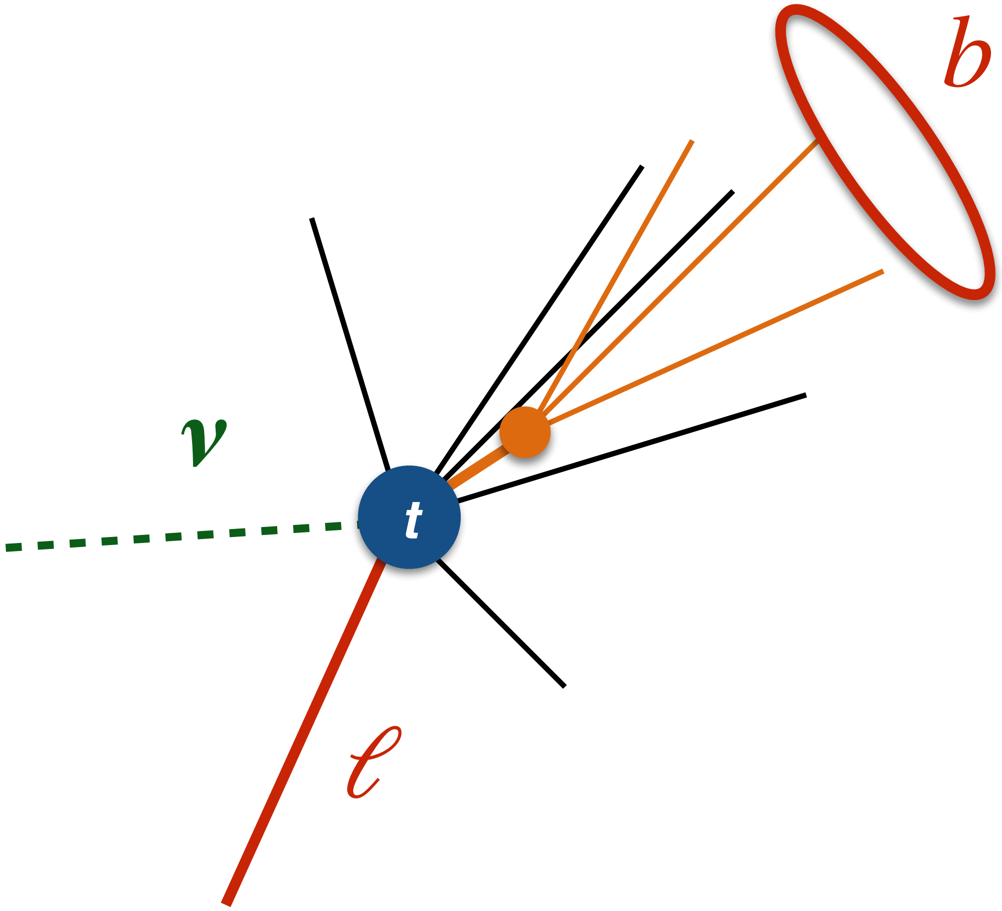
measuring b-fragmentation in $t\bar{t}$ events with ATLAS



Chris Pollard
on b-half of the ATLAS collaboration

LHC Top WG meeting
2020 11 23

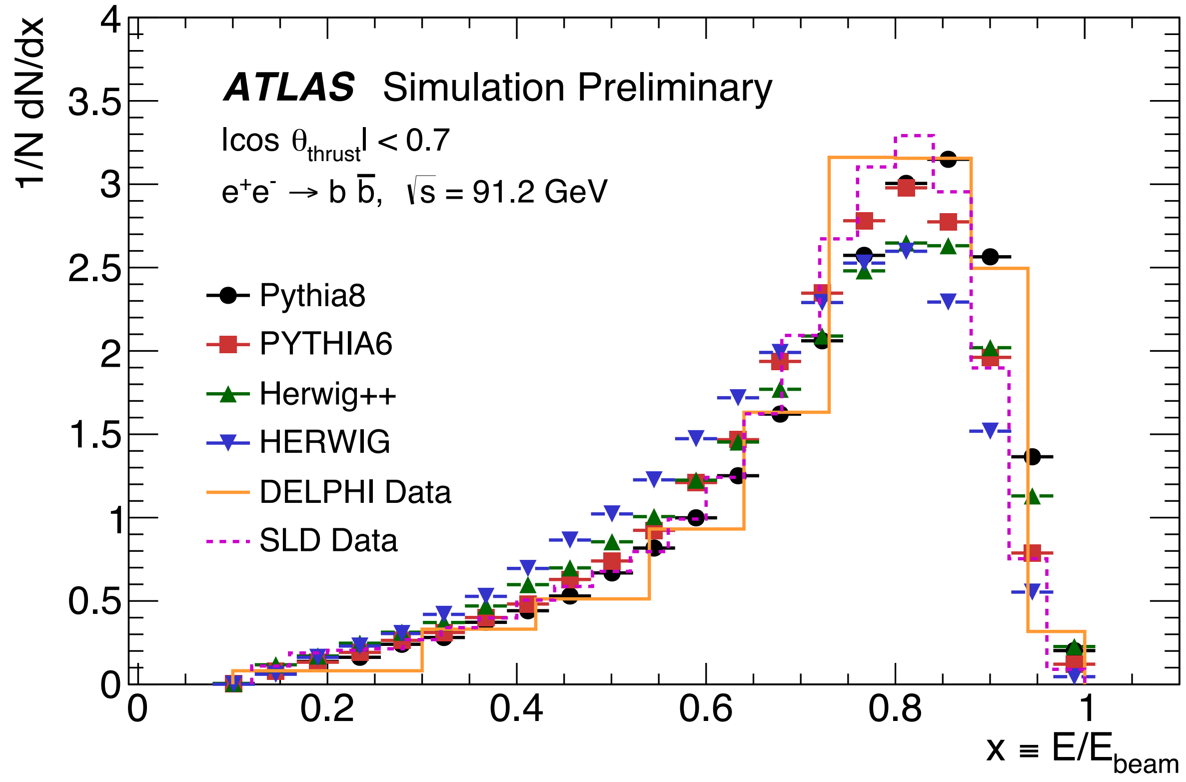
https://cds.cern.ch/record/2730444

ATLAS-CONF-2020-050

motivation I

the fragmentation of b-quarks into hadrons is of interest for many reasons.

- b-hadrons leave a *striking experimental signature* and…
- a *unique correspondence* to the originating b-quarks
  - ergo a *precise probe of QCD*
- b-fragmentation currently *tuned to e^+e^- data* …
- then *extrapolated to the LHC* environment
  - to what degree is this correct?

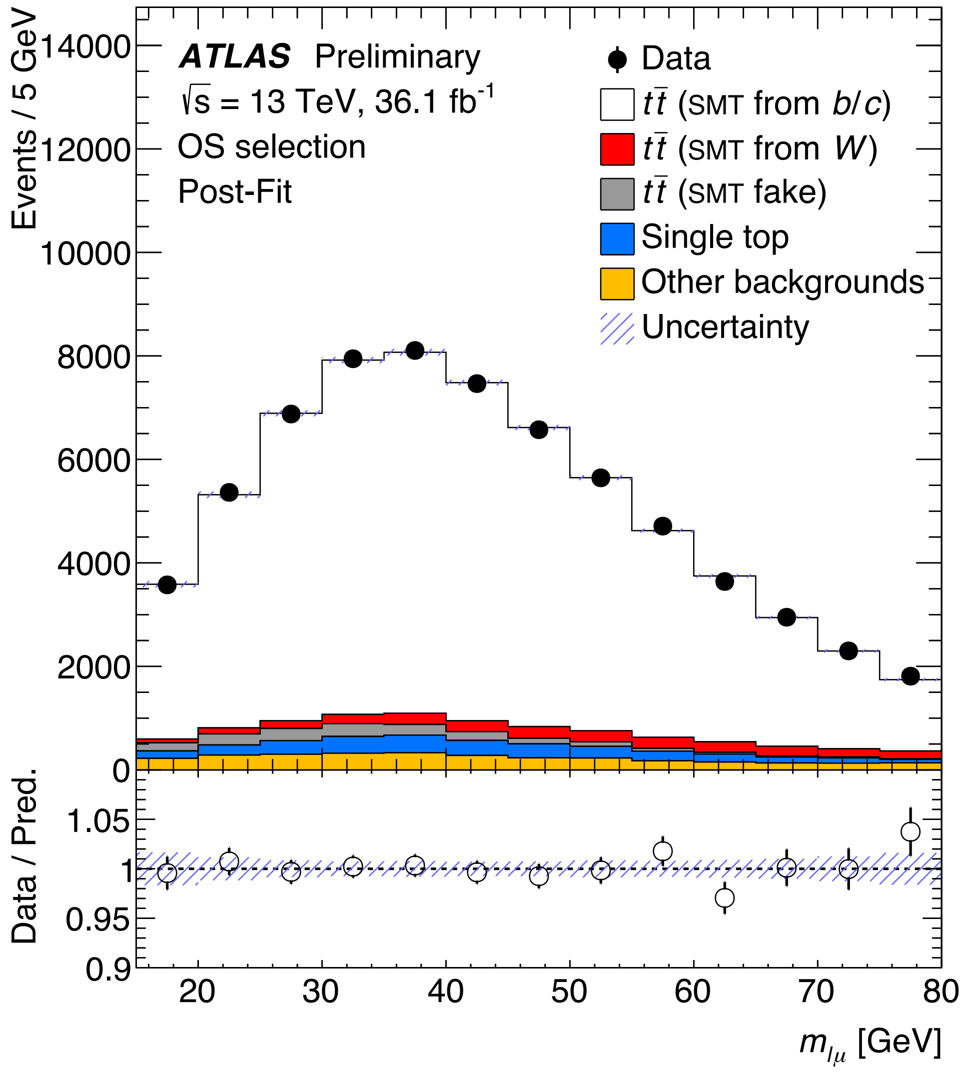


ATLAS-PHYS-PUB-2014-008

some tension between $e^+e^- \rightarrow bb$ measurements of b-fragmentation

parton-shower generators are also *not in good agreement*

motivation II

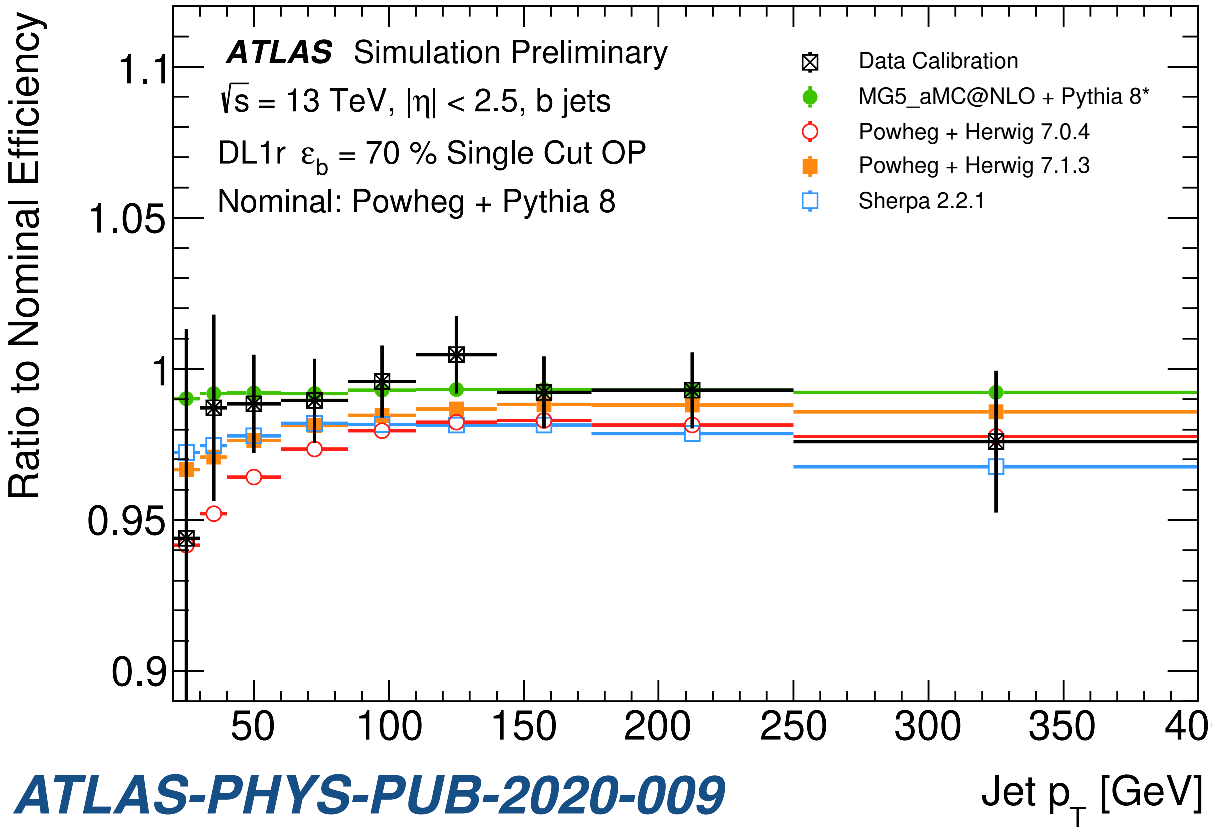


ATLAS-CONF-2019-046

$b \rightarrow \mu$ momentum transfer key for new top mass measurements

critical for delivering the best LHC physics results with b-jets.

b-tagging efficiency and b-jet response are very sensitive to fragmentation.



