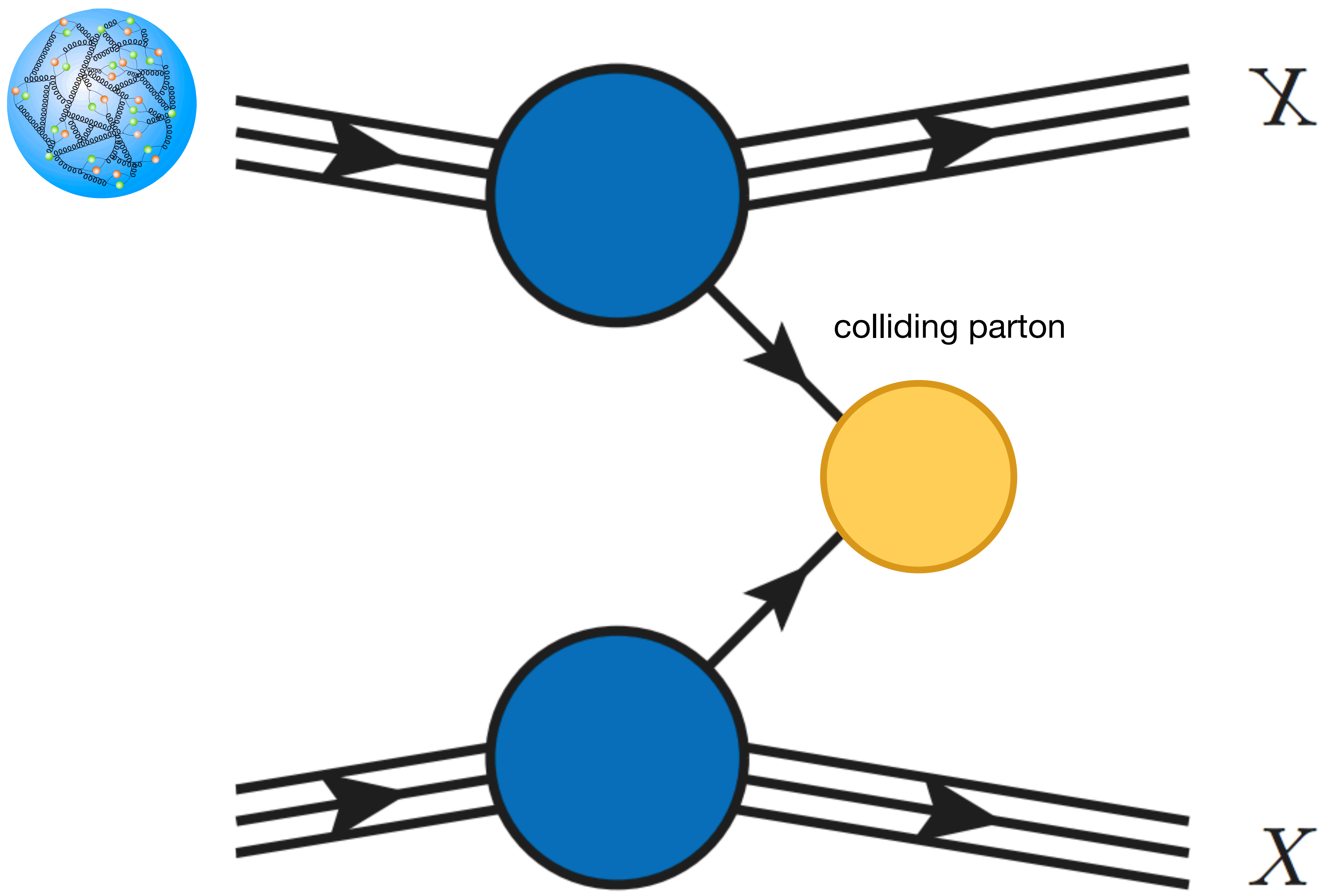
Recent results on PDF determinations

Juan M. Cruz-Martinez

LHCP 11 - Belgrade, May 2023

Collinear unpolarized PDFs

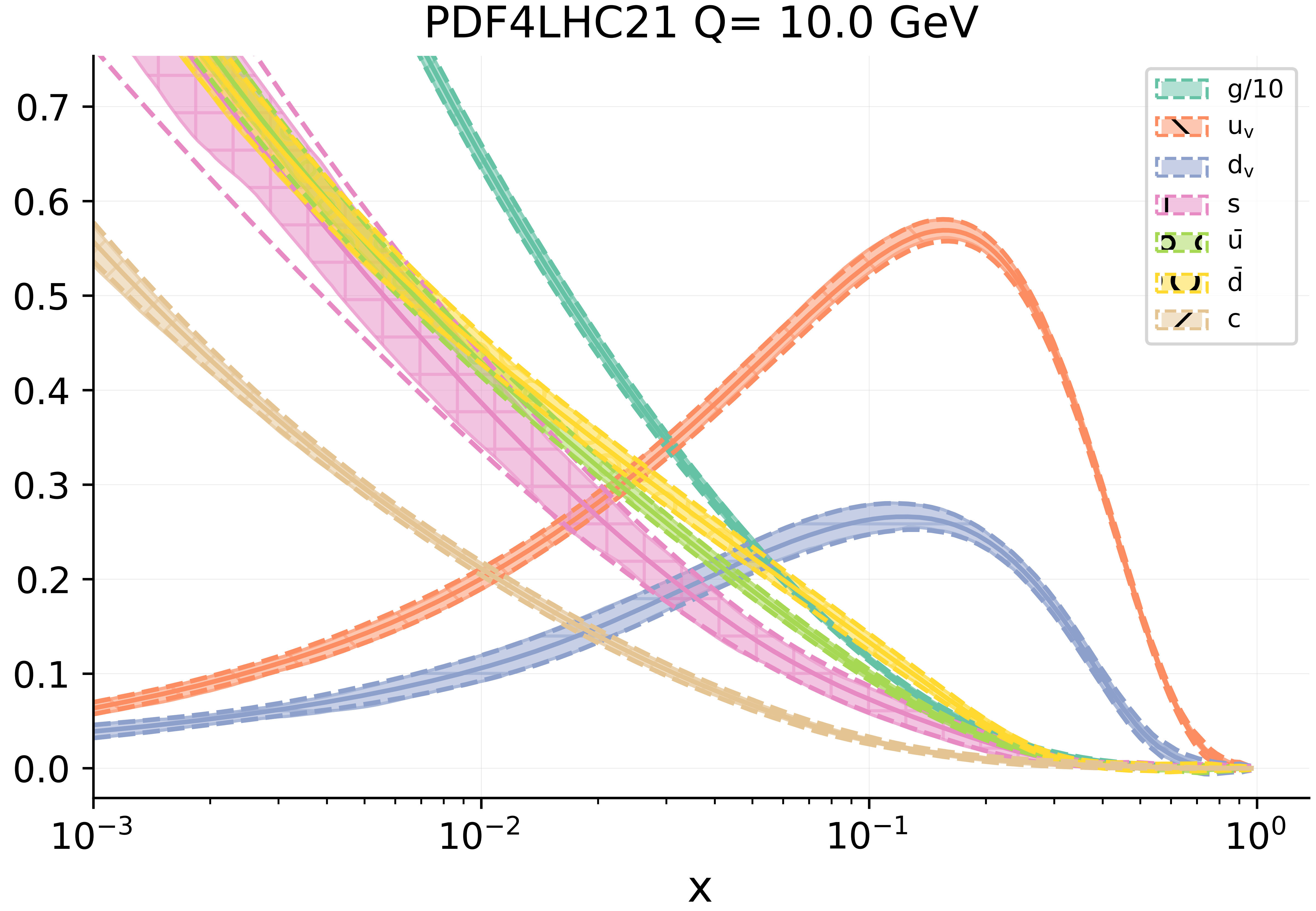


- Crucially important to compare Standard Model predictions to data and for BSM searches.
- Necessary for the extraction of physical parameters such as α_s or the mass of the W.
- A dominant source of uncertainty in precision physics.
- Cannot be computed exactly -> determination limited to fits of well known experimental data.

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

PDF4LHC21 combination

- PDFs determined from $\mathcal{O}(3500)$ datapoints coming from many different type of processes and kinematics. DIS, Fixed-Target and hadronic collider data.
- Latest community combination and benchmark is **PDF4LHC21** hep-ph/2203.05506 and includes:
  - NNPDF3.1 (1706.00428)
  - CT18 (1912.10053)
  - MSHT20 (2012.0468)
- NNLO corrections when available
- NNLO DGLAP evolution
- $\alpha_s(m_Z) = 0.118$
- Update with respect to PDF4LHC15



State of the art

NNPDF4.0, new generation PDF from NNPDF

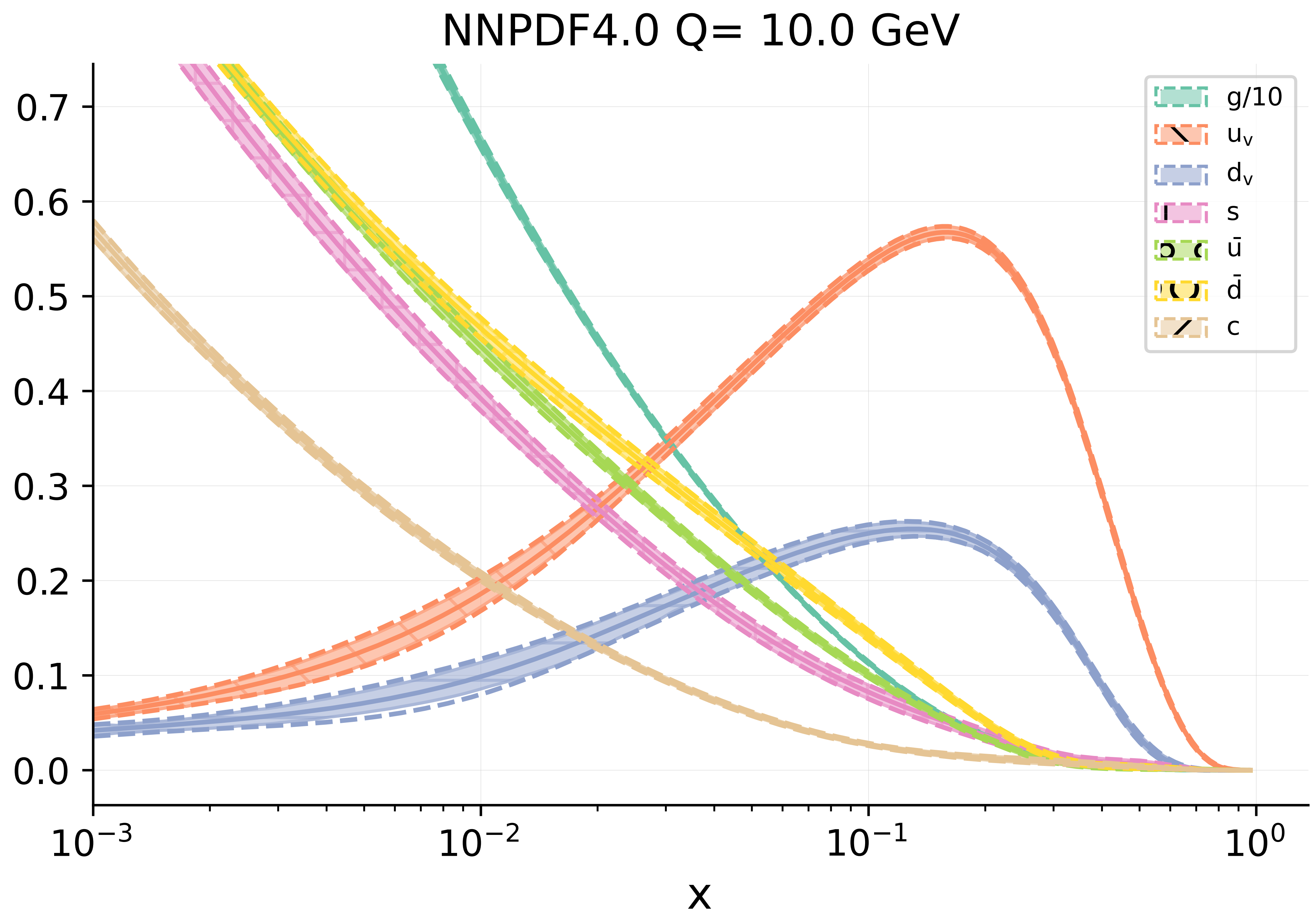
1. Enhanced methodology and tests for reliability of results.
2. Fit from $\mathcal{O}(4000)$ datapoints (with $\mathcal{O}(40)$ new datasets, mostly from LHC!)
3. Integrability and positivity imposed during the fit

