Measurements of jet substructure observables using the ATLAS detector

Amal Vaidya*

University College London

E-mail: amal.vaidya@cern.ch

Theoretical calculations for jet substructure observables with accuracy beyond leading-logarithm have recently become available. Such observables are significant not only for probing a new regime of QCD at hadron colliders, but also for improving the understanding of jet substructure properties that are used in many searches for physics beyond the Standard Model. Some first measurements of such quantities are discussed. The soft drop mass is measured in dijet events with the ATLAS detector at $\sqrt{s} =$13 TeV and compared to Monte Carlo simulations and high accuracy fixed order calculations. A measurement of substructure variables in $t\bar{t}$ and inclusive jet events, using data collected by the ATLAS experiment at $\sqrt{s} =$13 TeV is also presented. The measurements are performed with large-radius jets. They are corrected for detector effects, represented as particle-level distributions and are compared to the predictions of various Monte Carlo event generators.

---

*Speaker.

1. Introduction

Jet substructure techniques are now used ubiquitously within ATLAS for the identification of boosted jets from massive hadronically decaying particles and for the separation of quark and gluon initiated jets [1]. Variables designed to be sensitive to the internal structure of a jet are sensitive to the energy and angular resolution of jet constituents. It is an experimentally challenging regime where the relevant variables are sensitive to pile-up and multi-parton interactions (MPI) and non-perturbative QCD effects. There has been a significant effort within ATLAS to improve the performance of jet reconstruction, pile-up removal and identification [2, 3]. Consequently there has been a growing interest in providing detector corrected measurements of a number of commonly used jet substructure variables using jets from a variety of final states. Furthermore, the recent development of soft-drop grooming, one of several methods designed to remove the contributions of pile-up and other noise from a jet, have helped facilitate high accuracy calculations [4].

This has motivated a number of measurements of various substructure variables. A measurement of the soft-drop groomed jet mass in dijet events at $\sqrt{s} = 13$ TeV will be discussed which includes comparisons to resummed calculations performed at NLO+LO (Next-to Leading Order + Next-to Leading Log) and LO+NNLL (Leading Order + Next-to-Next-to Leading Log) [5]. In addition to this a measurement of a number of jet substructure variables calculated in three distinct jet topologies will also be discussed. A variety of substructure variables used for top and massive boson jet discrimination are measured in topologies containing light jets from multijet events and jets from the hadronic days of top quarks and W bosons [6]. These regions of phase space represent jets with distinct classes of substructure which are often studied for classification purposes. Both the measurements required the use of a novel method for obtaining systematic uncertainty which involved applying the systematic variations directly to calorimeter clusters.

2. Soft-drop

The soft-drop declusttering method can be used to classify jets as well as groom them [4]. A detailed description of the method will not be given here but its properties as a groomer have facilitated the resummed calculations mentioned above. Soft-drop groomed variables are formally insensitive to contributions from non-global logarithms, resummation terms associated with particles which radiate out of and then back into a jet. This property is what has allowed higher accuracy calculations [7]. The behaviour of the grooming algorithm itself is determined by two parameters z_{cut} which controls the scale of the energy removed β, which controls the angular weighting.

3. Soft-drop mass measurement

A number of MC event generators and state of the art calculations at NLO+NLL [8, 9] and LO+NNLL [10, 11] are compared to detector corrected measurements of the soft-drop jet mass. A single jet trigger is used to select events with at least two large radius ($R = 0.8$) jets. In order to minimise the contribution from multijet events an additional cut is placed on the p_{T} balance of the two hardest jets in the event ($p_{\text{T1}}/p_{\text{T2}} < 1.5$). These jets are both used with different and

three different soft-drop grooming parameters were considered $\beta = 0, 1, 2$ with a fixed value of $z_{\text{cut}} = 0.1$. The measurement considers a variable $\log(\rho^2)$ where

$$\rho = m^{\text{soft-drop}}/p_{\text{T}}^{\text{ungroomed}}. \tag{3.1}$$

Unlike the jet mass, ρ is only weakly correlated with the jet p_{T}. Furthermore, when the soft-drop β parameter is set at 0 the jet p_{T} is collinear unsafe whereas the ungroomed p_{T} isn't. One particularly interesting aspect of the $\log(\rho^2)$ spectrum is that different regions of the phase space are sensitive to different parts of the analytical calculation effects. At large values, $\log(\rho^2) > -1.7$, where the jet mass is large relative to the p_{T}, the distribution is more sensitive to fixed order matrix element effects rather than resummation or non-perturbative effects. In the so called resummation region, $-3.7 < \log(\rho^2) < -1.7$, resummation effects dominate while at larger values, $\log(\rho^2) > -3.7$, non-perturbative effects dominate. This is especially useful when comparing the various calculations or MC generators.

