Measurement of jet substructure observables using the ATLAS detector

Miguel Villaplana Perez[*][†]

Department of Physics, University of Alberta, Edmonton AB, Canada
E-mail: Miguel.Villaplana.Perez@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 a hadron collider, but also for improving the understanding of jet substructure properties that are used in many searches for physics beyond the Standard Model. In this talk, we discuss a first measurement of such jet substructure quantities. The soft drop mass is measured in dijet events with the ATLAS detector at 13 TeV, unfolded to particle-level and compared to Monte Carlo simulations. We also present a measurement of substructure variables in $t\bar{t}$ and inclusive jet events, using data collected by the ATLAS experiment at 13 TeV. 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. New event generator configurations for the modelling of jet production, derived using ATLAS data will also be presented.

---

[*]Speaker.

[†]On behalf of the ATLAS Collaboration.

1. Introduction

Jet substructure techniques are paramount to deal with highly boosted objects produced by the Large Hadron Collider (LHC) and recorded by the ATLAS experiment [1], and contribute to solidifying our understanding of the internal structure of jets and the theory of quantum chromo-dynamics (QCD). A larger fraction of heavy particles are being produced with large transverse momenta with the increasing centre-of-mass energy of proton-proton collisions of the LHC. This leads to collimated decay products which are usually reconstructed in a large-radius jet, whose internal structure shows interesting features that can be used to identify the particle that initiated the jet formation. The sensitivity of jet substructure to soft and collinear radiation allows us to test QCD in a regime where a fixed-order calculations are insufficient. A precise analytic calculation of substructure variables was historically not possible due to the presence of non-global logarithmic resummation terms related to particles radiating out of, and then into, the jet. However, a technique was found that can remove energy related to soft parton emission and is formally insensitive to non-global logarithms, Soft Drop [2]. For a given pair of constituents of a jet of radius R, the softer constituent is removed unless the following condition is met:

$$\frac{min(p_{\mathrm{T},1}, p_{\mathrm{T},2})}{p_{\mathrm{T},1} + p_{\mathrm{T},2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R} \right)^{\beta} , \tag{1.1}$$

where $p_{\mathrm{T},i}$ are the transverse momenta of the constituents, ΔR_{12} is their distance in the rapidity-azimuth plane, z_{cut} sets the scale of energy removal, and β determines the sensitivity to wide-angle radiation. Soon after the appearance of the soft drop algorithm, the distribution of the soft-drop jet mass was calculated with unprecedented precision [3]. Figure 1 shows a comparison of such calculations to data from the ATLAS collaboration [4]. In Ref. [4] the dimensionless mass parameter $\rho = m^{\mathrm{softdrop}}/p_{\mathrm{T}}^{\mathrm{ungroomed}}$ has been measured in dijet topologies, $p_{\mathrm{T},1}/p_{\mathrm{T},2} < 1.5$ for different configurations of the soft drop algorithm. Monte Carlo (MC) generators are found to do an excellent job of describing data over the entire mass range. A good agreement is seen as well between data and analytic calculations in the resummation and fixed-order regions. Including non-perturbative effects is found to improve the accuracy of the next-to-leading order + next-to-leading logarithm (NLO+NLL) prediction. The largest difference between MC and analytic predictions is found in the non-perturbative regime. As expected, this difference is larger for higher β, which configures the soft drop algorithm to remove a smaller fraction of soft energy.