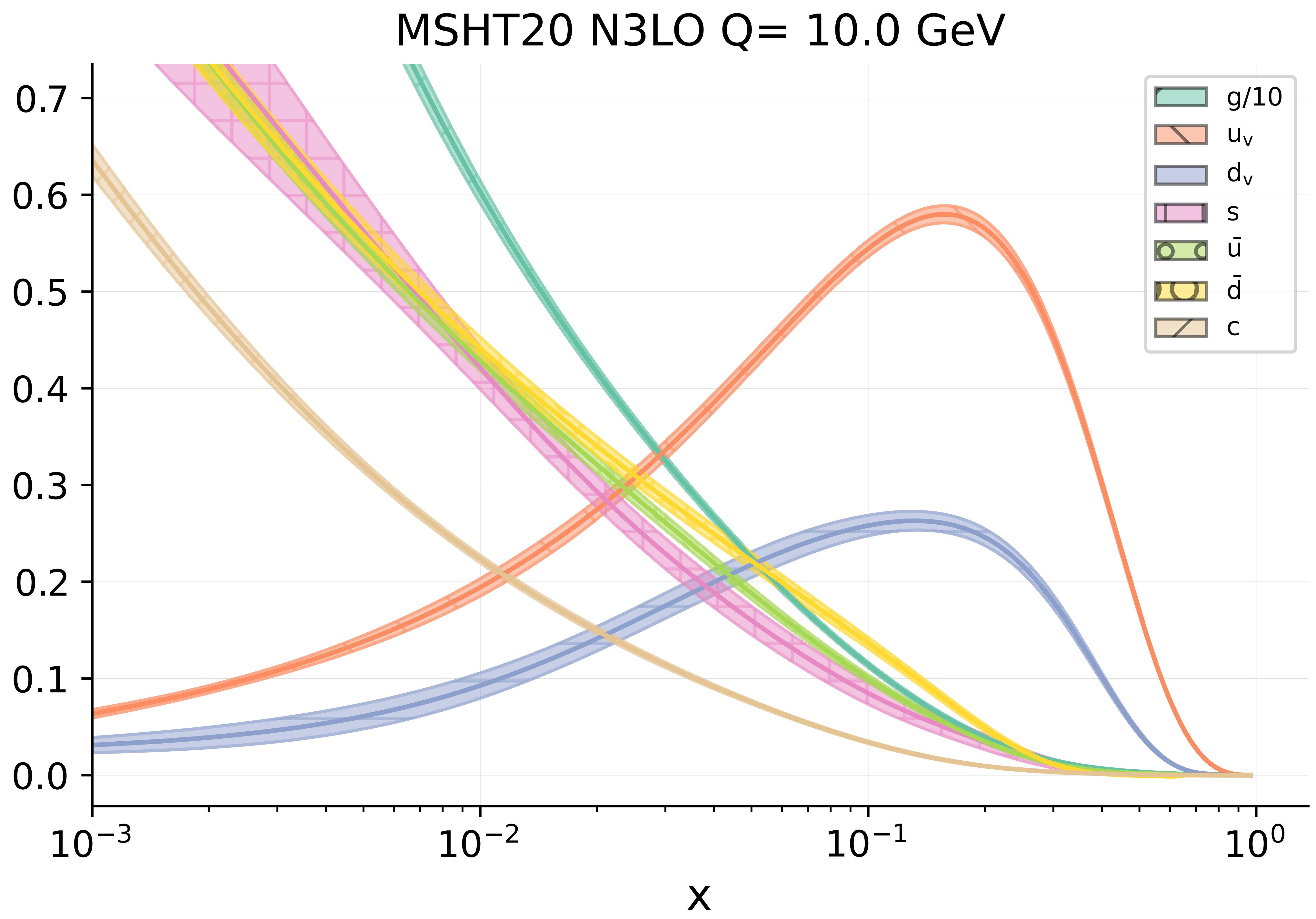
—> hep-ph/2109.02653

Approximated N3LO (in α_s) results by the MSHT group: **MSHT20 aN3LO**

1. Exploit available knowledge of N3LO processes and splitting function
2. Approximate unknown pieces and estimate uncertainty

-> hep-ph/2207.04739



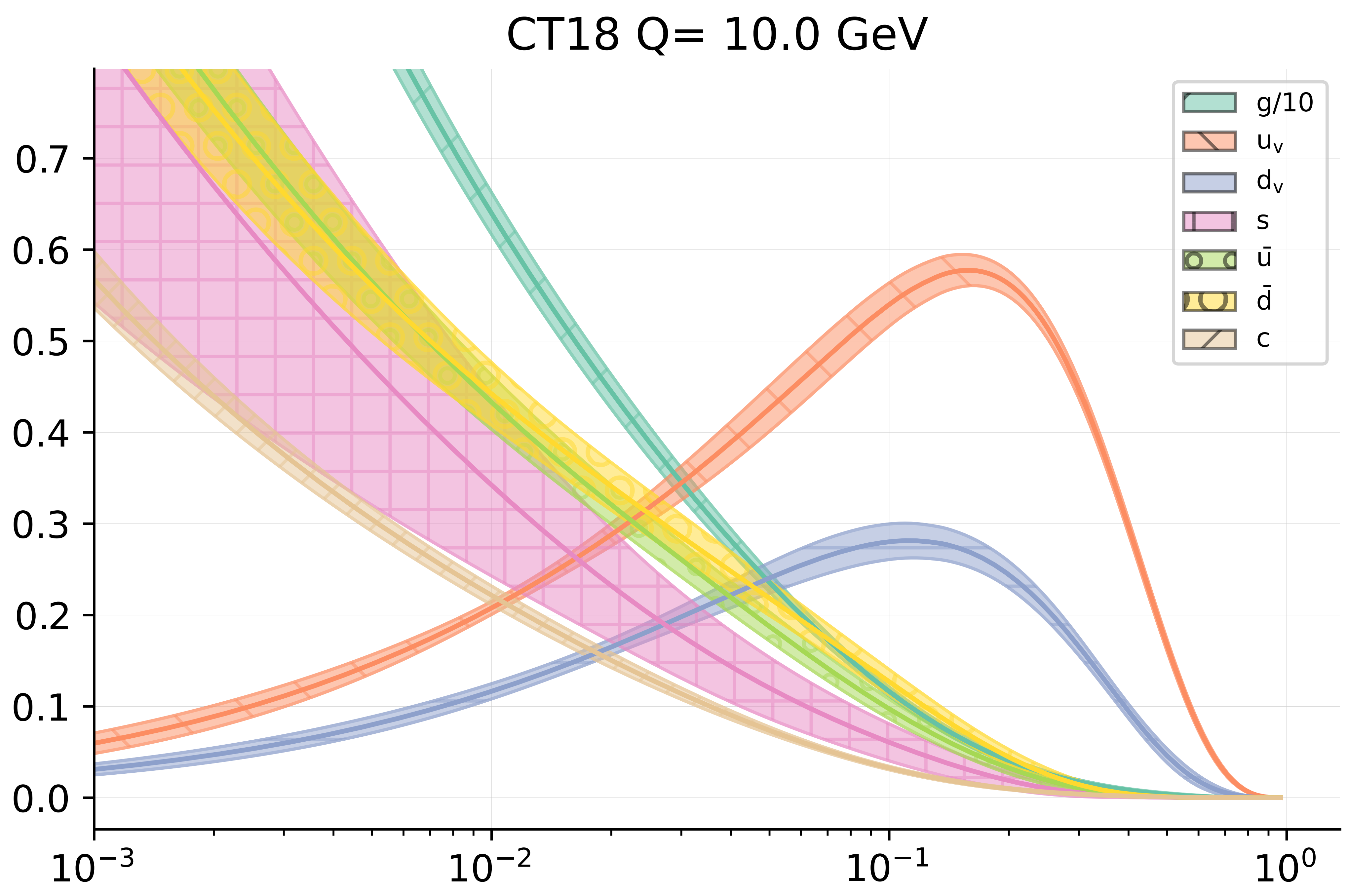


Global NNLO PDFs

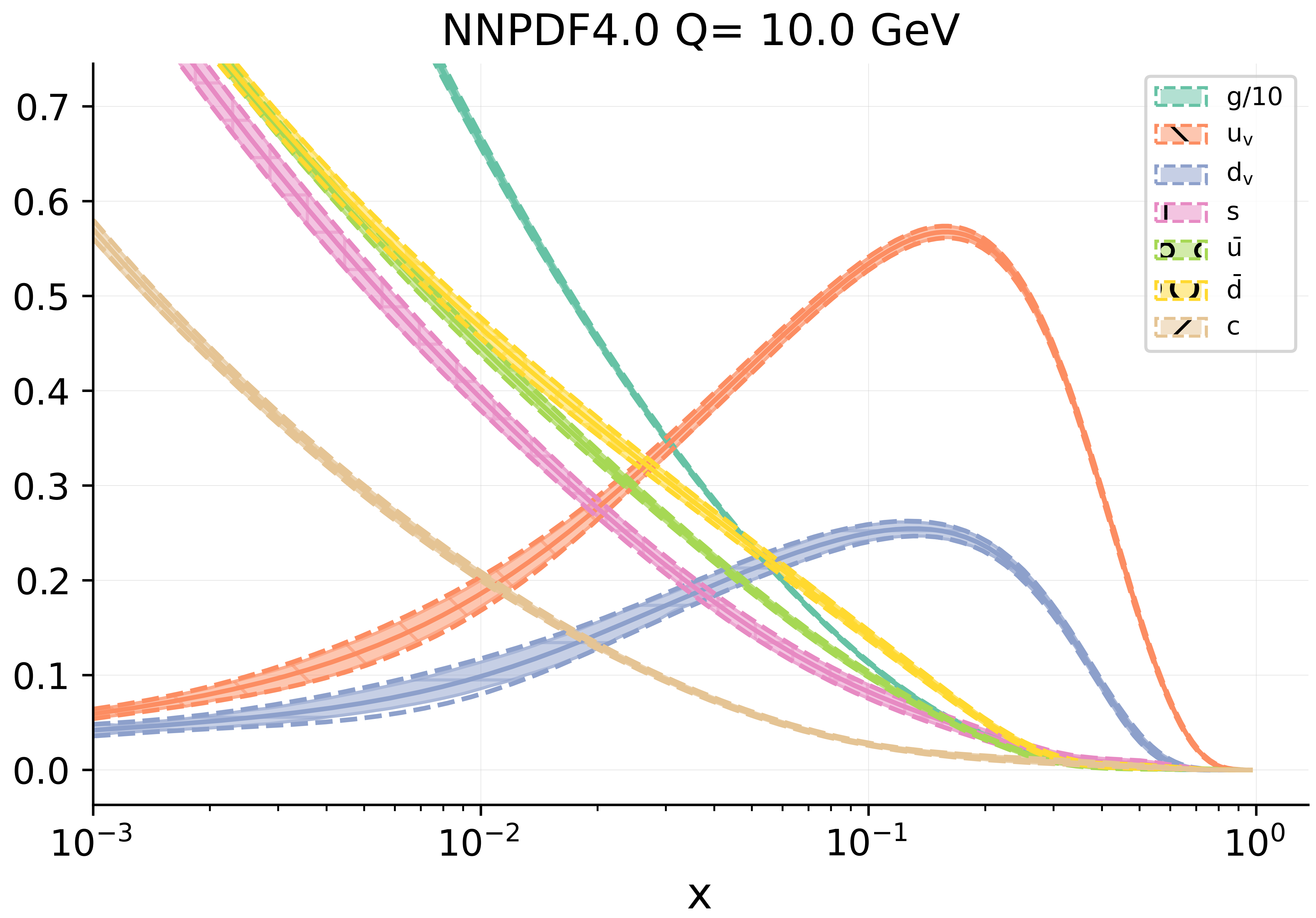
The last releases of the three biggest collaborations:

- **CT18** [hep-ph] 1912.10053
  -> perturbative charm, hessian, tolerance
- **MSHT20** [hep-ph] 2012.04684
  -> perturbative charm, hessian, dynamic tolerance
- **NNPDF4.0** [hep-ph] 2109.02653
  -> fitted (intrinsic) charm, monte carlo

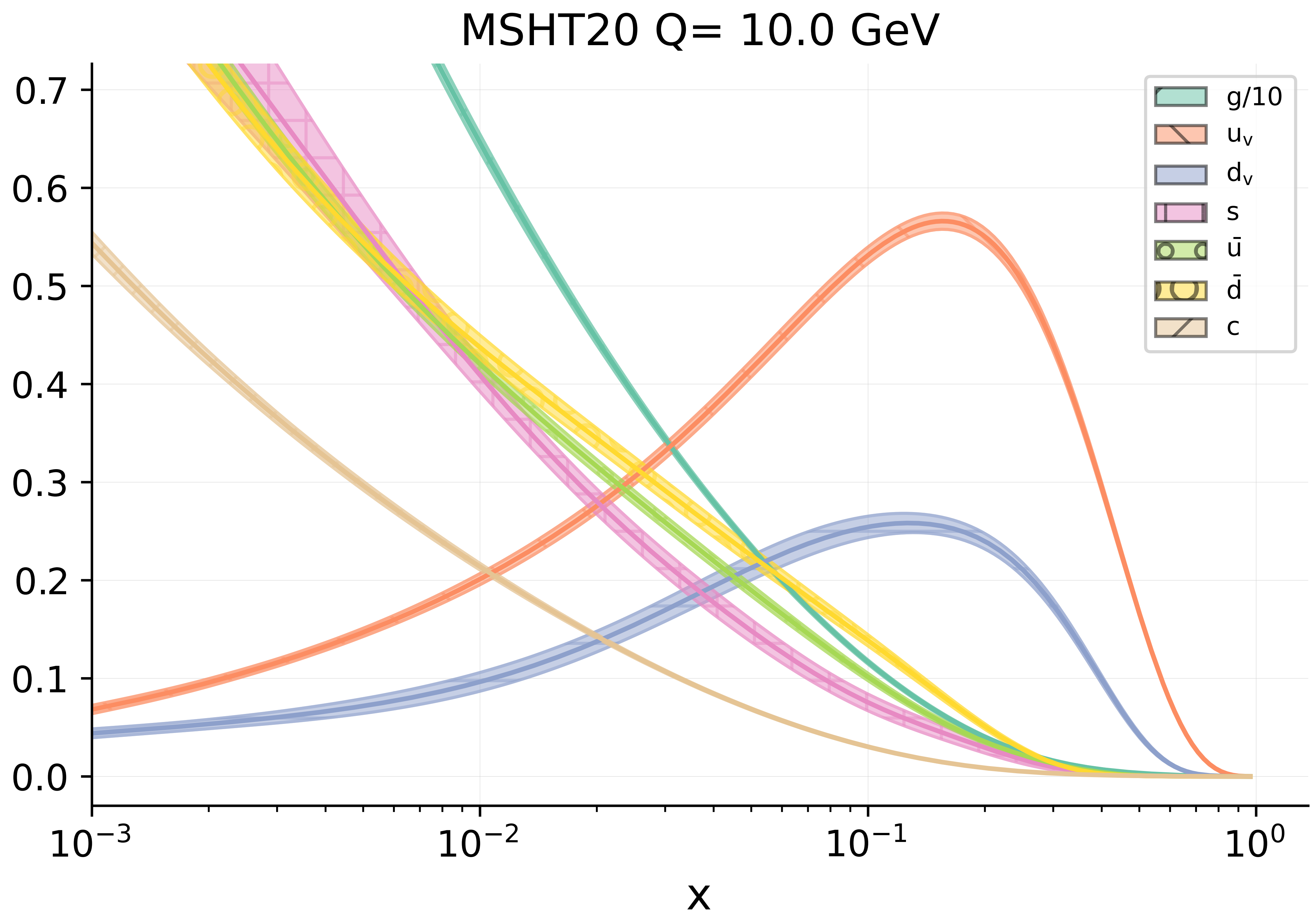
Entering the precision era of Parton Distribution Functions!



