The Kavli Institute for Theoretical Physics
University of California, Santa Barbara

European Research Council
Established by the European Commission


MARIA UBIALI
UNIVERSITY OF CAMBRIDGE


# DISCUSSION – PART III
# PROGRESS AND ISSUES IN THE PRECISION PDF DETERMINATION

# OUTLINE

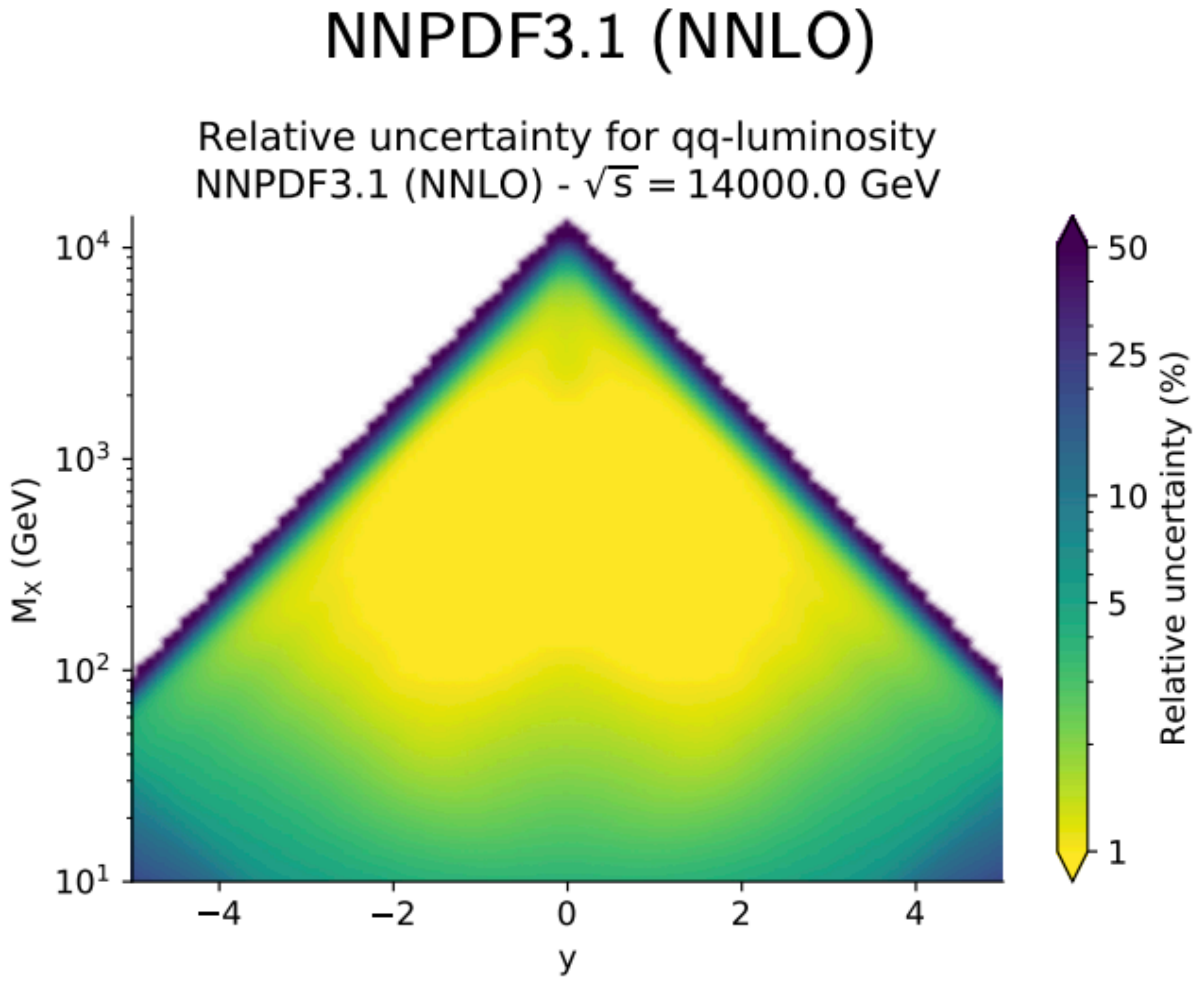
- Towards 1% accurate sets?
  - ➡ Experimental data
  - ➡ Methodological advances
- The way ahead
  - ➡ Theory: missing higher order uncertainties & electroweak corrections
  - ➡ Data/Interpretation: Interplay between PDFs and high mass tails
- Points for discussion

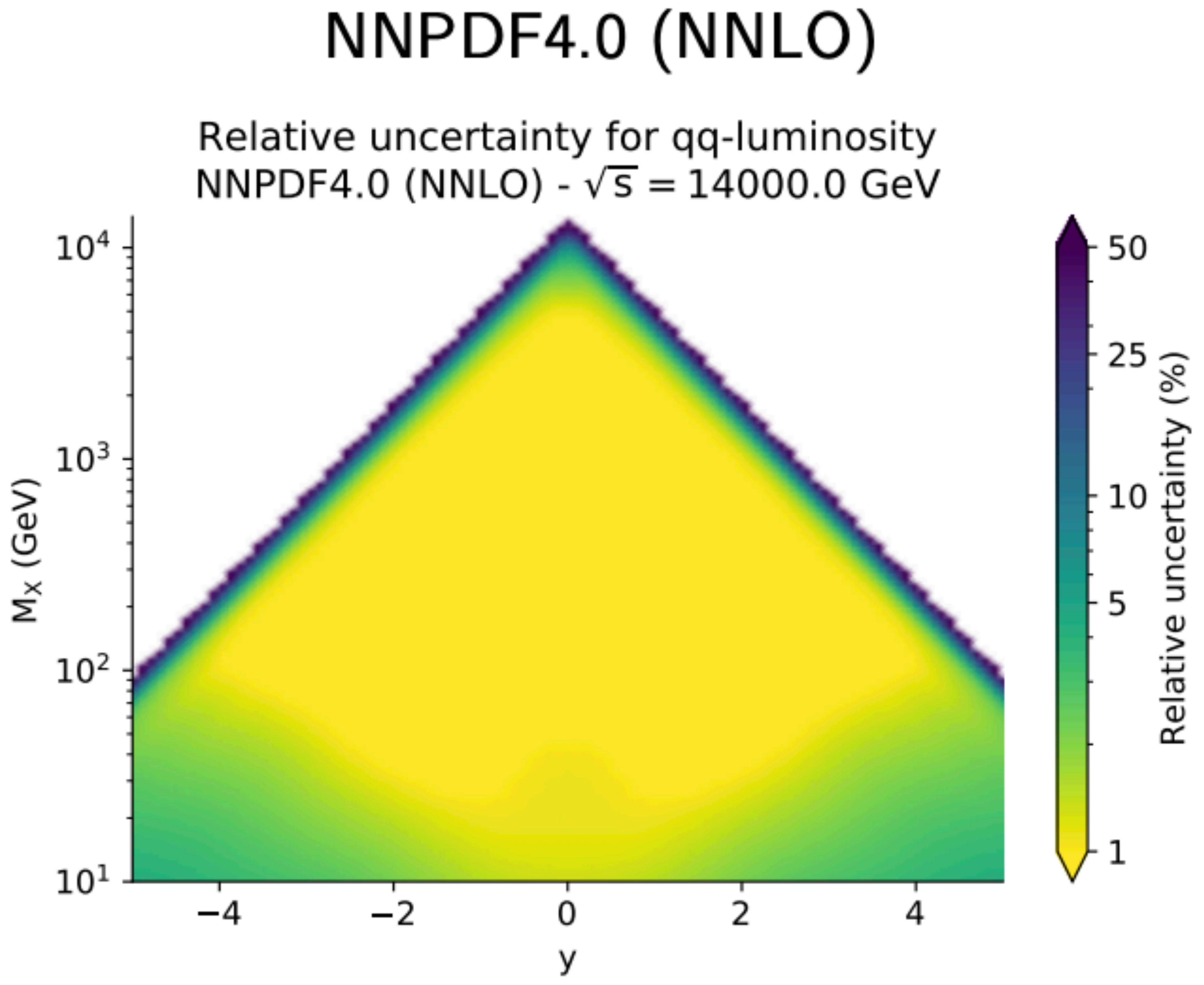
# 1% ACCURATE SETS?

# TOWARDS 1% ACCURATE PDFS

$$\mathcal{L}_{ij}(M_X, y, \sqrt{s}) = \frac{1}{s} \sum_{i,j} f_i\left(\frac{M_X e^y}{\sqrt{s}}, M_X\right) f_j\left(\frac{M_X e^{-y}}{\sqrt{s}}, M_X\right)$$

SINGLET



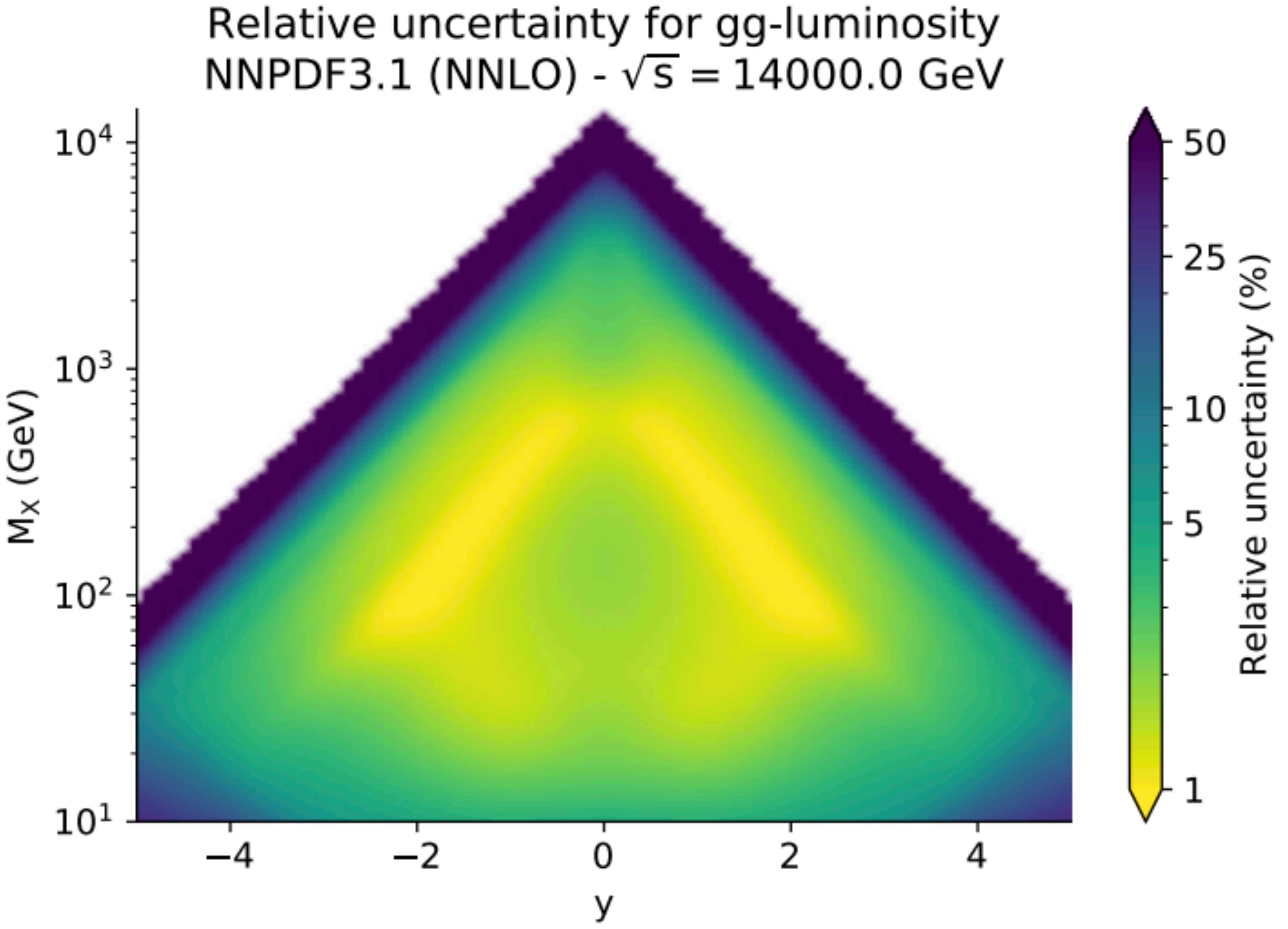



# TOWARDS 1% ACCURATE PDFS

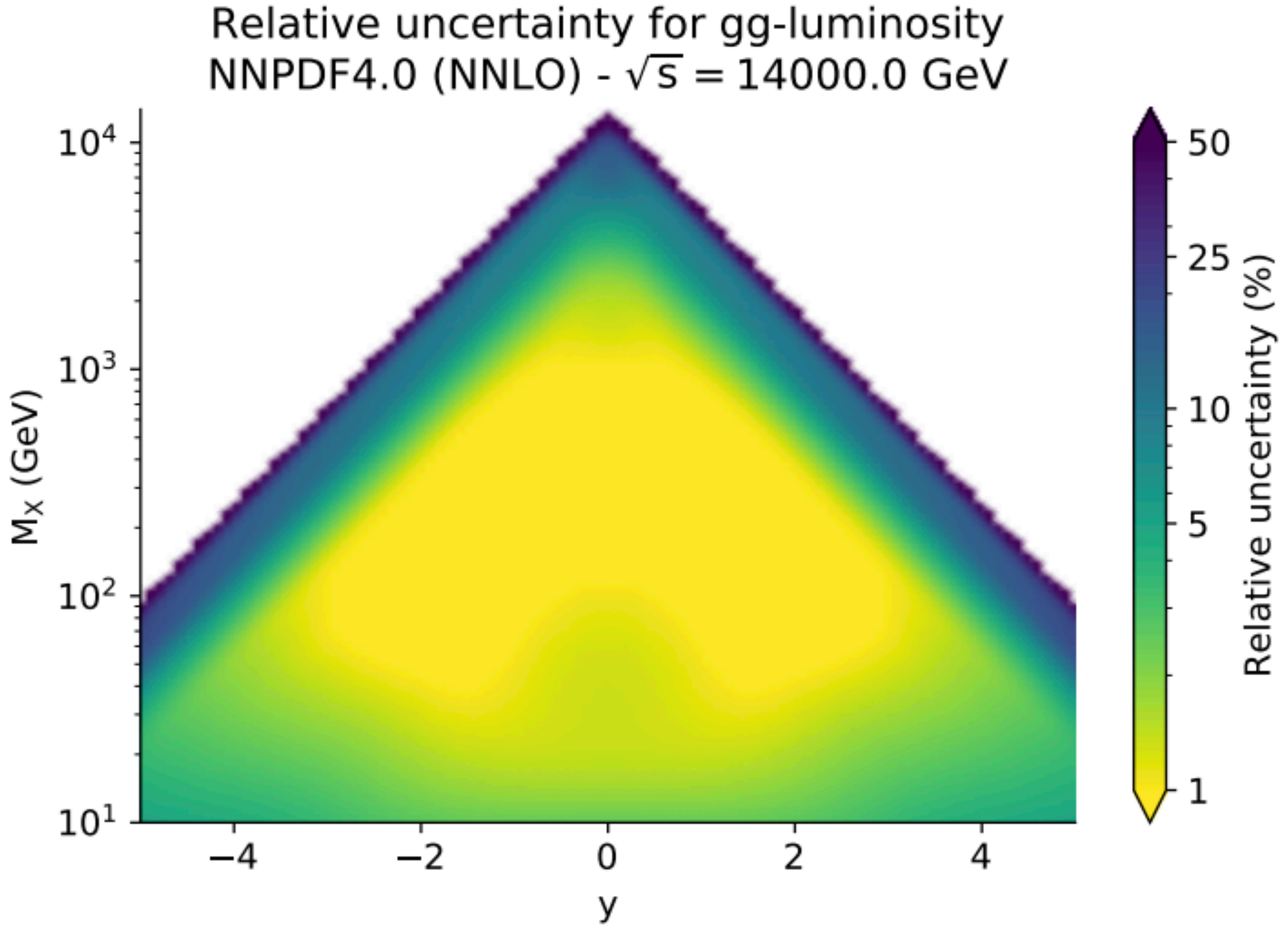
$$\mathcal{L}_{ij}(M_X, y, \sqrt{s}) = \frac{1}{s} \sum_{i,j} f_i \left( \frac{M_X e^y}{\sqrt{s}}, M_X \right) f_j \left( \frac{M_X e^{-y}}{\sqrt{s}}, M_X \right)$$

GLUON

NNPDF3.1 (NNLO)