2. ATLAS measurement of softdrop-groomed jet-substructure observables

Identifying the large-radius jets coming from boosted heavy particles is usually a two step procedure. In the first step, termed jet grooming, the effect of soft, uncorrelated radiation contained in the large-radius jet is reduced. Then jet substructure observables, which describe the spatial energy distribution inside the jets, are used to classify the jets originating from different particles. Most of these variables have never been measured in data, and performing a proper unfolded measurement is a common request from the theory community. Measuring these observables will help in optimising and developing current and future substructure taggers, as well as tuning hadronisation models in the important but still relatively unexplored regime of jet substructure.
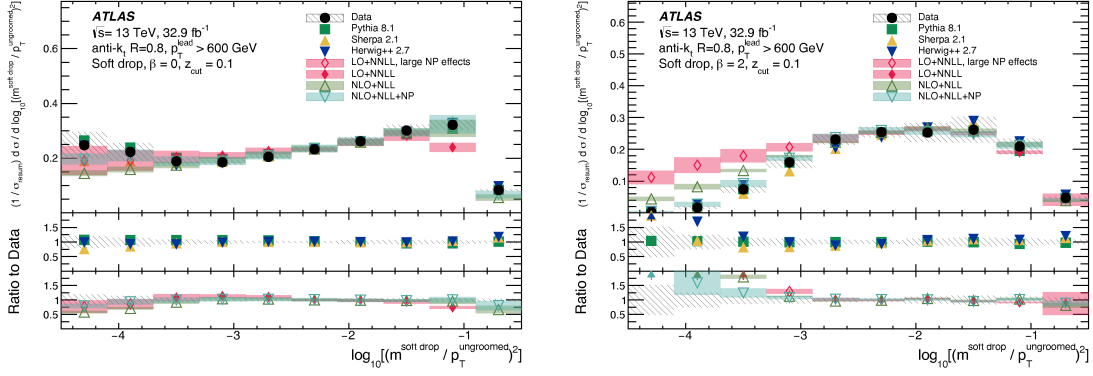
Figure 1: The unfolded $\log(\rho^2)$ distribution for anti-k_t R = 0.8 jets with $p_T^{\text{lead}} > 600$ GeV, after the soft-drop algorithm is applied for $\beta = 0$ (left) and $\beta = 2$ (right) in data compared to Pythia, Sherpa, Herwig++ particle-level simulations, and NLO + NLL and LO + NNLL theory predictions. The LO+NNLL calculation does not have non-perturbative corrections; the region where these are expected to be large is shown in a open marker (but no correction is applied), while regions where they are expected to be small are shown with a filled marker [4].

The ATLAS collaboration has performed measurements of a set of softdrop-groomed variables prioritising jet shapes commonly used in jet tagging, as well as those most useful for model tuning [5]. The measurements are performed separately for light quark events, and *W* and top jets in $t\bar{t}$ events. For both the top and *W* selections, events are collected with a set of single-muon triggers that become fully efficient for muon $p_T > 28$ GeV. The top quarks and the *W* bosons are identified from their decay products. A geometrical separation between the decay products of the two top quark candidates is required. Additional requirements are applied to separate large-radius jets containing all decay products of the top quark from those where the large-radius jet only contains the hadronic *W* boson decays, with the *b*-tagged small-radius jet reconstructed independently.

In Figures 2 and 3 the subjet multiplicity and the Les Houches angularity (LHA) inside the large-radius jets from the three different selections are compared with different MC predictions and the data. The measurement of additional observables including energy correlation functions and ratios of N-subjetiness can be found in [5]. In general, Pythia8 for light-quark/gluon large-radius jet observables, and Powheg+Pythia8, Sherpa as well as MG5 aMC@NLO+Pythia8 for top quark and *W* boson large-radius jet observables, describe the data better than other models. The different hadronisation models in Sherpa in the dijet selection result in similar predictions. For most observables, Herwig7 in the dijet selection, and Powheg+Herwig7 in the top and *W* selections do not describe the data well.

