

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)



Submitted to: EPJC

CERN-EP-2022-226
30th November 2022

ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset

The ATLAS Collaboration

The flavour-tagging algorithms developed by the ATLAS Collaboration and used to analyse its dataset of $\sqrt{s} = 13$ TeV pp collisions from Run 2 of the Large Hadron Collider are presented. These new tagging algorithms are based on recurrent and deep neural networks, and their performance is evaluated in simulated collision events. These developments yield considerable improvements over previous jet-flavour identification strategies. At the 70% b -jet identification efficiency operating point, light-jet (charm-jet) rejection factors of 600 (11) are achieved in a sample of simulated Standard Model $t\bar{t}$ events; similarly, at a c -jet identification efficiency of 30%, a light-jet (b -jet) rejection factor of 70 (9) is obtained.

Contents

1	Introduction	3
2	The ATLAS detector	3
3	Monte Carlo Samples	4
4	Key ingredients for flavour-tagging	5
5	Low-level b-taggers	6
5.1	Algorithms based on impact parameters	7
5.2	Track-based recurrent neural network tagger	7
5.3	Secondary-vertex-tagging algorithm	10
5.4	Topological multi-vertex finding algorithm	11
6	High-level flavour-taggers, the DL1 series	14
7	Flavour-tagging performance	16
7.1	b -tagging performance	16
7.2	Charm-tagging performance	19
7.3	MC generator dependence	27
7.4	Overtraining checks	27
8	Conclusion	30

1 Introduction

The separation of jets containing b - and c -hadrons (b -jets and c -jets, respectively) against jets containing neither b - or c -hadrons (light-flavour jets) is of major importance in many areas of the physics programme of the ATLAS experiment [1] at the Large Hadron Collider (LHC) [2]. Flavour-tagging has been decisive in observations of the Higgs boson decay into bottom quarks [3] and of its production in association with a top-quark pair [4], and plays a crucial role in a large number of Standard Model (SM) precision measurements, studies of Higgs boson properties, and searches for new phenomena.

The ATLAS Collaboration uses various algorithms to identify b - and c -jets [5], referred to as flavour-tagging algorithms, when analysing data from pp collisions recorded during Run 2 of the LHC (2015–2018) at $\sqrt{s} = 13$ TeV. These algorithms exploit the long lifetime, high mass and high decay multiplicity of b - and c -hadrons as well as the properties of heavy-quark fragmentation. Given a lifetime of the order of 1.5 ps ($\langle c\tau \rangle \approx 450 \mu\text{m}$), energetic b -hadrons have a significant mean flight length, $\langle l \rangle = \beta\gamma c\tau$, in the detector before decaying, generally leading to at least one vertex displaced by a few mm from the hard-scatter collision point.

The strategy developed by the ATLAS Collaboration is based on a two-stage approach. Low-level algorithms reconstruct the characteristic features of the heavy-flavour jets via two complementary approaches: one that uses the properties of individual charged-particle tracks (referred to as ‘tracks’) associated with a hadronic jet, and a second which combines the tracks to explicitly reconstruct displaced vertices. Then, in order to maximise performance, the results of low-level algorithms are combined in high-level algorithms consisting of multivariate classifiers. The analysis of the data from Run 2 of the LHC is marked by improvements and retuning of the low-level algorithms [6], first introduced during Run 1, but also by the introduction of new low- and high-level algorithms respectively based on recurrent and deep neural networks. This yields considerable improvements over previous work, which was based on boosted decision trees or likelihood discriminants.

This paper is organised as follows. Section 2 introduces the ATLAS detector. The simulated Monte Carlo events used in this work are described in Section 3. Section 4 contains the description of the objects reconstructed in the detector which are key inputs to flavour-tagging algorithms, while Sections 5 and 6 describe the low- and high-level tagging algorithms respectively. Finally, their performance, evaluated on simulated event samples, is presented in Section 7.

2 The ATLAS detector

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

The ID consists of a high-granularity silicon pixel detector which covers the vertex region and typically provides four measurements per track. The innermost layer, known as the insertable B-layer (IBL) [7], was added in 2014 and provides high-resolution hits at small radius to improve the tracking performance. For a fixed b -jet efficiency, the incorporation of the IBL improves the light-flavour jet rejection of the b -tagging algorithms by up to a factor of four [8]. The silicon pixel detector is followed by a silicon microstrip tracker that typically provides eight measurements from four strip double layers. These silicon detectors are

complemented by a transition radiation tracker (TRT), which enables radially extended track reconstruction up to a pseudorapidity¹ of $|\eta| = 2.0$. The TRT also provides electron identification information based on the fraction of hits (typically 33 in the barrel and up to an average of 38 in the endcaps) above a higher energy-deposit threshold corresponding to transition radiation. The ID is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the pseudorapidity range $|\eta| < 2.5$.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) sampling calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadronic calorimetry is provided by a steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadronic endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic measurements, respectively.

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

A two-level trigger system [9] is used to select interesting events. The first level of the trigger is implemented in hardware and uses a subset of detector information to reduce the event rate to a design value of at most 100 kHz. It is followed by a software-based trigger that reduces the event rate to a maximum of around 1 kHz for offline storage.

An extensive software suite [10] 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 Monte Carlo Samples

The optimisation of the ATLAS Run 2 b -tagging algorithms is performed with jets from a ‘hybrid’ sample composed of a mixture of simulated SM $t\bar{t}$ and high-mass $Z' \rightarrow q\bar{q}$ events. The Z' events do not correspond to a single resonance but have a broad Z' mass spectrum in order to optimise the b -tagging performance at high jet momentum transverse to the beam-line (p_T). The final hybrid sample for training is obtained by mixing all b -jets from the available $t\bar{t}$ events, if the corresponding b -hadron p_T is below 250 GeV, with all jets containing a b -hadron with $p_T > 250$ GeV from the Z' sample. A similar strategy, based on the jet p_T , is applied for c -jets and light-flavour jets. No attempt is made to distinguish quark-initiated jets from gluon-initiated jets in these samples.

The $t\bar{t}$ simulation sample was produced using POWHEG BOX v2 [11–14], which yields matrix elements at next-to-leading order (NLO) in the strong coupling constant α_s for top-quark pair production. The first-gluon-emission cut-off scale parameter h_{damp} was set to $1.5m_t$, with $m_t = 172.5$ GeV used for the top-quark mass. POWHEG BOX was interfaced to PYTHIA 8.230 [15] with the A14 set of tuned parameters [16]

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

and NNPDF3.0_{NNLO} (NNPDF2.3_{LO}) [17, 18] parton distribution functions in the matrix elements (parton shower). This set-up was found to produce the best modelling, out of a number of available generator configurations, of the multiplicity of additional jets and of both the individual top-quark and $t\bar{t}$ system p_T [19, 20]. The $t\bar{t}$ events with at least one leptonic W -boson decay are considered, which ensures that a sufficiently large fraction of b -jets, c -jets, and light-flavour jets are present in the jet population.

To train and to evaluate the performance of the b -tagging algorithms at high jet p_T , a Z' sample was generated using PYTHIA 8.212 with the A14 set of tuned parameters for the underlying event and the leading-order NNPDF2.3_{LO} [18] parton distribution functions. The cross-section of the hard-scattering process was modified by applying an event-by-event weighting factor to broaden the natural width of the resonance. The branching fractions of the decays were set to be one-third each for the $b\bar{b}$, $c\bar{c}$ and light-flavour quark pairs to obtain a sample uniformly populated by jets of each flavour. This results in a fairly flat jet p_T spectrum between 250 GeV and 1.5 TeV for b -jets, c -jets, and light-flavour jets, with the falling tail of the p_T distribution populated to 3 TeV for each flavour.

The EVTGEN [21] package was used to simulate the decay of heavy-flavour hadrons. All simulated events have additional overlaid minimum-bias interactions generated with PYTHIA 8.160 with the A3 set of tuned parameters [22] and NNPDF2.3_{LO} parton distribution functions to simulate pile-up background.² These events are weighted to reproduce the observed distribution of the average number of interactions per bunch crossing in the corresponding data sample. The simulated events were processed through the full ATLAS detector simulation [23] based on GEANT4 [24]. Interactions of b -hadrons with the detector material were not simulated; c -hadrons and τ -leptons are similarly affected. This may result in sub-optimal identification algorithm performance in data for jets with very energetic particles that decay beyond the IBL. ATLAS analyses apply dedicated correction factors and corresponding uncertainties to account for the difference in performance between data and simulation.

4 Key ingredients for flavour-tagging

Jet flavour identification relies upon a number of more fundamental objects reconstructed in the ATLAS detector. The most important of these are charged-particle tracks, primary vertices (PVs), and hadronic jets ('jets'). Limited information about the hadrons contained within hadronic jets from the simulated event's record is also used during algorithm optimisation and performance studies. A short description of each of these key ingredients is given below, and more detailed information is available in Section 3 of Ref. [6].

Charged-particle tracks are reconstructed in the ID [25]. Only tracks with $p_T > 500$ MeV are used in jet flavour tagging, with further selection criteria applied to reject fake and poorly measured tracks [26]. The total tracking efficiency for charged pions with $p_T > 4$ GeV ranges from 90% for $|\eta| < 1$ to 70% in the forward region ($2.3 < |\eta| < 2.5$) of the detector. Additional selection criteria for reconstructed tracks are applied in the low-level b -tagging algorithms described in Section 5 to maintain a high efficiency for charged particles from heavy-flavour hadron decays while rejecting tracks originating from pile-up.

Primary vertex reconstruction [27, 28] on an event-by-event basis is particularly important for b -tagging since it defines the reference point from which track and vertex displacements are computed. Reconstructed PVs are constrained to be within the luminous region of the colliding LHC beams. A longitudinal vertex position resolution of about 30 μm is achieved for events with a high multiplicity of reconstructed tracks,

² Pile-up interactions correspond to additional pp collisions accompanying the hard-scatter pp interaction in proton bunch collisions at the LHC.

while the transverse resolution ranges from 10 to 12 μm , depending on the LHC running conditions that determine the beam-spot size. Flavour tagging requires at least one PV in each event, and the PV with the highest sum of squared transverse momenta of contributing tracks is selected as the primary interaction point or ‘hard-scatter vertex’. Charged-particle tracks originating from b -hadron decays often have large transverse and longitudinal impact parameters, d_0 and z_0 respectively, where d_0 is the distance of closest approach of the track to the PV in the transverse plane, and z_0 is the longitudinal separation between the PV and the point on the track where d_0 is measured, referred to below as the ‘perigee’.

Hadronic jets are built from ‘particle-flow objects’ constructed in the ATLAS calorimeters and ID [29] using the anti- k_t algorithm with radius parameter $R = 0.4$ [30] implemented in FASTJET [31]. Jet transverse momenta are further corrected to the corresponding particle-level jet p_T , based on the simulation [32]. Remaining differences between simulated events and observed data are evaluated using in situ techniques, which exploit the transverse momentum balance between a jet and a reference object such as a photon, Z boson, or multi-jet system in data; corrections are applied to simulated jets to bring them in line with data [32]. Jets with $p_T < 20$ GeV or $|\eta| \geq 2.5$ are not considered for jet flavour identification, as low- p_T jets are outside the valid calibration range and high- $|\eta|$ jets are outside the tracking fiducial volume. In order to reduce the number of jets with large energy fractions from pile-up collision vertices, the ‘jet vertex tagger’ (JVT) algorithm is used [33]. The JVT procedure builds a multivariate discriminant for each jet within $|\eta| < 2.4$ based on the ID tracks ghost-associated with the jet [34]; in particular, jets with a large fraction of high-momentum tracks from pile-up vertices are less likely to pass the JVT requirement. The JVT efficiency for jets originating from the hard pp scattering is 92% in the simulation, but the rate of pile-up jets with $p_T \geq 60$ GeV is sufficiently small that the JVT requirement is removed above this threshold. The reconstructed jet direction and transverse momentum are especially important inputs to flavour-tagging, as they will determine which charged-particle tracks should be considered for jet flavour identification.

Tracks are matched to jets by setting a maximum allowed angular separation ΔR between the track momenta, defined at the perigee, and the jet axis. Given that the decay products from higher- p_T b -hadrons are more collimated, the ΔR requirement varies as a function of jet p_T , being wider for low- p_T jets (0.45 for jet $p_T = 20$ GeV) and narrower for high- p_T jets (0.26 for jet $p_T = 150$ GeV); if more than one jet fulfils the matching criteria, the closest jet is preferred. The jet axis is also used to assign signed impact parameters to tracks, where the sign is defined to be positive if the track intersects the jet axis in the transverse plane in front of the primary vertex, and negative if the intersection lies behind the primary vertex [5].

