<td>$</td>
<td>\eta^{\text{jet}}</td>
<td>&lt;1.2$</td>
<td>43 ± 1</td>
<td>68.8 ± 0.6</td>
<td>71.0 ± 0.4</td>
<td>72.6 ± 0.2</td>
<td>70.2 ± 0.1</td>
<td>63.4 ± 0.1</td>
</tr>
<tr>
<td>MV2c10Flip 85%, HF frac. (%)</td>
<td>$</td>
<td>\eta^{\text{jet}}</td>
<td>&lt;1.2$</td>
<td>1.2 &lt; $</td>
<td>\eta^{\text{jet}}</td>
<td>&lt; 2.5$</td>
<td>40 ± 3</td>
<td>58.9 ± 0.6</td>
<td>61.9 ± 0.4</td>
</tr>
<tr>
<td>MV2c10 60%, HF frac. (%)</td>
<td>$</td>
<td>\eta^{\text{jet}}</td>
<td>&lt;1.2$</td>
<td>13 ± 1</td>
<td>30 ± 1</td>
<td>40.9 ± 0.7</td>
<td>47.5 ± 0.3</td>
<td>51.0 ± 0.2</td>
<td>47.1 ± 0.2</td>
</tr>
<tr>
<td>MV2c10Flip 60%, HF frac. (%)</td>
<td>$</td>
<td>\eta^{\text{jet}}</td>
<td>&lt;1.2$</td>
<td>1.2 &lt; $</td>
<td>\eta^{\text{jet}}</td>
<td>&lt; 2.5$</td>
<td>10 ± 1</td>
<td>24 ± 1</td>
<td>32.6 ± 0.8</td>
</tr>
</tbody>
</table>

Table 3 summarises the ranges of $K_{HF}^{MC}$ and $K_{LL}^{MC}$ in the various $p_T^{\text{jet}}$/$|\eta^{\text{jet}}|$ categories after a selection corresponding to the 85% and 60% WP are reported in Table 2 for the MV2c10 and MV2c10Flip algorithms. Also the normalised LF-jets MV2c10 distribution substantially differs from the normalised LF-jet MV2c10Flip distribution at values close to 1. The $K_{HF}^{MC}$ and $K_{LL}^{MC}$ correction factors are therefore respectively smaller and higher than unity for all WPs. Both effects increase at higher MV2c10 output values such that the deviations from unity of $K_{HF}^{MC}$ and $K_{LL}^{MC}$ get higher for tight WPs. Table 3 summarises the ranges of $K_{HF}^{MC}$ and $K_{LL}^{MC}$. The individual $K_{HF}^{MC}$ and $K_{LL}^{MC}$ values for tight WPs (70%, 60%) are affected by large MC statistical uncertainties (up to 50%) due to the tiny LF-jet mistag rate of the MV2c10 algorithm in this range (see Table 1).

<table>
<thead>
<tr>
<th>WP</th>
<th>$K_{HF}^{MC}$ range</th>
<th>$K_{LL}^{MC}$ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>85%</td>
<td>0.6-0.9</td>
<td>1.3-2.8</td>
</tr>
<tr>
<td>77%</td>
<td>0.3-0.8</td>
<td>1.9-5.5</td>
</tr>
<tr>
<td>70%</td>
<td>0.1-0.8</td>
<td>2.9-12.5</td>
</tr>
<tr>
<td>60%</td>
<td>0.04-0.6</td>
<td>7-45</td>
</tr>
</tbody>
</table>

Table 3: $K_{HF}^{MC}$ and $K_{LL}^{MC}$ range for the 85%, 77%, 70% and 60% working points in the various $p_T^{\text{jet}}$/$|\eta^{\text{jet}}|$ bins considered in the calibration. Events generated with the Pythia 8 event generator are used.

The MV2c10 and MV2c10Flip output distributions in data and in the Pythia 8 multijet sample (all flavours) are shown in Figure 4. In order to avoid any bias from jet kinematics, the number of MC events is scaled to the number of events observed in data in each $p_T^{\text{jet}}$, $|\eta^{\text{jet}}|$ bin considered in the measurement. A rising slope in the data to MC ratio is observed for both taggers, indicating differences in the mistag rate in data and simulation.

5.4 Results

The LF-jet mistag rate measurement obtained by using Equation 1 allows to derive data to simulation correction factors, also called scale factors $S_{LF} = \frac{c_{LF}}{c_{MC}}$. These are shown for the 85%, 77%, 70% and
Figure 3: Normalised MV2c10 (left) and MV2c10Flip (right) output distribution for LF, c- and b-jets in multijet events generated with the Pythia 8 event generator. The highest $p_T$ jet in the event is considered, when this jet satisfies $p_T > 60$ GeV.

Figure 4: MV2c10 (left) and MV2c10Flip (right) output distribution for data and multijet events generated with the Pythia 8 event generator. The MC events are split by jet flavour. The number of MC events is scaled to the total number of data events in each $p_T$, $|\eta|$ bin considered in the measurement regardless of the jet flavour. The highest $p_T$ jet in the event is shown, when this jet satisfies $p_T > 60$ GeV.

60% WPs of the MV2c10 algorithm in Figure 5. They are approximately constant as a function of $p_T$ and $|\eta|$, around a value of 2.

The corresponding LF-jet mistag rate measurements are presented and compared to the nominal rates obtained from the Pythia 8 multijet sample (i.e. without any calibration or rescaling applied) in Figure 6 for the 85%, 77%, 70% and 60% WPs of the MV2c10 algorithm. The LF-jet mistag rate measurements range from 4 to 16% for the 85% WP, 1.5 to 6% for the 77% WP, 0.4 to 2.3% for the 70% WP and 0.1
to 0.4% for the 60% WP. The LF-jet mistag rates are relatively stable as a function of $p_T^{\text{jet}}$ for $p_T^{\text{jet}} < 800$ GeV and then increase significantly. A higher LF-jet mistag rate (by +20 to +50%) is measured in $1.2 < |\eta^{\text{jet}}| < 2.5$ with respect to $|\eta^{\text{jet}}| < 1.2$, expected from the degradation of track impact parameter resolution at high $|\eta|$. Uncertainties affecting the measurement that originate from the finite numbers of data and MC events are considered as well as systematic uncertainties affecting the $K^{\text{MC}}_{\text{HF}}$ and $K^{\text{MC}}_{\text{LL}}$ correction factor values. The bootstrap resampling technique [35] is used to assess the statistical uncertainties by creating an ensemble of statistically equivalent measurements using event weights, randomly chosen for each data and MC event from a Poisson distribution with a mean of 1. The standard deviation of the distribution of these measurements is taken as the statistical uncertainty. The systematic uncertainties are derived by varying a parameter in the simulated events, recomputing $K^{\text{MC}}_{\text{HF}}$ and $K^{\text{MC}}_{\text{LL}}$ with this varied parameter and taking as uncertainty the difference between the new evaluation of Equation 2 and the nominal measurement.

The uncertainties on the LF-jet mistag rate measurements obtained with the negative-tag method are summarised in Tables 4 to 7. The total uncertainty, obtained by summing all the listed uncertainties in quadrature, is dominated by systematic uncertainties. The total uncertainty ranges from 9-23% (15-34%, 18-39%, 22-76%) for the 85% (77%, 70%, 60%) WP. The total uncertainty increase from loose to tight WPs is expected due to the large increase of the $K^{\text{MC}}_{\text{HF}}$ and $K^{\text{MC}}_{\text{LL}}$ correction factors introduced at the end of the previous section. In the following, the labels between quotes refer to the uncertainty categories appearing in Tables 4 to 7.