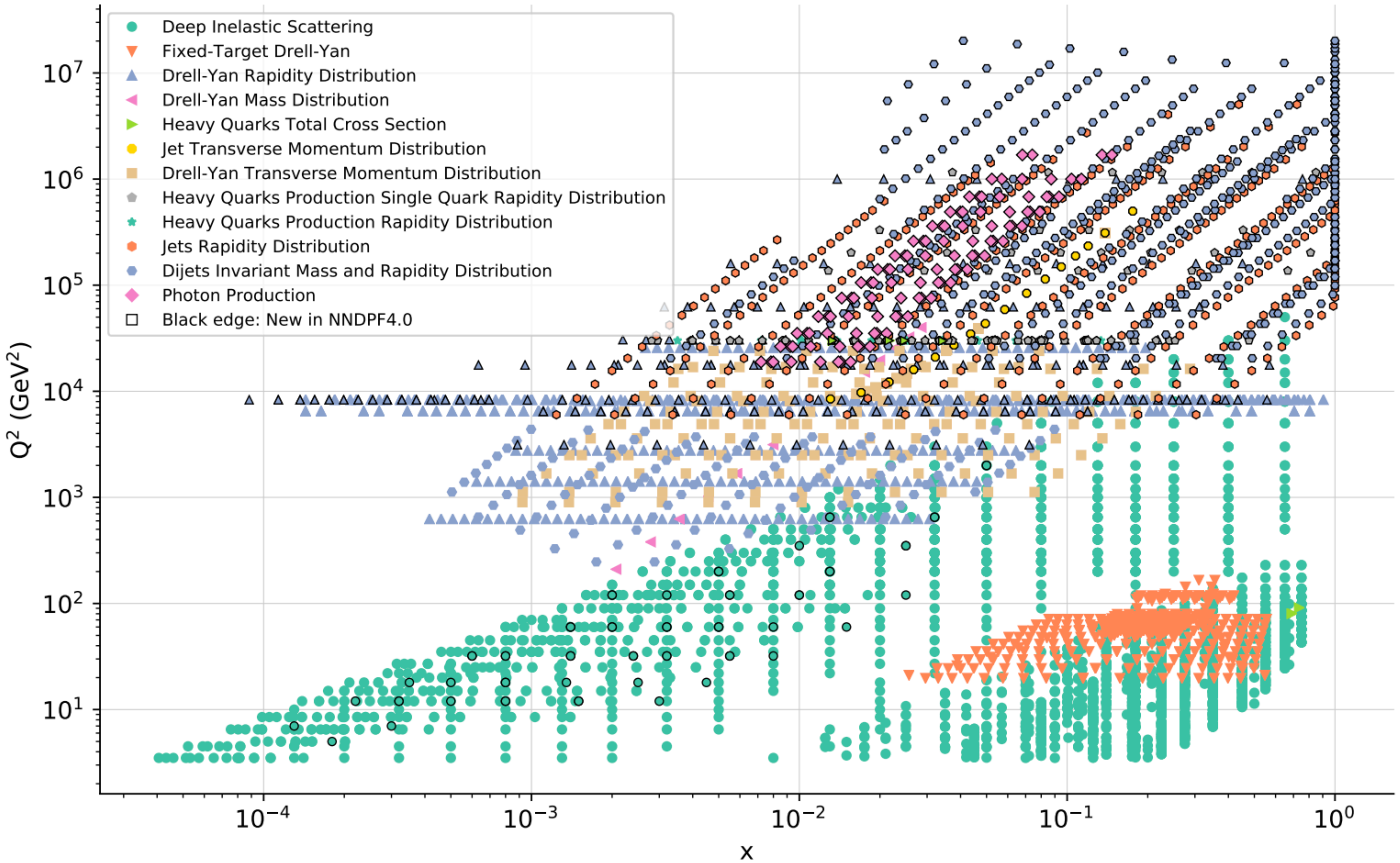



NNPDF4.0 (NNLO)




Steady progress towards 1% relative uncertainty on $L_{ij}$ on a broad kinematic range

# EXPERIMENTAL DATA




**NNPDF4.0**

- O(50) data sets investigated
- O(400) data points more than in NNPDF3.1

# EXPERIMENTAL DATA

| Data set | NNPDF4.0 | NNPDF3.1 | ABMP16 | CT18 | MSHT20 |
|---|---|---|---|---|---|
| ATLAS $W, Z$ 7 TeV (2010) | ✓ | ✓ | ✓ | ✓ | ✓ |
| ATLAS $W, Z$ 7 TeV (2011) | ✓ | ✓ | ✗ | ✓ | ✓ |
| ATLAS low-mass DY 7 TeV | ✓ | ✓ | ✗ | ✗ | ✗ |
| ATLAS high-mass DY 7 TeV | ✓ | ✓ | ✗ | ✗ | ✓ |
| ATLAS $W$ 8 TeV | ✓ | ✗ | ✗ | ✗ | ✓ |
| ATLAS DY 2D 8 TeV | ✓ | ✗ | ✗ | ✗ | ✓ |
| ATLAS high-mass DY 2D 8 TeV | ✓ | ✗ | ✗ | ✗ | ✓ |
| ATLAS $\sigma_{W,Z}$ 13 TeV | ✓ | ✗ | ✓ | ✗ | ✗ |
| ATLAS $W^+$+jet 8 TeV | ✓ | ✗ | ✗ | ✗ | ✓ |
| ATLAS $Z$ $p_T$ 8 TeV | ✓ | ✓ | ✗ | ✓ | ✓ |
| ATLAS $\sigma_{tt}^{tot}$ 7, 8 TeV | ✓ | ✓ | ✓ | ✗ | ✗ |
| ATLAS $\sigma_{tt}^{tot}$ 13 TeV | ✓ | ✓ | ✓ | ✗ | ✗ |
| ATLAS $t\bar{t}$ lepton+jets 8 TeV | ✓ | ✓ | ✗ | ✓ | ✓ |
| ATLAS $t\bar{t}$ dilepton 8 TeV | ✓ | ✗ | ✗ | ✗ | ✓ |
| ATLAS single-inclusive jets 7 TeV, R=0.6 | ✗ | ✓ | ✗ | ✓ | ✓ |
| ATLAS single-inclusive jets 8 TeV, R=0.6 | ✓ | ✗ | ✗ | ✗ | ✗ |
| ATLAS dijets 7 TeV, R=0.6 | ✓ | ✗ | ✗ | ✗ | ✗ |
| ATLAS direct photon production 13 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |
| ATLAS single top $R_t$ 7, 8, 13 TeV | ✓ | ✗ | ✓ | ✗ | ✗ |
| ATLAS single top diff. 7, 8 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |
| ATLAS single top diff. 8 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |

| Data set | NNPDF4.0 | NNPDF3.1 | ABMP16 | CT18 | MSHT20 |
|---|---|---|---|---|---|
| LHCb $Z$ 940 pb | ✓ | ✓ | ✗ | ✗ | ✓ |
| LHCb $Z \to ee$ 2 fb | ✓ | ✓ | ✓ | ✓ | ✓ |
| LHCb $W, Z \to \mu$ 7 TeV | ✓ | ✓ | ✓ | ✓ | ✓ |
| LHCb $W, Z \to \mu$ 8 TeV | ✓ | ✓ | ✓ | ✓ | ✓ |
| LHCb $Z \to \mu\mu, ee$ 13 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |

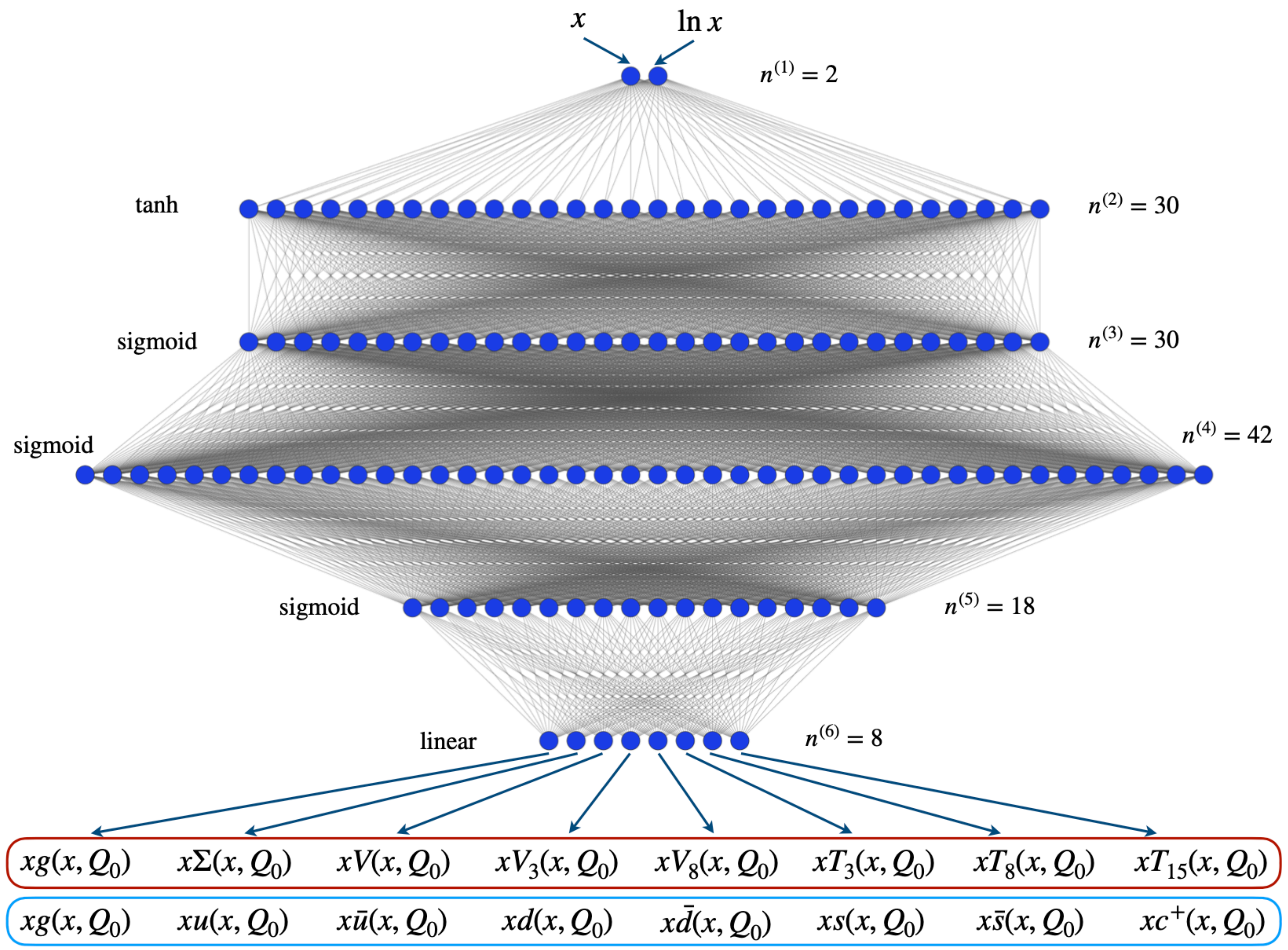
| Data set | NNPDF4.0 | NNPDF3.1 | ABMP16 | CT18 | MSHT20 |
|---|---|---|---|---|---|
| CMS $W$ electron asymmetry 7 TeV | ✓ | ✓ | ✗ | ✓ | ✓ |
| CMS $W$ muon asymmetry 7 TeV | ✓ | ✓ | ✓ | ✓ | ✗ |
| CMS Drell-Yan 2D 7 TeV | ✓ | ✓ | ✗ | ✗ | ✓ |
| CMS $W$ rapidity 8 TeV | ✓ | ✓ | ✓ | ✓ | ✓ |
| CMS $Z$ $p_T$ 8 TeV | ✓ | ✓ | ✗ | ✓ | ✗ |
| CMS $W + c$ 7 TeV | ✓ | ✓ | ✗ | ✗ | ✓ |
| CMS $W + c$ 13 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |
| CMS single-inclusive jets 2.76 TeV | ✗ | ✓ | ✗ | ✗ | ✓ |
| CMS single-inclusive jets 7 TeV | ✗ | ✓ | ✗ | ✓ | ✓ |
| CMS dijets 7 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |
| CMS single-inclusive jets 8 TeV | ✗ | ✗ | ✗ | ✓ | ✓ |
| CMS 3D dijets 8 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |
| CMS $\sigma_{tt}^{tot}$ 5 TeV | ✓ | ✗ | ✓ | ✗ | ✗ |
| CMS $\sigma_{tt}^{tot}$ 7, 8 TeV | ✓ | ✓ | ✓ | ✗ | ✓ |
| CMS $\sigma_{tt}^{tot}$ 13 TeV | ✓ | ✓ | ✓ | ✗ | ✗ |
| CMS $t\bar{t}$ lepton+jets 8 TeV | ✓ | ✓ | ✗ | ✗ | ✓ |
| CMS $t\bar{t}$ 2D dilepton 8 TeV | ✓ | ✗ | ✗ | ✓ | ✓ |
| CMS $t\bar{t}$ lepton+jet 13 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |
| CMS $t\bar{t}$ dilepton 13 TeV | ✓ | ✗ | ✗ | ✗ | ✗ |
| CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV | ✓ | ✗ | ✓ | ✗ | ✗ |
| CMS single top $R_t$ 8, 13 TeV | ✓ | ✗ | ✓ | ✗ | ✗ |

- Inclusion of di-jets preferred over jets based on results on [Khalek et al, Eur.Phys.J.C 80 (2020) 8, 797]
- + HERA combined reduced c and b cross sections
- + DIS jets (HERA) included via reweighting
- + NOMAD neutrino data included via reweighting

# AN HYPER-OPTIMISED METHODOLOGY

- New methodology based on modern machine-learning techniques [EPJC79 (2019) 676]

    ➡ Single neural network to parametrise 8 independent PDF combinations (g, u, d, s, u~, d~, s~, c=c~)
    ➡ Check independence of results on the chosen parametrisation basis
    ➡ New optimisation strategy based on gradient descent rather than genetic algorithm
    ➡ Hyper-optimised methodology: scan of the hyper parameter space to find optimal minimisation settings (optimiser, initialiser, stopping patience, number of layers, learning rate, epochs, activation function) [Carrazza et al, Eur.Phys.J.C 79 (2019) 8, 676]
    ➡ Complete statistica validation of PDF uncertainties via closure tests (data region) and future test (extrapolation region) [J. Cruz-Martinez et al, Acta Phys.Polon.B 52 (2021) 243]

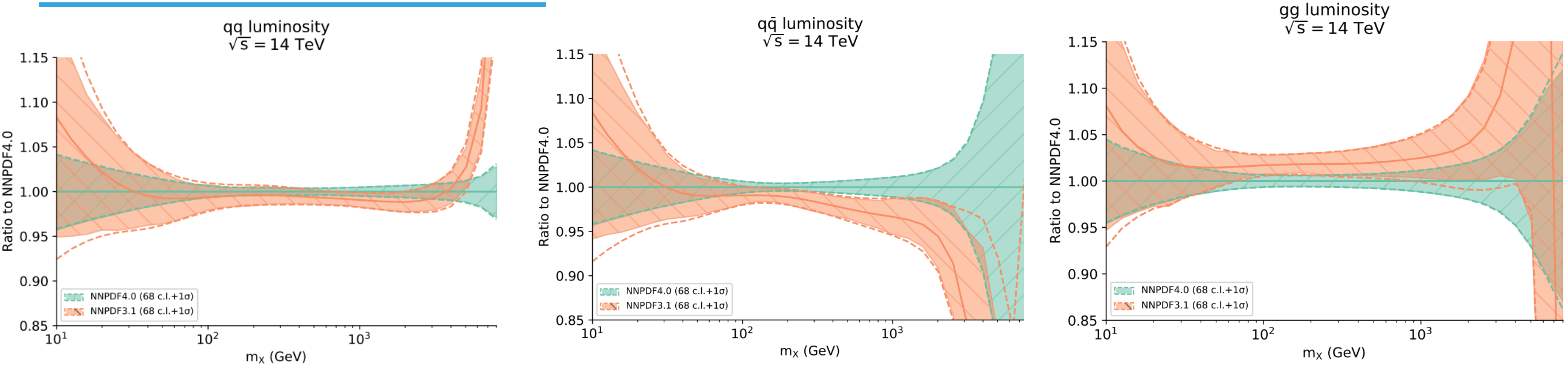



- Refined theoretical framework

    ➡ Nuclear uncertainties fit deuteron and heavy ions included as theory uncertainty [Ball et al, Eur.Phys.J.C 81 (2021) 1, 37]
    ➡ Extended positivity constraints at the PDF level for light quarks/anti-quarks and gluons [Candido et al, JHEP 11 (2020) 129]
    ➡ Integrability constraints
    ➡ EW corrections carefully assessed and data cuts consistently applied [Carrazza et al, JHEP 12 (2020) 1]

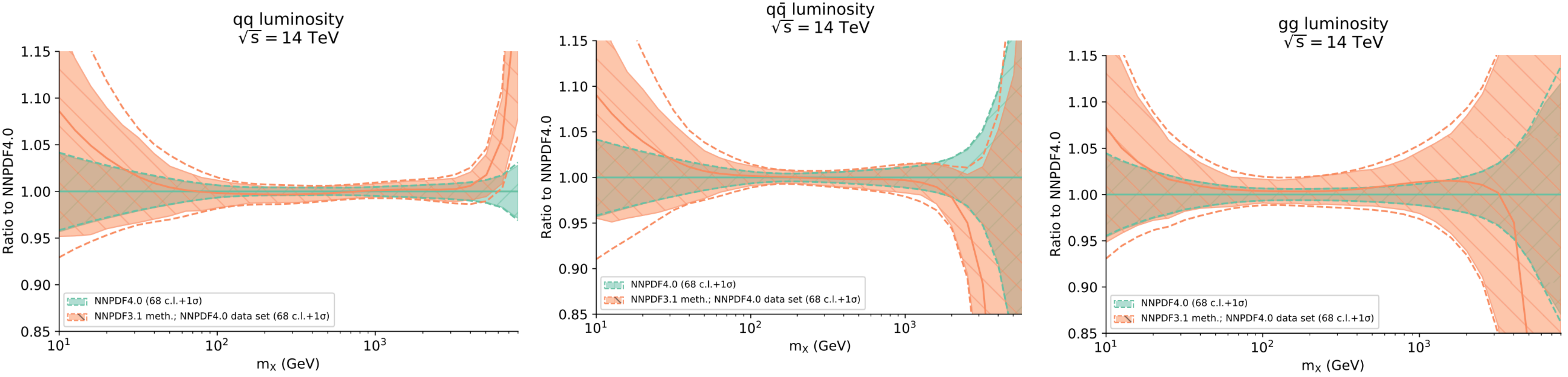
# NNPDF4.0 RESULTS

## NNPDF3.1 vs NNPDF4.0




| DATA \ METHODOLOGY | NNPDF3.1 | NNPDF4.0 |
|---|---|---|
| NNPDF3.1 (4093) | 1.19 | 1.12 |
| NNPDF4.0 (4491) | 1.25 | 1.17 |

