Optimisation and performance studies of the ATLAS *b*-tagging algorithms for the 2017-18 LHC run

The ATLAS Collaboration

The optimisation and performance of the ATLAS *b*-tagging algorithms for the 2017-18 data taking at the LHC are described. This note presents the use of additional taggers to further enhance the discrimination between *b*-, *c*- and light-flavour jets, and on new studies for more performant training of the algorithms and for assessing the universality of the training campaign in typical physics processes where flavour tagging plays a crucial role. Particular attention is paid to the inclusion of novel taggers, namely a Soft Muon Tagger, based on the reconstruction of muons from the semileptonic decay of *b*/*c*-hadrons, and a Recurrent Neural Network Impact-Parameter tagger that exploits correlations between tracks within the jet. New variants of the high-level discriminant, based on boosted decision trees and modern deep learning techniques, are also presented. The overlap between the jets tagged by the various *b*-tagging algorithms is studied, and the dependence of the tagging performance on the physics process producing the jets is explored. Comparisons between Monte Carlo simulation and 2016 data for both the input variables and the output *b*-tagging discriminants are also shown.

1 Introduction

Flavour tagging, namely the identification of jets containing b and c hadrons, is key to many fundamental measurements and searches for new phenomena performed by the ATLAS experiment [1] at the LHC. This note describes the flavour tagging algorithms to be used for the analysis of the data sample being recorded in 2017-18.

The main ingredients used in flavour tagging algorithms are the trajectories of charged particles reconstructed in the Inner Detector (ID) and associated to jets. The ATLAS ID system [2] consists of a silicon pixel detector, a silicon micro-strip detector (SCT) and the straw tubes of the transition radiation tracker (TRT). It is used to reconstruct the particle tracks and determine their parameters, i.e. their trajectories and momenta in the region of acceptance $|\eta| <2.5$ [1]. The ID is placed inside a solenoid magnet providing a 2T axial magnetic field. In order to improve the track extrapolation and the vertex reconstruction resolution, a fourth pixel layer, the Insertable B-Layer (IBL) [3] was added at the beginning of Run 2. This additional layer is placed at small radius (~ 3.3 cm) with respect to the beam pipe and increased the ATLAS tracking performance by providing a high-resolution space point at a radius close to the LHC interaction region. This upgrade improves b-tagging performance, mostly in the low-to-medium-jet p_{T} region, due to the reduced lever arm of particle trajectory extrapolation. The full chain of the flavour tagging algorithms has been updated to account for the inclusion of the IBL at the beginning of the Run 2 data-taking in 2015 [4] and subsequently optimised for the 2016 data-taking conditions [5].

Lead/liquid-argon (LAr) [6] sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity and longitudinal segmentation. A hadron (steel/scintillator-tile) calorimeter [7] covers the central pseudorapidity range ($|\eta| < 1.7$). The endcap and forward regions are instrumented with LAr calorimeters for both EM and hadronic energy measurements up to $|\eta| < 4.9$. A large air-core Muon Spectrometer (MS) [8], providing standalone precision muon momentum reconstruction in the range $|\eta| <2.7$, uses a combination of drift tubes and resistive plate chamber technologies and is equipped with dedicated detectors for triggering and precise timing. A system of one barrel and two endcap magnet toroids provides a bending power ranging between 1 T·m and 7.5 T·m.

Several improvements have been implemented in the ATLAS detector simulation and event reconstruction algorithms for the 2017-18 data-taking and corresponding simulation, in particular for the reconstruction of tracks in the ID. A more realistic charge deposition model, based on the Bichsel model [9], was introduced in the simulation of the Pixel detector. This model results in a broader distribution of the reconstructed position of charge deposition that improves agreement with that observed in data and affects the track extrapolation resolution. The pixel clustering is based on the response of a series of Neural Network (NN) algorithms [10] computing the hit spatial position, its uncertainties and the splitting of pixel clusters resulting from the charge deposition of multiple tracks. This algorithm has also been improved in the 2017 ATLAS software version. These features improve the tracking performance and the detector modelling in simulation in terms of data/MC agreement.

The improvements in simulation and reconstruction are augmented by novel b-tagging approaches ex-

---

[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 η is defined in terms of the polar angle θ as η=-$\log \tan(\frac{\theta}{2})$ while $\Delta R=\sqrt{\Delta \eta^2 + \Delta \phi^2}$.

ploited for the 2017 data-taking. The traditional track-based impact parameter taggers IP2D and IP3D, based on a log-likelihood ratio (LLR) discriminant separating tracks associated to jets according to whether or not they are compatible to the primary vertex hypothesis, are complemented by a new technique to take into account the track-to-track correlations. Such an approach has been introduced through a Recurrent Neural Network in the RNNIP tagger [11], as discussed in Section 3.1.2. The presence of muons coming from semileptonic decays of b- and c- hadrons is a distinctive feature that can be exploited to enhance the b-tagging efficiency. This is implemented by a dedicated Soft Muon Tagger (SMT) combining a set of variables to identify muons in jets (see Section 3.3). Vertex-based taggers use both an inclusive secondary vertex approach (SV1) and a topological decay reconstruction (JetFitter). In addition, dedicated algorithms for c-tagging, based also on the topological properties of the single displaced charm hadron decay vertex reconstructed by the JetFitter algorithm (Section 3.2.3) are implemented.

The discriminating variables produced by all these b-tagging algorithms are used as inputs to high-level tagger algorithms, i.e. multivariate discriminants for b- and c- tagging. The training strategy used for these algorithms is described in Section 4.1 and the multivariate-based taggers are discussed in Sections 4.2 and 4.3. The algorithm performance of such b-tagging algorithms is discussed in Section 5. In parallel, a multivariate classification algorithm based on deep learning techniques (DL1) is also developed and is discussed in Section 4.3. Checks on the modelling of variables in MC simulation are performed in order to assess the robustness of the algorithms and to identify possible differences between the performance in data and in the simulation, which might result in significant differences in the b-tagging efficiencies and c/light-flavour jets mistag rates. Data-MC comparisons for the low-level taggers and the high-level algorithm outputs are shown in Sections 3, 4 and 5. The data used in these comparisons are from 2016, but reprocessed using the same reconstruction software and algorithms as will be used for data collected in 2017. The performance of the new b-tagging algorithms is evaluated on benchmark physics processes with b-jets in the final state in Section 6.

2 Data and Monte Carlo Samples, Objects and Event Selection

The results discussed in this note are based on MC simulation samples of $t\bar{t}$ and Z' produced in proton-proton collisions with a centre-of-mass-energy of 13 TeV. The $t\bar{t}$ sample is simulated with Powheg [12] and interfaced with Pythia8 [13] for parton shower and hadronisation. Only $t\bar{t}$ decays with at least one lepton from a subsequent W boson decay are simulated. The CT10 [14, 15] parton distribution function set is used. Events with the associated production of a single top quark and a W boson, referred to as Wt events, form a significant background to $t\bar{t}$ production and are also simulated with Powheg interfaced to Pythia6 [16]. The high-level tagger training uses a dedicated sample of Z' decaying to jets to optimise the performance at high jet p_T. Pythia8 is used to generate this sample with the A14 [17] set of tuned parameters for the underlying event and the leading-order NNPDF2.3 [18] parton distribution function. The cross-section of the hard-scattering process is modified by applying an event-by-event weighting factor to broaden the natural width of the resonance and widen the transverse momentum (p_T) distribution of the jets produced in its hadronic decays. The branching fractions of these decays are set to be one-third each for the bb, cc and light-flavour quark pairs, to give a p_T spectrum uniformly populated in jets of all flavours. $Z \rightarrow \mu^+ \mu^-$ + jets events are generated with Powheg and interfaced with Pythia8. This sample is used for comparing simulation to data in regions where light-flavour jets are dominant. The fraction of b-jets in this sample is $\sim 7\%$ in the simulation. Events are reweighted to reproduce the distribution of the number of jets from MadGraph, which is expected to give a better description of of Z + jet data. In addition, jets are also re-

weighted from Pythia8 to MadGraph to account for the difference in the b-jet yield in the two generators.

For the physics sample comparisons discussed in Section 6, the generation of the direct pair production of scalar bottom and top quarks is performed using MadGraph5 interfaced with Pythia8 for parton shower and hadronisation. The NNPDF2.3 PDF set is used. Z' events, with nominal mass of $m_0 = 4$ TeV, are generated with Pythia8 and the same PDF set, ZH events are generated with Powheg and hadronised using Pythia8. $V+$ jets events, considered as a background to VH production, are generated with Sherpa 2.2 [19] and the NNPDF2.3 PDF set.

The EvtGen [20] program is used to model the decay of b- and c- hadrons. Additional inelastic collisions (pile-up) are included with the hard scattering event in the simulation using minimum bias interactions generated with Pythia8. The average number of pile-up events per bunch crossing, $< \mu >$, in the simulation is 24.4. The propagation of particles through the detector is performed using GEANT4 [21] and the ATLAS reconstruction is applied to the simulated events.

The hard-scatter primary vertex (PV) is defined to be the one with the largest sum of squared transverse momenta of the associated tracks matched to the vertex.

Jets are reconstructed from topological clusters in the calorimeters using the anti-k_t algorithm [22] with radius parameter R=0.4. The selection applied in this note requires jets with $p_T > 20$ GeV and $|\eta| < 2.5$. In addition, a requirement on the output of the Jet Vertex Tagger algorithm (JVT) [23] is also applied. This algorithm allows jets originating from pile-up interactions to be suppressed using track-based variables included in a multivariate discriminant. Jets with $p_T < 60$ GeV and $|\eta| < 2.4$ are removed if the JVT output is smaller than 0.59. This cut is 92% efficient for jets originating from the hard scatter vertex, with a residual rate for pile-up jets of approximately 2%.

The labeling scheme used to define the flavour of the jet (b-, c- or light-flavour jets) is performed by matching the jets to b- and c-hadrons with $p_T > 5$ GeV within a cone of radius $\Delta R = 0.3$ around the jet axis. First, b-hadrons within the cone are searched for; if no b-hadron is found, the same procedure is repeated for c-hadrons and for τ leptons. The remaining jets with no such matching hadrons or τ leptons are assigned to the light-flavour jet category. Overlaps between jets and leptons, caused by electrons reconstructed as jets, are suppressed by removing jets which are within $\Delta R < 0.3$ of true electrons or muons coming from W/Z boson decays.

2.1 Data/Monte Carlo comparisons

Given the complexity of the taggers, which use a large number of variables from low-level taggers probing different kinematical and topological features of tracks in jets, it is particularly important to ensure that simulation correctly models the relevant properties. Imperfections in the modelling of heavy and light flavour jet fragmentation and in the production and decay of heavy flavour hadrons can produce discrepancies between data and simulation which need to be understood to minimise systematic uncertainties. The modeling of variables is explored using two dilepton samples selected in data - an $e\mu$ sample dominated by $t\bar{t}$ events which are a rich source of b-jets, and a $Z \rightarrow \mu\mu$ + jets sample where the majority of the jets are light flavour.

Subsets of the data collected by the ATLAS detector from proton-proton collisions in 2016 at a centre of

mass energy of 13 TeV are used for the studies in this note. These sets of data are chosen to uniformly sample the data collected at different pile-up conditions. Events with two charged leptons were selected online using either a single electron or single muon trigger. The thresholds were set to be almost fully efficient for leptons with $p_T > 25$ GeV passing offline selection requirements on the lepton identification and isolation [24]. The selected $e\mu$ + jets events correspond to a total integrated luminosity of 2.5 fb^{-1} and the selected $Z \to \mu\mu$ + jets events to 0.5 fb^{-1}. In both cases, simulation events have been reweighted to reproduce the pile-up distributions of the data samples.

The $t\bar{t}$-dominated e-μ sample is selected by requiring an opposite charge sign e-μ pair, with the leading lepton p_T >25 (20) GeV, for electrons (muons) and the p_T of the second lepton > 15 GeV. The invariant mass of the dilepton pair is required to be greater than 10 GeV. Events are required to have between two and seven jets passing the selection discussed in Section 2. The subleading jet is required to be identified as a b-jet using the MV2 discriminant (see Section 4.2) at a working point with an identification efficiency of 77%. The other selected jets are used to study the flavour tagging algorithm response. Simulation predicts b-jets to account for ~25 % of the other selected jets. Data selected in this way are compared with $t\bar{t}$ and Wt simulated samples, properly rescaled to their relative cross-sections, for data/MC agreement studies.

$Z \to \mu\mu$ events are selected by requiring two oppositely signed muons with leading (subleading) p_T > 25 (20 GeV) and a dimuon invariant mass satisfying $76 < m_{\mu\mu} < 106$ GeV. In addition at least one jet with $p_T > 20$ GeV, $|\eta| < 2.5$ is required. No b-tagging requirement is imposed on the jets and only the two highest p_T jets in the event are kept.

The data-MC comparisons shown in the note are meant to provide an assessment of the overall agreement and to identify localised effects and possible mis-modellings. Only statistical uncertainties are taken into account and systematic effects that may arise from the different flavour compositions in data and MC simulation are neglected in the following studies. The final tagging performance will be calibrated in data as reported for Run-1 data in Ref. [25]. The label MC16 in all the figures included in this note indicates the simulation corresponding to the 2017 reconstruction and software configurations.

3 Low level taggers

The identification of jets containing b-hadrons is based on distinctive features exploited in different algorithm strategies: impact-parameter based algorithms (IP3D and IPRNN, described in Sections 3.1.1 and 3.1.2) and inclusive secondary vertex reconstruction algorithms discussed in Section 3.2.1 and 3.2.2. The information on muons in the jet, produced by the semileptonic decay of the b- and c-hadrons, is included in a dedicated Soft Muon Tagger (SMT) algorithm described in Section 3.3. The outputs of these algorithms are combined in multivariate discriminants, a Boosted Decision Tree (BDT)-based MV2.

3.1 Track-based b-tagging algorithms

Reconstructed tracks are parametrised with five perigee parameters. Reconstructed tracks are parametrised by using five perigee parameters. Among those, the transverse and longitudinal impact parameters

correspond to the distance of closest approach to the PV in the r-ϕ plane (d_0) and in the longitudinal plane (z_0), respectively.

Track-based b-tagging algorithms rely on the information from these impact parameters, and their significances, defined as the impact parameter divided by its estimated uncertainty. Due to the long lifetime of b-hadrons (~1.5 ps, $c\tau$ ~450 μm) , the b-hadron charged decay products are characterised by displaced vertices and thus large impact parameter values, given by $\beta\gamma c\tau\sin\alpha$, where α is the particle angle of emission w.r.t. the original heavy hadron direction. Therefore, tracks generated from b-hadron decays have impact parameter significances which differ significantly from zero, whereas those from tracks in light jets are typically consistent with zero.

The track selection applied for the impact parameter algorithm is based on the following requirements:

- p_T >1 GeV;

- $|d_0|$ <1 mm and $|z_0 \sin\theta|$ <1.5 mm;

- seven or more silicon hits with at most two silicon holes, at most one of which is in the pixel detector, where a hole is defined as a hit expected to be associated with the track but not present [10].

Tracks are associated to the nearest jet by exploiting the angular separation ΔR between the track and the jet axis. The ΔR requirement varies as a function of jet p_T, being wider for low p_T values (0.45 for jet p_T of 20 GeV) and narrower for high values (0.26 for jet p_T of 150 GeV) since decay products from energetic heavy-flavoured hadrons are more collimated [25]. A lifetime signing convention is adopted for the track-based tagger algorithms. The impact parameter is given a positive sign if the position of crossing of the track with the jet axis is located upstream to the PV along the jet axis which defines the direction, and negative otherwise. Distributions of the values of the transverse impact parameter significance, d_0/σ_{d_0} (denoted as S_{d_0}), and longitudinal impact parameter significance, z_0/σ_{z_0} (denoted as S_{z_0}) for tracks in the leading jet of Z + jets events are shown in Figure 1, comparing the response of the 2016 and 2017 configurations on the same data collected in 2016 (reprocessed with 2016 and 2017 reconstruction algorithms). The 2017 simulation configuration includes the Bichsel model for energy deposition in the pixels. A significant improvement in the agreement between data and MC simulation, dominated by contributions from particles originating at the PV, is observed.

