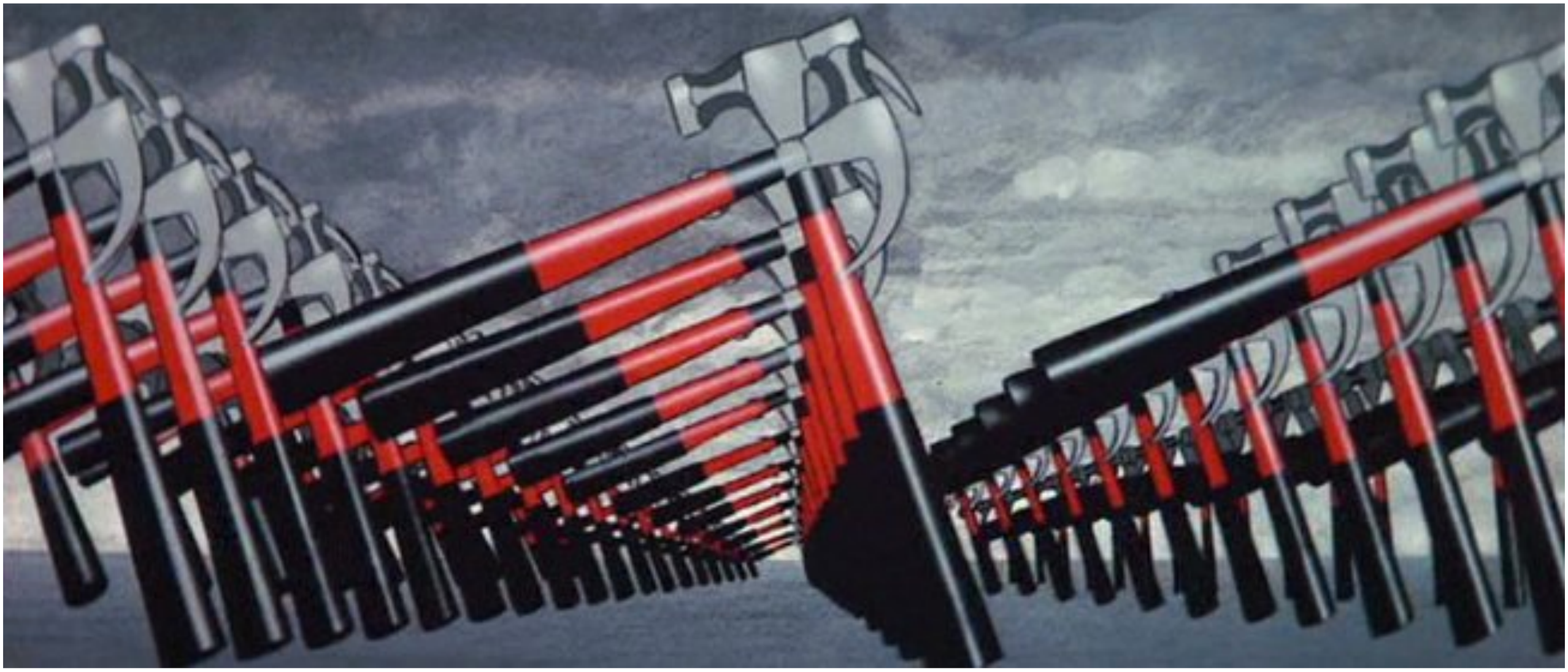


Hammering the systematics wall

P. Ferreira da Silva(CERN)

Top! Hammertime (a CMS top quark workshop)

Tuesday, 14th Nov. 2017

Outlook

What do the data tell us so far? unexpected findings and their impact

Current approach to systematics in Top

Challenges ahead

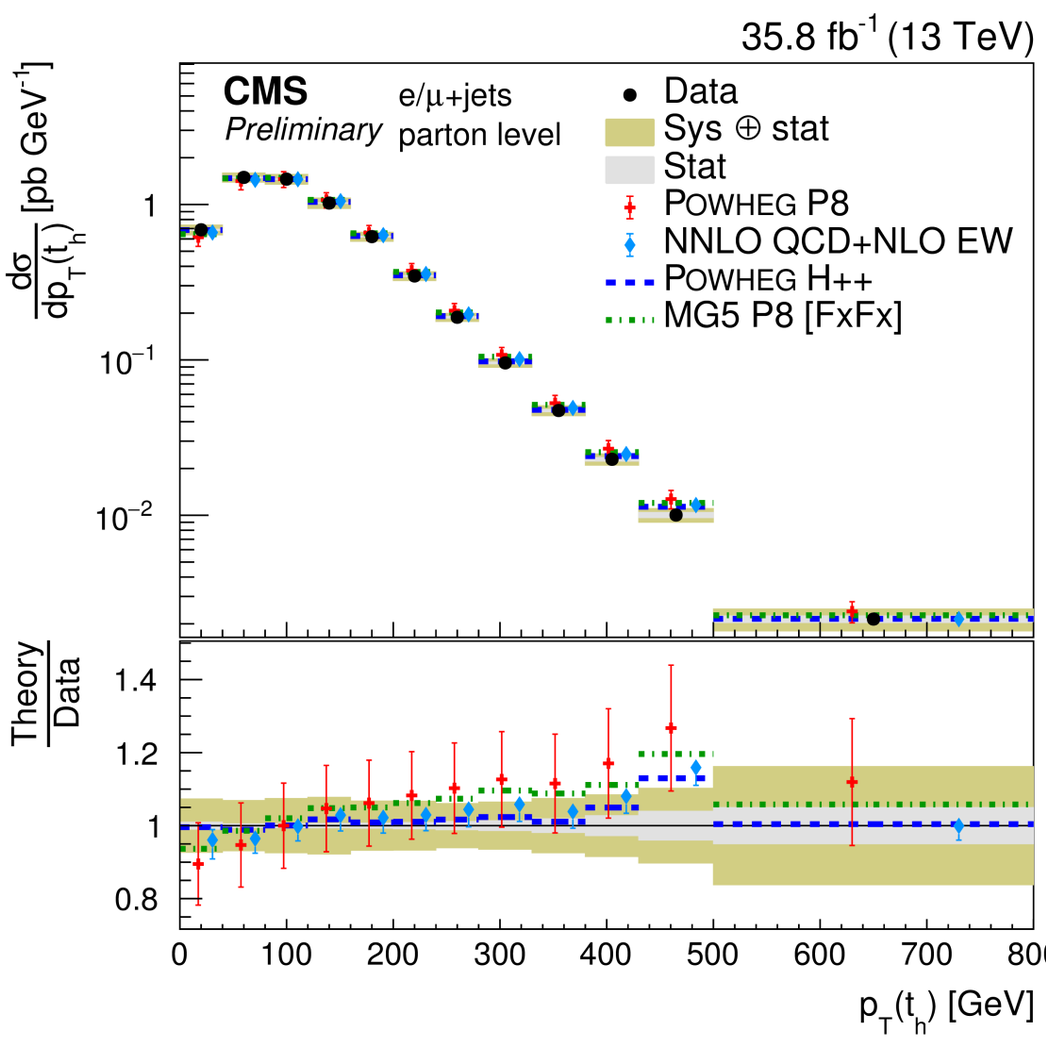
Preamble

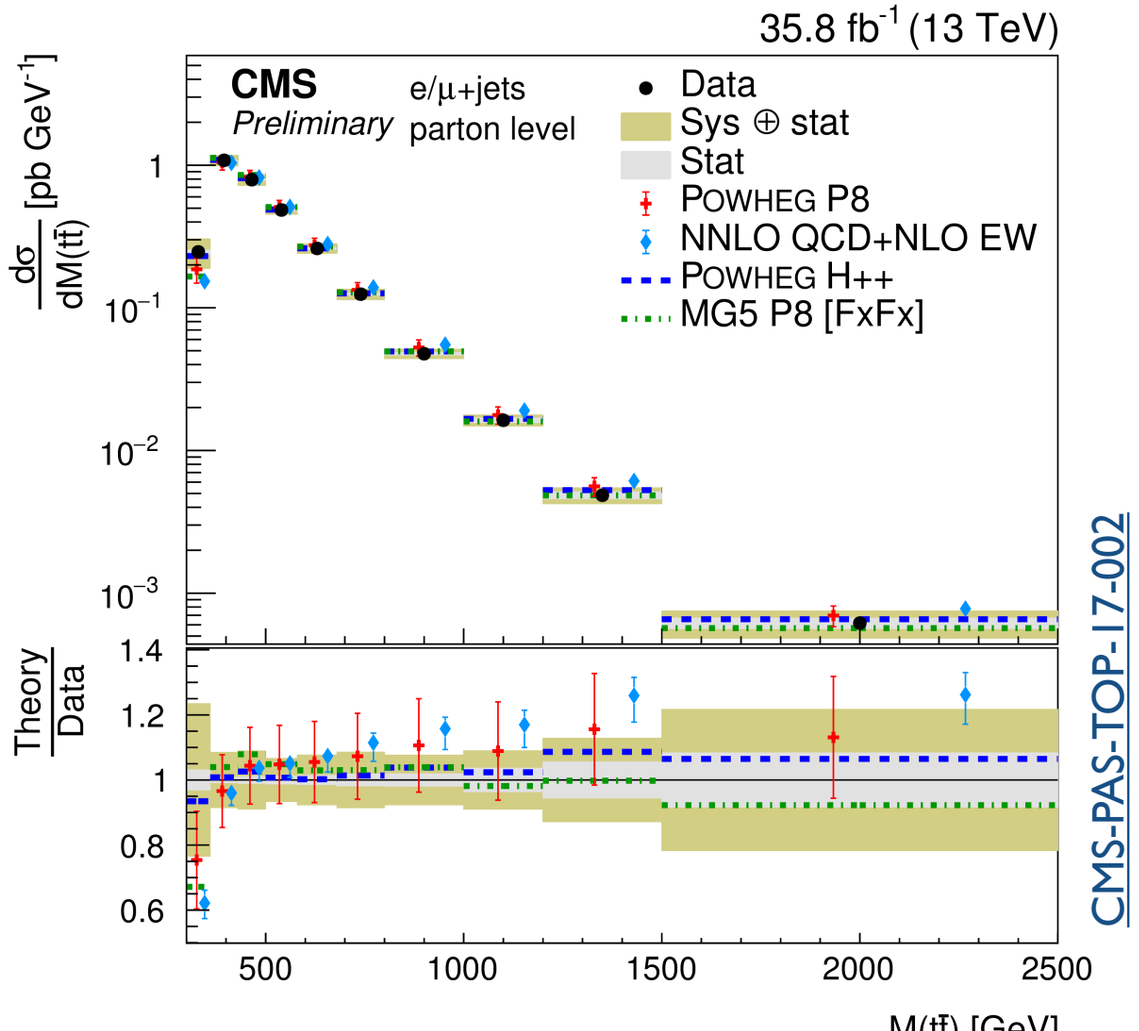
- **There is no general recipe to be given that covers every single analysis**
  - final state, kinematics range, detectors involved, dataset ⇒ **look for specific experimental effects**
  - dependency on a simulated reference ⇒ **question specific calibration assumptions**
  - re-interpretation of the measurement ⇒ **gauge the uncertainties in the theory predictions**
- **As a rule of thumb all analyses can be forecasted based on a simulation**
  - with *standard* data taking scenario, object definitions prescriptions
  - useful to identify the culprits on where to improve before final result is unblinded
  - but we are aiming for more than the average (=standard)
- **"state-of-the-art" means rather "starting point"**
  - the CMS.TopSystematics twiki is a good starting point but it's not a recipe book
  - it's absolutely dated to the knowledge/limitations at this point in time
- **Next slides: highlight main challenges to make the best out of systematics for top in Run 2**

