Light-jet mis-tag efficiency calibration with the ATLAS Experiment

L. Toffolin on behalf of the ATLAS Collaboration (*)
Dipartimento di Fisica, Università degli Studi di Trieste - Trieste, Italy
INFN, Sezione di Trieste, Gruppo Collegato di Udine - Udine, Italy

Summary. — In the ATLAS Experiment, an important source of systematic uncertainty for many physics analyses, especially in the top quark sector and in searches for high mass resonances, is constituted by those light jets, originated by up, down and strange quarks, and gluons, which can be wrongly identified as b-jets. It is therefore crucial to determine the rate and the efficiency of mis-identification of light jets with heavy flavour jets. The light jet efficiency calibration is challenged by the high rejection factors of the state-of-art b-tagging algorithms. It is therefore difficult to find a filtered sample of events pure enough in light jets following b-tagging selections. This problem is particularly challenging due to the excellent performances in b-jet identification of the modern algorithms widely used in ATLAS, based on deep neural networks (`DL1r`, `DL1d`) and graph neural networks (`GN2`). A possible solution for the low statistics issue after the b-tagging selection is the so-called "negative tag" method, which consists in enriching the $Z+$ jets data sample artificially via the inversion of the sign of the impact parameter of the tracks associated to the jets.
In this contribution, the algorithm and the strategies adopted in the light jet calibration in ATLAS are presented, with particular focus on the calibration with $Z+$ jets events. The most recent ATLAS results obtained for LHC Run-2 and Run-3 are finally discusses.

1. – Calibration

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

(*) Email: `leonardo.toffolin@cern.ch`

in order to correct for efficiency differences between data and simulation that may arise from an imperfect description of the data, e.g. in the parton shower and fragmentation modelling or in the detector and response simulation.

ATLAS analyses use selection requirements defining a lower bound on the tagger discriminant to select b-jets with a certain efficiency. Four of these so-called single-cut operating points (OPs) are defined, corresponding to b-jet selection efficiencies of 85%, 77%, 70% and 60%. The OPs are evaluated in a sample of b-jets from simulated $t\bar{t}$ events.

The efficiency of identifying a b-jet (ϵ^b) and the mis-tagging efficiencies (ϵ^c and ϵ^{light}), which are the probabilities that other jets are wrongly identified by the b-tagging algorithms as b-jets, are measured in data and compared with the predictions of the simulation. These tagging and mis-tagging efficiencies are defined as:

$$\epsilon^f = \frac{N^f_{\text{pass}}}{N^f_{\text{all}}} \tag{1}$$

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

$$SF^f = \frac{\epsilon^f_{\text{data}}}{\epsilon^f_{\text{MC}}} \tag{2}$$

The calibration factors correct the efficiencies and mis-tagging efficiencies in simulation to better reproduce performance obtained in data and are applied to all physics analyses in ATLAS that use b-tagging. The b-jet efficiency (ϵ^b) is calibrated using the method described in Ref. [2], where the SFs^b are extracted from a sample of events containing top-quark pairs decaying into a final state with two charged leptons and two b-jets. The c-jet mis-tagging efficiency (ϵ^c) is calibrated via the method described in Ref. [3], where the SFs^c are extracted from events containing top-quark pairs decaying into a final state with exactly one charged lepton and several jets. The events are reconstructed using a kinematic likelihood technique and include a hadronically decaying W boson, whose decay products are rich in c-jets.

2. – Light-jet mis-tag efficiency calibration: the "negative tag" method

The mis-tagging efficiency for light jets, ϵ^{light}, is difficult to calibrate because after applying a b-tagging requirement, the resulting sample of jets is strongly dominated by b-jets. Thus, the fraction of light-flavour jets passing a loose selection on the b-tagging score is too low to estimate $\epsilon^{\text{light}}_{\text{data}}$. In order to extract an unbiased and precise scale factor for light jets, SF$^{\text{light}}$, a sample enriched in mis-tagged light-flavour jets is required.