The main systematic uncertainties arise from differences between data and MC in terms of track impact parameter resolution and fake track rate (“IP, fakes”) and from the uncertainty in the fraction of HF-jets present in the data sample before and after negative tag (“HF-related”). They will be described in further detail in the following paragraph. The uncertainty in jet calibration (“calibration”) and pileup modelling [36] (“pileup”) is negligible, except for the tight WPs, 70% and 60%, for the lowest $p_T^{\text{jet}}$ bin, $20 < p_T^{\text{jet}} < 60$ GeV. The uncertainty in the amount of displaced vertices in jets (“vertices”) due to material interactions and strange hadron decays, evaluated by increasing their rate by 10% (material interactions) and 30% (strange hadron decays), reaches up to 6% for the 60% WP but remains negligible with respect to the other sources in all other bins. These variations are motivated by previous ATLAS measurements [37, 38]. The stability of the measurement is investigated by repeating the measurement based on the accompanying jet, i.e. the leading jet in the lowest $p_T^{\text{jet}}$ bin (20-60 GeV) and the subleading jet otherwise ($p_T^{\text{jet}} > 60$ GeV). A conservative uncertainty of ±5% is adopted in all the bins to cover for the small differences observed (“sample”). The measurement is also repeated using correction factors extracted from the HERWIG++ sample in order to evaluate the dependence of the correction factors on the parton shower model. No significant deviations are observed and therefore no extra uncertainties are considered.

The uncertainty in track IP resolution and fake track rate is evaluated by adjusting the MC to reflect the results of the data-driven efficiency and fake rate measurements described in Section 6.2. The uncertainty in the fraction of HF-jets in the data sample before and after negative tag is evaluated by varying sequentially the HF-jet fraction in the MC before tag and the HF-jet negative-tag efficiency. The HF-jet fraction is varied on the basis of the MV2c10 output distribution agreement with data before tag: variation of the $b$-jet ($c$-jet) predicted fractions of up about 20% (200%), depending on the $p_T^{\text{jet}}$ and $|\eta^{\text{jet}}|$, were observed to improve the data/MC agreement and therefore taken as uncertainties. The HF-jet negative-tag efficiency variations consist in applying the $b$- and $c$-jet calibration scale factors observed for MV2c10 [11, 39–41] after doubling or halving their deviations from unity. The two types of uncertainties show a significant
asymmetric behaviour going in opposite directions: the track IP resolution and fake rate are respectively over- and underestimated in the nominal MC simulation, therefore tracking uncertainties tend to increase the mistag rate measurement by increasing $K_{HF}^{MC}$. On the contrary, data/MC comparison of MV2c10 in the various $p_T^{\text{jet}}/|\eta^{\text{jet}}|$ bins used in the measurement before tag shows that a better data/MC agreement is obtained when increasing the HF-jet fraction in the Pythia 8 sample. Reducing that fraction in the MC leads to a decrease of $K_{HF}^{MC}$, which translates into a negative variation of the mistag rate.

<table>
<thead>
<tr>
<th>$p_T^{\text{jet}}$ [GeV]</th>
<th>20-60</th>
<th>60-100</th>
<th>100-200</th>
<th>200-300</th>
<th>300-500</th>
<th>500-750</th>
<th>750-1000</th>
<th>1000-3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta^{\text{jet}}</td>
<td>$ range</td>
<td>cen</td>
<td>fwb</td>
<td>cen</td>
<td>fwb</td>
<td>cen</td>
<td>fwb</td>
</tr>
<tr>
<td>mistag (%)</td>
<td>4.3</td>
<td>6.6</td>
<td>4.2</td>
<td>6.1</td>
<td>4.7</td>
<td>6.1</td>
<td>4.8</td>
<td>7.1</td>
</tr>
<tr>
<td>data/MC</td>
<td>1.7</td>
<td>2.0</td>
<td>1.8</td>
<td>2.0</td>
<td>1.9</td>
<td>2.0</td>
<td>1.9</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 4: LF-jet mistag rate measurements with the negative-tag method for the MV2c10 85\% WP and main source of uncertainties. “cen” (“fwd”) denotes $|\eta^{\text{jet}}| < 1.2$ ($1.2 < |\eta^{\text{jet}}| < 2.5$) and “data $\oplus$ MC” means that data and MC statistical uncertainties are added in quadrature. “total” corresponds to the sum in quadrature of “data $\oplus$ MC” and all the systematic uncertainties, labelled “stat+syst” in Figures 5 and 6. One-sided uncertainties imply that the opposite variation is equal to zero.
<table>
<thead>
<tr>
<th>$p_T^{\text{jet}}$ [GeV]</th>
<th>20-60</th>
<th>60-100</th>
<th>100-200</th>
<th>200-300</th>
<th>300-500</th>
<th>500-750</th>
<th>750-1000</th>
<th>1000-3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta^{\text{jet}}</td>
<td>$ range</td>
<td>cen fwd</td>
<td>cen fwd</td>
<td>cen fwd</td>
<td>cen fwd</td>
<td>cen fwd</td>
<td>cen fwd</td>
</tr>
<tr>
<td>mistag (%)</td>
<td>1.3 2.2</td>
<td>1.3 1.9</td>
<td>1.2 1.8</td>
<td>1.4 2.3</td>
<td>1.7 2.4</td>
<td>1.6 2.2</td>
<td>3.0 3.9</td>
<td>4.7 5.9</td>
</tr>
<tr>
<td>data/MC</td>
<td>1.9 2.3</td>
<td>2.1 2.3</td>
<td>1.9 2.1</td>
<td>1.9 2.0</td>
<td>1.9 1.9</td>
<td>1.9 2.2</td>
<td>2.1 2.2</td>
<td>2.2 2.3</td>
</tr>
</tbody>
</table>

**statistical uncertainties (\%, relative)**

| data @ MC | ±9 ±10 | ±9 ±10 | ±4 ±5 | ±2 ±3 | ±1 ±2 | ±1 ±2 | ±1 ±3 |
| jet uncertainties (\%, relative) | | | | | | | |
| calibration | +7 −9 | +9 −10 | +6 −5 | +3 −2 | ±2 ±2 | ±2 ±2 | ±3 −1 |
| HF-related | +8 −14 | +5 −14 | +7 −9 | +11 −13 | +12 −16 | +10 −17 | +11 −19 | +8 −25 |

**tracking uncertainties (\%, relative)**

| IP, fakes | +13 −1 | +10 −3 | +30 −2 | +25 −2 | +20 −4 | +17 −5 | +13 −7 | +14 −9 |
| vertices | +4 −1 | +3 −2 | +6 −2 | +5 −4 | +5 −4 | +5 −4 | +4 ±4 | ±4 ±3 |
| ±2 ±1 | | | | | | | | |

**other uncertainties (\%, relative)**

| sample | ±5 ±5 | ±5 ±5 | ±5 ±5 | ±5 ±5 | ±5 ±5 | ±5 ±5 | ±5 ±5 | ±5 ±5 |
| pileup | ±11 −6 | ±1 ±2 | ±3 −3 | ±2 ±2 | < 1 −1 | < 1 −1 | < 1 ±2 | < 1 ±1 |
| total | ±21 −20 | ±29 −18 | ±26 −15 | ±34 −17 | ±28 −17 | ±29 −18 | ±24 −19 | ±25 −19 |

