How to fit PDFs in the presence of new physics?

A systematic study of the interplay of PDFs and BSM signals in global fits

[PBSP, 2307.10370, JHEP]

[PBSP, 2402.03308]

[Hammou et Ubiali, forthcoming]

European Research Council

Established by the European Commission

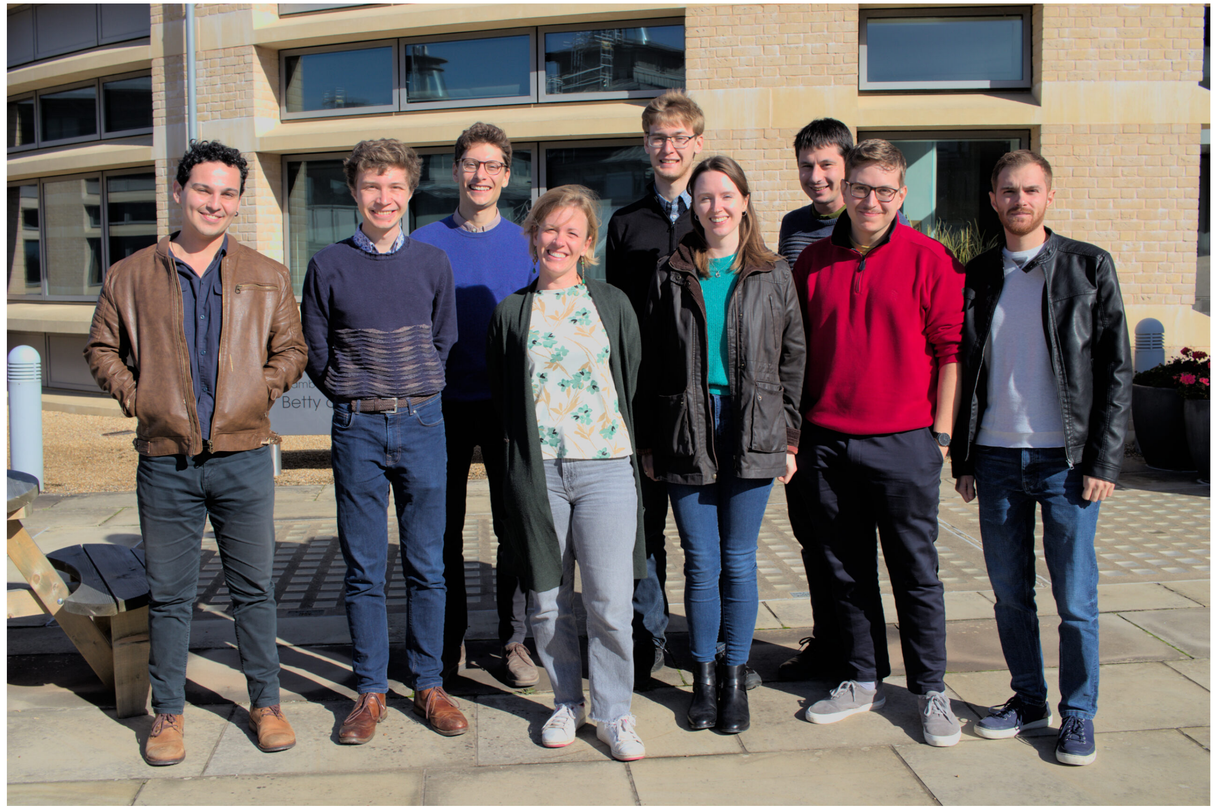
PBSP 

Elie Hammou, University of Cambridge

Rencontres de Moriond QCD 2024

Our group: PBSP

Physics Beyond the Standard Proton

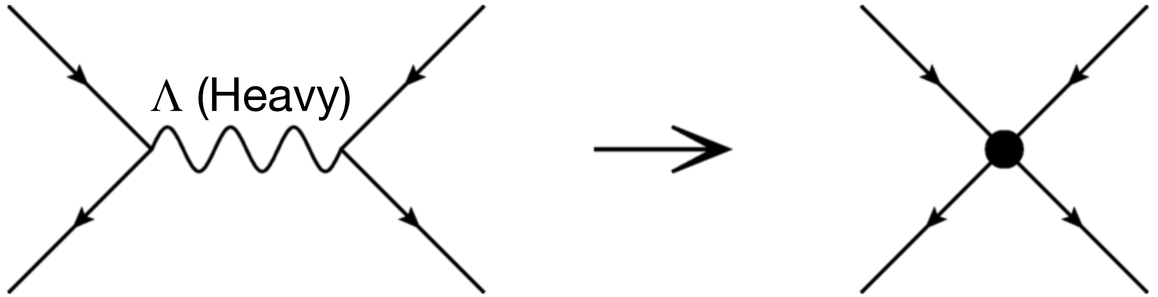


- Led by Maria Ubiali
- Based in Cambridge
- Focus on PDF and EFT interplay
  - Indirect search for heavy new physics
  - Investigate robust fitting methods

Heavy New Physics: from UV to SMEFT

Integrate heavy fields out:

UV Model



EFT Model

[10.1007/s10773-021-04723-1]

Obtain model independent Lagrangian:

$$\mathscr{L}^{\text{UV}} = \mathscr{L}^{\text{SM}} + \mathscr{L}^{\text{Heavy}} \longrightarrow \mathscr{L}^{\text{SMEFT}} = \mathscr{L}^{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \dots$$

- Dim 6 EFT operators with SM fields: $\mathcal{O}_i^{(6)}$
- Wilson coefficients fittable from data: $\frac{c_i}{\Lambda^2}$

Background on Parton Distribution Functions

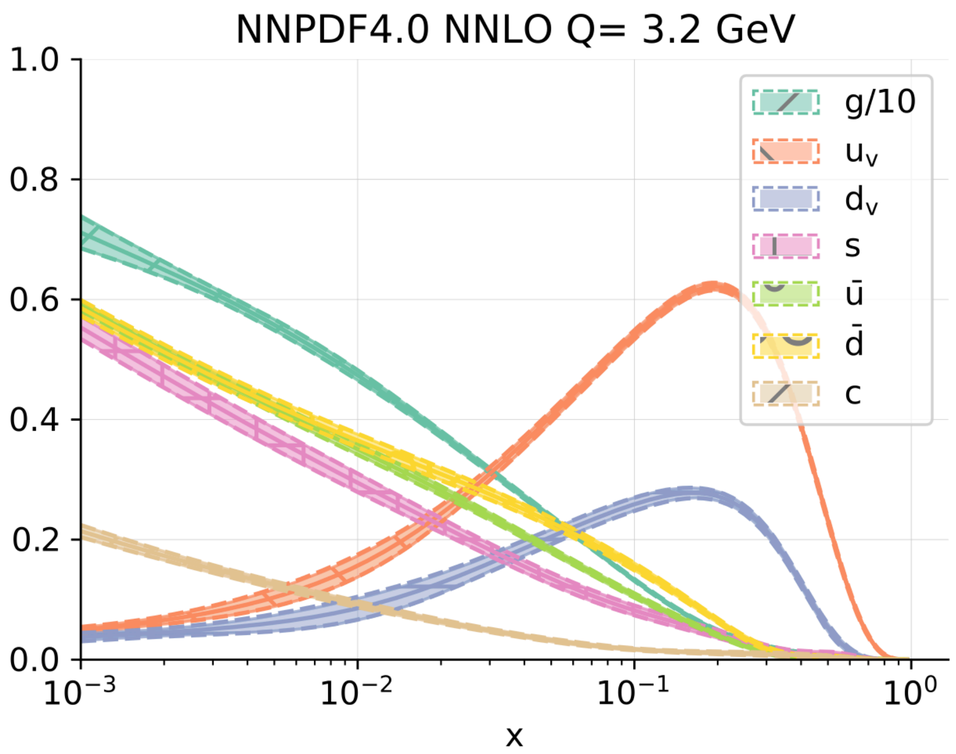
Hadron collider observable: $\sigma = \hat{\sigma} \otimes f_1 \otimes f_2$

PDFs in a nutshell:

- describe proton in terms of partonic content
- $f(x, Q)$
- x dependance: non-perturbative QCD

