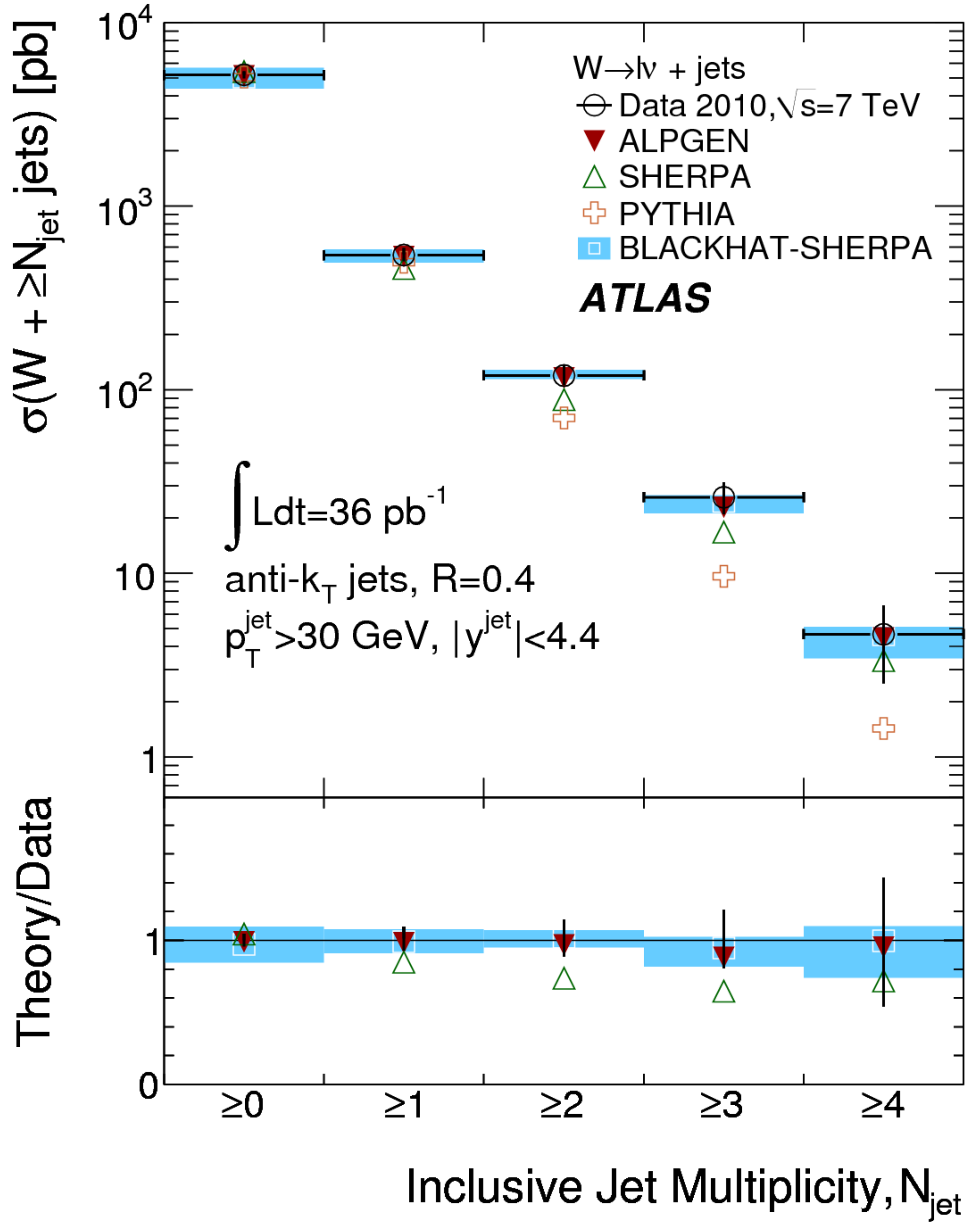
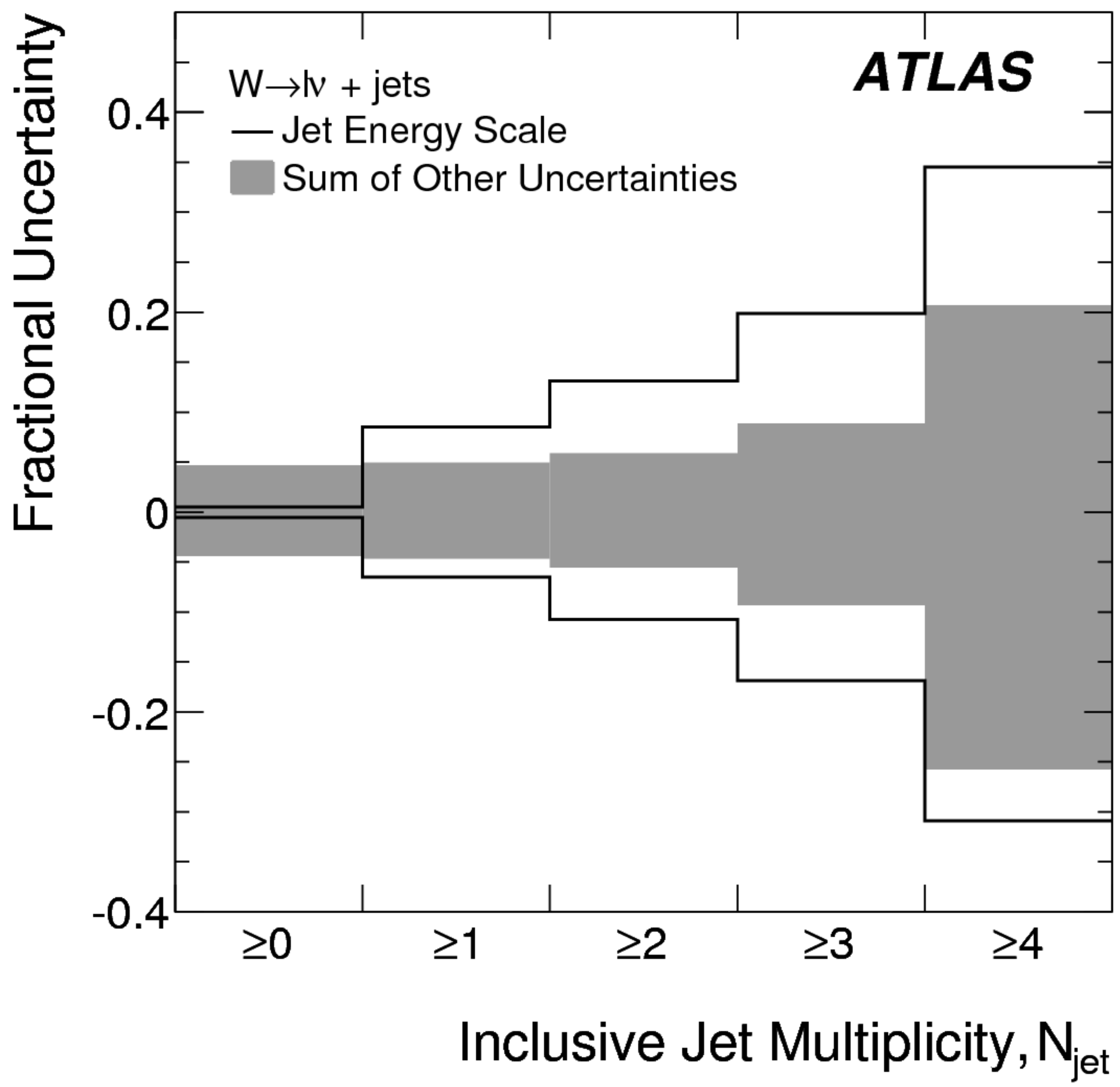
Future experimental paths for high p_T QCD
or
Fit everything
Both a story and a lesson

Monica Dunford - Heidelberg University
Sept 24th, 2024

The world when the LHC started…



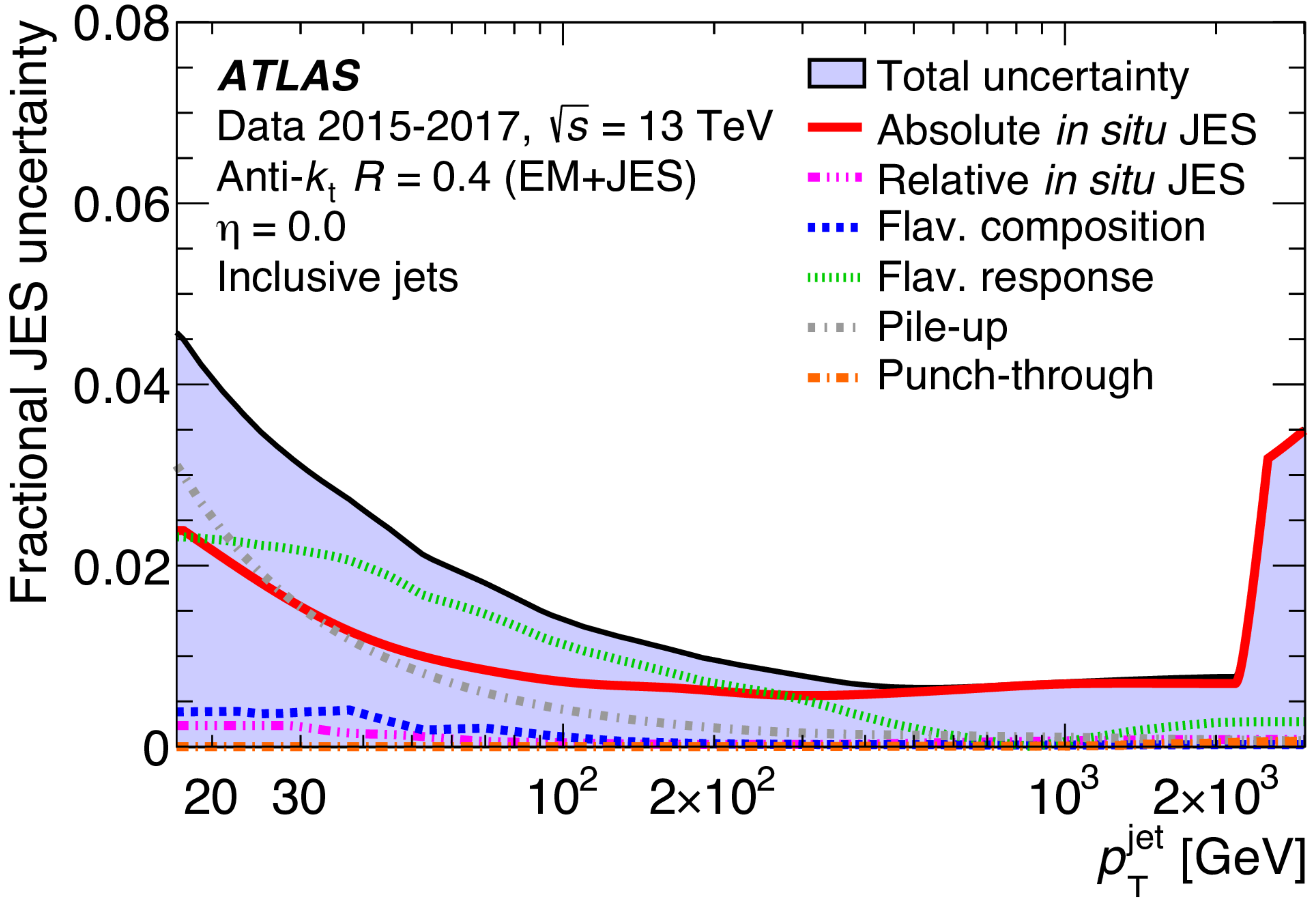


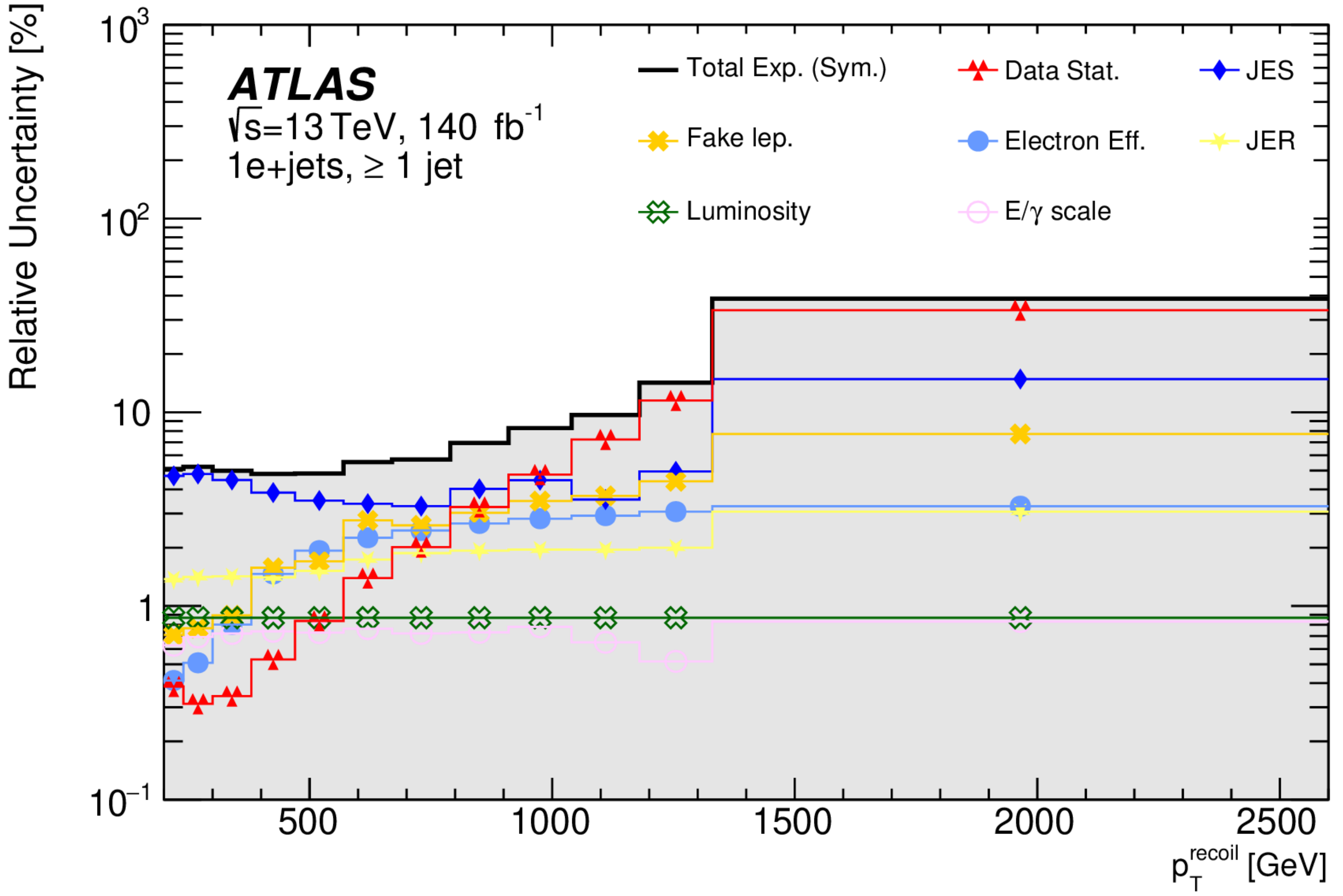
This was precision

The precision of jets

- The detector design goal was 1% precision on the jet energy scale (in the central region and at mid-pT i.e. ~60 GeV)

However, we are still at basically at the stage of JES and everything else





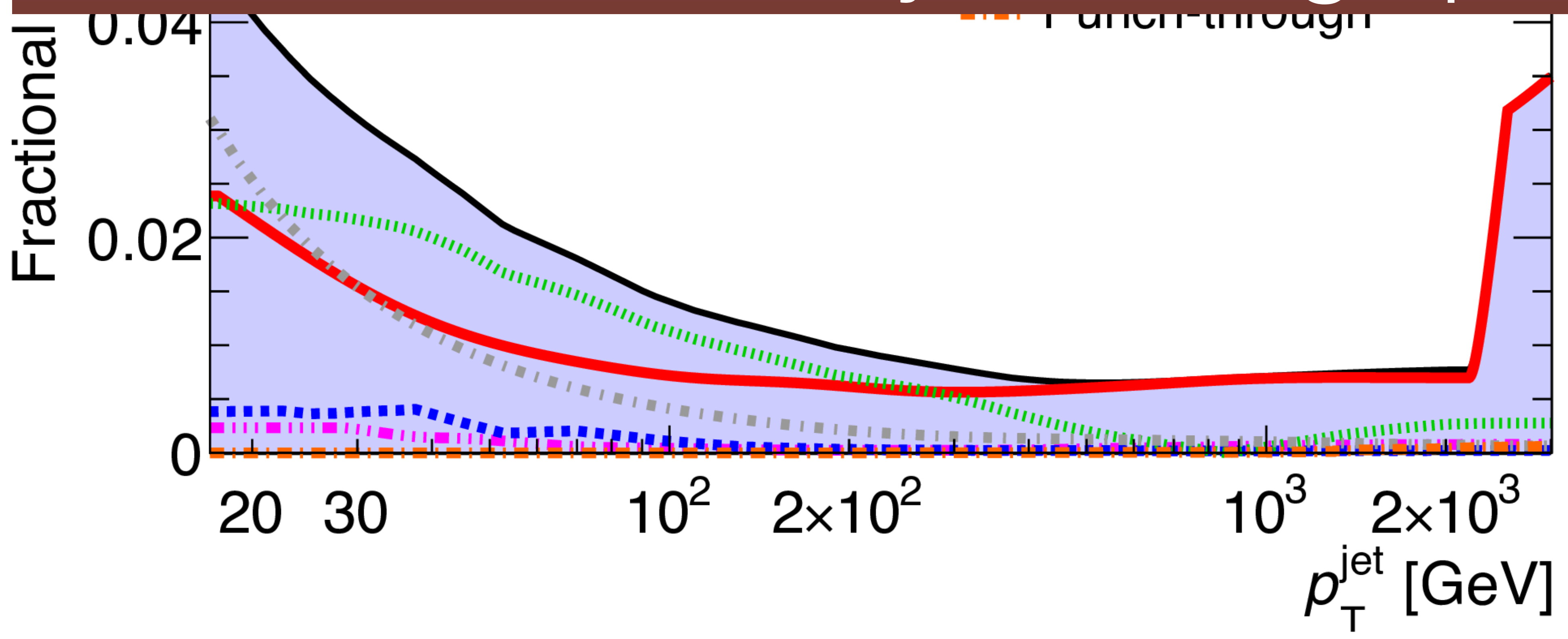
The precision of jets

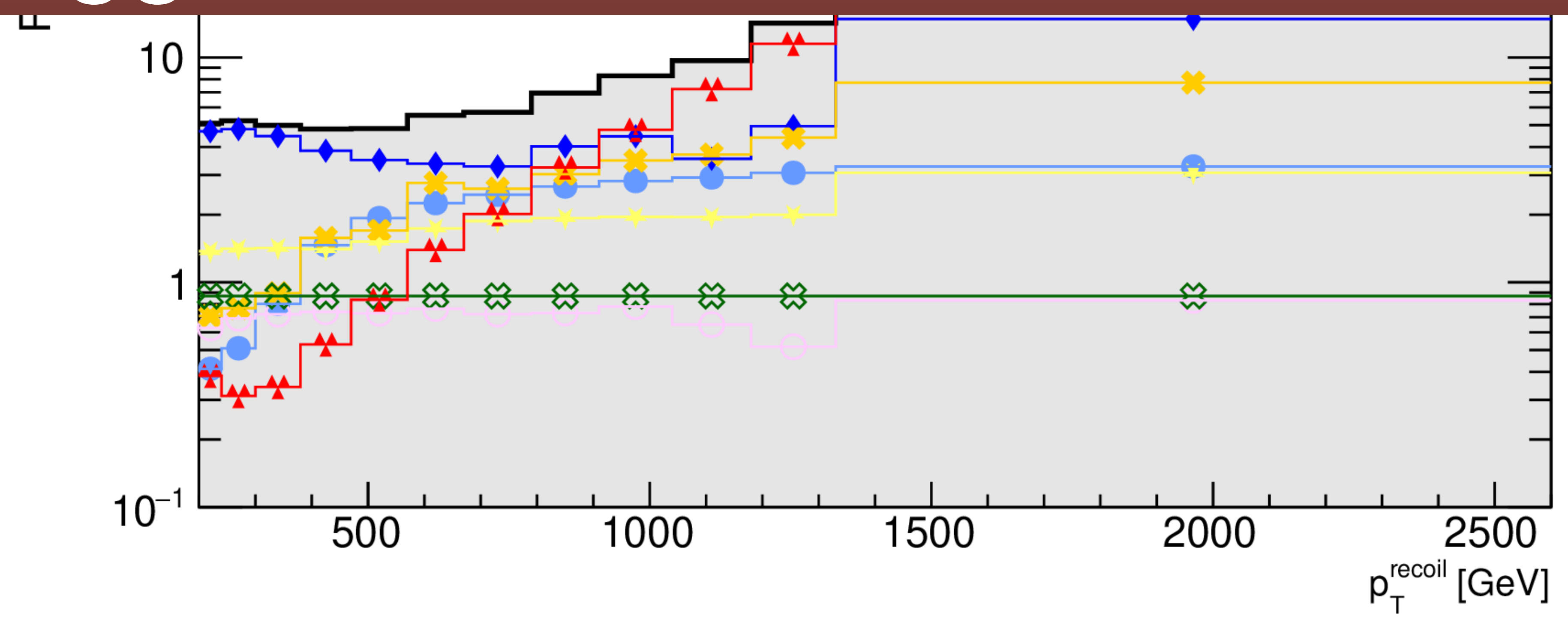
- The detector design goal was 1% precision on the jet energy scale (in the central region and at mid-pT i.e. ~60 GeV)

However, we are still at basically at the stage of JES and everything else

What if that paradigm changes? What if we went from jet uncertainties of 1-5% to those of 0.1-0.5%?

From today’s vantage point, this suggestion is not ridiculous



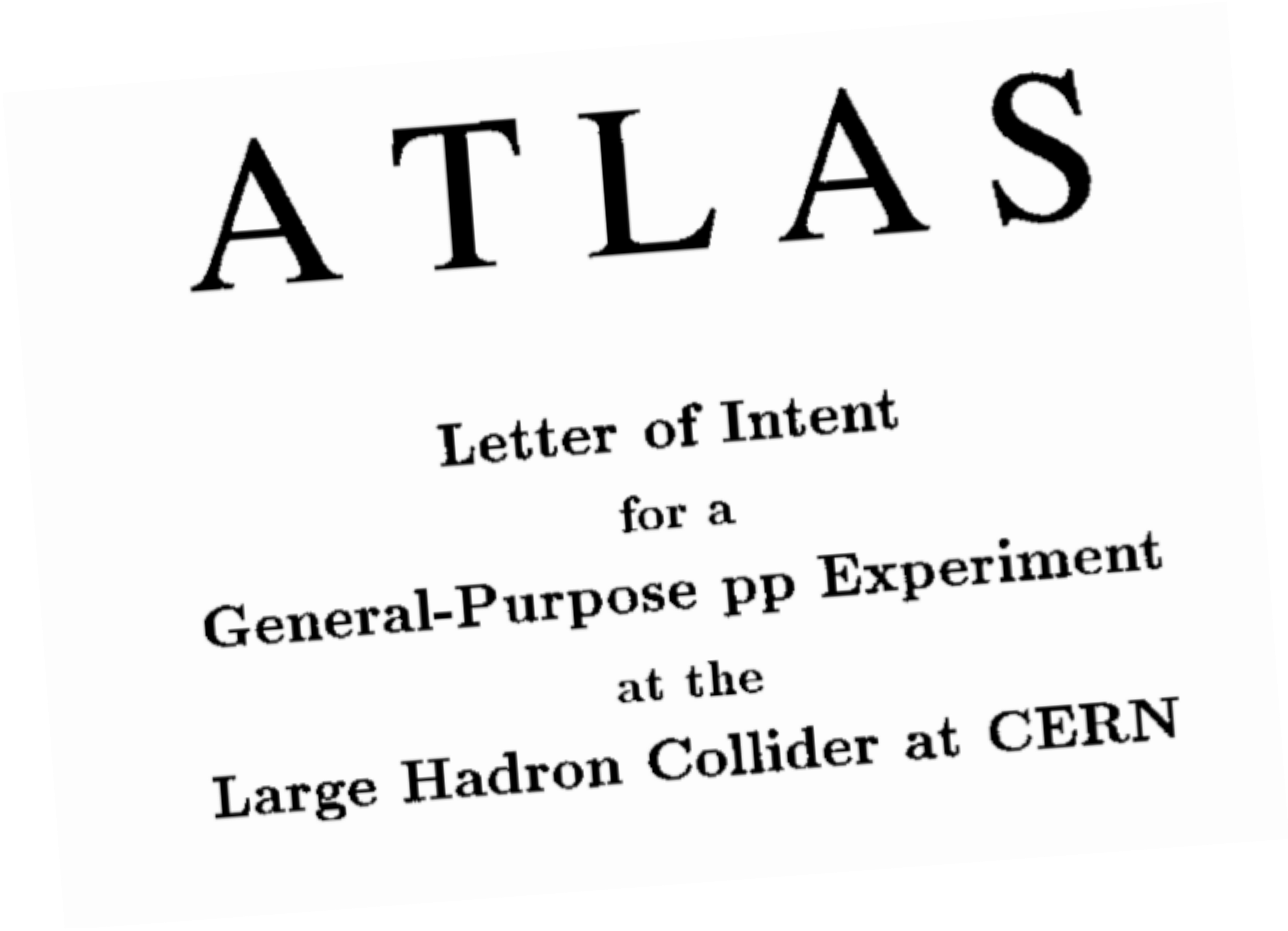


Why and how this paradigm can be broken…

- How to break the paradigm?
  - Our reconstruction was designed for the computers of the 90s
  - We have a huge amount of data
- Will go into two longish examples

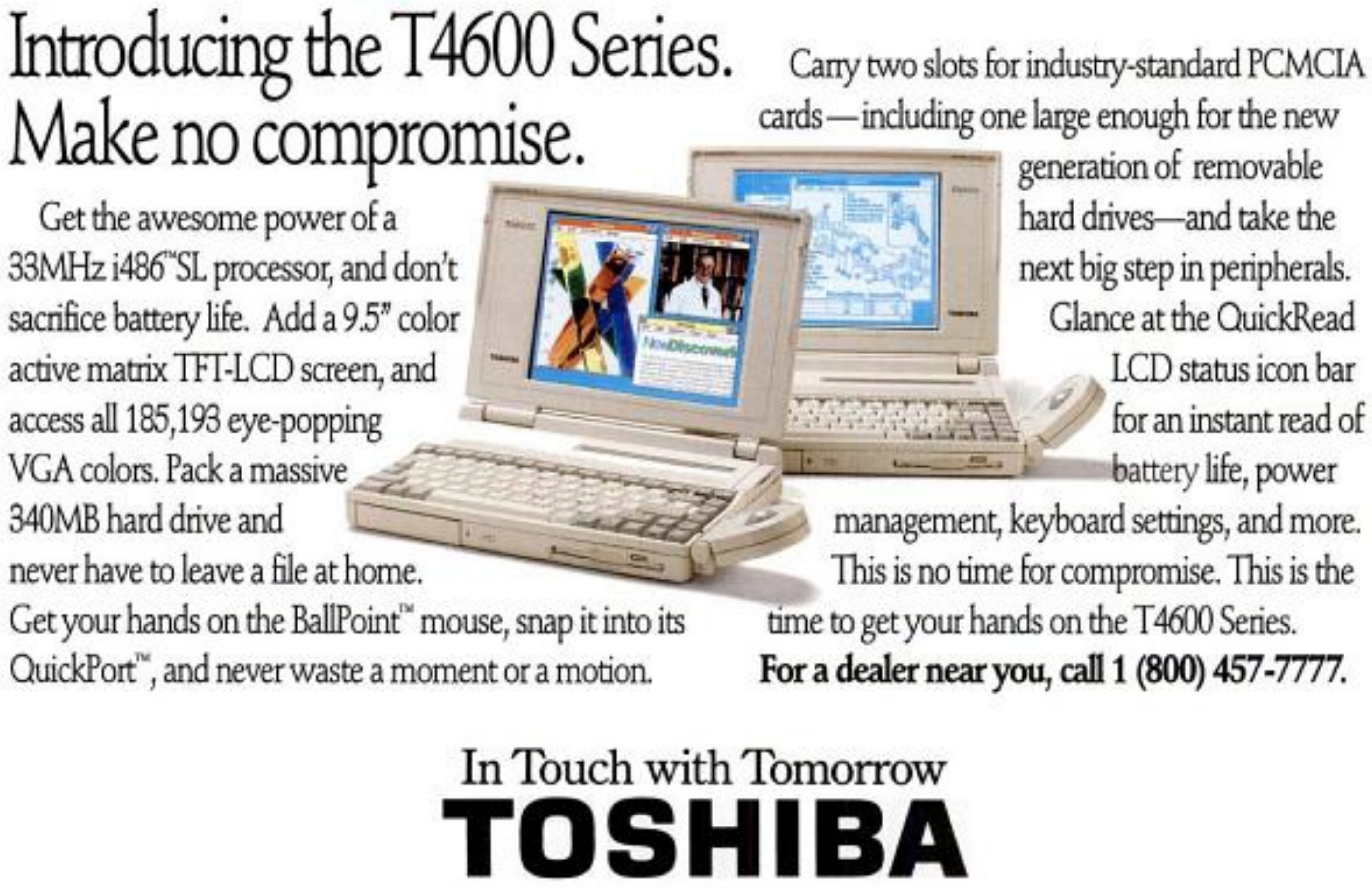
Computers from the 90s…

- There is a huge interest and effort to apply machine learning to object calibration and reconstruction



ATLAS

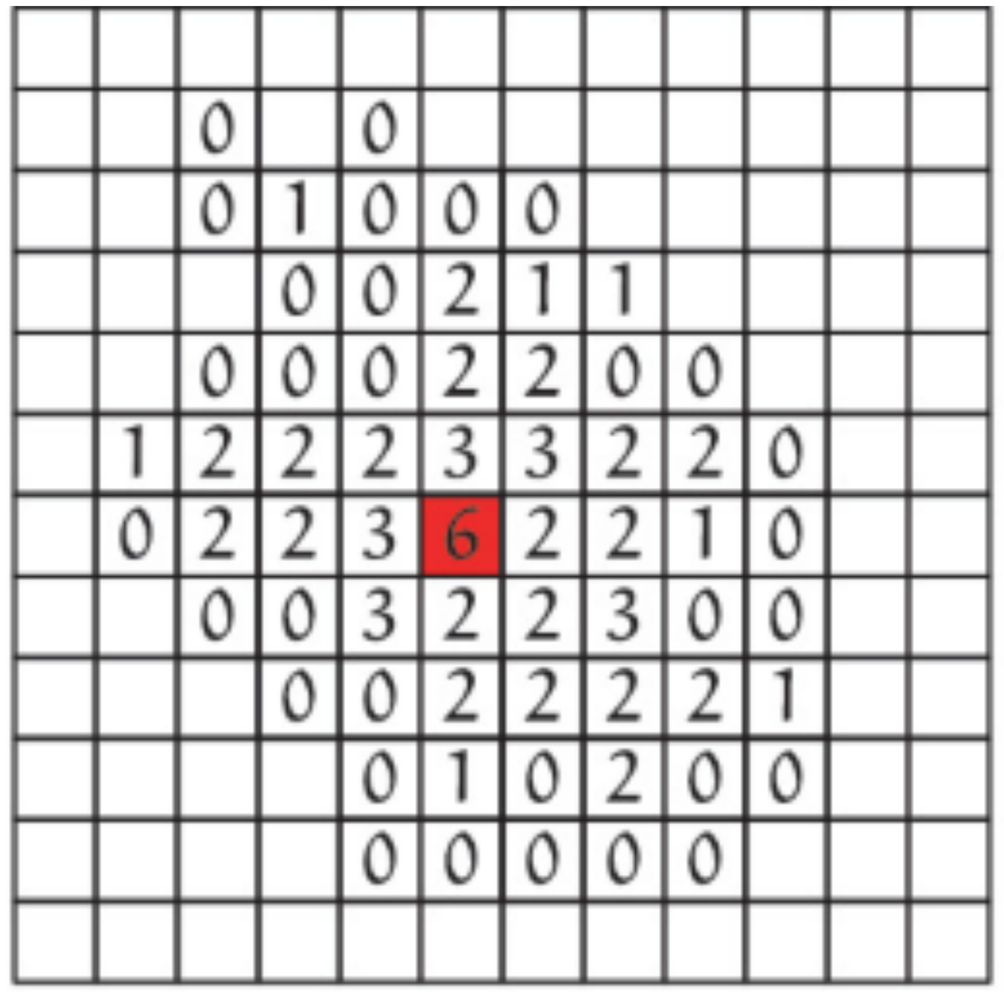
Letter of Intent
for a
General-Purpose pp Experiment
at the
Large Hadron Collider at CERN



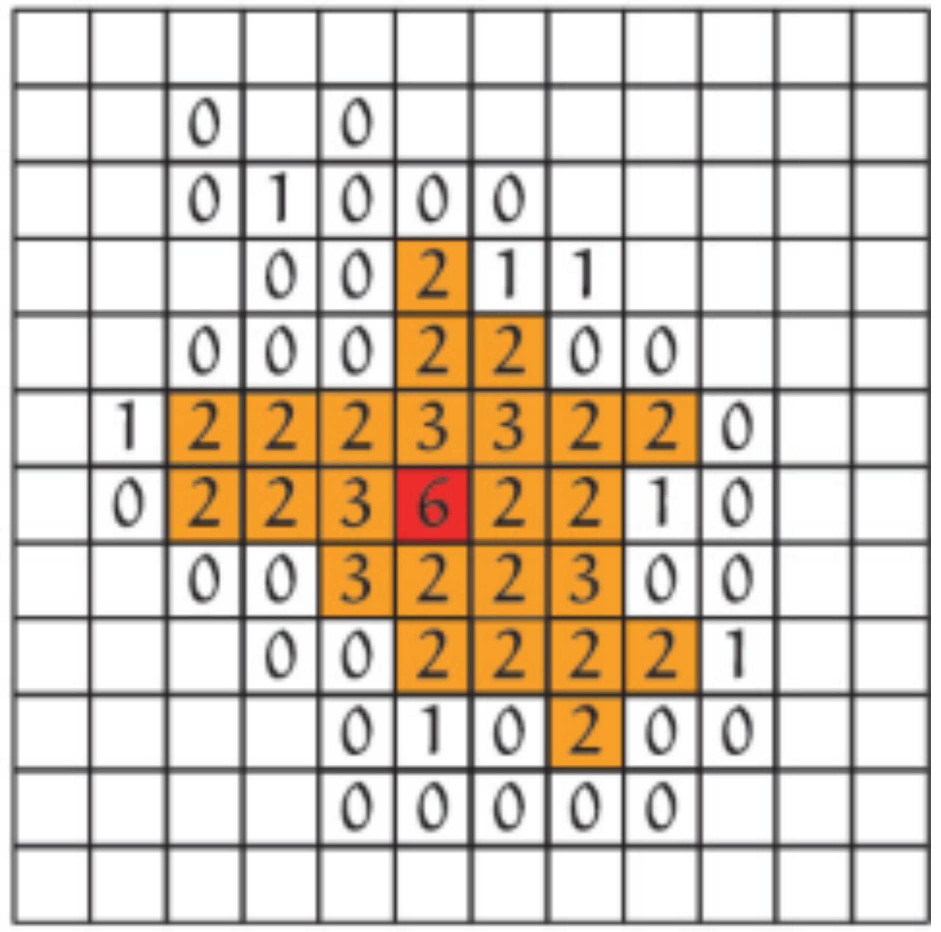
1992

Staying with examples of jets…

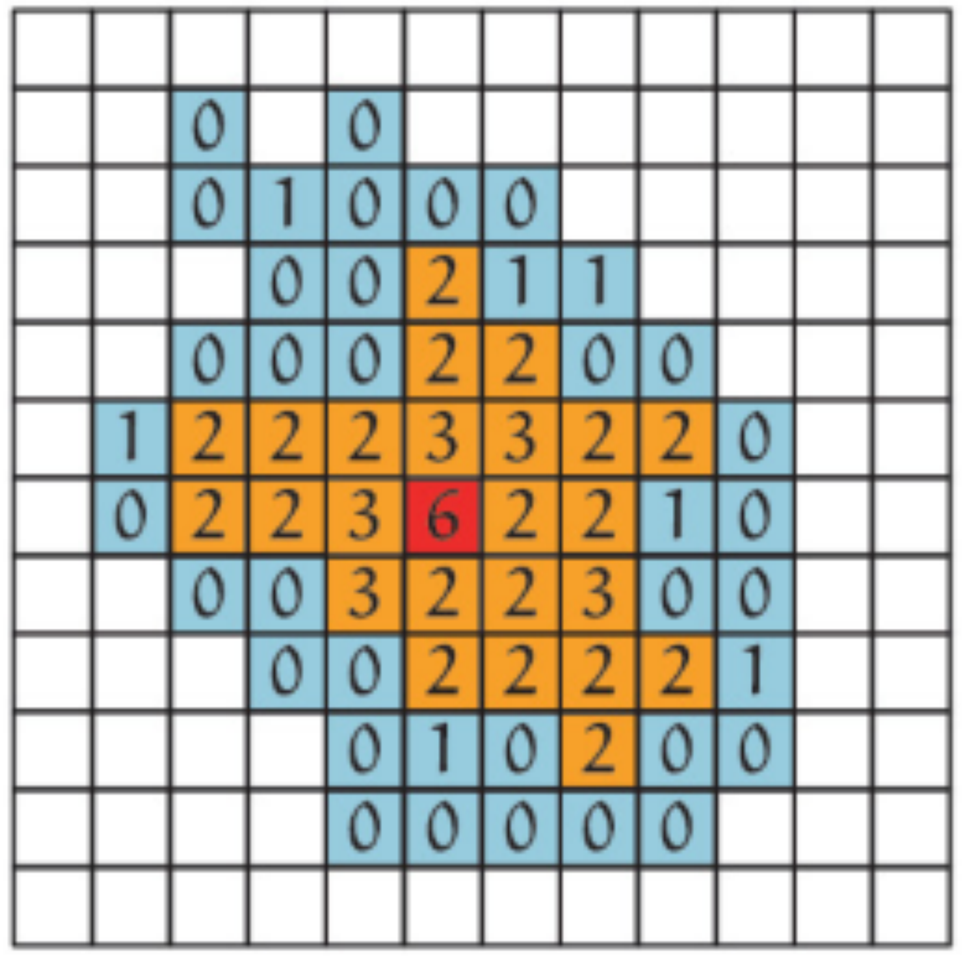
- We take calorimeter cells, build topo-clusters and then these to build jets



(a) Clustering of $|\epsilon_{\text{cell}}^{\text{EM}}| > 4$ cells.