CT18 Q= 10.0 GeV



NNPDF4.0 Q= 10.0 GeV

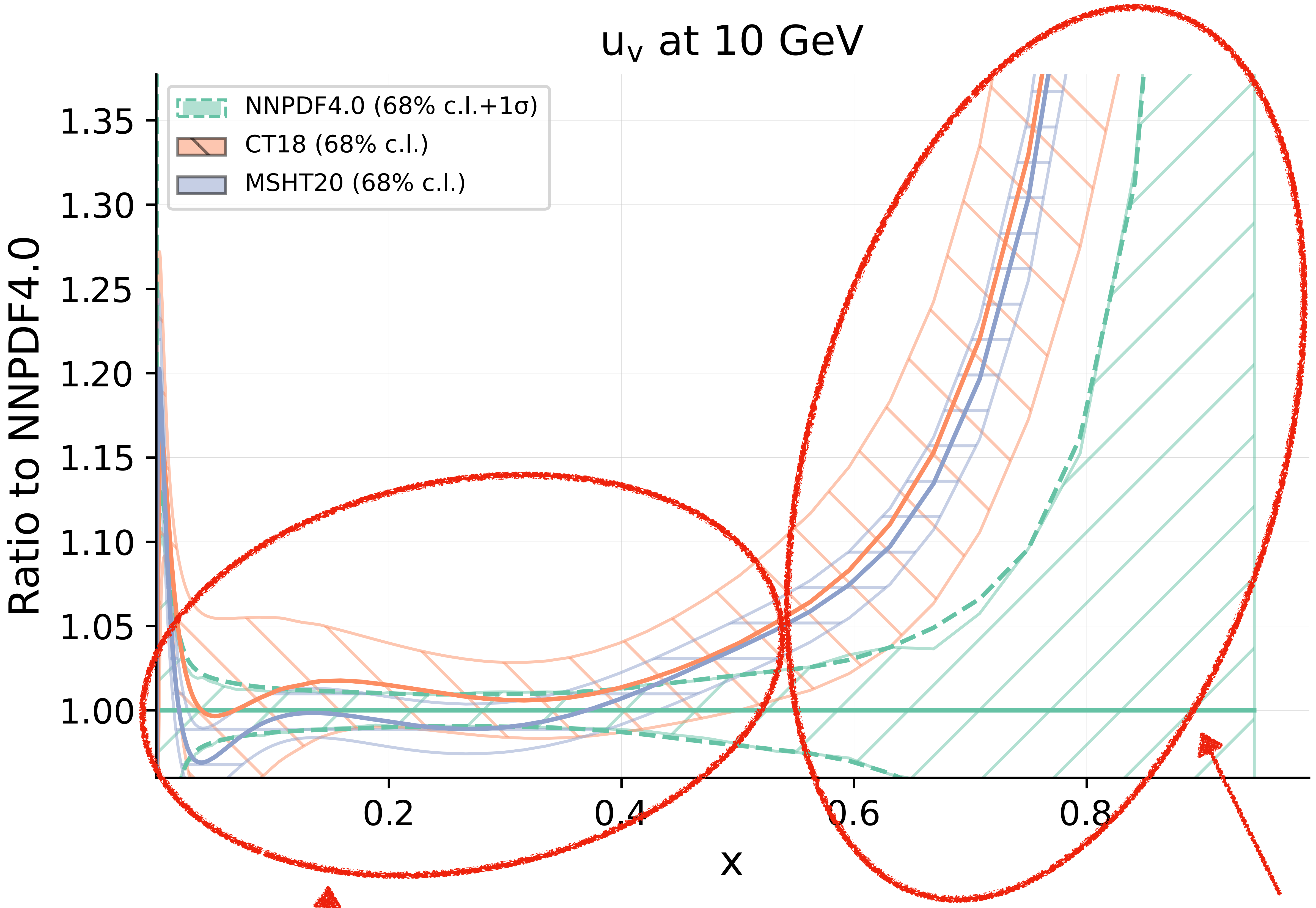


MSHT20 Q= 10.0 GeV

The precision follows the data

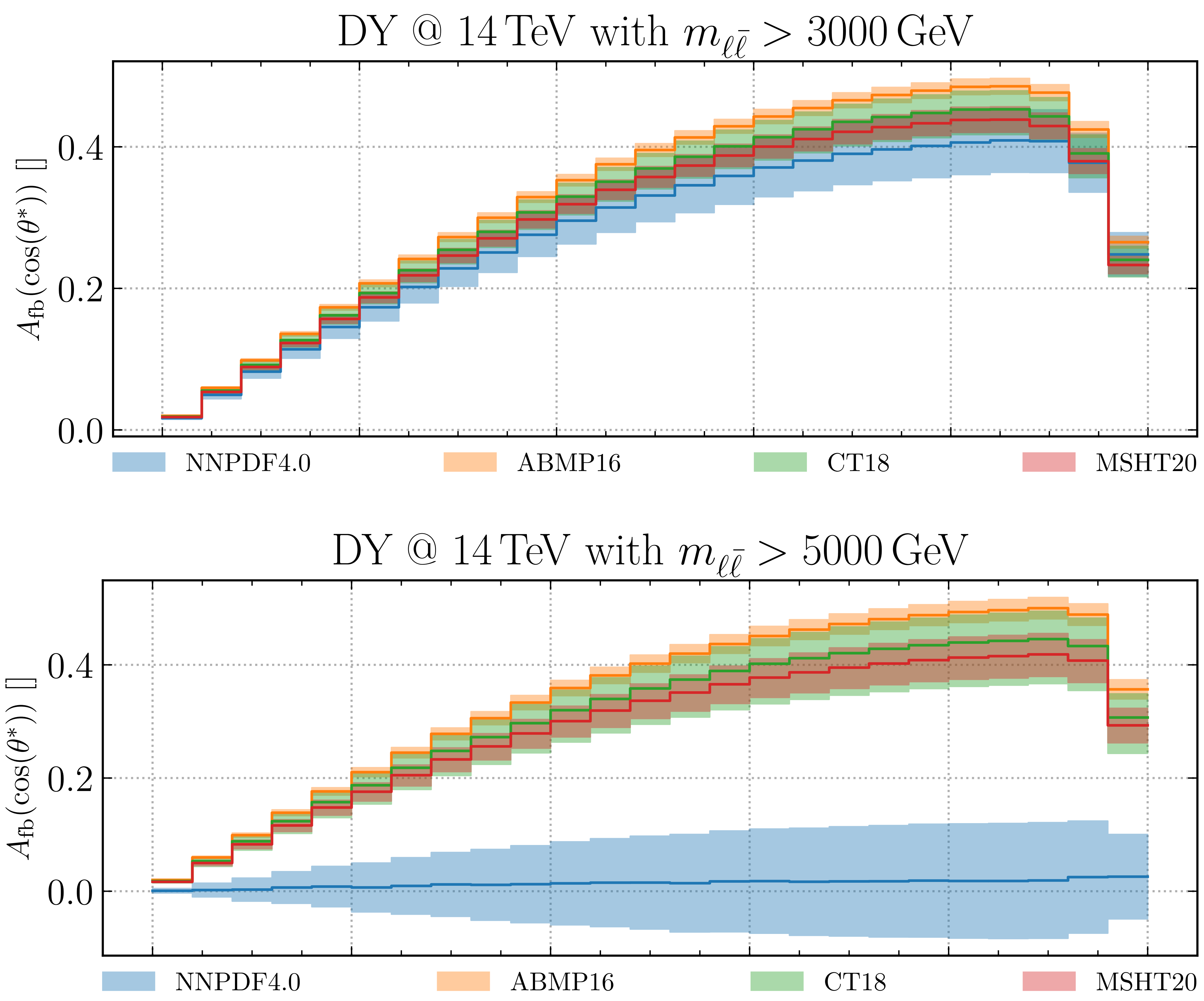
based on

Not all regions are equally well determined, for PDFs the “data region” ends at around x~0.5



Data region: reasonable agreement aiming for both accuracy & precision

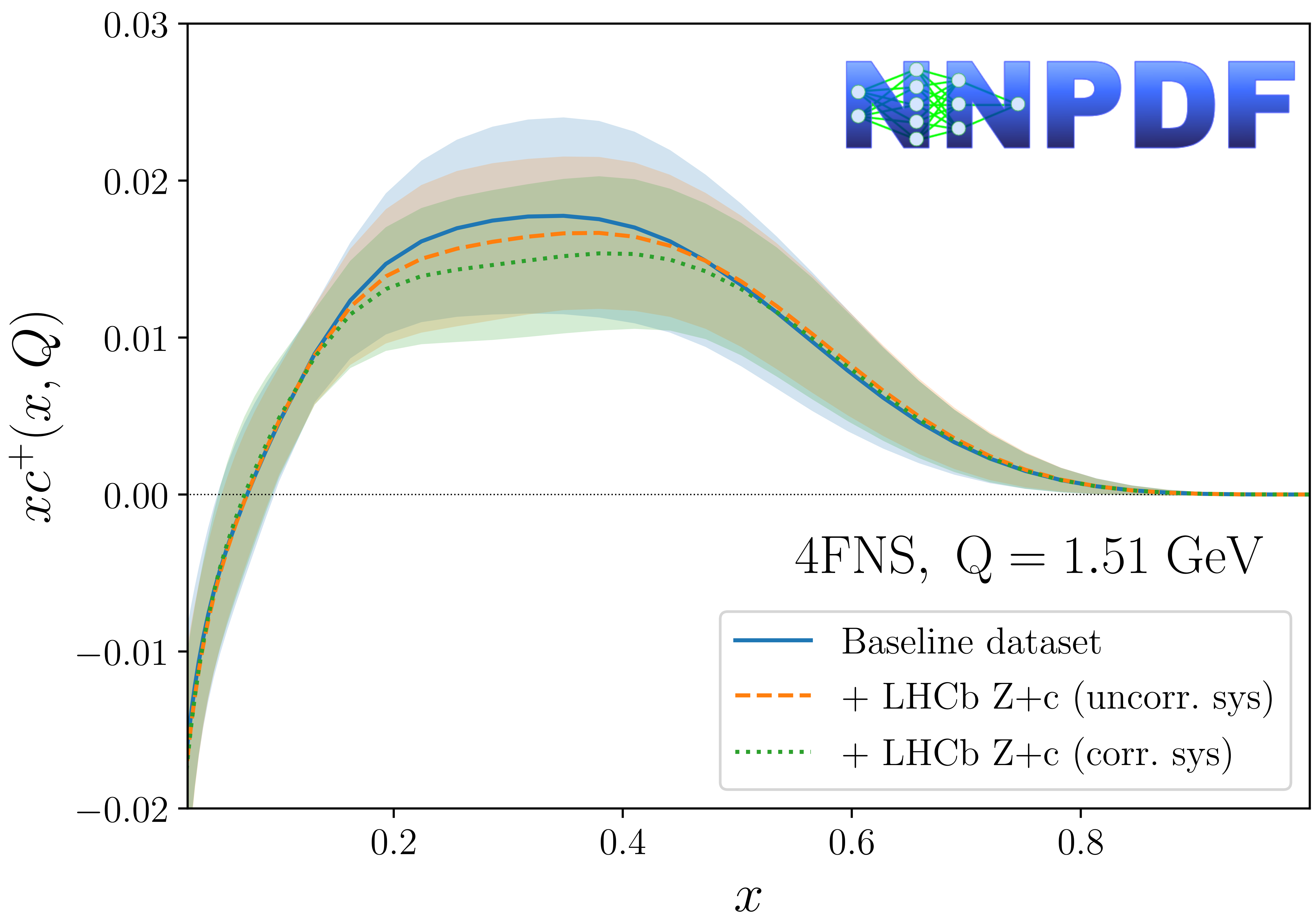
Extrapolation region hic sunt dracones!



In hep-ph/2209.08115 it was demonstrated how a too restrictive parametrization can lead to the extrapolation behaviours not justified by the available data!