➡ **Fitted from data**

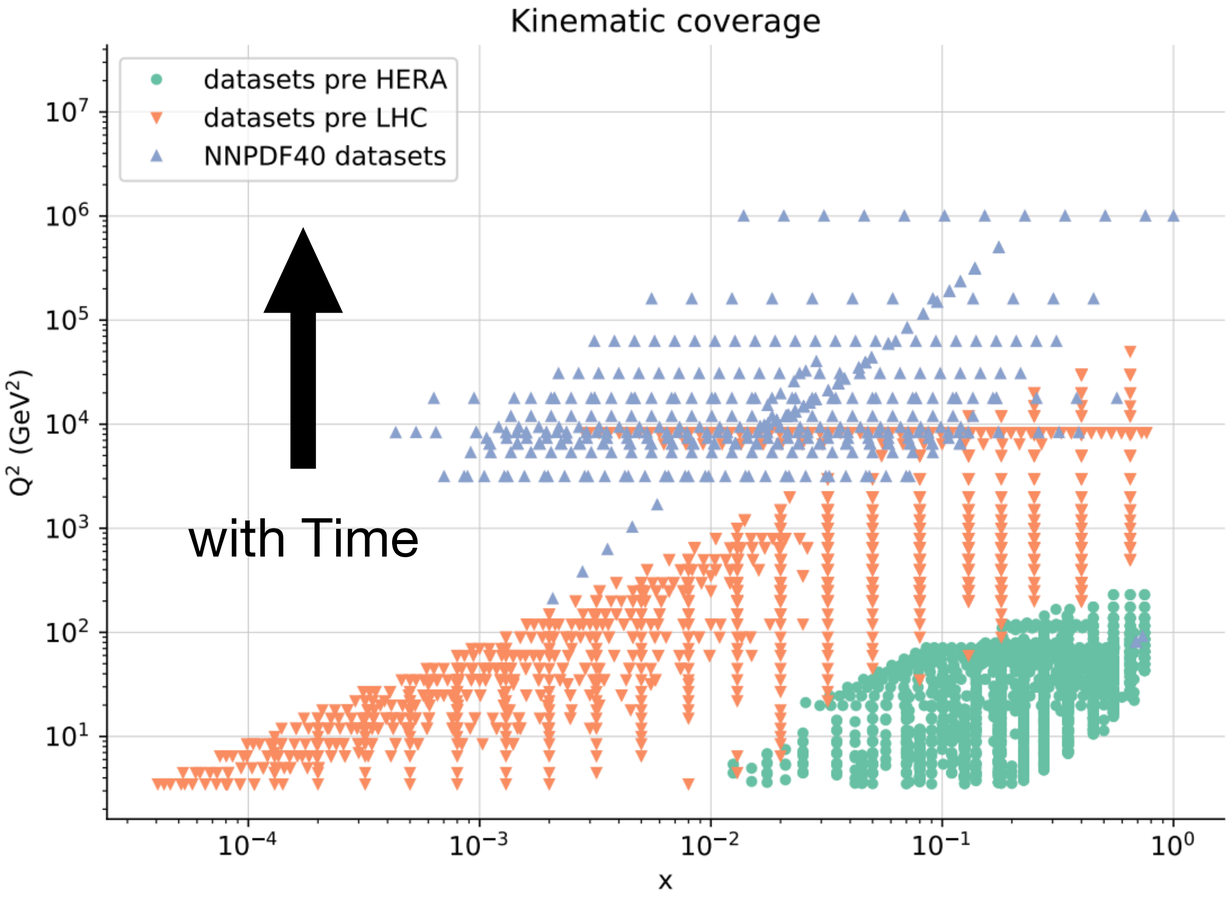
Using NNPDF methodology



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

Fitting PDF from data

The dataset used by NNPDF



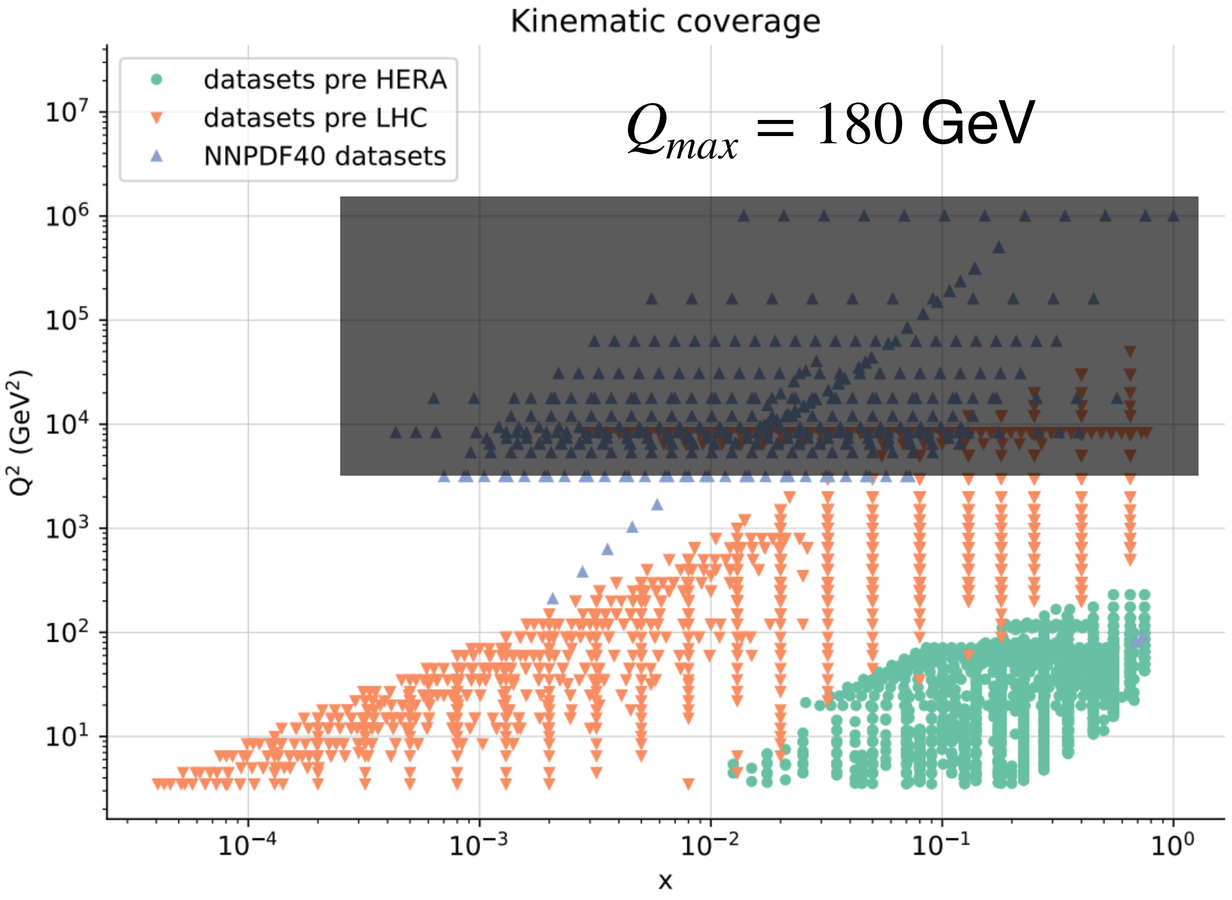
[Ball et al., NNPDF4.0, 2109.02653]

Evolution of the dataset through time:

- Moved toward higher energies
- 30% is LHC data
- More to come with HL-LHC run

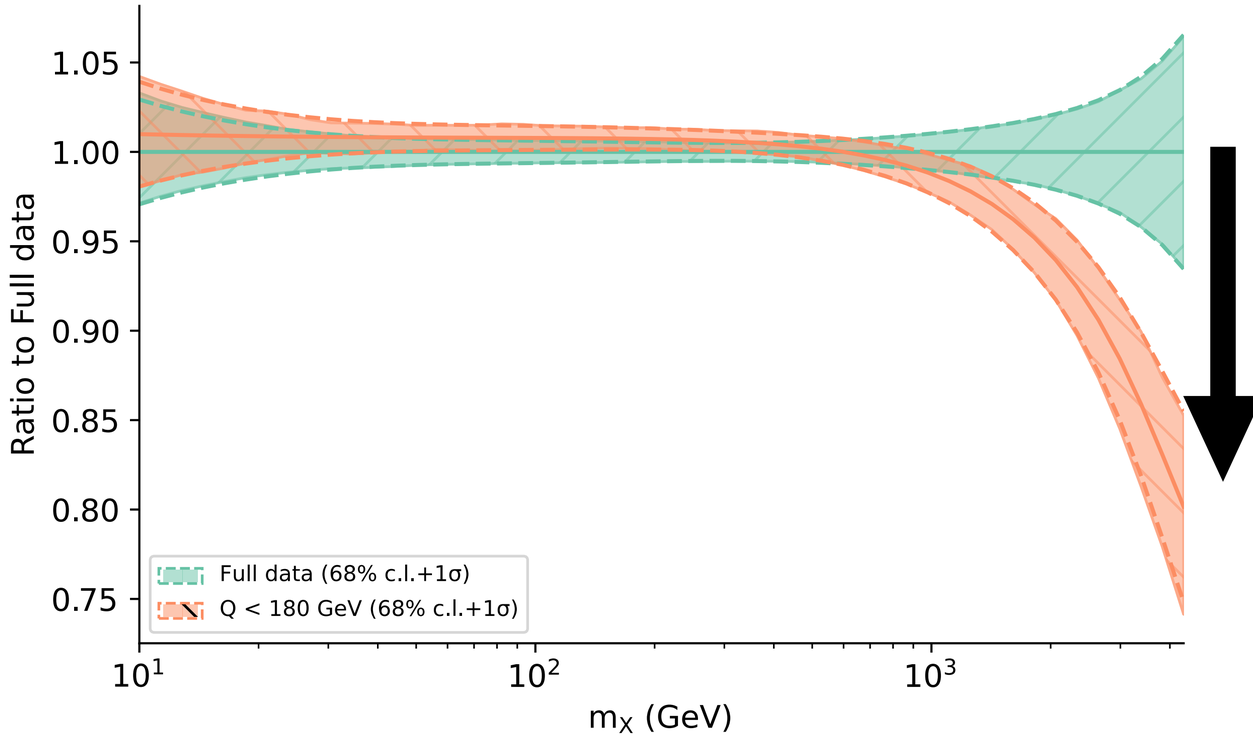
Discrepancy between low and high-energy data fits

Comparison of full data and no LHC PDF fit



Impact of energy cut

gg luminosity $\mathscr{L}_{gg} = f_g \otimes f_g$
$\sqrt{s} = 13$ TeV



Risk of absorbing new physics in PDFs?