(b) Clustering of $|\epsilon_{\text{cell}}^{\text{EM}}| > 2$ cells.

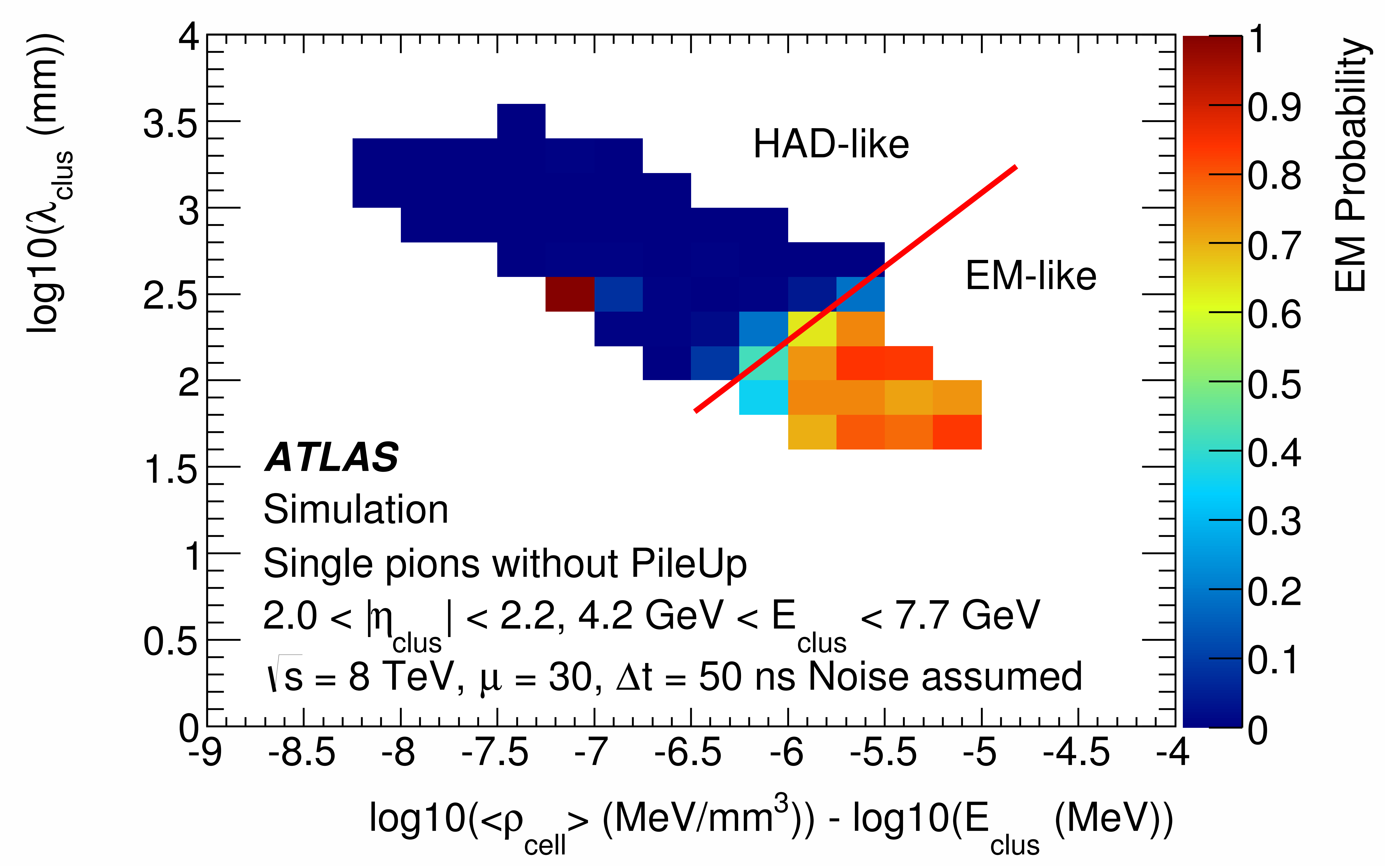


(c) Clustering of $|\epsilon_{\text{cell}}^{\text{EM}}| > 0$ cells.

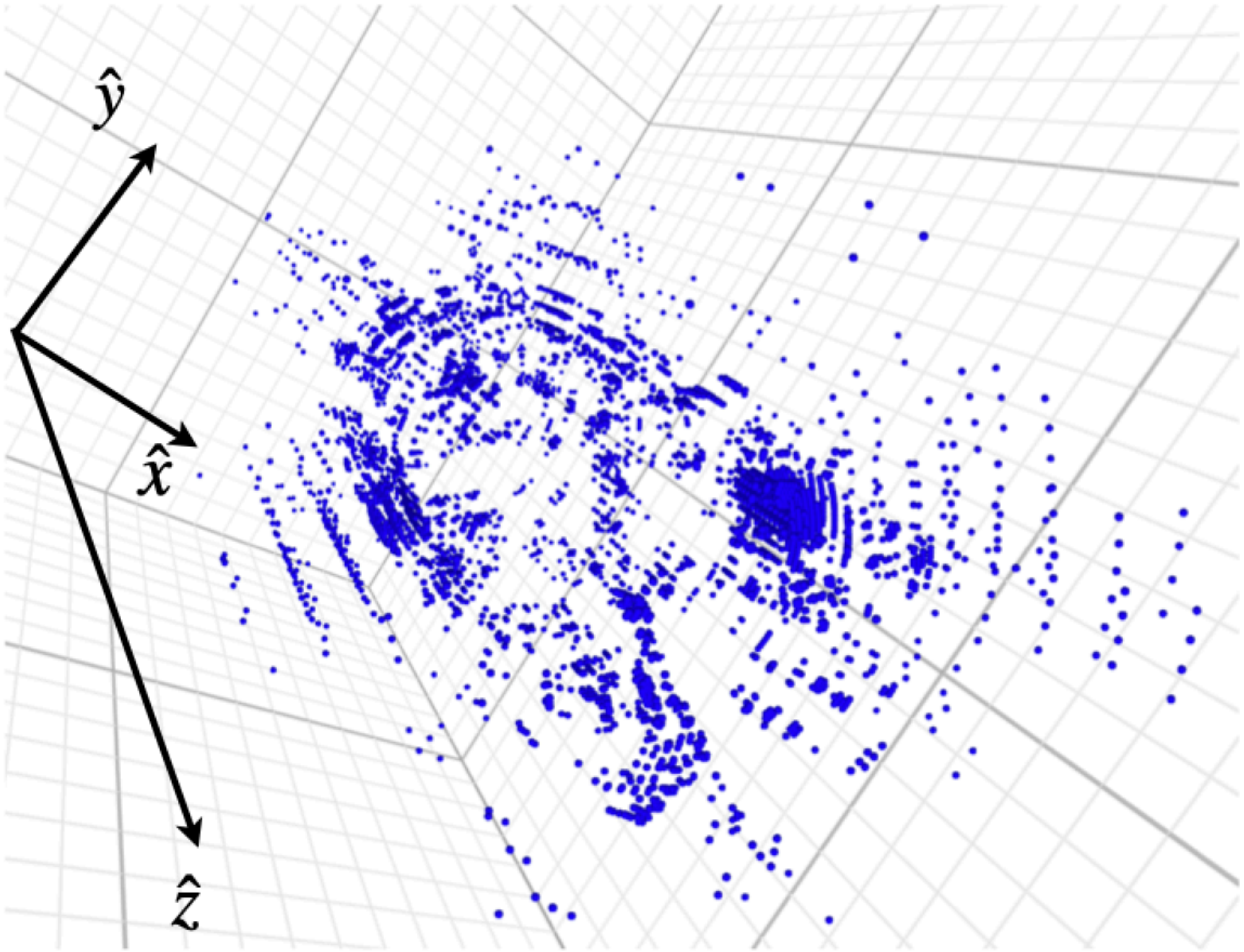
$$|E_{\text{seed, cell}}^{\text{EM}}| > S\sigma_{\text{noise, cell}}^{\text{EM}} \rightarrow \frac{|E_{\text{seed, cell}}^{\text{EM}}|}{\sigma_{\text{noise, cell}}^{\text{EM}}} > 4$$

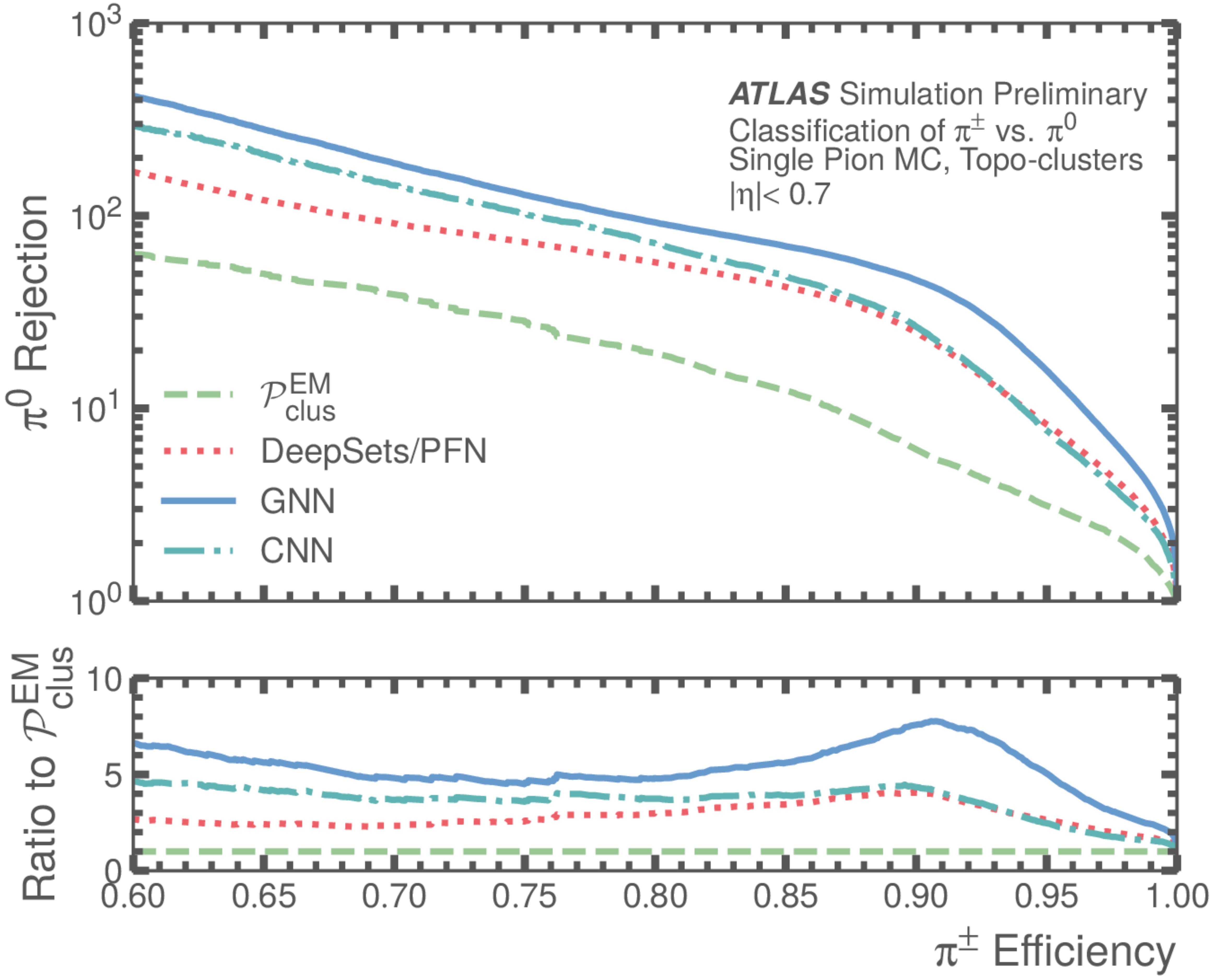
Staying with examples of jets…

- We consolidate (lose) information at each step
- To calibrate we define the probability a cluster is EM-like - based on i.e. energy-weighted average signal densities, longtitudal depth, first energy-weighted moment

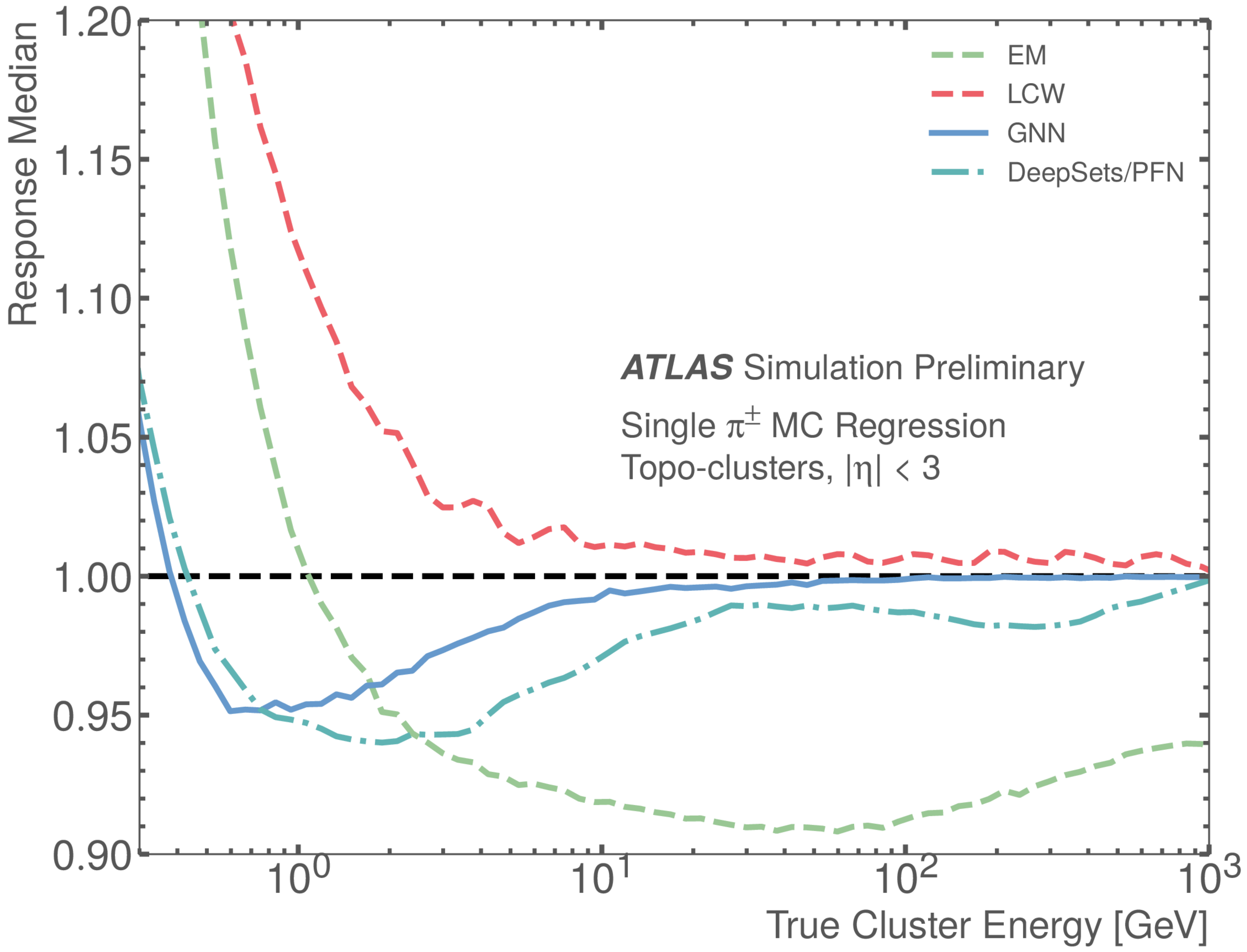


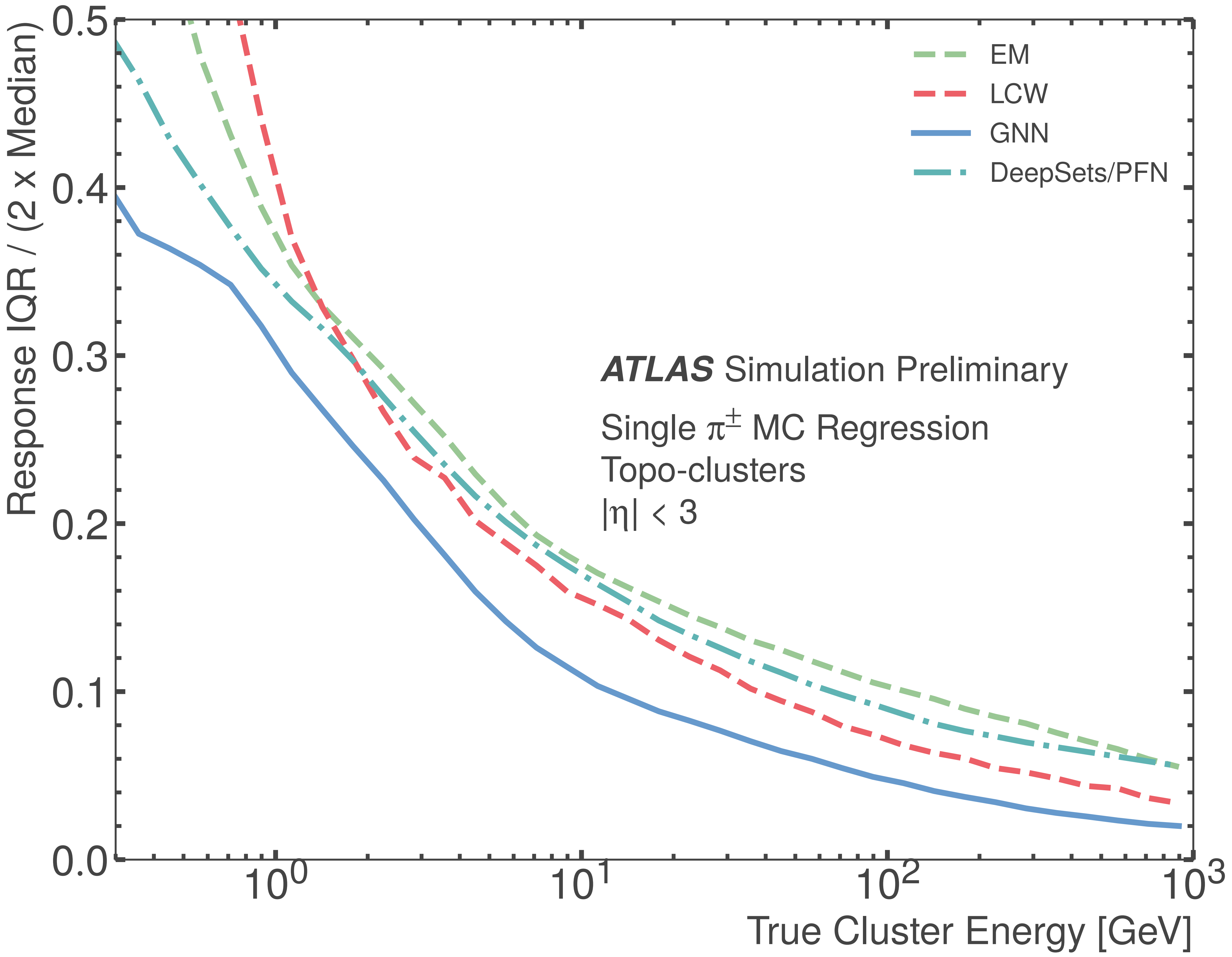
Exploring machine learning at the cell level





Exploring machine learning at the cell level

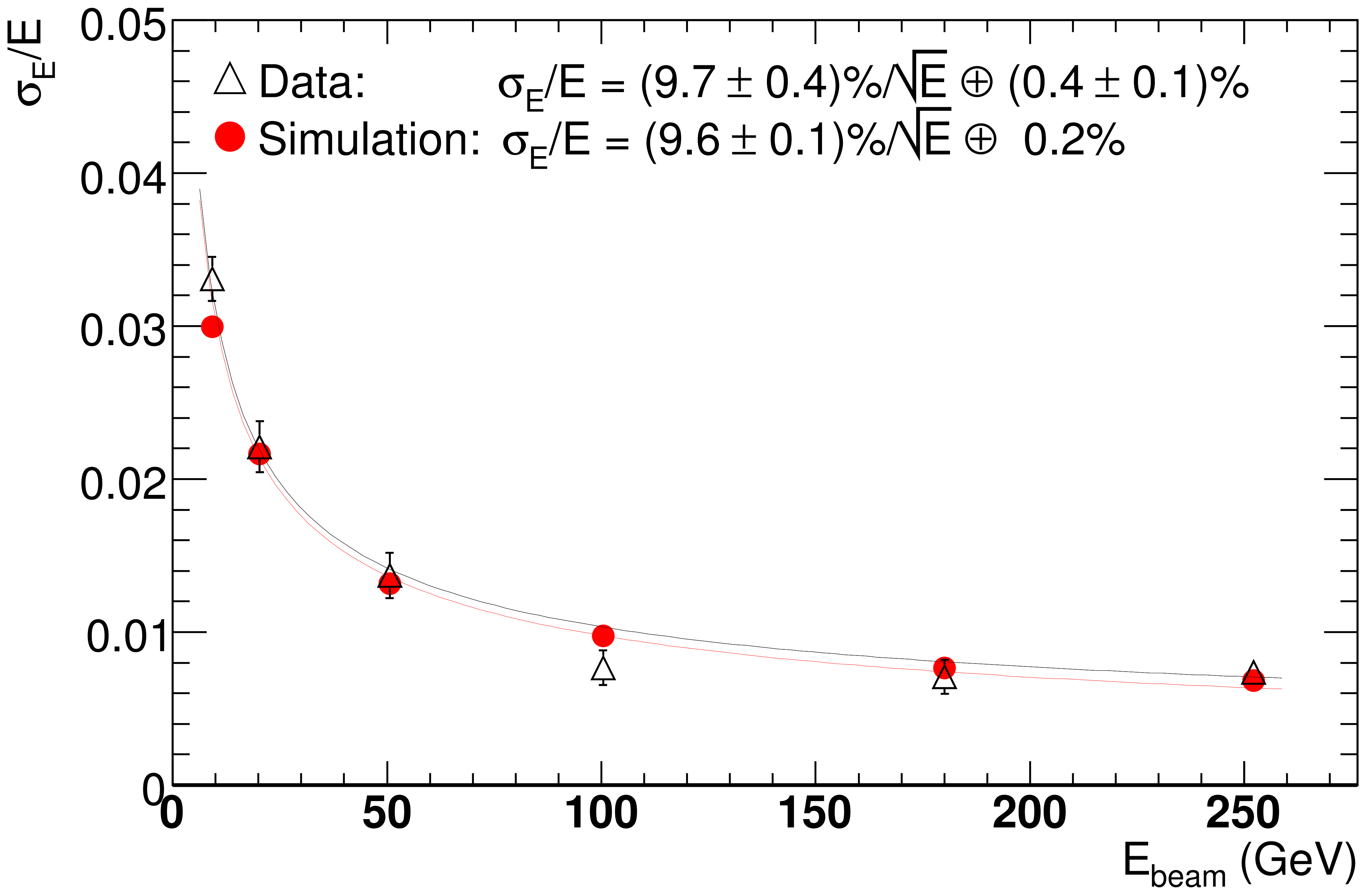


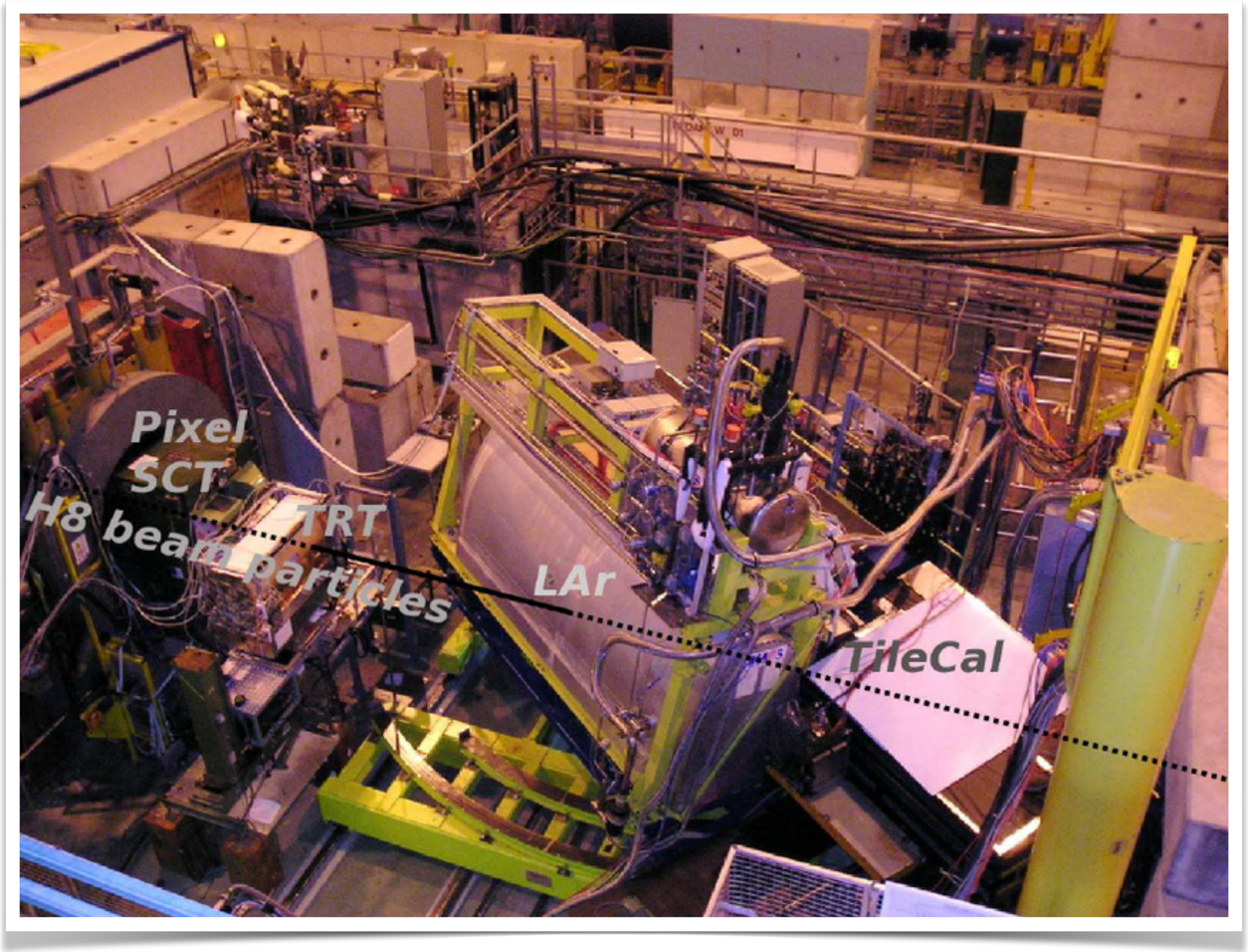


Very promising results - also many ideas on improved reconstruction. Stay tuned!

The power of data

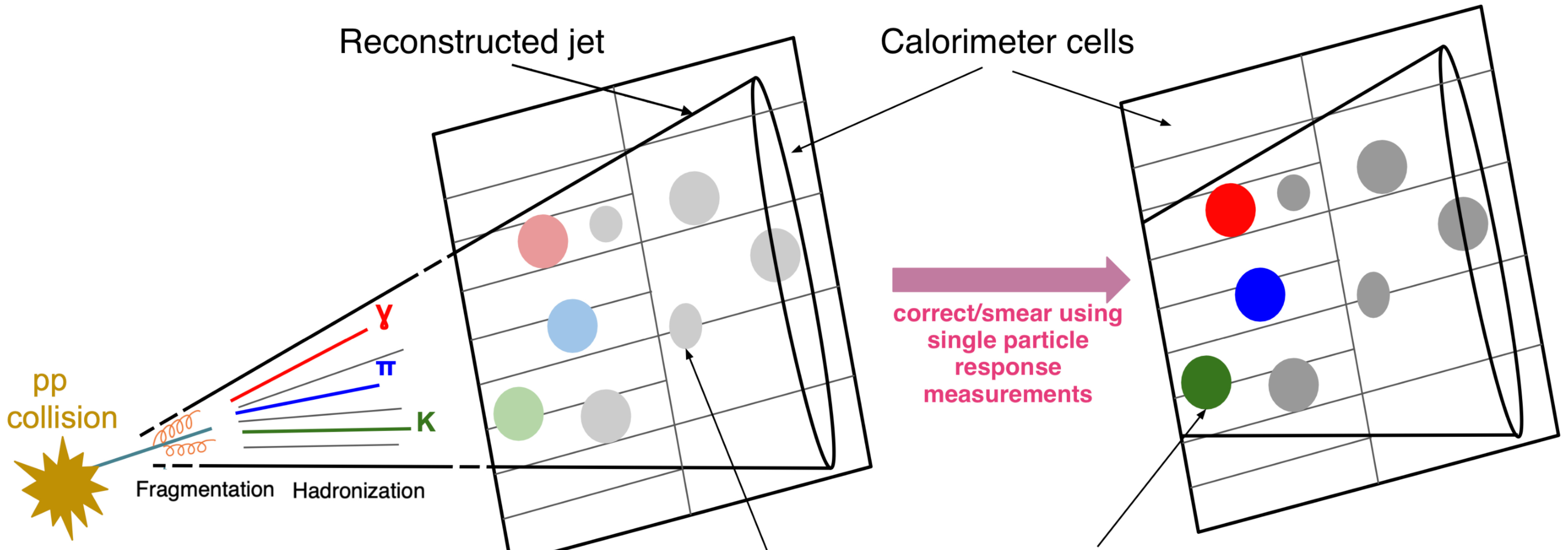
- In determining the JES, we rely on a series of standard candles (like Z+jets, gamma+jets) and testbeam data (taken largely in 2004)

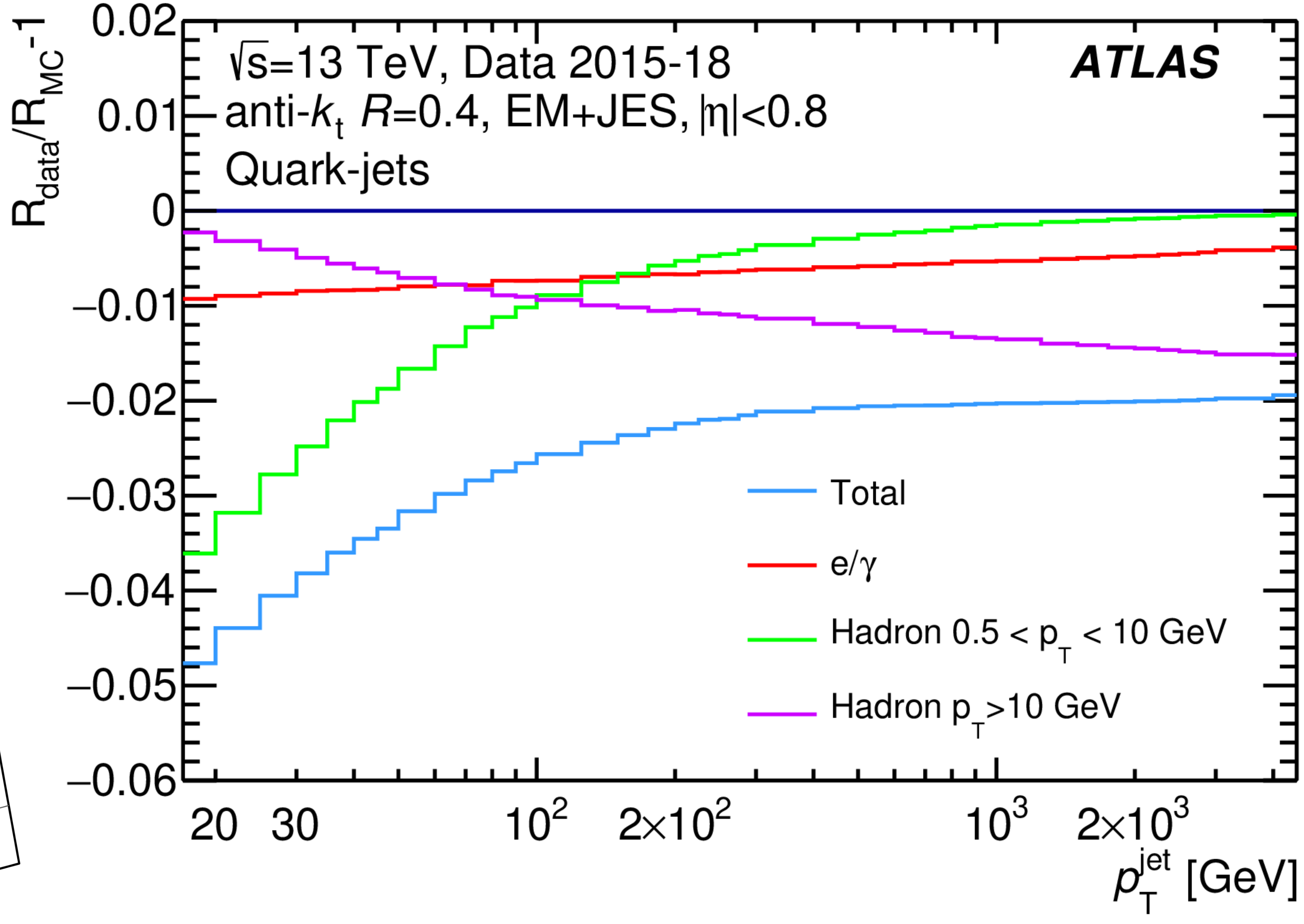




The power of data

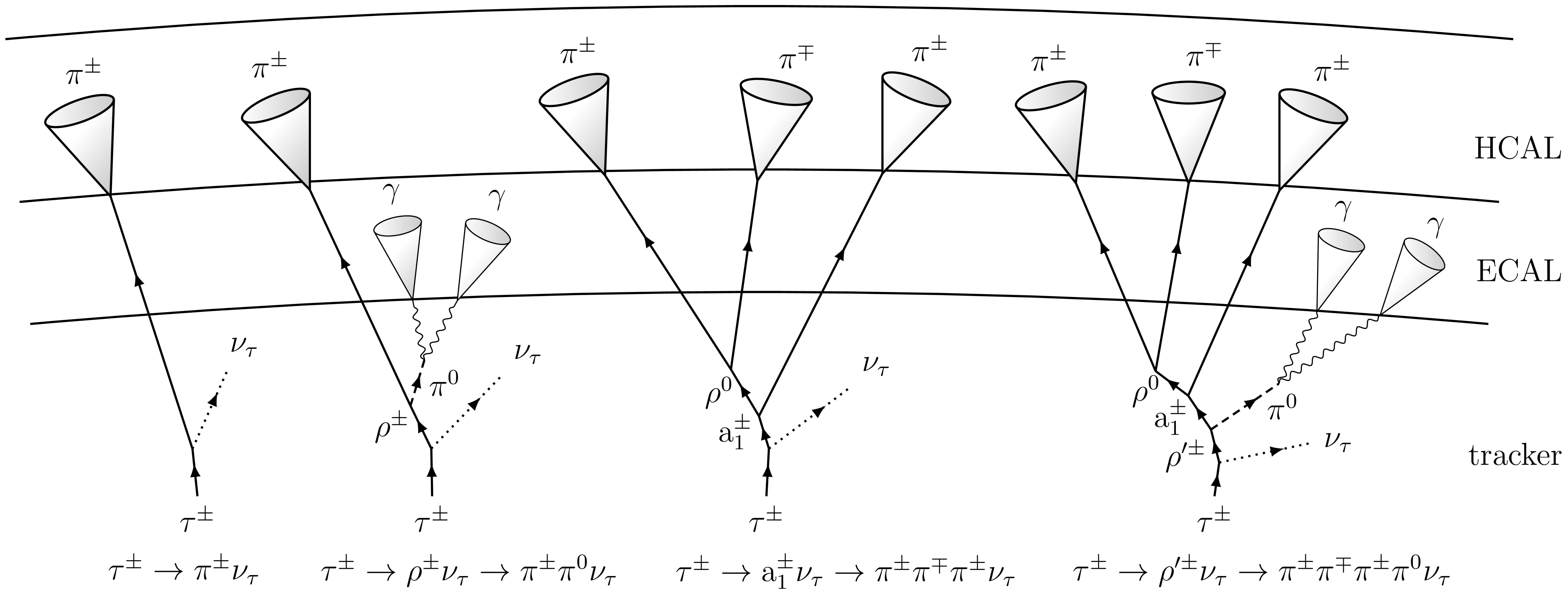
- Measure the momentum in the tracker to the energy in the calorimeter (e/p)
- Apply that correction to single particles in MC to correct the response w.r.t data

