PROCEEDINGS OF SCIENCE

Measurement of the b-tagging efficiency using multi-jet events in ATLAS.

Dr. Valentina Vecchio *[a]

[a] *Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester, United Kingdom*

E-mail: valentina.vecchio@cern.ch

The identification of jets containing b-hadrons, b-tagging, plays an important role in many physics analyses in ATLAS. Several different machine learning algorithms have been deployed for the purpose of b-tagging. These tagging algorithms are trained using Monte-Carlo simulation samples, as such their performance in data must be measured. The b-tagging efficiencies have been measured in data using $t\bar{t}$ events in the past and this work presents the measurements in multijet events using data collected by the ATLAS detector at $\sqrt{s}$ = 13 TeV. This offers several key advantages over the $t\bar{t}$ based calibrations, including a higher precision at low jet p_T and the ability to perform measurements of ε_b at significantly higher jet p_T. Two approaches are applied and for both a profile likelihood fit is performed to extract the number of b-jets in samples passing and failing a given b-tagging requirement. The b-jets yields are then used to determine ε_b in data and from that scale factors to the efficiency measured in the Monte-Carlo. The two approaches differ primarily in the discriminating variable used in the fit. At low jet p_T the variable muon p_T^{rel} is used, while for high jet p_T the signed impact parameter significance is used. Both calibrations give measurements of the scale factors as a function of the jet p_T.

The Tenth Annual Conference on Large Hadron Collider Physics - LHCP2022
16-20 May 2022
online

*on behalf of the ATLAS Collaboration

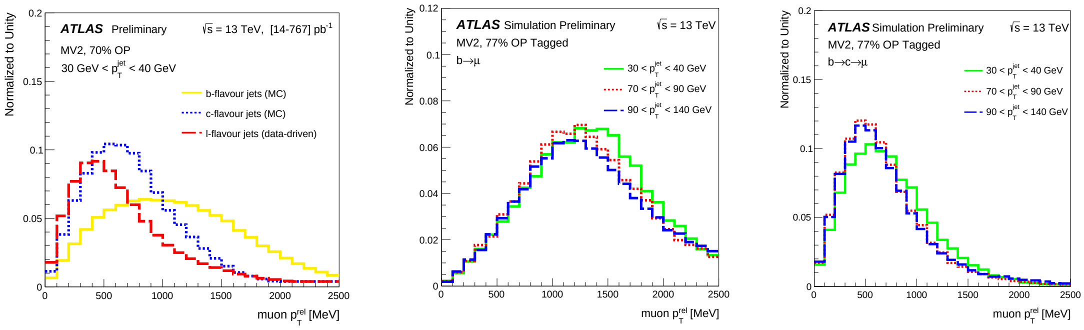
PoS(LHCP2022)323

1. Introduction

The identification of jets containing b-hadrons, b-tagging, is an essential tool for the measurements of the Higgs Boson properties [1–3], precision measurements in the top-quark sector [4] and in searches for physics beyond the Standard Model [5]. The flavour-tagging algorithms Ref. [8] are trained using Monte-Carlo (MC) simulation samples, therefore the estimation of their tagging efficiency in data, hereafter referred to as *calibration*, plays a critical role to have a thorough understanding of their performance. The outcome of a calibration measurement is the ratio between the b-tagging efficiency (ε_b) measured in data and that obtained using simulated events, the scale factor (SF). In ATLAS, the baseline b-tagging efficiency measurement relies on $t\bar{t}$ events [8]. Despite its robustness, the $t\bar{t}$ based calibration [8] presents some limitations. In the low jet p_T ($p_{\mathrm{T}}^{\mathrm{jet}}$) range its uncertainty is dominated by the systematic uncertainty contribution while in the high $p_{\mathrm{T}}^{\mathrm{jet}}$ spectrum it is limited by the low available statistics. In this work two calibration methods are presented, both relying on multijet events. These multijet-based calibrations are subject to different sources of uncertainties and are more precise, particularly in the low p_T range. Two multijet calibrations are presented in this work both using the full Run 2 data set of proton-proton collisions at $\sqrt{s}$ = 13 TeV.In both cases events with at least one jet spatially matched to muon are selected. To select data events the first calibration relies on prescaled muon-in-jet triggers while the second calibration uses a non-prescaled single-jet trigger. In both calibrations a profile likelihood fit is performed to extract the number of b-jets in samples passing and failing a given b-tagging requirement. The b-jets yields are then used to determine ε_b in data and from that scale factors. For a more detailed description of the two calibrations please refer to Ref. [10] and Ref. [11]