ATLAS-PHYS-PUB-2020-009

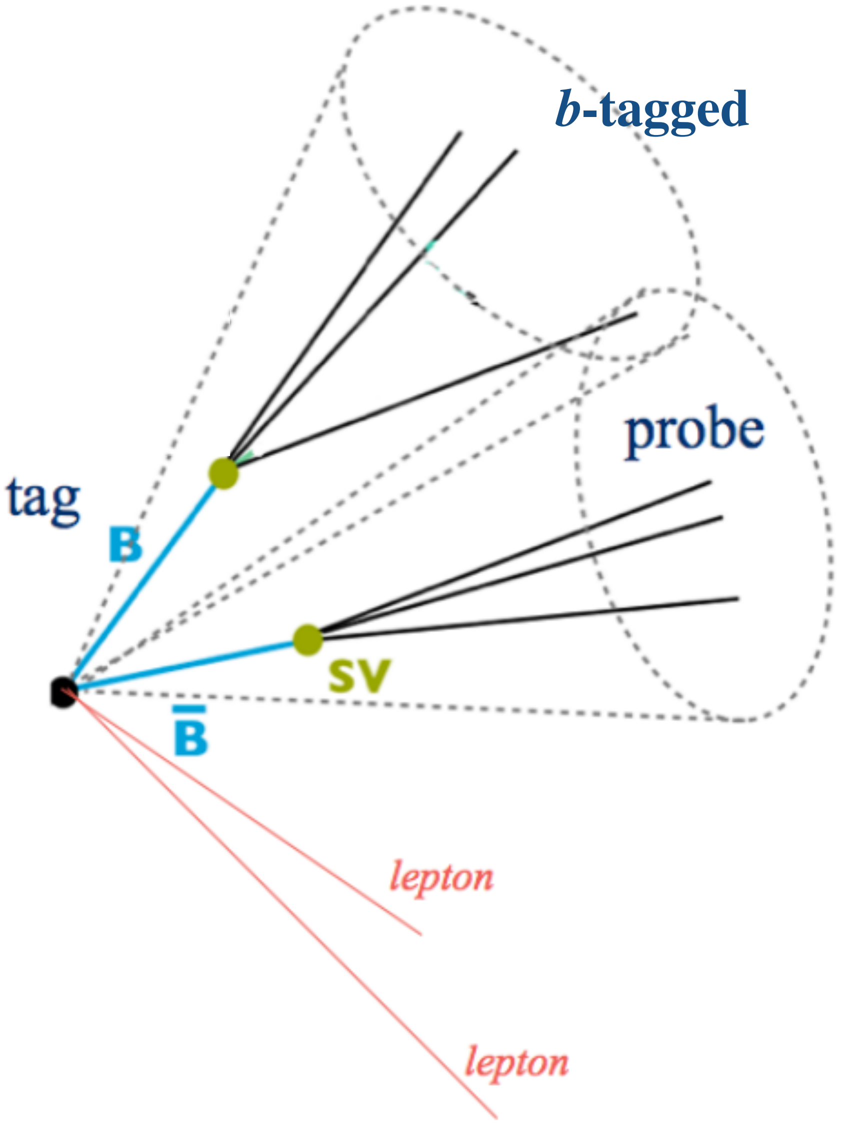
analysis strategy

1. obtain sample of b-jets from $t\bar{t} \rightarrow e\mu\nu\nu bb$ events
2. identify b-hadron candidates using ATLAS single secondary vertex finder (SSVF)
3. measure properties of the b-hadron sensitive to the modeling of b-quark fragmentation

overview: detector-level event selection

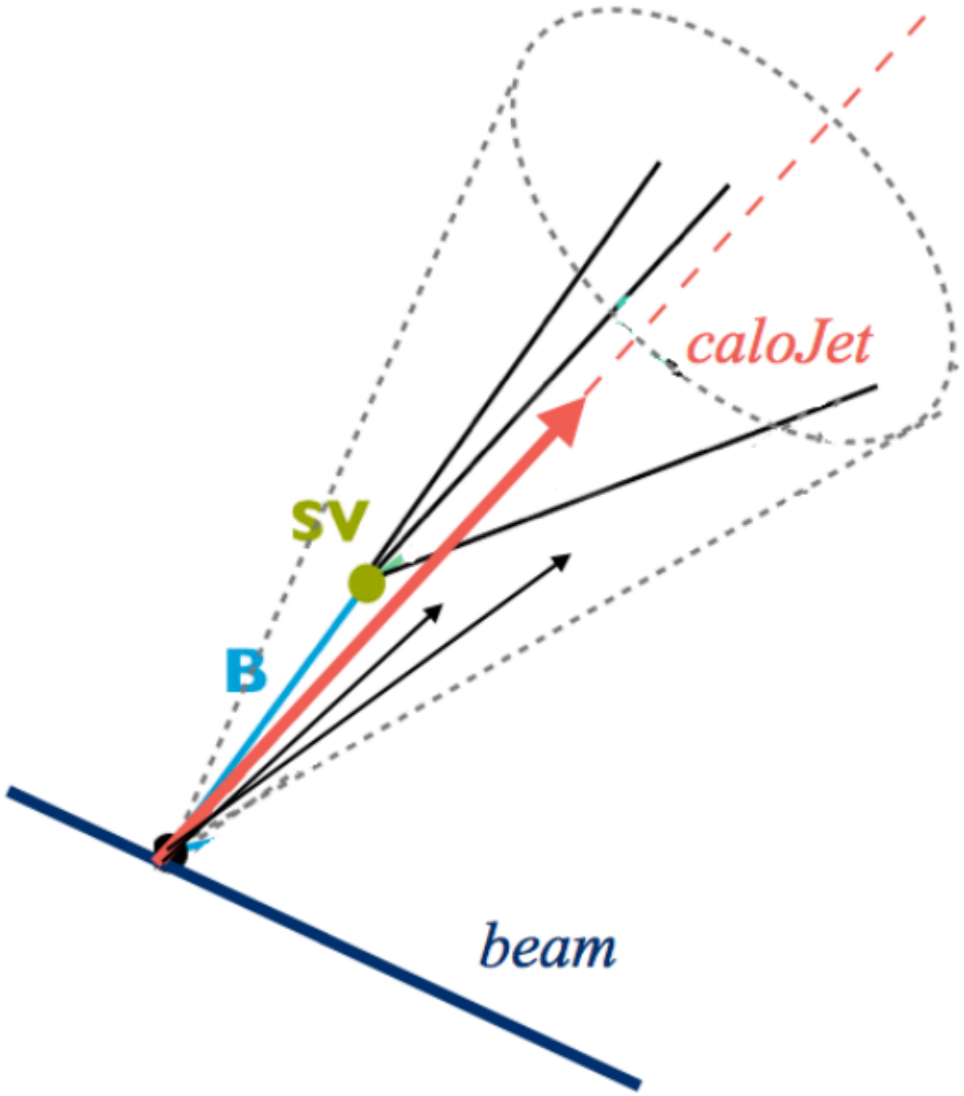
obtain sample of b-jets from $t\bar{t} \rightarrow e\mu\nu\nu bb$ events

- require ==1 e and ==1 μ with opposite charge
- ==2 anti-k_t $R = 0.4$ calorimeter jets
  - $|\eta| < 2.5$, $p_T > 30$ GeV
  - at least **one jet must be *b*-tagged** ("the tag")
  - we measure the other jet ("the probe")
  - if both jets are b-tagged, **both jets are measured**



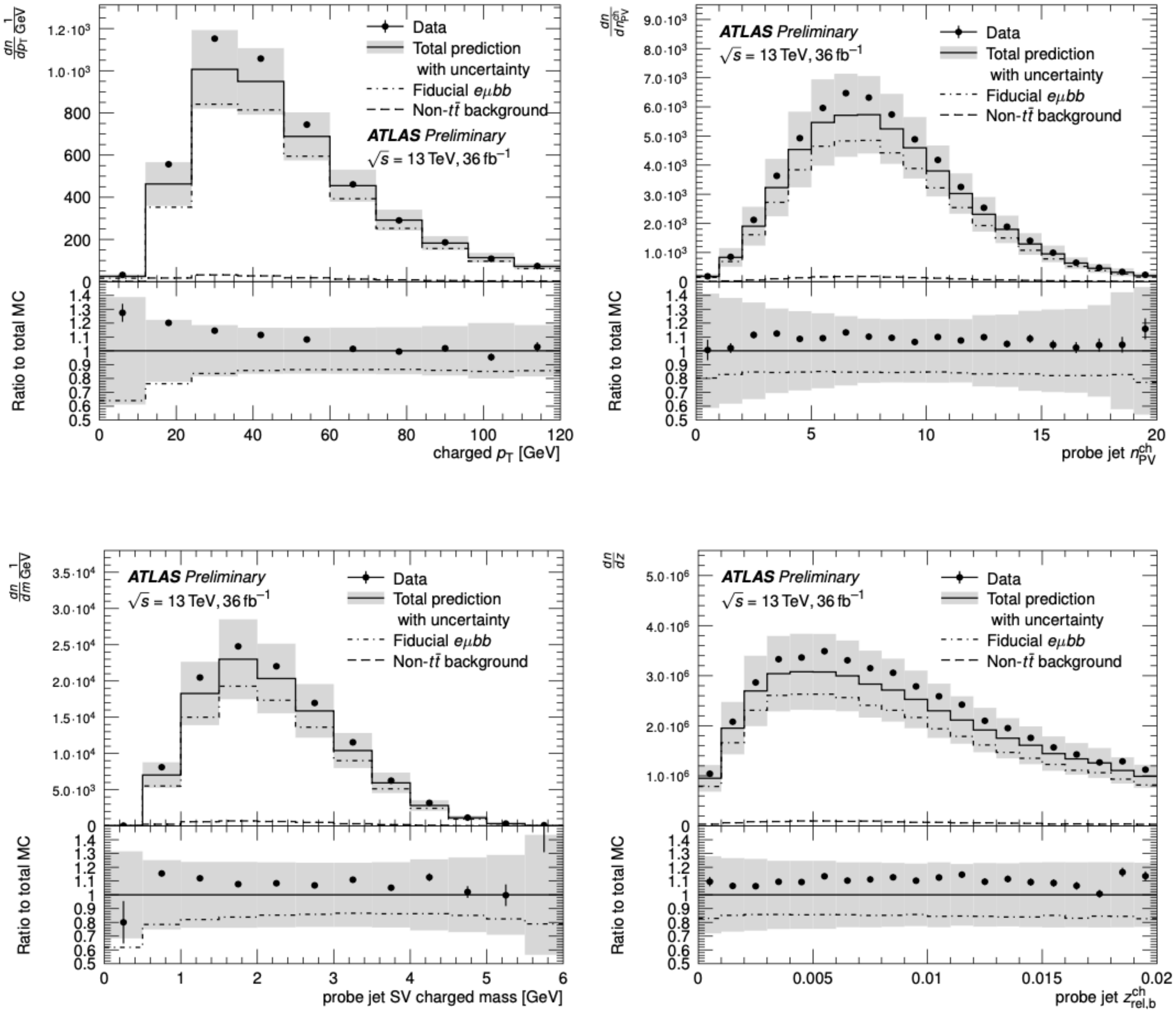
b-hadron reconstruction

identify b-hadron candidates using ATLAS's single secondary vertex finder (SSVF)



- we require a reconstructed SSVF vertex with ≥ 3 tracks for "probe" jets
- this is *nearly* a "second *b*-tag" and significantly reduces non-b backgrounds.
  - **95+% b-jet purity**
- the SSVF vertex **only has tracking information**
- track constituents of the SV make up the b-hadron candidate.
- tracks ghost-associated to the jet make up the jet charged component.

detector-level control plots



charged component of jet p_T has similar slope to jet p_T (data vs MC taken as uncertainty)

other relevant PV and SV quantities are well-modeled

SV mass and relative momentum transverse to to jet axis,

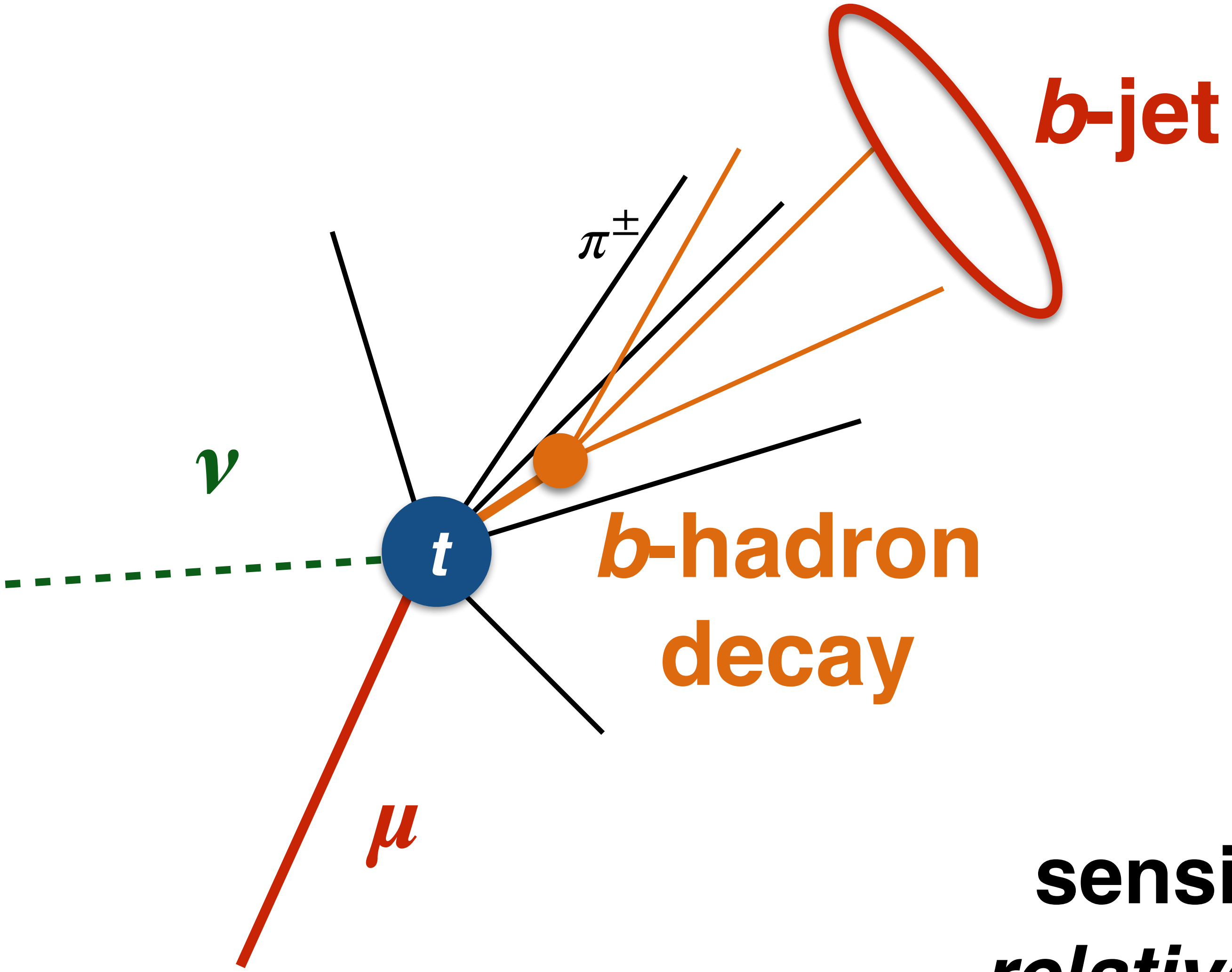
$$z^{\rm ch}_{{\rm rel},b} = (\vec{p}_b^{\,\rm ch} \times \vec{p}_{\rm jet}^{\,\rm ch})/|\vec{p}_{\rm jet}^{\,\rm ch}|^2$$

are also shown.

