Measuring Jet Substructure Observables at the ATLAS Experiment

Diffraction and Low-x 2018,

Reggio Calabria, 26.08.18

Andrey Minaenko, IHEP (Protvino)

on behalf of the ATLAS Collaboration

Introduction

- **There are several current studies in ATLAS on jet substructure observables but by now the only one has been completed: "A measurement of the soft-drop jet mass in pp collisions at $\sqrt{s} = 13\ TeV$ with ATLAS detector", arXiv:1711.083441v2 [hep-ex].**
- **Recent advances in soft collinear effective theory have shown that jet substructure tests QCD in a regime where a fixed-order picture is insufficient to describe the relevant physical processes. General procedures now exist for an analytical understanding of infrared- and collinear-safe observables at leading-logarithm (LL) accuracy with (approximate) higher-order resummation in particular cases.**
- **A complete prediction for mass or other variables beyond LL has not been possible due to the presence of non-global logarithms (NGLs): resummation terms associated with particles that radiate out of, and then radiate back into, a jet. These terms have prevented full comparisons of observables beyond LL.**
- **However, using insights from modern analytical methods, recently it was introduced a new jet grooming procedure that is formally insensitive to NGLs. This procedure was then extended to form the soft-drop grooming algorithm. The calculation of the masses of jets that have the soft-drop procedure applied is insensitive to NGLs. The distribution of the soft-drop mass has now been calculated at both next-to-leading order (NLO) with NLL and leading order (LO) with next-to-next-to-leading-logarithm (NNLL) accuracy.**