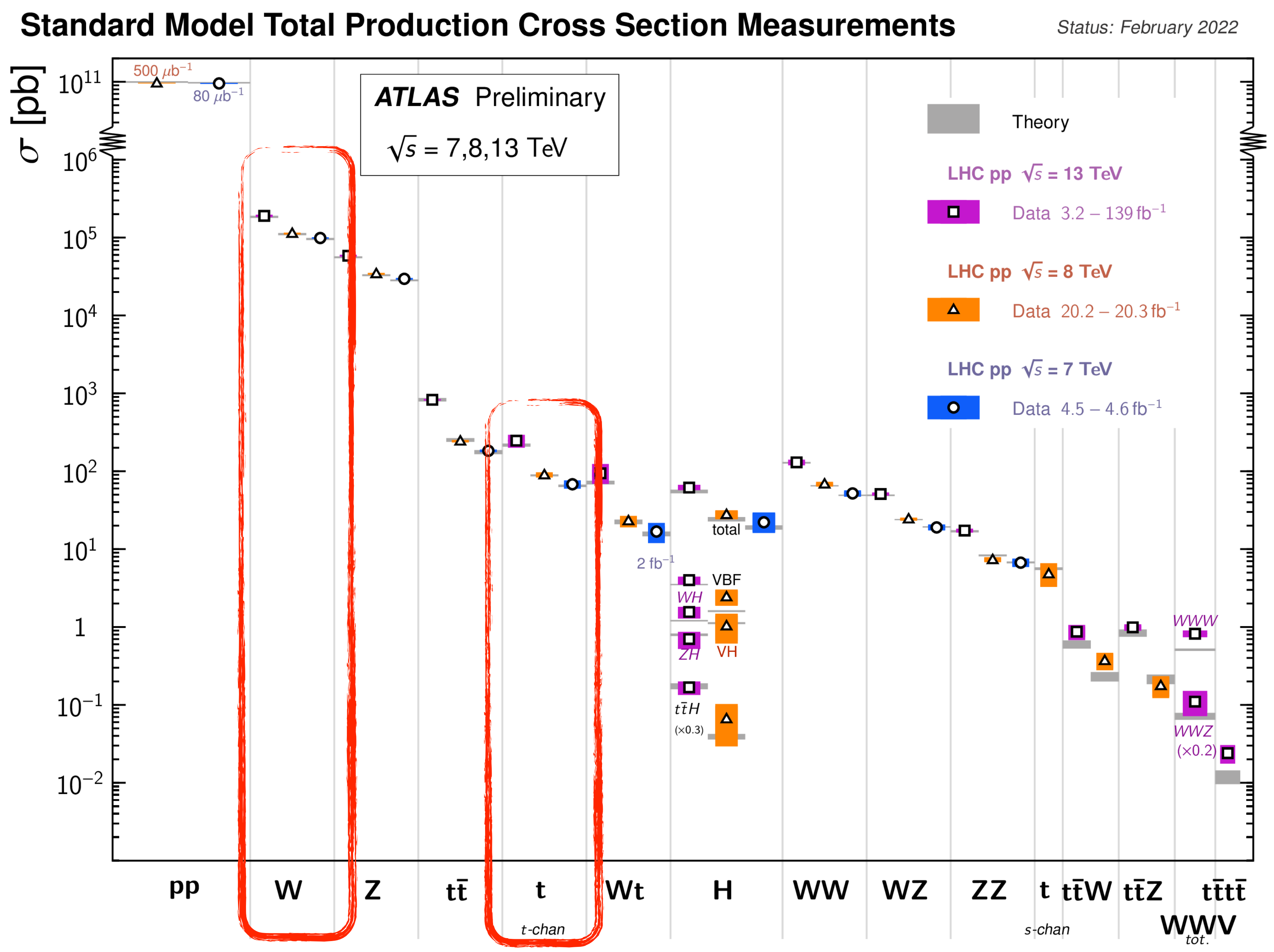




The LHC as ‘test beam’

- The LHC has a natural sample of single particles - taus (and we produce a lot of them



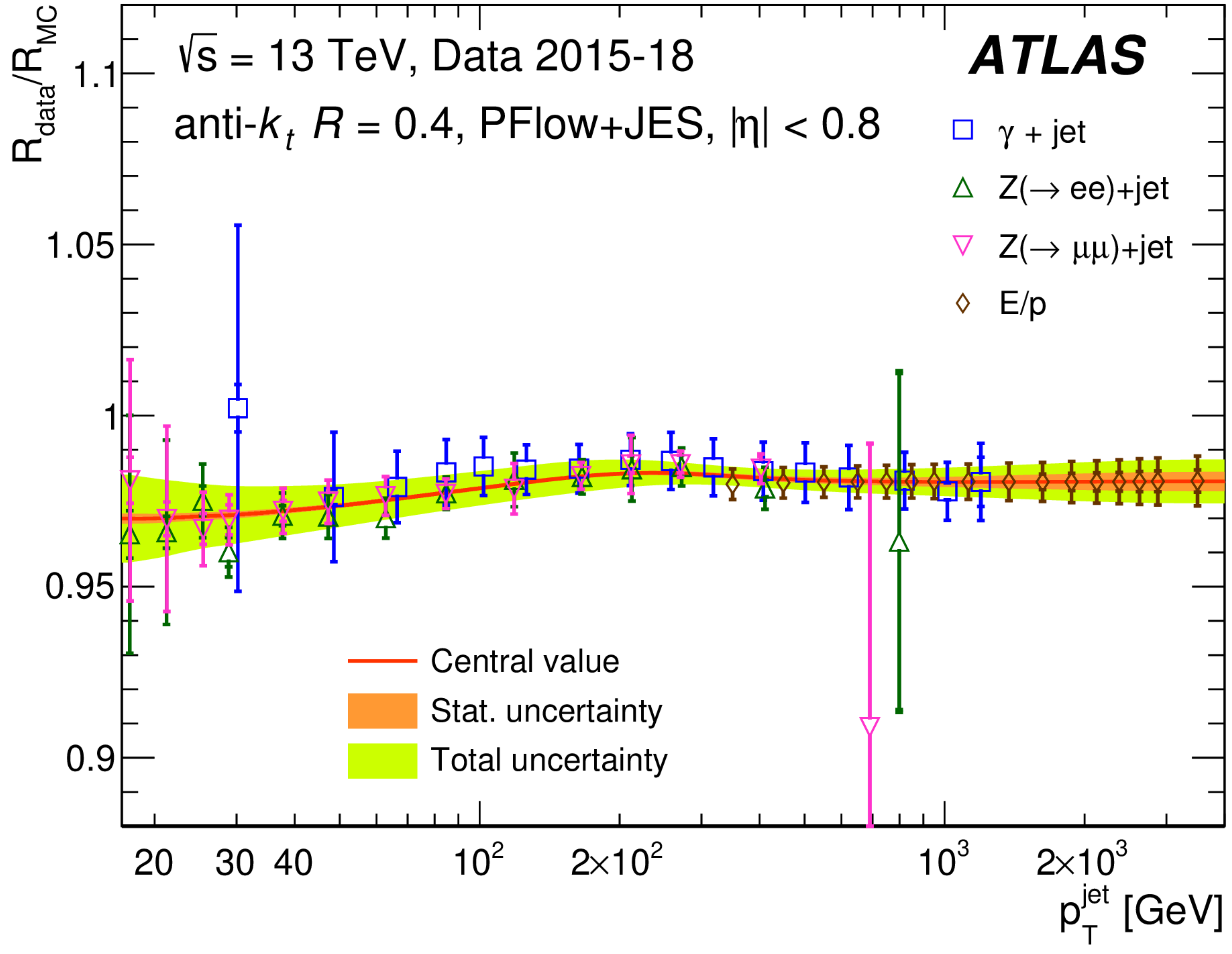


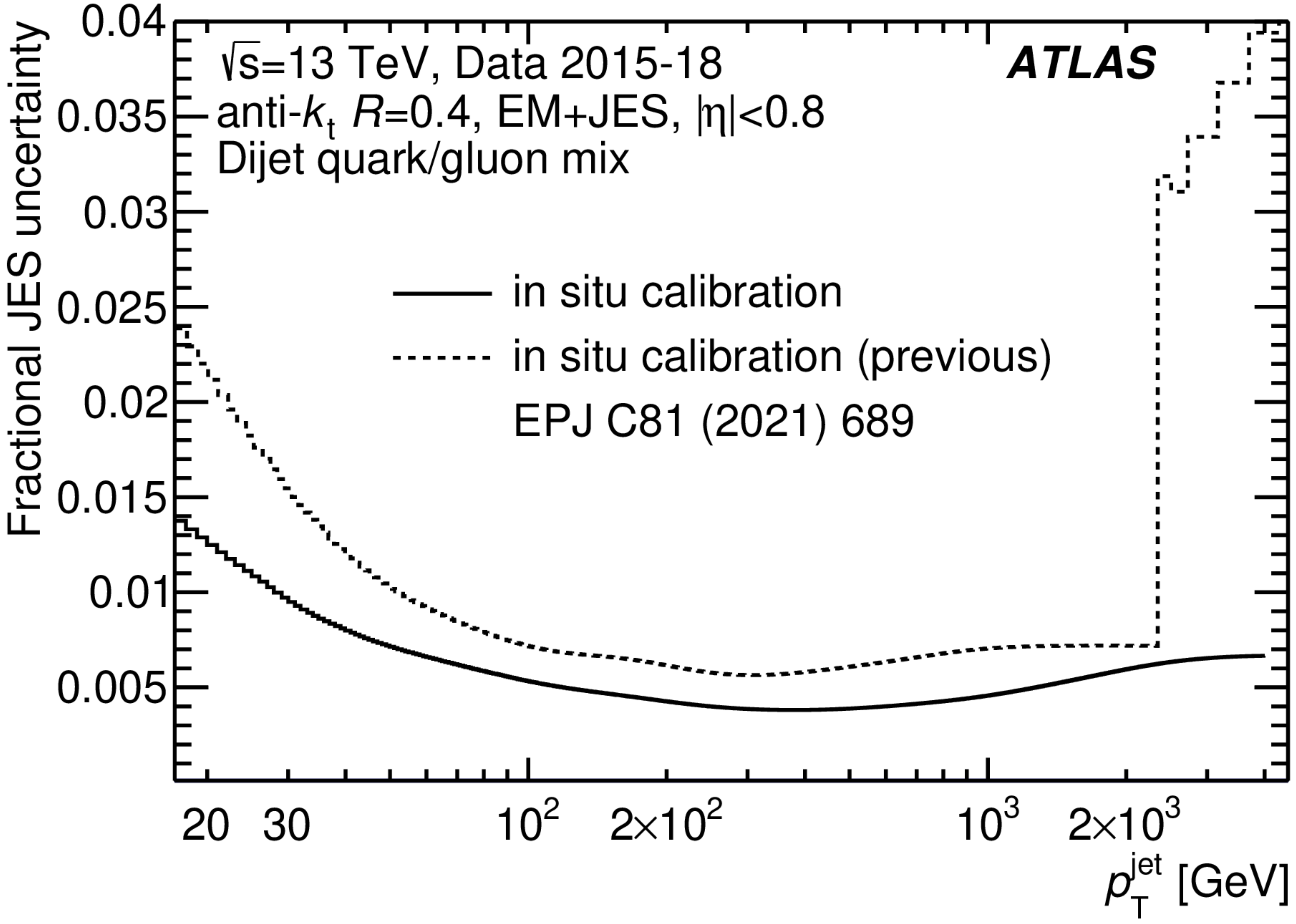
Note this figure is wrong!

[Paper link]

The e/p results

- E/p results give better precision compared to other standard candles
- Big improvement to the overall uncertainty



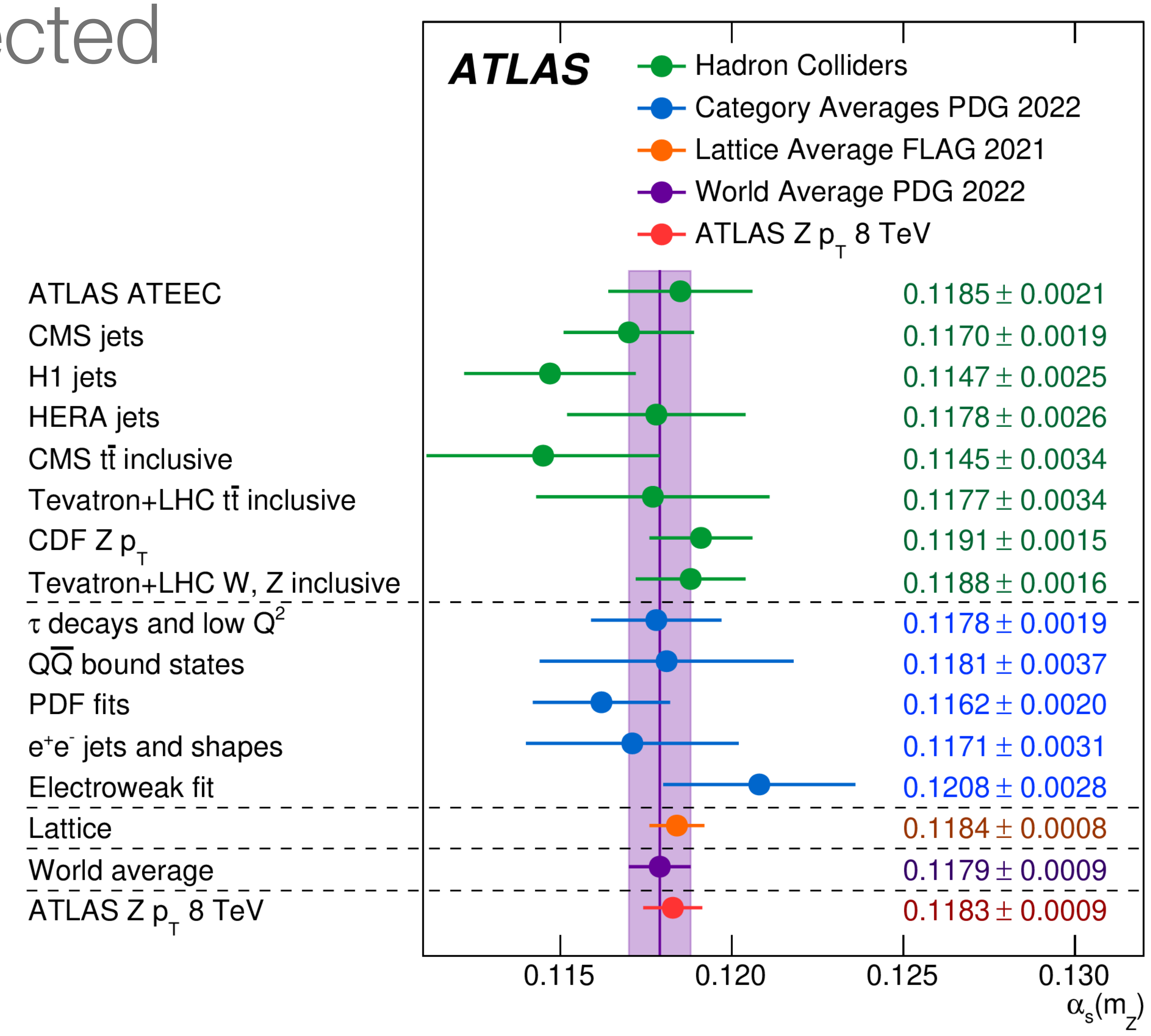


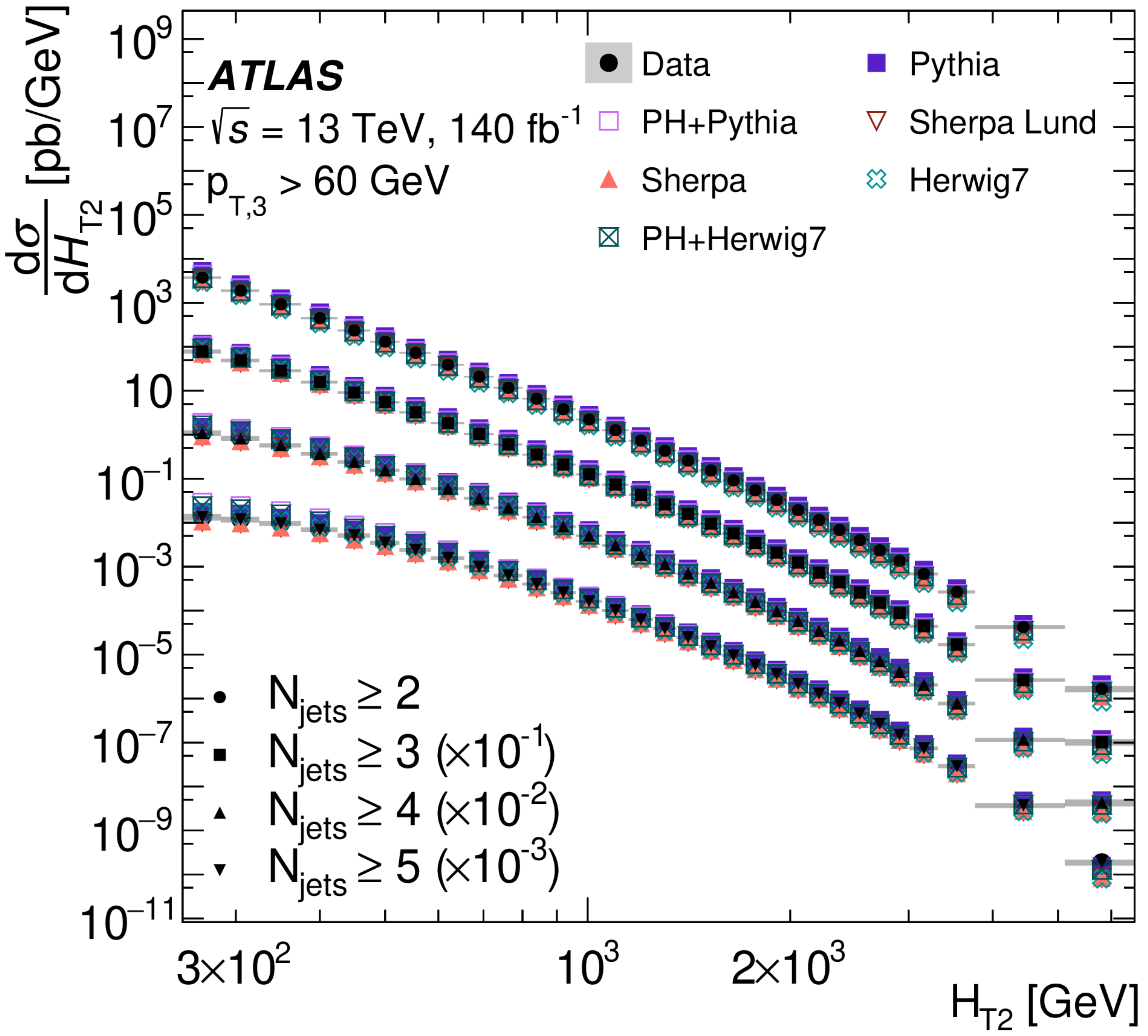
A lesson to be learned here

- You would be surprised to learn how much of our modelling still replies on the very early LHC data
  - Heavy-flavour, jet shapes, fragmentation, forward energy flow mostly done on old data (although this is changing in places), tuning (we have been talking about it for a long time)
  - ATLAS had a huge measurement program at 7 TeV for low pT/inclusive processes, some at 8 TeV and very few at 13 TeV
  - The shortcomings of this will start coming up more and more
- New methodologies aside, much of our data is untapped even using established methodologies
- Support from the theory community that measurements like these are important is critical to raising their visibility within the experimental collaborations

A direct application

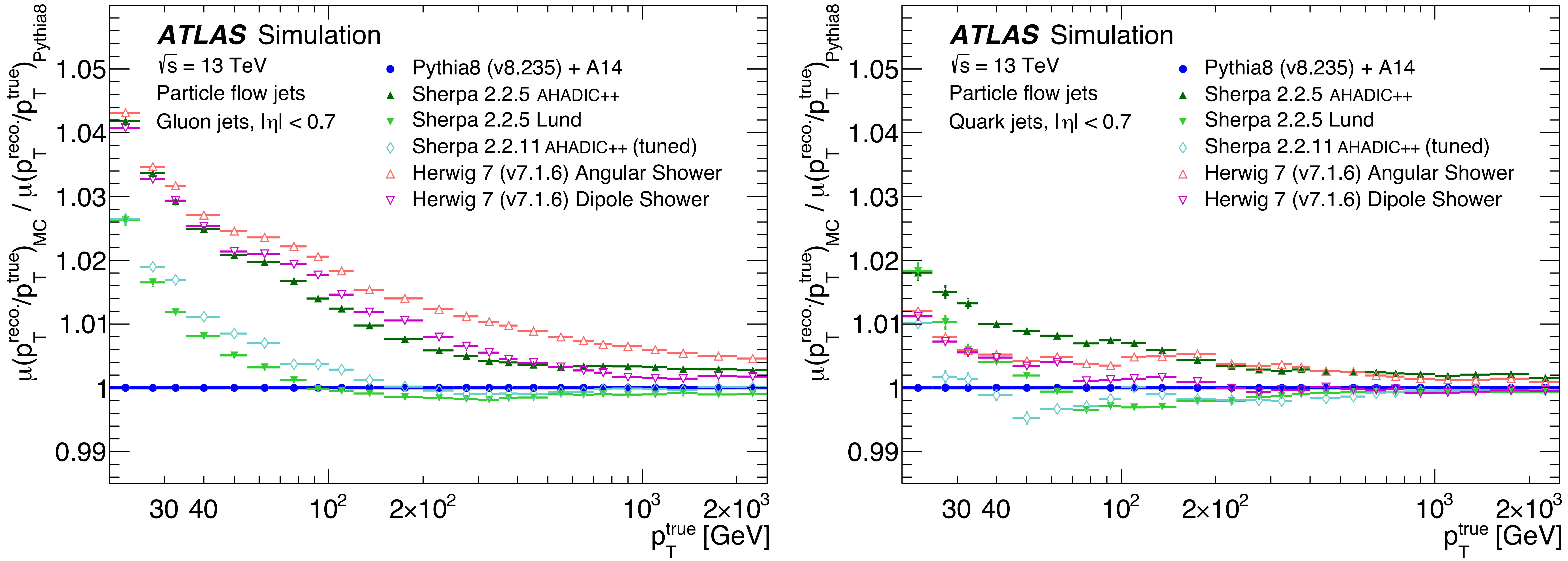
- Jet cross sections are very powerful but are dominated by jet uncertainties, as expected





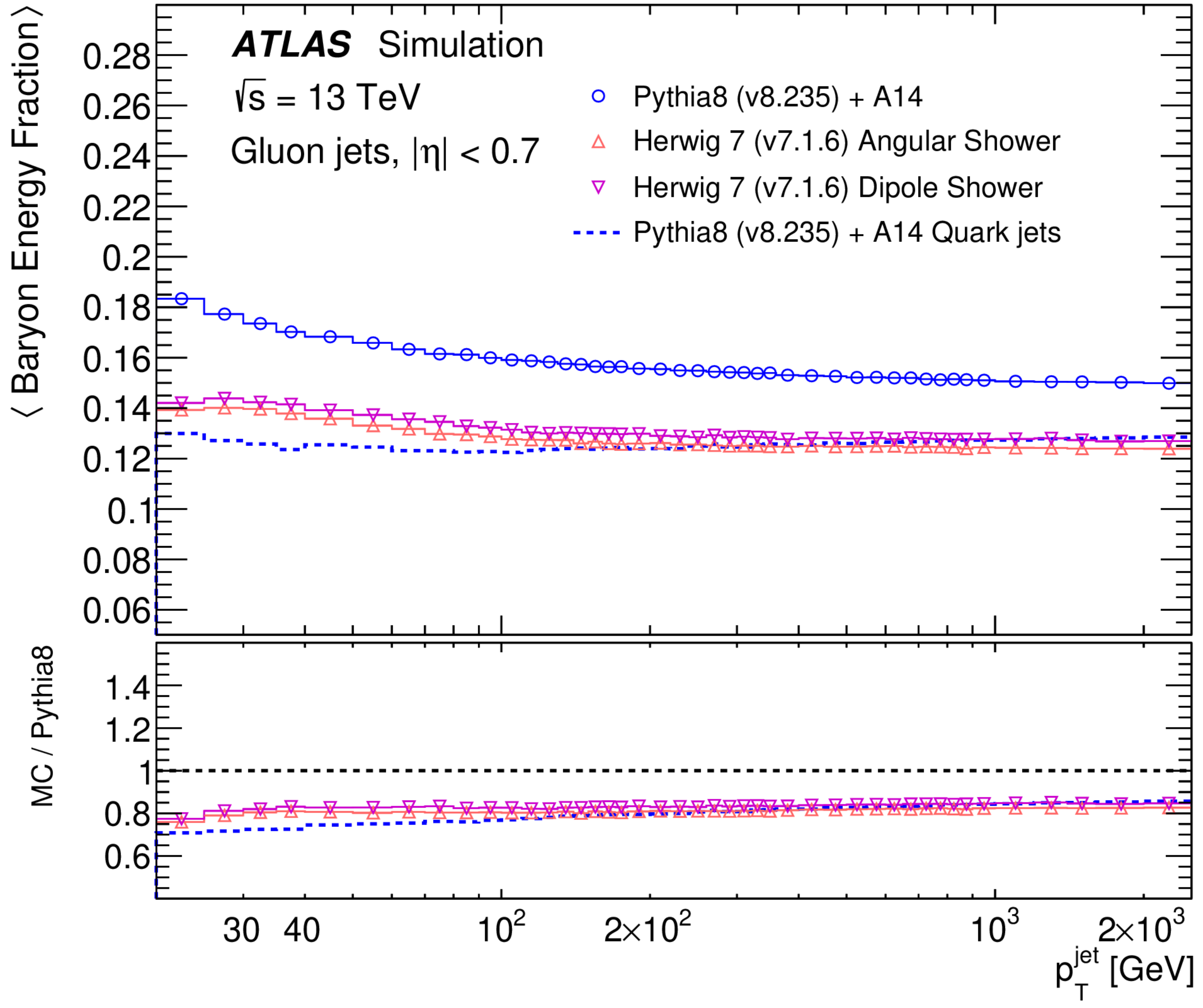
Improving how we handle MC differences

- Big uncertainties because of the relative jet energy response between different MC generators
- Particle momentum and compostion also plays a big role as the calorimeter response is very sentitive to it



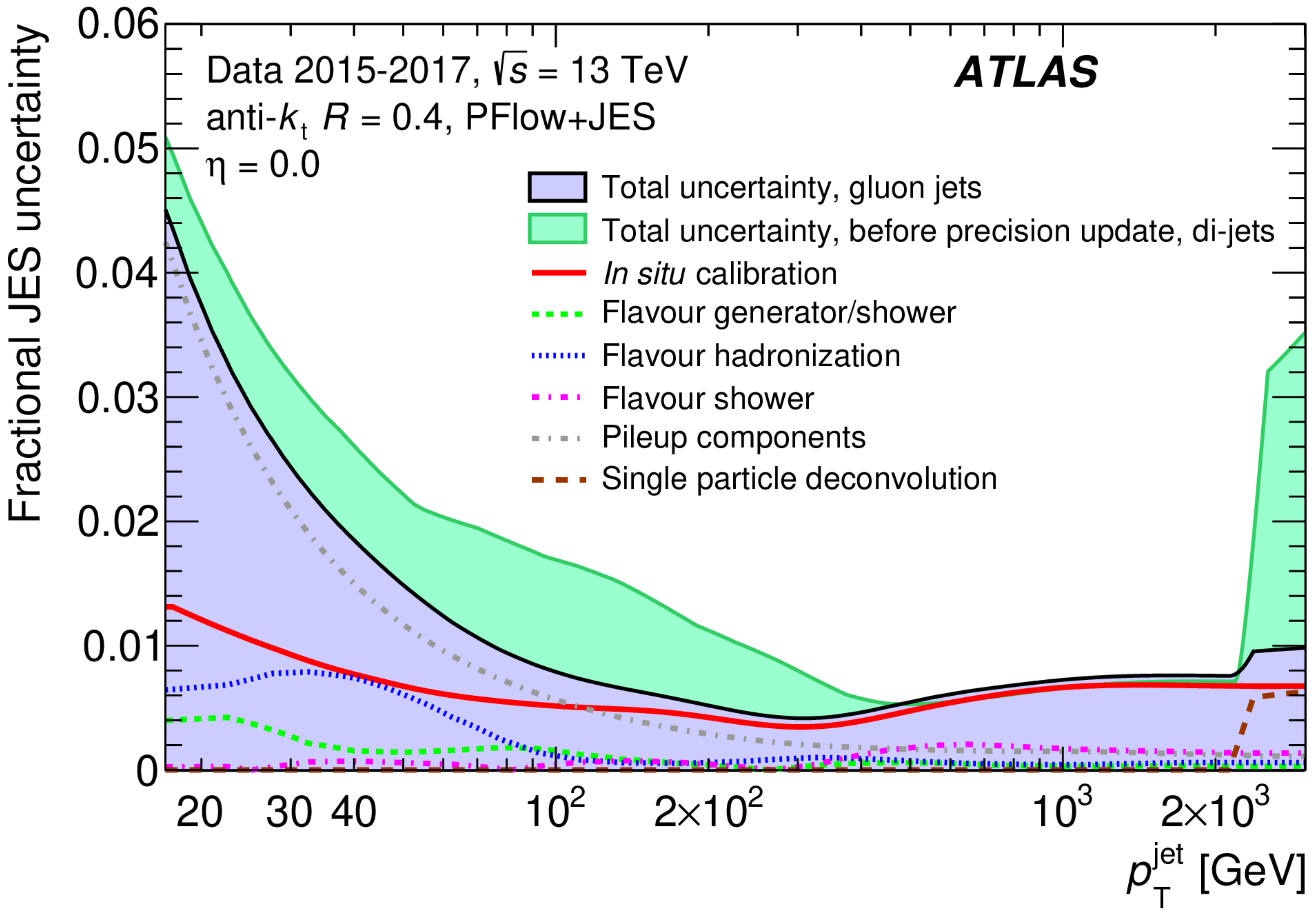
Improving how we handle MC differences

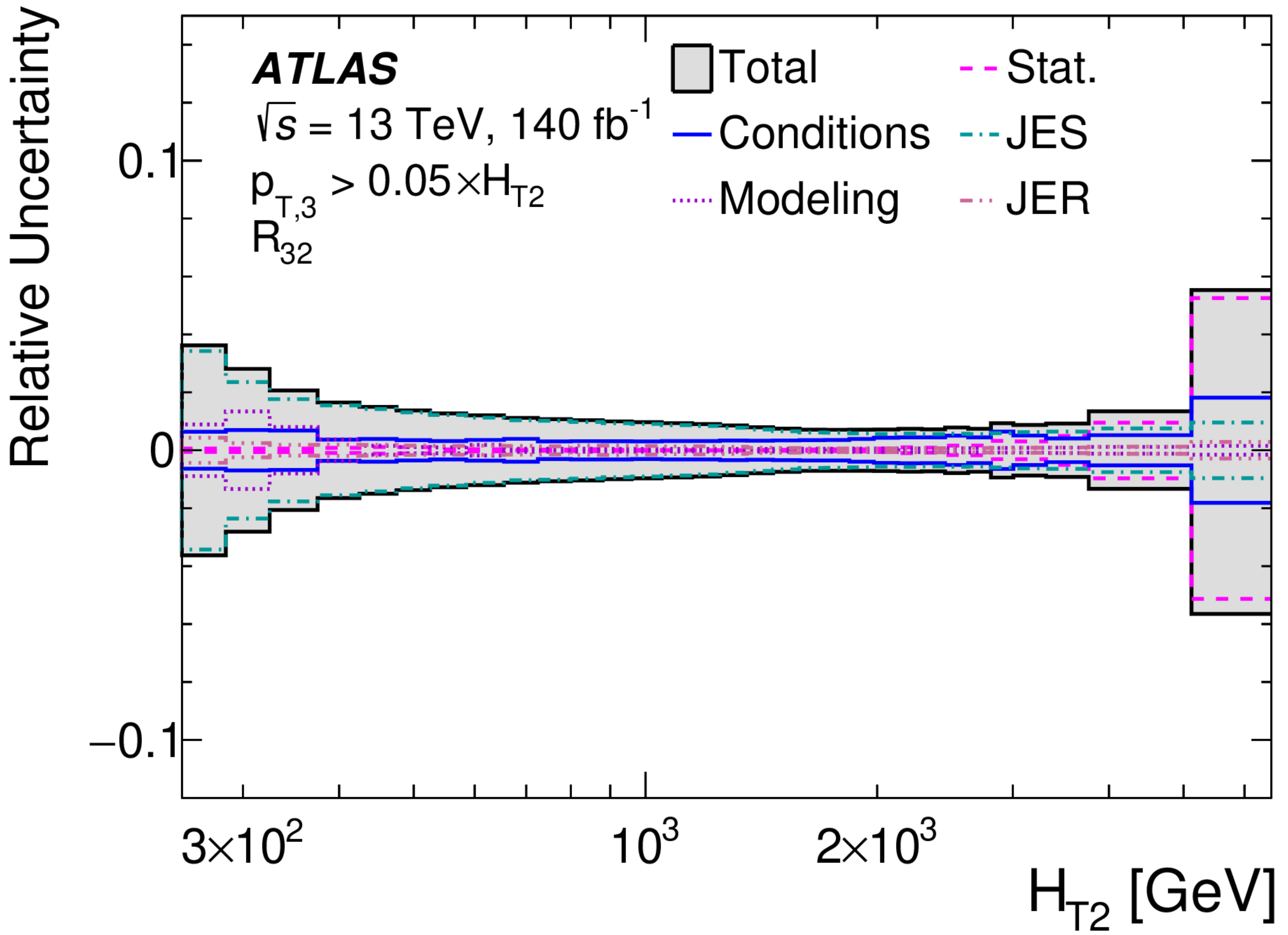
- Detailed studies to reduce the modelling dependences
- Allowed us to apply a more intellegent treatement for MC differences



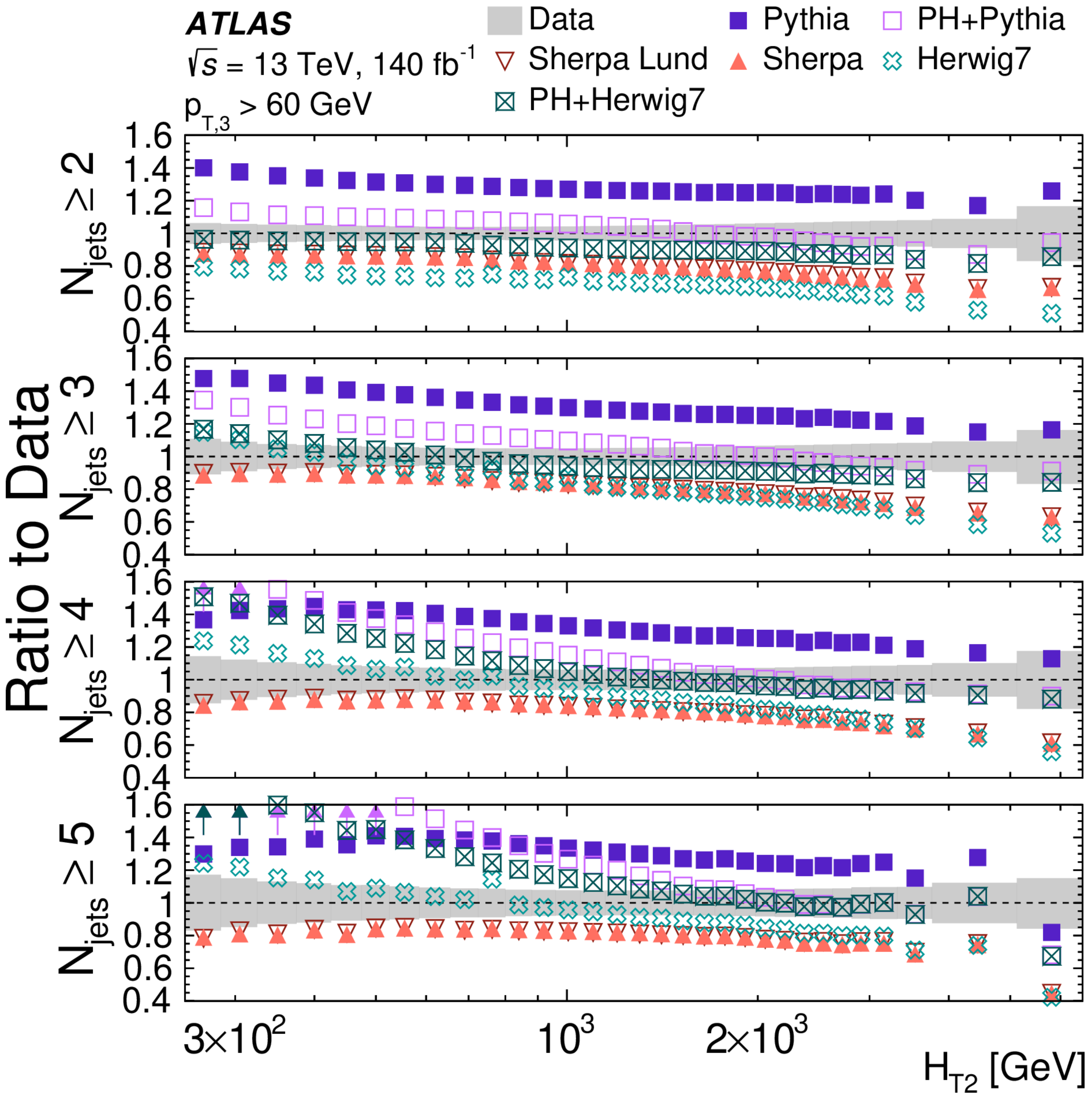
The results

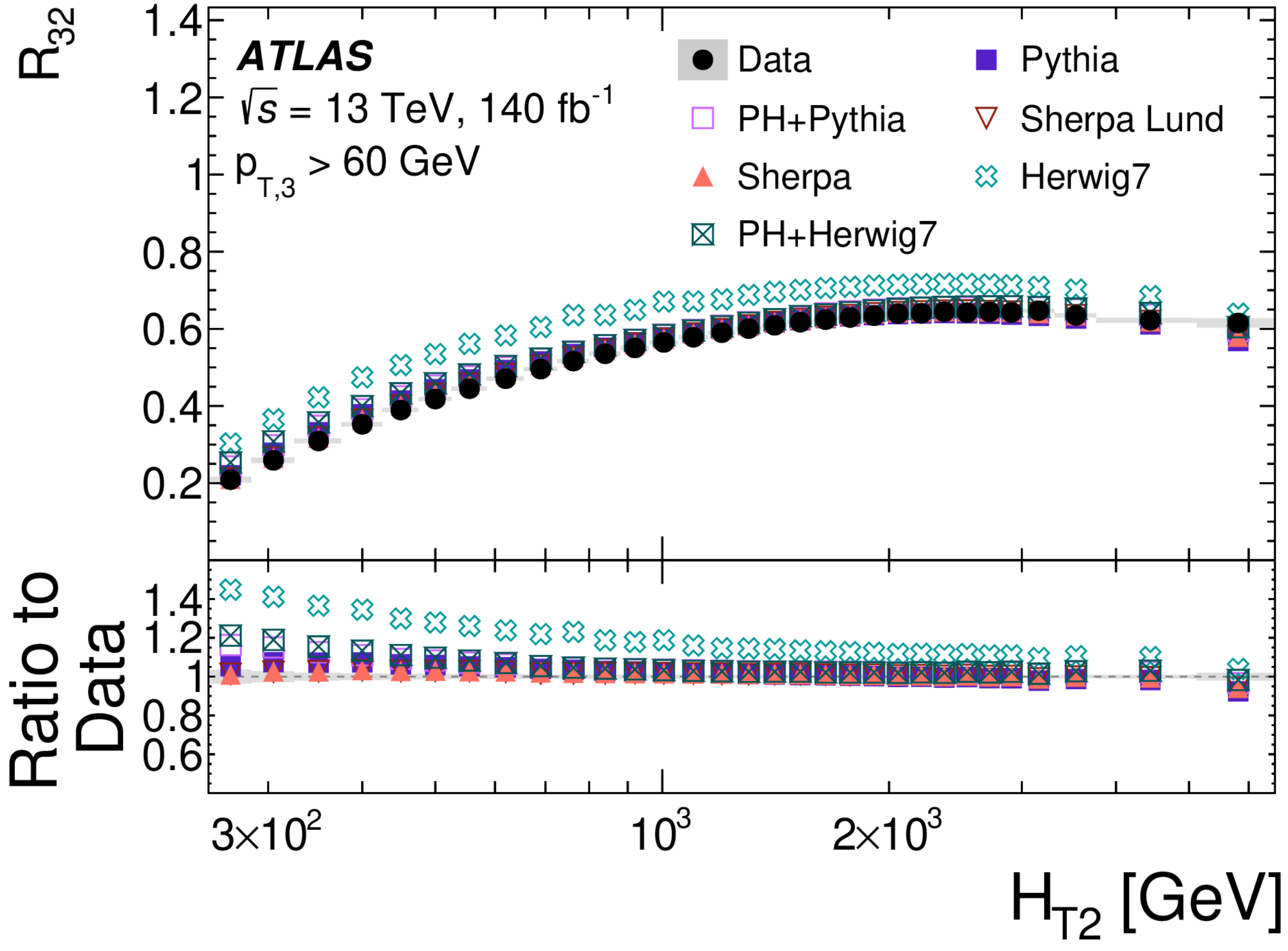
- Together with the improvements from e/p have a factor of 2-5 reduction in the uncertainties





