

MSHT Updates

Lucian Harland-Lang, University College London

QCD@LHC26, Oxford, 28 September 2026

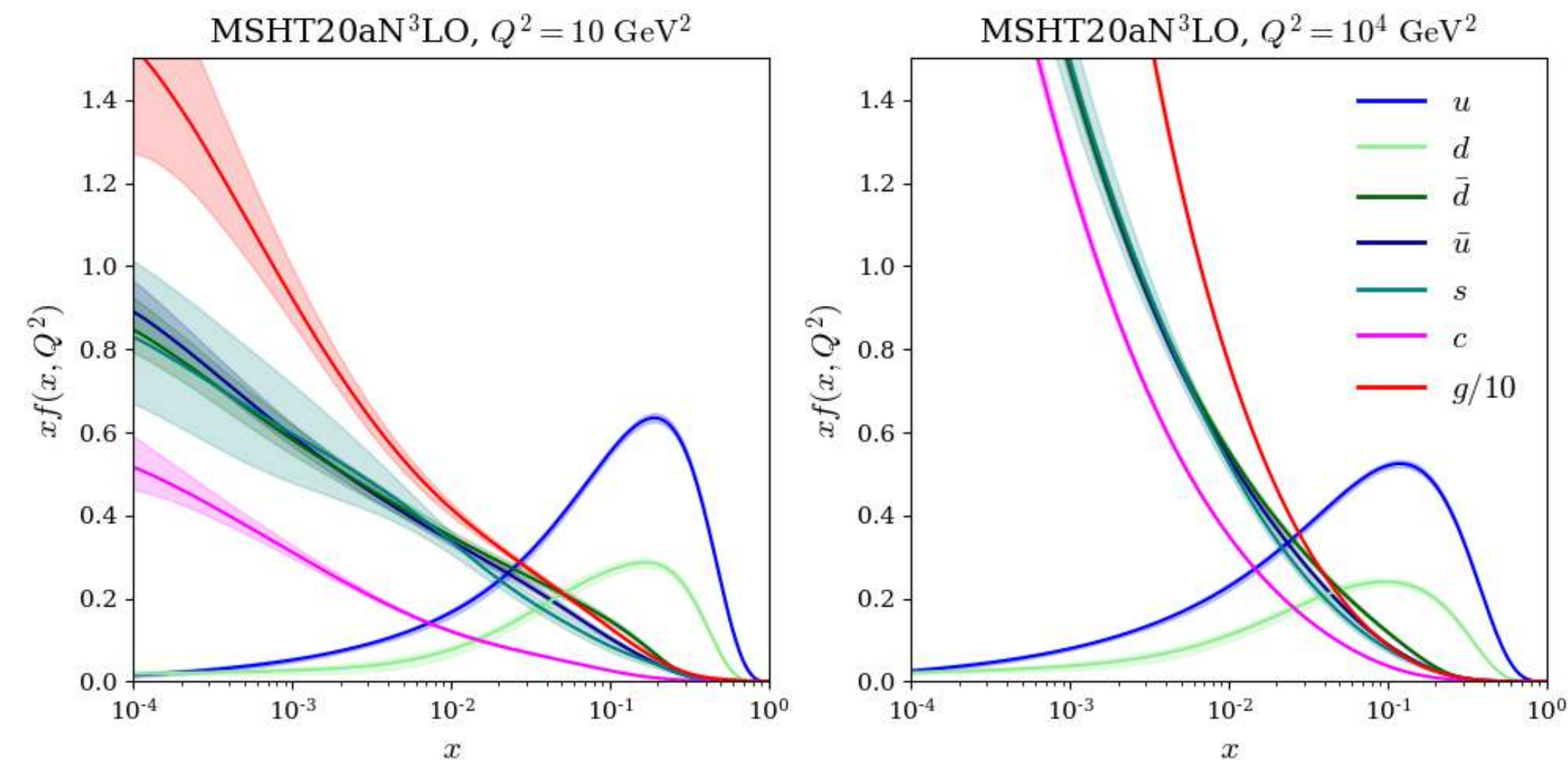
*In collaboration with MSHT colleagues - Robert Thorne, Tom Cridge and
Matt Reader*

...and thanks to T. Cridge for stolen slides!



MSHT News

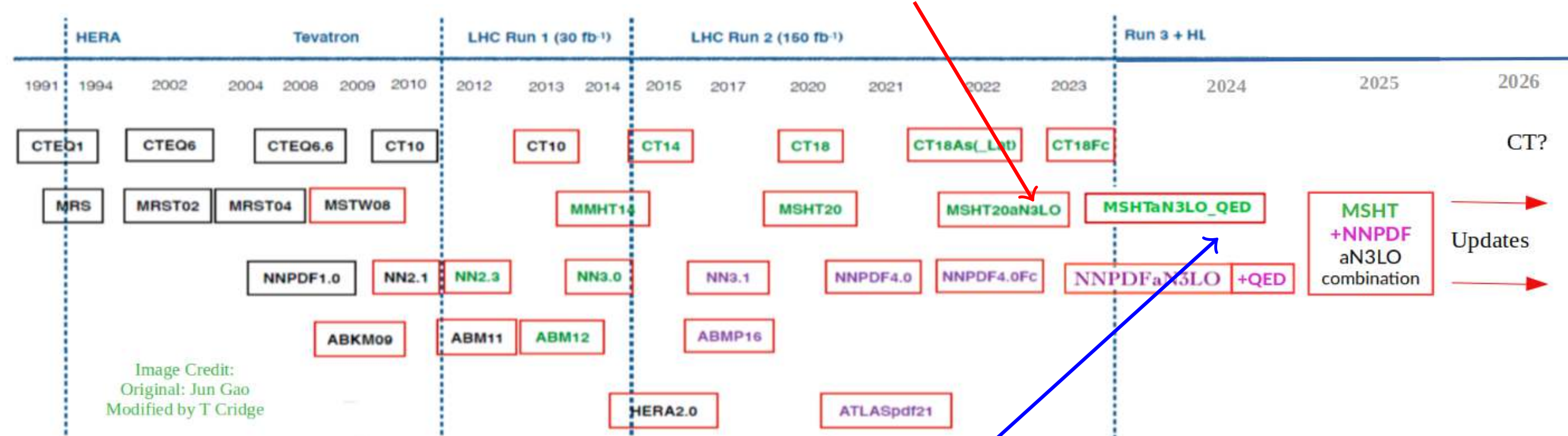
- ★ N3LO updates.
- ★ Work towards MSHT2n: impact of 13 TeV jet/dijet data.
- ★ Errors on errors.
- ★ FPDF!
- ★ Study on high x region: impact of **target mass corrections, higher twist...**



N³LO updates

Approximate N3LO PDFs Available

- World first aN3LO PDF set $\Rightarrow$ MSHT20aN3LO available for ~ 4 years.



- Also available for $\gtrsim 2.5$ years: MSHT20an3lo+QED
- First time at approximate N3LO (“aN3LO”) and first inclusion of theory uncertainties in PDFs at highest order.

Basic Idea:

- Include all known N3LO terms into the PDFs (a lot!)
- Include theoretical uncertainty for the missing ingredients
 $\Rightarrow$ aN3LO PDFs + theory uncertainty.

What is now known for N3LO PDFs?

- Need to know:

▶ **Splitting functions** - at 4-loop to evolve PDFs in (x, Q^2) : Mellin moments, small x , high x limits
New moments for $P_{qg, gq, gg}, P_{qq}^{PS}$ (FHRMUVV) ← $P(x, \alpha_s) = \alpha_s P^{(0)}(x) + \alpha_s^2 P^{(1)}(x) + \alpha_s^3 P^{(2)}(x) + \alpha_s^4 P^{(3)}(x) + \dots$

(Also now NS via 2604.09534) ▶ **Transition Matrix Elements (TMEs)** - at 3-loop to change number of PDF flavours at heavy quark mass (m_h) thresholds. Mellin moments, small x , high x limits

$A_{Hg}^{(3)}, A_{gg, H}$ since determined. (Blümlein et al) ← $f_{\alpha}^{n_f+1}(x, Q^2) = [A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2)](x)$

▶ **Coefficient Functions for DIS** - at 3-loop to determine structure functions. Light flavour known, heavy flavour high Q^2 known, approx for low Q^2 .

No further info. ← $F_2(x, Q^2) = \sum_{\alpha \in H, q, g; \beta \in q, H} (C_{\beta, \alpha}^{VF, n_f+1} \otimes A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2))$

▶ **Hadronic cross-section k-factors** - at N3LO. Very little known, PDFs need differential with cuts.
N3LO DY y distribution, (Chen et al) ← $\sigma = \sigma_0 + \sigma_1 + \sigma_2 + \sigma_3 + \dots \equiv \sigma_{N3LO} + \dots$

References in backup

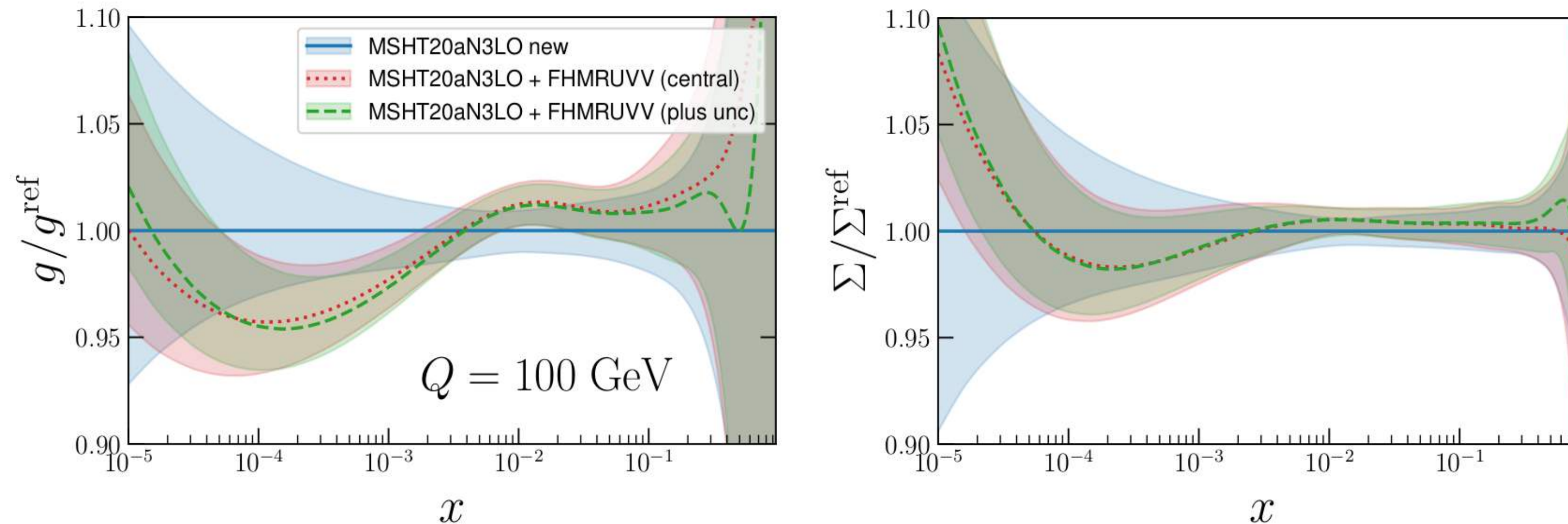
- Much already known, **only a few remaining missing pieces.**

MSHT - status, updates and future developments:

N3LO Ingredients		MSHT20aN3LO 2207.04739	Evolution Benchmarking 2406.16188	DIS2025 proceedings 2510.09321	DIS2026 updates	Future (Ongoing)
Splitting Functions	Singlet $qq_{PS}, qg,$ gq, gg	4 moments for all (own approximations)	10 moments: $qq_{PS}, qg,$ 5 moments: gq, gg	10 moments for all FHMURVV central	10 moments for all FHMURVV central + error	11 moments for all (FHMURVV 2512.10783). Further work on uncertainty in nuisance parameter approach
	Non-Singlet $qq_{NS+,-,S}$	8 moments (own qq_{NS+})	8 moments (Moch et al qq_{NS+})	8 moments (own qq_{NS+})	8 moments (own qq_{NS+})	Exact - see later (Gehrmann et al 2604.09534)
Transition Matrix Elements		$A_{Hq}^{PS}, A_{gq,H}$ known $A_{gg,H}, A_{Hg}, A_{qq,H}^{NS}$ approximated	-	All known (included via grids as in 2411.05373)	All known (included via grids as in 2411.05373)	Include Exact expressions (Ablinger et al 2510.02175) (consistency checks performed)
DIS Coefficient Functions	Light Mass	Known ZM-VFNS				
	Heavy Mass	FFNS known LL small- x , unknown NLL and x -dependence. Known mass threshold info $f(\ln(W^2/4M_H^2 - 1))$, unknown Q^2 interpolation to ZM-VFNS - interpolated based on lower orders				Improve estimate of uncertainty for $\mathcal{O}(\alpha_S^3)$ low Q^2 FFNS pieces

Approximate N3LO - Update Splitting Functions

- Update to FHMRUVV approximations based on 10 moments (cf 4 in original MSHT20aN3LO) for P_{qq}^{PS} , P_{qg} , P_{gq} , P_{gg} . Preliminary!
- First take central values (no uncertainty), now add uncertainty:



- 1% increase in gluon at $x \sim 10^{-2}$. Reduced uncertainty at low x .
- How to account for uncertainty? Apply TNP with penalty $\Delta\chi^2 = 1$:

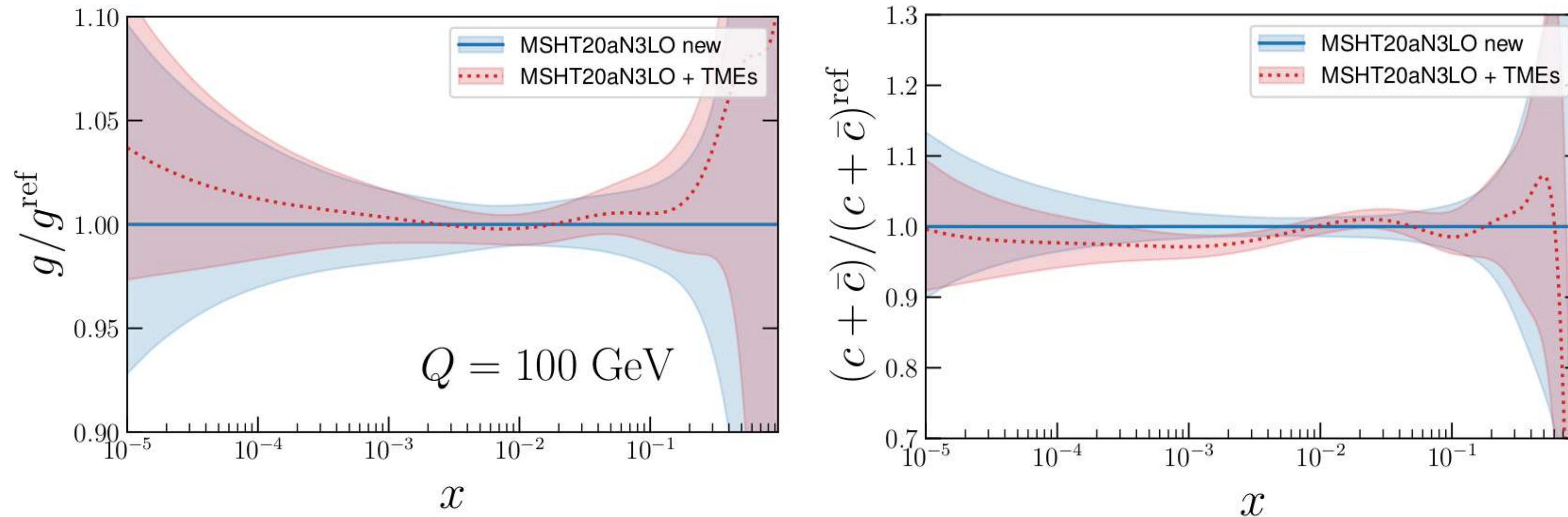
$$P_{ij} = P_{ij}^{\text{FHMRUVV}}[\text{central}] + \theta_{ij}(P_{ij}^{\text{FHMRUVV}}[\text{up}] - P_{ij}^{\text{FHMRUVV}}[\text{down}])$$

- Very little change in PDFs, but some fit quality improvement (later).

Approximate N3LO - Update TMEs

- Update Transition Matrix Elements (TMEs) to exact $A_{gg,H}$, A_{Hg} , $A_{qq,H}$, $A_{qg,H}$.

Via grids, as in TC et al (MSHT): 2510.09321

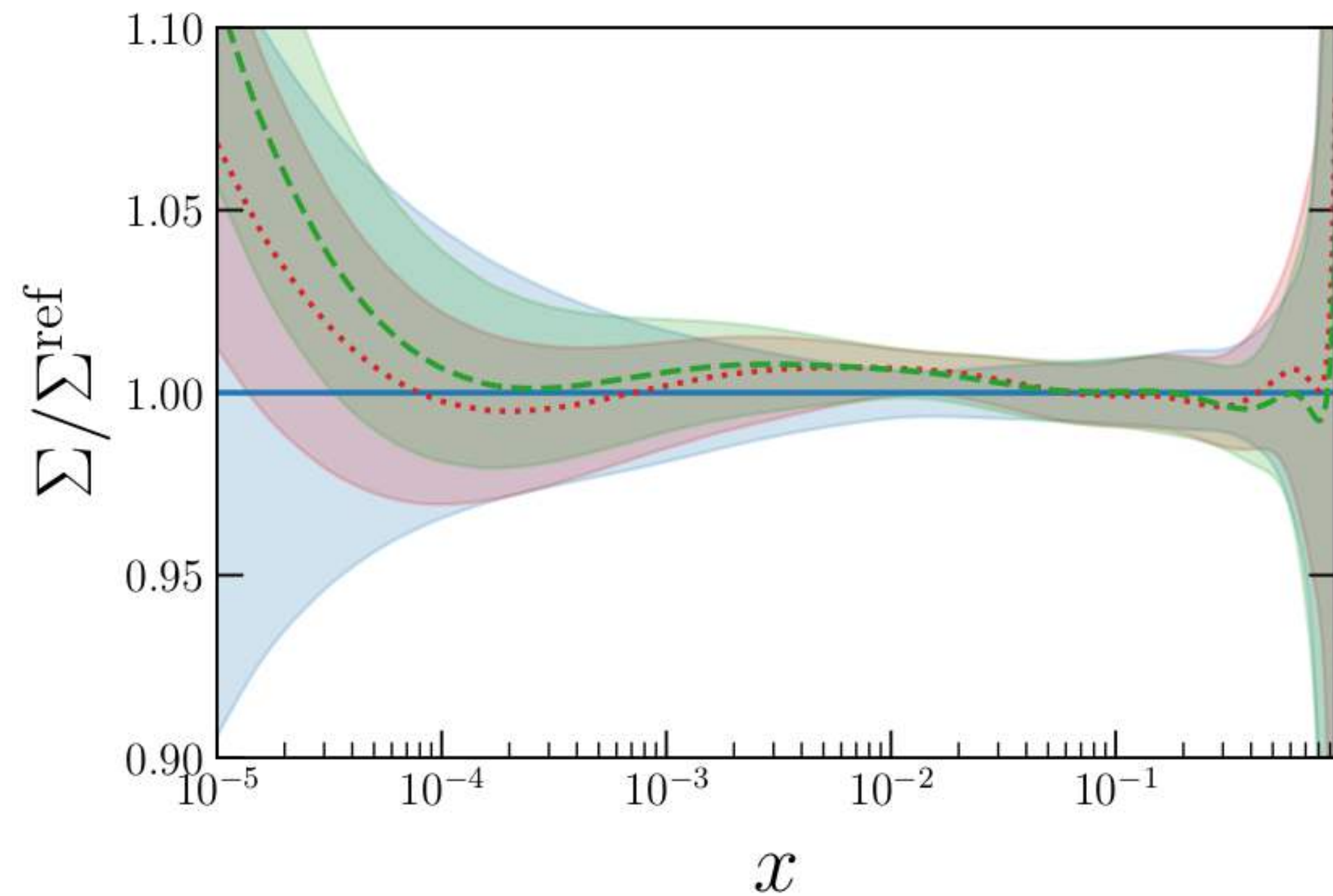
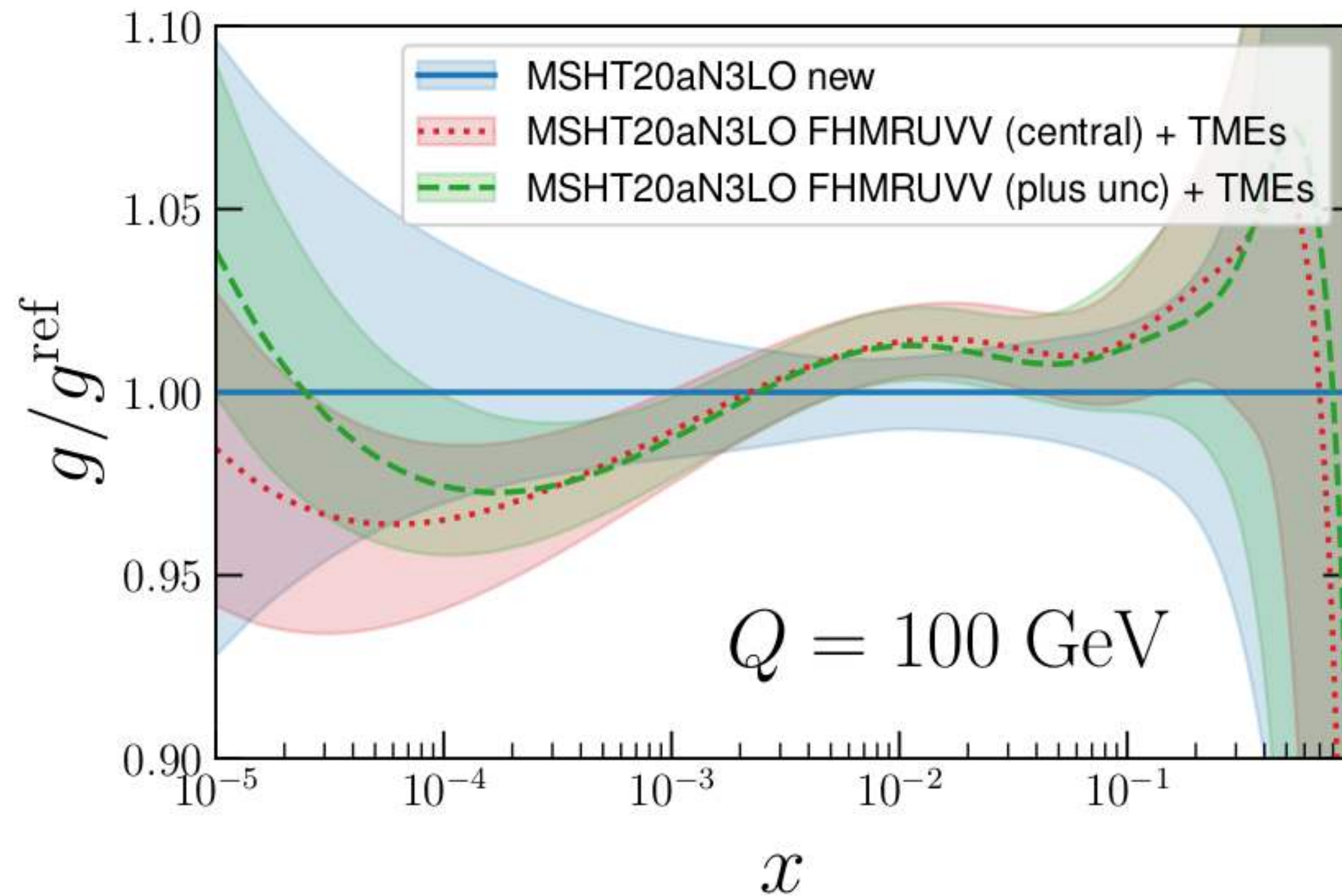


- Limited effects on PDF central values, within uncertainties.
- Effects on heavy quarks larger near threshold, though still consistent.

Approximate N3LO - Update FHMRUVV + TMEs

- Update Transition Matrix Elements (TMEs) to exact $A_{gg,H}$, A_{Hg} , $A_{qq,H}$, $A_{qg,H}$.

Via grids, as in TC et al (MSHT): 2510.09321

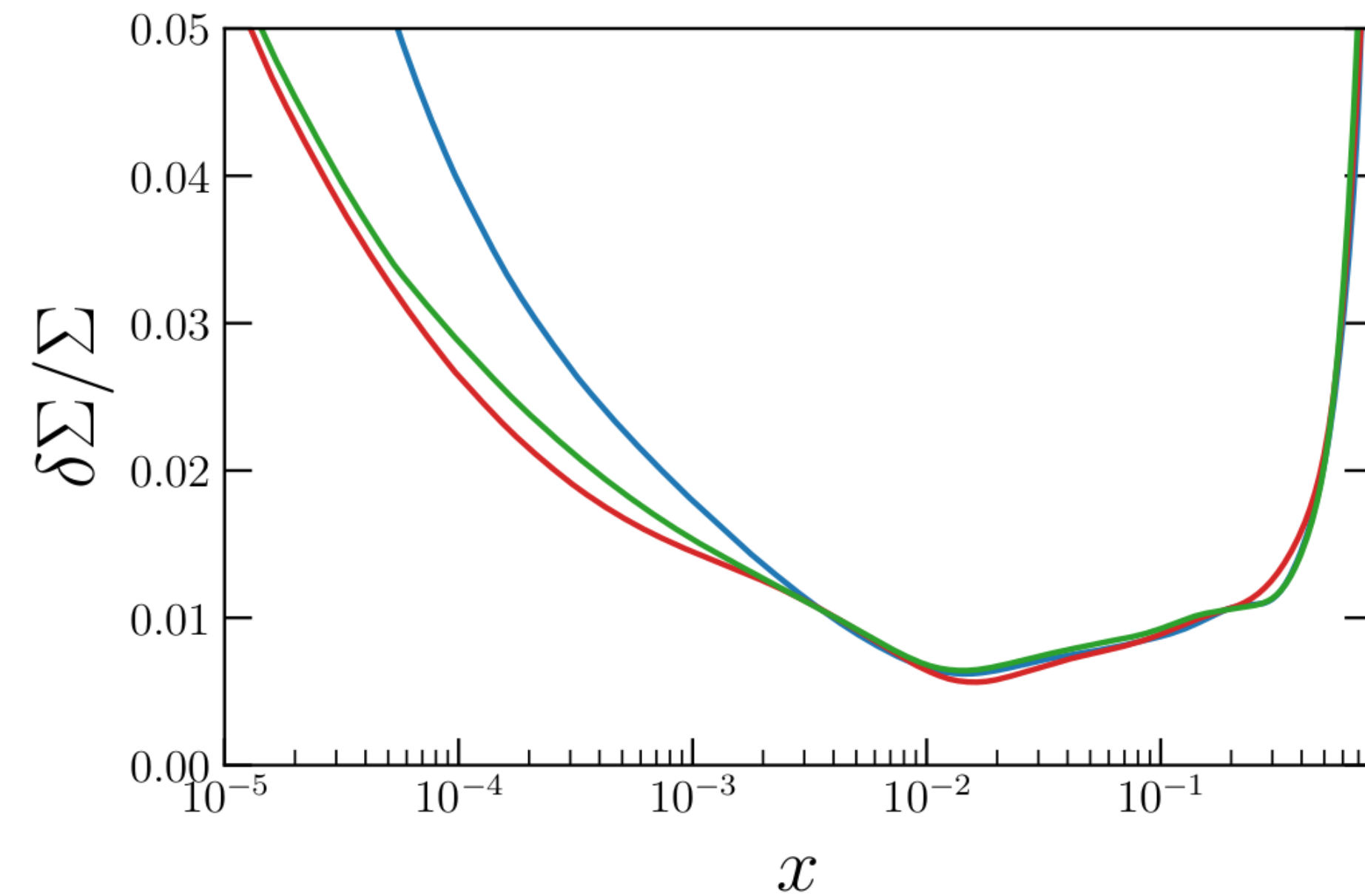
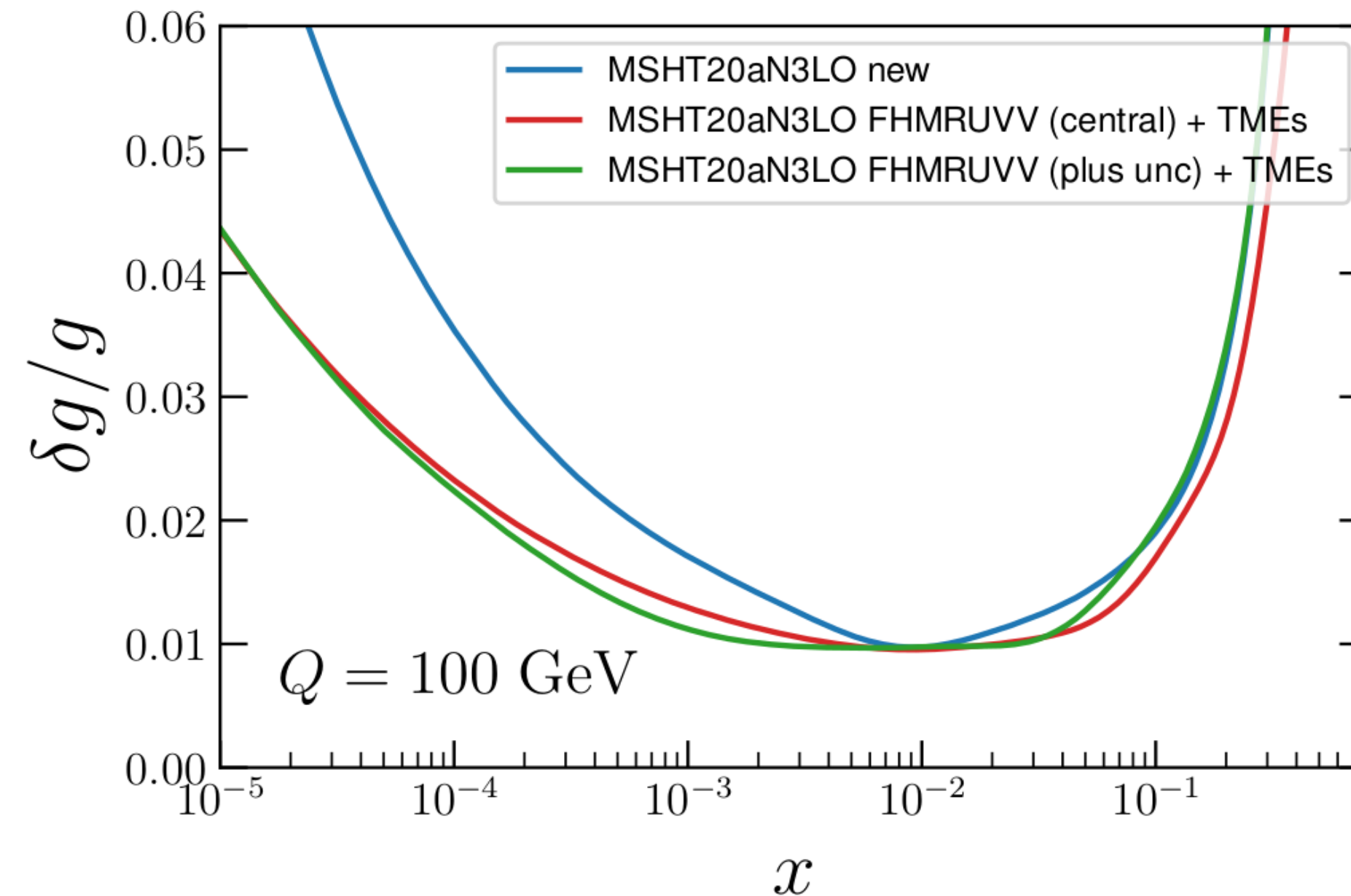


- Combine **TME + FHMRUVV** updates: first central, then + uncertainty.
- Limited effects on PDFs within uncertainties**, some increase in small x gluon.

Approximate N3LO - Update FHMRUVV + TMEs

- Update Transition Matrix Elements (TMEs) to exact $A_{gg,H}$, A_{Hg} , $A_{qq,H}$, $A_{qg,H}$.