3.1.1 IP2D and IP3D

The IP2D tagger makes use of the transverse impact parameter significance, d_0/σ_{d_0}, as discriminating variable whereas IP3D uses both the transverse and the longitudinal impact parameter significance, $z_0 \sin\theta/\sigma_{z_0 \sin\theta}$, in a two-dimensional template to account for their correlation. Details of the algorithms, as well as the distributions of the transverse and longitudinal impact parameter significance values for tracks from b-, c- and light-flavour jets, can be found in Ref. [5]. The IP2D and IP3D algorithms are based on a LLR discriminant. This discriminant is computed as the sum of per-track contributions, $\sum_{i=1}^{N} \log\left(\frac{p_b}{p_u}\right)$, where N is the number of tracks for a given jet and p_b, p_u are the template probability density functions (PDF) for the b- and light-flavour jet flavour hypotheses, respectively, assuming no correlation among the various tracks contributing to the sum. These template probability density functions are obtained from reference histograms for the transverse and longitudinal impact parameter significances and are derived from MC simulation. They are separated into exclusive categories (the track grade) that depend on the hit pattern of a given track to increase the discriminating power, and used to calculate ratios of the b- and light-flavour jet probabilities. This computation is performed on a per-track basis. Different
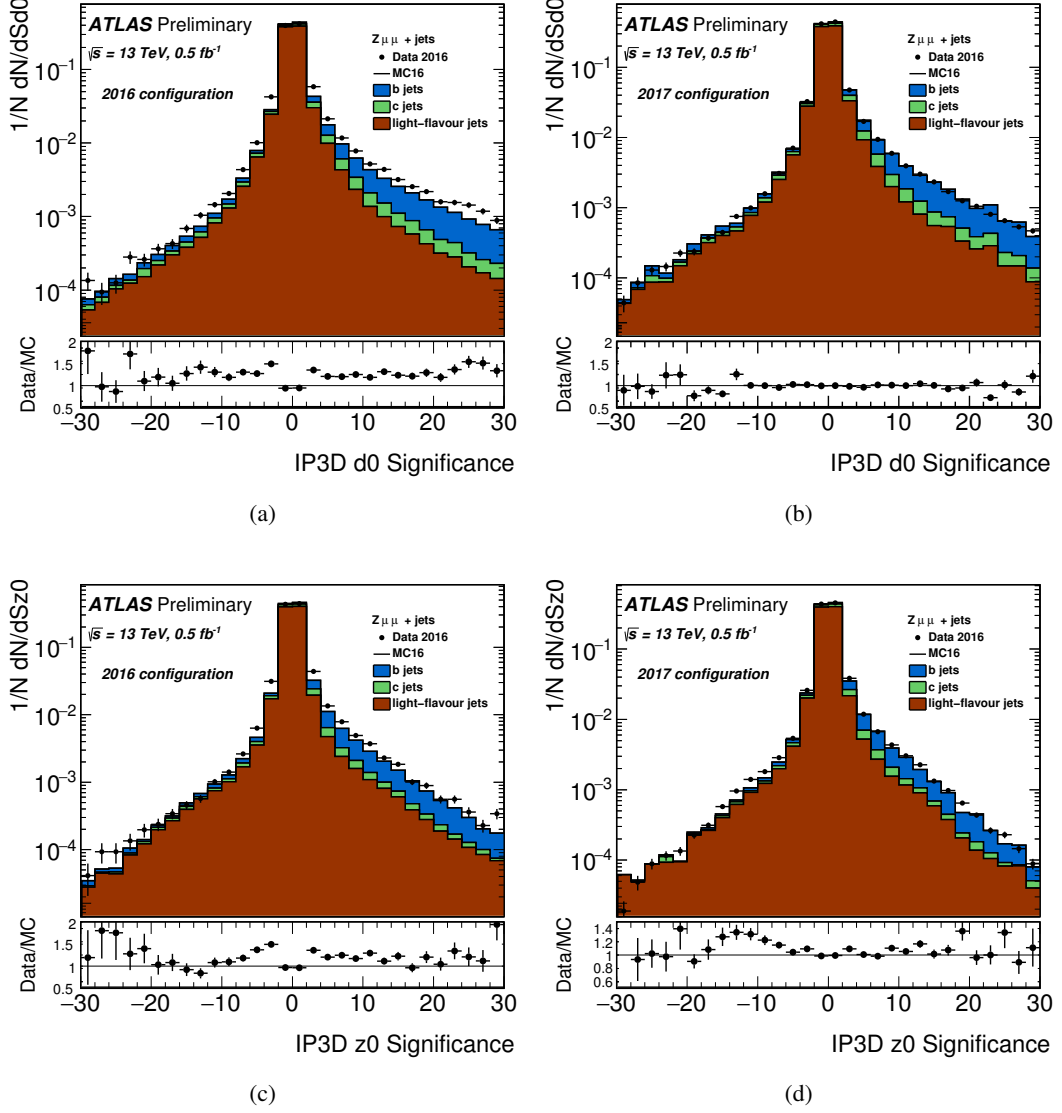
Figure 1: Data-MC comparisons of the transverse (a,b) and longitudinal (c,d) impact parameter significance values for IP3D selected tracks in the leading jet of the $Z \rightarrow \mu\mu$ + jets dominated sample. The plots in the upper row show the transverse impact parameter while the longitudinal impact parameters are shown in the lower row. The plots on the left are for the 2016 configuration and on the right for the 2017 configuration discussed in the text.

PDFs are used for the various track categories and for each jet-flavour hypothesis (b, c, light).

Unlike the 2016 optimization (Ref. [5]) where templates were defined on a mixture of $t\bar{t}$ and $Z' \to t\bar{t}$ (50%/50%) for the two categories with no hits in the first two layers despite them being expected, these PDFs are now produced using a 50%/50% mixture of a $Z' \to b\bar{b}, c\bar{c}, q\bar{q}$ sample described in Section 2 and a $t\bar{t}$ sample. These categories are particularly populated by long-lived b hadrons which make it through the first detector layers before they decay. The $t\bar{t}$ sample is used to populate all remaining categories. Figure 2 shows data-MC comparisons for the IP3D b vs u log-likelihood ratio in $t\bar{t}$ and $Z \to \mu^+\mu^-$ + jets selected events with the different b-, c- and light-flavour jets components highlighted. Simulation describes the
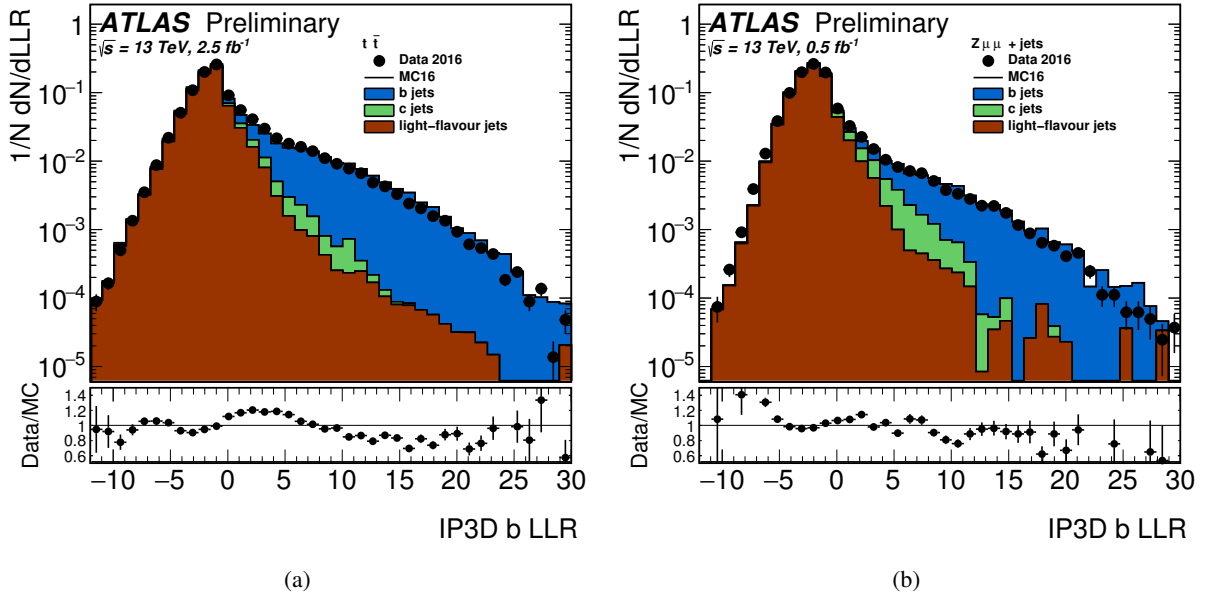
Figure 2: Data-MC comparison of the log-likelihood ratio used to discriminate the b- from the light-flavour jet hypotheses in the IP3D b-tagging algorithm using the selected $t\bar{t}$-dominated $e\mu$ sample (a) and $Z \to \mu^+\mu^-$ + jets-dominated sample (b).

response of the data within around 20 % with some localised discrepancies due to the corresponding difference in response in the impact parameter significance distributions shown in Figure 1.

3.1.2 Recurrent Neural Network Track-based tagger

Within the decay of a b-hadron, several charged particles can emerge from the secondary (or tertiary) decay vertex with large impact parameters. These impact parameters are intrinsically correlated: if one track is found with a large impact parameter then finding a second track with 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 should not exist. The baseline IP based b-tagging algorithm, IP3D, described in Section 3.1.1 uses likelihood templates to compute per-flavour conditional likelihoods. Due to the large sample statistics needed to compute such templates, IP3D assumes that the properties of each track in a jet are independent of all other tracks, which limits its ability to fully model the properties of b-jets.

Recurrent neural networks can be used to overcome this challenge.

Recurrent neural networks (RNN) are used to directly learn sequential dependencies between the variables in sequences of arbitrary length [26], as is the case for the tracks in a jet. For each selected track, the transverse (S_{z_0}) and longitudinal (S_{d_0}) impact parameters, 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 ($\Delta R(\text{track}, \text{jet})$) and the track grade defined previously as in the IP2D and IP3D algorithms, are fed to the network (see Ref. [11] for more details). The tracks are ordered by the $|S_{d_0}|$ values and passed to a neural network cell which transforms the track sequence to a vector. The outputs provided by the network correspond to the b-jet, c-jet, light-flavour jet, and τ-jet probabilities. The four b-tagging probabilities, p_b, p_u, p_c and p_τ are used as input variables in the high level taggers as discussed in Sections 4.2 and 4.3.

The outputs of the RNN are combined into the discriminant function for the b-jet hypothesis:

$$D_{\text{RNN}}(b) = \ln \frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau)p_{\text{light}}} \tag{1}$$

where f_c and f_τ are the c- and τ-jet fractions, respectively. The relative importance of c-jet, τ-jet and light-flavour jet rejection can be changed by varying the corresponding weights to the output nodes of the network. These parameters are fixed to f_c=0.07 and f_τ=0 based on the $t\bar{t}$ training sample. Background rejection vs signal efficiency curves are shown in Figure 3. The RNNIP obtains approximately twice the light-rejection and 1.2 times the c-rejection compared to IP3D. A constant-efficiency 70% working point is adopted to distinguish the effect of varying b-tagging efficiency from that of light-flavour and c-jet mis-identification probability as in the standard working points defined with a fixed cut on the discriminant. In this case, all taggers have a 70% efficiency for b-jets across the p_T spectrum while the rejection factors vary. The RNNIP algorithms outperforms IP3D across the full p_T range, as shown in Figure 4.

The IP3D and RNNIP taggers are based on similar input information, namely the impact parameters of each track. To determine if they exploit complementary features of this information the rates of jets with a positive tag by only one of the two algorithms (*exclusive tag*) has been studied for a b-tag efficiency of 77%. Studies show that there is a sizeable fraction of b-jets tagged by only one of the two algorithms. This is due to a significant difference in the algorithm sensitivity to the heavy hadron charged decay multiplicity and decay distance. Decays with low track multiplicity and with short decay distances, within the radius of the innermost pixel layer, are efficently tagged by IP3D. Decays with higher multiplicity and those with longer decay distances are more efficiently tagged by RNNIP. As a result, the use of the input from both algorithms in the MV2 training is beneficial to its performance as discussed in Section 5. The complementarity of the two taggers for the MV2 performance is tested by removing the IP3D inputs from the set of variables used by MV2 and re-evaluating its performance: a drop of 20% in the light-flavoured jet rejection at the 77% b-efficiency working point is observed compared to that obtained when both IP3D and RNNIP are used.

Data-MC comparisons are presented in Figure 5 of the $D_{\text{RNN}}(b)$ combined probabilities for the b-jet hypothesis obtained for jets in $t\bar{t}$ and $Z \rightarrow \mu^+\mu^-$ + jets events. As for the distribution for IP3D, the agreement between data and simulation is within 20% over the full range.

3.2 Secondary vertex-based b-tagging algorithms

ATLAS uses two algorithms for secondary vertex reconstruction; one reconstructs a single displaced vertex (SV1), and the other performs a topological decay reconstruction along the jet axis (JetFitter).
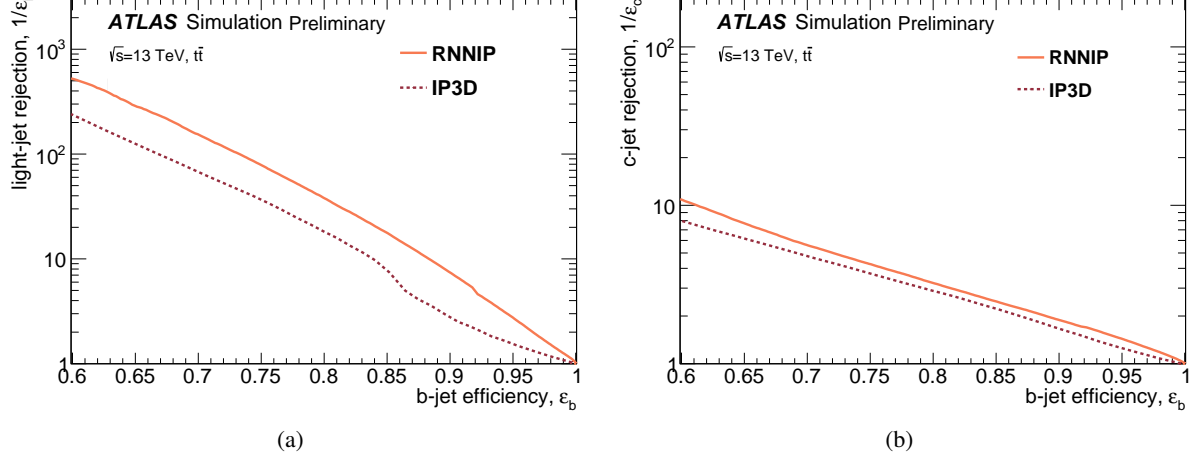
Figure 3: Light-flavour jet (a) and c-jet (b) rejection as a function of the b-tagging efficiency comparing the IP3D and RNNIP b-tagging algorithms for selected jets in simulated $t\bar{t}$ events.
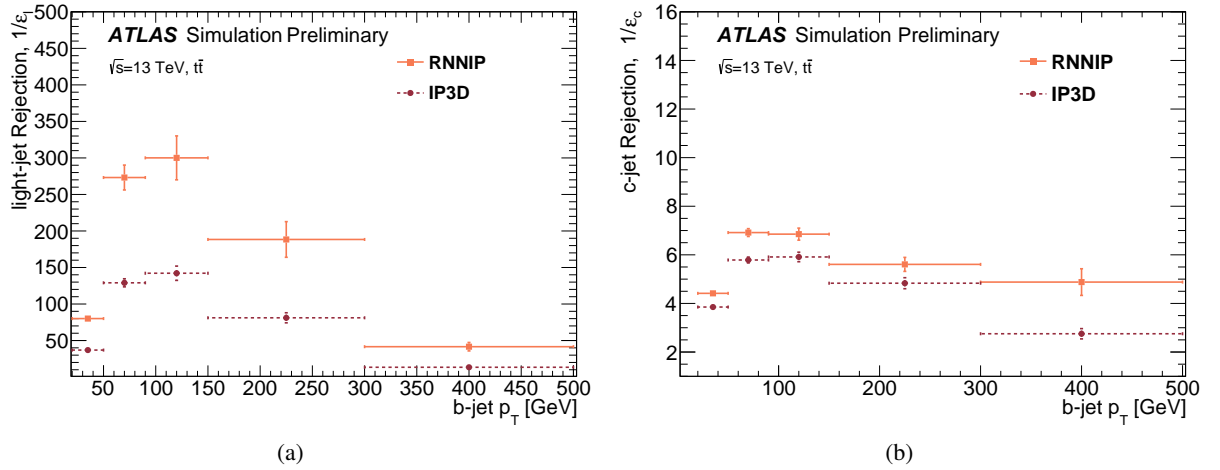

Figure 4: Light-flavour jet rejection (a) and c-jet rejection (b) at constant b-tagging efficiency of 70% as a function of the jet p_T comparing the IP3D and RNNIP b-tagging algorithms for selected jets in simulated $t\bar{t}$ events.
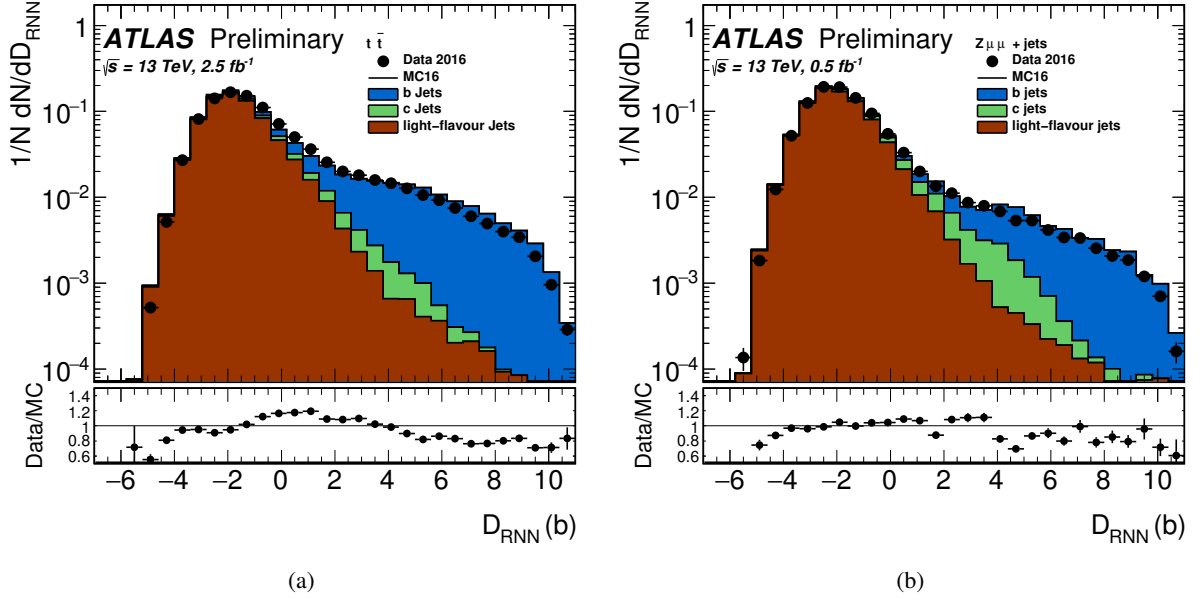
Figure 5: Data-MC comparison of the log-likelihood ratio used to discriminate the b- from the light-flavour jet hypotheses in the RNNIP b-tagging algorithm using the $t\bar{t}$-dominated $e\mu$ sample (a) and $Z \to \mu^+\mu^- +$ jets-dominated sample (b).