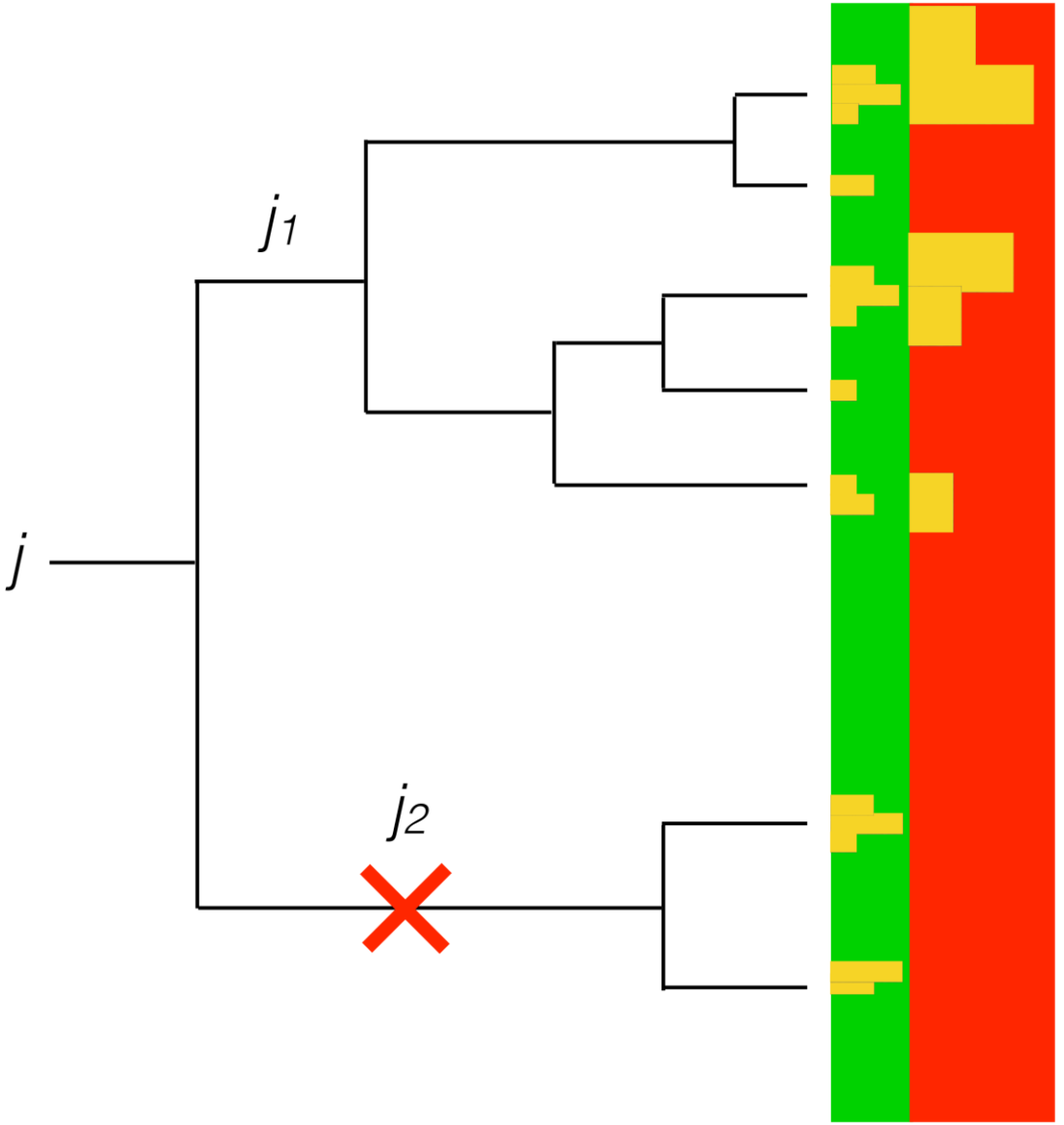
Soft Drop Procedure

- **The soft-drop procedure acts on the clustering history of a sequential recombination jet algorithm. In these algorithms, all inputs to jet-finding start as a proto-jet and are combined pairwise using a distance metric in *y*-φ space. The clustering history is the sequence of pairwise combinations that lead to a particular jet.**
- **Jets in ATLAS are usually clustered using the anti-k_t algorithm, which has the benefit of producing regularly shaped jets in *y*-φ space. The soft-drop algorithm starts by re-clustering an anti-k_t jet's constituents with the Cambridge/Aachen (C/A) algorithm which forms clustering tree with an angular-ordered structure.**
- **Next, the clustering tree is traversed from the latest branch to the earliest and at each node the following criterion is applied to proto-jets j_1 and j_2:**
  - $$\frac{min(p_{T,j_1},\, p_{T,j_2})}{(p_{T,j_1} + p_{T,j_2})} > z_{cut} \cdot \left(\frac{\Delta R_{12}}{R}\right)^{\beta},$$
- **where p_T is the momentum of a jet transverse to the beam pipe, z_{cut} (0.1) and β (0, 1, 2) are algorithm parameters, and $\Delta R_{12} = \sqrt{(\Delta y)^2 + (\Delta\varphi)^2}$ is the distance in y-φ between the proto-jets. The parameter z_{cut} sets the scale of the energy removed by the algorithm; β tunes the sensitivity of the algorithm to wide-angle radiation.**

Soft Drop Procedure

- **If the soft-drop condition is not satisfied, then the branch with the smaller p_T is removed. The procedure is then iterated on the remaining branch. If the condition is satisfied at any node, the algorithm terminates.**
- **As β increases, the fraction of branches where the condition is satisfied increases, reducing the amount of radiation removed from the jet**
- **The mass of the resulting jet is referred to as the soft-drop jet mass, $m^{soft\ drop}$**

$$\frac{min(p_{T,j_1},\ p_{T,j_2})}{(p_{T,j_1}+\ p_{T,j_2})} > z_{cut} \cdot \left(\frac{\Delta R_{12}}{R}\right)^{\beta}$$



Experimental Procedure

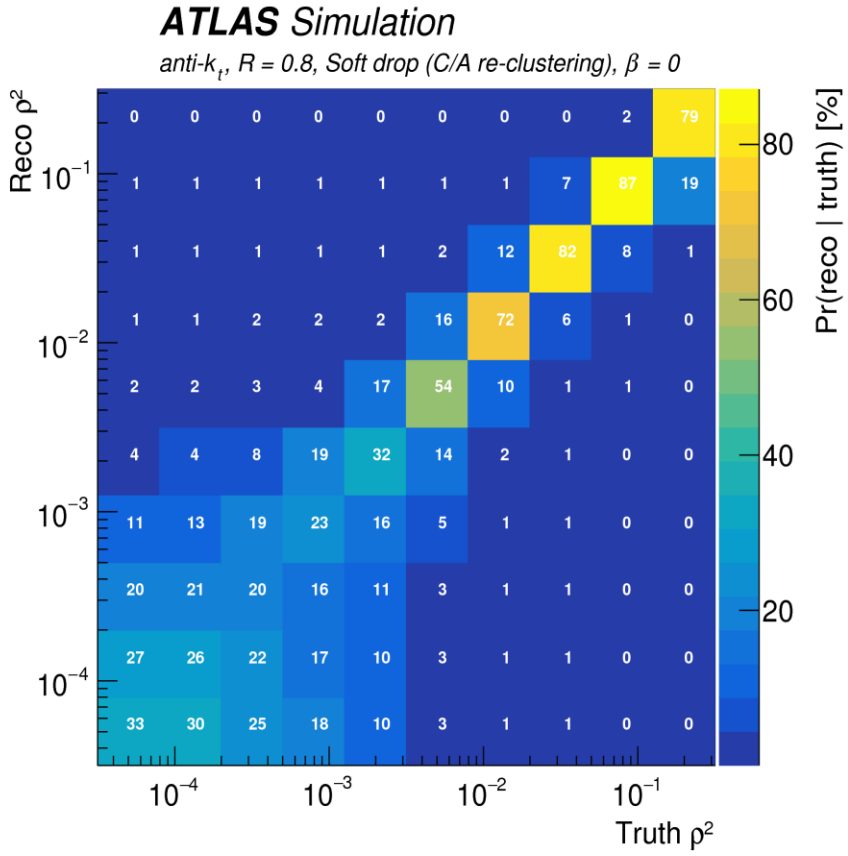
- **The study presents a measurement of the soft-drop jet mass using 32.9 fb^{-1} of $\sqrt{s}$ = 13 TeV pp data collected in 2016 by the ATLAS detector, and the first comparison to predictions of jet substructure that are formally more accurate than the LL PS approximation.**
- **ATLAS is a particle detector designed to achieve nearly a full 4π coverage in solid angle. The inner tracking detector (ID) is inside a 2 T magnetic field and is designed to measure charged-particle trajectories up to $|\eta|$ = 2.5. Surrounding the ID are electromagnetic and hadronic calorimeters, which are used mostly for the current analysis.**
- **For this study, jets are clustered using the anti-k_t jet algorithm with radius parameter R = 0.8. The inputs are topological 3D calorimeter-cell clusters calibrated using the local cluster weighting algorithm.**
- **Events were selected online using a two-level trigger system that is hardware-based at the first level and software-based for the second level. Events are required to have a minimum of two jets, at least one of which has p_T > 600 GeV. In addition, a dijet topology is imposed by requiring that the leading two-ordered jets satisfy $p_{T,1}/p_{T,2}$ > 1.5: this removes events with additional energetic jets.**