Methodology for risk assessment

Perform a “Contamination test”:

1. Choose a BSM model
2. Produce BSM pseudodata
3. Fit PDFs from pseudodata assuming SM
4. Compare results with baseline PDFs (no BSM physics)

[PBSP, 2307.10370]

Contamination criteria:

- Incompatible with baseline
- Fit quality does not deteriorate

➡ $\chi^2 = (Dat - Th)^{\top} \cdot \Sigma_{cov}^{-1} \cdot (Dat - Th)$

PDF contamination:

➡ **PDFs have absorbed new physics signals**

New physics scenario: W'

$pp \to l^- \bar{\nu}$ $M_{W'} = 13.8$ TeV

Generation of the pseudodata

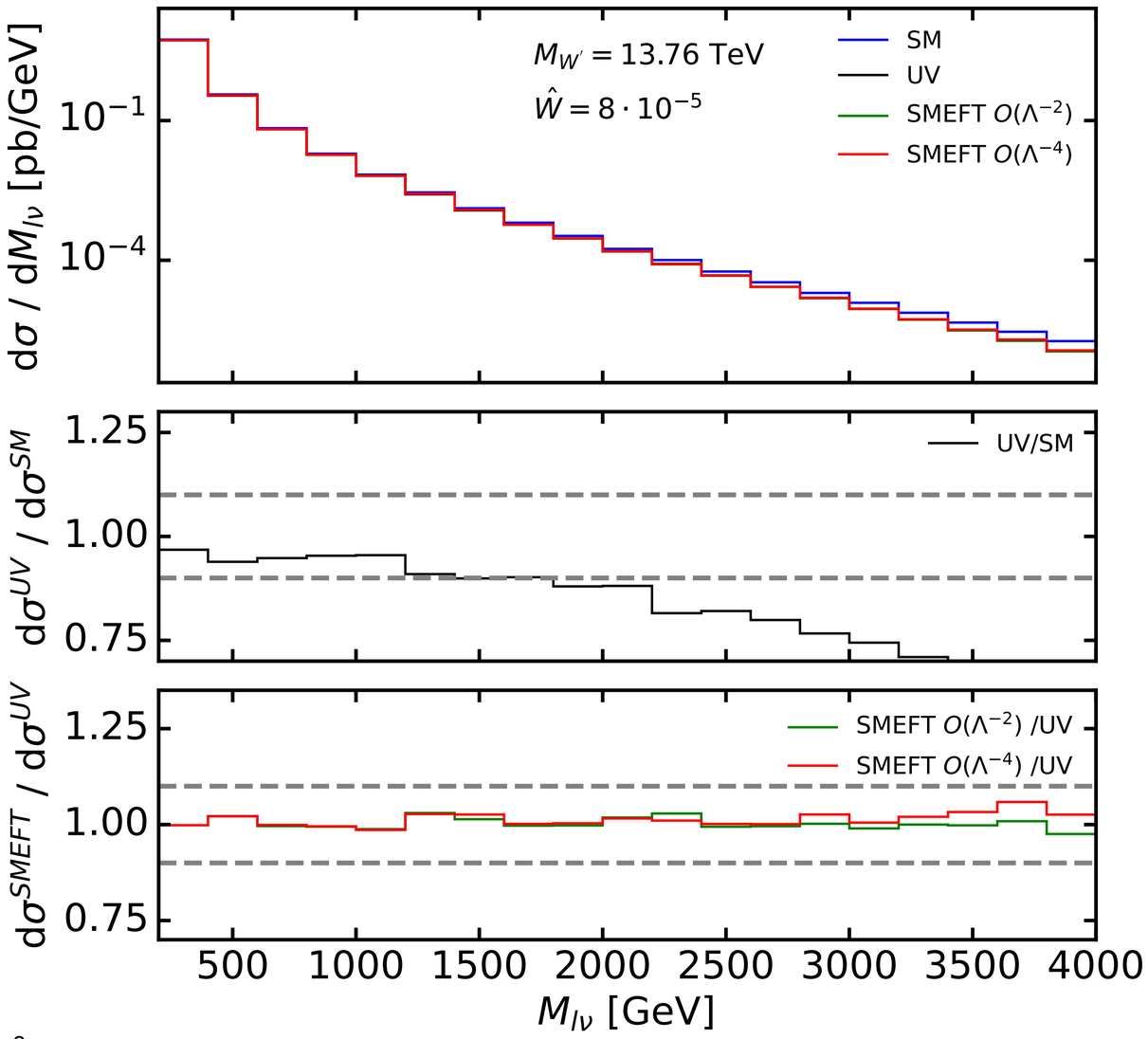
$$\mathscr{L}_{UV}^{W'} = \mathscr{L}_{SM} - \frac{1}{4} W'^{a}_{\mu\nu} W'^{a,\mu\nu} + \frac{1}{2} M^2_{W'} W'^{a}_{\mu} W'^{a,\mu} - g_{W'} W'^{a,\mu} \sum_{f_L} \bar{f}_L T^a \gamma^\mu f_L$$

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

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

Impacts Drell-Yan

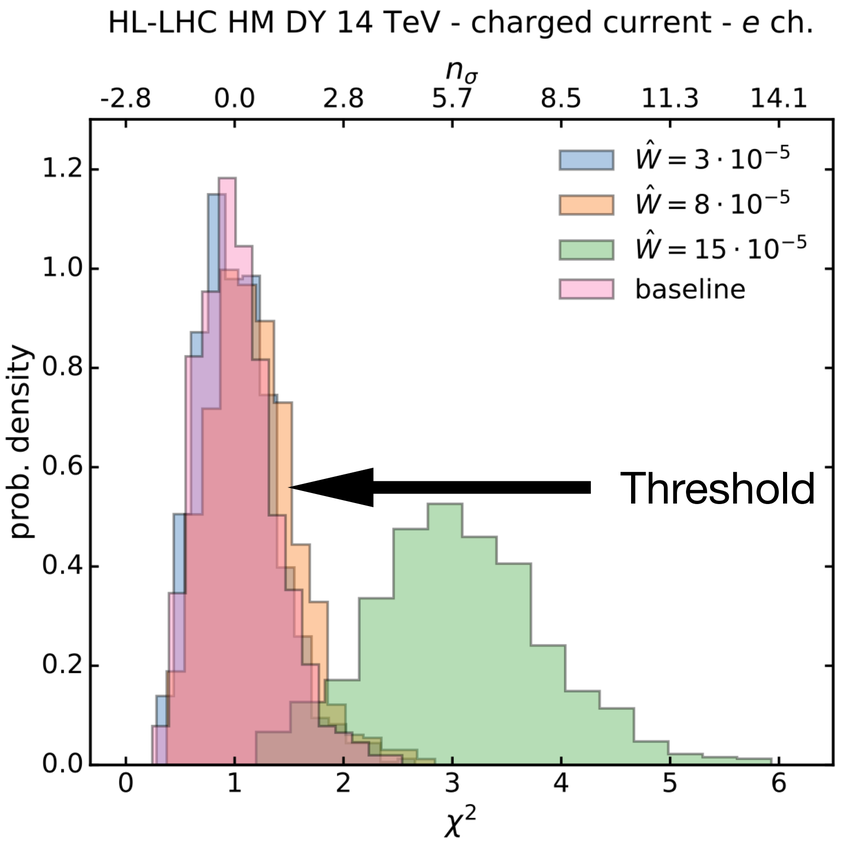
HL-LHC Projections



Impact of contamination: missing new physics

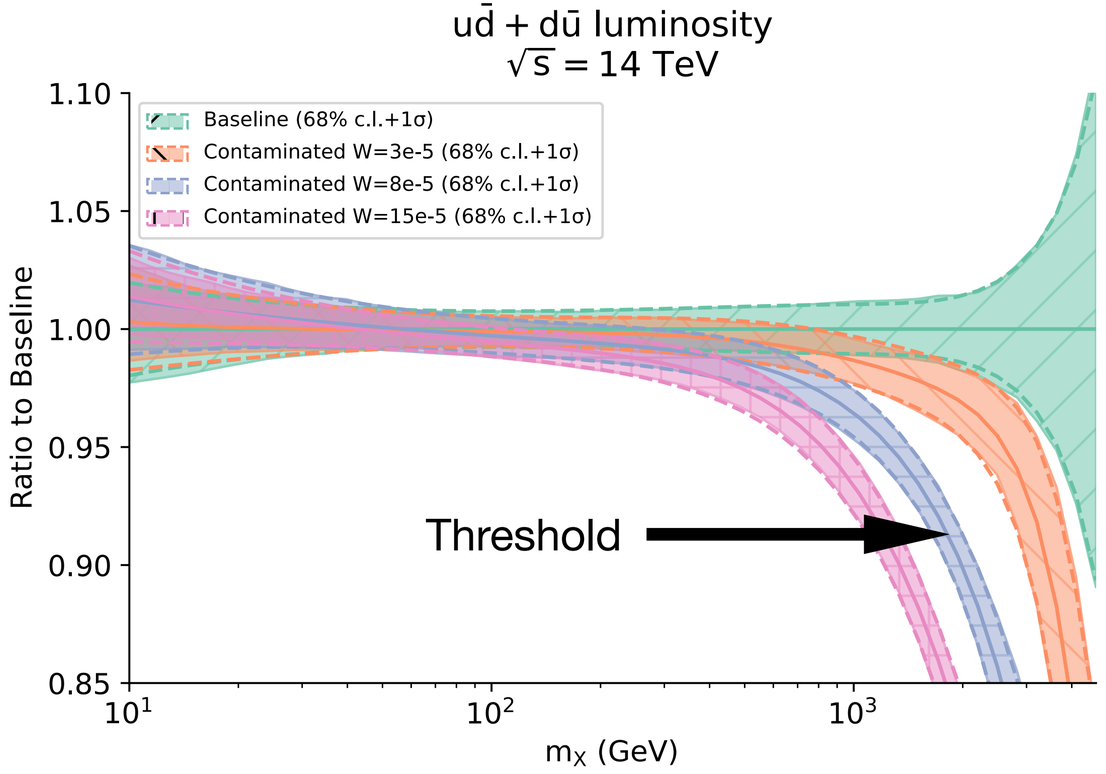
Comparison between contaminated and Baseline PDFs