These are based on the topology of tracks associated to jets as discussed above. The track selection, based on their impact parameters d_0 and z_0 and the number of associated hits in the tracking detectors, has been optimised independently for each of the algorithms.

3.2.1 Secondary Vertex Finder algorithm

The secondary vertex finding algorithm (SV1) [27] reconstructs a single displaced secondary vertex in a jet. The reconstruction starts from the set of all possible two-track vertices and rejects tracks that are compatible with the decay of long-lived particles (K_s or Λ), photon conversions or hadronic interactions with the detector material. An inclusive secondary vertex is then reconstructed from the accepted tracks and outlier tracks are iteratively removed from the fit, while still allowing decay products from b and c vertices to be associated to a single common secondary vertex. More details on the reconstruction strategy and distributions of observables of the reconstructed secondary vertex comparing vertices from b-, c- and light flavour jets can be found in Ref. [5]. Eight discriminating variables, including the number of tracks associated to a SV1 vertex, the invariant mass of the secondary vertex, its energy fraction (defined as the total energy of all the tracks associated to the secondary vertex divided by the energy of all the tracks associated to the jet), and the three-dimensional decay length significance are used as inputs to the high-level taggers, as described in Sections 4.2 and 4.3. The distributions are compared in data and MC simulation for the $t\bar{t}$-dominated $e\mu$ sample in Figure 6. A good modelling of the number of tracks associated to the SV1 vertex, their invariant mass and decay length significance is observed while the fraction of the jet energy carried by this vertex shows some disagreement at low and high values. This is also observed for the other secondary vertex fitter, which may be correlated with the estimate of the jet

energy. The rate of jets tagged by the SV1 algorithm, in the selected $Z \to \mu^+\mu^-$ + jets-dominated sample and $t\bar{t}$-dominated $e\mu$ sample is summarised in Table 1.

3.2.2 Topological Multi-Vertex Algorithm (JetFitter)

JetFitter (JF) is an inclusive vertex algorithm that reconstructs the topological vertex structure of heavy hadron decays inside a jet. The topological algorithm is based on a modified Kalman Filter. It uses the intercepts of the particle tracks with the jet axis in order to reconstruct the full decay topology, also including single-prong vertices. Several improvements have been introduced in the current version of the JetFitter algorithm to increase the efficiency for tertiary vertex reconstruction and to mitigate the effect of pile-up tracks. Eight discriminating variables, including the track multiplicity at the JetFitter vertex, the mass of the vertex, its energy fraction and the three-dimensional decay significance, are used as inputs to the high-level taggers. As shown in Figure 7, good agreement between data and simulation is observed, with the exception of the JF vertex energy fraction, which reproduces the pattern observed also for SV1. The rate of jets being tagged by the Jet Fitter algorithm, in the selected $Z \to \mu^+\mu^-$ + jets-dominated sample and $t\bar{t}$-dominated $e\mu$ sample is summarised in Table 1.

Tagger ($Z \to \mu^+\mu^-$+jets)	Rate in data	Rate in MC
SV1	0.109±0.001	0.108±0.001
JF secondary vertex	0.145±0.001	0.147±0.001
JF tertiary vertex	0.0247±0.0005	0.0249±0.0005
Tagger ($t\bar{t}$)	Rate in data	Rate in MC
SV1	0.189±0.001	0.187±0.001
JF secondary vertex	0.229±0.001	0.217±0.001
JF tertiary vertex	0.0467±0.0006	0.0504±0.0005

Table 1: Reconstruction rates for the SV1 and the JetFitter algorithms described in Section 3.2 for $Z \to \mu^+\mu^-$ + jets-dominated sample (first set) and for $t\bar{t}$-dominated $e\mu$ sample (second set). Only multi-prong vertices have been taken into account for the JetFitter algorithm and the quoted uncertanties are statistical.

3.2.3 c-tagging

The discrimination of c-jets from both b- and light-flavour jets is important in a number of physics studies, including the study of the Higgs boson in the H$\to c\bar{c}$ final states. This discrimination is helped by the topology and kinematics of the secondary vertex reconstructed by the JetFitter algorithm. A set of twelve variables used by the MV2 charm tagger and DL1 discriminant (Section 4.3) is given in Table 2. These variables exploit the fact that charm hadron decays have a single secondary vertex with intermediate charged decay multiplicity and a decay distance comparable to that of b-hadrons. The charm tagging variables are built by using exclusively the jets where only one displaced vertex is reconstructed by JetFitter. They include the number, invariant mass and energy of the particles associated to the secondary vertex as well as their rapidity, computed with respect to the jet axis and the vector defined between the primary and secondary vertex. The track rapidity distribution has already been used by the experiments at LEP [28] for discriminating heavy flavour decay products from primary particle tracks. Given the smaller decay multiplicity of charm hadrons compared to b-hadrons, the average energy per c-hadron
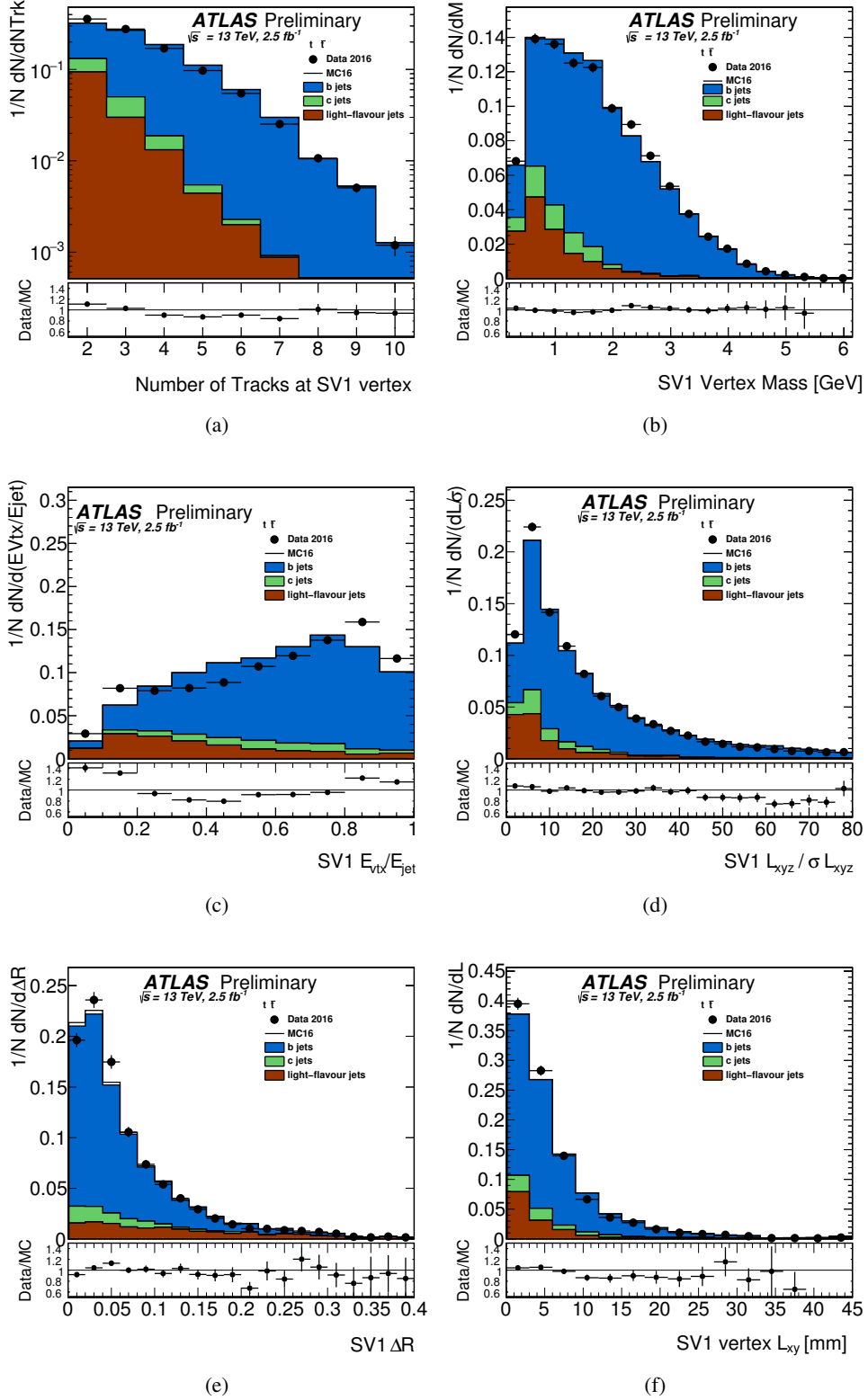
Figure 6: Data-MC comparisons for SV1 vertex variables: the number of tracks associated with a SV1 secondary vertex (a), the secondary vertex mass (b), the secondary vertex energy fraction (c) the significance of three dimensional decay distance (d), the ΔR between the PV to secondary vertex axis and the jet axis (e) and the secondary vertex decay distance in the transverse plane (f) for the $t\bar{t}$-dominated $e\mu$ sample.
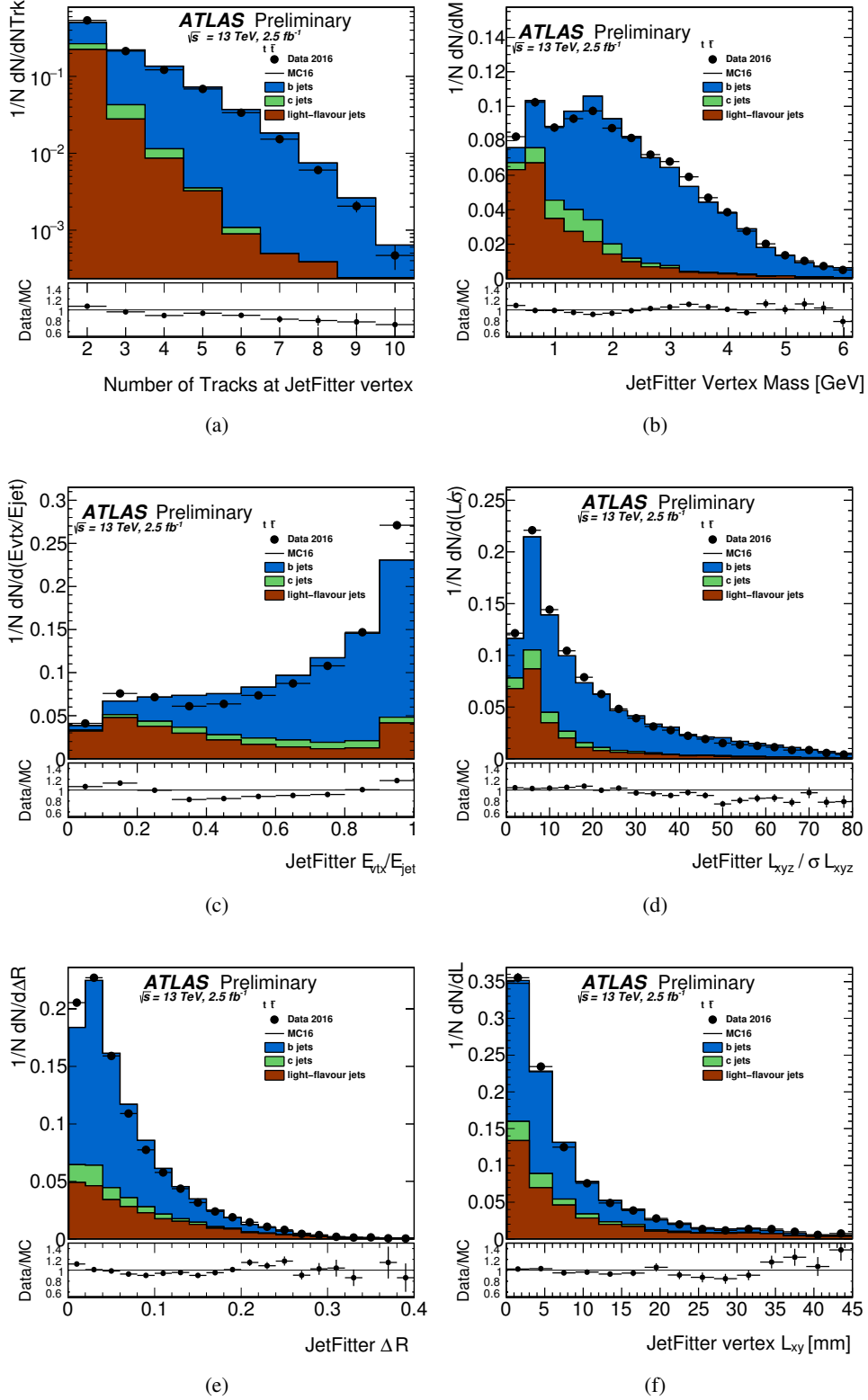
Figure 7: Data-MC comparisons for JetFitter variables: the number of tracks associated with a JetFitter secondary vertex (a), the secondary vertex mass (b), the secondary vertex energy fraction (c) the significance of the three dimensional decay distance (d), the ΔR between the reconstructed B-hadron flight direction and the jet axis (e) and the secondary vertex decay distance in the transverse plane (f) for the $t\bar{t}$-dominated $e\mu$ sample.

decay products is larger than that for b decay products, resulting in a larger rapidity [29]. This feature allows the charm and b-hadron decay products to be separated. The charm tagging variables are used as additional inputs in two variants of the MV2 algorithm, MV2c100 and MV2cl100, and in DL1 as described in Sections 5.1 and 4.3. The list of variables used by the MV2 algorithm for b-tagging can be found in Ref. [4].

Variable Name	Description
L_{xyz}	Three-dimensional displacement of secondary vertex from the primary vertex
L_{xy}	Transverse displacement of the secondary vertex
$Y_{\mathrm{trk}}^{\min}, Y_{\mathrm{trk}}^{\max}, Y_{\mathrm{trk}}^{\mathrm{avg}}$	Min, Max and Avg. track rapidity of tracks in jet
$Y_{\mathrm{trk}}^{\min}, Y_{\mathrm{trk}}^{\max}, Y_{\mathrm{trk}}^{\mathrm{avg}}$ (2$^{\mathrm{nd}}$ vtx)	Min, Max and Avg. track rapidity of tracks at secondary vertex
m	Invariant mass of tracks associated to secondary vertex
E	Energy of charged tracks associated to secondary vertex
f_{E}	Energy fraction of charged tracks (from all tracks in the jet) associated to secondary vertex
N_{trk}	Number of tracks associated to the secondary vertex

Table 2: Summary of new input variables, used for charm tagging, in addition to those used by the baseline MV2 b-tagging algorithm [4]. The track pseudo-rapidity Y_{trk} is computed as $Y_{\mathrm{trk}} \equiv -\log\left(\tan\left(\frac{\theta}{2}\right)\right)$, where θ is the angle between the track momentum and the hadron flight path reconstructed by JetFitter.

Several of the charm tagging variables are shown in Figure 8 for b-, c- and light-flavour jets. The response in data and MC simulation for the most discriminating charm tagging inputs, namely the invariant mass of the tracks, the number of tracks at the vertex, the energy fraction and the three-dimensional displacement of the secondary vertex is shown in Figure 9, where the observed agreement reproduces the pattern also observed for JetFitter and SV1.