fiducial definition

- $t\bar{t}$ events
  - single top is treated as background for ease of interpretation
- == 1 μ, == 1 e with dressed $p_T > 25$ GeV, $|\eta| < 2.5$
- == 2 b-jets with $|\eta| < 2.1$, $p_T > 30$ GeV, $\Delta R(jet, jet) > 0.5$
- b-jet defined as
  - anti-k_t $R = 0.4$ jet of stable particles
    - prompt charged leptons and all neutrinos are excluded
  - ==1 weakly-decaying b-hadron ($p_T > 5$ GeV) within ΔR(jet, hadron) < 0.3
- **stable charged particles with $p_T > 500$ MeV used to calculate observables**
  - stable charged b-hadron decay products make up $\vec{p}_b^{\,\mathrm{ch}}$
  - stable charged particles ghost-associated to jet make up $\vec{p}_{\mathrm{jet}}^{\,\mathrm{ch}}$

observables I



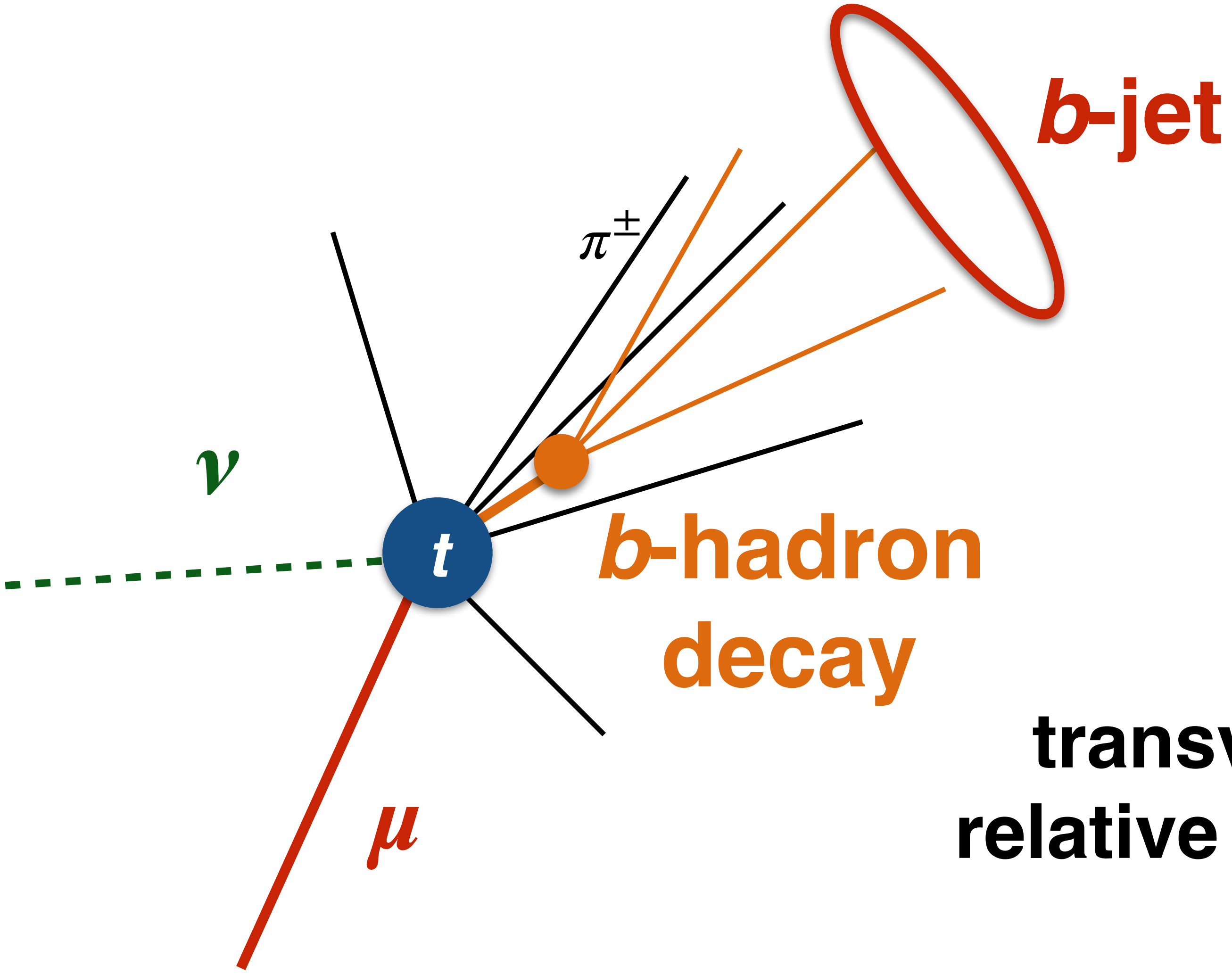
$$z^{\mathrm{ch}}_{\mathrm{L,b}} = \frac{\vec{p}^{\,\mathrm{ch}}_{\mathrm{b}} \cdot \vec{p}^{\,\mathrm{ch}}_{\mathrm{jet}}}{|p^{\mathrm{ch}}_{\mathrm{jet}}|^2},$$

$$z^{\mathrm{ch}}_{\mathrm{T,b}} = \frac{p^{\mathrm{ch}}_{\mathrm{T,b}}}{p^{\mathrm{ch}}_{\mathrm{T,jet}}},$$

sensitive to momentum of b-hadron
relative to the nearby hadronic activity

seeded by the particle jet direction, but **only charged particles are used to calculate fiducial observables**

observables II



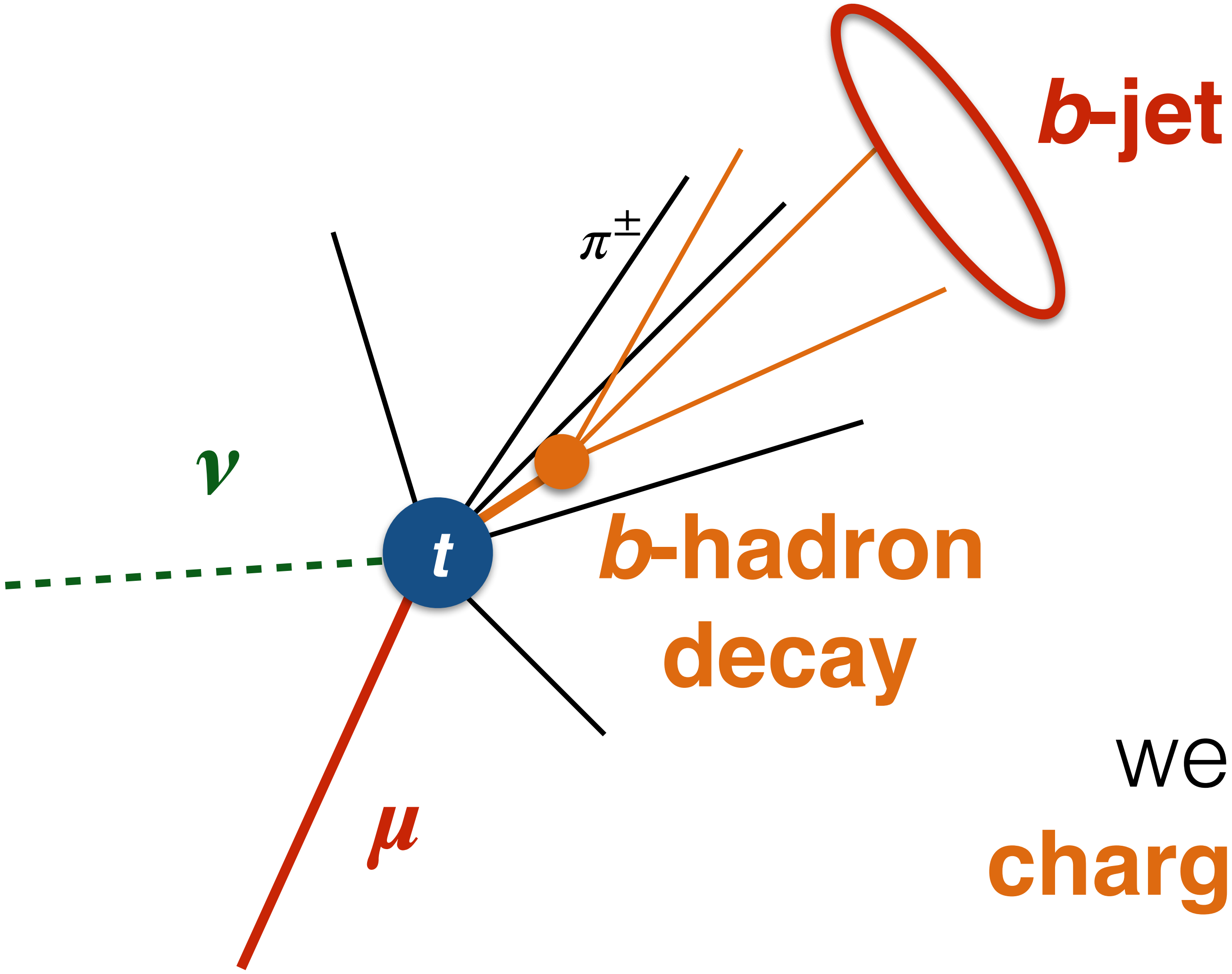
$$\rho = \frac{2p_{\mathrm{T,b}}^{\mathrm{ch}}}{p_{\mathrm{T}}^{e} + p_{\mathrm{T}}^{\mu}}$$

transverse momentum of *b*-hadron relative to the average pt of the leptons in the event

sensitive to choice of e.g. $\alpha_S(FSR)$ but ***not*** $\alpha_S(ISR)$

sensitivity also to wide-angle radiation
in top-quark decays

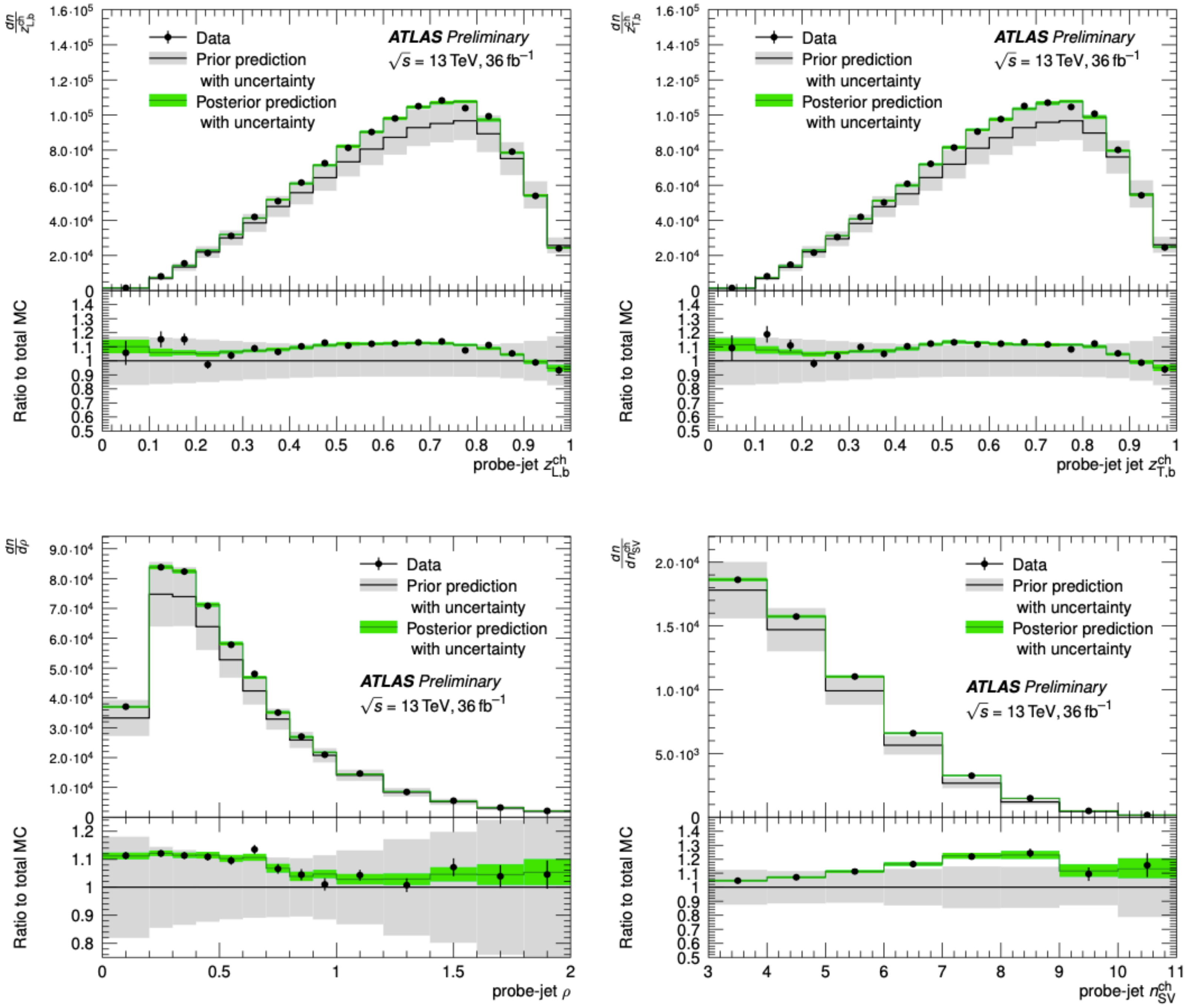
observables III



we also unfold the **number of charged, stable b-hadron children with $p_T > 500$ MeV**.

sensitive to the of b-hadron species production rates

detector-level data/MC comparison of observables



for now please ignore posterior prediction and green uncertainty bands (we'll return to this)

reasonable agreement observed, with some small systematic differences

unfolding (FBU)

- **based on the "Fully Bayesian Unfolding" technique (arxiv: 1201.4612)**
- we use the following likelihood to unfold the detector effects:

$$\mathcal{L}(d|\sigma,\Lambda) = \prod_{i \,\in\, \mathrm{recobins}} \mathrm{Poiss}(d_i|x_i(\sigma,\Lambda))$$

$$x_i(\sigma,\Lambda) = L(\Lambda)\times(b_i(\Lambda)+M_{ij}(\Lambda)\,\sigma_j)$$

