Interplay between PDFs and new physics

A systematic study of new physics contaminations in PDF fits

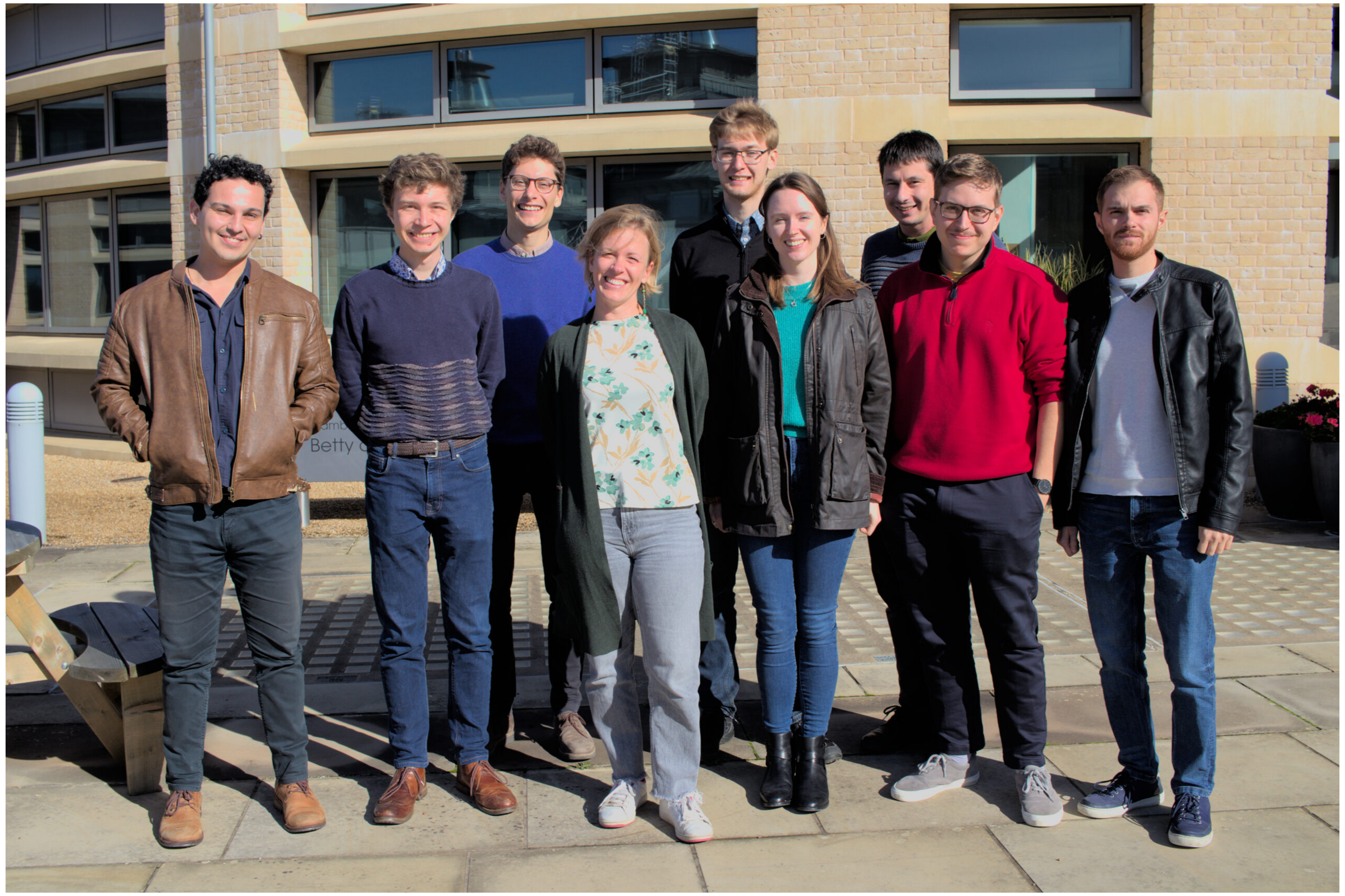
European Research Council
Established by the European Commission

PBSP

Elie Hammou / 34th Rencontres de Blois
Supervisor: Prof Maria Ubiali

Our group: PBSP

Physics Beyond the Standard Proton



- Led by Maria Ubiali
- Based In Cambridge
- Working on interpretation of LHC data
  - Indirect search for heavy new physics
  - Interplay of PDF and EFT

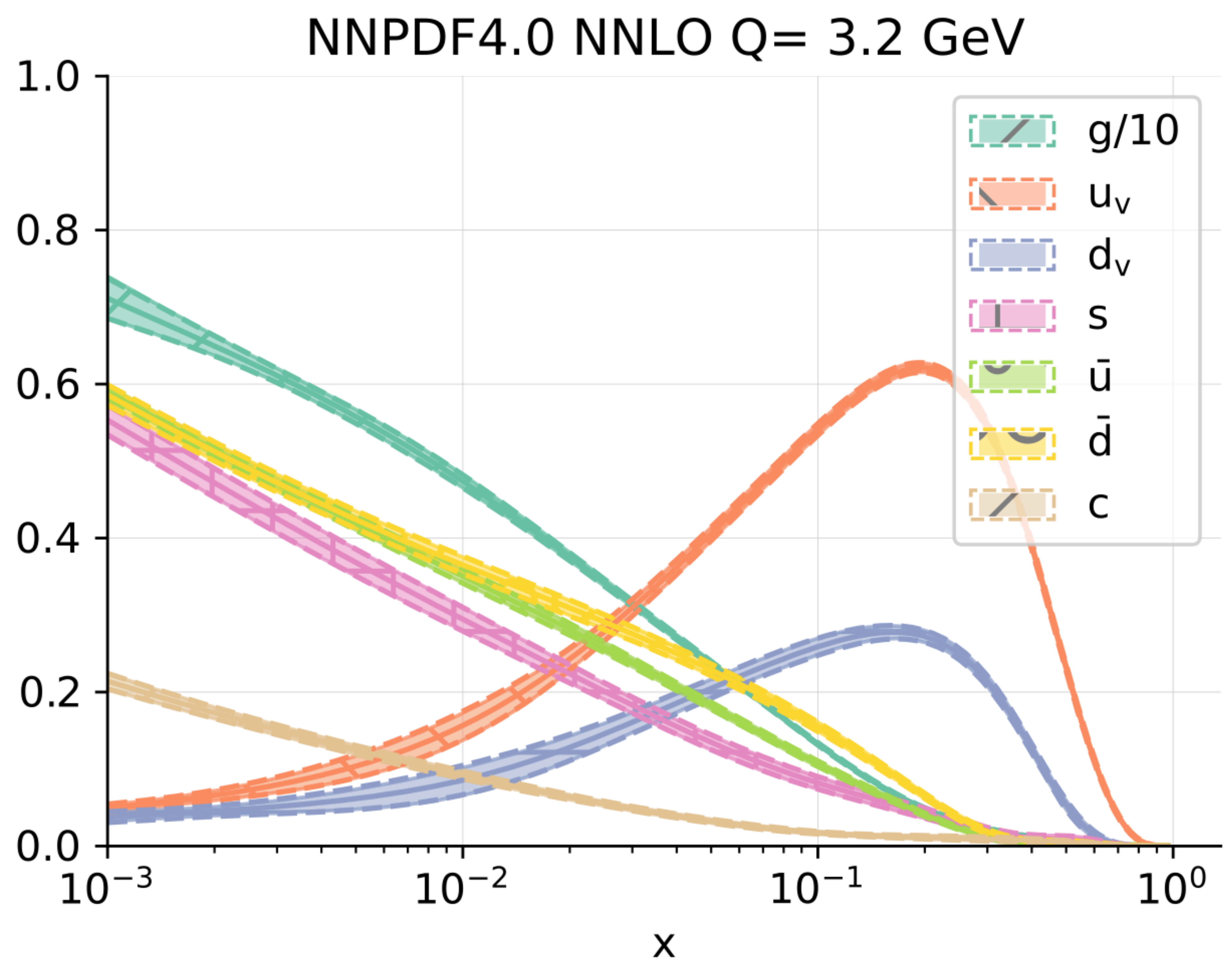
Background on Parton Distribution Functions

$$\sigma = \hat{\sigma} \otimes f$$

- PDFs: describe proton in terms of partonic content
- Very important in hadron colliders
- Non-perturbative QCD

