PROCEEDINGS OF SCIENCE

Top tagging at ATLAS

Julien CAUDRON*, **on behalf of the ATLAS Collaboration**
Johannes-Gutenberg-Universitaet Mainz (DE)
E-mail: julien.caudron@cern.ch

Studies of top physics in the boosted sector have known a fast-growing development with the arrival of high-energy data at the LHC. Different techniques to identify high-p_T top quarks based on substructure analyses of large radius jets have been developed for Run-1 and Run-2 data. New results are presented on the optimization and performances, for the different techniques (HEPTopTagger, Shower Deconstruction and substructure variables cut-based taggers), using pp collision data and MC simulations at $\sqrt{s} = 8$ TeV in ATLAS.

8th International Workshop on Top Quark Physics
14-18 September, 2015
Ischia, Italy

*Speaker.

PoS(TOP2015)061

1. Introduction

Top quark plays an important role in many beyond-the-standard-model scenarios. When produced by a massive new particle, the top quark will have a very high transverse momentum, and will be referred to as boosted top. In this regime, the study of the substructure of the large radius jet that fully contains the hadronic decay products is more efficient than the resolved reconstruction. The algorithms that identify these decays are called top-taggers.

2. Top-tagging algorithms

The ATLAS collaboration [1] has studied intensively three types of top-tagging algorithms [2].

The **substructure variable taggers** are based on simple rectangular cuts on one or several substructure variables, such as the mass m of the trimmed jet [3], the k_t splitting scale [4] ($\sqrt{d_{12}}$ or $\sqrt{d_{23}}$) obtained during the reclustering of the jet constituents with the k_t algorithm, and the N-subjettiness ratio [5] (τ_{32} and τ_{21}) obtained from the N-subjettiness variables τ_N. Six substructure variable taggers have been studied: tagger I ($\sqrt{d_{12}} > 40$ GeV), tagger II ($m > 100$ GeV), tagger III ($m > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV), tagger IV ($m > 100$ GeV, $\sqrt{d_{12}} > 40$ GeV and $\sqrt{d_{23}} > 10$ GeV), tagger V ($m > 100$ GeV, $\sqrt{d_{12}} > 40$ GeV and $\sqrt{d_{23}} > 20$ GeV) and W' top-tagger ($\sqrt{d_{12}} > 40$ GeV, $0.4 < \tau_{21} < 0.9$ and $\tau_{32} < 0.65$).

In the **shower deconstruction** algorithm (SD) [6], likelihoods are calculated for the case the jet originates from the hadronic decay of a top quark (signal) and for the case the jet originates from a background process where hard gluons split into $q\bar{q}$ (background). For signal and background, the likelihoods are calculated from theoretical first principles, including the effect of the parton shower. Subjets of the large-R jet are identified with partons and a weight is calculated for each possible shower history that leads to the observed subjet configuration. This weight is proportional to the probability that the assumed initial particle generates the final configuration, taking into account the standard model amplitude for the underlying hard process and the Sudakov form factors for the parton shower. A final discriminant variable χ is calculated as the ratio of the sum of the signal-hypothesis weights and the sum of the background-hypothesis weights. The large-R jet is considered as tagged if the χ value is higher than a given threshold.

The **HEPTopTagger** algorithm [7] tests the compatibility of the hard structure of a C/A $R = 1.5$ jet with the 3-prong pattern of the hadronic top quark decay. Firstly, a mass drop criterion is used to decompose the large-R jet into a collection of subjets with mass lower than a given value. All possible triplets from this collection are then filtered to reduce contamination from underlying events and pile-up and tested for compatibility with a hadronic top quark decay, based on kinematic requirements on the reclustered three subjets and on the top candidate jet built from this procedure. The large-R jet is considered as tagged if there is at least one triplet satisfying this compatibility.

3. Results

The performance of these top-taggers have been studied using simulation and approximately 20 fb^{-1} of $\sqrt{s} = 8$ TeV Run-1 data.

From simulation, the considered large-R jets from different techniques can be geometrically matched to the corresponding trimmed anti-k_t $R = 1.0$ particle-level jet. For a given transverse momentum and pseudorapidity region, the background rejection can be plotted as a function of the tagging efficiency, as visible in Fig.1. The lines correspond to a scan on the threshold of the discriminating variable used in the tagger.

The efficiency in background-subtracted data and in simulation is shown in Fig.2. The vertical error bars indicate the statistical uncertainty on the efficiency measurement and the data uncertainty band shows the systematic uncertainties. The ratio between the efficiencies in data and simulation is shown at the bottom of each subfigure and the error bar indicates the statistical uncertainty and the band the systematic uncertainty (taking into account the correlations).

4. Conclusion

The ATLAS collaboration has studied intensively three types of top-tagging algorithms, using approximately 20 fb^{-1} of $\sqrt{s} = 8$ TeV Run-1 data. The comparisons with data of the discriminating variables show that the substructure techniques used in those top-taggers are well-modeled. The efficiencies and the mistag rates have been measured, and performance comparisons in simulation have been provided. Those can be used to evaluate which top-taggers are optimal for a given analysis, this choice being dependent on, for example, the considered p_T range, the background contamination, or the final state.

References

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

[2] Atlas Collaboration, *Identification of high transverse momentum top quarks in pp collisions at* $\sqrt{s}$ = *8 TeV with the ATLAS detector*, ATLAS-CONF-2015-036, http://cdsweb.cern.ch/record/2043862

[3] D. Krohn, J. Thaler and L. T. Wang, *Jet Trimming*, JHEP **1002** (2010) 084 [arXiv:0912.1342 [hep-ph]].