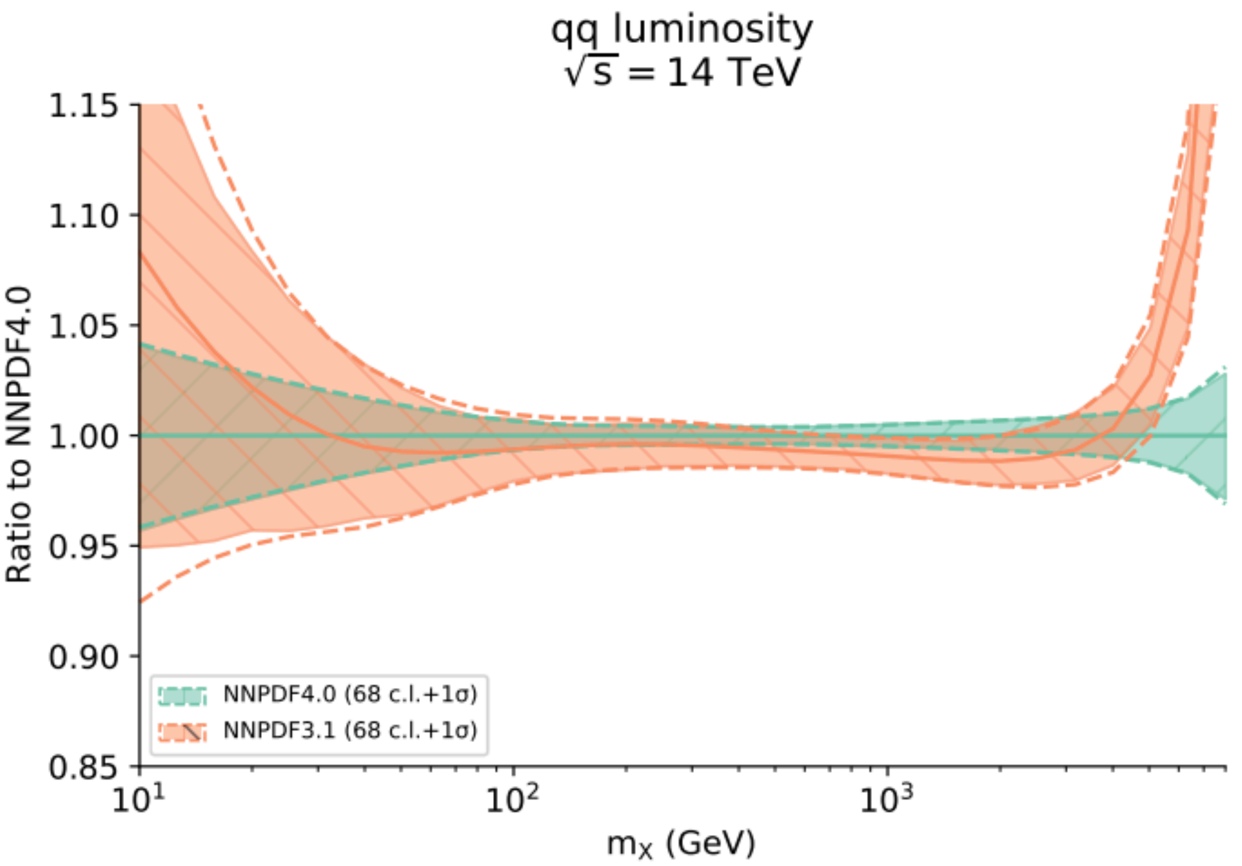
## NNPDF4.0 data - NNPDF3.1 vs NNPDF4.0 methodology

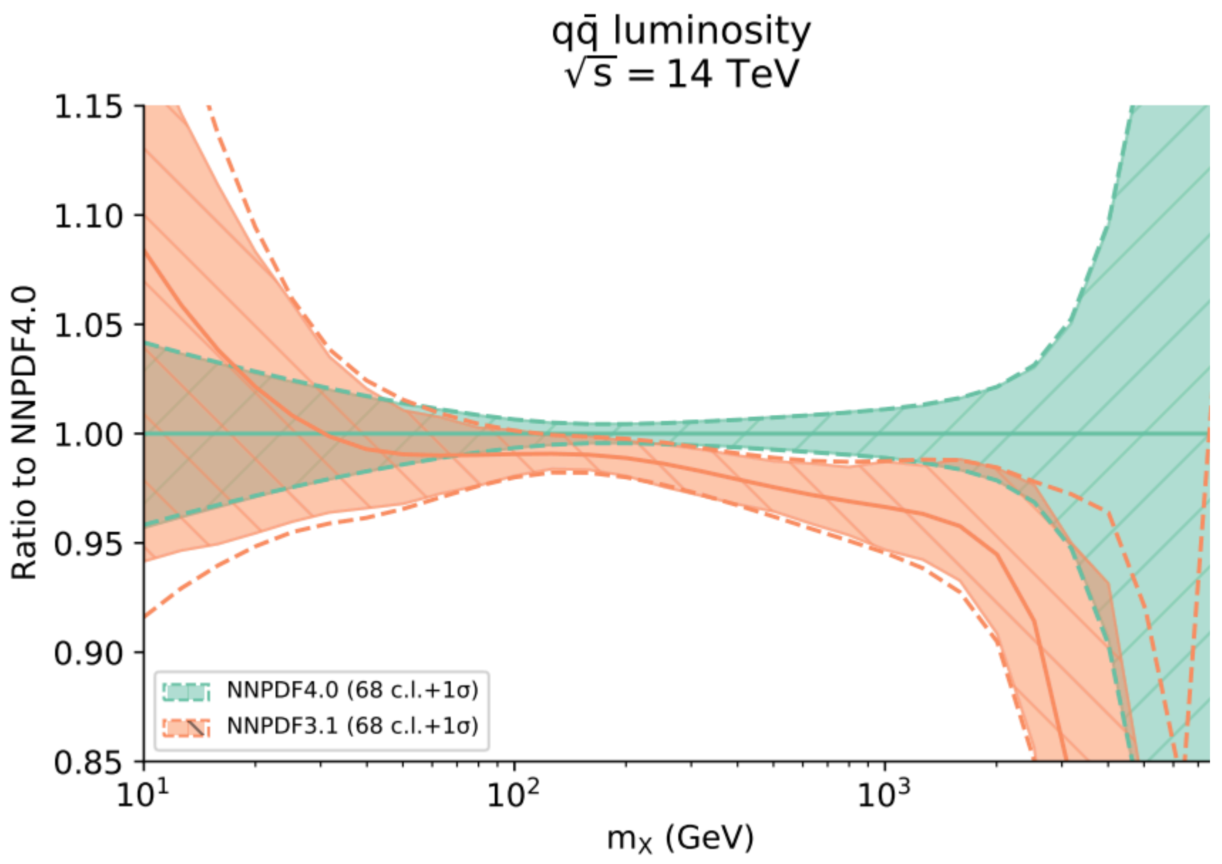


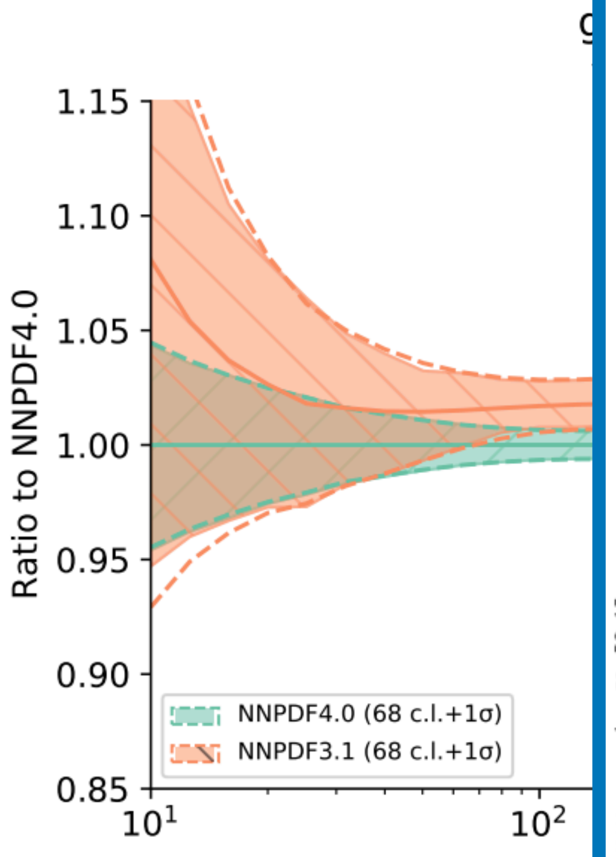

| DATA \ METHODOLOGY | NNPDF3.1 | NNPDF4.0 |
|---|---|---|
| NNPDF3.1 (4093) | 1.19 | 1.12 |
| NNPDF4.0 (4491) | 1.25 | 1.17 |

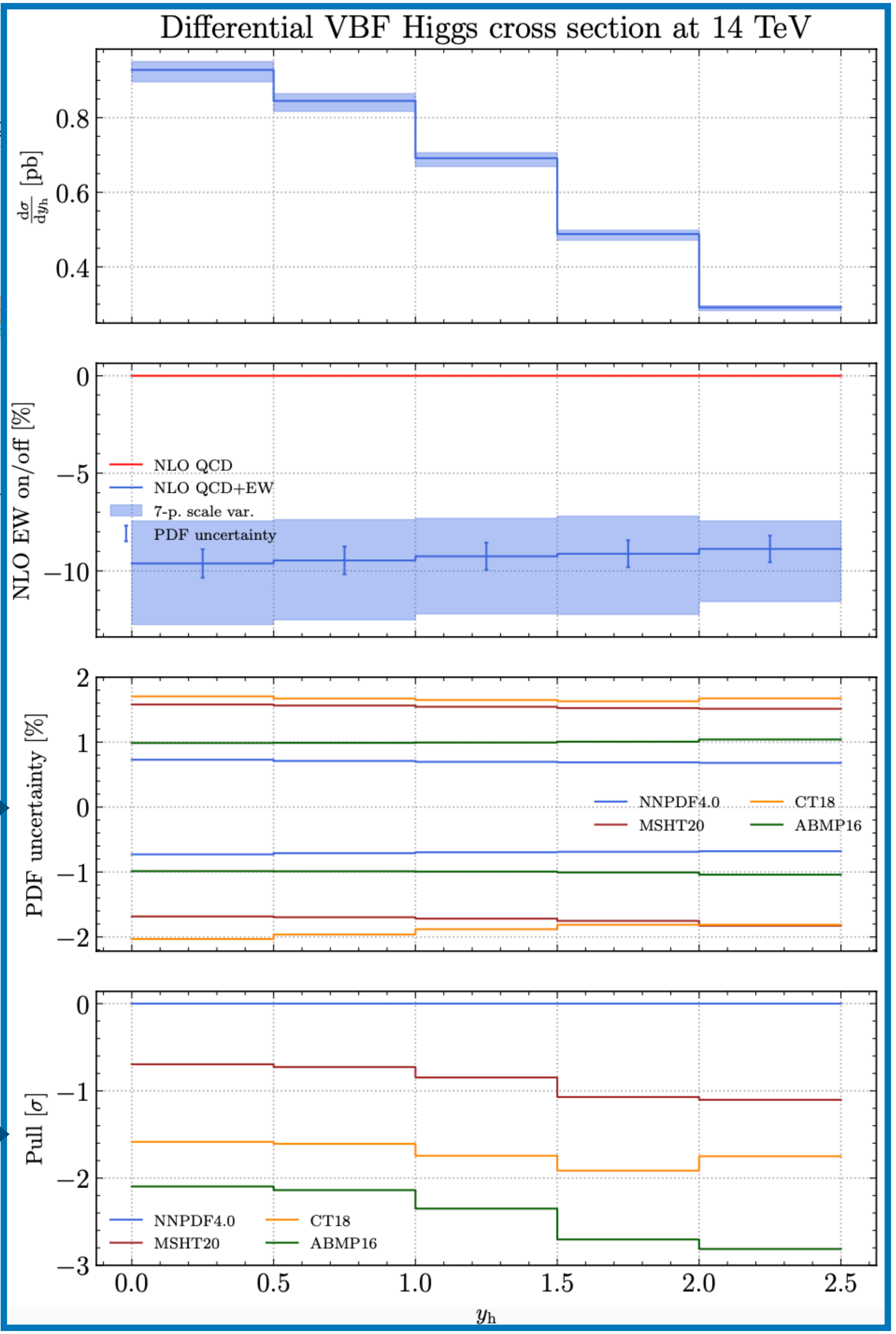
- Shift in parton luminosities mostly due to inclusion of O(500) more data points
- Parton luminosities based on same dataset are consistent with each other but 4.0 methodology displays smaller uncertainty than 3.1 methodology: NNPDF4.0 more accurate and superior to previous methodology

# NNPDF4.0 RESULTS




At the level of observables the NNPDF4.0 results yield < 1% PDF uncertainties in a number of observables

Pull between NNPDF4.0 and other global determinations enhanced by smaller uncertainties

# THEORY FRONTIERS

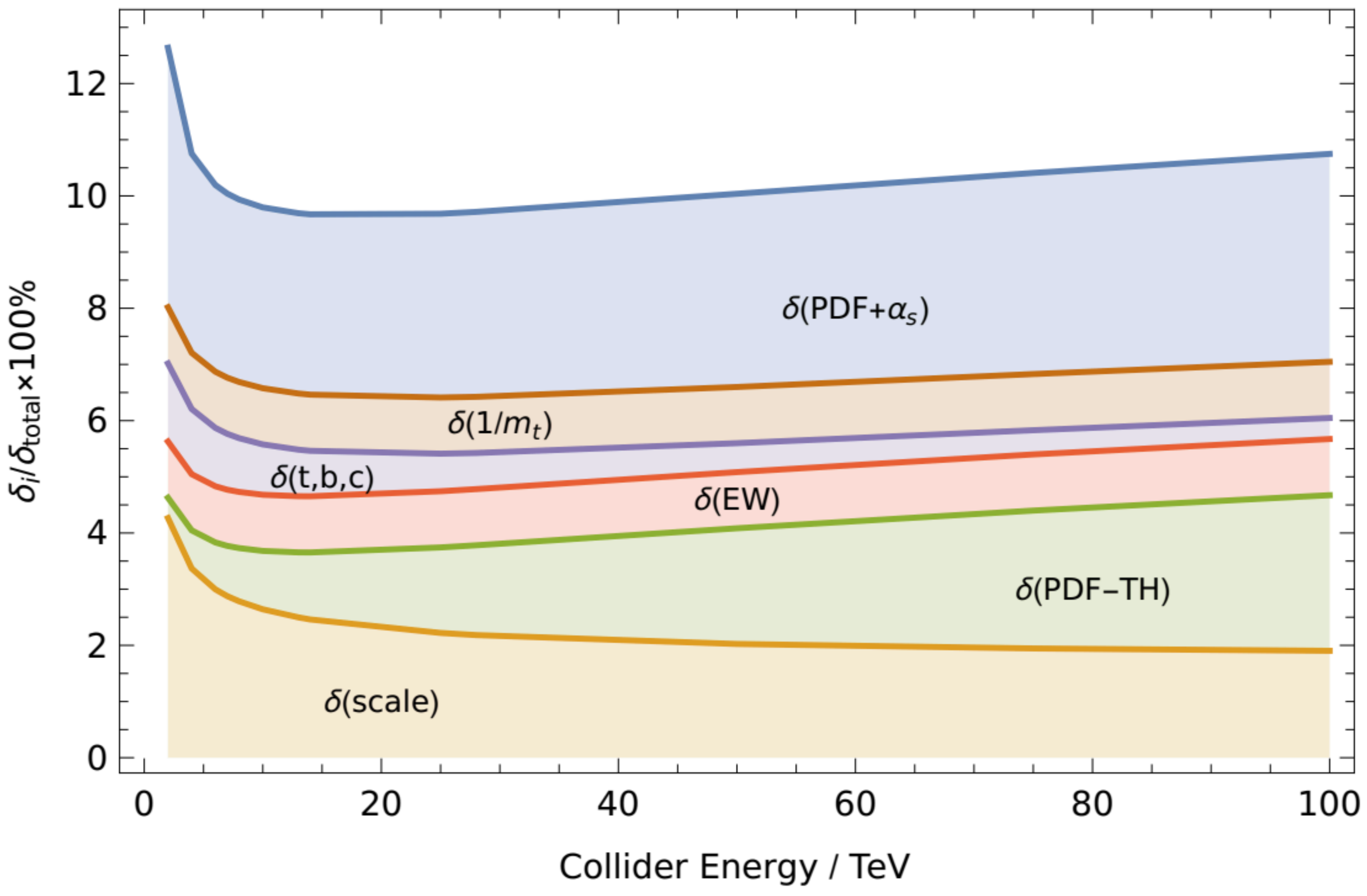
# THEORY UNCERTAINTIES IN PDF FITS

$$\sigma = \alpha_s^p \sigma_0 + \alpha_s^{p+1}\sigma_1 + \alpha_s^{p+2}\sigma_2 + \mathcal{O}(\alpha_s^{p+3})$$

- Standard global PDF fits based on fixed-order NNLO QCD calculations (using fast interpolation grid for NLO predictions accompanied by local K-factors for NNLO)
- N3LO is now the precision frontier for partonic cross sections (N3LO splitting functions partially known in Nc→∞ limit)
- Mismatch between perturbative order of partonic cross section and PDFs becoming significant source of uncertainty