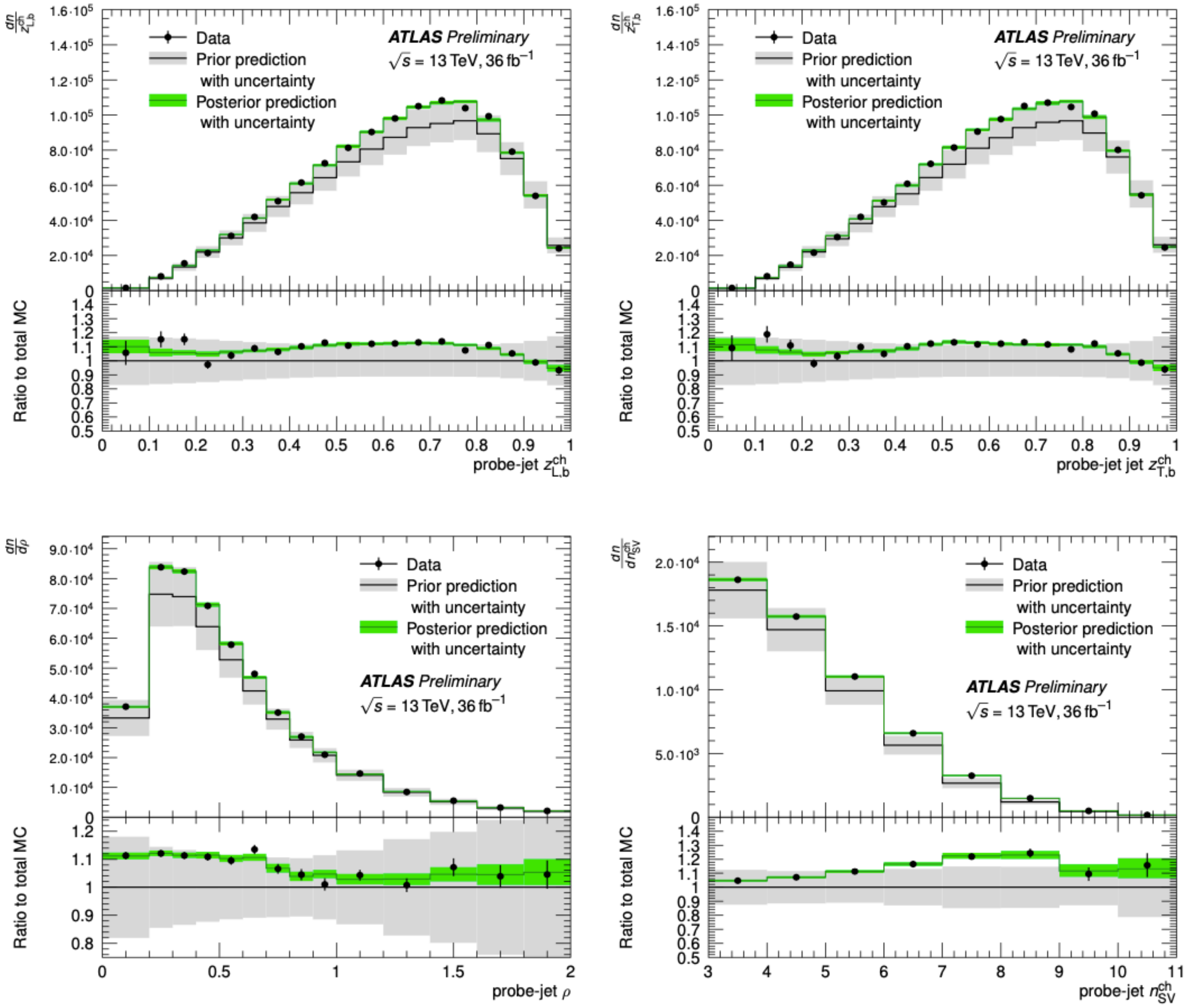
- where $\overrightarrow{x}$ is the total number of predicted events in detector-level bins, L is the luminosity, $\overrightarrow{b}$ are predicted background events, M_{ij} is probability that a particle-level jet in bin i is reconstructed in detector-level bin j, *and* Λ are the systematic nuisance parameters.
- **we marginalize over nuisance parameters through MCMC sampling to build the marginal posteriors on $\overrightarrow{\sigma}$, the parameters-of-interest.**

systematic uncertainties

the dominant uncertainties in the unfolding model come from

- track efficiency and impact-parameter resolutions
- $t\bar{t}$ modeling: non-fiducial background and detector response
  - Pythia vs Herwig
  - A14 α_S^{FSR} tune variations
  - jet charged p_T and n_{SV}^{trk} mismodeling
- pileup $\langle\mu\rangle$ distribution

detector-level data/MC comparison before/after unfolding



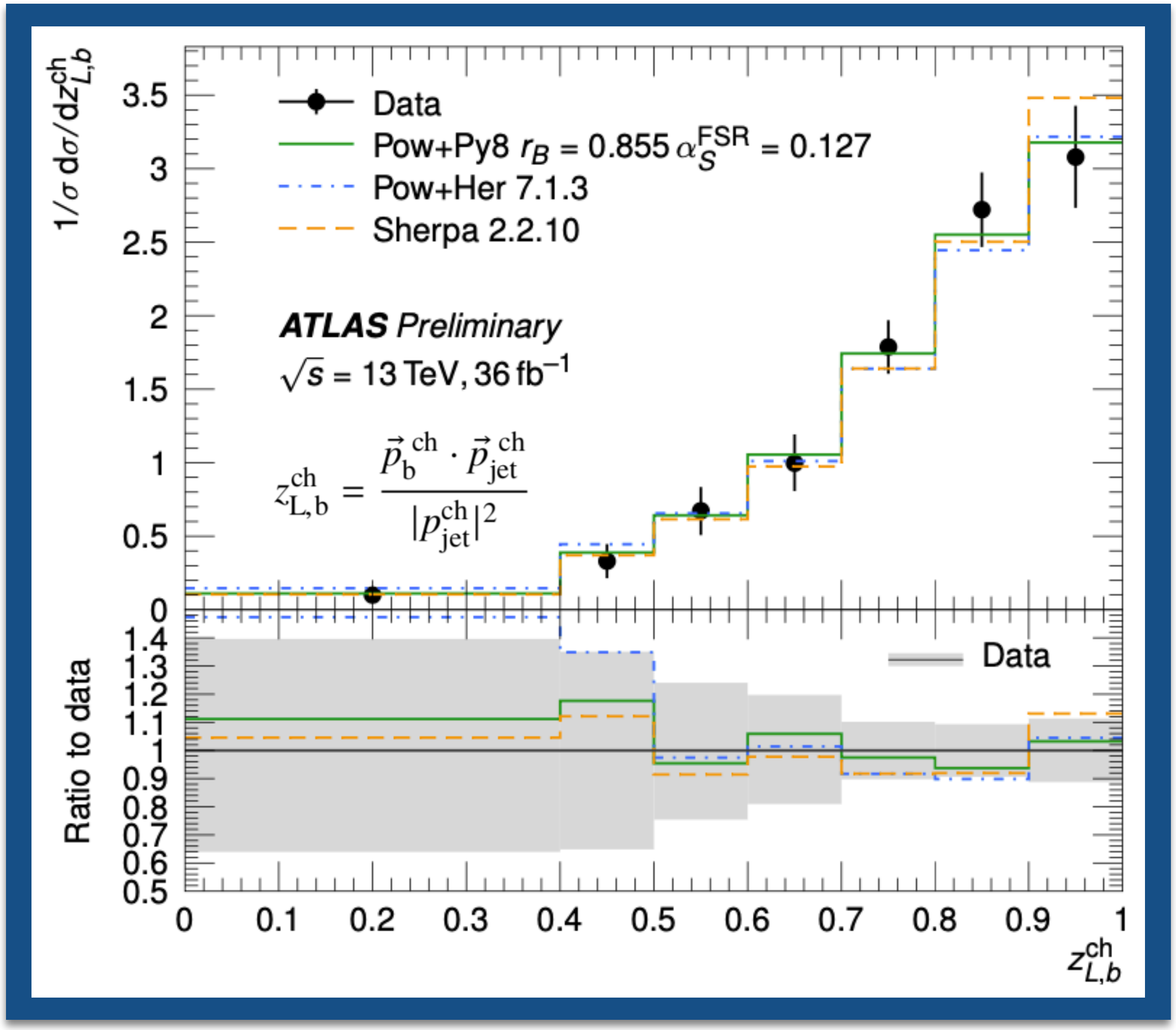
detector-level observables
before (prior) and
after (posterior) unfolding

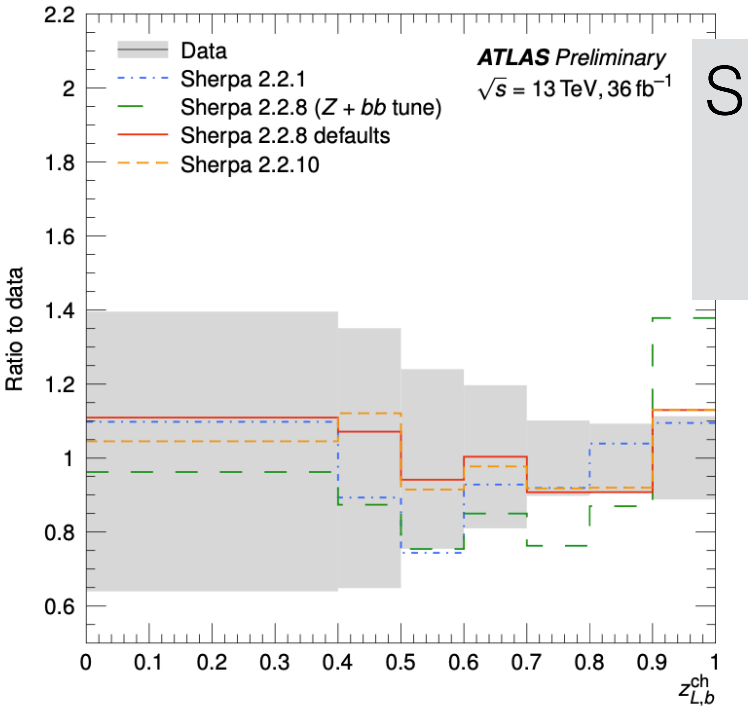
unfolding model is capable of describing observed data

parameter anti-correlations cause reduction in total uncertainty

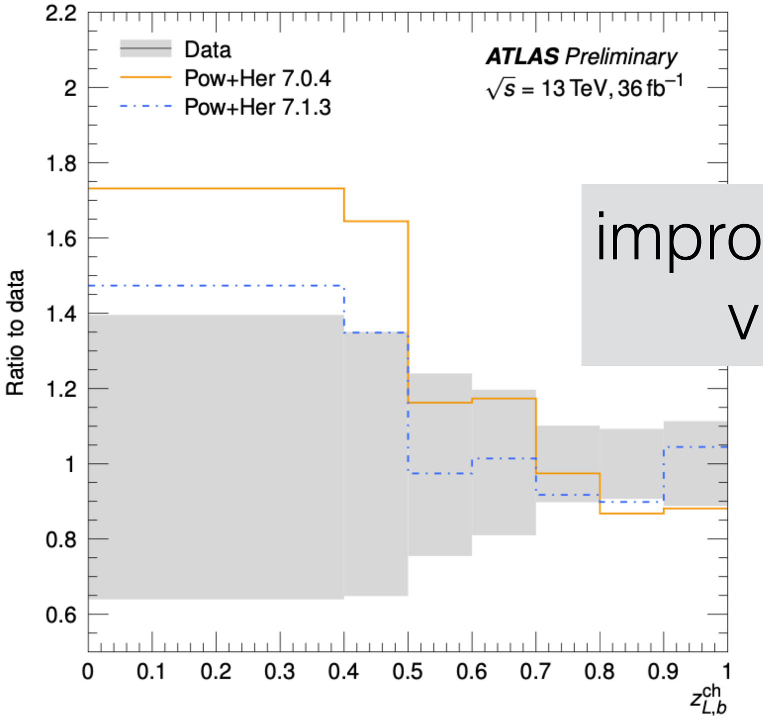
posterior uncertainty closely follows data stat uncertainty

fiducial data/MC comparisons I





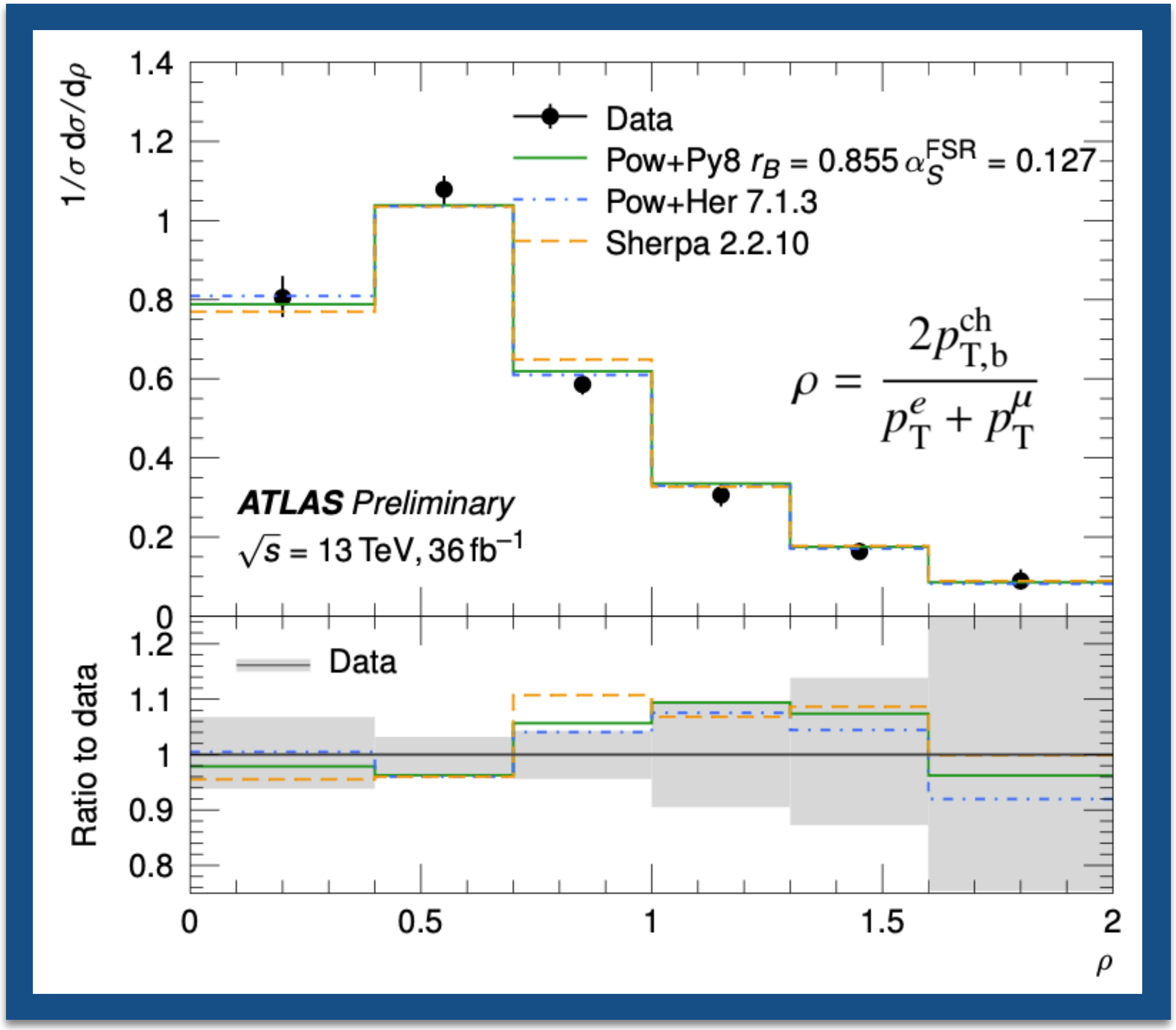
Sherpa: 2.2.10 with default tune looks good.



improvements in newer versions of H7

recent generator tunes to $e^+e^- \to Z \to bb$ do a reasonable job.