Unexpected findings in tt kinematics…

- **In general NLO+PS MC describe well the data (rate and shapes)**
  - most challenging: describe both top p_T and M(tt) simultaneously
  - Herwig++/Sherpa seem to be more accurate
- **Higher order corrections fail to predict both correctly**
  - see more details in the talks by J. Keaveney and A. Mitov

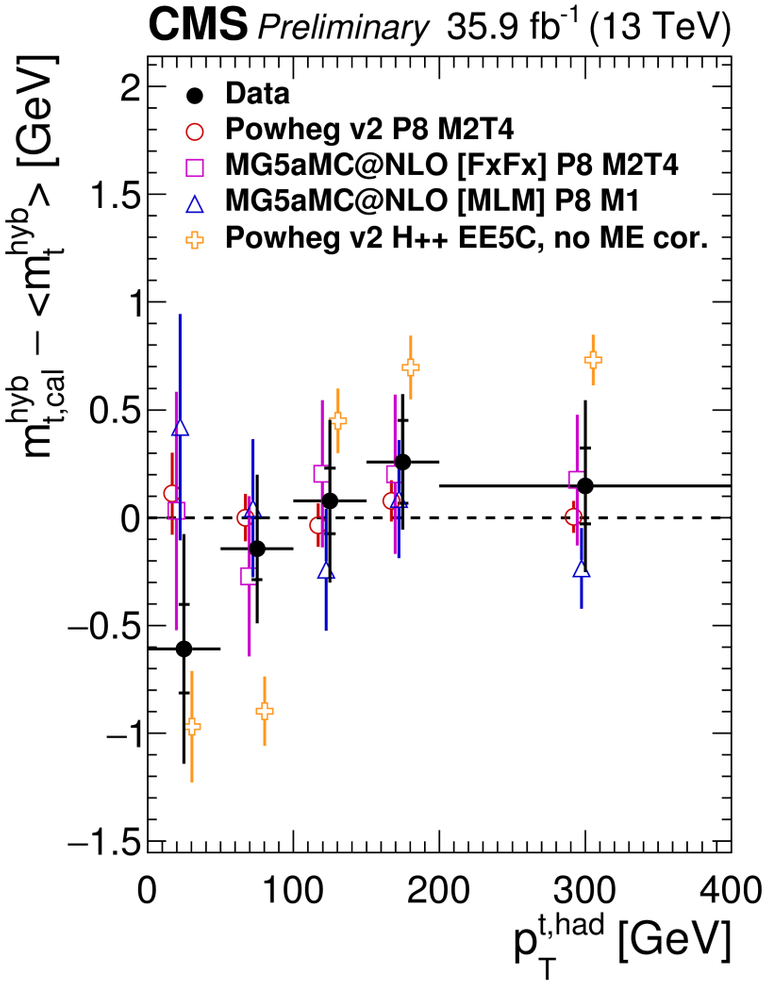
p-values	$p_T(t_h)$	M(tt)
Powheg+PY8	0.02	0.314
aMC@NLO+PY8	0.53	0.275
Powheg+HW++	0.955	0.940
Sherpa	0.335	0.985
NNLO QCD + NLO EW	0.953	<0.01

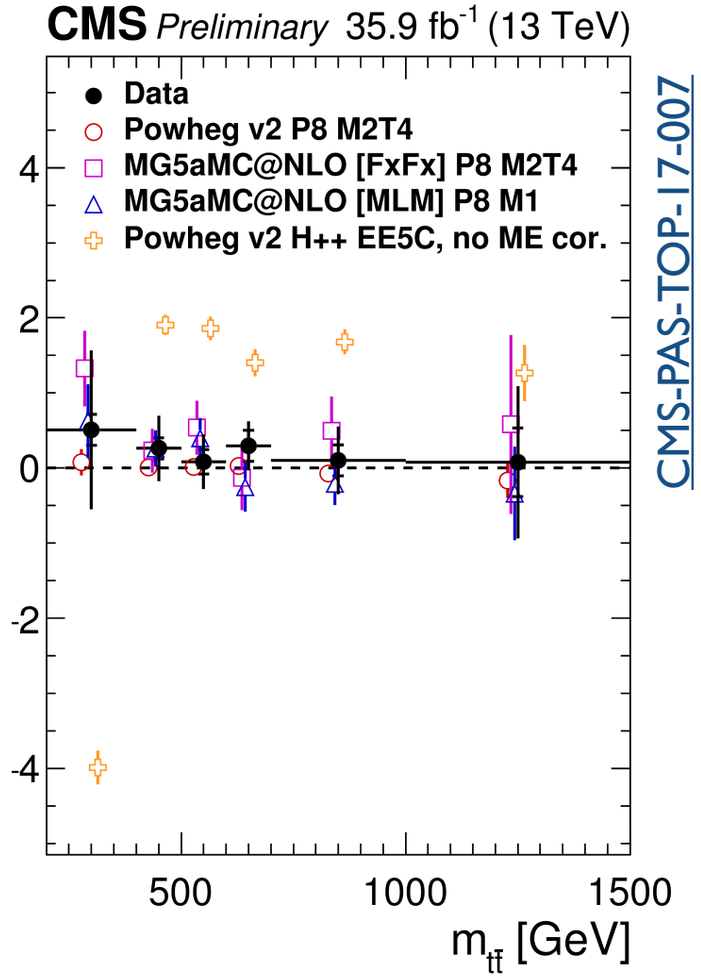


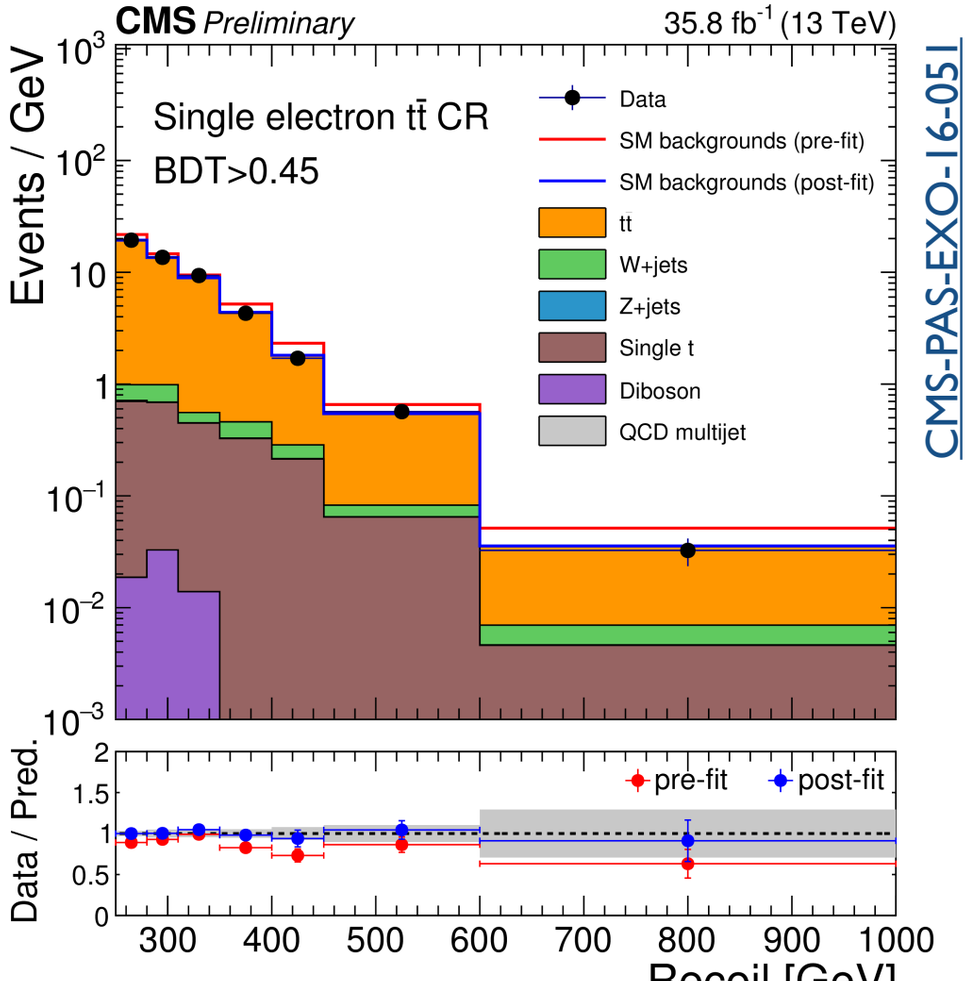


Unexpected findings in tt kinematics…and their impact

- **If measurements are based on Lorentz-invariant quantities, impact is mitigated**
  - e.g. top mass δm_t<500 MeV
  - despite the better agreement for HW++, the measurement disfavours the used setup
- **Searches: usually take into account mis-modelling by including a nuisance in the fit**
  - impact mitigated by constraining the nuisances in the control regions