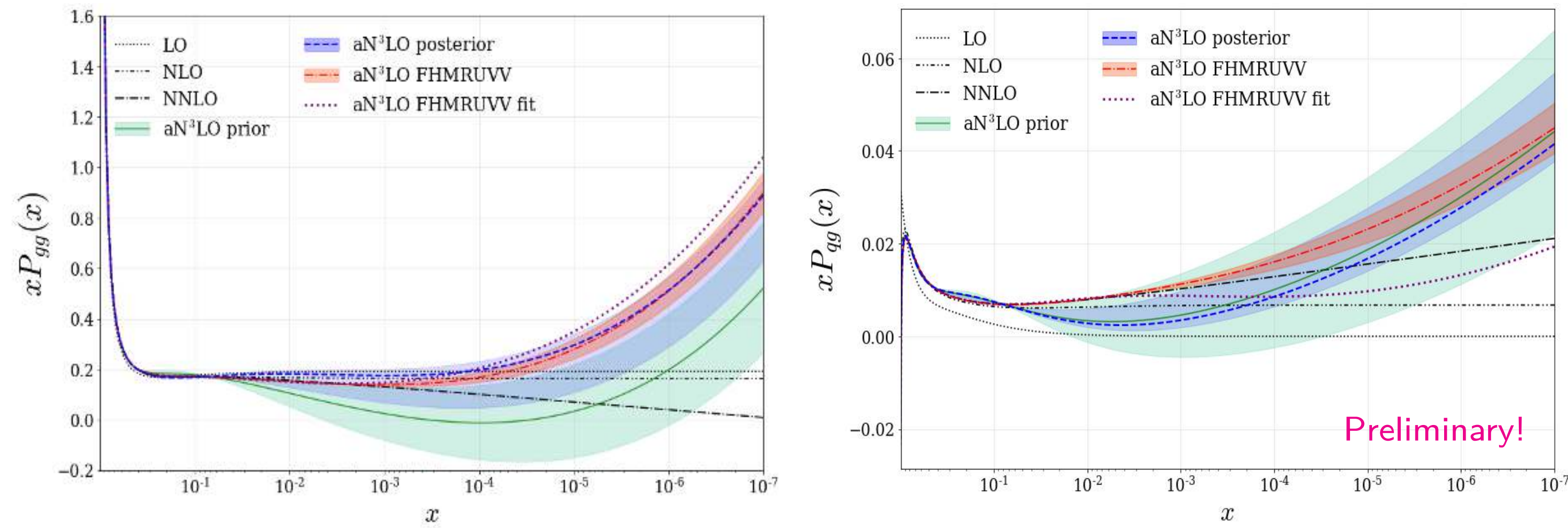
Via grids, as in TC et al (MSHT): 2510.09321



- Overall updates result in reduced PDF uncertainties, mainly at low x .

Approximate N3LO - Update FHMRUVV + TMEs

- When you include a theory uncertainty in fit you allow the central theory to move (whether via TNPs or theory cov matrix*). *in Gaussian, linear limit
- How do post-fit splitting functions look? Second with TME update:



- Splitting Functions follow FHMRUVV central down to $x \sim 10^{-3}$, slight shifts at lower x .
- P_{qg} most pulled, particularly after adding TME updates, still at low x .
- Work ongoing to investigate more general approach via TNPs for splitting functions.

Approximate N3LO - Update FHMRUVV + TMEs

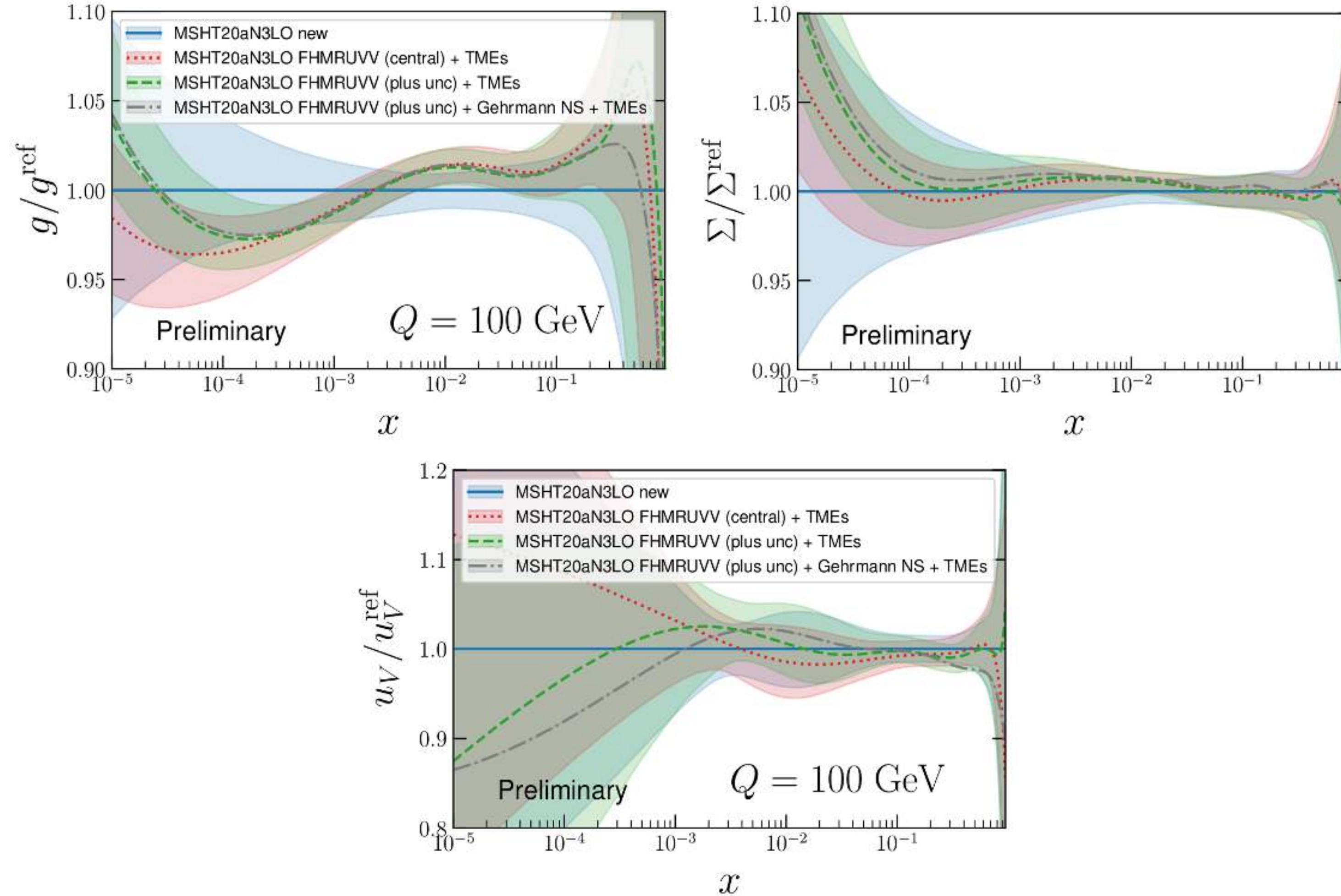
- Fit qualities, $\Delta\chi^2$ compared to new baseline NNLO fit: Preliminary!

MSHT new fits Δ^2 cf NNLO	aN3LO	aN3LO + FHMRUVV central	aN3LO + FHMRUVV plus unc	aN3LO + TMEs	aN3LO + FHMRUVV central + TMEs	aN3LO + FHMRUVV plus unc + TMEs
DIS	-41.0	17.1	-4.5	44.6	96.2	37.2
DY	-133.0	-99.2	-130.0	-105.1	-110.1	-121.5
Jets	-24.4	-25.7	-26.4	-23.8	-23.8	-21.8
Top	-4.8	-4.3	-4.9	-4.2	-4.2	-4.0
Theory penalties (none at NNLO)	13.8	10.7	11.5	23.1	9.0	17.3
Total ($N_{pts} = 4892$)	-189.5	-101.5	-154.3	-65.3	-32.9	-92.8

- In all cases aN3LO fit quality notably better than NNLO.
- Reduces as less uncertainty on aN3LO ingredients.
- Slight worsening of DIS (F_2^C) at aN3LO offset by improvement of DY data (mainly ATLAS 8TeV Zp_T , ATLAS 7TeV W, Z and CMS 7TeV Z).

Approximate N3LO - Add also Exact Non-Singlet:

- Following 2604.09534 we can replace NS with exact results: Preliminary!



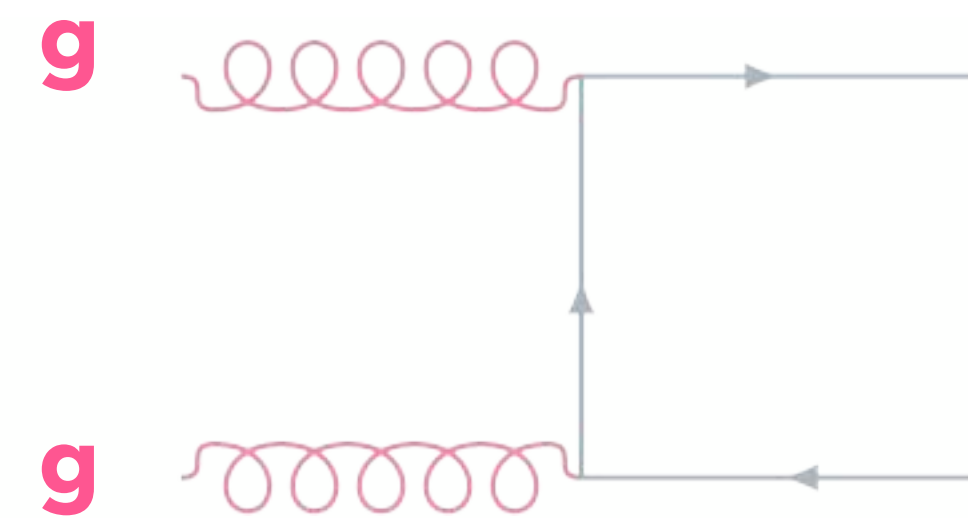
- As expected very limited impact, focused on high x . $\Delta\chi^2 < 2$ units.

Impact of 13 TeV jet data

Overview

- Jet production a **key ingredient** in modern PDF fits.
- By pushing to larger jet $p_{\perp}$ (dijet m_{jj}) go to larger x .
- Quark-initiated contribution tends to be better constrained $\rightarrow$ particularly relevant for **gluon** at high x .
- **NNLO QCD** (and **NLO EW**) theory available
 - full colour - for both inclusive and dijet data.
- But also studied using state-of-the-art aN3LO theory + missing higher order uncertainty/K-factor.
- Question of whether to include inclusive jets vs. dijets has been studied (ideally both), but triple differential dijets particularly promising - pins down initial-state kinematics completely.

$$x_{1,2} = \frac{p_{\perp}}{\sqrt{s}} \left(e^{\pm y_j} + e^{\pm y_{j'}} \right)$$



Earlier MSHT results

- Impact of jets/dijets studied by different groups. Highlight some ($\sim$ representative) MSHT results from dedicated paper on this.

- **Inclusive jets:** $0.0 < |y| < 2.5 - 3.0$ **T. Cridge, LHL, R.S. Thorne, *Eur.Phys.J.C* 84 (2024) 4, 446**
 - ★ CMS 2.76 TeV: 81 points — 5.43 pb⁻¹ — $74 < p_{\perp} < 592$ GeV
 - ★ CMS 7 TeV: 158 points — 5.0 fb⁻¹ — $74 < p_{\perp} < 2500$ GeV
 - ★ CMS 8 TeV: 174 points — 19.7 fb⁻¹ — $60 < p_{\perp} < 1300$ GeV
 - ★ ATLAS 7 TeV: 140 points — 4.5 fb⁻¹ — $100 < p_{\perp} < 2000$ GeV
 - ★ **ATLAS 8 TeV:** 171 points — 20.2 fb⁻¹ — $70 < p_{\perp} < 2500$ GeV

→ 724 points in total, v.s. ~ 4500 in global MSHT fit (inc.).
- **Dijets:**
 - ★ ATLAS 7 TeV: 90 points — 4.5 fb⁻¹ — $\frac{d^2\sigma/dm_{jj}d|y_{\max}|}{0.26 < m_{jj} < 5.04 \text{ TeV}}$
 - ★ CMS 7 TeV: 54 points — 5.0 fb⁻¹ — $\frac{d^2\sigma/dm_{jj}d|y^*|}{0.25 < m_{jj} < 4.48 \text{ TeV}}$
 - ★ CMS 8 TeV: 122 points — 19.7 fb⁻¹ — $\frac{d^3\sigma/dp_{\perp,avg}dy_bdy^*}{143 < p_{\perp,avg} < 1638 \text{ GeV}}$

→ 266 points in total, v.s. ~ 4000 in global MSHT fit (inc.).

New data in fit

- Inclusive jets:

★ ATLAS 13 TeV: 177 points - 3.2 fb^{-1} - $0.1 < p_{\perp} < 3 \text{ TeV}$ *ATLAS, JHEP 05 (2018) 195*

★ CMS 13 TeV: 78 points - 36.3 fb^{-1} - $0.1 < p_{\perp} < 3.1 \text{ TeV}$ *CMS, JHEP 02 (2022) 142*

- Dijets:

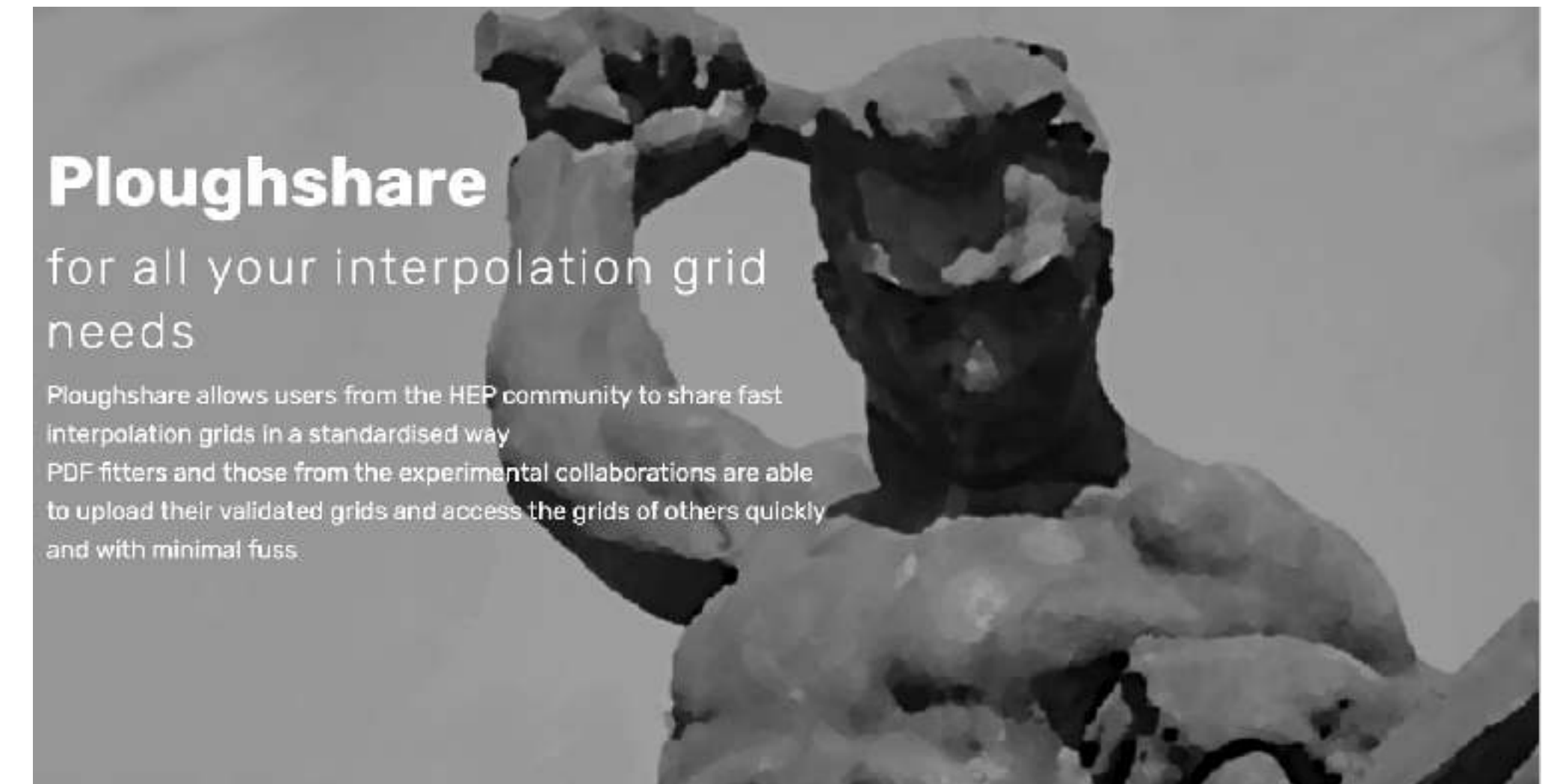
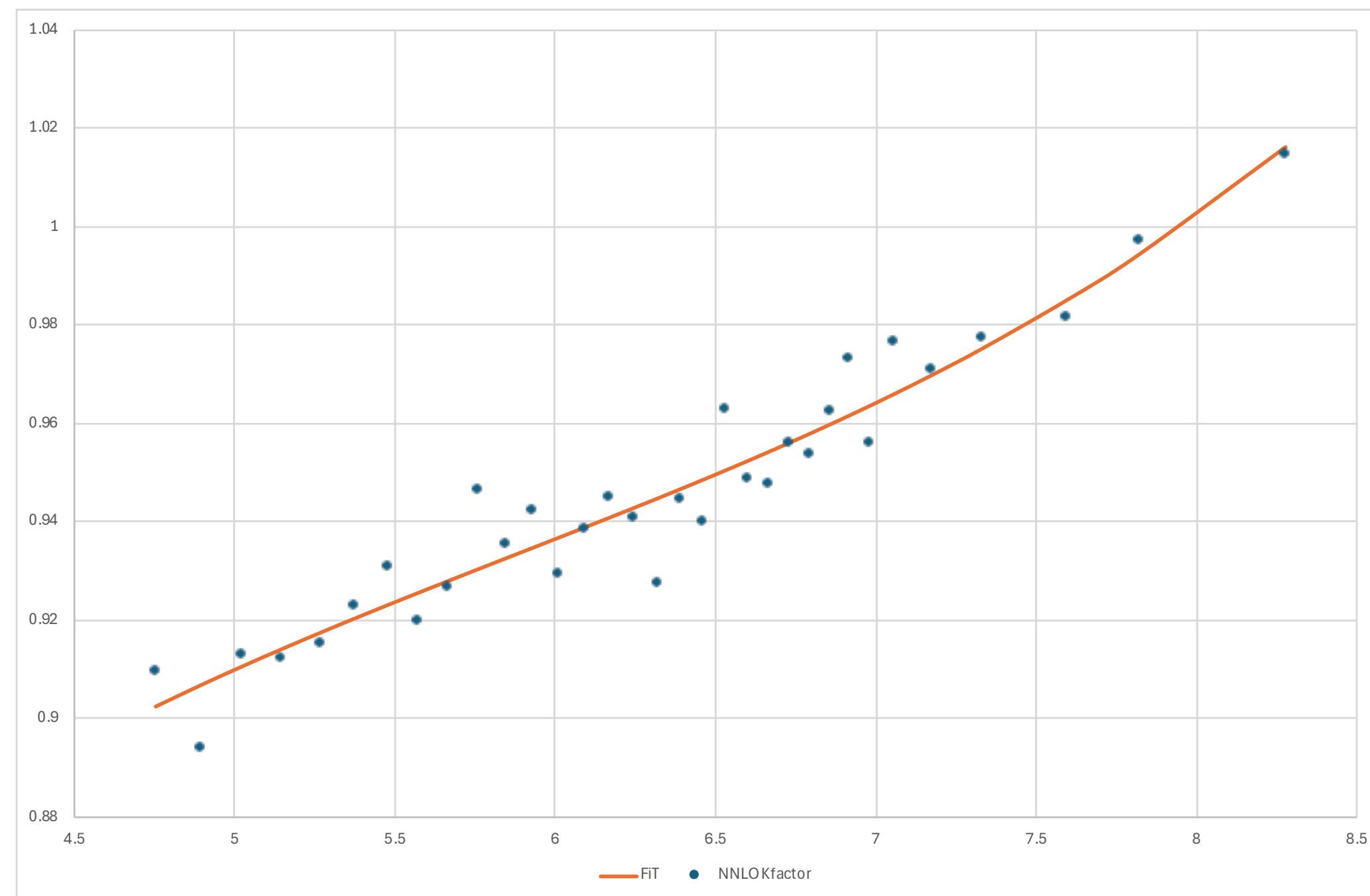
★ ATLAS 13 TeV: 136 points - 3.2 fb^{-1} - $\frac{d^2\sigma}{dm_{jj}d|y^*|}$ - $0.3 < m_{jj} < 9 \text{ TeV}$ *ATLAS, JHEP 05 (2018) 195*

★ CMS 13 TeV: 164 points - 36.3 fb^{-1} - $\frac{d^3\sigma}{dm_{jj}dy_bdy^*}$ - $0.3 < m_{jj} < 6 \text{ TeV}$ *CMS, Eur.Phys.J.C 85 (2025) 1, 72*

- $\sim 250 - 300$ new datapoints in total.
- In dijet case CMS again available (and fitted) triple differentially.
- aN3LO fits performed with most recent updates.

Technical Aside - NNLO theory

- NNLO predictions all calculated using full interpolations grids - not K-factors.



- But as always these do contain non-negligible MC uncertainties.
- Can either smooth or extract directly from grids.
- Unlikely to impact pheno that much, but worth bearing in mind!

Fit quality - NNLO

		NNLO				
	N_{pts}	$\sqrt{2/N_{\text{pts}}}$	< 13 TeV jets	< 13 TeV dijets	13 TeV jets	13 TeV dijets
< 13 TeV jets	724	0.05	1.62	1.69	1.64	1.66
< 13 TeV dijets	266	0.09	1.43	1.26	1.41	1.38
ATLAS 13 TeV jets	78	0.16	3.08	2.93	2.90	2.87
CMS 13 TeV jets	177	0.11	1.96	1.85	1.76	1.80
13 TeV jets	255	0.09	2.74	2.60	2.55	2.54
ATLAS 13 TeV dijets	136	0.12	1.25	1.34	1.25	1.25
CMS 13 TeV dijets	164	0.11	2.76	3.05	2.61	2.60
13 TeV dijets	300	0.08	2.08	2.27	2.00	1.99
Non-jets	4313	0.02	1.19	1.19	1.19	1.19
Total (fit)	-	-	1.25	1.22	1.32	1.27

- Fit quality rarely good in statistical sense (ongoing issue). Some improvement on refitting.

Fit quality - NNLO

	N_{pts}	NNLO				
		$\sqrt{2/N_{\text{pts}}}$	< 13 TeV jets	< 13 TeV dijets	13 TeV jets	13 TeV dijets
< 13 TeV jets	724	0.05	1.62	1.69	1.64	1.66
< 13 TeV dijets	266	0.09	1.43	1.26	1.41	1.38
ATLAS 13 TeV jets	78	0.16	3.08	2.93	2.90	2.87
CMS 13 TeV jets	177	0.11	1.96	1.85	1.76	1.80
13 TeV jets	255	0.09	2.74	2.60	2.55	2.54
ATLAS 13 TeV dijets	136	0.12	1.25	1.34	1.25	1.25
CMS 13 TeV dijets	164	0.11	2.76	3.05	2.61	2.60
13 TeV dijets	300	0.08	2.08	2.27	2.00	1.99
Non-jets	4313	0.02	1.19	1.19	1.19	1.19
Total (fit)	-	-	1.25	1.22	1.32	1.27

- Fit quality rarely good in statistical sense (ongoing issue). Some improvement on refitting.
- Improvement also occurs when fitting the alternative 13 TeV dataset - compatability.

Fit quality - NNLO

	N_{pts}	NNLO				
		$\sqrt{2/N_{\text{pts}}}$	< 13 TeV jets	< 13 TeV dijets	13 TeV jets	13 TeV dijets
< 13 TeV jets	724	0.05	1.62	1.69	1.64	1.66
< 13 TeV dijets	266	0.09	1.43	1.26	1.41	1.38
ATLAS 13 TeV jets	78	0.16	3.08	2.93	2.90	2.87
CMS 13 TeV jets	177	0.11	1.96	1.85	1.76	1.80
13 TeV jets	255	0.09	2.74	2.60	2.55	2.54
ATLAS 13 TeV dijets	136	0.12	1.25	1.34	1.25	1.25
CMS 13 TeV dijets	164	0.11	2.76	3.05	2.61	2.60
13 TeV dijets	300	0.08	2.08	2.27	2.00	1.99
Non-jets	4313	0.02	1.19	1.19	1.19	1.19
Total (fit)	-	-	1.25	1.22	1.32	1.27

- Fit quality rarely good in statistical sense (ongoing issue). Some improvement on refitting.
- Improvement also occurs when fitting the alternative 13 TeV dataset - compatability.
- No big shifts in earlier jets when fitting 13 TeV - no significant tensions.

Fit quality - N3LO

		aN ³ LO			
	N_{pts}	< 13 TeV jets	< 13 TeV dijets	13 TeV jets	13 TeV dijets
< 13 TeV jets	724	1.61	1.66	1.62	1.64
< 13 TeV dijets	266	1.21	1.01	1.21	1.30
ATLAS 13 TeV jets	78	3.03	2.87	2.84	2.80
CMS 13 TeV jets	177	2.22	2.02	1.90	1.93
13 TeV jets	255	2.78	2.60	2.55	2.54
ATLAS 13 TeV dijets	136	1.40	1.76	1.36	1.46
CMS 13 TeV dijets	164	3.52	4.01	3.24	2.62
13 TeV dijets	300	2.56	3.00	2.39	2.09
Non-jets	4313	1.16	1.16	1.16	1.16
Total (fit)	-	1.23	1.18	1.29	1.25

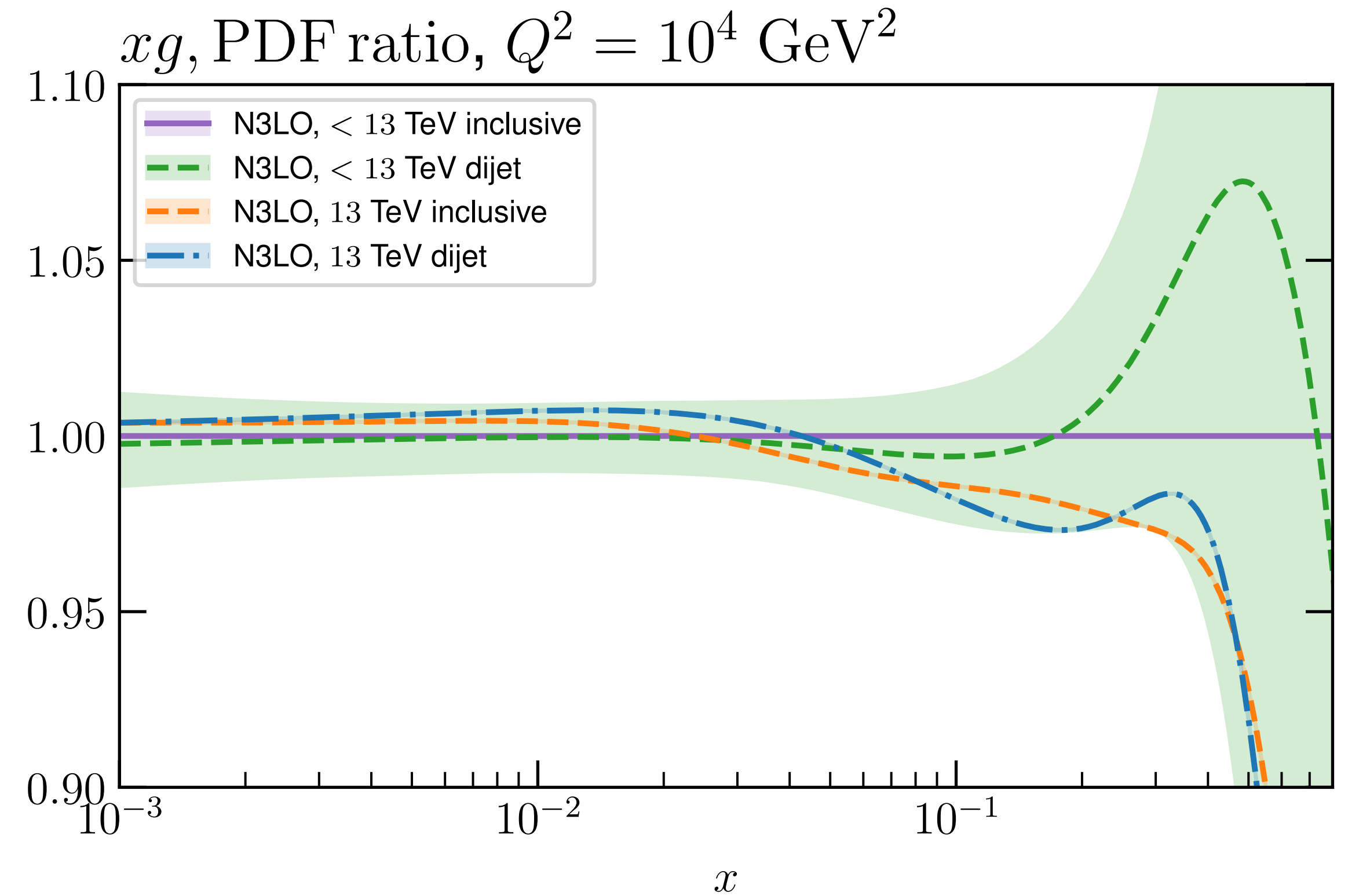
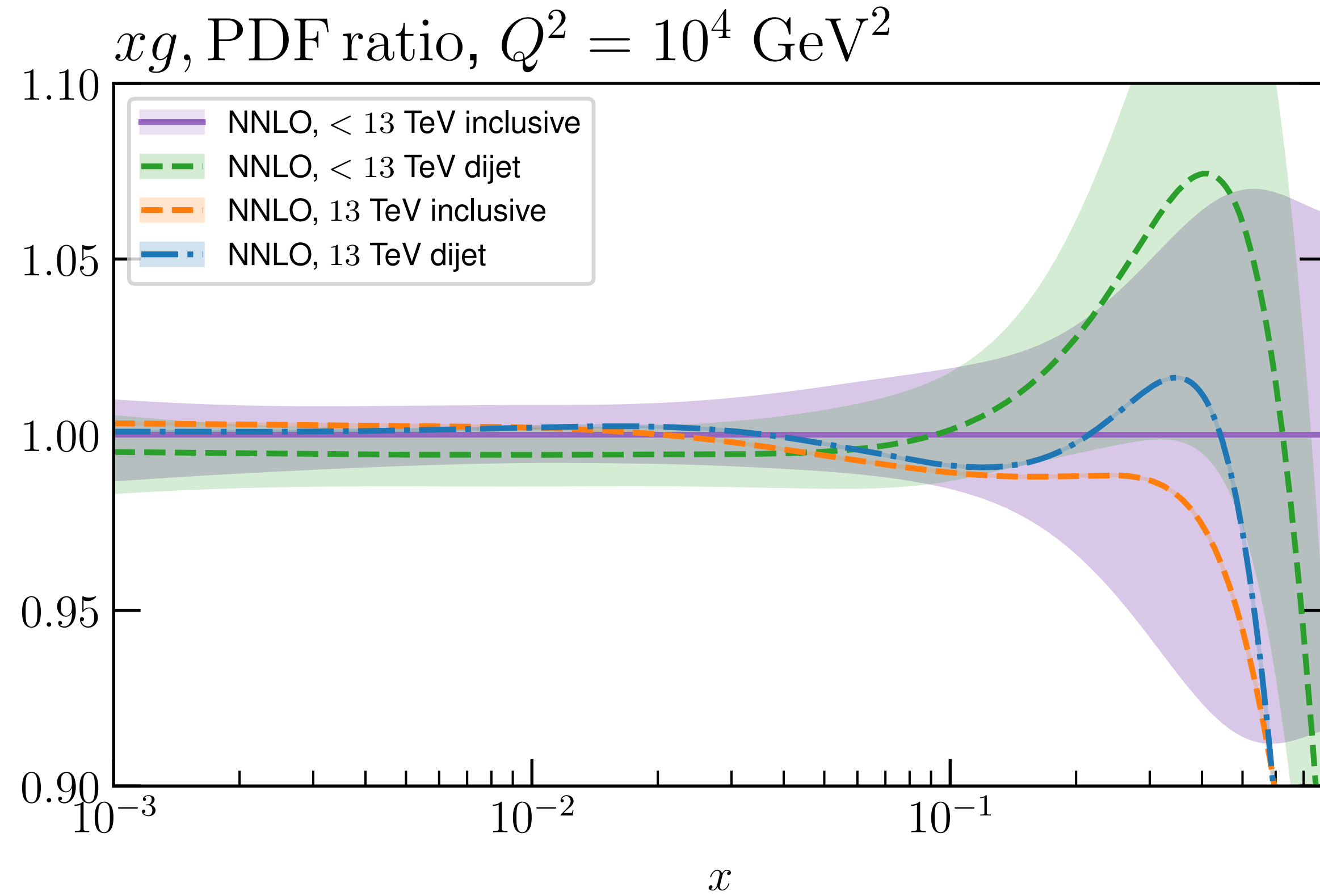
- No significant change in fit quality vs. NNLO.