Details of the Soft Drop Procedure

- **The soft-drop algorithm is then run on the leading two jets in the selected events. Three different values of β = {0, 1, 2} are considered. The value of z_{cut} is fixed at 0.1.**
- **The dimensionless mass $\rho = m^{soft\ drop}/p_T^{ungroomed}$ is the observable of interest. ρ is a dimensionless quantity that only weakly depends on p_T.**
- **For each β value, $\log_{10}(\rho^2)$ is constructed from the jet's mass after the soft drop algorithm and its p_T before (referred to as $p_T^{ungroomed}$).**
- **Studying the distribution in log-scale allows this region to be studied more closely.**

Unfolding procedure

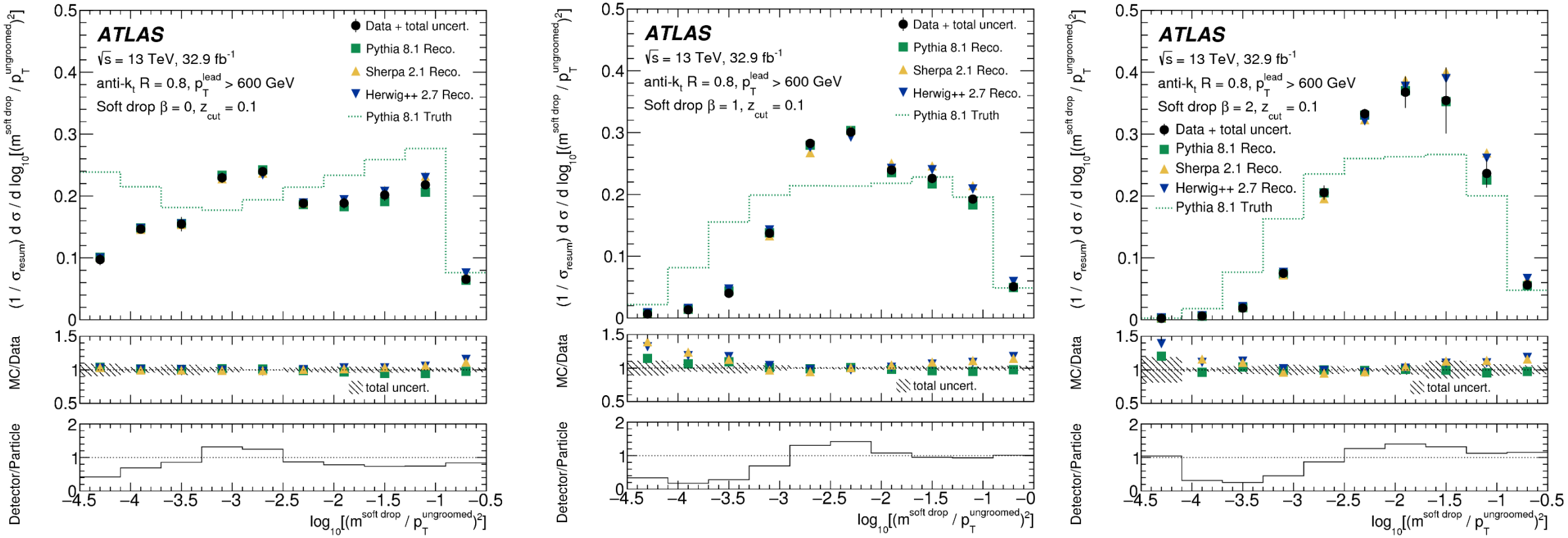
- After the event selection, the data are unfolded to correct for detector effects and obtain particle-level distributions. Monte Carlo (MC) simulations are used to perform the unfolding and for comparisons with the corrected data.
- Particle-level MC jets are built using the same algorithm as for detector-level jets, and particle-level events must pass the same dijet requirement.
- The $\log_{10}(\rho^2)$ and p_T distributions are simultaneously unfolded. After correcting for the acceptance of the event selection, the full two-dimensional distribution is unfolded using an iterative Bayesian (IB) technique with four iterations.
- An example of the folding matrix for $\log_{10}(\rho^2)$ and inclusive jet p_T. The z-axis corresponds to the probability of a reco bin being reconstructed from a jet in a particular truth bin.



Monte-Carlo (MC) samples used

- **Several MC simulations are used to unfold and compare to the data. Dijet events were generated at LO using Pythia 8.186, with the $2 \rightarrow 2$ matrix element (ME) convolved with the NNPDF2.3LO parton distribution function (PDF) set, and using the A14 set of tuned PS and underlying-event model parameters. Additional radiation beyond the ME was simulated in Pythia 8 using the LL approximation for the p_T-ordered PS.**
- **To provide several comparisons to data, additional dijet samples were simulated using different generators. Sherpa 2.1.1 generates events using multi-leg $2 \rightarrow 3$ matrix elements, which are matched to the PS following the CKKW prescription. These Sherpa events were simulated using the CT10 PDF set and the default Sherpa event tune.**
- **Herwig++ 2.7.1 events were generated with the $2 \rightarrow 2$ matrix element, convolved with the CTEQ6L1 PDF set and configured with the UE-EE-5 tune. Both Sherpa and Herwig++ use angular ordering in the PS and a cluster model for hadronization.**
- **All MC samples use Pythia 8 minimum bias events (MSTW2008LO PDF set and A2 tune) to simulate pileup.**
- **The MC samples were processed using the full ATLAS detector simulation based on Geant4.**