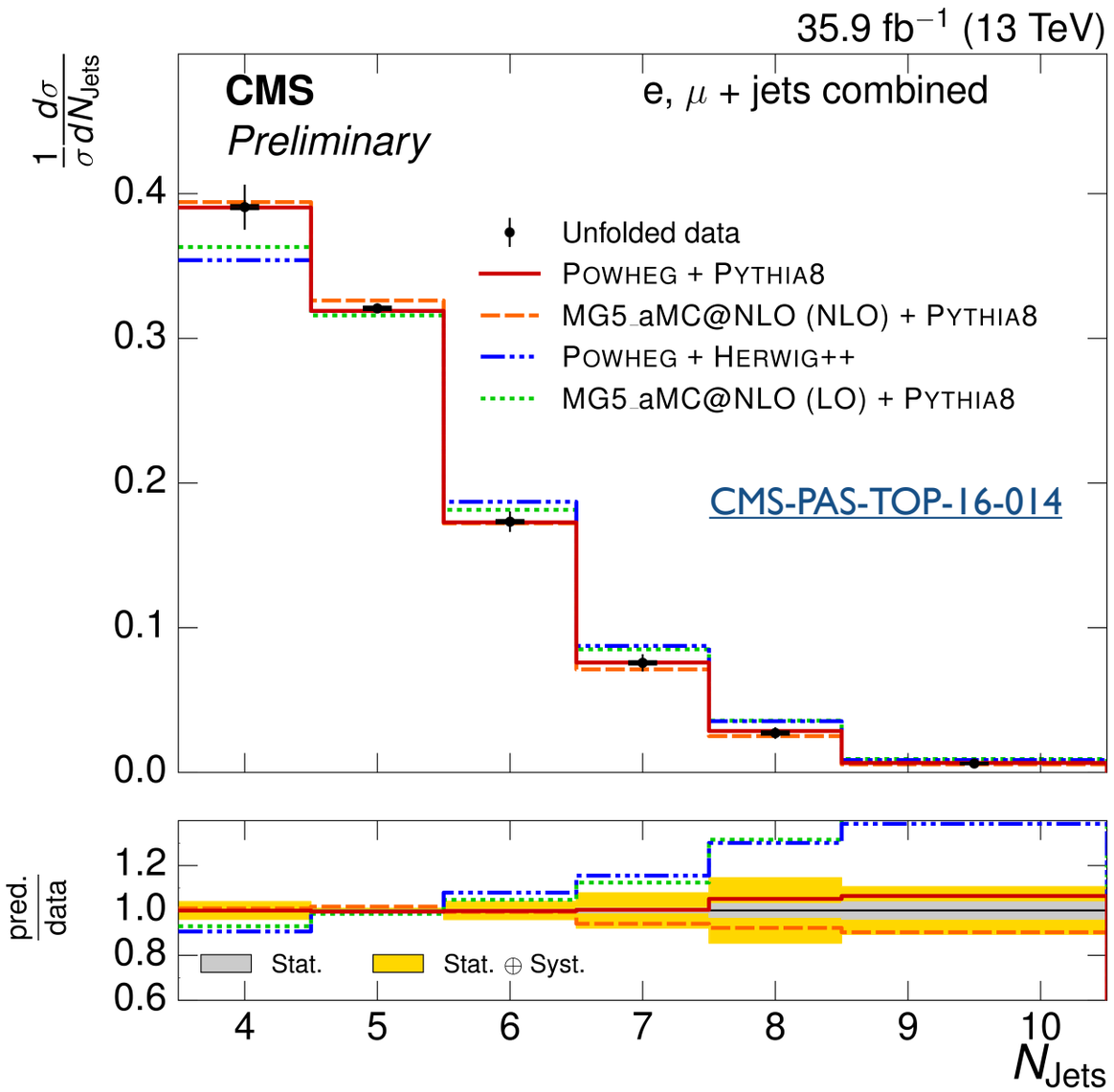


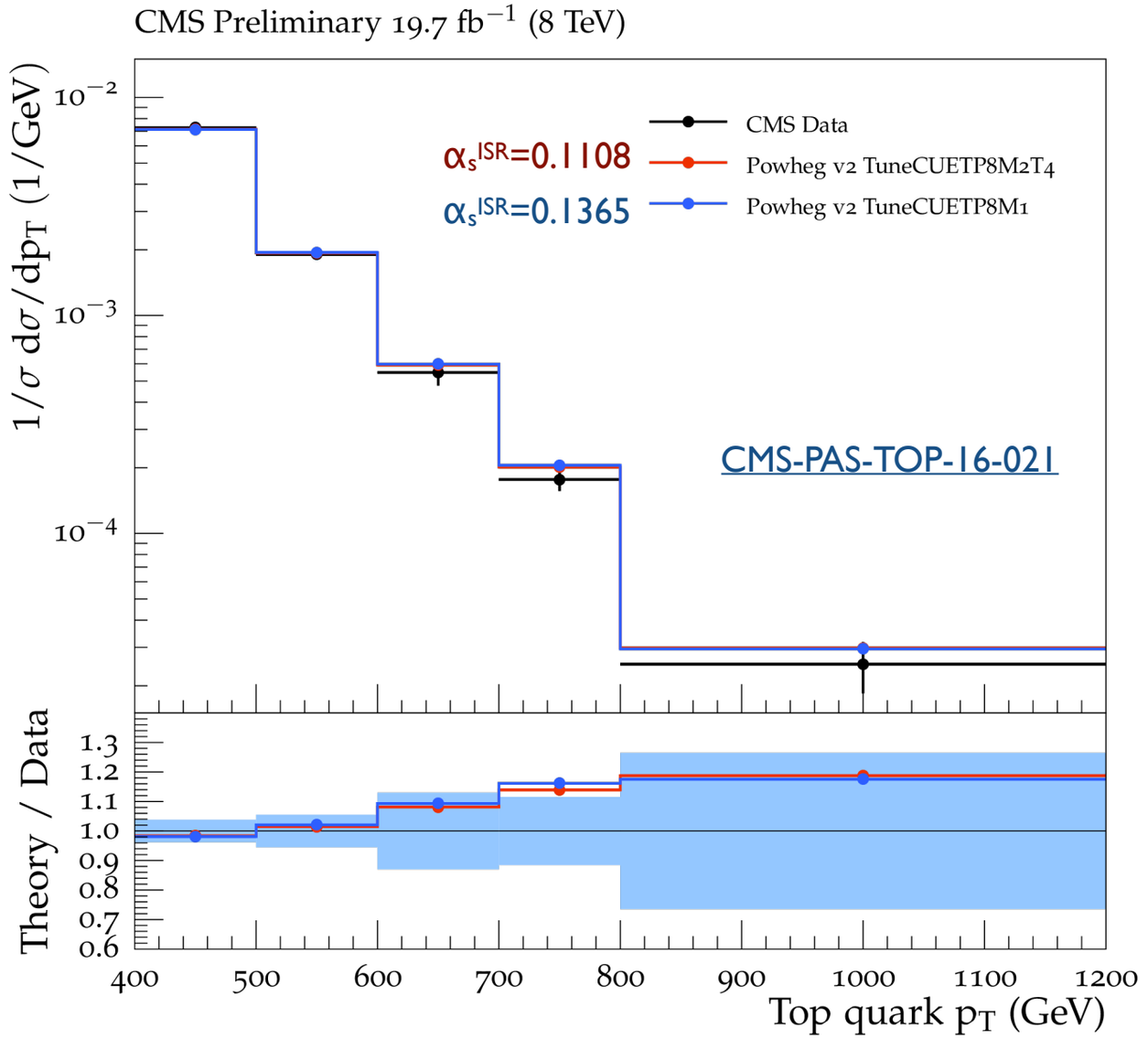




Unexpected findings in jet activity/underlying event…

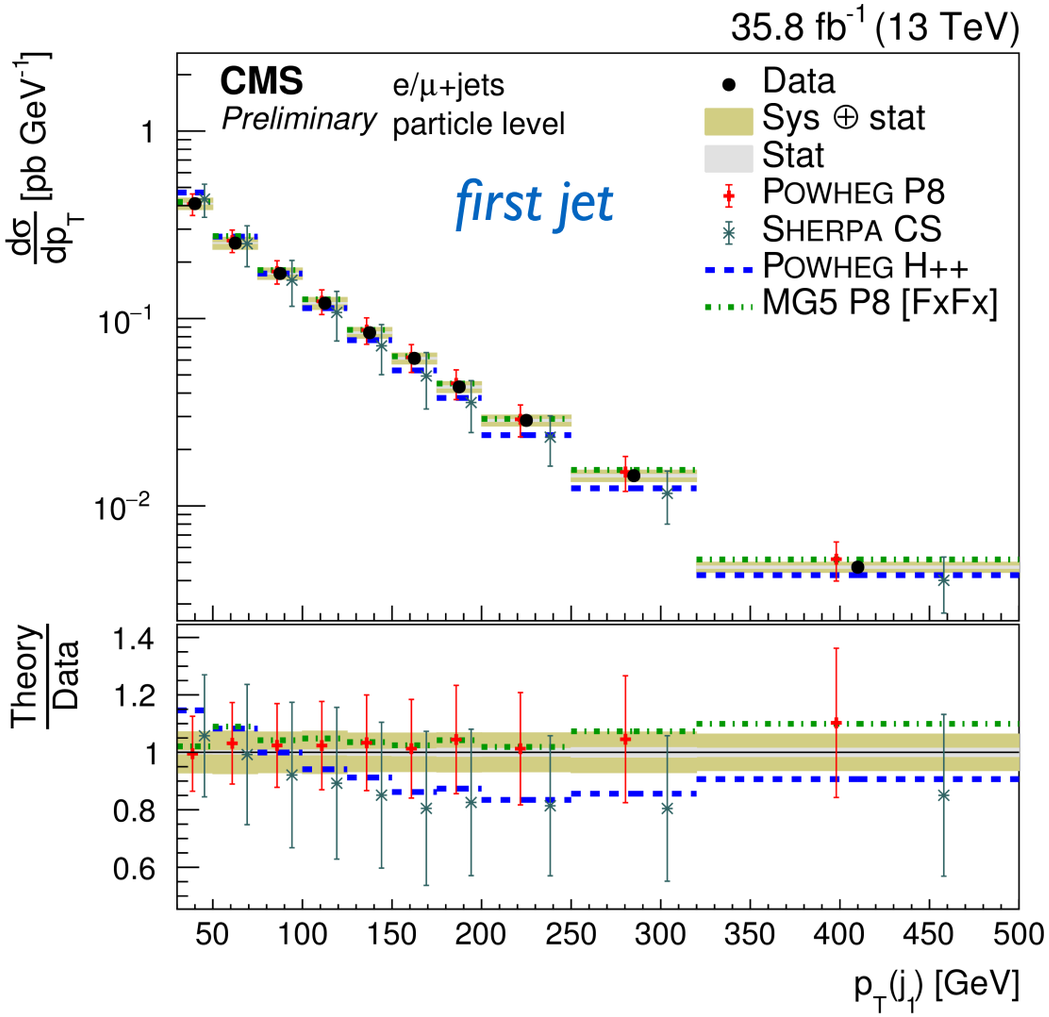
- **The current baseline MC has a strong dependency on the parton shower**
  - emission rates for p_T>30 GeV are reproduced in Pythia8 after a retune of α_s^{ISR} to 8 TeV data
  - tuning is expected to have negligible impact on kinematics/global observables (H_T, E_T^{miss}, etc.)

