Measurement of jet substructure observables using the ATLAS detector

EPS-HEP 2019
Ghent, Belgium

Miguel Villaplana
INFN

on behalf of the ATLAS Collaboration

July 12, 2019

Outlook

- **The Soft Drop algorithm**
- **Measurement of the soft-drop jet mass**
- **Measurement of substructure variables in ttbar and inclusive jet events**
- **New event generator configurations for the modelling of jet production**

Jet substructure

- Jet substructure techniques are paramount to deal with boosted objects in the LHC

- New measurements of jet substructure are also solidifying our understanding of the internal structure of jets and the theory of QCD

  - k_T splitting scales in $Z \rightarrow ll$ events at 8 TeV with the ATLAS detector. (JHEP08 (2017) 26)
  - Colour flow using jet-pull observables in $t\bar{t}$ events at 13 TeV. (Eur. Phys. J. C 78 (2018) 847)
  - **Soft-drop jet mass at 13 TeV. (Phys. Rev. Lett. 121, 092001)**
  - **Measurement of substructure variables in ttbar and inclusive jet events. (arXiv:1903.02942)**
  - **Multijet simulation for 13 TeV ATLAS analyses. (ATL-PHYS-PUB-2019-017)**

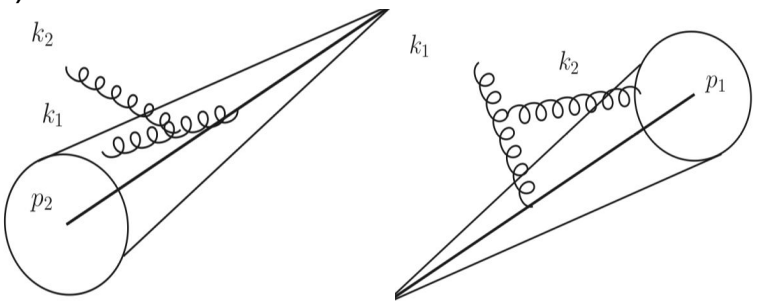
- Non-negligible differences are observed between data and MC simulations

- Can constrain both analytic calculations in perturbative regime and soft-hadronic activity in non-perturbative region

Soft Drop

- Jet substructure tests QCD in a regime where a fixed-order is insufficient
  - Sensitive to soft and collinear radiation

- A precise analytic calculation of substructure variables (beyond leading log) not possible due to the presence of non-global logarithmic resummation terms (NGLs).
  - Related to particles radiating out of and then into jet



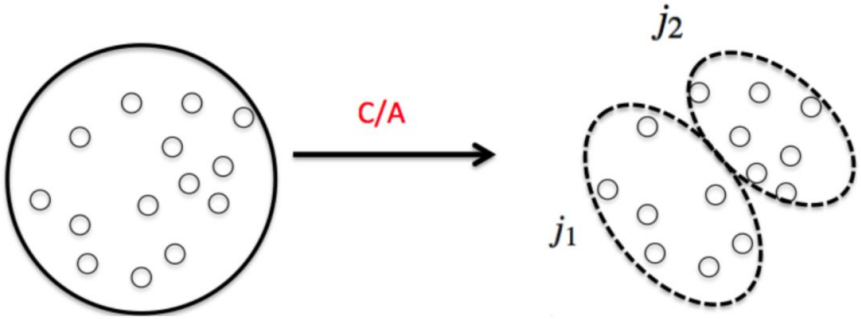
- A perfect example is jet mass

- Soft drop. JHEP 1405 (2014) 146
  - Jet grooming procedure that removes energy related to soft parton emission and pile-up
  - Formally insensitive to NGLs

- The distribution of the soft-drop mass has now been calculated at
  - NLO with NLL. JHEP07(2016)064
  - LO with NNLL. JHEP07(2017)132

The Soft Drop algorithm

- Take a jet, re-cluster its constituents with C/A, and go backwards in the C/A clustering sequence



- If $\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1}+p_{T,2}} > z_{cut}\left(\frac{\Delta R_{12}}{R}\right)^{\beta}$ then the jet is a soft drop jet.

- Otherwise, the highest p_T sub-jet is taken as a new candidate and the procedure is iterated.

- z_{cut} sets the scale of energy removal. Higher z_{cut} means more energy removed by grooming.

- β determines the sensitivity to wide-angle radiation.
  - Larger β means smaller fraction of soft small-angle radiation removed -> less grooming.

ATLAS measurement of the soft-drop jet mass

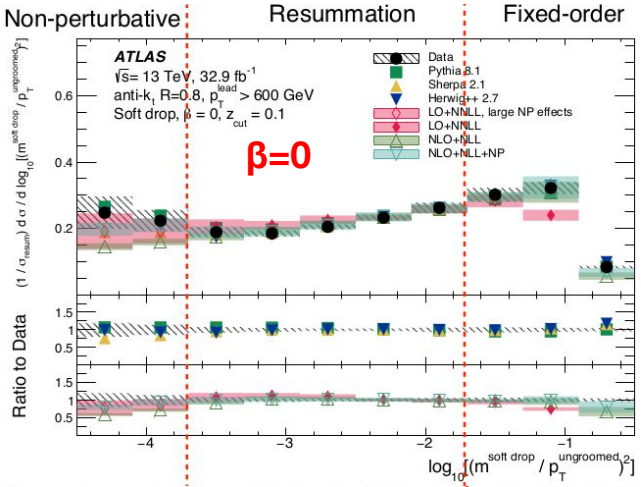
Phys. Rev. Lett. 121, 092001 (2018)

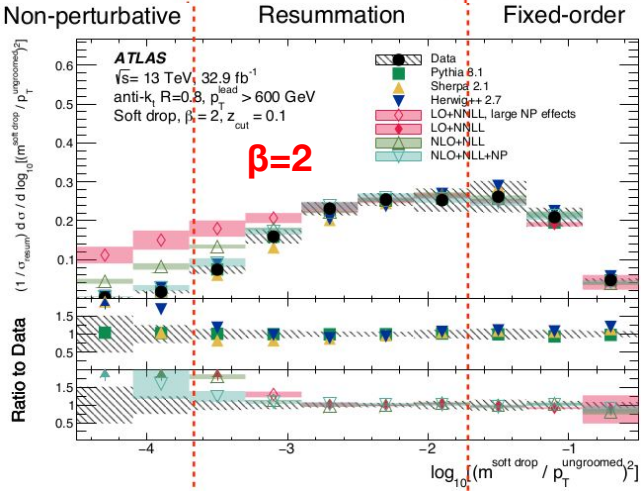
Soft-drop jet mass measurement

- Using anti-k_T R=0.8 jets built from locally calibrated calorimeter-cell clusters
- Lowest un-prescaled trigger (400 GeV) and $p_{T,1}$> 600 GeV
- Dijet topologies: $p_{T,1}/p_{T,2} < 1.5$ for two leading jets
- Measuring dimensionless mass parameter, $\rho = m_{softdrop} / p_T^{ungroomed}$
  - Weak dependence on p_T
  - Distribution of $\log_{10}(\rho^2)$ studied for β =0,1,2 and $z_{cut} = 0.1$
- Simultaneously unfolding in $\log_{10}(\rho^2)$ and jet p_T distributions using Pythia LO predictions
- Three distinct regions
  - Non-perturbative region: $\log_{10}(\rho^2) < -3.7$ (soft and collinear emissions)
  - Resummation region: $-3.7 < \log_{10}(\rho^2) < -1.7$ (resummation dominates)
  - Fixed-order region: $\log_{10}(\rho^2) > -1.7$ (wide-angle hard gluon emissions)

Results

- Distributions normalised to data in resummation region
- MC generators do an excellent job of describing data over entire mass range
- Good agreement between data and analytic calculations in resummation and fixed-order regions
- Largest difference between MC and analytic predictions in non-perturbative regime
  - Effect larger for higher β (smaller fraction of soft energy removed)
- Including non-perturbative effects improves the accuracy of the NLO+NLL prediction





ATLAS measurement of jet-substructure observables in ttbar and inclusive jet events

arXiv:1903.02942

Overview

- JSS variables are used extensively for top/W/Z and Higgs tagging
- No ATLAS unfolded measurement of variables at 13 TeV
- Useful for MC model and tagger developments
- A set of softdrop-groomed variables can be calculated analytically
  - strong theoretical motivation
- Large-radius jet substructure for: W, top, and light quark jets!
  - Semi-leptonic selections for W and top jets in ttbar events
  - Dijet selection for light quark jets in QCD events
- 8 observables for top, W and QCD jets

Observables

- **(generalised) N-subjetiness**

$$\tau_N^{(\beta)} = \frac{1}{d_0} \sum_i p_{Ti} \min\left\{(\Delta R_{1,i})^\beta, (\Delta R_{2,i})^\beta \ldots (\Delta R_{N,i})^\beta\right\}$$

  - A set of N subjet axis are defined using the exclusive k_t algorithm.

