Global PDF Analyses: Overview and Latest Results

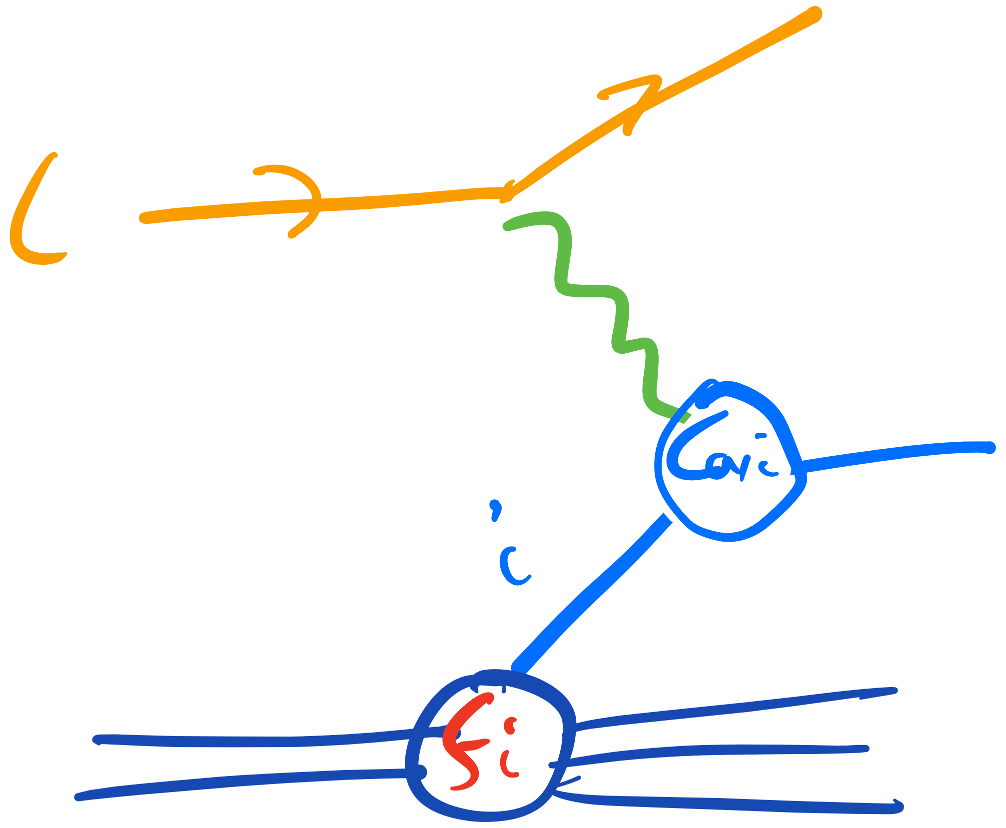
Lucian Harland-Lang, University College London

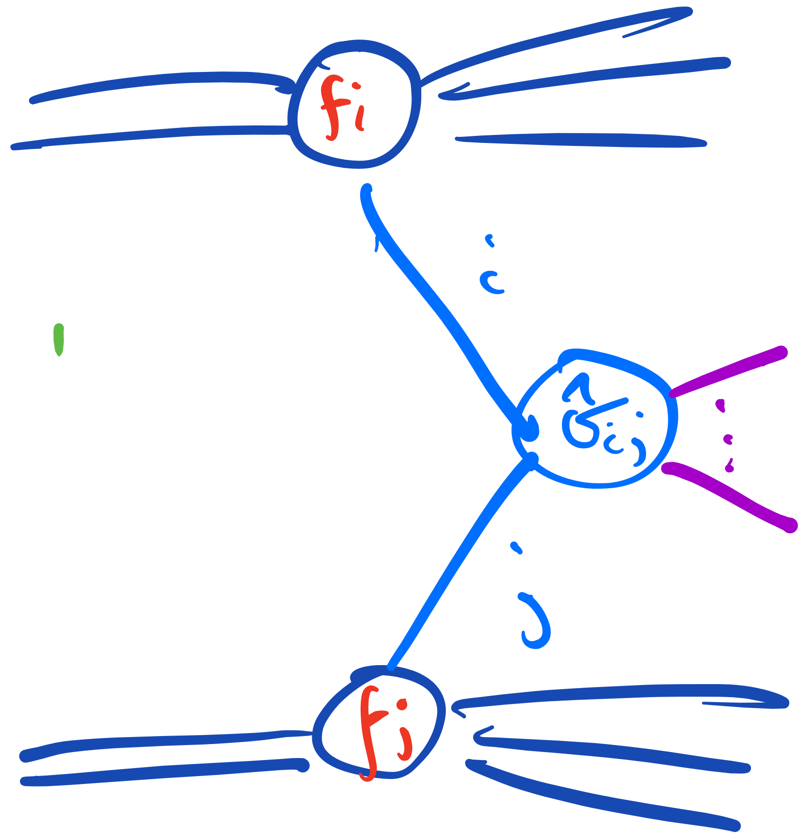
PDFLattice 2024, Nov 18 2024

Collinear, unpolarized

Setting the Scene…

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





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

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

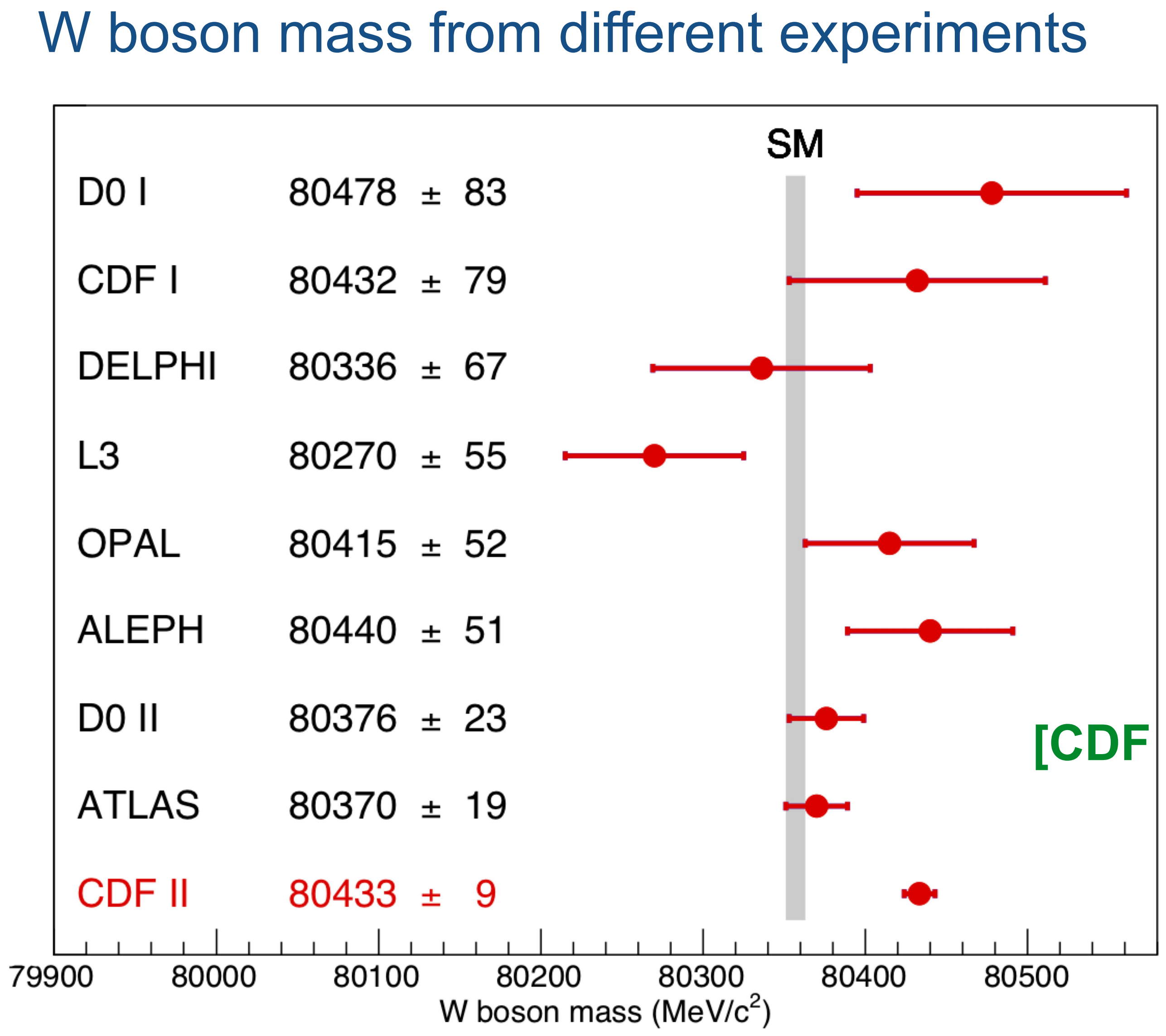
$$\text{Factorization} \Rightarrow f_i^{\mathrm{DIS}}(x,Q^2) \equiv \{f_i^{\mathrm{Collider}}(x,Q^2)\} \longleftarrow$$ **Drell Yan, Jets, Higgs…**

- 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. Aim to constrain PDFs to high precision for all flavours (q, $\bar{q}$, g …) over a wide x region.

Why do we care about (precision) 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)

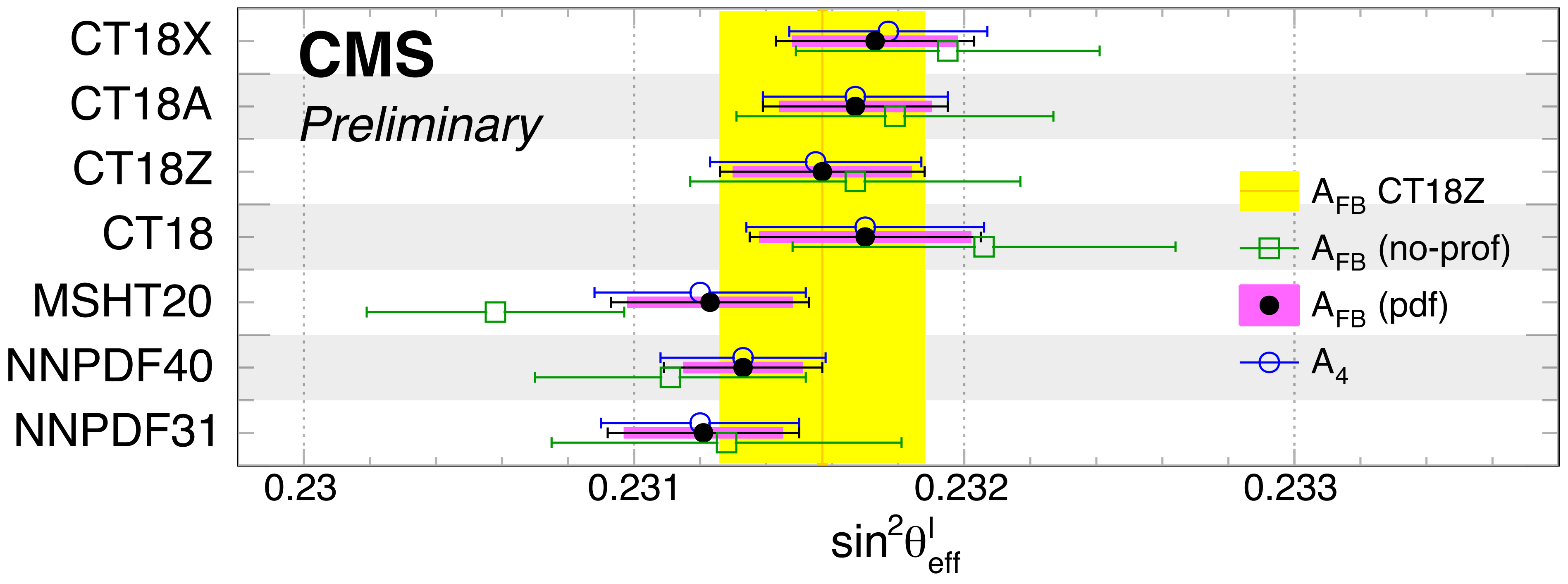
LHCb measurement : $M_W = 80{,}354 \pm 23_{\text{stat}} \pm 10_{\text{exp}} \pm 17_{\text{theory}} \pm 9_{\text{PDF}}$ *[JHEP* **2022,** 36 (2022)]

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

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

(up to)

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

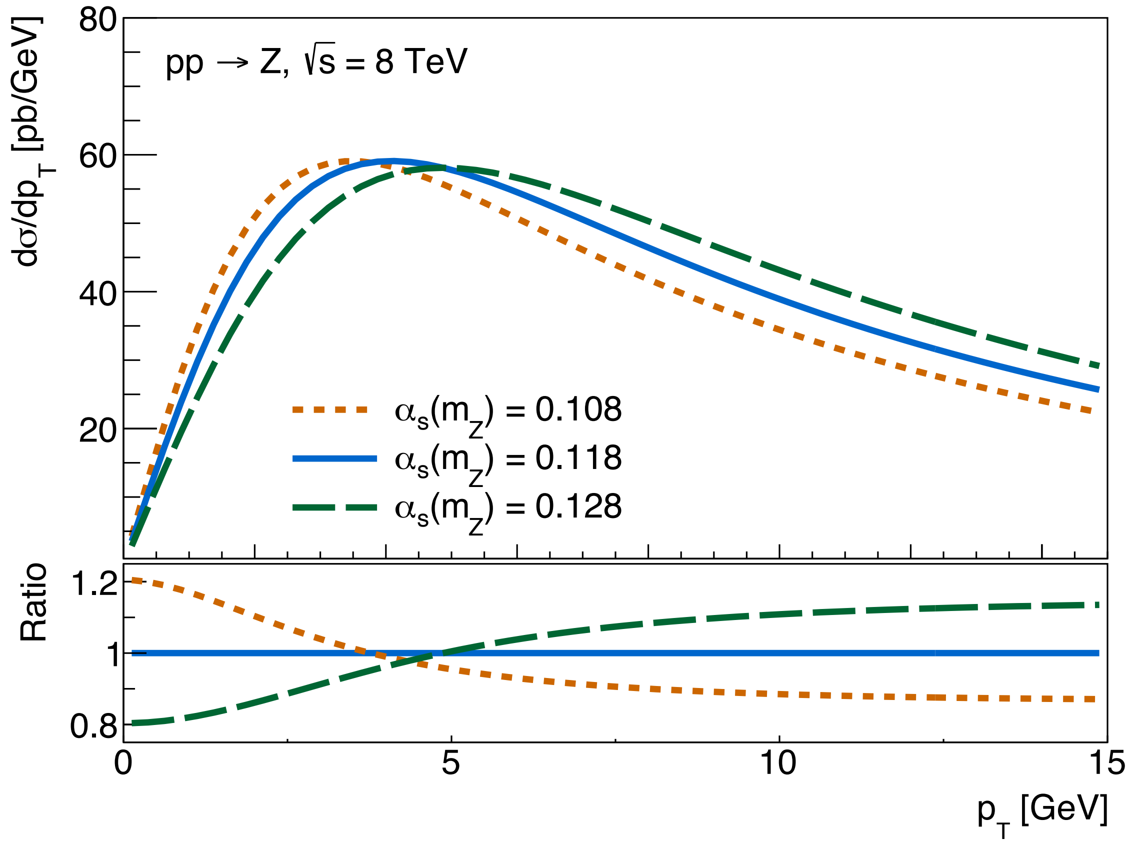


CMS PAS SMP-22-010

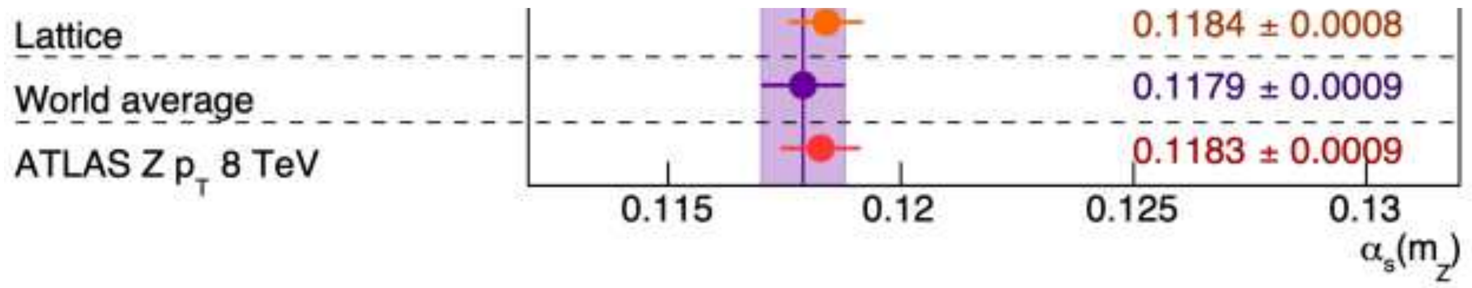
$$\sin^2\theta^\ell_{\text{eff}} = 0.23157 \pm 0.00010(\text{stat}) \pm 0.00015(\text{syst}) \pm 0.00009(\text{theo}) \pm 0.00027(\text{PDF}),$$

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

$$\alpha_S$$



ATLAS, 2309.12986

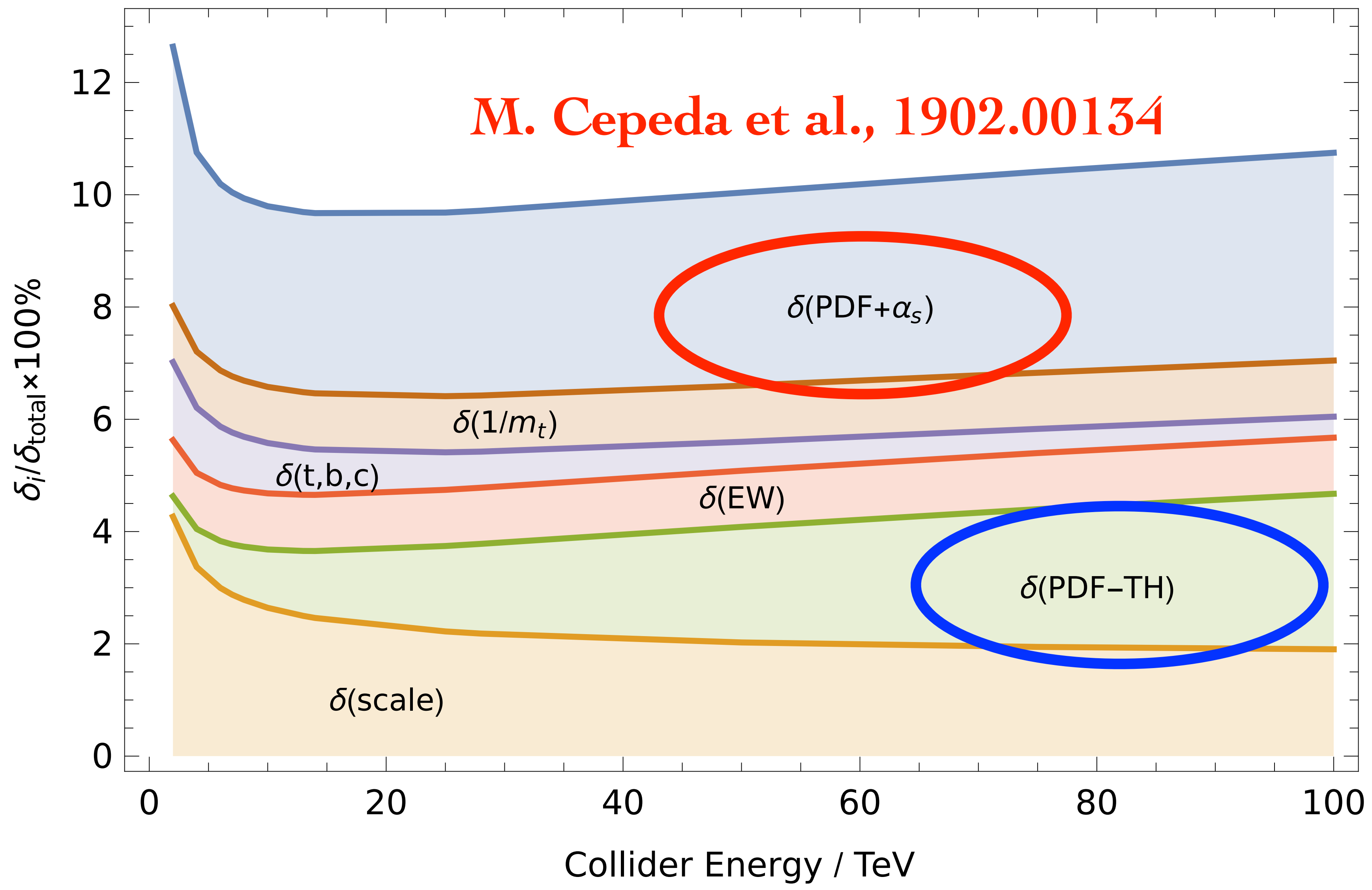


$$\sigma_{\text{PDF}} \sim \sigma_{\text{tot}}/2$$

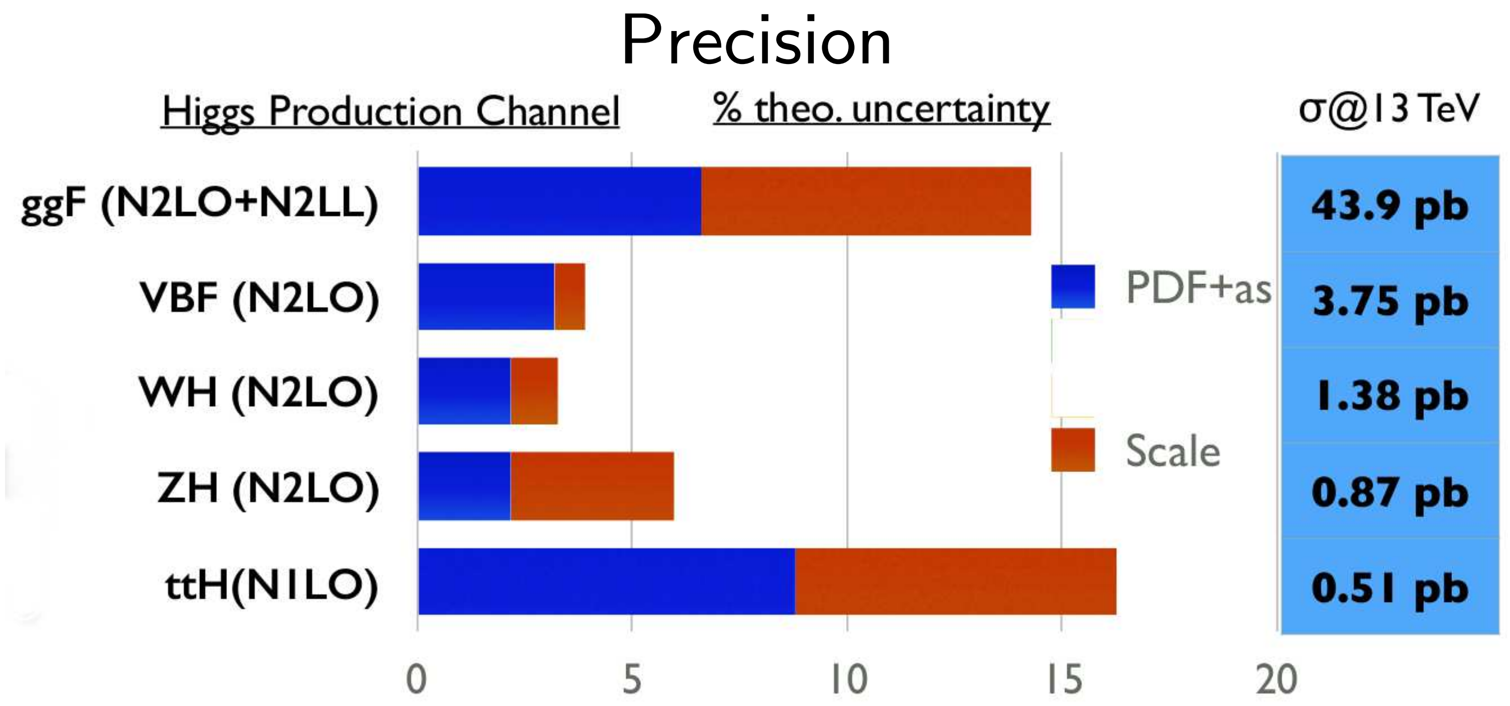
Disclaimer: will generally refer to papers by their arxiv number, even if published.

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

Image Credit: Emanuele Nocera


M. Cepeda et al., 1902.00134

(progress here)



- 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.

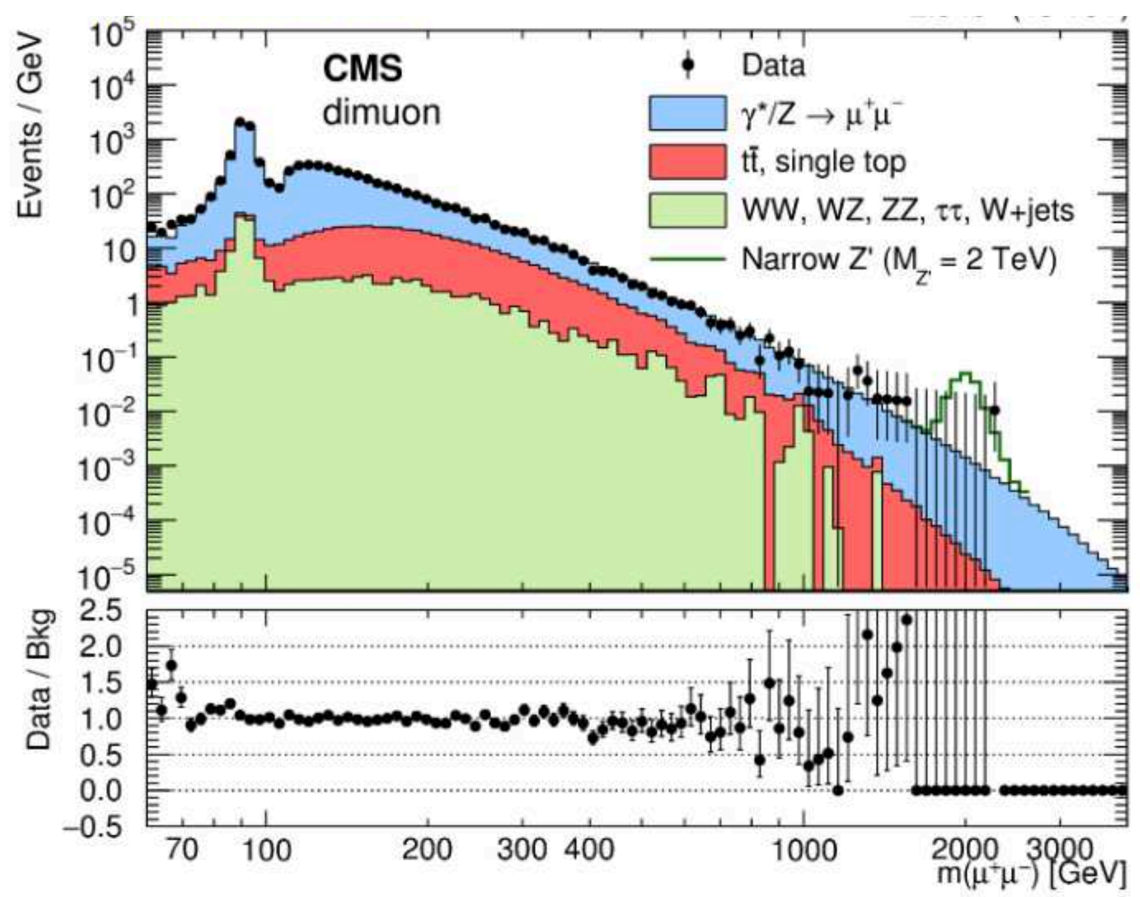
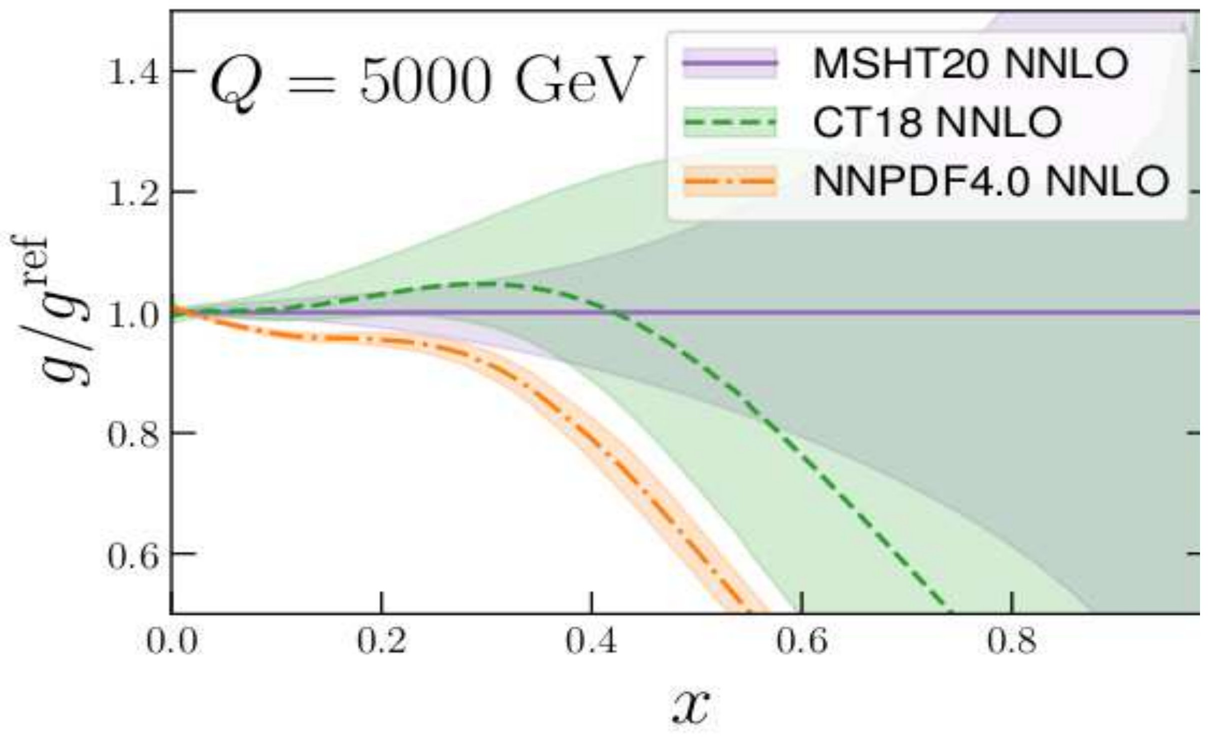


Image Credit: Tom Cridge



Global PDF Fits

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

measure **fit** **predict**

Parameterise PDFs at low scale Q_0

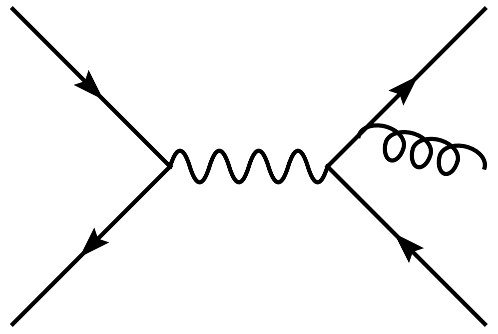
$f_i(x,Q_0):\ A_f x^{a_f}(1-x)^{b_f} \times \longrightarrow \sum_{i=1}^{n} \alpha_{f,i} P_i(y(x))$, **CT, MSHT…**

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

↓

Select data & theory

Perturbative QCD for parton-level theory



$+ \ldots$

↓

Perform fit

Minimise $\chi^2(\{a\},\{\lambda\}) = \sum_{k=1}^{N_{pt}} \frac{1}{s_k^2}\left(D_k - T_k - \sum_{\alpha=1}^{N_\lambda} \beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha^2$,

DGLAP: $f(x,Q_0) \to f(x,\mu_{\mathrm{data}})$

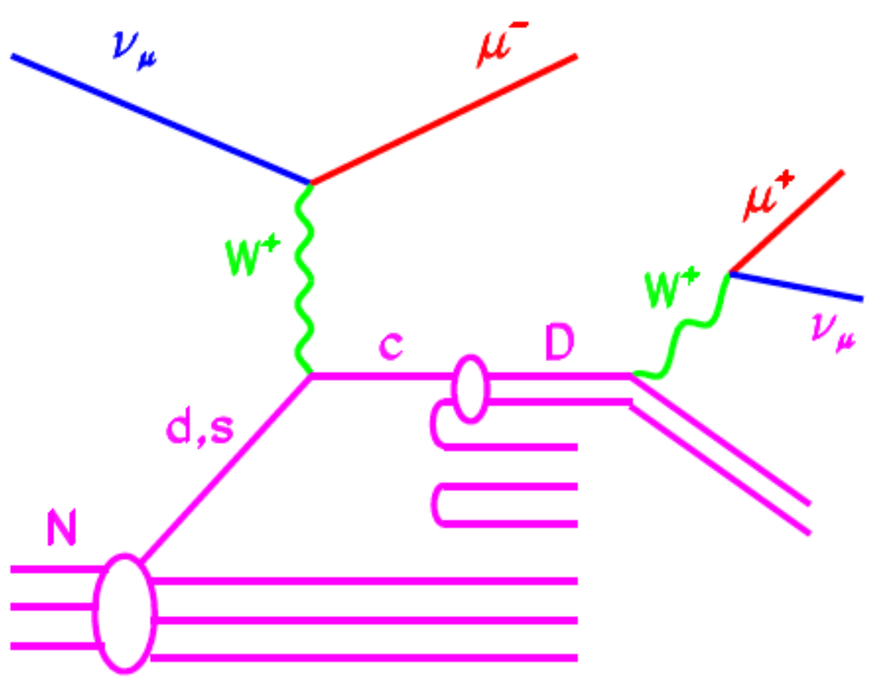
↓

Output PDFs

$f(x,\mu) \pm \Delta(x,\mu)$

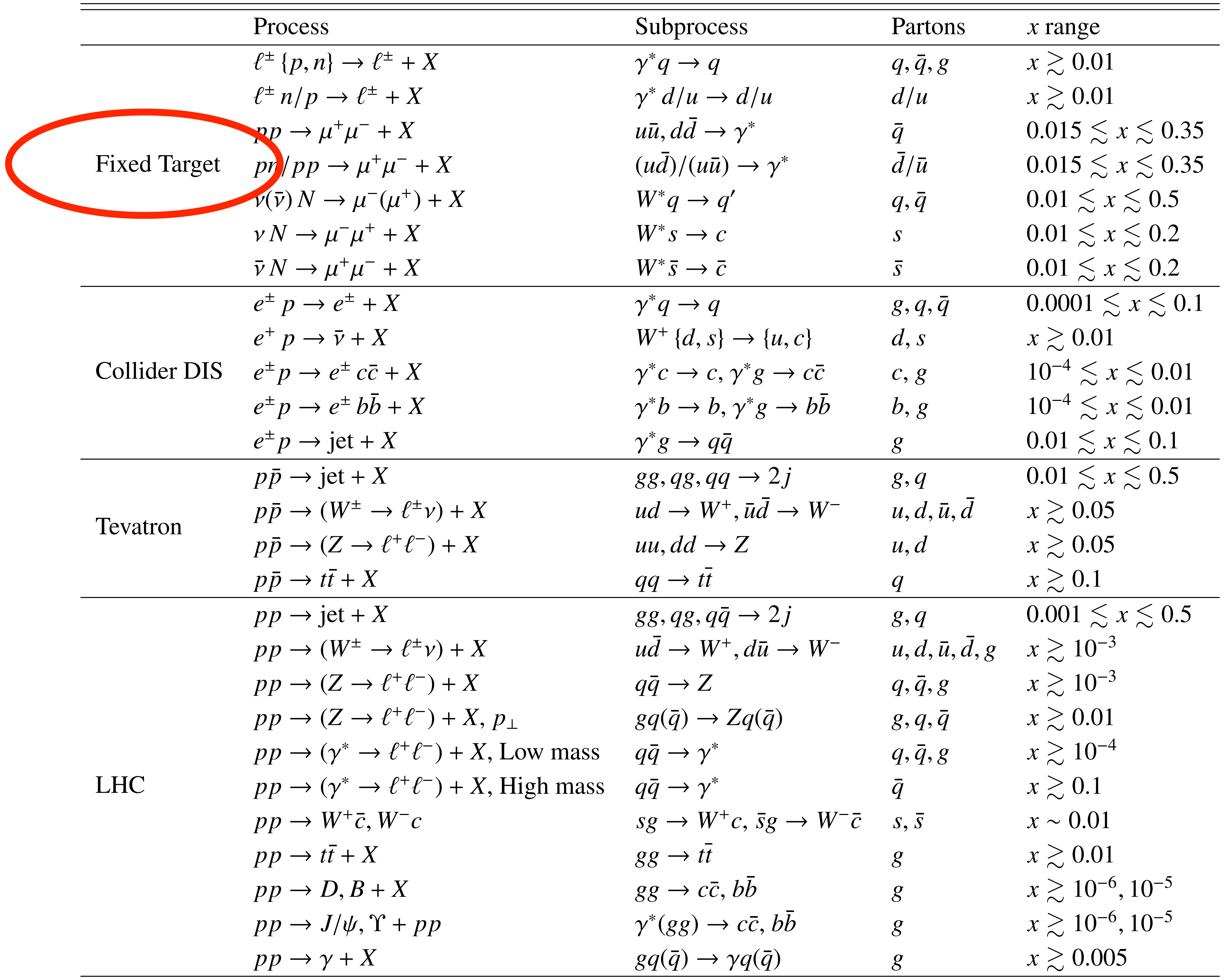
- In a bit more detail…

Global PDF fits: datasets



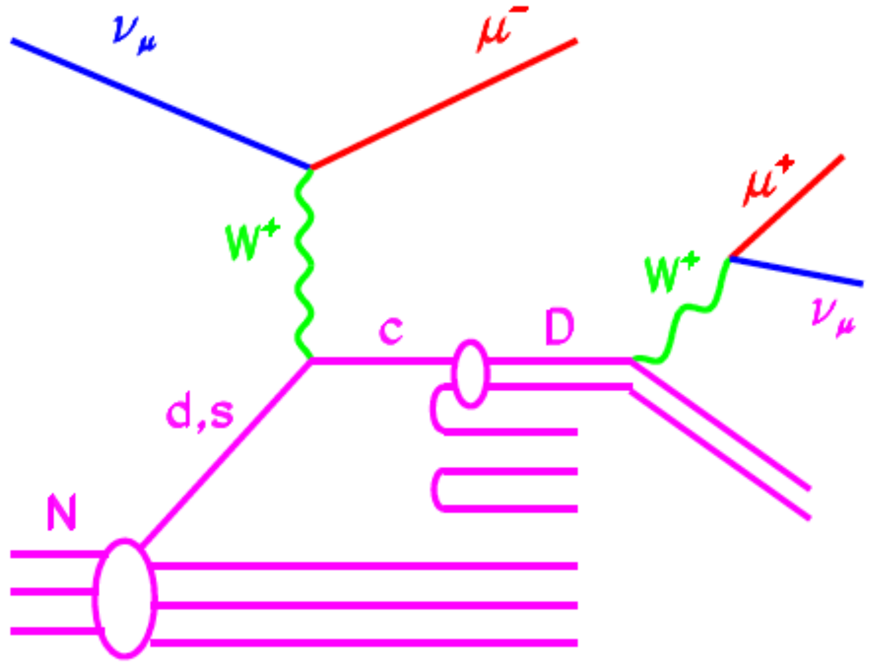
$$lh \to l + X$$

- Fixed target: sensitive to higher x .
- Different lepton/ hadron beams - flavour decomposition.

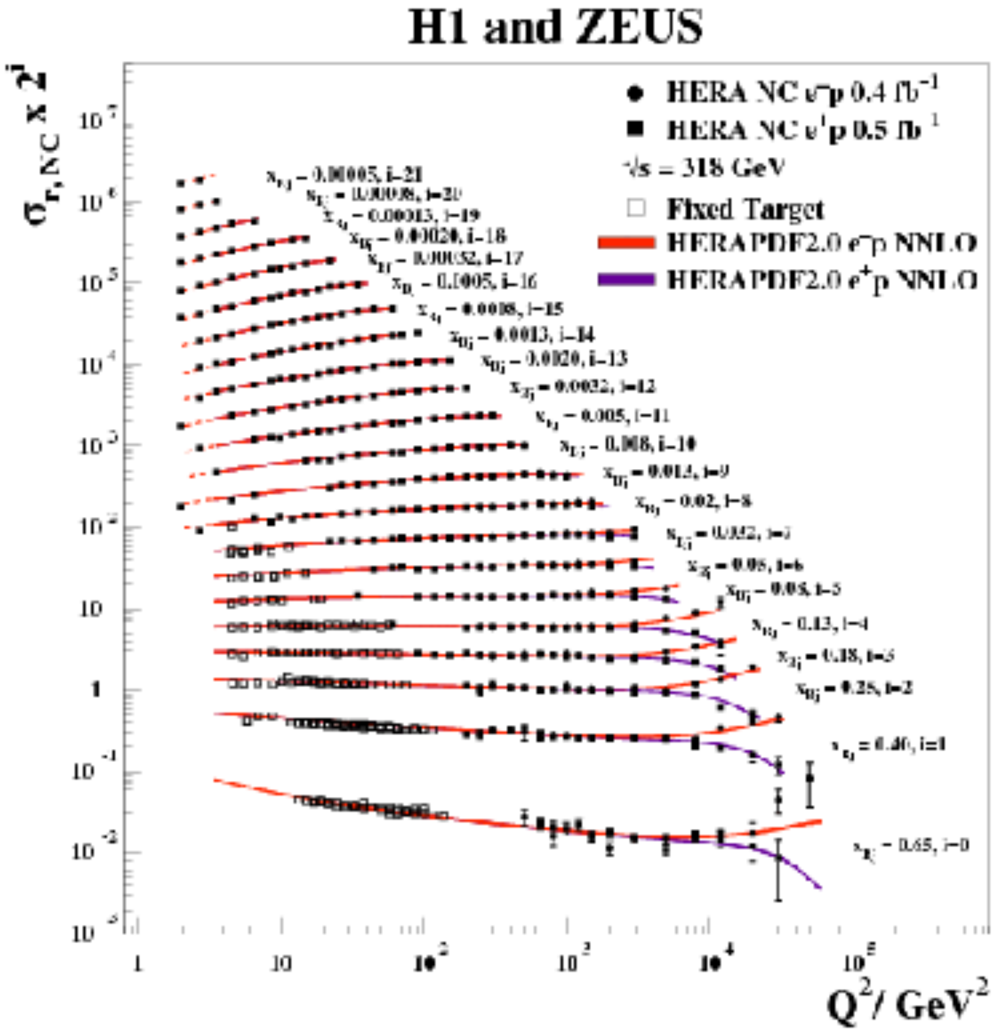


	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm\,\{p,n\} \to \ell^\pm + X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm\, n/p \to \ell^\pm + X$	$\gamma^*\, d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu})\, N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu\, N \to \mu^-\mu^+ + X$	$W^* s \to c$	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	$x \gtrsim 0.005$

Global PDF fits: datasets



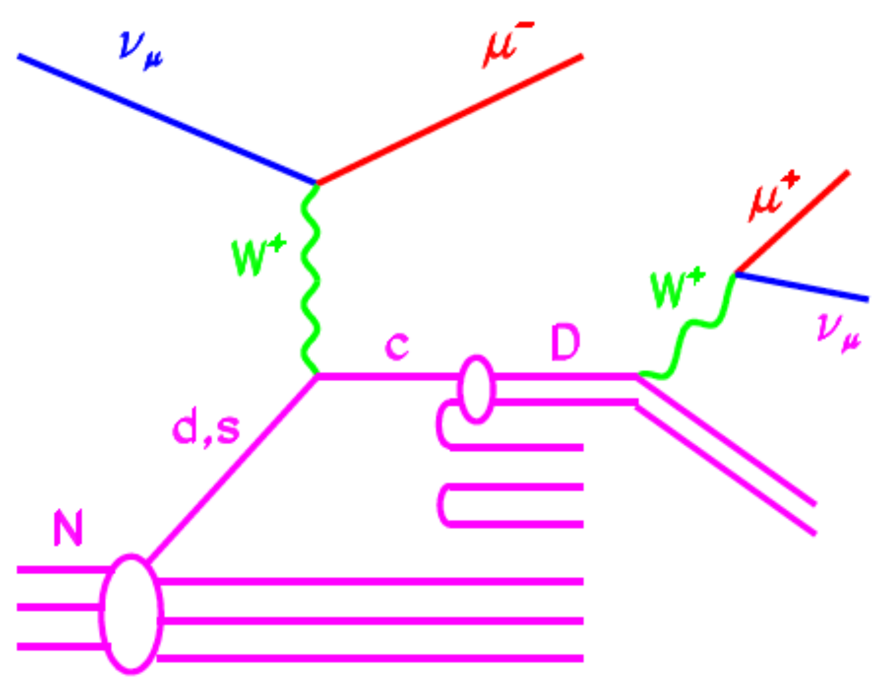
	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm\{p,n\} \to \ell^\pm + X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm n/p \to \ell^\pm + X$	$\gamma^* d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu})\, N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu\, N \to \mu^-\mu^+ + X$	$W^* s \to c$	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	$x \gtrsim 0.005$



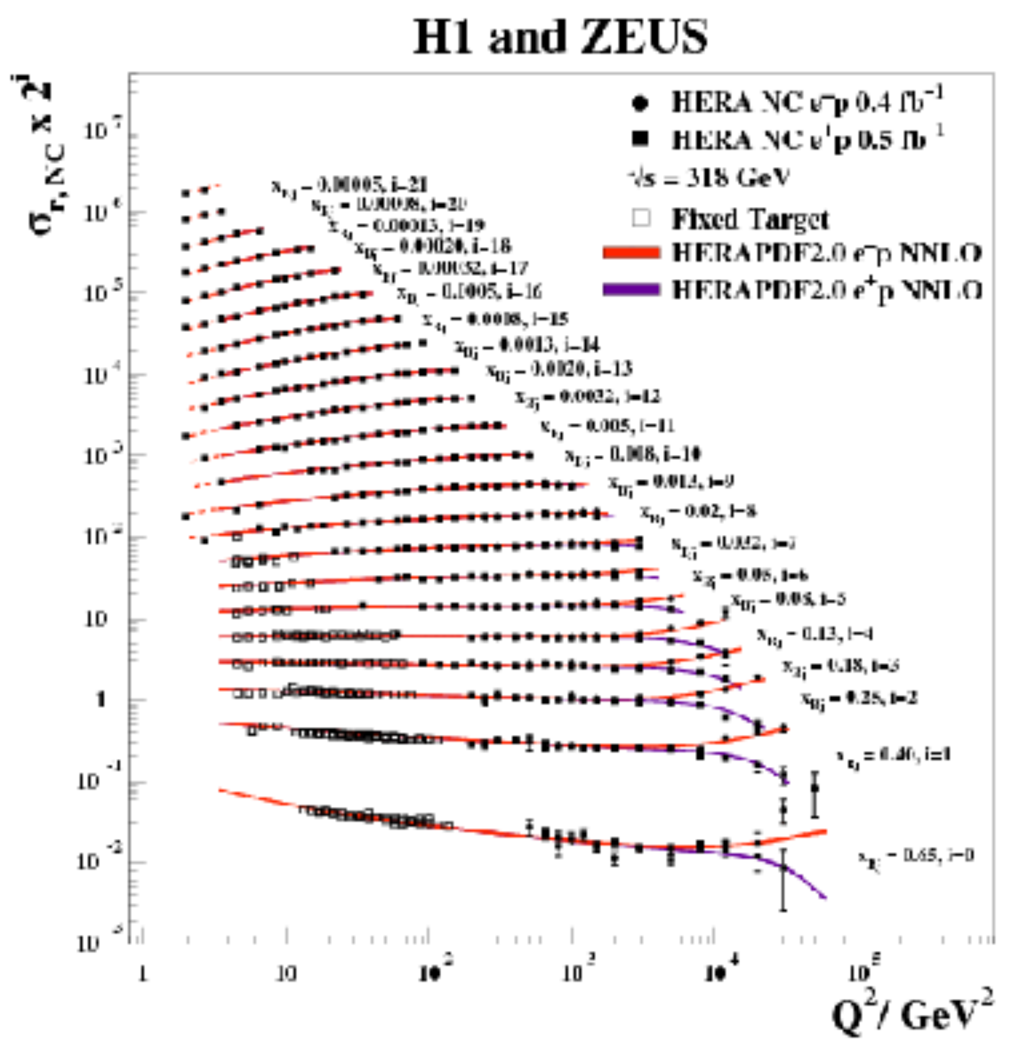
$$lp \to l + X$$

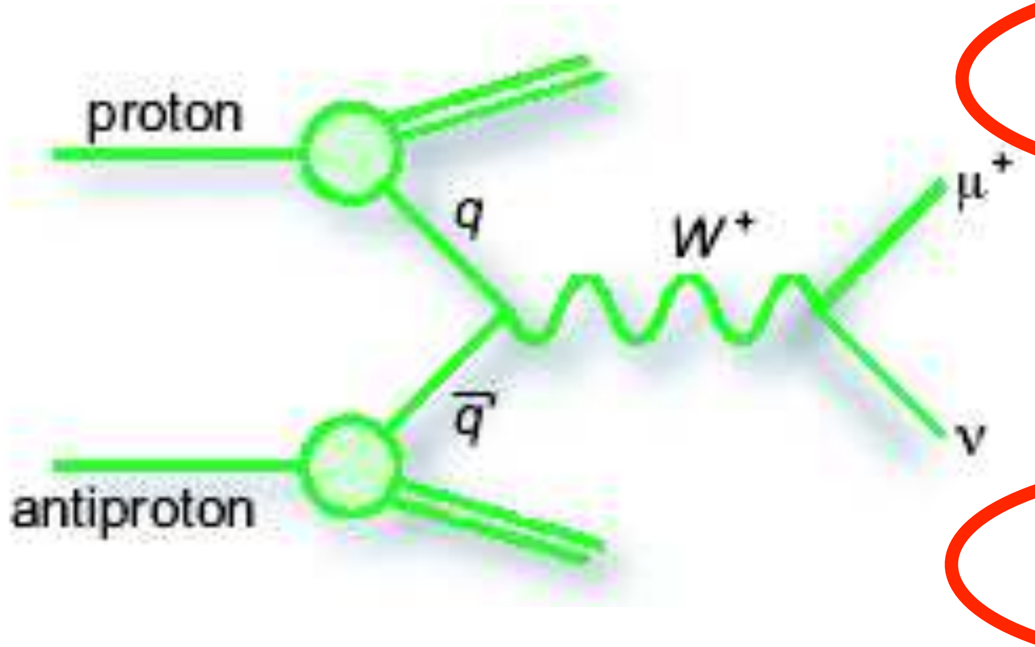
- HERA collider DIS: backbone of PDF fits.
- Gluon and flavour structure at low to intermediate x .

