Tailored PDFs for new physics searches

[Hammou et al. (PBSP + Mangano), 2307.10370, JHEP]

[Costantini et al. (PBSP), 2402.03308, Eur.Phys.J.C]

[Cole et al. (PBSP), forthcoming]

European Research Council

Established by the European Commission

Funded by
the European Union

PBSP

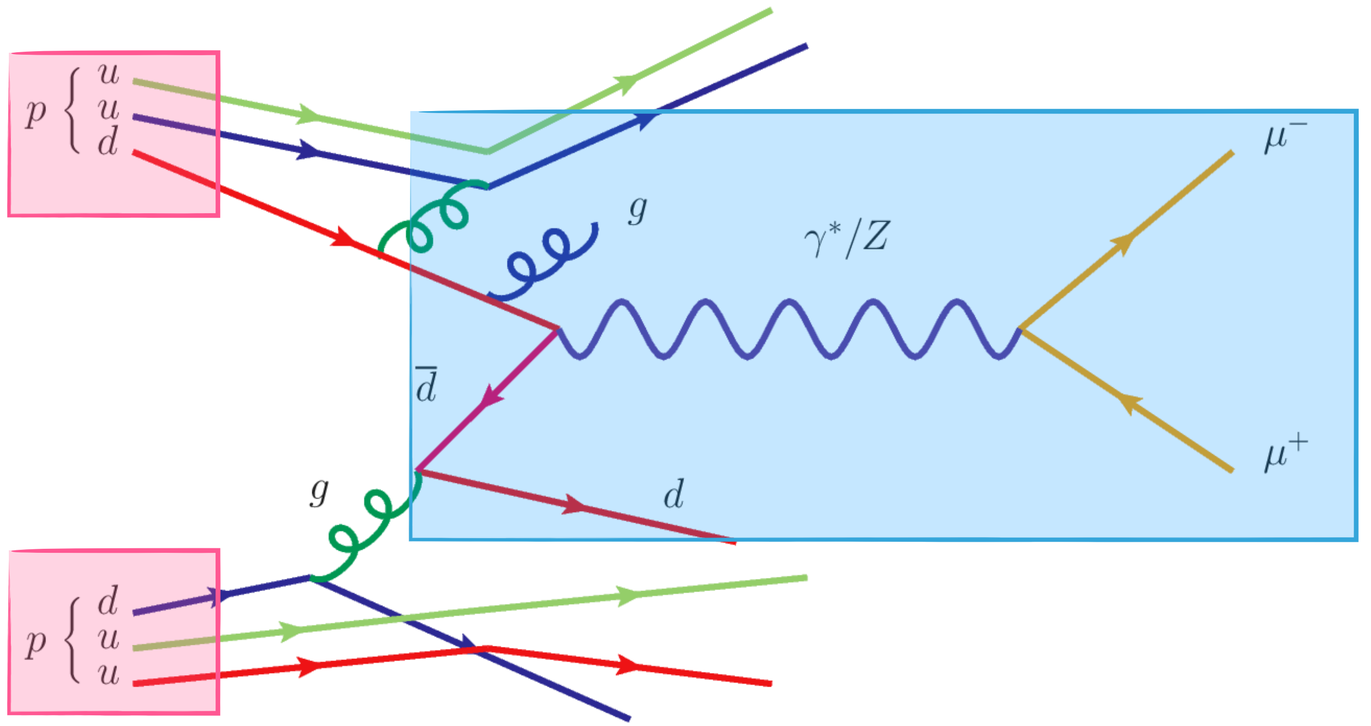
Elie Hammou

HEFT, CERN, June 2025

Hadron colliders, SMEFT and PDFs

Collinear factorization theorem

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



Parton Distribution Functions

Hard Scattering: Perturbative QCD + EW

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

PDFs overview

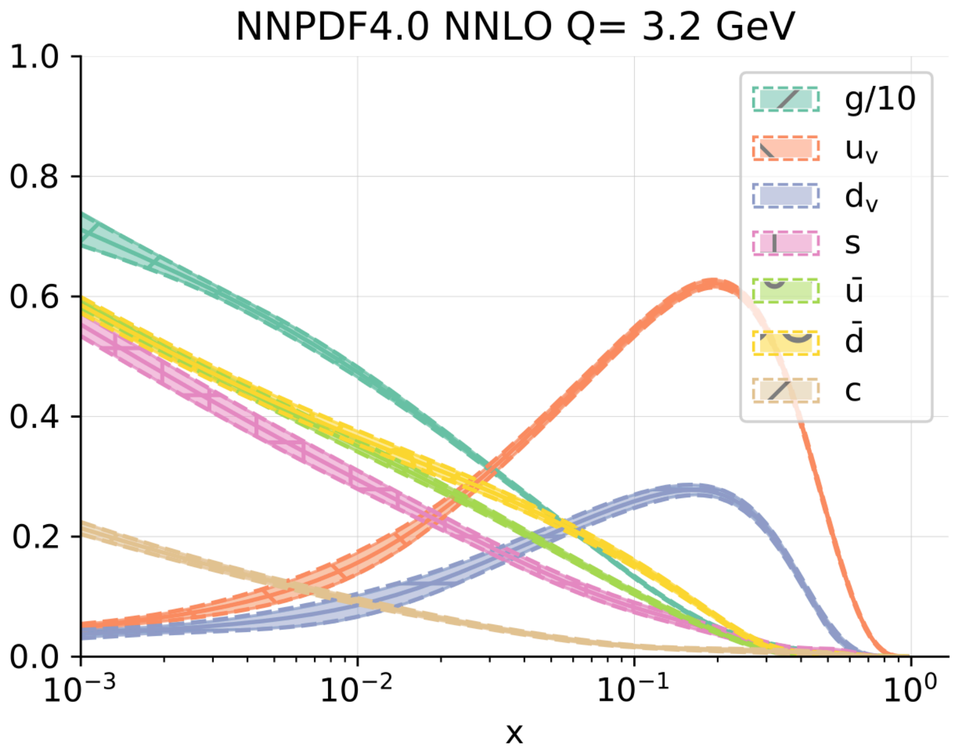
Hadron collider observable: $\sigma = f_1 \otimes f_2 \otimes \hat{\sigma}$

PDFs in a nutshell:

- describe proton's partonic content
- $f(x, Q)$
- Q dependance: DGLAP equation ✓
- x dependance: non-perturbative QCD ✗ (no lattice constrain)

➡ **Fitted from data**

Using NNPDF methodology

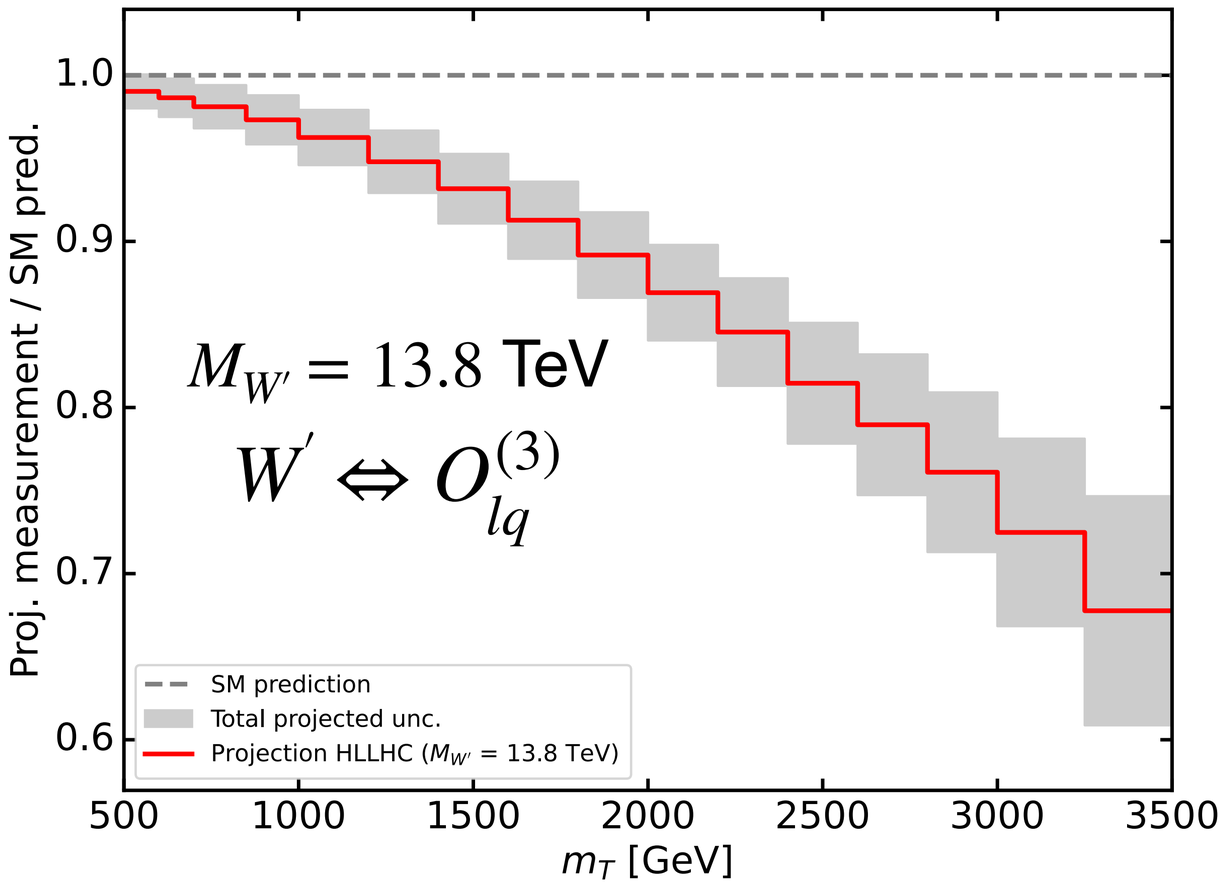


[Ball et al., NNPDF4.0, 2109.02653]

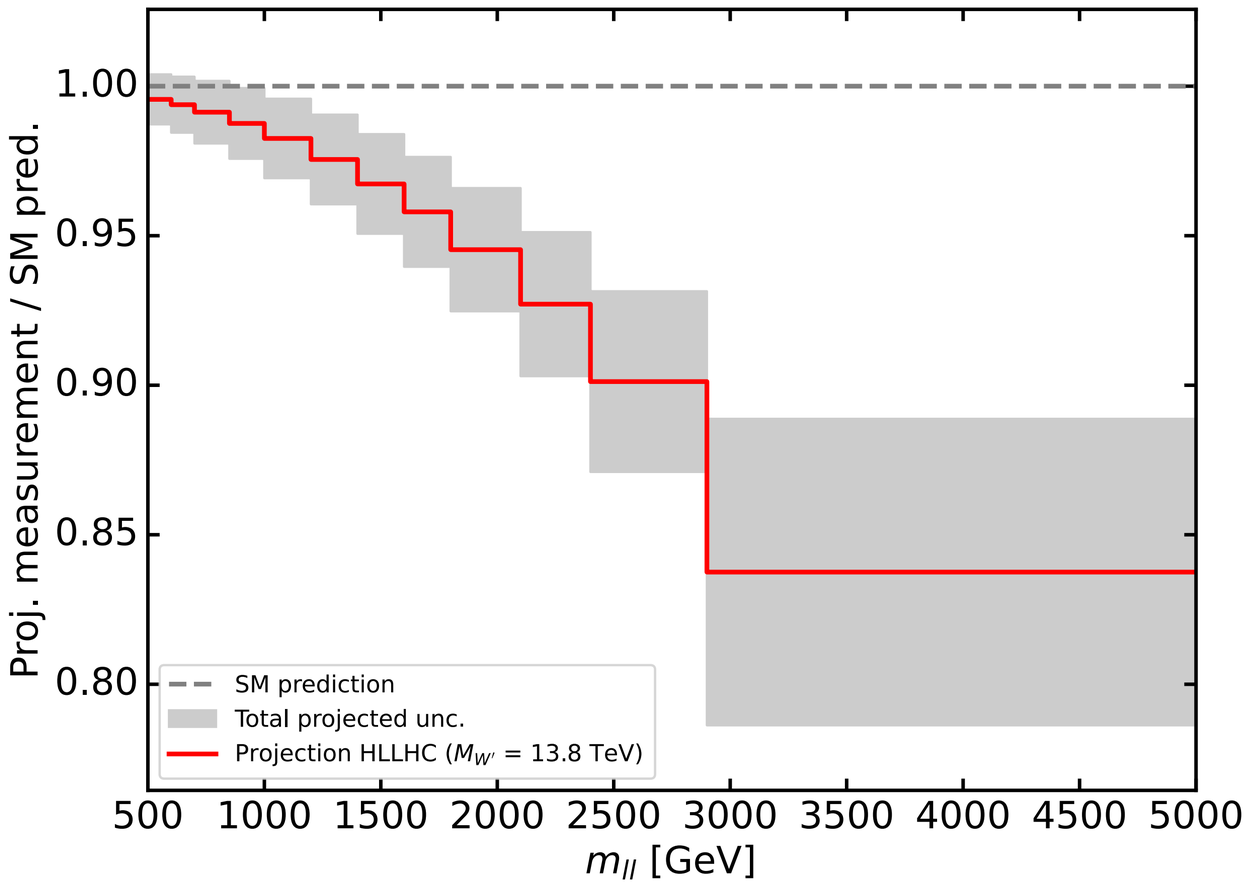
Injecting new physics in HL-LHC projections (DY)