Fit quality



$M_{W'} = 13.8$ TeV

Impact on PDFs



$$\hat{\sigma}_{BSM} \otimes \mathcal{L}_{baseline} \approx \hat{\sigma}_{SM} \otimes \mathcal{L}_{cont}$$

Impact of contamination: fake deviations

SM predictions with:

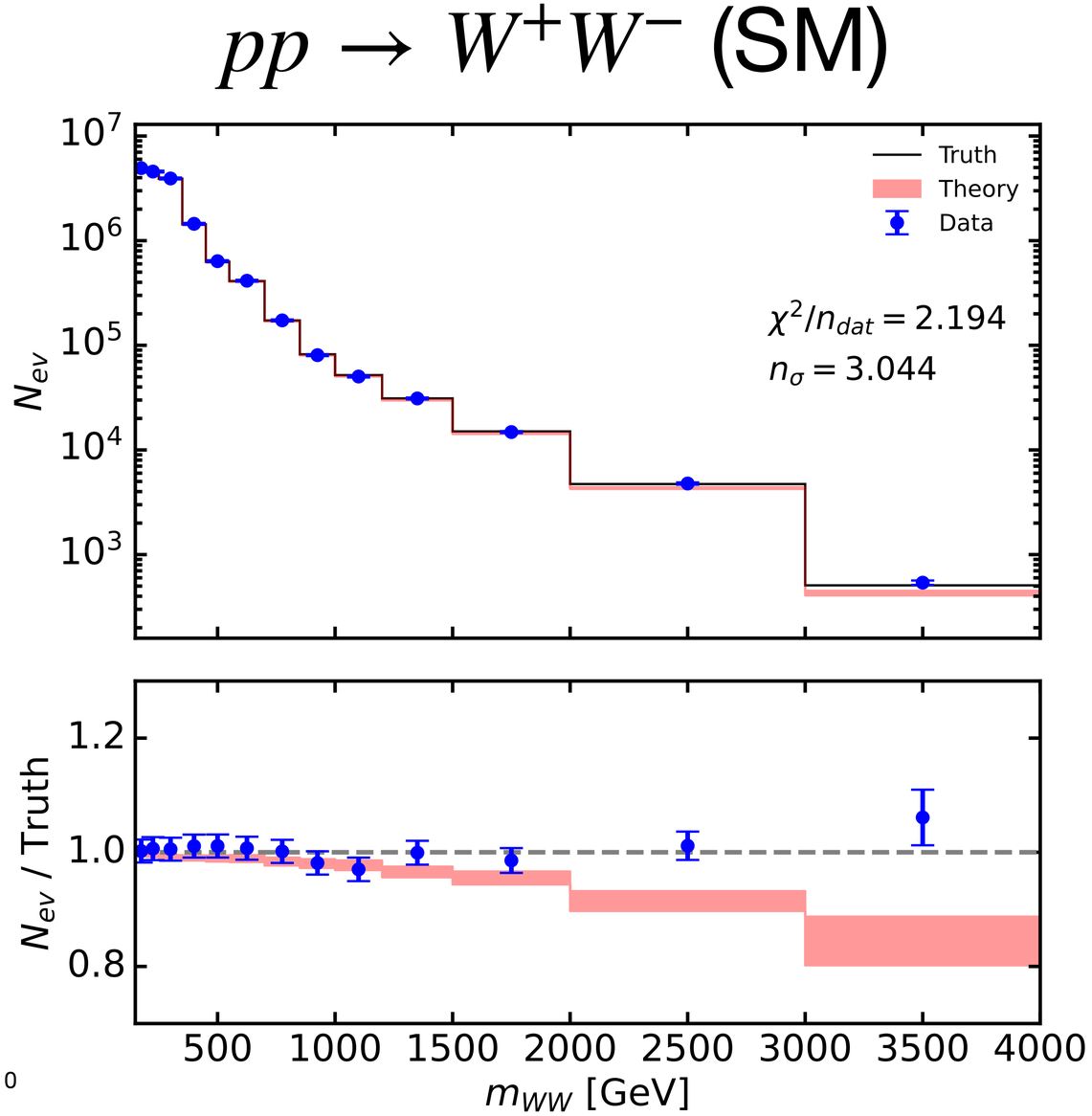
- Contaminated PDFs (red)
- Baseline PDFs (black)

➡ Fake deviation in other sectors

Also seen in:

WH, WZ, ZH production

HL-LHC Projections



Synergy of high and low-energy data

Adding low-energy dataset constraining the large-x region

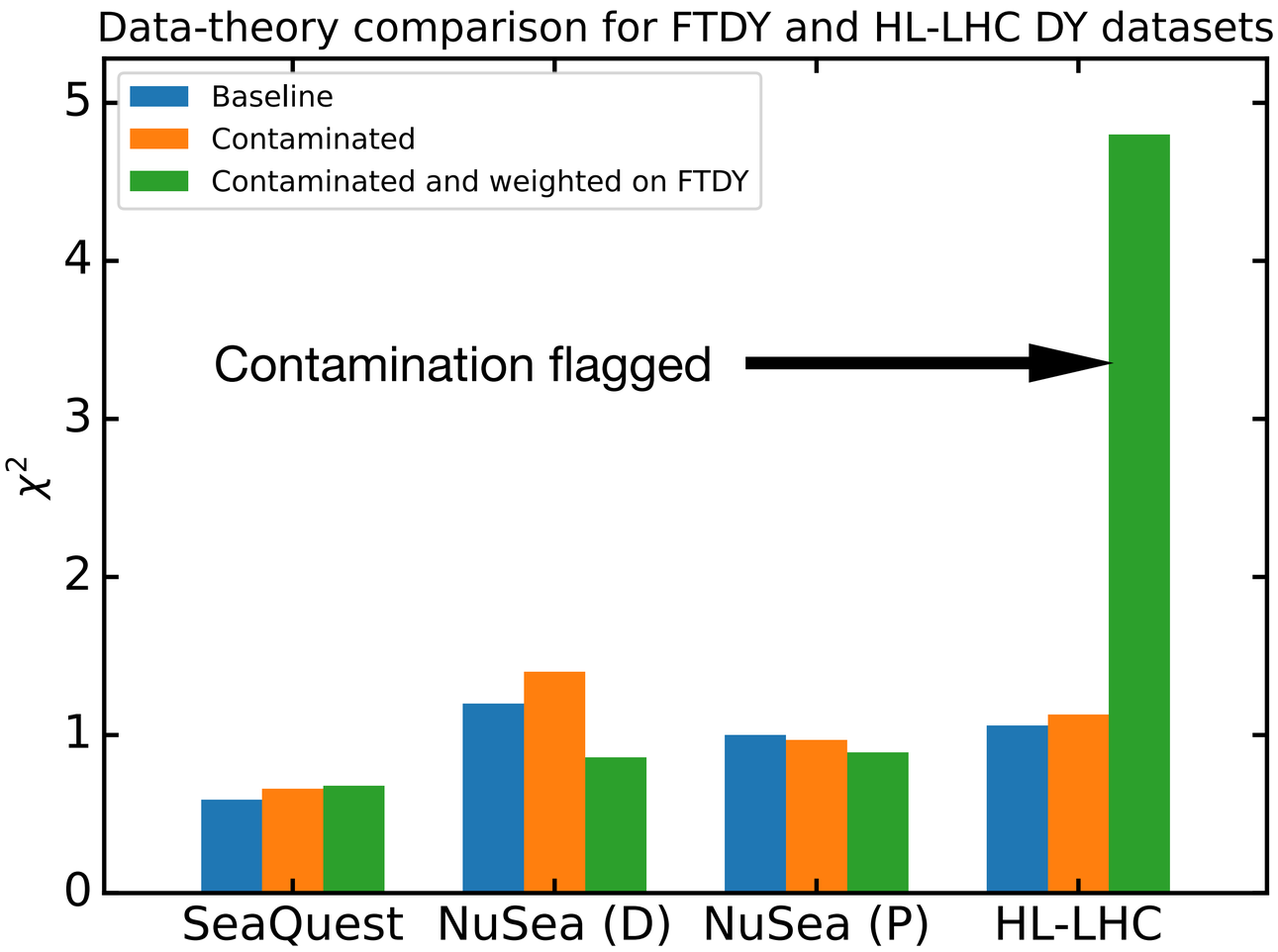
Excessive antiquark PDF flexibility in large-x region:

- ➡ Accommodates real data and BSM pseudodata
- ➡ Allows contamination

Including low-energy large-x data:

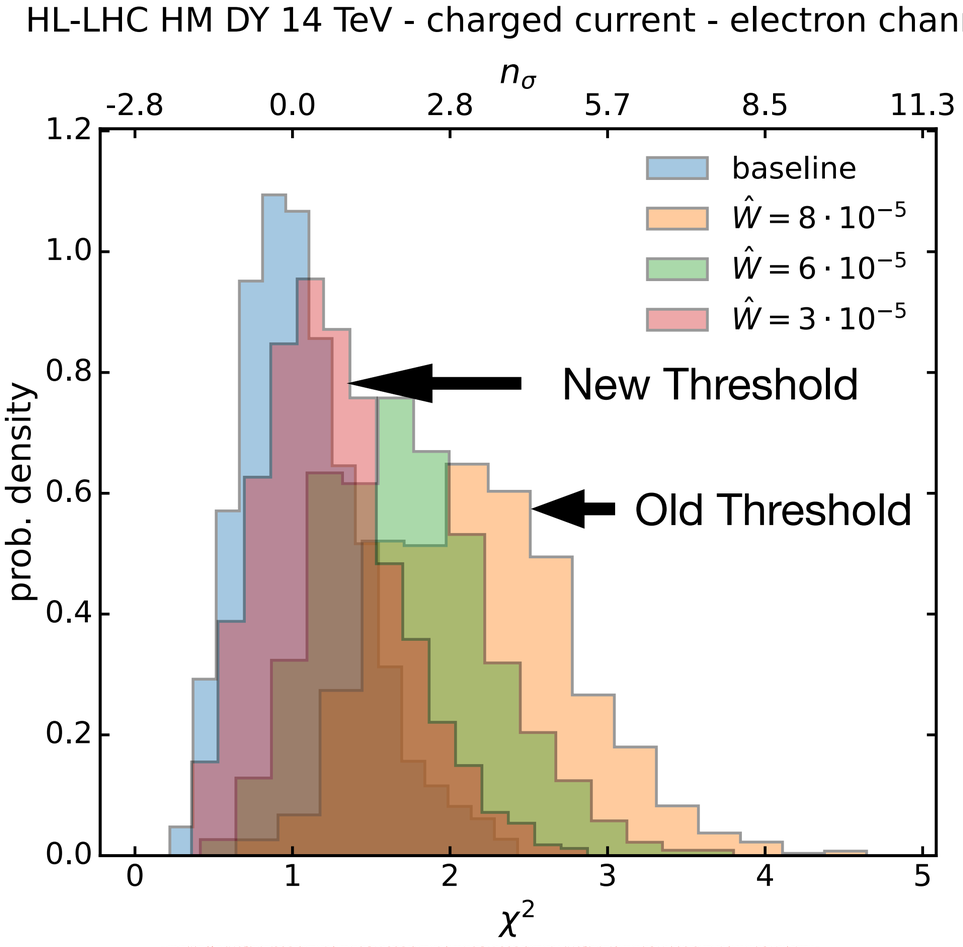
- Constraint large-x region
- Safe from BSM contamination

[Hammou et Ubiali, incoming]

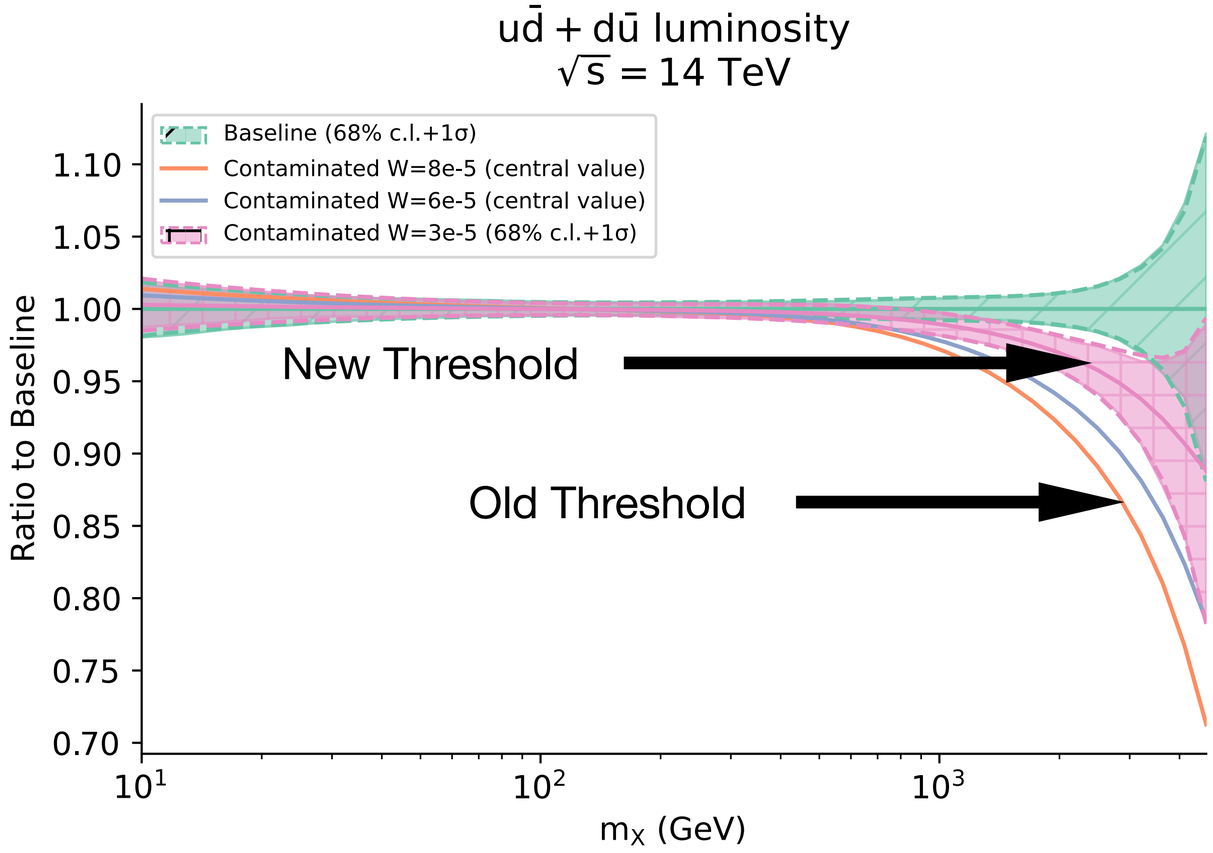


Impact of Forward Physics Facility

Projection data from FASER, FASER2, SND and AdvSND,



$M_{W'}$: 13.8 → 22.5 TeV



Reduces fake deviations

Simultaneous fit of PDF and new physics

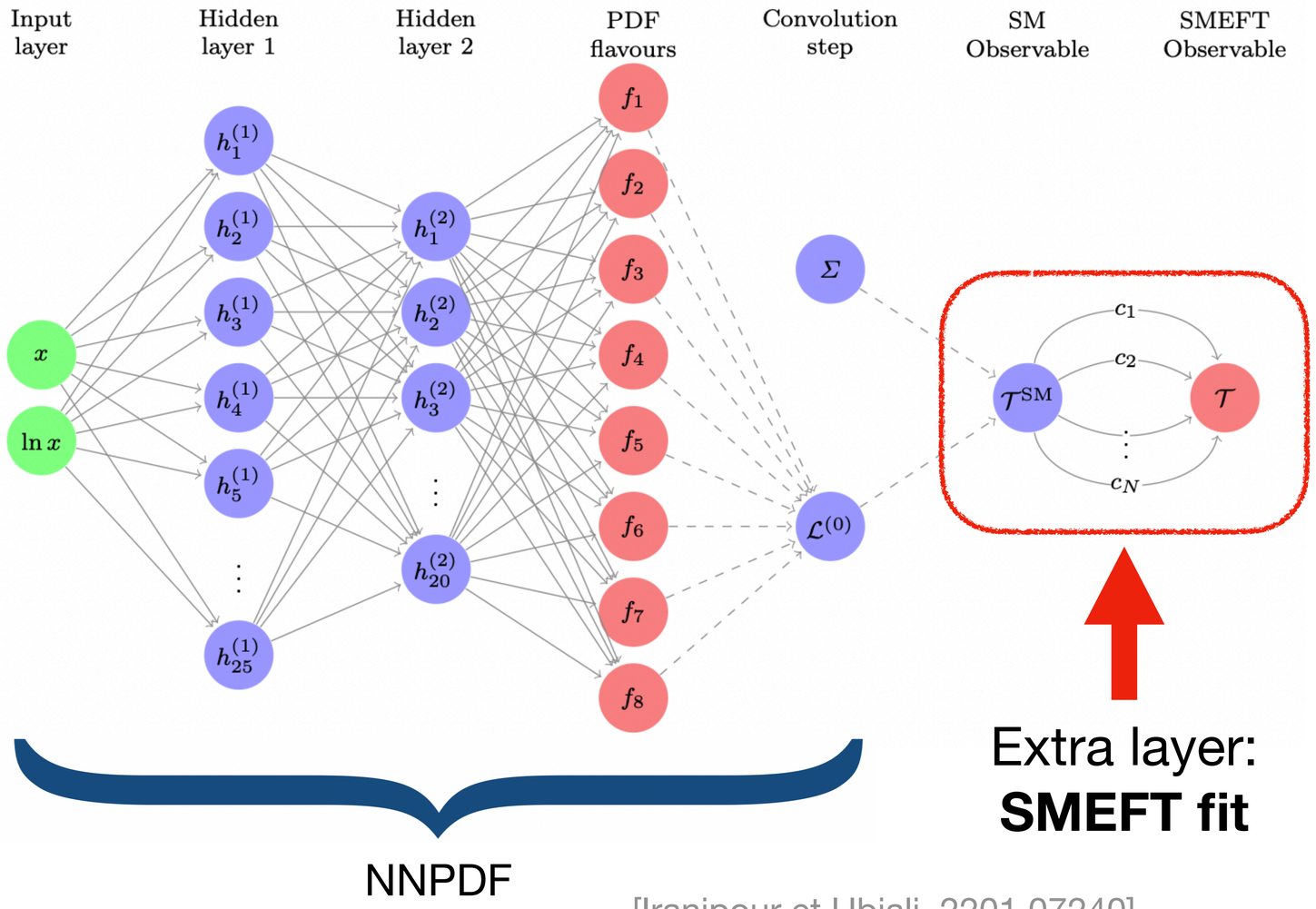
Presentation of the tool: SIMUnet

SIMUnet:

- Open-source tool: github.com/HEP-PBSP/SIMUnet

  [PBSP, 2402.03308]