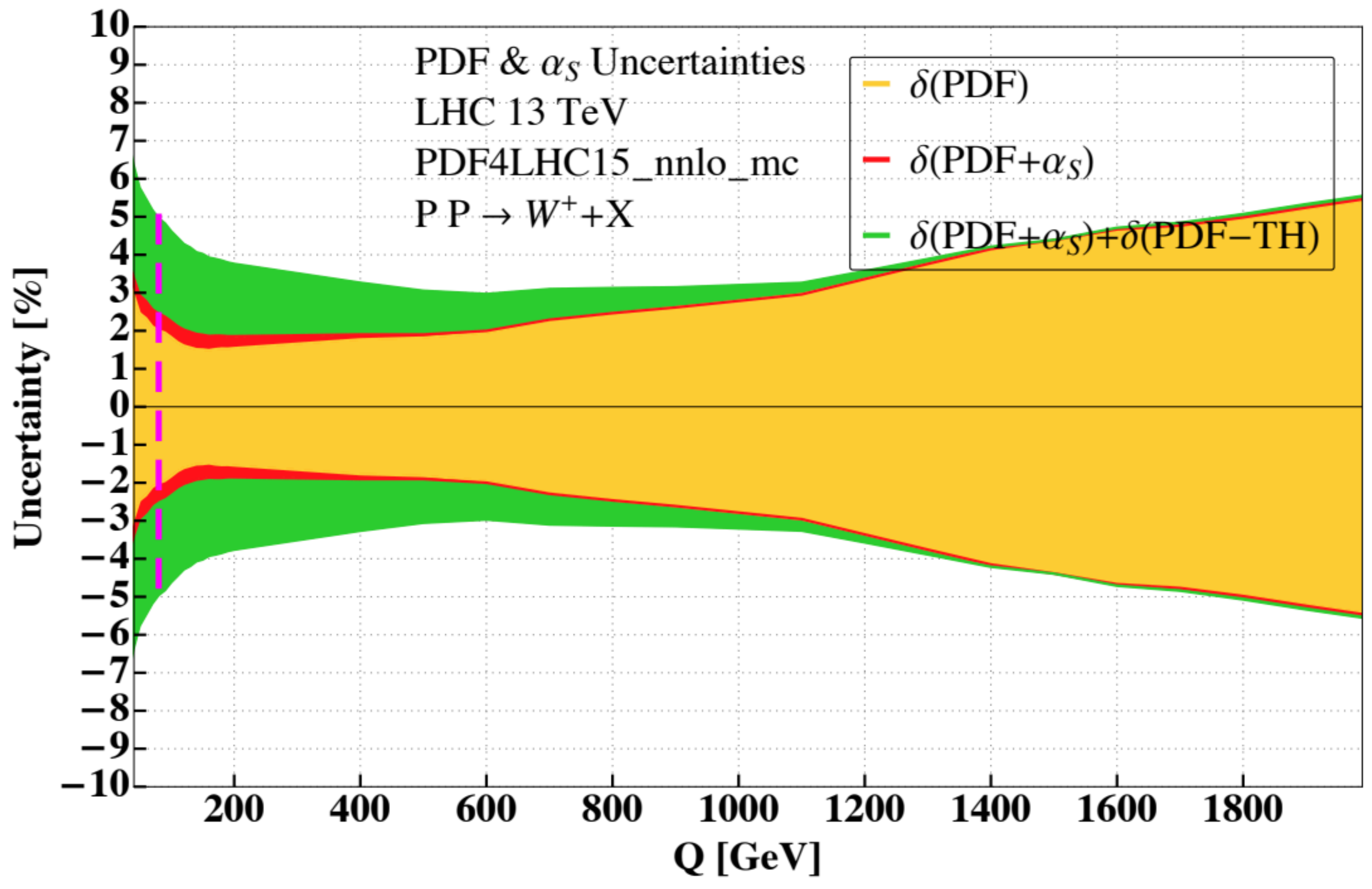
$$\delta(PDF-TH) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\mathrm{NNLO-PDFs}} - \sigma^{(2)}_{\mathrm{NLO-PDFs}}}{\sigma^{(2)}_{\mathrm{NNLO-PDFs}}}\right|$$

**Gluon-gluon fusion into Higgs**




M. Cepeda et al. [HL/HE WG2 group], arXiv:1902.00134

**Drell-Yan**



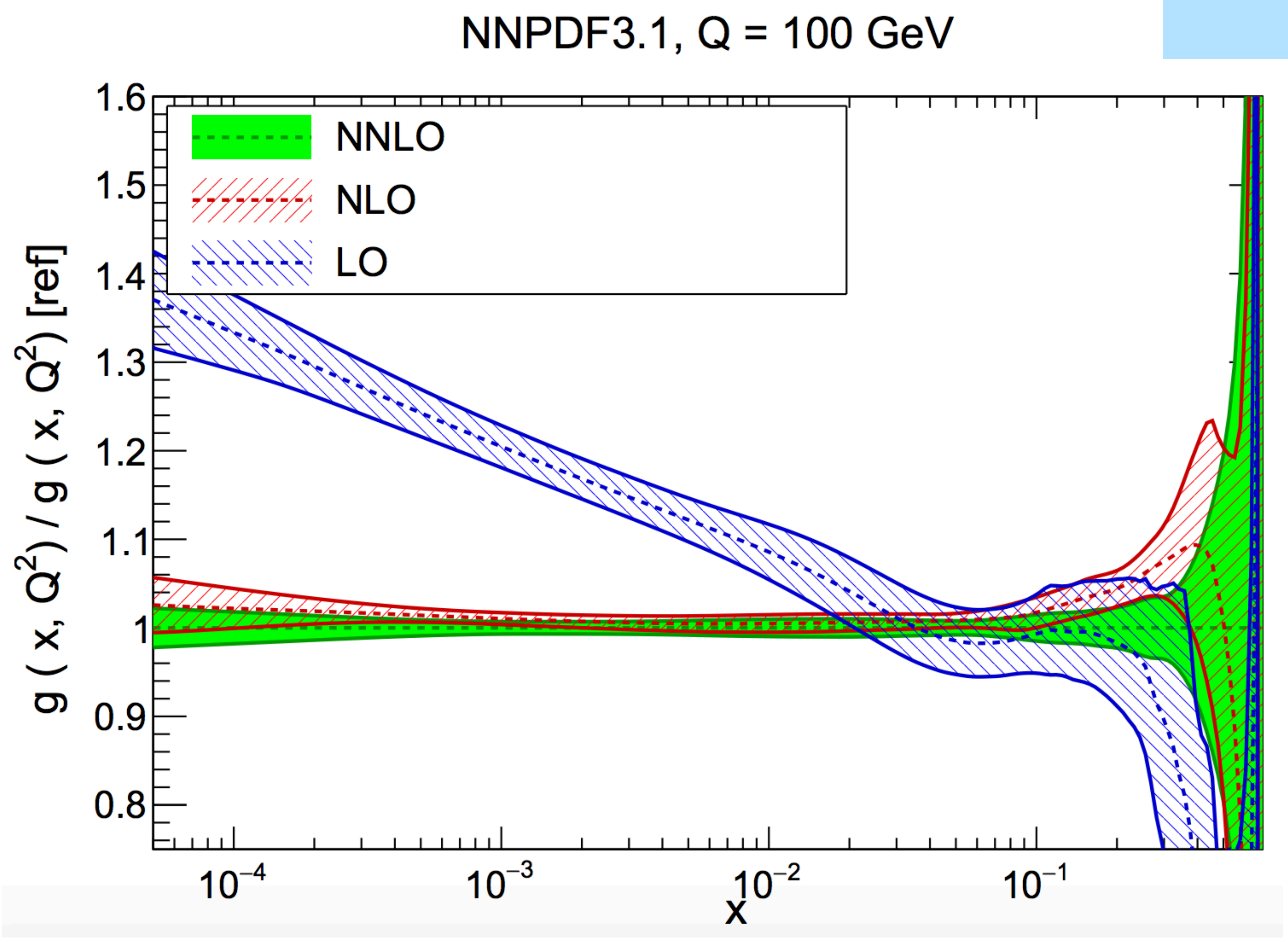

C. Duhr et al, *JHEP* 11 (2020) 143

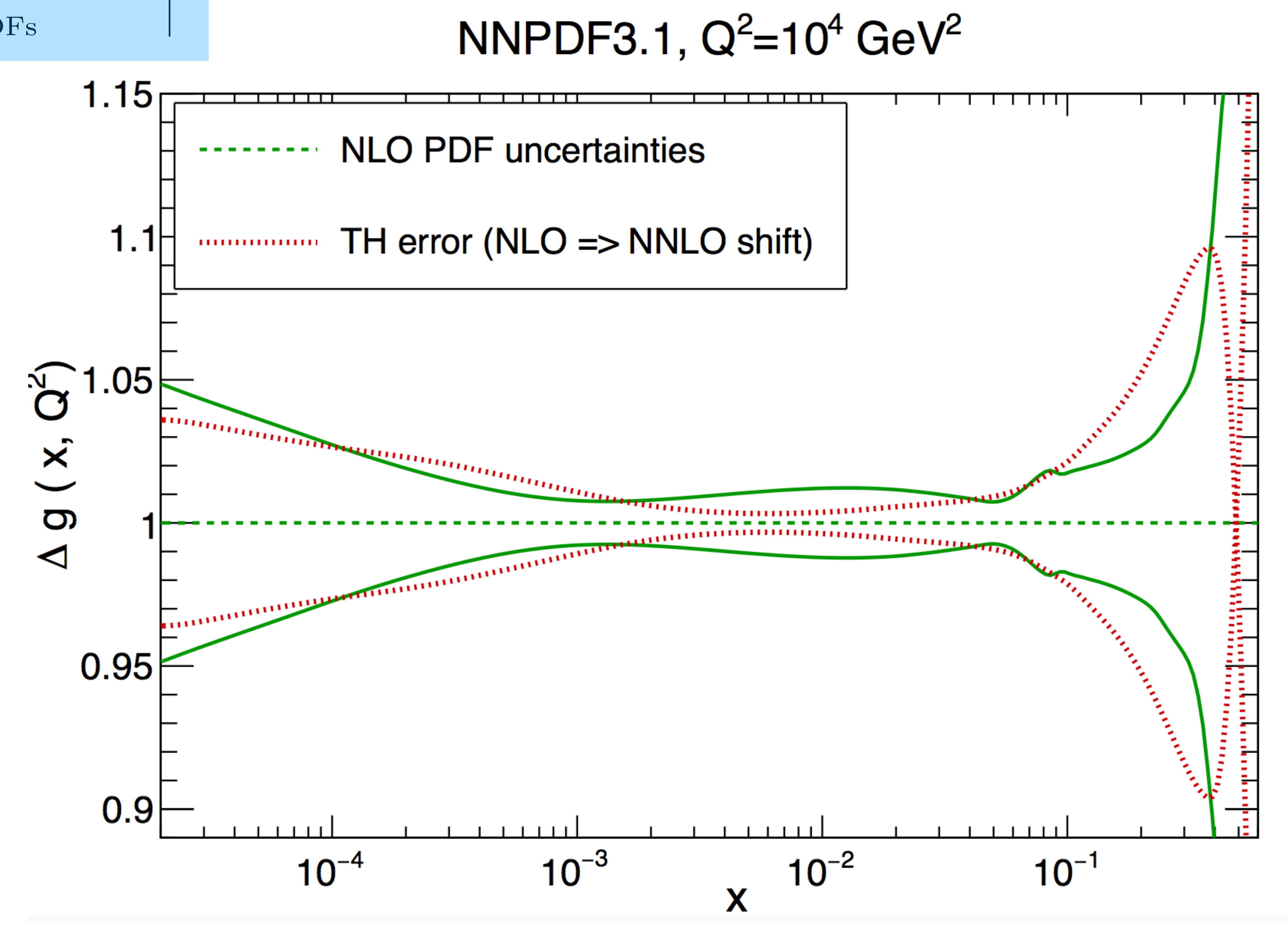
# THEORY UNCERTAINTIES IN PDF FITS

$$\sigma = \alpha_s^p \sigma_0 + \alpha_s^{p+1}\sigma_1 + \alpha_s^{p+2}\sigma_2 + \mathcal{O}(\alpha_s^{p+3})$$

- Standard global PDF fits based on fixed-order NNLO QCD calculations (using fast interpolation grid for NLO predictions accompanied by local K-factors for NNLO)
- N3LO is now the precision frontier for partonic cross sections
- Mismatch between perturbative order of partonic cross section and PDFs becoming significant source of uncertainty

$$\delta(PDF - TH) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\mathrm{NNLO-PDFs}} - \sigma^{(2)}_{\mathrm{NLO-PDFs}}}{\sigma^{(2)}_{\mathrm{NNLO-PDFs}}}\right|$$



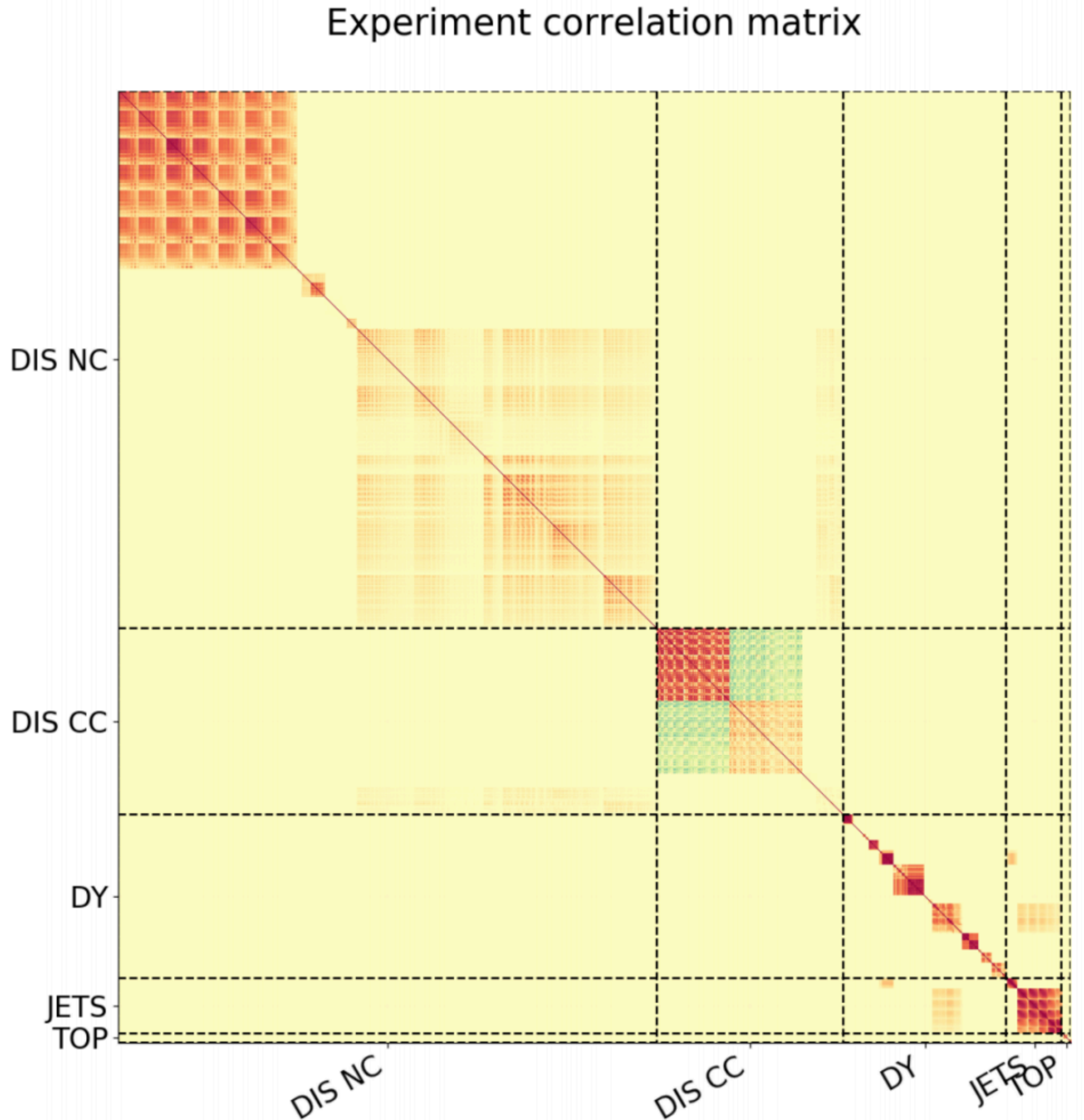

R.D. Ball et al, EPJ C77 (2017) 663

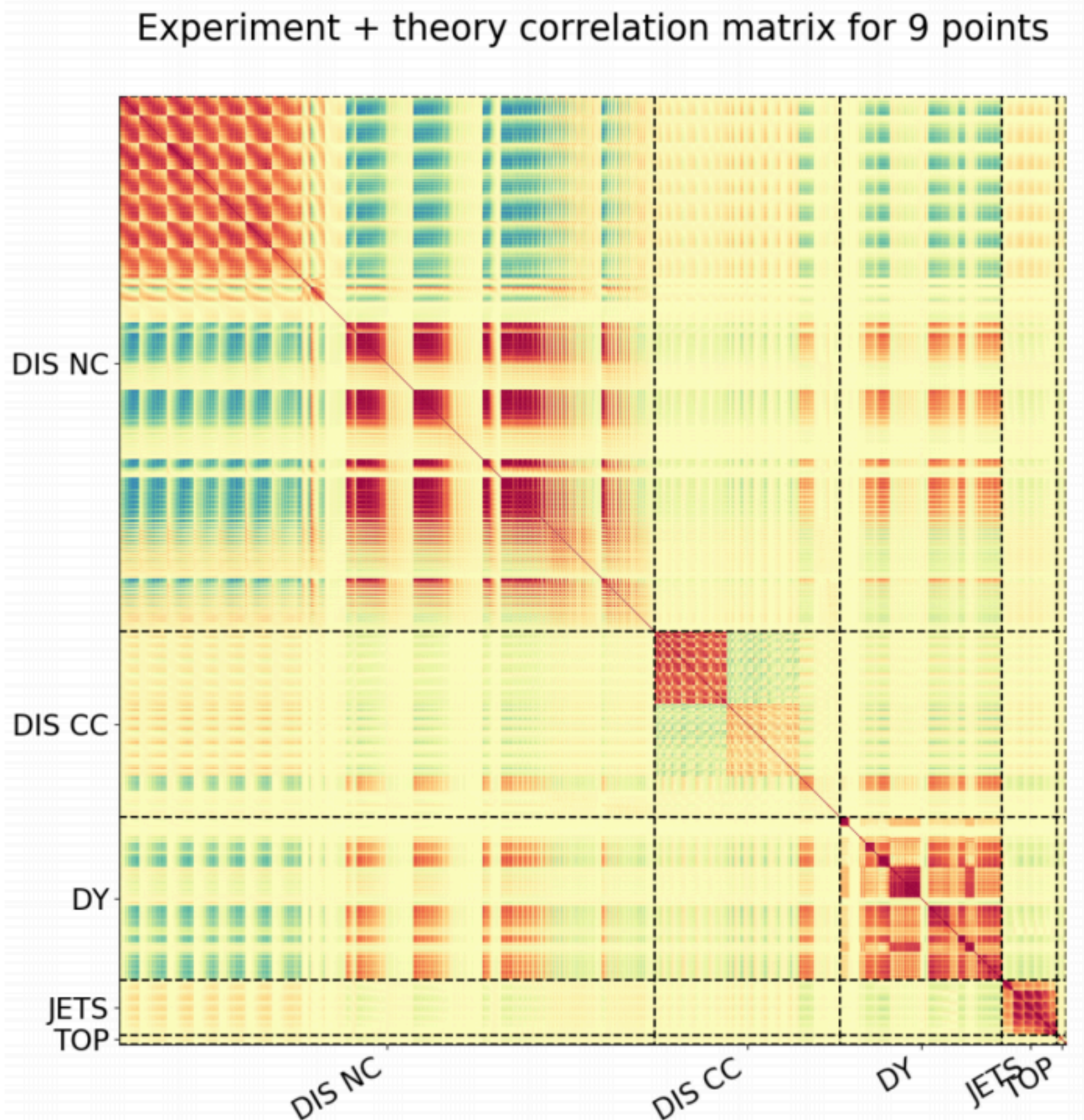
# THEORY UNCERTAINTIES IN PDF FITS

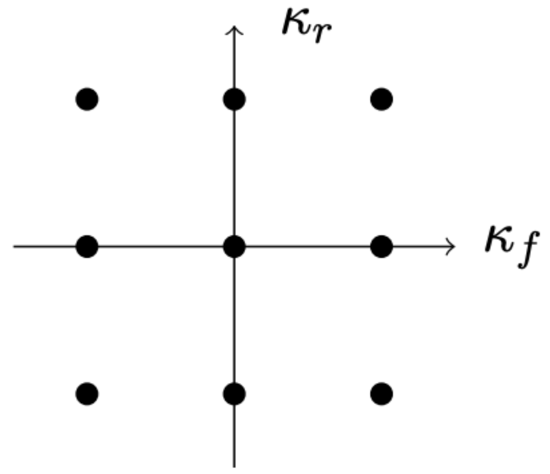
- ➡ How to include Missing Higher Order Uncertainties in a PDF fit?
- ➡ Construct a theory covariance matrix from scale-varied cross sections and combine it with the experimental covariance matrix [R.D. Ball et al, Eur.Phys.J.C 79 (2019) 11, 931, Eur.Phys.J. C (2019) 79:838]

$$\chi^2 = \sum_{m,n=1}^{N} (d_m - t_m)(\mathrm{cov}_{\mathrm{exp}} + \mathrm{cov}_{\mathrm{th}})^{-1}_{mn}(d_n - t_n)$$

