

Monte Carlo Event Generator Modeling and Validation with Jet Measurements in CMS

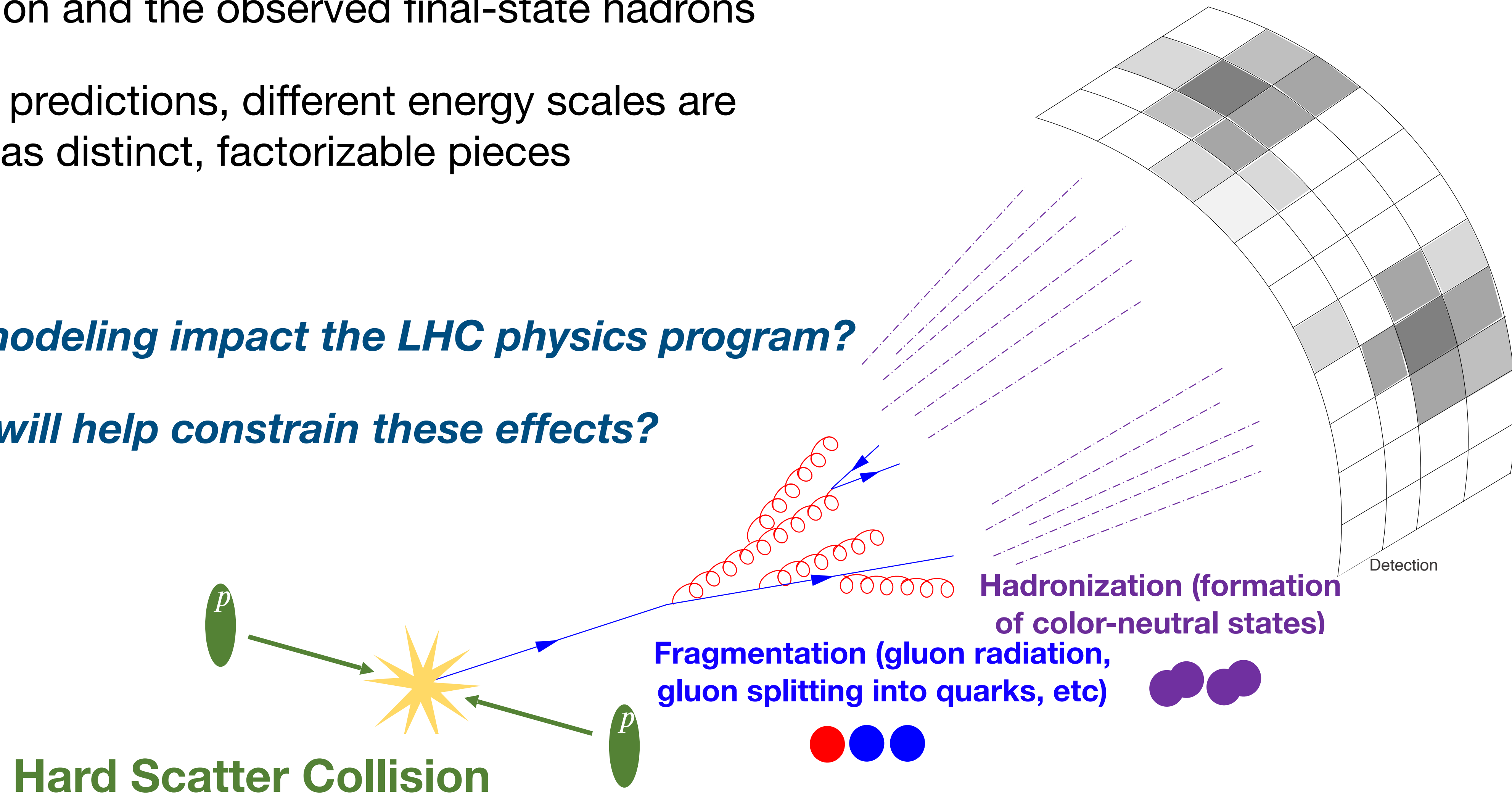
Jennifer Roloff, on behalf of the CMS collaboration
QCD@LHC, 2026



BROWN

introduction

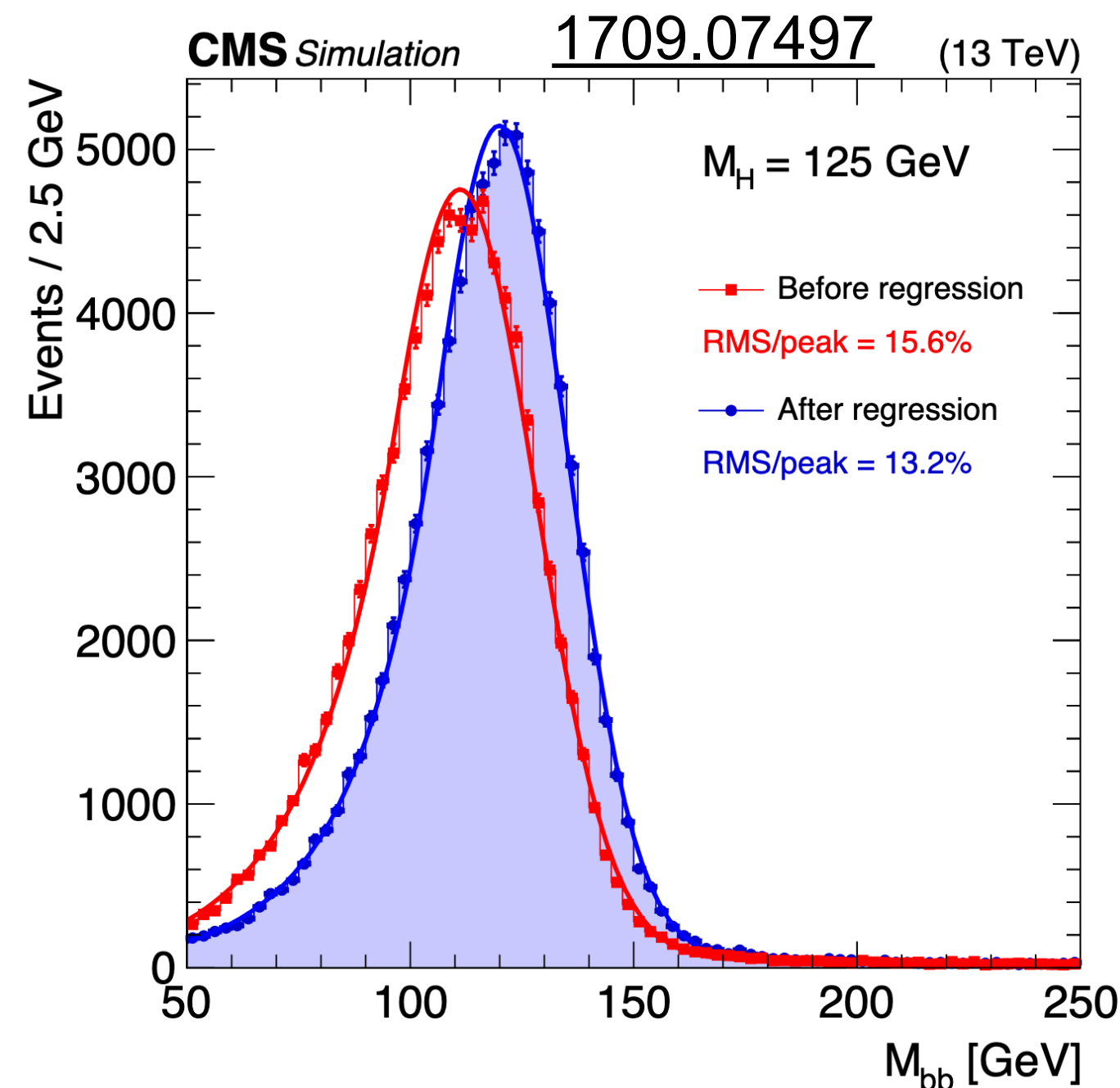
- ▶ Important to be able to describe the transition between the hard scatter collision and the observed final-state hadrons
- ▶ For Monte Carlo predictions, different energy scales are typically treated as distinct, factorizable pieces
- ▶ Today's talk:
 - ▶ *How does mismodeling impact the LHC physics program?*
 - ▶ *What analyses will help constrain these effects?*



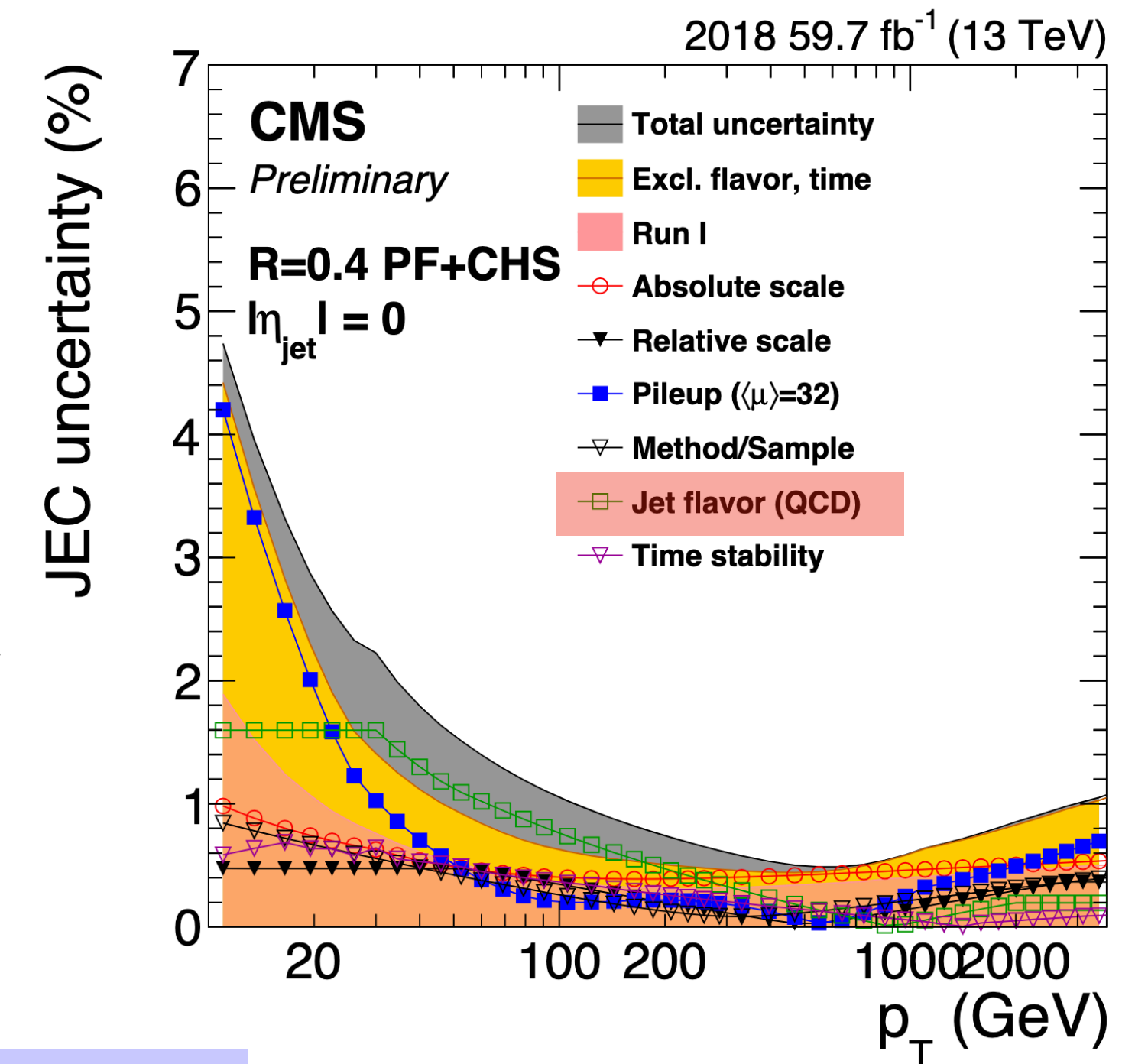
impact on calibrations

<https://cds.cern.ch/record/2715872>

- ▶ Any analysis involving jets includes *jet energy scale uncertainties*
 - ▶ Jet flavor uncertainty dominates between ~30-200 GeV
 - ▶ Based on calibration differences between quark and gluon jets
- ▶ In-situ uncertainties are often driven by modeling of additional jets in the event
- ▶ Analyses limited by *jet energy corrections* are typically limited by *jet modeling*