- Fits PDFs and WC simultaneously
- Performs contaminated PDF fits

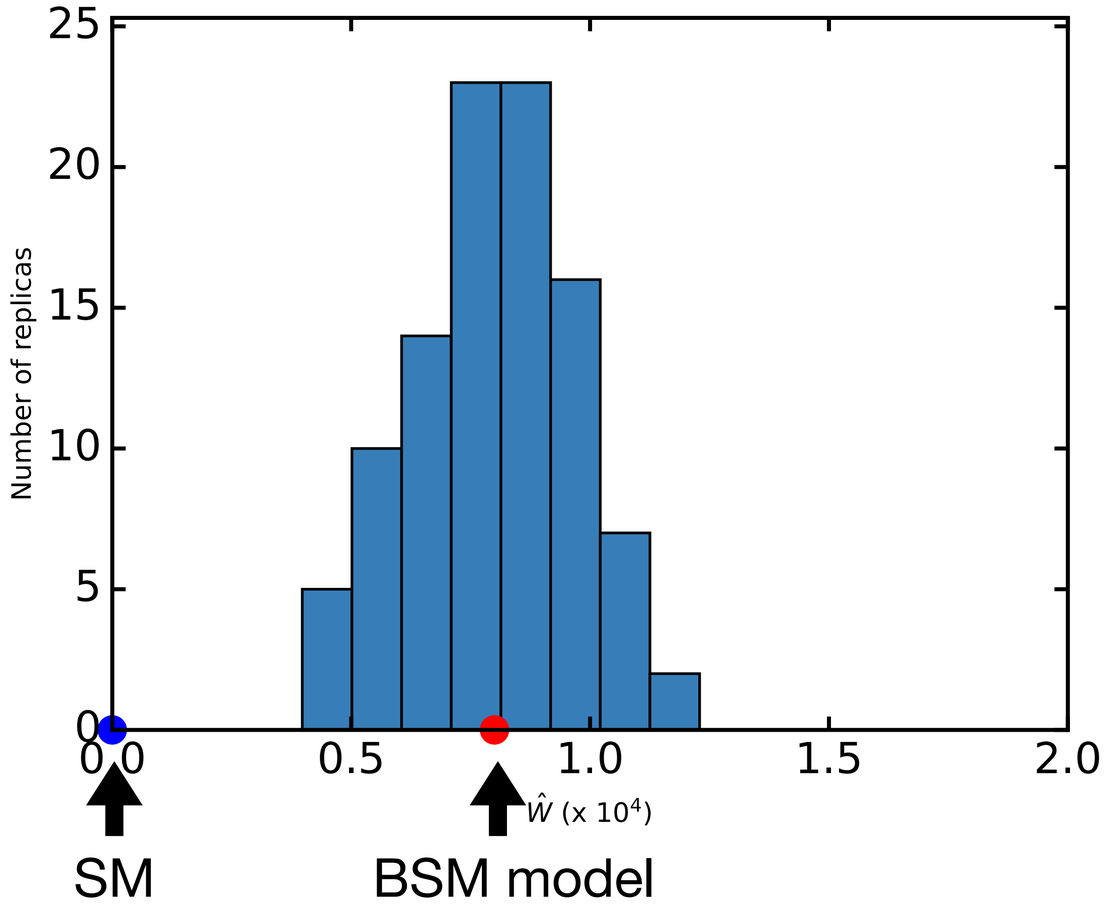


[Iranipour et Ubiali, 2201.07240]

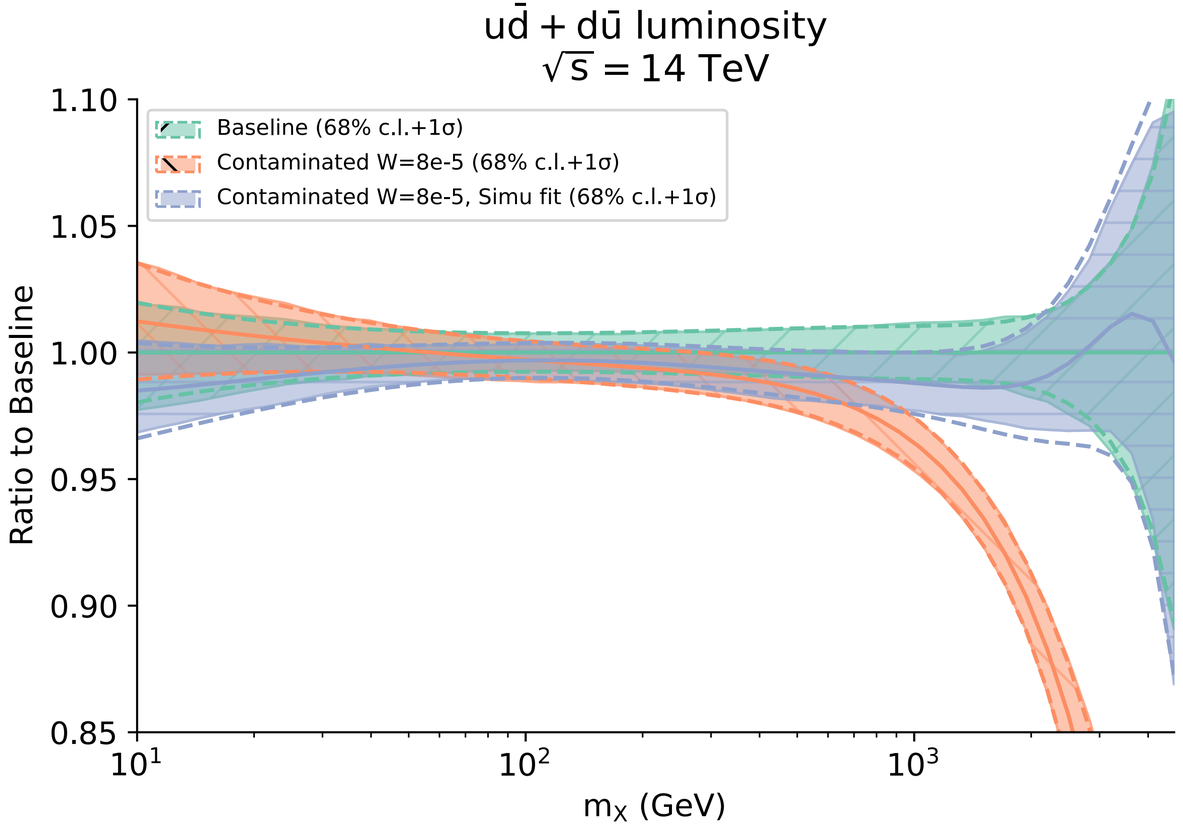
Simultaneous fit of PDF and new physics

Disentangling PDF contamination

SMEFT Fit



PDF Fit



Summary and outlook

- PDF contamination: BSM model dependant
  - Not seen for Z'
  - Ongoing study for gluon sector
- Signs of W' got fitted away in PDF parametrisation
  - Missed new physics
  - Introduced fake deviations in other sectors
- Solution to prevent contamination:
  - Add precise large-x low-energy datasets into fits: FTDY, FPF, EIC… (forthcoming)
  - Fitting simultaneously PDF and new physics: **SIMUnet** tool available

You can contact me at:
eh651@cam.ac.uk

Thank you for your attention!

Extra slides

New physics scenarios: Z'

Generation of the pseudodata

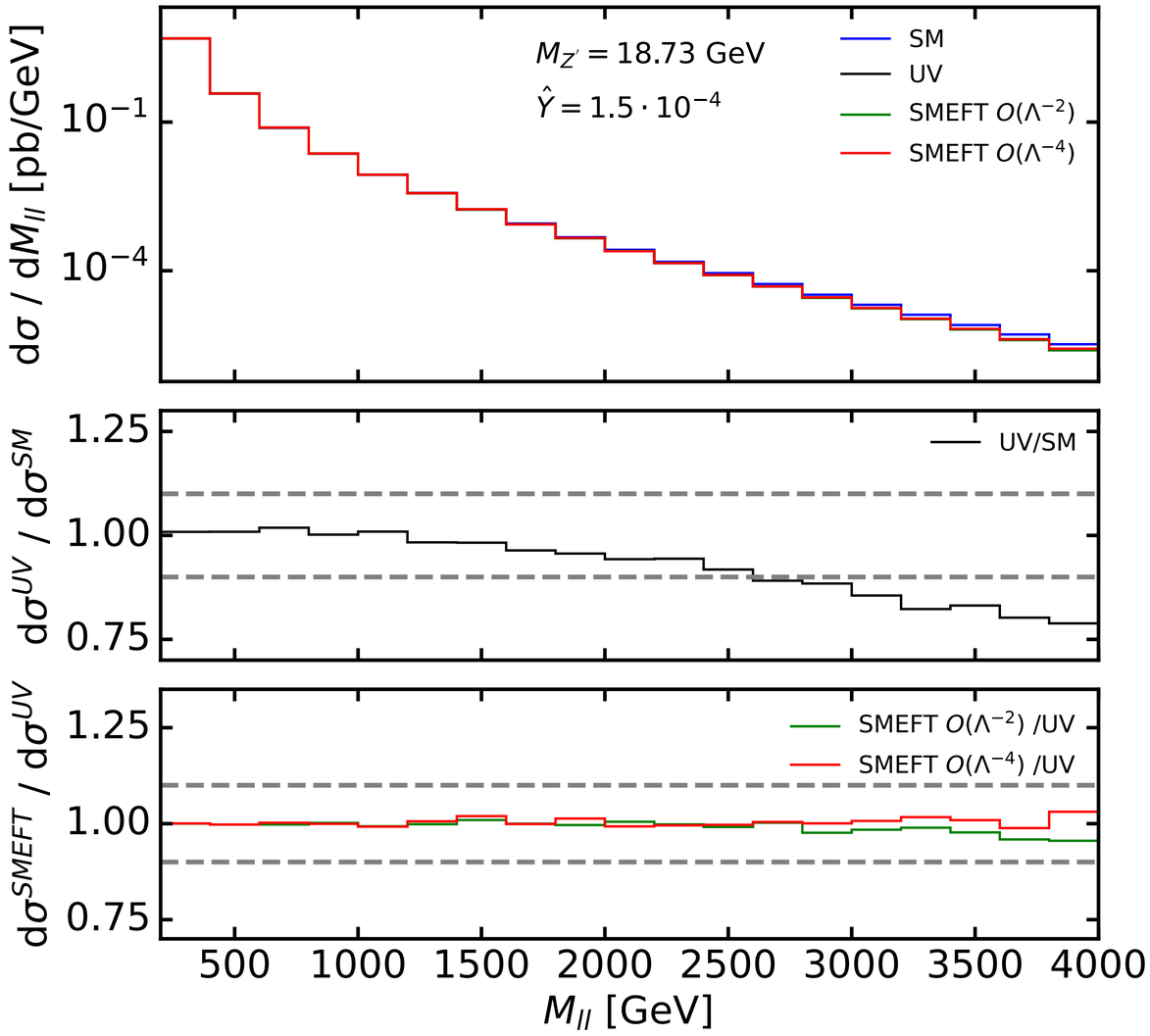
$$\mathscr{L}^{Z'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2_{Z'}}{2M^2_{Z'}} J^{\mu}_{Y} J_{Y,\mu}$$