The flavour of a jet in simulation is determined by the nature of the hadrons it contains. Jets are labelled as b -jets if at least one weakly decaying b -hadron having $p_T \geq 5$ GeV is found within a cone of size $\Delta R = 0.3$ around the jet axis. If no b -hadrons are found, c -hadrons and then τ -leptons are searched for, based on the same selection criteria. The jets matched to a c -hadron (τ -lepton) are labelled as c -jets (τ -jets). The remaining jets are labelled as light-flavour jets. For jets with more than one heavy-flavour hadron, e.g. from gluon splitting to $b\bar{b}$ or $c\bar{c}$, the procedure above is still followed, and $b\bar{b}$ ($c\bar{c}$) jets will receive a b (c) label.

5 Low-level b -taggers

This section describes the different low-level algorithms used for b -jet identification in the ATLAS Run 2 dataset. These algorithms, designed to reconstruct the characteristic features of b -jets, fall into two broad categories. The first approach, implemented in the IP2D and IP3D algorithms [35] and described in

Section 5.1, is inclusive and based on exploiting the large impact parameters of the tracks originating from the b -hadron decays. The new RNNIP [36] algorithm, developed during Run 2 and described in Section 5.2, exploits a recurrent neural network [37] to learn track impact-parameter correlations in order to further improve the jet flavour discrimination. The second approach explicitly reconstructs displaced vertices. The SV1 algorithm [38], discussed in Section 5.3, attempts to reconstruct an inclusive secondary vertex, while the JetFitter algorithm [39], presented in Section 5.4, aims to reconstruct the full b - to c -hadron decay chain.

5.1 Algorithms based on impact parameters

Two impact-parameter-based (IP-based) algorithms, IP2D and IP3D [35], are used by ATLAS. The IP2D tagger makes use of the signed transverse impact parameter significance of tracks to construct a discriminating variable, whereas IP3D uses both the signed transverse and signed longitudinal impact parameter significances³ in a two-dimensional template to account for their correlation. The signed transverse and longitudinal impact parameter significances are shown in Figure 1. Probability density functions (pdfs) obtained from reference histograms of the signed transverse and longitudinal impact parameter significances of tracks associated with b -jets, c -jets and light-flavour jets are derived from MC simulation. The pdfs are computed in exclusive categories that depend on the hit multiplicities of the tracks in the different ID layers to increase the discriminating power. In particular, the IBL hit pattern expectations and the second-innermost layer's information are fully exploited to improve the b -tagging performance in LHC Run 2. A set of template pdfs is produced using an equal mix of simulated $t\bar{t}$ and Z' events for track categories with no hits in the first two layers, which are populated by long-lived b -hadrons traversing the first layers before they decay. The $t\bar{t}$ sample is used to populate all remaining categories. The pdfs are used to calculate ratios of the b -jet, c -jet and light-flavour jet probabilities on a per-track basis. Log-likelihood ratio (LLR) discriminants are then defined as the sum of the logarithms of the per-track probability ratios for each jet-flavour hypothesis, e.g. $\sum_{i=1}^N \ln(p_b(S_i)/p_u(S_i))$ for the b -jet and light-flavour jet hypotheses, where N is the number of tracks associated with the jet and $p_b(S_i)$ ($p_u(S_i)$) is the template pdf for the b -jet (light-flavour jet) hypothesis, with S_i the impact parameter significance of the track i . The flavour probabilities of the different tracks contributing to the sum are assumed to be independent of each other. The log-likelihood ratios separating b -jets from light-flavour jets for the IP2D and IP3D b -tagging algorithms are shown in Figure 2. In addition to the LLR separating b -jets from light-flavour jets, two extra LLR discriminants are defined to separate b -jets from c -jets and c -jets from light-flavour jets, respectively. These three likelihood discriminants for both the IP2D and IP3D algorithms are used as inputs to the high-level taggers.

5.2 Track-based recurrent neural network tagger

In the case of a b -hadron decay, several charged particles can emerge from the secondary (or tertiary) decay vertex with large impact parameters. The impact parameters of these hadron-decay tracks are intrinsically correlated: if one track is found with a large impact parameter then finding a second track with a large impact parameter is more likely. If no displaced decay is present, as in light-flavour jets, then such a correlation for large impact parameter significance is not expected. The baseline IP-based b -tagging algorithms, described in Section 5.1 uses likelihood templates to compute per-flavour conditional likelihoods. Due to the large event sample needed to compute such templates and therefore the difficulty in

³ The impact parameter significance S_i of a track i is defined as the ratio of the track impact parameter to its uncertainty.

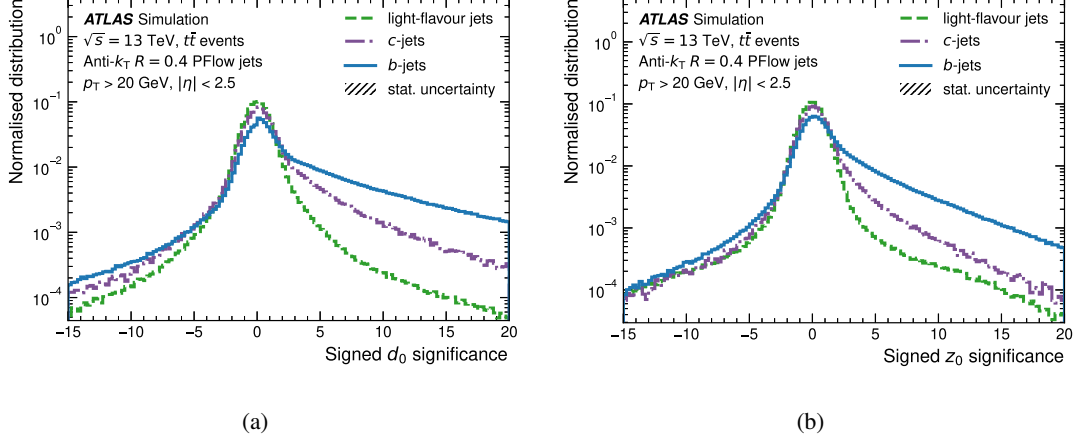


Figure 1: The (a) transverse and (b) longitudinal signed impact parameter significance of tracks for b -jets, c -jets and light-flavour jets in $t\bar{t}$ events. The first (last) bin in the distribution does not account for underflow (overflow).

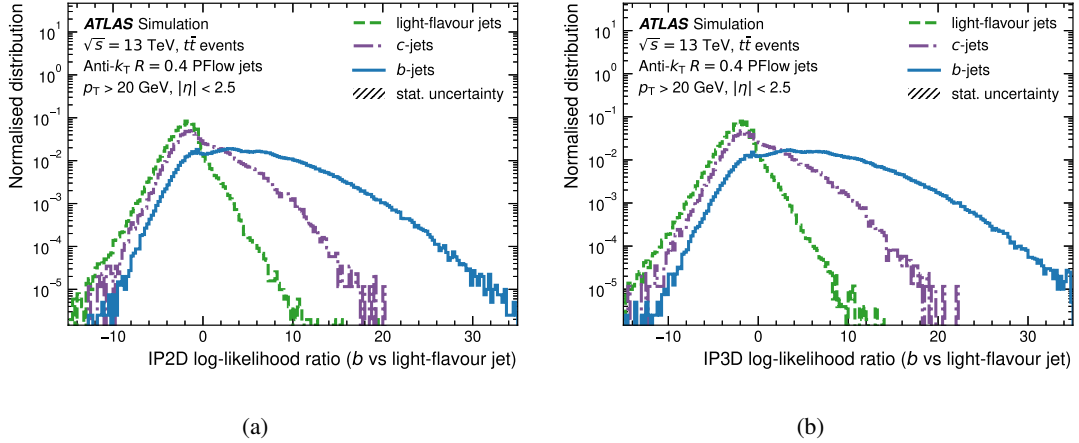


Figure 2: The log-likelihood ratio for the (a) IP2D and (b) IP3D b -tagging algorithms for b -jets, c -jets and light-flavour jets in $t\bar{t}$ events. The log-likelihood ratio shown here is computed as the ratio of the b -jet to light-flavour jet hypothesis probabilities. Jets with no tracks are not included in the plot, but assigned a large negative value. The first (last) bin in the distribution does not account for underflow (overflow).

exploiting more input variables as additional histogram axes, IP-based b -tagging algorithms assume that the properties of each track in a jet are independent of all other tracks, which limits their ability to fully model the properties of b -jets. Recurrent neural networks (RNN) can be used to overcome this challenge by directly learning sequential dependencies between the variables in sequences of arbitrary length, as is the case for the tracks in a jet. For each selected track, the lifetime-correlated signed transverse (S_{d0}) and longitudinal (S_{z0}) impact parameter significances, the fraction of transverse momentum carried by the track relative to the jet p_T (p_T^{frac}), the angular distance between the track and the jet-axis ($\Delta R(\text{track}, \text{jet-axis})$) and the hit multiplicities of the tracks in the different ID layers are fed into the network [36]. The tracks are ordered by the $|S_{d0}|$ values to form a sequence emphasising the particular importance this kinematic feature. This sequence is then passed to the neural network cells as a vector of the ordered-track features. During the training, the b -jet and c -jet p_T spectra are separately reweighted to the light-flavour jet spectrum so as to prevent the RNN from learning to discriminate directly from sample- and flavour-specific momentum distributions. The outputs provided by the network correspond to the b -jet, c -jet, and light-flavour jet probabilities. Figure 3 shows a schematic of the network architecture used for this algorithm, known as RNNIP.

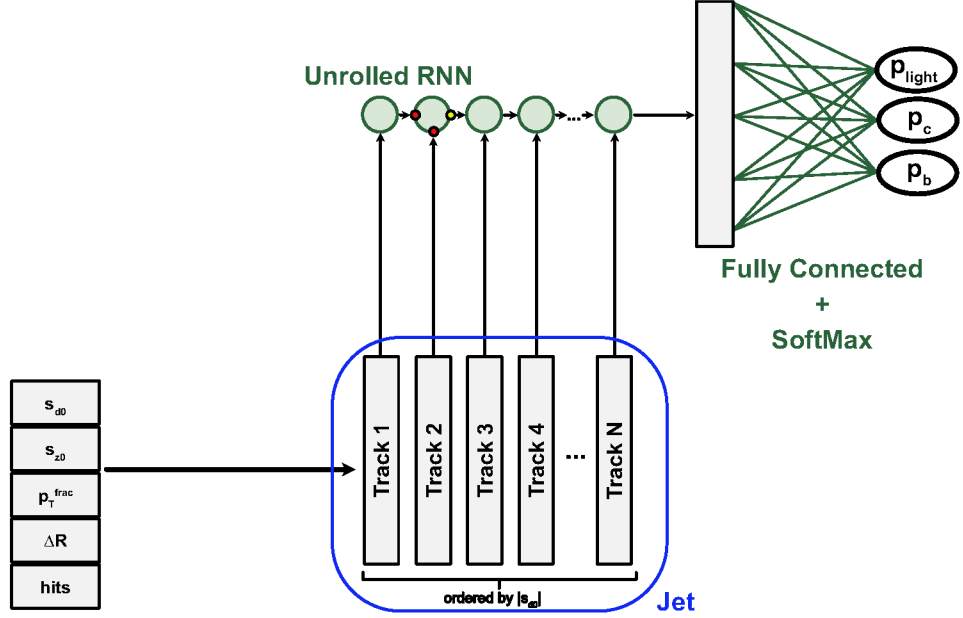


Figure 3: Schematic drawing of the RNNIP neural network architecture. The S_{d0} and S_{z0} input variables correspond to the lifetime-correlated signed transverse and longitudinal impact parameter significances, while p_T^{frac} and ΔR represent the fraction of transverse momentum carried by the track relative to the jet and the angular distance between the track and the jet axis, respectively.

The outputs of the RNN are combined into the b -tagging discriminant function as defined as:

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

where f_c is the c -jet fraction. The relative importance of c -jet and light-flavour-jet rejection can be changed by varying f_c . In this paper an optimised c -jet fraction of 0.07 is used in order to evaluate the performance of the RNNIP b -tagging algorithm in $t\bar{t}$ and $Z' \rightarrow q\bar{q}$ events. This value is chosen as a compromise to ensure good rejection factors for both c -jets and light-flavour jets in a large b -tagging efficiency range. The output

discriminant of the RNNIP b -tagging algorithms for b -jets, c -jets and light-flavour jets in the baseline $t\bar{t}$ simulated events is shown in Figure 4.