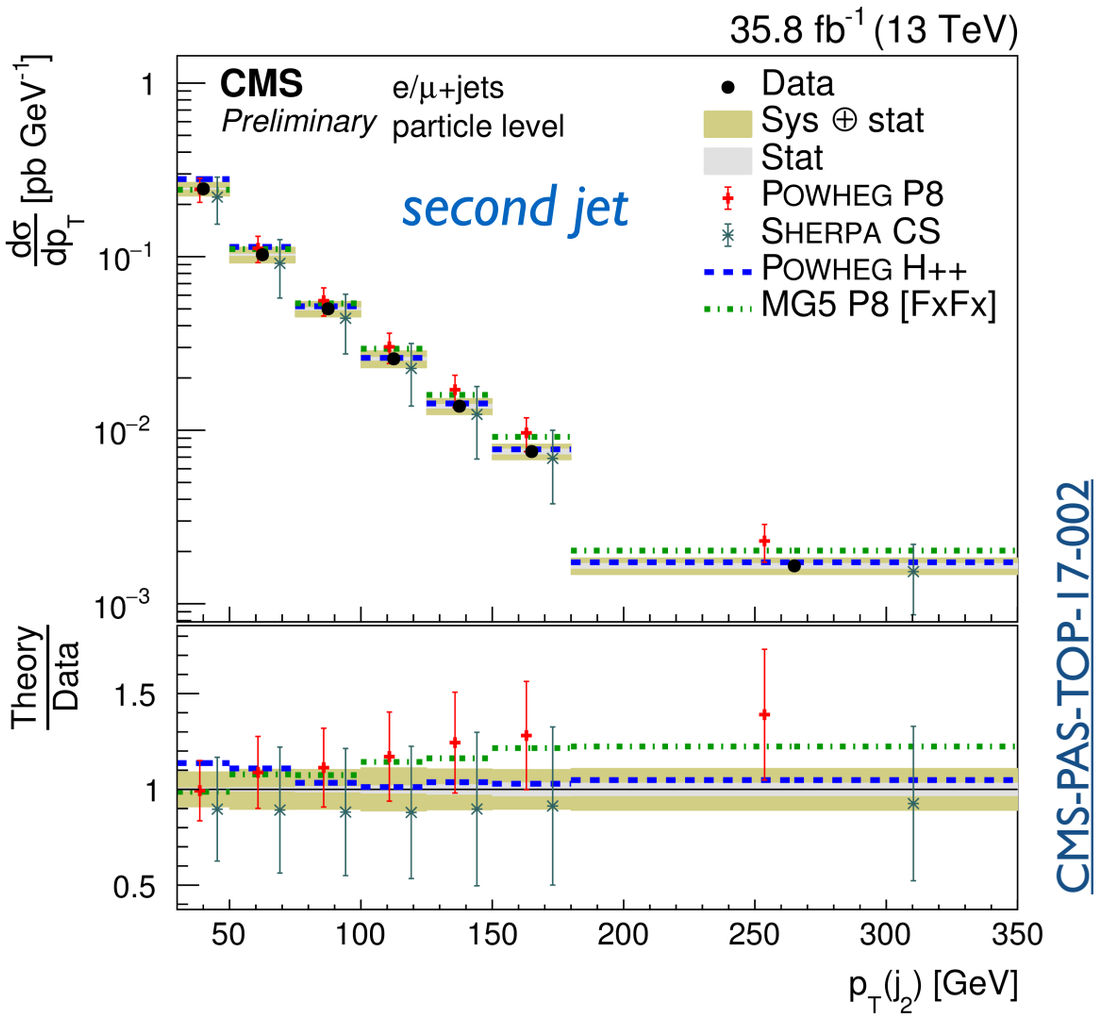




Unexpected findings in jet activity/underlying event…

- **The current baseline MC has a strong dependency on the parton shower**
  - emission rates for p_T>30 GeV are reproduced in Pythia8 after a retune of α_s^{ISR} to 8 TeV data
  - tuning is expected to have negligible impact on kinematics/global observables (H_T, E_T^{miss}, etc.)
  - but additional jet p_T spectra not well reproduced in general (just covered by variations in PS)
  - in general **p-values<0.001 for all NLO+PS generators**, when compared to our data!



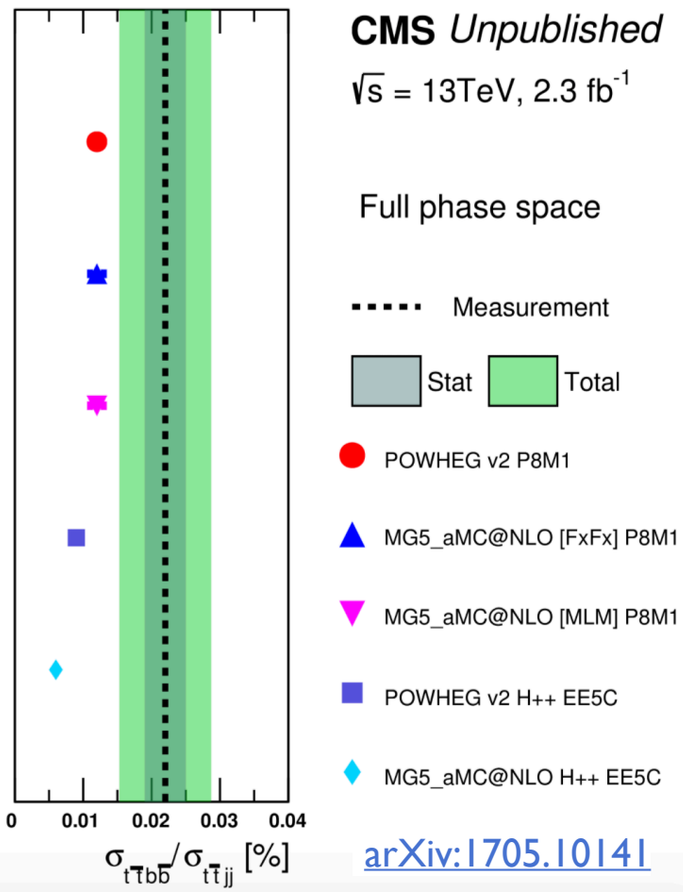


CMS-PAS-TOP-17-002

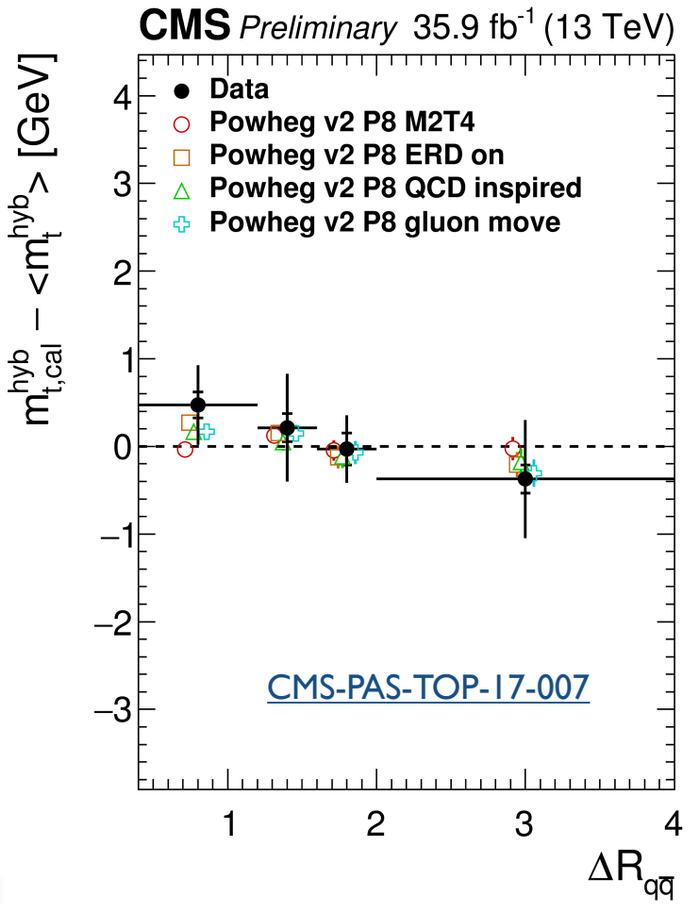
Unexpected findings in jet activity/underlying event…and their impact

- **Control regions can sometimes be used to correct MC**
  - e.g. count extra jets in 2ℓ+2b = mostly dominated by ISR
  - dedicated measurements of tt+bb with corresponding uncertainties

  check: are scales/PDFs/etc. comparable in CR and SR?
  - hard to distinguish different np-QCD models: larger impact on m_t measurements
- **Using only MC: non-negligible impact can be expected from PS scale choices**
  - change of response, kinematics, jet activity lead in general to large variations

