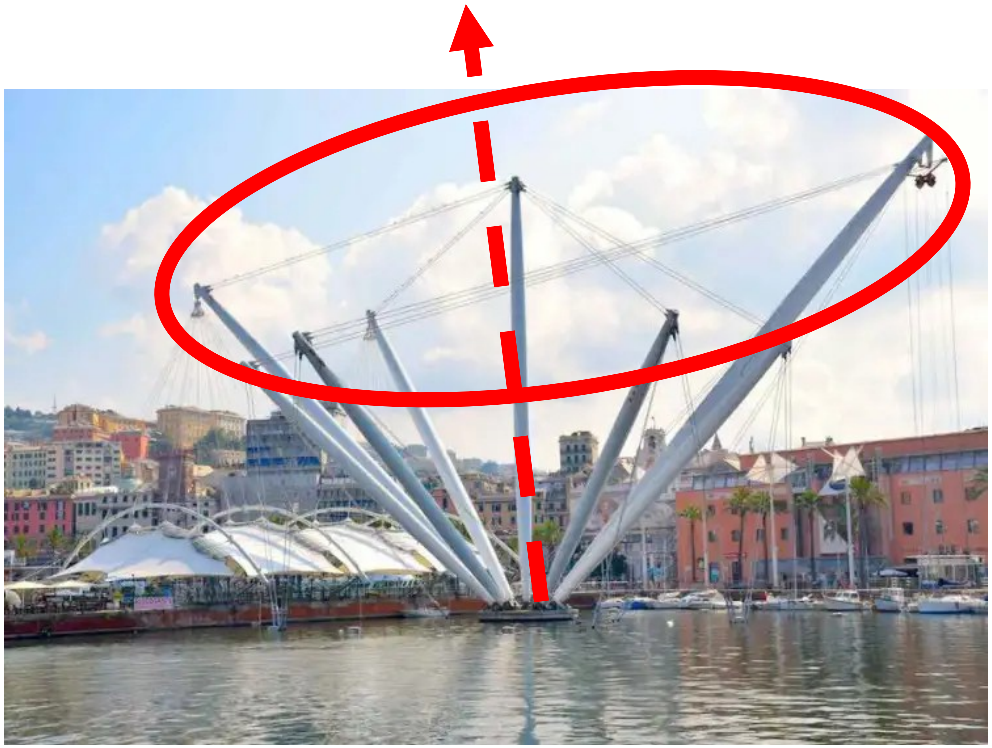
Top-quark jet substructure measurements with ATLAS

Mario Campanelli (UCL)
On behalf of the
ATLAS collaboration

BOOST 2024
Genova



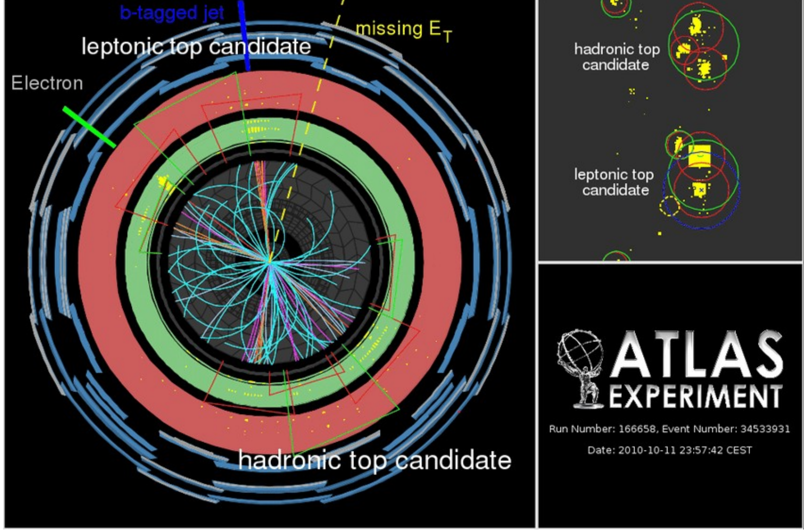
Why measure substructure for top quarks

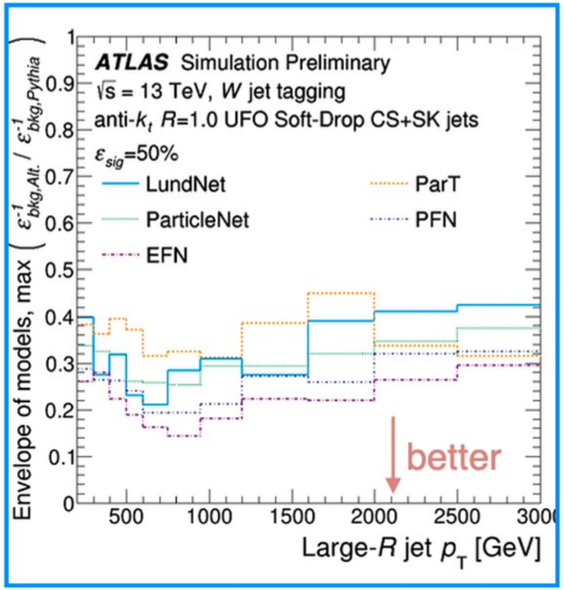
Jets from hadronic top decays are complex and interesting, with a b-jet (secondary vertex) and a color-singlet W.

The BSM preference for the third generation made top tagging a hot topic for several years

Complexity of final state means measurements are fundamental to reduce tagging systematics, especially for the most sophisticated ML tools

eg. for W tagging, the more complex (and powerful) the tool, the larger the modelling systematics





Final states for different measurements

Top quarks can be selected with high purities and final state can be optimised for the study of interest

- For b-jet studies (like b fragmentation), fully leptonic decays offer an environment free of extra hadronic radiation; displaced tracks offer a proxy for the b hadron decay products
- To measure classic substructure variables in the top jet, use hadronic decays of one top (with the other hadronic or leptonic)
- For separate studies of the top and W jets (Lund Plane), select a large-R jet in events with b-tagging, and distinguish if b is inside jet (top) or outside (W)

B-hadrons fragmentation in b-jet

PRD 106, 032008 (2022)

Why measuring them again after e^+e^- (aren't they universal?):

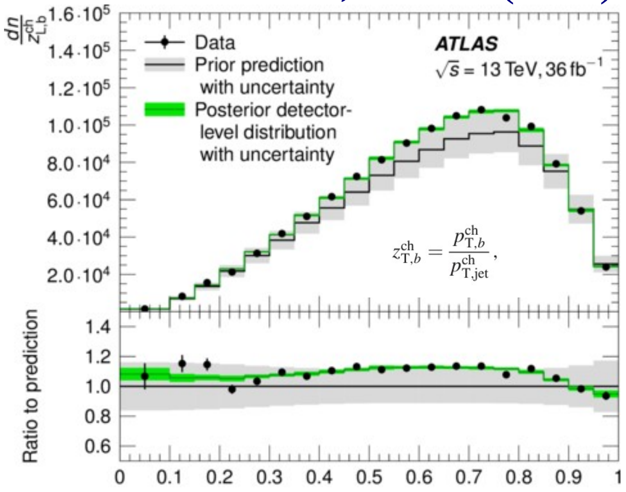
- Color connections between top and b
- Energy of b- quarks has wide range
- Better MC/unfolding tools