Comparison with uncorrected data

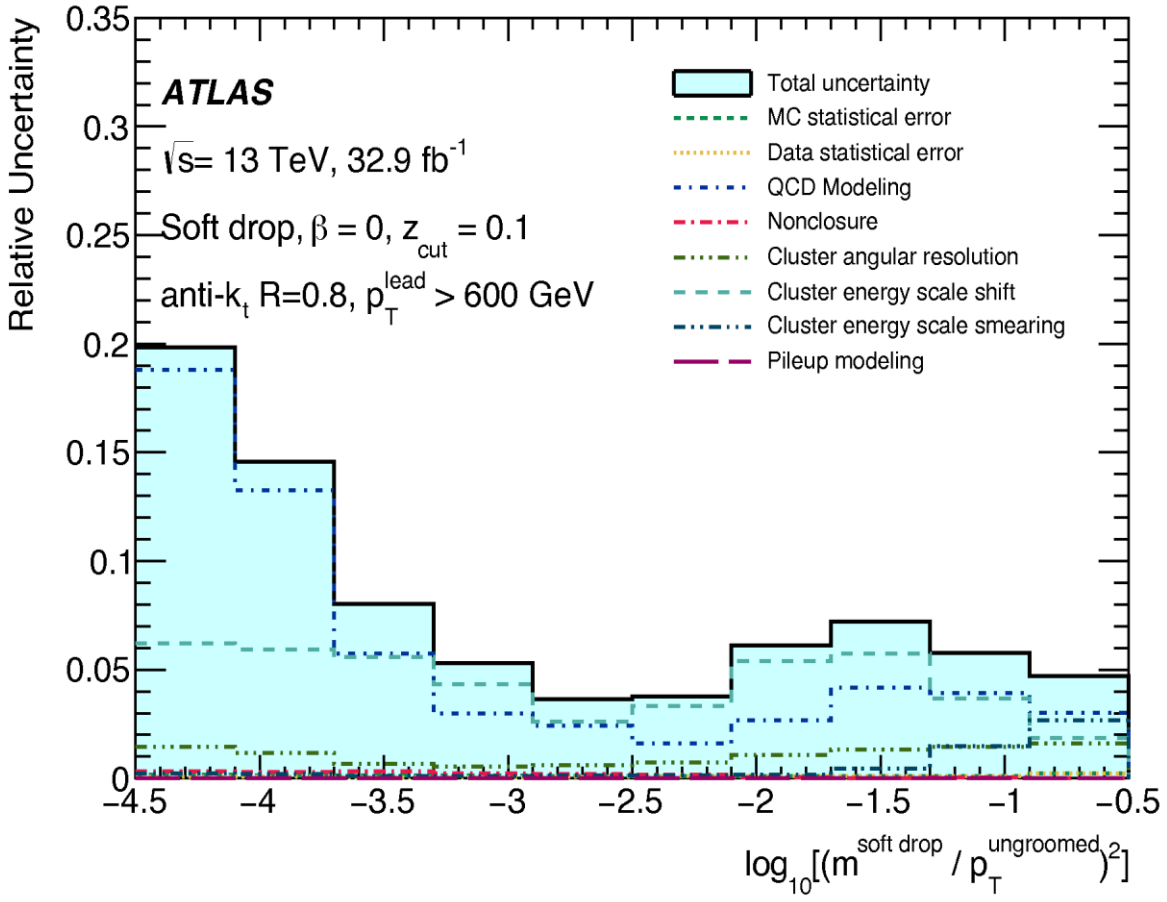


- **Distributions of $\log_{10}(\rho^2)$ in data compared to reconstructed detector-level Pythia, Sherpa, and Herwig++ and particle-level (Truth) Pythia simulations for β = 0, 1, 2. The ratio of the three detector-level MC predictions to the data is shown in the middle panel, and the size of the detector→ particle-level corrections for Pythia is shown as the ratio in the bottom panel. The error bars on the data points and in the first ratio include the experimental systematic uncertainties in the cluster energy, angular resolution, and efficiency. The distributions are normalized to the integrated cross-section, σ_{resum}, measured in the resummation region, $-3.7 < \log_{10}(\rho^2) < -1.7$. There are substantial migrations between the detector- and particle-level distributions, which cause large off-diagonal terms in the unfolding matrix especially at low values of $log_{10}(\rho^2)$.**

Systematic Uncertainties

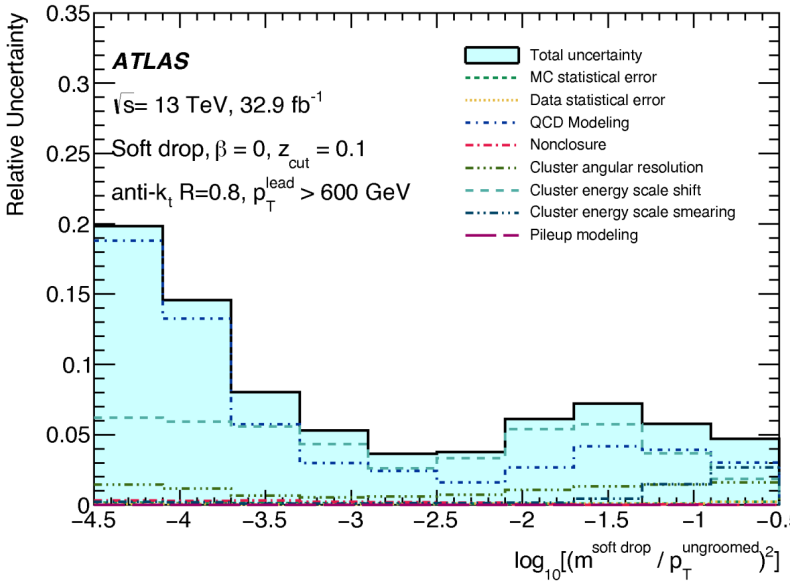
- **Experimental uncertainties are due to limitations in the accuracy of the reconstruction of calorimeter-cell cluster energies and positions as well as their reconstruction efficiency, and are evaluated as follows. Isolated calorimeter-cell clusters are matched to tracks; the mean and standard deviation of the energy-to-momentum ratio *(E/p)* is used for the cluster energy scale and resolution uncertainties, and the standard deviation of the relative position is used for the cluster angular resolution.**
- **The reconstruction efficiency is studied using the fraction of tracks without a matched calorimeter-cell cluster.**
- **A series of validation studies are performed to ensure that these uncertainties are valid also for non-isolated clusters.**
- **One of the dominant uncertainties is due to the theoretical modelling of jet fragmentation (QCD modelling). In particular, as dijet simulation is used to unfold the data, the results of the analysis are sensitive to the choice of MC generator used for this procedure. The Pythia generator is used for the nominal sample, and comparisons are made with Sherpa and Herwig++. The impact of this uncertainty is assessed by unfolding the data with the alternative response matrix.**