➡ Fit functions from data

- NNPDF methodology: MC replica and NN parametrisation



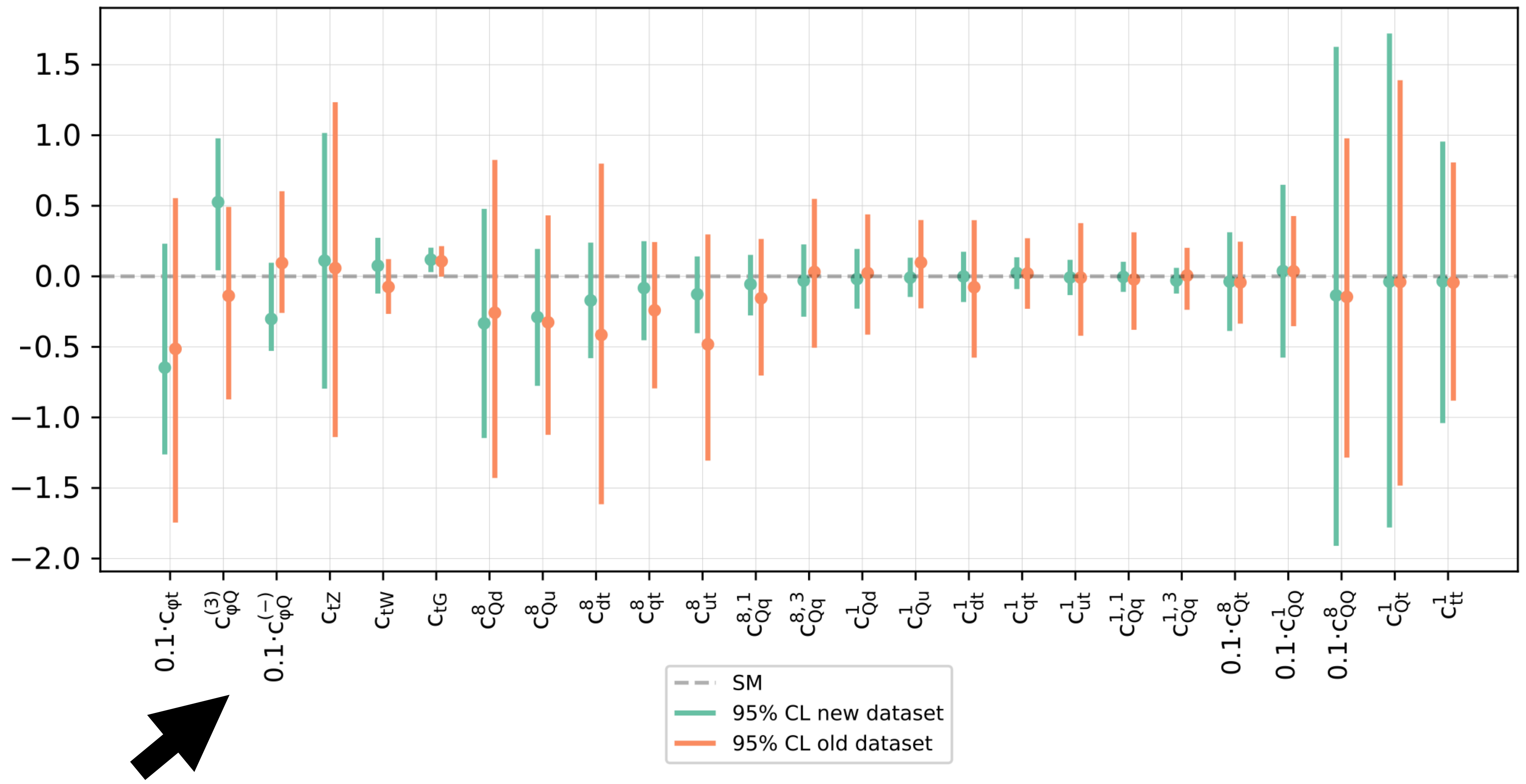
[Ball et al., NNPDF4.0, 2109.02653]

Background on SMEFT

$$\mathscr{L}^{\text{SMEFT}} = \mathscr{L}^{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i + \ldots$$

- Parametrisation of heavy new physics
- Dimension 6 operators with SM fields
- Model-independent
  - ➡ Fit Wilson coefficients from data
- Tools such as SMEFiT, Fitmaker

$\frac{c_i}{\Lambda^2}[TeV^{-2}]$



c_i

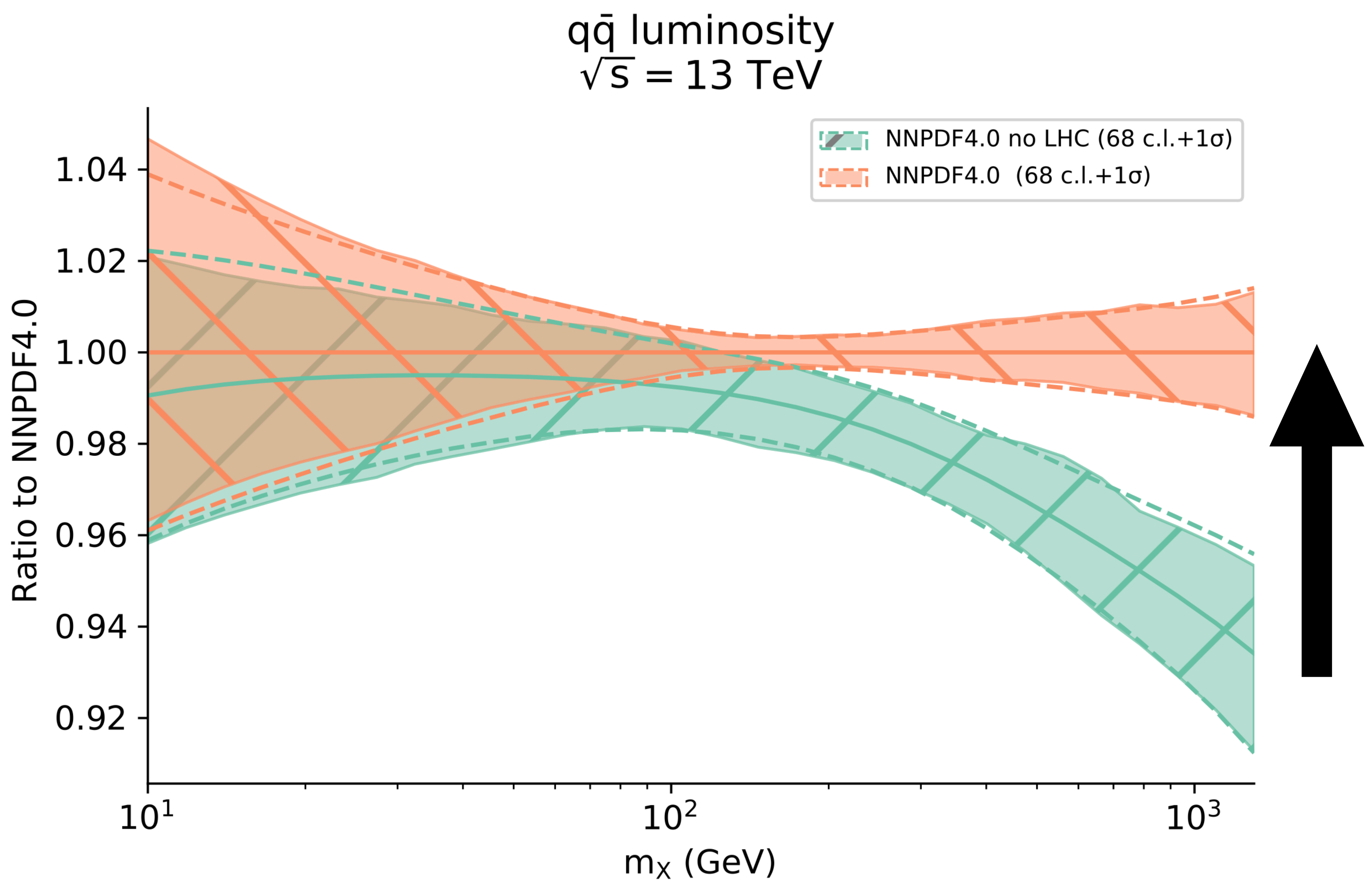
[PBSP, 2303.06159]

Problem: Can we mix them up?