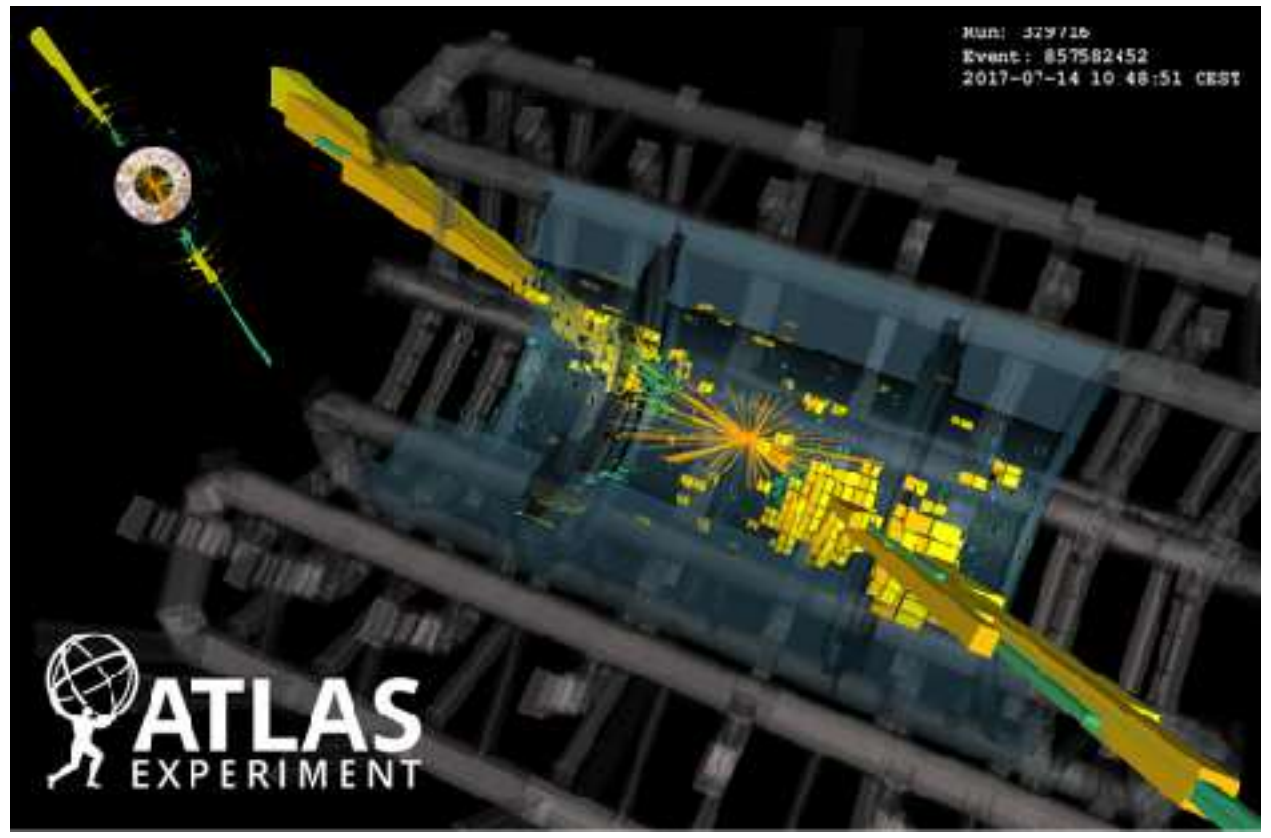
Global PDF fits: datasets



	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	$x \gtrsim 0.005$

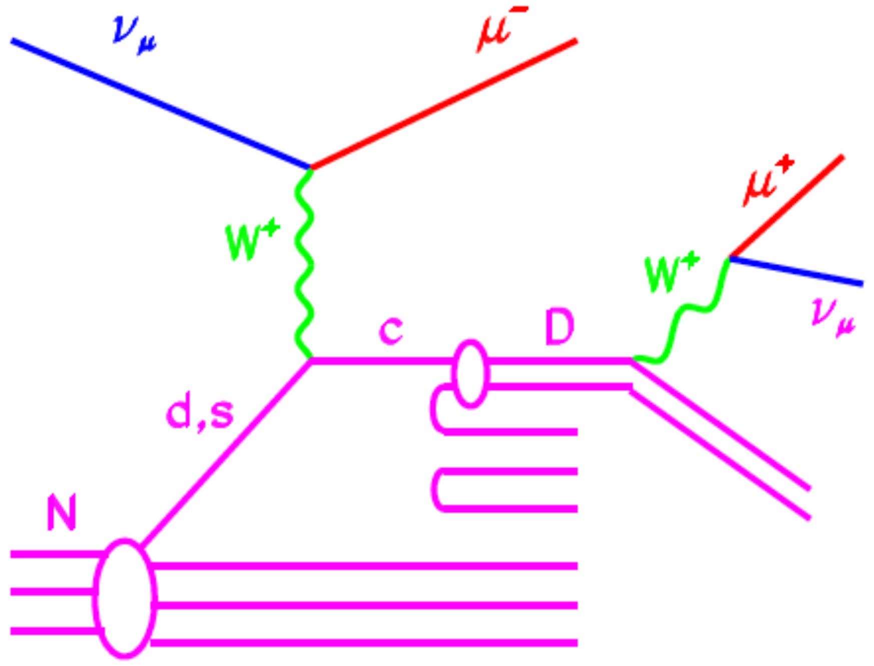


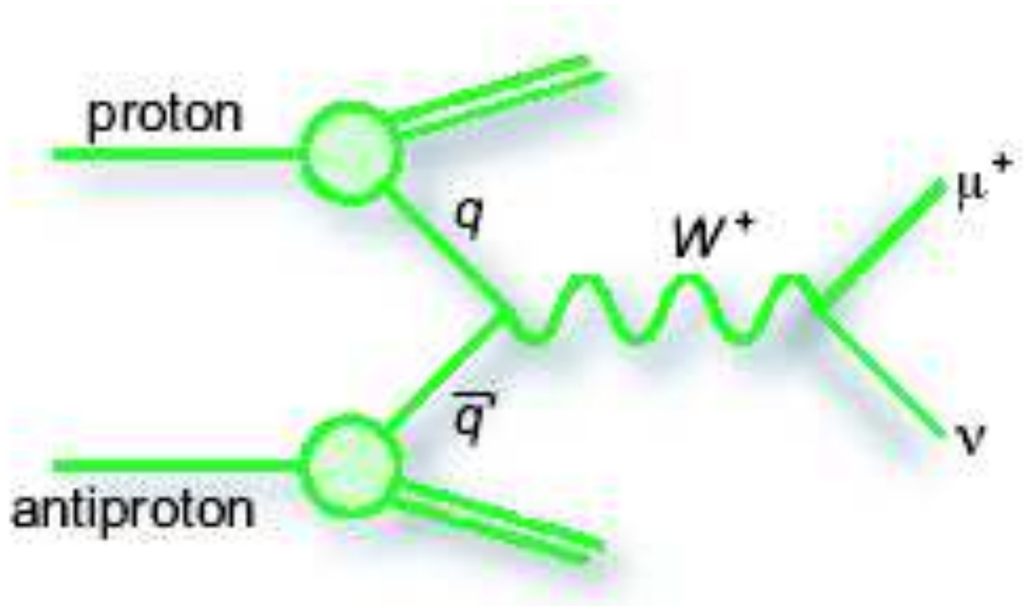




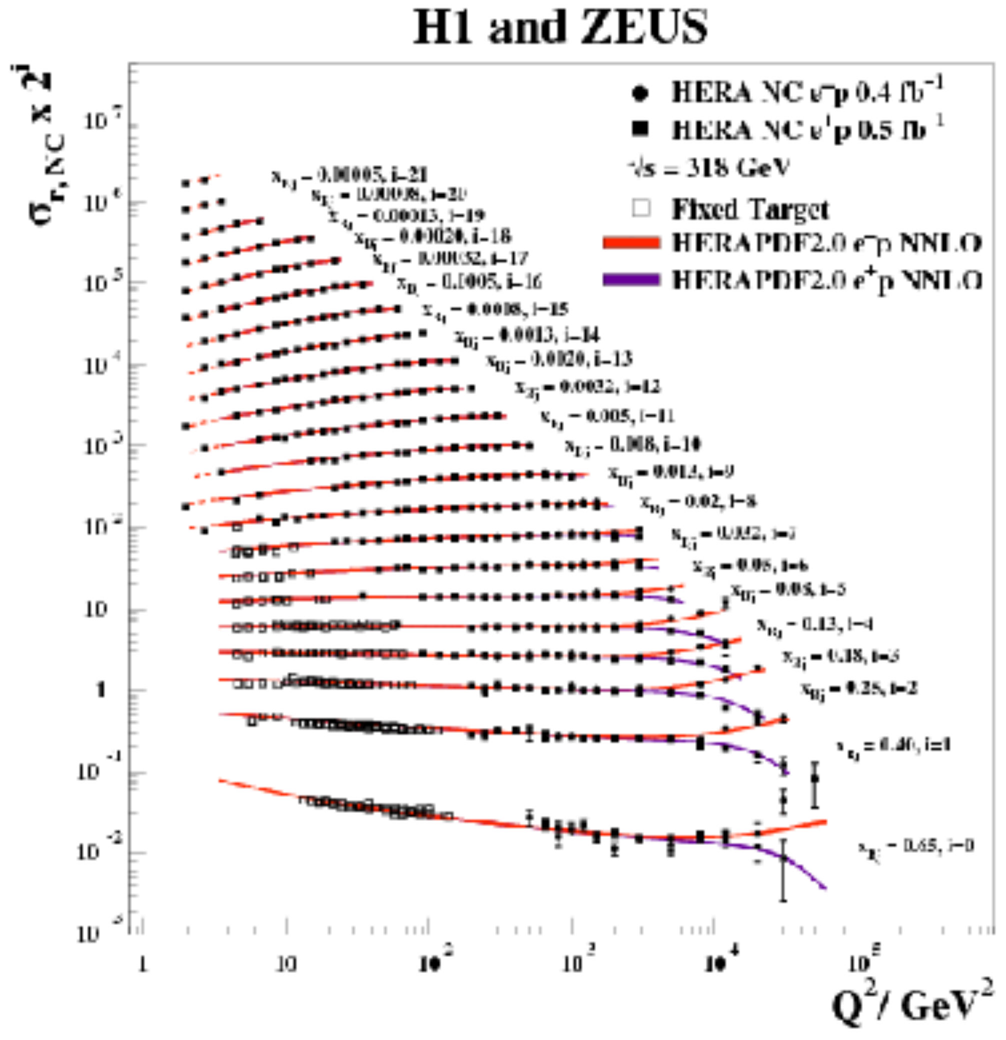
- LHC (Tevatron) collider data.
- Precision Drell-Yan - flavour structure. Top quark, jets - high x gluon…

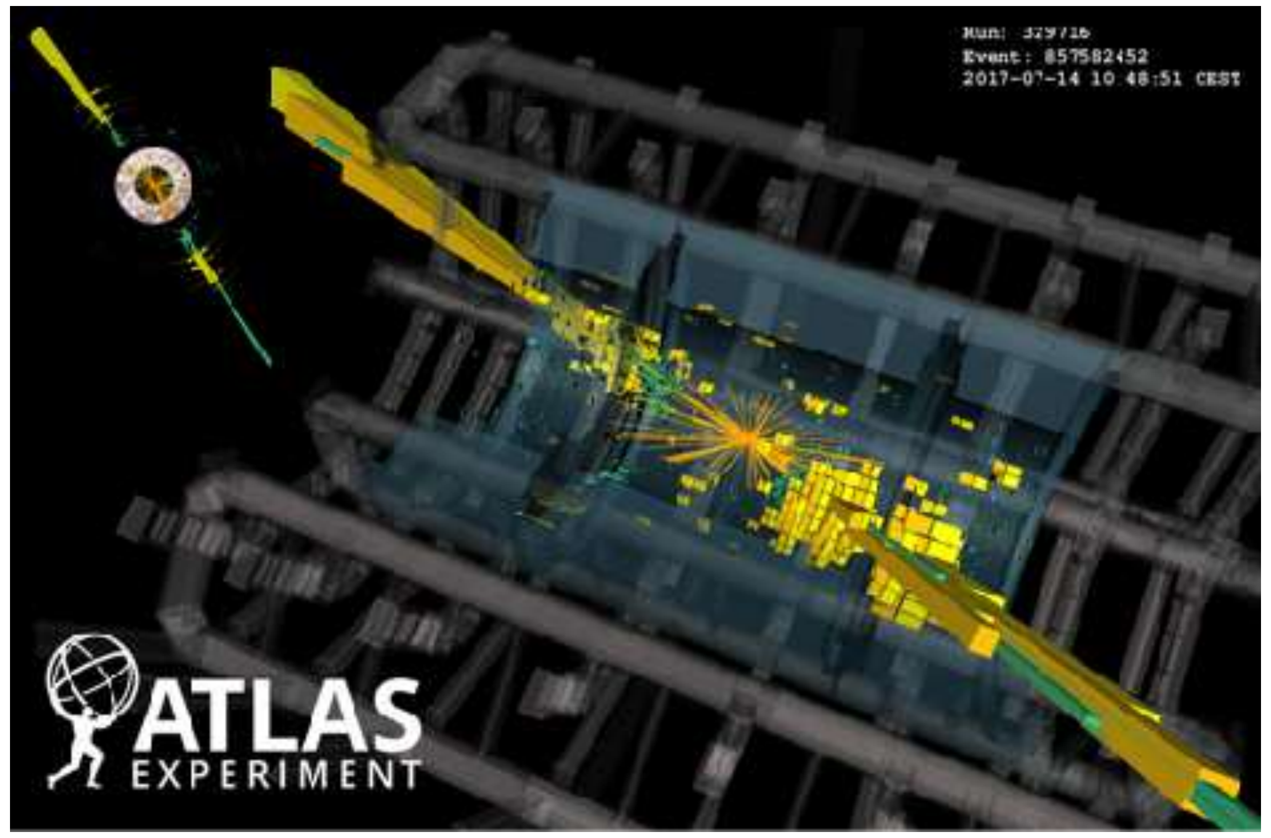
Global PDF fits: datasets





	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	$x \gtrsim 0.005$



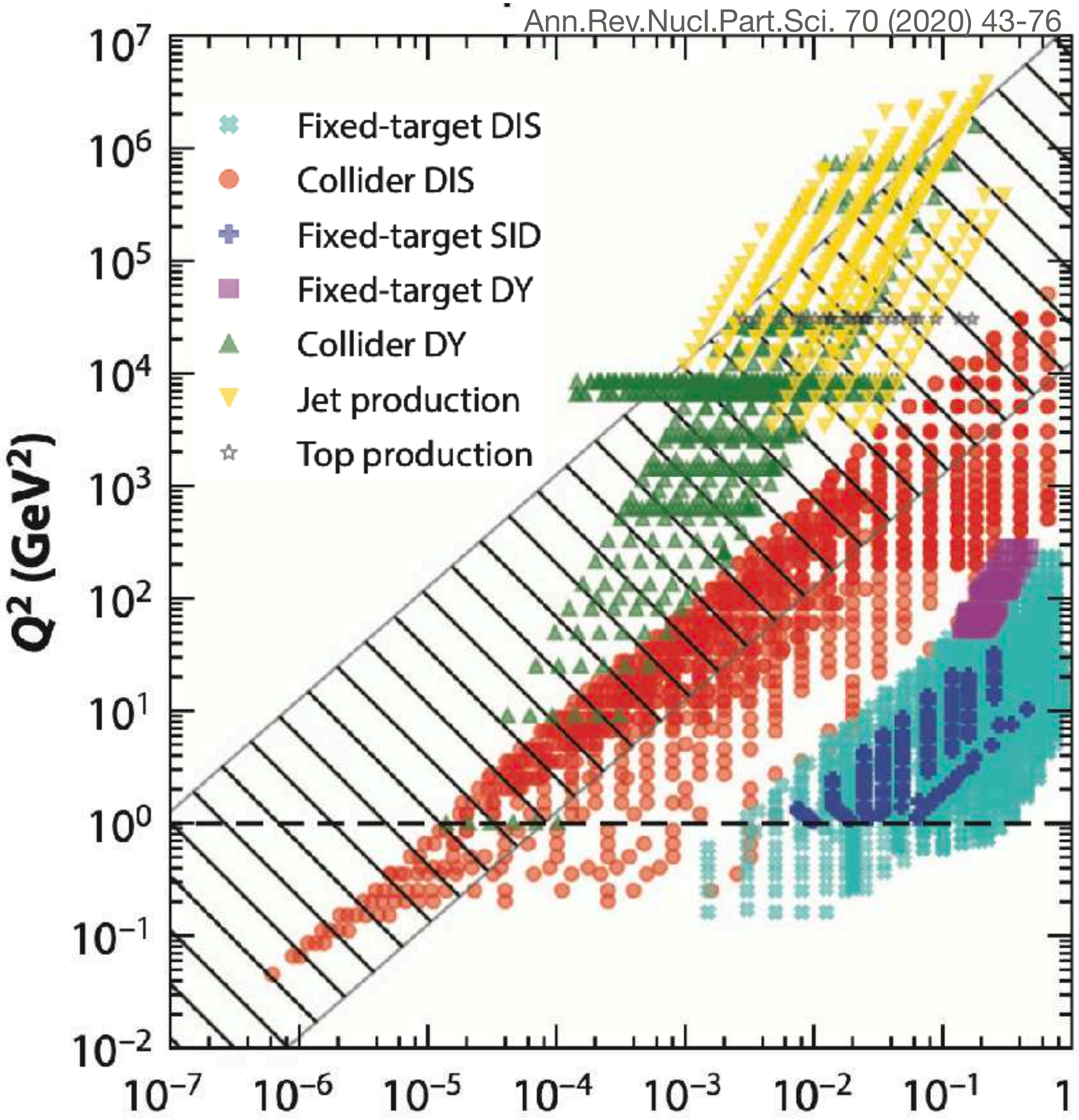


- From fixed target, to HERA DIS and collider LHC data.

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

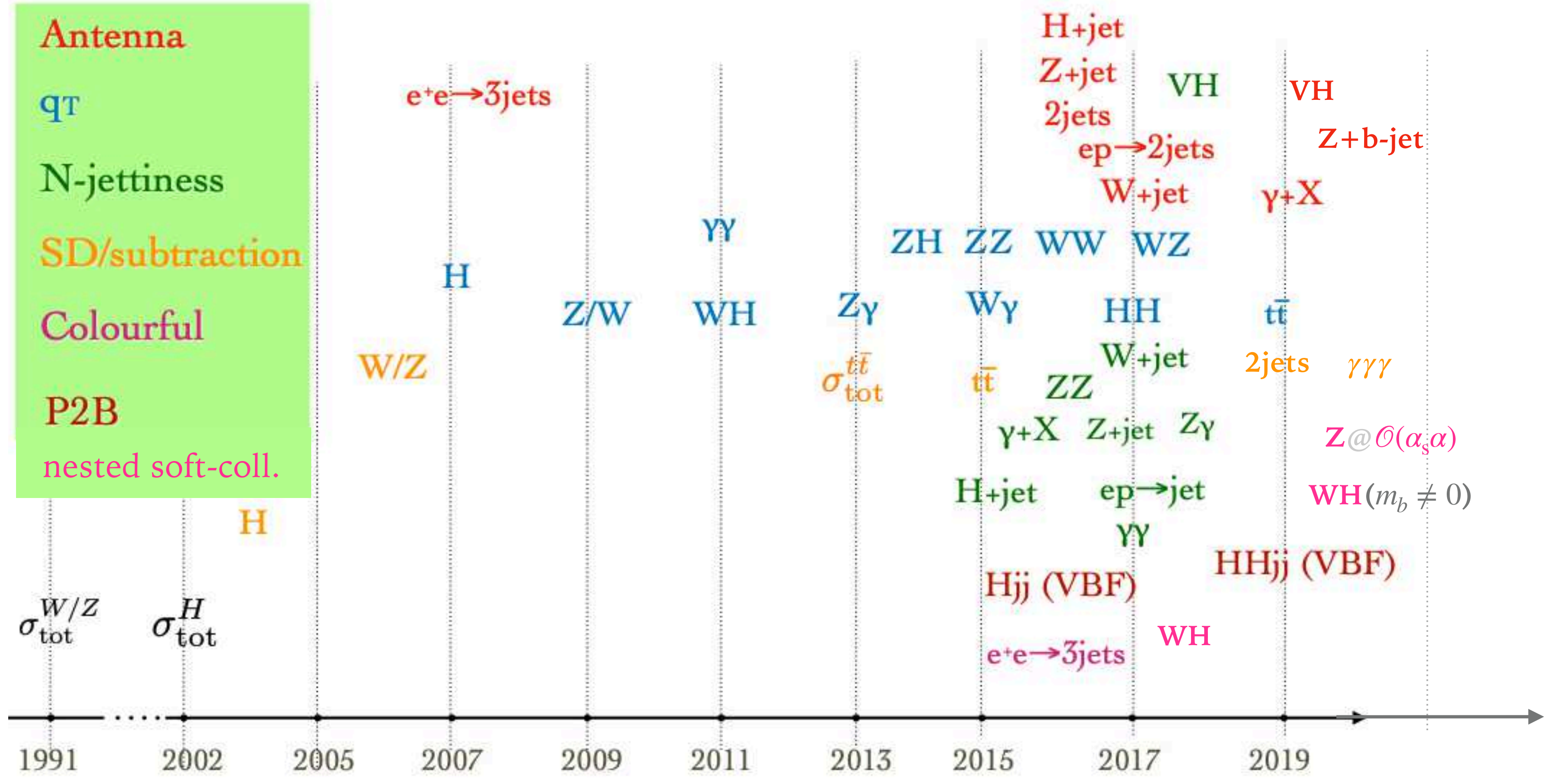
Kinematic Coverage

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



Precision Theory

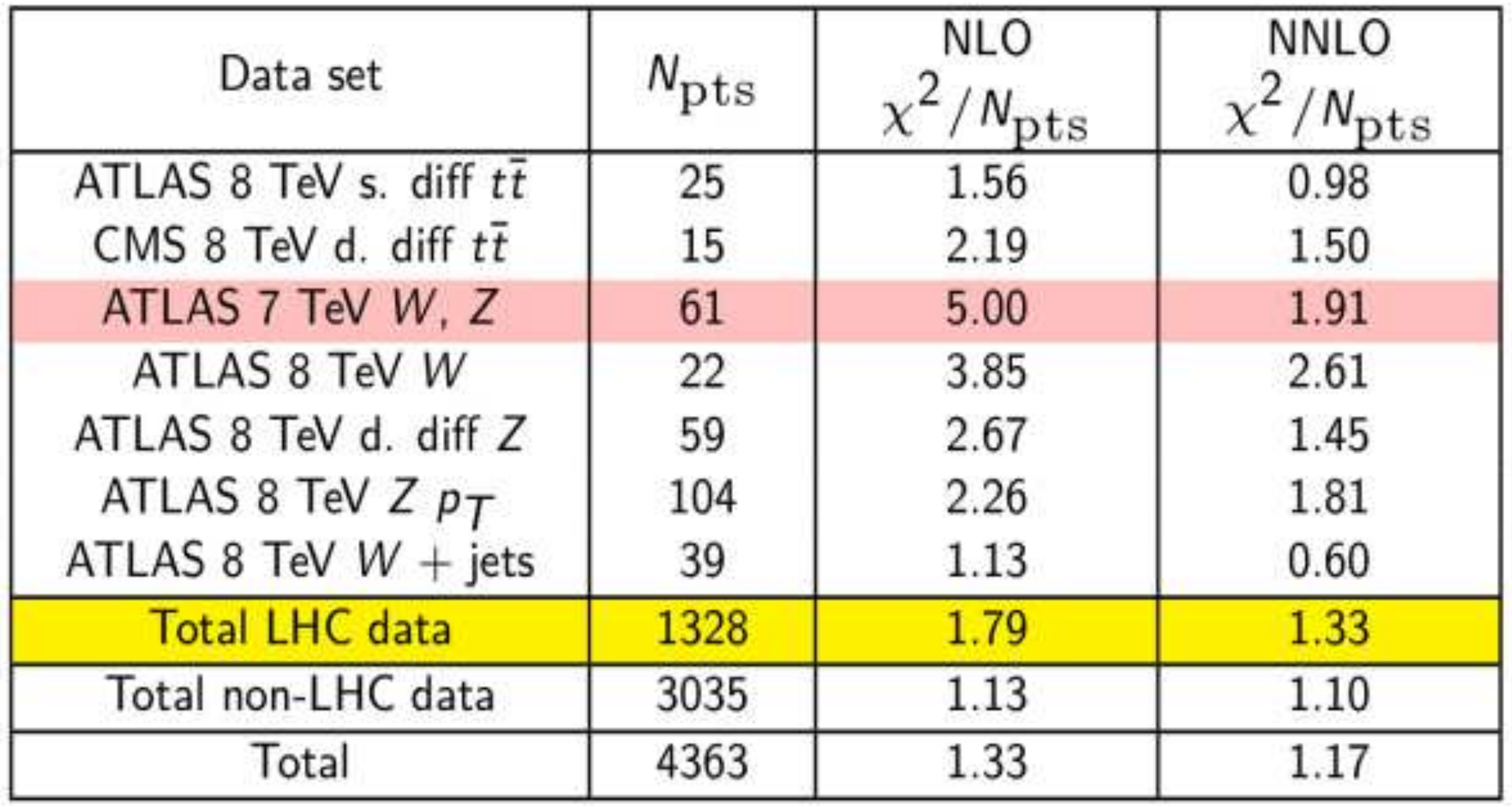
- Has been significant progress in perturbative calculations: **NNLO QCD + NLO EW** now long been the standard (~1% level precision).



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

- 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).

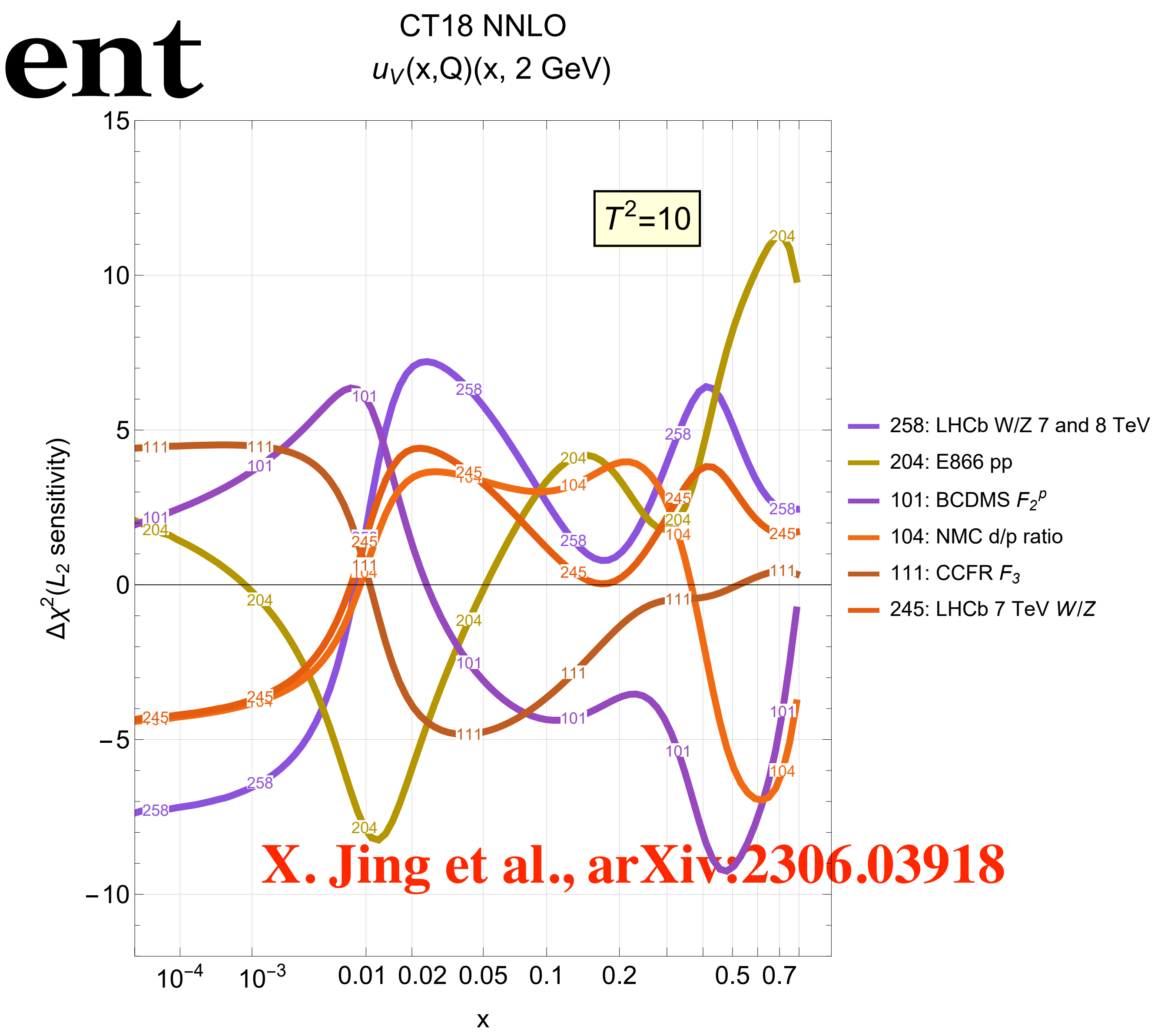
MSHT20



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

Image Credit: Tom Cridge

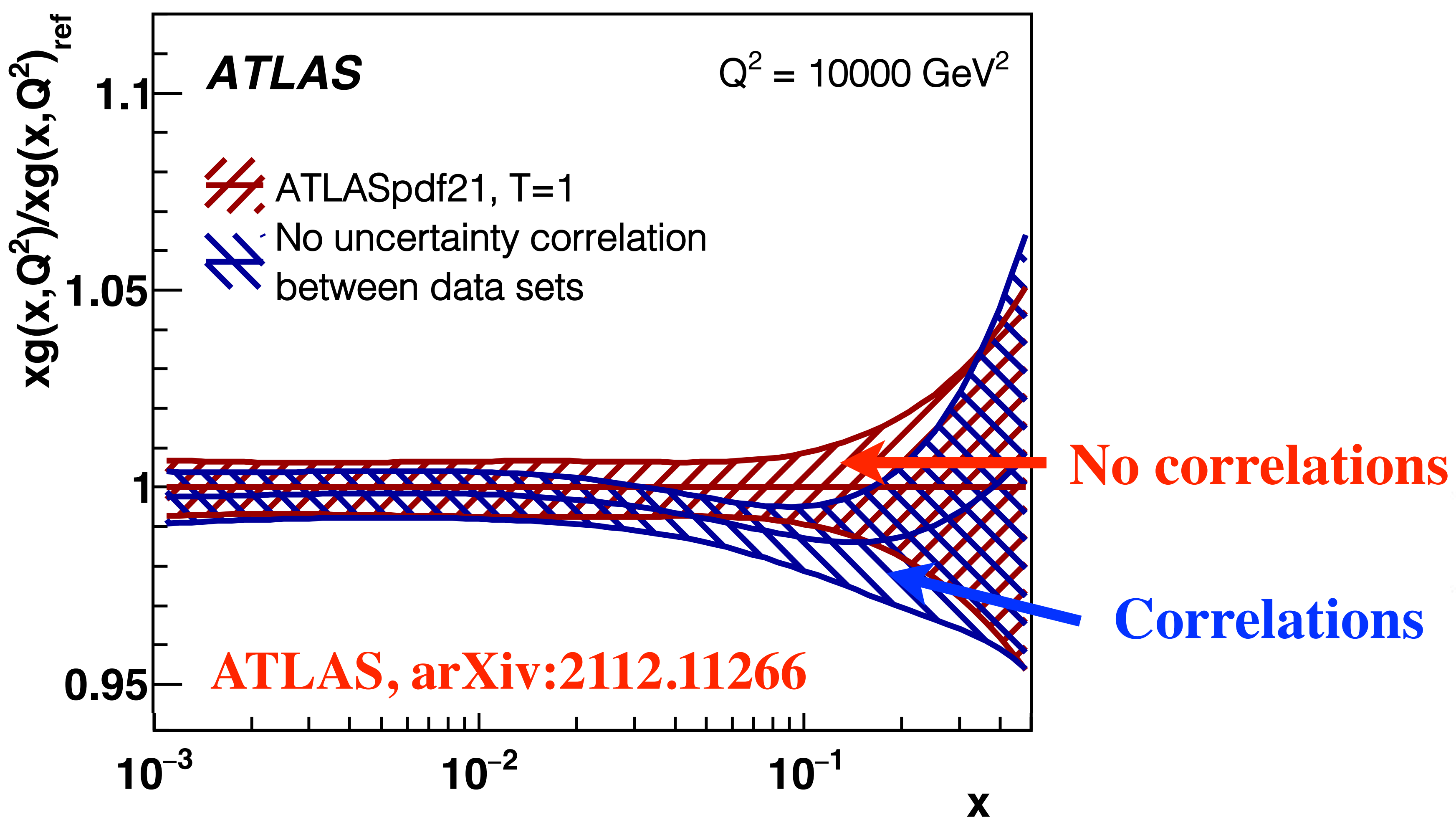
Precision fits - a challenging environment



X. Jing et al., arXiv:2306.03918

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

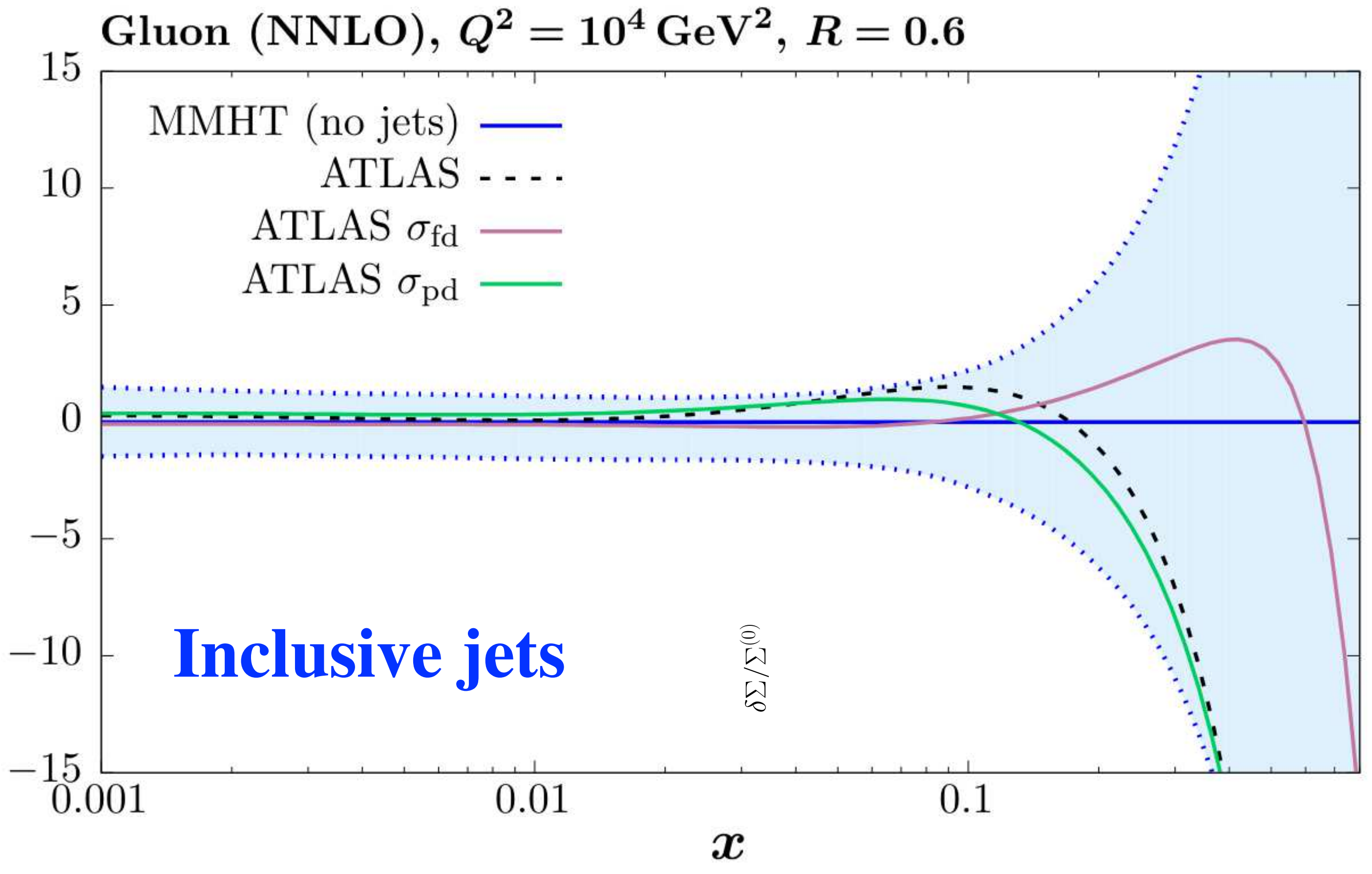
- ★ As LHC datasets become increasingly **systematics** dominated, sensitive to precise treatment of these, and their correlations as well as correlations between datasets!



No correlations

Correlations

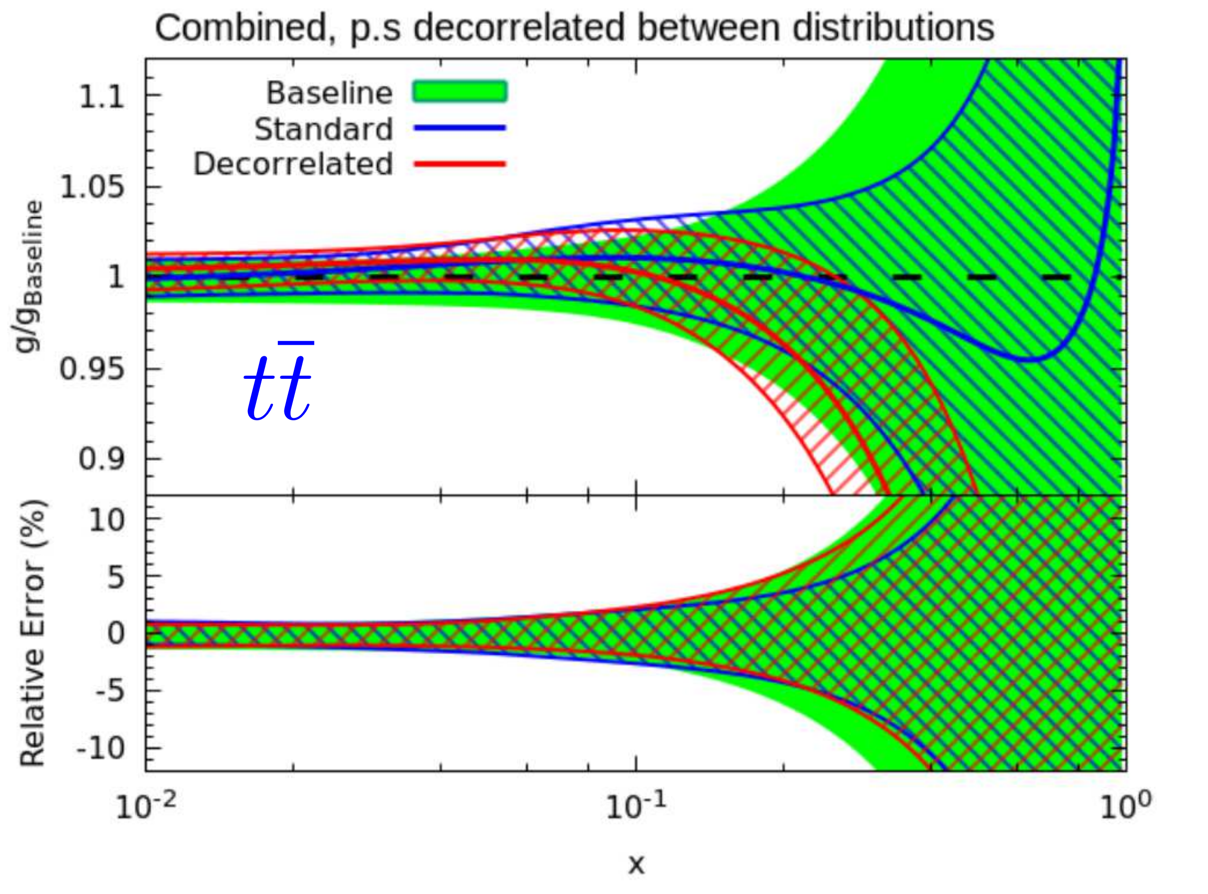
ATLAS, arXiv:2112.11266



Inclusive jets

S. Bailey et al., arXiv:2012.04684

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



S. Bailey and LHL, arXiv:1909.10541

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

Global PDF fits: parameterisation

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

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

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

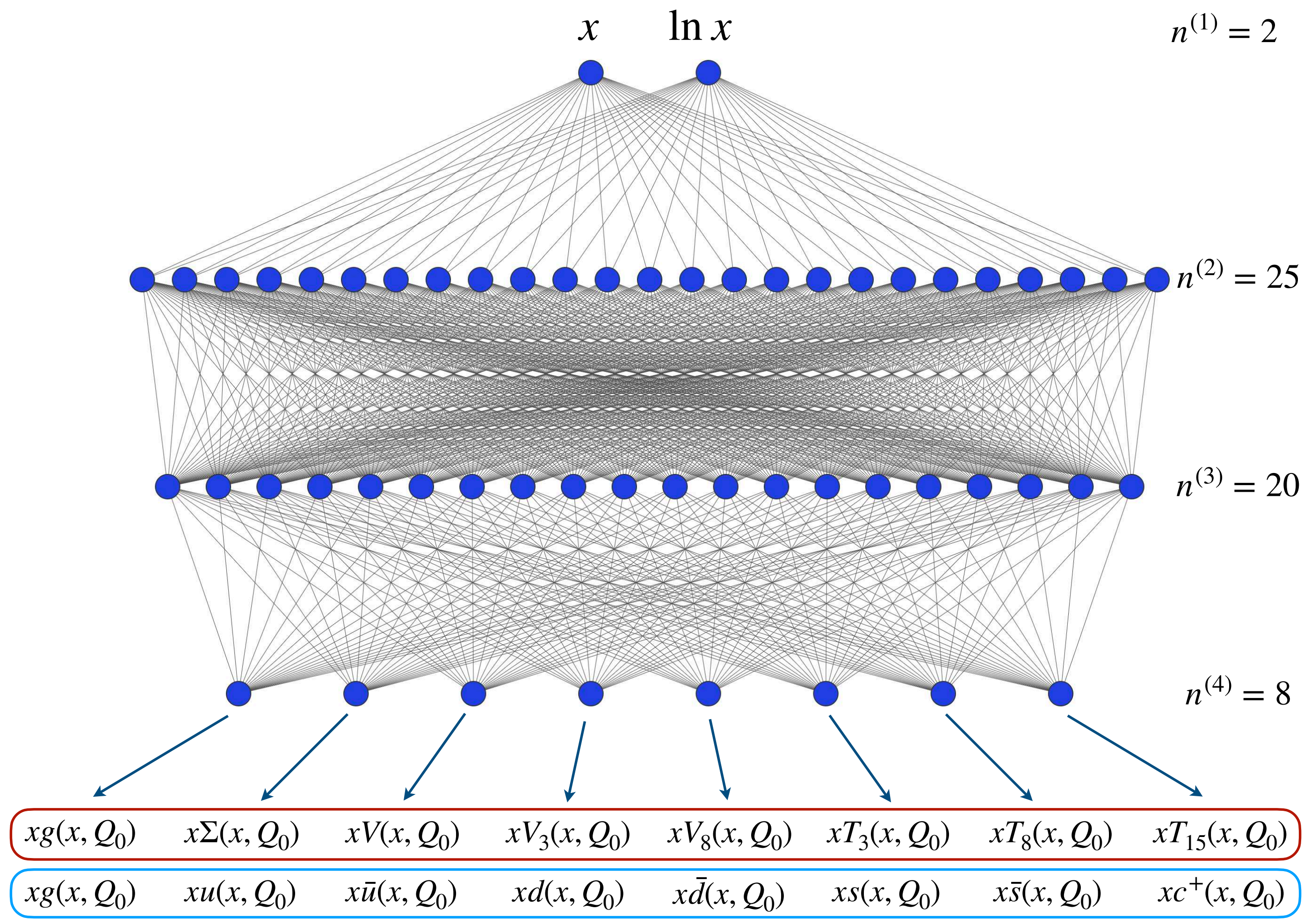
✦ MSHT20: **52** free parameters in terms of Chebyshev polynomials.

$$xf(x,Q_0)=Ax^{\delta}(1-x)^{\eta}\left(1+\sum_{i=1}^{6}a_iT_i(y(x))\right)\ ,$$

✦ CT18: **29** free parameters in term of Bernstein polynomials. Variations considered and PDF uncertainty expanded to account for these.

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

★ NNPDF4.0: ***763*** free parameter Neural Net.



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

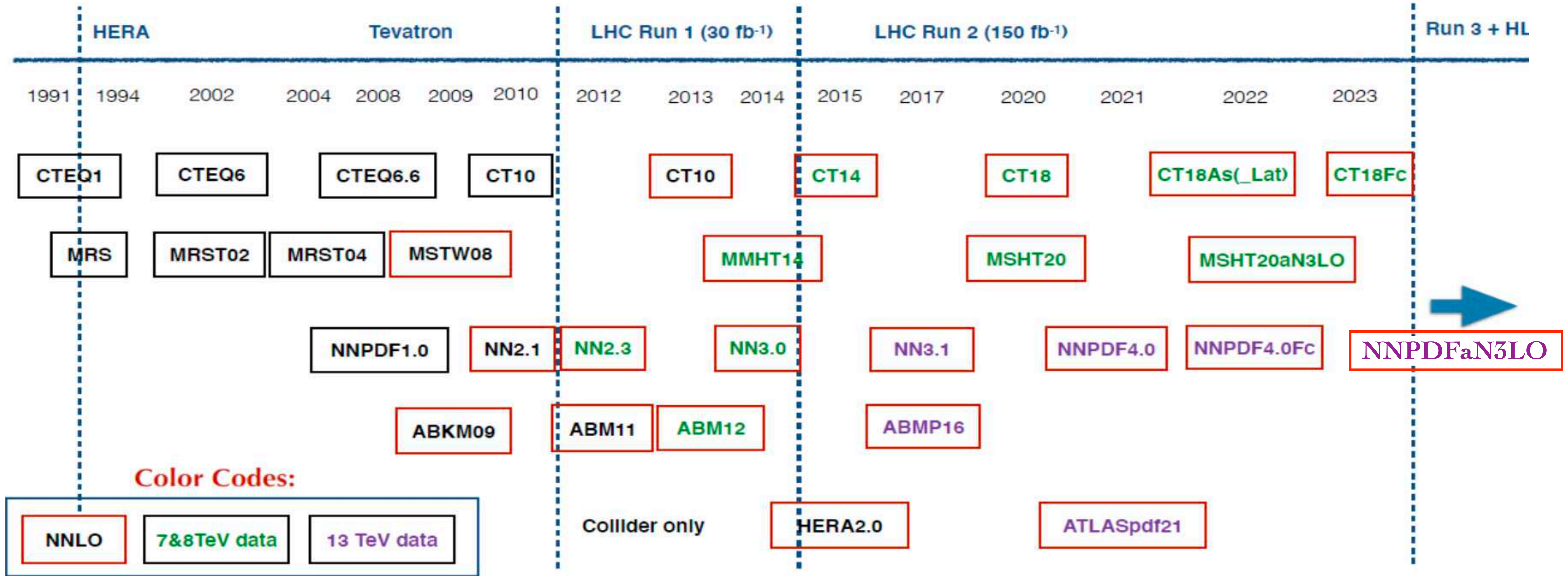
Major PDF Analyses

- 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



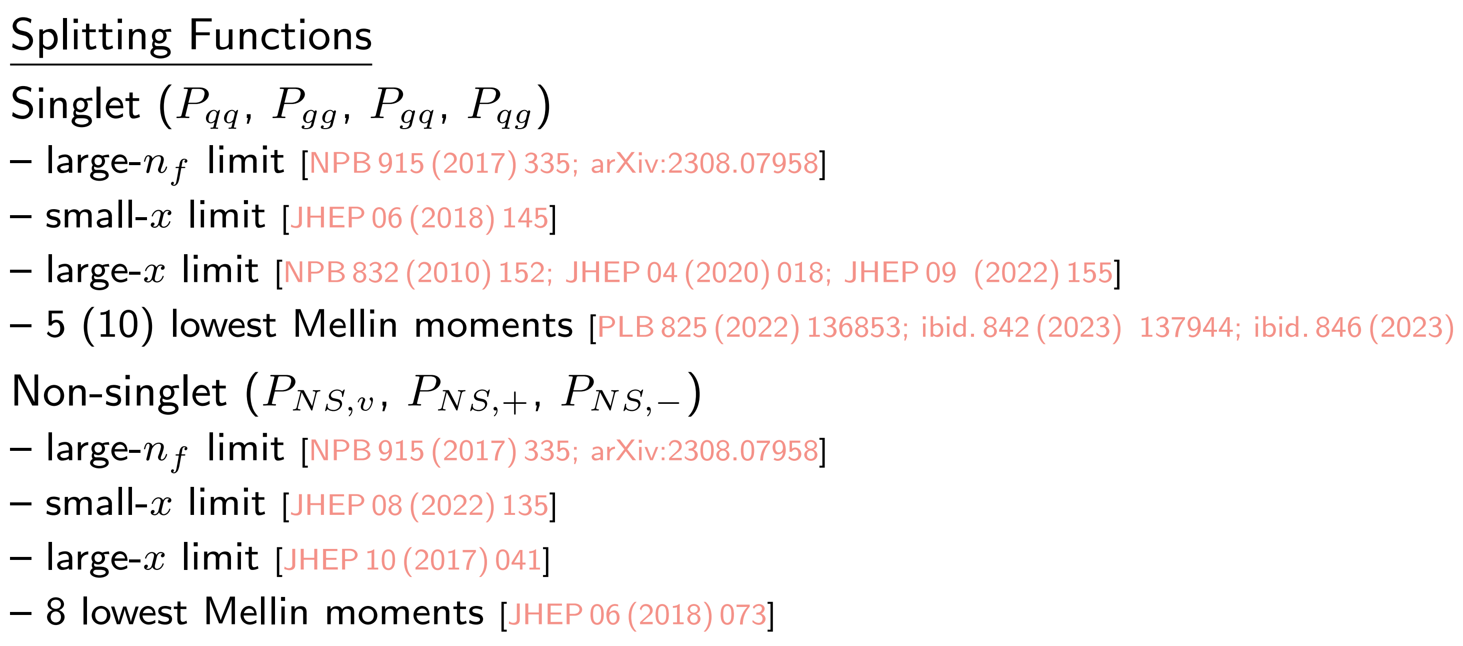
- These advances all build towards next generation of releases.

Image Credit:
Jun Gao

Theory updates : aN3LO and missing higher orders

- MSHT20aN3LO and NNPDF4.0aN3LO now go to approximate N3LO in QCD.
- Approximate ≠ poorly known! Much information about splitting functions + DIS. Little about hh cross sections (MHO uncertainty here).

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



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)

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]

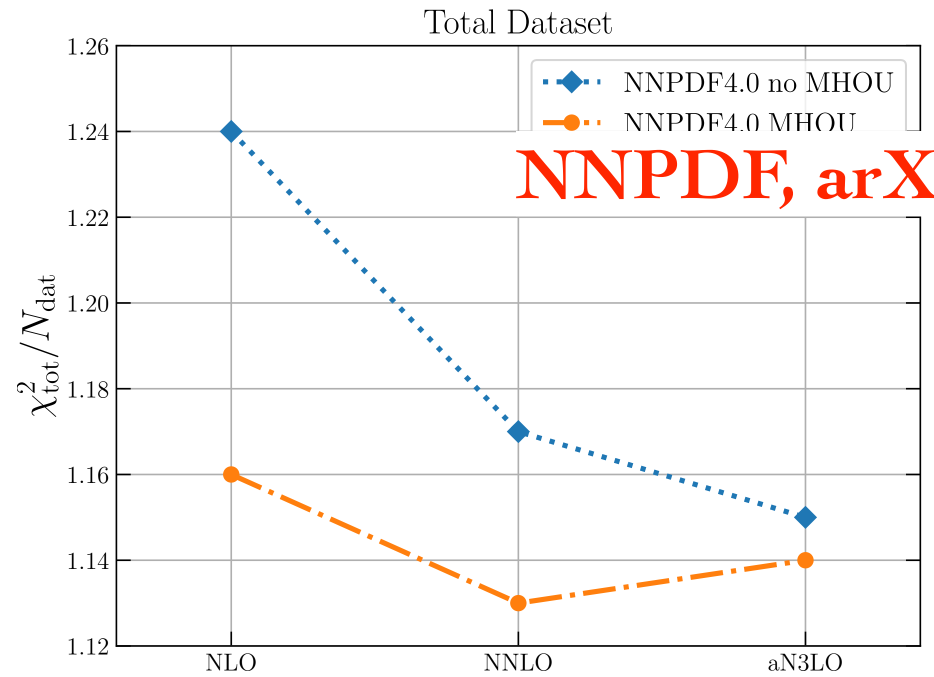
– large-x limit [JHEP 10 (2017) 041]

– 8 lowest Mellin moments [JHEP 06 (2018) 073]

- Gives improvement in fit quality…

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

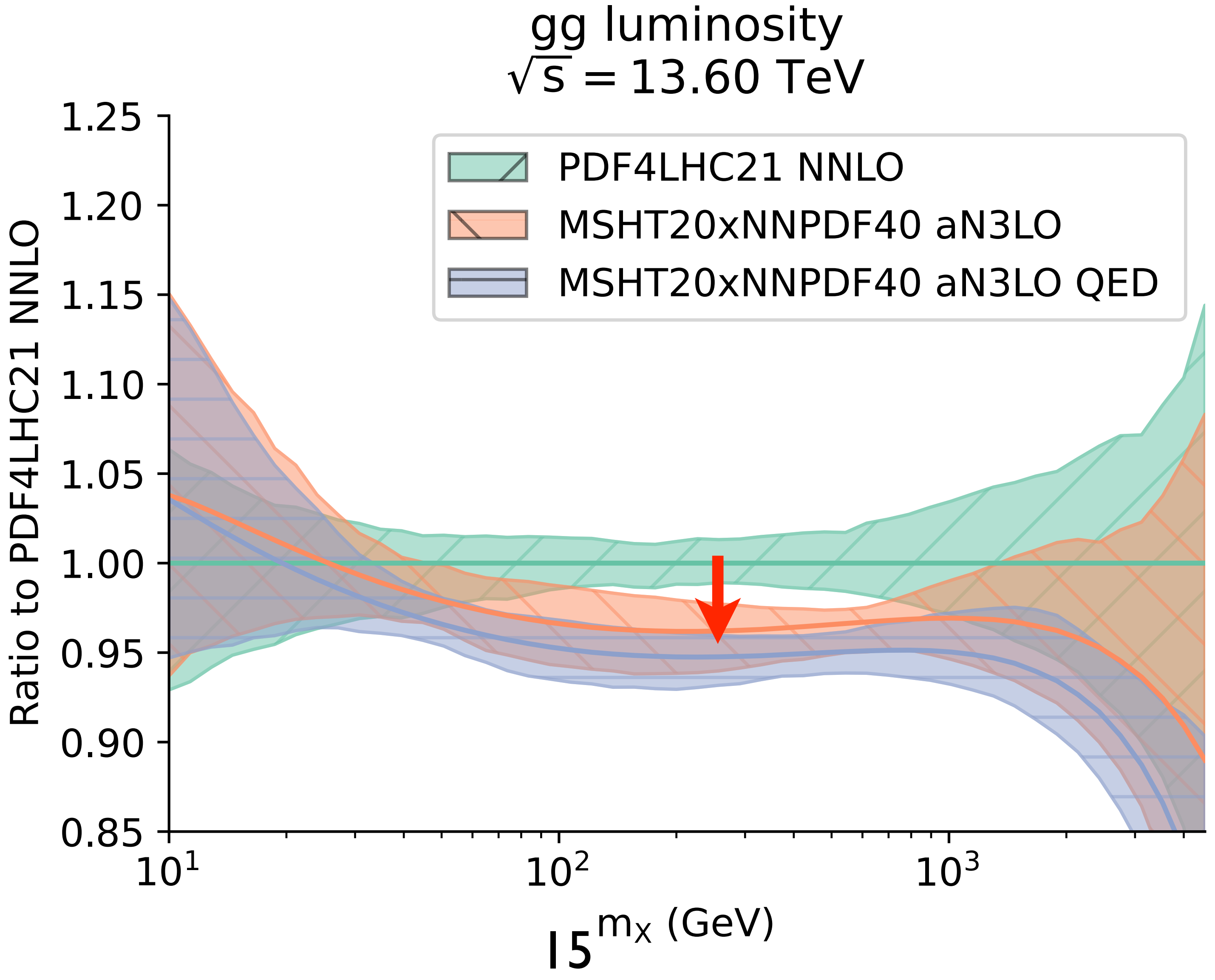
MSHT, arXiv:2207.04739

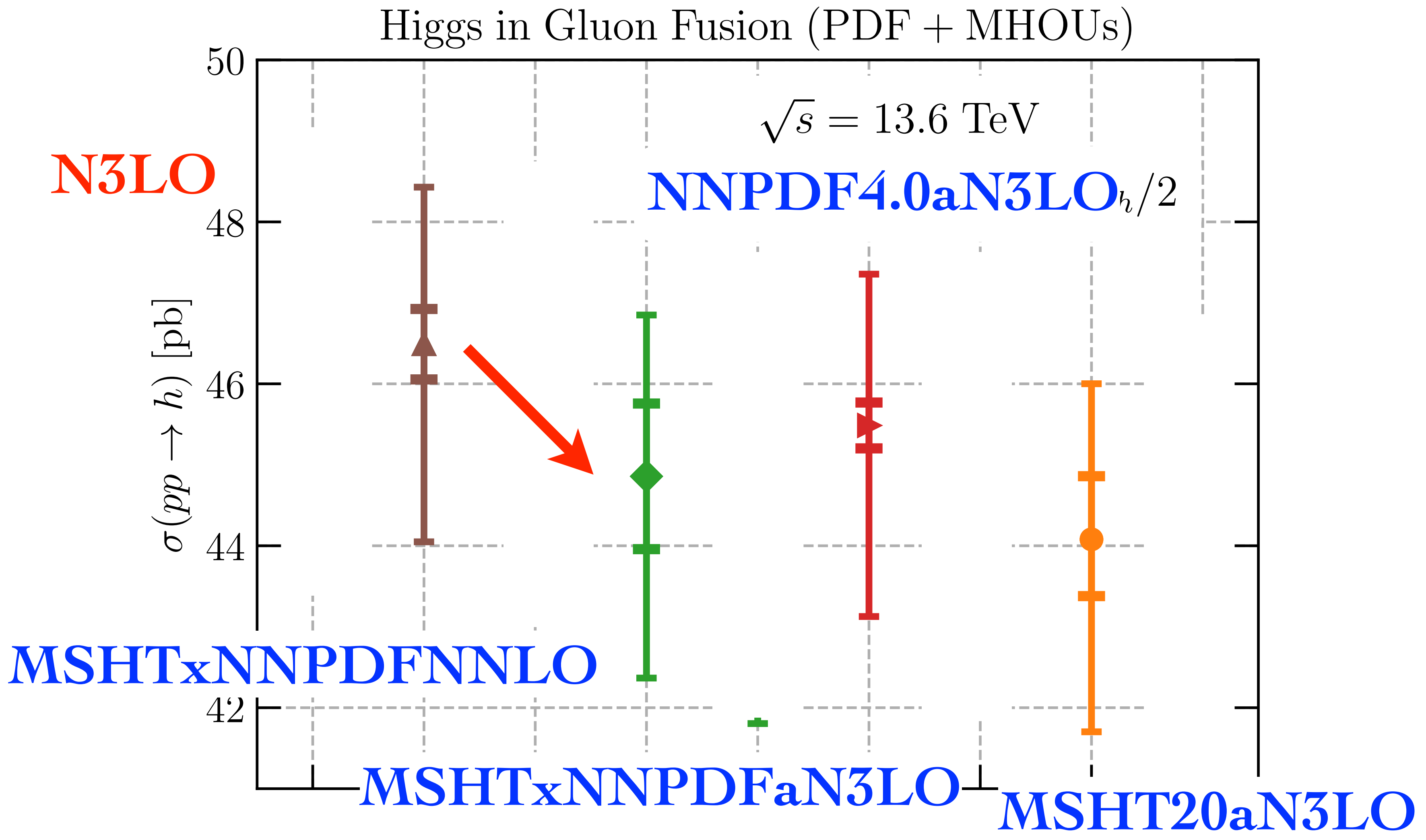


NNPDF, arXiv:2402.18635

- And moderate but non-negligible impact on PDFs with implications for e.g. ggH.

MSHT & NNPDF, arXiv:2411.05373



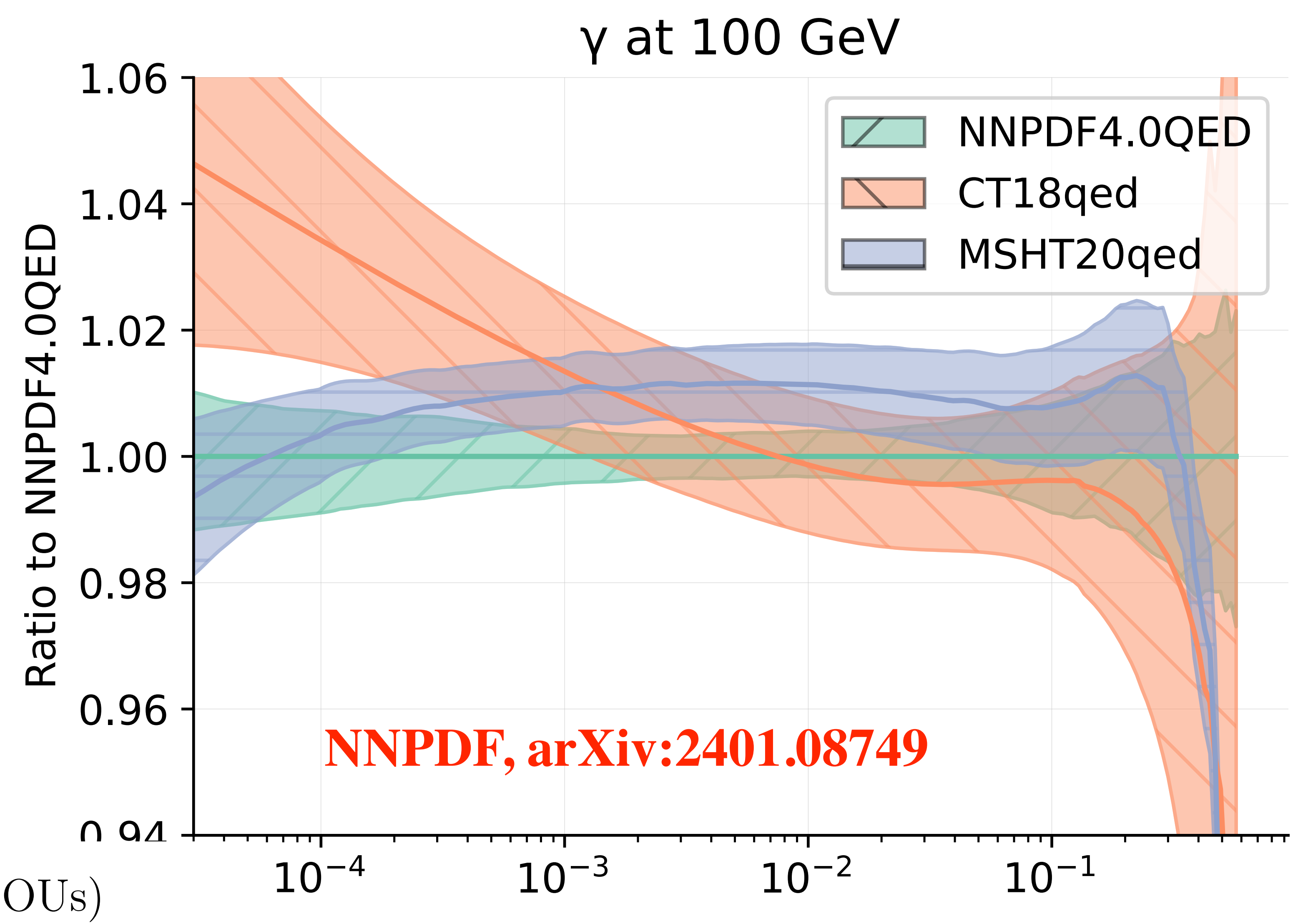


Theory updates : QED corrections

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

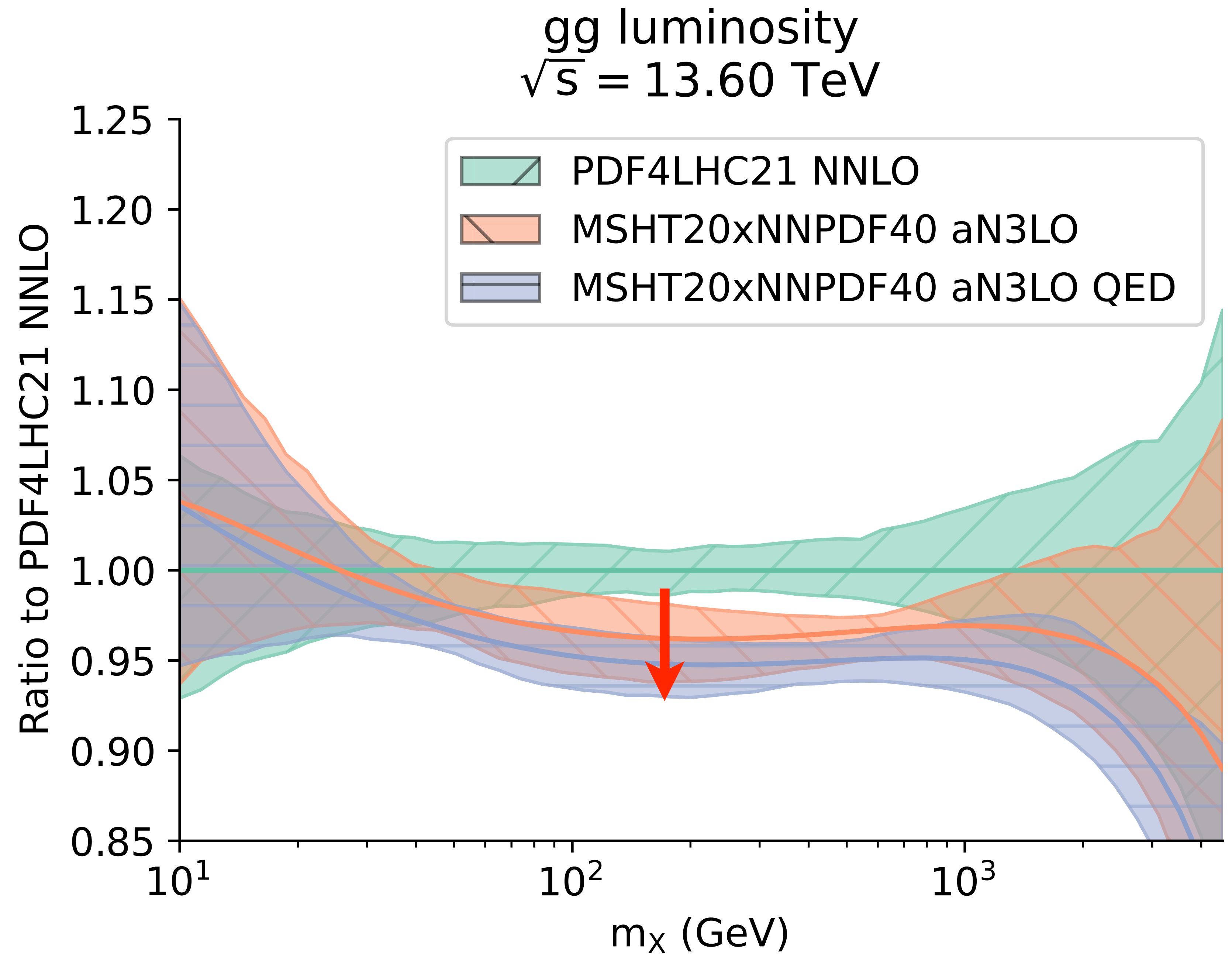
A. Manohar et al., JHEP 1712 (2017) 046

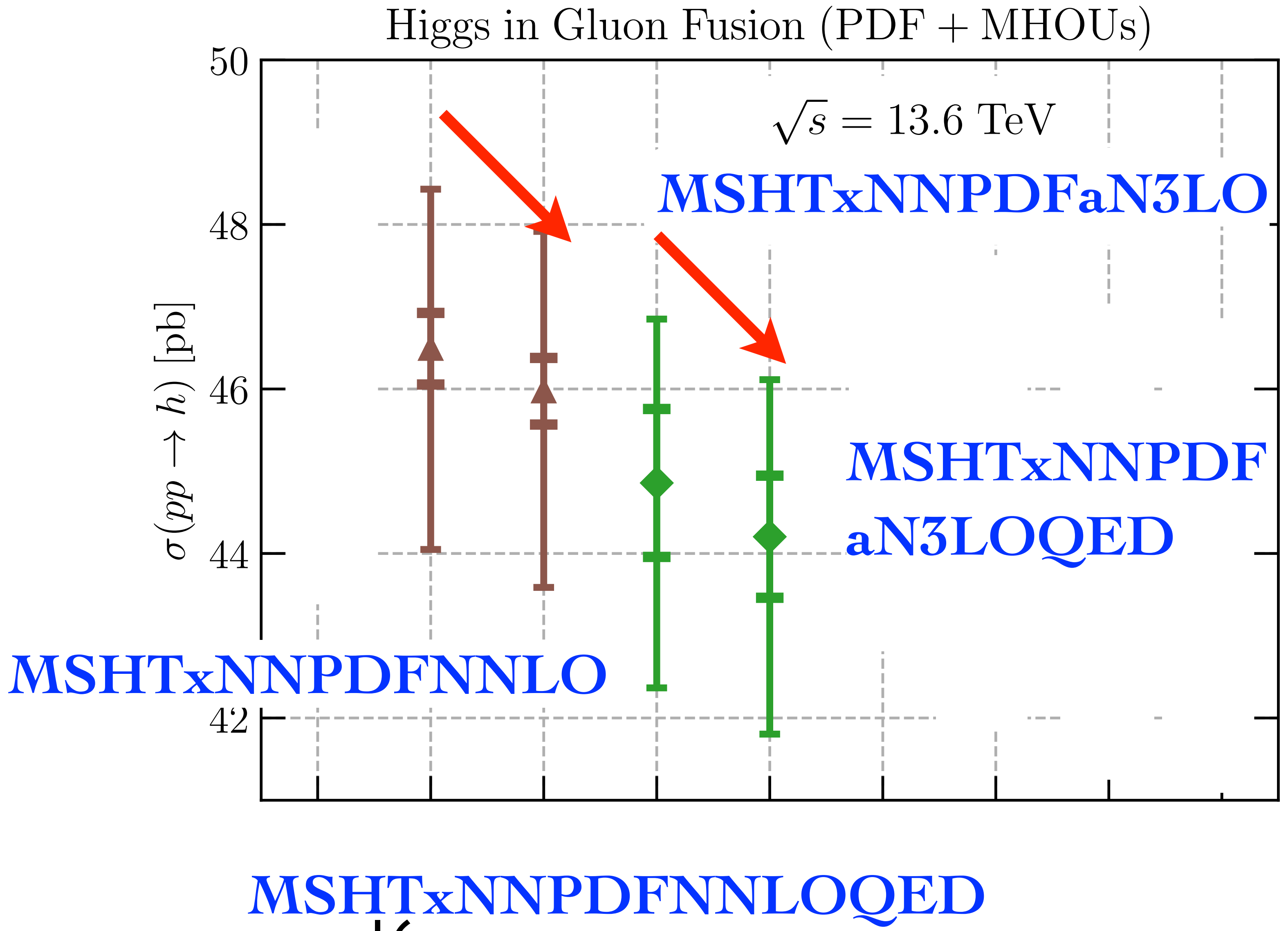
  - Now included in all fits, and in future will/should be by default.