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

Impacts neutral current Drell-Yan processes

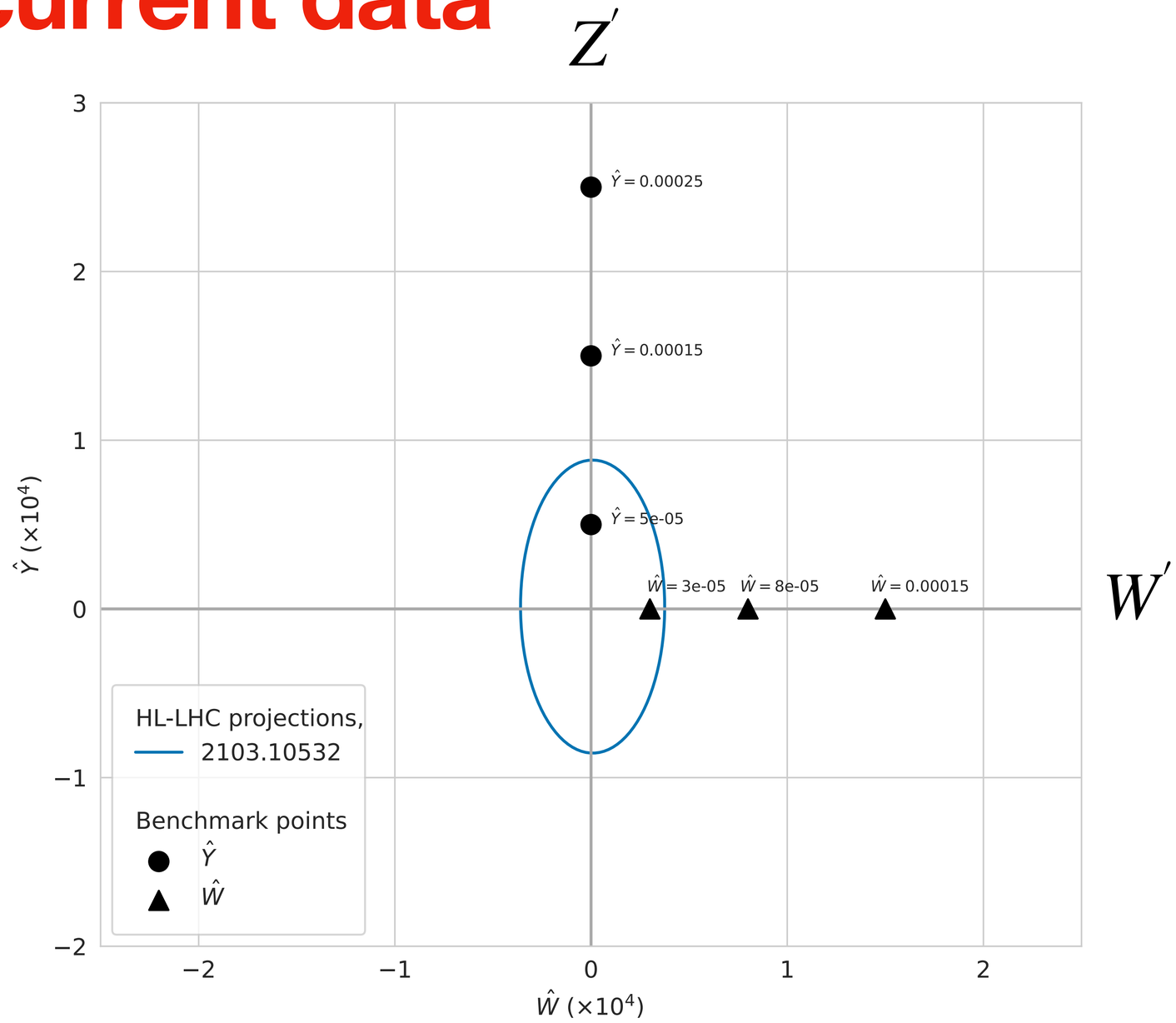
$$p\bar{p} \rightarrow l^+ l^-$$

$M_{Z'} = 18.7$ TeV



Constraints from current data

- New physics scenarios compared to constraints at 95% CL



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

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

PDF fitting: selection test

Do our contaminated datasets pass the selection criteria?

Z'

Selection test: ✗

➡ Excluded from PDF fit

No impact on PDFs

W'

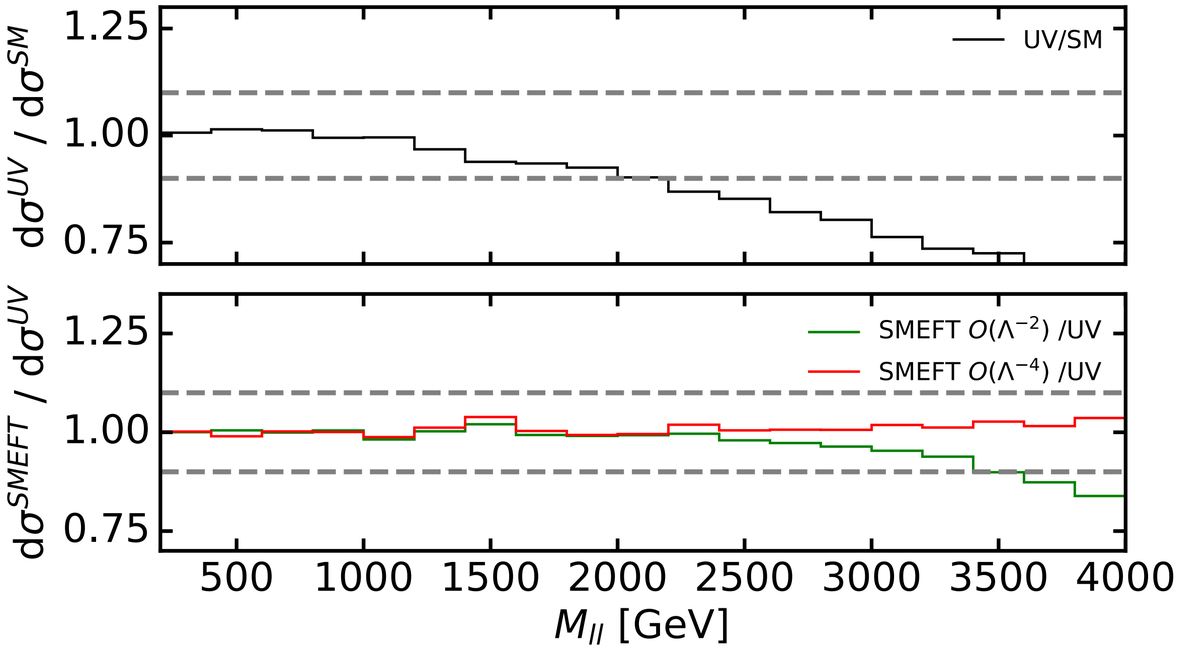
Selection test: ✓

➡ Included in PDF fit

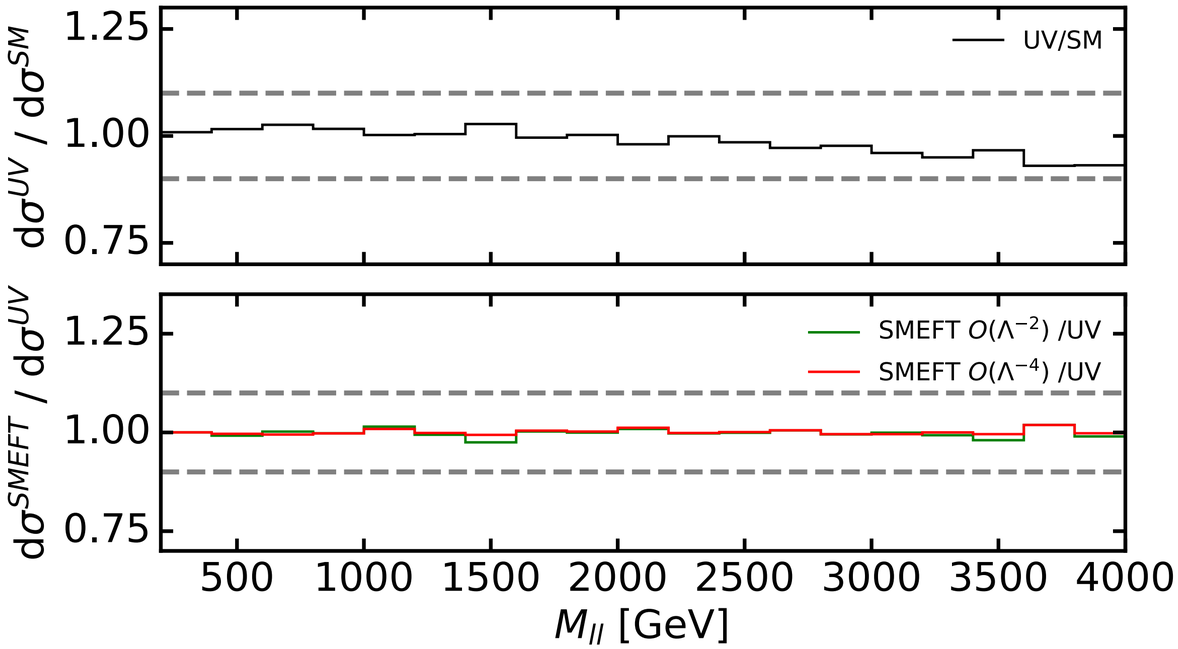
PDFs contaminated

New physics scenarios: Z'