Intrinsically charming

Evidence for intrinsic charm quarks in the proton
hep-ph/2208.08372



NNPDF is the only collaboration which fits charm by default, i.e., $c^+ = c + \bar{c} \neq 0$ at the fitting scale which means the contribution is not limited to DGLAP evolution

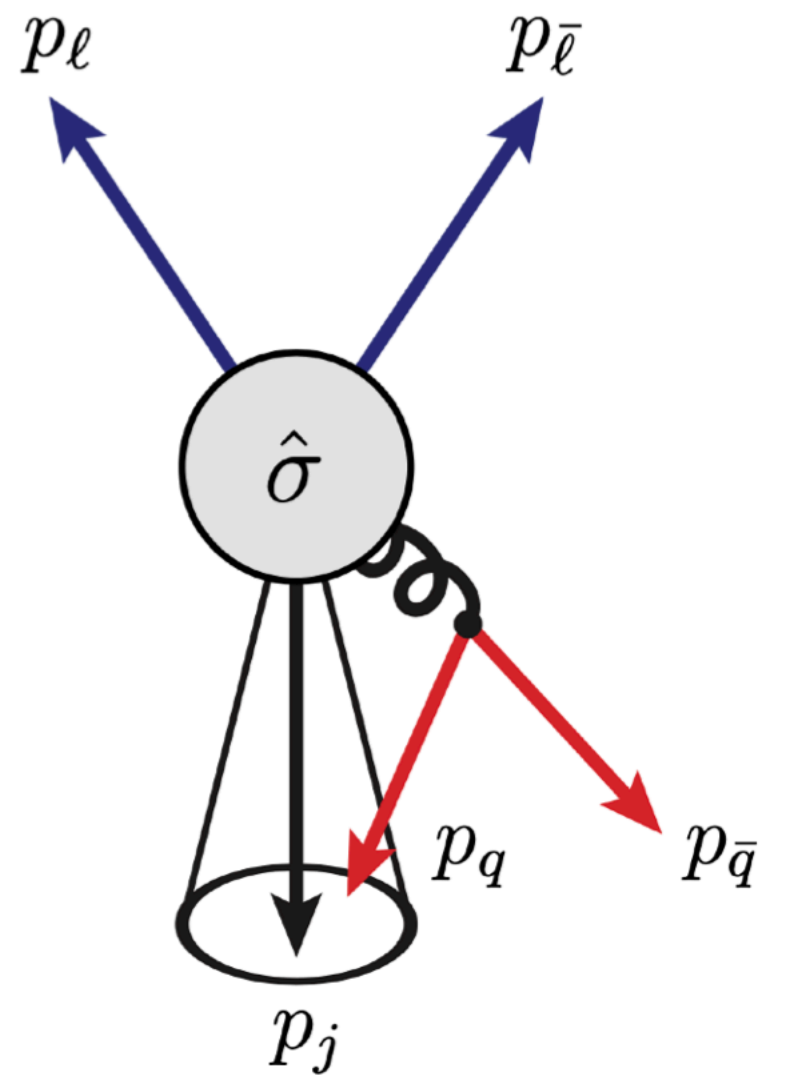
Open challenges:

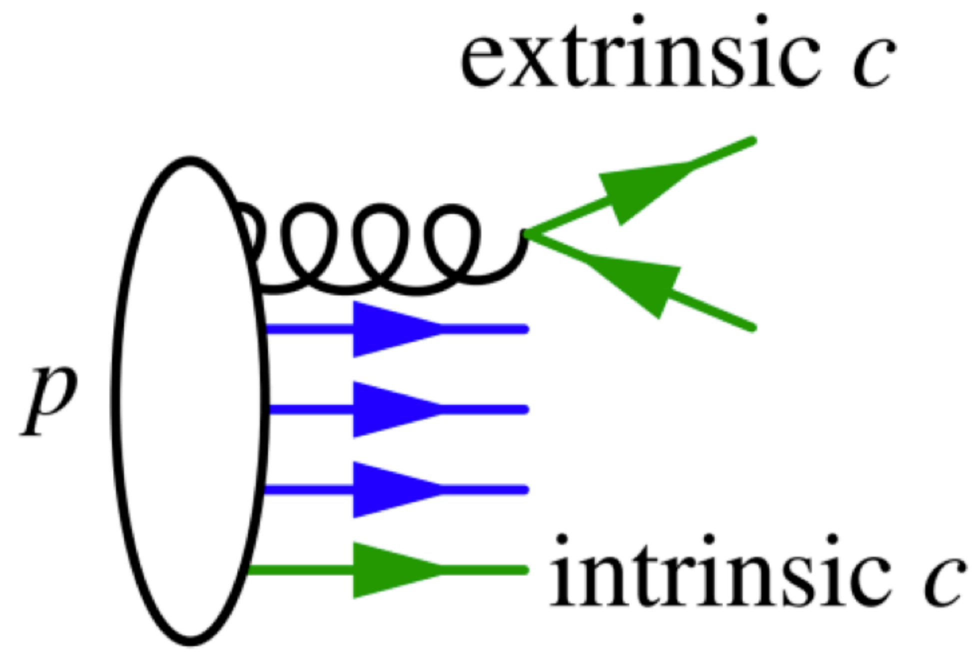
- Better grasp of MHOU
- Improved jet algorithms

in order to match data and predictions…

collinear-safe jet algorithms need to be used

See talks on Monday session for more about the theoretical and experimental challenges on this topic

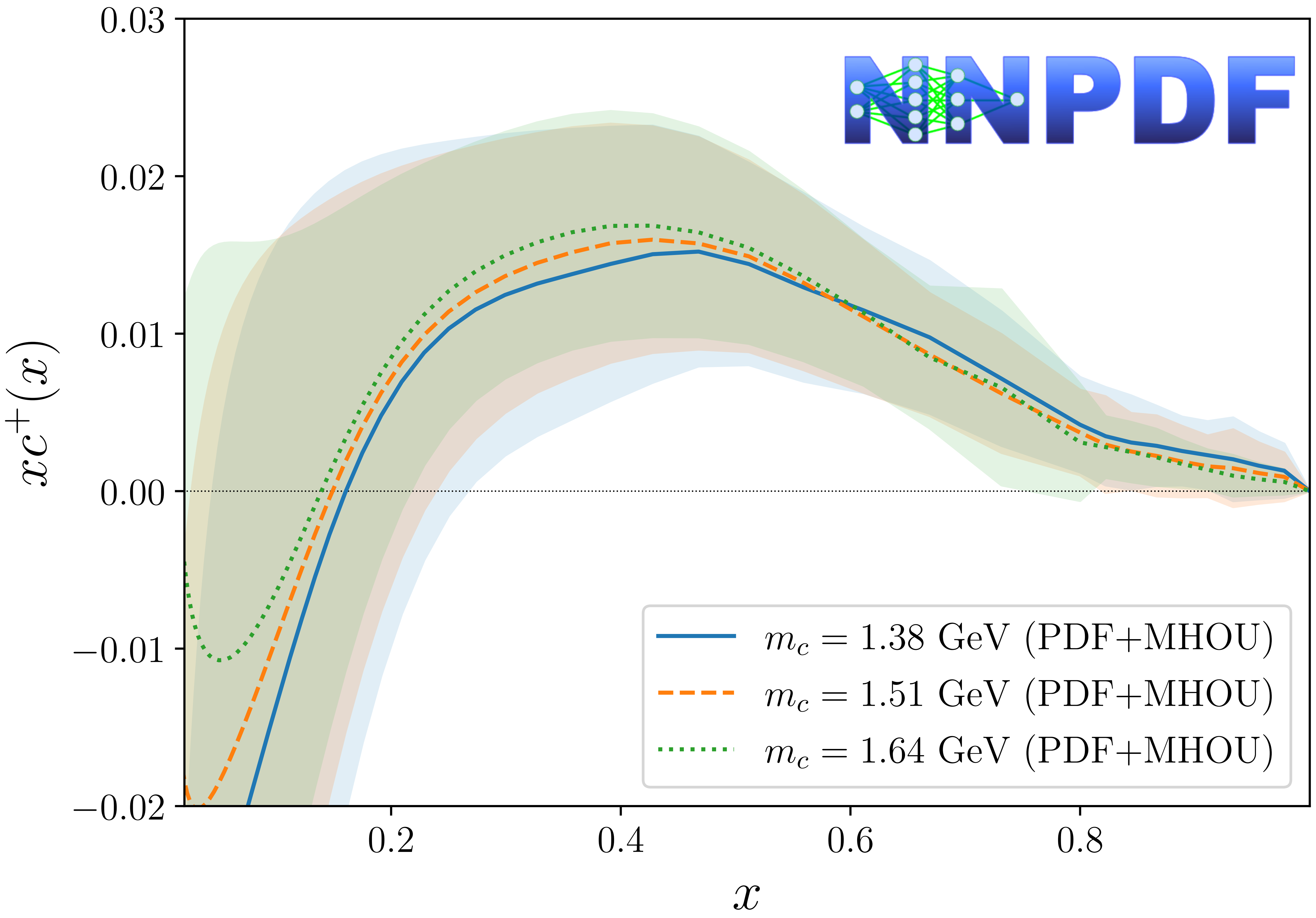




Diagrams taken from talks by G. Stagnitto and D. Zuliani

Intrinsically charming

Evidence for intrinsic charm quarks in the proton
hep-ph/2208.08372



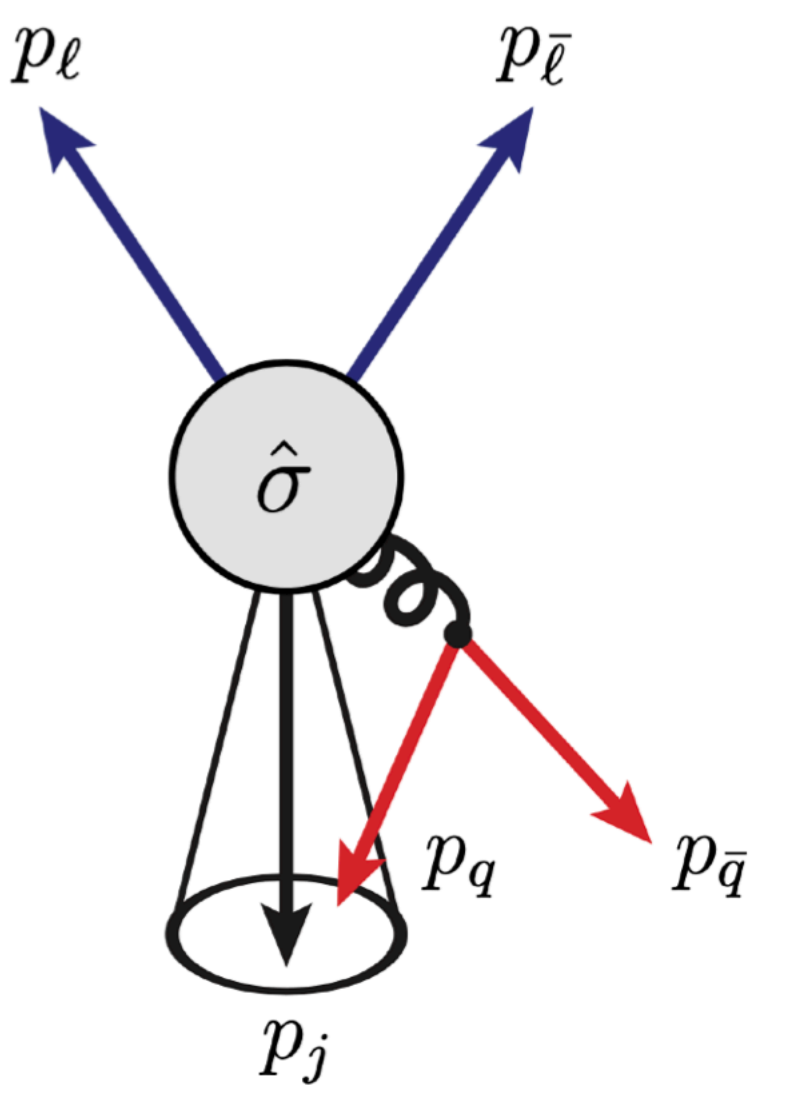
NNPDF is the only collaboration which fits charm by default, i.e., $c^+ = c + \bar{c} \neq 0$ at the fitting scale which means the contribution is not limited to DGLAP evolution

Open challenges:

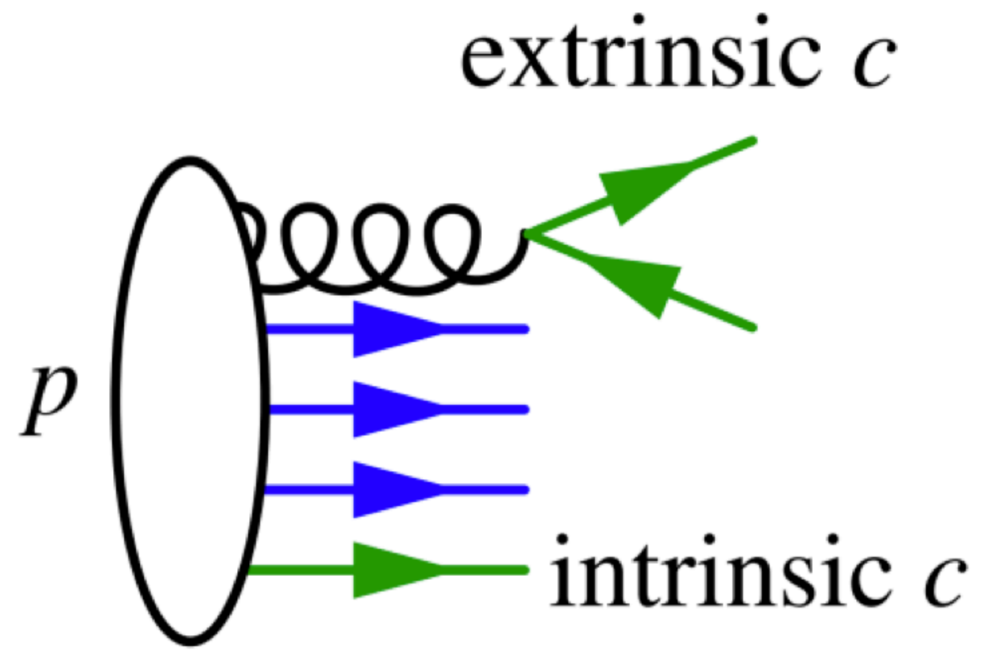
- Better grasp of MHOU
- Improved jet algorithms