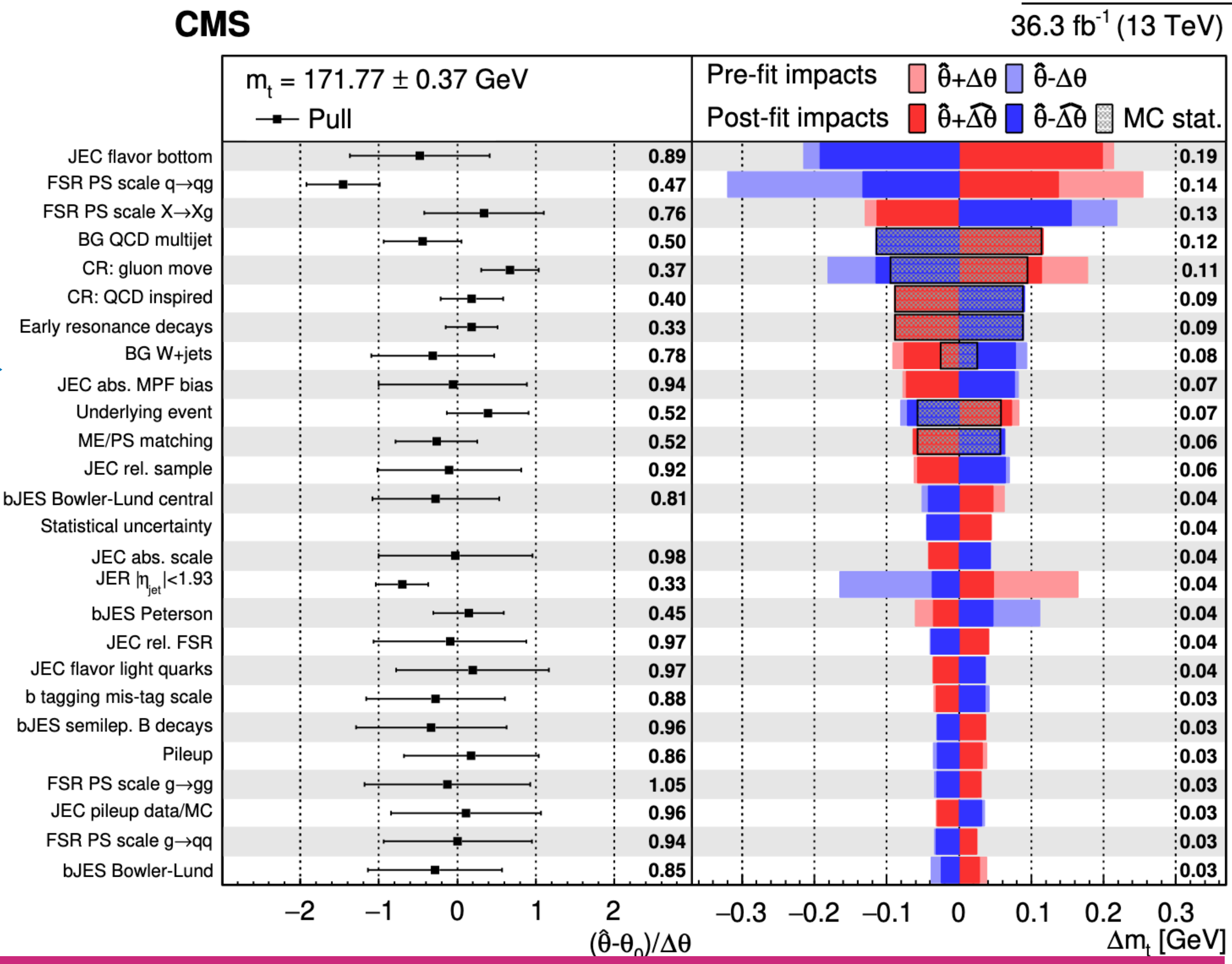
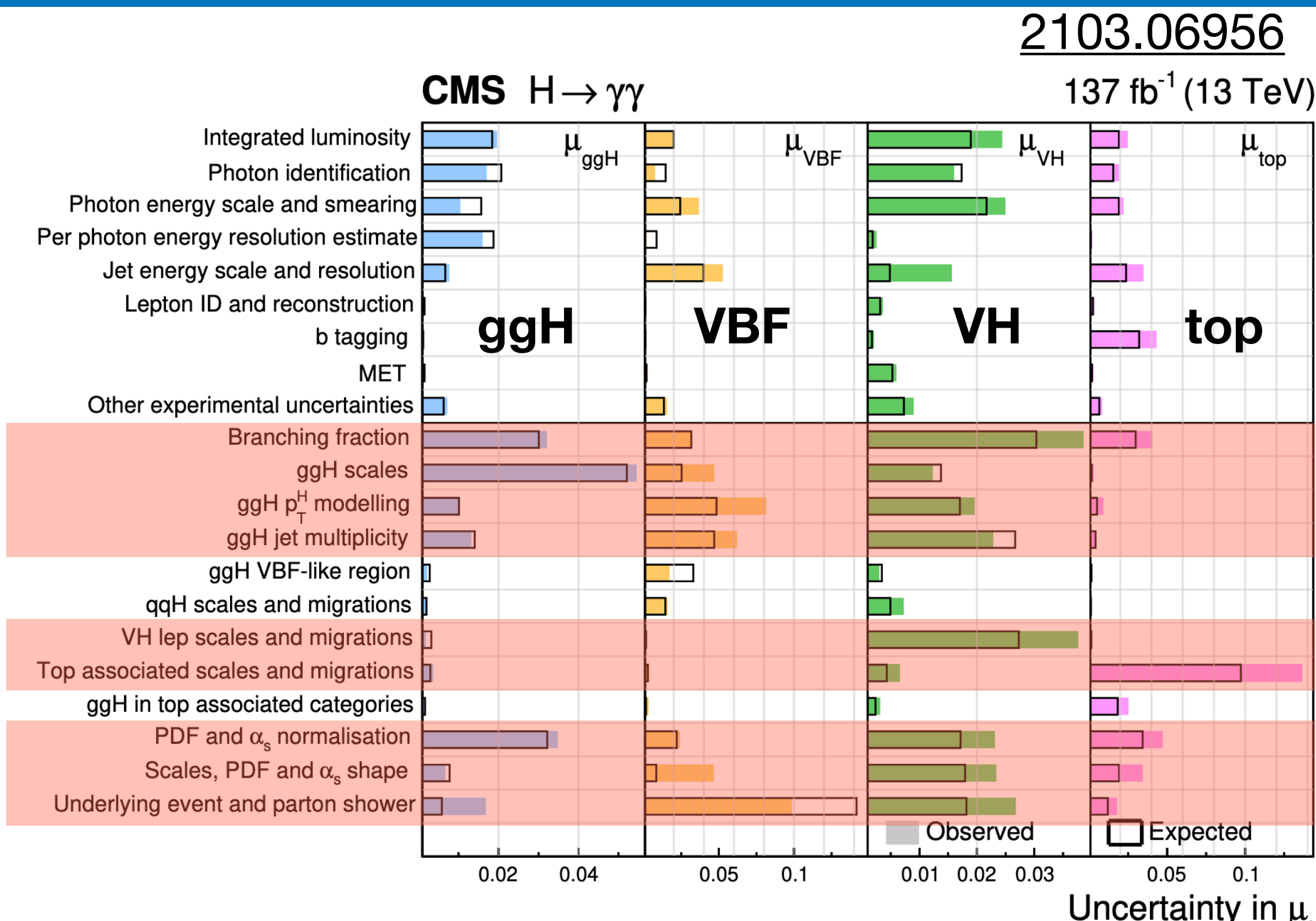
- ▶ Additional corrections to the b-jet energy scale are important to improving the resolution of the jets for $H \rightarrow b\bar{b}$
- ▶ Targeted corrections can often help reduce impact of modeling on results



See also Matt's talk from Tuesday!