- **Normalised energy correlation functions (and ratios)**

$$e_2^\beta = \frac{1}{p_{T\,J}^2} \sum_{i<j\in J} p_{Ti} p_{Tj} R_{ij}^\beta \qquad e_3^\beta = \frac{1}{p_{T\,J}^3} \sum_{i<j<k\in J} p_{Ti} p_{Tj} p_{Tk} R_{ij}^\beta R_{jk}^\beta R_{ik}^\beta$$

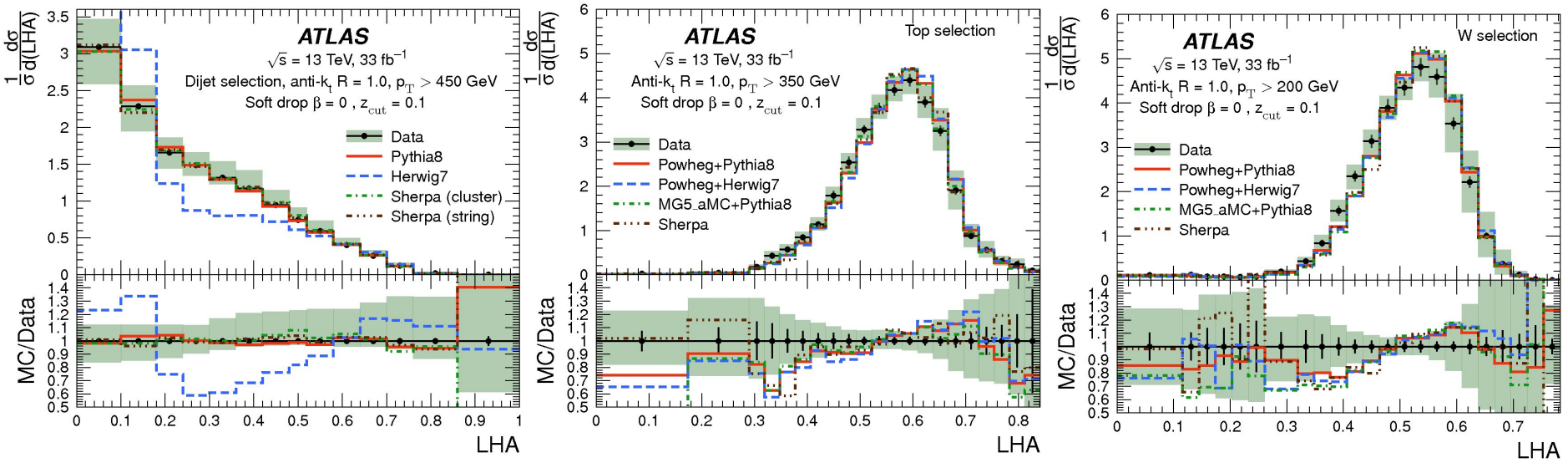
  - Subjet-independent way to discriminate between a “two-pronged” jet and a single prong parton initiated jet.
  - Ratio of energy correlation functions: $D_2^\beta = \frac{e_3^\beta}{\left(e_2^\beta\right)^3}, C_2^\beta = \frac{e_3^\beta}{\left(e_2^\beta\right)^2}$

- **Les Houches angularity (LHA)**
  - Used in various quark/gluon discrimination studies with κ=1 and β = 0.5

$$\lambda_\beta^\kappa = \sum_{i\in jet} \left(\frac{p_{Ti}}{\sum_{i\in jet} p_{Ti}}\right)^\kappa \left(\frac{R_{i\hat{n}}}{R}\right)^\beta$$

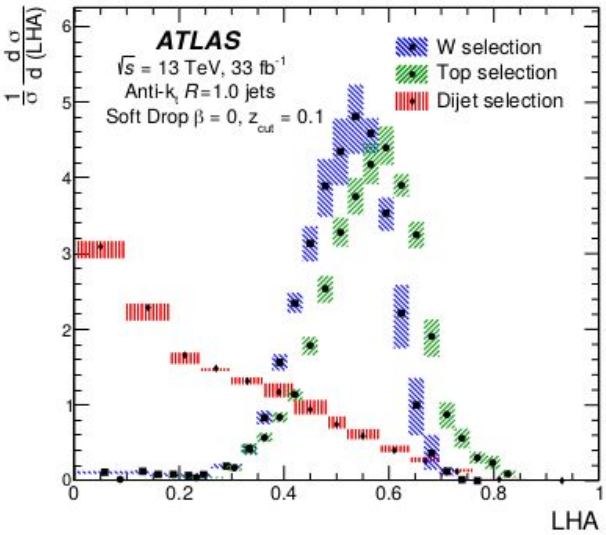
Results

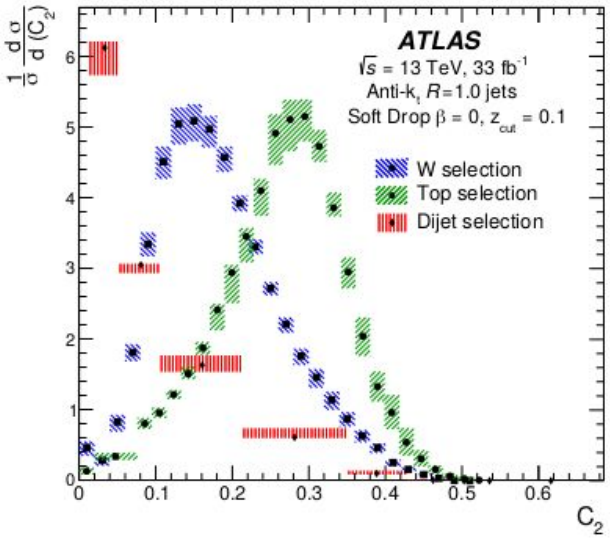
- Unfolded to particle level with Bayesian method



- Pythia8 does best for light quark/gluon observables
- Herwig7 very different from data
- No meaningful differences seen between different hadronization models used in Sherpa
- Shifted peak suggest excess gluon radiation in predictions

Results





- Important discrimination for BSM physics tagging
- Clear differences in boosted jet shapes and substructures
- Detailed measurements and MC tuning essential for optimizing physics performance of both measurements and searches

Multijet simulation for 13 TeV ATLAS Analyses

ATL-PHYS-PUB-2019-017

Overview

- Measurement of the inclusive jet spectrum at 13 TeV
  - provide interesting inputs for the understanding of basic physics modelling features such as the parton shower or the hadronization model

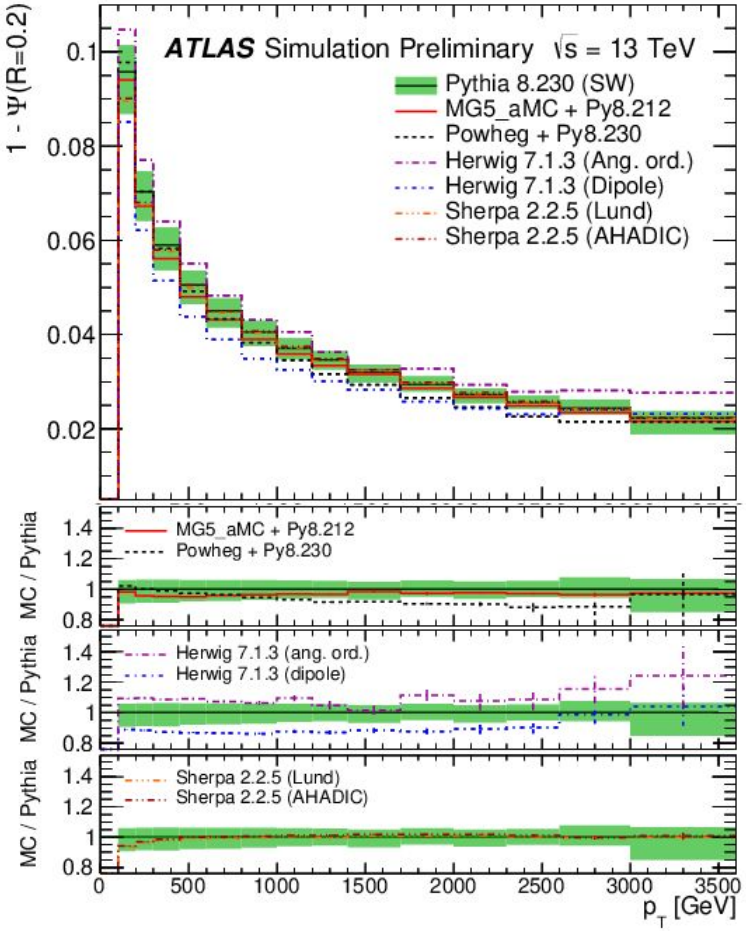
- Measurement in distributions of global event observables and jet substructure variables

- Compared to different MC samples:
  - Pythia 8.230
  - Madgraph + Pythia 8.212
  - Herwig 7.1.3
    - PS: angular-ordered and dipole
  - Powheg + Pythia 8.230
  - Sherpa 2.2.5
    - Hadronization: cluster fragmentation (AHADIC) and string fragmentation (Lund)

- For this analysis, all jets (R=0.4) fulfilling $p_T > 100$ GeV and $|\eta| < 2.5$ are considered.

Results

- Fraction of p_T in the outer band of the jets (r=0.2)
  - The green bands show the PS uncertainties for Pythia
  - The vertical error bars represent the stat. uncertainties
- MG5_aMC@NLO and Powheg predict narrower jets than Pythia 8
  - but compatible within the Pythia 8 shower uncertainties
- Herwig:
  - dipole parton shower predicts systematically narrower jets than Pythia 8
  - angular-ordered showers give, on average, wider energy distributions inside the jet cone
- Both Sherpa samples provide a very similar prediction to that of Pythia 8



Conclusions

- Jet substructure studies are essential to find new physics in post-Higgs era
- Proper estimation of uncertainties, and robustness against pile-up is critical
- Need measurements, and best possible MC modeling

