Jet Substructure and Tagging in ATLAS

Chris Malena Delitzsch

University of Arizona

On behalf of the ATLAS Collaboration

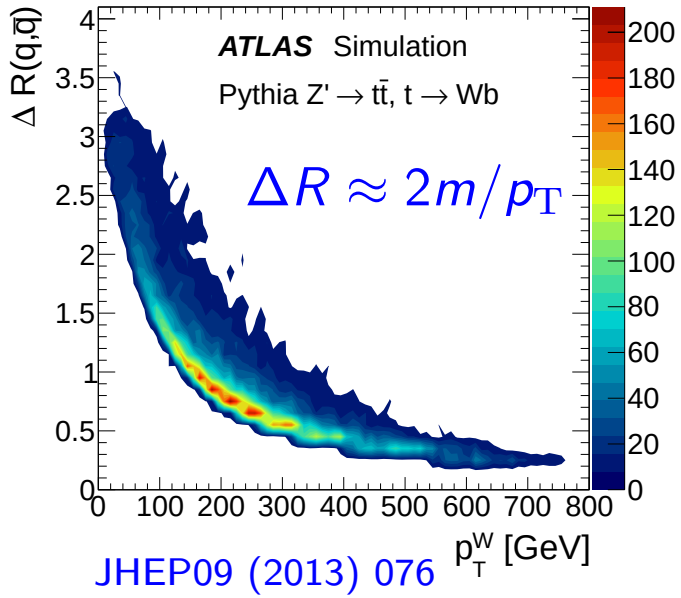
LHCP 2018

4-9 June 2018

Introduction

- High centre-of-mass energy allows abundant production of $W/Z/H$ bosons and top quarks with $p_{\mathrm{T}} >> m$
  - Decay products of initiating particle are collimated
  - $\rightarrow$ reconstruction as one single large-radius jet

- Large-R jet more susceptible to pile-up
  $\rightarrow$ grooming
- Use internal structure of jet to discriminate between initiating particle $\rightarrow$ tagging
- Substructure techniques heavily used in searches beyond SM but measurements are catching up!!

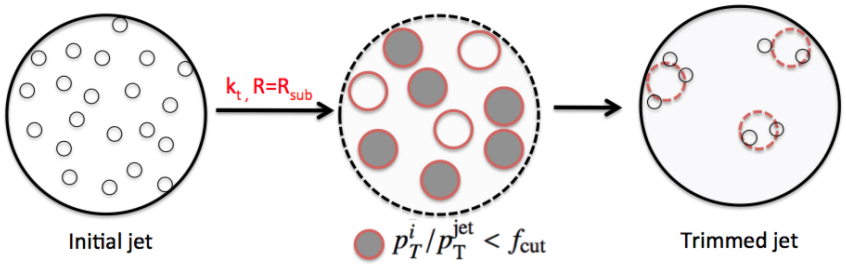


JHEP09 (2013) 076

- Measurements of jet properties/structure crucial to constrain SM in new energy regimes and important test of perturbative calculations
- Cross-section measurements provide constraints on PDFs and α_s

A few basics before getting started

Trimming JHEP 1002:084,2010



- Removes subjet (size R_{sub}) if:
  $p_{\mathrm{T}}^{i}/p_{\mathrm{T}}^{\mathrm{jet}} < f_{\mathrm{cut}}$
- Typical parameters in ATLAS:
  - $R_{\mathrm{sub}} = 0.2$, $f_{\mathrm{cut}} = 5\%$

N-subjettiness JHEP 1103:015,2011

- Describes how likely it is that a jet is composed out of N subjets:

$$\tau_N = \frac{\sum_k p_{\mathrm{T},k}(\min(\Delta R_{1,k}, R_{2,k}, \ldots, R_{N,k}))^{\beta}}{\sum_k p_{\mathrm{T}}(R_0)^{\beta}}$$

- Powerful discrimination:
  τ_2/τ_1 (W/Z), τ_3/τ_2 (top)

Soft-drop JHEP 1405 (2014) 146

- Preferred by theory community, calculation available at NNLL

1. Recluster constituents of large-R jet with C/A algorithm
2. Go through C/A clustering history and check soft drop condition:

$$\frac{\min(p_{\mathrm{T1}}, p_{\mathrm{T2}})}{p_{\mathrm{T1}} + p_{\mathrm{T2}}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

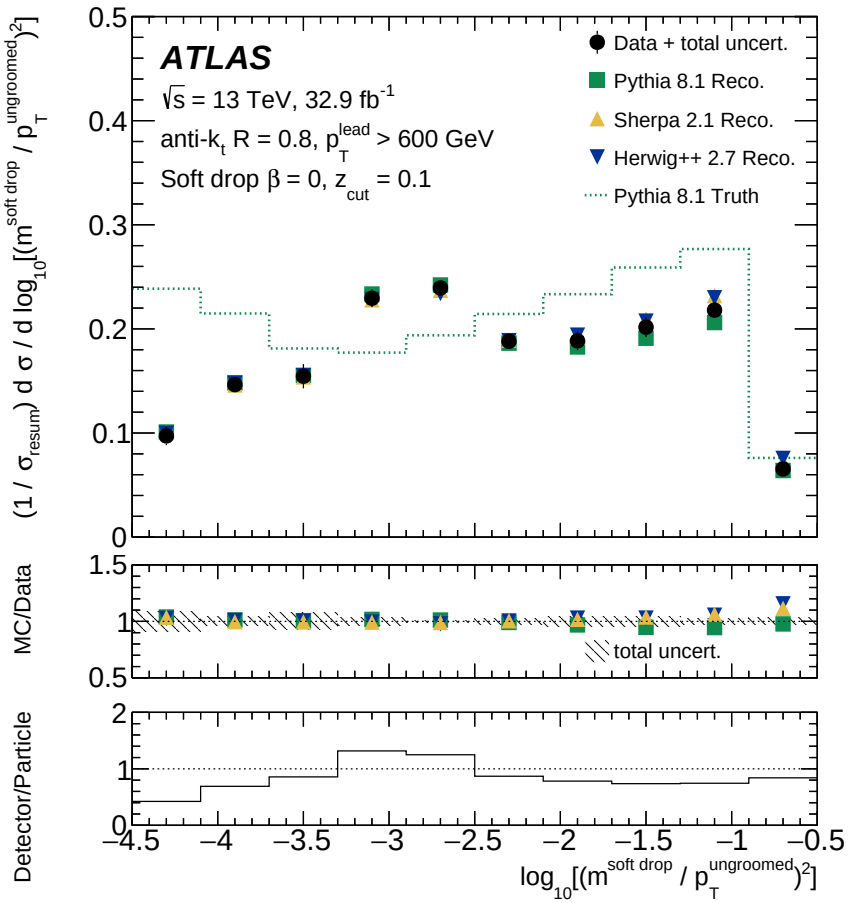
Jet mass

- Calculated from jet constituents

$$m^2 = \left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2$$

Soft drop mass measurement

- Soft-drop mass predicted at NLO + NLL and LO + NNLL accuracy
- Mass measurement performed with 32.9 fb^{-1} of 13 TeV pp collisions (arXiv:1711.08341)

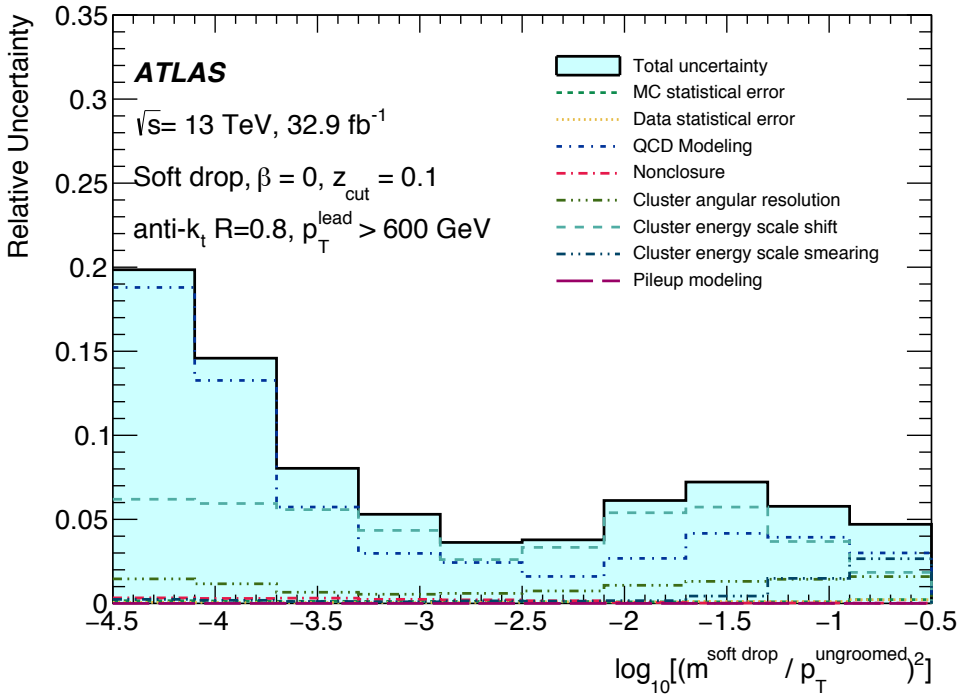


Event selection

- anti-k_t $R = 0.8$ jets
- $p_{\mathrm{T},1} > 600$ GeV, $p_{\mathrm{T},1}/p_{\mathrm{T},2} < 1.5$
- $z_{\mathrm{cut}} = 0.1$, $\beta = \{0, 1, 2\}$