One of the strategies widely adopted by the ATLAS Collaboration is the so called "negative tag" method [4, 5]. It consists in calibrating data via a modified version of the recommended b-tagging algorithms (`DL1d` [6], `GN2` [7]) mentioned above, that achieves lower ϵ^b to ϵ^{light} light ratios without changing ϵ^{light} significantly. This strategy is based on some assumptions related to the tracks assigned to a certain jet. In general, tracks matched to b-jets have relatively large and positively signed impact parameters (IPs)

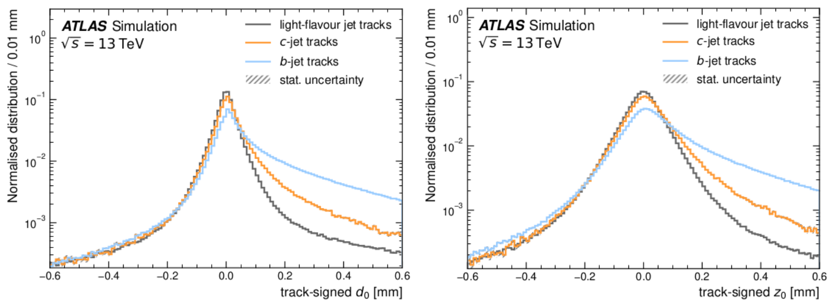

Fig. 1. – Signed transverse impact parameter d_0 (left) and signed longitudinal impact parameter z_0 (right) distributions for tracks matched to b-jets, c-jets and light-flavour jets in simulated $t\bar{t}$ events. The selected tracks are matched to particle-flow jets with $p_T > 20$ GeV and $|\eta| < 2.5$, and pass the jet-vertex tagger selection. The distributions are normalised to unity. Statistical uncertainties are also shown. Plots are taken from [8].

due to the long lifetime of the b-hadrons and the presence of displaced decay vertices. In contrast, tracks matched to light-flavour jets typically have IP values consistent with zero within the IP resolution such that a more symmetric IP distribution ([1]) is expected. The expected IP distributions of the tracks associated with b-jets, c-jets or light-flavour jets are shown in Fig. 1.

The negative tag method assumes that the probability for a light flavour jet to be mis-tagged remains almost the same when inverting the IP signs of all tracks and displaced vertices. This is based on the assumption that light flavour jets are misidentified as b-jets mainly due to resolution effects in the track reconstruction, which result in tracks with positive IPs inside the jet. Given the symmetric IP distributions, the fractions of tracks and vertices from tracks with positive IPs remain stable after inverting the IP signs of all tracks and vertices. The presence of the positive tail in the IP distribution challenges this assumption and its impact is taken into account by a dedicated "extrapolation uncertainty".

Therefore, the adoption of this method allows to enrich artificially the fraction of light jets in the considered data sample via inverting the sign of the impact parameter of the tracks associated with the jets. It can be finally applied to both `DL1d` and `GN2` taggers, which namely become `DL1dFlip` and `GN2Flip`, for the mainstream calibration provided by the ATLAS Flavour Tagging group in $Z+$ jets events, as well as for alternative calibrations in di-jet events.

3. – Preliminary results on Run-2 and Run-3 data

The calibration of the light-jet mis-tagging efficiency is performed independently in jet p_T intervals in order to account for the p_T dependence of ϵ^{light}. A simultaneous binned fit to the distribution of the mass of the secondary vertex, m_{SV}, in each pseudo-continuous

---

([1]) The tracks matched to light-flavour jets have a slight bias towards positive-sign values due to the presence of some long-lived particles (e.g. K_S or Λ), as shown in Fig. 1. The contribution from the mis-modelling of long-lived particles is expected to be negligible relative to the mis-modelling of the d_0 and z_0 resolutions [4].

