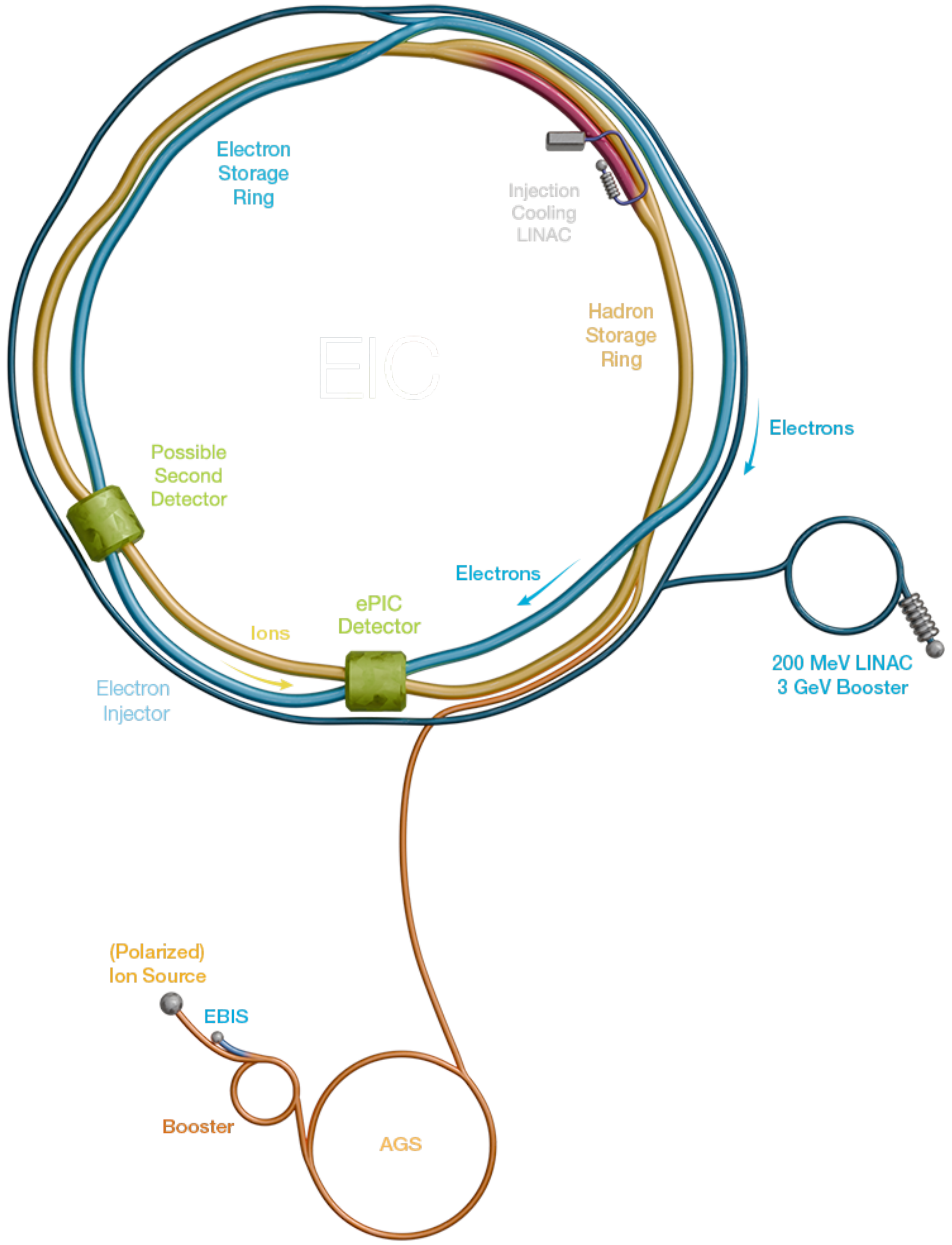
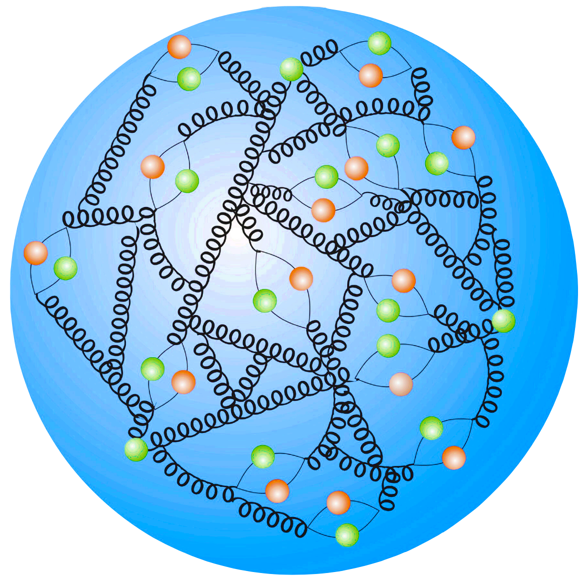




PDF determination and the EIC

impact and opportunities

Juan M. Cruz Martínez

CERN TH Department - juan.cruz.martinez@cern.ch

Uncovering New Laws of Nature at the EIC - BNL November 2024

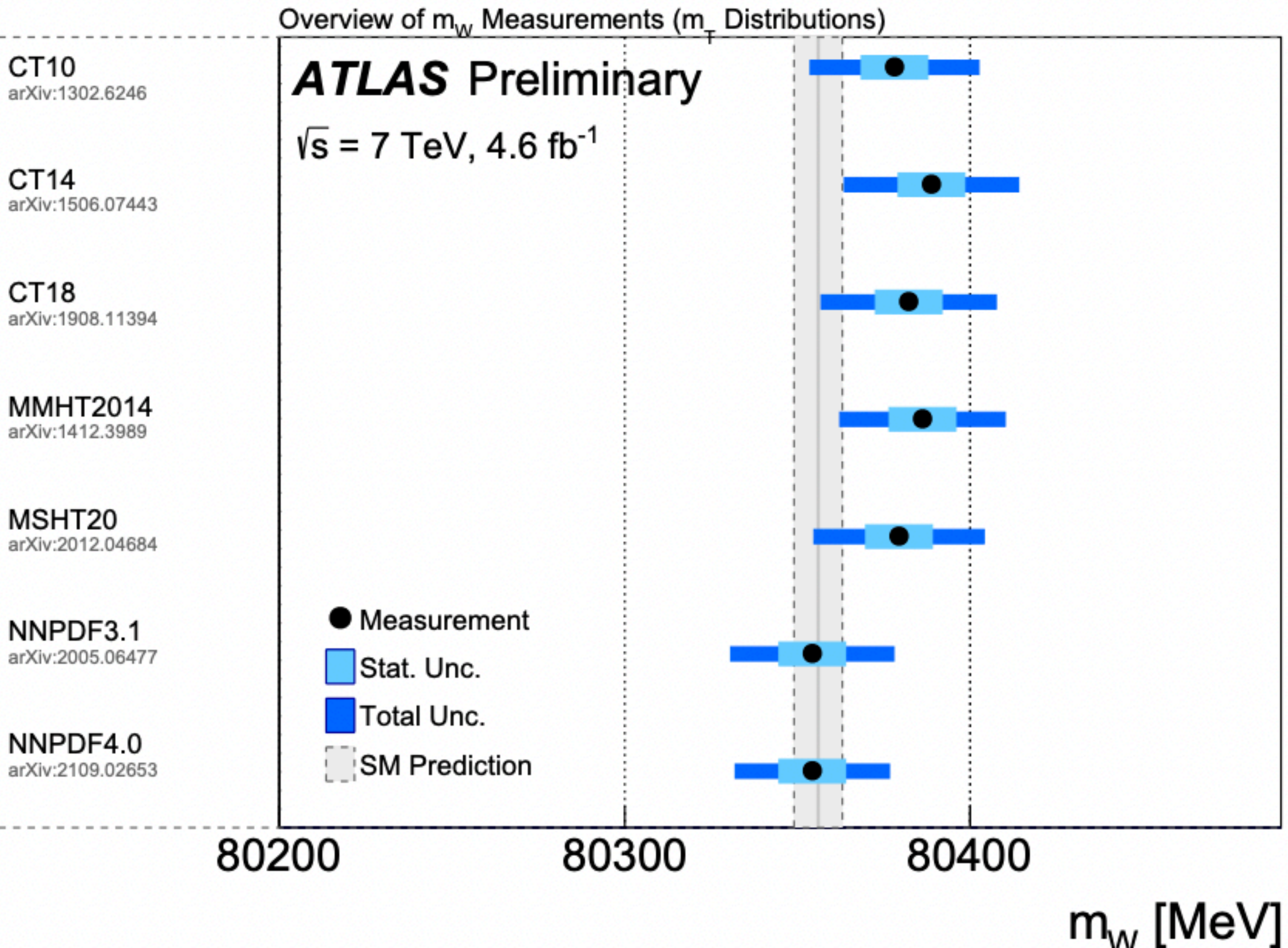
In this talk

- Mainly about unpolarized PDFs, focus on new opportunities form the EIC.
- Polarized PDFs, news from NNPDFpol2.0!

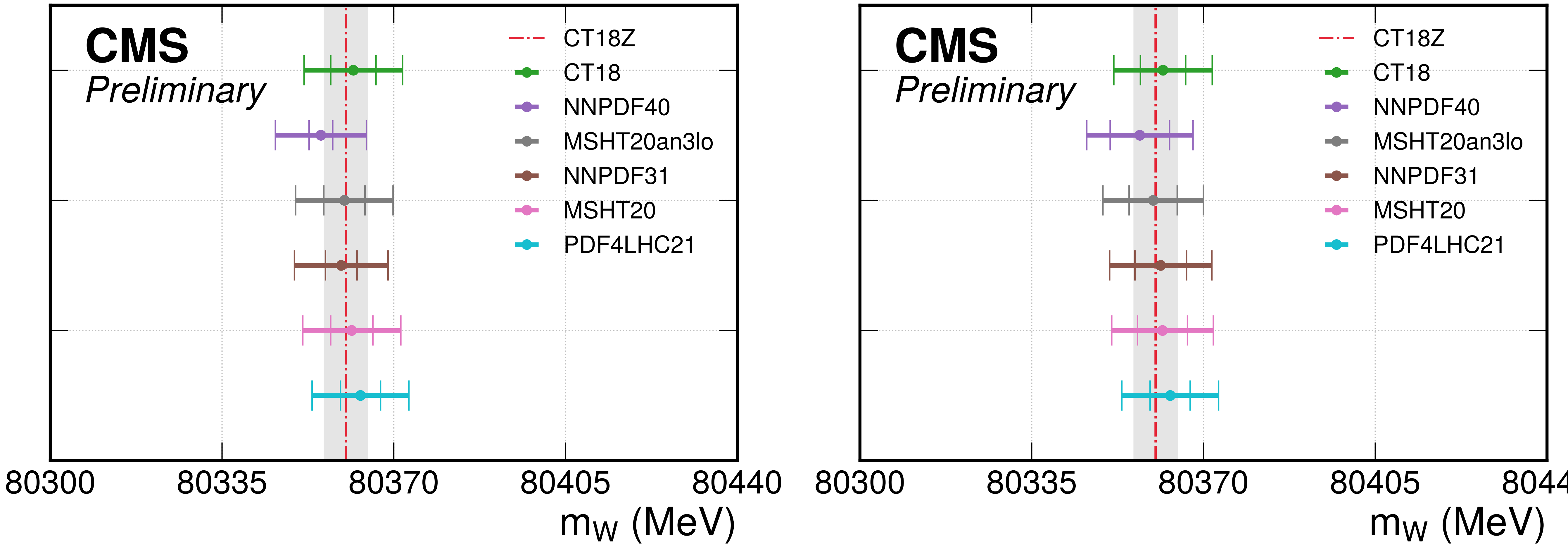
Why do we still care about PDFs?

As the recent measurement of the W mass highlighted, the PDF uncertainty is dominant still today!

The PDF is a very relevant theoretical input with a non-negligible phenomenological impact, and its determination strongly depend on both the theory and experimental data available.



CMS-PAS-SMP-23-002

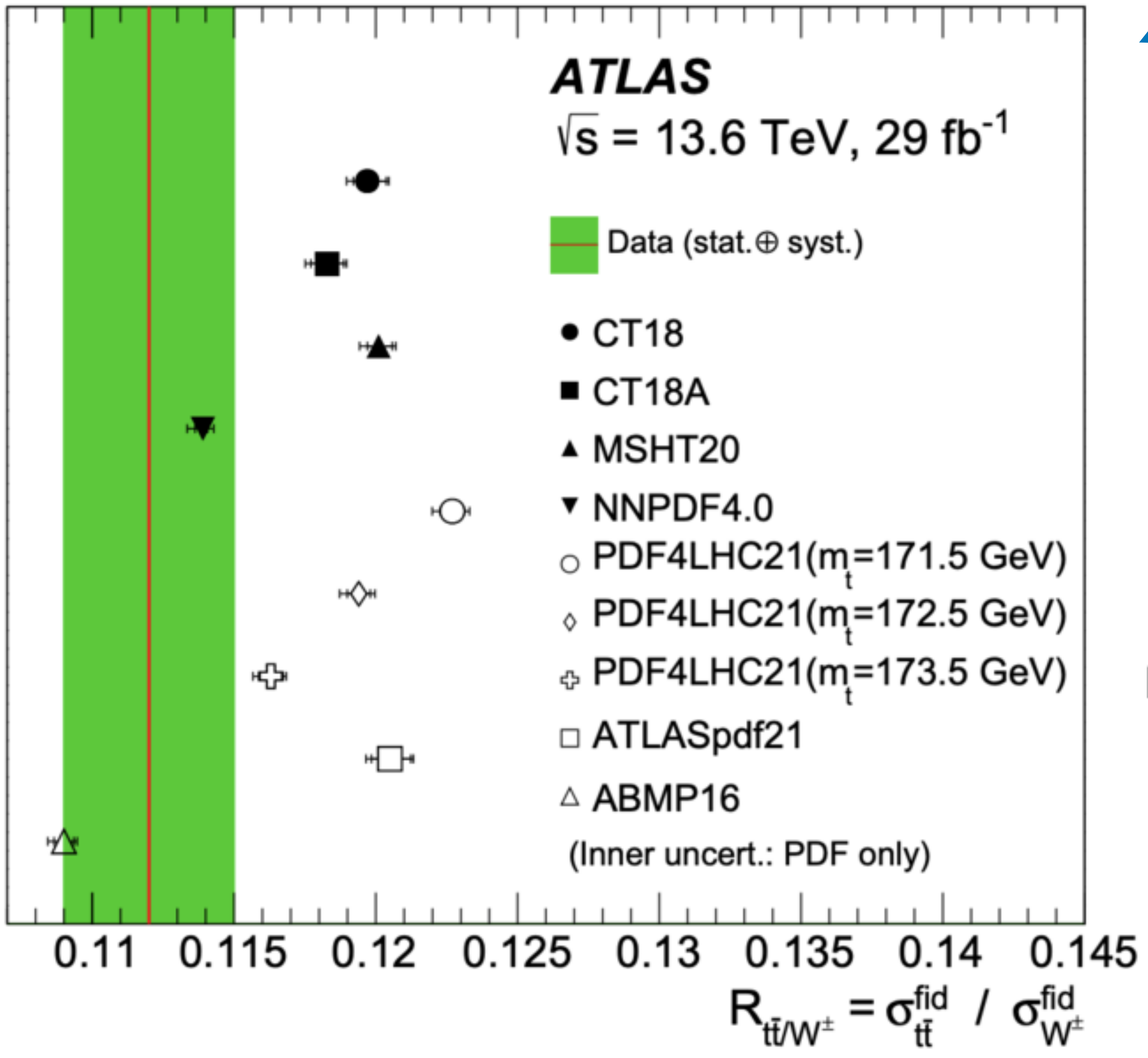


PDF set	Extracted m_W (MeV) Original σ_{PDF}	Extracted m_W (MeV) Scaled σ_{PDF}
CT18Z	$80\,360.2 \pm 9.9$	
CT18	$80\,361.8 \pm 10.0$	
PDF4LHC21	$80\,363.2 \pm 9.9$	
MSHT20	$80\,361.4 \pm 10.0$	$80\,361.7 \pm 10.4$
MSHT20aN3LO	$80\,359.9 \pm 9.9$	$80\,359.8 \pm 10.3$
NNPDF3.1	$80\,359.3 \pm 9.5$	$80\,361.3 \pm 10.4$
NNPDF4.0	$80\,355.1 \pm 9.3$	$80\,357.0 \pm 10.8$

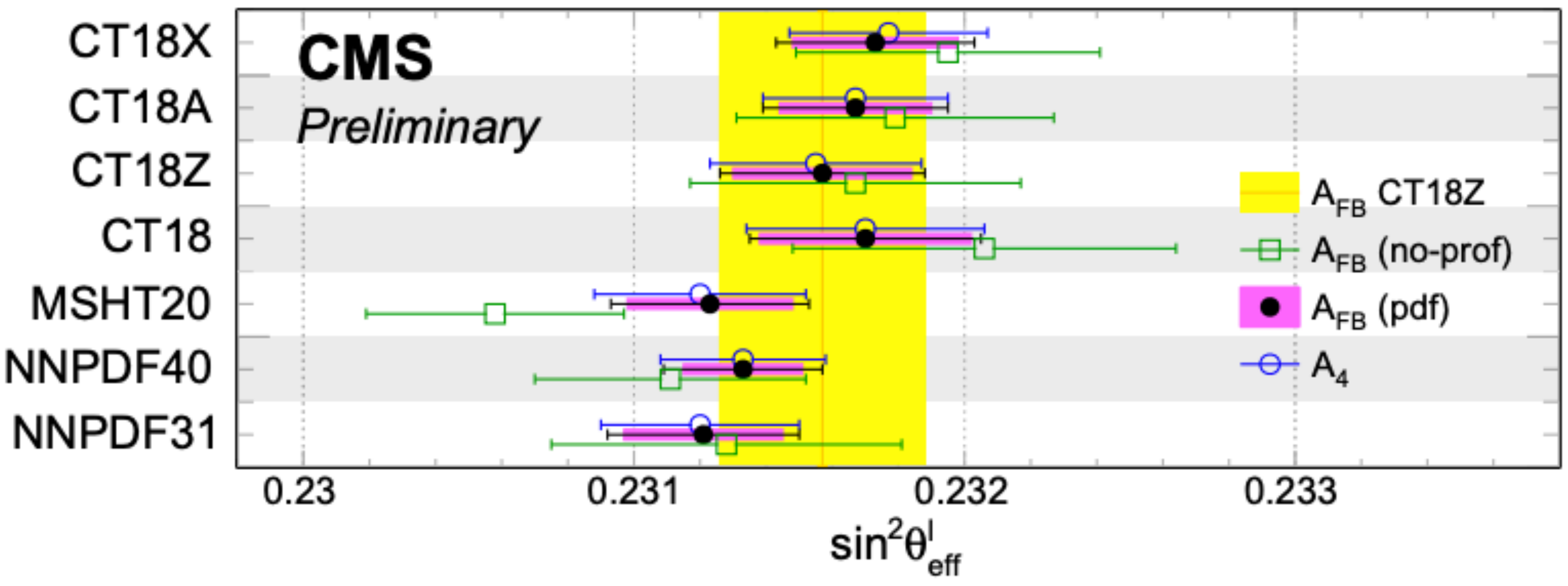
Why do we still care about PDFs?

More examples

Ratio of ttb / W cross section by ATLAS



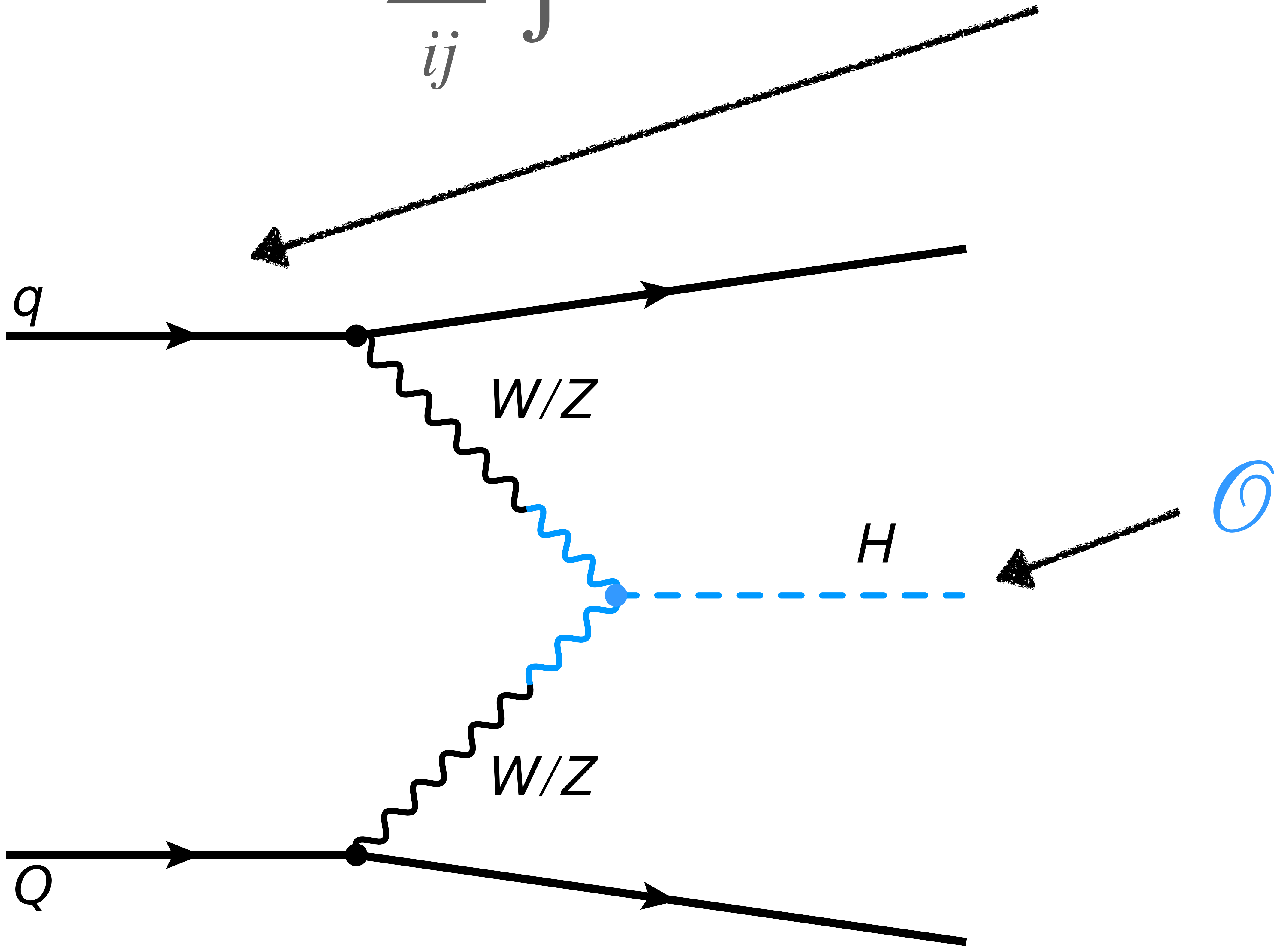
CMS determination of the weak mixing angle

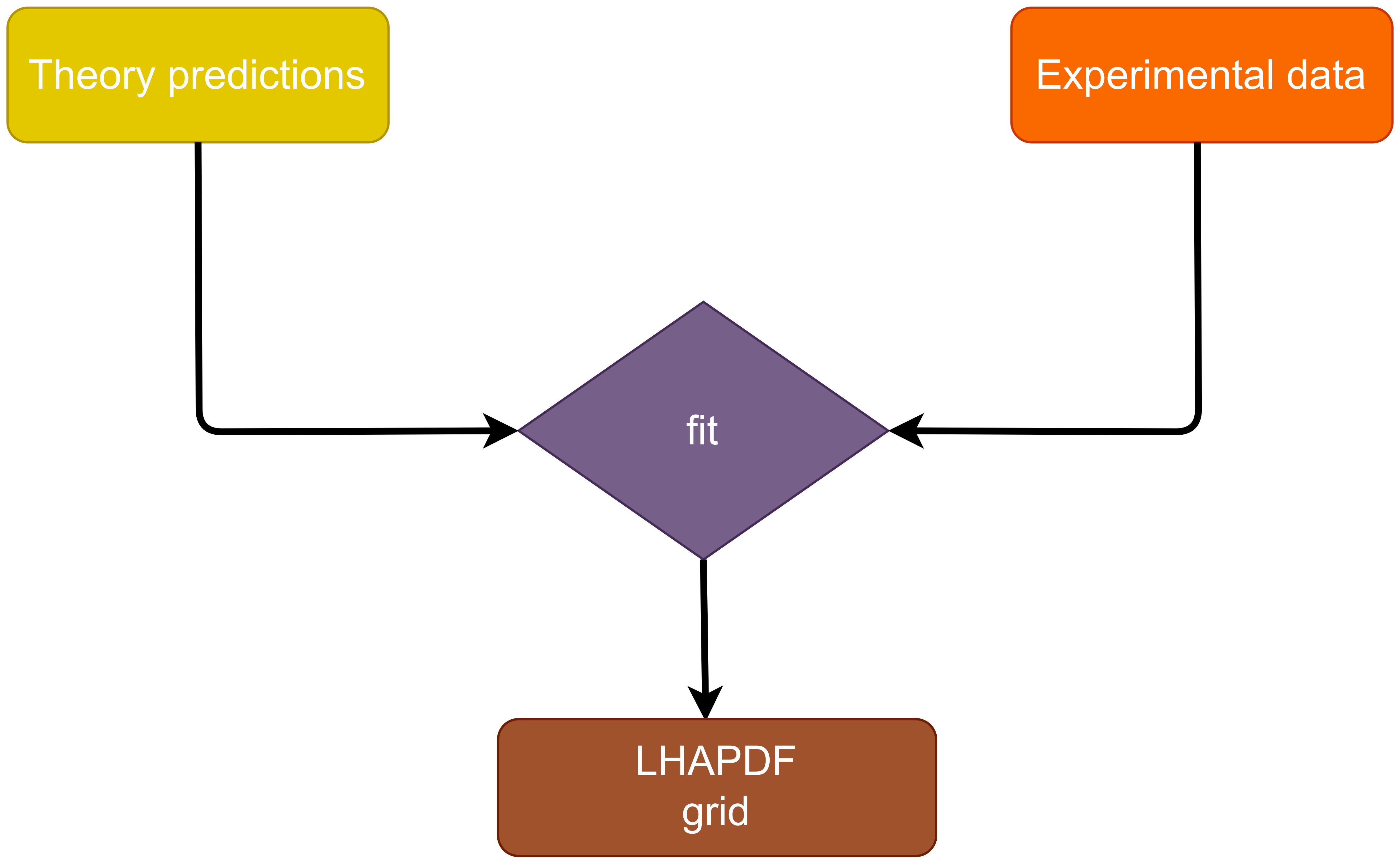


PDF determination ingredients

PDFs cannot be computed analytically from first principles…so their determination depend on the comparison of (fixed order) observables against experimental data.

$$\mathcal{O} = \sum_{ij} \int dx_1 dx_2 \; f_i(x_1, \mu_F) \; f_j(x_2, \mu_F) \; \hat{\sigma}_{ij}(x_1, x_2, \mu_R, \mu_F)$$