Table 5: LF-jet mistag rate measurements with the negative-tag method for the MV2c10 77\% WP and main source of uncertainties. "cen" ("fwd") denotes $|\eta^{\text{jet}}| < 1.2 (1.2 < |\eta^{\text{jet}}| < 2.5)$ and "data ⊕ MC" means that data and MC statistical uncertainties are added in quadrature. "total" corresponds to the sum in quadrature of "data ⊕ MC" and all the systematic uncertainties, labelled "stat+syst" in Figures 5 and 6. One-sided uncertainties imply that the opposite variation is equal to zero.
Figure 5: Ratio between the LF-jet mistag rate measured from data with the negative-tag method and as simulated using the Pythia 8 multijet sample for the 85% (top row), 77% (second row from top), 70% (third row from top) and 60% WPs (bottom row) as a function of $p_T$ for $|\eta^{\text{jet}}| < 1.2$ (left) and $1.2 < |\eta^{\text{jet}}| < 2.5$ (right). The negative-tag measurements include simulation-based corrections for HF-jet contamination and LF-jets with true secondary vertices. The statistical uncertainty represents the sum in quadrature of data and MC statistical uncertainties.
Figure 6: LF-jet mistag rate for the 85% (top row), 77% (second row from top), 70% (third row from top) and 60% WPs (bottom row) as a function of $p_T^{\text{jet}}$ for $|\eta^{\text{jet}}| < 1.2$ (left) and $1.2 < |\eta^{\text{jet}}| < 2.5$ (right) as measured with the negative-tag method and as simulated using the Pythia 8 multijet sample without any calibration or rescaling applied. The negative-tag measurements include simulation-based corrections for HF-jet contamination and LF-jets with true secondary vertices. The statistical uncertainty associated with the negative-tag method measurements represents the sum in quadrature of data and MC statistical uncertainties.
Table 6: LF-jet mistag rate measurements with the negative-tag method for the MV2c10 70% WP and main source of uncertainties. “cen” (“fwd”) denotes $|\eta^{jet}| < 1.2$ ($1.2 < |\eta^{jet}| < 2.5$) and “data ⊕ MC” means that data and MC statistical uncertainties are added in quadrature. “total” corresponds to the sum in quadrature of “data ⊕ MC” and all the systematic uncertainties, labelled “stat+syst” in Figures 5 and 6. One-sided uncertainties imply that the opposite variation is equal to zero.
<table>
<thead>
<tr>
<th>$p_T^{j}$ [GeV]</th>
<th>20-60</th>
<th>60-100</th>
<th>100-200</th>
<th>200-300</th>
<th>300-500</th>
<th>500-750</th>
<th>750-1000</th>
<th>1000-3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta</td>
<td>^{j}$ range</td>
<td>cen</td>
<td>fwd</td>
<td>cen</td>
<td>fwd</td>
<td>cen</td>
<td>fwd</td>
</tr>
<tr>
<td>mistag (%)</td>
<td>0.35</td>
<td>0.15</td>
<td>0.13</td>
<td>0.17</td>
<td>0.12</td>
<td>0.15</td>
<td>0.14</td>
<td>0.20</td>
</tr>
<tr>
<td>data/MC</td>
<td>3.3</td>
<td>1.4</td>
<td>1.5</td>
<td>1.6</td>
<td>1.7</td>
<td>1.5</td>
<td>1.4</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Table 7: LF-jet mistag rate measurements with the negative-tag method for the MV2c10 60% WP and main source of uncertainties. “cen” (“fwd”) denotes $|\eta|^{j}$ $<$ 1.2 (1.2 $<$ $|\eta|^{j}$ $<$ 2.5) and “data $\oplus$ MC” means that data and MC statistical uncertainties are added in quadrature. “total” corresponds to the sum in quadrature of “data $\oplus$ MC” and all the systematic uncertainties, labelled “stat+syst” in Figures 5 and 6. One-sided uncertainties imply that the opposite variation is equal to zero.
6 Calibration with the adjusted Monte Carlo method

6.1 Method Description

As previously discussed, the \(b\)-tagging LF-jet mistag-rate is largely determined by the track-reconstruction resolution and fake-rate, where differences between MC simulation and data result in unequal \(b\)-tagging performance. The necessary correction factors can be obtained by adjusting the simulation of the tracking observables to match data measurements and comparing the \(b\)-tagging performance in the “standard” and in the “adjusted” MC environments. This novel bottom-up calibration procedure, named “adjusted-MC method”, can be summarised as follows:

1. Accurate measurements of performance of the ATLAS charged-particle tracking system in both data and simulation in terms of track impact-parameter modelling, track reconstruction efficiency, fake-rate, long-lived hadron fraction, and particles produced by interaction with the detector material.

2. Modification of the MC to reproduce the tracking performance observed in data and re-evaluation of the \(b\)-tagging algorithm output weight for each jet.

3. Evaluation of the mistag rate in the standard simulation and in the adjusted simulation in order to obtain two different LF-jet tagging efficiencies, \(\varepsilon_{\text{LF}}^{\text{MC}}\) and \(\varepsilon_{\text{LF}}^{\text{MC-adj}}\), respectively. The LF-jet calibration scale-factor, \(SF_{\text{LF}}\), is given by:

\[
SF_{\text{LF}} \equiv \frac{\varepsilon_{\text{LF}}^{\text{MC-adj}}}{\varepsilon_{\text{LF}}^{\text{MC}}}
\]

The adjusted-MC calibration method relies on data-driven calibrations, but measured before the \(b\)-tagging algorithm evaluation. This is a complementary approach to the negative-tag method, described in Section 5, where the calibration of the \(b\)-tagging algorithm is done directly on a data control sample enriched in LF-jets.

The adjusted-MC method is well suited for the calibration of the mistag rate because experimental effects, like the impact of the IP resolution in the measurement of the prompt tracks or the presence of fake tracks associated to a LF-jet, are the major contributors to the LF-jet tag-rate.

A thorough investigation of other possible sources of mismodelling of the LF-jet tag-rate was also performed to ensure that all major effects are considered in the calibration. This resulted in a few additional systematic uncertainties concerning, for example, the simulation of the interaction of charged particles with the detector material or the LF-jet hadronization model.

6.2 Data-driven Impact Parameter Measurements and Calibration

Precise measurement of charged particles track IP and correct modelling in simulation is fundamental for \(b\)-tagging performance. The IP of charged particle tracks of \(p_T \approx 1\) GeV \((p_T \approx 20\) GeV) is measured by the ATLAS tracking system with a resolution of \(O(100)\) \(\mu m\) \((O(10)\) \(\mu m\)). This results from a convolution of many effects, for example: intrinsic single-hit resolution, level of alignment of the tracking components, multiple-scattering inside the detector material, accuracy of track reconstruction algorithms. These effects are complex to simulate and, as a result, the track IP reconstruction performance in MC may need additional tuning [42] to precisely reproduce the data. The measurement in di-jet data and simulation of the IP resolution and the fraction of non-Gaussian tails in the IP distribution, two quantities of primary
importance for the likelihood of a LF-jet to be mistagged, is described in the following. The fraction of
non-Gaussian tails in the IP distribution is measured for the first time in the ATLAS experiment.

Tracks used for this study are associated to jets using the ghost-matching method [43] and selected
according to tight or loose identification criteria, described in detail in Ref. [44]. The two selection
criteria differ in the minimum number of required silicon pixel and strip hits associated to the track, with
tight being a sub-set of loose. Furthermore the tight criteria require the presence of one hit in the innermost
pixel detector layer. The tight and loose track selection criteria are meant to represent the various track
quality requirements used by the lower-level tagging algorithms listed Section 4. The results reported in
the following have been found to agree, within uncertainties, for the tight and loose track selections.