3.3 Soft Muon Tagger Algorithm

The Soft Muon Tagger (SMT) is based on the reconstruction of muons coming from semileptonic decays of heavy-flavour hadrons. An earlier version of the SMT tagger was based on a simple cut on a χ^2 distribution [30]. These muons usually have a sizable transverse momentum (though smaller than the typical p_{T} of leptons from electroweak bosons decays, hence the label "Soft"), as well as a large transverse momentum relative to the jet axis, $p_{\mathrm{T}}^{\mathrm{rel}}$. The presence of a muon is enhanced in b-jets with respect to c- and light-flavour jets due to the significant semileptonic decay branching ratio of b hadrons ($BR(b \to \mu \nu X) \approx 11\%$) and c-hadrons produced by the b-hadron decay (sequential semileptonic decay, $BR(b \to c \to \mu \nu X) \approx 10\%$). SMT, while intrinsically limited by the semileptonic branching ratio, provides a useful complement to impact parameter- and vertex-based taggers for jets where their performance is affected by the semileptonic decay, which reduces the number of charged decay particles, their invariant mass and energy.

The development of the new SMT is based on calorimetric jets and combined muons, i.e. muon candidates with matching tracks reconstructed in the ID MS systems. Muons are associated to the closest selected jet by requiring an angular separation between the muon and the jet-axis of $\Delta R < 0.4$. Muons are required to have $p_{\mathrm{T}} > 5$ GeV, $d_0 < 4$ mm and $|\eta| < 2.5$. Since minimum ionising particles lose on average ~ 3 GeV in the ATLAS calorimeter system, the efficiency for reconstructing muons with p_{T} below this value is low. The fraction of b-jets in the $t\bar{t}$ simulated sample described in Section 2 with a reconstructed muon passing
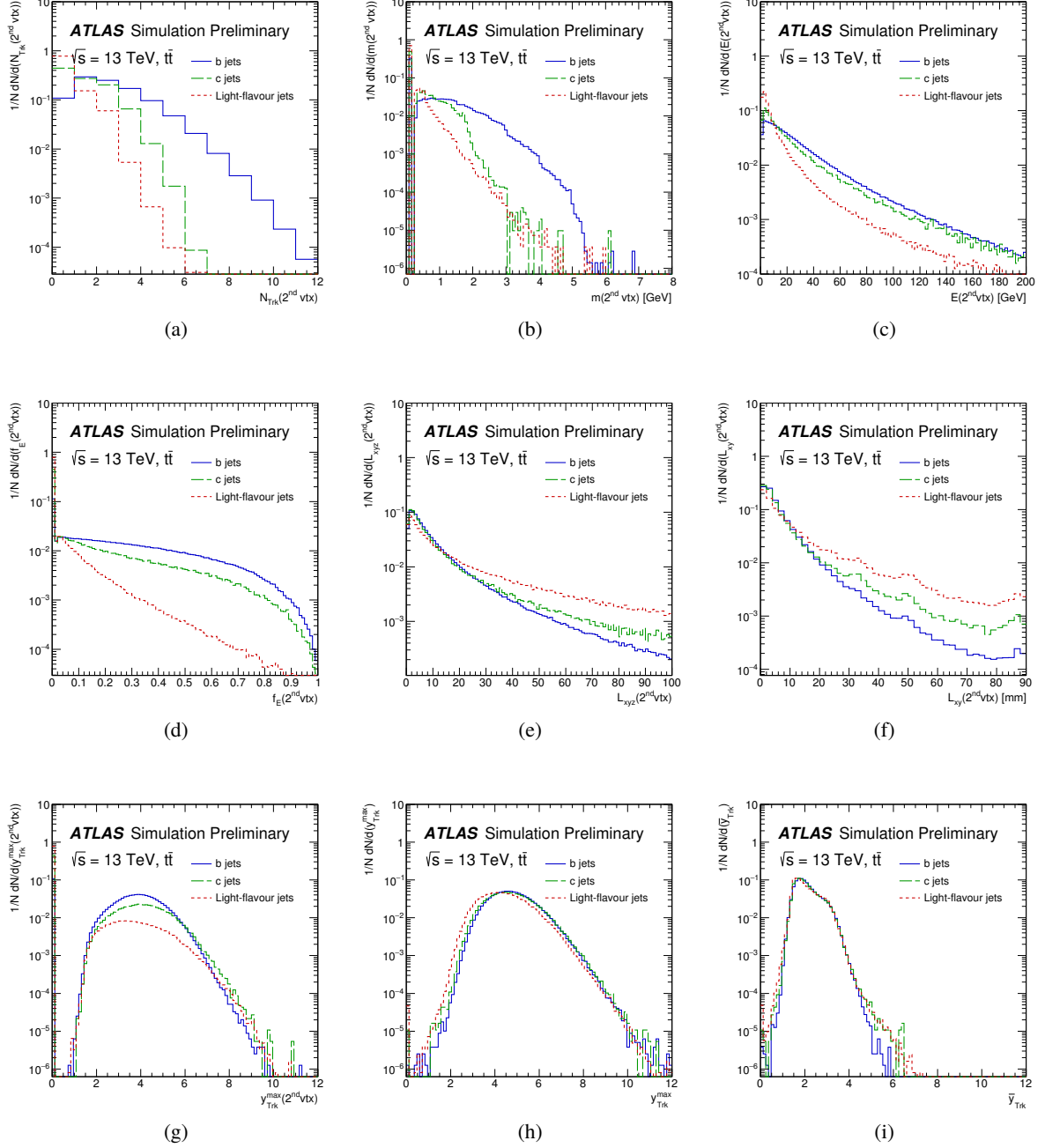
Figure 8: Distributions of input charm tagging variables based on the JetFitter decay reconstruction for *b*-, *c*- and light-flavour jets for a $t\bar{t}$ simulated sample. Definitions are given in Table 2.
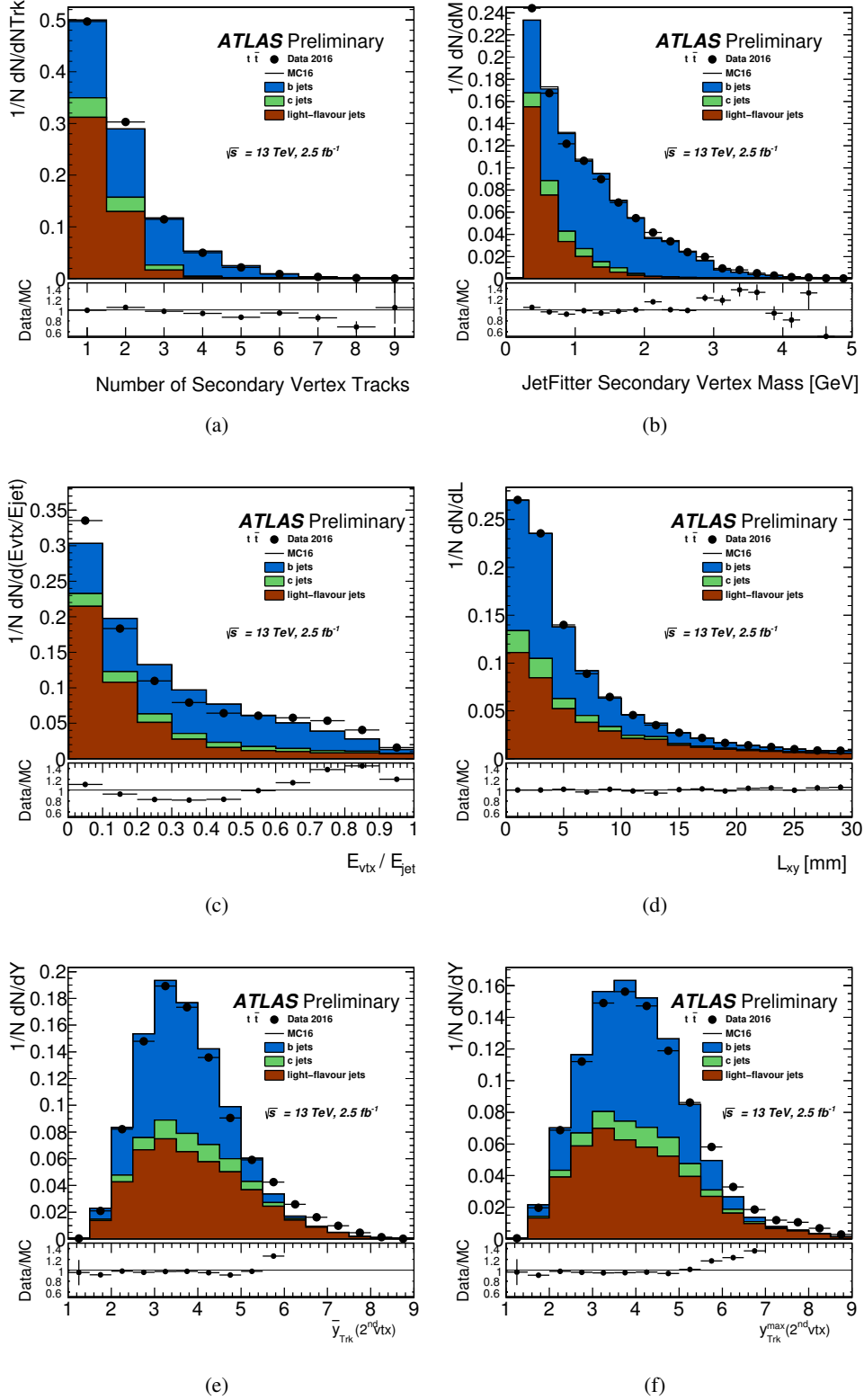
Figure 9: Data-MC comparisons for charm variables: the number of tracks (a), the invariant mass of the tracks (b) associated to the secondary vertex, the energy fraction of charged tracks (c) associated to the secondary vertex w.r.t. the energy, the transverse decay distance (d), the average (e) and maximum (f) rapidity of tracks associated to the secondary vertex using the $t\bar{t}$-dominated $e\mu$ sample.

the requirements listed above is ~12%.

There are three main background sources in light jets that give rise to muon candidates passing these requirements and they are evaluated in the $t\bar{t}$ simulated sample:

- prompt muons from the nearby W boson randomly associated to light jets ($\sim 1\%$ contamination);

- muons coming from the decay in flight of light hadrons, mostly pions and kaons ($\sim 1\%$ contamination);

- energetic hadrons ("punch-through") that travel through the calorimeter system and reach the MS ($< 0.1\%$ contamination).

Three variables separating these muons from those from b- or c- hadron decays are used: ΔR, d_0 and p_T^{rel}. The p_T^{rel} value of decay products is correlated to the parent particle mass. Therefore, muons from direct b decays tend to have a larger p_T^{rel} value than those from background. Muons from the cascade $b \to c$ decay are more difficult to identify since they tend to be softer in the p_T^{rel} spectrum. Muons from π and K decays in flight have tracks that do not in general extrapolate close to the primary vertex. Most of these muons do not represent a background source. Nevertheless, if the decay occurs at a small angle, the decaying hadron and the muon can be reconstructed as a single track and the muon candidate can pass the track selection. Figure 10 shows these discriminating variables for b-, c- and light-flavour jets in simulated $t\bar{t}$ events.
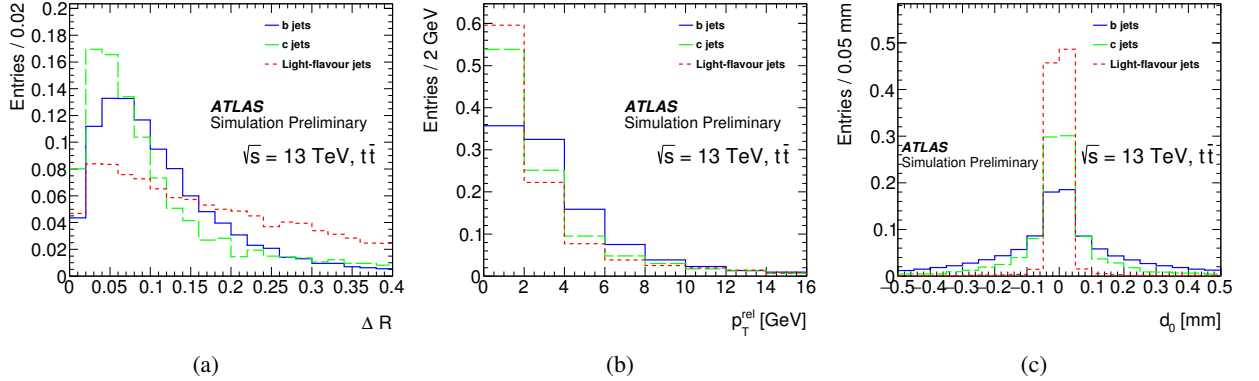
(a) (b) (c)

Figure 10: ΔR (a), p_T^{rel} (b) and d_0 (c) normalised distributions for reconstructed muons associated to b-jets (blue), c-jets (green) and light-flavour jets (red).

In addition to ΔR, d_0 and p_T^{rel}, the SMT algorithm makes use of three observables defining the quality of the muon track. The scattering neighbour significance ($\mathcal{S}$) is a variable introduced to quantify the significance of a kink along the track, a change in trajectory expected in the presence of a pion or kaon decaying to a muon. This quantity is computed by considering pairs of adjacent hits along the track (treating them as neighbouring scattering centers) and evaluating the significance of the angular difference $\Delta\phi$ between the two half tracks ending/starting at each of them, summing significances up over the whole track times the particle charge (q):

$$\mathcal{S} = q \times \sum_i \frac{\Delta\phi_{\text{scat}}^i}{\sigma_{\Delta\phi_{\text{scat}}^i}}. \tag{2}$$

The scattering neighbour significance can be considered as a measurement of how many hits in the tracker do not fit well with the reconstructed track. A higher significance is more likely to correspond to a π or K decay in flight. The momentum imbalance significance ($\mathcal{M}$) is defined as follows:

$$\mathcal{M} = \frac{p_{\text{ID}} - p_{\text{MS}}^{\text{extr}}}{\sigma_{E_{\text{loss}}}}, \tag{3}$$

where p_{ID} is the muon momentum measured by the Inner Detector, $p_{\text{MS}}^{\text{extr}}$ is the momentum measured by the MS and extrapolated to the vertex and $\sigma_{E_{\text{loss}}}$ is the uncertainty on the energy loss measured by the calorimeters. The momentum imbalance significance is useful to reject K and π decays in flight. Another variable sensitive to muons originating from these decays through the p_{ID} to p_{MS} comparison is:

$$\mathcal{R} = \frac{(q/p)_{\text{ID}}}{(q/p)_{\text{MS}}}, \tag{4}$$

where $(q/p)_{\text{ID}}$ is the charge-to-momentum-ratio, i.e. track curvature, measured by the Inner Detector and $(q/p)_{\text{MS}}$ is the same measured by the MS. The three track quality variables are shown in Figure 11.

In quantitative terms, the muon reconstruction efficiency inside jets is ~65% for b-jets containing a muon, with a light-jet mis-identification probability of ~1.8%, evaluated in $t\bar{t}$ events. In the SMT tagging the full set of six variables are used as inputs in a dedicated MVA, using a BDT. Figure 12 shows the new discriminant output for b-, c- and light-flavour jets containing a muon candidate passing the SMT selection. By selecting on the SMT discriminant (BDT>-0.15), the efficiency for accepting a jet having a candidate SMT muon is 85% for b-jets and 15% for light-flavoured jets in $t\bar{t}$ events; for a b-jet tagging efficiency of 10%, the mistag rate is 0.02%.
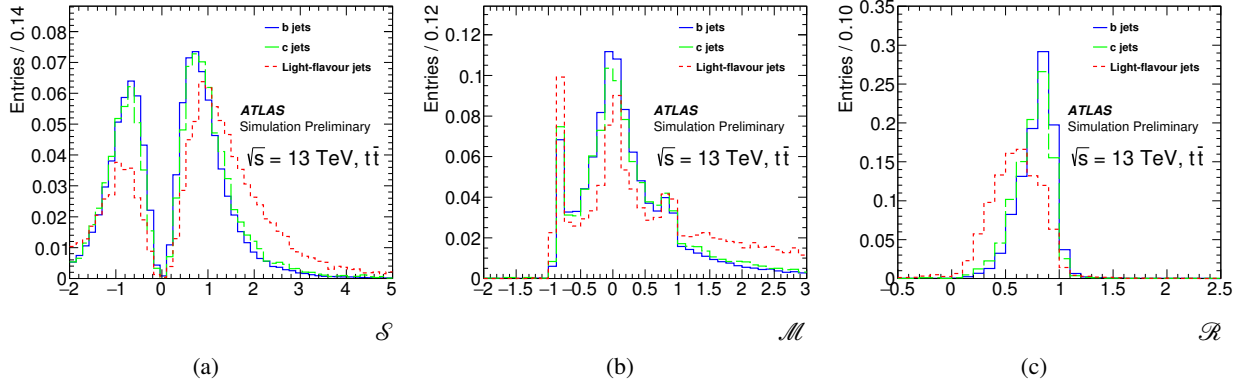

Figure 11: Scattering neighbour significance (a), momentum imbalance significance (b) and q/p ratio (c) normalised distributions for reconstructed muons associated to b-jets (blue), c-jets (green) and light-flavour jets (red). The long tails in $\mathcal{S}$ and $\mathcal{M}$ as well as the low values in $\mathcal{R}$ for muons associated to light-jets are mostly due to in-flight decays; the resolution on $\sigma_{E_{\text{loss}}}$ for low p_{T} muons causes the spikes in $\mathcal{M}$.