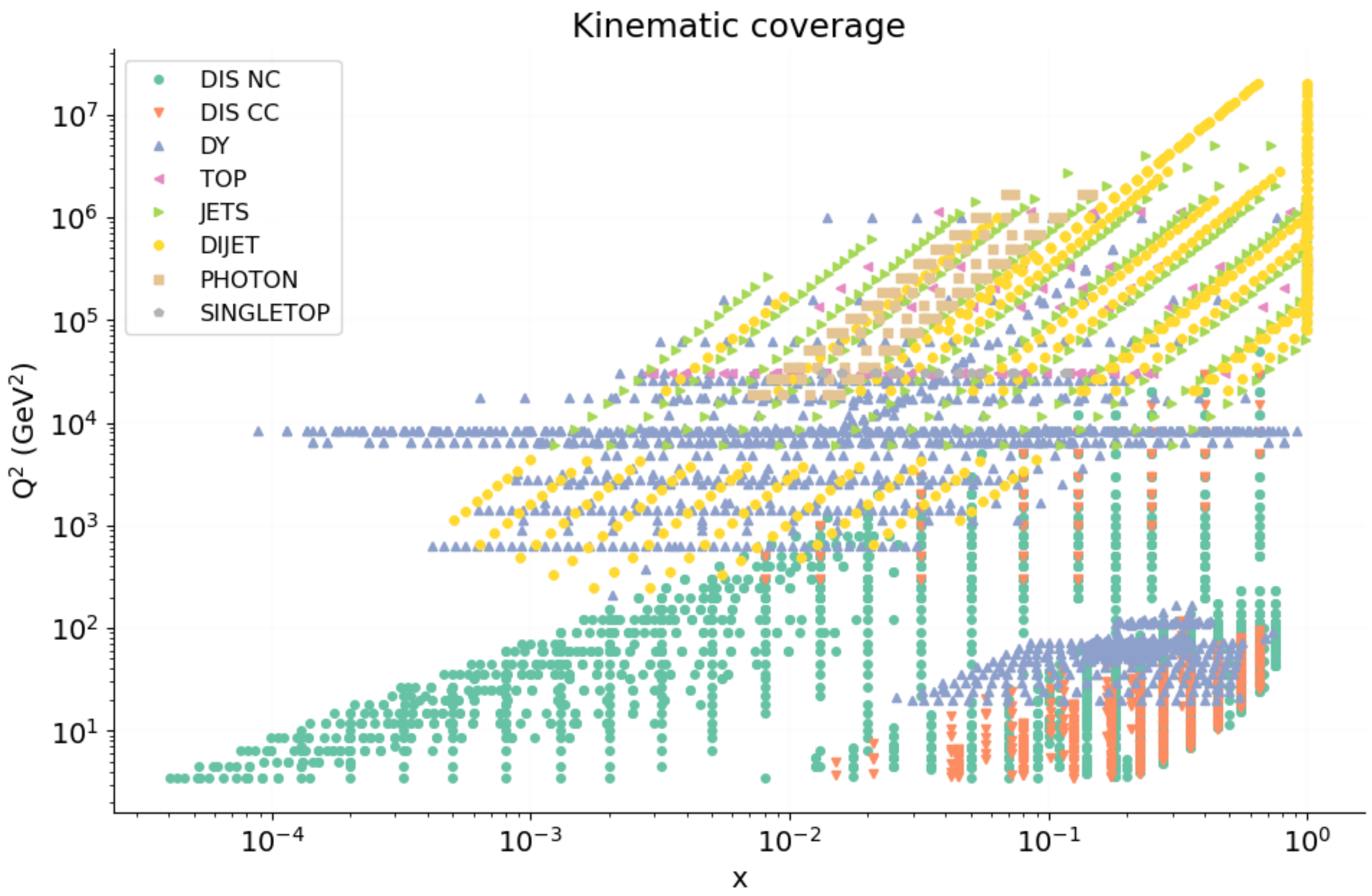


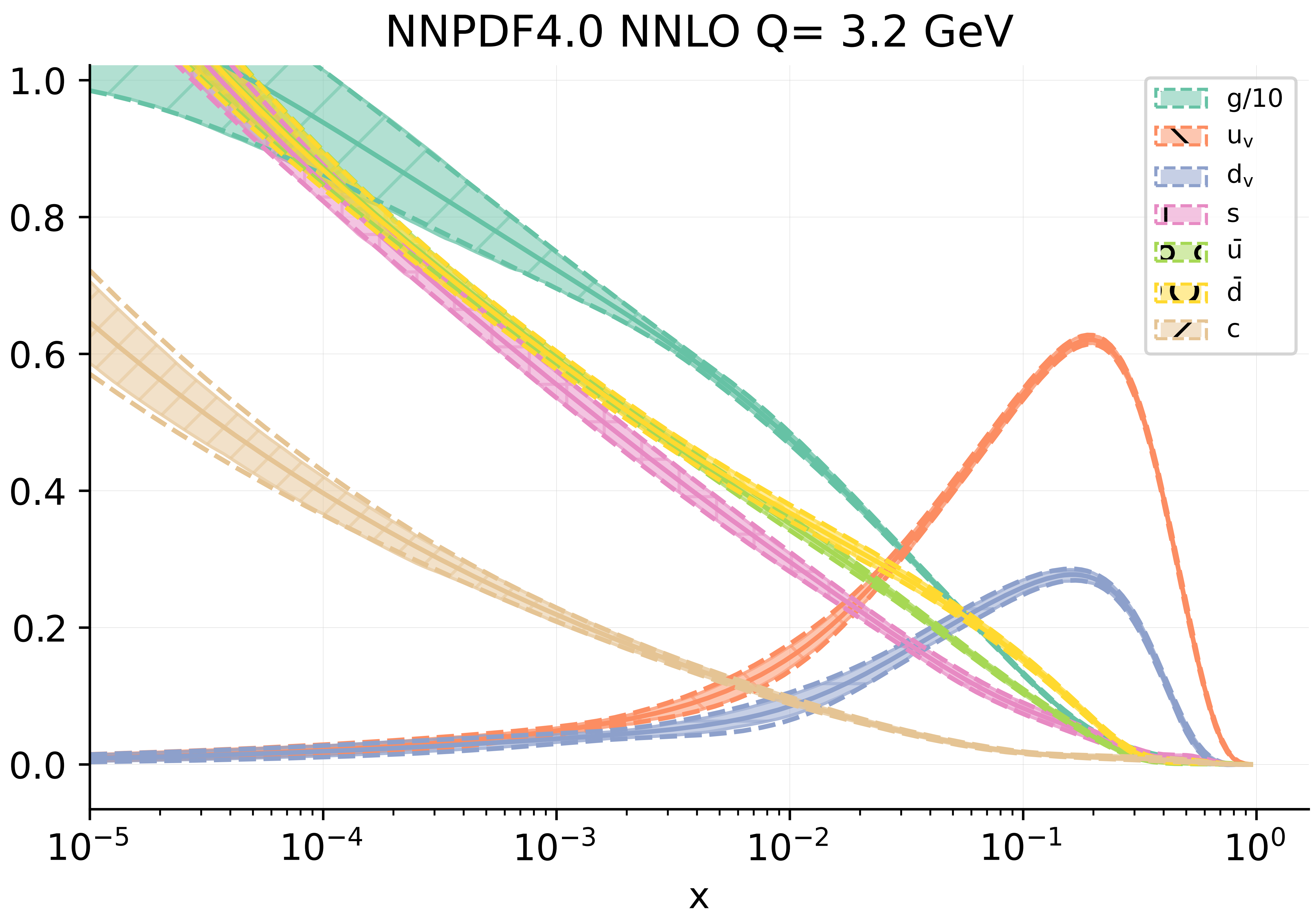


PDF determination ingredients

PDFs cannot be computed analytically from first principles…so their determination depend on the comparison of (fixed order) observables against experimental data.

$$\mathcal{O} = \sum_{ij} \int dx_1 dx_2 \; f_i(x_1, \mu_F) \; f_j(x_2, \mu_F) \; \hat{\sigma}_{ij}(x_1, x_2, \mu_R, \mu_F)$$

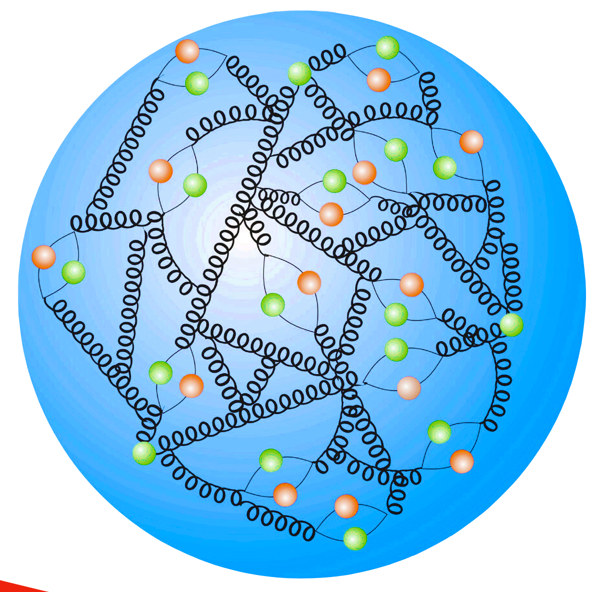




The NNPDF methodology

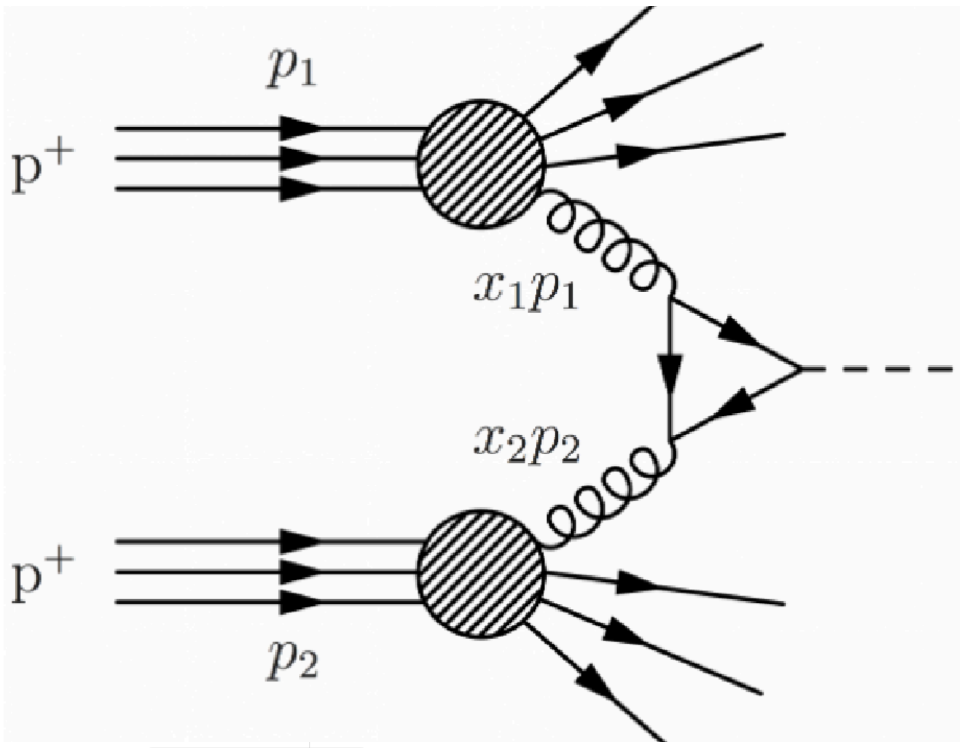
$$f_i(x, \mu_F^0) = A_i x^{\alpha_i}(1-x)^{\beta_i} NN(x)$$

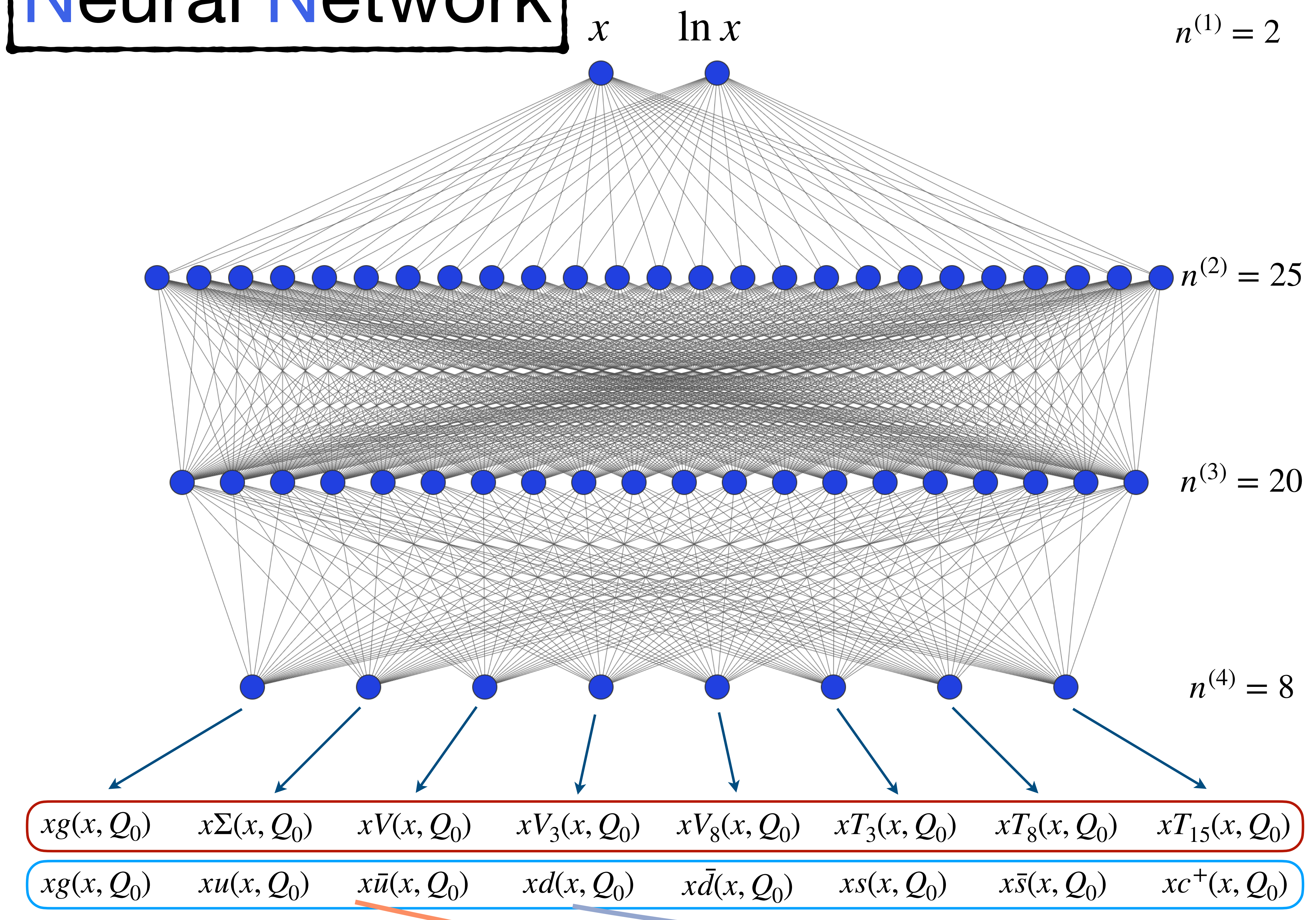
NNPDF

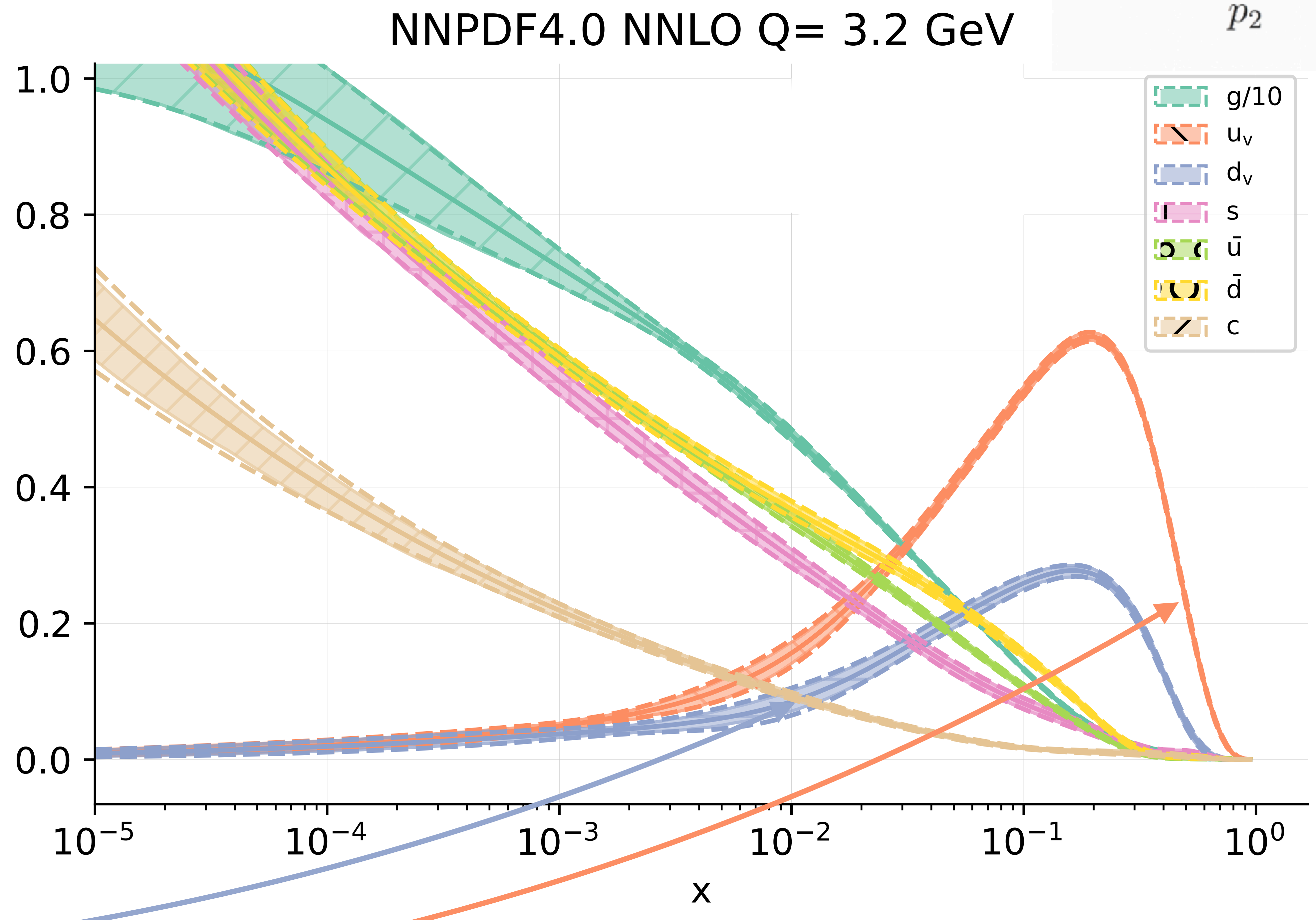


Neural Network

Parton Distribution Function





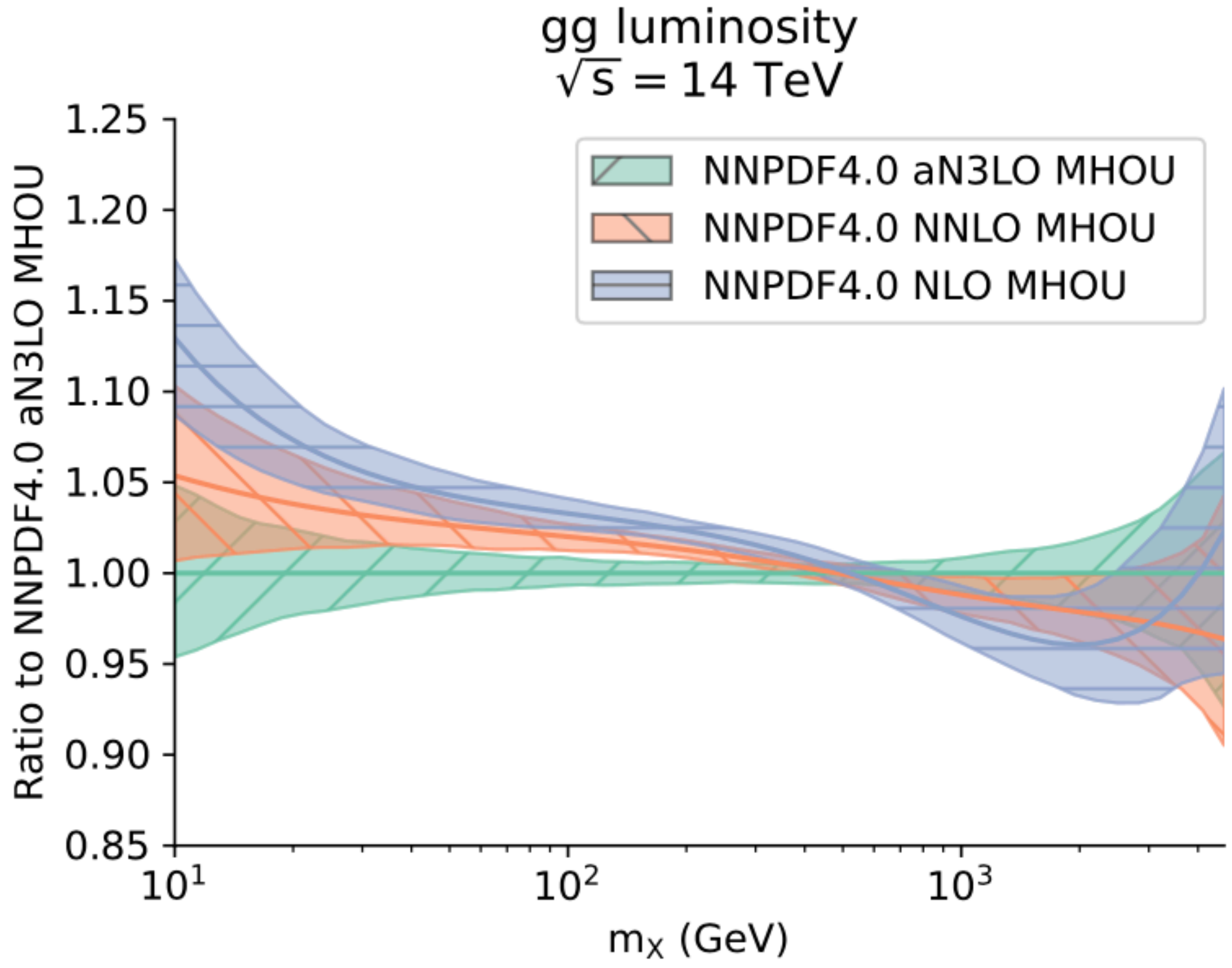


Constraining the PDF accuracy

The PDF can never be more precise or accurate than the data and the predictions that enter as ingredients.

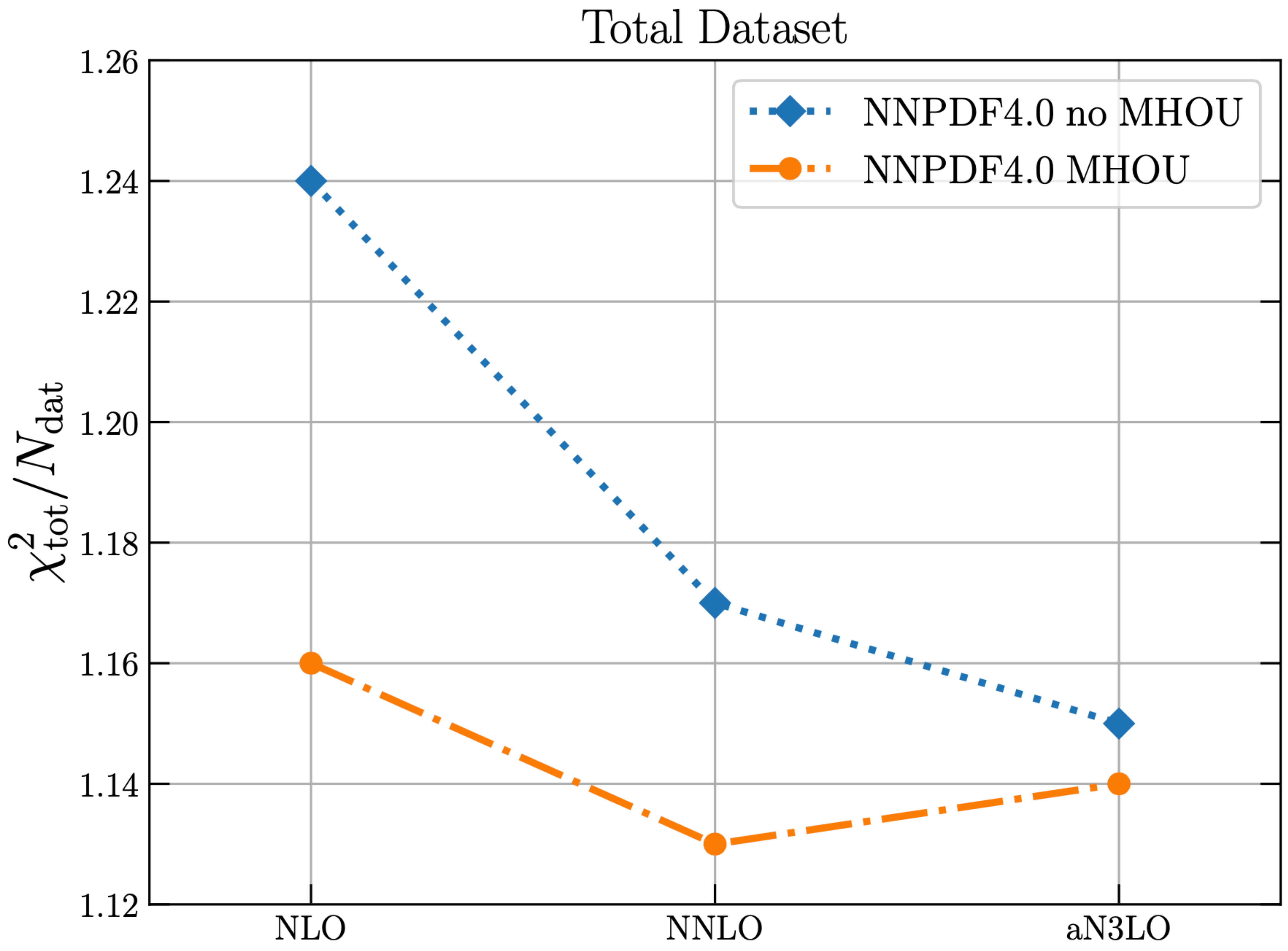
The current state of the art is (approximated) N3LO.

Still work to be done, but perturbative convergence and agreement between groups points in the right direction.



The path to N3LO parton distributions
NNPDF collaboration - hep-ph/2402.18635

- ✳ Splitting functions are not exact, but approximated
- ✳ Hadronic data is only NNLO accurate (NLO grids, NNLO k-factors), with an extra source of uncertainties through scale variations.



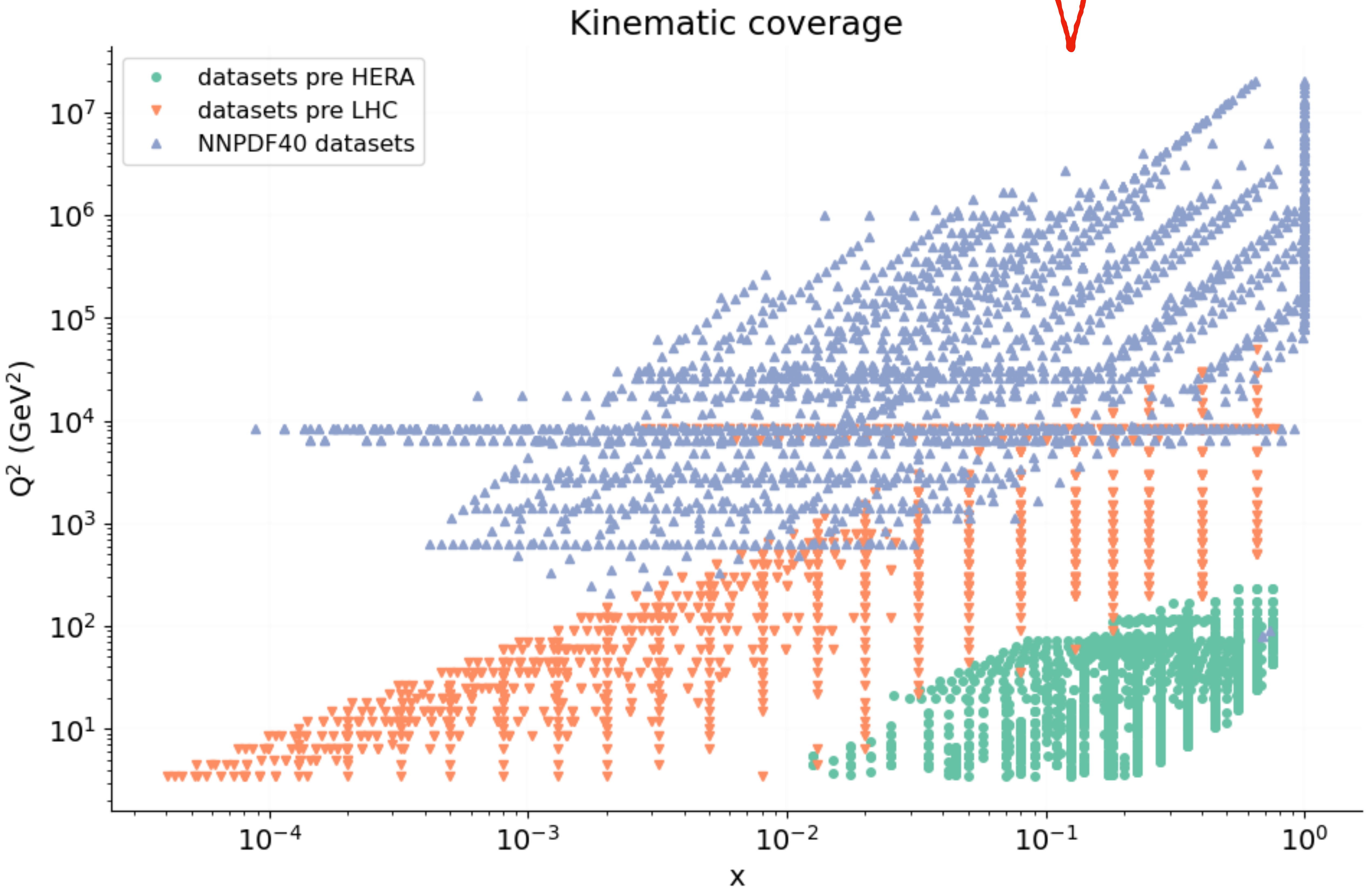
The precision follows the data

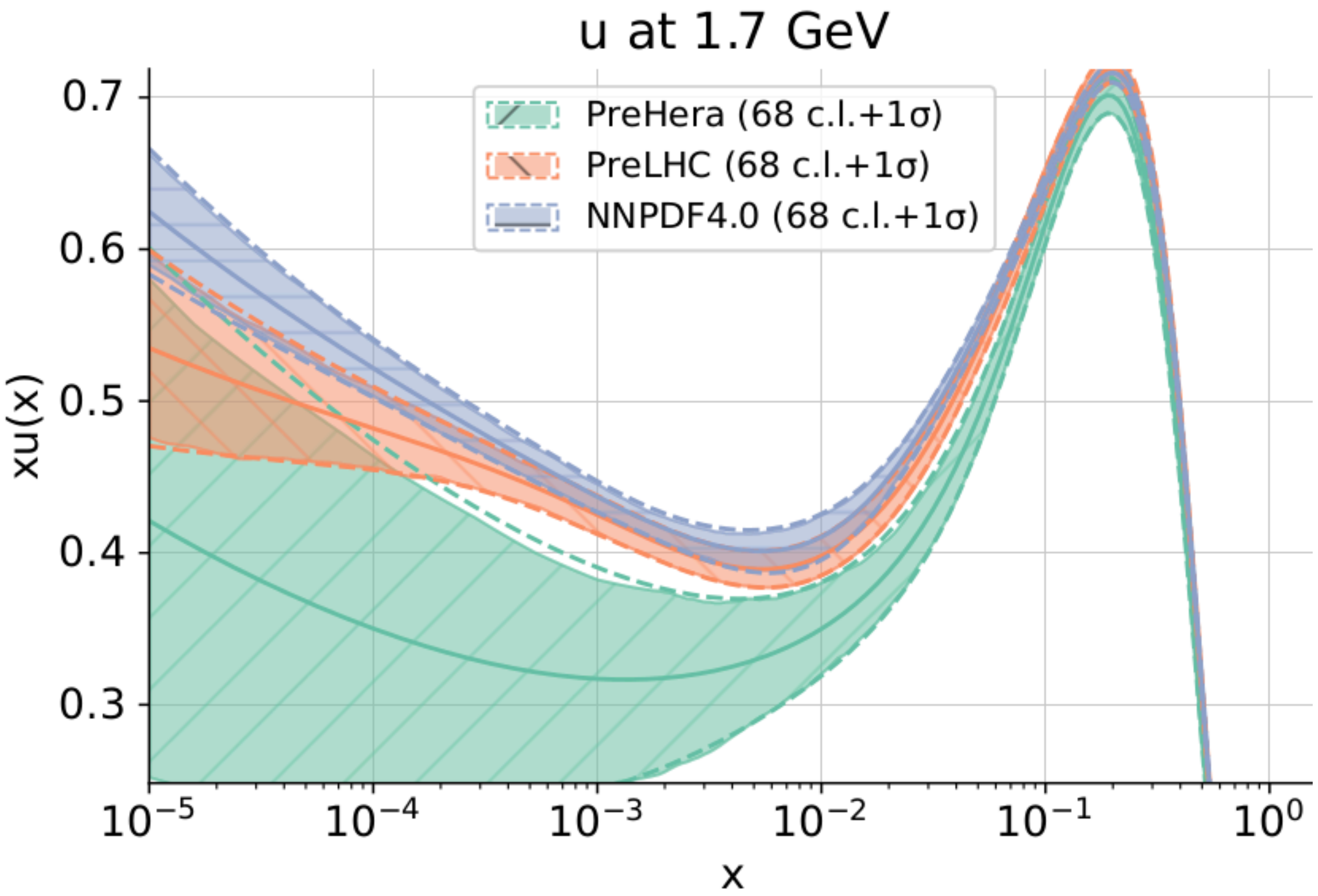
The precision that a PDF can achieve is determined, first and foremost, by the precision and the kinematic coverage of the experimental data that enter the determination.