The resolution of track IP measurements is directly related to the capability of single-track reconstruction
of the ATLAS ID, however the resolution of the event-by-event reconstructed PV is also convoluted with
the IP resolution measurement. The contribution from the PV resolution, of the order or 10 μm in the
transverse plane, is sizeable for track-PT above a few GeV and it needs to be removed to determine the
track intrinsic IP resolution. The track IP resolution is “unfolded”, in both data and simulation, using an
iterative deconvolution of the PV resolution. This procedure is described in detail in Ref. [45] and it is
based on the following steps (the same considerations are valid for both d0 and z0):

- For a given track pT−η, the distribution of the measured d0,unbias of prompt tracks, R(d0,unbias),
  follows with good approximation the equation:

\[
R(d_{0,\text{unbias}}) = \int \exp\left[-\frac{1}{2} \frac{d_{0,\text{unbias}}^2}{\sigma^2(d_{0,\text{trk}}) + \sigma^2(d_{0,\text{PV}})}\right] P(\sigma(d_{0,\text{PV}})) \, d\sigma(d_{0,\text{PV}}),
\]

where the Gaussian distribution has a width equal to the sum in quadrature of the track intrinsic
resolution, σ(d0,trk), and of the PV resolution in the transverse plane. This is integrated over the
distribution, P(σ(d0,PV)), of PV resolution values originating from the considered tracks.

- The dependence on σ(d0,PV) is removed from Equation 4 using the following change of variables:

\[
d_{0,\text{unbias}} = d_{0,\text{unfold}} \sqrt{\frac{\sigma^2(d_{0,\text{trk}}) + \sigma^2(d_{0,\text{PV}})}{\sigma^2(d_{0,\text{trk}})}},
\]

which, when substituted in Equation 4, allows the integral to be factorised and a simple Gaussian
distribution of width σ(d0,trk) to be obtained.

- The mentioned change of variables relies on the knowledge of the unknown observable of interest,
  σ(d0,trk). The iterative method, described in Ref. [45], allows to obtain the approximate value of
  σ(d0,trk) starting from the d0,unbias variable. For each track the variable d0,unbias is corrected with
  an unfolding factor:

\[
d_{0,\text{unfold},(i)} \rightarrow d_{0,\text{unbias}} \sqrt{\frac{K_{(i-1)}^2 \sigma^2(d_{0,\text{unbias}})}{K_{(i-1)}^2 \sigma^2(d_{0,\text{unbias}}) + K_{PV}^2 \sigma^2(d_{0,\text{PV}})}},
\]

where the index (i) indicates the i-th iteration, σ(d0,PV) and σ(d0,unbias) are the errors on the
individual track and PV fits, and K_{(i)} = \frac{\hat{\sigma}(d_{0,\text{unbias},(i)})}{\sigma(d_{0,\text{unbias},(i)})}. The “\hat{\sigma}” quantities in the definition of K_{(i)}
are extracted using an iterative Gaussian fit of the 1.5σ “core” of the distribution of d0,unfold, (i) and
d0,unbias for the set of tracks of given pT−η. For the first iteration the original d0,unbias is used and
the iterations stop when \( K^{(i)} \approx 1 \) within a few percent, which results in \( \hat{\sigma}(d_{0,\text{unfold}}^{(i)}) \approx \sigma(d_{0,\text{trk}}) \). The correction factor \( K_{PV} \), equal to \( 1.25 \pm 0.08 \), accounts for differences between the measured and expected PV resolution.

- **Uncertainties on the di-jet IP measurement:** uncertainties, evaluated in both data and simulation, are due to: propagation of the unfolding statistical uncertainty, variation of the “core” definition of the resolution fits, propagation of the \( K_{PV} \) uncertainty, and track-density effects. Track density effects are evaluated by comparing the standard IP-resolution measurement and the measurement in the region of \( \Delta R \) between a track and the jet axis less than 0.02. The IP-resolution measurement extracted with a similar procedure in samples of \( Z \rightarrow \mu\mu \), for track \( p_T > 20 \text{ GeV} \), and minimum-bias [46], for track \( 0.4 < p_T < 10 \text{ GeV} \), was also performed; the difference with respect to the di-jet measurement gives an additional systematic uncertainty.

The measured track IP resolution was found to be stable, with \( K^{(i)} \approx 1 \) within a few percent, after two iterations and the results are summarised in Figure 7, showing the result for data and simulation, as a function of the track-\( p_T \), after averaging the 2-dimensional \( p_T-\eta \) results along \( \eta \).

![Figure 7: Measured track IP resolution after unfolding the contribution of the PV, in the transverse direction, \( \sigma(d_{0,\text{unfold}}) \) (left) and in the longitudinal direction, \( \sigma(z_{0,\text{unfold}}) \) (right). The top panels show the results of the PV deconvolution procedure for data (black) and simulation (red) as a function of the track-\( p_T \), after averaging the 2-dimensional \( p_T-\eta \) distribution along \( \eta \). Systematic plus statistic uncertainties are shown as a dashed area. The bottom panels show the ratio between data and MC data with corresponding uncertainties; correlations between systematic uncertainties are taken into account.](image)

The simulation can be corrected to match the IP resolution in data by using the bin-by-bin quadrature difference between data and MC resolutions, i.e.:

\[
\sigma_{\text{cor}}(d_{0}) = \sqrt{\sigma_{\text{Data}}^{2}(d_{0}) - \sigma_{\text{MC}}^{2}(d_{0})},
\]

where \( \sigma_{\text{cor}}(d_{0}) \) (and similarly for \( \sigma_{\text{cor}}(z_{0}) \)) is the additional resolution term to be added to the simulation of the track transverse (longitudinal) IP value to match the IP resolution of the MC to the one in data. The correction is obtained with a smearing procedure where a random value extracted from a Gaussian
distribution of zero mean and width equal to $\sigma_{\text{cor}}(d_0)$ ($\sigma_{\text{cor}}(z_0)$) is added to the $d_0$ ($z_0$) of each simulated track. Figure 8 shows the values of $\sigma_{\text{cor}}(d_0)$ and $\sigma_{\text{cor}}(z_0)$ averaged over the $\eta$ bins. The uncertainties on the di-jet IP measurement, summarised the last bullet of the preceding list, are propagated: systematic sources are added in quadrature and correlation between the same systematic sources in data and MC are taken into account, the uncertainty accounting for the difference in the IP resolution measurements in di-jet events and in other processes as $Z \rightarrow \mu\mu$ and minimum-bias [46] production\(^2\) is kept as an independent uncertainty.

As discussed, a single Gaussian function describes well the core of the track-IP distribution for a small enough interval of track $p_T-\eta$. A fraction of track-IP measurements do not follow this resolution function and populate the tails of the IP distribution. This can be due to a number of reasons which are challenging to simulate precisely: reconstruction issues, contamination from poorer quality tracks, presence of secondary particles due to hadronic interaction with the material, and long-lived HF-hadrons. The following studies aim to analyse the total shape of the IP distribution in data and simulation, quantify the fraction of tracks populating its tails, and correct possible mismodelling in the prompt\(^3\) component of the simulated tracks.

The analysis of the IP-tail distribution is done for sets of tracks of given $p_T-\eta$ associated to jets in data and MC. As the goal of the study is the correction of the prompt track component in MC, this category of tracks is studied separately. The fraction of tracks classified as prompt in simulation varies with $p_T$ and $\eta$, from a minimum of 40% at $p_T$ of the order of 1 GeV to a maximum of 75% at $p_T$ of 100 GeV, reach 60% at about 20 GeV.

The IP distribution is parameterized using the fit of a function described by a double Gaussian plus expo-

---

\(^2\) The minimum-bias events were collected under different detector and running conditions.