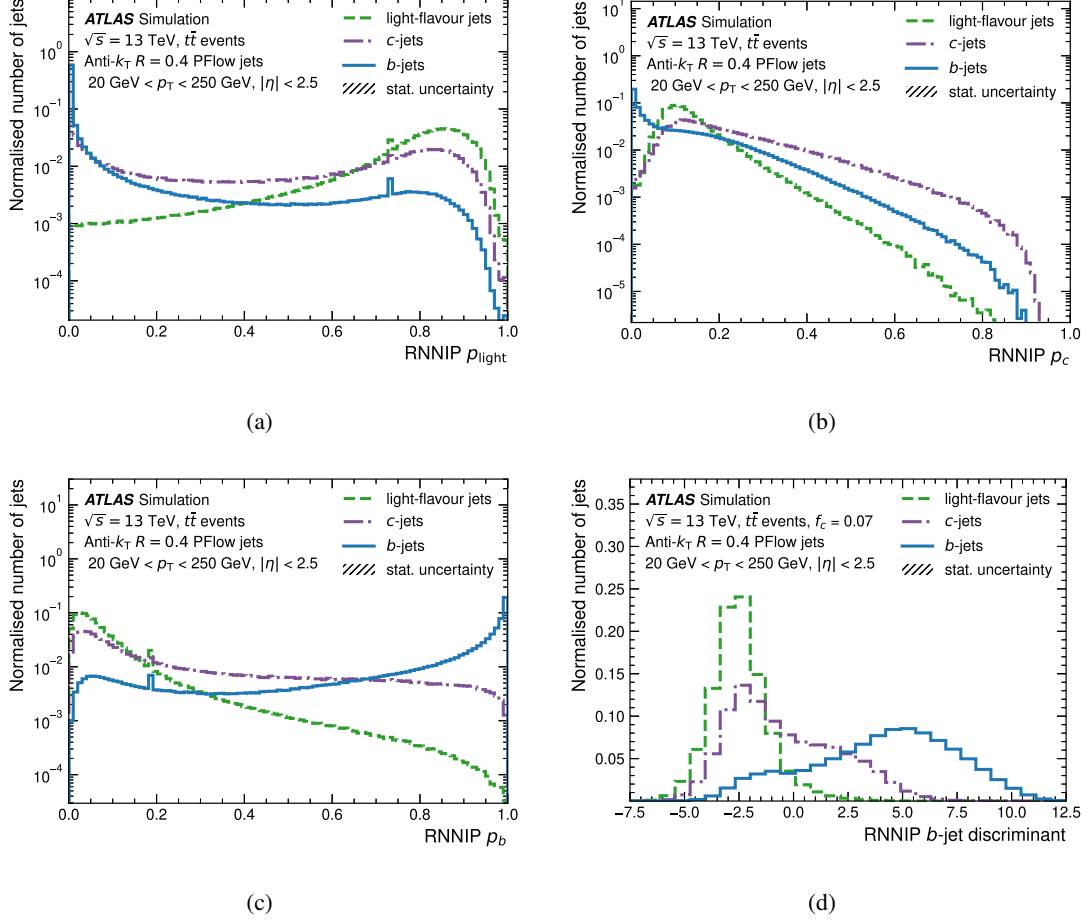


Figure 4: Distributions of the outputs of the RNNIP b -tagging algorithm for b -jets, c -jets and light-flavour jets in the baseline $t\bar{t}$ simulated events: (a) p_{light} , (b) p_c , (c) p_b , and (d) the RNNIP discriminant. The spikes at $p_u \approx 0.73$ and $p_b \approx 0.18$ originate from jets with no tracks. The location of these spikes is determined by the probability that a jet of a given flavour has no associated tracks: since light-flavour jets are more likely than b -jets to contain no tracks, $p_u > p_b$ in such cases. The corresponding spike at $p_c \approx 0.09$ is not easily visible since it occurs within a steeply falling or rising region of the p_c distribution.

5.3 Secondary-vertex-tagging algorithm

The secondary-vertex-tagging algorithm, SV1 [38], reconstructs a single displaced secondary vertex in a jet. The reconstruction starts by identifying the possible two-track vertices built with all tracks associated with the jet, while rejecting any tracks making two-track vertices compatible with K_S^0 or Λ decays, photon conversions or hadronic interactions with the detector material. The SV1 algorithm runs iteratively on all tracks contributing to the selected two-track vertices, trying to fit one secondary vertex. In each iteration, the track-to-vertex matching is evaluated using a χ^2 test. The track with the largest χ^2 is removed and the vertex fit is repeated until an acceptable vertex χ^2 and a vertex invariant mass less than 6 GeV are obtained.

With this approach, the decay products from b - and c -hadrons are typically assigned to a single common secondary vertex. The SV1 algorithm also benefits from several improvements [38] introduced during Run 2, and resulting in increased pile-up rejection and an overall enhancement of the performance at high jet p_T . Among the various algorithm improvements, additional track-cleaning requirements are applied to jets in the high-pseudorapidity region ($|\eta| \geq 1.5$). The fake-vertex rate is also better controlled by limiting the algorithm to only consider the 25 highest- p_T tracks in the jets. Finally, eight discriminating variables, including the jet p_T and η , the number of tracks associated with the SV1 vertex, the invariant mass of the secondary vertex, its energy fraction (defined as the total energy of all the tracks associated with the secondary vertex divided by the energy of all the tracks associated with the jet), and the three-dimensional decay length significance, are used as inputs to the high-level taggers. Six of these variables are illustrated in Figure 5.

In order to quantify the flavour-tagging performance of the SV1 algorithm, a simple feed-forward neural network was trained by exclusively using the outputs of the algorithm and the p_T and η of the input jets. The network was trained in the same way as the main high-level tagger and on identical training samples, as explained in Section 6. This new low-level tagger, defined only to illustrate the performance of SV1 relative to other algorithms described in this paper, is referred to as SVKine.

5.4 Topological multi-vertex finding algorithm

The topological multi-vertex finding algorithm, **JetFitter** [39], 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 modified Kalman filter [40] is used to find a common line on which the primary, b - and c -hadron decay vertices lie, approximating the b -hadron flight path as well as the vertex positions. With this approach, it is possible to resolve the b - or c -hadron decay vertex even if only a single track is attached to it. The **JetFitter** algorithm also benefits from several improvements [39] introduced during Run 2. These include a reoptimisation of the track selection to better mitigate the effect of pile-up tracks, an improvement in the rejection of interactions with detector material, and the introduction of a vertex-mass-dependent selection during the decay chain fit to increase the efficiency for tertiary vertex reconstruction. Eight discriminating variables, including the track multiplicity at the **JetFitter** displaced vertices, the invariant mass of tracks associated with these vertices, their energy fraction and their average three-dimensional decay length significance, are used as inputs to the high-level taggers. Six of these variables are illustrated in Figure 6. Finally, the discrimination of c -jets from b -jets and light-flavour jets is further improved by more specifically exploiting jets for which only a single secondary vertex is reconstructed with intermediate charged decay multiplicity and comparable decay distance to b -hadrons in jets. A set of nine additional variables [35], among which the number, the invariant mass and the energy of the tracks associated with the secondary vertex as well as their rapidity, computed with respect to the jet axis and the vector defined between the primary and secondary vertices, are used as inputs to the high-level b -tagging algorithms.

Similarly to the SV1 algorithm, the flavour-tagging performance of **JetFitter** is assessed from a simple feed-forward neural network trained by exclusively using the outputs of **JetFitter** and the input jets' p_T and η . This training was performed in the same way as for the main high-level tagger and on identical training samples, as explained in Section 6. This new low-level tagger, defined only to illustrate the performance of **JetFitter** relative to other algorithms described in this paper, is referred to as JFKine.

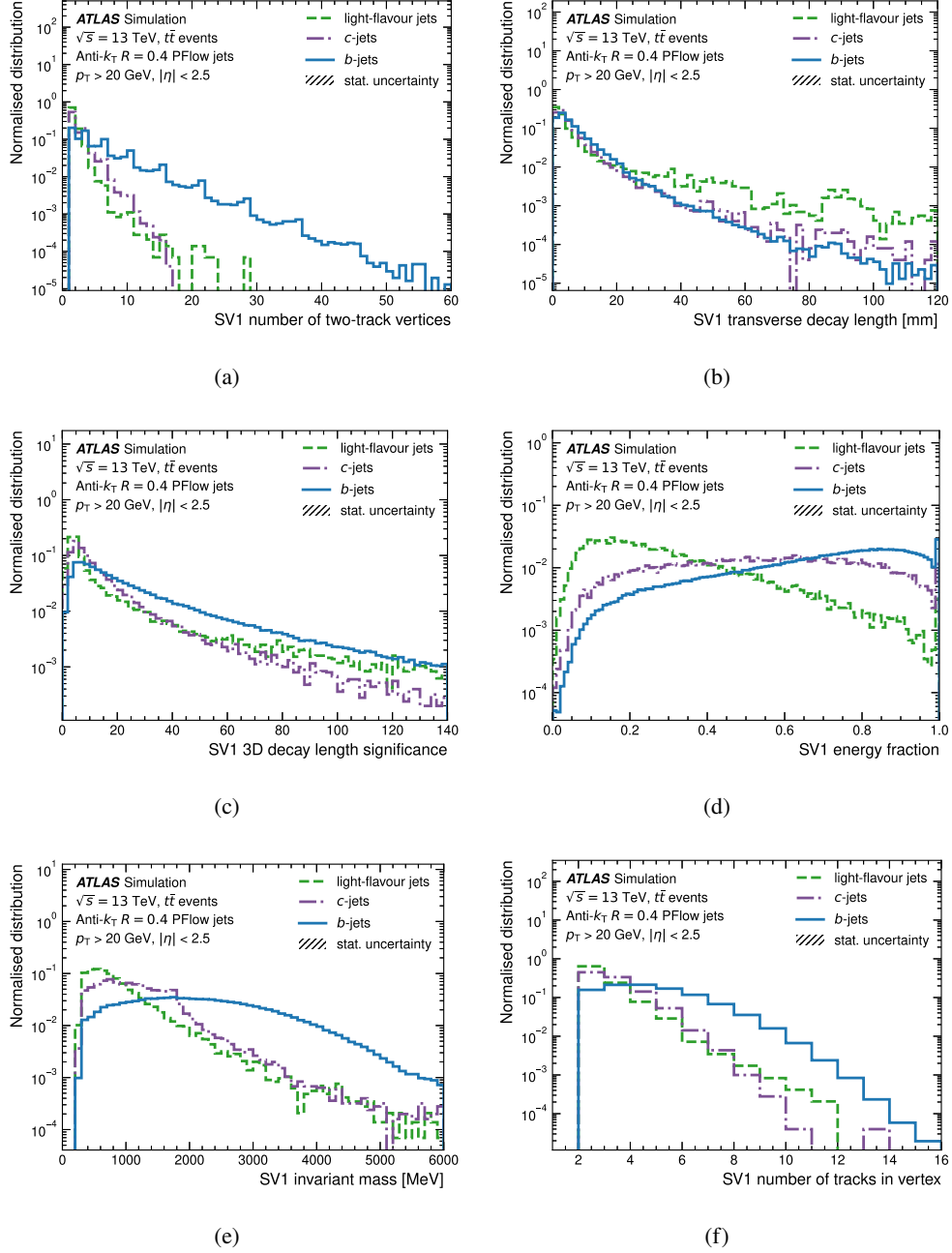


Figure 5: Properties of secondary vertices reconstructed by the SV1 algorithm for b -jets, c -jets and light-flavour jets in the baseline $t\bar{t}$ simulated events: (a) the number of two-track vertices reconstructed within the jet, (b) the transverse decay length, (c) the 3D decay length significance defined as the significance of the distance between the primary vertex and displaced vertex, (d) the energy fraction, defined as the energy of the tracks in the displaced vertex relative to the energy of all tracks reconstructed within the jet, (e) the invariant mass and (f) the number of tracks associated with the vertex. The increased rate of light-flavour jets at high transverse decay length values is due to residual interactions with detector material. The jumps in the frequency of reconstructed two-track vertices (a) originates from combinatorics. Expecting $(N \cdot (N - 1))/2$ possible track pairs created from a set of N tracks, this number is reduced due to track selection criteria, resulting in low-side tails on each spike. The first (last) bin in the distribution does not account for underflow (overflow).