Fit quality - N3LO

	N_{pts}	aN ³ LO			
		< 13 TeV jets	< 13 TeV dijets	13 TeV jets	13 TeV dijets
< 13 TeV jets	724	1.61	1.66	1.62	1.64
< 13 TeV dijets	266	1.21	1.01	1.21	1.30
ATLAS 13 TeV jets	78	3.03	2.87	2.84	2.80
CMS 13 TeV jets	177	2.22	2.02	1.90	1.93
13 TeV jets	255	2.78	2.60	2.55	2.54
ATLAS 13 TeV dijets	136	1.40	1.76	1.36	1.46
CMS 13 TeV dijets	164	3.52	4.01	3.24	2.62
13 TeV dijets	300	2.56	3.00	2.39	2.09
Non-jets	4313	1.16	1.16	1.16	1.16
Total (fit)	-	1.23	1.18	1.29	1.25

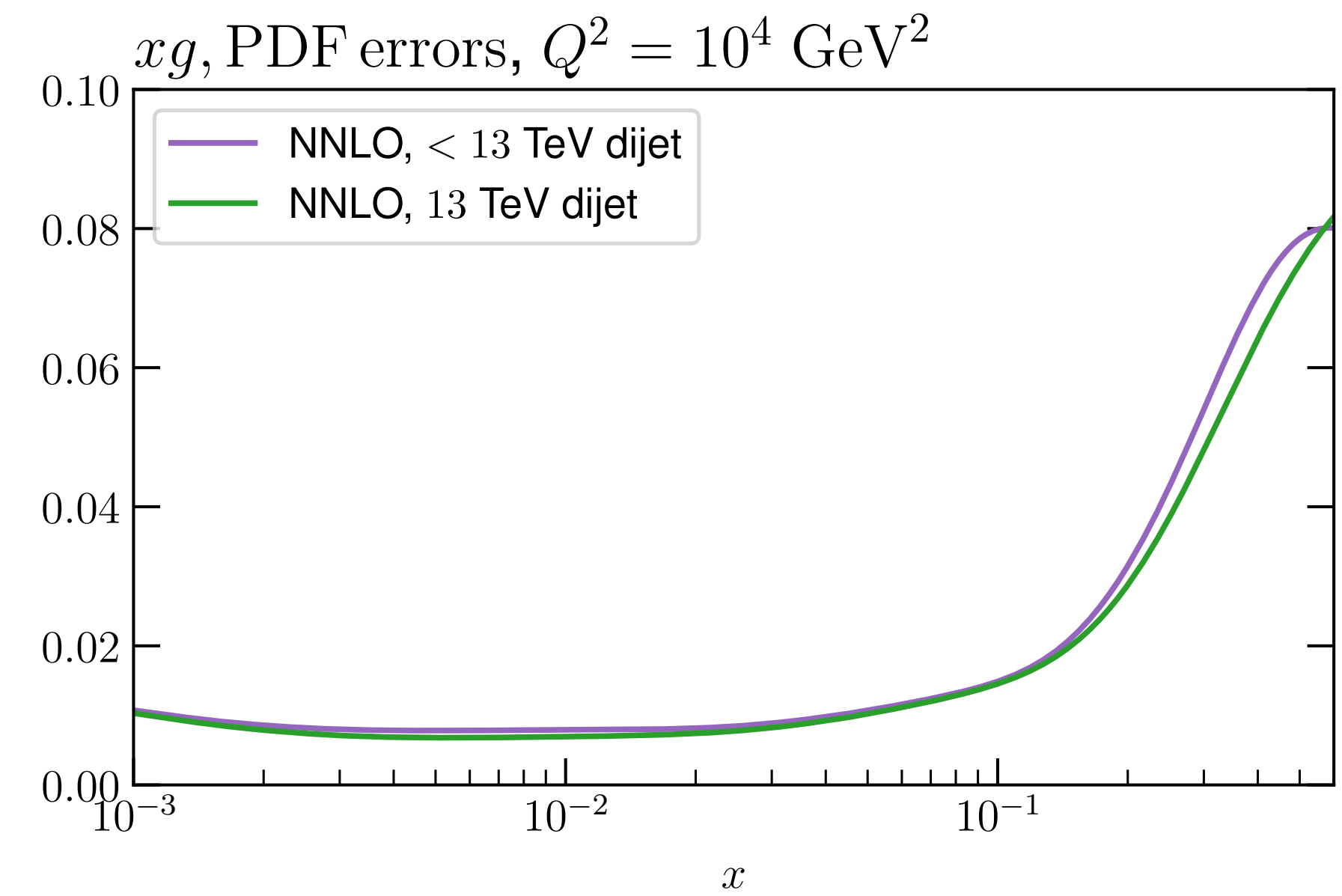
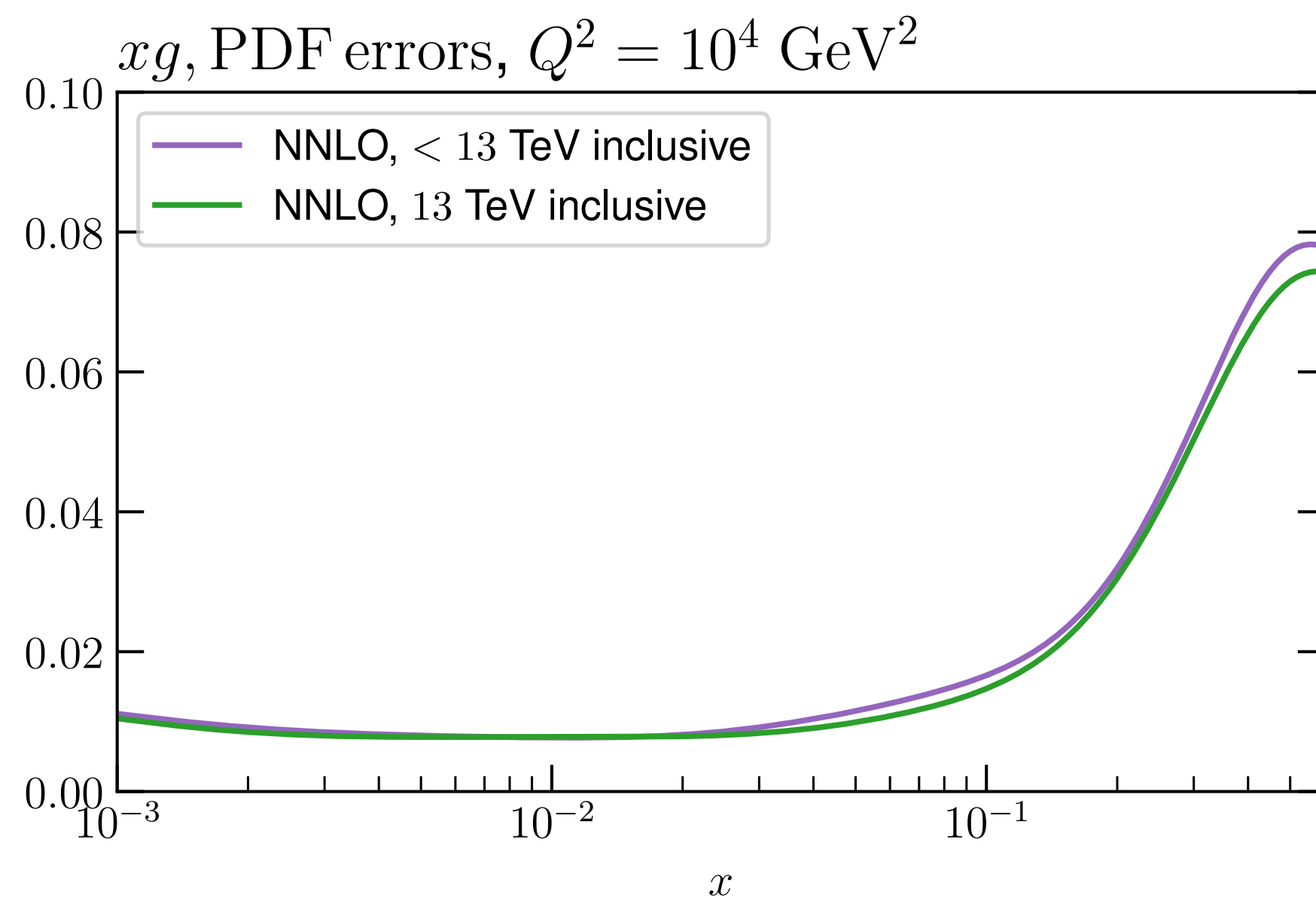
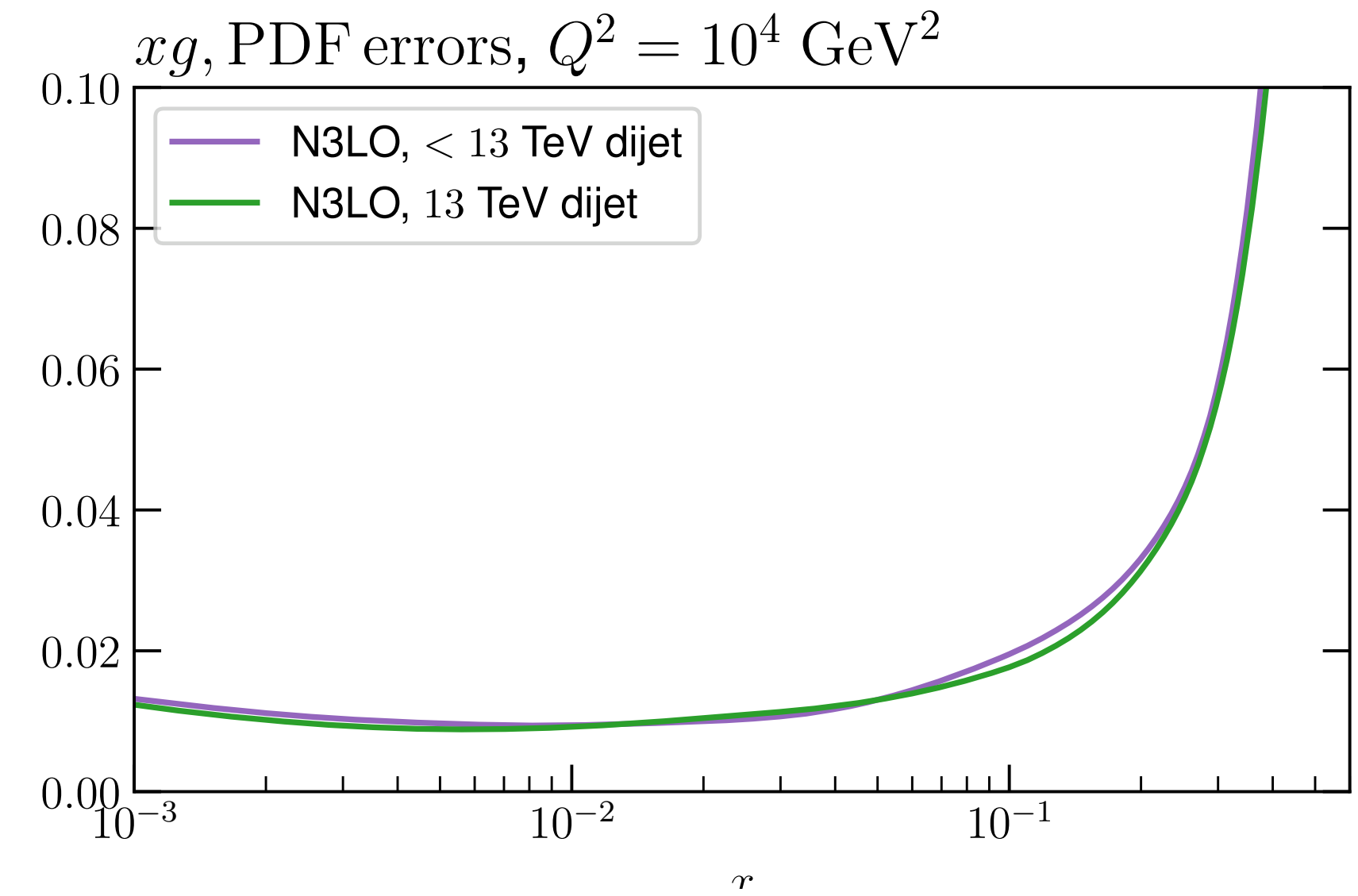
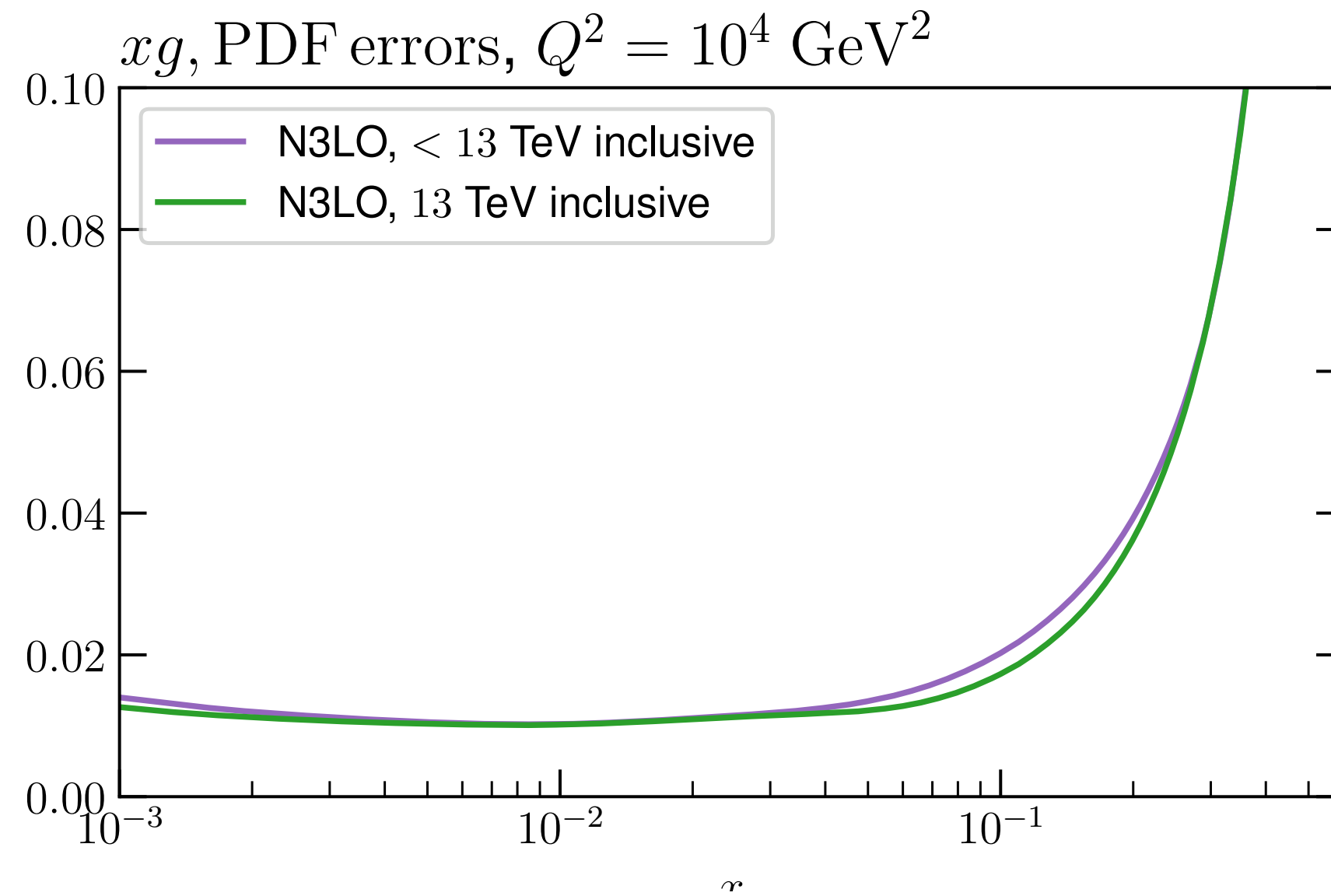
- No significant change in fit quality vs. NNLO.
- Other trends similar - somewhat larger improvement for 13 TeV dijets upon refitting.

PDFs



- Focus on gluon at higher x .
- Some difference in jet vs. dijet pull at NNLO, largely gone at aN3LO.
- Moderate trend for the 13 TeV data to improvement compatibility - consistent with fit quality.
- Overall moderate pulls.

- Consistent, but moderate improvement in uncertainty...



- Two questions from this:
 - ★ What about the relatively poor fit quality?
 - ★ Given adding in all this data is having relatively moderate impact, are there other more fundamental questions that we need to focus on?
- Addressing both of these is ongoing work...

Errors on Errors

The importance of uncertainties (1)

$\chi^2_{\text{exp+th,orig}}$	$\sqrt{2/n_{\text{dat}}}$	<i>CT18Z</i>	<i>MSHT20</i>	<i>NNPDF3.1</i>	<i>NNPDF4.0</i>
ATLAS 13 TeV single-inclusive jets $\frac{d^2\sigma}{dp_T d y }$	0.11	2.49	2.86	2.17	2.80
CMS 13 TeV single-inclusive jets ($R = 0.4$) $\frac{d^2\sigma}{dp_T d y }$	0.16	2.63	2.94	2.48	2.70
ATLAS 13 TeV di-jets $\frac{d^2\sigma}{dm_{jj} d y^* }$	0.12	1.52	1.72	1.53	1.70

A. Chiefa et al., *JHEP* 07 (2025) 067

$\sqrt{2n_{\text{dat}}} \sim 18$

χ^2/ndf	$p_{\text{T}}^{\text{jet}}$ $R = 0.4$ $R = 0.6$	
$p_{\text{T}} > 70 \text{ GeV}$		
CT14	340/171	392/171
HERAPDF2.0	405/171	418/171
NNPDF3.0	350/171	393/171
MMHT2014	354/171	399/171

ATLAS, *JHEP* 09 (2017) 020

See also LHL, A. D. Martin, R. S. Thorne, *Eur.Phys.J.C* 78 (2018) 3, 248

- Quite a few ‘subtleties’ beneath these results relating to uncertainties.
- In particular a now old, but continuing, story of often rather poor fit qualities to many LHC jet datasets - seen again in our 13 TeV fits!
- Often multiple (!) standard deviations away from a good fit. What is going on? And what can we do?

Errors on Errors

- Ongoing study on this by MSHT. **Dedicated talk on Thursday** - just advertise here.

- Potential lack of knowledge of **size** of experimental uncertainties accounted for via (unknown) ‘error-on-error’ in extended likelihood.
- Impact on fit quality and potential contribution to tolerance studied.

- Model the estimated variances, v_i , of $\sigma_{u_i}^2$, as Gamma distributed gives:

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

$$\alpha_i = \frac{1}{4r_i^2} \quad \beta_i = \frac{1}{4r_i^2 \sigma_{u_i}^2}$$

- r_i is defined as the relative uncertainty in the estimate of the systematic error. The parameters r_i can therefore be referred to as the “error on errors”.

- Potential lack of knowledge of correlations of experimental uncertainties accounted for via no ‘serial correlation’ approach.
- Provides arguably more reasonable decorrelation of uncertainties.

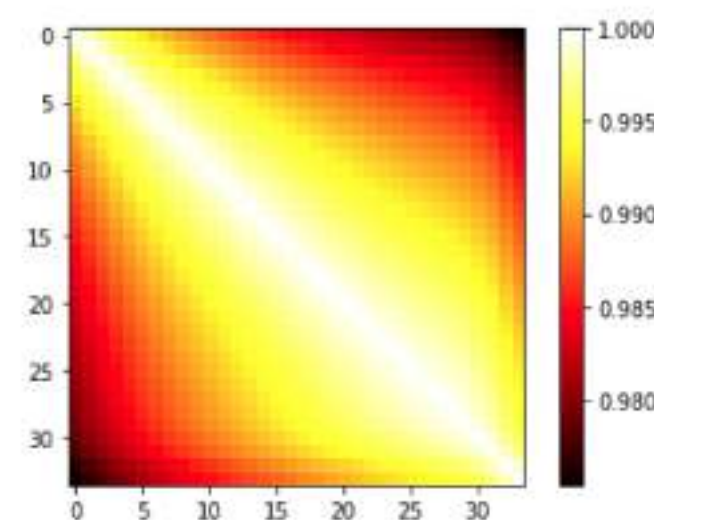
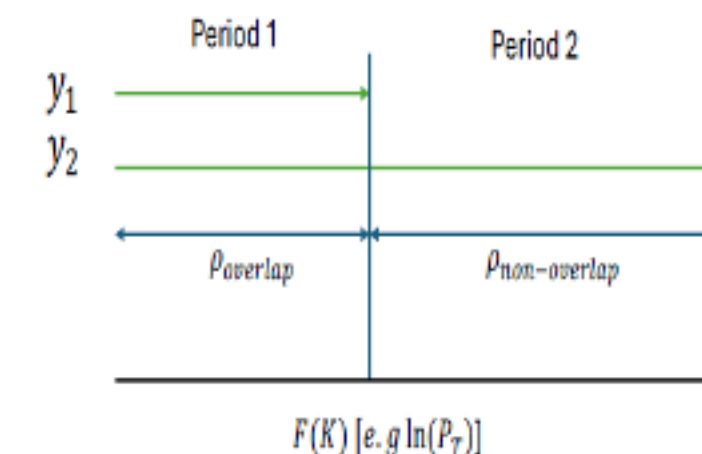
- For a given kinematic variable, K , we may want the following behaviour for a 2-point error [$z_i \sim N(0, 1)$]:

$$y_1 \sim \sqrt{F(K_1)} z_1$$

$$y_2 \sim \sqrt{F(K_1)} z_2 + \sqrt{F(K_2) - F(K_1)} z_3$$

with correlations over the 2 periods:

$$\rho_{z_1, z_2} = \rho_{\text{overlap}} \quad \rho_{z_2, z_3} = \rho_{\text{non-overlap}}$$



FPPDF

FPPDF

- New study: publicly available NNPDF code can be suitably adapted to allow fit to be performed with a fixed parameterisation, but to NNPDF data/theory.
- Original study in [2407.07944](#). Now released as open source tool:



<https://github.com/FPPDF/fppdf>

hep-ph/2602.07118

In collaboration with: L. Harland-Lang, T. Giani

- Key point: this allows us to do both a NN and a fixed parameterisation fit to exactly the same data/theory settings. The **only difference** is **input parameterisation**.
- For the first time we can begin to fully disentangle fitting methodology (parameterisation) vs. data/theory choices.

- ▶ Which differences are driven by physics and which are driven by methodological choices? (is there a clear delimitation between both?)
- ▶ Can we identify general trends when going from NNLO to N3LO? Are they driven by the methodology or the data/theory?
- ▶ What is the impact of N3LO corrections in PDF fits?
Can we give an answer which does not depend on the fitting methodology?

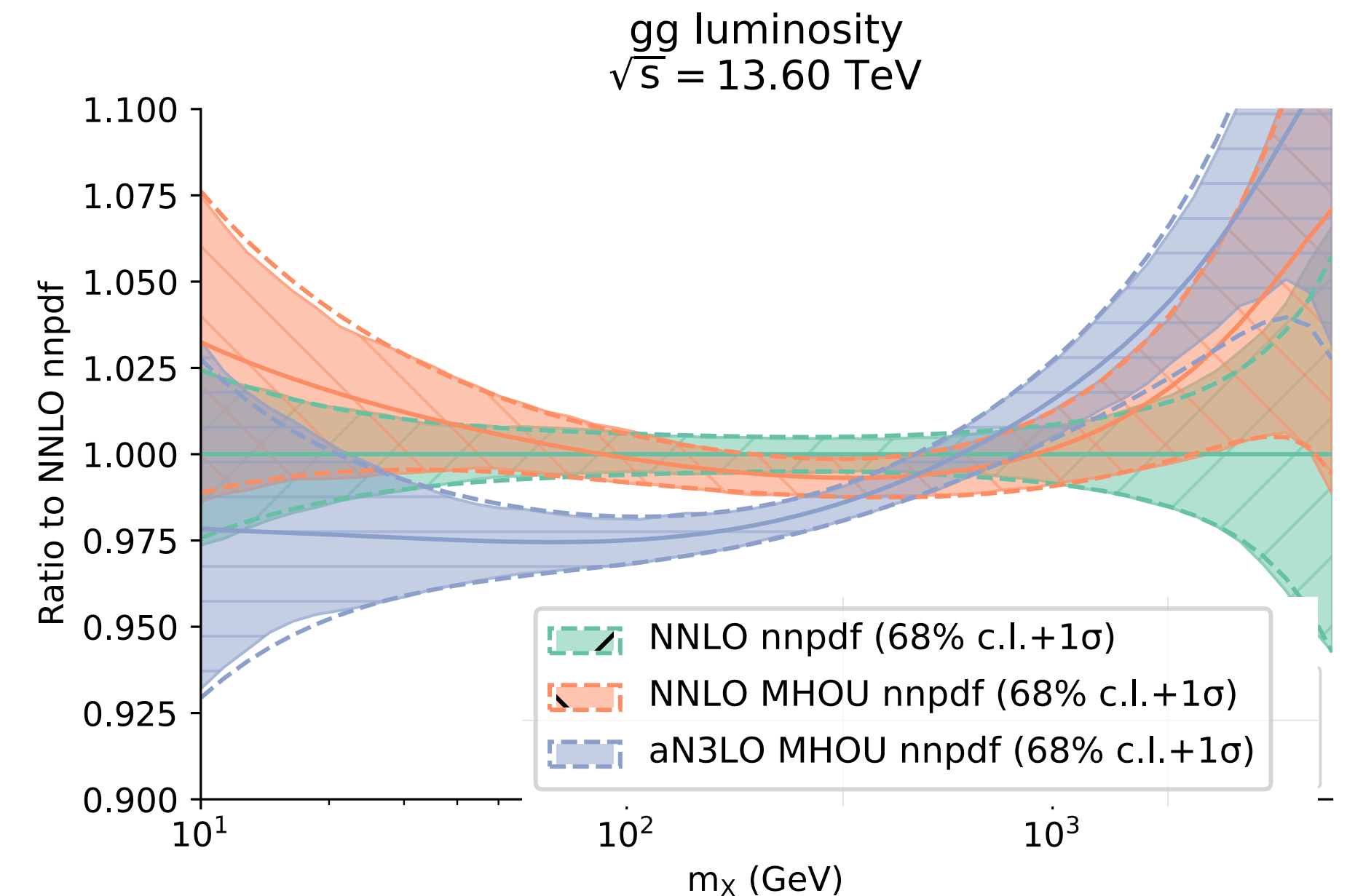
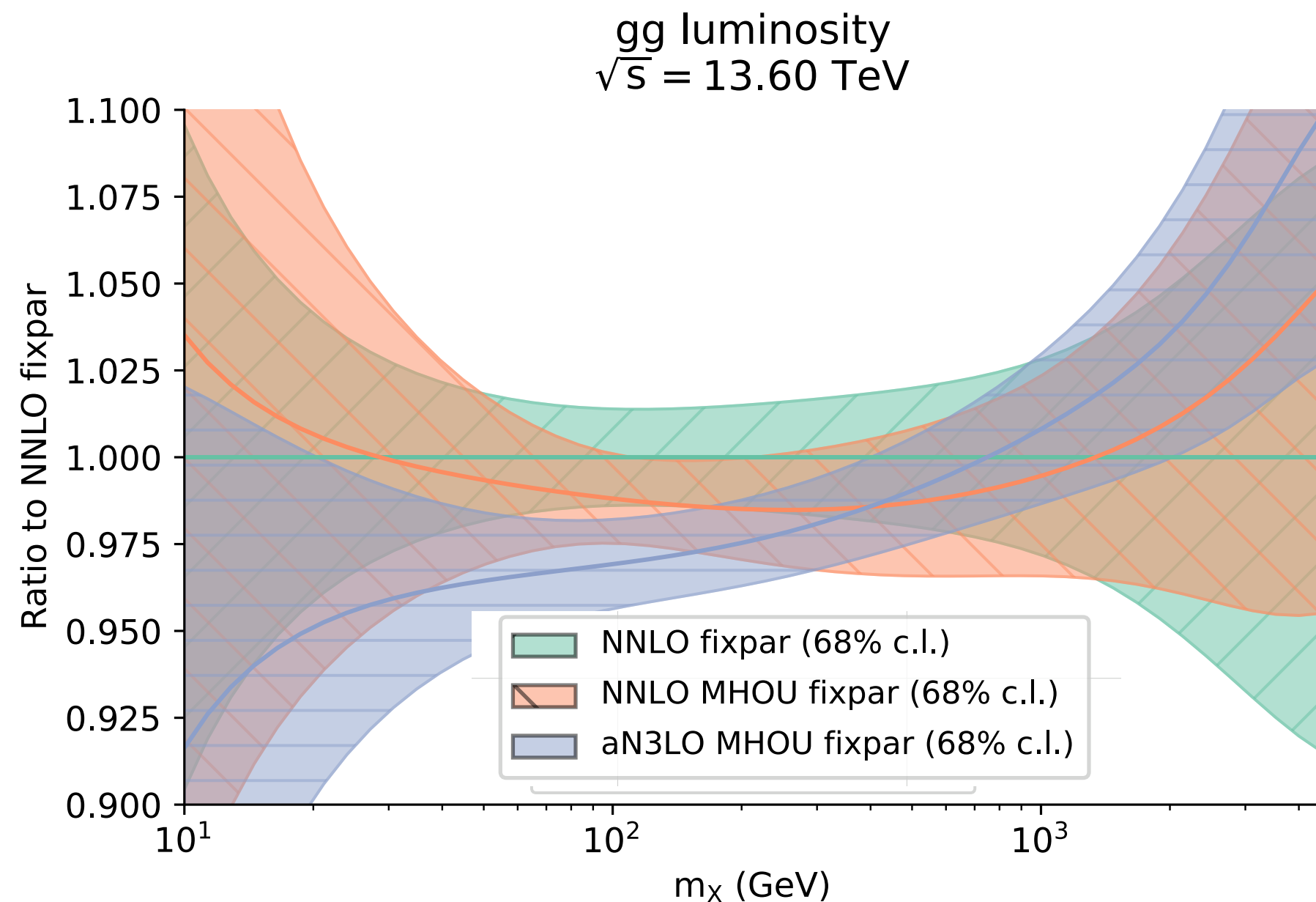
→ **FPPDF** to the rescue!?

- Perform fits to NNPDF4.0 data/theory settings (fitted charm) as before, but now at up to aN3LO, with and without missing higher order uncertainties (MHOUUs).
- Similar trends in **fit qualities** for NNLO vs. aN3LO in both methodologies. What about PDFs?

