PROCEEDINGS OF SCIENCE

# Performance and calibration for the identification of boosted Higgs bosons decaying into beauty quark pairs in ATLAS

**Eleanor Jones*on behalf of the ATLAS Collaboration**

*University of Warwick,*
*Coventry, CV4 7AL, UK*

*E-mail:* eleanor.jones@cern.ch

These proceedings present the performance and calibration of the $X \to b\bar{b}$ tagger, developed by the ATLAS experiment to identify boosted Higgs bosons decaying in the dominant decay channel to a beauty quark-antiquark pair. Data-MC scale factors for the signal efficiency are calculated using $Z\gamma$ and $Z+$jets events, in four regions of $p_{\mathrm{T}}$, and vary between $1.45^{+0.47}_{-0.45}$ and $0.55^{+0.23}_{-0.22}$, whilst the mis-tag efficiency scale factors are measured using single-lepton $t\bar{t}$ events, also in four regions of $p_{\mathrm{T}}$, and vary between $1.1 \pm 0.12$ and $1.0 \pm 0.16$. The modelling of the large-$R$ jet kinematics is evaluated using multijet events enriched in $g \to b\bar{b}$ splitting, achieved by requiring a non-isolated muon contained inside the large-$R$ jet.

**** Particles and Nuclei International Conference - PANIC2021 ****
**** 5 - 10 September, 2021 ****
**** Online ****

*Speaker





PoS(PANIC2021)403

The discovery of the Higgs boson in 2012 [1, 2] was a landmark discovery for particle physics. The last of the Standard Model (SM) fundamental particles to be observed, the properties of the Higgs boson continue to be extensively studied by the ATLAS experiment [3] and precision measurements are crucial for understanding the nature of this particle and testing SM theory predictions. There are many Beyond Standard Model (BSM) scenarios where high-transverse momentum Higgs bosons are produced in the decay of new heavy particles. In addition, BSM effects could modify the production differential cross-section, particularly in the high-$p_\mathrm{T}$ regime.

The SM Higgs boson decays preferentially to a beauty quark-antiquark pair ($b\bar{b}$), with a branching fraction of 58% for a Higgs boson mass of 125 GeV. As the $p_\mathrm{T}$ of the Higgs boson increases, the $b$-jets become progressively more collimated and are resolved inside a large-radius (large-$R$) jet. A dedicated algorithm, the $X \to b\bar{b}$ tagger [4], is being introduced to identify these jets and this document reports on its performance and calibration.

Currently in ATLAS, two algorithms identifying jets from $b$-quarks are maintained: MV2, which is a boosted decision tree, and DL1r, which is a deep neural network [5]. These taggers take as inputs the output discriminants of 'low-level' algorithms, which each exploit one of the distinct features of $B$-hadron decays. As the $p_\mathrm{T}$ of the Higgs boson increases, the performance of the standard $b$-taggers degrades, due in part to the difficulty in resolving the individual track-jets when they become collimated inside the large-$R$ jet. Therefore, the $X \to b\bar{b}$ tagger has been developed that aims to identify boosted $H \to b\bar{b}$ decays by tagging the large-$R$ jet, which can contain up to three variable radius (VR) track-jets [6]. The algorithm is a feed-forward neural network that combines the individual DL1r scores of the track-jets, $p_u$, $p_c$ and $p_b$, with the $p_\mathrm{T}$ and $\eta$ of the large-$R$ jet. Three outputs are produced, corresponding to the probabilities for Higgs ($p_\mathrm{Higgs}$), multijet ($p_\mathrm{multijet}$) and top ($p_\mathrm{top}$) process hypotheses. These outputs are combined into a single discriminant, $D_\mathrm{Xbb}$,

$$D_\mathrm{Xbb} = ln \frac{p_\mathrm{Higgs}}{f_\mathrm{top} \cdot p_\mathrm{top} + \left(1 - f_\mathrm{top}\right) \cdot p_\mathrm{multijet}} , \tag{1}$$

where $f_\mathrm{top}$ determines the fraction of top background and is set to 0.25.

