Disentangling PDFs and SMEFT in global fits

[PBSP, 2307.10370, JHEP]

[Hammou et Ubiali, 2410.00963]

[PBSP, forthcoming]

European Research Council

Established by the European Commission

Funded by
the European Union

PBSP

Elie Hammou

Seminar, Durham, May 2025

Talk overview

1. Background of SMEFT and PDFs
2. Risk assessment: absorbing signs of new physics in PDFs [PBSP, 2307.10370, JHEP]
3. Solution 1: Bridging “blind spot” in dataset [Hammou et Ubiali, 2410.00963]
4. Solution 2: Simultaneous fit of PDFs and SMEFT [PBSP, forthcoming]
5. Conclusions and future prospects

Background on SMEFT and PDFs

Physics beyond the Standard Model?

Limits and unsolved puzzles: motivation for new physics

Motivation for BSM physics:

- Dark matter
- Matter/anti-matter asymmetry
- Flavour structure and anomalies
- CP problem
- Hierarchy problem…

A fair amount of questions

Extension of the Lagrangian:

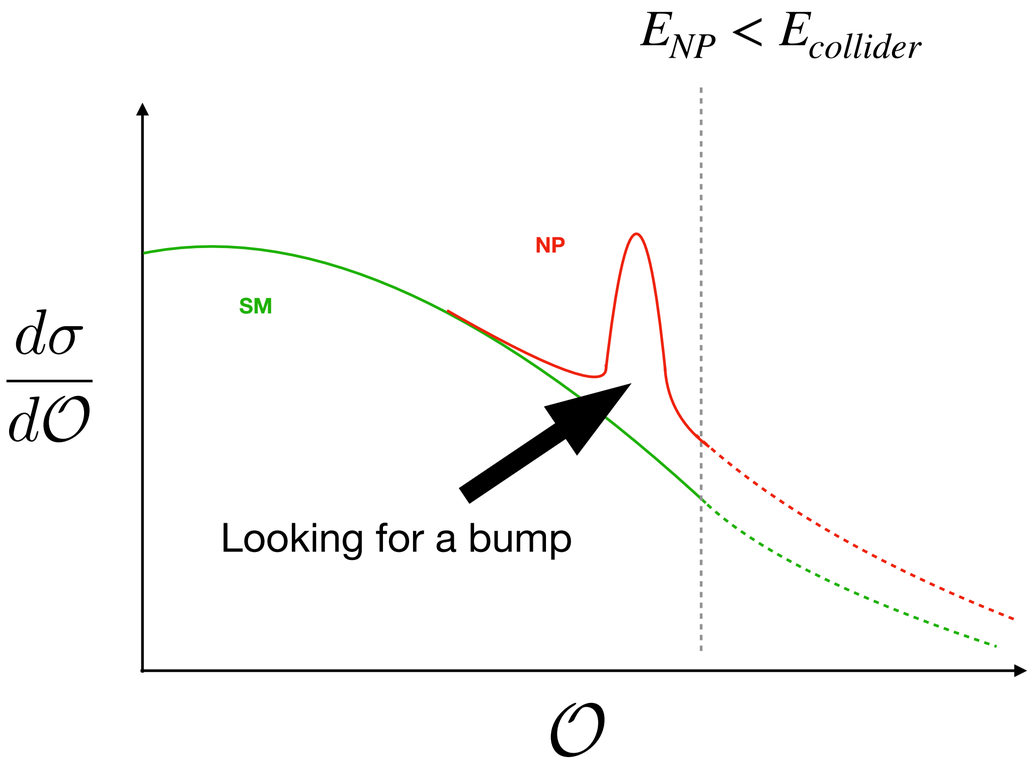
- New gauge symmetry?
- Right-handed neutrino?
- More Higgs?
- SUSY?
- ALPs?…

A lot of possible models

New physics searches

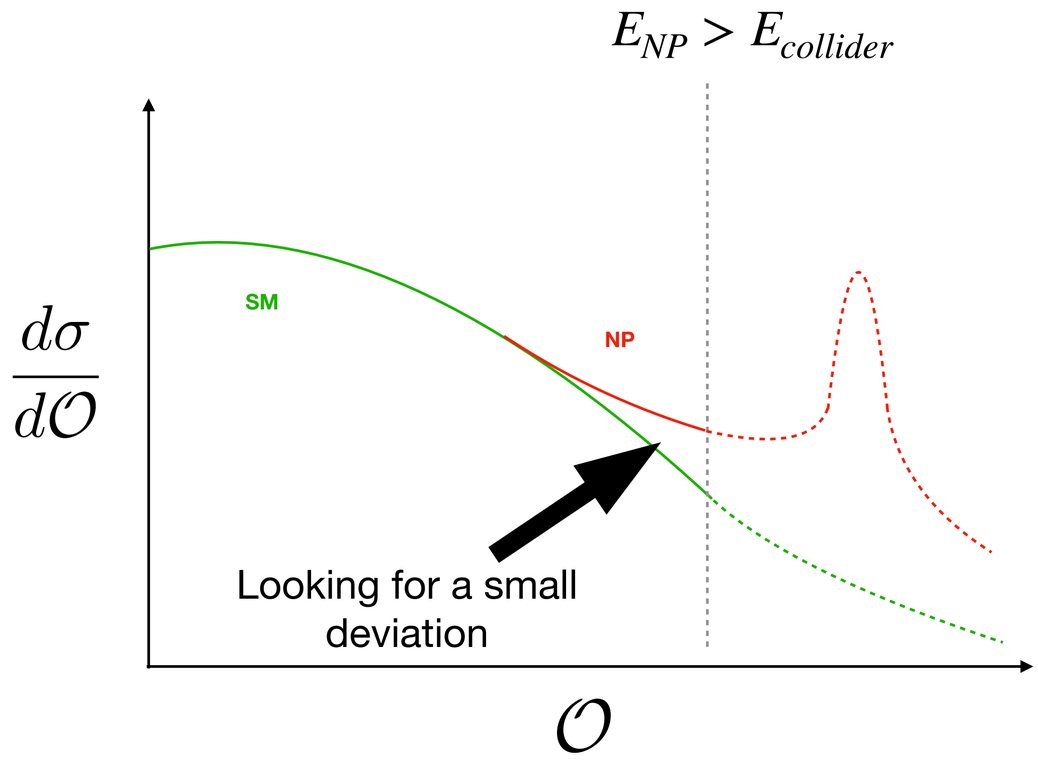
Looking toward higher energy scales and indirect searches

Direct searches



No luck so far…

Indirect searches



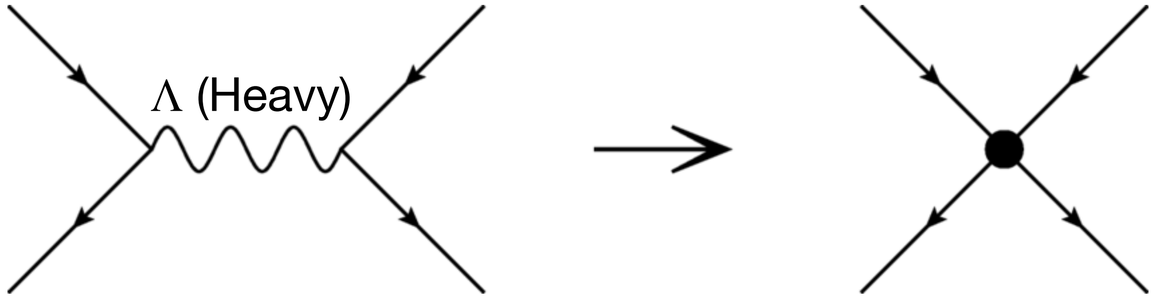
Requires precision

Indirect searches and Effective Field Theories

The Standard Model EFT (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: $\frac{c_i}{\Lambda^2}$

The SMEFT

Dimension-6 operators

Operator basis

2499 operators
[Grzadkowski et al, arXiv:1008.4884]

Reduced with symmetry assumptions:

- e.g. baryon number conservation : 59 operators

Presented in the Warsaw basis

Corrections

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

$$\sigma^{\text{SMEFT}} = \sigma^{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2} |\mathscr{A}^{SM}\mathscr{A}^{\mathcal{O}_i}| + \sum_{i,j} \frac{c_i c_j}{\Lambda^4} |\mathscr{A}^{\mathcal{O}_i}\mathscr{A}^{\mathcal{O}_j}|$$

Linear

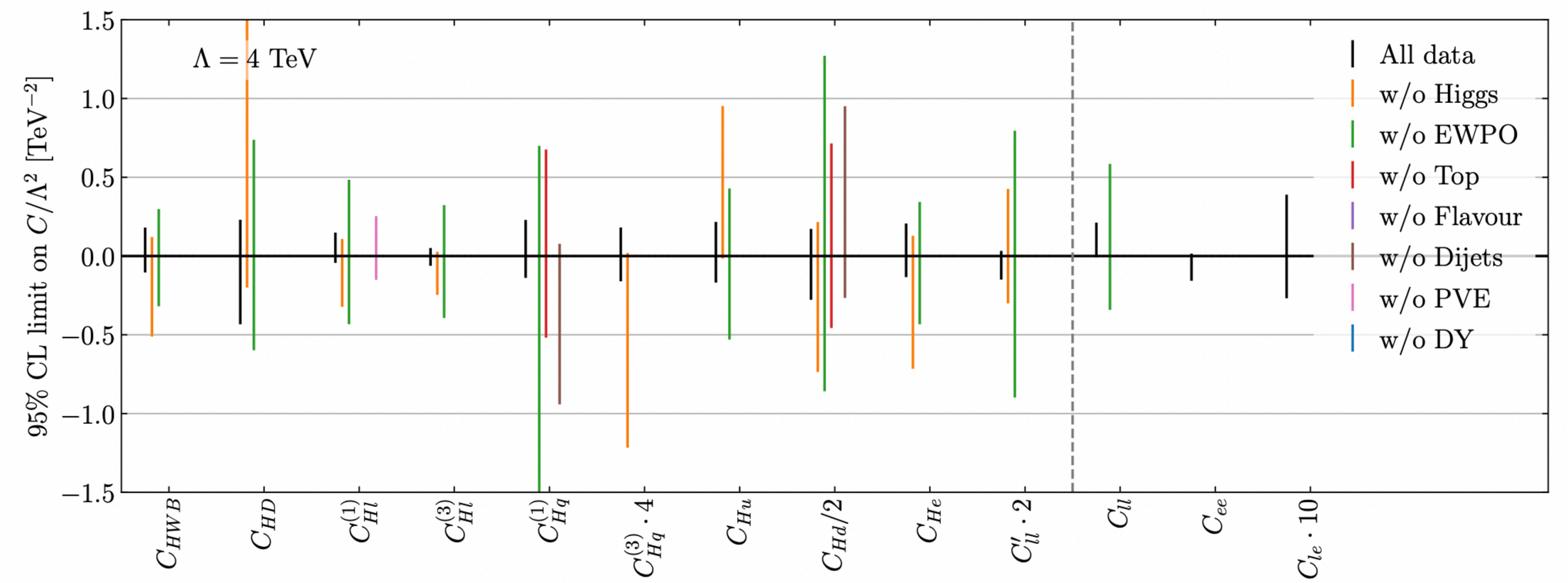
$$\mathcal{O}\left(\frac{c_i}{\Lambda^2}\right)$$

Quadratic

$$\mathcal{O}\left(\frac{c_i c_j}{\Lambda^4}\right)$$

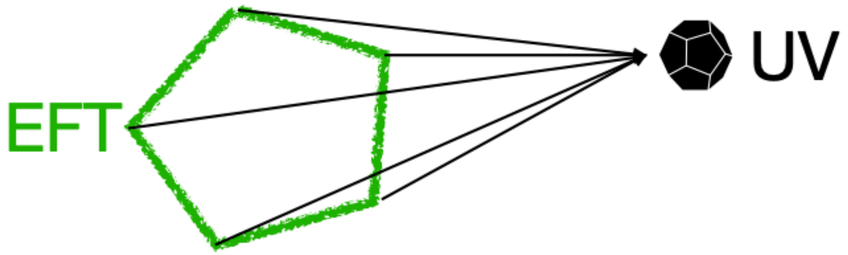
SMEFT fit from data

[SMEFiT, 2302.06660]



Can fit $\{\frac{c_i}{\Lambda^2}\}$:

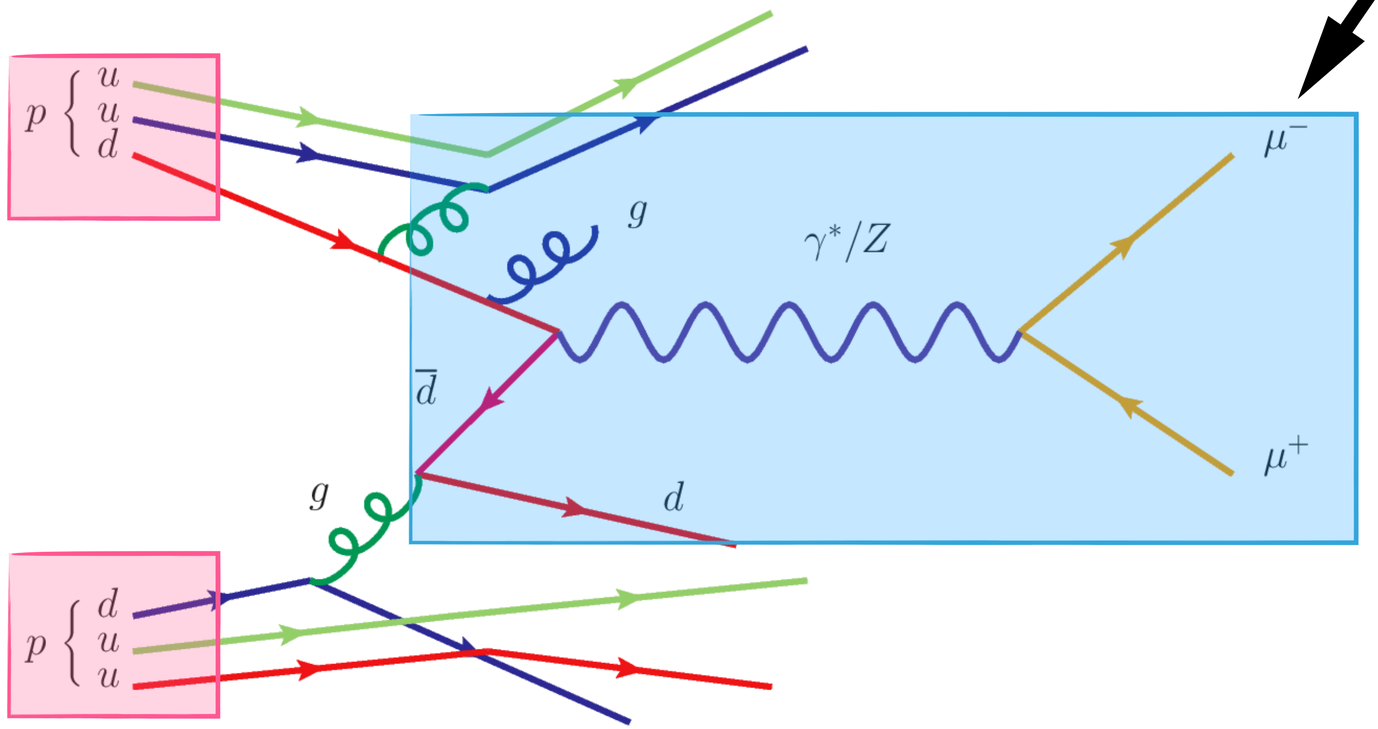
➡ can then be matched to a UV model



Hadron colliders and PDFs

Collinear factorization theorem

SMEFT corrections



Parton Distribution Functions

Hard Scattering: Perturbative QCD + EW

$$d\sigma^{pp\to ab} = \sum_{i,j} f_i \otimes f_j \otimes d\hat{\sigma}^{ij\to ab} + \ldots$$

PDFs overview

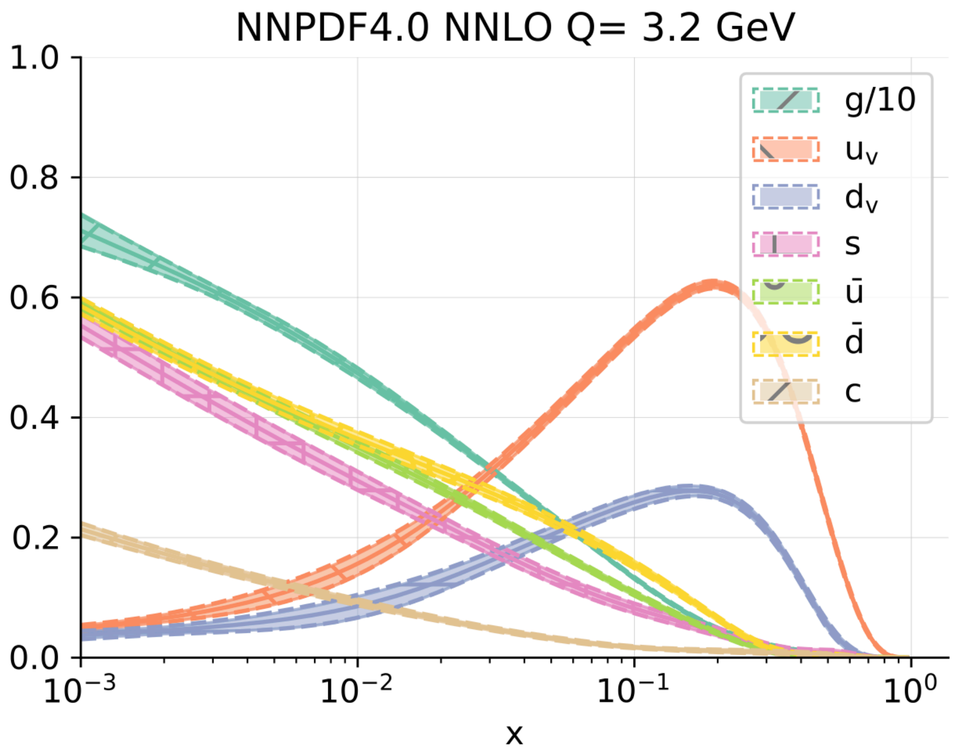
Hadron collider observable: $\sigma = f_1 \otimes f_2 \otimes \hat{\sigma}$

PDFs in a nutshell:

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

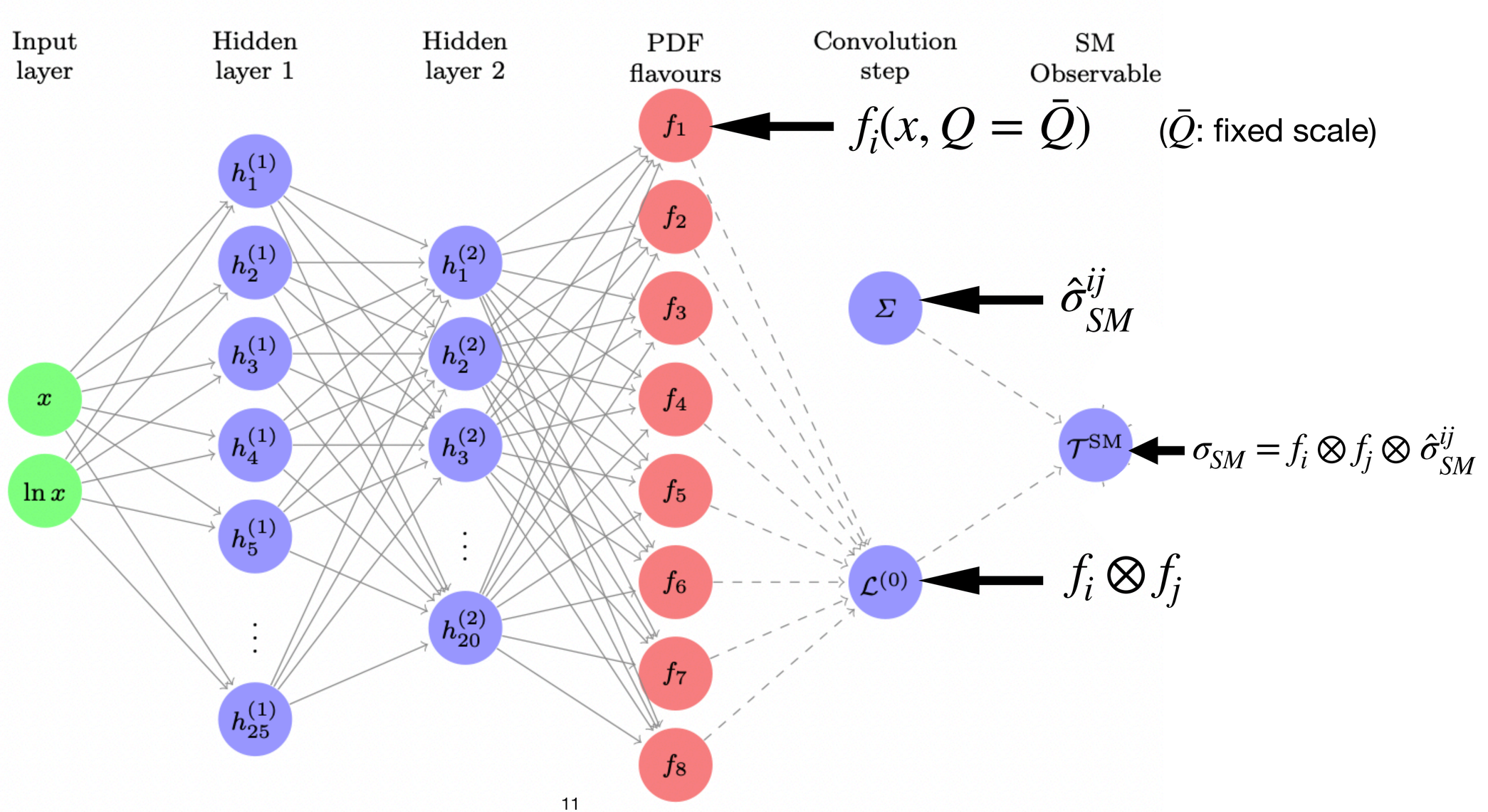
➡ **Fitted from data**

Using NNPDF methodology



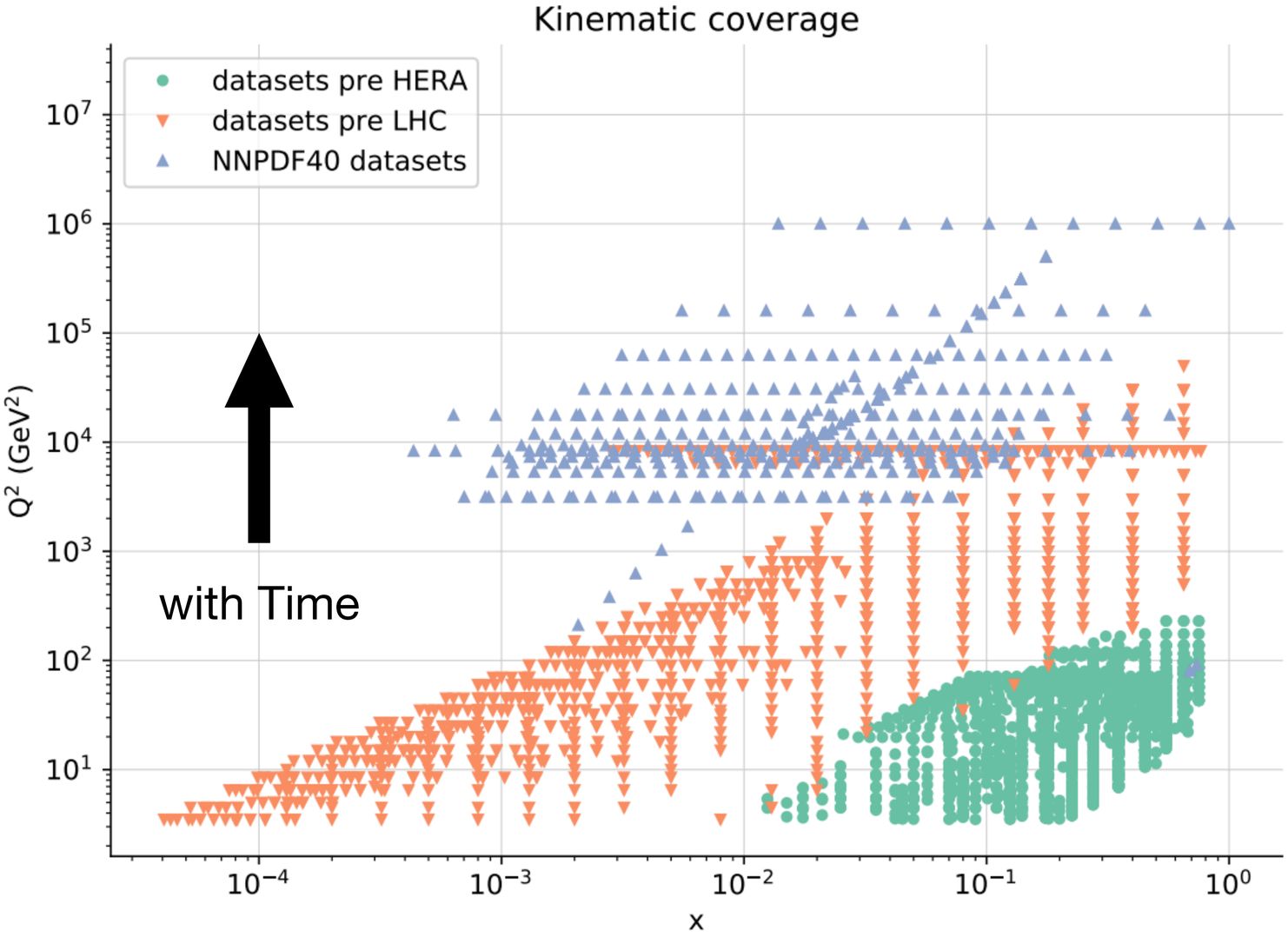
[Ball et al., NNPDF4.0, 2109.02653]

NNPDF fitting methodology



Bjorken-x
(Momentum fraction)

NNPDF dataset



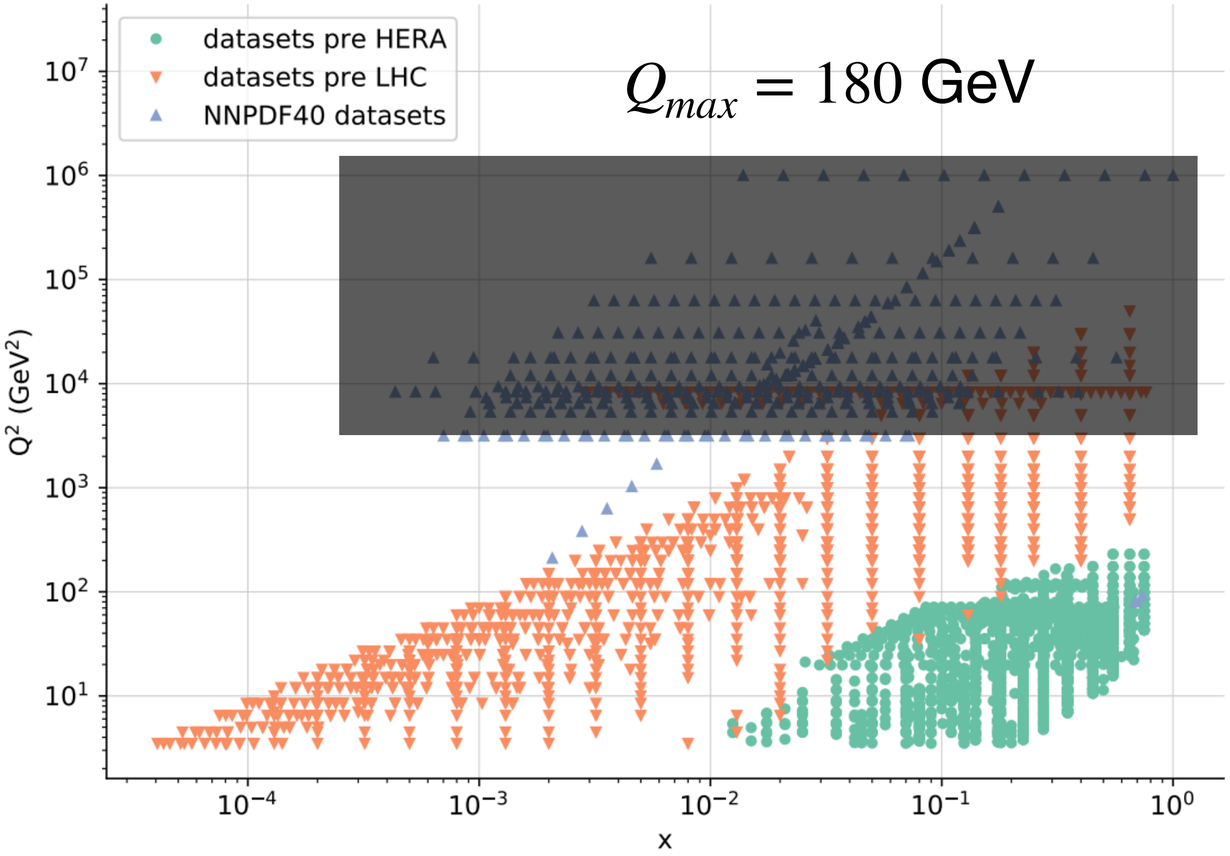
Evolution of the dataset through time:

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

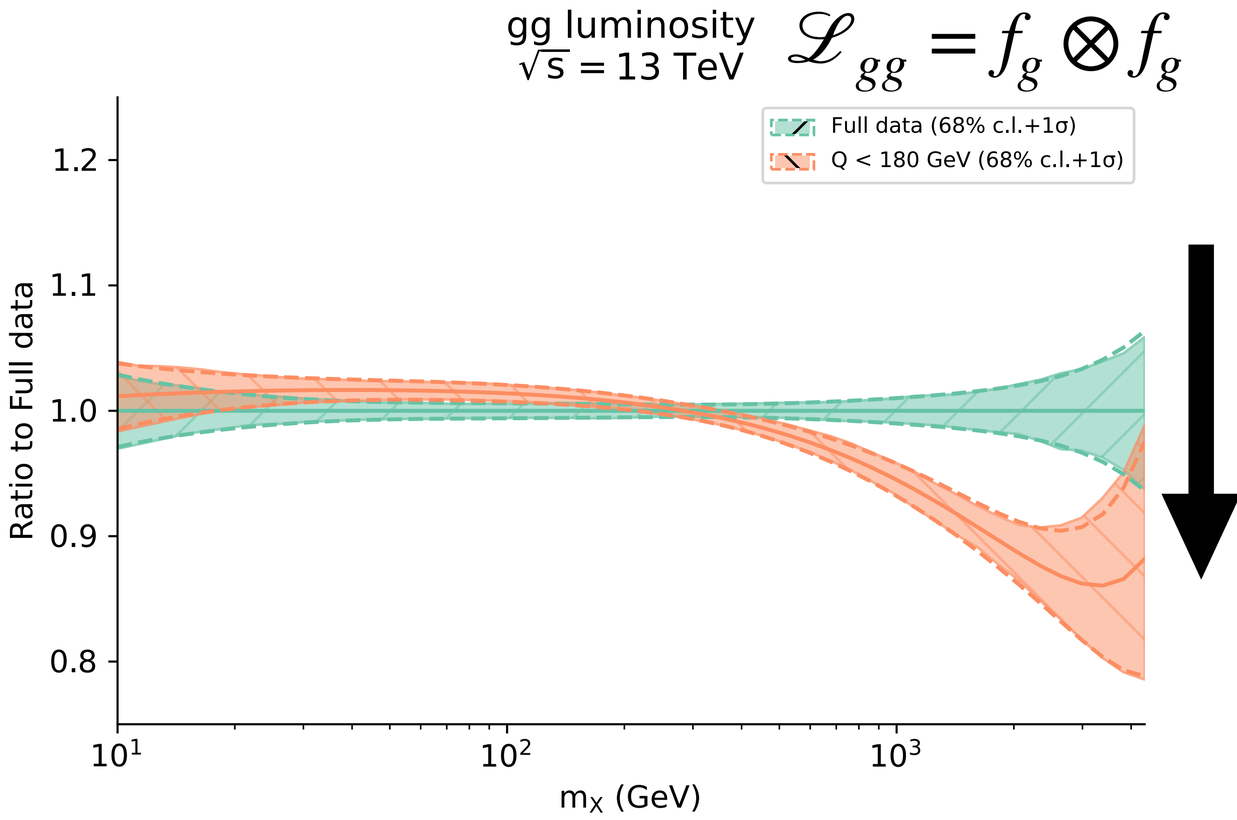
[Ball et al., NNPDF4.0, 2109.02653]

Discrepancy between low and high-energy data fits

Kinematic coverage



Impact of energy cut



Risk assessment

Absorbing signs of new physics in the PDFs

[PBSP, 2307.10370, JHEP]

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 scenarios: W'

From UV to the SMEFT

Heavy triplet under $SU(2)_L$: W'

$$\mathscr{L}_{UV}^{W'} = \mathscr{L}_{SM} - \frac{1}{4} W'^{a}_{\mu\nu} W'^{a,\mu\nu} + \frac{1}{2} M^2_{W'} W'^{a}_{\mu} W'^{a,\mu} - g_{W'} W'^{a,\mu} \sum_{f_L} \bar{f}_L T^a \gamma^\mu f_L - g_{W'} (W'^{a,\mu} \varphi^\dagger T^a i D_\mu \varphi + \text{h.c.})$$

➡ Creates two charged particles: W'^{+} / W'^{-} and a neutral one: W'_3

Matching to the SMEFT:

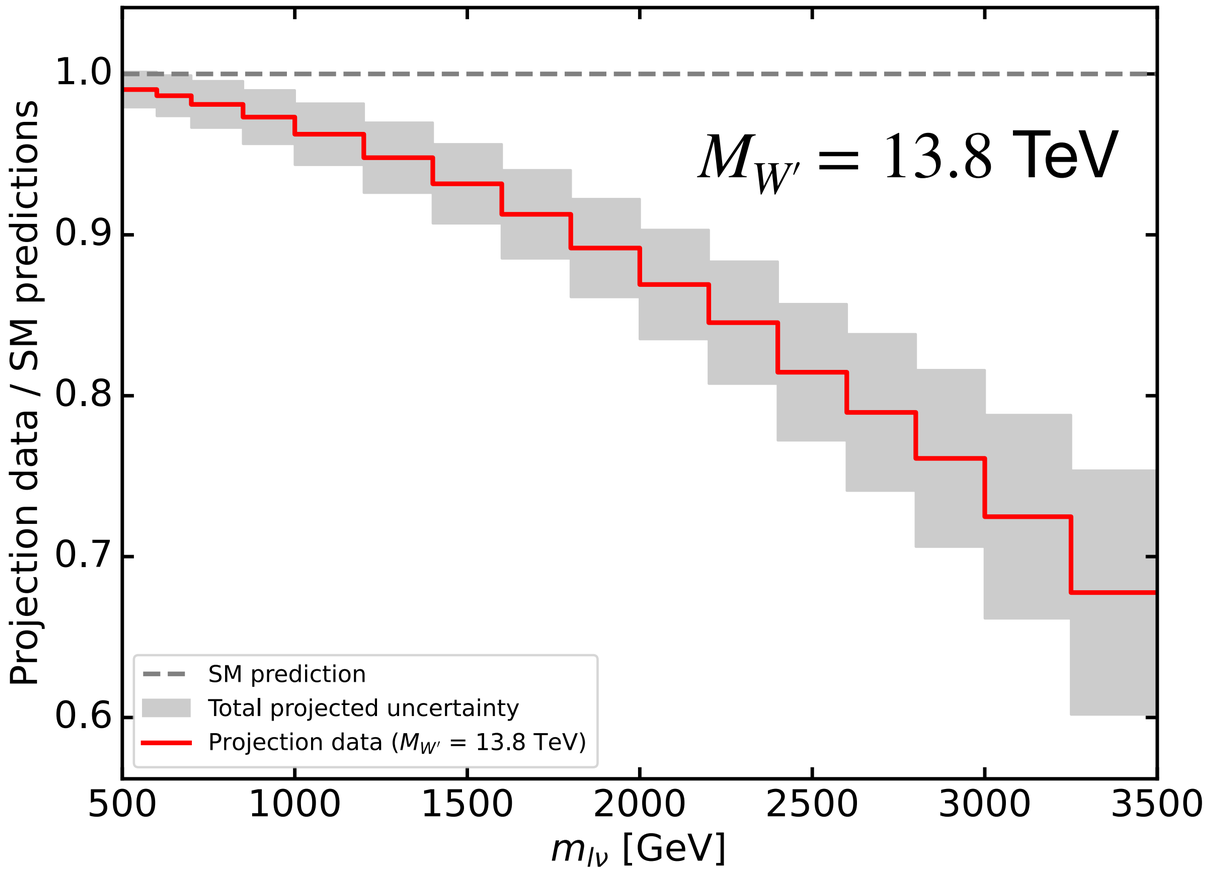
$$\mathscr{L}_{SMEFT}^{W'} = \mathscr{L}_{SM} - \frac{g^2_{W'}}{2M^2_{W'}} J_L^{a,\mu} J^a_{L,\mu} \qquad J_L^{a,\mu} = \sum_{f_L} \bar{f}_L T^a \gamma^\mu f_L$$

→ $$\mathscr{L}_{SMEFT}^{W'} = \mathscr{L}_{SM} - \frac{g^2 \hat{W}}{2m^2_W} J_L^{a,\mu} J^a_{L,\mu} \qquad \hat{W} = \frac{g^2_{W'}}{g^2} \frac{m^2_W}{M^2_{W'}} \propto \frac{c}{\Lambda^2}$$

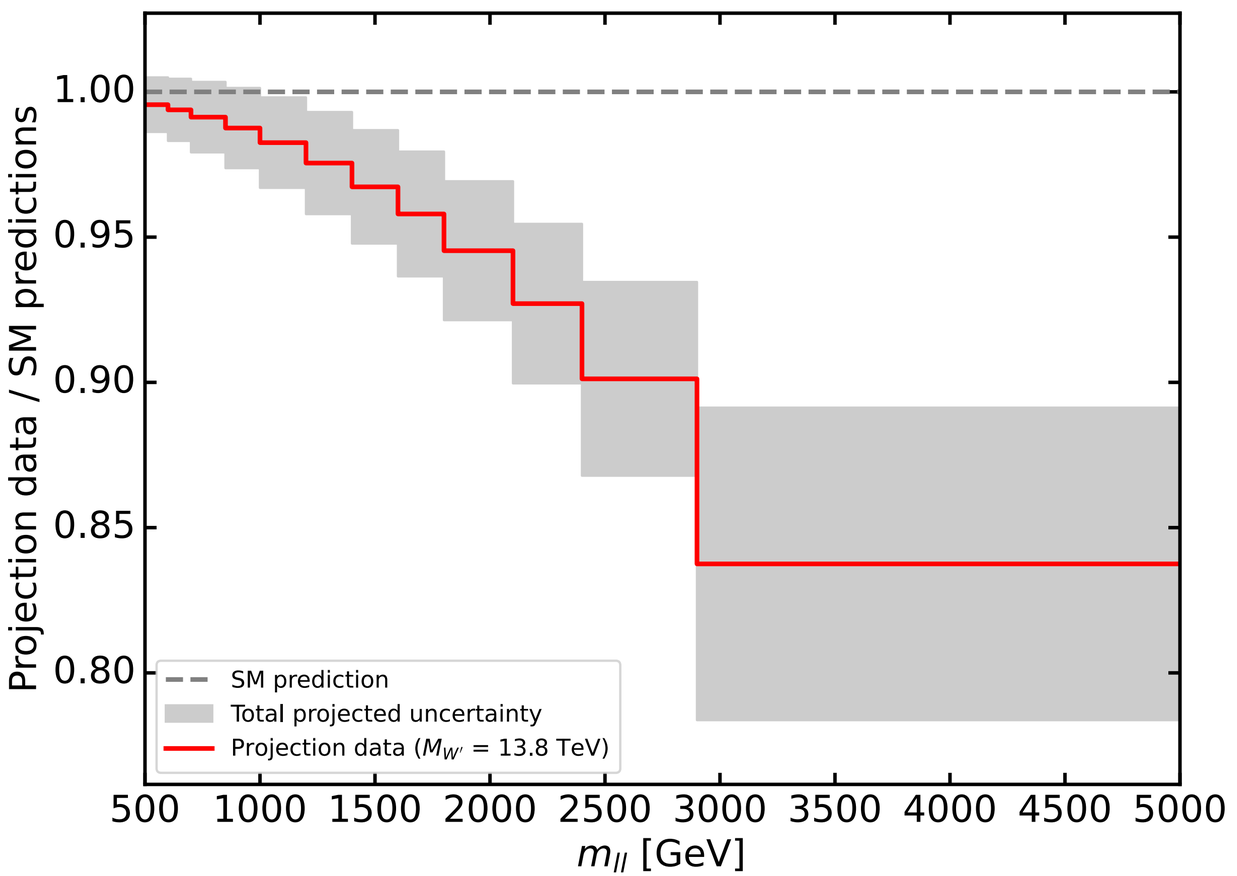
New physics parameter

Impact on HL-LHC HMDY projections: W'