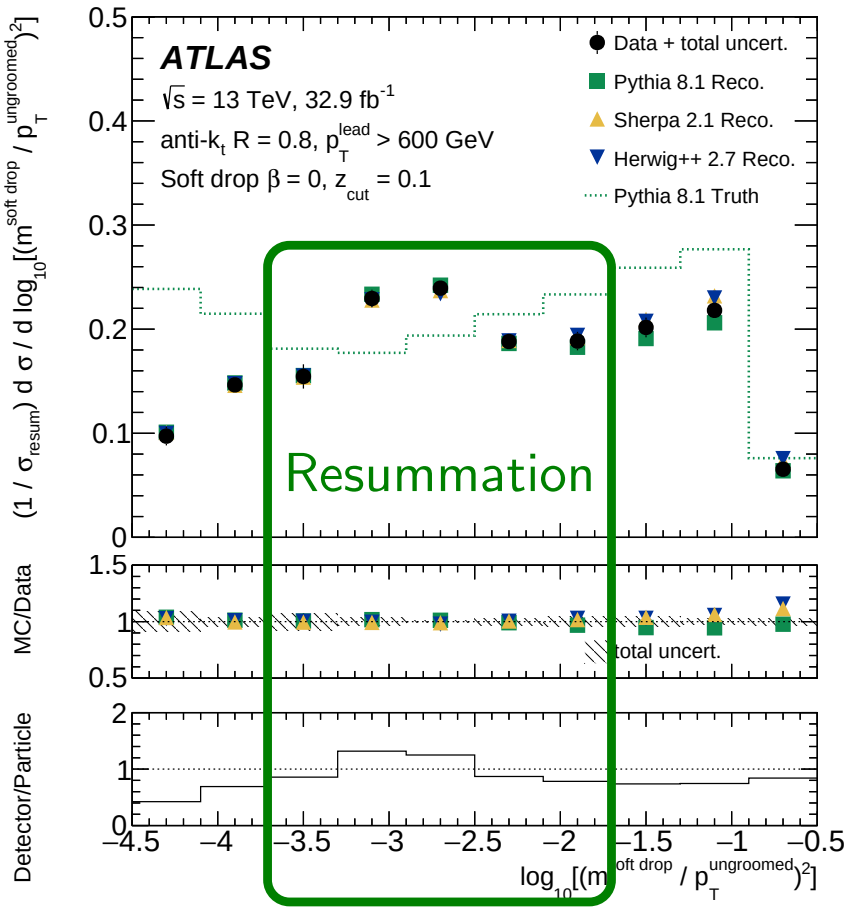
Dominating systematic uncertainties

- Calorimeter cell cluster energy scale
- Modelling (PYTHIA vs. SHERPA)



Soft drop mass measurement

- Soft-drop mass predicted at NLO + NLL and LO + NNLL accuracy
- Mass measurement performed with 32.9 fb^{-1} of 13 TeV pp collisions (arXiv:1711.08341)

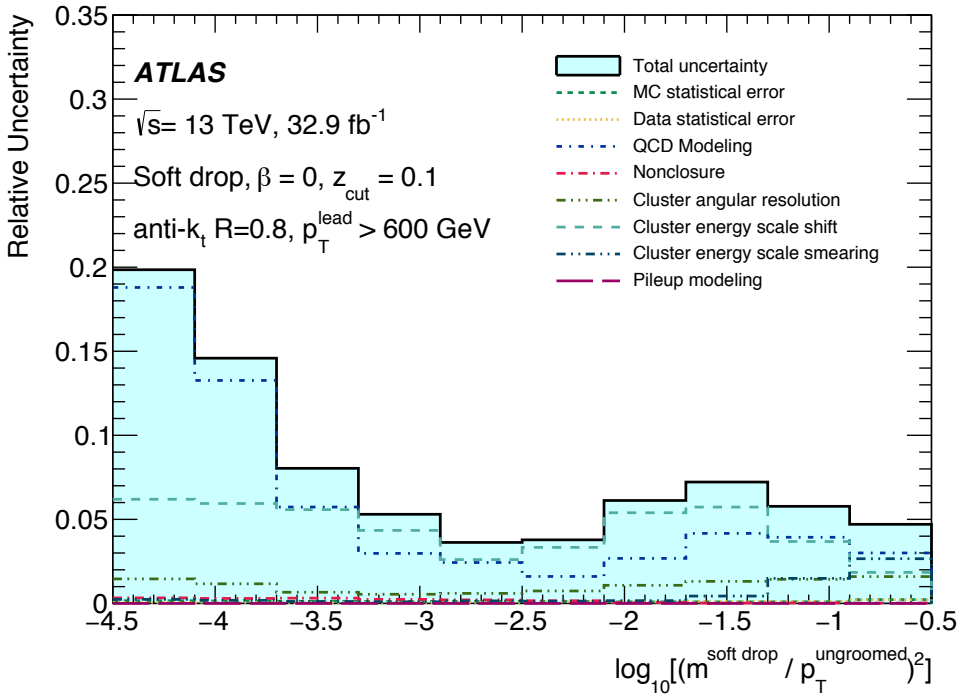


Event selection

- anti-k_t $R = 0.8$ jets
- $p_{\mathrm{T},1} > 600$ GeV, $p_{\mathrm{T},1}/p_{\mathrm{T},2} < 1.5$
- $z_{\mathrm{cut}} = 0.1$, $\beta = \{0, 1, 2\}$

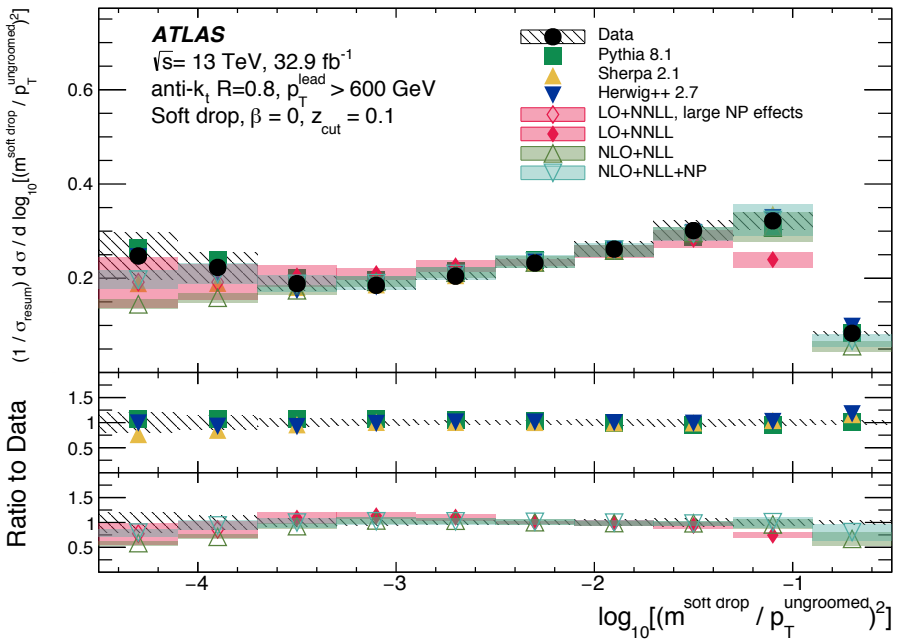
Dominating systematic uncertainties

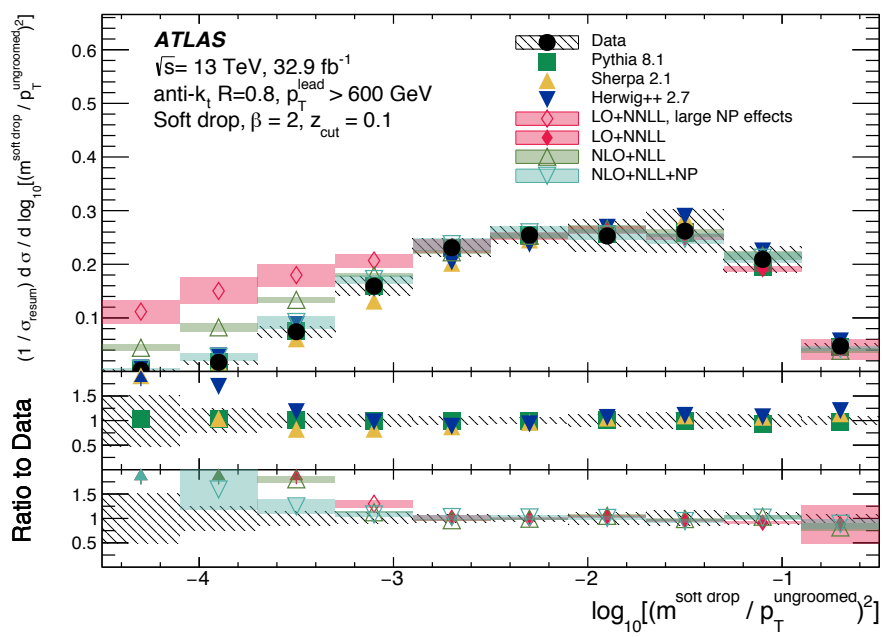
- Calorimeter cell cluster energy scale
- Modelling (PYTHIA vs. SHERPA)