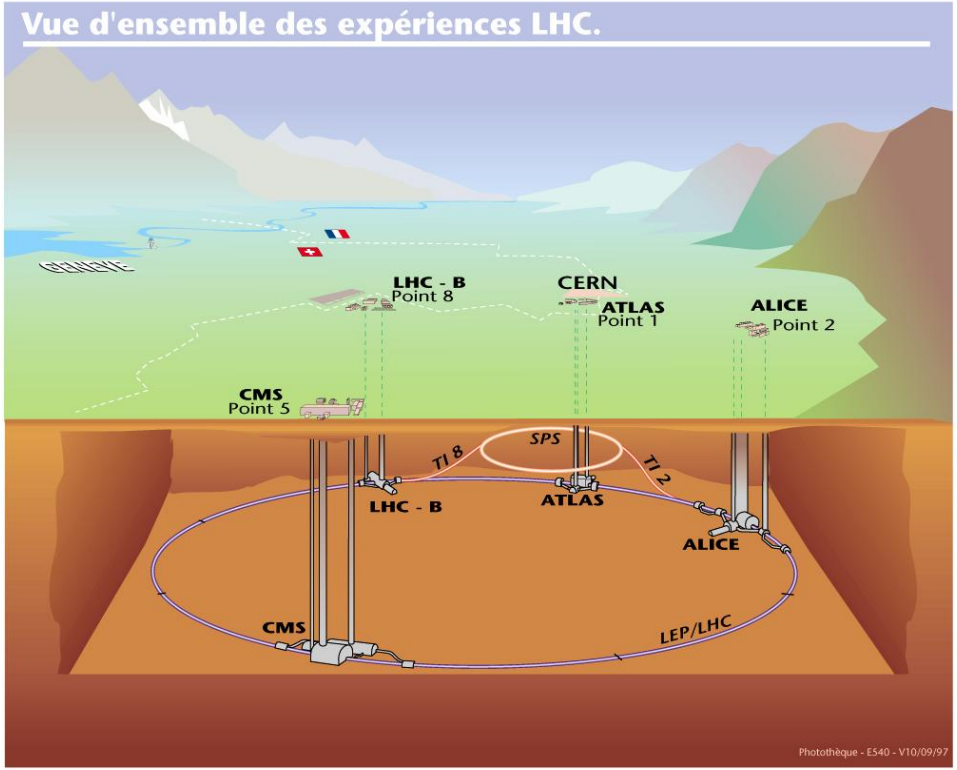
Presented recent ATLAS measurements of substructure observables

- **Soft-drop jet mass at 13 TeV.** (Phys. Rev. Lett. 121, 092001)
  - Good agreement between data and calculations in resummation and fixed-order regions
  - MC generators do better in non-perturbative region
  - Results to be used to constrain future calculations and MC generator predictions
- **Measurement of jet substructure observables at 13 TeV.** (arXiv:1903.02942)
  - Uncertainties small enough to have discriminating power between the various MC models
    - useful to improve modeling
  - Help understand limitations of searches with boosted topologies
  - Improve the design and performance of tagging algorithms
- **Multijet simulation for 13 TeV ATLAS Analyses.** (ATL-PHYS-PUB-2019-017)
  - Significant differences in the description of substructure observables such as the jet mass found at low p_T in both hadronization models implemented in Sherpa.
  - Both Herwig samples give very different descriptions of the jet shapes
    - the dipole parton shower giving significantly wider jets than the angular-ordered counterpart

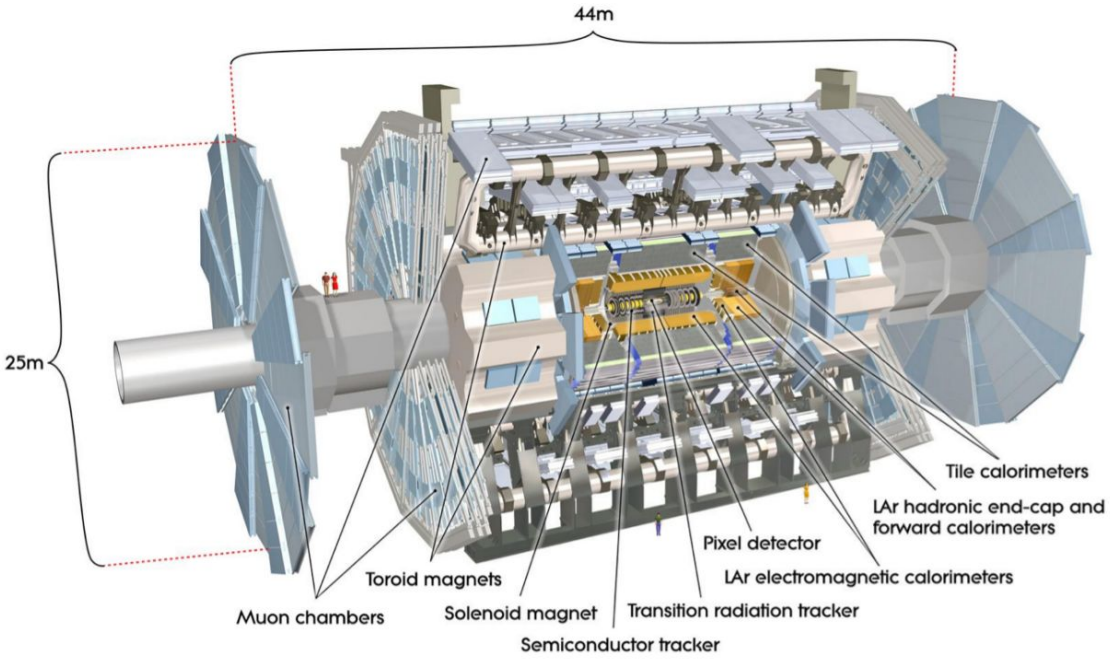
Thanks

home.infn.it

The Large Hadron Collider (LHC)

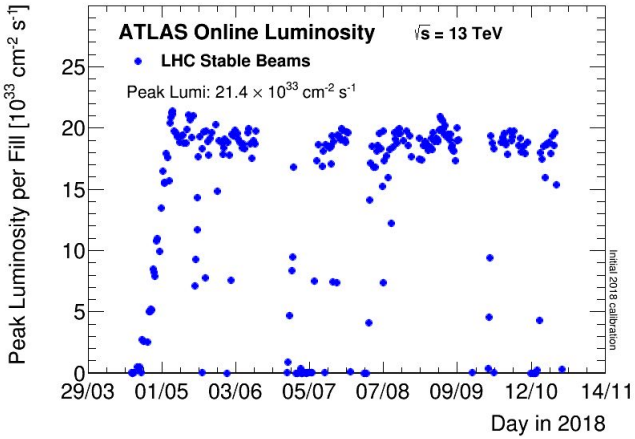


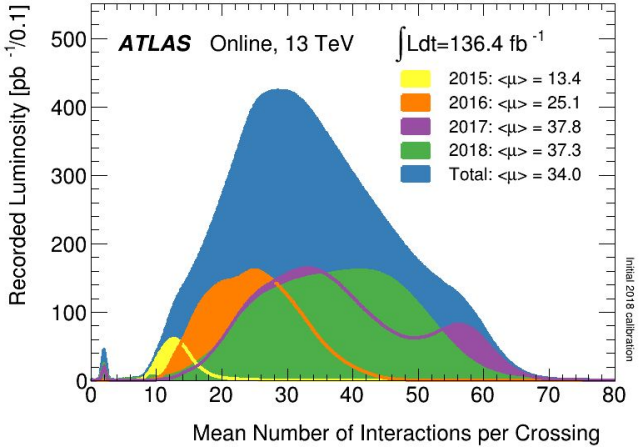
The ATLAS detector



- **A T**oroidal **L**HC **A**pparatu**S**
  - 44 m long, 25 m of diameter
  - 4 layers of detectors
- Inner detector
  - pixel, strip, TRT
- Electromagnetic calorimeter
- Hadronic calorimeter
- Muon detector

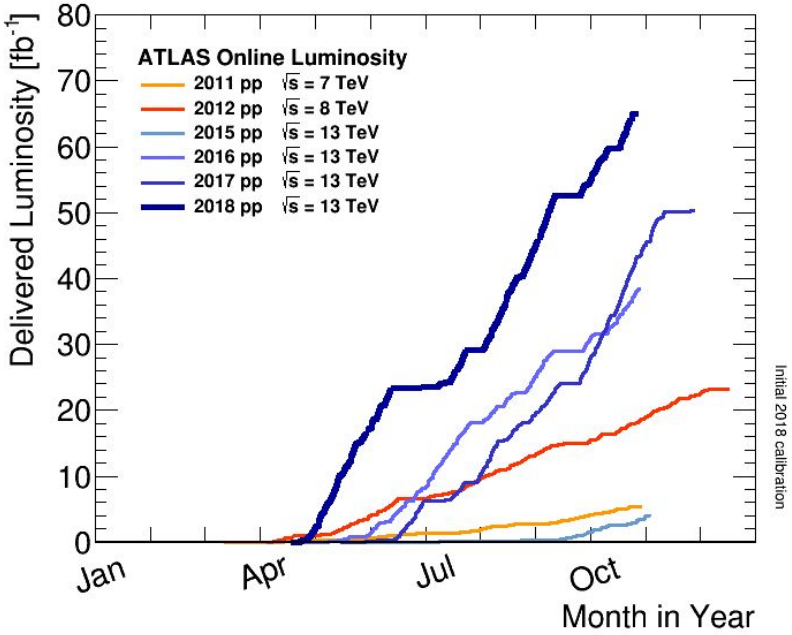
LHC and ATLAS performance





- At LHC two bunches of protons collide every 25 ns (40 MHz)
- LHC design instantaneous luminosity: 10^{34} cm^{-2} s^{-1}
- Hard collisions -> between two elementary components of the protons (q or g)
- Other components of the same hadrons produce “underlying event”
- Several collision events per proton bunch crossing “pile-up events”

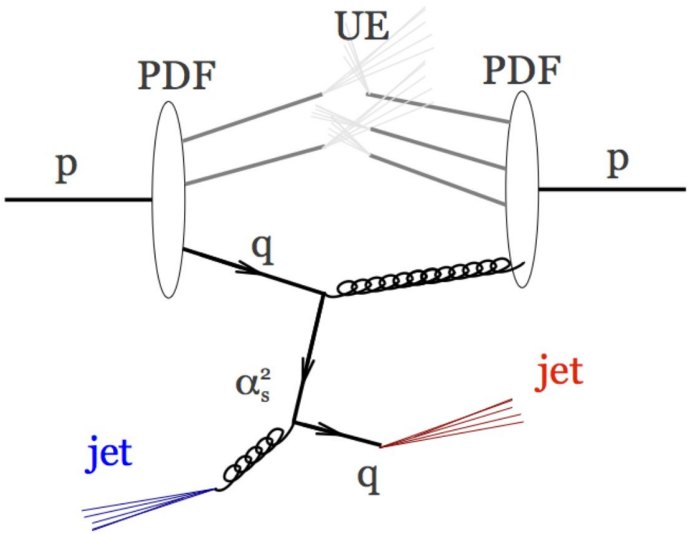
LHC and ATLAS performance



- Current center of mass energy $\sqrt{s}$=13 TeV
- Originally expected to get up to 120 fb^{-1} by the end of 2018
- ~ 160 fb^{-1} (!!) of proton-proton collision data registered by ATLAS in Run-II

What is a jet?