NNPDF, arXiv:2401.08749

- Gives direct impact via photon-initiated production, but also indirect impact on QCD partons is non-negligble:





MSHTxNNPDFaN3LO

MSHTxNNPDF aN3LOQED

MSHTxNNPDFNNLO

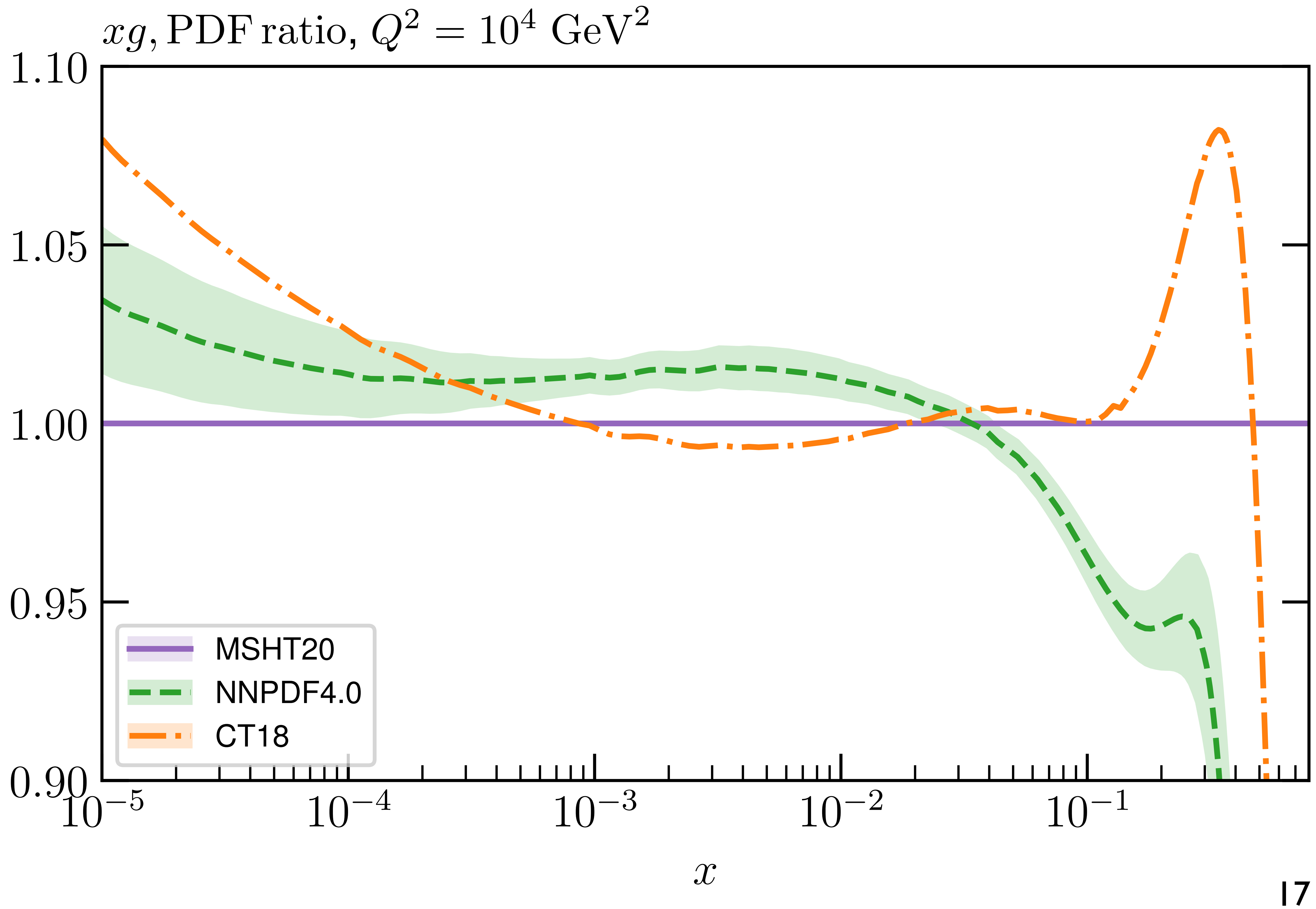
MSHTxNNPDFNNLOQED

- Universality of these corrections currently under investigation.

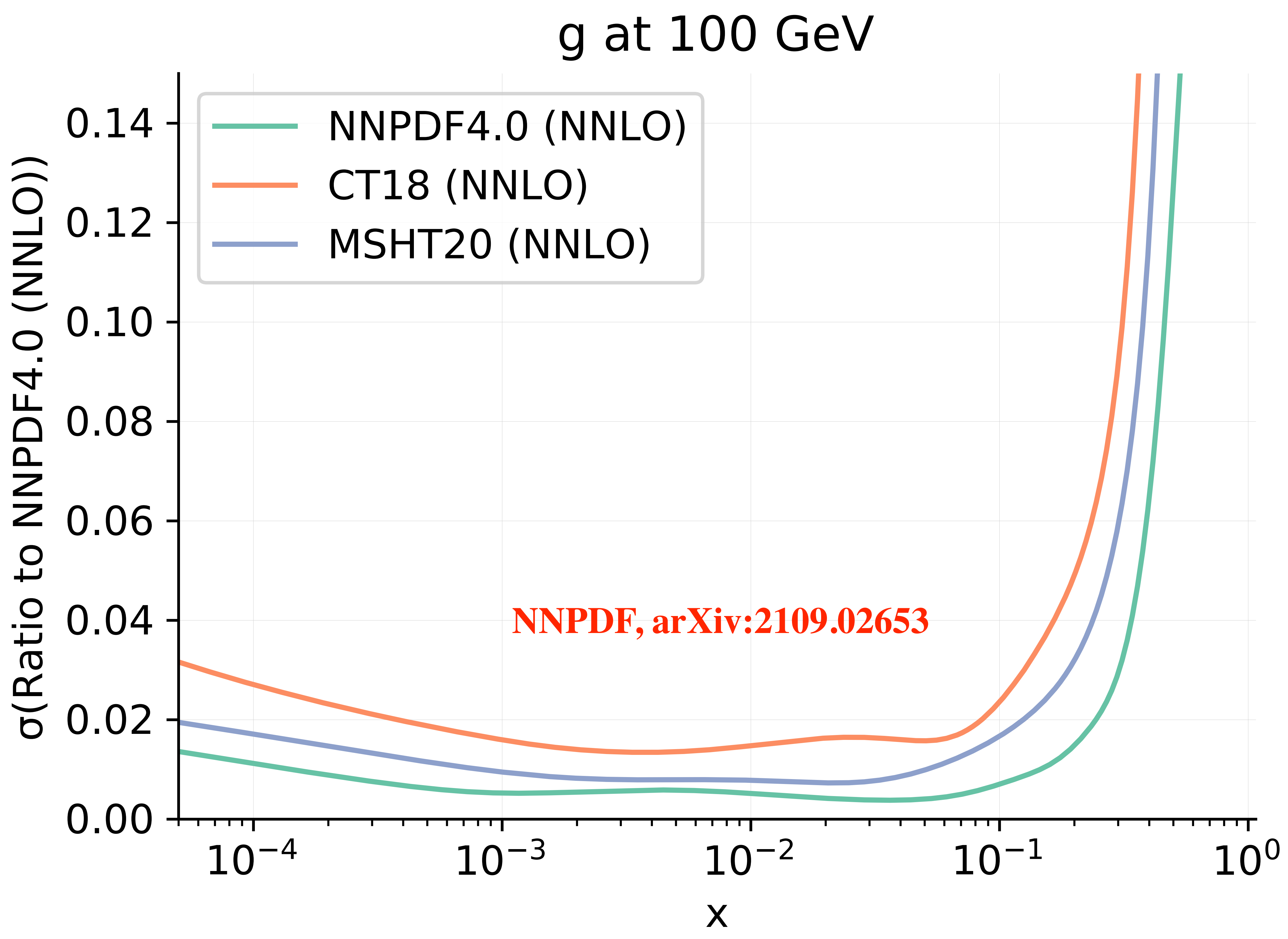
CT&MSHT joint study

Where do we stand?

- New data & theory updates in fits has clear impact - PDF uncertainties continue to reduce - and theoretical precision continues increase.
- But agreement between sets at level of PDF & uncertainties not always as good
- Due in part to data/theory inputs but also methdology.



xg, PDF ratio, $Q^2 = 10^4$ GeV^2



g at 100 GeV

NNPDF, arXiv:2109.02653

PDF uncertainty quantification: comments and recent results

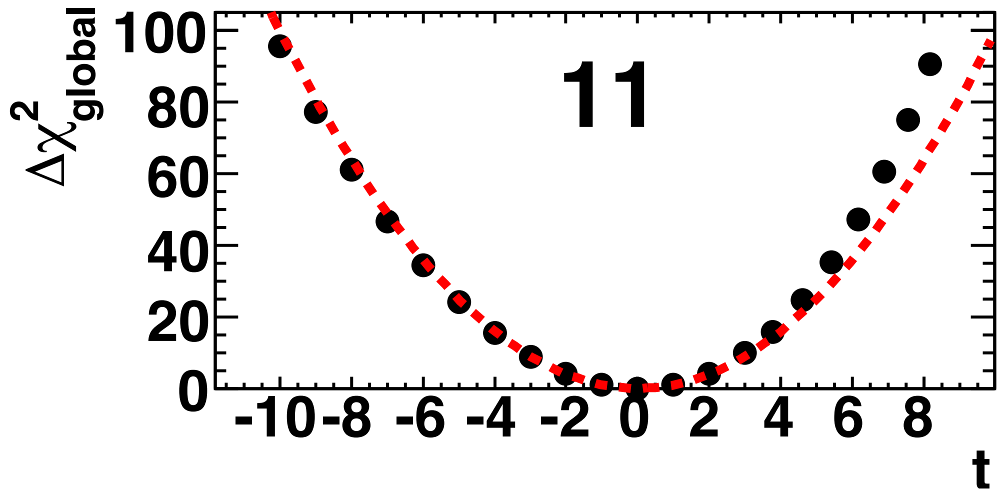
PDF Uncertainties

$$\text{Data} = \text{PDF} \otimes \sigma_H \longrightarrow f(x,\mu) \pm \Delta(x,\mu) \ ?$$

- Different approaches to PDF uncertainties taken by each group:

★ Fixed parameterisation (MSHT/CT): $\chi^2_{\text{global}} \sim \frac{(D_{\text{ata}} - T_{\text{heory}})^2}{\sigma^2}$ $H_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2_{\text{global}}}{\partial a_i \partial a_j}\Big|_{\text{min}}$

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



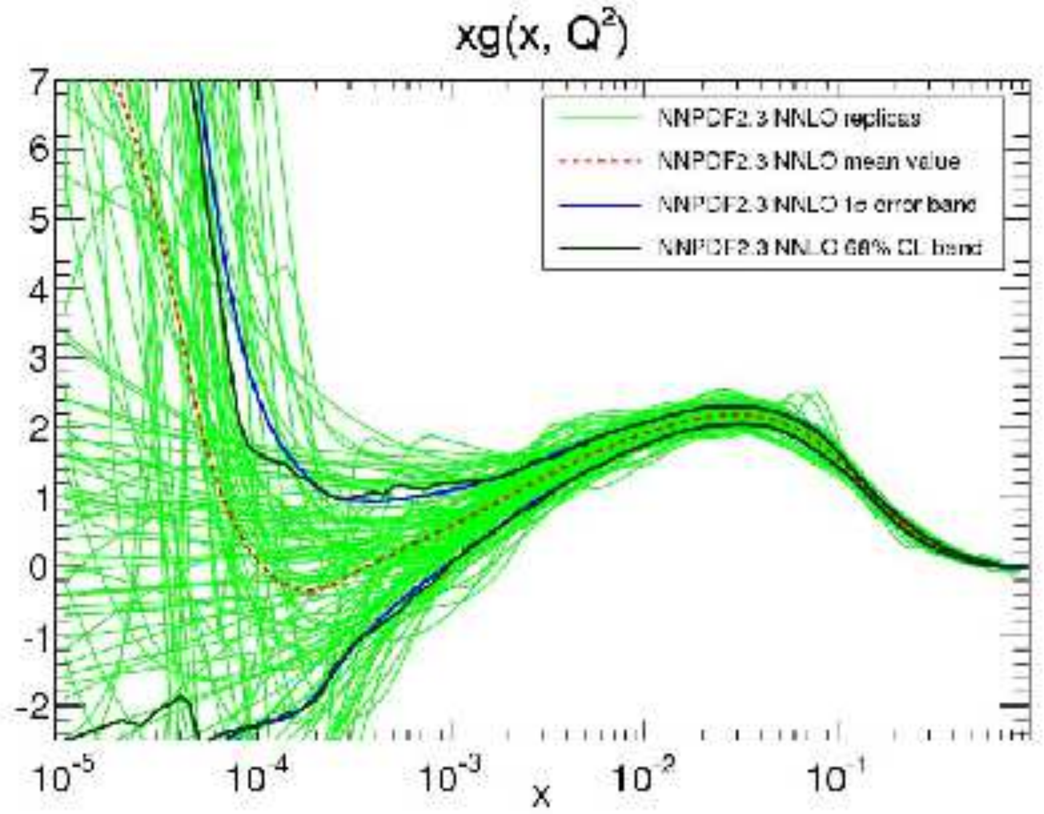
✦ Parameter shifts corresponding to given $\Delta\chi^2$ criteria given in terms of these

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

where ‘**tolerance**’ $T \neq 1$ in general for 68% C.L. uncertainties.

★ NNPDF: ✦ Generate set of MC `replicas’ by shifting data by errors.

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

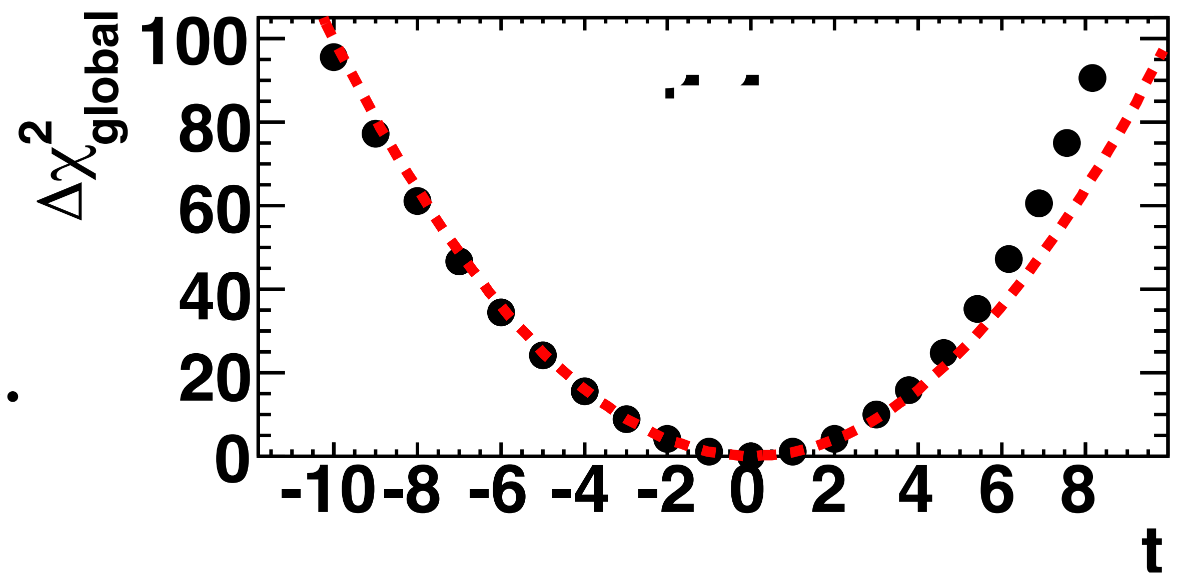


G. Watt and R. Thorne, arXiv:1205.4024

✦ 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? **See Pavel's talk**
  - $T = 1$: `textbook' criterion for 68% C.L., would apply if: $N_{\text{dataset}} \sim 50 - 60$
    - ★ Complete statistical compatibility between multiple datasets entering fit.
    - ★ Completely faithful evaluation of experimental uncertainties within each dataset.
    - ★ Theoretical calculations that match these exactly.



  - Good evidence that first two points do not always hold, while last point known not be true (though much recent progress on 3rd point).
  - **Simple observations:** global fit quality bad strictly speaking, distribution amongst datasets not as expected.

$$\frac{\chi^2}{N_{\text{pts}}} \gg 1 + \sigma(N_{\text{pts}}) \sim 1.02$$

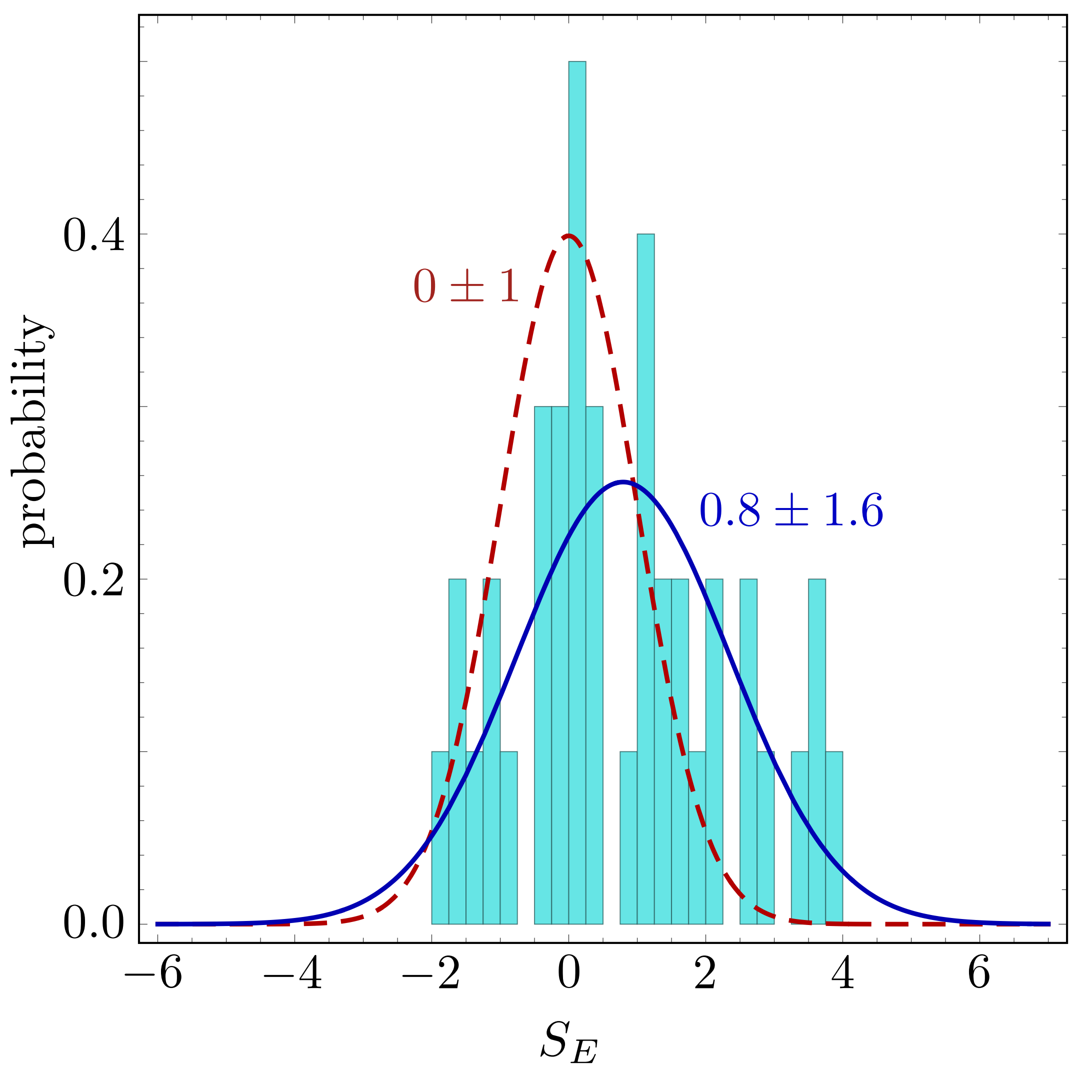
$$N_{\text{pts}} \sim 4000 - 5000$$

MSHTaN3LO

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

And similar for other fits!

NNPDF3.1 NNLO (40 data sets)



$$S_E = \sqrt{2\chi^2} - \sqrt{2N - 1}$$

K. Kovarich et al., arXiv.1905.06957

G. Watt and R. Thorne, arXiv:1205.4024

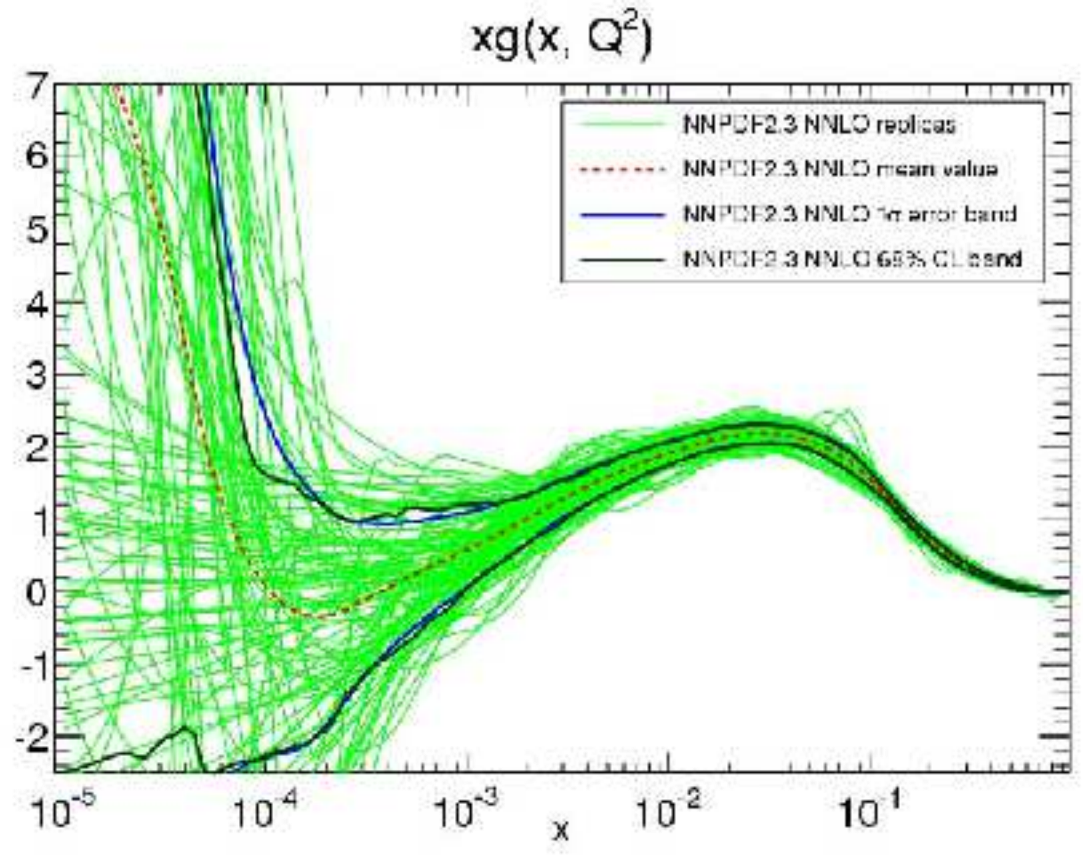
M. Yan et al., arXiv.2406.01664

J. Pumplin, arXiv:0909.0268

- 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 $T > 1$. Procedure for choosing this 'tolerance' T differs between MSHT/CT.

Backup

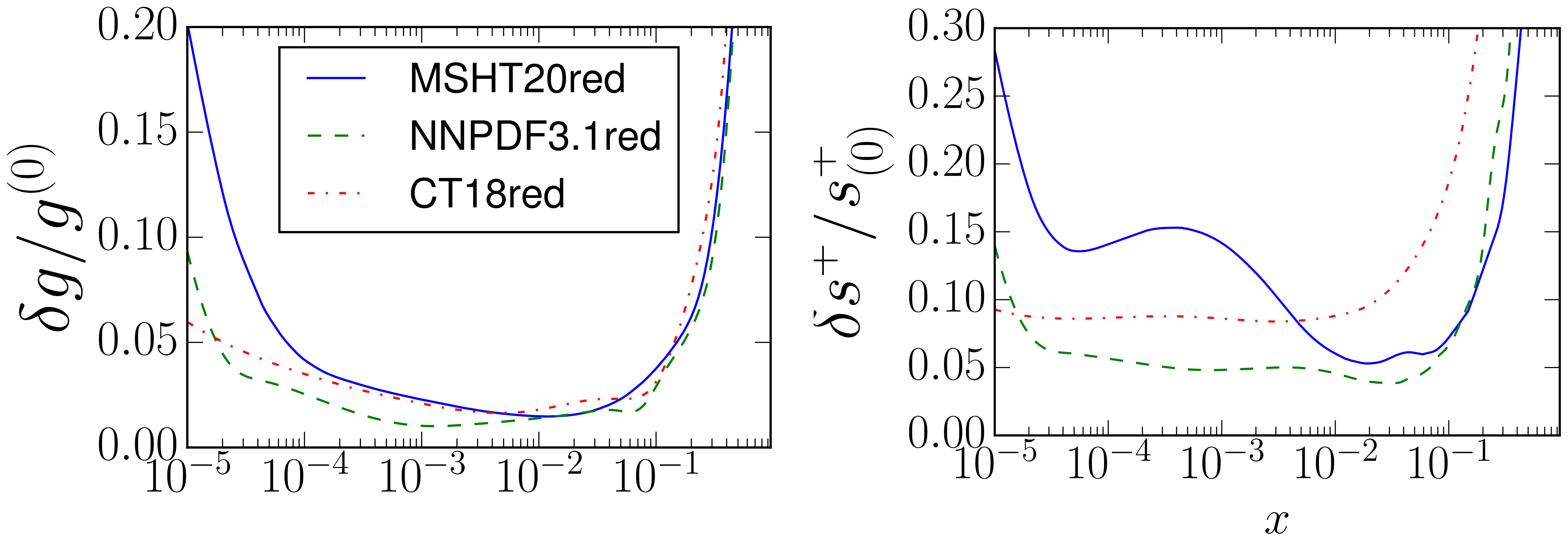
- However no such **explicit** tolerance in NNPDF procedure. Implicitly what is going on less clear, given complex nature of NN.
- What we do know is that in previous benchmarking NNPDF uncertainties generally smallest for ~ same data and theory in fit!



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

Benchmark = similar data/settings

PDF4LHC21, arXiv:2203.05506



- With larger difference again expected in more recent NNPDF4.0 fits

★ Suggests three possibilities:

1. **NN** uncertainty not conservative enough (**too small**).
2. Fixed par. (**MSHT)** uncertainty too conservative (**too large**).
3. Fixed par. (**MSHT)** fit less accurate, due to parameterisation inflexibility, and hence enlarged errors needed (**less precise**).

- Understanding which essential to reach true accuracy. What to do?

Closure Tests

- First, to answer third point:

★ Suggests three possibilities:

1. **NN** uncertainty not conservative enough (**too small**).
2. Fixed par. (**MSHT**) uncertainty too conservative (**too large**).
3. Fixed par. (**MSHT**) fit less accurate, due to parameterisation inflexibility, and hence enlarged errors needed (**less precise**).

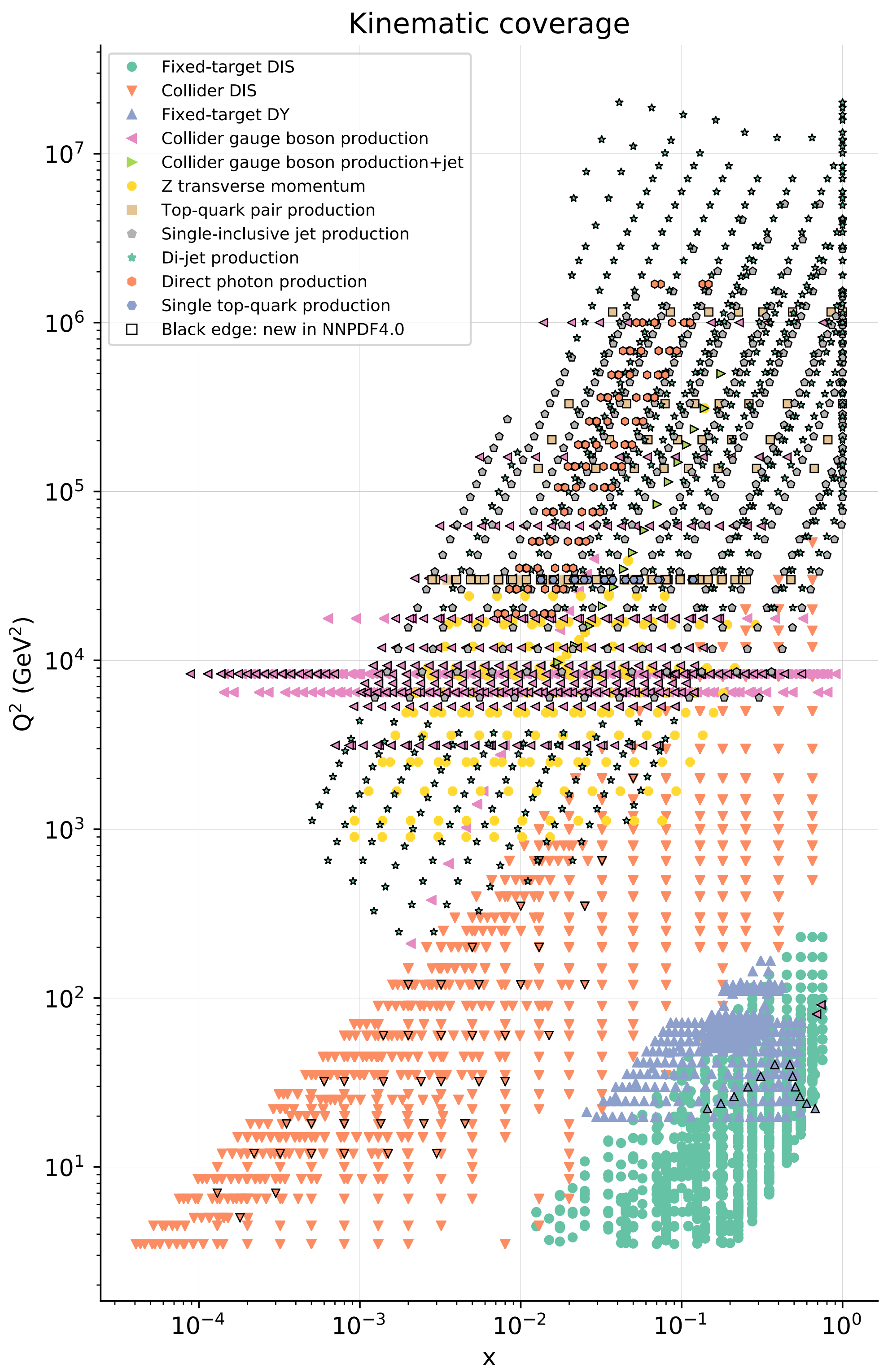
- Perform **closure test:** generate pseudodata = theory predictions with given input PDF set corresponding to global PDF fit (in this case NNPDF4.0).

$$\sigma_{PD} = T + \delta$$

- Then perform fit to this but with MSHT20 parameterisation. True values known - can test agreement and hence faithfulness of MSHT.

- How do we do?

NNPDF, arXiv:2109.02653

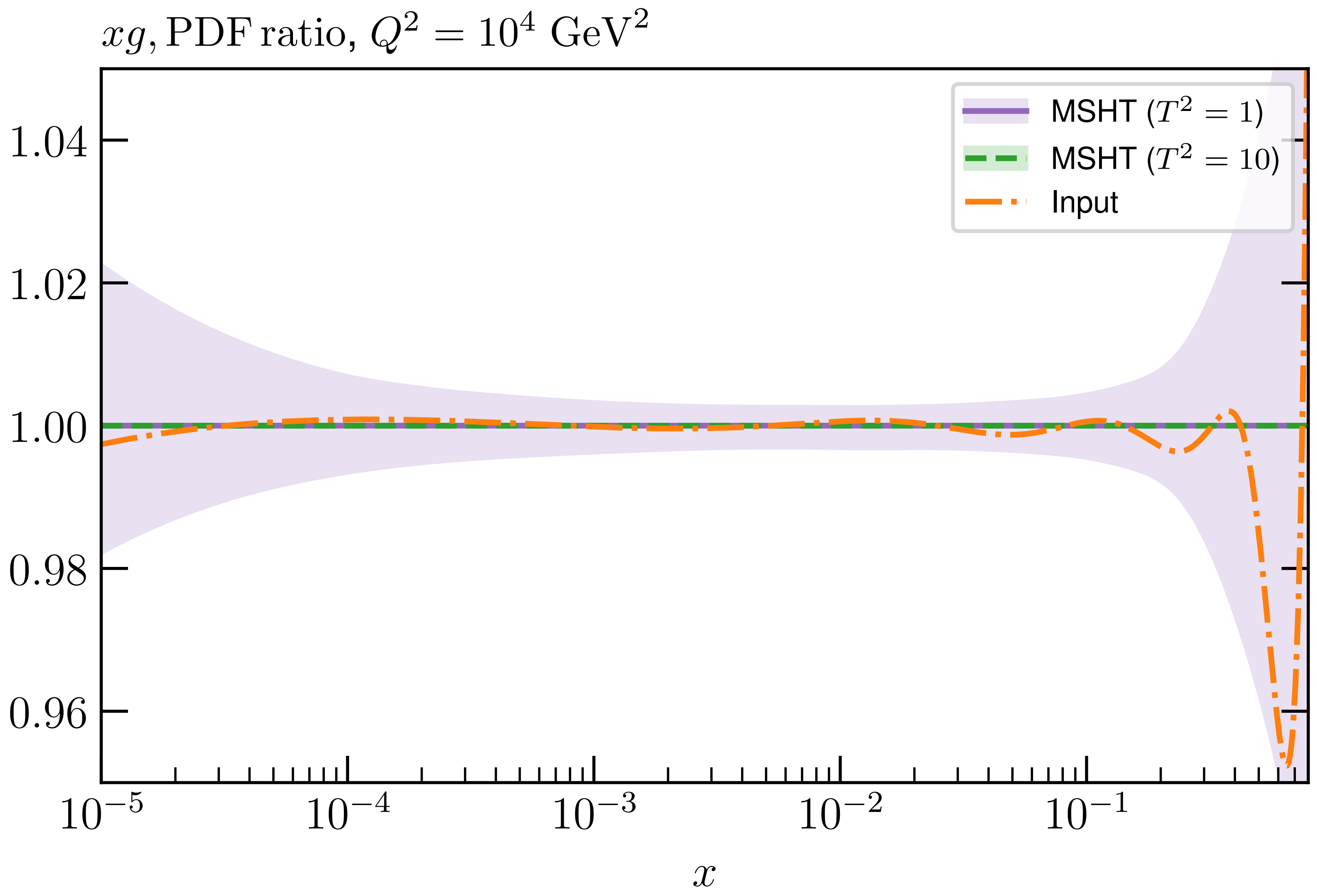


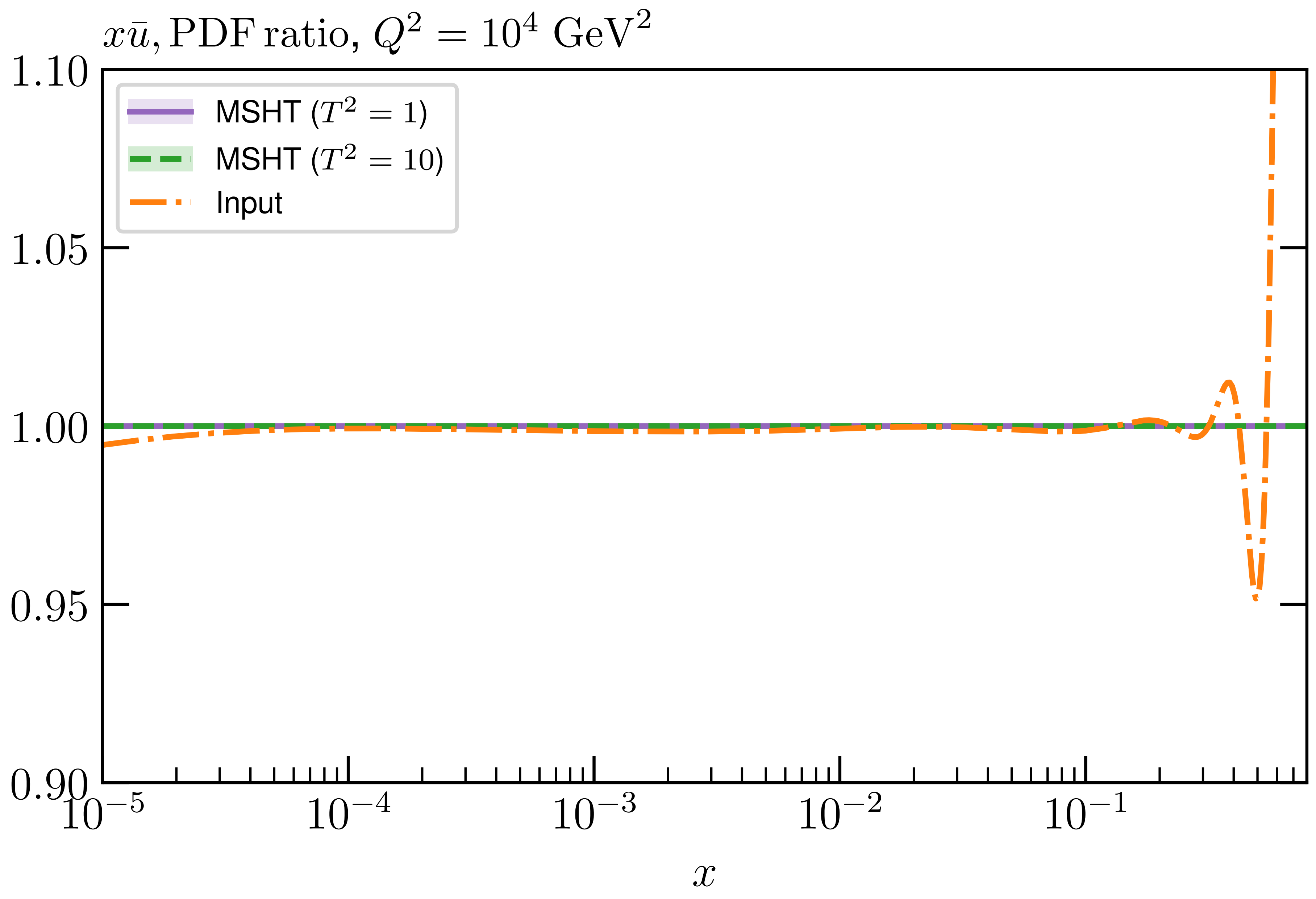
Always NNLO, NNPDF4.0 pch as input

- Generate unshifted pseudodata for 4.0 global dataset ($N_{\text{pts}} = 4627$). In principle exact agreement possible, with $\chi^2 = 0$. Won't get exactly due to finite parameterisation. Find:

	χ^2	χ^2/N_{pts}
Fit quality:	2.4	0.0005

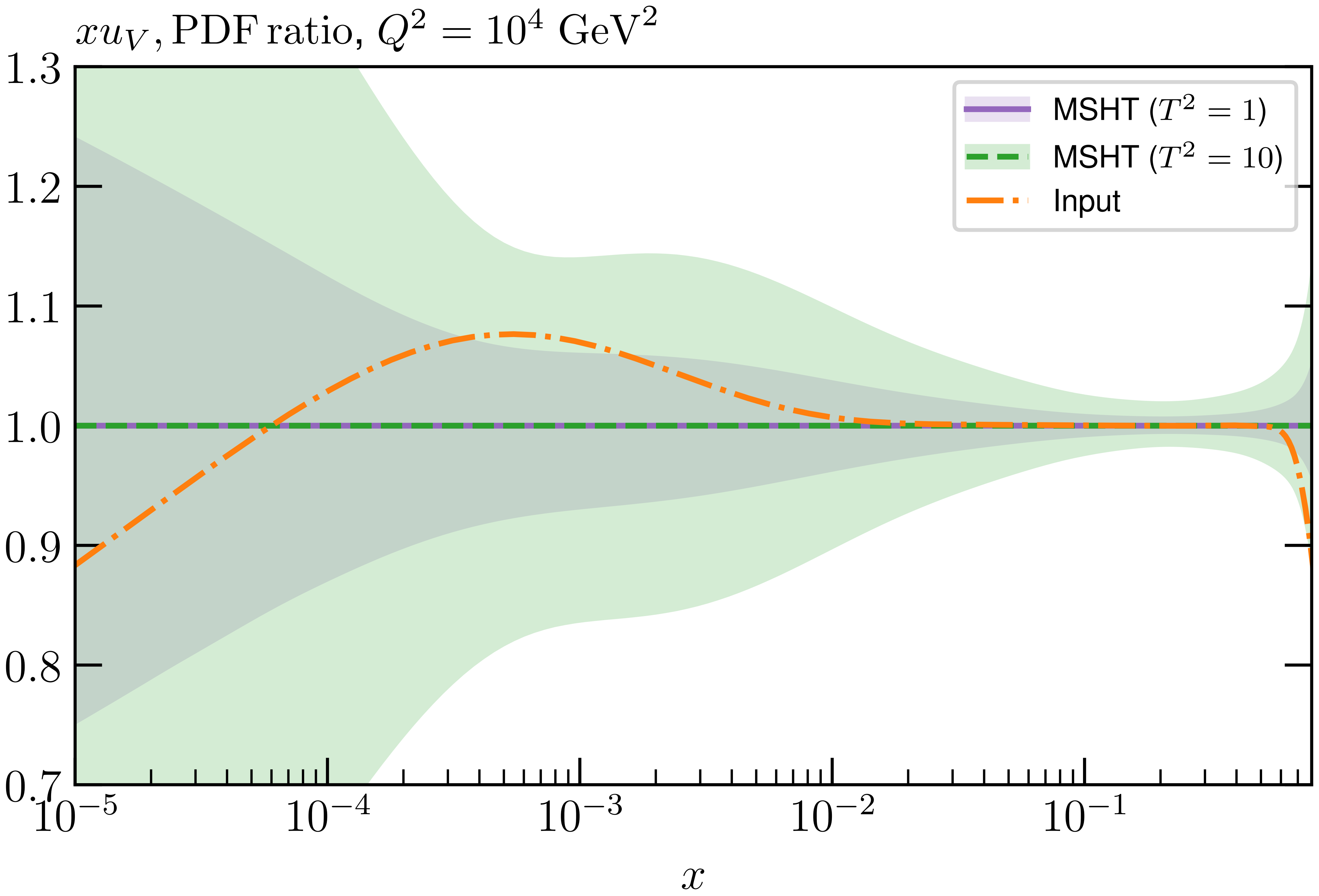
- Looks very good! How about PDFs?

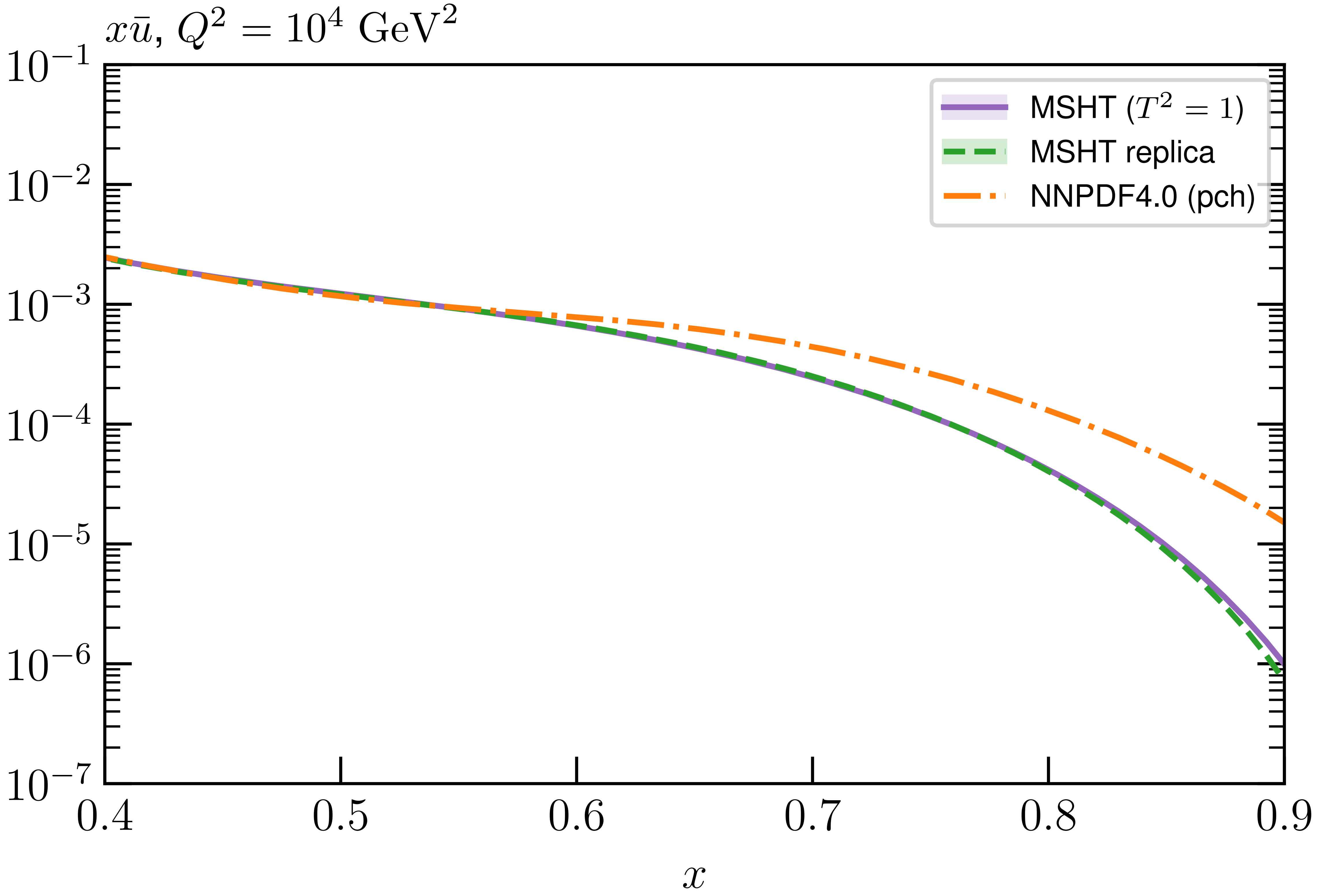




$\longrightarrow$ In **data region** input PDF matched very well, and much better than $T^2 = 1$ uncertainties. **No immediate evidence** that the increased tolerance is driven by **parameterisation inflexibility** for MSHT. Similar results found for more complete 'fluctuated' set of closure tests.

Backup

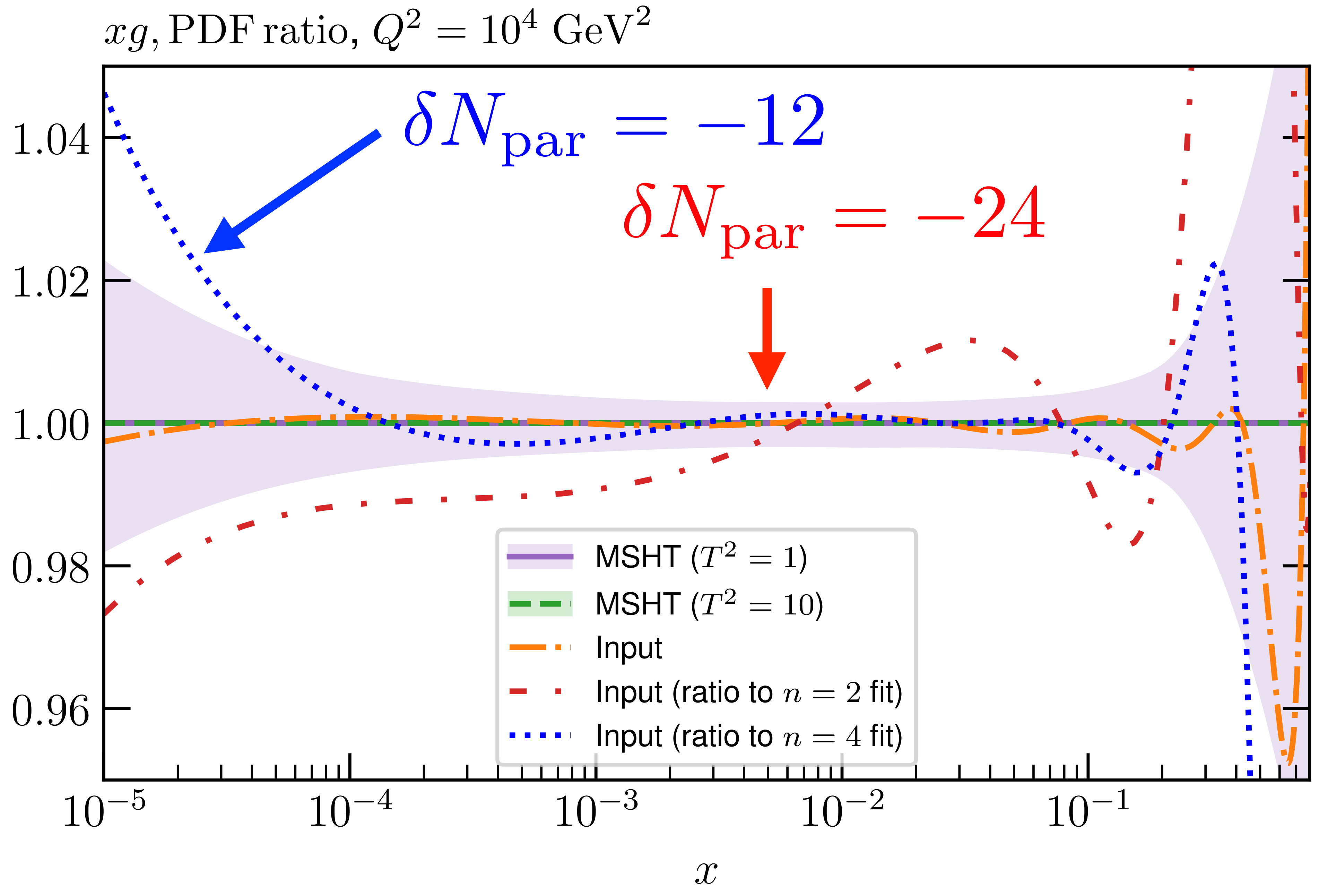




- In less well constrained regions deviation larger, e.g for u_V, d_V at low and high x and the $\overline{u}, \overline{d}$ at high x .
- Hence in extrapolation region input not always consistent within uncertainties
- As ~ outside data region not inconsistent (errors driven by data), but indicates more conservative error definition in these regions may be desirable (as tends to happen in NN approach).
- At high x MC replica uncertainty gives more conservative uncertainty.
- Though arguably no ‘right’ answer in true extrapolation region (too conservative vs. over-conservative).

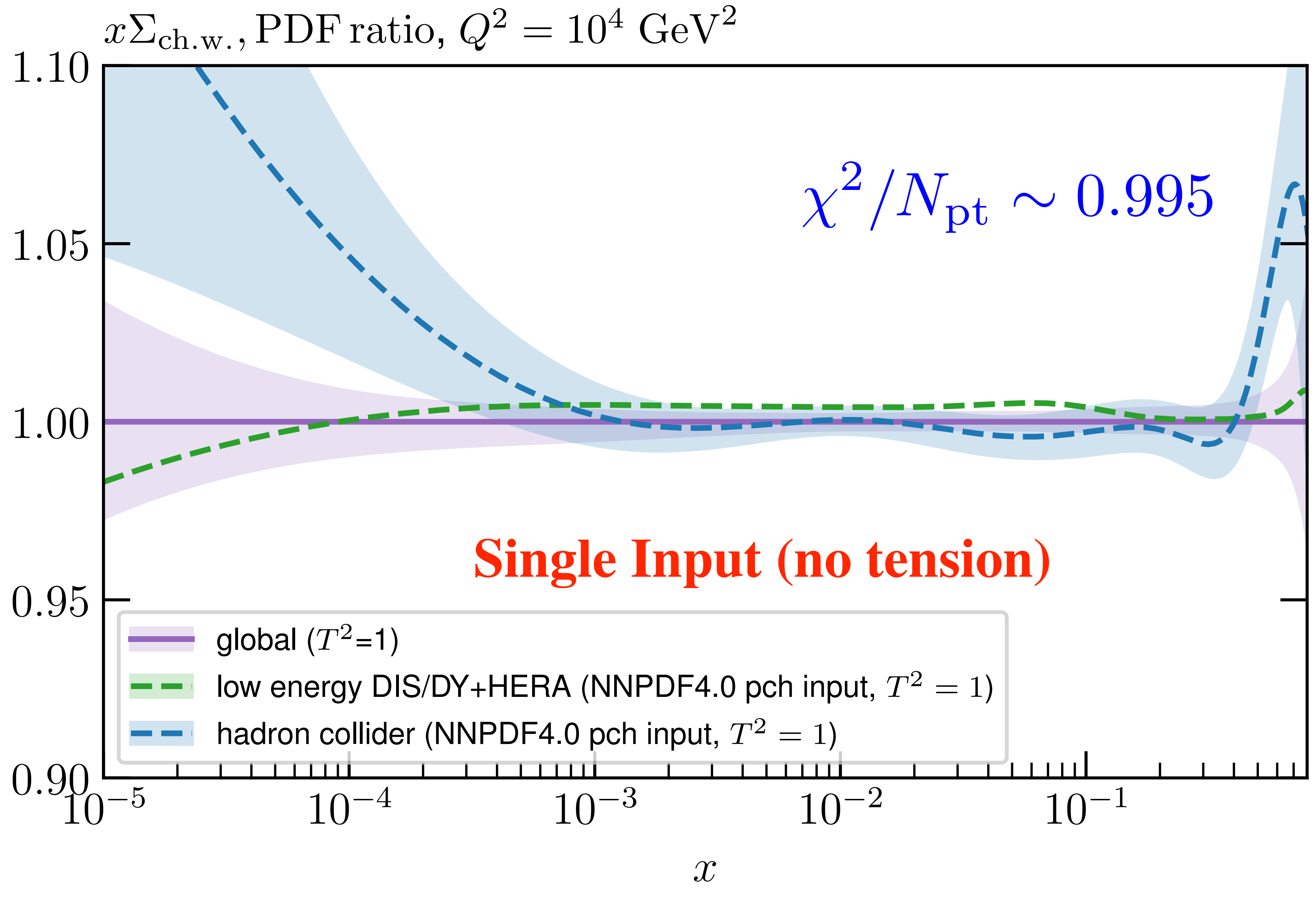
- Can play all sorts of games with this (no time to go into here). In brief:

  ★ **Restrict** number of free PDF parameters - start to fail closure test (as expected!). Guide to number of free PDF parameters needed.

