Challenges in the EFT interpretation of unfolded cross section measurements

Andrew Gilbert

26 April 2024 | LPC EFT Workshop

Introduction

- **Disclaimer**: this talk does not include any new or groundbreaking ideas, but maybe a few thoughts for discussion
- We will reinterpret our measurements for a long time
  - EWPO observables measured at LEP ⇒ part of our combined EFT fits decades later
- The nature of the reinterpretation may change over time
  - Even if still EFT, we may consider (more) (higher dimensional) operators, different bases, flavour assumptions, higher order QCD, EW, ..
- Unfolded measurements have already been used extensively in EFT fits (especially outside the expt. collaborations)
  - Non-exhaustive examples: Fitmaker [2012.02779], SMEFiT - [2105.00006], EFTfitter [1605.05585]
- **Challenges:**
  - Usually only unfold 1-3 observables simultaneously ⇒ lack complete information about the process in question
  - Backgrounds often assumed to be SM and subtracted
  - For EFT often interested in high energy tails ⇒ low stats. ⇒ difficult to unfold / Gaussian regime not valid
  - Reinterpretation assumes $\epsilon \times A$ in bin same for SM and EFT
- This talk: a few analysis examples facing these issues

Current EFT approaches

CMS

Fiducial/differential measurements with EFT interpretation

Unfolding, with **likelihood fit** … or **matrices**

$L(\text{data} \mid \sigma_i)$

$L(\text{data} \mid \sigma_i(c_j))$

$$(\vec{y} - \mathbf{K}\vec{\sigma})^{\mathrm{T}}\mathbf{V}^{-1}(\vec{y} - \mathbf{K}\vec{\sigma}) + \delta P(\vec{\sigma})$$

Can recast from $\boldsymbol{\sigma_i}$ to give $\mathbf{c_j}$, or other parametrization of $\boldsymbol{\sigma_i}$

Direct EFT constraints (w/ optimised analysis) aka "full-sim"

One direct fit:

$L(\text{data} \mid c_j)$

Can (in principle) recast from c_j to other congruent EFT basis

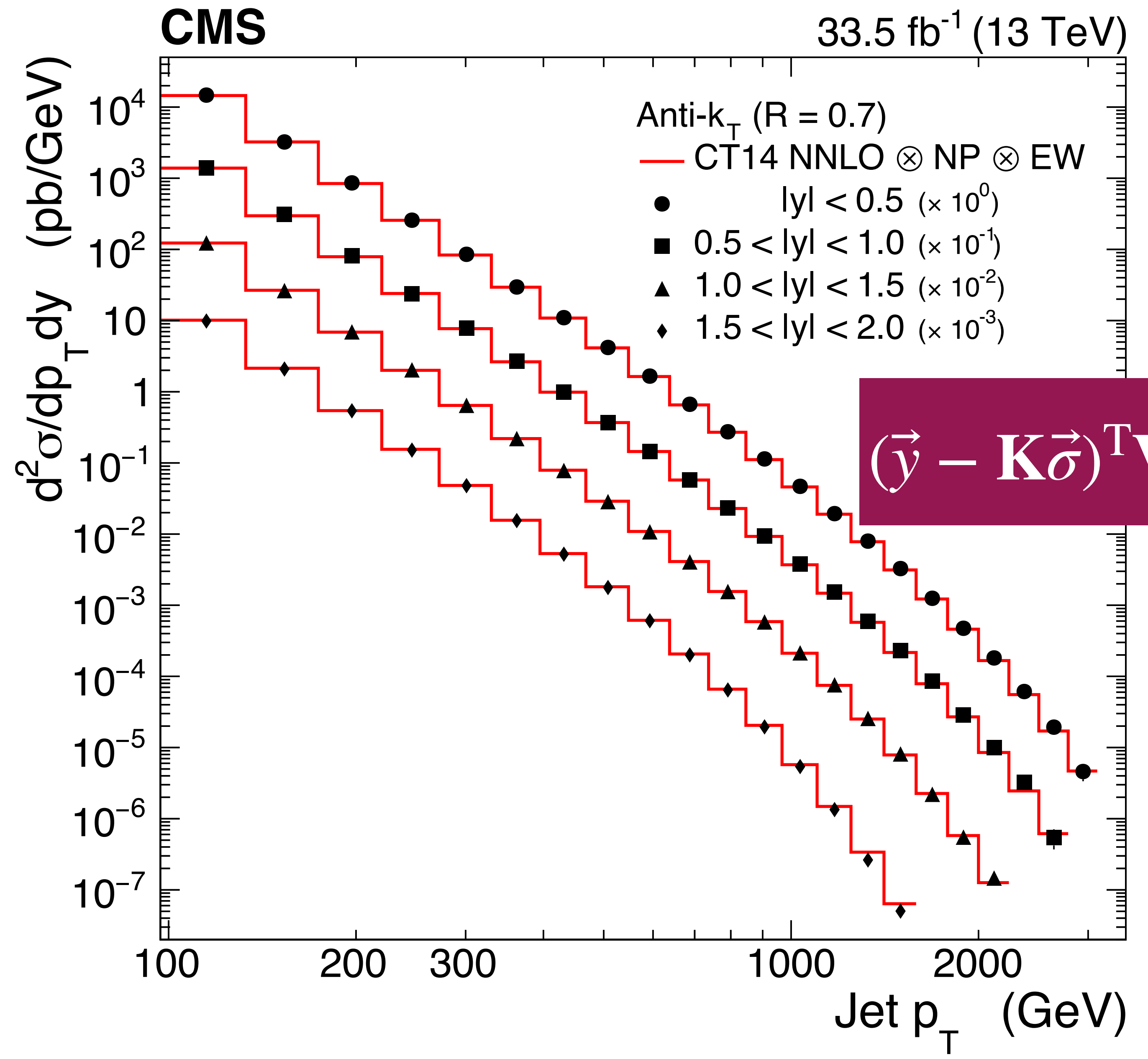
Often simplified info. made public:

$X^2(\sigma_i\, ;\, \sigma_i^{\text{meas.}})$

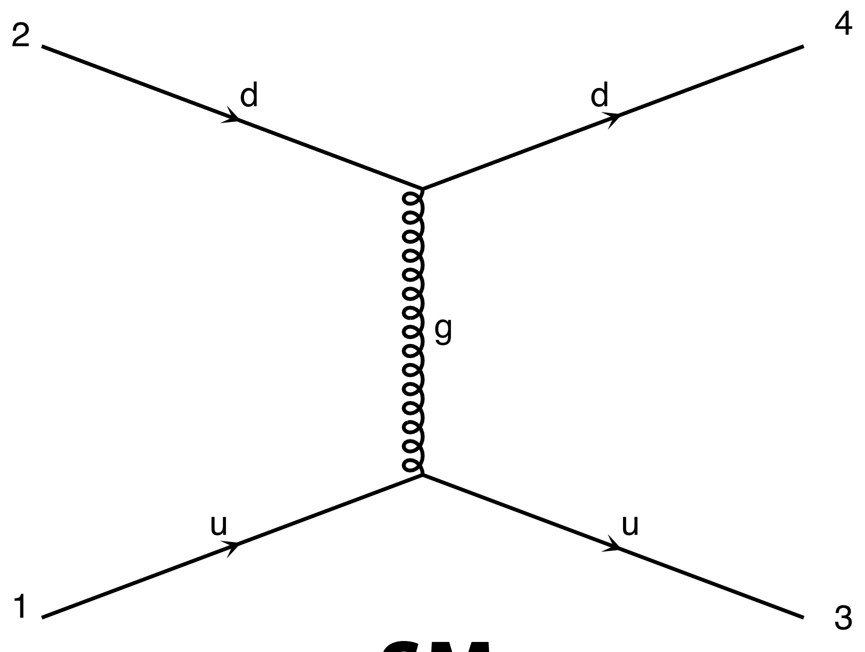
Discussed extensively in other talks!
E.g. talk from Sergio

Example: inclusive jets measurement

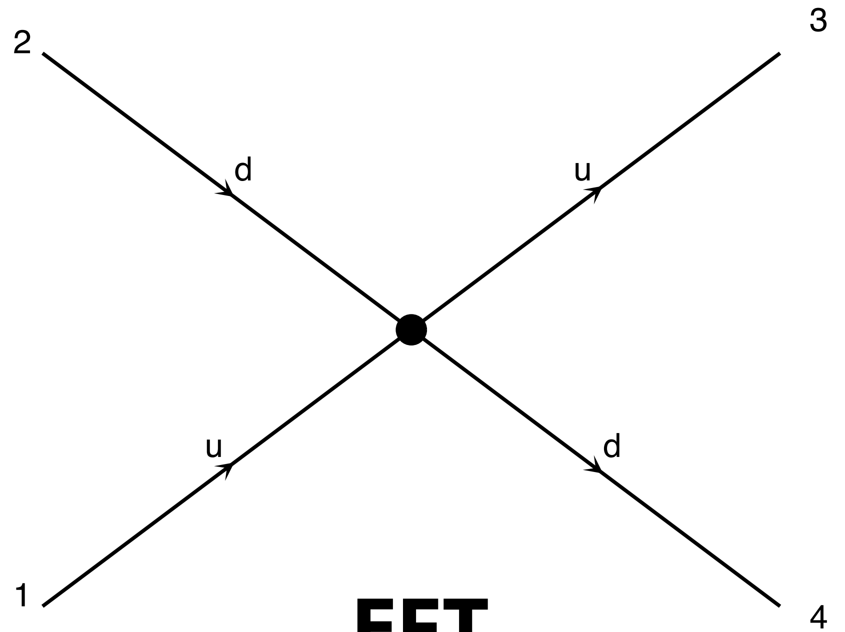
- SMP-20-011: measurement of inclusive jet production in 2016 data (36.3 fb^{-1})
- Double-differential cross section in jet p_T (97 GeV - 3.1 TeV) and $|y| < 2.0$
- Why interesting for EFT constraints?
  - Mostly through sensitivity to four-quark operators



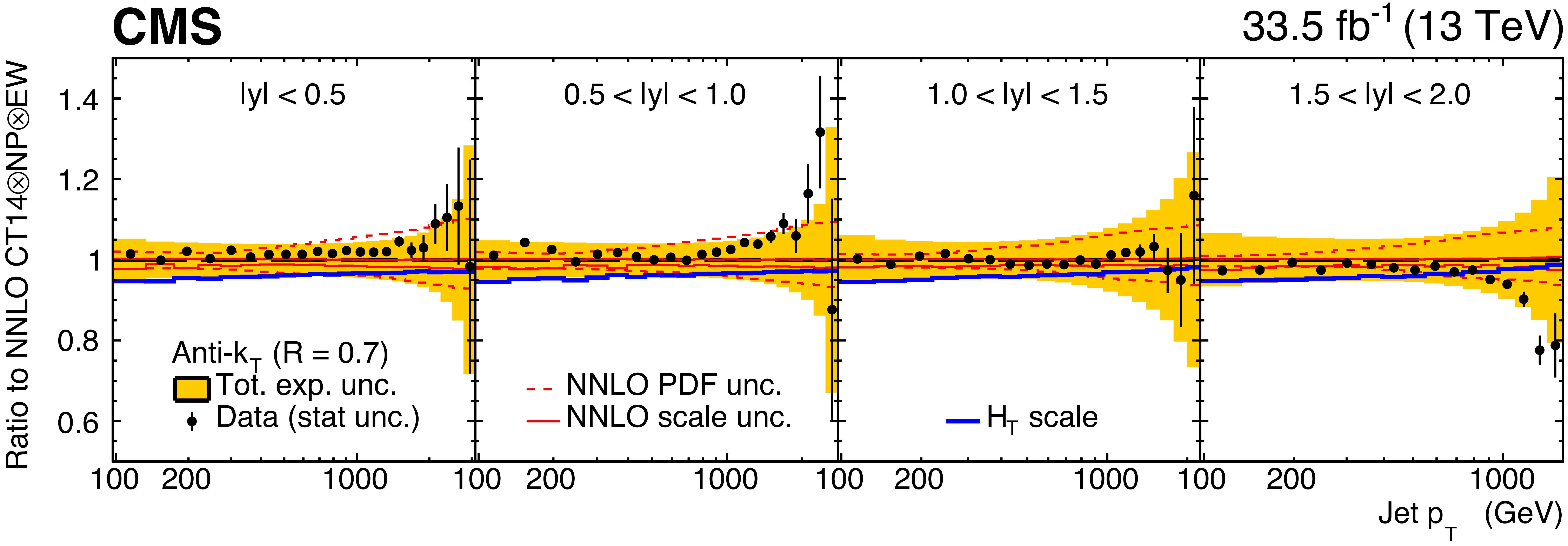
$$(\vec{y} - \mathbf{K}\vec{\sigma})^{\mathrm{T}}\mathbf{V}^{-1}(\vec{y} - \mathbf{K}\vec{\sigma}) + \delta P(\vec{\sigma})$$



