Global PDF Analyses and Precision Physics

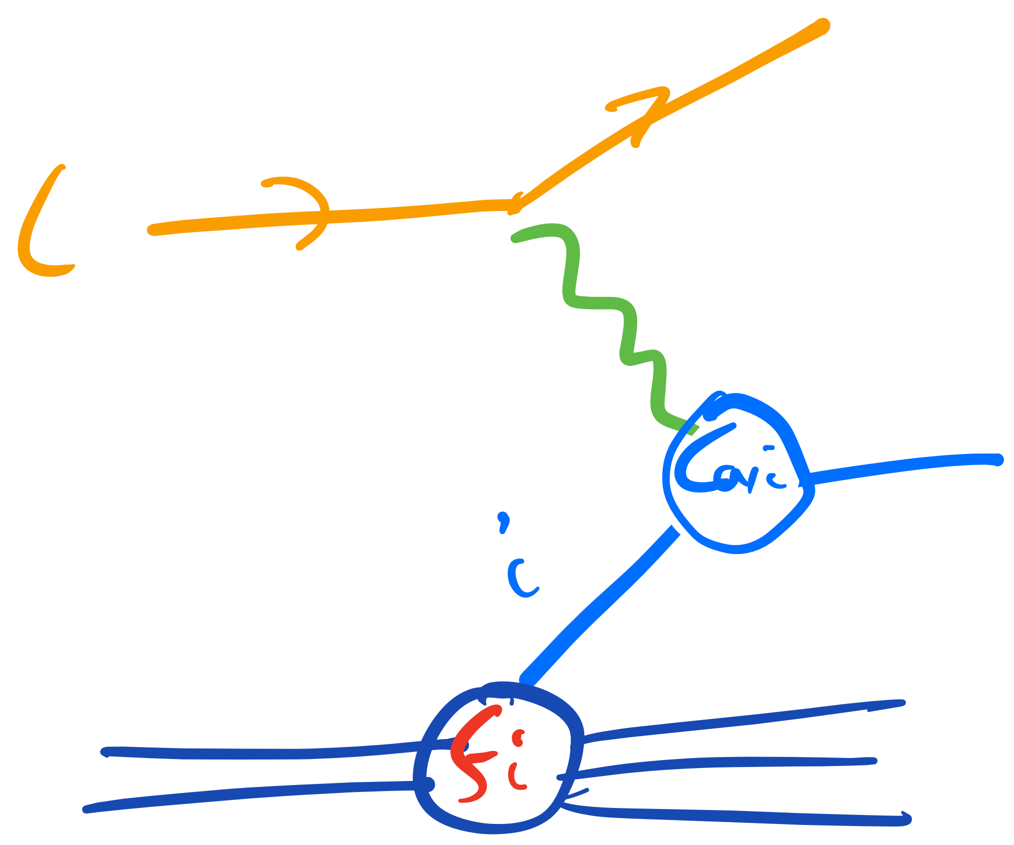
Lucian Harland-Lang, University College London

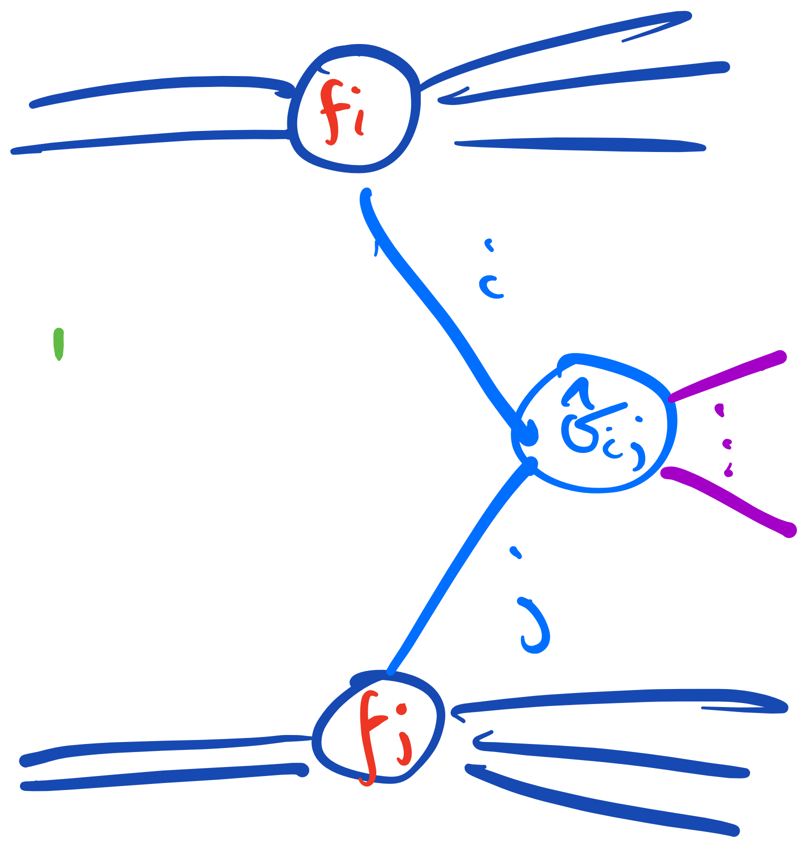
Lattice@CERN 2024, July 9th 2024

Setting the Scene…

Collinear, unpolarized

- Parton distribution functions (PDFs): a key ingredient in hadron collider physics!
- QCD factorization: perturbative physics separated from **universal** non-perturbative PDFs





$$F_a(x,Q^2) = \sum_{i=q,\bar{q},g} \int_0^1 \frac{dz}{z} f_i(z,Q^2) C_{a,i}\left(\frac{x}{z}, \alpha_S(Q^2)\right) + \mathcal{O}\left(\frac{\Lambda^2_{QCD}}{Q^2}\right)$$

$$\sigma = \sum_{ij} \int_{x_{min}}^1 dx_1 dx_2\, f_i(x_1,\mu_f^2) f_j(x_2,\mu_f^2) \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, Q, \mu_F^2)$$

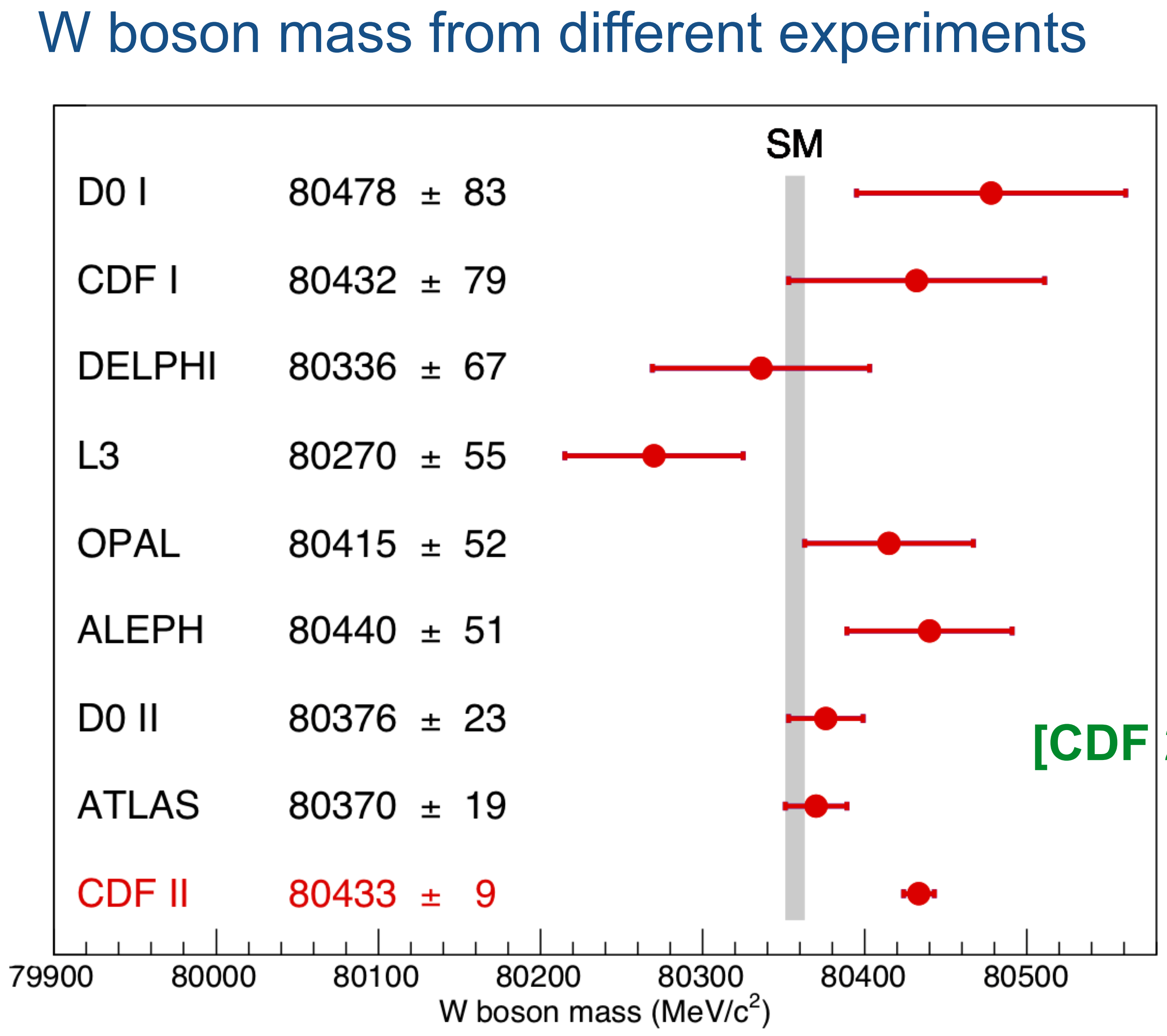
$$\text{Factorization} \Rightarrow f_i^{\text{DIS}}(x,Q^2) \equiv \{f_i^{\text{Collider}}(x,Q^2)\}$$ ← **Drell Yan, Jets, Higgs…**

- PDFs at different scales connected by DGLAP evolution $\frac{\partial f_q^{NS}(x,\mu^2)}{\partial \log \mu^2} = \frac{\alpha_S}{2\pi} \int_x^1 \frac{dz}{z} f_q^{NS}(z,\mu^2) P_{qq}^{NS}(x/z)$ **etc…**
- Foundation of global PDF fits: use data at different scales and processes to extract PDFs.

Why do we care about PDFs?

- The LHC is a Standard Model precision machine, and PDFs are a key ingredient in this. Increasingly a limiting factor:

W mass



[CDF 2022]

Fig. 5

SM expectation: $M_W = 80{,}357 \pm 4_{\text{inputs}} \pm 4_{\text{theory}}$ (PDG 2020)

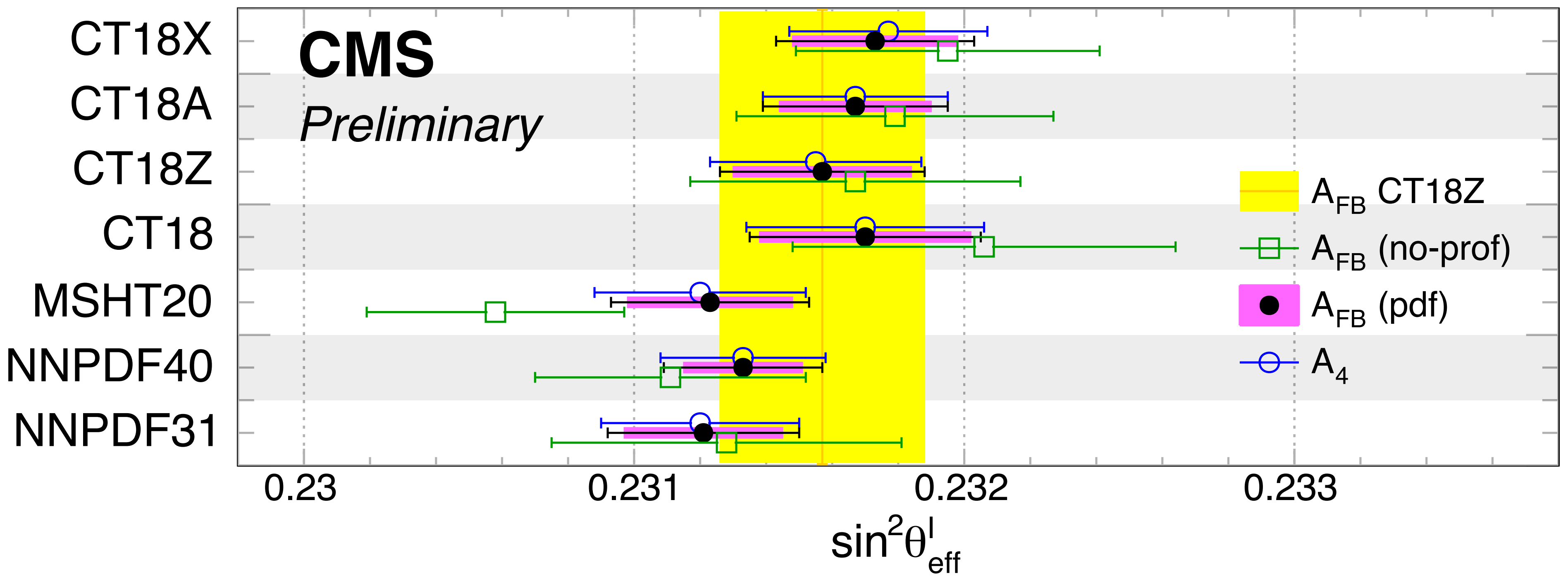
LHCb measurement : $M_W = 80{,}354 \pm 23_{\text{stat}} \pm 10_{\text{exp}} \pm 17_{\text{theory}} \pm 9_{\text{PDF}}$ *[JHEP* **2022,** 36 (2022)]

PDF unc. of CDF / ATLAS / LHCb: 3.9 / 8 / 9 MeV

$$\sigma_{\rm PDF} \sim \sigma_{\rm tot}/2$$

(up to)

$\sin^2\theta^l_{\rm eff}$

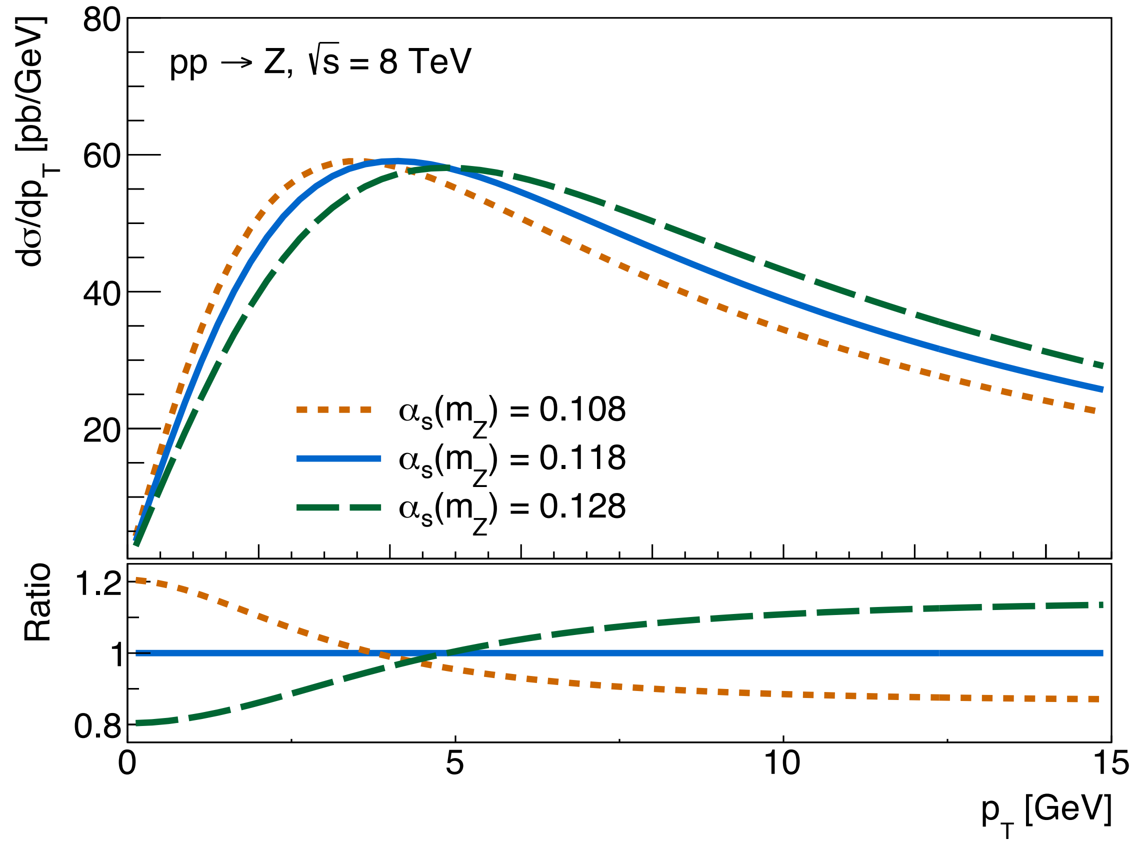


CMS PAS SMP-22-010

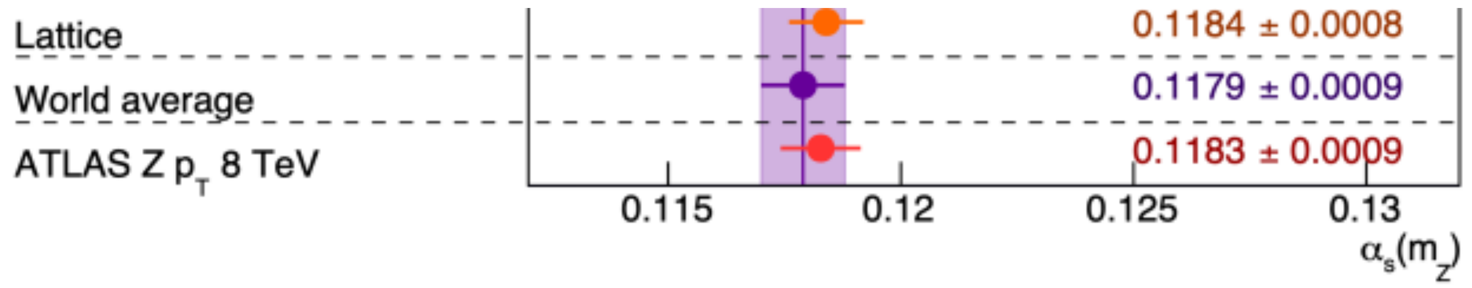
$$\sin^2\theta^\ell_{\rm eff} = 0.23157 \pm 0.00010({\rm stat}) \pm 0.00015({\rm syst}) \pm 0.00009({\rm theo}) \pm 0.00027({\rm PDF}),$$

$$\sigma_{\rm PDF} \sim \sigma_{\rm tot}$$

α_S



ATLAS, 2309.12986

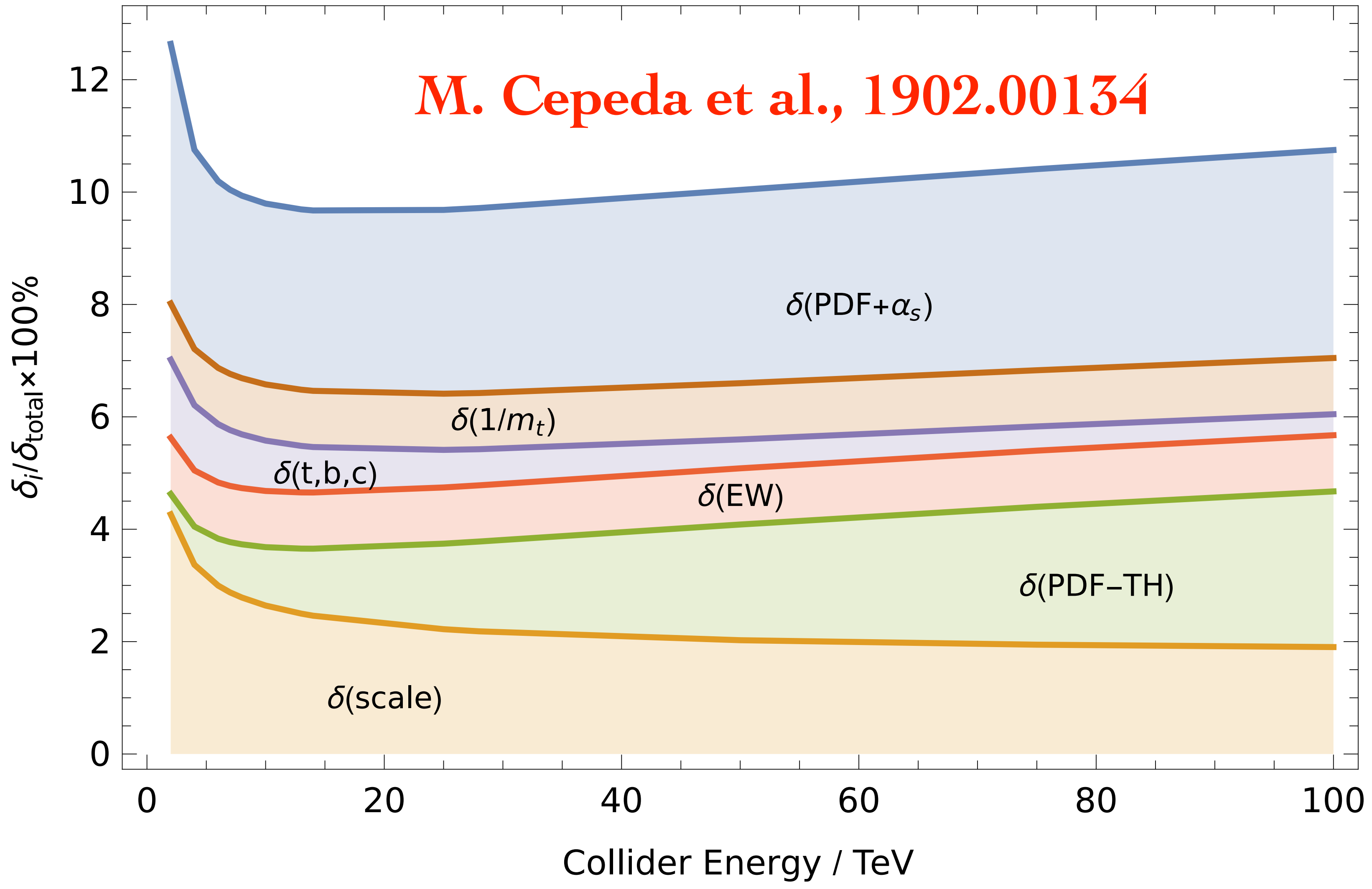


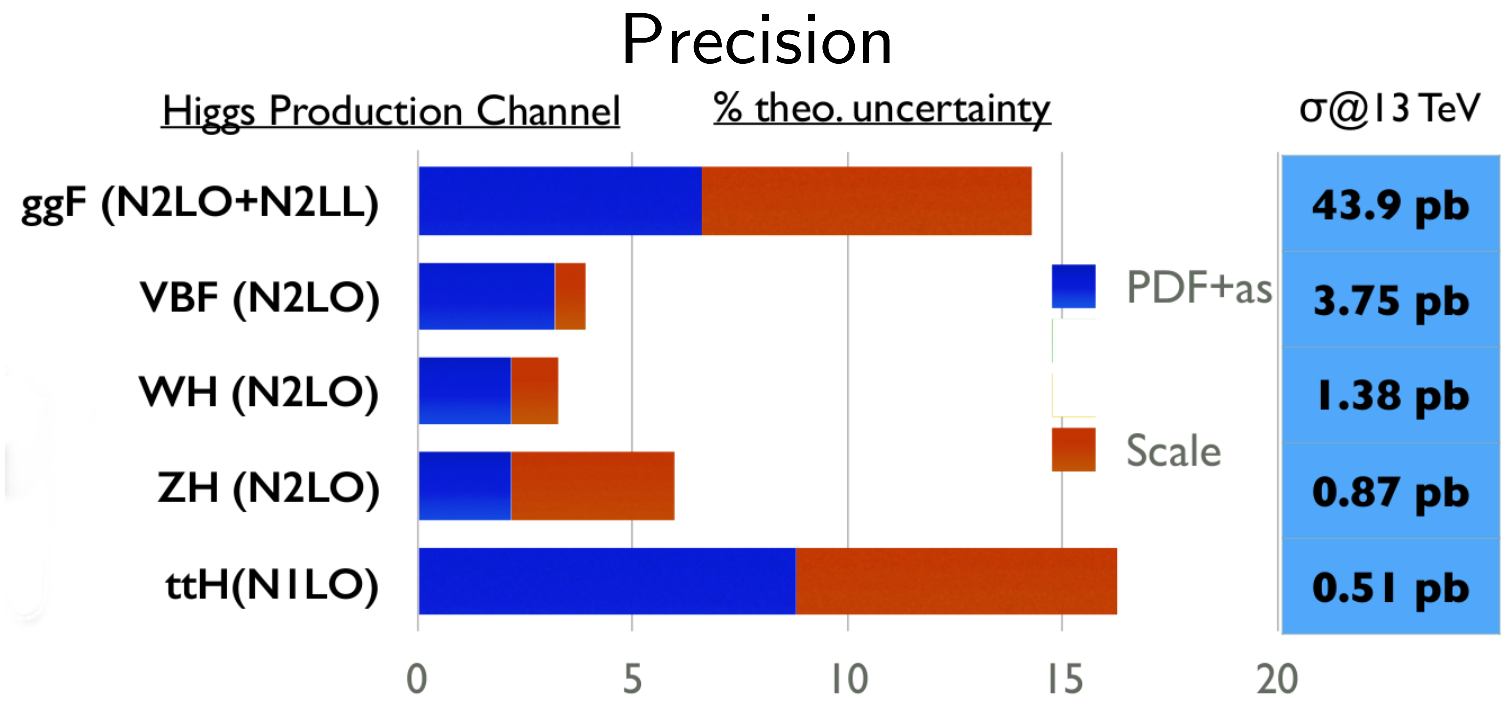
$$\sigma_{\rm PDF} \sim \sigma_{\rm tot}/2$$

Disclaimer: will generally refer to papers by their arxiv number, even if published.

- The LHC is a **Higgs** factory: PDFs play a key role here.

Image Credit: Emanuele Nocera





- The LHC is a **BSM** search machine. Often need PDFs here.
- High mass = high x , where PDFs are less well known. Key when looking for small/smooth deviations.

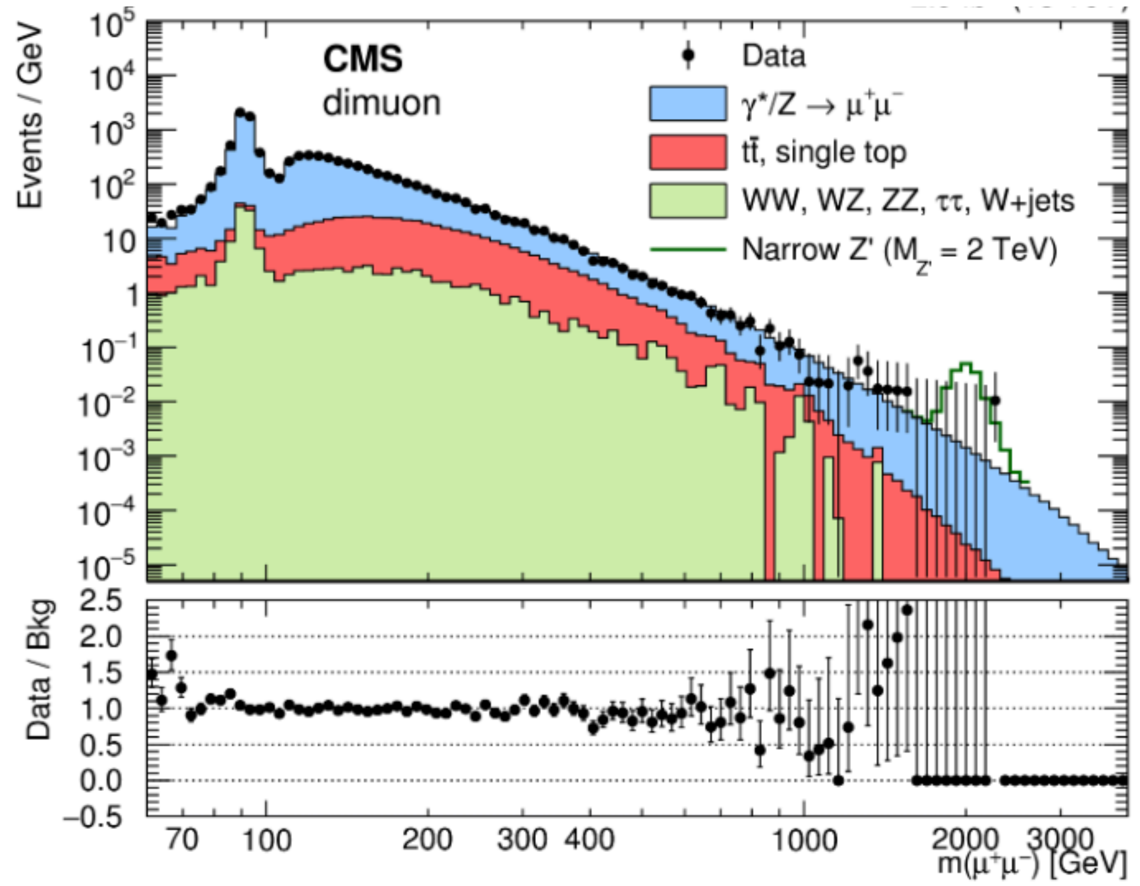
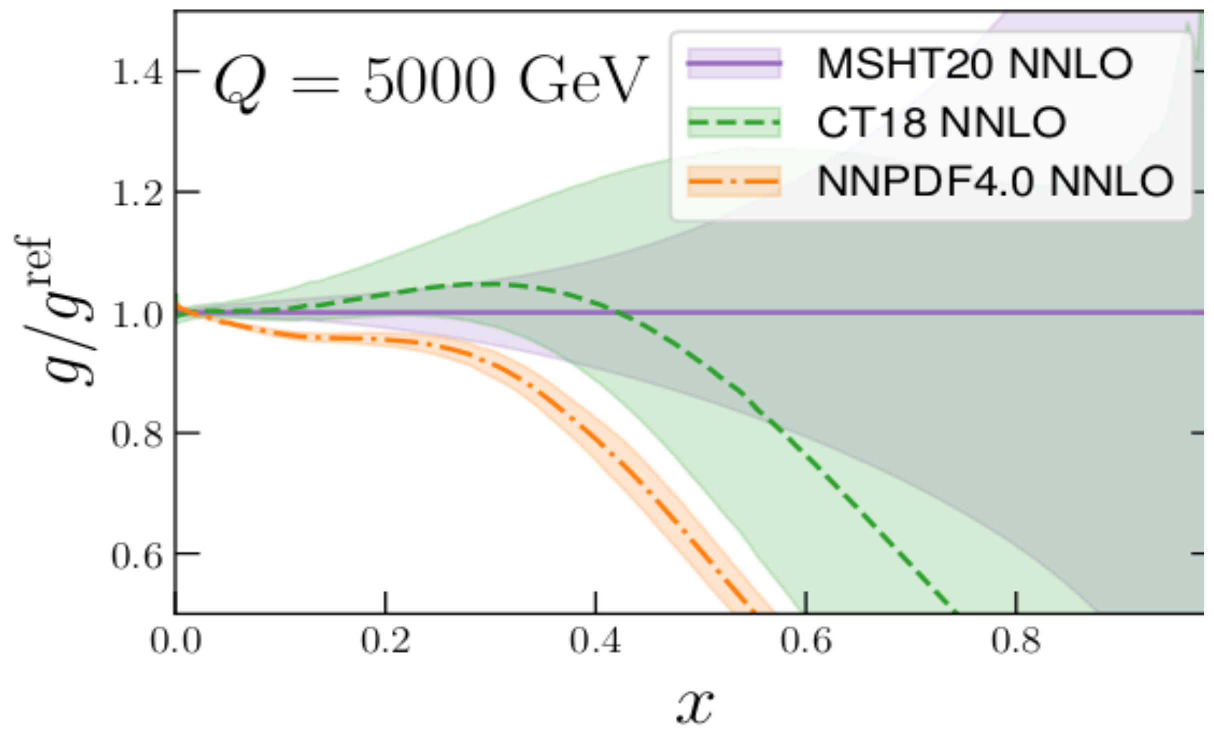


Image Credit: Tom Cridge



Global PDF Fits

- Basic idea is simple: $\text{Data} = \text{PDF} \otimes \sigma_H$ but many ingredients enter! Three key areas:

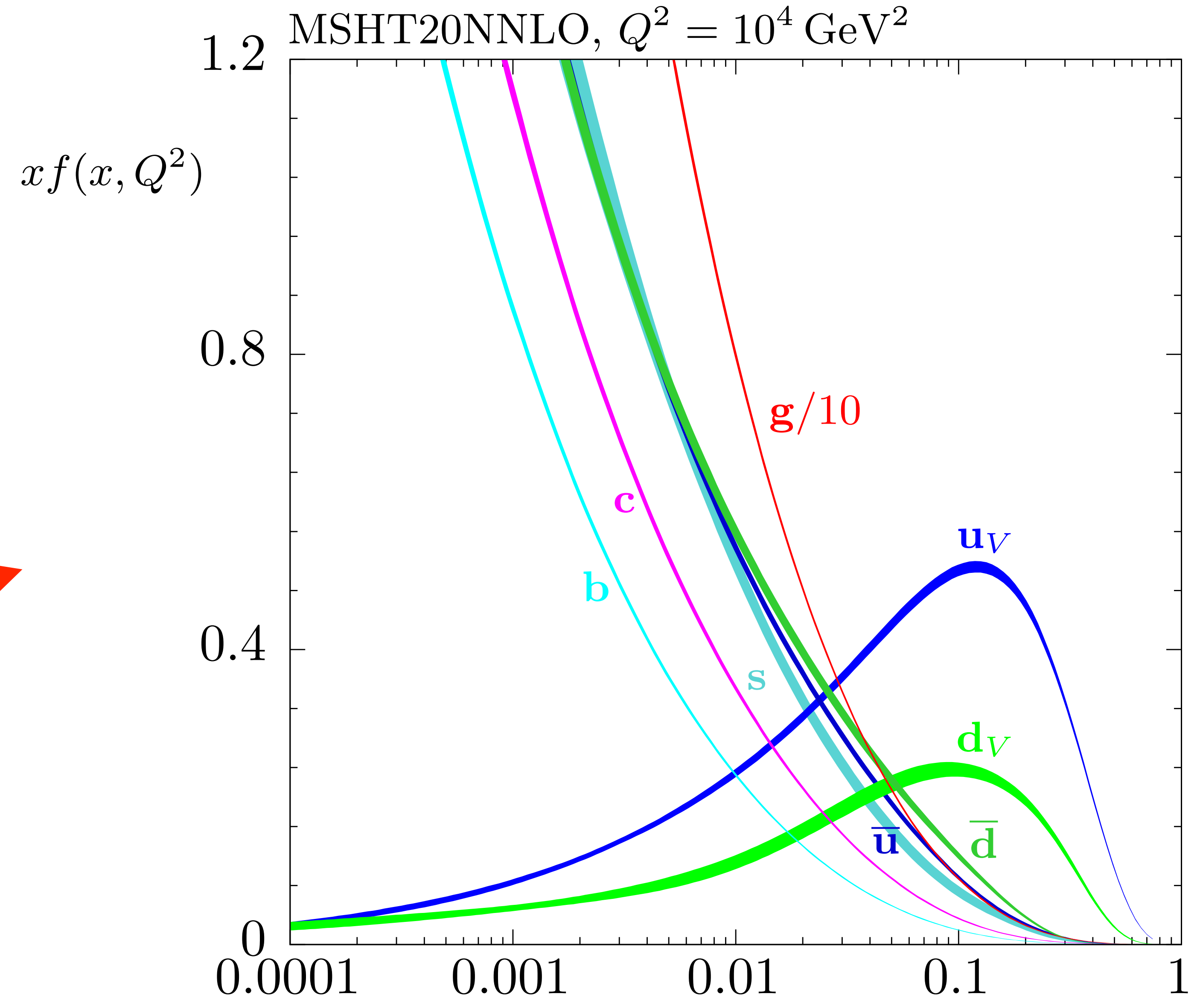
measure → Data; fit → PDF; predict → σ_H

Methodology

- PDF parameterisation, uncertainty prescription…

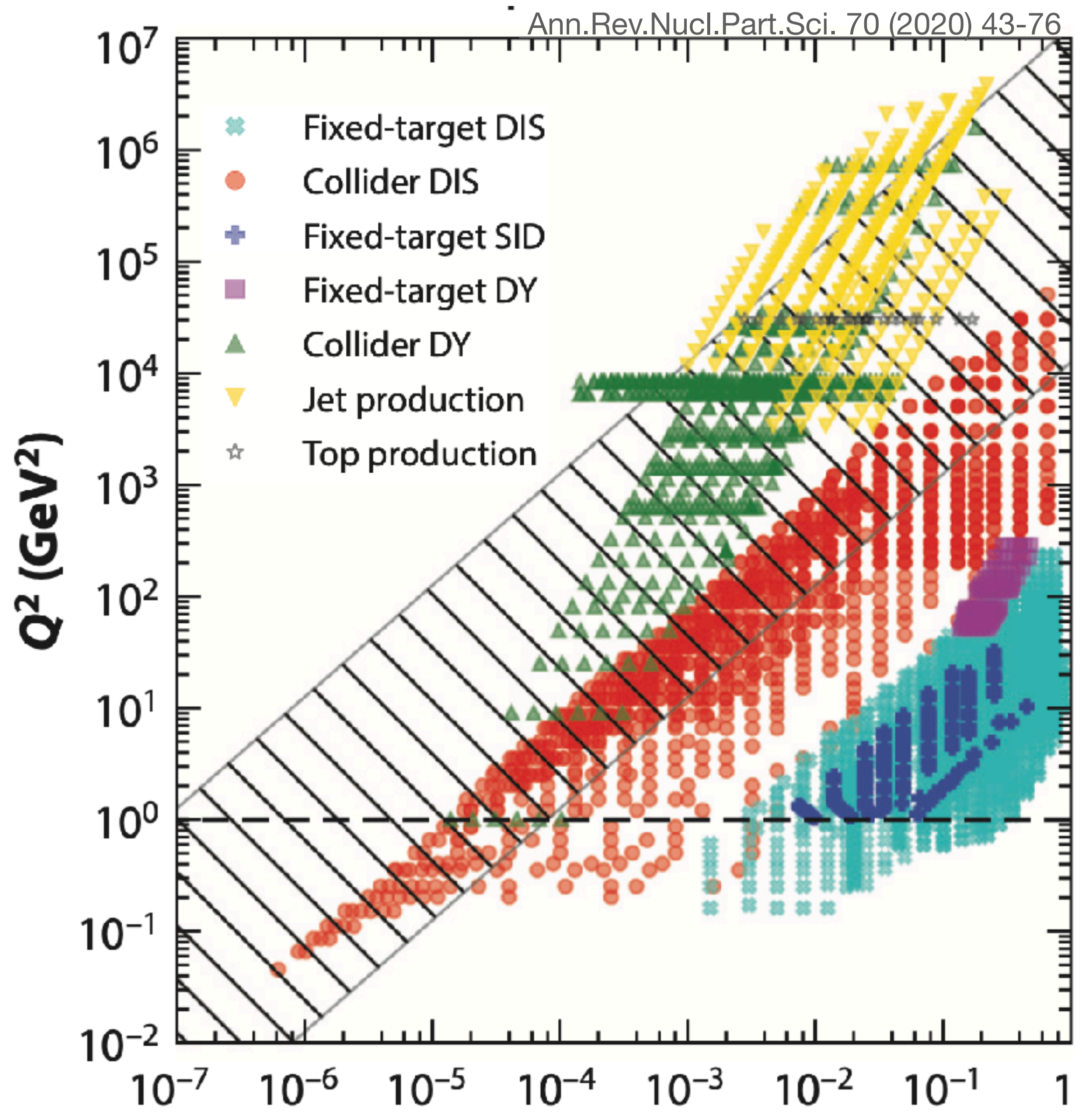
Theory

- High precision: NNLO QCD + NLO EW the standard



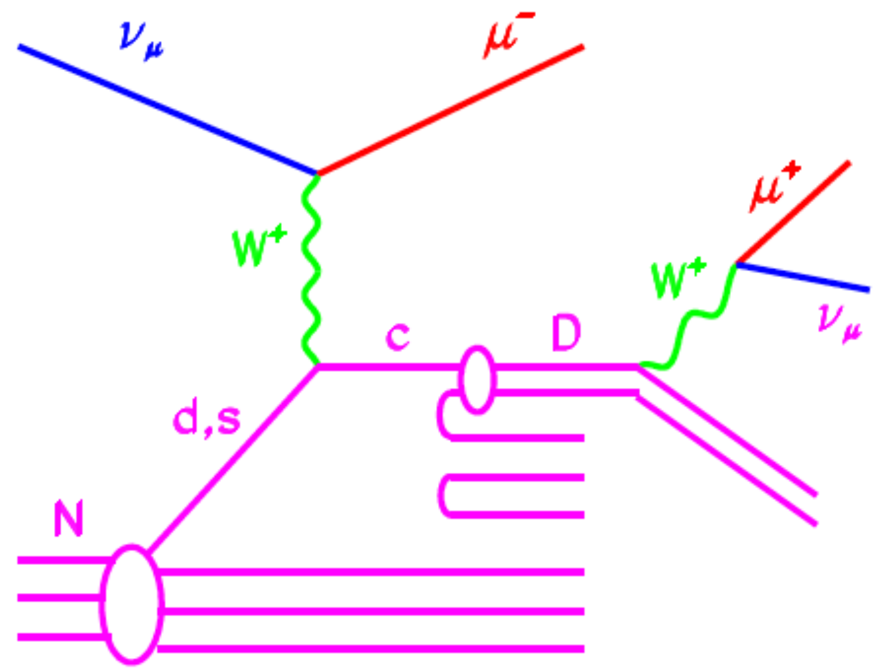
Data

- From fixed target, to HERA DIS and collider. LHC data increasingly important.