Goal: recover it with a SMEFT fit

Charged current HMDY

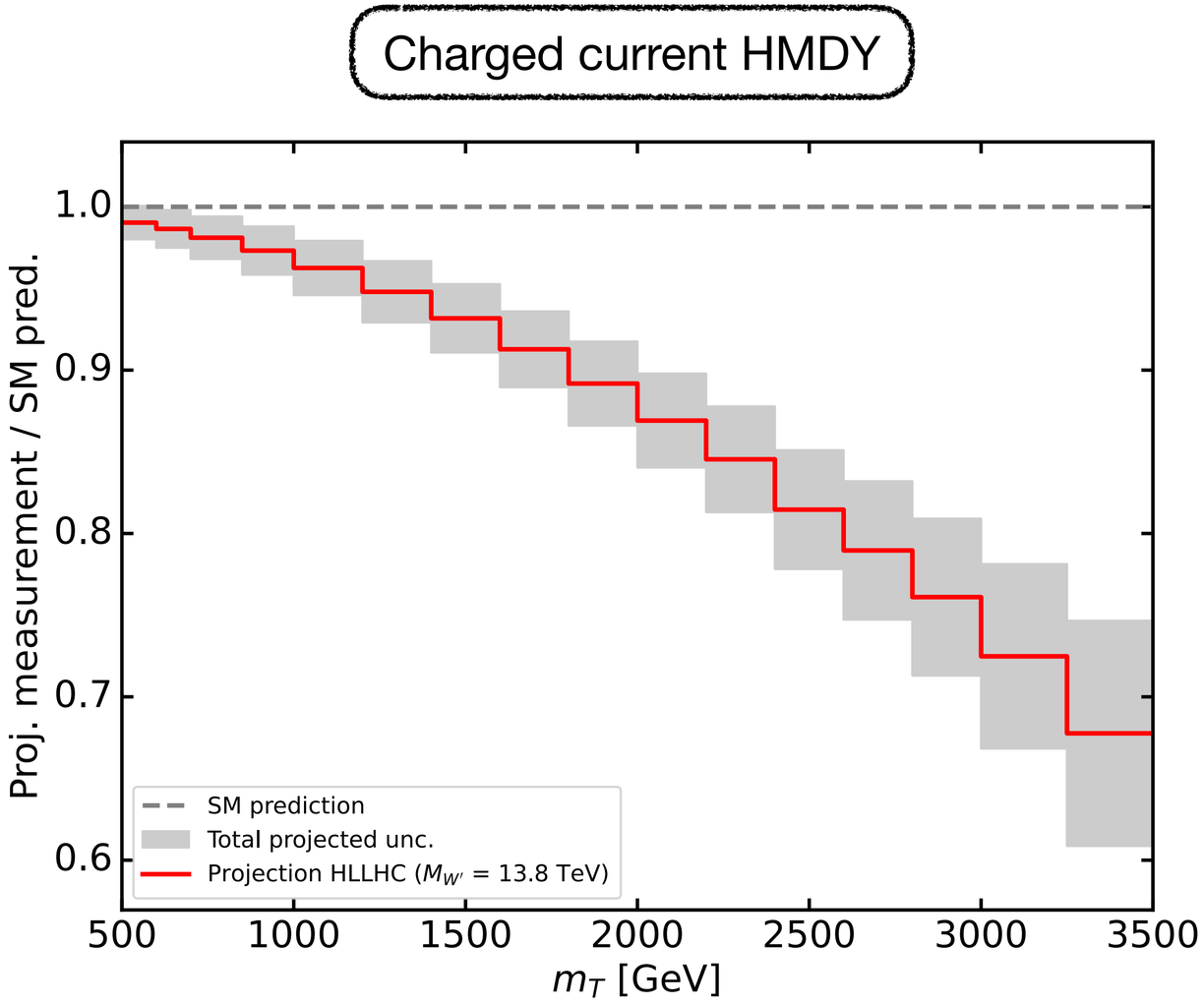


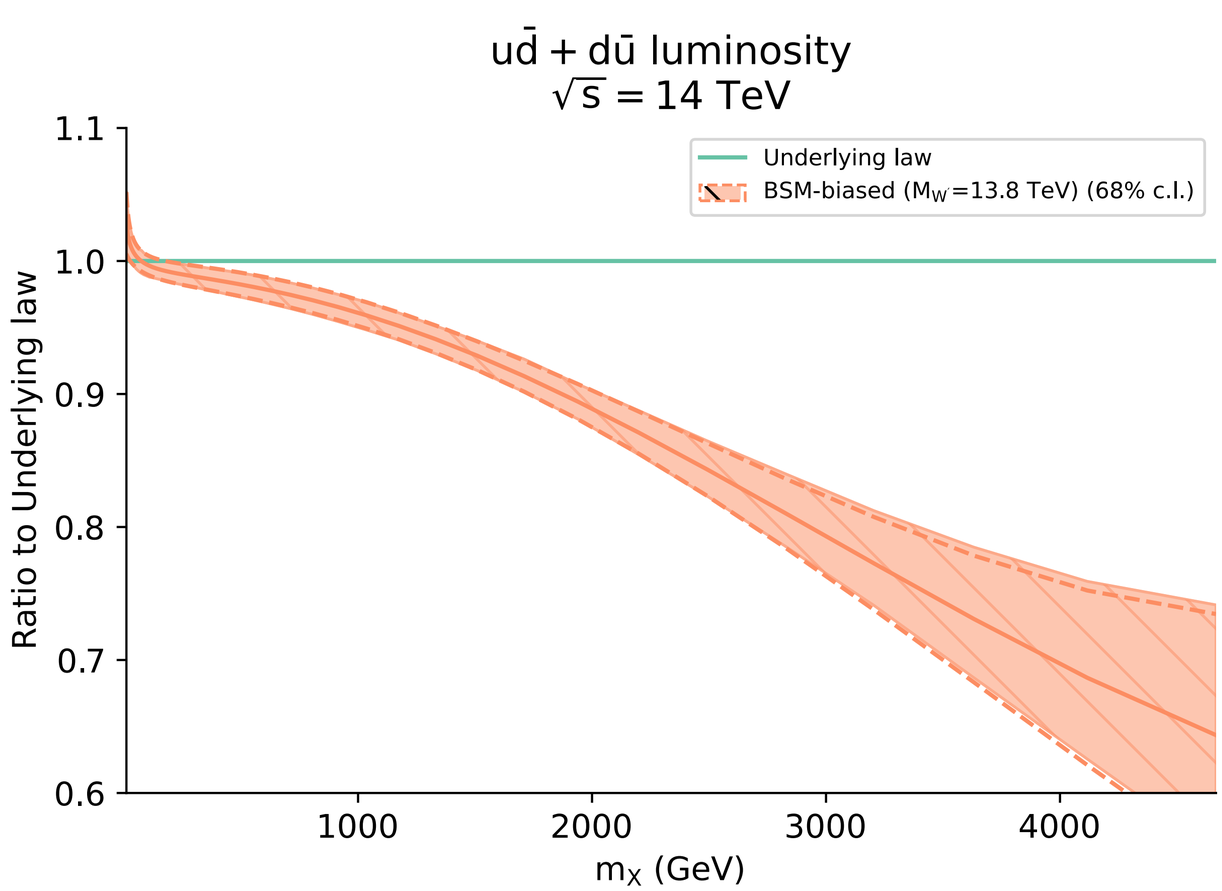
Neutral current HMDY



“BSM-biased” PDF

PDFs are mimicking the SMEFT corrections



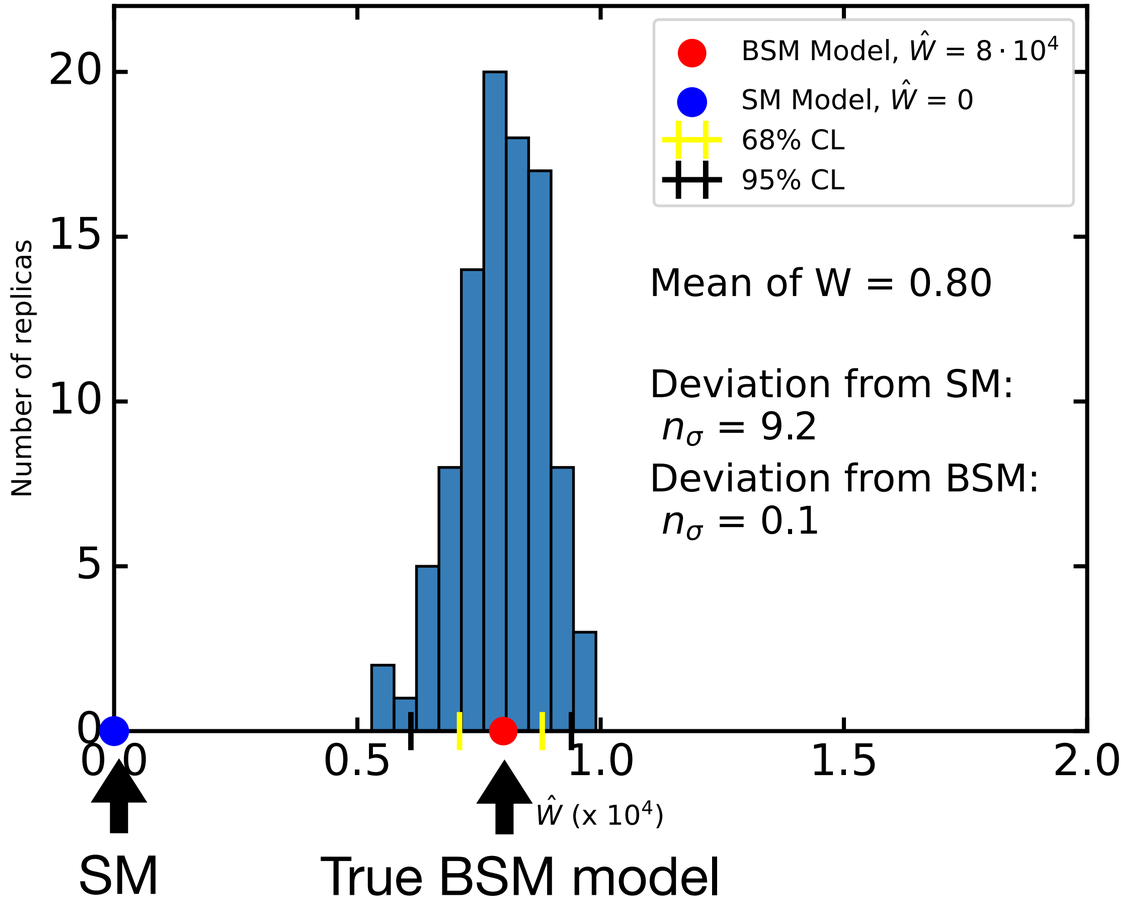


$$\mathscr{L}_{true} \otimes \hat{\sigma}_{BSM} \approx \mathscr{L}_{BSM-biased} \otimes \hat{\sigma}_{SM}$$