impact on top and higgs

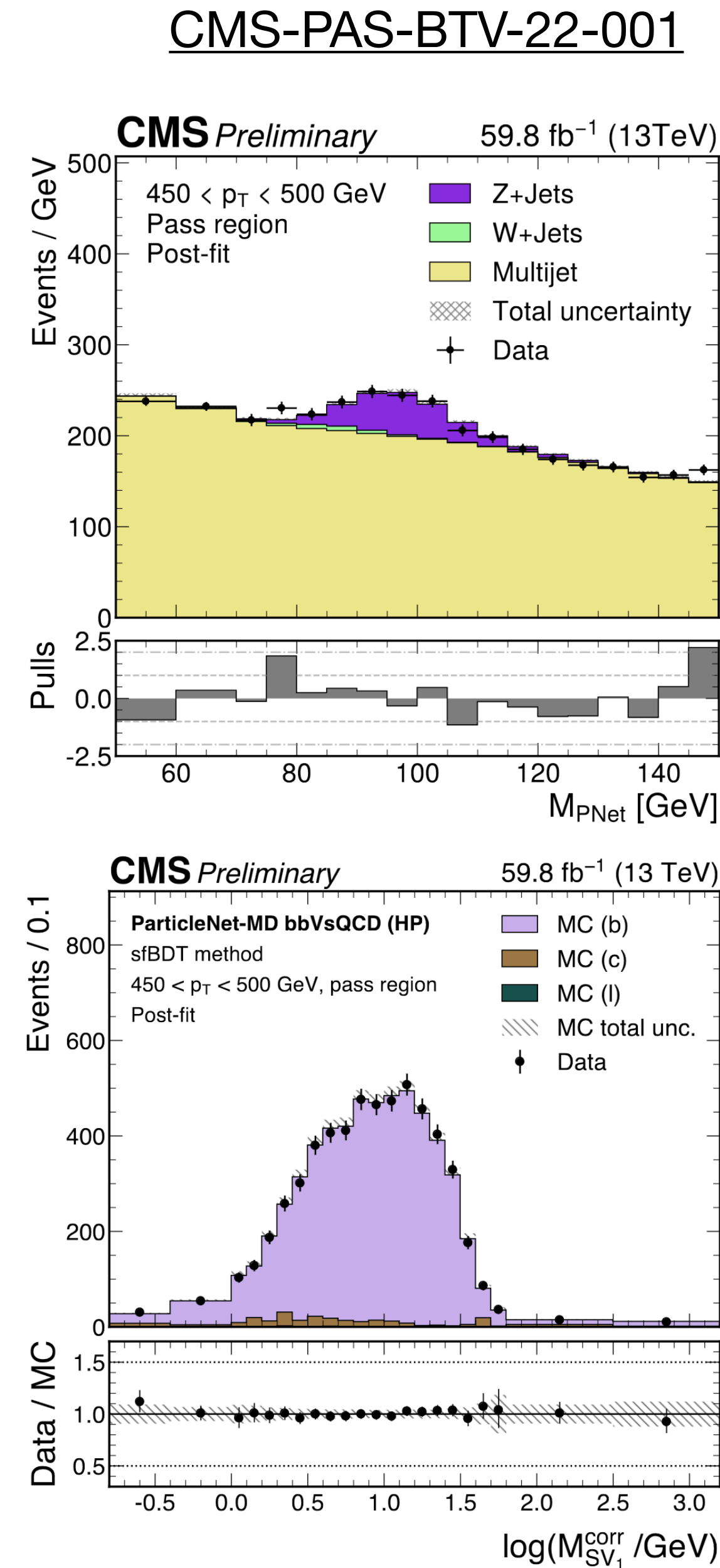
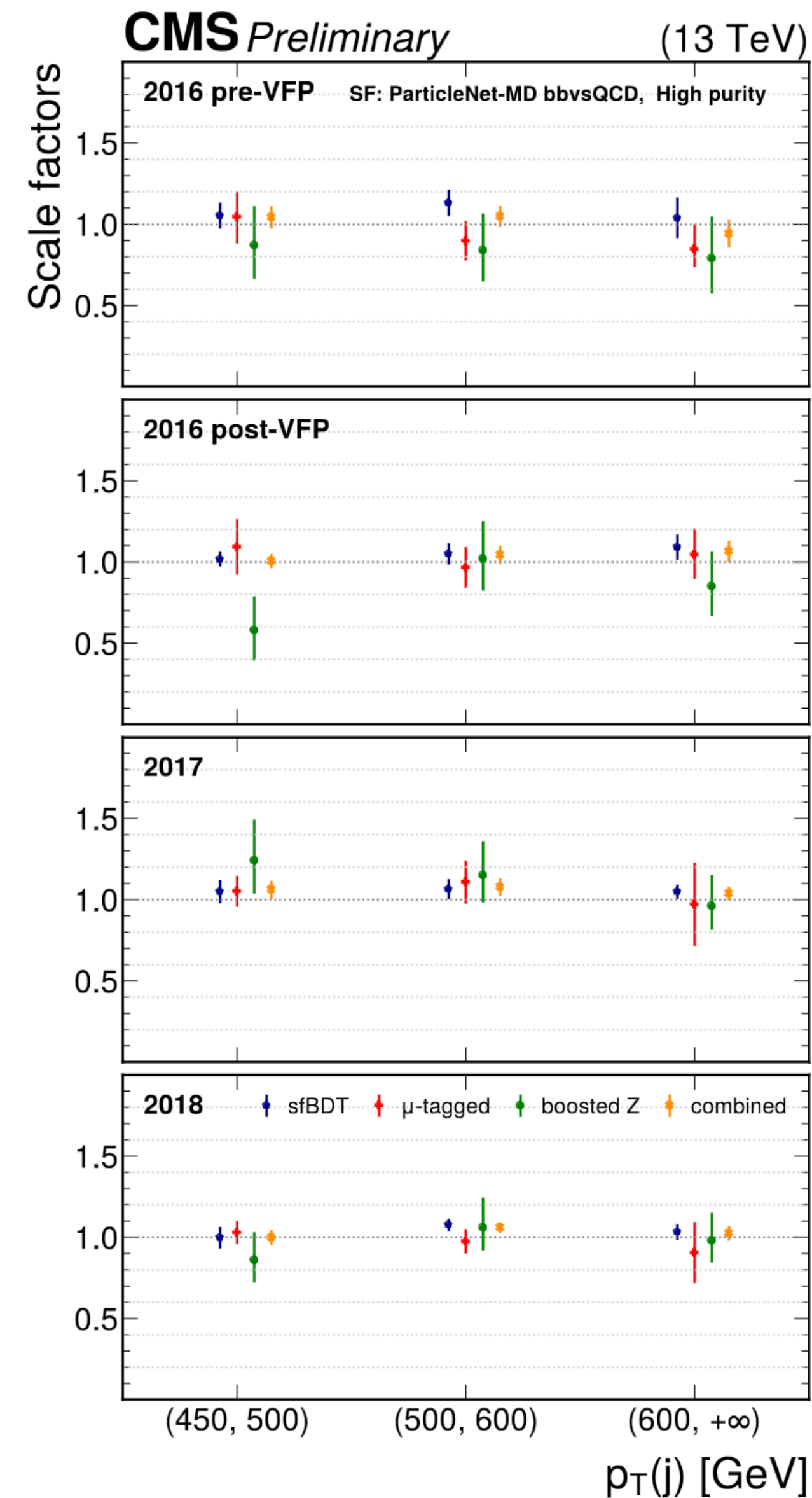
- Modeling uncertainties dominate direct top mass measurement
- Several of these uncertainties are constrained through auxiliary measurements that provide additional information



- ggH*: scale uncertainties, α_s , PDFs
- VBF**: underlying event, parton showers
- Top production**: top modeling

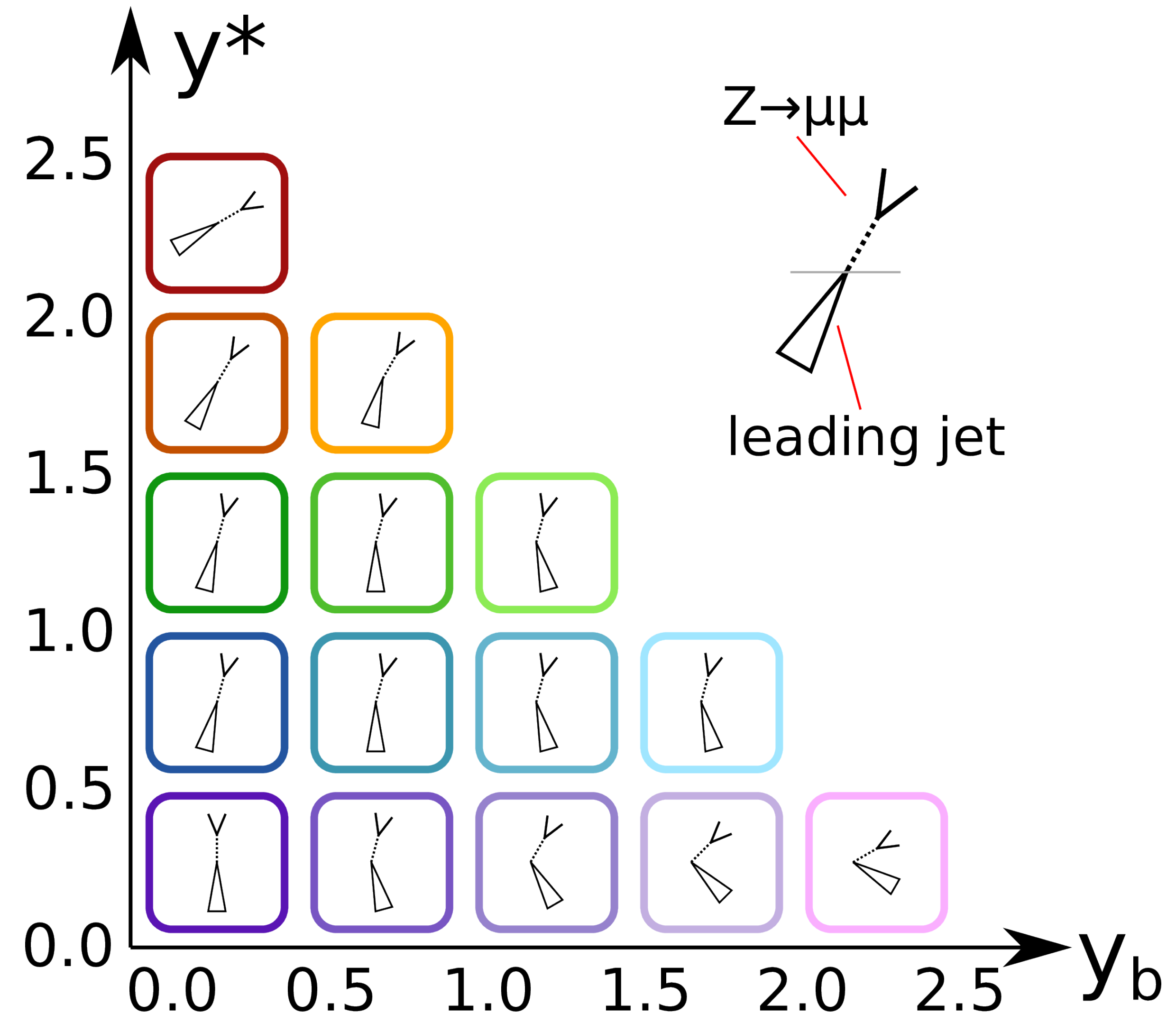
impact on taggers

- ▶ Scale factors describe difference in tagging performance between data and MC
 - ▶ Difficult to get scale factors for rare processes
- ▶ Boosted $H \rightarrow b\bar{b}$ combines scale factors from multiple sources ($g \rightarrow b\bar{b}$ and $Z \rightarrow b\bar{b}$)
 - ▶ Multiple calibrations provides extra validation from compatibility of methods
 - ▶ $Z \rightarrow b\bar{b}$ is currently statistically limited $\rightarrow$ expect better precision moving forward
 - ▶ For $g \rightarrow b\bar{b}$, using two methods
 - ▶ Tagging μ in jet (μ -tagged)
 - ▶ Using BDT to select $g \rightarrow b\bar{b}$ events (sfBDT)
 - ▶ BDT trained on N-subjettiness, subjet properties, and secondary vertex information