A relatively novel bottom-up method was used in order to determine the systematic uncertainty on the measured distributions. Uncertainties were obtained by directly varying the kinematics of the calorimeter clusters that are the jet constituents. The energy scale of each cluster is scaled up and down, and smeared independently. To account for the uncertainty on the angular resolution, the position of clusters is changed by 5 mrad. Additionally a fraction of low energy clusters are dropped to account for the reconstruction efficiency. This is a flexible approach that allows a large variety of substructure variables to be measured. The Iterative Bayesian unfolding [12] method is used to correct the measured distributions to particle level. In order to assess how sensitive the unfolded distributions are to the modelling of QCD the unfolding was performed with an alternate generators and the difference was taken as an uncertainty. Pythia8 was used as the nominal generator and Herwig++ was used for comparison. The largest sources of systematic uncertainty are related to the cluster uncertainties and QCD modelling. Other sources of systematic uncertainty are relatively small.
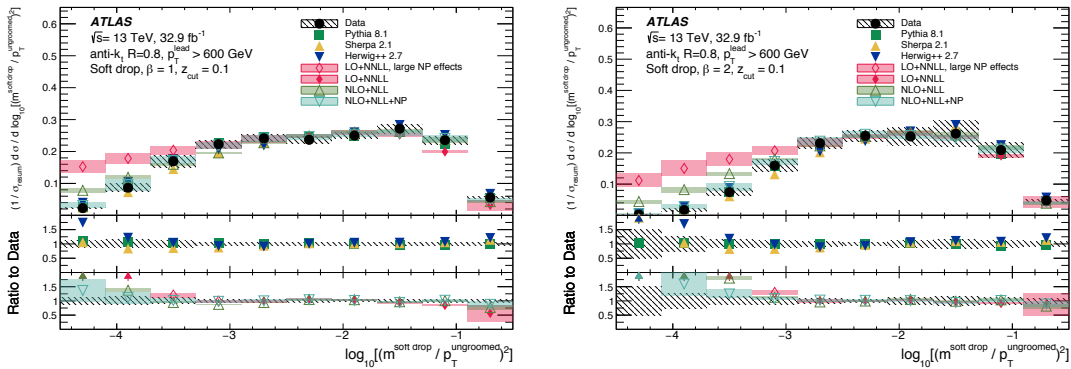
Figure 1: The unfolded $\log(\rho^2)$ distributions for the selected anti-k_t=0.8 jets with soft-drop $\beta = 1$ on the left and $\beta = 2$ on the right. The distributions are compared to state of the art LO+NNLL and NLO+NLL calculations with additional NP effects. The particle level distributions predicted by the Pythia, Sherpa and Herwig++ generators are also shown.

Figure 1 shows the unfolded results and the various predictions for comparison. In the resummation region the analytical results tend to describe the data well. Curiously there are some bins in this central region where the NNLL predictions are outperformed by the NLL. As expected in the low $\log(\rho^2)$ region the analytical predictions start to deviate from the measured data. At high values the agreement between the data and the analytical prediction is also slightly worse. The MC generators tend to predict the measured distributions well, with Herwig++ showing slightly more disagreement than the others. The behaviour is similar for the $\beta = 3$ case.

4. Measuring the substructure of multi-prong jets

Measurements of the groomed substructure of jets from three distinct topologies are provided. Large anti-k_t $R = 1.0$ jets from topologies containing light jets from multijet events and top quark and W boson jets from $t\bar{t}$ events are groomed using the trimming [13] and soft-drop algorithms. The variables selected, the energy correlation variables and their ratios C_2 and D_2 [14] and the N-subjettiness variables and their ratios [15] are designed to discriminate between the 2 and 3 prong jets associated with hadronic W and top jets and the single pronged light quark and gluon jets. Additionally the Les Houches Angularity (LHA) [16] variable, designed to discriminate between quark and gluon jets and the distribution of the number of subjets within each jets were also measured.