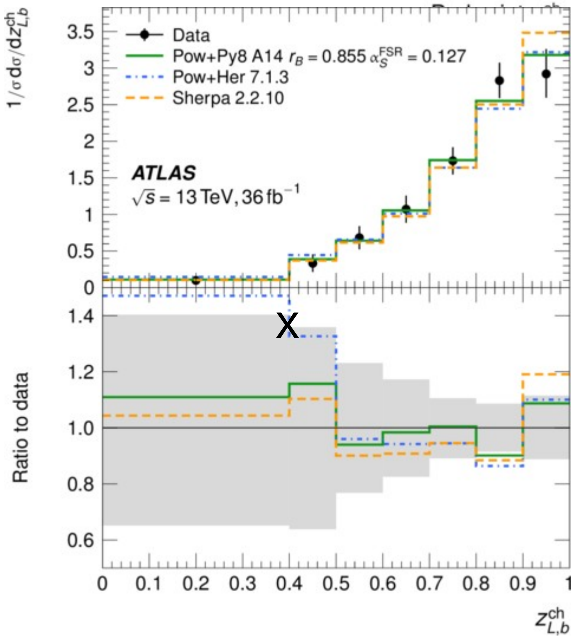
Analysis used 36/fb of data from 2015 and 2016, in the very pure exclusive eμννbb final state.

Require one jet as b-tagged (tag) and the other to have a secondary vertex with > 2 tracks and fully contained in tracker (probe).

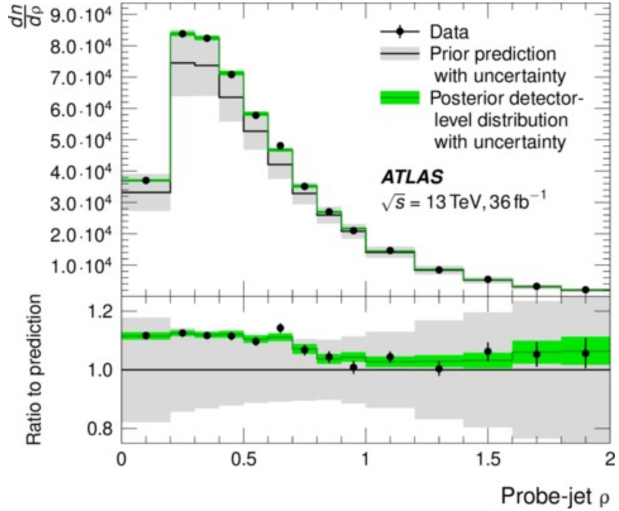
Measurement performed on charged particles of probe jets (0, 1 or 2 per event)

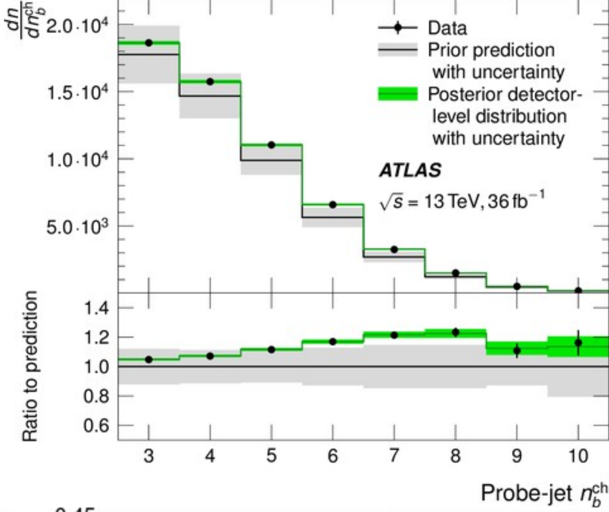
Fully Bayesian unfolding, with rescaled MC as prior. Clearly observe a shift towards lower z values in detector





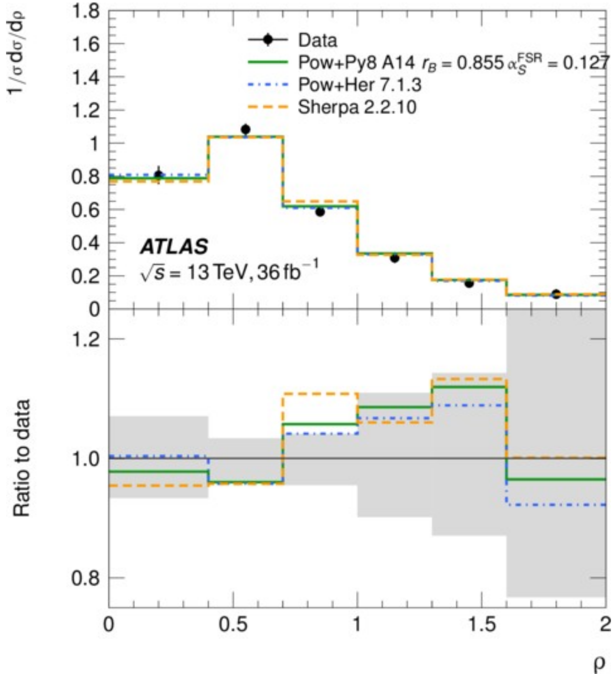
“QCD” observables
Detector level



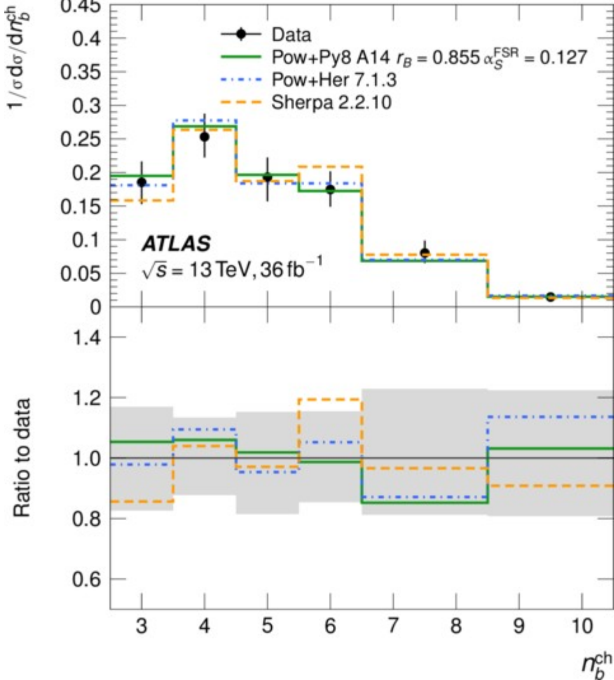


$$\rho = \frac{2 p_{\mathrm{T},b}^{\mathrm{ch}}}{p_{\mathrm{T}}^{e} + p_{\mathrm{T}}^{\mu}},$$