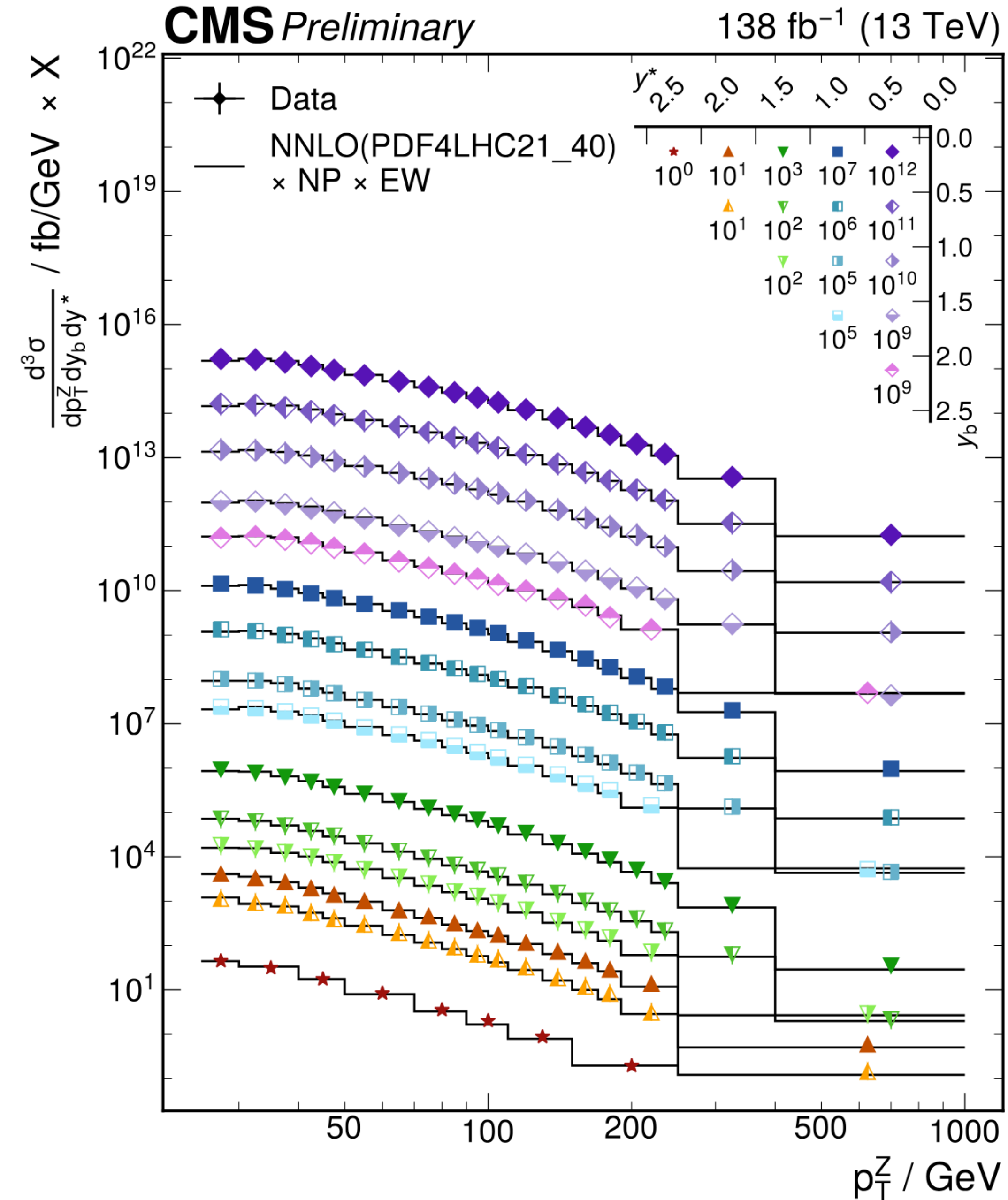
***Modeling remains an important challenge
for improving the LHC physics program!***

- $Z + \text{jets}$ events provide a clean probe of PDF effects
- Can use the rapidity separation (y^*) and boost (y_b) to improve sensitivity to PDFs
- *Similar to what is often done for dijet cross-section measurements*
- $y^* = \frac{1}{2} |y_Z - y_{\text{jet}}|$
- $y_b = \frac{1}{2} |y_Z + y_{\text{jet}}|$
- Use muon channel to get clean Z reconstruction
- Dominant backgrounds from $t\bar{t}$ and single- t production only contribute $\sim 1\%$ $\rightarrow$ very pure selection!



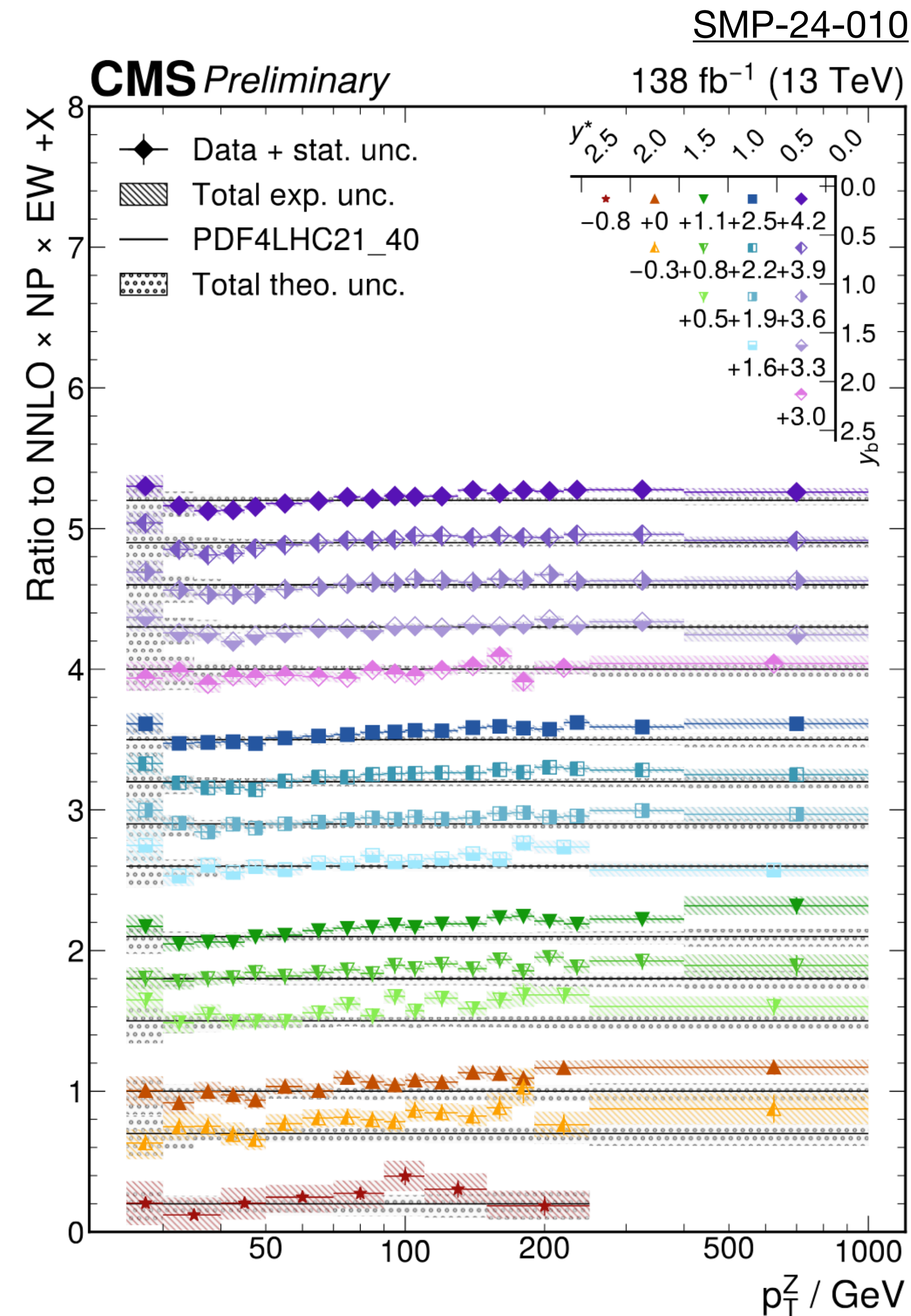
z+jets

- ▶ Comparing against fixed order theory prediction
- ▶ $\sigma_{\text{QCD+EW}}^{\text{NNLO,nLO}} = \sigma^{\text{LO}} + \delta\sigma_{\text{QCD}}^{\text{NLO}} + \delta\sigma_{\text{QCD}}^{\text{NNLO}} + \delta\sigma_{\text{EW}}^{\text{nLO}}$
- ▶ Impressive agreement between data and NNLO theory predictions across a wide range of boosts, p_T^Z , and rapidity separation
- ▶ *Predictions scaled to enable visualization on a single plot*
- ▶ Spans many orders of magnitude for the cross-section



Z+jets

- ▶ $\sigma_{\text{QCD+EW}}^{\text{NNLO,nLO}} = \sigma^{\text{LO}} + \delta\sigma_{\text{QCD}}^{\text{NLO}} + \delta\sigma_{\text{QCD}}^{\text{NNLO}} + \delta\sigma_{\text{EW}}^{\text{nLO}}$
- ▶ Excellent experimental precision across full measurement
- ▶ NNLO predictions tend to overestimate cross-section at high p_T^Z
 - ▶ nLO EW corrections tend to overestimate full NLO predictions, which could explain this ([1705.04664](#))
- ▶ Slight underproduction at low p_T^Z , aside from the lowest bin
- ▶ Trends are fairly consistent across y_b , slight worsening in agreement for larger y^*