\(^3\) In this note tracks are classified as prompt if they are matched to a LF-jet but they do not originate from secondary hadronic interactions, converted photons, decay of strange-hadrons, nor from pile-up interaction vertices.
natural $C^1$-continuation\textsuperscript{4}: the first Gaussian corresponds to the core of the IP distribution, accommodating PV and track IP resolution, while the Gaussian with exponential continuation [47], dubbed Gauss-ExpC1, describes the IP-tail distribution. The exact form of this function is:

$$ g(d_0) = \begin{cases} I \frac{(1-f_T)}{\sqrt{2\pi}\sigma_C} \exp \left[ -\frac{1}{2} \frac{(d_0-\mu)^2}{\sigma_C^2} \right] + I \frac{1}{\sqrt{2\pi}\sigma_T} \exp \left[ -\frac{1}{2} \frac{(d_0-\mu)^2}{\sigma_T^2} \right], & \text{for } |d_0 - \mu| < K\sigma_T, \\ (1) \frac{(1-f_T)}{\sqrt{2\pi}\sigma_C} \exp \left[ -\frac{1}{2} \frac{(d_0-\mu)^2}{\sigma_C^2} \right] + I \frac{1}{\sqrt{2\pi}\sigma_T} \exp \left[ -\frac{1}{2} \frac{(d_0-\mu)^2}{\sigma_T^2} \right], & \text{for } |d_0 - \mu| \geq K\sigma_T, \end{cases} \tag{8} $$

where $I$ is the integral of the $d_0$ distribution, the core Gaussian is characterised by the width $\sigma_C$, the fraction of events included in the “tail” part of the function is described by $f_T$, so that $1 - f_T = f_C$ is the complementary fraction of events contained in the core Gaussian, $\mu$ parameterises any offset in the $d_0$ distribution, and finally the Gauss-ExpC1 tail function is characterised by the width $\sigma_T$ and the parameter $K$ describing the number of sigmas at which the Gaussian switches to the exponential decay in a continuous and $C^1$ way (these smoothness requirements ease the convergence of the fits); the normalisation factor $\frac{1}{A}$ is fixed by:

$$ A = \text{erf} \left( \frac{K}{\sqrt{2}} \right) + \frac{e^{-\frac{K^2}{2}}}{\sqrt{\pi}} \int_0^\infty \frac{1}{\sigma_T} \exp \left[ -\frac{1}{2} \frac{(d_0-\mu)^2}{\sigma_T^2} \right] \, \mathrm{d}d_0. \tag{9} $$

Figure 9 shows examples of the $d_0$ distribution fits for one interval of track $p_T$ for data, for simulation, and for simulation of only prompt-tracks. The function fits show satisfactory description of both data and simulation, with the residual small discrepancies having a negligible impact on the calibration discussed in the following.

The result of the parameterisation fits are summarised in Figure 10 by the $\eta$-averaged values of $f_T$ and of the exponential-tail decay constant $\sigma_T/K$ for track-IP measurements in data, in the inclusive simulation and in the simulation of only prompt tracks. The data and simulation agree within about 20% in the shape of the IP-tail distribution, summarised by the $\sigma_T/K$ value. On the other hand, the fraction of track-IP measurements populating the tails of the distribution, $f_T$, shows a sizable discrepancy of about 30%-40%. A correction scale factor is derived for this latter quantity:

$$ SF_{IP-tail} = \frac{f_{data}^{d_0}}{f_{MC}^{d_0}}. \tag{10} $$

The $SF_{IP-tail}$ correction has been found stable under the $\pm 20\%$ variations of the HF content in di-jet MC (see Section 5.4), and, as discussed in the following, separate studies are used to constrain the amount of interactions in the detector material or the reconstruction of fake tracks (both possible sources of large IP values). Therefore the $SF_{IP-tail}$ correction is expected to account mainly for mismodeling of prompt tracks in MC. The correction factor is propagated to prompt-tracks matched to LF-jets using the following procedure:

1. Each prompt-track of given $p_T$ is classified as “tail-track” or “core-track” thanks to the corresponding parameterisation of prompt-track IP distribution derived using Equation 8 (see also the red dashed line in Figure 10). The $d_0$ of the track is used to evaluate the fractional contribution, $p_{\text{tail}}$, of the Gauss-ExpC1 function of the Equation 8 with respect to the total curve, then a random number $r$ is thrown in the interval $[0, 1]$. If $r < p_{\text{tail}}$ the track is classified as a “tail-track”.

2. A “tail-track” is discarded from simulation with a probability $p_{\text{drop}} \equiv 1 - SF_{IP-tail}$. 

\textsuperscript{4} A continuous function is $C^1$ when also its first derivative is continuous.
Figure 9: Blue crosses show the distributions, and statistical uncertainty, of the $d_0$ of tracks matched to jets in data (top left), tracks matched to di-jet simulation (top right), and prompt-tracks matched to di-jet simulation (bottom centre), in the range $13 < p_T < 15$ GeV and $0.75 < \eta < 1.0$. The distributions are compared to the functional form described by Equation 8 (green line) and to its sub-components: the Gaussian function used for the description of the core of the distribution (magenta) and the Gauss-ExpC1 function describing the tails of the distribution (red). The bottom panels show the bin-by-bin difference between the data and the fits scaled by the statistical error on data. The reduced $\chi^2$ of the fit and the parameters of the tail parameterisation (i.e. $f_T, \sigma_T, K$), are reported in labels at the top of each figure.

3. Effects due to the removed tracks on variables of interest, as on the LF-jet MV2c10 output in the case discussed in the following, are symmetrised. The mentioned approach is necessary because new tracks can not be created at the last stage of simulation. It was tested that the fraction of jets with an MV2c10 output score exceeding a given cut depends linearly on the “tail-track” fraction.

The IP-tail correction extraction has also been studied also using alternative parameterisation of the $d_0$-distribution, different track selection, and the analysis of the composition of the non-prompt track component in simulation. The different tests show roughly consistent results, however a conservative uncertainty equal to doubling or removing the $S_F$IP-tail correction is used to cover any such variation.
6.3 Propagation of MC-Adjustments to the MV2c10 Distribution

The baseline mistag-rates for LF-jets, $\epsilon_{LF}^{MC}$, are derived analysing the MV2c10 output distribution for the various selection WPs (see Table 1), as a function of the jet $p_T-\eta$, for jets simulated using the di-jet Pythia 8 MC (see Section 3.2). The MV2c10 distribution is then re-evaluated after adjusting the simulation to account for several individual effects which might give data/MC performance differences. Each of this new evaluations is used to derive the “adjusted” LF-jet mistag-rate, $\epsilon_{LF}^{MC-adj,i}$, associated to the $i$-th effect. The analysed effects are:

1. Track IP resolution and tails corrections propagated to the reconstructed tracks using the methods described in the previous section. These effects account for a significant data/MC discrepancy and they have a sizeable impact on the MV2c10 output distribution of LF-jets.

2. Probability of reconstructing fake tracks for tight or loose track selection criteria [48]. These effects account for a significant data/MC discrepancy and they have a sizeable impact on the MV2c10 distribution of LF-jets.

3. LF-jets physics modelling associated to the strange-hadron fraction in Pythia 8 simulation. It has been found that increasing by 30% the fraction of long-lived strange-hadrons (e.g. $K_s$ and $\Lambda_s$) improves the data/MC agreement in distributions sensitive to $K_s$ and $\Lambda_s$ contributions. The correction is applied in the following together with a conservative uncertainty of $\pm 30\%$. This effect has a sizeable impact on the MV2c10 distribution.

4. The impact of the detector material in the production rate of photon conversions and secondary hadronic interactions [37]. The production rates are in agreement between data and simulation within an uncertainty of 10%, however this uncertainty has a sizeable impact on the MV2c10 distribution of LF-jets.
5. LF-jet physics modelling associated to the parton-shower. This effect has been tested by analysing the MV2c10 output distribution of LF-jets with an alternative di-jet MC generator, HERWIG++ presented in Section 3.2. The parton-shower model choice has a sizeable impact on the MV2c10 distribution of LF-jets.