2. Muon p_T^{rel} calibration

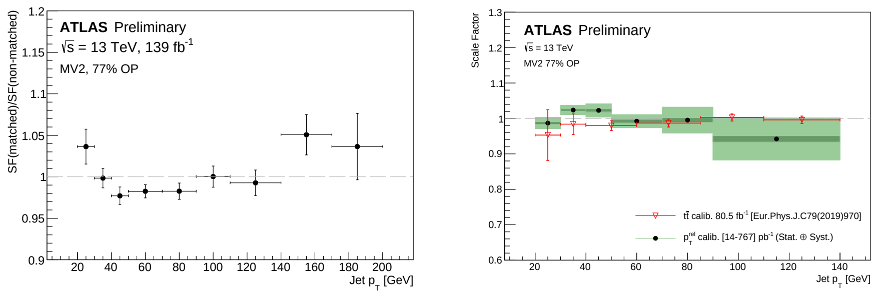
Given the larger masses of b-hadrons relative to hadrons containing only charm or light quarks, their decay products have larger momenta in the rest frame of the decaying hadron, referred to as p^*. The momentum of the muon projected onto the plane that is transverse to the combined muon and jet axis, referred to as p_T^{rel}Figure 1a, is correlated with its p^* and it is a quantity that can be measured in the laboratory frame. The p_T^{rel} calibration method [10] exploits the discrimination power of the muon p_T^{rel} distribution to estimate the b-tagging efficiency in data. Beyond the optimised range ($p_{\mathrm{T}}^{\mathrm{jet}}$ = [20-140] GeV) the muon p_T^{rel} distribution for b-jets becomes similar to that of c-jets and *light*-jets, making this method not applicable. The dependency of the muon p_T^{rel} from the $p_{\mathrm{T}}^{\mathrm{jet}}$ is due to the different kinematics of muons from a direct b-hadron decay, $b \to \mu\nu + X$, or a cascade decay, $b \to c \to \mu\nu + X$. As shown in Figure 1b and 1c, respectively, the muon p_T^{rel} in a direct decay is harder than the cascade decay case. The fraction of cascade decays muons increases with the $p_{\mathrm{T}}^{\mathrm{jet}}$ as they are more likely to pass the selections threshold. Several effects play a role in shaping the muon p_T^{rel} distribution and they are accounted for as systematic uncertainties. The biggest source of uncertainty in the low $p_{\mathrm{T}}^{\mathrm{jet}}$ range up to 50 GeV is the modelling of the muon p^*. This systematic is evaluated by comparing the muon p^* of the ATLAS MC with the unfolded measurement produced by the DELPHI experiment [12]. The reason behind the big impact of the muon p^* modelling is in the kinematical differences between the phase-space explored by the DELPHI experiment, which probes muons and b-hadrons softer than those probed with the ATLAS detector. In the high $p_{\mathrm{T}}^{\mathrm{jet}}$



(a) (b) (c)

Figure 1: Muon p_T^{rel} spectra for the b-, c- and light-jet in yellow, blue and red, respectively. The jets have a p_T =[30-40] GeV and are b-tagged at the 70% OP of the MV2 tagger (a). Muon p_T^{rel} spectra for muons from a direct decay (b) and cascade decay (c) for jets in different $p_{\mathrm{T}}^{\mathrm{jet}}$ ranges passing the 77% OP of the MV2 tagger [10].

regime the main source of uncertainty stems from the modelling of $g \to b\bar{b}$ enclosed in small jet. These events smear the jet axis and hence result in a much harder muon p_T^{rel}. The $g \to b\bar{b}$ increases with the $p_{\mathrm{T}}^{\mathrm{jet}}$ and it is observed to be at least 2.5 times higher in data than the ATLAS MC prediction. Finally, the third most important systematic uncertainty is the one associated with the extrapolation of the b-tagging SF from the semileptonic b-hadrons sample used in the p_T^{rel} calibration to the inclusive one. A sample of $t\bar{t}$ events is used to derive ε_b SF in the muon-in-jet case and the case of no muons matched to jet. The ratio of the two SFs is shown in Figure 2a as a function of the $p_{\mathrm{T}}^{\mathrm{jet}}$. The central value is used to correct the SF derived with the muon p_T^{rel} method and the uncertainty is summed in quadrature. The final SFs are shown in Figure 2b in comparison with the SF derived by the $t\bar{t}$ based calibration. The muon p_T^{rel} b-tagging efficiency SFs are more precise in the $p_{\mathrm{T}}^{\mathrm{jet}}$ range