Figure 13 shows data/MC comparisons for some of the variables used in the SMT discriminant and the BDT output on $t\bar{t}$ events; good agreement between data and MC simulation is observed in the SMT input variables, $p_{\text{T}}^{\text{rel}}$ and ΔR as well as on the final BDT discriminant. Some localised discrepancies between
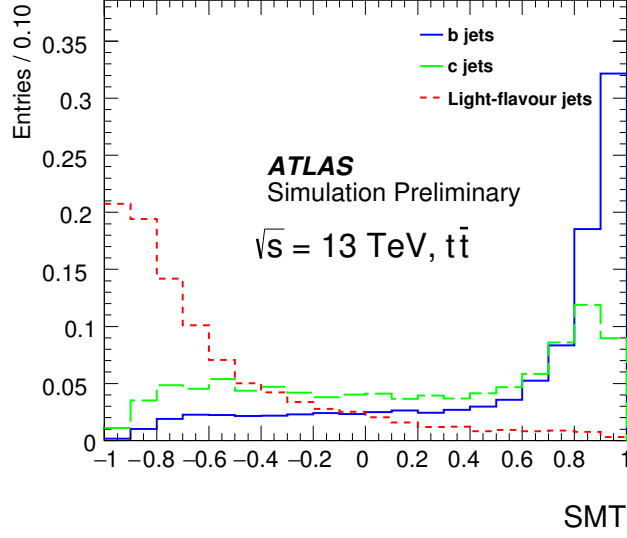
Figure 12: Normalised BDT response in simulated $t\bar{t}$ events of SMT for reconstructed muons associated to b-jets (blue), c-jets (green) and light-flavour jets (red).

data and MC are found in the low BDT region (BDT<-0.3) where the observed difference is of the order of 20-25%.

4 High-level taggers

The high-level taggers combine the inputs from several of the low-level taggers using a multivariate classifier to maximise the b-tagging performance. In the 2017 configuration ATLAS has adopted two variants of high-level taggers. The first is an evolution of the Boosted Decision Tree (BDT) discriminant already used in the past (MV2), and the second is a new tagger (DL1) based on Deep Learning NN.

4.1 Algorithm training samples

The flavour tagging performance depends on the number of charged decay products from b- and c-hadrons, their separation from the event primary vertex and the fraction of the jet energy that they carry. The discrimination of charged decay products improves with increasing heavy hadron energy while their fraction in the jet decreases with increasing jet energy.

Previous versions of the ATLAS multivariate-based taggers were trained using $t\bar{t}$ samples [5]. The jet p_{T} distribution of this sample is steeply falling, making the number of events with jets above 250 GeV insufficient for an effective training. This resulted in sub-optimal performance in the medium-to-high p_{T} range. A broad Z' sample, described in Section 2, enhances the jet population in the kinematic region not accessible when training on a $t\bar{t}$ sample. Figure 14 shows the p_{T} distributions for b- and light-flavour jets in the $t\bar{t}$ and Z' samples.
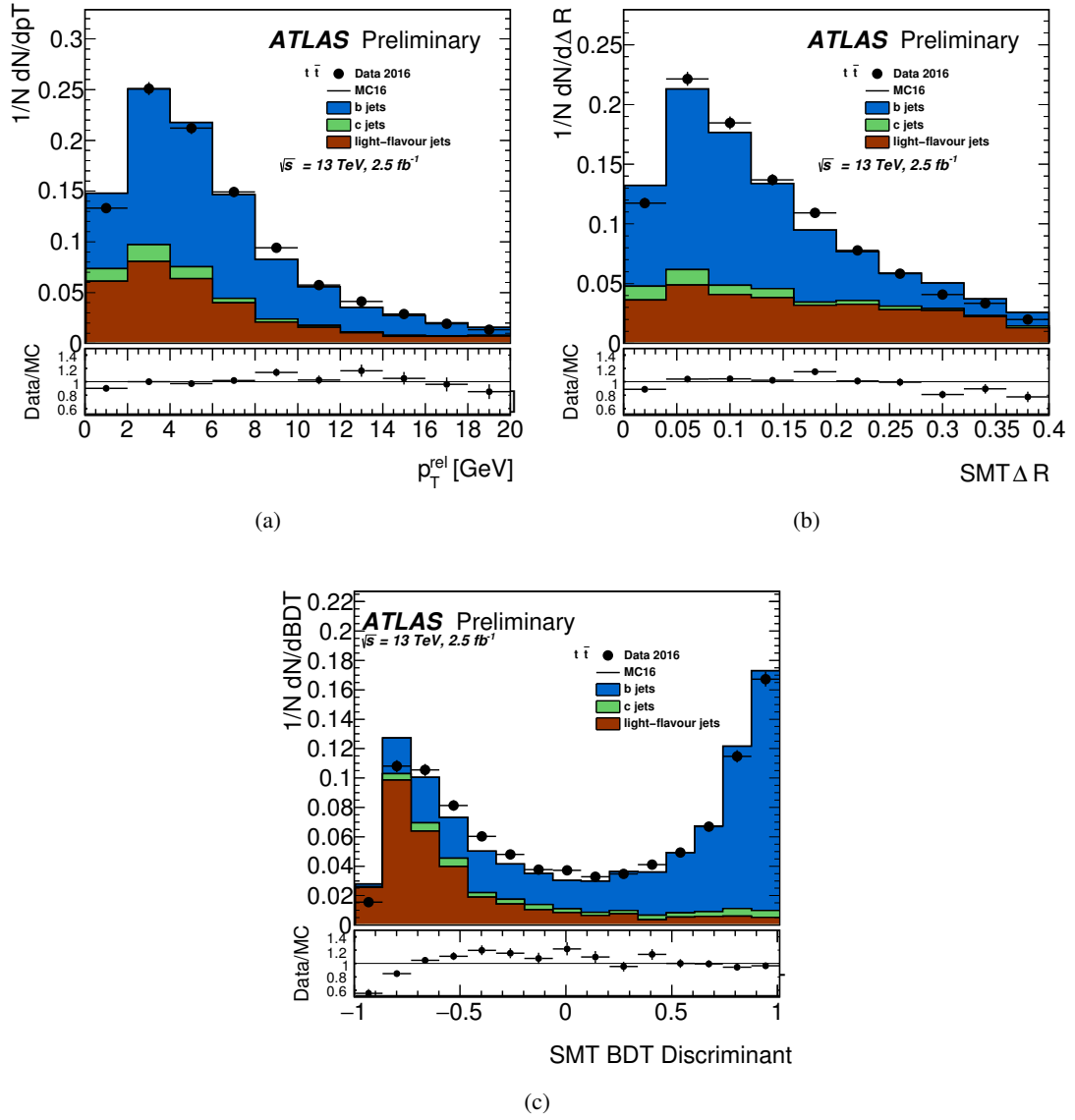
(a)

(b)

(c)

Figure 13: Data-MC comparisons of p_T^{rel} (a), ΔR (b) and the SMT BDT output (c) for $t\bar{t}$-dominated $e\mu$ sample.
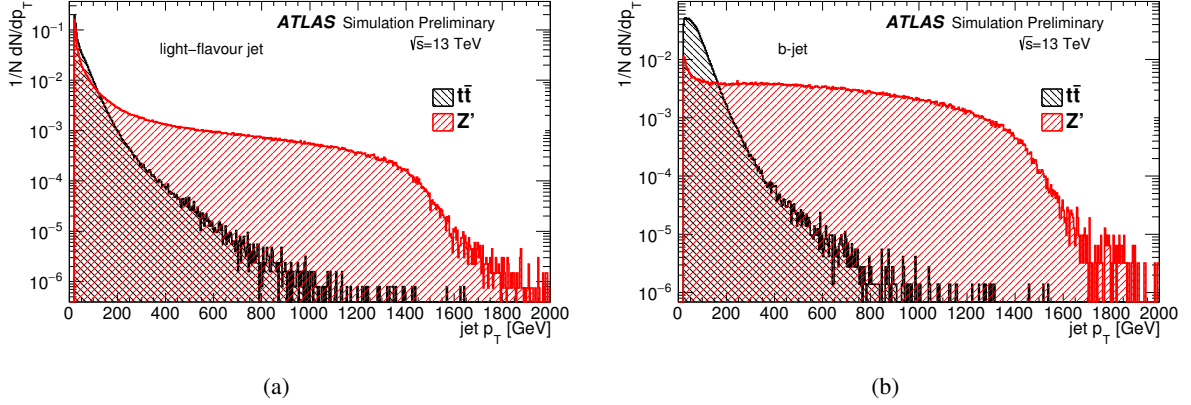
Figure 14: Distribution of the jet transverse momentum of the Z' sample compared to that of $t\bar{t}$ events for light-flavour (a) and b-jets (b).

The relation between the p_T and energy of the jet to those of the original b-quark and hadron is key for understanding the flavour tagging response as a function of the jet p_T. Two different regimes can arise, as shown in Figure 15. In $t\bar{t}$ events the b-jets originate from a relatively low-mass (m_t) state. This results in small p_T transferred to the heavy b-hadron. This leads to a correlation between jet p_T and heavy hadron p_T for $p_\mathrm{T} \lesssim m_t$. For jet $p_\mathrm{T} \gtrsim m_t$, the jet transverse momentum is determined by nearby hadronic activity unrelated to the heavy hadron and the correlation is therefore reduced. Instead, b-hadrons produced in the broad Z' decays with large p_T yield a high degree of correlation with the jet p_T. It is important to ensure that the training of the flavour taggers covers both of these regimes.
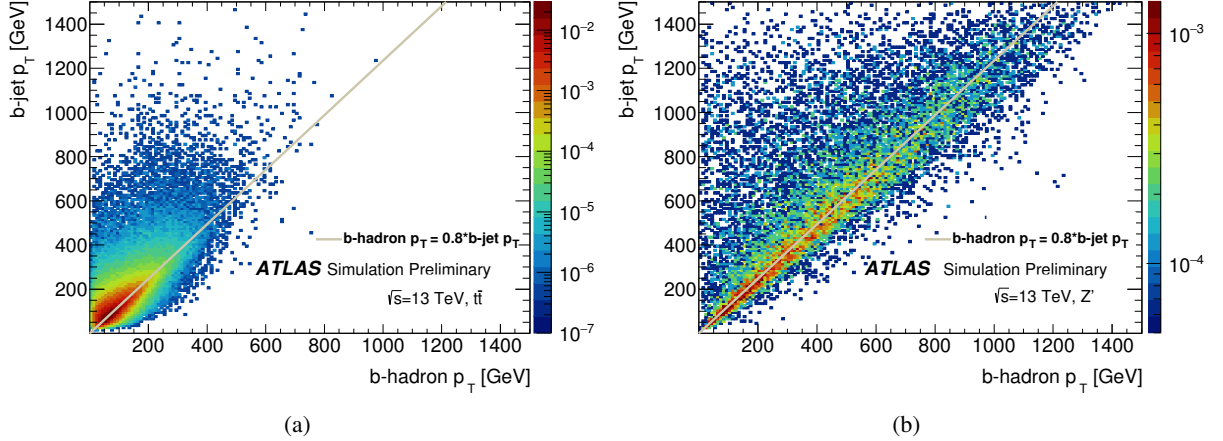

Figure 15: Two-dimensional correlations between the b-hadron and the b-jet transverse momenta for $t\bar{t}$ (a) and Z' (b) events. The continuous line represents the most probable value of the b quark fragmentation function used in the simulation that controls the relation between the b quark and the b hadron energy.

A new training strategy has been employed using a sample made both of $t\bar{t}$ events, to characterise the low p_T region, and broad Z' events, to probe the high p_T regime. This new sample, referred to as the

hybrid sample in the following, is obtained by including *b*-jets from $t\bar{t}$ if the corresponding *b*-hadron p_T <250 GeV and from the Z' sample if the *b*-hadron p_T >250 GeV. For *c*- and light-flavour jets, the same mixing strategy is applied, moving from $t\bar{t}$ events to Z' events for values of the jet p_T above 250 GeV. The full statistics of the $t\bar{t}$ (5M events) and the Z' (3M events) samples is used for the training of the multivariate algorithms. Figure 16 shows the resulting p_T distribution of *b*-jets in this training sample.
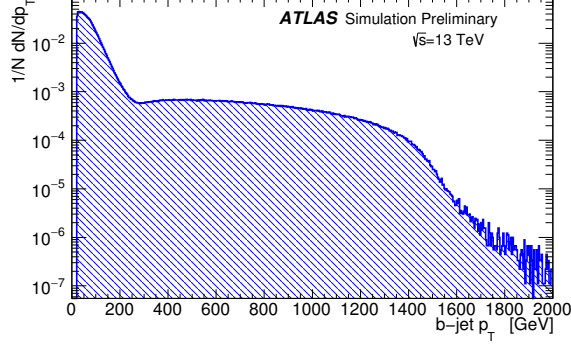
Figure 16: *b*-jet p_T distribution of the hybrid sample used for the training of the multivariate-based taggers.

4.2 MV2 tagger

The first high-level tagger is a BDT discriminant that combines the output of the low-level taggers described in Section 3. The BDT algorithm is trained using the ROOT Toolkit for Multivariate Data Analysis (TMVA) [31] on the hybrid sample. *b*-jets are considered as signal, *c*- and light-flavour jets as background. The performance is evaluated separately on the $t\bar{t}$ and Z' samples. The list of input variables also includes the p_T and η of the jets, as they give useful information in interpreting the separation power of the variables from the low-level taggers.

The *b*- and *c*-jets are reweighted in p_T and η to match the spectrum of the light-flavour jets. This procedure avoids differences in the kinematic distributions of signal and background that can be interpreted as discriminating by the training. The complete list of variables used for the training, as well as the features of the algorithm and the optimization of the training parameters are discussed in Ref. [5]. The *c*-jet fraction in the training is set to 7% and that of light-flavoured jets to 93%. These values achieve a suitable balance between light-flavour and *c*-jet rejection.

Given the availability of new low-level taggers, several variants of the MV2 taggers have been developed:

- a reference option, based on the standard impact parameter (IP2D and IP3D) and secondary vertex-based (SV1 and JetFitter) input variables is adopted for the 2016 data-analysis in Ref. [5] (MV2);

- a new option, including in addition the SMT (MV2Mu);

- a full option with the SMT and the RNNIP inputs in addition to the standard variables (MV2MuRnn).

The use of the full set of input variables from the SMT algorithm and the output of the SMT multivariate BDT discriminant have both been tested in the MV2 training. The MV2 BDT configuration is significantly simplified by using a single output variable from SMT instead of six inputs. Furthermore, it is found that

this approach leads to better performance for light-flavour and c-jet rejection than when using the full set of SMT variables. The effect of the finite number of events available for the MC samples for the MV2 training is estimated by performing cross-training tests where independent sub-samples are used as input to the BDT classifier. The tagging performance for light-flavour and c-rejection is observed to be stable at the level of a few %, and the statistical uncertainty on the values of the typical light-flavour and c-jet rejections on the evaluation sample is found to be negligible. The performance of this b-tagging algorithm will be discussed in Section 5.

4.3 DL1 tagger

The second high level tagger (DL1) is based on an Artificial Deep Neural Network. DL1 is trained using KERAS with the THEANO [32] backend and the ADAM optimizer [33]. The DL1 NN has a multidimensional output corresponding to the probabilities for a jet to be a b-, c- or light-flavour jets. Its topology consists of a mixture of fully-connected hidden layers and maxout layers [34].

The input variables to DL1 consist of those used for the MV2 BDT, with the full set of SMT observables replacing the SMT BDT output, and the addition of the JetFitter c-tagging variables, discussed in Section 3.2.3.

The parameters of DL1 are optimised, including the architecture of the NN, the number of training epochs, the learning rates and training batch sizes. Batch Normalisation [35] is added by default since it is found to improve the performance. Training with multiple output nodes offers additional flexibility when constructing the final cut discriminant by combining the b, c and light probability, as done for the RNNIP tagger without the presence of the τ node (see Equation 1). By treating all flavours equally during training, the trained network can be used for both b-jet and c-jet tagging. Therefore, the final determination of the c- and b-jets fraction can be chosen a posteriori according to the desired performance. Iso-efficiency curves of the DL1 discriminant for c- and b-tagging showing the light-flavour jet rejection vs c- and b-jet rejection for different charm fractions in the training are shown in Figure 17.
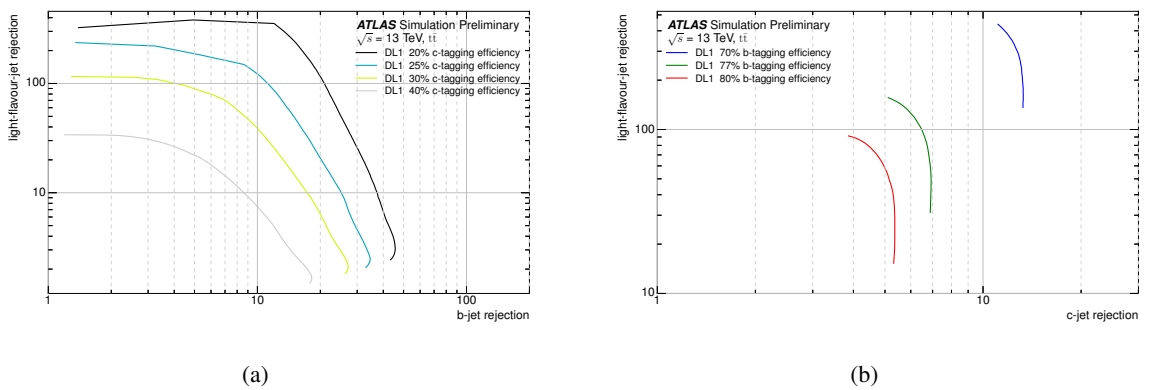

Figure 17: DL1 light-flavour vs b-jet rejection (a) and c-jet rejection (b) for a set of working points corresponding to different values of the c-jet and b-jet efficiency, respectively. The evaluation is performed on $t\bar{t}$ events.