- ➡ Assumptions: experimental and theoretical errors independent and Gaussian
- ➡ Accounting for the theory covariance matrix will modify the relative weight that each of the datasets carries in the global fit: processes with higher MHOUs will be ``de-weighted''
- ➡ Assumptions on correlation of scales and scale ratio will determine the form of the covariance matrix

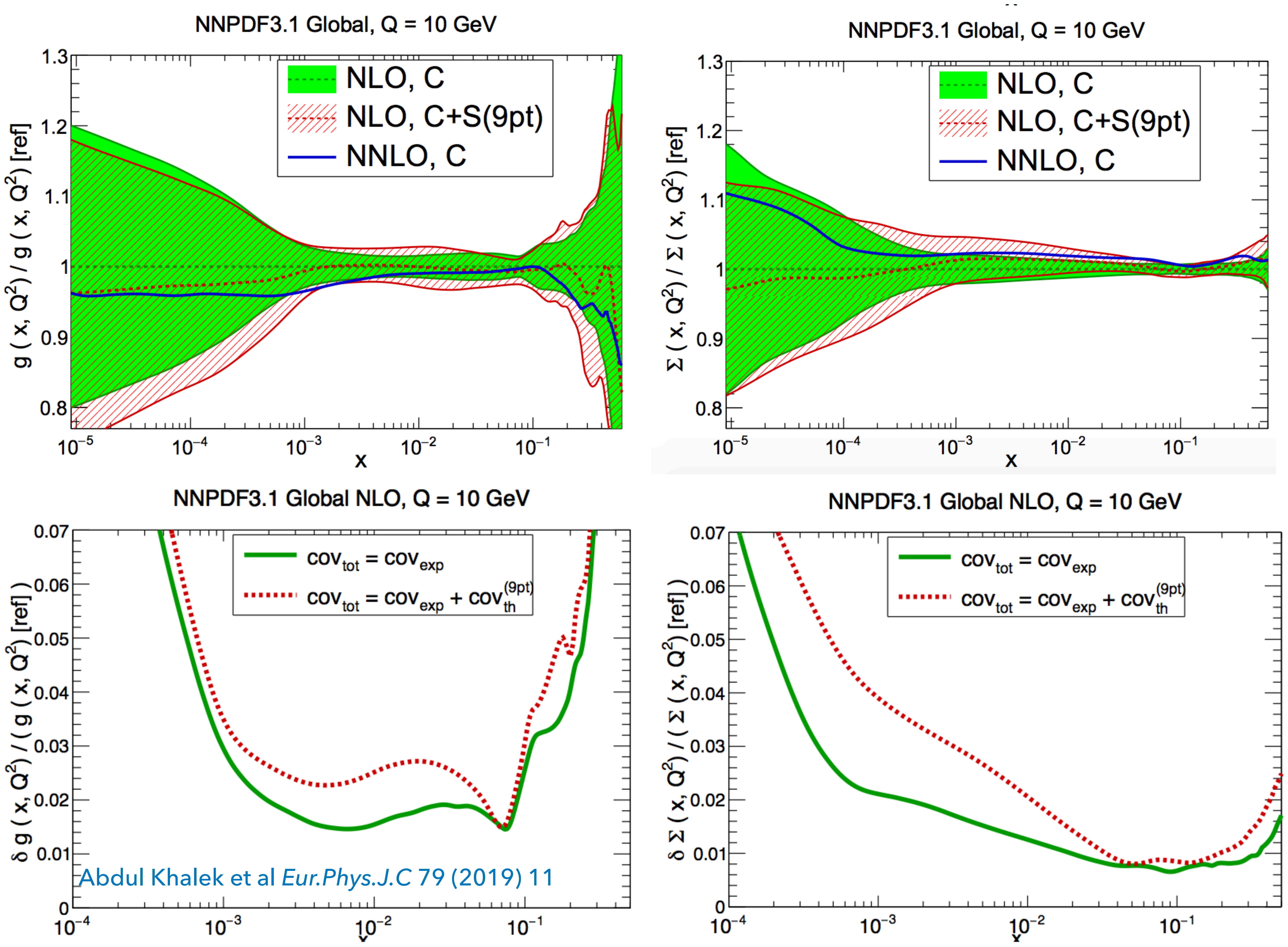


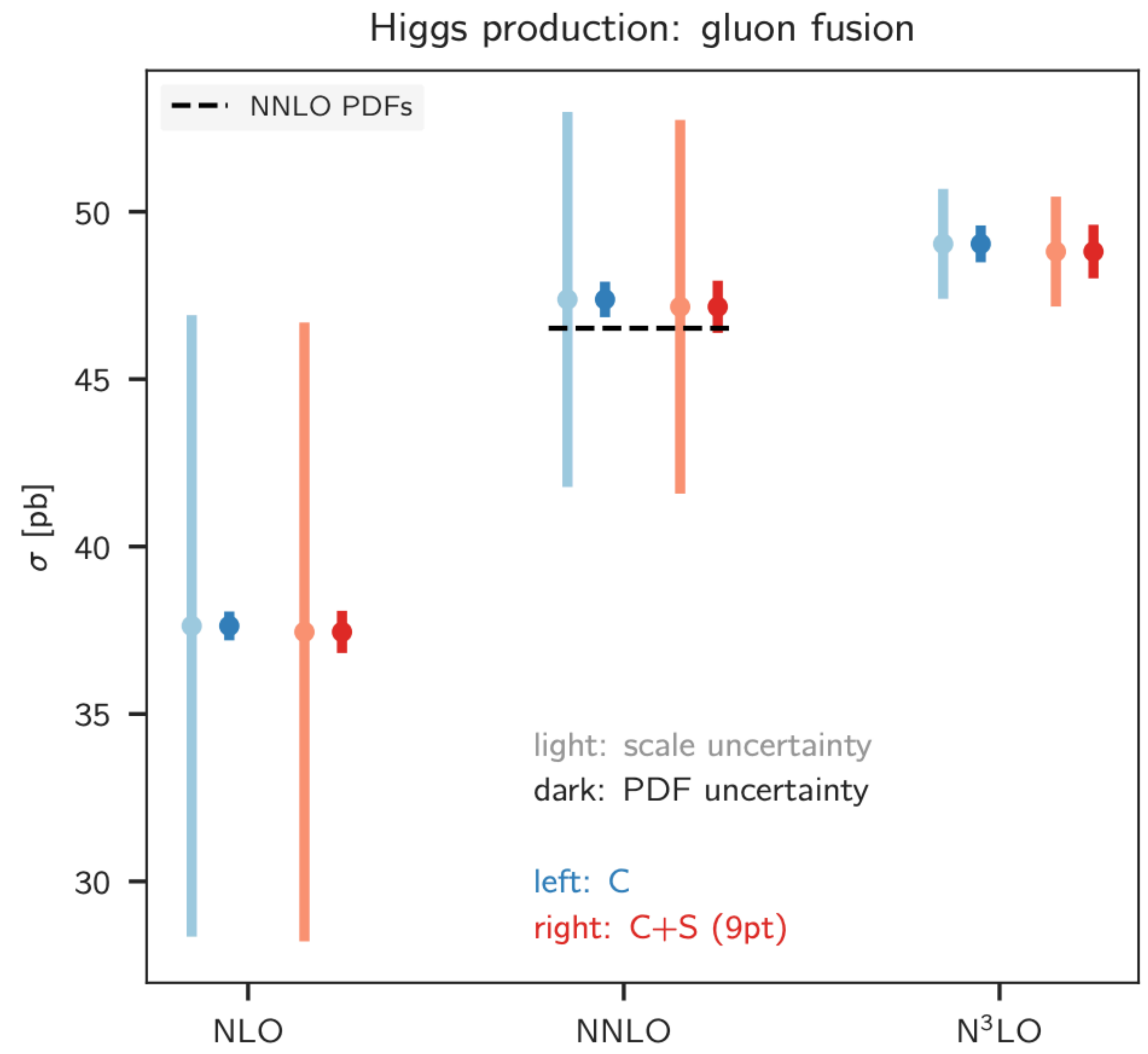



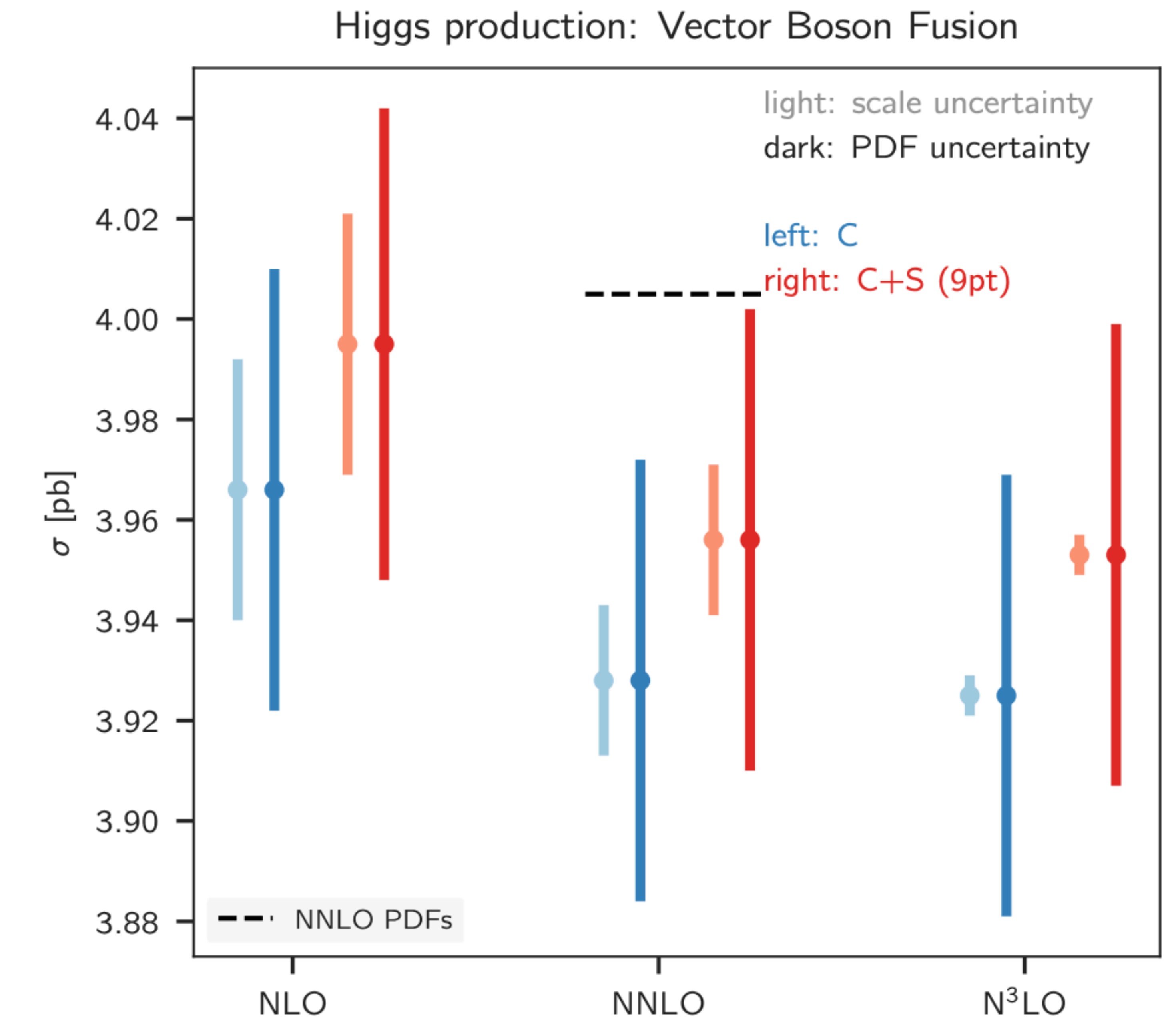

Abdul Khalek et al
*Eur.Phys.J.C 79 (2019) 11*

# THEORY UNCERTAINTIES IN PDF FITS




Abdul Khalek et al *Eur.Phys.J.C* 79 (2019) 11




➡ Missing correlation between scale variation in PDF fits and hard cross sections [Harland-Lang, Thorne Eur.Phys.J.C 79 (2019) 3, 225] might yield conservative estimate

# ELECTROWEAK CORRECTIONS

# NLO EW CORRECTIONS FOR PDF FITS

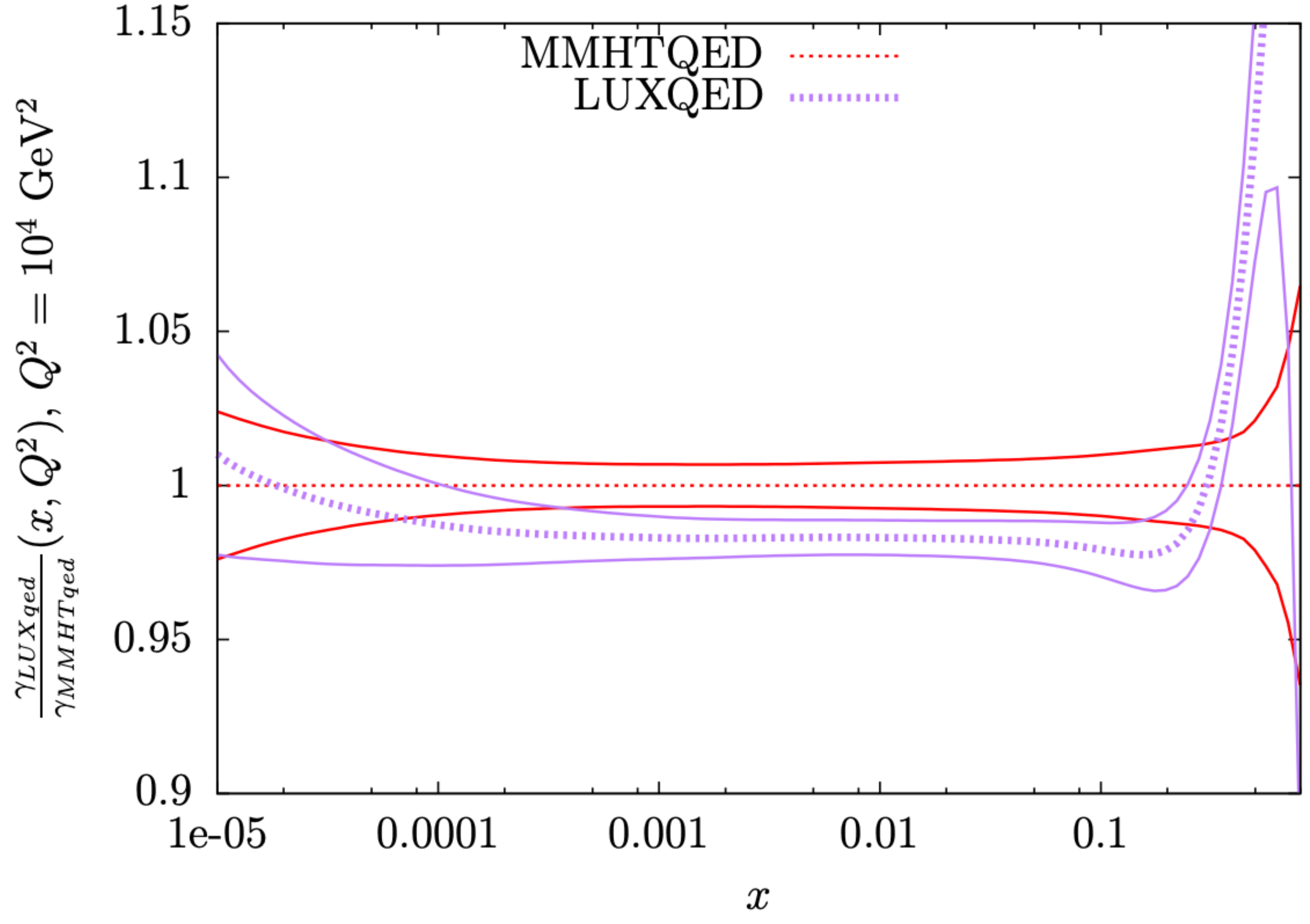
- If we aim to 1% accuracy NLO EW corrections do matter.
- As the LHC reaches higher invariant mass regions and pT regions EW corrections increasingly relevant
- Currently NLO EW corrections included for some observables in MSHT20 (via K-factors)
- NLO EW corrections not included by default in NNPDF4.0 nor CT18 (large invariant mass regions, large Z pT excluded => less constraints to large-x PDFs)