Do we risk absorbing new physics signals in PDF fitting?

Motivation for concern:

- Both are fitted from data
- PDF parametrisation is very flexible
- LHC data shifts PDFs

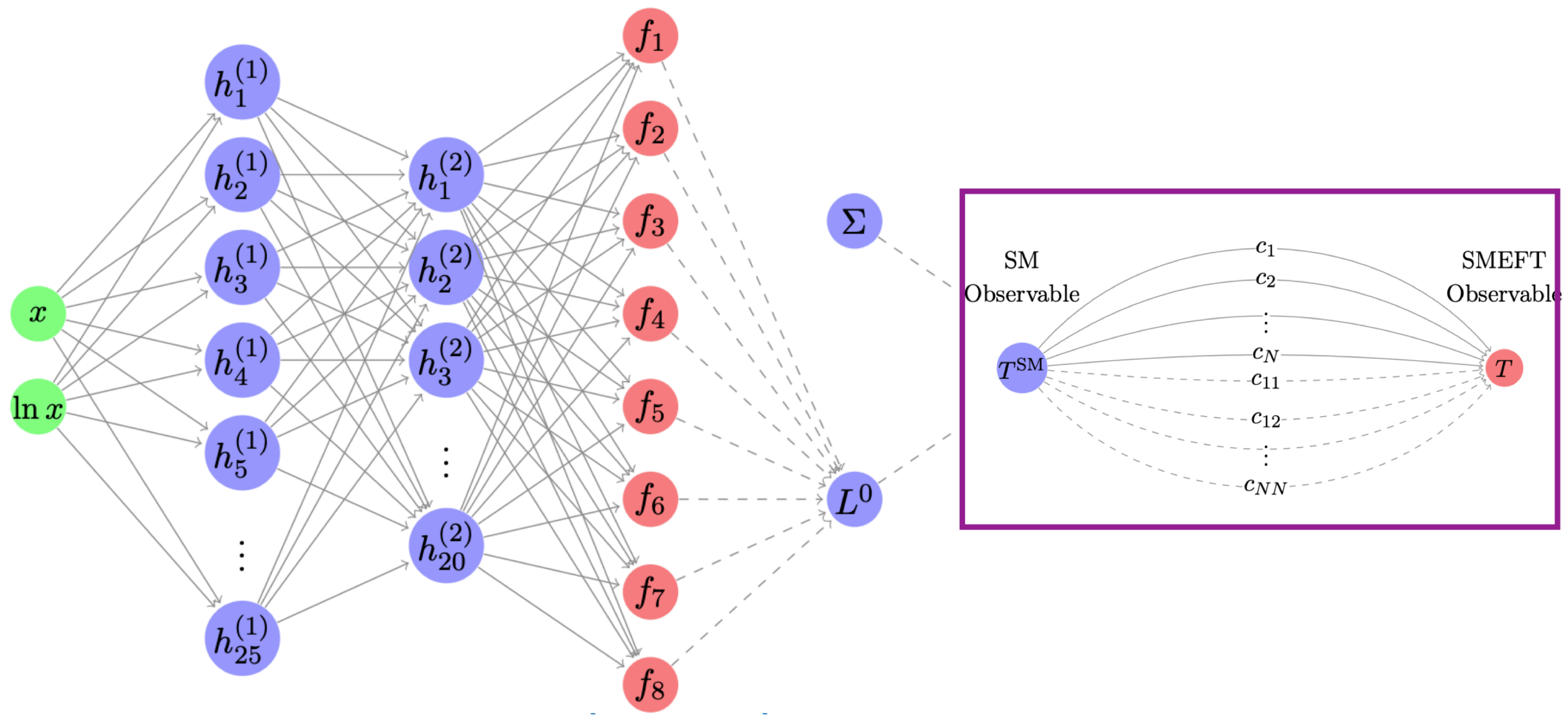


[Ball et al., NNPDF4.0, 2109.02653]

Don't mix apples and oranges

Need robust framework to disentangle EFT and PDF signals

- Simultaneous fits:
  - SIMUnet, [*The top quark legacy of the LHC Run II for PDF and SMEFT analyses, 2303.06159*]
- Conservative dataset:
  - Prevent contamination



PDFs **Wilson coefficients (SMEFT)**

Focus of the talk: Risk assessment

Do we risk absorbing new physics in PDF fitting?

Perform a "Contamination test":

- Produce pseudodata using SM PDFs and NP
- Fit PDFs from pseudodata assuming SM

Can we get "contaminated PDFs"?

New physics scenarios: Z'