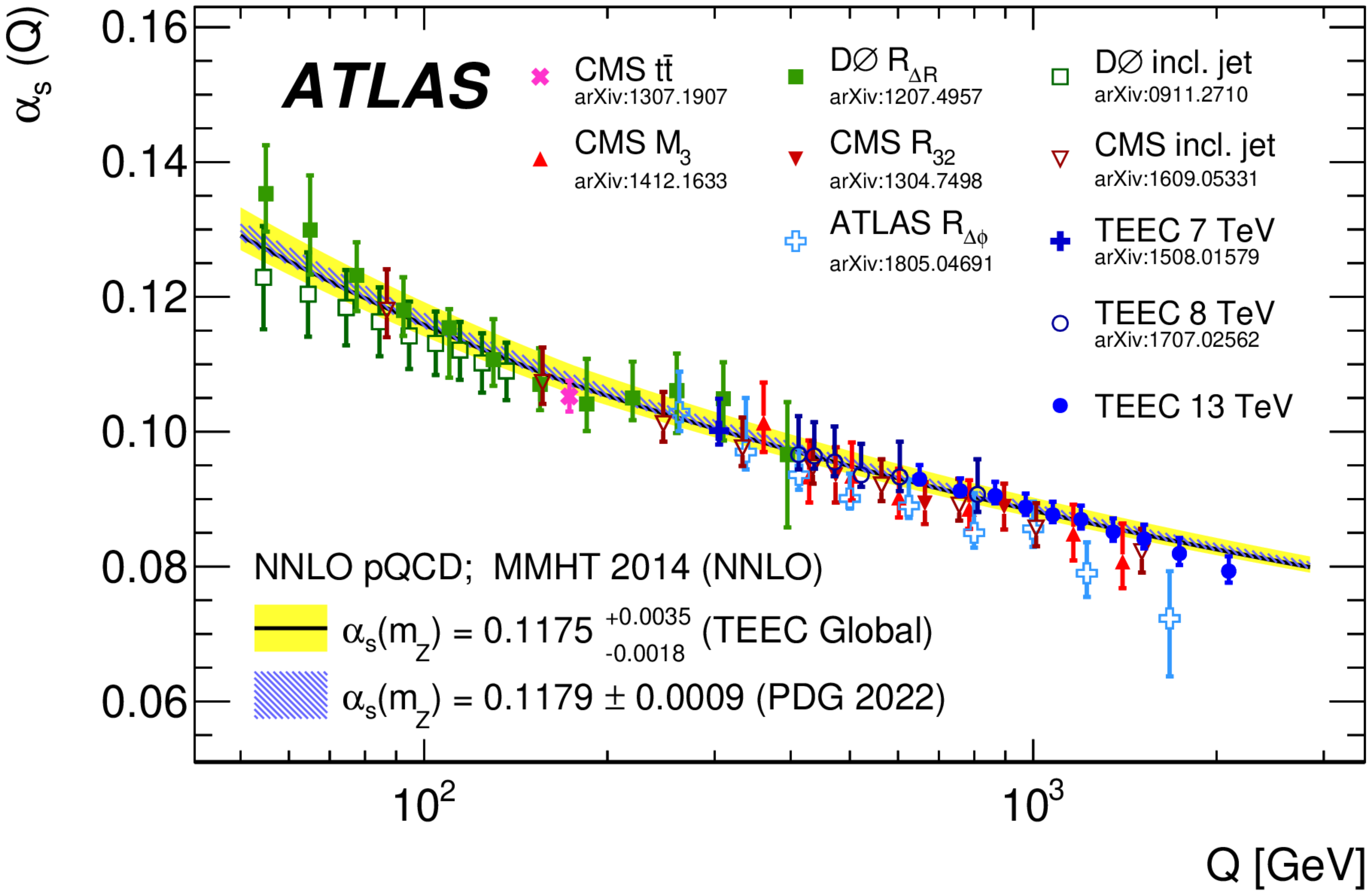
Reaching percent-level uncerainties in the cross sections

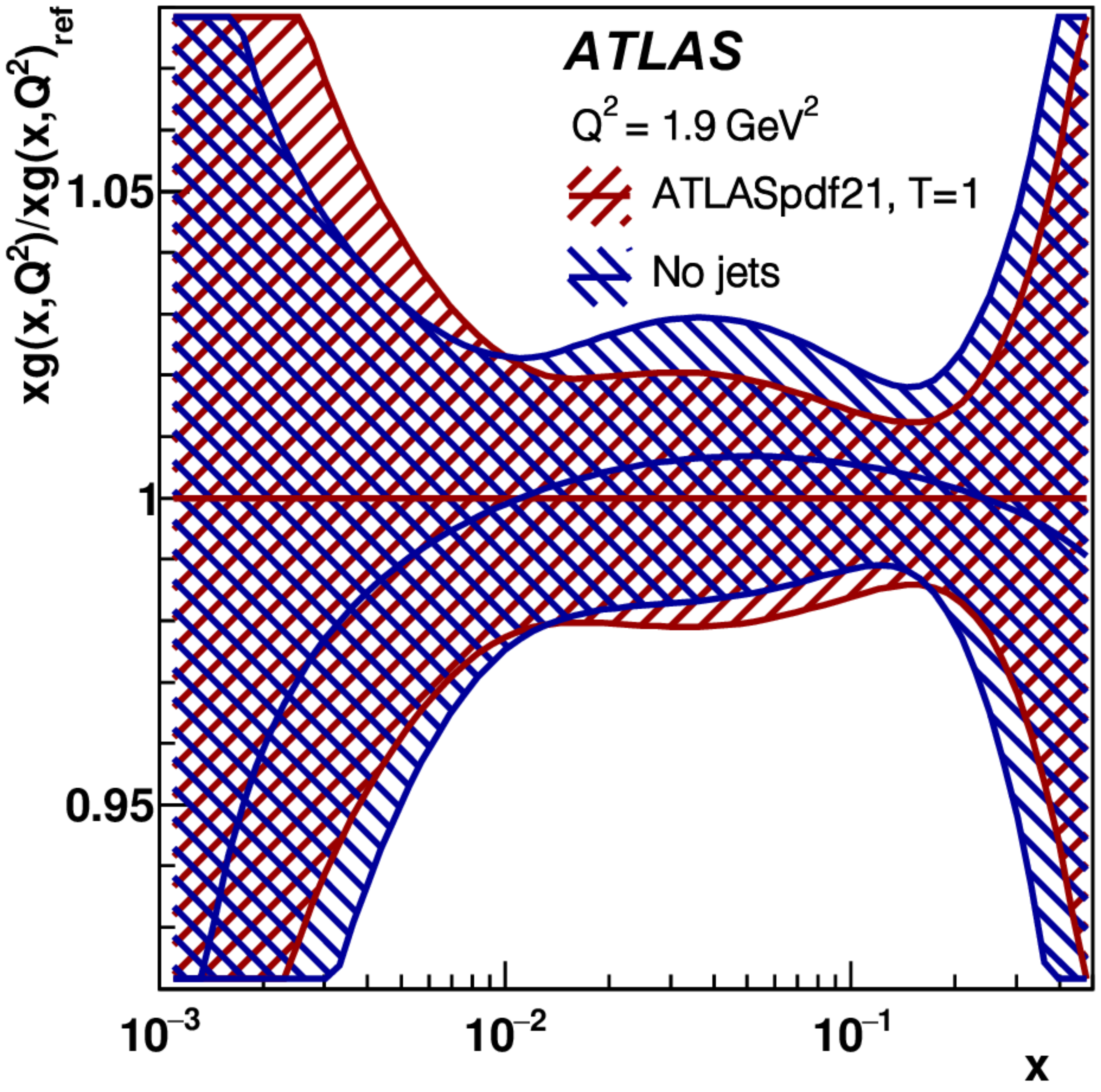




Impact of precision results

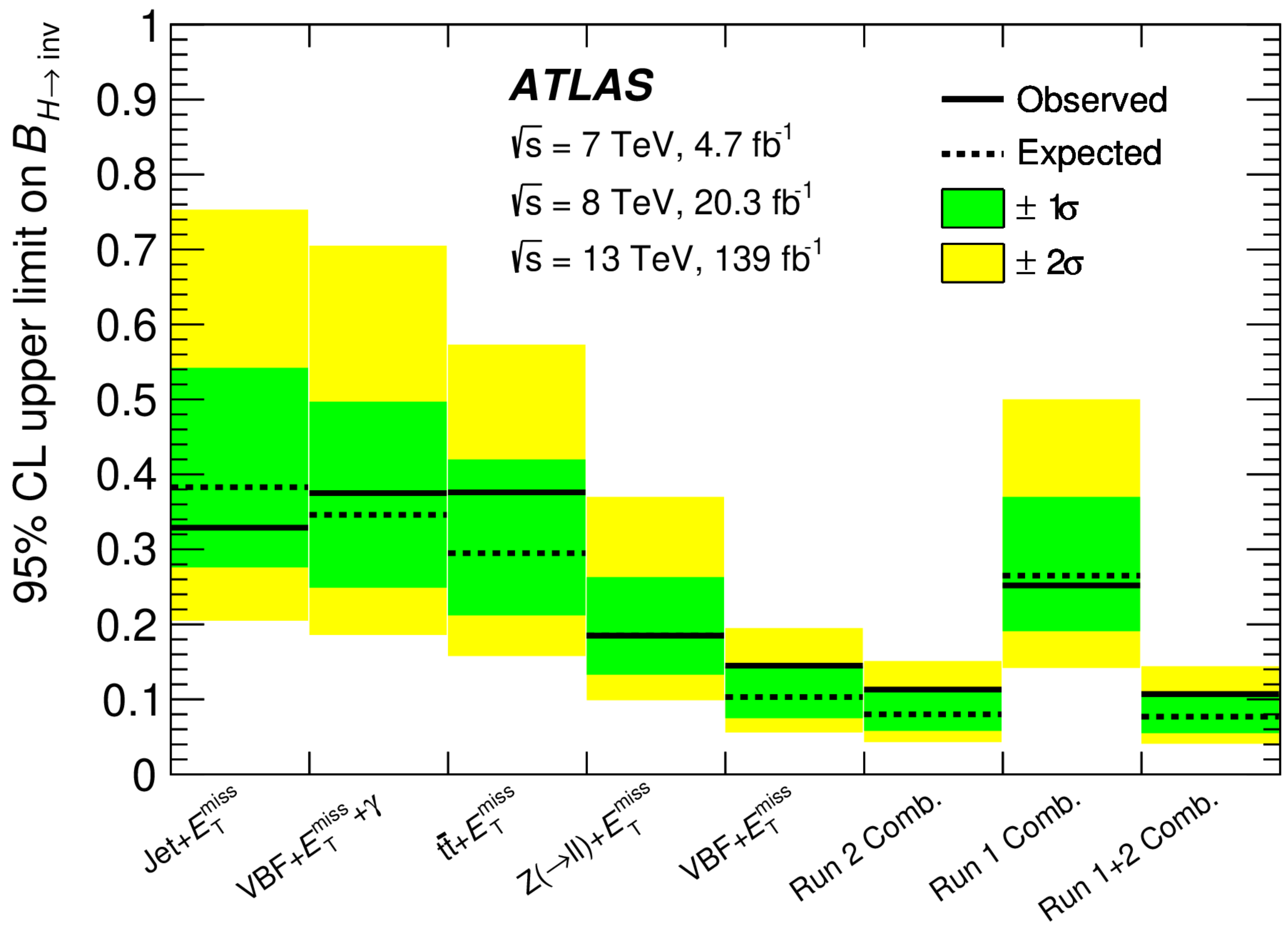
- This data has a variety of uses from PDFs to alpha_s



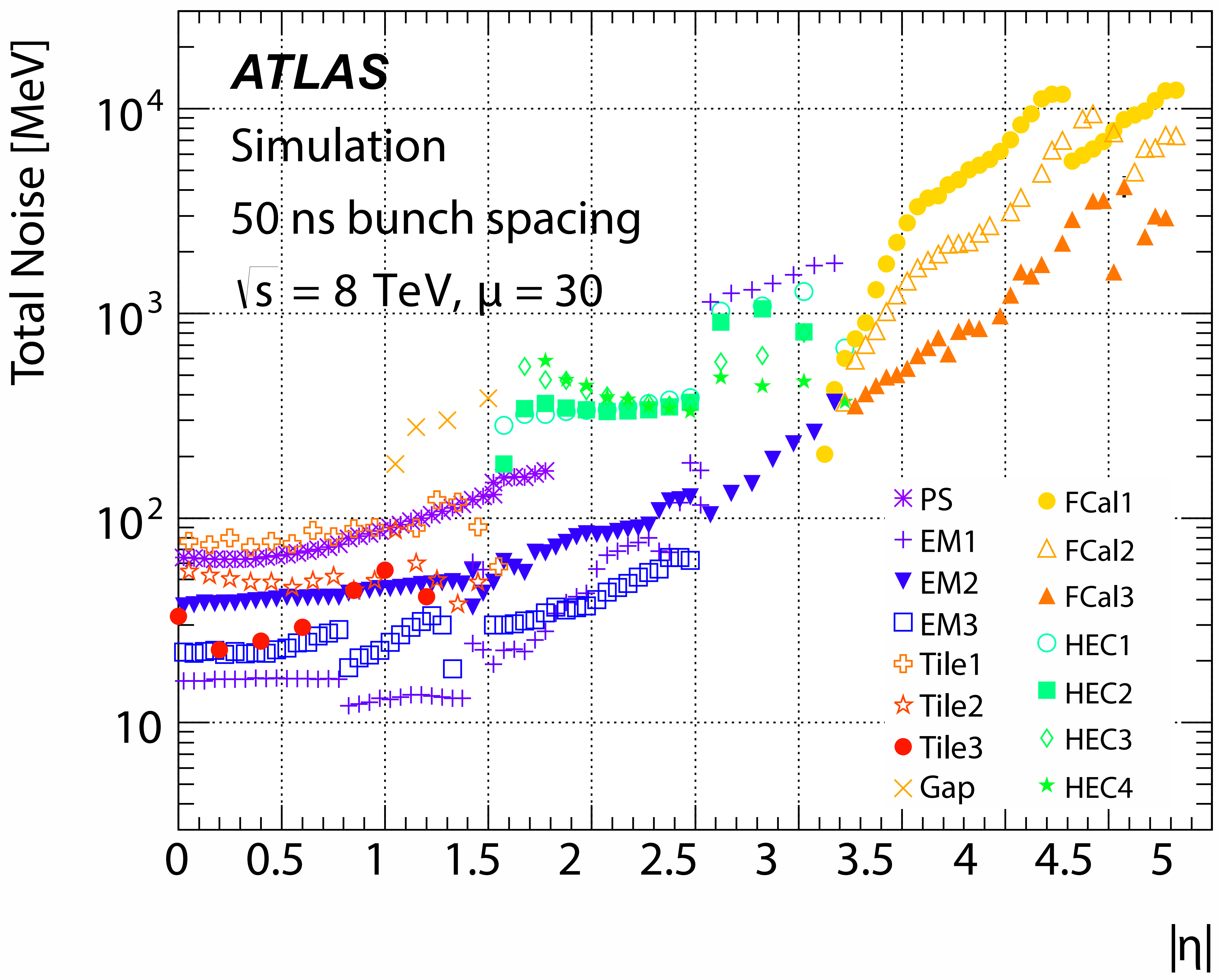


A special shout out to the forward region

- Forward region is critical - but very hard.



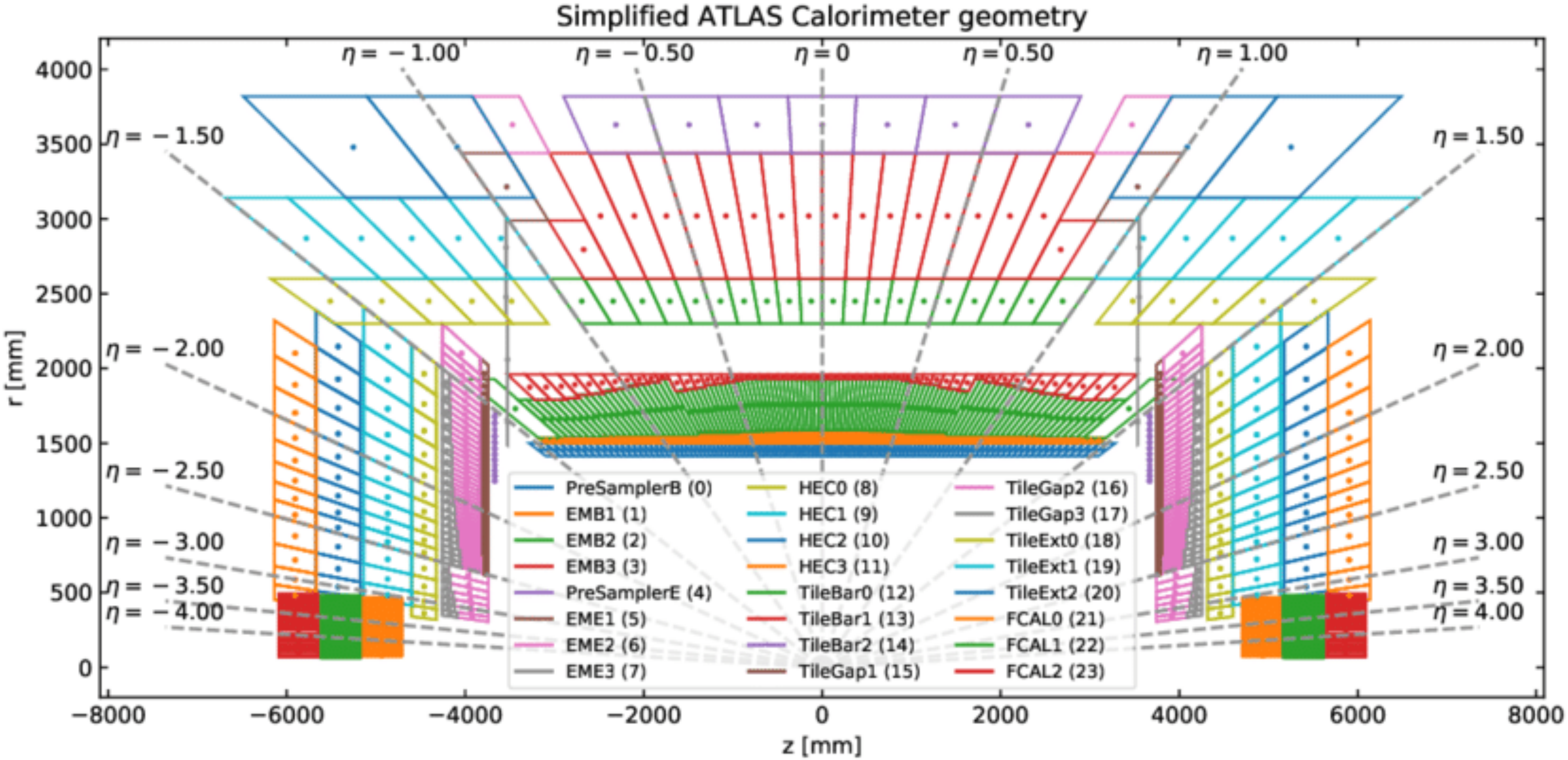
An example of Higgs to invisible



(c) $\sigma_{noise}(|\eta|)$ in 2012 (μ = 30)

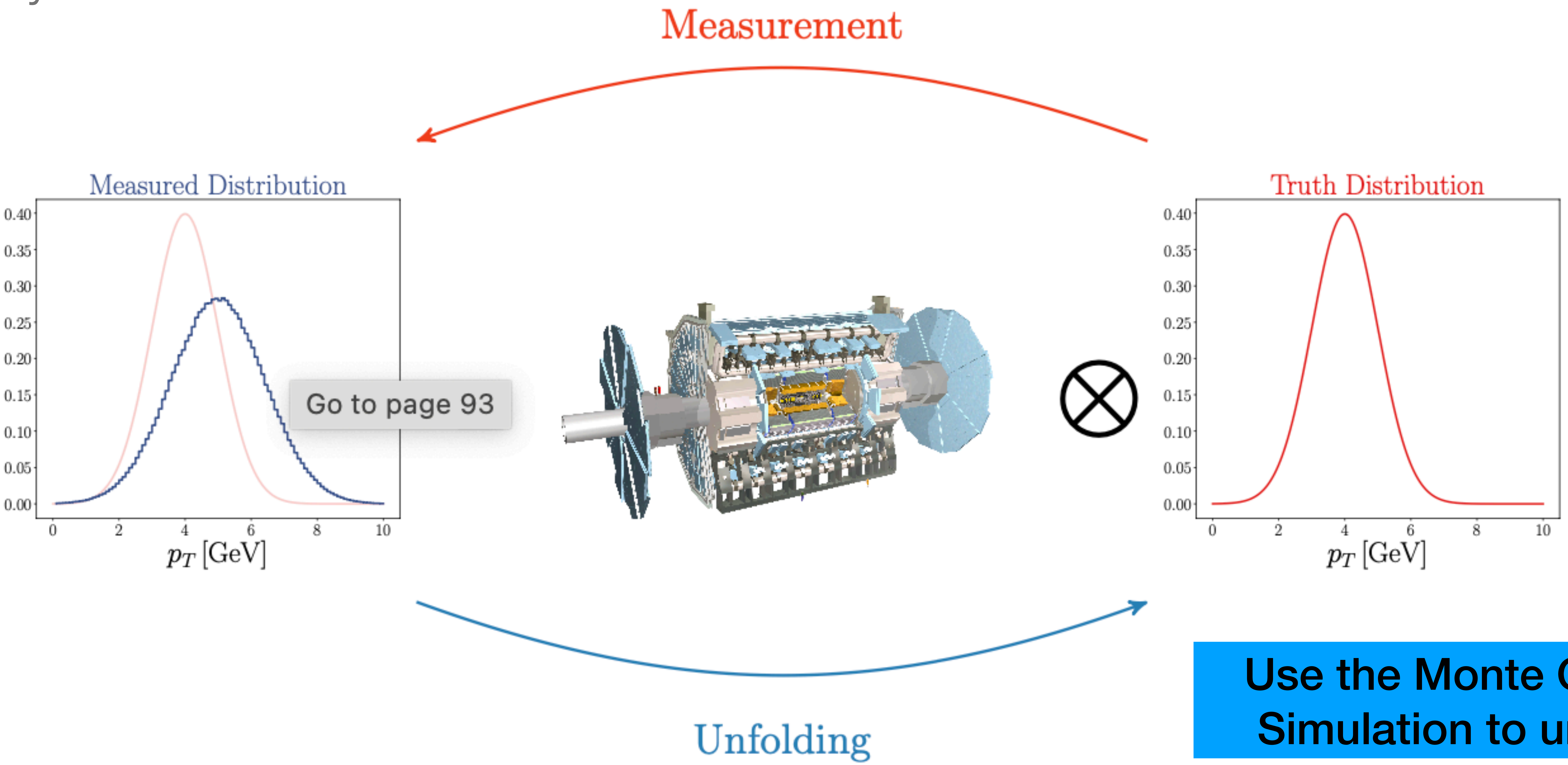
A special shout out to the forward region

- There are a few groups working on ML in the forward region, but not enough (IMO)



If the JES uncertainties are reduced, what hits us next?

- Depends on the final state but unfolding is surprising a large uncertainty in many measurements

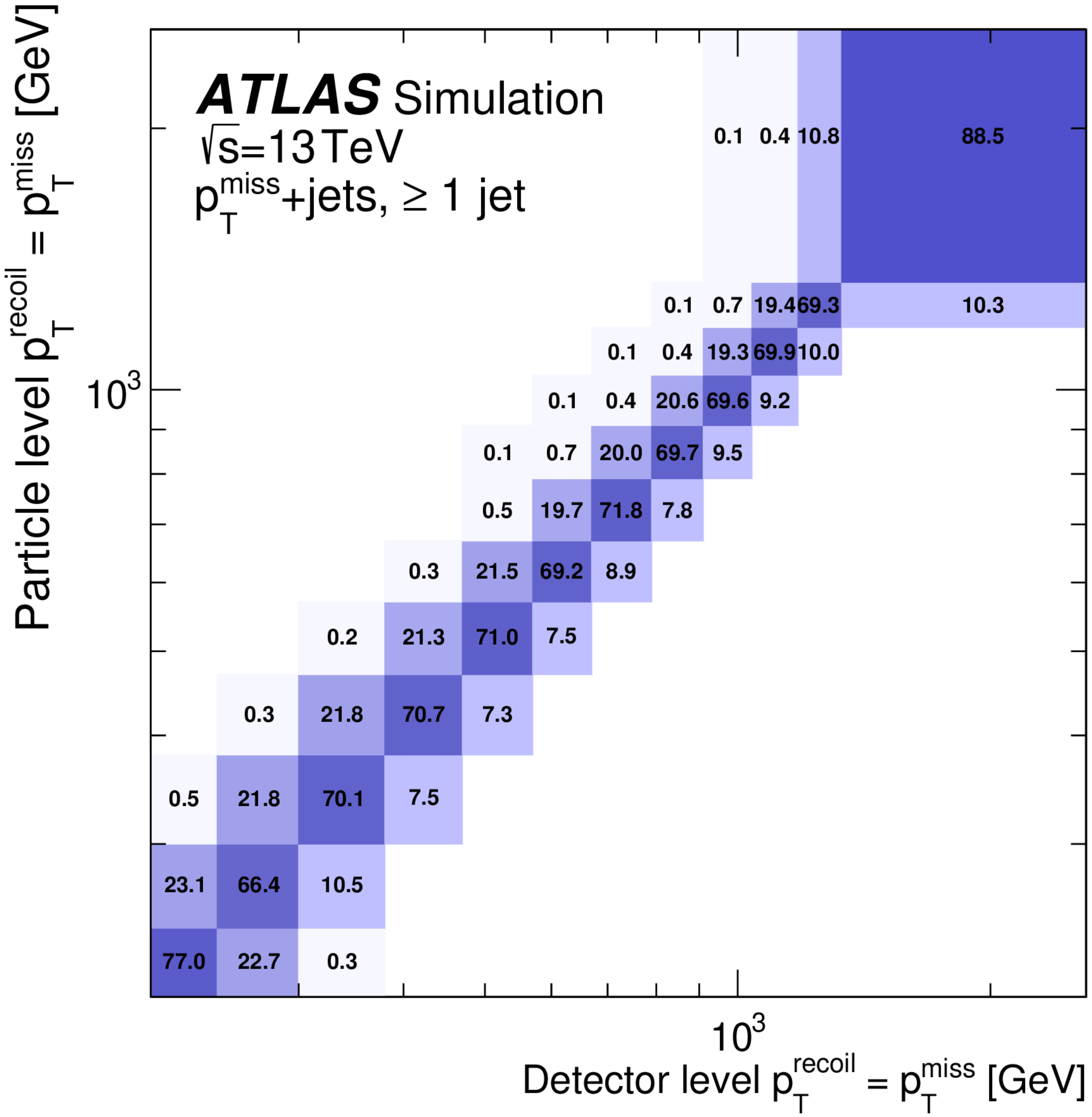


Go to page 93

Use the Monte Carlo/ Simulation to unfold

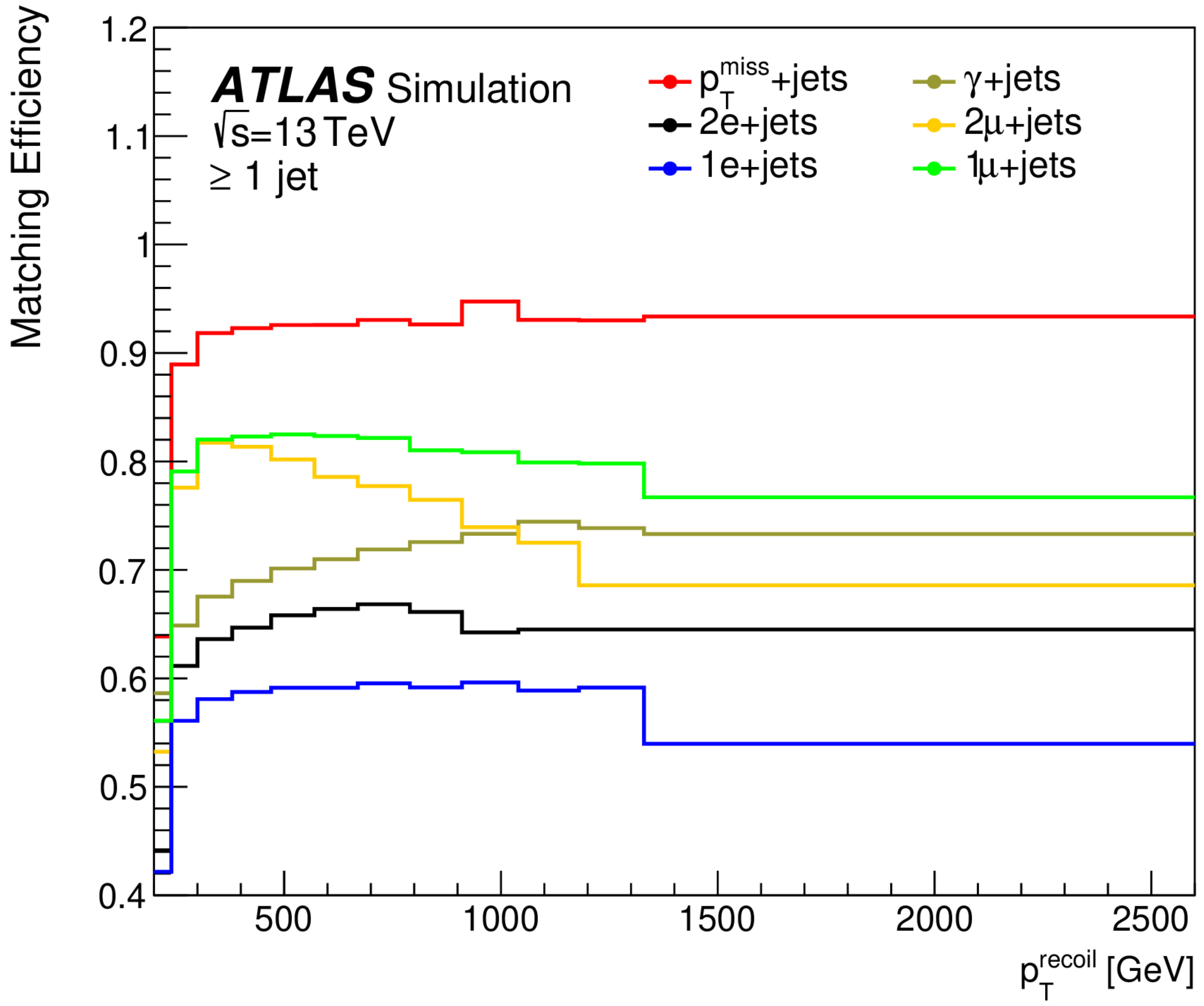
Unfolding - some basics

- Migration matrix
  - Fewer migrations is better
  - Rule of thumb is bin size = 2x the detector resolution

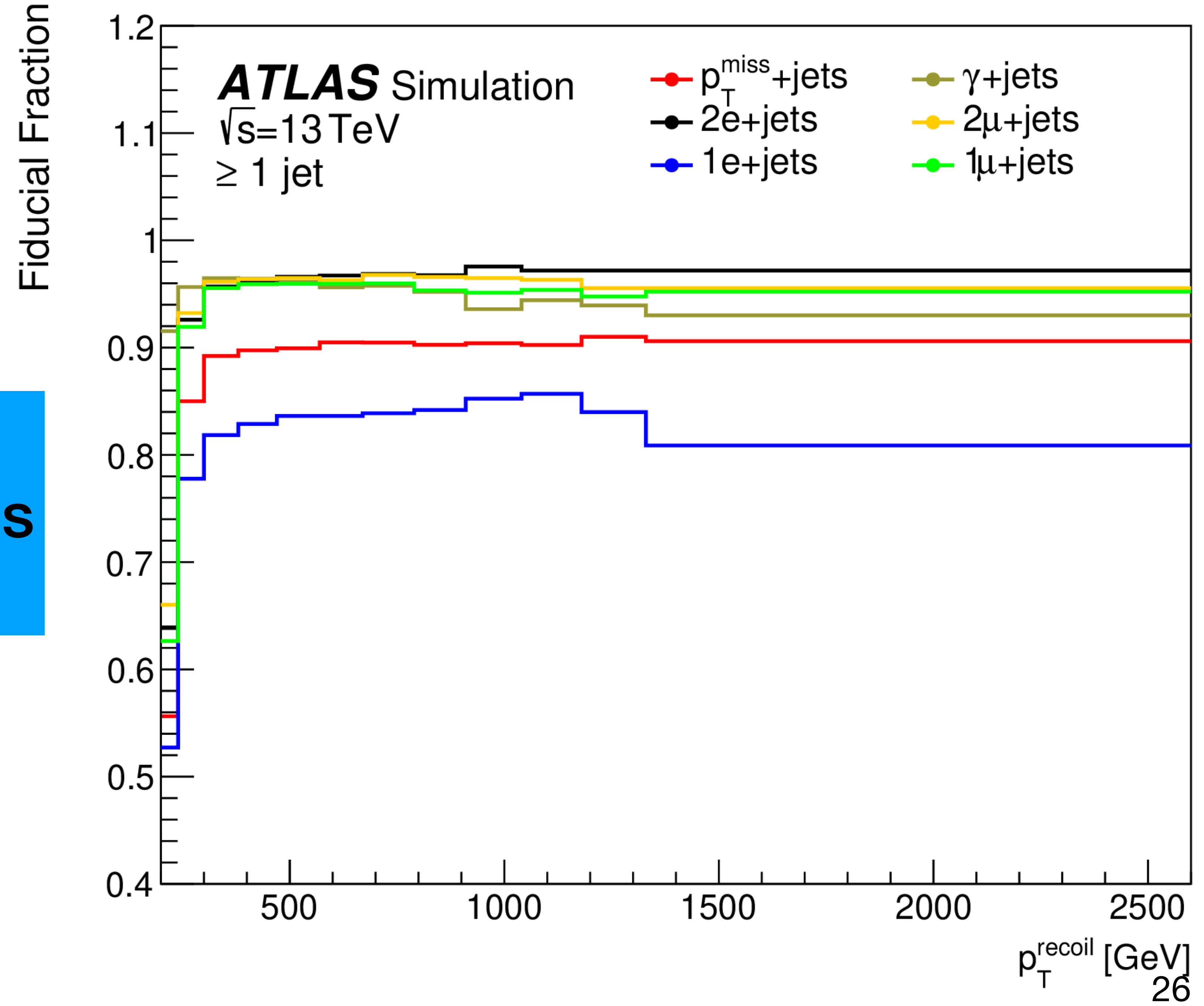


Unfolding - some basics

- Additional corrections for
  - Events in truth phase that are not at detector level (matching efficiency)
  - Events at detector-level and also in truth the phase space
  - Closer to 1.0 is better
  - These matching effiencies can dominate the unfolding uncertainty