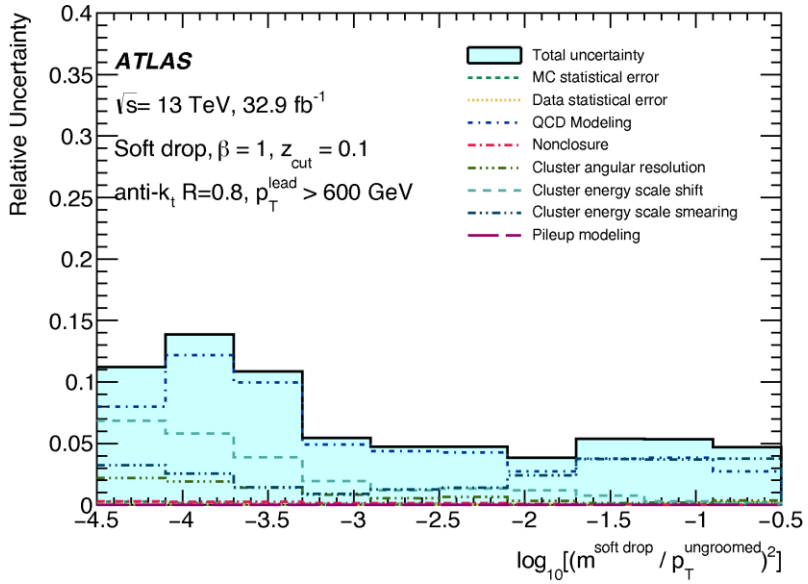
Systematic Uncertainties, β = 0

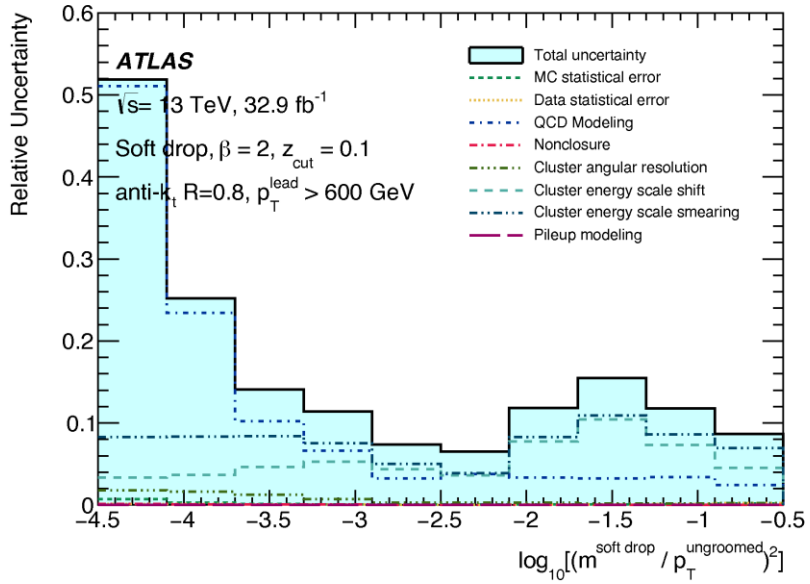


- A summary of the relative sizes of the various systematic uncertainties for β = 0 shown in the figure.
- The uncertainties are dominated by QCD modelling and the cluster energy scale. The former are largest ($\lesssim 20\%$) at low $log_{10}(\rho^2)$ where non-perturbative effects introduce a sensitivity to the $log_{10}(\rho^2)$ distribution prior, and are $\lesssim 10\%$ for the rest of the distribution.
- Cluster energy uncertainties are large ($\lesssim 5\%$) at low $log_{10}(\rho^2)$ where the cluster multiplicity is low and also at high $log_{10}(\rho^2)$ where the energy of the hard prongs, rather than their opening angle, dominates the mass resolution.
- Other sources of uncertainty are typically below 5% across the entire distribution.