3. Multijet simulation for 13 TeV ATLAS Analyses.

Measurements of inclusive jet cross sections [6]; topological observables, such as angular distances between jets; event shape variables such as thrust or sphericity [7] and of the internal structure of jets [8] provide interesting inputs for the understanding of basic physics modelling such as the parton shower or the hadronisation model.
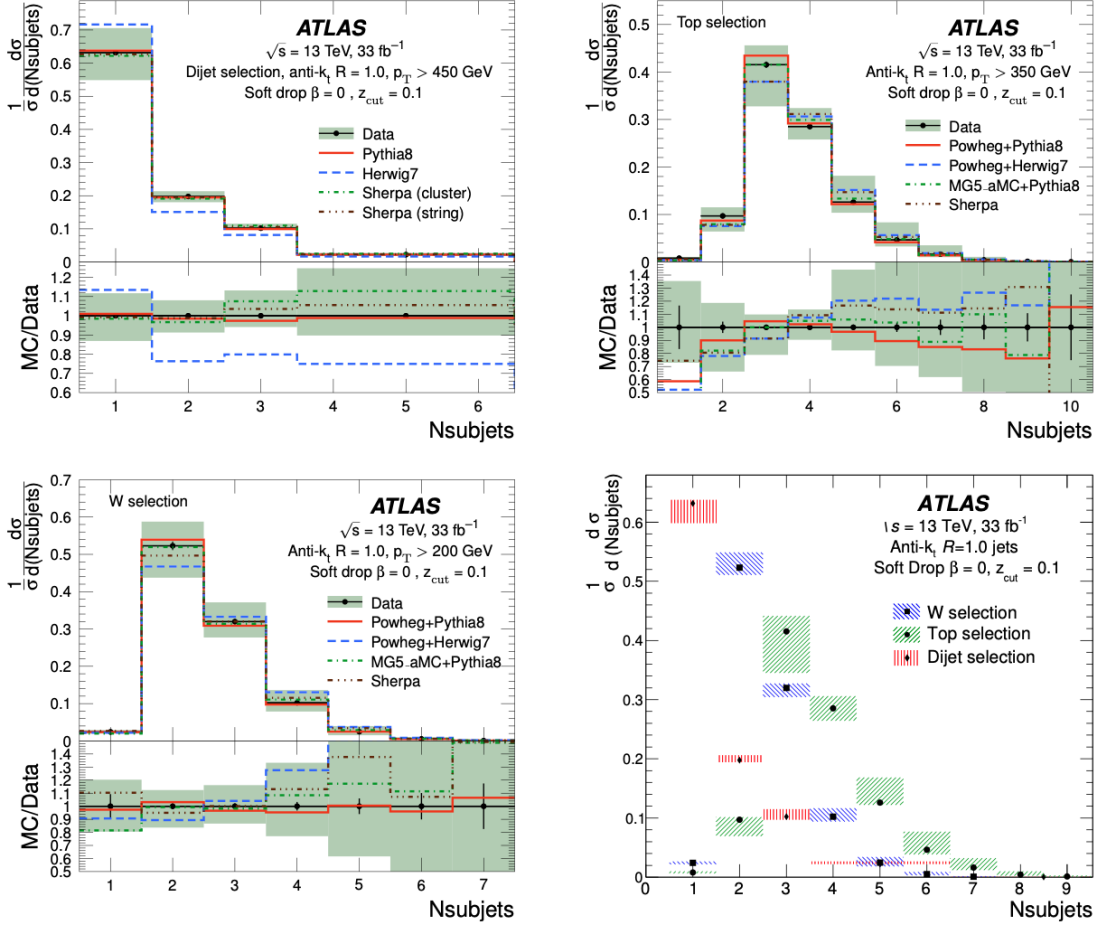
Figure 2: Subjet multiplicity distributions compared with different MC predictions for soft-dropped large-radius jets from dijet (top left), top quark (top right), and *W* (bottom left) selections. For the dijet selection, Sherpa is tested with two different hadronisation models. Data are compared between the soft-dropped large-radius jets for the three selections mentioned above (bottom right). The shaded bands represent the total uncertainty, while the error bars show the statistical uncertainty, except in the bottom right plot, where the shaded areas represent the total uncertainty [5].