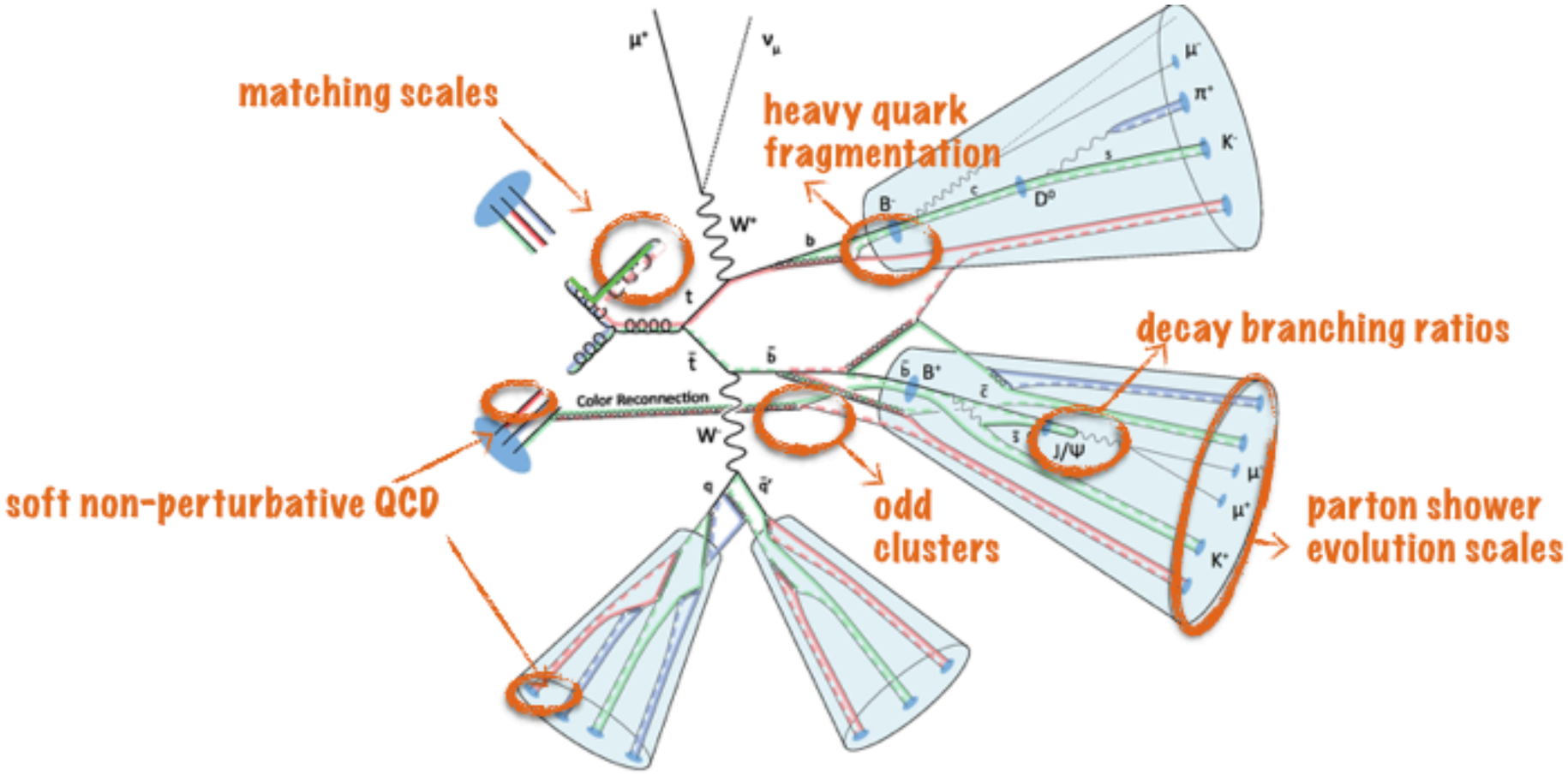


arXiv:1705.10141



ISR/FSR impact [%]	σ(tt) ℓ+j arXiv:1711.03143	σ(tt) all jets TOP-16-013	dσ/dp$_T$ TOP-17-002	4 tops arXiv:1710.10614
Acceptance				6
Shape				15
Unfolding			8	
Fit	4	3 (resolved) 5 (boosted)		

Factorised parton shower/hadronization uncertainties I



- **Test stability of the methods used with independent variations**
  - identify as many as possible using a single MC setup model (well defined "N-1" comparison)
  - must be complemented with ME QCD scale choice, PDF variations
  - in addition: shortcomings of the description of top p_T vs m(tt) must be tested

    time to start assuming (NNLO QCD + NLO EW)/(NLO QCD)?

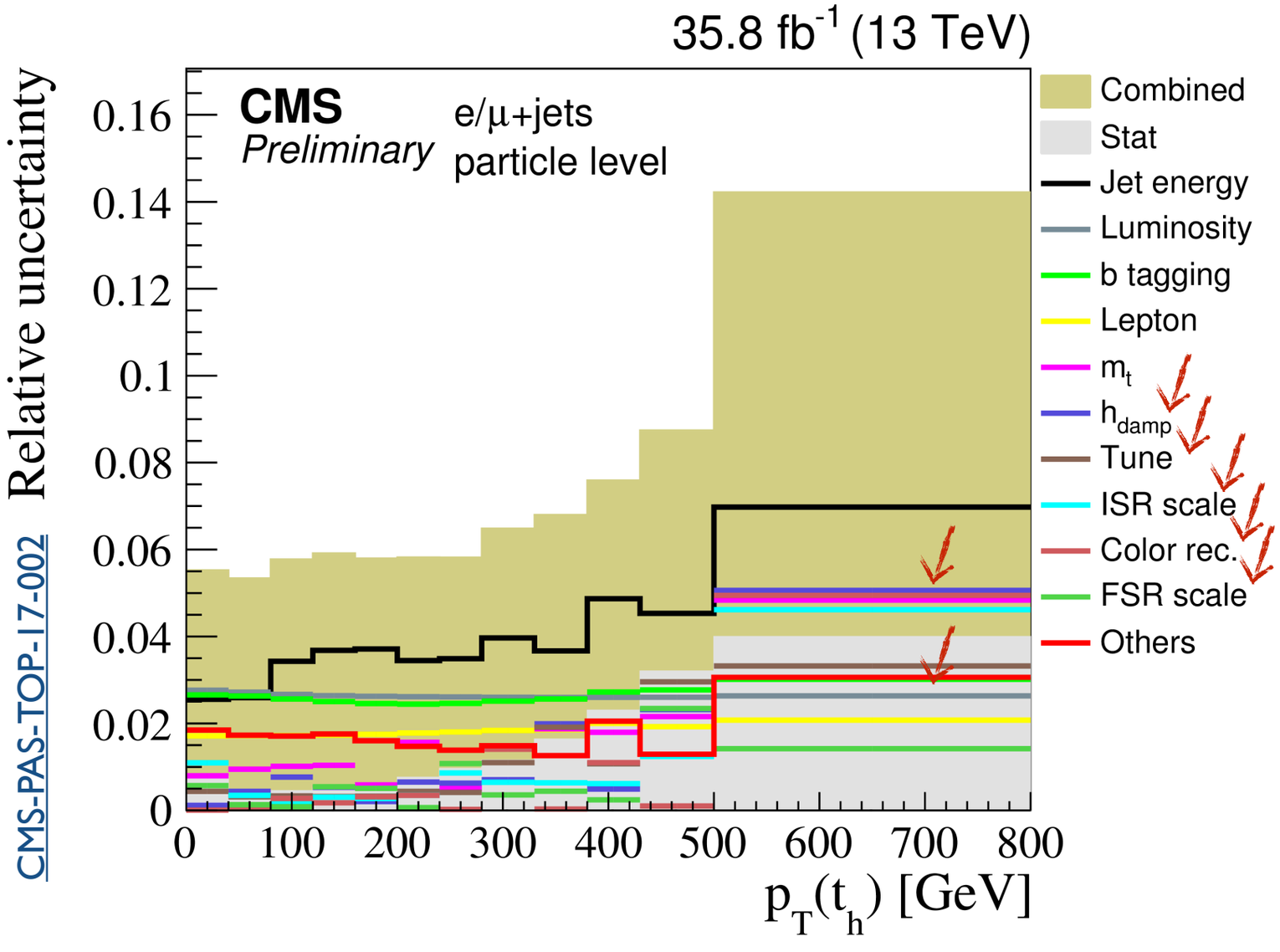
Factorised parton shower/hadronization uncertainties II

- **The "master table" currently in use is based on the Powheg+Pythia8 model**
  - most handles are common to tt(+X) and single top(+X) topologies

Source	Handle	Weights?	Variation	Note
Shower scales	ISR,FSR {Space,Time}Shower: renormMultFac	n	1/2 - 2	scale down FSR by $\sqrt{2}$ (LEP-based?)
Resummation	hdamp	n	$(1.58^{+0.66}_{-0.59})m_t$	TOP-16-021
Soft QCD	MPI/CR strength ERD off	n	uncertainties from tune	resonance decays not affected
Colour reconnection	Enable resonance decays in CR	n	MPI,QCD-based, gluon flip models	
Fragmentation	$x_b=p_T(B)/p_T(b\ jet)$	y	Vary Bowler-Lund, Peterson within LEP/SLD fits	TOP-16-022
Decay tables	b hadron decays	y	re-weight default decay tables to PDG uncs.	
Hadronization	Pythia vs Herwig	n	Included in JES (independent for light, g, c or b jets)	
ME QCD scales	(μ_R,μ_F)	y	1/2 - 2	exclude (1/2,2) or (2,1.2)
PDFs	single set	y	depends on set	preferably use non top PDF, Hessian uncs.

Impacts (examples)

- **PS systematics rank high (experimental uncertainties discussed this afternoon - link)**
  - particular visible in "jet busy" phase spaces (compare e.g. ISR/FSR in tt and tW)
  - having separate sources ⇒ study incremental improvements separately



CMS-PAS-TOP-17-002

Source	Uncertainty (%)
Trigger efficiencies	2.7
Muon efficiencies	3.1
Electron efficiencies	3.2
Jet energy scale	3.2
Jet energy resolution	1.8
b tagging efficiency	1.4
Mistagging rate	0.2
Pileup	3.3
$t\bar{t}$ μ_R and μ_F scale	2.5
tW μ_R and μ_F scale	0.9
Underlying event	0.4
ME/PS matching	1.8
Initial state radiation	0.8
Final state radiation	0.8
Color reconnection	2.0
PDF	1.5
DR-DS	1.3
VV normalization	0.4
Drell–Yan normalization	1.1
Non-W/Z leptons normalization	1.6
$t\bar{t}V$ normalization	0.1
MC statistics	1.6
Full phase space extrapolation	2.9
Total systematic (excluding integrated luminosity)	9.5
Integrated luminosity	3.3
Statistical	2.8
Total	10.5

CMS-PAS-TOP-17-018

Critical appraisal of current approach to systematics

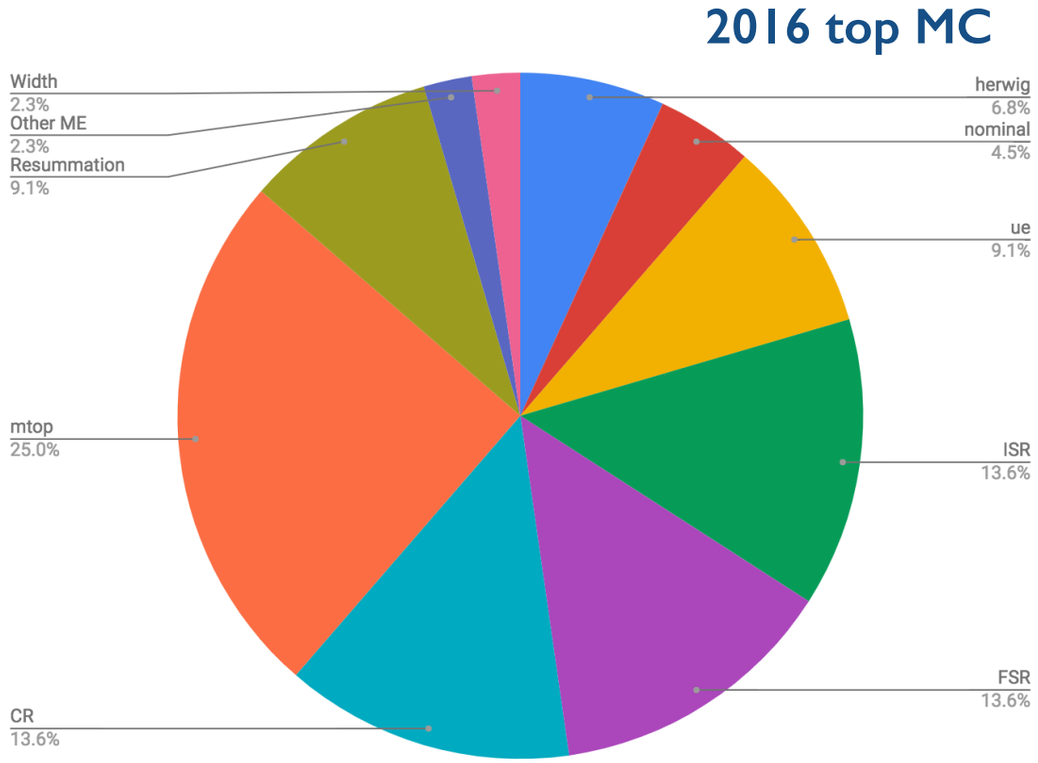
- **Calculations are useful if uncertainty bands are trustworthy**
  - most of the techniques we use need validation through MC calculations
- **Varying the salient parameters**
  - identify dominant sources and approach better coverage of real uncertainty
  - require resources (CPU, storage) and increases analysis complexity
  - which parameters? which ranges? correlated or not?
  - does model allow to cover all possible variations?
- **Comparing model A to model B**
  - at most indicative of possible effects
  - effective: single point-to-point and it's done
  - but if they agree: common limitation in the phase space under study? accidental?
- **Part of the answers are (will be) in the data** (examples in the next slides)