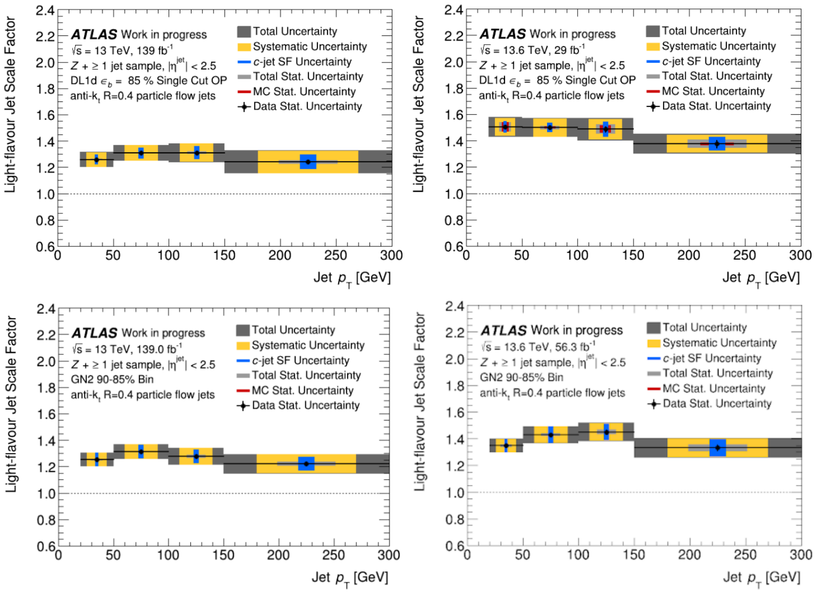

Fig. 2. – SFs to calibrate the ϵ^{light} of the pseudo-continuous OPs for the `DL1d` (first row) and `GN2` tagger (second row) on the full set of Run-2 data (left) and a 56.3 fb^{-1} sample (top right) and a 29 fb^{-1} sample (bottom right) of Run-3 data collected by the ATLAS detector. The size of the uncertainty bands in the direction of the jet p_T axis is arbitrary and corresponds to the choice of the calibration intervals in p_T.

interval of the $\epsilon^{b}_{\text{data}}$ and $\epsilon^{\text{light}}_{\text{data}}$ in the b-tagging discriminant is performed in order to simultaneously determine $\epsilon^{b}_{\text{data}}$ and $\epsilon^{c}_{\text{data}}$ discriminant intervals. The sensitivity of the fit does not allow the SFs of all three jet flavours to be derived simultaneously. Therefore, $\epsilon^{c}_{\text{data}}$ is constrained to the MC predictions and SF^{c} is fixed to unity within an uncertainty of 30%, as suggested by studies of the c-jet mis-tagging efficiency calibration [3].

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

$$N_i(m_{\text{SV}}) = N_{i,\text{MC}} \cdot C \cdot \sum_{f=\text{light},c,b} F^f \cdot SF_i^f \cdot \epsilon^f_{i,\text{MC}} \cdot P_i^f(m_{\text{SV}}) \tag{3}$$

where $N_{i,\text{MC}}$ is the predicted flavour-inclusive event yield for each discriminant interval; C is a global normalisation factor and F^f are the *jet-flavour fractions*; $P_i^f(m_{\text{SV}})$ is the probability density function of m_{SV} for jet flavour f in the i-th tagger discriminant interval, taken from simulation. The $P_i^f(m_{\text{SV}})$ is defined in such a way to integrate an additional bin ($m_{\text{SV}} < 0$ GeV) representing the number of events where no secondary vertex is found. The m_{SV} distribution has been obtained with tracks with nominal sign as input to the secondary vertex finder algorithm. The C, F^f and SF$^{b,\text{light}}$ parameters are allowed to float in the fit, while $N_{i,\text{MC}}$ and $\epsilon^f_{i,\text{MC}}$ are fixed to the predictions from simulated events and SF^c is set to 1.0 ± 0.3.

Preliminary results on the Run-2 and Run-3 scale factors for the `DL1d` and `GN2` taggers are presented in Fig. 2 for the 85% and 90% OPs respectively.

