Hide and Seek: How PDFs can conceal New Physics

A systematic study of new physics contaminations in PDF fits

European Research Council

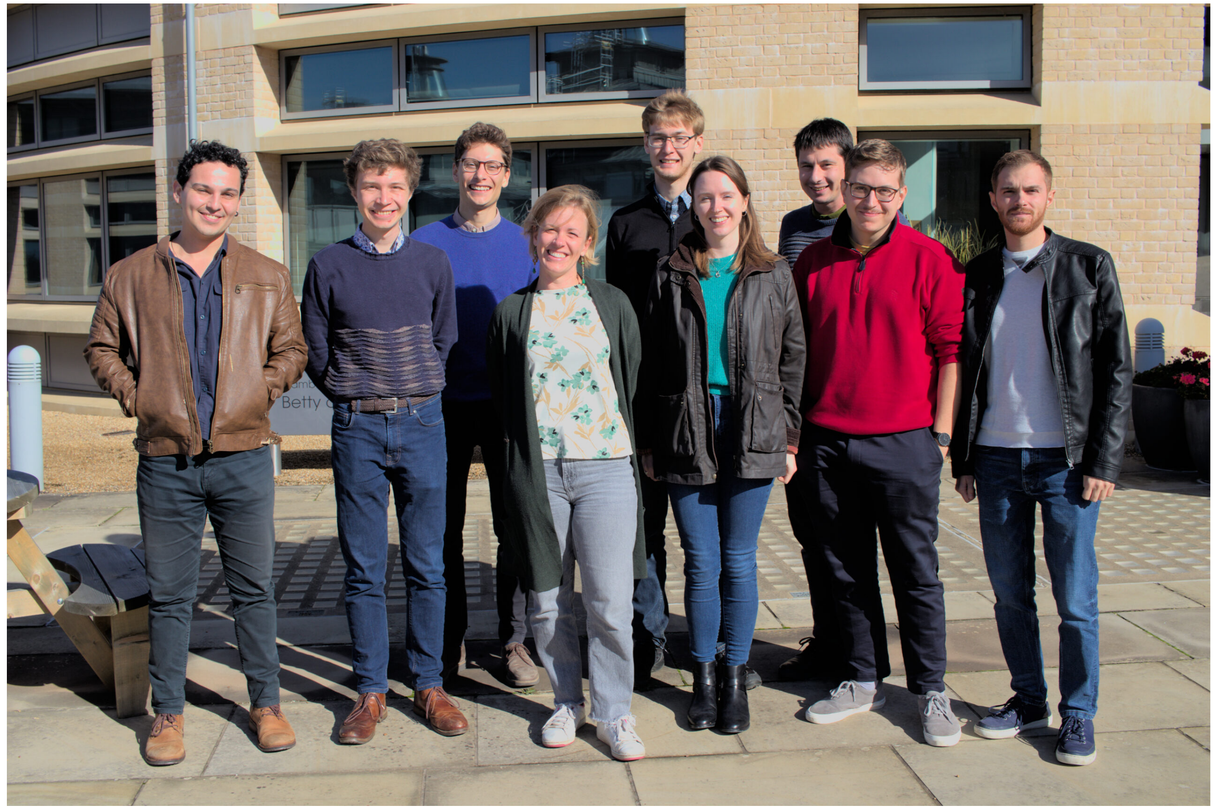
Established by the European Commission

PBSP

Elie Hammou / Brookhaven Forum 2023
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

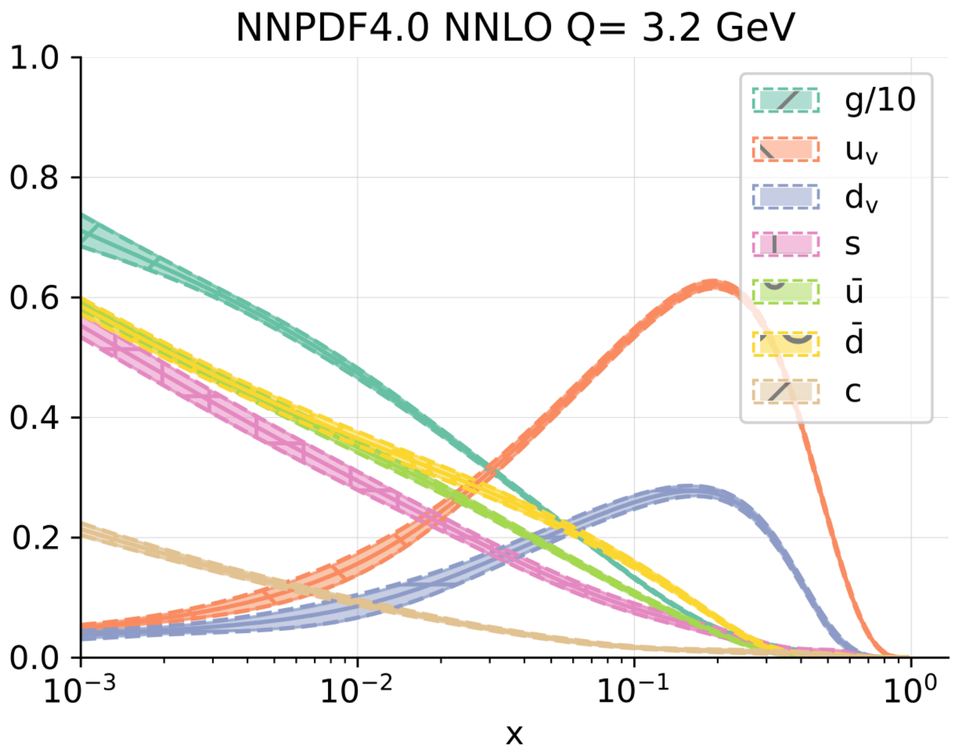
Background on Parton Distribution Functions

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

- PDFs: describe proton in terms of partonic content
- Indispensable at hadron colliders
- Non-perturbative QCD

  ➡ Fitted from data

- NNPDF methodology



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

Heavy New Physics: UV vs SMEFT

- Integrating UV heavy fields out:

  ➡ Dim 6 EFT operators with SM fields

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

- Advantages to choose BSM parameters:

  ➡Predictions polynomial in Wilson coefficients c_i

- Model-independent:

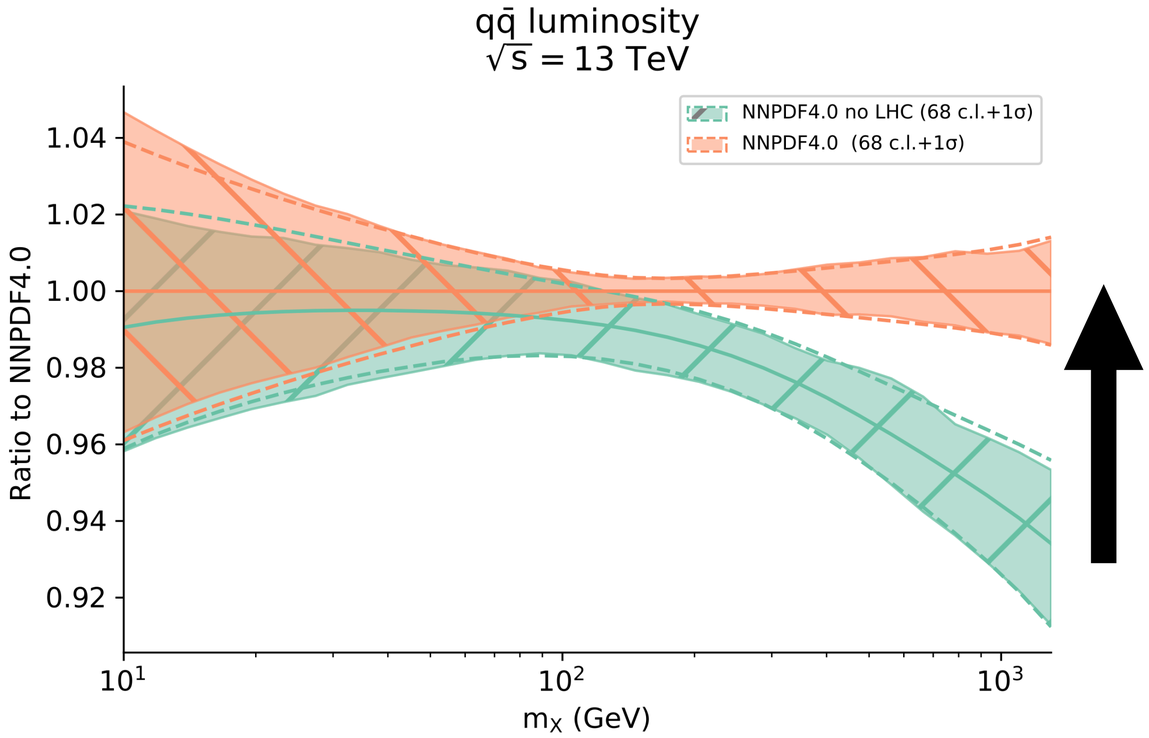
  ➡ Fit Wilson coefficients from data

Problem: Can New Physics contaminate PDFs?

Do we risk absorbing new physics signals in PDF fitting?

Motivation for concern:

- Neither is predicted by theory
- PDF parametrisation is very flexible
- LHC data shifts PDFs
- **Risk of missing new physics**



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

Focus of the talk: Risk assessment

Methodology

Perform a “Contamination test”:

[Hammou, Kassabov, Madigan, Mangano, Mantani, Moore, Morales Alvarado and Ubiali, 2307.10370]

1. Produce pseudodata with BSM physics
2. Fit PDFs from pseudodata assuming SM
3. Compare results with baseline PDFs

If results incompatible with baseline:

➡ **PDFs are contaminated by new physics**

New physics scenarios: Z'

Generation of the pseudodata

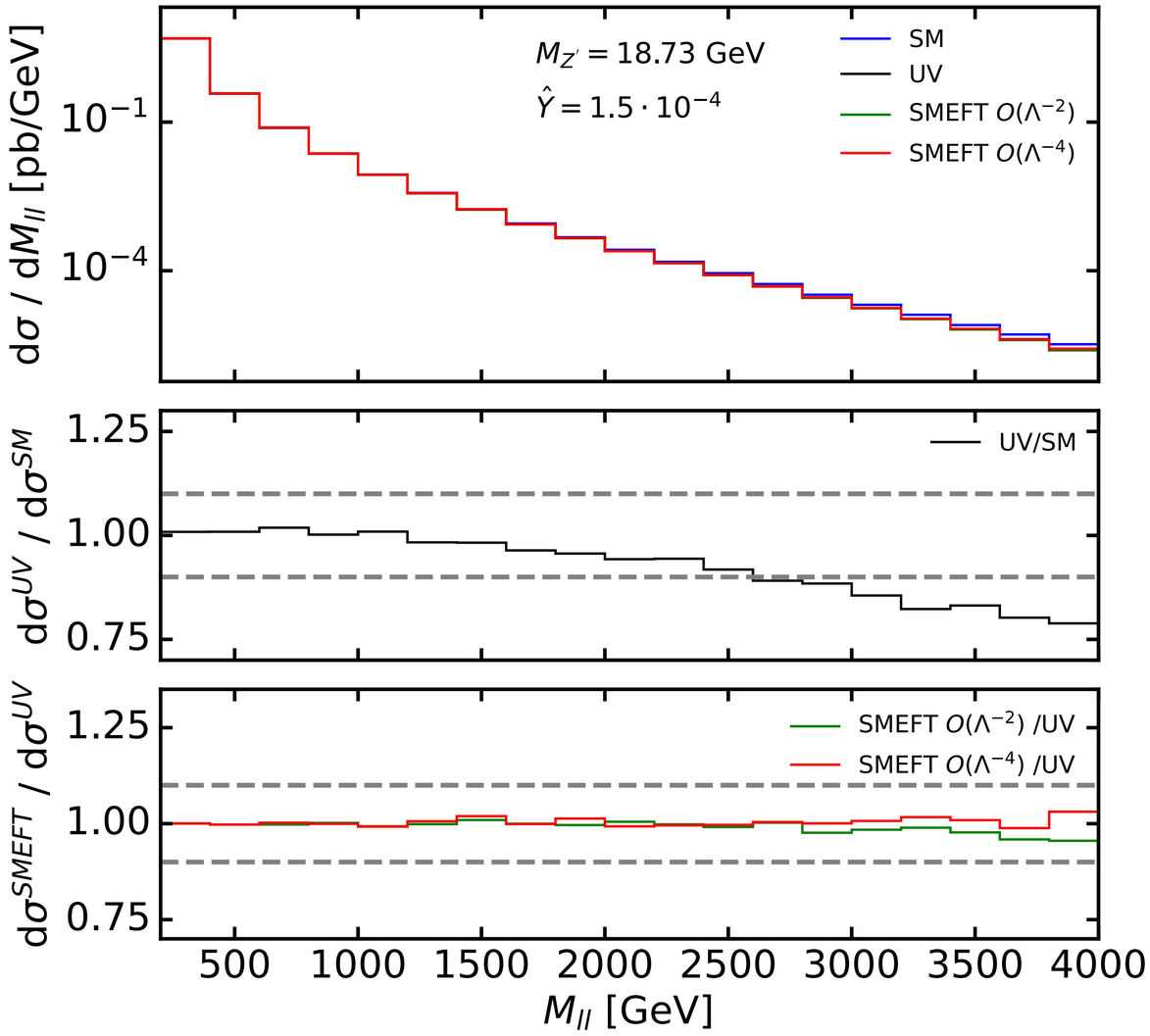
$$\mathscr{L}^{Z'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2_{Z'}}{2M^2_{Z'}} J^{\mu}_{Y} J_{Y,\mu}$$