PDF errors included only in the **red** results

dataset \ fit	Full dataset of NNPDF4.0	pre-LHC	pre-HERA
pre-HERA			0.87
pre-LHC		1.18	**1.22**
Full dataset	1.15	**1.30**	**1.38**

χ^2/N

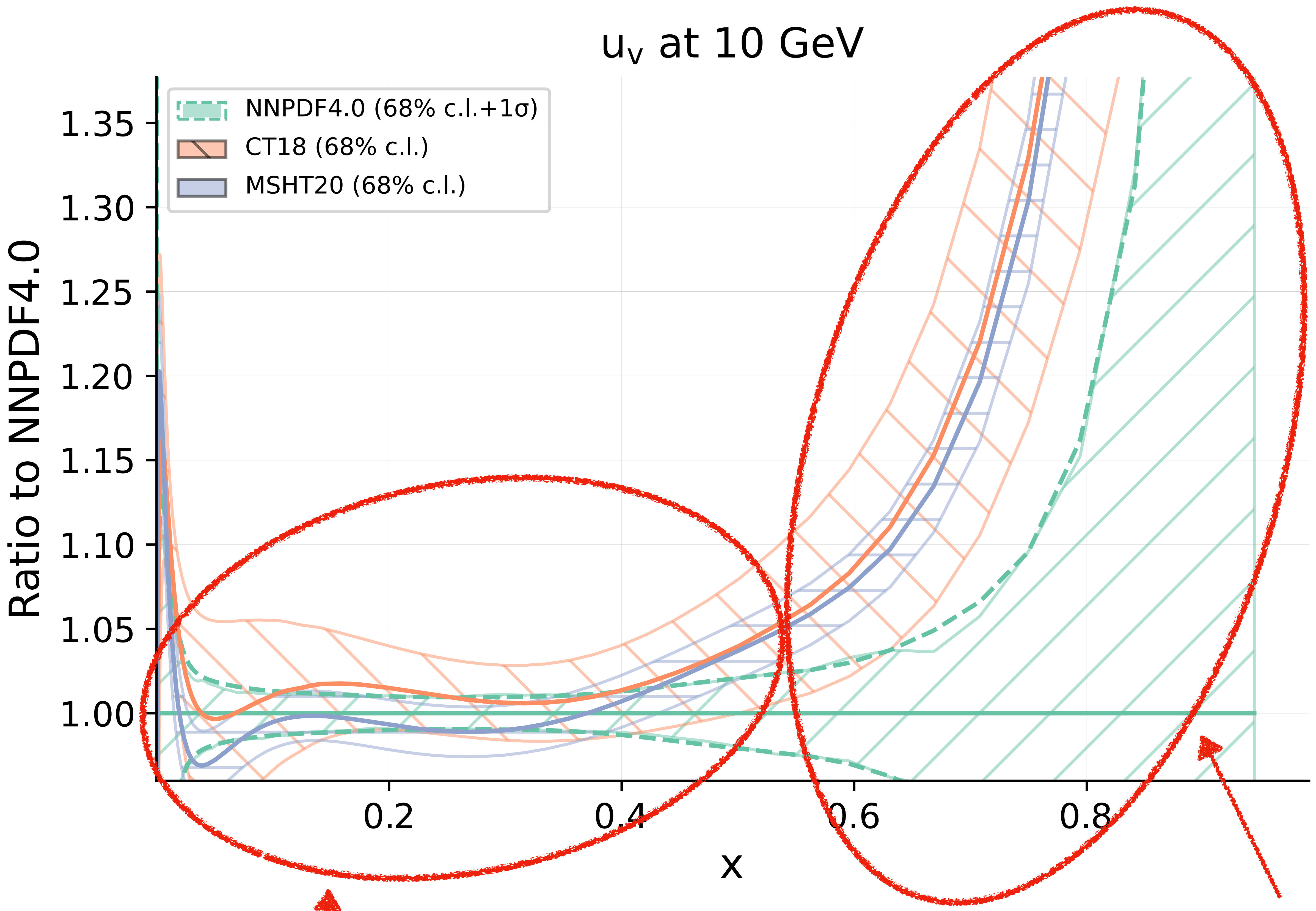




Using for this example the NNPDF dataset and open source fitting code

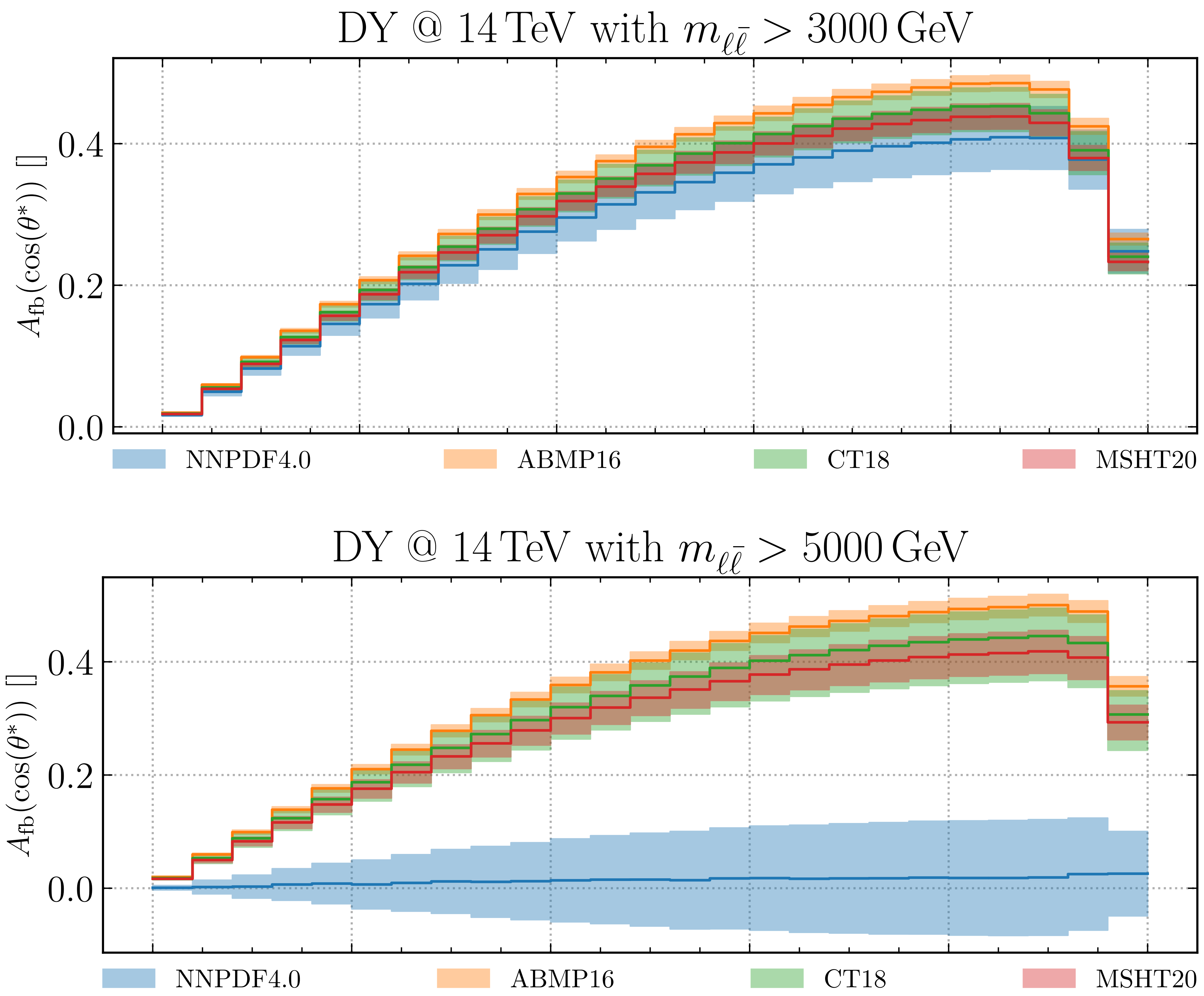
The precision follows the data

Not all regions are equally well determined, for PDFs as we go to higher values of x we leave the “data region”

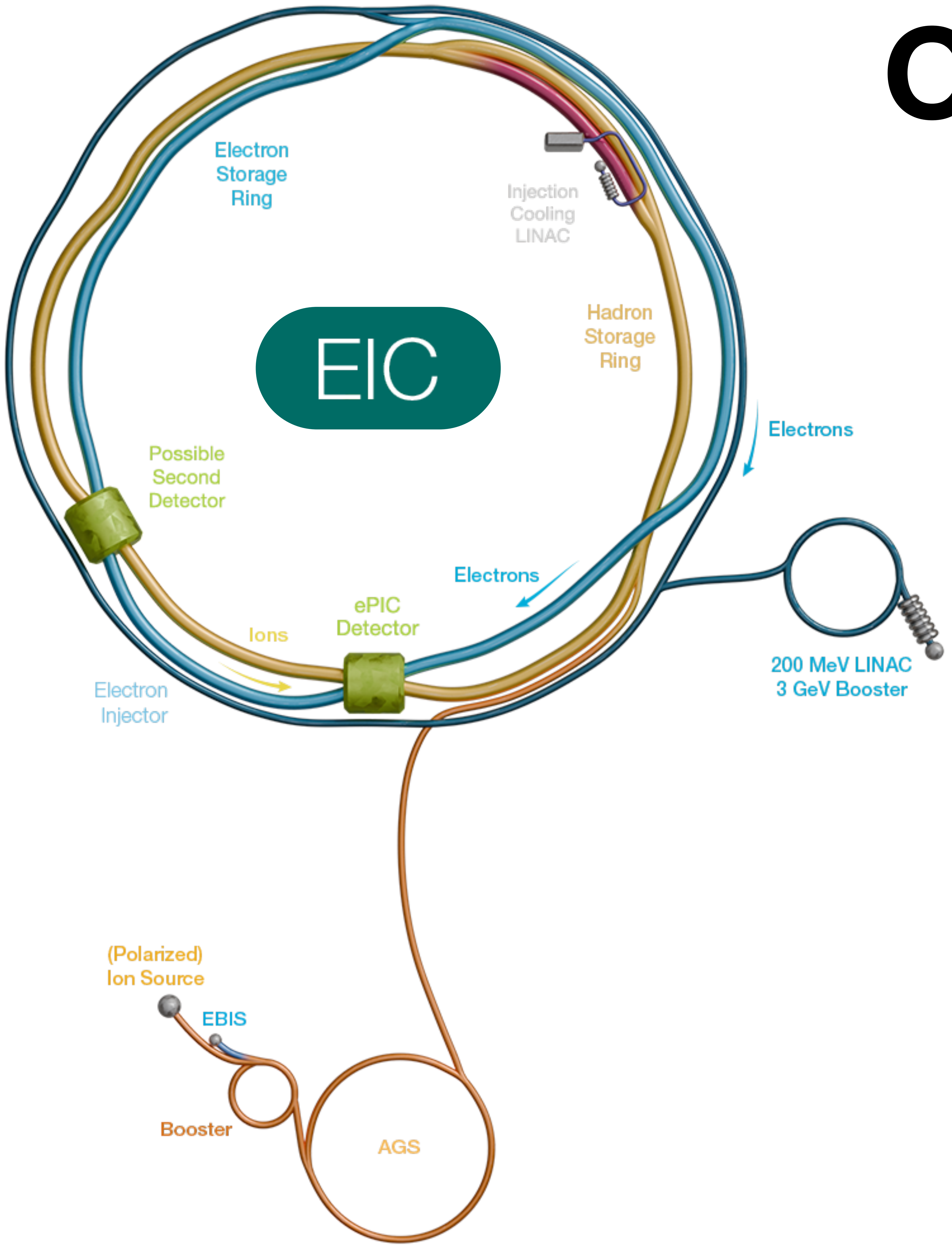


Data region: reasonable agreement between different PDF sets even though they include different datasets. aiming for both accuracy & precision

Extrapolation region hic sunt dracones!

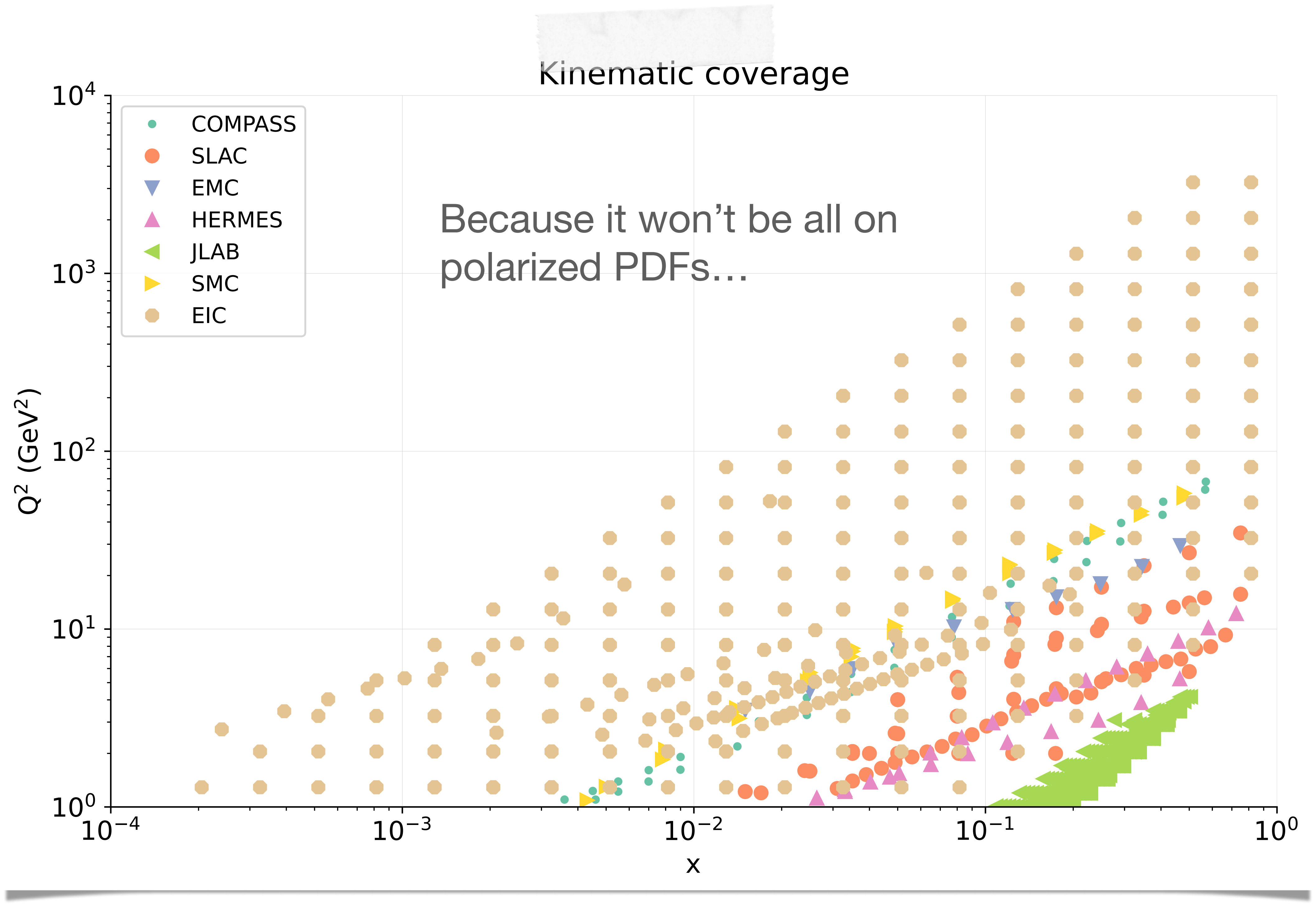


Parton distributions and new physics searches: the Drell–Yan forward–backward asymmetry as a case study [hep-ph] 2209.08115
R. Ball, A. Candido, S. Forte, F. Hekhorn, E. Nocera, J. Rojo, C. Schwan



Opportunities at EIC

- Flavour-tagged structure functions
- The high-x impact



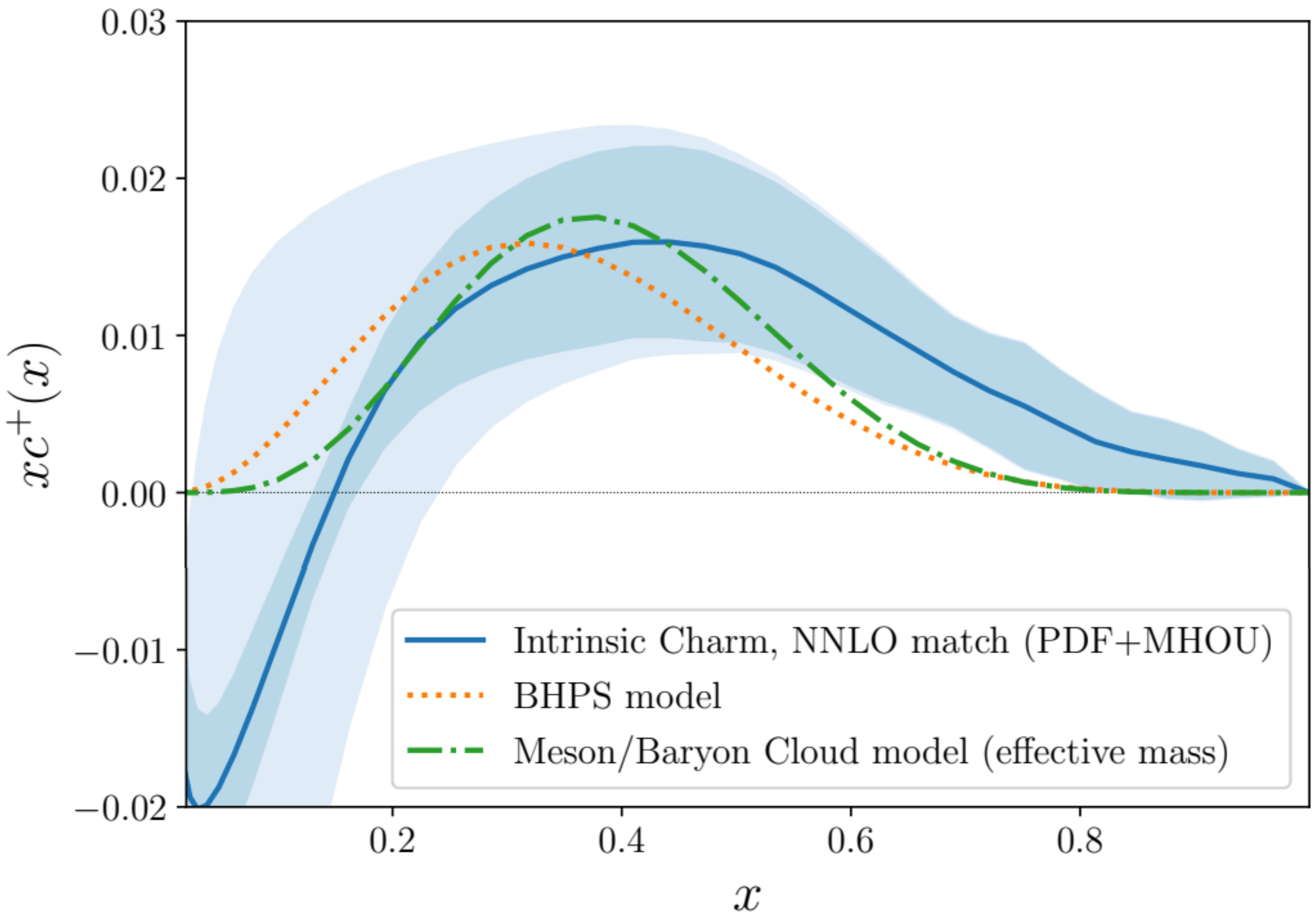
The charm content of the proton: flavour tagged structure functions

PDFs are commonly fitted under the assumption of a proton wave function with no charm content.

We can instead parametrize the PDF above the charm mass and then remove the perturbative component.

- ✓ The PDF describes the data better than when the charm is considered purely perturbative.
- ✓ Non-0 contribution from the charm quark when evolving below the charm threshold!

Can the EIC find more evidence of this non-perturbative component?



THE INTRINSIC CHARM OF THE PROTON

S.J. BRODSKY [1]
Stanford Linear Accelerator Center, Stanford, California 94305, USA

and

P. HOYER, C. PETERSON and N. SAKAI [2]
NORDITA, Copenhagen, Denmark

Received 22 April 1980

Recent data give unexpectedly large cross-sections for charmed particle production at high x_F in hadron collisions. This may imply that the proton has a non-negligible uudc$\bar{c}$ Fock component. The interesting consequences of such a hypothesis are explored.

Evidence for intrinsic charm quarks in the proton [hep-ph] 2208.08372
R. Ball, A. Candido, JCM, S. Forte, T. Giani, F. Hekhorn, K. Kurdashkin, G. Magni, E. Nocera, J. Rojo

The charm content of the proton: flavour tagged structure functions

PDFs are commonly fitted under the assumption of a proton wave function with no charm content.

We can instead parametrize the PDF above the charm mass and then remove the perturbative component.

✓ The PDF describes the data better than when the charm is considered purely perturbative.

✓ Non-0 contribution from the charm quark when evolving below the charm threshold!

THE INTRINSIC CHARM OF THE PROTON

S.J. BRODSKY [1]
Stanford Linear Accelerator Center, Stanford, California 94305, USA

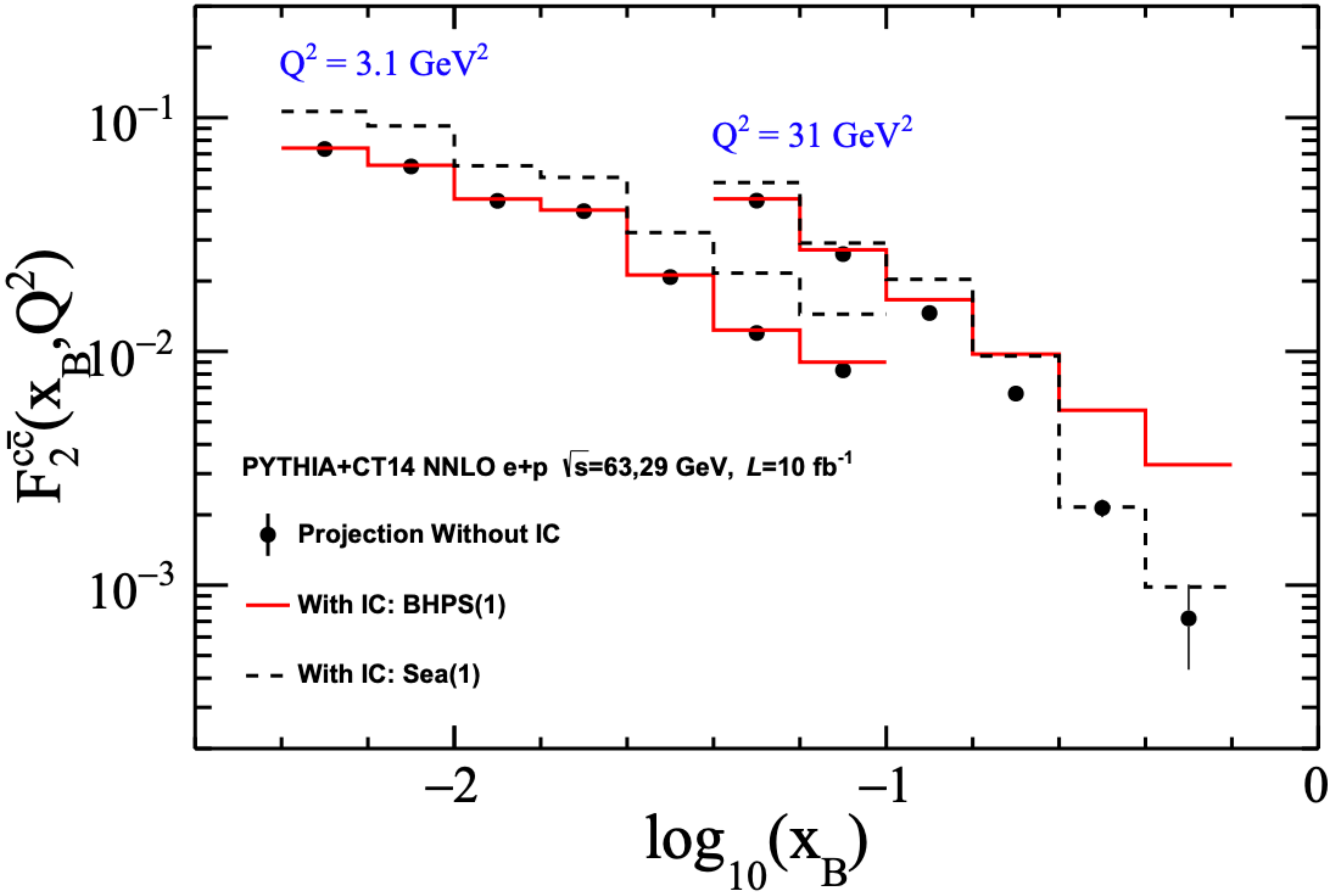
and

P. HOYER, C. PETERSON and N. SAKAI [2]
NORDITA, Copenhagen, Denmark

Received 22 April 1980

Recent data give unexpectedly large cross-sections for charmed particle production at high x_F in hadron collisions. This may imply that the proton has a non-negligible uudc$\bar{c}$ Fock component. The interesting consequences of such a hypothesis are explored.

Can the EIC find more evidence of this non-perturbative component?