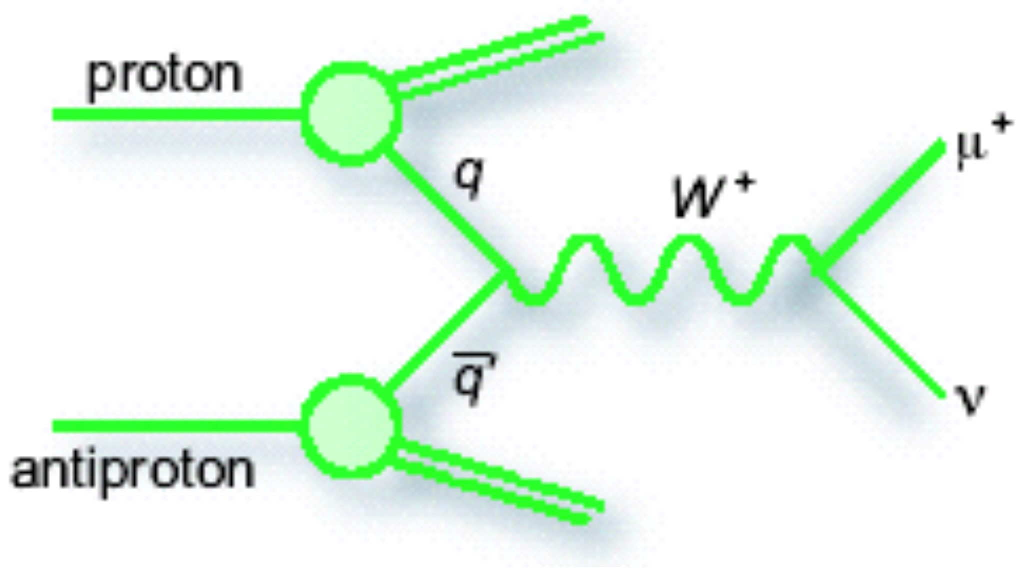


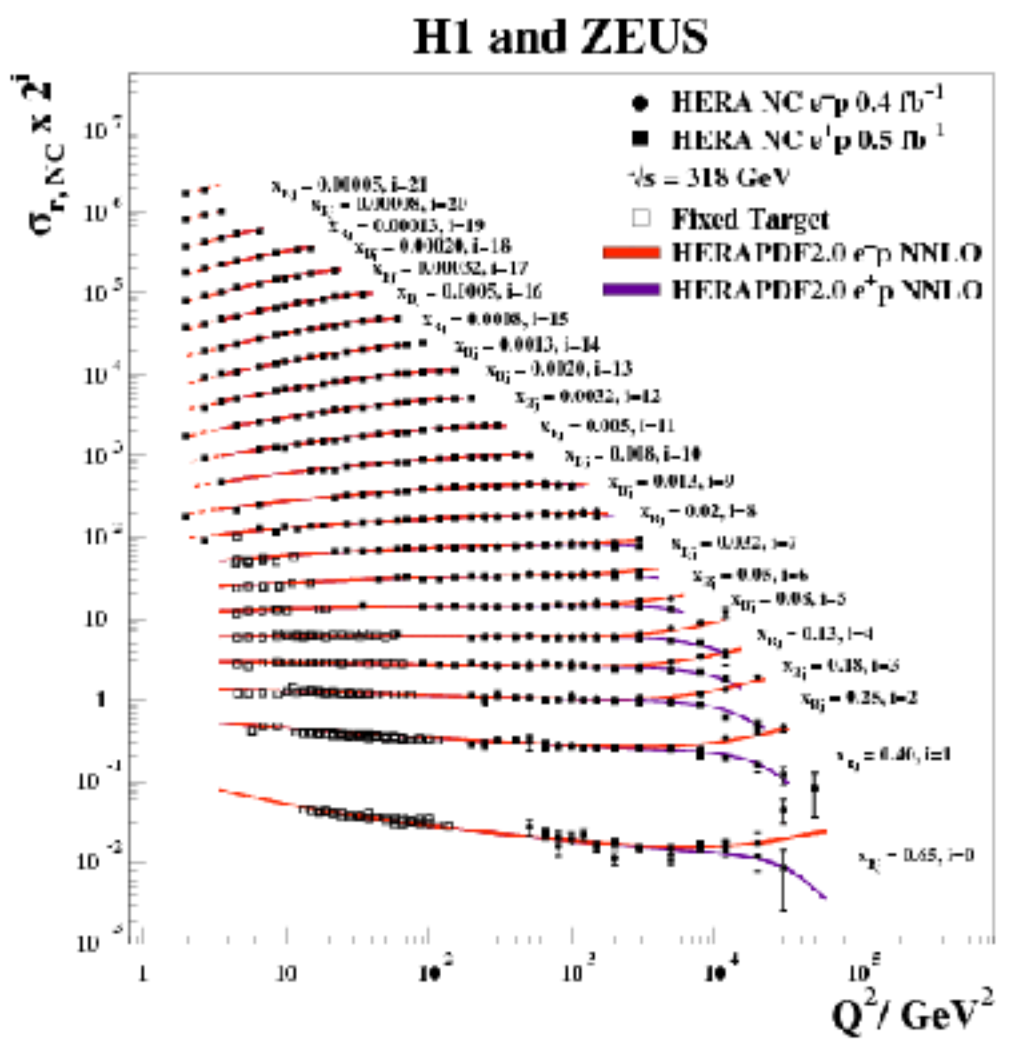
- Aim: high precison theory + wide range of data $\longrightarrow$ **precise** + **accurate** PDFs
- Alternative/complementary route: input from lattice.

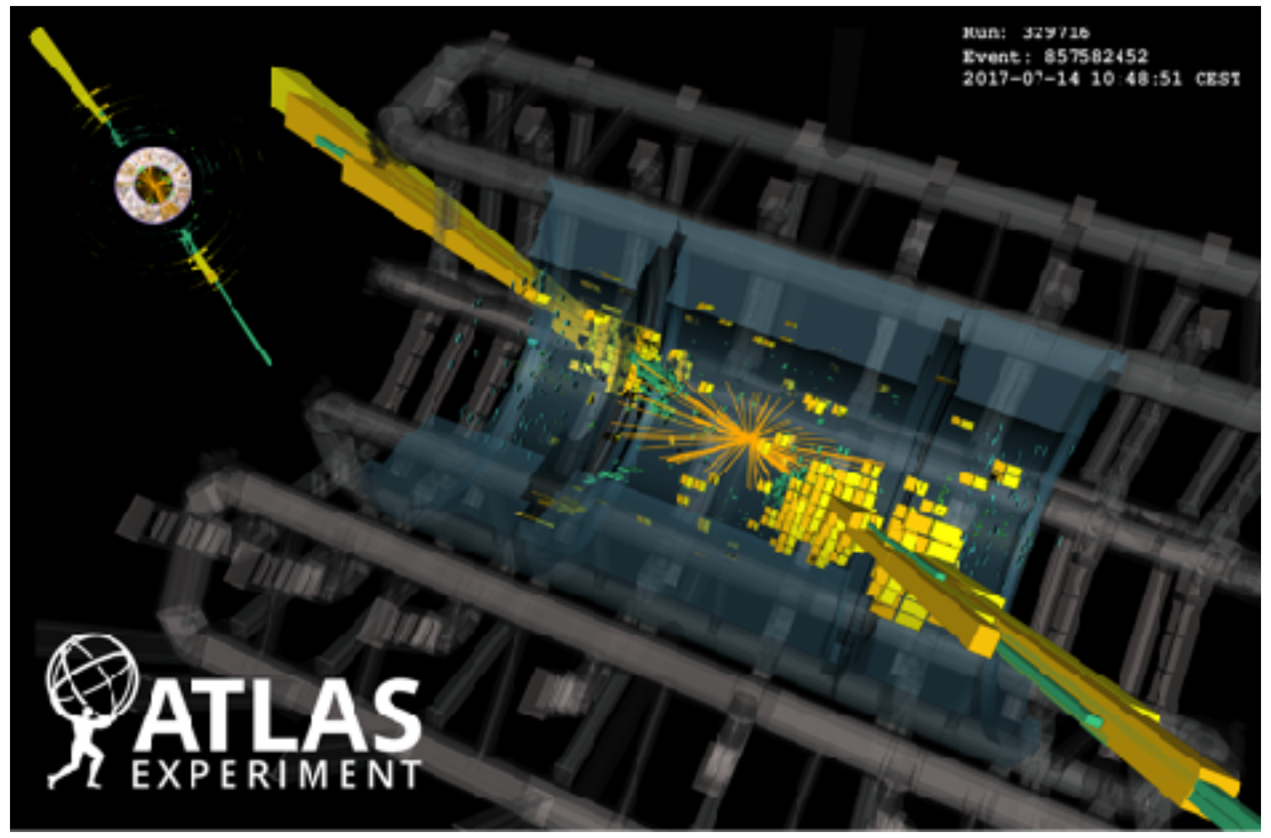
Global PDF fits: datasets



	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm\{p,n\} \to \ell^\pm + X$	$\gamma^* q \to q$	$q,\bar{q},g$	$x \gtrsim 0.01$
	$\ell^\pm\, n/p \to \ell^\pm + X$	$\gamma^*\, d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu})\,N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q,\bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu\,N \to \mu^-\mu^+ + X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu}\,N \to \mu^+\mu^- + X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
Collider DIS	$e^\pm p \to e^\pm + X$	$\gamma^* q \to q$	$g,q,\bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
	$e^+ p \to \bar{\nu} + X$	$W^+\{d,s\} \to \{u,c\}$	d,s	$x \gtrsim 0.01$
	$e^\pm p \to e^\pm\, c\bar{c} + X$	$\gamma^* c \to c,\ \gamma^* g \to c\bar{c}$	c,g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to e^\pm\, b\bar{b} + X$	$\gamma^* b \to b,\ \gamma^* g \to b\bar{b}$	b,g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
Tevatron	$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g,q	$0.01 \lesssim x \lesssim 0.5$
	$p\bar{p} \to (W^\pm \to \ell^\pm\nu) + X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u,d,\bar{u},\bar{d}$	$x \gtrsim 0.05$
	$p\bar{p} \to (Z \to \ell^+\ell^-) + X$	$uu, dd \to Z$	u,d	$x \gtrsim 0.05$
	$p\bar{p} \to t\bar{t} + X$	$qq \to t\bar{t}$	q	$x \gtrsim 0.1$
LHC	$pp \to \text{jet} + X$	$gg, qg, q\bar{q} \to 2j$	g,q	$0.001 \lesssim x \lesssim 0.5$
	$pp \to (W^\pm \to \ell^\pm\nu) + X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u,d,\bar{u},\bar{d},g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X$	$q\bar{q} \to Z$	$q,\bar{q},g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X,\ p_\perp$	$gq(\bar{q}) \to Zq(\bar{q})$	$g,q,\bar{q}$	$x \gtrsim 0.01$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, Low mass	$q\bar{q} \to \gamma^*$	$q,\bar{q},g$	$x \gtrsim 10^{-4}$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, High mass	$q\bar{q} \to \gamma^*$	$\bar{q}$	$x \gtrsim 0.1$
	$pp \to W^+\bar{c}, W^-c$	$sg \to W^+c,\ \bar{s}g \to W^-\bar{c}$	$s,\bar{s}$	$x \sim 0.01$
	$pp \to t\bar{t} + X$	$gg \to t\bar{t}$	g	$x \gtrsim 0.01$
	$pp \to D, B + X$	$gg \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to J/\psi, \Upsilon + pp$	$\gamma^*(gg) \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to \gamma + X$	$gq(\bar{q}) \to \gamma q(\bar{q})$	g	$x \gtrsim 0.005$





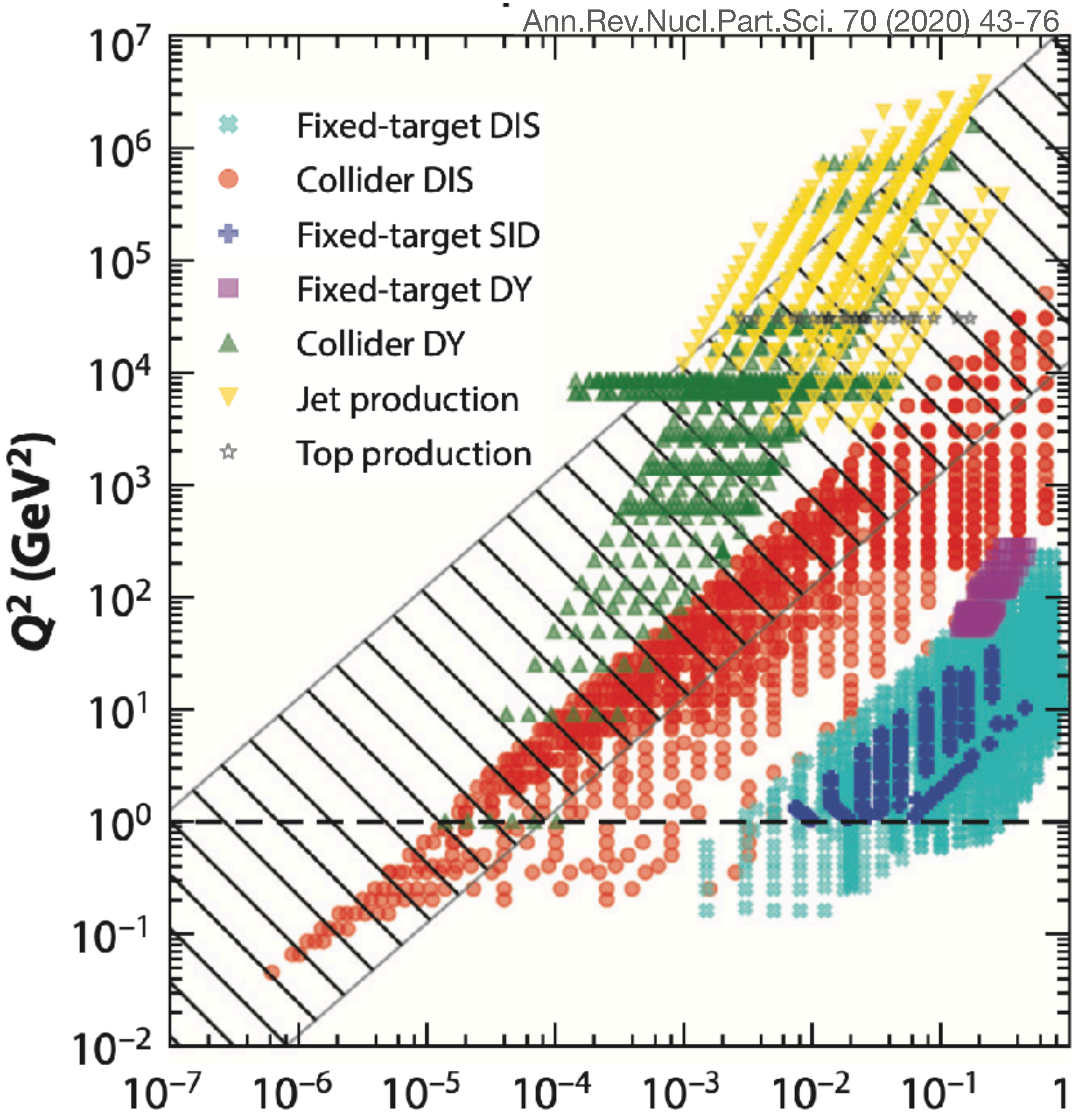


- From fixed target, to HERA DIS and collider. LHC data increasingly important.

$N_{\text{dataset}} \sim 50 - 60$ $\quad N_{\text{pts}} \sim 4000 - 5000$

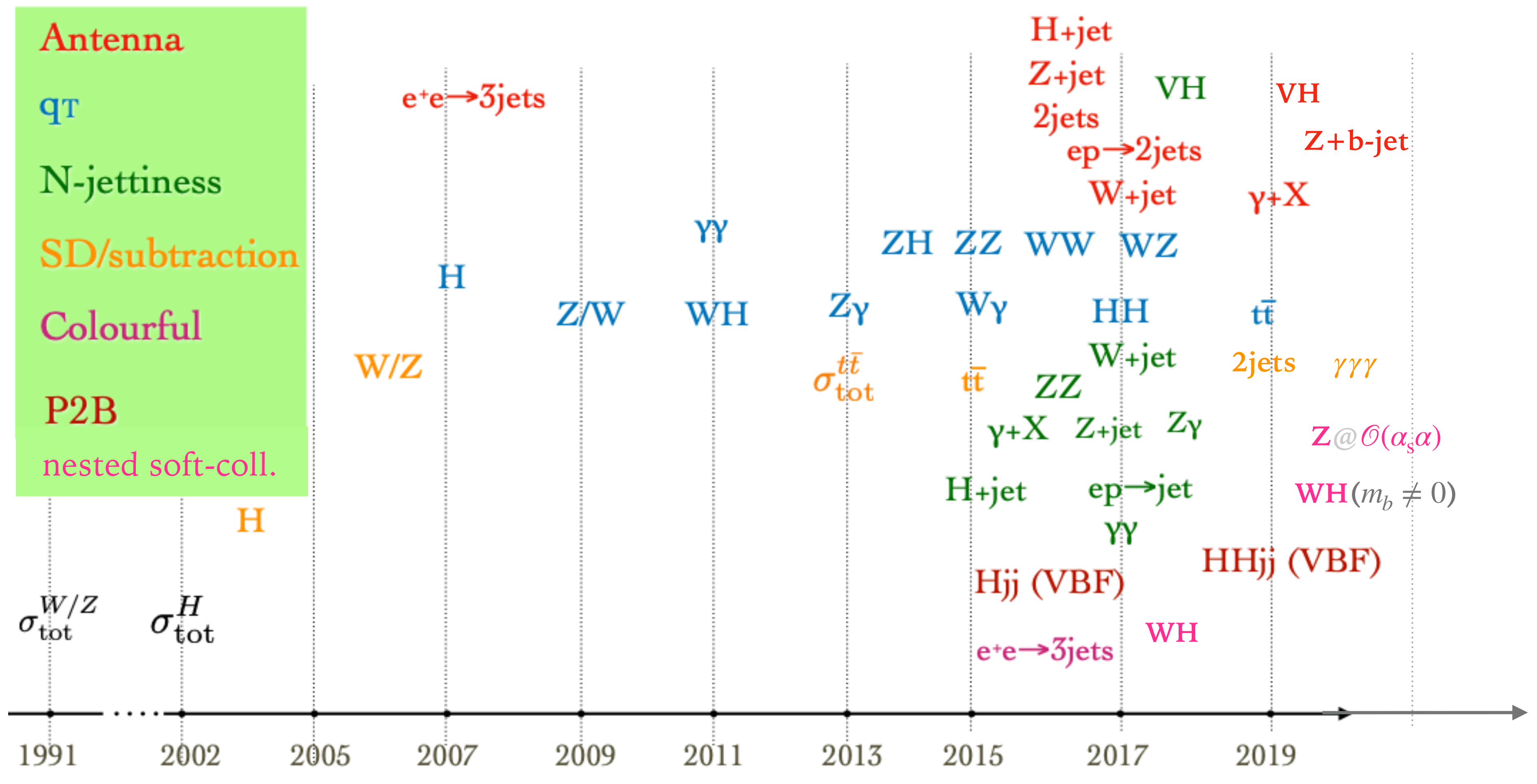
Kinematic Coverage

- Global fits achieve **broad coverage** from low to high Q^2, and over many orders of magnitude in x .



Precision Theory

- Has been significant progress in perturbative calculations: **NNLO QCD + NLO EW** now long been the standard.



[based on slide by M. Grazzini; QCD@LHC 2019]

MSHT20

Data set	$N_{\rm pts}$	NLO $\chi^2/N_{\rm pts}$	NNLO $\chi^2/N_{\rm pts}$
ATLAS 8 TeV s. diff $t\bar{t}$	25	1.56	0.98
CMS 8 TeV d. diff $t\bar{t}$	15	2.19	1.50
ATLAS 7 TeV W, Z	61	5.00	1.91
ATLAS 8 TeV W	22	3.85	2.61
ATLAS 8 TeV d. diff Z	59	2.67	1.45
ATLAS 8 TeV Z p_T	104	2.26	1.81
ATLAS 8 TeV W + jets	39	1.13	0.60
Total LHC data	1328	1.79	1.33
Total non-LHC data	3035	1.13	1.10
Total	4363	1.33	1.17

- Not simply question of adding in a bit more precision. E.g. **NNLO QCD** is essential to give good description to global data set (LHC in particular).

Image Credit: Tom Cridge

Methodology

- Two distinct methodologies on the market to parameterising PDFs: **Neural Nets** (NNPDF) or **Explicit Parameterisation** (CT, MSHT).

$$f_i(x,Q_0):\ A_f x^{a_f}(1-x)^{b_f}\times \longrightarrow \sum_{i=1}^{n}\alpha_{f,i}P_i(y(x))\,,\quad \textbf{CT, MSHT...}$$

$$\searrow\ \mathrm{NN}_i(x)\quad \textbf{NNPDF}$$

  - ✦ MSHT: **52** free parameters in terms of Chebyshev polynomials.

$$u_V(x,Q_0^2)=A_u(1-x)^{\eta_u}x^{\delta_u}\left(1+\sum_{i=1}^{6}a_{u,i}T_i(y(x))\right)$$

$$d_V(x,Q_0^2)=A_d(1-x)^{\eta_d}x^{\delta_d}\left(1+\sum_{i=1}^{6}a_{d,i}T_i(y(x))\right)$$

$$S(x,Q_0^2)=A_S(1-x)^{\eta_S}x^{\delta_S}\left(1+\sum_{i=1}^{6}a_{S,i}T_i(y(x))\right)$$

$$s_+(x,Q_0^2)=A_{s_+}(1-x)^{\eta_{s+}}x^{\delta_S}\left(1+\sum_{i=1}^{6}a_{s_+,i}T_i(y(x))\right)$$

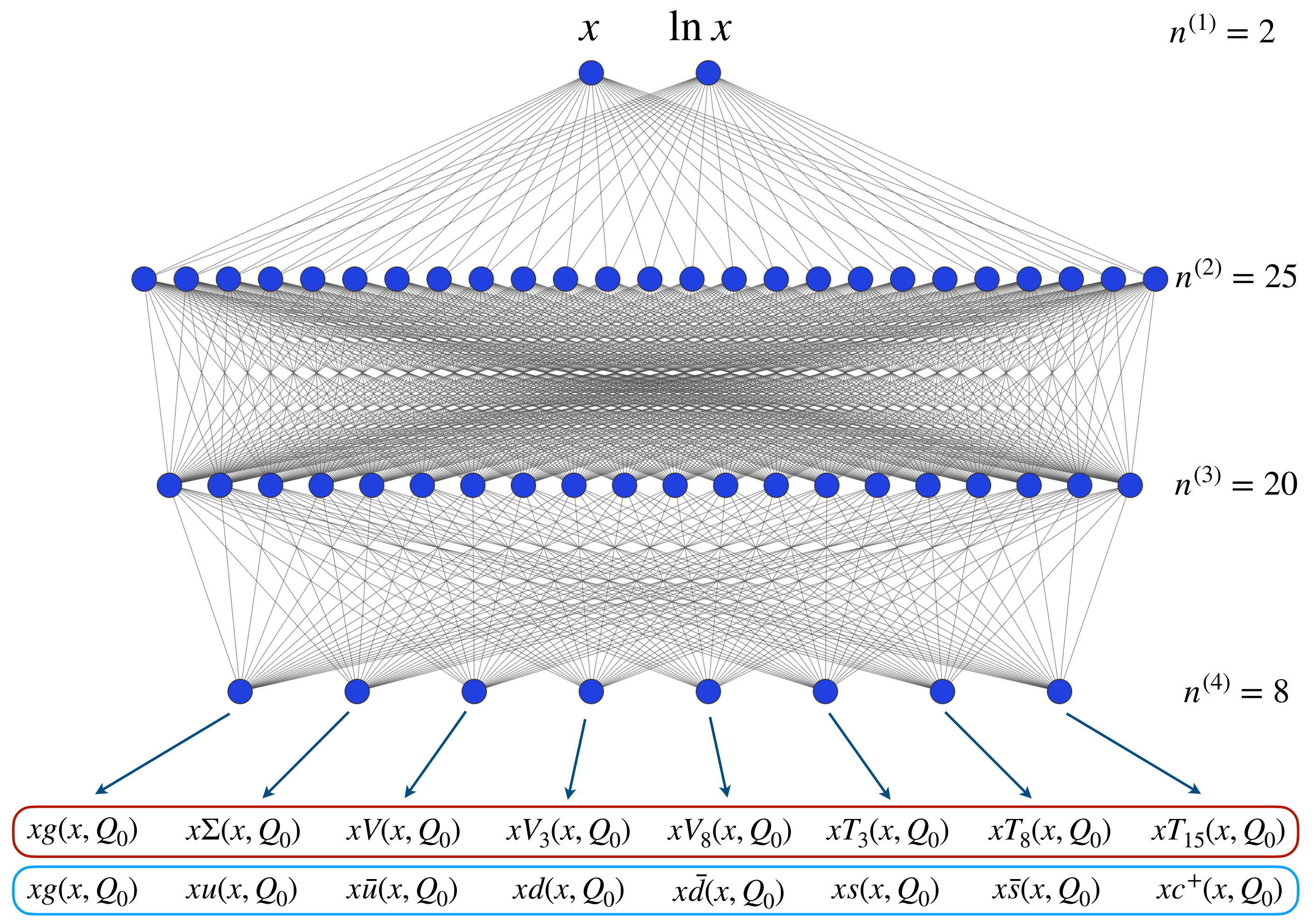
$$g(x,Q_0^2)=A_g(1-x)^{\eta_g}x^{\delta_g}\left(1+\sum_{i=1}^{4}a_{g,i}T_i(y(x))\right)+A_{g-}(1-x)^{\eta_{g-}}x^{\delta_{g-}}$$

$$s_-(x,Q_0^2)=A_{s_-}(1-x)^{\eta_{s-}}(1-x/x_0)x^{\delta_{s-}}$$

$$(\bar{d}/\bar{u})(x,Q_0^2)=A_\rho(1-x)^{\eta_\rho}\left(1+\sum_{i=1}^{6}a_{\rho,i}T_i(y(x))\right)$$

  - ✦ Less flexible in general - need to be sure flexible enough! Allows direct handle on uncertainties in Hessian framework.

- ★ NNPDF: **763** free parameter Neural Net.



- ★ Increased flexibility, but needs robust optimisation + stopping (avoid over and under fitting).

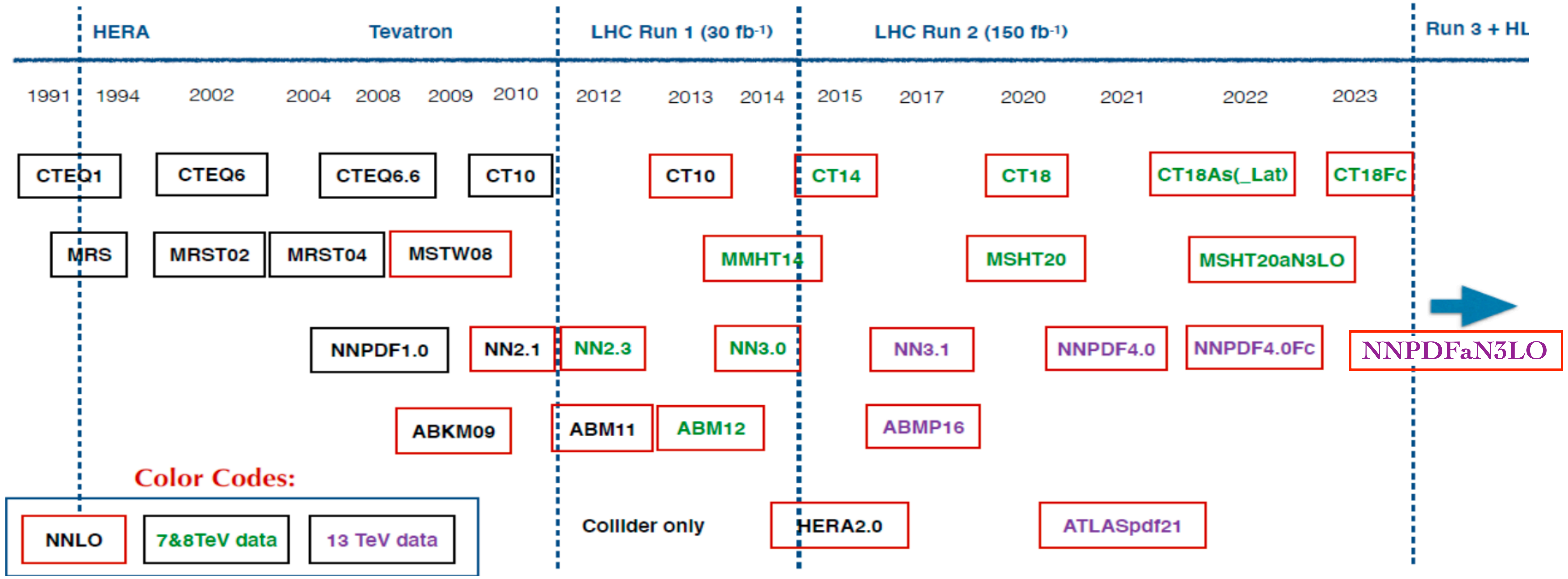
Major PDF Analyses

- Multiple PDF analyses, with different methodogies and datasets. Cannot cover these all here!
- Major releases from 3 global fitters (CT, MSHT, NNPDF) ~ 2 or more years ago. But they have been busy:

★ Major push to approximate N3LO + theoretical uncertainties

★ QED/EW corrections standard

★ Many dedicated studies



- These advances all build towards next generation of releases.

Image Credit:
Jun Gao

New Data

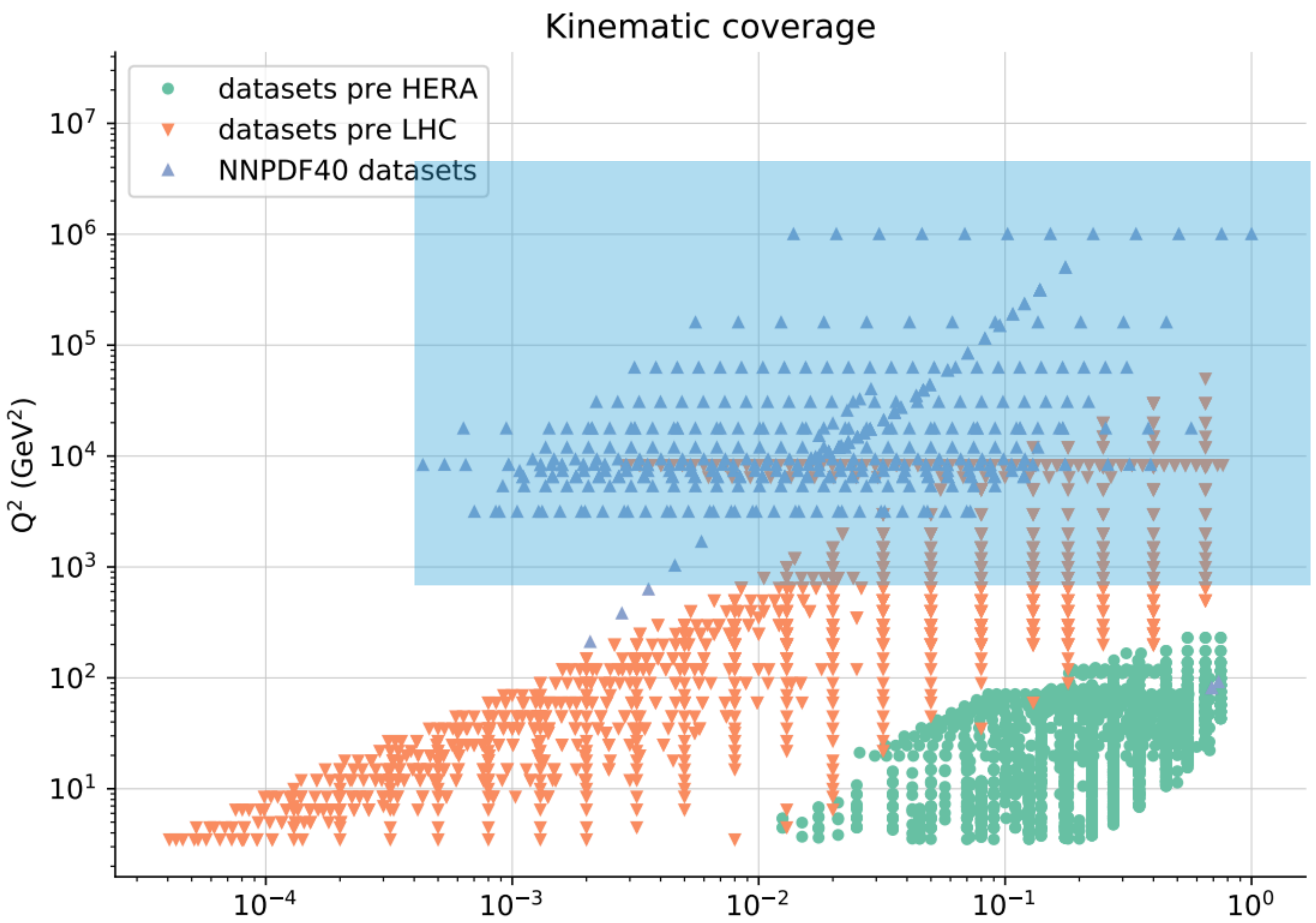
LHC data!

Maria Ubiali, ep/ea synergy workshop, CERN Feb 24

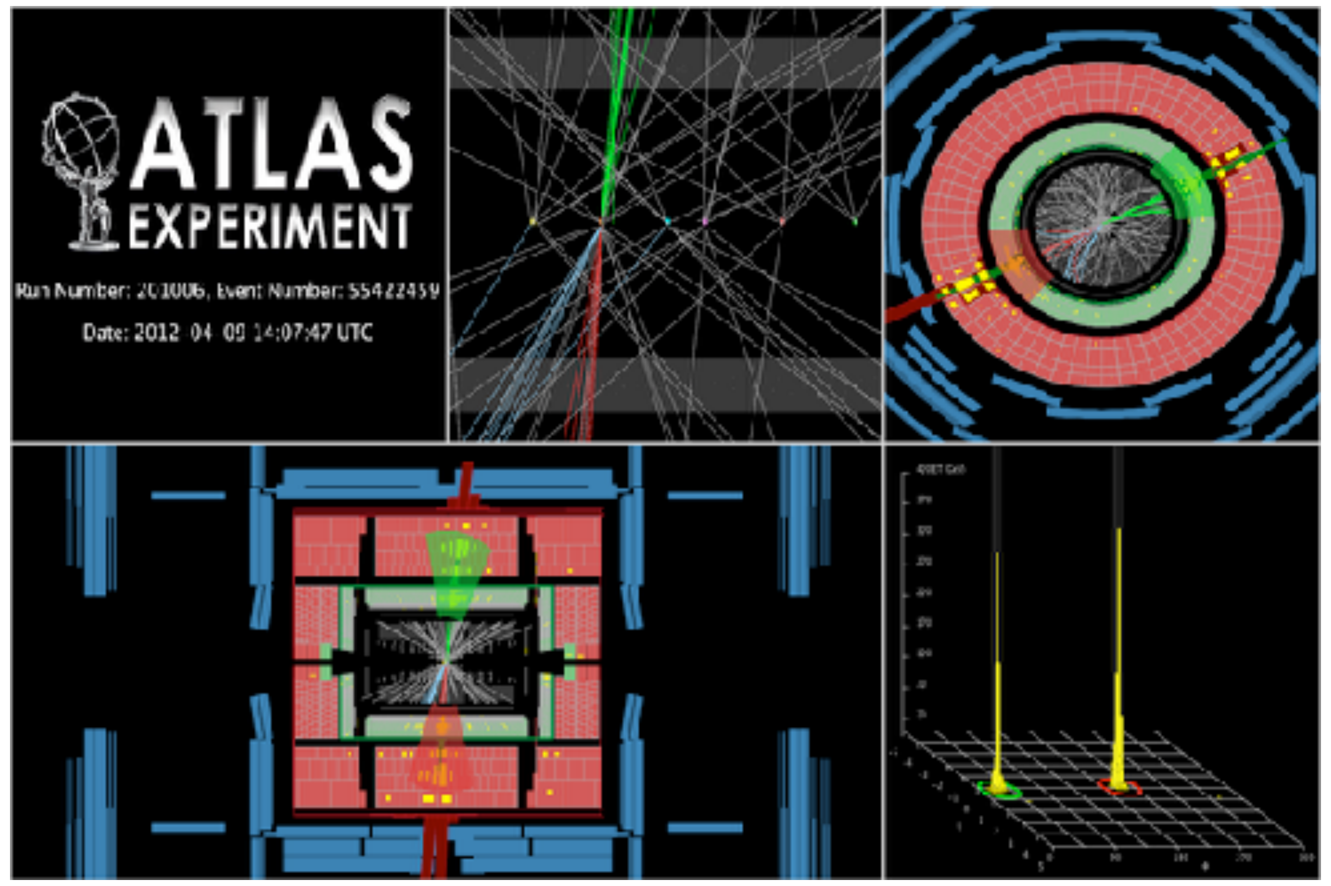
- A wealth of data from the **LHC**, playing a significant role in PDF fits.

- Two key categories:

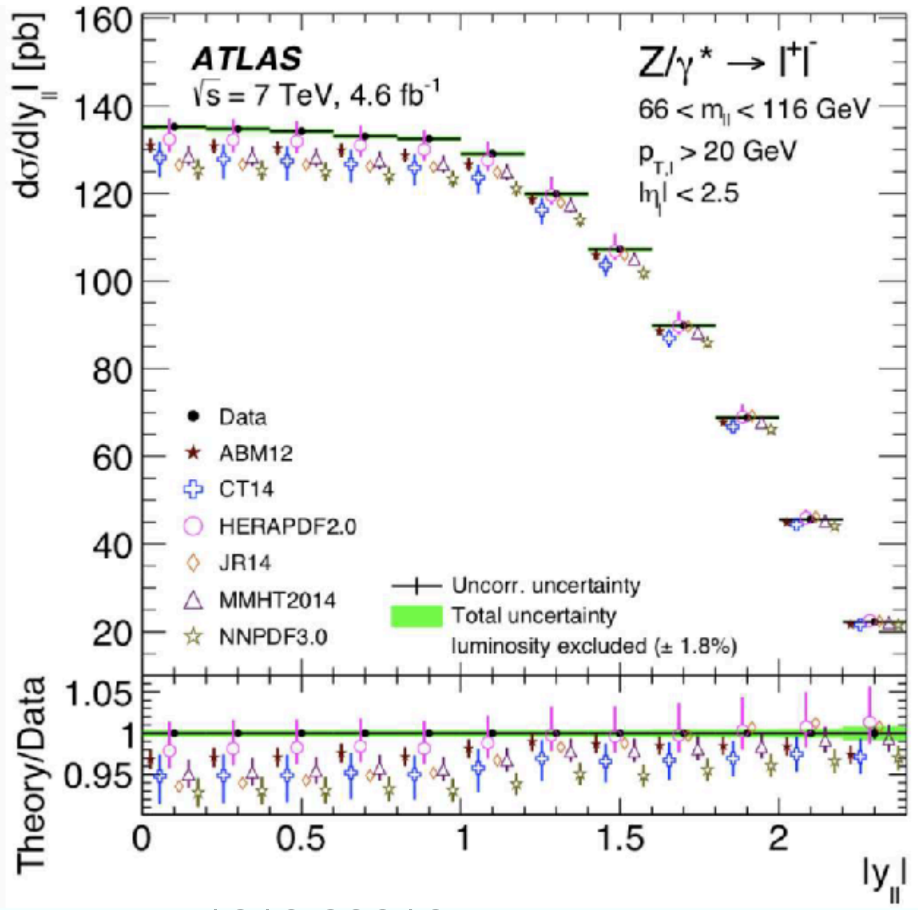
  ★ **High energy** data - probing the high x region. Jets, top quarks, vector boson $p_\perp$…



NNPDF4.0: About 30% of input data are LHC data!