- At short distances quarks and gluons move as quasi-free particles (asymptotic freedom)
- When they are energetic, they produce bremsstrahlung cascades of gluons and qqbar pairs, which then hadronise
- We see jets of tightly collimated hadrons



- There are a lot at the LHC! (dominant high p_T process)
- Jet properties reflect those of the quarks and gluons which originated them
- A good handle to test the QCD sector of the SM over several orders of magnitude
  - Proton structure (PDF)
  - Strong coupling constant, α_s
  - Perturbative QCD effects
  - Fragmentation/Hadronization effects

Jet reconstruction

- Sequential recombination algorithms most popular in the LHC era (G. Salam, arXiv:0906.1833)
- Collinear and infrared safe!

- The clustering inverts the parton shower by combining the constituents of the jet according to subsequent 'distance' criteria

$$
\begin{aligned}
d_{ij} &= \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \times \frac{\Delta R_{i,j}^2}{R^2} \\
d_{iB} &= p_T^{2n}
\end{aligned}
$$

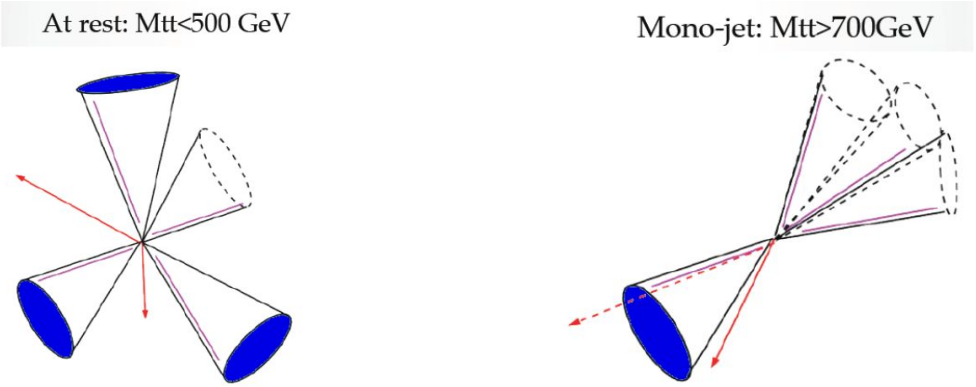
- Inclusive jet reconstruction: clustering continues until the minimum distance is found to be d_{iB}

- **(n=1) k_t :** Softest pair of constituents clustered first. Follows IR and collinear splittings.
- **(n=0) Cambridge-Aachen (C/A):** Closest pair of constituents clustered first. Mimics angular-ordered parton shower.
- **(n=-1) anti-k_t :** Hardest constituent clustered with closest neighbour. Regularly shaped jets.

Jet substructure

- Classical “resolved” algorithms run into problems for highly boosted final states

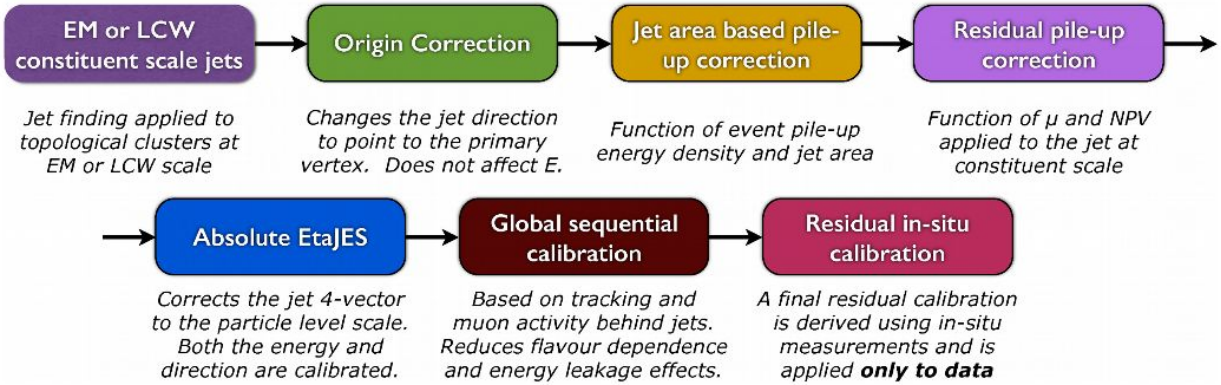
- A large radius jet of $R > 2m/p_T$ can contain all decay products of a given particle
  - Top quark, Higgs/W/Z bosons, new heavy particles ...



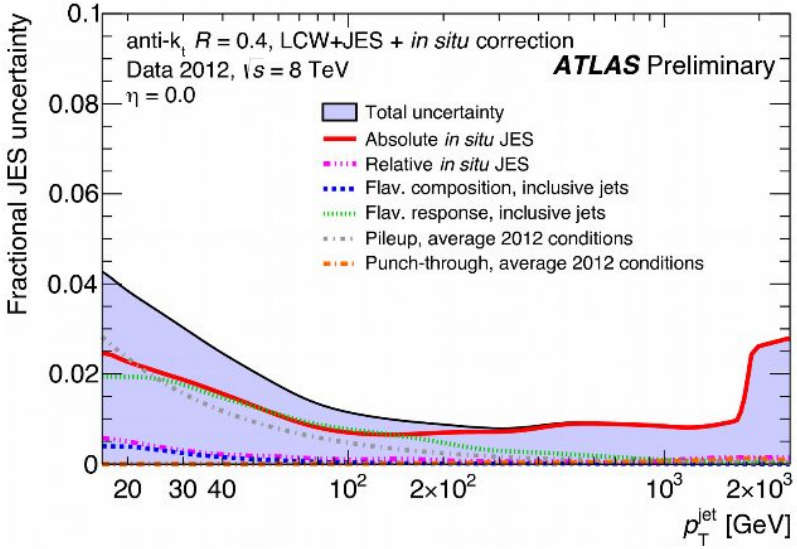
- Internal structure of the large R jet shows interesting features that can be used to identify the origin of the jet
  - distinguish multi-jet background from signals

Jets in ATLAS

- Jet production is the dominant high-p_T process in the LHC
- Jet observables play an important role in the study of:
  - The structure of the proton
  - The color interaction and its coupling strength α_s



- Anti-k_T jets
- Built considering topological clusters of calorimeter cells
- Clusters corrected for pileup prior to jet building
- Multi-stage calibration scheme
- Larger energy scale uncertainty than photons

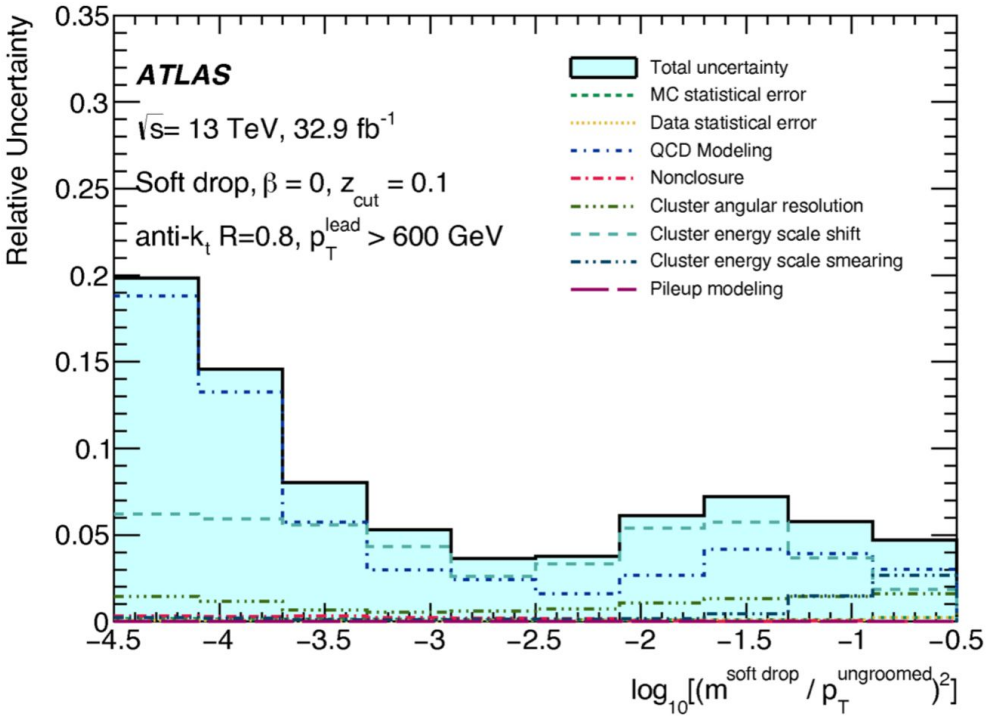


ATLAS measurement of the soft-drop jet mass

Phys. Rev. Lett. 121, 092001 (2018)

Uncertainties

- Many uncertainties cancel since ρ is a ratio
- QCD modeling uncertainties dominate
  - Particularly large at low mass where non-perturbative effects are largest
- Cluster energy uncertainties
  - Large at lower masses
    - Low cluster multiplicity
  - Also important at higher masses
    - Energy of hard prongs dominates the mass resolution instead of the opening angle
- Other uncertainties are subdominant
  - Pile-up negligible as expected