$$J^{\mu}_{Y} = \sum_{f} Y_f \bar{f} \gamma^{\mu} f$$

Impacts neutral current Drell-Yan processes

$$p\bar{p} \to l^+ l^-$$

$M_{Z'} = 18.7$ TeV



New physics scenarios: W'

$$M_{W'} = 13.8 \text{ TeV}$$

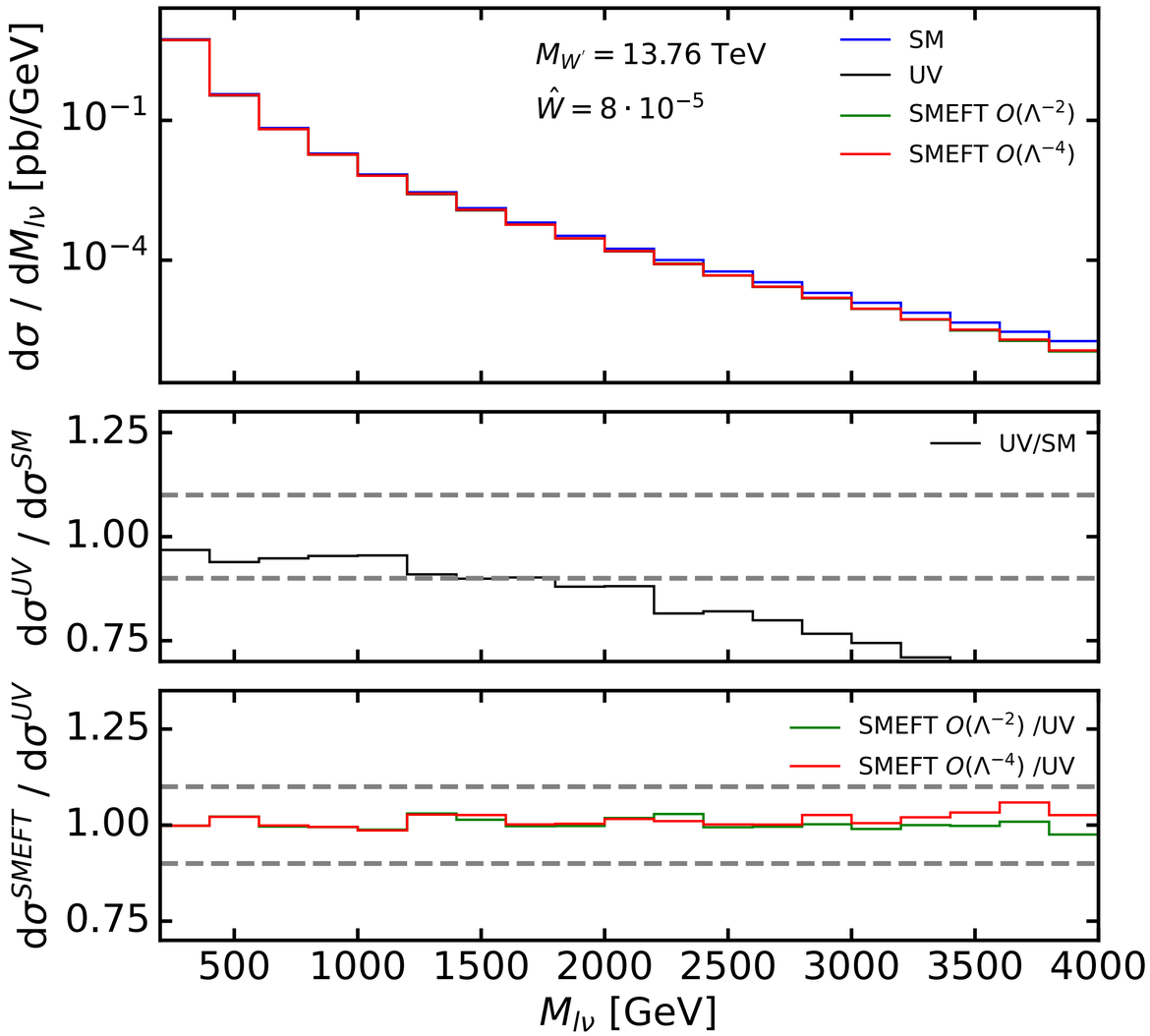
Generation of the pseudodata

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