What is needed

➡ QED corrections in DGLAP evolution [Bertone et al, APFEL, Comput.Phys.Commun. 185 (2014)]

➡ Photon PDFs [Manohar et al Phys.Rev.Lett. 117 (2016), Bertone et al SciPost Phys. 5 (2018) 1, 008, Harland-Lang JHEP 79 (2019) 10]

➡ NLO EW corrections for all processes included in PDF fits [Frederix et al,JHEP 07 (2018) 185]




Harland-Lang JHEP 79 (2019) 10

# NLO EW CORRECTIONS FOR PDF FITS

- If we aim to 1% accuracy NLO EW corrections do matter.
- As the LHC reaches higher invariant mass regions and pT regions EW corrections increasingly relevant
- Currently NLO EW corrections included for some observables in MSHT20 (via K-factors)
- NLO EW corrections not included by default in NNPDF4.0 nor CT18 (large invariant mass regions, large Z pT excluded => less constraints to large-x PDFs)

What is needed

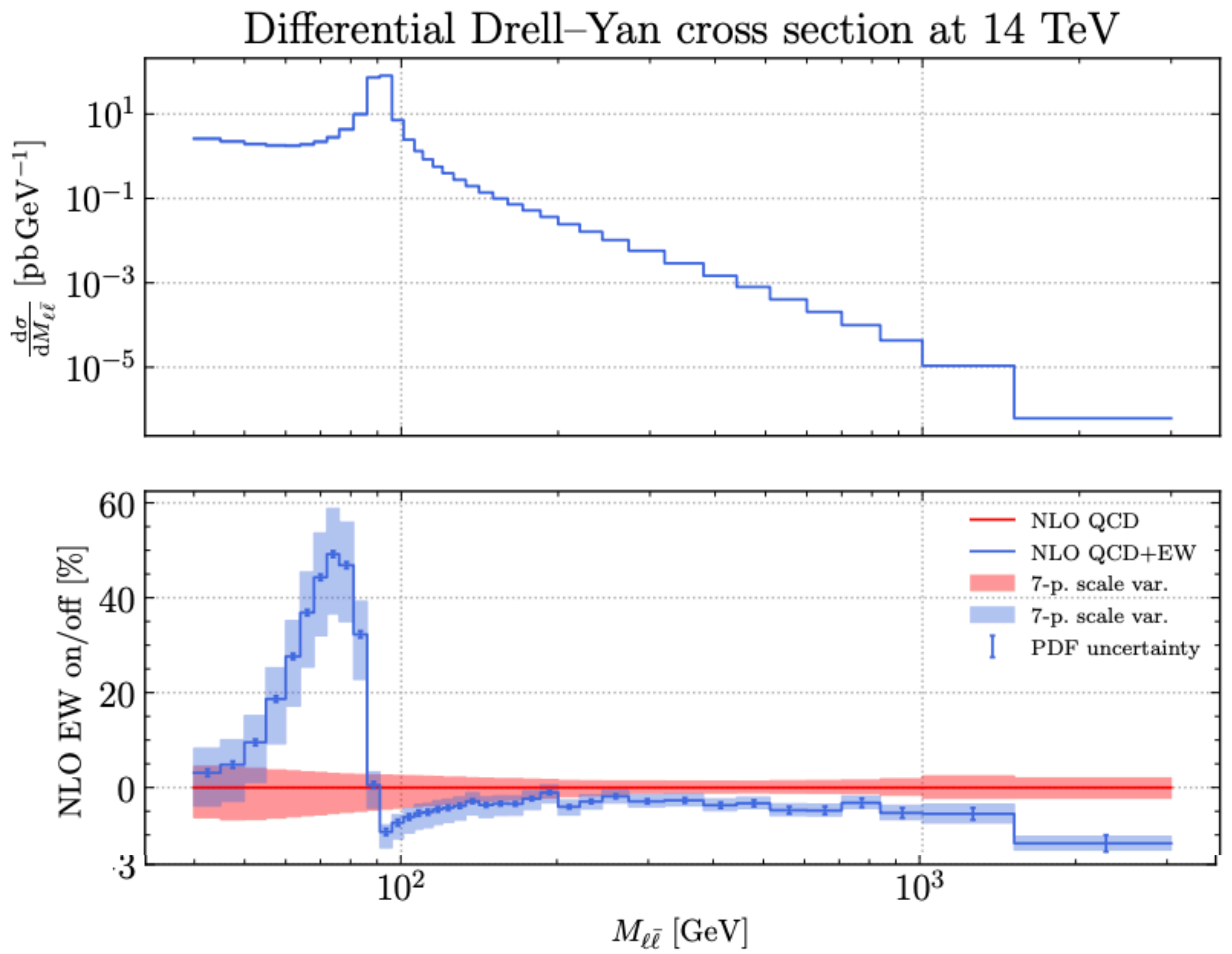
➡ QED corrections in DGLAP evolution [Bertone et al, APFEL, Comput.Phys.Commun. 185 (2014)]

➡ Photon PDFs [Manohar et al Phys.Rev.Lett. 117 (2016), Bertone et al SciPost Phys. 5 (2018) 1, 008, Harland-Lang JHEP 79 (2019) 10]

➡ NLO EW corrections for all processes included in PDF fits [Frederix et al,JHEP 07 (2018) 185]

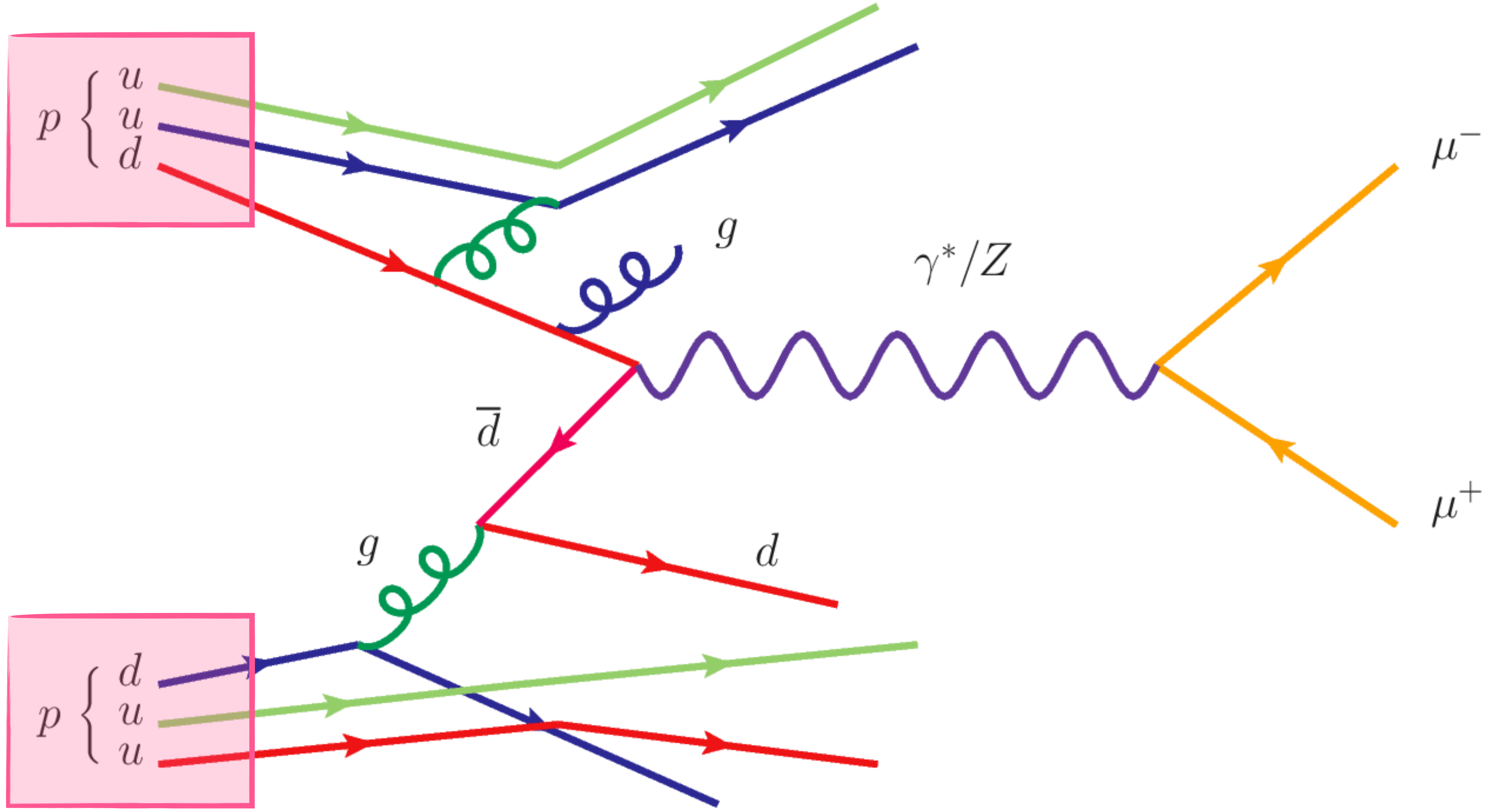
➡ Fast interpolating grids: **PineAPPL** [Carrazza, Nocera, Schwan, Zaro JHEP 12 (2020) 108]

➡ Careful scrutiny of data (no FSR nor photon-initiated subtraction & observable definition in experimental data) is probably going to be the point requiring more work