Challenges ahead I

- **The current recipe will hardly scale with the resources for higher luminosity**
  - need alternatives to O(10^9) MC productions

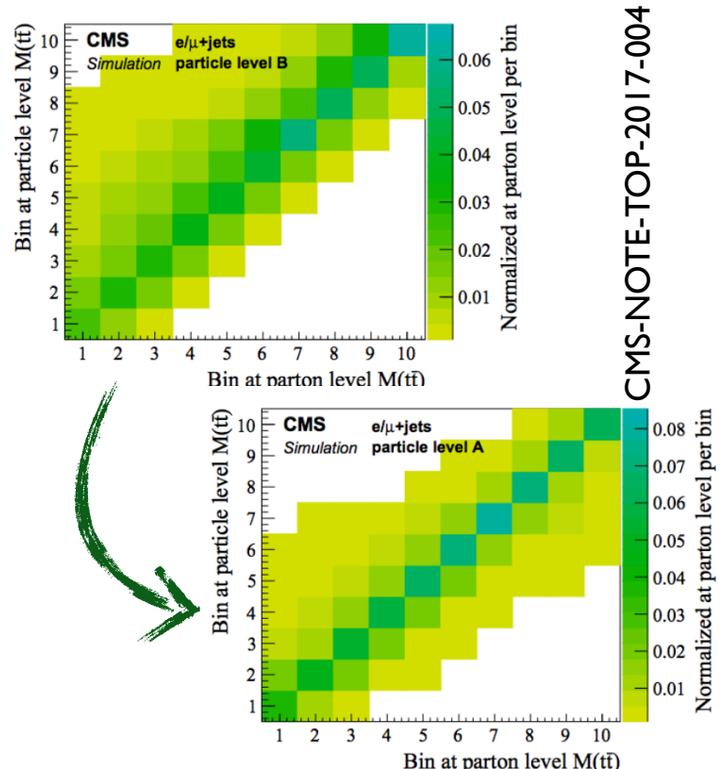
- **PS (+UE) weighting potential to reduce ~28% (50%)**
  - at the same time need to improve accuracy (e.g. N_{jets})
  - negative weights reduce $\mathcal{L}_{eq} = \frac{N}{\sigma} \cdot (1 - 2 \cdot f_{neg})^2$

    (84% for 30% negative weights)



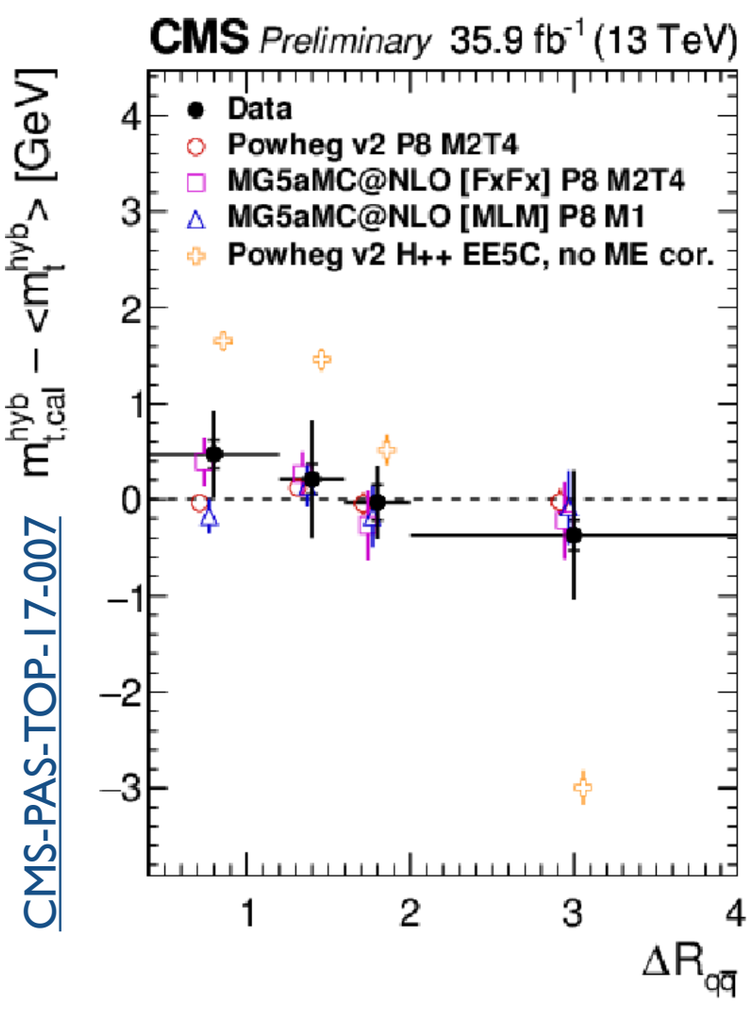
- **What can be taken at generator level/ from fixed order?**
  - folding or re-weighting to predict reconstruction level: accurate enough?
  - maybe yes if analyses designed to be as close as possible to hard process

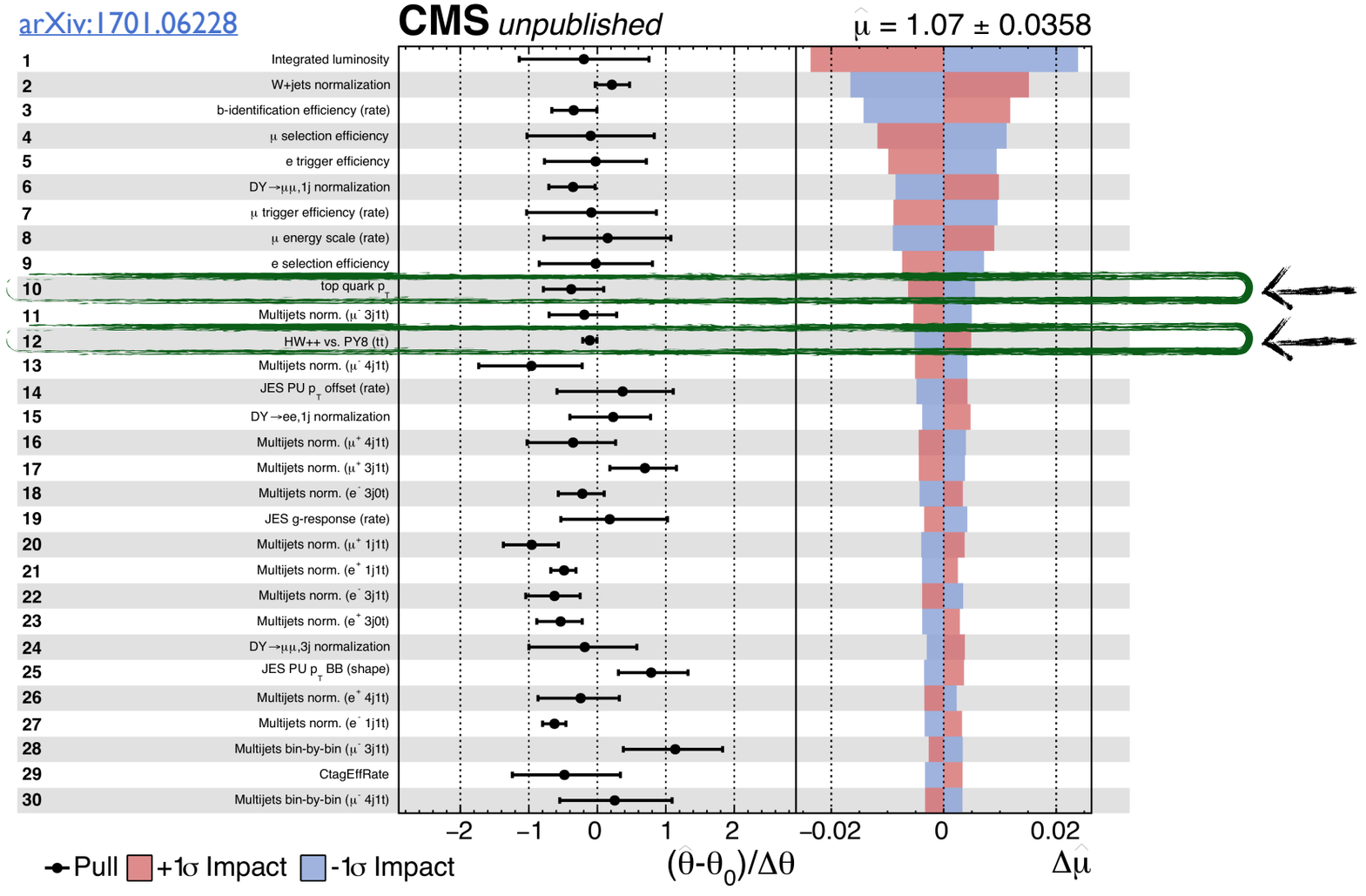
    (reduce combinatorics, improve correct assignments, etc.)