fiducial data/MC comparisons II



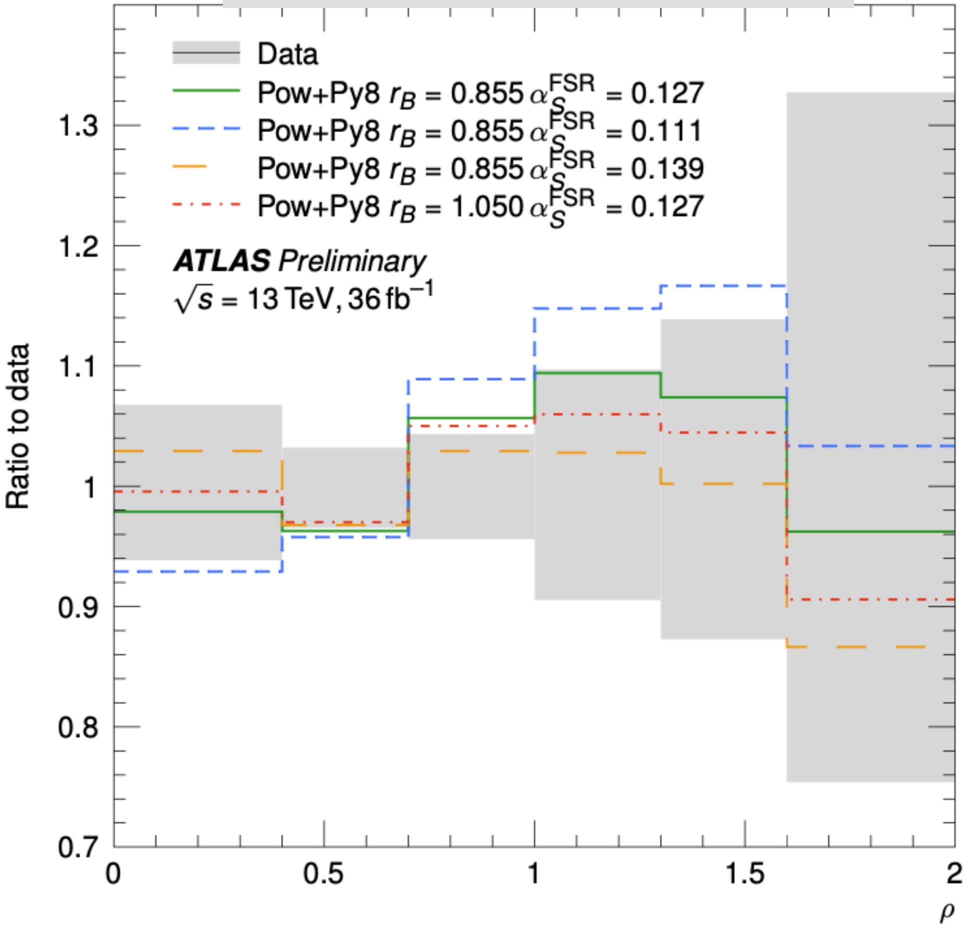
$\sim 1\sigma$ tensions

with recent generators

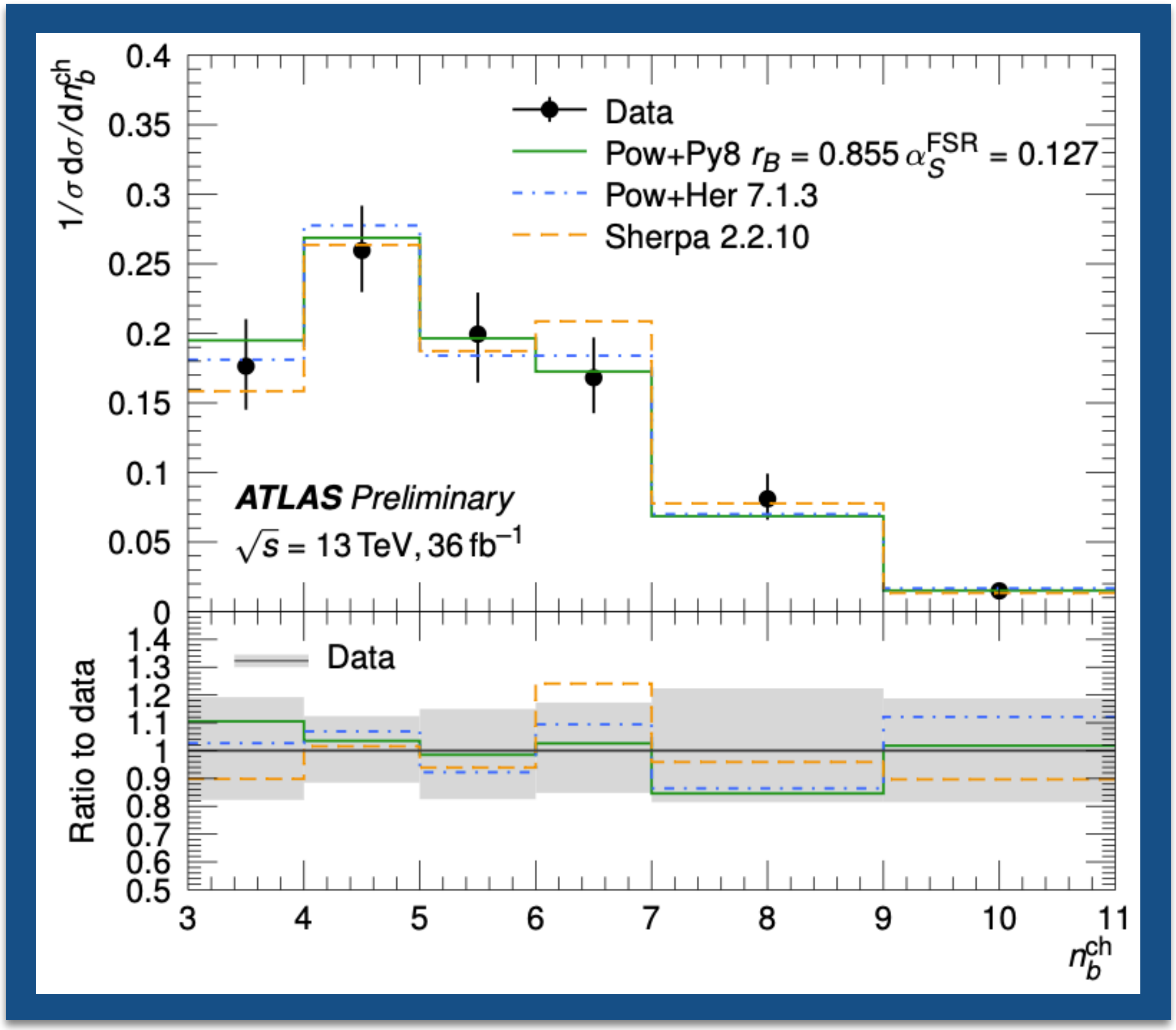
→ further investigation.

large impact from A14 α_S^{FSR} variations

no sensitivity to α_S^{ISR}

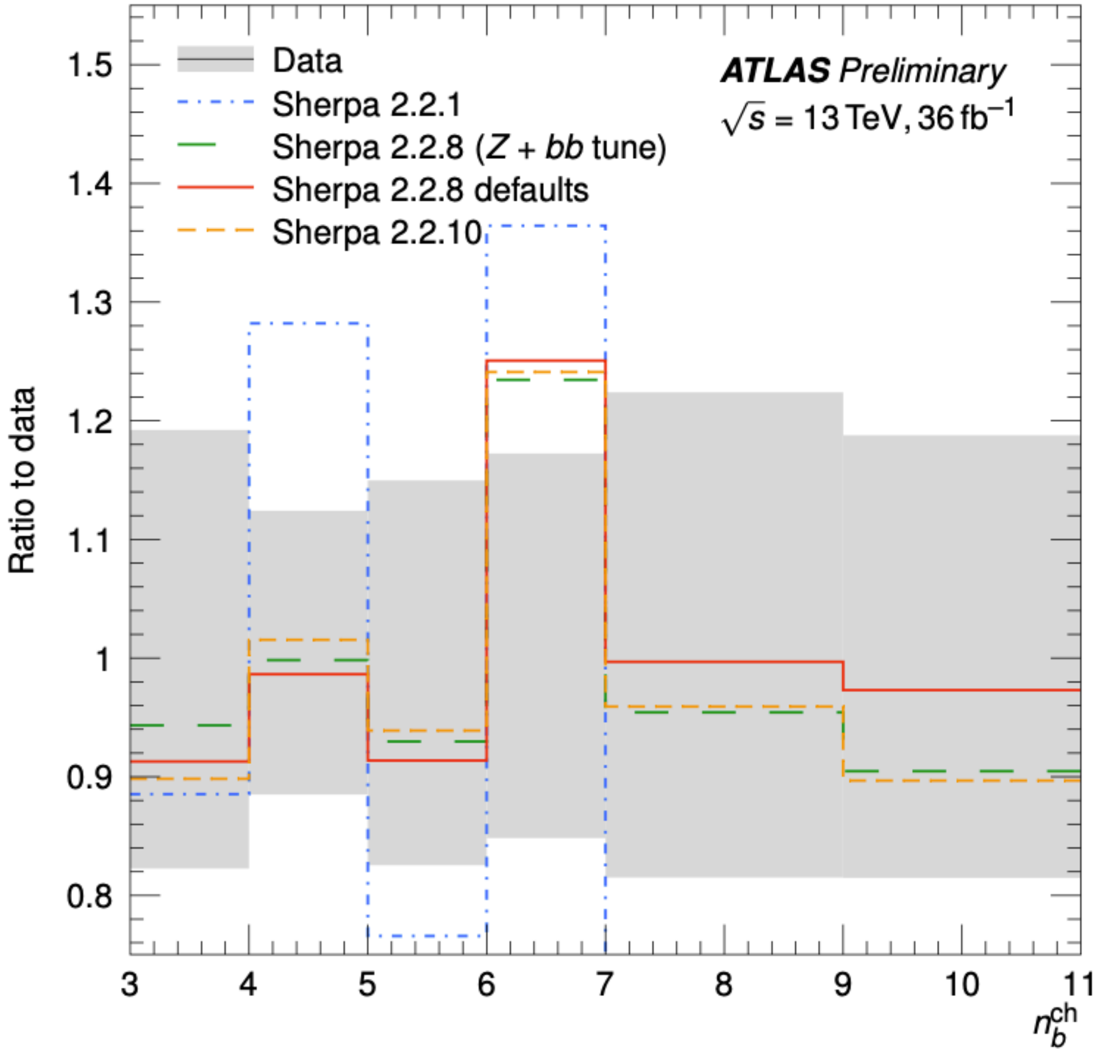


fiducial data/MC comparisons III



poor choice of heavy baryon enhancement in ATLAS's Sherpa 2.2.1 setup has now been fixed.

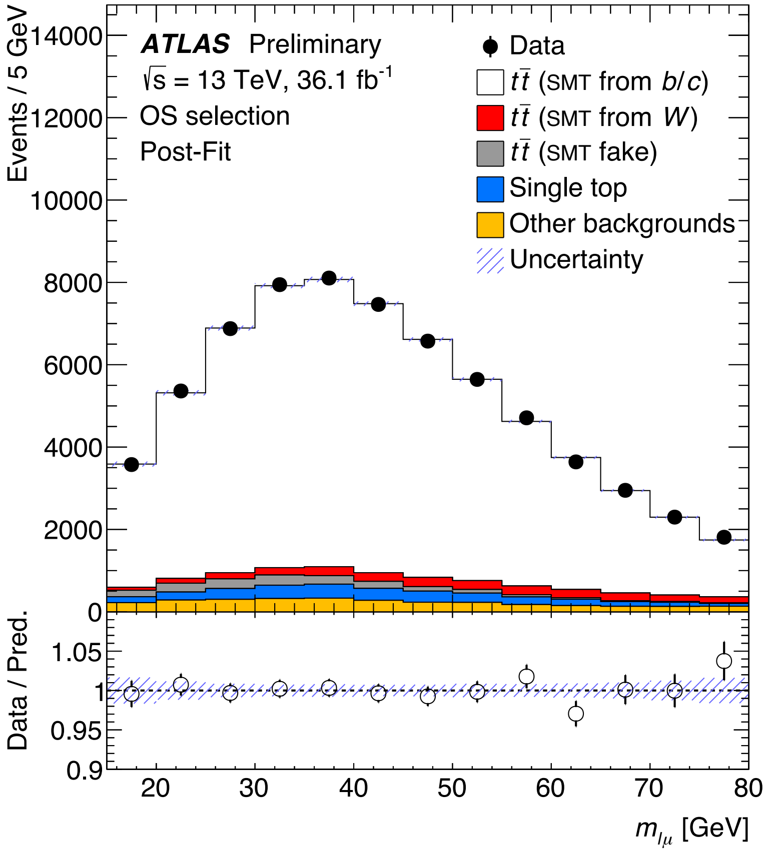
Py8/Her7 + EvtGen and Sherpa 2.2.10 all perform admirably.



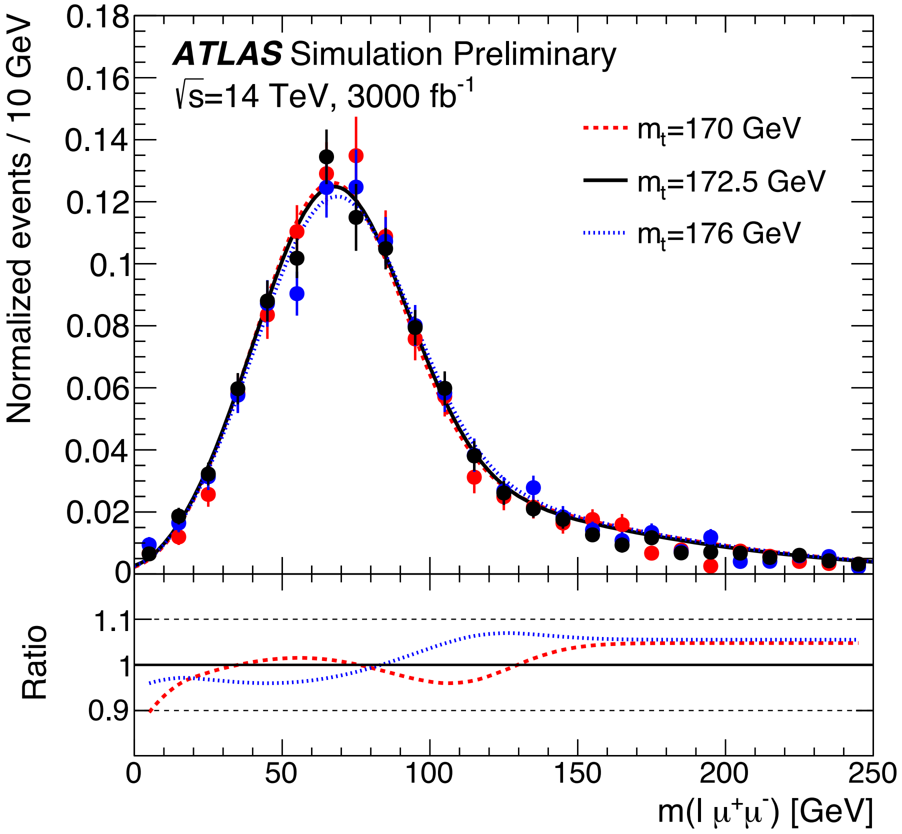
impact: top mass

at our current precision, **generators *well-tuned* to $e^+e^- \to bb$ data yield good predictions in $t\bar{t}$ events and top-quark decays**.