Unfolding

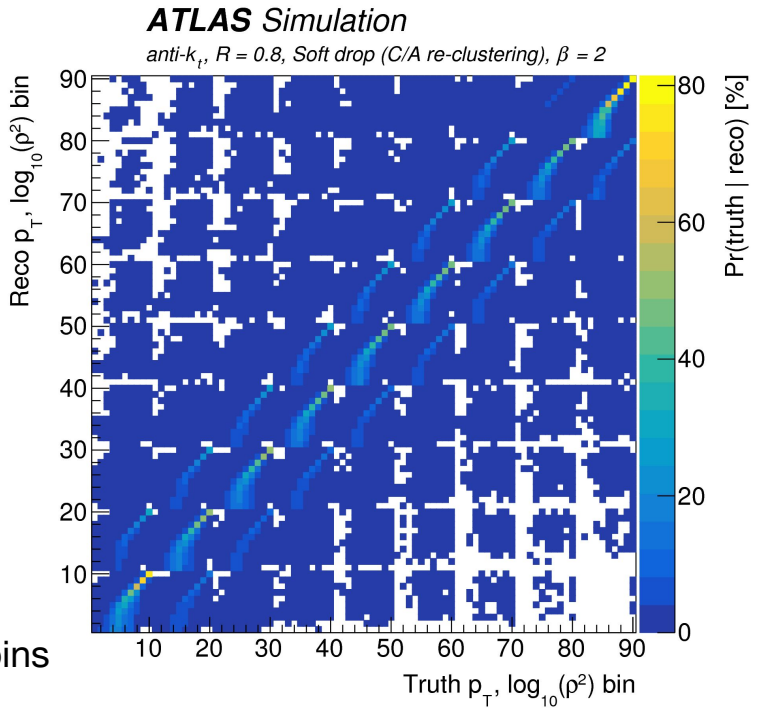
- Pythia used as nominal
  - Sherpa and Herwig++ to evaluate uncertainty

- Particle-level selection as close as possible to detector-level
  - Jets built using the same algorithm
  - Events must pass the same dijet requirement
  - Additional correction for the acceptance included

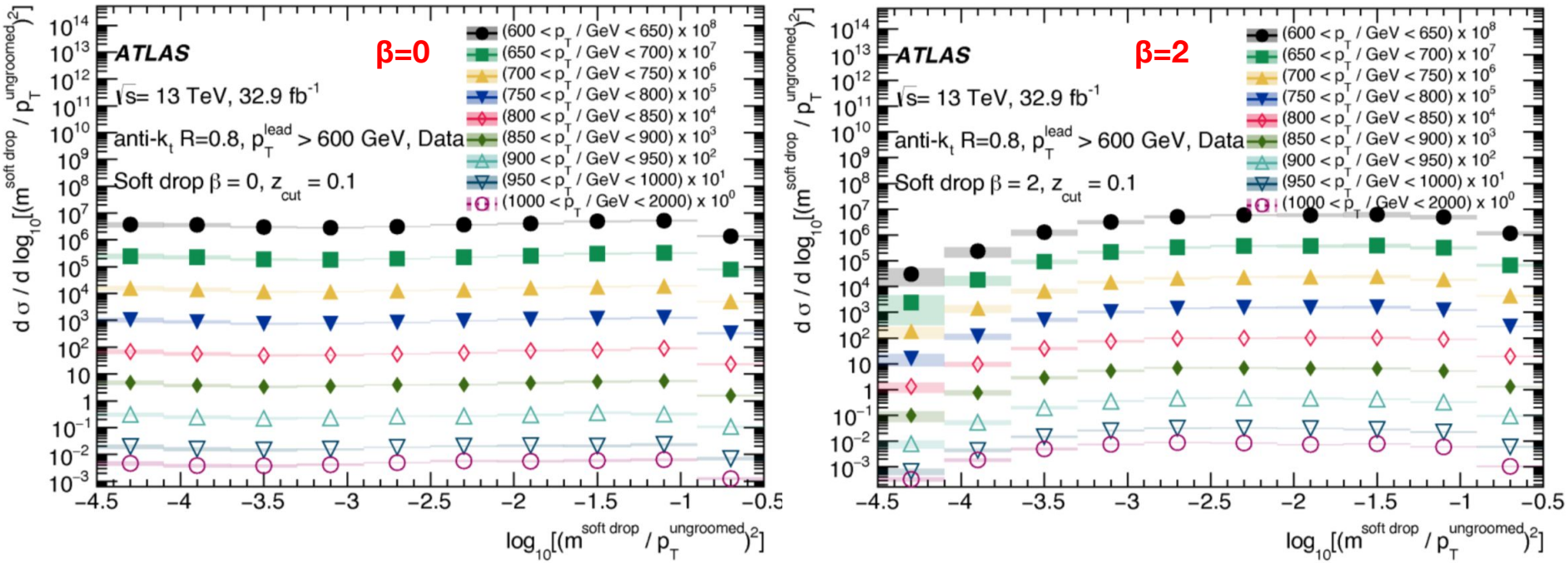
- $\log_{10}(\rho^2)$ and p_T unfolded simultaneously

- Example of response matrix for the combined p_T and $\log_{10}(\rho^2)$ bins
  - Each group of 10 bins corresponds to a different p_T bin
  - Each bin within the p_T bin corresponds to 10 evenly spaced bins in $\log_{10}(\rho^2)$
  - The bins are normalized so that the z-axis corresponds to the probability of a jet lying in a particular truth bin, given its reconstructed bin