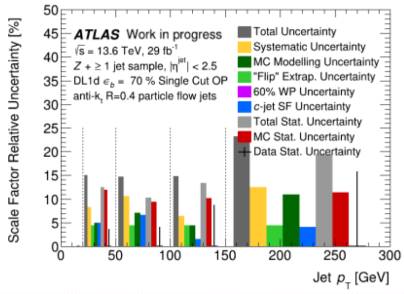

Fig. 3. – `DL1d` mis-tagging efficiency calibration uncertainties for light jets with a 29 fb^{-1} sample of Run-3 data from 2022. The breakdown of the different uncertainties is shown for the 70% OP.

The calibration SFs$^{\text{light}}$ of the `DL1d` and `GN2` algorithms are presented. The 90%, 85%, 77% and 70% OPs have been successfully calibrated in data by using the negative Tag method. However, it is not feasible to calibrate the 60% OP in data because of insufficient statistics and the relatively large contamination by heavy-flavour jets.

The measurement of SF$^{\text{light}}$ is affected by four types of uncertainties, including those due to experimental effects, the modelling of the $b+$ jets and background processes, and the limited number of events in data and simulation. For each source of uncertainty, one parameter of the fit model is varied at a time, and the effect of this variation on SF$^{\text{light}}$ is evaluated. The relative contribution of the various uncertainties is shown in Fig. 3 for the `DL1d` tagger only, since those for `GN2` tagger are still under validation at the time of writing, and are listed below:

- Monte Carlo Modelling (dark green): modelling of fit template;
- "flip" extrapolation (light green): uncertainty to cover potential differences between `DL1(d)Flip` and direct `DL1(d)` calibration;
- calibration of `DL1(d)Flip` for other jet flavours (blue): this includes contribution of other jet flavours in selected sample, for instance the one by c-jets, which appears to be subdominant;
- MC statistical uncertainty (red): it is related to the amount of simulated data and gives subdominant contributions, except for high jet p_T and tight OPs (in particular, 60%).

In addition to the systematic uncertainties listed above, a relative 50% uncertainty is assigned by hand to the 60% OP, in order to keep track on the fact that the fit for that OP is failing.

4. – Conclusions

The light-flavour jet mis-tagging efficiency ϵ^{light} of the `DL1d` b-tagging algorithm has been measured with a 139 fb^{-1} sample of $\sqrt{s} = 13$ TeV collision events recorded during 2015–2018 and a 56.3 fb^{-1} sample of $\sqrt{s} = 13.6$ TeV collected during 2022 and 2023 by the ATLAS detector at the LHC. The measurement is based on an improved method applied to a sample of Z+jets events. The negative tag method, based on the application of an alternative b-tagging algorithm, designed to facilitate the measurement of the light-flavour jet mis-tagging efficiency, is used. Data-to-simulation scale factors for correcting $\epsilon^{\mathrm{light}}$ in simulation are measured in different jet transverse momentum intervals, ranging from 20 to 300 GeV, for four separate quantiles of the b-tagging discriminant. The total uncertainties range from 11% to around 25%, and the scale factors do not exhibit any strong dependence on jet transverse momentum.

REFERENCES

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

[2] 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*, **82** (2011) 95, arXiv: 2109.10627 [hep-ex].

[3] ATLAS Collaboration, *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, *Eur. Phys. J. C*, **82** (2021) 95, arXiv: 2109.10627 [hep-ex].

[4] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046, 2014, `url`: https://cds.cern.ch/record/1741020.

[5] D0 Collaboration, *Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 1.96$ TeV using secondary vertex b-tagging*, *Phys. Rev. D*, **74** (2006) 112004.

[6] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, *Eur. Phys. J. C*, **83** (2023) 7,681.

[7] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-047, `url`: https://cds.cern.ch/record/2811135.

[8] ATLAS Collaboration, *Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb^{-1} of ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C*, **83** (2023) 728.