SM

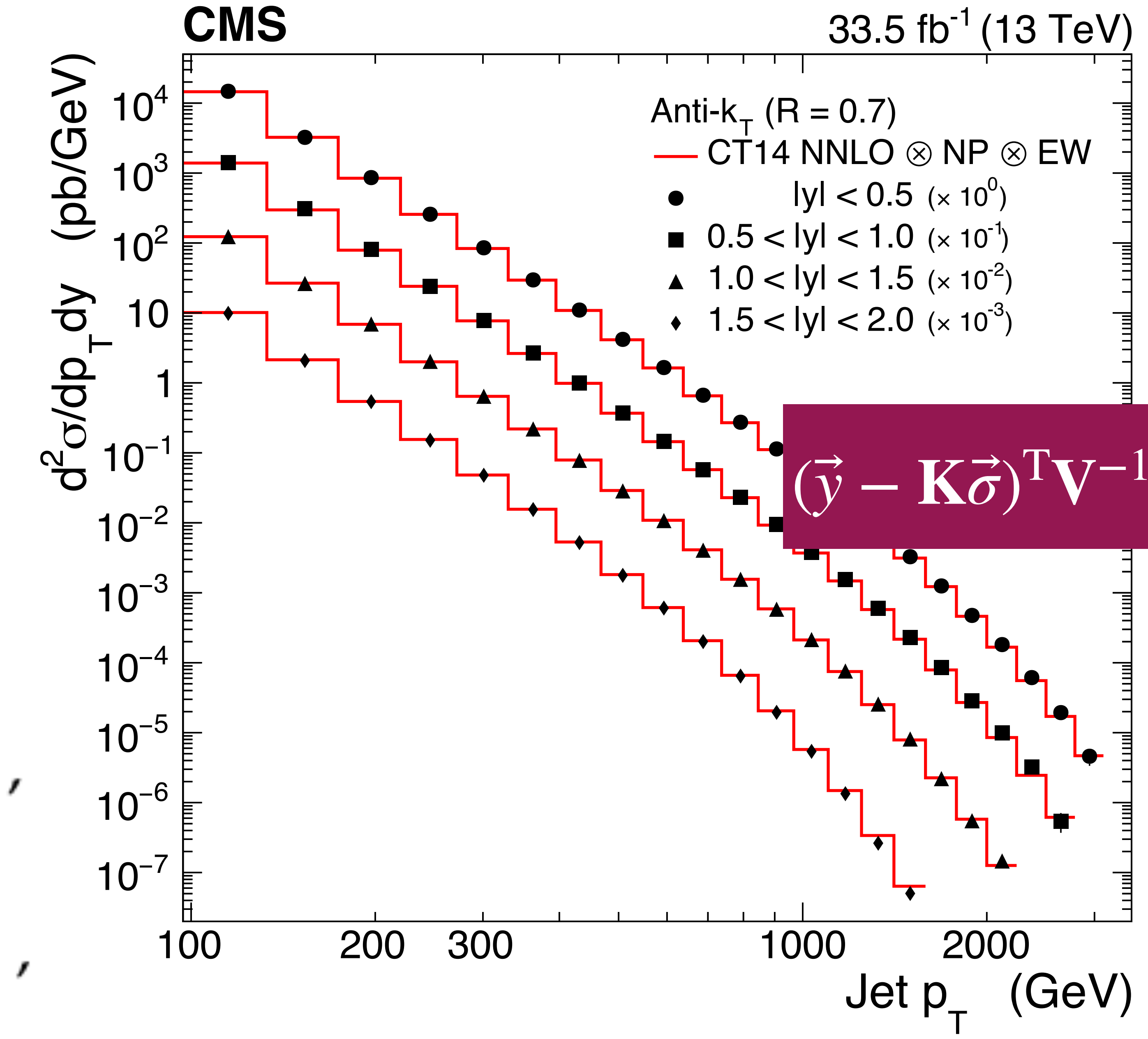


EFT

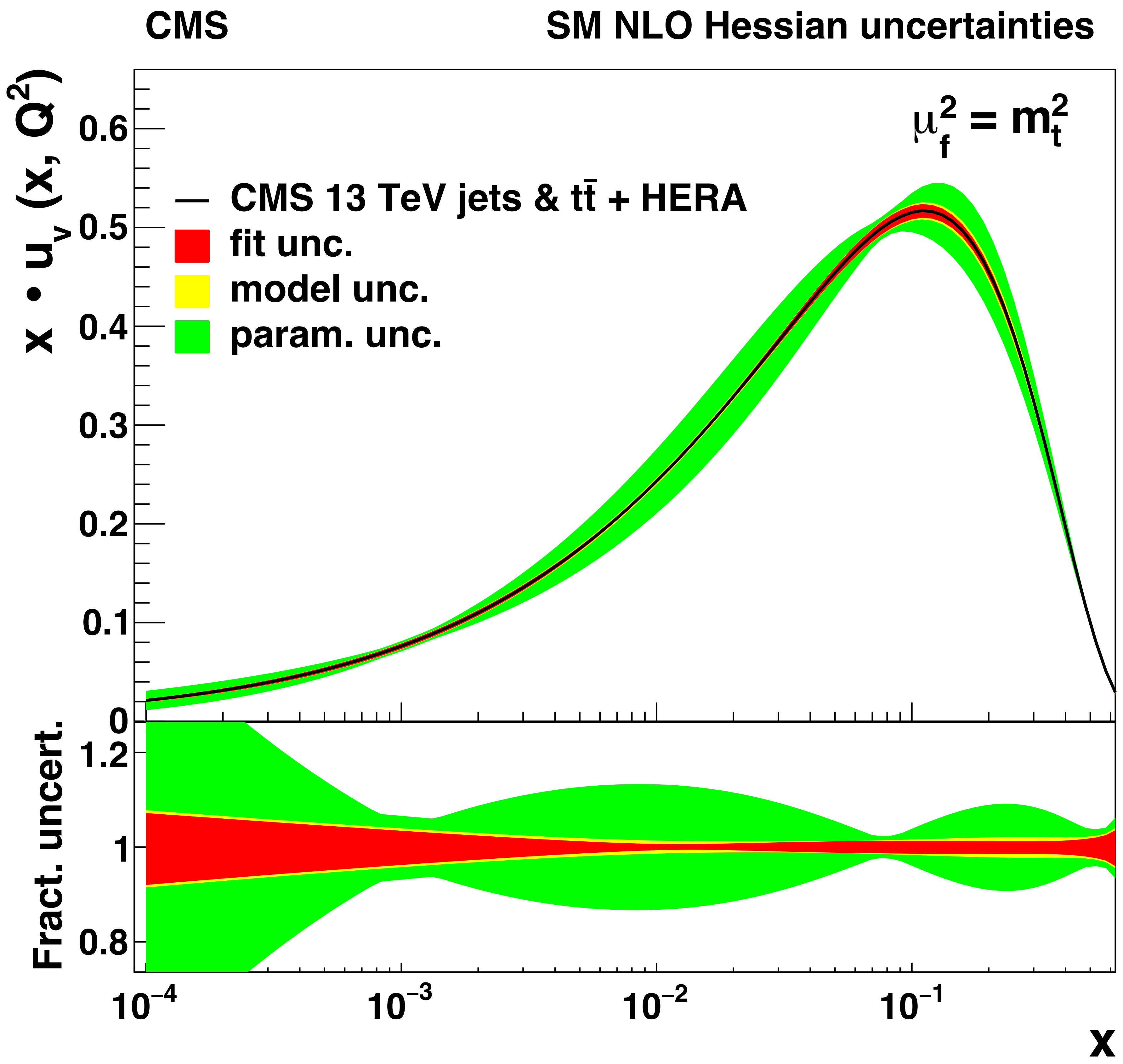


Example: inclusive jets measurement

- Using unfolded cross section and associated covariance matrix, perform simultaneous PDF and SMEFT fit using xFitter framework



$$(\vec{y} - \mathbf{K}\vec{\sigma})^{\mathrm{T}}\mathbf{V}^{-1}(\vec{y} - \mathbf{K}\vec{\sigma}) + \delta P(\vec{\sigma})$$

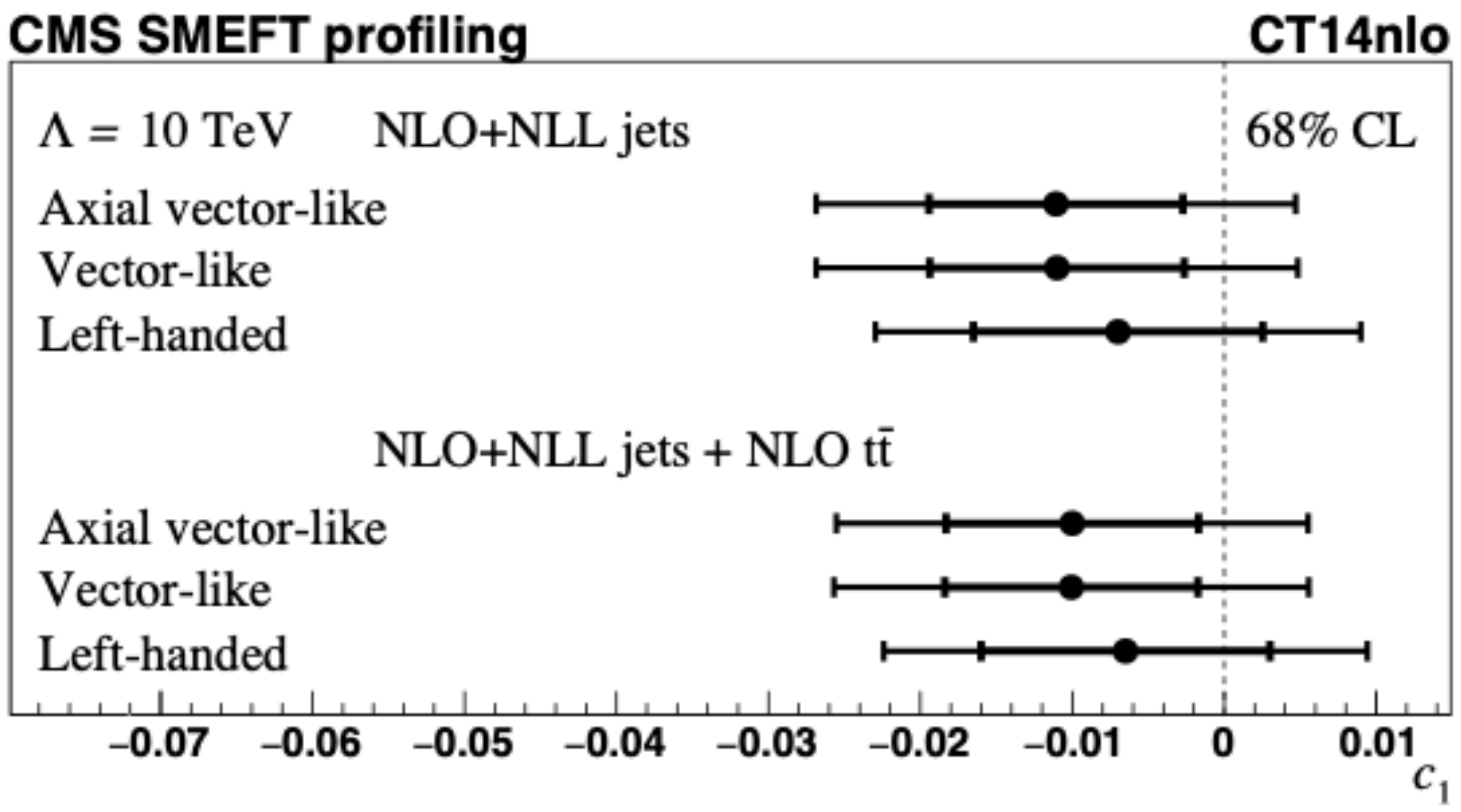


$$O_1 = \delta_{ij}\delta_{kl}\left(\sum_{c=1}^{3}\bar{q}_{Lci}\gamma_\mu q_{Lcj}\sum_{d=1}^{3}\bar{q}_{Ldk}\gamma^\mu q_{Ldl}\right),$$

$$O_3 = \delta_{ij}\delta_{kl}\left(\sum_{c=1}^{3}\bar{q}_{Lci}\gamma_\mu q_{Lcj}\sum_{d=1}^{3}\bar{q}_{Rdk}\gamma^\mu q_{Rdl}\right),$$

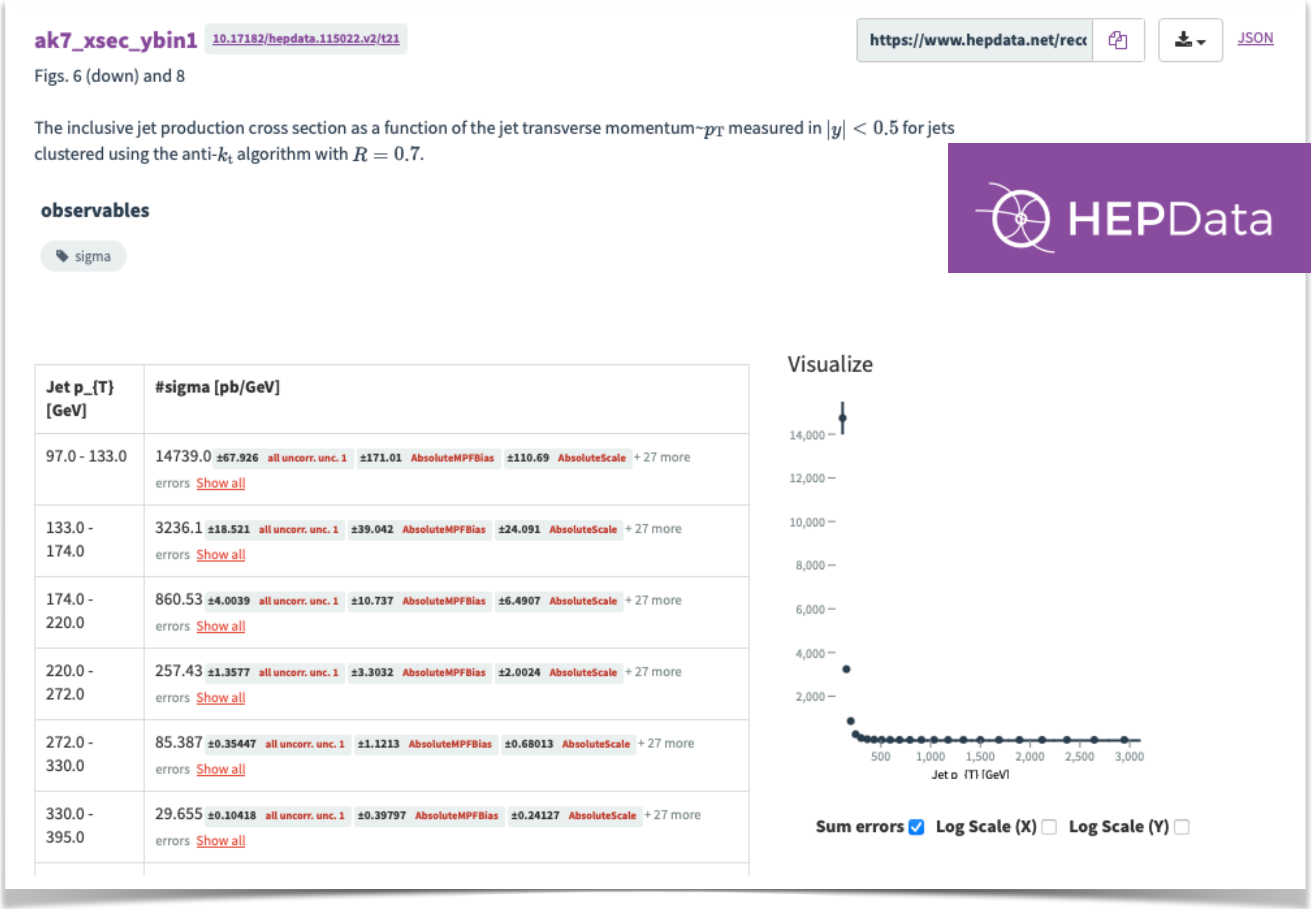
$$O_5 = \delta_{ij}\delta_{kl}\left(\sum_{c=1}^{3}\bar{q}_{Rci}\gamma_\mu q_{Rcj}\sum_{d=1}^{3}\bar{q}_{Rdk}\gamma^\mu q_{Rdl}\right),$$

- Fit combinations of four quark operators



How we could reinterpret?