Soft drop mass measurement - Results

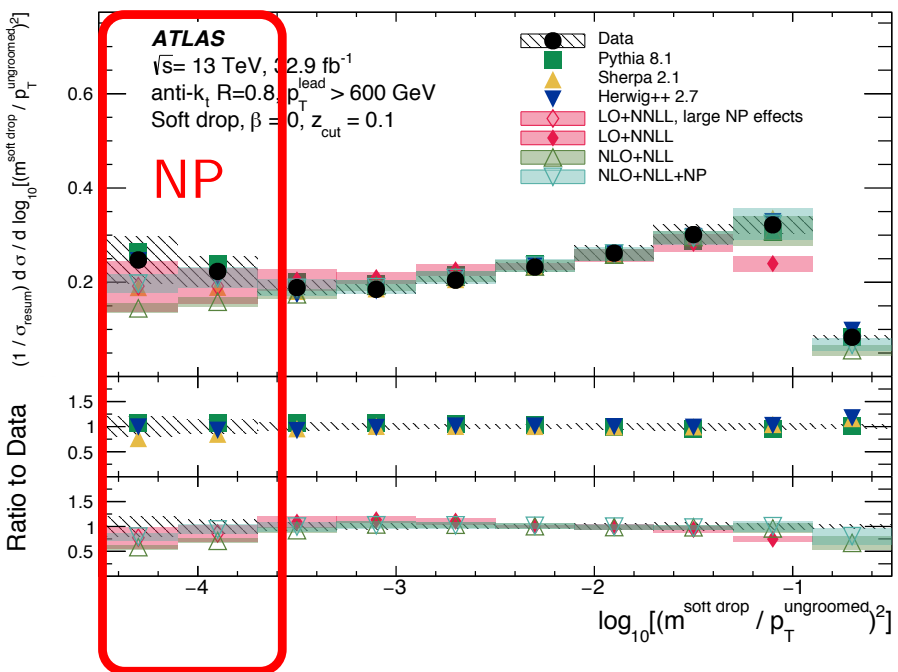
- Iterative Bayesian unfolding to particle-level
- $\beta = 2$: less soft-radiation removed compared to $\beta = 0$
- Parameter of interest: $\rho = m^{\text{soft drop}}/p_{\text{T}}^{\text{ungroomed}}$

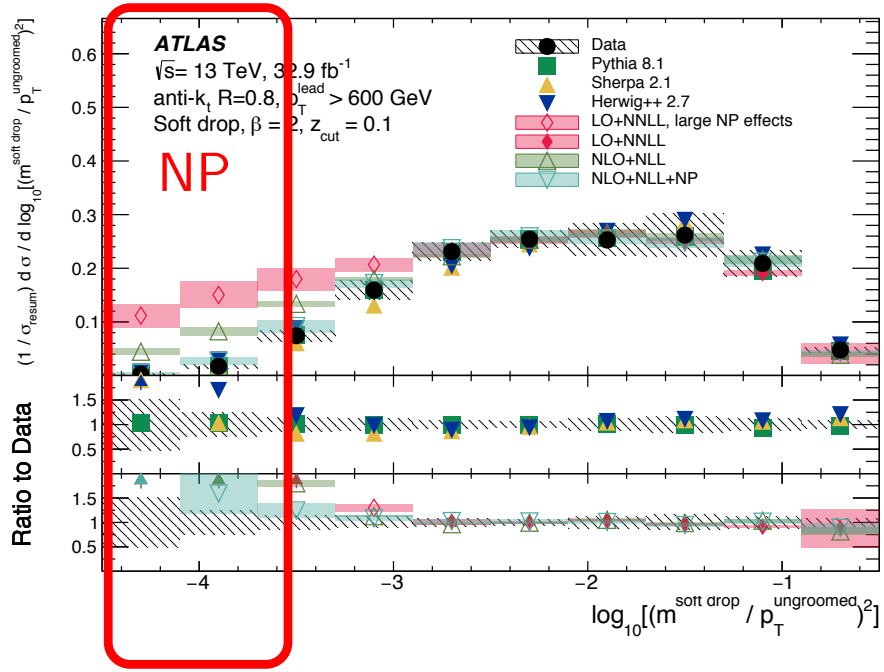




Soft drop mass measurement - Results

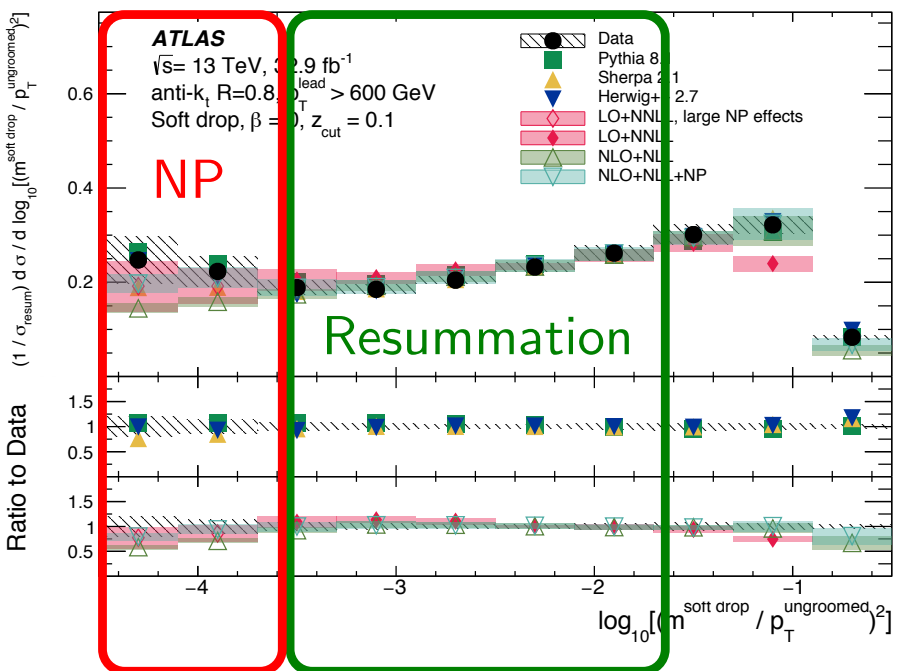
- Iterative Bayesian unfolding to particle-level
- $\beta = 2$: less soft-radiation removed compared to $\beta = 0$
- Parameter of interest: $\rho = m^{\text{soft drop}}/p_{\text{T}}^{\text{ungroomed}}$
- Non-perturbative effects large for $\log \rho^2 < -3.7$ (small angle, soft emission)
  - NP effects play larger role for $\beta = 2$