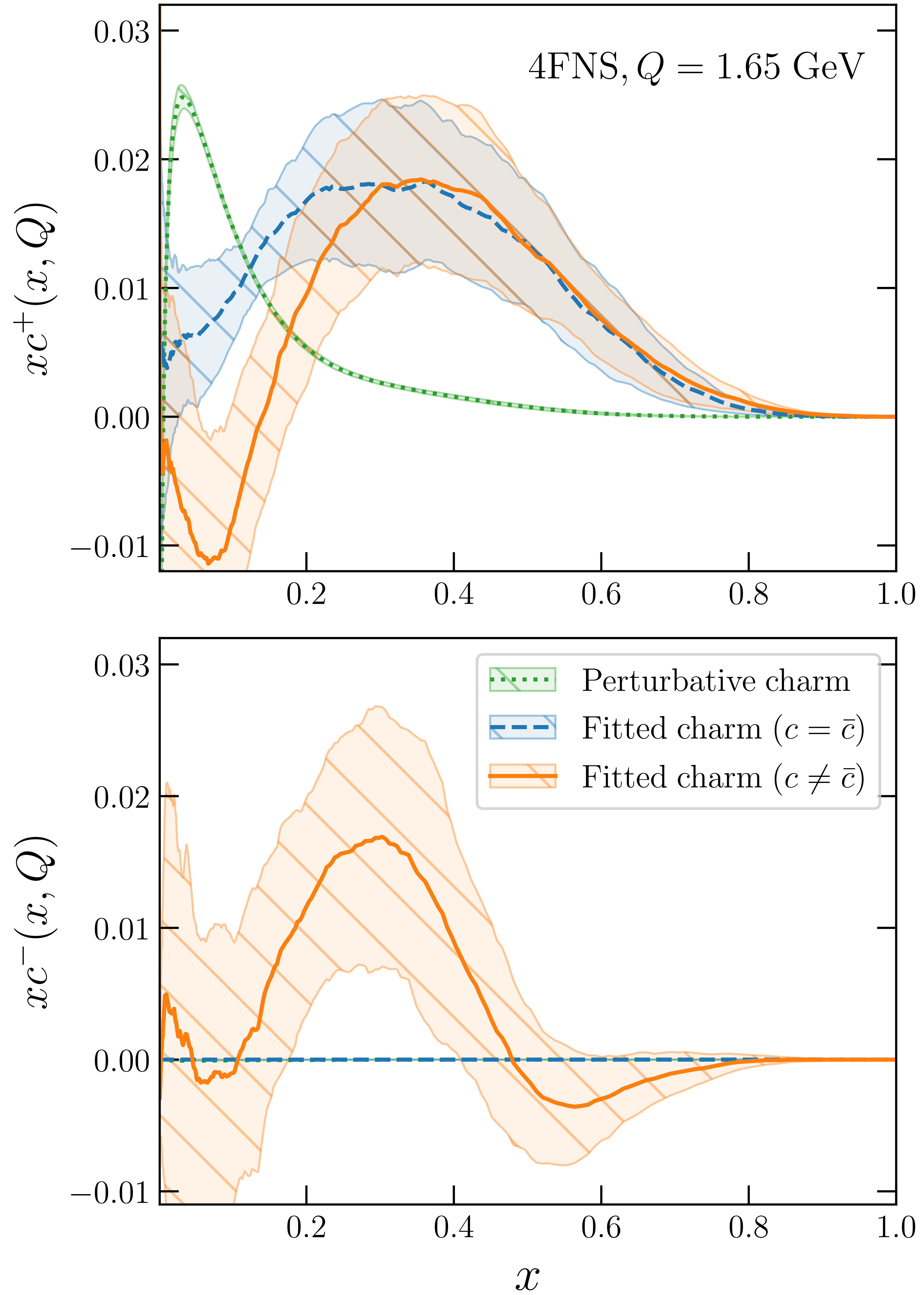
From [arXiv: 2107.05632]

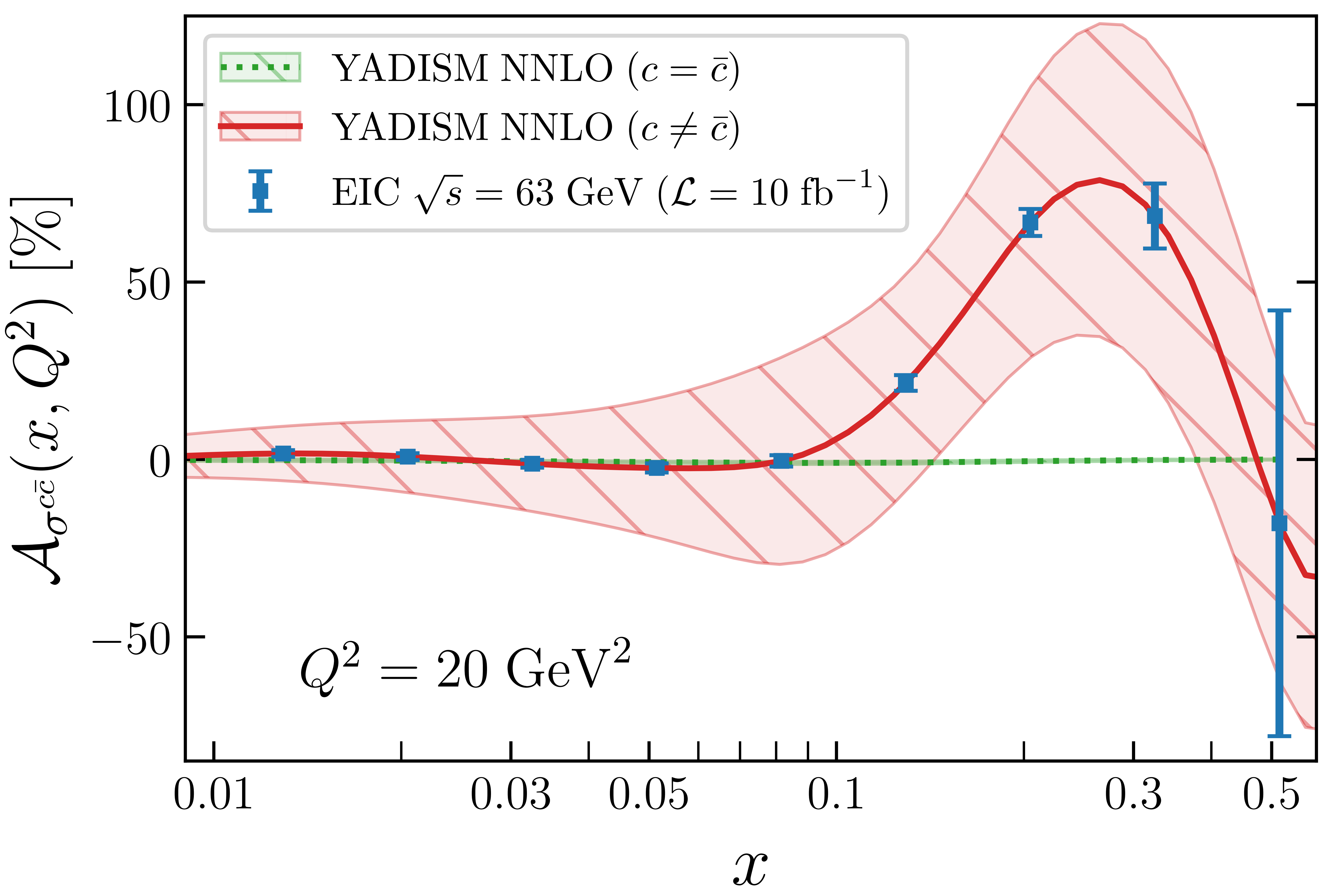
Evidence for intrinsic charm quarks in the proton [hep-ph] 2208.08372
R. Ball, A. Candido, JCM, S. Forte, T. Giani, F. Hekhorn, K. Kurdashkin, G. Magni, E. Nocera, J. Rojo

Intrinsic charm asymmetry in the EIC

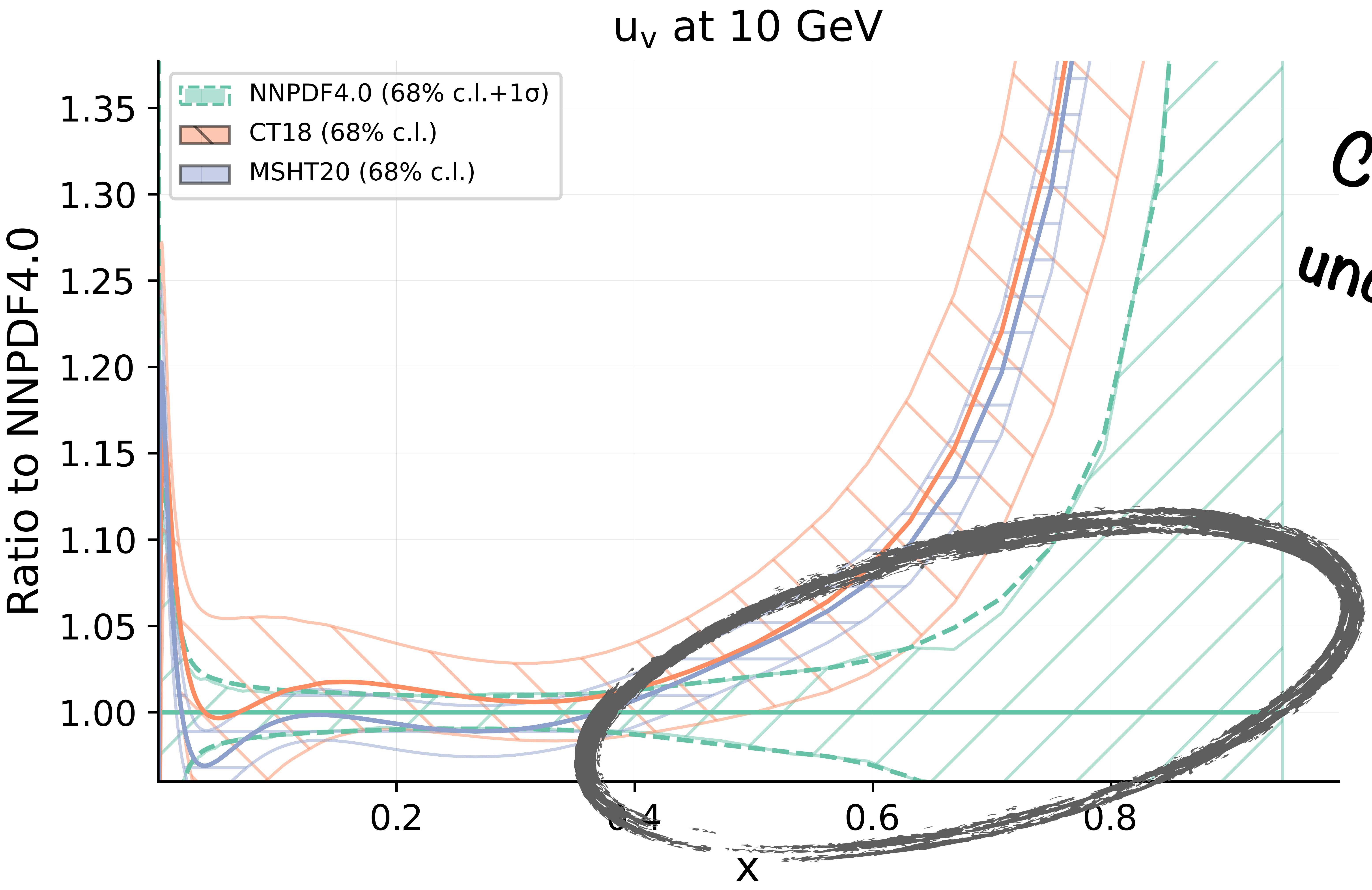
through favour-tagged structure functions

Intrinsic charm quark valence distribution of the proton [hep-ph] 2311.00743
R. Ball, A. Candido, JCM, S. Forte, T. Giani, F. Hekhorn, G. Magni, E. Nocera,
J. Rojo, R. Stegeman





And what about the large x region?

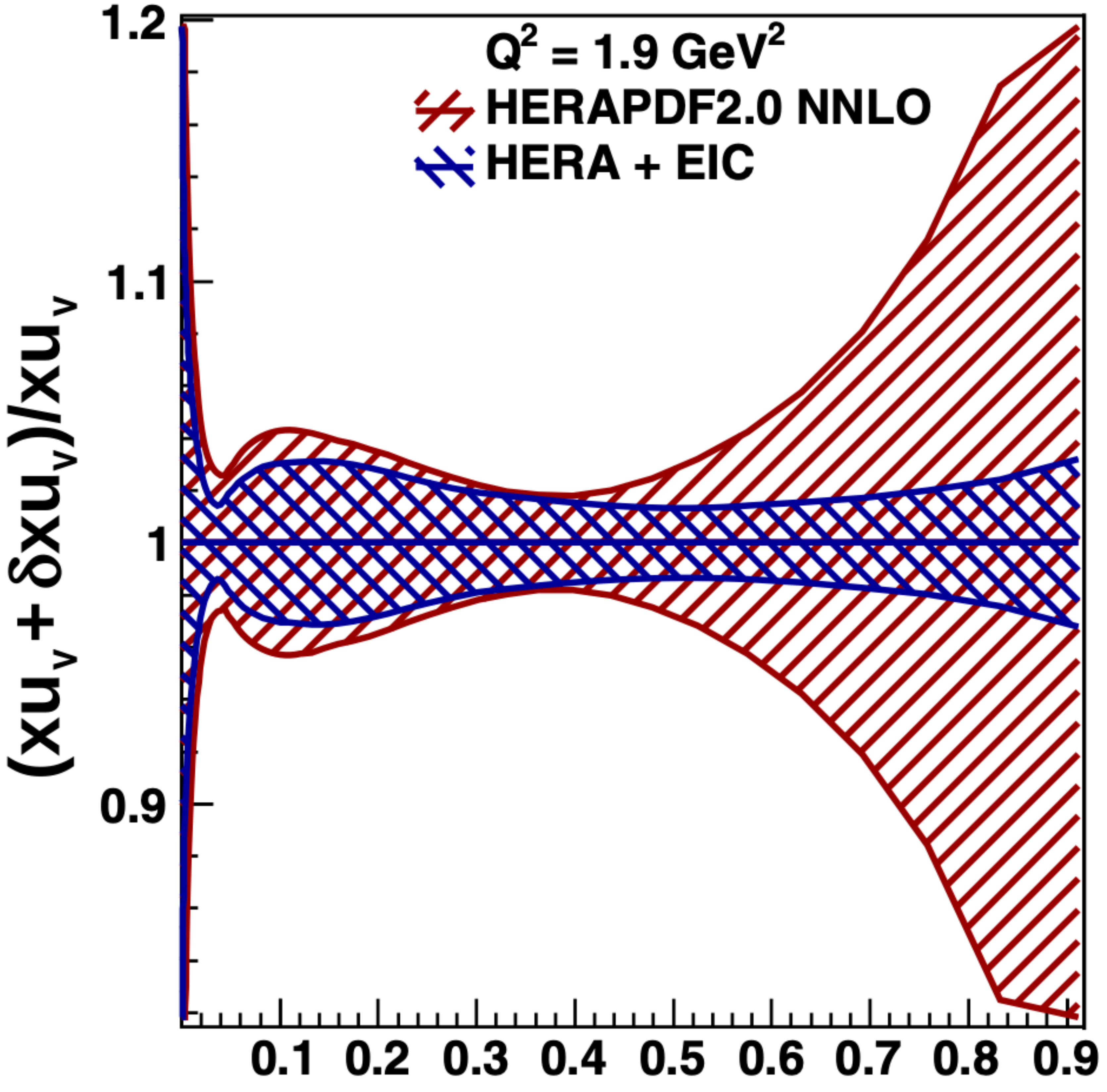


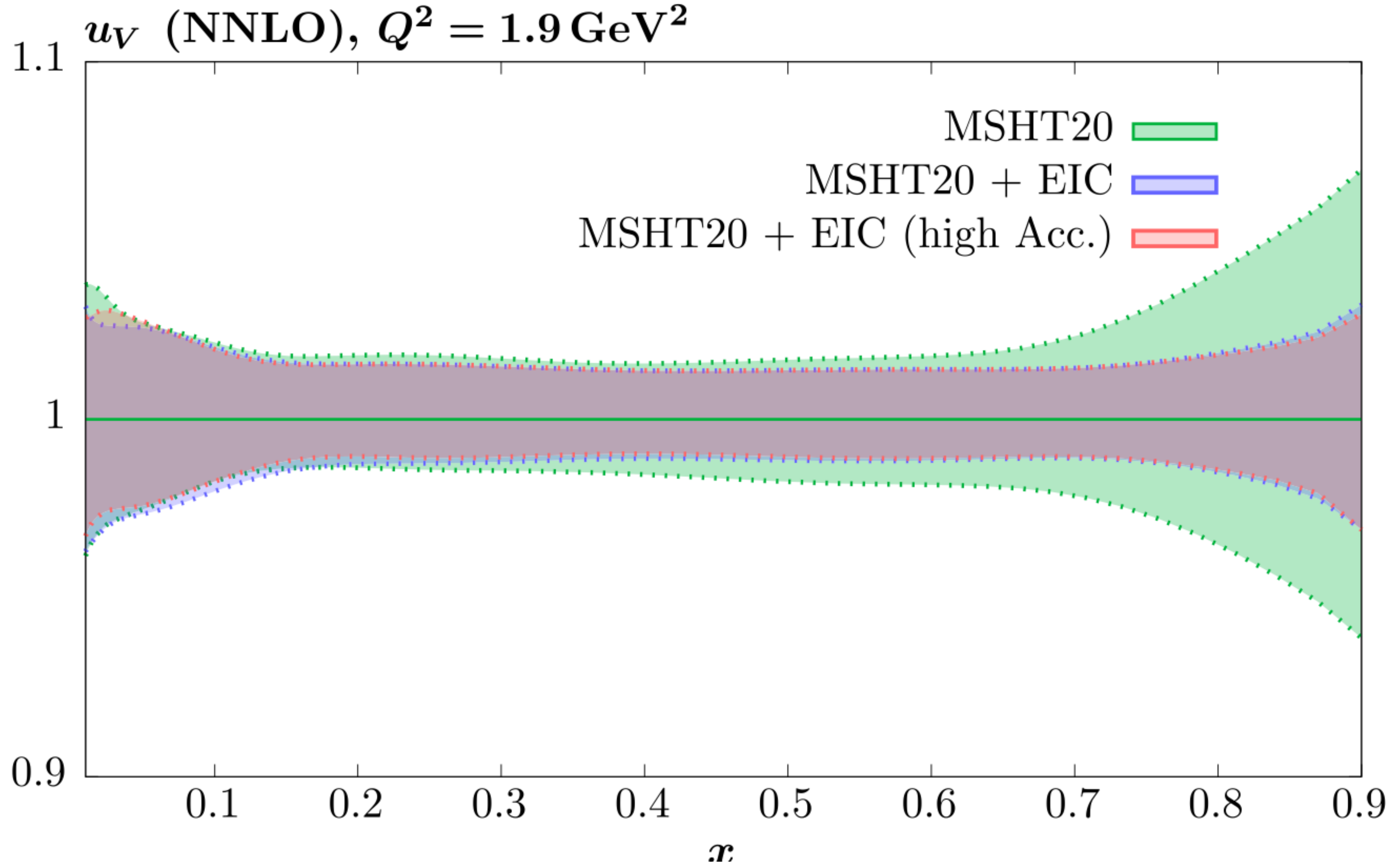
Can we get it under control?

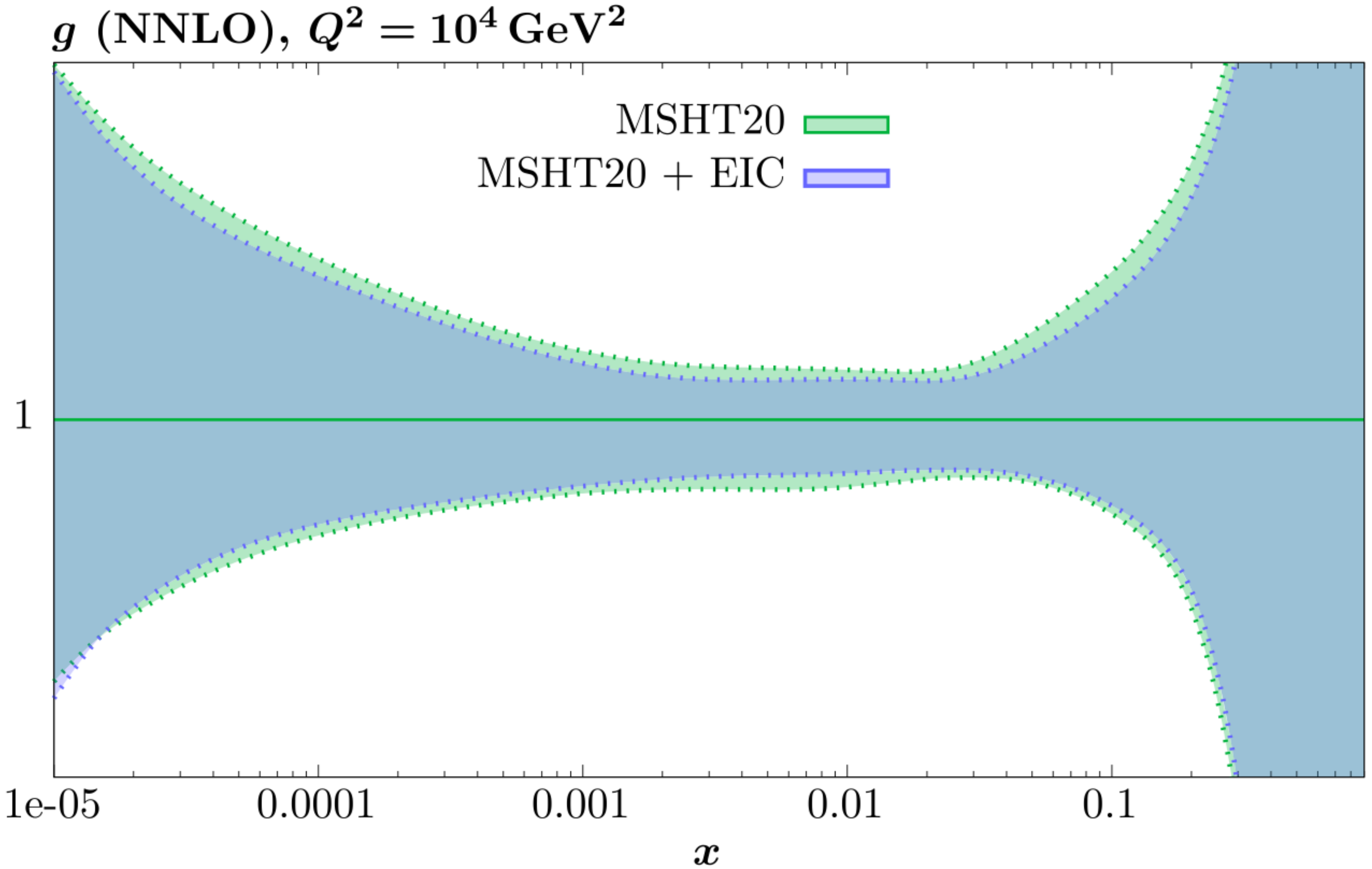
Silently tiptoeing over higher-twist corrections which may be relevant for this large-x low-Q region…

but every crisis is an opportunity: very precise probe of higher-twist effects!

Impact of EIC in global fits



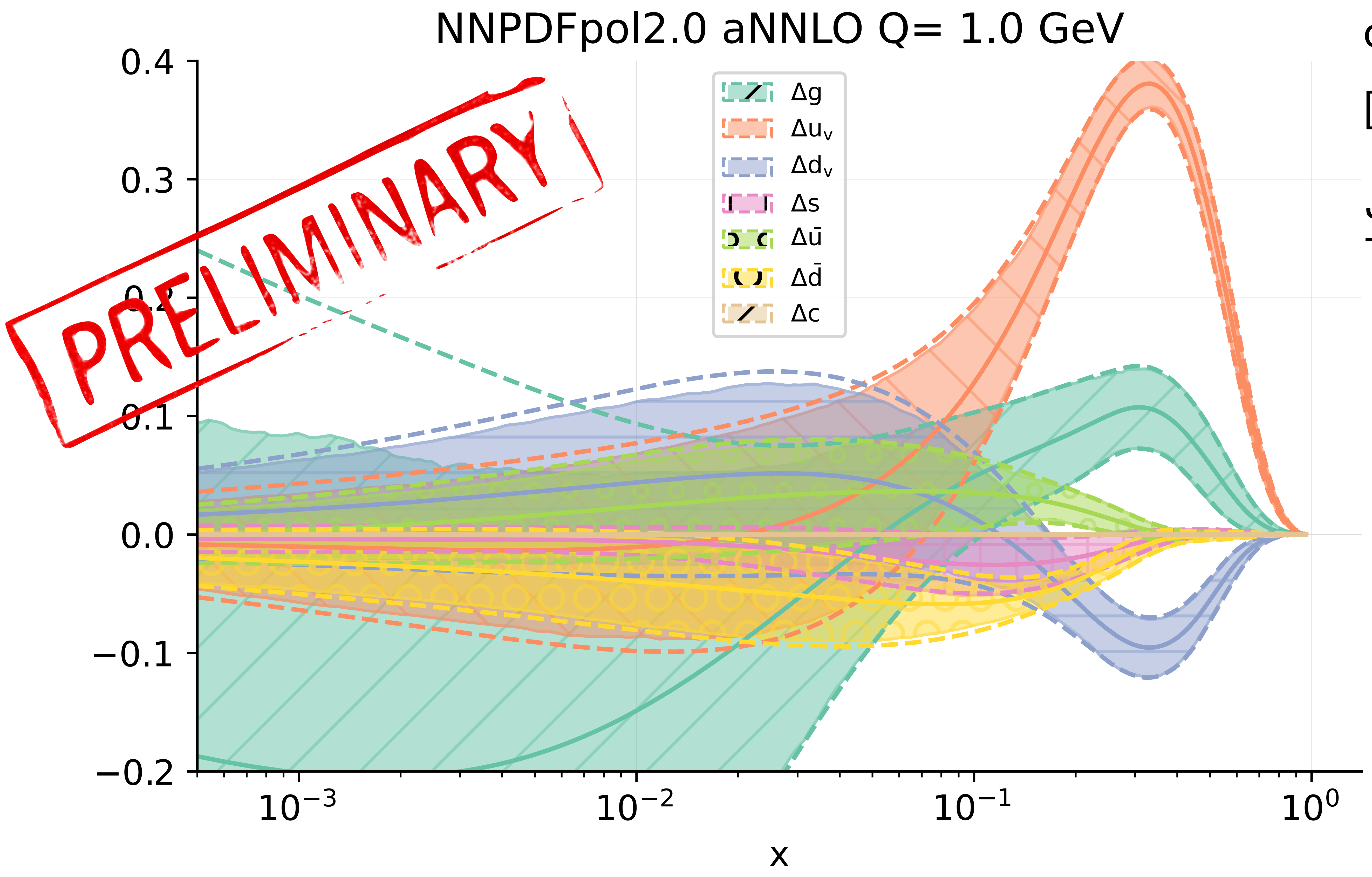




Impact of inclusive electron ion collider data on collinear parton distributions
[hep-ph] 2309.11269
N. Armesto, T. Cridge, F. Giuli, L. Harland-Lang, P. Newman, B. Schmookler, R. Thorne, K. Wichmann

(longitudinally) polarized PDFs

the upcoming NNPDFpol 2.0!