Data, theory, calculations

Impact of N3LO corrections to parton luminosities

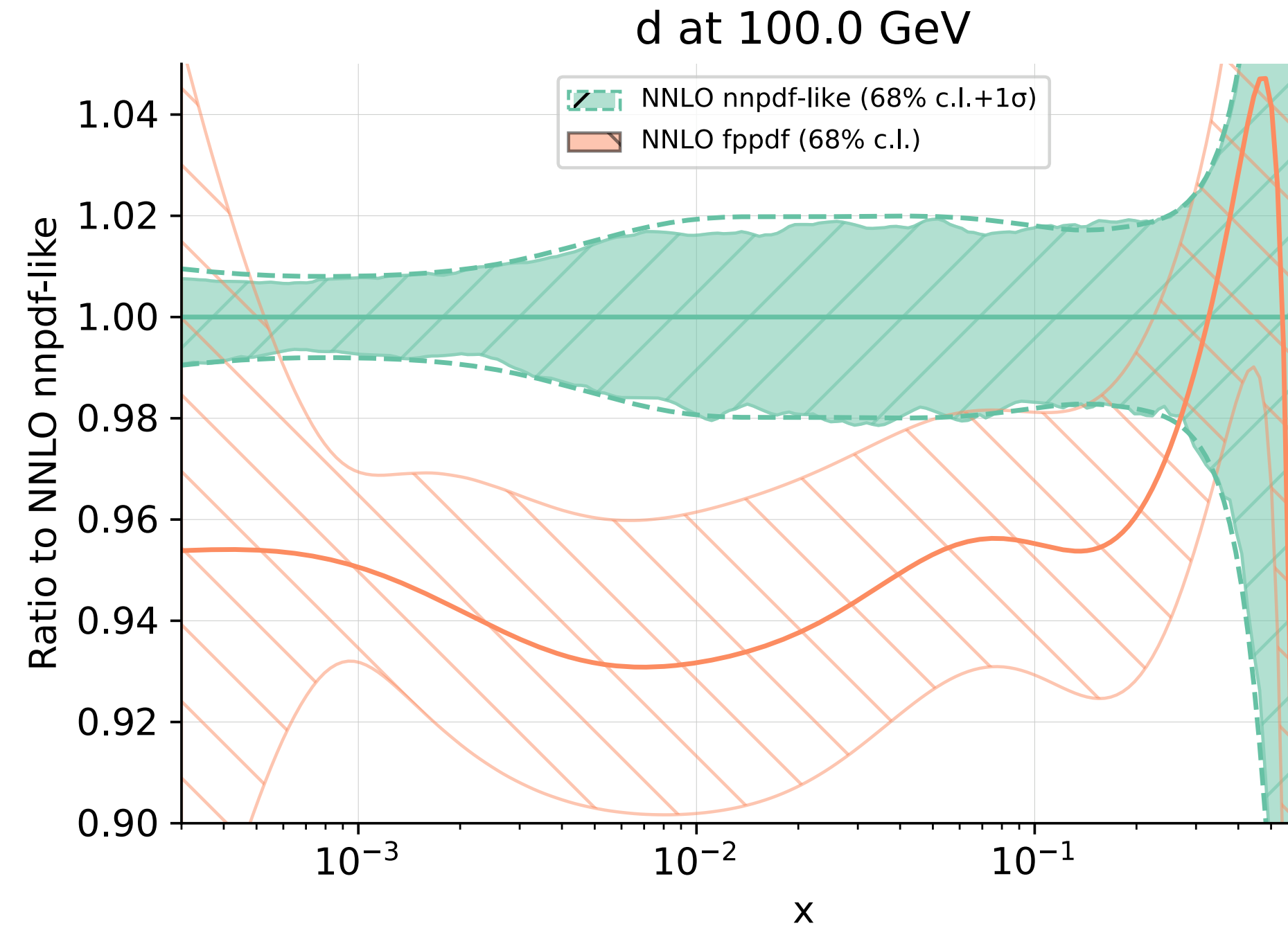
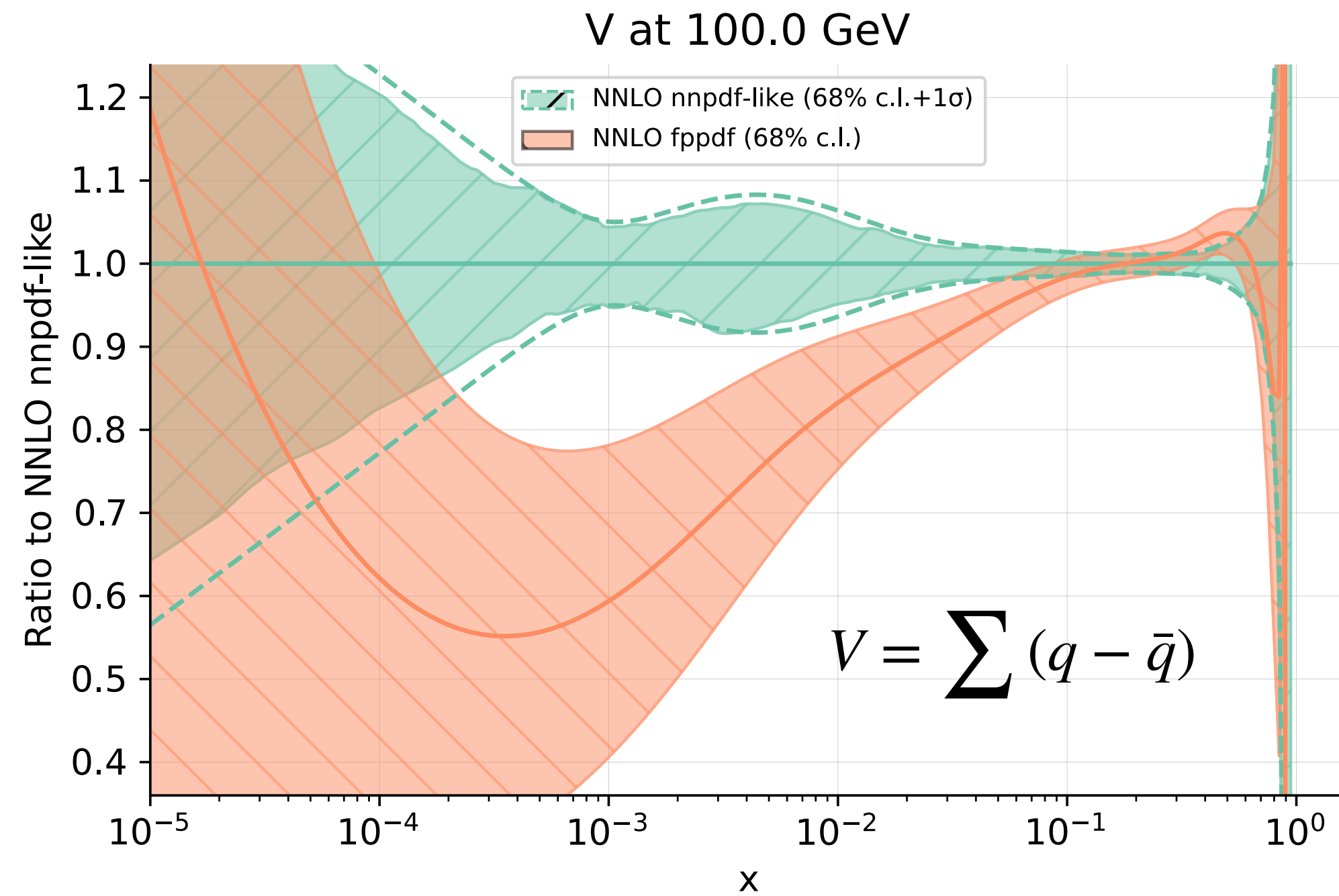
NNPDF



- Very similar in terms of central values of PDFs - no strong parameterisation dependence here.

→ NNLO to aN3LO rather stable w.r.t parameterisation (MSHT vs. NN). Any differences here rather driven by other - data, theory - differences.

- But, methodologies do not always agree.

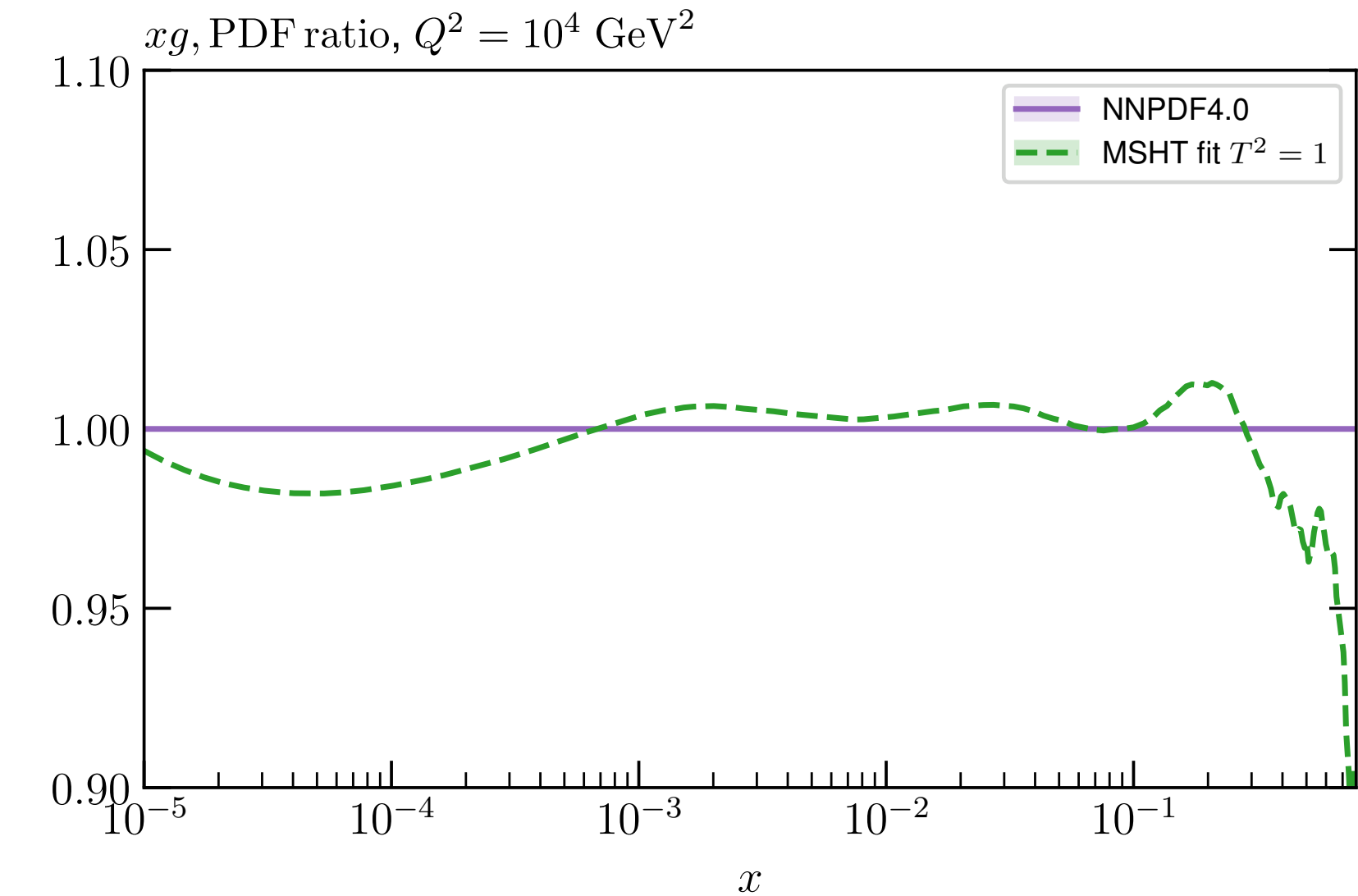
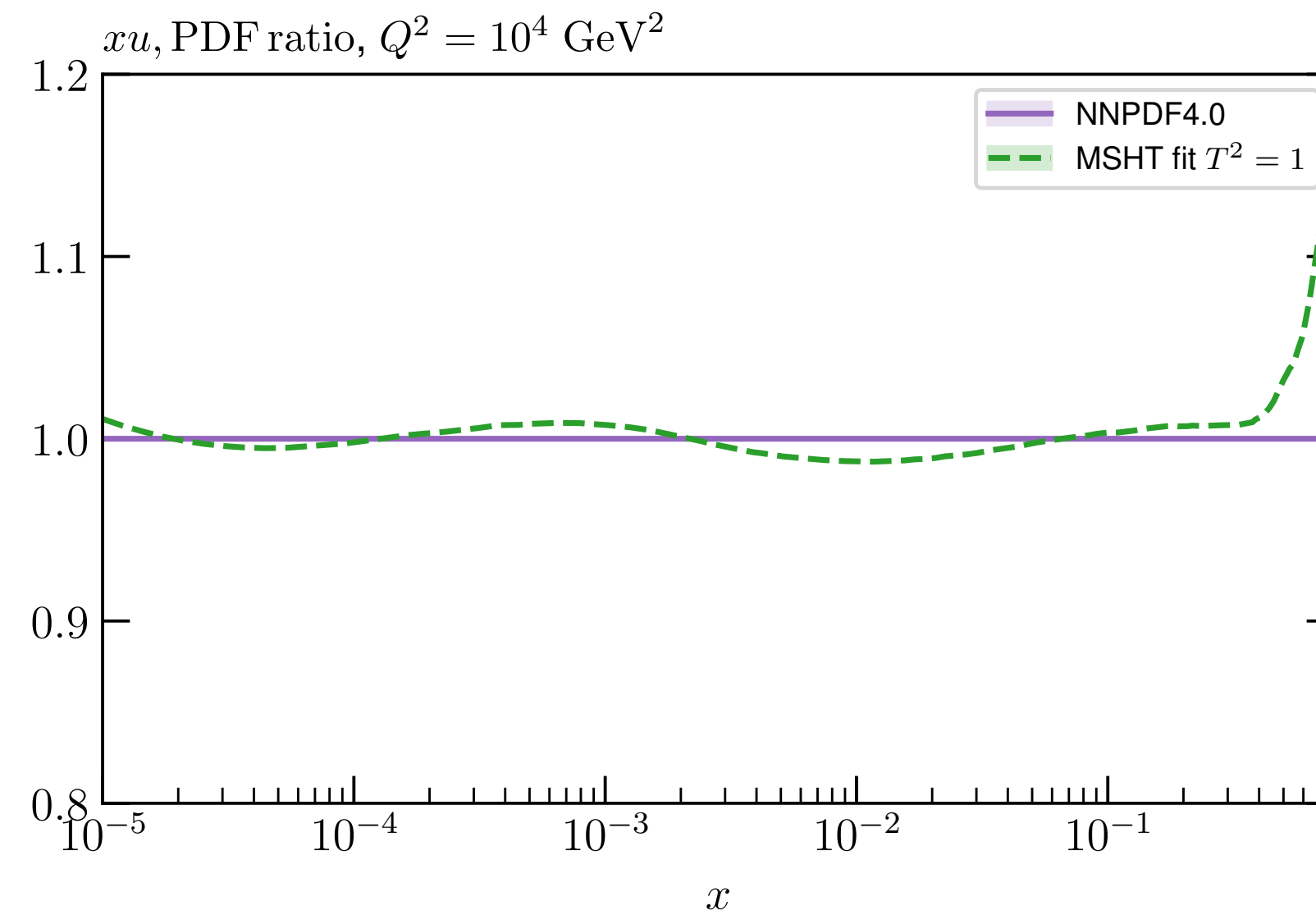
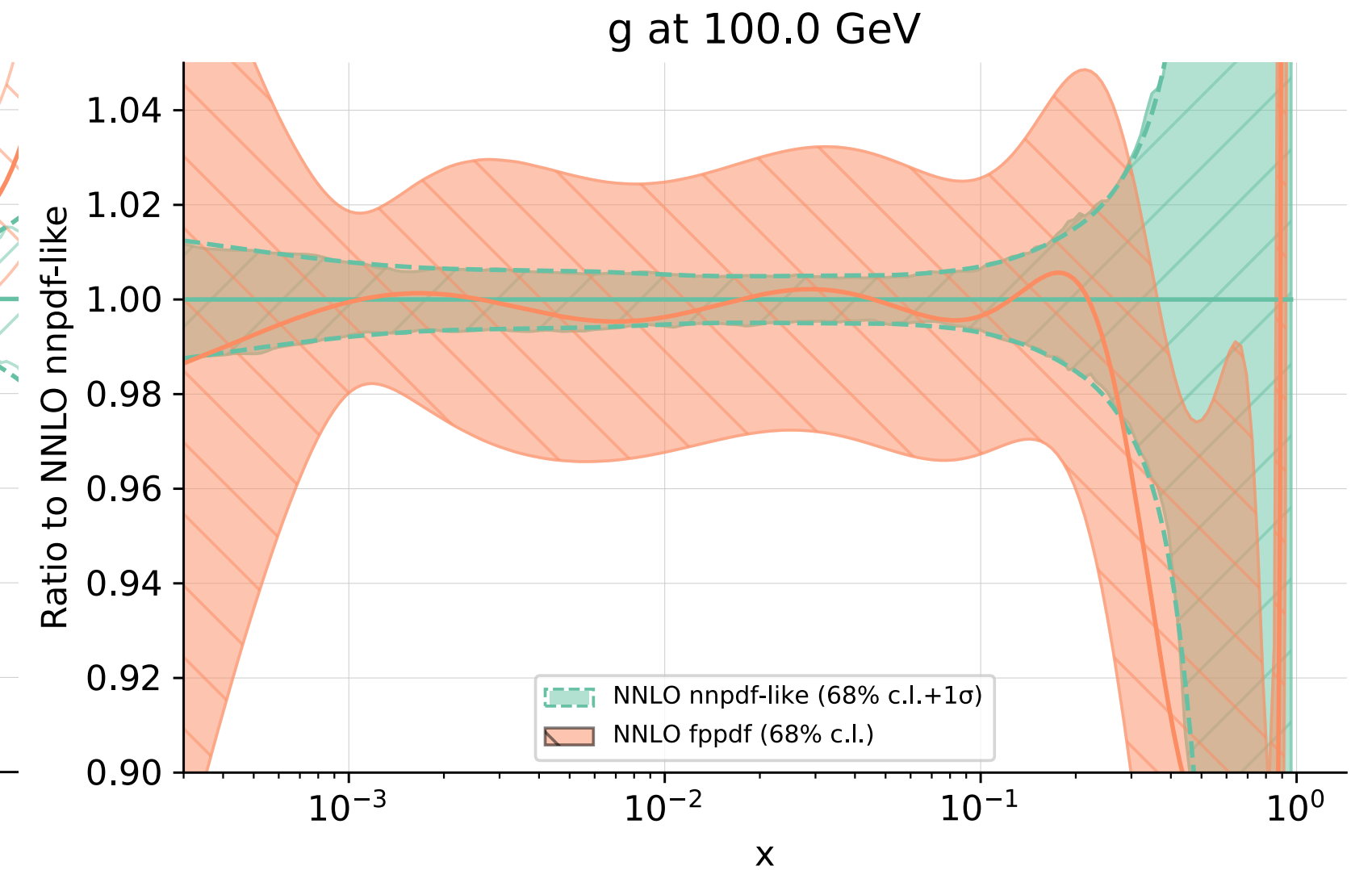
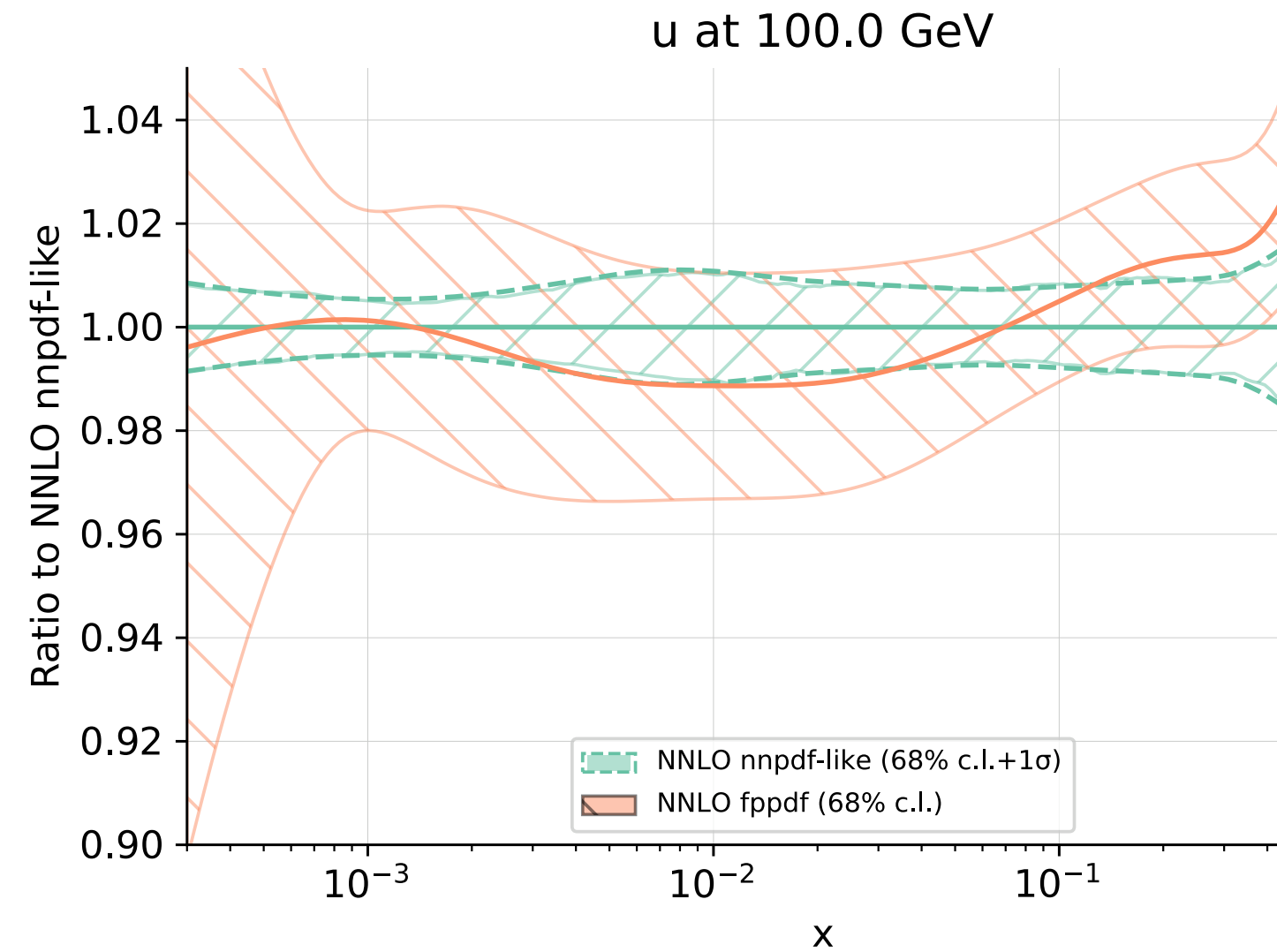


- But agreement not as good in some regions, particularly where constraints are lower. Genuine methodological differences here.
- May originate from freedom in space of PDF solutions, which different methodologies tend to lock onto.

Uncertainties

- **Elephant in room** - uncertainties.

- Despite good agreement in central values, these differ significantly.
- Almost entirely driven by the enlarged tolerance definition used by MSHT:



High x revisited

PDFs and high x

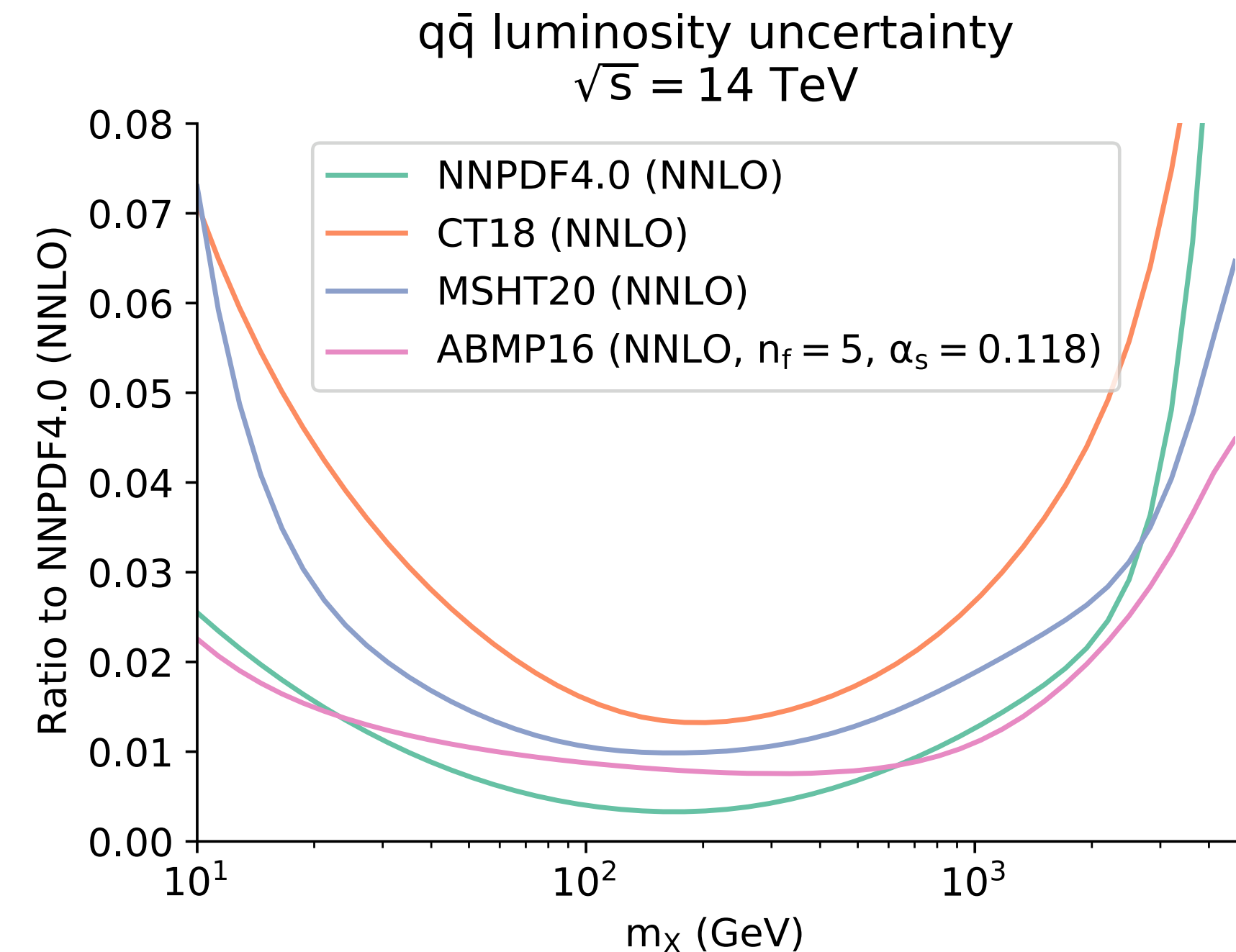
R. D. Ball et al., arXiv:2109.02653

- High x quark/antiquarks:
 - ★ Region of much interest for **BSM**...
 - ★ ...but uncertainties (+ methodology sensitivity) large!
- Constraints either come from:
 - ★ **Collider data** - high mass DY, jets etc...
 - ★ But statistics low and interplay with BSM non-trivial

J. Gao et al., arXiv:2211.020194

E. Hammou et al., arXiv:2307.10370

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



- ★ **Fixed target data.**
 - ★ Tends to be older - questions of data/error treatment.
 - ★ Non-perturbative (higher twist, target mass) corrections can be relevant.
 - Given increased PDF **precision** requirements, worth revisiting - study in preparation to reconsider this...

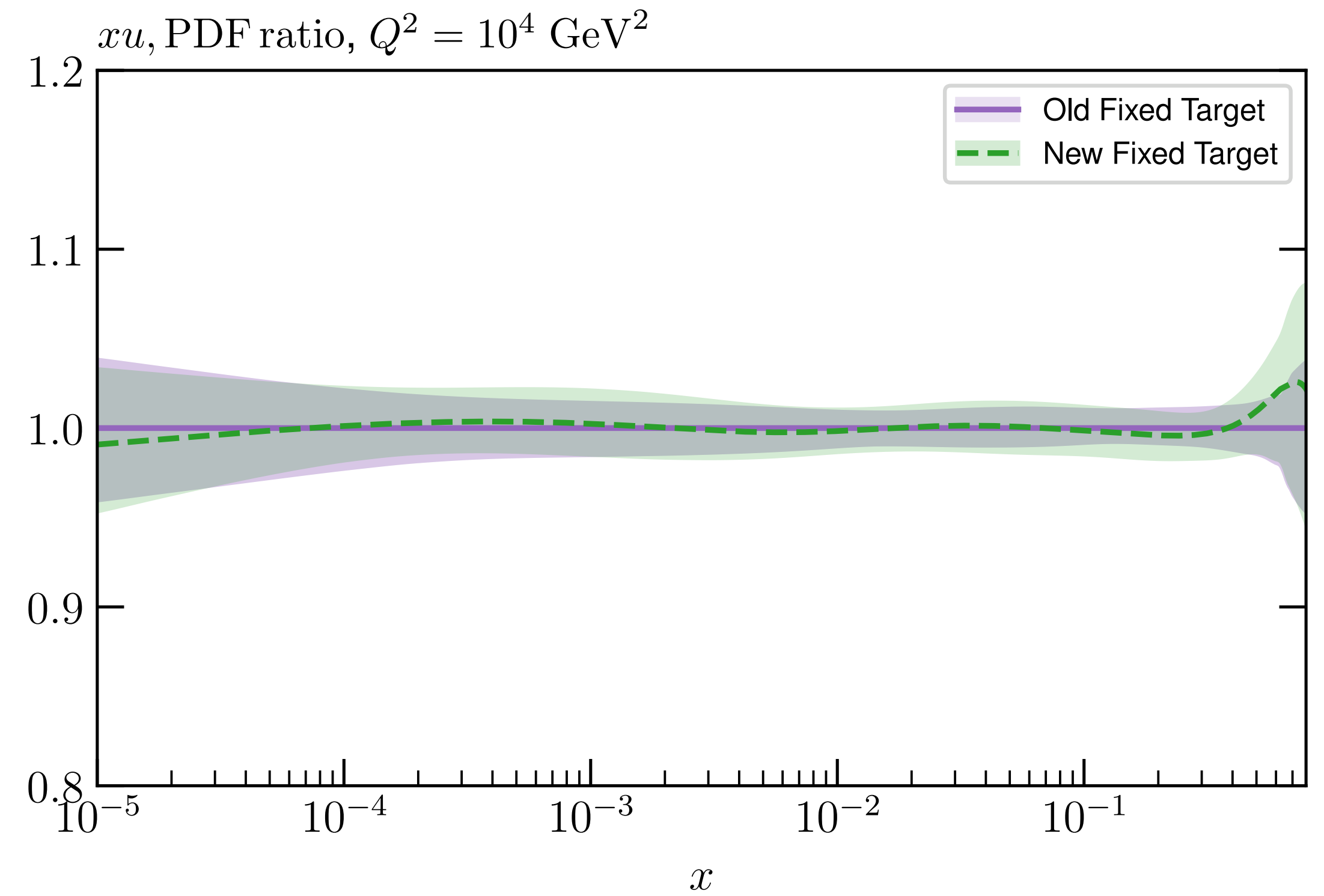
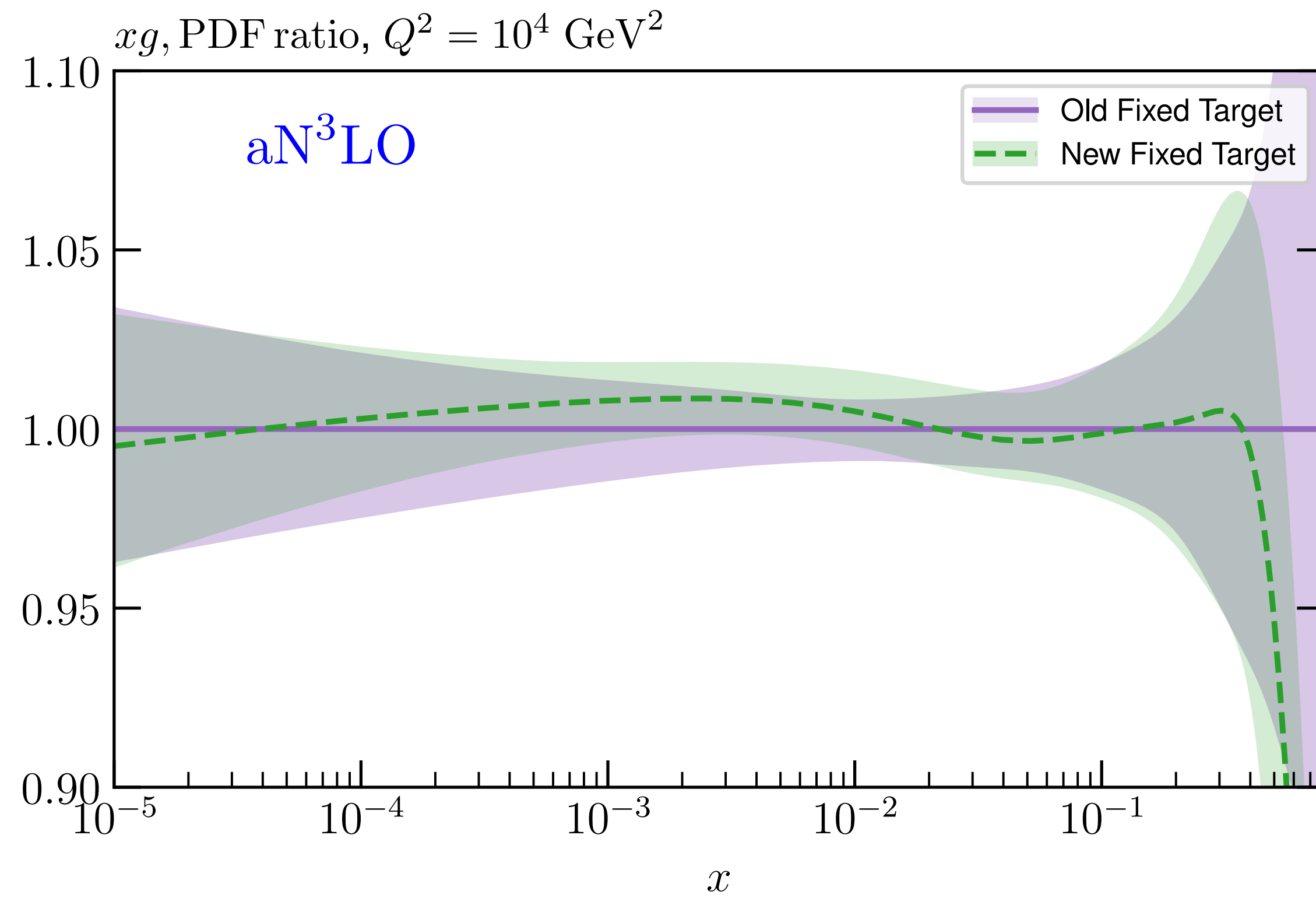
Fixed target data: updates

- Previous **MSHT** fits range of (BCDMS, NMC, E665) fixed target DIS tended to be taken averaged over beam energies and without full treatment of systematic correlations.
- Reasonable for older fits, but not for current high precision standard.
- Updating these leads to increase of ~ 500 datapoints (~ 4500 to 5000), increasing size of FT data in fit.

- Some **fit quality deterioration...**

	Old FT (aN ³ LO)	New FT (aN ³ LO)
Fixed Target	<u>1424.8 (0.96)</u>	<u>2201.1 (1.11)</u>
HERA	1593.3 (1.26)	1626.6 (1.29)
Hadron Collider	2391.9 (1.34)	2387.1 (1.33)
Global	<u>5410.0 (1.19)</u>	<u>6214.7 (1.23)</u>

- (Similar story at NNLO)



- ...but PDF impact **moderate**.
- Largest (shown) in gluon and high x up quark, as we would expect.