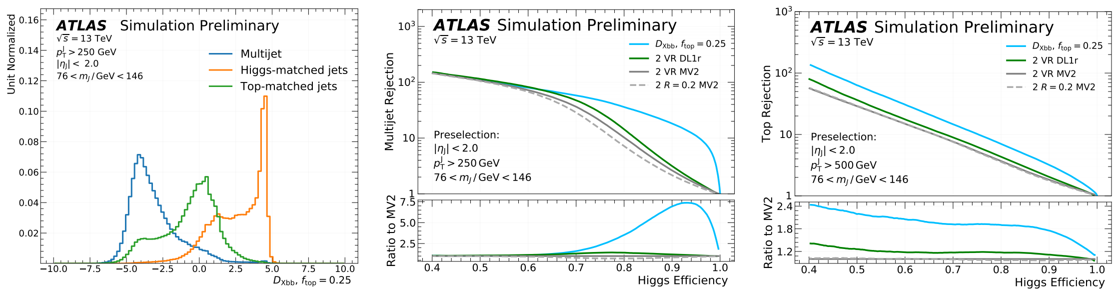
The distribution of the discriminant defined in Eq. (1) is shown in Figure 1(a). The Higgs and top jet candidates are required to be ghost-associated to exactly one truth Higgs boson or top-quark, respectively. Good discrimination between the signal (Higgs-matched jets) and the backgrounds (top-matched jets and multijet) is observed. Large-$R$ jets are considered tagged if the value of the discriminant is above a certain threshold. The performance of the tagger is evaluated using the Higgs efficiency, $\epsilon$, defined as the number of tagged jets divided by the total number of $H \to b\bar{b}$ jets. Figures 1(b) and 1(c) show the rejection for the multijet and top backgrounds, respectively, as a function of the Higgs efficiency. A significant improvement in performance over the standard $b$-taggers can be seen.

Data-to-simulation scale factors need to be derived to match the efficiency measured in data to that predicted by simulation. The scale factor is calculated as shown in Eq. (2), where $\mu_\text{pre-tag}$ and $\mu_\text{post-tag}$ are the signal strengths relative to the SM expectation before and after tagging, respectively.

$$\mathrm{SF} = \frac{\epsilon^\mathrm{data}}{\epsilon^\mathrm{MC}} = \frac{\frac{N^\mathrm{data}_\mathrm{passed}}{N^\mathrm{data}_\mathrm{total}}}{\frac{N^\mathrm{MC}_\mathrm{passed}}{N^\mathrm{MC}_\mathrm{total}}} = \frac{\frac{N^\mathrm{data}_\mathrm{passed}}{N^\mathrm{MC}_\mathrm{passed}}}{\frac{N^\mathrm{data}_\mathrm{total}}{N^\mathrm{MC}_\mathrm{total}}} = \frac{\mu_\text{post-tag}}{\mu_\text{pre-tag}} . \tag{2}$$

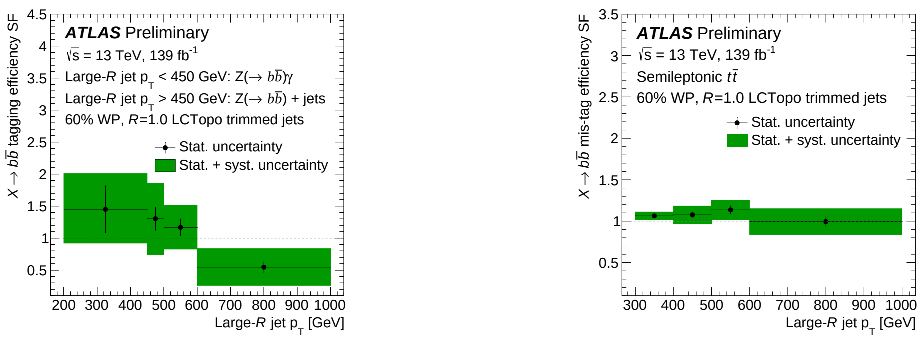


PoS(PANIC2021)403




**(a)** **(b)** **(c)**

**Figure 1:** (a) The discriminant distribution for the $X \to b\bar{b}$ algorithm, normalised to unity. (b) The multijet and (c) top background rejection as a function of the Higgs efficiency for large-$R$ jet $p_T > 500$ GeV for the $X \to b\bar{b}$ (blue). The performance of the DL1r (green) and variants of the MV2 (grey) taggers are also shown for comparison [4].




**(a)** **(b)**

**Figure 2:** (a) Signal efficiency scale factors and (b) mis-tag efficiency scale factors for the $X \to b\bar{b}$ tagger at the 60% operating point [7].