The ATLAS collaboration has recently probed configurations for the Pythia8, MG5_aMC@NLO + Pythia8, Herwig7, Powheg + Pythia8 and Sherpa Monte Carlo event generators, which are used for the modelling of jet production at 13 TeV [9]. The MC predictions for the different samples are investigated in terms of the observables mentioned above and recipes for estimating uncertainties both from the perturbative and non-perturbative aspects of the calculations are discussed. In these studies, all anti-k_t jets (R = 0.4) fulfilling $p_T > 100$ GeV and $|\eta| < 2.5$ are considered.

Figure 4 shows that the difference between both hadronisation models implemented in Sherpa is, as expected, very small for the high-p_T scale of most of these studies. It is also important to note that both Herwig 7.1.3 samples give very different descriptions of the jet shapes, with the dipole parton shower giving significantly wider jets than the angular-ordered counterpart. However, it is observed that for low p_T there are significant differences in the description of substructure
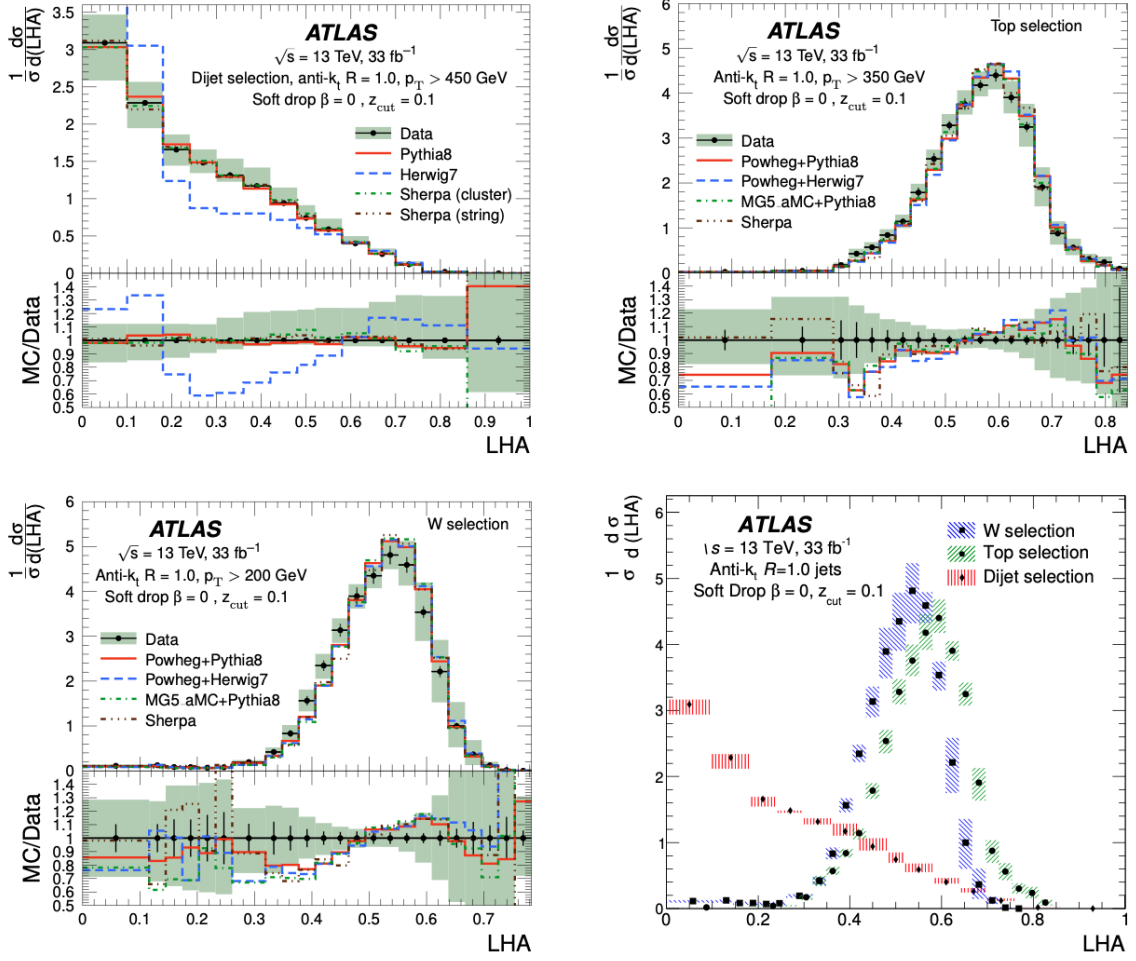
Figure 3: Les Houches angularity is compared with different MC predictions for soft-dropped large-radius jets from dijet (top left), top quark (top right), and *W* (bottom left) selections. For the dijet selection, Sherpa is tested with two different hadronisation models. Data are compared between the soft-dropped large-radius jets for the three selections mentioned above (bottom right). The shaded bands represent the total uncertainty, while the error bars show the statistical uncertainty, except in the bottom right plot, where the shaded areas represent the total uncertainty [5].