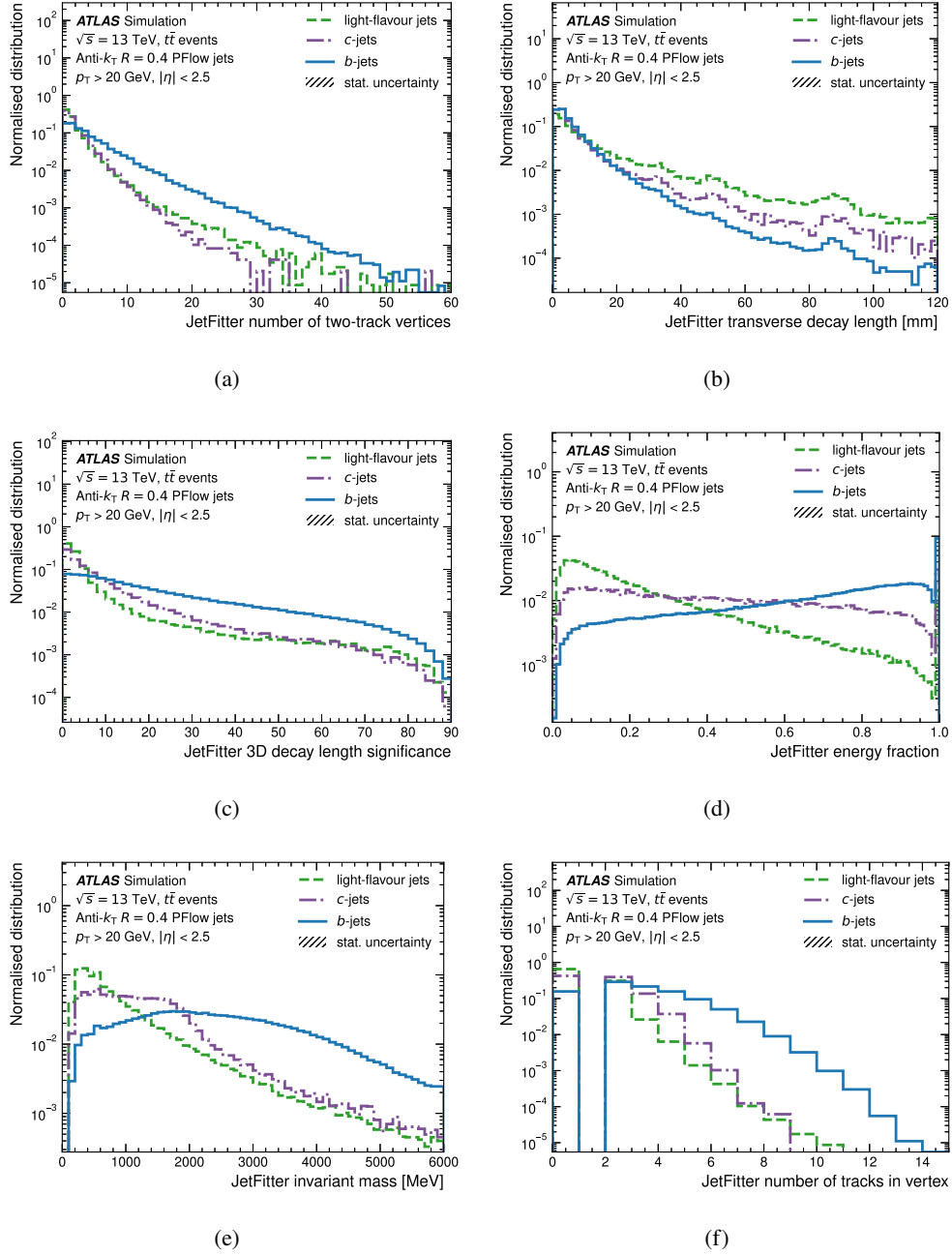


Figure 6: Properties of secondary vertices reconstructed by the JetFitter algorithm for b -jets, c -jets and light-flavour jets in the baseline $t\bar{t}$ simulated events: (a) the number of two-track vertices reconstructed within the jet, (b) the transverse decay length, (c) the average 3D decay length significance, defined as the significance of the average distance between the primary vertex and displaced vertices, (d) the energy fraction, defined as the energy of the tracks in the displaced vertex relative to the energy of all tracks reconstructed within the jet, (e) the invariant mass and (f) the number of tracks associated with the vertex. The increased rate of light-flavour jets at high transverse decay length values is due to residual interactions with detector material. Expecting jumps in the frequency of reconstructed two-track vertices originating from combinatorics as in the case of SV1, the $(N \cdot (N - 1))/2$ possible track pairs created from a set of N tracks is reduced due to track selection criteria, resulting in a smooth distribution. The first (last) bin in the distribution does not account for underflow (overflow).

6 High-level flavour-taggers, the DL1 series

To maximise the flavour-tagging performance for Run 2, the output quantities of the low-level algorithms are combined using deep-learning classifiers, based on fully connected multi-layer feed-forward neural networks (NN) [41], forming the so-called DL1 algorithm series.

These algorithms are trained with a hybrid training sample, for which 70% of the jets in the sample are from $t\bar{t}$ events and the remaining 30% are from $Z' \rightarrow q\bar{q}$ events, using TensorFlow [42] with the Keras [42] front-end and the Adam optimiser [43]. The DL1 algorithm, introduced at the beginning of Run 2 in Ref. [6], exploits as input the IP2D, IP3D, SV1 and JetFitter algorithm outputs, while the DL1r algorithm also includes the jet RNNIP output probabilities.

In addition, the kinematic properties of the jets, namely p_T and $|\eta|$, are included in the training in order to take advantage of the correlations with the other input variables. However, to avoid differences between the kinematic distributions of signal and background jets being used to discriminate between the different jet flavours, the input training dataset is resampled. The resampling procedure ensures that jets in the training sample are uniformly distributed in jet p_T and η for each flavour class. No kinematic resampling is applied at the evaluation stage of the algorithms. Table 1 presents a detailed list of input variables used by each algorithm.

The DL1r NN has a multidimensional output corresponding to the probabilities for a jet to be a b -jet, a c -jet or a light-flavour jet. The use of a multi-class network architecture provides the algorithm with a smaller memory footprint than the previous ATLAS MV2c10 algorithm [6] based on boosted decision trees (BDTs). The topology of the network consists of a mixture of fully connected hidden layers. The DL1r algorithm parameters, listed in Table 2, include the architecture of the NN, the number of training epochs, the learning rates and training batch size. Each of these parameters is optimised in order to maximise the b -tagging performance. Batch normalisation [44] is added by default since it is found to improve the performance.

Training with multiple output nodes offers additional flexibility when constructing the final output discriminant by combining the b -jet, c -jet and light-flavour jet probabilities. Since all flavours are treated equally during training, the trained network can be used for both b -jet and c -jet tagging. The final DL1r b -tagging discriminant is defined as:

$$D_{\text{DL1r}} = \ln \left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}} \right),$$

where p_b , p_c , p_{light} and f_c represent respectively the b -jet, c -jet and light-flavour jet probabilities, and the effective c -jet fraction in the background hypothesis. Using this approach, the c -jet fraction in the background can be chosen a posteriori in order to optimise the performance of the algorithm at physics analysis level. In this paper, an optimised c -jet fraction of 0.018 is used to evaluate the performance of the DL1r b -tagging algorithm in simulated $t\bar{t}$ and $Z' \rightarrow q\bar{q}$ events. This value is chosen as a compromise to ensure good rejection factors for both c -jet and light-flavour jets in a large b -tagging efficiency range across a number of analyses. In particular, the ATLAS measurements of VH , $H \rightarrow b\bar{b}$ production [45] and $t\bar{t}H$, $H \rightarrow b\bar{b}$ production [46] were considered in this optimisation.

Similarly, the DL1r c -tagging discriminant is defined as:

$$D_{\text{DL1r}}^c = \ln \left(\frac{p_c}{f_b \cdot p_b + (1 - f_b) \cdot p_{\text{light}}} \right), \quad (1)$$

Table 1: Input variables used by the SVKine and JFKine, DL1 and DL1r algorithms.

Input	Variable	Description	SVKine	JFKine	DL1	DL1r
Kinematics	p_T	Jet p_T	✓	✓	✓	✓
	η	Jet $ \eta $	✓	✓	✓	✓
IP2D, IP3D	$\log(P_b/P_{\text{light}})$	Likelihood ratio of the b -jet to light-flavour jet hypotheses			✓	✓
	$\log(P_b/P_c)$	Likelihood ratio of the b -jet to c -jet hypotheses			✓	✓
	$\log(P_c/P_{\text{light}})$	Likelihood ratio of the c -jet to light-flavour jet hypotheses			✓	✓
RNNIP	P_b	b -jet probability				✓
	P_c	c -jet probability				✓
	P_{light}	light-flavour jet probability				✓
SV1	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass	✓		✓	✓
	$f_E(\text{SV})$	Jet energy fraction of the tracks associated with the secondary vertex	✓		✓	✓
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex	✓		✓	✓
	$N_{2\text{TrkVtx}}(\text{SV})$	Number of two-track vertex candidates	✓		✓	✓
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertices	✓		✓	✓
	$L_{xyz}(\text{SV})$	Distance between the primary and secondary vertices	✓		✓	✓
	$S_{xyz}(\text{SV})$	Distance between the primary and secondary vertices divided by its uncertainty	✓		✓	✓
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.	✓		✓	✓
JetFitter	$m(\text{JF})$	Invariant mass of tracks from displaced vertices		✓	✓	✓
	$f_E(\text{JF})$	Jet energy fraction of the tracks associated with the displaced vertices		✓	✓	✓
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices		✓	✓	✓
	$S_{xyz}(\text{JF})$	Significance of the average distance between PV and displaced vertices		✓	✓	✓
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks from multi-prong displaced vertices		✓	✓	✓
	$N_{2\text{TrkVtx}}(\text{JF})$	Number of two-track vertex candidates (prior to decay chain fit)		✓	✓	✓
	$N_{1\text{-trk vertices}}(\text{JF})$	Number of single-prong displaced vertices		✓	✓	✓
	$N_{\geq 2\text{-trk vertices}}(\text{JF})$	Number of multi-prong displaced vertices		✓	✓	✓
	$L_{xyz}(2^{\text{nd}})(\text{JF})$	Distance of 2^{nd} vertex from PV		✓	✓	✓
	$L_{xy}(2^{\text{nd}})(\text{JF})$	Transverse displacement of the 2^{nd} vertex		✓	✓	✓
	$m_{\text{Tk}}(2^{\text{nd}})(\text{JF})$	Invariant mass of tracks associated with the 2^{nd} vertex		✓	✓	✓
	$E(2^{\text{nd}})(\text{JF})$	Energy of the tracks associated with the 2^{nd} vertex		✓	✓	✓
	$f_E(2^{\text{nd}})(\text{JF})$	Jet energy fraction of the tracks associated with the 2^{nd} vertex		✓	✓	✓
	$N_{\text{TrkAtVtx}}(2^{\text{nd}})(\text{JF})$	Number of tracks associated with the 2^{nd} vertex		✓	✓	✓
	$\eta_{\text{trk}}^{\text{min,max,avg}}(2^{\text{nd}})(\text{JF})$	Min., max. and avg. pseudorapidity of tracks at the 2^{nd} vertex		✓	✓	✓

Table 2: List of optimised hyperparameters used in the DL1r flavour-tagging algorithms.

Hyperparameter	Value
Number of input variables	31
Number of hidden layers	8
Number of nodes [per layer]	[256, 128, 60, 48, 36, 24, 12, 6]
Learning rate	0.01
Training batch size	15 000
Activation function	ReLU
Number of training epochs	200
Free (trainable) parameters	59 275
Fixed parameters	1 140
Training sample size	22 M jets

where f_b represents the effective b -jet fraction in the background training sample. A b -jet fraction of 0.2 is used to evaluate the performance of the DL1r c -tagging algorithm in this paper in simulated $t\bar{t}$ and $Z' \rightarrow q\bar{q}$ events. Larger than the f_c fraction presented above, the f_b fraction is chosen here to maximise the b -jet rejection factor at the given c -tagging efficiency rates of 20% and 30%. The output probabilities of the DL1r b -tagging algorithms for b -jets, c -jets and light-flavour jets in the baseline $t\bar{t}$ simulated events are shown in Figures 7(a)–(c); the corresponding b -tagging and c -tagging discriminants are also shown in Figures 7(d) and 7(e).

7 Flavour-tagging performance

The performance of a flavour-tagging algorithm is characterised by the probability or efficiency of tagging a signal jet, ε , and the probability of mistakenly identifying a background jet, referred to as the mis-tag rate. In this paper, the performance of the algorithms is quantified in terms of background-jet rejection factors, defined as $1/\varepsilon$ for background jets.