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

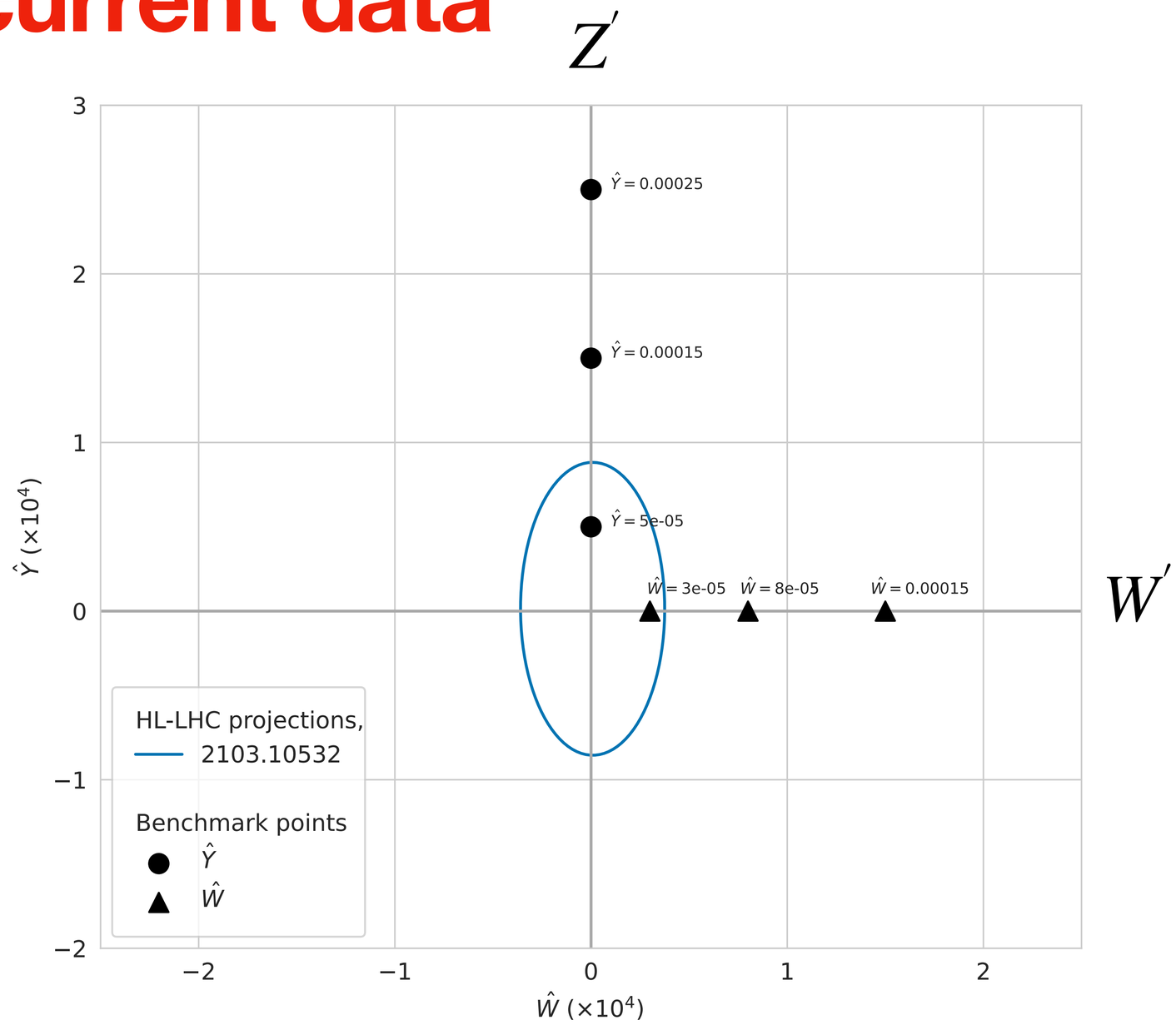
Impacts charged current Drell-Yan processes

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



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: missing new physics

Comparison between contaminated and Baseline PDFs

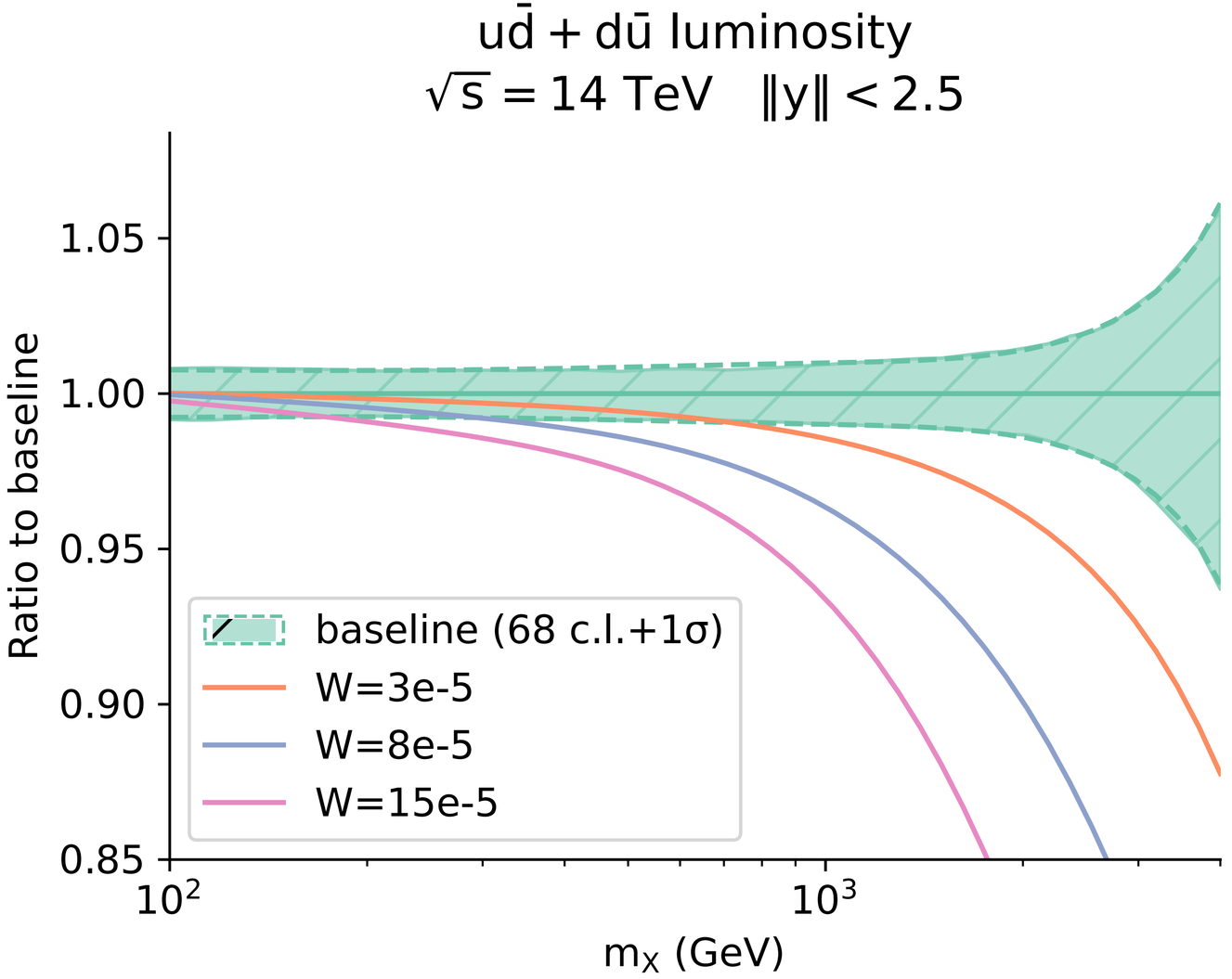
Comparing:

- “Contaminated PDFs” : BSM data
- Baseline PDFs: SM data

Incompatible…

➡ Contamination occurred

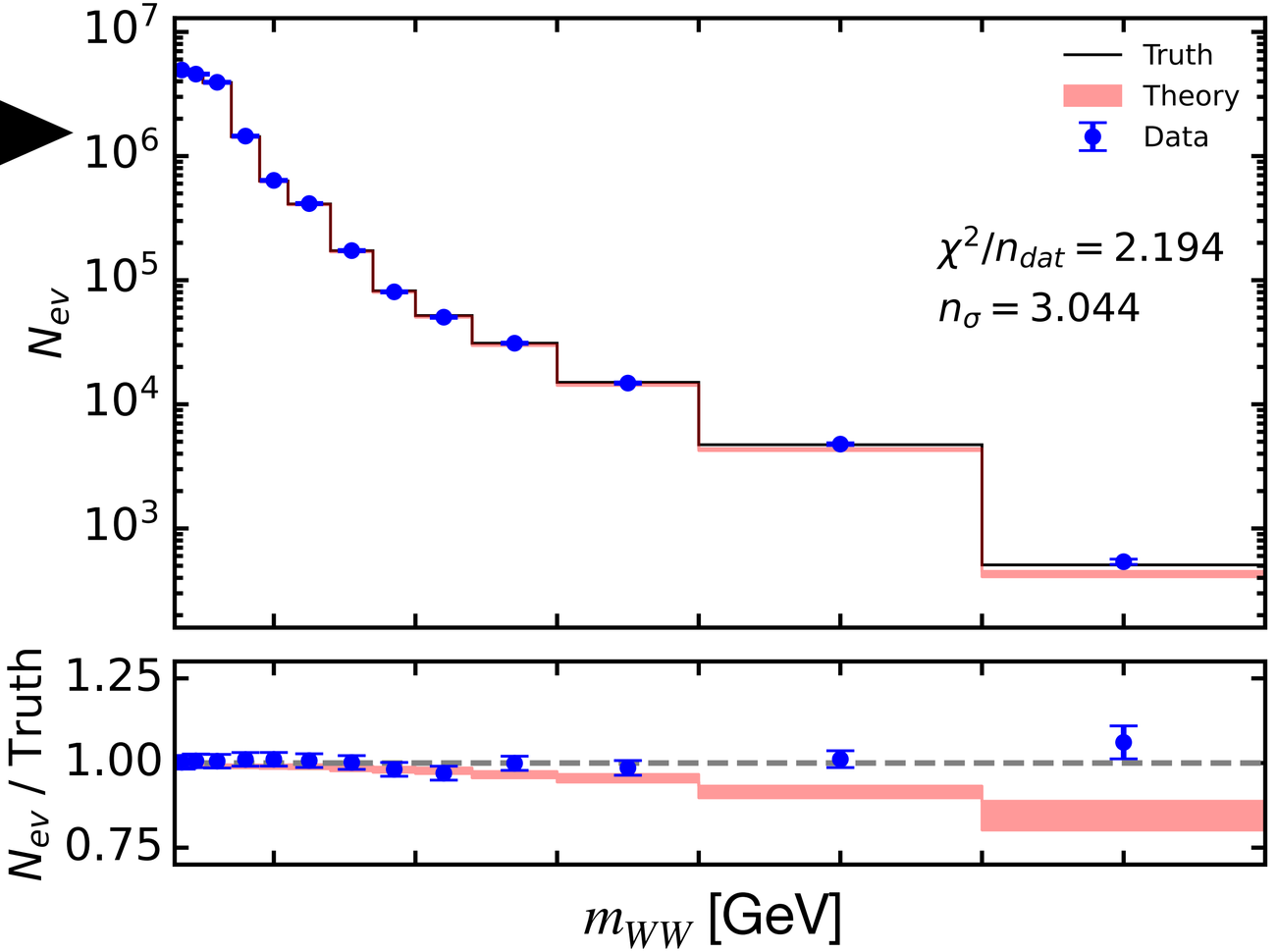
➡ Risk of missing NP



Impact of contamination: fake deviations

Analysis of contaminated predictions for HL-LHC data

$$p\bar{p} \rightarrow W^+W^-$$



Comparison of SM prediction with:

- Contaminated PDFs (red)
- Baseline PDFs (black)

➡ Fake deviation induced by PDFs

Identifying contamination

Study of ratio of observable with same parton channels

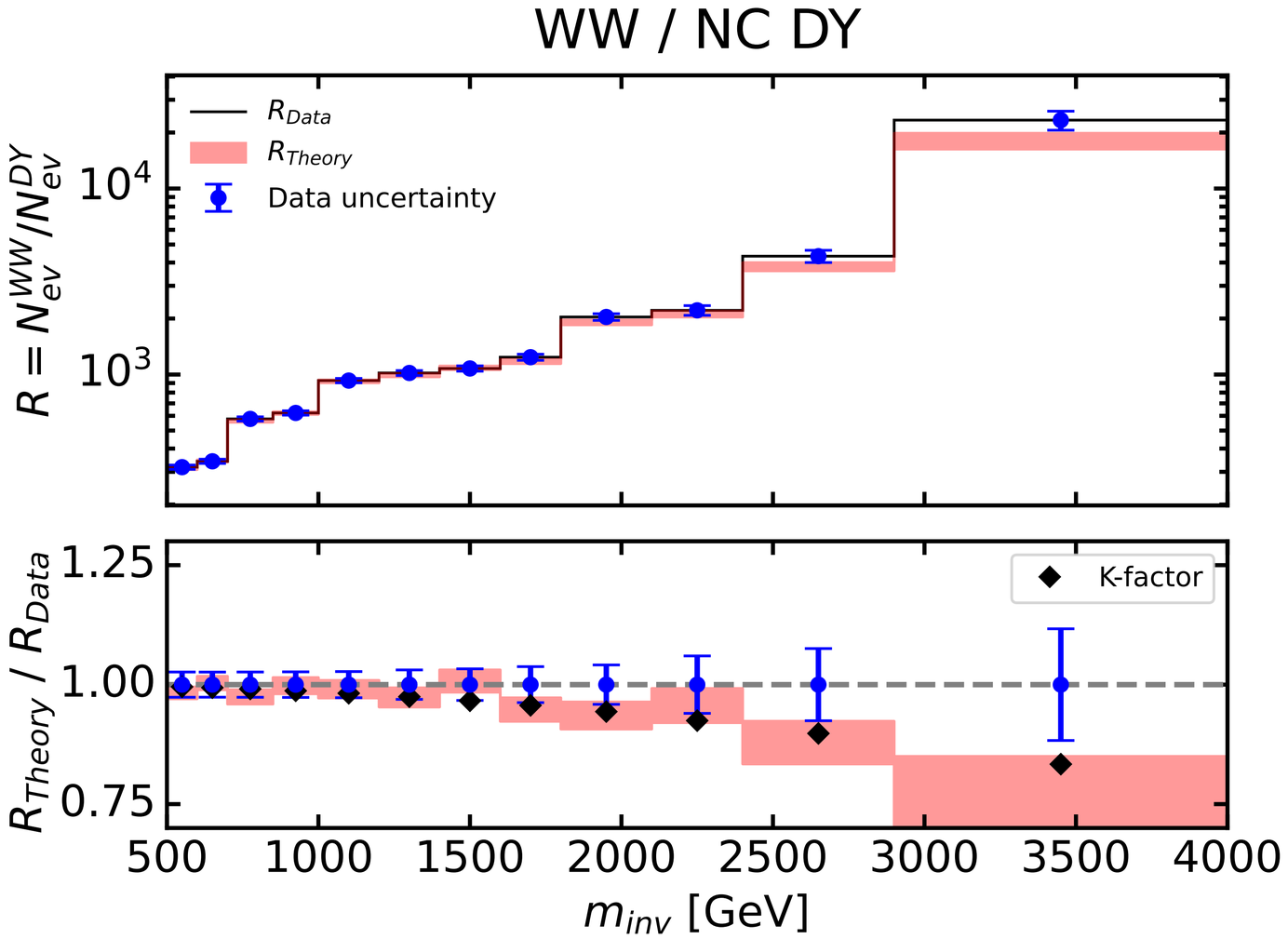
Taking ratio of:

- $p\bar{p} \rightarrow W^+W^-$
- $p\bar{p} \rightarrow l^+l^-$

Suppresses impact of PDFs

Deviation observed:

➡ New physics in the data



Preventing contamination

Adding low-energy dataset in the large-x region

Source of contamination:

- Excessive antiquark PDF flexibility in large-x region
- Accommodates real data and BSM pseudodata

Including low-energy large-x data:

- Constraint large-x region
- Safe from BSM contamination

Data-Theory comparison

		Baseline	Contaminated
	Data points (ndata)	χ^2/ndata	χ^2/ndata
NuSea (2001)	15	1.350	1.823
NuSea (2003)	89	0.8017	0.9769
SeaQuest	6	0.4192	1.034
D0 detector	9	2.385	3.046
Total	119	0.9699	1.239

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 fake deviations

- Solution to prevent contamination:
  - Consider observable ratios
  - Add large-x low-energy datasets into fits

Thank you for your attention!

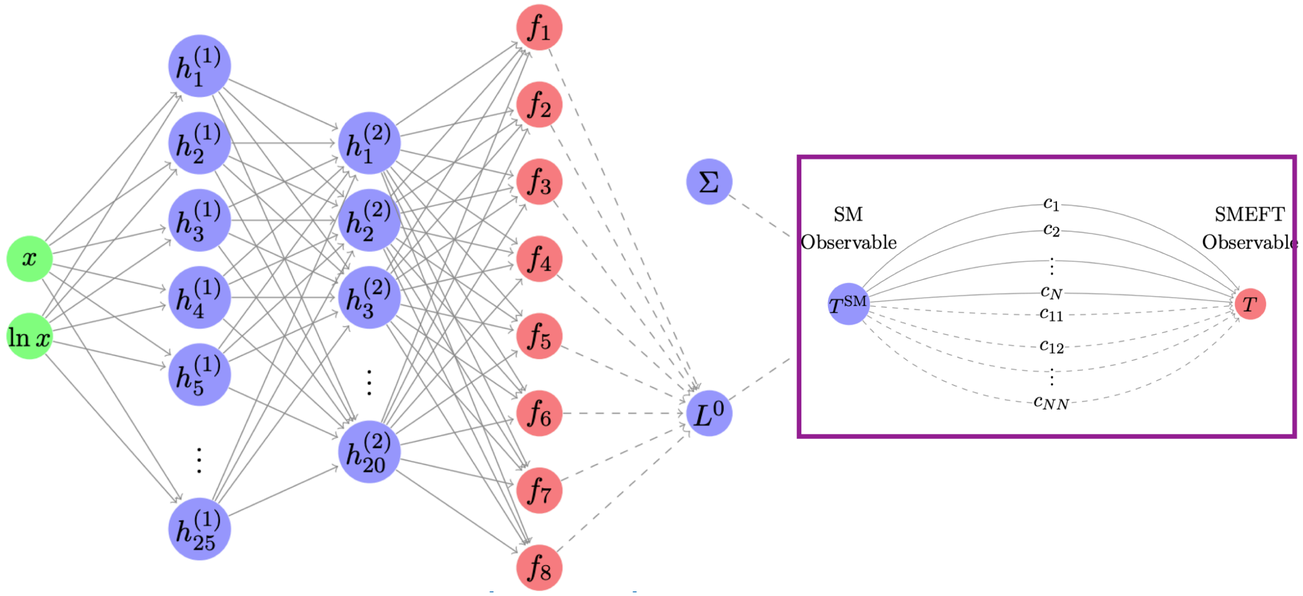
Extra slides

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)**

PDF fitting: selection criteria

Exclusion of incompatible datasets (NNPDF criteria)

Two criteria:

- χ^2-statistics:

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

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

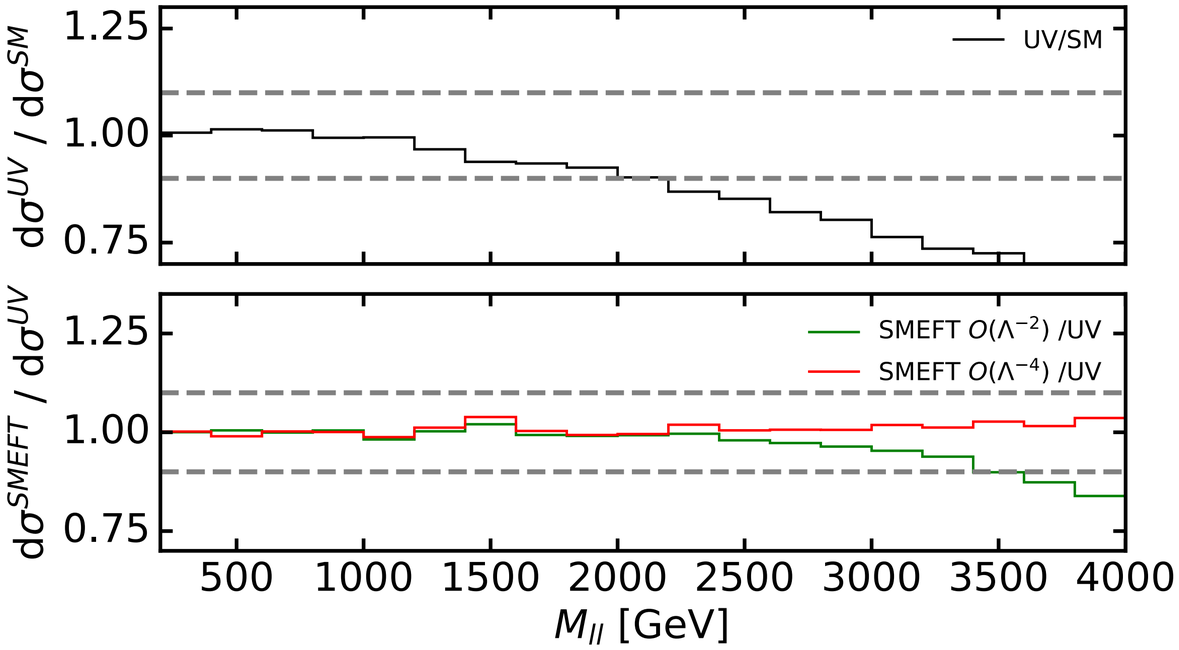
- n_σ standard deviation:

  - $n_\sigma > 2$ → excluded

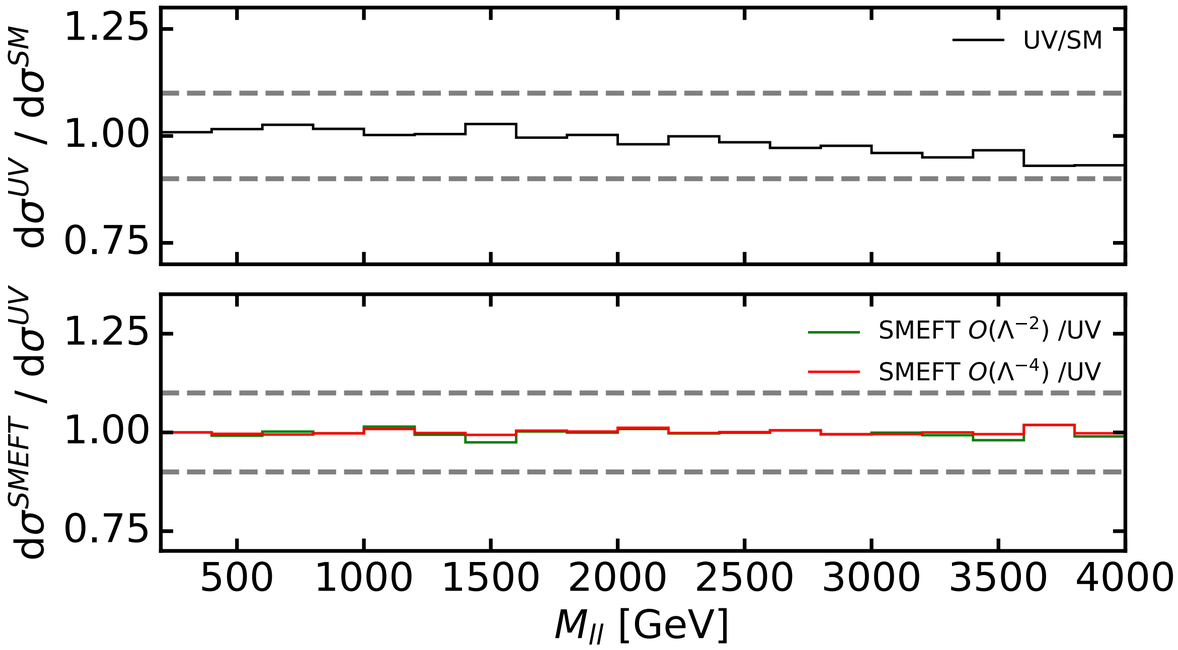
$$n_\sigma = \frac{\chi^2 - 1}{\sigma_{\chi^2}}$$

New physics scenarios: Z'