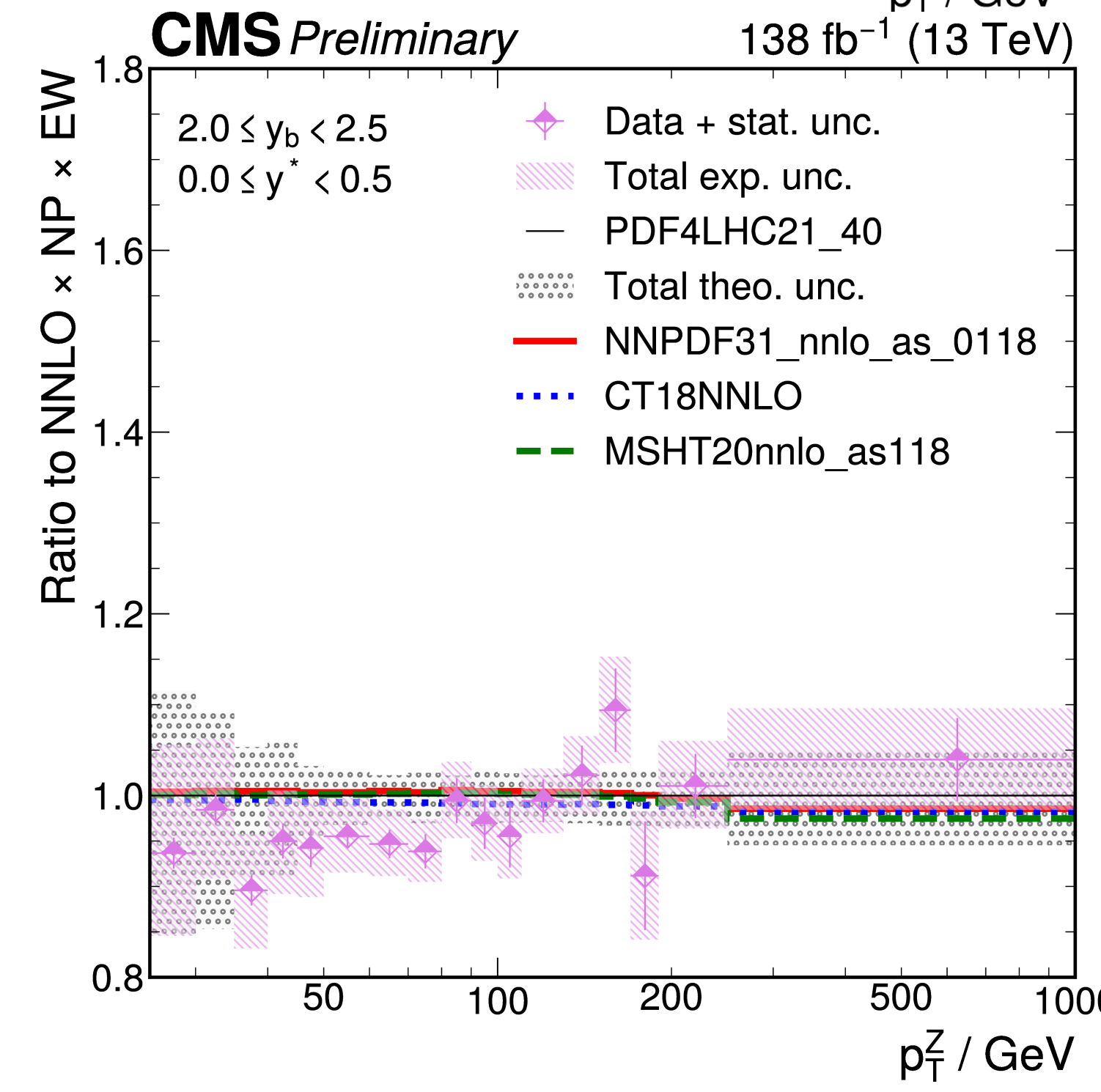
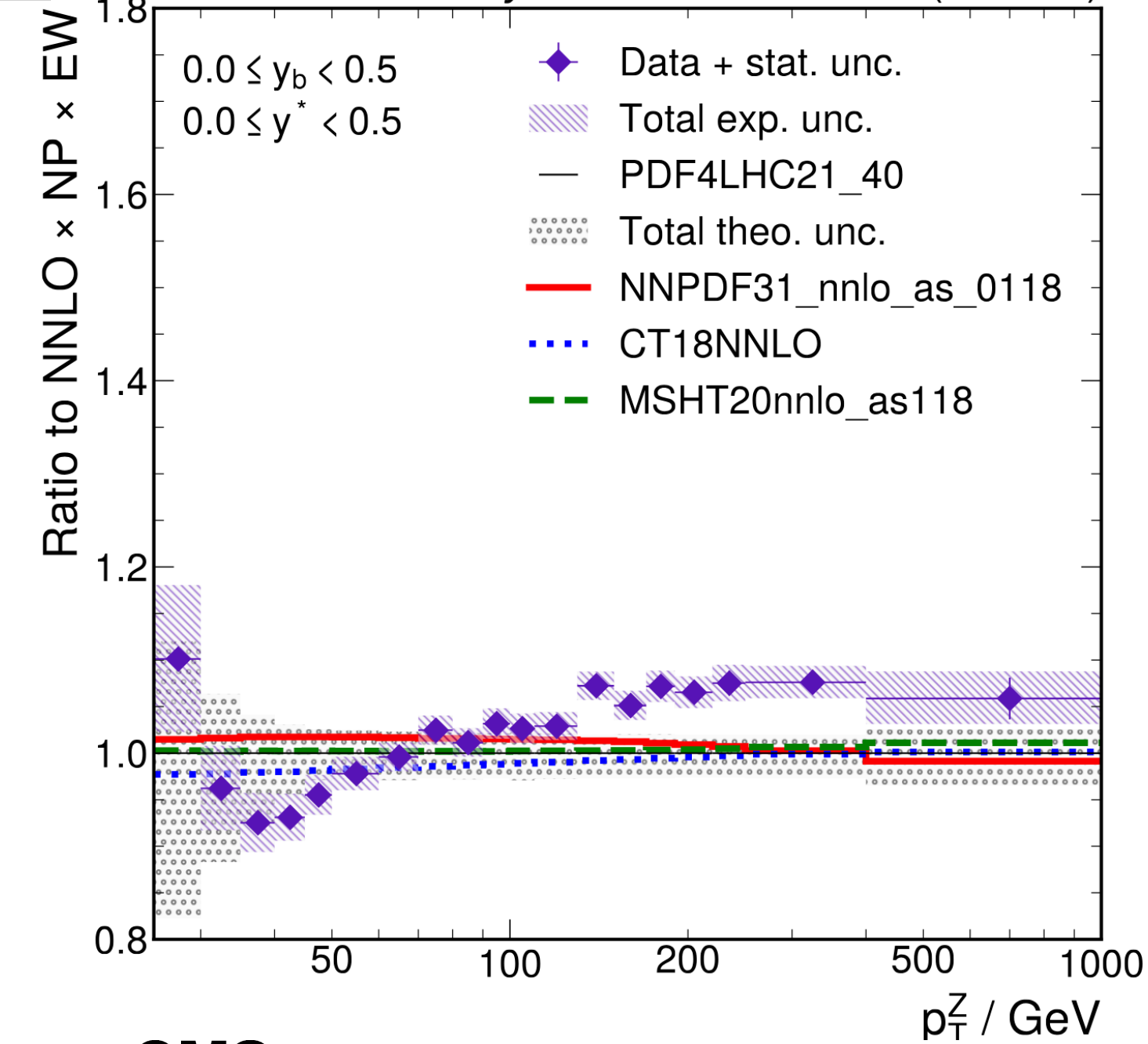
z+jets

- ▶ Relatively small differences between modern PDF sets, particularly for larger boosts
- ▶ *Expect this to be most relevant for high- x gluon constraints, based on previous fits*
- ▶ Small p_T^Z is impacted by resummation effects, and higher p_T^Z is impacted by electroweak effects → expect the intermediate p_T range to be the most powerful (See CTEQ discussion)

SMP-24-010

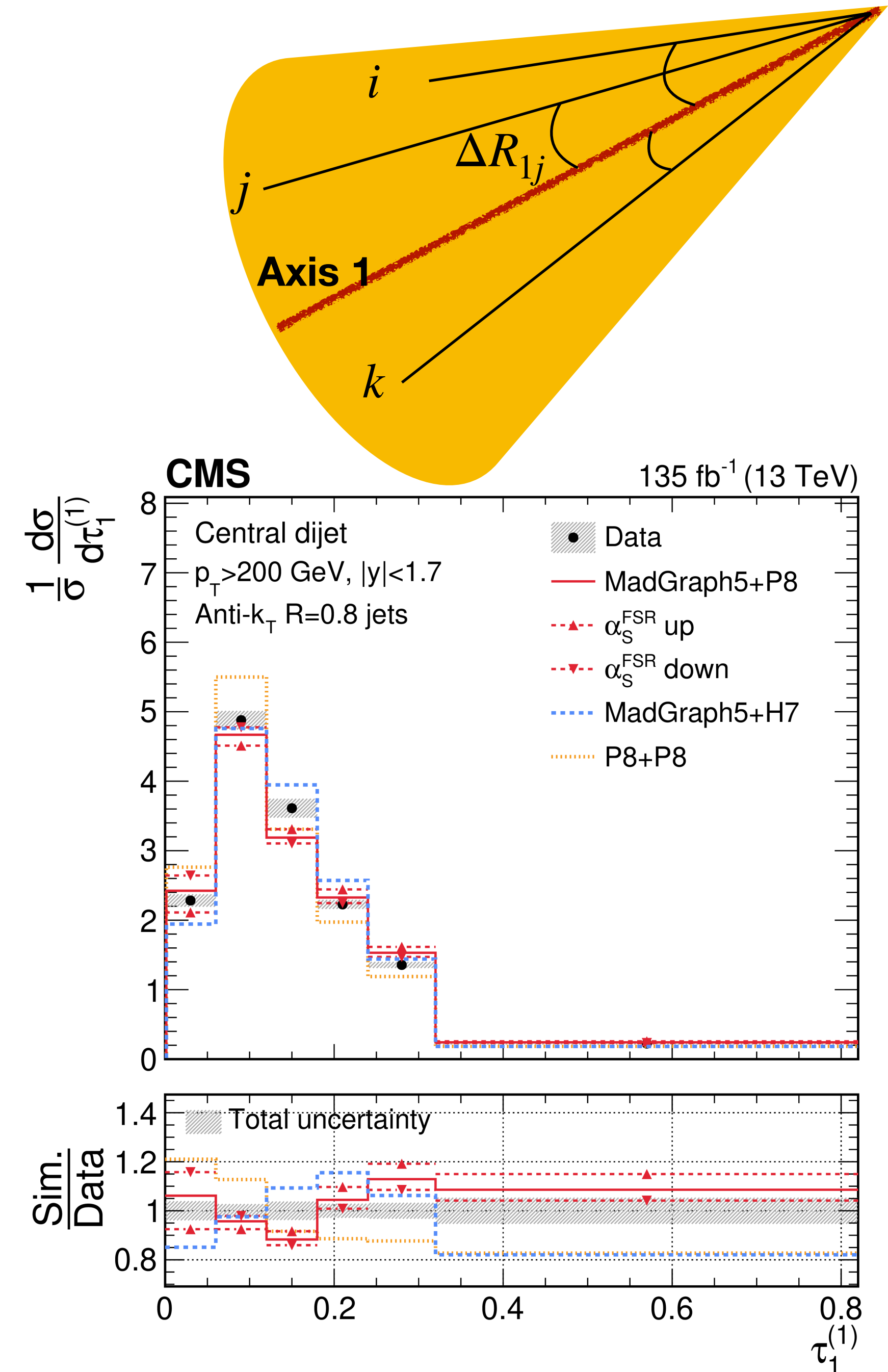
CMS Preliminary

138 fb⁻¹ (13 TeV)



$N_{\text{subjettiness}}$

- ▶ $N_{\text{subjettiness}}$ encodes information about the *pronginess* of a jet
- ▶
$$\tau_N^\beta = \frac{1}{d_0} \sum_i p_{T_i} \min(\Delta R_{1i}^\beta, \Delta R_{2i}^\beta, \dots, \Delta R_{Ni}^\beta)$$
- ▶ Lower values mean more consistent with N -prong substructure
- ▶ Used for early W and top taggers, now for $g \rightarrow b\bar{b}$ scale factor determinations
- ▶ Well-chosen set of $(3M-4)$ τ_N^β fully describes space of M -body jets
- ▶ $(M-1)$ p_T fractions and $(2M-3)$ angles ([1704.08249](#))
- ▶ Use E -scheme to ensure spanning of space
- ▶ Simultaneously unfolded 24 observables in 3 different event topologies (*overcomplete basis*)