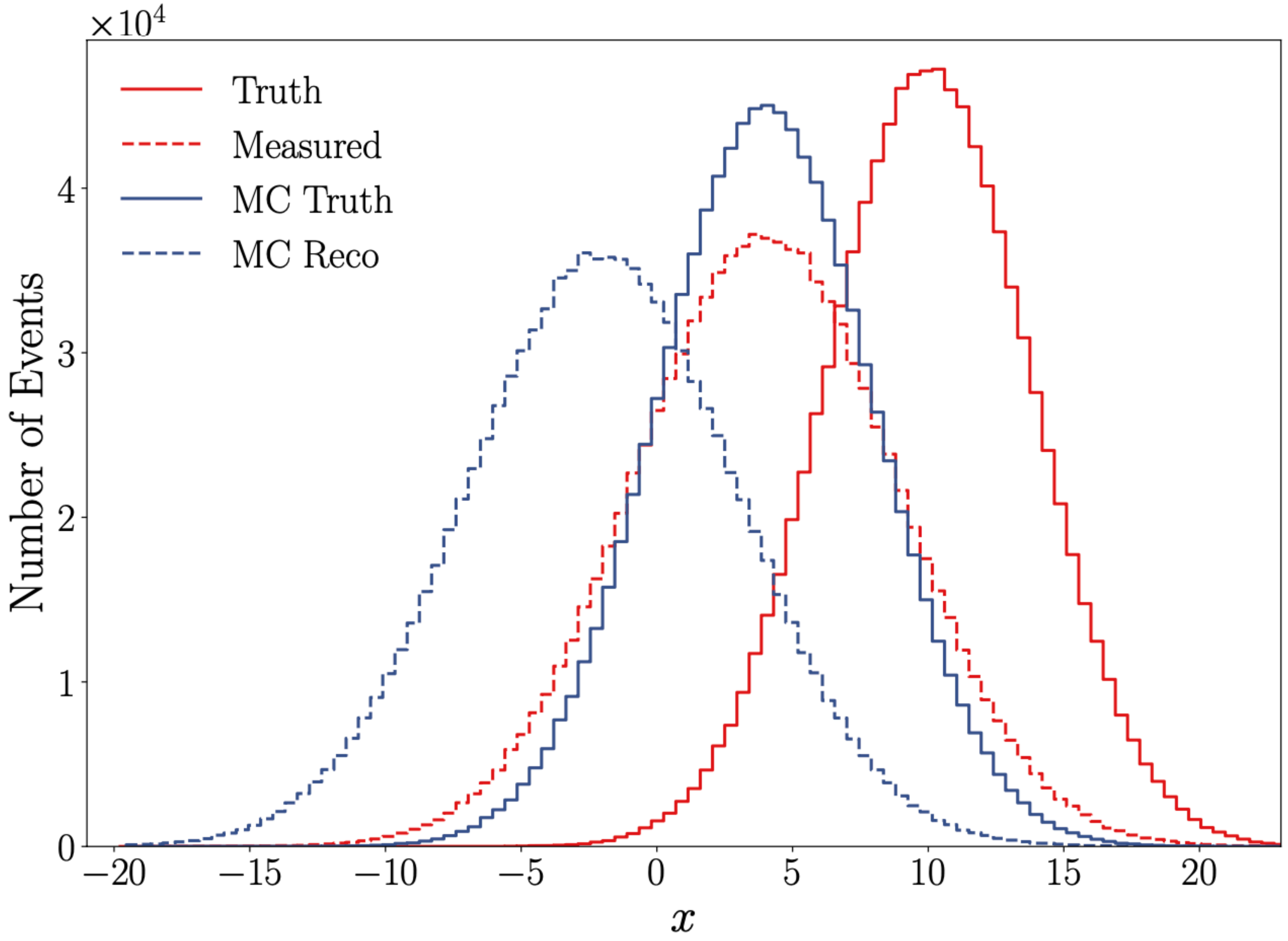


A poor choice of phase space definition can affect this

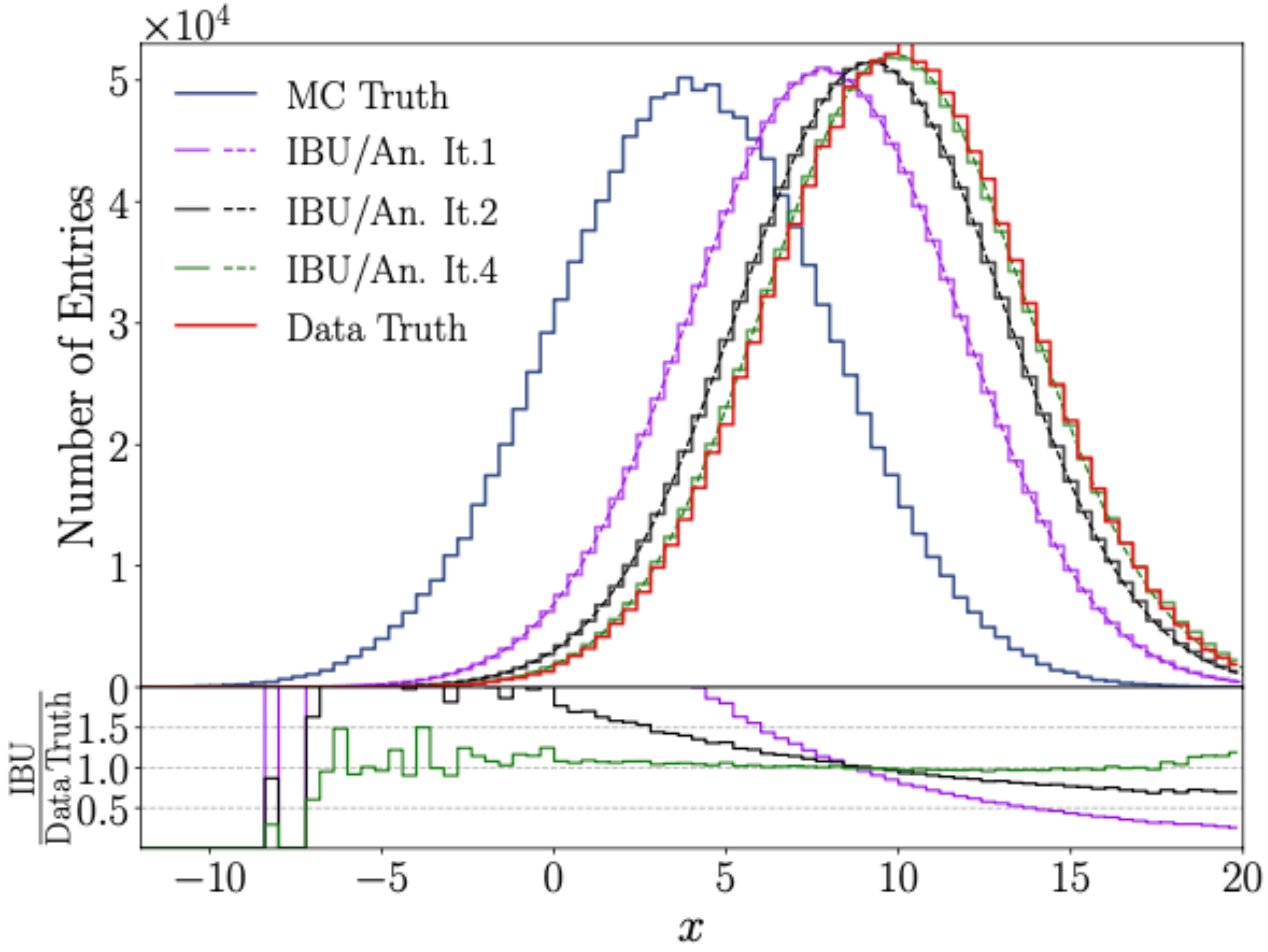


Poor choice of binning, poor detector resolutions can affect this

A poor MC/simulation model example



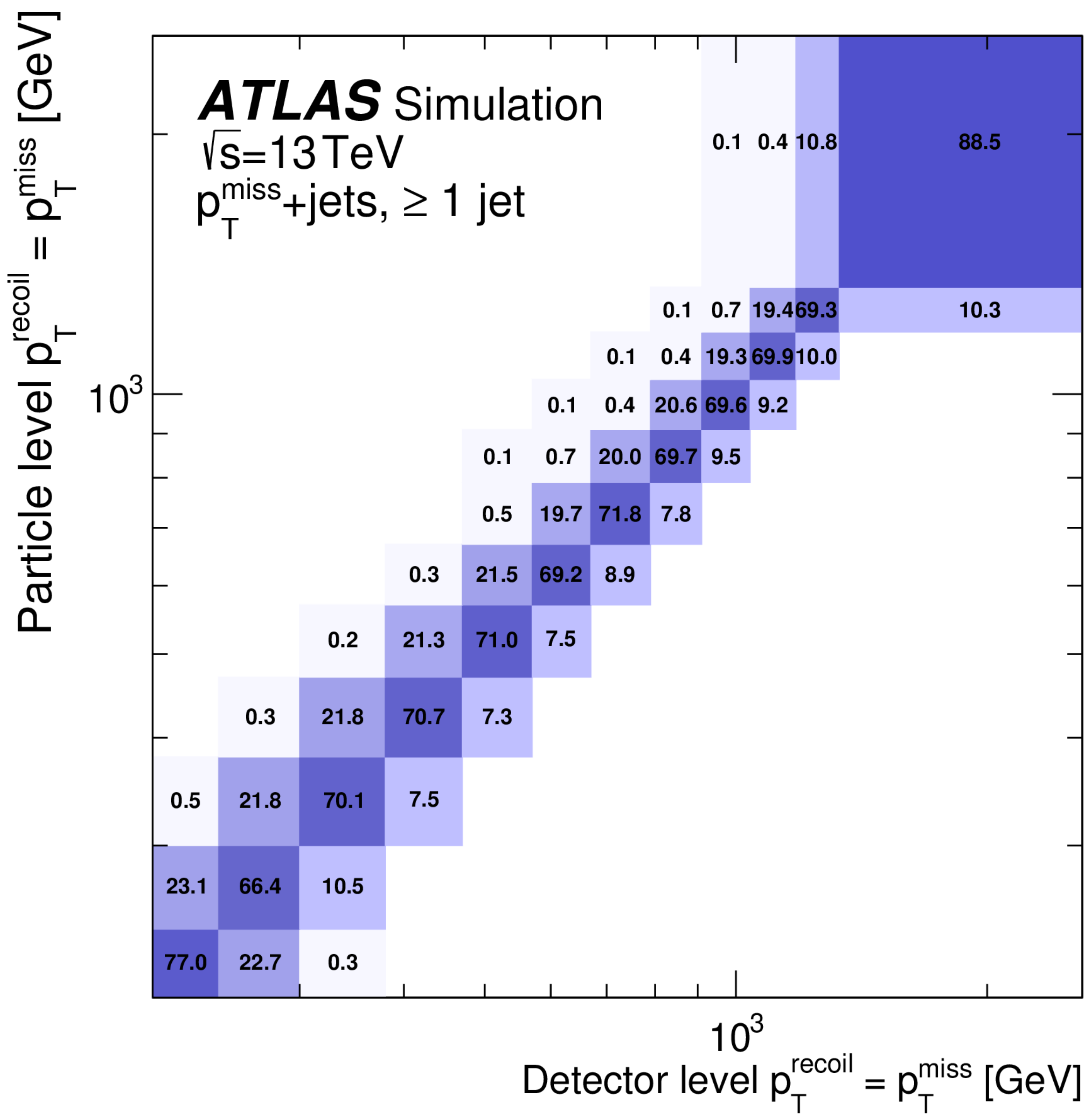
A toy example of a poor model

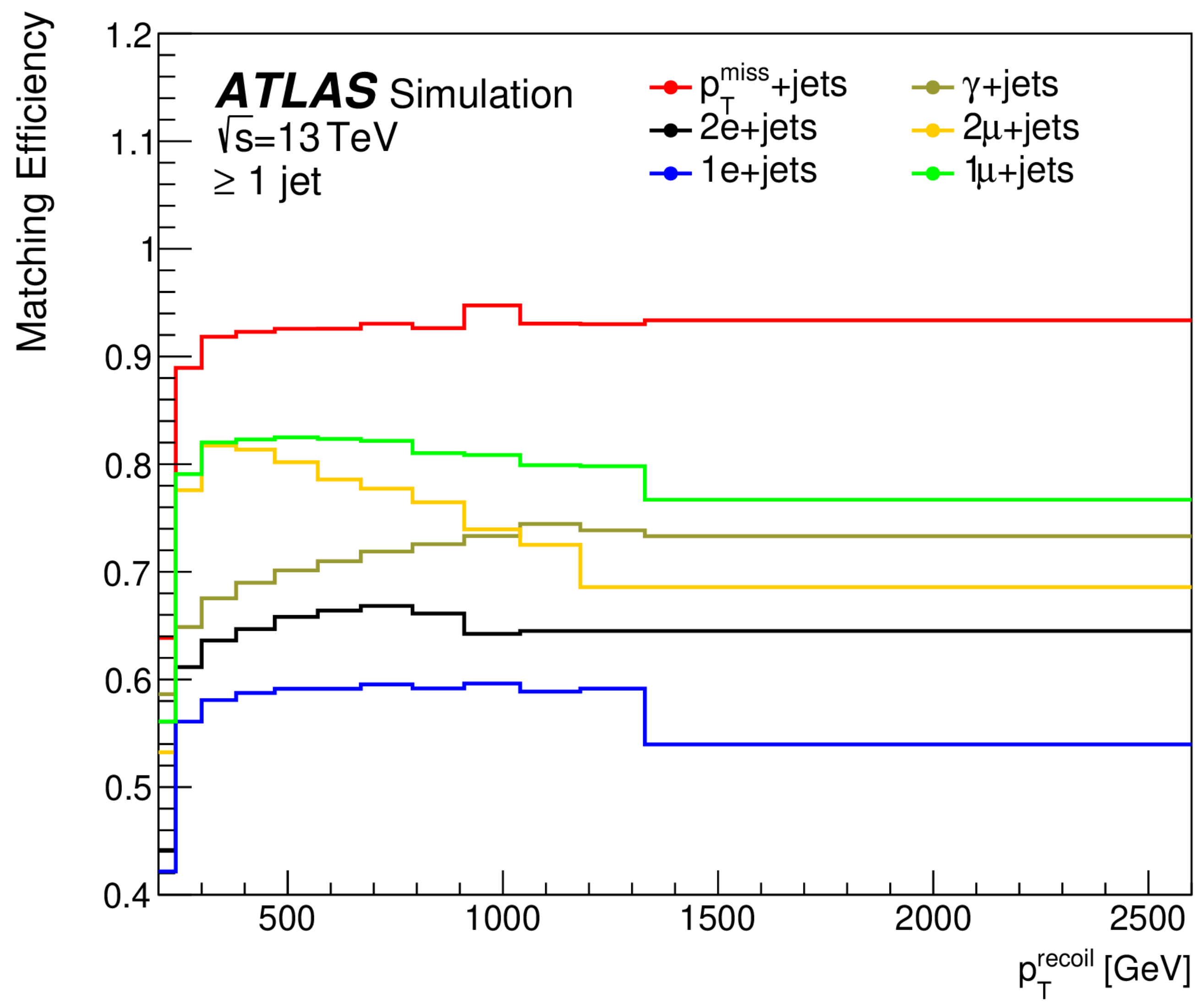


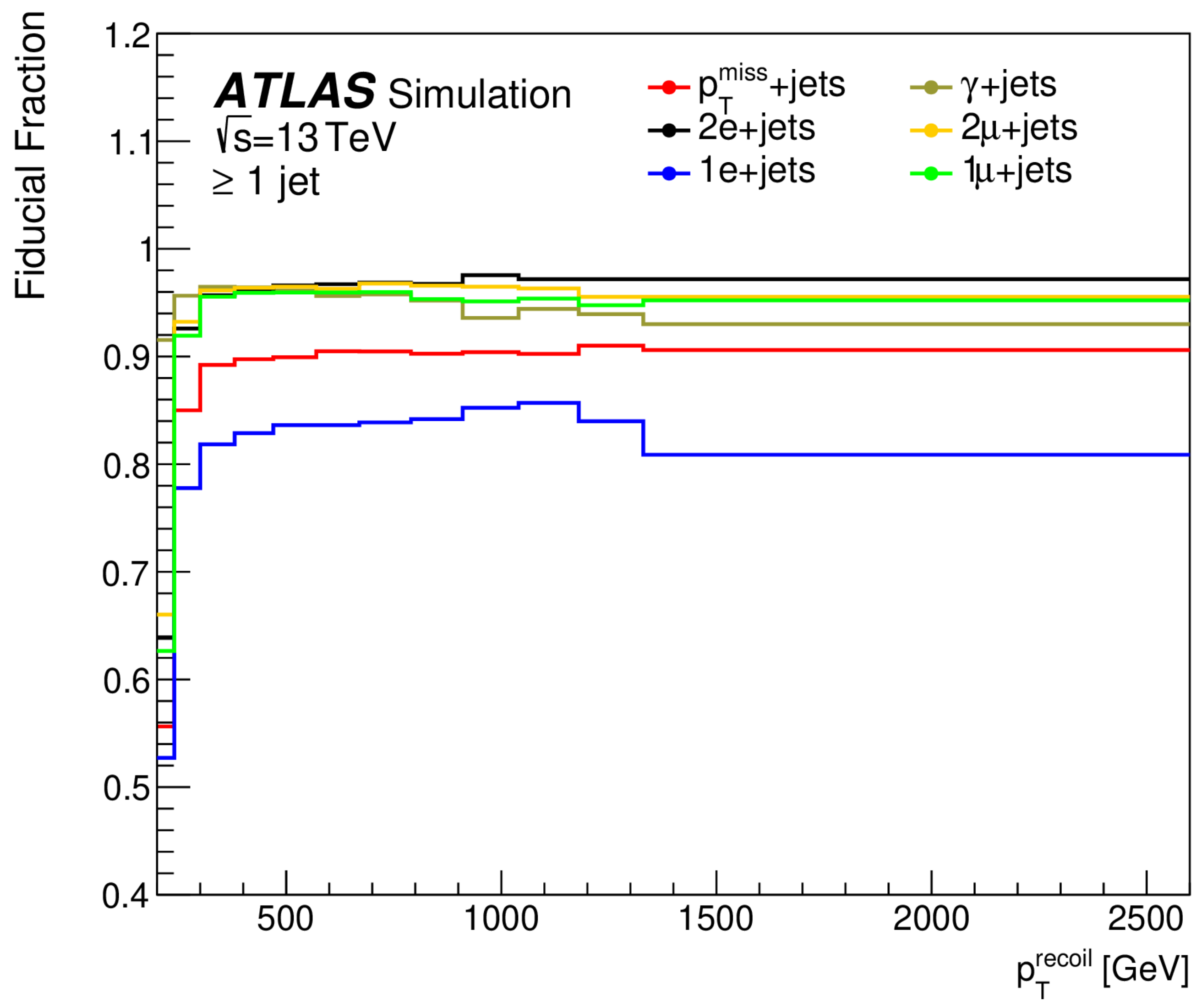
To avoid model dependences on the measure, do an iterative approach

A word of caution

- We use iterative unfolding to reduce dependencies on the underlying model





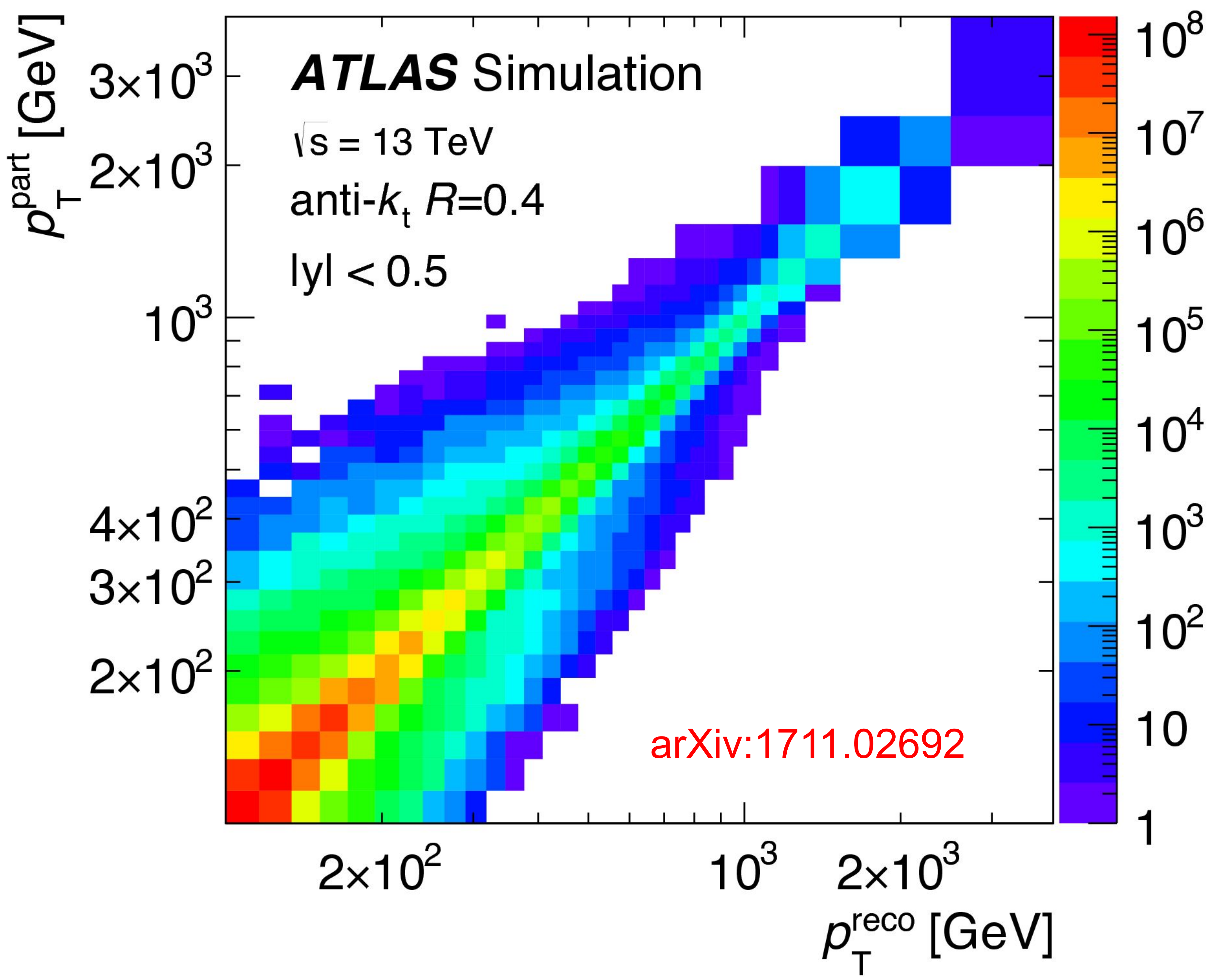


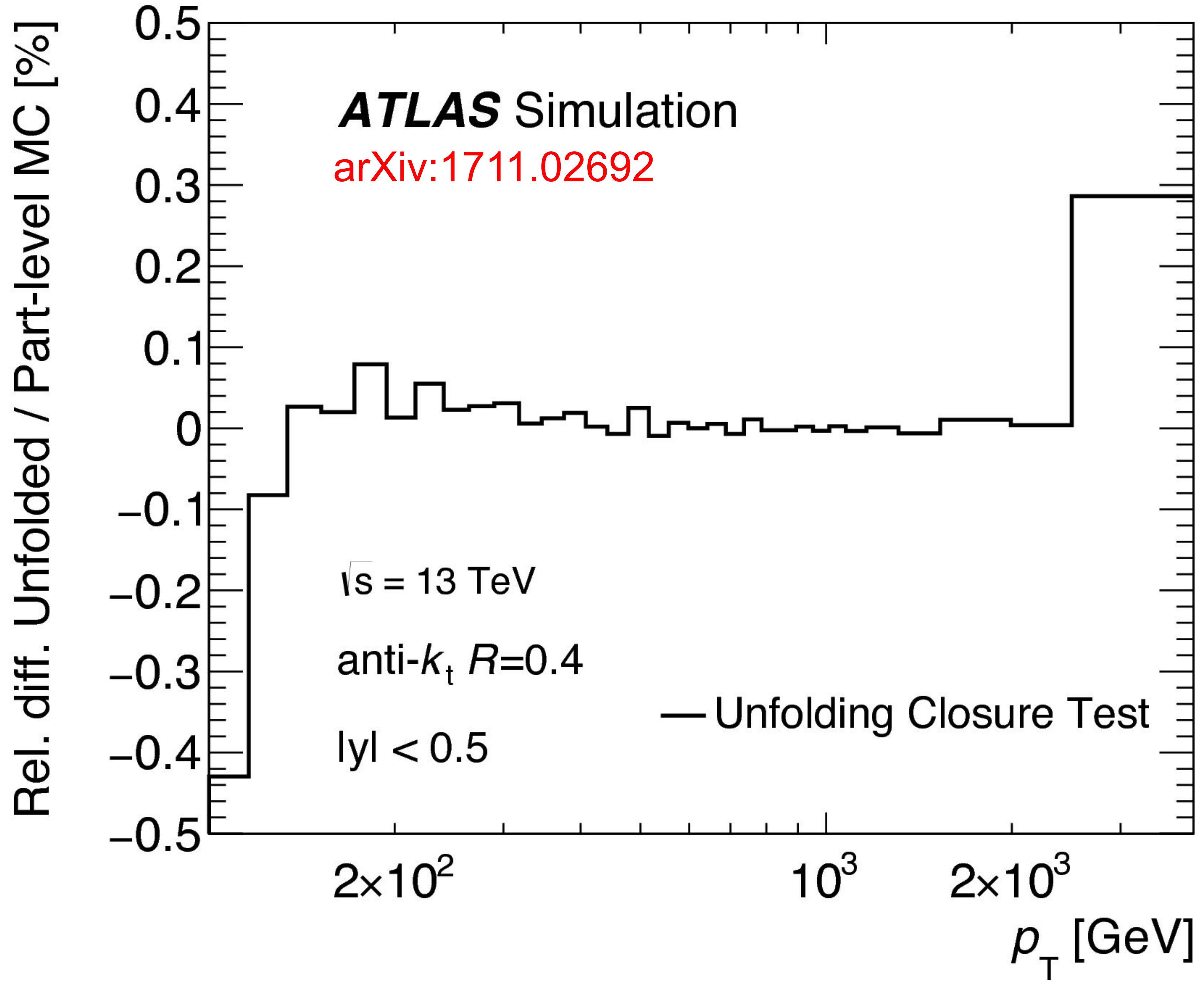
Iterations largely improve the migration matrix

These are more susceptible to model dependencies and therefore the optimal phase space and measurement definition for the question at hand is critical

Unfolding uncertainties

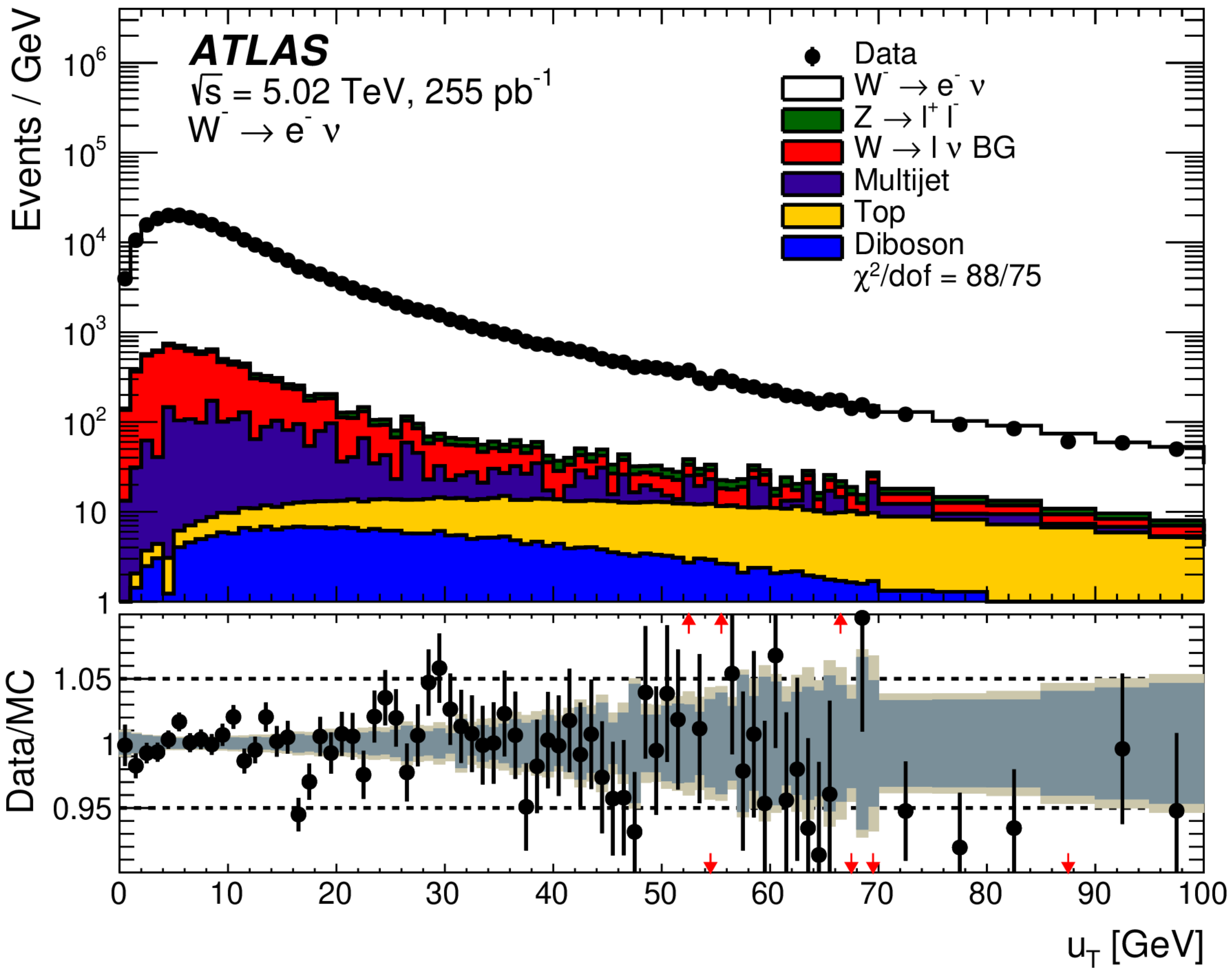
- Reweight to test data/MC differences
- We worry about ‘hidden variables’, i.e. differences in MCs in observables that we are not measuring - test this by unfolding with a different MC, reweighted to say match the jet energy scale

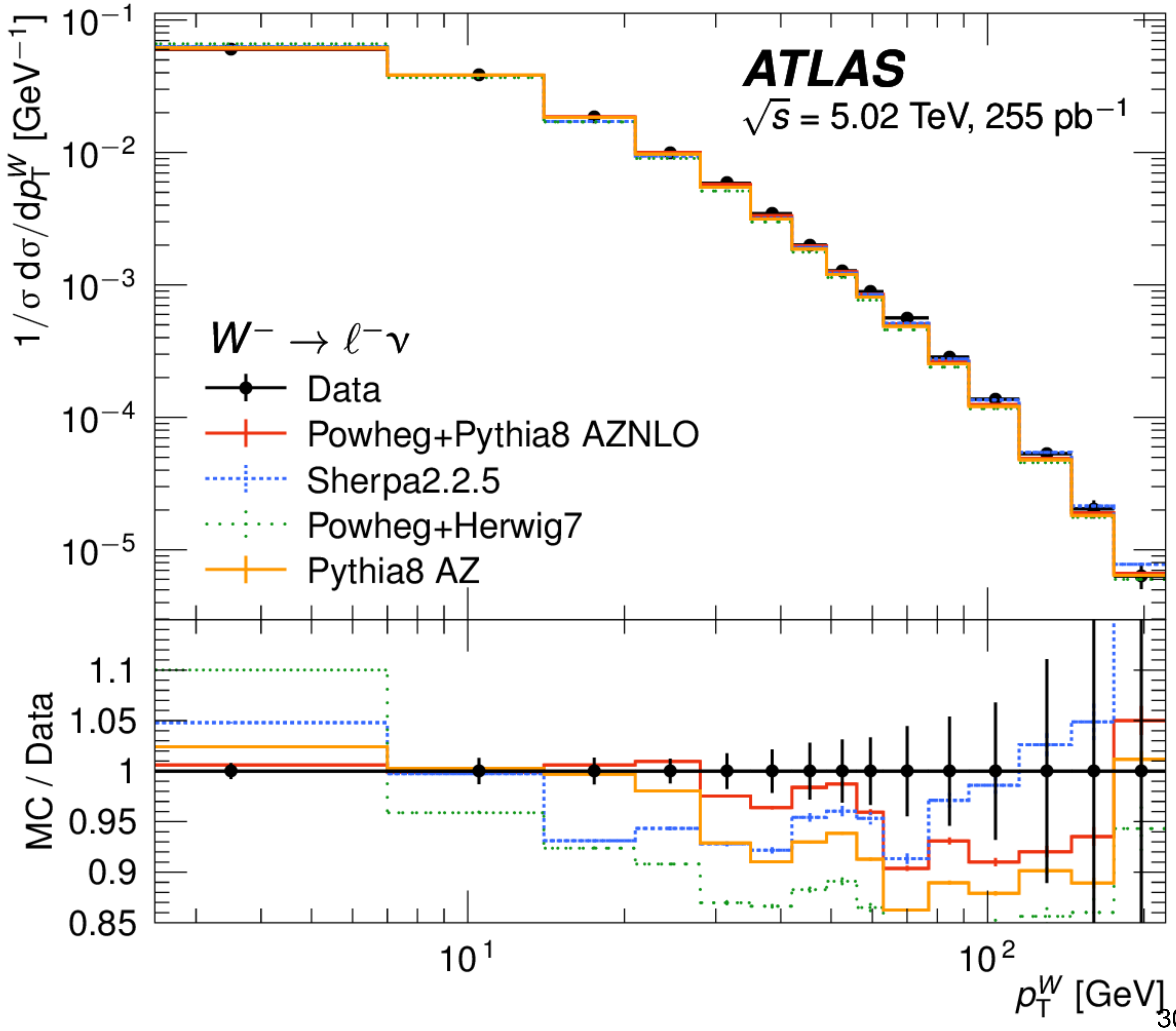




An example of a challenge

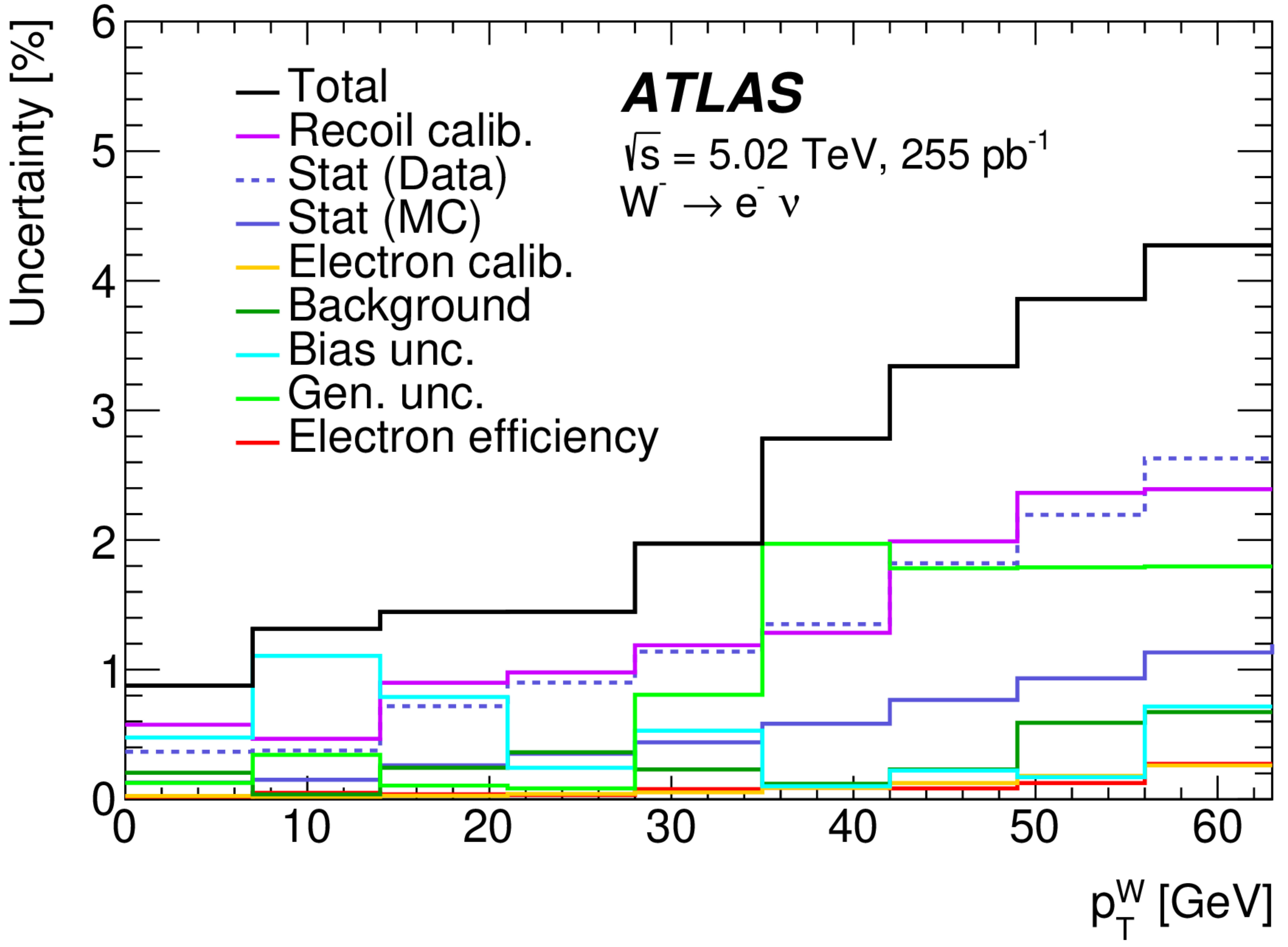
- A measurement of the W pT using low mu data





W pT measurement

- At low pT, the unfolding uncertainties are large
- Is there a better way to do this?