$M_{Z'} = 14.5$ TeV

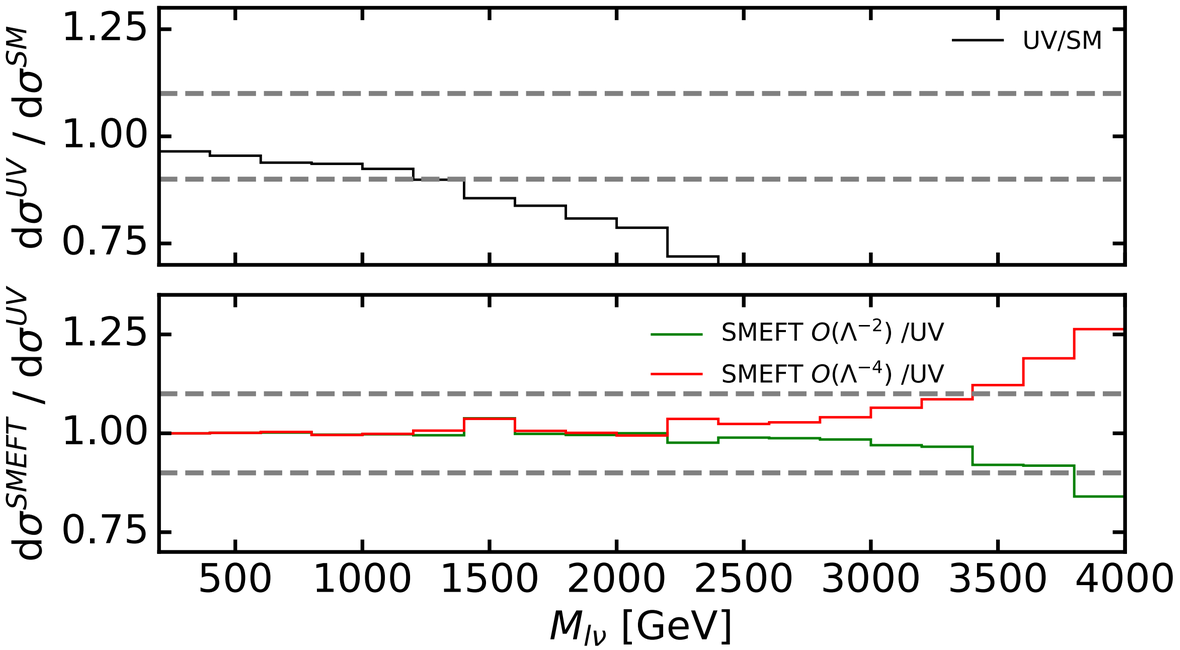


$M_{Z'} = 32.5$ TeV

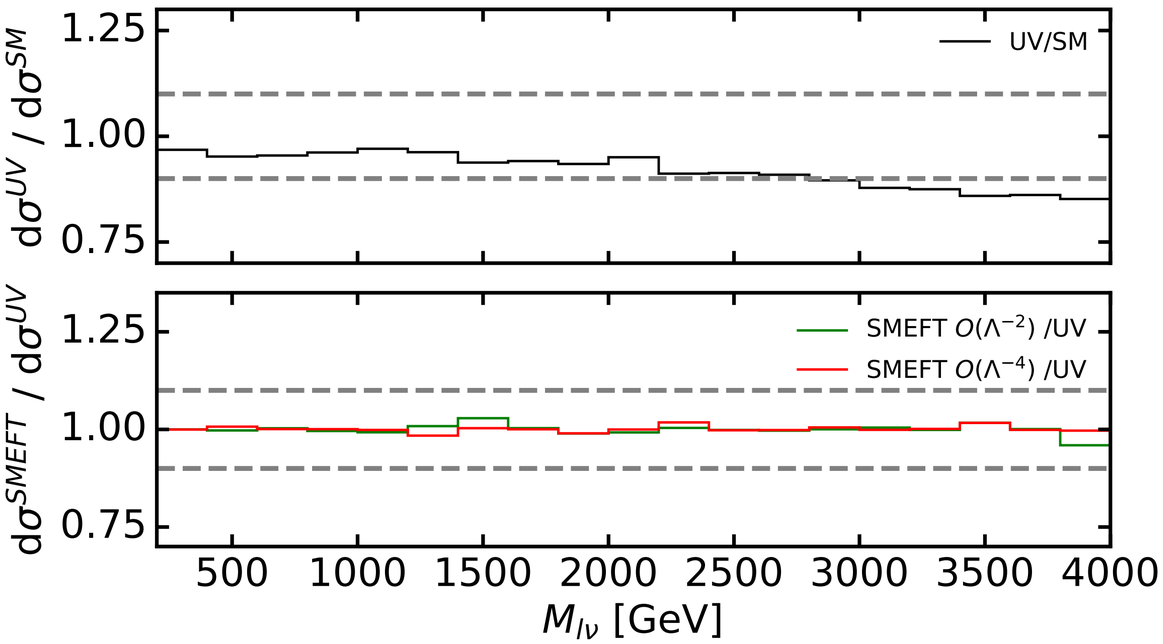


New physics scenarios: W'

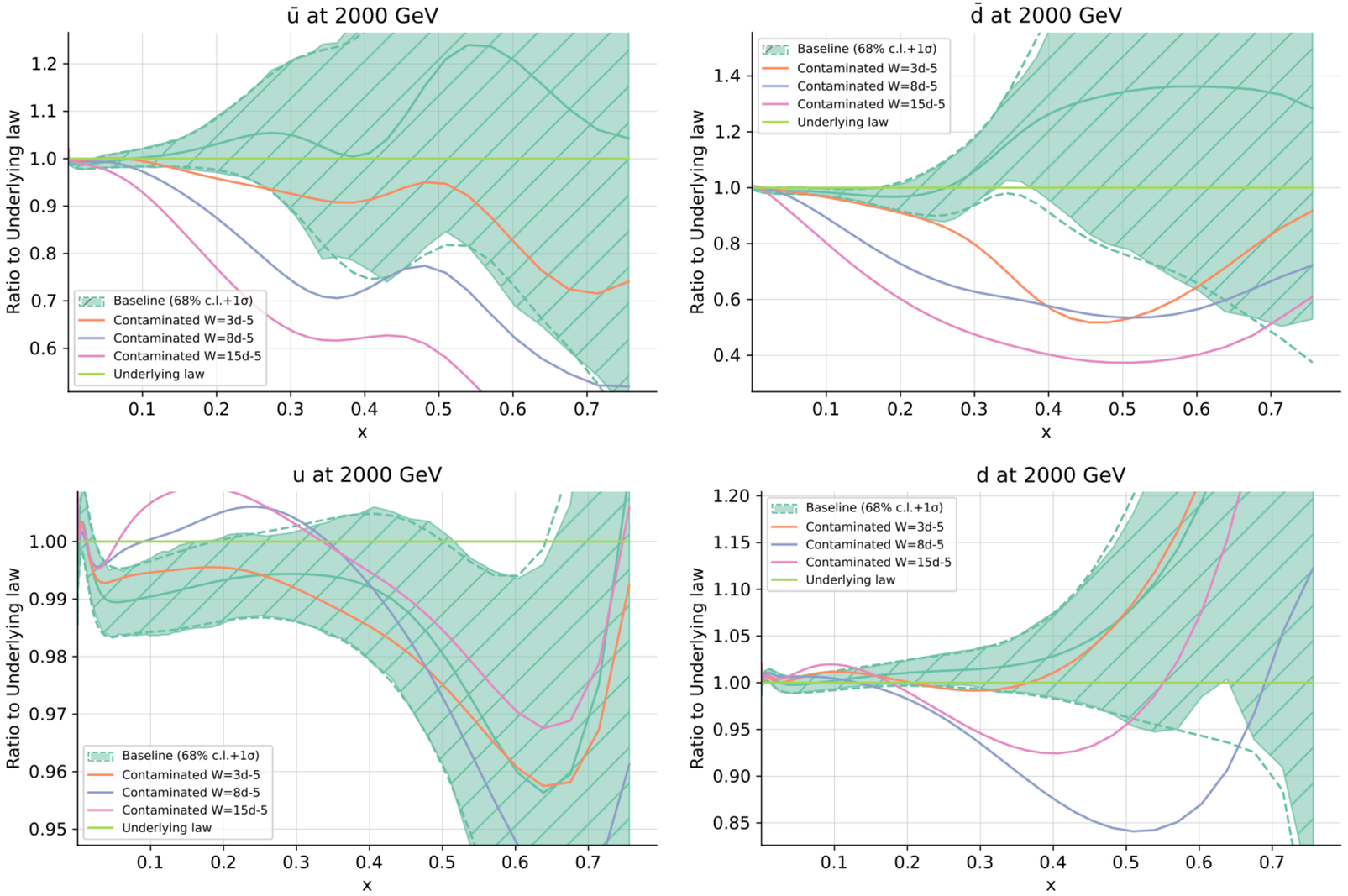
$M_{W'} = 10$ TeV



$M_{W'} = 22.5$ TeV



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

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 $\rightarrow$ H	0.70	-0.74	0.62	-0.90