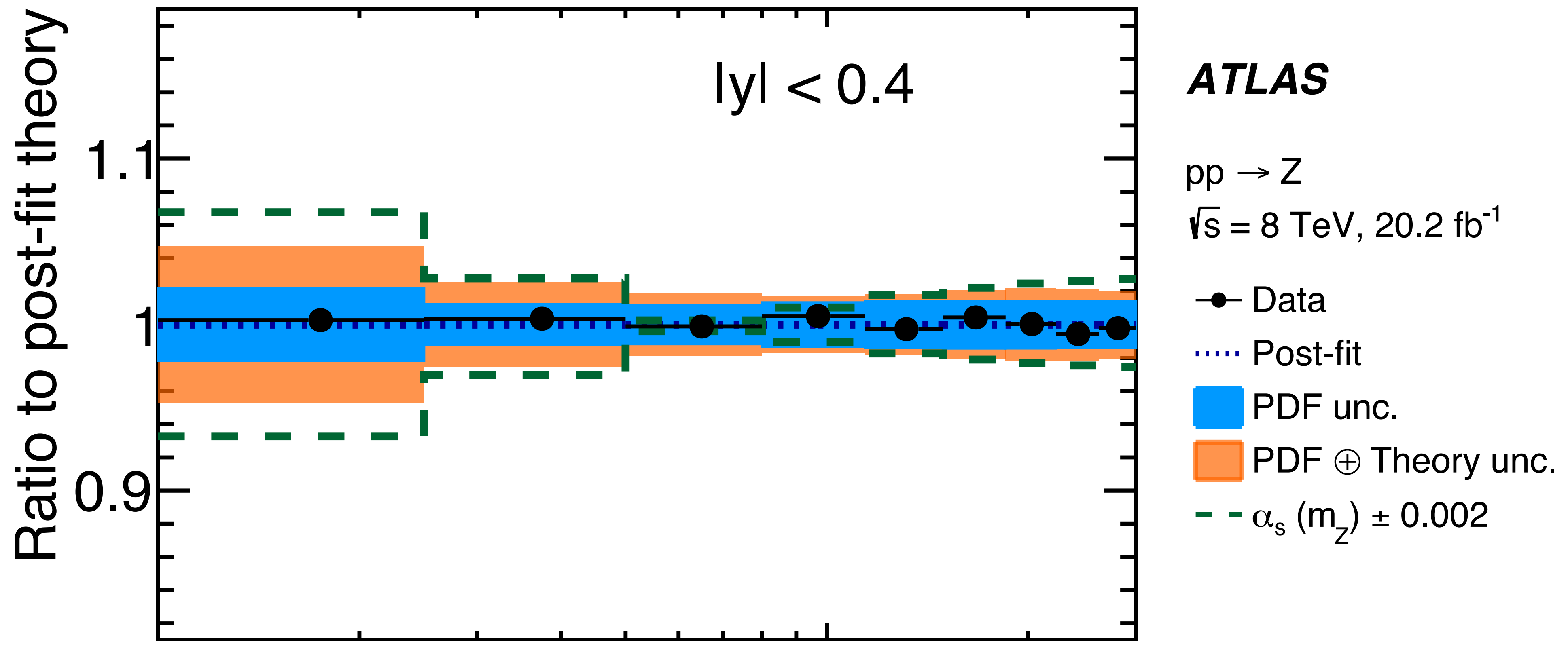
★ **High precision** data - Drell Yan and flavour structure



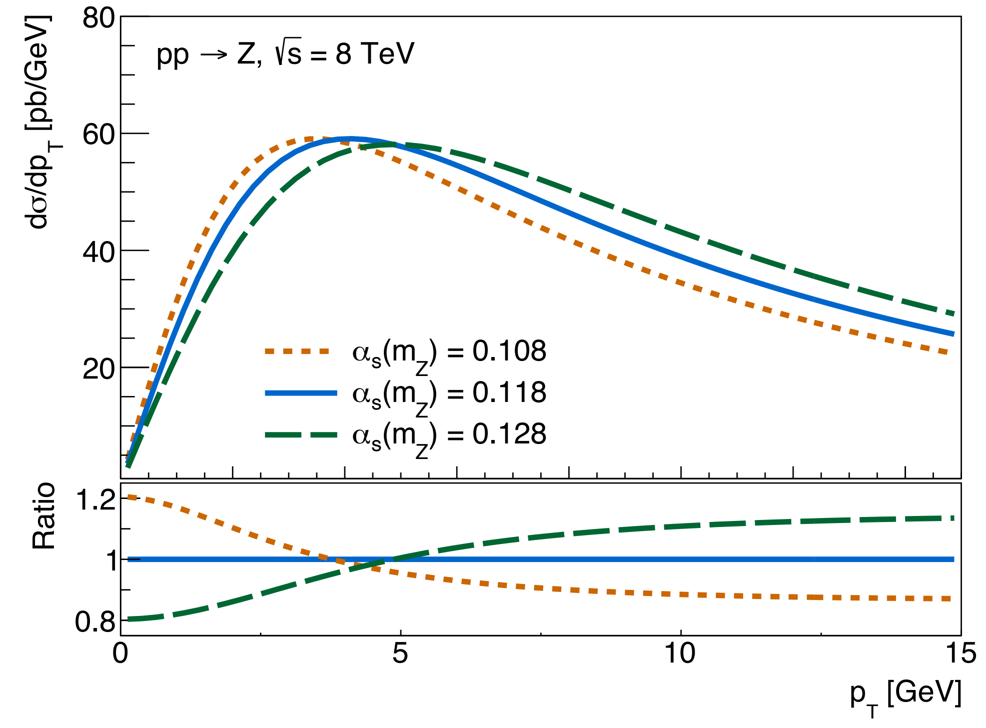
1612.03016

LHC data - Some Examples

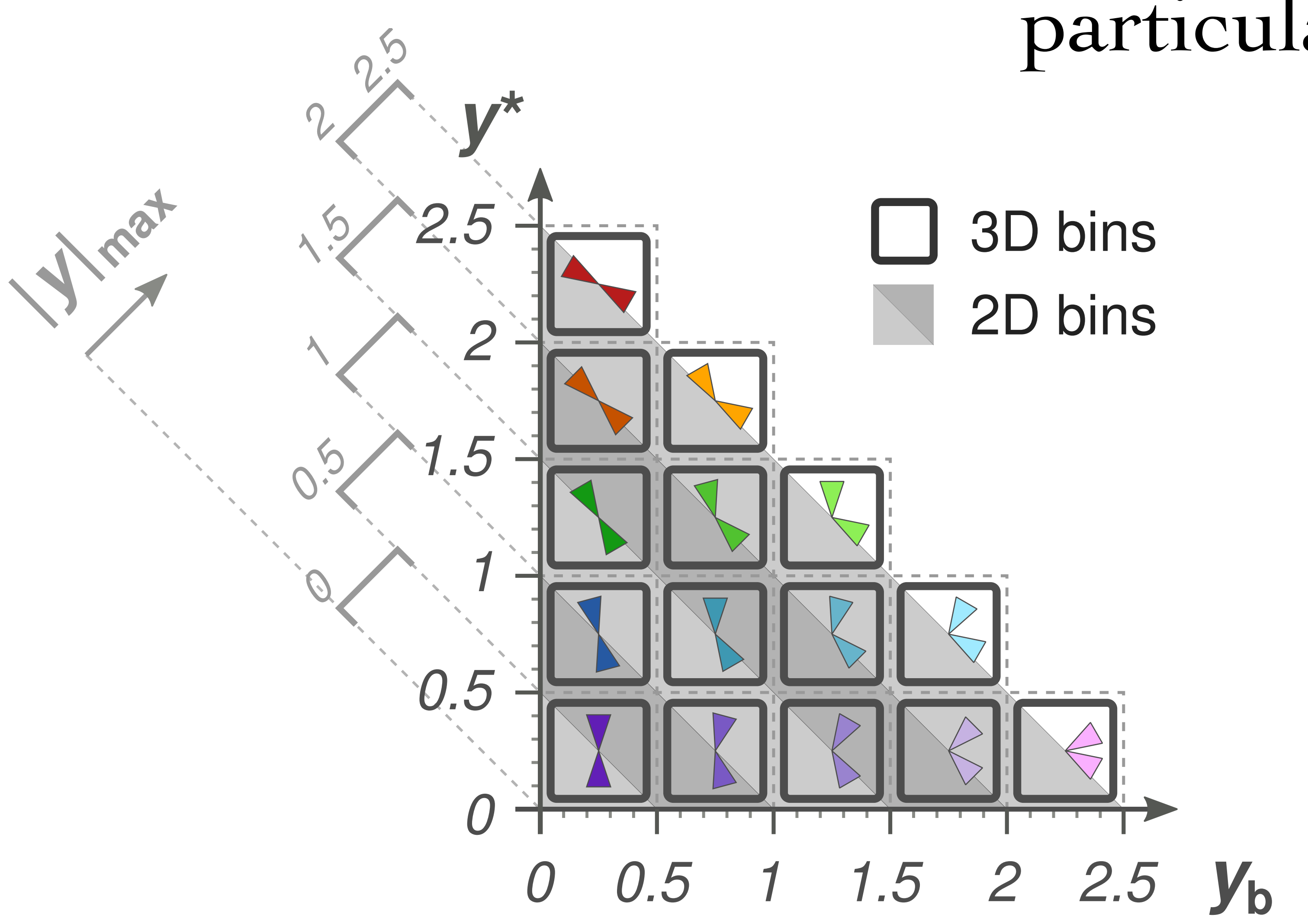
ATLAS, arXiv:2309.12986

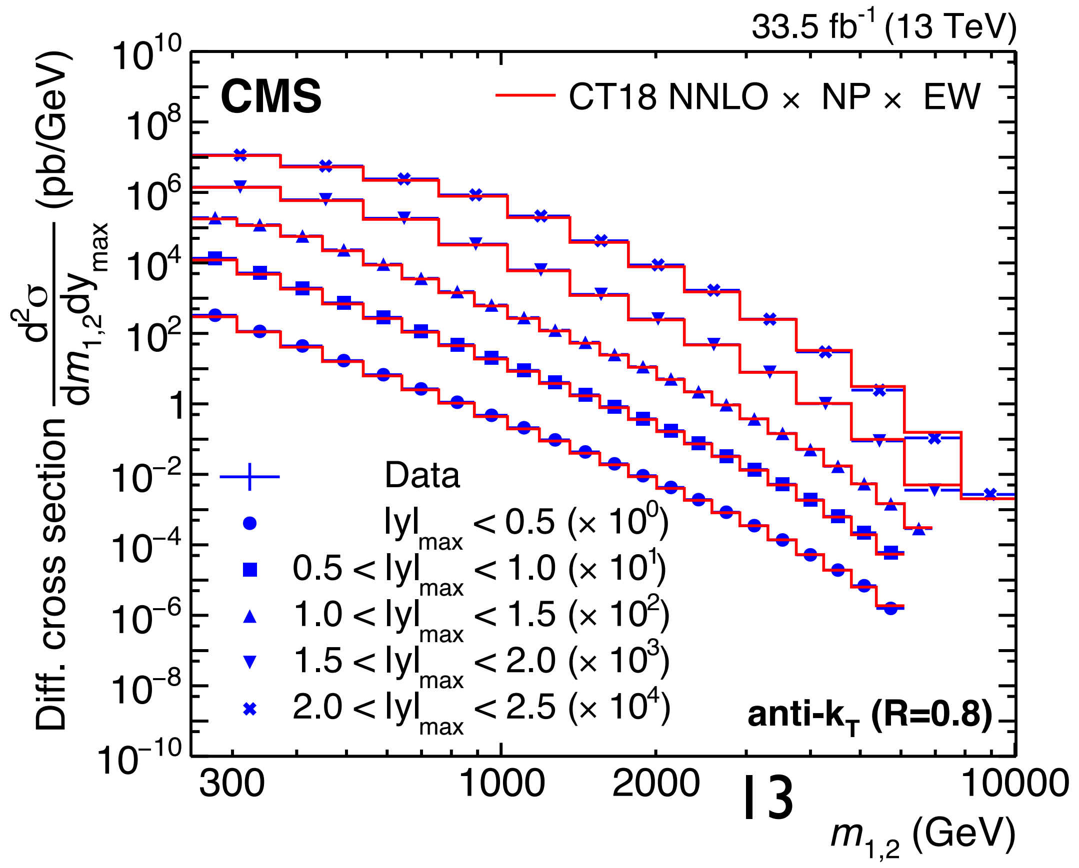


- ATLAS first extraction of strong coupling from Sudakov peak of $Z\,p_\perp$ distribution, at up to aN3LO.
- Also displays PDF sensitivity to e.g. gluon at intermediate x .



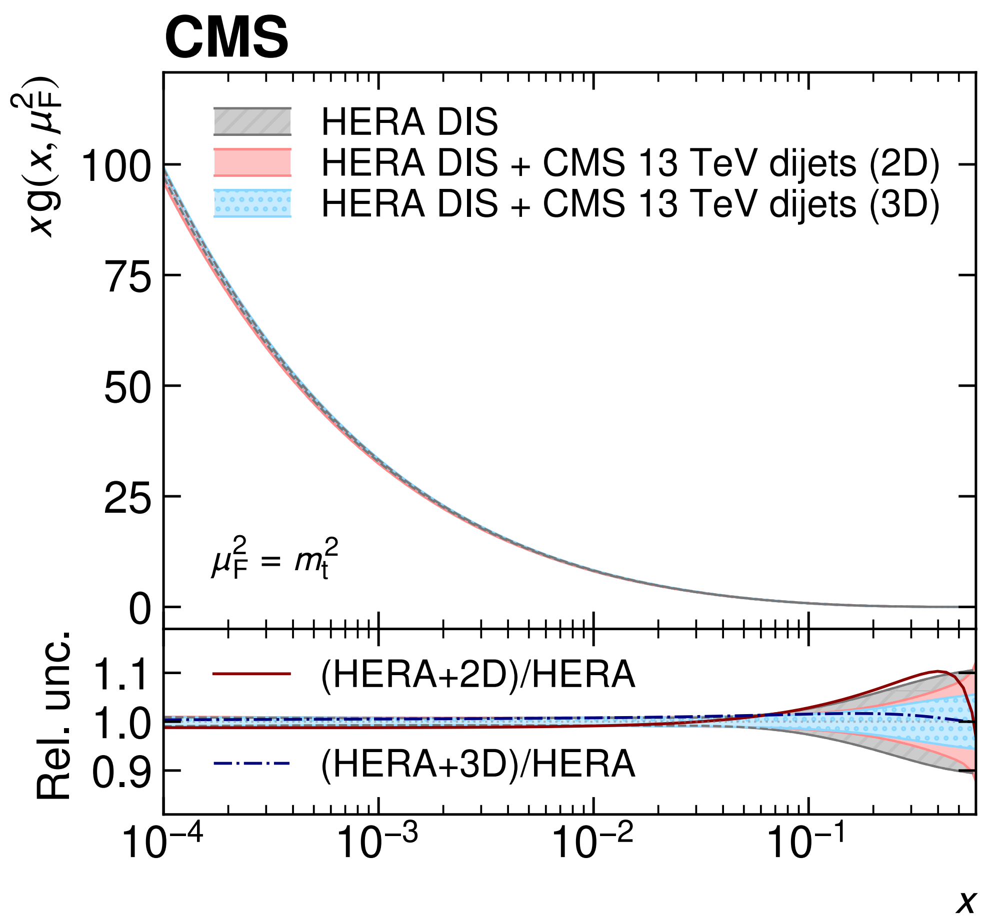
- Extensive range of jet measurements: triple differential dijet production particularly promising. More differential = more constraining!





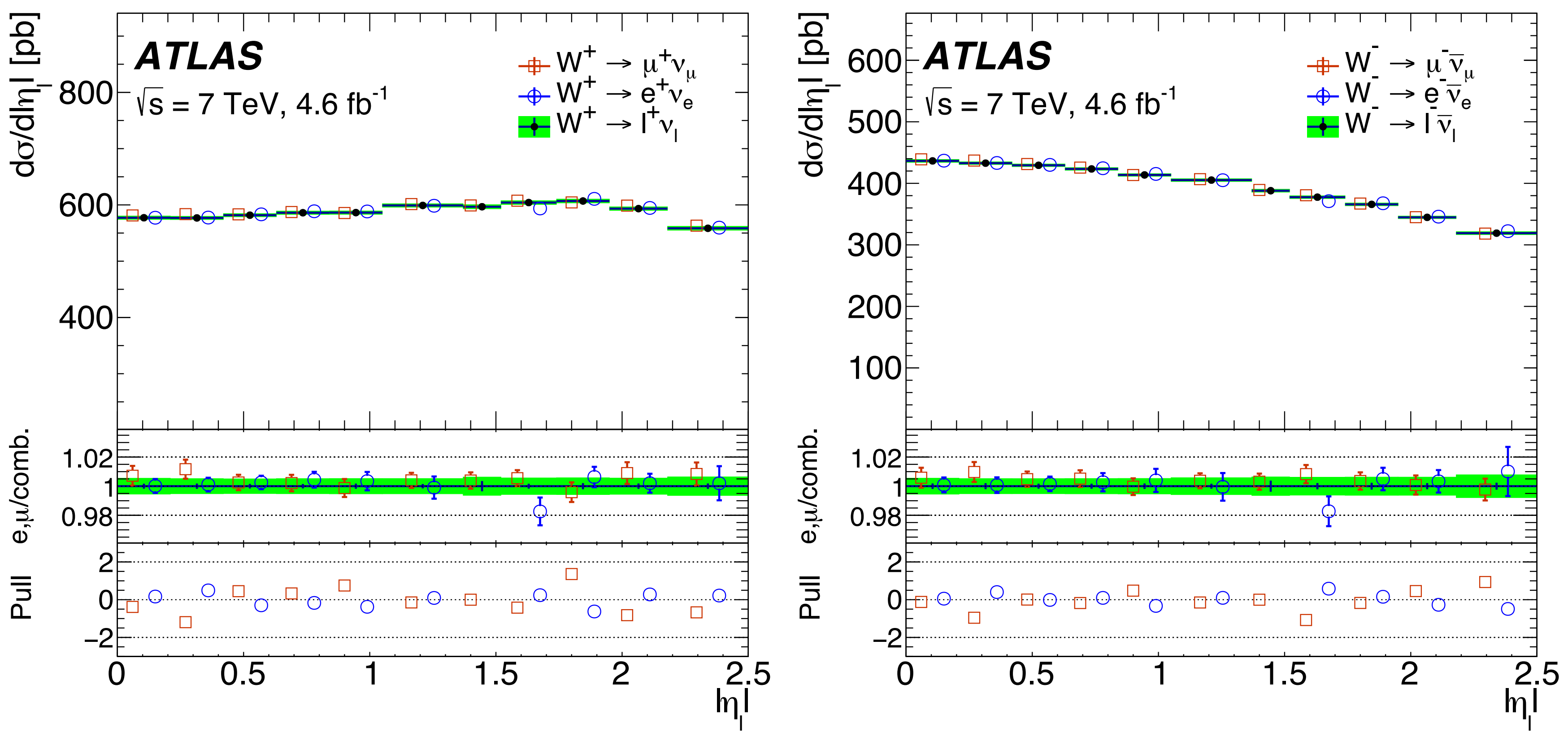
- Sensitivity to e.g. gluon at high x .

CMS, arXiv:2312.16669

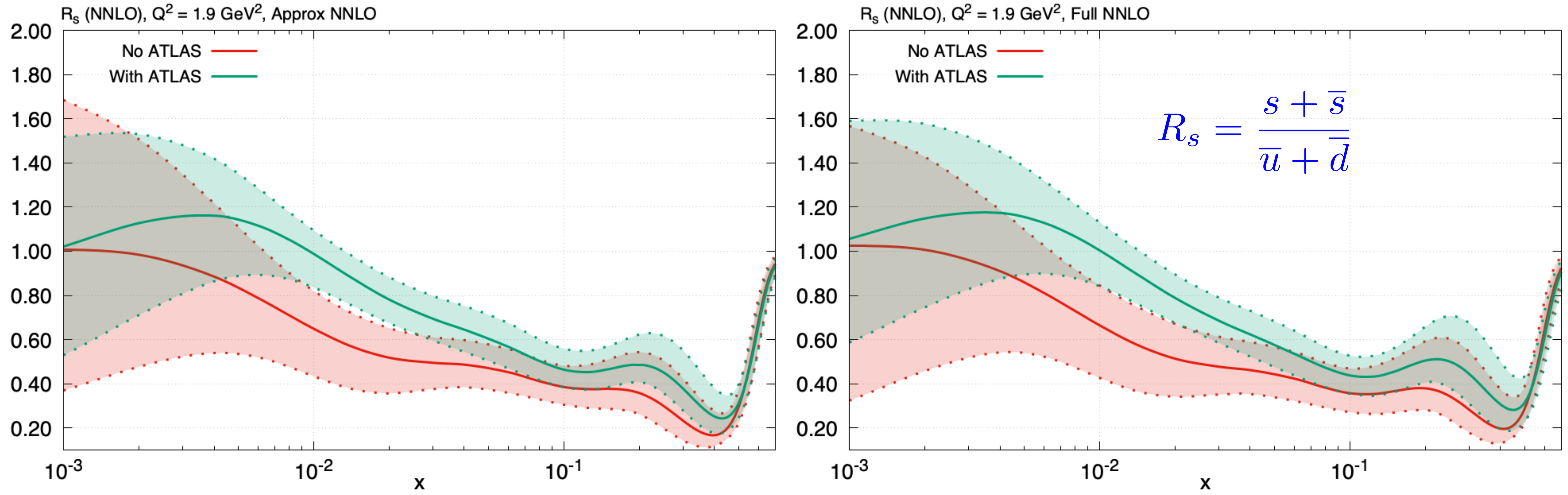


- Extensive range of multi-differential Drell Yan measurements.

  - ATLAS high precision DY data even from Run I had significant impact on PDF fits, notably strangeness.



ATLAS, Eur. Phys. J C77 (2017) 367



$$R_s = \frac{s+\bar{s}}{\bar{u}+\bar{d}}$$

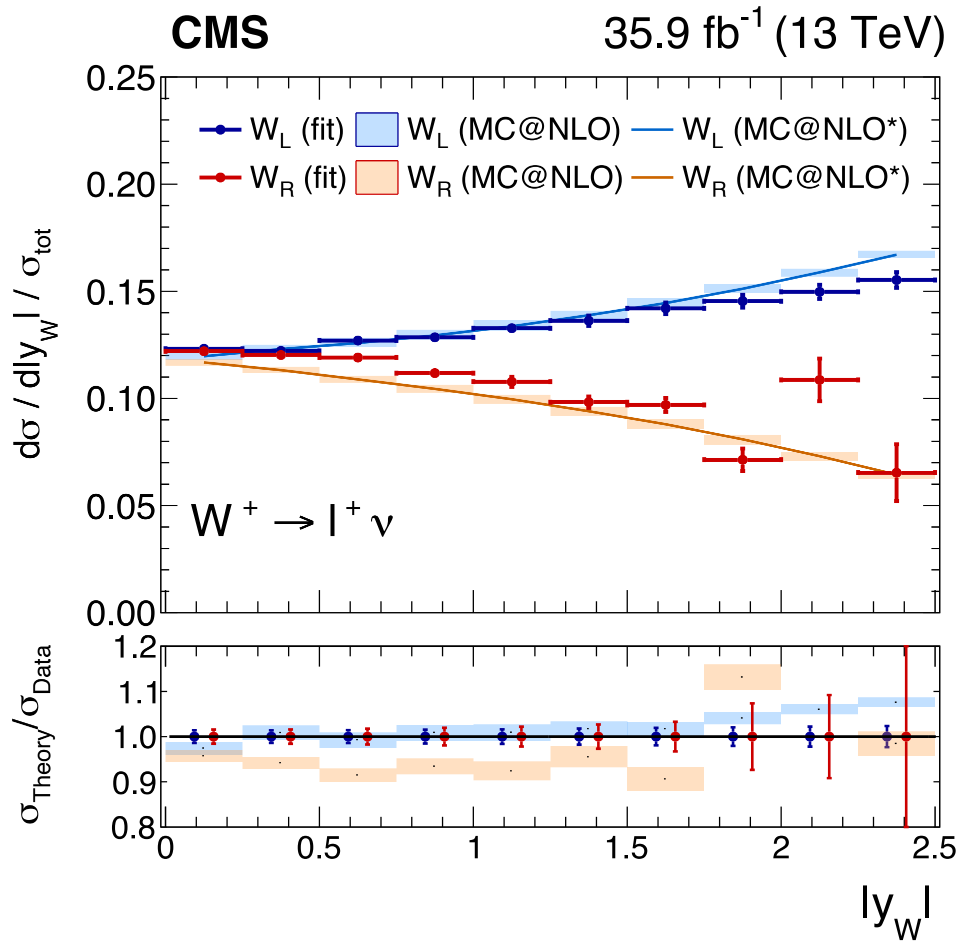
CMS, *Phys.Rev.D* 102 (2020) 9, 092012

- Many more measurements since then….

  - Recent example - polarised W boson production.
  - Improved sensitivity to valence quarks and strangeness.

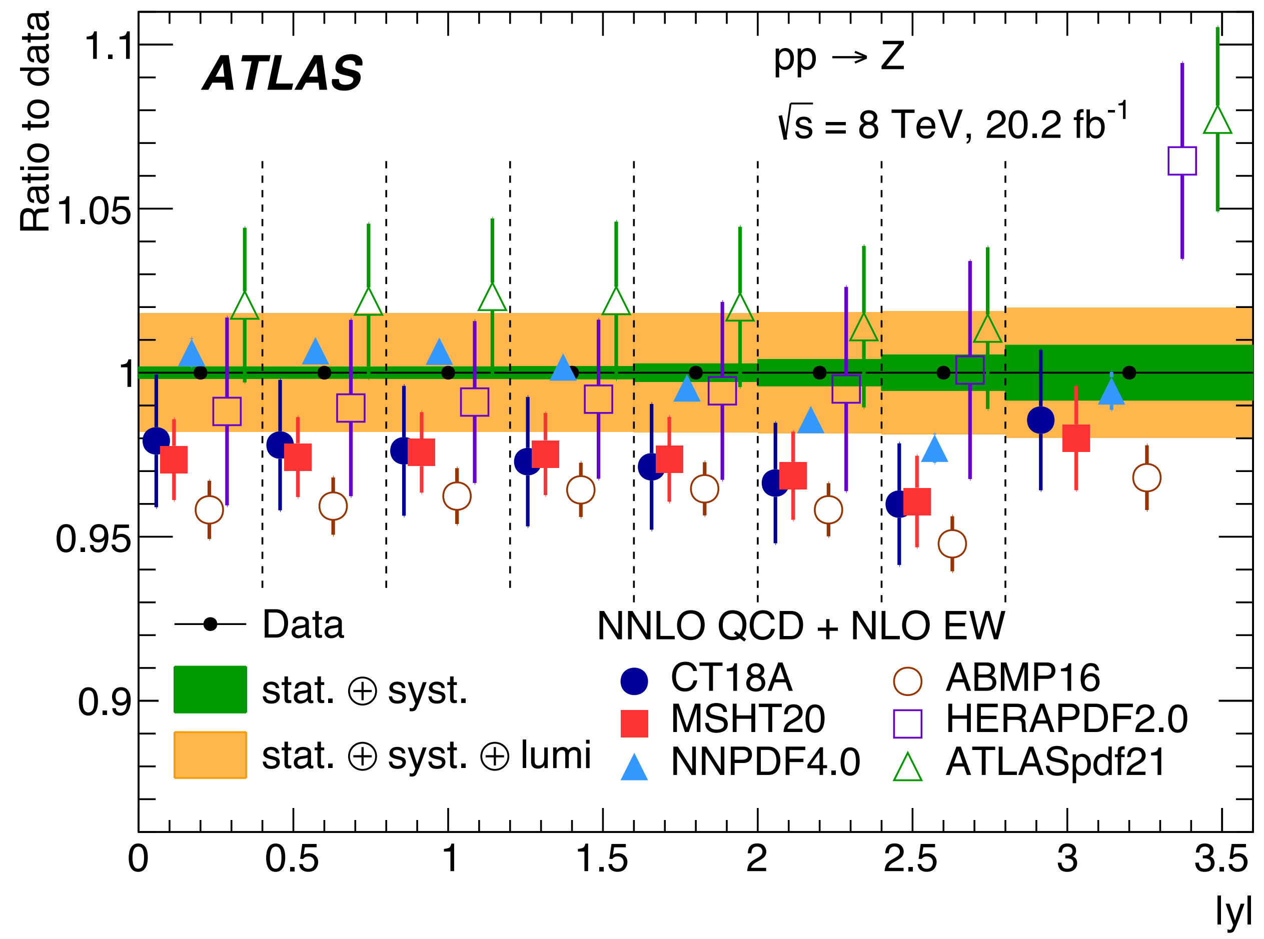
Measurements of the W boson rapidity, helicity, double-differential cross sections, and charge asymmetry in pp collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration*



- Many other new observables/data from other experiments also available for future fits:

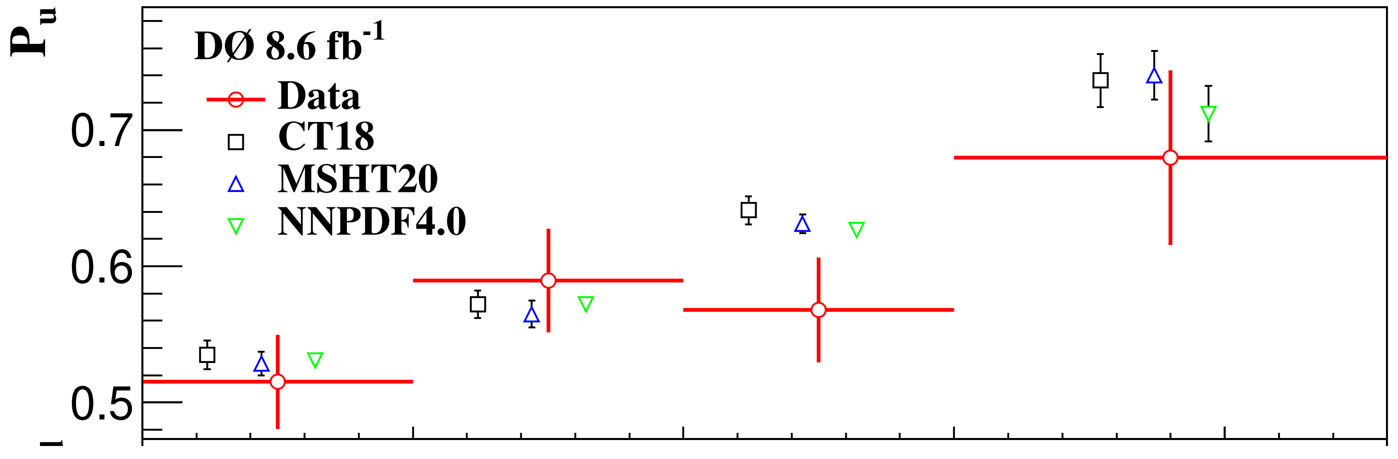
★ **New observables:** l^+l^-corrected to full phase space. Angular coefficients - limit extrapolation uncertainty.

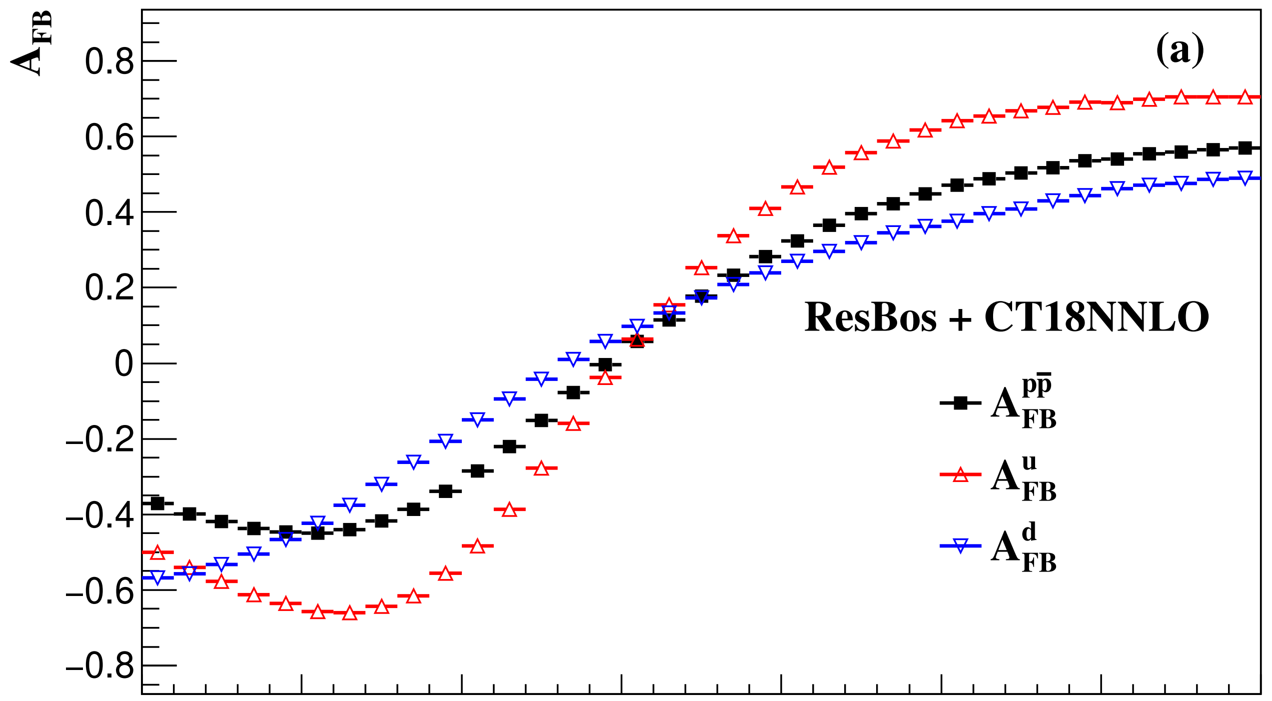


PDF set	Total χ^2 / d.o.f.	χ^2 p-value	Pull on luminosity
MSHT20aN^{3}LO [58]	13/8	0.11	1.2 ± 0.6
CT18A [59]	12/8	0.17	0.9 ± 0.7
MSHT20 [60]	10/8	0.26	0.9 ± 0.6
NNPDF4.0 [61]	30/8	0.0002	0.0 ± 0.2
ABMP16 [62, 63]	30/8	0.0002	1.8 ± 0.4
HERAPDF2.0 [64]	22/8	0.005	−1.3 ± 0.8
ATLASpdf21 [65]	20/8	0.01	−1.1 ± 0.8

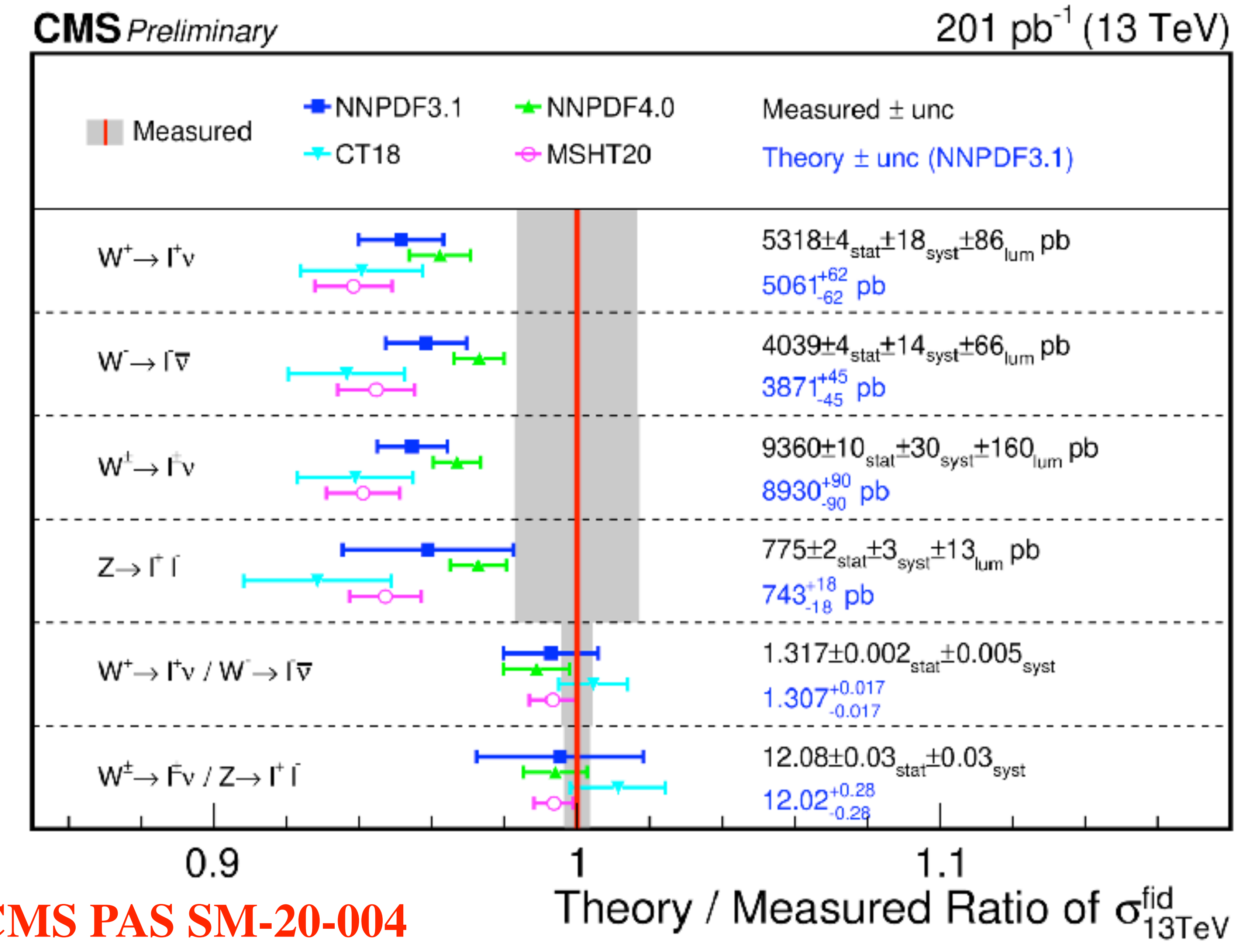
ATLAS, arXiv:2309.09318

★ **New data:** Not just the LHC. Recent **D0** measurement of dilepton AFB. Sensitivity to high x flavour structure.





★ **New ratios:** low lumi runs. Ratios at different energies increase PDF sensitivity.

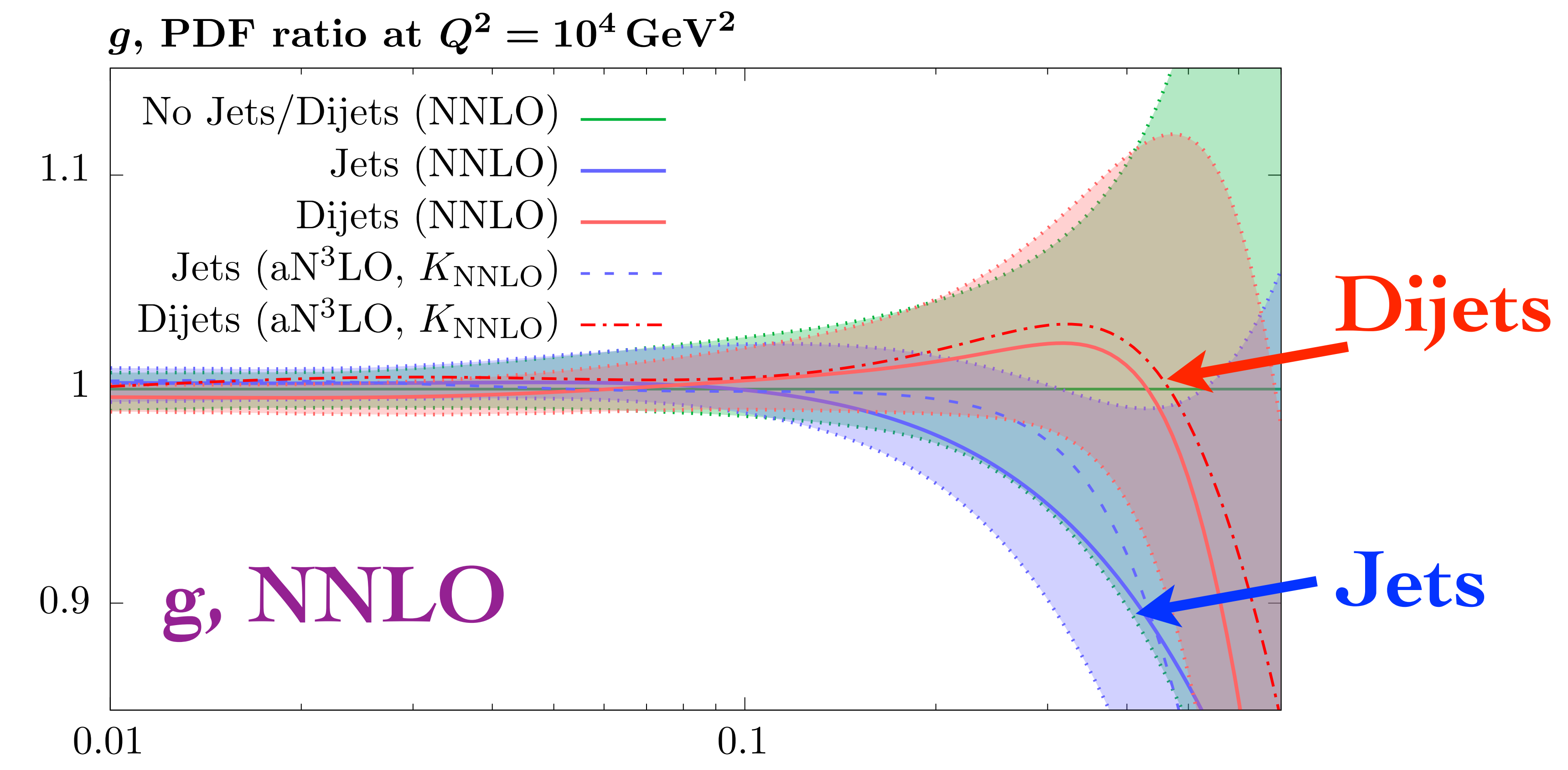


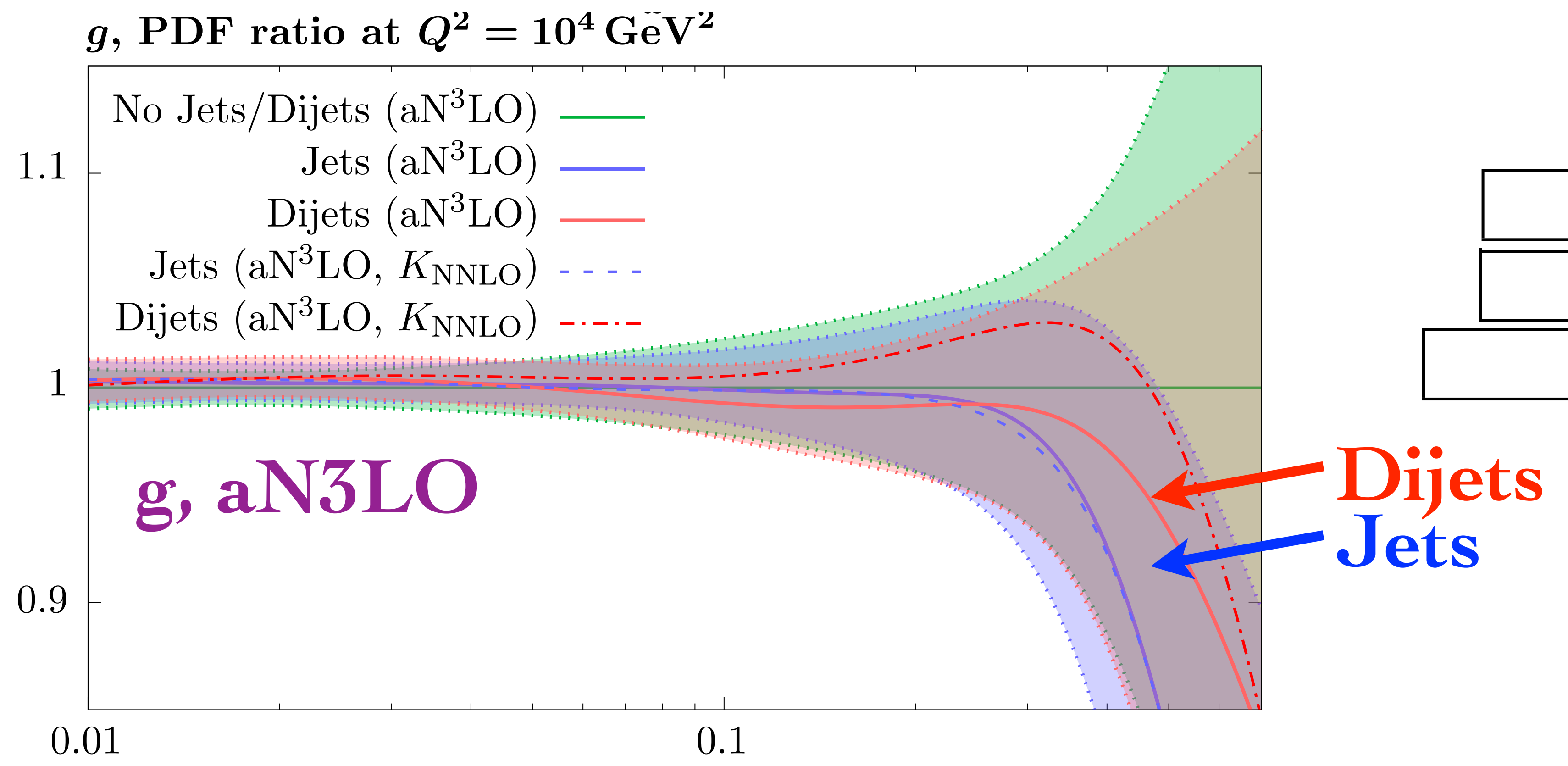
CMS PAS SM-20-004

- And much more not shown here.

Impact of New Data

- Impact of newer (13 TeV) data being assessed, and older (7-8 TeV) data within new theory approaches:
  - ★ New study of impact of **jet** vs. **dijet** production at up to **aN3LO** order (more later).
  - ★ Preference for dijet data, and for aN3LO. PDF impact depends on order (NNLO vs aN3LO).