observables, such as the jet mass displayed in Figure 5.

    Moreover, large differences in the description of observables sensitive to azimuthal decorrelations, such as $\Delta\phi_{12}$ or the transverse energy-energy correlation function, are observed between Powheg and the rest of the generators studied, as can be seen in Figure 6.

References

[1] ATLAS collaboration, 2008 JINST 3 S08003.

[2] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, Jesse Thaler, JHEP05 (2014) 146.

[3] Frye, C., Larkoski, A.J., Schwartz, M.D. et al. JHEP07 (2016) 064,
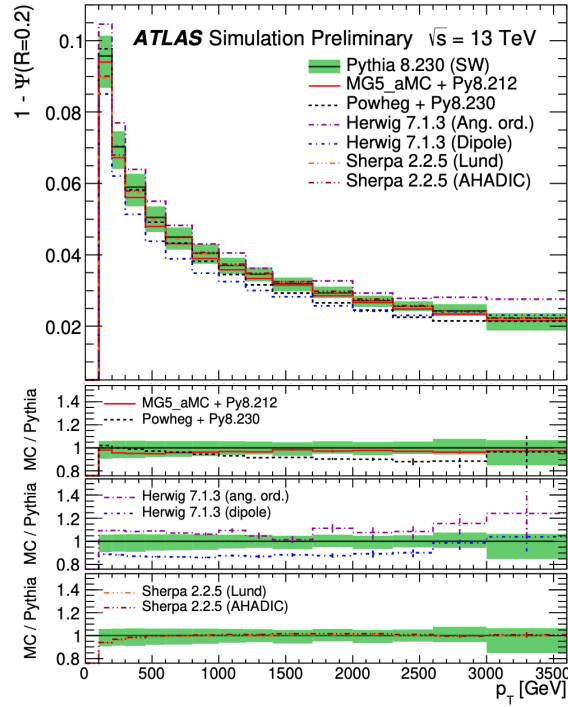Marzani, S., Schunk, L., Soyez, G. JHEP07 (2017) 132.

Figure 4: Dependence of the momentum fraction in the outer jet band, $1 - \Psi(r = 0.2)$, as a function of the jet p_T. Each subfigure shows the predictions obtained from Pythia8, MG5_aMC@NLO+Pythia8, Powheg + Pythia8, Herwig7 with both angular-ordered and dipole parton showers and Sherpa with both cluster and string hadronisation models. The green bands show the parton shower uncertainties for Pythia8. The vertical error bars represent the statistical uncertainties. The lower panels show the ratio between the generators and the Pythia8 prediction [9].

[4] ATLAS collaboration, 2018 Phys. Rev. Lett. 121, 092001.

[5] ATLAS collaboration, JHEP08 (2019) 033.

[6] ATLAS collaboration, JHEP05 (2018) 195.

[7] ATLAS collaboration, Eur. Phys. J. C 72 (2012) 2211.