# PDFS AND NEW PHYSICS INTERPLAY IN HIGH ENERGY TAILS

# PDF AND SMEFT FIT INTERPLAY




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

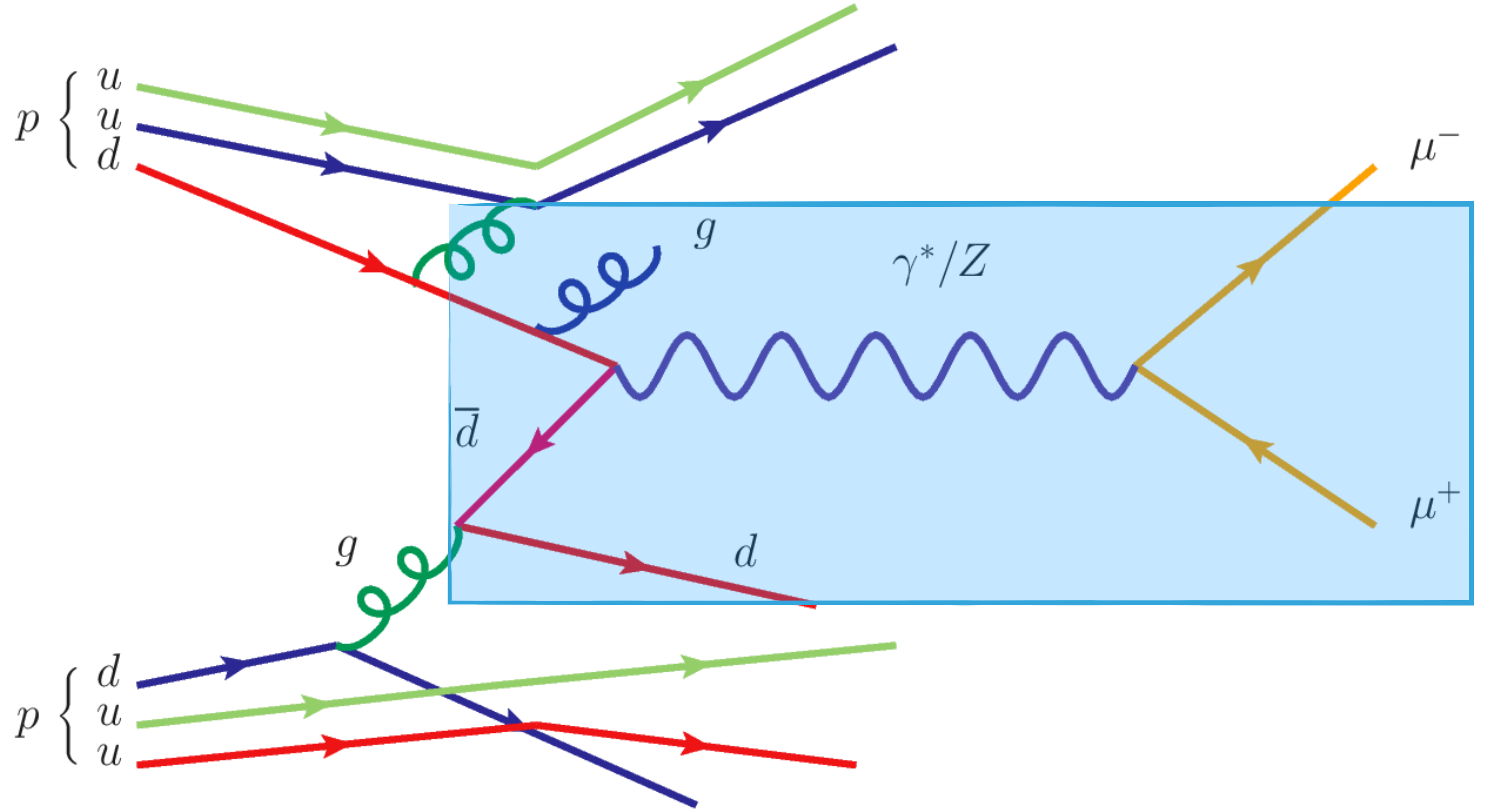
Fit PDF parameters $\{\theta_l\}$

LHC Data

$f_i(\{\theta_k\})$

Assume SM in theory predictions

# PDF AND SMEFT FIT INTERPLAY




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

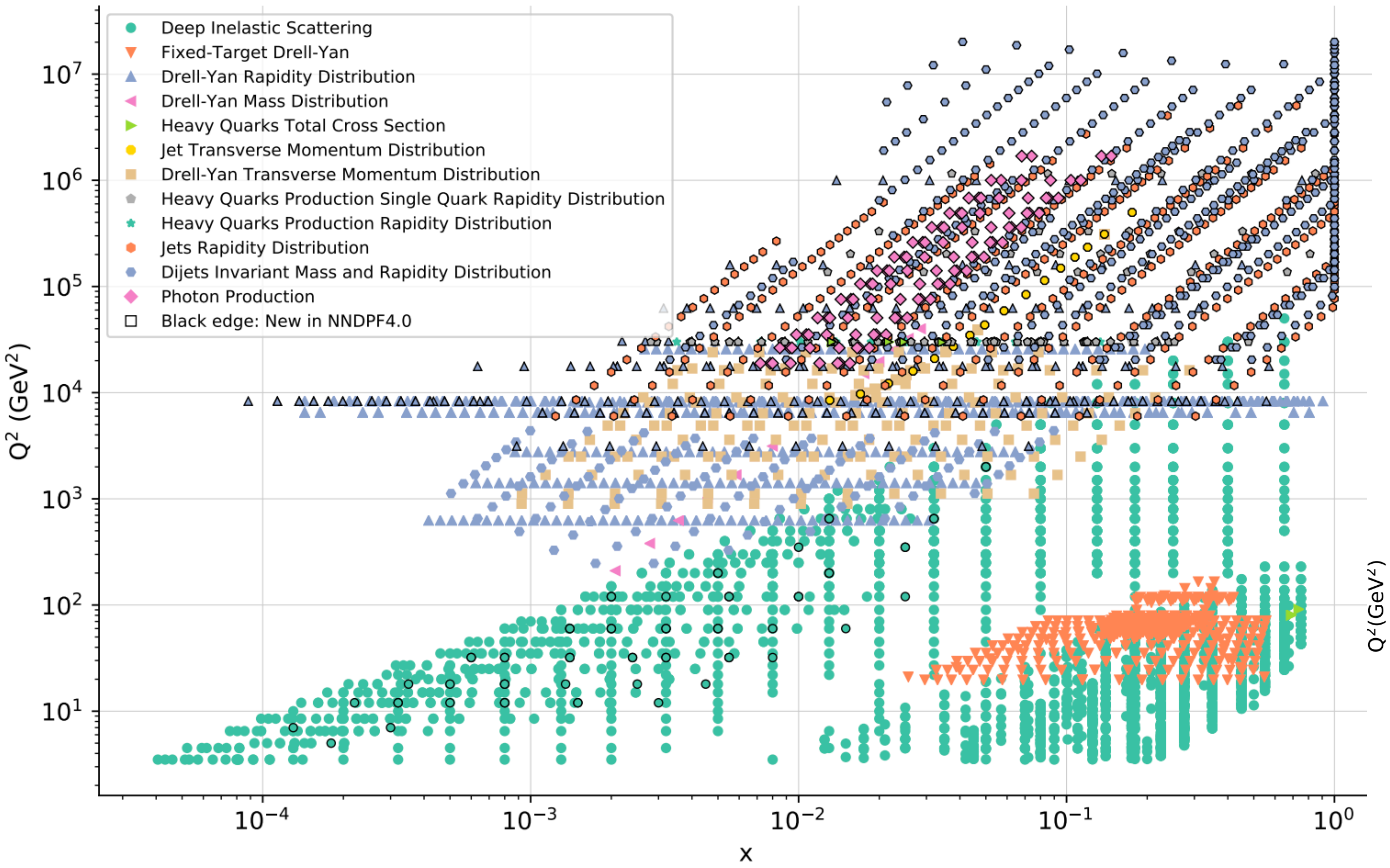
Fit of Wilson Coefficients $\{c_i\}$

LHC Data

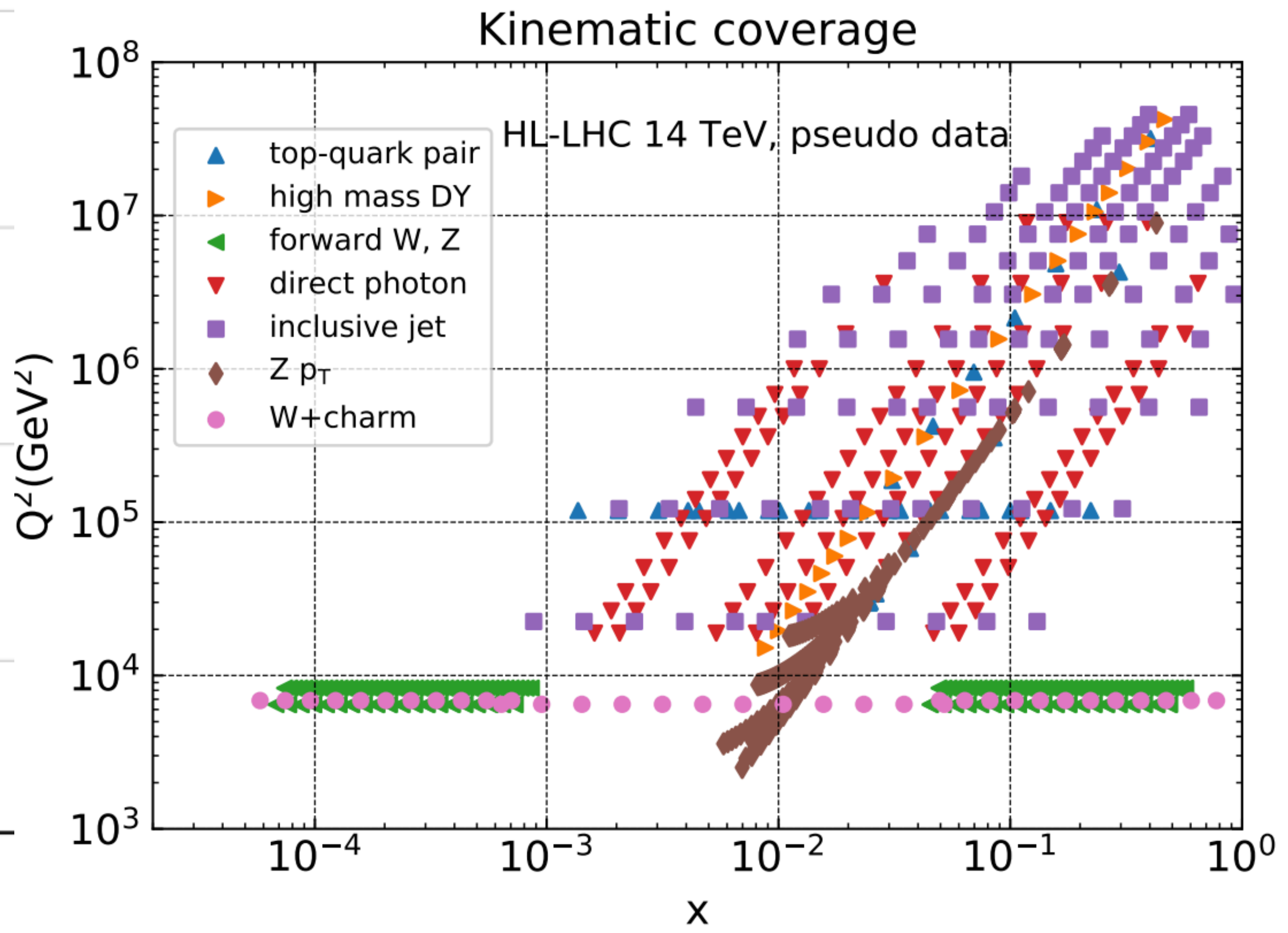
Assume SM PDFs

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

# KINEMATIC OVERLAP




- ➡ Top pair production and single top data included in [Hartland et al 1901.05965]
- ➡ Dijets data in [Bordone et al 2103.10332] [Alioli et al 1706.03068 ]
- ➡ Drell-Yan data in [Farina et al 1609.08157, Torre et al 2008.12978]
- ➡ Inclusive jets in [Alte et al 1711.07484]
- ➡ Overlap enhanced in HL-LHC projections [Abdul Khalek et al1810.03639]

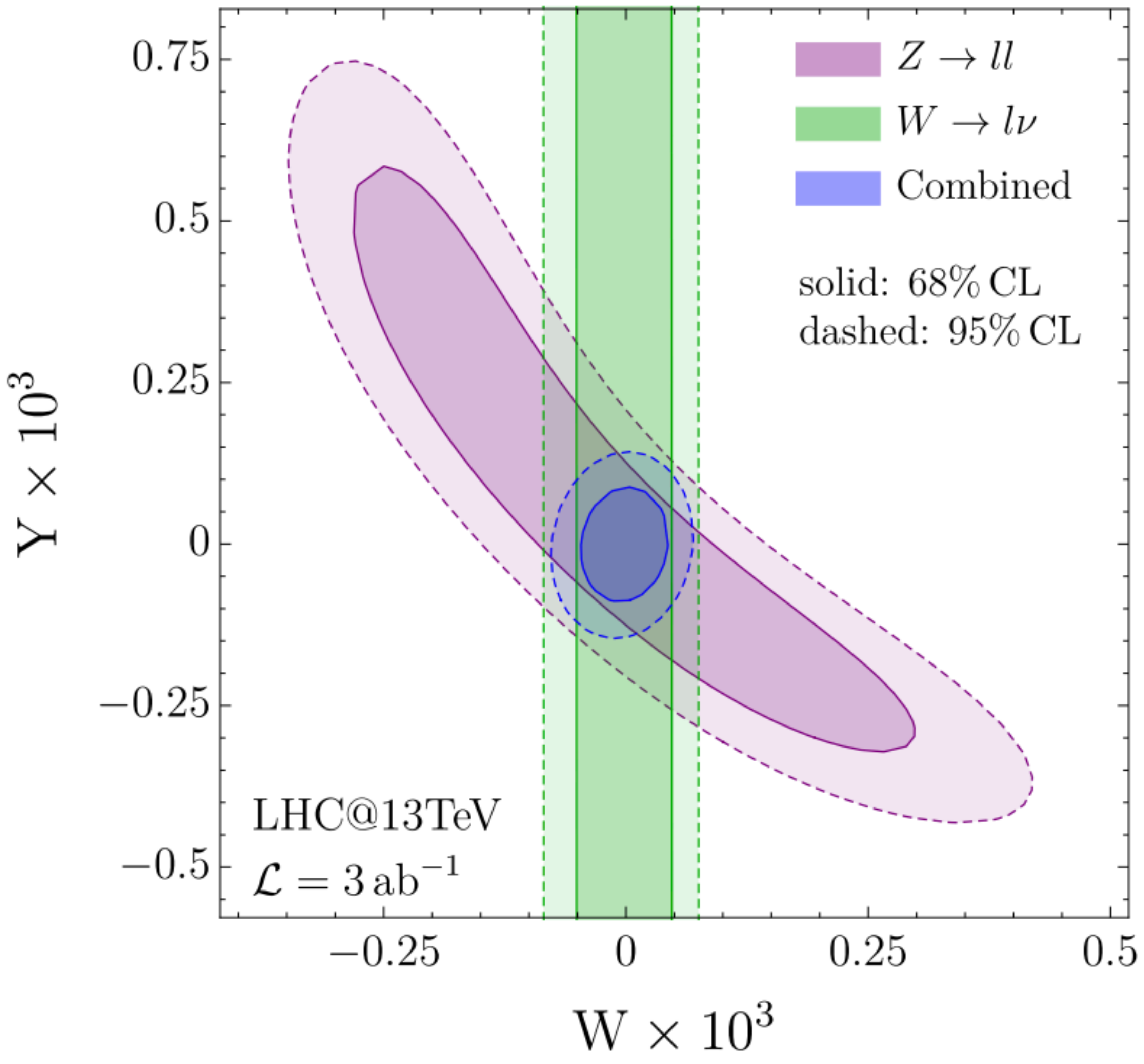


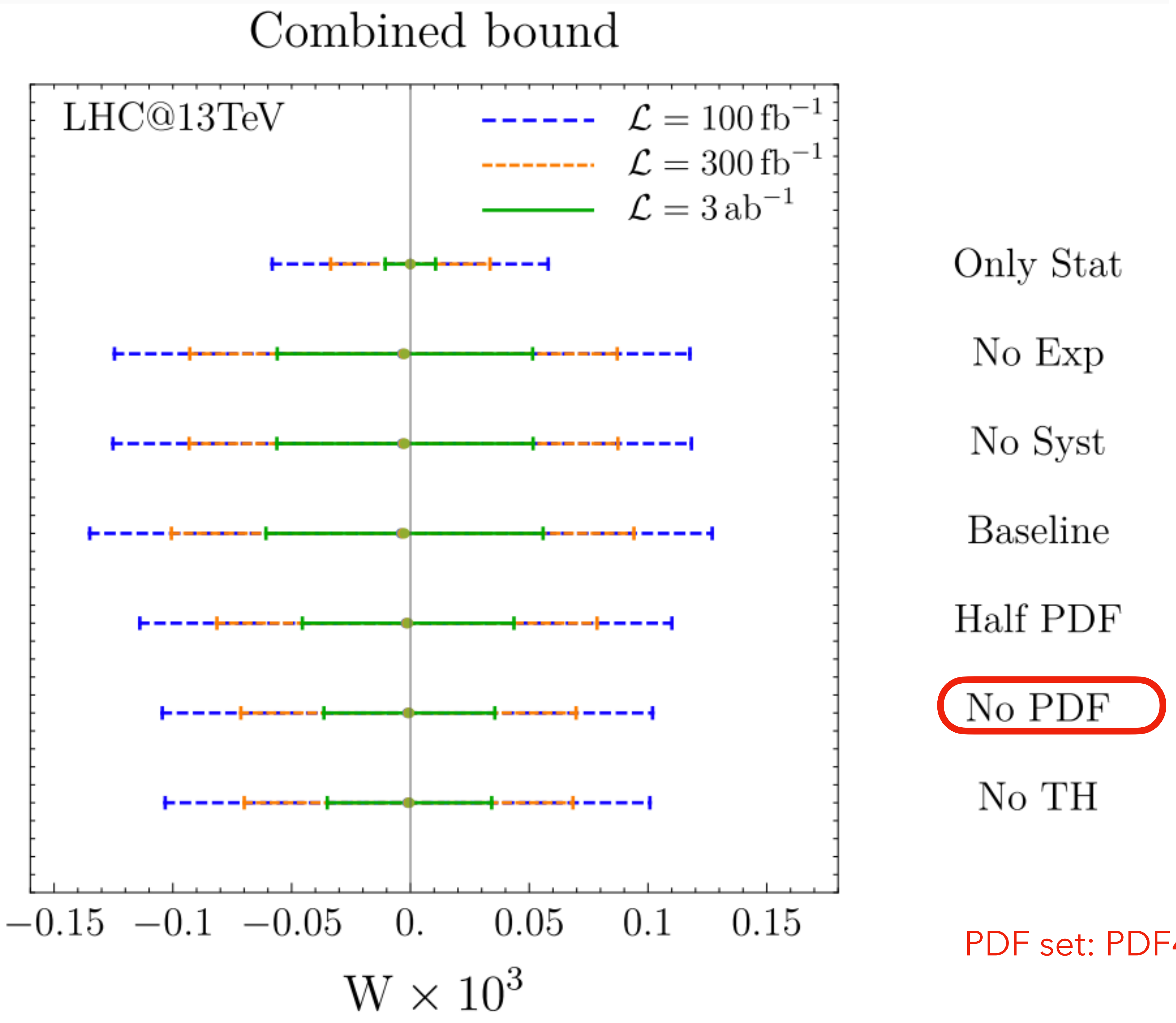
# SHOULD WE WORRY ABOUT PDF AND SMEFT INTERPLAY?

- First study of interplay in case of DIS data
  [Carrazza et al Phys.Rev.Lett. 123 (2019) 13, 132001]
- Case study: EW oblique corrections in high-mass NC and CC Drell-Yan tails
- They are powerful probes of quark-lepton contact interactions that produce growing-with-energy effects. [Torre et al, 2008.12978]

$$\mathscr{L}_{\text{SMEFT}} \supset -\frac{\hat{W}}{4m_W^2}(D_\rho W^a_{\mu\nu})^2 - \frac{\hat{Y}}{4m_W^2}(\partial_\rho B_{\mu\nu})^2$$

Torre et al, 2008.12978

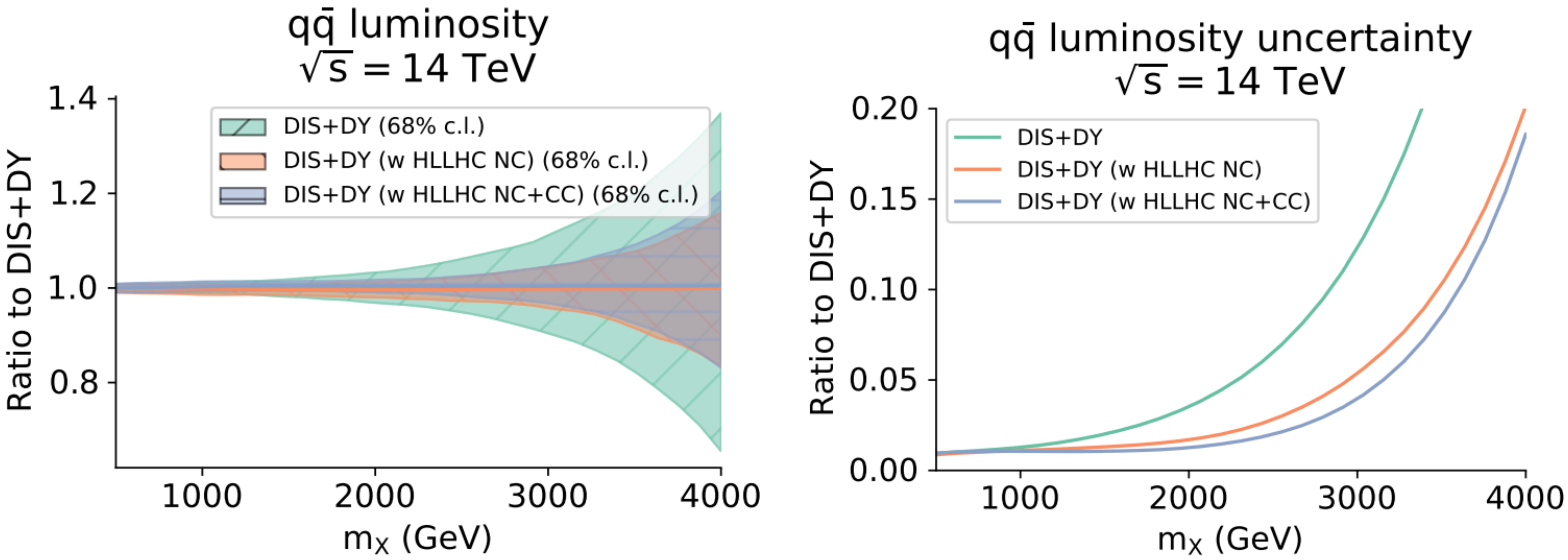



PDF set: PDF4LHC15_nlo_30_pdfas

# SHOULD WE WORRY ABOUT PDF AND SMEFT INTERPLAY?

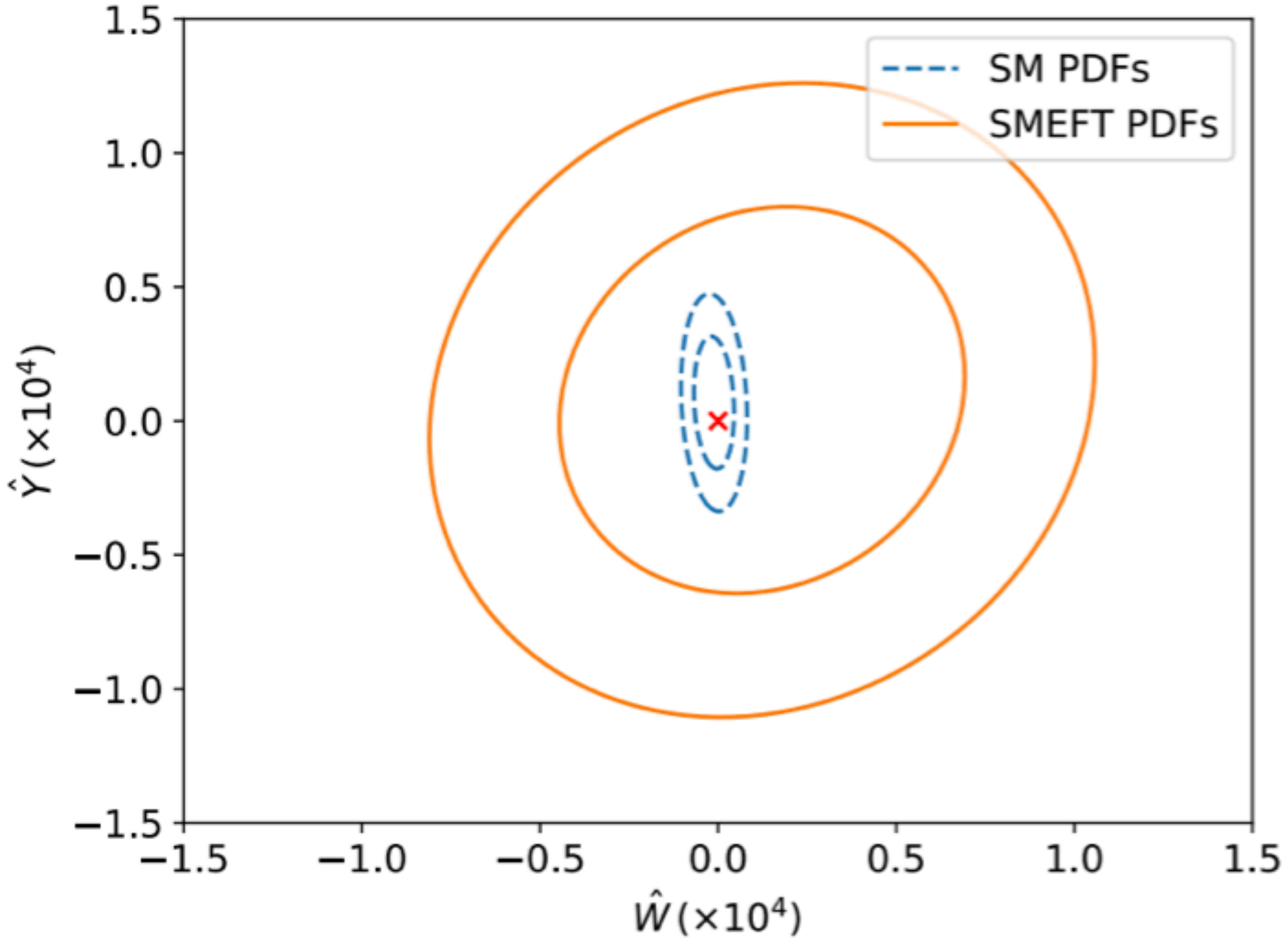
- A similar analysis has been performed, now with emphasis on PDF and their interplay with bounds on oblique operators in [Greljo et al, 2104.02723]
- Settings:
  - ➡ PDF fit based on DIS, Drell-Yan on-shell and low-mass data from ATLAS, CMS and LHCb
  - ➡ + Run I and II ATLAS and CMS high mass NC Drell-Yan data
  - ➡ + HL-LHC projections for NC and CC Drell-Yan data
  - ➡ SM predictions at NNLO QCD + NLO EW and SMEFT corrections added via local K-factors