Charged current HMDY



Neutral current HMDY



New physics scenarios: Z'

From UV to the SMEFT

Heavy boson charged under $U(1)_Y$: Z'

$$\mathscr{L}^{Z'}_{\text{UV}} = \mathscr{L}_{\text{SM}} - \frac{1}{4} Z'_{\mu\nu} Z'^{\mu\nu} + \frac{1}{2} M^2_{Z'} Z'_\mu Z'^\mu - g_{Z'} Z'_\mu \sum_f Y_f \bar{f} \gamma^\mu f - Y_\varphi g_{Z'} (Z'_\mu \varphi^\dagger i D^\mu \varphi + \text{h.c.})$$

➡ Creates one neutral particle

Matching to the SMEFT:

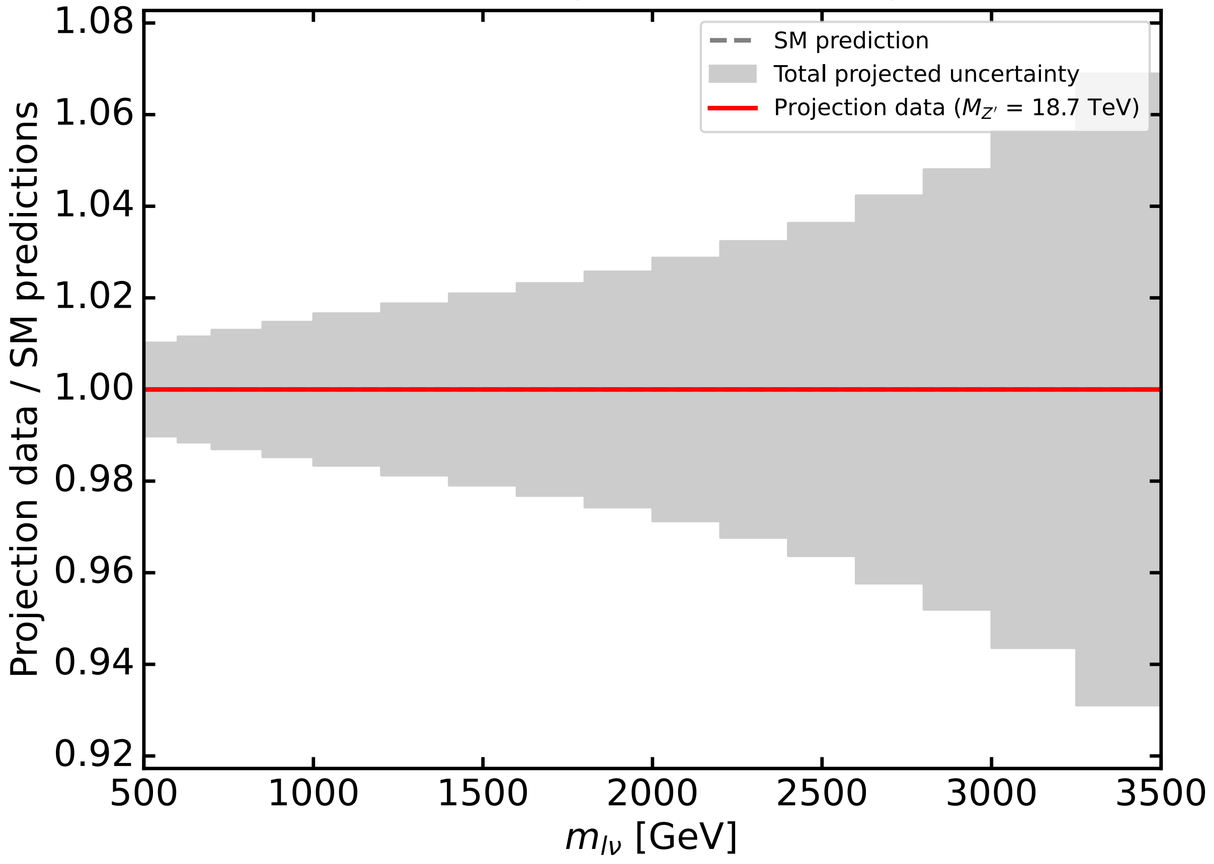
$$\mathscr{L}^{Z'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2_{W'}}{2M^2_{W'}} J^{a,\mu}_Y J^a_{Y,\mu} \qquad J^{a,\mu}_L = \sum_{f_L} Y_f \bar{f}_L \gamma^\mu f_Y$$

→ $$\mathscr{L}^{Z'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2 \hat{Y}}{2m^2_W} J^{a,\mu}_Y J^a_{Y,\mu} \qquad \hat{Y} = \frac{g^2_{Z'}}{g^2} \frac{m^2_W}{M^2_{Z'}} \propto \frac{c}{\Lambda^2}$$

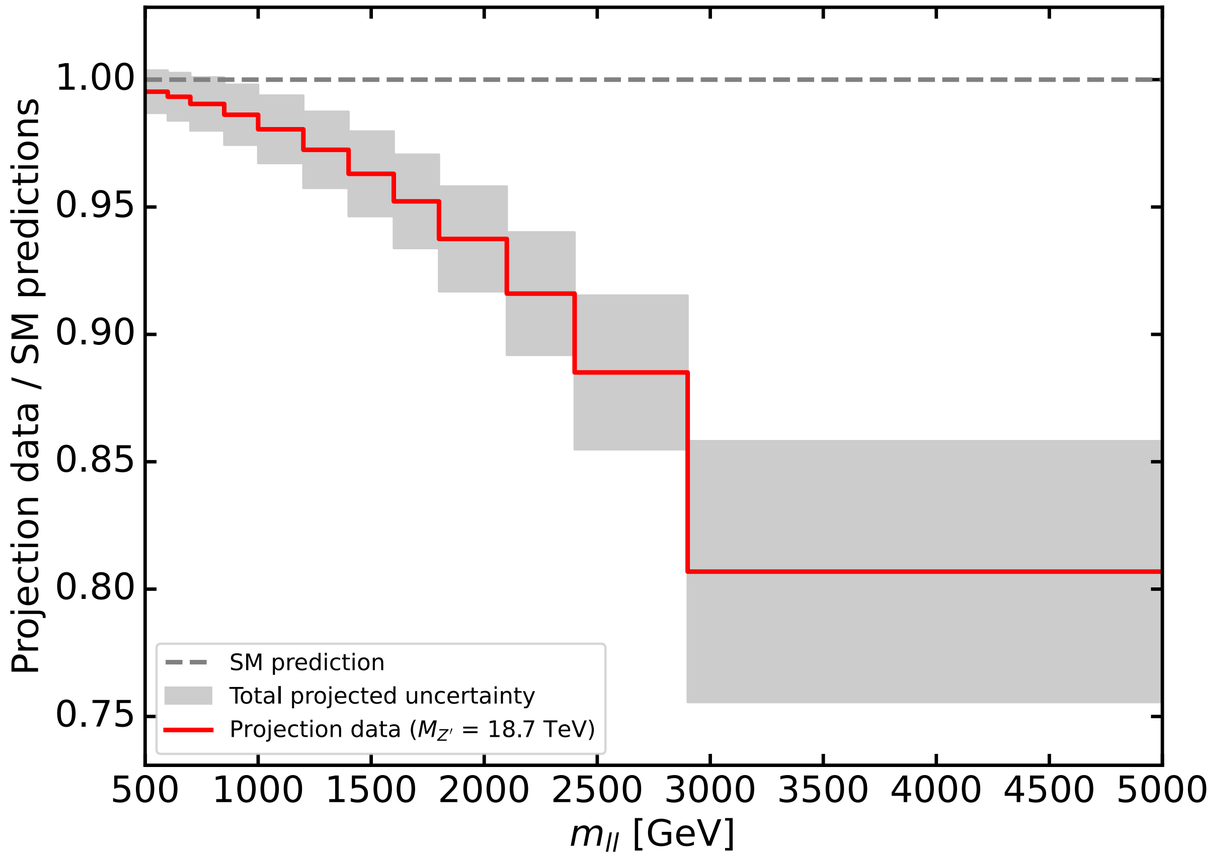
New physics parameter

Impact on HL-LHC HMDY projections: Z'

Charged current HMDY



Neutral current HMDY



PDF fitting: selection test

Do our BSM datasets have a good χ^2?

Z'

Selection test: ✗

➡ Excluded from PDF fit

No impact on PDFs

W'

Selection test: ✓

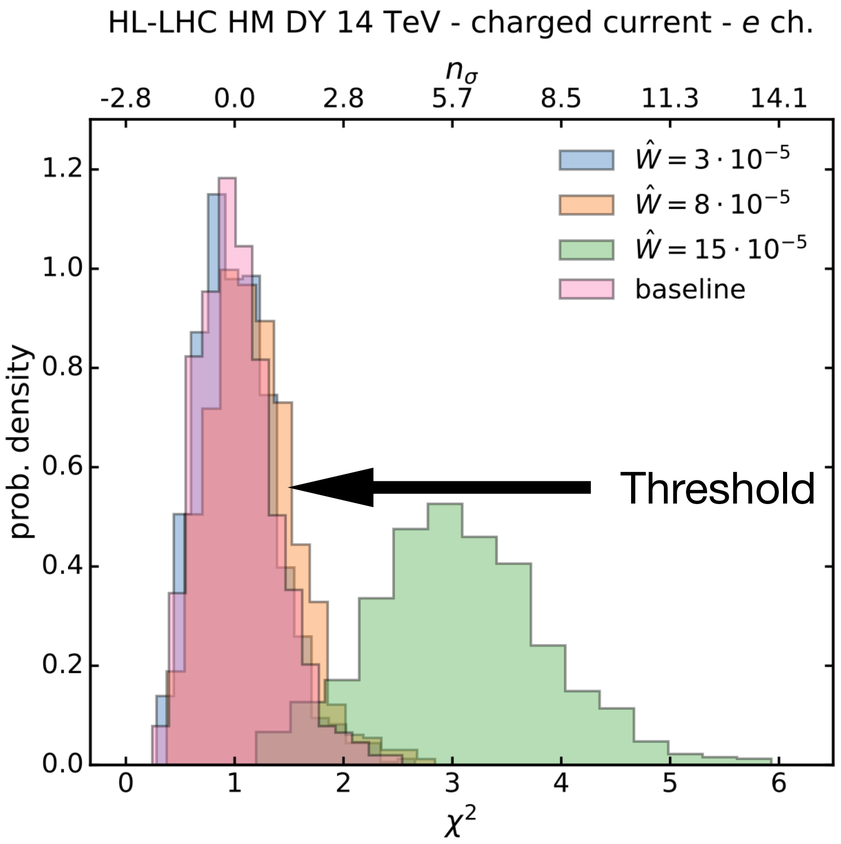
➡ Included in PDF fit

Absorbed by PDFs

Impact of contamination: the PDFs

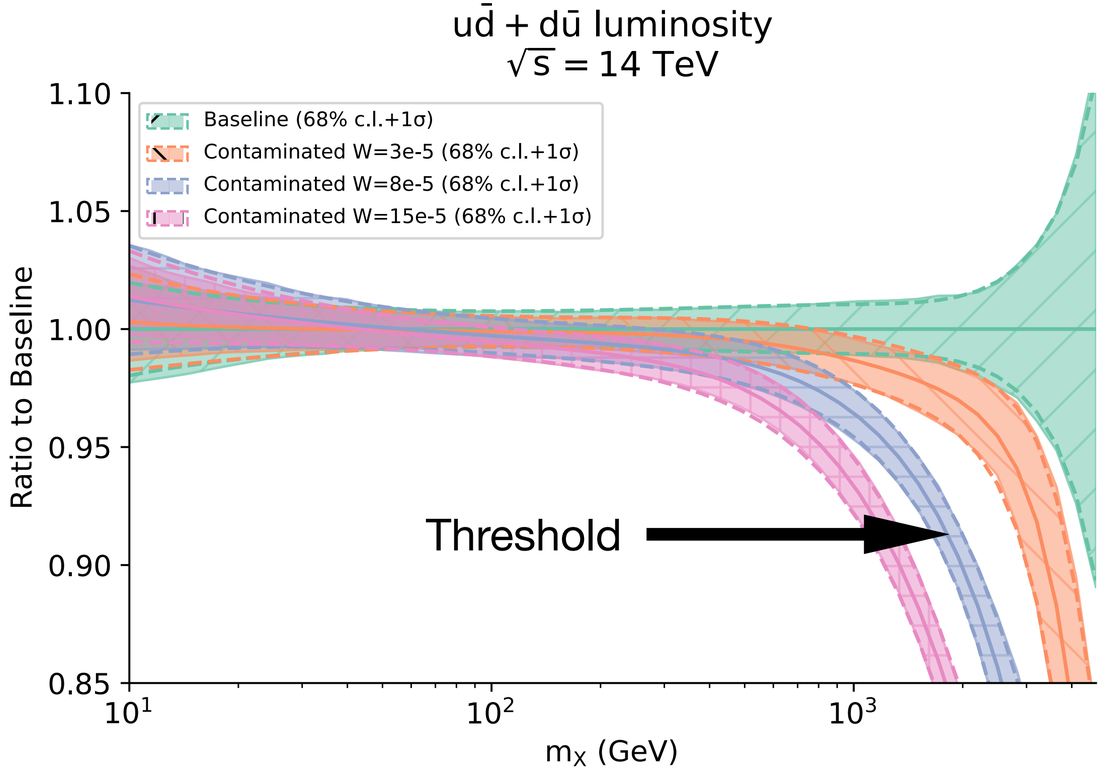
Comparison between contaminated and Baseline PDFs

Fit quality



$M_{W'} = 13.8$ TeV

Impact on PDFs

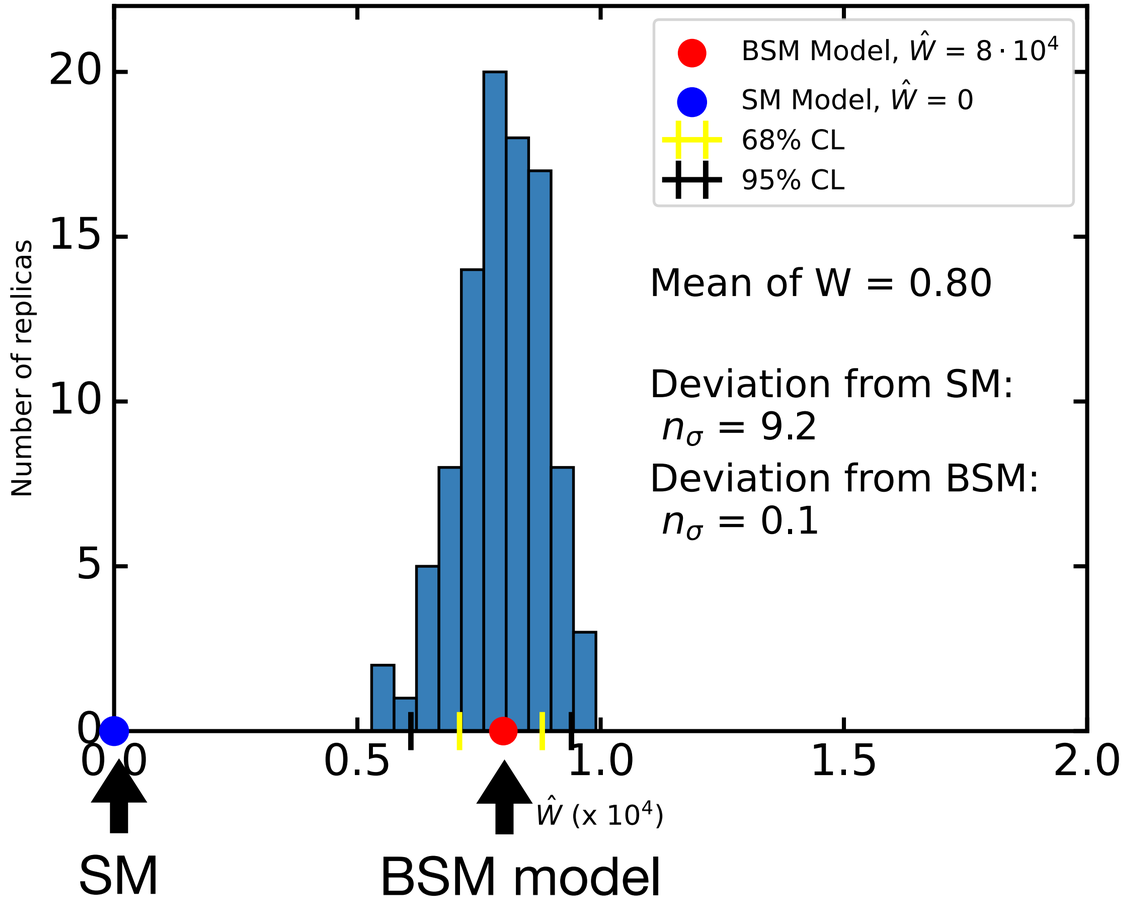


$$\sigma_{obs} = \hat{\sigma}_{BSM} \otimes \mathcal{L}_{SM} \approx \hat{\sigma}_{SM} \otimes \mathcal{L}_{BSM-biased}$$

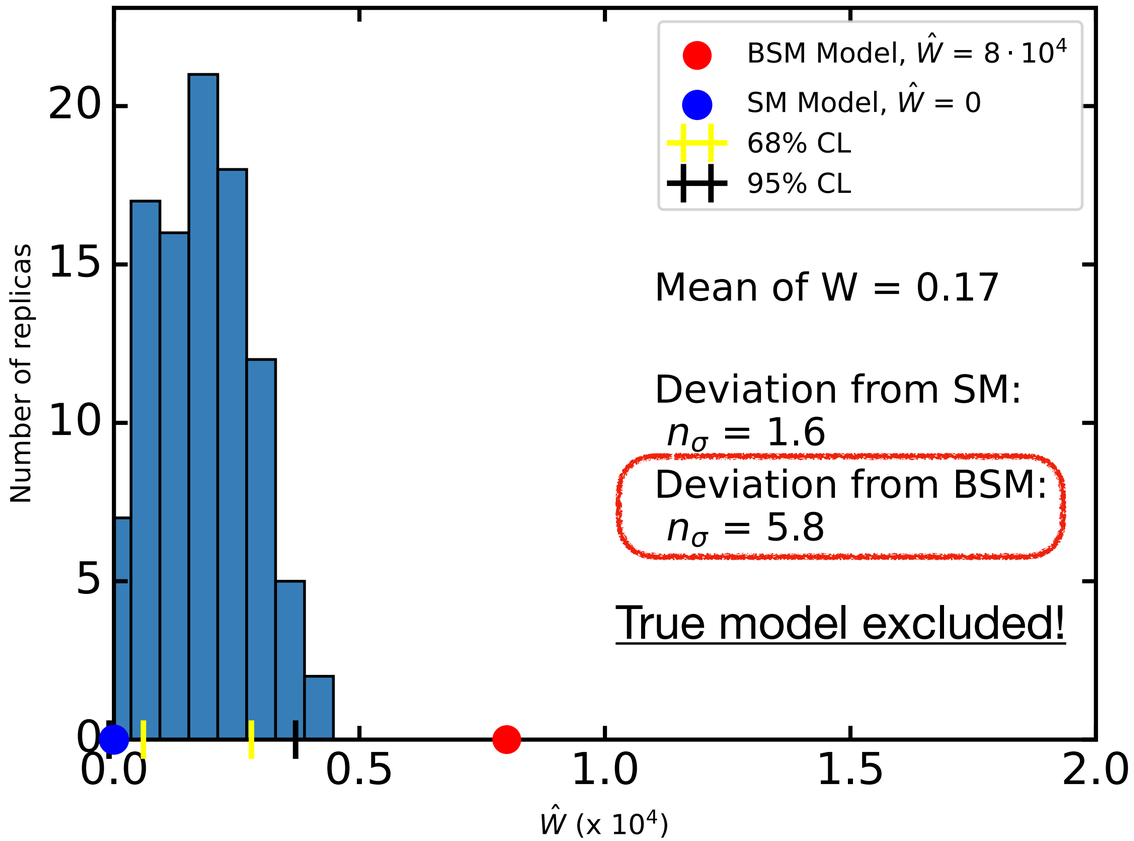
Missing new physics

Impact of the NP absorption in PDFs on SMEFT fits

SMEFT Fit with true PDF



SMEFT Fit with BSM PDF



PDF contamination: summary

- BSM data in PDF fit:
  - At best: BSM data flagged and excluded
  - At worst: BSM signal absorbed by the PDF

- Consequences of PDF contamination:
  - New physics is hidden (model can be ruled out)

➡ Possible solutions?