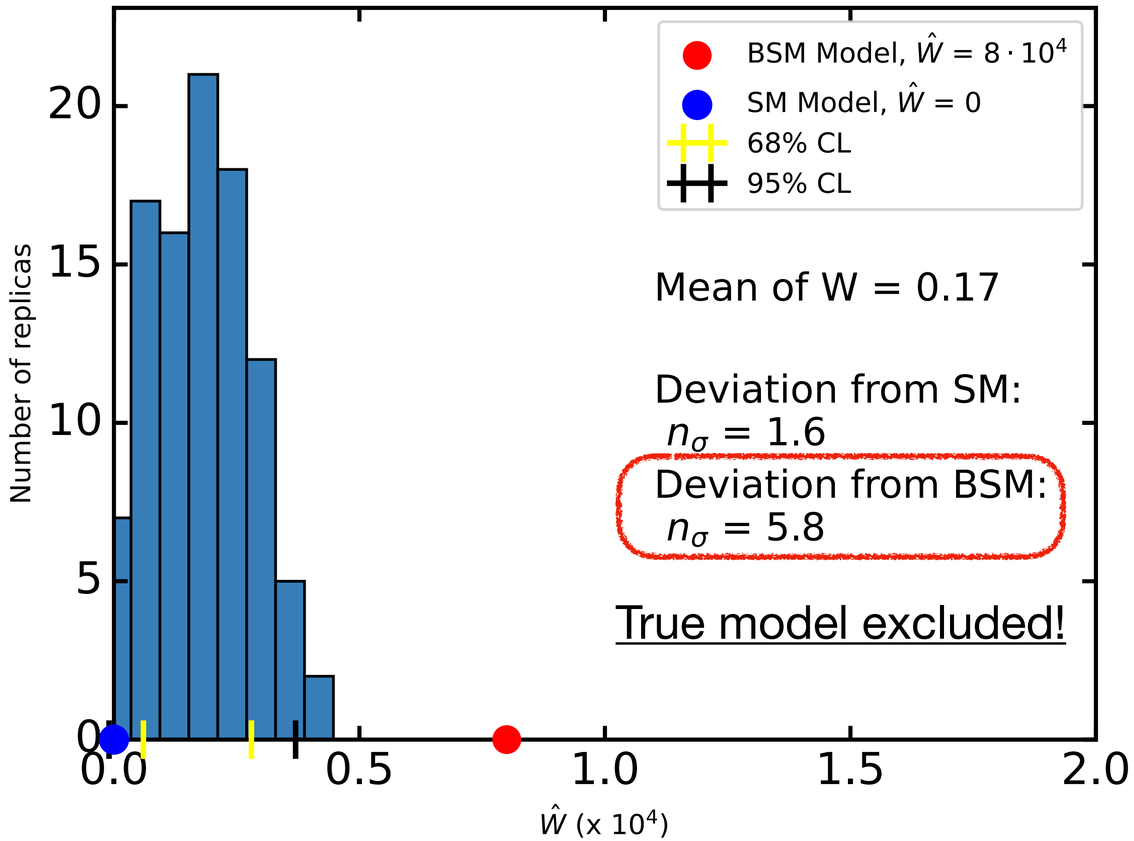
Missing new physics in Drell-Yan

Impact of the BSM-biased PDFs on SMEFT fits

SMEFT fit with true PDF



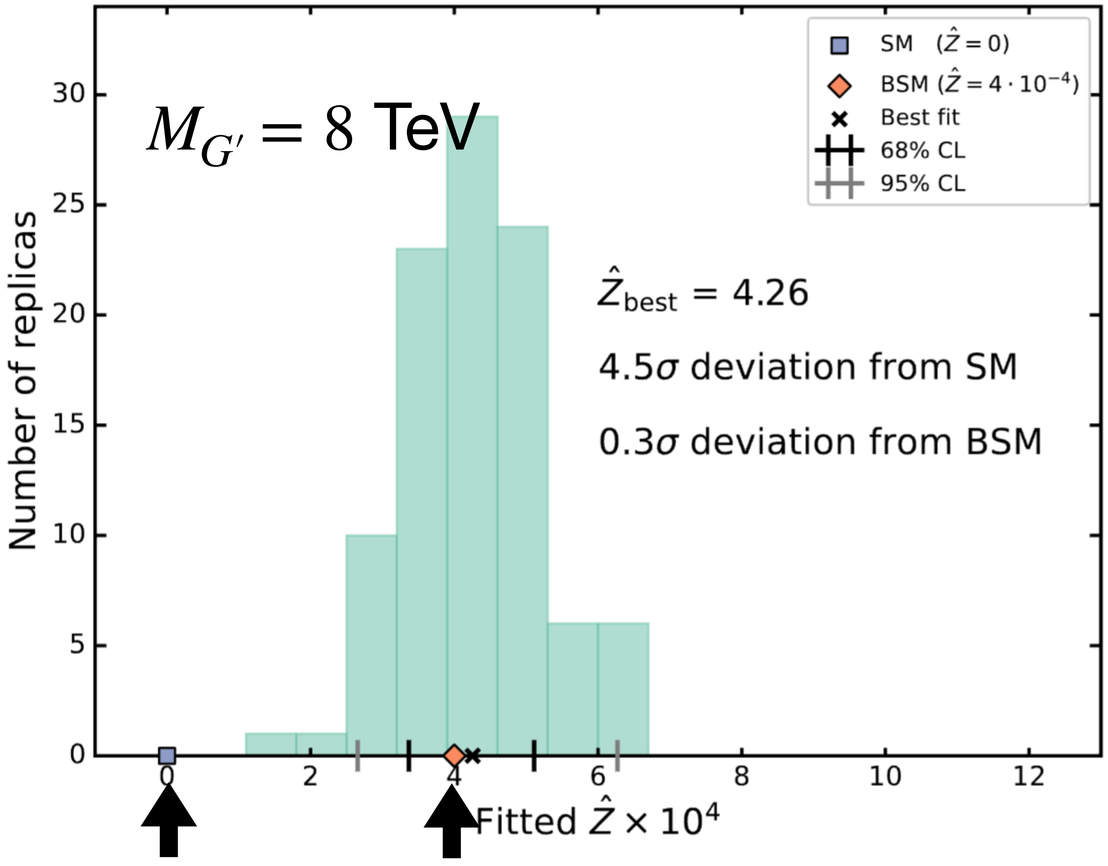
SMEFT fit with BSM-biased PDF



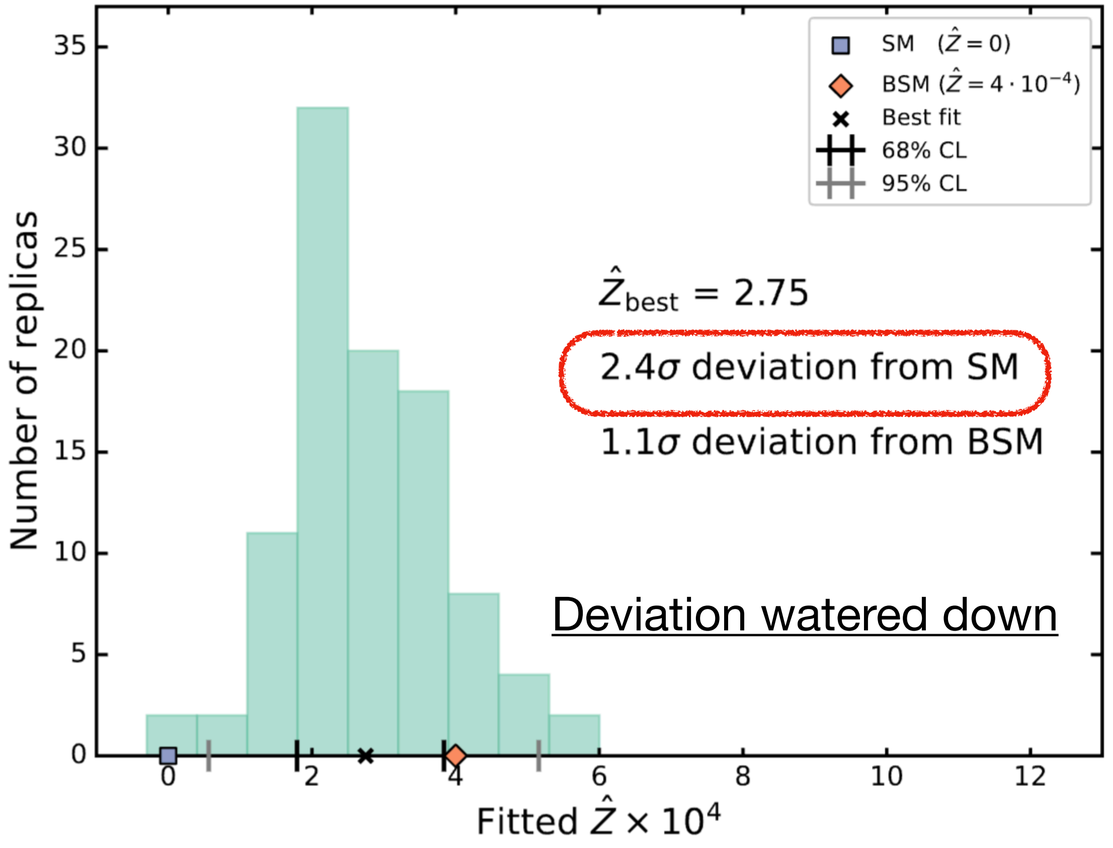
Diluting new physics deviation in the top sector

Same exercise with heavy gluon G' in ttbar at HL-LHC

SMEFT fit with true PDF



SMEFT fit with BSM-biased PDF



Separate and simultaneous fits of PDF and SMEFT

Comparison of the methodologies

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:

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

➡ $\bar{c}$

Need to make sure SM assumption is true

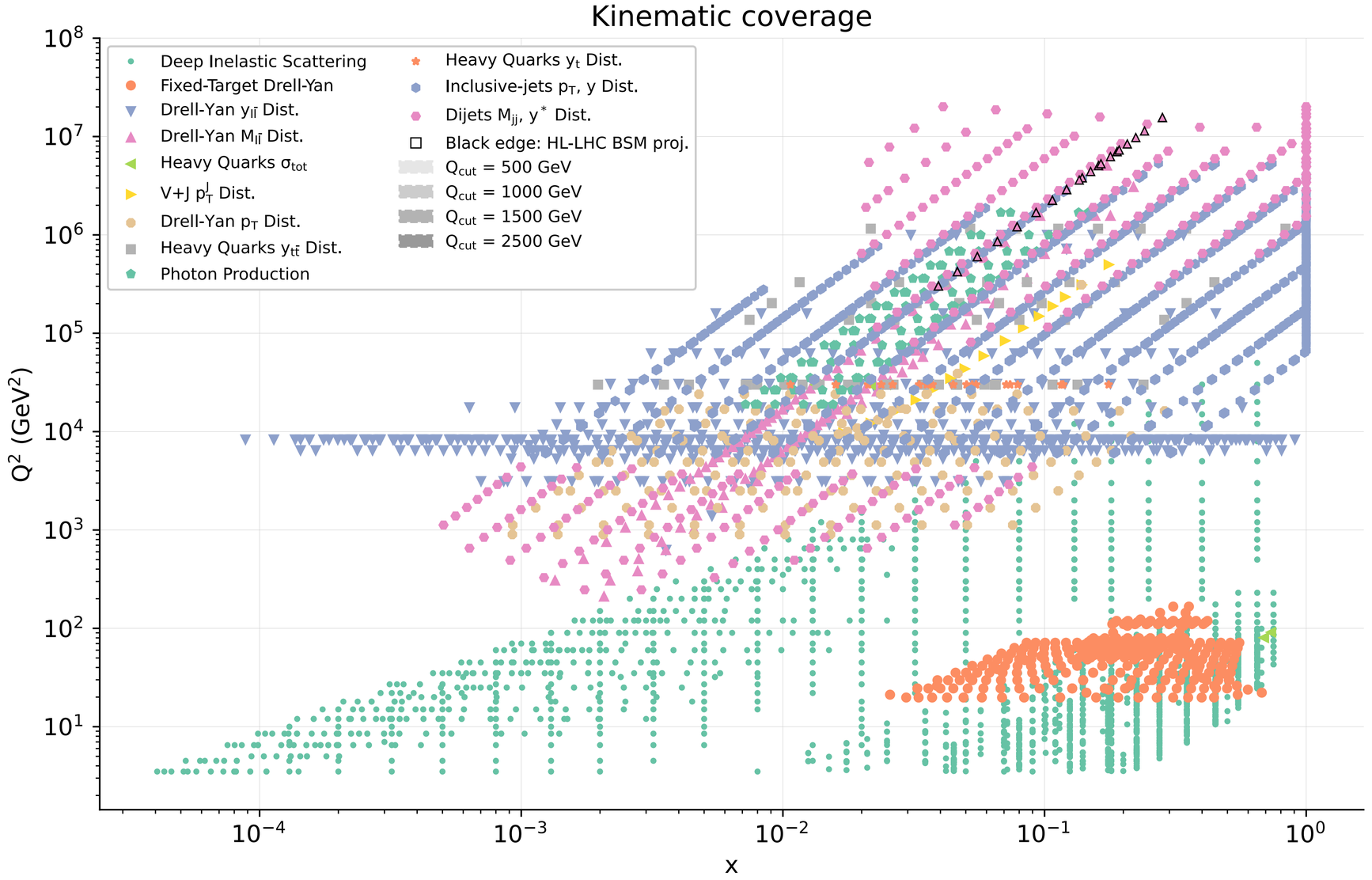
Simultaneous fits

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

➡ $\{\bar{\theta}, \bar{c}\}$

Removes assumption-based bias

Safe separate fits: “conservative” PDFs

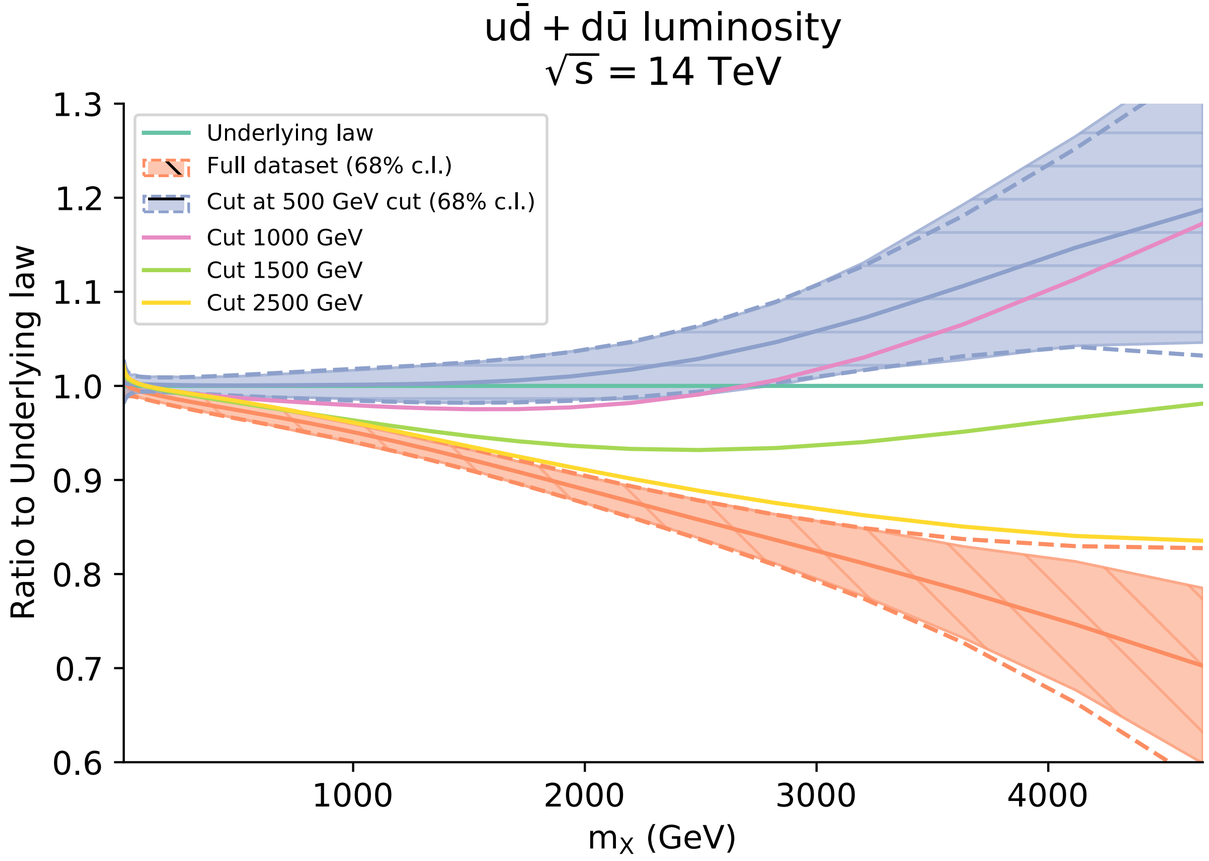


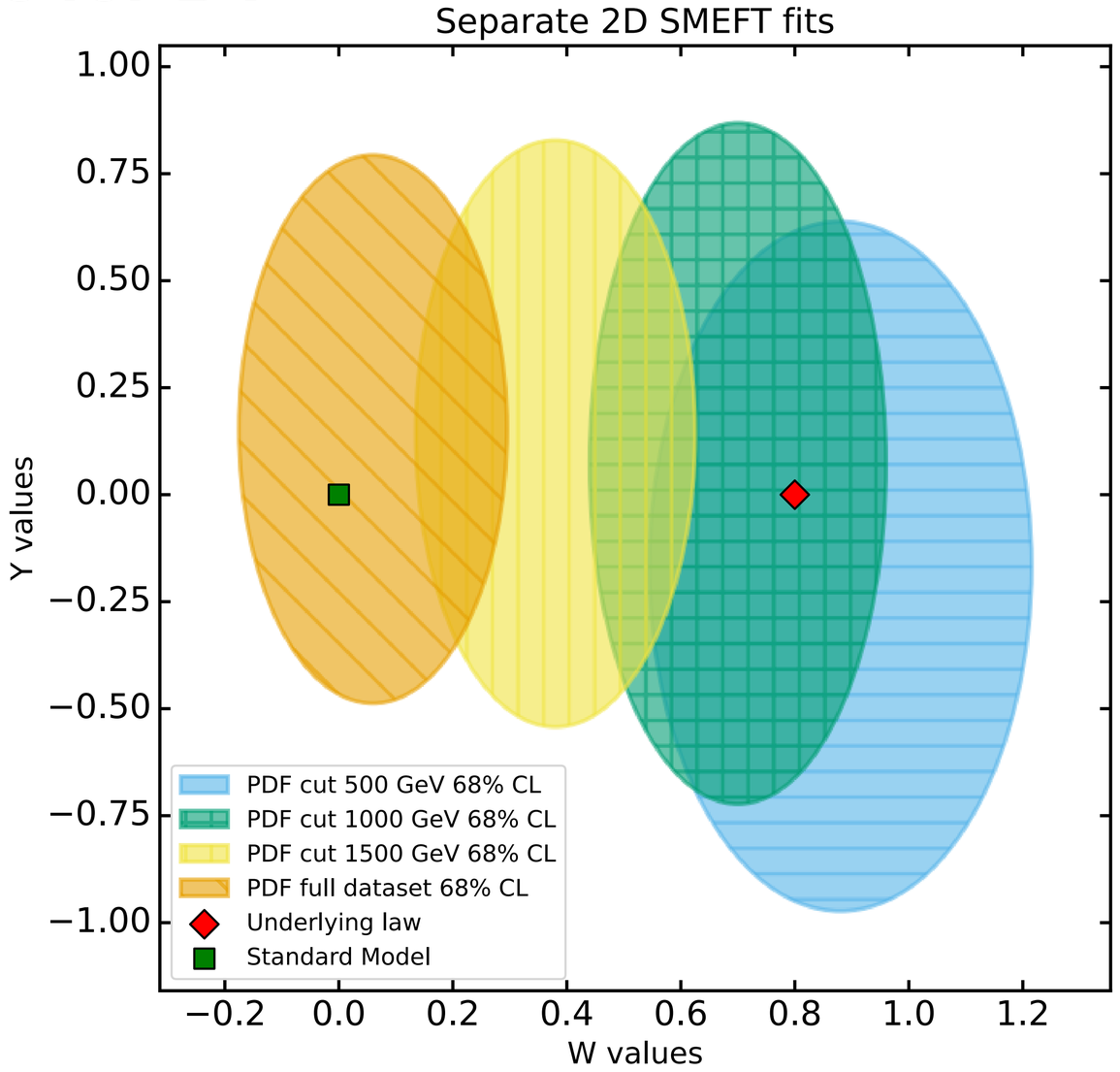
Agnostic energy cuts

“Conservative” separate fits

[PBSP, forthcoming]