$$M_{Z'} = 18.7 \text{ TeV}$$

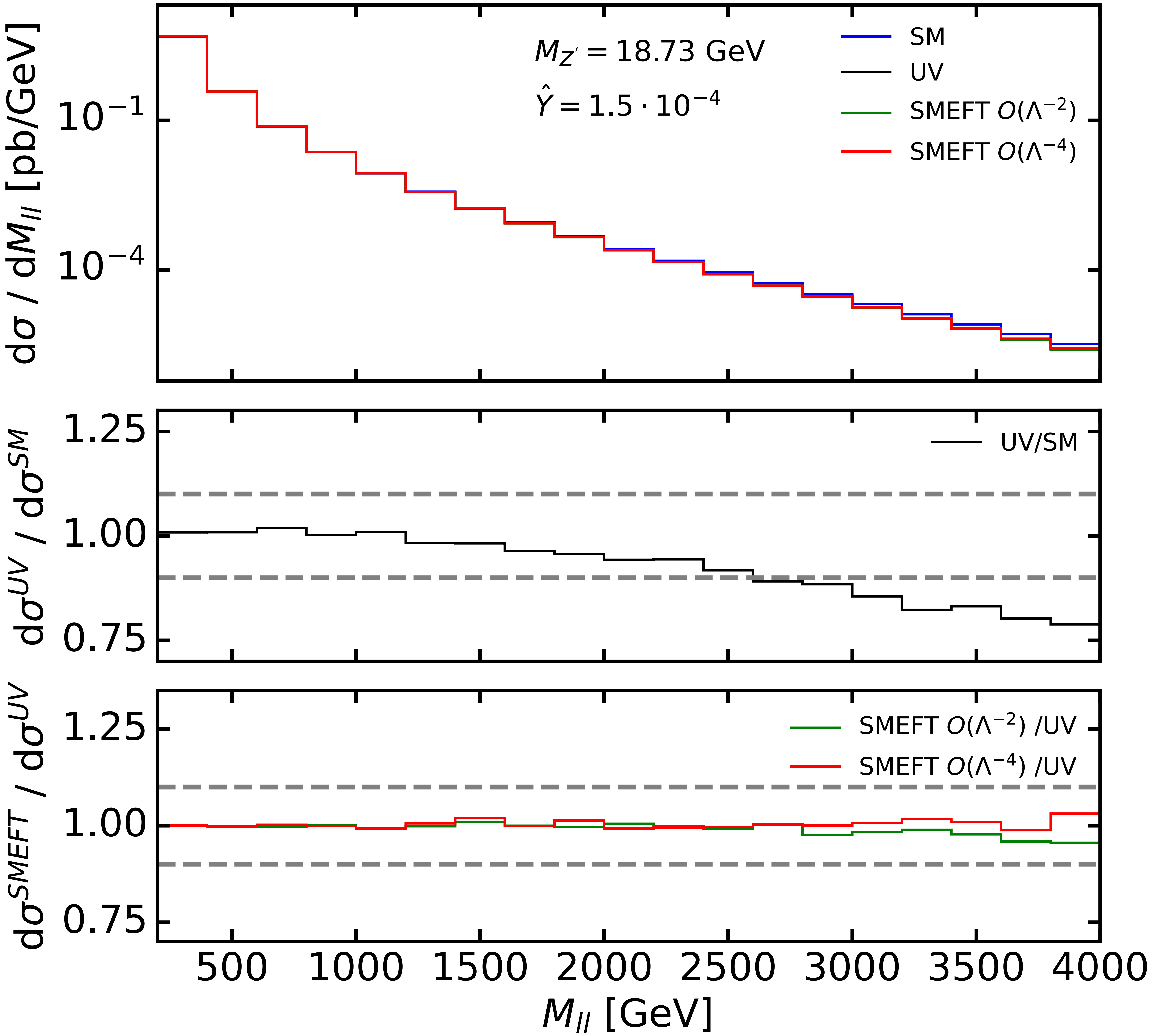
Generation of the pseudodata

$$\mathscr{L}^{Z'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2_{Z'}}{2M^2_{Z'}} J^\mu_Y J_{Y,\mu}$$

$$J^\mu_Y = \sum_f Y_f \bar{f} \gamma^\mu f$$

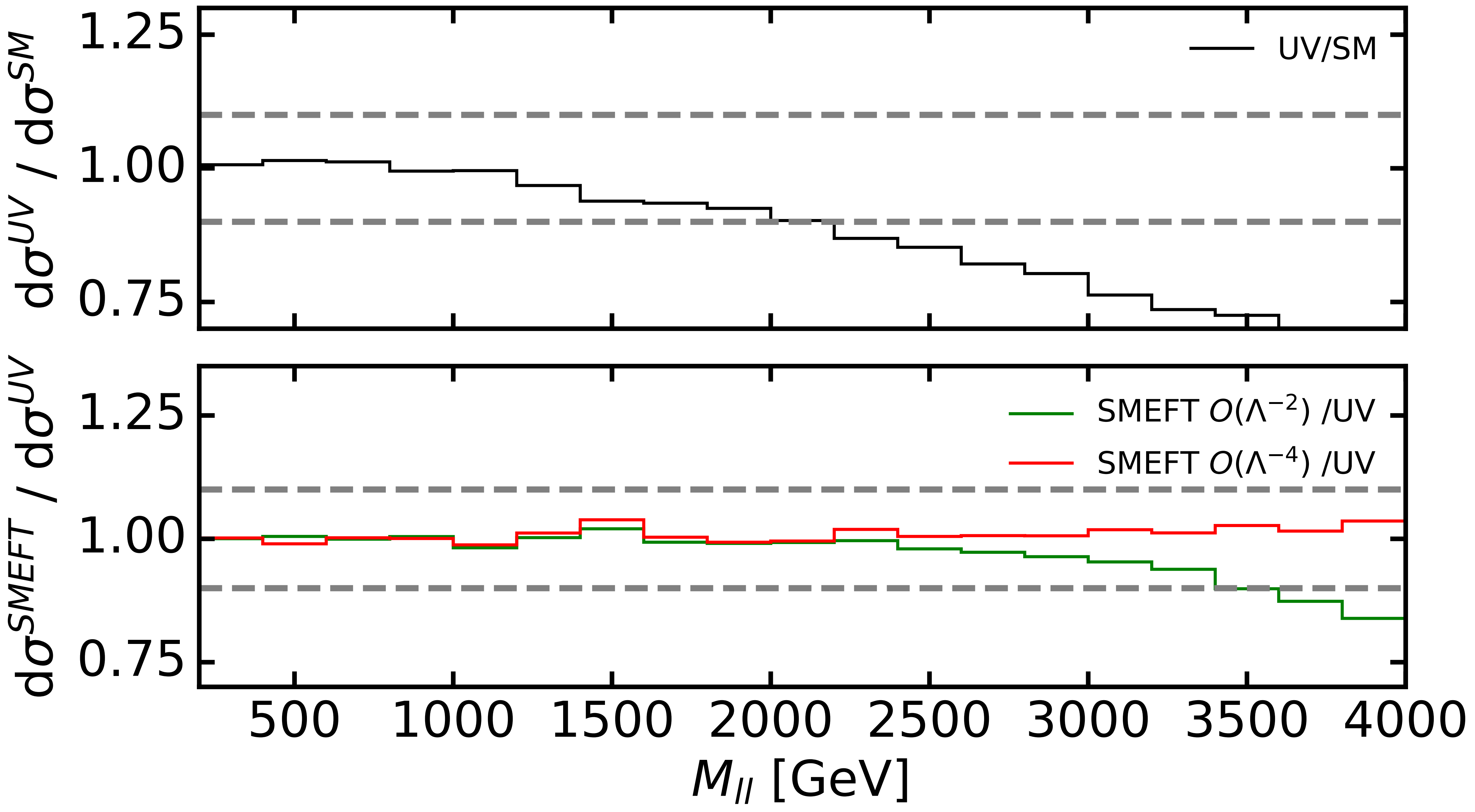
Impacts neutral current Drell-Yan processes