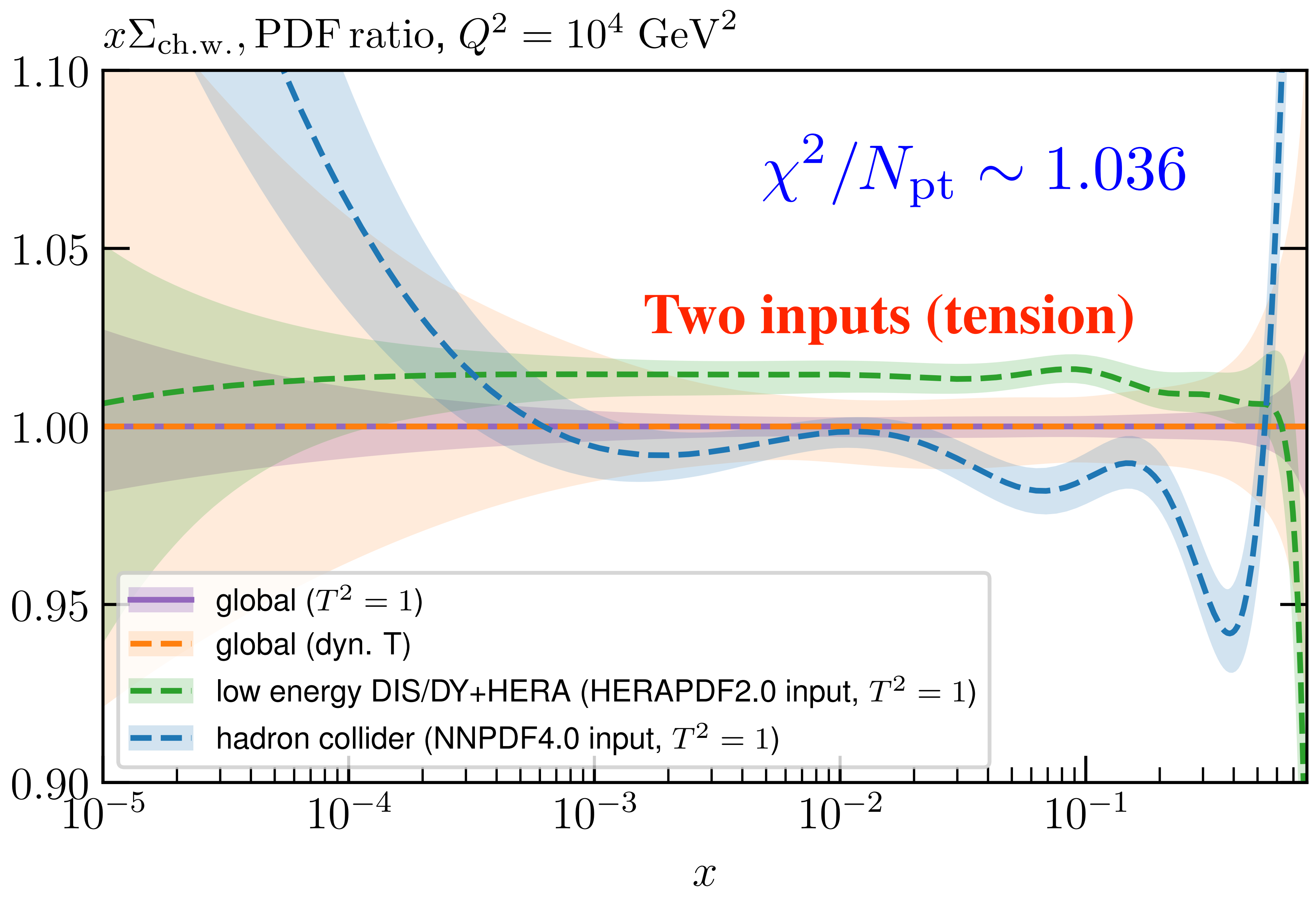

$\delta N_{\rm par} = -12$

$\delta N_{\rm par} = -24$

★ Inject **tension** into closure test - break (pseudo)dataset up into two.


$\chi^2/N_{\rm pt} \sim 0.995$

Single Input (no tension)


$\chi^2/N_{\rm pt} \sim 1.036$

Two inputs (tension)

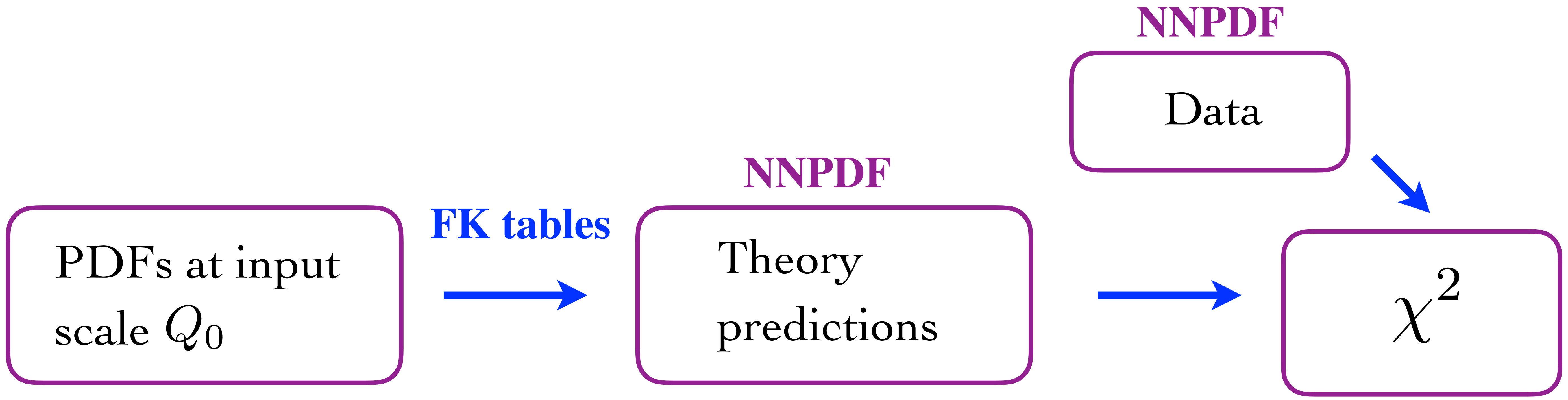
Direct Comparison

- Possible for first time using of publicly available NNPDF code - suitably adapted to allow fit to be performed with a (MSHT) fixed parameterisation, but to NNPDF data/theory - **only difference** is **input parameterisation.**

Data + theory fixed to NNPDF4.0

Neural net ⟶ **Fixed parameterisation**

- How does fit quality compare? How to PDFs and their uncertainties compare?



PDFs at input scale Q_0

FK tables

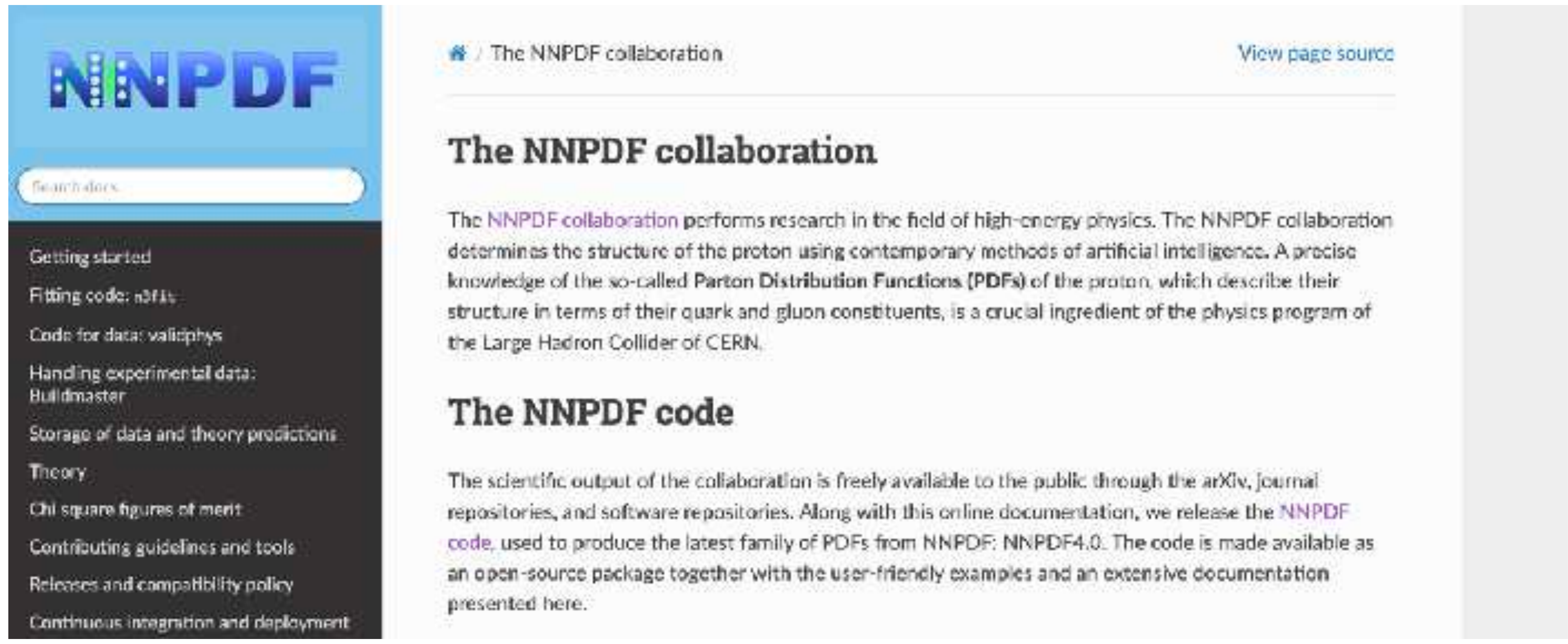
NNPDF

Theory predictions

NNPDF

Data

χ^2



The NNPDF collaboration

The NNPDF collaboration performs research in the field of high-energy physics. The NNPDF collaboration determines the structure of the proton using contemporary methods of artificial intelligence. A precise knowledge of the so-called **Parton Distribution Functions (PDFs)** of the proton, which describe their structure in terms of their quark and gluon constituents, is a crucial ingredient of the physics program of the Large Hadron Collider of CERN.

The NNPDF code

The scientific output of the collaboration is freely available to the public through the arXiv, journal repositories, and software repositories. Along with this online documentation, we release the NNPDF code, used to produce the latest family of PDFs from NNPDF: NNPDF4.0. The code is made available as an open-source package together with the user-friendly examples and an extensive documentation presented here.

https://docs.nnpdf.science/

Fit Quality

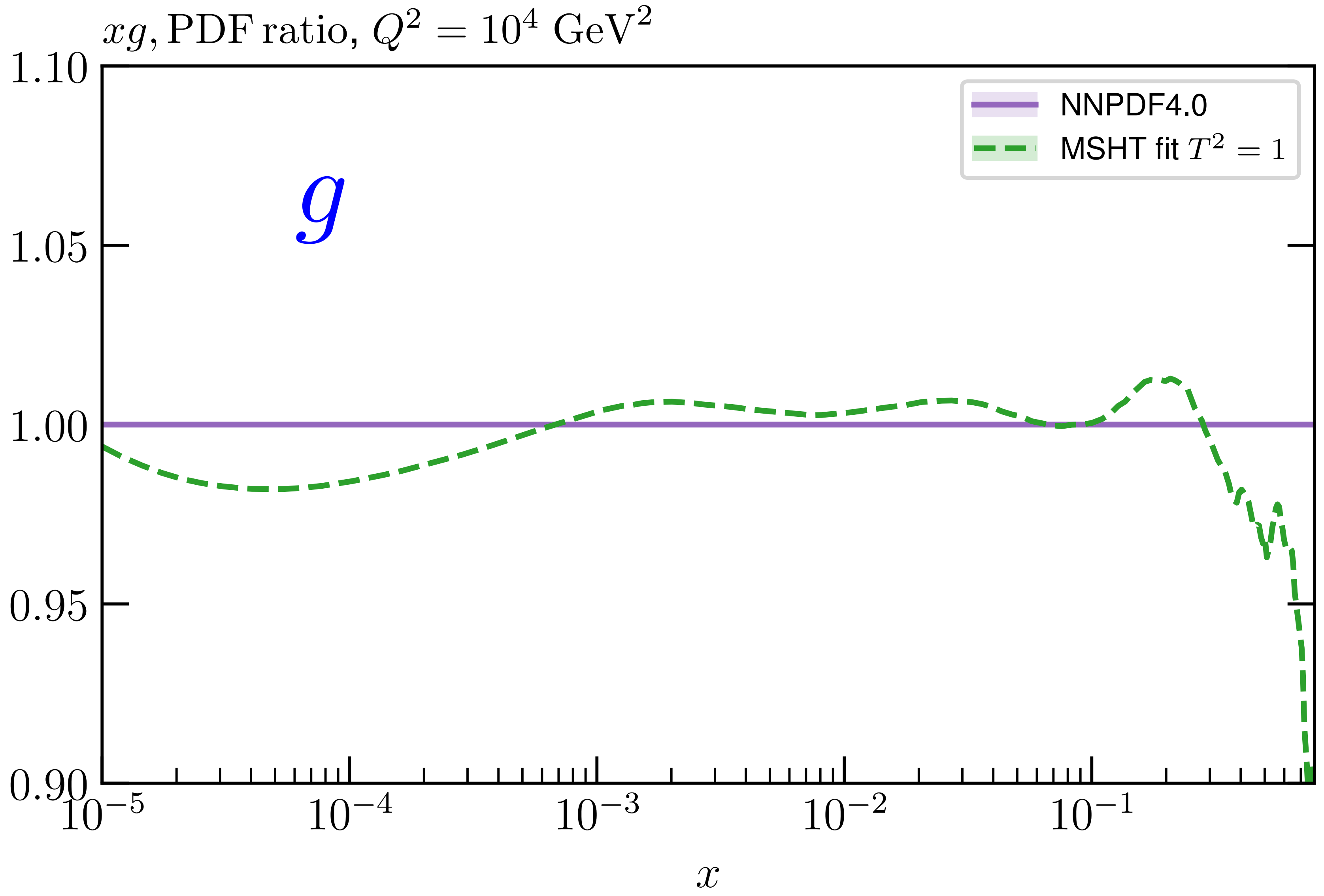
Fitted Charm case - similar results for p. charm

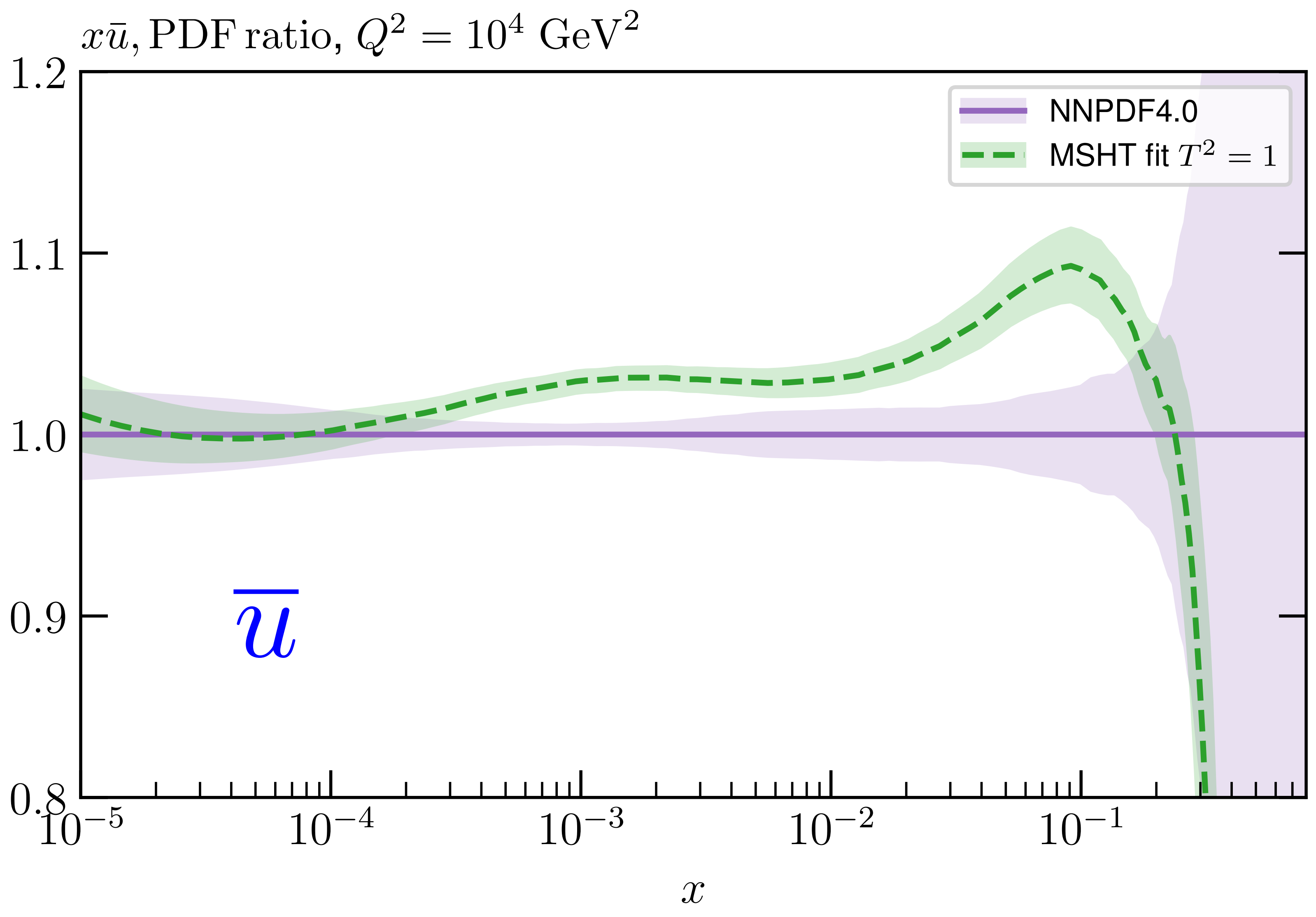
- Fit quality with **MSHT** parameterisation similar to (slightly better) than NNPDF central replica result.

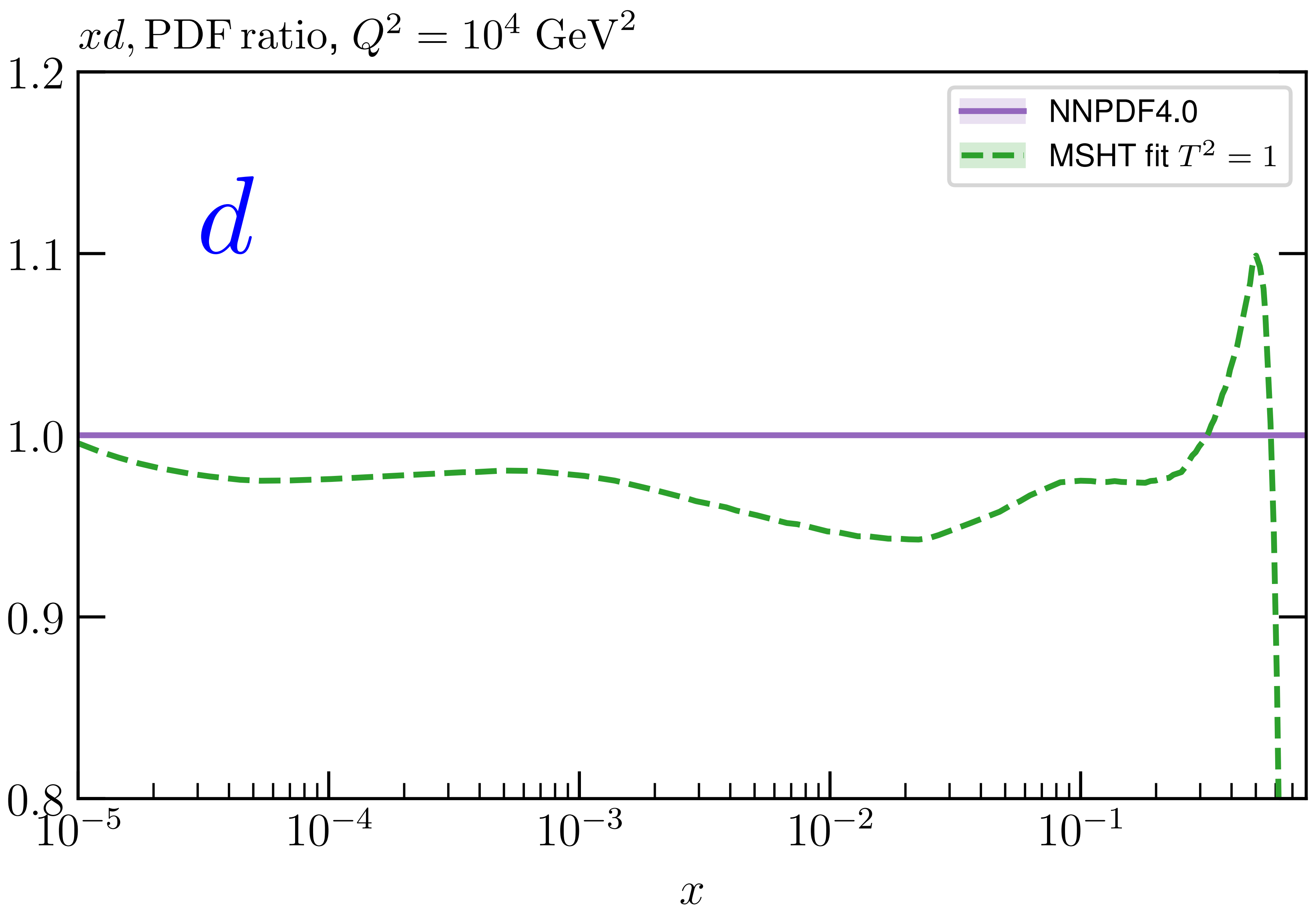
	NNPDF4.0	MSHT fit
χ^2	5692.1 (1.233)	5645.2 (1.222)

$\Delta\chi^2$: $-46.9\,(0.011)$

- PDF in some cases similar, but not always consistent.







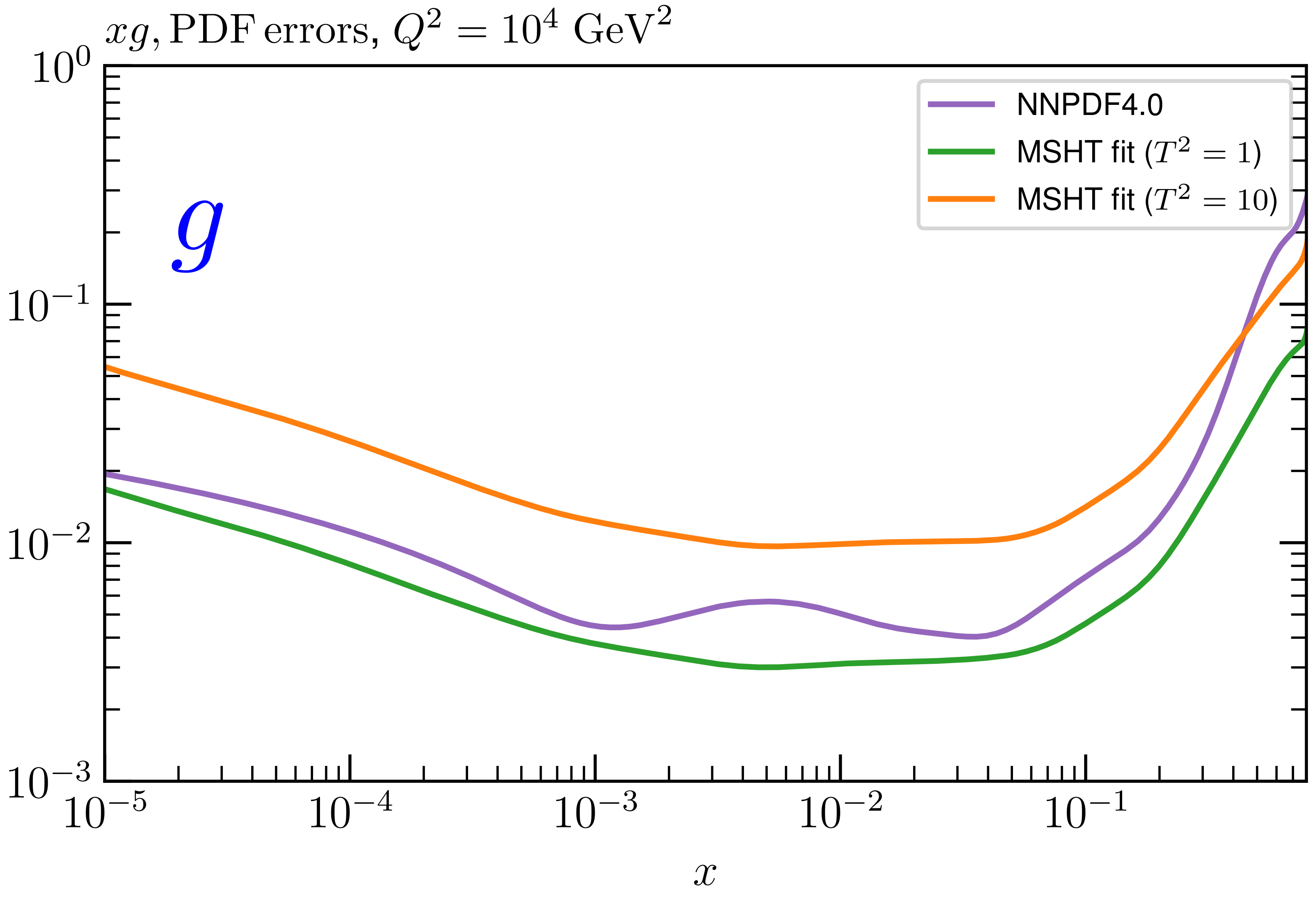
Comparison to NNPDF uncertainties

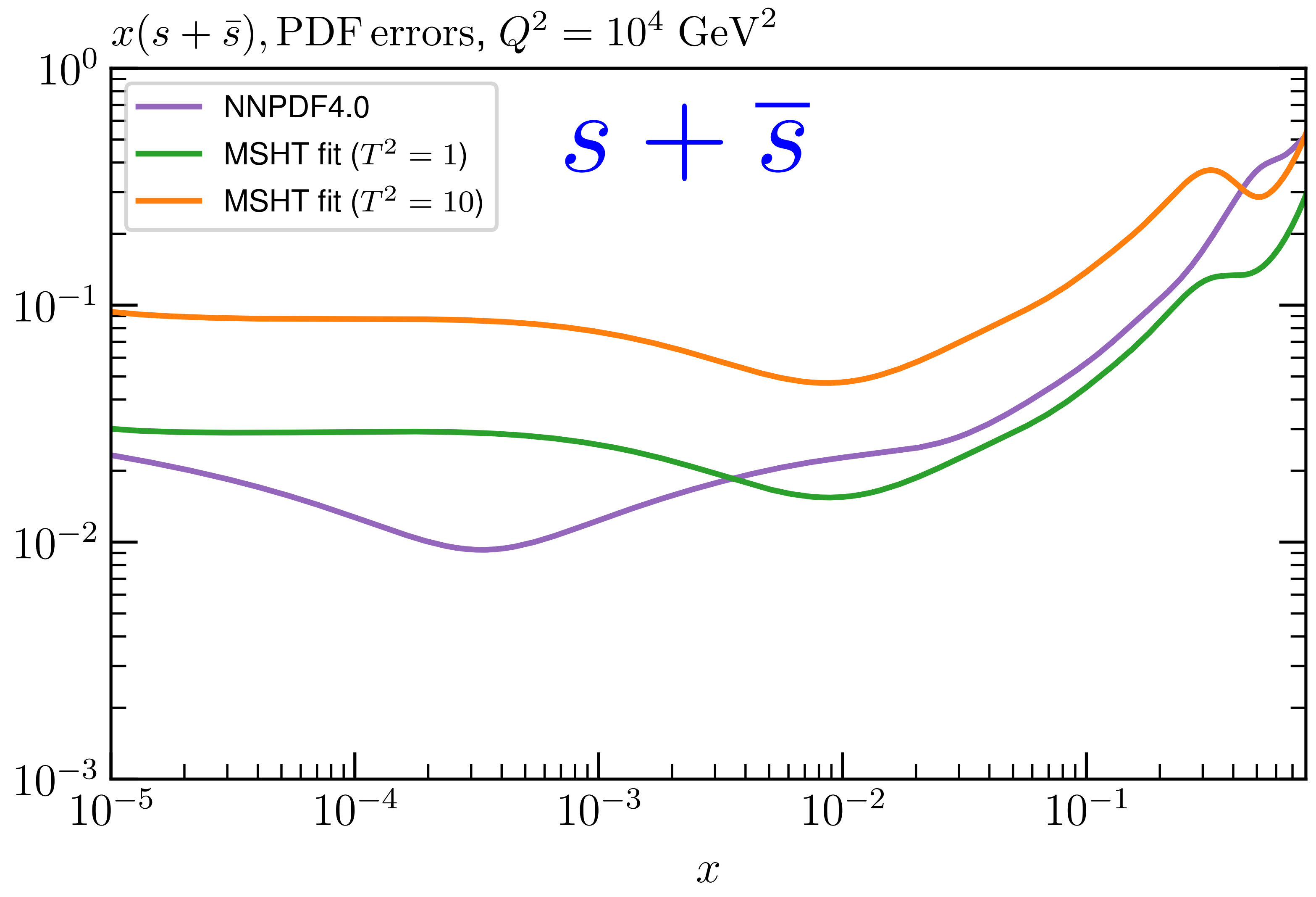
- Completely like-for-like comparision of NN and fixed parameterisation uncertainties. Find indeed that neural net approach ~ similar to fixed parameterisation but with $\Delta\chi^2 = 1$

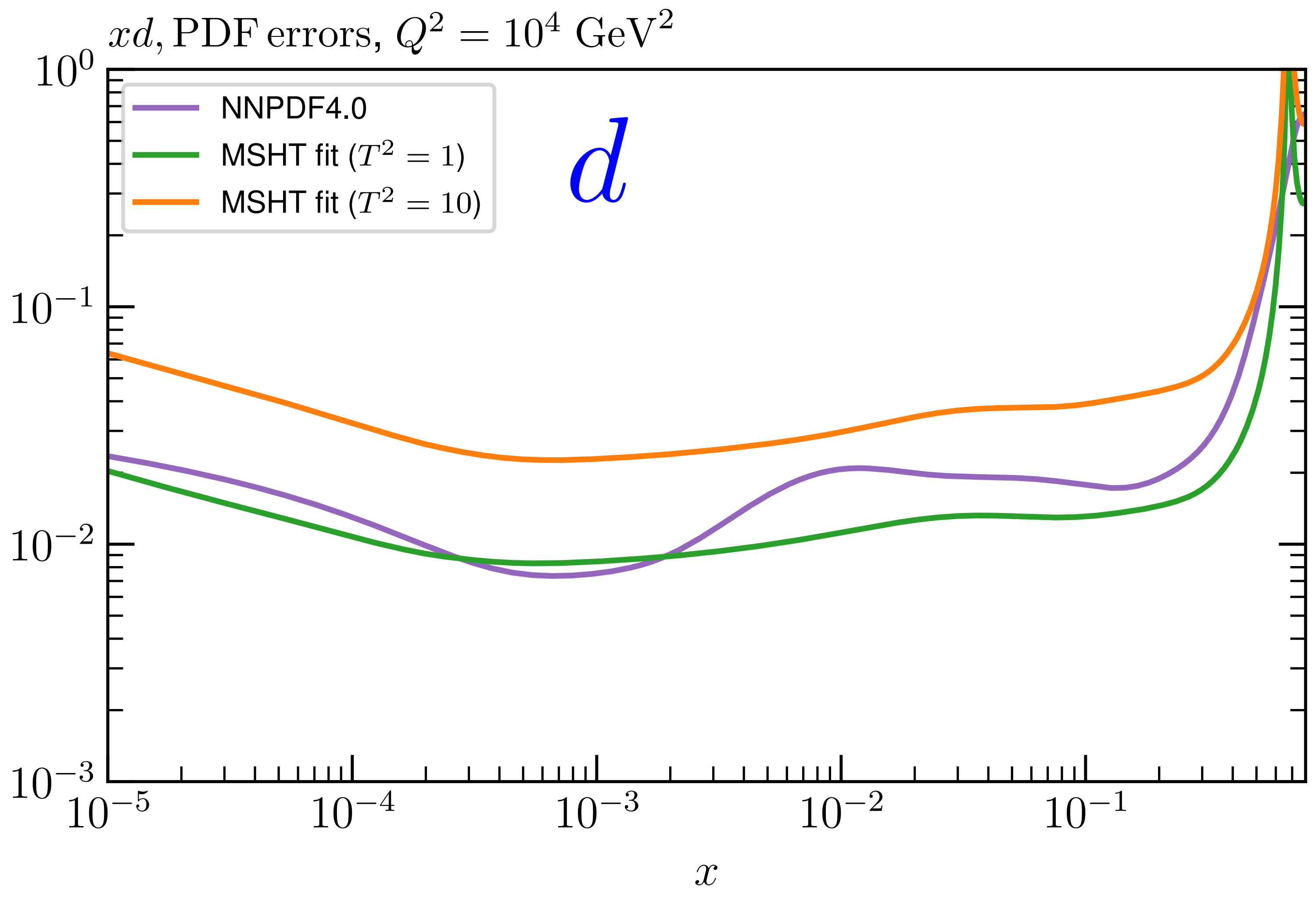
★ Quark flavour decomposition: $\sigma(\text{NNPDF}) \sim \sigma(\text{MSHT}, T^2 = 1)$

★ Gluon (singlet at intermediate x): $\sigma(\text{MSHT}, T^2 = 1) \lesssim \sigma(\text{NNPDF}) \lesssim \sigma(\text{MSHT}, T^2 = 10)$

NNPDF4.0 MSHT, $T^2 = 10$ MSHT, $T^2 = 1$







What to conclude?

- Returning to the original possibilities. Successful closure test appear to exclude 3rd option:

1. NNPDF4.0 uncertainty not conservative enough (too small).
2. MSHT uncertainty too conservative (too large).
3. ~~MSHT fit less accurate, due to parameterisation inflexibility, and hence enlarged errors needed (less precise).~~

- And so we are in situation where two approaches appear to work equally well but approach applied to defining uncertainties gives different answers.
- But ongoing work to clarify the picture further is certainly need. Priorities for future…?

Future Goals/Challenges for UQ

- Some questions to address/being addressed:

★ Why do we need a tolerance?

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

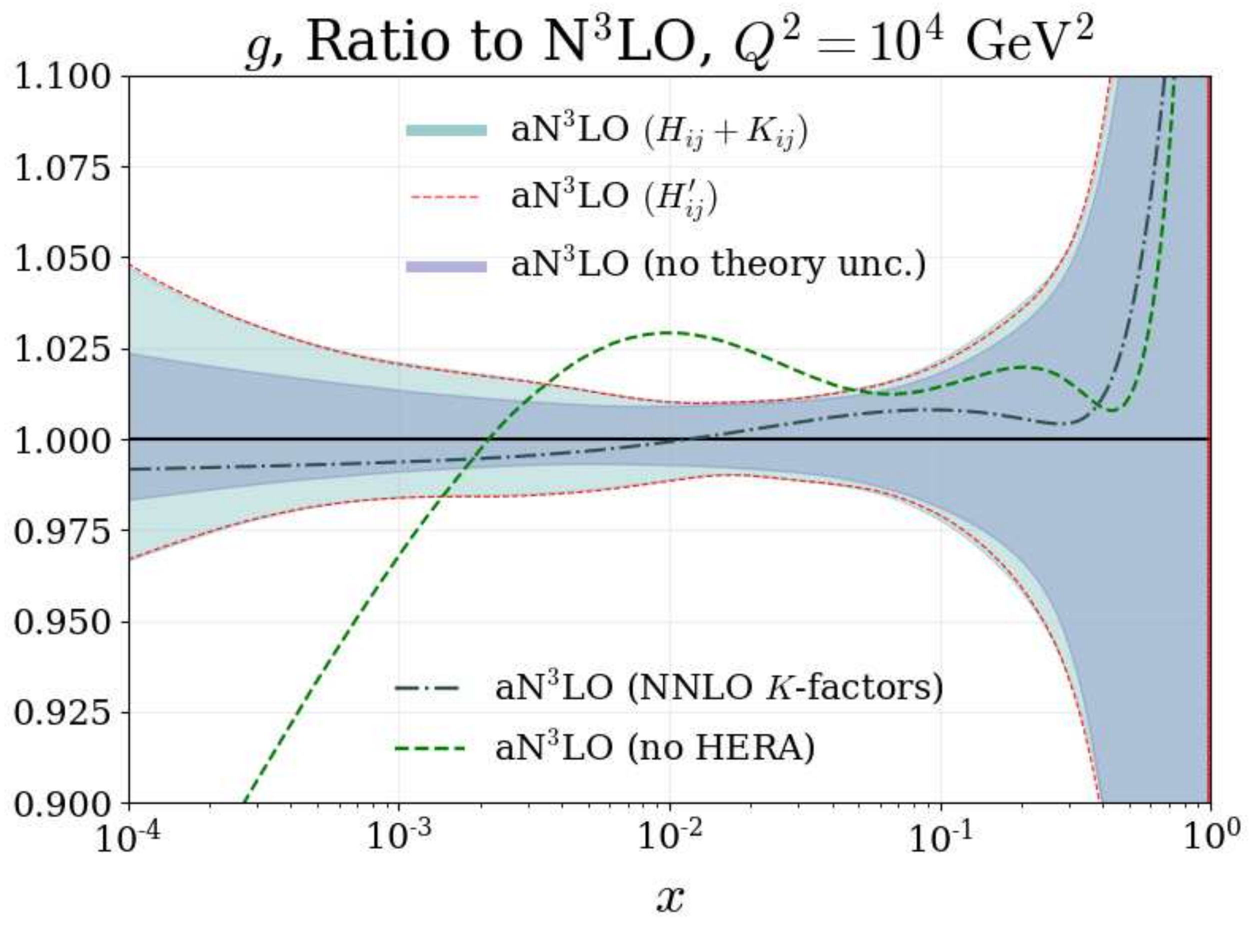
E. Canonero & G. Cowan, arXiv:2407.05322…

✦ Ongoing studies on 'error on error' for data.

✦ Much progress towards including theoretical uncertainties.

✦ Can we identify those datasets that are most in tension?

✦ Further input from closure tests?



MSHT20aN3LO, arXiv:2207.04739

NNPDF4.0aN3LO, arXiv:2402.18635

NNPDF, arXiv:2109.02653

by addressing these questions may hope to reduce need for it and certainly refine treatment.

- Use new tools to continue with further like-for-like comparisons between NN and fixed parameterisation approaches:
    - Expand **closure tests** to include more **statistical estimators** (bias to variance ratio etc) to give more fine grained picture within both approaches.

$$\xi_{n\sigma}^{(\mathrm{pdf})} = \frac{1}{n_{\mathrm{flav}}} \frac{1}{n_x} \frac{1}{n_{\mathrm{fit}}} \sum_{i=1}^{n_{\mathrm{flav}}} \sum_{j=1}^{n_x} \sum_{l=1}^{n_{\mathrm{fit}}} I_{[-n\sigma^{i(l)}(x_j), n\sigma^{i(l)}(x_j)]} \left(q^{i(l)}(x_j) - q_{\mathrm{input}}^{i(l)}(x_j) \right) .$$

	Total
$\xi_{1\sigma}^{\mathrm{PDF}}$	$\mathbf{0.65 \pm 0.01}$

    - Expand **closure tests** to include other fixed parameterisation approaches (CT…).
    - How might we account for a **tolerance** (enlarged error definition) in NNPDF - **MC replica** - approach?
    - How can we improve fixed parameterisation fits in **extrapolation** regions, where they work less well (e.g. exploring MC replica generation…)? Input from lattice here?
    - Questions of **generalisability** etc - how well do PDFs work when tested against new data or as part of future tests. Again role of lattice inputs here?
- Some of these questions clearly somewhat open-ended/longer term. Would suggest key is in direct comparisons and benchmarking to help understand the situation more clearly.

Conclusions

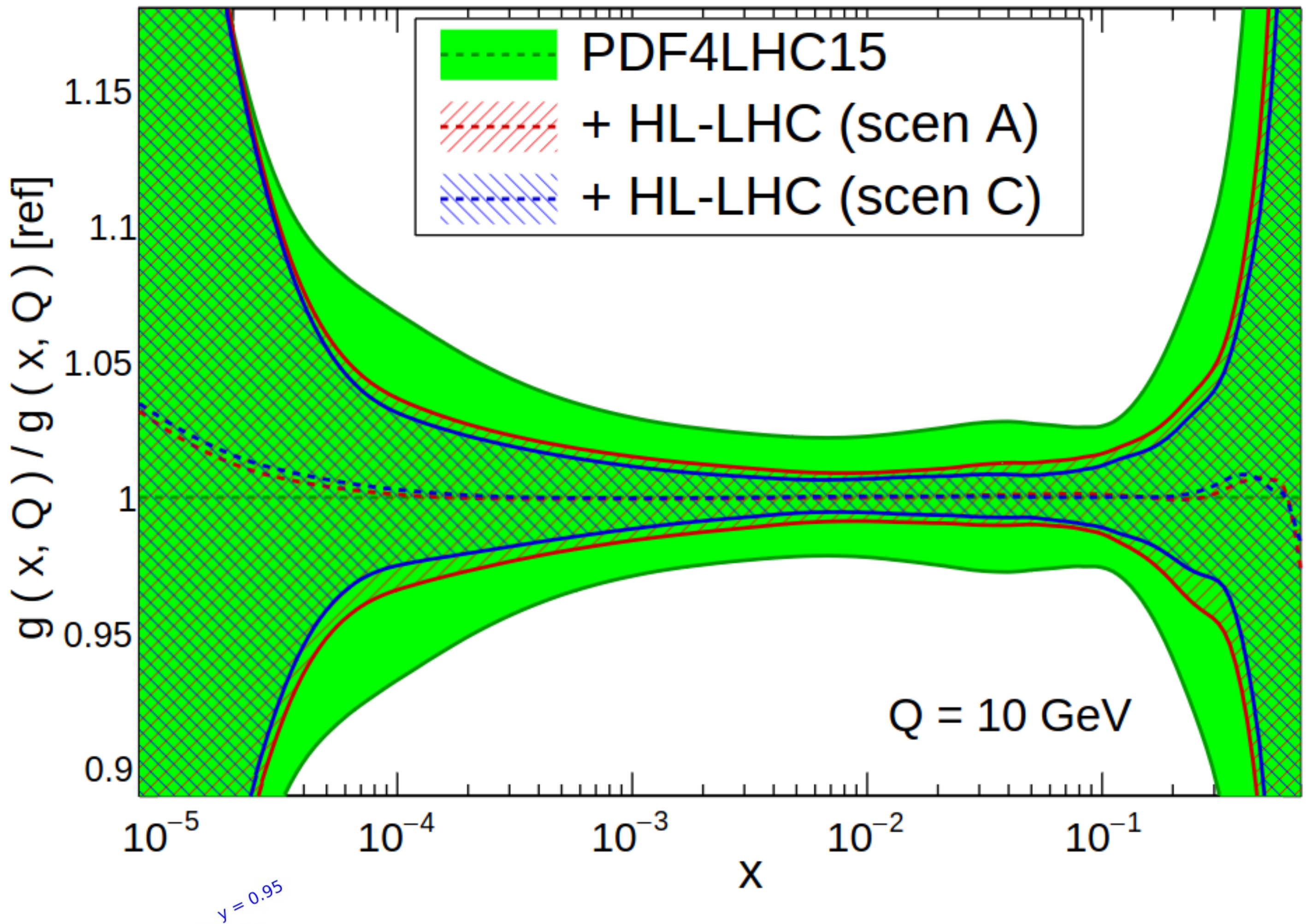
- ★ Parton Distribution Functions a key input in the precision physics programme of LHC and beyond.
- ★ Precise and accurate PDF determination crucial. Global PDF fits currently the best way to achieve this.
- ★ A significant deal of experimental and theoretical progress: high precision LHC data driving 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!

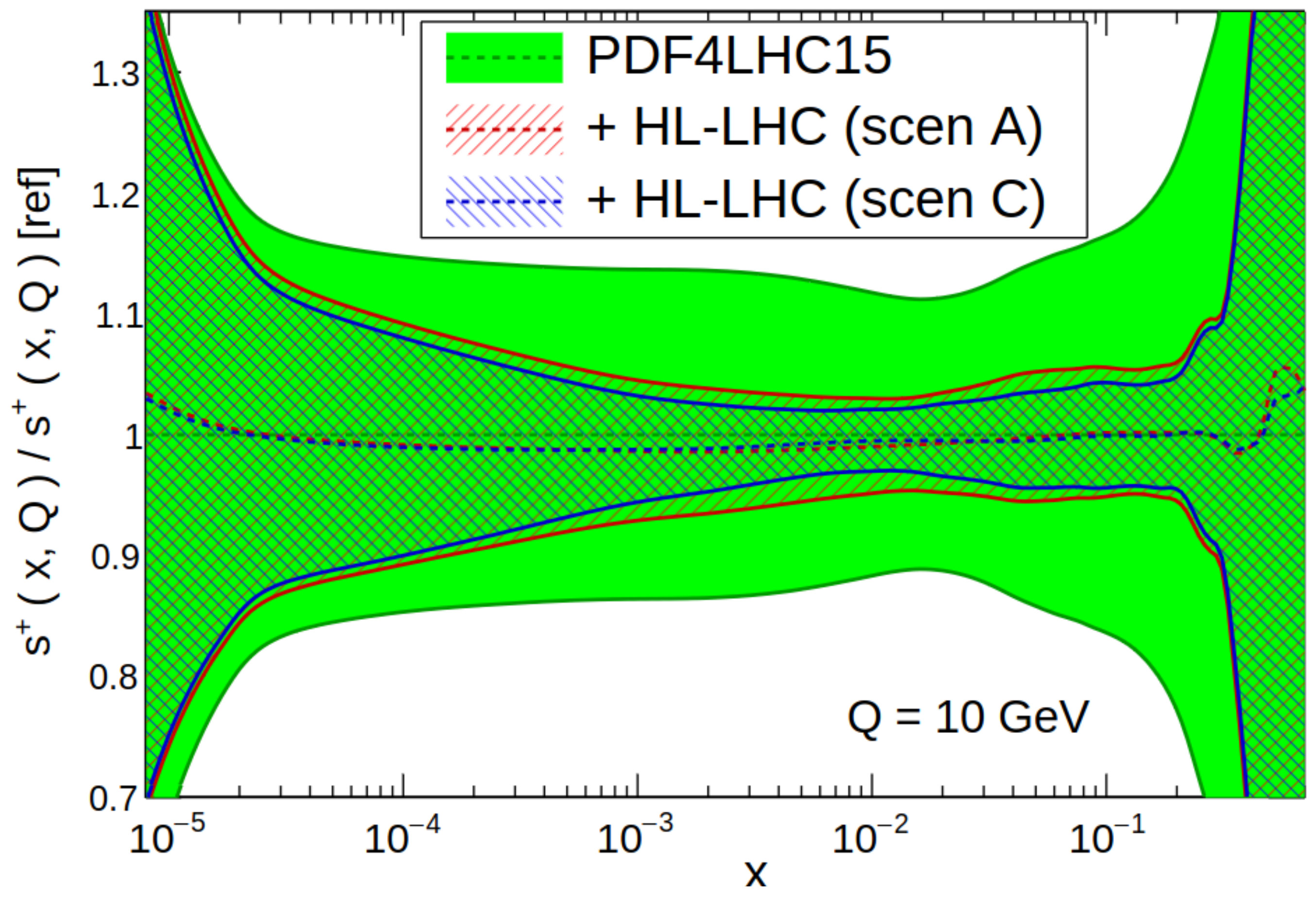
Thank you for listening!

Backup

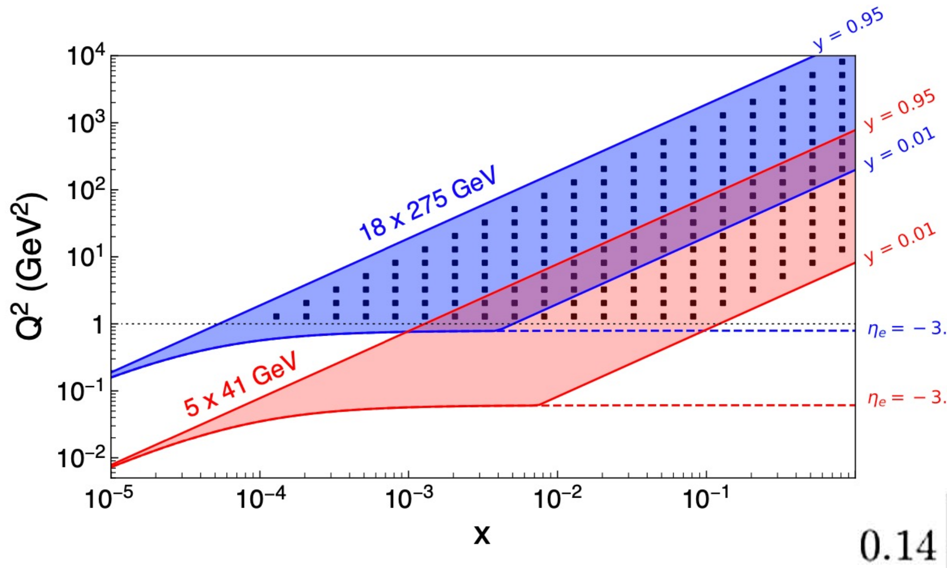
Where might we go? Future Data

- 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.



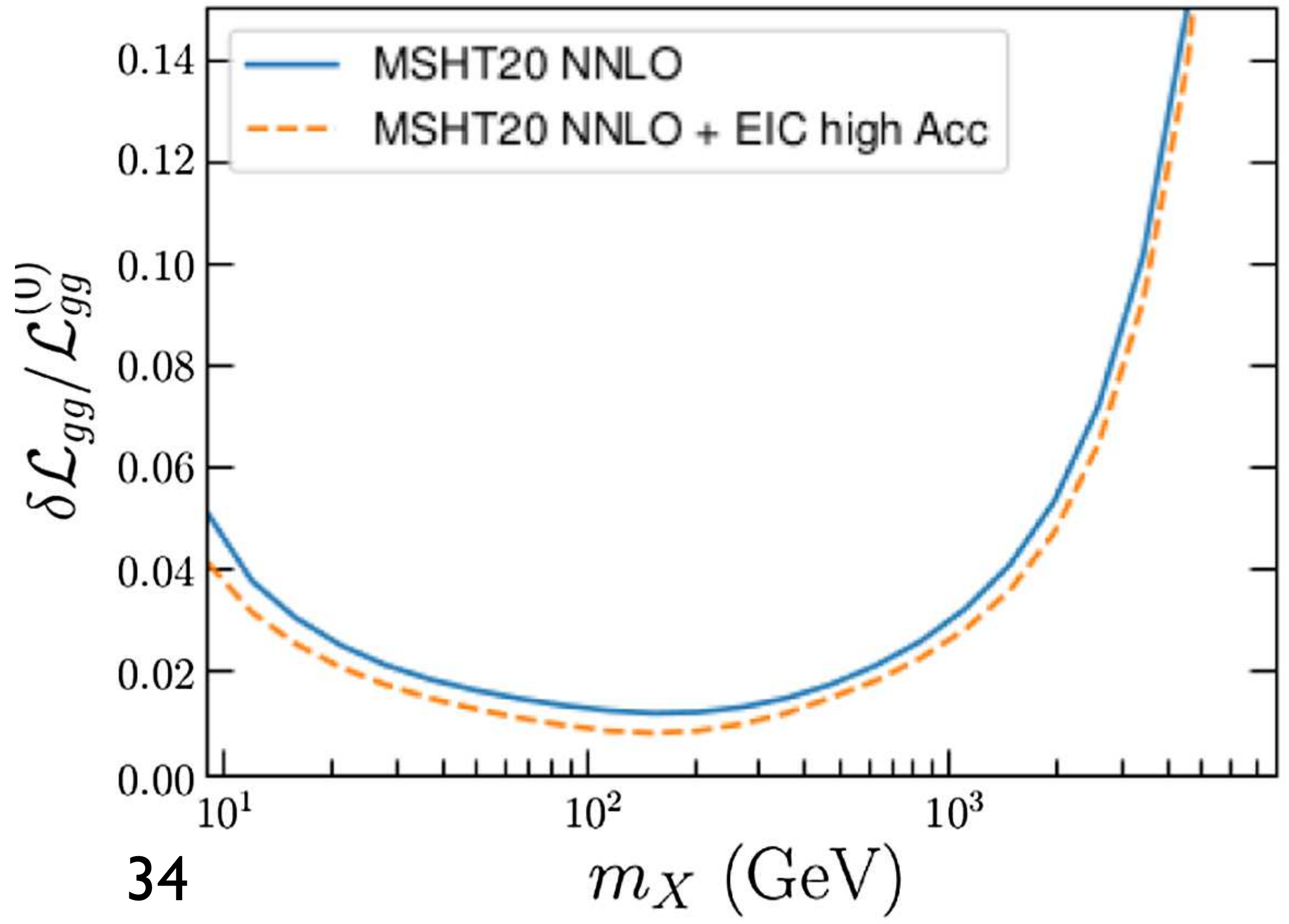


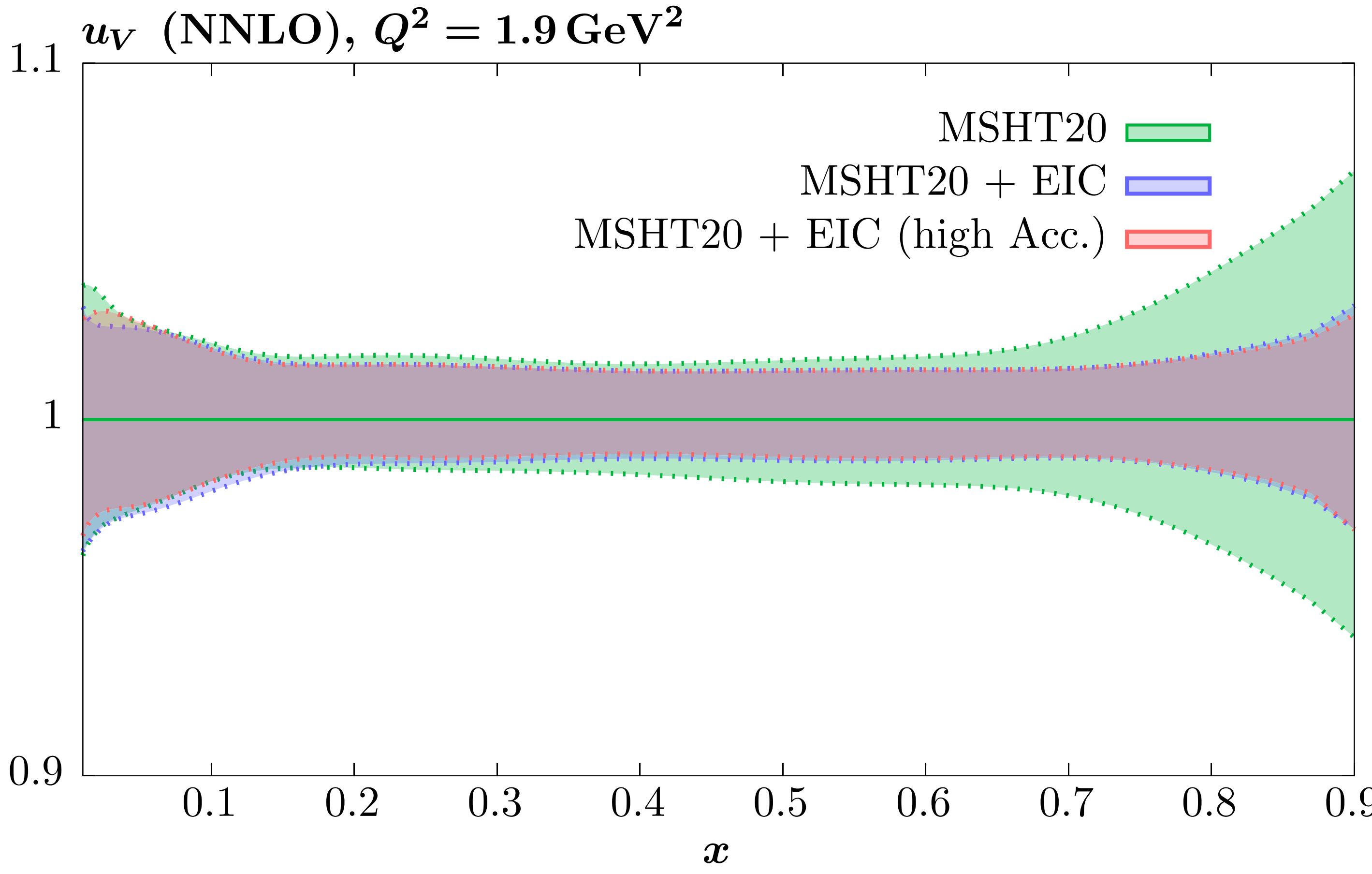
- Amongst many things the **EIC** will give us are better constraints on high x PDFs.



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

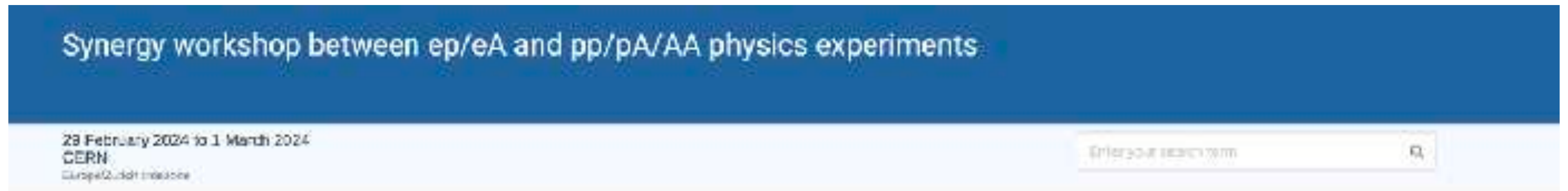
N. Armesto et al., arXiv:2309.11269





Future Data?