6. Several other tested effects were found to have negligible impact on the MV2c10 distribution of LF-jets, these are: effect of tracking in dense environment [49], track reconstruction efficiency, and alignment weak-modes [44].

Figure 11 shows the impact of the first three effects, the ones correcting the observed data/MC discrepancies, on the MV2c10 distribution of LF-jets with $p_T > 20$ GeV and $|\eta| < 2.5$.

### 6.4 Results and Systematic Uncertainties

As discussed in Section 6.3, the following MC-adjustments have a sizable effect on the LF-jet mistag-rate and account for data/MC discrepancies: IP resolution in $d_0$ and $z_0$, fraction of tracks in the IP distribution tails, fake-track reconstruction probability, and strange-hadron fraction. Each of them induces mistag-rate
variation, $\varepsilon_{\text{MC-adj},i}$ \,(with $i$ running on all relevant adjustments), with respect to the nominal mistag-rate, $\varepsilon_{\text{MC-LF}}$, therefore, assuming uncorrelated effects, the total mistag-rate correction scale-factor can by obtained by:

$$SF_{\text{MC-adj}}^{\text{MC-LF}} = \prod_i \frac{\varepsilon_{\text{MC-adj},i}}{\varepsilon_{\text{MC-LF},i}} = \prod_i SF_{\text{LF}}^{\text{MC-adj},i} = SF_{\text{d}_0-\text{smearing}}^{\text{LF}} \cdot SF_{\zeta_0-\text{smearing}}^{\text{LF}} \cdot SF_{\text{fake-track-rate}}^{\text{LF}} \cdot SF_{\text{hadrons}}^{\text{LF}} \cdot SF_{\text{IP-tail}}^{\text{LF}},$$

(11)

allowing the evaluation of the LF-jet calibration SF for different WPs and for bins of jet $p_T$ and $\eta$.

Several systematic uncertainties are considered: MC statistics, $d_0$ and $z_0$ smearing uncertainties, $d_0-z_0$ correlations in the IP smearing, variation of the IP-tail correction fraction, uncertainty on the rate of fake-tracks (evaluated from the difference in the measurements using the tight and loose track selections), variation of the fraction of strange-hadrons, uncertainty on the rate of material interactions affecting the production of conversions and the secondary hadronic interactions with the detector material. For each of the listed systematic uncertainties, a modified version of Equation 11 is evaluated and the difference with respect to the nominal value of $SF_{\text{LF}}$ is taken as uncertainty. Parton-shower variations are also considered to take into account the uncertainty on the LF-jet hadronisation. This is done by re-evaluating Equation 11 using an alternative di-jet MC simulation based on the HERWIG++ generator (see Section 3.2) and comparing these results to the baseline calibration. The main systematic error, ranging from approximately 10% to 20% across all the selection WPs, arises from the propagation of the uncertainty of track-$d_0$ smearing; the uncertainty on the fraction of IP tails has comparable size for jets of $p_T$ below 200 GeV, while the uncertainty on the rate of fake tracks becomes of comparable size for jets of $p_T$ above 100 GeV. For the purest WP, the 60% efficiency selection, uncertainties connected to the hadronization model and to the fraction of strange-hadrons become large, at the level of 10% to 20%, for jets of $p_T$ below 100 GeV. The breakdown of the different systematic uncertainties, as a function of the LF-jet $p_T$, can be seen in Figure 12, and in Tables 8 and 11, for two example selections of the MV2c10 $b$-tagging algorithm, corresponding to the 85% and 60% WPs as listed in Table 1 (the uncertainty breakdown for the remaining WPs is reported in the Auxiliary material).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta</td>
<td>^{\text{jet}}$ range</td>
<td>cen</td>
<td>fwd</td>
<td>cen</td>
<td>fwd</td>
<td>cen</td>
<td>fwd</td>
</tr>
<tr>
<td>SF</td>
<td>1.8</td>
<td>1.7</td>
<td>2.1</td>
<td>1.9</td>
<td>2.2</td>
<td>1.9</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>Stat. error</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Syst. error</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>Total error</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Parton-shower (%)</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0.1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>$d_0$ smearing unc. (%)</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>13</td>
<td>7</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>$\zeta_0$ smearing unc. (%)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Fake-track rate (%)</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Strange-had. frac. (%)</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Material interaction (%)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>IP-tail variations (%)</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>$d_0-\zeta_0$ correlation (%)</td>
<td>3</td>
<td>0.6</td>
<td>0.1</td>
<td>0.3</td>
<td>1</td>
<td>0.2</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>Total relative err. (%)</td>
<td>16</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>19</td>
<td>16</td>
<td>20</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 8: Light-flavour jet SF measurements obtained with the adjusted-MC method for the MV2c10 85% WP and its uncertainty sources. Results are reported in the “cen” (“fwd”) calibration regions corresponding to $|\eta|^{\text{jet}} < 1.2$ (1.2 < $|\eta|^{\text{jet}} < 2.5$).
Figure 12: The blue crosses show the adjusted-MC LF-SF and statistical error, as a function of the LF-jet $p_T$ and inclusive in $\eta$, for the 85% efficiency (left) and 60% efficiency (right) WPs defined from the cuts on the MV2c10 output listed in Table 1. The various systematic sources are shown as dashed lines of different colours, the blue continuous lines show the envelop of the “up” or “down” systematic uncertainties added in quadrature.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta^{\text{Jet}}</td>
<td>$ range</td>
<td>cen</td>
<td>fwd</td>
<td>cen</td>
<td>fwd</td>
<td>cen</td>
<td>fwd</td>
</tr>
<tr>
<td>SF</td>
<td>2.0</td>
<td>2.0</td>
<td>2.6</td>
<td>2.1</td>
<td>2.5</td>
<td>2.2</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Stat. error</td>
<td>0.09</td>
<td>0.08</td>
<td>0.1</td>
<td>0.07</td>
<td>0.1</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Syst. error</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Total error</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Parton-shower (%)</td>
<td>10</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>$d_0$ smearing unc. (%)</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>10</td>
<td>13</td>
<td>6</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>$z_0$ smearing unc. (%)</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Fake-track rate (%)</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Strange-had. frac. (%)</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Material interaction (%)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>IP-tail variations (%)</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>$d_0, z_0$ correlation (%)</td>
<td>8</td>
<td>0.4</td>
<td>3</td>
<td>0.1</td>
<td>4</td>
<td>0.3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Total relative err. (%)</td>
<td>24</td>
<td>20</td>
<td>20</td>
<td>19</td>
<td>20</td>
<td>16</td>
<td>21</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 9: Light-flavour jet SF measurements obtained with the adjusted-MC method for the MV2c10 77% WP and its uncertainty sources. Results are reported in the “cen” (“fwd”) calibration regions corresponding to $|\eta^{\text{Jet}}| < 1.2$ (1.2 < $|\eta^{\text{Jet}}| < 2.5$).

All the derived LF-jet $b$–tagging calibration SFs are reported in Figure 13, together with the statistical and total uncertainties, for all the WPs listed in Table 1: 85%, 77%, 70%, and 60% efficiencies. The SFs are of about 2, reaching up to about 3 for the tighter selection WPs in the $p_T$ range between 50 and 100 GeV; uncertainties are dominated by systematic errors and range from about 15% to 50%, depending on the selection WP and $p_T$ of the jets.