The use of a deep NN is motivated by the interest in exploring whether this architecture might be better

suited to exploit input correlations than a BDT. This is of particular interest for a more performant c-jet discrimination w.r.t. light-flavour jets. While the performance of the DL1 and MV2 discriminants for tagging b-jets is found to be very similar and these algorithms are found to tag a highly correlated sample of b-jets when training on the same input variables, DL1 does provide specific advantages in terms of possible future R&D. In addition, DL1 is amenable to new training procedures using Adversarial Networks that could help mitigate the effects of modelling systematics at training time [36]. Finally, while DL1 and RNNIP are currently trained separately, the RNN could in principle be connected directly into the DL1 and the entire system trained at once in an end-to-end fashion.
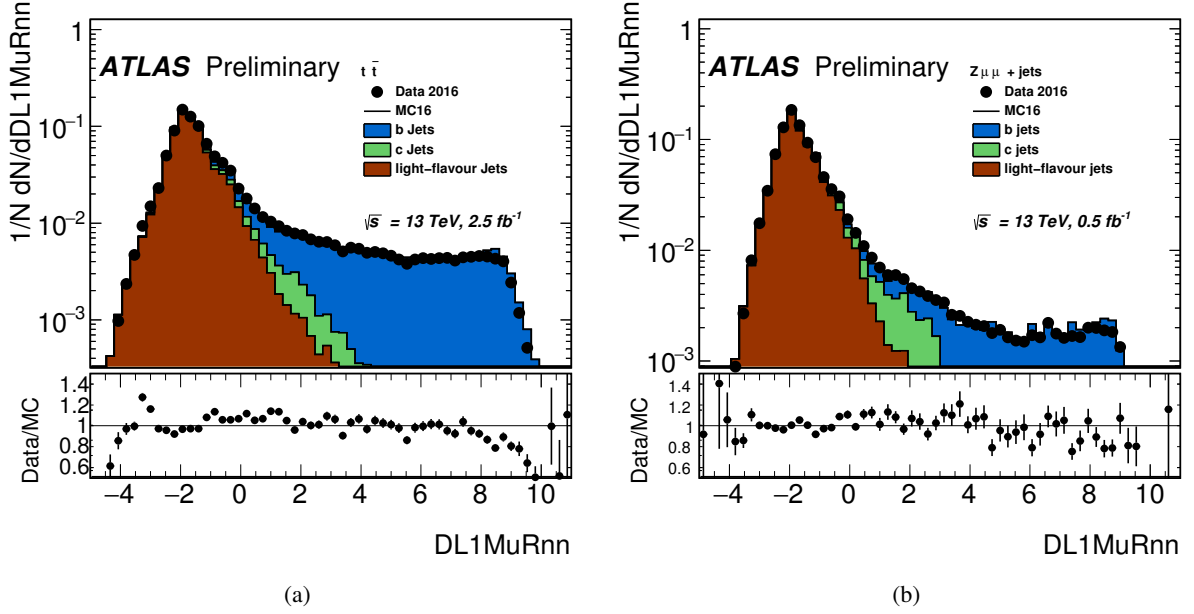

(a)   (b)

Figure 18: Data-MC comparison of the DL1MuRnn output discriminant for b-, c- and light-flavour jets using selected $t\bar{t}$-dominated $e\mu$ sample and $Z \to \mu^+\mu^-$ + jets-dominated sample.

Figure 18 shows the response of the DL1MuRnn discriminant for the $t\bar{t}$-dominated $e\mu$ sample and $Z \to \mu^+\mu^-$+jets-dominated sample. This discriminant is built by assigning the same c-fraction as in the MV2 training. The final c-jet fraction assignment will be tuned to best satisfy the analysis needs in terms of rejecting charm jets versus light-flavour jets. The good separation of the highly-enriched b-jet sample is noticeable in the distributions and the simulation describes the response of data within 20% with some localised differences for low and high values of the DL1MuRnn discriminant especially in the $t\bar{t}$-dominated $e\mu$ sample.

5 b-tagging algorithm performance

The performance of MV2 trained on the hybrid sample (Section 4.1) is evaluated on $t\bar{t}$ and Z' events. Figure 19 shows the light-flavour rejection as a function of the b-jet efficiency evaluated on $t\bar{t}$ and Z' simulation when performing the training of the MV2 algorithm with $t\bar{t}$ events alone or with the hybrid configuration. The latter achieves similar performance in the $t\bar{t}$ sample where the jet p_T is typically below

250 GeV but outperforms it by a factor ~2.2 at 77% b-jet efficiency in the Z' sample where the jet p_T is typically larger than 250 GeV. Similar results are observed for the c-jet rejection.

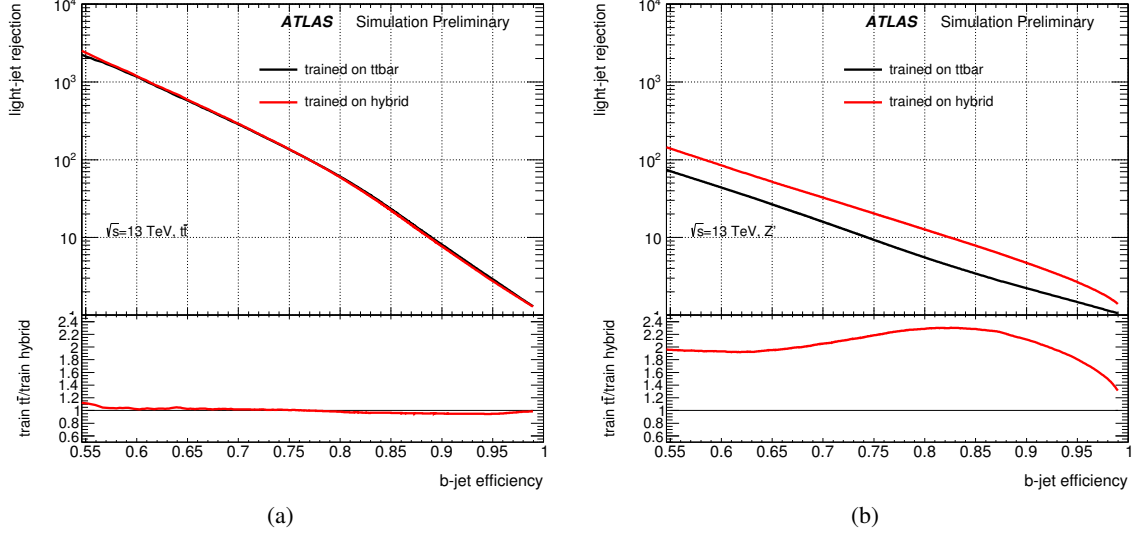(a)   (b)

Figure 19: Light-flavour jet rejection as a function of b-jet efficiency using the MV2 b-tagging algorithm when comparing the nominal $t\bar{t}$ training (black line) with a training using the hybrid configuration (red line), evaluated on the $t\bar{t}$ (a) and the Z' (b) samples. The ratio reported on the bottom of the figure is calculated between the $t\bar{t}$-based training and the hybrid-based configuration.

The performance of the different variants of MV2 (MV2, MV2Mu and MV2MuRnn) trained on the hybrid sample and evaluated on the $t\bar{t}$ and Z' samples are shown in Figures 20-23. The light-flavour and charm rejection are evaluted as a function of the b-jet efficiency in Figures 20 and 22, and the light-flavour and c- jet rejections as a function of the jet p_T for a constant b-jet efficiency of 77% are given in Figures 21 and 23.

Figure 23 also shows the performance of the baseline MV2 tagger in the 2016 configuration. At low jet p_T, the changes in the tracking between the 2016 and the 2017 configurations, as described in Sections 1 and 3.1, lead to a degradation of the MV2 light-jet rejection performance of around 20%. In the higher-p_T region, the tracking effects are more than compensated by the tagger response improvements from the new training strategy. In addition, the MV2 response on data/MC for the 2016 and 2017 configuration shows an improved modelling for the new simulation setup as shown in Figure 24 for the Z + jets dominated sample.

The inclusion of the SMT BDT output improves the light-jet rejection in the low-medium p_T (below 1 TeV) range where the additional information of the presence of the muon from the semileptonic decay in jets is beneficial to enhance the b-to-light jet separation, while this improvement fades away at higher p_T values. The b-to-c separation does not profit from the muon-in-jet information at efficiency value larger that ~ 50% since leptons are produced in b- and c decays alike as shown in Figure 23 (b). An effect of the b-to-c separation is observed at an efficiency comparable to the muon in-jet efficiency, ~ 15%.

The MV2MuRnn setup provides further improvements in light and c-jet rejection, in particular in the high-p_T region where the use of the RNNIP tagger is most effective. For example, for jet p_T = 1 TeV, the improvement in performance brought by the use of RNNIP as a function of jet p_T is approximately a factor
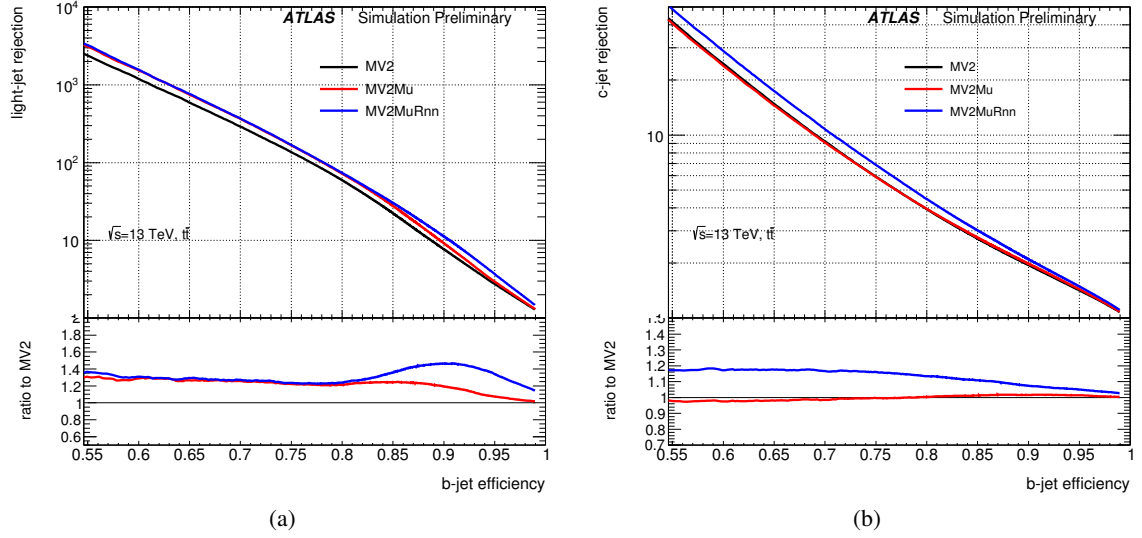
Figure 20: Light-flavour and *c*-jet rejection as a function of *b*-jet efficiency for MV2 (black line), MV2Mu (red line), MV2MuRnn (blue line). The algorithm evaluation is performed on $t\bar{t}$ events. The ratio reported on the bottom of the figure is calculated for each MV2 variant (MV2Mu, MV2MuRnn) with respect to MV2.
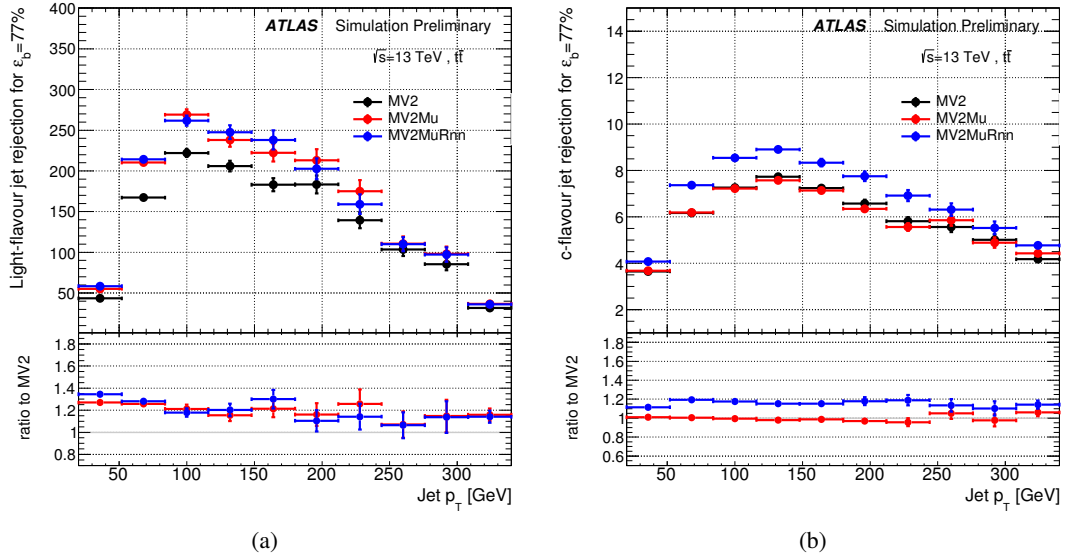

Figure 21: Light-flavour (a) and *c*-jet (b) rejection as a function of the jet transverse momentum for MV2 (black markers), MV2Mu (red markers), MV2MuRnn (blue markers). The algorithm evaluation is performed on $t\bar{t}$ events for a flat *b*-jet efficiency of 77% for each p_T bin. The ratio reported on the bottom of the figure is calculated for each MV2 variant (MV2Mu, MV2MuRnn) with respect to MV2.
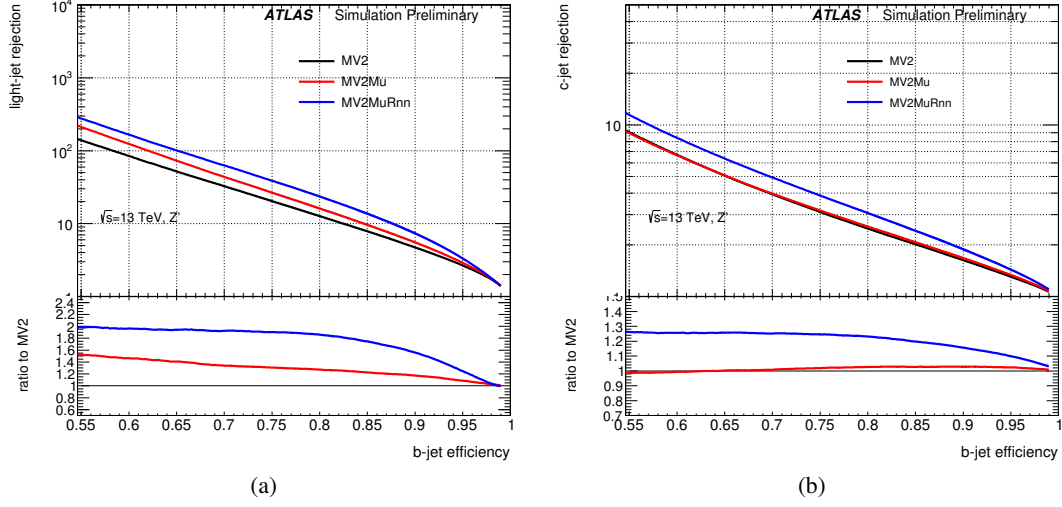
Figure 22: Light-flavour and c-jet rejection as a function of b-jet efficiency for MV2 (black line), MV2Mu (red line), MV2MuRnn (blue line). The algorithm evaluation is performed on Z' events. The ratio reported on the bottom of the figure is calculated for each MV2 variant (MV2Mu, MV2MuRnn) with respect to MV2.
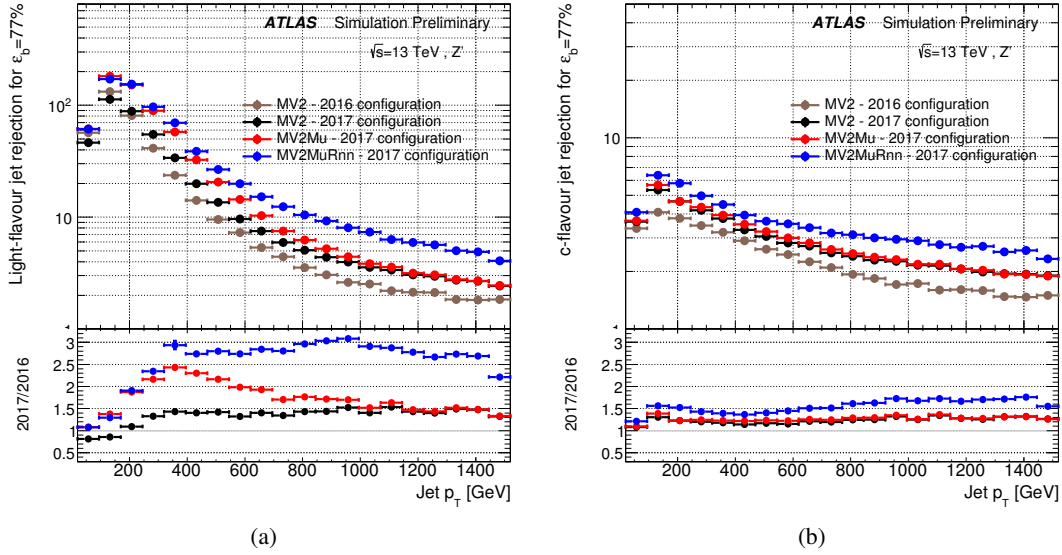

Figure 23: Light-flavour (a) and c-jet (b) rejection as a function of the jet p_T for MV2 in the 2016 configuration (brown markers), MV2 (black markers) in the 2017 configuration, MV2Mu (red markers) and MV2MuRnn (blue markers). The algorithm evaluation is performed on Z' events for a flat b-jet efficiency of 77% for each p_T bin. The ratio is calculated with respect to the 2016 configuration.