Systematic Uncertainties, Different β = 0, 1 ,2

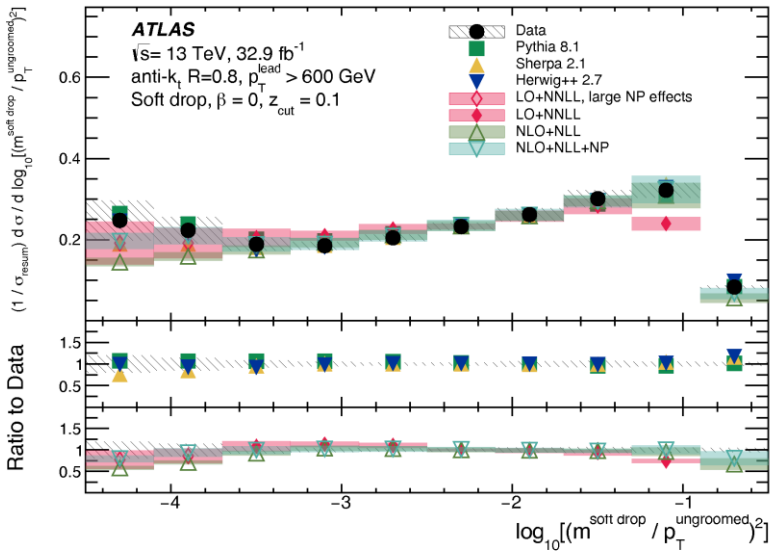


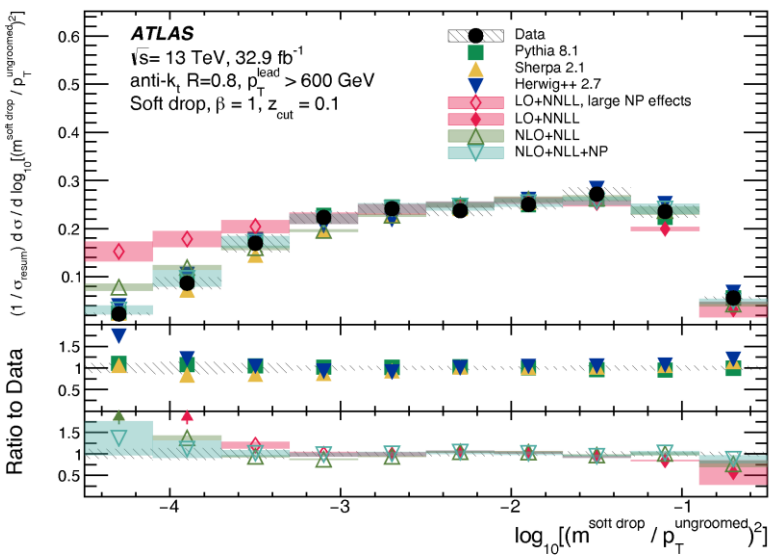


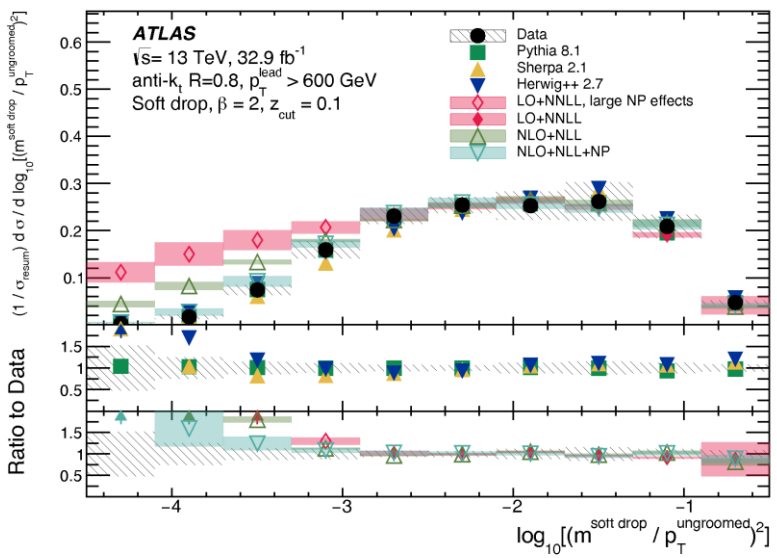


- **The relative sizes of the different sources of systematic uncertainty are similar for β = 1 and β = 2, except that the large uncertainty at low $log_{10}(\rho^2)$ values spans a larger range.**