Challenges ahead II

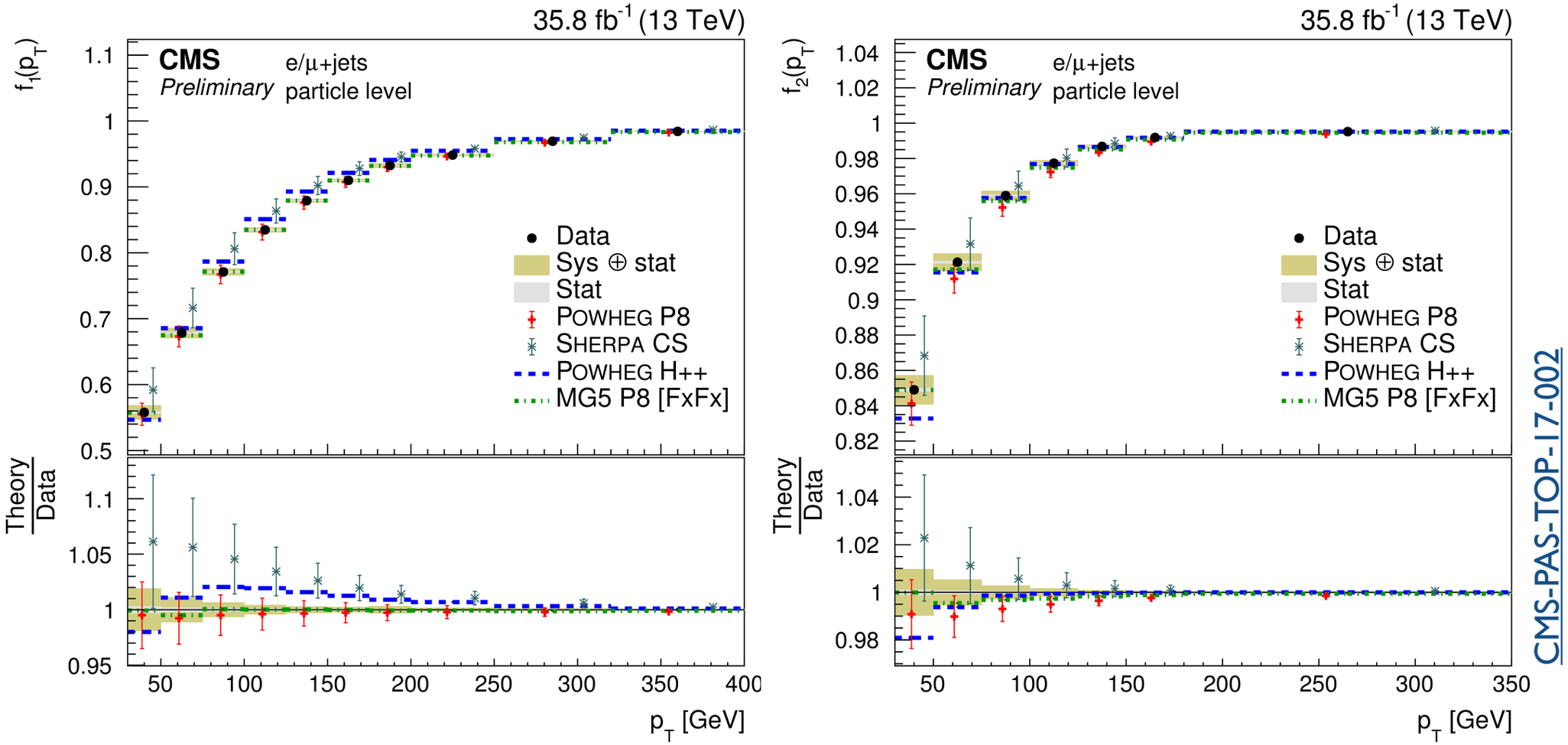
- **Can only benefit from statistics by going differential, taking ratios to references, etc.**
- **Full likelihoods needed parametrising effect of systematics through nuisances**
  - can be included in unfolding-like procedures (see talks by J. Kieseler and J. Bendavid)
  - but need to be careful: single nuisance to rule full phase space is often incomplete (different scales, sensitivity to PDFs, detector effects, etc.)





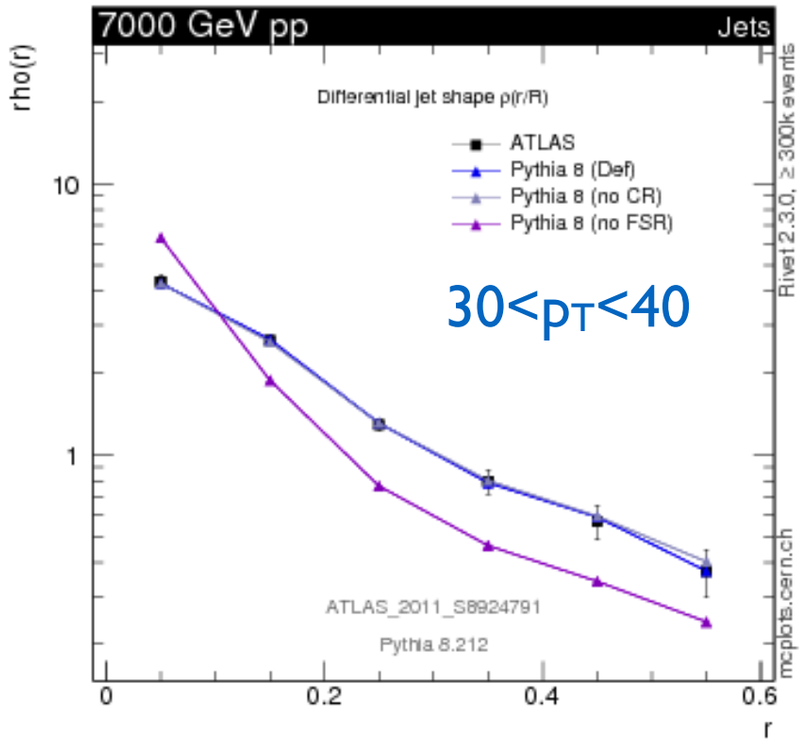
ISR scale through extra jet activity

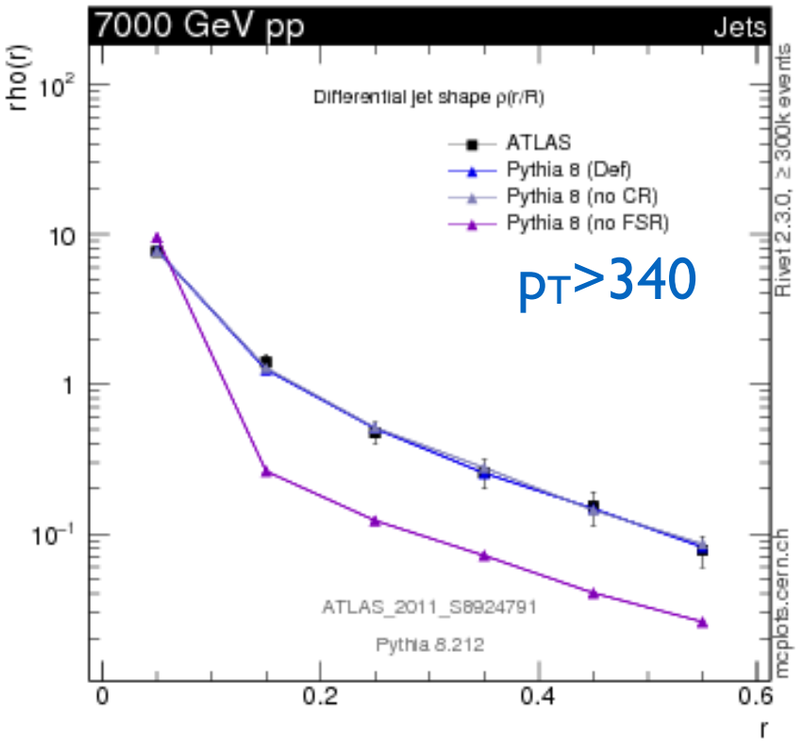
- **Jet multiplicity proved useful to tune α_s^{ISR} but gives incomplete picture**
  - bound to fixed p_T thresholds used
- **Gap fractions are precise measurements**
  - cancel out several systematics while preserving the sensitivity
- **Measurements in the dilepton events**
  - cleaner definition of ISR, profit from high statistics

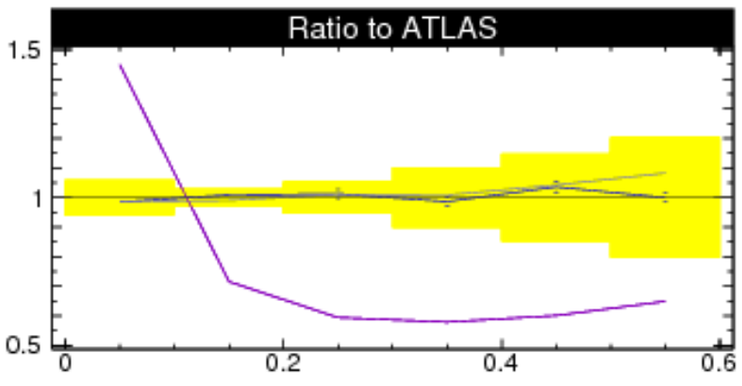


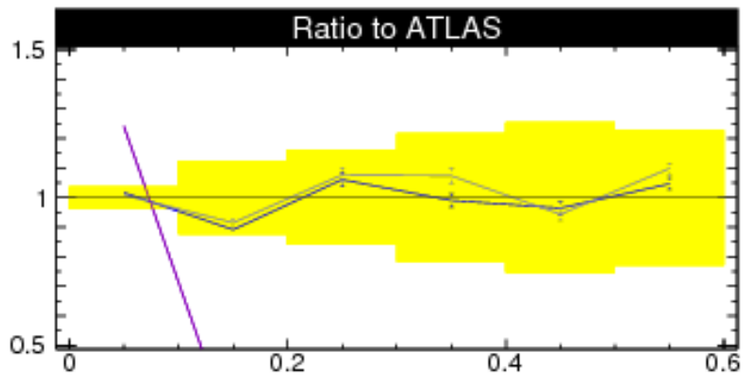
FSR scales : jet shapes

- **Measuring the flow of particles in the jets has obvious sensitivity to FSR**
  - split by kinematics and flavour, but also check different cone sizes (more in M.Voutilainen's talk)
  - several new variables since ATLAS result back in 2010
  - pileup may be challenging for the inclusion of neutrals⇒ fragmentation can be factorised for b's