Impact on PDF and SMEFT bounds for DY





Simultaneous fit of PDF and SMEFT

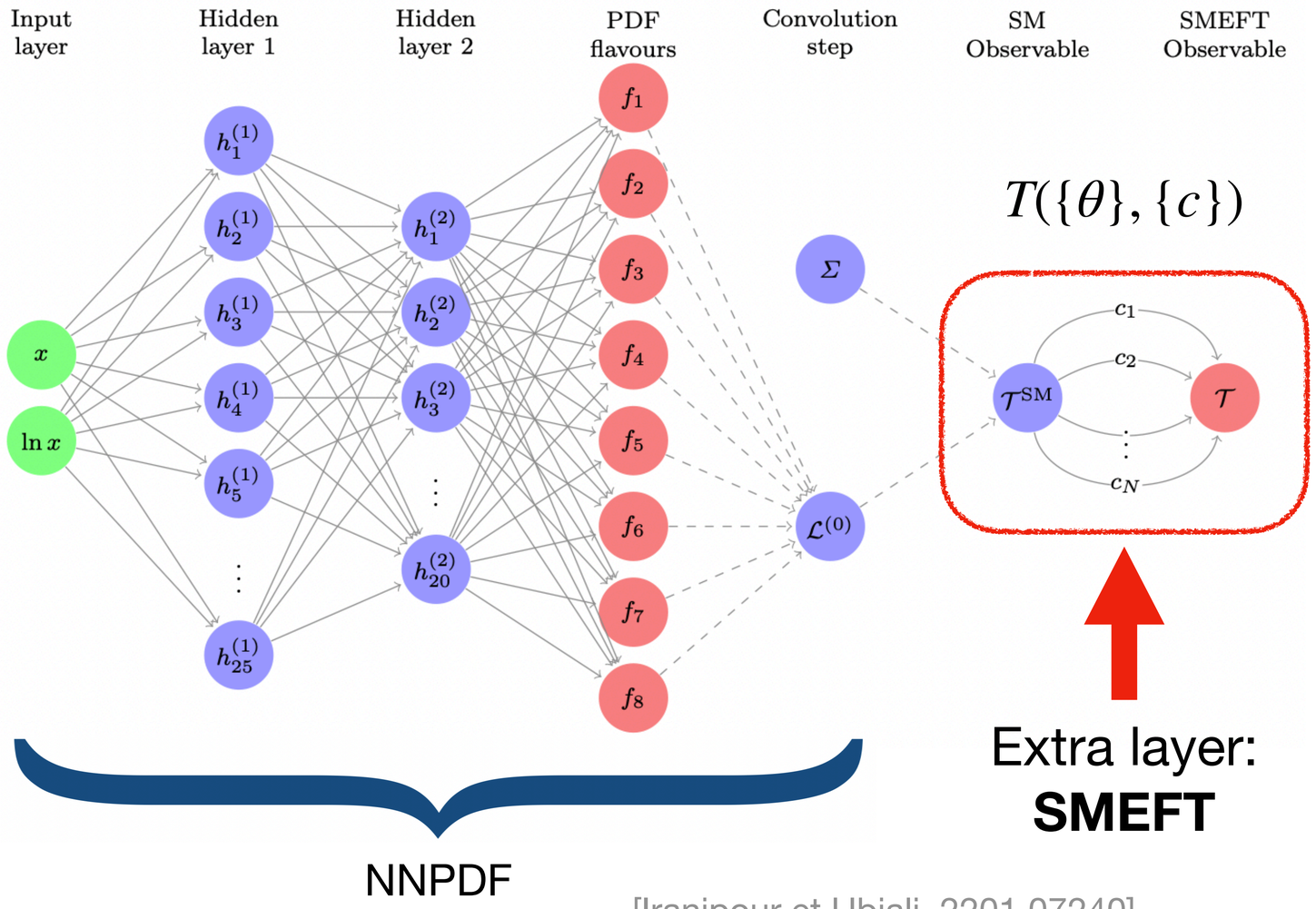
Presentation of the tool: SIMUnet

SIMUnet:

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

  [PBSP, 2402.03308]

- Fits PDFs and WC simultaneously
- Performs BSM closure tests

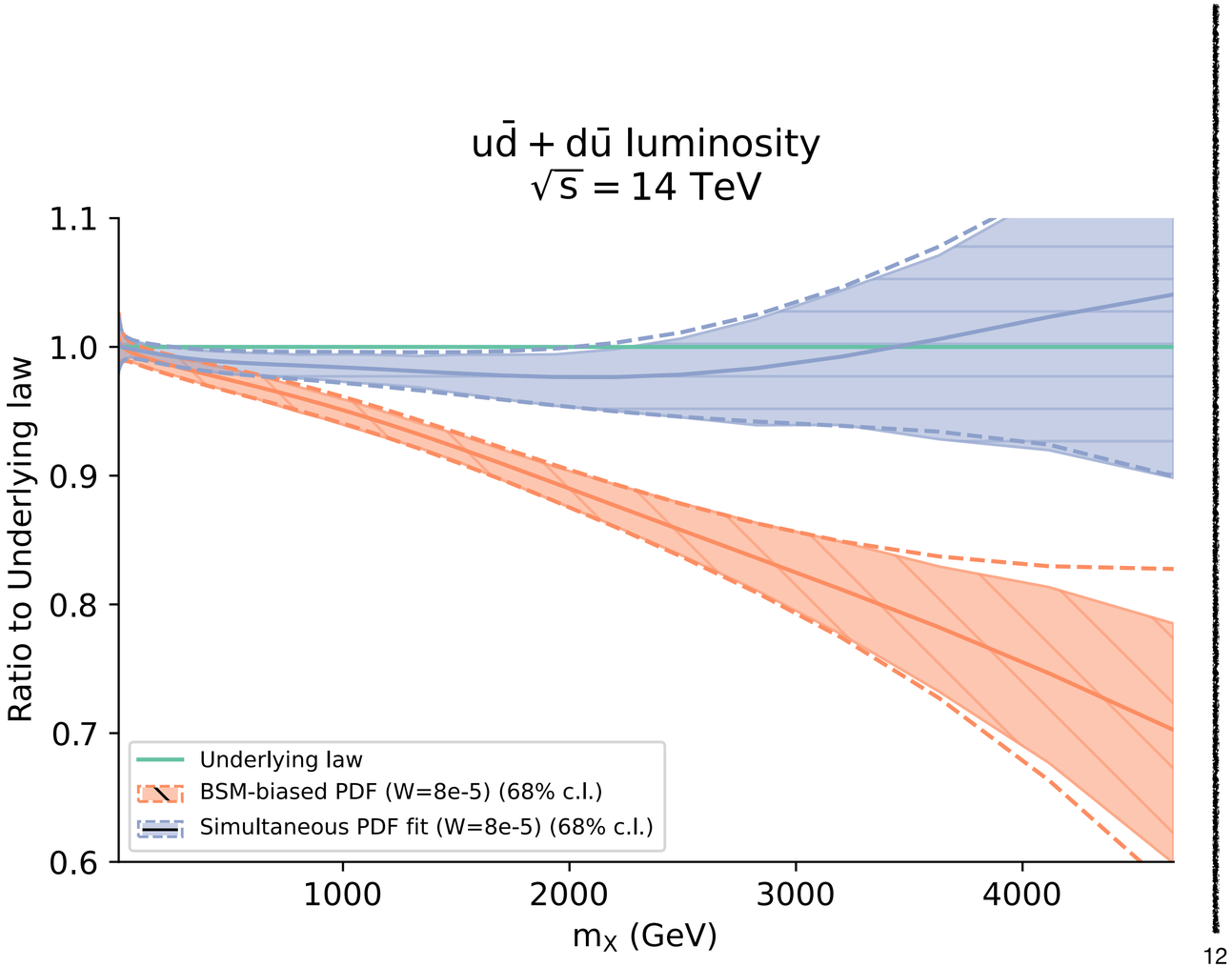


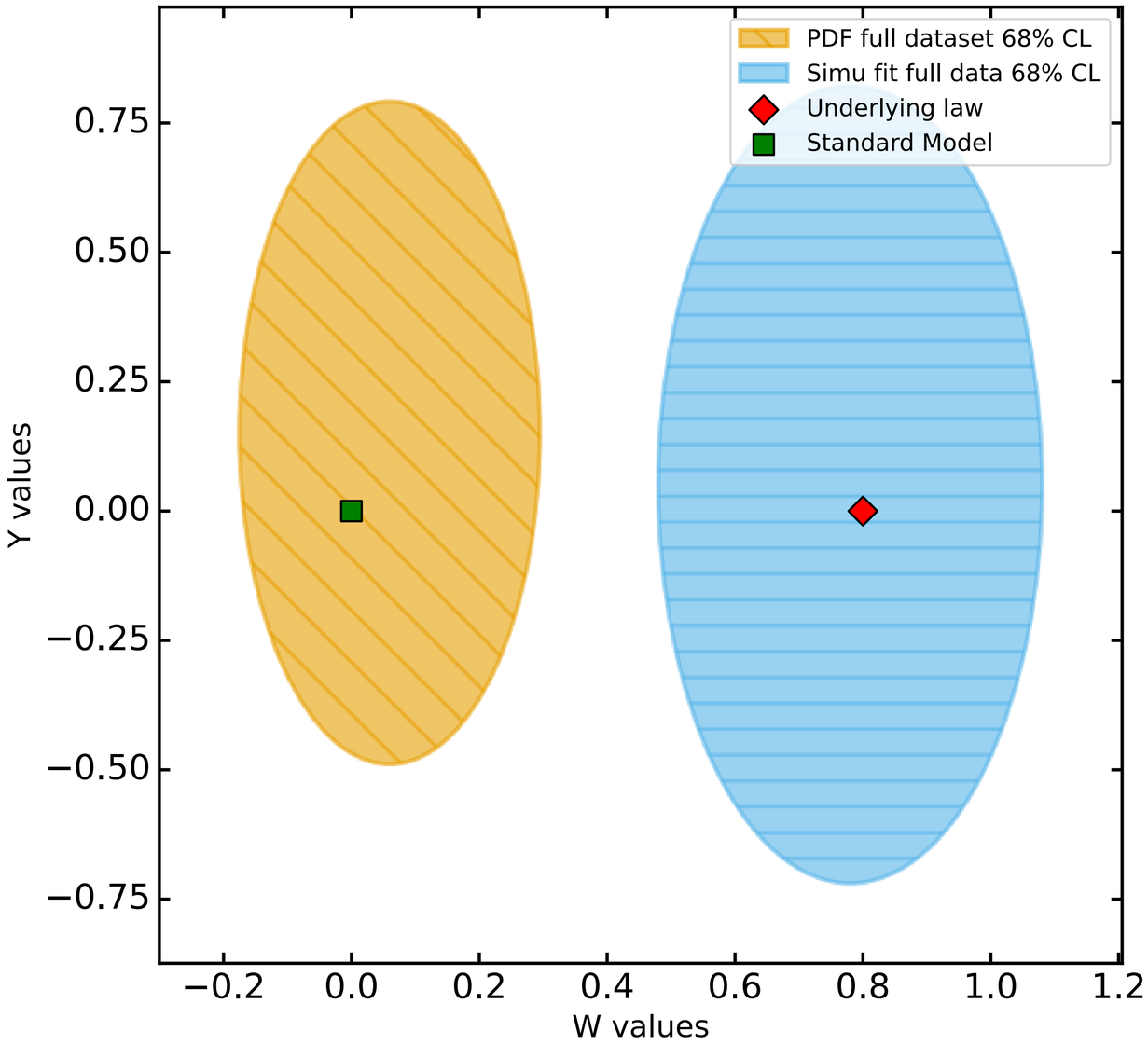
[Iranipour et Ubiali, 2201.07240]

Simultaneous fit of PDF and SMEFT

[PBSP, forthcoming]