The adjusted-MC is not affected by the limitation connected to large HF-contamination present for the purer WPs because, by construction, the calibration does not rely on a data-driven control-sample. This has made possible the calibration of an extremely tight WP, characterized by a 50% efficiency on $b$-jet selection and LF-jet mistag rates of about $10^{-4}$. The calibration SF for this last WP range between about
\[ p_{\text{T}}^{\text{jet}} \text{ [GeV]} \]
\[
\begin{array}{cccccccccccc}
|\eta^{\text{jet}}| \text{ range} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} \\
\hline
SF & 2.2 & 2.1 & 3.2 & 2.2 & 2.7 & 2.3 & 2.4 & 2.0 & 1.7 & 1.5 & 1.6 & 1.5 & \text{1.6} & \text{1.6} & \text{1.8} & \text{1.7} \\
Stat. error & 0.2 & 0.2 & 0.3 & 0.1 & 0.2 & 0.1 & 0.1 & 0.08 & 0.03 & 0.03 & 0.02 & 0.02 & \text{0.02} & \text{0.04} & \text{0.01} & \text{0.03} \\
Syst. error & 0.5 & 0.5 & 0.9 & 0.5 & 0.6 & 0.4 & 0.5 & 0.4 & 0.4 & 0.3 & 0.3 & 0.4 & 0.4 & \text{0.4} & \text{0.5} & \text{0.5} \\
Total error & 0.5 & 0.6 & 1.0 & 0.5 & 0.7 & 0.4 & 0.5 & 0.4 & 0.4 & 0.3 & 0.4 & 0.4 & 0.4 & \text{0.4} & \text{0.5} & \text{0.5} \\
\hline
Parton-shower (%) & 2 & 0.1 & 16 & 5 & 0.3 & 5 & 12 & 2 & 10 & 6 & 10 & 3 & 3 & 8 & 3 & 0.7 \\
d_\zeta \text{ smearing unc.} (%) & 6 & 17 & 13 & 10 & 17 & 8 & 10 & 8 & 12 & 5 & 9 & 6 & 9 & 5 & 9 & 7 \\
\zeta_0 \text{ smearing unc.} (%) & 13 & 7 & 0.6 & 4 & 4 & 2 & 5 & 2 & 7 & 6 & 5 & 5 & 4 & 4 & 3 & 4 \\
Fake-track rate (%) & 3 & 5 & 5 & 7 & 4 & 9 & 8 & 11 & 9 & 12 & 11 & 15 & 16 & 20 & 24 & 27 \\
Strange-had. frac. (%) & 8 & 4 & 9 & 5 & 7 & 7 & 7 & 7 & 6 & 3 & 4 & 3 & 4 & 2 & 2 & 2 \\
Material interaction (%) & 0.4 & 1 & 3 & 5 & 4 & 2 & 7 & 5 & 7 & 6 & 6 & 6 & 4 & 4 & 4 & 2 \\
d_{0,0} \text{ correlation} (%) & 3 & 4 & 12 & 1 & 6 & 0.1 & 3 & 3 & 11 & 6 & 5 & 6 & 1 & 2 & 3 & 2 \end{array}
\]

Table 10: Light-flavour jet SF measurements obtained with the adjusted-MC method for the MV2c10 70% WP and its uncertainty sources. Results are reported in the “cen” (“fwd”) calibration regions corresponding to |\eta^{\text{jet}}| < 1.2 (1.2 < |\eta^{\text{jet}}| < 2.5).

\[ p_{\text{T}}^{\text{jet}} \text{ [GeV]} \]
\[
\begin{array}{cccccccccccc}
|\eta^{\text{jet}}| \text{ range} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} & \text{cen} & \text{fwd} \\
\hline
SF & 2.0 & 1.8 & 3.2 & 2.8 & 3.1 & 2.3 & 2.0 & 2.1 & 1.4 & 1.3 & 1.5 & 1.4 & \text{1.5} & \text{1.4} & \text{1.7} & \text{1.7} \\
Stat. error & 0.4 & 0.3 & 0.6 & 0.4 & 0.6 & 0.3 & 0.2 & 0.2 & 0.05 & 0.04 & 0.03 & 0.04 & \text{0.04} & \text{0.08} & \text{0.02} & \text{0.08} \\
Syst. error & 1.0 & 1.0 & 0.9 & 0.9 & 0.6 & 0.6 & 0.4 & 0.4 & 0.2 & 0.2 & 0.2 & 0.3 & \text{0.3} & \text{0.3} & \text{0.5} & \text{0.5} \\
Total error & 1.0 & 1.0 & 1.1 & 1.0 & 0.9 & 0.7 & 0.5 & 0.5 & 0.2 & 0.2 & 0.2 & 0.3 & \text{0.3} & \text{0.3} & \text{0.5} & \text{0.5} \\
\hline
Parton-shower (%) & 38 & 42 & 17 & 19 & 12 & 16 & 14 & 14 & 6 & 6 & 6 & 6 & 9 & 9 & 6 & 6 \\
d_\zeta \text{ smearing unc.} (%) & 9 & 9 & 0.5 & 0.6 & 6 & 8 & 7 & 6 & 4 & 4 & 7 & 7 & 8 & 8 & 9 & 9 \\
\zeta_0 \text{ smearing unc.} (%) & 3 & 3 & 5 & 6 & 5 & 6 & 0.7 & 0.7 & 4 & 4 & 5 & 5 & 2 & 2 & 2 & 2 \\
Fake-track rate (%) & 8 & 9 & 5 & 6 & 5 & 7 & 10 & 9 & 9 & 10 & 9 & 9 & 13 & 14 & 23 & 24 \\
Strange-had. frac. (%) & 11 & 12 & 8 & 9 & 8 & 11 & 8 & 8 & 4 & 5 & 3 & 3 & 3 & 3 & 3 & 3 \\
Material interaction (%) & 0.8 & 0.9 & 2 & 3 & 4 & 6 & 8 & 7 & 6 & 7 & 7 & 7 & 6 & 7 & 5 & 5 \\
IP-tail variations (%) & 12 & 13 & 10 & 12 & 11 & 15 & 2 & 2 & 1 & 1 & 6 & 6 & 3 & 3 & 1 & 1 \\
d_{0,0} \text{ correlation} (%) & 23 & 25 & 18 & 21 & 0.6 & 0.8 & 2 & 1 & 7 & 7 & 3 & 3 & 5 & 5 & 0.9 & 1 \\
Total relative err. (%) & 53 & 57 & 34 & 37 & 28 & 31 & 23 & 23 & 16 & 17 & 17 & 17 & 20 & 22 & 26 & 28 \end{array}
\]

Table 11: Light-flavour jet SF measurements obtained with the adjusted-MC method for the MV2c10 60% WP and its uncertainty sources. Results are reported in the “cen” (“fwd”) calibration regions corresponding to |\eta^{\text{jet}}| < 1.2 (1.2 < |\eta^{\text{jet}}| < 2.5).

1 and 2, with uncertainties ranging from about 20% to 100%, as reported in Table 12.