Target mass and higher twist corrections

- **Target mass corrections** - $O(m^2/Q^2)$ corrections to standard leading twist calculation of SFs. Calculated using known results.

$$F_2^{\text{TMC}}(x, Q^2) \approx \frac{x^2}{\xi^2 r^3} F_2^{(0)}(\xi, Q^2) \left[1 + \frac{6x\mu\xi}{r} (1 - \xi)^2 \right] ,$$

$$\xi = \frac{2x}{1 + \sqrt{1 + 4x^2 M^2/Q^2}} .$$

- **Higher twist** - more general class of $O(1/Q^2)$ correction. Models exist, but we choose to parameterise this freely in current study.

$$F_2^{\text{HT}}(x, Q^2) = F_2^{\text{LT}}(x, Q^2) \left(1 + \frac{D_i(x)}{Q^2} \right) ,$$

- In standard (baseline) MSHT fits cuts are imposed to reduce impact of these, but can investigate impact of reducing W^2 cut in particular, as well as above effects for baseline case.

$$Q^2 > 2 \text{ GeV}^2 \text{ and } W^2 > 15 \text{ GeV}^2$$

TMCs

- Consider including TMCs alone. Negligible with baseline cuts, but consider lowering it:

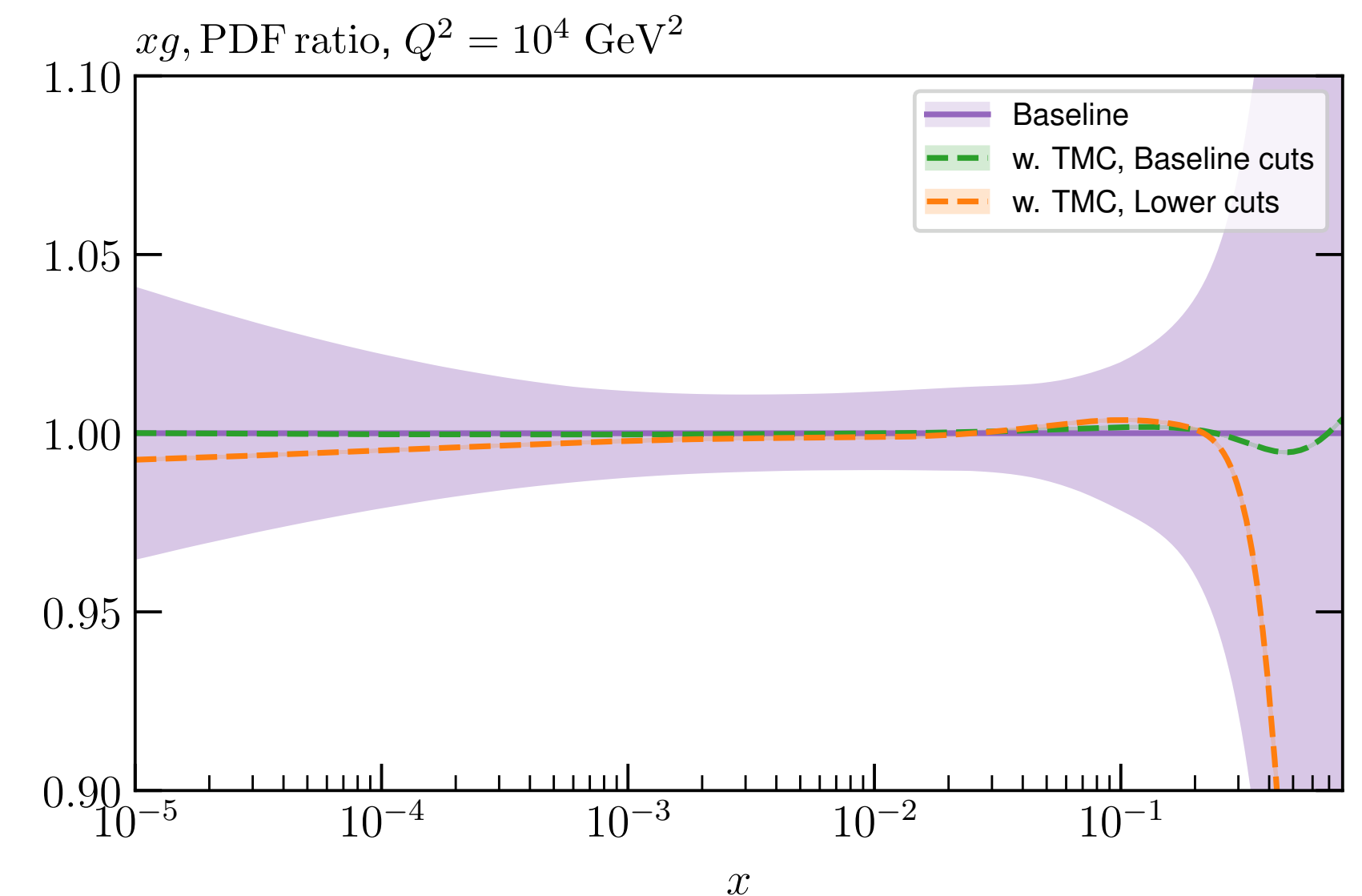
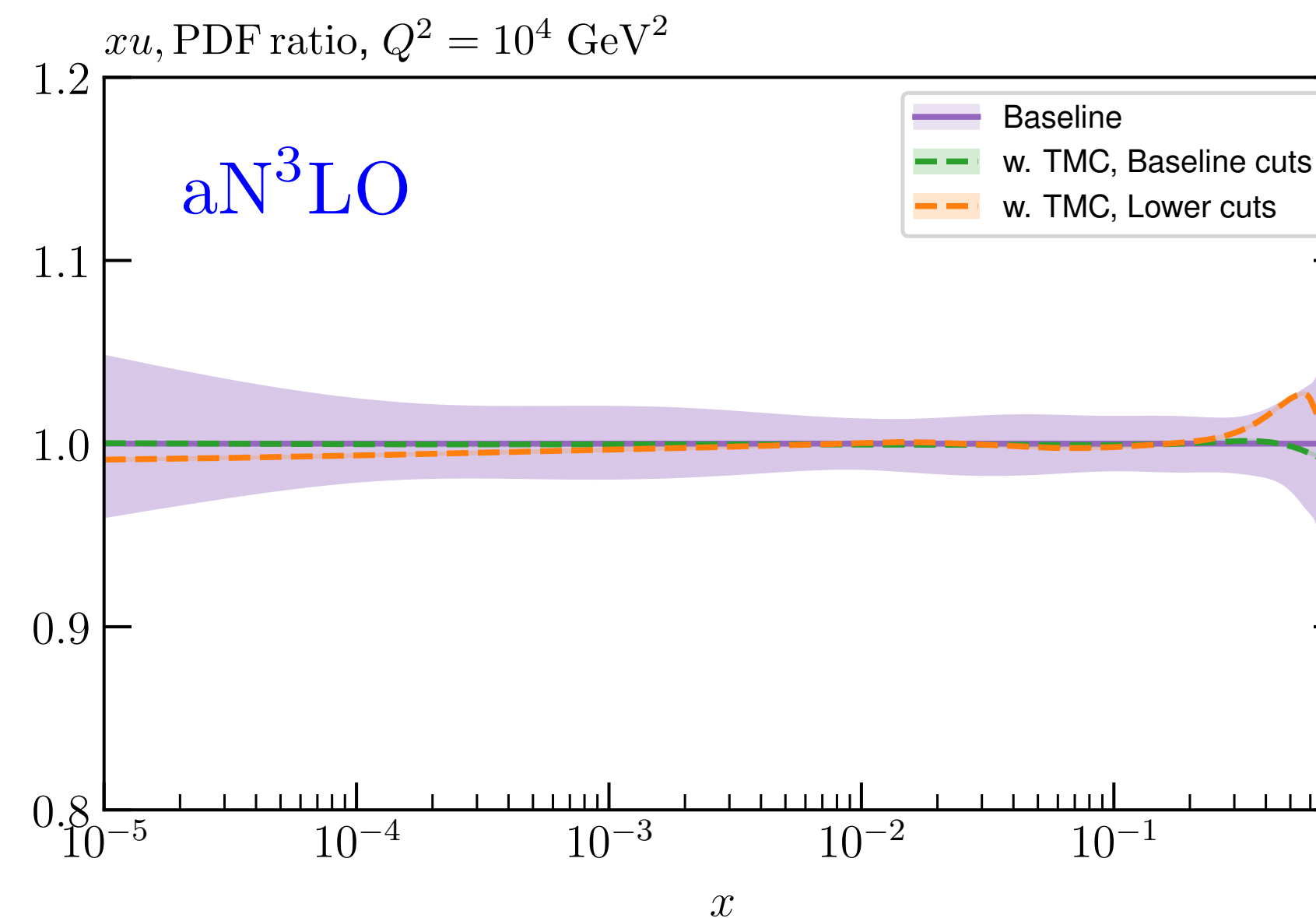
$$W^2 > 15 \text{ GeV}^2$$

$$W^2 > 5 \text{ GeV}^2$$

	Baseline cuts, no TMC	no TMC, lower W^2 cut	w. TMC, lower W^2 cut
Fixed Target	2201.1 (1.11)	2590.7 (1.15)	2511.1 (1.12)
Global	6214.7 (1.23)	6646.4 (1.26)	6526.6.4 (1.23)

- TMCs** certainly **needed** in this case - once included give comparable fit quality to baseline!

- PDF impact very mild. Nice **stability**.
- Lower cut has ~ 250 more points, so not guaranteed!



Higher twist

$$F_2^{\text{HT}}(x, Q^2) = F_2^{\text{LT}}(x, Q^2) \left(1 + \frac{D_i(x)}{Q^2} \right),$$

- Including these (freely parameterised) gives **improvement** in **fit quality** for lower W^2 cut:

$$W^2 > 5 \text{ GeV}^2$$

w. TMC, lower W^2 cut	Lower W^2 (w. HT, w. TMC)
6526.64 (1.23)	6469.8 (1.22)

$$W^2 > 15 \text{ GeV}^2$$

- ...but also for **baseline** case!

Baseline cuts (no HT, no TMC)	Baseline cuts (w. HT, w. TMC)
6214.7 (1.23)	6121.7 (1.21)

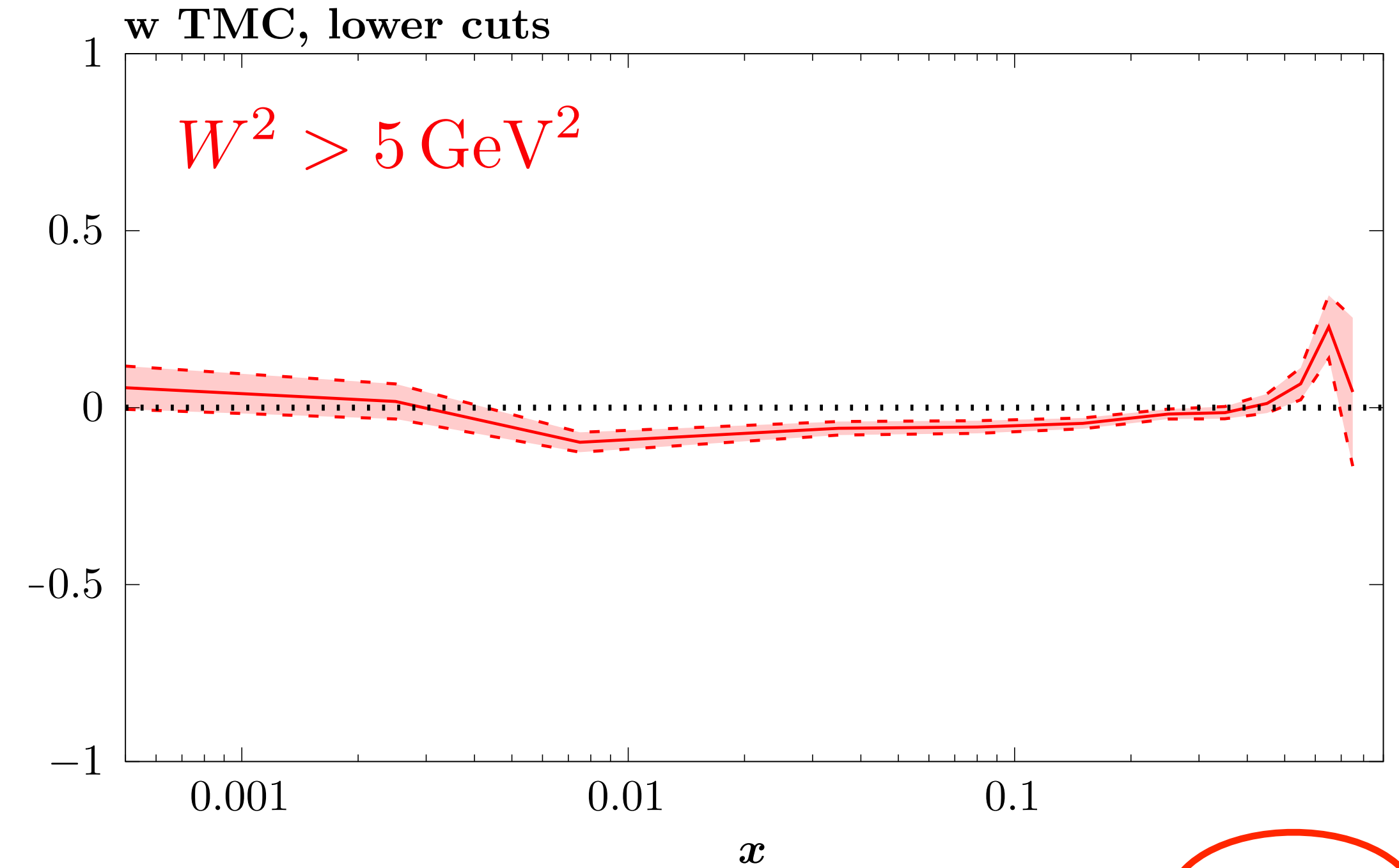
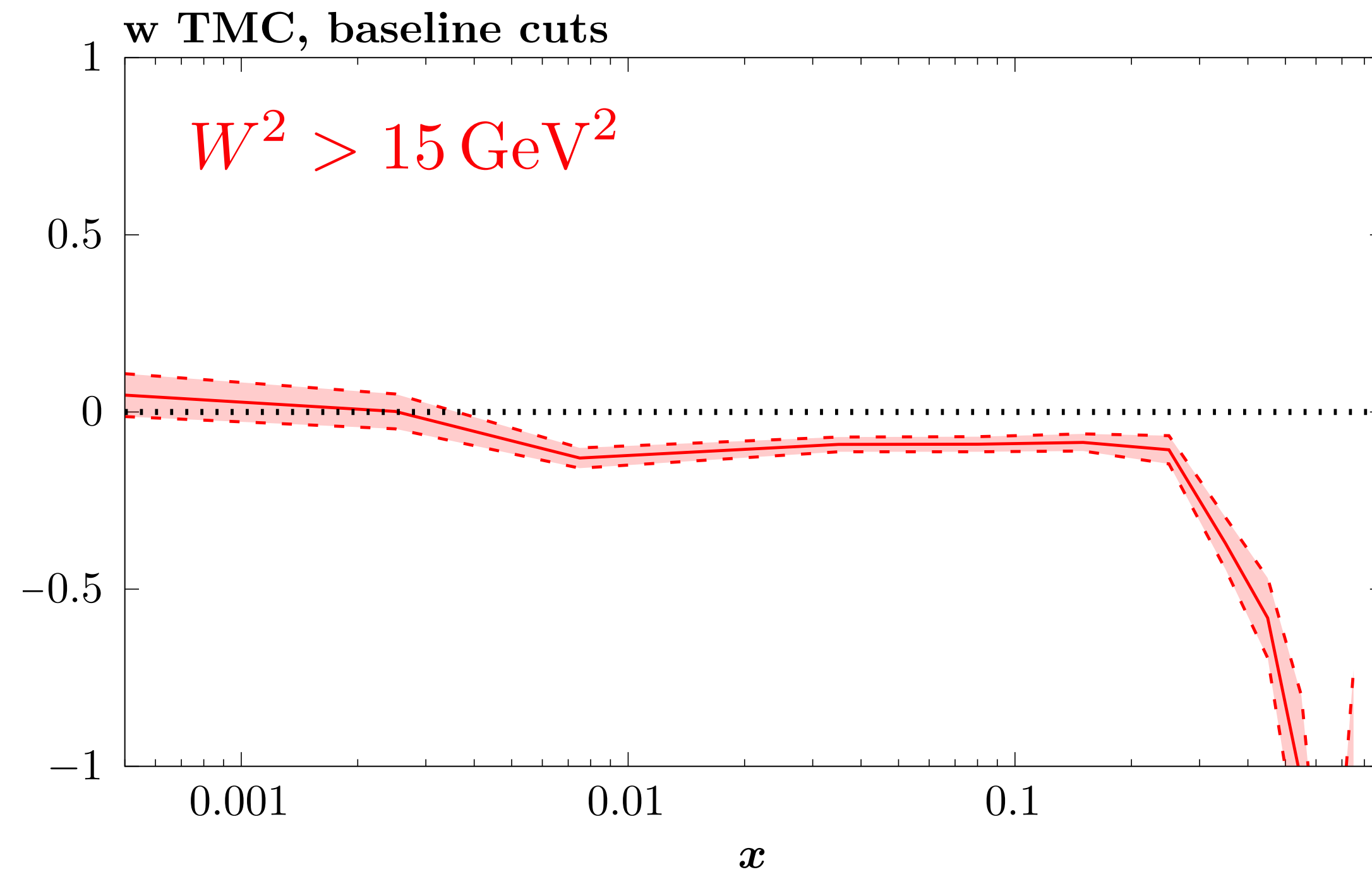
$$Q^2 > 2 \text{ GeV}^2 \text{ and } W^2 > 15 \text{ GeV}^2$$

- Is baseline fit sensitive to HT? Fixing HTs to lower W^2 case gives milder improvement:

Baseline cuts (w. HT lower, w. TMC)
6175.5 (1.23)

- While taking HTs from baseline fit and using in lower W^2 fit gives **terrible** fit quality ~ 1.55 .

- Fit qualities indicate difference in preferred HT corrections for two cut choices, and indeed this is seen...



- What is going on?

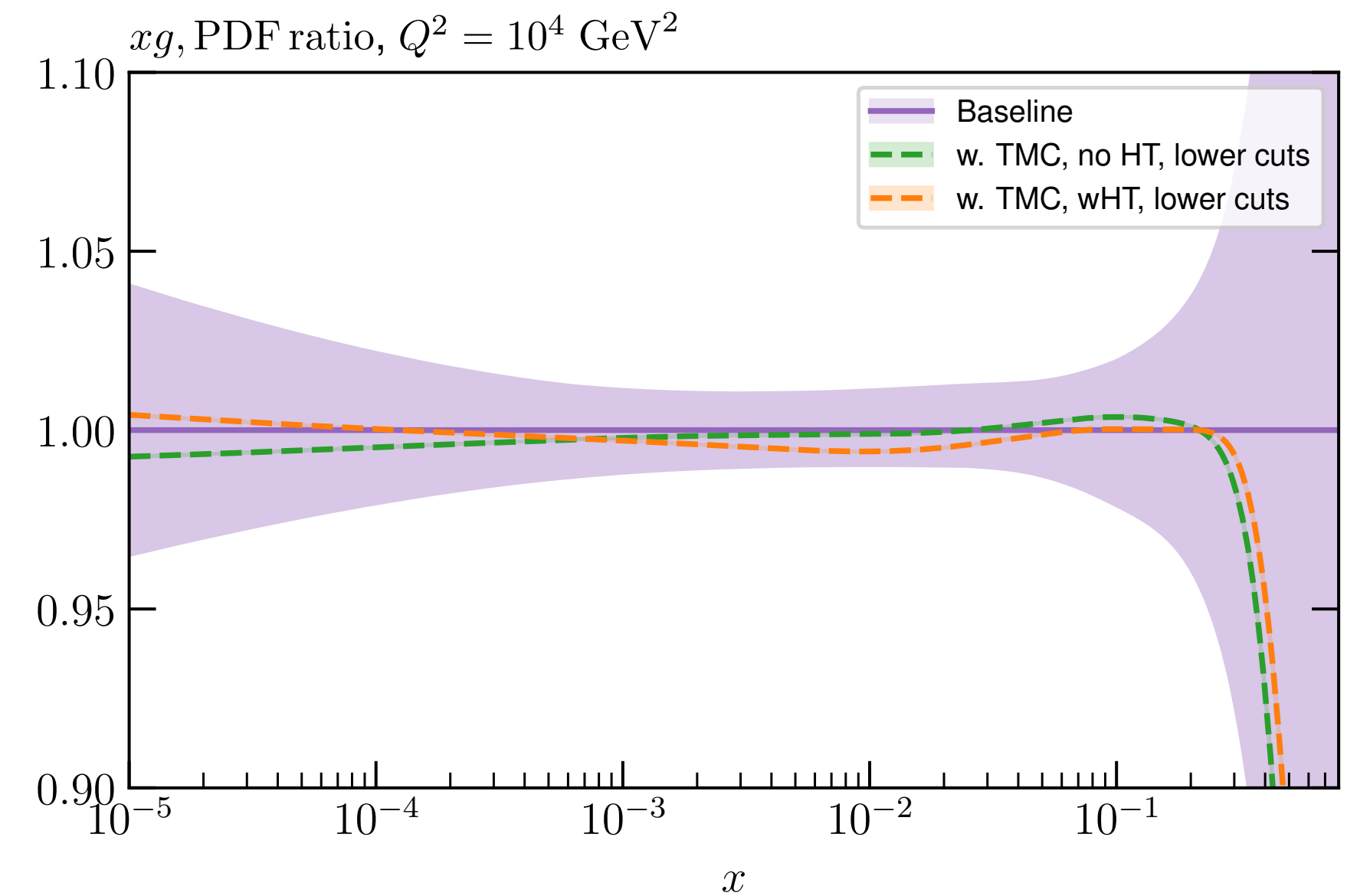
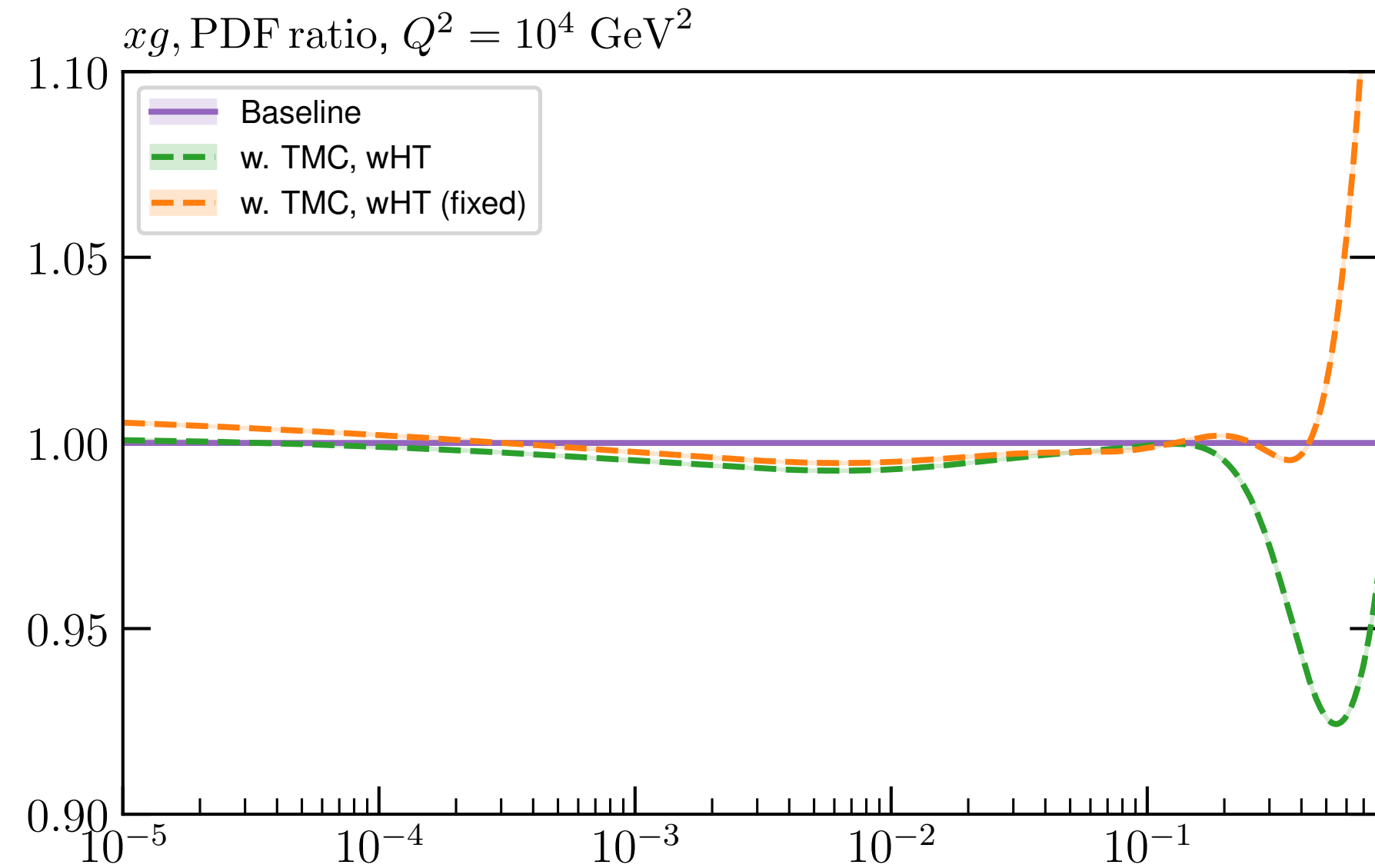
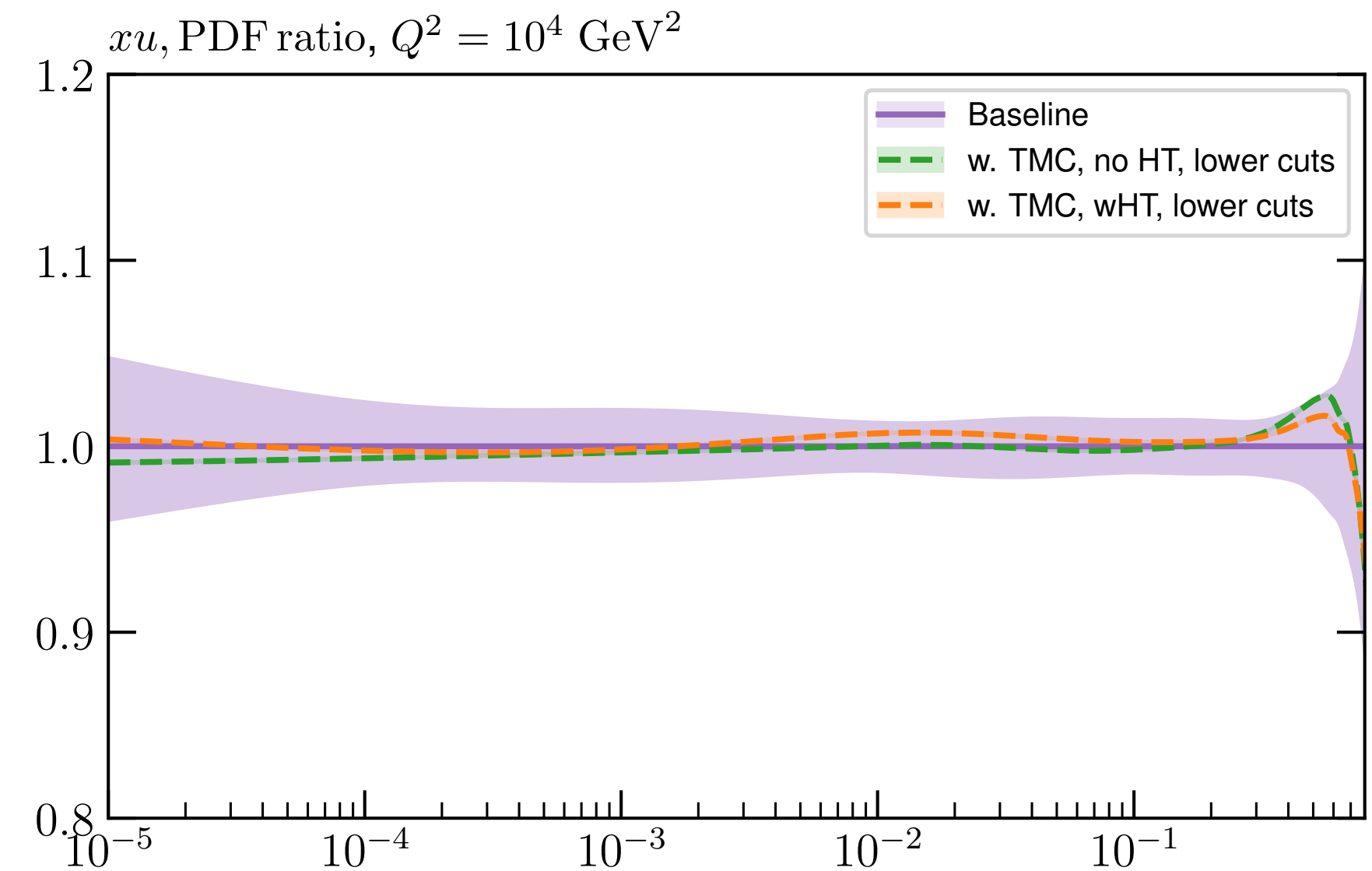
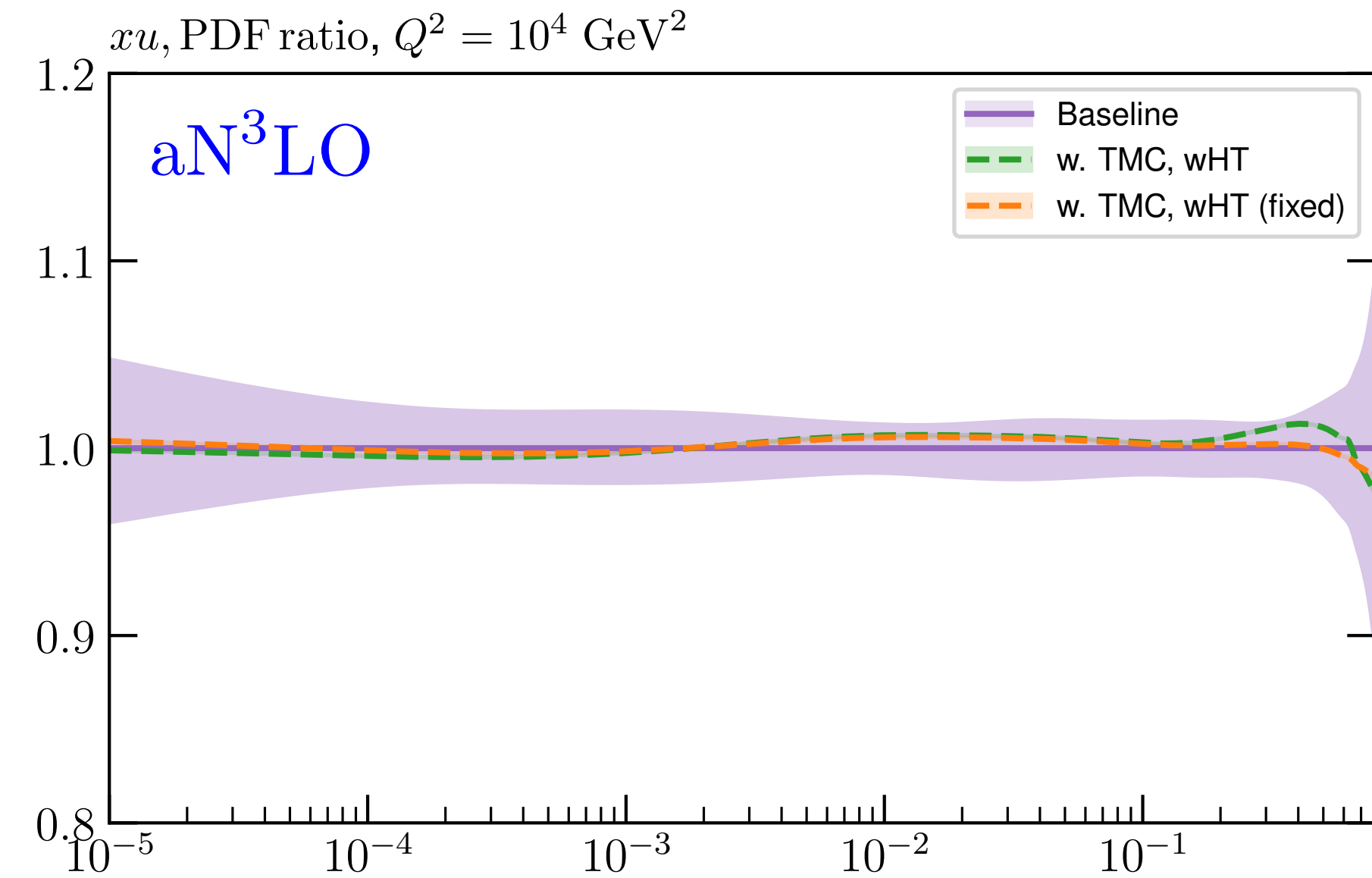
$$F_2^{\text{HT}}(x, Q^2) = F_2^{\text{LT}}(x, Q^2) \left(1 + \frac{D_i(x)}{Q^2} \right),$$

- Indication that 'higher twist' corrections in baseline case fit ~ same order as QCD corrections, i.e. mimicking **missing higher order corrections**.
- Not so in lower cut case.
- Care needed** in interpreting such contributions.