in order to match data and predictions…

collinear-safe jet algorithms need to be used



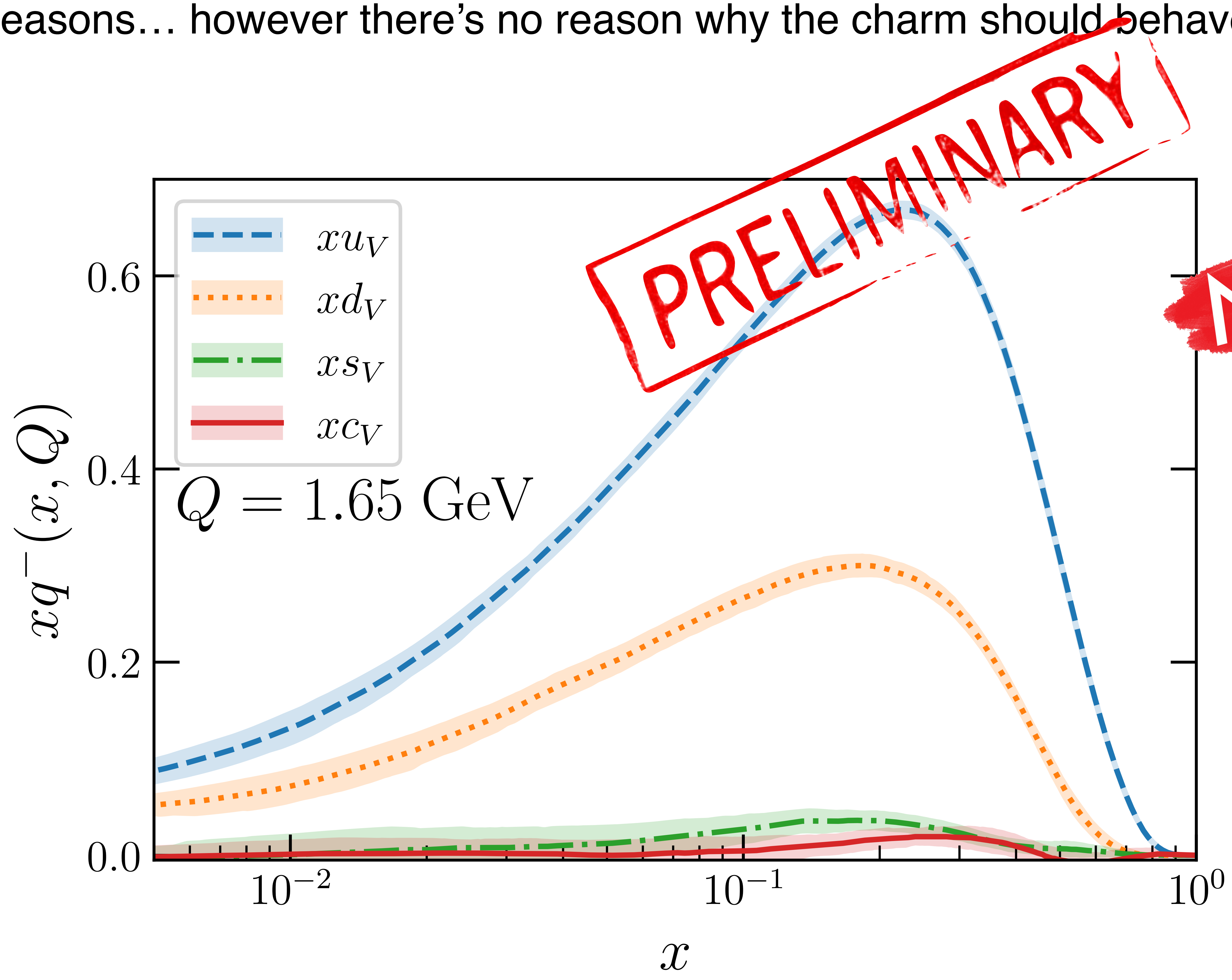
See talks on Monday session for more about the theoretical and experimental challenges on this topic

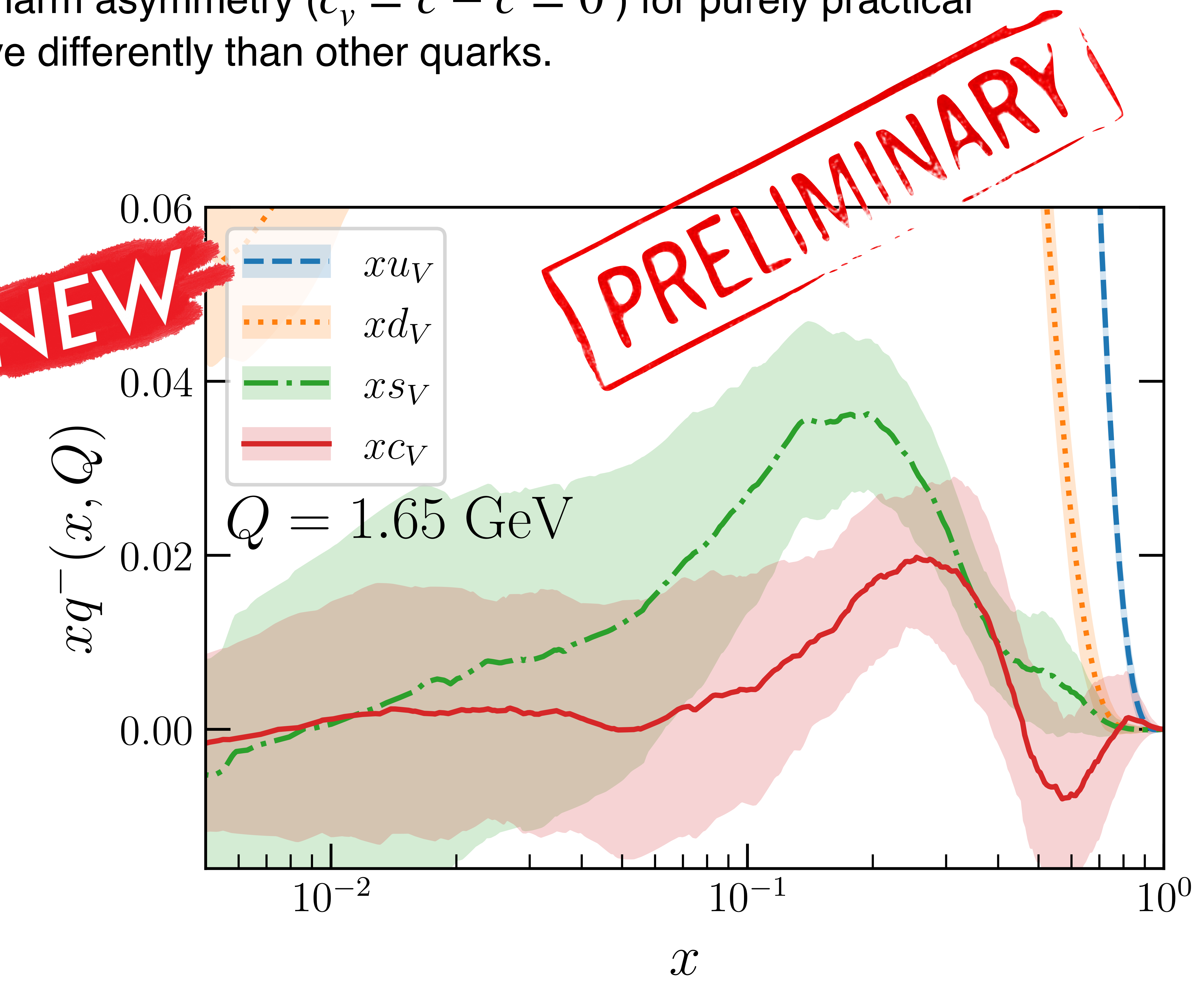


Diagrams taken from talks by G. Stagnitto and D. Zuliani

~~Intrinsically~~ Asymmetrically charming

The determination of the charm content of the proton assumed 0 charm asymmetry ($c_v = c - \bar{c} = 0$) for purely practical reasons… however there's no reason why the charm should behave differently than other quarks.





Beyond the NNLO QCD PDF

The *true* NNLO PDF

While technically the PDFs are truly NNLO orders (two extra orders in α_s) this come with caveats:

- The evolution from the fitting scale to the process scale is performed exactly at NNLO
- … but the prediction often relies on the “k-factor approximation”…

i.e., grids exact up to only NLO with the NNLO contribution applied bin-by-bin on the experimental data: integrating out flavour decomposition or x-dependence.

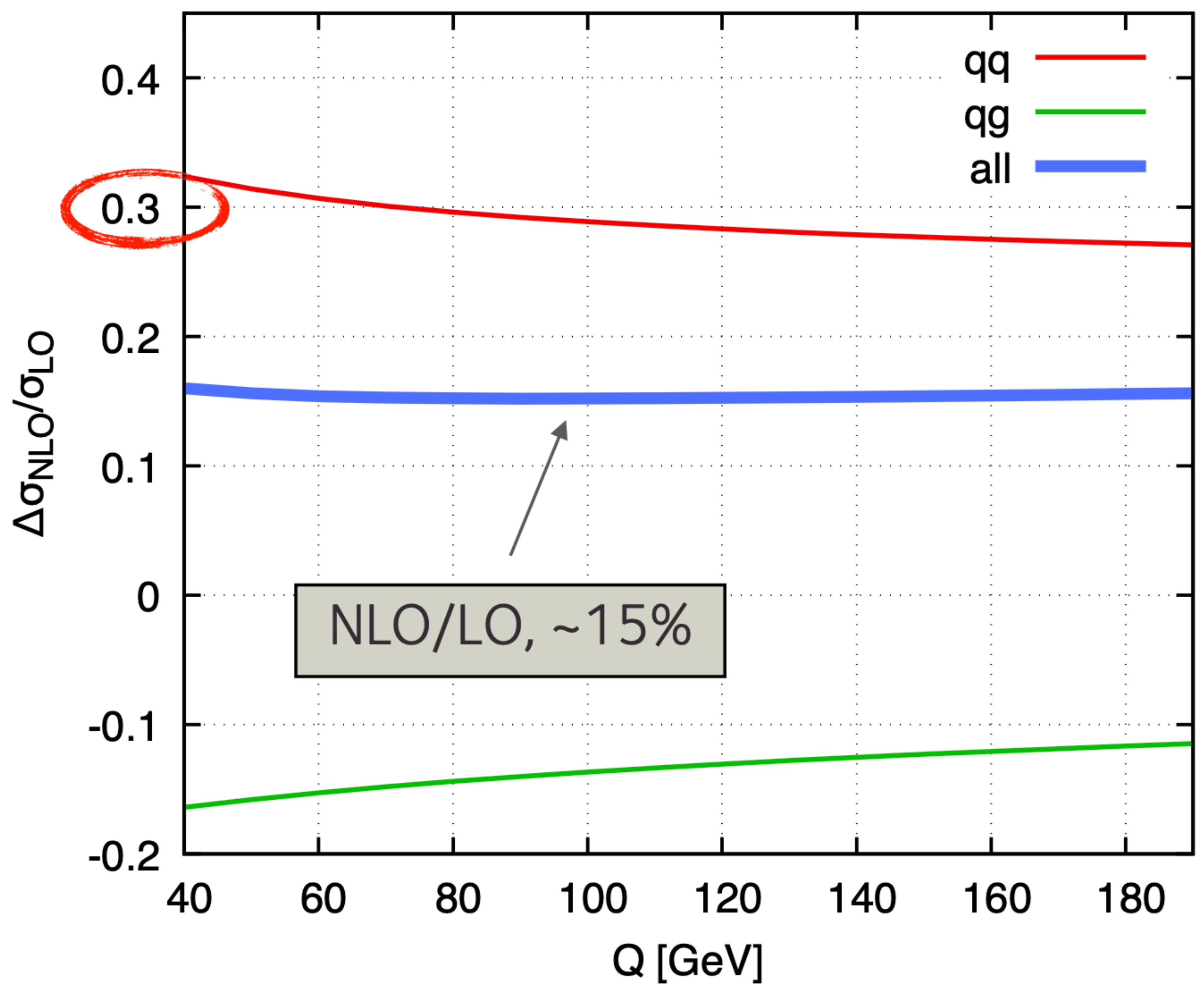
We need NNLO-accurate grids, differential in x, flavour (and hopefully Q)!

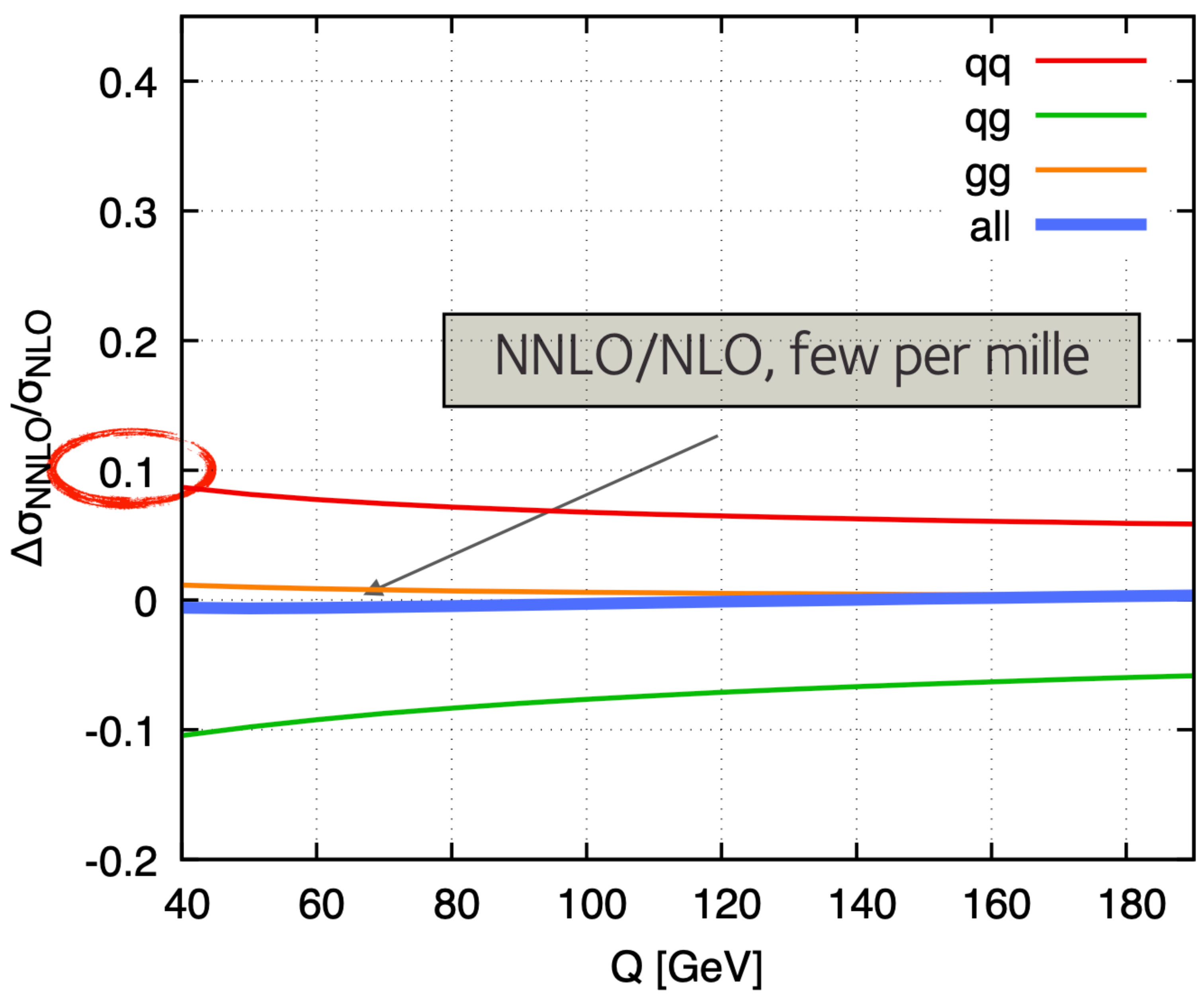
see how in hep-ph/2302.12124

Some active efforts to make grids available for the community:

[Plougshare](#)

[Pineline](#)

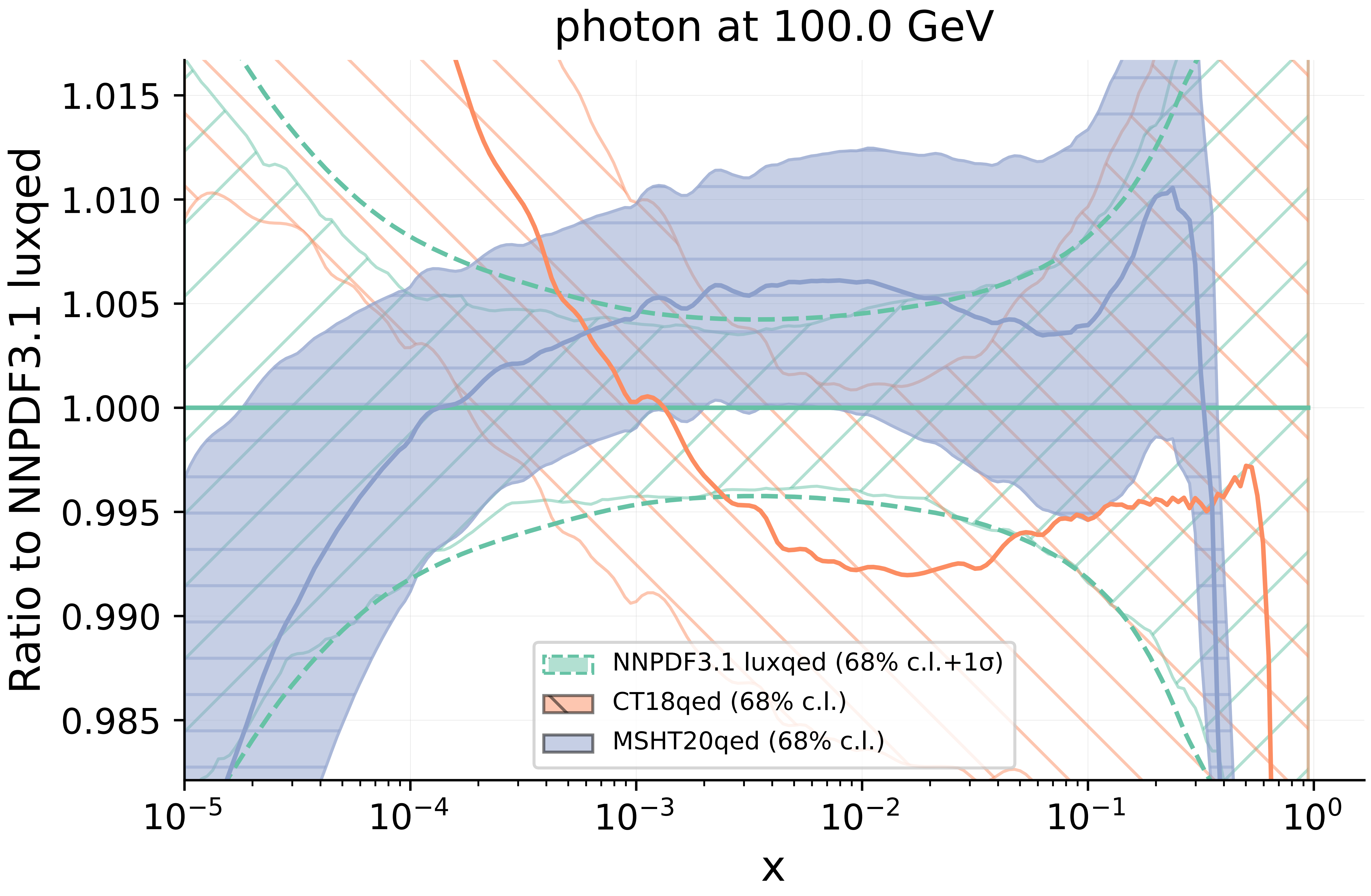


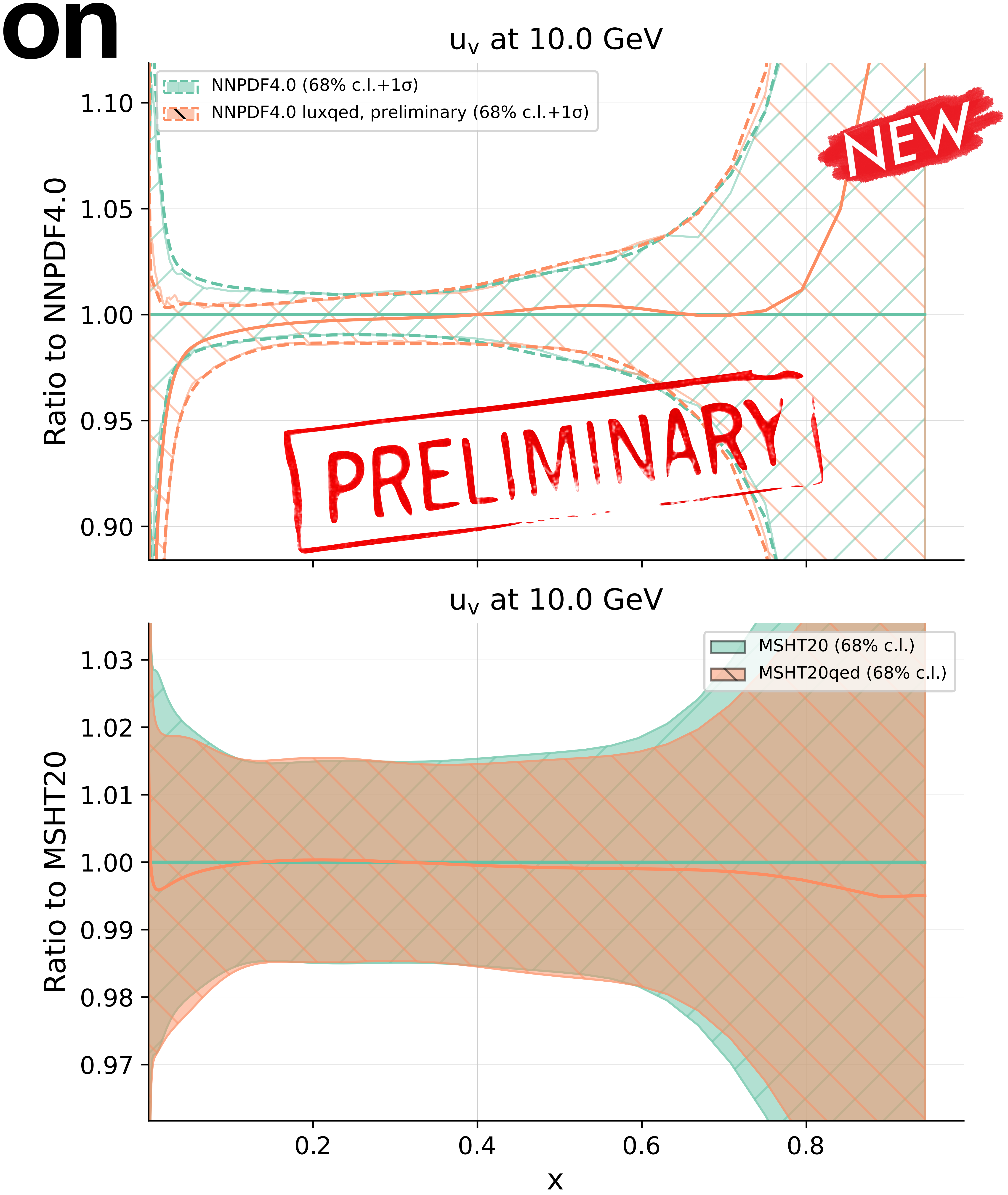


plots by F. Caola [link]

The photon inside the proton

All three groups utilize the LUXqed formalism hep-ph/1607.04266 to generate the photon.





Missing Higher Orders

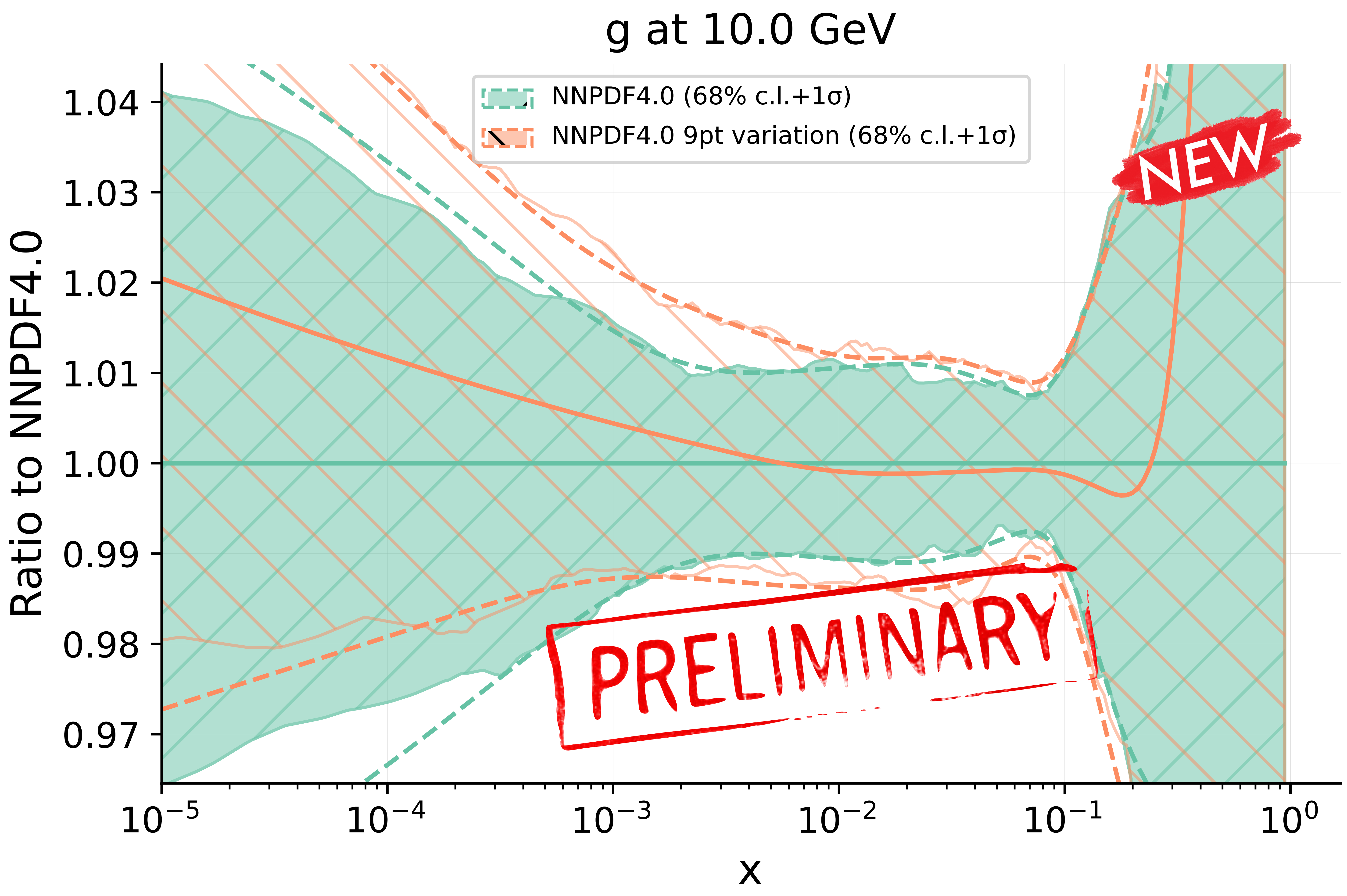
Uncertainties beyond the data

PDF uncertainties are propagate mainly from the data but this is just half of the story, fixed-order prediction also contain uncertainties:

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

A spurious dependence on unphysical scales (renormalization, factorization) is kept. This is exploited to generate a theory uncertainty. Two possible approaches:

- Modifying the covariance matrix 1906.10698
- Monte Carlo sampling of scales 2207.07616