χ^2/N_{pts}

	N_{pts}	NNLO	aN^{3}LO
Total Jets	643	1.67	1.63
Total Dijets	266	1.13	1.04

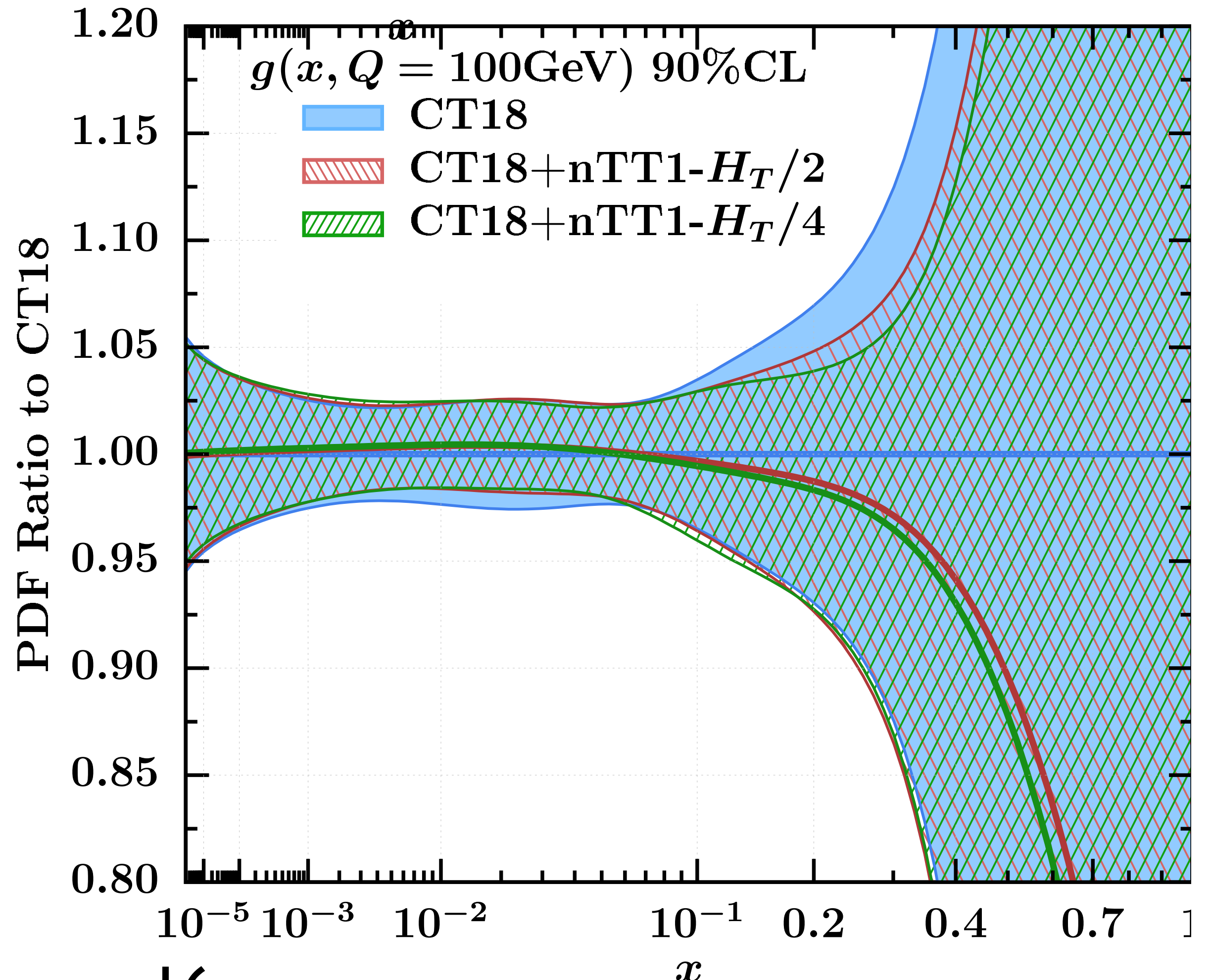
MSHT, arXiv:2312.12505

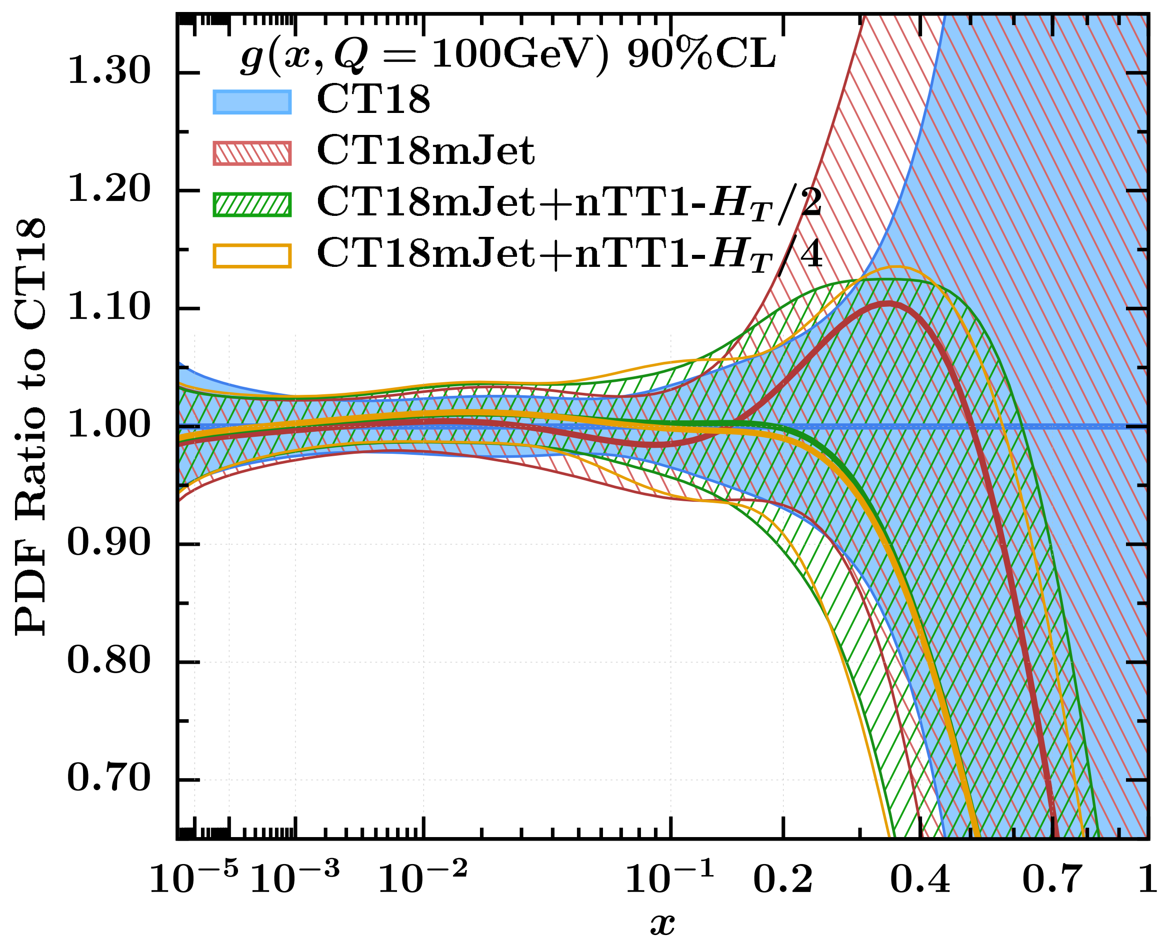
✦ 13 TeV $t\bar{t}$: study within **CT** global PDF fit.

✦ Impact moderate but non-negligible.

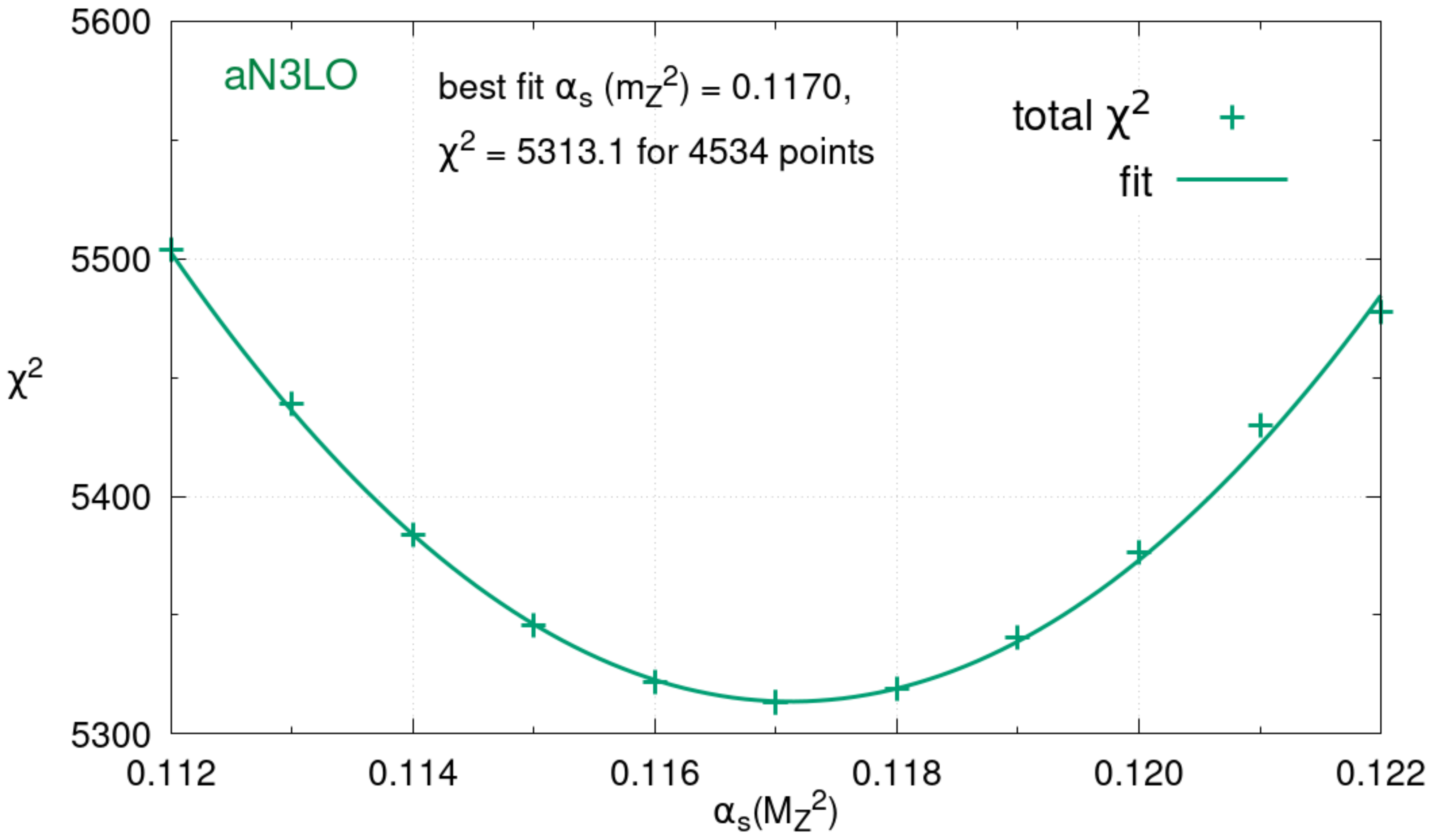
✦ Complementarity with LHC jet data highlighted.

A. Ablat et el., arXiv:2307.11153





★ First simultaneous α_S + global PDF extraction at aN3LO.



MSHT, arXiv:2404.02964

$$\alpha_S(M_Z^2)(\text{NNLO}) = 0.1171 \pm 0.0014$$

$$\alpha_S(M_Z^2)(\text{aN}^3\text{LO}) = \underline{0.1170 \pm 0.0016}$$

★ Nice **convergence** from NNLO to aN3LO. Fully consistent with PDG.

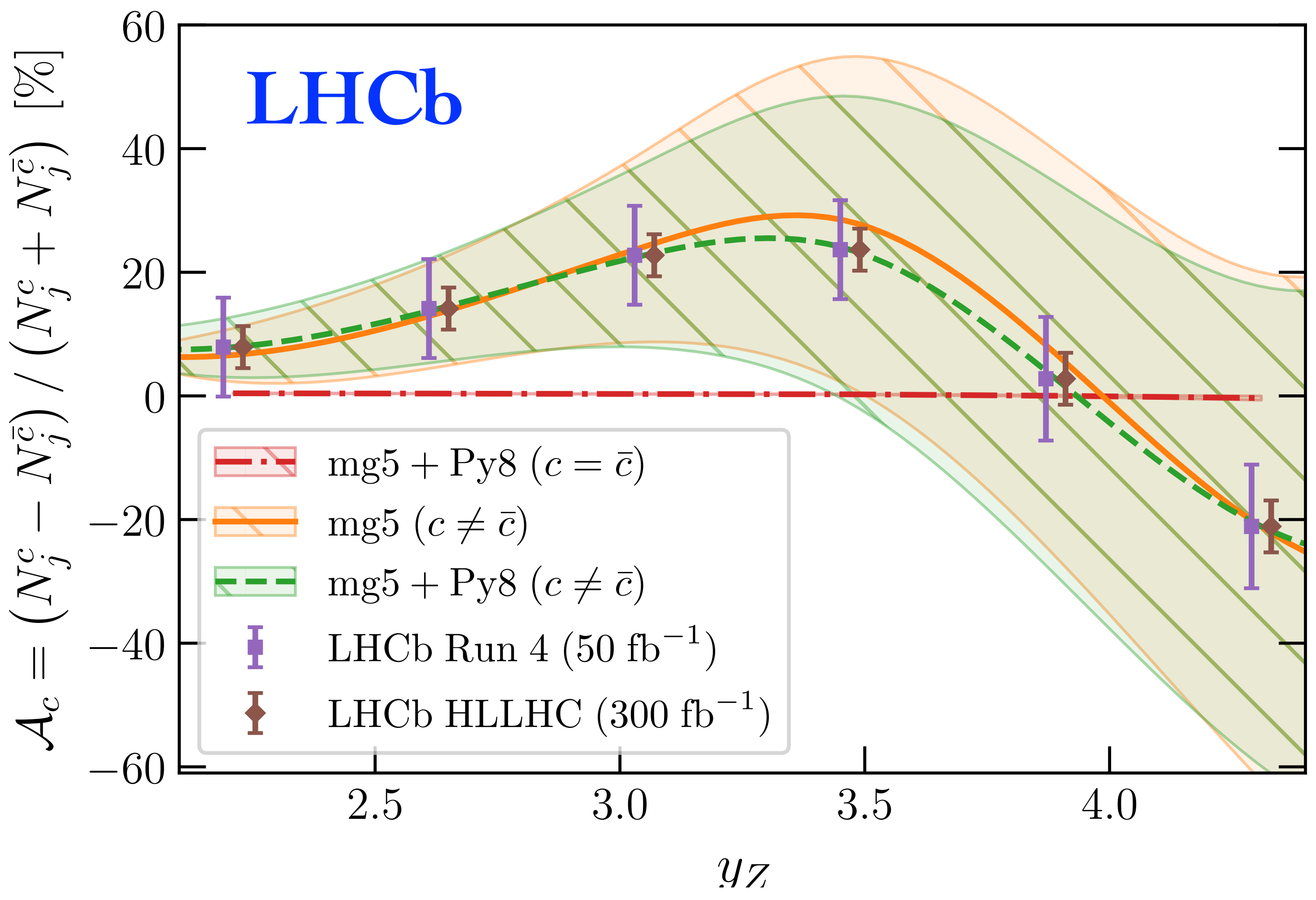
★ Errors slightly larger (more accurate) due to MHO uncertainty.

✦ New data - and theory - **'valence' charm** in proton. Evidence for non perturbatively generate charm and charm difference - `instrinsic' charm.

✦ Evidence from global fit quality, but particular LHC (+ EIC) data sets can have further impact.

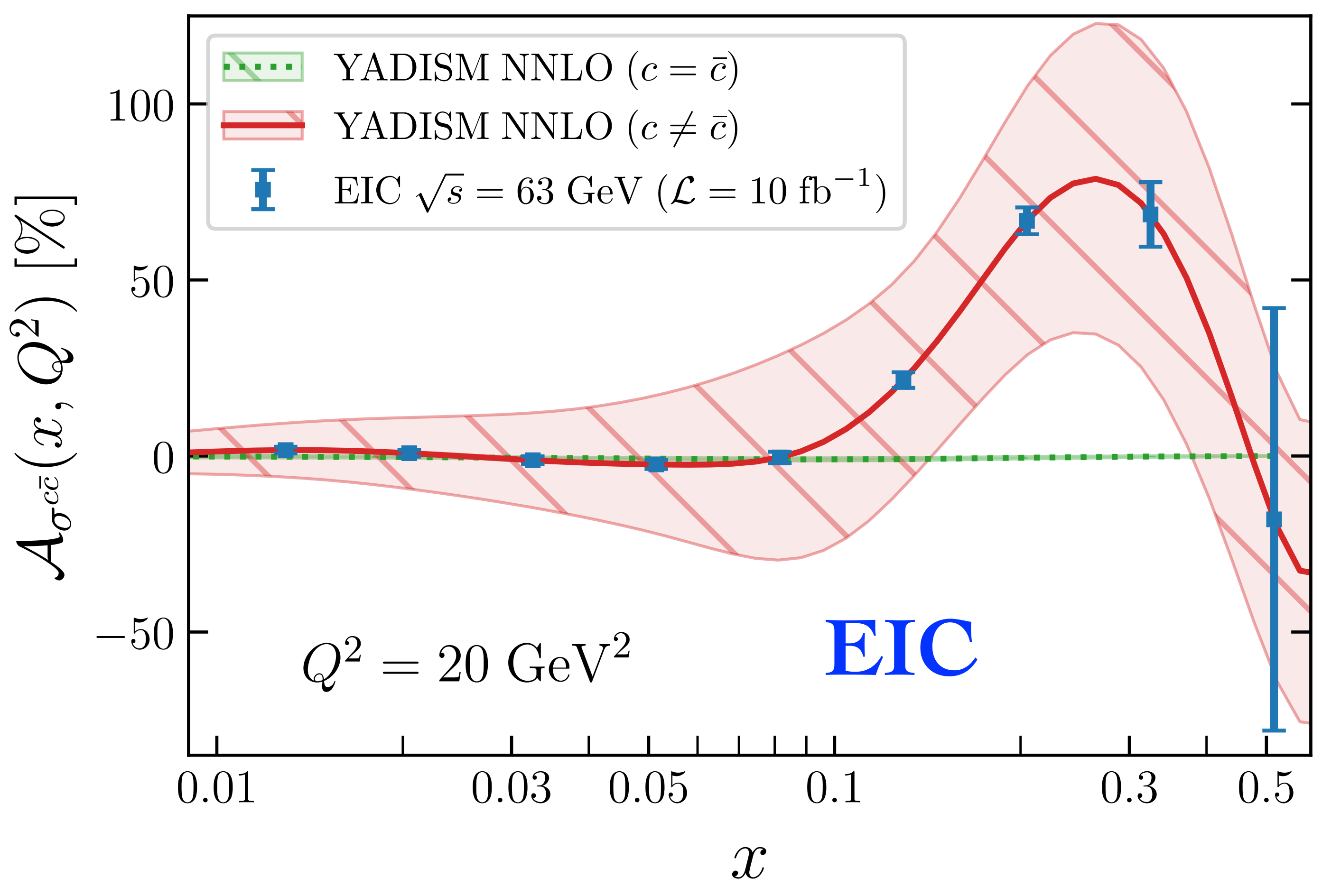
See talk by J. Rojo

R. Ball et al., arXiv:2311.00743



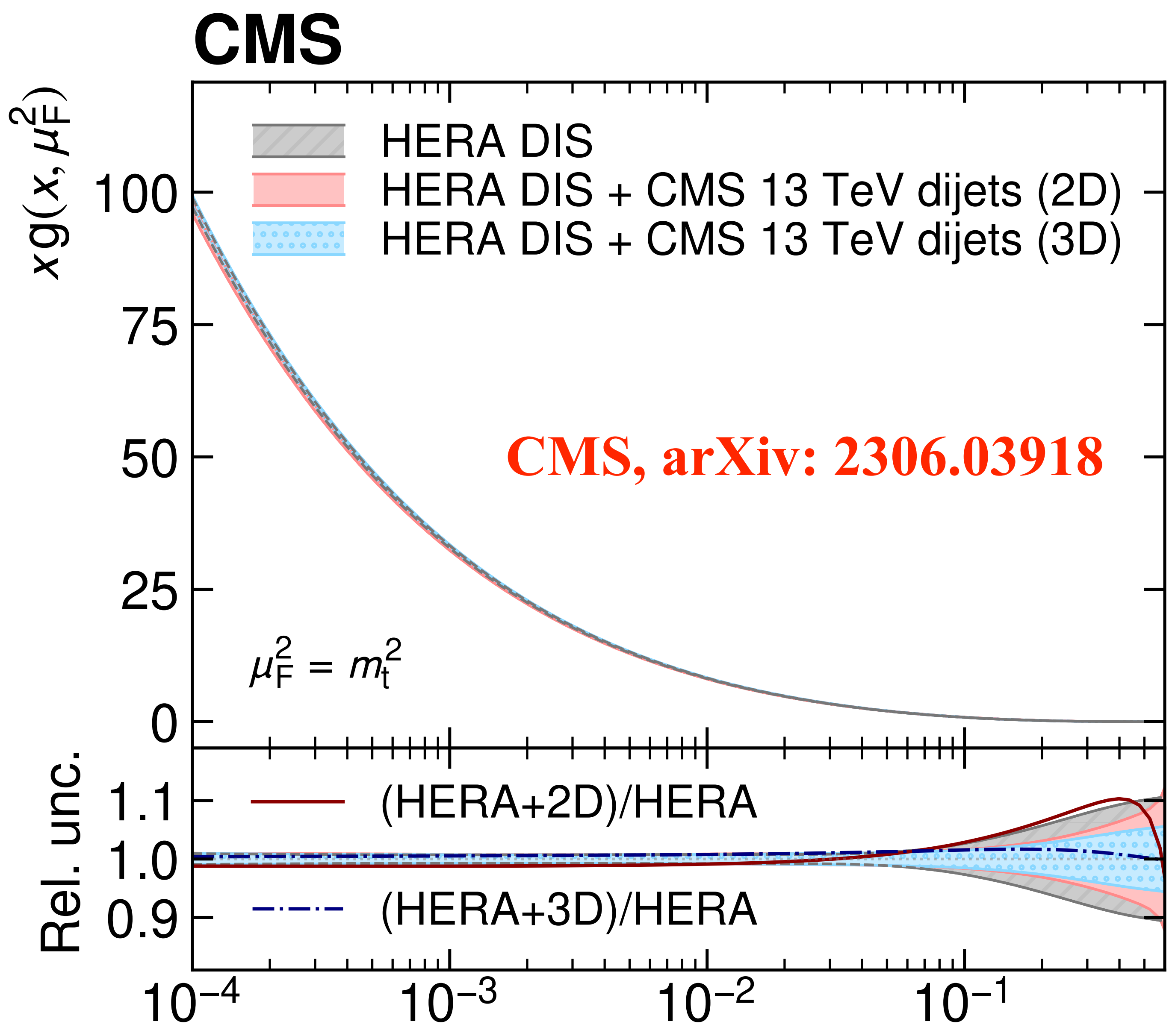
$$\mathcal{A}_c(y_Z) \equiv \frac{N_j^c(y_Z) - N_j^{\bar{c}}(y_Z)}{N_j^c(y_Z) + N_j^{\bar{c}}(y_Z)}$$

$$\mathcal{A}_{\sigma^{c\bar{c}}}(x, Q^2) \equiv \frac{\sigma^c_{\text{red}}(x, Q^2) - \sigma^{\bar{c}}_{\text{red}}(x, Q^2)}{\sigma^{c\bar{c}}_{\text{red}}(x, Q^2)}\,.$$

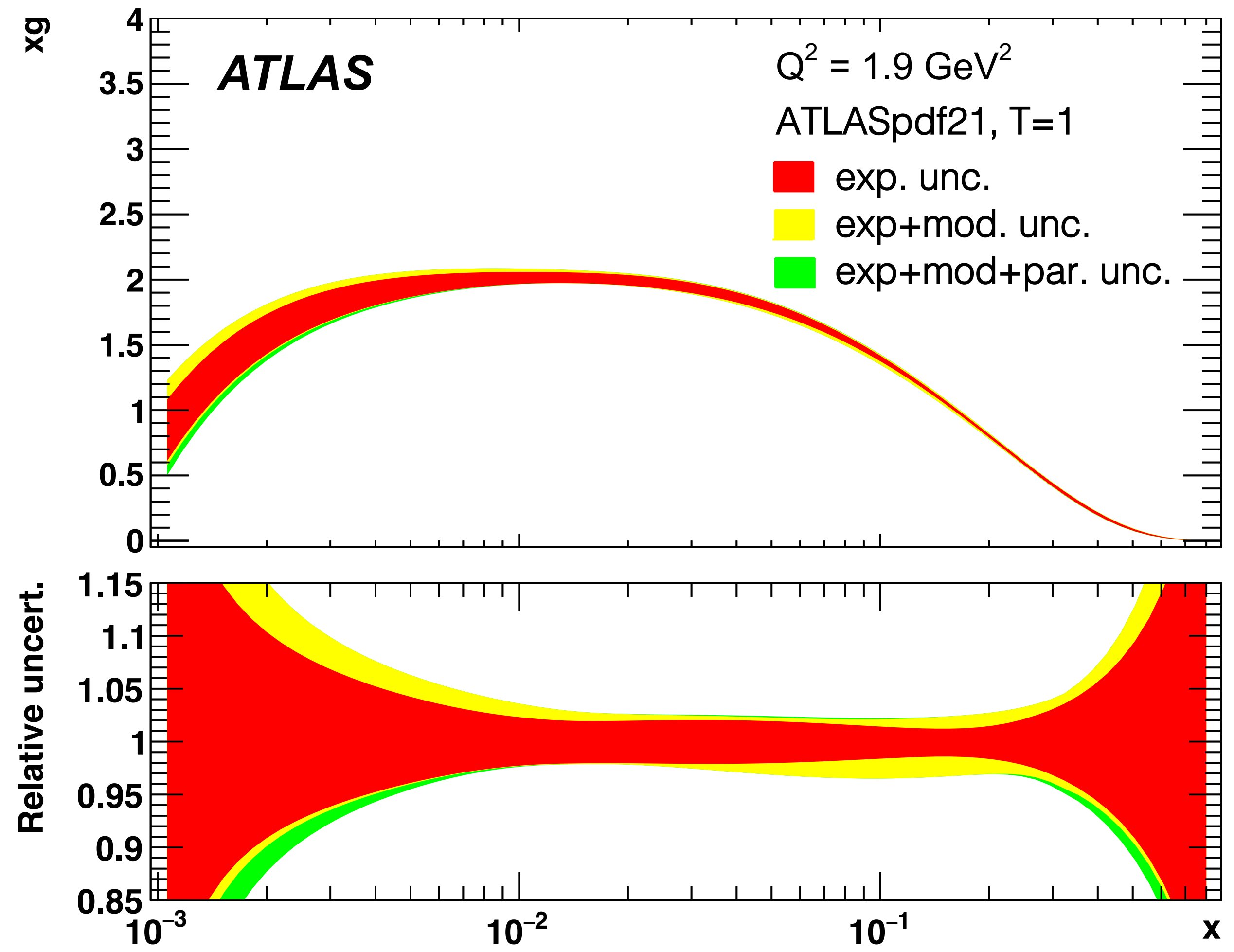


★ Not just global fitters - experimental collaborations also busily assessing impact of their data on PDFs.

★ Though some caution often needed: impact not the same as in global fit.



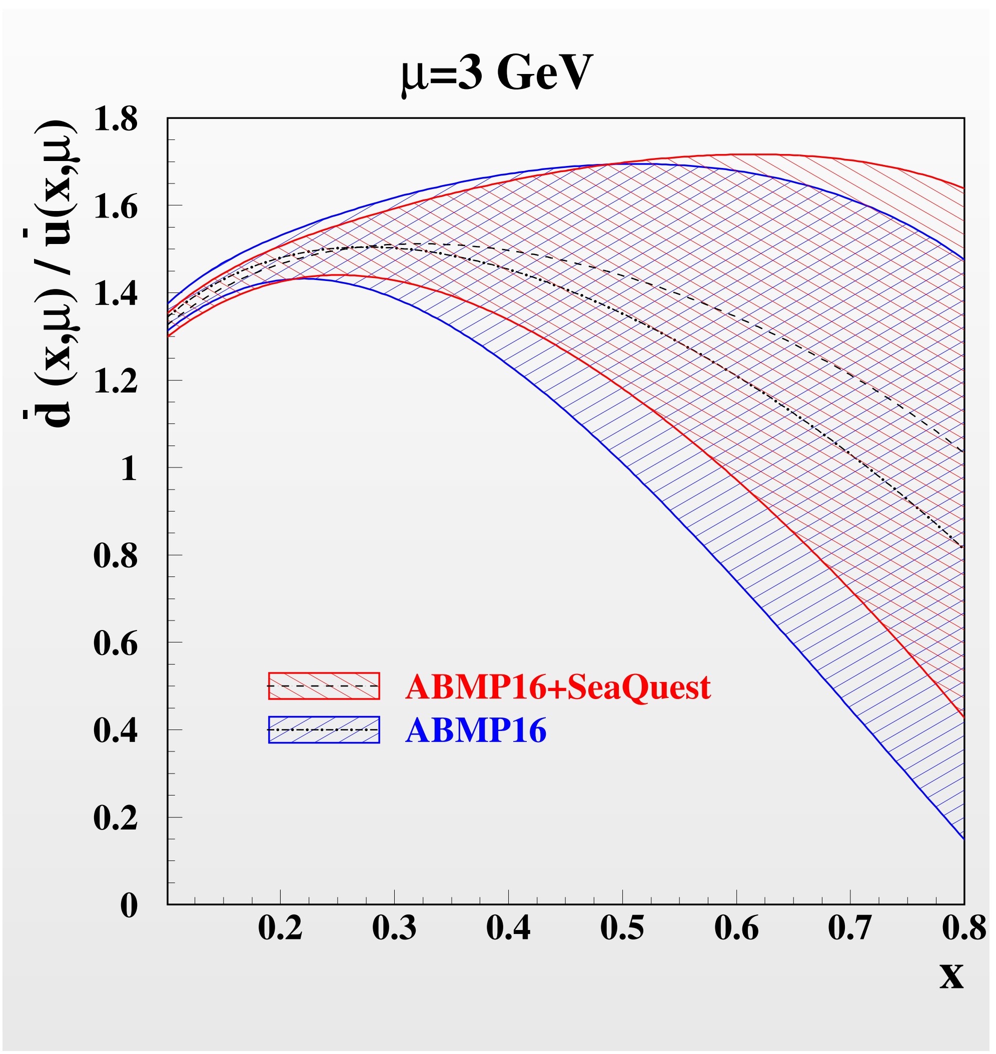
ATLAS, arXiv:2306.03918

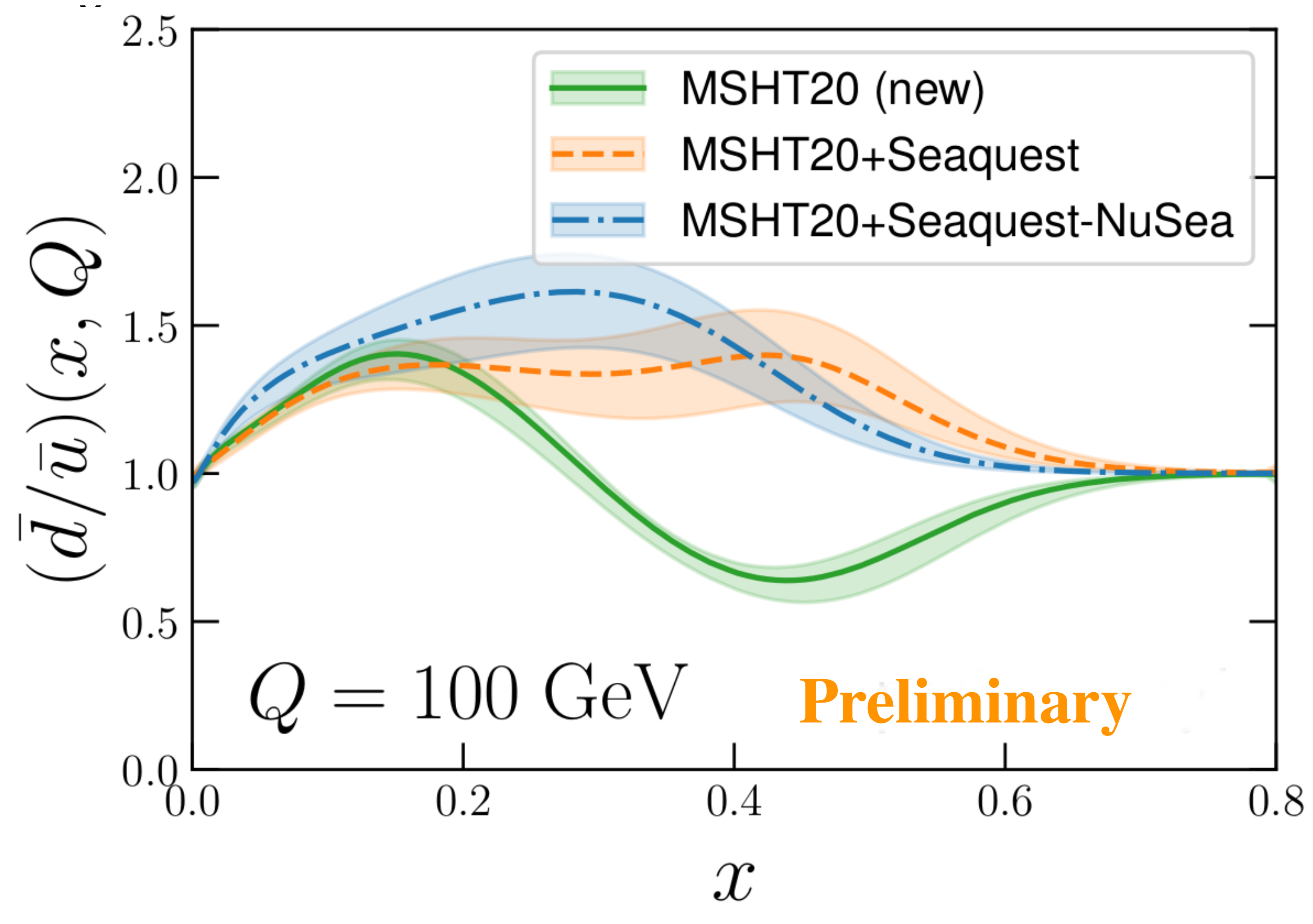


ATLASpdf21

★ Not just LHC data. Seaquest example of non-LHC dataset with important impact (high x flavour decomposition).

★ Though some tension with other (NuSea) data!





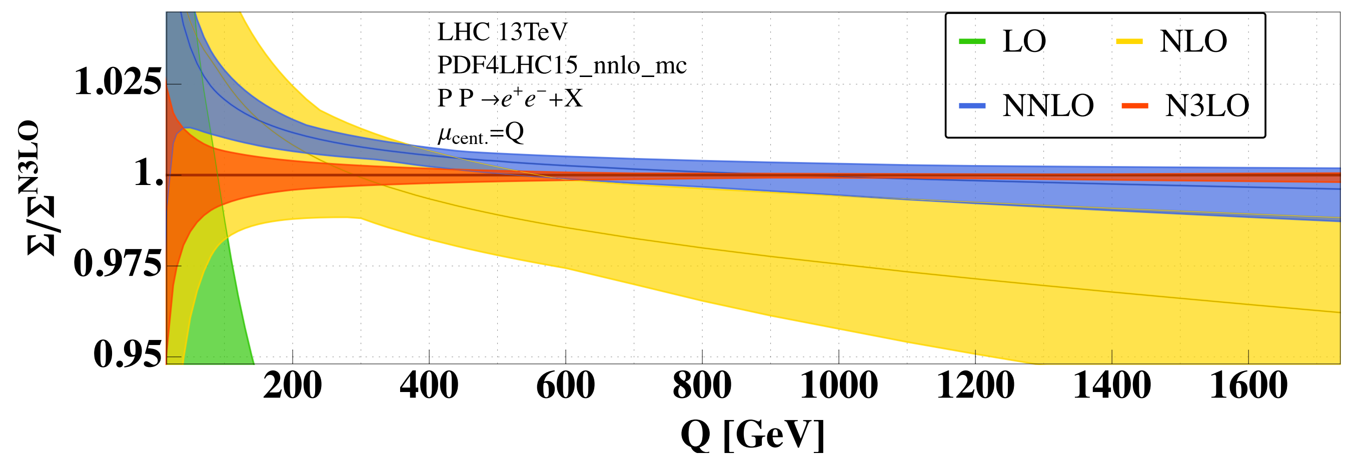
M. V. Garzeili, PDF4LHC23 meeting

New Developments

Motivation

- **N3LO:**
  - ★ State of the art is NNLO for PDF fits but a lot known at N3LO about DGLAP evolution and DIS (light + heavy flavours). Why not use this?
  - ★ For hadron colliders less is known but already quite a bit
  - **Uncertainty** due to lack of N3LO PDFs a key factor $\Rightarrow$ need to - and can - go to N3LO!

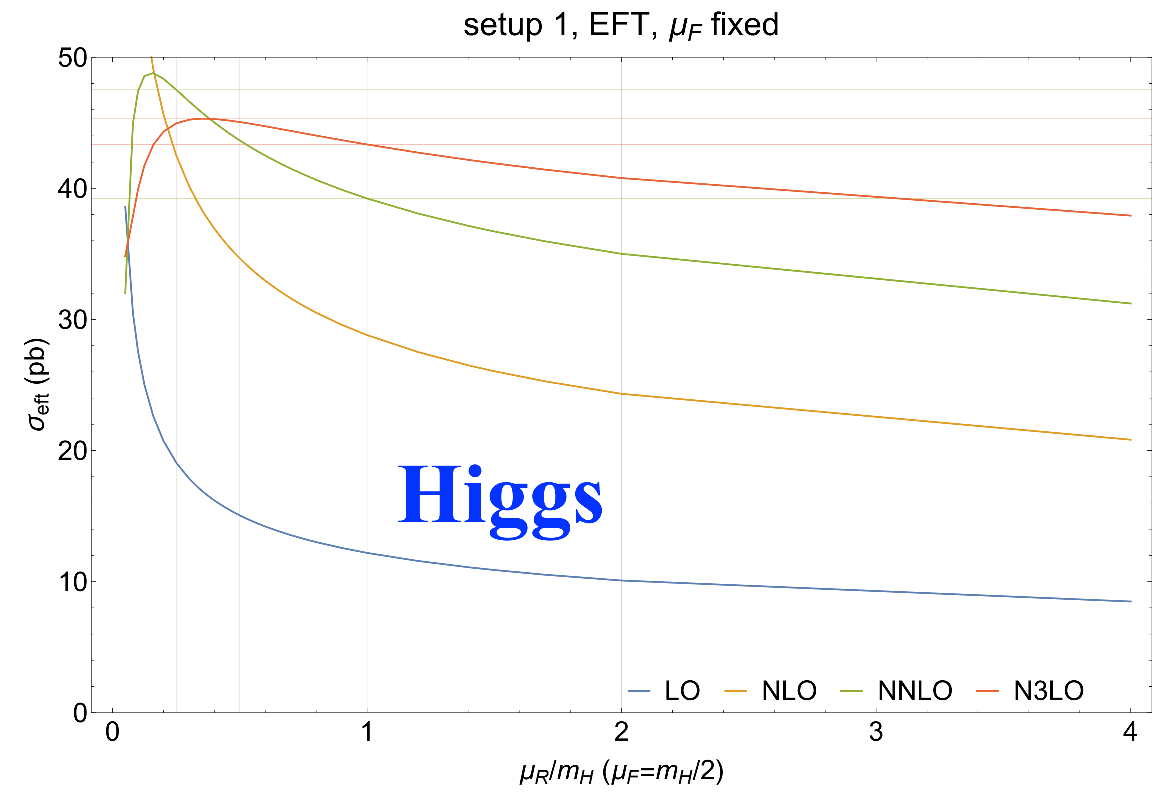
Drell Yan



C. Duhr and B. Mistleberger, arXiv:2111.10379

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

C. Anastasiou et al., arxiv:1602.00695



- **Missing higher orders:**
  - ★ As (LHC) data becomes ever more precise sensitivity to any data/theory mismatch increases.

$$\chi^2 \sim \sum \frac{(D-T)^2}{\sigma^2_{\mathrm{exp}}} \qquad T_{\mathrm{N}^x\mathrm{LO}} \neq D \Rightarrow \chi^2 \to \infty \text{ as } \sigma_{\mathrm{exp}} \to 0$$

  - ★ Need to account for this missing higher order uncertainty:
    - More accurate PDF uncertainty.
    - Weight datasets correctly in fit (less well known $\Rightarrow$ larger uncertainty).

New Developments : aN3LO and missing higher orders

- Two approximate N3LO (aN3LO) global PDF sets available: **MSHTaN3LO** and **NNPDF4.0aN3LO**.

Emanuele Nocera, Forward Physics and QCD at the LHC and EIC, Bad Honnef 23

  - Approximate splitting functions built up from known information. Approximate $\neq$ poorly known. A lot of information available:

Splitting Functions

Singlet (P_{qq}, P_{gg}, P_{gq}, P_{qg})
- large-n_f limit [NPB 915 (2017) 335; arXiv:2308.07958]
- small-x limit [JHEP 06 (2018) 145]
- large-x limit [NPB 832 (2010) 152; JHEP 04 (2020) 018; JHEP 09 (2022) 155]
- 5 (10) lowest Mellin moments [PLB 825 (2022) 136853; ibid. 842 (2023) 137944; ibid. 846 (2023) 138215]

Non-singlet ($P_{NS,v}$, $P_{NS,+}$, $P_{NS,-}$)
- large-n_f limit [NPB 915 (2017) 335; arXiv:2308.07958]
- small-x limit [JHEP 08 (2022) 135]
- large-x limit [JHEP 10 (2017) 041]
- 8 lowest Mellin moments [JHEP 06 (2018) 073]

- And a great deal of progress recently!