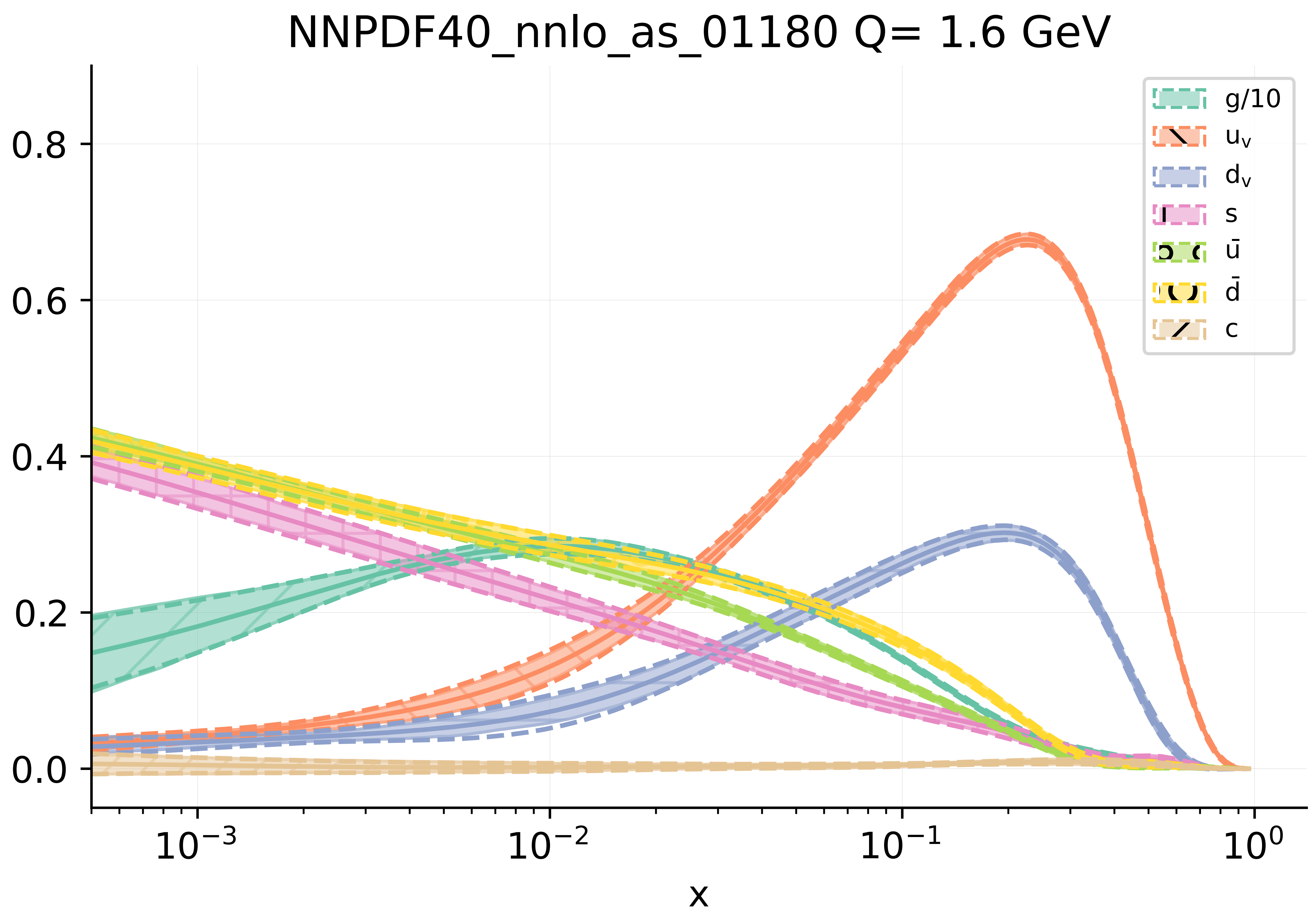
NNPDFpol2.0: a first global determination of polarised parton distributions at NNLO accuracy with theory uncertainties

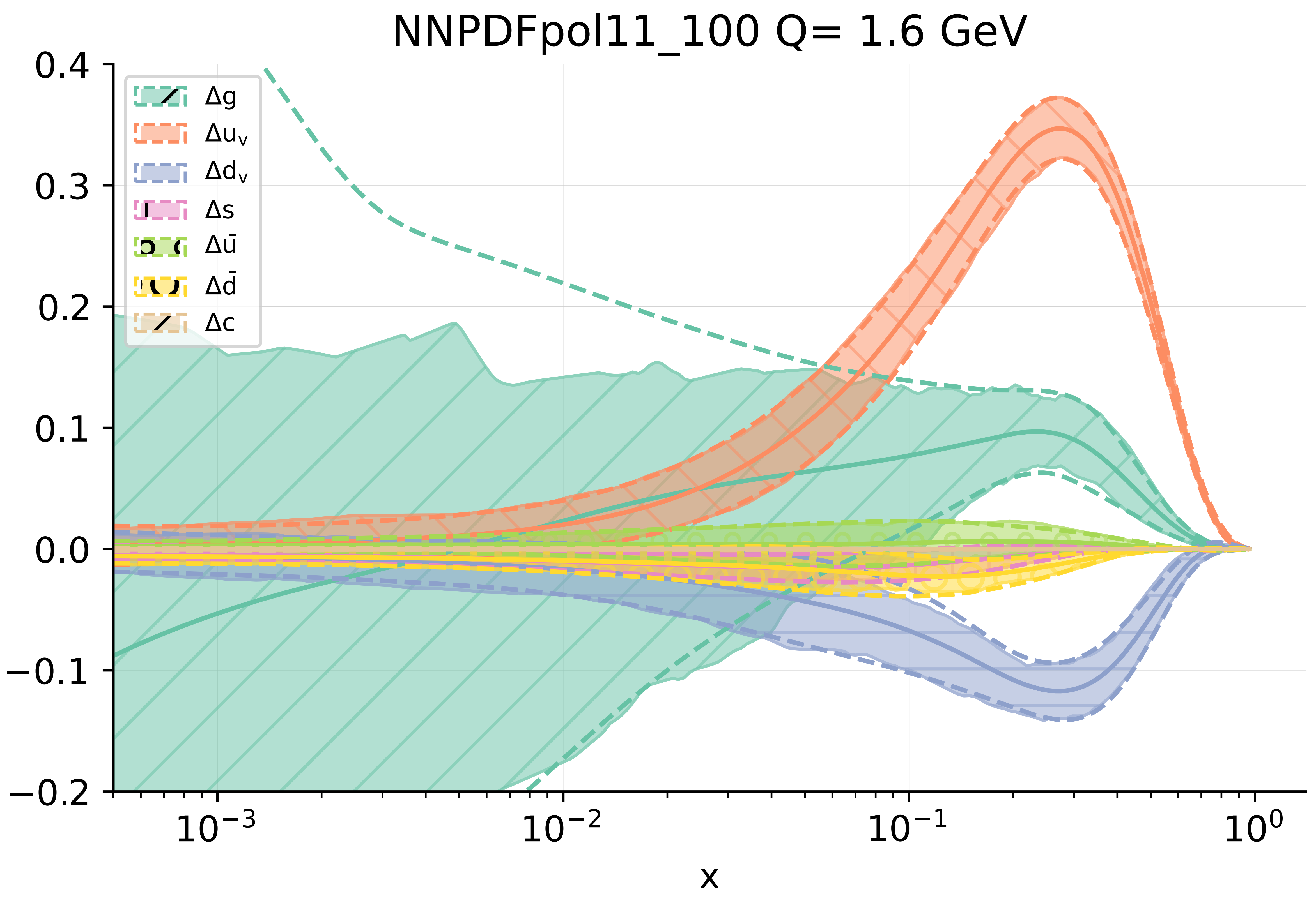
[hep-ph] (in the coming weeks)

JCM, T. Hasenack, F. Hekhorn, G. Magni, E. Nocera,
T. Rabemananjara, J. Rojo, G. van Seeventer,

Polarized vs unpolarized

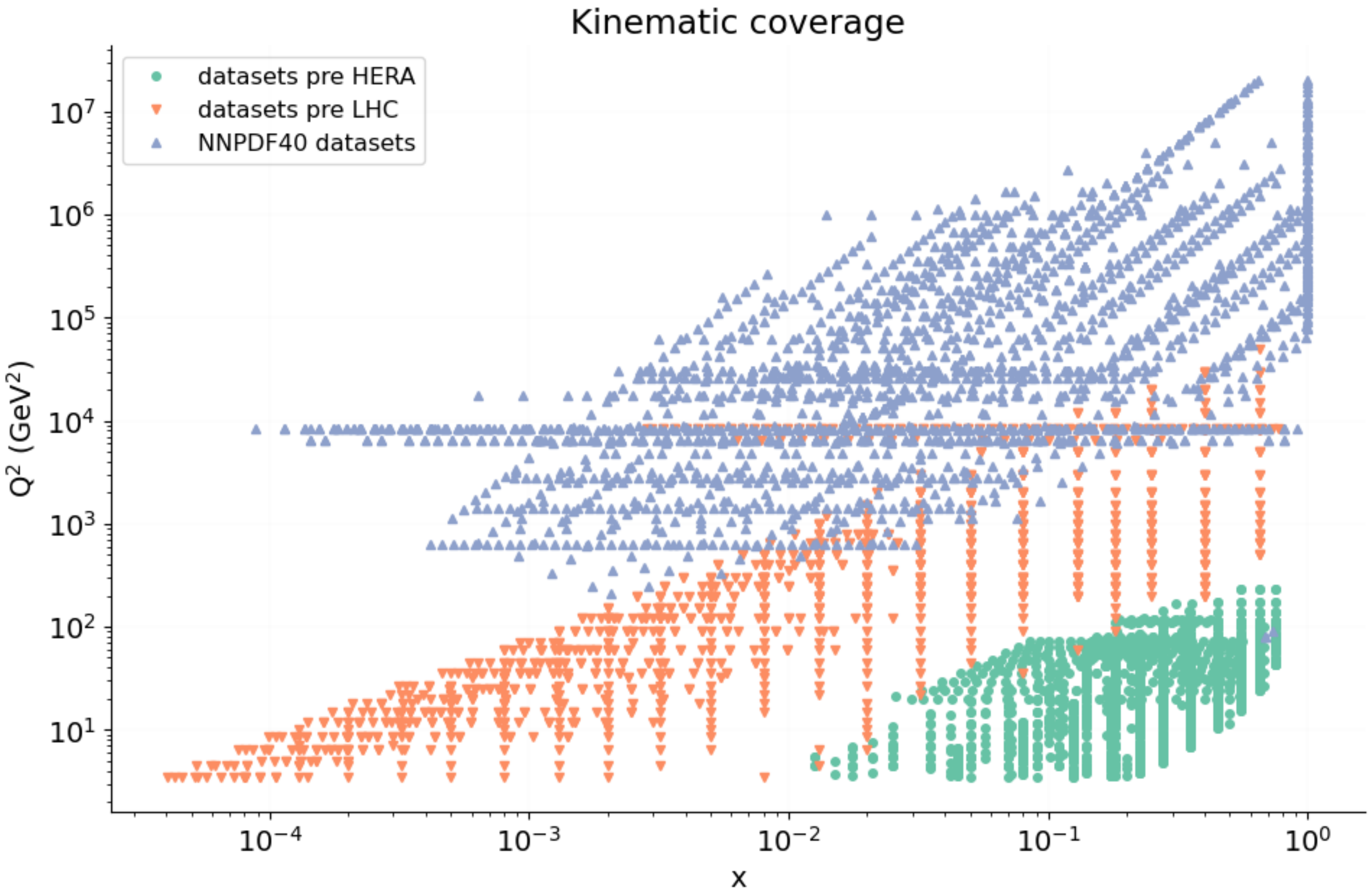
While, thanks to the large amout of data available, longitudinal unpolarized PDFs are reaching the target of %-level uncertainties. The knowledge of unpolarized PDFs is still far from there.

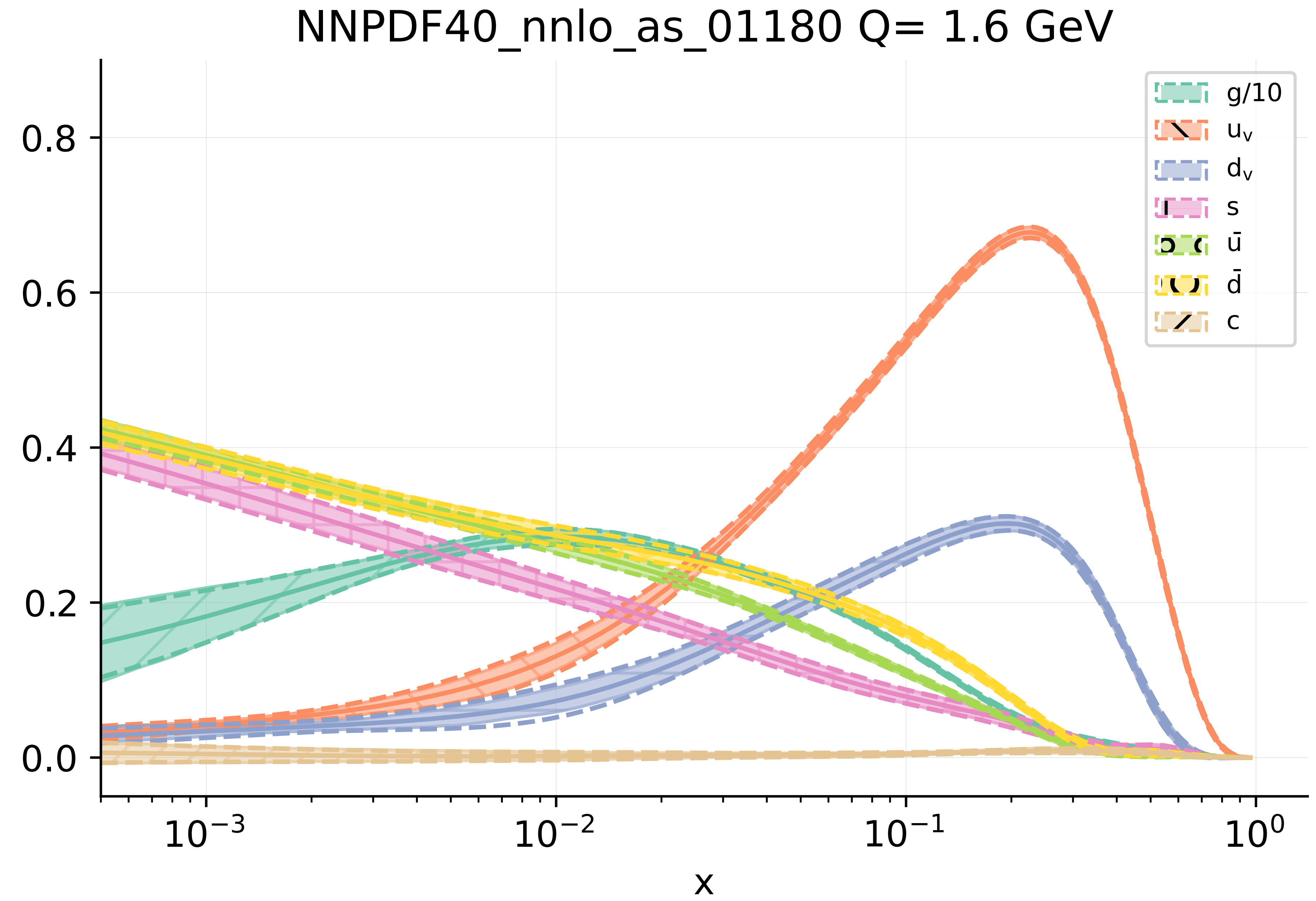


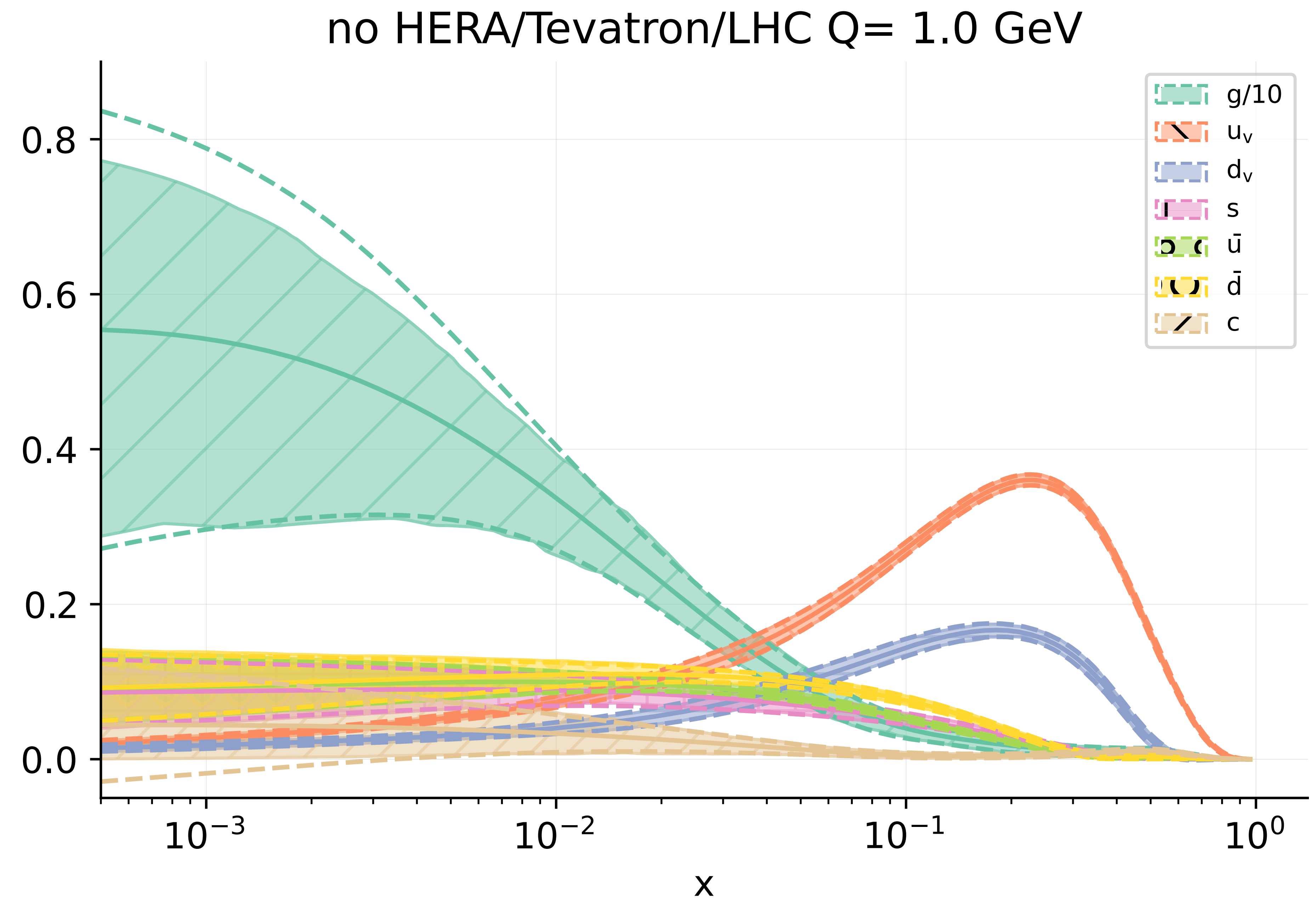


Polarized vs unpolarized

Will the EIC be the HERA of Polarized PDFs?







Why do we care about (polarized) PDFs?

Longitudinal PDFs are spin averaged, i.e., the information of the spin content of the proton is lost. But the proton has spin 1/2… how is it distributed?

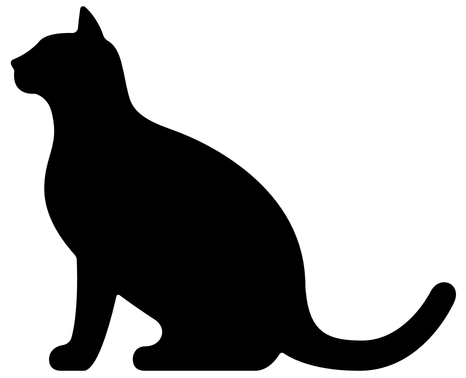
The integral of the the polarised PDFs is related to the total proton spin:

$$S_P \equiv \frac{1}{2} = \frac{1}{2}\sum_q \eta_q(Q^2) + \eta_g(Q^2) + \langle \sum_q \mathscr{L}_q + \mathscr{L}_g \rangle (Q^2) \qquad \eta_k = \int_0^1 dx \Delta f_k(x)$$

From HERMES, COMPASS, EMC, SMC… it can only contribute to a fraction of the total proton spin

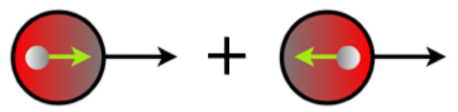
Not well constrained and compatible with 0

Where does the proton spin come from?
How are partons distributed inside the proton (and other nuclei)?



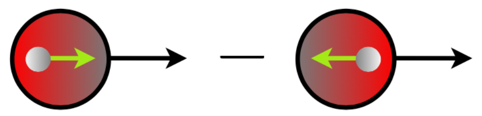
Spin averaged (usual) PDFs

$$q_k(x) \equiv q_k^{\uparrow\uparrow}(x) + q_k^{\uparrow\downarrow}(x)$$



Longitudinally polarised PDFs

$$\Delta q_k(x) \equiv q_k^{\uparrow\uparrow}(x) - q_k^{\uparrow\downarrow}(x)$$



Plenty of information necessary for polarized PDFs has been made available in the last few years:

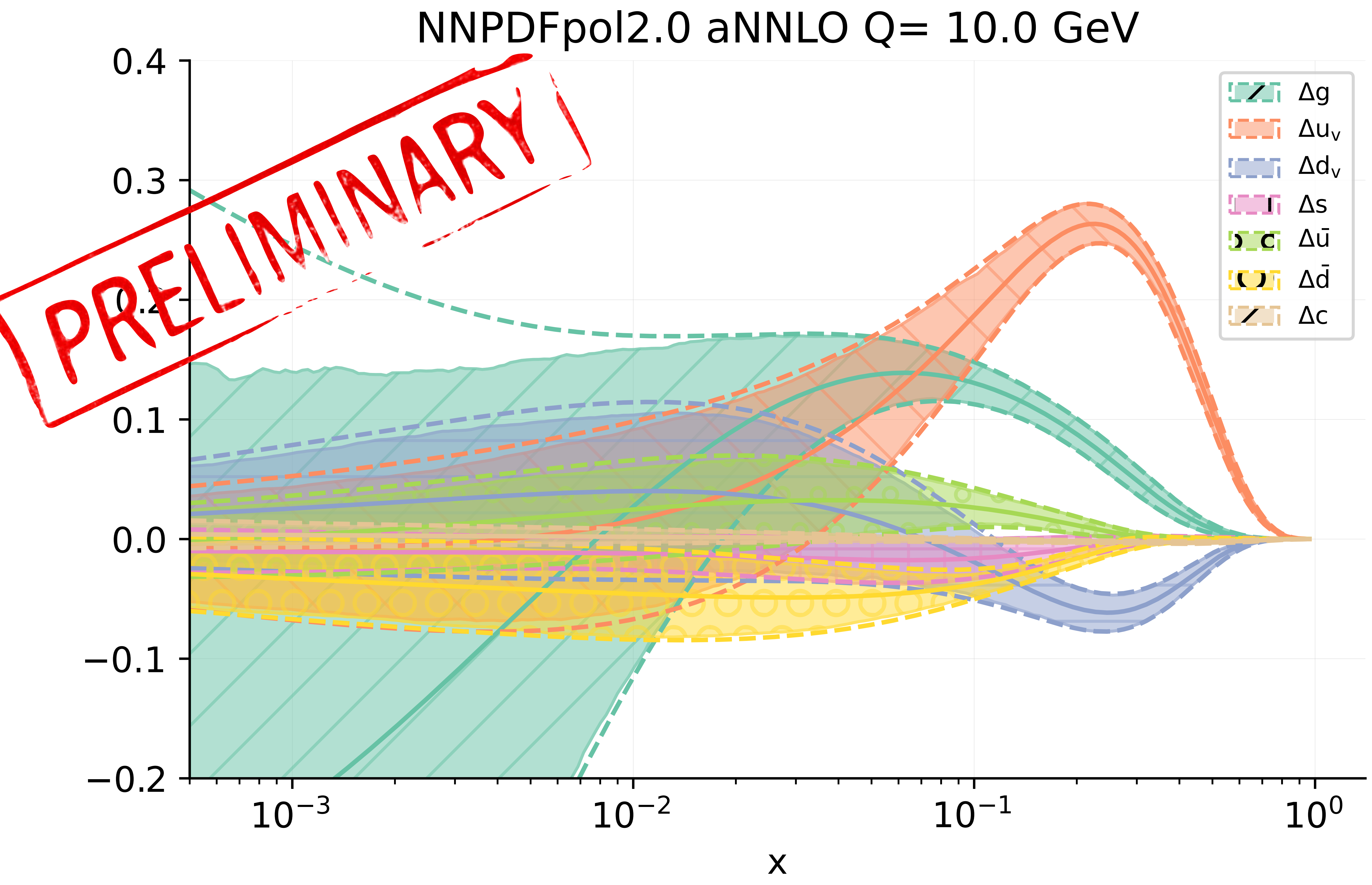
- NNLO corrections for polarized W production [hep-ph] 2101.02214
- NNLO polarized DGLAP and matching conditions [hep-ph] 1409.5131 1506.04517 2107.06267 2111.12401 2211.15337
- FONLL variable flavour number scheme [hep-ph] 2401.10127

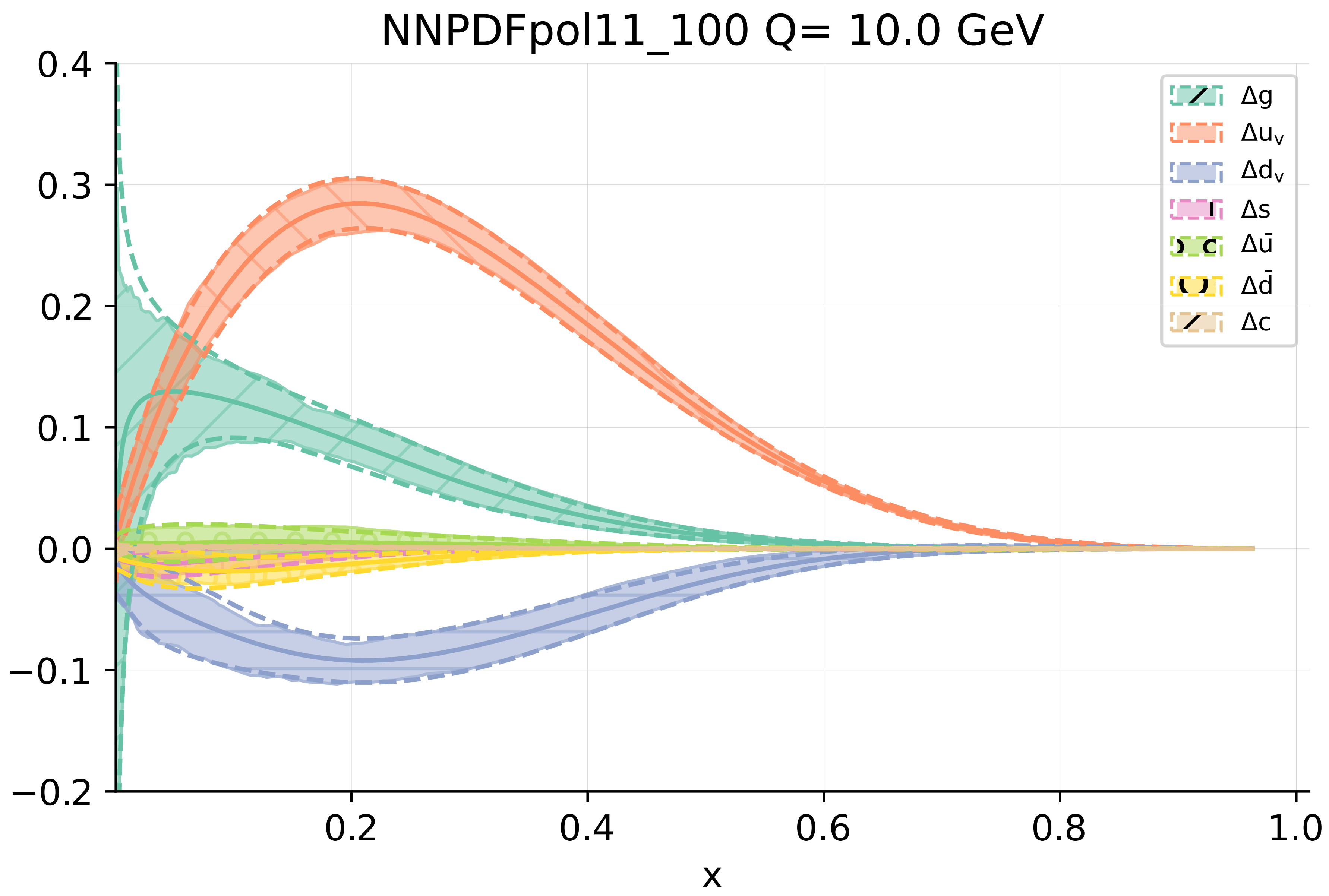
PDFs *for* EIC: NNPDFpol2.0!