First solution

Bridging dataset “blindspot”

[Hammou et Ubiali, 2410.00963, accepted in PRD]

Synergy of high and low-energy data

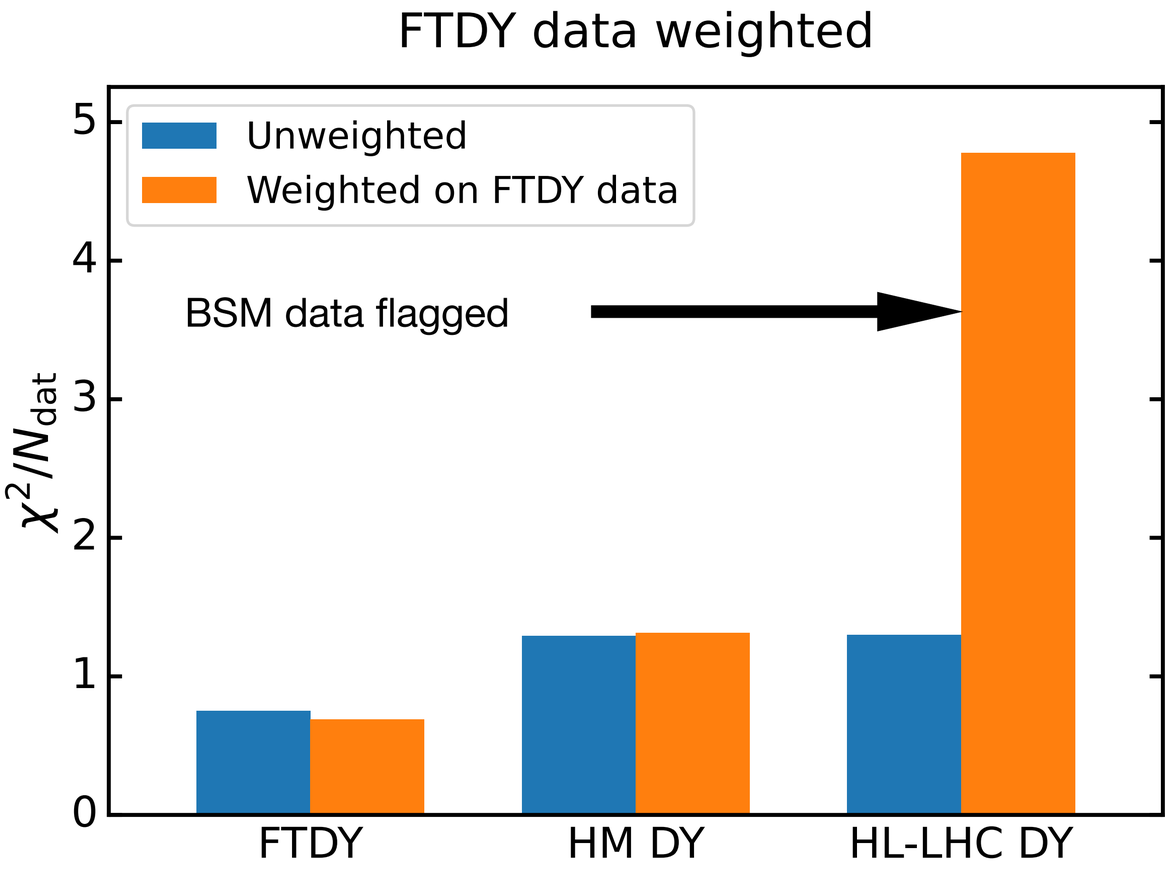
Adding low-energy dataset constraining the large-x region

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 effects



Future low energy observables

Presentation of the future DIS programmes

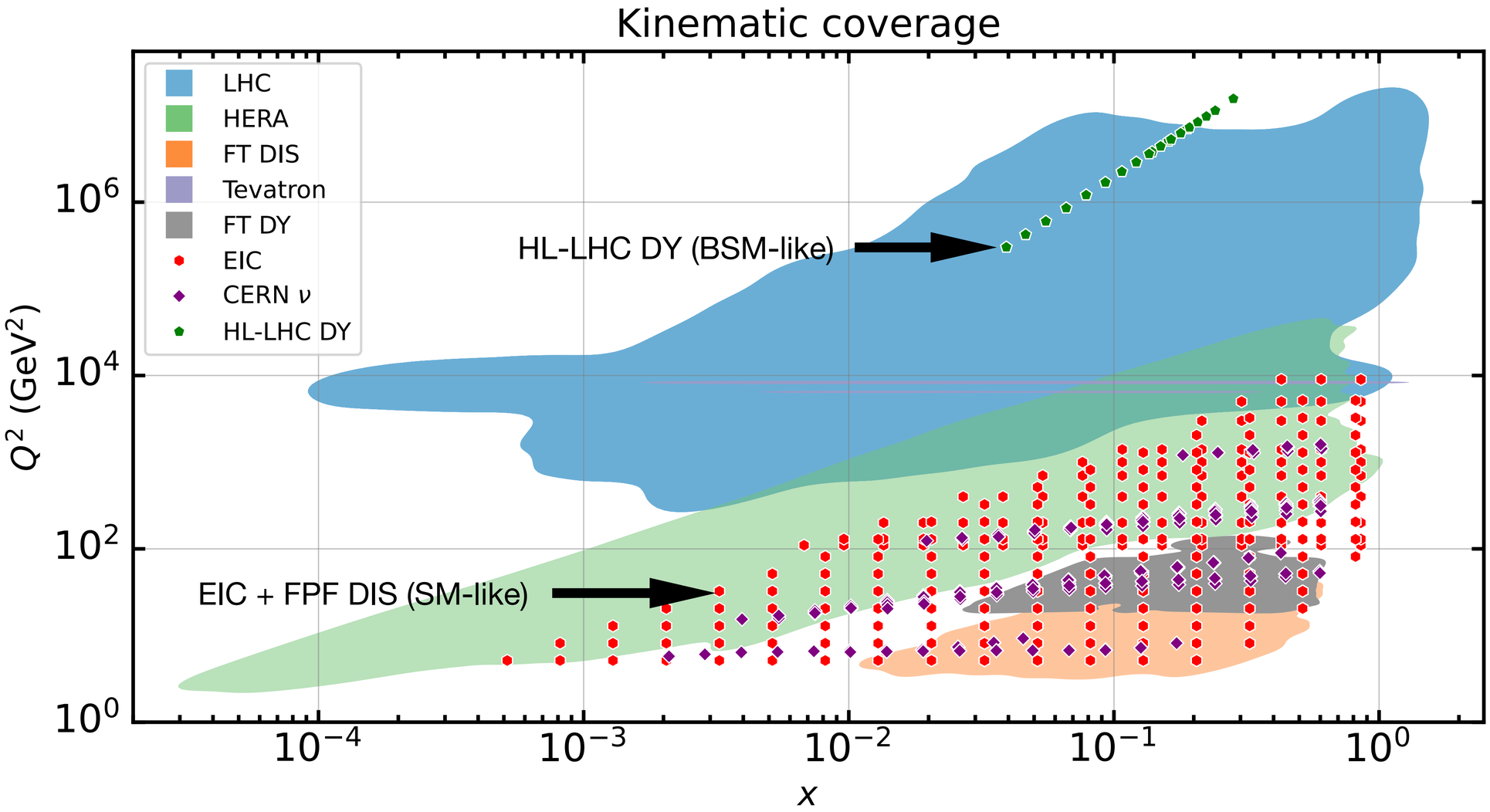
Electron Ion Collider

- e^+/e^- projectiles
- proton, deuteron and heavy ions targets
- Hosted in Brookhaven
- Planned for 2030s
- Probes large-x, low-energy

Forward Physics Facility

- “Neutrino Ion collider” at the LHC
- $\nu/\bar{\nu}$ projectiles from proton beam
- proton, neutron and other nuclear targets
- FASERν and SND@LHC already running
- Proposed expansion for HL-LHC run (FASERν2 , AdvSND, FLArE)
- Probes large-x, low-energy
- Constrain large-x antiquarks

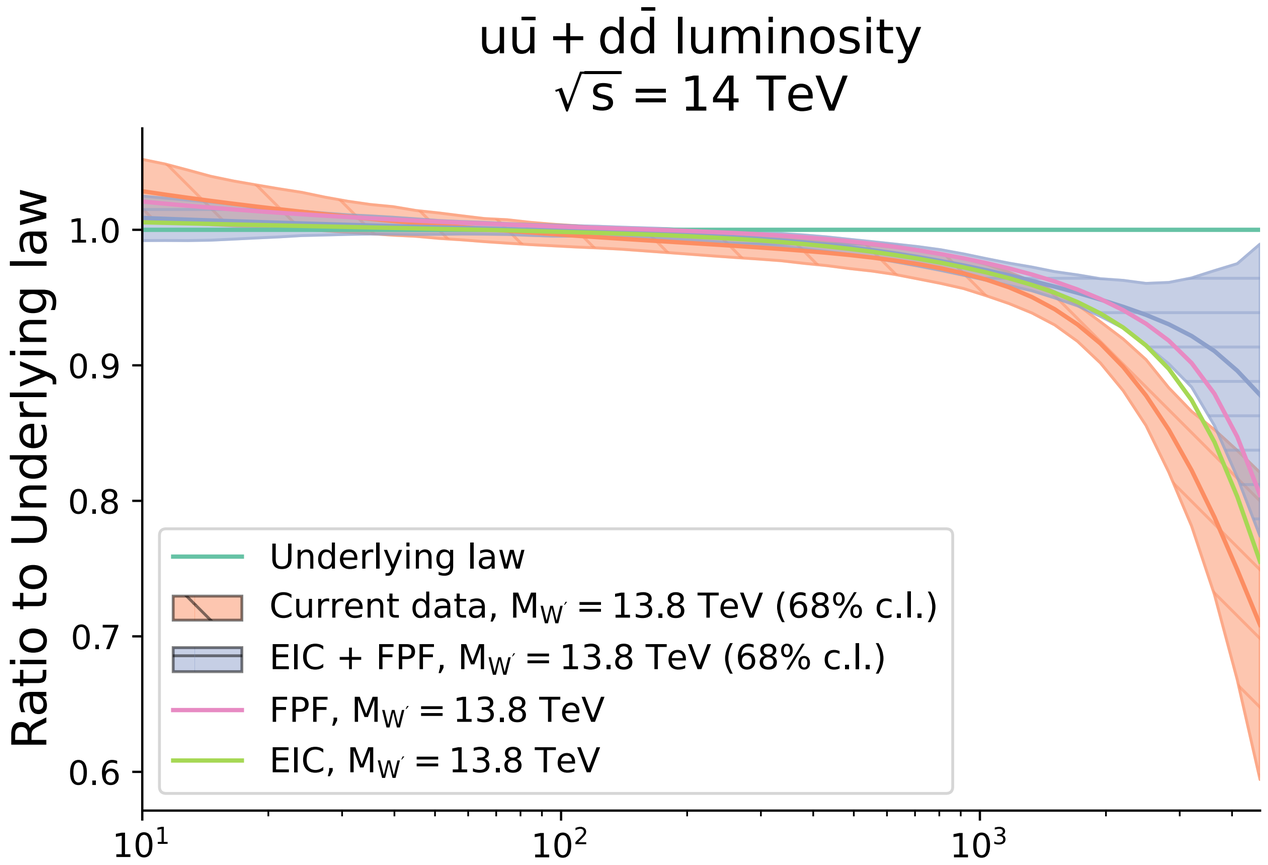
EIC, FPF and HL-LHC kinematics



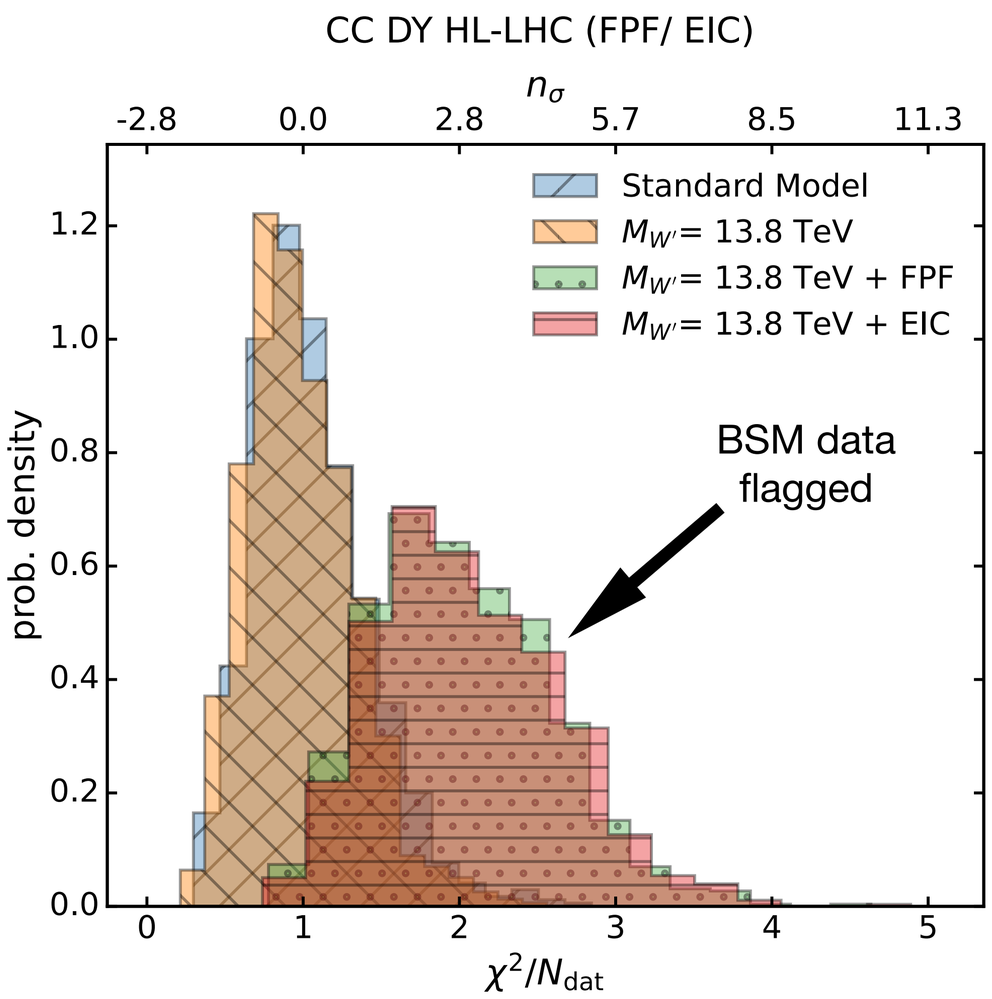
Impact on the PDF contamination

Flagging the BSM data

Pull on PDFs



Impact on fit quality



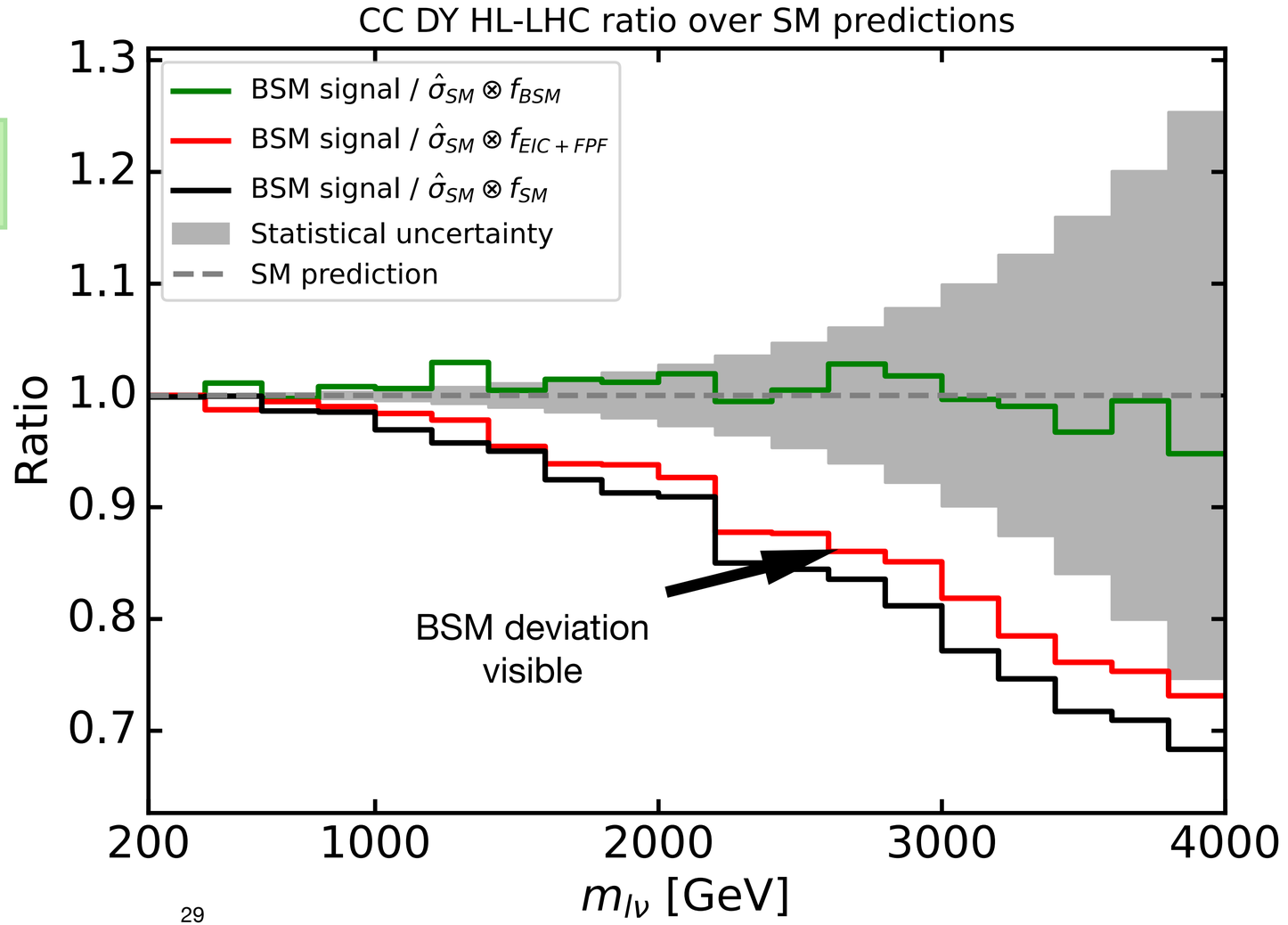
Recovering the signs of new physics

BSM data versus SM theory predictions

$$\hat{\sigma}_{BSM} \otimes \mathcal{L}_{SM} \approx \hat{\sigma}_{SM} \otimes \mathcal{L}_{BSM-biased}$$

$$\hat{\sigma}_{BSM} \otimes \mathcal{L}_{SM} \neq \hat{\sigma}_{SM} \otimes \mathcal{L}_{EIC+FPF}$$