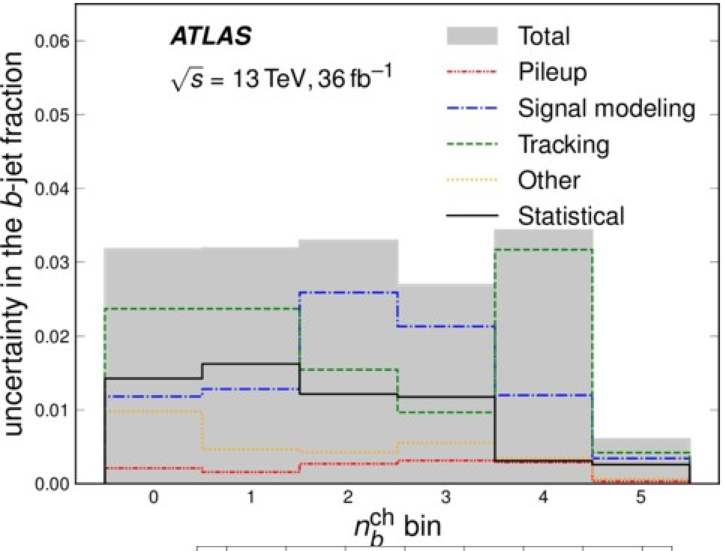
Particle level



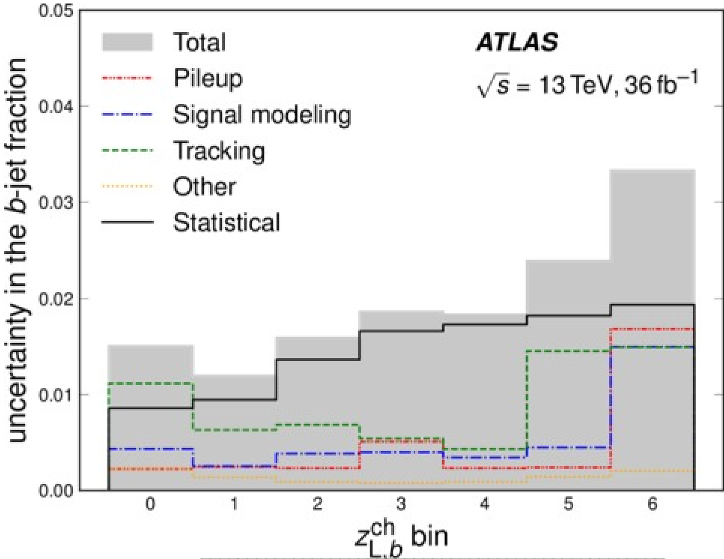
Number of charged particles in b hadron decay



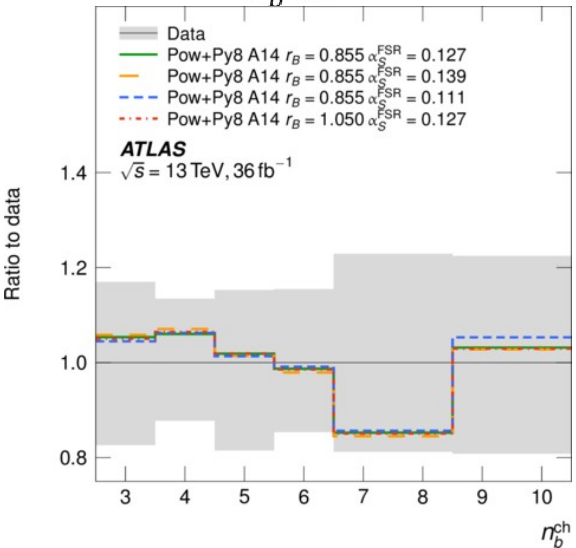
Systematics and theory comparison



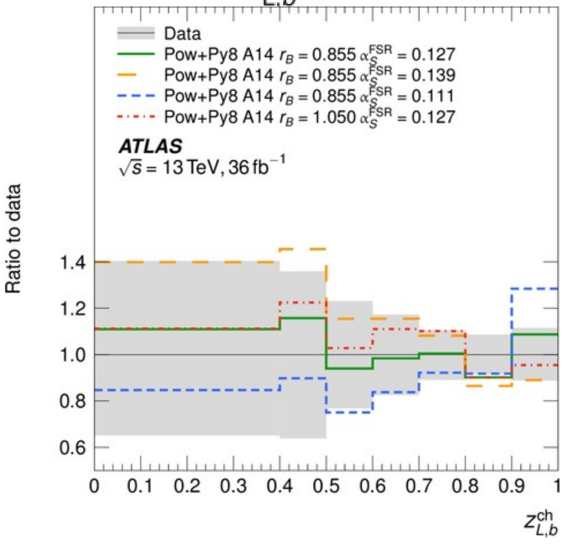
Tracking and modelling dominate for track multiplicity



But largely cancel out in ratios, where statistics dominates



MC fragmentation parameters largely tuned on LEP data, changing the FSR αs and the parameter rb connected to the b mass has no effect on track multiplicity but quite a lot on z



"Classic" substructure

PRD 109, 112016 (2024)

Even if nowadays just used as a reference for top tagging, variables like n-subjettiness, LH Angularity, Energy Correlation functions are still among the best to describe jet inner structure (and bring nostalgic souvenirs to the oldest among us).

Definitions:

$$\lambda_\beta^\kappa = \sum_{i \in J} z_i^\kappa \left(\frac{\Delta R(i, \hat{n})}{R} \right)^\beta .$$

Generalised angularities
λ_2^0 and $\lambda_1^{0.5}$ are pT dispersion and Les Houches Angularity

$$\mathrm{ECF}(N) = \sum_{i_1 < i_2 < \ldots < i_N \in J} \left(\prod_{a=1}^{N} p_{\mathrm{T},i_a} \right) \times \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^{N} \Delta R(i_b, i_c) \right).$$

Energy Correlation Fractions, with N the number of prongs

$$\mathrm{ECF2} = \frac{\mathrm{ECF}(2)}{\mathrm{ECF}(1)^2} .$$

$$C_3 = \frac{\mathrm{ECF}(4)\,\mathrm{ECF}(2)}{\mathrm{ECF}(3)^2} ,$$

$$D_2 = \frac{\mathrm{ECF}(3)\,\mathrm{ECF}(1)^3}{\mathrm{ECF}(2)^3} .$$

C_3 (D_2) close to 0 means 3-body (2-body) structure

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T},k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\},$$
$$\text{with} \quad d_0 = \sum_k p_{\mathrm{T},k} R_0 .$$

Similarly, indicates the number of subjets In a jet. Usually used in ratios, like $\tau_{32} = \tau_3/\tau_2$
Indicating a 3-prong structure

Semileptonic vs fully hadronic events

Analysis performed in the cleaner semileptonic channel, and in the fully hadronic one, to test BG modelling and go to higher pT.

To minimise bias, jets are selected opposite to the lepton or to a top-tagged hadronic jet.