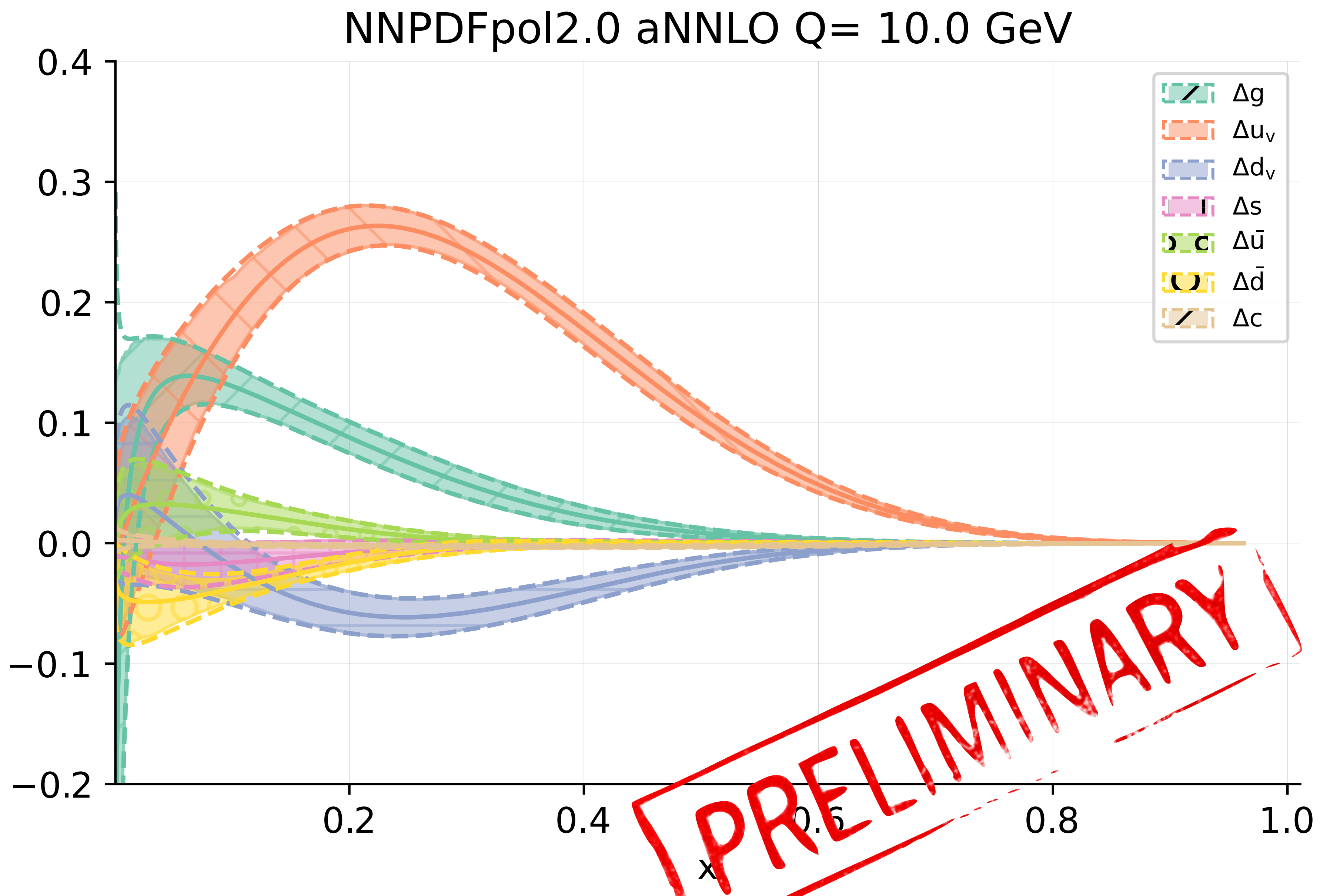
1. Global dataset:
   - ➡ Polarized DIS from CERN, HERA, JLab and SLAC
   - ➡ STAR: W boson production and Jet/Dijet production
2. Missing Higher Order Uncertainties considered as a 7pt factorization and renormalization scale variations.
3. NNLO accurate predictions for (almost) all observables in the form of PineAPPL interpolation grids
   - ➡ DIS at NNLO predictions computed with YADISM ([hep-ph] 2401.15187 and references therein)
   - ➡ W boson production at NNLO thanks to H. T. Li for access to the code from [hep-ph] 2101.02214
   - ➡ Jet/Dijet production at NLO, with MHOU included as an extra source of unc. thanks to W. Vogelsang Li for access to the code from [hep-ph] 0404057
4. Renewed fitting methodology: more faithful uncertainty estimation (future tested methodology!), optimization of the hyperparameters of the Machine Learning methodology, GPU computing, open-source framework.

PDFs *for* EIC: NNPDFpol2.0!

To be published in the coming weeks! Stay tuned!





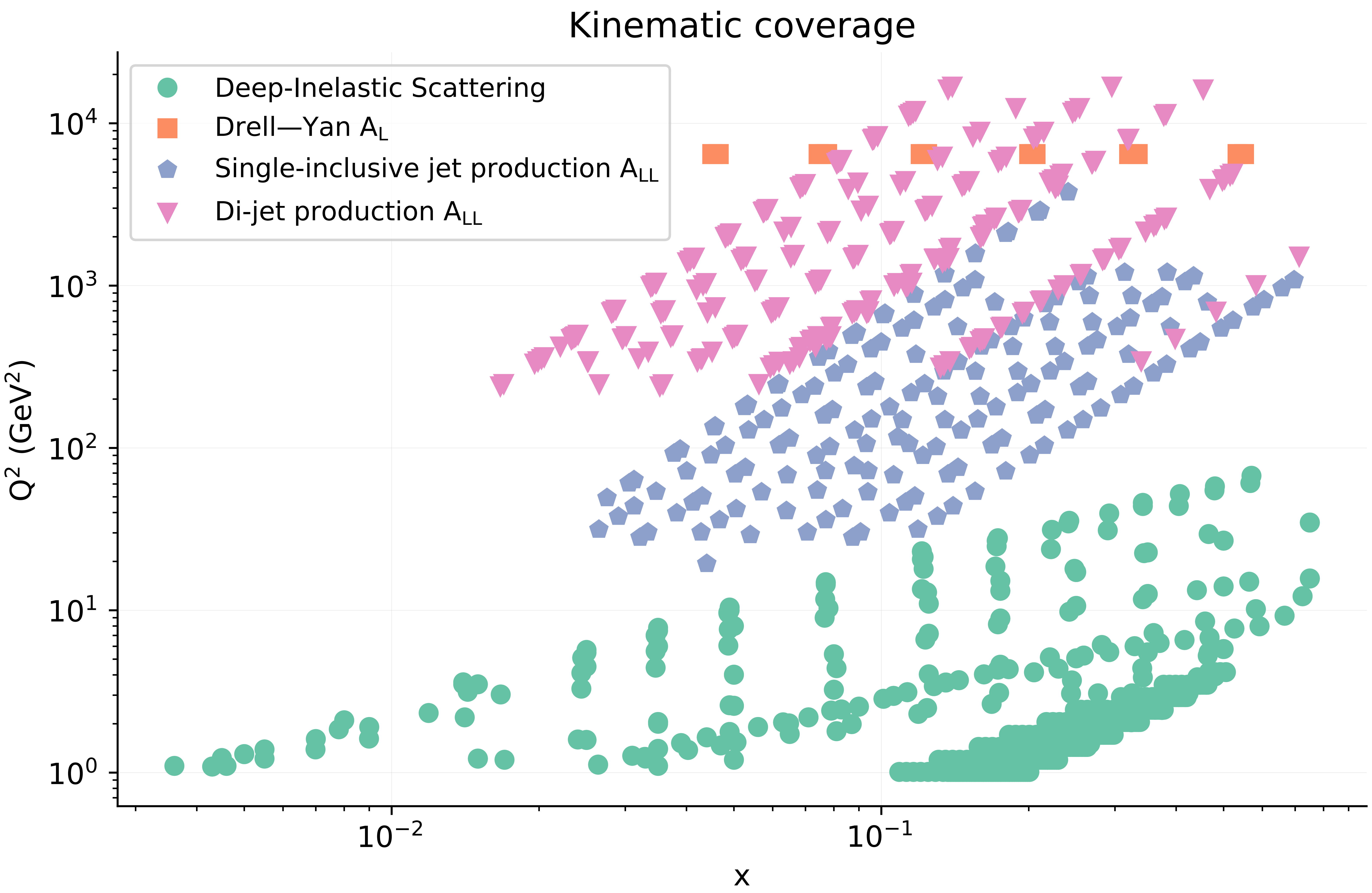


Kinematic Coverage of NNPDFpol 2.0

- Extended kinematic coverage with respect to previous PDF sets
- Total number of datapoints: **951**
- **Sensitivity** down/up to $x \sim (4 \cdot 10^3,\, 0.75)$
- **Kinematic cuts** to minimize non-perturbative QCD effects and possible higher-twist corrections:

$$Q^2 \geq 1\ \mathrm{GeV}^2$$

$$W^2 = M^2 - Q^2\left(1 - \frac{1}{x}\right) \geq 4\ \mathrm{GeV}^2$$

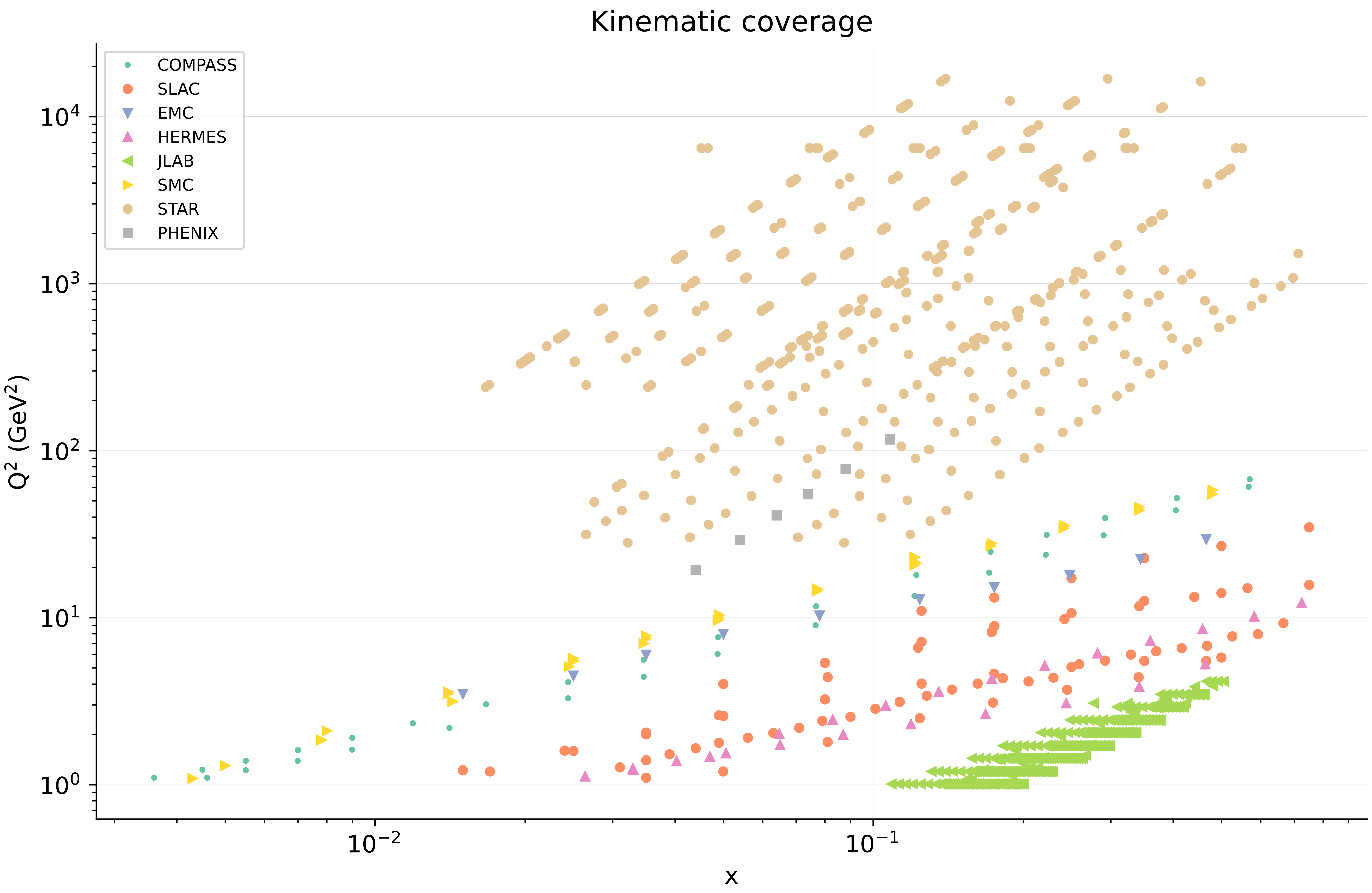


Kinematic Coverage of NNPDFpol 2.0

- Extended kinematic coverage with respect to previous PDF sets
- Total number of datapoints: **951**
- **Sensitivity** down/up to $x \sim (4 \cdot 10^{3},\ 0.75)$
- **Kinematic cuts** to minimize non-perturbative QCD effects and possible higher-twist corrections:

$$Q^2 \geq 1\ \mathrm{GeV}^2$$

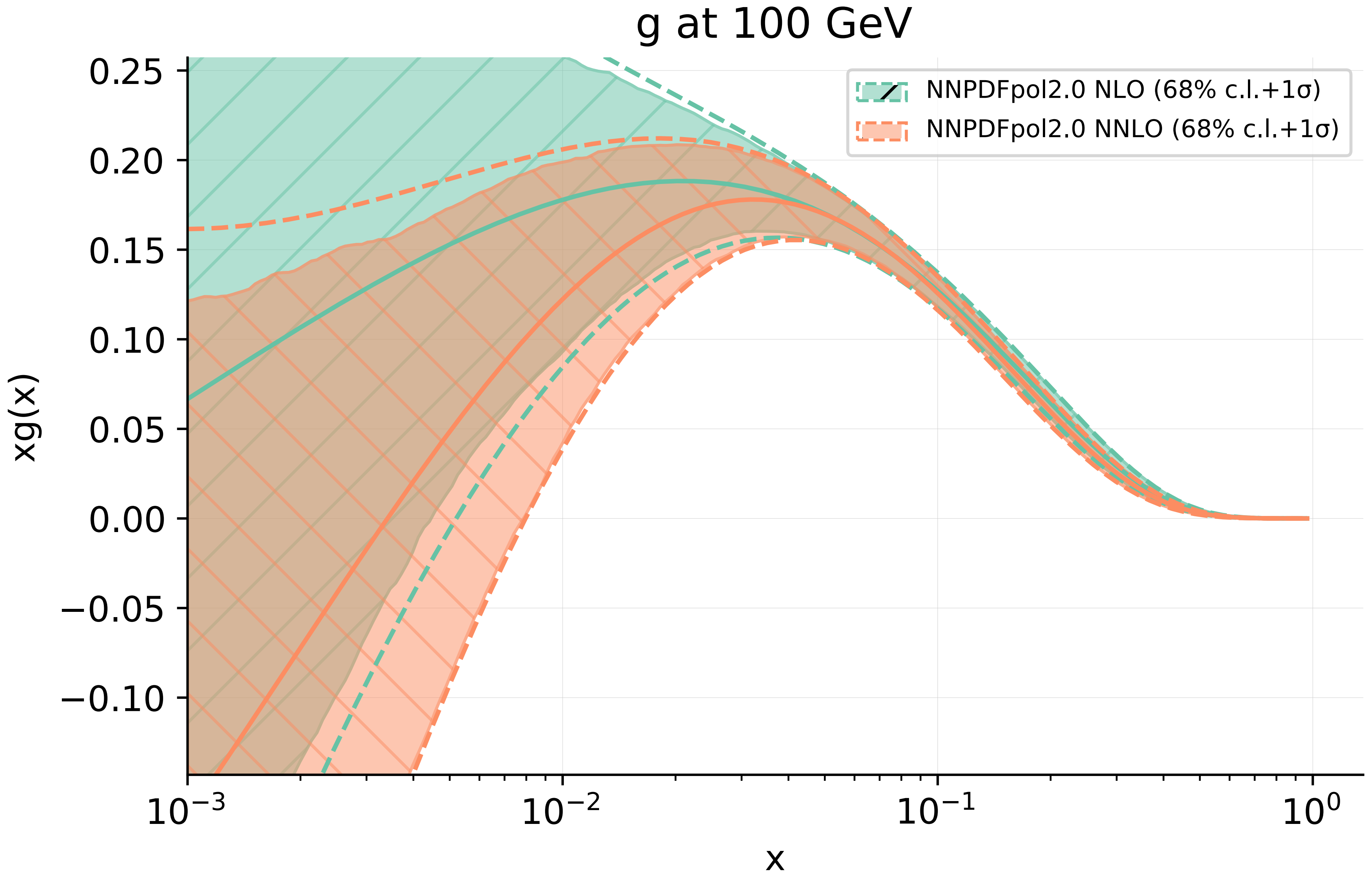
$$W^2 = M^2 - Q^2 \left(1 - \frac{1}{x}\right) \geq 4\ \mathrm{GeV}^2$$

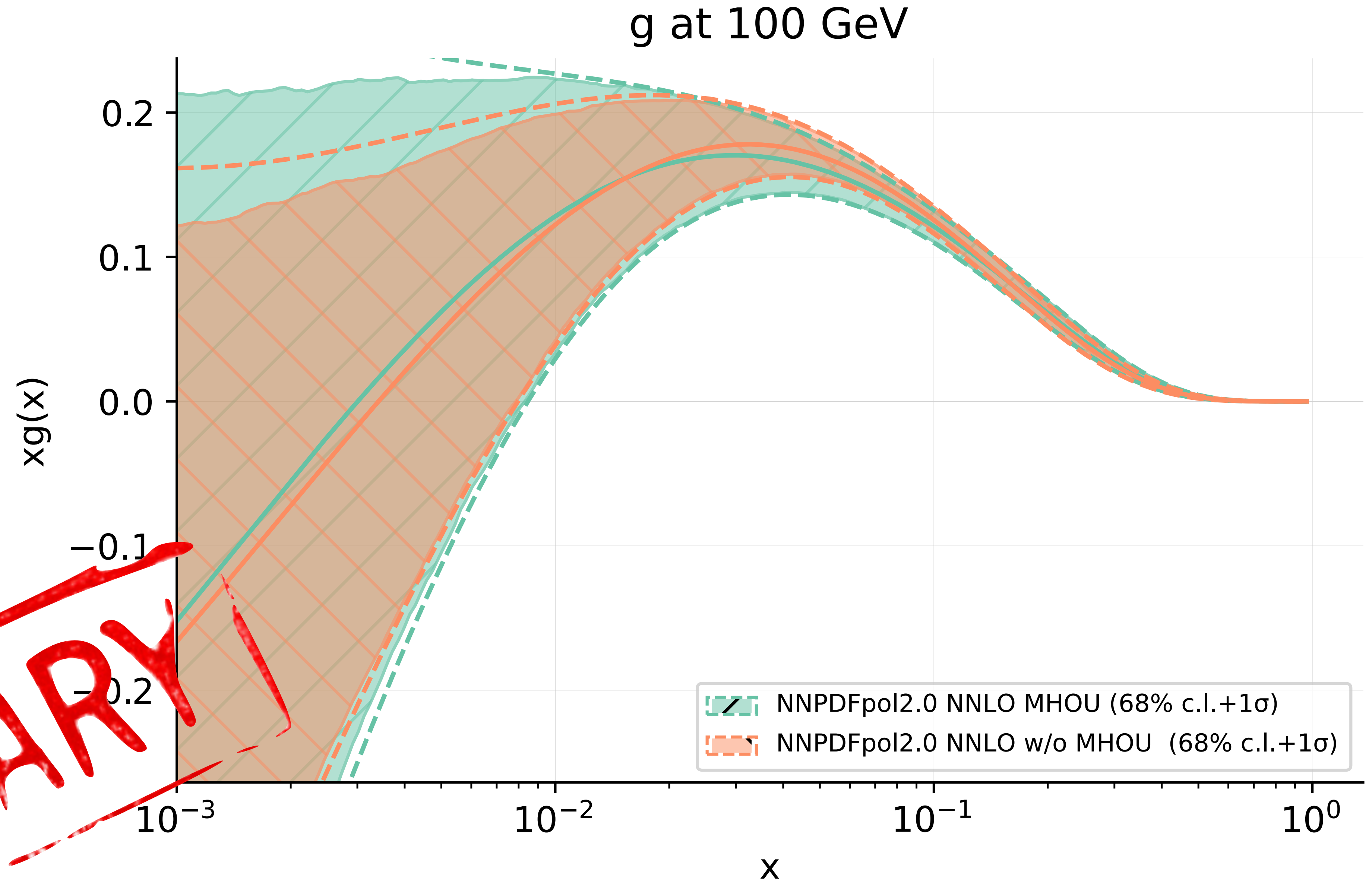


NNPDFpol 2.0: the perturbative impact

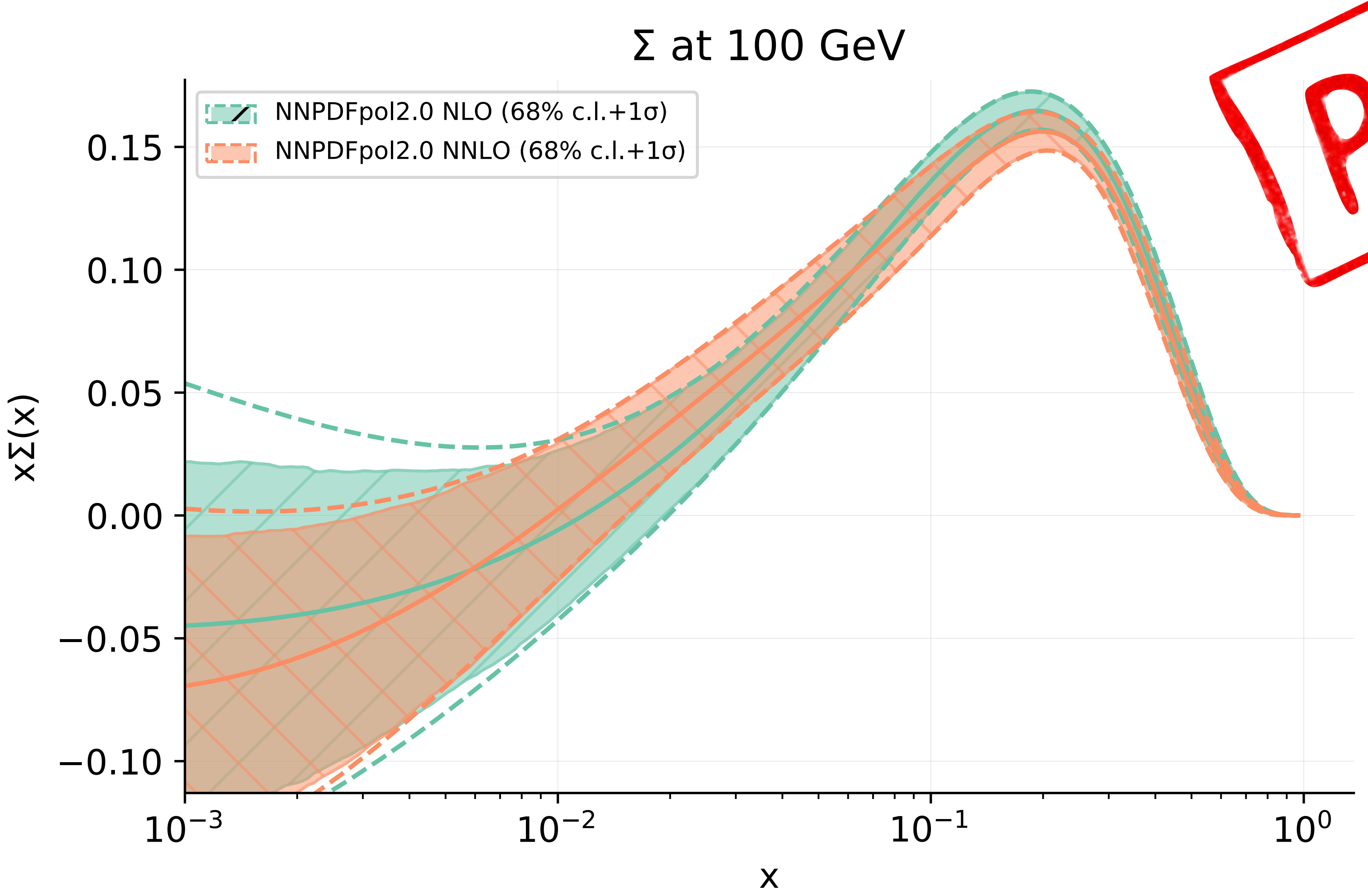
NLO vs NNLO

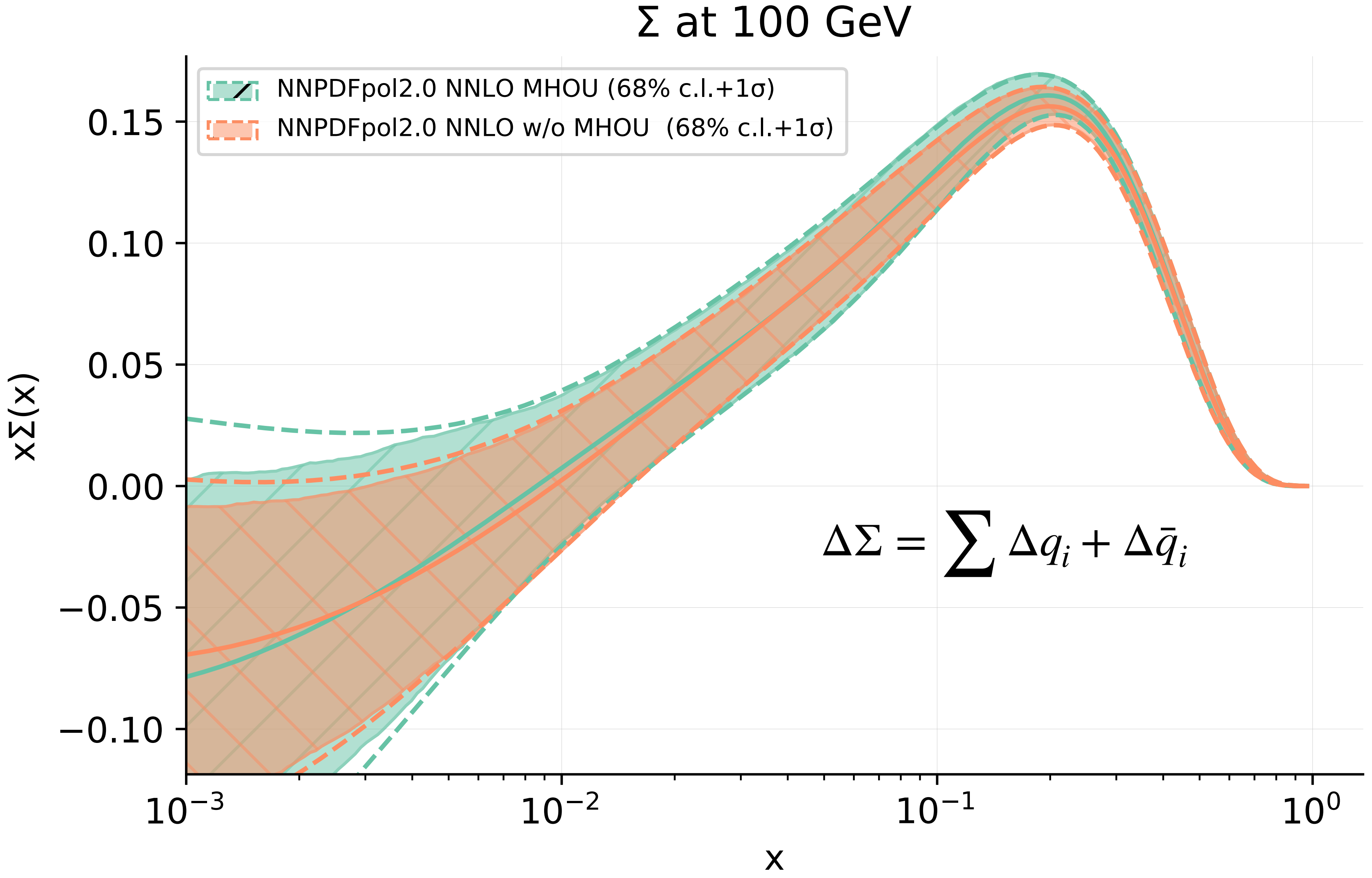
NNLO, with and w/o MHOU





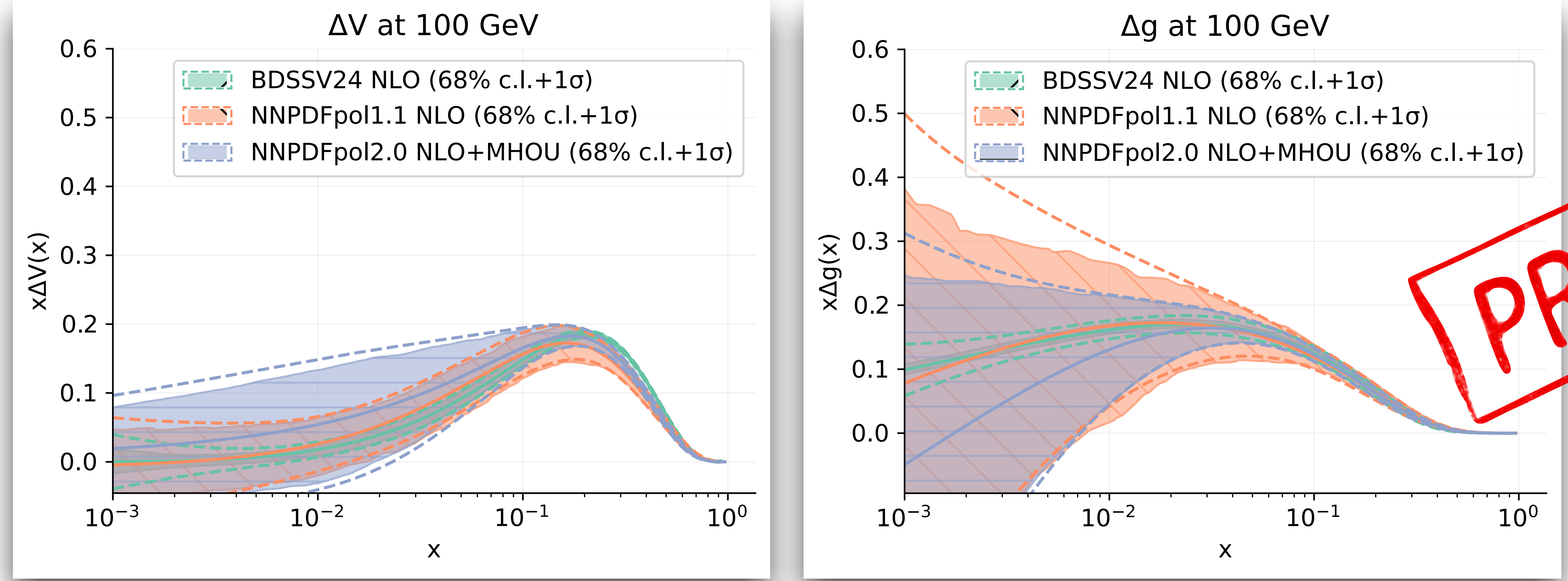
PRELIMINARY



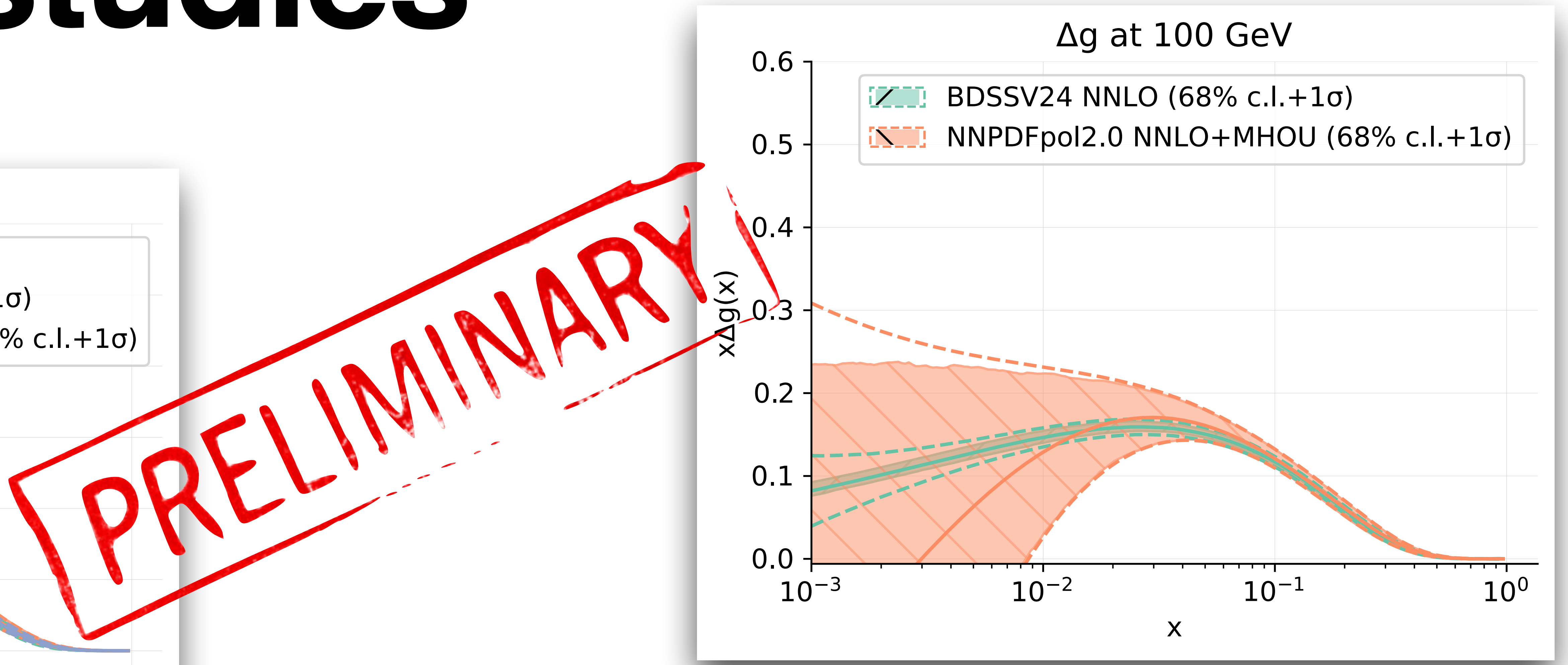


Comparison to other studies

NLO



NNLO



Comparison to NNPDFpol1.1, at NLO.