$M_{Z'} = 14.5$ TeV

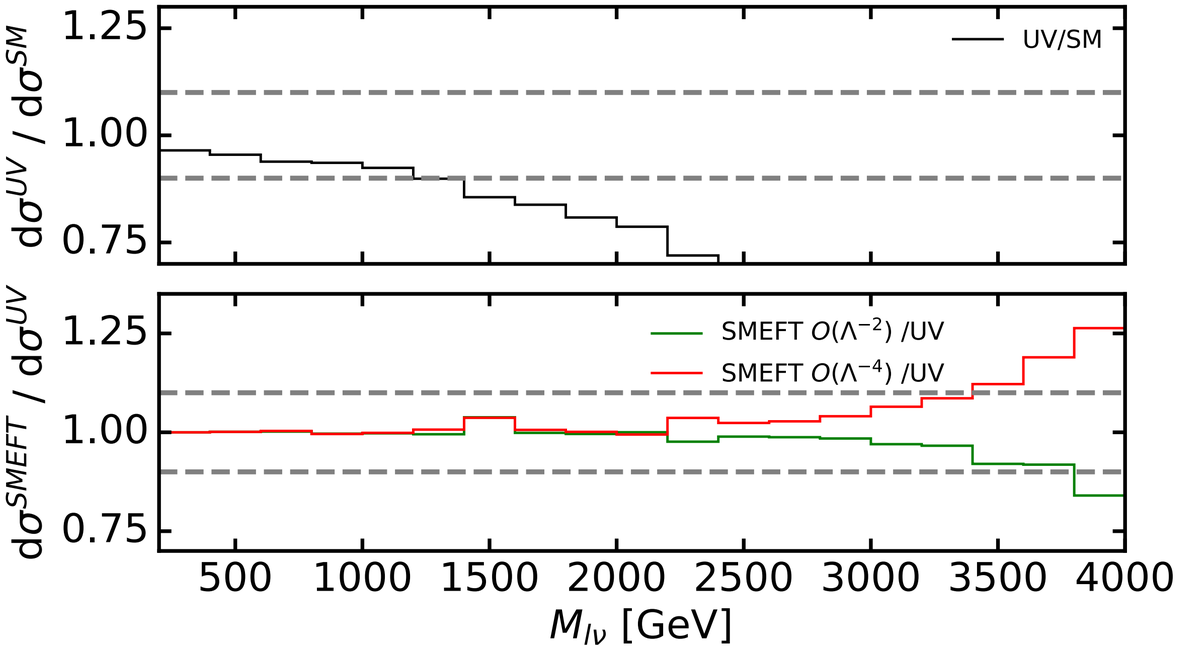


$M_{Z'} = 32.5$ TeV

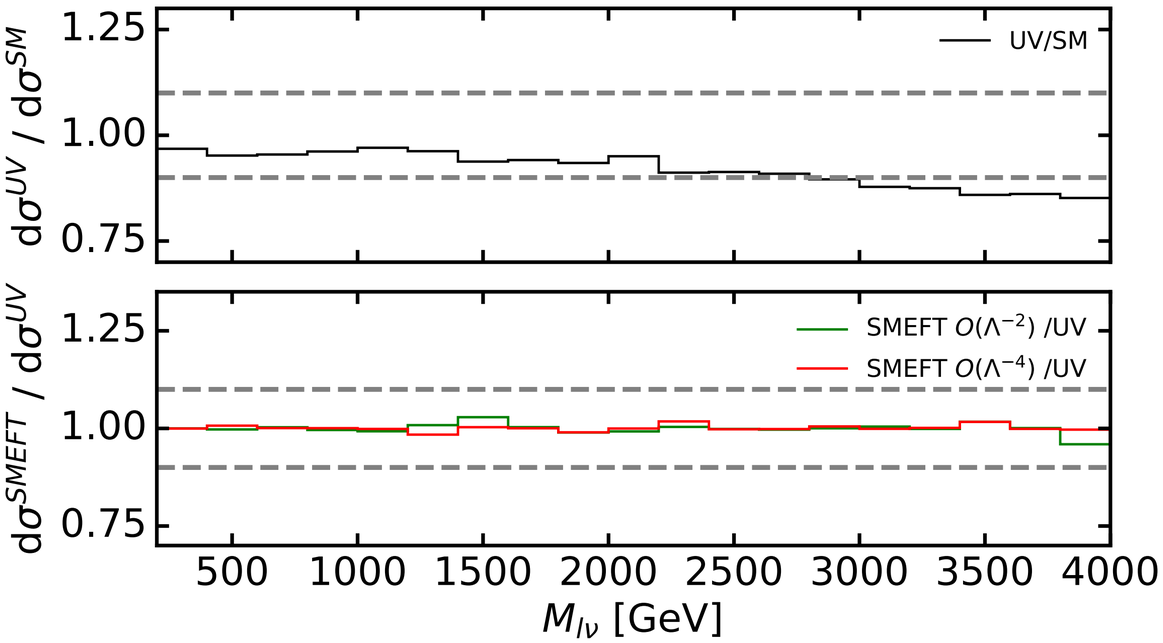


New physics scenarios: W'

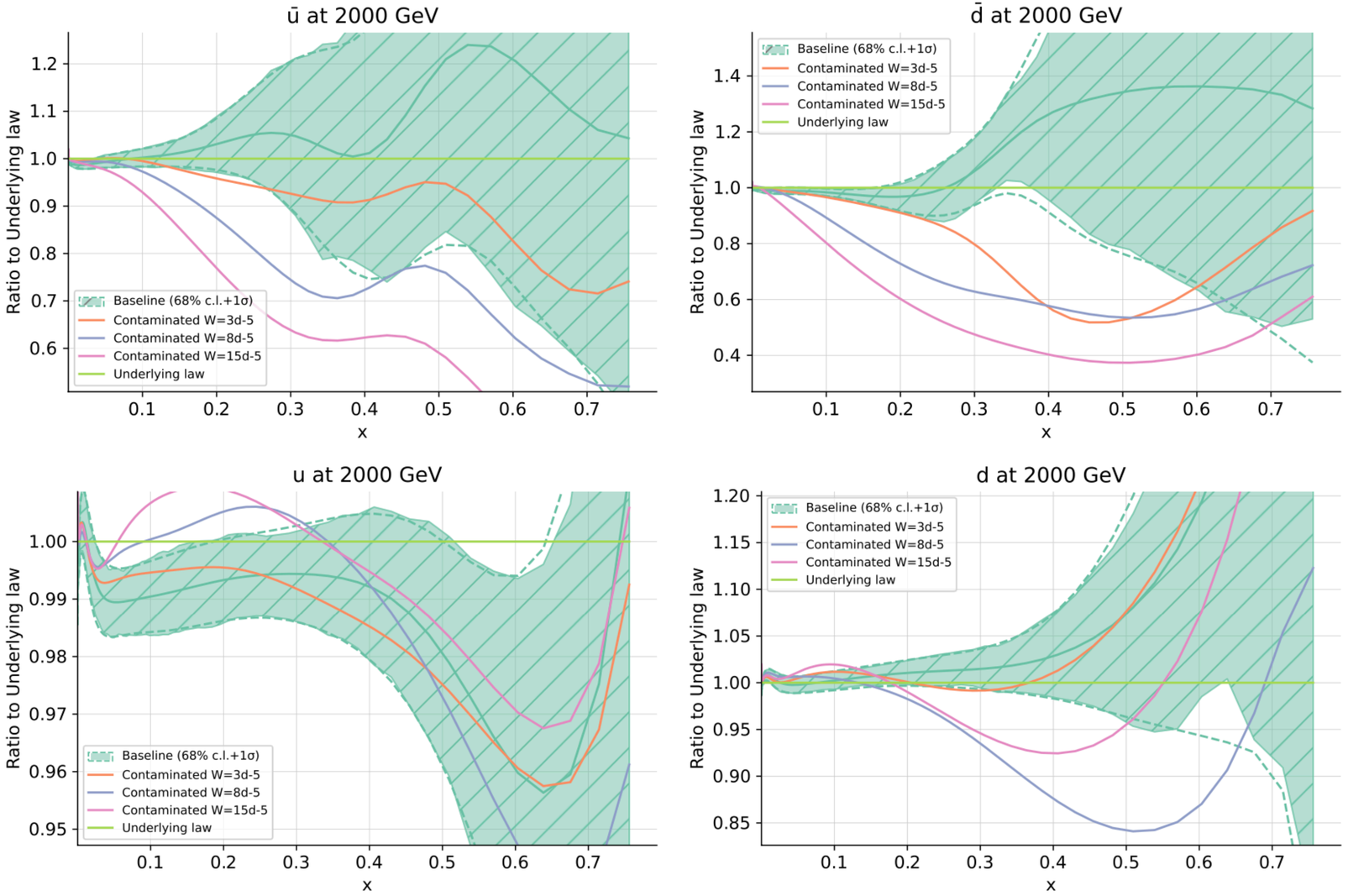
$M_{W'} = 10$ TeV



$M_{W'} = 22.5$ 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