N3LO: the next frontier

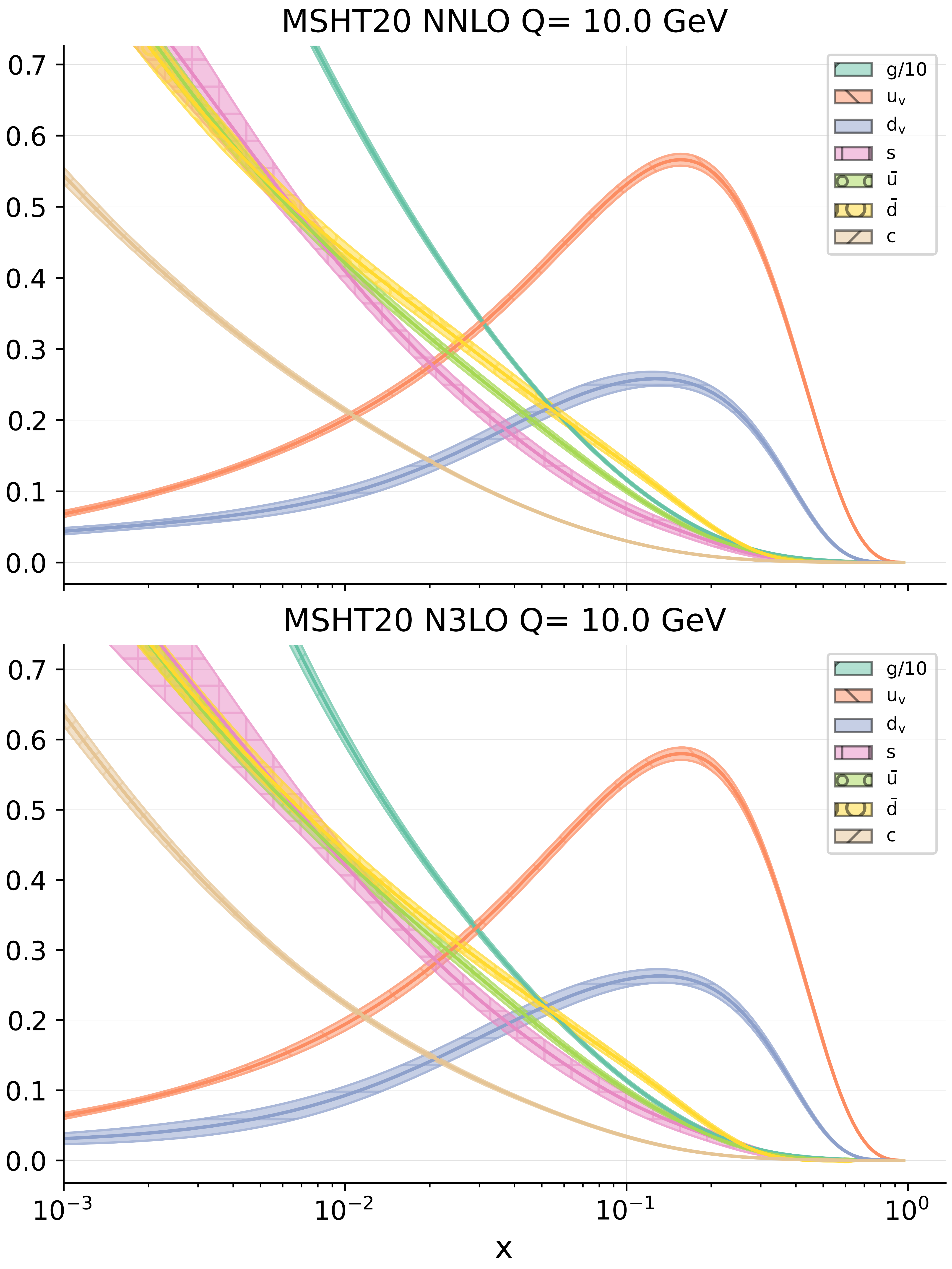
Open challenges:

- Exact N3LO evolution
- NNLO grids, instead of k-factors
- N3LO grids, instead of k-factors!
- Computationally very complex (diminishing returns on precision gain Vs computational cost)

First approximated results by MSHT collaboration

- Using K-factors for fixed-order predictions
- Splitting functions only partially known, approximated evolution.
- Exploit knowledge about N3LO results to constrain the PDF and fit (with uncertainties) the unknown pieces.

See talks by F. Buccioni and G. Zanderighi for more on recent progresss in N3LO calculations.



N3LO: news from NNPDF

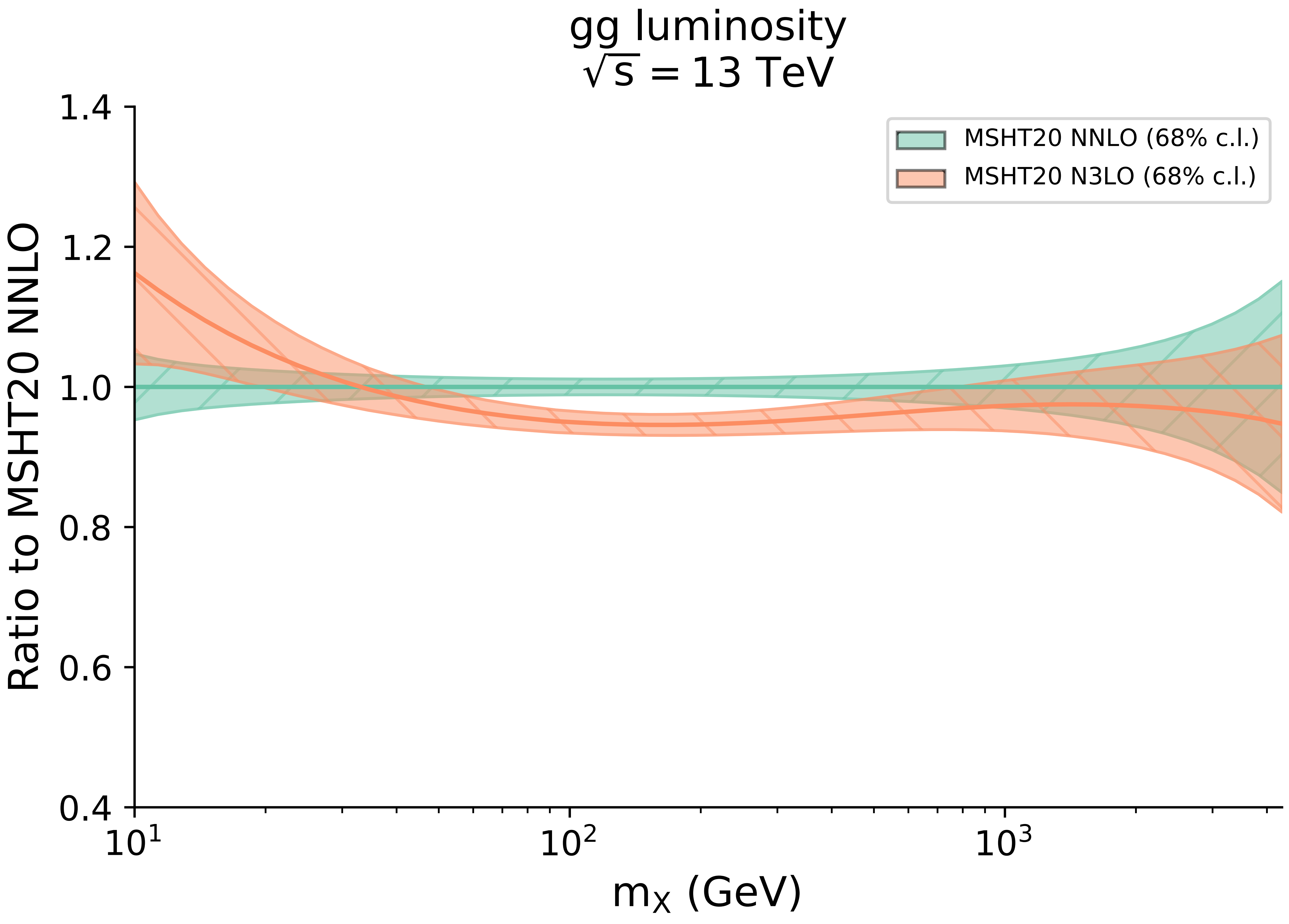
Ongoing work on the implementation of the known bits for the splitting functions at N3LO in

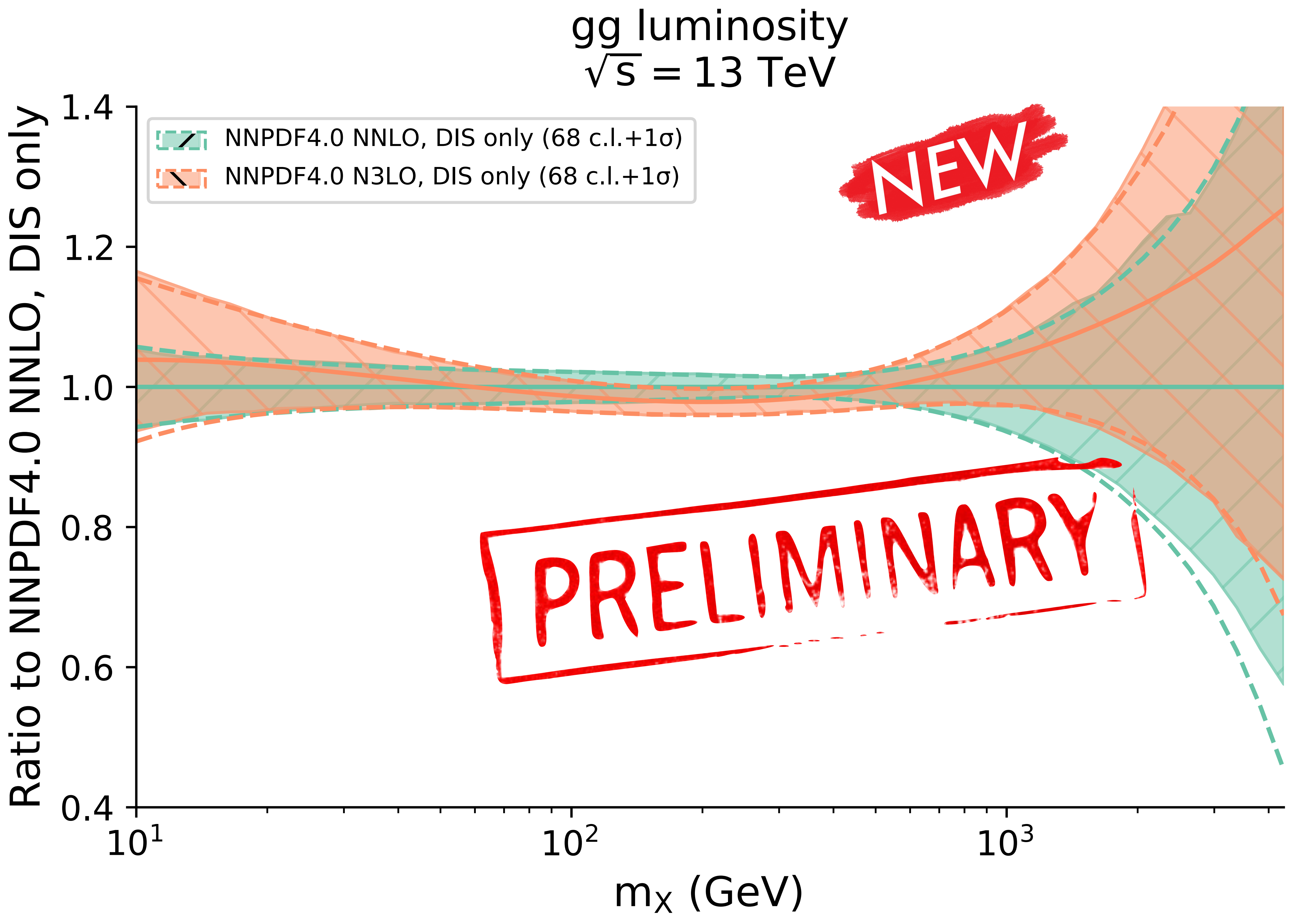
EKO
Evolution Kernel Operators

(see EKO documentation and references therein)

Caveats for NNPDF plot:

- DIS only fit (both N3LO and NNLO)
- Includes MHOU at both N3LO and NNLO
- Uncertainties include also incomplete knowledge of the N3LO splitting functions





Conclusions

Enough data available that we can start talking about the precision age of Parton Distribution Functions.

1. Uncertainties in the data region can be of order 1%
2. Including an estimation of the uncertainties due to Missing Higher Order Uncertainties becomes important.
3. N3LO will bring the determination of PDFs to the same level of accuracy of Higher Order Predictions (ingredients missing!)

With great power comes great responsibility:

1. A close scrutiny of PDF uncertainties is important: theory, data.
2. For that more public codes are necessary from PDF fitters (like NNPDF and xfitter) but also Monte Carlo generators.
3. Systematic ways of testing fitting methodologies for flexibility and reliability (e.g., closure tests)

Thank you!

Backup

Why are grids needed?