- Better handle on the valence quarks ($\Delta V = \sum \Delta q_i - \Delta \bar{q}_i$) in the data region, with bigger uncertainties in the extrapolation regions.
- Smaller uncertainties in the gluon PDF thanks to the improved treatment of the jet data.

Comparison to BDSSV24, at NNLO

- Remarkable agreement on the central gluon despite of the difference on treatment of the NNLO corrections for jet data. Big differences in terms of uncertainties however.

	NNPDF 1.1	BDSSV24
DIS	✓	✓
SIDIS	✗	✓
Proton-Proton	✓	✓
Perturbative Order	NLO	NLO and NNLO
Statistical Treatments	Monte Carlo	Monte Carlo
Parametrisation	Neural Networks	Fixed functional form

The EIC in the context of NNPDFpol 2.0

Projections from:

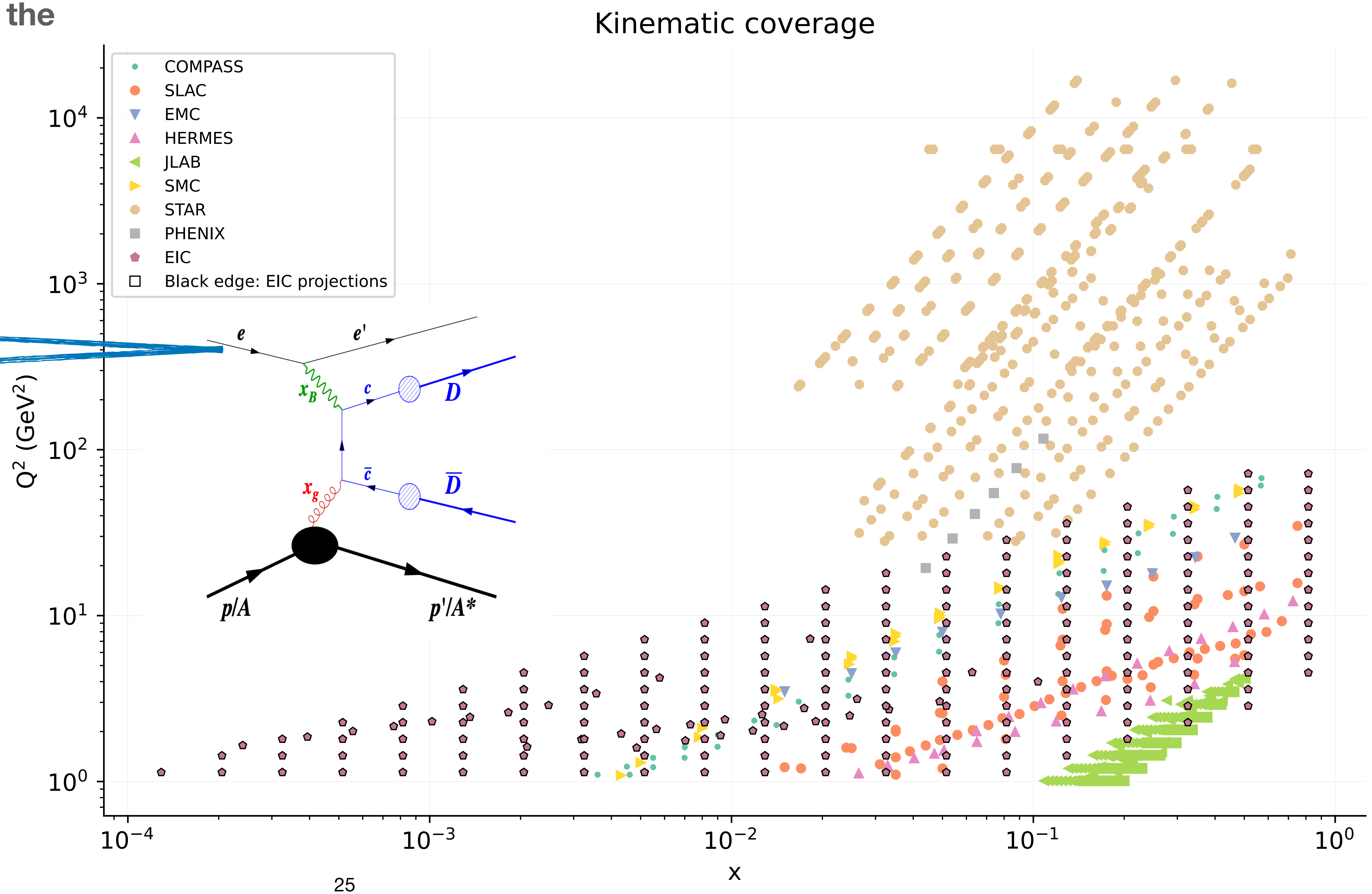
Probing gluon helicity with heavy flavor at the Electron-Ion Collider 2110.04489 [hep-ex]

ATHENA Detector Proposal 2210.09048 [physics.ins-det]

Studying the polarization content of the proton in a much greater kinematic range!

And the kinematic coverage is only half of the story!
Tagging charm hadrons means that we have *VIP* access to Δg, a very important piece of the puzzle!

Greater kinematic coverage, high precision at high-x: great opportunities for a more precise determination of the (spin) content of the proton!



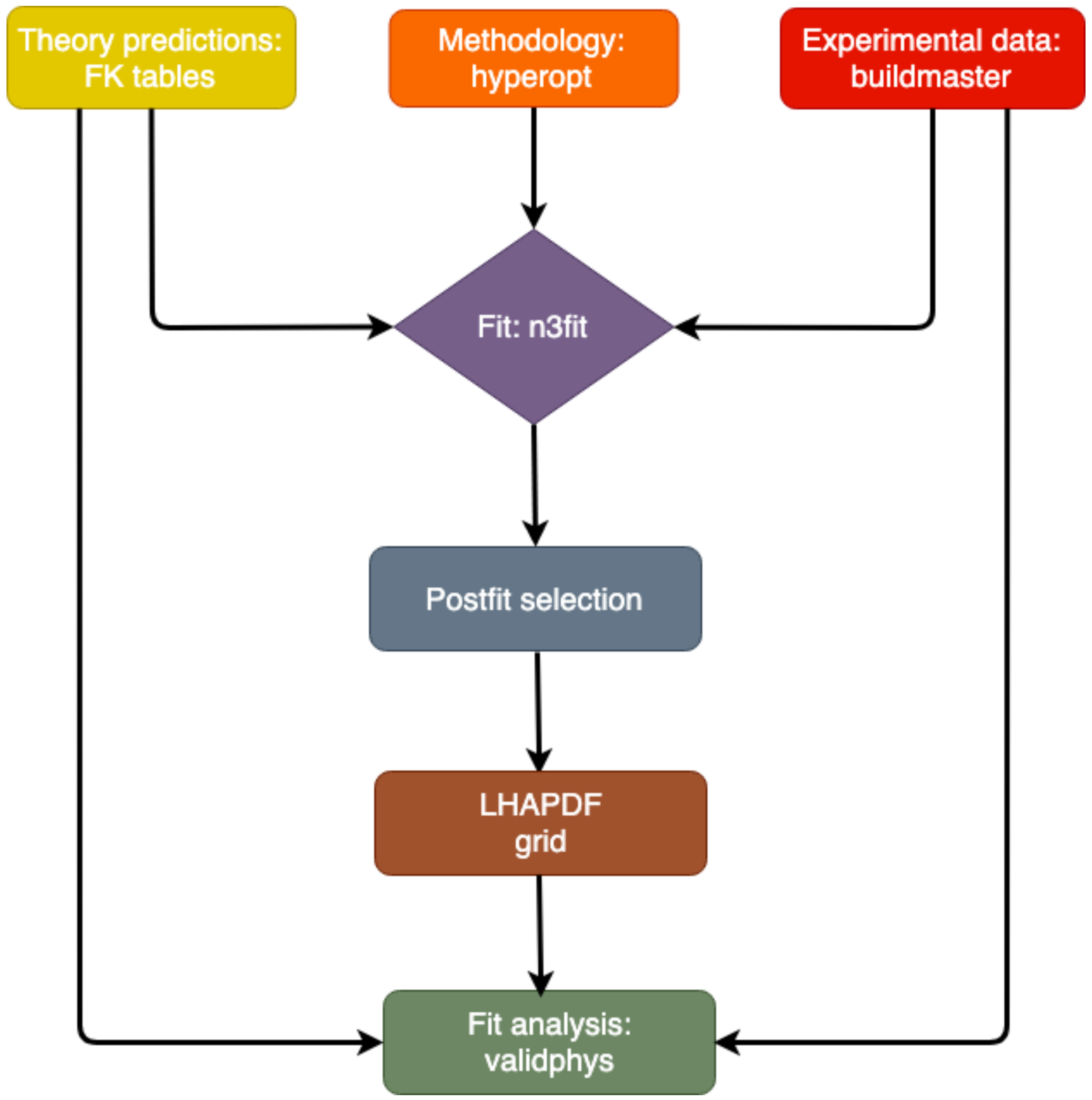
Conclusions

- ☑ PDF determinations already achieve good precision in a wide kinematic range, and it will improve in the coming years (data from Run II, III, HL-LHC).
- ☑ Room for improvement from EIC data: large-x region or charm content
- ☑ The EIC physics program will benefit considerably from the new NNPDFpol 2.0
- ☑ Polarized PDF determinations will benefit considerably from the new EIC

Thanks!

Backup

NNPDF fitting framework summary



The ingredients necessary to complete a global PDF fits are:

- Experimental data and uncertainties (hepdata)
- Theory predictions in the form of interpolation tables (plougshare, madgraph): Fast Kernel Tables
- Fitting framework (n3fit) -> PDF at scale Q_0
- DGLAP evolution for any value of Q (Apfel, EKO, Apfel++)
- Postfit selection (eliminate outliers, underlearnt or wiggly replicas and double-check physical constraints)
- Final output: LHAPDF grid
- (optional) an analysis framework to facilitate creating nice plots and presentations

An open-source machine learning framework for global analyses of parton distributions
NNPDF collaboration - [hep-ph] 2109.02671

Validation and testing

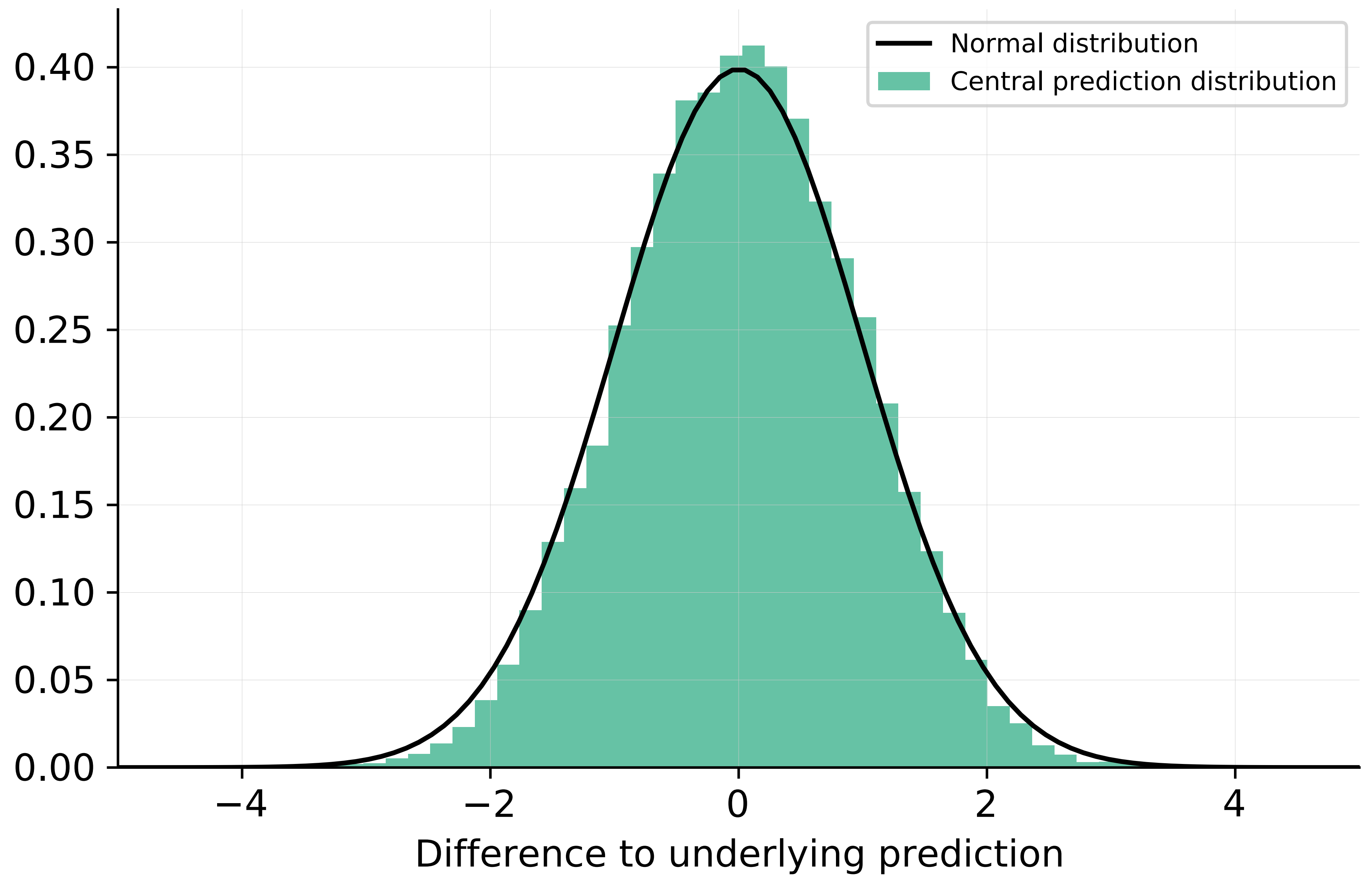
Closure tests

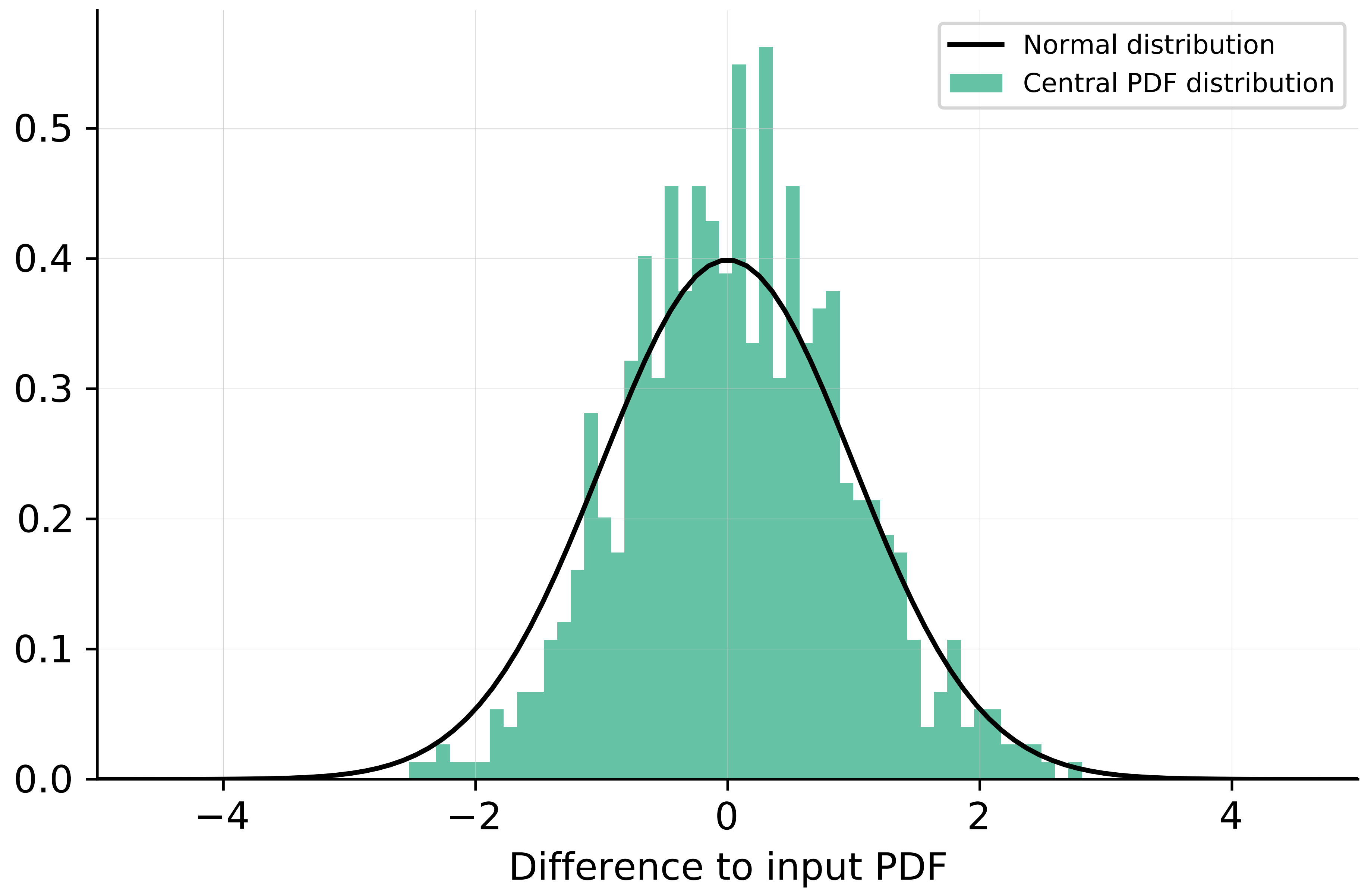
1. Select some other PDF as the truth (an NNPDF replica or a fit from another group)
2. Generate fake data according to the theoretical predictions used in the fit
3. Generate variations of the data using the experimental uncertainties

➡ Check whether the parametrization is flexible enough

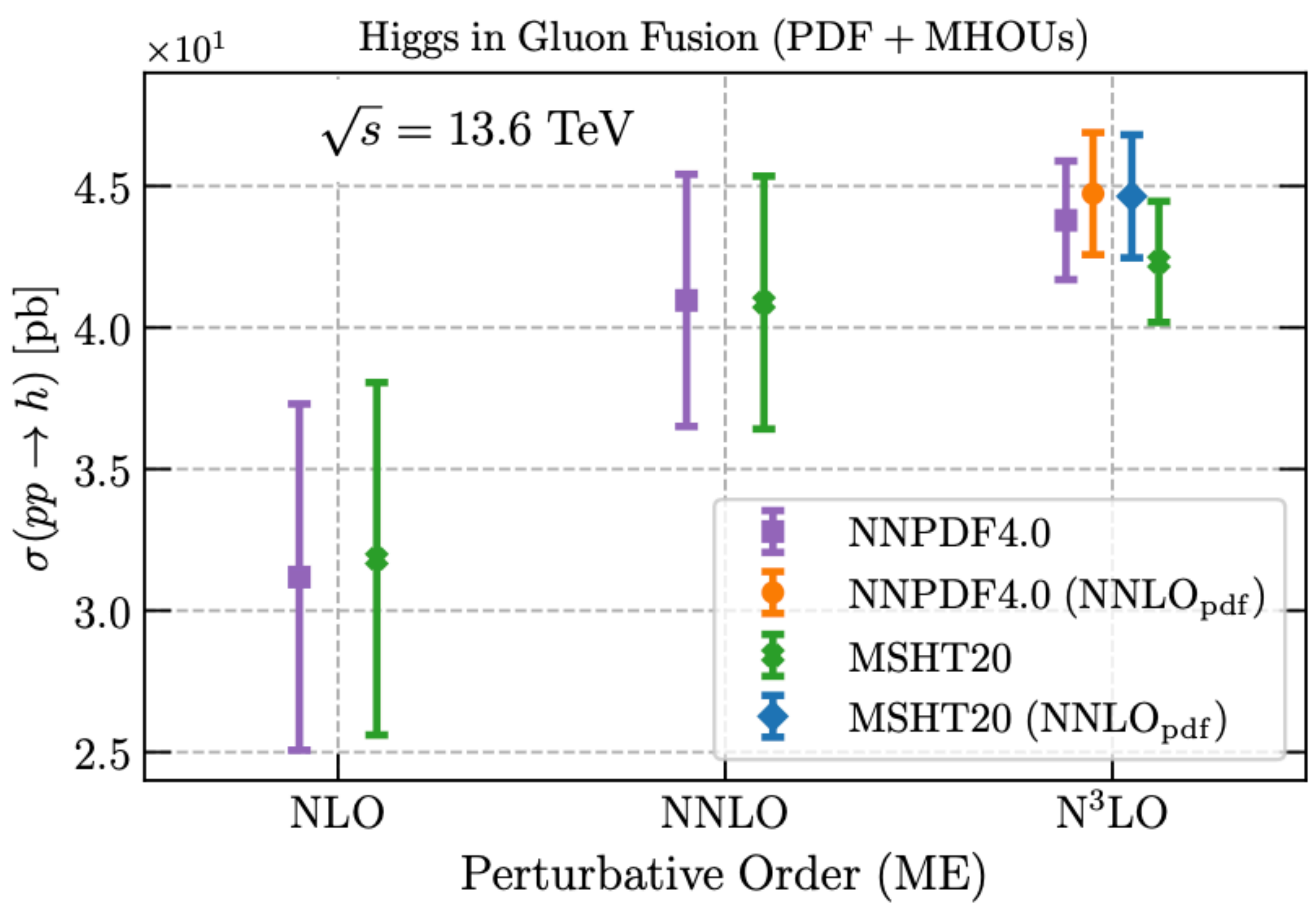
➡ Check whether we can reproduce the “true” PDF if it were known

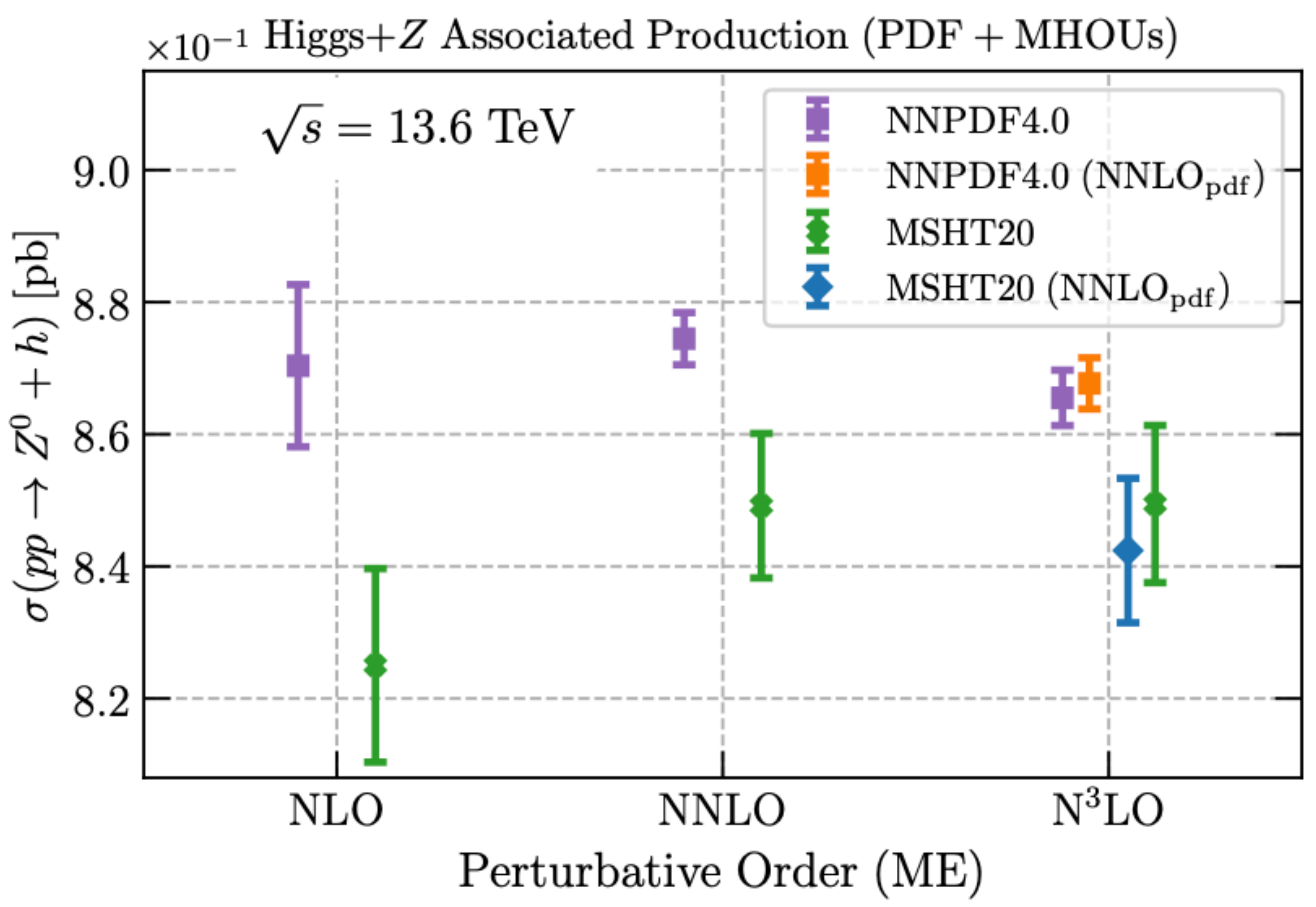
➡ Do all of that in an environment in which everything is consistent and no theoretical knowledge is missing (no MHOU needed)