Results

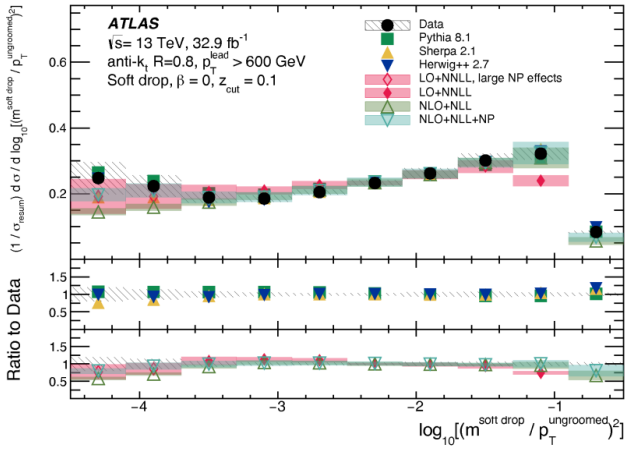


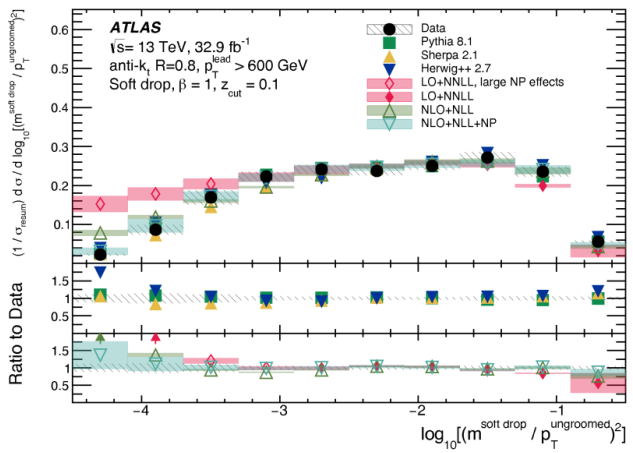


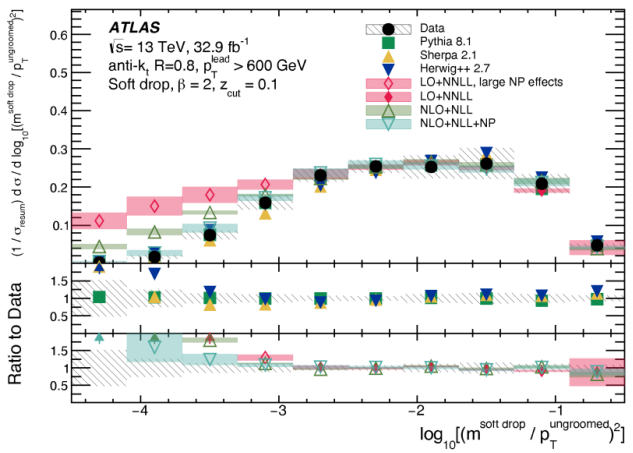


- The unfolded $log_{10}(\rho^2)$ distribution for anti-k_t R=0.8 jets with p_T^{lead} > 600 GeV, after the soft drop algorithm is applied for β = {0, 1, 2}, in data compared to Pythia, Sherpa, and Herwig++ particle-level, and NLO+NLL+NP and LO+NNLL theory predictions.
- The LO+NNLL calculation does not have non-perturbative (NP) corrections; the region where these are expected to be large is shown in a open marker, while regions where they are expected to be small are shown with a filled marker.
- The distributions are normalized to the integrated cross section, σ_{resum}, measured in the resummation region, -3.7 < $\log_{10}(\rho^2)$ < -1.7.

Results

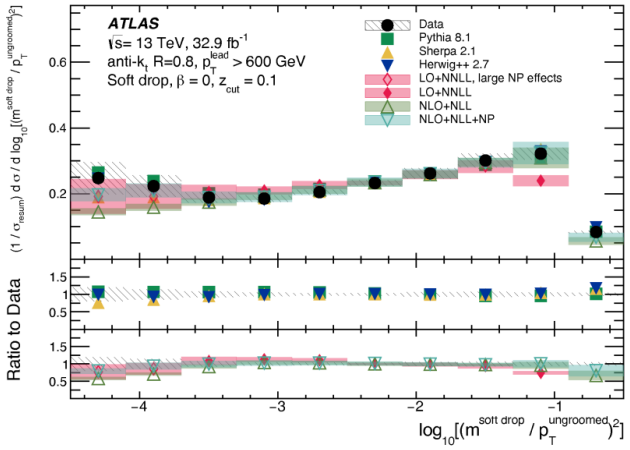


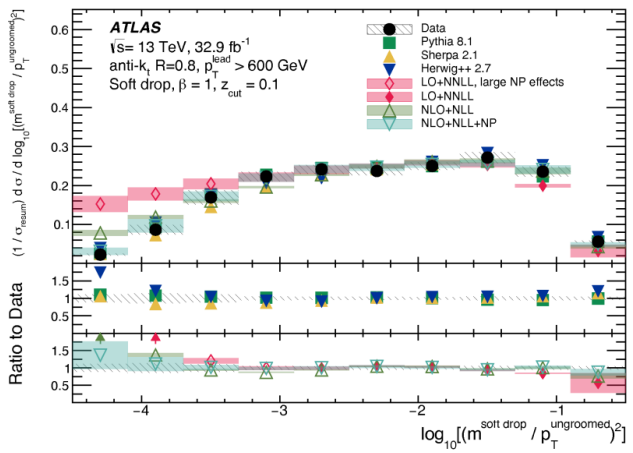


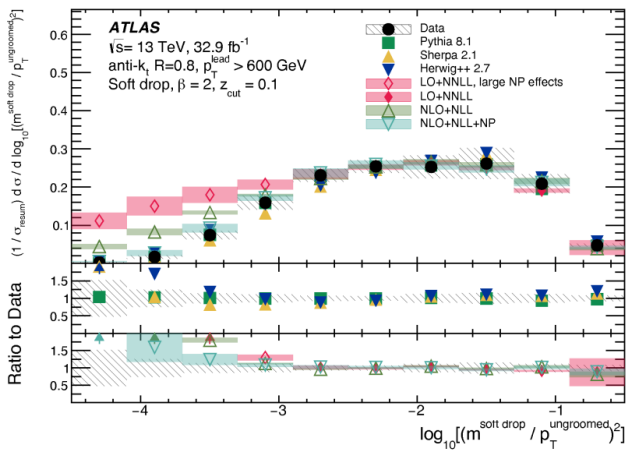


- **In general, non-perturbative effects are large for $log_{10}(\rho^2)$ < -3.7 (where small-angle or soft gluon emissions dominate) and small for -3.7 < $log_{10}(\rho^2)$ < -1.7 where resummation dominates.**
- **Fixed, higher-order corrections are expected to be important for $log_{10}(\rho^2)$ > -1.7, where large-angle gluon emission can play an important role.**
- **This implies that the region -3.7 < $log_{10}(\rho^2)$ < -1.7 (the resummation region) should have the most reliable predictions for both the MC generators and the LO+NNLL analytical calculation, while the NLO+NLL calculation should also be accurate for $log_{10}(\rho^2)$ > -1.7.**
- **For all values of β, the measured and predicted shapes agree well in the resummation region, and the data and NLO+NLL prediction continue to agree well at higher values of $log_{10}(\rho^2)$.**