G. Falcioni et al., arXiv:2307.04158

$$
\begin{aligned}
\gamma^{(3)}_{qg}(N=2) &= -654.4627782\,n_f + 245.6106197\,n_f^2 - 0.924990969\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=4) &= 290.3110686\,n_f - 76.51672403\,n_f^2 - 4.911625629\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=6) &= 335.8008046\,n_f - 124.5710225\,n_f^2 - 4.193871425\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=8) &= 294.5876830\,n_f - 135.3767647\,n_f^2 - 3.609775642\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=10) &= 241.6153399\,n_f - 135.1874247\,n_f^2 - 3.189394834\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=12) &= 191.9712464\,n_f - 131.1631663\,n_f^2 - 2.877104430\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=14) &= 148.5682948\,n_f - 125.8231081\,n_f^2 - 2.635918561\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=16) &= 111.3404252\,n_f - 120.1681987\,n_f^2 - 2.443379039\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=18) &= 79.51561588\,n_f - 114.6171354\,n_f^2 - 2.285486861\,n_f^3\,,\\
\gamma^{(3)}_{qg}(N=20) &= 52.24329555\,n_f - 109.3424891\,n_f^2 - 2.153153725\,n_f^3\,.
\end{aligned}
$$

G. Falcioni et al., arXiv:2302.07593

$$
\begin{aligned}
\gamma^{(3)}_{ps}(N=2) &= -691.5937093\,n_f + 84.77398149\,n_f^2 + 4.466956849\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=4) &= -109.3302335\,n_f + 8.776885259\,n_f^2 + 0.306077137\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=6) &= -46.03061374\,n_f + 4.744075766\,n_f^2 + 0.042548957\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=8) &= -24.01455020\,n_f + 3.235193483\,n_f^2 - 0.007889256\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=10) &= -13.73039387\,n_f + 2.375018759\,n_f^2 - 0.021029241\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=12) &= -8.152592251\,n_f + 1.819958178\,n_f^2 - 0.024330231\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=14) &= -4.840447180\,n_f + 1.438327380\,n_f^2 - 0.024479943\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=16) &= -2.751136330\,n_f + 1.164299642\,n_f^2 - 0.023546009\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=18) &= -1.375969240\,n_f + 0.960873318\,n_f^2 - 0.022264393\,n_f^3\,,\\
\gamma^{(3)}_{ps}(N=20) &= -0.442681568\,n_f + 0.805745333\,n_f^2 - 0.020918264\,n_f^3\,.
\end{aligned}
$$

$$\int_0^1 \mathrm{d}x\, x^{N-1} P(x)$$

Up to $N = 20$ even moments in quark sector

N3LO - What do we know?

- **Approximate** $\neq$ **poorly known**!

Emanuele Nocera, Forward Physics and QCD at the LHC and EIC, Bad Honnef 23

$$P(x,\alpha_s) = \alpha_S P^{(0)}(x) + \alpha_S^2 P^{(1)}(x) + \alpha_S^3 P^{(2)}(x) + \alpha_S^4 P^{(3)}(x) + \ldots$$

Splitting Functions

Singlet (P_{qq}, P_{gg}, P_{gq}, P_{qg})
- large-n_f limit [NPB 915 (2017) 335; arXiv:2308.07958]
- small-x limit [JHEP 06 (2018) 145]
- large-x limit [NPB 832 (2010) 152; JHEP 04 (2020) 018; JHEP 09 (2022) 155]
- 5 (10) lowest Mellin moments [PLB 825 (2022) 136853; ibid. 842 (2023) 137944; ibid. 846 (2023) 1

Non-singlet ($P_{NS,v}$, $P_{NS,+}$, $P_{NS,-}$)
- large-n_f limit [NPB 915 (2017) 335; arXiv:2308.07958]
- small-x limit [JHEP 08 (2022) 135]
- large-x limit [JHEP 10 (2017) 041]
- 8 lowest Mellin moments [JHEP 06 (2018) 073]

★ **Splitting functions:** a wealth of information. Moments & various limits, with much recent further progress.

G. Falcioni et al., arXiv:2307.04158, arXiv:2302.07593

$$F_2(x,Q^2) = \sum_{\alpha\in H,q,g;\beta\in q,H} \left(C^{VF,n_f+1}_{\beta,\alpha} \otimes A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2)\right)$$

★ **DIS:** massless coefficient functions known (+ massive high Q^2). Massive low Q^2 approx. known.

$$f_\alpha^{n_f+1}(x,Q^2) = [A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2)](x)$$

★ **Heavy Flavour:** again wealth of information. Moments & various limits, with much recent progress.

$$\sigma = \sigma_0 + \sigma_1 + \sigma_2 + \sigma_3 + \ldots \equiv \sigma_{N3LO} + \ldots$$

★ **Hadronic Cross Sections:** while much progress made, thus far not useable in PDF fits.

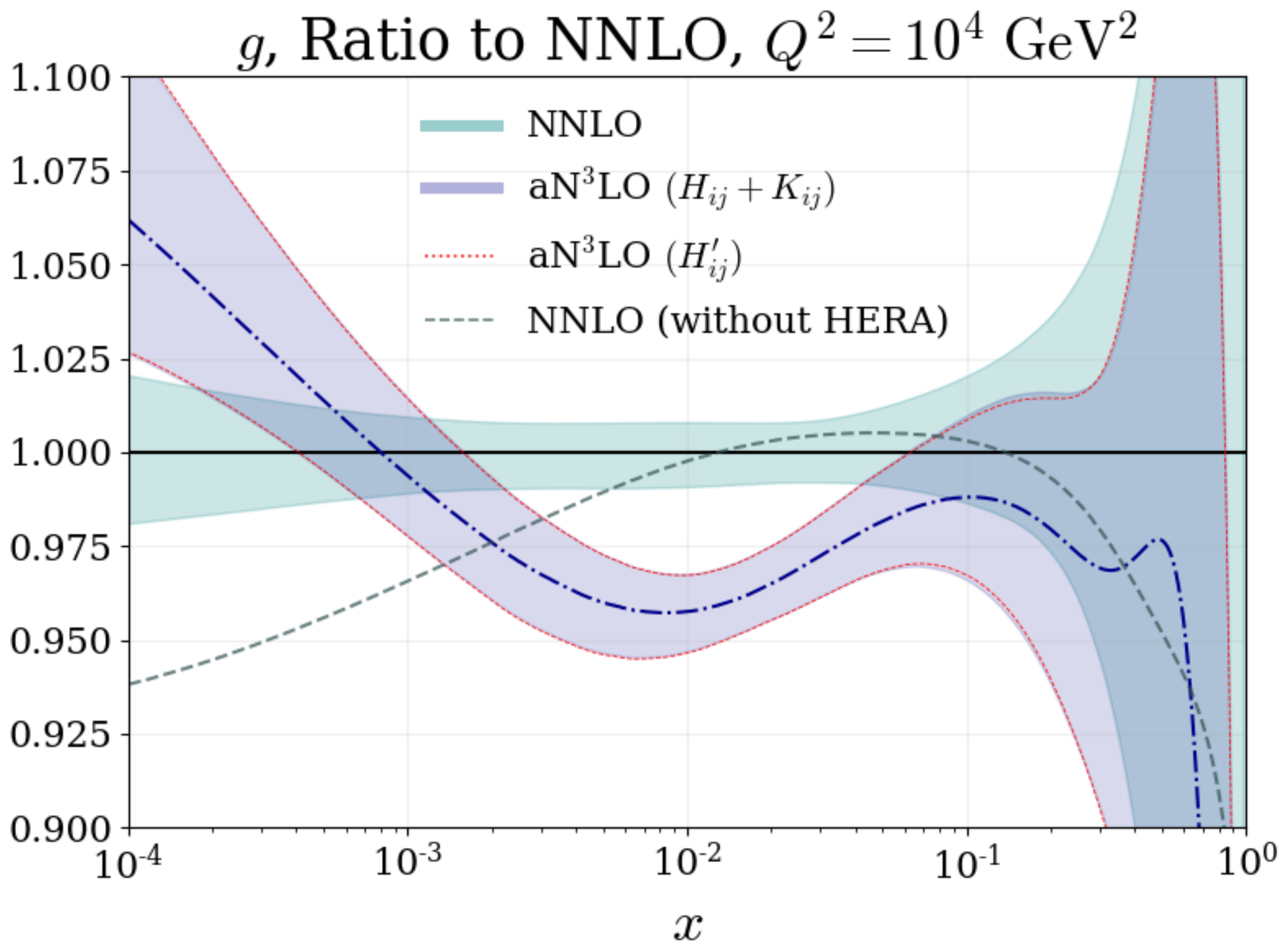
- First three ingredients now largely known with sufficient precision to give close to a N3LO fit. Final ingredient clearly the bottleneck for that - approximation + uncertainty required.

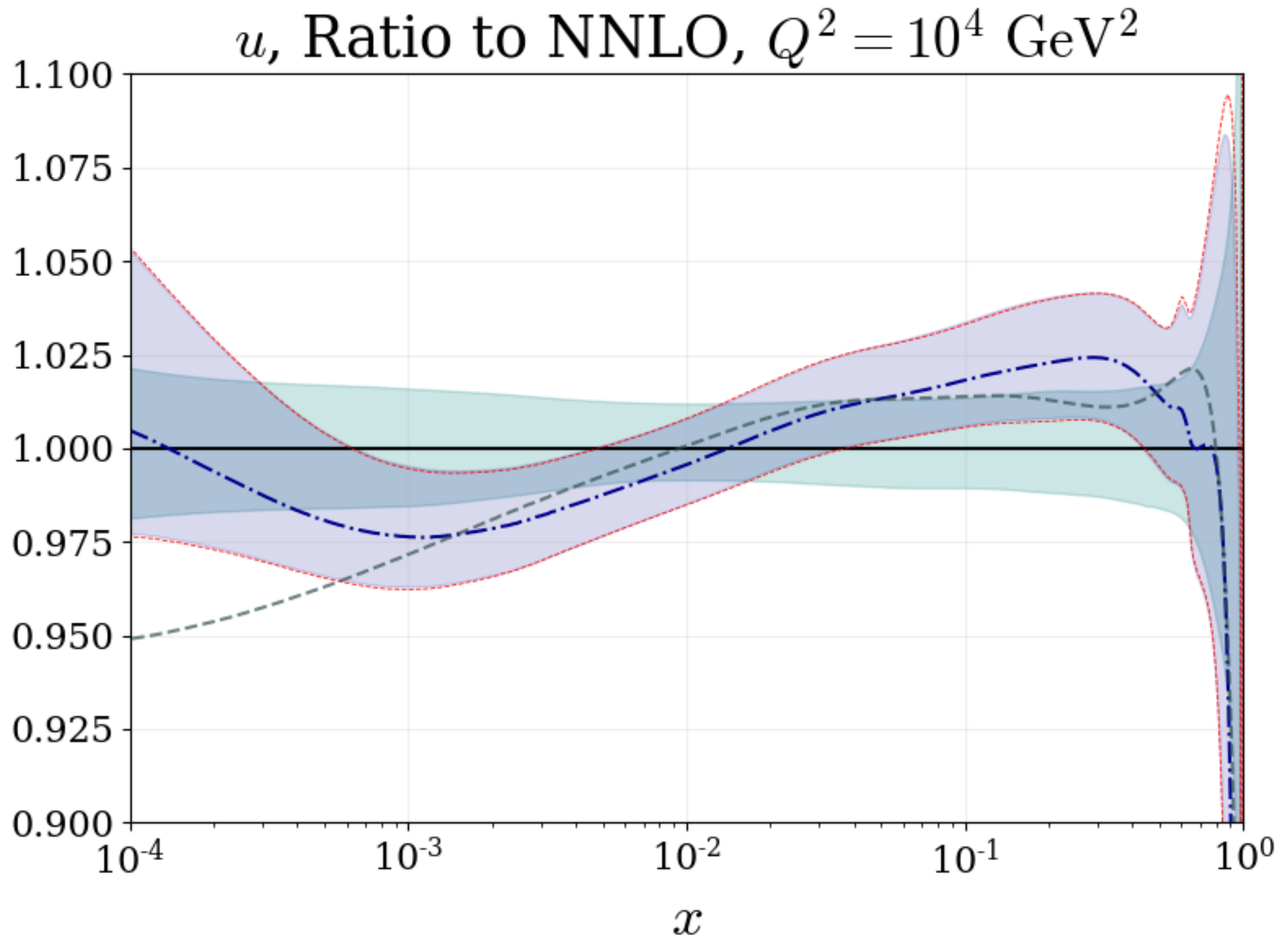
MSHTaN3LO

- First global aN3LO analysis - **MSHT20aN3LO**. Released ~ 2 years ago.
- Main bottleneck to 'real' N3LO is hadronic cross sections. Include via nuisance parameters:
- Clear improvement in fit quality, ~ driven by known N3LO.

	LO	NLO	NNLO	N^3LO
$\chi^2_{N_{pts}}$	2.57	1.33	1.17	1.14

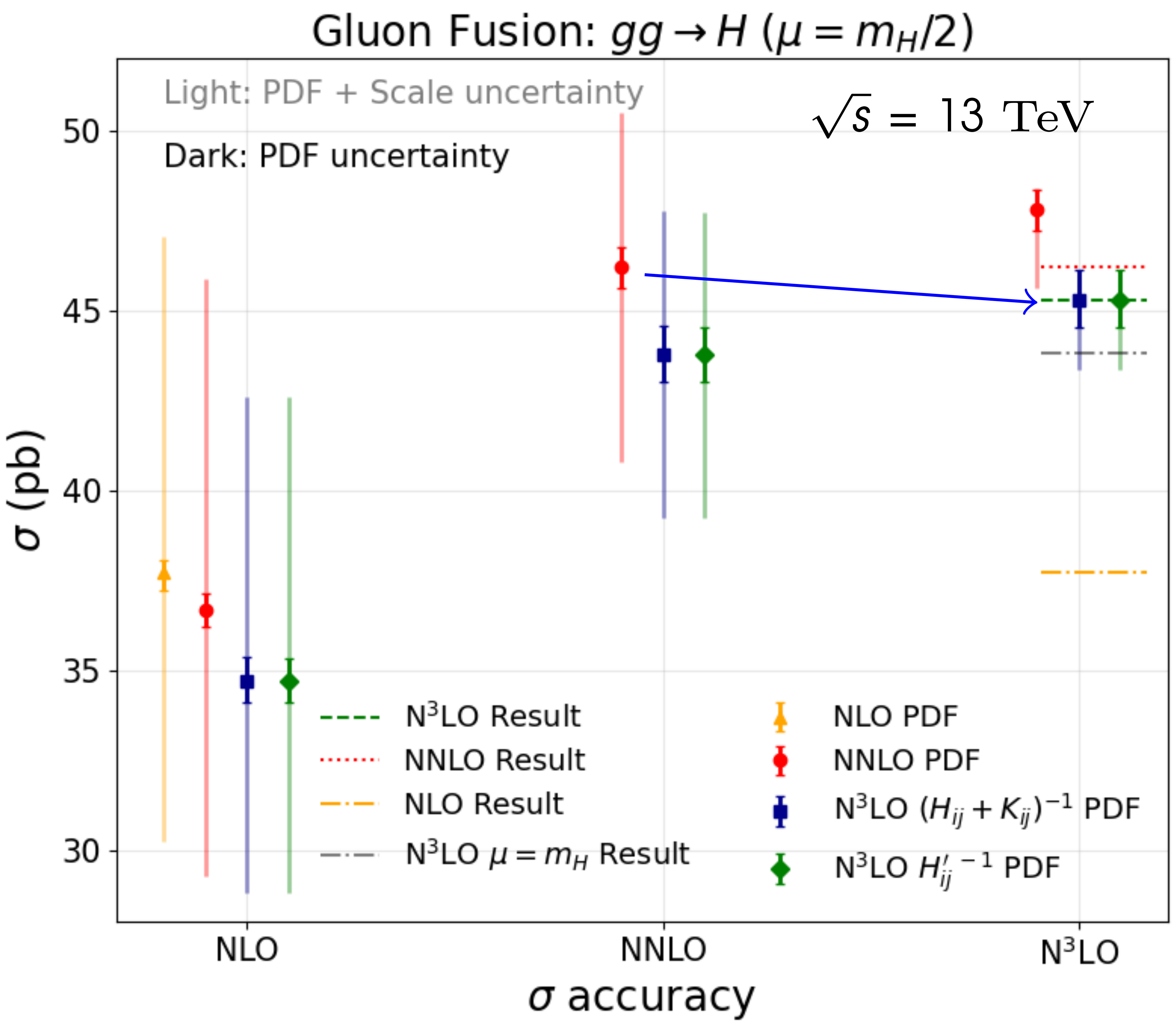
- Evidence that aN3LO reduces tensions between low and high x regions.
- Largest change is in gluon at low and intermediate x. Some change in e.g. quarks at high x.



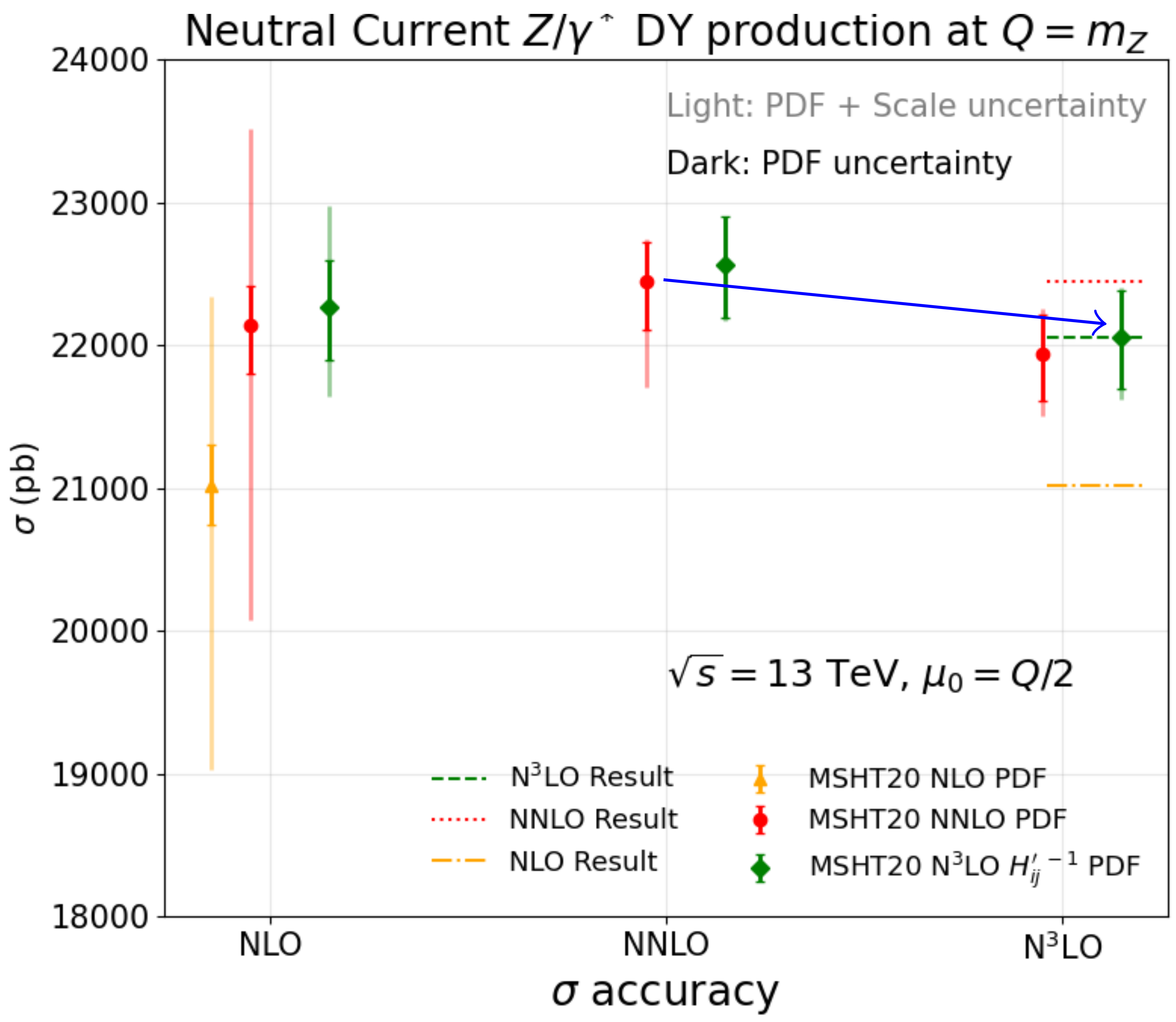


- More recent **NNPDF4.0aN3LO** analysis sees qualitatively similar results, some quantitative differences.

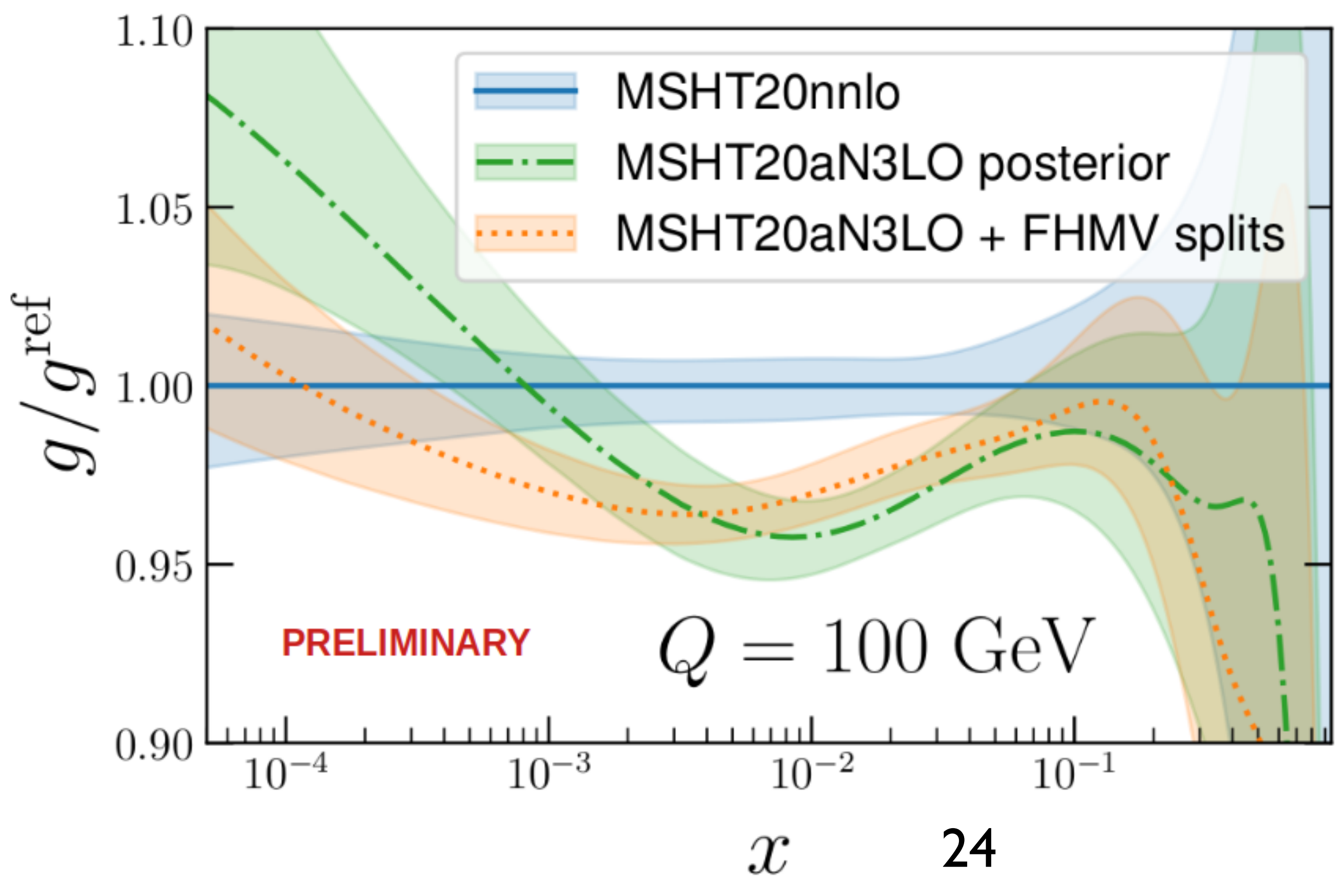
- Change in gluon corresponds to reduction in e.g. ggH at N3LO - improves stability.



- Some increase in NC DY - again mild improvement in stability.



- Have studied impact of newer splitting function information. Moderate impact, within uncertainties.

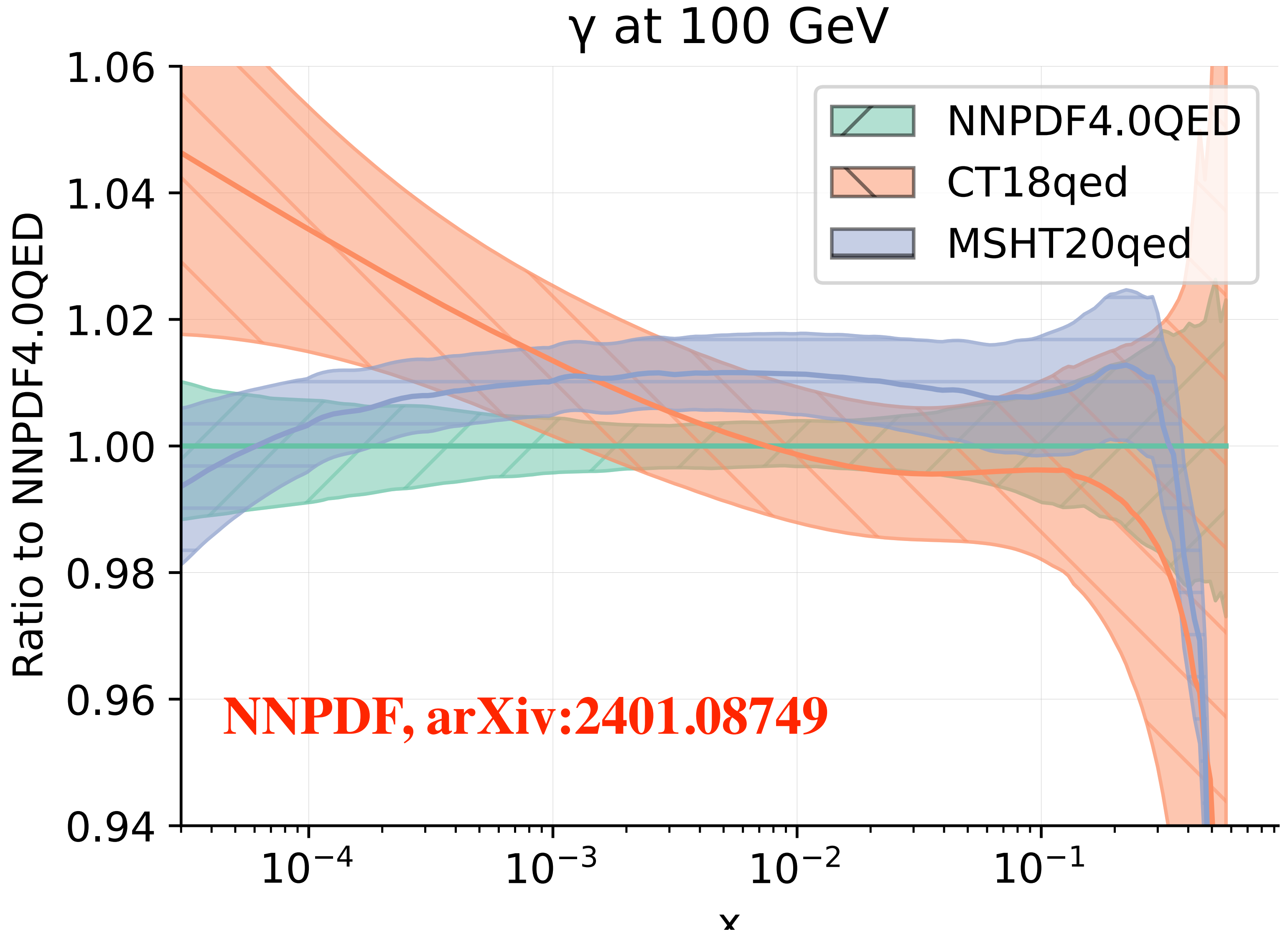


New Developments : QED corrections

- QED corrections key part of PDF fit. Requires the **photon** be included as another parton of the proton: **LUXqed** breakthrough enables high precision. Some recent highlights:

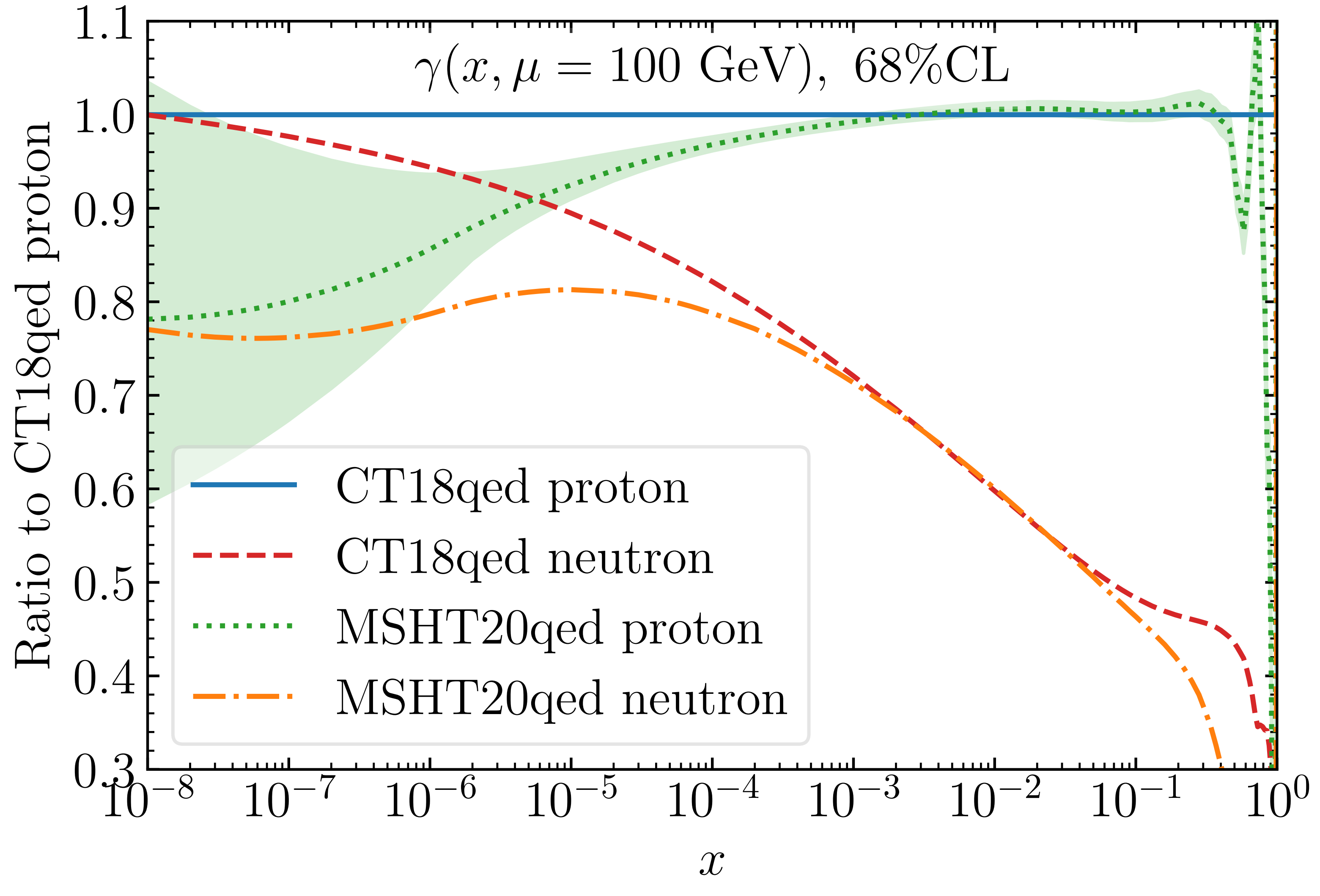
A. Manohar et al., JHEP 1712 (2017) 046

★ **NNPDF4.0QED**: latest addition to QED baseline sets



NNPDF, arXiv:2401.08749

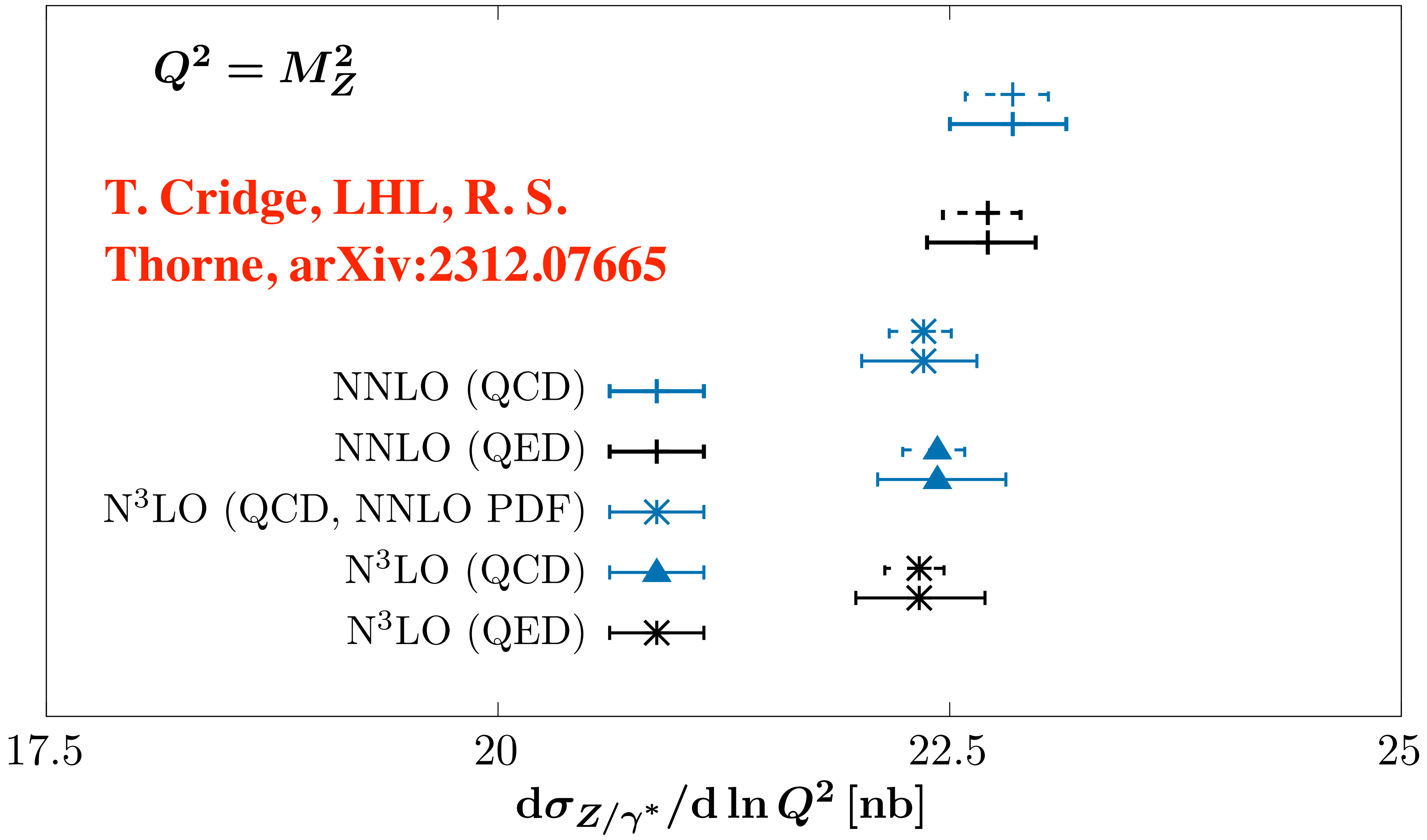
★ Not just the proton! Recent CT study on photon content of the **neutron**.



K. Xie, B. Zhou and T. Hobbs, arXiv:2305.10497

★ First combined **QED + aN3LO** set - **MSHT**. In future should/will be the standard.

★ QED on PDFs effects smaller then or ~ aN3LO. Need both!



T. Cridge, LHL, R. S. Thorne, arXiv:2312.07665

New Developments : New Physics + PDFs

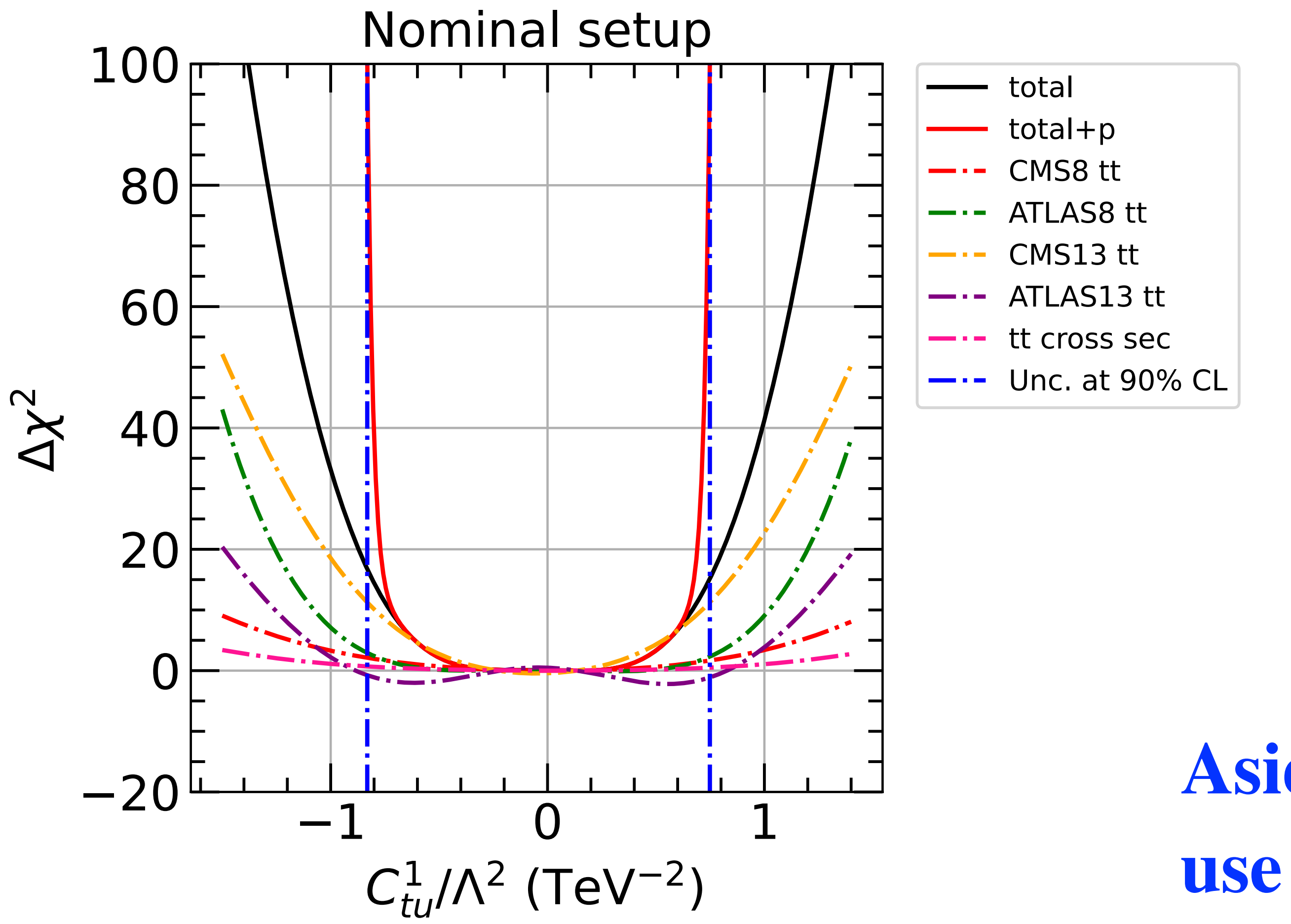
- Key element of LHC precision physics: looking for indirect signs of new physics in high energy data. Parameterise in **SMEFT**:

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \ldots,$$

- When constraining BSM with **SMEFT** fits, in principle need to account for interplay with **PDF** fit.

★ CT study - joint fit to SMEFT + PDF parameters.

★ For current LHC data PDF - SMEFT correlation small (safely fit SMEFT with fixed PDFs).



J. Gao et al., arXiv:2211.020194

Aside: both of these studies rely on use of Neural Networks!

`Physics Beyond the Standard Proton'

✦ NNPDF (**PBSP**) study: similar conclusion for current data, but what about the HL-LHC?

✦ HL-LHC pseudodata study: could new physics be absorbed in PDF fit, and if so what to do?

E. Hammou et al., arXiv:2307.10370

'Reality'

- Predictions are formed from **TRUE** PDFs, and **TRUE** New Physics parameters:

$$\sigma = \hat{\sigma}_{\rm SM+NP} \otimes f_{\rm true}$$

Result of fit

- Predictions are formed from **CONTAMINATED** PDFs, and **NO** New Physics parameters:

$$\sigma = \hat{\sigma}_{\rm SM} \otimes f_{\rm cont}$$

See also A. Anataichuk et al., arXiv:2310.19638

New Developments : New Physics + PDFs

★ HL-LHC pseudodata study: could **new physics** might be **absorbed** in PDF fit?

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \ldots,$$

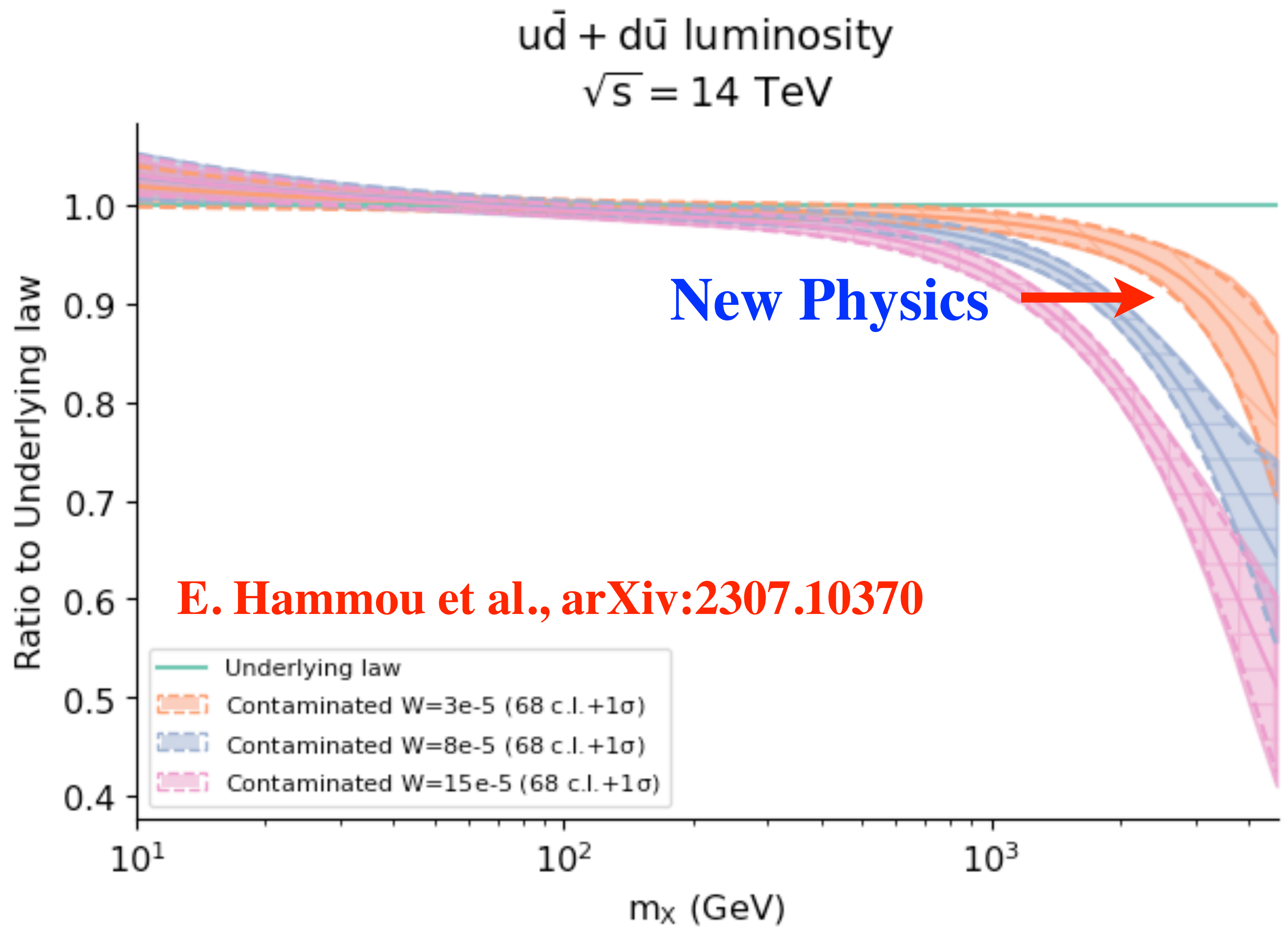
- For certain models **it can**. New physics in high mass DY absorbed into PDFs, with still reasonable fit quality.