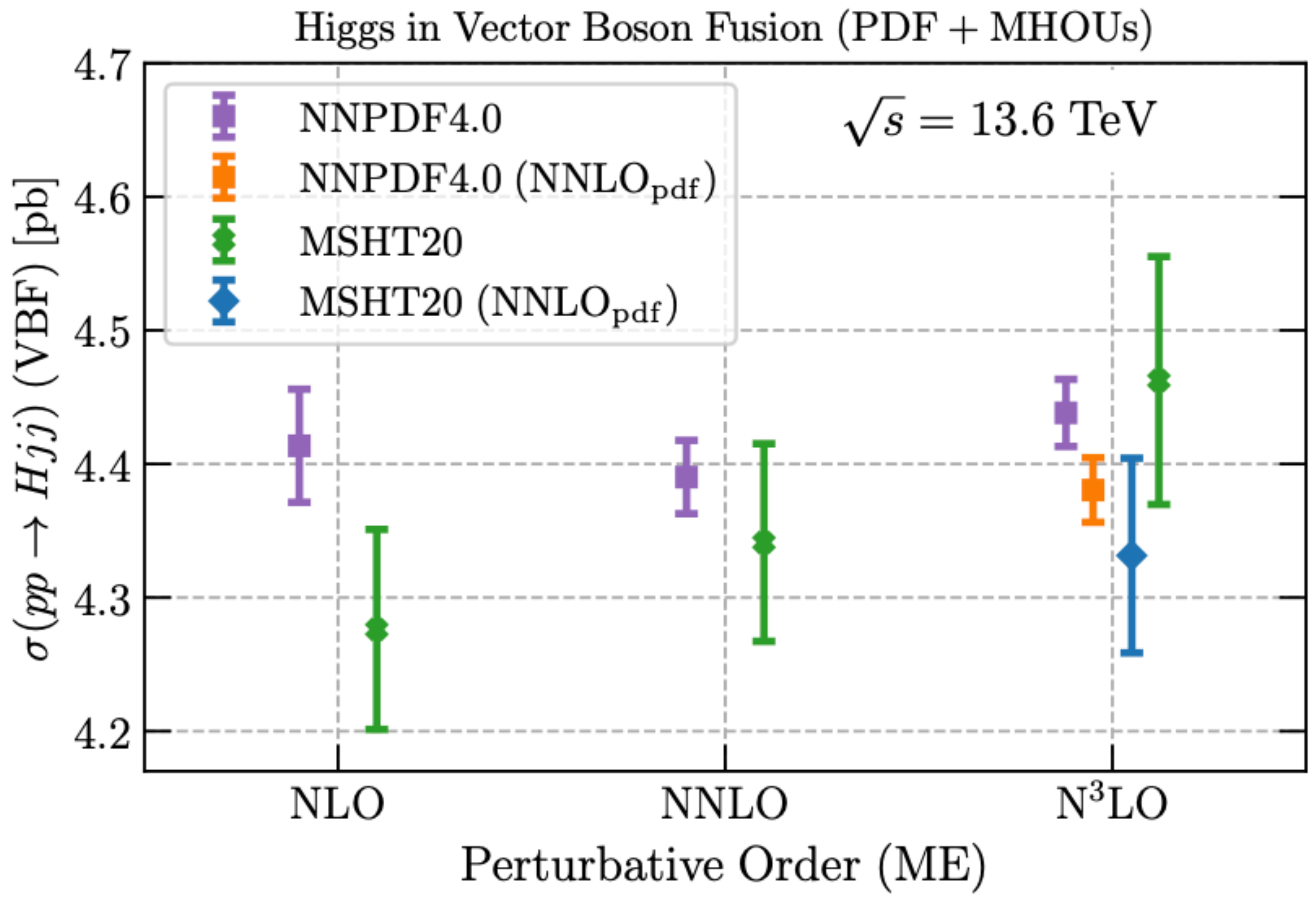




LHC phenomenology: Higgs production







- N^3LO PDF corrections to **Higgs in gluon fusion small,** with a 1.5% suppression wrt NNLO PDFs
- N^3LO corrections improve agreement between NNPDF4.0 and MSHT20 for ***hZ***
- **Higgs VBF** also receives large corrections (in units of the very small N^3LO scale error)

Theoretical Input of NNPDFpol 2.0

Slide by G. Magni

DGLAP evolution

$$\mu^2 \frac{d\Delta f_i}{d\mu^2} = \Delta P_{ij}(x, \alpha_s) \otimes \Delta f_j(x, \mu^2)$$

- RGE for PDFs. It describes how quarks and gluon mixes into each other and fixes the μ^2 dependency.
- Splitting Functions are **analytically known** up to **NNLO** $\mathcal{O}(\alpha_s^3)$ Moch, Vermaseren, Vogt [arxiv:1409.5131] [arxiv:1506.04517], Blümlein, Schneider, Schönwald [arxiv:2111.12401], Gluck, Reya, Stratmann, Vogelsgand [arxiv:9508347]
- **Helicity conservation** implies that the first moment of the gluon-to-quark splitting function vanishes: $\int_1^0 dx\, x\Delta P_{qg}(x, \alpha_s) = 0$

Matching conditions

$$f_i^{(n_f+1)}(x, \mu^2) = A_{ij}(x, \alpha_s) \otimes f_j^{(n_f)}(x, \mu^2)$$

- Describe how massive and massless schemes are matched.
- Matching Condition matrices ΔA_{ij} are **known analytically** up to **NNLO** $\mathcal{O}(\alpha_s^2)$ Bierenbaum, Blümlein, Freitas, Goedicke, Klein, Schönwald [arxiv:2211.15337]

Partonic Matrix elements

$$\sigma(x, Q^2) = \sum_{i,j} f_j\left(\mu^2\right) \otimes f_i\left(\mu^2\right) \otimes \hat{\sigma}_{ij}\left(\alpha_s, Q^2, \mu^2, \right)$$

- **For Drell-Yan and DIS**: all the needed partonic matrix elements $\hat{\sigma}_{ij}(x, Q^2)$ are available at **NNLO**.
- **For Jets**: $\hat{\sigma}_{ij}(x, Q^2)$ are only at NLO. We include **NLO MHOU** computed with *3pt-renormalisation* scale variations **as proxy** of the unknown NNLO contribution.

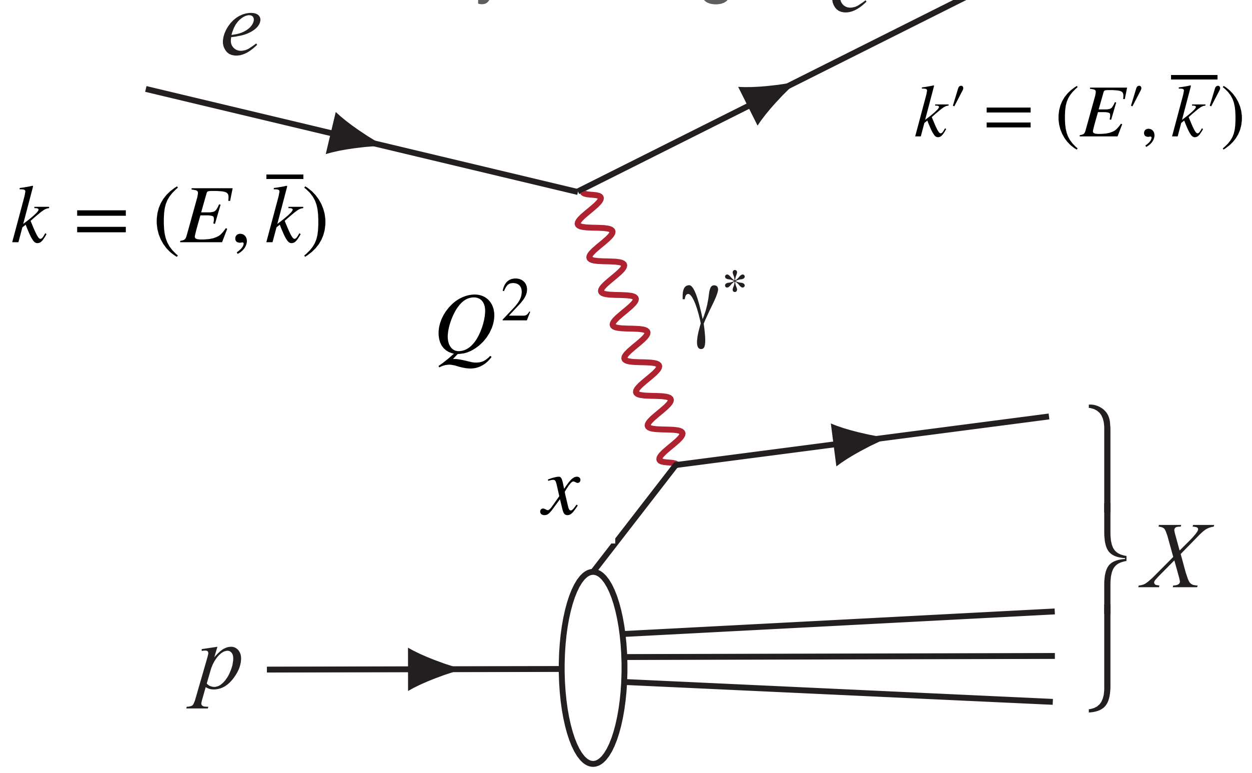
Other optional **theory uncertainties** are computed through *7pt-factorisation and renormalisation* scale variations and propagated to the fit with a **covariance matrix**:

$$Cov_{tot} = Cov_{exp} + Cov_{Jets,MHOU} \quad (+Cov_{MHOU})$$

Deep Inelastic Scattering

Slide by G. Magni

$$Q^2 = -(k'-k)^2$$
$$y = 1 - \frac{E'}{E}$$
$$x = \frac{Q^2}{2M(E-E')}$$



$$\frac{d\sigma^{\uparrow\downarrow}}{dxdy} + \frac{d\sigma^{\uparrow\uparrow}}{dxdy} = \frac{8\pi ME}{Q^4}\left[\left(2y - y^2 - \frac{Mxy}{E}\right) 2xF_1(x,Q^2) - \frac{4M}{E}x^2y\, F_2(x,Q^2)\right]$$

$$\frac{d\sigma^{\uparrow\downarrow}}{dxdy} - \frac{d\sigma^{\uparrow\uparrow}}{dxdy} = \frac{8\pi ME}{Q^4}\left[\left(2y - y^2 - \frac{Mxy}{E}\right) 2xg_1(x,Q^2) - \frac{4M}{E}x^2y\, g_2(x,Q^2)\right]$$

The cross section $d\sigma^{\uparrow\uparrow}$ $(d\sigma^{\uparrow\downarrow})$ corresponds to the incoming lepton being longitudinally polarised to parallel (anti-parallel) to the nucleon. Define the **asymmetry**:

$$A = \frac{d\sigma^{\uparrow\downarrow} - d\sigma^{\uparrow\uparrow}}{d\sigma^{\uparrow\downarrow} + d\sigma^{\uparrow\uparrow}} \xrightarrow{E^2 \gg M^2} A_1 \approx \frac{g_1(x,Q^2)}{F_1(x,Q^2)}$$

For photon exchange the structure function can be decomposed as:

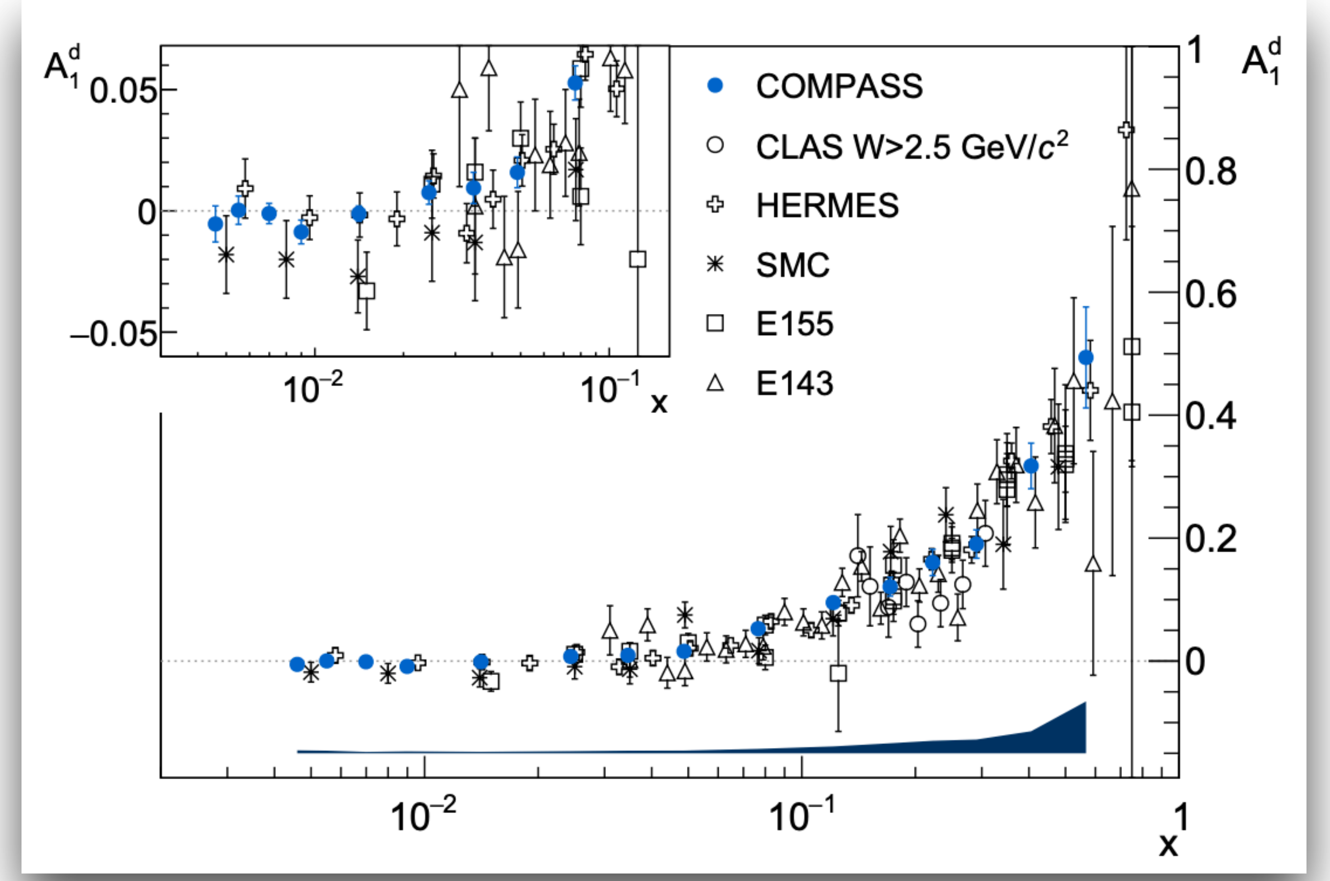
$$g_1(x,Q^2) = \sum_q e_q^2 \Delta C_q \otimes (\Delta q + \Delta \bar{q}) + \langle e^2 \rangle \Delta C_g \otimes \Delta g$$

$$\Delta\Sigma + \Delta T_8,\ \Delta T_3\ \left[\mathcal{O}(a_s^0)\right]$$

$$\Delta g\ \left[\mathcal{O}(a_s)\right]$$

$$\Delta\Sigma = \Delta u^+ + \Delta d^+ + \Delta s^+$$
$$\Delta T_3 = \Delta u^+ - \Delta d^+$$
$$\Delta T_8 = \Delta u^+ + \Delta d^+ - 2\Delta s^+$$

- Measurements available for **proton and deuterium targets** from: COMPASS, SLAC, HERMES, CLAS, and SMC.



- Theory predictions available up to **QCD NNLO**. Ziljstra, Van Neerven [iNSPIRE:353973]
- Heavy quark $\frac{m_h^2}{Q^2}$ effects known. Hekhorn, Stratmann [arxiv:1805.09026], Blümlein et al. [arxiv:1504.08217][arxiv:1912.02536] [arxiv:2101.05733]
- Full fledge **FONLL variable flavour number scheme**. [arxiv:2401.10127]

Longitudinal asymmetries in pp collisions

Slide by G. Magni

Charged boson production

$p\ p \to W^{\pm} \to l^{\pm}\nu$ with one longitudinal polarised beam:

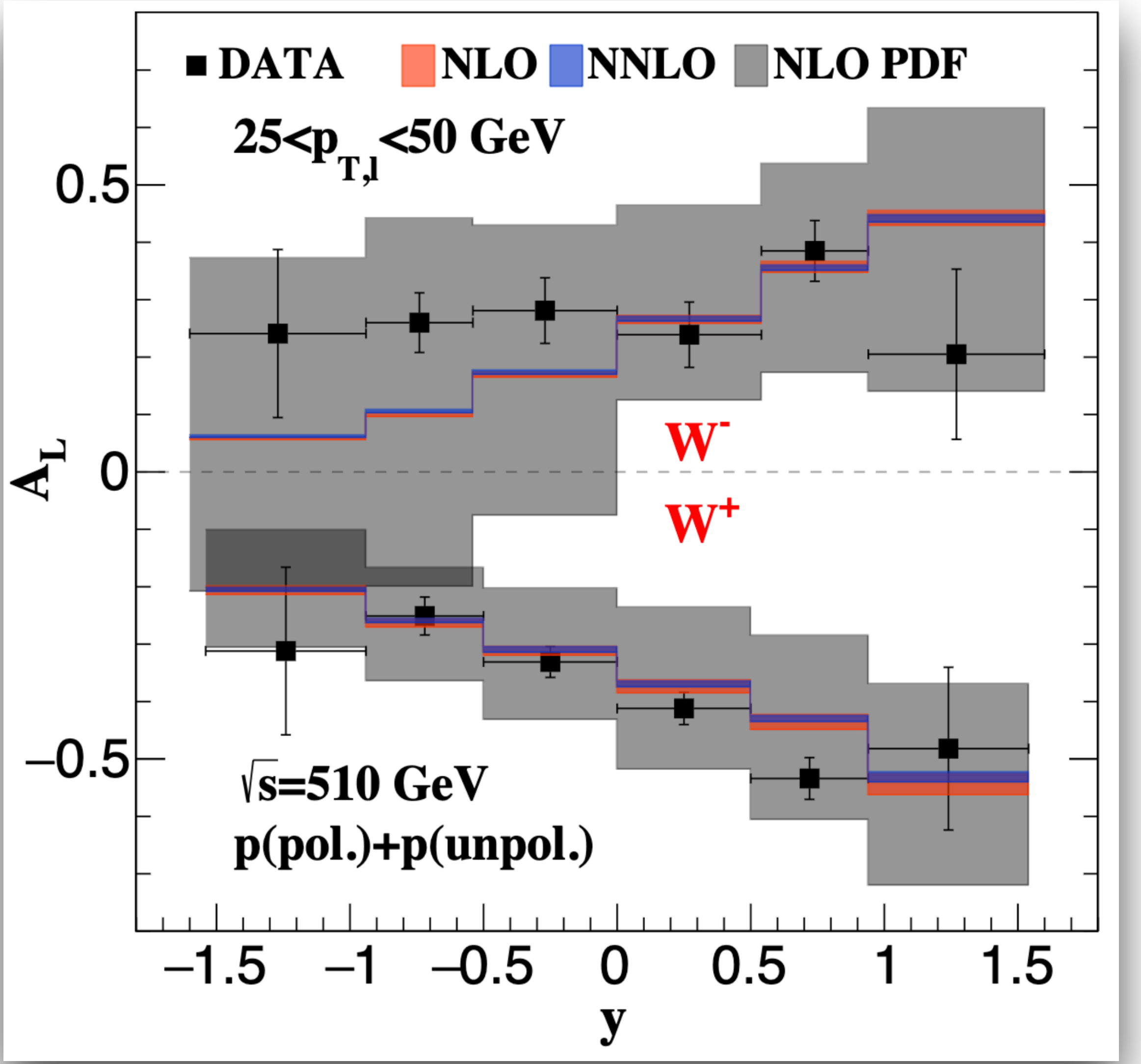
$$A_L = \frac{d\sigma^{\uparrow} - d\sigma^{\downarrow}}{d\sigma^{\uparrow} + d\sigma^{\downarrow}}$$

$$A_L^{W^+} \approx \frac{\Delta\bar{d}(x_1)u(x_2) - \Delta u(x_1)\bar{d}(x_2)}{\bar{d}(x_1)u(x_2) - u(x_1\bar{d}(x_2)} \qquad A_L^{W^-} \approx \frac{\Delta\bar{u}(x_1)d(x_2) - \Delta d(x_1)\bar{u}(x_2)}{\bar{u}(x_1)d(x_2) - d(x_1)\bar{u}(x_2)}$$

$$\downarrow$$

$$\Delta u^-, \Delta d^-$$

- Measurements from STAR as function of boson rapidity.
- Dominated by **statistical** uncertainties.



Included in the fit at NNLO through a grid interpolation implementation of a modified MCFM 8 implementation from [arxiv:2101.02214]

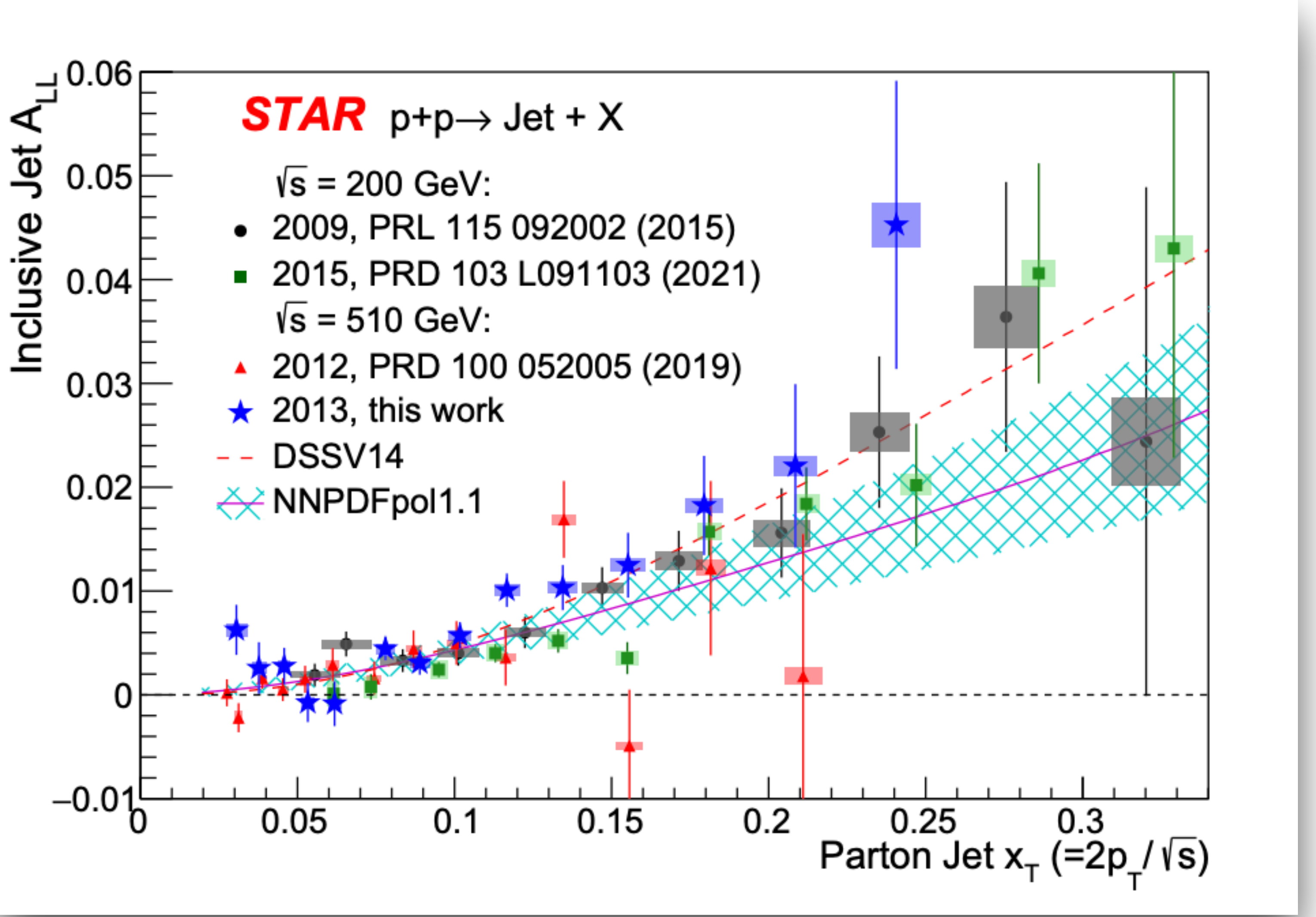
Inclusive (Di)Jets

$p\ p \to j\ (j) + X$ with two longitudinal polarised beam

$$A_{LL} = \frac{d\sigma^{\uparrow\uparrow} - d\sigma^{\uparrow\downarrow}}{d\sigma^{\uparrow\uparrow} + d\sigma^{\uparrow\downarrow}} \approx \frac{\Delta g(x_1)\Delta g(x_2)}{g(x_1)g(x_2)} \longrightarrow \Delta g$$

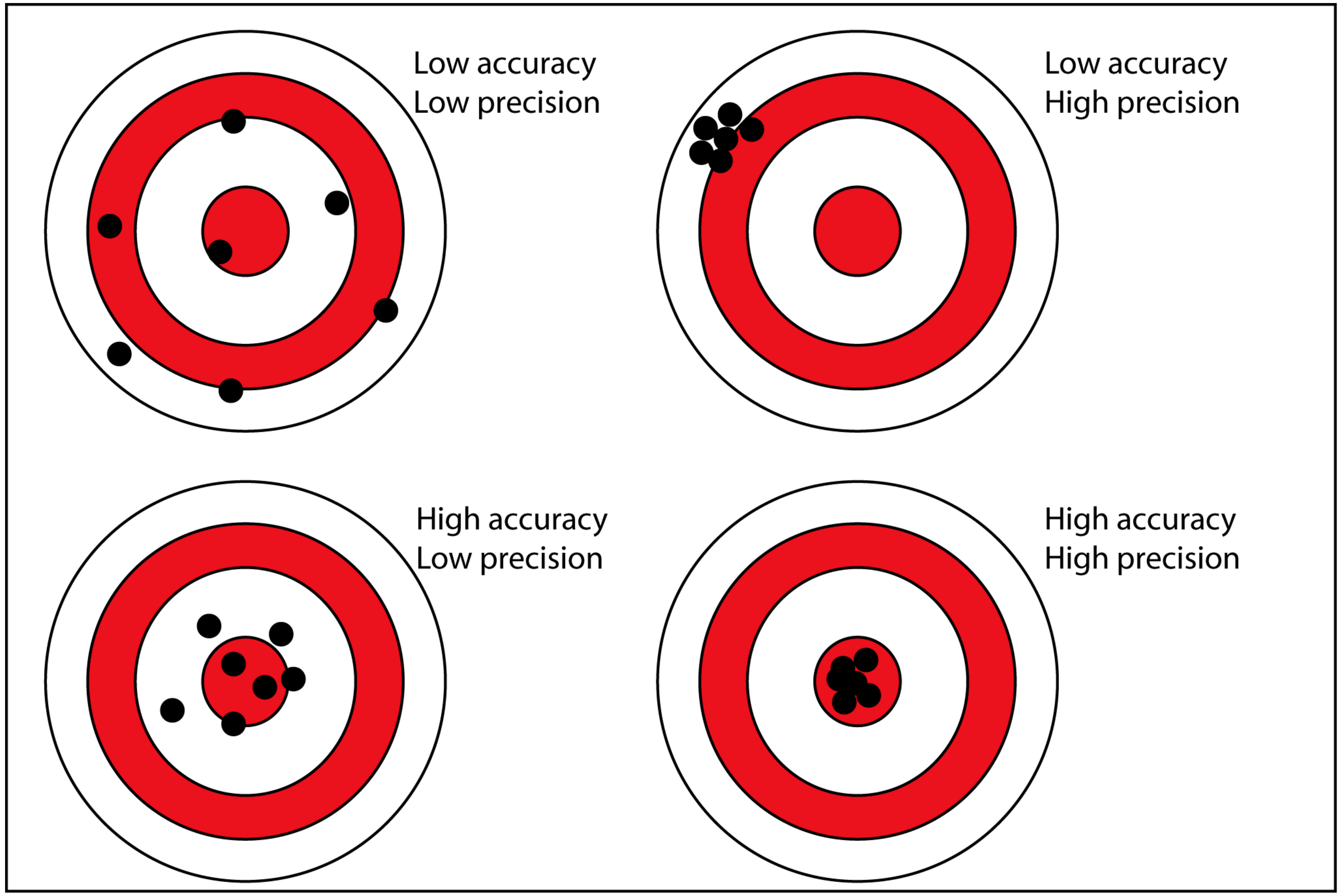
- Measurements from STAR, PHENIX as function of jet transverse momentum and dijets invariant mass.
- Dominated by **systematic (correlated)** uncertainties.

Included in the fit at NLO through a grid interpolation implementation of the codes from [arxiv: 0980.8262] and [arxiv: 0404.057]



Accuracy and precision

The PDF is precise because its input is precise. But what about the *missing* contributions: photon PDF, scale variations, missing higher order contributions, exact NNLO/N3LO predictions…



$$\sigma_{NNLO} = \sigma_0 + \alpha_s\sigma_1 + \alpha_s^2\sigma_2 + \mathcal{O}(\alpha_s^3)$$