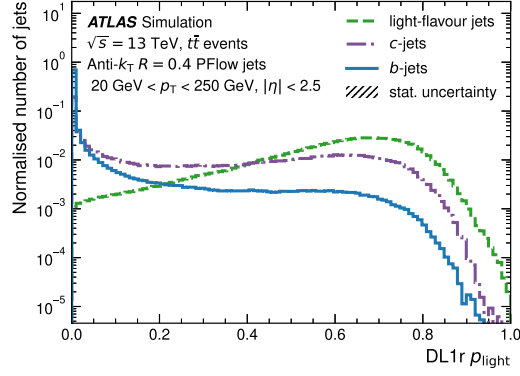
7.1 b -tagging performance

When analysis of LHC data requires the identification of b -jets, the tagging efficiency of a given requirement on the b -tagging discriminant is denoted by ε_b , and the charm-jet and light-flavour jet rejection factors are $1/\varepsilon_c$ and $1/\varepsilon_l$, respectively. Many ATLAS analyses of Run 2 LHC data use a requirement on the DL1r discriminant, or bins built from several such requirements, that do not vary with jet kinematics. These are referred to as ‘fixed-cut operating points’, and they are labelled according to their inclusive efficiency for the population of b -jets present in the $t\bar{t}$ sample used to train DL1r; for example, the DL1r discriminant value for which 77% of the b -jets in a $t\bar{t}$ sample have a higher score is called the ‘77% operating point’.

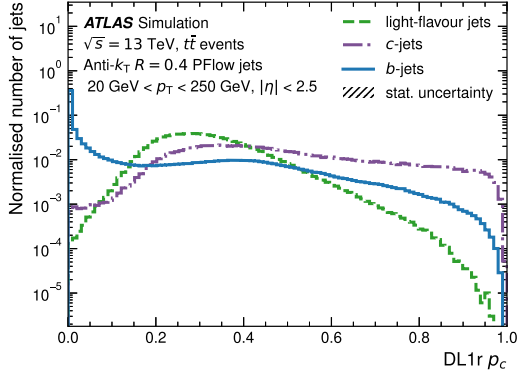
Figures 8 and 9 show the light-flavour jet and c -jet rejection factors as a function of ε_b for a variety of low- and high-level b -taggers. At high-efficiency operating points, RNNIP provides the best rejection among the low-level taggers, while at low efficiency the secondary-vertex finders SVKine and JFKine achieve the highest background rejection. DL1r substantially outperforms all low-level taggers across the ε_b range: the low-level algorithms exploit different jet properties, and combining these produces large gains. DL1r also exceeds the performance of the BDT-based MV2c10 tagger and the DL1 tagger [6].

It is also important to gauge the b -tagging performance across a broad p_T range, because the ATLAS physics programme relies on excellent background rejection in a variety of situations, depending on the needs of a given analysis of the LHC data. In Figure 10, the background rejection achieved at a fixed b -tagging efficiency of 77% is shown in bins of jet p_T ; this fixed-efficiency is obtained by choosing the appropriate b -tagging discriminant requirement in each p_T bin.

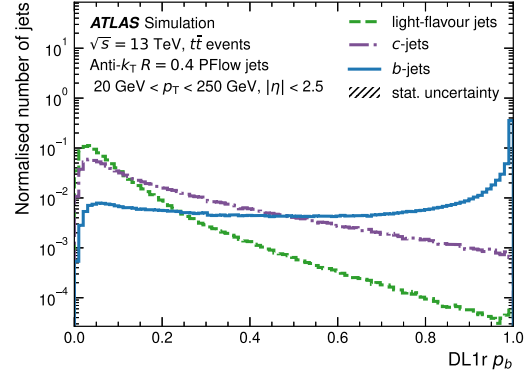
Figures 11 and 12 show the ε_b values and background rejection factors for jets from simulated SM $t\bar{t}$ and flattened Z' samples, respectively, in bins of jet p_T ; several high-level b -taggers are compared at the 77% operating point. DL1r performs significantly better than previous ATLAS b -taggers across a broad range of jet p_T , although some common patterns are worth noting: (1) the b -tagging efficiency is highest at around $p_T \approx 200$ GeV and falls with p_T above this point, and (2) the light-flavour jet rejection falls until about 1 TeV, above which it is approximately constant. However, while MV2c10 maintains a nearly flat c -jet rejection versus jet p_T , the DL1 and DL1r rejection factors improve with p_T . The enhanced performance for highly energetic jets has yielded substantially stronger tests of the Standard Model with the ATLAS



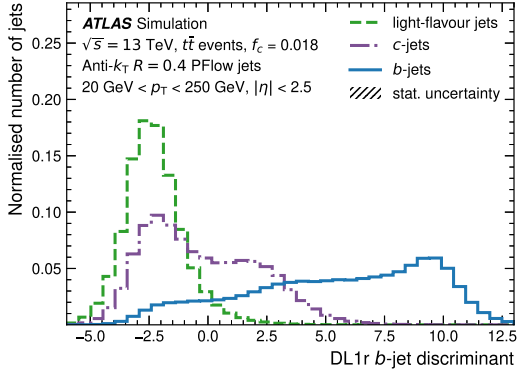
(a)



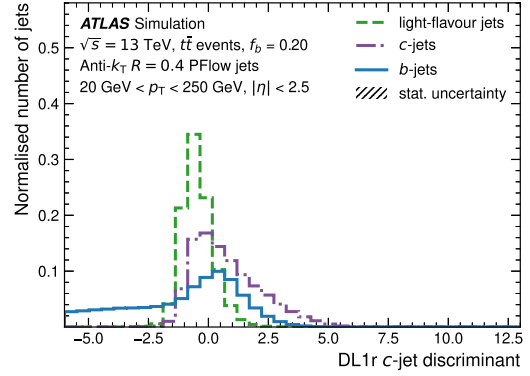
(b)



(c)



(d)



(e)

Figure 7: Distributions of the outputs of the DL1r b -tagging algorithm for b -jets, c -jets and light-flavour jets in $t\bar{t}$ simulated events: (a) p_{light} , (b) p_c , (c) p_b , (d) D_{DL1r} , and (e) D_{DL1r}^c .

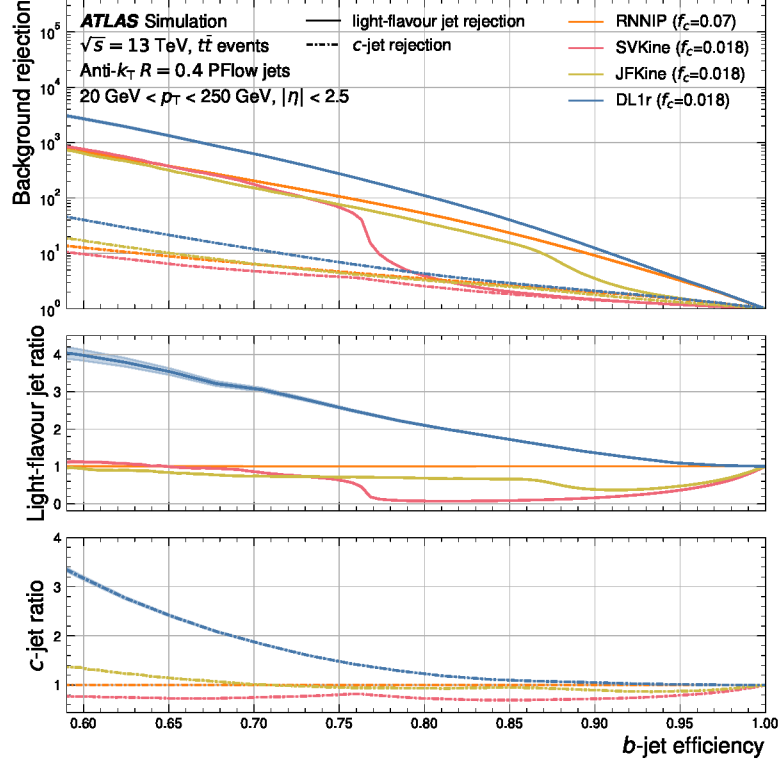


Figure 8: The light-flavour jet and c -jet rejection factors as a function of ε_b for DL1r as well as the low-level b -taggers RNNIP, SVKine, and JFKine. The statistical uncertainties of the rejection factors are calculated using binomial uncertainties and are indicated as coloured bands. The lower two panels show the ratio of the light-flavour jet rejection and the c -jet rejection of the algorithms to RNNIP. SV1 and JetFitter have secondary-vertex-finding efficiencies of approximately 80% and 90%, respectively; this causes the rapid growth in light-flavour jet rejection around these values of ε_b for SVKine and JFKine.

data. For example, recent searches for new resonances decaying into $b\bar{b}$ pairs using the DL1r b -tagger achieved about a factor of 3 stronger limits on new narrow resonances decaying into $b\bar{b}$ than predicted via luminosity-scaling of previous results using MV2c10 [47].

Similarly, the b -tagging efficiencies and background-jet rejection factors vary with the jet pseudorapidity η , in large part due to the poorer track d_0 and z_0 resolutions at high $|\eta|$ [48]. Figure 13 shows ε_b and the background-jet rejection as a function of jet η . The b -tagging efficiency is higher for all compared high-level taggers in the central region than at high $|\eta|$; the light-flavour jet rejection performance also deteriorates for more forward jets. However, DL1r consistently outperforms DL1 and MV2c10 across jet η ranges.

The ATLAS flavour-tagging algorithms are stable versus the number of pile-up interactions accompanying the hard-scatter collisions, as shown in Figure 14. While there is a small slope in the b -tagging efficiency versus the average number of interactions per bunch crossing $\langle\mu\rangle$, ε_b only changes by about 2% over the range $10 < \langle\mu\rangle < 70$ for the inclusive 77% efficiency operating point. The light-flavour jet and c -jet rejection factors also show little dependence on $\langle\mu\rangle$.

The 60%, 70%, 77%, and 85% operating points are commonly used in ATLAS physics analyses interpreting the LHC Run 2 pp dataset [6]. Figure 15 shows the b -tagging efficiency and background-jet rejection

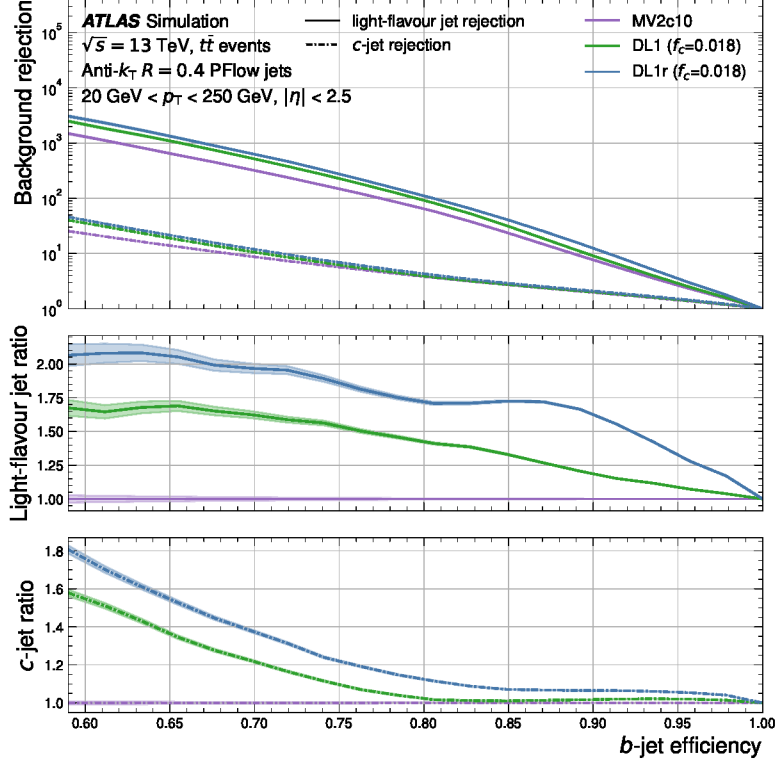


Figure 9: The light-flavour jet and c -jet rejection factors as a function of ε_b for the high-level b -taggers MV2c10, DL1, and DL1r. The lower two panels show the ratio of the light-flavour jet rejection and the c -jet rejection of the algorithms to MV2c10. The statistical uncertainties of the rejection factors are calculated using binomial uncertainties and are indicated as coloured bands.

factors versus jet p_T for DL1r at these commonly used operating points. It is worth noting that relatively small changes in b -tagging efficiency operating point of the order of 10% result in very different background rejection factors, which range from ~ 10 to $\sim 10^3$ for light-flavour jets and from ~ 3 to ~ 40 for charm jets.

7.2 Charm-tagging performance

A growing number of ATLAS analyses of LHC data require identification of c -jets in order to probe physics processes involving final-state charm quarks. For example, the search for Higgs boson decays into charm quarks aims to obtain direct evidence of the charm-quark Yukawa coupling parameter by examining a sample of events with two c -tagged jets [49]. An algorithm's performance is indicated by its b -jet and light-flavour jet rejection factors, $1/\varepsilon_b$ and $1/\varepsilon_l$, at a given charm-tagging efficiency, ε_c . Due to charmed hadrons having shorter lifetimes and smaller masses than b -hadrons, the identification of charm jets is challenging, and the efficiency of the optimal operating point tends to be much lower than for b -tagging.