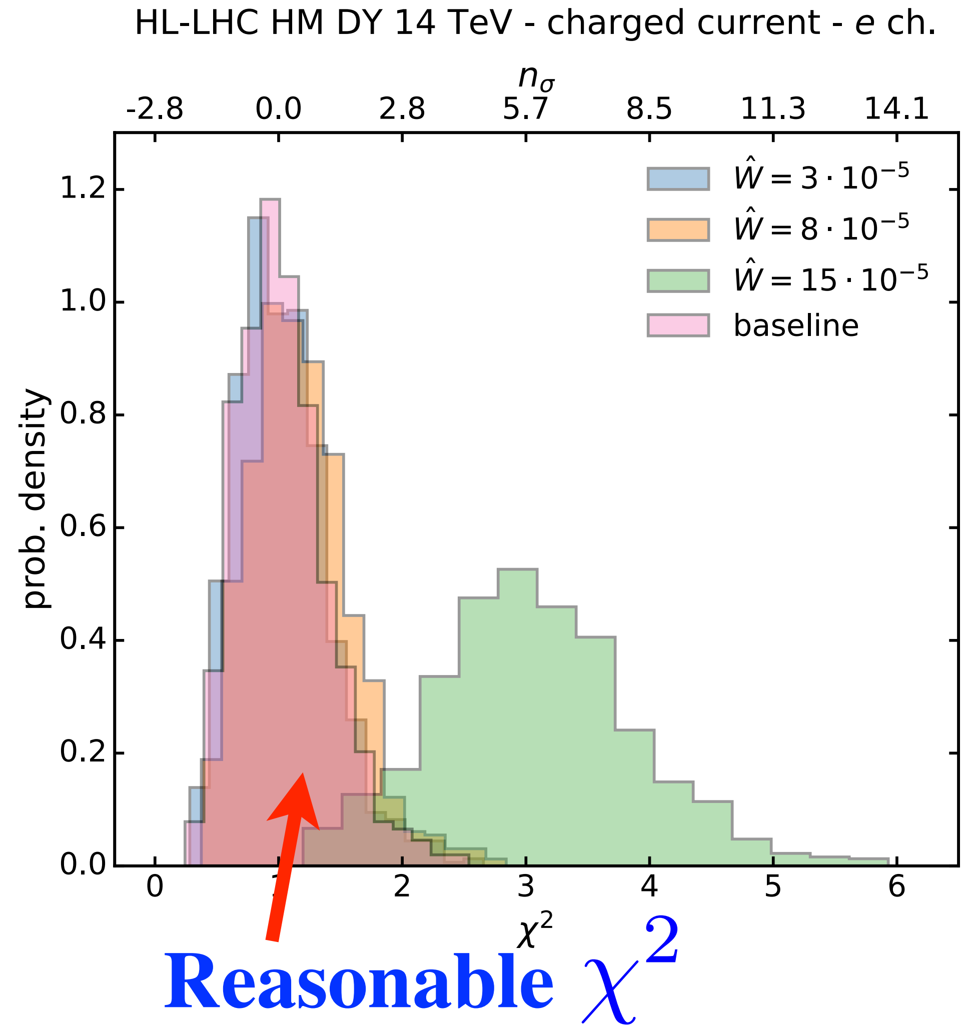
Public `SIMUnet' tool for this:

https://hep-pbsp.github.io/SIMUnet

PBSB collab., arXiv:2402.03308

- Solutions?





✦ Limit PDF fits to lower mass? **Too Naive**: if high mass deviation seen, first place would look is PDFs. Suitable choice of observables, e.g. cross section ratios better.

✦ **LHC(b)** forward data: yes, unclear for high x antiquarks $\Rightarrow$ Low energy future data (**EIC**), lattice?

✦ More broadly highlights benefit of **global fit**: different data constraints can limit above effect. Of course relies on a good understanding of PDF absent BSM…

Challenges and Methodology

New Challenges

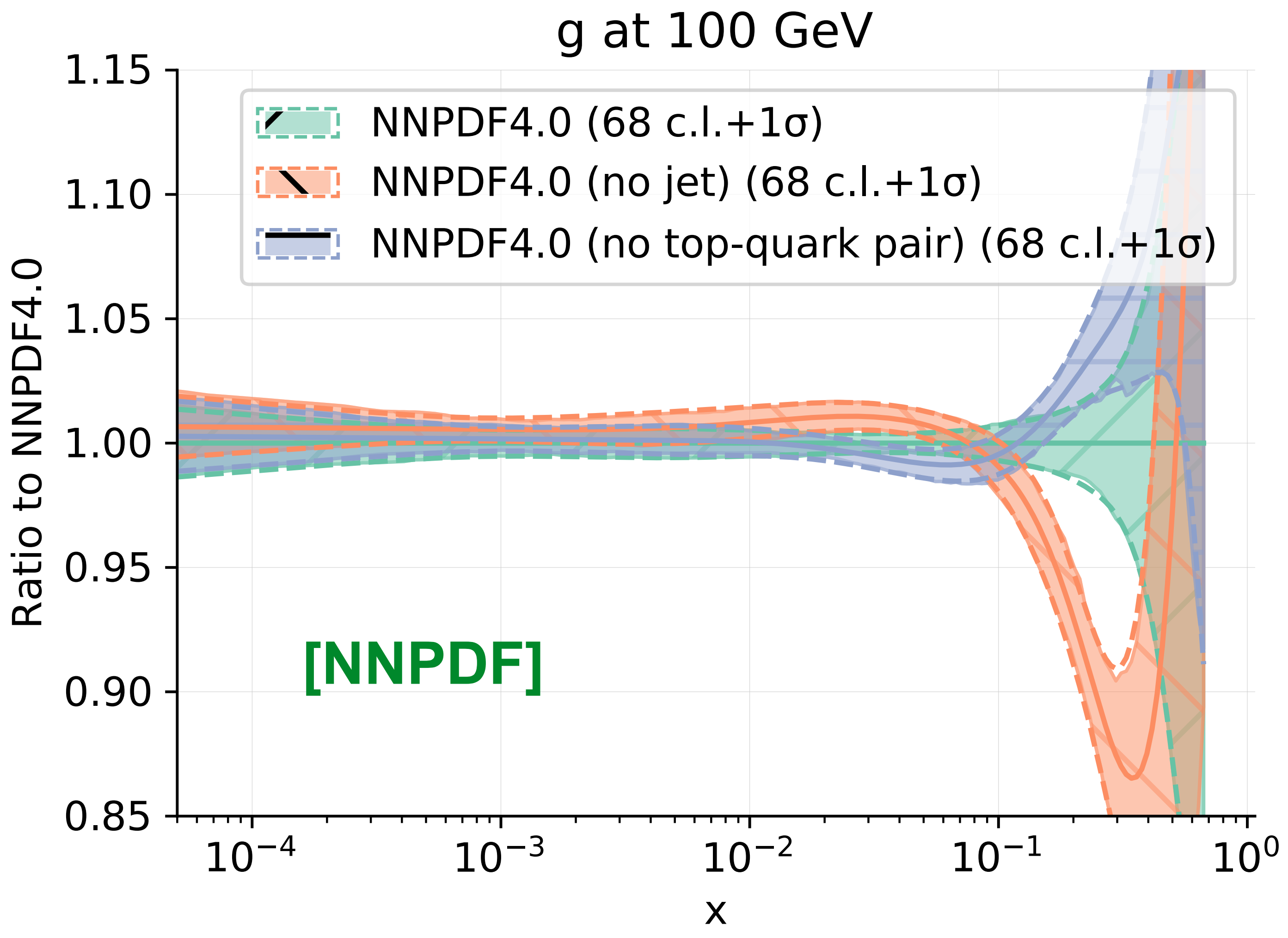
MSHTaN3LO

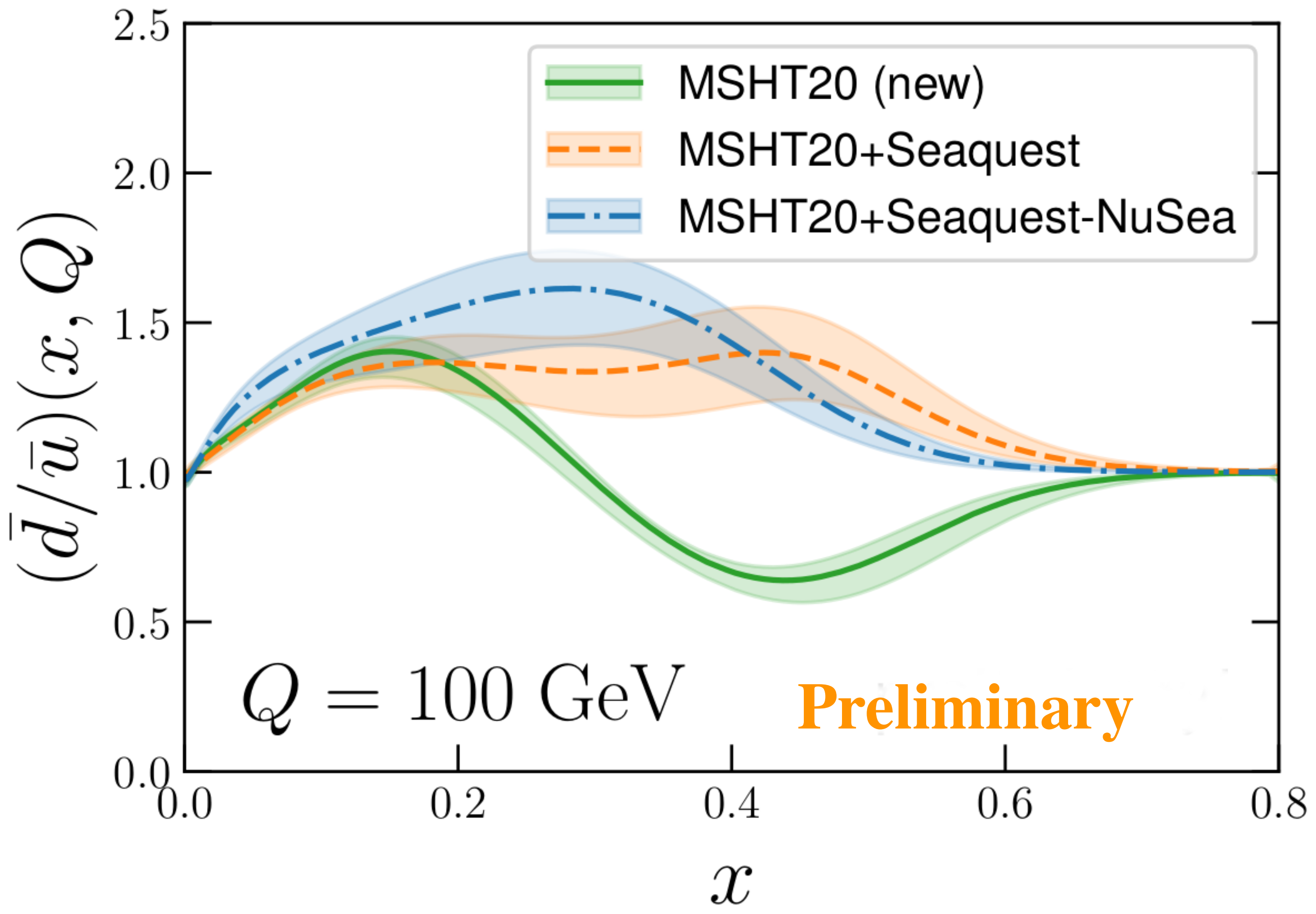
	LO	NLO	NNLO	N^3LO
$\chi^2_{N_{pts}}$	2.57	1.33	1.17	1.14

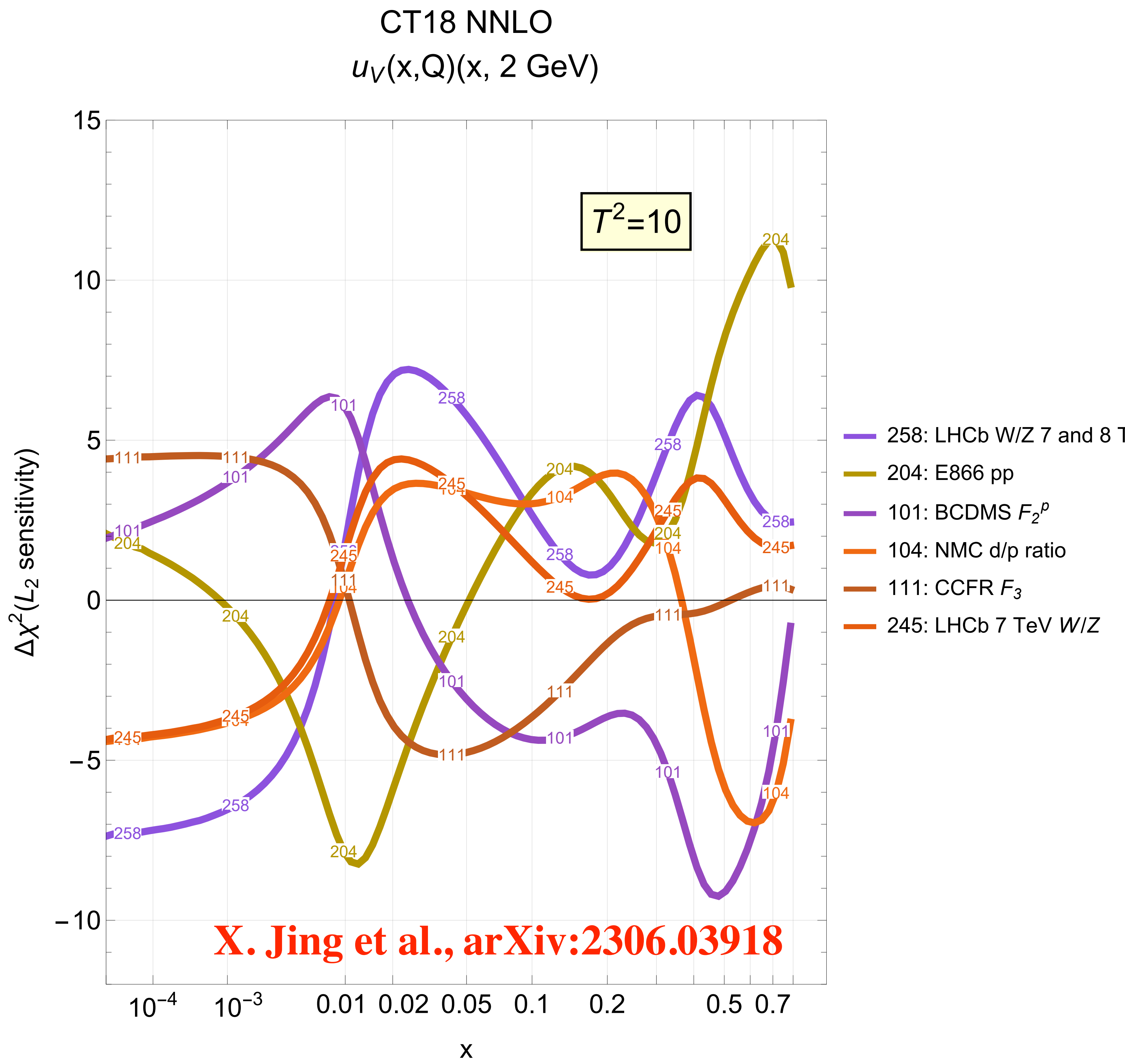
And similar for other fits!

- PDF fitting is a **challenging** environment. $\frac{\chi^2}{N_{\rm pts}} \gg 1 + \sigma(N_{\rm pts}) \sim 1.02$
  - Global PDF **fit qualities** not good by textbook definition. Many reasons for this:

★ Global fits adversarial - different datasets pull in different directions and **tensions** exist. See recent study on `L2' sensitivity



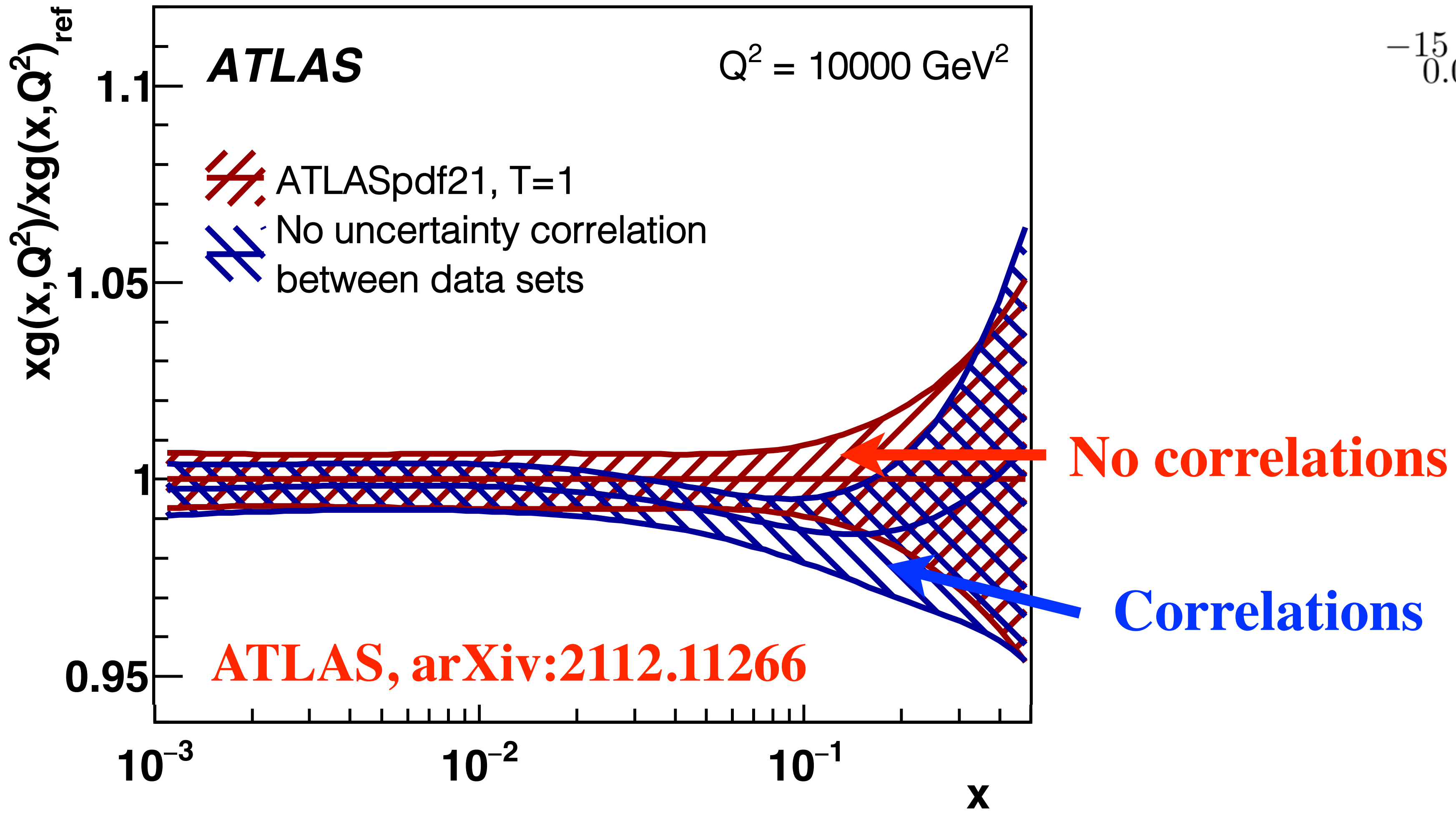




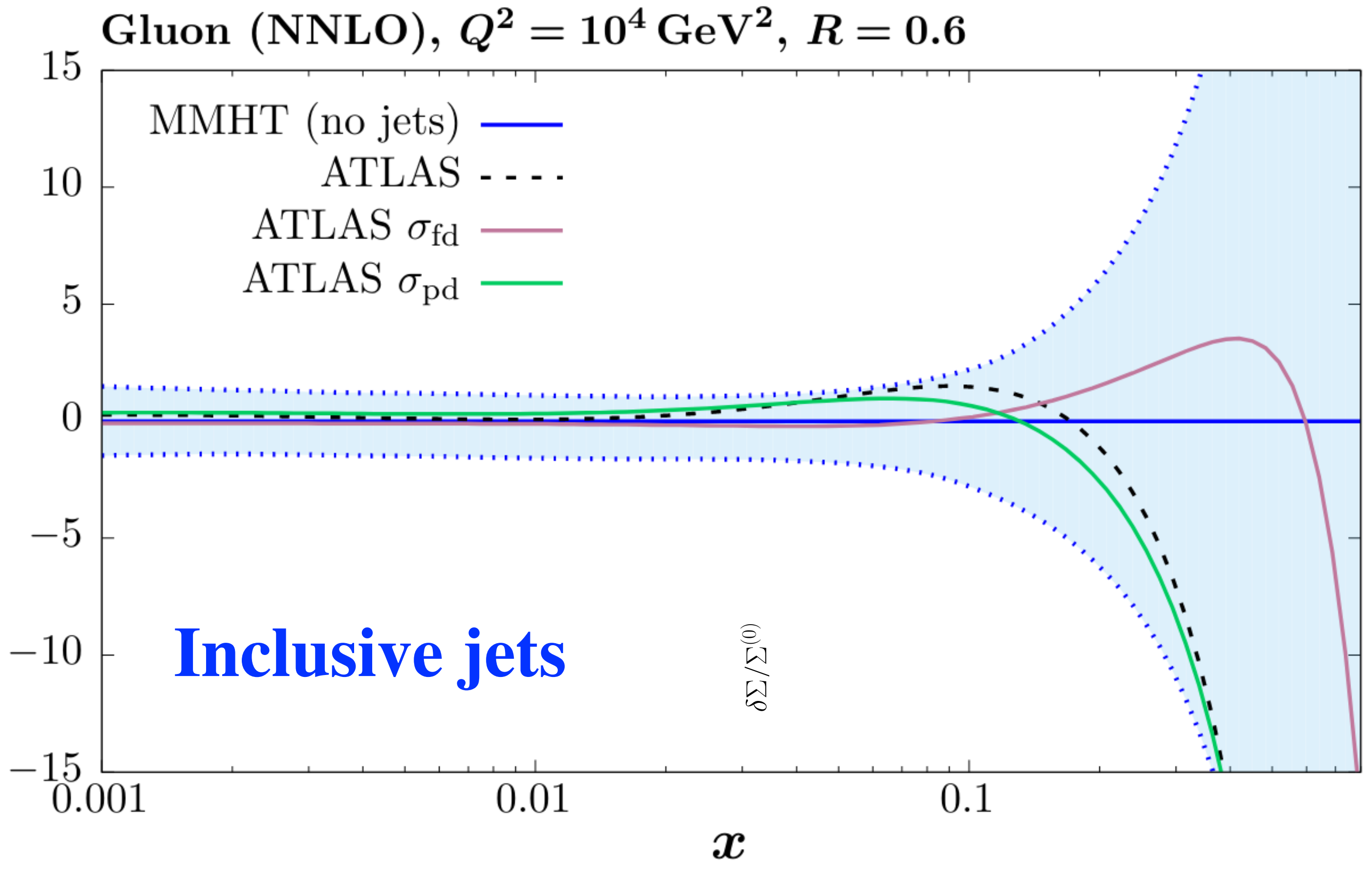
X. Jing et al., arXiv:2306.03918

★ As LHC datasets become increasingly **systematics** dominated, sensitive to precise treatment of these, and their correlations…

★ …as well as correlations between datasets!

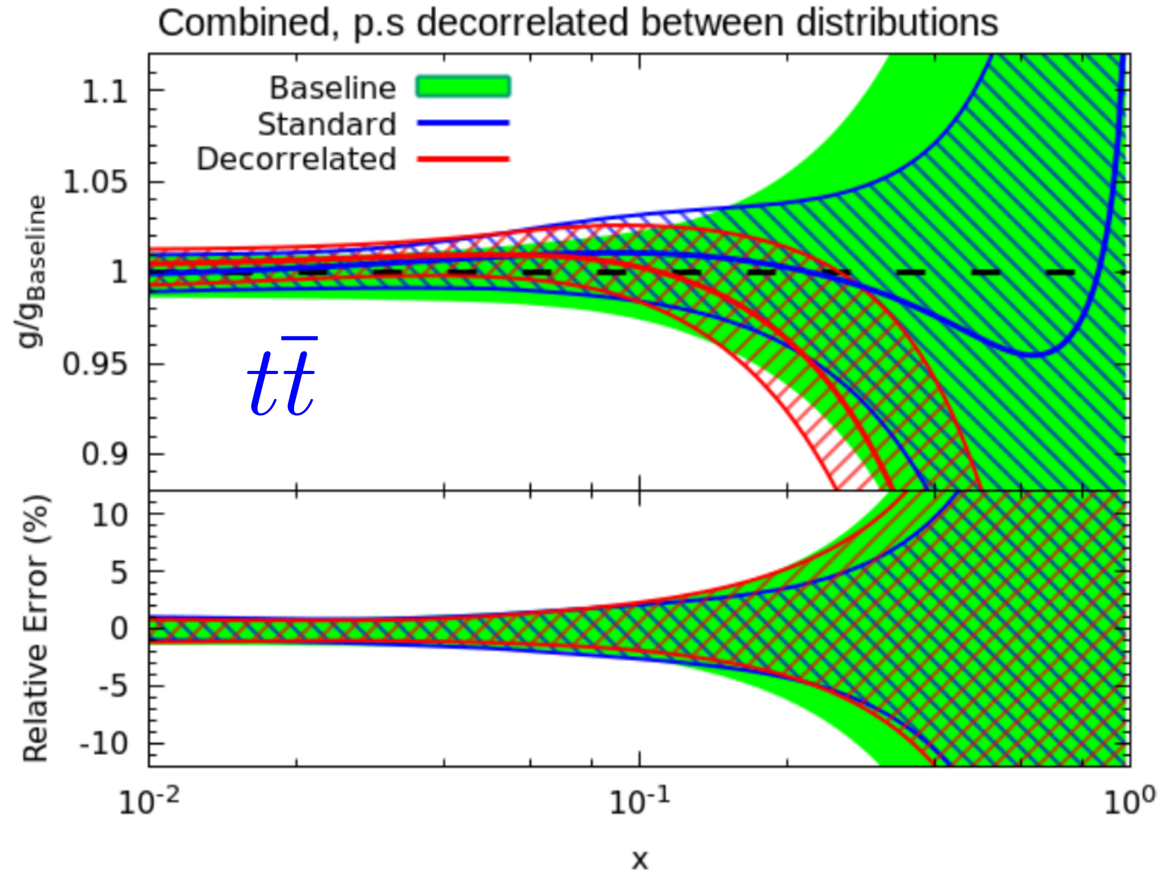


ATLAS, arXiv:2112.11266



S. Bailey et al., arXiv:2012.04684

$n_{\rm dat}$	default	part. decorr.	full decorr.
140	1.89	1.28	0.83

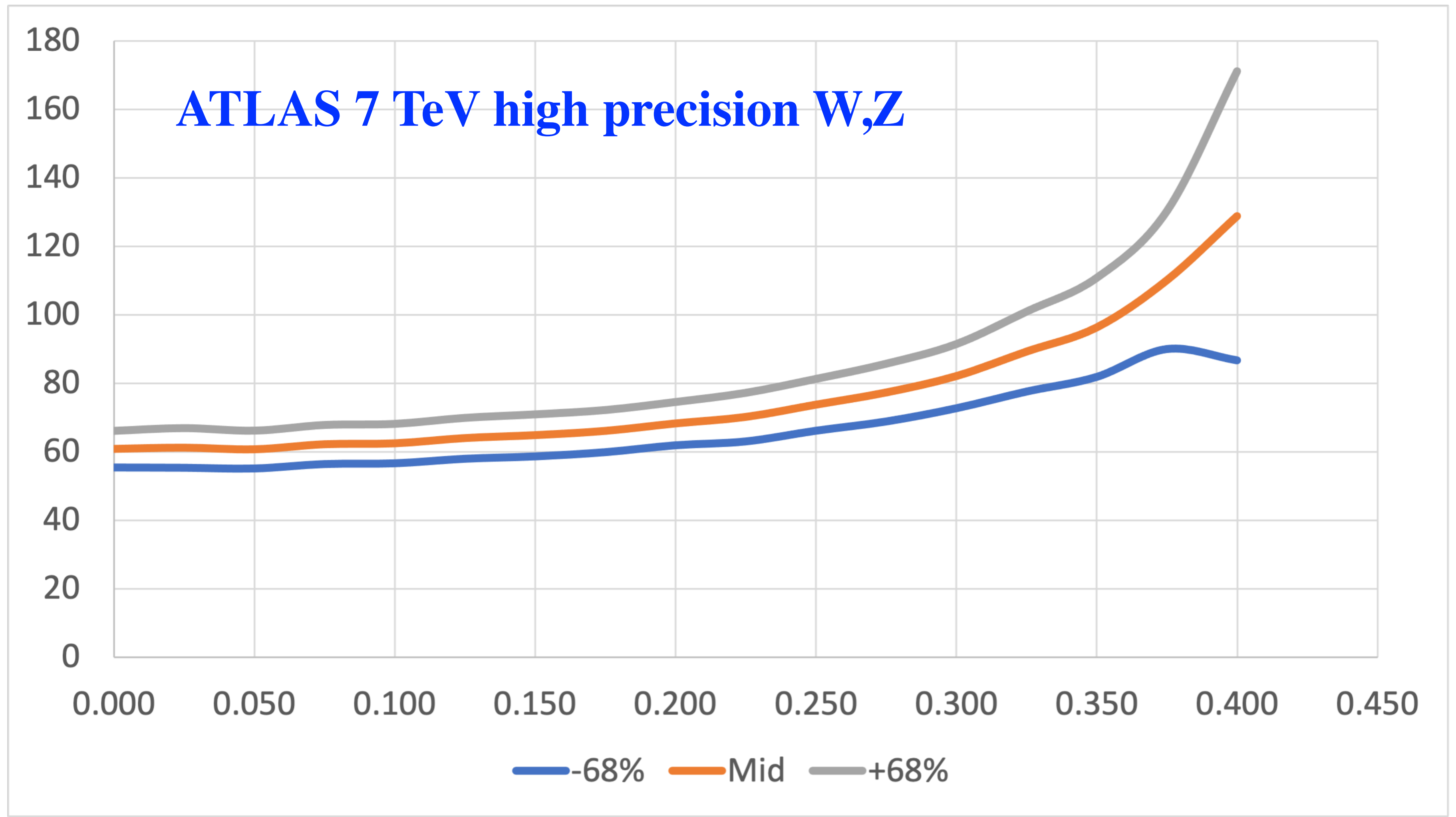


S. Bailey and LHL, arXiv:1909.10541

$n_{\rm dat}$	default	stat. uncorr.	p.s. uncorr
25	7.00	3.28	1.80

★ Could be that a more systematic approach is needed here: include errors on the errors fit quality.

★ Monitor what size of error is needed to match observed χ^2 for high precision datasets.



M. Reader et al., In Preparation

$$L(\mu,\theta,\sigma_{u_i}^2) = P(y|\mu,\theta)\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma_{u_i}}e^{-(u_i-\theta_i)^2/2\sigma_{u_i}^2}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}v_i^{\alpha_i-1}e^{-\beta_i v_i}$$

PDF Uncertainties

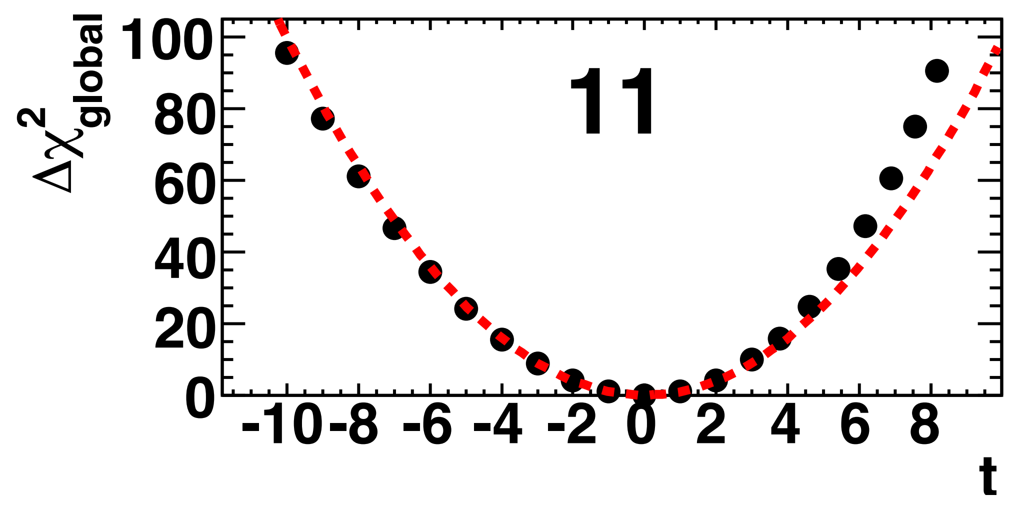
- Above issues (and others) feed into question of how we define our PDF uncertainties. All 3 groups do this differently:

★ Fixed parameterisation (MSHT/CT):

$$\chi^2_{\text{global}} \sim \frac{(D_{\text{ata}} - T_{\text{heory}})^2}{\sigma^2} \qquad H_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2_{\text{global}}}{\partial a_i \partial a_j}\Bigg|_{\text{min}}$$

✦ Find global minimum of χ^2 and evaluate eigenvectors of Hessian matrix at this point.

✦ Parameter shifts corresponding to given $\Delta\chi^2$ criteria given in terms of these



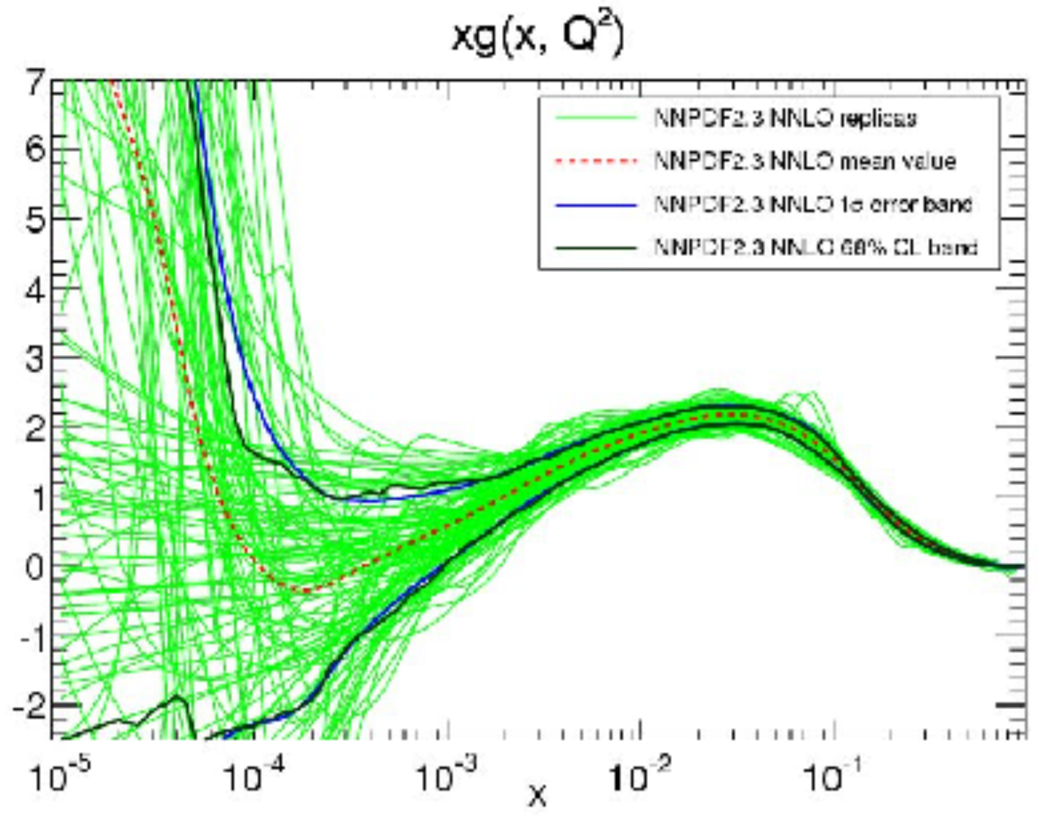
$$a_i(S_k^{\pm}) = a_i^0 \pm t\, e_{ik}, \qquad \text{with } t \text{ adjusted to give desired } T = \Delta\chi^2_{\text{global}}$$

✦ Procedure for choosing this 'tolerance' T differs between MSHT/CT.

★ NNPDF:

✦ Generate set of MC `replicas' by shifting data by errors.

Each D_i gives f_i and from $\{f_i\} \Rightarrow$ PDF errors



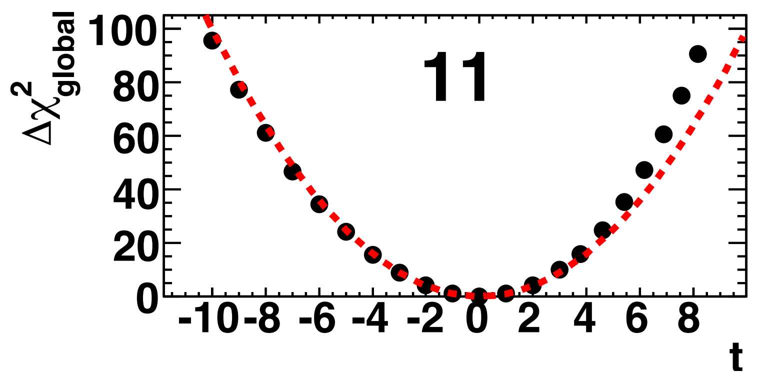
G. Watt and R. Thorne, arXiv:1205.4024

✦ Note not specific to NNs: can apply in fixed parameterisation as well: shown to be ~ equivalent to Hessian $\Delta\chi^2 = 1$ in that case.

✦ However, in NN approach direct correspondence is lost as Hessian approach does not apply.

- Why introduce a tolerance?

$$a_i(S_k^{\pm}) = a_i^0 \pm t\, e_{ik}, \qquad \text{with } t \text{ adjusted to give desired } T = \Delta\chi^2_{\rm global}$$

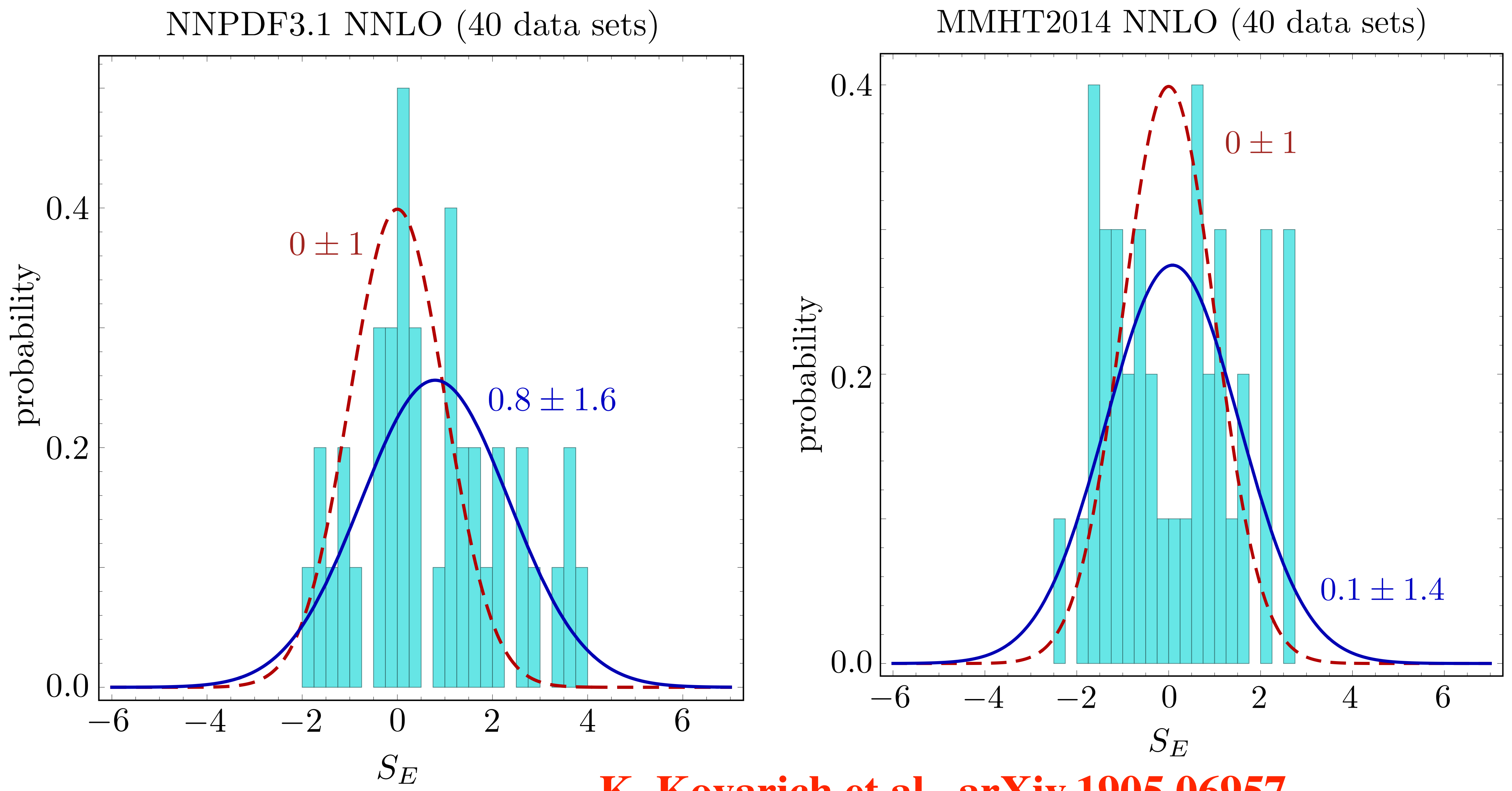


Fixed target, DIS, Tevatron, LHC

$N_{\rm dataset} \sim 50-60$

- $T = 1$: `textbook' criterion for 68% C.L., would apply if:
  - ⋆ Complete statistical compatibility between multiple datasets entering fit.
  - ⋆ Completely faithful evaluation of experimental uncertainties within each dataset.
  - ⋆ Theoretical calculations that match these exactly.