Marginal distributions for DY

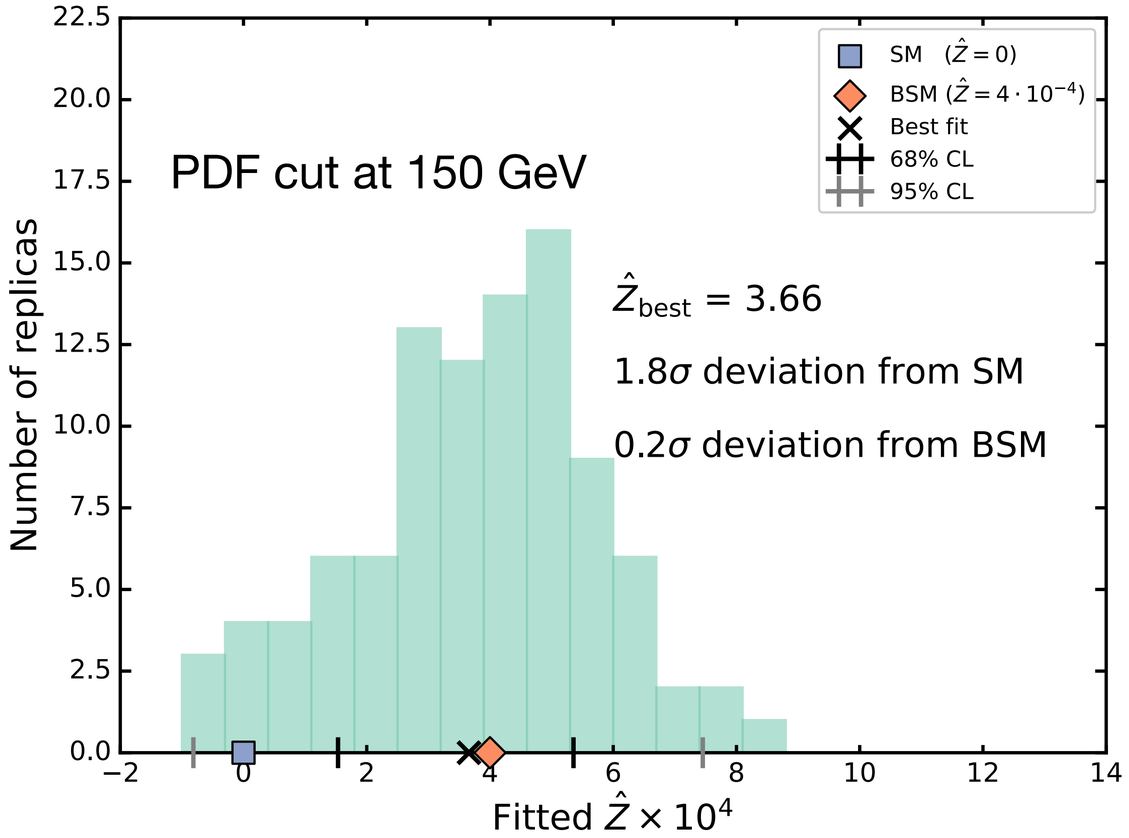




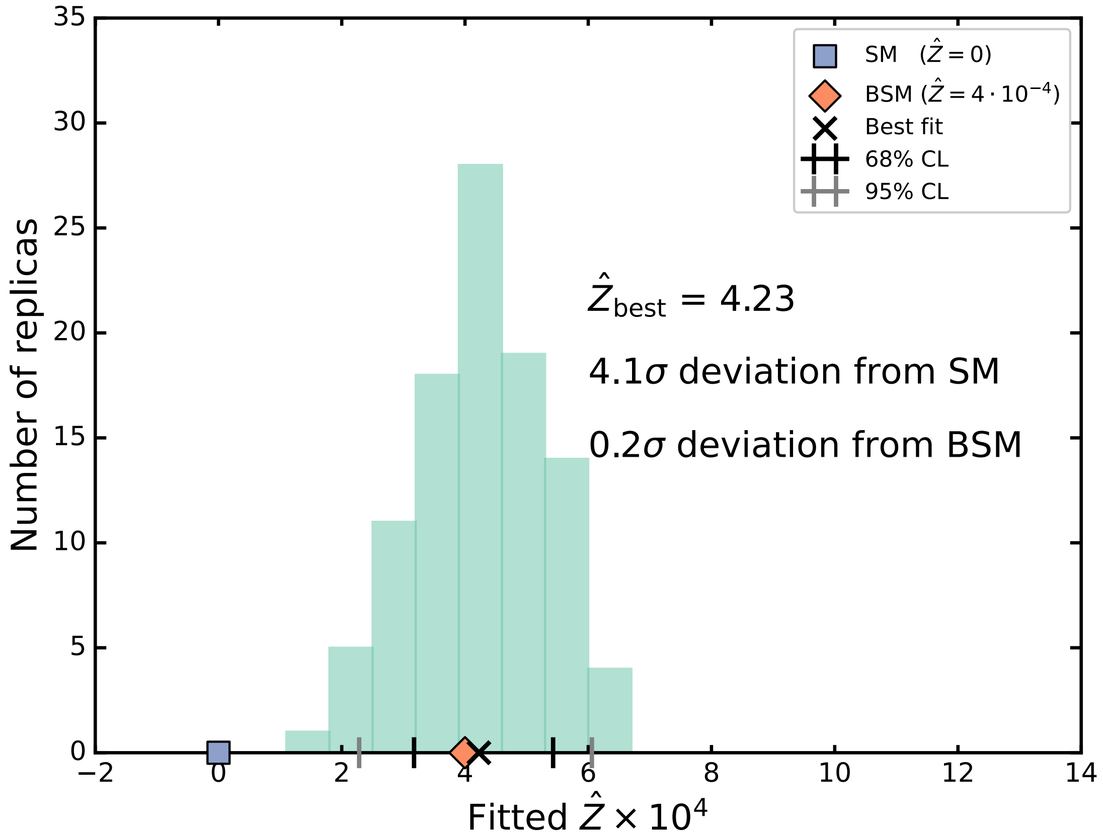
SMEFT fits comparison in the top sector

Same exercise with heavy gluon in ttbar at HL-LHC

Conservative separate fit



Simultaneous fit



Comparing conservative and simultaneous fits

Conservative separate fits

Pros:

- Easier
- Less parameters per fit

Cons:

- Difficult to figure out optimal cutoff (manageable)
- **Cannot use precise high-energy observables to constrain PDFs**

Simultaneous fits

Pros:

- Entire dataset constrains PDF and SMEFT
- High-energy observables constrain PDF

Cons:

- More parameters -> more uncertainty (manageable)
- **Risk to absorb SM error as SMEFT signal**

Summary and outlook

- Signs of new physics fitted away in PDF parametrisation
  - Exclude true underlying law (DY) / water down BSM deviation (ttbar)
  - **Must be aware of this source of bias**
- Practical recommendations:
  - Conservative PDFs for SMEFT fits
  - Simultaneous fits of PDFs and SMEFT (**SIMUnet** tool already available)
- Ongoing work:
  - Ongoing study on real jet data [Greljo, Hammou, Merlotti, Smolkovic, Ubiali, forthcoming]
  - Developing a bayesian framework [Costantini, Moore, Mantani, Schutze, Ubiali, forthcoming (PDF)]
  - Allowing choice in PDF modelisation [Hammou, ter Hoeve, Shutze, in progress (SMEFT)]

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

Thank you for your attention!

Extra slides

Potential pitfalls of PDF fitting

- No theory constraint available (e.g. from lattice)

  ➡ Only determined from measurements

- Very flexible parametrisation by necessity

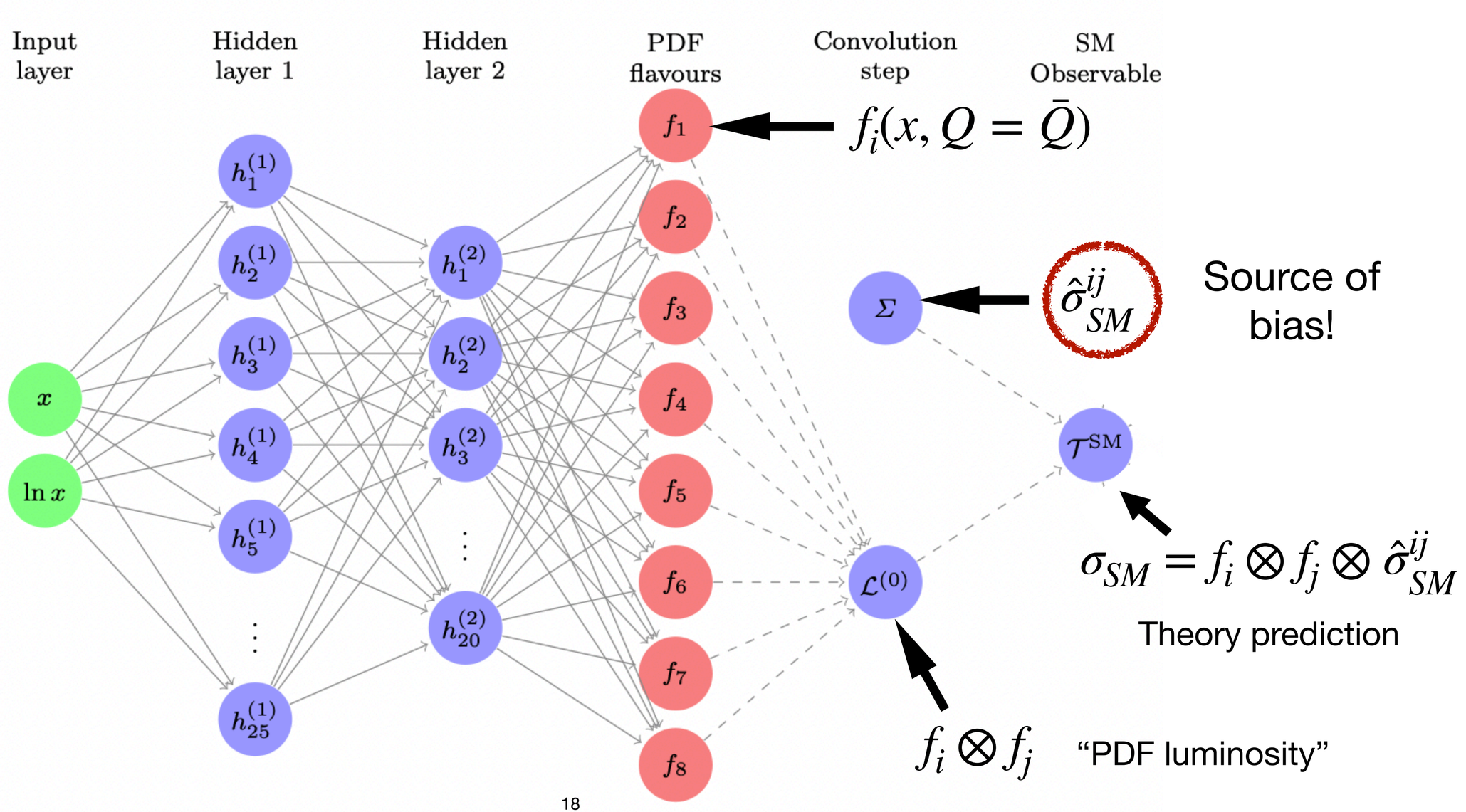
  ➡ Can mimic deviations in the observables

- Assumes SM in theory predictions

- Large-x PDF mostly constrained by high-energy measurements

  ➡ Risk to absorb BSM deviations!

NNPDF fitting methodology



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]

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

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

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:

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

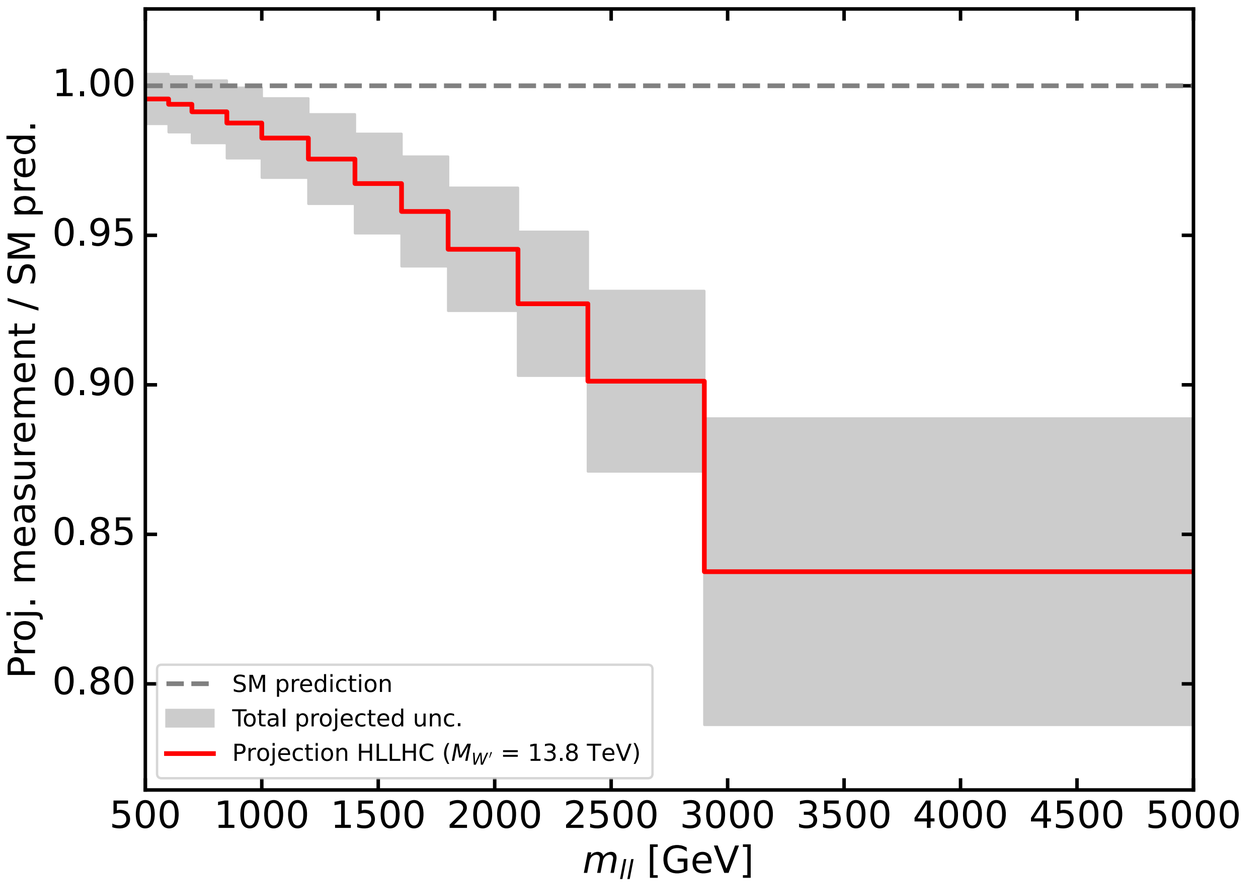
→ $$\mathscr{L}_{SMEFT}^{W'} = \mathscr{L}_{SM} - \frac{g^2 \hat{W}}{2m^2_W} J_L^{a,\mu} J^a_{L,\mu} \qquad \hat{W} = \frac{g^2_{W'}}{g^2} \frac{m^2_W}{M^2_{W'}} \propto \frac{c}{\Lambda^2}$$