Figure 16 presents the b -jet and light-flavour jet rejection factors as a function of the c -tagging efficiency, evaluated in a population of jets taken from a sample of simulated $t\bar{t}$ events. Figure 17 shows the light-flavour jet and b -jet rejection factors attained by the DL1 and DL1r algorithms for a fixed charm-tagging efficiency, again evaluated in jets from a $t\bar{t}$ sample. These ‘iso-efficiency’ curves are obtained by varying the parameter f_b used to define the c -tagging discriminant introduced in Eq. (1).

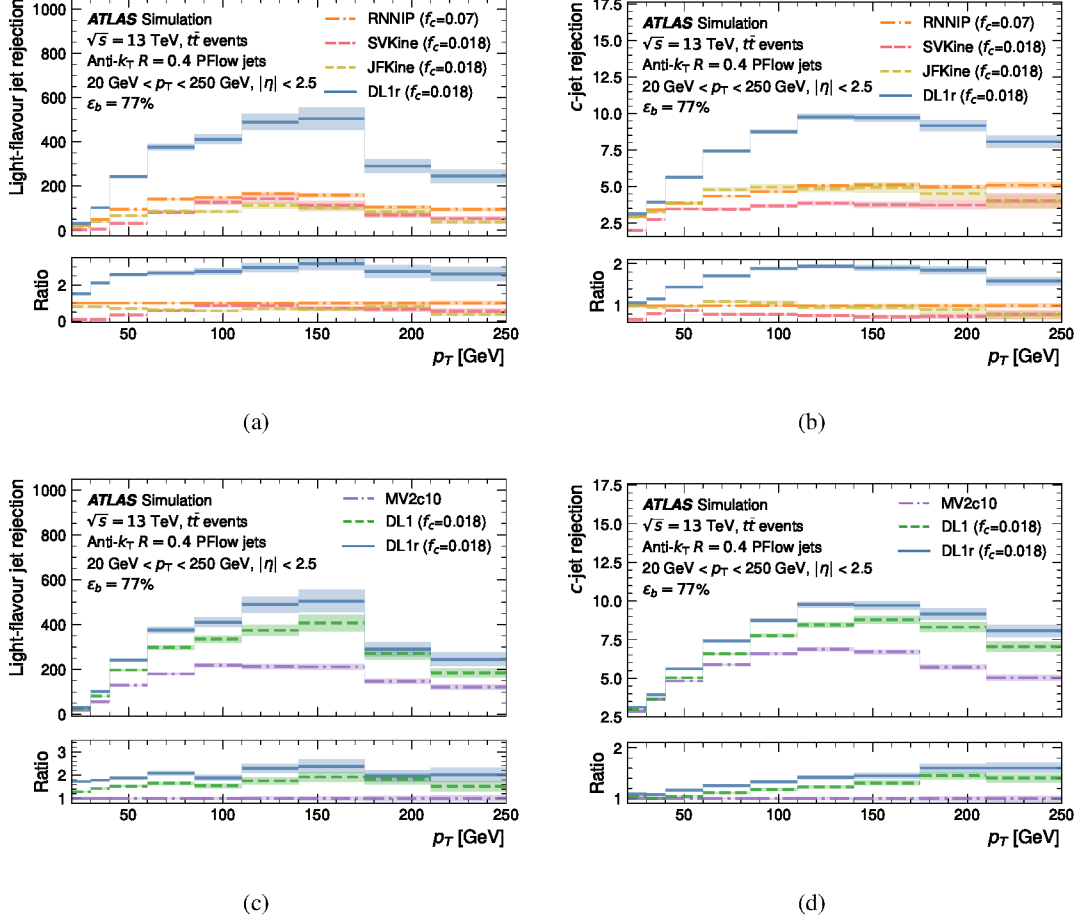
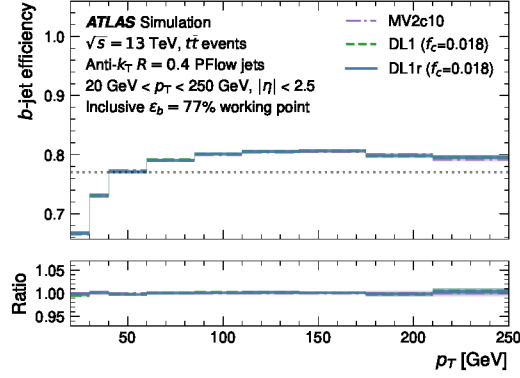
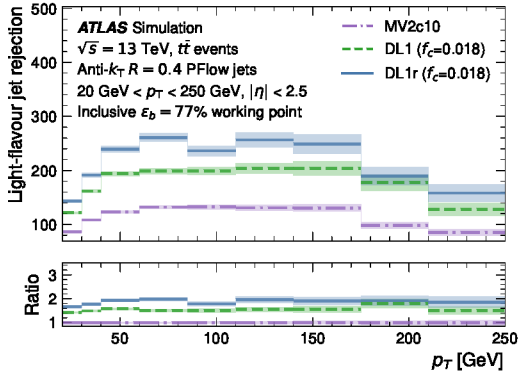


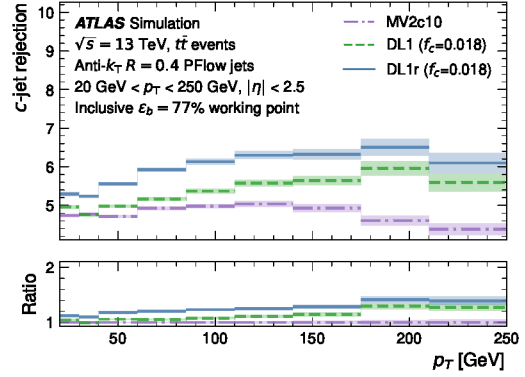
Figure 10: The light-flavour jet and c -jet rejection factors at a fixed b -tagging efficiency ε_b as a function of jet p_T . In each bin, the b -tagging efficiency is set to 77%, and the resulting background rejection is shown. (a) shows the light-flavour jet rejection for several low-level tagging algorithms and DL1r, (b) shows the c -jet rejection for the same taggers. (c) and (d) show light-flavour jet and c -jet rejection, respectively, for the DL1r, DL1, and MV2c10 high-level taggers. The lower panels of (a) and (b) show the ratio of each algorithm's background rejection to that of RNNIP; the lower panels of (c) and (d) show the corresponding ratios to MV2c10. The statistical uncertainties of the efficiencies and rejection factors are calculated using binomial uncertainties and are indicated as coloured bands. The drop in light-flavour jet rejection at $p_T \sim 175 \text{ GeV} \sim m_{\text{top}}$ is due to the overlap of decay products from boosted top-quark decays.



(a)

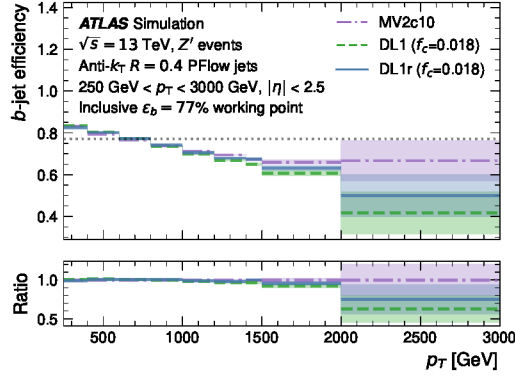


(b)

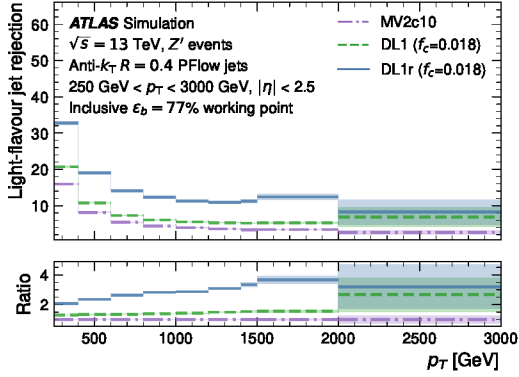


(c)

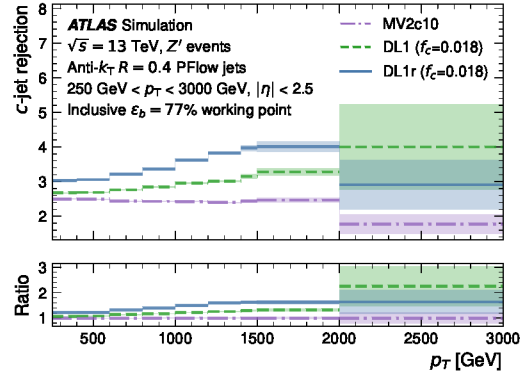
Figure 11: The b -tagging efficiency, ϵ_b , and background-jet rejection factors for several high-level b -taggers, including DL1r, as a function of jet p_T , evaluated using simulated $t\bar{t}$ events. (a) shows ϵ_b , (b) shows the light-flavour jet rejection, and (c) shows the c -jet rejection, all for a range of jet p_T bins at the inclusive 77% efficiency operating point, commonly used in ATLAS analyses of LHC Run 2 pp collision data. The lower panels show the ratio of each algorithm's performance to that of MV2c10. The statistical uncertainties of the efficiencies and rejection factors are calculated using binomial uncertainties and are indicated as coloured bands.



(a)

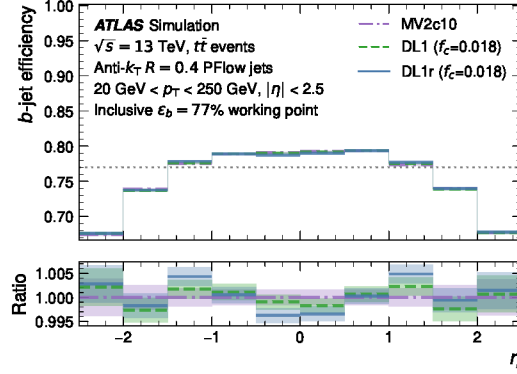


(b)

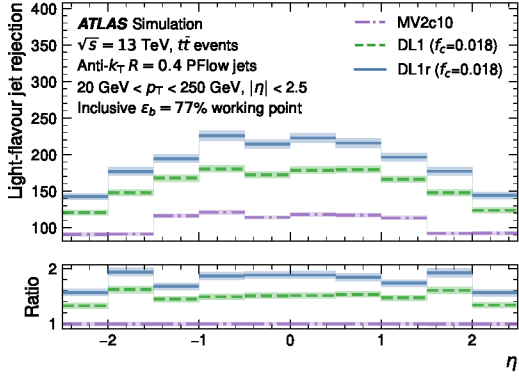


(c)

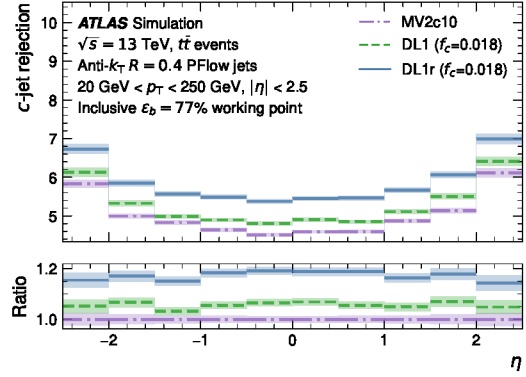
Figure 12: The b -tagging efficiency, ε_b , and background-jet rejection factors for several high-level b -taggers, including DL1r, as a function of jet p_T , evaluated using simulated Z' events, which mostly provide high- p_T jets. (a) shows ε_b , (b) shows the light-flavour jet rejection, and (c) shows the c -jet rejection, all for a wide range of jet p_T bins at the inclusive 77% efficiency operating point, commonly used in ATLAS analyses of LHC Run 2 pp collision data. The lower panels show the ratio of each algorithm's performance to that of MV2c10. The statistical uncertainties of the efficiencies and rejection factors are calculated using binomial uncertainties and are indicated as coloured bands.