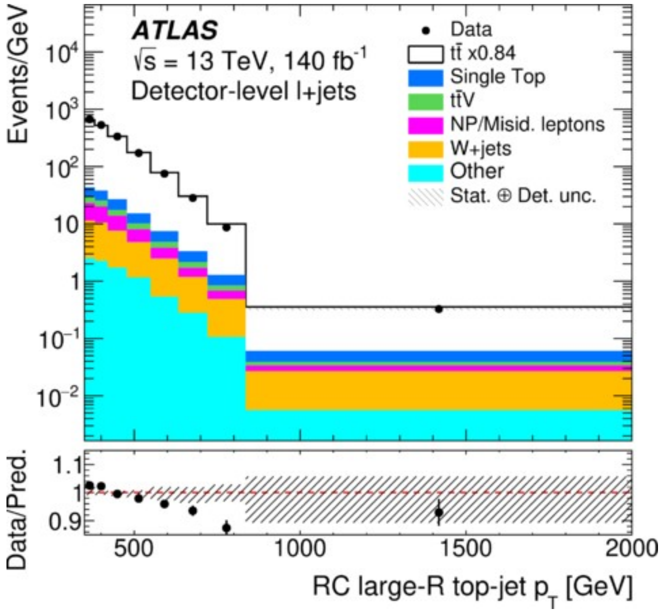
For semileptonic events, top jet is obtained by reclustering (RC) R=0.4 jets, hadronic events use R = 1 jets.