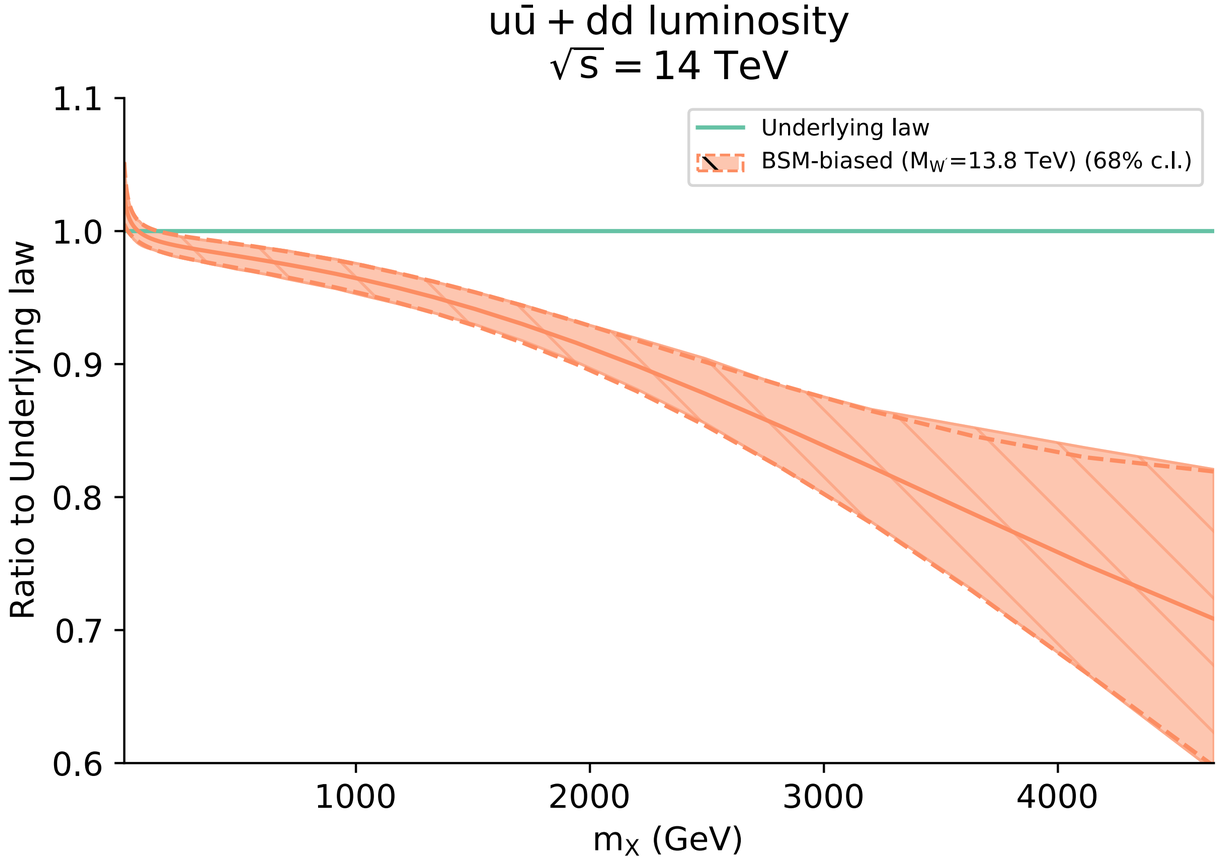
New physics parameter

“BSM-biased” PDF

PDFs are mimicking the SMEFT corrections

Neutral current HMDY

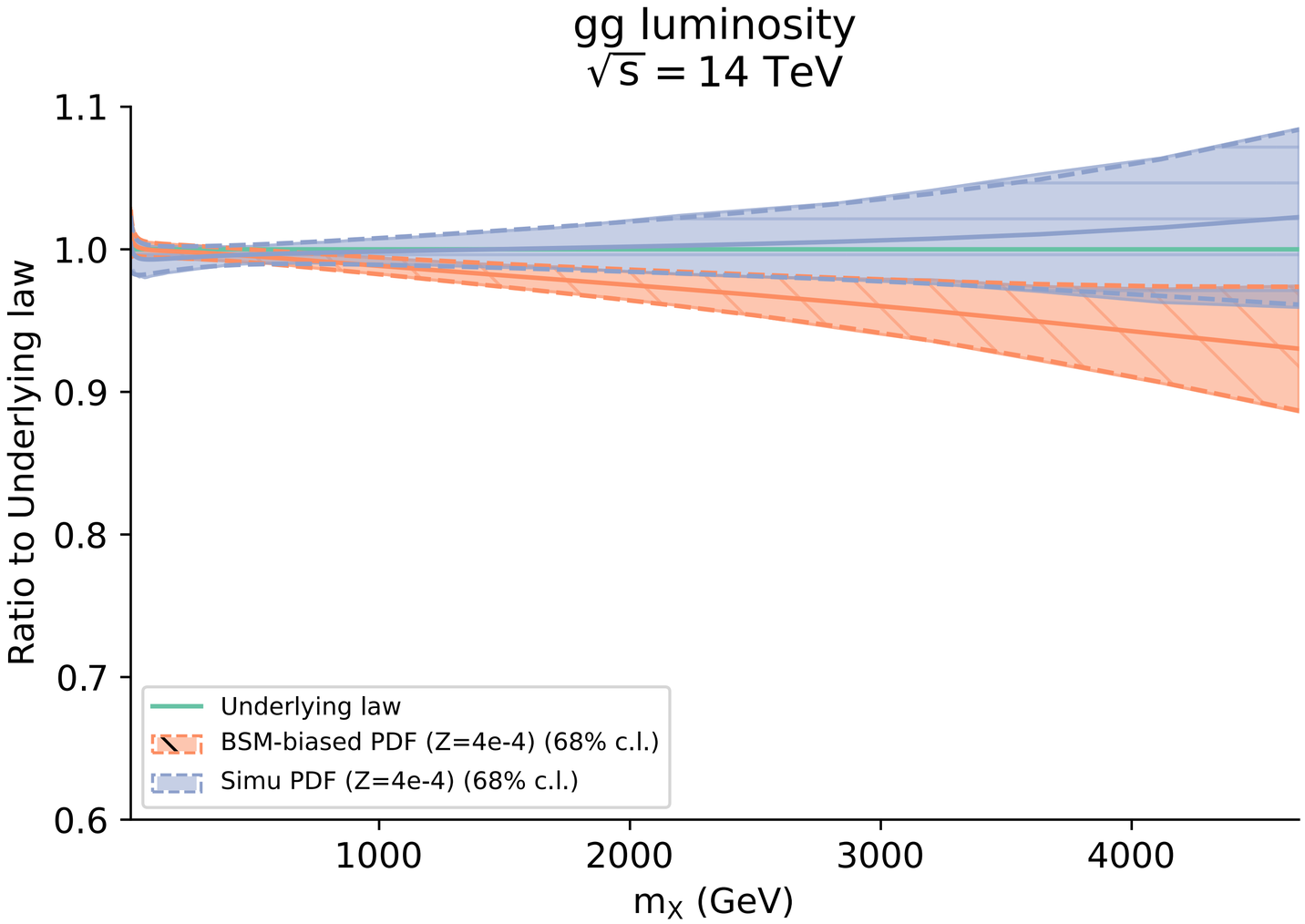




$$\mathscr{L}_{true} \otimes \hat{\sigma}_{BSM} \approx \mathscr{L}_{BSM-biased} \otimes \hat{\sigma}_{SM}$$

“BSM-biased” PDF for ttbar

PDFs are mimicking the SMEFT corrections



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

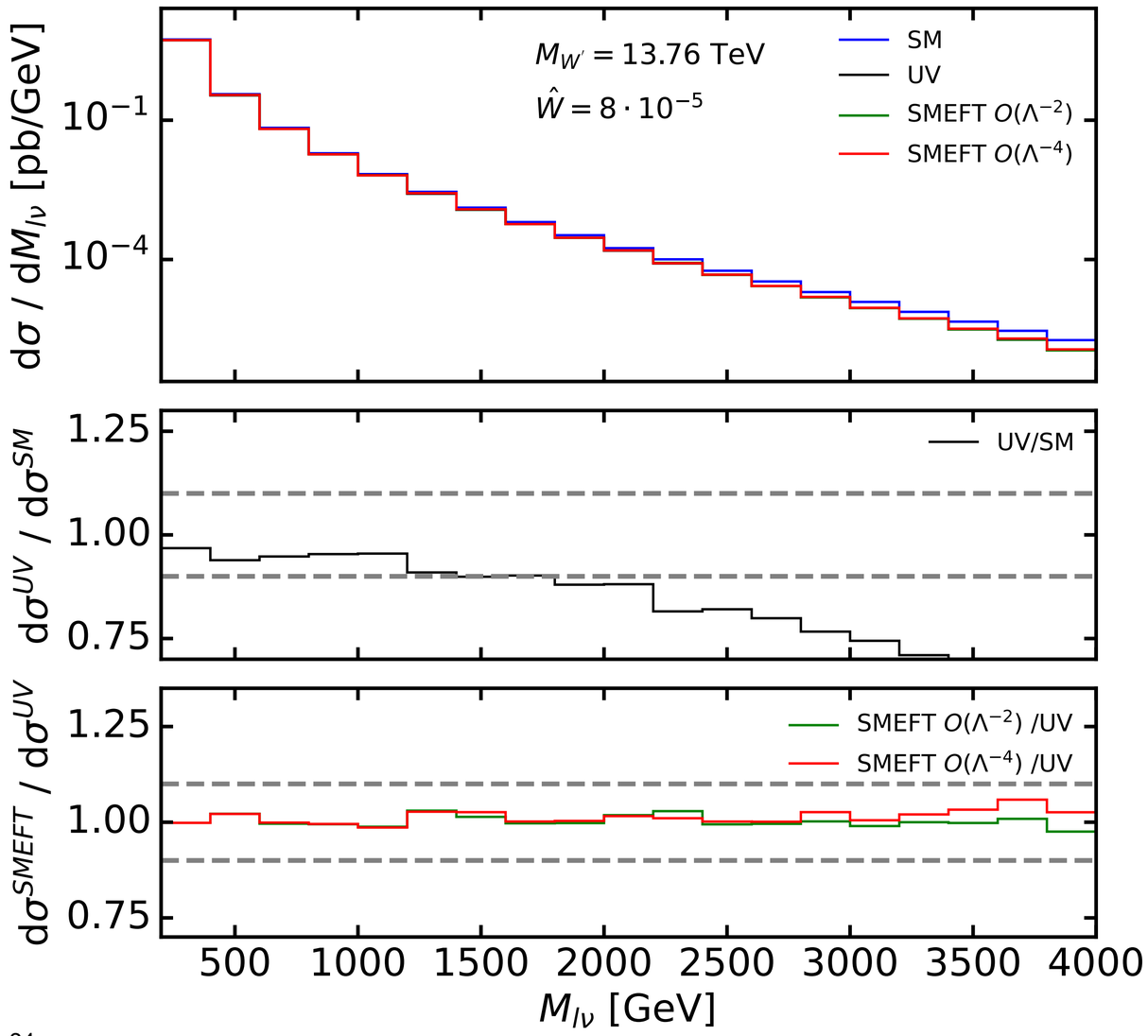
Generation of the pseudodata

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

➡ Impacts CC and NC Drell-Yan

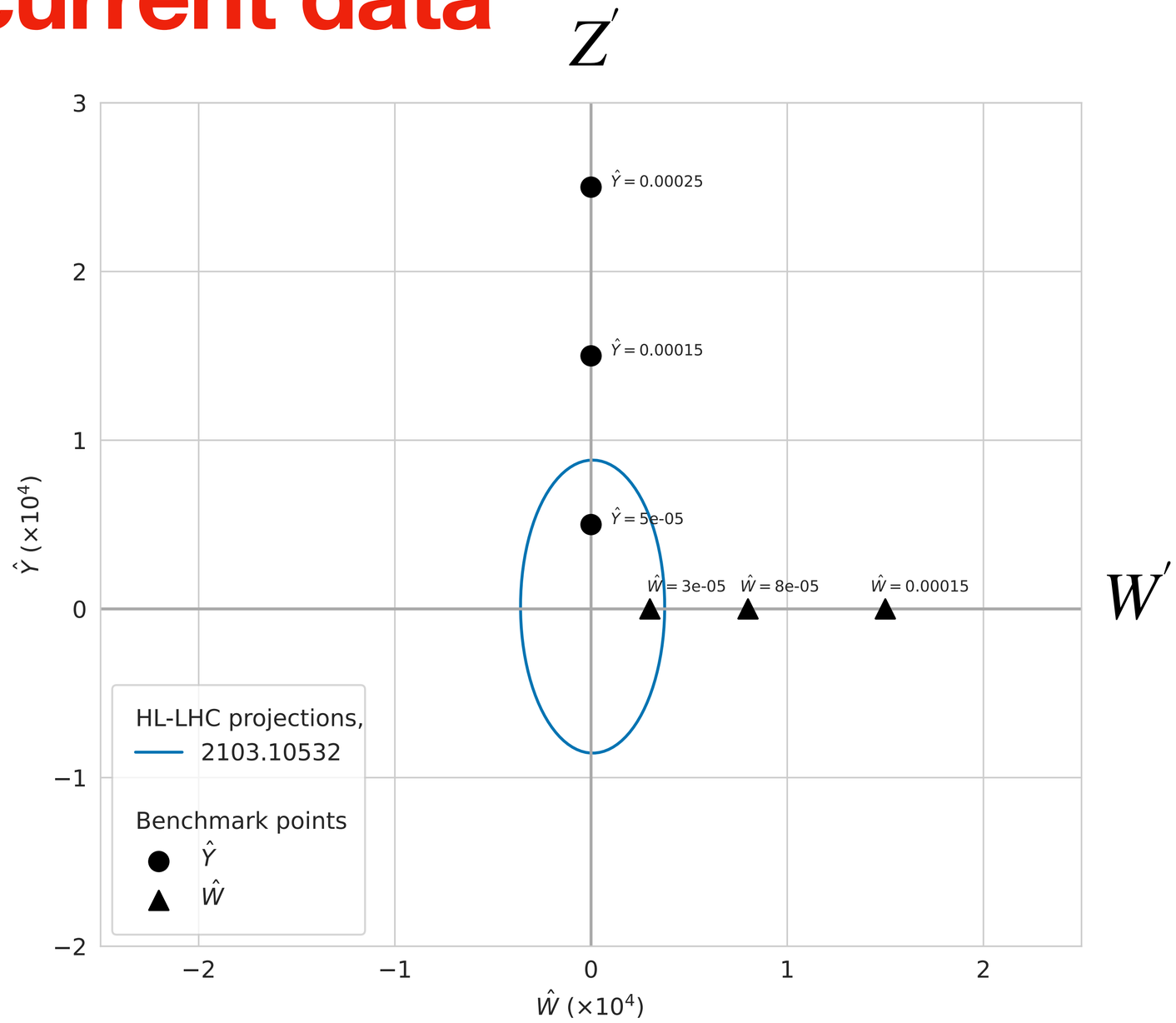
$$\hat{W} = \frac{g_{W'}^2}{g^2}\frac{m_W^2}{M_{W'}^2} \propto \frac{c}{\Lambda^2}$$

$$\hat{W} \leftrightarrow M_{W'} \quad (g_{W'} = 1)$$



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)

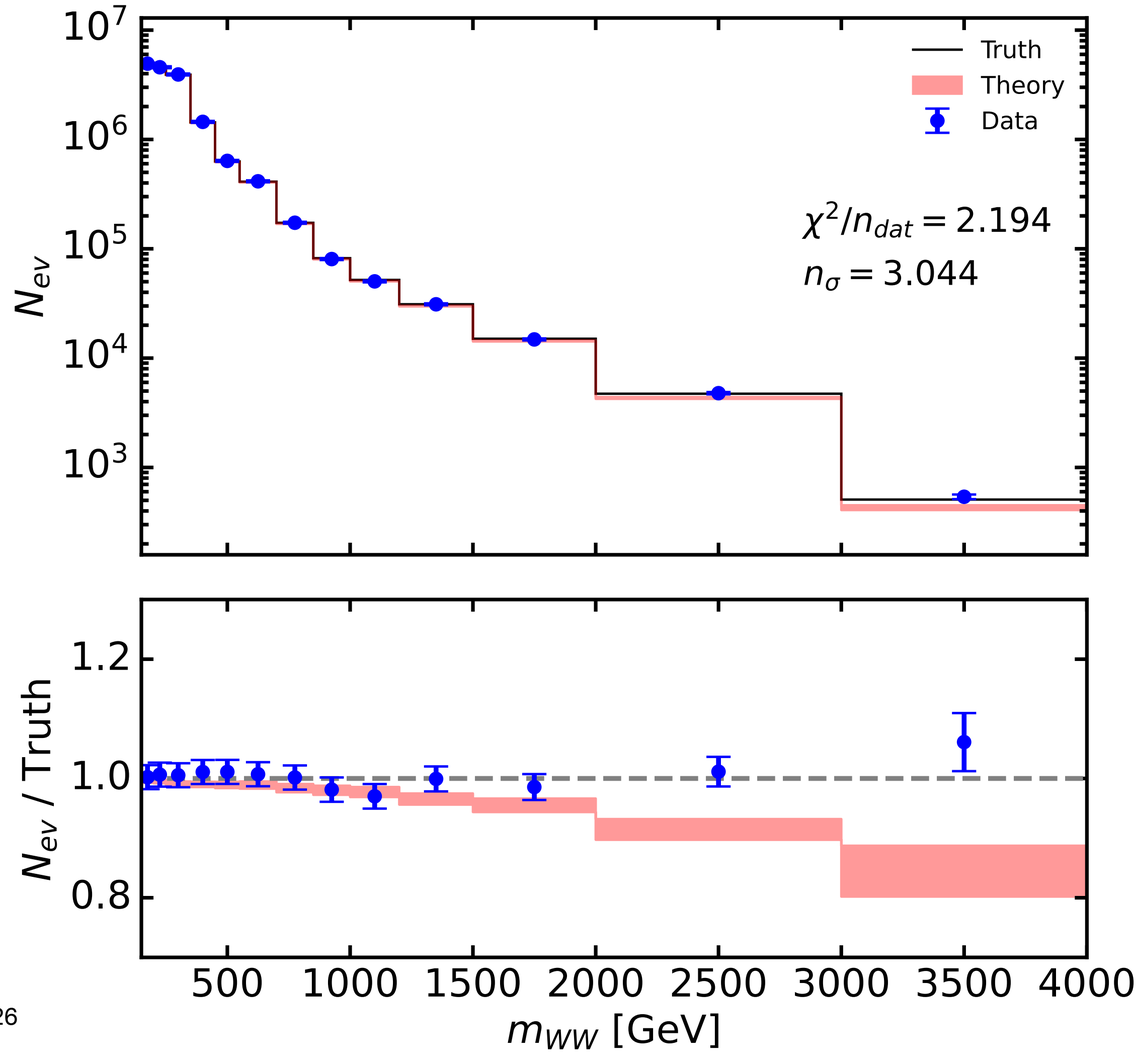
➡ Fake deviation in other sectors

Also seen in:

WH, WZ, ZH production

HL-LHC Projections

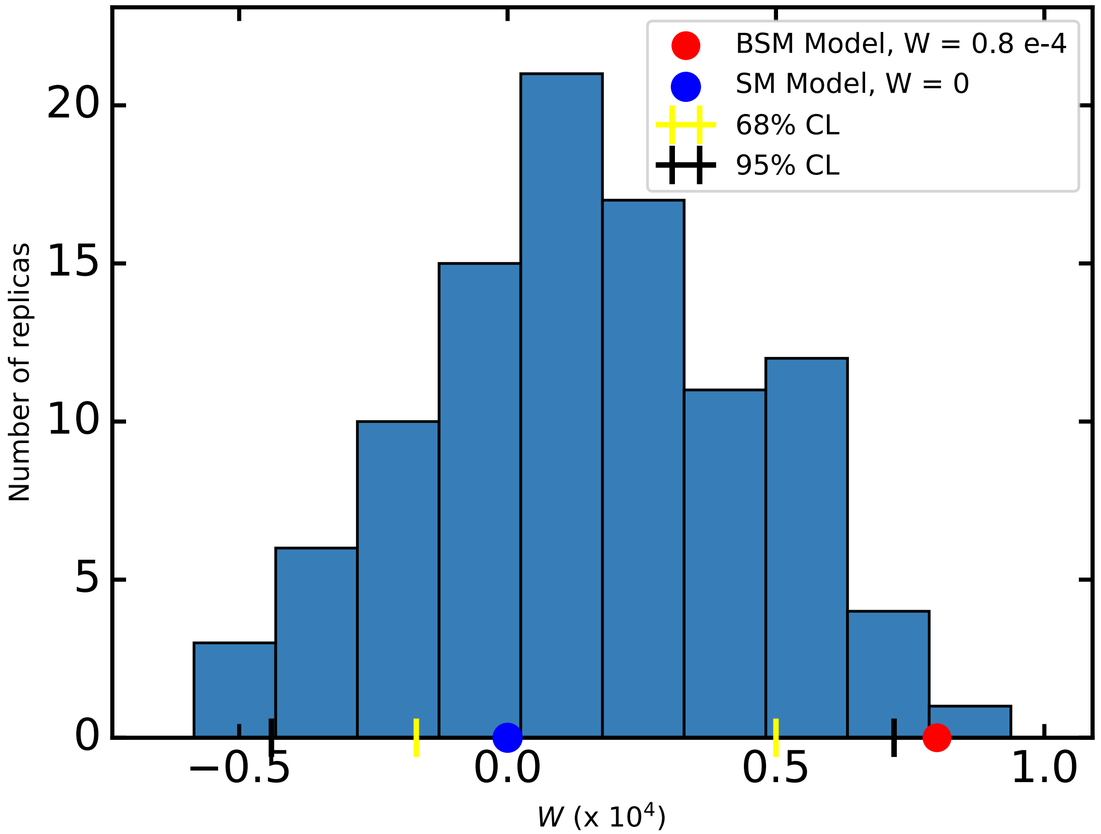
$pp \to W^+W^-$ (SM)



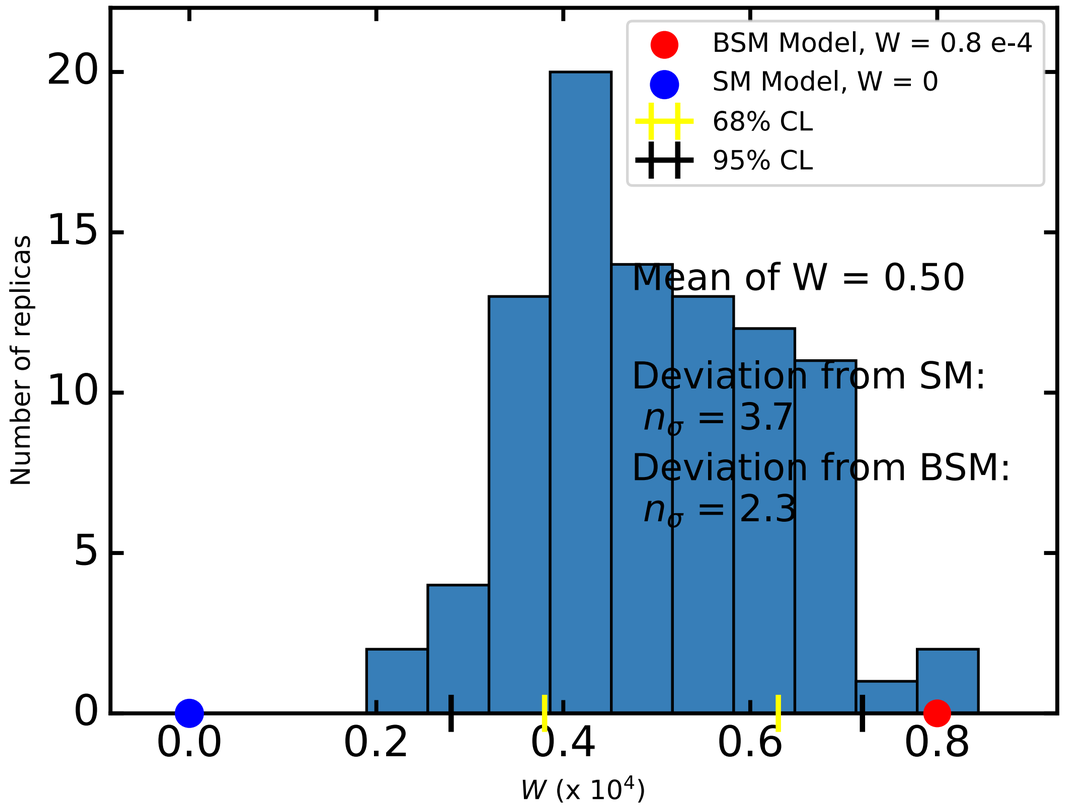
Missing new physics

Impact of the NP absorption in PDFs on SMEFT fits

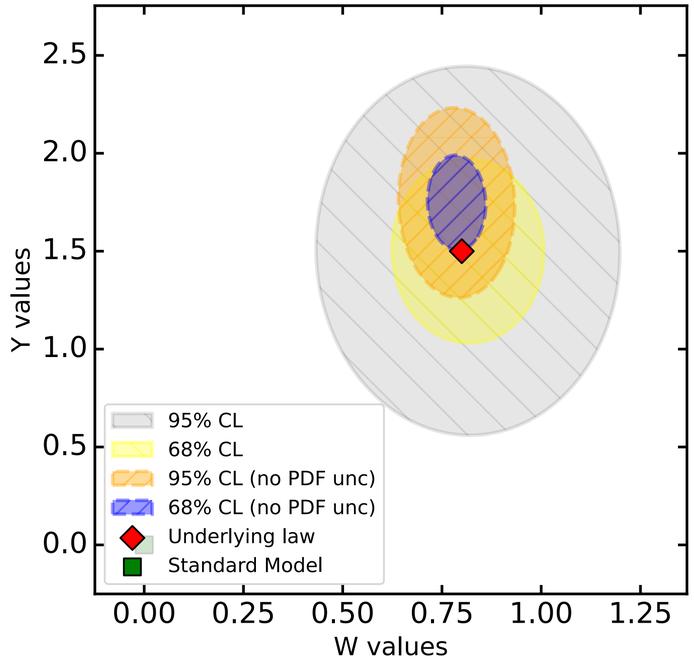
CC in PDF / NC in SMEFT fits



NC in PDF / CC in SMEFT fits



PDFs for new physics searches

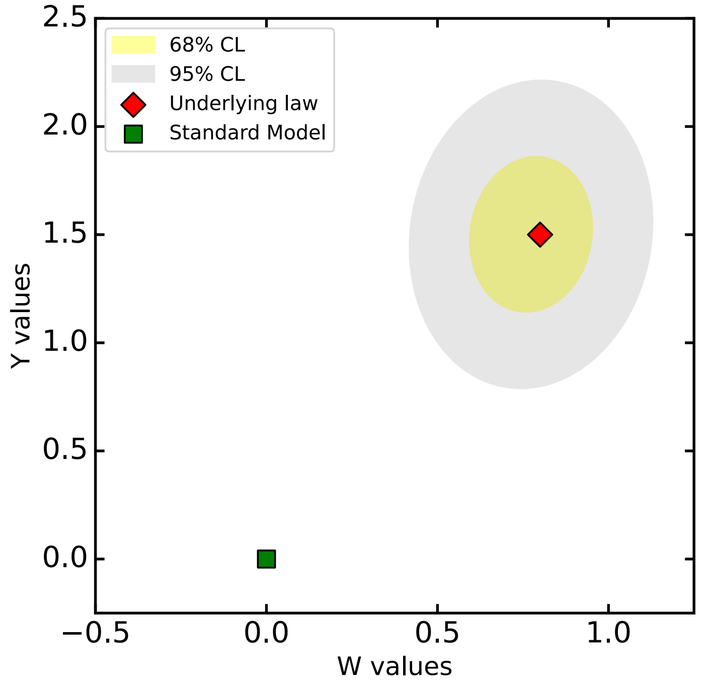


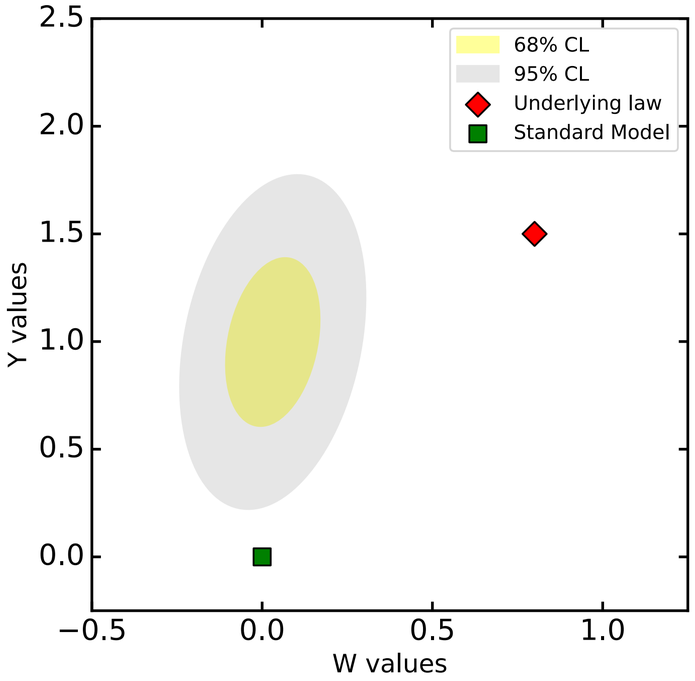
SMEFT only fit
(True PDF)

✗ Impossible

Simultaneous fit
(no PDF assumption)

✓ Doable



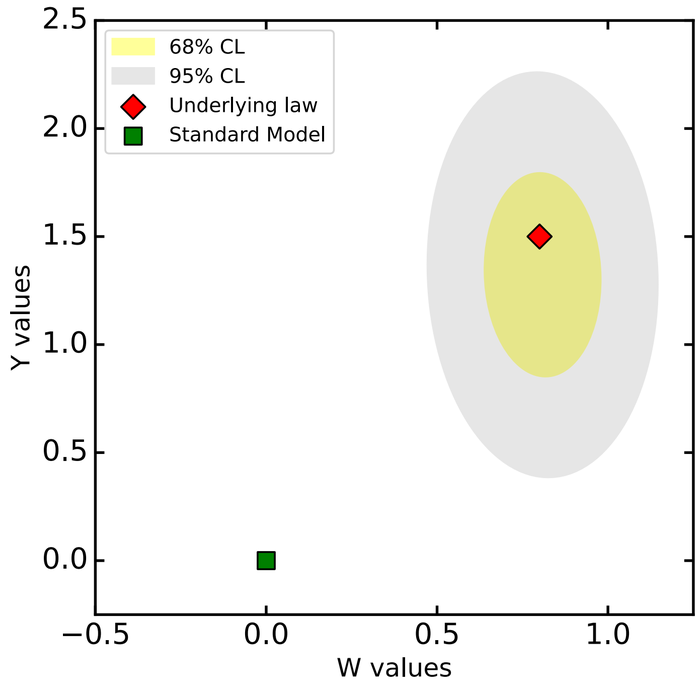


SMEFT only fit
(BSM-biased PDF)

✗ Wrong

SMEFT only fit
(Conservative PDF)