$M_{W'}$: 13.8 TeV



Shift of the contamination threshold

From the fit quality

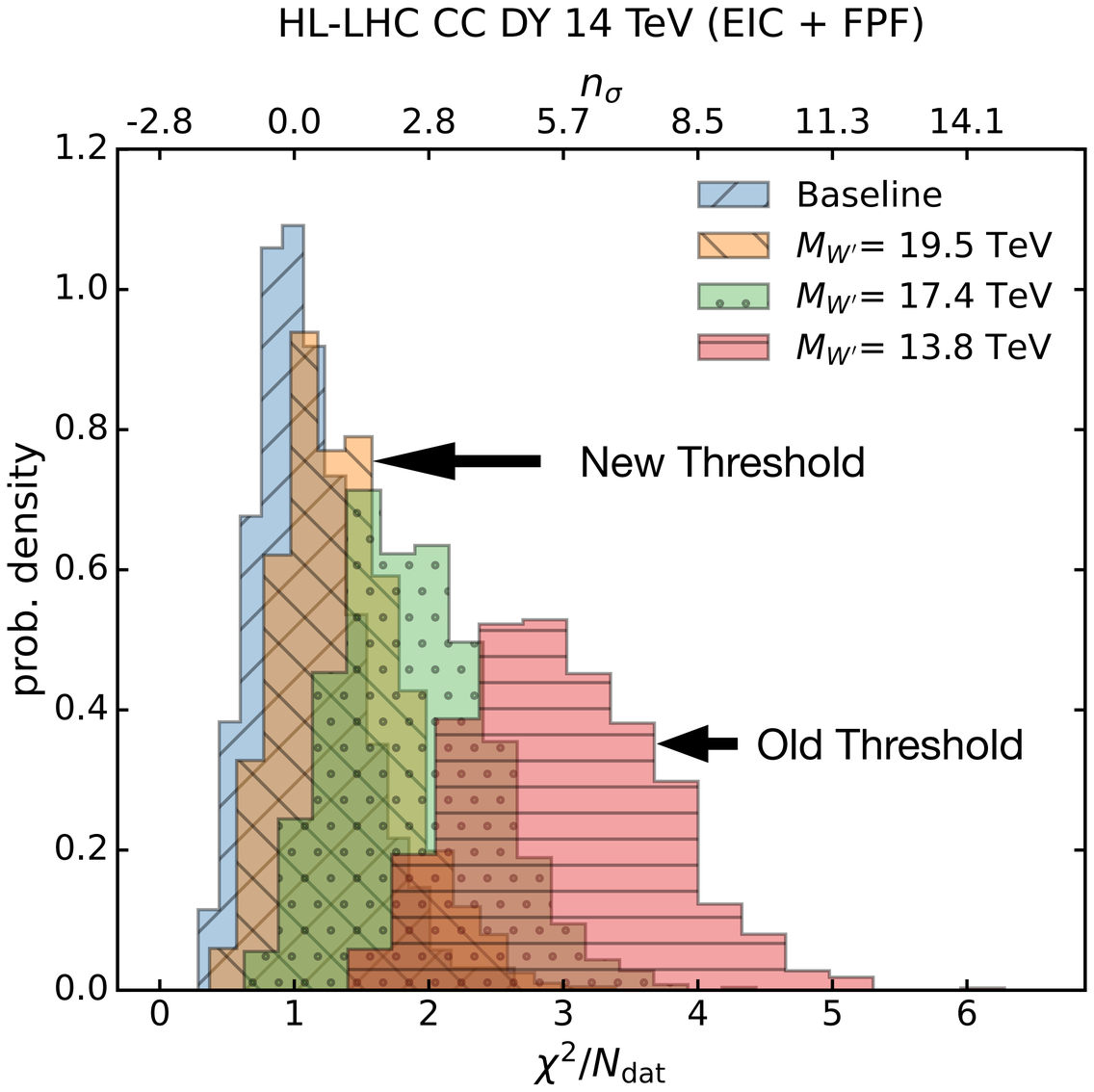
Not a complete solution:

Smaller deviations can still be absorbed

➡ risk at higher BSM mass

Reduction of the “blindspot”:

$$M_{W'} : 13.8 \rightarrow 19.5 \text{ TeV}$$

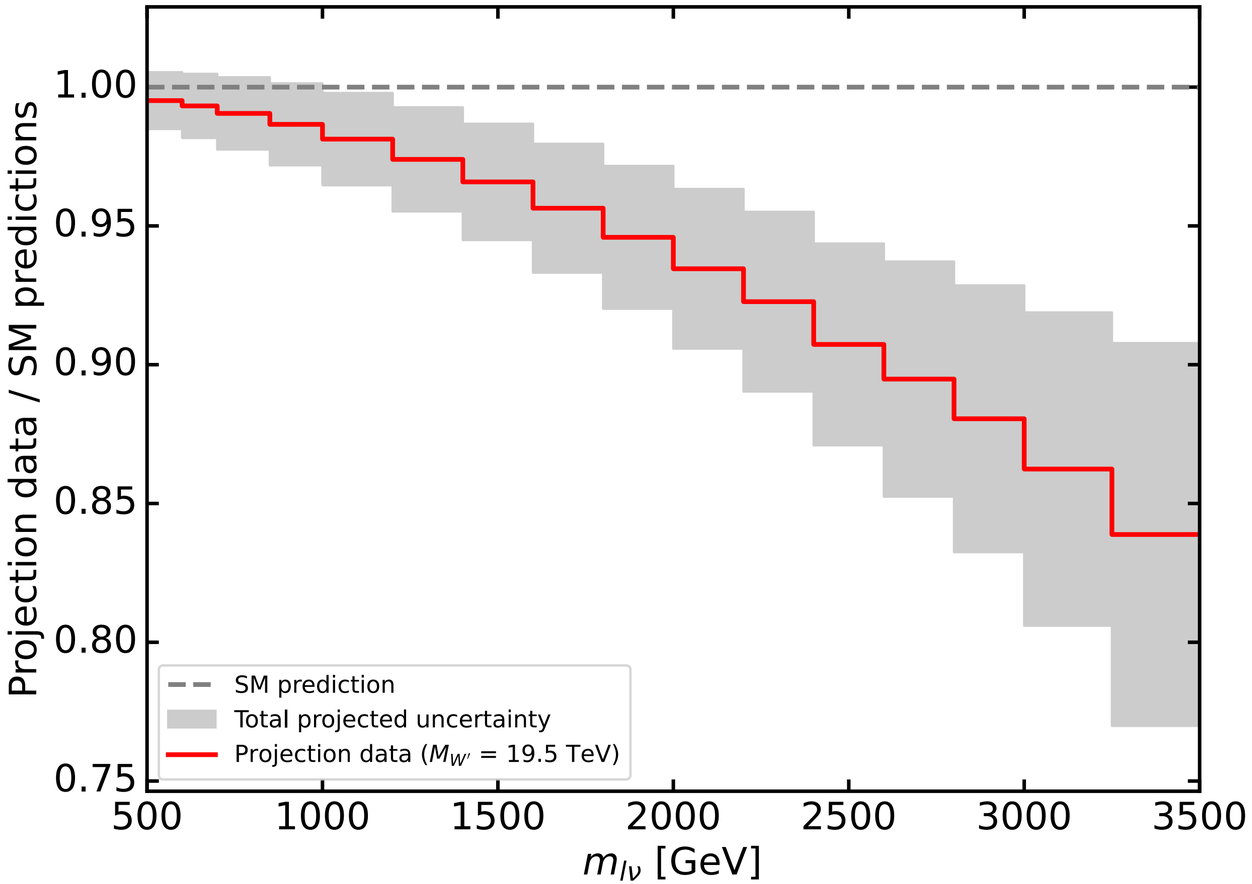


Shift of the contamination threshold

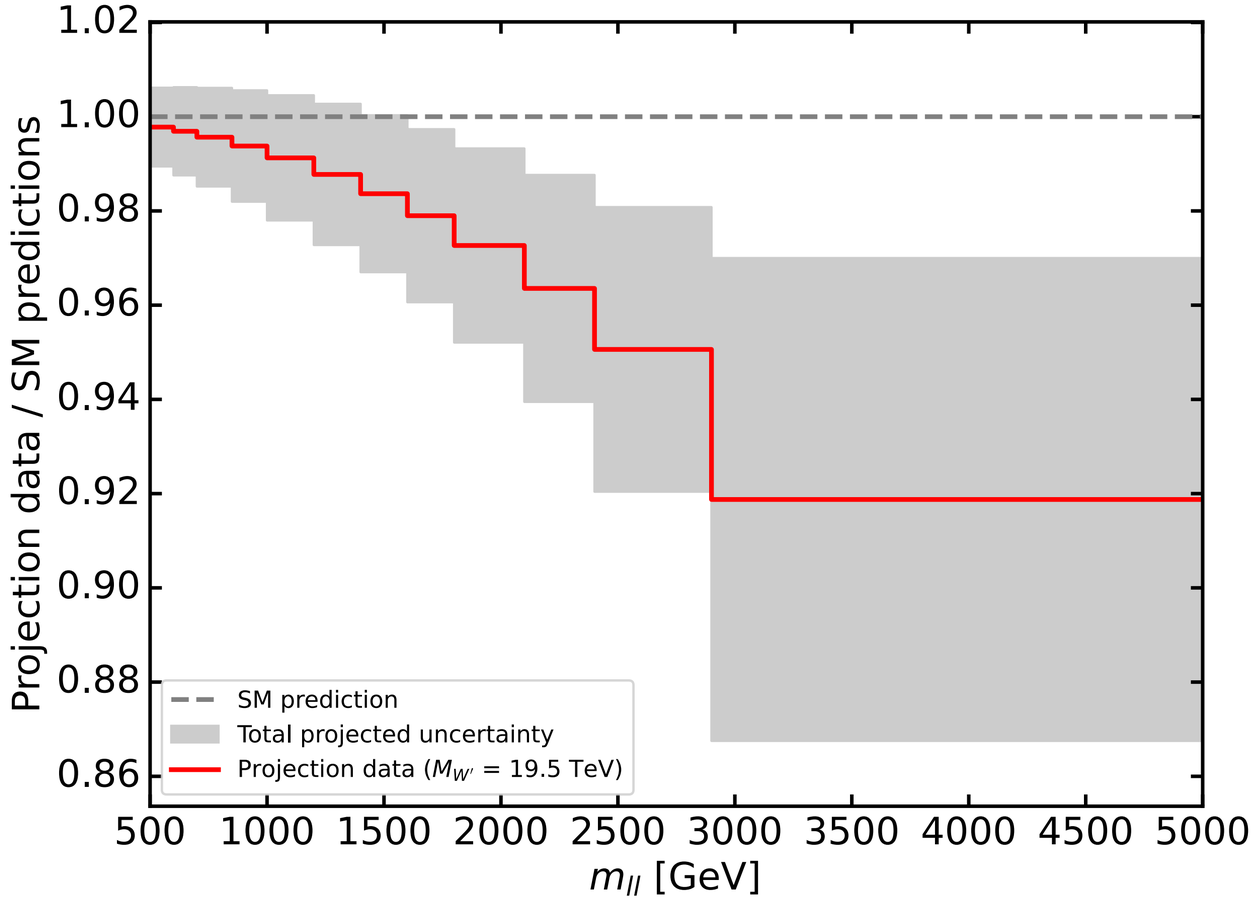
Impact on the HLHC HMDY measurements

$M_{W'} = 19.5$ TeV

Charged current HMDY

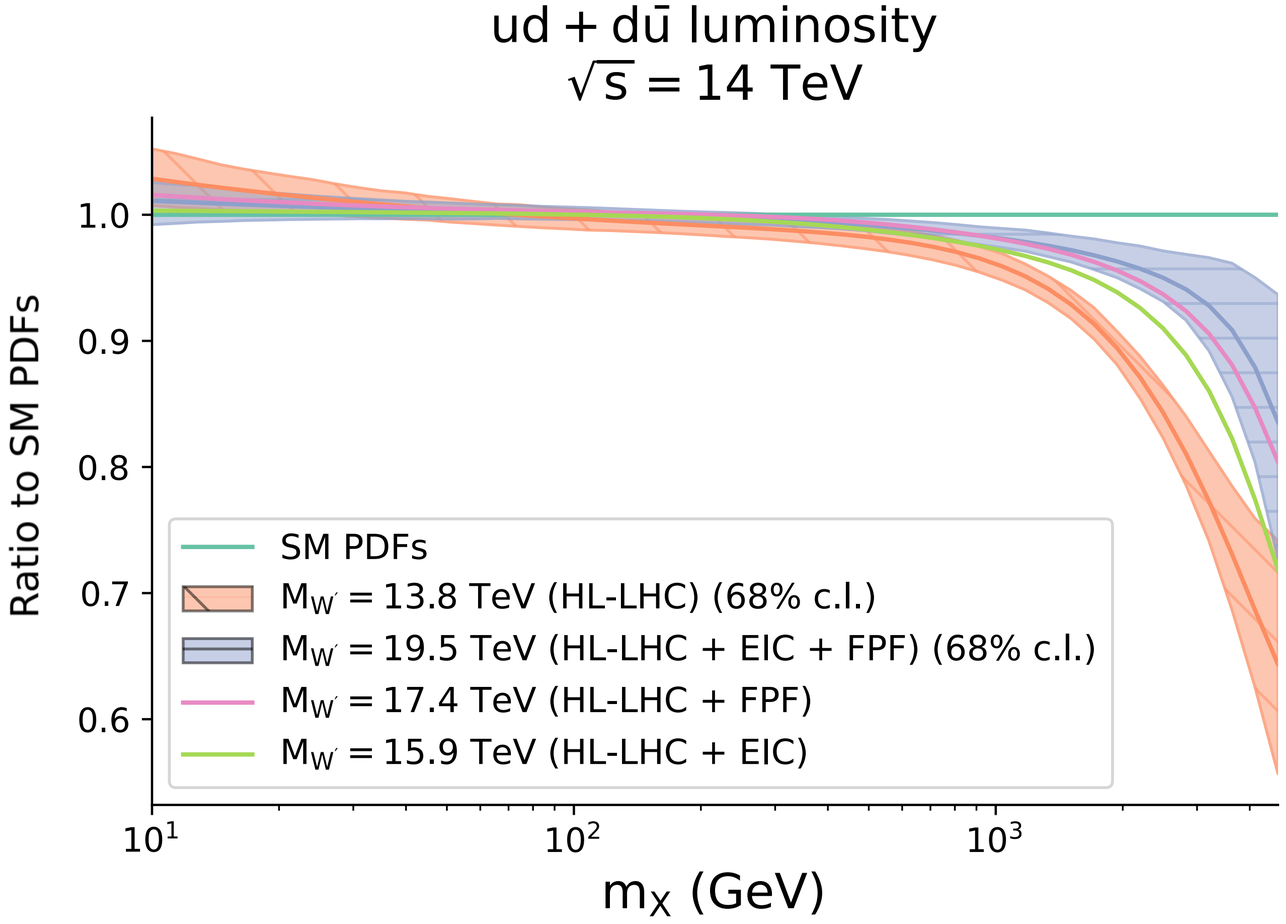


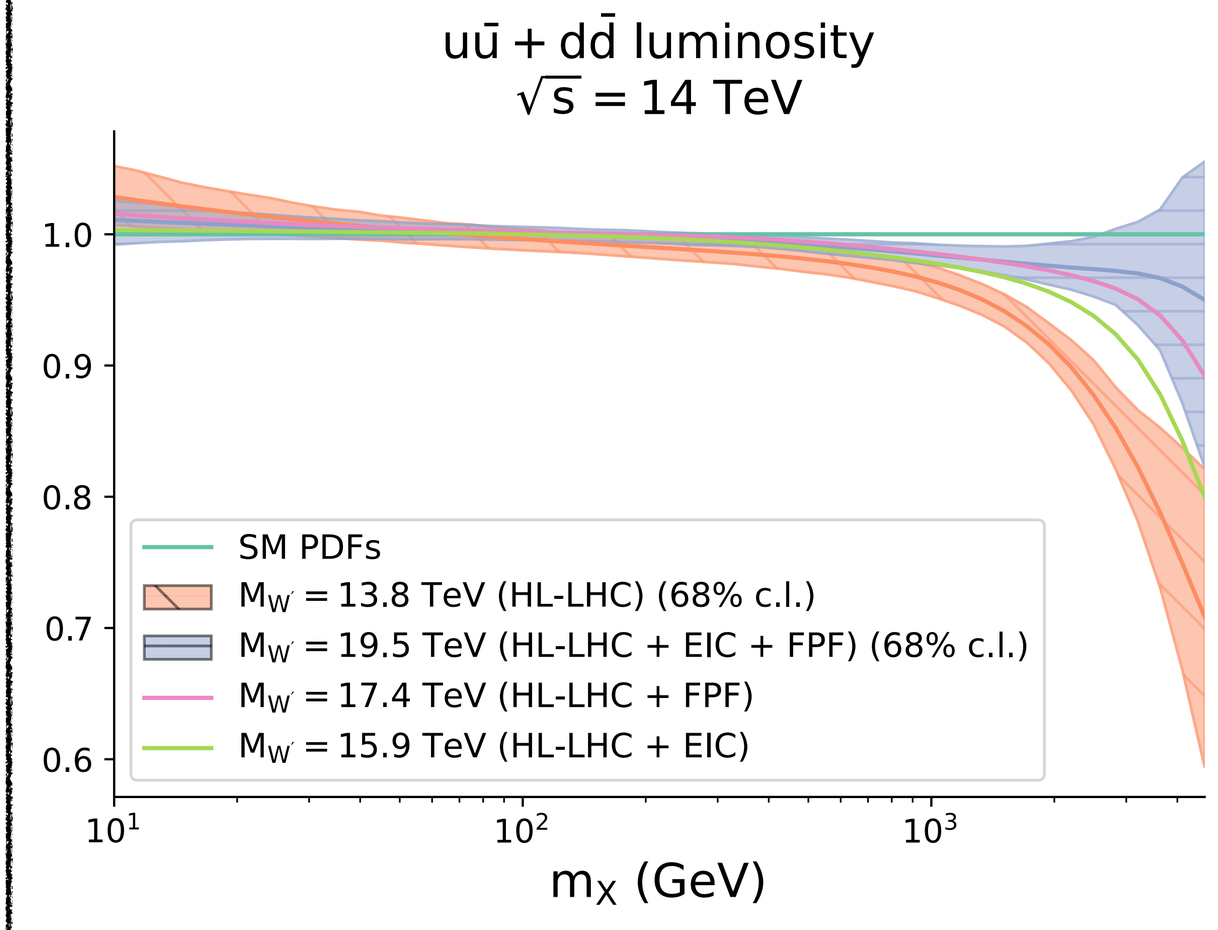
Neutral current HMDY



Shift of the contamination threshold

Impact on PDF luminosities





Second solution

Simultaneous fits of PDFs and SMEFT

[PBSP, 2402.03308, Eur.Phys.J.C]

[PBSP, forthcoming]

Simultaneous fit of PDF and new physics

Separate versus simultaneous fits

Separate fits

PDF fit:

$$T(\{\theta\}, \{c = 0\}) = \text{PDF}(\{\theta\}) \otimes \hat{\sigma}(\{c = 0\})$$

➡ $\bar{\theta}$

Assumes SM: source of bias

SMEFT fit:

$$T(\{\theta = \bar{\theta}\}, \{c\}) = \text{PDF}(\{\theta = \bar{\theta}\}) \otimes \hat{\sigma}(\{c\})$$

➡ $\bar{c}$

Simultaneous fits

$$T(\{\theta\}, \{c\}) = \text{PDF}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

➡ $\{\bar{\theta}, \bar{c}\}$

Removes assumption-based bias

Simultaneous fit of PDF and new physics

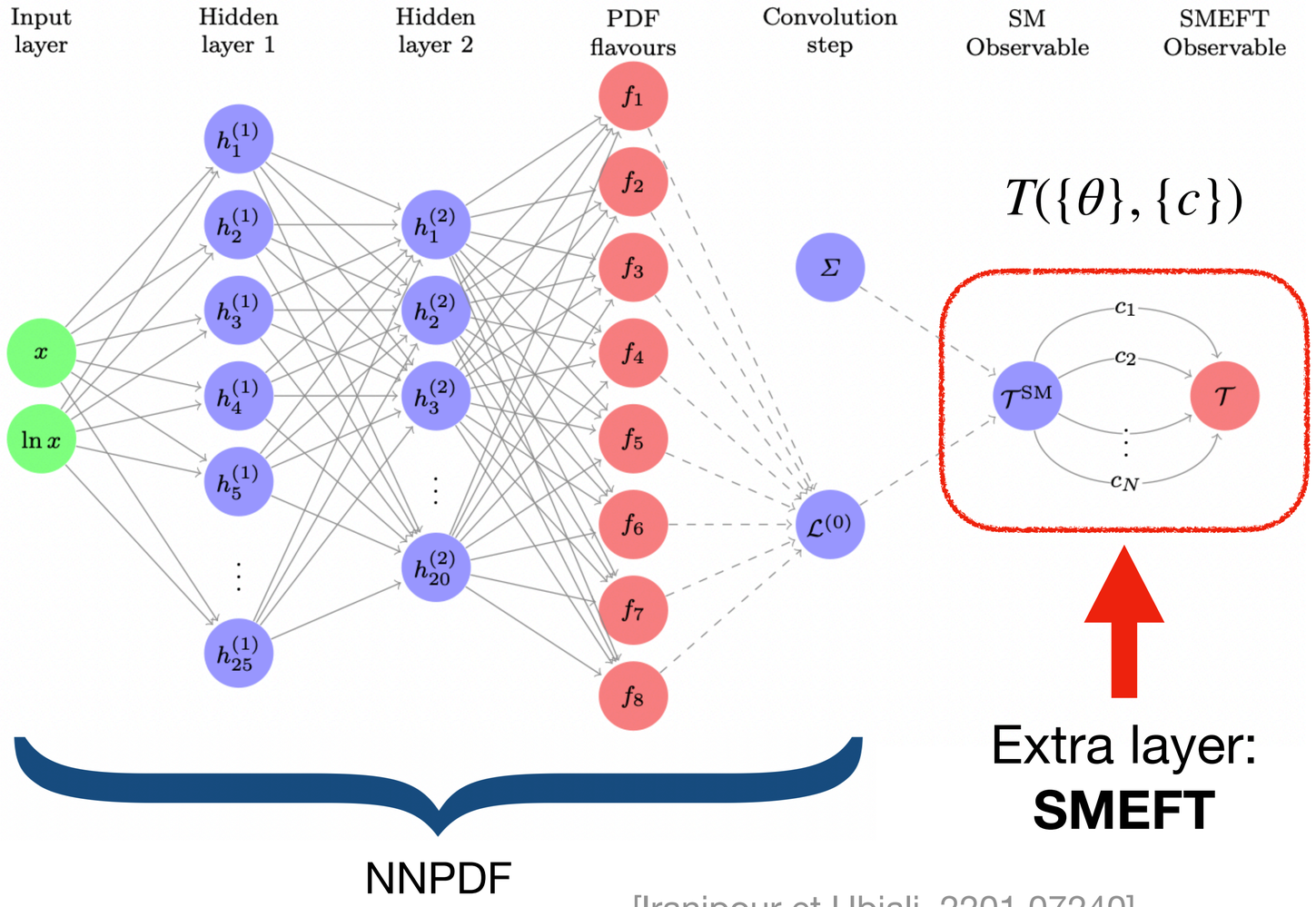
Presentation of the tool: SIMUnet

SIMUnet:

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

  [PBSP, 2402.03308]

- Fits PDFs and WC simultaneously
- Performs contaminated PDF fits



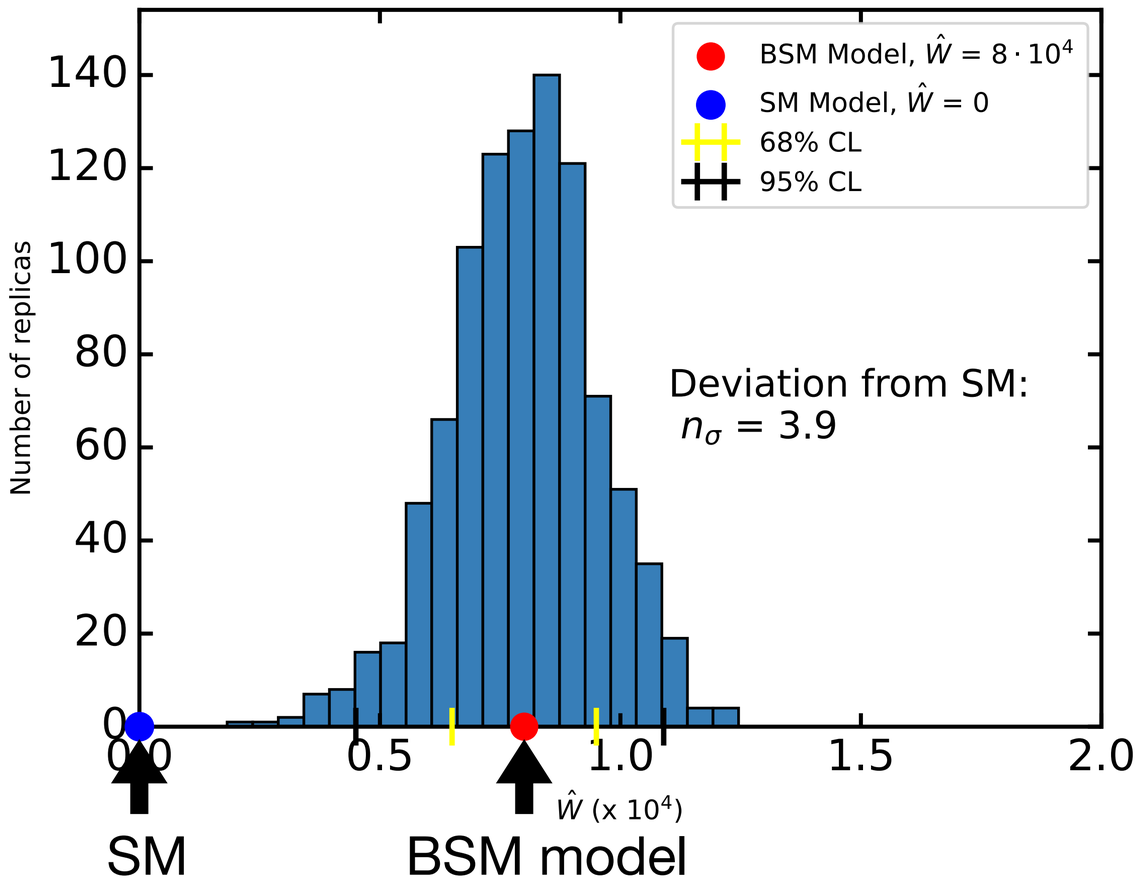
[Iranipour et Ubiali, 2201.07240]

Application to the Drell-Yan sector

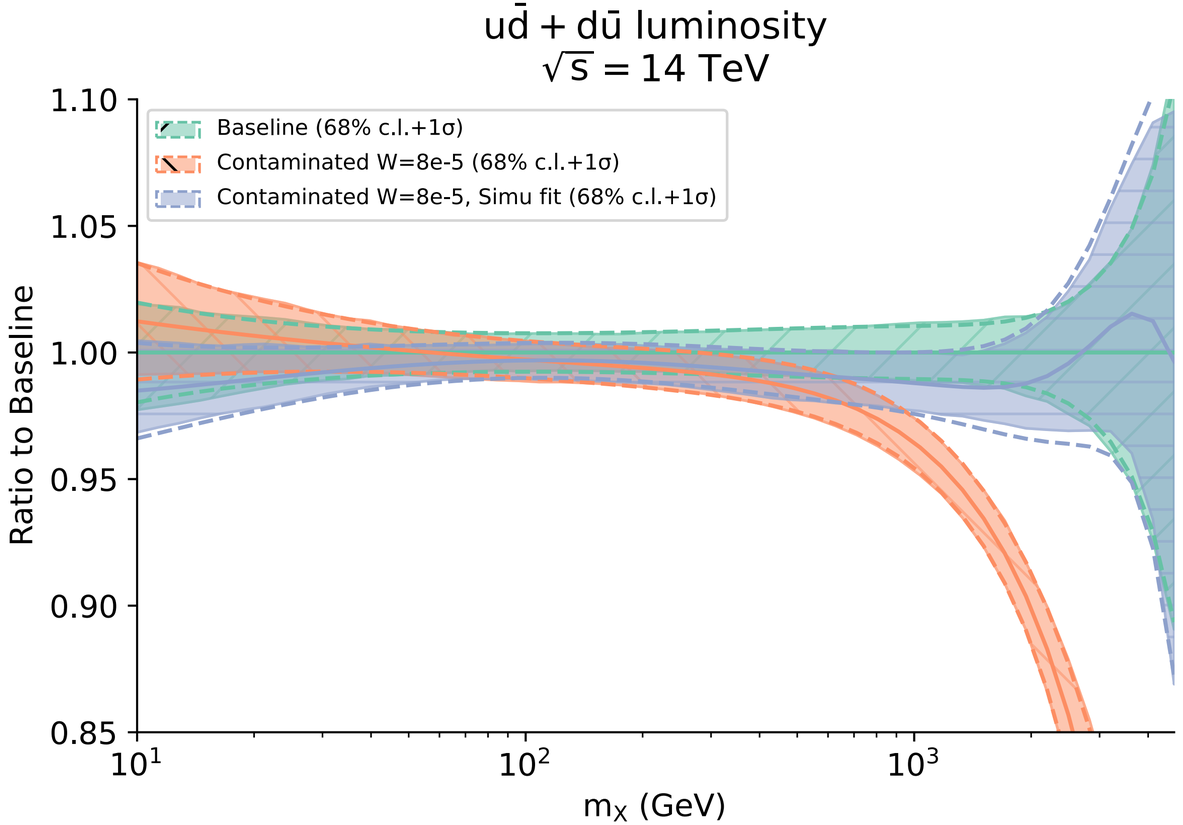
Disentangling PDF contamination

[PBSP, forthcoming]

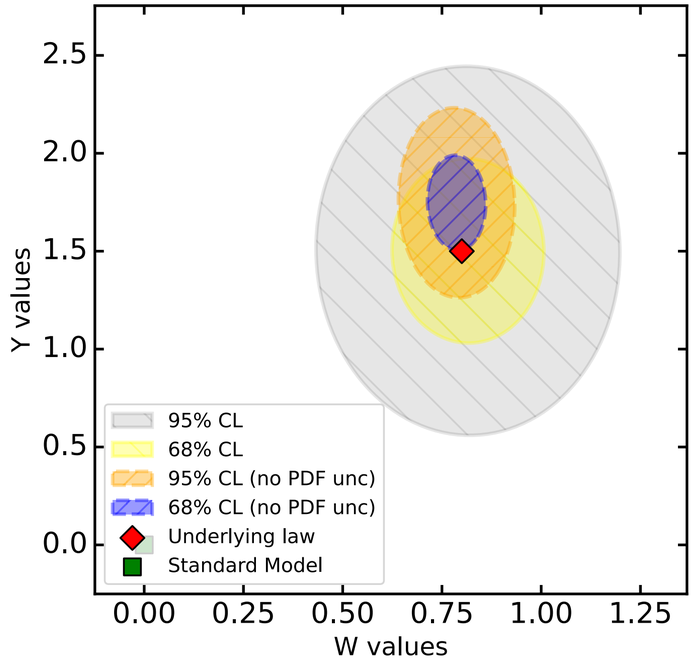
SMEFT marginal distribution



PDF marginal distribution



PDFs for new physics searches

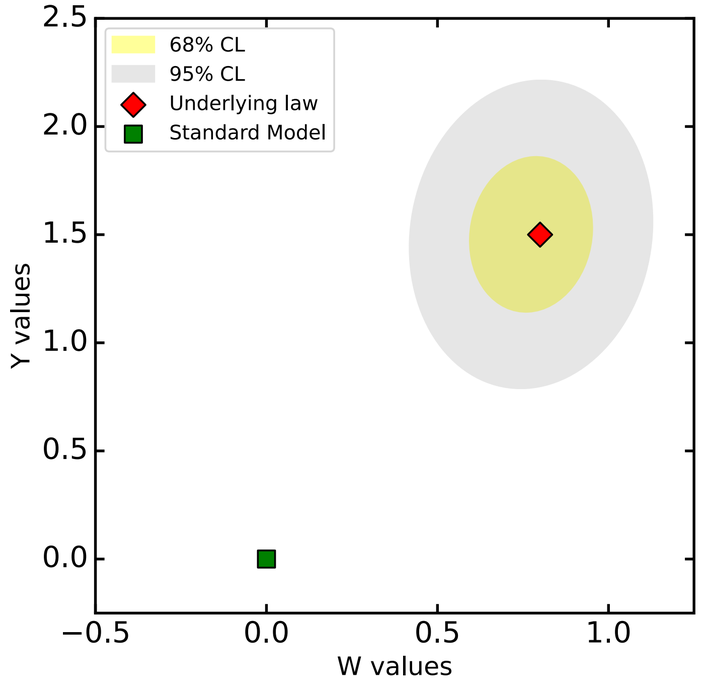


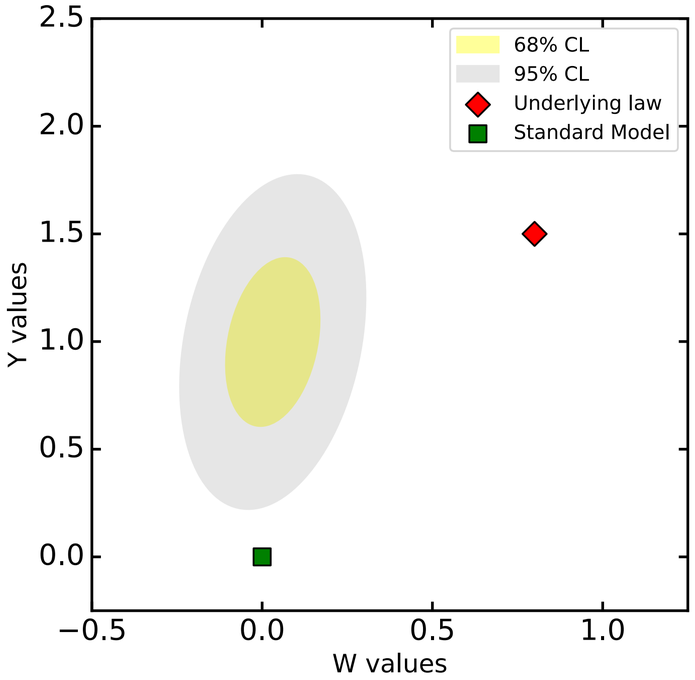
SMEFT only fit
(True PDF)

✗ Impossible

Simultaneous fit
(no PDF assumption)

✓ Doable



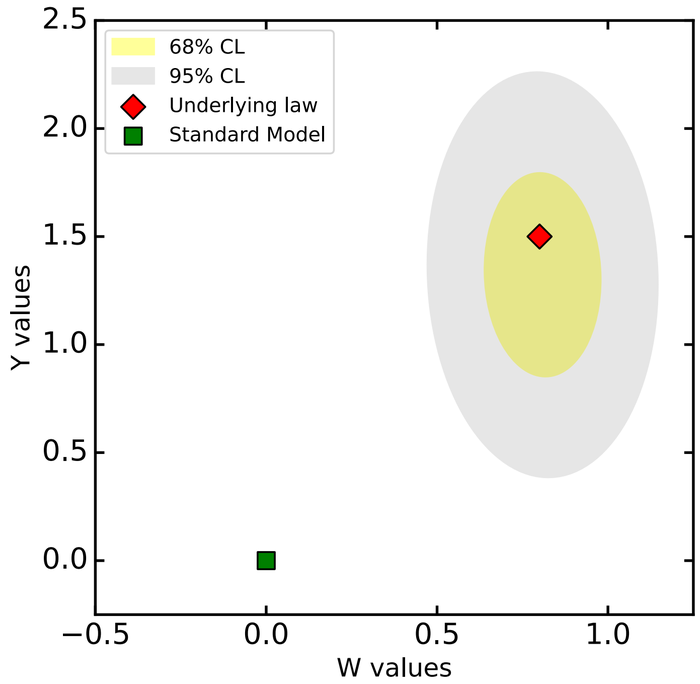


SMEFT only fit
(Contaminated PDF)

✗ Wrong

SMEFT only fit
(Conservative PDF)

✓ Doable



Limits of the simultaneous fits

- Technical limits:
  - Can only fit linear SMEFT corrections (issue with Monte Carlo replica method)
    - Working on an alternative bayesian method

- Fundamental limits:
  - More difficult than PDF fit
  - Correlation between PDFs and SMEFT?
  - **Risk to absorb error as SMEFT signal**

Summary and outlook

- Signs of W' got fitted away in PDF parametrisation
  - Missed new physics
  - Exclude true underlying law
- Bridge the dataset “blind spot”:
  - Add precise large-x low-energy datasets into fits: FPF + EIC
  - Pushes the PDF/BSM mixing threshold toward higher energies
- Simultaneous fits of PDFs and SMEFT:
  - Fitting simultaneously PDF and new physics: **SIMUnet** tool already available
  - Ongoing study on real jet data [Greljo, Hammou, Merlotti, Ubiali, forthcoming]
  - Developing a bayesian framework [Costantini, Moore, Mantani, Schutze, Ubiali, forthcoming (PDF)] [Hammou, ter Hoeve, Shutze, in progress (SMEFT)]

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

Thank you for your attention!

Extra slides

PARTON DISTRIBUTION FUNCTIONS

$$f_i(x,\mu)$$

Perturbative QCD

$$\frac{d}{dt}\begin{pmatrix} q_i(x,t) \\ g(x,t) \end{pmatrix} = \frac{\alpha_s(t)}{2\pi}\int_x^1 \sum_{j=q,\bar{q}} \frac{d\xi}{\xi} \begin{pmatrix} P_{ij}\left(\frac{x}{\xi},\alpha_s(t)\right) & P_{ig}\left(\frac{x}{\xi},\alpha_s(t)\right) \\ P_{gj}\left(\frac{x}{\xi},\alpha_s(t)\right) & P_{gg}\left(\frac{x}{\xi},\alpha_s(t)\right) \end{pmatrix} \otimes \begin{pmatrix} q_j(\xi,t) \\ g(\xi,t) \end{pmatrix}$$

Dokshitzer - Gribov - Lipatov - Altarelli - Parisi

DGLAP evolution equation

- Impressive progress in amplitude computations leading towards solution of DGLAP evolution equations up to N^{3}LO in perturbative QCD, plus NLO-coupled QED. Many ingredients made available, some still missing

  ➡ 4-loop DGLAP Splitting Functions P_{ij} to evolve PDFs

  non-singlet - large n_F limit [NPB 915 (2017) 335; arXiv:2308.07958]
  - small-x [JHEP 08 (2022) 135] and large-x [JHEP 10 (2017) 041] limits
  - lowest 8 Mellin moments [JHEP 06 (2018) 073]

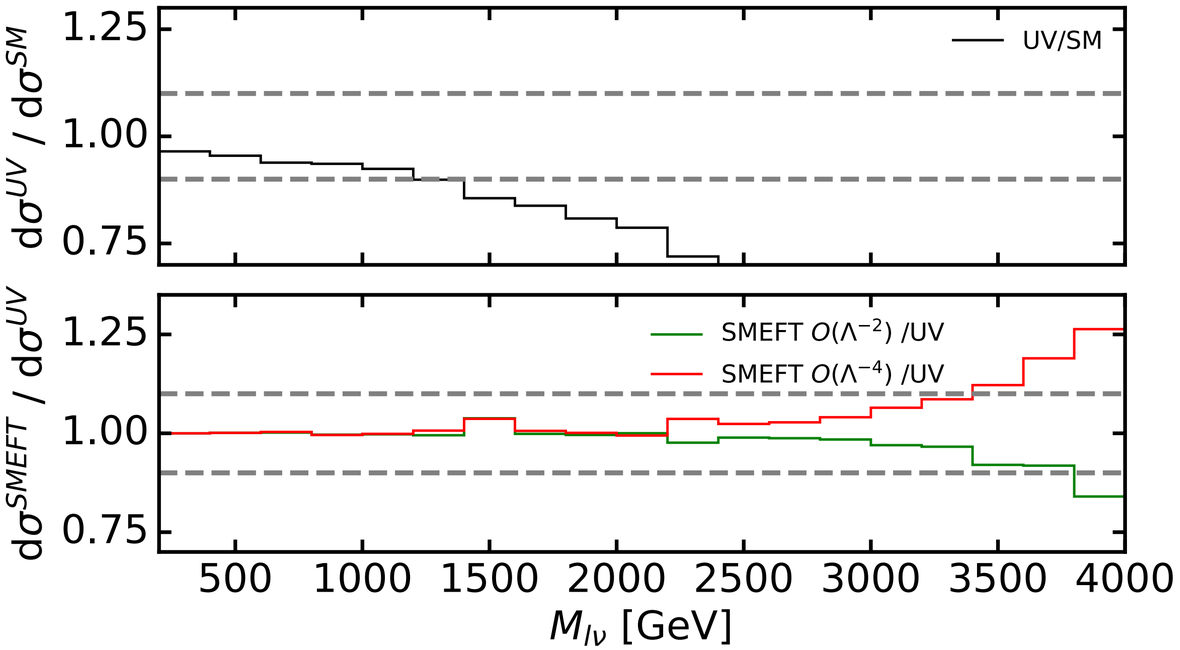
  singlet
  - large n_F limit [NPB 915 (2017) 335; arXiv:2308.07958, arXiv:2310.01245]
  - small-x [JHEP 06 (2018) 145] and large-x [NPB 832 (2010) 152; JHEP 04 (2020) 018; JHEP 09 (2022) 155] limits
  - lowest 5 (10) Mellin moments [PLB 825 (2022) 136853; ibid. 842 (2023) 137944; ibid. 846 (2023) 138215]

  ➡ Deep Inelastic Structure Functions (hard scattering coefficient functions for DIS)
  - DIS NC (massless) [NPB 492 (1997) 338; PLB 606 (2005) 123; NPB 724 (2005) 3]
  - DIS CC (massless) [NPB 813 (2009) 220]
  - Massive from param. combining known limits and damping functions [NPB 864 (2012) 399]