Results

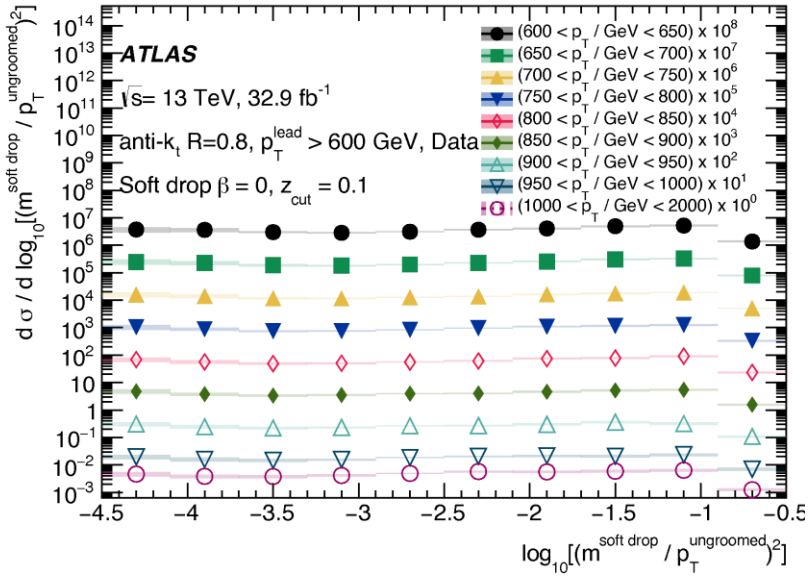


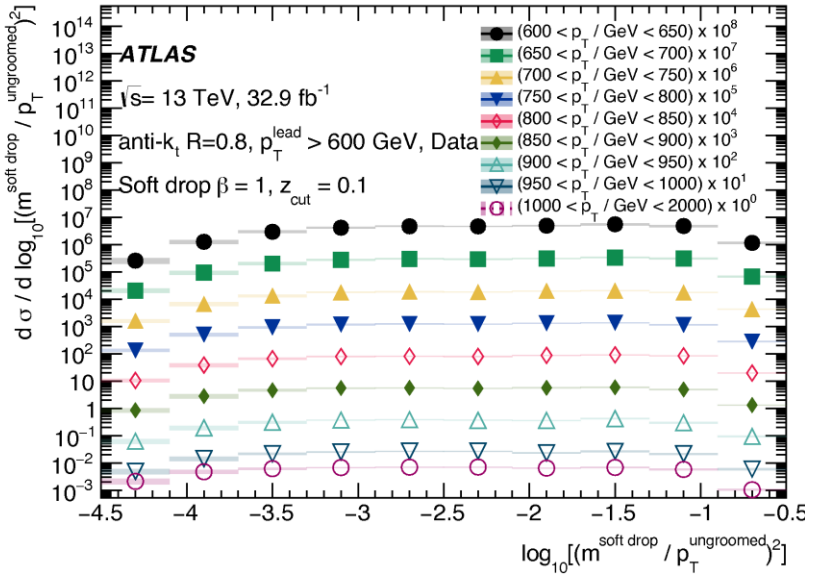


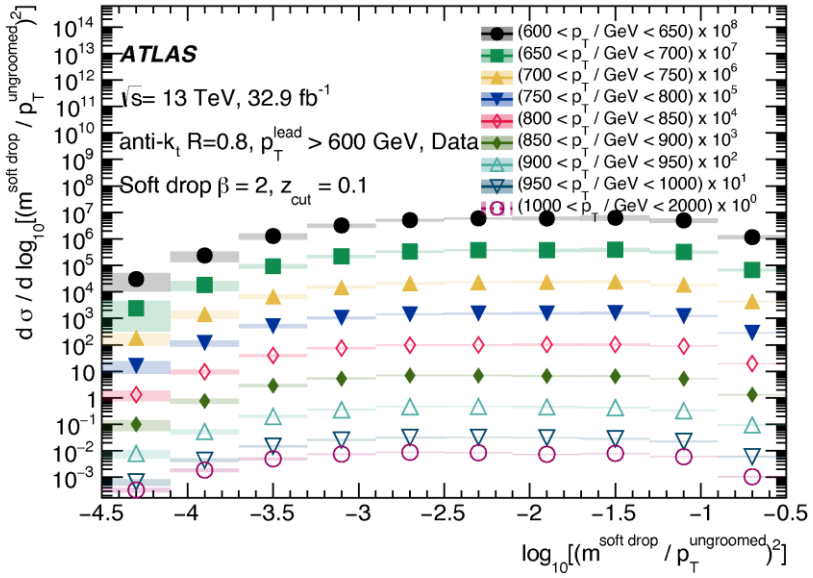


- **At more negative values of $log_{10}(\rho^2)$, non-perturbative effects lead to distinctly different predictions between the MC generators and the LO+NNLL calculation; the data fall below the predictions for all β values.**
- **The NLO+NLL calculation contains NP corrections and continues to agree well for more negative values of $log_{10}(\rho^2)$.**
- **As β increases, the fraction of radiation removed by soft-drop grooming decreases and the impact of non-perturbative effects grows larger, so the range over which the analytical calculations are accurate also decreases.**
- **The degree of agreement between data and all the calculations for $log_{10}(\rho^2)$ < -3 does substantially worsen for β = {1, 2}, especially when NP corrections are not included.**
- **Agreement between the data and the MC generators remains generally within uncertainties for all values of β.**

p_T-dependence







- **The unfolded $log_{10}(\rho^2)$ distribution in data for each of the $p_T$bins in the analysis, overlaid with a multiplicative factor indicated in the legend.**
- **Covered p_T region spreads from 600 up to 2000 GeV.**
- **As expected, there is no strong dependence of the shape on p_T.**

Summary

- **In summary, a measurement of the soft-drop jet mass is reported.**
- **The measurement provides a comparison of the internal properties of jets between 32.9 fb collision data collected by the ATLAS detector at the LHC and precision QCD calculations accurate beyond leading logarithm.**
- **Where the calculations are well defined perturbatively, they agree well with the data; in regions where non-perturbative effects are expected to be significant, the calculations disagree with the data and the predictions from MC simulation are better able to reproduce the data.**
- **A normalized fiducial dijet differential cross section is presented as a function of the $log_{10}(\rho^2)$, allowing the results to be used to constrain future calculations and MC generator predictions.**