(a)

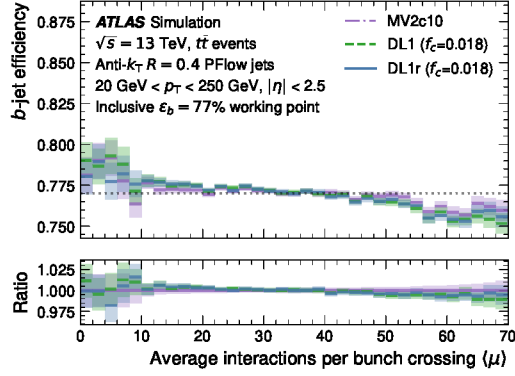


(b)

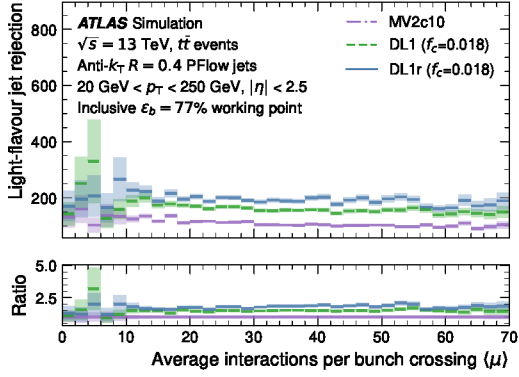


(c)

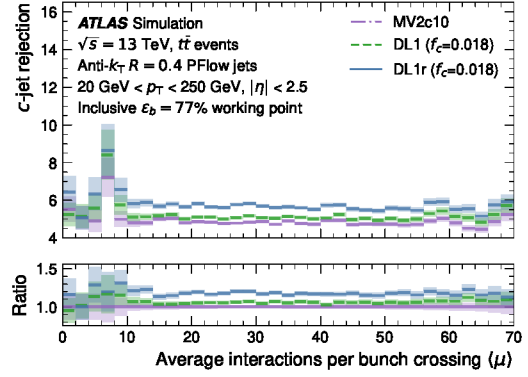
Figure 13: The b -tagging efficiency, ϵ_b , and background-jet rejection factors for several high-level b -taggers, including DL1r, as a function of jet η . (a) shows ϵ_b , (b) shows the light-flavour jet rejection, and (c) shows the c -jet rejection, all at the inclusive 77% efficiency operating point, commonly used in ATLAS analyses of LHC Run 2 pp collision data. The lower panels show the ratio of each algorithm's performance to that of MV2c10. The statistical uncertainties of the efficiencies and rejection factors are calculated using binomial uncertainties and are indicated as coloured bands.



(a)

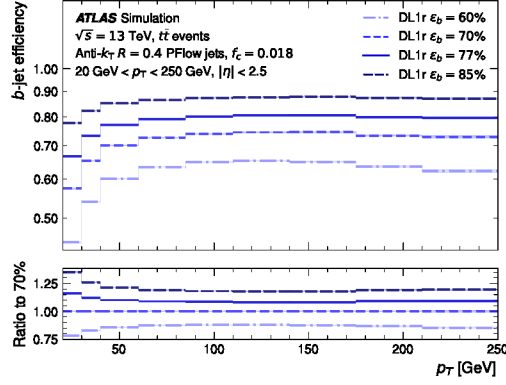


(b)

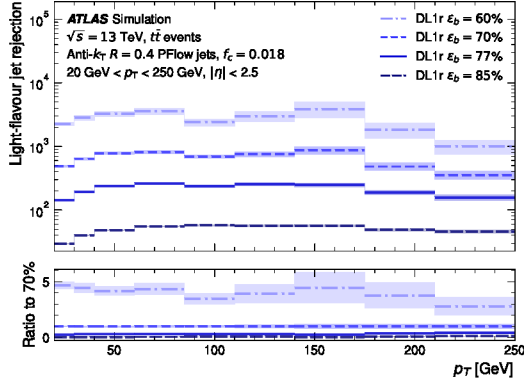


(c)

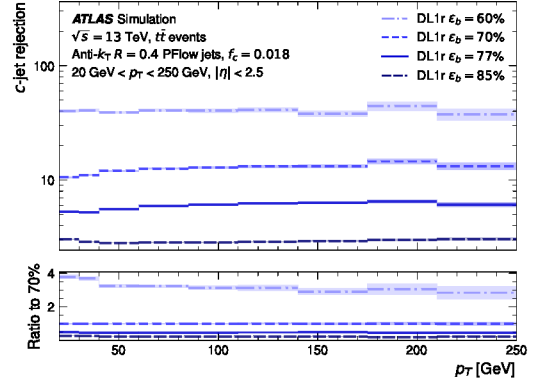
Figure 14: The b -tagging efficiency, ϵ_b , and background-jet rejection factors for several high-level b -taggers, including DL1r, as a function of the average instantaneous luminosity $\langle\mu\rangle$. (a) shows ϵ_b , (b) shows the light-flavour jet rejection, and (c) shows the c -jet rejection, all for a wide range of jet p_T bins at the inclusive 77% efficiency operating point, commonly used in ATLAS analyses of LHC Run 2 pp collision data. The lower panels show the ratio of each algorithm's performance to that of MV2c10. The statistical uncertainties of the efficiencies and rejection factors are calculated using binomial uncertainties and are indicated as coloured bands.



(a)



(b)



(c)

Figure 15: The b -tagging efficiency ε_b and background-jet rejection factors for several operating points as a function of jet p_T . (a) shows the DL1r ε_b , (b) shows the light-flavour jet rejection, and (c) shows the c -jet rejection, all for a wide range of jet p_T bins and for efficiency operating points commonly used in ATLAS analyses of LHC Run 2 pp data. The lower panels show the ratio of each operating point's performance to that of the 70% operating point. The statistical uncertainties of the efficiency (rejection) are calculated using binomial uncertainties and are indicated as coloured bands.

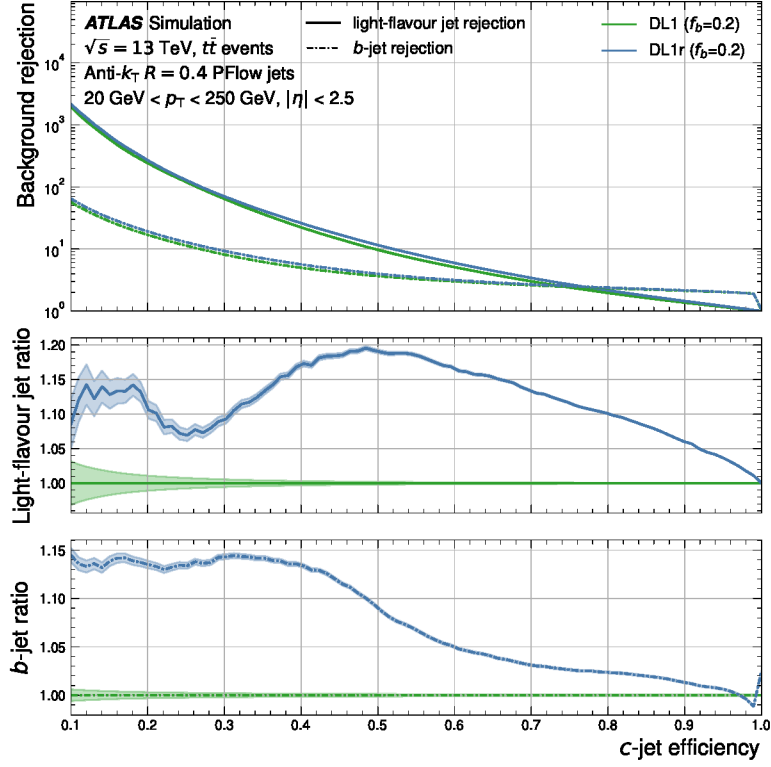


Figure 16: The light-flavour jet and b -jet rejection factors as a function of c -jet efficiency for DL1 and DL1r. The lower two panels show the DL1r-to-DL1 ratios of the light-flavour jet rejection and the c -jet rejection. The statistical uncertainties of the rejection are calculated using binomial uncertainties and are indicated as coloured bands.

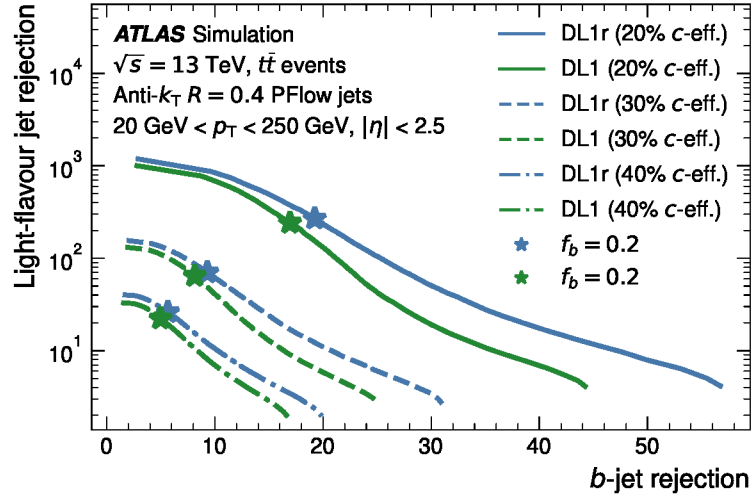


Figure 17: The light-flavour jet rejection as a function of b -jet rejection for inclusive 20%, 30%, and 40% c -jet efficiency operating points for the DL1 and DL1r high-level c -taggers. Each point on a curve corresponds to a particular choice of f_b , the b -jet background fraction in the log-likelihood ratio that defines the tagging discriminant; the star symbols indicate the $f_b = 0.2$ point.

7.3 MC generator dependence

A variety of MC event generators are used in ATLAS analyses to model various production processes at the LHC. PYTHIA [15], HERWIG [50], and SHERPA [51] are commonly used parton-shower and hadronisation programs that describe the structure and composition of hadronic jets. ATLAS PYTHIA and HERWIG MC samples also utilise the EVTGEN package to simulate the decay chains of b - and c -hadrons. The choice of parton-shower generator, hadronisation model, and hadron-decay model affects the predicted performance of the ATLAS flavour-taggers. Tagging efficiencies and mis-tag rates have been studied using various MC generator versions, settings, and tuned parameter values [52]. These effects are taken into account by applying generator-specific corrections to the simulation.

7.4 Overtraining checks

In order to correct the tagging rates of jets in simulation, corresponding rates are measured in data through a variety of procedures [6, 53–56], reaching a precision as good as 1% in b -tagging efficiency for b -jets with $p_T \sim 100$ GeV; the most precise measurement of the c -jet (light-flavour jet) mis-tag rate has uncertainties of approximately 5% (15%). Efficiencies and mis-tag rates in simulation are adjusted via per-jet weights or ‘scale factors’ such that they reflect the performance measured in data. However, $t\bar{t}$ MC events used to train the RNNIP and DL1r algorithms are also used in the efficiency measurements and in physics analyses utilising flavour-tagging. For the calibration procedure to result in a properly corrected simulation, the DL1r performance on the ‘training’ sample, used to optimise the tagging algorithms, and a ‘test’ (or ‘validation’) sample, comprising MC events not contained in the training sample, must be consistent to within the uncertainty associated with the corresponding data-based measurement.

Figure 18 shows the background rejection factors versus the b -tagging efficiency separately for the training and test samples of $t\bar{t}$ events. Figure 19 presents background rejection factors and b -tagging efficiency versus the jet p_T , and the expected performance in the training and validation samples is again compared. No discrepancy is observed that is significant relative to the precision of efficiency measurements, indicating that it is safe to re-use the events used in DL1r training in physics analyses.

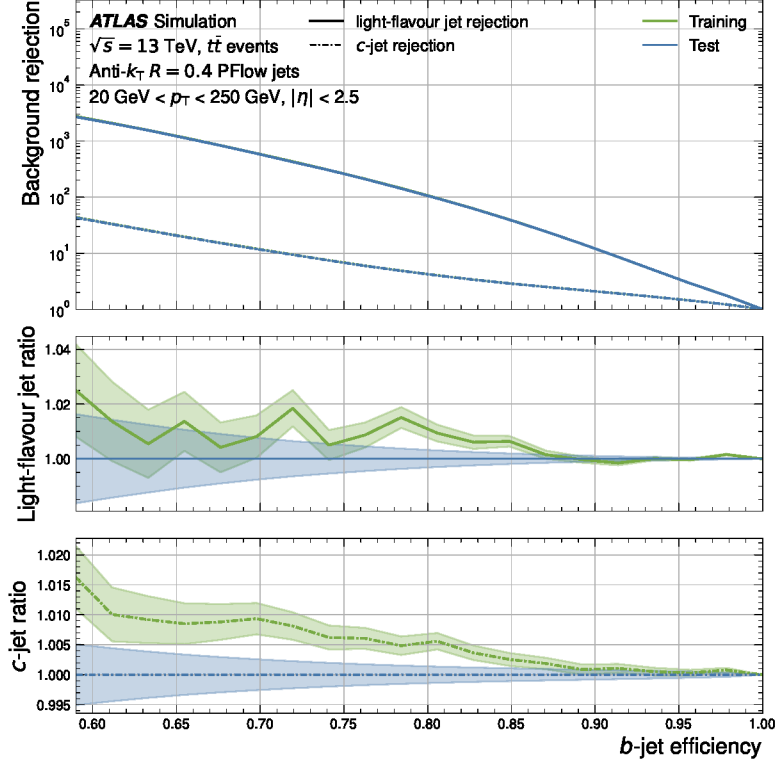


Figure 18: A comparison of the c -jet and light-flavour jet background rejection factors versus the b -tagging efficiency for training and test samples for the DL1r b -tagging algorithm. The training sample contains events used to optimise the DL1r algorithm, while the test sample comprises a statistically distinct set of events that were not used during training. Efficiencies and rejection factors are both derived separately for each sample. The lower two panels show the ratios of training sample to validation sample light-flavour jet rejection and c -jet rejection for DL1r. The statistical uncertainties of the rejection are calculated using binomial uncertainties and are indicated as coloured bands. No difference in performance is observed above the 2% level, which is below the precision of the calibration measurements for these quantities.