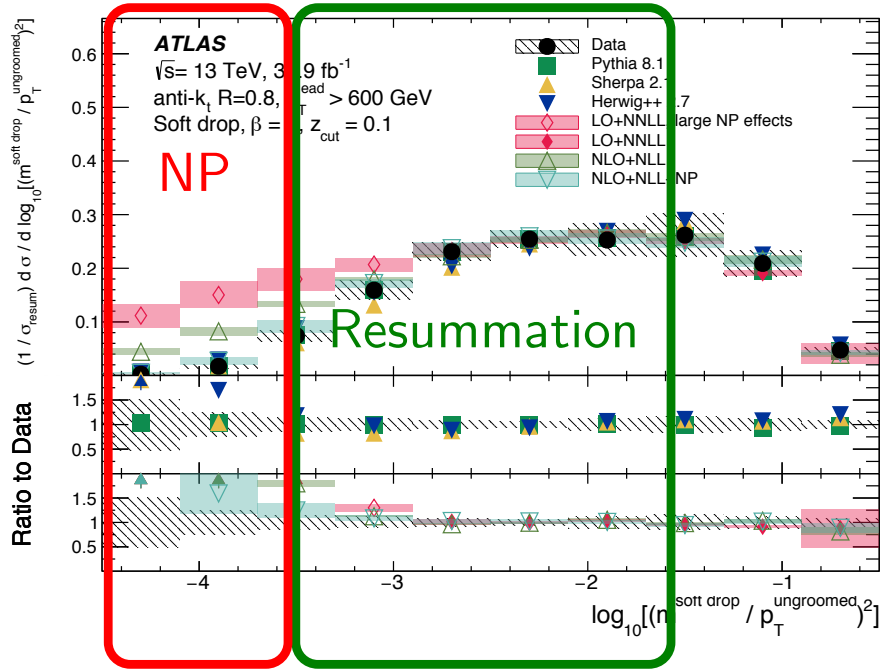




Soft drop mass measurement - Results

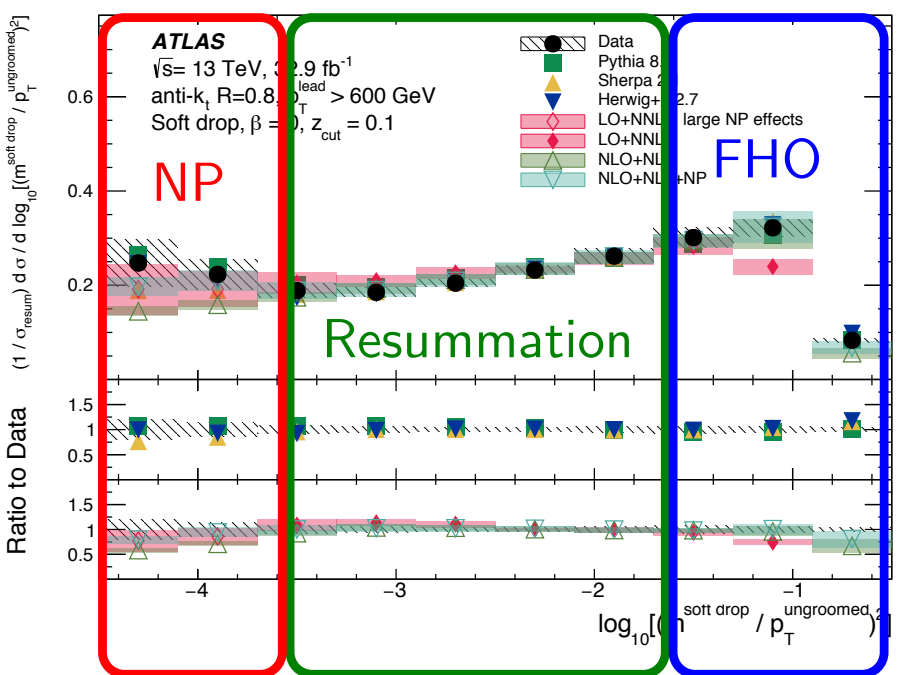
- Iterative Bayesian unfolding to particle-level
- $\beta = 2$: less soft-radiation removed compared to $\beta = 0$
- Parameter of interest: $\rho = m^{\text{soft drop}}/p_{\text{T}}^{\text{ungroomed}}$
- Non-perturbative effects large for $\log \rho^2 < -3.7$ (small angle, soft emission)
  - NP effects play larger role for $\beta = 2$
- Resummation region for $-3.7 < \log \rho^2 < -1.7$

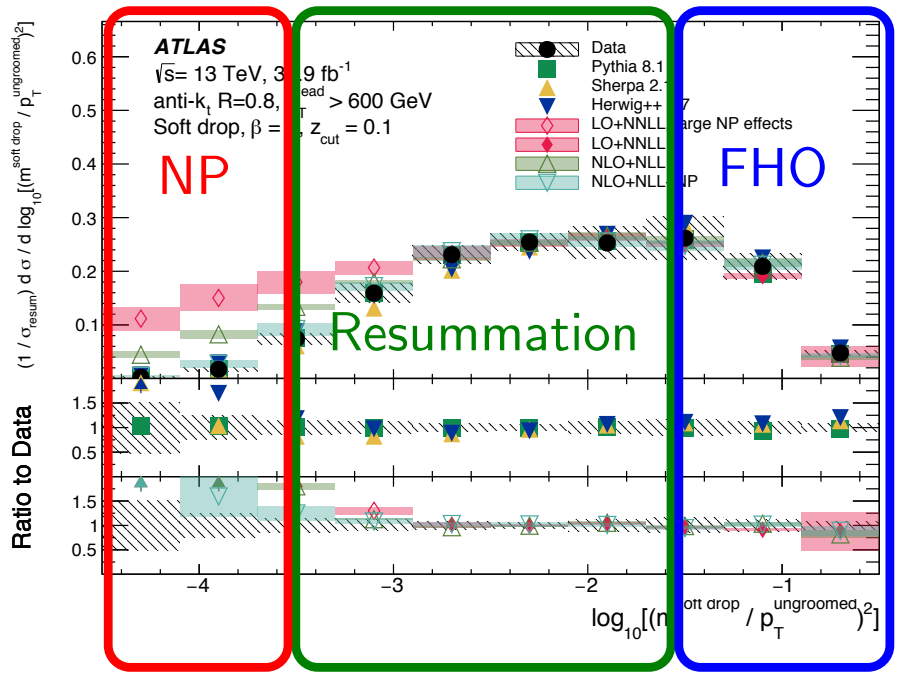




Soft drop mass measurement - Results

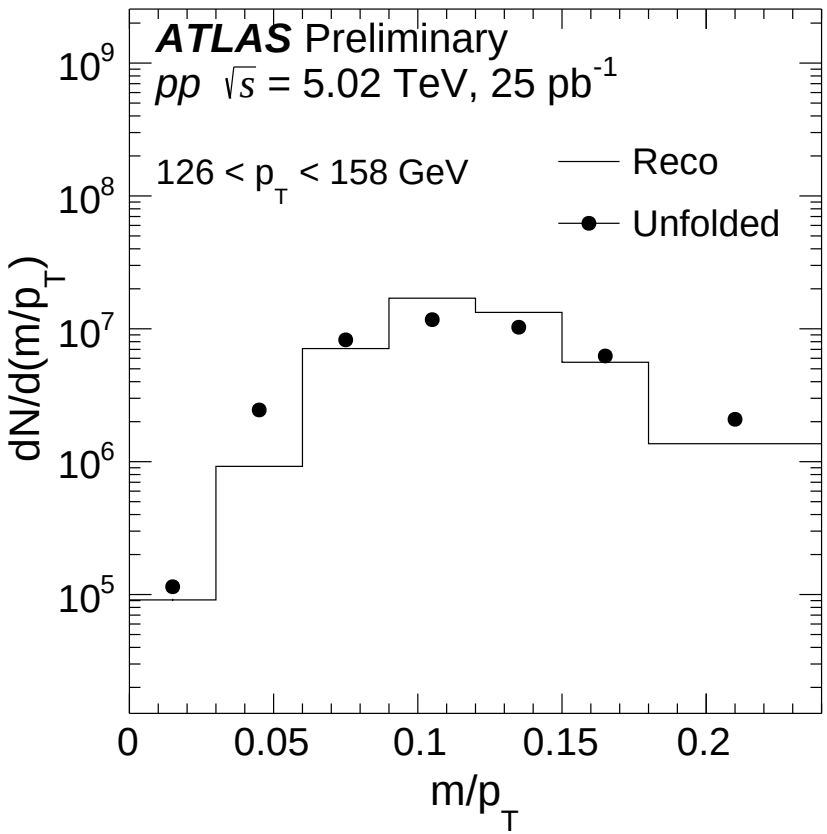
- Iterative Bayesian unfolding to particle-level
- $\beta = 2$: less soft-radiation removed compared to $\beta = 0$
- Parameter of interest: $\rho = m^{\text{soft drop}}/p_{\text{T}}^{\text{ungroomed}}$
- Non-perturbative effects large for $\log \rho^2 < -3.7$ (small angle, soft emission)
  - NP effects play larger role for $\beta = 2$
- Resummation region for $-3.7 < \log \rho^2 < -1.7$
- Fixed, higher-order corrections for $\log \rho^2 > -1.7$ (large angle gluon emission)