$$pp \to l^+ l^-$$

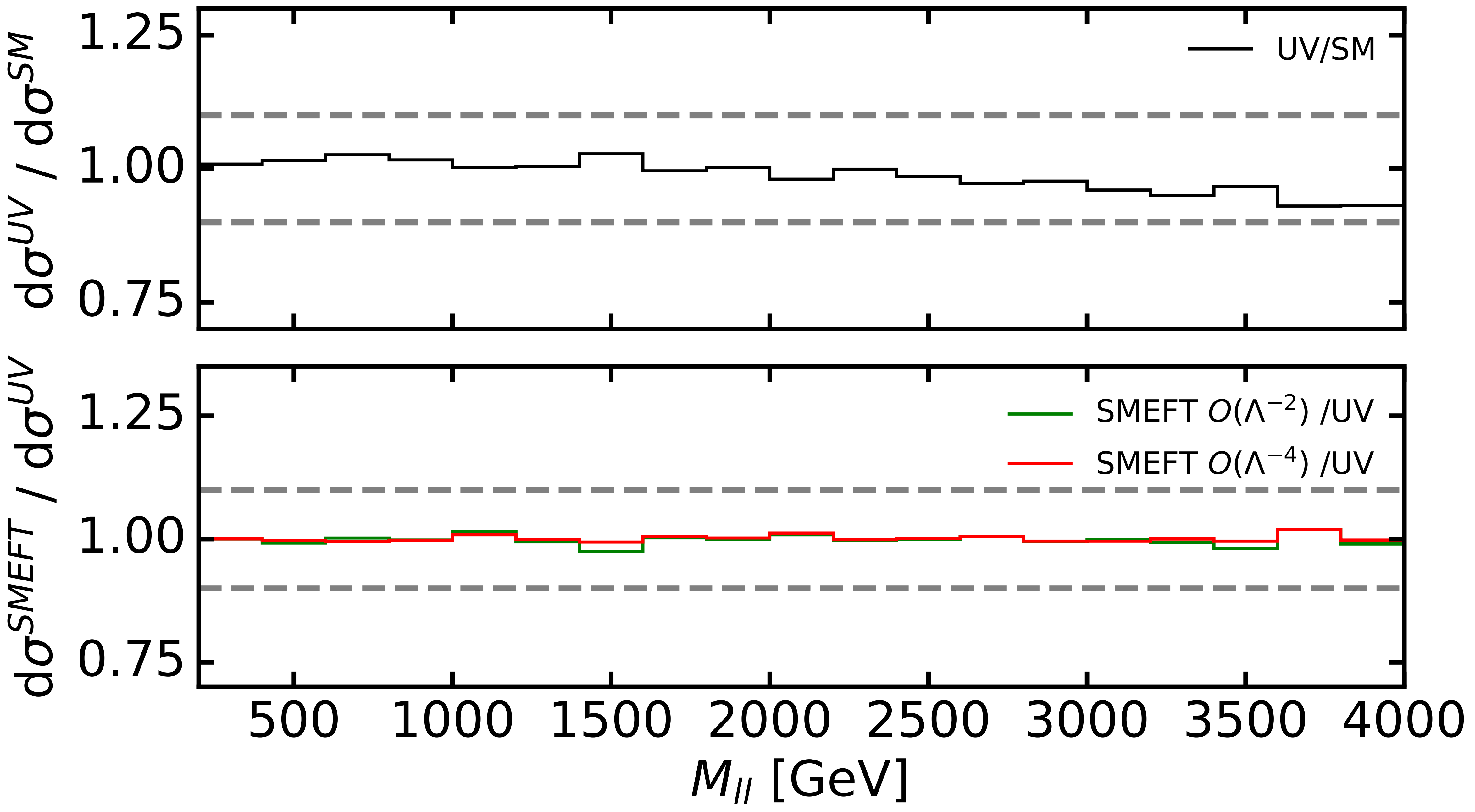


New physics scenarios: Z'

$M_{Z'} = 14.5$ TeV



$M_{Z'} = 32.5$ TeV



New physics scenarios: W'

$M_{W'} = 13.8$ TeV

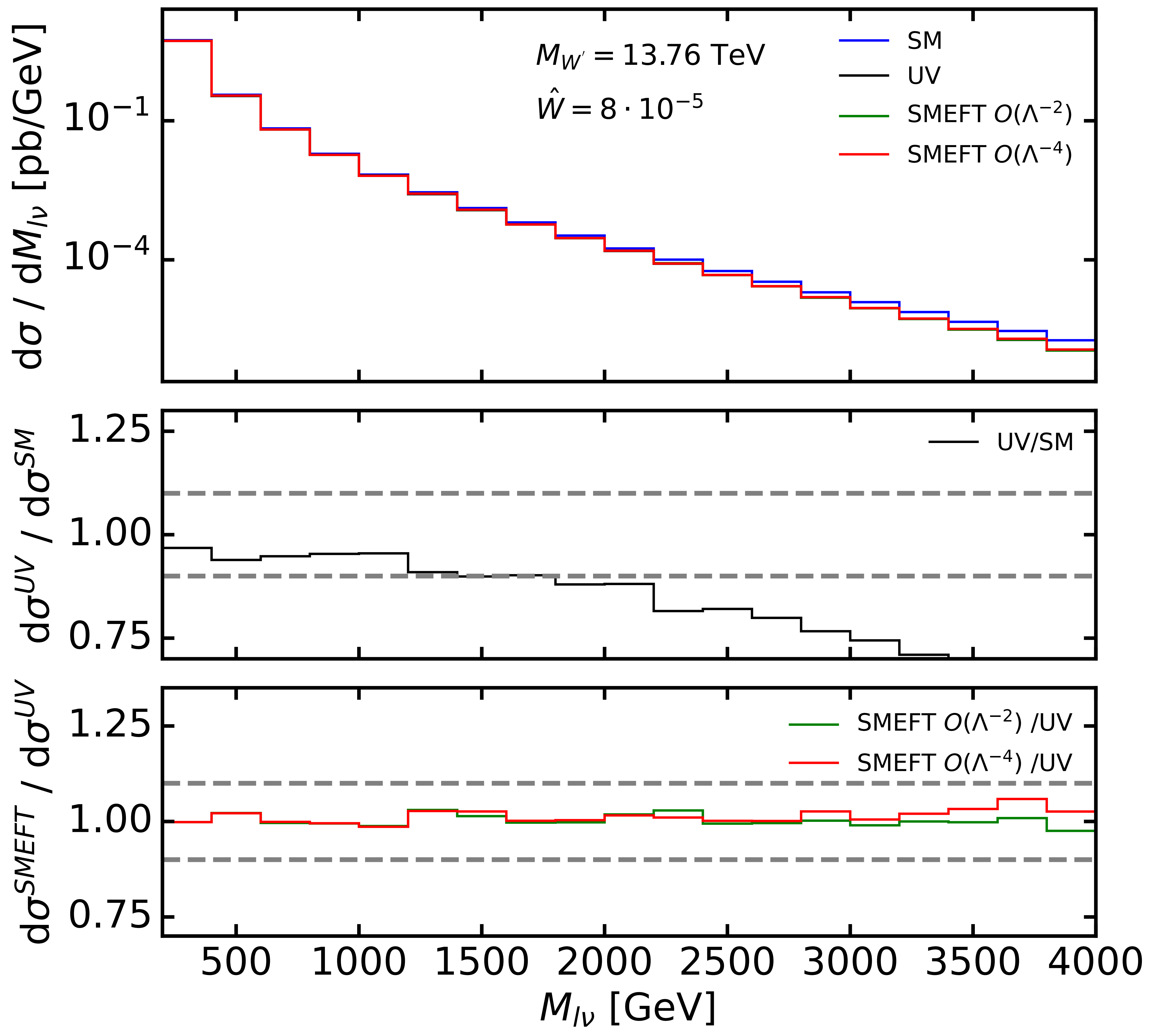
Generation of the pseudodata

$$\mathscr{L}^{W'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2_{W'}}{2M^2_{W'}} J^{a,\mu}_L J^a_{L,\mu}$$

$$J^{a,\mu}_L = \sum_{f_L} \bar{f}_L T^a \gamma^\mu f_L$$

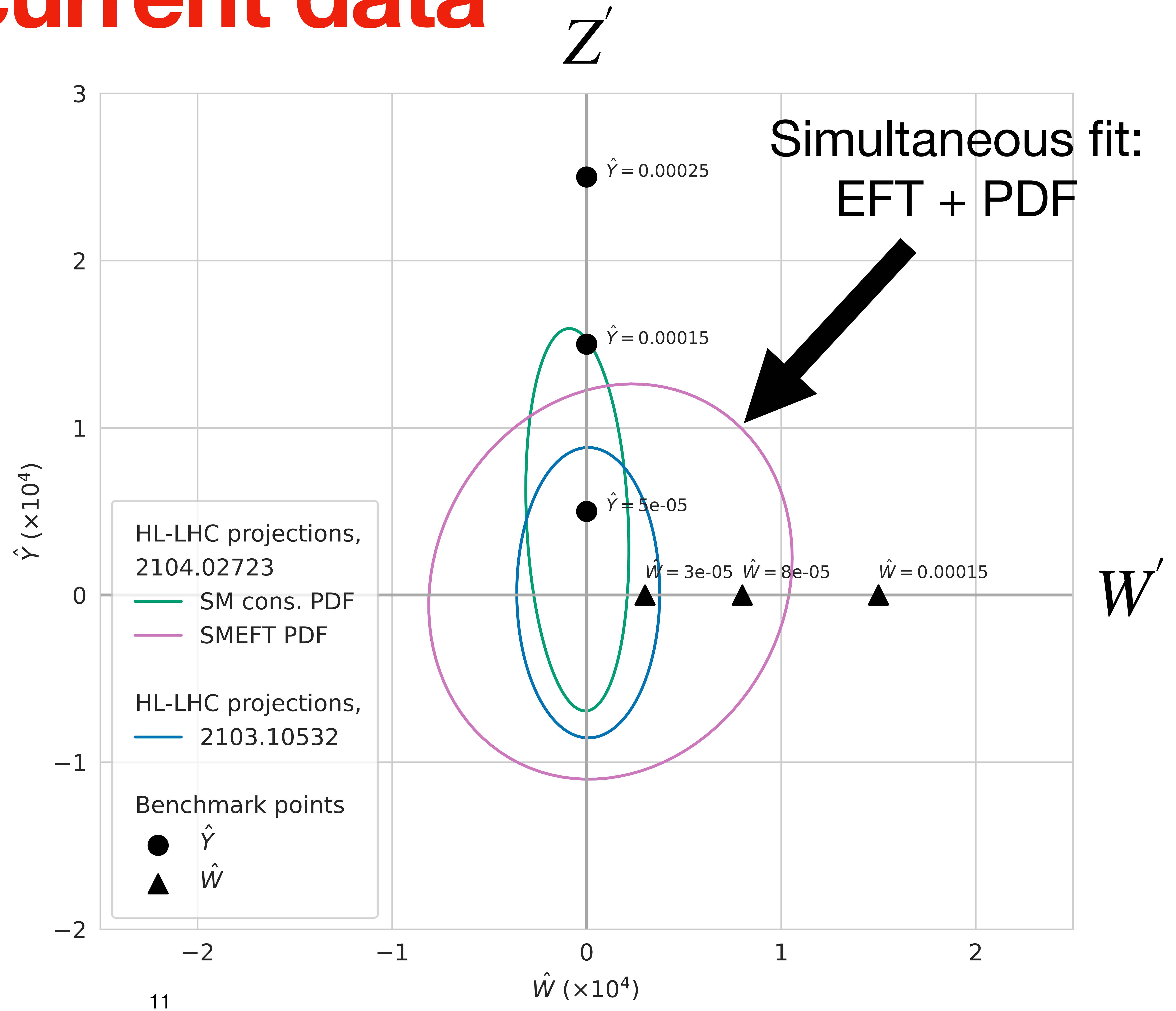
Impacts charged current Drell-Yan processes