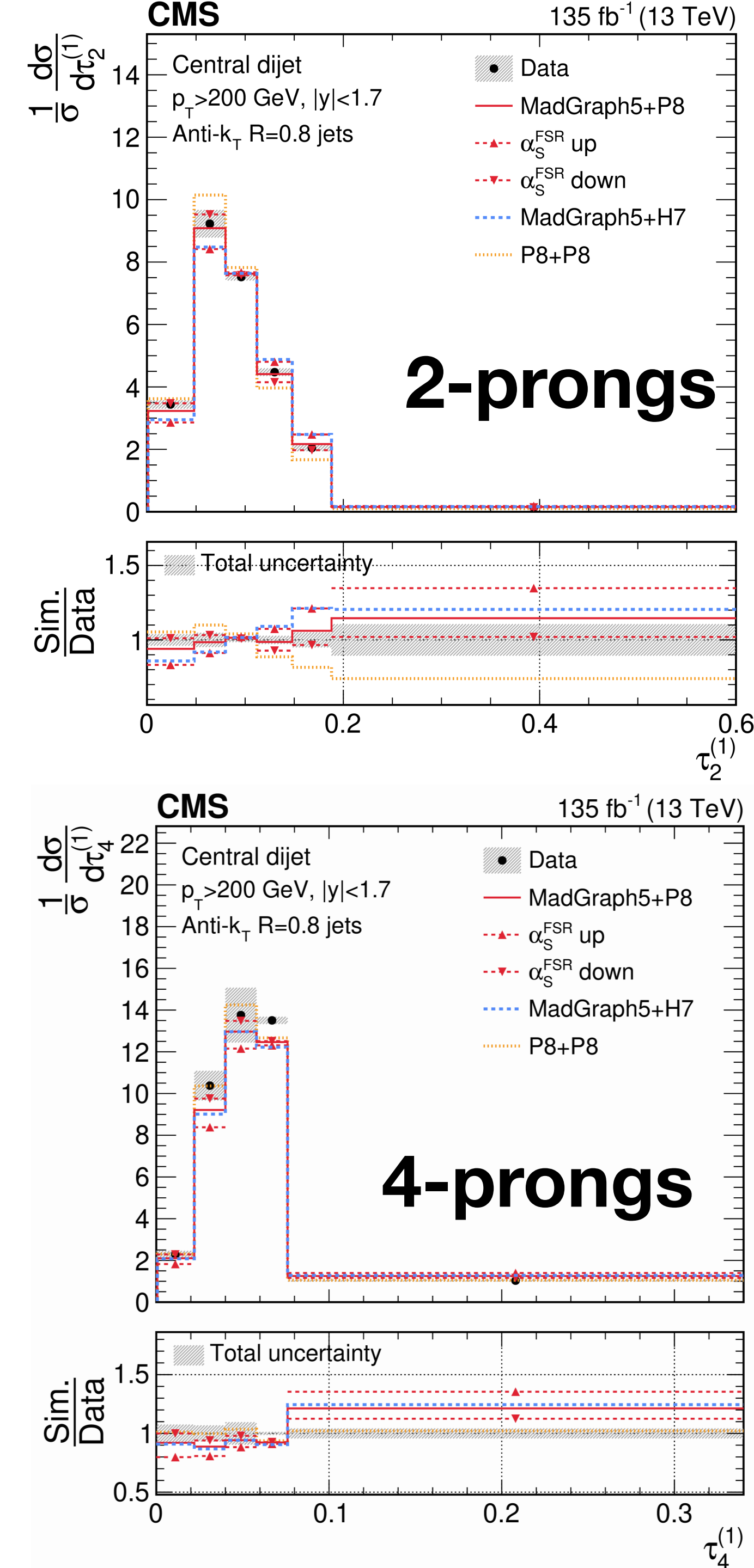


$N_{\text{subjettiness}}$

- ▶ $N_{\text{subjettiness}}$ encodes information about the *pronginess* of a jet

$$\tau_N^\beta = \frac{1}{d_0} \sum_i p_{T_i} \min(\Delta R_{1i}^\beta, \Delta R_{2i}^\beta, \dots, \Delta R_{Ni}^\beta)$$

- ▶ Worse experimental precision for higher N $\rightarrow$ fewer bins
- ▶ Systematics dominated by modeling uncertainty
- ▶ Modeling depends on matrix element (**MadGraph** vs **P8**), α_s (**dashed** vs. **solid**), and PS/hadronization (**P8** vs **H7**)
- ▶ For $\tau_2^{(1)}$, low values depend more on PS/hadronization, high values depend on ME calculation
 - ▶ *i.e. collinear effects are especially impacted by PS and hadronization, as expected*
- ▶ 4-prong substructure dominated by α_s effects



the lund jet plane

2312.16343

- ▶ A jet may be approximated as soft emissions around a hard core which represents the originating quark or gluon
- ▶ Emissions may be characterized by
 - k_t = transverse momentum of emission relative to the jet core
 - ΔR = angle of emission relative to the jet core

1. Jet Finding:

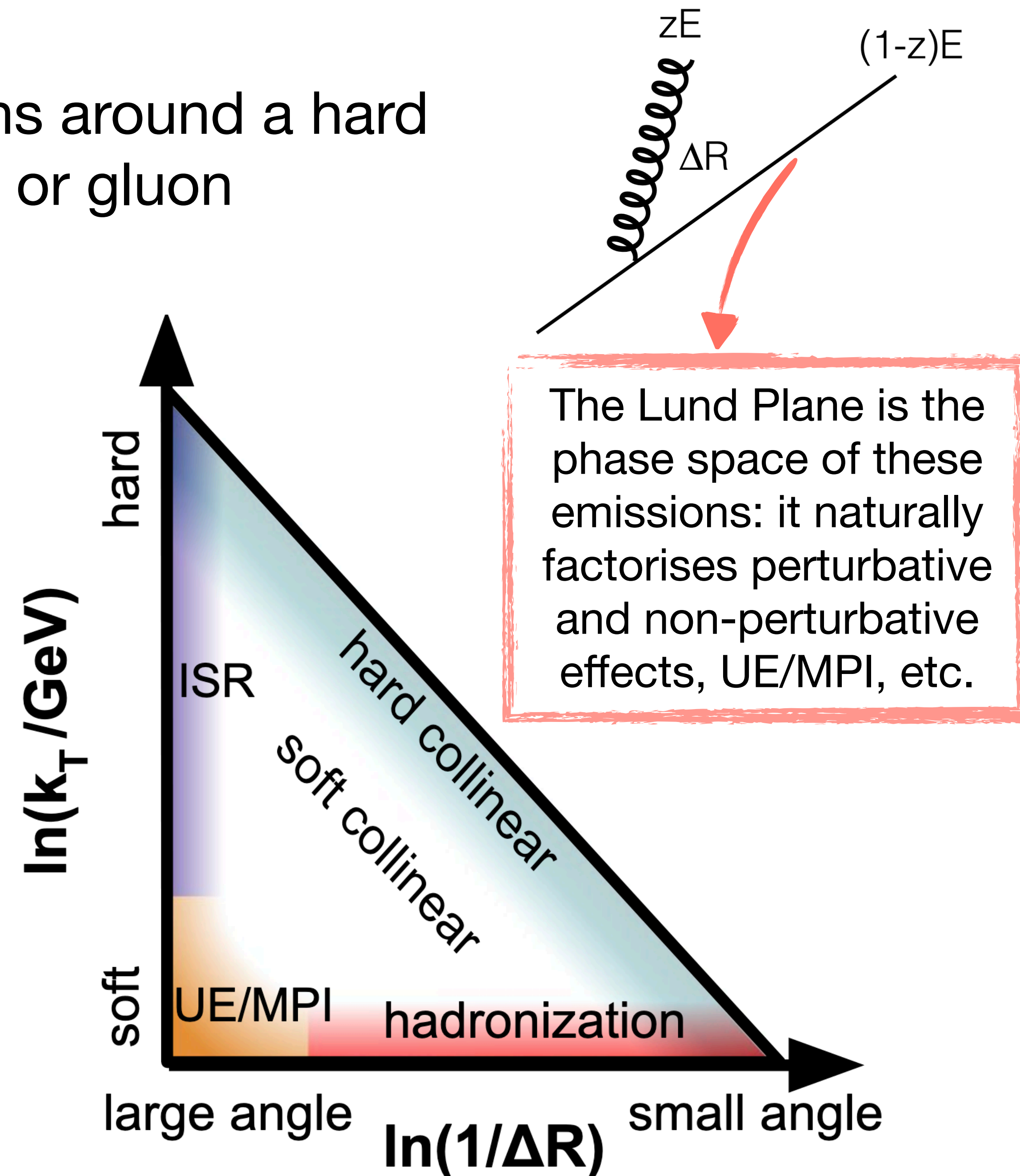
Cluster jets using your favorite jet algorithm