- we now have unfolded data to guide our choice of related uncertainties.
- precise b-fragmentation → smaller uncertainties on the b-quark to b-jet transfer
- a better understanding of $t \to b^{\text{quark}} \to b^{\text{hadron}}$ crucial for fully-leptonic template mass extractions ($m_{\ell\mu}$ or $m_{\ell,J/\psi}$)
  - joint fragmentation+mass measurements in leptonic modes **should be investigated**



ATLAS-CONF-2019-046

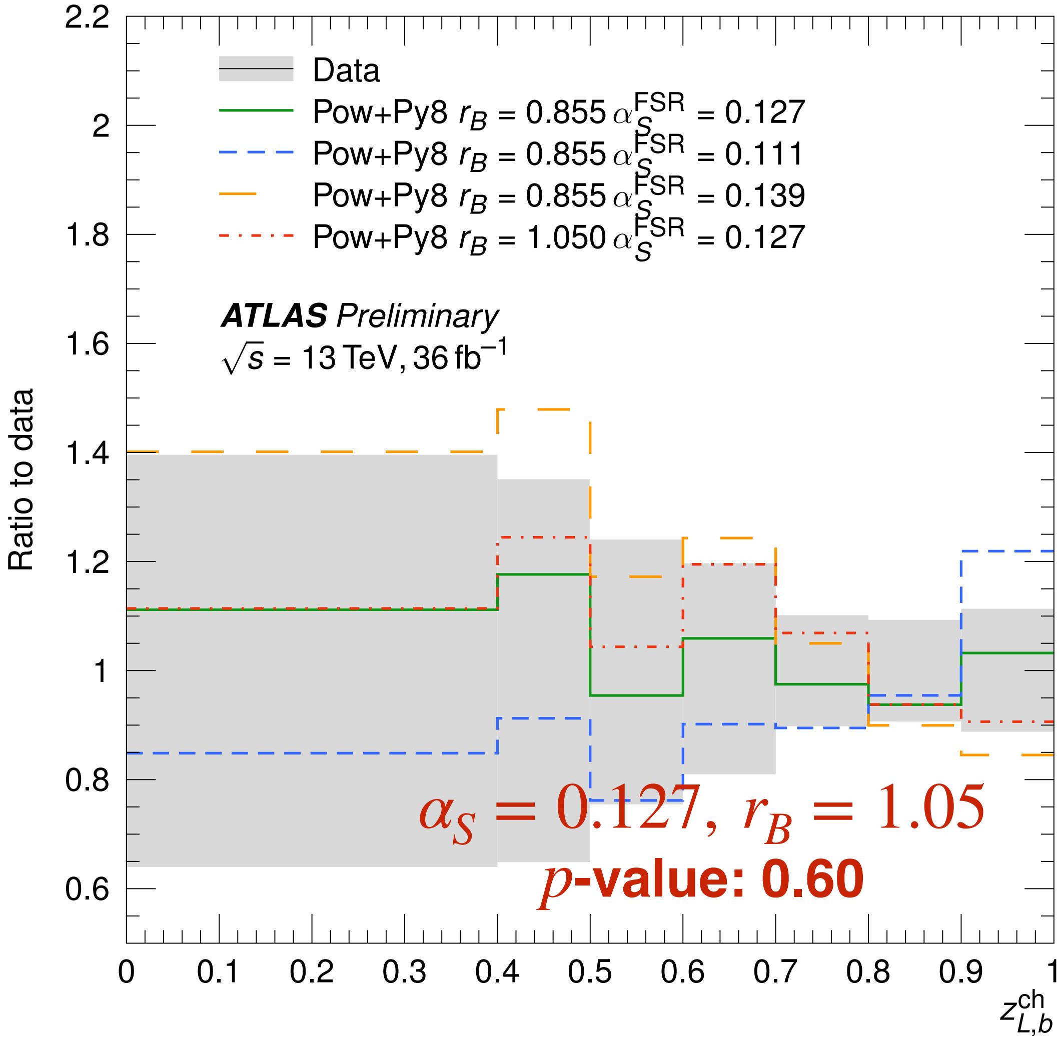


ATLAS-CONF-2018-042

impact: top mass

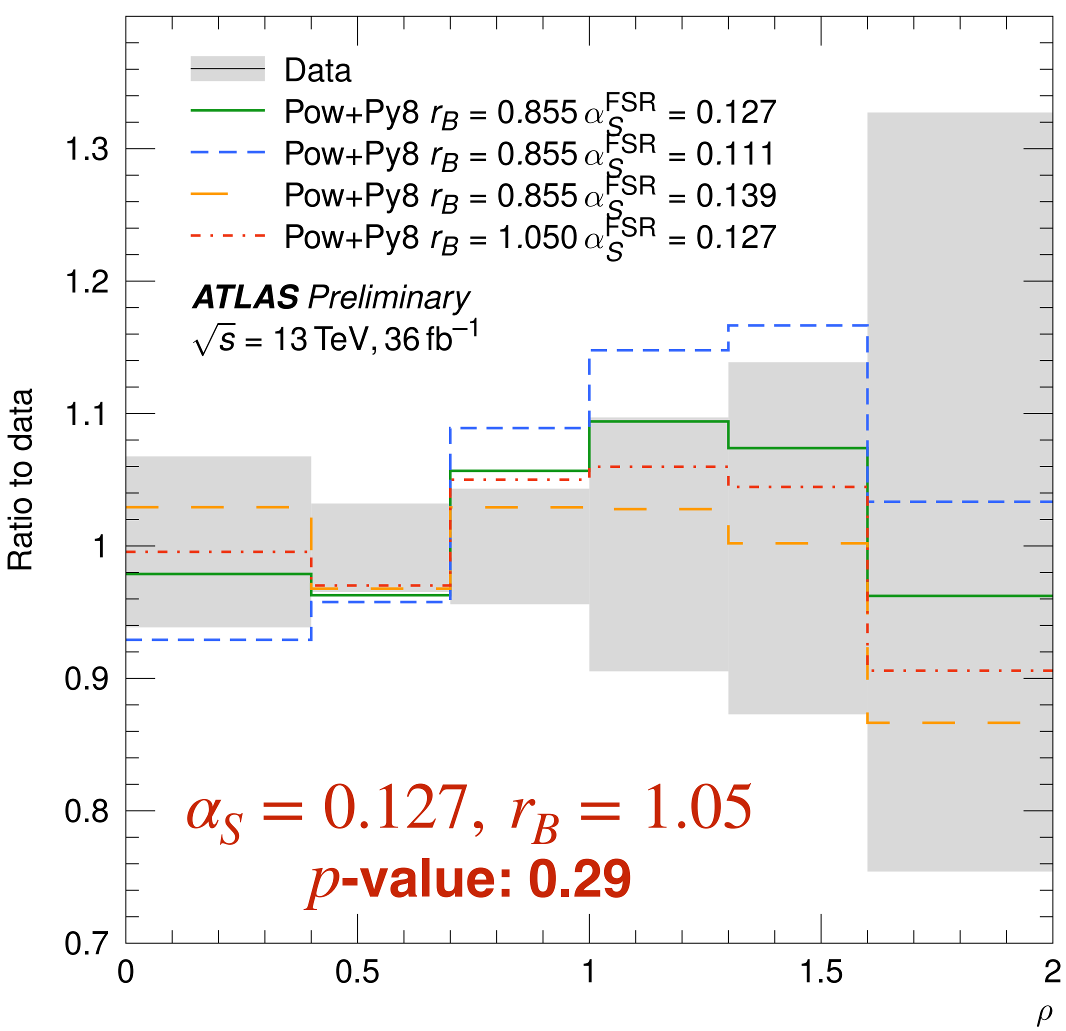
ATLAS retuned the Lund-Bowler r_B parameter to LEP data with the A14 α_S

- used in preliminary mass measurement ATLAS-CONF-2020-046
- this combination of $\alpha_S = 0.127,\ r_B = 1.05$ agrees with observed data to current precision. **(dash-dotted red curve)**



$\alpha_S = 0.127,\ r_B = 1.05$

p-value: 0.60

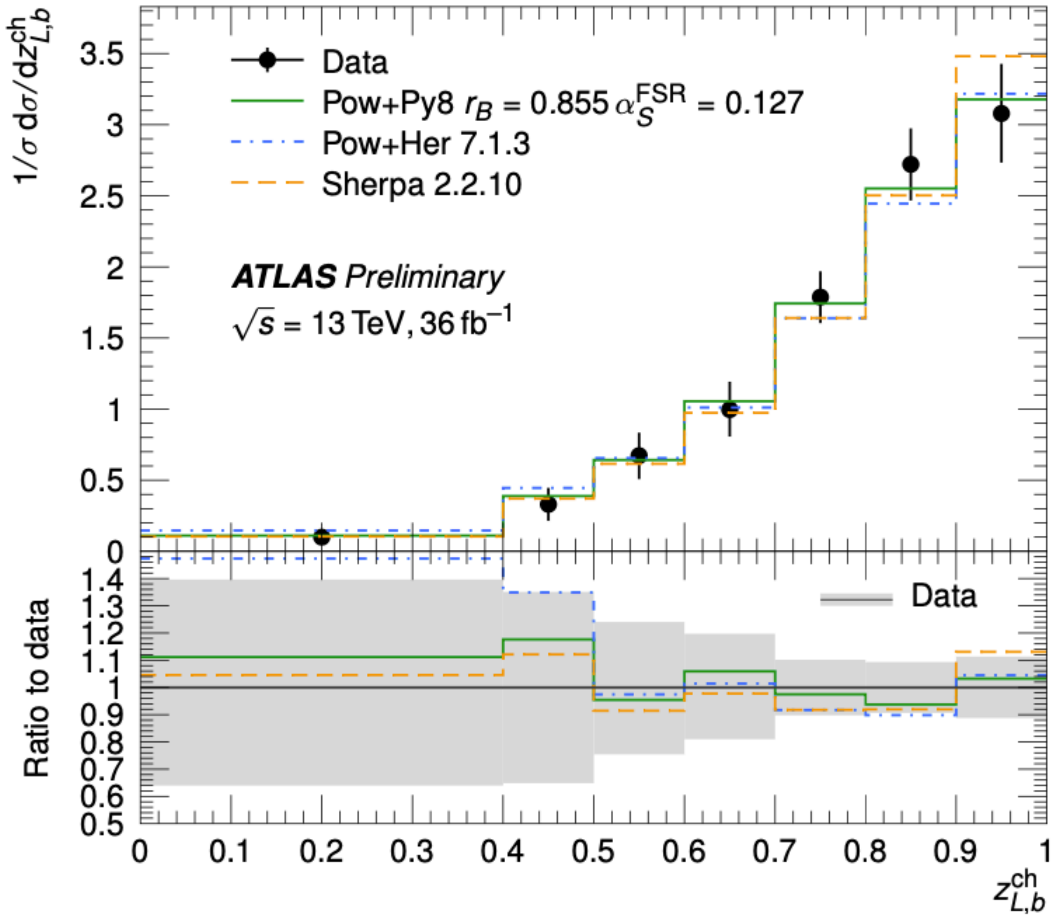


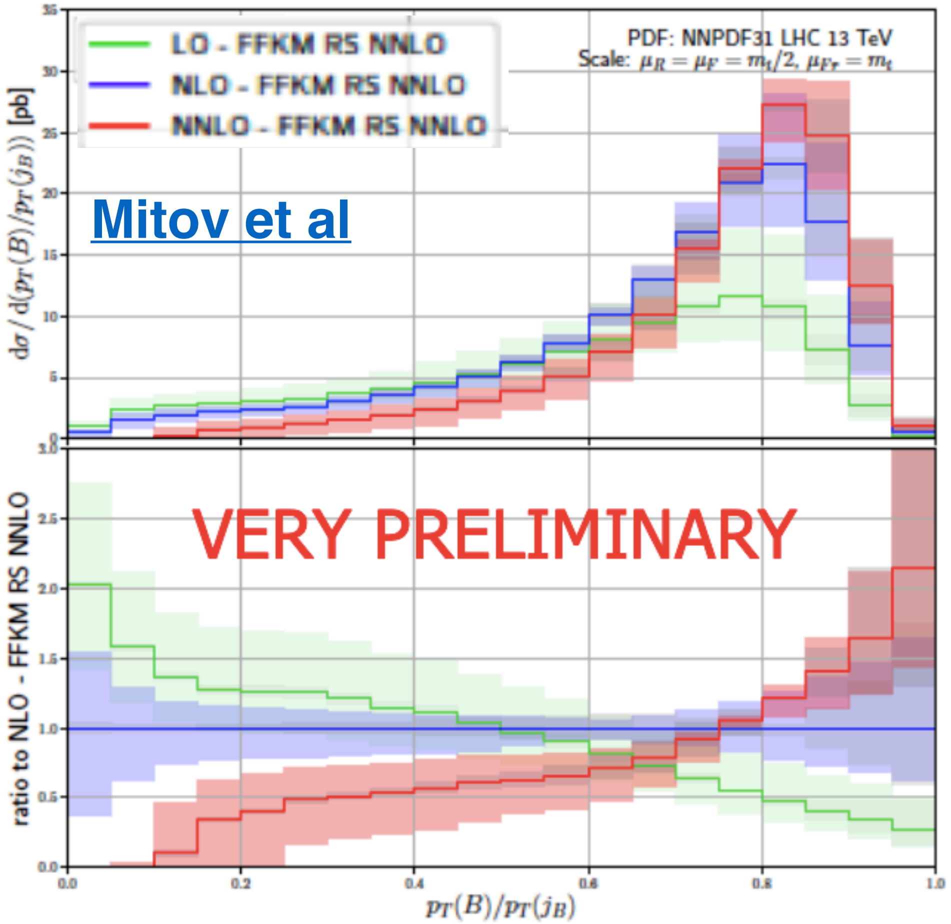
$\alpha_S = 0.127,\ r_B = 1.05$

p-value: 0.29

status of ~analytic calculations

- Mitov et al. showed the first quasi-analytic calculations of similar distributions at TOP2020.
- will be very interesting to compare to unfolded data
- challenges for comparison:
  - calculations are for production of exclusive B-hadron species
  - jets are defined at parton-level in calculation
  - ATLAS measurement is to stable, charged particles