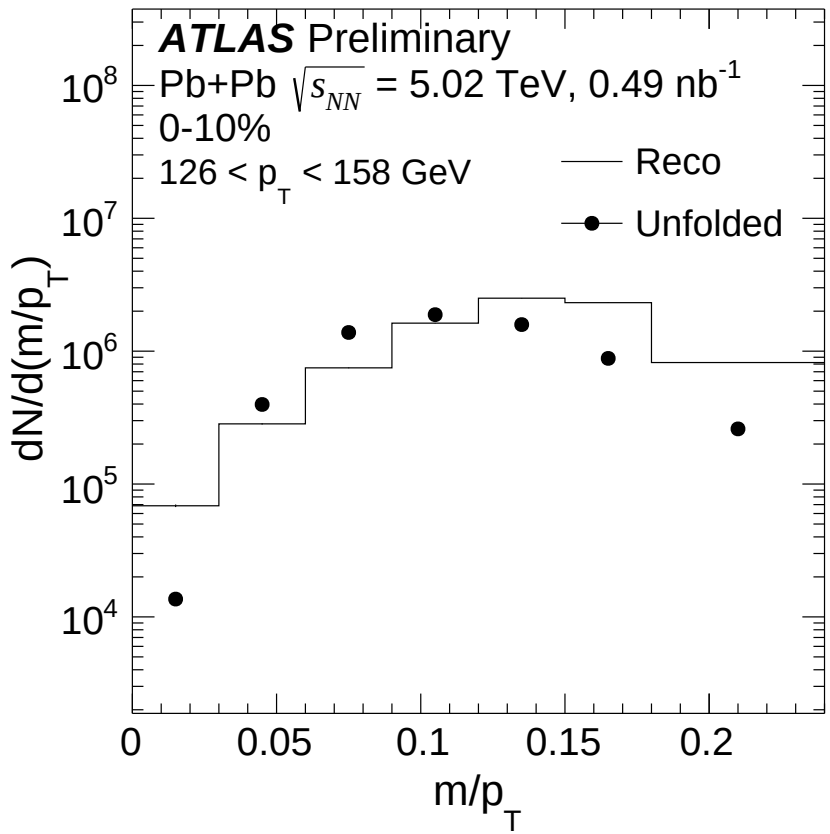




Jet mass measurement in pp and Pb+Pb collisions

- Measurement of m/p_{T} in pp and PbPb collisions at $\sqrt{s} = 5.02$ TeV ATLAS-CONF-2018-014
- Jet mass measurement in PbPb gives insight in the modification of the jet while passing through quark-gluon plasma
- Jets are reconstructed with anti-k_t $R = 0.4$ algorithm from 0.1×0.1 towers
  - Underlying event is subtracted from towers before calibrations are applied



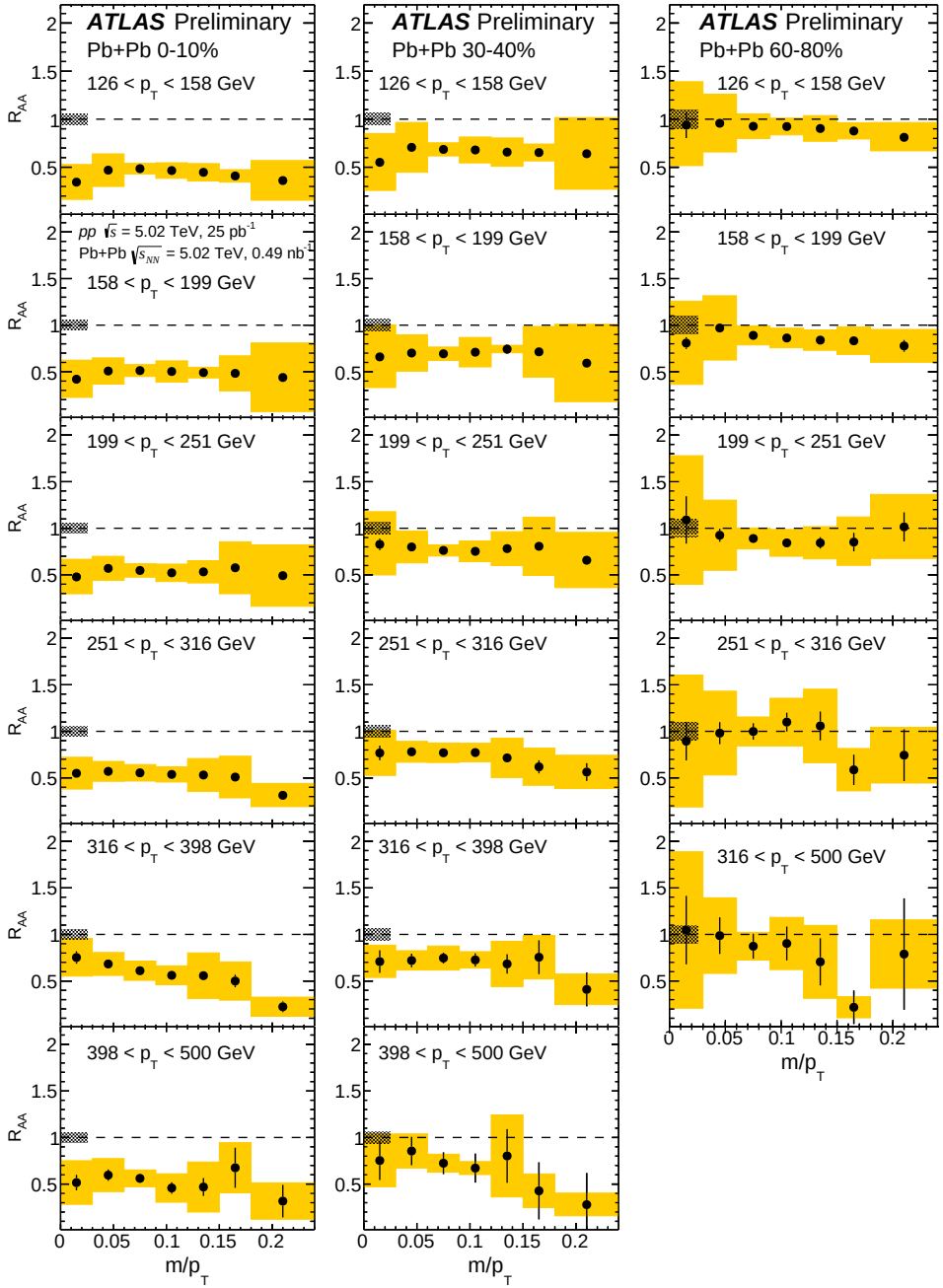


Jet mass measurement in Pb+Pb collisions

- **Nuclear modification factor** R_{AA}:
  - Quantifies modification of jet yields PbPb collisions relative to *pp*

$$R_{\mathrm{AA}}(m/p_{\mathrm{T}}, p_{\mathrm{T}}) = \frac{\left.\frac{1}{N_{\mathrm{evt}}} \frac{\mathrm{d}N_{\mathrm{jet}}^{\mathrm{Pb+Pb}}}{\mathrm{d}(m/p_{\mathrm{T}})}(p_{\mathrm{T}})\right|_{\mathrm{cent}}}{\langle T_{\mathrm{AA}} \rangle \frac{\mathrm{d}\sigma_{\mathrm{jet}}^{pp}}{\mathrm{d}(m/p_{\mathrm{T}})}(p_{\mathrm{T}})}$$

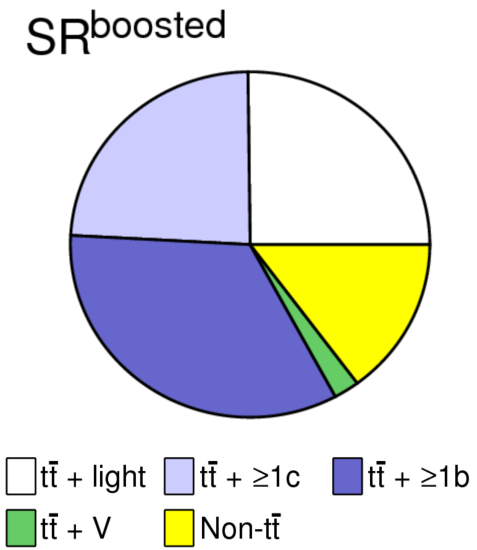
- $\langle T_{\mathrm{AA}} \rangle$: nuclear thickness function
- R_{AA} measured for various p_{T} and centralities (based on E_{T} in FCal)
- Modification factor is relatively flat in m/p_{T} but shows suppression of jets in PbPb collisions