[8] ATLAS collaboration, Phys. Rev. D 83 (2011) 052003.
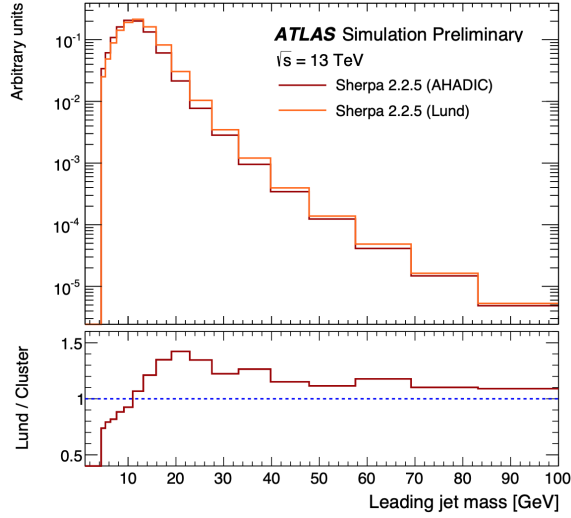
[9] ATLAS collaboration, ATL-PHYS-PUB-2019-017, https://cds.cern.ch/record/2672252.

Figure 5: The distribution of the leading jet mass in a dijet selection comparing the predictions from the Sherpa generator with the default AHADIC hadronisation model (red curve) with the Lund string model (orange curve). The lower panel shows the ratio between the two displayed predictions, while the blue dashed is there to guide the eye [9].
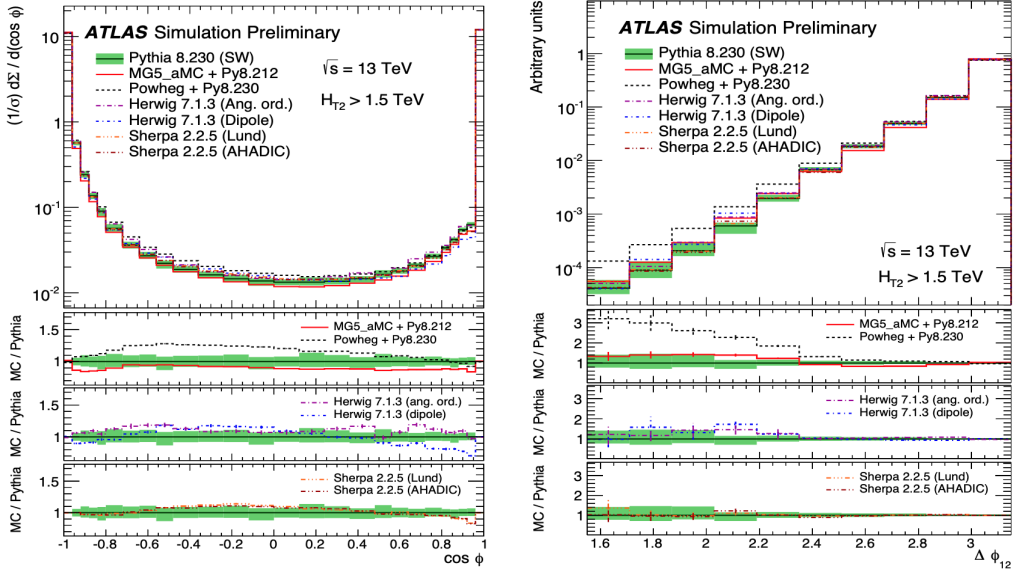

Figure 6: Comparison of the predictions on the transverse energy-energy correlation function distribution (left), and the azimuthal difference $\Delta\phi$ between the first and second jets (right) for events with $H_{T2} > 1.5$ TeV. The bottom panels show the predictions obtained from Pythia8, MG5_aMC@NLO+Pythia8, Powheg + Pythia8, Herwig7 with both angular-ordered and dipole parton showers, and Sherpa with both cluster and string hadronisation models. The green bands show the parton shower uncertainties for Pythia8. The vertical error bars represent the statistical uncertainties. The lower panels show the ratio between the generators and the Pythia8 prediction [9].