- 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 $\log_{10}(\rho^2)$

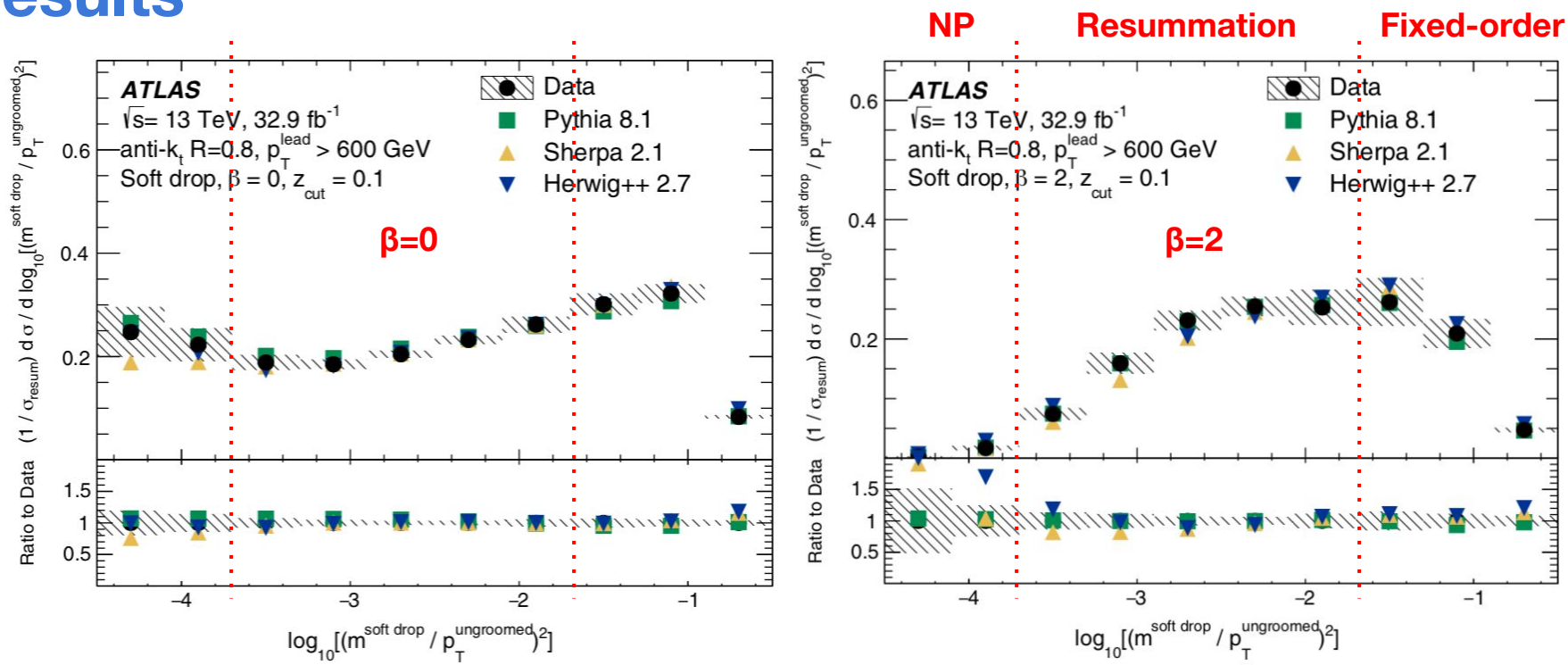


Results: $\log_{10}(\rho^2)$ vs p_T



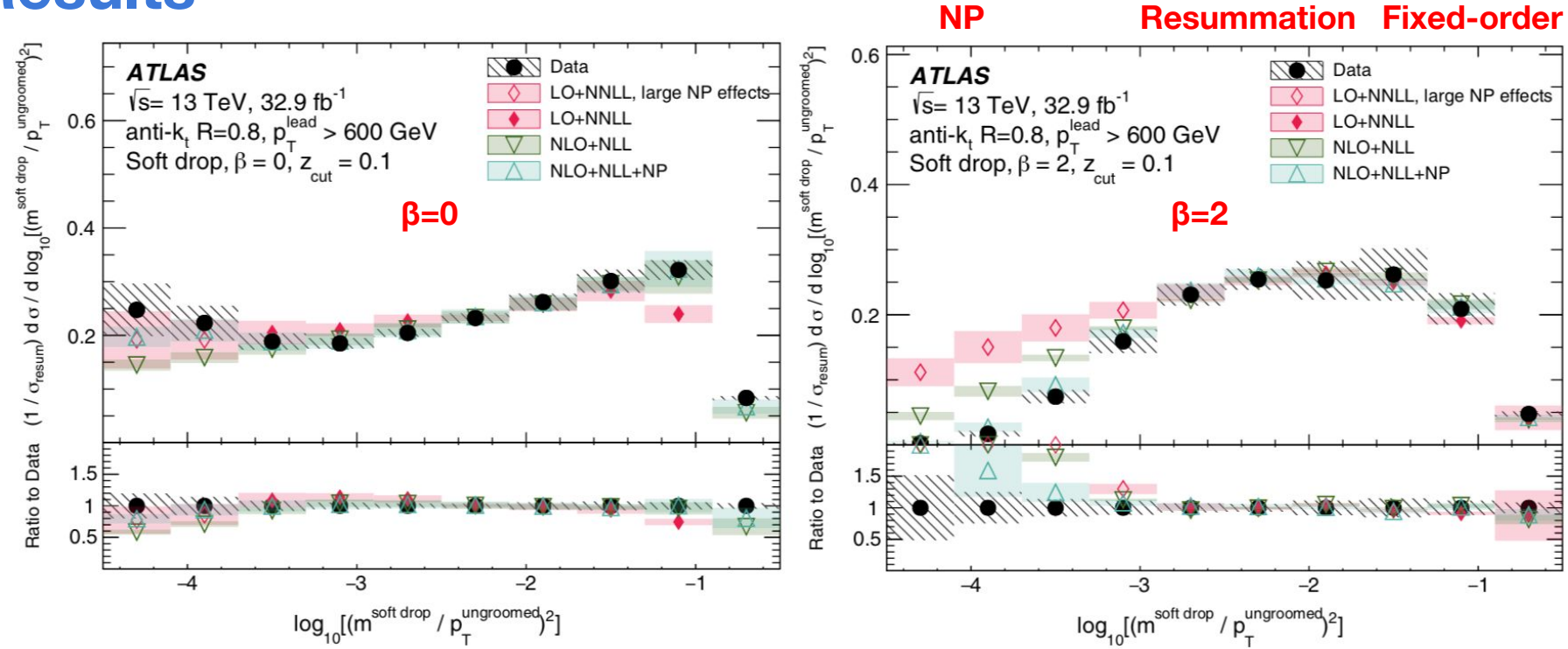
- $\log_{10}(\rho^2)$ for the p_T bins used in the analysis (from 600 GeV up to 2000 GeV)
- As expected, there is no strong dependence on p_T

Results



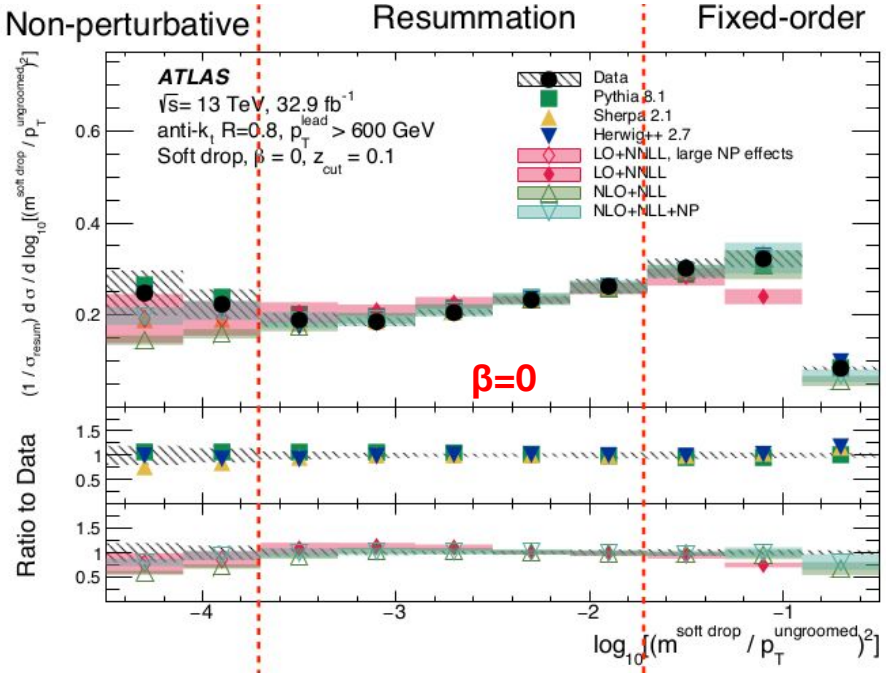
- Distributions normalised to data in resummation region
- MC generators do an excellent job of describing data over entire mass range

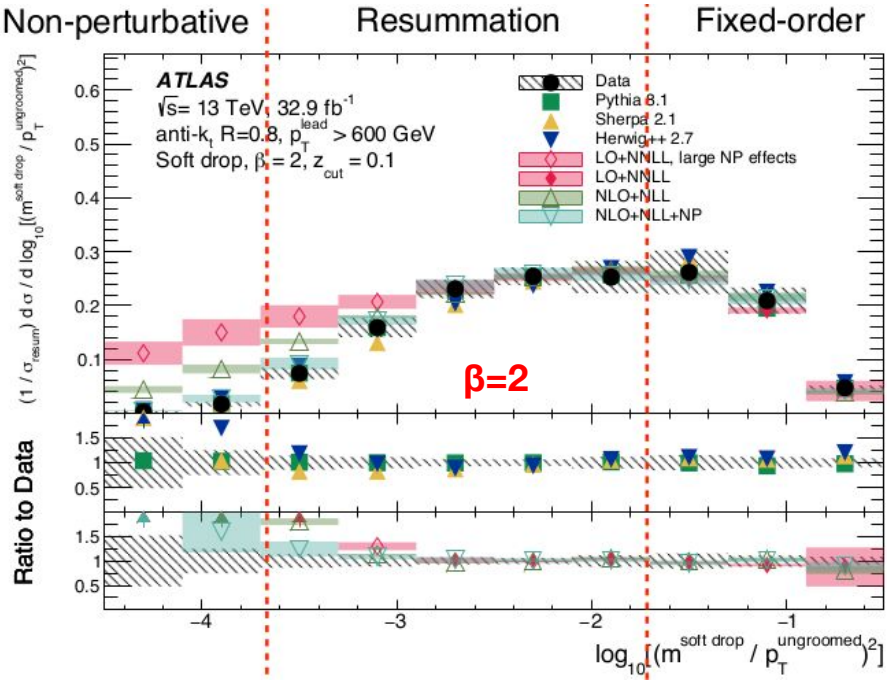
Results



- Good agreement between data and analytic calculations in resummation region
  - and fixed-order region for NLO calculations
- Including non-perturbative effects improves the accuracy of the NLO+NLL prediction

Results





- Largest difference between data and MC/analytic predictions in non-perturbative regime
  - Effect larger for higher β (smaller fraction of soft energy removed)

ATLAS measurement of jet-substructure observables in ttbar and inclusive jet events

arXiv:1903.02942

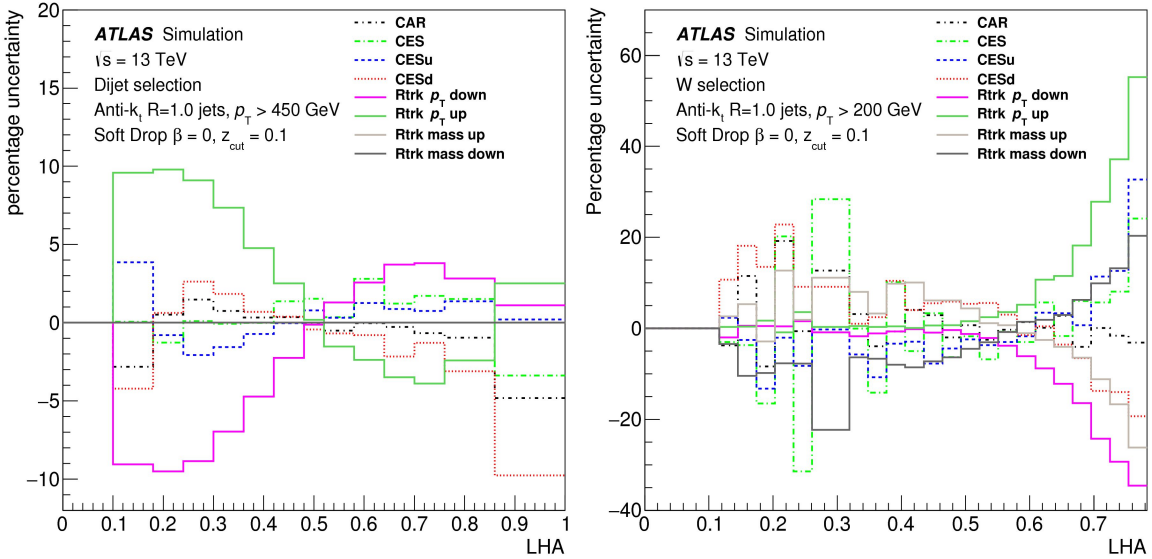
Simulations

Process	Generator	Version	PDF	Tune	Use
Dijet	Pythia8 [27, 28]	8.186	NNPDF23LO [29]	A14 [30]	Nominal for unfolding
	Sherpa [31]	2.2.1	CT10 [32]	Default	Validation of unfolding (with two different hadronisation models)
	Herwig7 [33]	7.0.4	MMHT2014	H7UE [33]	Comparison
$t\bar{t}$	Powheg [34]	v2	NNPDF30NLO		Nominal for unfolding
	+ Pythia8	8.186	NNPDF23LO	A14	
	Powheg	v2	CT10		Validation of unfolding
	+Herwig++ [35]	2.7	CTEQ6L1	UE-EE-5 tune [36]	
	Powheg	v2	CT10		Comparison
	+Herwig7	7.0.4	MMHT2014	H7UE	
	MG5_aMC@NLO [37]	2.6.0	NNPDF30NLO		Comparison
	+ Pythia8	8.186	NNPDF23LO	A14	
	Sherpa	2.2.1	CT10	Default	Comparison
Single top	Powheg	v1	CT10		Nominal for unfolding
	+ Pythia6 [38, 39]	6.428	CTEQ6L1 [32]	Perugia2012 [40]	
Z+jets	Sherpa	2.2.1	CT10	Default	Background estimation
W+jets	Sherpa	2.2.1	CT10	Default	Background estimation (nominal)
W+jets	MG5_aMC@NLO	2.2.5	CT10		Background estimation (cross-check)
	+ Pythia8	8.186	NNPDF23LO	A14	
Diboson	Sherpa	2.2.1	CT10	Default	Background estimation