[4] J. M. Butterworth, B. E. Cox and J. R. Forshaw, *WW scattering at the CERN LHC*, Phys. Rev. D **65** (2002) 096014 [hep-ph/0201098].

[5] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, JHEP **1103** (2011) 015 [arXiv:1011.2268 [hep-ph]].

[6] D. E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, Phys. Rev. D **87** (2013) 054012 [arXiv:1211.3140 [hep-ph]].

[7] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, *Stop Reconstruction with Tagged Tops*, JHEP **1010** (2010) 078 [arXiv:1006.2833 [hep-ph]].

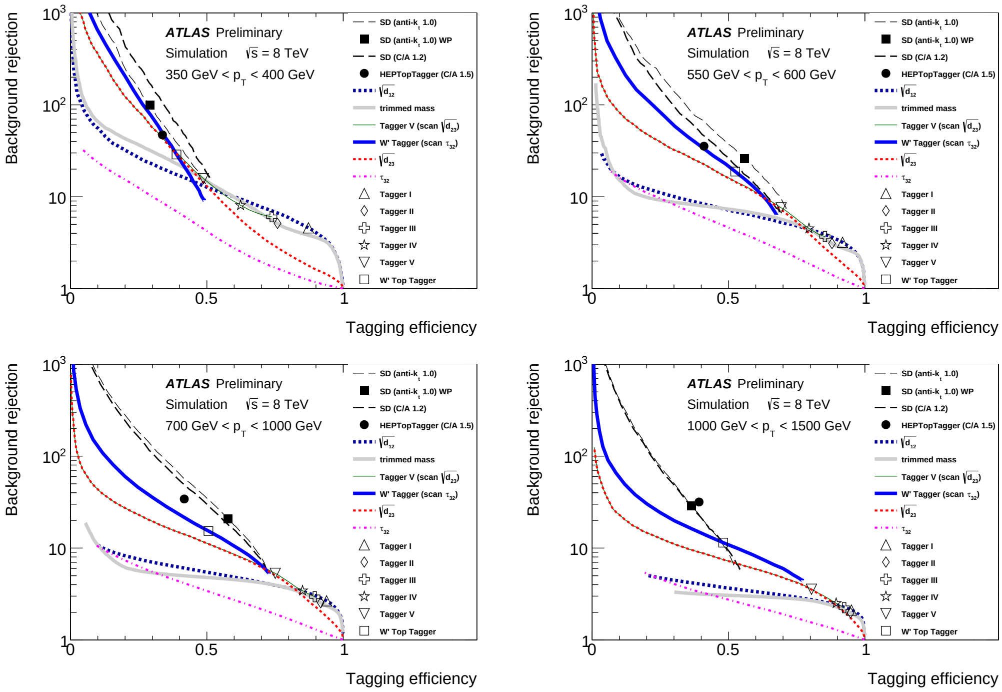

Figure 1: Background rejection as a function of the tagging efficiency of large-R jets, as obtained from MC simulations for different p_T ranges for trimmed anti-k_t $R = 1.0$ particle-level jets to which the large-R jets are geometrically matched. [2]

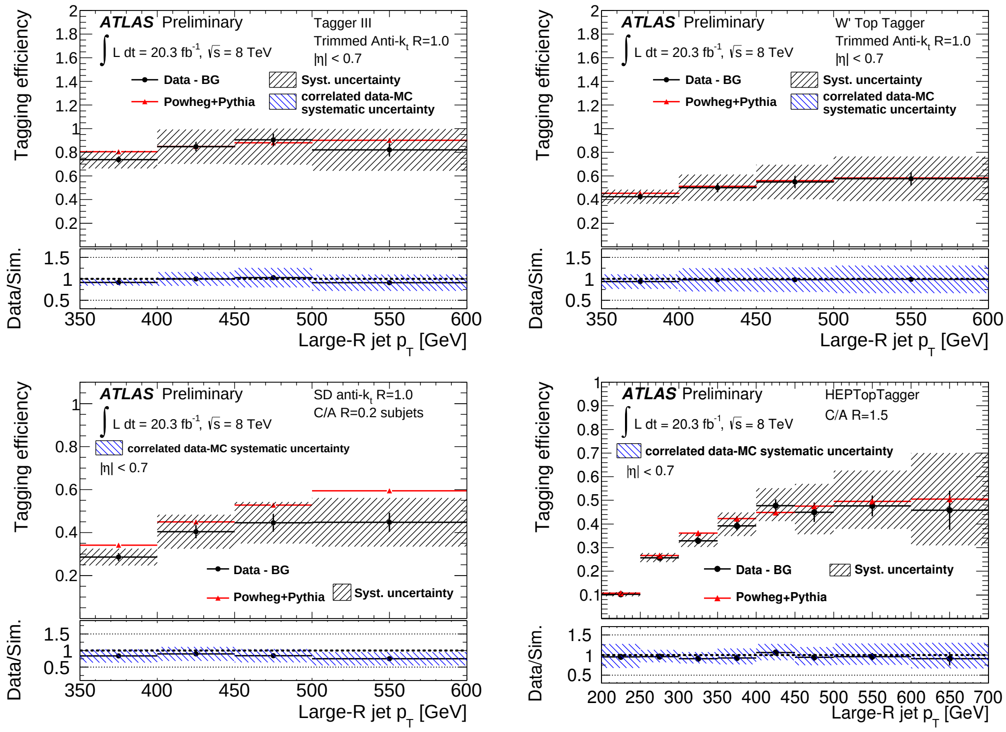

Figure 2: Tagging efficiency of large-R jets in the central region, for different top-taggers (substructure tagger III, substructure tagger W', shower deconstruction and HEPTopTagger), from data and simulation. [2]