- Let's pretend we want to make a new EFT fit to this data, can we do it?
  - A lot of what we need is already public:



ak7_xsec_ybin1 10.17182/hepdata.115022.v2/t21

https://www.hepdata.net/recc JSON

Figs. 6 (down) and 8

The inclusive jet production cross section as a function of the jet transverse momentum~p_T measured in $|y| < 0.5$ for jets clustered using the anti-k_t algorithm with $R = 0.7$.

HEPData

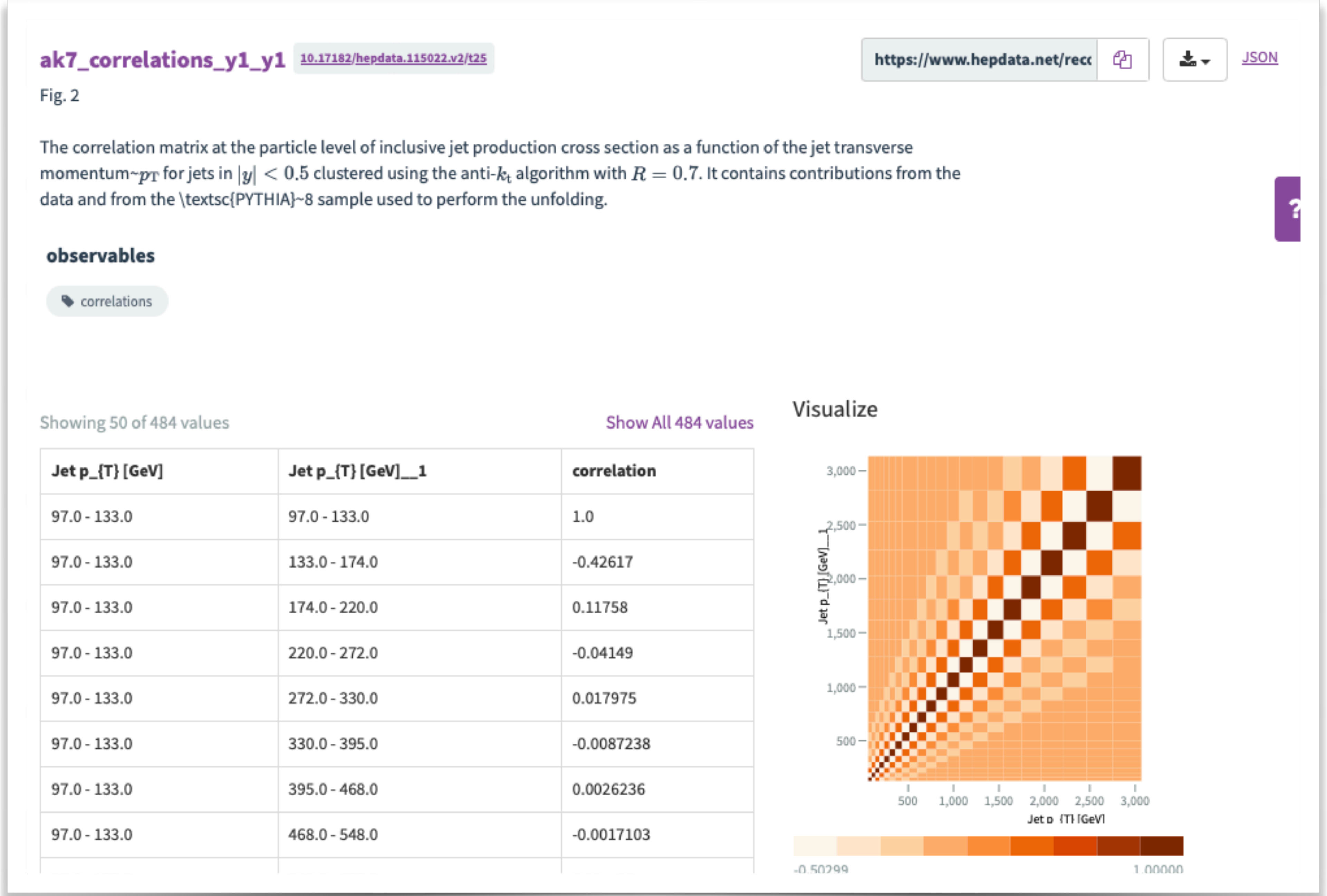
observables

sigma

Jet p_{T} [GeV]	#sigma [pb/GeV]
97.0 - 133.0	14739.0 ±67.926 all uncorr. unc. 1 ±171.01 AbsoluteMPFBias ±110.69 AbsoluteScale + 27 more errors Show all
133.0 - 174.0	3236.1 ±18.521 all uncorr. unc. 1 ±39.042 AbsoluteMPFBias ±24.091 AbsoluteScale + 27 more errors Show all
174.0 - 220.0	860.53 ±4.0039 all uncorr. unc. 1 ±10.737 AbsoluteMPFBias ±6.4907 AbsoluteScale + 27 more errors Show all
220.0 - 272.0	257.43 ±1.3577 all uncorr. unc. 1 ±3.3032 AbsoluteMPFBias ±2.0024 AbsoluteScale + 27 more errors Show all
272.0 - 330.0	85.387 ±0.35447 all uncorr. unc. 1 ±1.1213 AbsoluteMPFBias ±0.68013 AbsoluteScale + 27 more errors Show all
330.0 - 395.0	29.655 ±0.10418 all uncorr. unc. 1 ±0.39797 AbsoluteMPFBias ±0.24127 AbsoluteScale + 27 more errors Show all

Ingredients needed:
- Cross section measurements ✅
- Experimental covariance ✅
- Reference theory predictions
- Theory covariance
- EFT paramterization

Uncertainties divided into fully-correlated (across bins) components, and partially correlated (statistical), for which correlation matrices are provided ⇒ can reconstruct V_{expt}



ak7_correlations_y1_y1 10.17182/hepdata.115022.v2/t25

https://www.hepdata.net/recc JSON

Fig. 2

The correlation matrix at the particle level of inclusive jet production cross section as a function of the jet transverse momentum~p_T for jets in $|y| < 0.5$ clustered using the anti-k_t algorithm with $R = 0.7$. It contains contributions from the data and from the \textsc{PYTHIA}~8 sample used to perform the unfolding.

observables

correlations

Showing 50 of 484 values Show All 484 values

Jet p_{T} [GeV]	Jet p_{T} [GeV]__1	correlation
97.0 - 133.0	97.0 - 133.0	1.0
97.0 - 133.0	133.0 - 174.0	-0.42617
97.0 - 133.0	174.0 - 220.0	0.11758
97.0 - 133.0	220.0 - 272.0	-0.04149
97.0 - 133.0	272.0 - 330.0	0.017975
97.0 - 133.0	330.0 - 395.0	-0.0087238
97.0 - 133.0	395.0 - 468.0	0.0026236
97.0 - 133.0	468.0 - 548.0	-0.0017103

EFT parametrization

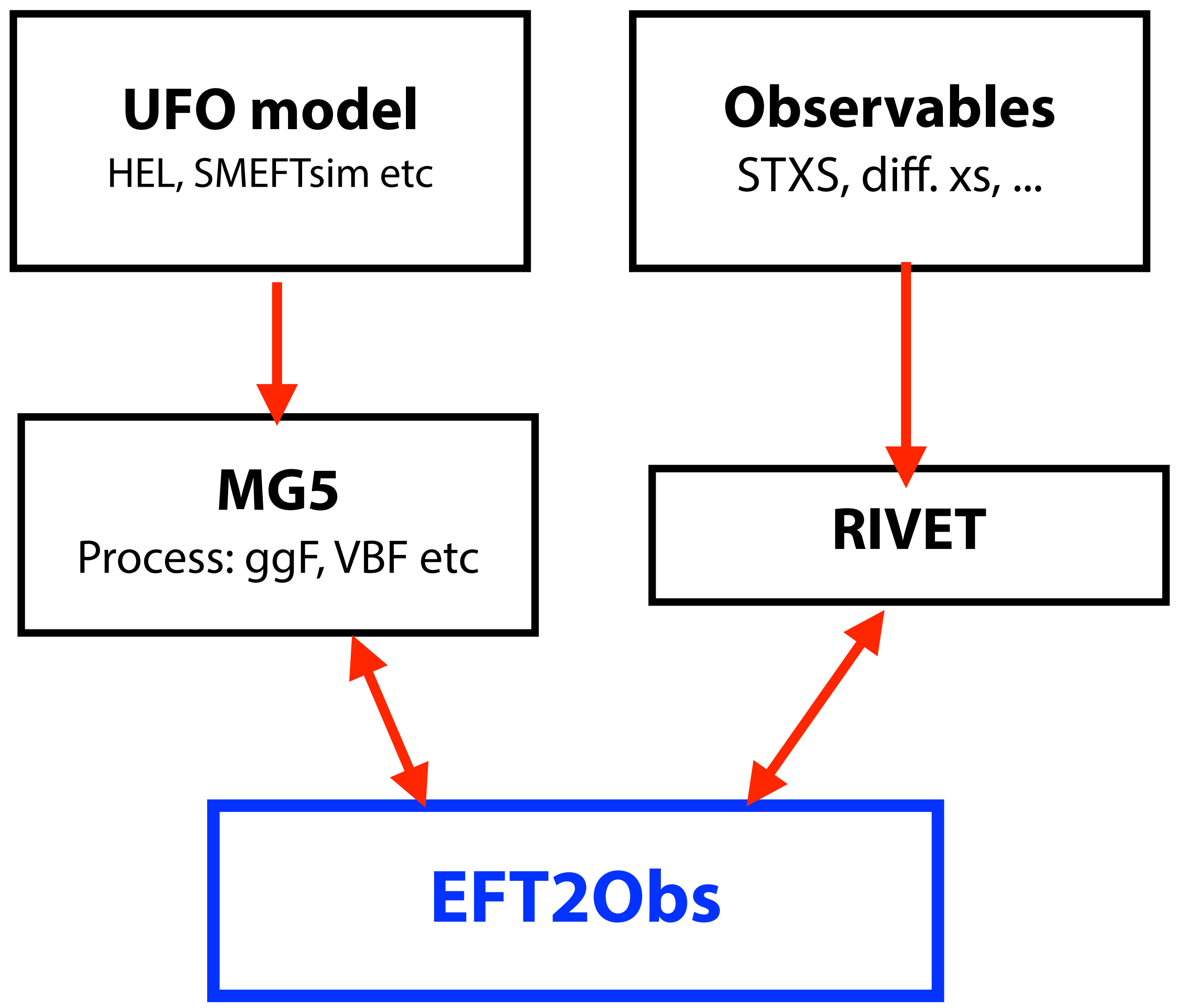
$$\sigma^i_{\mathrm{SMEFT}}(\vec{c}) = \sigma^i_{\mathrm{SM}} + \underbrace{\sigma^i_{\mathrm{int}}(\vec{c})}_{\sim\Lambda^{-2}} + \underbrace{\sigma^i_{\mathrm{BSM}}(\vec{c})}_{\sim\Lambda^{-4}} = \sigma^i_{\mathrm{SM}}\left(1 + \sum_j A^i_j c_j + \sum_{j,k} B^i_{jk} c_j c_k\right)$$