G. Watt and R. Thorne, arXiv:1205.4024 M. Yan et al., arXiv.2406.01664
J. Pumplin, arXiv:0909.0268



K. Kovarich et al., arXiv.1905.06957

$N_{\rm pts} \sim 4000-5000$

- Good evidence that first two points do not always hold, while last point known not be true (though progress towards missing higher order uncertainties made).
- Applying textbook tolerance to global dataset with tensions does not lead to automatically larger errors.
- Given complete statistical compatibility, global PDF fit very constraining. Danger is claimed (high) precision will increasingly not match accuracy with $T = 1$. Motivates enlarged tolerance $T > 1$.

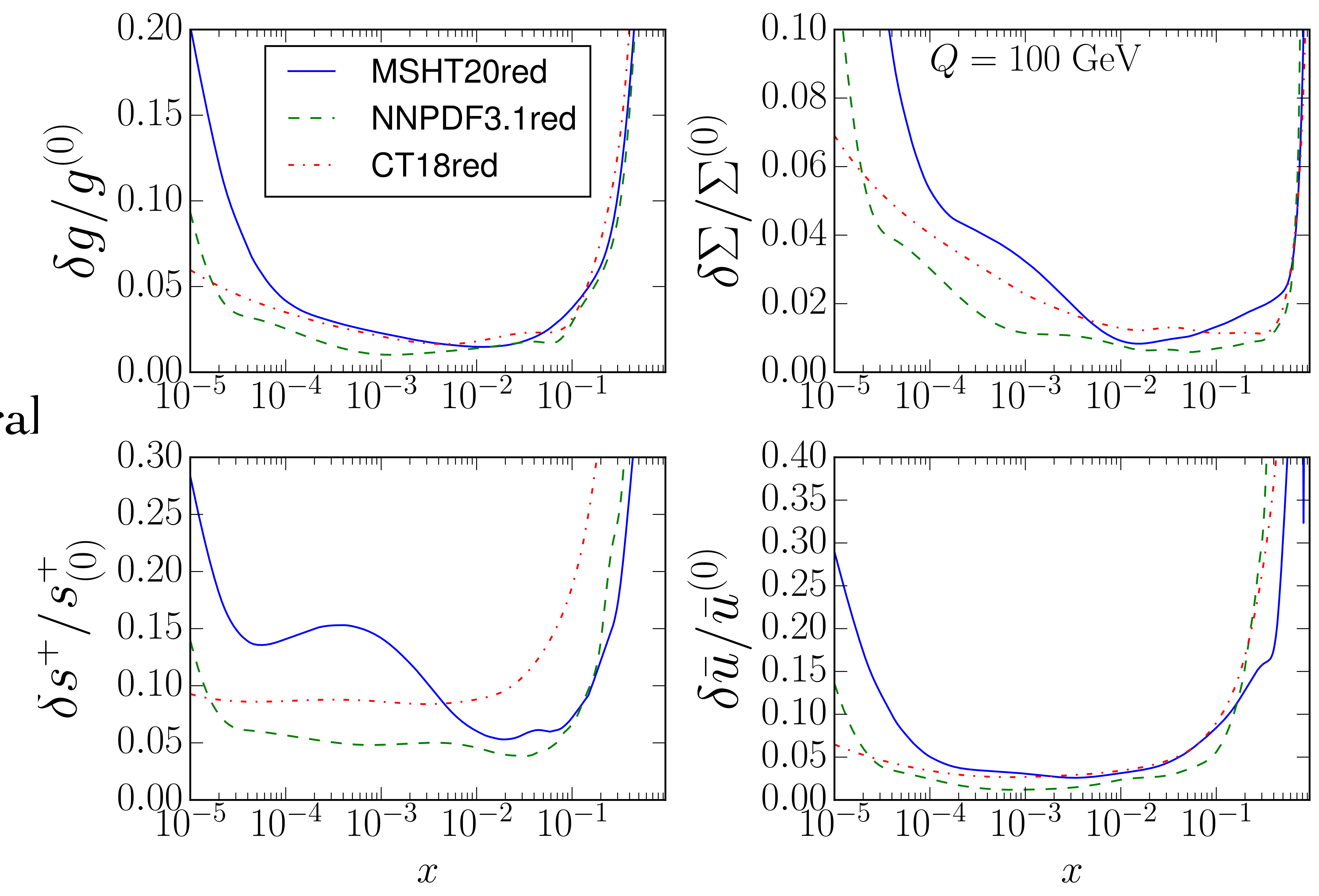
Impact on PDFs

PDF4LHC21, arXiv:2203.05506

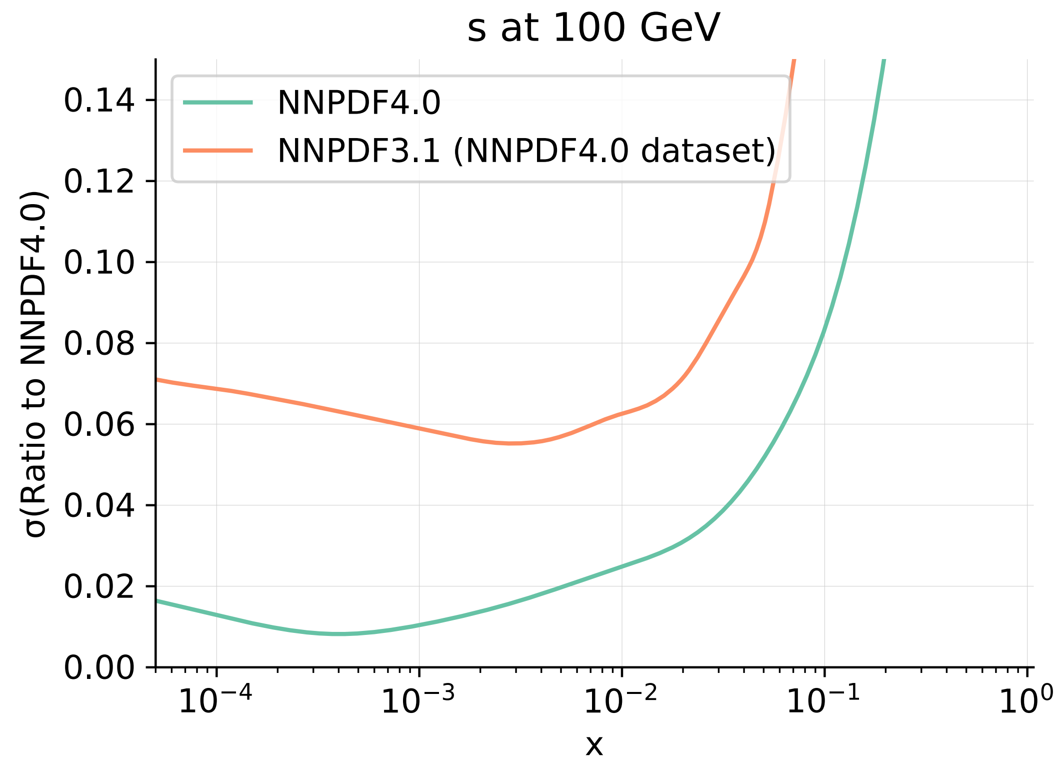
- Comparison/benchmarking of PDFs considered in PDF4LHC21 study. Fit representative subset of global dataset, unified between 3 global fits and with very close theory settings. Find:

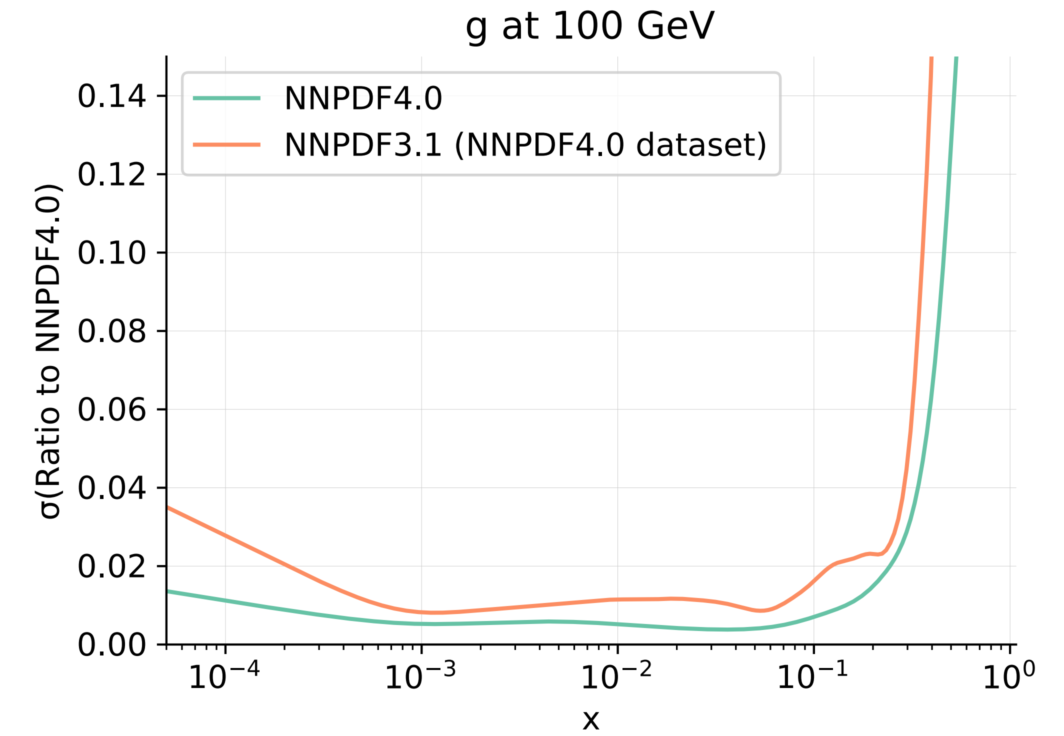
✦ Global fits give different errors in PDF4LHC21 benchmarking. NNPDF3.1 in general **smaller errors.**

Benchmark = similar data/settings



✦ And **4.0 methodology** gives further errors reduction.

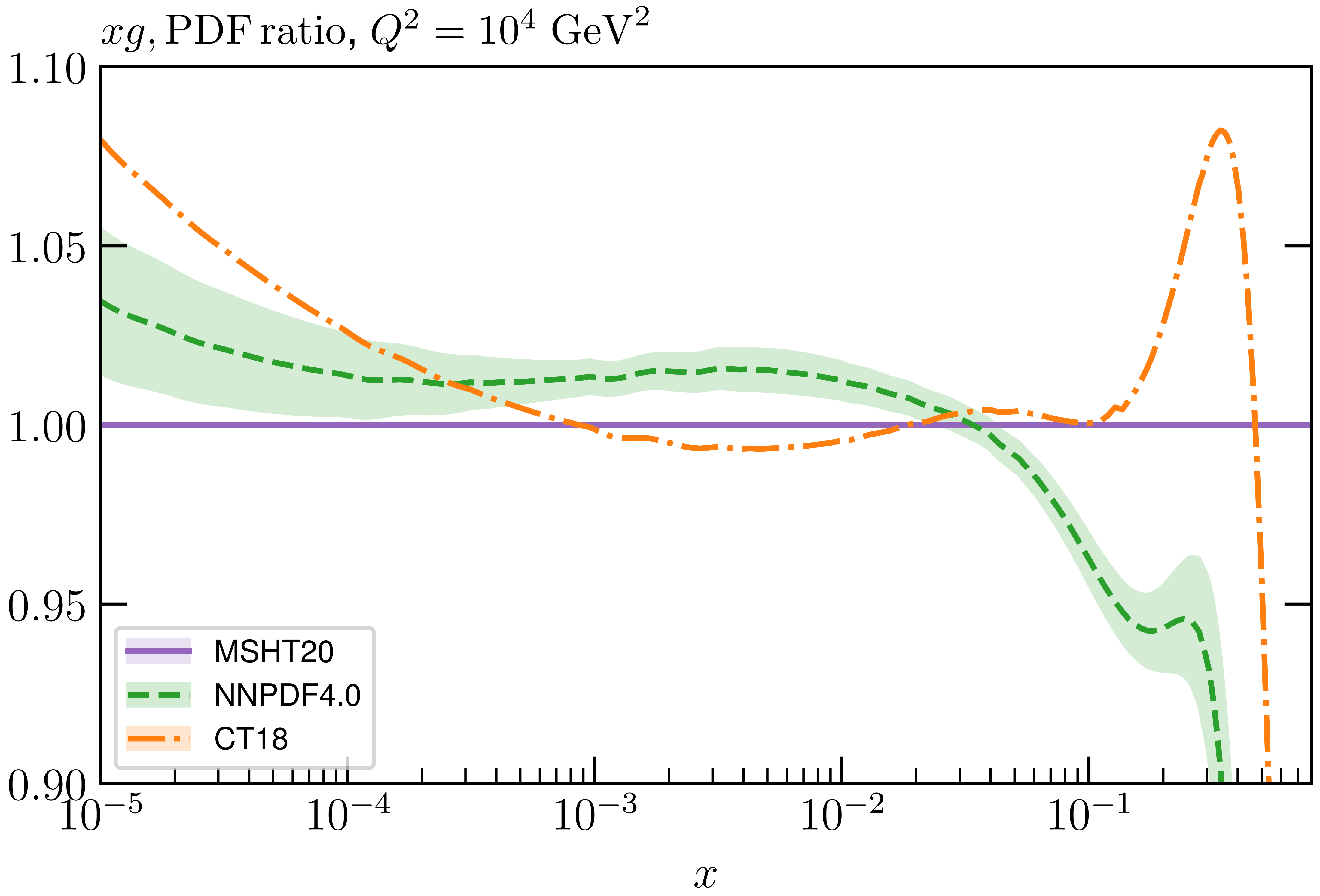


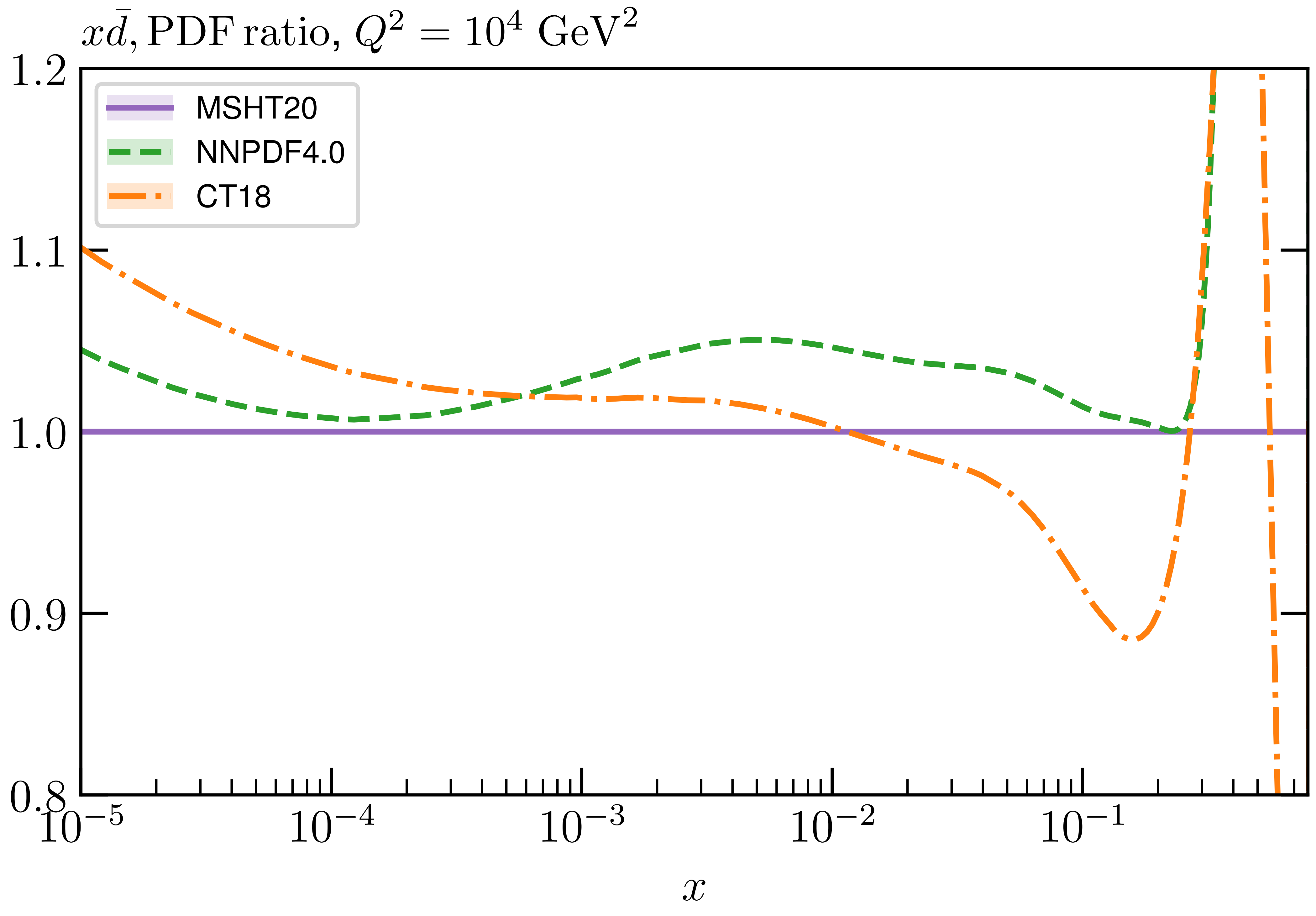


- Different methodologies giving different results.

Where do we stand?

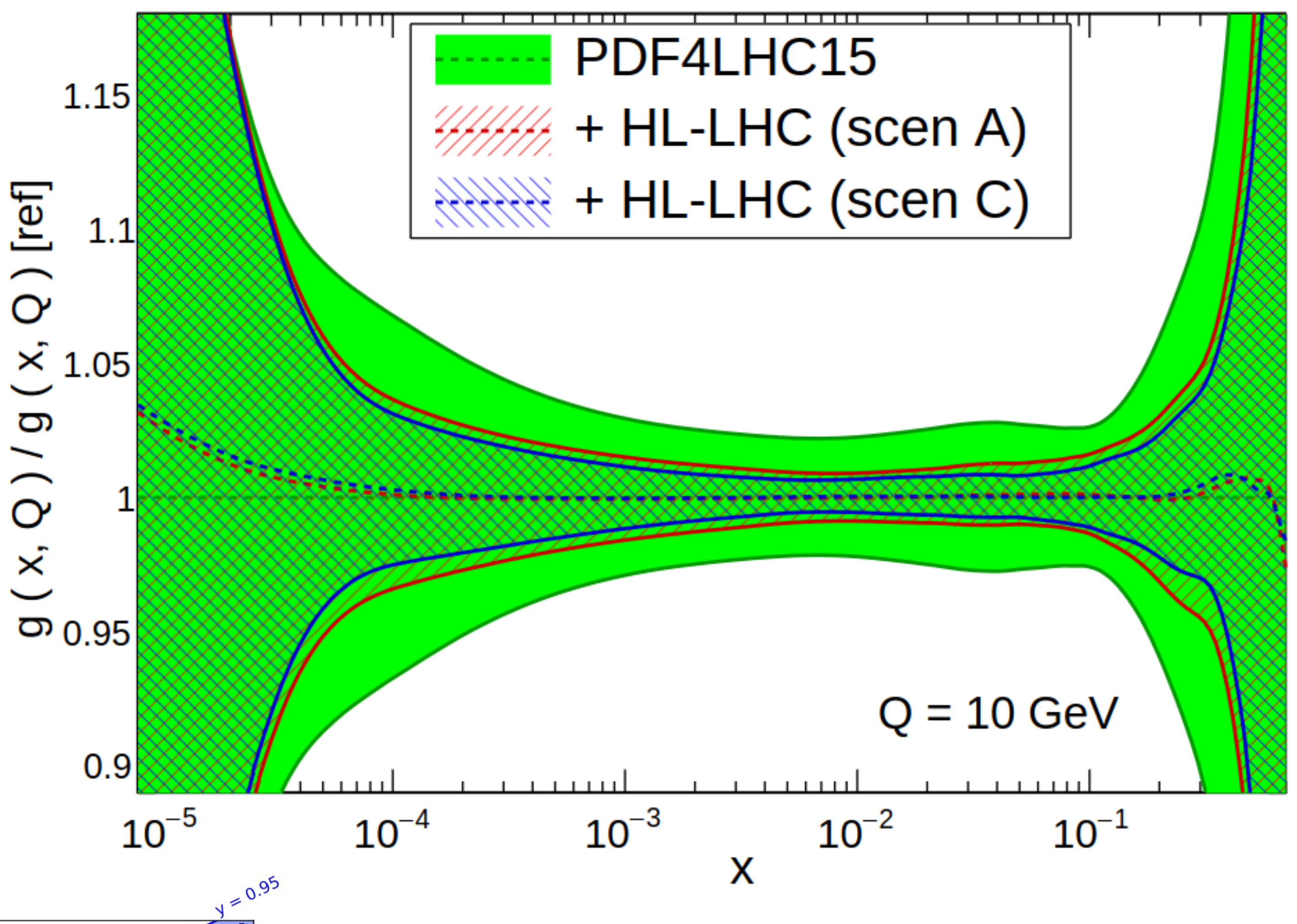
- New (LHC) data in fits has clear impact - PDF uncertainties continue to reduce - and theoretical precision continues increase, up to now aN3LO order.
- However, global PDF fits are complex: agreement between sets not always good at level of PDFs or uncertainties. Evidence that methodology can be just as important as data in fits.

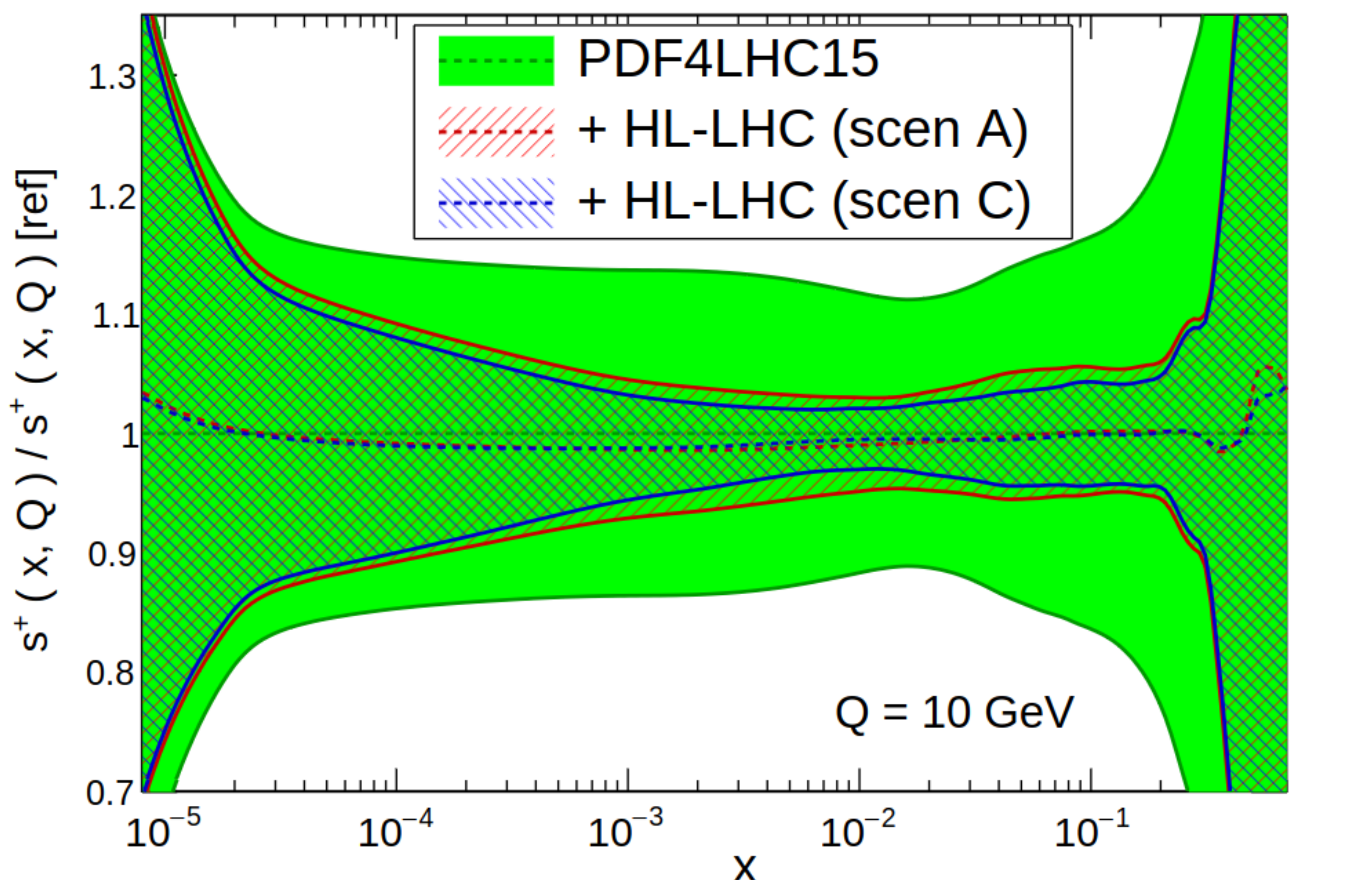




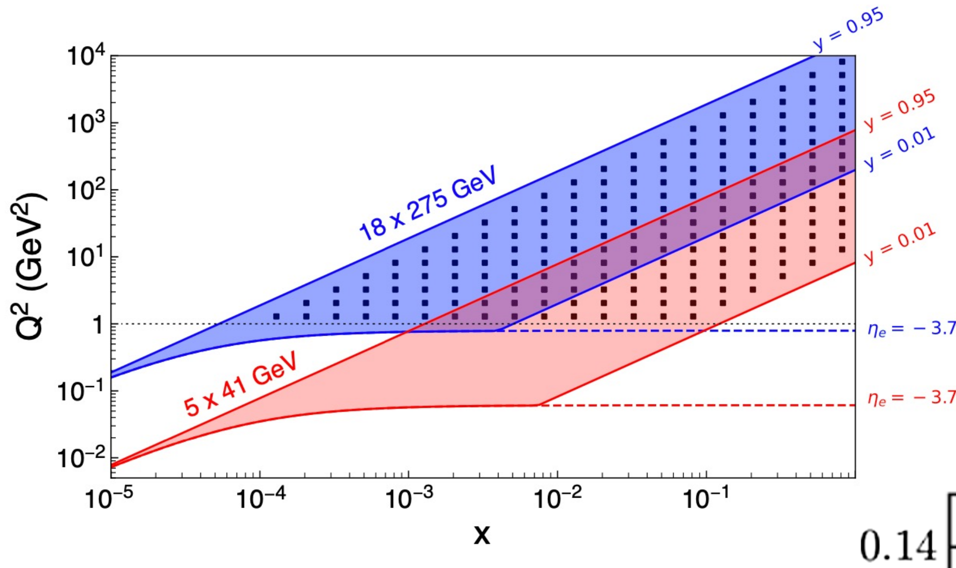
Where might we go? Future Data

- LHC continuing to have an impact, and **HL-LHC** projected to beyond that…
- …but these are only projections. Reality usually more complicated. Other experiments/colliders providing complementary information will be key.



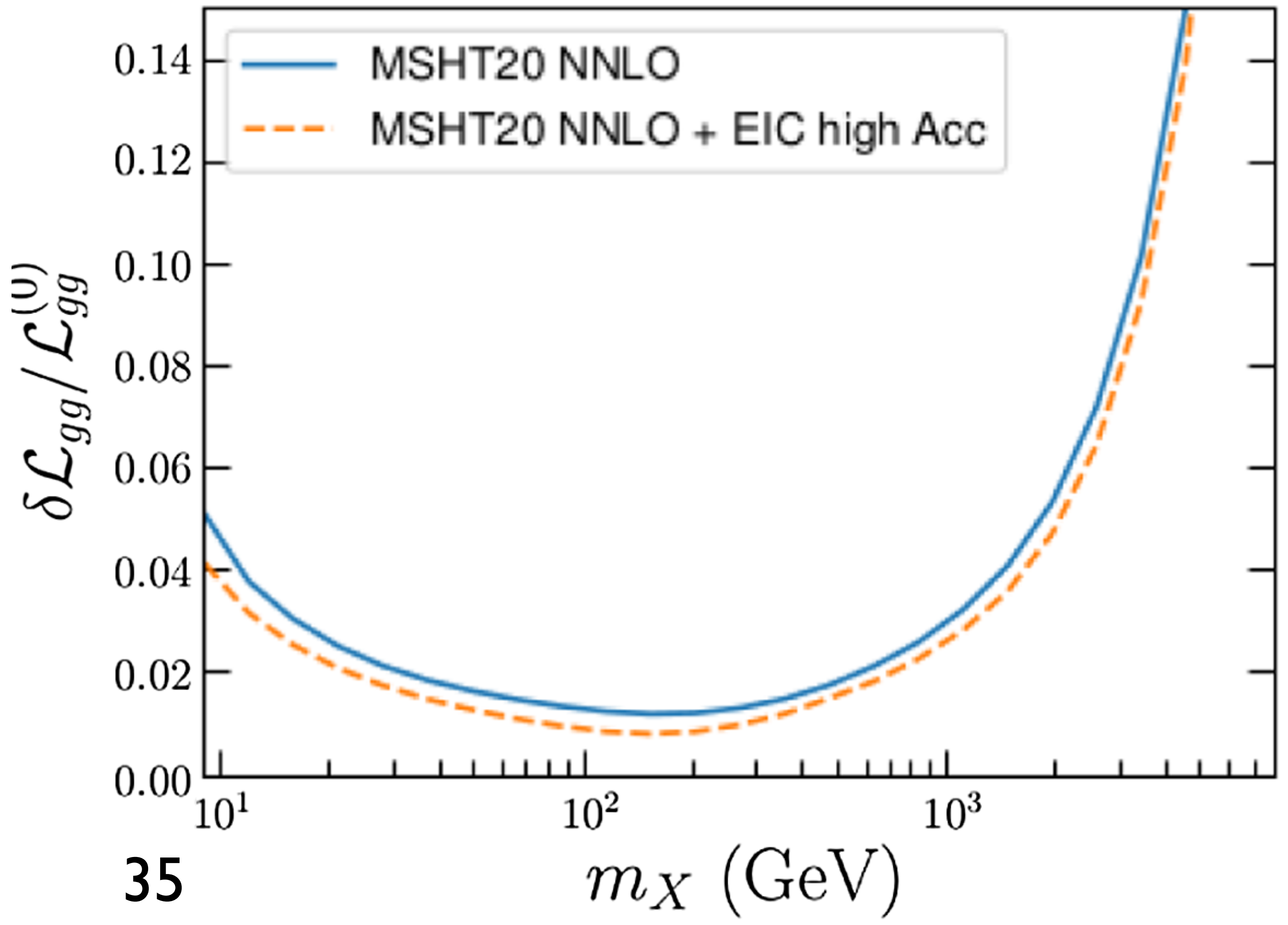


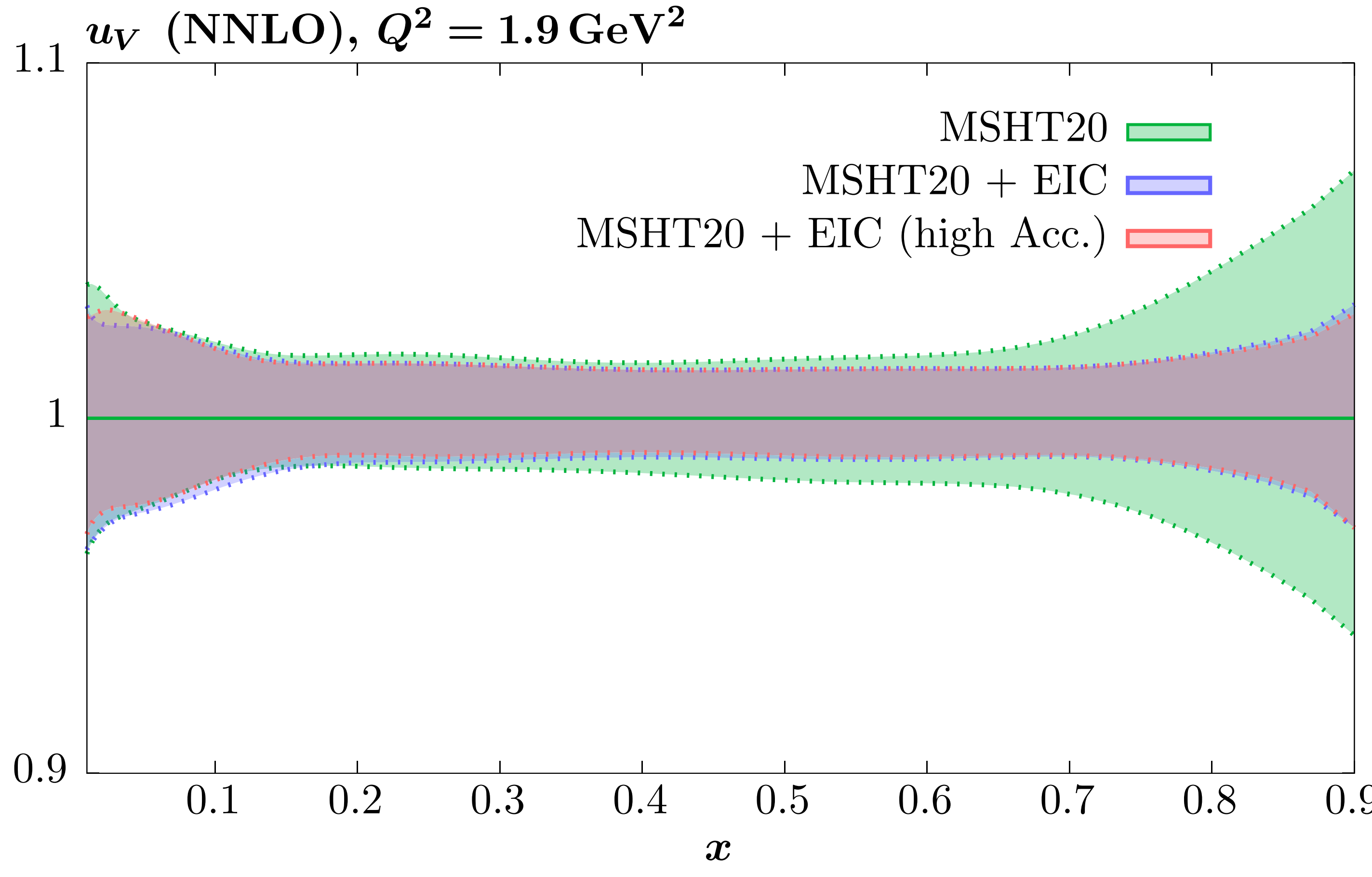
- Amongst many things the **EIC** will give us are better constraints on high x PDFs.



- Expected impact in global PDF context moderate **but** complementarity is key. See BSM studies - what if see a disagreement in high energy data?

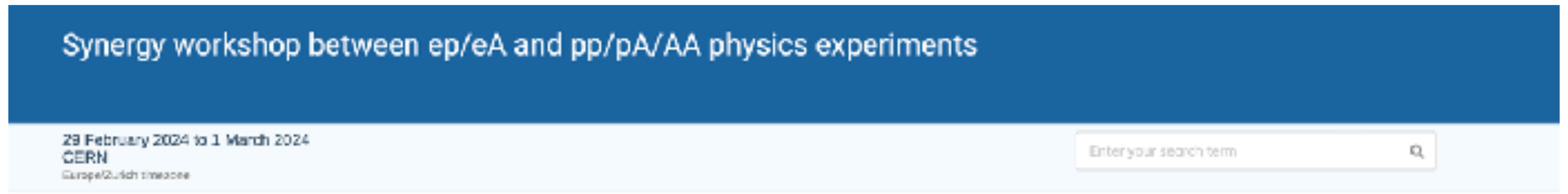
N. Armesto et al., arXiv:2309.11269



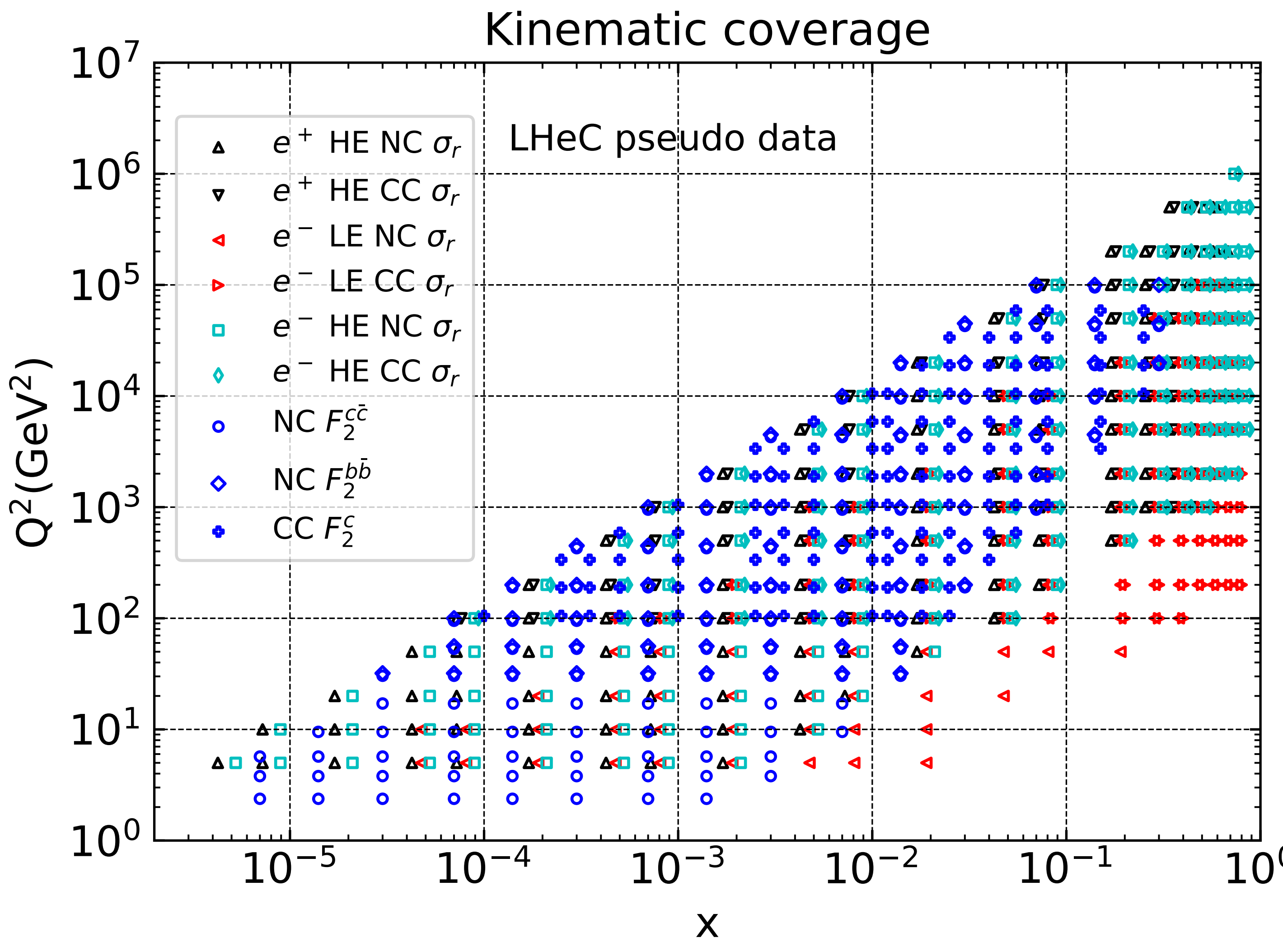


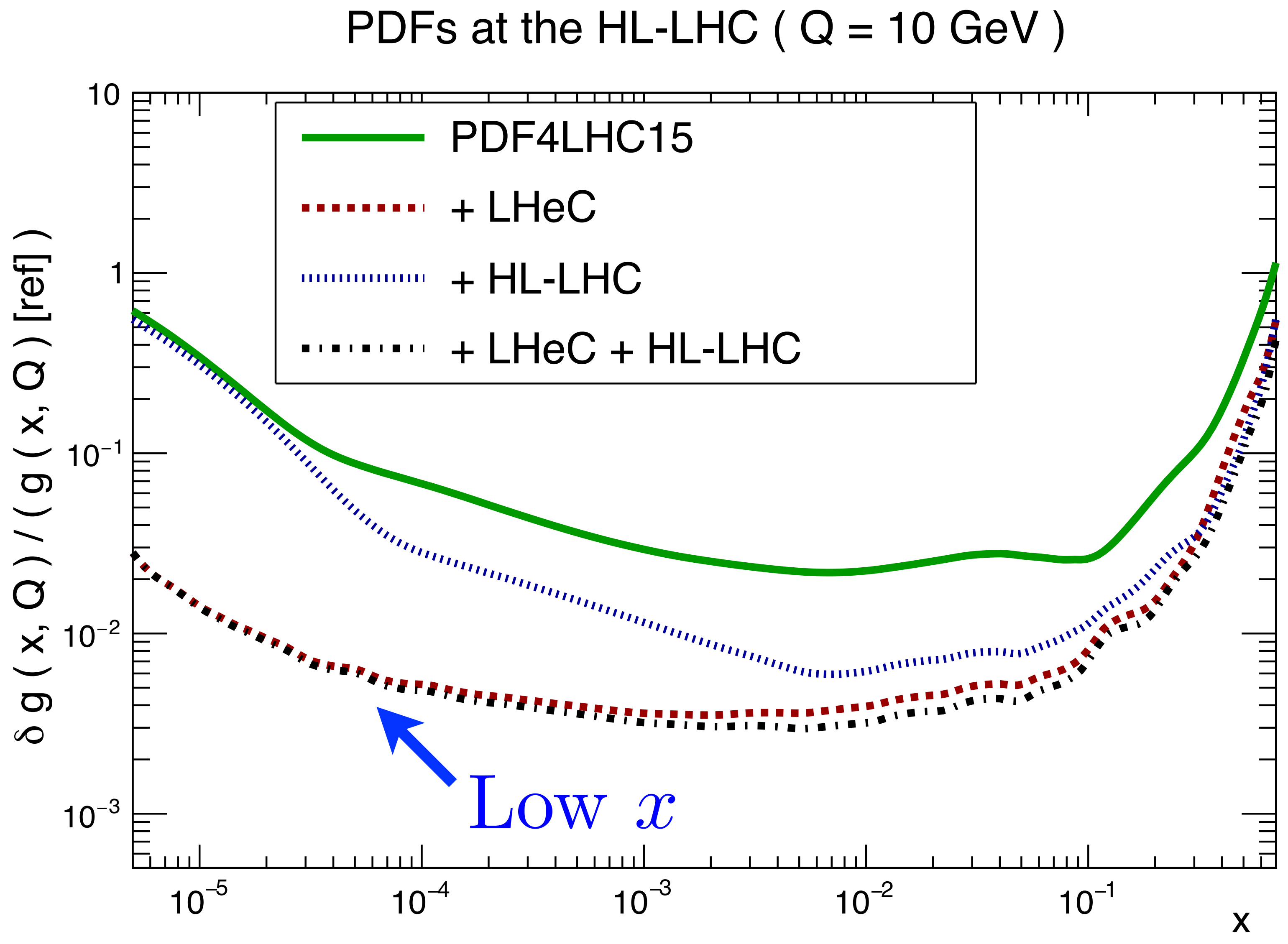
Future Data?

https://indico.cern.ch/event/1367865/overview

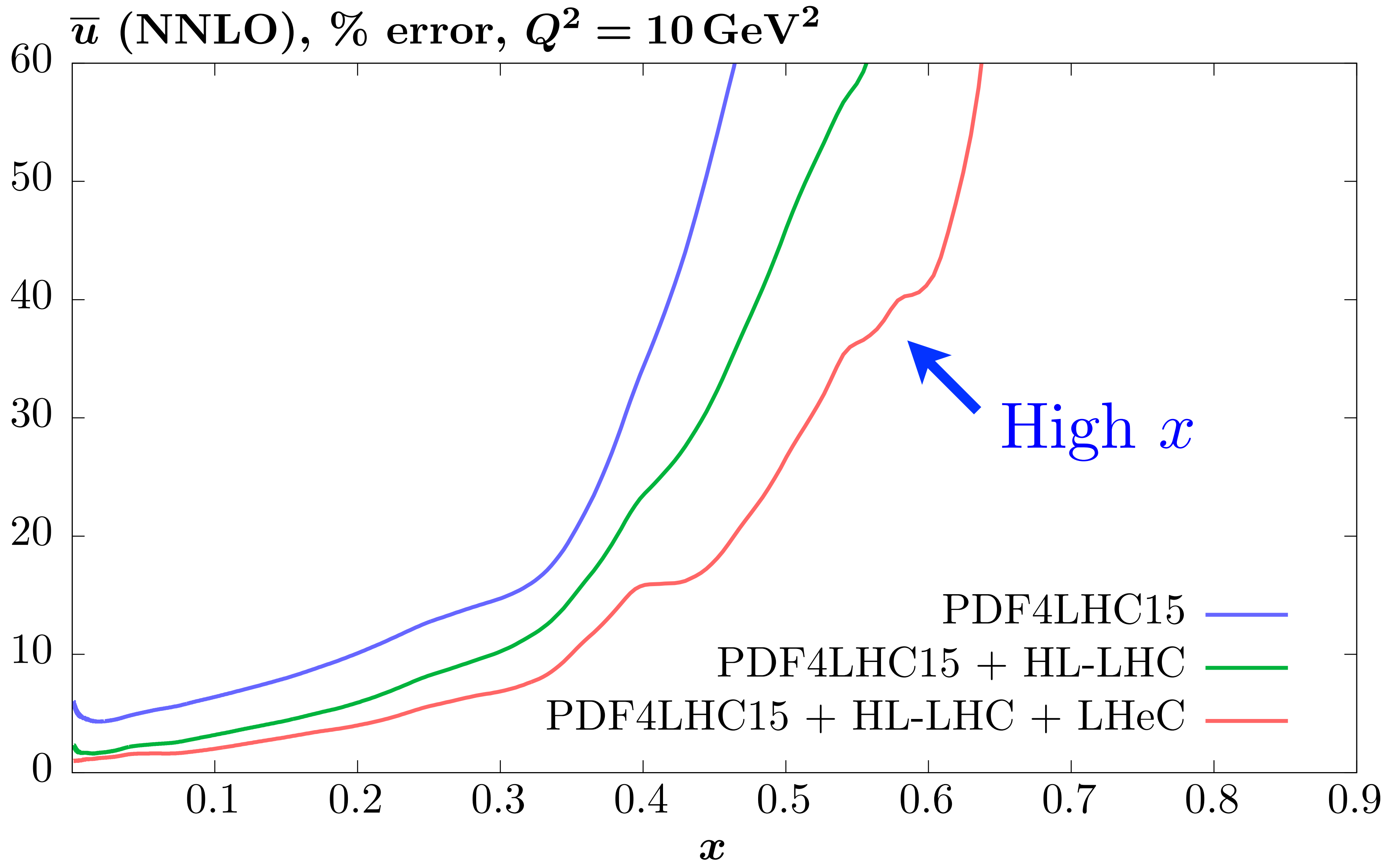


- In this context **LHeC** proposal also very advantageous.
- Clean and complementary ep data over wide region of phase space, with impressive PDF projections.