Search for SM Higgs production in association with $t\bar{t}$

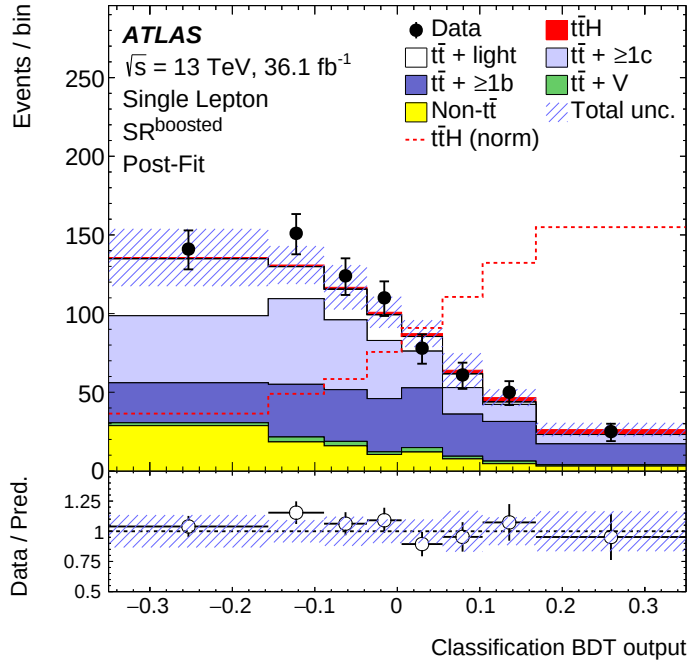
- Search for $t\bar{t}H$ production performed with 36.1 fb^{-1} of 13 TeV pp collisions arXiv:1712.08895, Phys. Rev. D 97, 072016 (2018)
- $t\bar{t}H$ process allows for direct measurement of the top quark Yukawa coupling
- Search designed for $H \to b\bar{b}$ decay and for dilepton & lepton+jets decay channel of top quark pair
- Search divided in different signal regions based on number of jets and number of b-tagged jets
- Dedicated **boosted** channel in analysis
  - Decay of top quark and Higgs boson each reconstructed as one anti-k_t $R = 1.0$ jet
    - Inputs are anti-k_t $R = 0.4$ jets $\to$ **reclustering**
  - Higgs candidate: two subjets that are b-tagged
  - Top candidate:
    - one subjet that is b-tagged
    - at least one other subjet that is not b-tagged

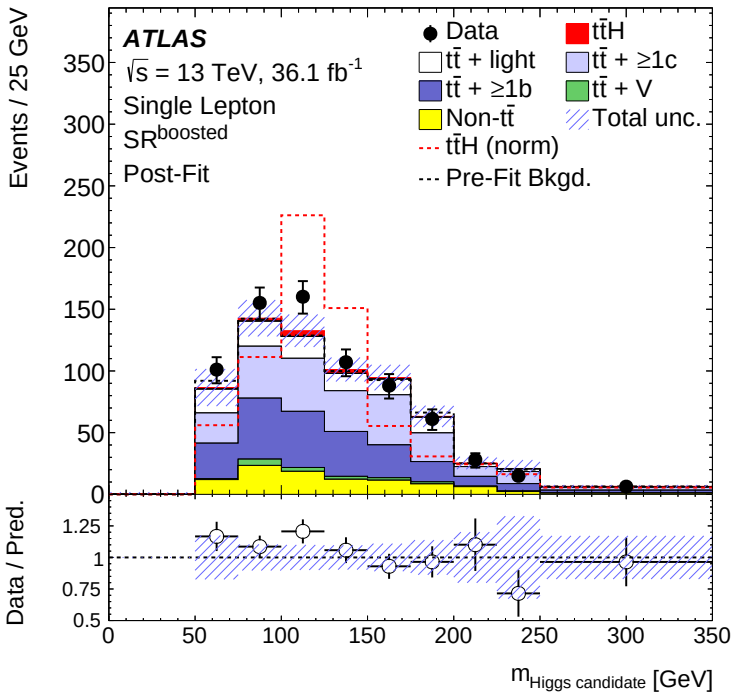
Bkg composition



Boosted $t\bar{t}H$ - multivariate analysis techniques

- Eight variables are combined in BDT
  - ΔR between objects, Higgs mass candidate, b-tagging weights
- Top three variables with highest discrimination:
  - k_{T} splitting scale of top candidate: $\sqrt{d_{12}}$
  - ΔR between top and Higgs
  - ΔR between top and additional b-jet





- Discriminants of various analysis categories combined in profile likelihood fit
- Best-fit μ value:
  $\mu = 0.84 \pm 0.29\ (\mathrm{stat.})\ ^{+0.57}_{-0.54}\ (\mathrm{syst.})$
- Observed data consistent with both the bkg-only hypothesis and with the SM $t\bar{t}H$ production

Boosted $t\bar{t}$ differential cross-section at $\sqrt{s} = 13$ TeV

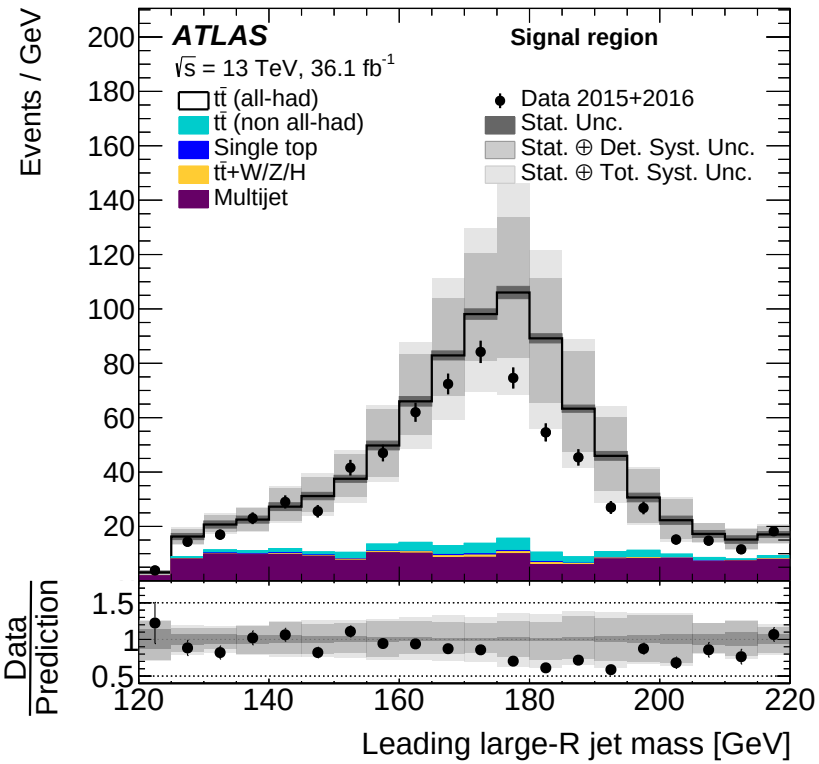
- Theoretical calculations in the TeV scale range still present large uncertainties
- Measurement of boosted $t\bar{t}$ cross-section targets this region of phase space
  - Performed in the all-hadronic channel using 36.1 fb^{-1} (arXiv:1801.02052)

Top quark candidate

- anti-k_t $R = 1.0$ trimmed jets
- $p_{\mathrm{T,j1(j2)}} > 500$ (350) GeV
- One associated small-R b-tagged jet
- Top tagged (based on mass and τ_{32})

Background estimation

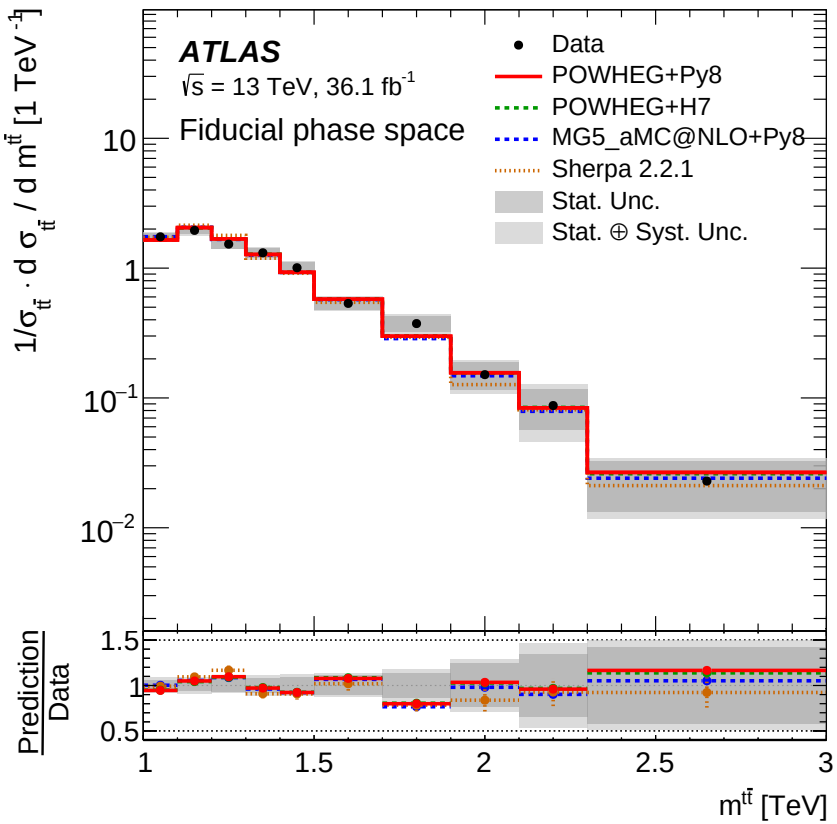
- MC: $t\bar{t}$ (non all-had), single top & $t\bar{t}$+X
- Data-driven multijet bkg estimate
  - 16 regions defined based on number of top tags and b-tags