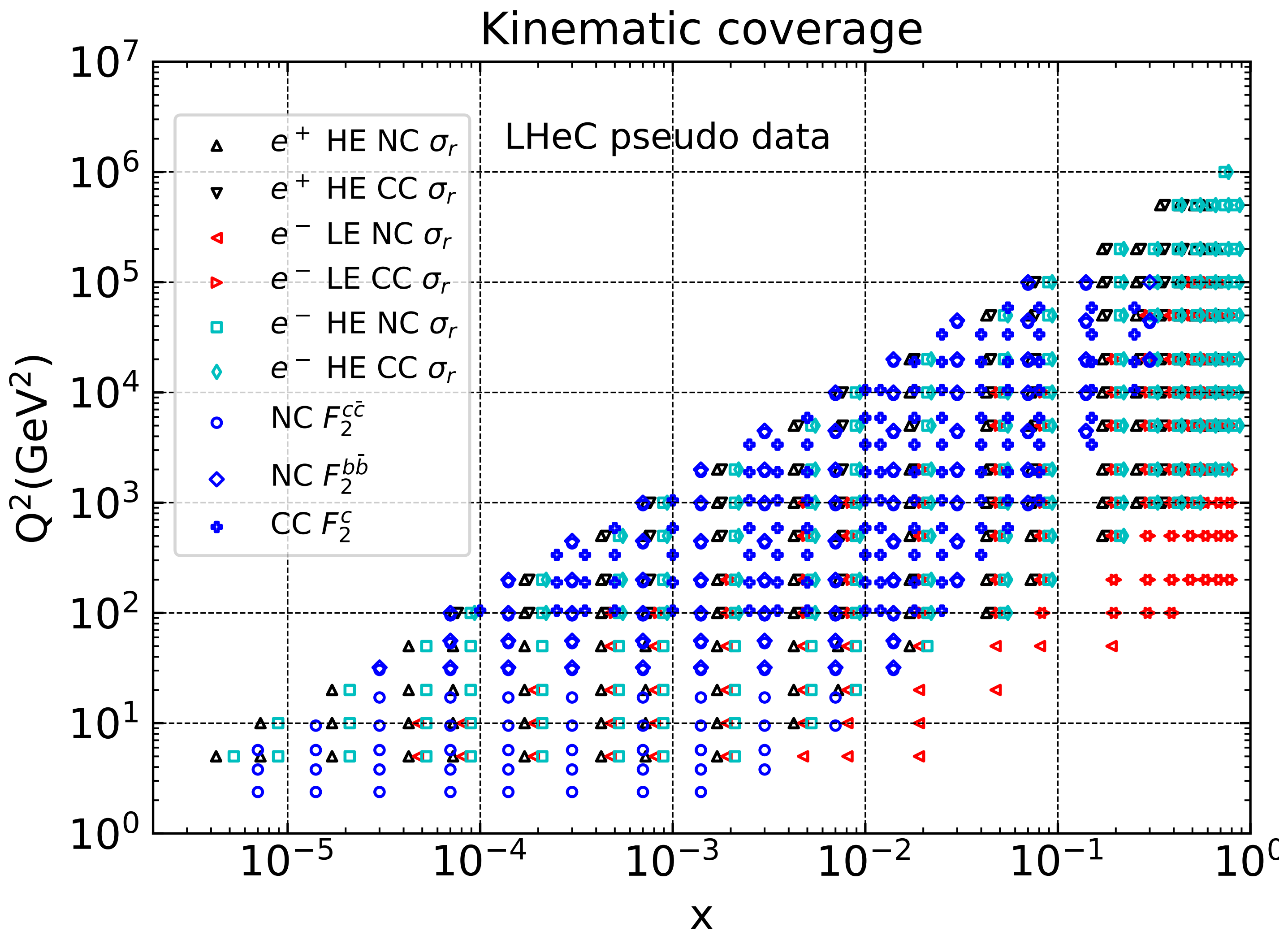
https://indico.cern.ch/event/1367865/overview

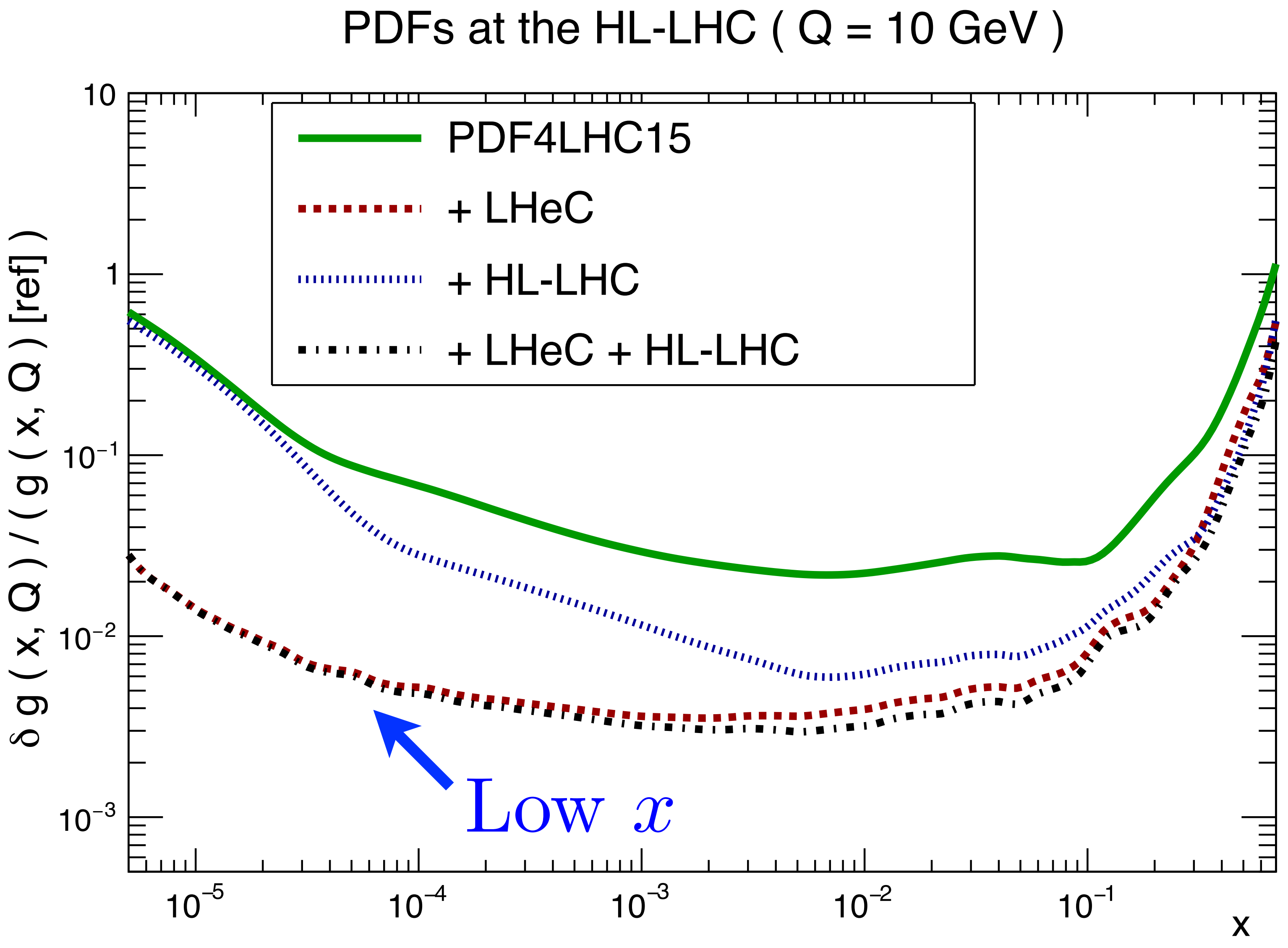


Synergy workshop between ep/eA and pp/pA/AA physics experiments

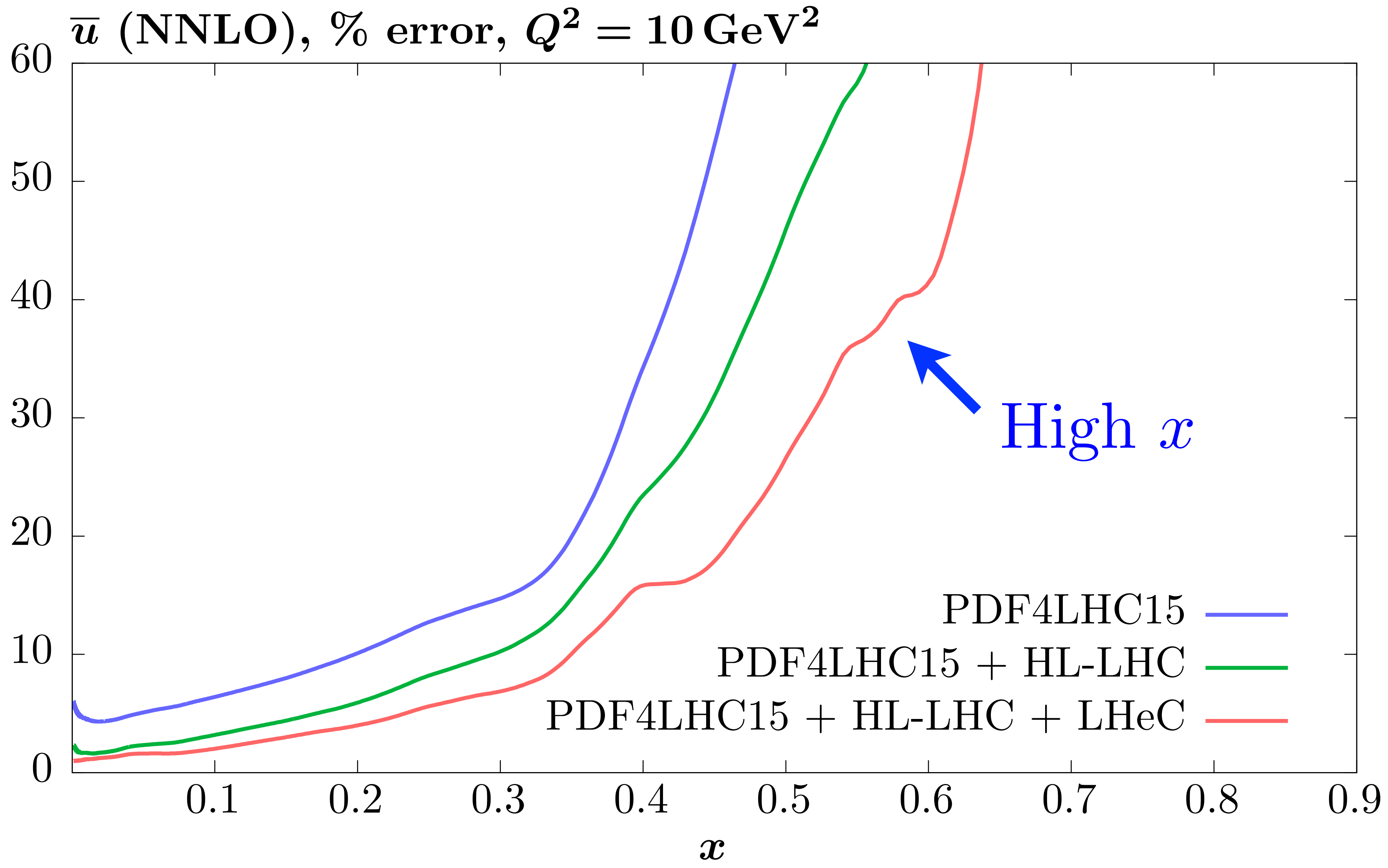
29 February 2024 to 1 March 2024
CERN

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



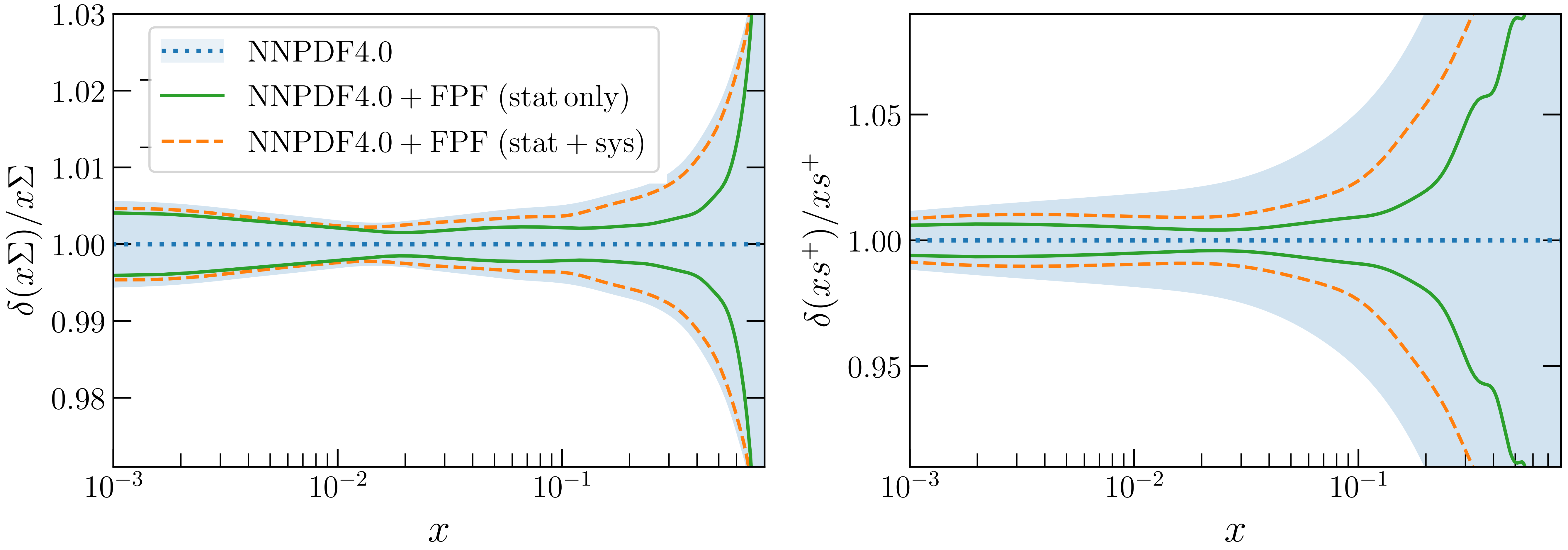


R. Abdul Khalek et al., arXiv:1906.10127

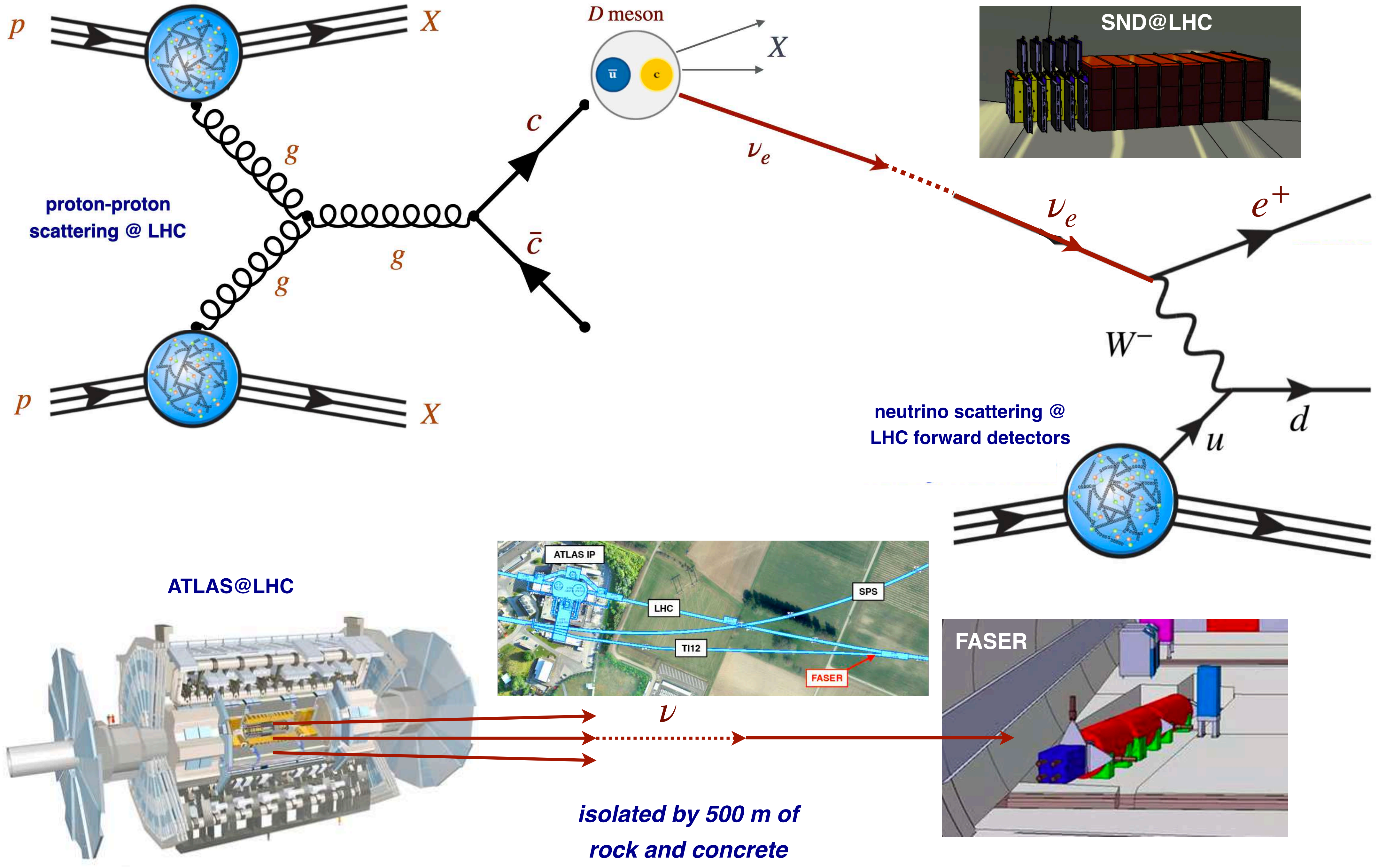


Future Data?

- **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

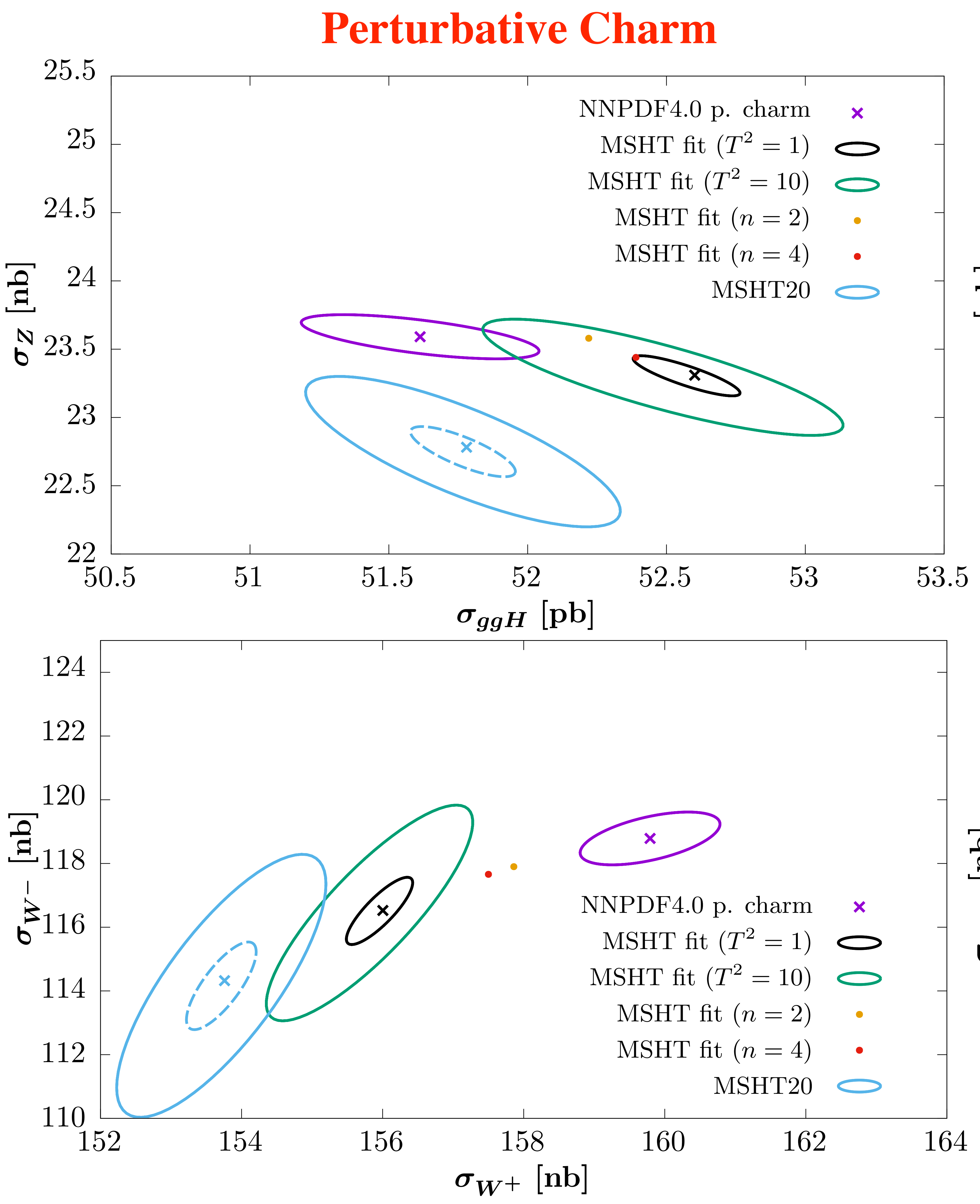
Cross Sections

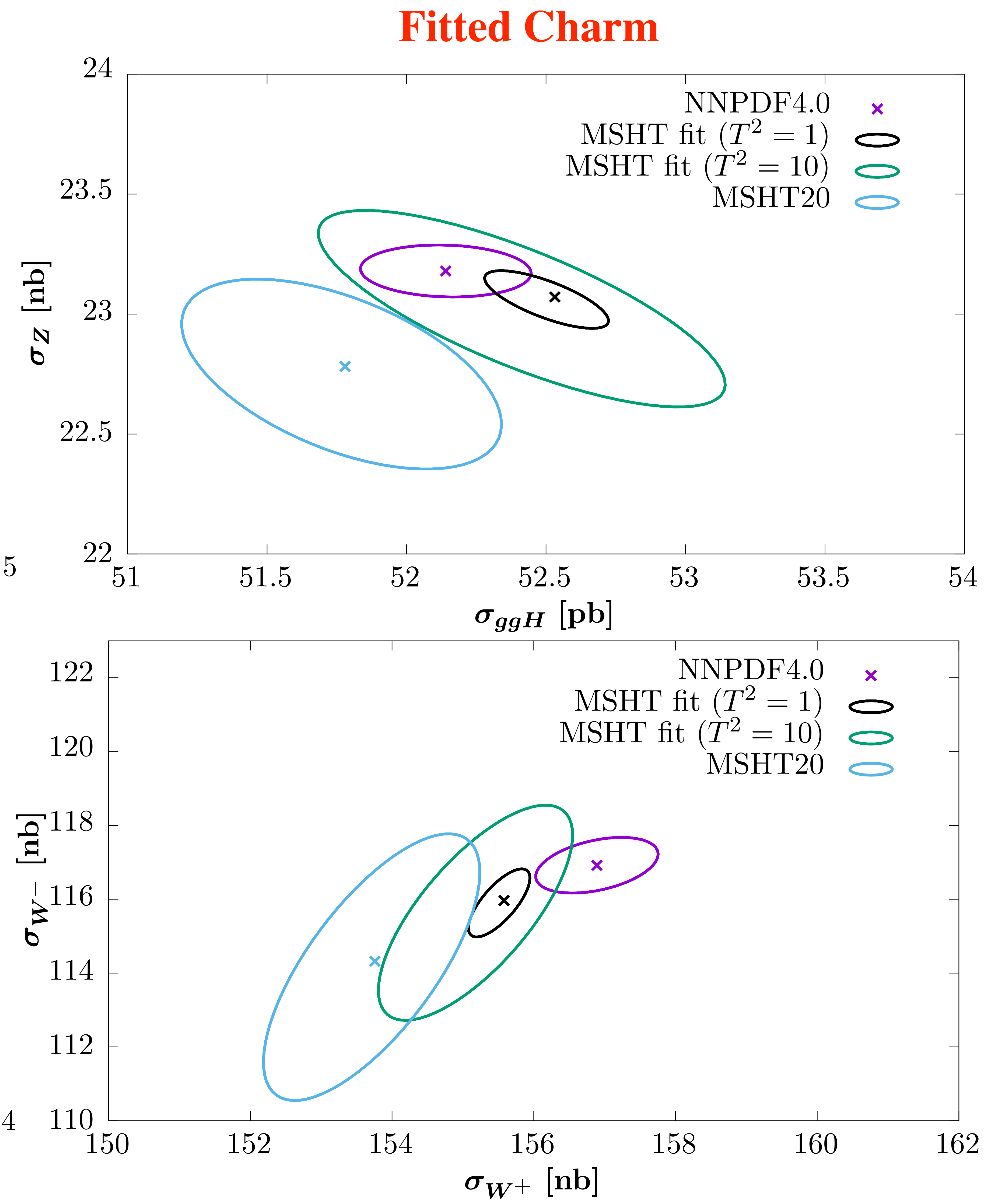
`n3loxs`

- Consider ggH, and W, Z cross sections (14 TeV):

MSHT, $T^2 = 10$ MSHT ($T^2 = 1$) NNPDF4.0

- ★ NNPDF uncertainties ~ MSHT ($T^2 = 1$) but **significantly smaller** than $T^2 = 10$
- ★ NNPDF and MSHT fit basically consistent within $T^2 = 1$ uncertainties but not relevant factor given fit qualities.





Cross Sections

`n3loxs`

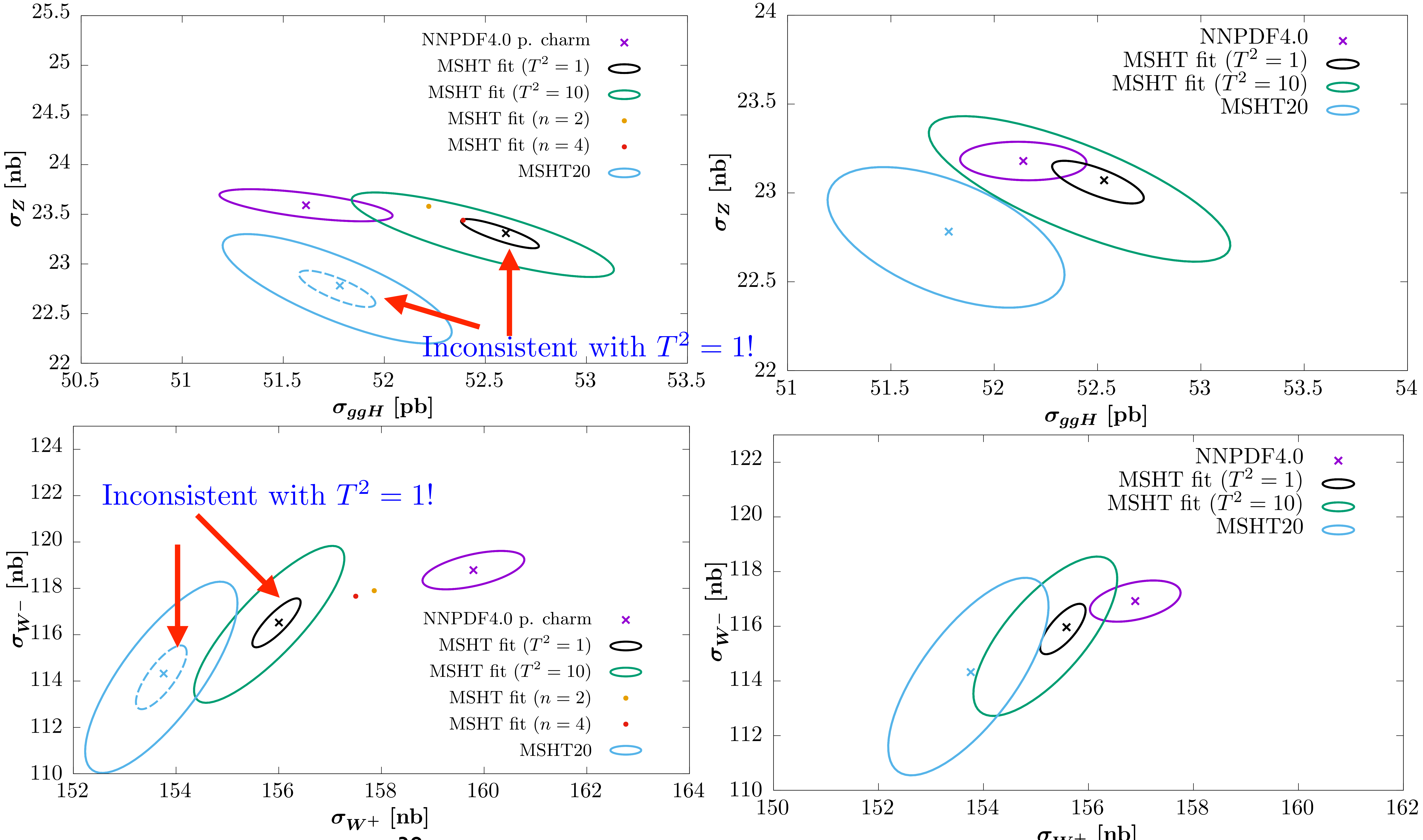
- Consider ggH, and W, Z cross sections (14 TeV):

MSHT, $T^2 = 10$ MSHT $(T^2 = 1)$ NNPDF4.0

- ★ NNPDF uncertainties ~ MSHT ($T^2 = 1$) but **significantly smaller** than $T^2 = 10$
- ★ NNPDF and MSHT fit basically consistent within $T^2 = 1$ uncertainties but not relevant factor given fit qualities.
- ★ Also shown is MSHT20 - impact of changing data/ theory alone. Need for **tolerance** clear!

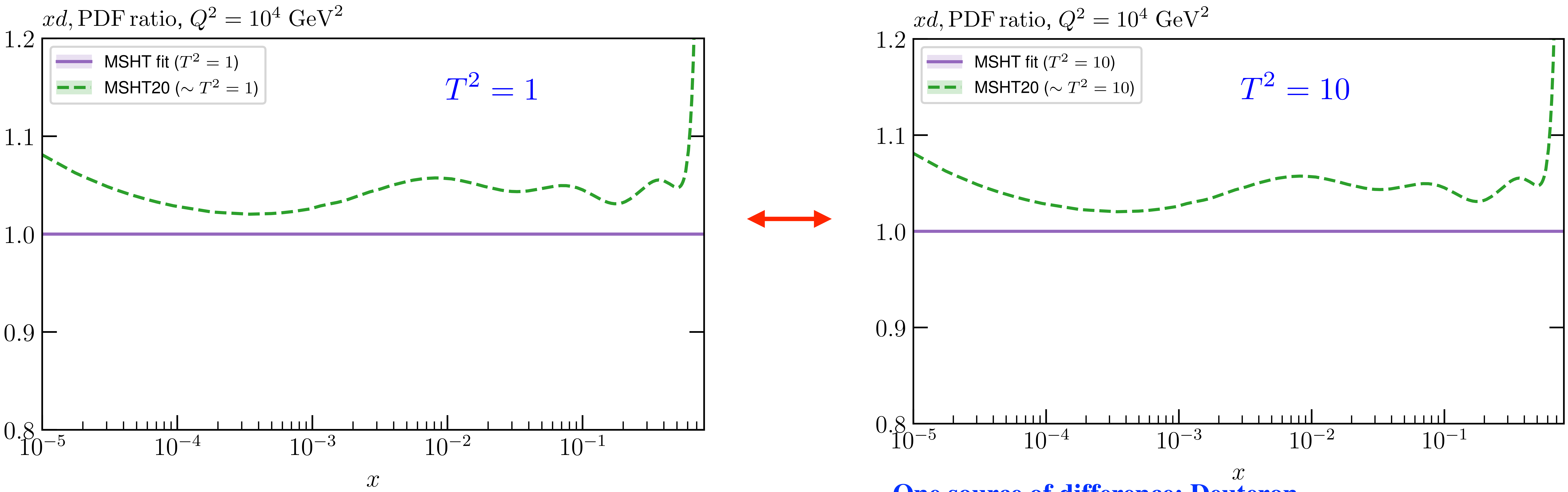
Perturbative Charm

Fitted Charm



Tolerance (and Again)

- Compare to our MSHT fit to NNPDF4.0 data + theory to public MSHT20 fit: only difference due to differing data + theory.

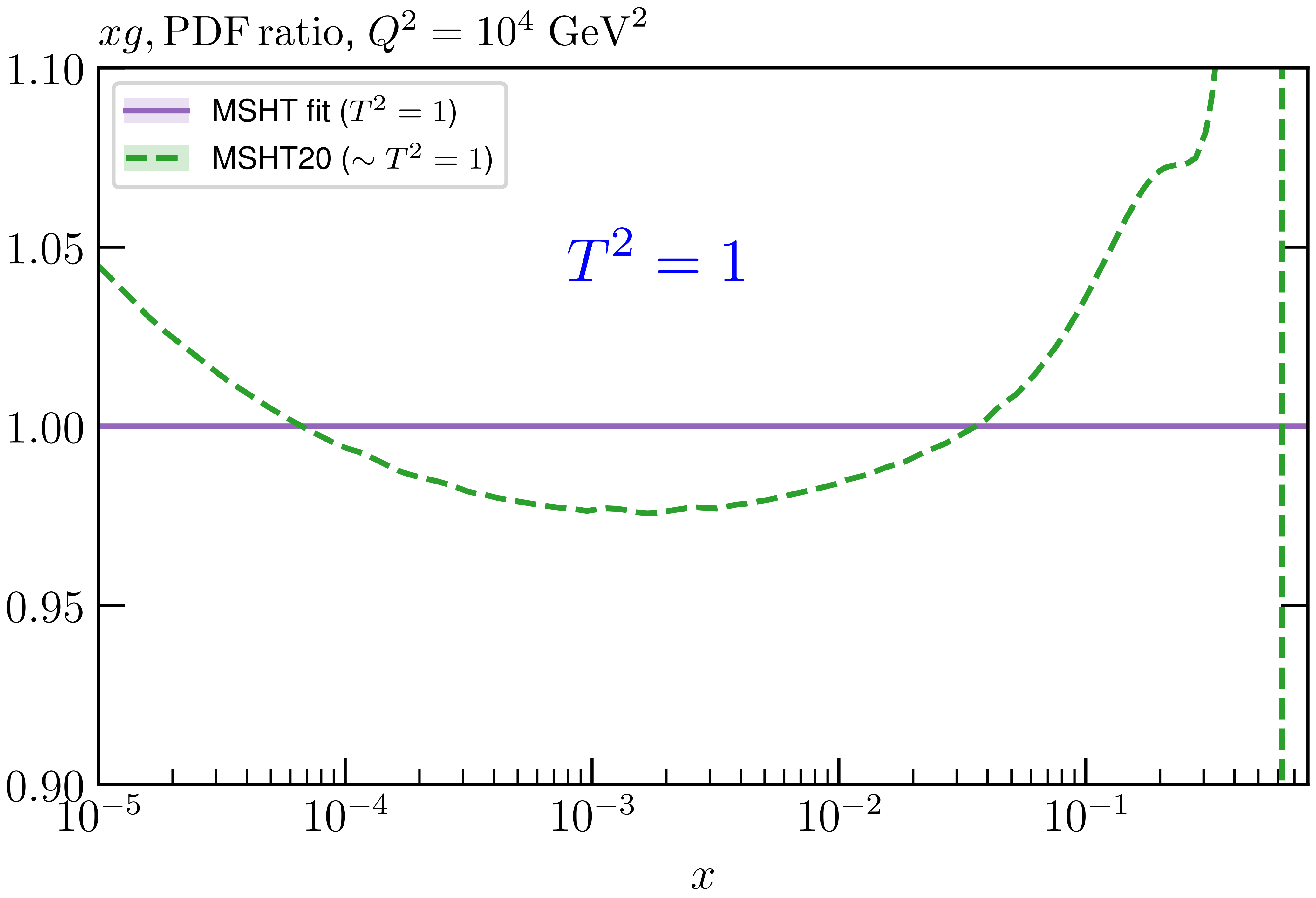


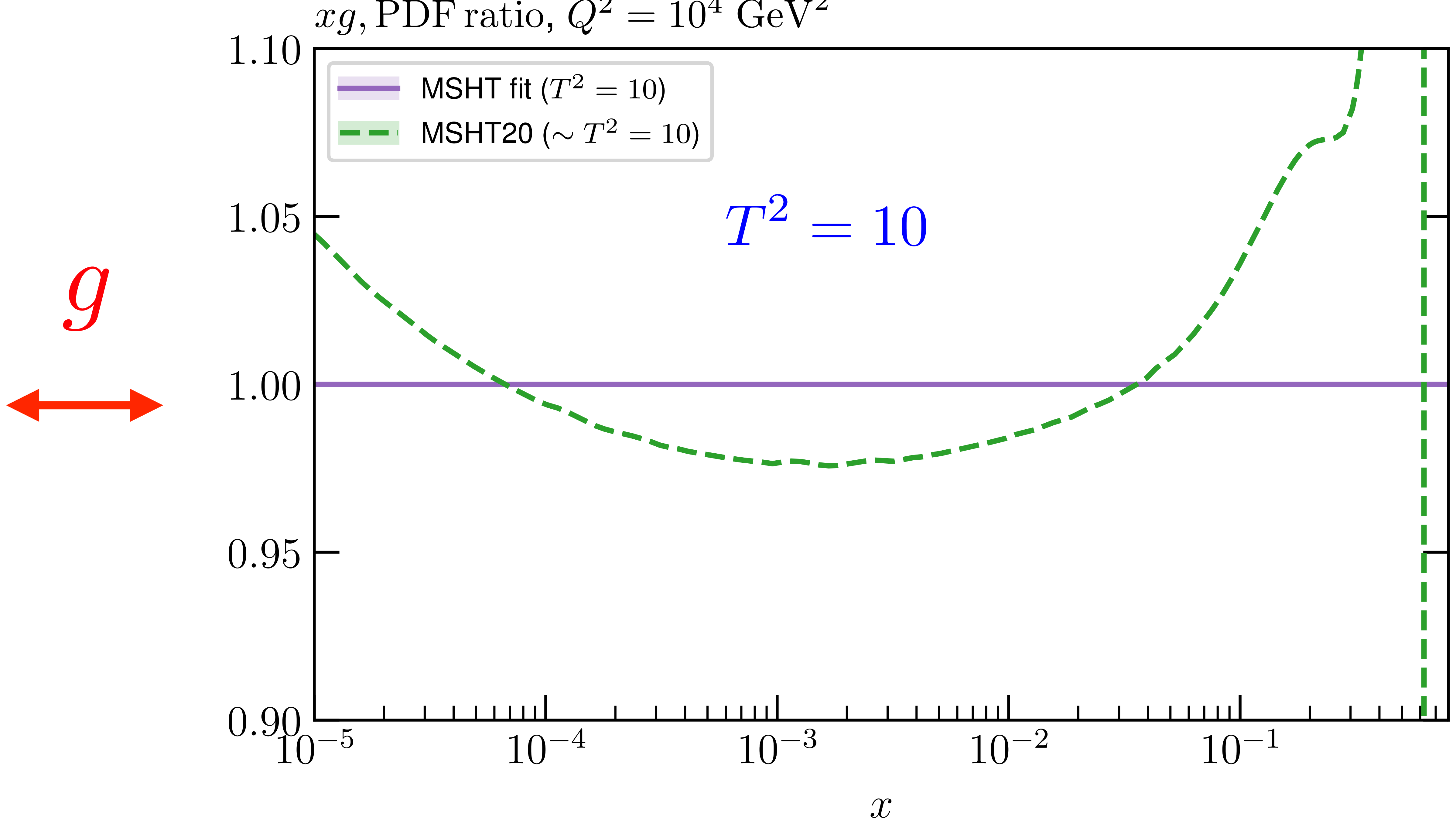
One source of difference: Deuteron Corrections

Tolerance (and Again)

- Compare to our MSHT fit to NNPDF4.0 data + theory to public MSHT20 fit: only difference due to differing data + theory.

One source of difference: LHC data sensitive to high x .





- Results arguably speak for themselves!

Restricted Parameterisations

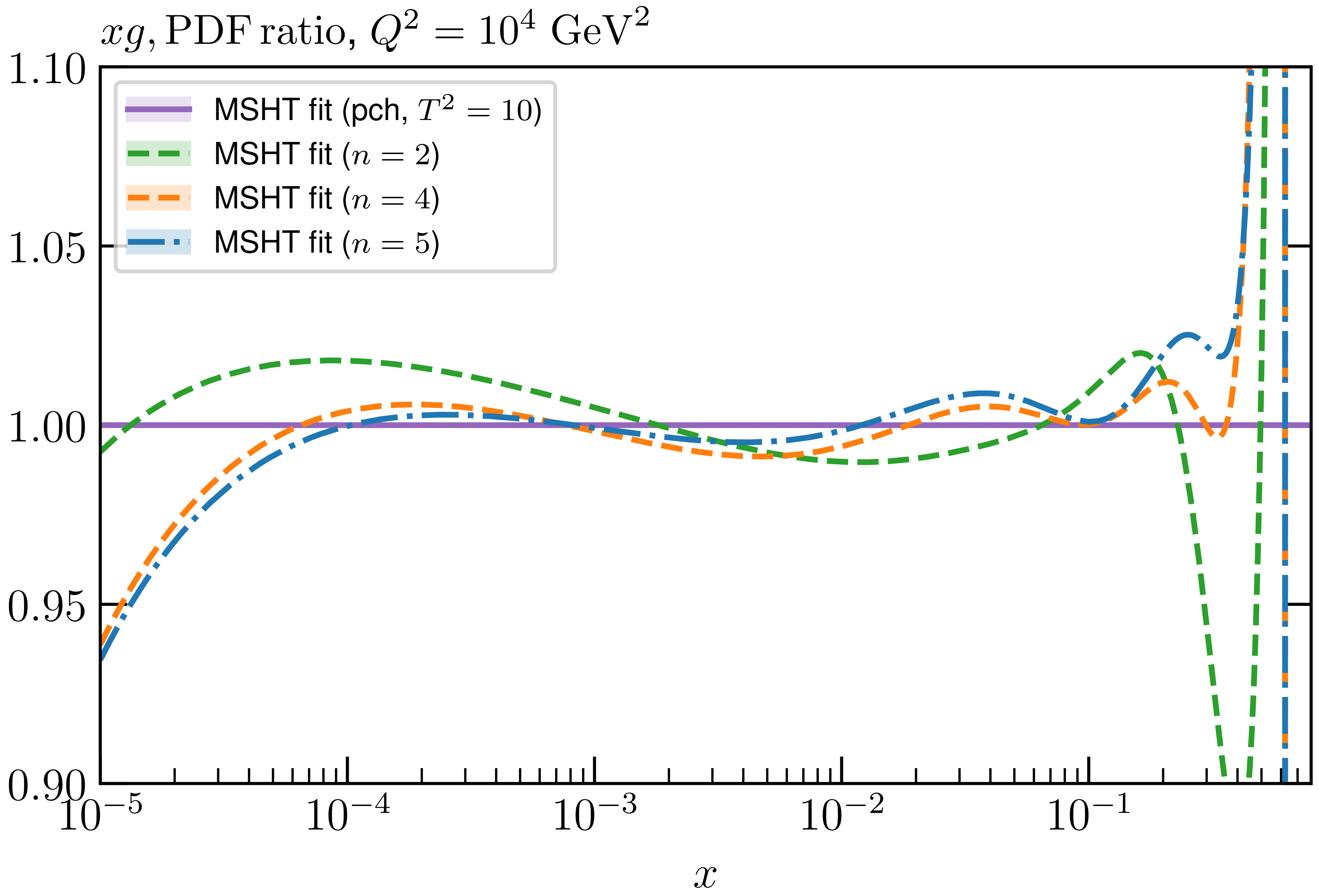
- Returning to fits - with apparently more restrictive parameterisation we are getting a better fit quality (and somewhat different PDFs) that NNPDF. What is going on? Try restricting parameterisation (as in closure test) - with p. charm.

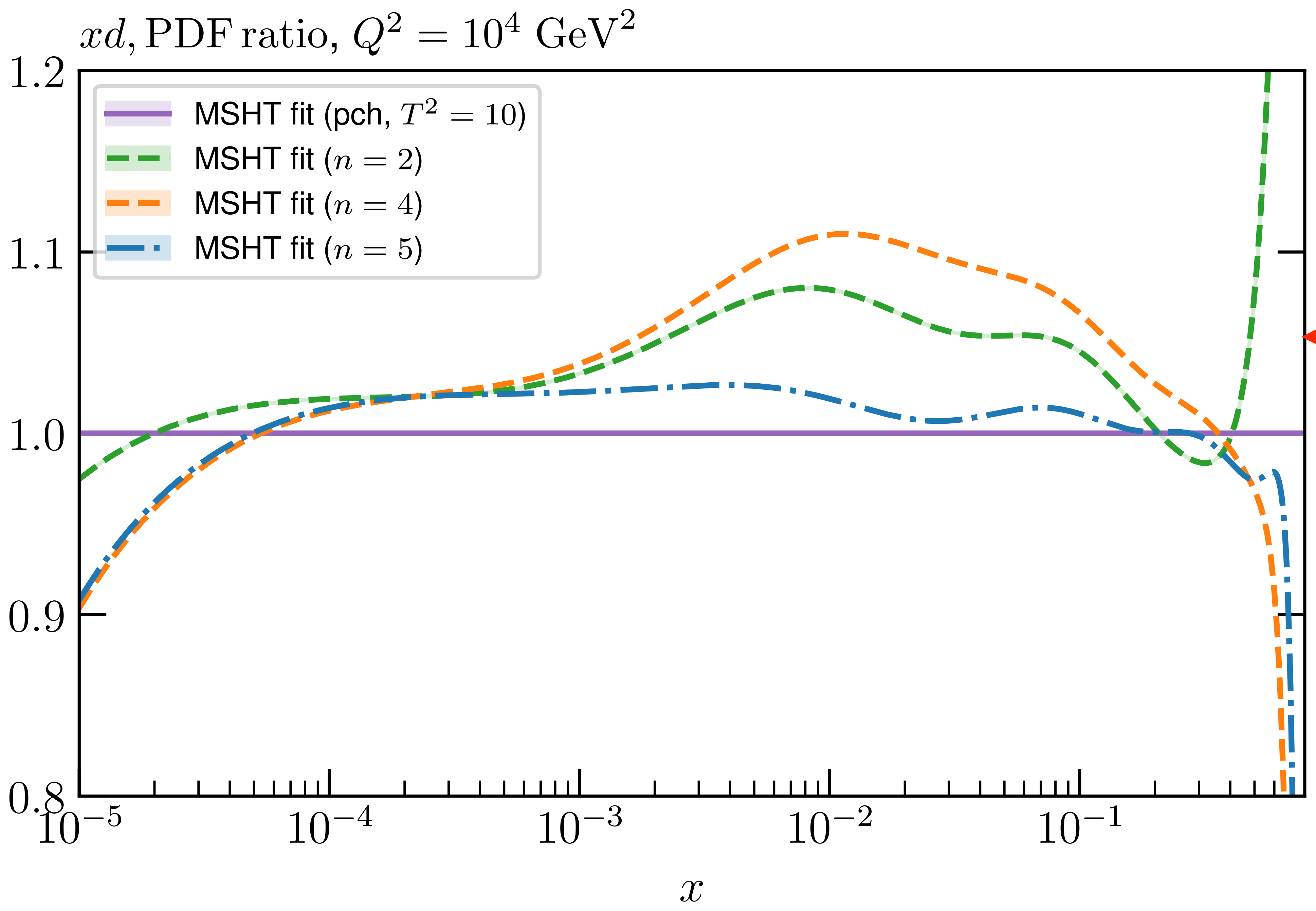
- Impact on fit quality slightly larger, e.g. ~ 120 point deterioration for $n = 4$ (~ 70 in closure)

$$xf(x,Q_0) = Ax^{\delta}(1-x)^{\eta}\left(1+\sum_{i=1}^{\boxed{n}} a_i T_i(y(x))\right) ,$$

$n = 2, 4 \quad \Rightarrow \quad n_{\mathrm{par}} = 28, 40$

$n = 6 \quad \Rightarrow \quad n_{\mathrm{par}} = 52$



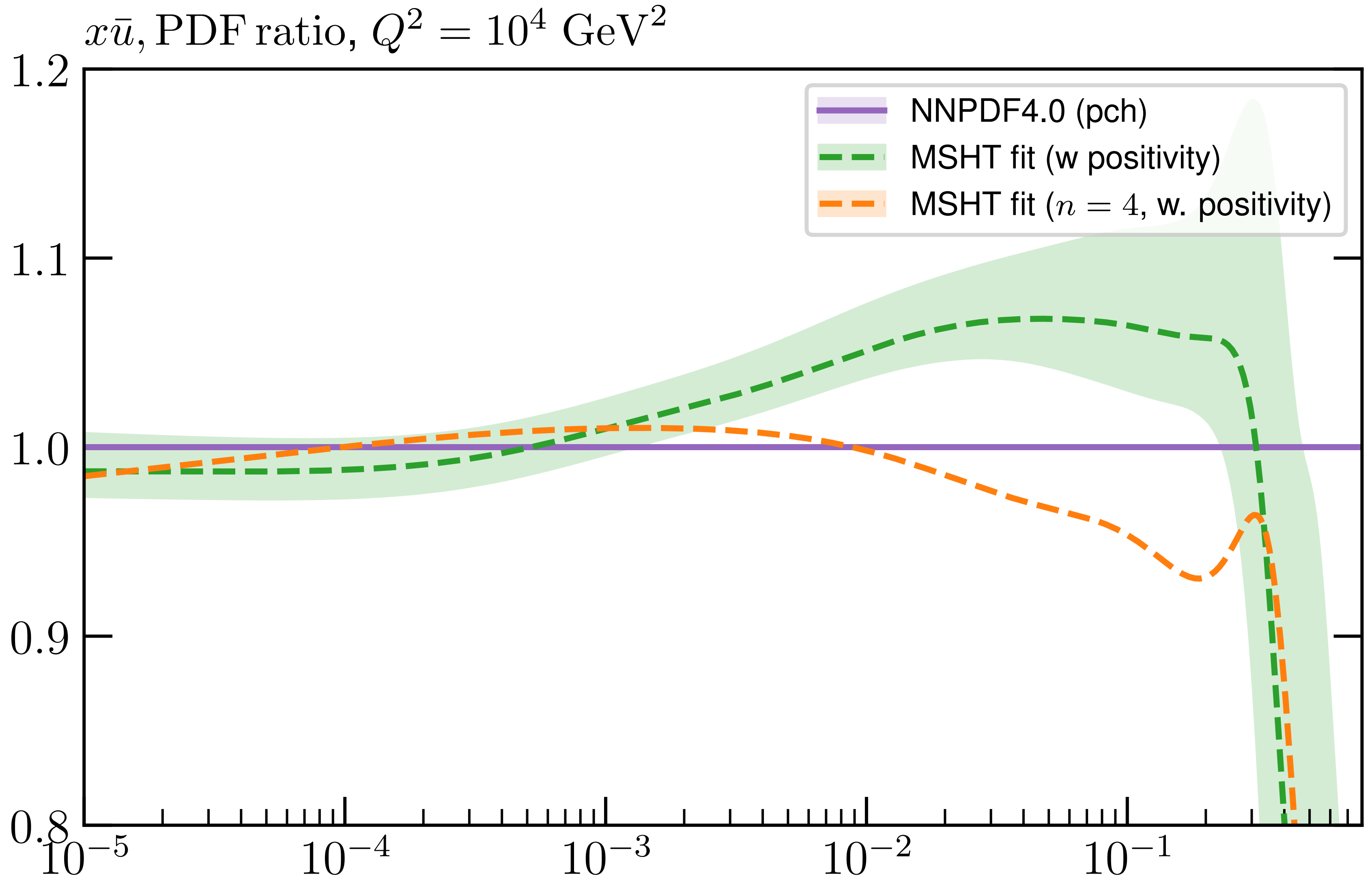


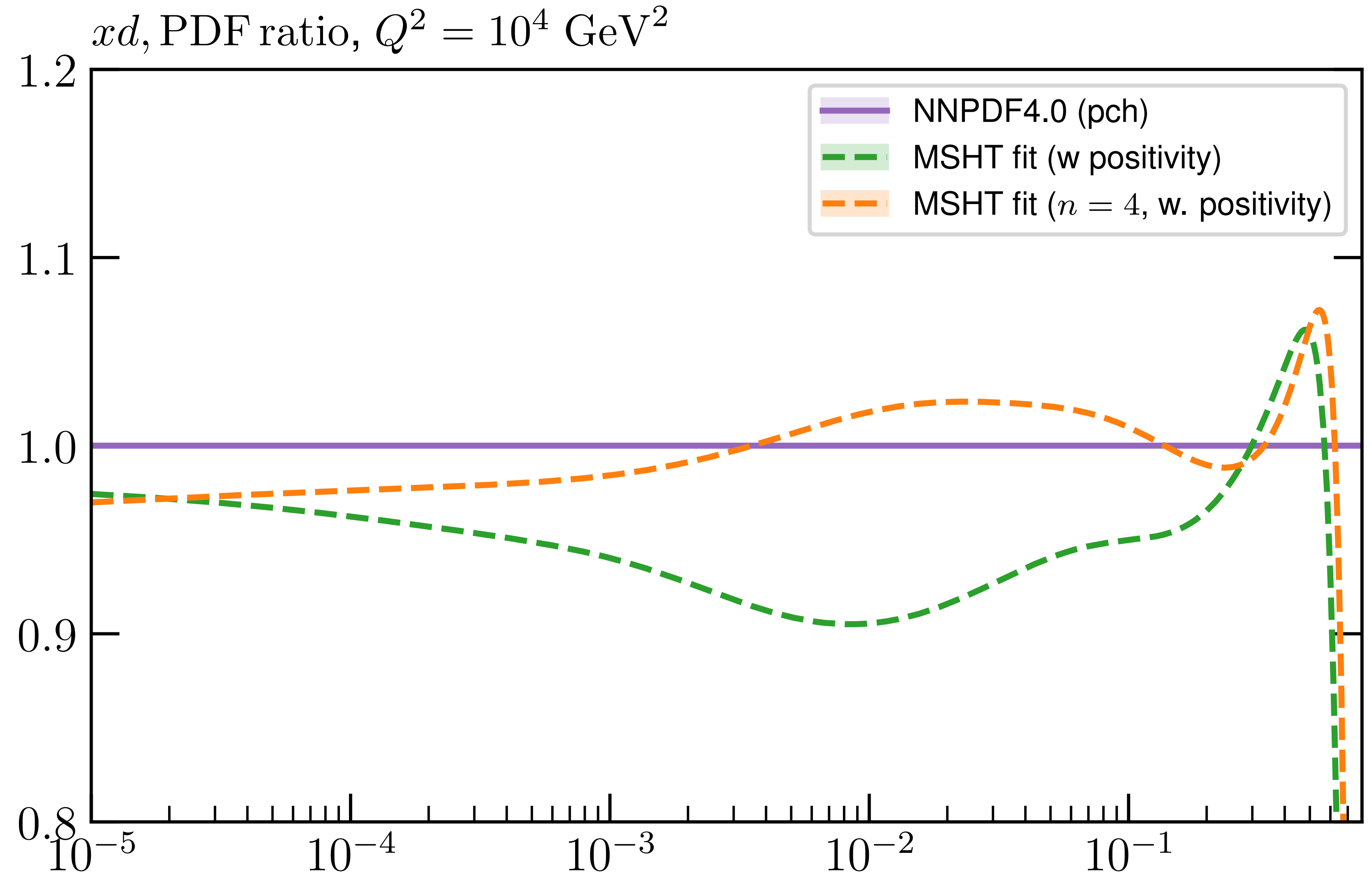
$T^2 = 10$

- Impact on PDFs somewhat larger than in closure test.