$$pp \to l^- \bar{\nu}$$



Constraints from current data

- New physics scenarios compared to constraints at 95% CL



PDF fitting: selection test

Do our contaminated datasets pass the selection criteria?

Z'

Selection test: ✗

➡ Excluded from PDF fit

No impact on PDFs

W'

Selection test: ✓

➡ Included in PDF fit

PDFs contaminated

Impact of contamination: PDFs

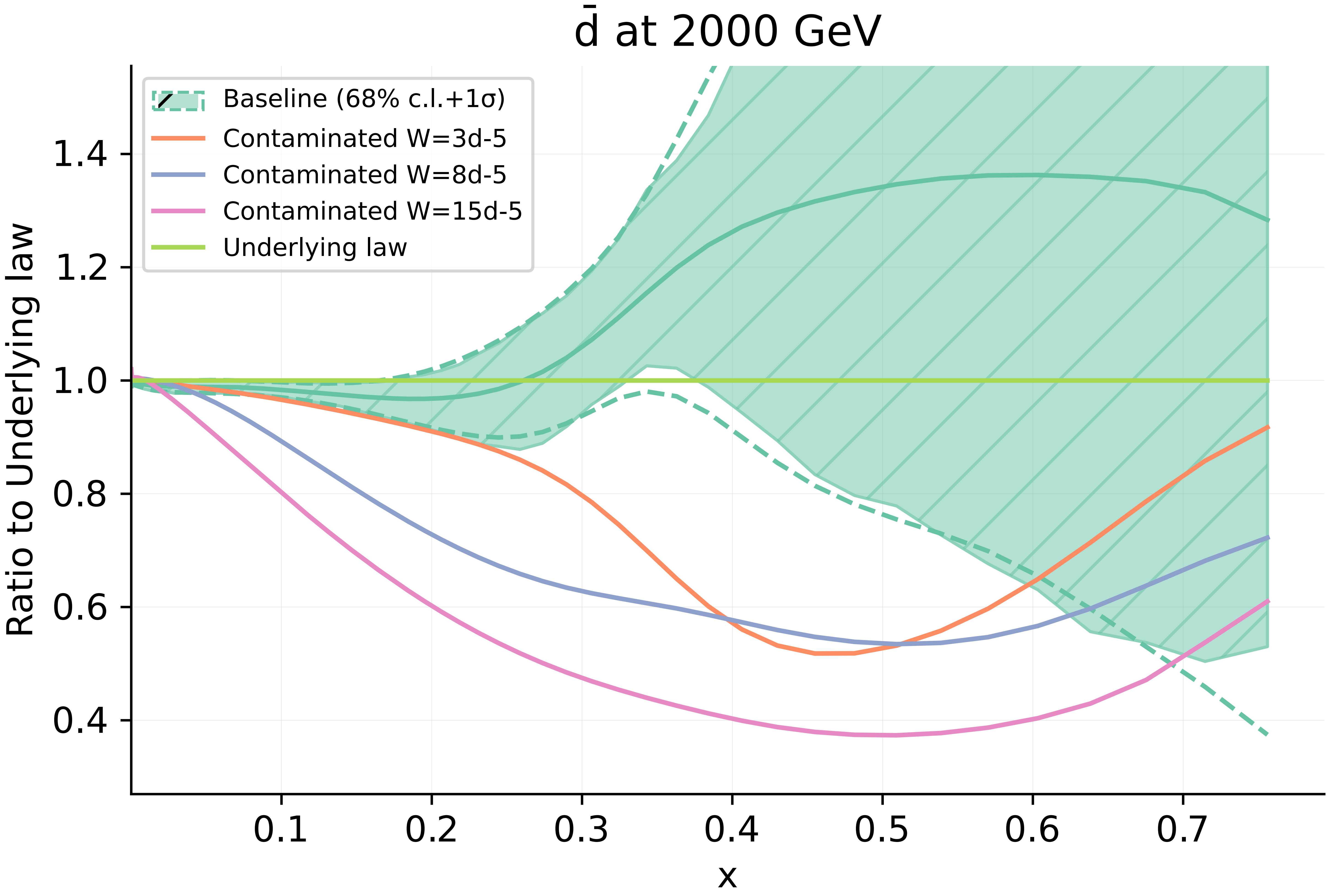
Comparison between contaminated and Baseline PDFs