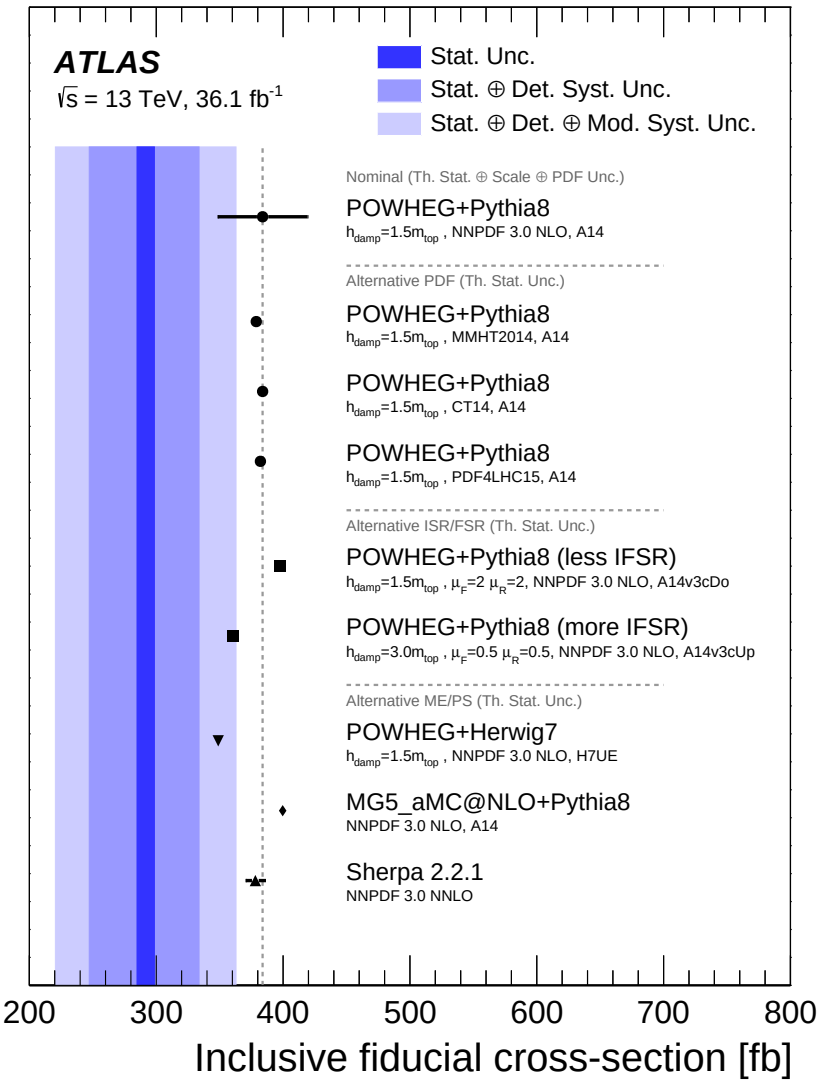


Boosted $t\bar{t}$ differential cross-section - Results I

- Detector-level distribution unfolded to particle-level **fiducial phase-space**
- Differential and inclusive fiducial cross-section measured
- Large-R jet energy scale & top tagging uncertainties dominate

$$\sigma_{\text{fid}} = 292 \pm 7\ (\text{stat}) \pm 76\ (\text{syst})\ \text{fb}$$

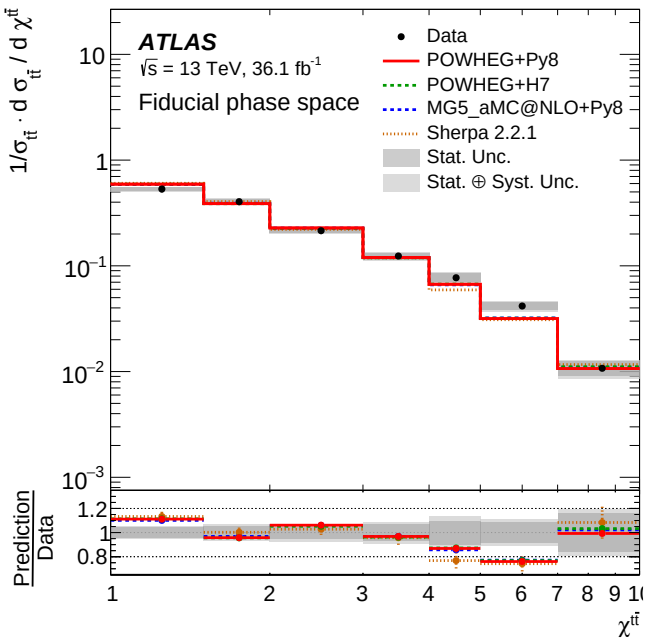




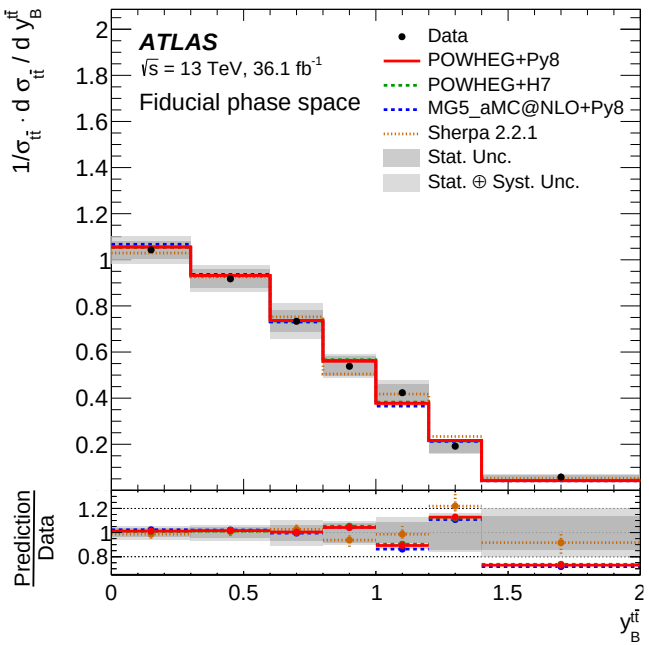
Boosted $t\bar{t}$ differential cross-section - Results II

- $\chi^{t\bar{t}} = \exp 2|y^*|,\ y^* = \frac{1}{2}(y^{t,1} - y^{t.2})$
  - Sensitive to processes not included in Standard Model at low $\chi^{t\bar{t}}$ values
- $y_{\mathrm{B}}^{t\bar{t}} = \frac{1}{2}(y^{t,1} + y^{t.2})$
  - Longitudinal motion of the $t\bar{t}$ system in the lab. frame is sensitive to PDFs
- $p_{\mathrm{out}}^{t\bar{t}}$
  - Out-of-plan momentum which is sensitive to additional radiation in the matrix element process