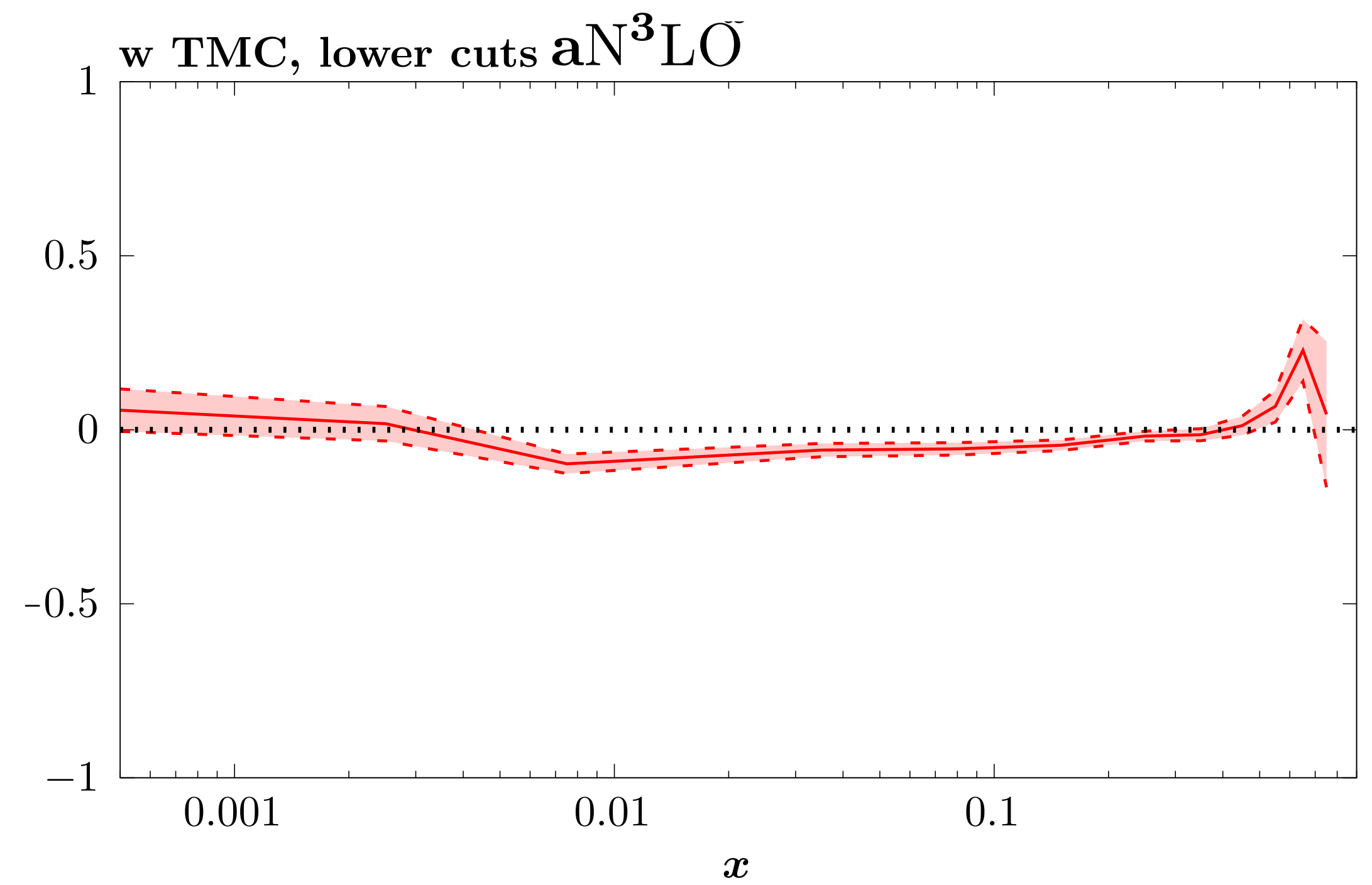
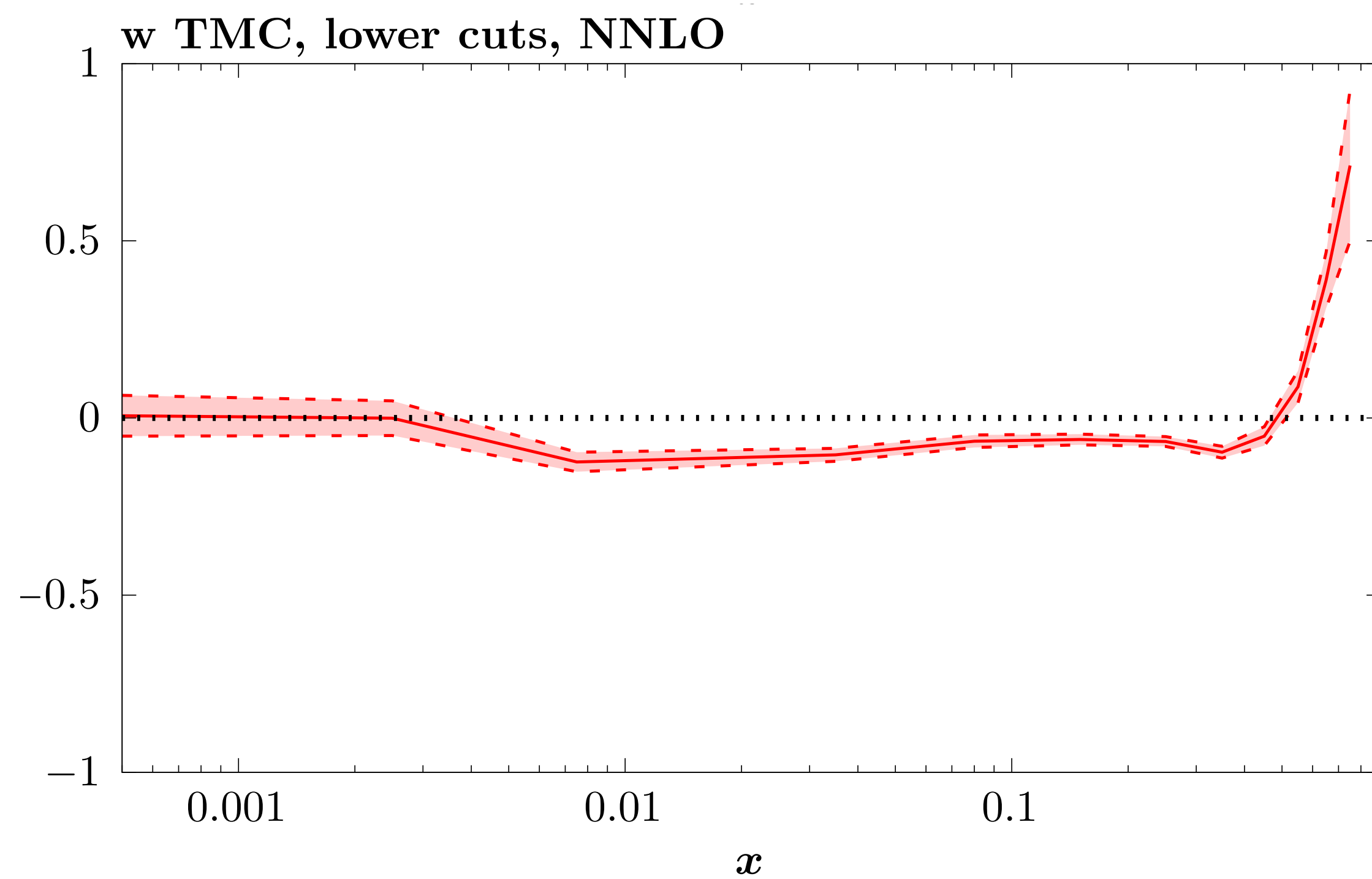
$$W^2 > 15 \text{ GeV}^2$$

PDFs

$$W^2 > 5 \text{ GeV}^2$$



- But in the end PDF impact rather moderate - nice **stability** though **some reduction in** ggH region.



- Compare **NNLO** and **aN³LO** fits - size of HT corrections decreases with order (in line with earlier observations).

Summary and Outlook

- ★ MSHT busily working on updates towards next release: aN³LO, new data, errors on errors...
- ★ Beyond that benchmarking and understanding methodological differences will be key.

Thank *you* for listening (and
goodbye)!

Backup

EIC pseudodata

Pseudodata Study

N. Armesto et al., arXiv:2309.11269

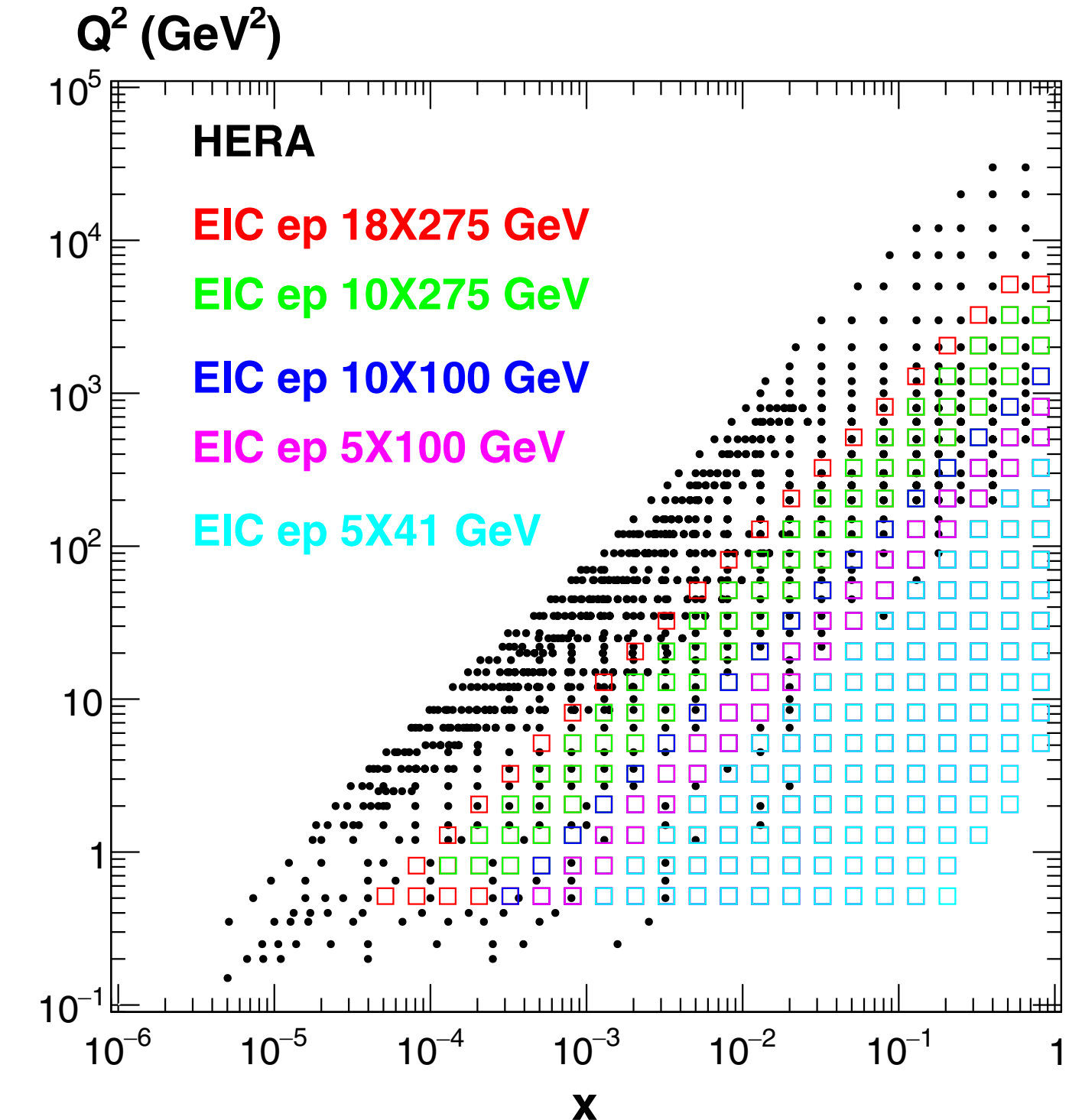
- Previous collaborative study: generate pseudodata (PD) with updated beam energies, configurations, lumis and uncertainty projections.
- Two (low/high) acceptance cases considered.
- Kinematic coverage (including MSHT cuts)

$$Q^2 > 2 \text{ GeV}^2, 0.01 < y < 0.95, \\ W^2 > 15 \text{ GeV}^2.$$

- Including sensible projections for main uncertainty sources, with both standard and aggressive scenarios:

- ★ **Standard:** 1% uncorrelated, 3.4% correlated
- ★ **Aggressive:** 0.5% uncorrelated, 0.5% correlated

- Impact on PDFs studied previously - moderate but complementary. Consider **strong coupling** here...



Determining the strong coupling

- $\alpha_S(M_Z^2)$ sensitivity in global PDF fit come from:

- ▶ Direct $\alpha_S(M_Z^2)$ dependence in coefficient functions.

$$C(\alpha_S) = \alpha_S^l [C_0 + \alpha_S C_1 + \alpha_S^2 C_2 + \alpha_S^3 C_3 + \dots]$$

- ▶ Indirect $\alpha_S(M_Z^2)$ dependence through PDF evolution.

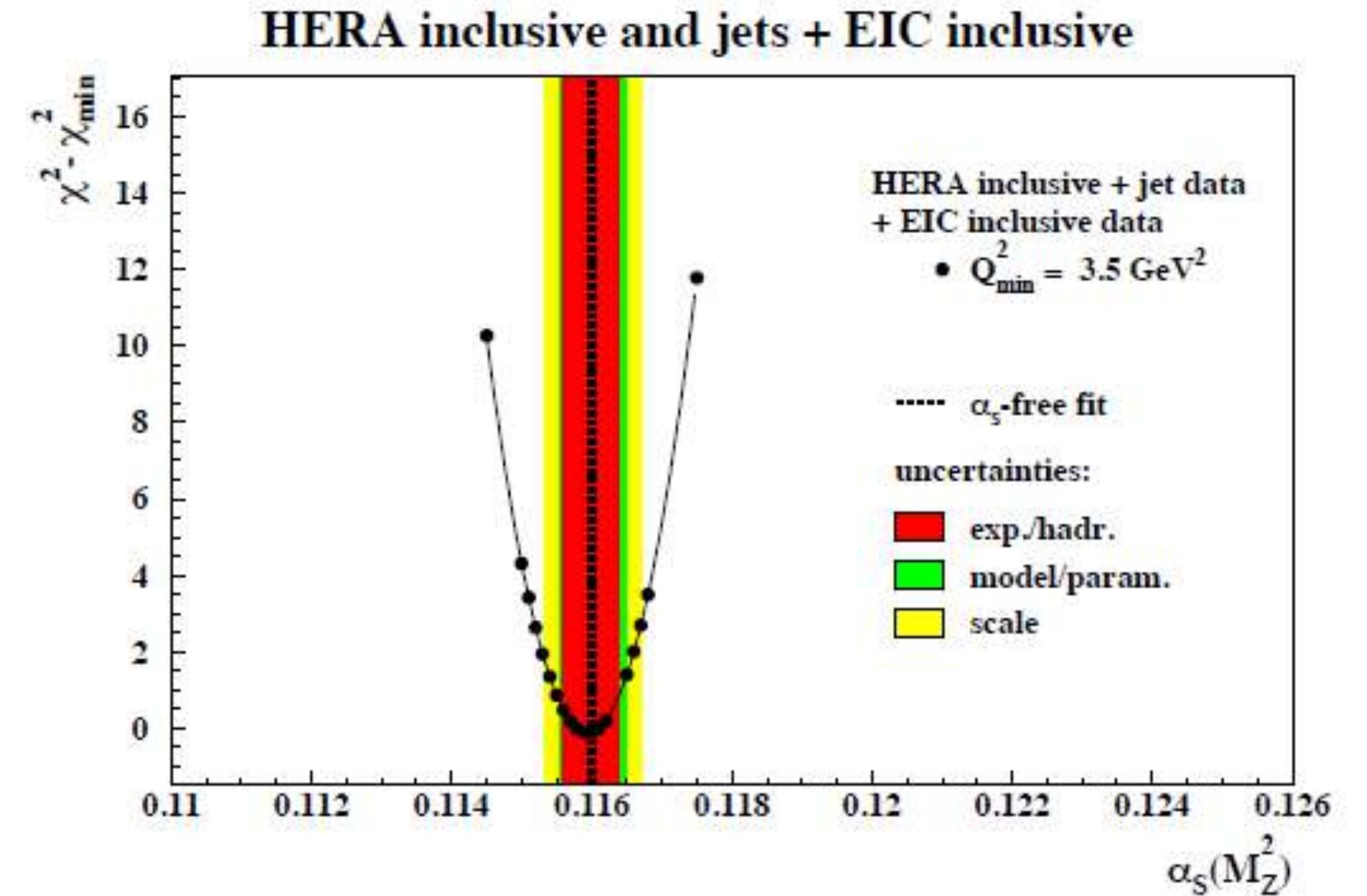
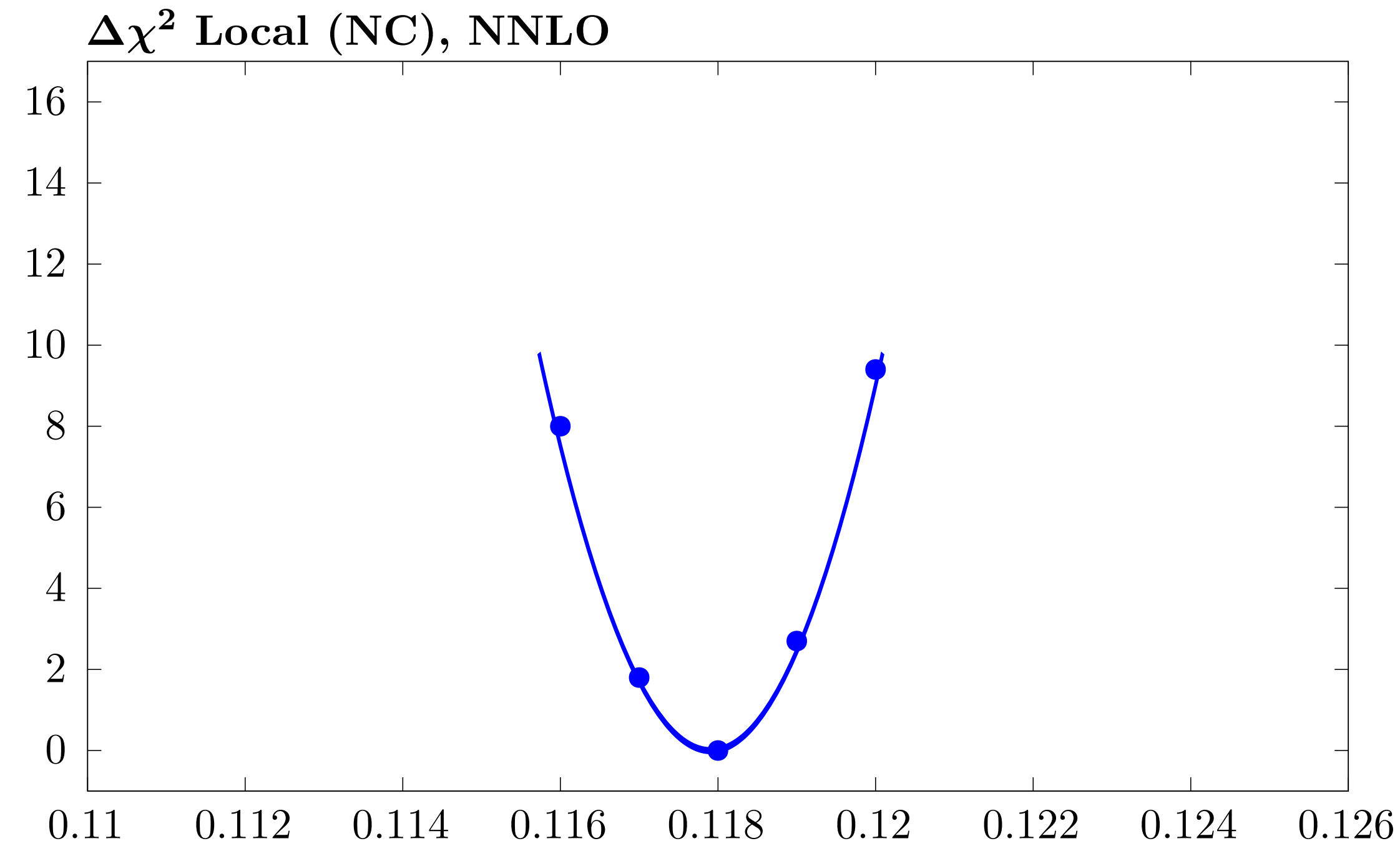
$$\frac{df}{d \log \mu_F^2} = \begin{bmatrix} P_{qq} & n_f P_{qg} \\ P_{gq} & P_{gg} \end{bmatrix} \begin{bmatrix} \Sigma \\ g \end{bmatrix}$$

- DIS has limited sensitivity indirectly via scaling violations.
- HERA at low/intermediate x driven by gluon splitting, hard to disentangle α_S .
- EIC at higher x driven by non-singlet splitting, so α_S less correlated to g .
- Improved precision + more datapoints on structure function evolution.

★ So what do we find?

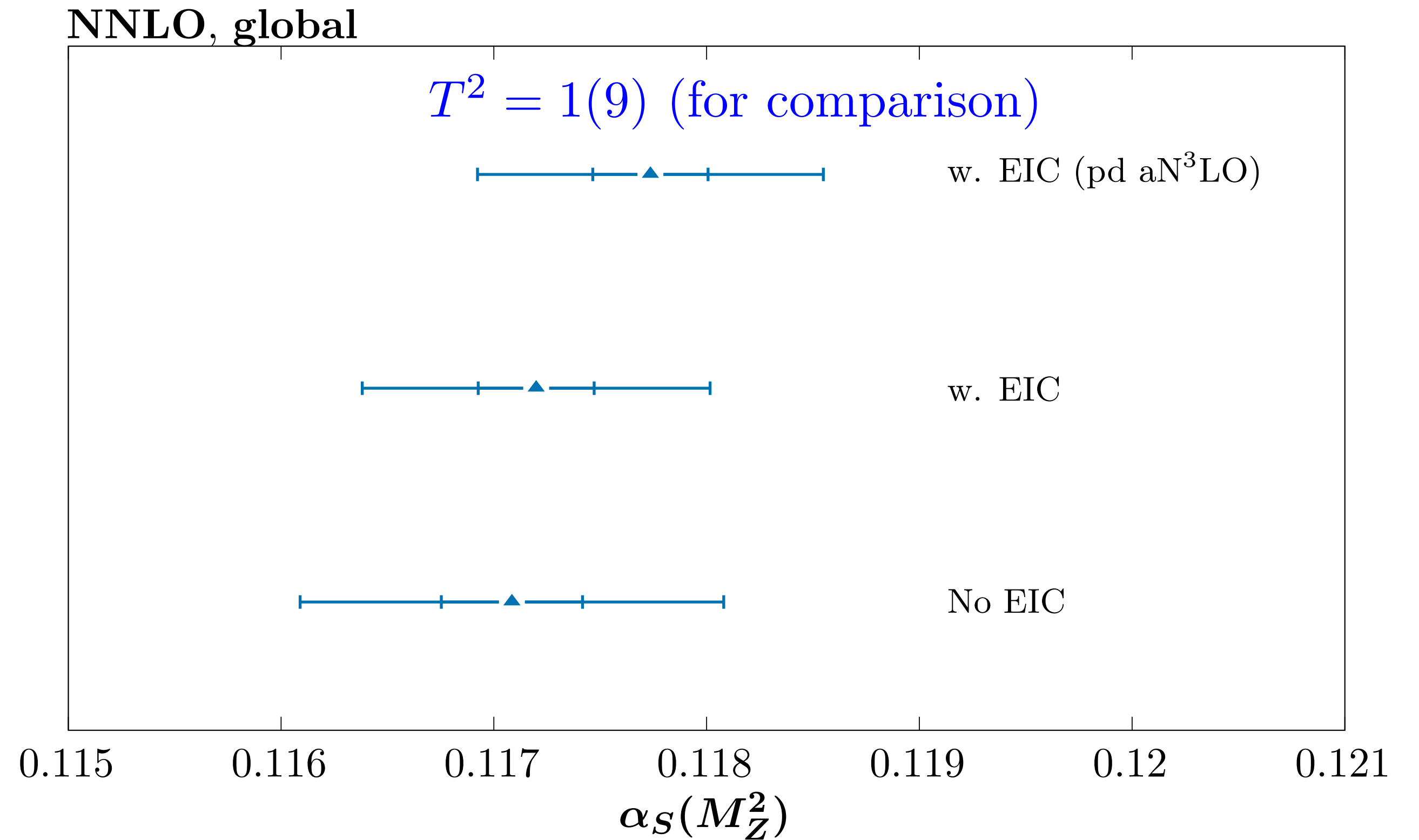
LHL et al., in prep.

NNLO Fit



- Looking at local profile - prefers $\alpha_s(M_Z^2) = 0.118$ by construction. Shape + constraining power is in fact very similar to previous HERA + EIC study.
- But of course now it is the global impact that is relevant...

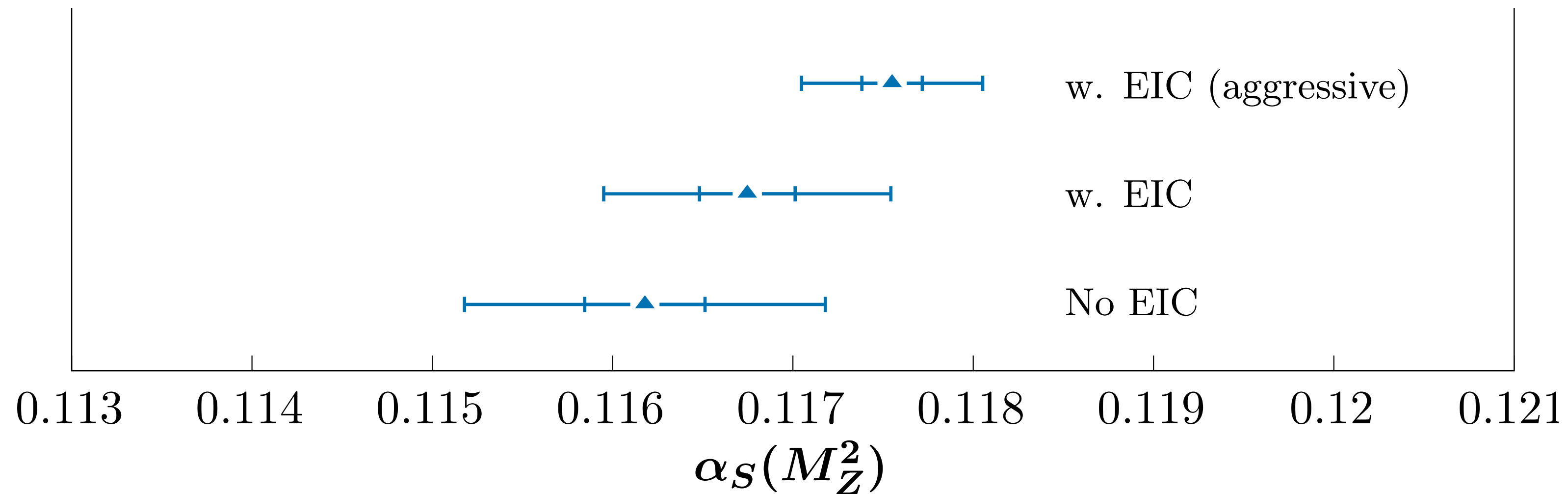
- What about **global** impact? Additional constraint clearly visible.
- Note precise impact depends on preferred previous value (~ 0.117) and that used in pseudodata generation (0.118).



- However, care needed here! Impact of this high precision EIC pseudodata in principle significant, but need to have accuracy in mind as well.
- For example, impact of finite QCD (NNLO) order - to assess generate PD at aN³LO (with MSHTaN³LO set) but include in NNLO fit. Provides estimate of **missing higher orders**.
- Even though PD generated with $\alpha_S(M_Z^2) = 0.118$ mismatch leads to some increase in best fit value.

aN3LO Fits

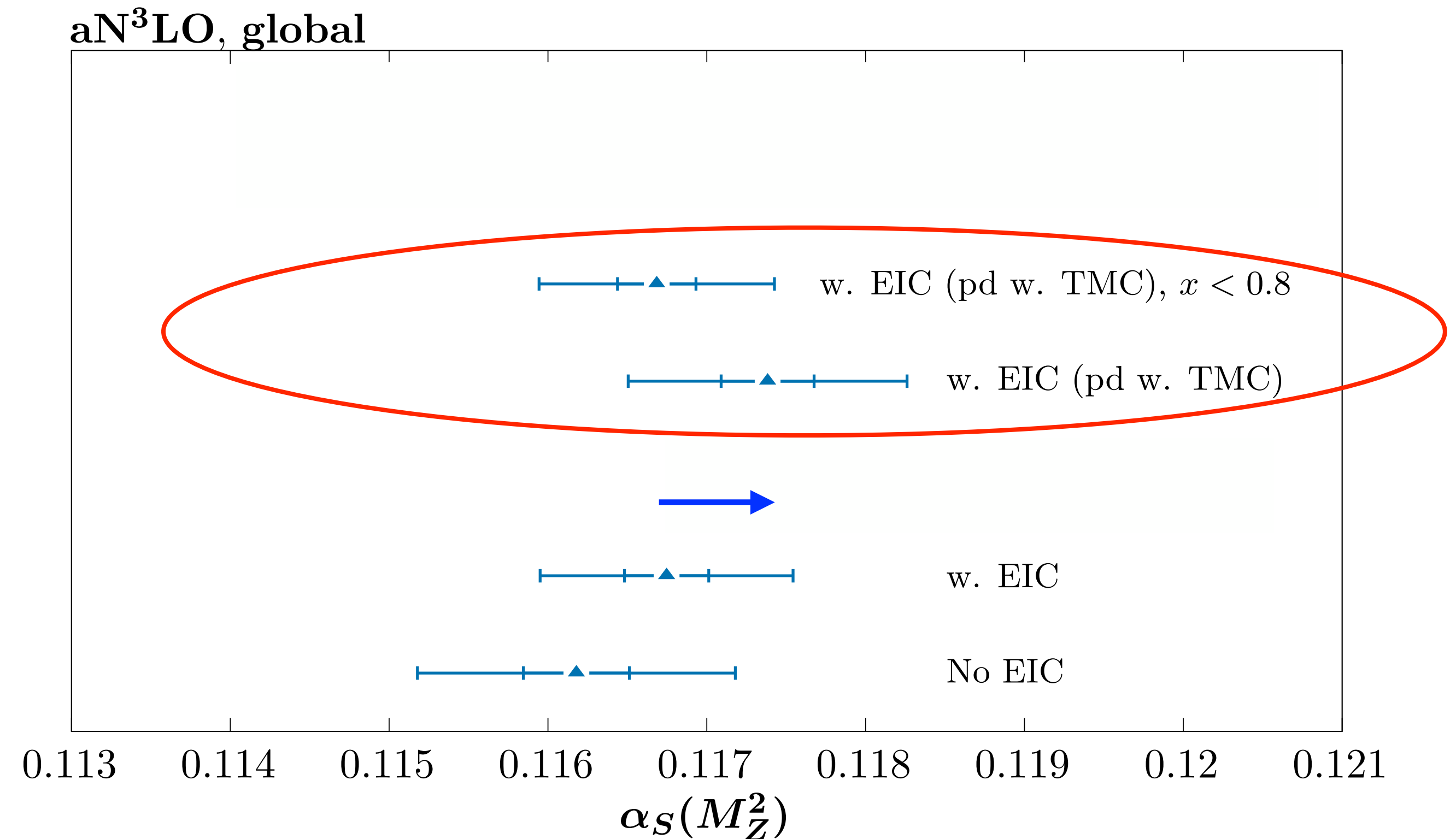
$T^2 = 1(9)$ (for comparison)



- Move to **aN3LO** - generate PD consistently, consider both ‘standard’ and ‘optimistic’ scenarios.
- Impact on global best fit and uncertainties is clear, and particularly dramatic in optimistic case!
- But is this the end of the story?