A single jet trigger was used to select events with at least two jets for the multijet selection. Semi leptonic $t\bar{t}$ events are selected using a single muon trigger. The leptonic top is identified and the recoiling hadronic jet is considered as the hadronic top candidate. W jets are separated from fully reconstructed top jets by cutting on the jet mass and by selecting jets where no b-tagged small radius jets are associated with the large radius jet. A pure top jet selection is obtained by inverting this last criterion. The same cluster based systematic uncertainties are used. In this case, due to the large variety of variables that were measured a variety of additional cross checks were conducted in order to ensure that the systematic uncertainties were fully constrained. It was found the that N-subjettiness variables in the multijet selection were particularly sensitive to effects that can cause clusters to be split or merged during their construction.

The distributions are unfolded using the iterative bayesian technique, as with the previous measurement. In addition to the cluster uncertainties the uncertainty associated with the MC predictions used for unfolding is constrained in the same manner as for the soft-drop mass measurement and is again found to be significant. Other significant sources of uncertainty include the uncertainty on the kinematics of the large radius jets. The Pythia8 generator is used as the in the multijet selection while Powheg interfaced with Pythia8 is used as the nominal for the $t\bar{t}$ selection.

Figure 2 shows a few unfolded distributions, one from each of the selections. A subset of variables are shown and it can be seen that the various MC generators tested perform differently in the different regions of phase space. One important thing to mention is that the Herwig7 version (7.0.4) used here is known to not perform optimally for substructure and has been superseded. There are in total 56 unfolded distributions which prove a broad range of variables in three distinct regions of

phase space. This large number of variables will be a useful tool for MC tuning purposes. Further-more, as more analytical calculations of soft-drop groomed observables become available several of the measured distributions can be used for direct comparison [6].
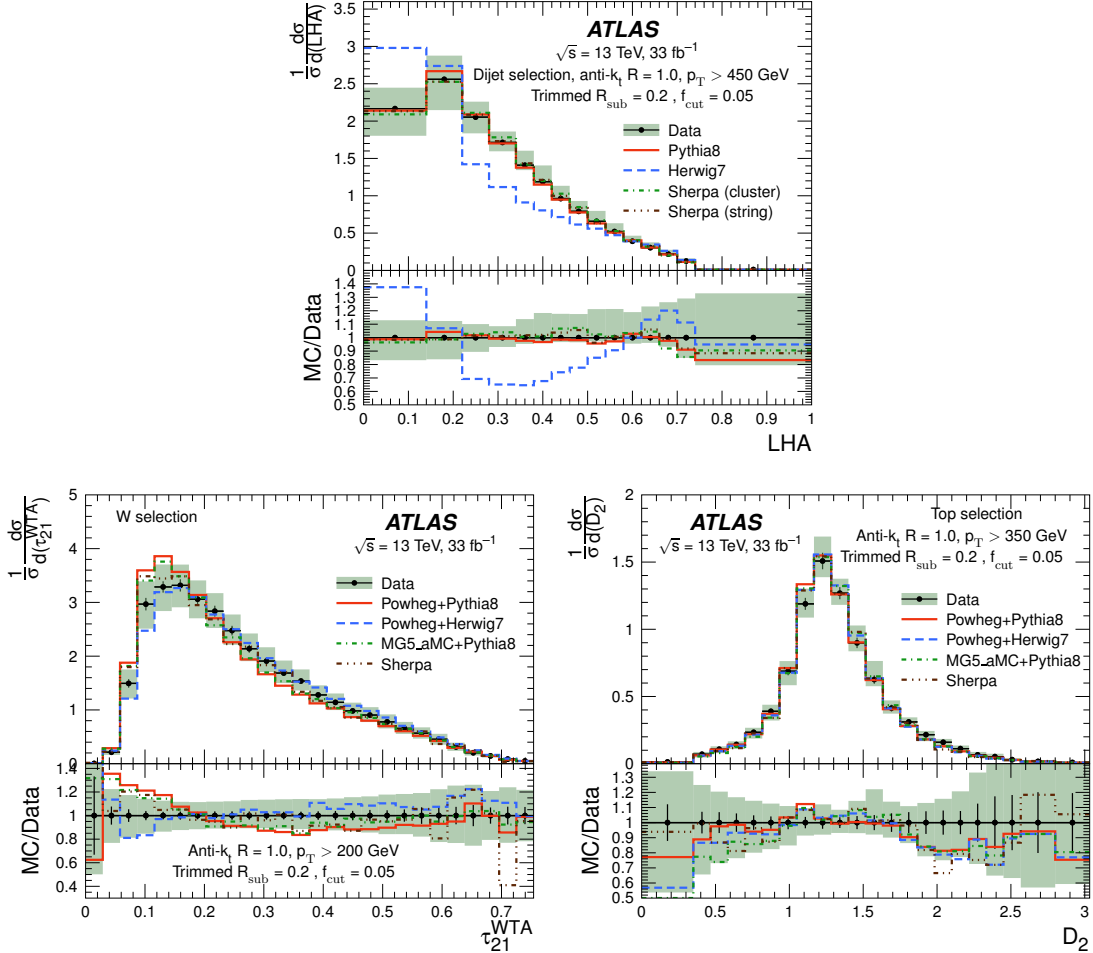
Figure 2: The unfolded LHA distribution form the multijet selection (top) is shown alongside the *N*-subjettiness ratio τ_{21} in the *W* selection (bottom left) and the D_2 distribution in the top selection (bottom right). In each case the unfolded distributions are compared to several MC generator predictions [6].