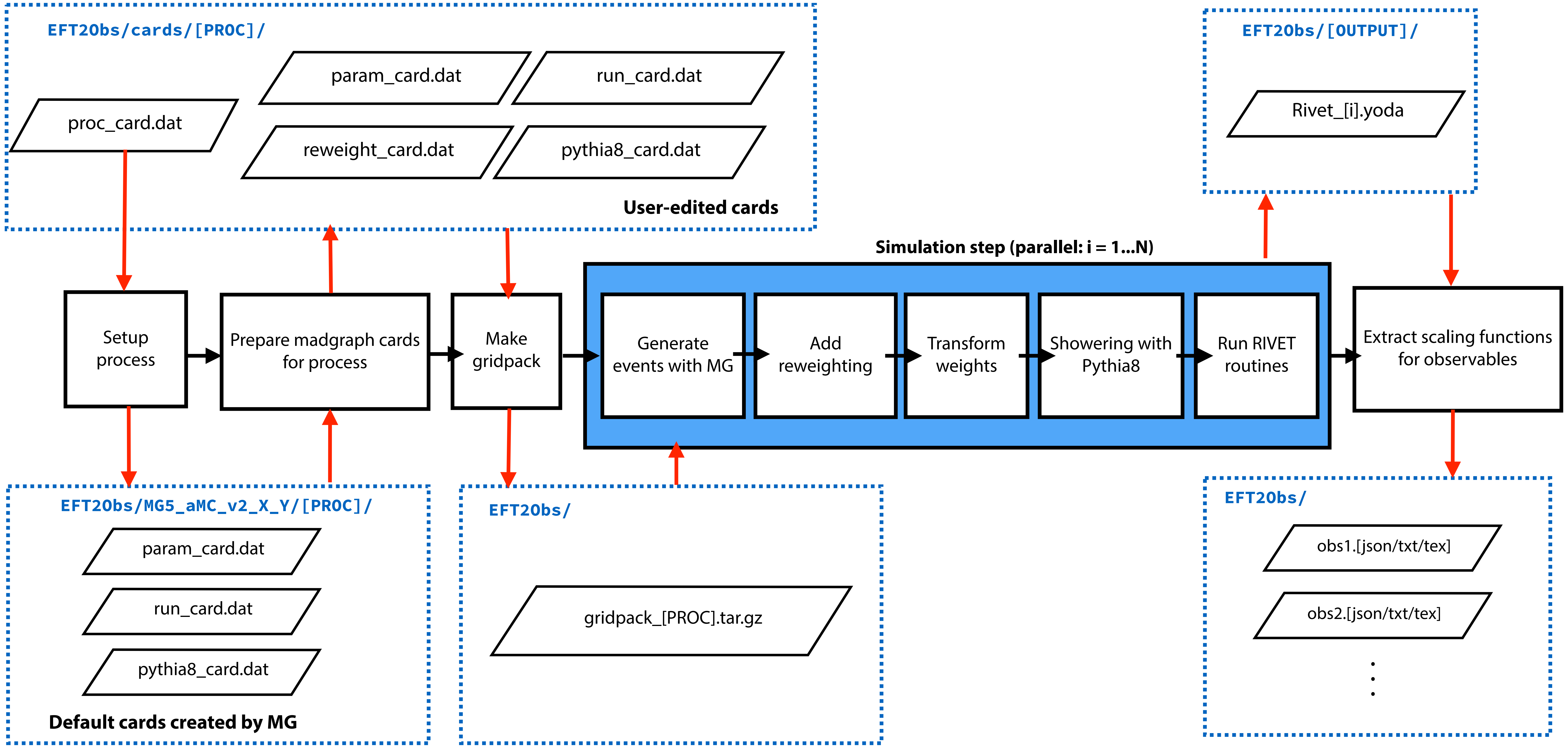
- Assume we will generate events with MG5_amc@NLO + our favourite UFO model (e.g. SMEFTsim, SMEFT@NLO, …)
- Need to implement the fiducial selection & observables
  - Best option: **RIVET** (validated, should match exactly what was defined for the analysis)
- Use the EFT2Obs tool to automate some of the steps

```
import model SMEFTsim_topU3l_MwScheme_UFO

generate p p > j j NP<=1 @0
add process p p > j j j NP<=1 @1

output Multijet-SMEFTsim3
```





Theory predictions

Ingredients needed:

- Cross section measurements ✅
- Experimental covariance ✅
- Reference theory predictions
- Theory covariance
- EFT paramterization ✅

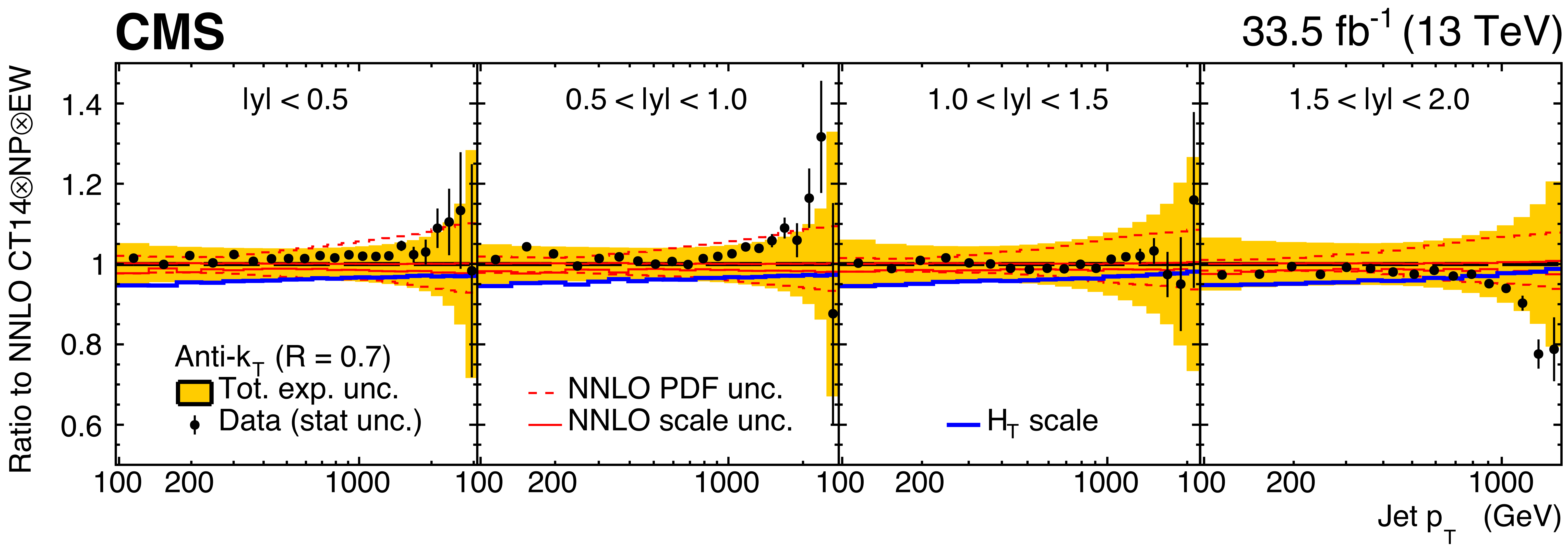
- Still one piece missing:

$$\sigma^i_{\mathrm{SMEFT}}(\vec{c}) = \sigma^i_{\mathrm{SM}} + \underbrace{\sigma^i_{\mathrm{int}}(\vec{c})}_{\sim\Lambda^{-2}} + \underbrace{\sigma^i_{\mathrm{BSM}}(\vec{c})}_{\sim\Lambda^{-4}} = \sigma^i_{\mathrm{SM}}\left(1 + \sum_j A^i_j c_j + \sum_{j,k} B^i_{jk} c_j c_k\right)$$

- We only have. LO prediction from our EFT2Obs simulation, want state-of-the-art (here NNLO QCD + NLO EWK)

- Most of our differential results come with comparison to theory prediction(s), but these are often not tabulated in HepData
  - Really need the "theory covariance matrix" ⇒ in this analysis the bin-to-bin correlations due to PDFs are clearly important

- Not a big problem! Tools exists to recalculate cross sections
  - In this case, fastNLO + interpolation tables for this measurement
  - For other processes, if we have the RIVET routine we can generate our state-of-the-art MC and derive systematics



Choice of observables

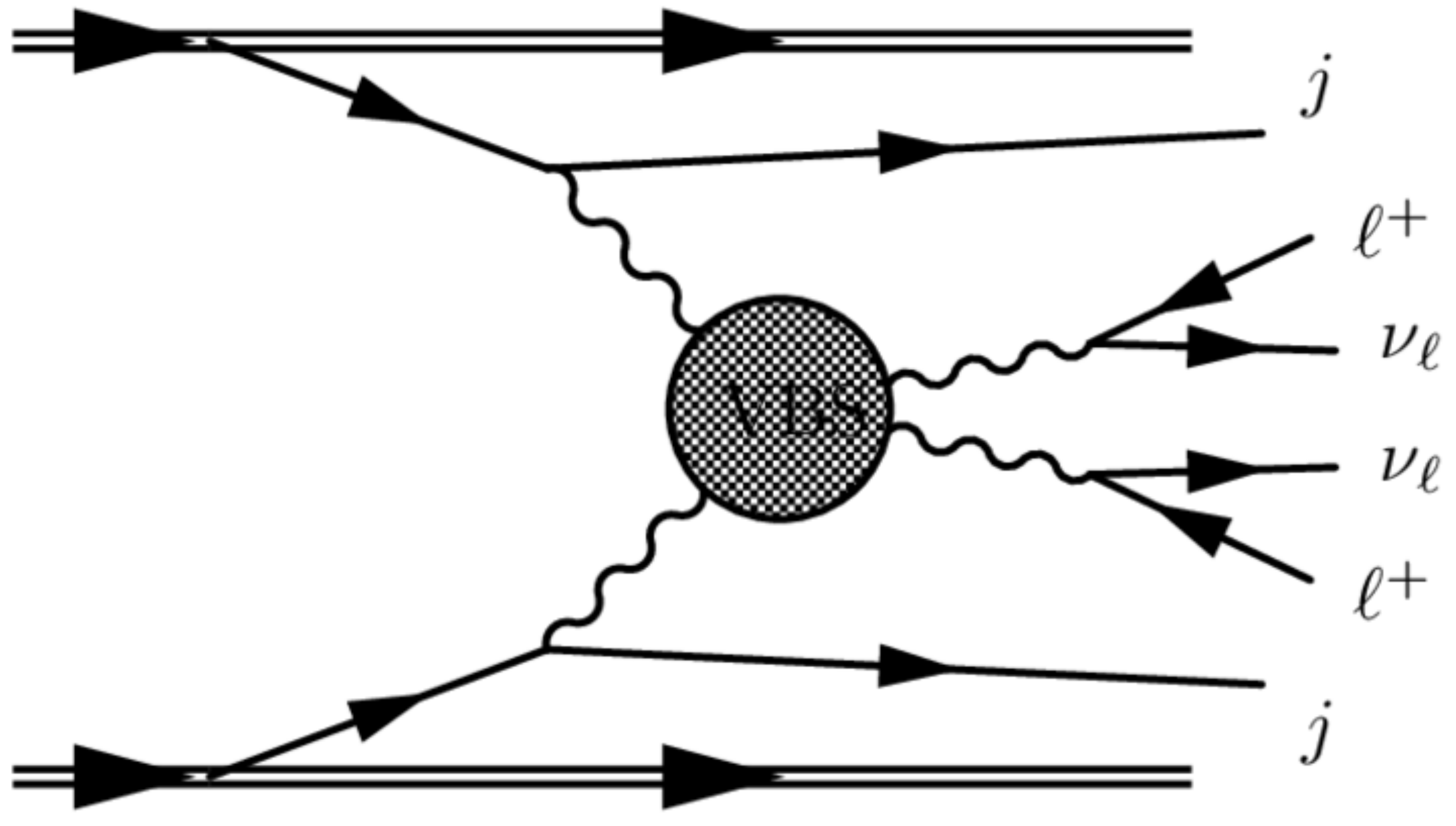
- For some processes can measure fully differential cross sections, e.g. $q\bar{q} \to \ell^+\ell^-$ at tree level:

$$\frac{d^3\sigma}{dm_{\ell\ell}^2\, dc_*\, dy} = \frac{\tau}{3\cdot 64\,\pi\, m_{\ell\ell}^4} \sum_q \Big\{ \left[(1+c_*)^2 \mathcal{L}_q(\tau,y) + (1-c_*)^2 \mathcal{L}_q(\tau,-y)\right] P_s^q(m_{\ell\ell}) + \left[(1-c_*)^2 \mathcal{L}_q(\tau,y) + (1+c_*)^2 \mathcal{L}_q(\tau,-y)\right] P_o^q(m_{\ell\ell}) \Big\}$$

[2103.10532]

- But in general not possible, e.g. VBS (2→6 process)
  - Can try to find most sensitive observables for EFT
  - Phenomenological study performed for all main VBS processes and relevant dim-6 operators:
  - Different observables provide different sensitivity to different operators [JHEP 05 (2022) 039]
  - No single "best" observable for simultaneous fit

Op.	SSWW+2j		OSWW+2j	
	L	L+Q	L	L+Q
$c_{Hl}^{(1)}$	—	m_{ll}	—	MET
$c_{Hl}^{(3)}$	$\Delta\eta_{jj}{}^\dagger$	$\Delta\eta_{jj}{}^\dagger$	$m_{jj}{}^\dagger$	$m_{jj}{}^\dagger$
$c_{Hq}^{(1)}$	p_{T,j^1}	p_{T,j^1}	m_{jj}	m_{ll}
$c_{Hq}^{(3)}$	$\Delta\phi_{jj}$	$\Delta\phi_{jj}$	m_{ll}	m_{ll}
$c_{qq}^{(3)}$	$m_{ll}{}^\dagger$	p_{T,j^2}	m_{jj}	p_{T,j^2}
$c_{qq}^{(3,1)}$	$\Delta\phi_{jj}$	p_{T,j^2}	m_{jj}	p_{T,j^2}