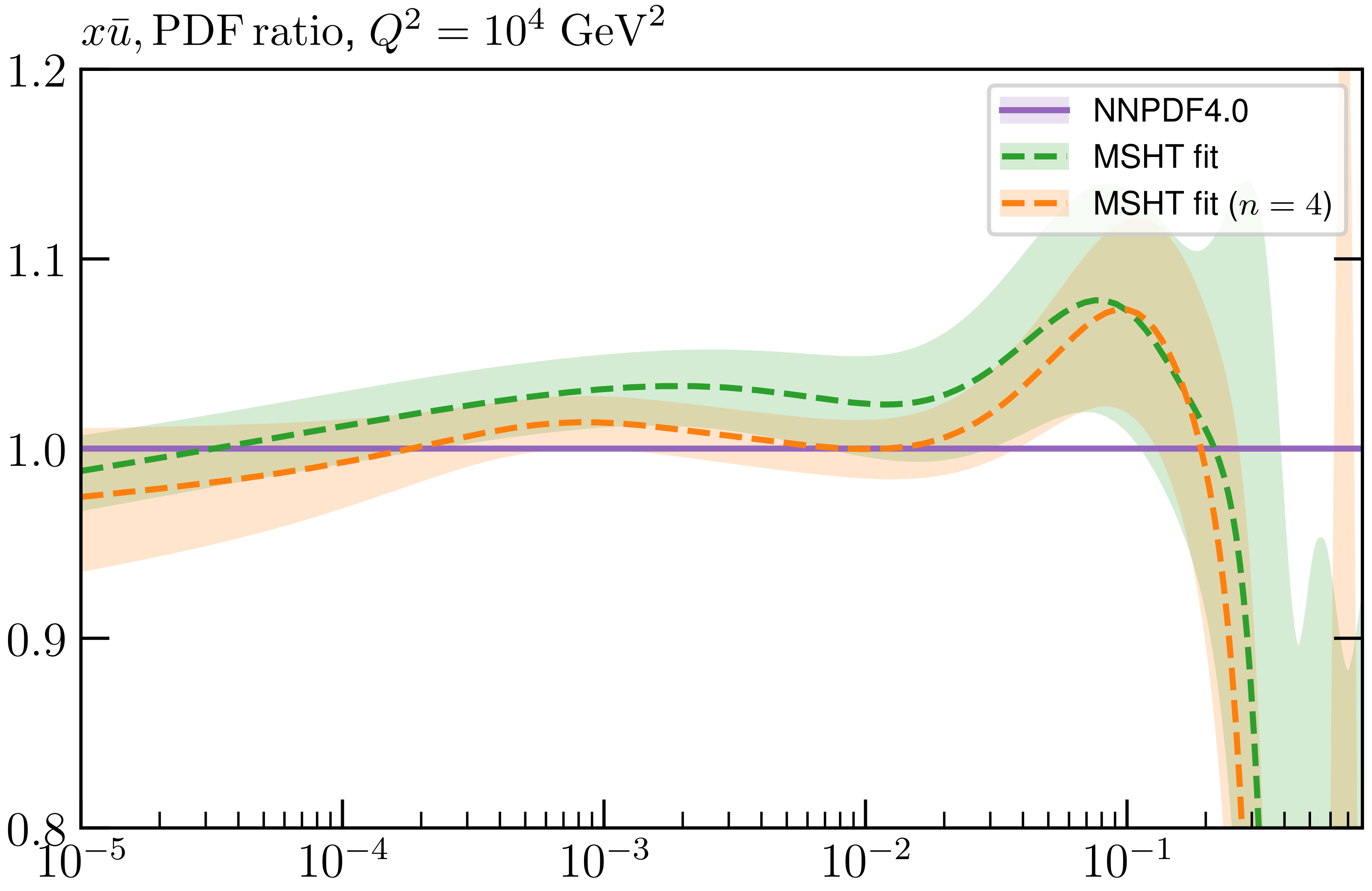
- Interestingly change in fit quality for $n = 4$ is similar to difference between MSHT and NNPDF fits…

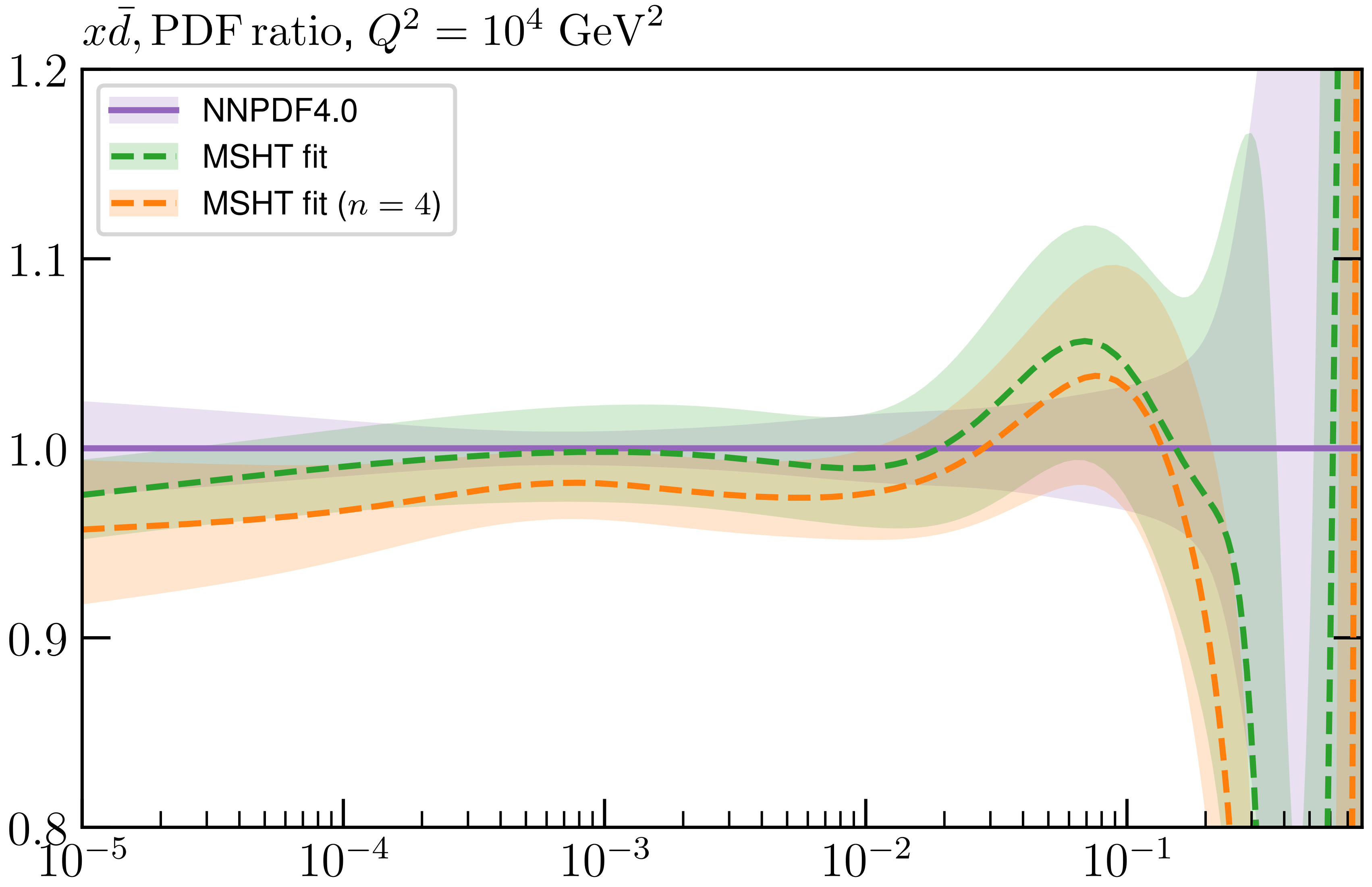
Pert. Charm





Fitted Charm

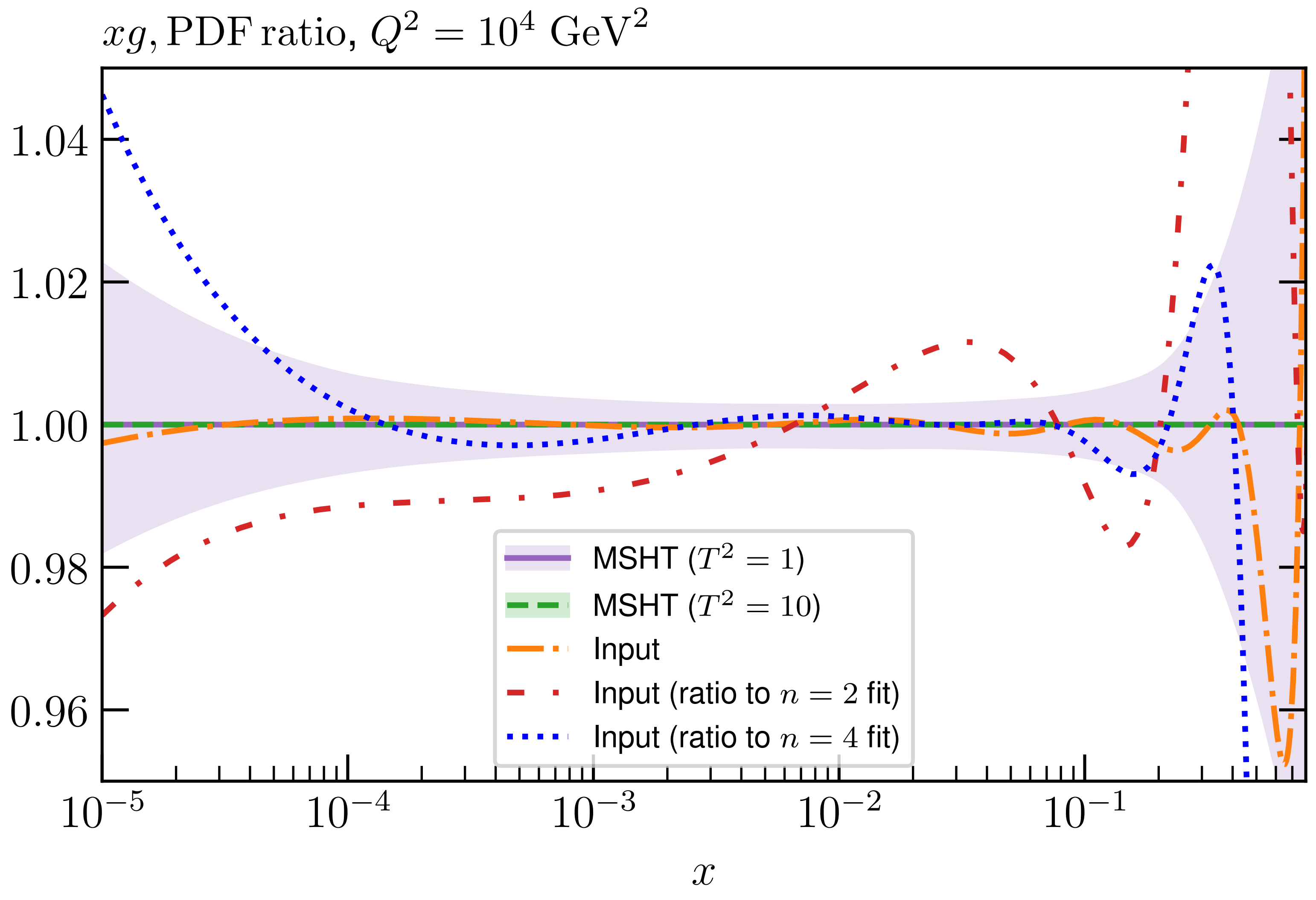




- Some indication that more restricted parameterisation ($\delta N_{\rm par} = -12$) gives PDFs closer to NNPDF, though less clear in fitted charm case. Less freedom consistent with poorer fit quality observed earlier.

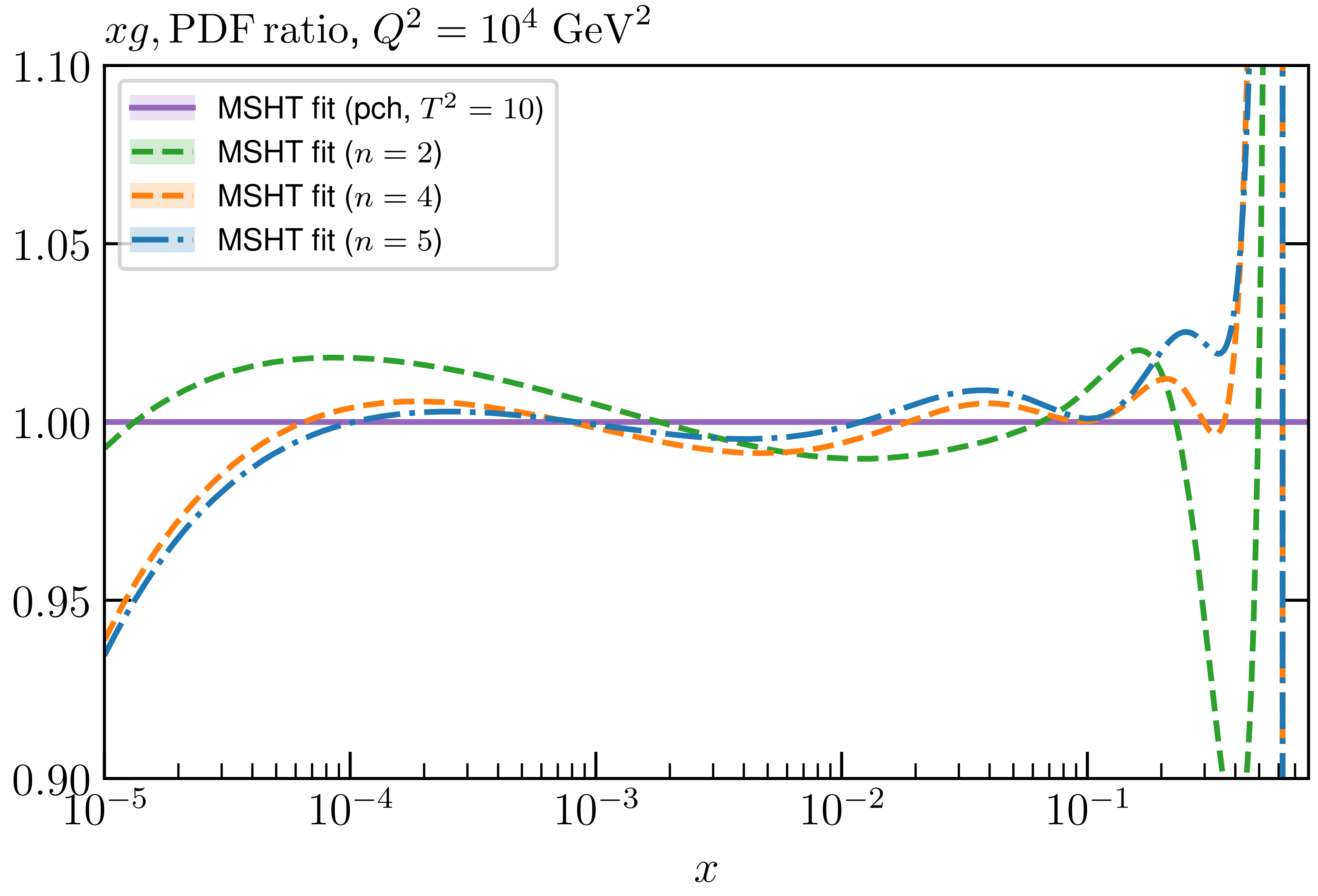
Overfitting?

- Various tests performed to examine whether the default MSHT parameterisation ($n_{\text{par}} = 52$) or more restricted ones results in overfitting, i.e. fitting noise in data.

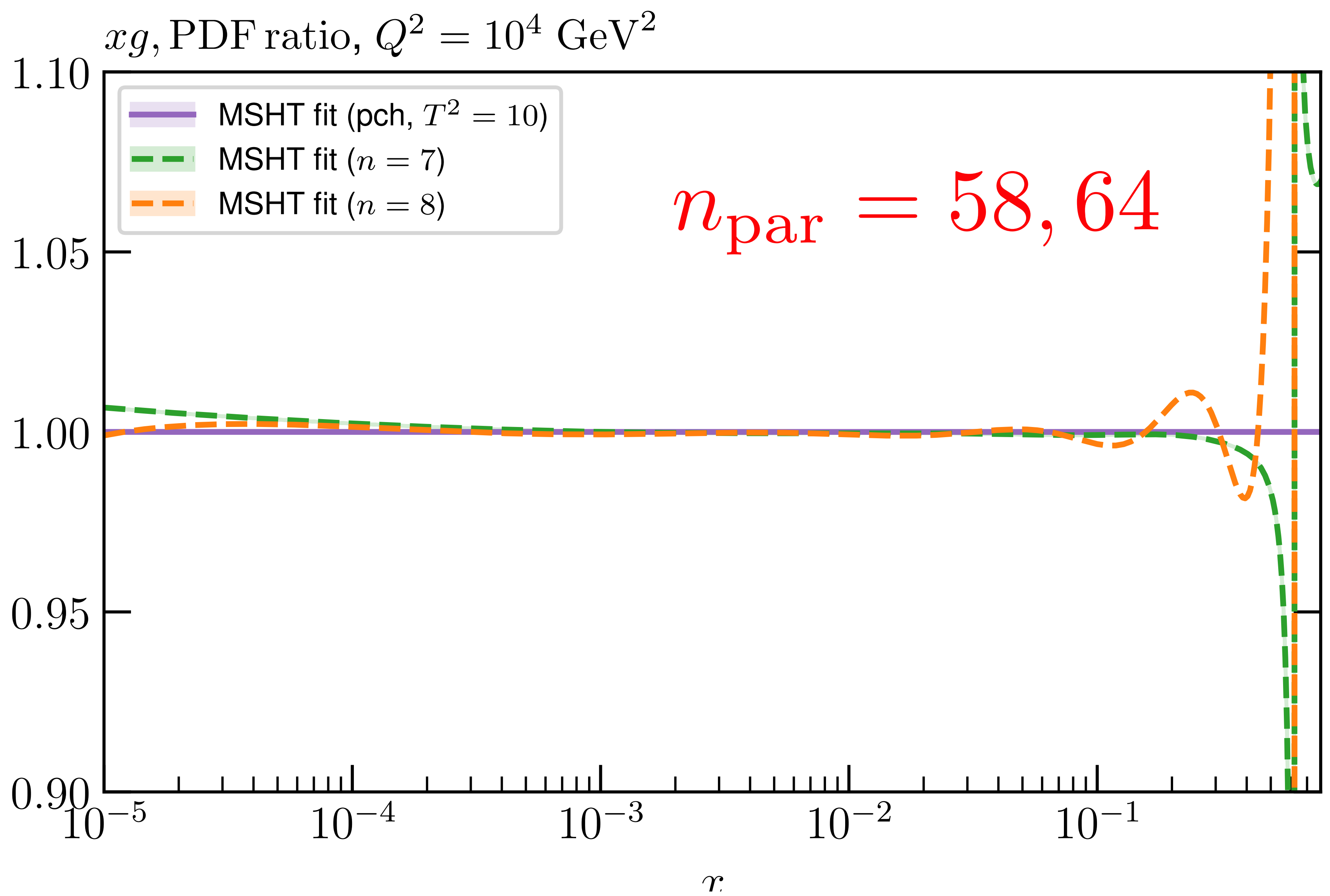


★ Closure test - restricted ($n_{\text{par}} = 28, 40$) parameterisation cannot match pseuodata well. Unlikely to result in overfitting in real fit!

★ Indeed precisely same effect seen in real fit…



- Indeed can try **extending** parameterisation, and find very mild impact on fit quality + PDFs - i.e. no clear trend for overfitting setting in with this ~ of parameters.



- In NNPDF fit extensive training/validation used to prevent overfitting (in context of NN fit). Fitting same data/theory with MSHT parameterisation we find similar fit qualities.

	NNPDF4.0 pch	MSHT fit (w positivity)
$\chi^2_{t_0}$	5928.3 (1.282)	5837.8 (1.262)

$\Delta\chi^2_{t_0}$: $-90.5\ (-0.02)$

NNPDF4.0	MSHT fit (w positivity)
5692.1 (1.233)	5651.0 (1.224)

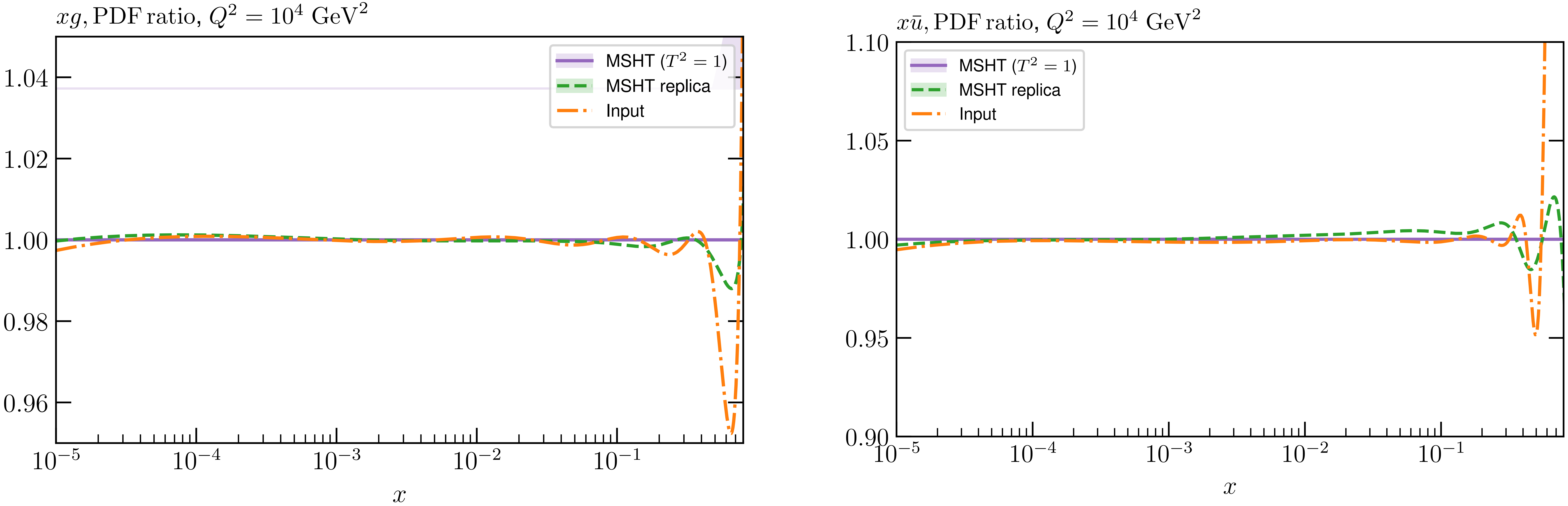
$\Delta\chi^2_{t_0}$: $-41.1\ (0.009)$

(After controlling for positivity)

- True that we find somewhat better fit quality, but improvement is spread across range of e.g. LHC DY datasets that independently constrain similar PDFs.

Global Fluctuated Closure Test

- Exactly the same closure test settings, but fluctuate pseudodata according to experimental uncertainties. Fit quality $\chi^2 \sim N_{\rm dat} \pm \sqrt{2N_{\rm dat}}$ expected (and found).
- Test faithfulness of MSHT parameterisation by producing MC replica set - perform 100 replica fits.



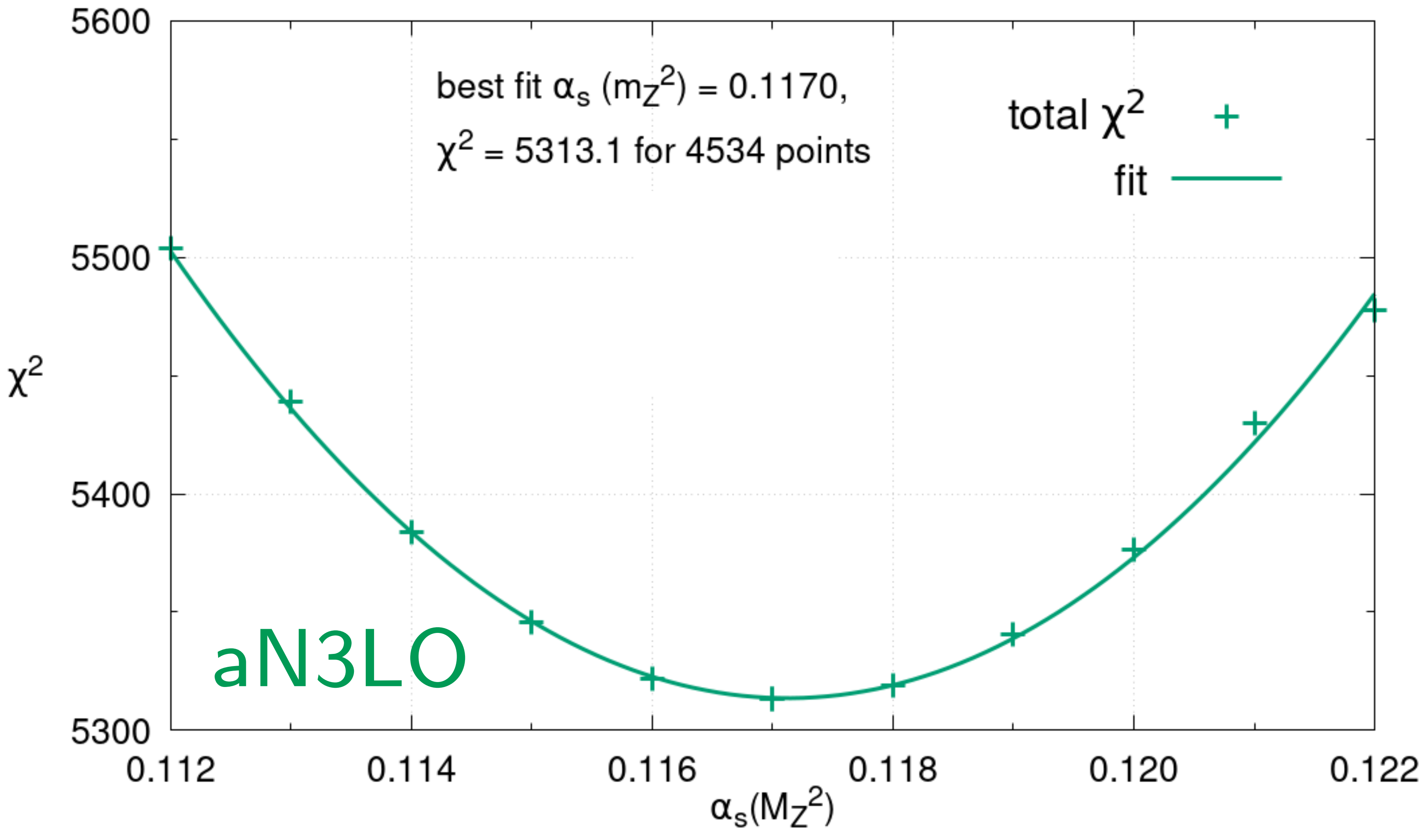
- Encouraging agreement between MC replica and Hessian uncertainties. In particular can use these result to show that for each replica fit, the truth lies within its 1σ uncertainty with 68% frequency - i.e. uncertainties are faithful in fluctuated fit.

Uncertainty Evaluation

- In textbook case, would simply take $\Delta\chi^2 = 1$ from minimum, to give (roughly):

$$\alpha_S = 0.1170 \pm 0.0005$$

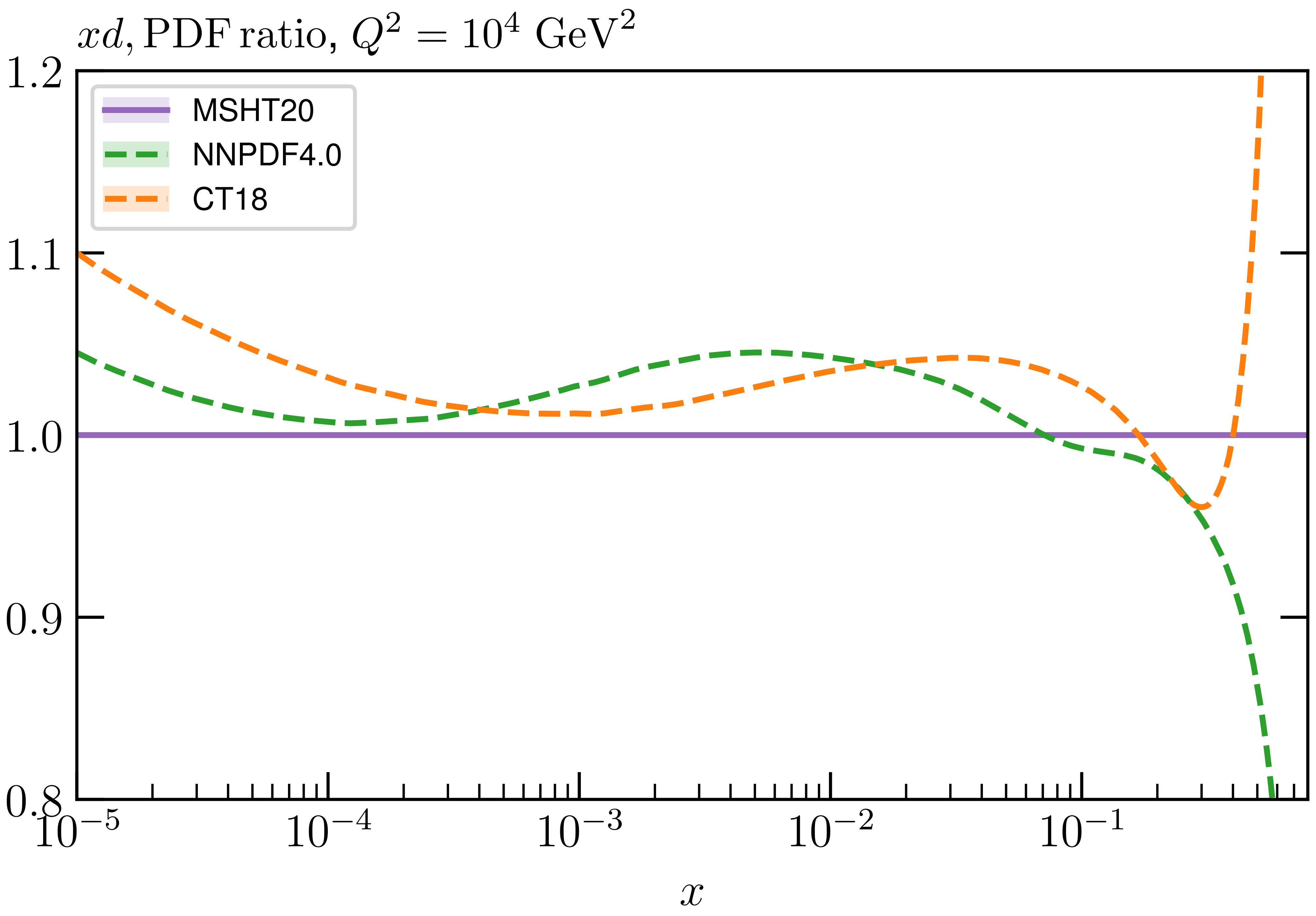
- However from discussion before, expect to be too aggressive. Enlarged tolerance needed.

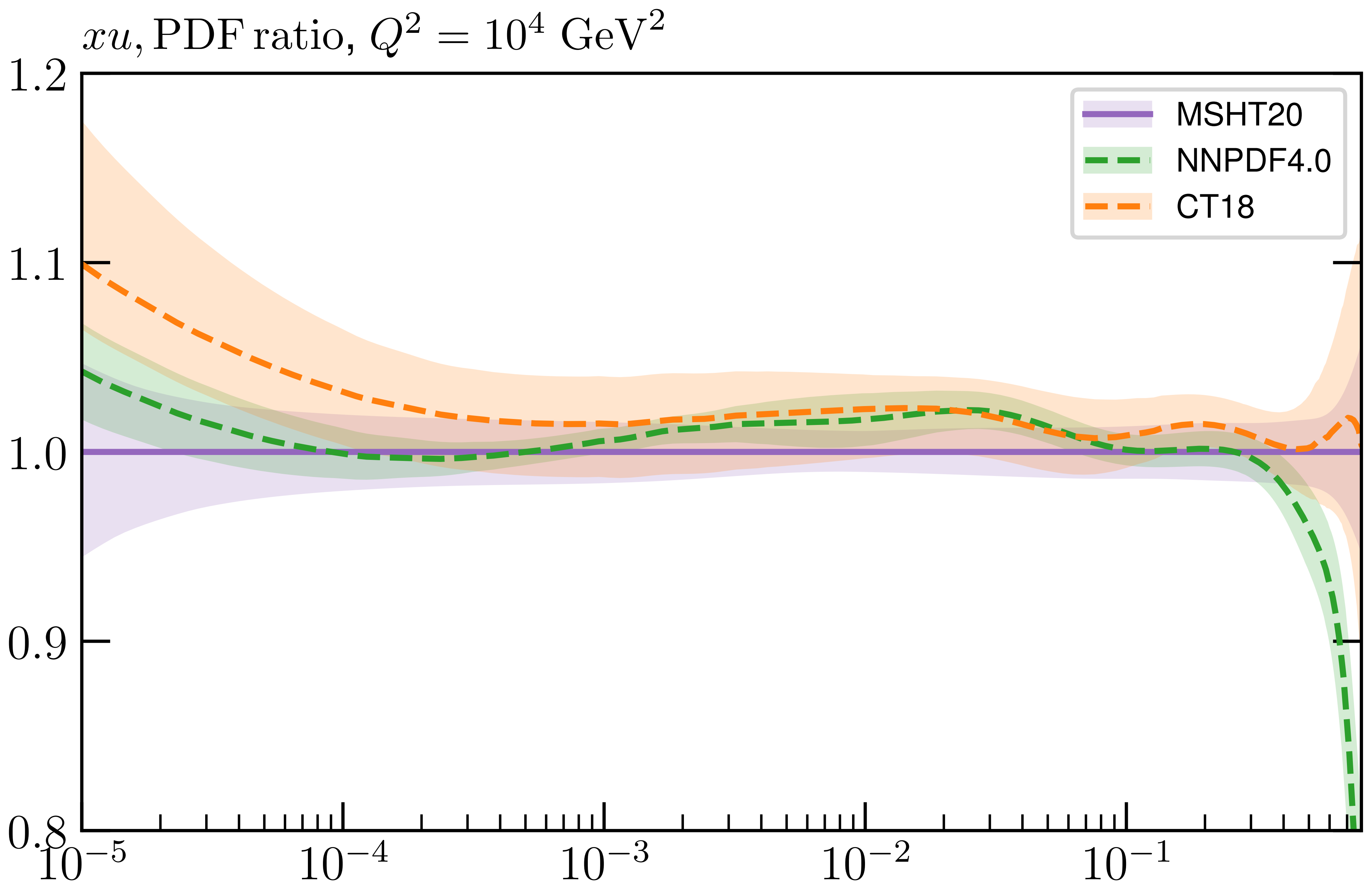


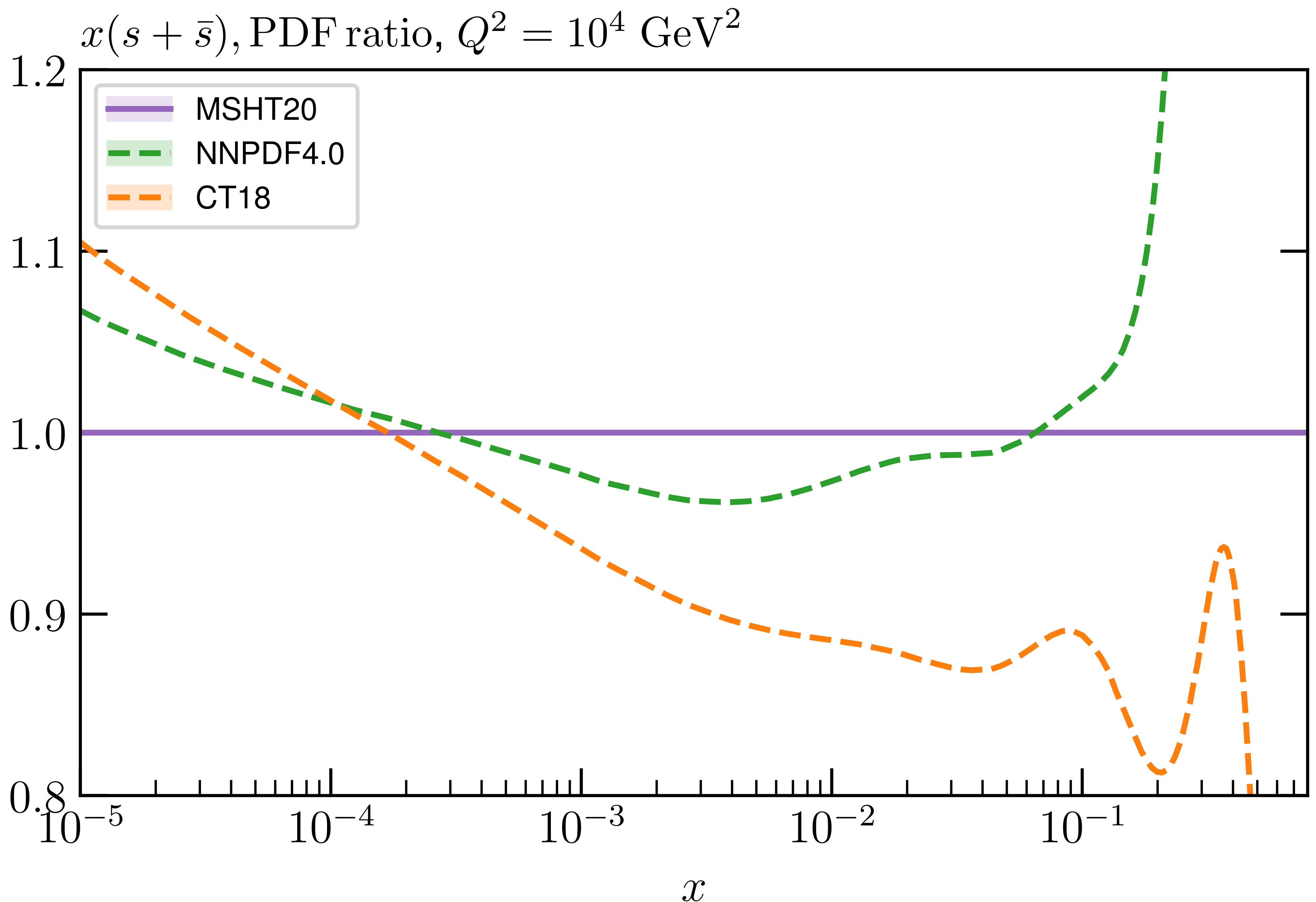
- In MSHT apply `dynamic tolerance' criterion. Briefly:
  - ⋆ Evaluate individual χ^2 profile for each dataset
  - ⋆ Deviation with α_S increasing/decreasing monitored and limited such that this does not exceed 'hypothesis testing' criterion $\Delta\chi^2 \lesssim \sqrt{2N}$ i.e. remains good according to this measure.
  - ⋆ In toy model can show given two datasets in tension that* PDF uncertainty $\propto$ difference (unlike $T = 1$).
  - ⋆ Will result in one dataset setting most stringent upper/lower limits, but find many others with similar limits, i.e. uncertainty not driven by a single (potentially problematic) dataset.
  - ⋆ Broadly corresponds to $T \sim 3$.

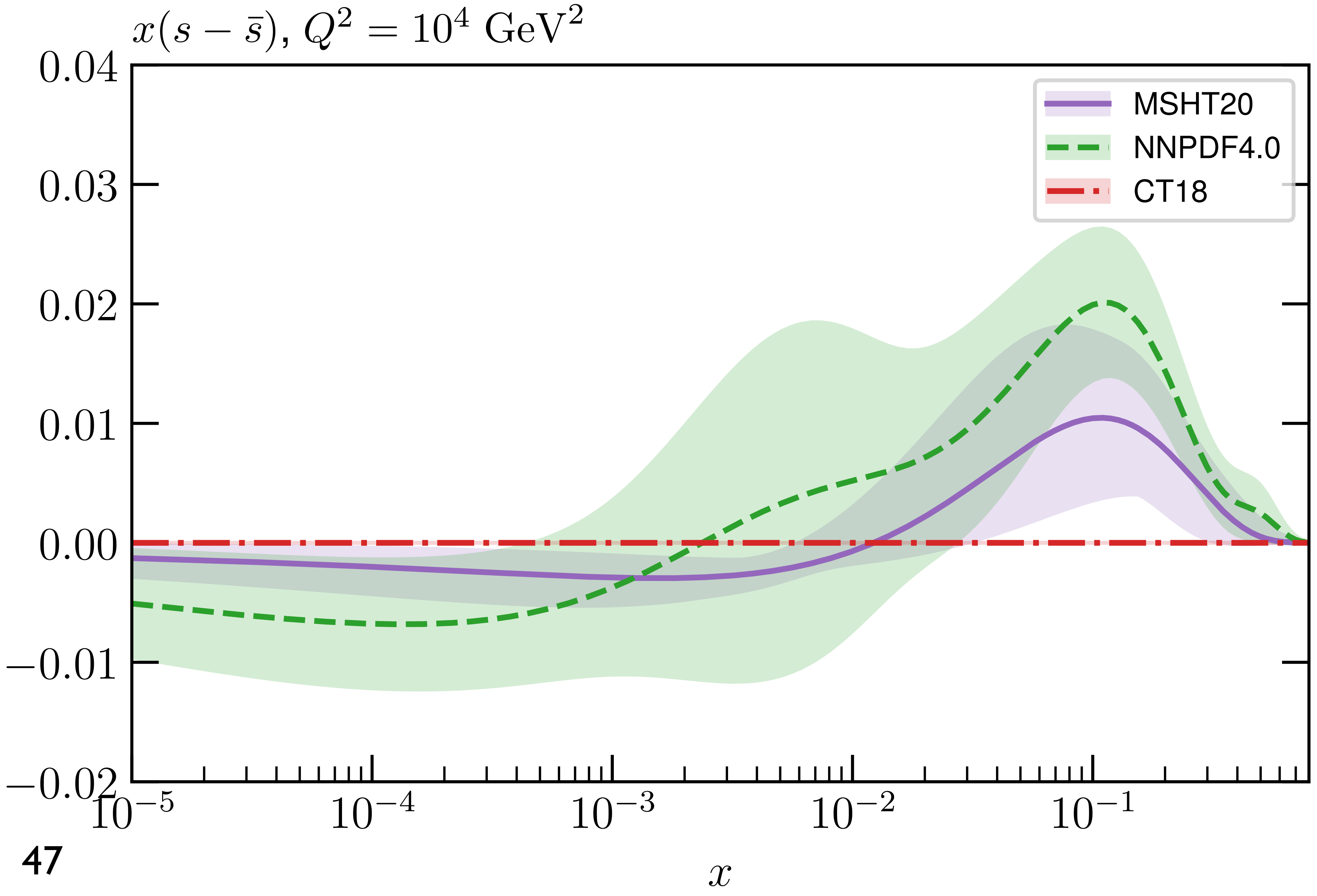
Backup

*For experts: in reality, limit is rescaled by best fit value: $\Delta\chi^2 \lesssim \chi^2_{n,0}\left(\frac{\xi_{68}}{\xi_{50}} - 1\right)$


xd, PDF ratio, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
1.2
1.1
1.0
0.9
0.8
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
x


xu, PDF ratio, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
1.2
1.1
1.0
0.9
0.8
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}


$x(s+\bar{s})$, PDF ratio, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
1.2
1.1
1.0
0.9
0.8
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
x


$x(s-\bar{s})$, $Q^2 = 10^4$ GeV^2
MSHT20
NNPDF4.0
CT18
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
x

New Developments : New Physics + PDFs

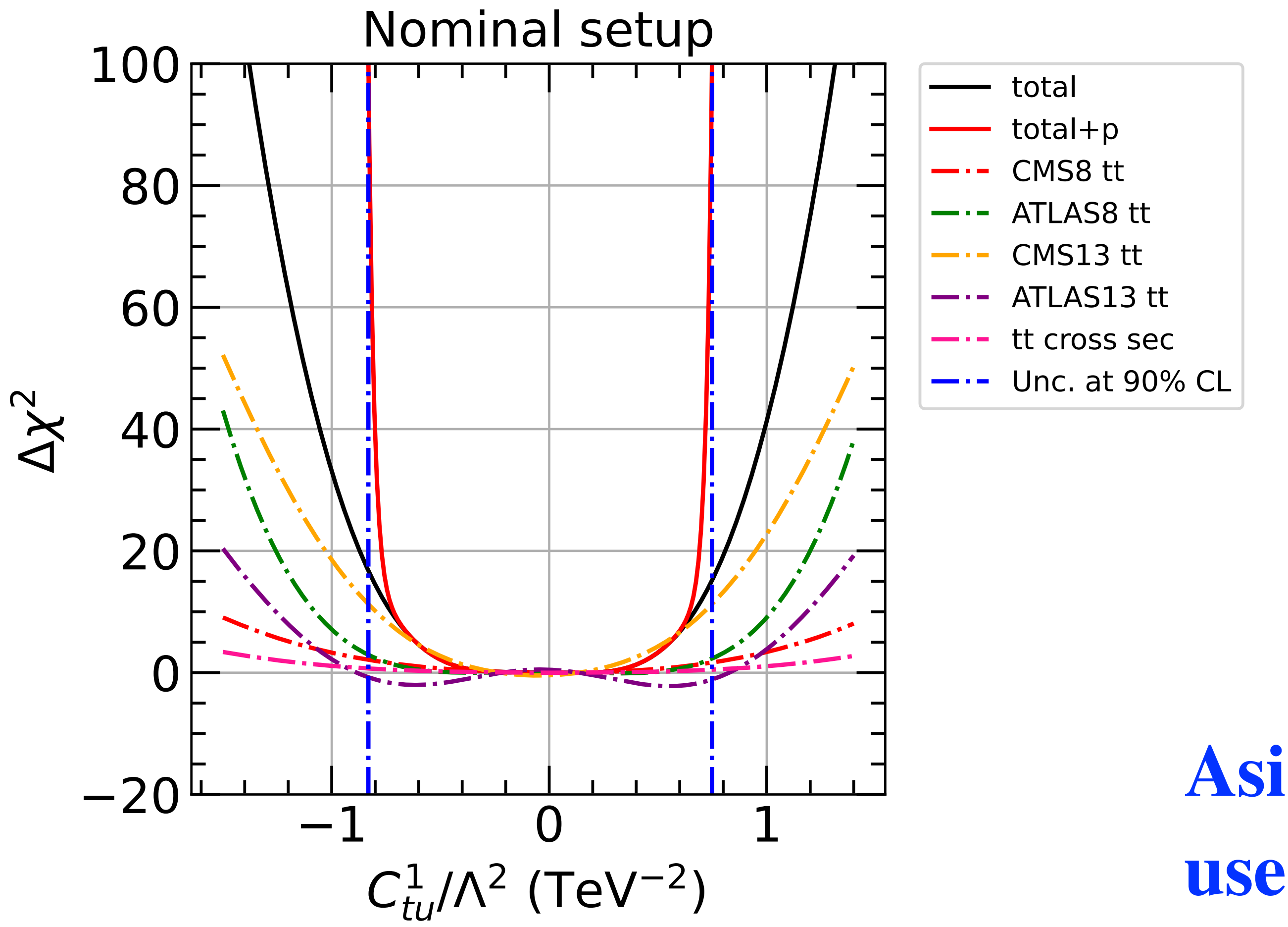
- Key element of LHC precision physics: looking for indirect signs of new physics in high energy data. Parameterise in **SMEFT**:

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

- 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?

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

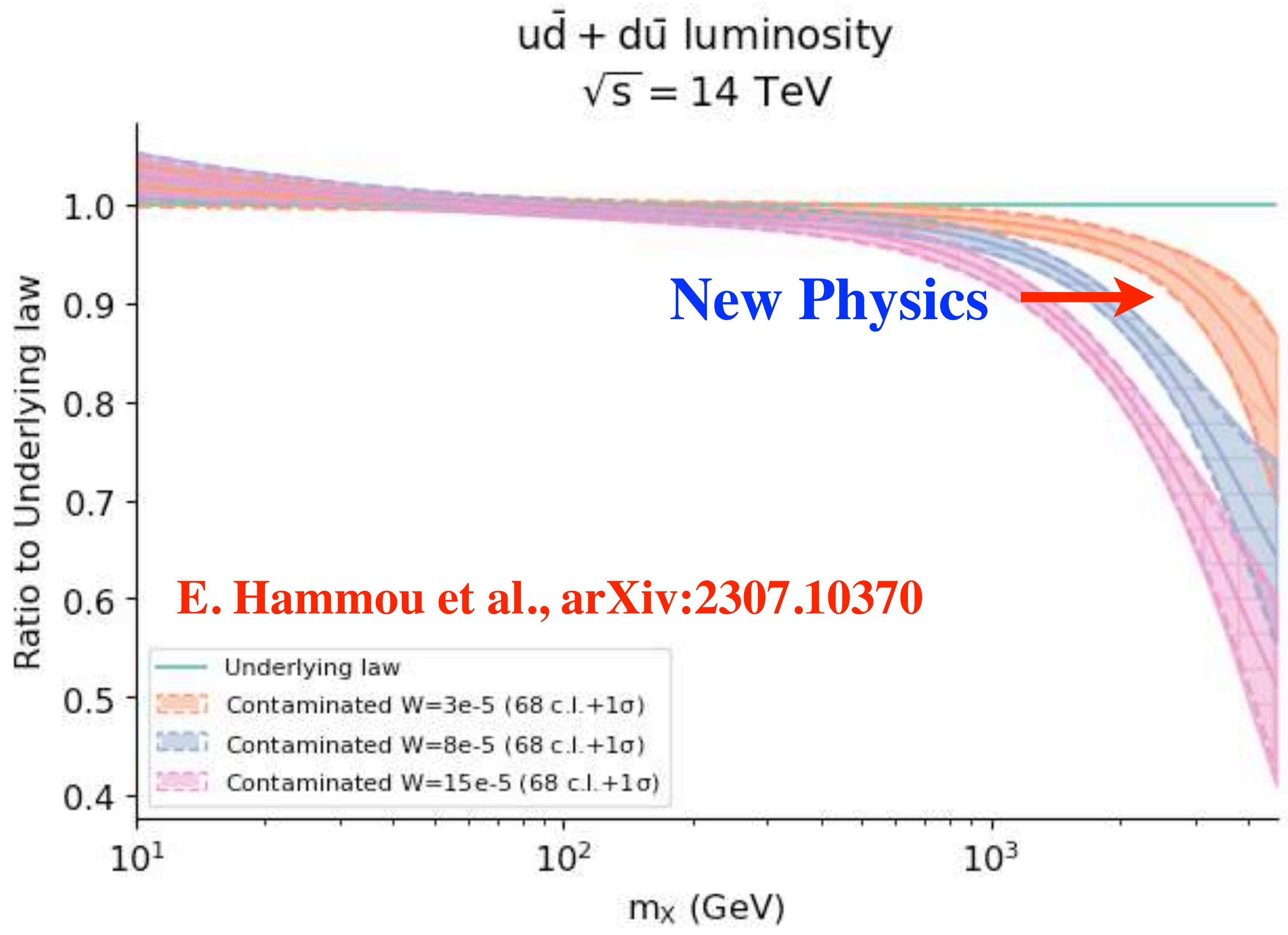
- For certain models **it can**. New physics in high mass DY absorbed into PDFs, with still reasonable fit quality.

Public `SIMUnet' tool for this:

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

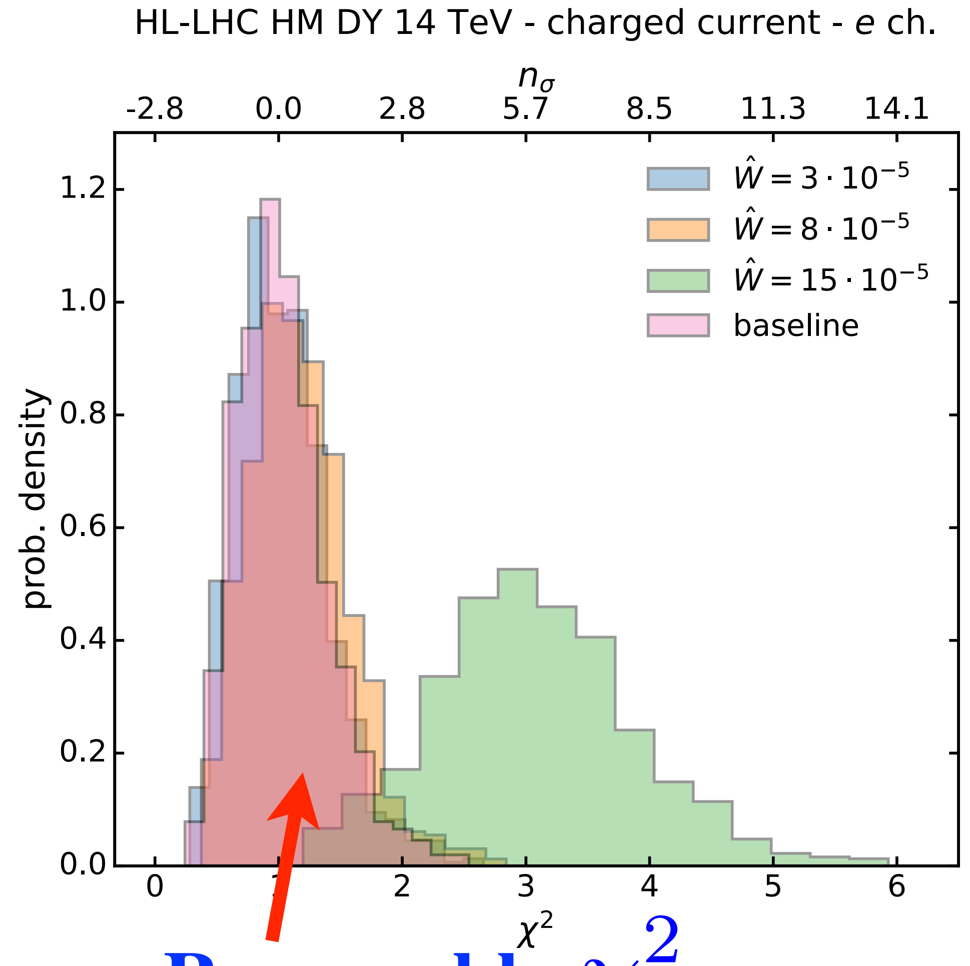
PBSB collab., arXiv:2402.03308

- Solutions?



New Physics

E. Hammou et al., arXiv:2307.10370



Reasonable χ^2

✦ 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…

N3LO - What do we know?

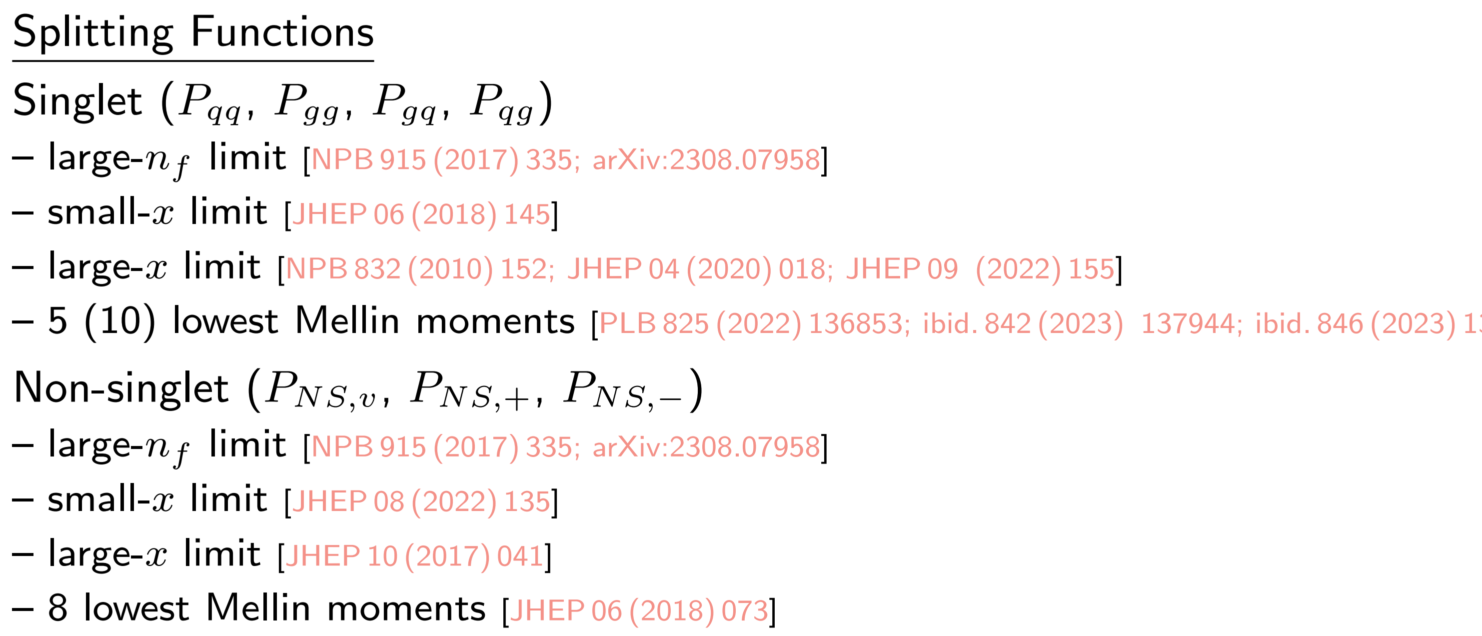
- **Approximate** $\neq$ **poorly known**!

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

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

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

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



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]

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

★ **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.

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

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

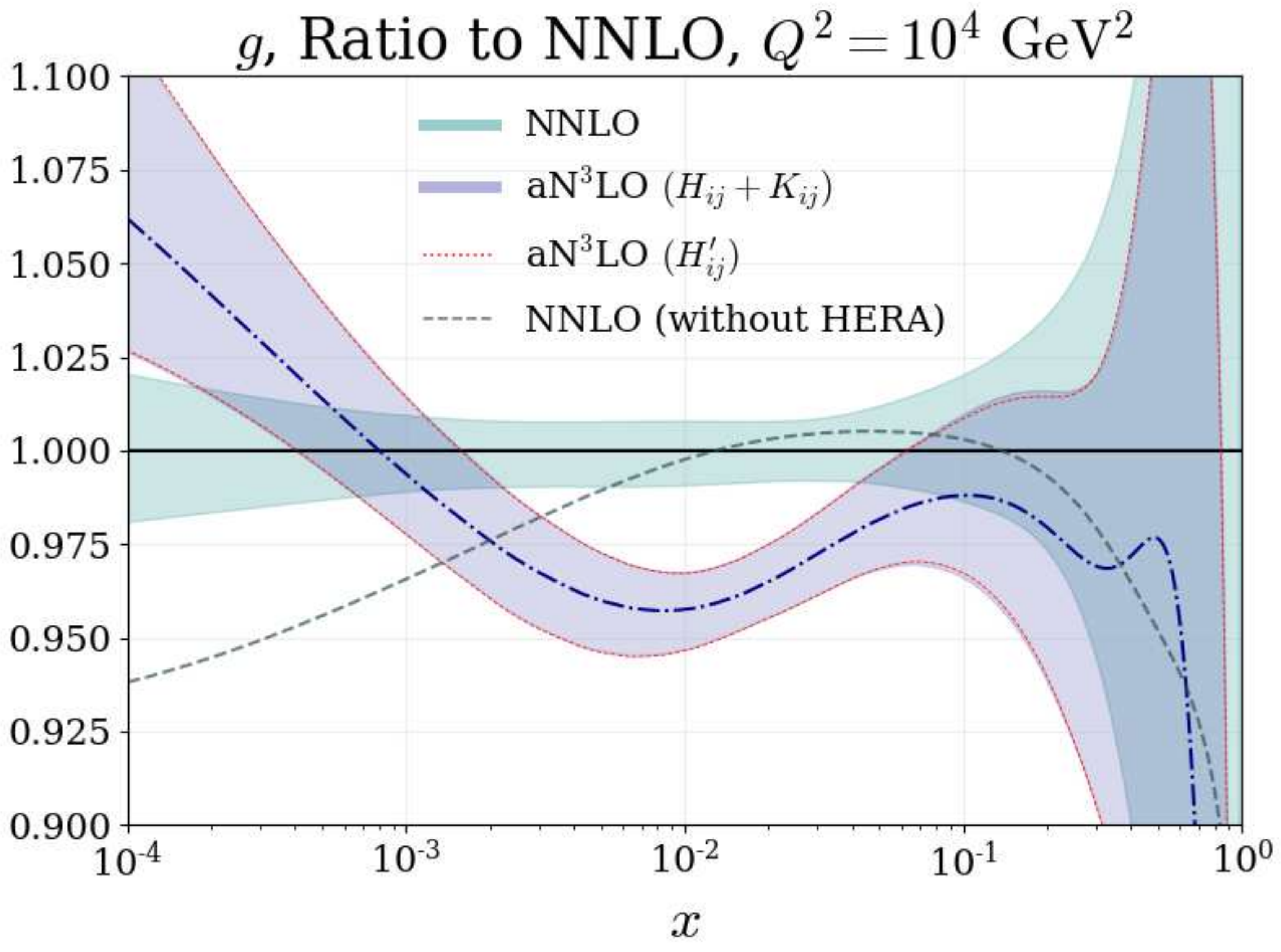
- First three ingredients now largely known with sufficient precision to give close to a N3LO fit. Final ingredient clearly the bottleneck for that - approximation + uncertainty required.