Substructure variables extracted from tracks in jets

Semileptonic

RC Jets

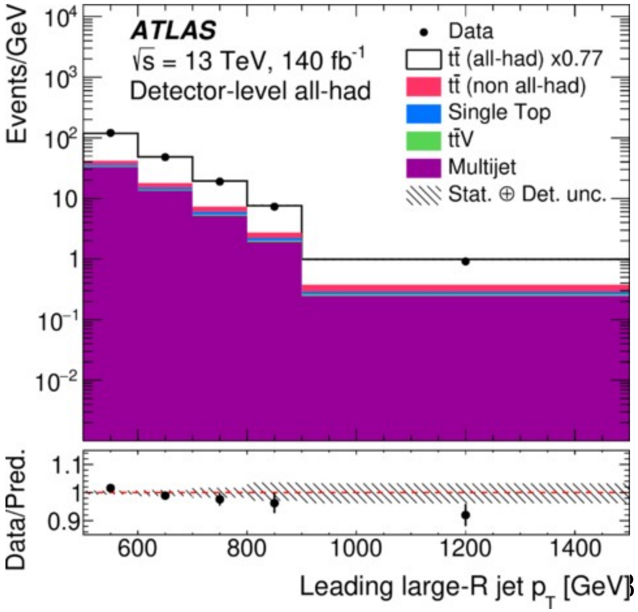
Less BG



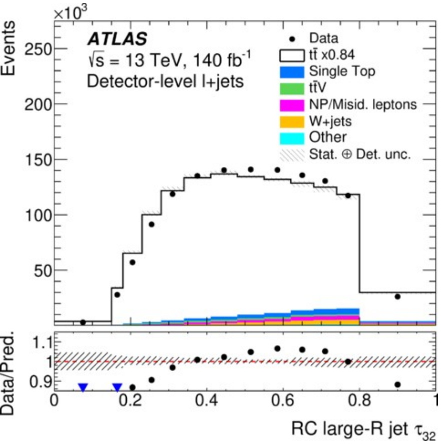
Fully Hadronic

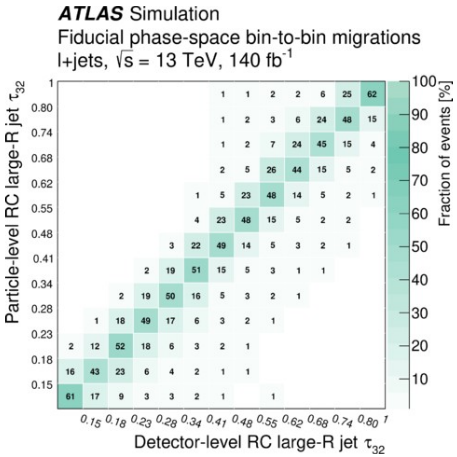
R = 1.0 jets

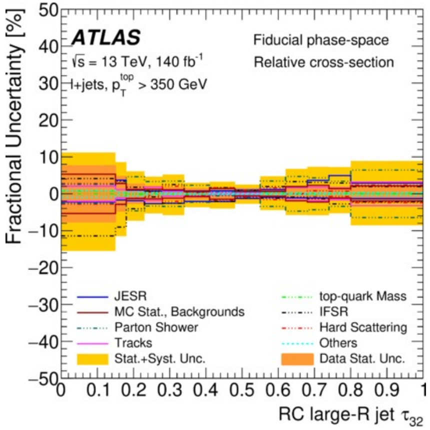
Higher pT reach

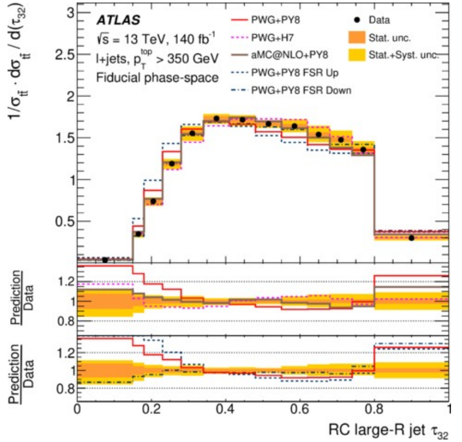


Substructure variables: the τ_{32} example

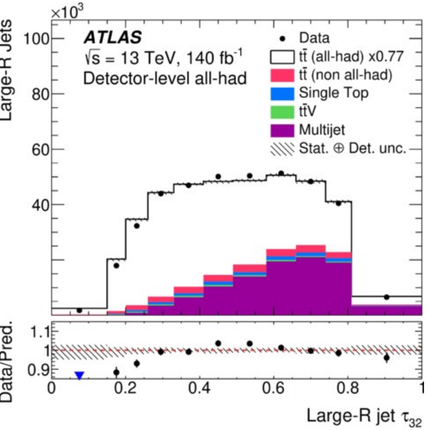


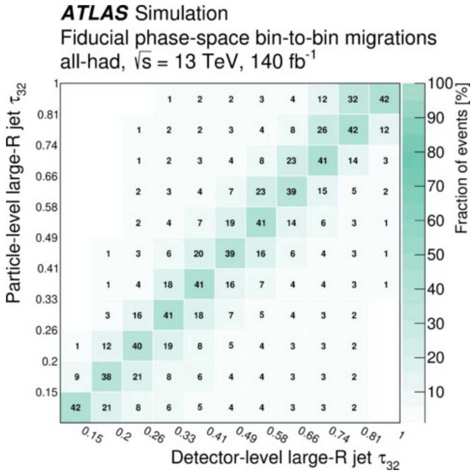


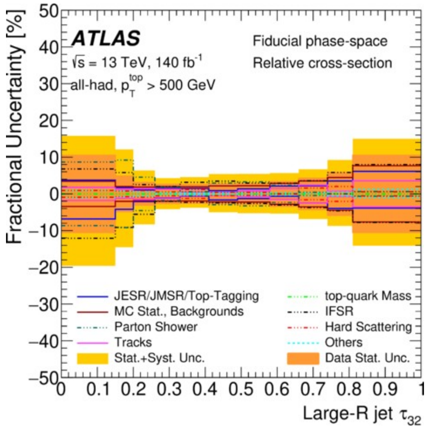


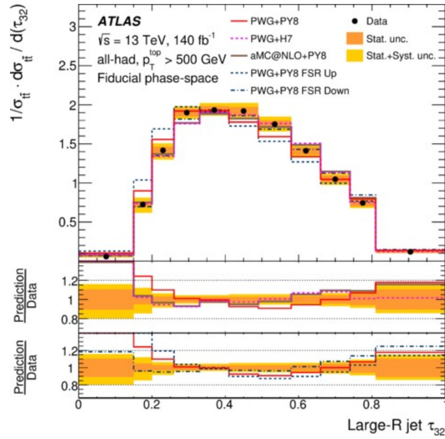


Detector level | Transfer matrix | Syst. Uncertainties | Particle level



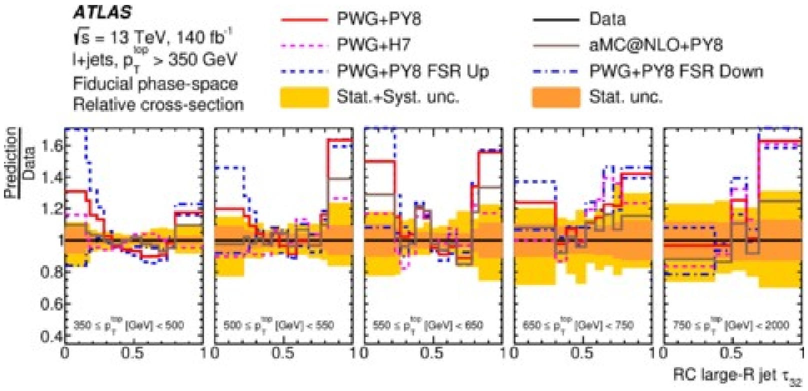


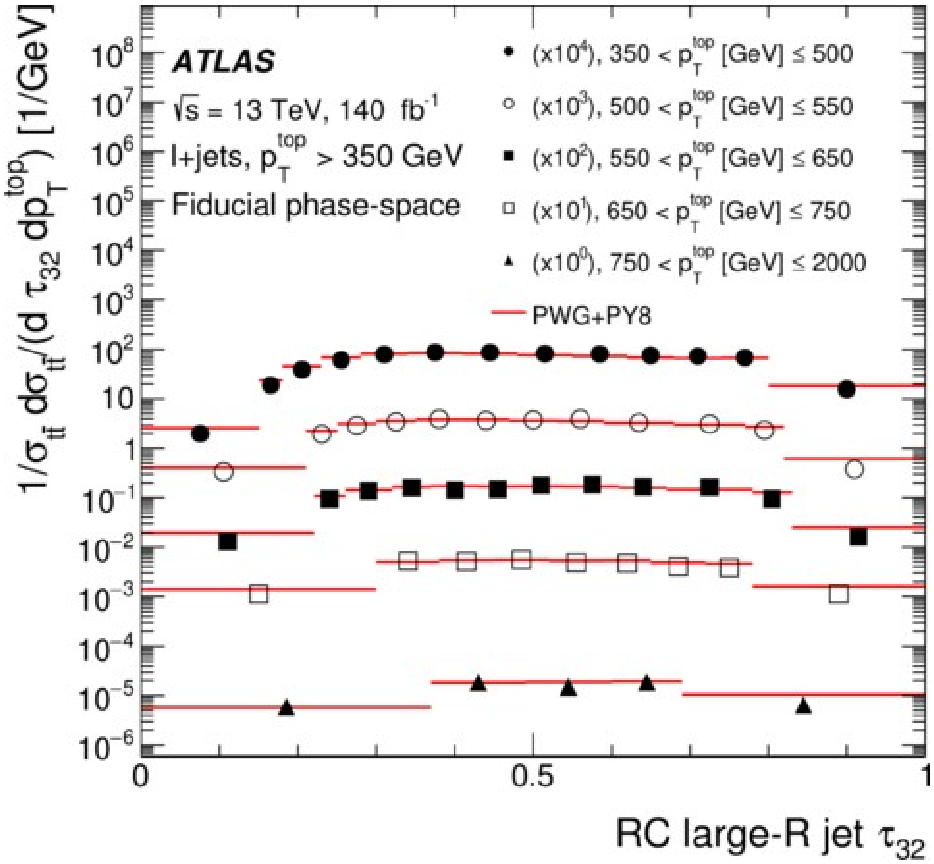




pT dependence

τ_{32} in pT bins





Similar results for mass dependence and for all other variables.

Quantitative model comparison

Observable	PWG+PY8		PWG+H7		AMC@NLO+PY8		PWG+PY8(FSR Up)		PWG+PY8(FSR Down)	
	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value
τ_{32}	54/12	<0.01	19/12	0.09	15/12	0.24	165/12	<0.01	40/12	<0.01
τ_{21}	14/14	0.41	7/14	0.92	16/14	0.32	42/14	<0.01	8/14	0.91
τ_3	36/11	<0.01	42/11	<0.01	14/11	0.23	130/11	<0.01	23/11	0.02
ECF2	25/18	0.13	13/18	0.78	15/18	0.69	31/18	0.03	24/18	0.14
D_2	20/16	0.20	17/16	0.39	20/16	0.20	37/16	<0.01	15/16	0.49
C_3	11/14	0.65	6/14	0.97	3/14	1.00	35/14	<0.01	3/14	1.00
$p_T^{d,*}$	27/12	<0.01	10/12	0.58	11/12	0.53	56/12	<0.01	24/12	0.02
LHA	14/17	0.65	9/17	0.92	20/17	0.29	14/17	0.69	19/17	0.32
D_2 vs. m^{top}	61/42	0.03	62/42	0.02	59/42	0.05	118/42	<0.01	44/42	0.37
D_2 vs. p_T^{top}	71/56	0.08	68/56	0.13	70/56	0.11	107/56	<0.01	93/56	<0.01
τ_{32} vs. m^{top}	153/42	<0.01	72/42	<0.01	56/42	0.07	413/42	<0.01	77/42	<0.01
τ_{32} vs. p_T^{top}	153/50	<0.01	103/50	<0.01	57/50	0.23	360/50	<0.01	114/50	<0.01

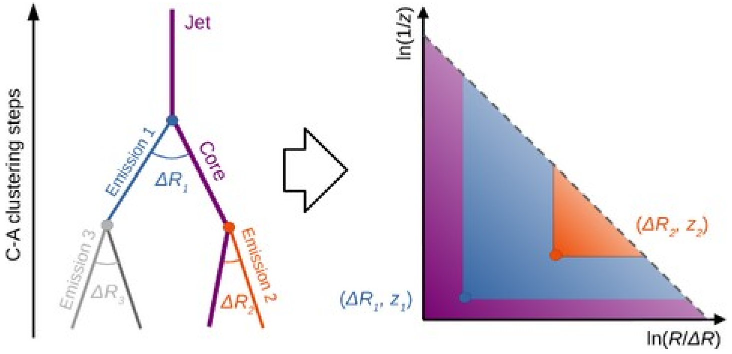
Semi-leptonic channel

Similar behaviour for all-hadronic channel.