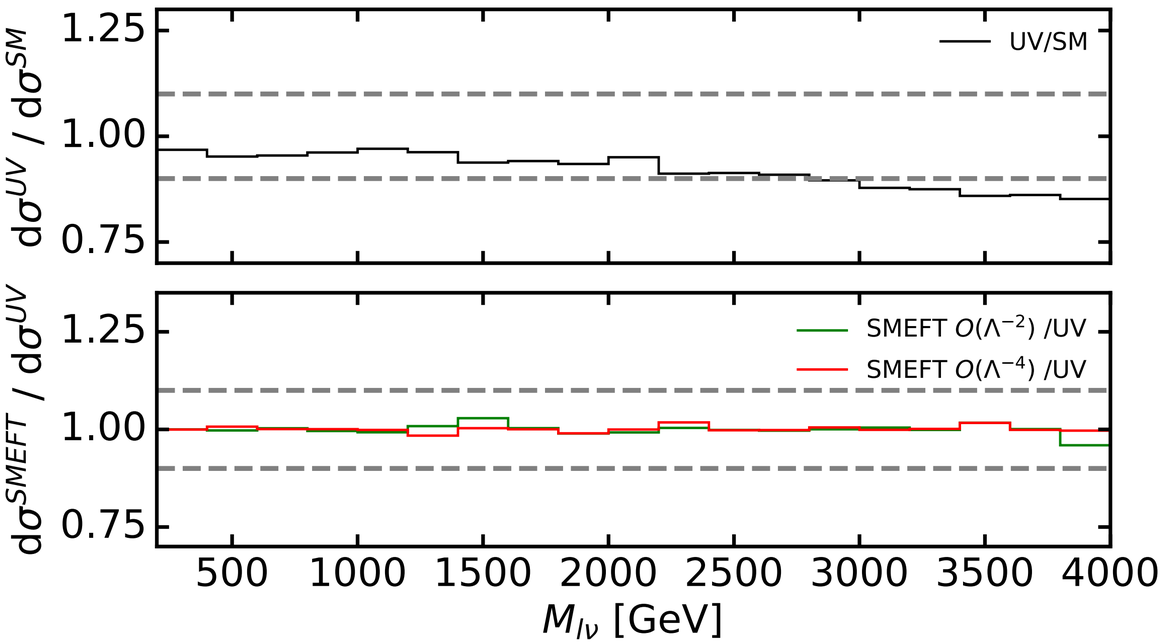
New physics scenarios: W'

Consideration of different masses

$M_{W'} = 10$ TeV

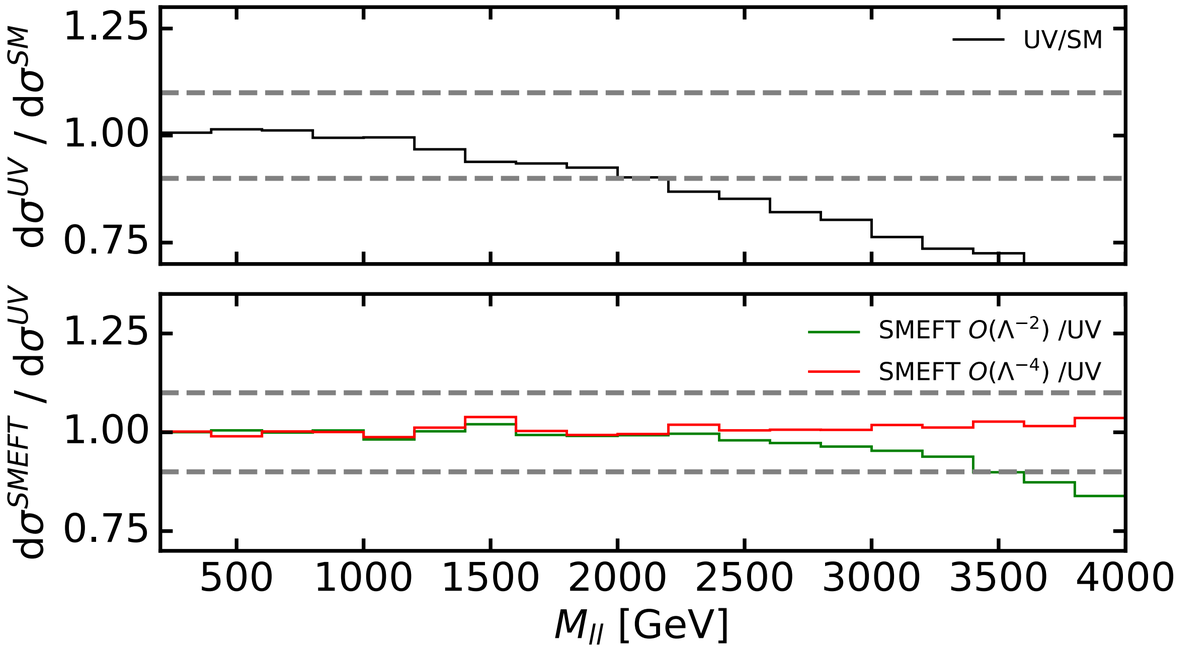


$M_{W'} = 22.5$ TeV

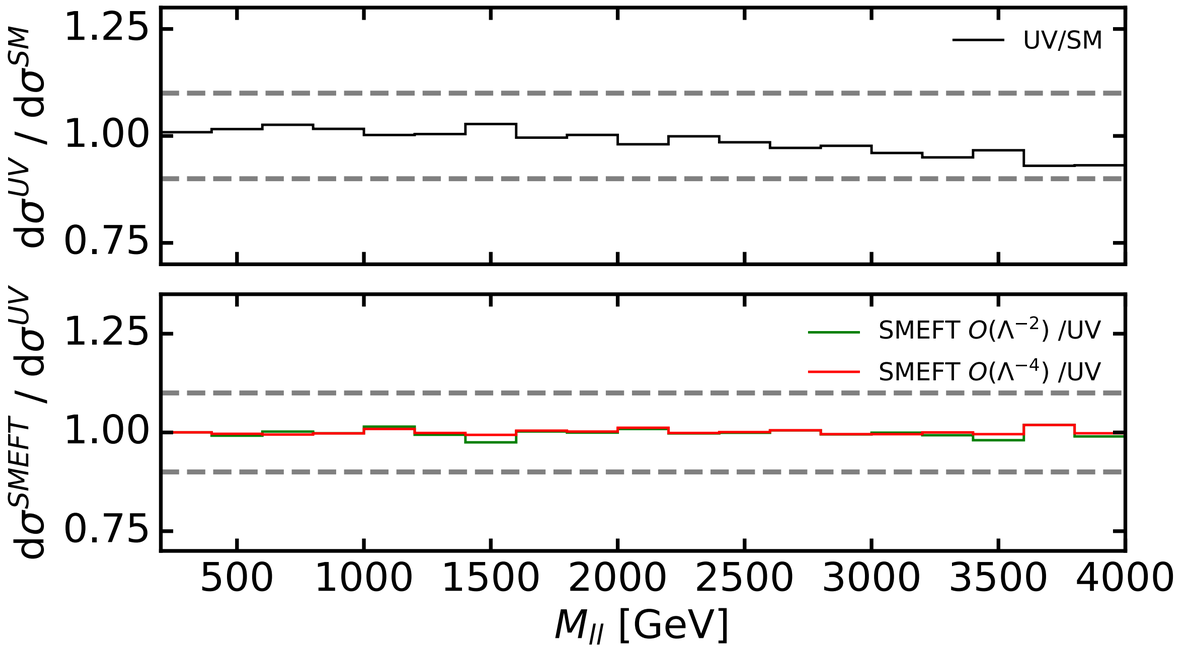


New physics scenarios: Z'

$M_{Z'} = 14.5$ TeV



$M_{Z'} = 32.5$ TeV



PDF fitting: selection criteria

Exclusion of incompatible datasets (NNPDF criteria)

Two criteria:

$$\chi^2 = (D - T_{SM})^T \cdot V_{cov}^{-1} \cdot (D - T_{SM})$$

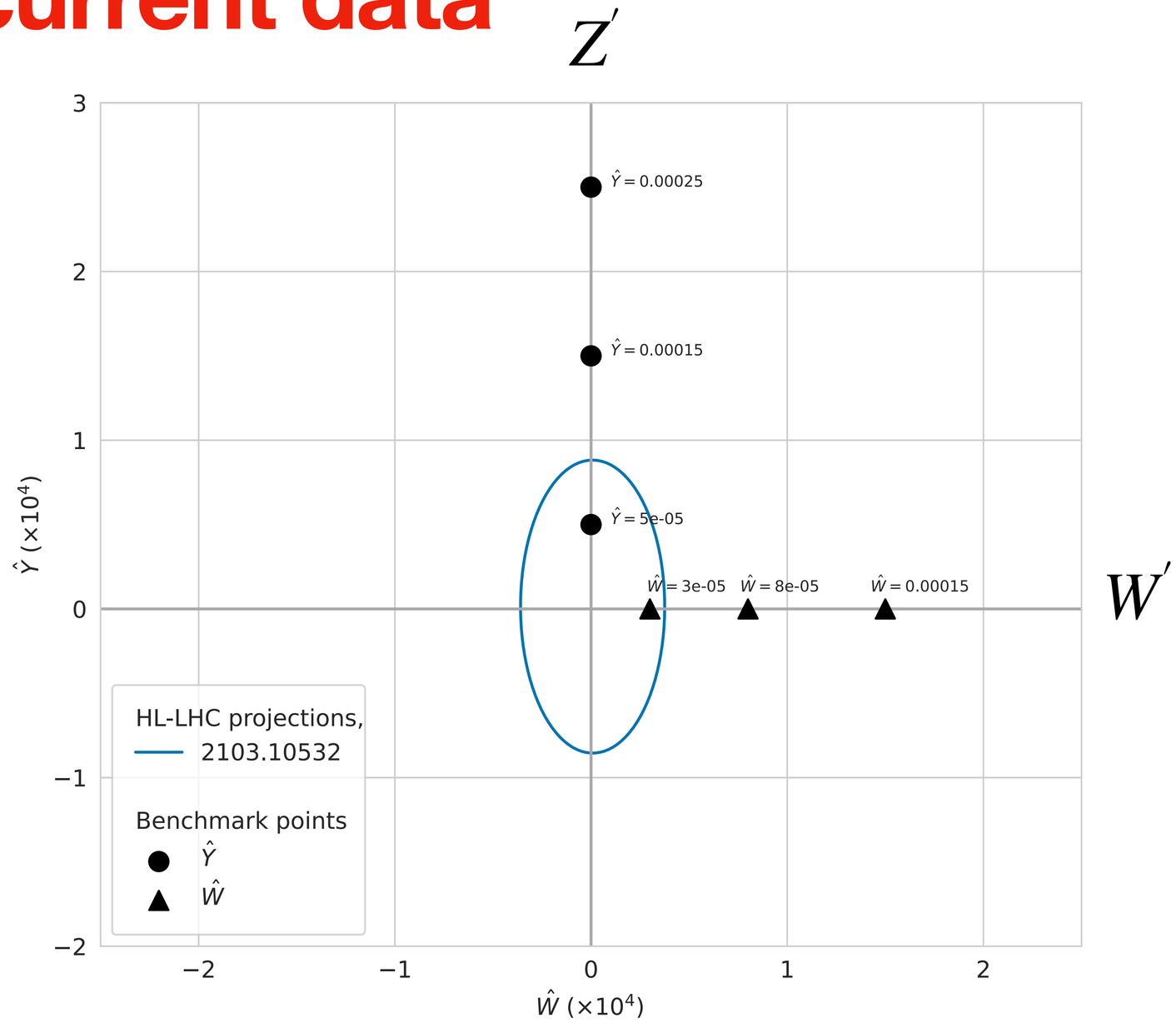
- χ^2-statistics:
  - $\frac{\chi^2}{n_{dat}} > 1.5$ → excluded

- n_σ standard deviation:
  - $n_\sigma > 2$ → excluded

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

Constraints from current data

- New physics scenarios compared to constraints at 95% CL



Impact of contamination: fake deviations

SM predictions with:

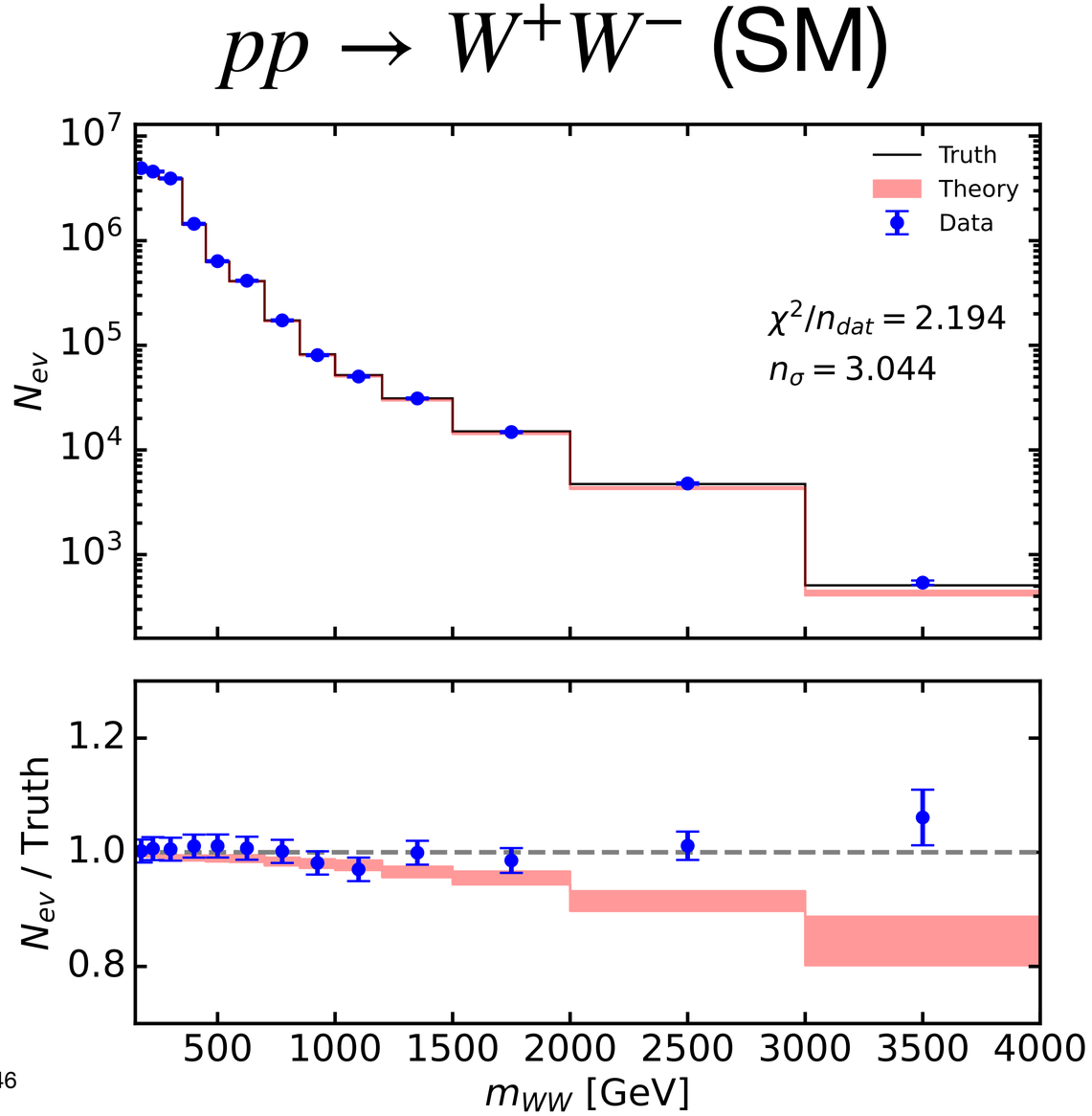
- Contaminated PDFs (red)
- True PDFs (black)

➡ Fake deviation in other sectors

Also seen in:

WH, WZ, ZH production

HL-LHC Projections



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

SIMUNET: INPUT DATA

Exp.	$\sqrt{s}$ (TeV)	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}
ATLAS and CMS	7 and 8	$\mu_{H\to\mu^+\mu^-}$	5 and 20	22
CMS	13	μ_H	35.9	24
ATLAS	13	μ_H	80	25
ATLAS	13	$\mu_{H\to Z\gamma}$	139	1
ATLAS	13	$\mu_{H\to\mu^+\mu^-}$	139	1

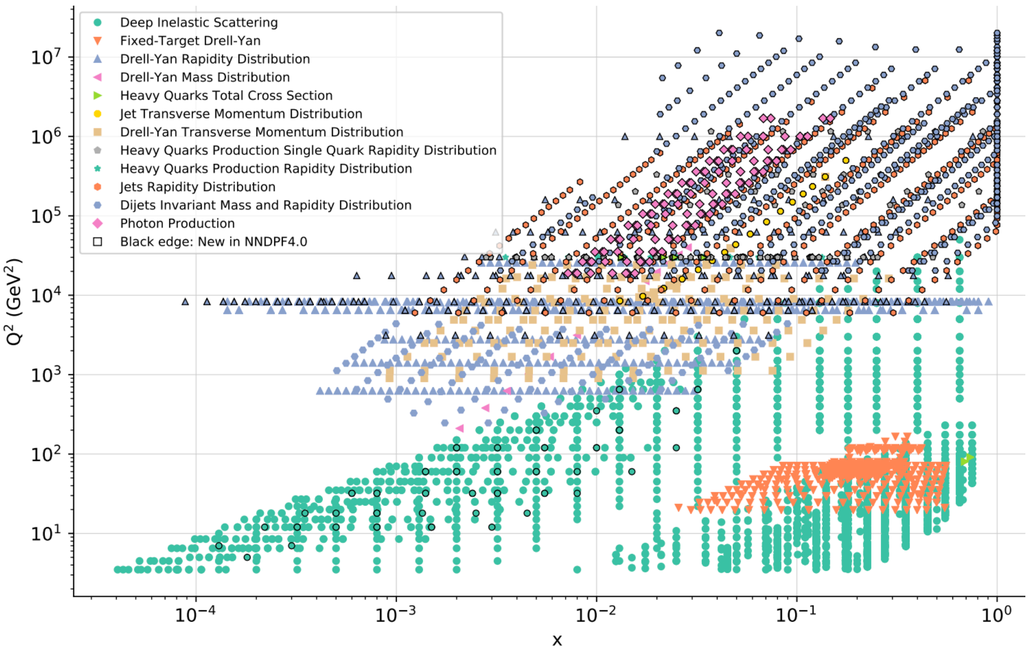
Higgs signal strength SMEFT only

Exp.	$\sqrt{s}$ (TeV)	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}
LEP	0.182	$d\sigma_{WW}/dcos(\theta_W)$	0.164	10
LEP	0.189	$d\sigma_{WW}/dcos(\theta_W)$	0.588	10
LEP	0.198	$d\sigma_{WW}/dcos(\theta_W)$	0.605	10
LEP	0.206	$d\sigma_{WW}/dcos(\theta_W)$	0.631	10
ATLAS	13	$d\sigma_{W^+W^-}/dm_{e\mu}$	36.1	13
ATLAS	13	$d\sigma_{WZ}/dm_T$	36.1	6
ATLAS	13	$d\sigma(Zjj)/d\Delta\phi_{jj}$	139	12
CMS	13	$d\sigma_{WZ}/dp_T$	35.9	11

Di-boson SMEFT only

Exp.	$\sqrt{s}$ (TeV)	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}
LEP	0.250	Z observables		19
LEP	0.196	$\mathcal{B}(W\to l^-\bar{\nu}_l)$	3	3
LEP	0.189	$\sigma(e^+e^-\to e^+e^-)$	3	21
LEP	0.209	$\hat{\alpha}^{(5)}(M_Z)$	3	1

EWPO SMEFT only



Total of ~ 5000 input datapoints, some constraining only SMEFT, some only PDFs, some both SMEFT & PDFs