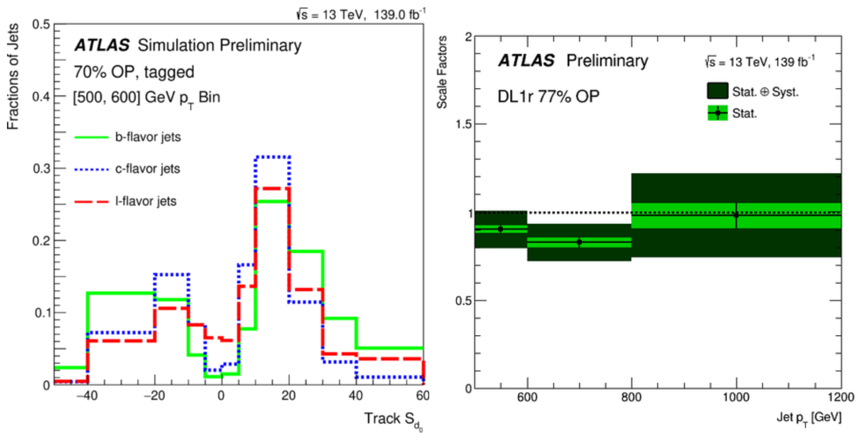
(a) (b)

Figure 2: Semileptonic to inclusive efficiency SF extrapolation uncertainty plot as function of the $p_{\mathrm{T}}^{\mathrm{jet}}$ (a) [10]. b-tagging efficiency scale factors derived with the p_T^{rel} method for jets passing the 77% OP of the MV2 tagger [10] compared to the $t\bar{t}$ based calibration scale factors [8](b).

between 20 and 40 GeV and in agreement with the $t\bar{t}$ SFs.

3. High p_T calibration

The high p_T calibration provides b-tagging SFs in the $p_{\mathrm{T}}^{\mathrm{jet}}$ range between 500 GeV to 1.2 TeV. The variable used to estimate ε_b in data is the signed impact parameter significance, S_{d_0}. Given the relative longer lifetime of b-hadrons compared to c- or *light*-hadrons, their decays produce tracks along the direction of flight with a larger impact parameter, d_0. A sign can be assigned for each track by looking at the cross product of the track vector and the closest approach vector on the transverse plane. Tracks in b-jets are more likely to have large positive values, which means they are produced in front of the primary vertex. To take the tracking resultion into account, S_{d_0} is defined dividing the signed d_0 by its uncertainty. The S_{d_0} of the track with the second largest $|S_{d_0}|$, shown in Figure 3a is used to derive the ε_b SF, shown in Figure 3b. A significant contribution to the



(a) **(b)**

Figure 3: Distribution of $S_{d_0}$for jets with $p_{\mathrm{T}}^{\mathrm{jet}}$ between 500 and 600 GeV, passing the 70% OP of the DL1r tagger (a). b-tagging efficiency SFs as a function of the $p_{\mathrm{T}}^{\mathrm{jet}}$ for the 77% OP of the DL1r tagger (b) [11].

final SF uncertainty in this high p_T calibration is caused by the modelling and track uncertainties as well as the MC statistical uncertainty.

4. Conclusions

Calibrations of the b-tagging efficiency based on multijet events are a powerful tool to assess the performances of the ATLAS b-tagging algorithms in data. The request of at least one jet matched to a muon is a simple yet effective way to enhance the b-jets component of the probe jets. In the $p_{\mathrm{T}}^{\mathrm{jet}}$ region between 20 and 40 GeV the muon p_T^{rel} calibration provides a more precise estimate of the ε_b scale factor. Thanks to the higher cross-section of the multijet process, such a good result is achieved with a lower effective luminosity, opening for the possibility of using this method in the early stages of the Run 3. This high cross-section allowed to probe a region that was never probed before, resulting in a data-driven estimate of the b-tagging performance for jets up to a p_T of 1.2 TeV.

References

[1] 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“, *J. High Energ. Phys.* 2022,97

[2] ATLAS Collaboration, “Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector“, *Eur. Phys. J. C*,81,178(2021)

[3] ATLAS Collaboration, “Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s}$ = 13 TeV“, *Eur. Phys. J. C*, 81,537(2021)

[4] ATLAS Collaboration, “Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s}$ = 13 TeV with the ATLAS experiment“, *Phys. Lett. B*, 810(2020)135797

[5] ATLAS Collaboration, “Search for resonances in the mass distribution of jet pairs with one or two jets identified as b-jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector“, *Phys. Rev. D* 98, 032016(2018)

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

[7] 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* 79,970(2019)

[8] ATLAS Collaboration,“Measurement of the b-jet identification efficiency with the p_T^{rel} method in multi-jet events using pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector“, *ATL-PHYS-PUB-2022-025*, 2022

[9] ATLAS Collaboration,“Meaurement of the b-jet identification efficiency for high p_T jets using multijet events in proton-proton collisions at $\sqrt{s}$ = 13 TeV recorded with the ATLAS detector“, *ATL-PHYS-PUB-2022-010*, 2022

[10] DELPHI Collaboration, “Determination of the heavy quark non-perturbative parameters from spectral moments in semileptonic B decays“, *Eur. Phys. J.* C45 (2006)35