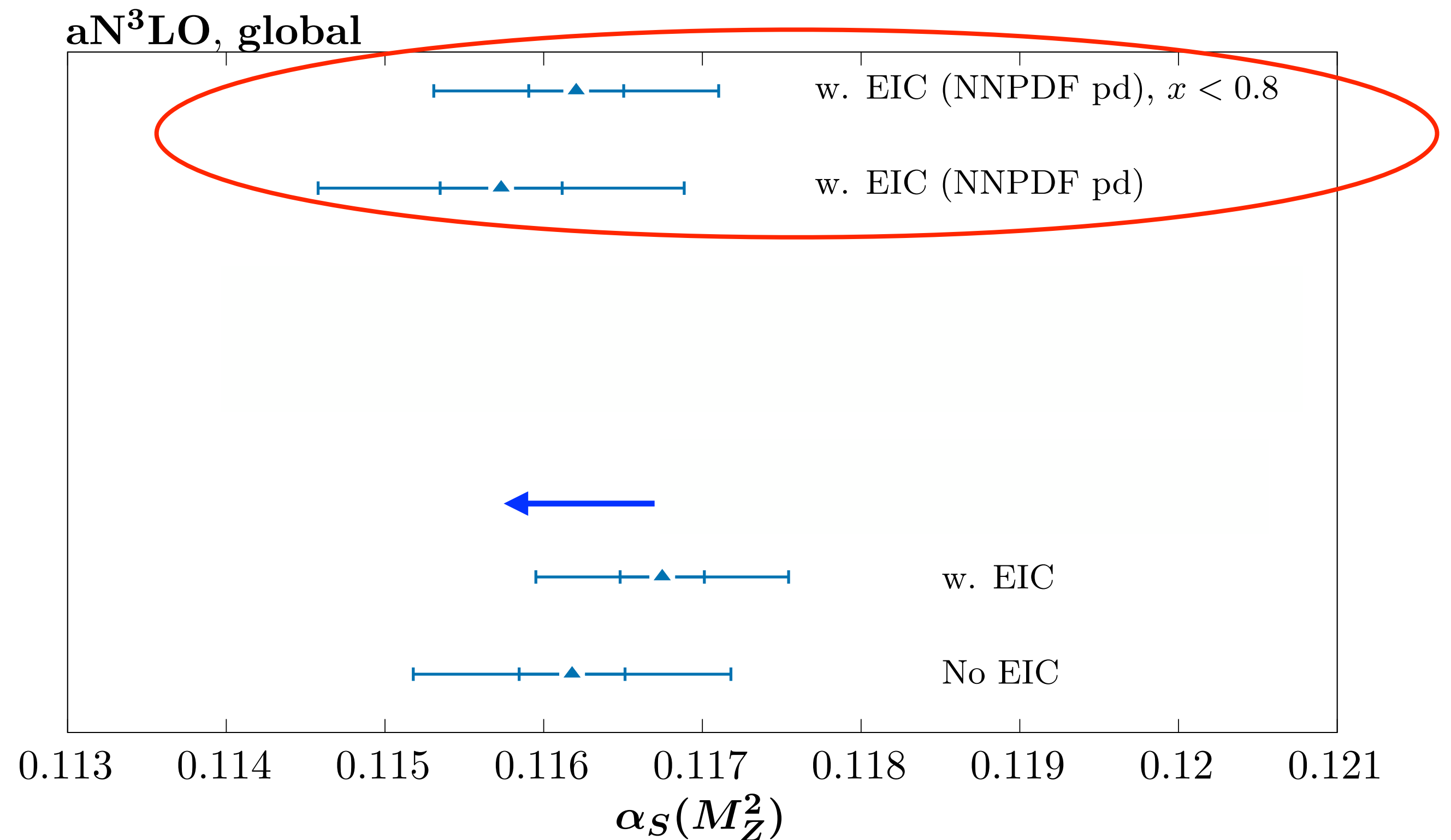
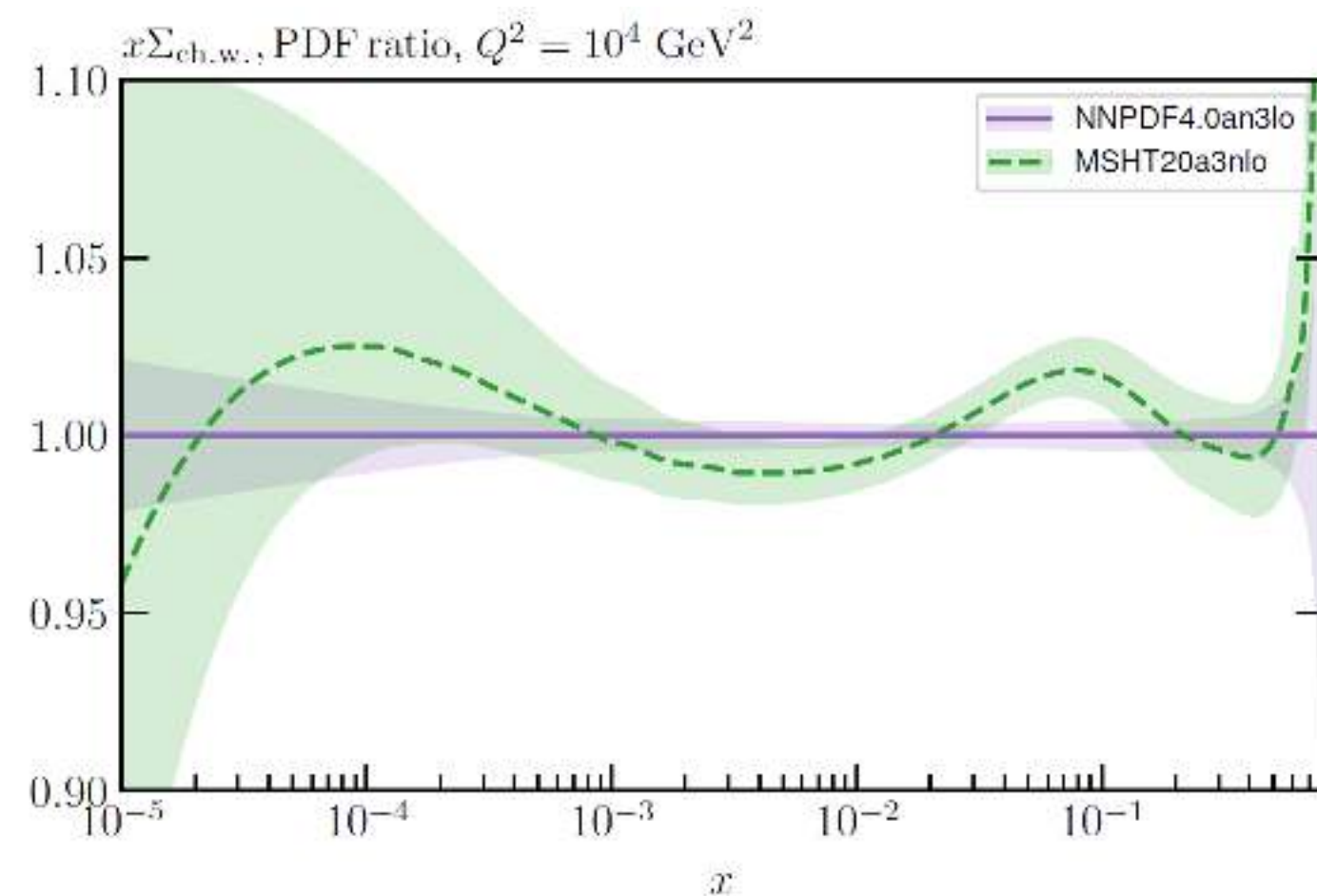
- What about other sources of **uncertainty**? Try two things:

- ★ Generate PD with **target mass corrections**, and fit without.
- ★ This inconsistency induces upwards shift in strong coupling.
- ★ Clearly should include (and known) but indicates broader issue here.
- ★ Size of shift at $T \sim 1$ level - well within enlarged (dynamic tolerance).

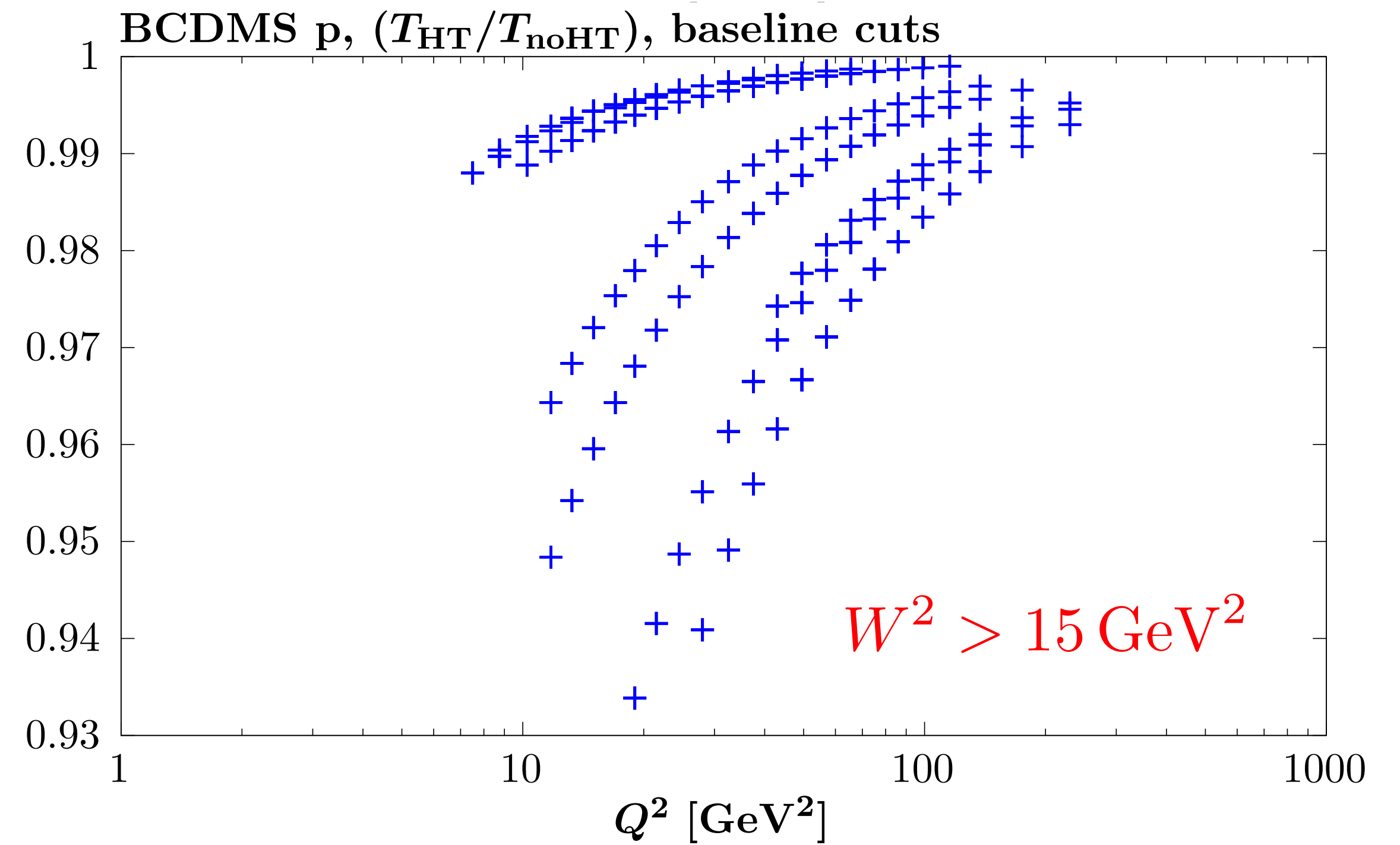
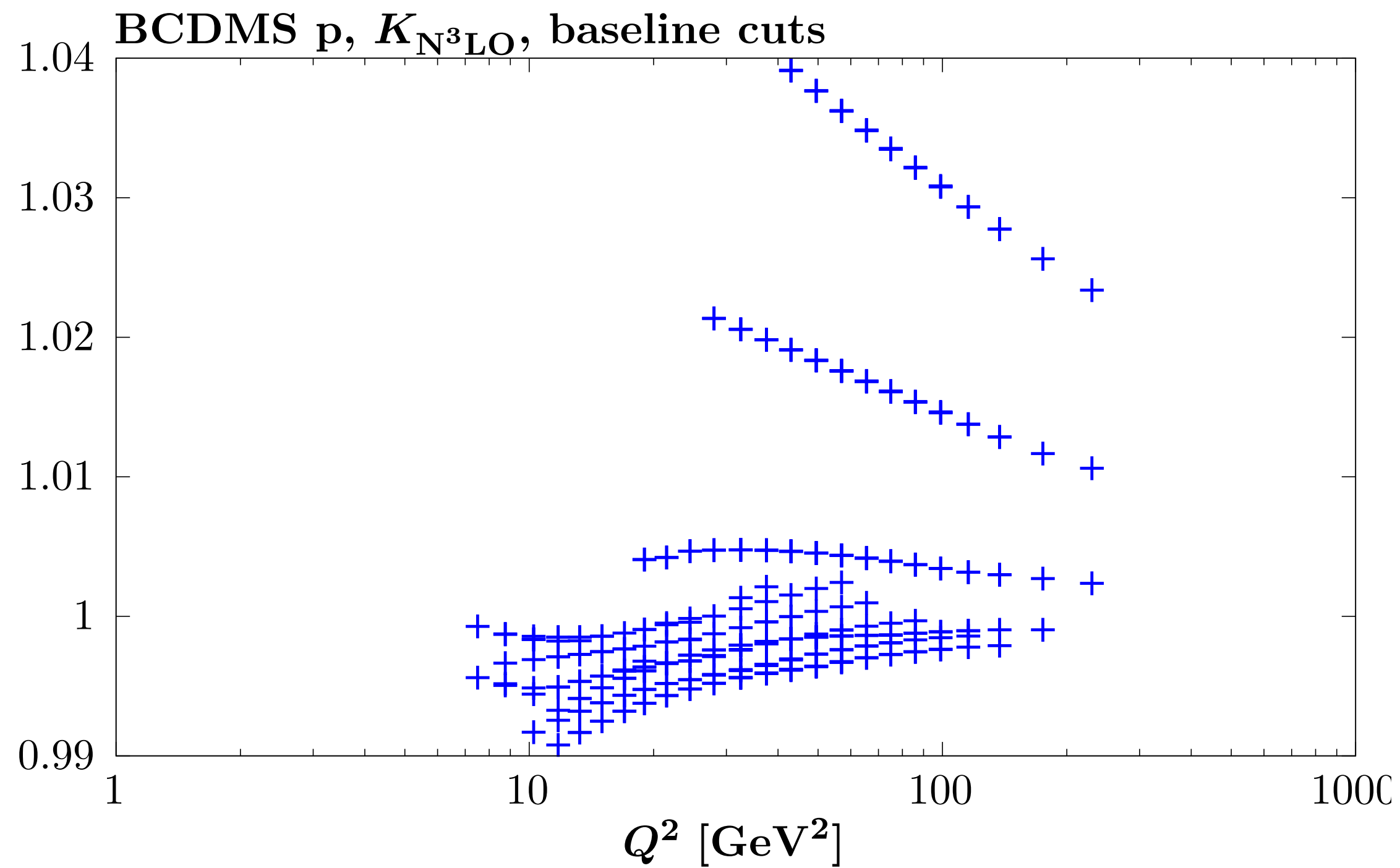


- What about other sources of uncertainty? Try two things:

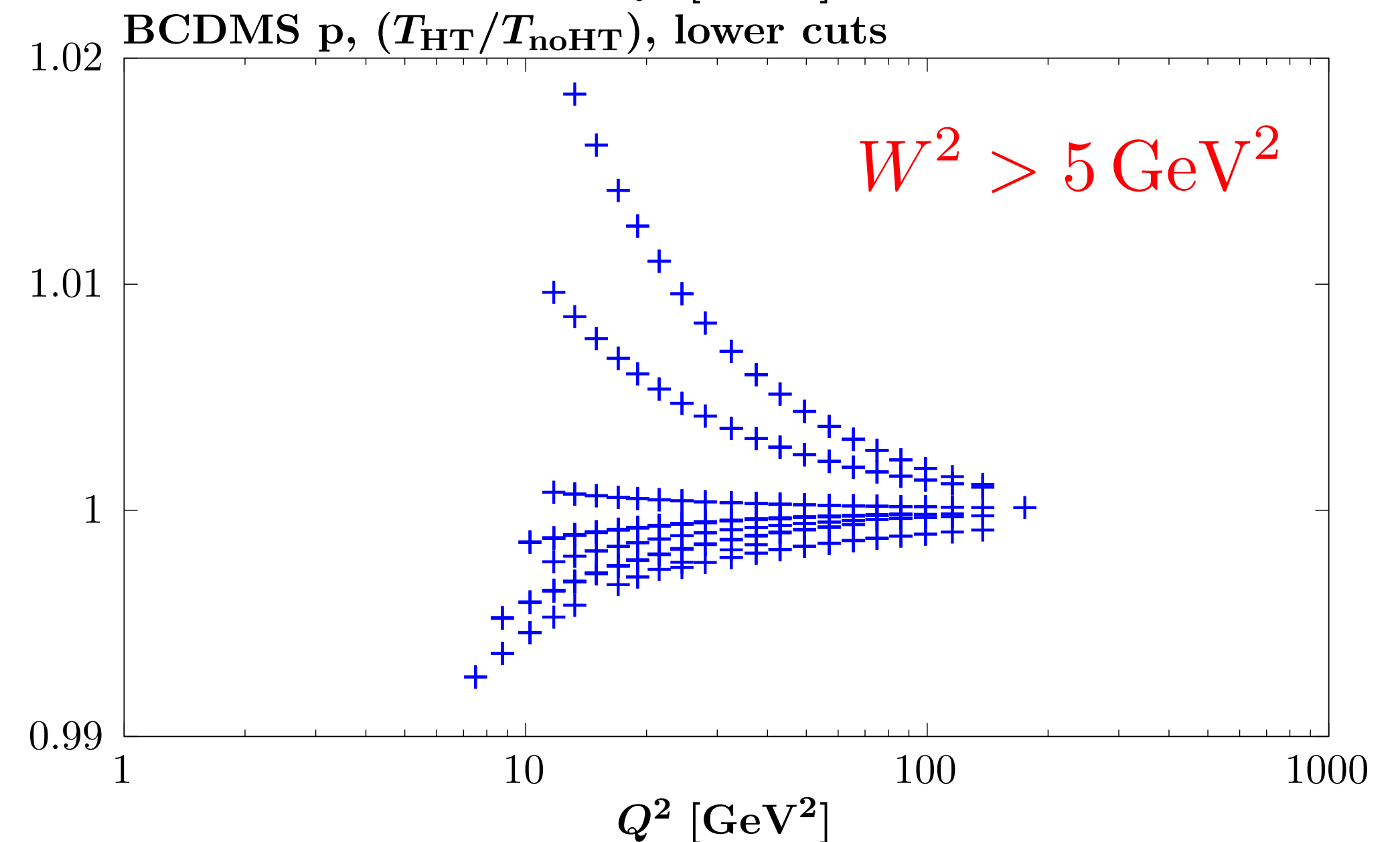
- ★ Generate PD with **NNPDF4.0**- difference in underlying (MSHT. vs. NNPDF) input modelling inconsistency in data.
- ★ This induces downwards shift in strong coupling.
- ★ Size of shift at $T \sim 1$ level - well within enlarged (dynamic tolerance).



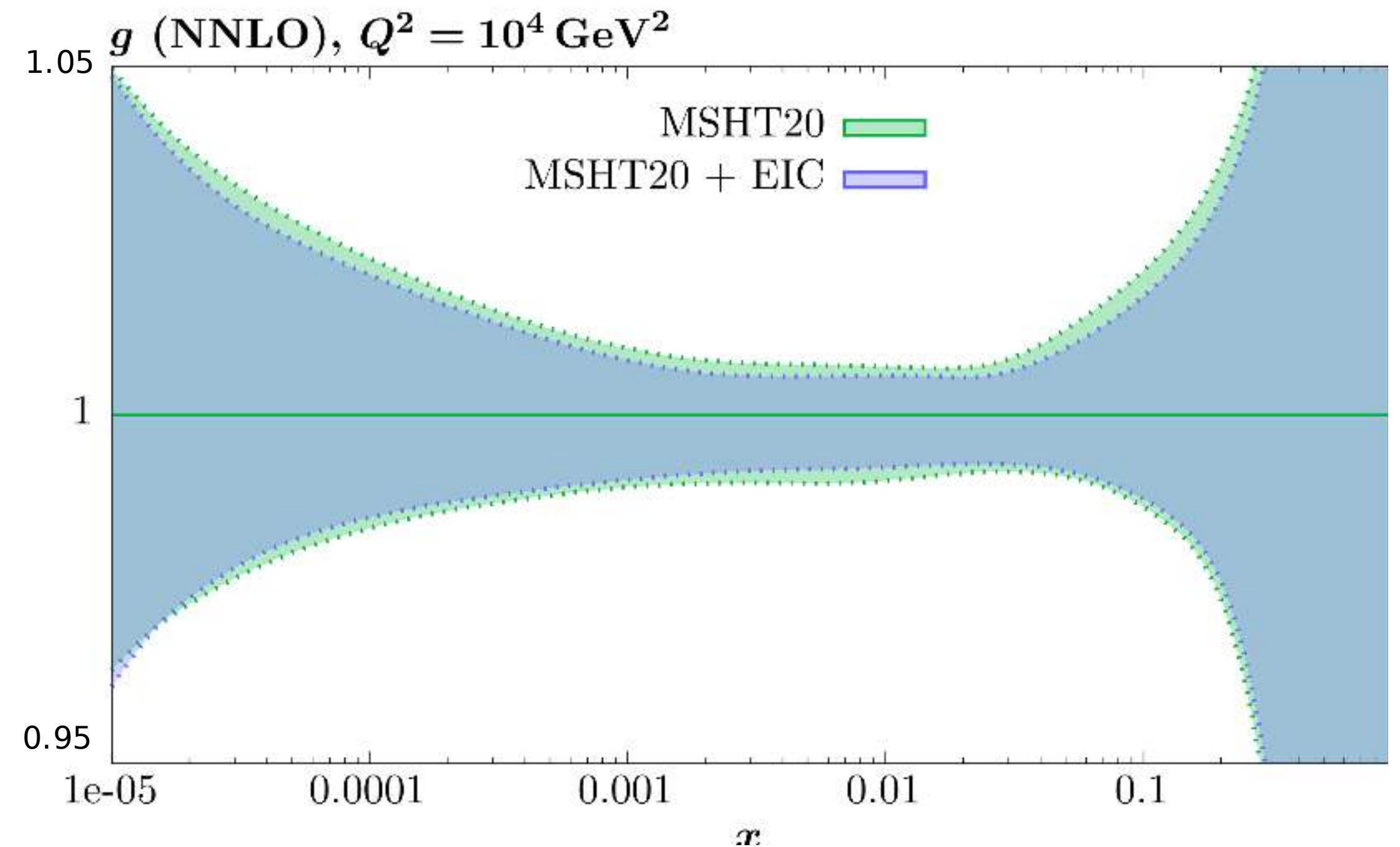
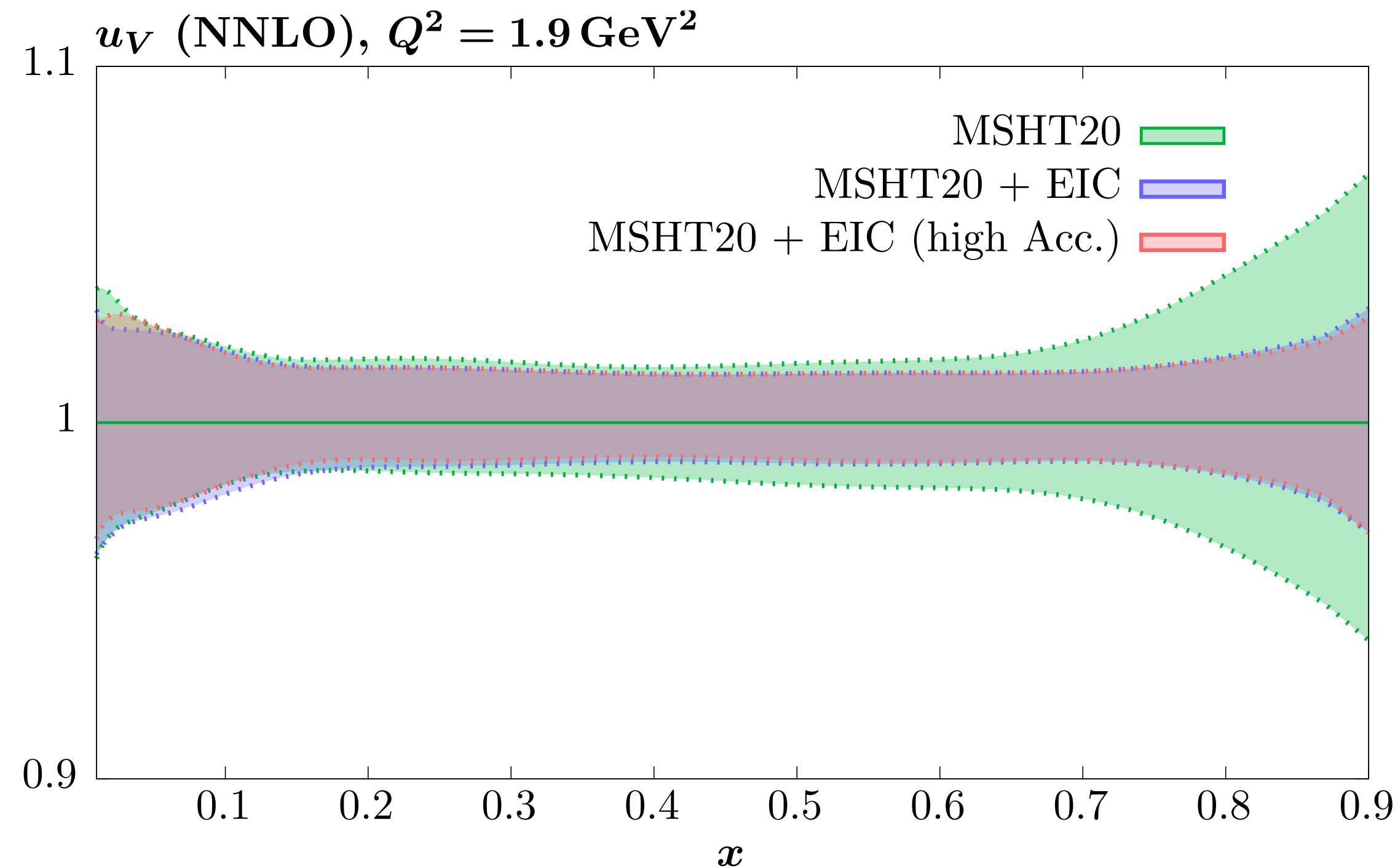
- Note these effects are not obviously evident 'post-hoc' by looking at EIC fit qualities - $\sim 1 - 2\sigma$ worse than ideal case, i.e. similar to PDF fits to real data.



- Indication that ‘higher twist’ corrections in baseline case fit $\sim$ same order as QCD corrections, i.e. mimicking **missing higher order corrections**.
- Not so in lower cut case.
- **Care needed** in interpreting such contributions.

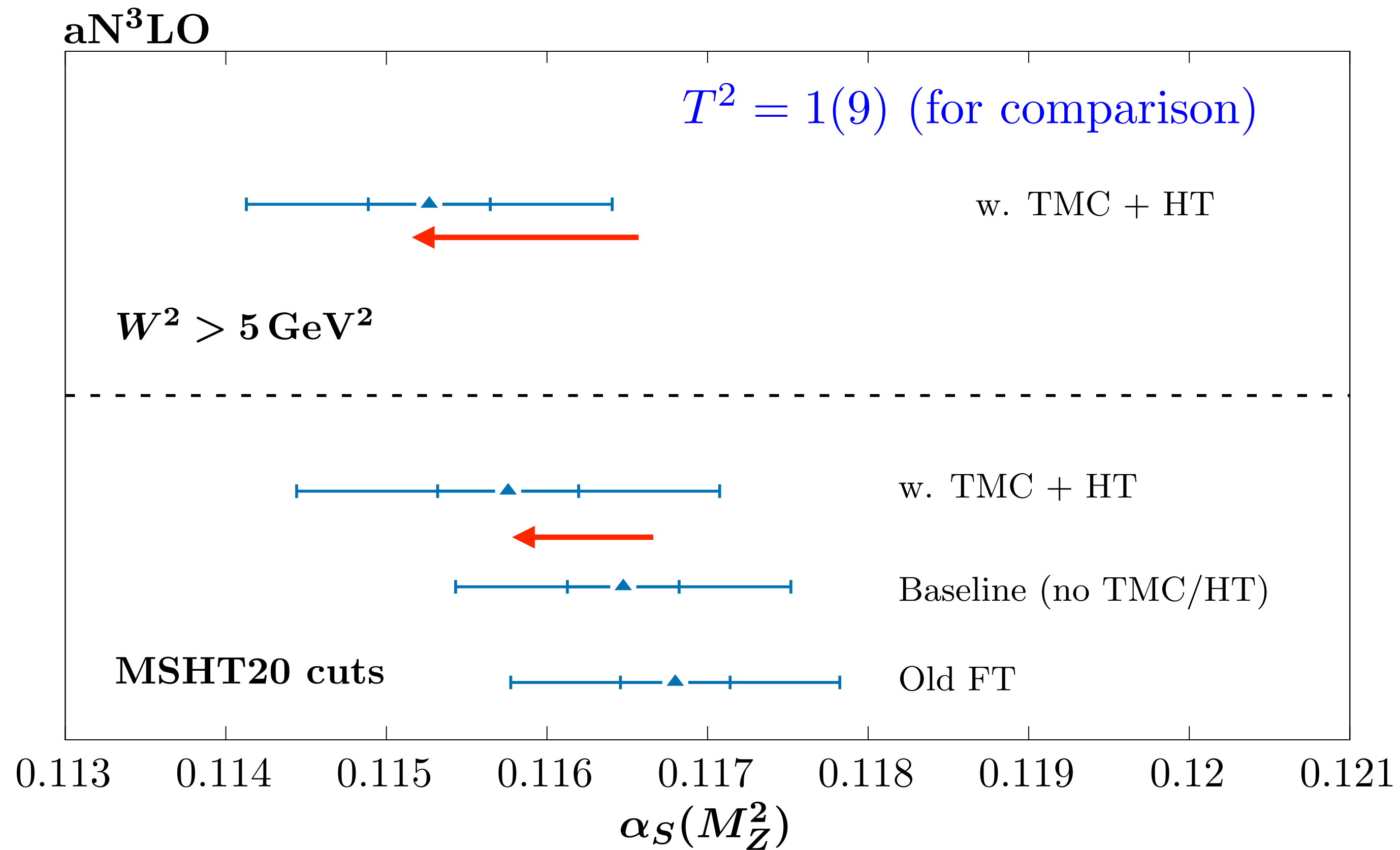


PDF impact



- Largest impact on u at large x as $\sigma_{e-p}^{\text{NC DIS}} \propto \sum_i Q_i^2 f_i(x)$.
- Though some care needed in interpretation at the highest x values: PDF extrapolation, TMC effects relevant...
- Moderate impact on gluon across x from scaling violations.
- Impact of larger y acceptance negligible as different beam energy configurations provide constraints.
- Always worth bearing in mind these tests assume consistency - in presence of tensions complementarity is key!

Strong Coupling



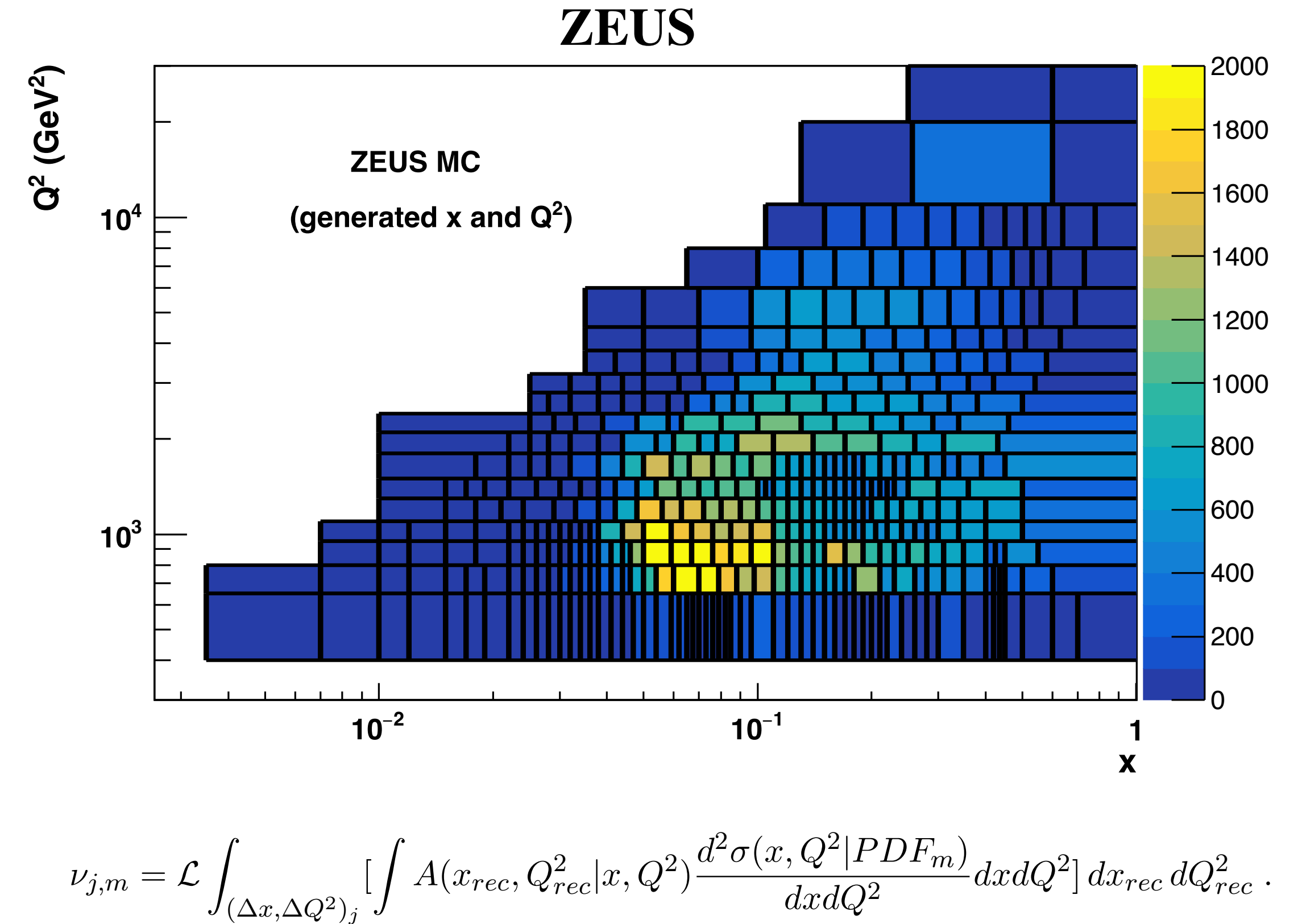
- Including ‘higher twist’ corrections in baseline and lower W^2 cut cases leads to some lowering of best fit values of **strong coupling**, but well within $\sim$ MSHT uncertainty.