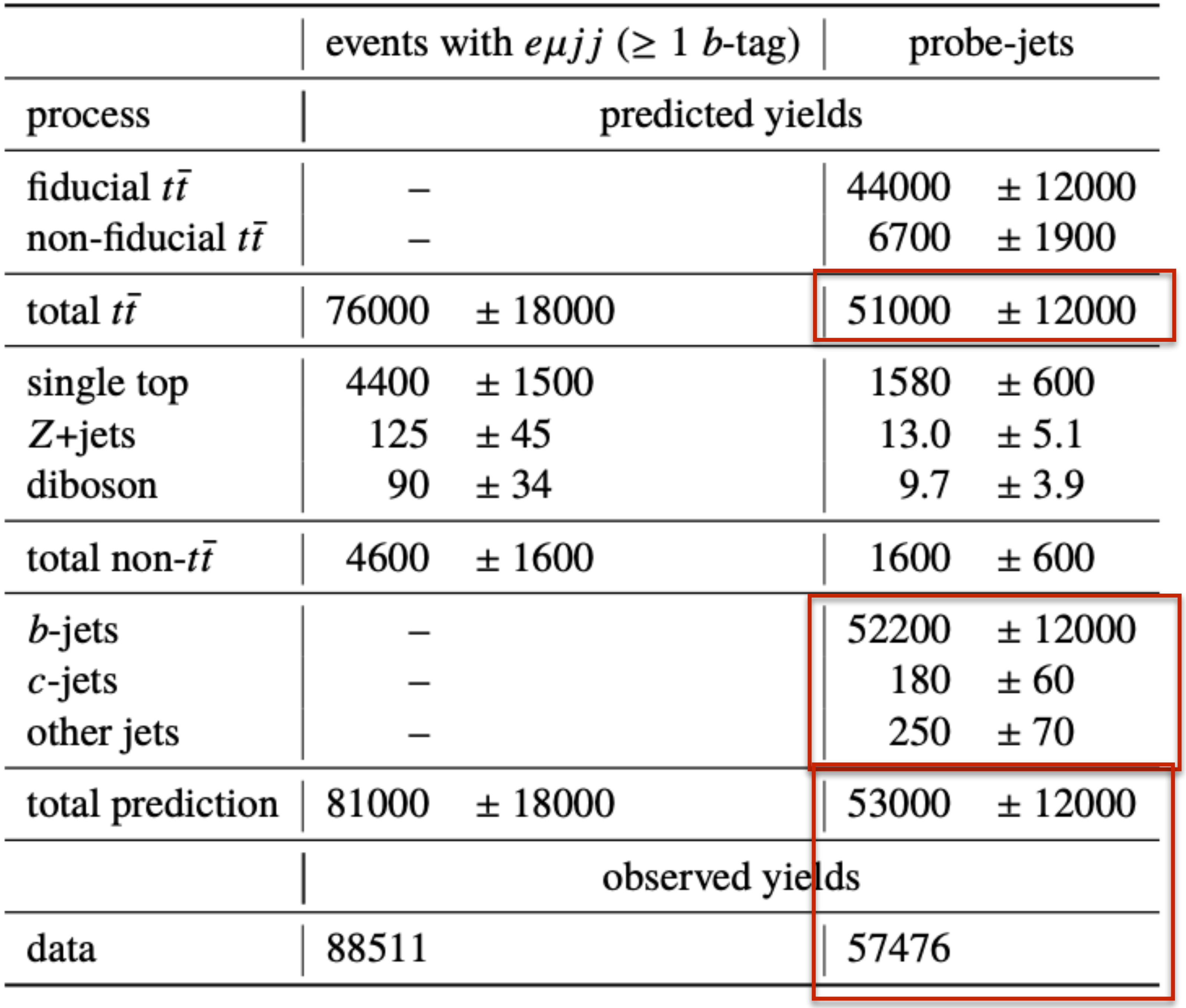


summary and outlook

- several differential observables sensitive to b-quark fragmentation have been measured in $tt \rightarrow e\mu bb$ events with the 2015+2016 ATLAS data
- comparisons to modern MC generator predictions are made
- we do not explicitly tune MC parameters in this paper
    - but these measurements will be used in future tunes.
- we will provide Rivet routine + HEPData entries for the unfolded data.
- we've identified several "sub-optimal" tunes that were used previously: hopefully we can prevent that in the future.
- many improvements possible: more precise measurements may be necessary for best "direct" top-quark mass extraction
- first ~analytic calculations of b-quark fragmentation in top-quark decays are on the horizon for comparison!

bonus

detector-level yields: predicted and observed



	events with $e\mu jj$ (≥ 1 b-tag)	probe-jets
process	predicted yields	
fiducial $t\bar{t}$	–	44000 ± 12000
non-fiducial $t\bar{t}$	–	6700 ± 1900
total $t\bar{t}$	76000 ± 18000	51000 ± 12000
single top	4400 ± 1500	1580 ± 600
Z+jets	125 ± 45	13.0 ± 5.1
diboson	90 ± 34	9.7 ± 3.9
total non-$t\bar{t}$	4600 ± 1600	1600 ± 600
b-jets	–	52200 ± 12000
c-jets	–	180 ± 60
other jets	–	250 ± 70
total prediction	81000 ± 18000	53000 ± 12000
	observed yields	
data	88511	57476

Table 1: Summary of expected and observed event and probe-jet yields.

tt dominant

high expected *b*-jet purity

7.5% data / MC discrepancy

unfolding procedure: background and response matrices

- to calculate the background predictions and response matrices at a given point in parameter space, we take linear combinations of the background predictions for nominal and the 1-sigma variation corresponding to a systematic uncertainty:

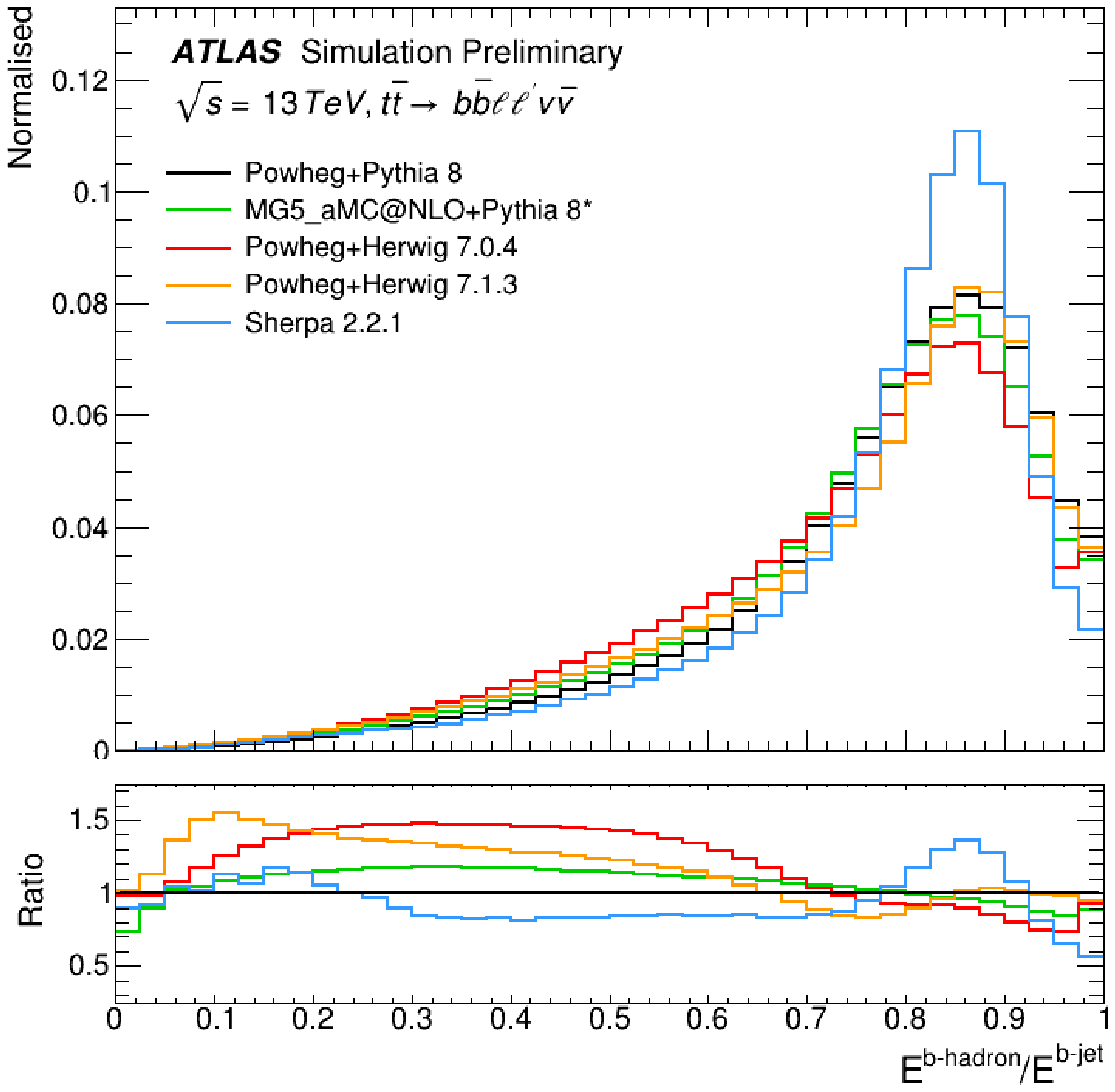
$$M_{ij}^{\mathrm{p}\to\mathrm{d}}(\vec{p}) = M_{ij,0}^{\mathrm{p}\to\mathrm{d}} + \sum_{k\,\in\mathrm{systematics}} p_k * \left(M_{ij,k}^{\mathrm{p}\to\mathrm{d}} - M_{ij,0}^{\mathrm{p}\to\mathrm{d}}\right)$$

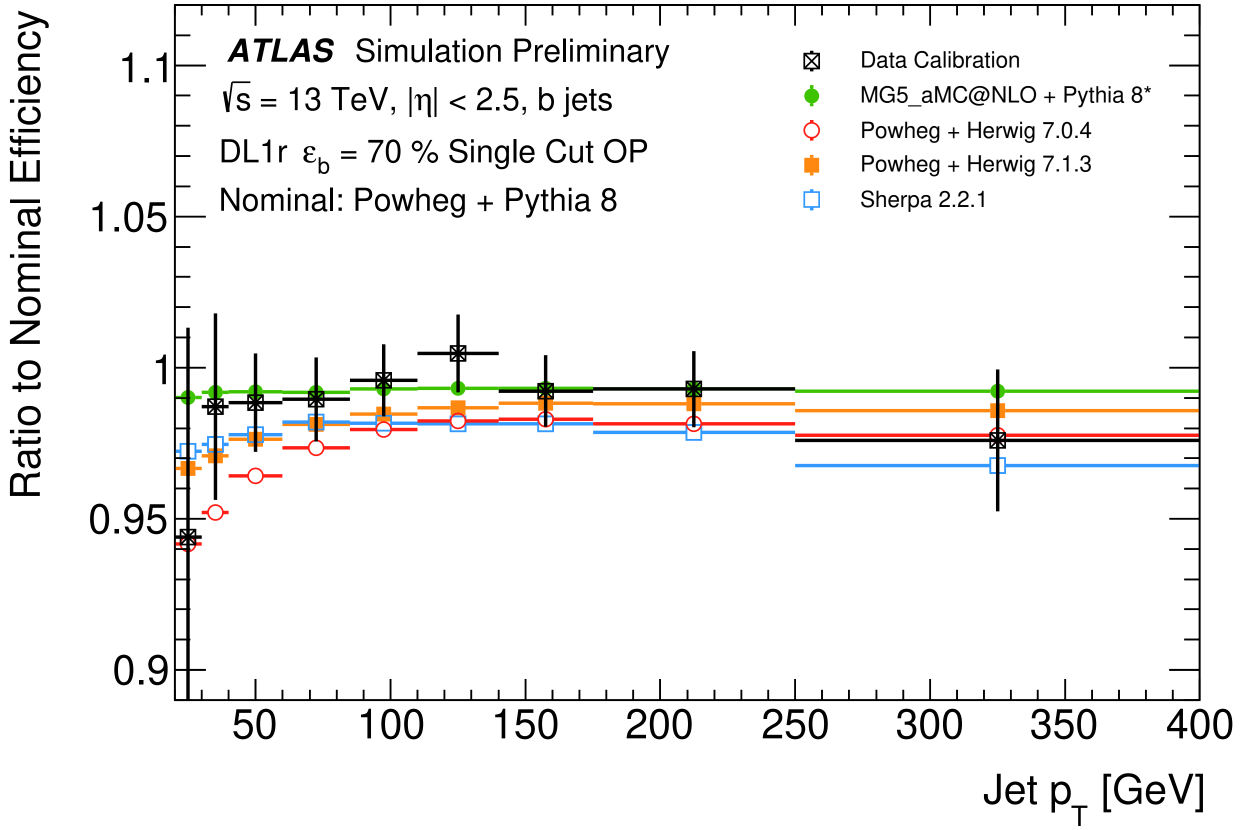
$$b_i(\vec{p}) = b_{i,0} + \sum_{k\,\in\mathrm{systematics}} p_k * \left(b_{i,k} - b_{i,0}\right)$$

where the $\overrightarrow{p}$ are the nuisance parameters at a particular point in parameter space.

- in this way we can coherently and smoothly vary the response matrices and background predictions corresponding to systematic variations.
- we're wandering around the space of possible response matrices and background predictions available given our understanding of the predictions.