28
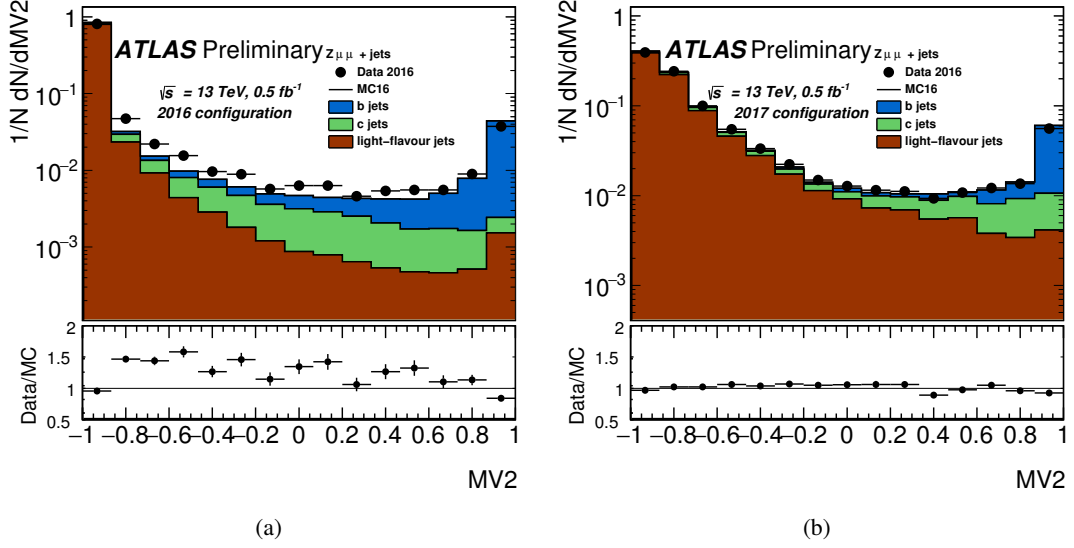
Figure 24: Data-MC comparison of the MV2 output of for the 2016 configuration (a) and the 2017 configuration (b) using $Z \to \mu^+ \mu^-$ + jets-dominated sample.

1.5 in light-flavour jet rejection and $\sim$ 1.3 for c-jet rejection compared to the MV2 2017 configuration when evaluating on Z' events.

Data-MC comparisons of the new MV2-based algorithms for the $t\bar{t}$ dominated sample and the Z + jets dominated sample are shown in Figure 25. Consistently good agreement is observed over the full range of the discriminant values and in particular at very low and very high values in the region dominated by light-flavour and b-jets. Discrepancies between the data and the simulation are at most 10%. This is particularly the case in the BDT region relevant for the 77% working point (MV2 >0.54 evaluated on the MV2MuRnn BDT version) where the data-MC agreement is found to be excellent.

MV2MuRnn performs better than the other MV2 BDT tagger algorithms and is used for further studies of flavour tagging performance. Figure 26 shows the distribution of its BDT discriminating output for the signal and the background components. Figure 27 shows the light-flavour and b-jet rejection of the MV2 algorithms for constant b-jet efficiency as a function of $|\eta|$. The decrease in rejection for light-flavour jets is mainly due to the decrease in tracking resolution and the increase in the amount of material in the ID at large $|\eta|$.

Due to the higher pile-up expected in the 2017 LHC run, it is interesting to consider the change in performance as a result of the increasing pile-up in the simulation. Figure 28 shows the change in the light-flavour jet rejection for a constant b-jet efficiency with different pile-up values in the simulation. The decrease in rejection with increasing pile-up is studied by examining the input variables to the MV2 discriminant. These variables do not show significant shape modifications as a function of $<\mu>$ for b-jets. In light-flavour and, to a lesser extent, in c-jets, they are shifted closer to the values obtained for b-jets with increasing pile-up, resulting in a decrease in MV2 performance. The drop in light-jet rejection may be partially mitigated in the future by performing a dedicated training with increased pile-up in the MC

simulation or by adding the pile-up information in the training process.

5.1 c-tagging with MV2 and algorithm performance

The MV2 algorithm has been extended to provide c-tagging functionality by training two additional variants, designed to separate c-jets from the other flavours. These versions differ from the standard MV2 versions by their input variables and by the composition of their training sets. MV2c100 is trained with b-jets as signal, and only c-jets as background, while MV2cl100 has c-jets as signal and light-flavour jets as background. Both variants use the standard input variables of MV2, the additional 12 variables described in section 3.2.3, and the RNNIP variables. The training was performed using the same hybrid sample as the standard MV2 variants. In order to tag a jet, threshold values are defined for the output of both MV2c100 and MV2cl100. This two-dimensional cut allows to choose the working points a posteriori according to desired performance with a range of b- and light-flavour jet rejections for a given c-jet tagging efficiency similarly to the functionality offered by DL1. The standard b-tagging efficiency curve is replaced by an iso-curve of constant charm-tagging efficiency in the b- and light-flavour jet rejection space. Figure 29 shows these curves evaluated on the $t\bar{t}$ and Z' test samples.

The response observed on selected data $e\mu$ + jets events and $t\bar{t}$ simulated events for these two additional MV2 variants are shown in Figure 30 . A comparison of the flavour composition in this region and that obtained by the MV2MuRnn tagger in Figure 25 shows an enrichment in charm in the highest signal-like bin of the discriminant.

6 b-tagging performance for selected physics processes

The use of the hybrid sample extends the range of b-jet p_T to very large values, therefore a large variety of jet and parton configurations can be probed. For this reason, it is interesting to assess the universality of the achieved performance in other samples, namely the dependance of efficiency and background rejection on the specific physics process of interest. This is studied with four physics processes that probe different kinematics and event topologies:

- direct pair production of scalar bottom quarks with $\tilde{b} \rightarrow b \tilde{\chi}_1^0$ ($m_{\tilde{b}}$ = 800 GeV, $m_{\chi_1^0}$=1 GeV),

- direct pair production of scalar top quarks with $\tilde{t} \rightarrow t \tilde{\chi}_1^0$ ($m_{\tilde{t}}$ = 800 GeV, $m_{\chi_1^0}$=1 GeV),

- Z' production with $Z' \rightarrow b\bar{b}$ ($m_{Z'}$ = 2 TeV),

- associated production of the Higgs boson with vector boson (VH) at m_H=125 GeV with $H \rightarrow b\bar{b}$.

The first three samples are dominated by high-p_T b-jets, and mainly probe the typical kinematic range of the broad Z' phase space. The $\tilde{b}$ and Z' decays have rather isolated jets with a large correlation between parton and jet p_T. In these events the fraction of b-jets having a second jet within $\Delta R < 1$ is less than 10%. The $\tilde{t}$ sample also contains high-p_T jets but with more significant nearby hadronic activity highlighted by the fact that more than 15% of b-jets have another jet within ΔR=1. This is comparable to the environment for high-p_T-jets ($p_T > 200$ GeV) in $t\bar{t}$ events. These samples are used to test the stability of the b-tagging efficiency vs jet p_T and possible distortions in the shape of the p_T spectrum of tagged jets. In order to avoid kinematic effects due to forward jets, the study is restricted to jets in the barrel region ($|\eta|$<1).
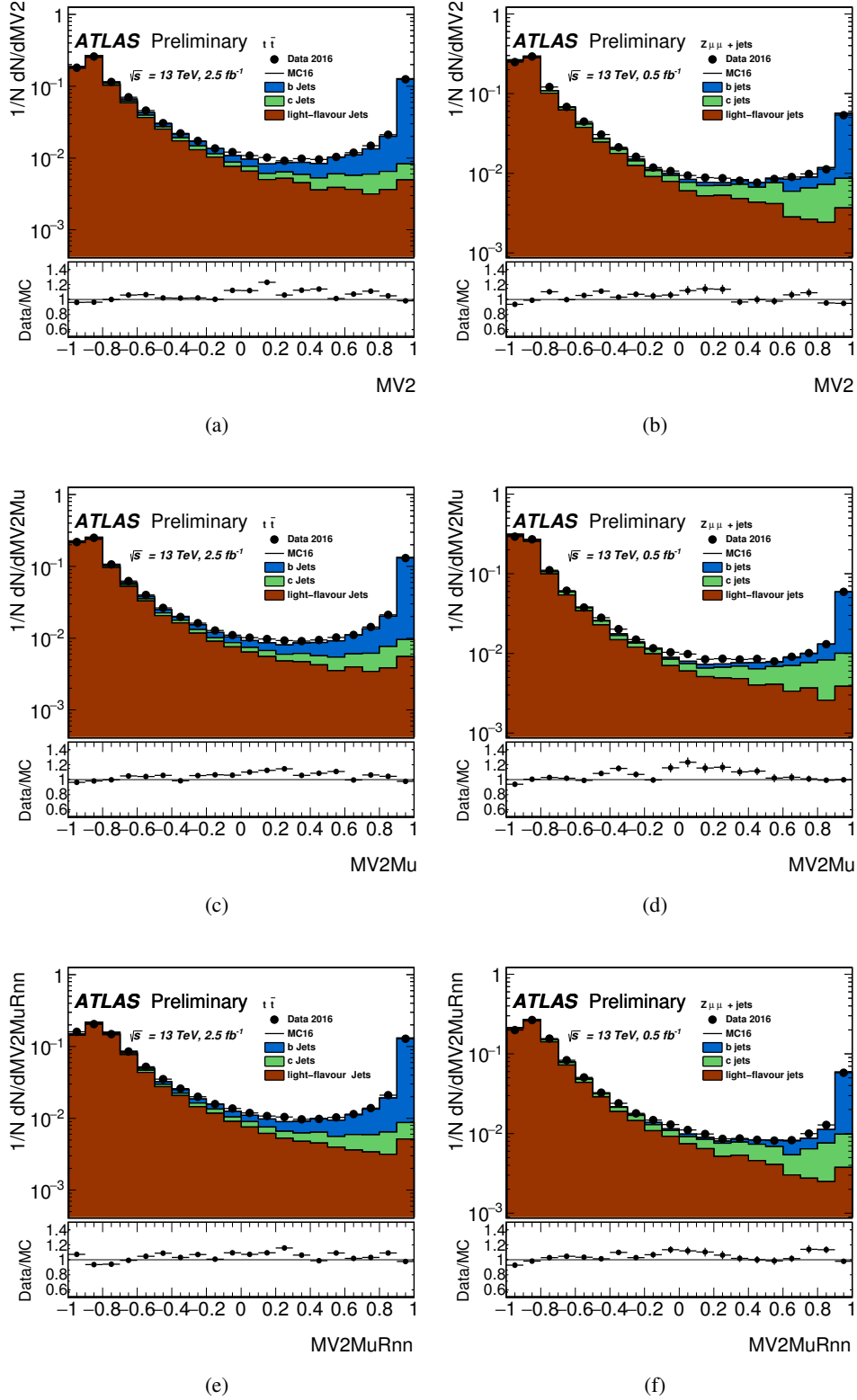
Figure 25: Data-MC comparisons of the MV2 b-tagging algorithms, MV2 (a-b), MV2Mu (c-d), MV2MuRnn (e-f) using the selected $t\bar{t}$-dominated $e\mu$ sample (left) and $Z \to \mu^+\mu^-$ + jets-dominated sample (right).
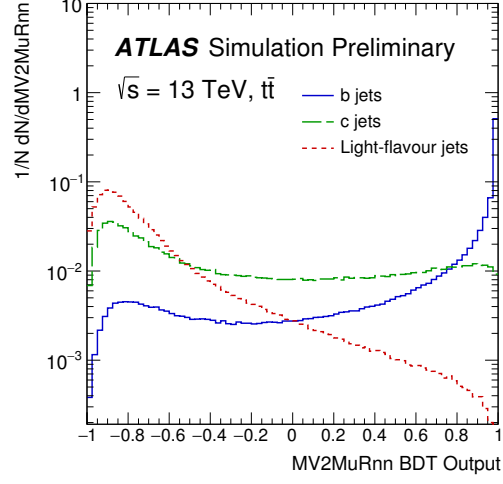
Figure 26: MV2MuRnn BDT output for *b*- (solid blue), *c*- (dashed green) and light-flavour (dotted red) jets evaluated with $t\bar{t}$ events.
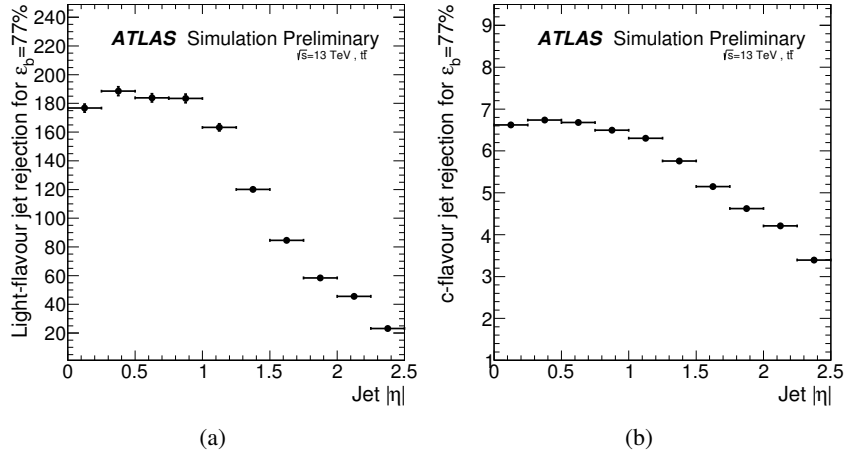

(a)   (b)

Figure 27: MV2MuRnn performance for light-flavour jet rejection (a) and *c*-jet rejection (b) at the 77% *b*-jet efficiency working points as a function of jet $|\eta|$.
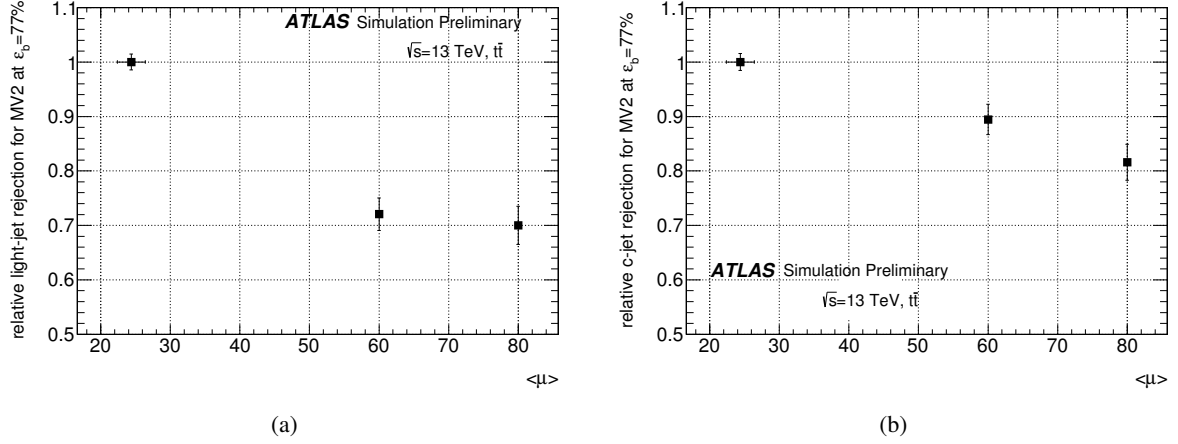
(a)   (b)