✓ Doable



Comparing SMEFT bounds for different PDFs

Injection of W' and Z' in Drell-Yan

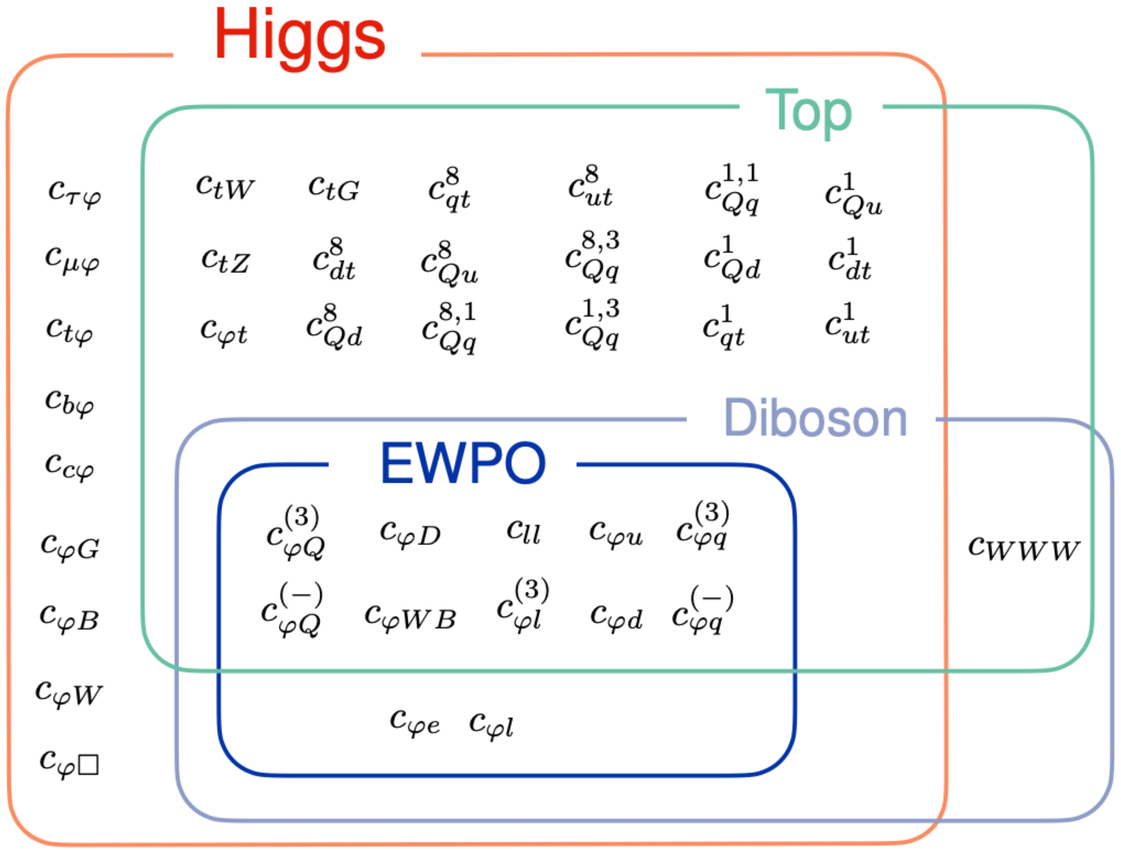
Fit	$\hat{W} \times 10^5$ (UL = 0.8)		$\hat{Y} \times 10^5$ (UL = 1.5)	
	95% CI	n_σ **(SM/UL)**	**95% CI**	n_σ **(SM/UL)**
True PDF (no unc)	(0.62, 0.92)	10.79 / 0.18	(1.3, 2.25)	7.24 / 1.03
True PDF (w unc)	(0.45, 1.12)	4.28 / 0.08	(0.65, 2.33)	3.20 / 0.00
BSM-biased PDF	(-0.22, 0.3)	0.23 / 5.74	(0.22, 1.77)	2.60 / 1.31
Cons. PDF	(0.45, 1.1)	4.86 / 0.05	(0.51, 2.35)	2.85 / 0.38
Simu. fit	(0.4, 1.08)	4.42 / 0.15	(0.77, 2.22)	4.27 / 0.00

Simultaneous fit of PDF and new physics

SMEFT operators implemented

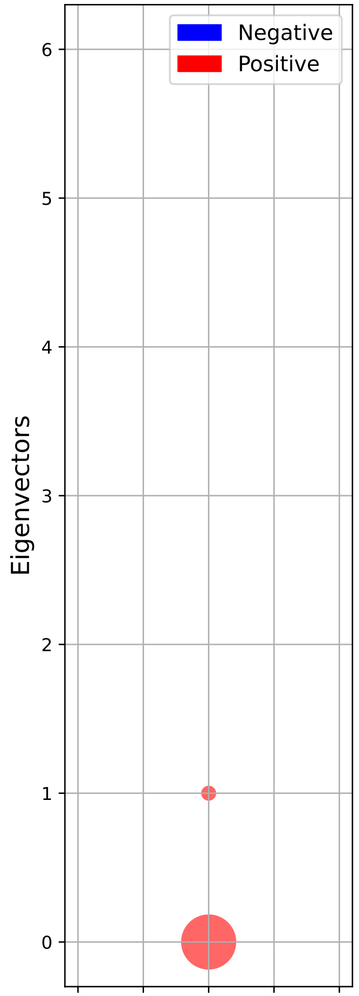
- 40 operators implemented

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

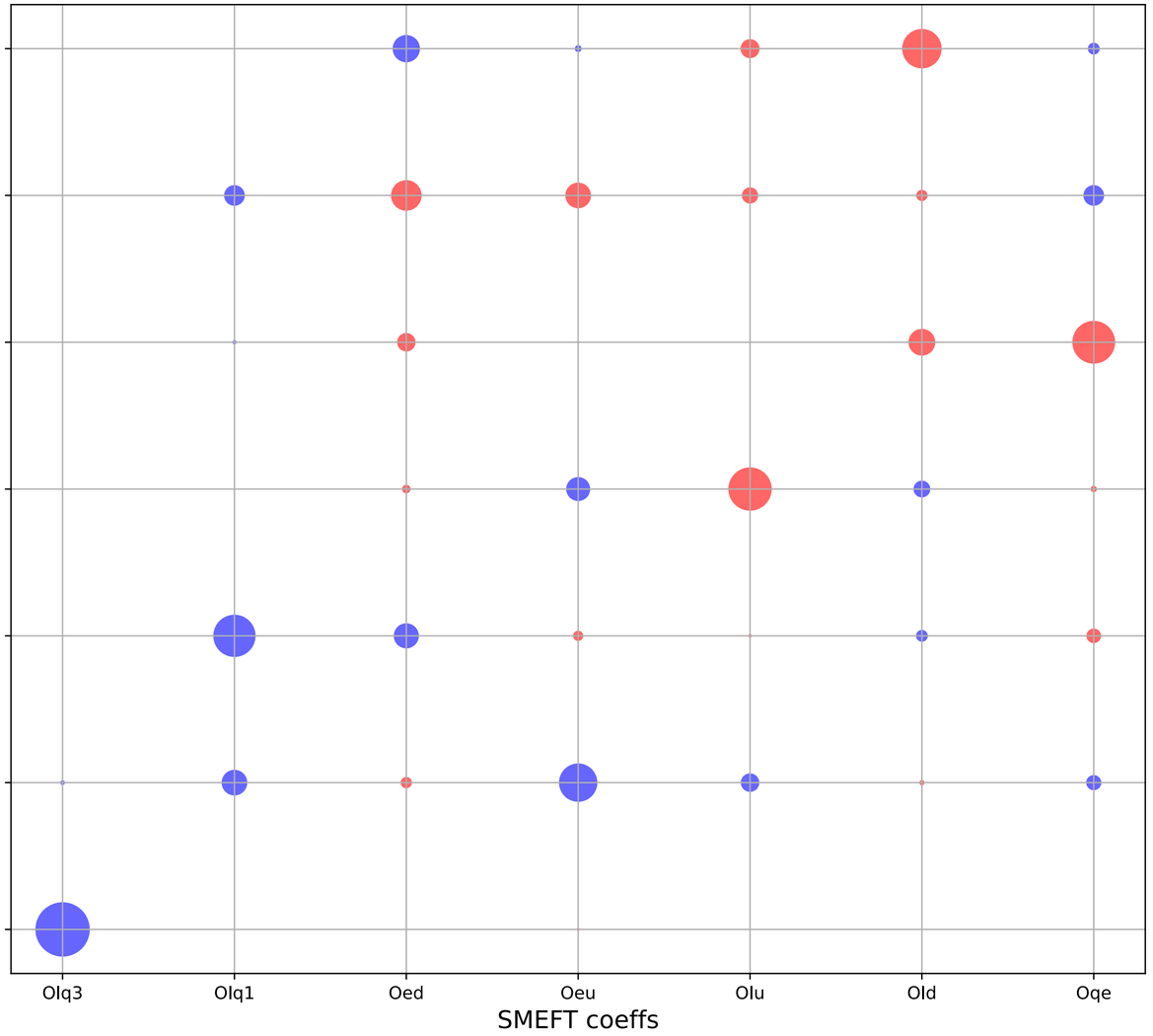


SMEFT PCA results

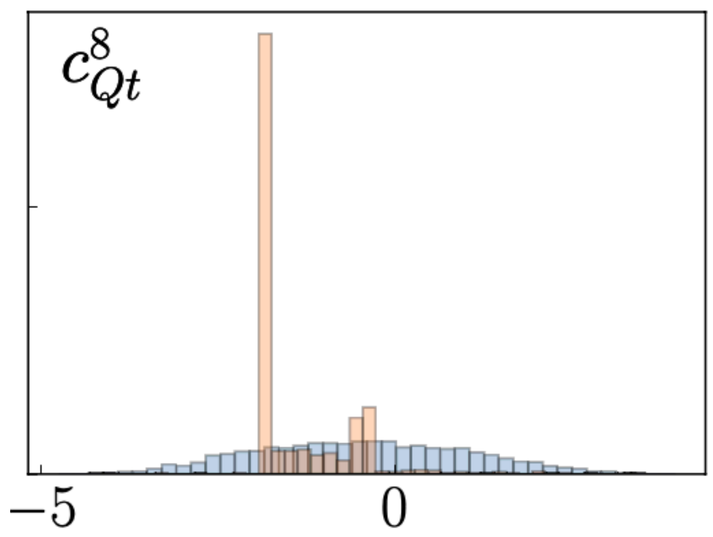
FIM Eigenvalues

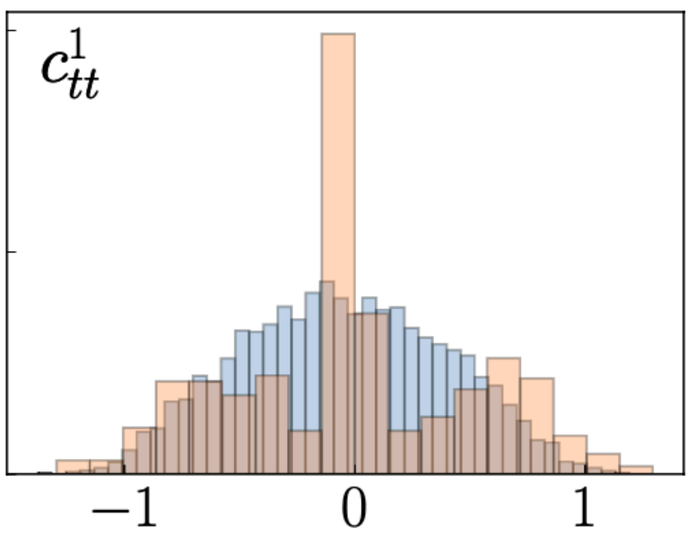


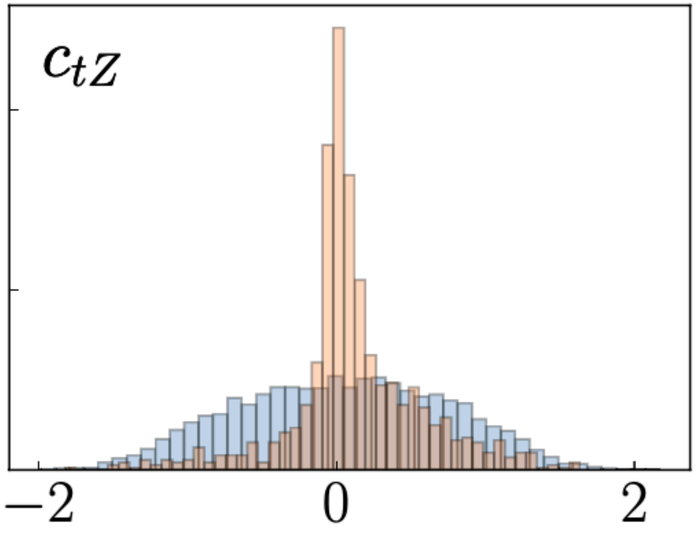
FIM eigenvectors and SMEFT operators



BEYOND MONTE CARLO AND NEURAL NETWORKS







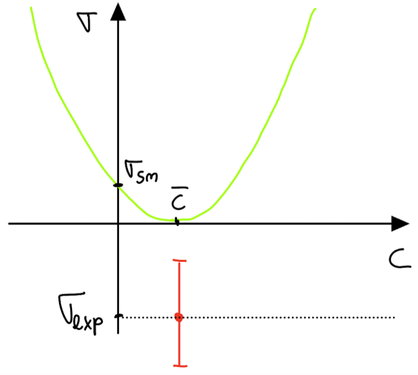
- In the quadratic SMEFT fit observed disagreement between MC method and Bayesian method. Very different posterior (hence different CLs)

- 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]

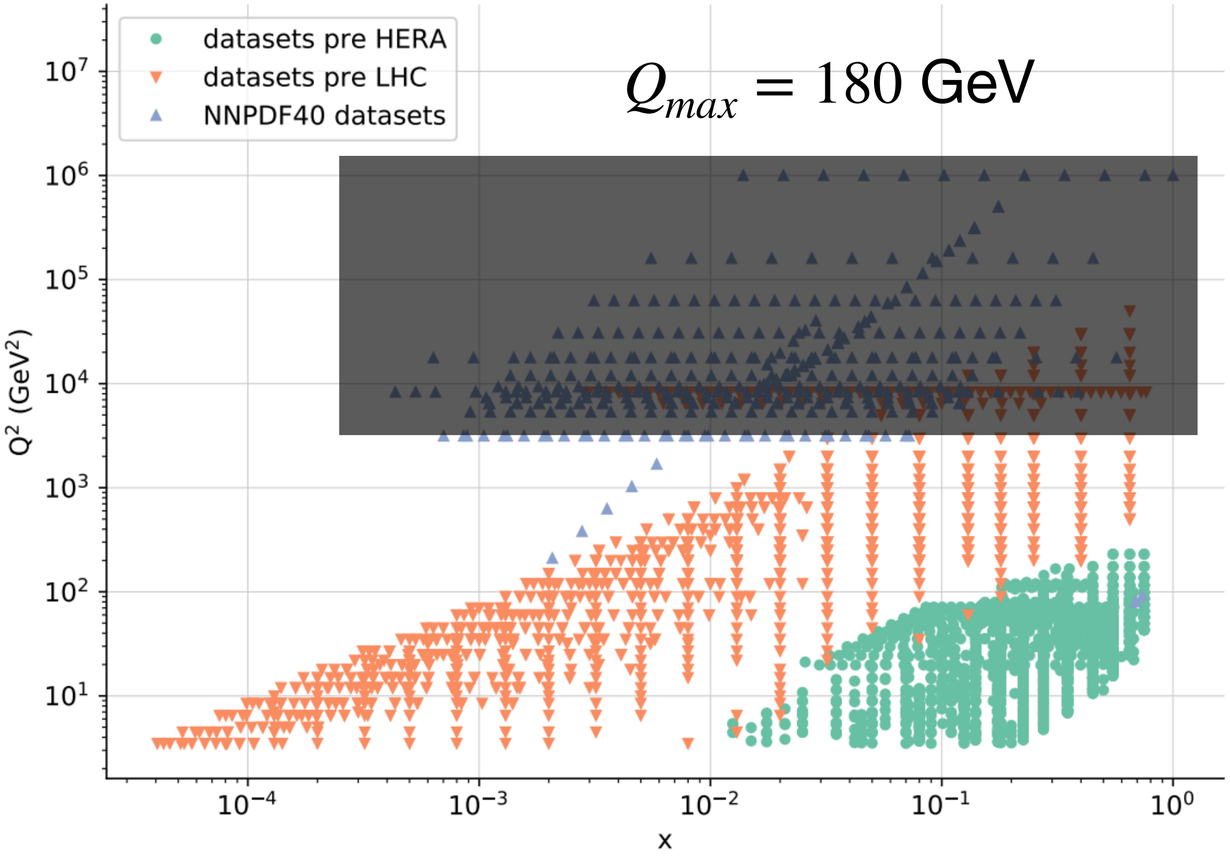
Let's consider a simple scenario: 1 operator, 1 datapoint

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



Discrepancy between low and high-energy data fits

Kinematic coverage



Impact of energy cut