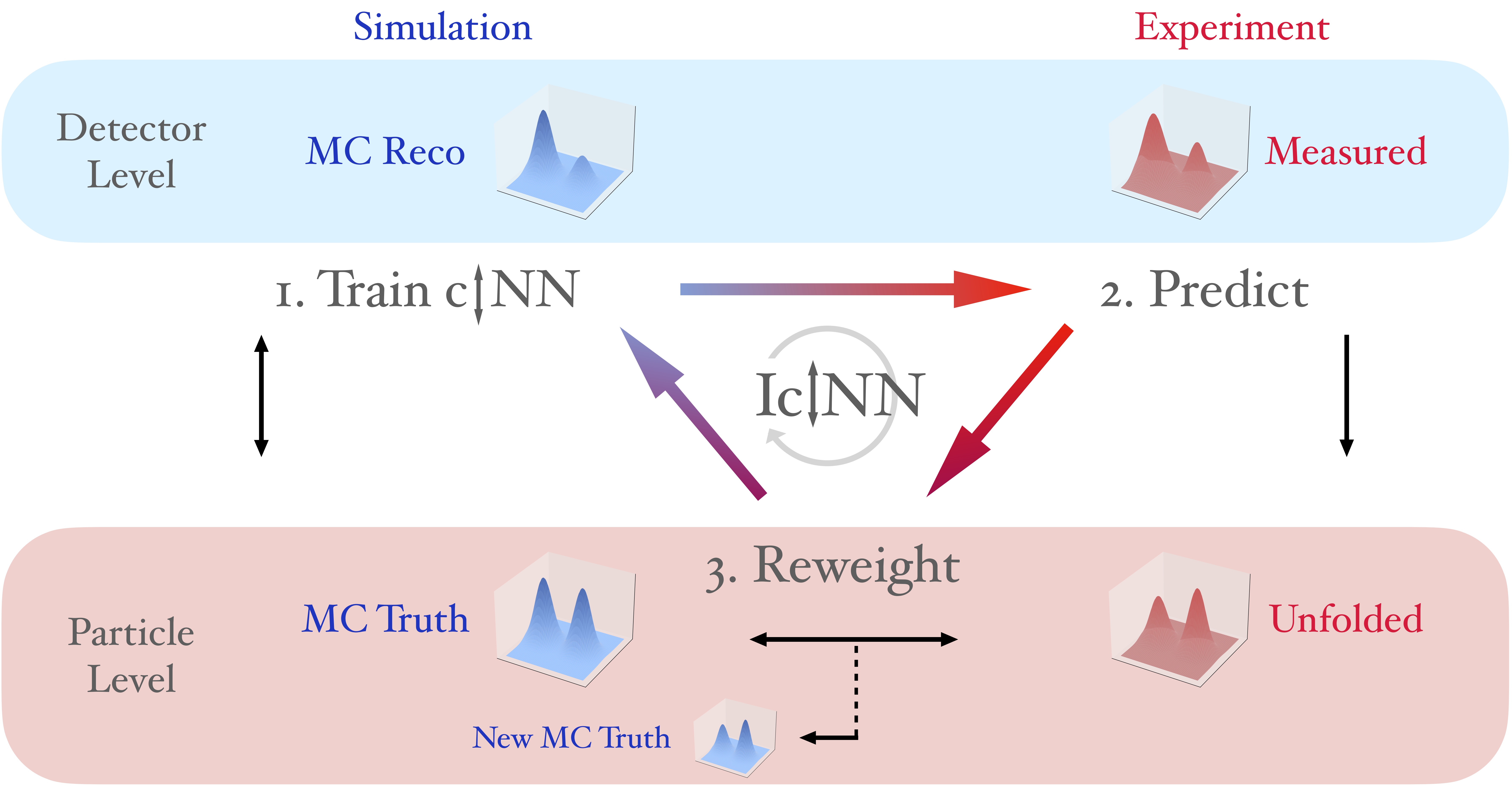


The future of data…

- Are cross section measurements something of the past?
- An example of machine learning unfolding

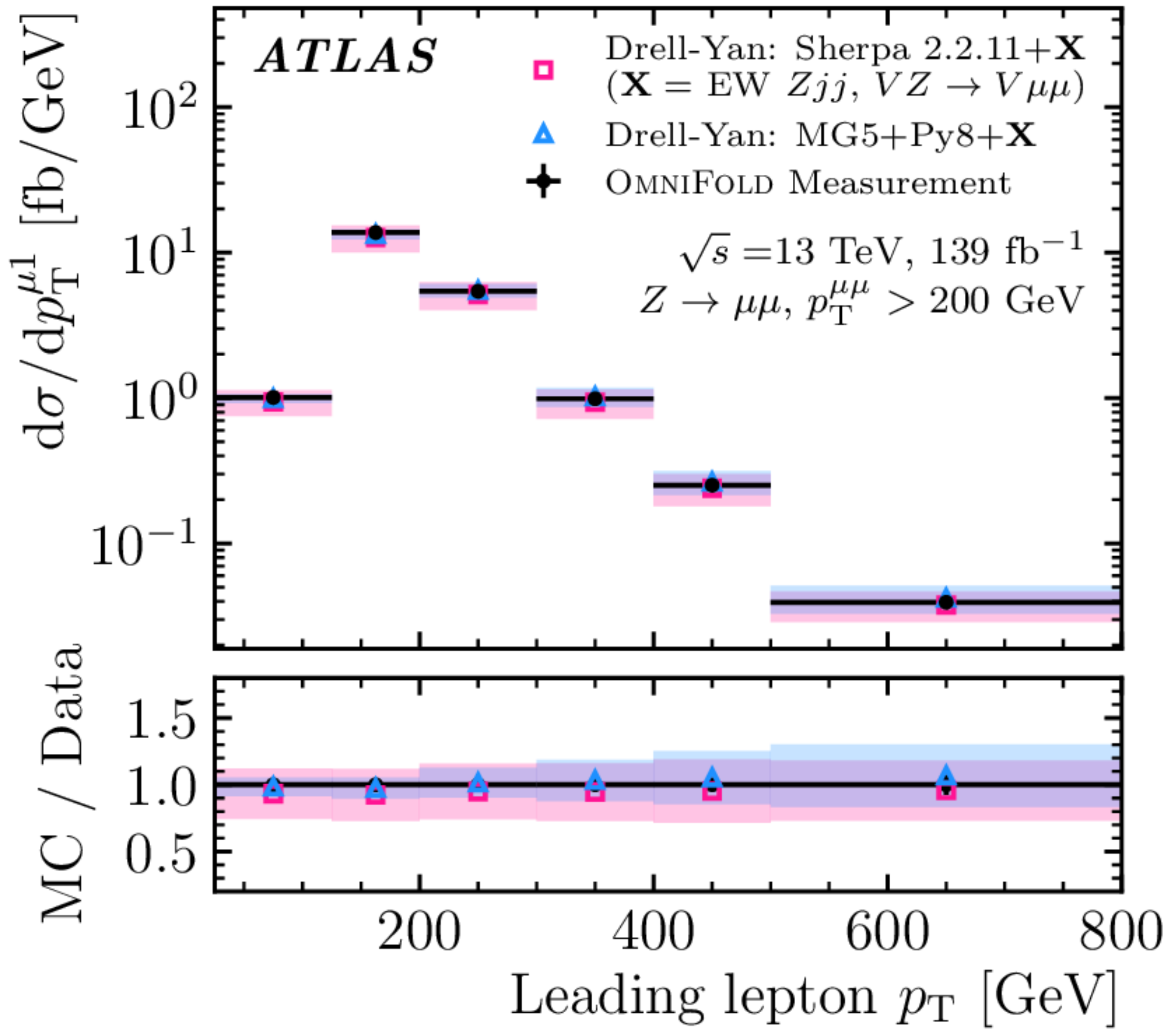
The future of data…

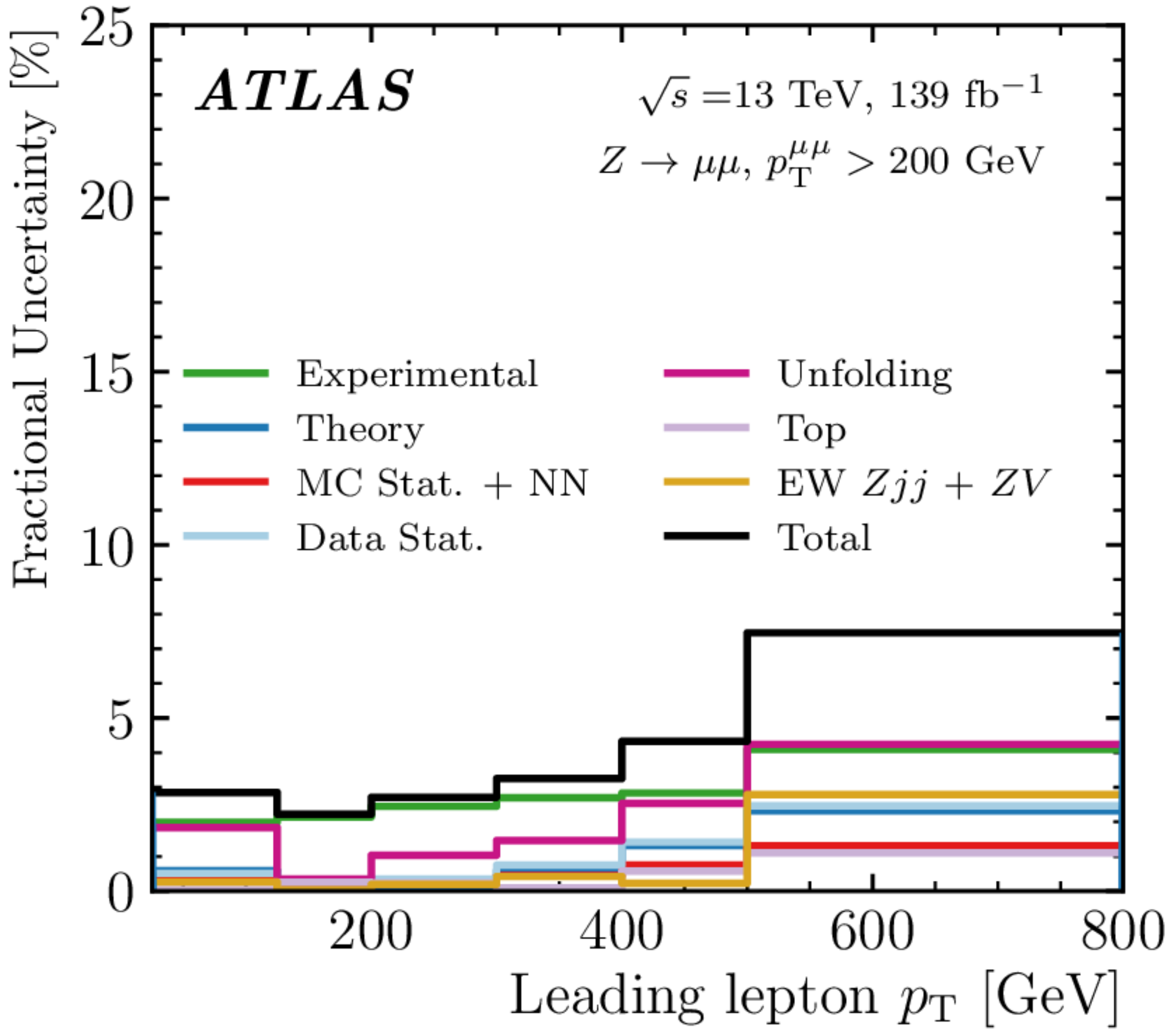
- Several alternative approaches available

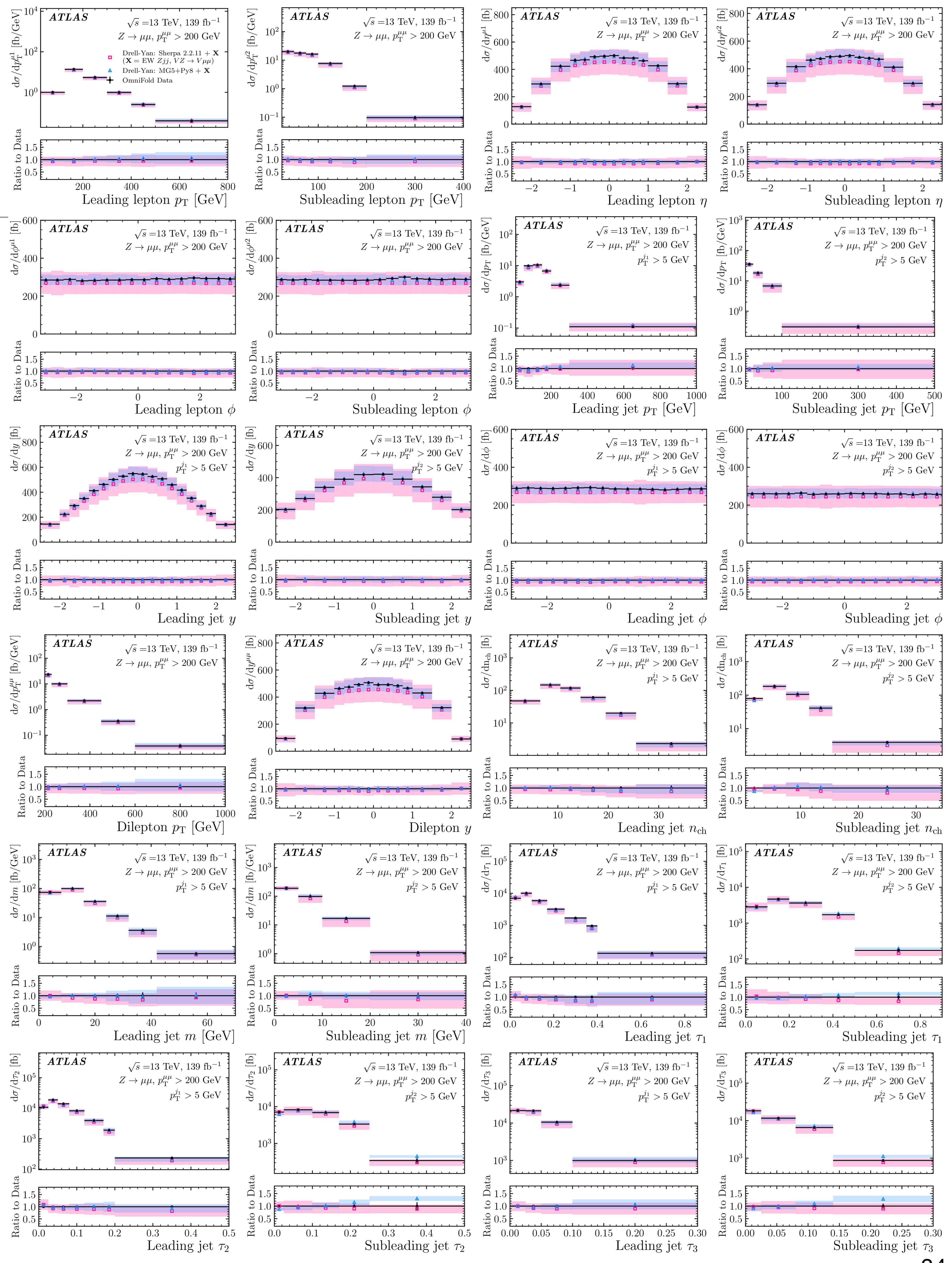


Unfolding Z+jets in 24 dimensions

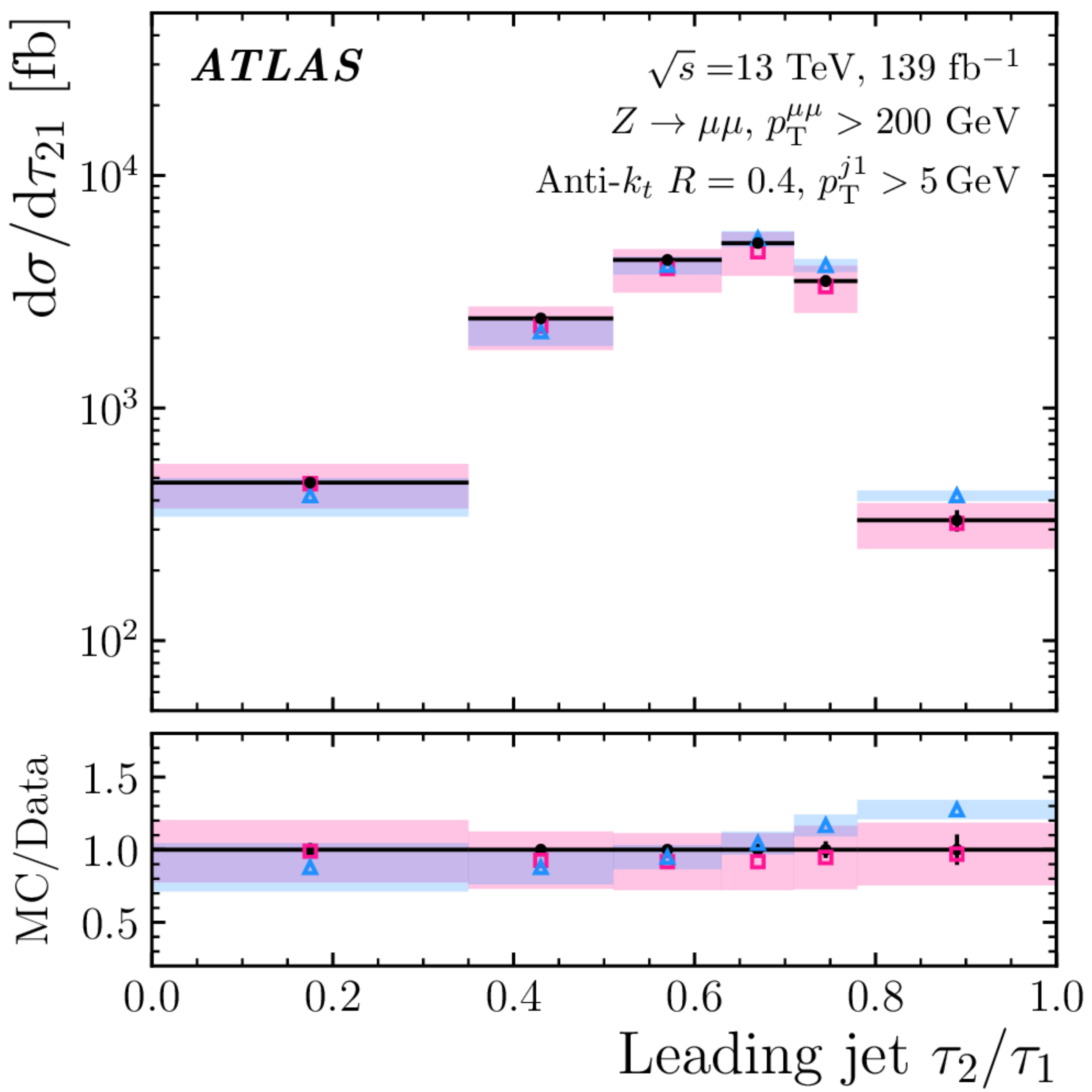
- A fully corrected samples, building other variables and applying other selection criteria is possible
- The unfolding uncertainties are still as a large as via traditional methods

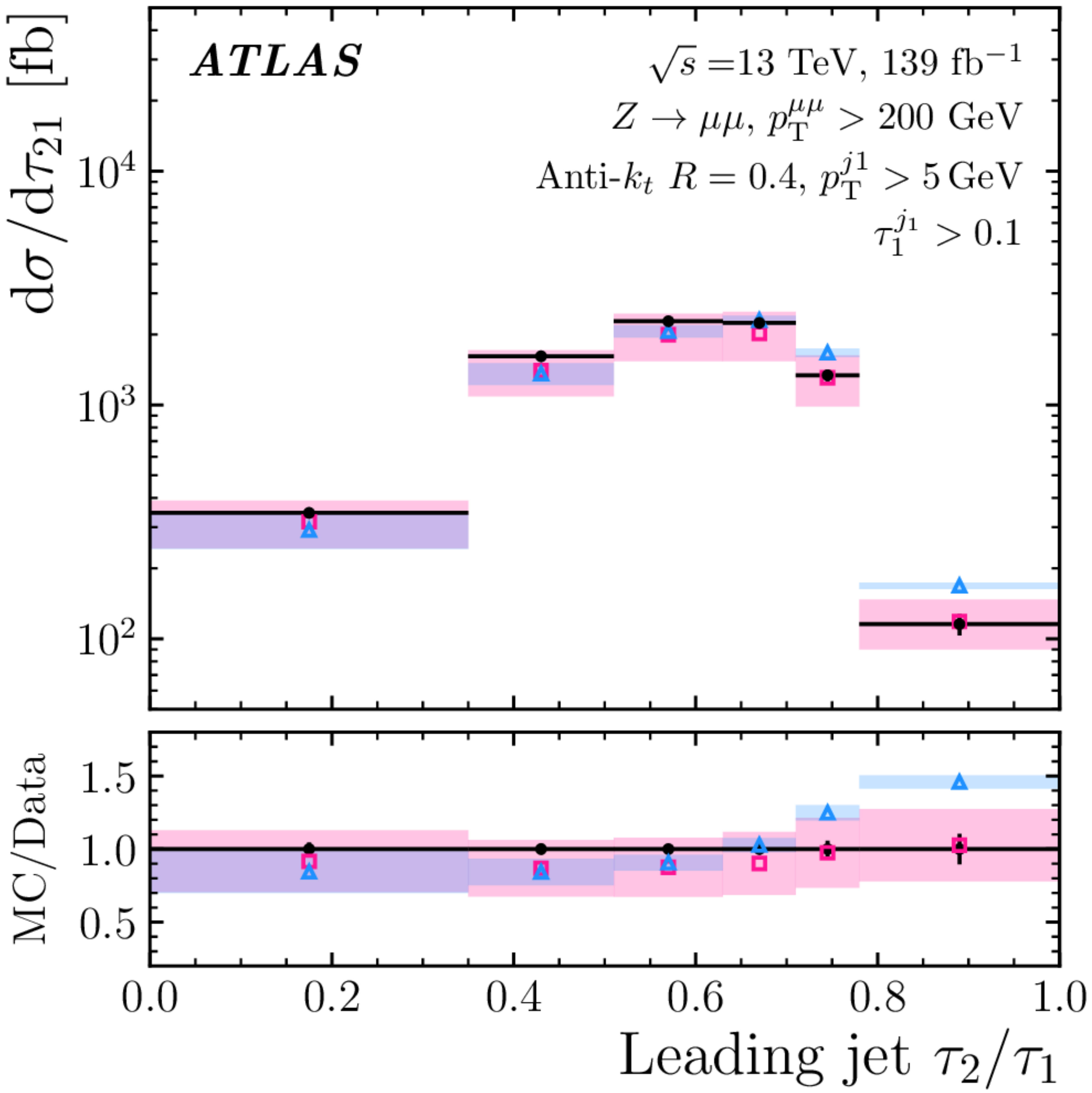






New variables and selections



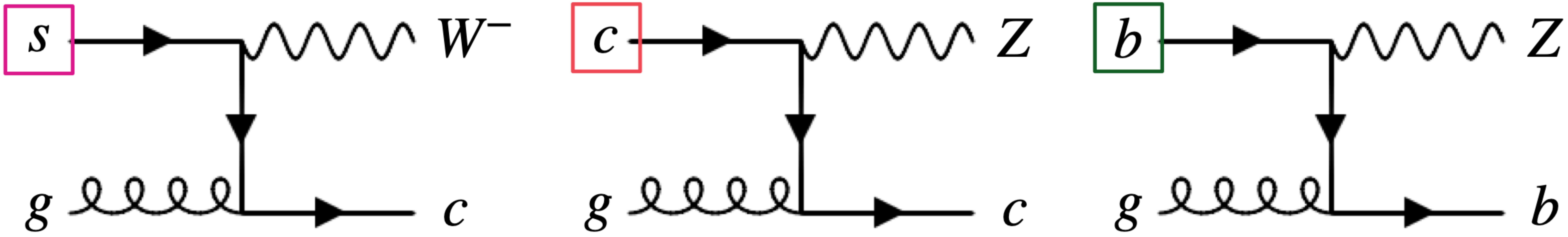


Some questions for you?

- Can't currently do inclusive phase spaces due to computational limitations
  - This analysis used events with Z pT > 200 GeV
- Also exploring unfolding lower level information like all tracks. What use cases could this have?
- How can this kind of data be used for the theory community?

Tapping into heavy flavour

- Results with heavy flavour are excellent data probes for many areas
    - Higher-order QCD, test of 4- vs 5-flavour schemes, PDFs, intrinsic charm
    - Have their own set of experiment challenges



Some heavy flavour challenges

Inclusive tagging

Using jet reconstruction and flavour tagging (FTAG) algorithms

- ✗ Rely on jet reconstruction, FTAG and the related uncertainties
- ✓ Large selection efficiency for *b*-jets and *c*-jets
- ✗ The information on the quark charge is lost
- ✓ Measurements can be used in PDF fits

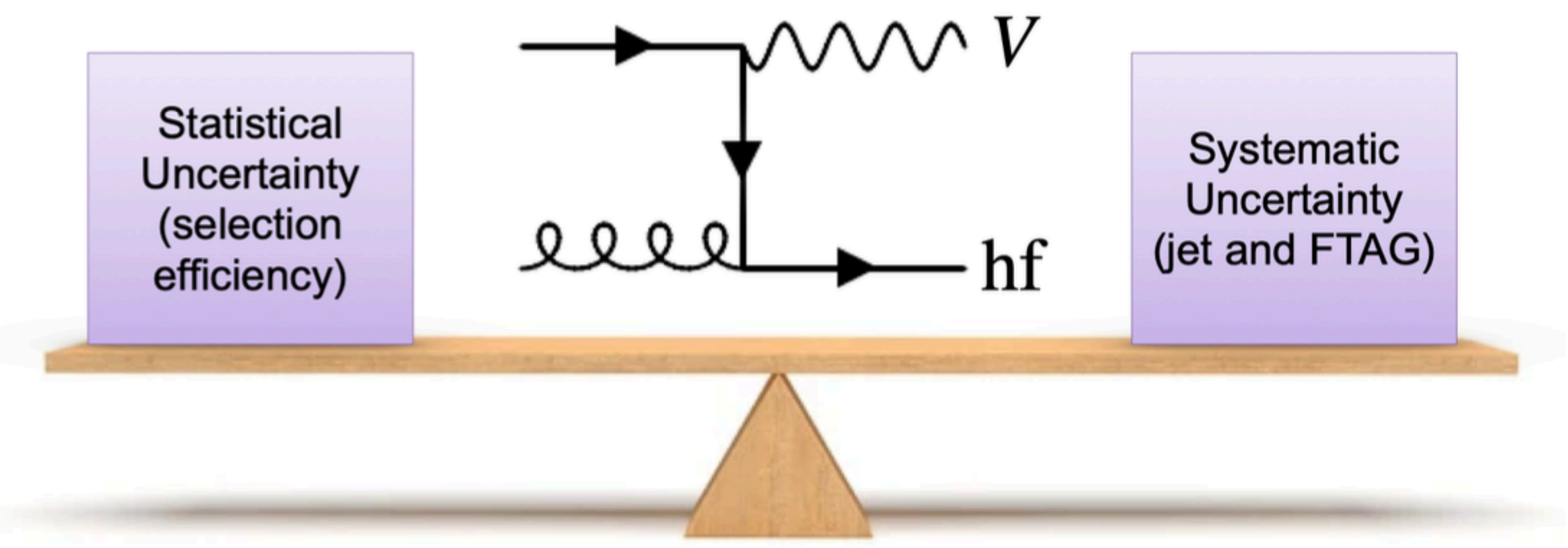
Exclusive tagging

Using specific heavy hadrons decays (e.g. $D \to \mu X$ or $D \to K\pi\pi$)

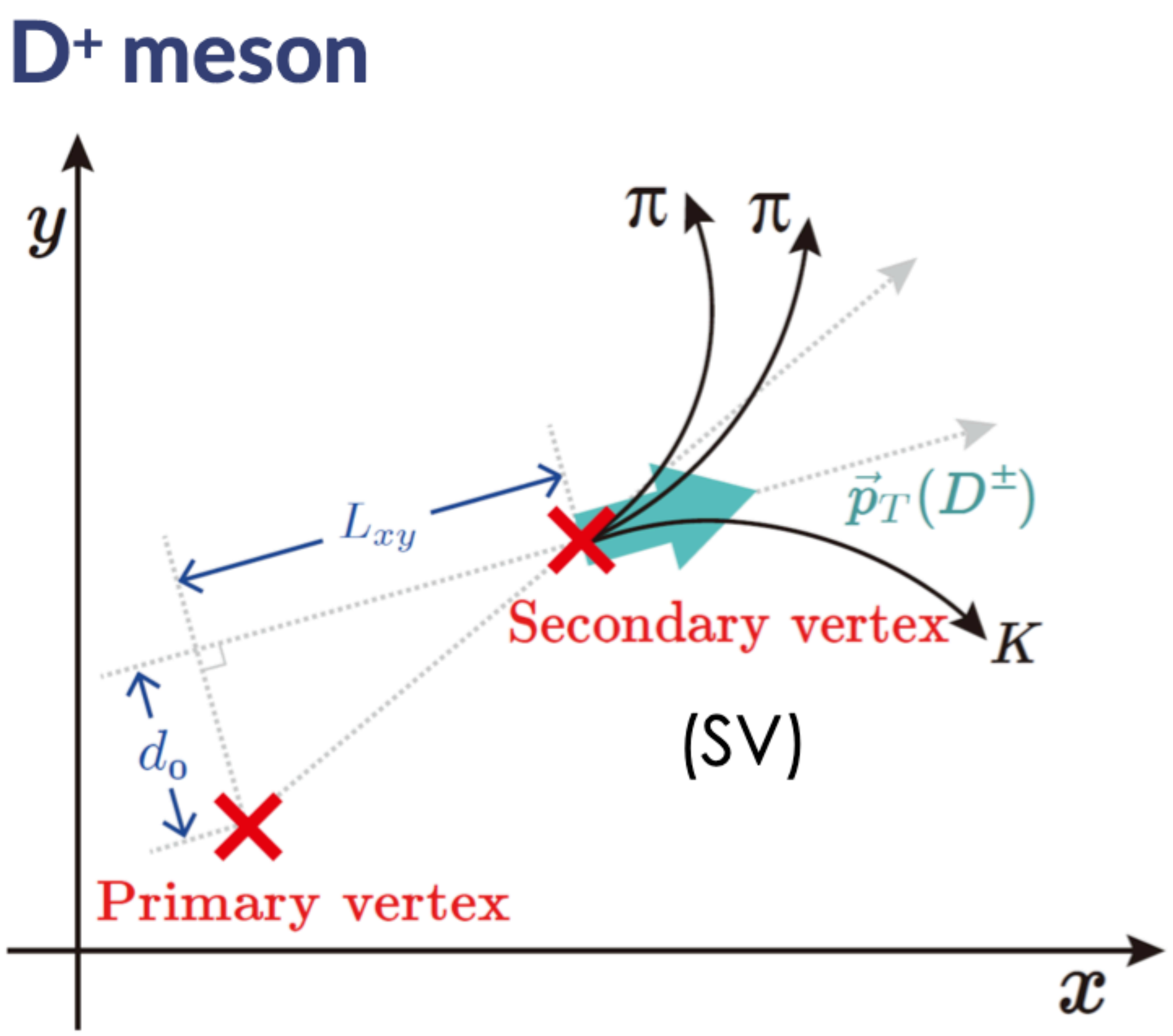
- ✓ No (less) need for jet reconstruction and FTAG
- ✗ Low selection efficiency
- ✓ Very reliable determination of the quark charge
- ✓ Measurements with high precisions, requires to include Fragmentation Functions

Ambiguity in the algorithm used to identify the jet-flavour

The definition of the jet-flavour is not infrared and collinear (IRC) safe - direct comparison with theoretical predictions not possible

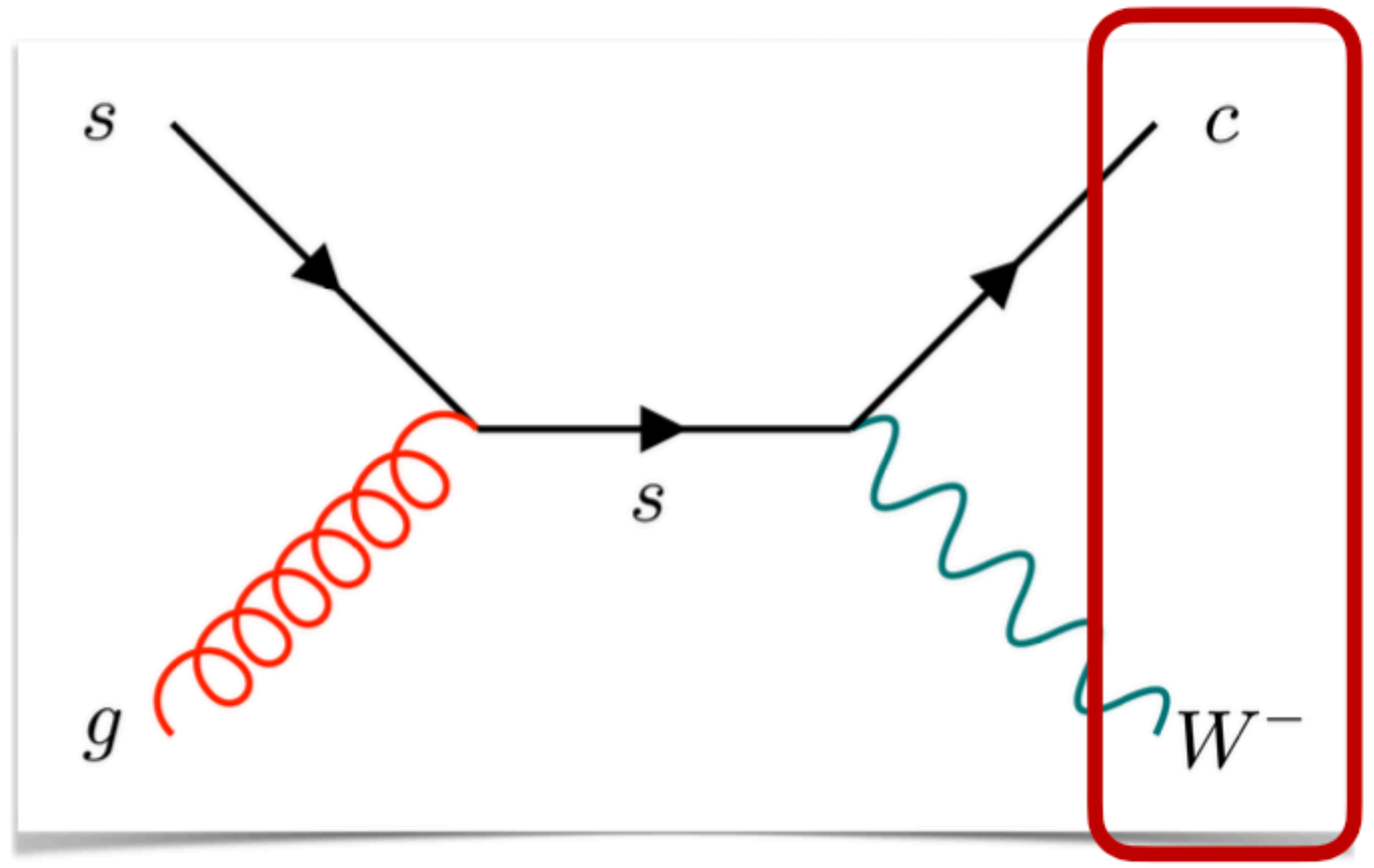


W+c via exclusive tagging



$$D^+ \rightarrow K^-\pi^+\pi^+$$

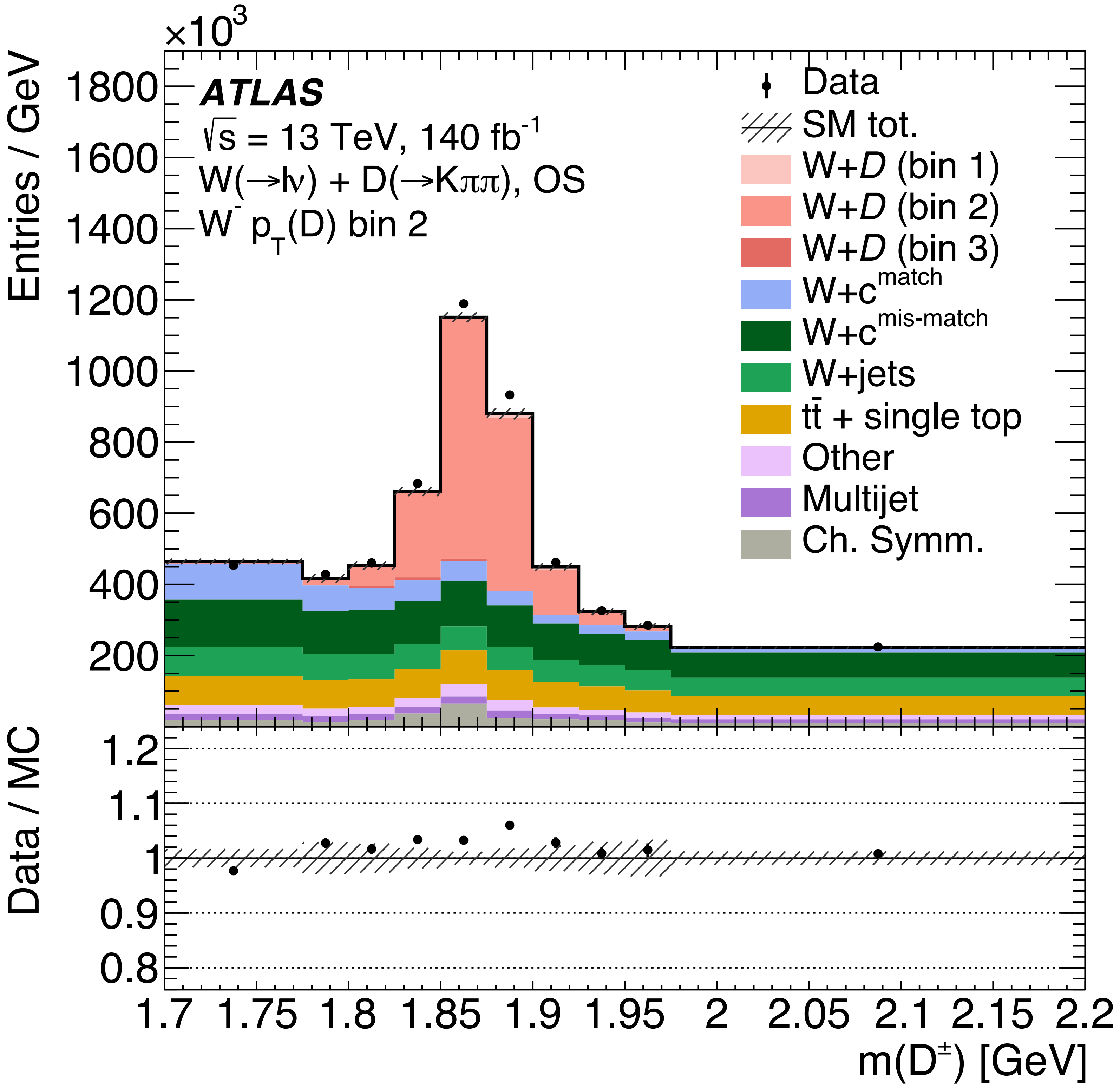
$$D^{*+} \rightarrow D^0\pi^+ \rightarrow (K^-\pi^+)\pi^+$$