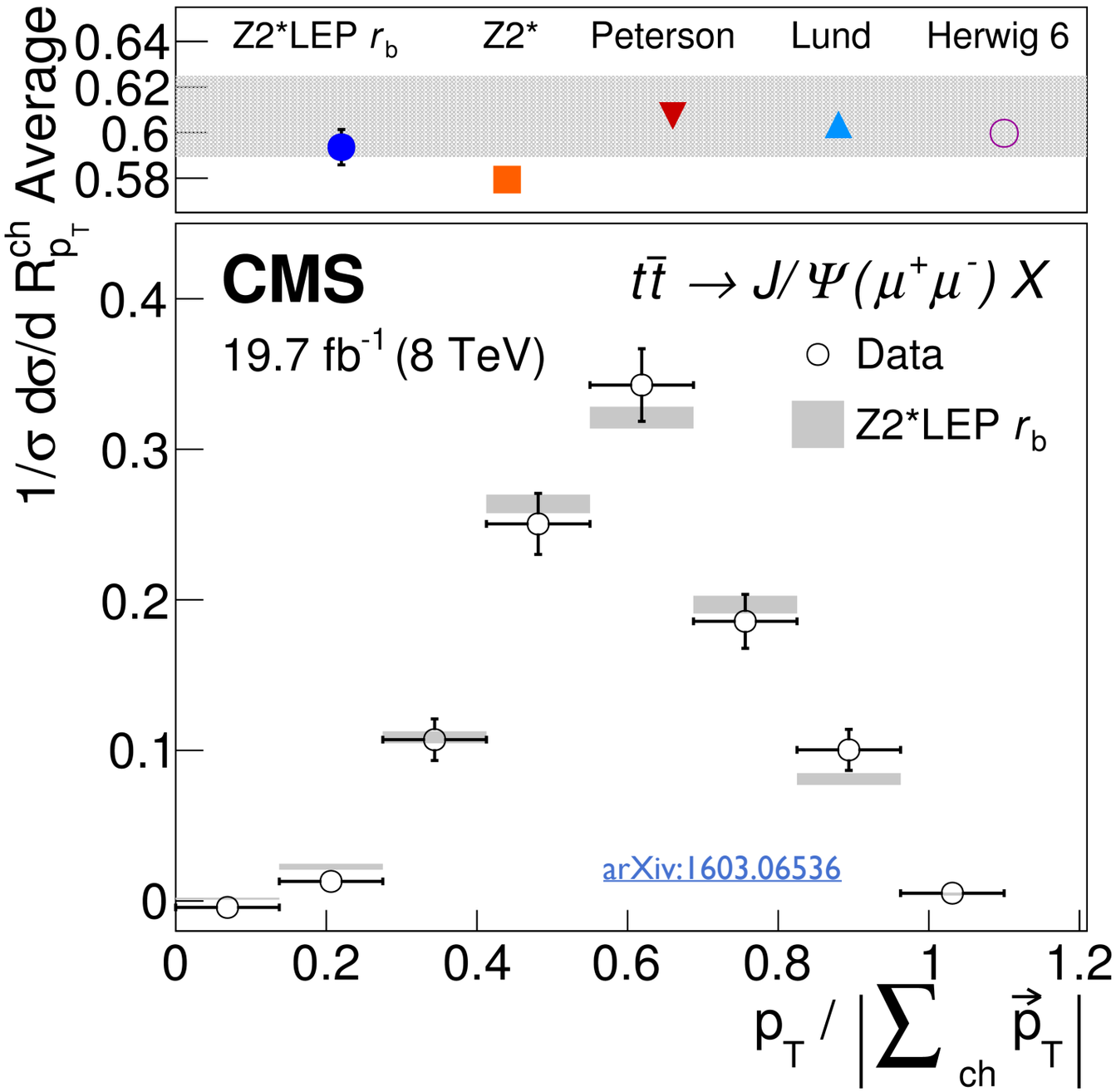


b-fragmentation

- **In practice, most top measurements sensitive to $\langle p_T(B)/p_T(b) \rangle$ only**

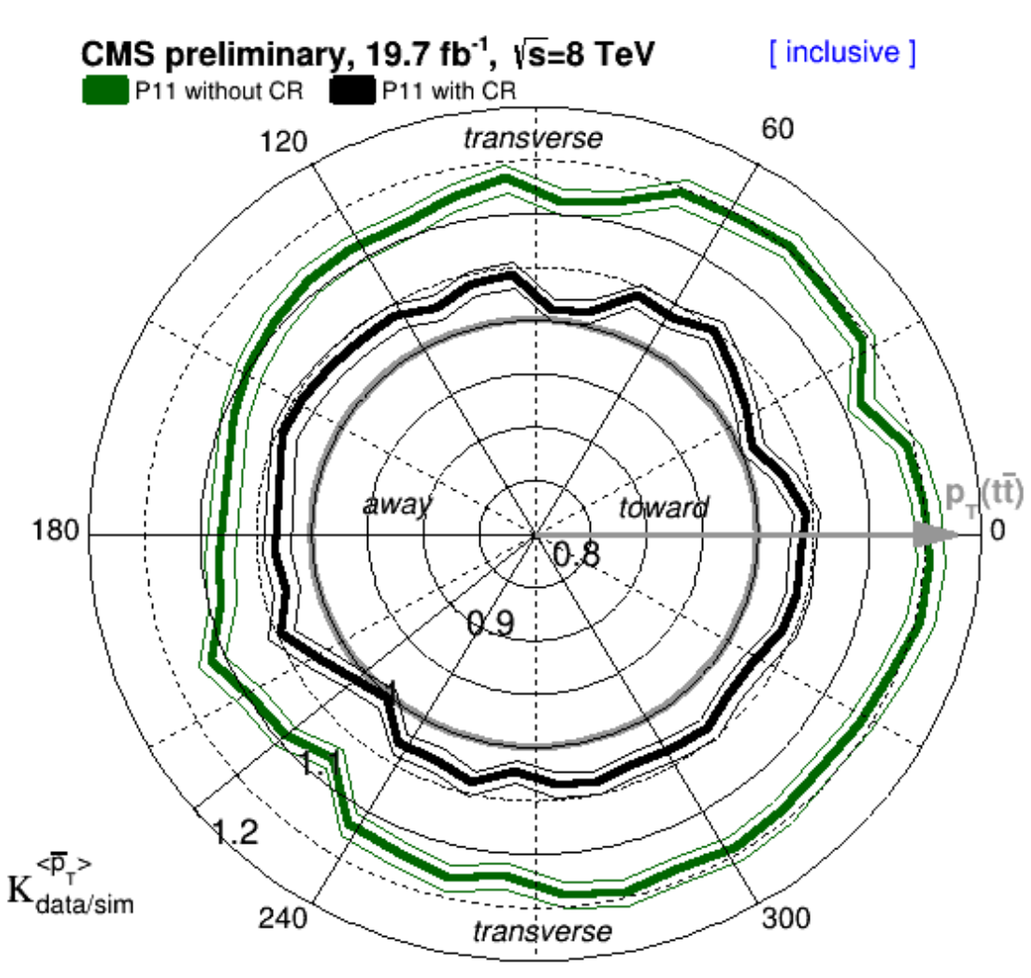
- **Use charmed mesons in tagged jets**
  - resonant peaks clearly separate signal
  - measure kinematics of the meson

- **Ratios to jet kinematics**
  - tracks only = PU resilient, high resolution
  - cancel several experimental systematics
  - preserve high resolution
  - sensitive to the fragmentation function

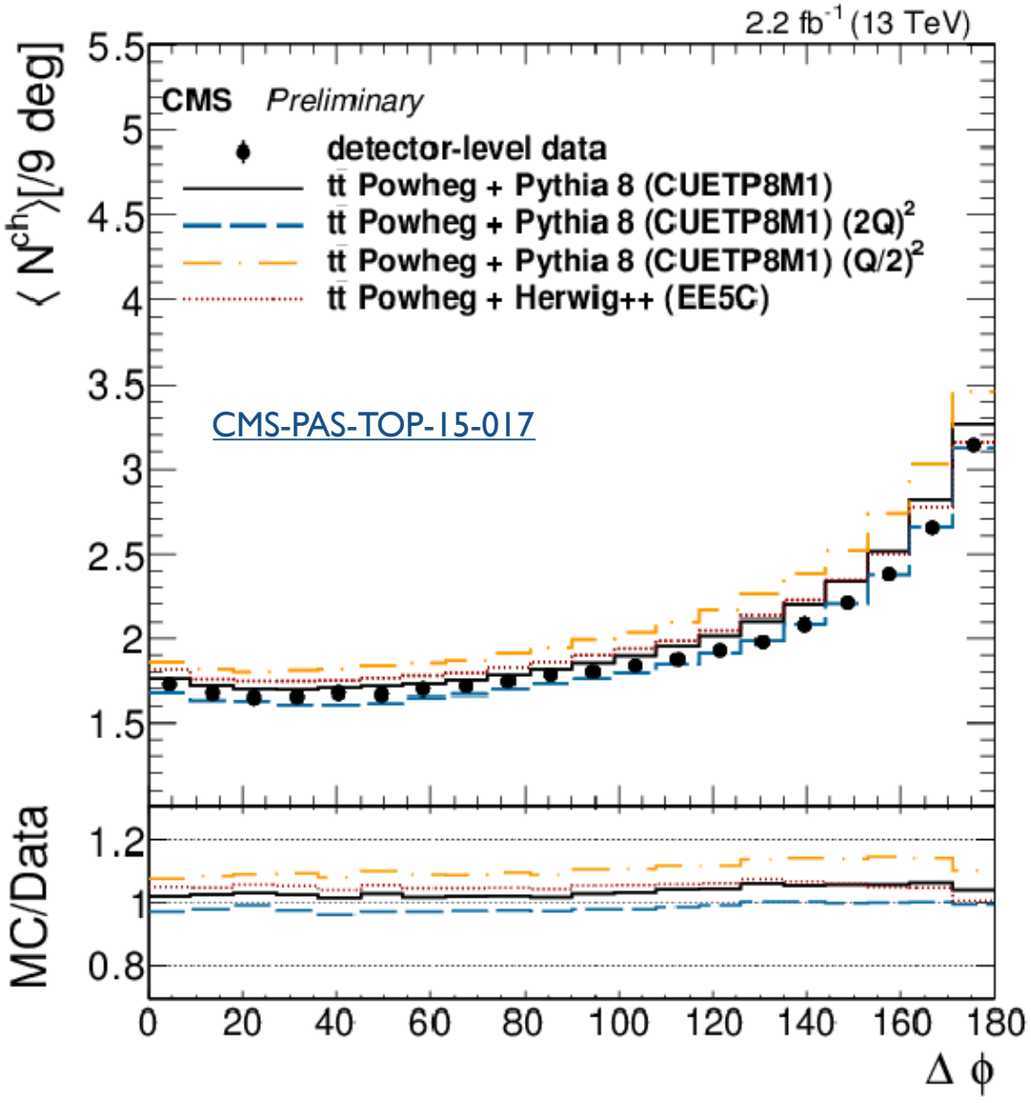


Underlying event: MPI and CR

- **Non-perturbative QCD effects can be probed in-situ**
  - so far no evidence that Min. Bias-based tune does not apply to top energy scales
  - sensitivity to CR with low p_T systems
  - sensitivity to PS scale: α_S^{FSR} or α_S^{MPI}? the latter bound to the PDF choice in Pythia



CMS-PAS-TOP-13-007



CMS-PAS-TOP-15-017

Summary

Summary

- **It is incomplete to say our data is in general well described by MC/fixed-order**
  - it is not: no single prediction to rule them all, simultaneous $p_T(t)$ and M(tt) seem challenging
  - but current PS-dominated approach to systematics provides a generous blanket
- **Benefits from our current approach are clear**
  - we can pin-point where we need to improve
  - but studies in the TMG group are badly needed!
  - Powheg+Pythia8 must be complemented by repeating factorized approach with other setups
- **Challenges ahead**
  - MC production will hardly cope with increase in statistics in data if strategy doesn't change
  - if MC is the only possibility: delegate to weights, generator level only
  - using data is superior: need to guarantee compatibility of control region with application region