Selection

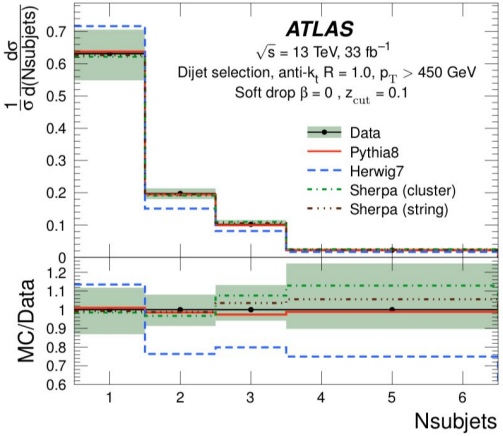
	Detector level	Particle level
Dijet selection:		
Two trimmed anti-k_t $R = 1.0$ jets	$p_T > 200$ GeV $\lvert\eta\rvert < 2.5$	$p_T > 200$ GeV $\lvert\eta\rvert < 2.5$
Leading-$p_T$trimmed anti-k_t $R = 1.0$ jet	$p_T > 450$ GeV	
Top and W selections:		
Exactly one muon	$p_T > 30$ GeV $\lvert\eta\rvert < 2.5$ $\lvert z_0 \sin(\theta)\rvert < 0.5$ mm and $\lvert d_0/\sigma(d_0)\rvert < 3$	$p_T > 30$ GeV $\lvert\eta\rvert < 2.5$
Anti-k_t $R = 0.4$ jets	$p_T > 25$ GeV $\lvert\eta\rvert < 4.4$ JVT output > 0.5 (if $p_T < 60$ GeV)	$p_T > 25$ GeV $\lvert\eta\rvert < 4.4$
Muon isolation criteria	If $\Delta R(\mu, \text{jet}) < 0.04 + 10$ GeV $/p_{T,\mu}$: muon is removed, so the event is discarded	None
E_T^{miss}, m_T^W	$E_T^{miss} > 20$ GeV , $E_T^{miss} + m_T^W > 60$ GeV	
Leptonic top	At least one small-radius jet with $0.4 < \Delta R(\mu, \text{jet}) < 1.5$	
Top selection:		
Leading-$p_T$trimmed anti-k_t $R = 1.0$ jet	$\lvert\eta\rvert < 1.5$, $p_T > 350$ GeV , mass > 140 GeV $\Delta R(\text{large-radius jet}, b\text{-tagged jet}) < 1$ $\Delta\phi(\mu, \text{large-radius jet}) > 2.3$	
W selection:		
Leading-$p_T$trimmed anti-k_t $R = 1.0$ jet	$\lvert\eta\rvert < 1.5$, $p_T > 200$ GeV , mass > 60 GeV and mass < 100 GeV $1 < \Delta R(\text{large-radius jet}, b\text{-tagged jet}) < 1.8$ $\Delta\phi(\mu, \text{large-radius jet}) > 2.3$	

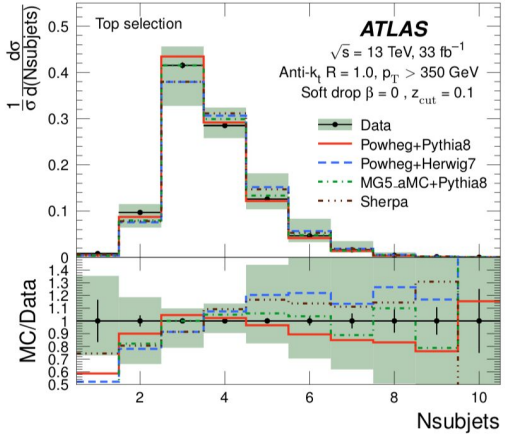
Uncertainties

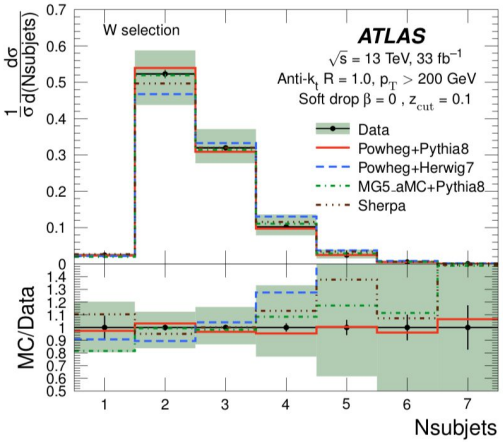


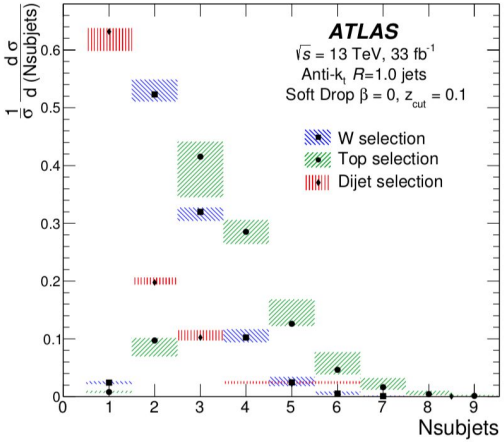
- Cluster and large R systematics dominate in all topologies
  - Largest effect on the majority of measured distributions from cluster energy smearing (CES)

Results

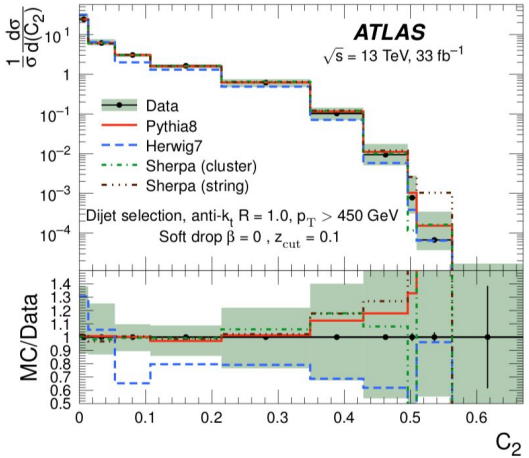

ATLAS
√s = 13 TeV, 33 fb⁻¹
Dijet selection, anti-kt R = 1.0, pT > 450 GeV
Soft drop β = 0 , zcut = 0.1
Data
Pythia8
Herwig7
Sherpa (cluster)
Sherpa (string)
1/σ dσ/d(Nsubjets)
MC/Data
Nsubjets

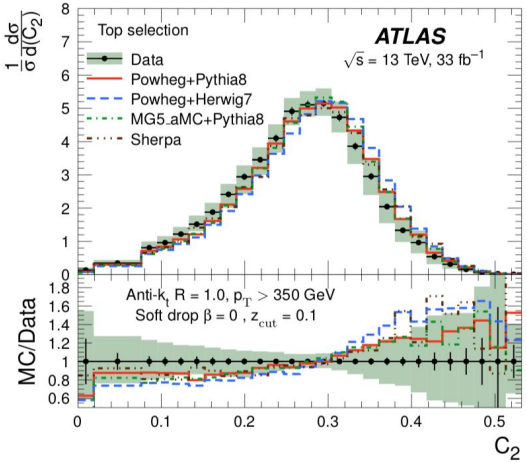

Top selection
ATLAS
√s = 13 TeV, 33 fb⁻¹
Anti-kt R = 1.0, pT > 350 GeV
Soft drop β = 0 , zcut = 0.1
Data
Powheg+Pythia8
Powheg+Herwig7
MG5_aMC+Pythia8
Sherpa
1/σ dσ/d(Nsubjets)
MC/Data
Nsubjets

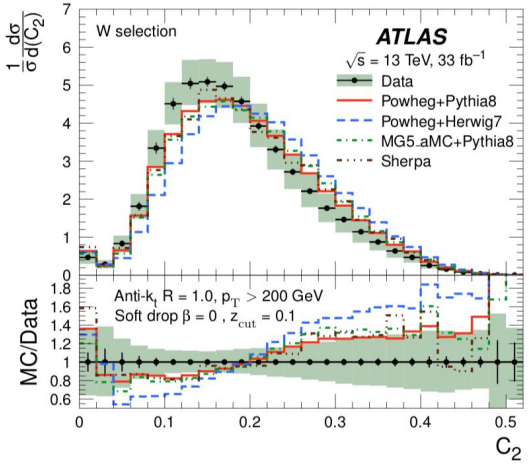

W selection
ATLAS
√s = 13 TeV, 33 fb⁻¹
Anti-kt R = 1.0, pT > 200 GeV
Soft drop β = 0 , zcut = 0.1
Data
Powheg+Pythia8
Powheg+Herwig7
MG5_aMC+Pythia8
Sherpa
1/σ dσ/d(Nsubjets)
MC/Data
Nsubjets

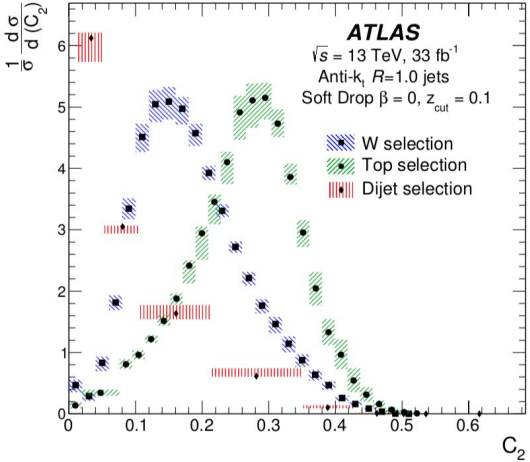

ATLAS
√s = 13 TeV, 33 fb⁻¹
Anti-kt R=1.0 jets
Soft Drop β = 0, zcut = 0.1
W selection
Top selection
Dijet selection
1/σ dσ/d(Nsubjets)
Nsubjets

