# Measuring the b-jet tagging efficiency using top anti-top events with ATLAS data

*Dominik Duda on behalf of the ATLAS Collaboration.*
*Bergische Universität Wuppertal*
*D-42119 Wuppertal, GERMANY*

The identification of jets originating from $b$-quarks is relevant for many physics analyses with the ATLAS detector [1]. Algorithms that allow to identify such jets are thus of great importance and it is crucial to understand their performance with data-driven measurements of efficiencies and fake rates.
Since the decay of a $t\bar{t}$ pair has a very clear signature and because the top quark almost exclusively decays to a $W$ boson and a $b$-quark, a sample of top anti-top events is ideal for calibrating the various $b$-tagging algorithms used in ATLAS analyses. The calibration results are presented in the form of data-to-simulation scale factors

$$\kappa_{\varepsilon_b}^{\text{data/sim.}} = \frac{\varepsilon_b^{data}}{\varepsilon_b^{sim.}} \tag{1}$$

in which the measured $b$-tagging efficiency in data is divided by the $b$-tagging efficiency in simulated events. Currently three different $t\bar{t}$ calibration methods are in use, where both the lepton+jets and the dileptonic decay channels are considered [2].
Using the kinematic selection method the $b$-tagging efficiency can be estimated by

$$\varepsilon_b = \frac{1}{f_{b-\text{jets}}} \cdot \left( f_{b-\text{tag}} - \varepsilon_c f_{c-\text{jets}} - \varepsilon_l f_{l-\text{jets}} - \varepsilon_{\text{fake}} f_{\text{fake}} \right) , \tag{2}$$

where $f_{b-\text{jets}}$, $f_{c-\text{jets}}$ and $f_{l-\text{jets}}$ are the fractions of $b$-, $c$- and light-flavour jets within the selected sample, while $f_{\text{fake}}$ is the fraction of jets coming from the fake lepton background. The mistag efficiencies $\varepsilon_c$ for $c$- and $\varepsilon_l$ for light-flavour jets as well as the fractions of the various jet flavours are calculated using all the selected events in simulation, while the tagging efficiency of the jets coming from the fake lepton background $\varepsilon_{\text{fake}}$ and the fraction of $b$-tagged jets $f_{b-\text{tag}}$ are obtained from data.
While the $b$-jet purity with the kinematic dileptonic selection reaches 40-80 %, the lepton+jets selection includes a requirement that at least one jet in each event is $b$-tagged, to enhance the signal purity. If the leading jet (the second leading jet) in the lepton+jets analysis passes that $b$-tagging requirement, the $b$-tagging efficiency is measured on the three subleading jets (the leading jet). In the dilepton channel the $b$-tagging efficiency is instead extracted from the two leading jets.

In case of the tag counting method a Likelihood fit is performed to estimate the $b$-tagging efficiency by assuming that the expected number of events containing $n$ $b$-tagged jets is given by

$$
\begin{aligned}
< N_n > = \sum_{i,j,k} \Bigg\{ & (\sigma_{t\bar{t}} \cdot \mathrm{BR} \cdot \mathcal{A}_{t\bar{t}} \cdot \mathcal{L} \cdot F_{ijk}^{t\bar{t}} + N_{\mathrm{bkg}} \cdot F_{ijk}^{\mathrm{bkg}}) \times \\
& \sum_{i'+j'+k'=n} \binom{i}{i'} {\varepsilon_b}^{i'} (1-\varepsilon_b)^{i-i'} \cdot \binom{j}{j'} {\varepsilon_c}^{j'} (1-\varepsilon_c)^{j-j'} \cdot \binom{k}{k'} {\varepsilon_l}^{k'} (1-\varepsilon_l)^{k-k'} \Bigg\}
\end{aligned}
\tag{3}
$$

where $i$, $j$ and $k$ are the number of $b$-, $c$- and light-flavour jets per event before applying $b$-tagging, while $i'$, $j'$ and $k'$ represent the number of those jets after $b$-tagging. $F_{ijk}$ is the fraction of events (before tagging) containing $i$ $b$-jets, $j$ $c$-jets and $k$ light-flavour jets. $BR$ is the branching ratio, $\mathcal{A}_{t\bar{t}}$ is the selection acceptance and $\mathcal{L}$ is the integrated luminosity. The $b$-tagging efficiency can then be determined by fitting this expected $n$-tag distribution to that observed in data.
The $b$-tagging calibration results for the kinematic selection and tag counting methods are shown in Figure 1.

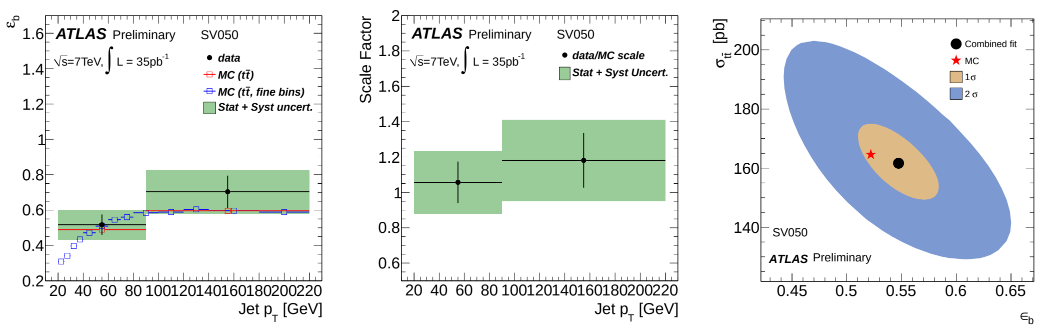


Figure 1: $b$-tagging efficiencies in data and simulation (left), the corresponding scaling factors $\kappa_{\varepsilon_b}^{\mathrm{data/sim.}}$ (middle) both measured with the kinematic selection method and the 2-dimensional contour of the $b$-tagging efficiency and the $t\bar{t}$ cross section from the tag counting method (right). All plots show results for the SV0 tagging algorithm at an operating point corresponding to 50 % signal efficiency in simulated $t\bar{t}$ events [2].

# References


[1] ATLAS Collaboration, 2008 JINST 3 S08003.

[2] ATLAS Collaboration, *Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb$^{-1}$ of Data with the ATLAS Detector*, ATLAS-CONF-2011-089, https://cdsweb.cern.ch/record/1356198.