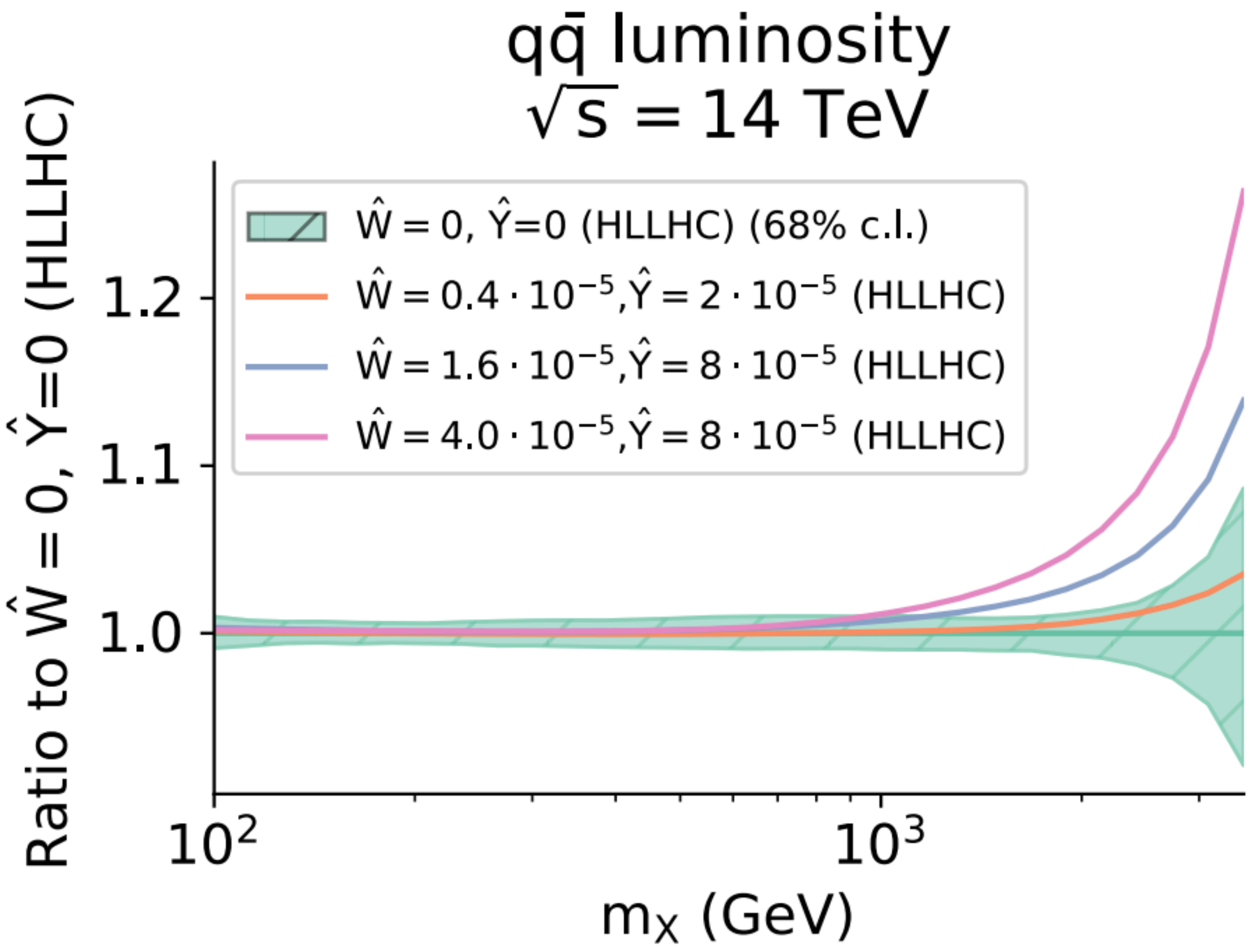


# SHOULD WE WORRY ABOUT PDF AND SMEFT INTERPLAY?

- While mild interplay with Run I and Run II data, first evidence of interplay at **HL-LHC**
- Compare Wilson coefficients bounds from HL-LHC projections assuming SM PDFs (that include NC+CC data) to the bounds on the same Wilson coefficients obtained from a simultaneous fit of PDFs and Wilson coefficients
- Not accounting for interplay (using PDFs as a black box) leads to over-constrained bounds

Greljo et al, 2104.02723

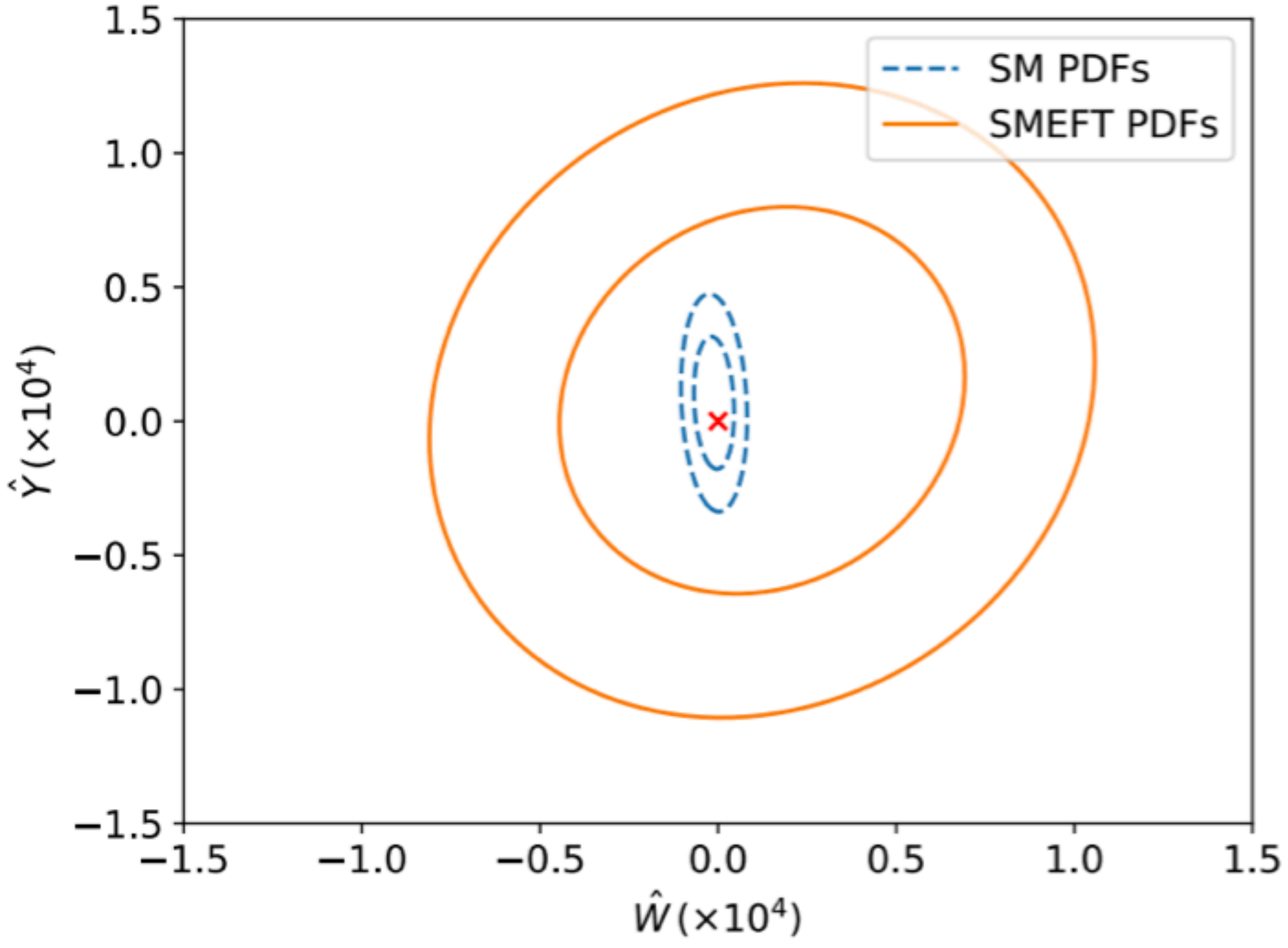


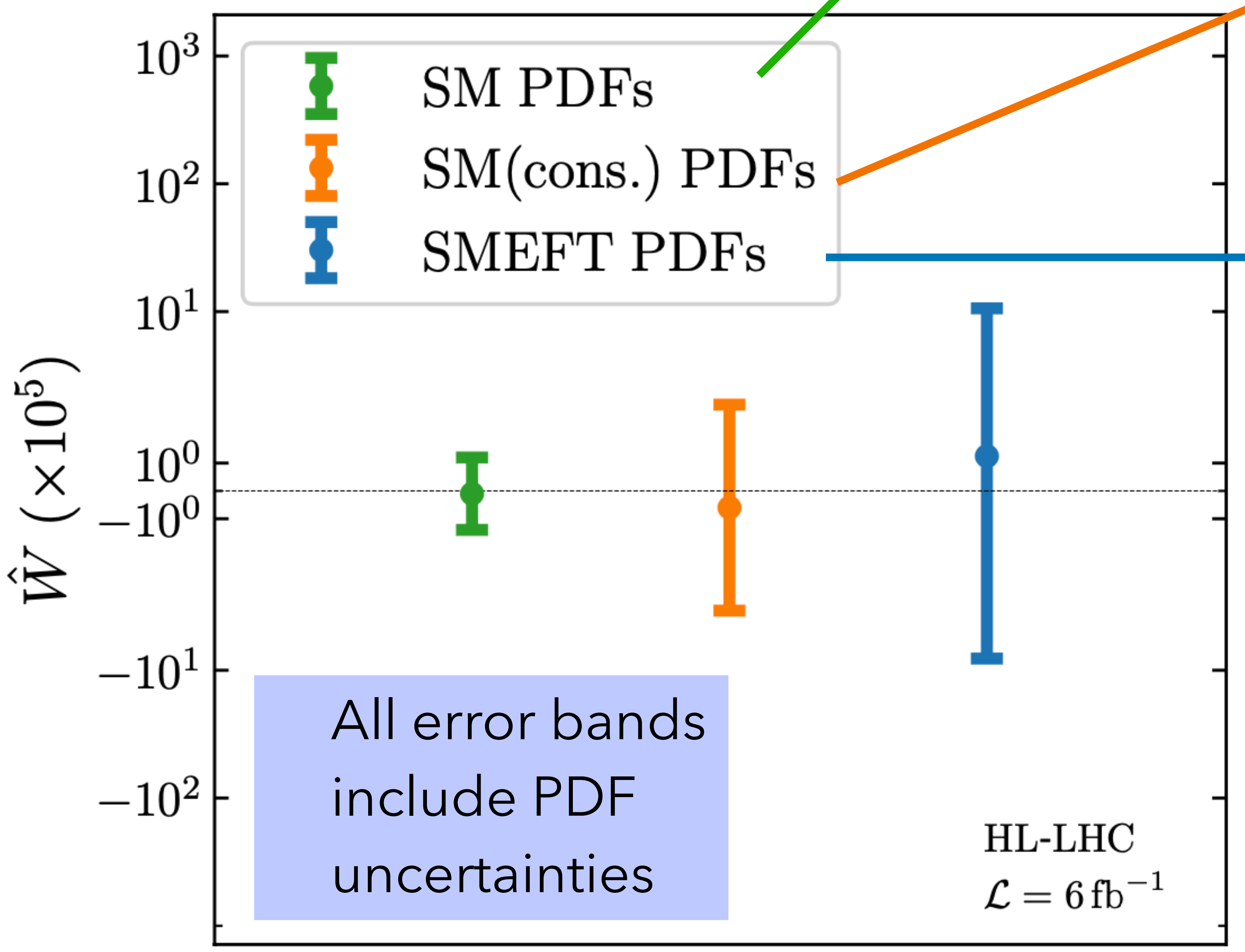



# SHOULD WE WORRY ABOUT PDF AND SMEFT INTERPLAY?

- While mild interplay with Run I and Run II data, first evidence of interplay at **HL-LHC**
- Compare Wilson coefficients bounds from HL-LHC projections assuming SM PDFs (that include NC+CC data) to the bounds on the same Wilson coefficients obtained from a simultaneous fit of PDFs and Wilson coefficients
- Not accounting for interplay (using PDFs as a black box) leads to over-constrained bounds
- Simultaneous fits? Conservative PDF set?

Greljo et al, 2104.02723




PDF set with all data in

"Conservative" PDF set without any high-mass DY data

Simultaneous PDF and SMEFT fit

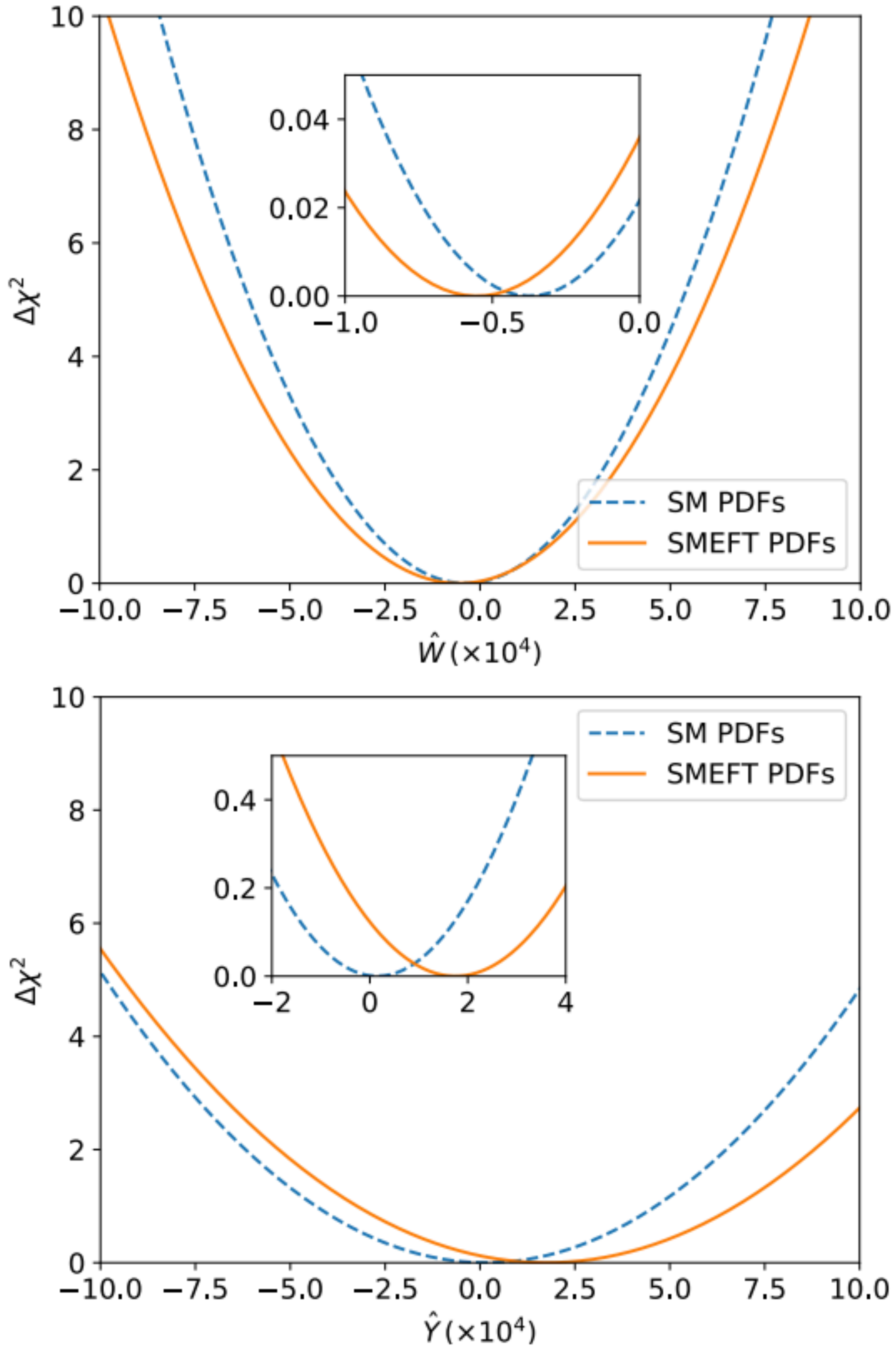
# ADDITIONAL POINTS FOR DISCUSSION

- The path towards 1% accurate PDFs opens up new fascinating challenges
- Precise and accurate determination of the proton structure is key to make progress in precision phenomenology
- Need: robust methodology & solid statistical tests of methodology
- NNPDF4.0 yields precise PDFs for phenomenology and not is beyond the sets included in PDF4LHC21 combination
- PDF combination versus individual sets?

- Current uncertainties partially underestimated due to theory uncertainty not accounted for at NNLO & other effects (fixed-order calculations, full account of electroweak corrections…)
- Proposal for inclusion of missing higher order uncertainties has been discussed (how important is the missing information on correlation between scale variation in processes used in PDF fit and those in the hard partonic cross section that one wishes to compute?)

- Time to study the interplay between indirect new physics and PDFs
- The preferred avenue ahead is to be able to perform simultaneous fits (like for PDFs and $\alpha_s$)
- In parallel a more careful investigation of definition of conservative PDF sets & account for PDF uncertainties
- Also, would be important to disentangle large-x from high-energy / low-energy (LHCb) as well as scaling behaviour (ratios at different centre of mass energies?)

# EXTRA MATERIAL




$$R_{\chi^2}(m_{\ell\ell}^{(\max)}, \hat{W}, \hat{Y}) \equiv \frac{\chi^2(m_{\ell\ell}^{(\max)}, \hat{W}, \hat{Y})}{\chi^2(120\ \text{GeV}, \hat{W}, \hat{Y})}$$