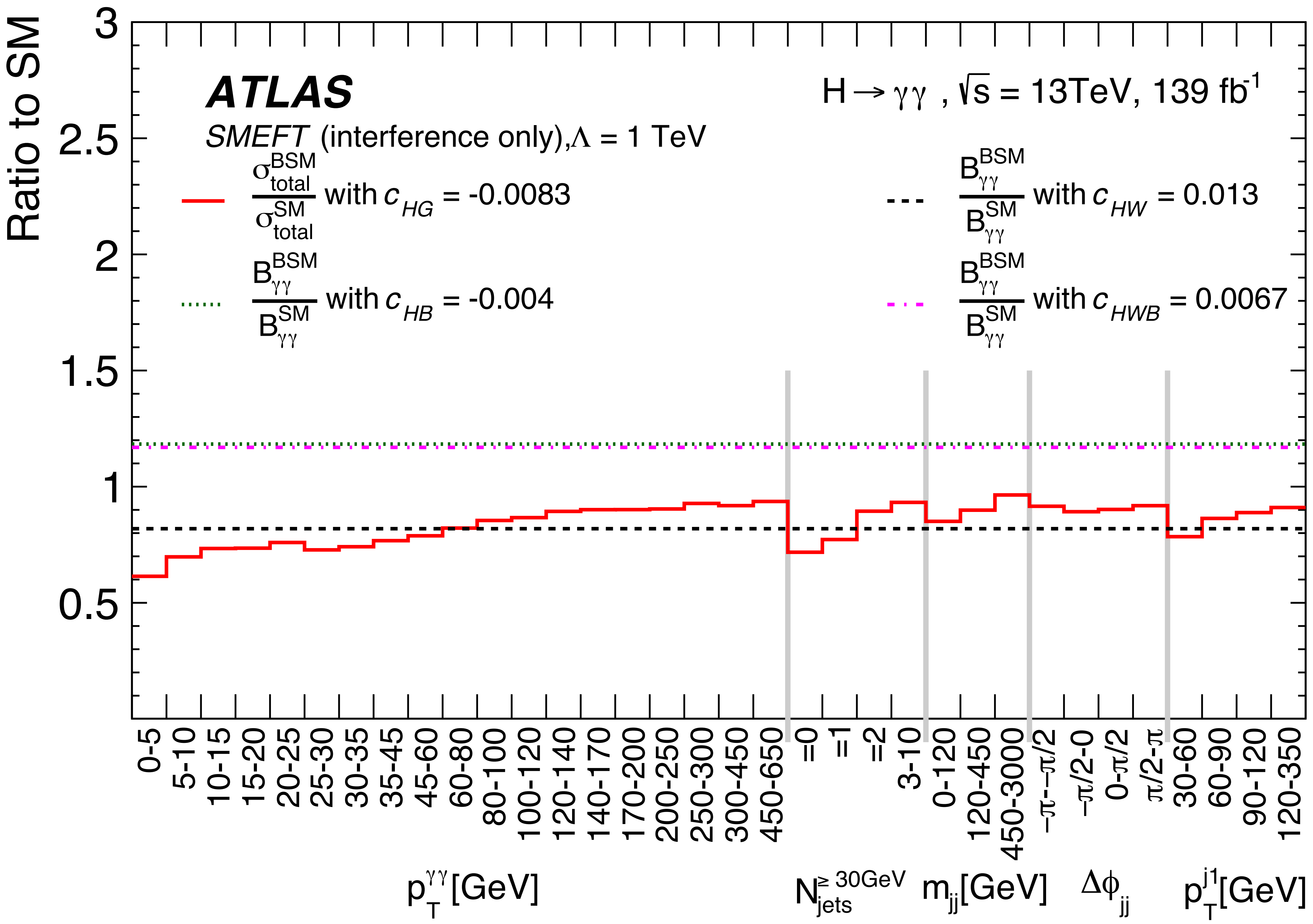


Choice of observables

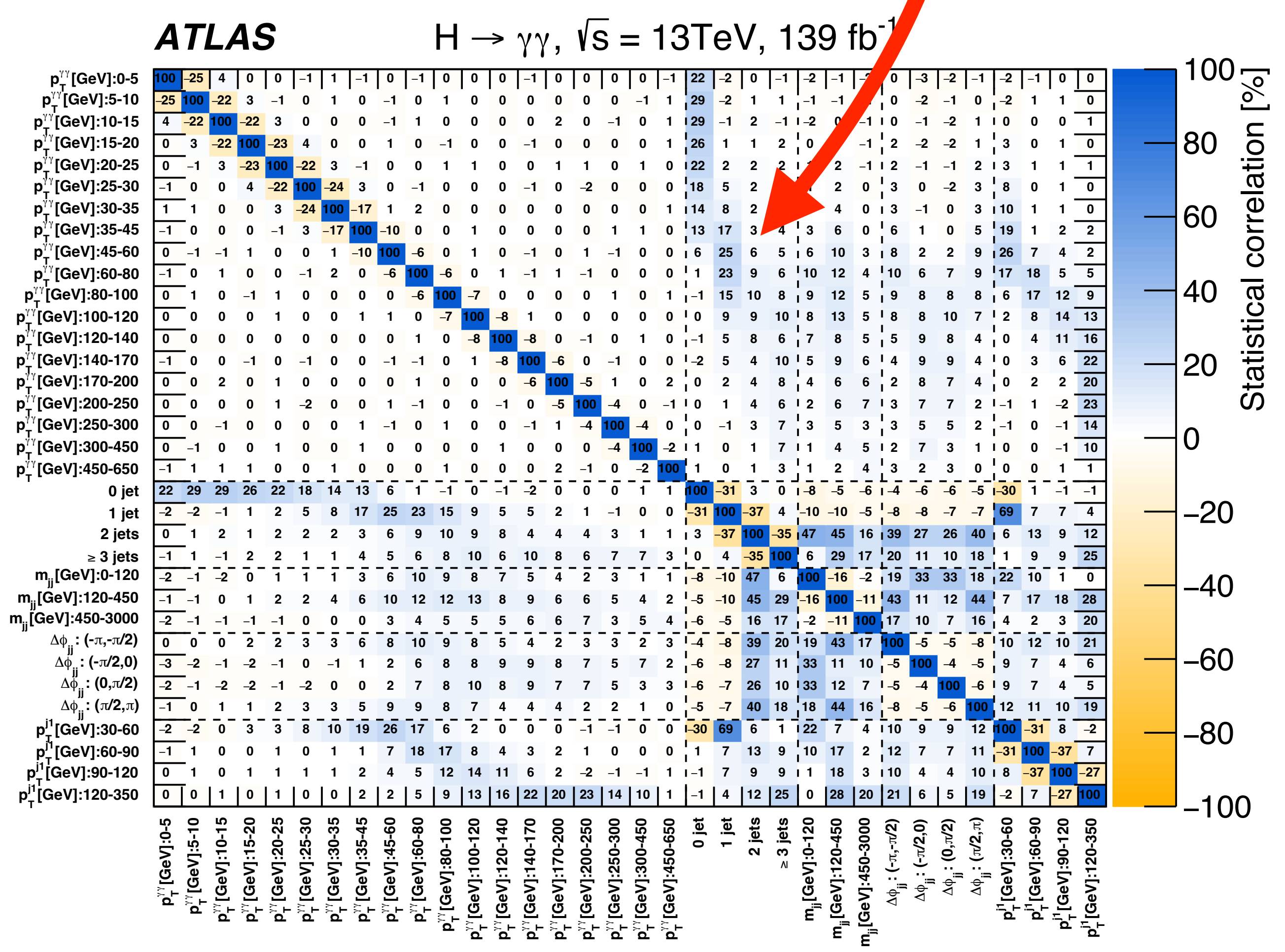
- One possible solution: measure N x 1D differential cross sections, and reinterpret simultaneously
  - Same events in each measurement ⇒ account for statistical correlations (e.g. by bootstrapping)

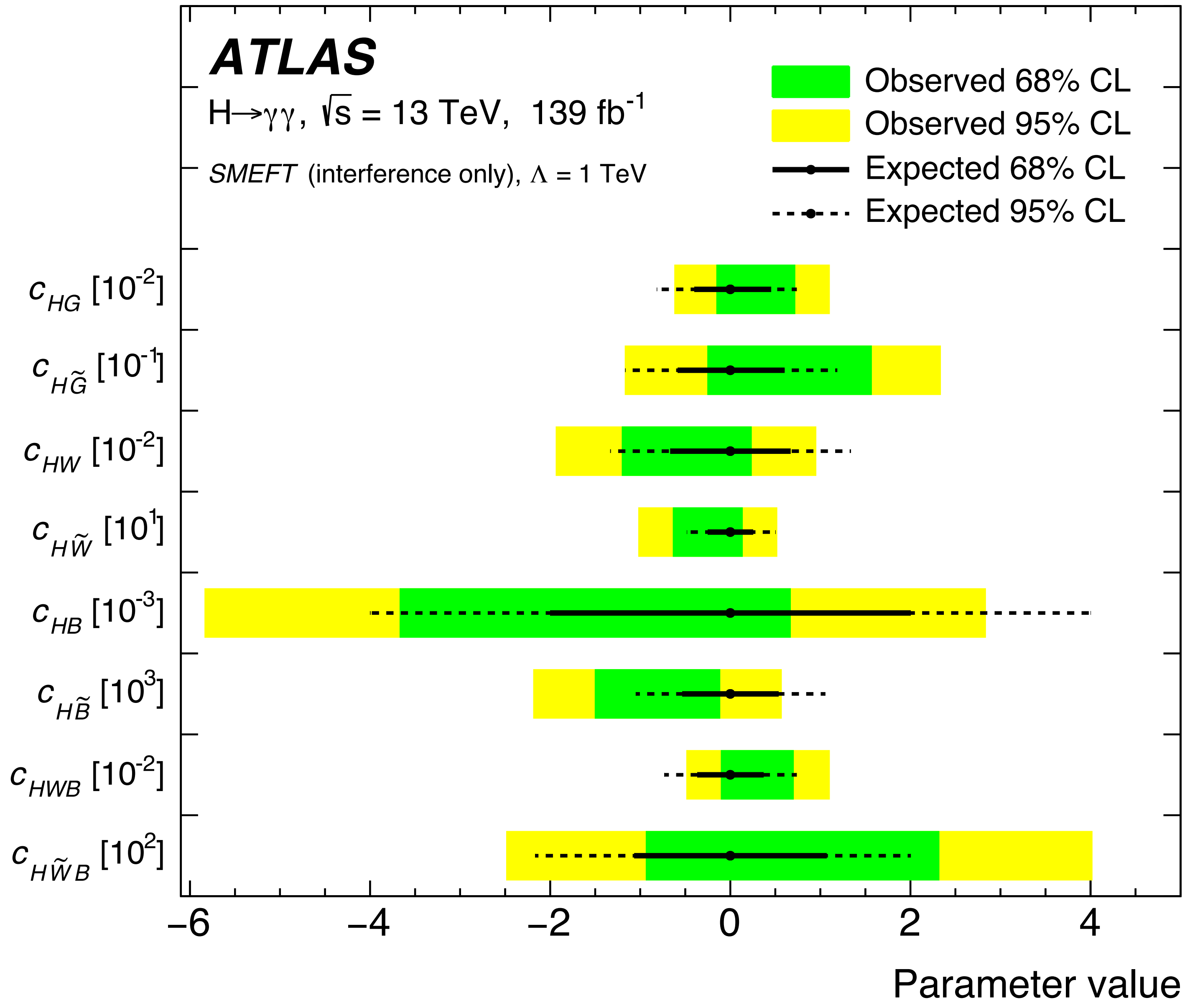
- Example, ATLAS H→γγ differential cross sections [JHEP 08 (2022) 027]

- Limitations: contains less information than full joint PDF, Gaussian approx may not be valid with small event counts



$$L = \exp\left[-\frac{1}{2}\left(\sigma_{\text{obs}} - \sigma_{\text{pred}}\right)^{T} C^{-1}\left(\sigma_{\text{obs}} - \sigma_{\text{pred}}\right)\right]$$