The use of a purely simulation-based estimate also simplifies the calibration of new algorithms, for example the adjusted-MC method has been used to calibrate the LF-jet response of a version of the MV2 algorithm trained for the identification of jets originating from c-quarks (c-tagging) [50].
Table 12: LF-jet SF measurements obtained with the adjusted-MC method for the MV2c10 50% WP and its uncertainty sources. Results are reported in the “cen” (“fwd”) calibration regions corresponding to $|\eta^{\text{jet}}| < 1.2$ (1.2 < $|\eta^{\text{jet}}| < 2.5$).
Figure 13: The blue crosses show the adjusted-MC LF-SF, and statistical error, extracted as a function of the LF-jet $p_T$ for central, $|\eta| < 1.2$, (left) and forward, $1.2 < |\eta| < 2.5$, (right) regions, for different selections on the MV2c10 output, from 85% to 60% WPs (from top to bottom), as listed in Table 1. The light-blue area shows the sum in quadrature of the considered systematic uncertainty variations.
7 Comparison between the two calibrations

The LF-jet mistagging rate calibration obtained using the negative-tag method (see Section 5) and adjusted-MC method (see Section 6) are compared in Figure 14 where LF-jet scale factors obtained from the two methods are shown for the different MV2c10 selection WPs, as a function of the LF-jet $p_T$ for central, $|\eta| < 1.2$, and forward, $1.2 < |\eta| < 2.5$, regions.

The SFs obtained by the two calibration methods are in good agreement within the systematic uncertainties. Uncertainties are in general of comparable size between the two methods with the main differences arising in the loosest or tightest of the $b$-tag efficiency WPs. Uncertainties in the negative-tag method are smaller than for the adjusted-MC method when comparing the 85% $b$-efficiency WP because the negative-tag calibration relies on a large sample where the HF-jet contamination is low. Uncertainties in the adjusted-MC method are smaller than for the negative-tag method when comparing the 60% WP because the adjusted-MC calibration does not suffer from issues related HF-jet contamination in the calibration sample.
Figure 14: Comparison of LF-jet calibration SFs for the negative-tag method and adjusted-MC method as a function of the LF-jet $p_T$ for central, $|\eta| < 1.2$, (left) and forward, $1.2 < |\eta| < 2.5$, (right) regions, for different selections on the MV2c10 output score, from 85% to 60% (from top to bottom), as listed in Table 1. Results from the Negative-tag method are shown with black lines (nominal values and statistical uncertainties) and green bands (total systematic uncertainties), results from the adjusted-MC method are shown with blue lines (crosses showing total systematic uncertainties).
8 Conclusion

Two methods for the calibration of the LF-jet mistag rate of flavour-algorithms are available to the ATLAS experiment and have been illustrated in this note:

- **The negative-tag calibration:** based on the comparison of the mistag rate between simulation and data in a LF-jet enriched data set, selected by requiring the jets to pass a “negative” tag requirement.
- **The adjusted-MC calibration:** a new methodology where data-driven tracking performance studies are propagated to the simulation and then to the mistag rate.

Although the two methods rely on different assumptions and techniques, they provide consistent LF-jet calibration SFs, of about 2. Uncertainties are ranging from 10% to 50%, with the negative-tag total error being smaller than the adjusted-MC one for the loosest $b$-tag efficiency WPs and the opposite case for the tightest WPs.

References


[7] ATLAS Collaboration, *Search for heavy resonances decaying to a $W$ or $Z$ boson and a Higgs boson in the $q\bar{q}^{(r)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2017), arXiv: 1709.06958 [hep-ex].


[29] ATLAS Collaboration, 
Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1, 


[31] ATLAS Collaboration, 
Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector, 

[32] ATLAS Collaboration, Jet energy scale measurements and their systematic uncertainties in 
proton–proton collisions at \( \sqrt{s} = 13 \) TeV with the ATLAS detector, (2017), 

[33] ATLAS Collaboration, 
Secondary vertex finding for jet flavour identification with the ATLAS detector, 

[34] ATLAS Collaboration, Characterisation and mitigation of beam-induced backgrounds observed 
in the ATLAS detector during the 2011 proton–proton run, JINST 8 (2013) P07004, 

[35] G. Bohm, Introduction to statistics and data analysis for physicists, 

[36] ATLAS Collaboration, Measurement of the Inelastic Proton–Proton Cross Section at \( \sqrt{s} = 13 \) TeV with the ATLAS Detector at the LHC, Phys. Rev. Lett. 117 (2016) 182002, 

[37] ATLAS Collaboration, Study of the material of the ATLAS inner detector for Run 2 of the LHC, 

[38] ATLAS Collaboration, \( K_0^* \) and \( \Lambda \) production in pp interactions at \( \sqrt{s} = 0.9 \) and 7 TeV measured with the ATLAS detector at the LHC, Phys. Rev. D 85 (2012) 012001, 

[39] ATLAS Collaboration, b-tagging calibration plots using dileptonic \( t\bar{t} \) events produced in pp 
collisions at \( \sqrt{s} = 13 \) TeV and a combinatorial likelihood approach, FTAG-2016-003, 2016, 

[40] ATLAS Collaboration, b-tagging efficiency calibration using a tag-and-probe technique with opposite-sign, different-flavour dileptonic \( tt \) events produced in pp collisions at \( \sqrt{s} = 13 \) TeV, 
FTAG-2017-001, 2017, 


[42] G. Borisov and C. Mariotti, 
Fine tuning of track impact parameter resolution of the DELPHI detector, 

[43] ATLAS Collaboration, Performance of jet substructure techniques for large-R jets in 
proton–proton collisions at \( \sqrt{s} = 7 \) TeV using the ATLAS detector, JHEP 09 (2013) 076, 
[44] ATLAS Collaboration,  
*Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV*, 

[45] ATLAS Collaboration,  
*Tracking Studies for $b$-tagging with 7 TeV Collision Data with the ATLAS Detector*, 

[46] ATLAS Collaboration,  
*Impact Parameter Resolution Using 2016 MB Data*, IDTR-2016-018, 2016, 


[48] ATLAS Collaboration,  
*Number of tracks vs. $\mu$ with full 2016 data*, IDTR-2016-015, 2016, 

[49] ATLAS Collaboration,  

[50] ATLAS Collaboration,  
*Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment*, 
HIGG-2017-01, 2017, 
Appendix

Figure 15: Normalized IP2D (left) and IP2DNeg (right) log-likelihood ratio of the b to LF hypothesis for LF, c- and b-jets in multijet events generated with the Pythia 8 event generator. The highest $p_T$ jet in the event is shown, when this jet satisfies $p_T > 60$ GeV. Jets with no tracks selected by the IP2D (left) and IP2DNeg (right) algorithm are not shown.

Figure 16: Normalized JetFitter (left) and JetFitterFlip (right) 1-track vertex multiplicity for LF, c- and b-jets in multijet events generated with the Pythia 8 event generator. The highest $p_T$ jet in the event is shown, when this jet satisfies $p_T > 60$ GeV.
Figure 17: Impact parameter smearing values of $\sigma^{\text{cor}}(d_0)$ (left) and $\sigma^{\text{cor}}(z_0)$ (right), defined in Eq. 7, for each $p_T$ and $\eta$ value of Tight-Primary tracks. The values are extracted from di-jet events.

Figure 18: The blue crosses show the adjusted-MC LF-SF, and statistical error, extracted as a function of the LF-jet $p_T$, inclusive in $\eta$, for the 77% efficiency WP. Systematic sources are shown as dashed lines of different colours, the blue continuous lines show the “up” or “down” systematic uncertainty variations added in quadrature. The blue crosses show the adjusted-MC LF-SF vs $p_T$ for the 77% efficiency WP $b$-tag selection, for central $|\eta| < 1.2$ (left) and $1.2 < |\eta| < 2.5$ (right). The blue lines show the “up” or “down” systematic uncertainty variations added in quadrature, where each systematic source has been symmetrized.
Figure 19: The blue crosses show the adjusted-MC LF-SF, and statistical error, extracted as a function of the LF-jet $p_T$ for central, $|\eta| < 1.2$, (left) and forward, $1.2 < |\eta| < 2.5$, (right) regions, for the selection on the MV2c10 output corresponding to the 50% WP listed in Table 1. The light-blue area shows the sum in quadrature of the considered systematic uncertainty variations.