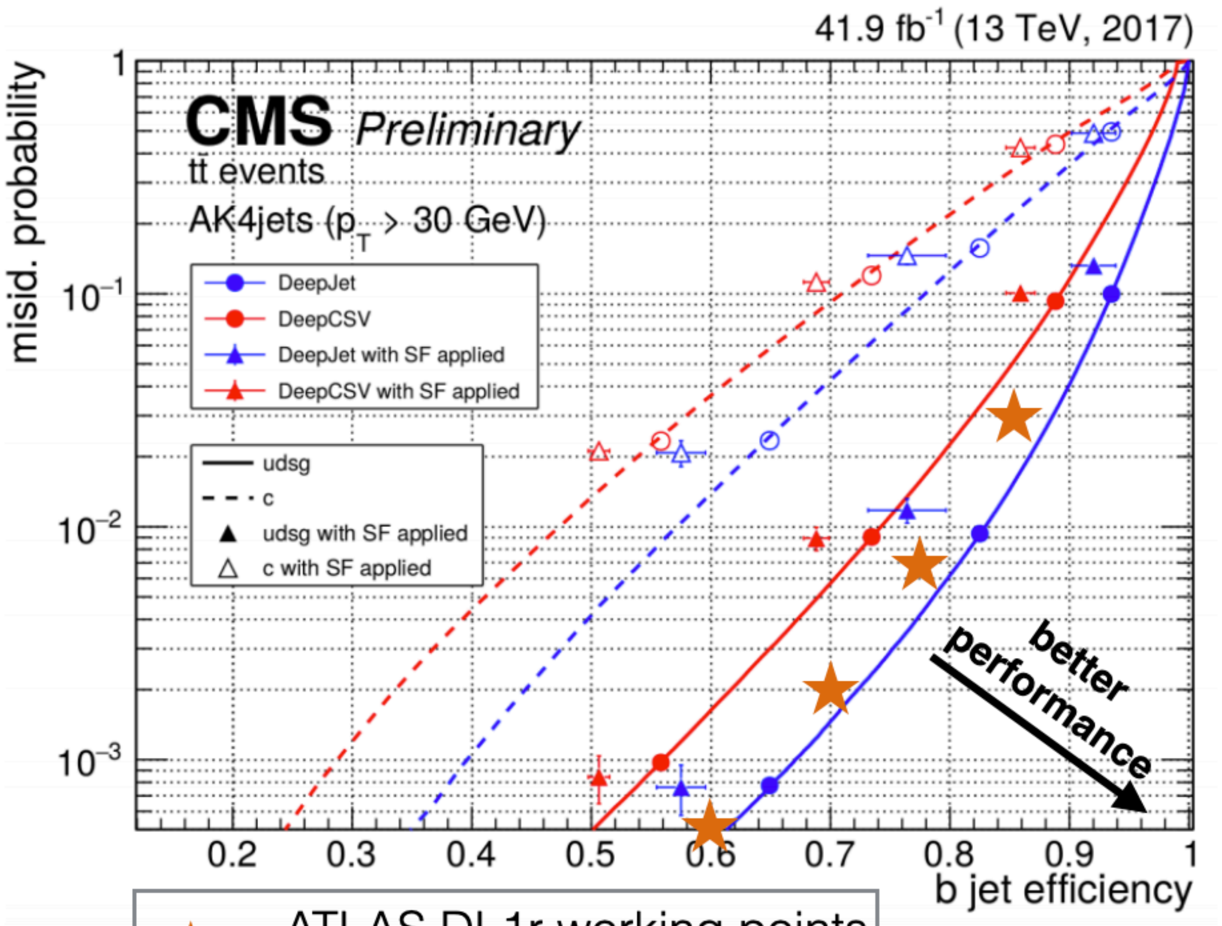
impact on *b*-tagging



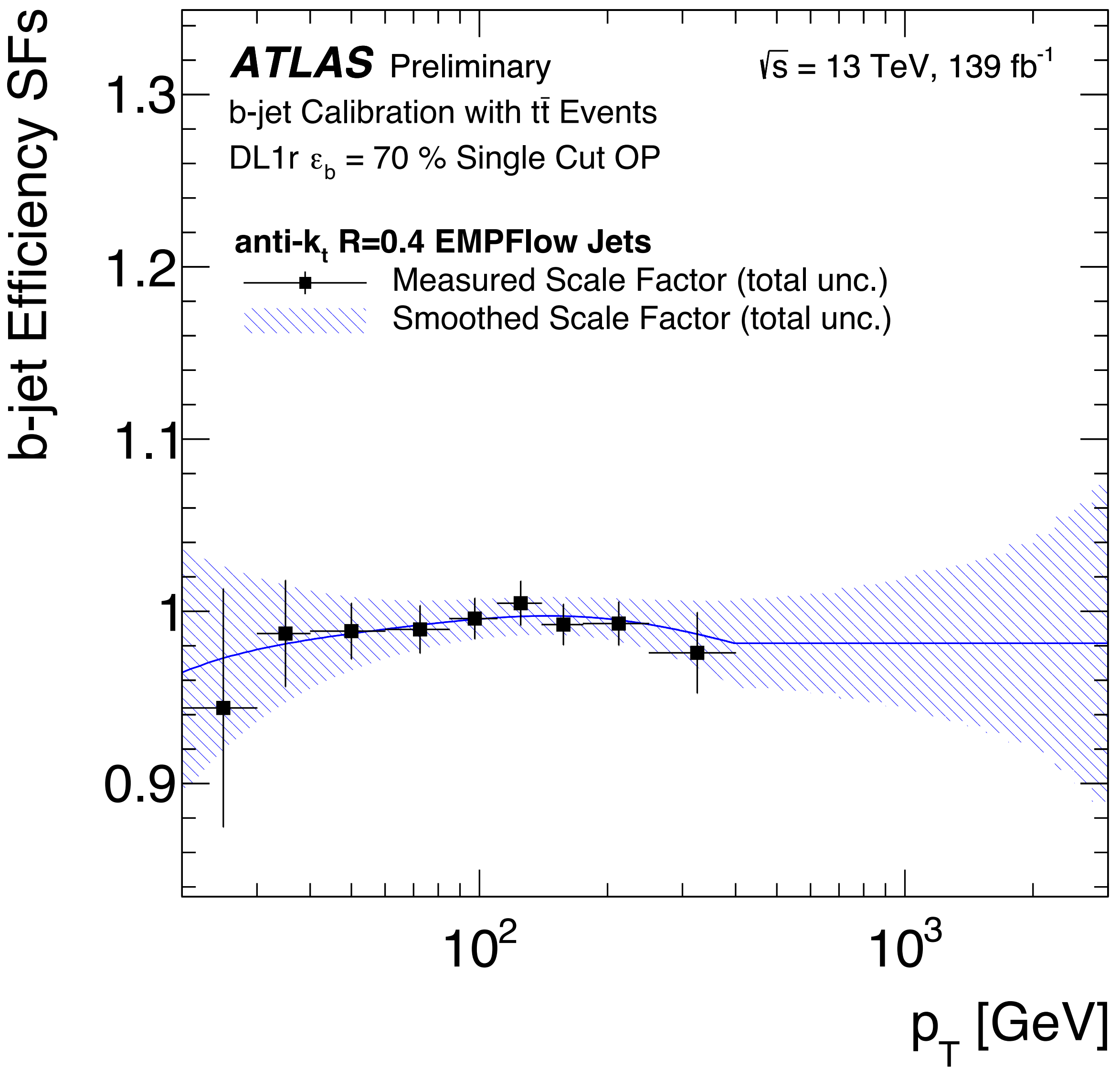


ATLAS-PHYS-PUB-2020-009

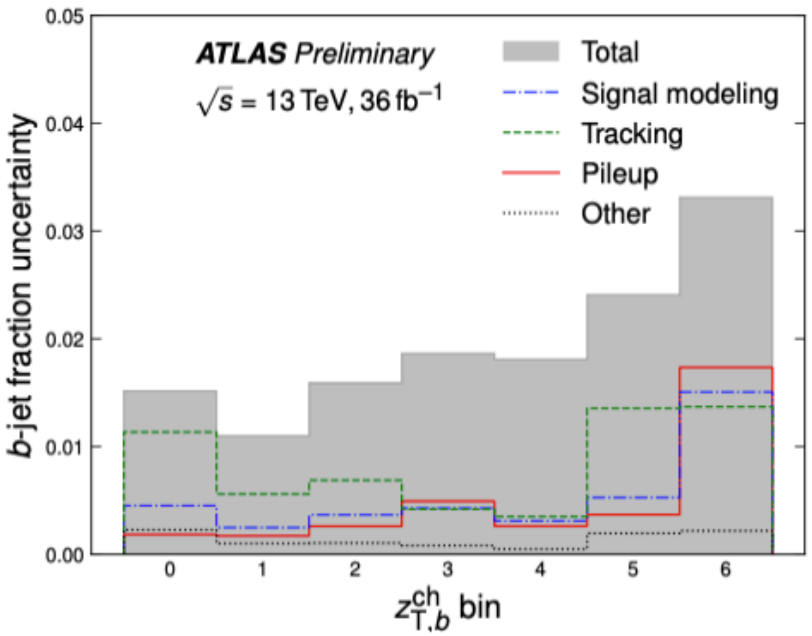
b-tagging SFs



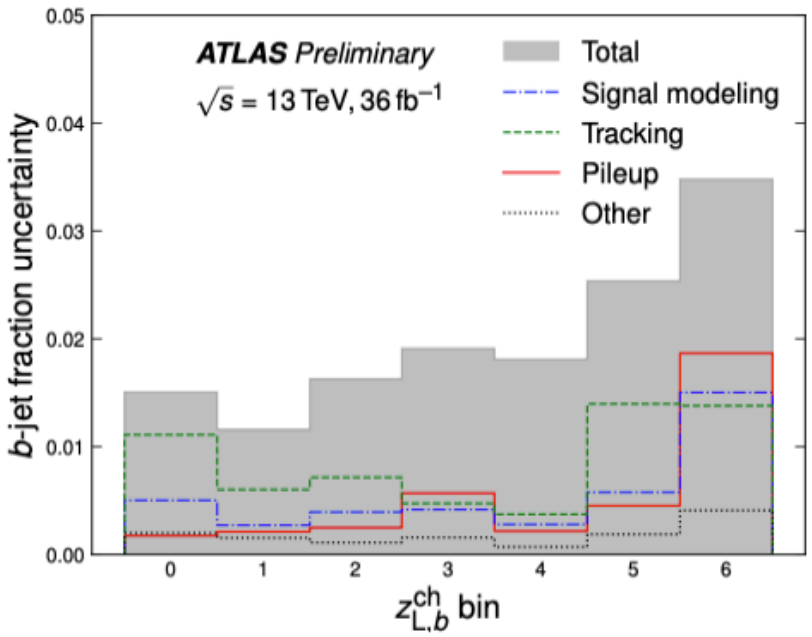
★ ATLAS DL1r working points **(addition mine)**



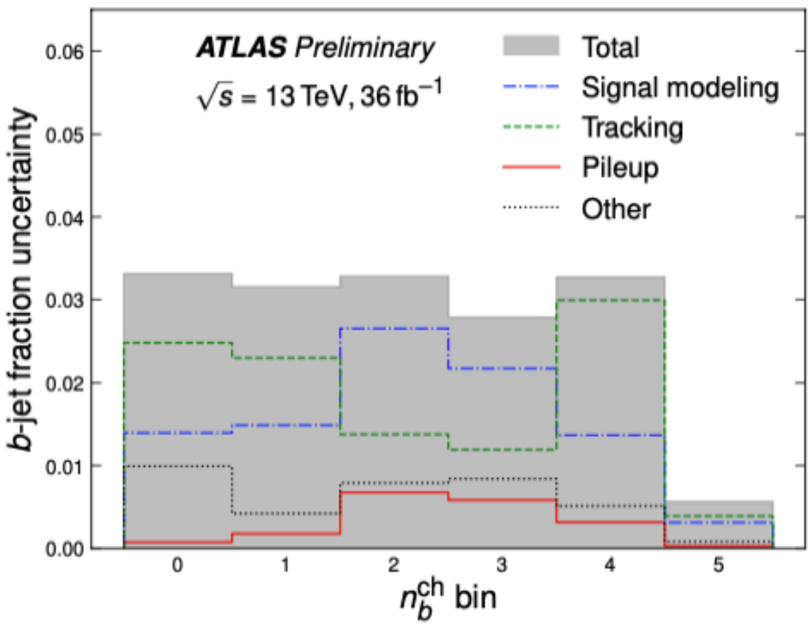
uncertainty breakdowns



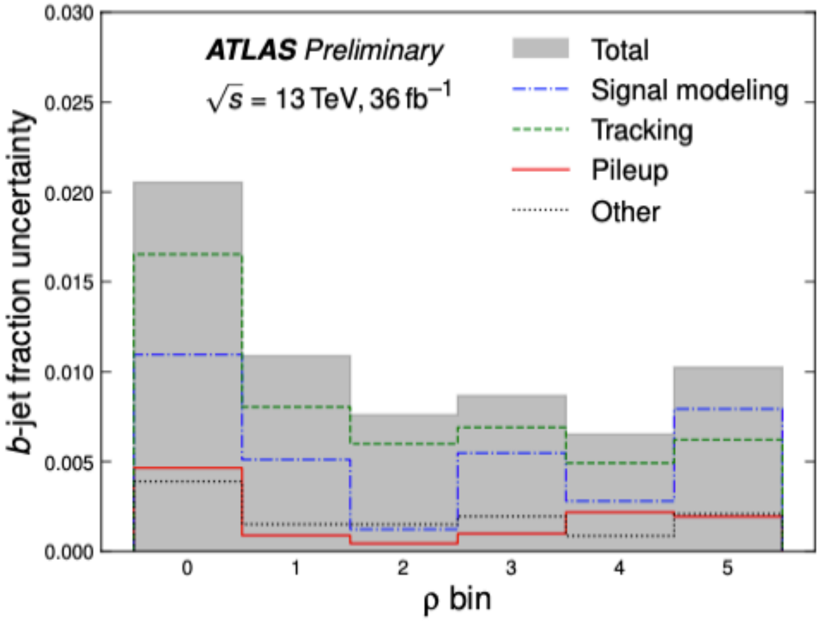
(a) $z^{\mathrm{ch}}_{\mathrm{T},b}$



(b) $z^{\mathrm{ch}}_{\mathrm{L},b}$

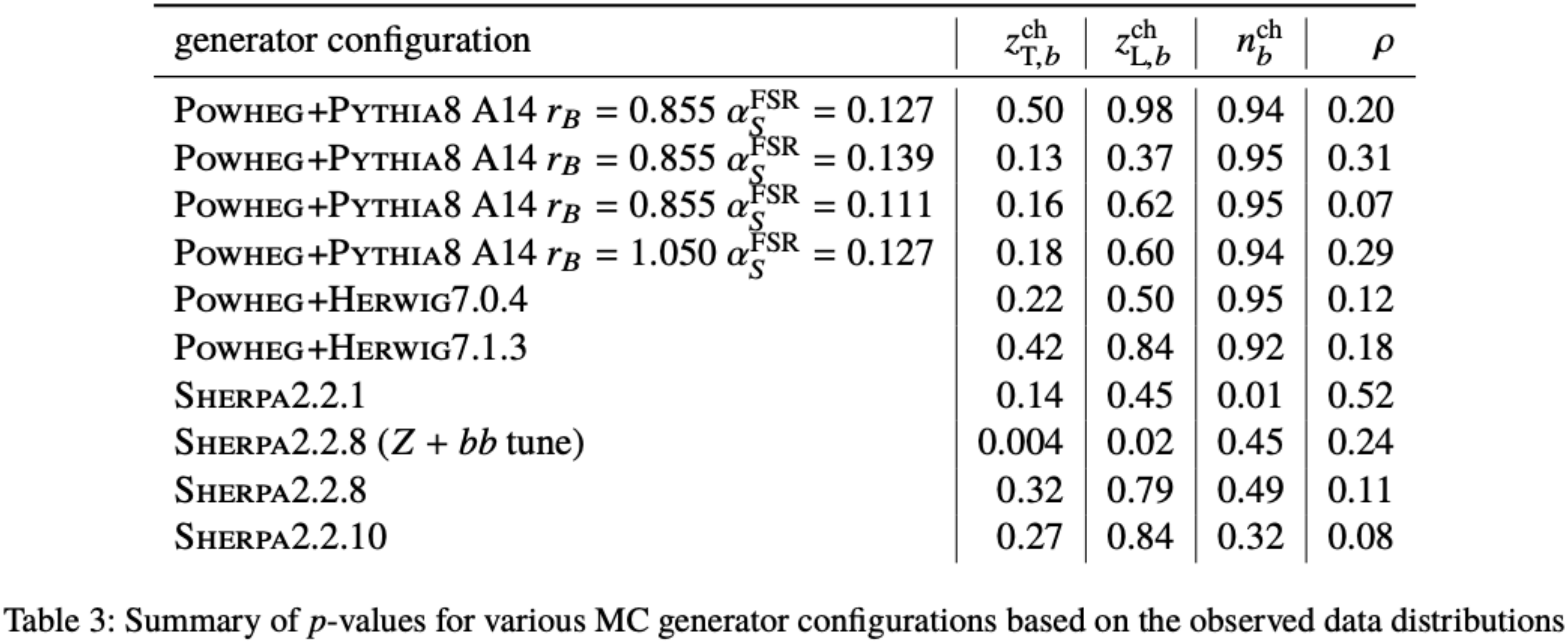


(c) n^{ch}_{b}



(d) ρ

p-values



generator configuration	$z^{\text{ch}}_{\text{T},b}$	$z^{\text{ch}}_{\text{L},b}$	n^{ch}_{b}	ρ
POWHEG+PYTHIA8 A14 $r_B = 0.855$ $\alpha_S^{\text{FSR}} = 0.127$	0.50	0.98	0.94	0.20
POWHEG+PYTHIA8 A14 $r_B = 0.855$ $\alpha_S^{\text{FSR}} = 0.139$	0.13	0.37	0.95	0.31
POWHEG+PYTHIA8 A14 $r_B = 0.855$ $\alpha_S^{\text{FSR}} = 0.111$	0.16	0.62	0.95	0.07
POWHEG+PYTHIA8 A14 $r_B = 1.050$ $\alpha_S^{\text{FSR}} = 0.127$	0.18	0.60	0.94	0.29
POWHEG+HERWIG7.0.4	0.22	0.50	0.95	0.12
POWHEG+HERWIG7.1.3	0.42	0.84	0.92	0.18
SHERPA2.2.1	0.14	0.45	0.01	0.52
SHERPA2.2.8 ($Z + bb$ tune)	0.004	0.02	0.45	0.24
SHERPA2.2.8	0.32	0.79	0.49	0.11
SHERPA2.2.10	0.27	0.84	0.32	0.08

Table 3: Summary of p-values for various MC generator configurations based on the observed data distributions