2. C/A Reclustering:

Combine closest pairs of **charged particles or tracks!**

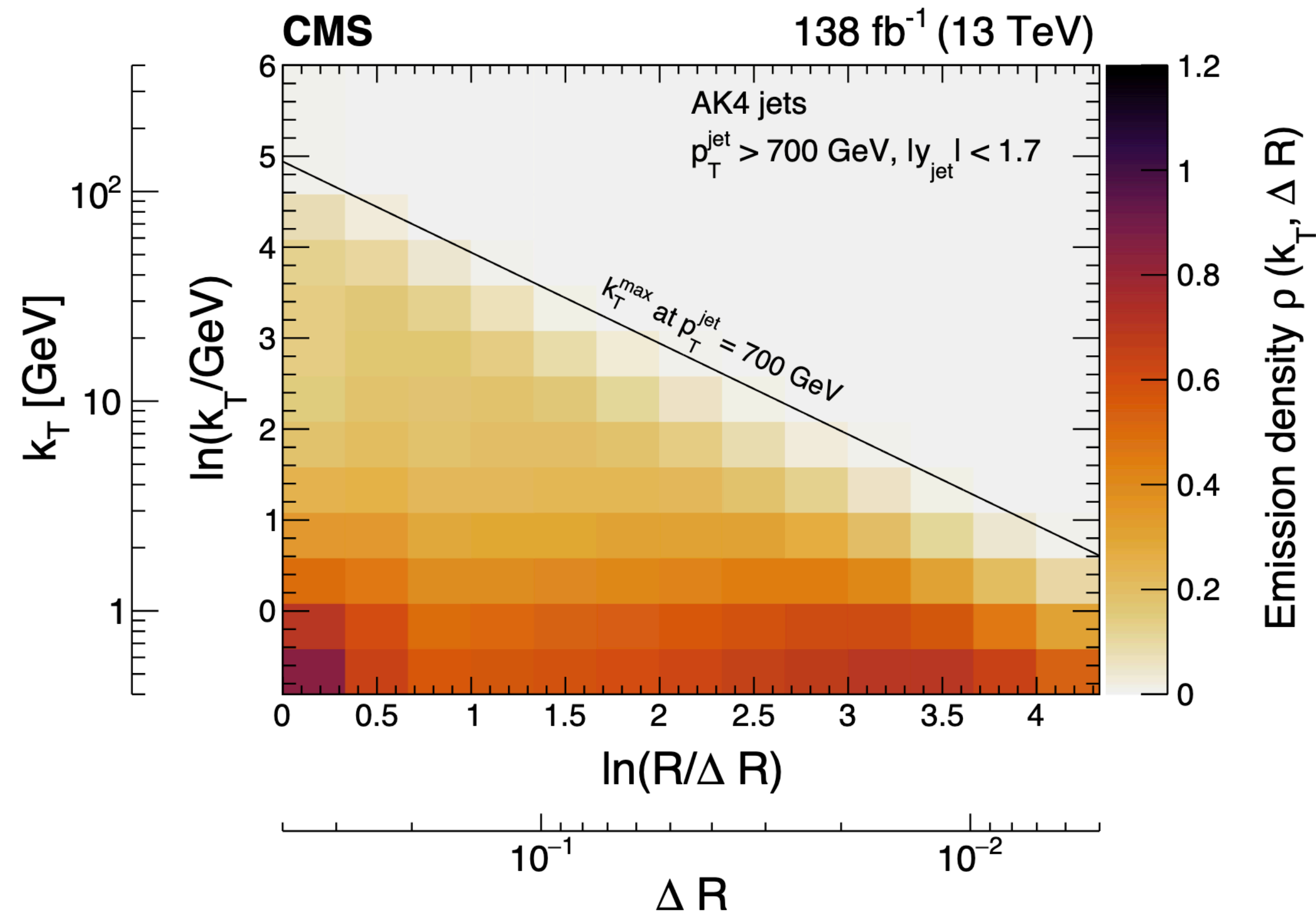
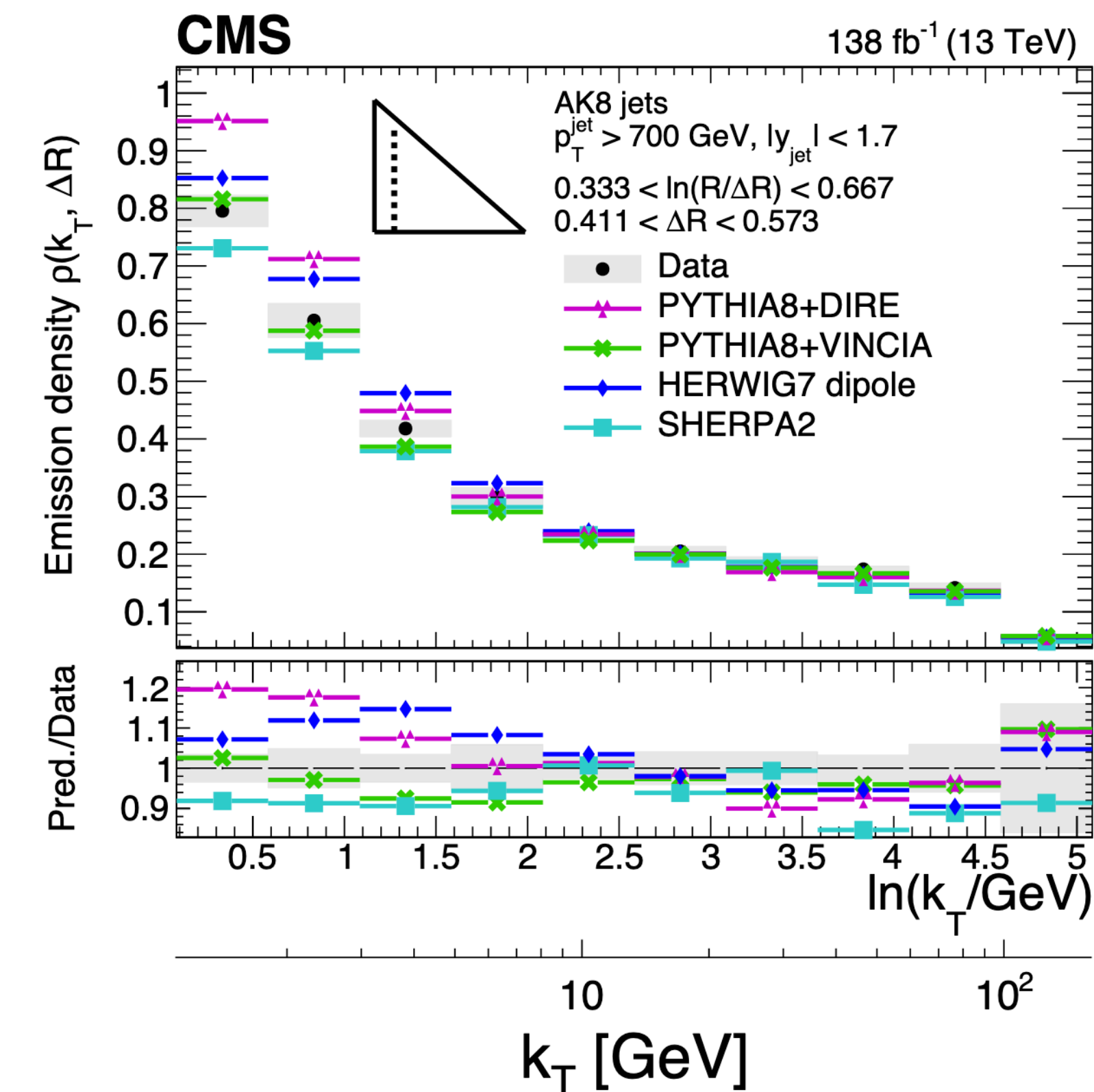
3. C/A Declustering:

Unwind, widest angles first. Each step is an **emission**, or, a point in the Lund Jet Plane!



the lund jet plane

- Measured the primary Lund jet plane for high p_T dijet events
- Higher jet p_T translates to a larger kinematic range
 - maximum k_t is determined by jet p_T



- Factorization of PS and hadronization effects seen by comparing MC predictions
- Largest parton shower differences observed in the higher k_t region where these effects are dominant

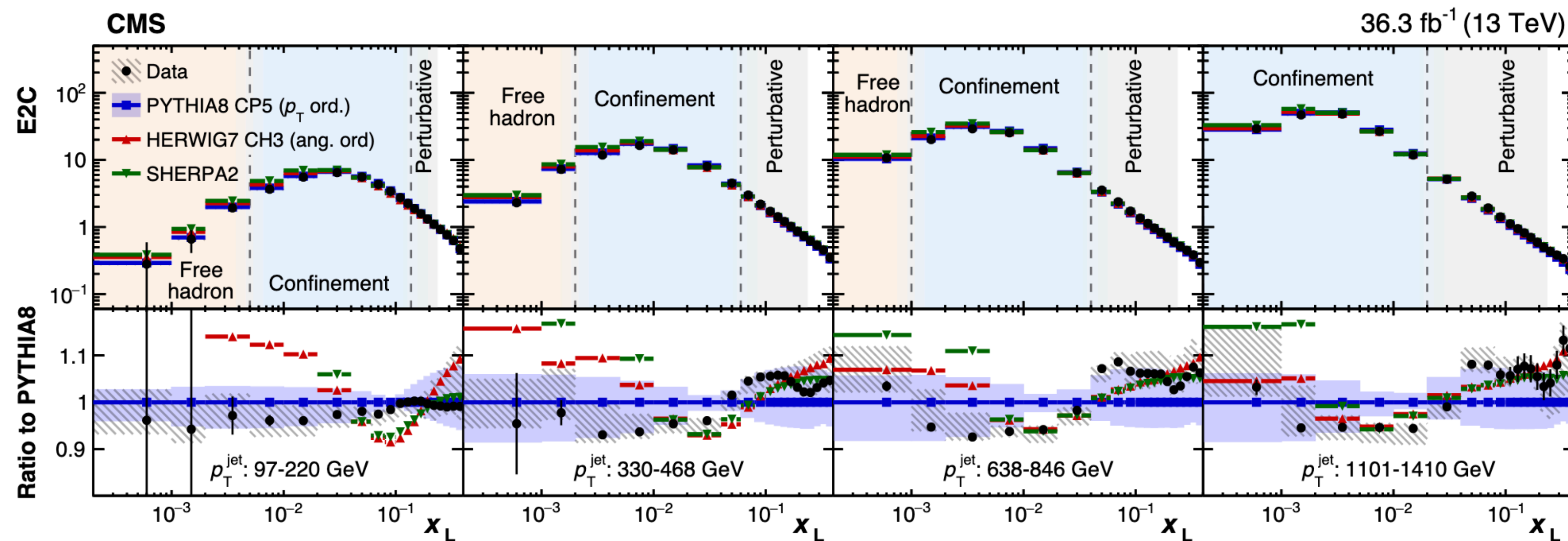
energy energy correlators

- Instead of constructing a clustering history, can use energy-weighting to study correlations

between particles:
$$\frac{d\sigma}{dx_L} = \int d\sigma \sum_{i,j \in \text{jet}} \frac{p_T^i p_T^j}{(\sum_k E_k)^2} \delta(x_L - x_{ij})$$

See also Wouter's talk from Tuesday!

- Naturally emphasizes the importance of high-energy contributions, inherently infrared- and collinear-safe
- CMS measurement shows significant disagreement between data and MC
- Provided a first extraction of α_s from jet substructure, with prospects for more precise extraction with experimental and theoretical advances