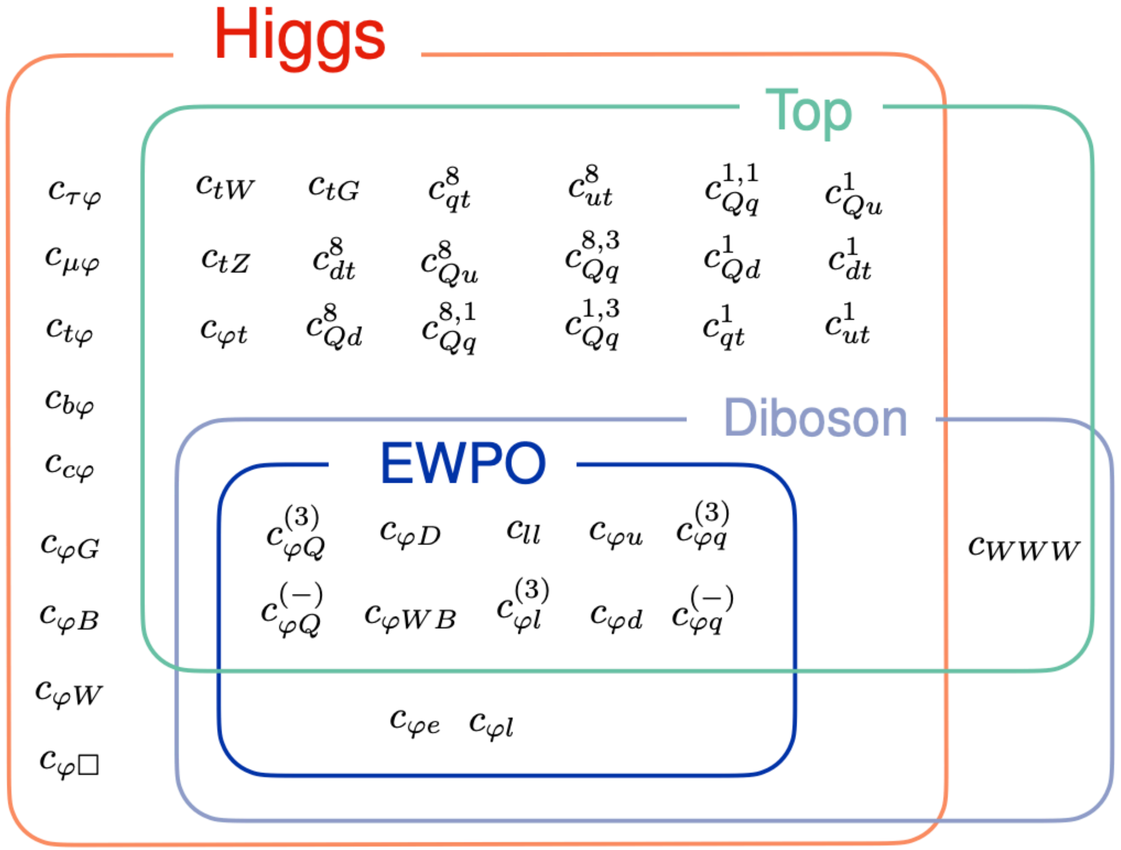
+O(4000) data from DIS, DY, jets, di-jets, W and Z production, Z pT - PDF only

Simultaneous fit of PDF and new physics

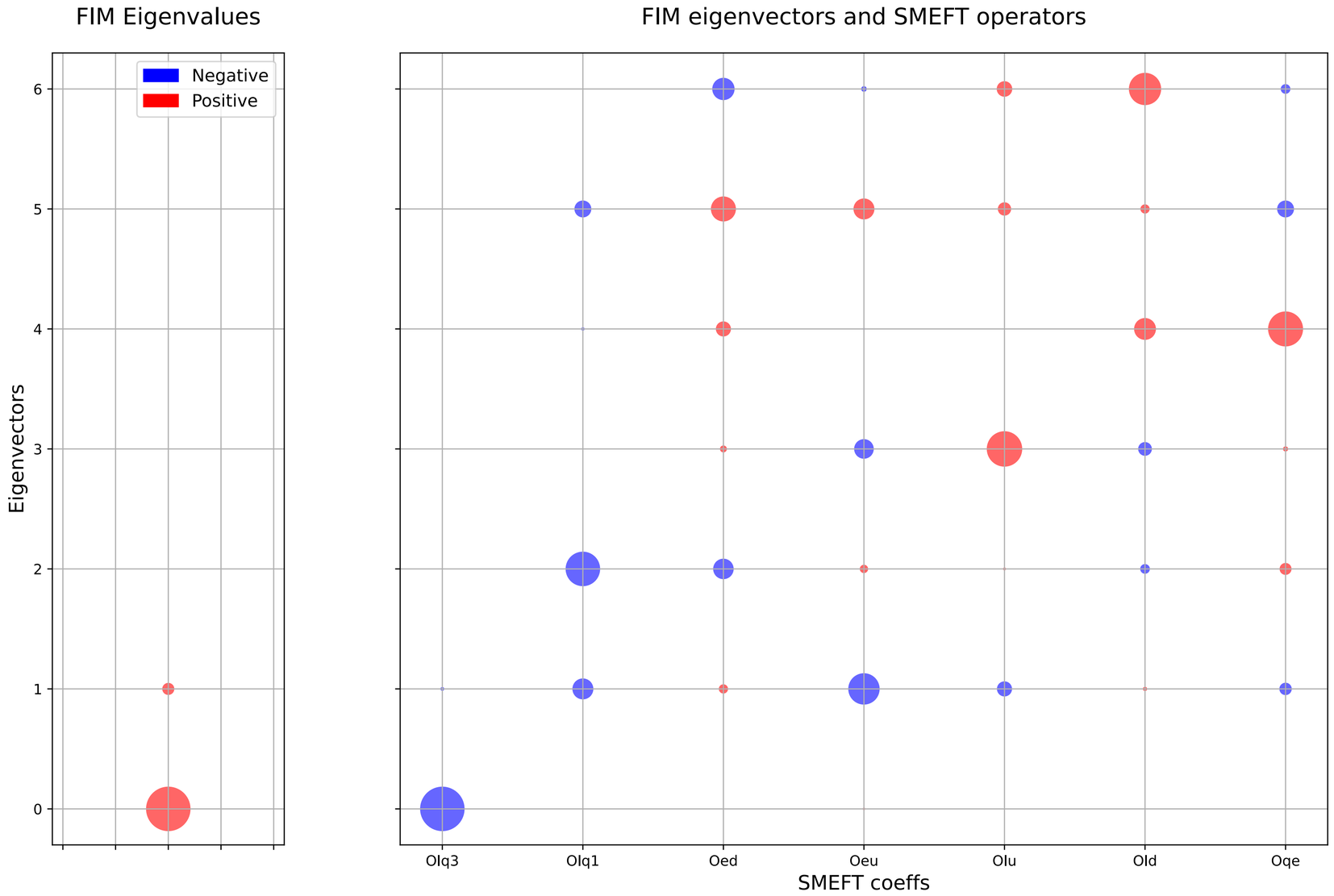
SMEFT operators implemented

- 40 operators implemented

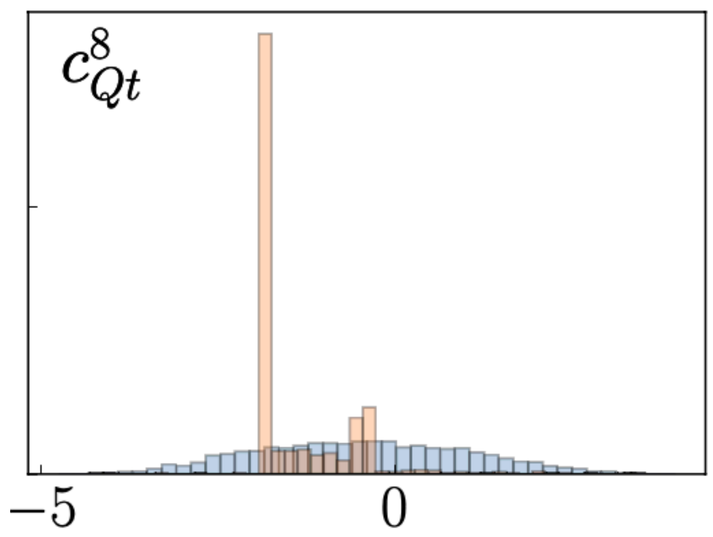
- Observables:
  - top sector
  - diboson
  - Higgs
  - Drell-Yan
  - EW Precision Observables

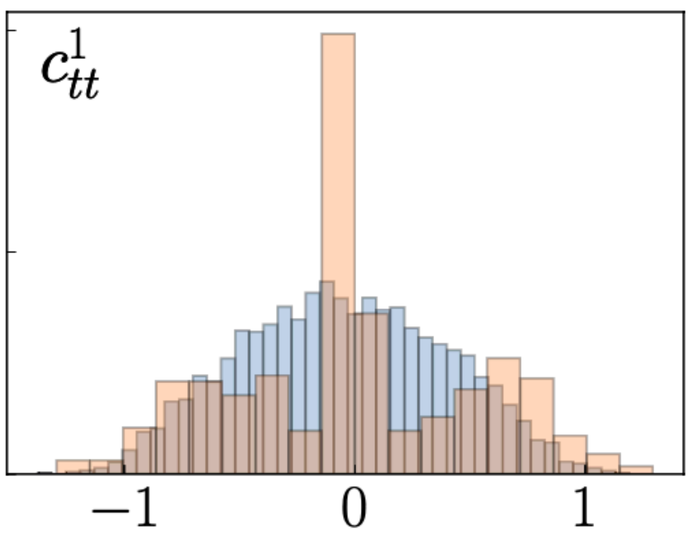


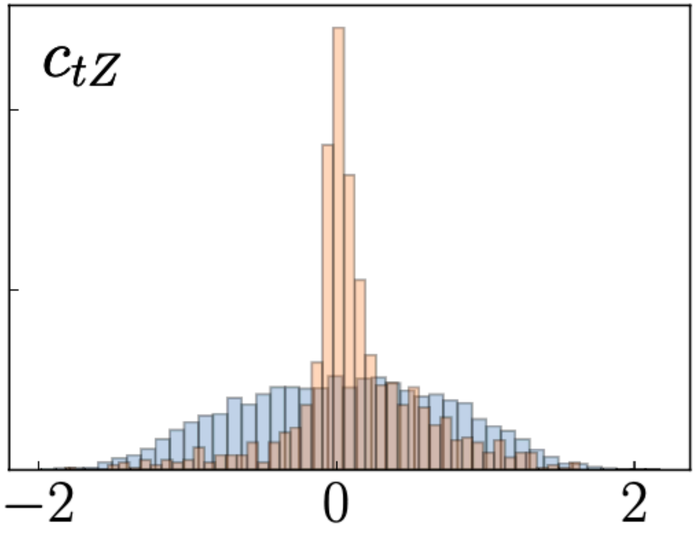
SMEFT PCA results



BEYOND MONTE CARLO AND NEURAL NETWORKS







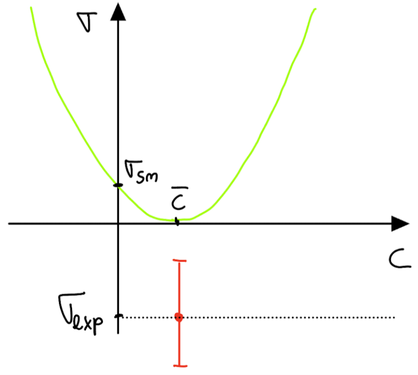
- In the quadratic SMEFT fit observed disagreement between MC method and Bayesian method. Very different posterior (hence different CLs)

- Study of MC versus Bayesian method based on nested sampling for PDF fits and SMEFT fits [Costantini, Madigan, Mantani, Moore arXiv:2404.10056]

- Towards a general Bayesian methodology for simultaneous fits [Costantini, Mantani, MU, in progress]

Let's consider a simple scenario: 1 operator, 1 datapoint

$$\chi^2 = \frac{(\sigma(c) - \sigma_{exp})^2}{\delta\sigma^2} \qquad \Delta\chi^2 = \chi^2 - \chi_{min} = 1 \quad \rightarrow \quad [c_-, c_+]$$

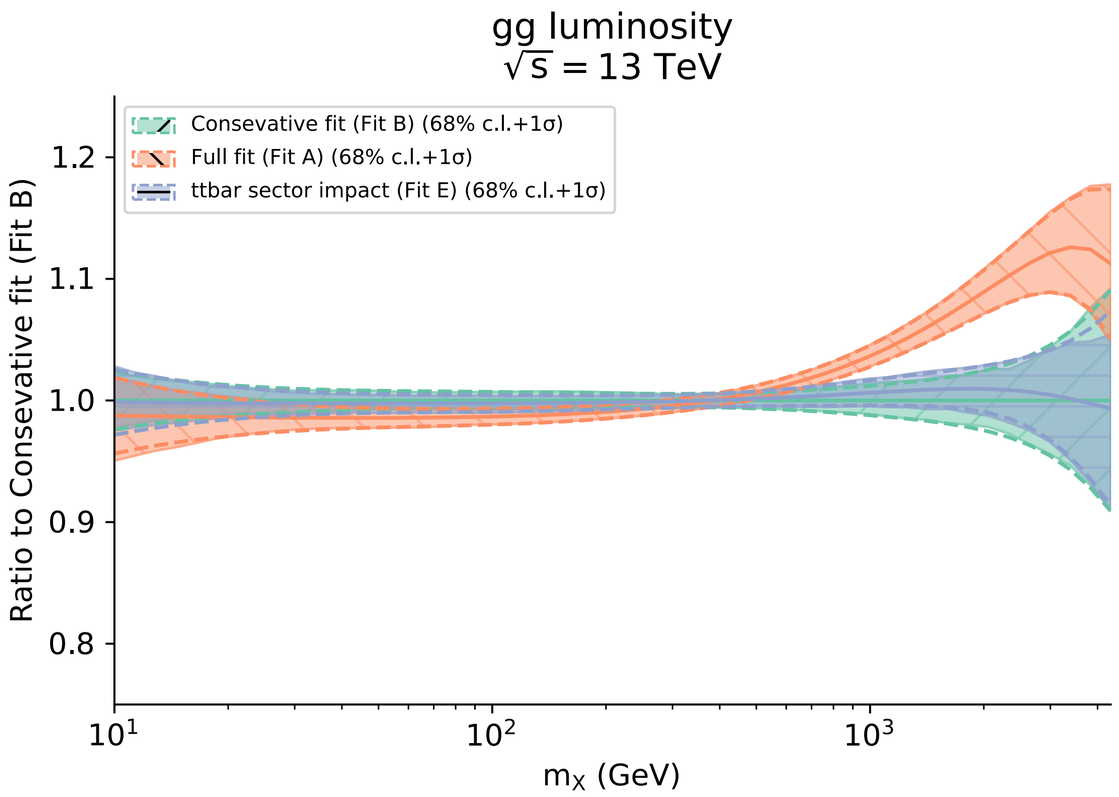


Study of other sectors (real data)

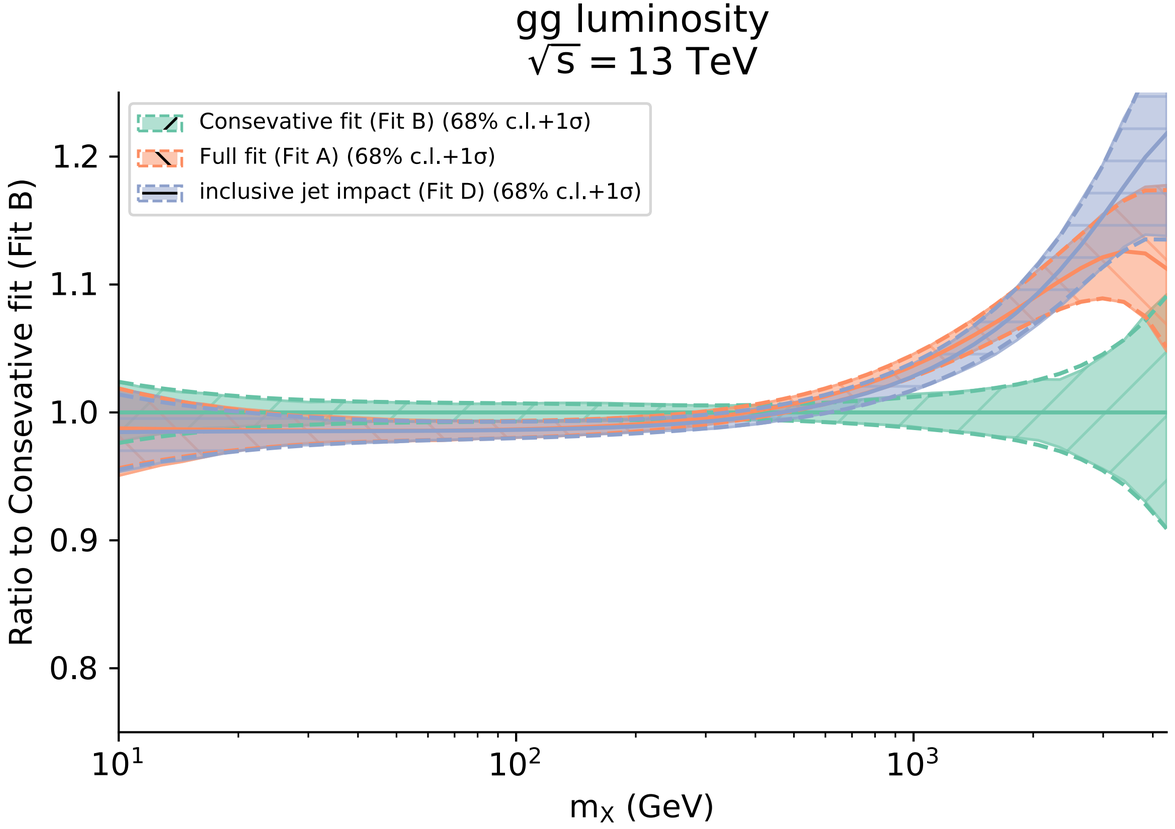
In progress

Closer look at the gg luminosity tension

Top sector



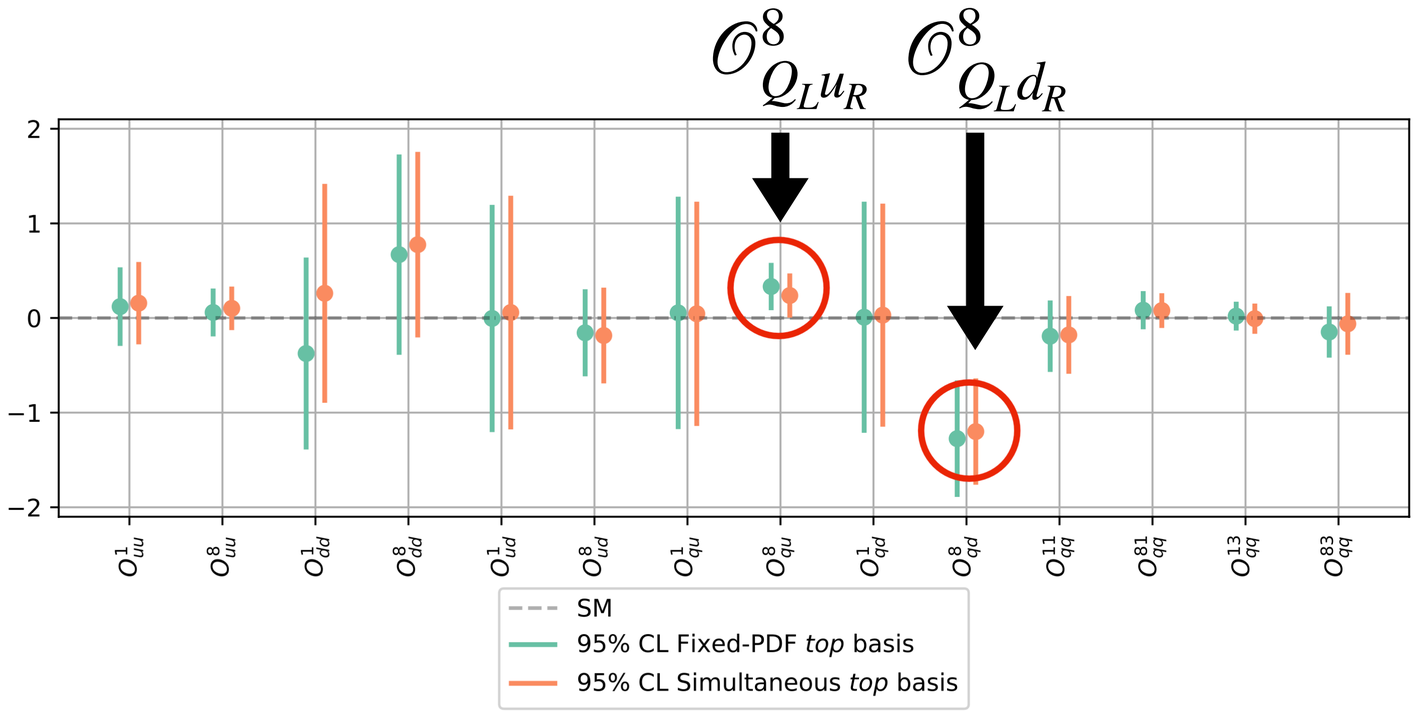
Jet sector



Application to the jet sector (real data)

In progress

SMEFT Fit



4-quark operators
(first two light families)

PDF Fit