ZEUS data

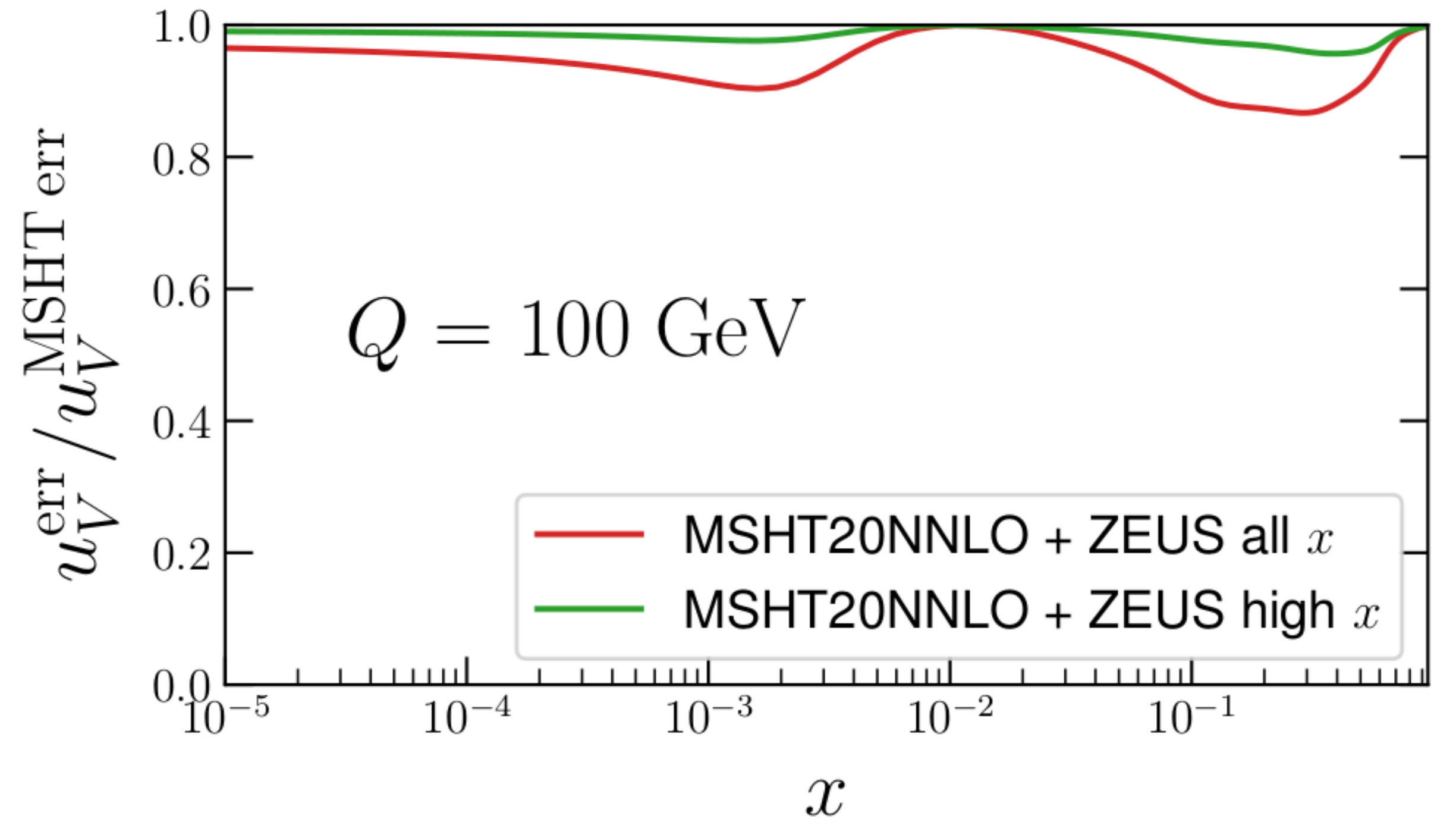
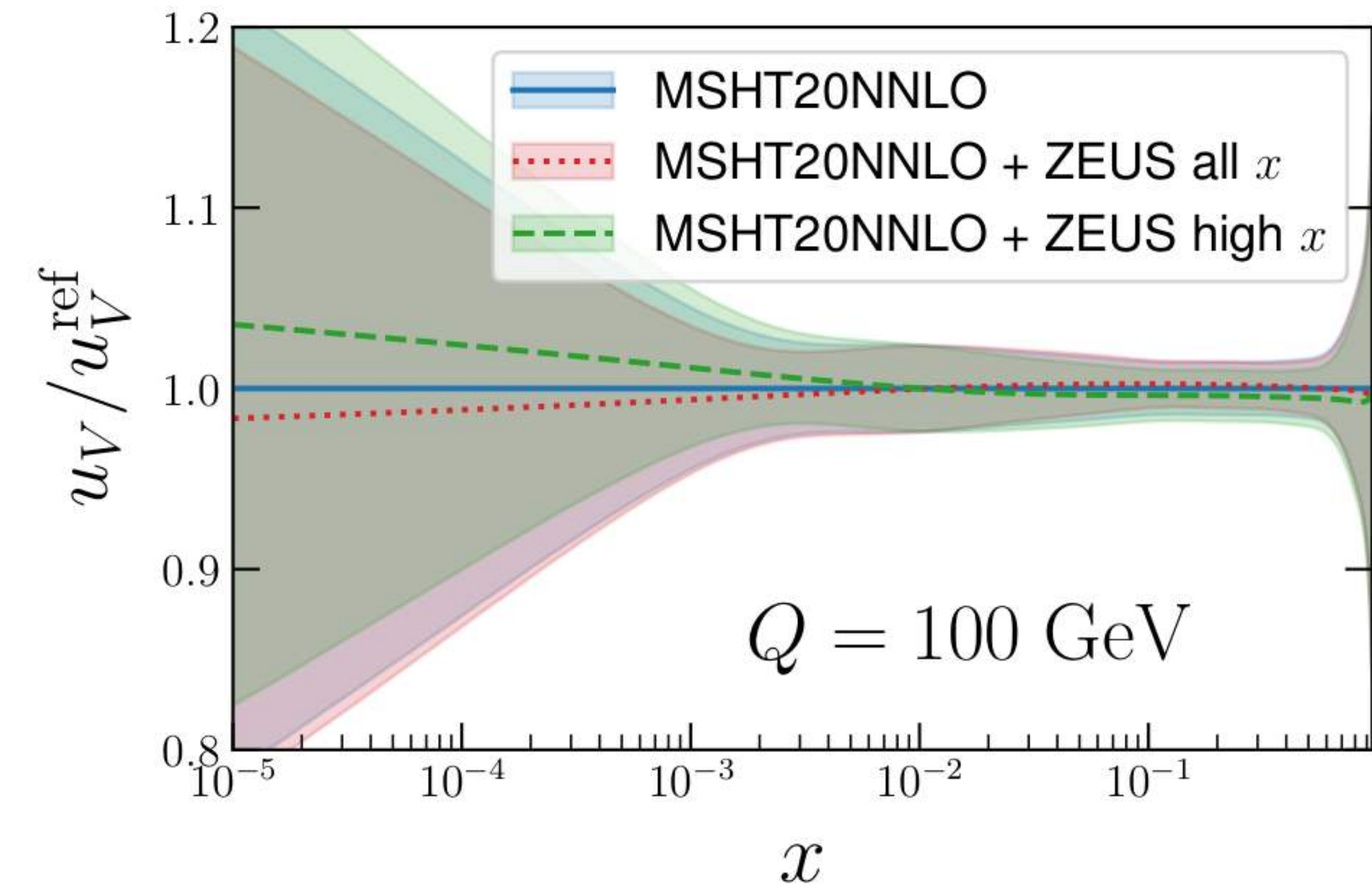
ZEUS, *Phys.Rev.D* 101 (2020) 11, 112009,

- Recent ZEUS re-analysis of inclusive DIS: finer binning and extends to $x \rightarrow 1$.
- Event counts can be very small - standard approach not applicable.
- Instead apply Poissonian probability:

$$\chi_{new}^2 \equiv \chi_0^2 + \sum_{k=1}^{N_{eig}} z_k^2 - 2 \sum_{j=1}^{N_{data}} \ln \left(\frac{e^{-\nu_j} \nu_j^{n_j}}{n_j!} \right) .$$



- And assess impact via Hessian reweighting on MSHT20 baseline.



- Impact very mild! In particular when lower x data excluded to avoid double counting with combined HERA data.

$$\chi_{\text{global, new}}^2 = \chi_{\text{new}}^2 + \chi_{\text{global, old}}^2$$

$$\sum_{i,j=1}^{N_{\text{dat}}} \left(D_i - t_i - \sum_k \Gamma_{i,k}^{\text{PDF}} \beta_{k,\text{PDF}} \right) C_{ij}^{-1} \left(D_j - t_j - \sum_m \Gamma_{j,m}^{\text{PDF}} \beta_{m,\text{PDF}} \right) + T^2 \sum_k \beta_{k,\text{PDF}}^2$$

New data
Data already in fit

- Can also investigate what things would look like if we (incorrectly) omit tolerance from profiling.

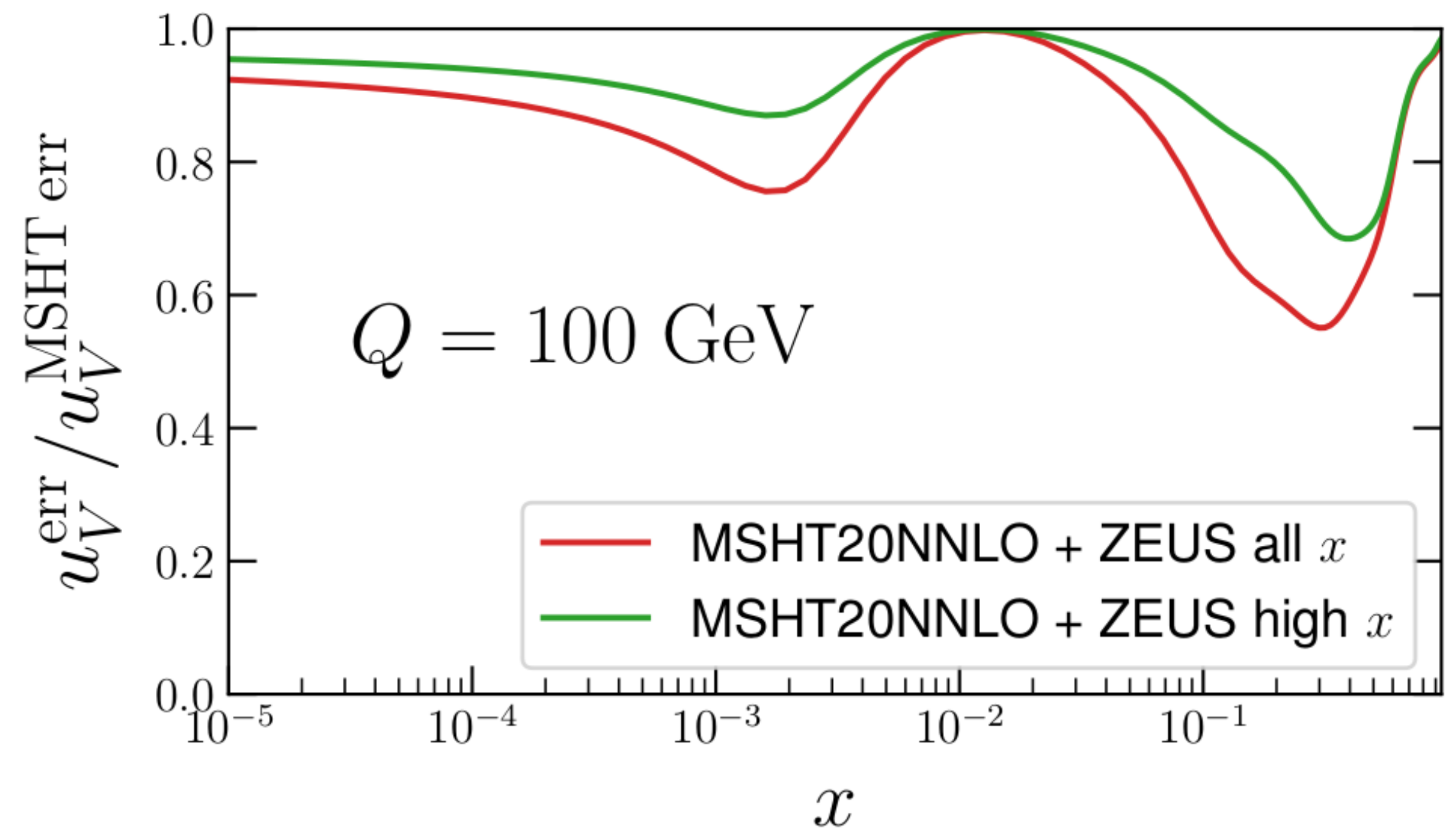
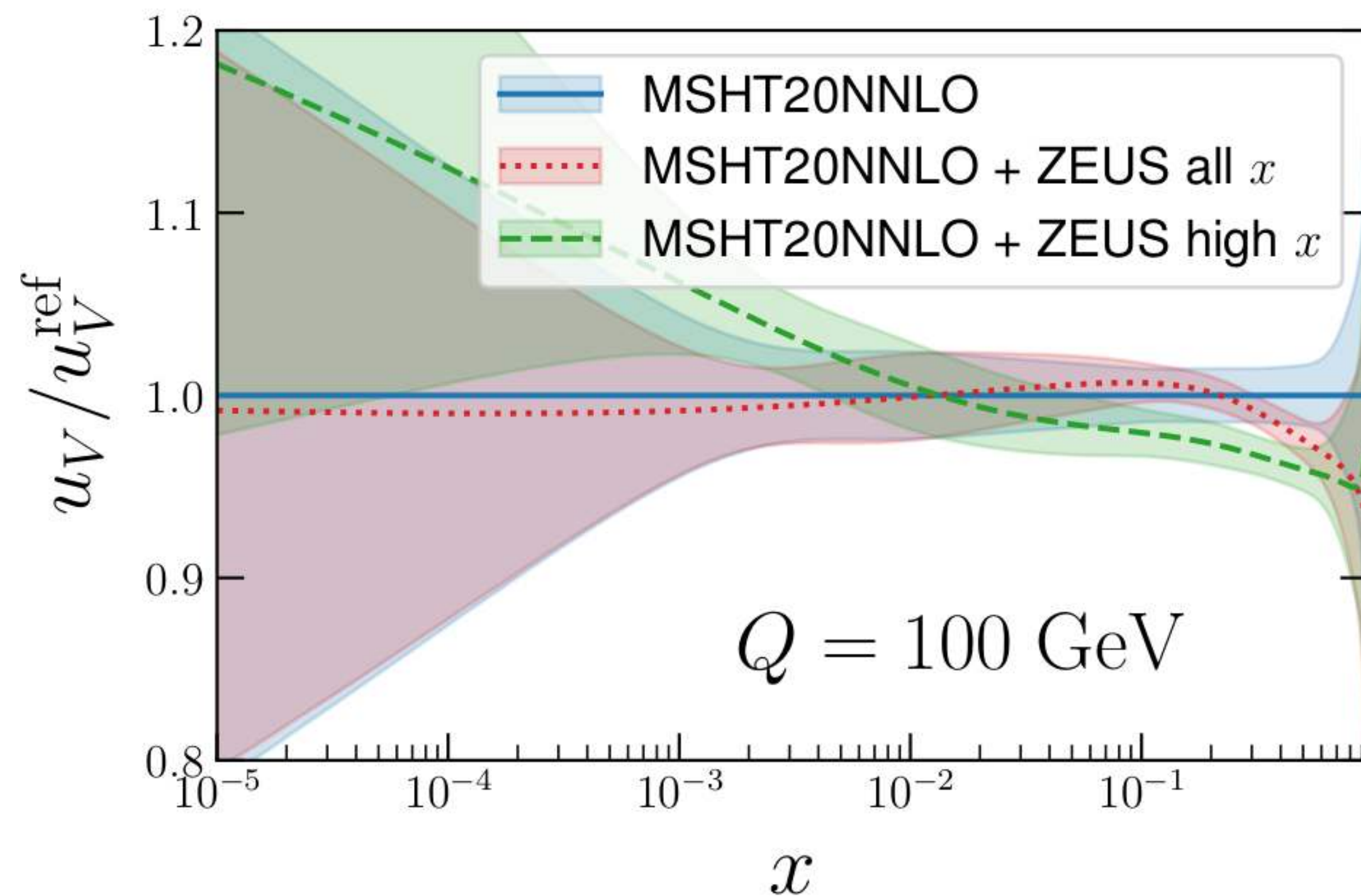
$$\chi_{\text{global, new}}^2 = \chi_{\text{new}}^2 + \chi_{\text{global, old}}^2$$

$$\sum_{i,j=1}^{N_{\text{dat}}} \left(D_i - t_i - \sum_k \Gamma_{i,k}^{\text{PDF}} \beta_{k,\text{PDF}} \right) C_{ij}^{-1} \left(D_j - t_j - \sum_m \Gamma_{j,m}^{\text{PDF}} \beta_{m,\text{PDF}} \right) + \cancel{\sum_k \beta_{k,\text{PDF}}^2}$$

New data

Data already in fit

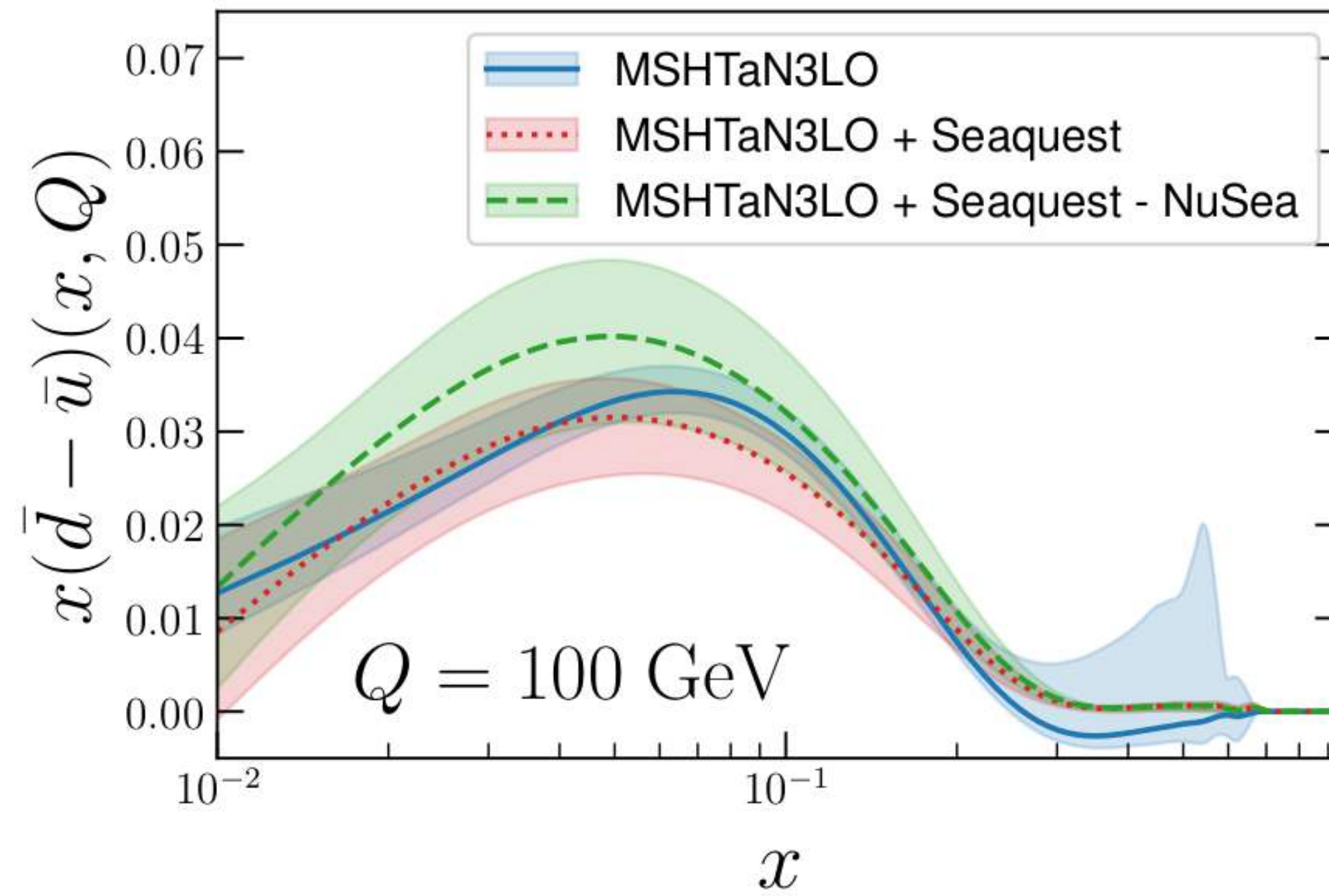
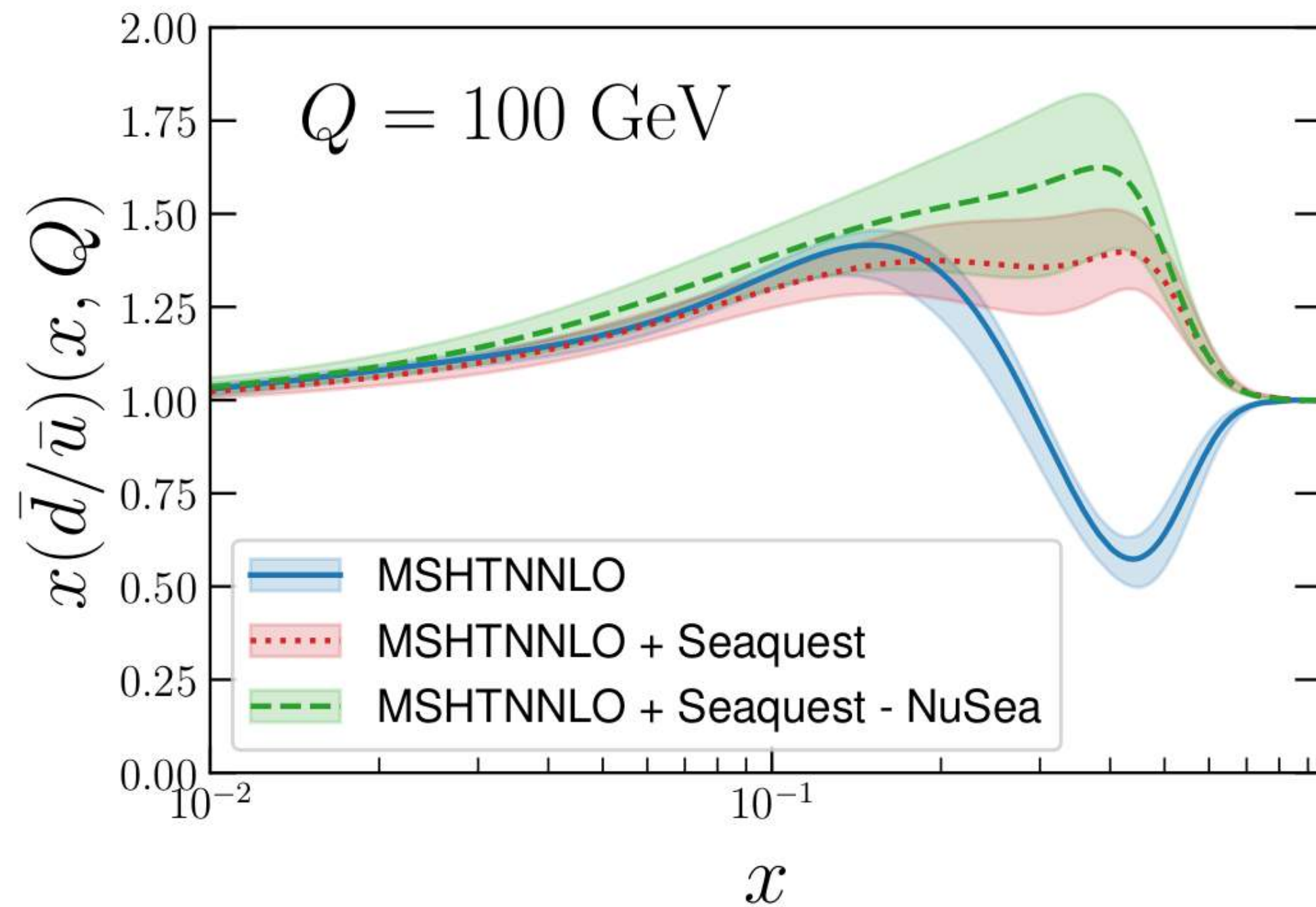
- Can also investigate what things would look like if we (incorrectly) omit tolerance from profiling.



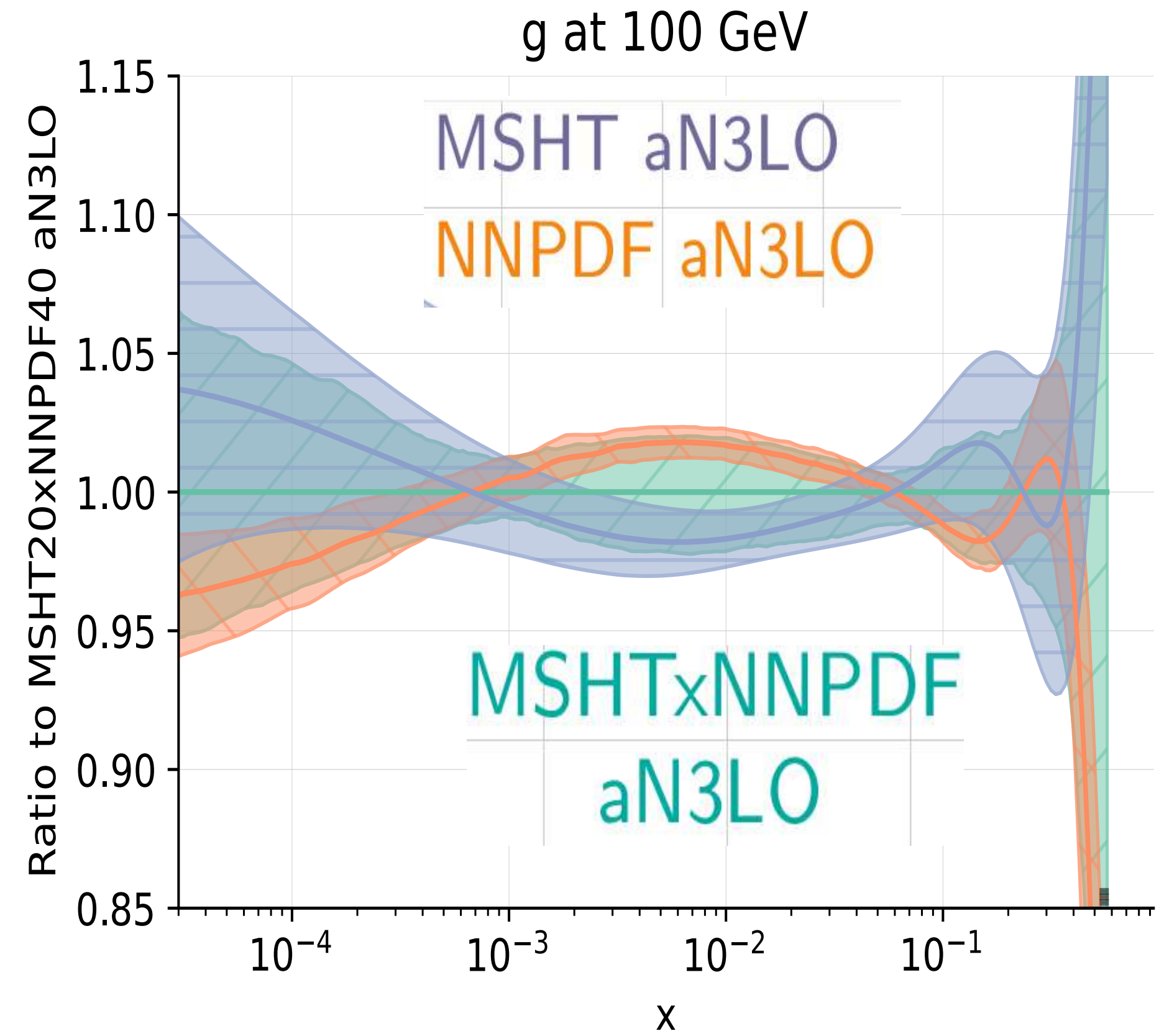
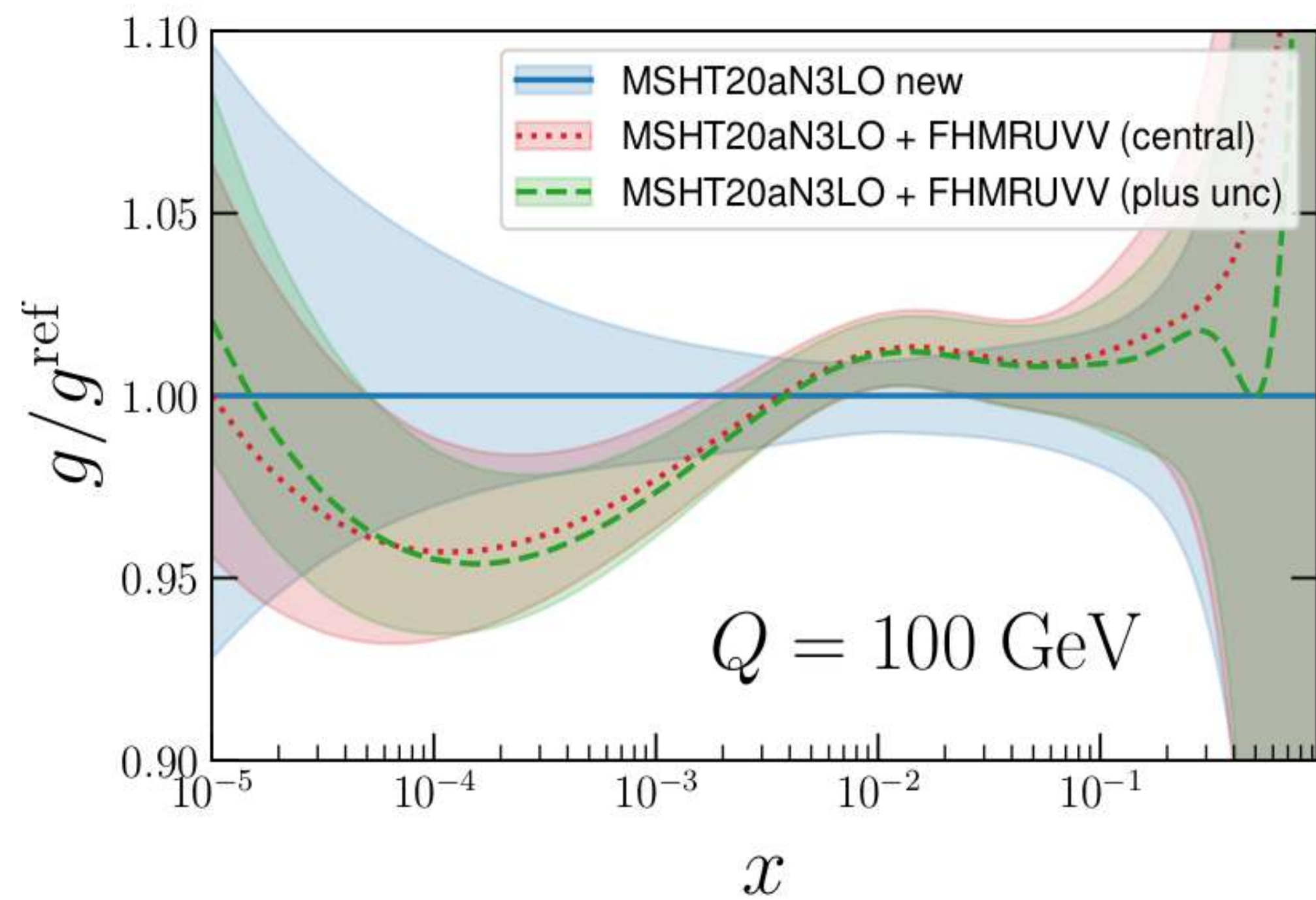
- Impact significantly overestimated! Serves as warning for other studies...

Seaquest Data

- Seaquest (E906) fixed target DY data - raises $\bar{d}/\bar{u}$, in some tension with NuSea (E866).
- Constraint on MSHT fit at NNLO and aN3LO clear.
- Tension with NuSea evident from fit qualities: NuSea $\chi^2/N \sim$ doubles with some deterioration elsewhere.



aN3LO gluon - comparison to NNPDF



(See TC et al - MSHT and NNPDF - 2411.05373)