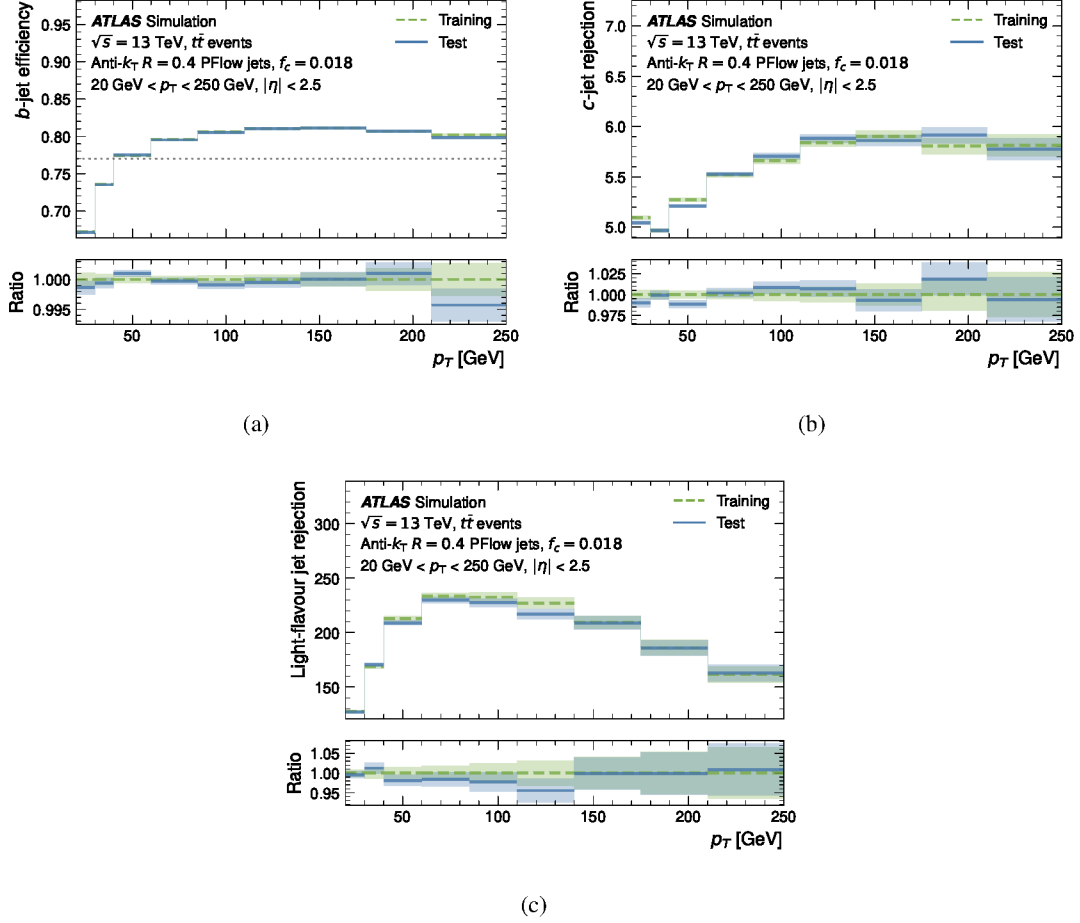


Figure 19: A comparison of performance in jet p_T bins between the training and test $t\bar{t}$ samples for DL1r. The training sample contains events used to optimise the DL1r algorithm, while the test sample comprises a statistically distinct set of events that were not used during training. The (a) b -tagging efficiency, (b) c -jet rejection and (c) light-flavour jet rejection are shown for a broad p_T range at the inclusive 77% efficiency operating point, commonly used in ATLAS analyses of LHC Run 2 pp data. This operating point is derived using the combined test and validation samples. The lower panels show the ratio of training sample performance to validation sample performance. No statistically significant difference in performance is observed at the level of the precision of data-based efficiency measurements, which is $\sim 1\%$ for b -jets, $\sim 5\%$ for c -jets, and $\sim 15\%$ for light-flavour jets. The statistical uncertainties of the efficiency (rejection) are calculated using binomial uncertainties and are indicated as coloured bands.

8 Conclusion

Several flavour-tagging algorithms are used to identify jets containing heavy-flavour hadrons in data recorded by the ATLAS experiment during Run 2 of the LHC. The recent ATLAS strategy combines the results of low-level algorithms (IP2D, IP3D, SV1, JetFitter and RNNIP) into high-level algorithms based on the DL1r feed-forward neural network classifier. The low-level algorithms either exploit the large impact parameters of tracks left by heavy-flavour hadron decay products or attempt to directly reconstruct their decay vertices. Large increases in background-jet rejection are obtained by the DL1r algorithms compared to each individual low-level algorithm and to previous tagging algorithms, illustrating the high complementarity of the low-level inputs. In a sample of simulated Standard Model $t\bar{t}$ events, light-flavour jet (charm-jet) rejection factors of 600 (11) are achieved at a b -jet identification efficiency of 70%; similarly, at a c -jet efficiency of 30%, the obtained light-flavour jet (b -jet) rejection factor is 70 (9).

References

- [1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, [JINST 3 \(2008\) S08003](#).
- [2] L. Evans and P. Bryant, *LHC Machine*, [JINST 3 \(2008\) S08001](#).
- [3] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector*, [Phys. Lett. B 786 \(2018\) 59](#), arXiv: [1808.08238 \[hep-ex\]](#).
- [4] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, [Phys. Lett. B 784 \(2018\) 173](#), arXiv: [1806.00425 \[hep-ex\]](#).
- [5] ATLAS Collaboration, *Performance of b -jet identification in the ATLAS experiment*, [JINST 11 \(2016\) P04008](#), arXiv: [1512.01094 \[hep-ex\]](#).
- [6] ATLAS Collaboration, *ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, [Eur. Phys. J. C 79 \(2019\) 970](#), arXiv: [1907.05120 \[hep-ex\]](#).
- [7] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633>, Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: <https://cds.cern.ch/record/1451888>.
- [8] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, [JINST 13 \(2018\) T05008](#), arXiv: [1803.00844 \[physics.ins-det\]](#).
- [9] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, [Eur. Phys. J. C 77 \(2017\) 317](#), arXiv: [1611.09661 \[hep-ex\]](#).
- [10] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187>.
- [11] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, [JHEP 11 \(2004\) 040](#), arXiv: [hep-ph/0409146](#).
- [12] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, [JHEP 11 \(2007\) 070](#), arXiv: [0709.2092 \[hep-ph\]](#).
- [13] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, [JHEP 09 \(2007\) 126](#), arXiv: [0707.3088 \[hep-ph\]](#).
- [14] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, [JHEP 06 \(2010\) 043](#), arXiv: [1002.2581 \[hep-ph\]](#).
- [15] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, [Comput. Phys. Commun. 178 \(2008\) 852](#), arXiv: [0710.3820 \[hep-ph\]](#).
- [16] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419>.
- [17] R. D. Ball et al., *Parton distributions for the LHC run II*, [JHEP 04 \(2015\) 040](#), arXiv: [1410.8849 \[hep-ph\]](#).

- [18] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: [1207.1303 \[hep-ph\]](#).
- [19] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: <https://cds.cern.ch/record/2216168>.
- [20] ATLAS Collaboration, *Study of top-quark pair modelling and uncertainties using ATLAS measurements at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2020-023, 2020, URL: <https://cds.cern.ch/record/2730443>.
- [21] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152.
- [22] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: <https://cds.cern.ch/record/2206965>.
- [23] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, arXiv: [1005.4568 \[physics.ins-det\]](#).
- [24] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.
- [25] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) 673, arXiv: [1704.07983 \[hep-ex\]](#).
- [26] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-051, 2015, URL: <https://cds.cern.ch/record/2110140>.
- [27] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) 332, arXiv: [1611.10235 \[hep-ex\]](#).
- [28] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: <https://cds.cern.ch/record/2037717>.
- [29] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466, arXiv: [1703.10485 \[hep-ex\]](#).
- [30] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, arXiv: [0802.1189 \[hep-ph\]](#).
- [31] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896, arXiv: [1111.6097 \[hep-ph\]](#).
- [32] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2020) 689, arXiv: [2007.02645 \[hep-ex\]](#).
- [33] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: <https://cds.cern.ch/record/1700870>.
- [34] M. Cacciari, G. P. Salam and G. Soyez, *The catchment area of jets*, *JHEP* **04** (2008) 005, arXiv: [0802.1188 \[hep-ph\]](#).

- [35] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: <https://cds.cern.ch/record/2273281>.
- [36] ATLAS Collaboration, *Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL: <https://cds.cern.ch/record/2255226>.
- [37] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, *Studies in Computational Intelligence* **385** (2012).
- [38] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, URL: <https://cds.cern.ch/record/2270366>.
- [39] ATLAS Collaboration, *Topological b -hadron decay reconstruction and identification of b -jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-PUB-2018-025, 2018, URL: <https://cds.cern.ch/record/2645405>.
- [40] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth. A* **262** (1987) 444.
- [41] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, *Chapter 10 - Deep learning*, (2017) 417, ed. by I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, URL: <https://www.sciencedirect.com/science/article/pii/B9780128042915000106>.
- [42] F. Chollet et al., *Keras*, 2015, URL: <https://keras.io>.
- [43] D.P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, (2014), arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.SC].
- [44] S. Ioffe; C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, (2015), arXiv: [1502.03167](https://arxiv.org/abs/1502.03167) [cs.LG].
- [45] ATLAS Collaboration, *Measurement of VH , $H \rightarrow b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector*, *JHEP* **05** (2019) 141, arXiv: [1903.04618](https://arxiv.org/abs/1903.04618) [hep-ex].
- [46] ATLAS Collaboration, *Measurement of Higgs boson decay into b -quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **06** (2021) 097, arXiv: [2111.06712](https://arxiv.org/abs/2111.06712) [hep-ex].
- [47] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs using 139 fb⁻¹ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **03** (2020) 145, arXiv: [1910.08447](https://arxiv.org/abs/1910.08447) [hep-ex].
- [48] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-018, 2015, URL: <https://cds.cern.ch/record/2037683>.
- [49] ATLAS Collaboration, *Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector*, (2022), arXiv: [2201.11428](https://arxiv.org/abs/2201.11428) [hep-ex].
- [50] J. Bellm et al., *Herwig 7.1 Release Note*, (2017), arXiv: [1705.06919](https://arxiv.org/abs/1705.06919) [hep-ph].
- [51] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, arXiv: [1905.09127](https://arxiv.org/abs/1905.09127) [hep-ph].

- [52] ATLAS Collaboration, *Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration*, ATL-PHYS-PUB-2020-009, 2020, URL: <https://cds.cern.ch/record/2718610>.
- [53] ATLAS Collaboration, *Measurement of the c -jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, *Eur. Phys. J. C* **82** (2021) 95, arXiv: [2109.10627](https://arxiv.org/abs/2109.10627) [hep-ex].
- [54] ATLAS Collaboration, *Calibration of the b -tagging efficiency on charm jets using a sample of $W + c$ events with $\sqrt{s} = 13$ TeV ATLAS data*, ATLAS-CONF-2018-055, 2018, URL: <https://cds.cern.ch/record/2652195>.
- [55] ATLAS Collaboration, *Calibration of the ATLAS b -tagging algorithm in $t\bar{t}$ semileptonic events*, ATLAS-CONF-2018-045, 2018, URL: <https://cds.cern.ch/record/2638455>.
- [56] ATLAS Collaboration, *Calibration of light-flavour b -jet mistagging rates using ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2018-006, 2018, URL: <https://cds.cern.ch/record/2314418>.