Overall, PWG+Py8, especially with FSR changes and differential distributions, have small p-values

Measuring the Lund Jet Plane

arXiv:2407.10879



The Lund Diagram ([Z. Phys. C 43, 625–632 (1989)] is an abstract representation of the jet formation, where each branching is a point in a ln(ΔR/R), ln(1/z), ln(kT) space, usually projected in a 2D plane

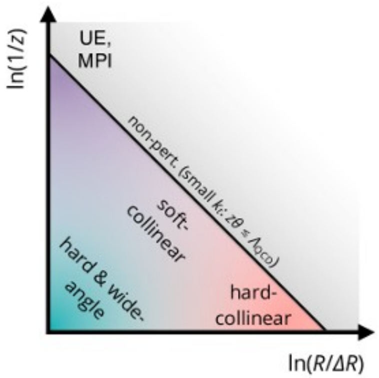
(notice, ATLAS uses ln(1/z) CMS ln(kT))

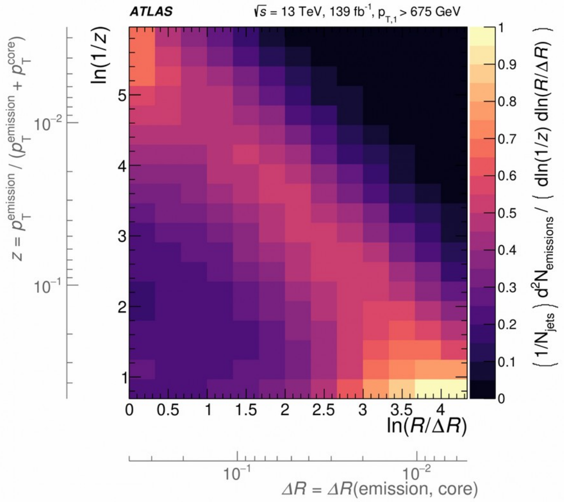
$$z = \frac{p_T^e}{p_T^e + p_T^c}; \qquad \Delta R = \sqrt{(y_e - y_c)^2 + (\phi_e - \phi_c)^2}$$

Experimentally, it can be reconstructed by running backwards the Cambridge /Aachen clustering algorithm,.

Each region of the plane corresponds to a different phase of jet evolution, allowing to disentangle them and analyse them separately.

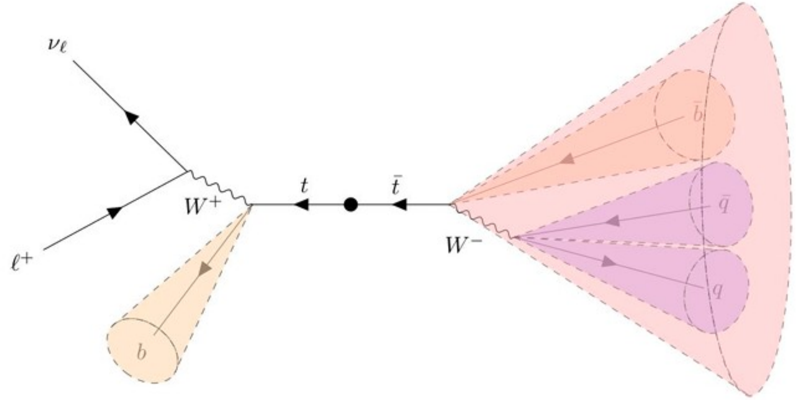
Charged particles LJP already measured by ATLAS for dijet events (Phys. Rev. Lett. 124, 222002)





Lund Jet Plane for Top and W jets

Semi-leptonic events, measuring the charged LJP on the hadronic side, reconstructed as a R = 1.0 jet



To avoid bias, no jet tagging applied

Is the R = 1.0 jet a top or W candidate?

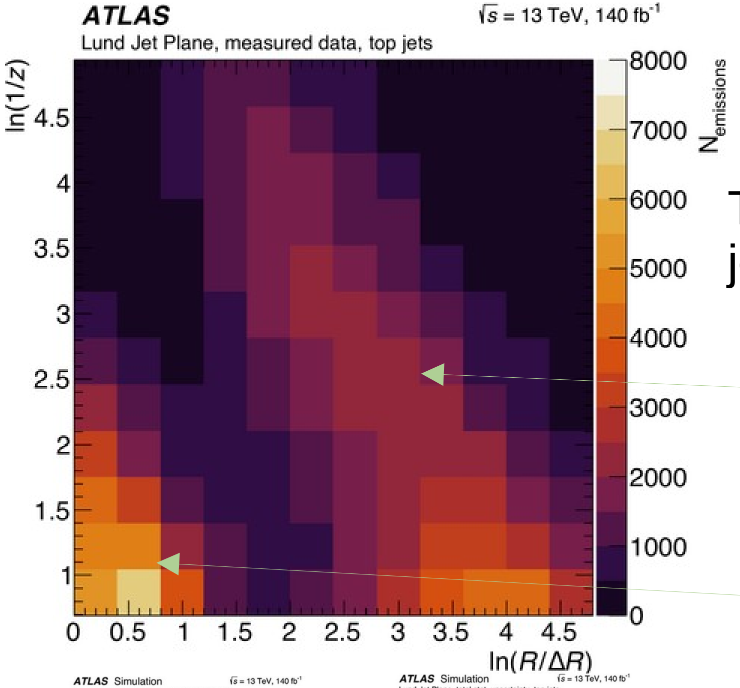
"top jet" selection

- $\Delta R(lepton,\ ljet) > 2.3$,
- $m_{ljet} > 140\,\mathrm{GeV}$,
- $+1$ b-tagged $R = 0.4$ jet, $\Delta R(bjet_2,\ ljet) < 1.0$.

"W jet" selection

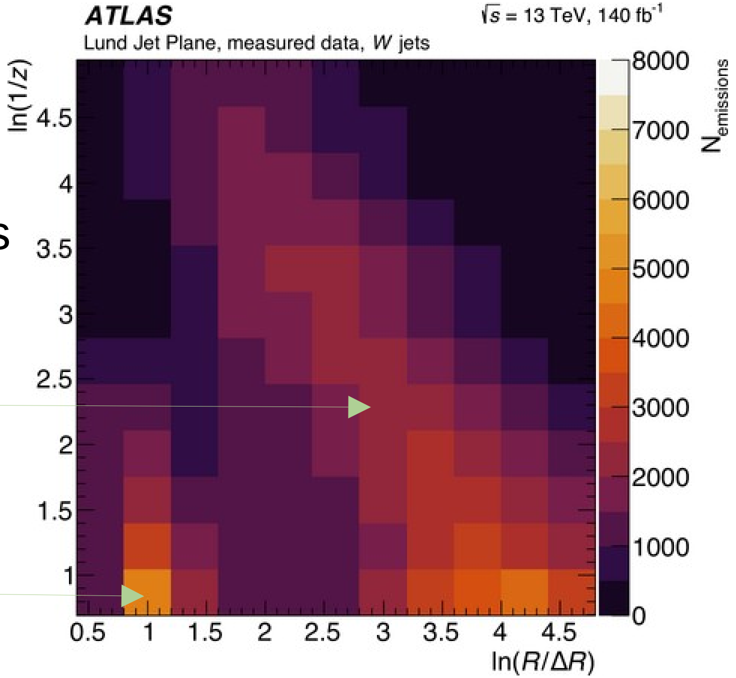
- $\Delta R(lepton,\ ljet) > 2.3$,
- $60\,\mathrm{GeV} < m_{ljet} < 100\,\mathrm{GeV}$,