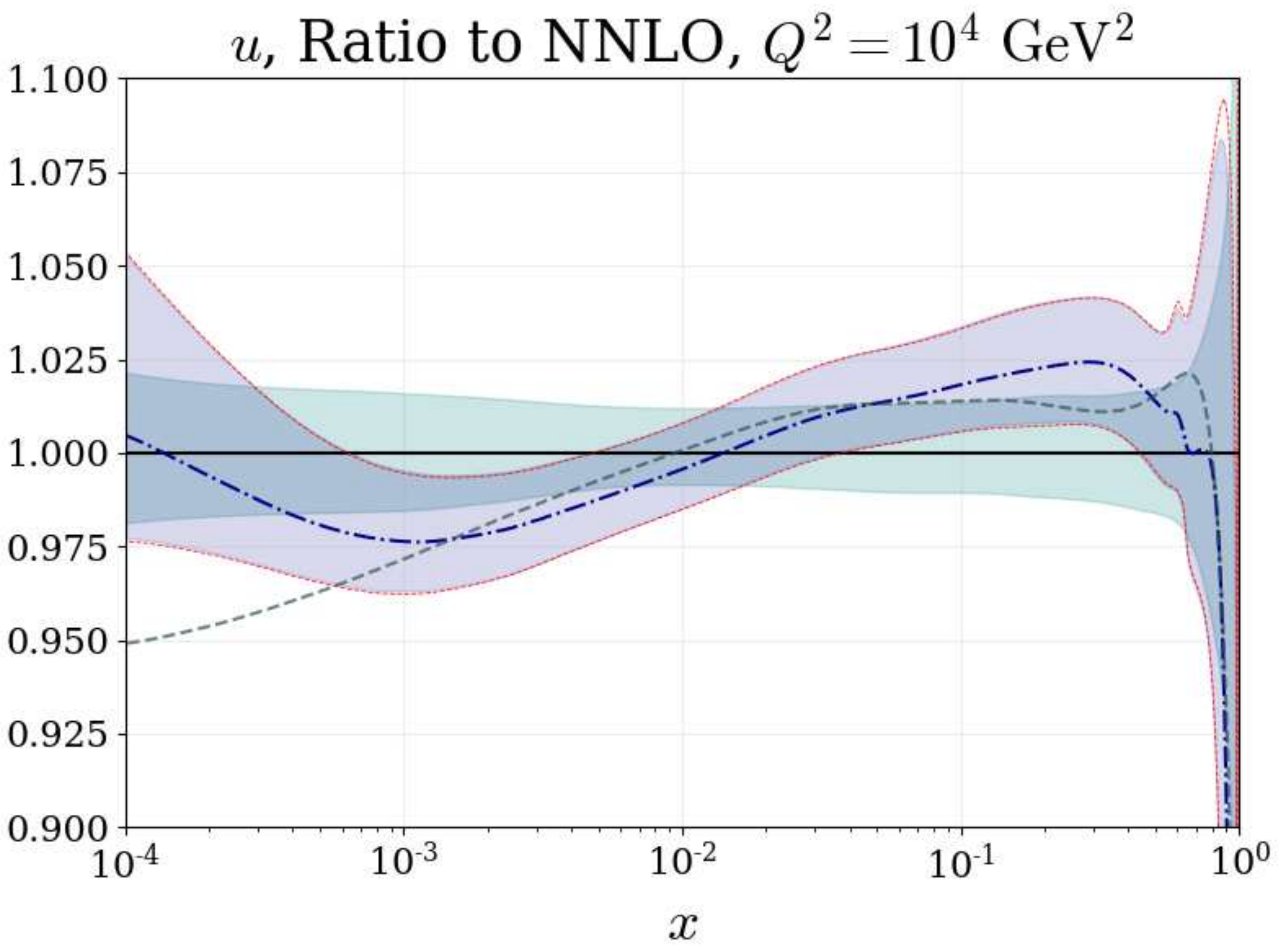
MSHTaN3LO

- 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

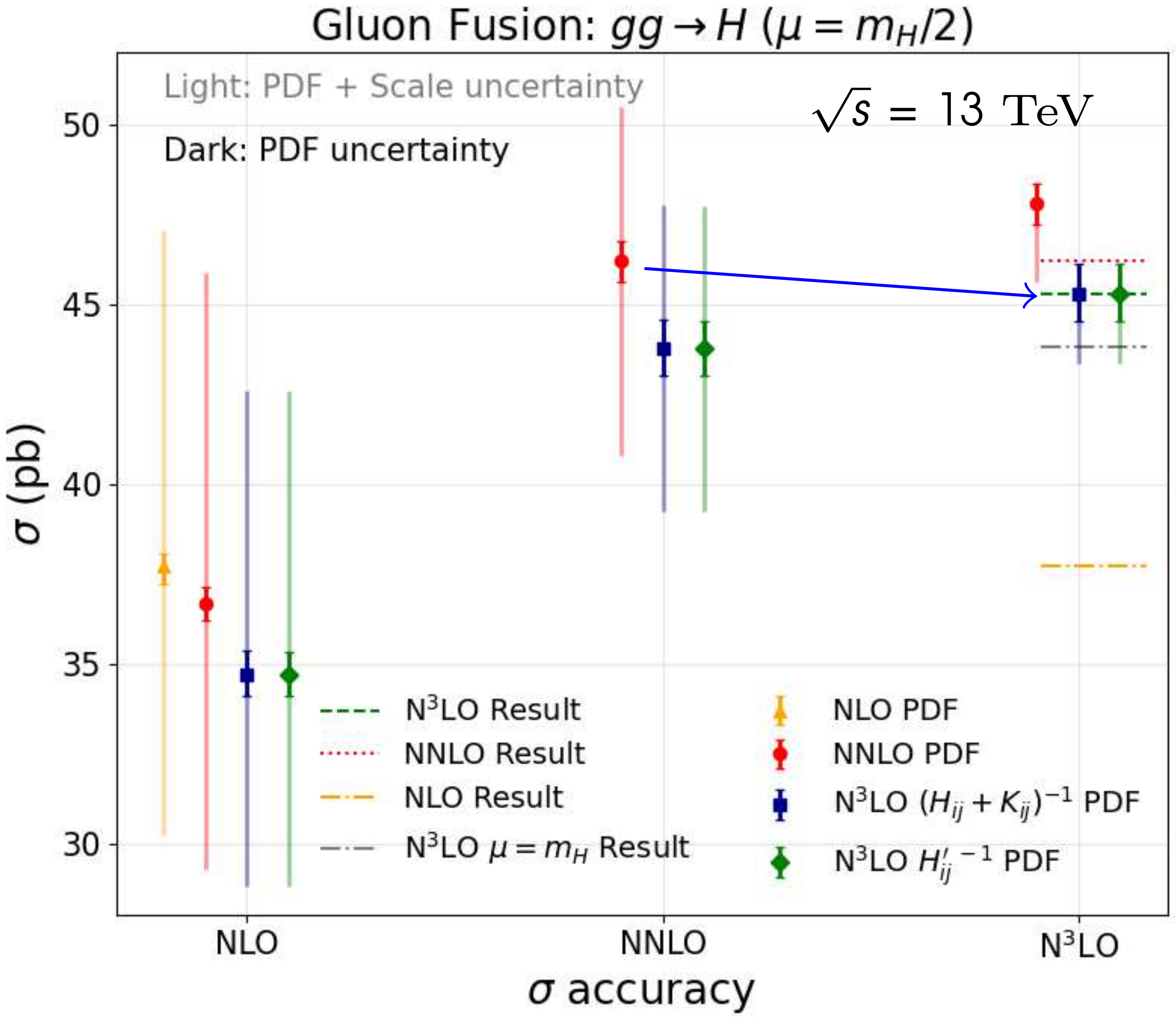
- Evidence that aN3LO reduces tensions between low and high x regions.
- Largest change is in gluon at low and intermediate x . Some change in e.g. quarks at high x .



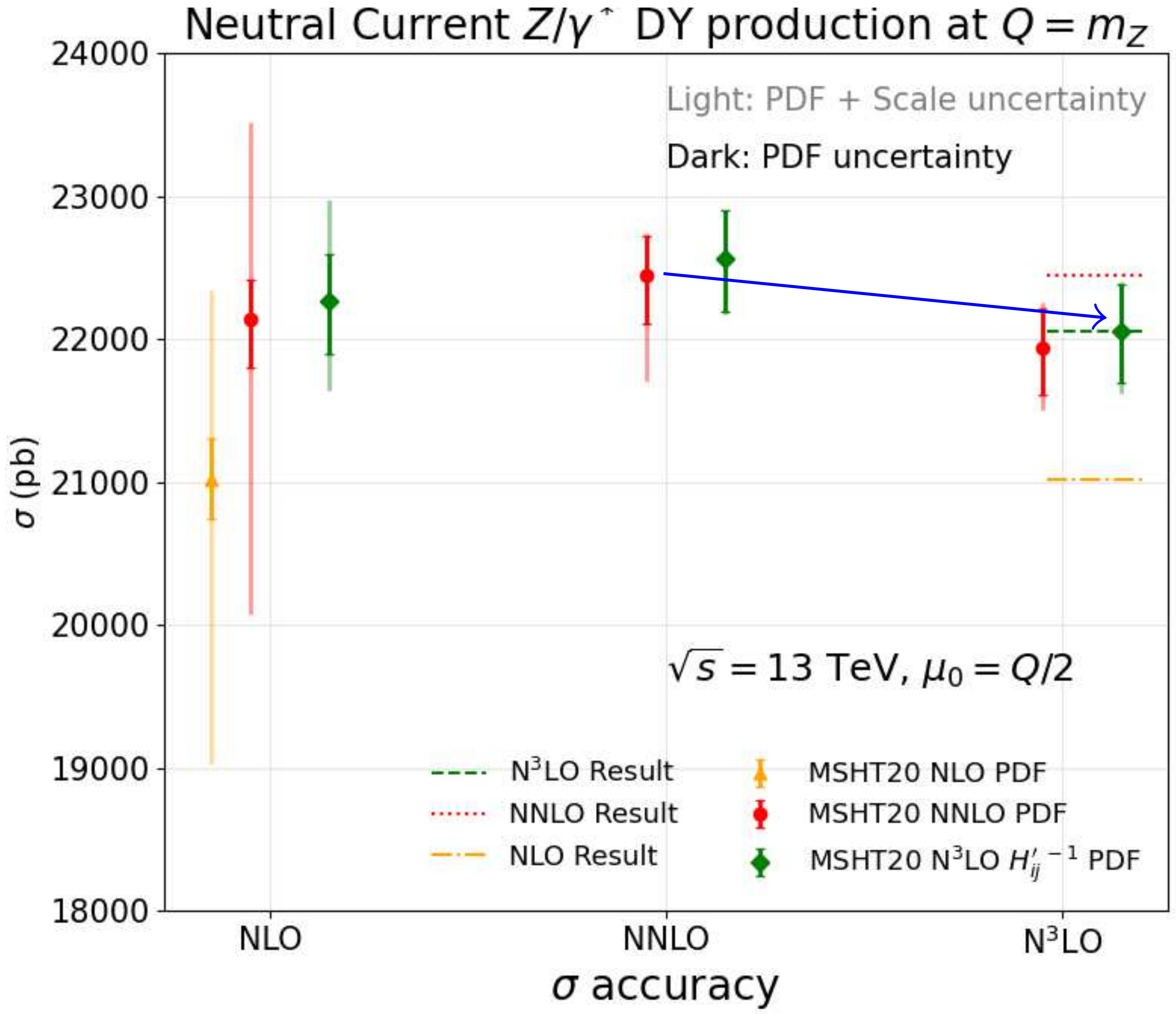


- More recent **NNPDF4.0aN3LO** analysis sees qualitatively similar results, some quantitative differences.

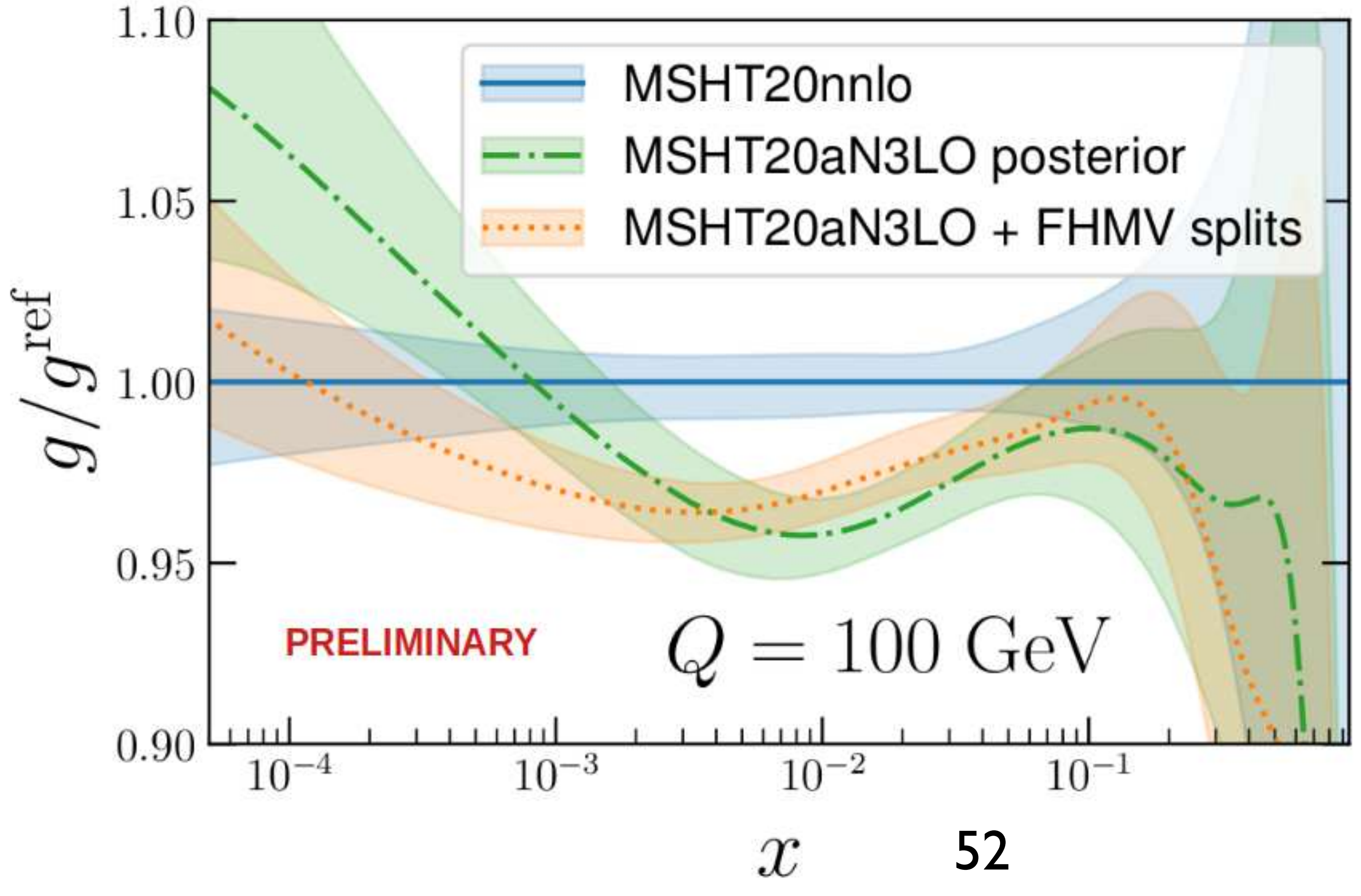
- Change in gluon corresponds to reduction in e.g. ggH at N3LO - improves stability.



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

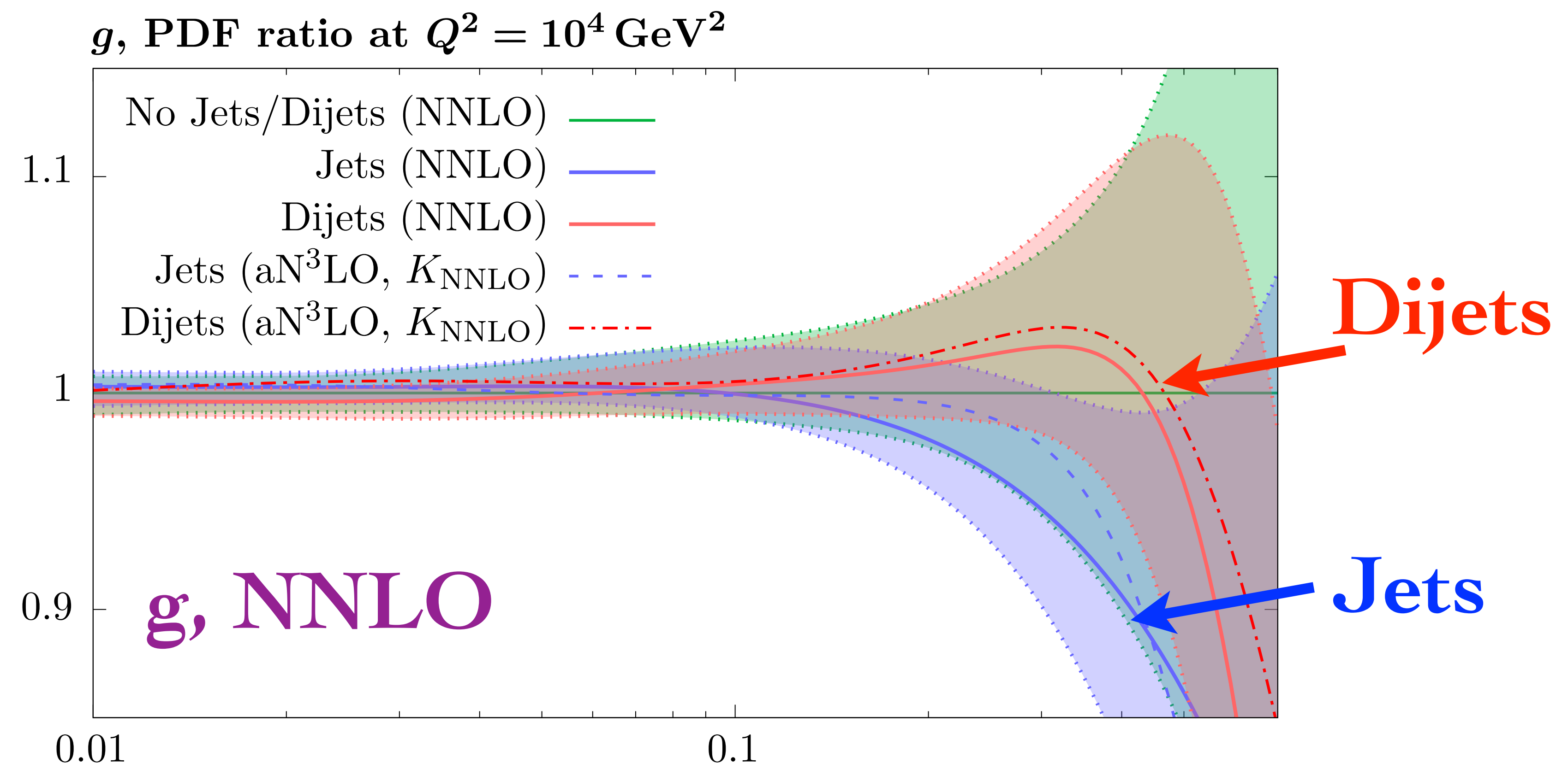


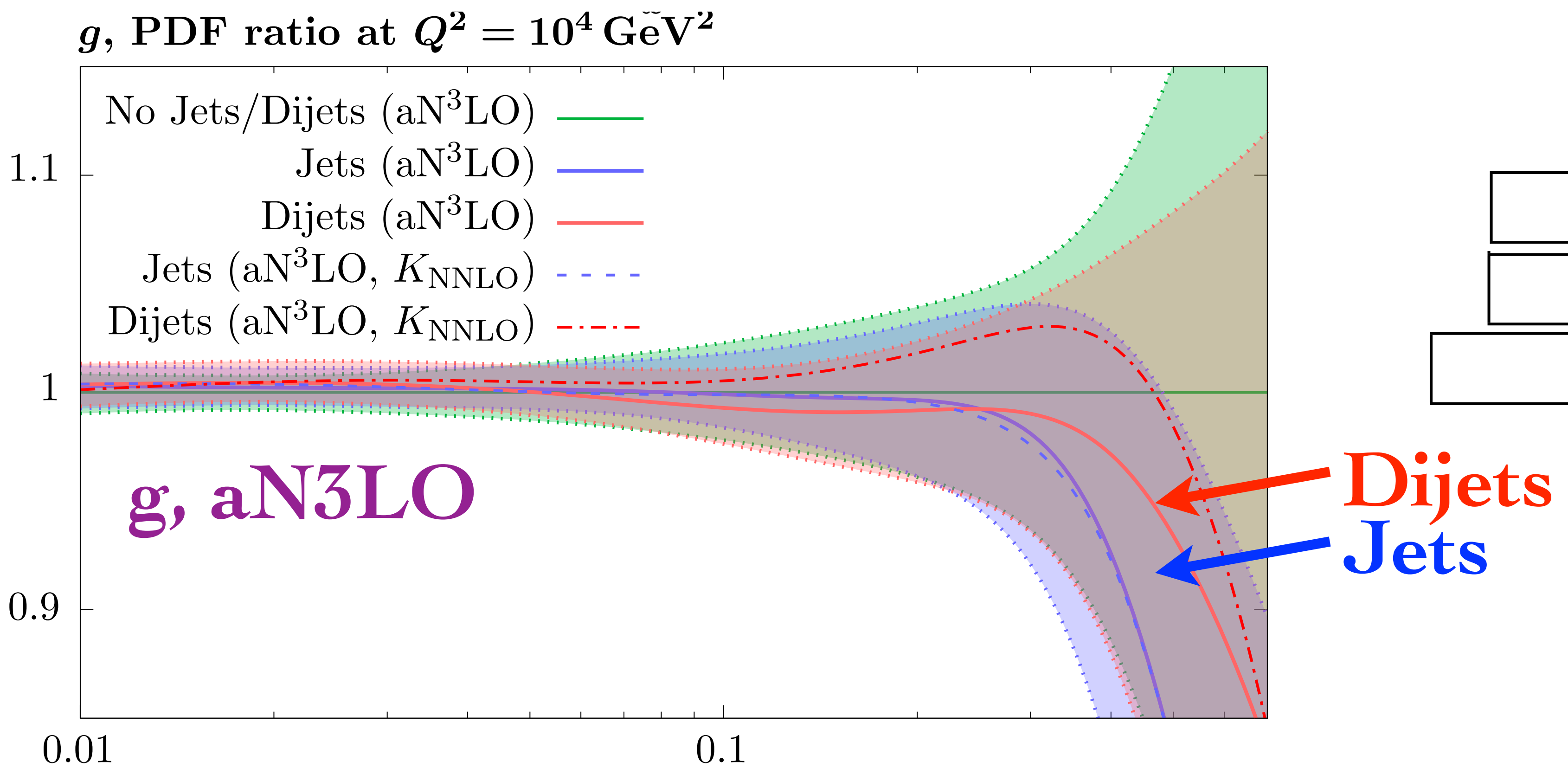
- Have studied impact of newer splitting function information. Moderate impact, within uncertainties.



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).





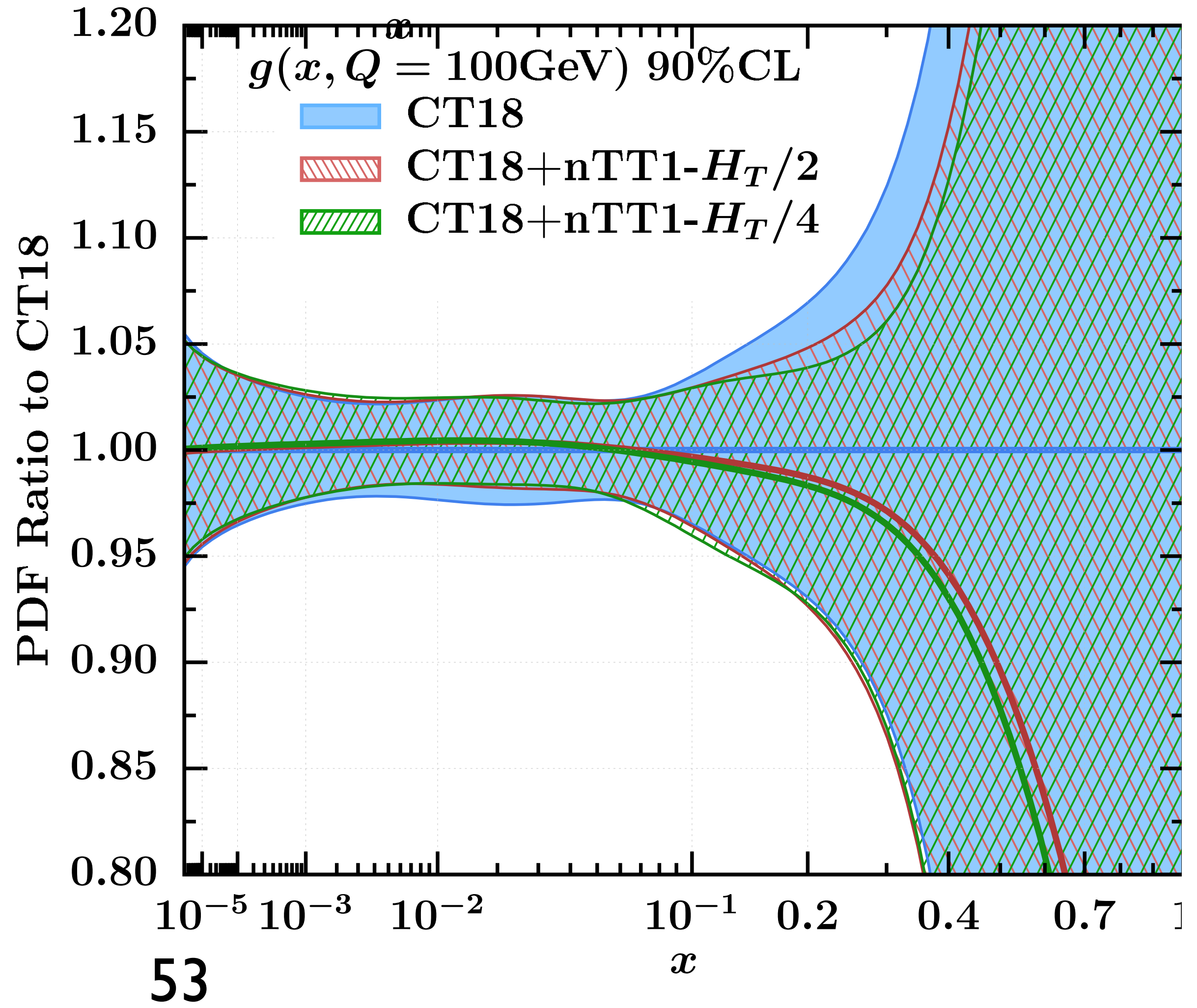
$\chi^2/N_{\rm pts}$

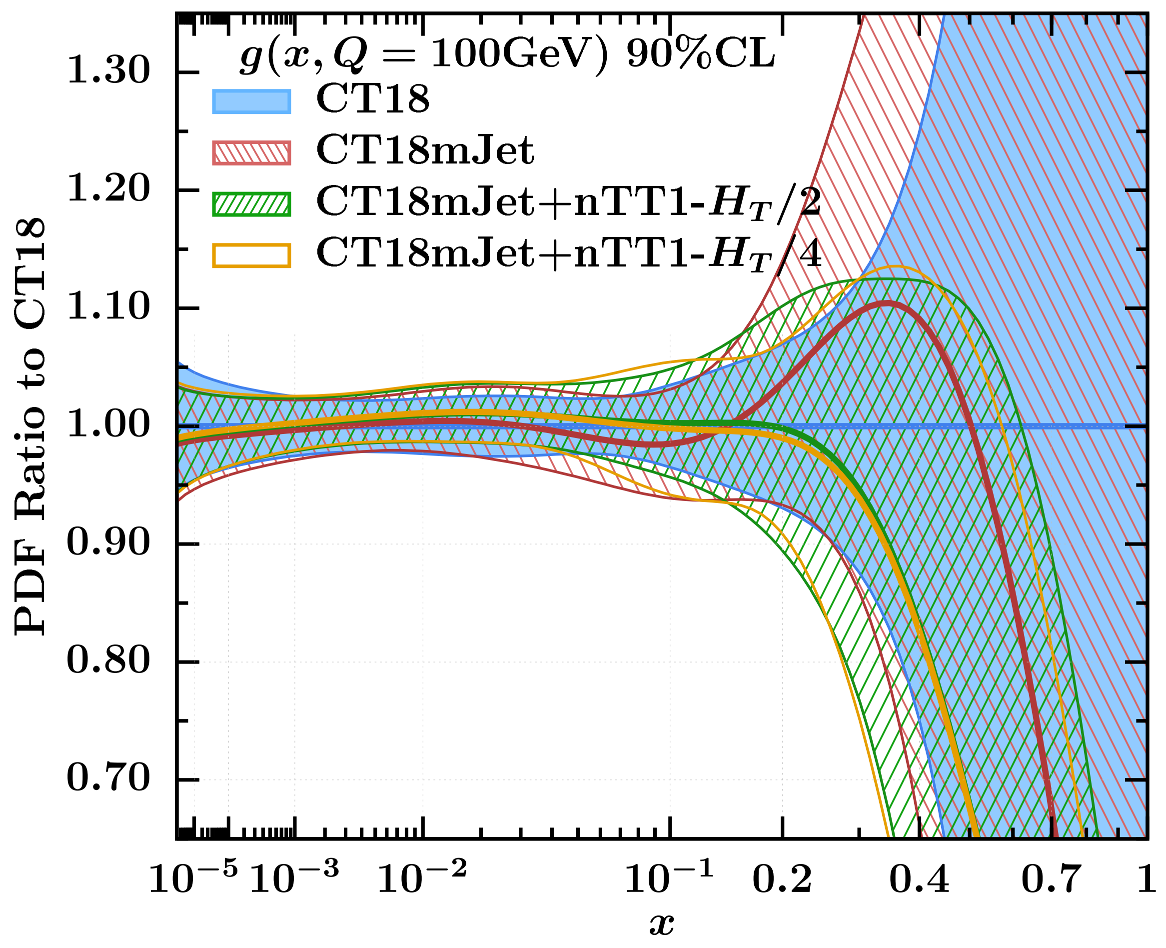
	$N_{\rm pts}$	NNLO	aN^{3}LO
Total Jets	643	1.67	1.63
Total Dijets	266	1.13	1.04

MSHT, arXiv:2312.12505

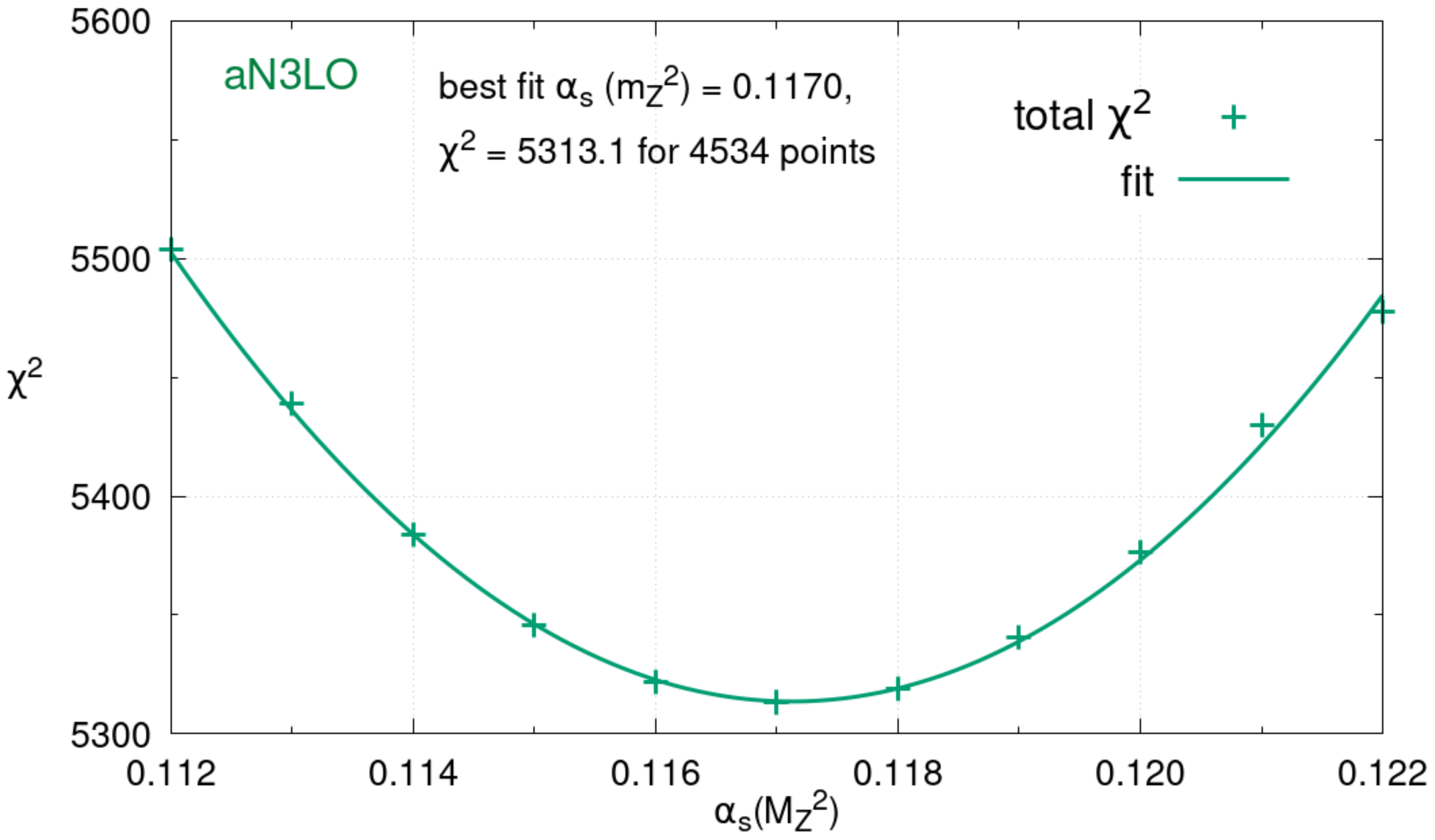
- ✦ 13 TeV $t\bar{t}$: study within **CT** global PDF fit.
- ✦ Impact moderate but non-negligible.
- ✦ Complementarity with LHC jet data highlighted.

A. Ablat et el., arXiv:2307.11153





★ First simultaneous α_S + global PDF extraction at aN3LO.



MSHT, arXiv:2404.02964

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

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

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

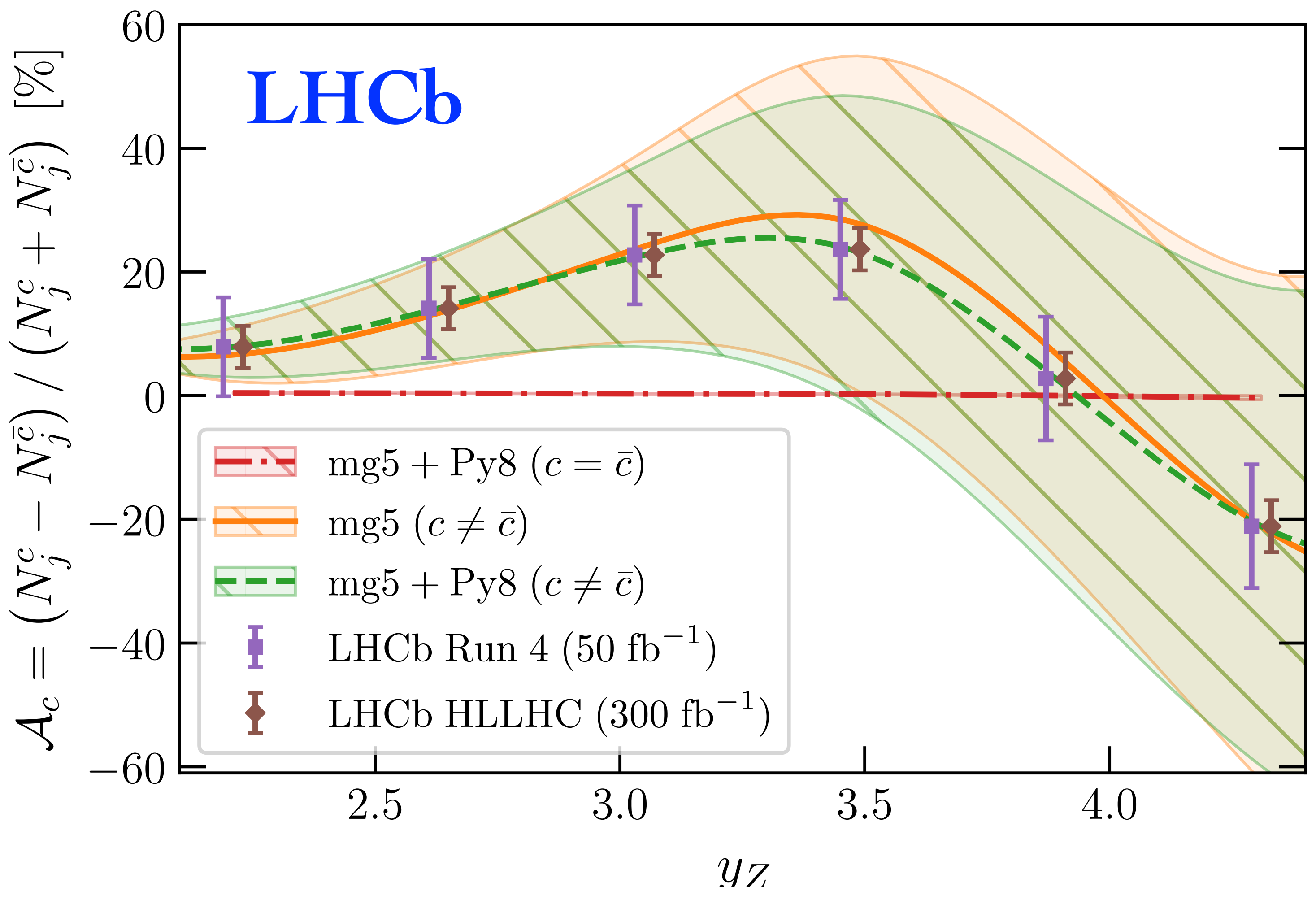
★ Errors slightly larger (more accurate) due to MHO uncertainty.

✦ 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.

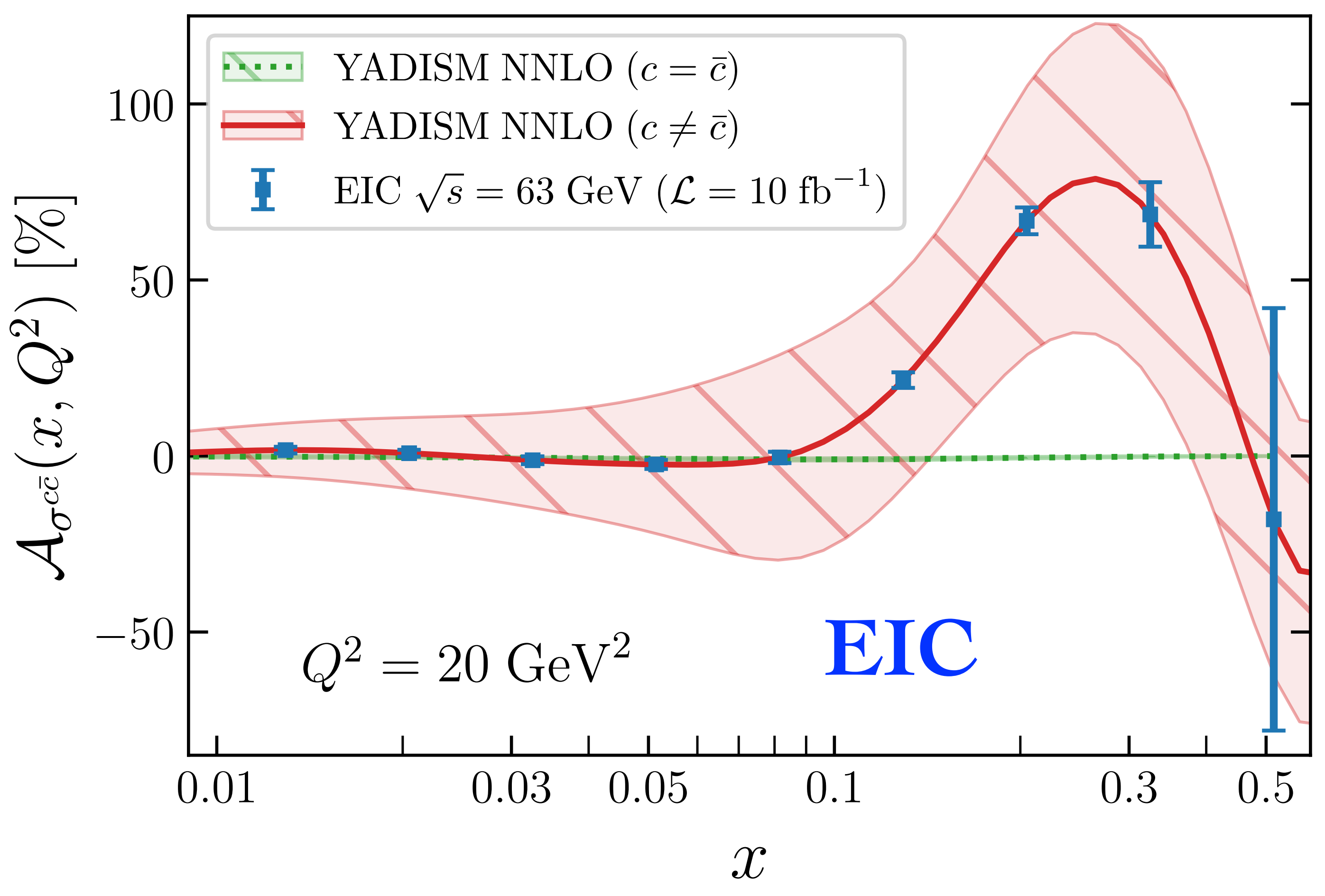
See talk by J. Rojo

R. Ball et al., arXiv:2311.00743



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

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

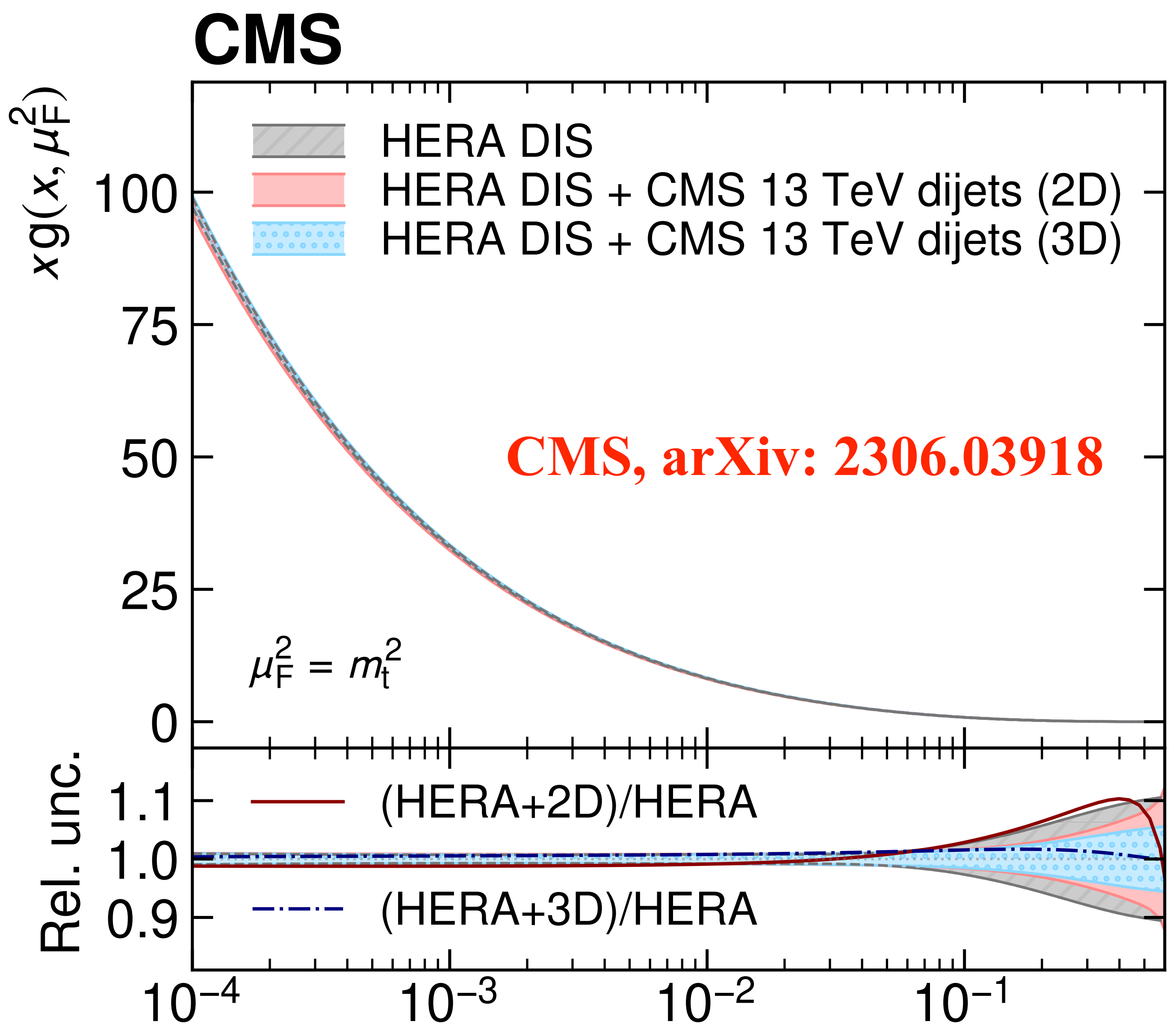


★ Not just global fitters - experimental collaborations also busily assessing impact of their data on PDFs.

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

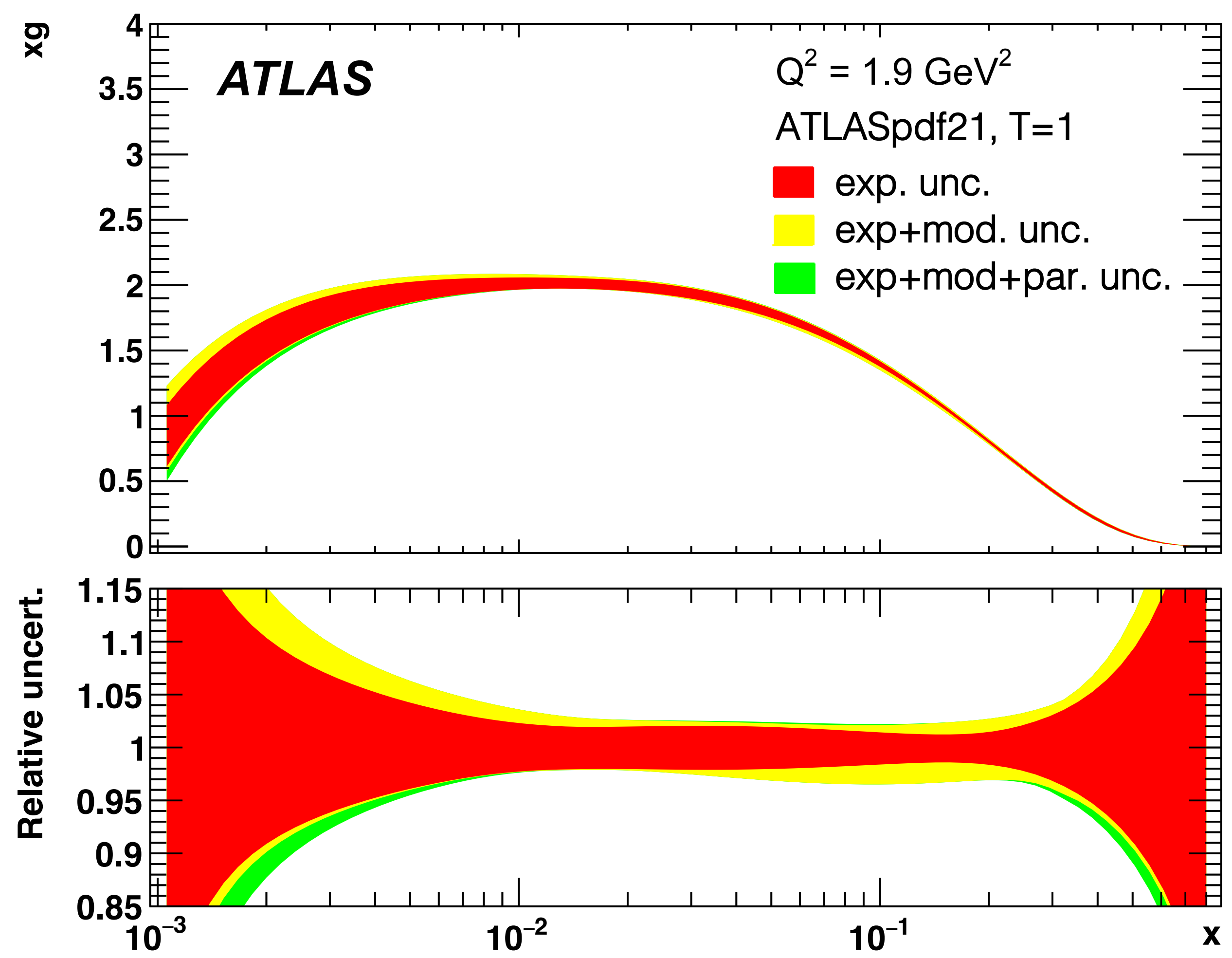
★ Not just LHC data. Seaquest example of non-LHC dataset with important impact (high x flavour decomposition).

★ Though some tension with other (NuSea) data!

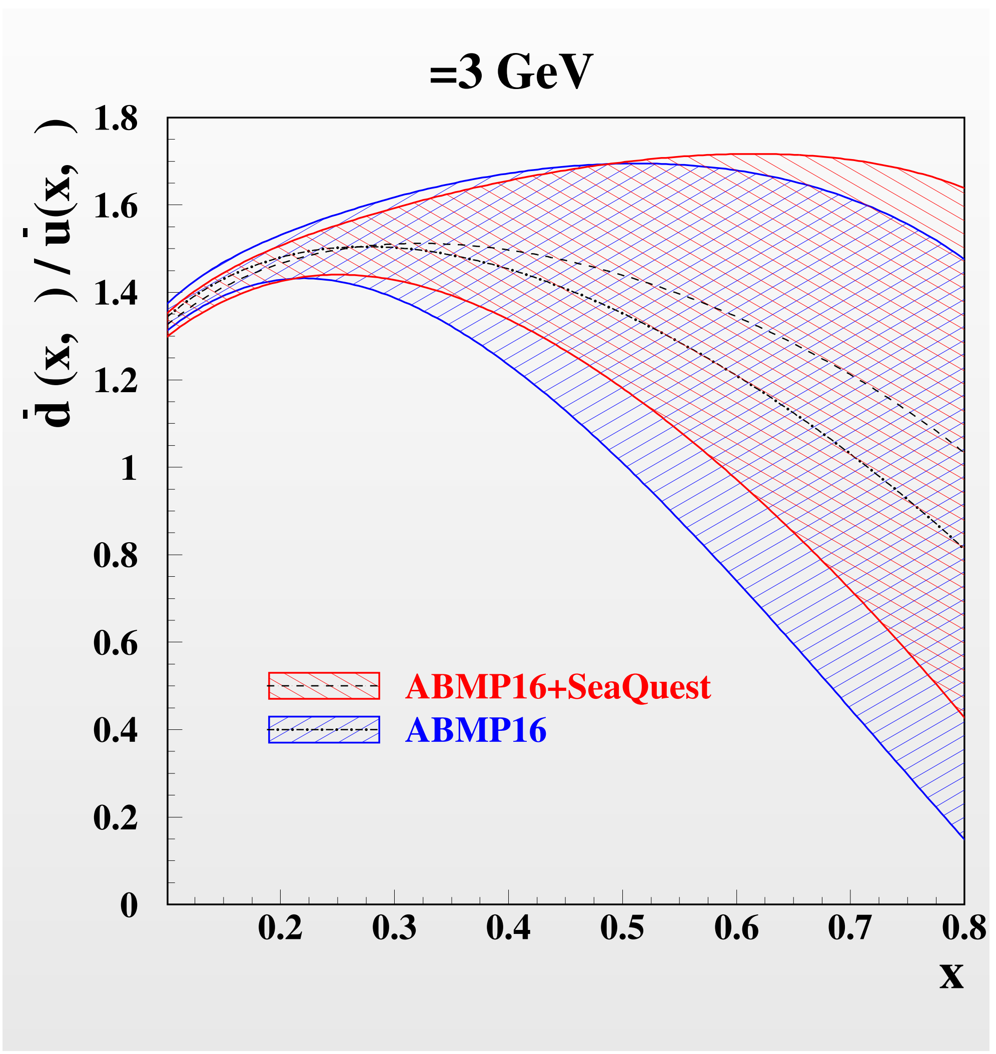


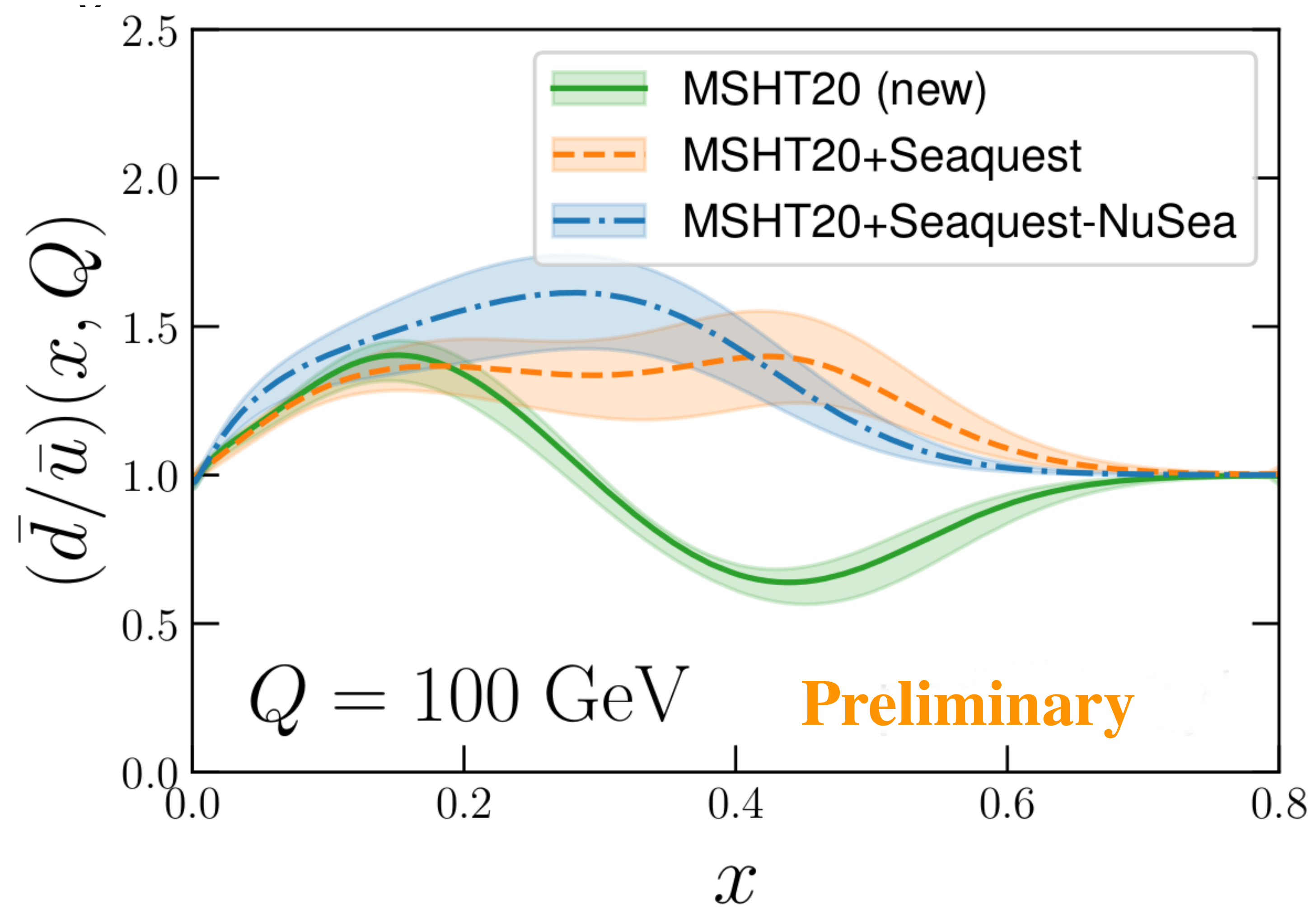
CMS, arXiv: 2306.03918

ATLAS, arXiv:2306.03918



ATLASpdf21

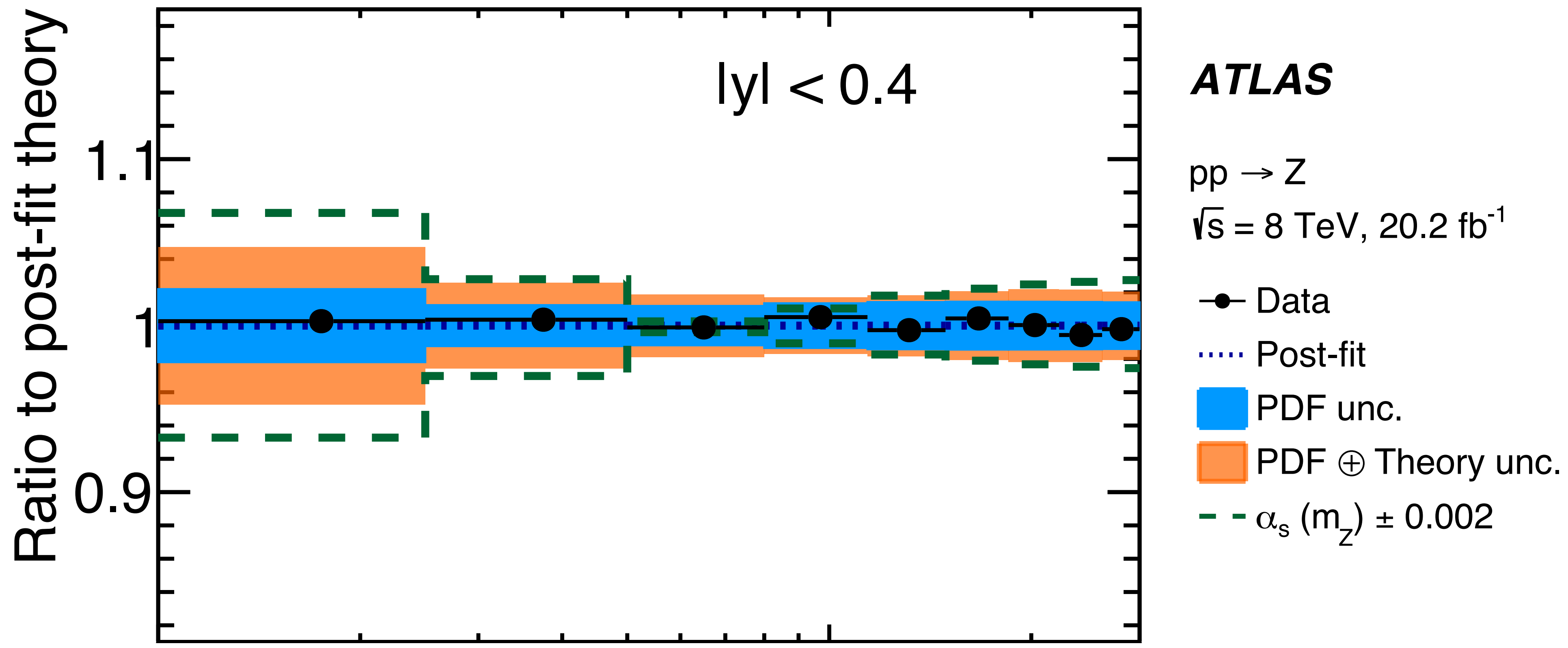




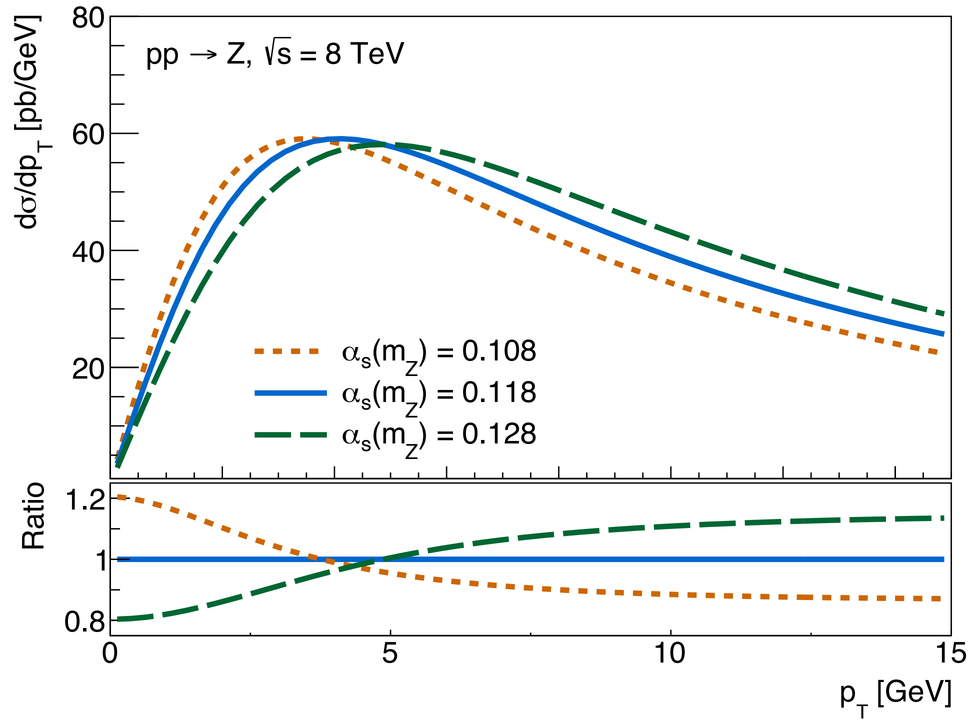
M. V. Garzeili, PDF4LHC23 meeting

LHC data - Some Examples

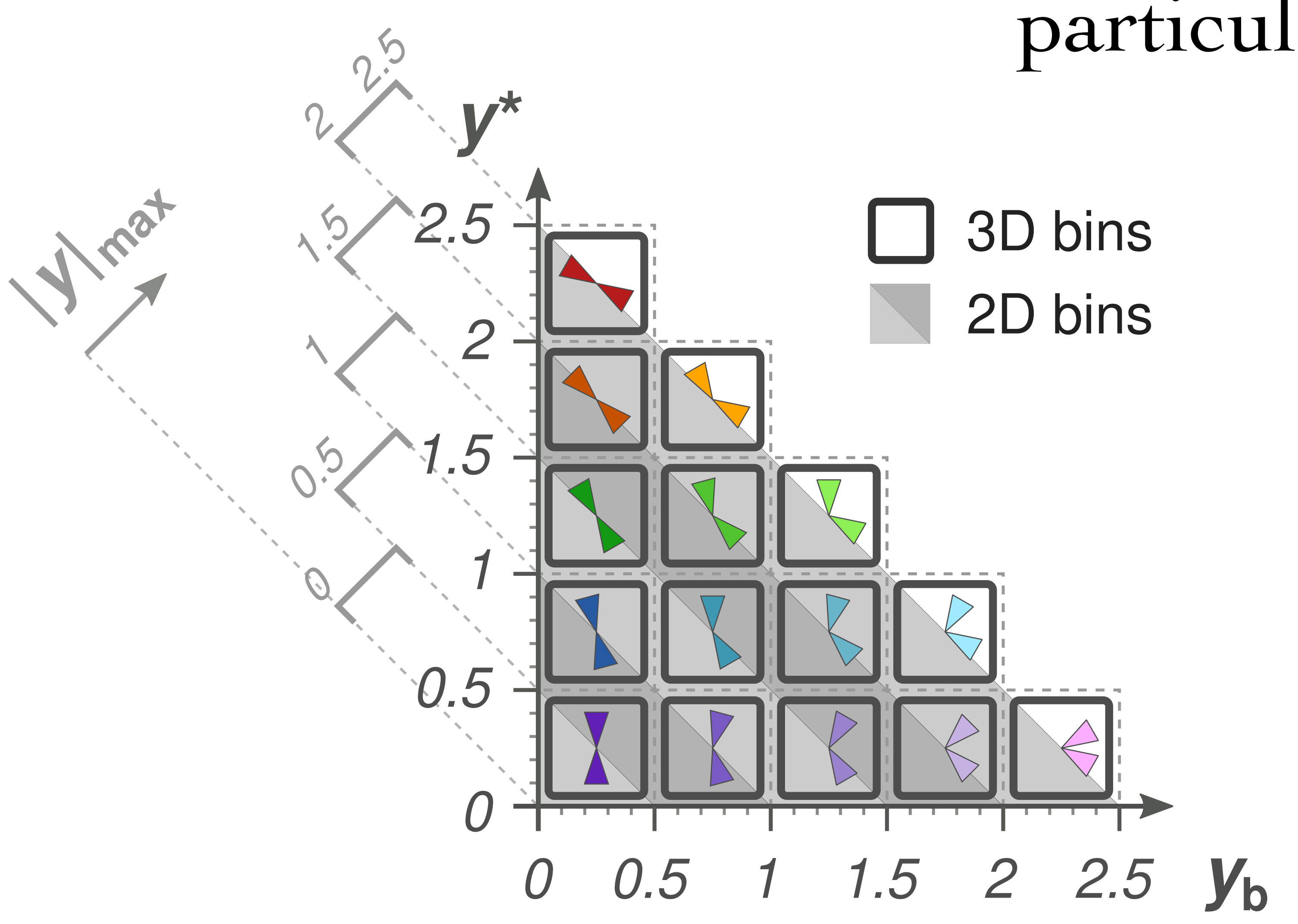
ATLAS, arXiv:2309.12986

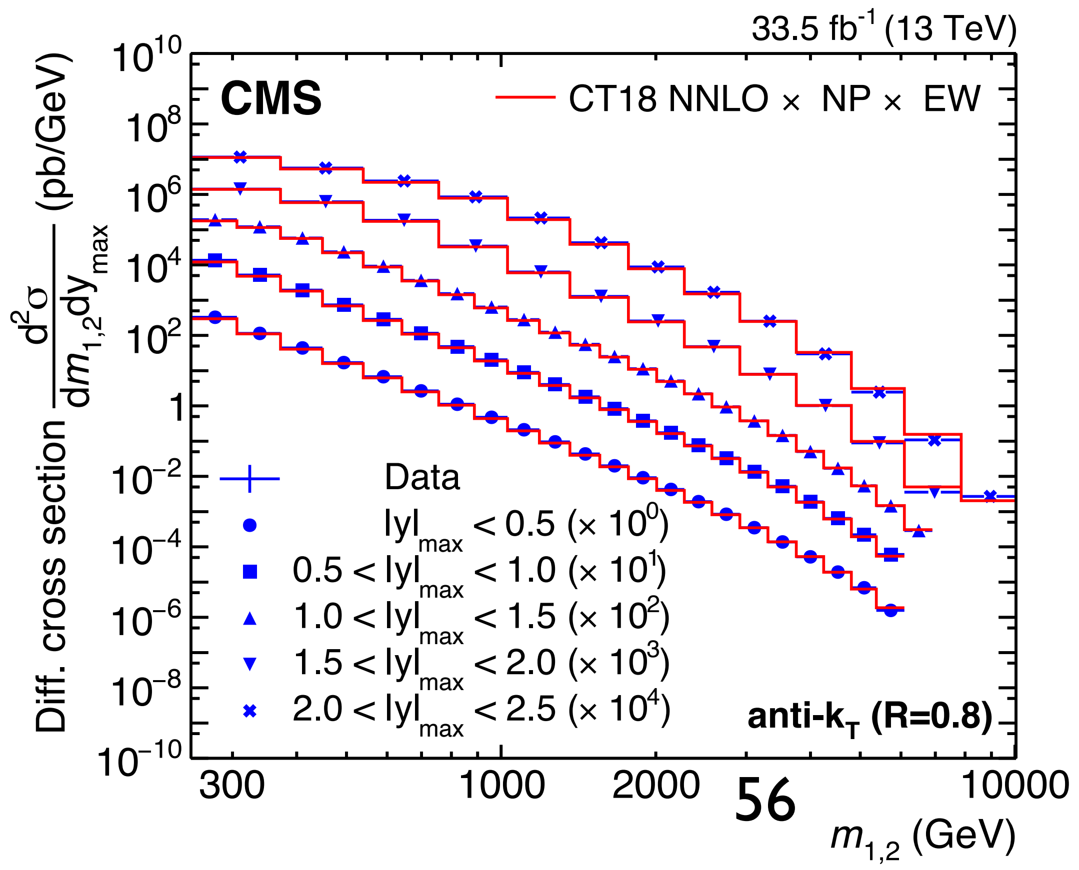


- ATLAS first extraction of strong coupling from Sudakov peak of $Z\, p_\perp$ distribution, at up to aN3LO.
- Also displays PDF sensitivity to e.g. gluon at intermediate x .