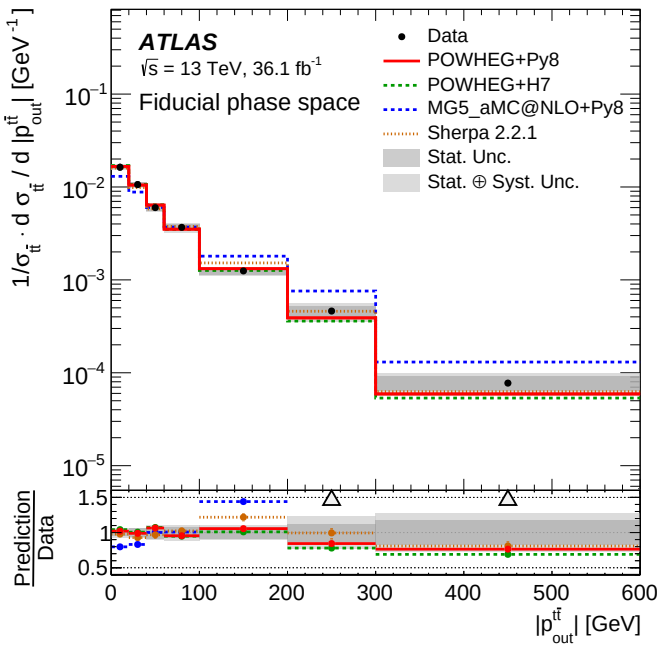
$\chi^{t\bar{t}}$



$y_{\mathrm{B}}^{t\bar{t}}$



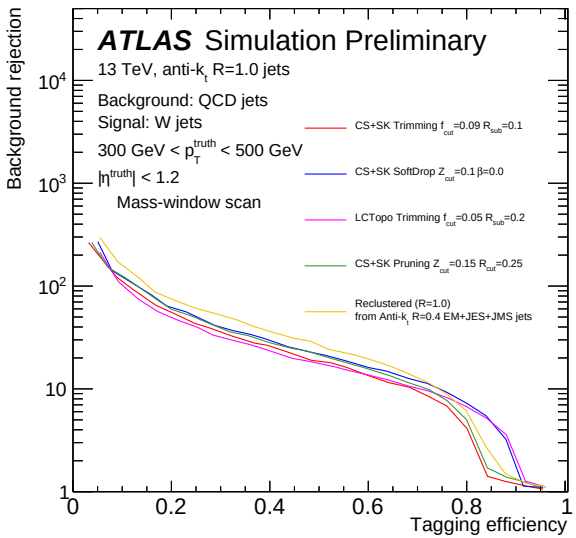
$p_{\mathrm{out}}^{t\bar{t}}$

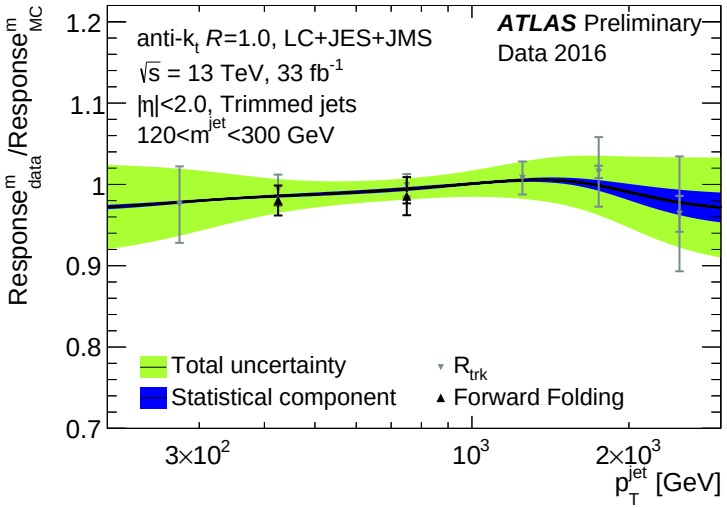


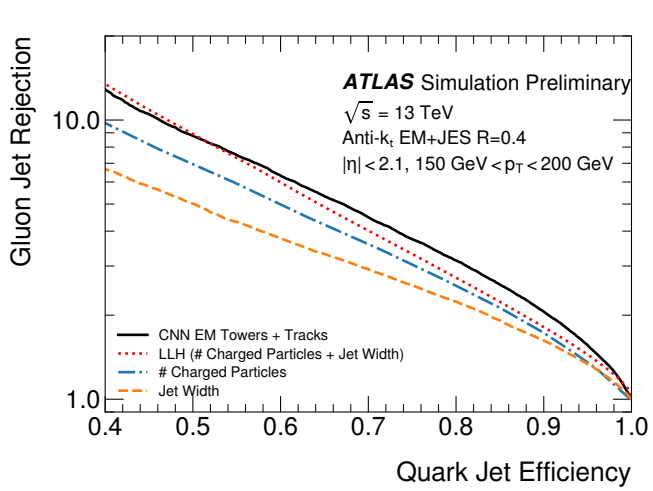
How to improve these measurements in the future

- Big effort on-going to define a new jet collection in conjunction with constituent-level pile-up suppression
- Large-R jet and modelling uncertainties dominating in many measurements
  - Suite of *in situ* calibration techniques deployed for large-R jets as for small-R
  - Ongoing work to connect *in situ* techniques on entire jets with low-level studies of jet constituents
- Improving tagging
  - Use state-of-the-art tagging techniques such as jet images for quark-gluon ID
  - Scale factors to correct for tagger efficiency differences in data and MC

ATL-PHYS-PUB-2017-020 ATLAS-CONF-2017-063 ATL-PHYS-PUB-2017-017







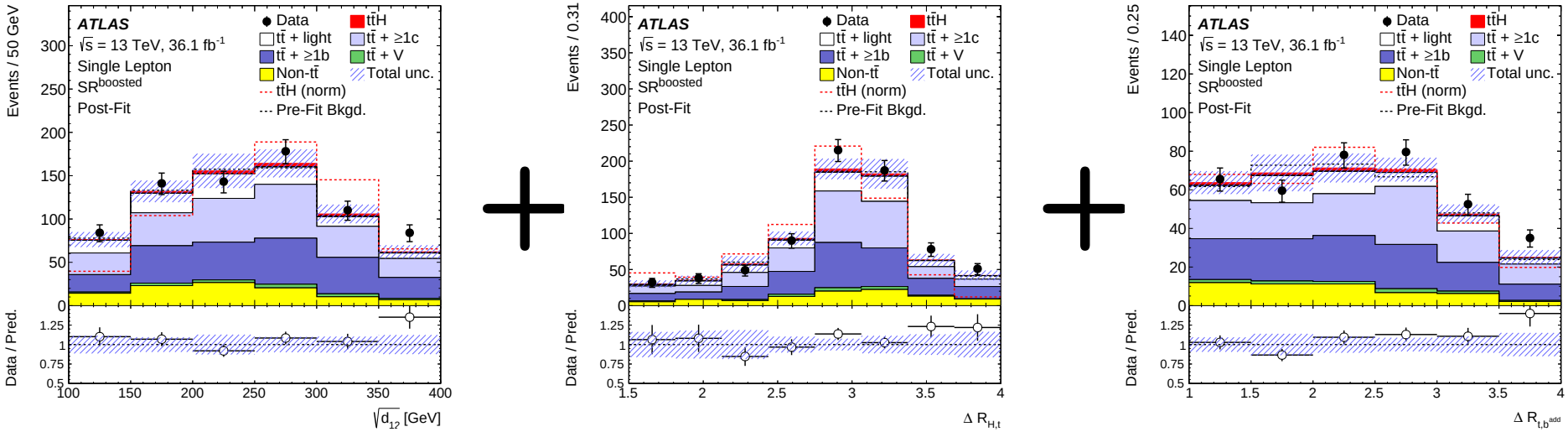
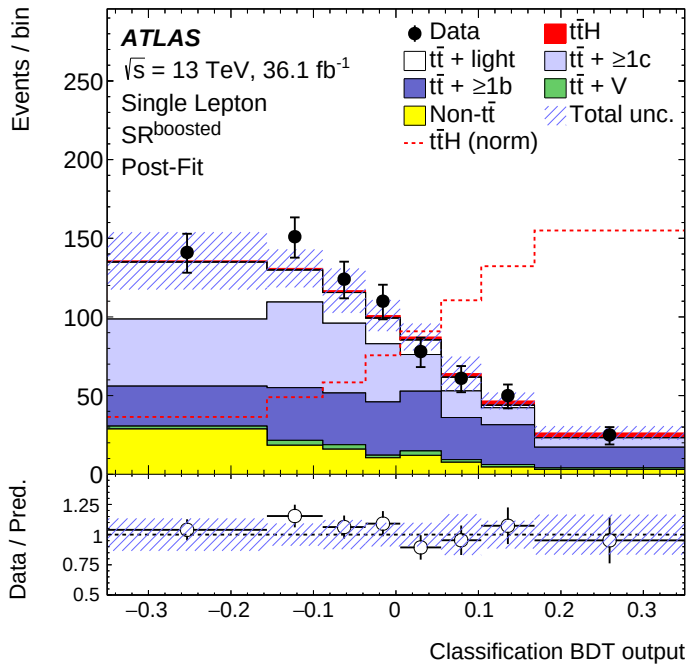
Summary

- Jet substructure techniques are now mainstream tools applied in QCD measurements, particle ID in searches for rare SM processes and new physics
- Jet substructure and tagging techniques are an exciting field with significant ongoing development
- The usage of substructure techniques allows us to explore higher p_{T} regimes that were not probed before to constrain the SM
- Measurements are mostly limited by experimental uncertainties related to the large-R jet kinematics and theory uncertainties
  - Expect significant reduction from *in situ* calibration techniques for the jet p_{T}
  - New *in situ* calibration techniques like forward folding or bottom up unc. used to understand jet mass scale uncertainties but more work needed
  - Work needs to be done to reduce modelling uncertainty dependence, communication with theory community is crucial

Backup

Boosted $t\bar{t}H$ - multivariate analysis techniques

$\sqrt{d_{12}}$ $\Delta R(t, H)$ $\Delta R(t, b)$



- Eight variables are combined in BDT
  - ΔR between objects, Higgs mass candidate, b-tagging weights
- Top three variables with highest discrimination:
  - k_{T} splitting scale of top candidate: $\sqrt{d_{12}}$
  - ΔR between top and Higgs
  - ΔR between top and additional b-jet