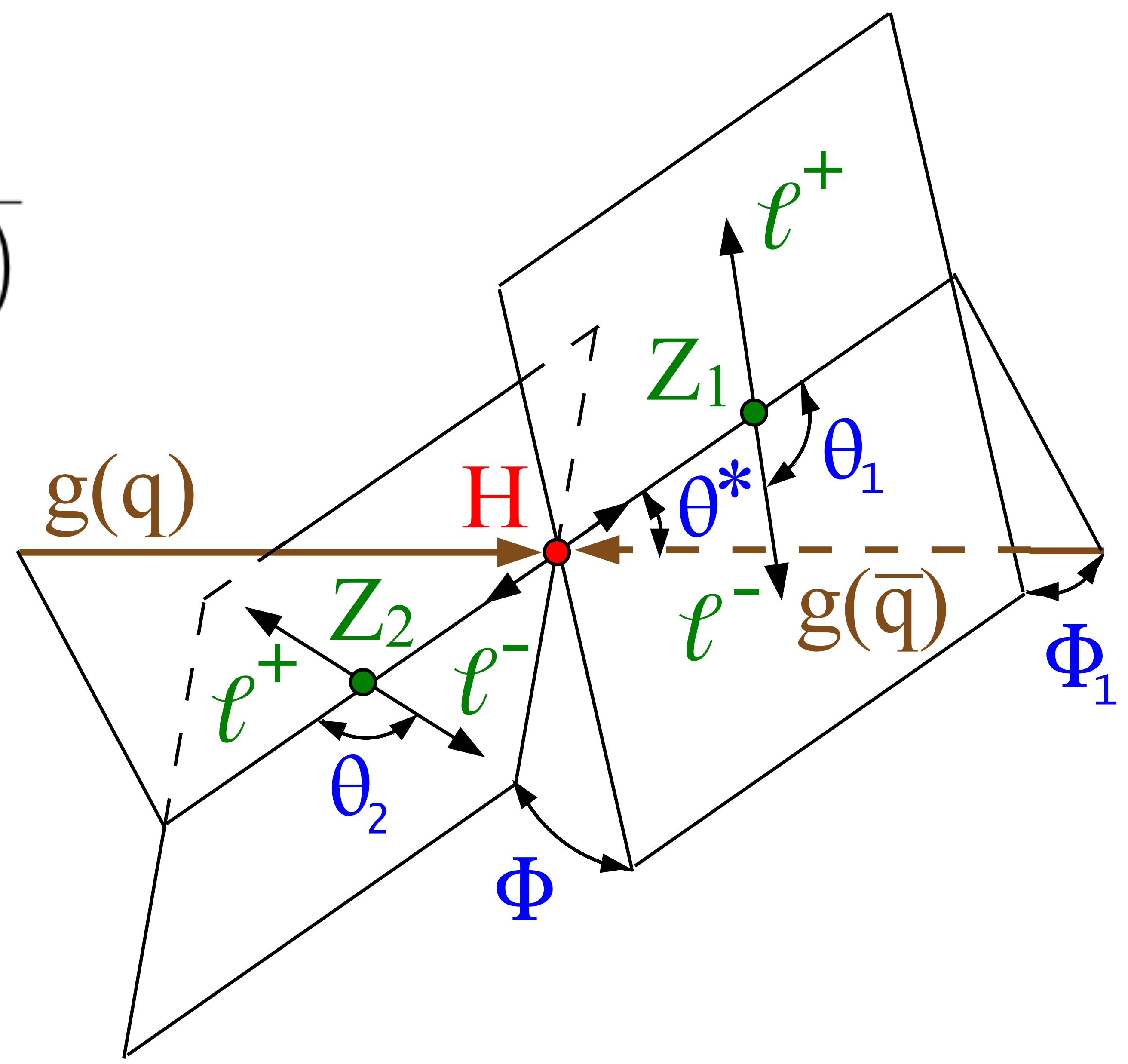


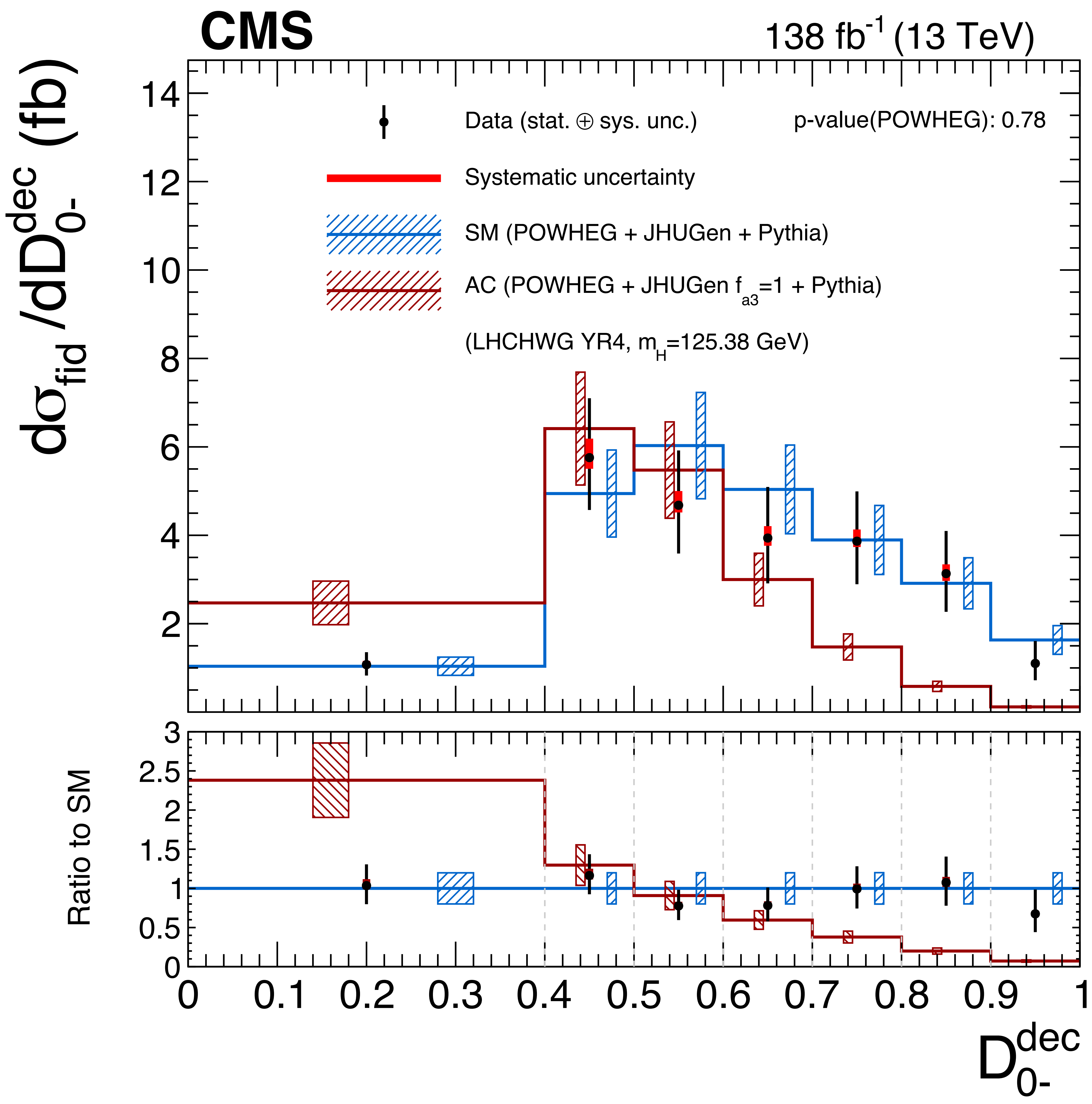


Alternative: "non-trivial" observables

- Well established that we can build optimal discriminators from ratios of squared matrix elements (or train a DNN to learn them)
- These discriminator distributions can be unfolded too
- Example: CMS $H \rightarrow ZZ \rightarrow 4l$ differential cross section [HIG-21-009]
- Possible limitations: given discriminator not optimal for all c_j in general

$$\mathcal{D}_{\text{alt}}\left(\vec{\Omega}\right) = \frac{\mathcal{P}_{\text{sig}}\left(\vec{\Omega}\right)}{\mathcal{P}_{\text{sig}}\left(\vec{\Omega}\right) + \mathcal{P}_{\text{alt}}\left(\vec{\Omega}\right)}$$

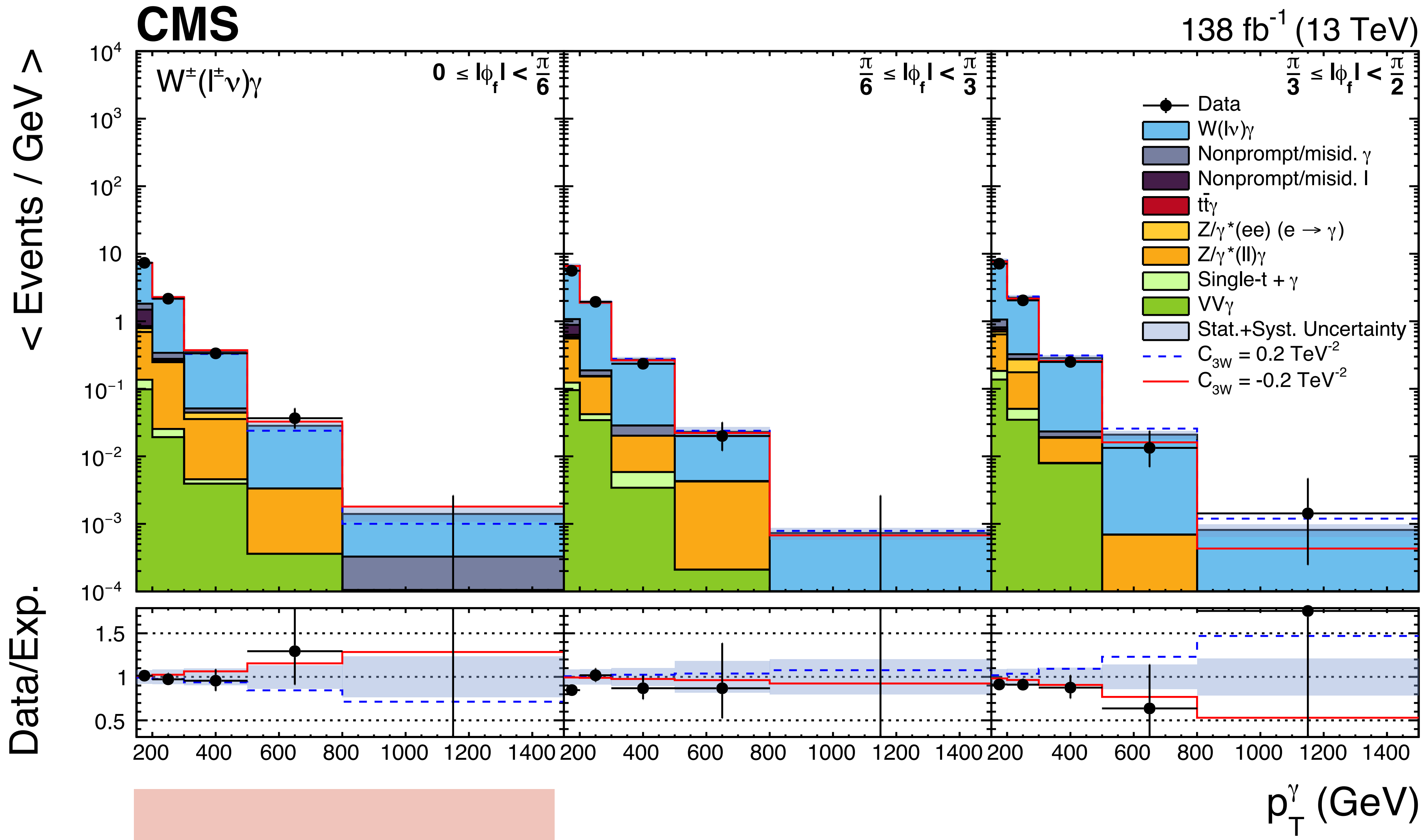


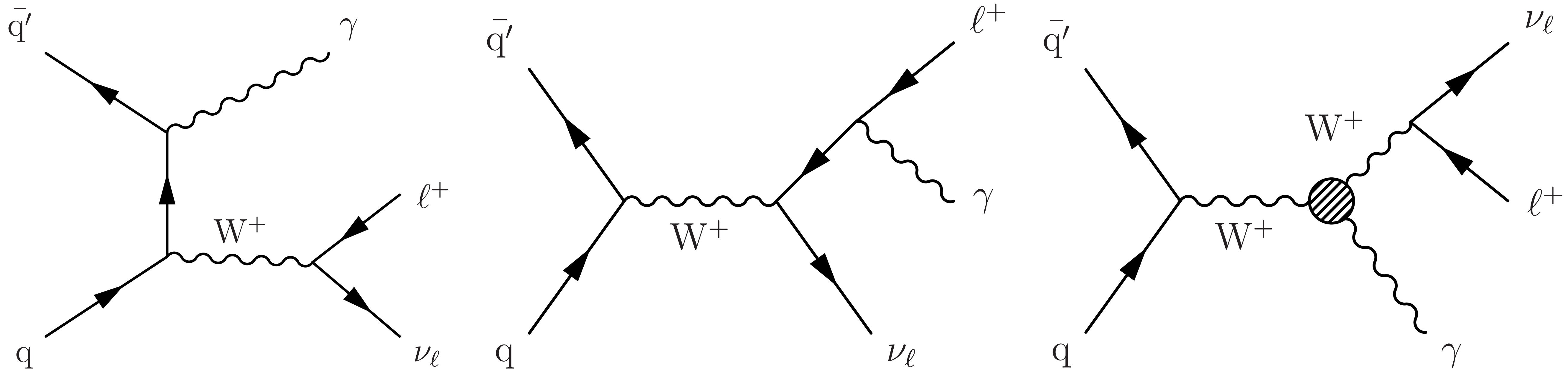


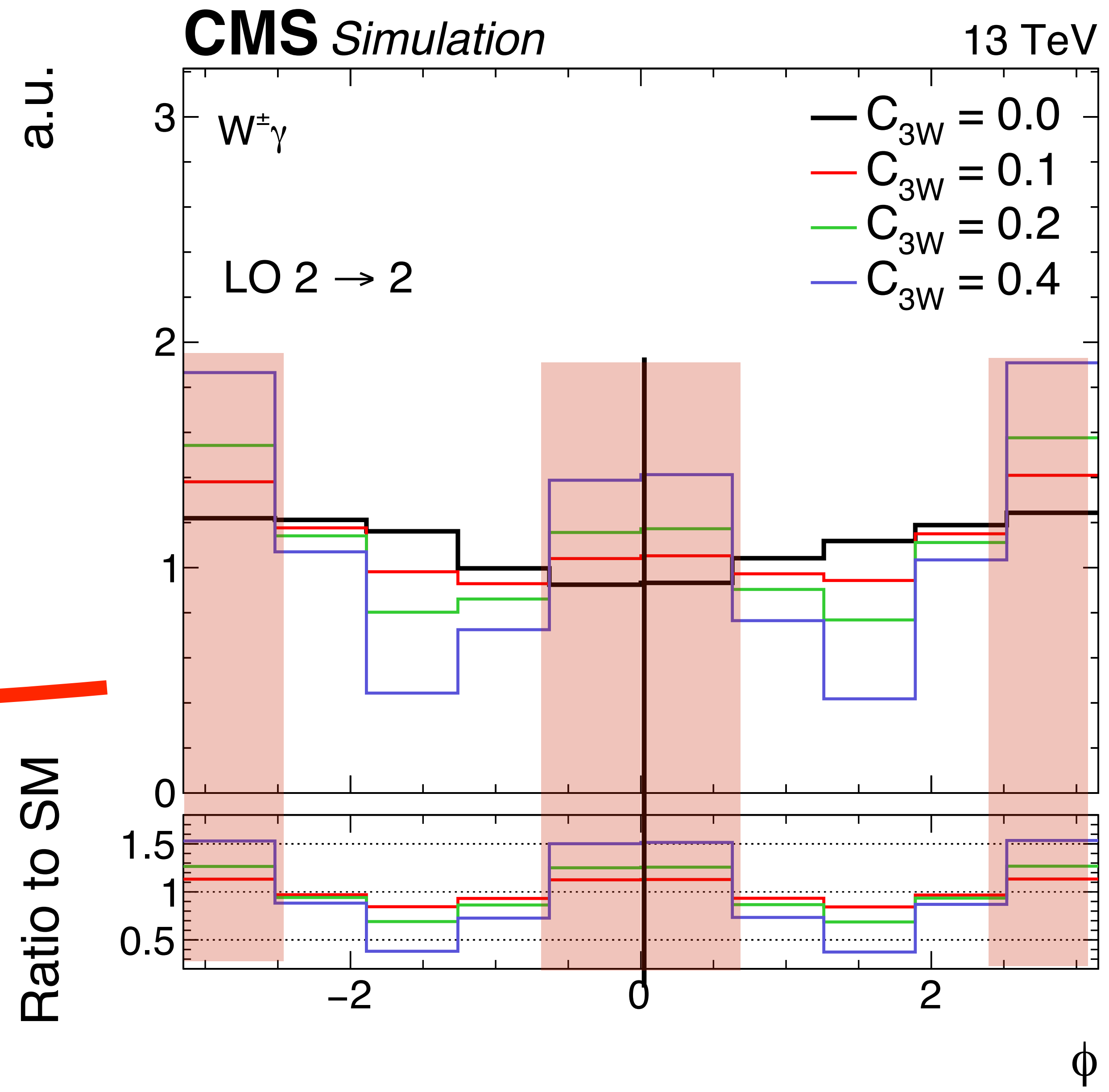
Another example: Wγ differential cross section