- Contaminated

  ➡ BSM Lagrangian

- Baseline

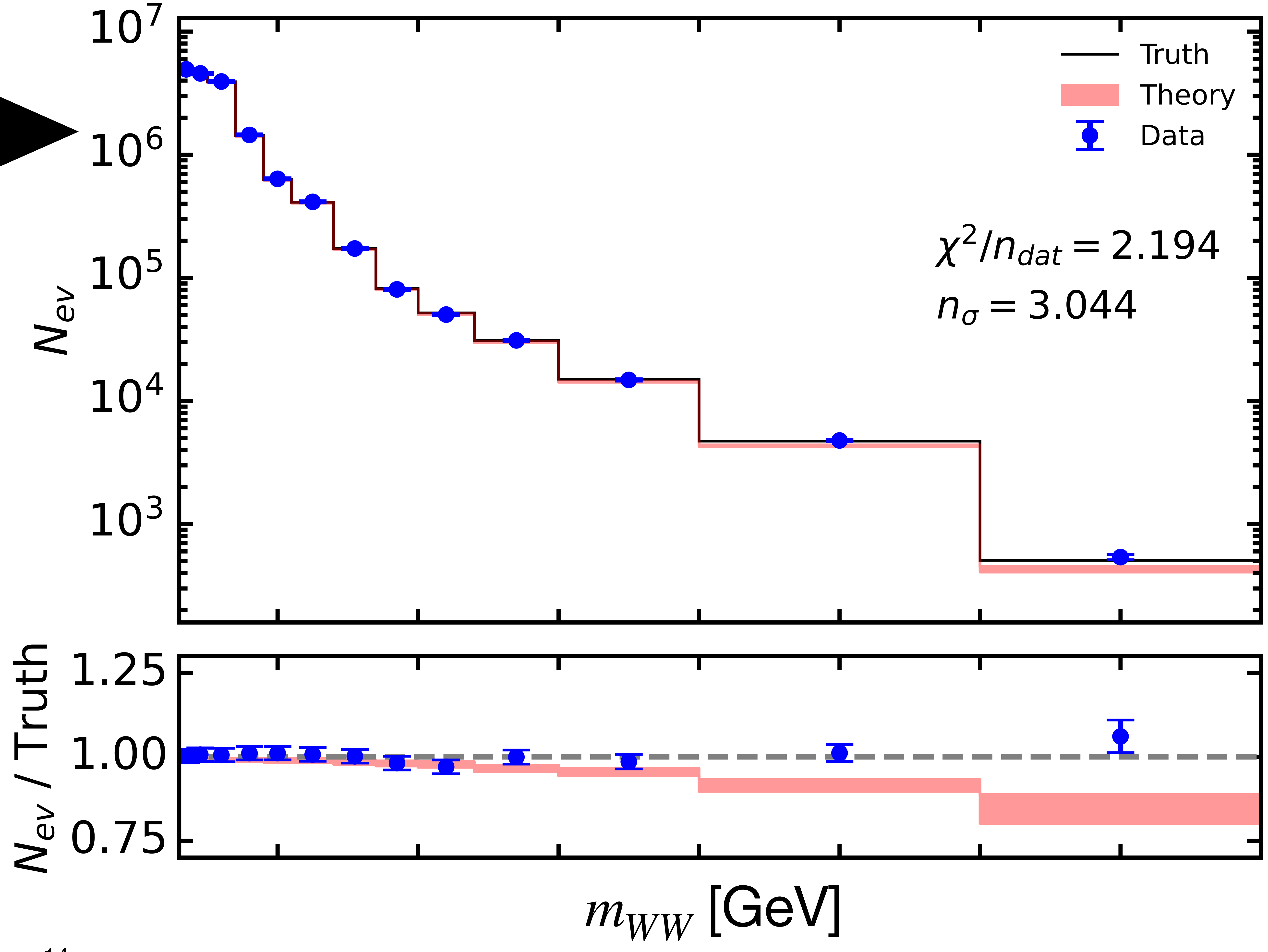
  ➡ SM Lagrangian



Impact of contamination: LHC predictions

Analysis of contaminated predictions for HL-LHC data

$$pp \to W^+W^-$$



- WW production

- Comparison between:
  - Contaminated PDFs (red)
  - Baseline PDFs (black)

What does it mean?

- Contamination effect
  - ➡ Miss new physics (W’ field)
  - ➡ Introduce fake deviations in other sectors
- Need way to identify contamination
  - ➡ Test on observable not included in PDFs fit
- Need way to prevent contamination
  - ➡ Additional selection criteria?

Summary and outlook

- Discussed two new physics scenarios: $Z^{'}$ and $W^{'}$. Both impact high-energy Drell-Yan

- Signs of $W^{'}$ got fitted away in PDF parametrisation

➡ Missed new physics

➡ Introduced deviations where they are not present

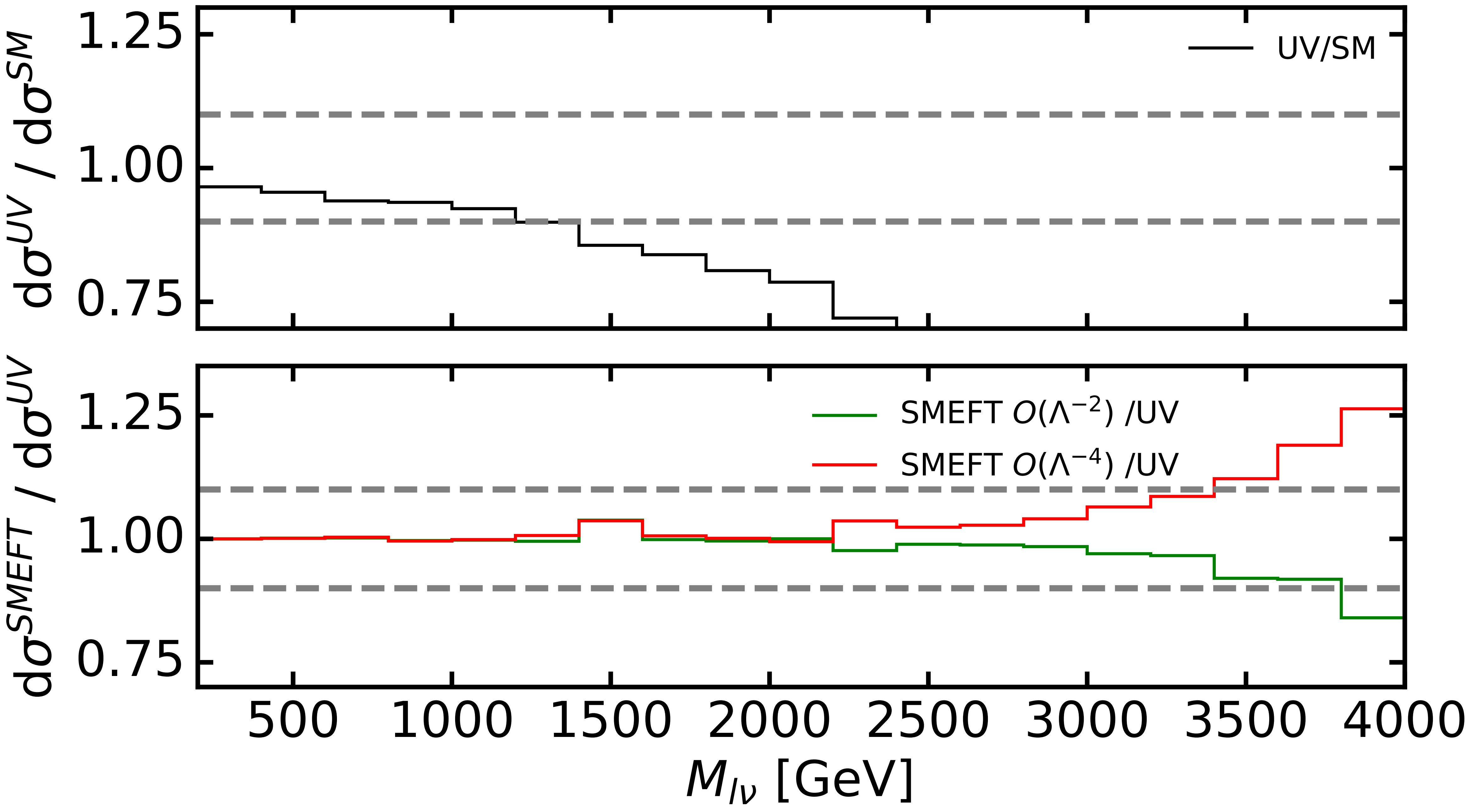
- Need a robust disentangling method for a precision study
  - Identify and prevent contamination

Thank you for your attention!