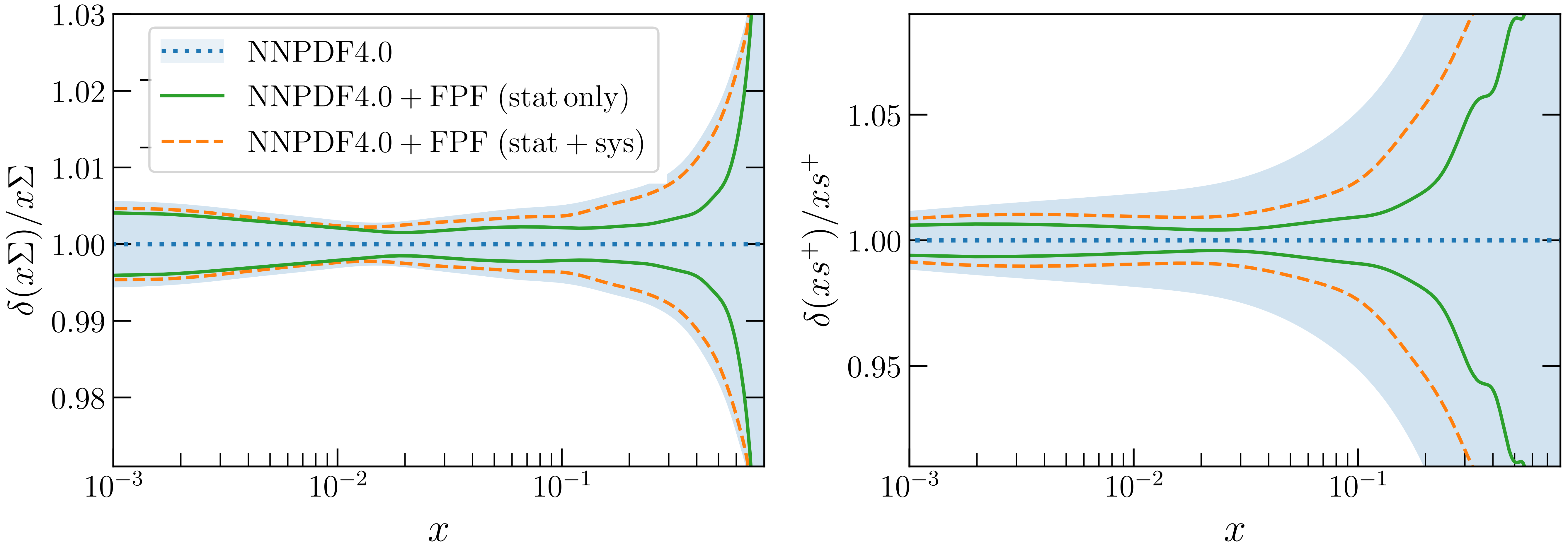


R. Abdul Khalek et al., arXiv:1906.10127

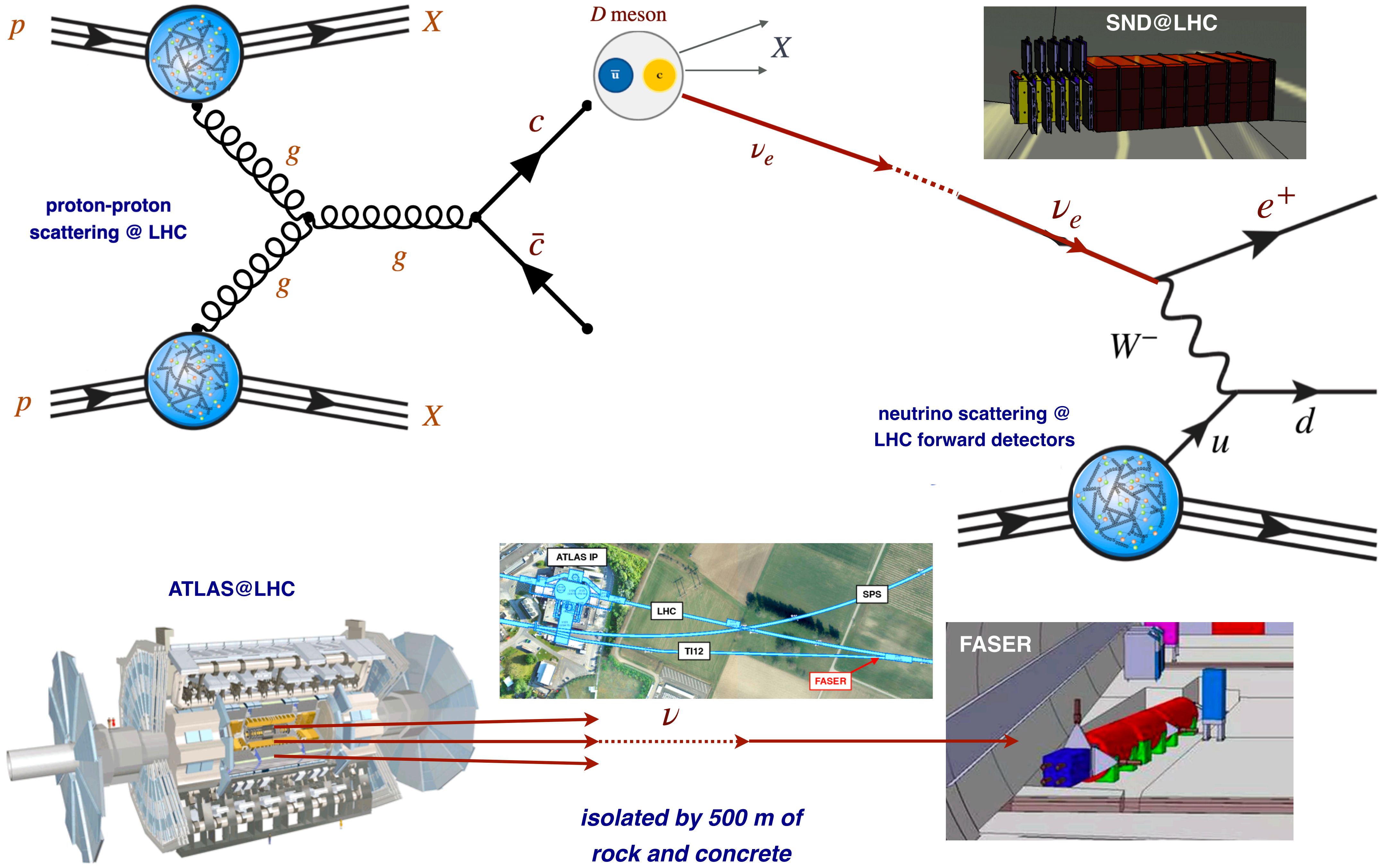


Future Data?

- **FPF** proposal to extend far forward detectors also shows promising potential for high energy (TeV) neutrino-induced DIS data at the LHC.
- Promising projected constraints on quark flavour structure.



J. Cruz-Martinez et al., arXiv:2309.09581

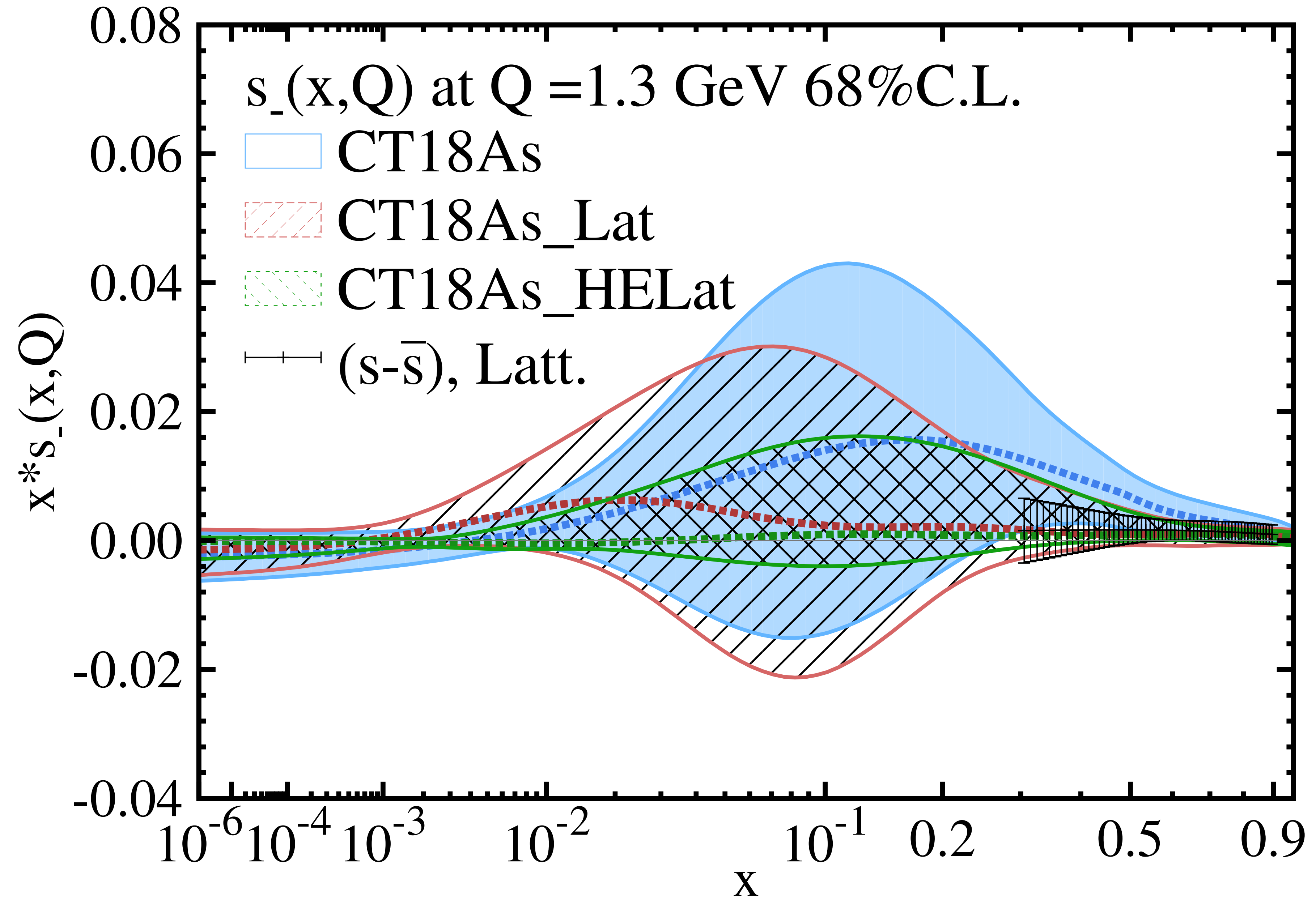


J. Rojo, PDF4LHC23

Impact of Lattice?

- Recent studies from CT and NNPDF groups.

★ CT: Impact of quasi-PDF lattice data on strangeness asymmetry promising at high x ?

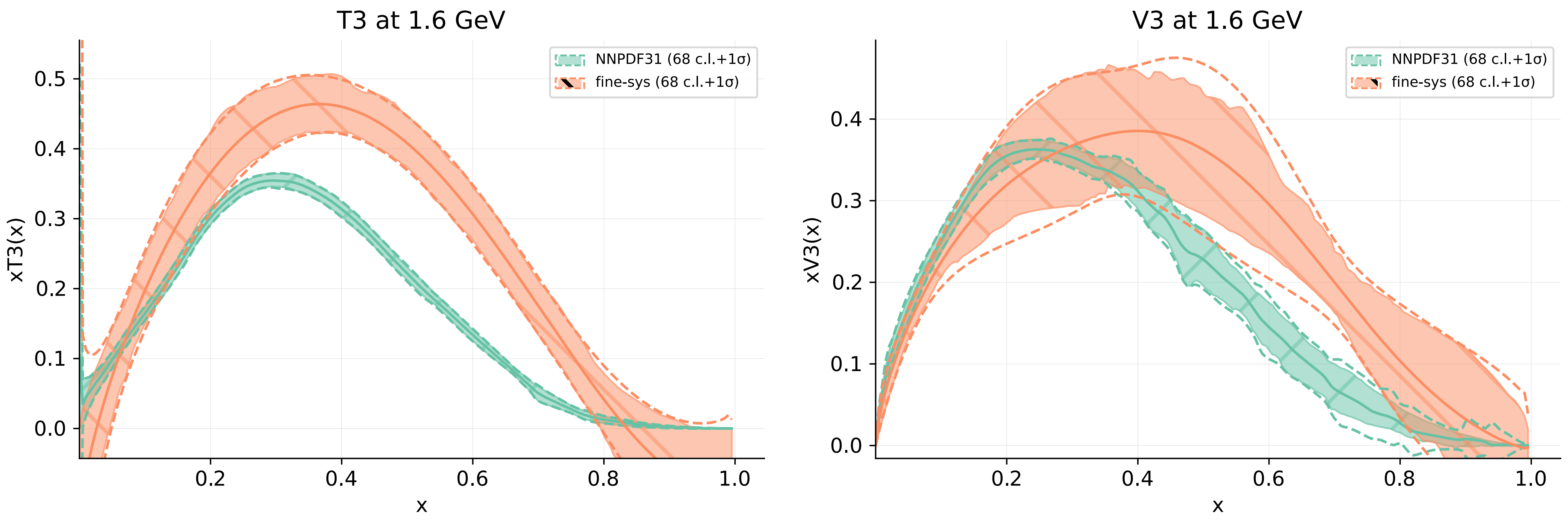


T-J Hou et al., arXiv:2211.11064

K. Cichy, L. Del Debbio and T. Giani, arXiv:1907.06037
L. Del Debbio et al., arXiv:2010.03996

$$T3 = (u + \overline{u}) - (d + \overline{d})$$

$$V3 = (u - \overline{u}) - (d - \overline{d})$$



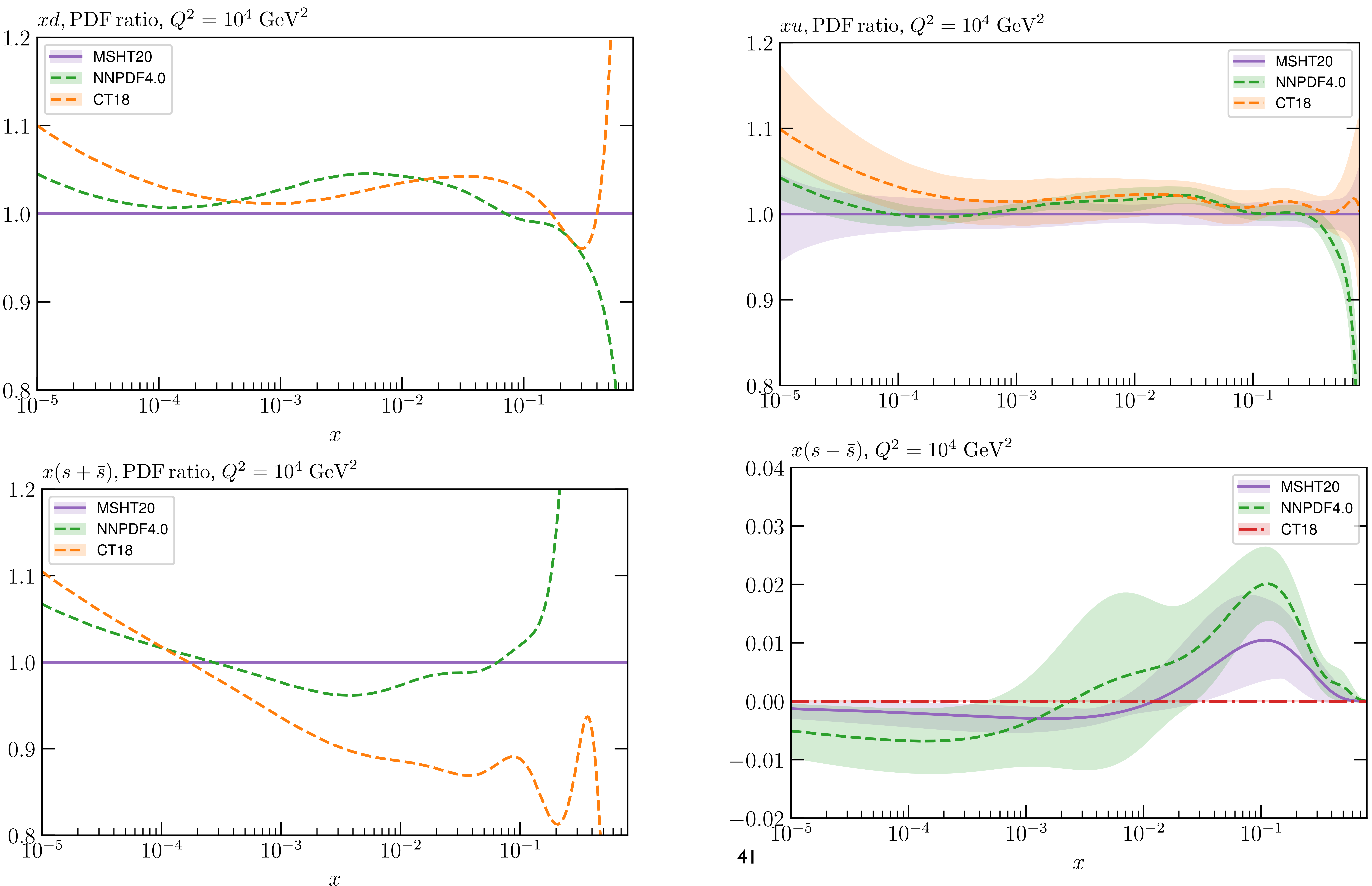
★ NNPDF: looking at well determined non-singlet quantities some further room for improvement?

Conclusions

- ⋆ Parton Distribution Functions a key input in the precision physics programme of LHC and beyond.
- ⋆ Precise and accurate PDF determination crucial. Global PDF fits currently the best way to achieve this.
- ⋆ A significant deal of experimental and theoretical progress: high precision LHC data driving PDF fits, and up to (approximate) N3LO will be the standard (+ NLO EW) for theory.
- ⋆ But path to achieving accuracy and precision is not an easy one: non-negligible differences between latest PDF fits. Clear understanding of uncertainties and comparison of methodologies essential.
- ⋆ One takeaway: not simply a question of looking at nominal precision of given PDF set to assess potential impact of lattice. May also help disentangle above differences? Complementarity is key!

Thank you for listening!

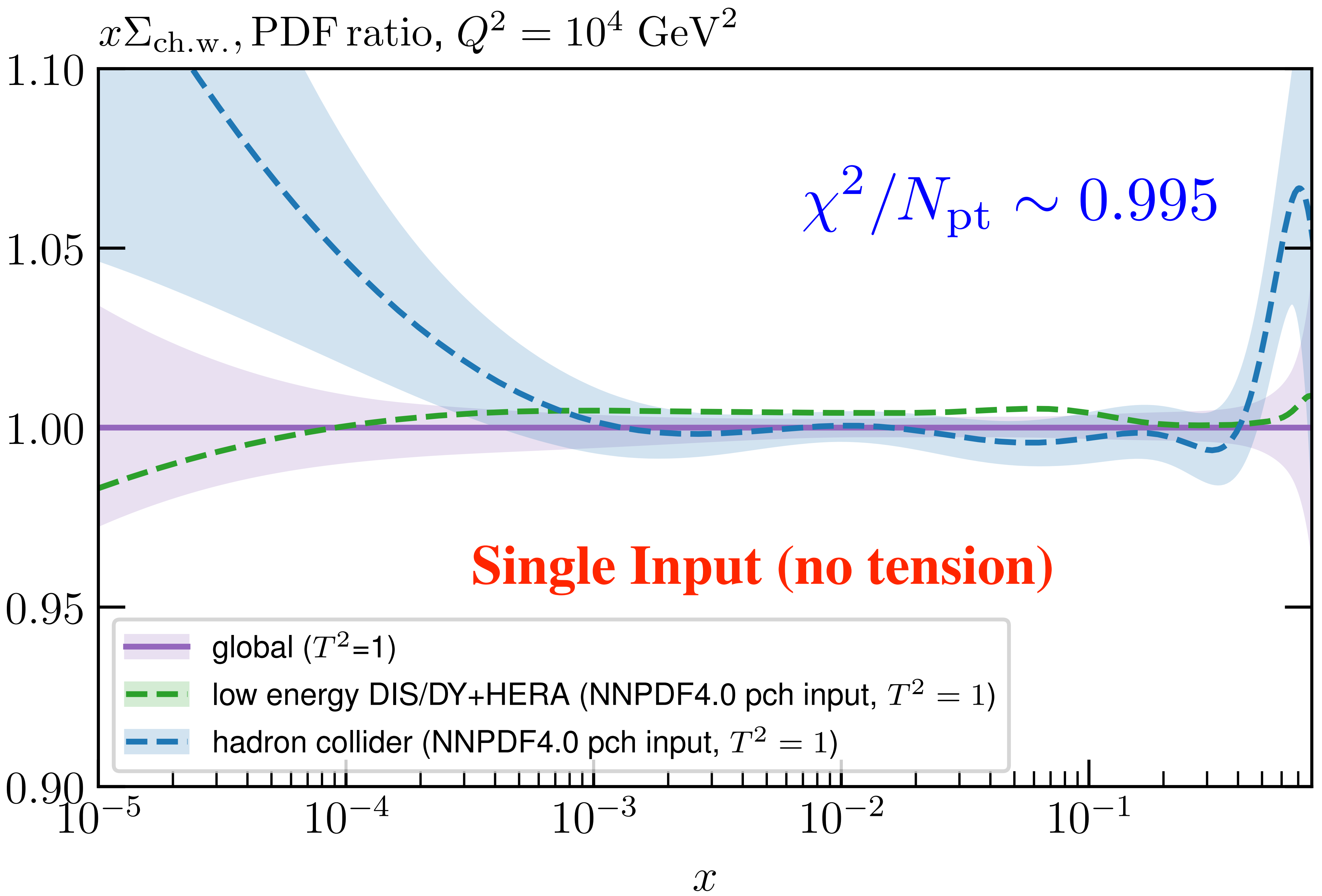
Backup

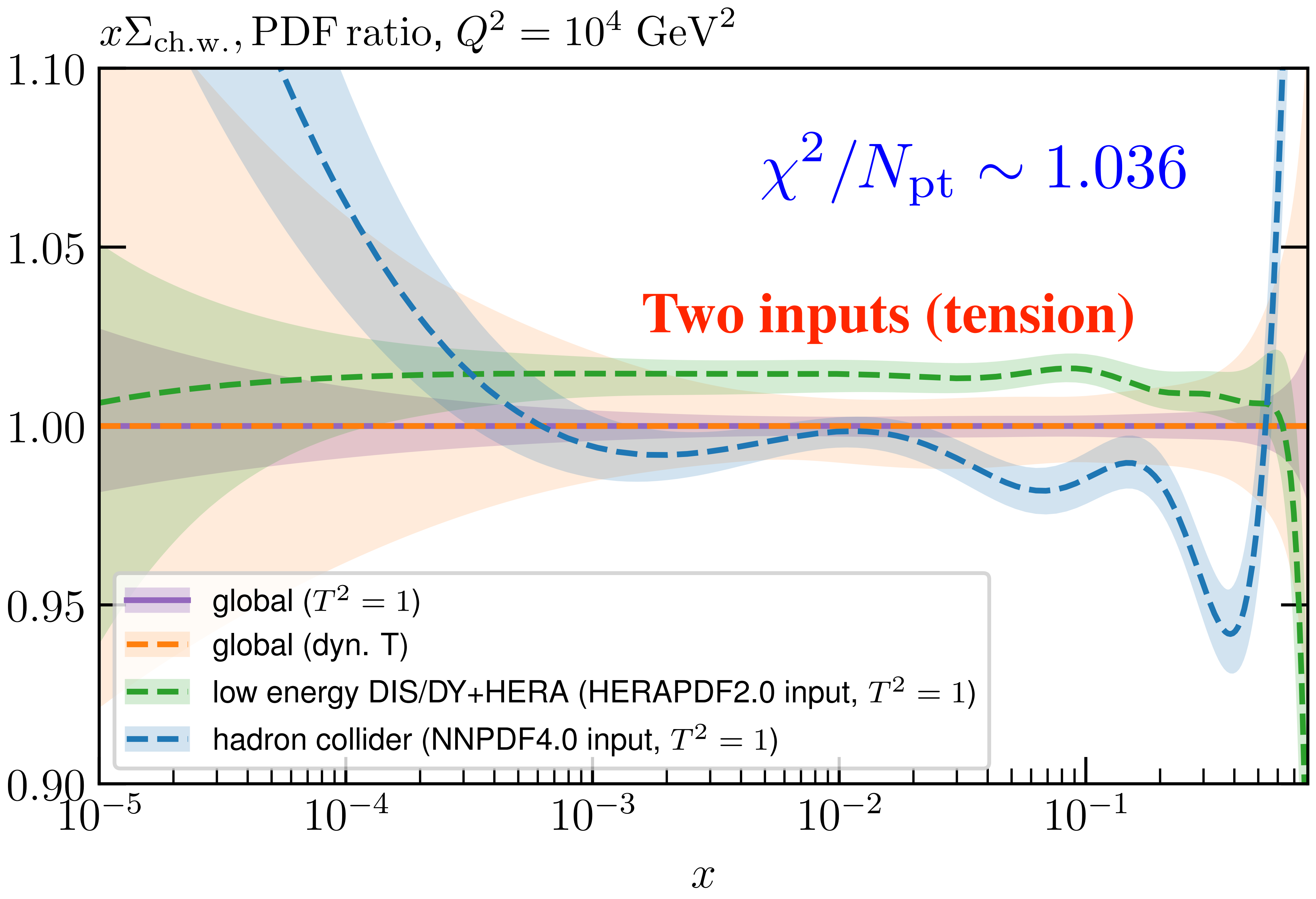

xd, PDF ratio, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
x
xu, PDF ratio, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
$x(s+\bar{s})$, PDF ratio, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
x
$x(s-\bar{s})$, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
x

Tolerance (Again)

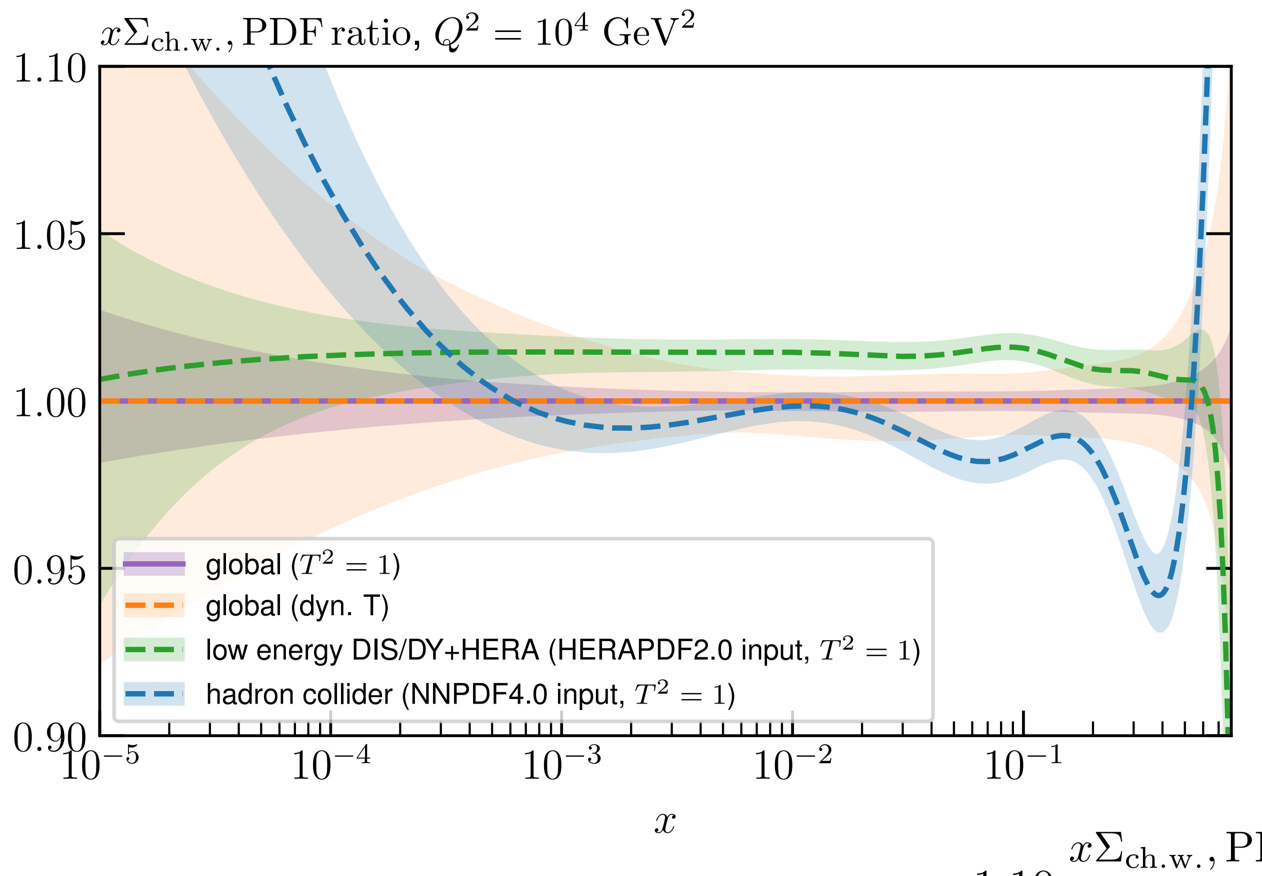
See also G. Watt and R. Thorne, arXiv:1205.4024

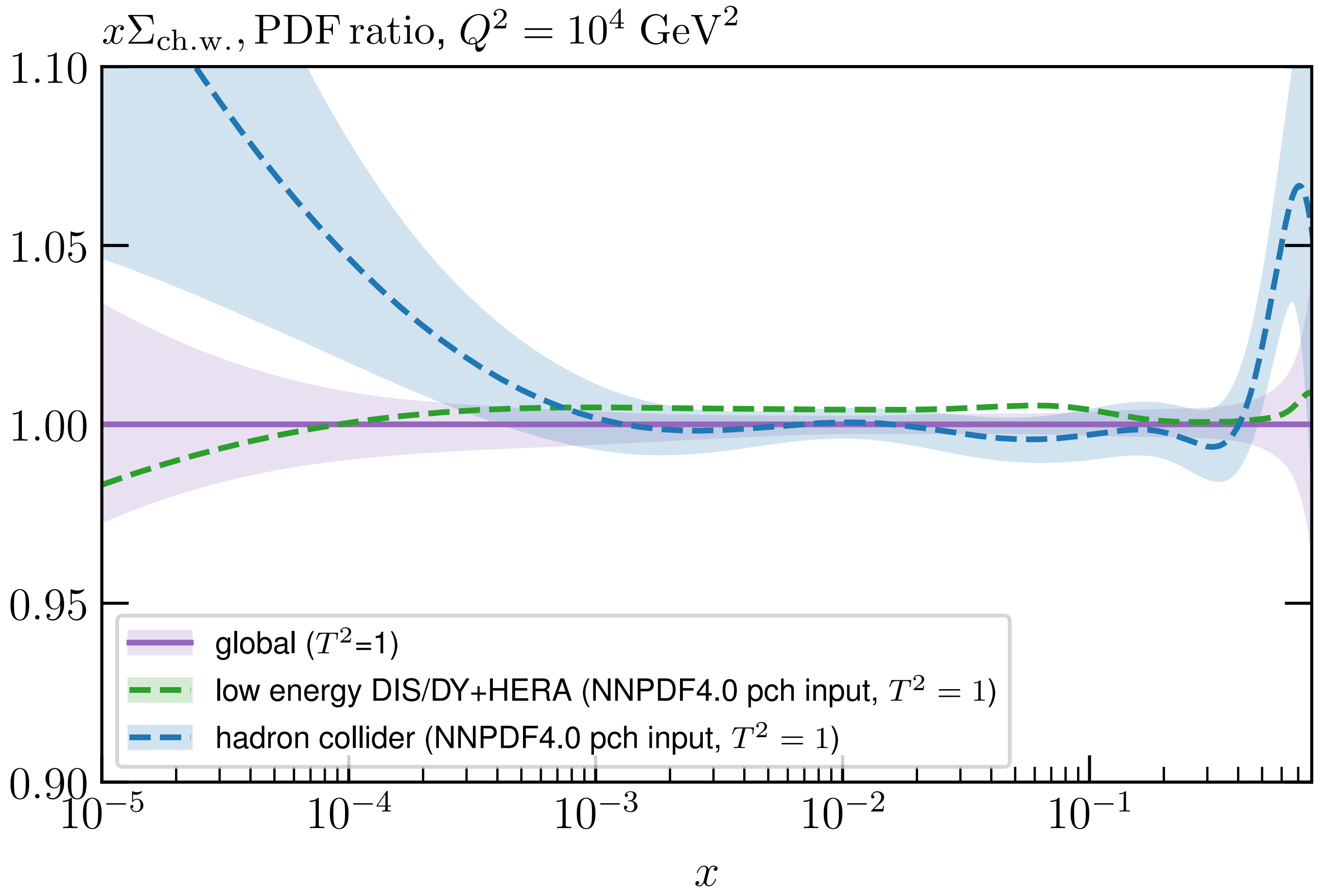
- Can also use closure test to motivate need for tolerance. Generate:
  - ★ **Fixed-Target DY + DIS** data with **HERAPDF2.0** input.
  - ★ **Hadron Collider** data with **NNPDF4.0** (pch) input.
- Inputs are indeed in tension for various PDFs - simply model of incompatibility in fit. What do we find?

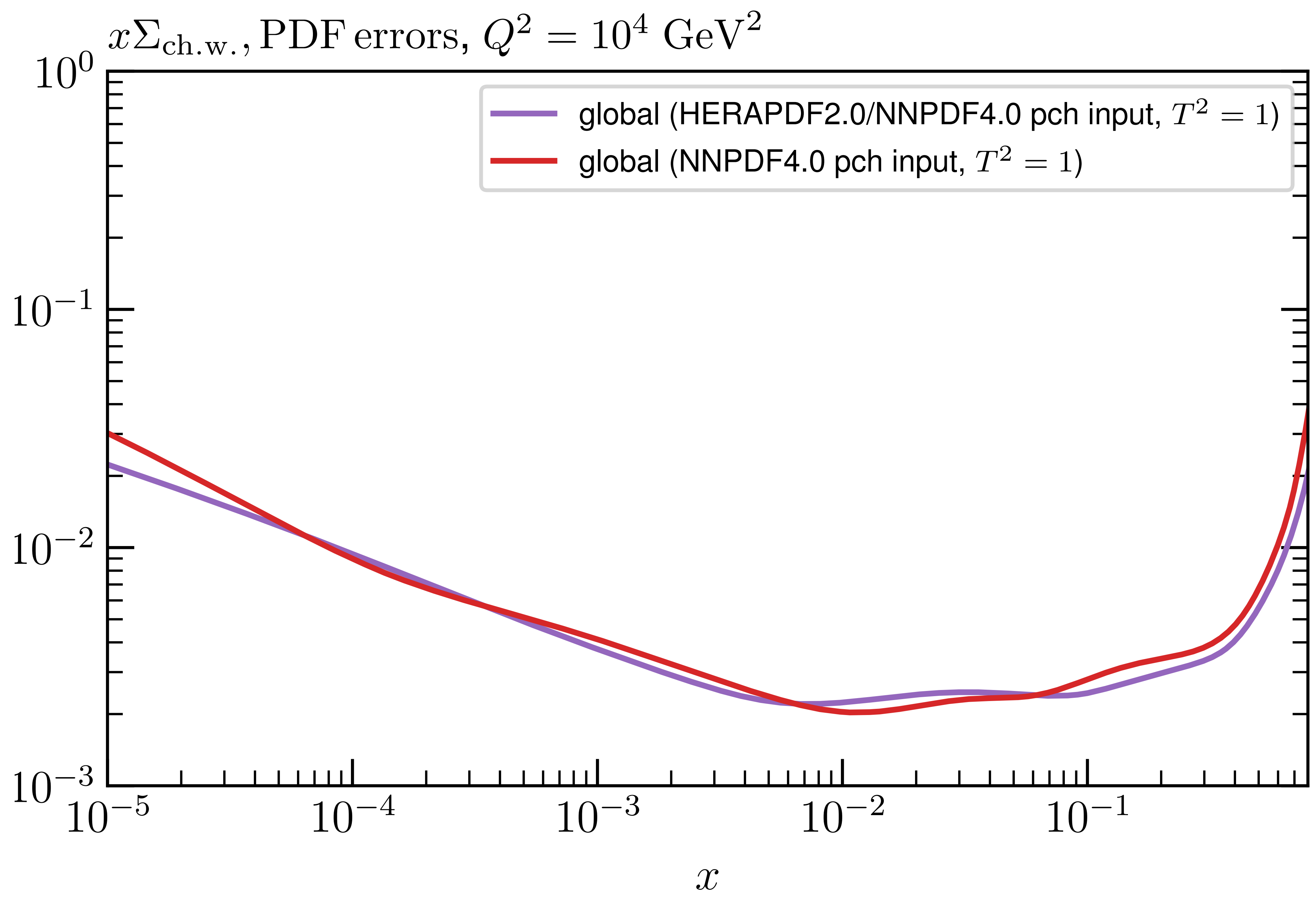




- Fit including tension lies ~ in the middle where tension appears. The $T = 1$ error clearly too small - enlarged MSHT tolerance does rather better. Crucially the $T = 1$ error v. similar between left + right…







Backup

- This effect is completely expected. Can show in simple toy model: PDF uncertainties driven by the quoted experimental (theoretical) uncertainties whether underlying fit is self—consistent or not.
- Naive application of $T = 1$ criterion in such a scenario will lead to overly **aggressive errors**.