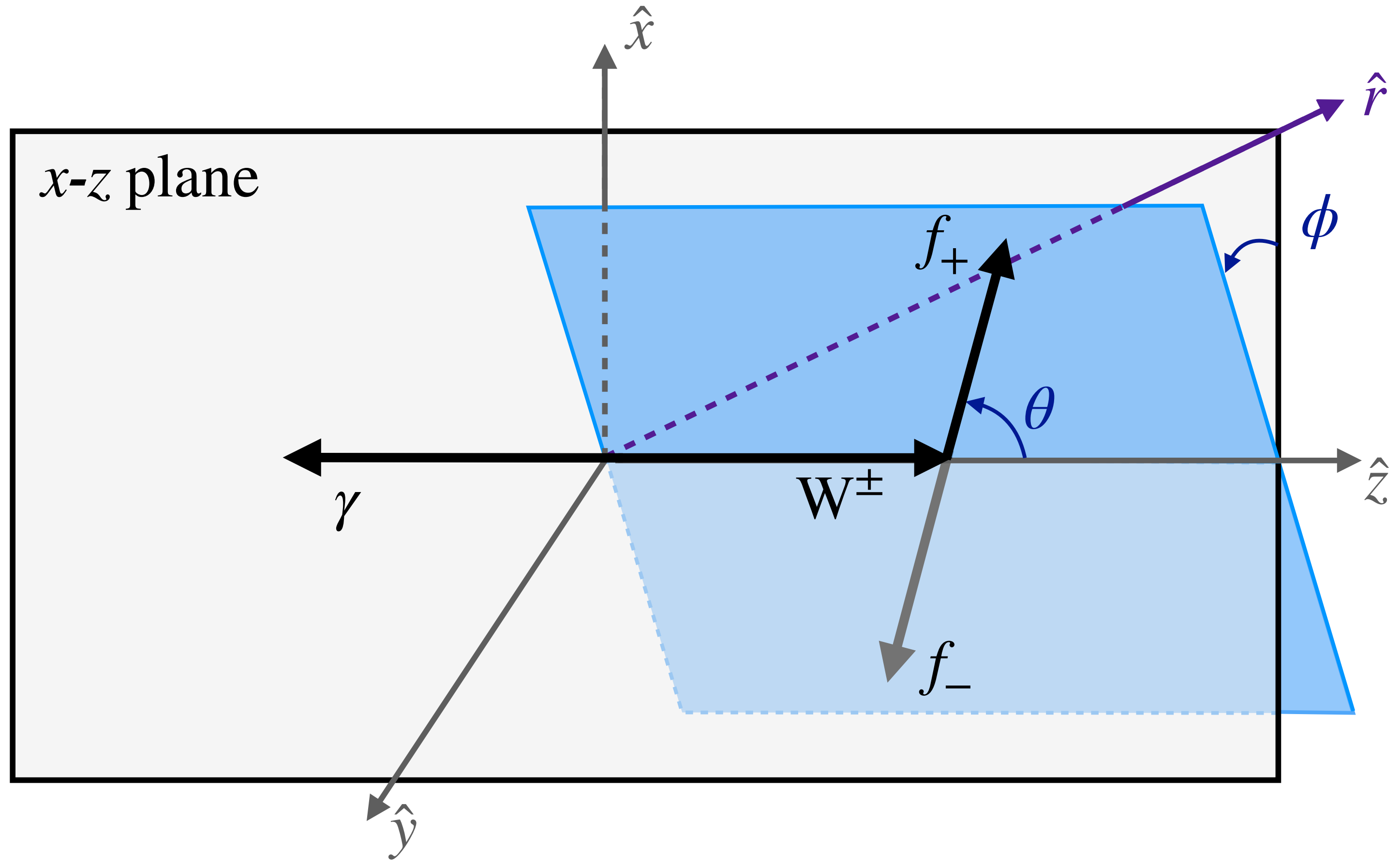
L(data | σ_i) → L(data | $\sigma_i(c_j)$)

- 2D measurement in $p_T(\gamma)$ [energy growth] and φ_f [interference resurrection]



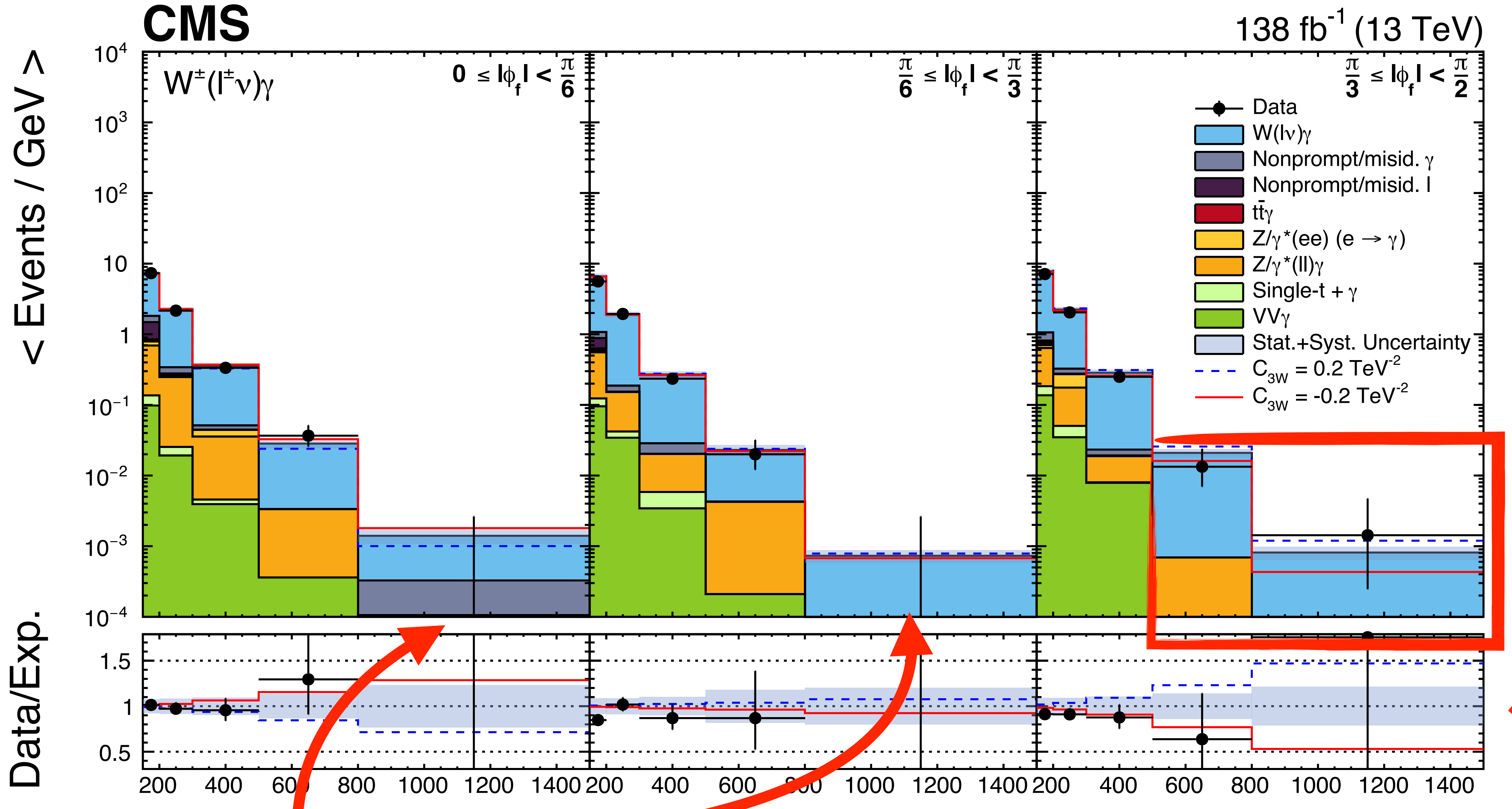






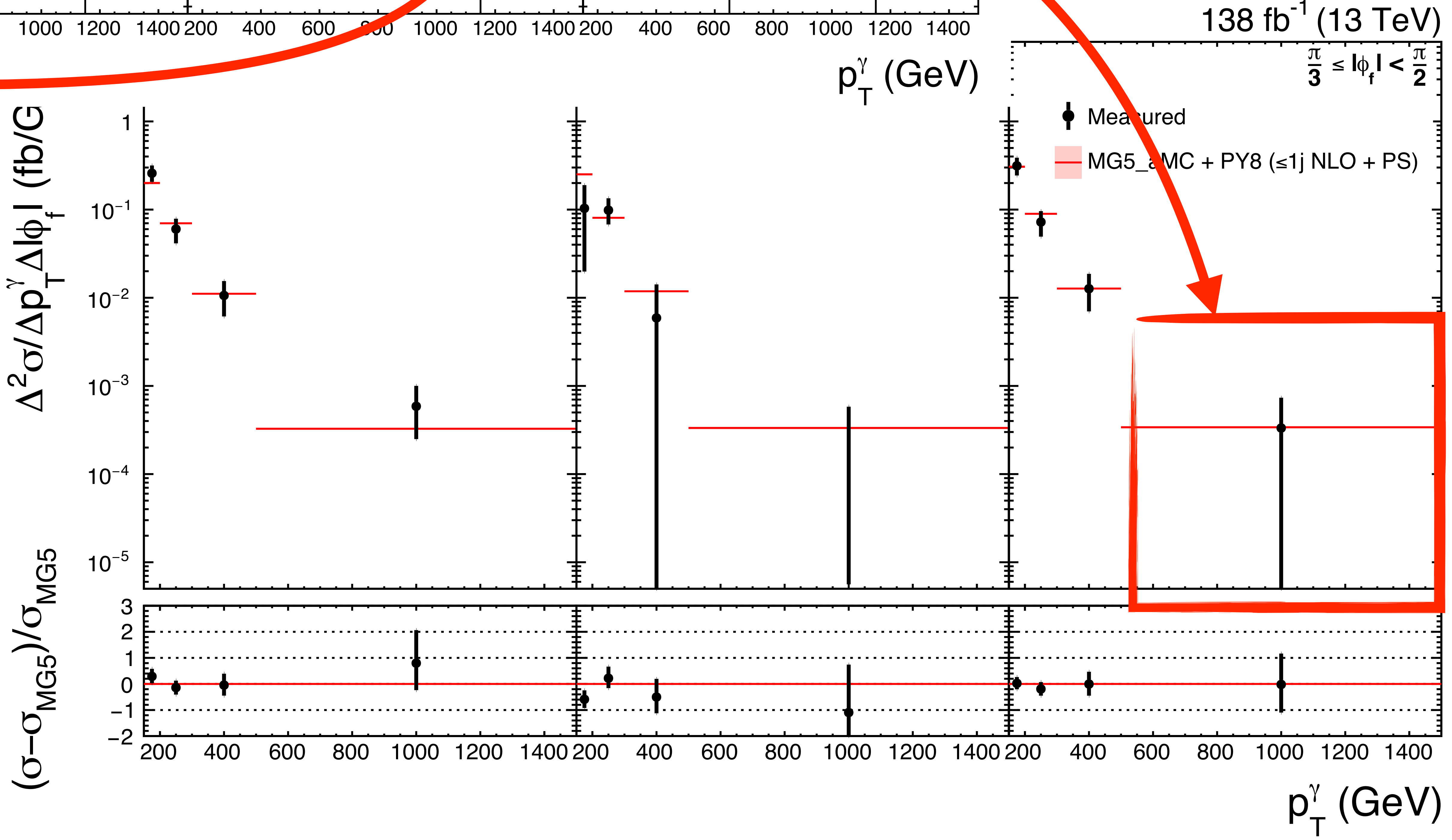
- Measure in three bins of $|\varphi_f|$. Limited by p_T^{miss} resolution