$\mu_{\text{post-tag}}$ is extracted in four different $p_T$ regions using $Z\,(\to b\bar{b}) + \gamma$ events for $200 < p_T < 450$ GeV and $Z\,(\to b\bar{b})$ + jets events for $450 < p_T < 500$ GeV, $500 < p_T < 600$ GeV and $600 < p_T < 1000$ GeV. Events are additionally required to have either exactly one photon with $p_T > 200$ GeV or at least one other large-$R$ jet with $p_T > 200$ for the $Z\,(\to b\bar{b}) + \gamma$ and the $Z\,(\to b\bar{b})$ + jets events, respectively. The $Z$-candidate is required to have a mass $50 < m_Z < 150$ GeV. A binned likelihood template fit and an unbinned likelihood fit to the $b\bar{b}$ invariant mass spectrum is performed for the $Z\,(\to b\bar{b}) + \gamma$ process and $Z\,(\to b\bar{b})$ + jets process, respectively. The $Z \to l^+l^-$ decay channel is used to extract $\mu_{\text{pre-tag}}$ in both measurements since overwhelming backgrounds make a pre-tag measurement unfeasible in the $Z \to b\bar{b}$ channel.

The scale factors are shown in Figure 2(a). The dominant uncertainty in the first $p_T$ bin is statistical, whereas in the other three $p_T$ bins, the uncertainty is dominated by either the uncertainty in the fit model due to the choice of mass range or the uncertainty coming from the modelling of the $Z$ + jets background. Figure 2(b) shows the scale factors derived for the mis-tag efficiency, which is defined as the probability to misidentify a top-quark jet as a signal event. The measurement is performed with a binned likelihood fit to the large-$R$ jet mass spectrum simultaneously in two

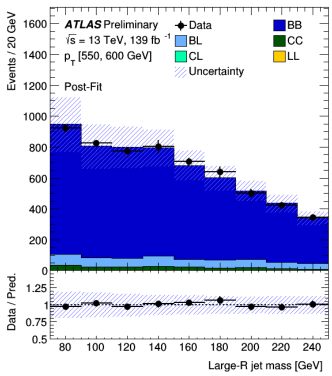



**Figure 3:** The large-$R$ jet mass after tagging at the 60% operating point in multijet events for $550 < p_{\mathrm{T}} < 600$ GeV [7].

signal regions (defined by whether the leading large-$R$ jet in the event passes or fails the tagging requirement) in four $p_{\mathrm{T}}$ regions, using single-lepton $t\bar{t}$ events. The measurement is limited by the systematic uncertainties, with the dominant source coming from the $t\bar{t}$ modelling. However, a precision in the range of 5–16% is achieved.

For many boosted Higgs analyses, the dominant background is from QCD multijet production, particularly in analyses probing all-hadronic final states. Consequently, the modelling of the large-$R$ jet kinematics after the application of the $X \to b\bar{b}$ tagger is an important validation to ensure that the tagger is not sculpting the mass distribution around the Higgs boson mass. A multijet sample, where the presence of a non-isolated muon is required to enhance the fraction of $g \to b\bar{b}$ events, are used to perform these checks. First, a correction to the flavour composition is required to match the flavour composition in simulation to that in data. This is performed using a template fit to the mean of the signed $d_0$ significance, $\langle s_{d_0} \rangle$, which is a flavour-sensitive variable. The correction is derived as a function of $p_{\mathrm{T}}$, according to the number of events observed in data, in three regions: $550 < p_{\mathrm{T}} < 600$ GeV, $600 < p_{\mathrm{T}} < 750$ GeV and $750 < p_{\mathrm{T}} < 1000$ GeV. Figure 3 shows the large-$R$ jet mass after tagging and after applying the flavour fraction correction for the first $p_{\mathrm{T}}$ bin. Good agreement between data and simulation is observed.

In conclusion, these proceedings present the performance and calibration of the $X \to b\bar{b}$ tagger, which has been developed as a new approach to tagging boosted $H \to b\bar{b}$ events in ATLAS. The signal efficiency calibration is performed using $Z\,(\to b\bar{b}) + \gamma$ and $Z\,(\to b\bar{b})$ + jets events and the mis-tag efficiency is measured in single-lepton $t\bar{t}$ events. The modelling of large-R jet kinematics is validated using multijet events enriched in $g \to b\bar{b}$ events.

## References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1–29

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30–61

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

[4] ATLAS Collaboration, *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*, ATL-PHYS-PUB-2020-019

[5] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in $pp$ collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019) 970

[6] D. Krohn, J. Thaler and L-T. Wang, *Jets with variable R*, *JHEP* **06** (2009) 059

[7] ATLAS Collaboration, *Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATL-PHYS-PUB-2021-035