Detector-level LJP and unfolding



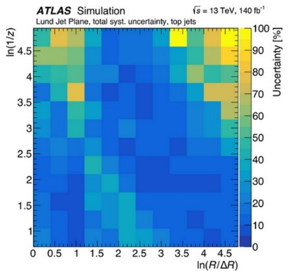
Top jets

W jets

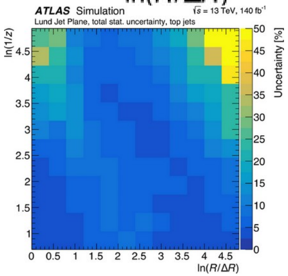


Soft collinear

Hard/ wide angle

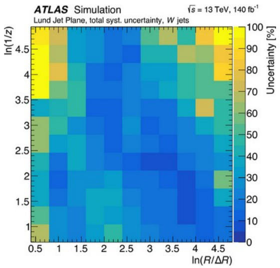


Systematics

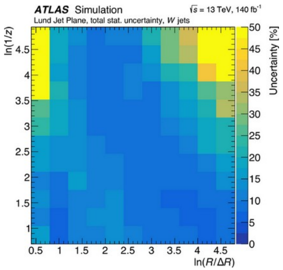


Statistics

Each emission is matched between detector and truth level according to distance in (η-φ) space. The plane is projected into a vector, corrected for efficiency and purity, and unfolded using a 4-iteration Bayesian unfolding.

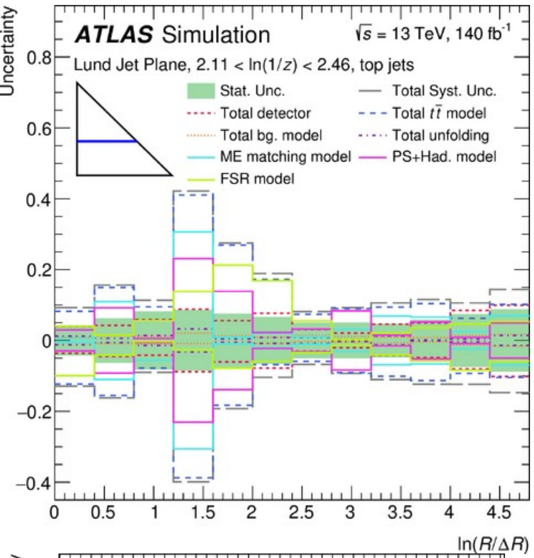


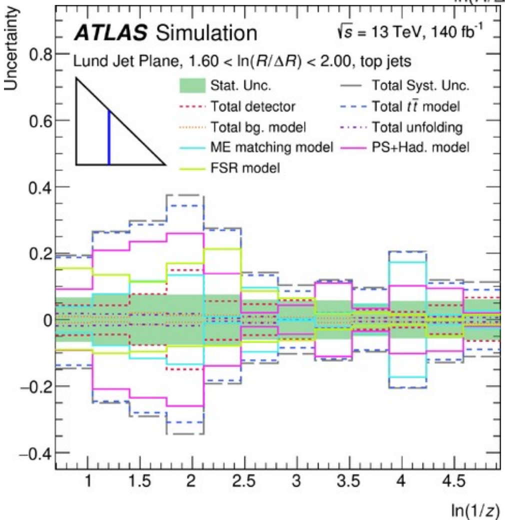
Systematics



Statistics

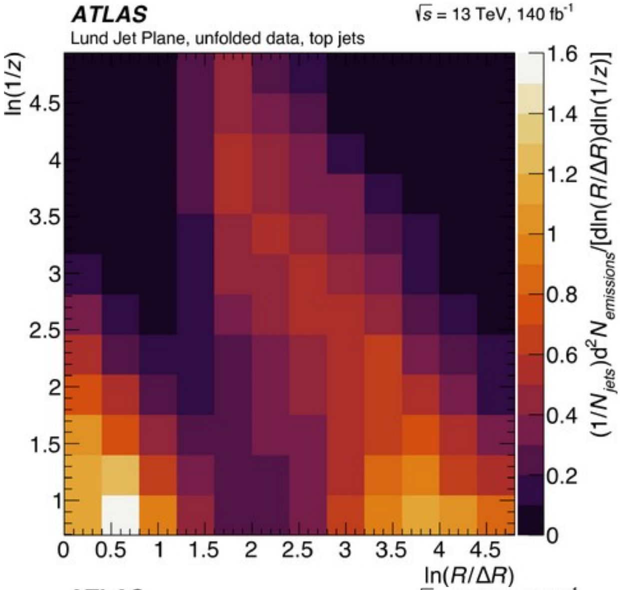
Systematic uncertainties and unfolded results



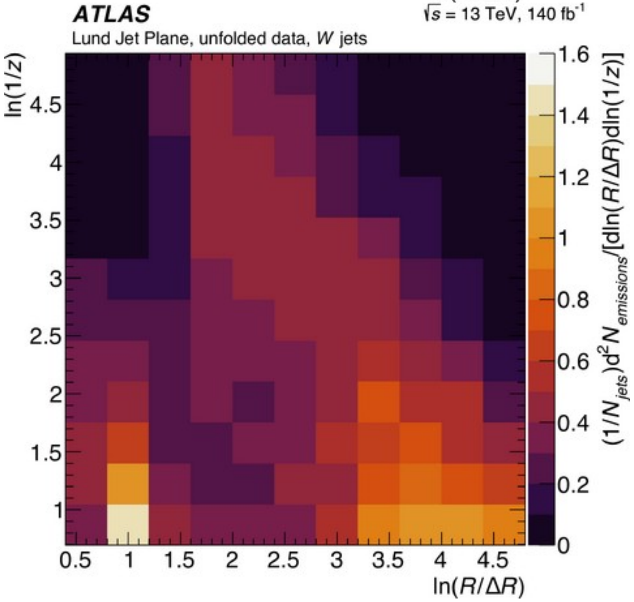


Dominated by modelling of the tt system and Parton Shower, obtained comparing different MC in the unfolding.