5. conclussions

New theoretical advances have motivated the above measurements, which have in turn devel-oped a range of new experimental techniques. There measurements will further facilitate jet sub-structure based MC tunes, which are currently being studied, and improve generator performance. Other MC improvements like the use of higher order parton showers can be tested using these results. The better understanding of systematic uncertainties associated with calorimeter cluster based observables will also aid any future measurement, which will potentially use a combination of reconstructed tracks and calorimeter clusters. Tools such as Rivet [17] and the HEPData repos-

itory [18] are necessary for the wider distribution of the results and preservation of the analysis methods for the wider particle physics community.

References

[1] L. Asquith et al., *Jet Substructure at the Large Hadron Collider : Experimental Review*, 1803.06991.

[2] ATLAS Collaboration collaboration, *Impact of Alternative Inputs and Grooming Methods on Large-R Jet Reconstruction in ATLAS*, Tech. Rep. ATL-PHYS-PUB-2017-020, CERN, Geneva, Dec, 2017.

[3] ATLAS collaboration, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, 1808.07858.

[4] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146 [1402.2657].

[5] ATLAS Collaboration, *Measurement of the Soft-Drop Jet Mass in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. Lett.* **121** (2018) 092001 [1711.08341].

[6] ATLAS collaboration, *Measurement of jet-substructure observables in top quark, W boson and light jet production in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Submitted to: JHEP* (2019) [1903.02942].

[7] A. J. Larkoski, I. Moult and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, 1709.04464.

[8] S. Marzani, L. Schunk and G. Soyez, *A study of jet mass distributions with grooming*, *JHEP* **07** (2017) 132 [1704.02210].

[9] S. Marzani, L. Schunk and G. Soyez, *The jet mass distribution after Soft Drop*, *Eur. Phys. J.* **C78** (2018) 96 [1712.05105].

[10] C. Frye, A. J. Larkoski, M. D. Schwartz and K. Yan, *Factorization for groomed jet substructure beyond the next-to-leading logarithm*, *JHEP* **07** (2016) 064 [1603.09338].

[11] C. Frye, A. J. Larkoski, M. D. Schwartz and K. Yan, *Precision physics with pile-up insensitive observables*, 1603.06375.

[12] G. D'Agostini, *Improved iterative bayesian unfolding*, arXiv preprint arXiv:1010.0632 (2010) .

[13] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084 [0912.1342].

[14] A. J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, *JHEP* **12** (2014) 009 [1409.6298].

[15] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [1011.2268].

[16] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer et al., *Systematics of quark/gluon tagging*, *Journal of High Energy Physics* **2017** (2017) 91.

[17] A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, J. Monk et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [1003.0694].

[18] E. Maguire, L. Heinrich and G. Watt, *HEPData: a repository for high energy physics data*, *J. Phys. Conf. Ser.* **898** (2017) 102006 [1704.05473].