Figure 28: Relative rejection degradation using the MV2 discriminant at the 77% b-jet efficiency working point for light-flavour (a) and c-jet (b) as a function of the average number of interactions per bunch-crossing evaluated for $t\bar{t}$ events normalized to the rejection at $< \mu >=24.4$. Vertical error bars show the statistical uncertainty on the relative light-flavour jet rejection while the horizontal error bars show the binning used to sample the distribution. The bin on the left is obtained by selecting in the region of the flat distribution of the pile-up profile, the points at $\mu=60$ and $\mu=80$ are for samples generated at a fixed μ value.
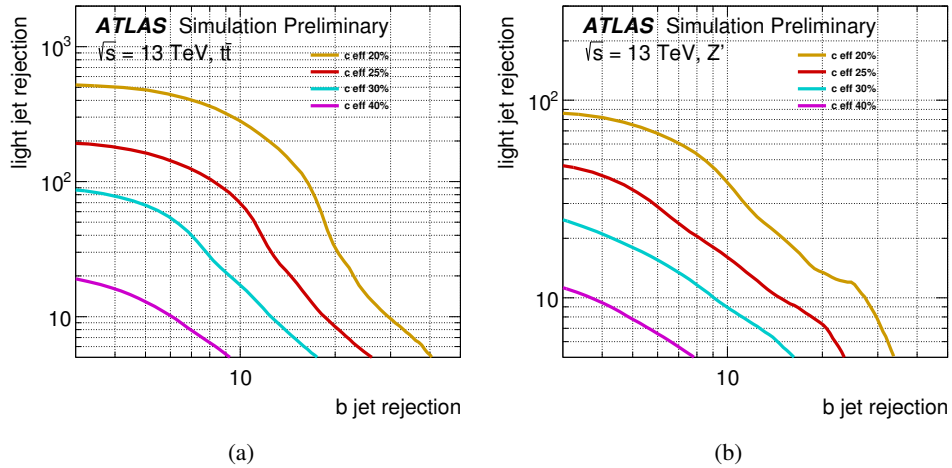

(a)   (b)

Figure 29: c-tagging performance of the MV2 variants MV2c100 and MV2cl100 evaluted on the $t\bar{t}$ (a) and the Z' (b) samples. The yellow, red, blue and purple lines define the c iso-efficiency curves for 20%, 25%, 30% and 40% c-jet efficiency, respectively.
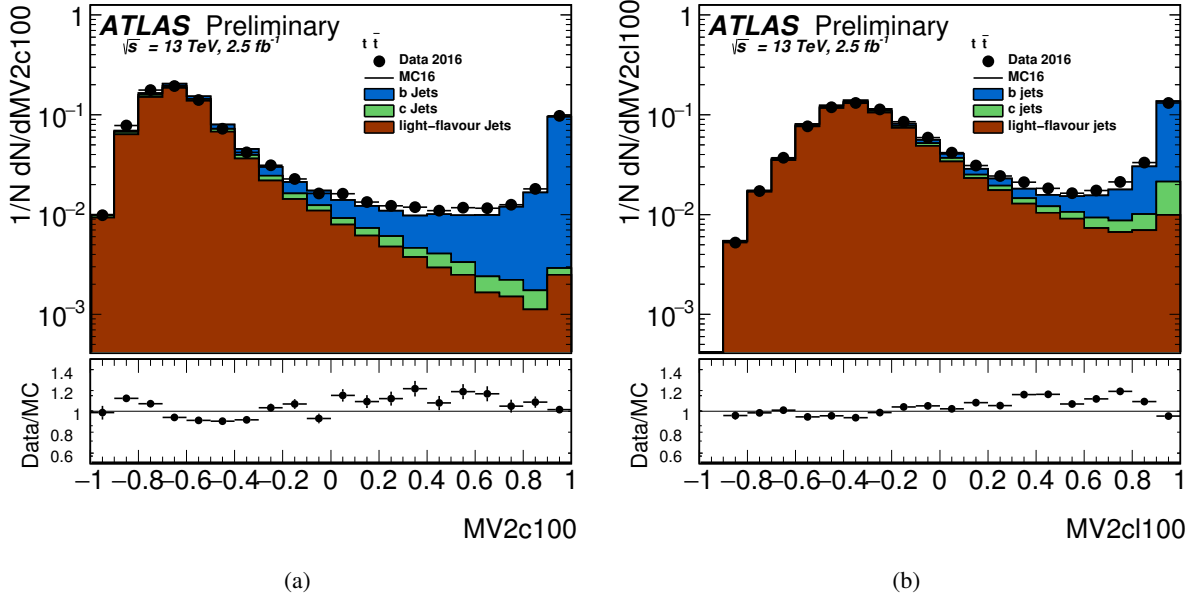
Figure 30: Data-MC comparisons of MV2c100 (a) and MV2cl100 (b) *b*-tagging algorithms using the selected $t\bar{t}$-dominated $e\mu$ sample.

The MV2MuRnn tagger performance, trained on the hybrid sample, is tested by studying the *b*-tagging efficiency on these samples. The MV2MuRnn cut working points yielding 77% *b*-jet efficiency on the hybrid sample is computed for several bins in jet p_T. These MV2MuRnn cuts are applied to the $\tilde{b}$, $\tilde{t}$ and Z' samples and the *b*-tagging efficiency is determined as a function of the jet p_T. It is worth noticing that lower values of *b*-jet efficiency correspond to a stricter working point in MV2 resulting in general in better performance in terms of light-flavour and *c*-jets rejection. Results are shown in Figure 31. At low jet-p_T the hadronic activity in the $\tilde{t}$ sample is comparable to the one of the $t\bar{t}$ sample used for training this phase-space region. The fraction of *b*-jets with another nearby jet within $\Delta R < 1$ is similar and the discrepancy in *b*-jet efficiency is less than 3%. Simulation shows that these nearby extra jets are responsible for the degradation of the tagging efficiency. On the other hand, at low p_T the Z' and the $\tilde{b}$ samples outperform the $t\bar{t}$ samples by around 25% in terms of light jet rejection due to their less dense hadronic environment. At high p_T the effect of hadronic activity is reduced and the results are compatible within 5% with respect to the sample used for training.

The VH process is characterised by relatively low jet p_T, corresponding to the $t\bar{t}$ region of the training. For this process, the flavour tagging performance is studied for both signal efficiency and background rejection. Signal *b*-jets from $H \rightarrow b\bar{b}$ and background events from Z + jets and W + jets without *b*-jets are considered. Events are required to have at least two jets with p_T in excess of 20 GeV, $|\eta| < 2.5$ and fulfilling JVT cuts, with the leading jet required to have $p_T > 45$ GeV. The events are also required to have two opposite-sign leptons with $81 < m_{\ell\ell} < 101$ GeV and significant missing transverse momentum. The jet pair giving invariant mass closest to the Higgs boson mass of 125 GeV, m_{jj}, is chosen. Jet pairs giving $90 < m_{jj} < 150$ GeV are tested with the MV2MuRnn tagger at the working point cut corresponding to 77% efficiency for a jet of the same p_T in the hybrid sample. The ratio of the selected *b*-jets in the VH signal sample to those measured in the hybrid sample is found to be 1.03±0.02 (stat.)
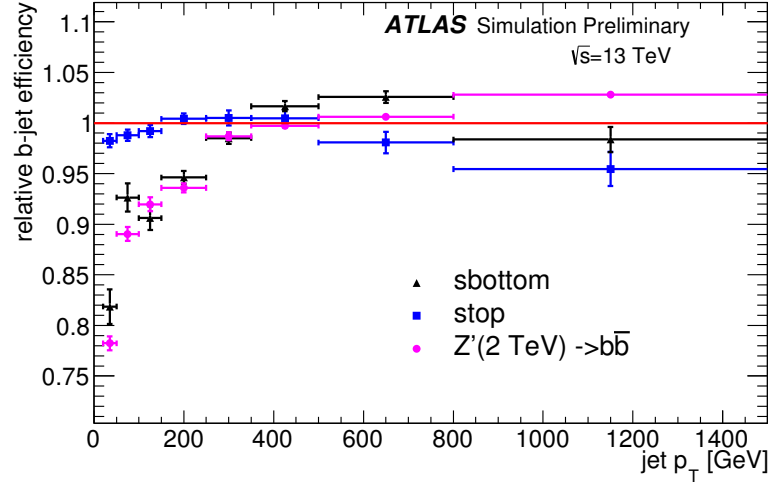
Figure 31: Relative b-jet efficiency using the MV2MuRnn algorithm with respect to the 77% working point extracted from the hybrid sample in bins of jet p_T for the $\tilde{t}$ (blue), $\tilde{b}$ (black) pair production and Z' (purple). The red line corresponds to the 77% b-jet efficiency on the hybrid sample.

inclusively in p_T. The misidentification probability of c- and light-flavour jets fulfilling the same selection described above is determined on the Z + jets and W + jets sample and compared to observed for c- and light-flavour jets in the training hybrid sample. The c- and light-flavour jet misidentification probabilities of those jets of the same p_T in the hybrid sample are found to be 0.87±0.08 (stat.) and 0.88±0.11 (stat).

In summary, the three main parameters influencing the flavour tagging performance are the jet p_T and η and the overlap between the b parton and additional hadronic activity. This additional hadronic activity may be due to a nearby hadronic jet, at low to intermediate jet p_T values, or gluon radiation within the jet, more notable for high p_T jets. The introduction of the hybrid sample in the multivariate tagger training is beneficial in maintaining a stable flavour tagging performance for quark jets across these different kinematical and topological regimes. Concerning light-flavour jets, the gluon-jet contribution at low p_T is not negligible and difficult to disentangle from the light quark jet component. This is particularly important in the training of the Z' sample at low values of jet p_T where the presence of gluon jets is larger than in the $t\bar{t}$ sample.

6.1 Sensitivity to b-hadron production and decay modelling

The effects of the modelling of the b-hadron production and decays on the efficiency of the MV2MuRnn tagger is studied. The b fragmentation function, defined as the fraction of the b quark energy (E_b) transferred to the weakly decaying b hadron (E_B), $x_B = E_B/E_b$, and the charged decay multiplicity, N_{dec}^{ch} is also considered. The b fragmentation function was studied in detail at LEP and SLC due to the favourable kinematics of the $e^+e^- \rightarrow Z \rightarrow b\bar{b}$ process at the resonance peak [37–40]. The LEP experiments measured $< x_B >= 0.7092 \pm 0.0025$ [41]. When b-jets in simulated $t\bar{t}$ events are linearly reweighted to reproduce the $\pm 1\sigma$ uncertainty of the LEP measurement, the corresponding relative change of the b-tag efficiency at 77% working point is $\pm 1\%$. The charged decay multiplicity was studied for B_u^- and B_d^0 mesons by CLEO

at the $\Upsilon(4S)$ [42] resonance and by DELPHI [43]. The measurement at LEP is sensitive to all b-hadron species and finds the charge decay multiplicity = 4.97±0.03±0.06. The uncertainty on this measurement, when propagated to b-jets in $t\bar{t}$ simulation, leads to a relative change of the b-tagging efficiency of 0.4%. The effects of uncertainties in b modelling on the predicted b-jet efficiency are found to be small compared to the typical uncertainties of calibration analyses in flavour-enriched control regions.

7 Conclusions

Improvements and innovations in the low-level taggers combined with new approaches to multivariate analysis and training samples have resulted in more performant flavour-tagging algorithms available for the analysis of the 2017 LHC collision data with ATLAS.

In particular, the new RNNIP tagger exploits information from the track-by-track impact parameter correlations within a jet, providing superior and complementary tagging capability to the simpler impact parameter-based IP2D and IP3D taggers. Soft muon tagging, based on the reconstruction of muons from semileptonic decays of b-hadrons offers additional discriminating power for jets with semileptonic decays. Furthermore, a c-tagging algorithm that is constructed from the topological features of the charm displaced secondary vertex is also available.

A new hybrid training sample including $t\bar{t}$ events and hadronic decays of a broad, heavy Z' state boosts the high-level tagger response at medium-to-high jet p_T. A NN-based algorithm, based on a Deep Learning architecture, has been developed in parallel with the more traditional, BDT-based, MV2 tagging algorithm. It is expected that it will enable a higher degree of integration of low-level and high-level taggers and the adoption of new training procedures able to mitigate the impact of modelling systematics.

In addition to the improvements in the algorithms employed, the optimisation of the flavour tagging response has included detailed comparisons of 2016 collision data and MC simulation, from track variables to low-level tagger observables and high-level tagger response for event samples enriched and depleted in b jets. It is found that the changes in the tracking simulation made available for the 2017 configuration have improved the data/MC agreement in the b-tagging input and output distributions.

The stability of the baseline MV2 tagger efficiency on physics samples of different kinematics, $\tilde{t}$ and $\tilde{b}$ pair production, $VH\ H \rightarrow b\bar{b}$ and $Z' \rightarrow b\bar{b}$ ($m_{Z'}$ = 2 TeV), has been verified. Finally, the sensitivity of the MV2 tagger to pile-up conditions and to the modelling of b production and decay have been characterised.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. **C70** (2010) 787.

[3] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013. ATLAS-TDR-19, URL: https://cds.cern.ch/record/1291633.

[4] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, URL: https://cds.cern.ch/record/2037697.

[5] ATLAS Collaboration, *Optimization of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, URL: http://cds.cern.ch/record/2160731.

[6] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, CERN-LHCC-96-041, URL: https://cds.cern.ch/record/331061.

[7] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*, CERN-LHCC-96-042, URL: https://cds.cern.ch/record/331062.

[8] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*, CERN-LHCC-97-022, URL: https://cds.cern.ch/record/331068.

[9] H. Bichsel, *Straggling in thin silicon detectors*, Rev.Mod.Phys **60** (1988) 663.

[10] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2015-018, URL: https://cds.cern.ch/record/2037683.

[11] ATLAS Collaboration, *Identification of jets containing b-hadrons with Recurrent Neural Network at the ATLAS experiment*, ATLAS-PHYS-PUB-2017-003, URL: http://cds.cern.ch/record/2255226.

[12] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

[13] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852.

[14] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024.

[15] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. **D89** (2014) 033009, arXiv: 1302.6246 [hep-ph].

[16] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026.

[17] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, (2011), ATL-PHYS-PUB-2014-021, URL: http://cdsweb.cern.ch/record/1966419.

[18] A. Martin, W. Stirling, R. Thorne and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189.

[19] T. Gleisberg et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189.

[20] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[21] S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[22] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063.

[23] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, URL: https://cds.cern.ch/record/1700870.

[24] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[25] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*,
JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex].

[26] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*,
Studies in Computational Intelligence 385, Springer, 2012.

[27] ATLAS Collaboration,
Secondary vertex finding for jet flavour identification with the ATLAS detector,
ATL-PHYS-PUB-2017-011, URL: https://cds.cern.ch/record/2270366.

[28] J. Abdallah et al., *b tagging in DELPHI at LEP*, Eur. Phys. J. **C32** (2004) 185,
arXiv: hep-ex/0311003 [hep-ex].

[29] P. Abreu et al., *Measurement of the charged particle multiplicity of weakly decaying B hadrons*,
Phys. Lett. **B425** (1998) 399.

[30] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*,
JINST **11** (2016) P04008.

[31] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040,
arXiv: physics/0703039 [PHYSICS].

[32] Theano Development Team,
Theano: A Python framework for fast computation of mathematical expressions,
arXiv - 1605.02688 (2016), URL: http://arxiv.org/abs/1605.02688.

[33] D.P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv - 1412.6980 (2014),
URL: http://arxiv.org/abs/1412.6980.

[34] I.J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio, *Maxout Networks*,
arXiv - 1302.4389 (2013), URL: http://arxiv.org/abs/1302.4389.

[35] S. Ioffe; C. Szegedy,
Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,
arXiv -1502.03167 (2015), URL: http://arxiv.org/abs/1502.03167.

[36] G. Louppe, M. Kagan and K. Cranmer, *Learning to Pivot with Adversarial Networks*, (2016),
arXiv: 1611.01046 [stat.ME].

[37] A. Heister et al., *Study of the fragmentation of b quarks into B mesons at the Z peak*,
Phys. Lett. **B512** (2001) 30, arXiv: hep-ex/0106051 [hep-ex].

[38] G. Abbiendi et al., *Inclusive analysis of the b quark fragmentation function in Z decays at LEP*,
Eur. Phys. J. **C29** (2003) 463, arXiv: hep-ex/0210031 [hep-ex].

[39] K. Abe et al., *Measurement of the b quark fragmentation function in Z0 decays*,
Phys. Rev. **D65** (2002) 092006, [Erratum: Phys. Rev.D66,079905(2002)],
arXiv: hep-ex/0202031 [hep-ex].

[40] J. Abdallah et al., *A study of the b-quark fragmentation function with the DELPHI detector at LEP
I and an averaged distribution obtained at the Z Pole*, Eur. Phys. J. **C71** (2011) 1557,
arXiv: 1102.4748 [hep-ex].

[41] P. Abreu et al., *A Precise measurement of the partial decay width ratio R(b)**0 = Gamma(b
anti-b) / Gamma(had)*, Eur. Phys. J. **C10** (1999) 415.

[42] G. Brandenburg et al., *Charged track multiplicity in B meson decay*,
Phys. Rev. **D61** (2000) 072002, arXiv: hep-ex/9907057 [hep-ex].

[43] P. Abreu et al., *Measurement of the charged particle multiplicity of weakly decaying B hadrons*, Phys. Lett. **B425** (1998) 399.