Results

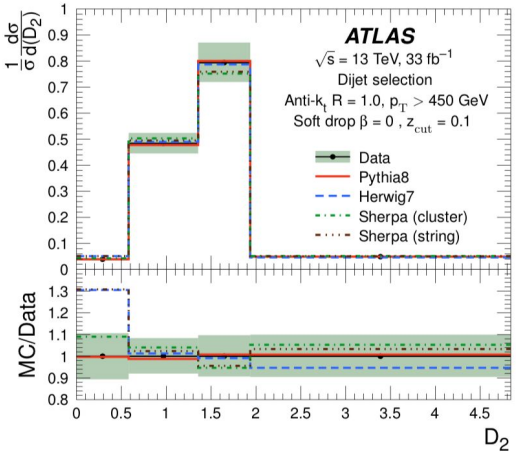

ATLAS
√s = 13 TeV, 33 fb⁻¹
Data
Pythia8
Herwig7
Sherpa (cluster)
Sherpa (string)
Dijet selection, anti-kt R = 1.0, pT > 450 GeV
Soft drop β = 0 , zcut = 0.1
MC/Data
C2

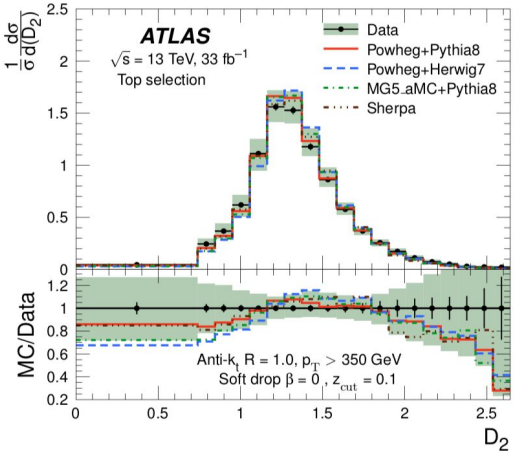

Top selection
ATLAS
√s = 13 TeV, 33 fb⁻¹
Data
Powheg+Pythia8
Powheg+Herwig7
MG5_aMC+Pythia8
Sherpa
Anti-kt R = 1.0, pT > 350 GeV
Soft drop β = 0 , zcut = 0.1
MC/Data
C2

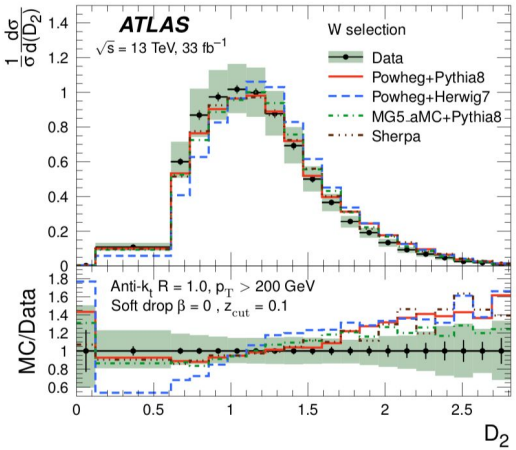

W selection
ATLAS
√s = 13 TeV, 33 fb⁻¹
Data
Powheg+Pythia8
Powheg+Herwig7
MG5_aMC+Pythia8
Sherpa
Anti-kt R = 1.0, pT > 200 GeV
Soft drop β = 0 , zcut = 0.1
MC/Data
C2

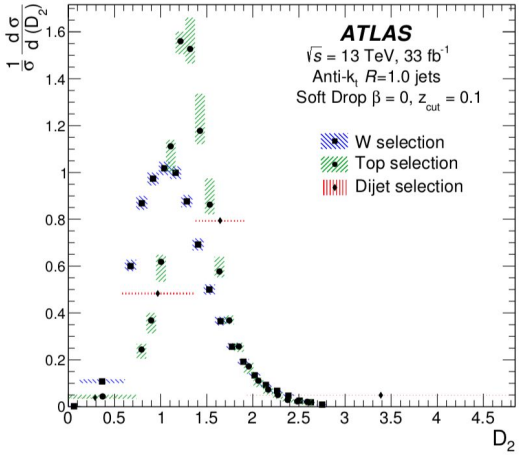

ATLAS
√s = 13 TeV, 33 fb⁻¹
Anti-kt R=1.0 jets
Soft Drop β = 0, zcut = 0.1
W selection
Top selection
Dijet selection
C2

Results


ATLAS
√s = 13 TeV, 33 fb⁻¹
Dijet selection
Anti-kt R = 1.0, pT > 450 GeV
Soft drop β = 0 , zcut = 0.1
Data
Pythia8
Herwig7
Sherpa (cluster)
Sherpa (string)
MC/Data
D2


ATLAS
√s = 13 TeV, 33 fb⁻¹
Top selection
Data
Powheg+Pythia8
Powheg+Herwig7
MG5_aMC+Pythia8
Sherpa
Anti-kt R = 1.0, pT > 350 GeV
Soft drop β = 0 , zcut = 0.1
MC/Data
D2


ATLAS
√s = 13 TeV, 33 fb⁻¹
W selection
Data
Powheg+Pythia8
Powheg+Herwig7
MG5_aMC+Pythia8
Sherpa
Anti-kt R = 1.0, pT > 200 GeV
Soft drop β = 0 , zcut = 0.1
MC/Data
D2


ATLAS
√s = 13 TeV, 33 fb⁻¹
Anti-kt R=1.0 jets
Soft Drop β = 0, zcut = 0.1
W selection
Top selection
Dijet selection
D2

Multijet simulation for 13 TeV ATLAS Analyses

ATL-PHYS-PUB-2019-017

Overview

- Jet shape: normalized p_T flow inside the jet cone as a function of the angular distance to the jet axis
- Two versions:
  - the differential jet shape is defined as the fraction of p_T in a circular crown of width Δr at a distance $r < R$ of the jet centroid

$$\rho(r) = \frac{1}{\Delta r} \frac{p_T(r - \Delta r/2, r + \Delta r/2)}{p_T(0, R)}$$

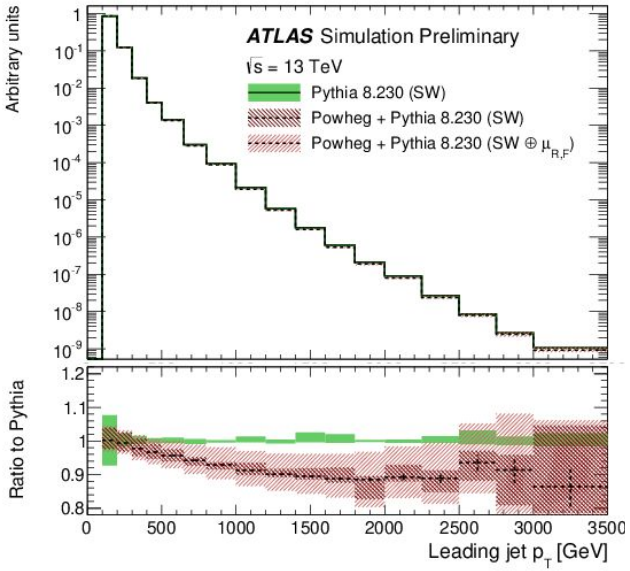
  - the integrated jet shape is defined as the cumulative distribution $\rho(r)$. Geometrically, it can be regarded as the fraction of momentum in circles of radius $r < R$ taken from the jet axis

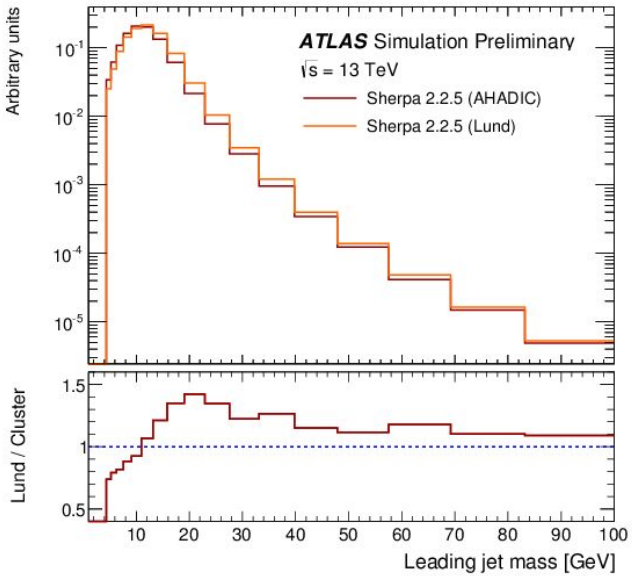
$$\Psi(r) = \frac{p_T(0, r)}{p_T(0, R)}$$

- Both are sensitive to the kinematics of the jet under study.
  - the radiation around the original parton becomes more collimated when increasing the jet p_T, giving higher values of $\rho(r)$ for small values of r and also higher values of $\Psi(r)$ overall

-

Results





- Pythia 8 shower uncertainty largest at low p_T, but negligible at mid/high p_T
- Powheg + Pythia 8
  - reduced impact of the combined ME+PS uncertainties at low p_T (one extra parton included at ME at LO)
  - uncertainty slightly increases as a function of the jet p_T (due to the increase in PDF uncertainty)
- Sherpa two different hadronization models show differences of up to 45%
  - low p_T jets described by the cluster model are more collimated than their string model counterparts
    - as the mass distribution peaks at lower values