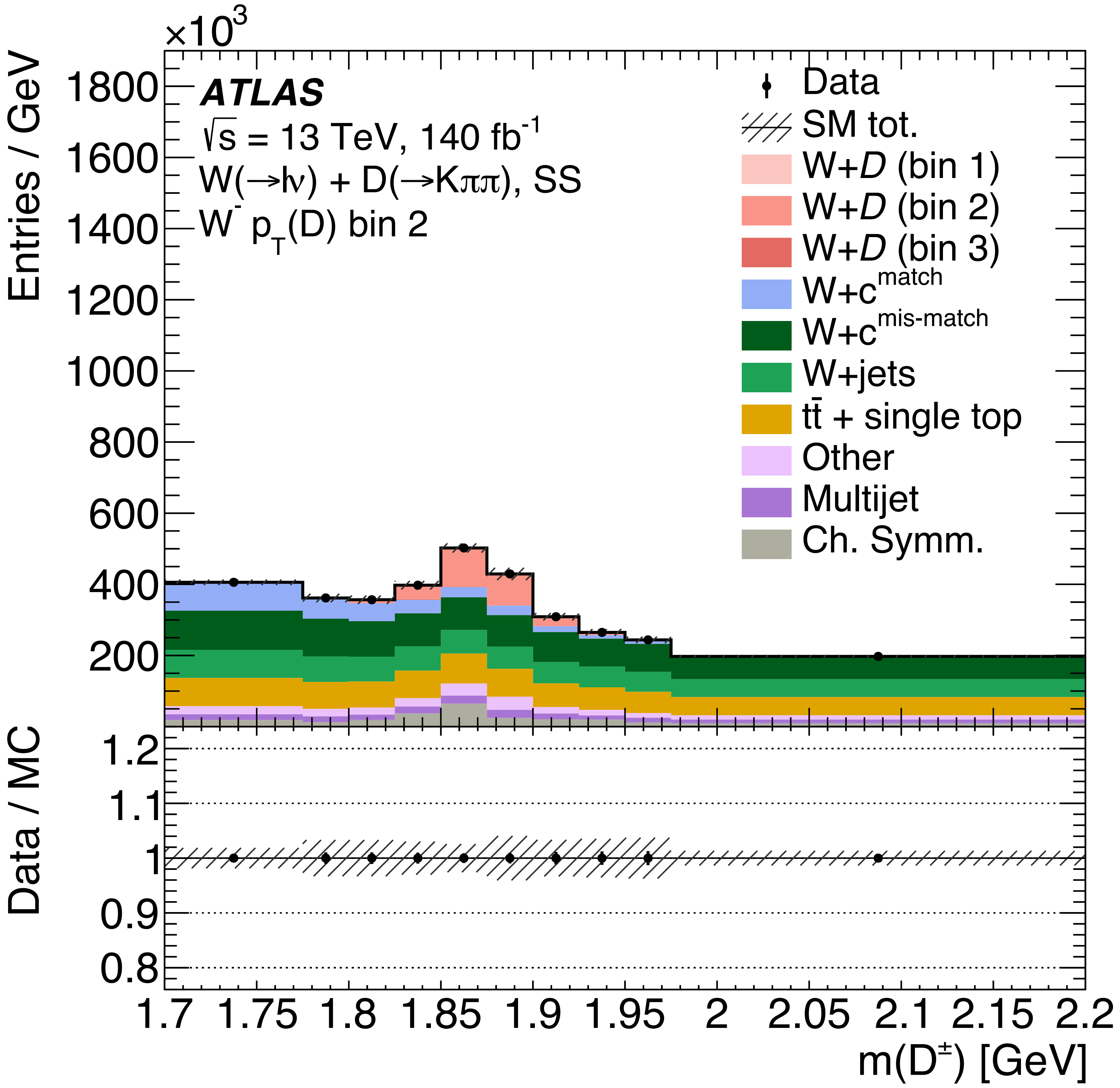


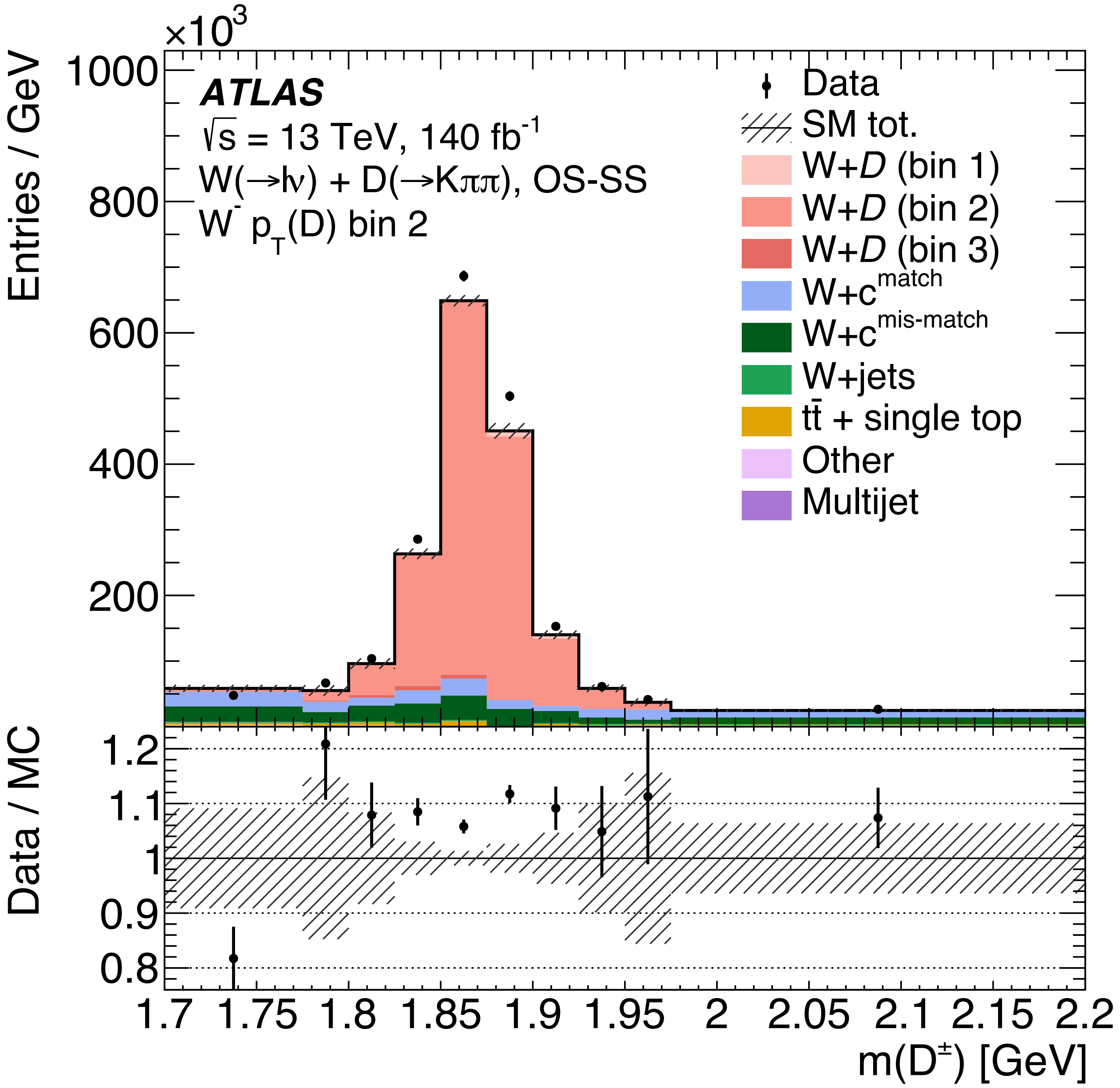
Utilise that the W and the charm are of opposite charge

W+c via exclusive tagging

- Backgrounds rates are symmetric in same-charge/opposite-charge



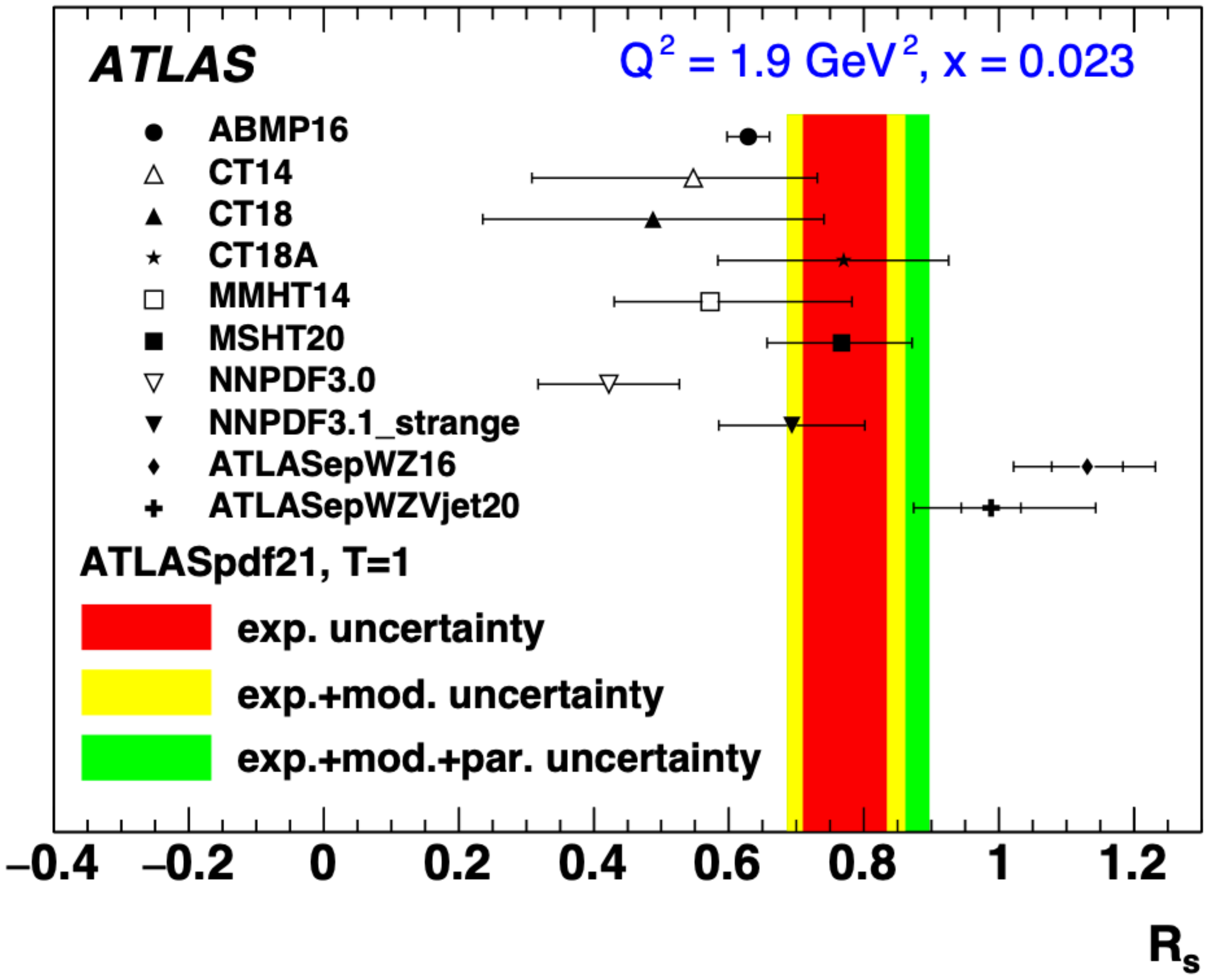




W+c and PDF sensitivity

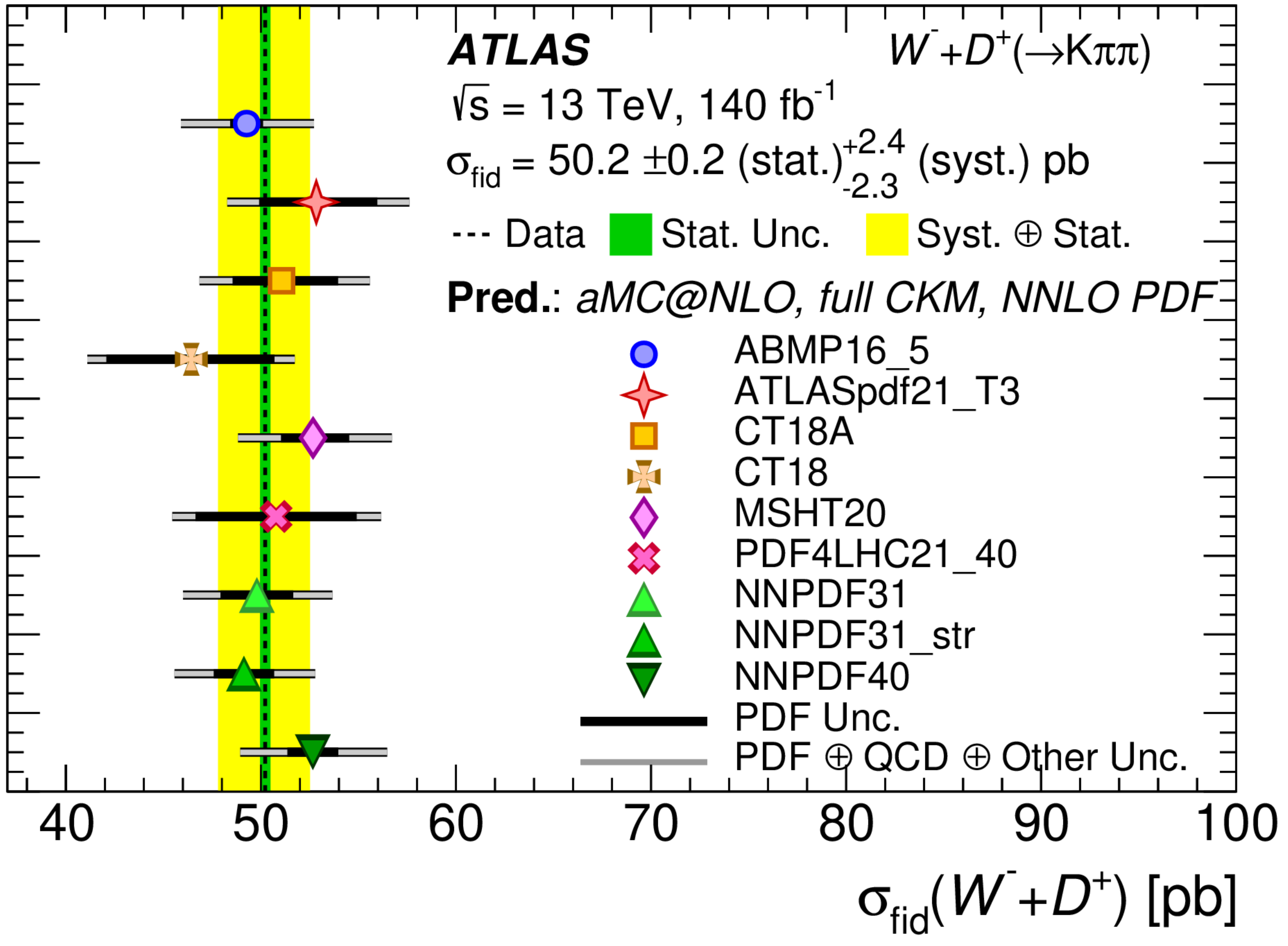
$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}}$$

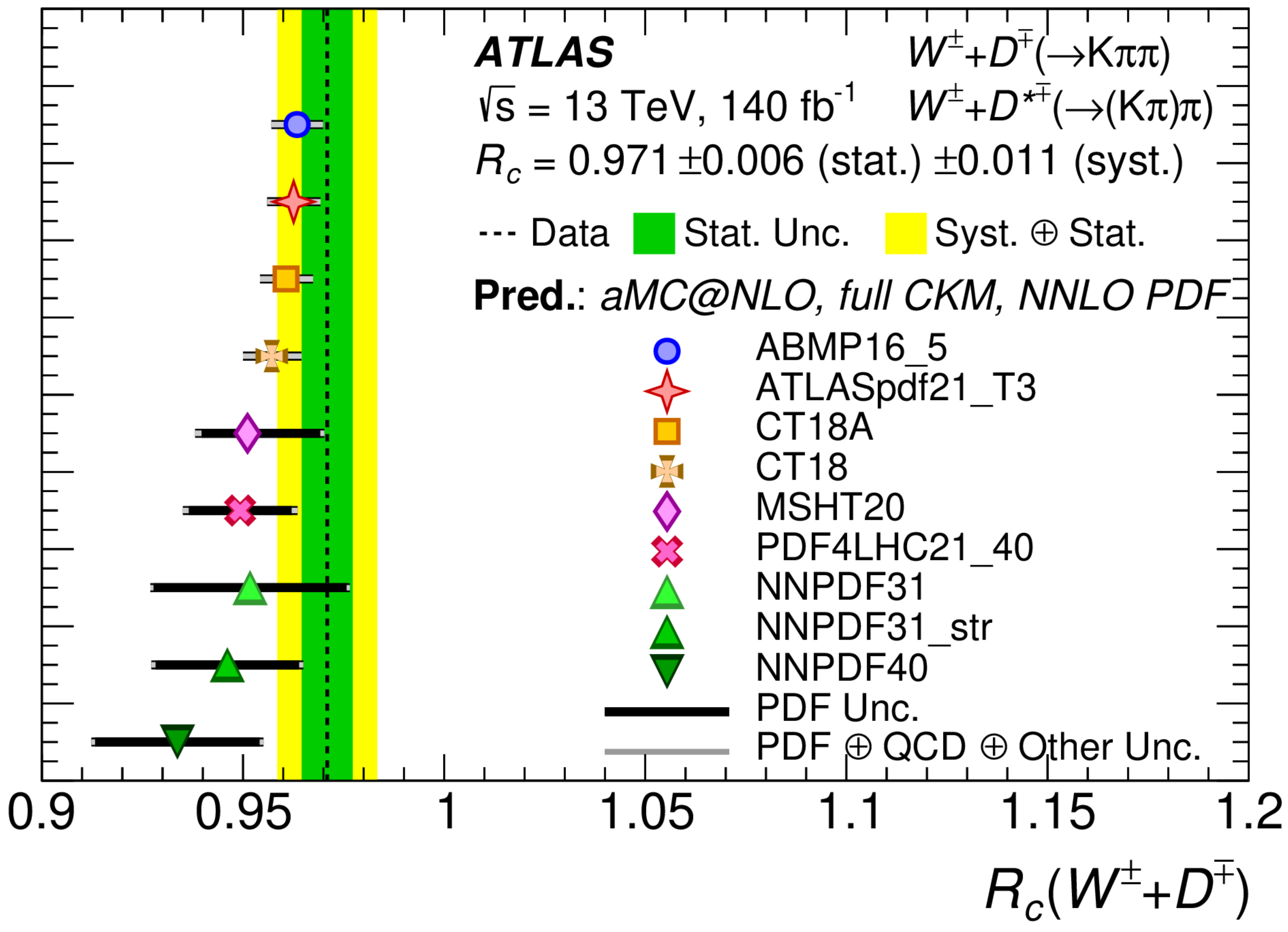
- W + c production sensitive to s-quark PDF in the proton
  - It allows to study the s-quark asymmetry at the initial scale in PDF evolution



W+c results

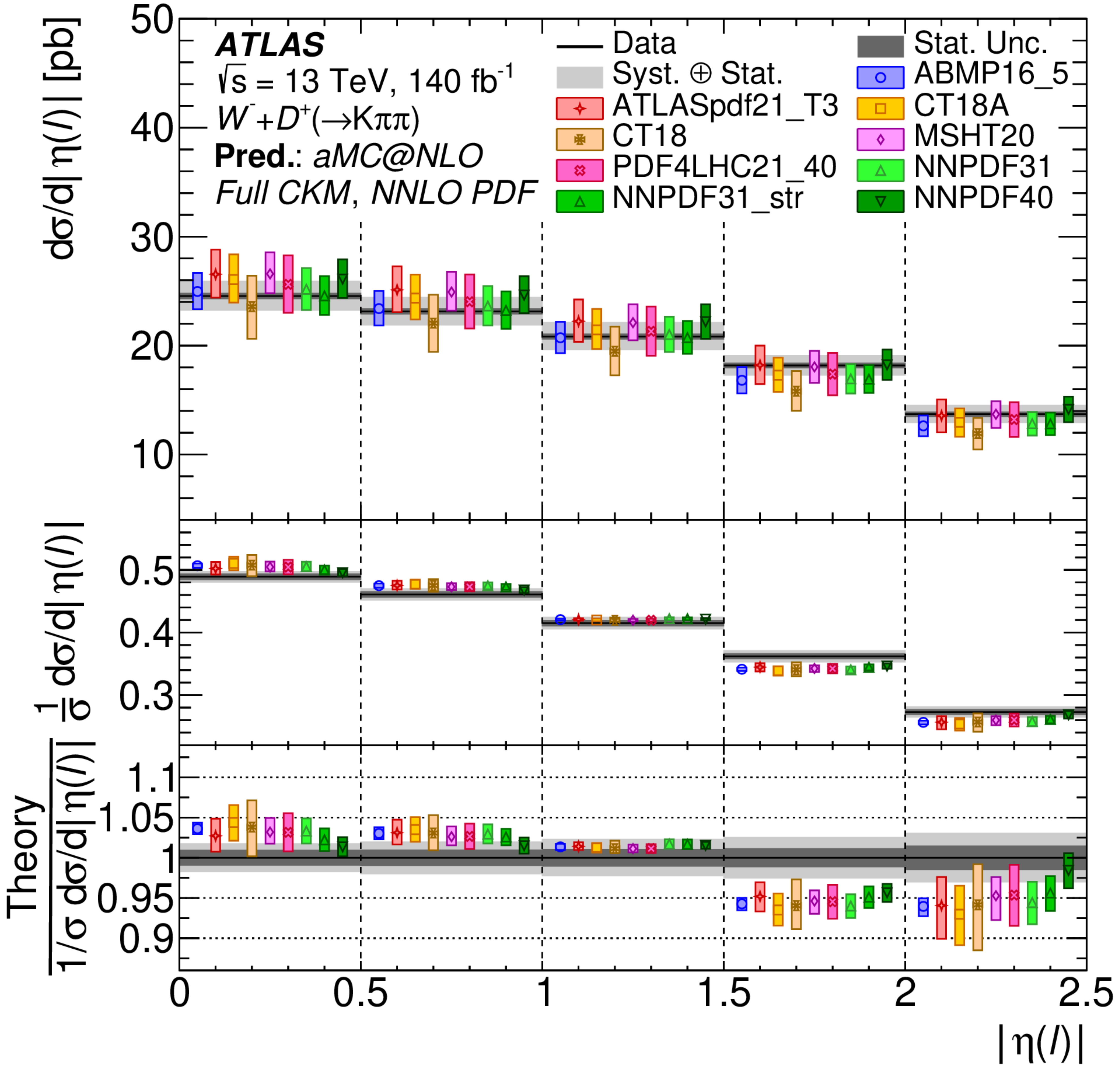
- Rc has an experimental precision of ~1% with compariable precision between statistical and systematic uncertainties





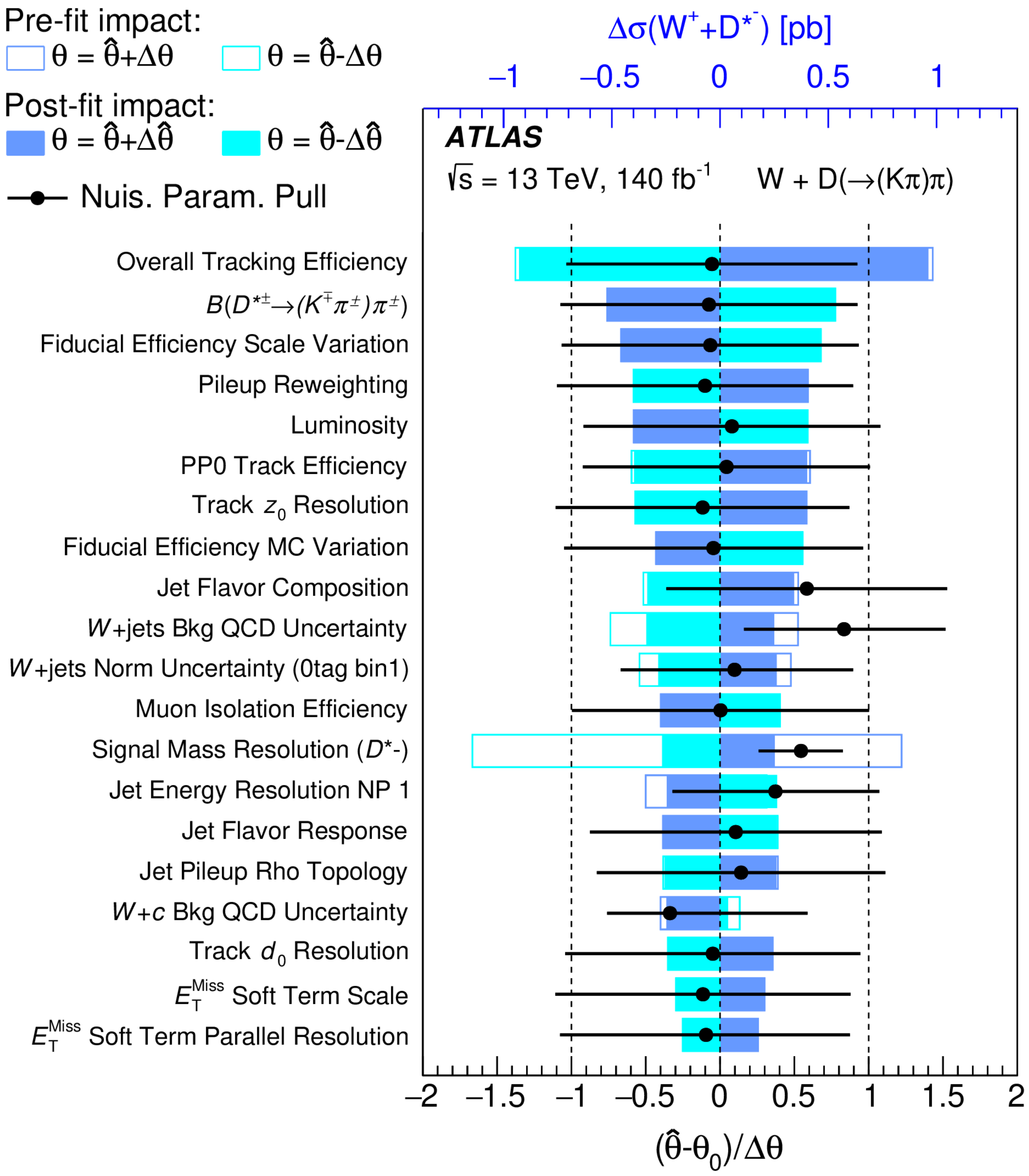
W+c results

- This is a very precise result
- Could be very power tests of PDFs
- Working now on ideas on how to measure the fragmentation functions



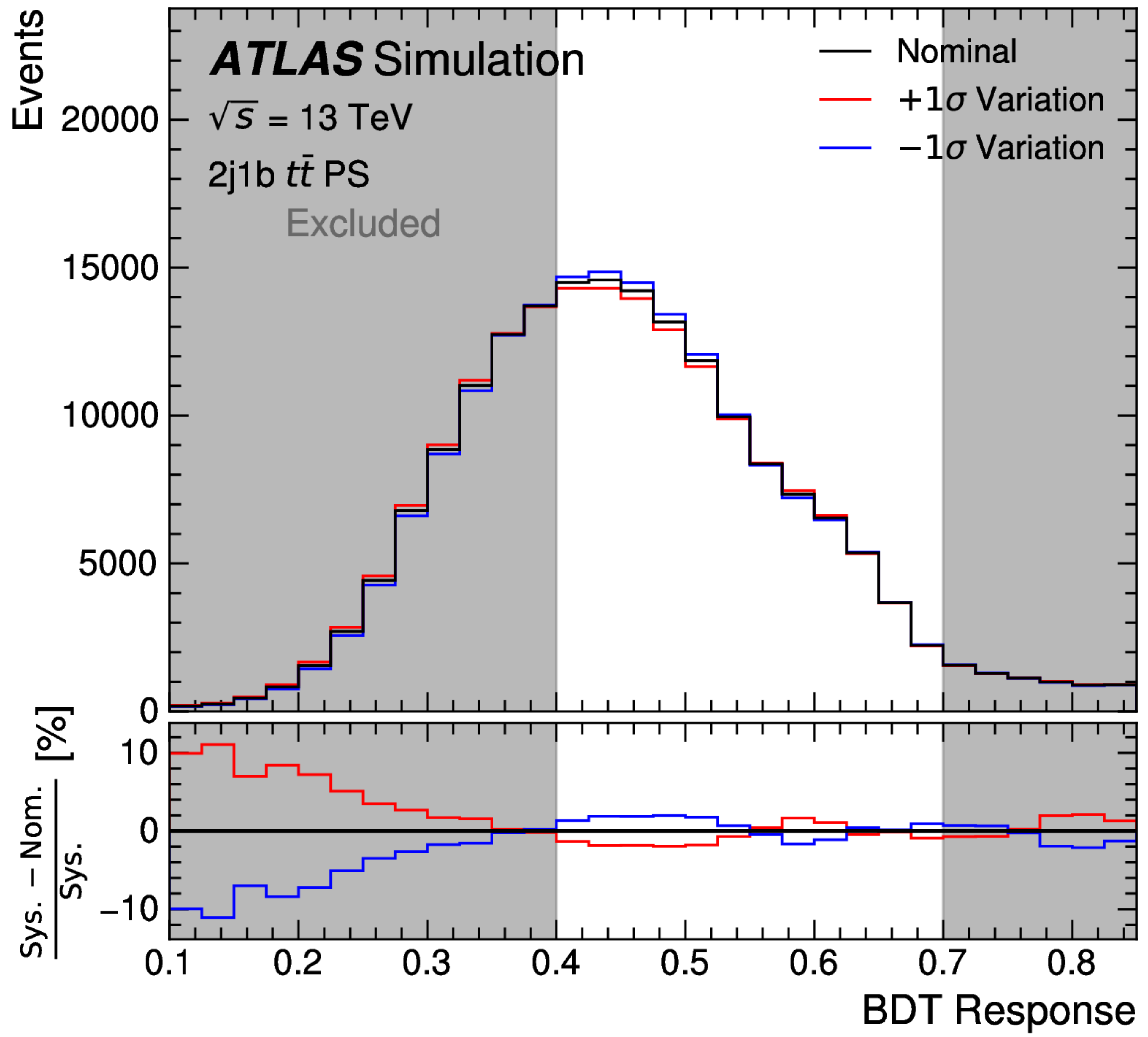
W+c results

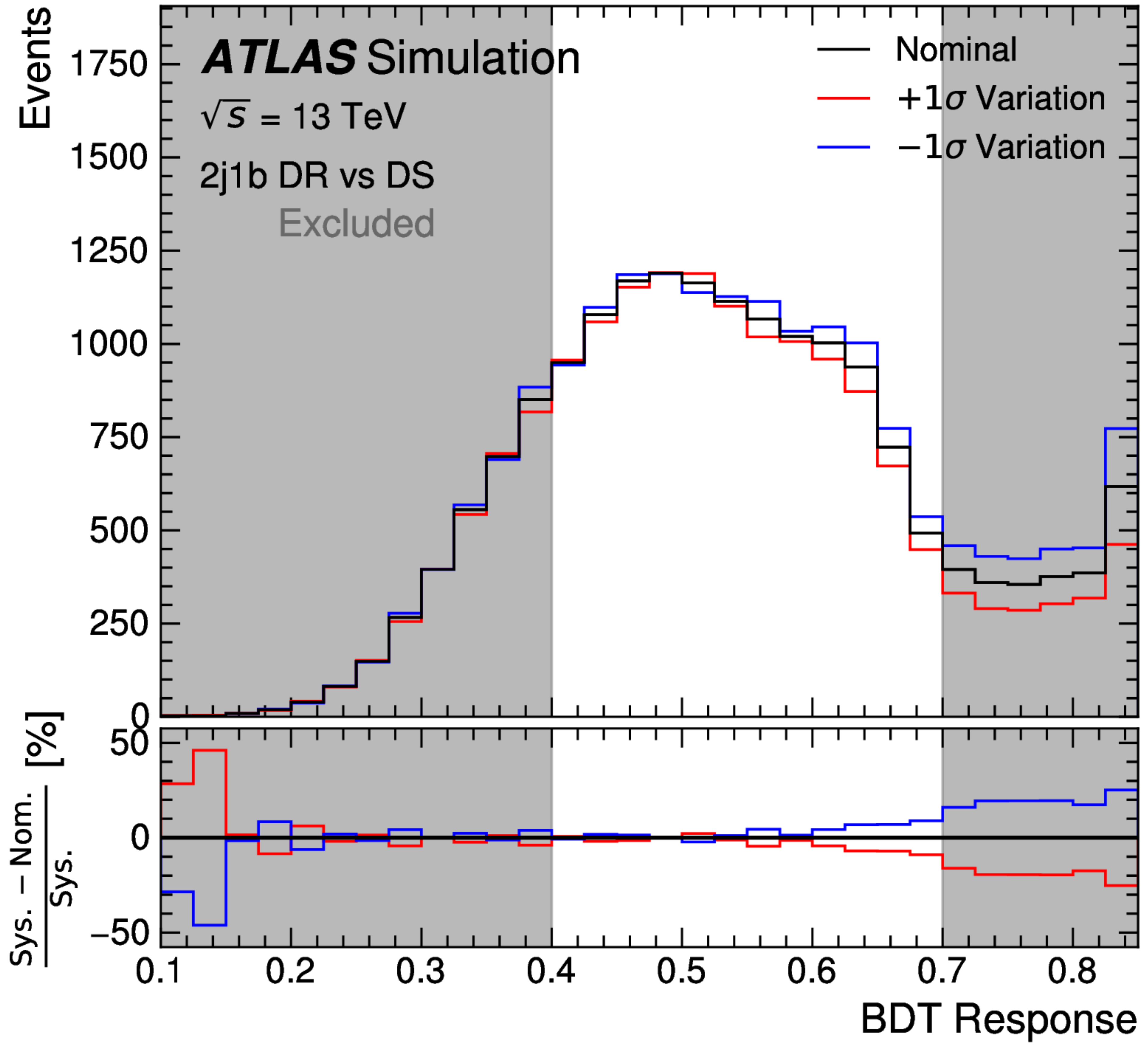
- Few major constraints and pulls



A small aside…

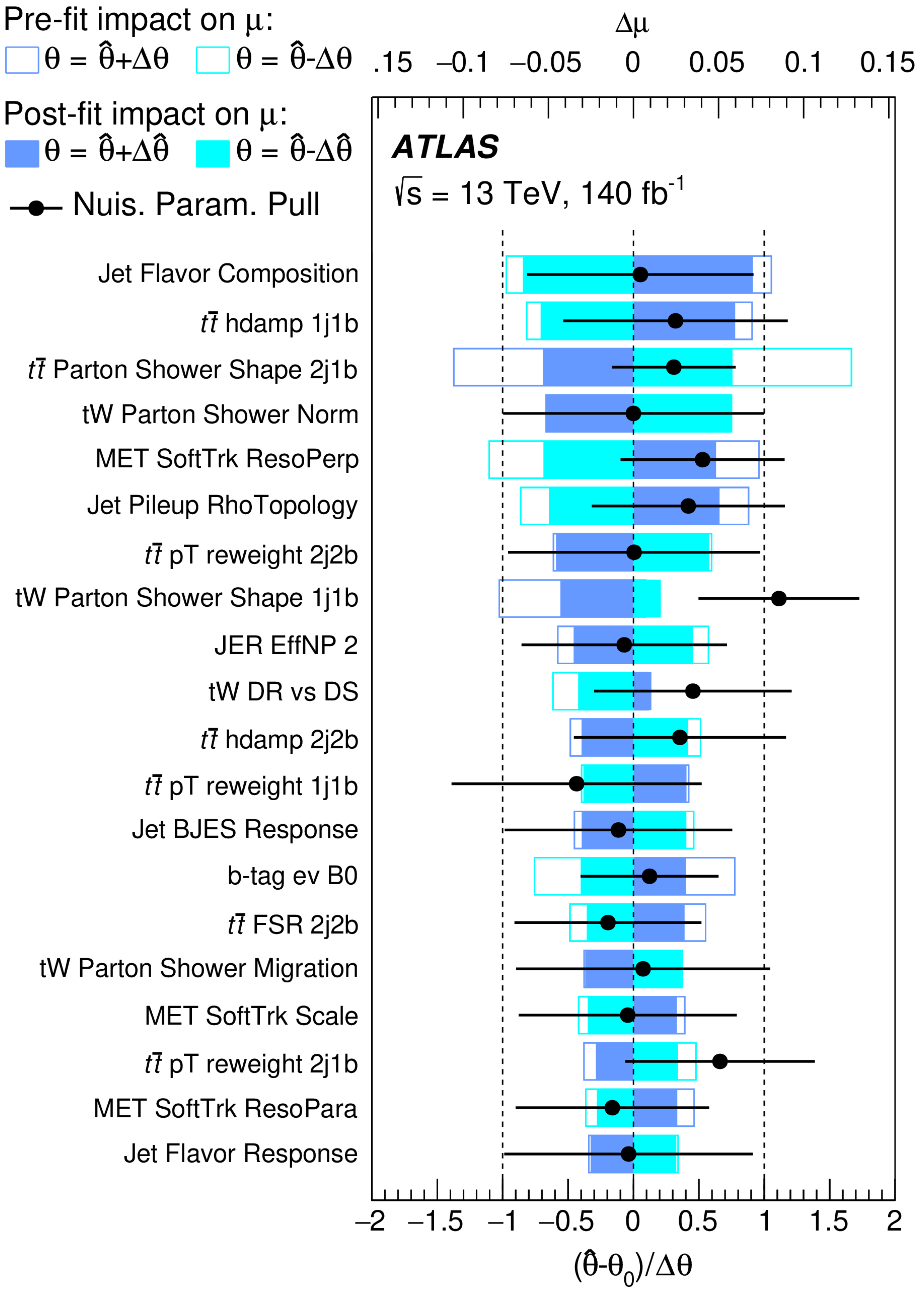
- A recent measurement of tW and fitting modelling uncertainties