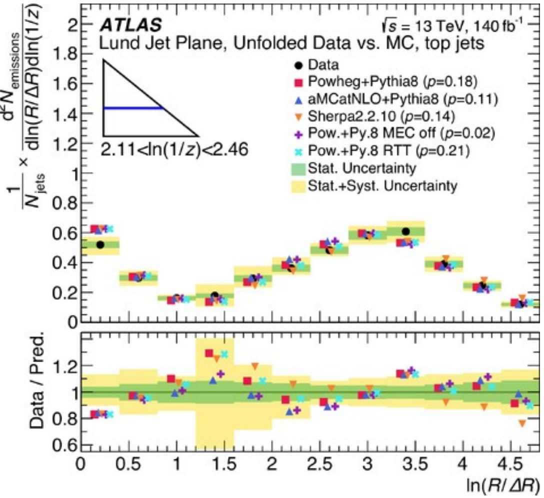
Top jets

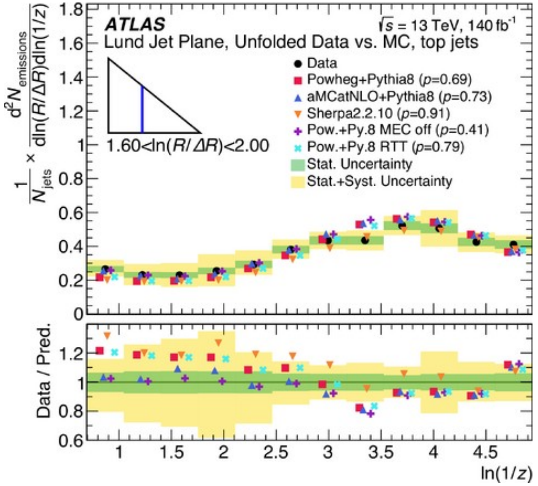


W jets



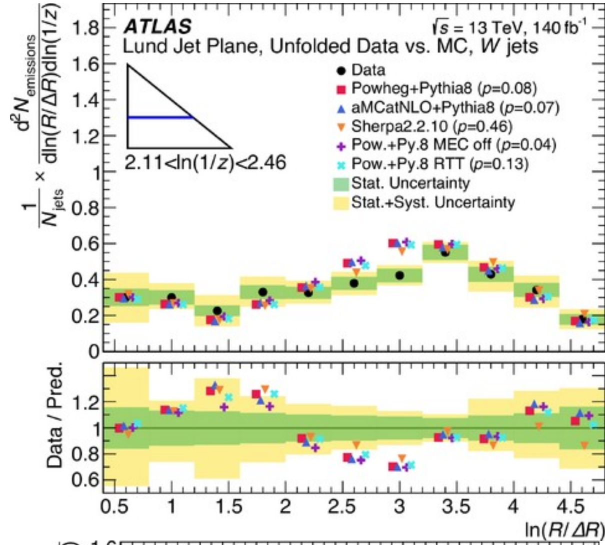
Selected slices of the planes

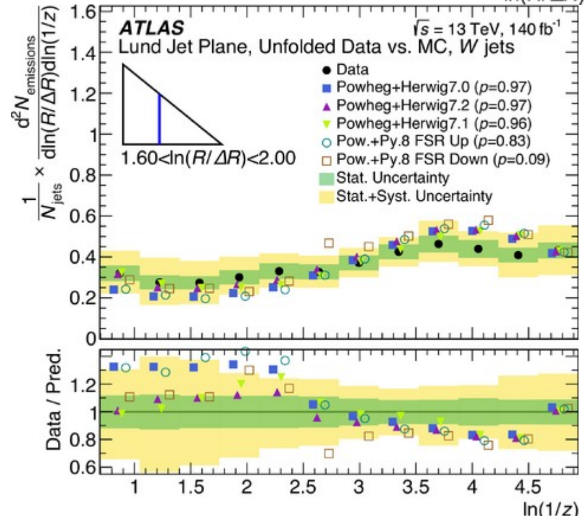




Top jets

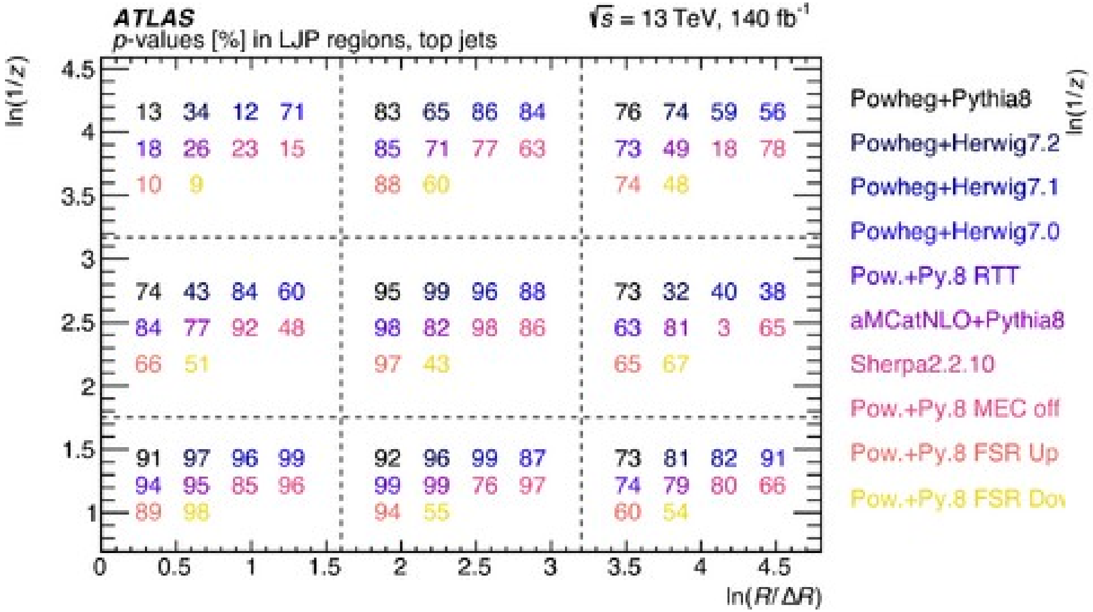
W jets



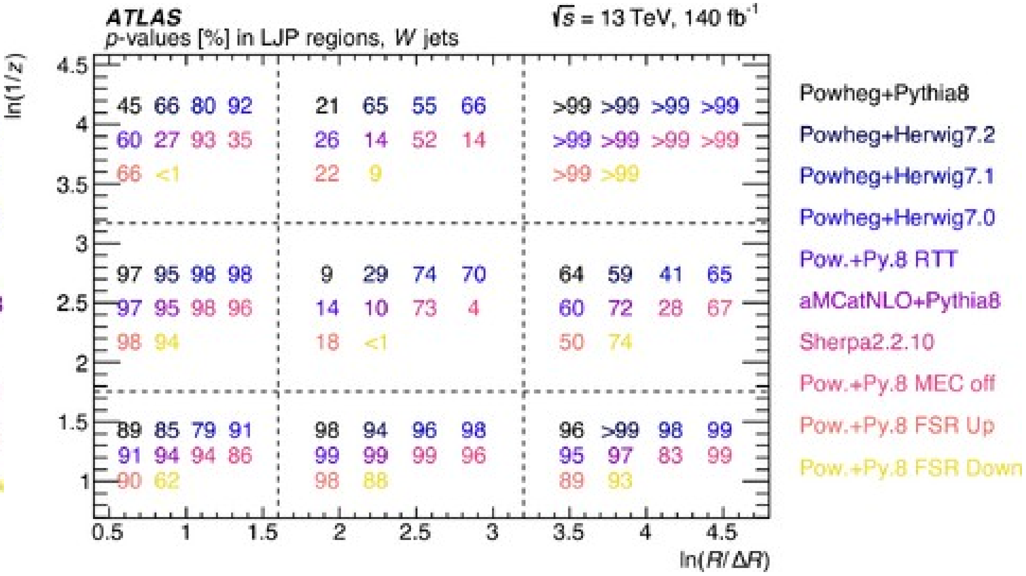


P-values for different models in different zones of the plane

Top jets



W jets



Usually good agreement for most models, some small p-values especially for W jets

Conclusions

- Always measure something before using it
- Apart for its use for tagging, substructure in top jets is complex and can teach a lot about QCD in various regimes
- Remarkable agreement for most of models and observables, but some corner of phase-space still to improve
- Ideal laboratory to study tuning and matching