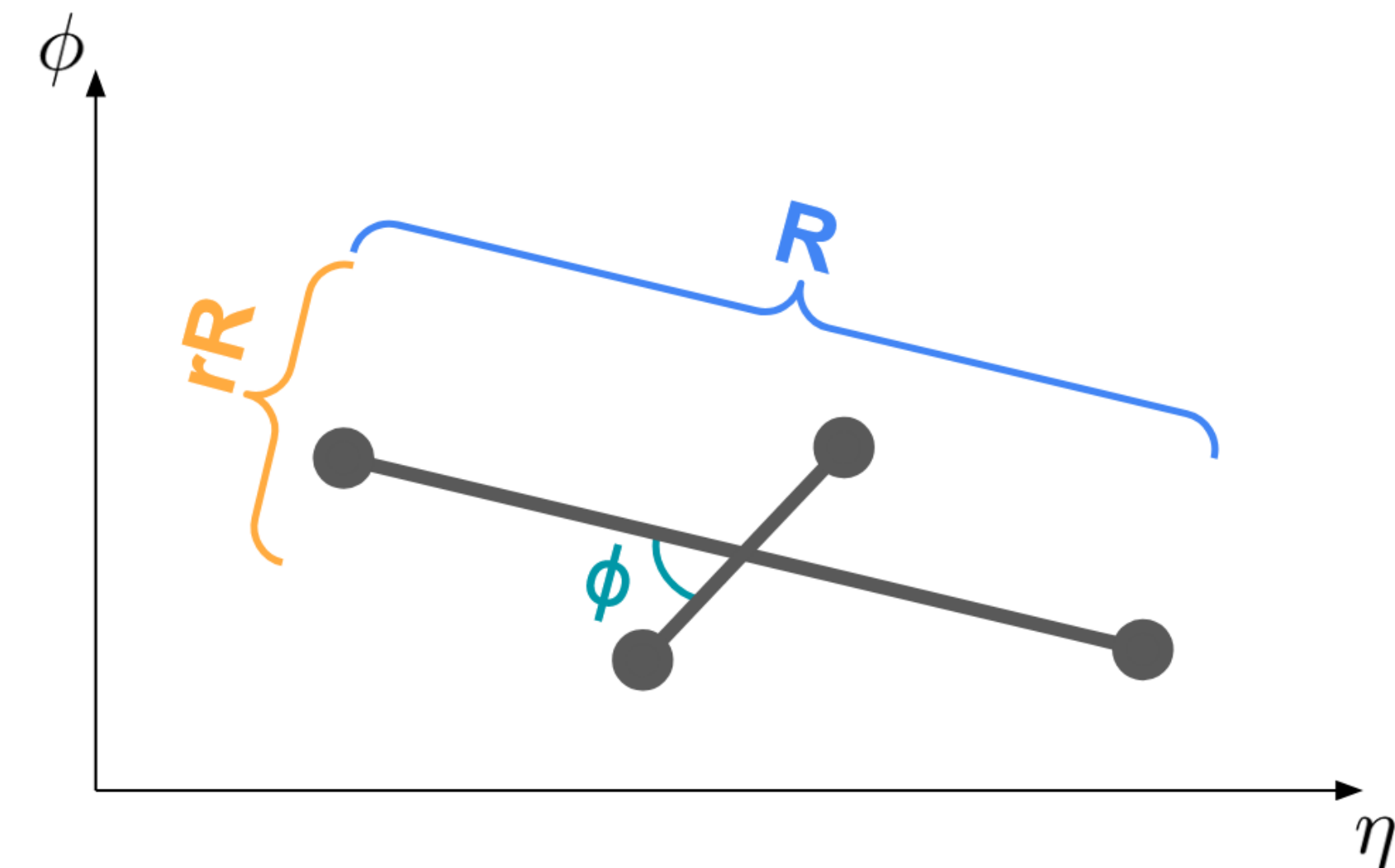


energy energy correlators

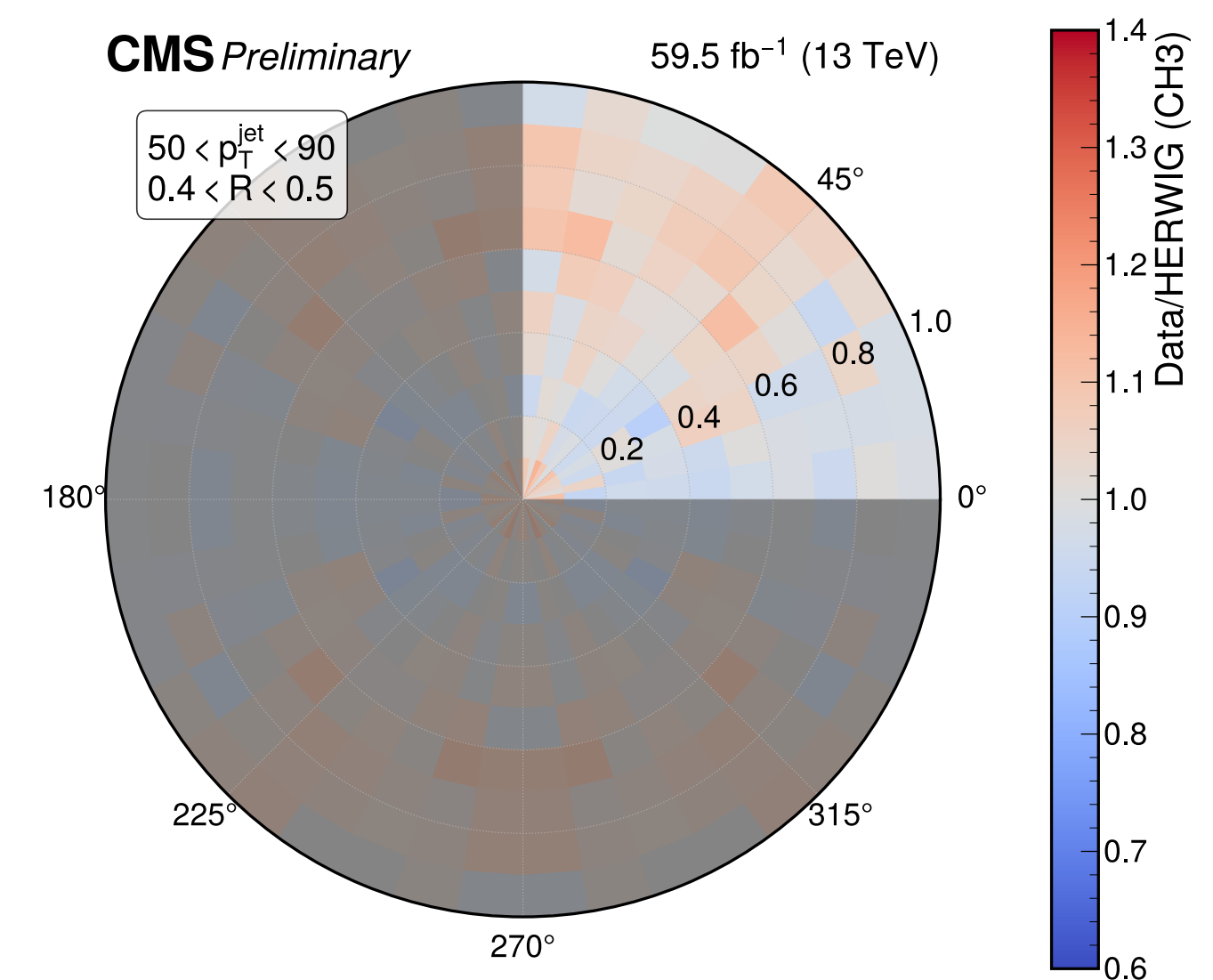
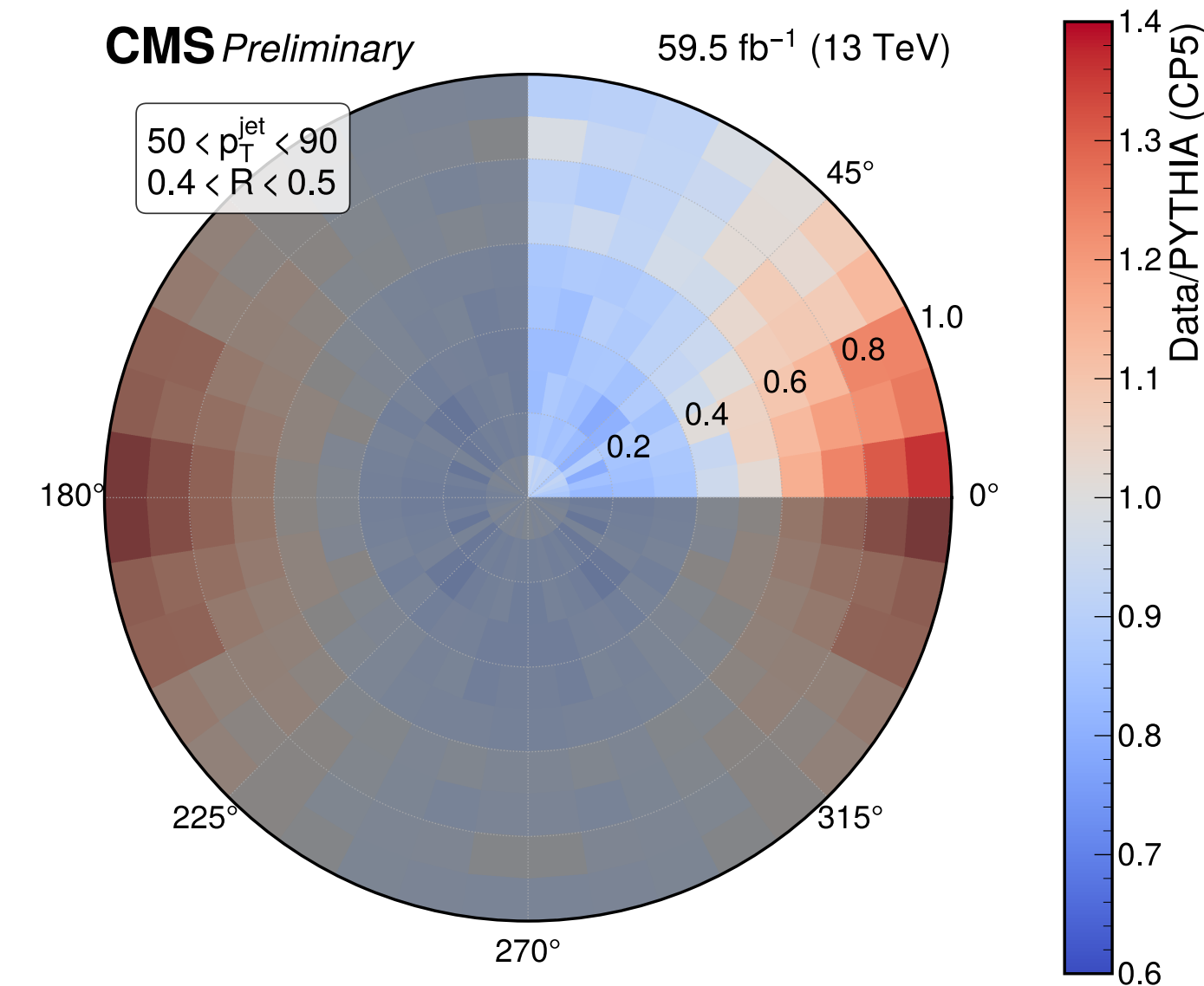
- Higher point energy-energy correlators encode complex correlations
- CMS has a preliminary measurement of fully resolved 4-point correlators

$$\frac{d\sigma^4}{d\vec{\Omega}} = \sum_{i,j,k,l \in \text{jet}} \frac{p_T^i p_T^j p_T^k p_T^l}{p_T^{(\text{jet}),4}} \delta(\vec{\Omega}_{ijkl} - \vec{\Omega})$$

- See Attila's talk for more details on this analysis!
- Hard to study high-dimensional phase-space → picked out a few configurations that are particularly interesting



- Sensitive to spin correlations, soft and collinear effects, higher-order splittings, and more
- Opposite behavior between Pythia and Herwig — more to be understood!



summary

- ▶ QCD modeling is a limiting factors in the precision of many important analyses
 - ▶ Need better constraints on all aspects of jet formation to improve the LHC physics program
 - ▶ *Relevant for top physics, Higgs physics (boosted and resolved), and much more*
- ▶ Jet measurements provide insight into different aspects of MC production
 - ▶ Jet substructure is powerful for characterizing jet formation and modeling
 - ▶ Many recent observables have good factorization properties that can target individual effects in parton shower or hadronization modeling
- ▶ More differential measurements will help pick out specific effects, both for tuning and tests of modeling
 - ▶ Need a wide variety of measurements to constrain the many different ways that jet modeling enters analyses