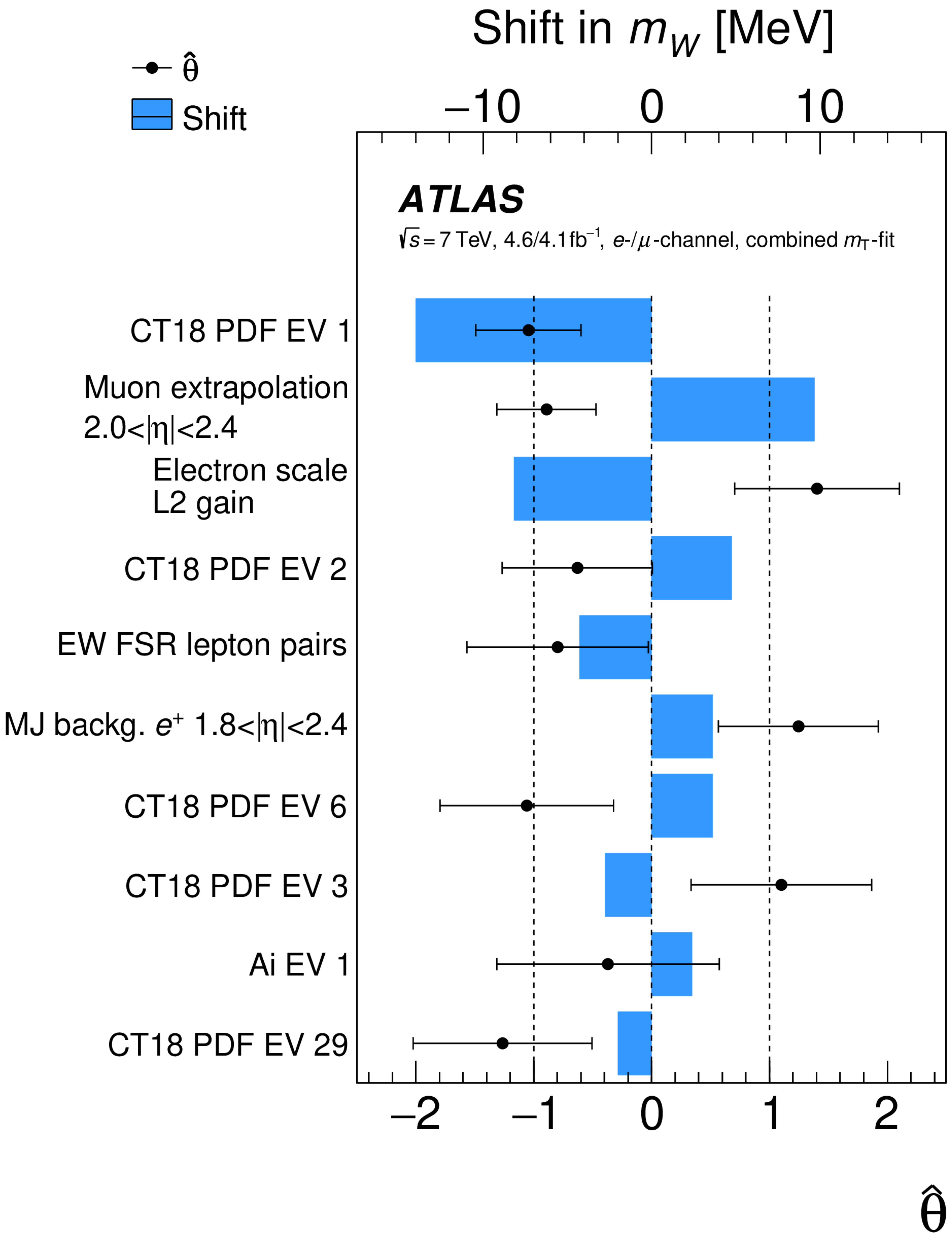
A small aside…

- Our data is very precise
- As a result, we obviously see strong pulls and we don't really trust the interpolation model



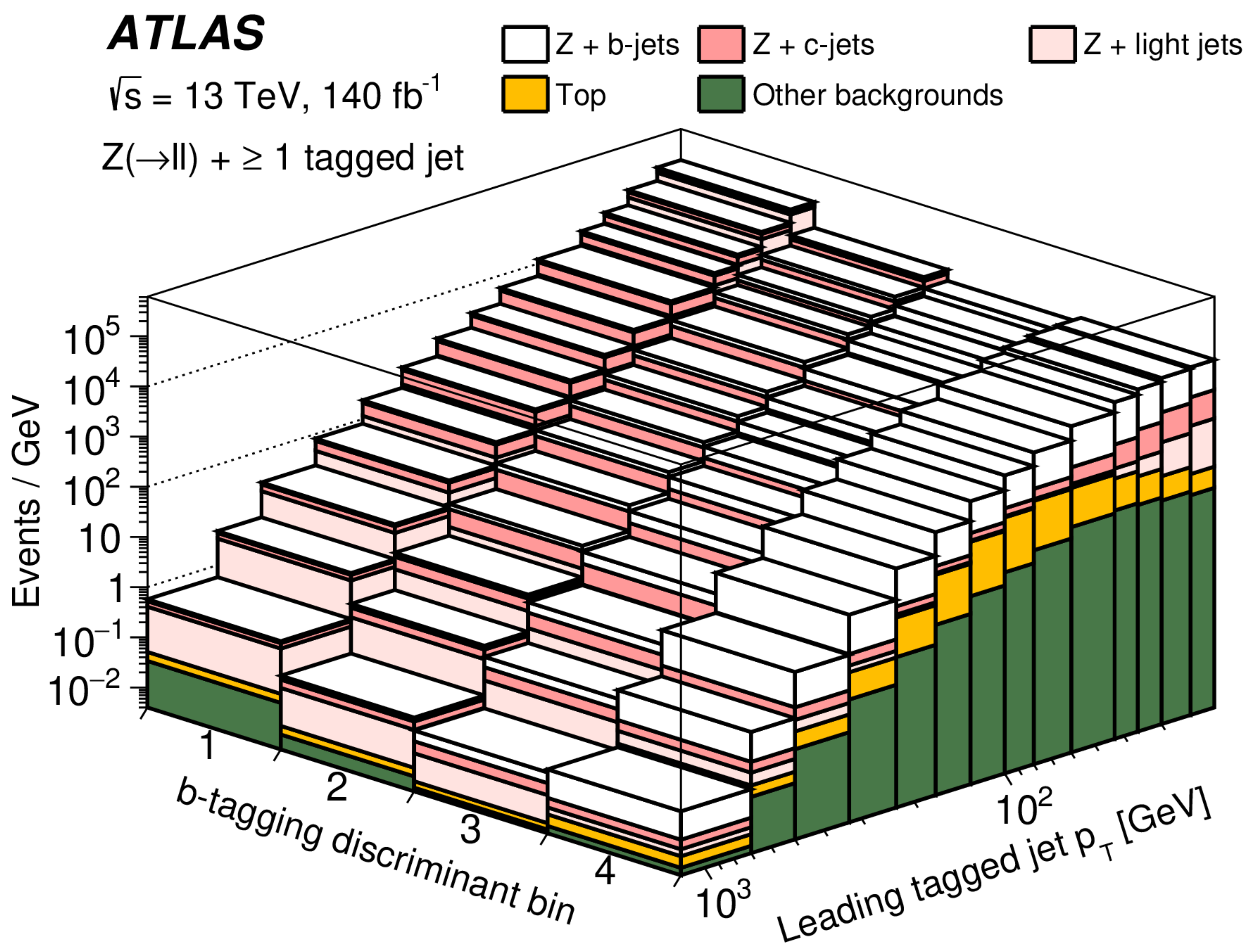
A small aside…

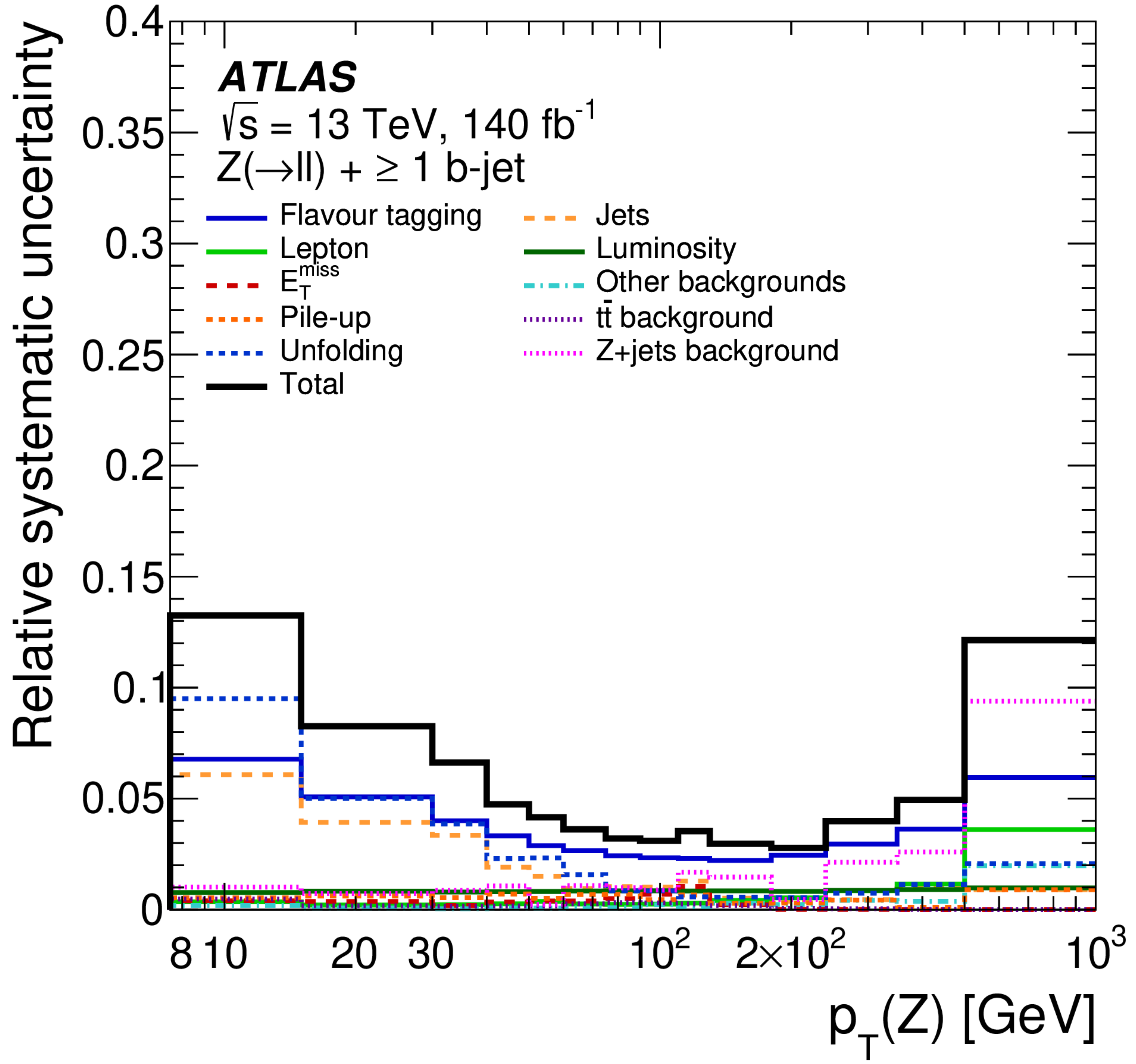
- And we really get nervous when the experimental uncertainties pull
- An example from our mW result



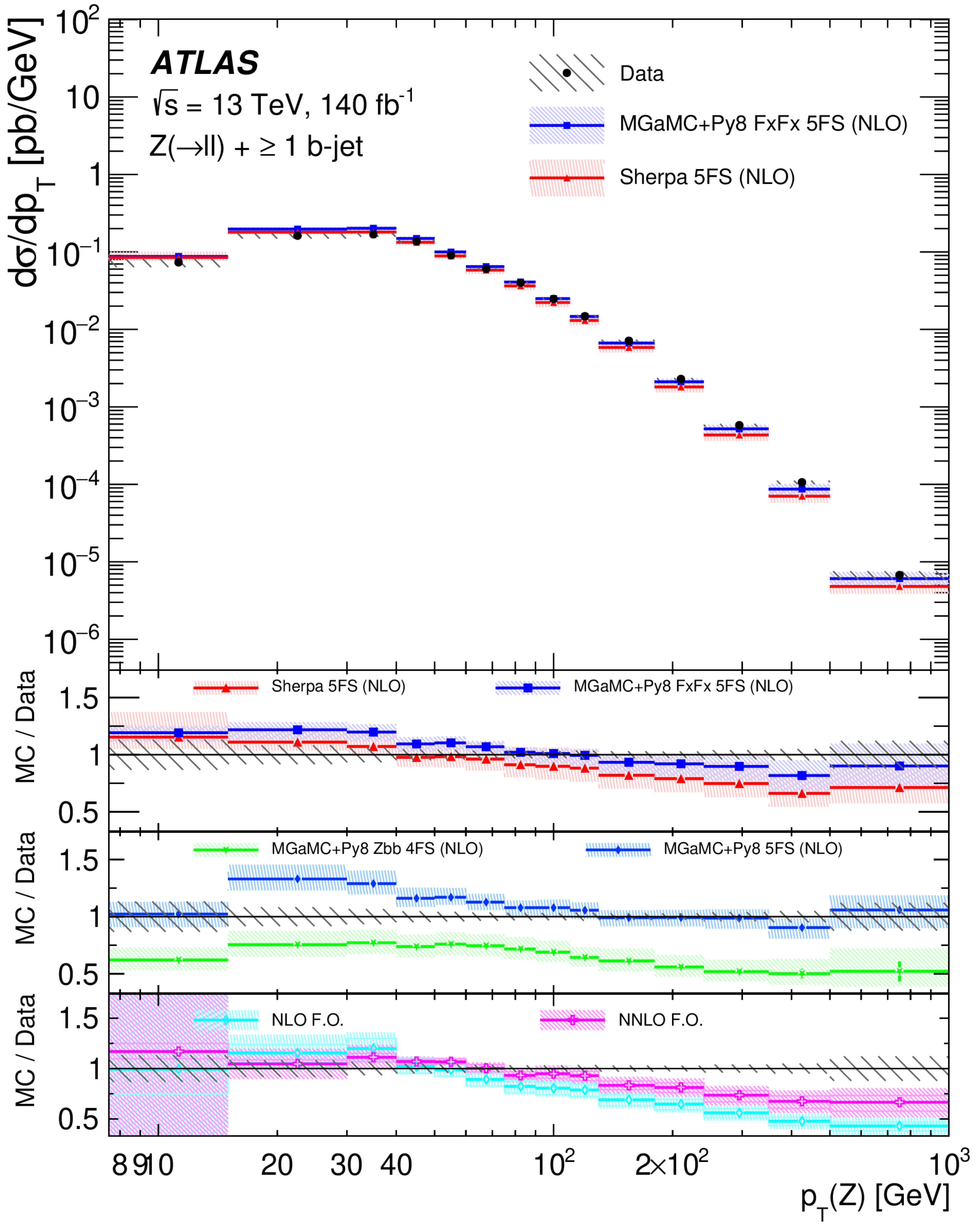
Z plus heavy flavor production

- Very comprehensive result on Z+1b, Z+2b and Z+c production
- Fit is performed in individual, optimised bins for each measured observable

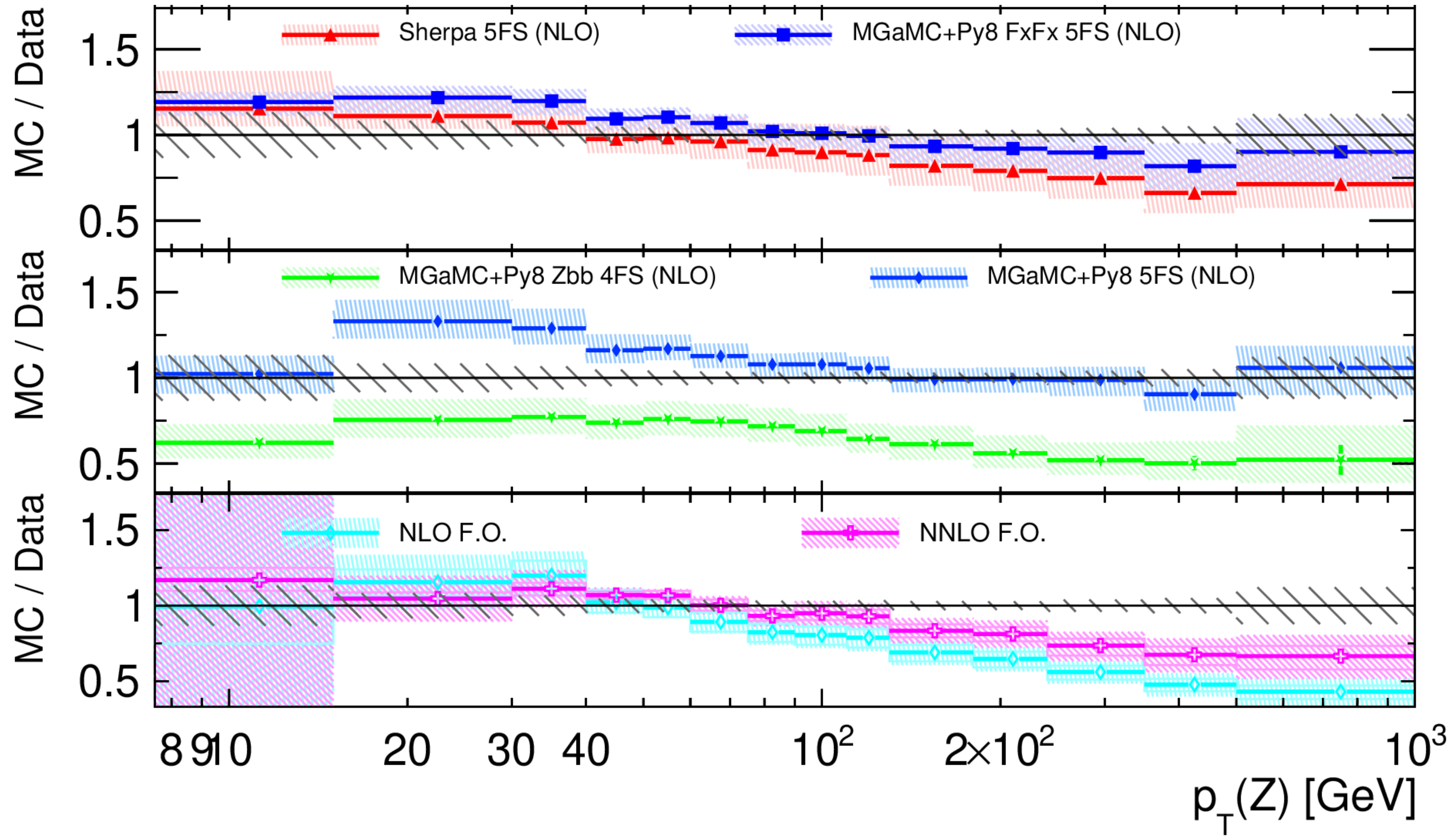




Some Z+heavy flavour results

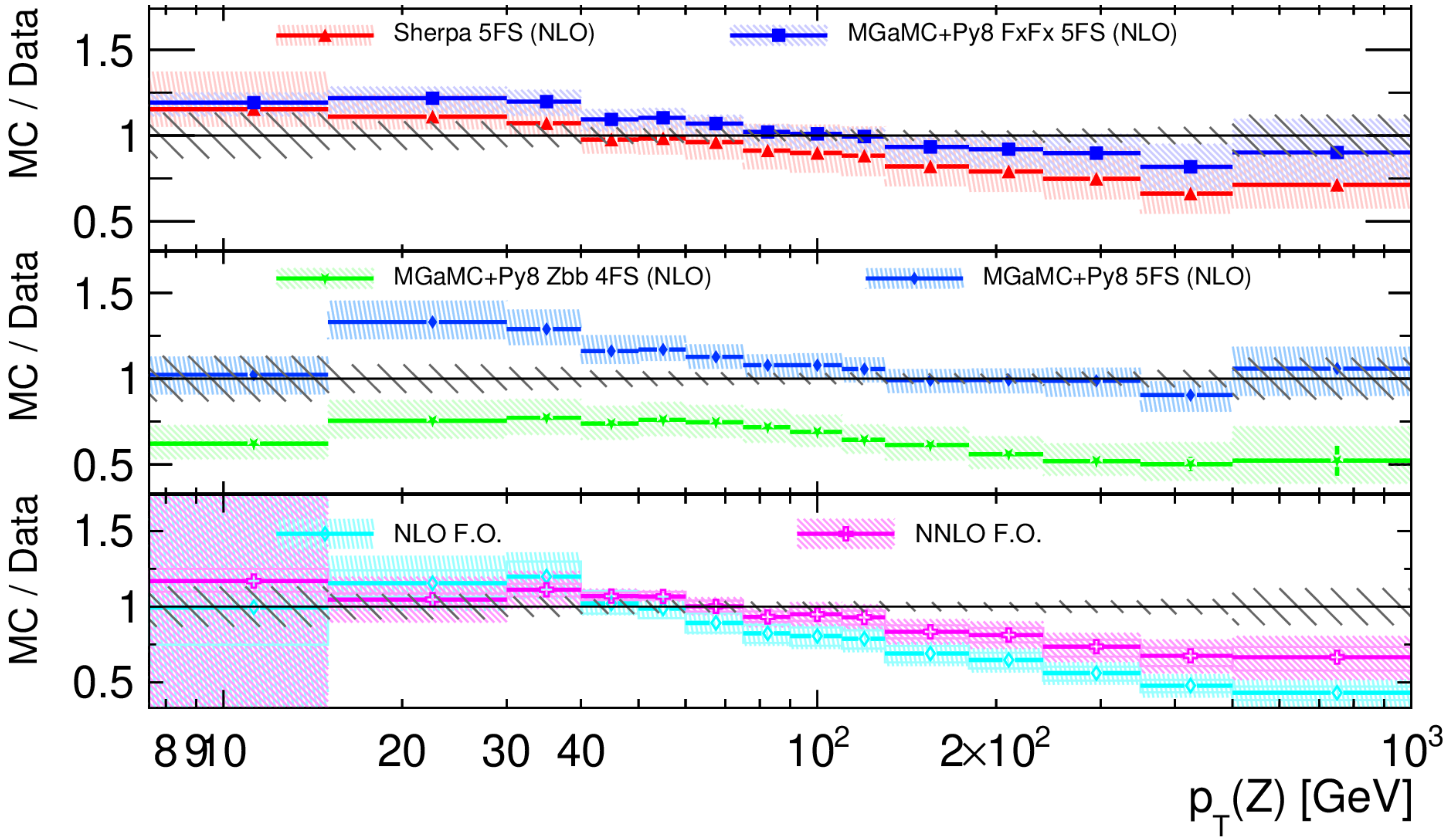


- To make these plots more manageable, will just focus on the ratios



Some Z+heavy flavour results

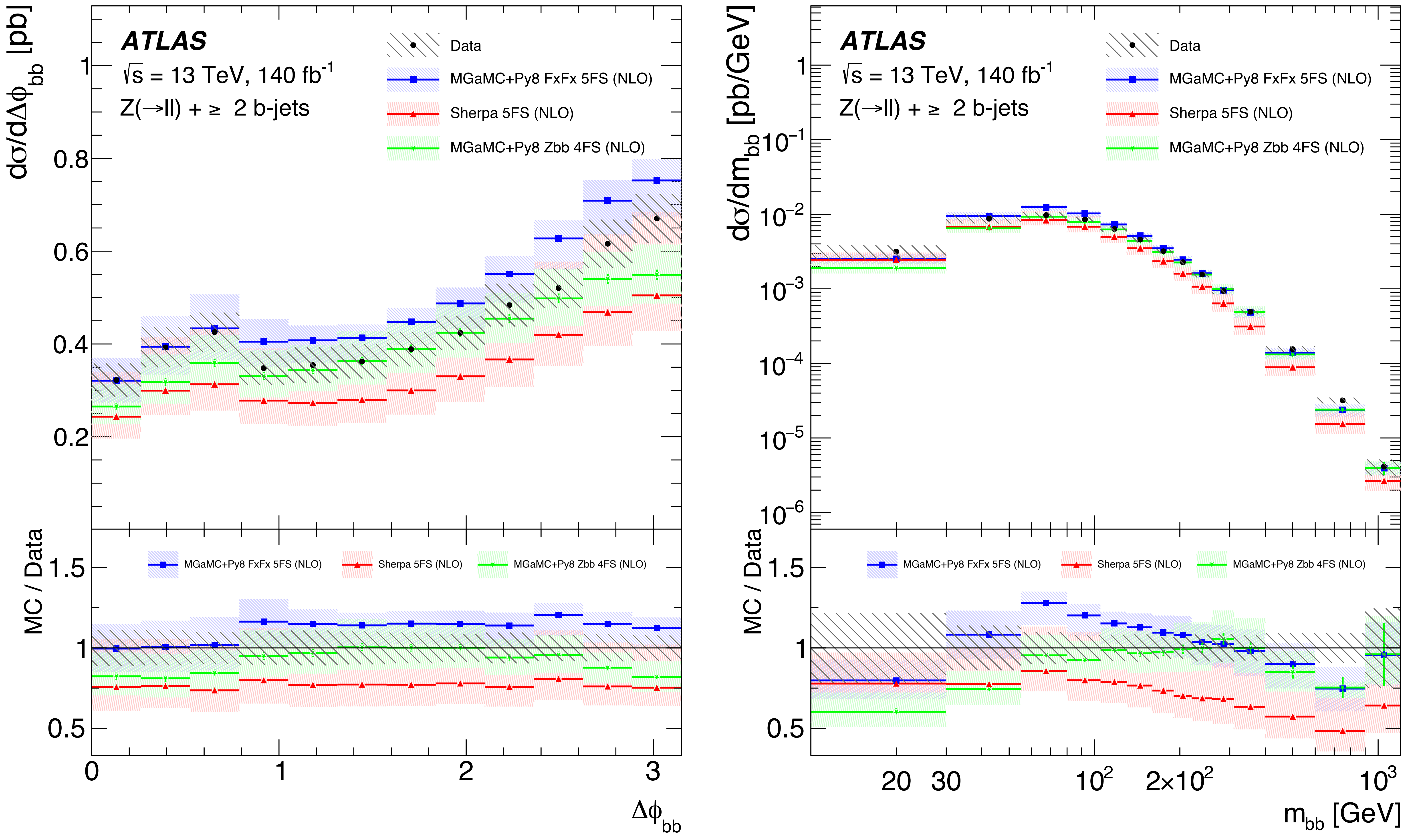
- Recall… 4FS: massive b-quarks
  - b-quarks do not contribute to proton wave functions and do not enter in pQCD calculations and PDF evolution
  - b-quarks can only be generated in the hard scattering by gluon splitting (g → bb)
  - suitable for kinematic region with energy scale Q ~ m_b^2
- 5FS: massless b-quarks → b-quark density is allowed in the initial state via a b-quark PDF
  - suitable for kinematic region with Q >> m_b^2
- The ambiguity between the FSs is expected to reduce including higher oder pQCD corrections



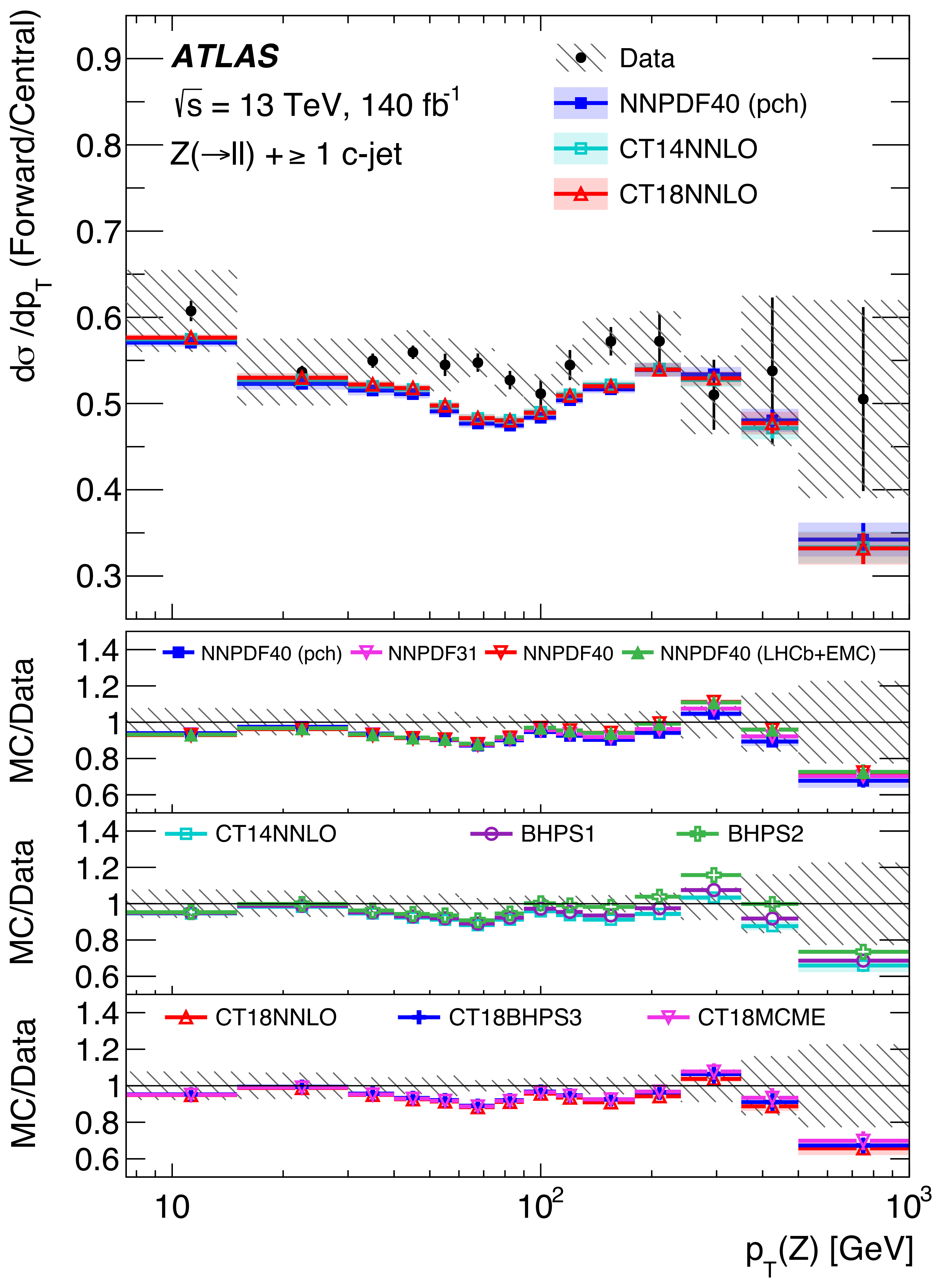
Large uncertainty on the NNLO due to different jet flavour algorithms

Some Z+heavy flavour results

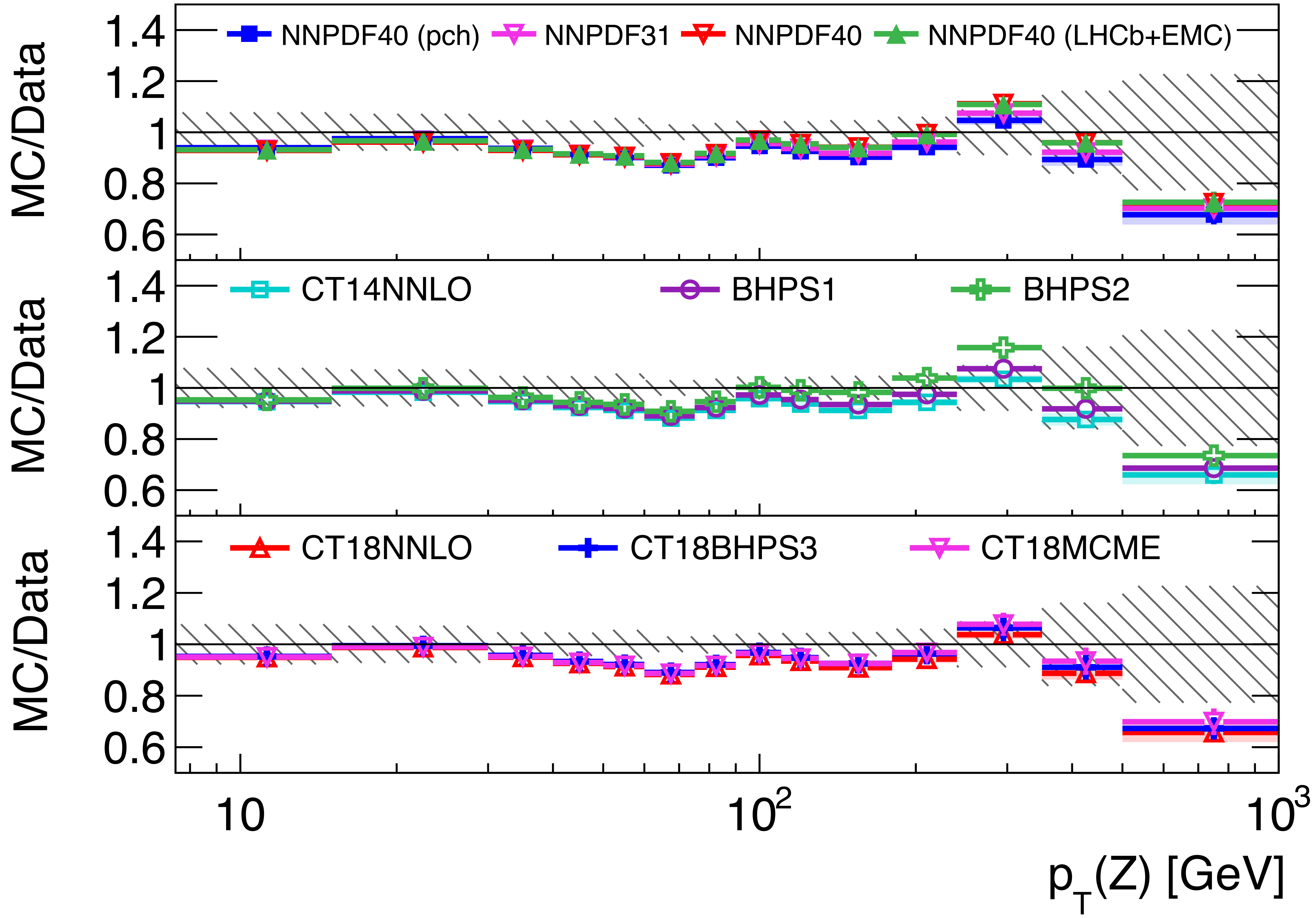
- Distributions like mbb are always a challenge



Some Z+heavy flavour results



- Similar trends for IC models - the measurements has a small sensitivity



My guesses for the future

- We will be breaking the jet energy uncertainty paradigm soon
- We have tons of data that is very constraining
- Many examples of how this has allowed us to make big steps in improvement to ‘traditional systematics dominanted’ results
- Machine learning allows us to rethink everything from calibration to cross sections
- Fitting everything has great power however the classic tails of caution still apply

Some references

- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2011-08/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2016-14/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2018-55/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2020-04
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2018-17/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2020-17/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2018-43/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-002/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-022/
- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-040/