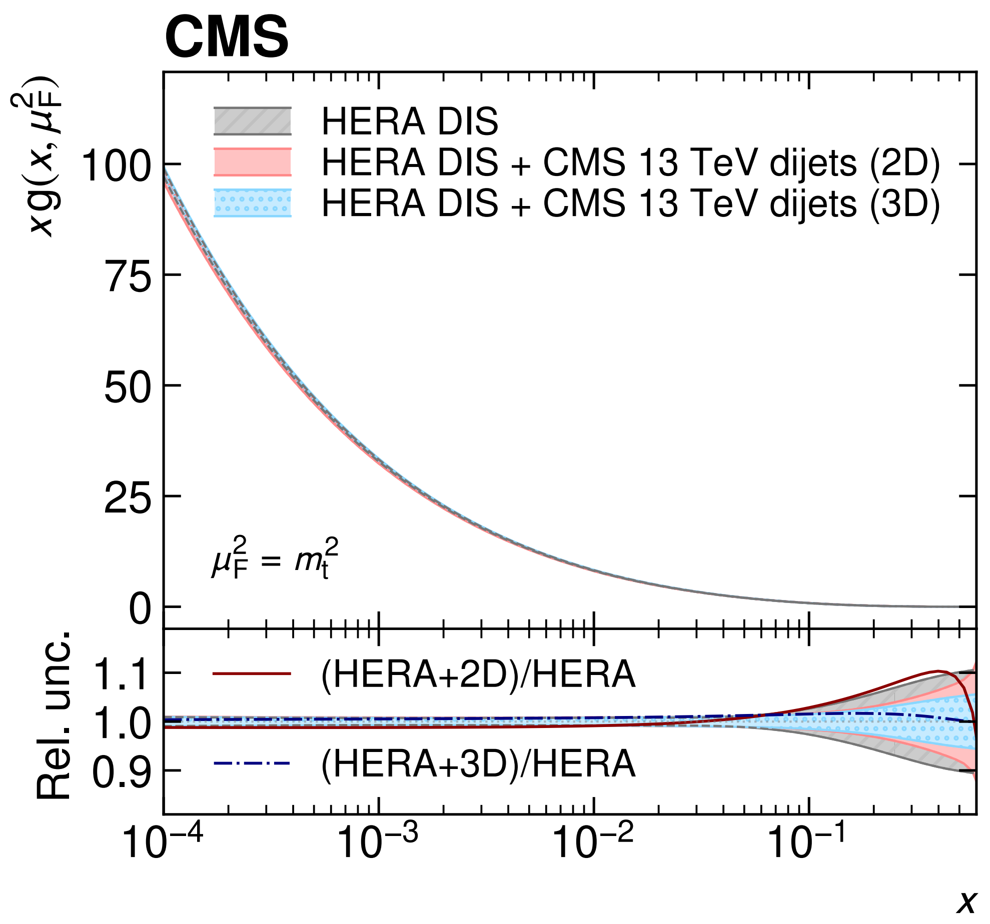
- Extensive range of jet measurements: triple differential dijet production particularly promising. More differential = more constraining!



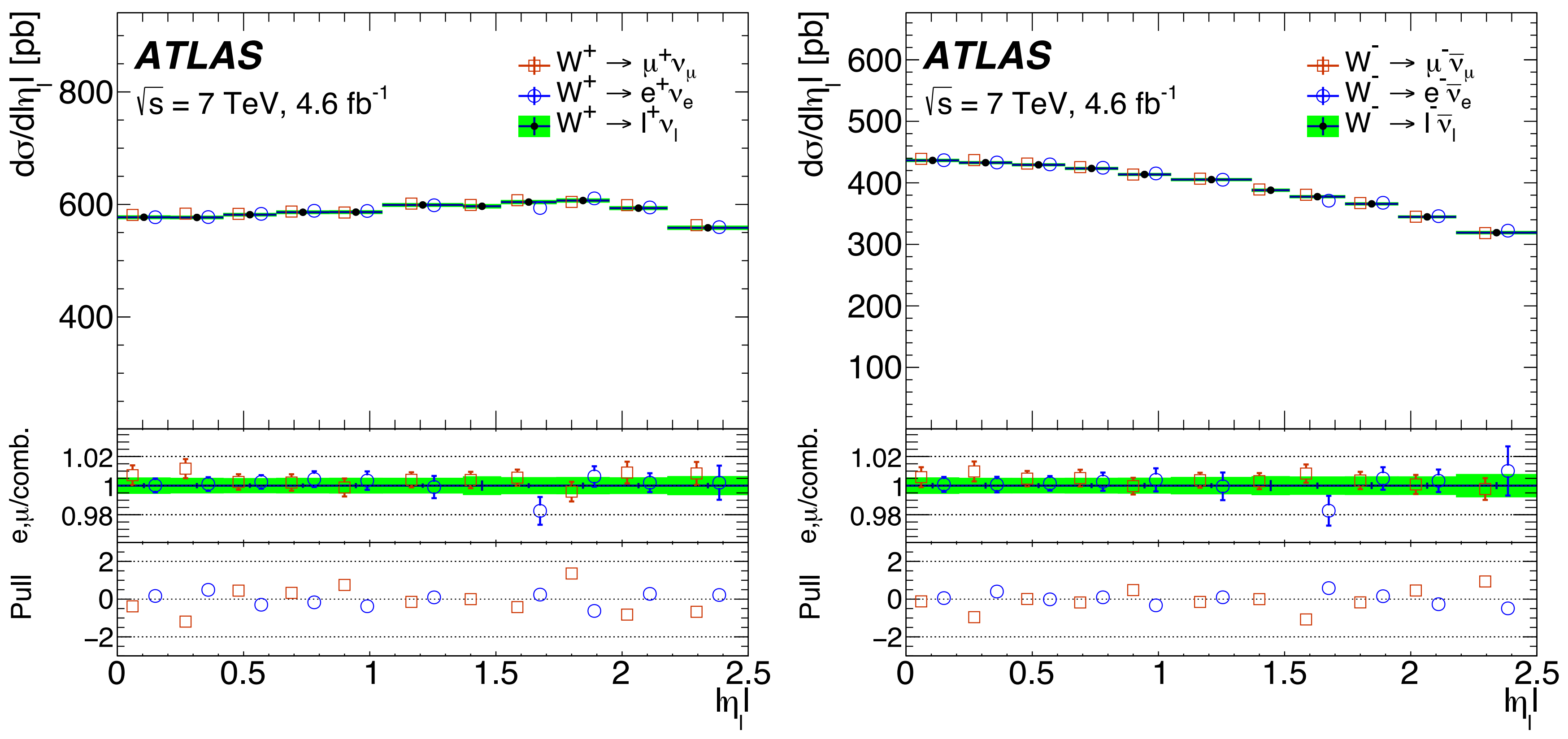


- Sensitivity to e.g. gluon at high x .

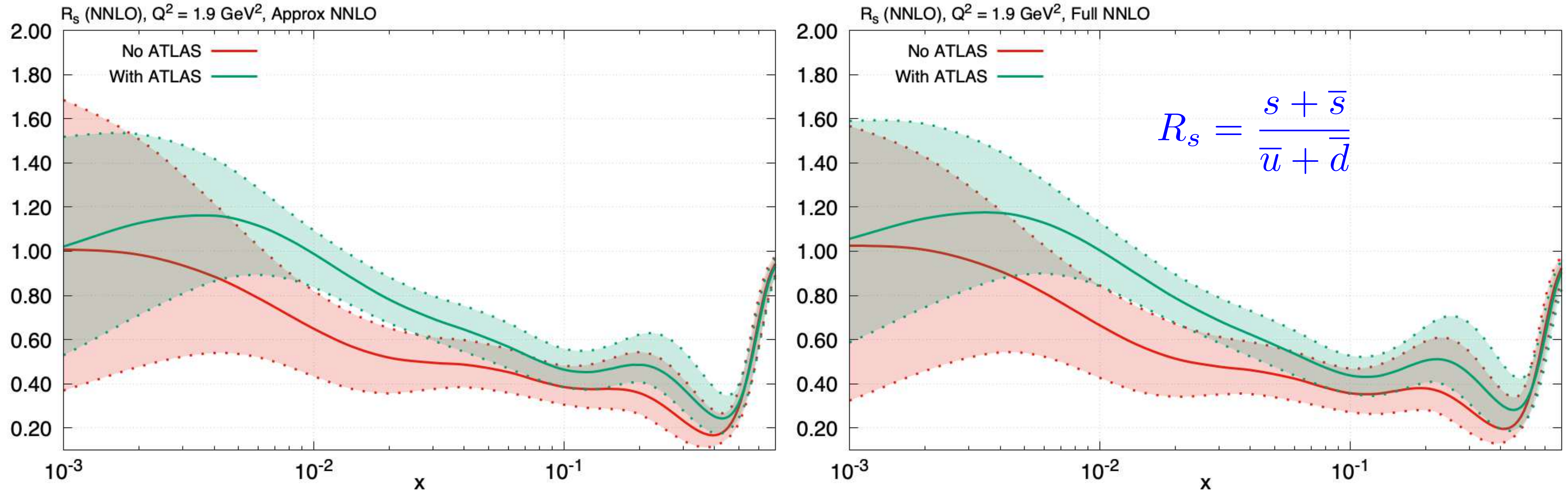
CMS, arXiv:2312.16669



- 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



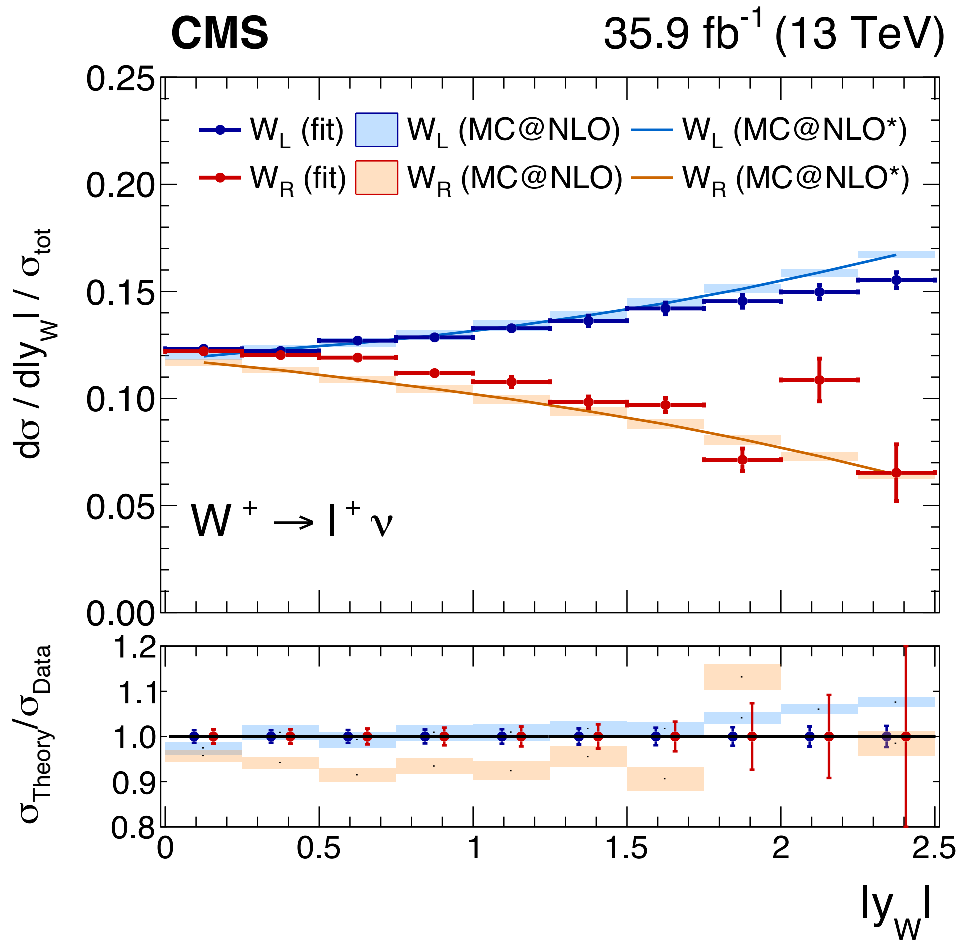
$$R_s = \frac{s+\bar{s}}{\bar{u}+\bar{d}}$$

- Many more measurements since then….

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*

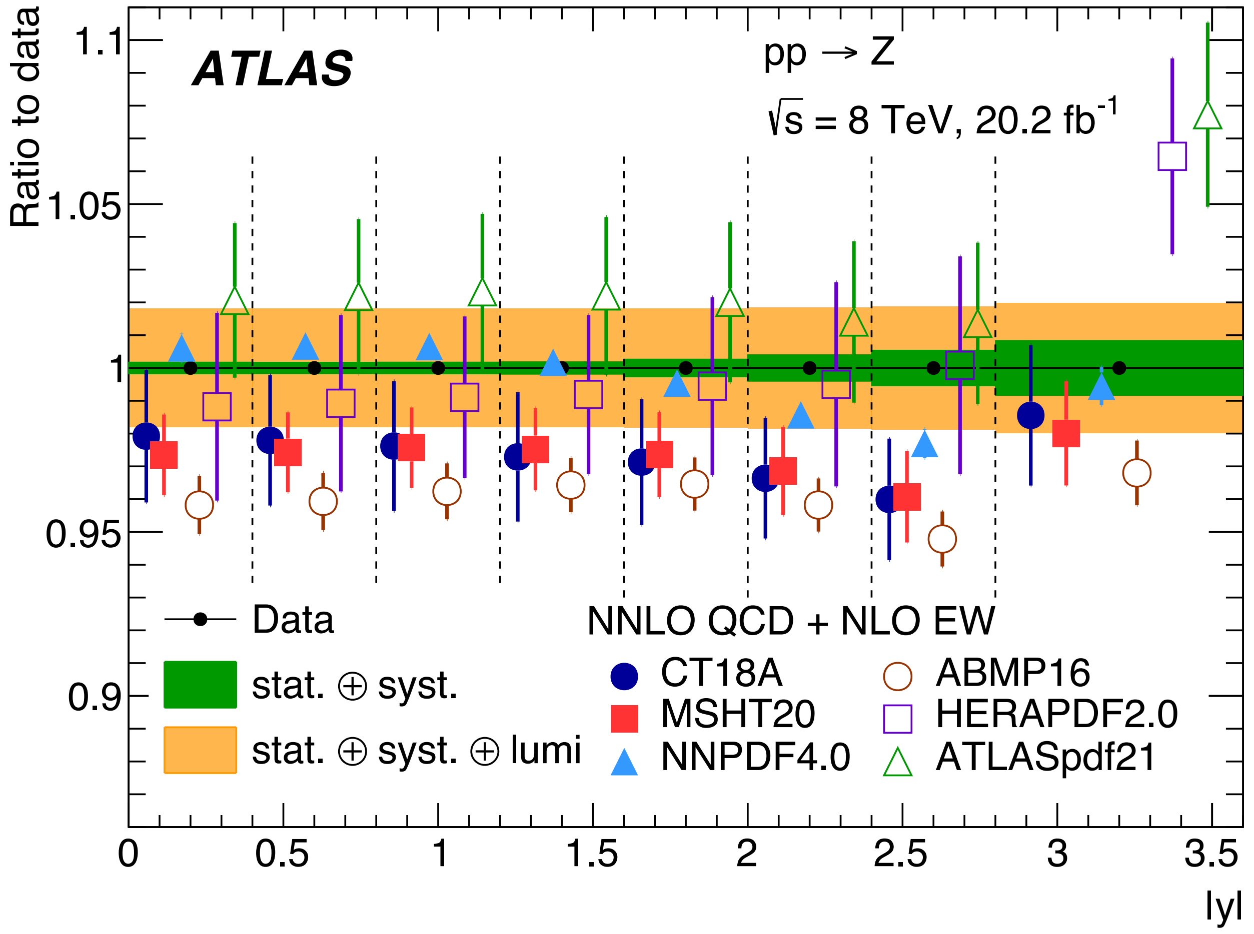
CMS, *Phys.Rev.D* 102 (2020) 9, 092012



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

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

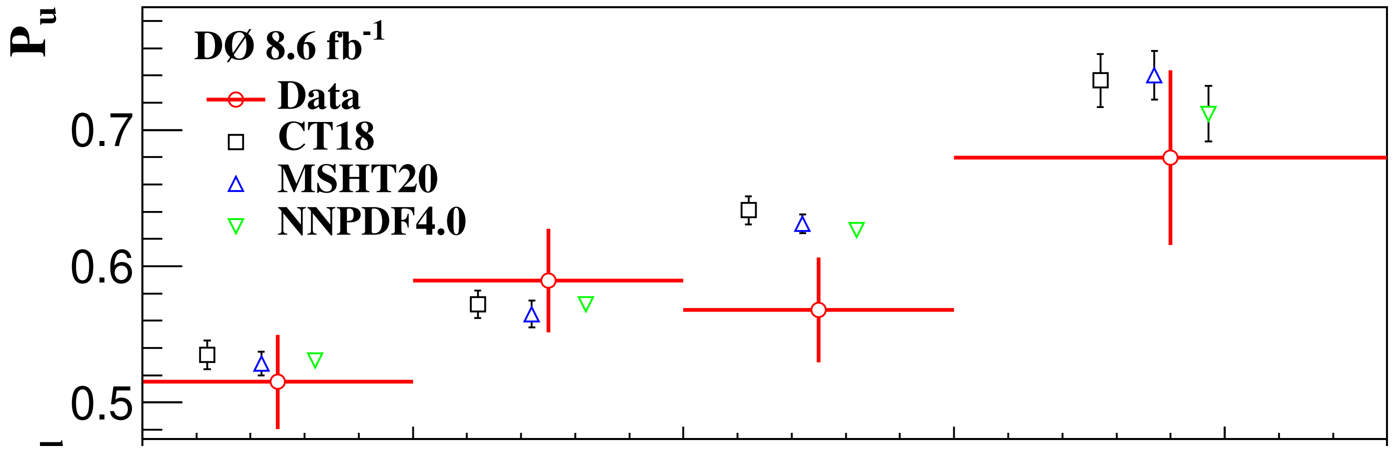
★ **New observables:** l^+l^-corrected to full phase space. Angular coefficients - limit extrapolation uncertainty.

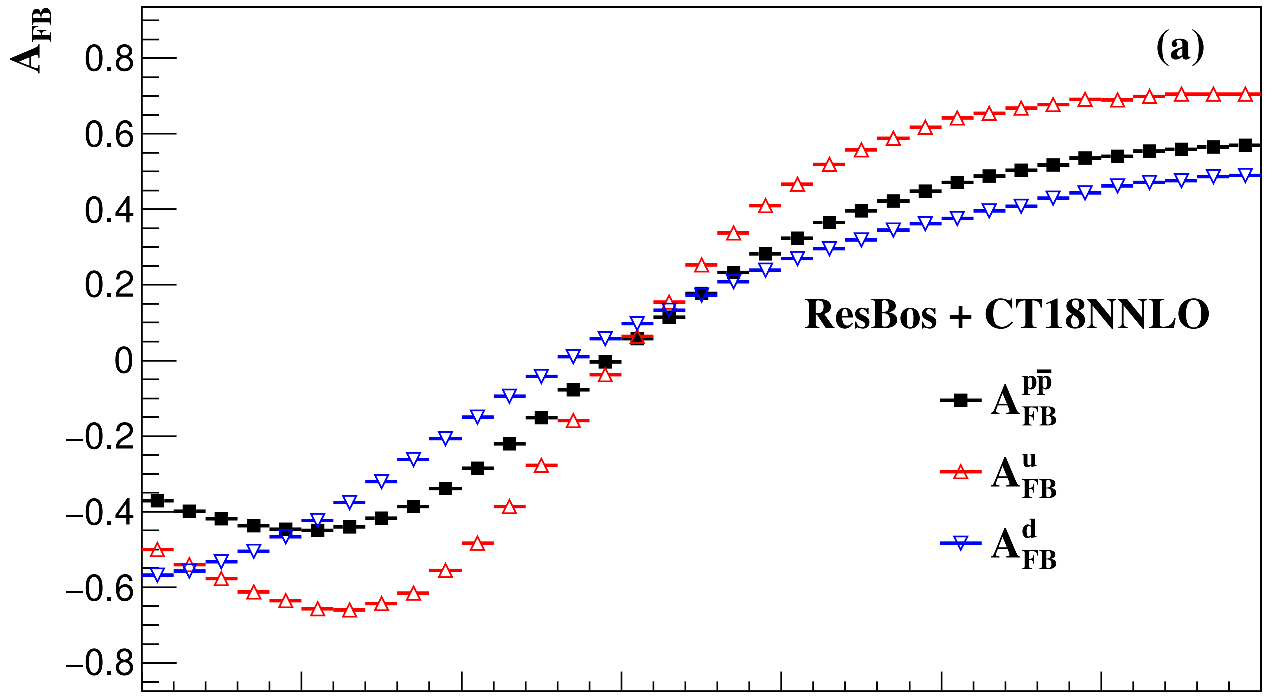


PDF set	Total χ^2 / d.o.f.	χ^2 p-value	Pull on luminosity
MSHT20aN^{3}LO [58]	13/8	0.11	1.2 ± 0.6
CT18A [59]	12/8	0.17	0.9 ± 0.7
MSHT20 [60]	10/8	0.26	0.9 ± 0.6
NNPDF4.0 [61]	30/8	0.0002	0.0 ± 0.2
ABMP16 [62, 63]	30/8	0.0002	1.8 ± 0.4
HERAPDF2.0 [64]	22/8	0.005	−1.3 ± 0.8
ATLASpdf21 [65]	20/8	0.01	−1.1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

ATLAS, arXiv:2309.09318

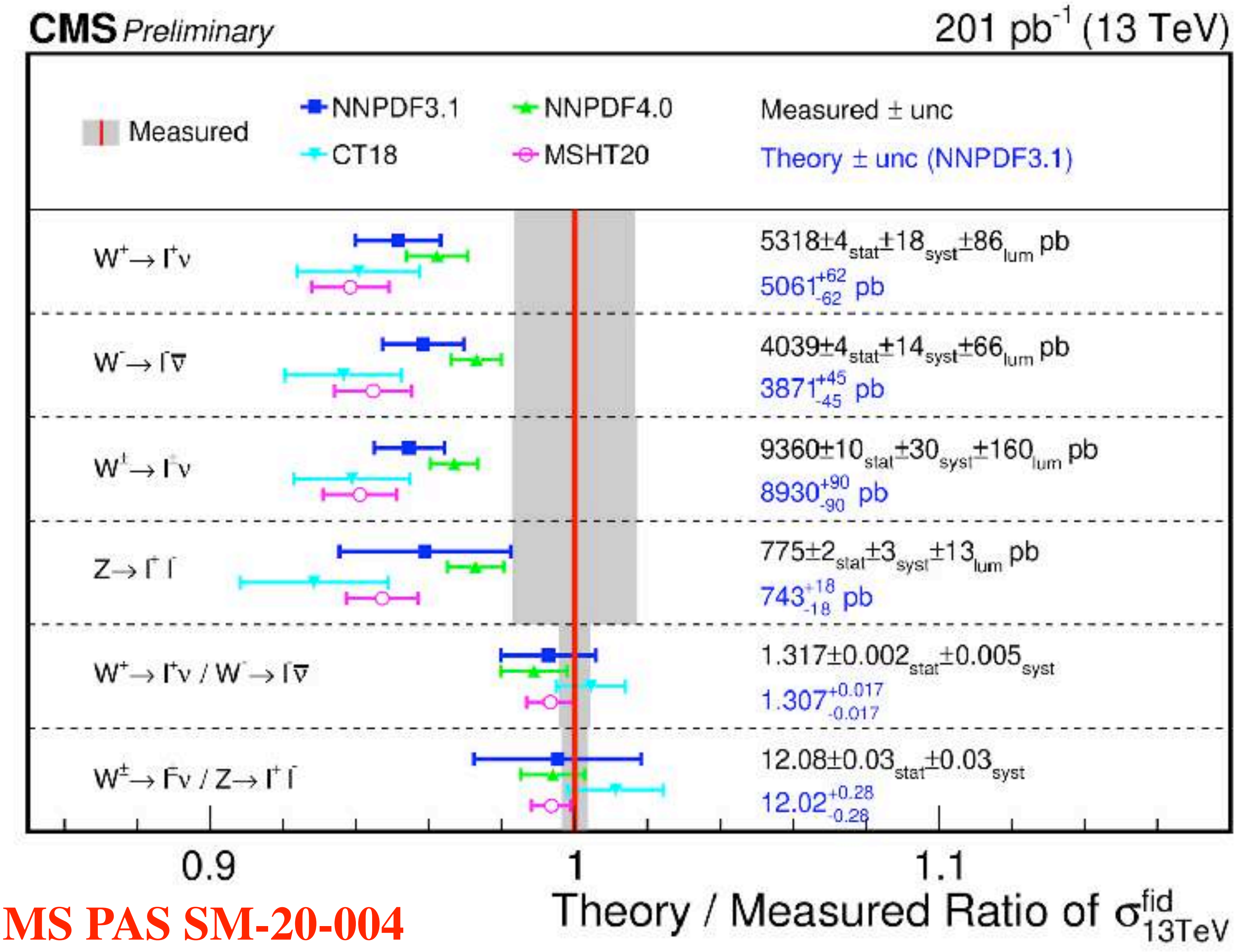
★ **New data:** Not just the LHC. Recent **D0** measurement of dilepton AFB. Sensitivity to high x flavour structure.





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

- And much more not shown here.

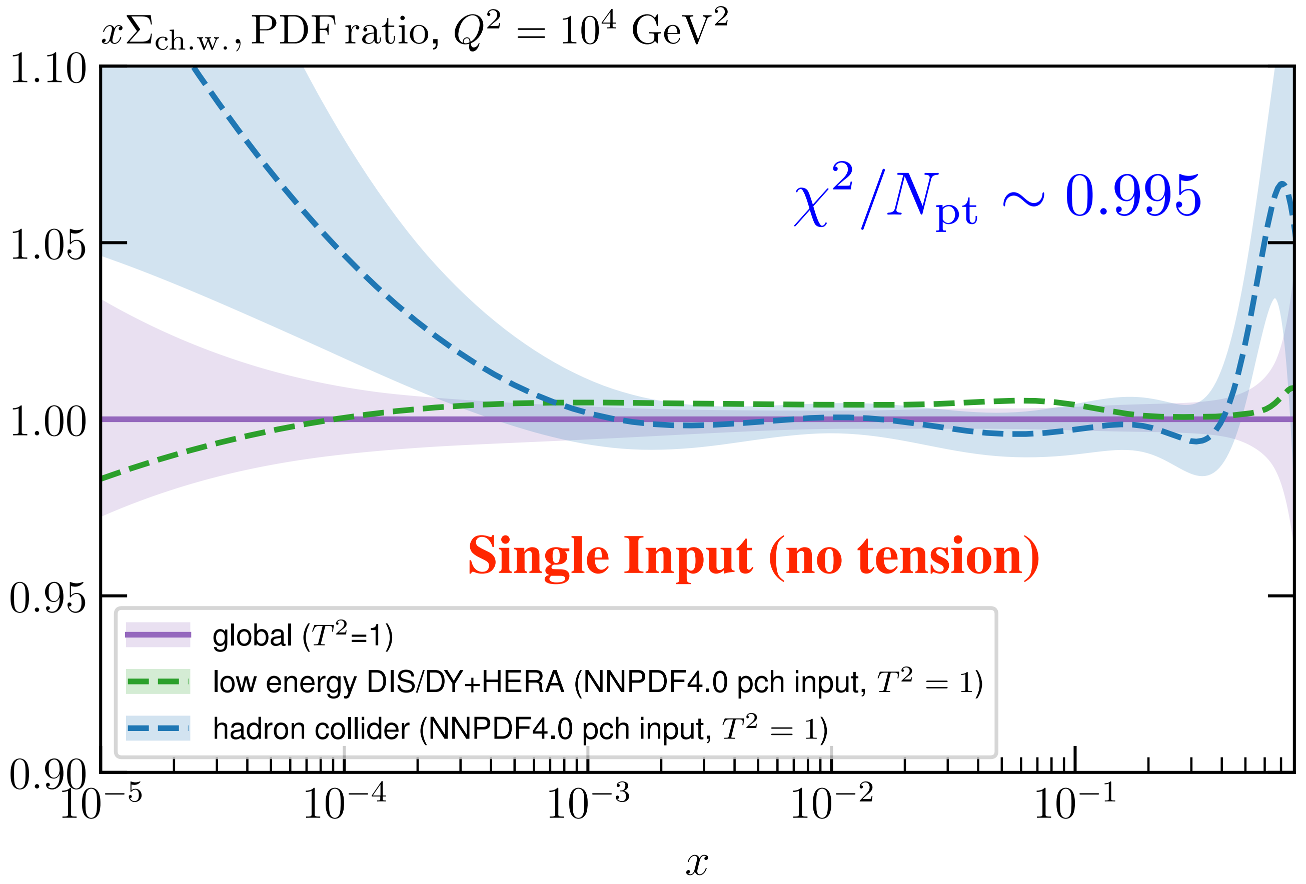


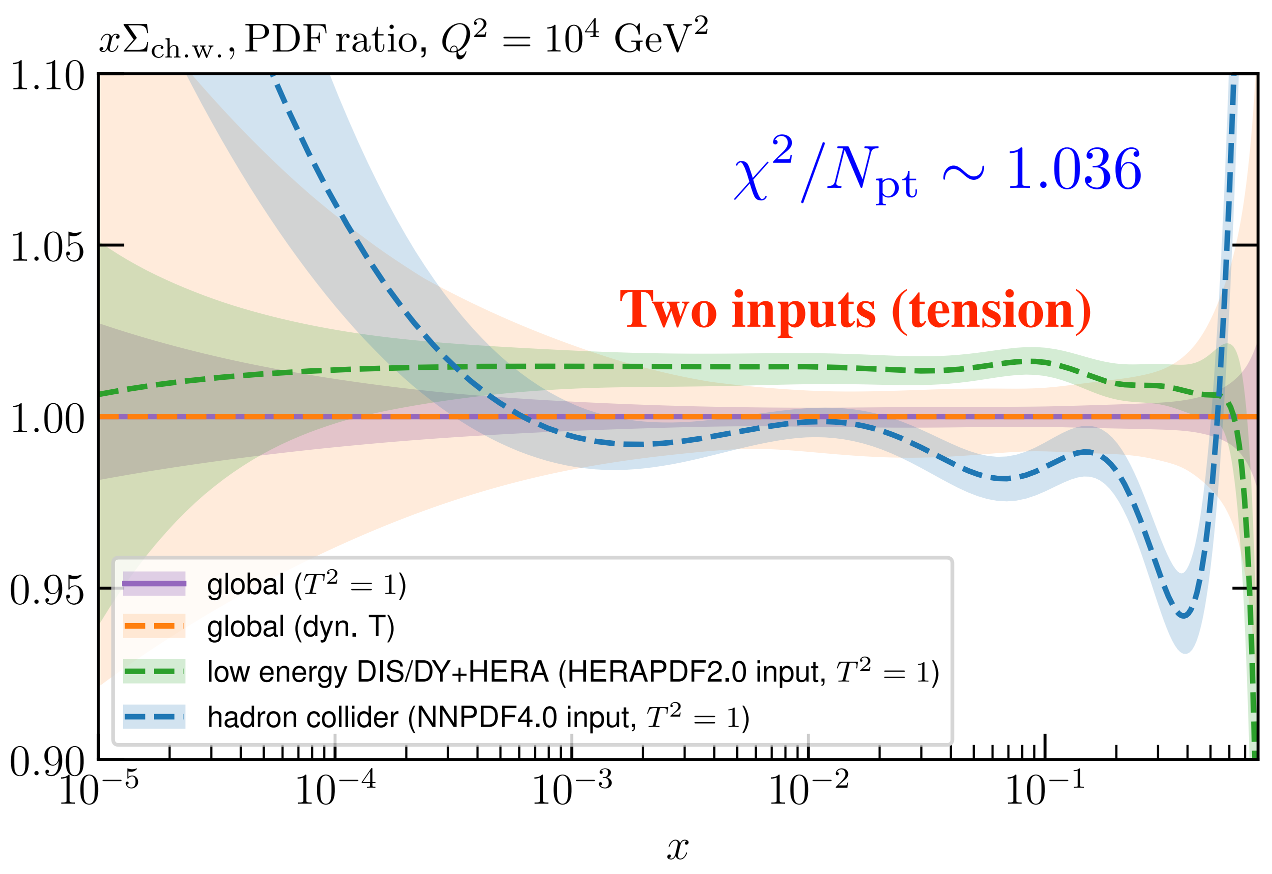
CMS PAS SM-20-004

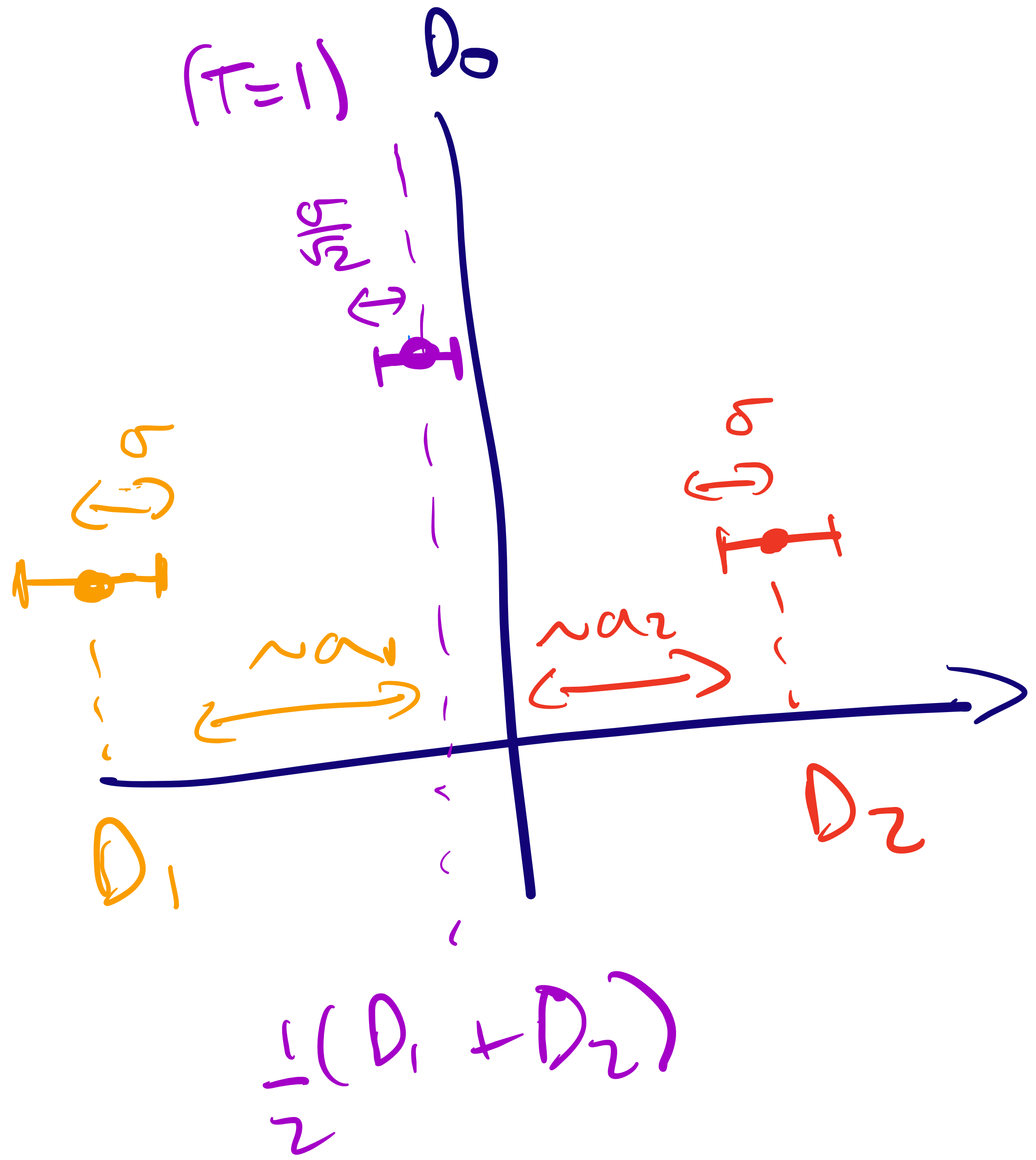
★ Inject **tension** into closure test - break (pseudo)dataset up into two.

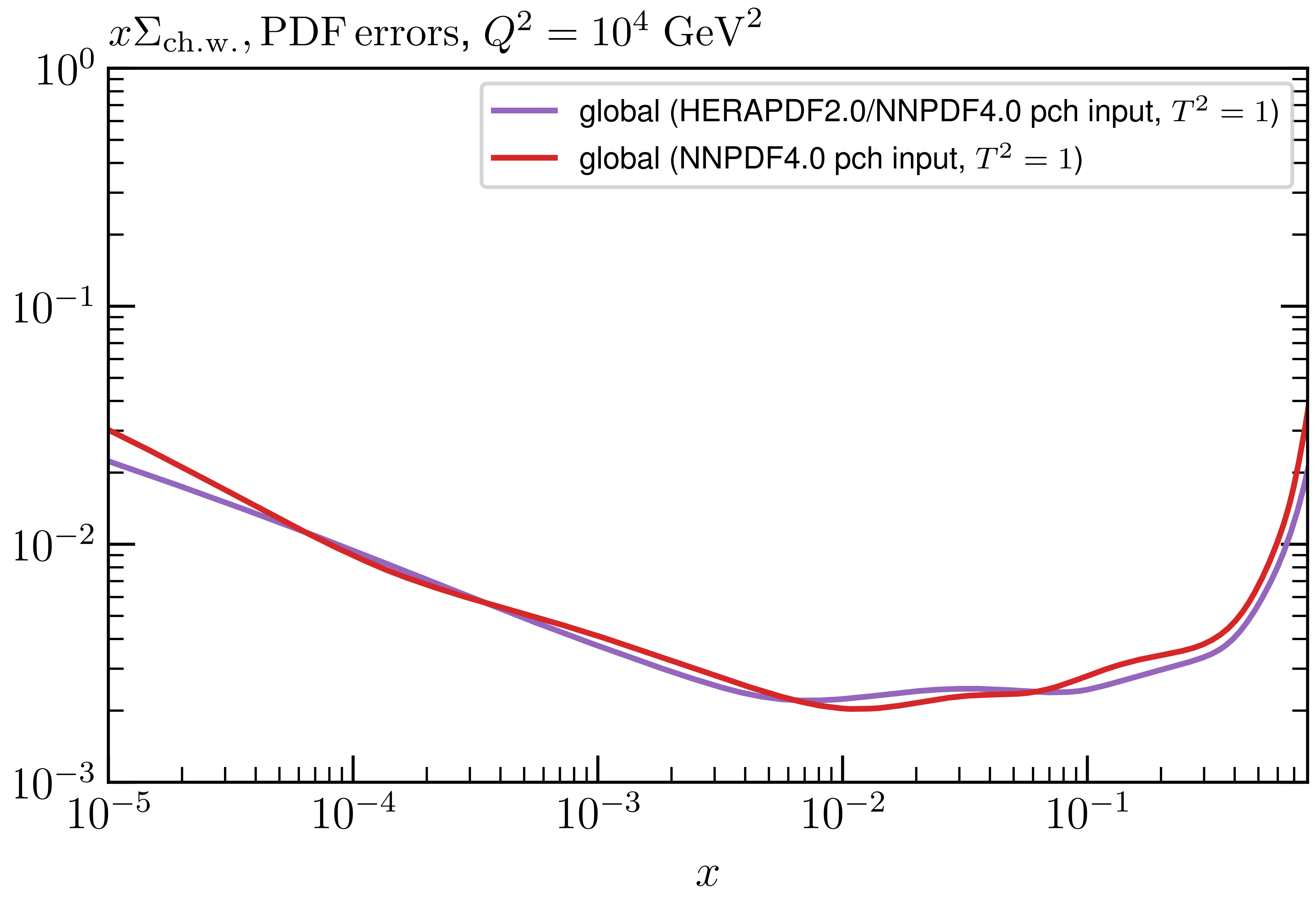
- Key point: $\Delta\chi^2 = 1$ uncertainty in global fit ~ same between consistent and inconsistent fits.

- PDF uncertainties driven by the quoted experimental (theoretical) uncertainties whether underlying fit is self—consistent or not.







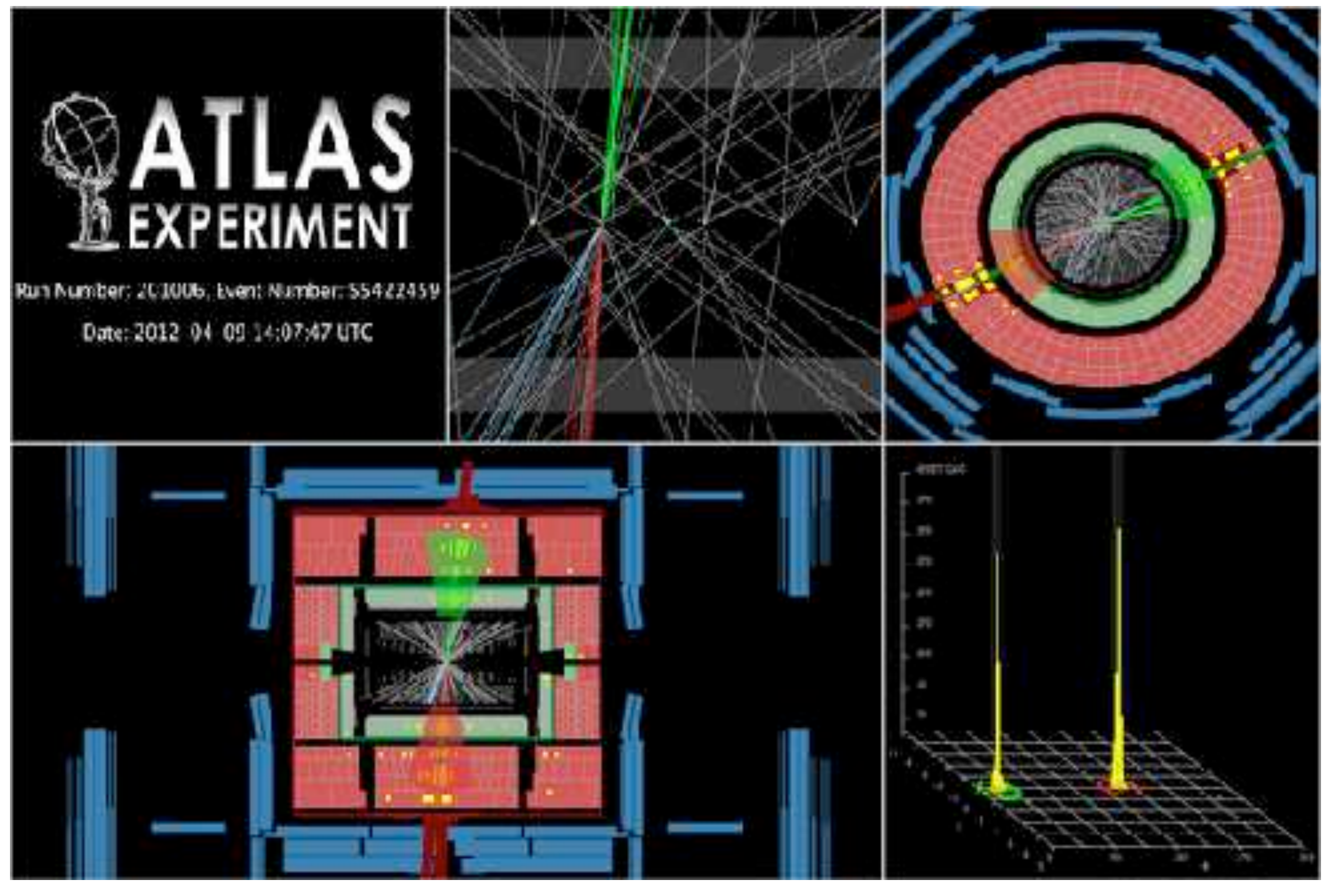


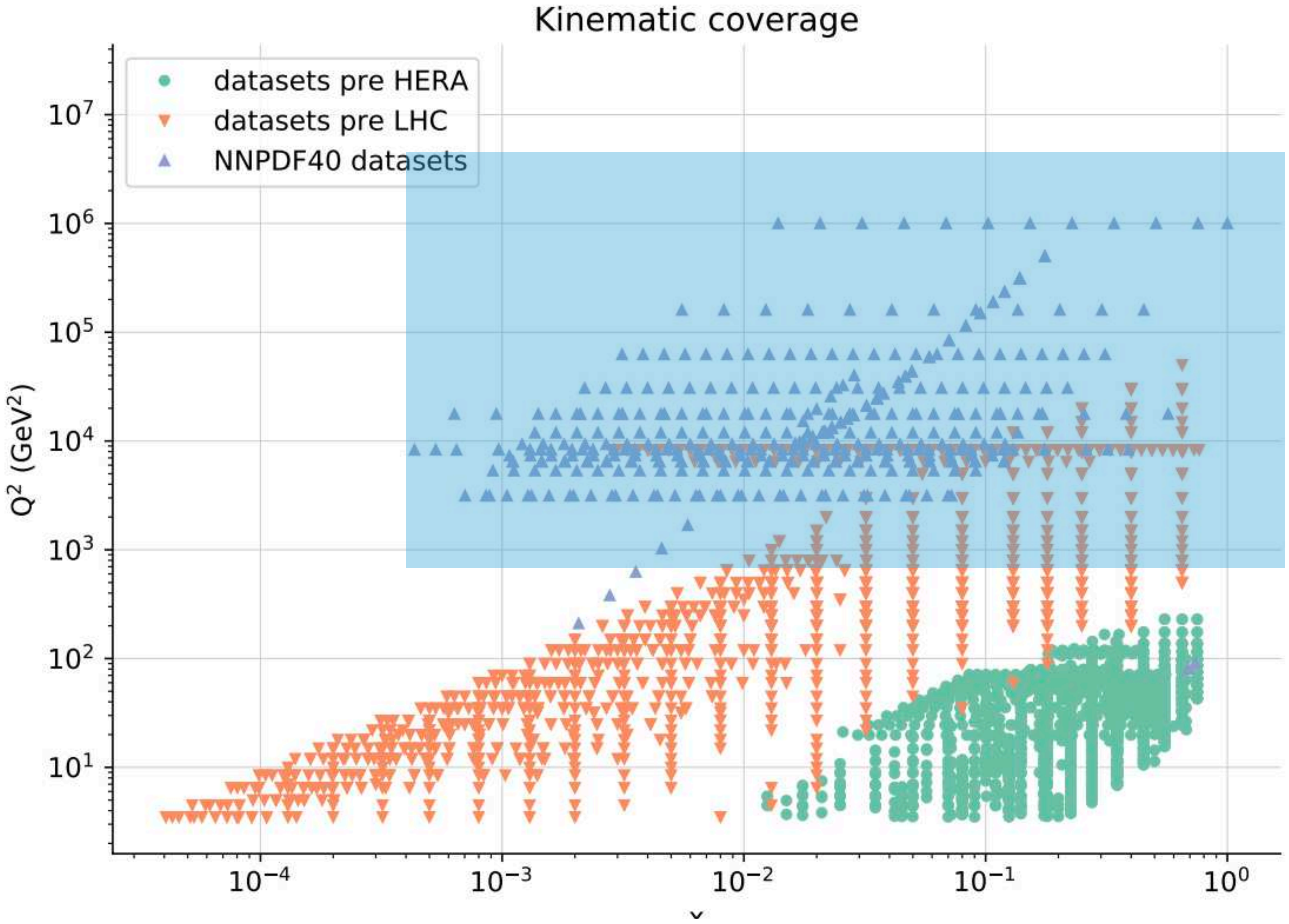
LHC data

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

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

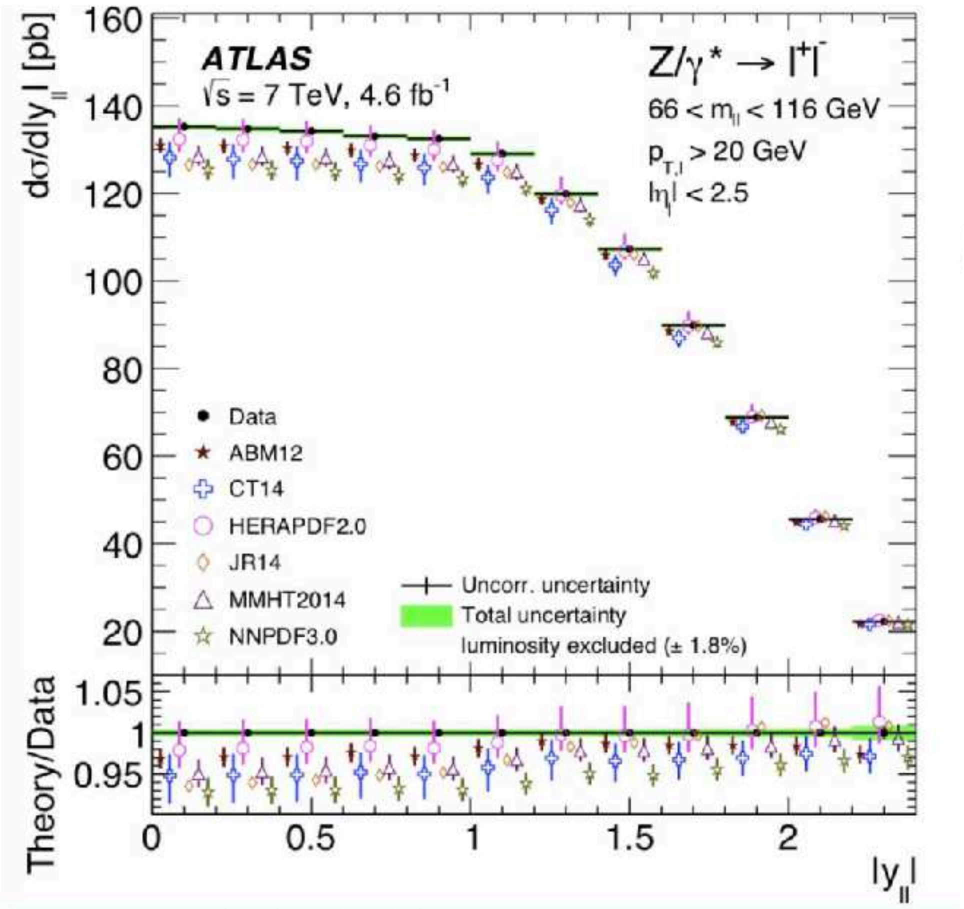
- Two key categories:
  - ★ **High energy** data - probing the high x region. Jets, top quarks, vector boson $p_\perp$…





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

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



1612.03016