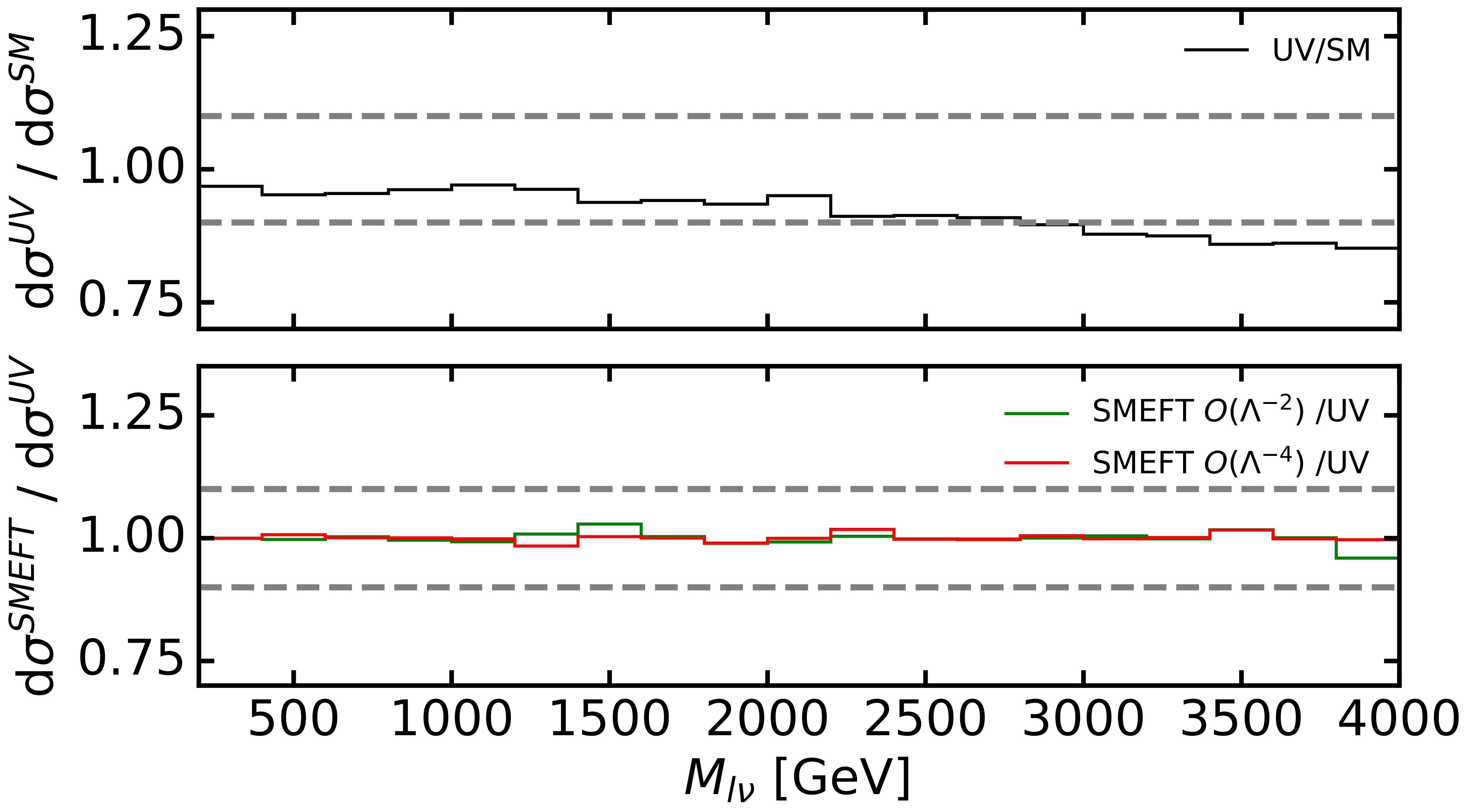
Extra slides

New physics scenarios: W'

$M_{W'} = 10$ TeV



$M_{W'} = 22.5$ TeV



PDF fitting: selection criteria

Exclusion of incompatible datasets (NNPDF criteria)

Two criteria:

- χ^2-statistics:

$$\chi^2 = (data - theory)^T \cdot V_{cov}^{-1} \cdot (data - theory)$$

  - $\dfrac{\chi^2}{n_{dat}} > 1.5$ → excluded

- n_σ standard deviation:

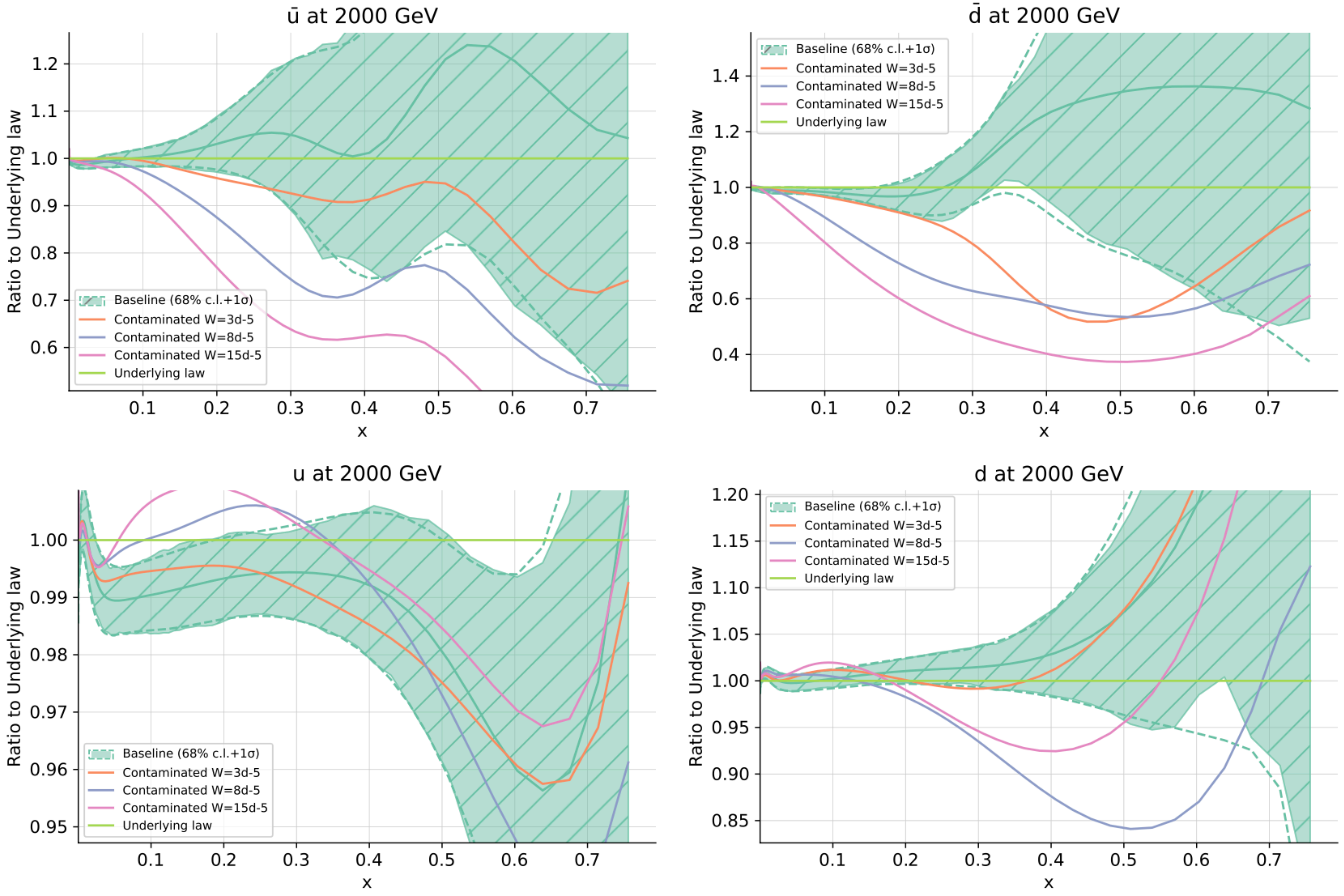
  - $n_\sigma > 2$ → excluded

$$n_\sigma = \frac{\chi^2 - 1}{\sigma_{\chi^2}}$$

List of deviations

	HL-LHC		Stat. improved	
Dataset	$\chi^2/n_{\rm dat}$	n_σ	$\chi^2/n_{\rm dat}$	n_σ
W^+H	1.17	0.41	1.77	1.97
W^-H	1.08	0.19	1.08	0.19
W^+Z	1.08	0.19	1.49	1.20
W^-Z	0.99	-0.03	1.02	0.05
ZH	1.19	0.44	1.67	1.58
W^+W^-	2.19	3.04	2.69	4.31
VBF $\to$ H	0.70	-0.74	0.62	-0.90

Quarks PDF



ū at 2000 GeV
d̄ at 2000 GeV
u at 2000 GeV
d at 2000 GeV
Ratio to Underlying law
x
Baseline (68% c.l.+1σ)
Contaminated W=3d-5
Contaminated W=8d-5
Contaminated W=15d-5
Underlying law