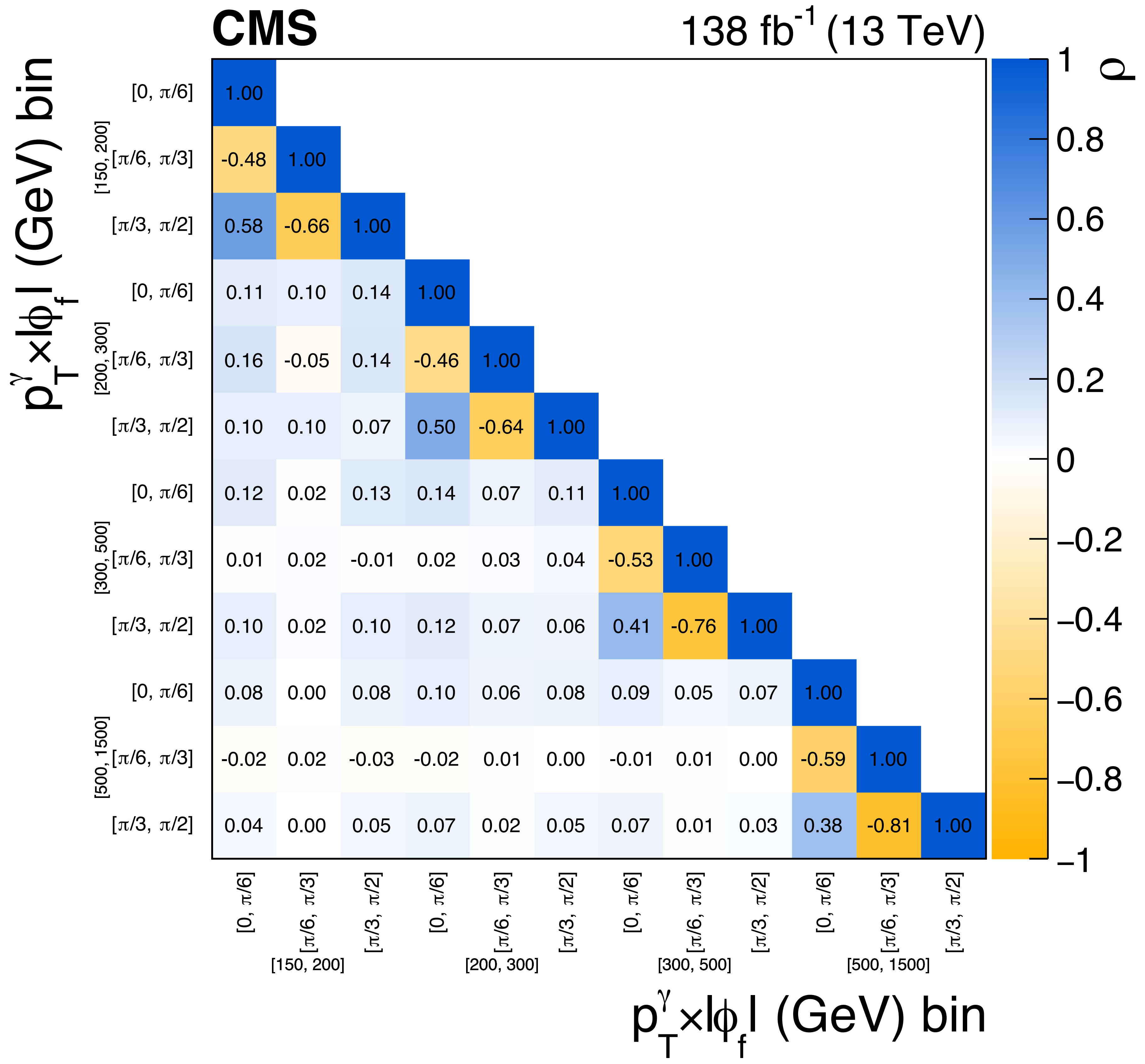
Binning limitations



two highest p_T bins merged for unfolded cross section

only ~ 1 event / bin expected

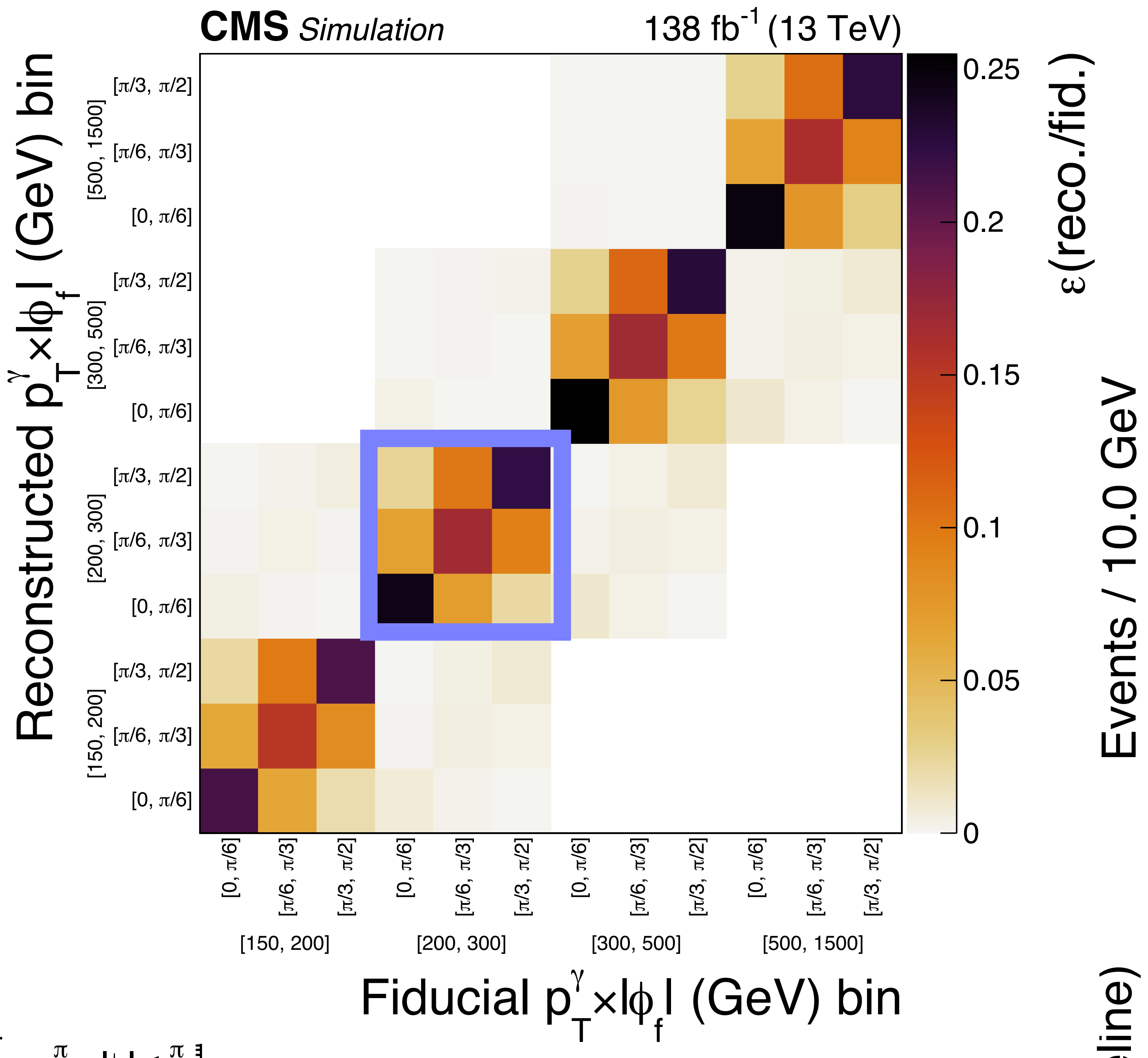


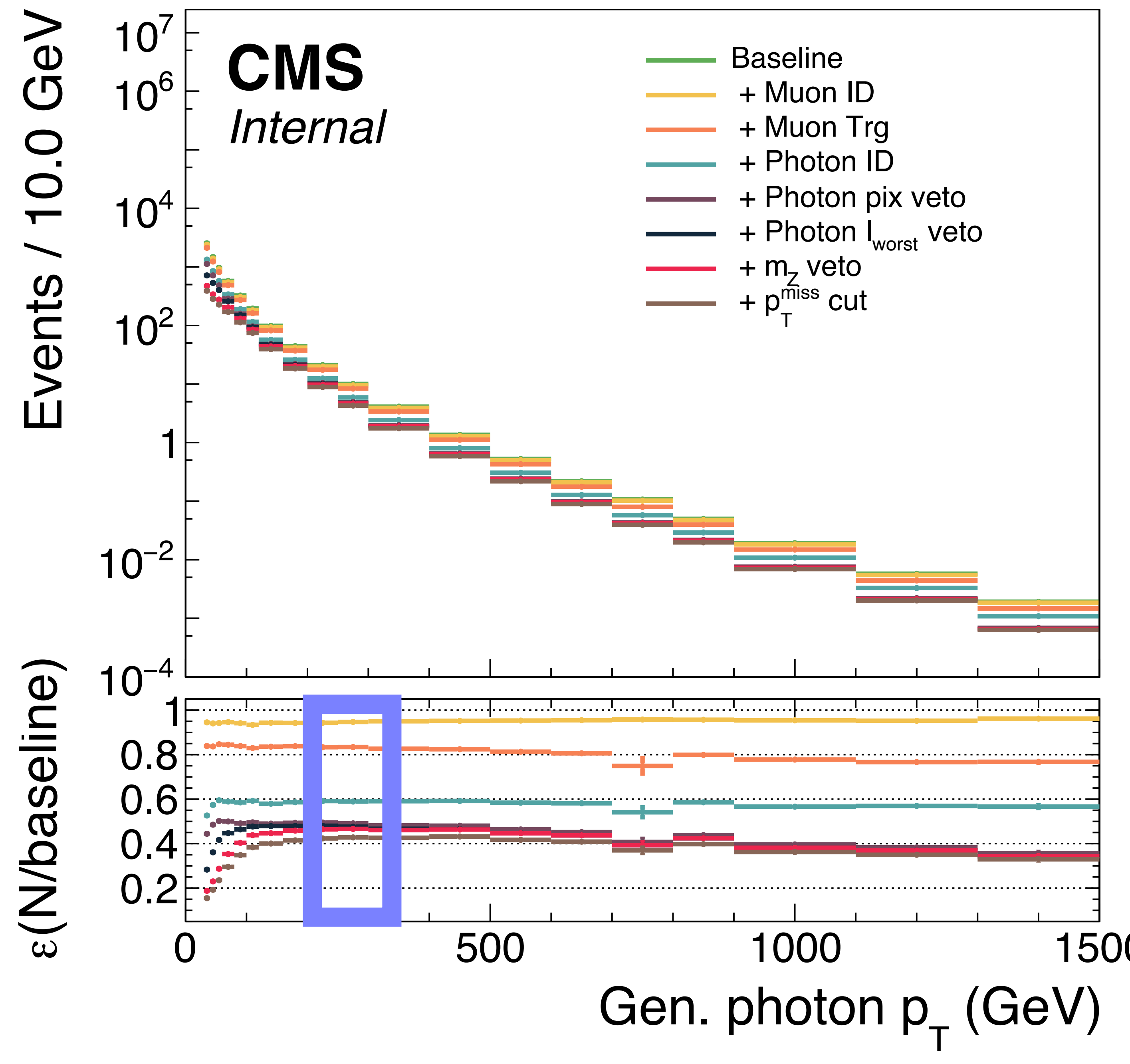


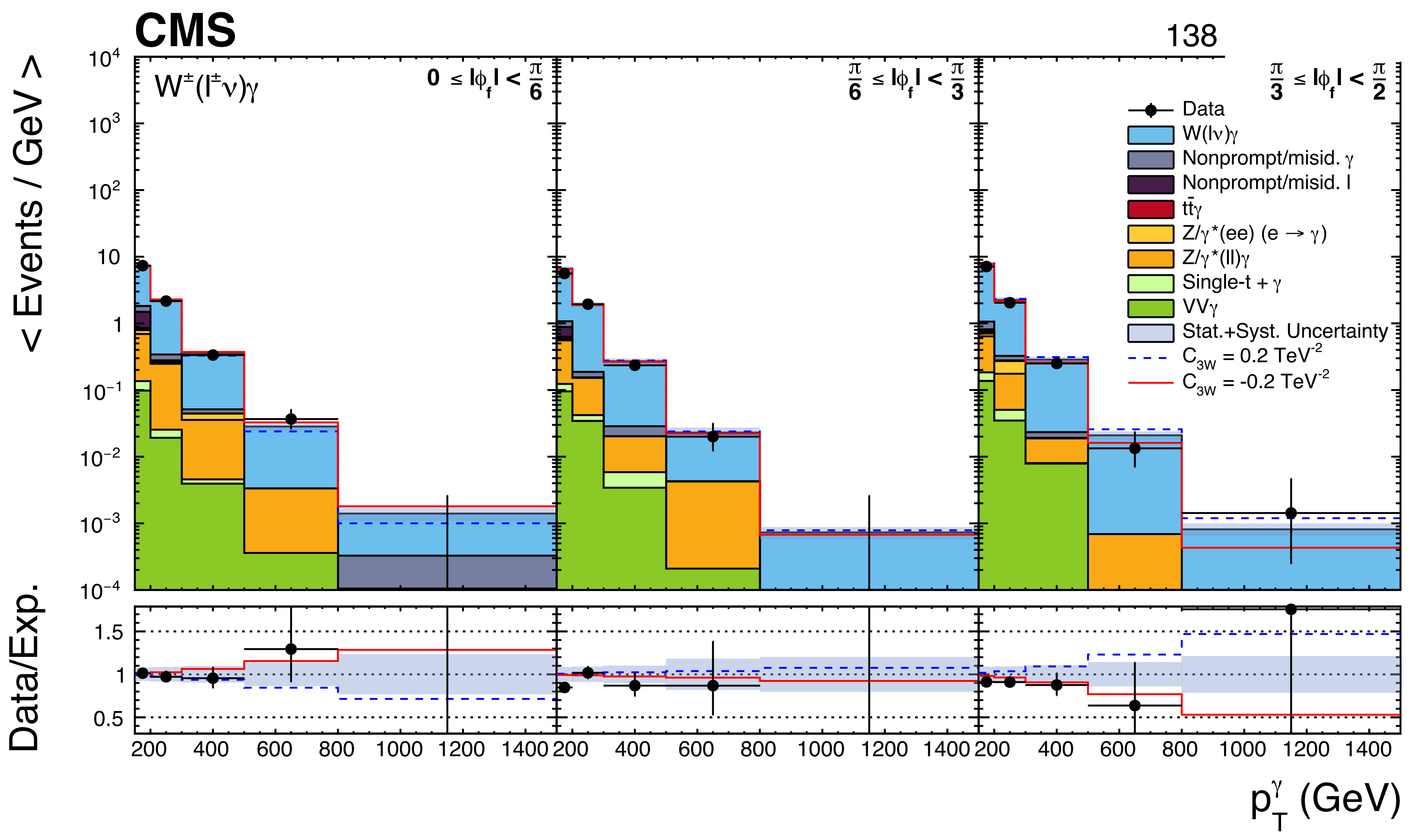
Reasonably strong correlations between $|\varphi_f|$ bins

Within-bin efficiency

- Built-in assumption that EFT does not change response matrix
- To 1st order, can check selection eff. does not change significantly within each bin
- But could be dependence on other variables we integrate over







- Backgrounds assumed to be SM-like, but in principle EFT-sensitive
- We could unfold sum of all contributions to prompt final-state (building in more assumptions about process composition and ε X A)

Reinterpretation & summary

- A big question with reinterpretation is who is doing it, and what information they have!
- As discussed in Kyle's talk, we are making good progress with preserving and releasing statistical models
- If in 30 years we can build a new : $L(\text{data} \mid \sigma_i(c^j))$, no problem with:
  - unfolded binning limitations
  - Gaussian approximations
  - Lack of systematics information for correlations
  - Presence of backgrounds (as long as we split the processes in the same fiducial bins)
- But concerns of efficiency / acceptance dependence remain, as does challenge of building optimal observables
- **Do we conclude unfolded cross sections are no longer a useful too for EFT interpretation?**
  - Clearly not the best we can do, but still provide a common language at the intersection of theory and experiment(s)
    - E.g. allow us to benchmark against fixed-order calculation
  - Consider it a good insurance option ⇒ the underlying machinery is ultimately the same (likelihood model), so the cost of providing these results is not high