Impossible to perform a numerical integral for every iteration of a fit.
We need some approximation

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

The evolution on the μ scales is exact (O(α^2)) and the PDF depend then only on the flavour and the momentum fraction: $\dfrac{d^2\hat{\sigma}_{ij}}{dx_1 dx_2}$

x-grid

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

flavours

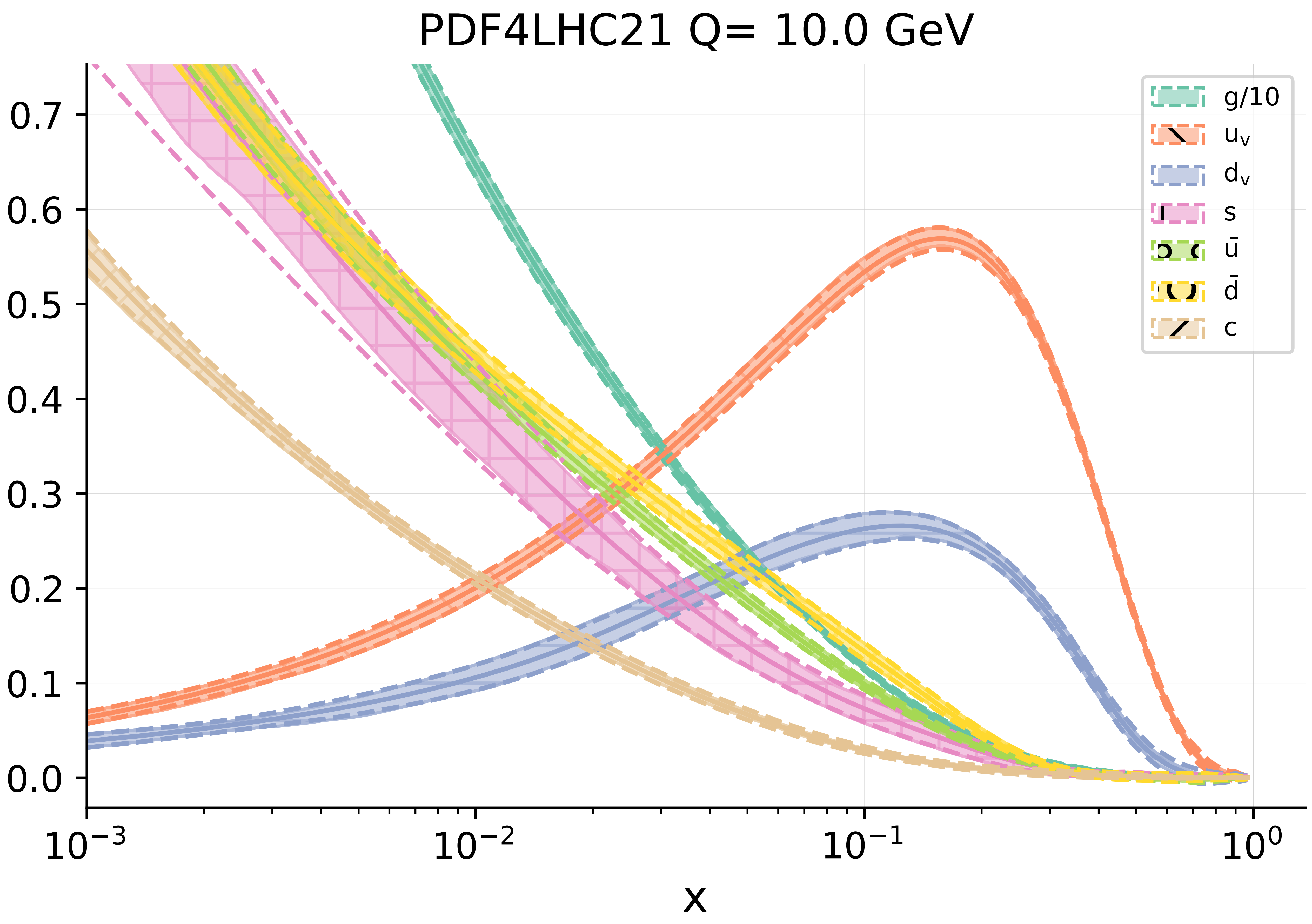
With O(50) points we can get a good representation of most observables, i.e., for each step of the fitting process we *just* need to contract the PDF with a tensor of *only* $4500 \times 50 \times 50 \times 14 \times 14 \simeq 10^9$ elements. Easy!

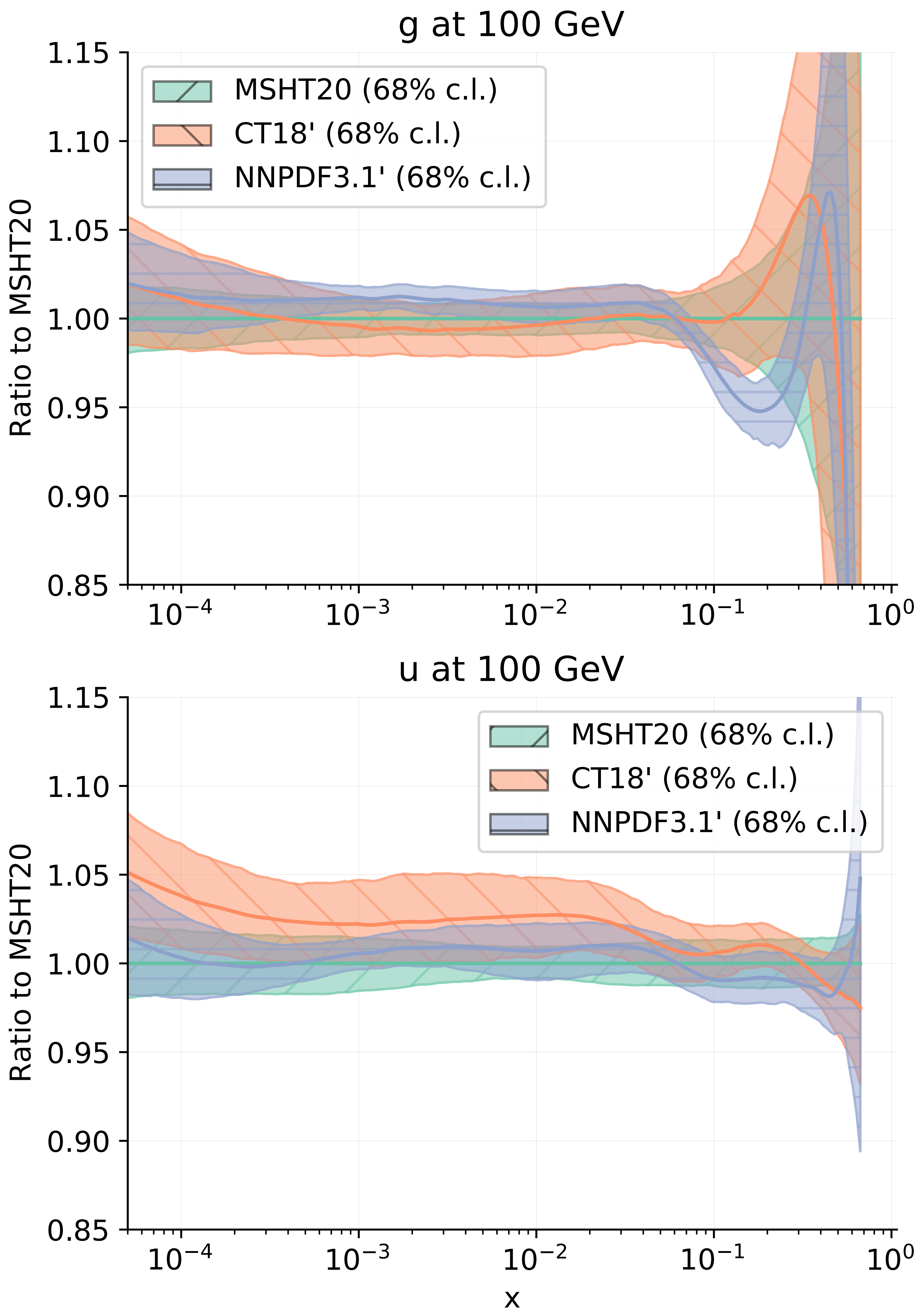
But actually, the convolution of such a big array with the luminosity takes roughly 1 second!

fastNLO APPLgrid

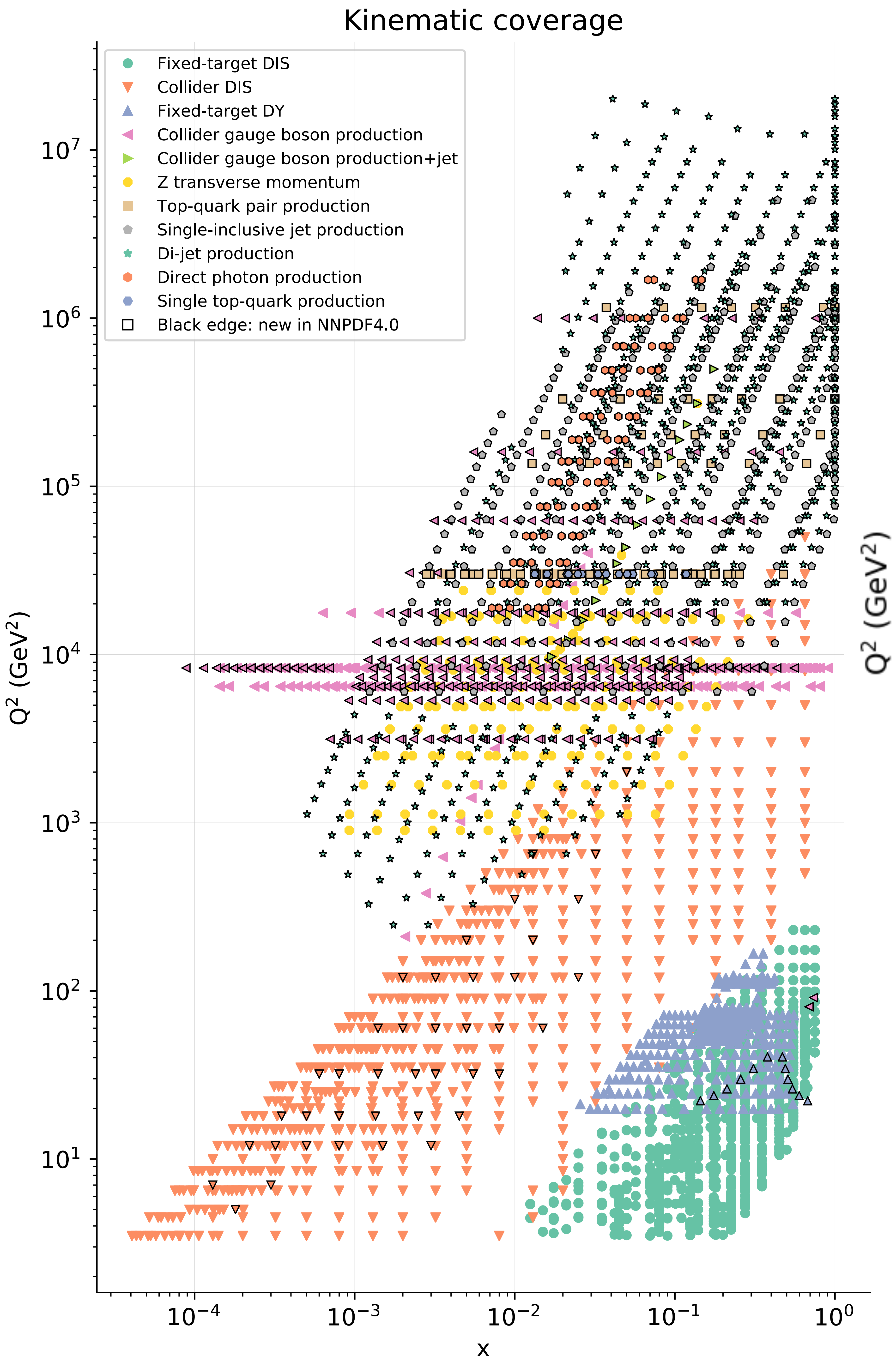
PineAPPL
General process, PDF-independent, grid storage

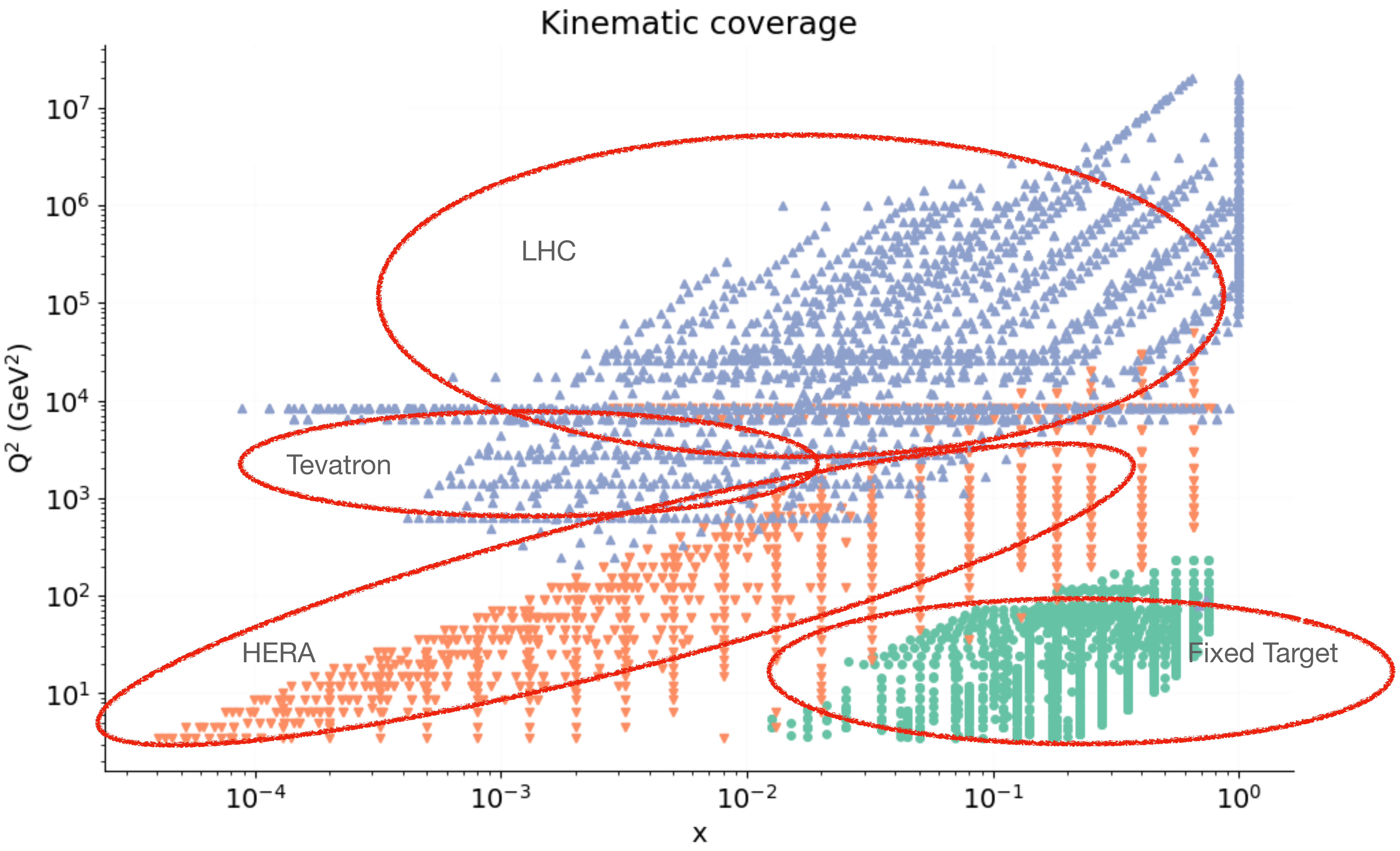
PDF4LHC combination





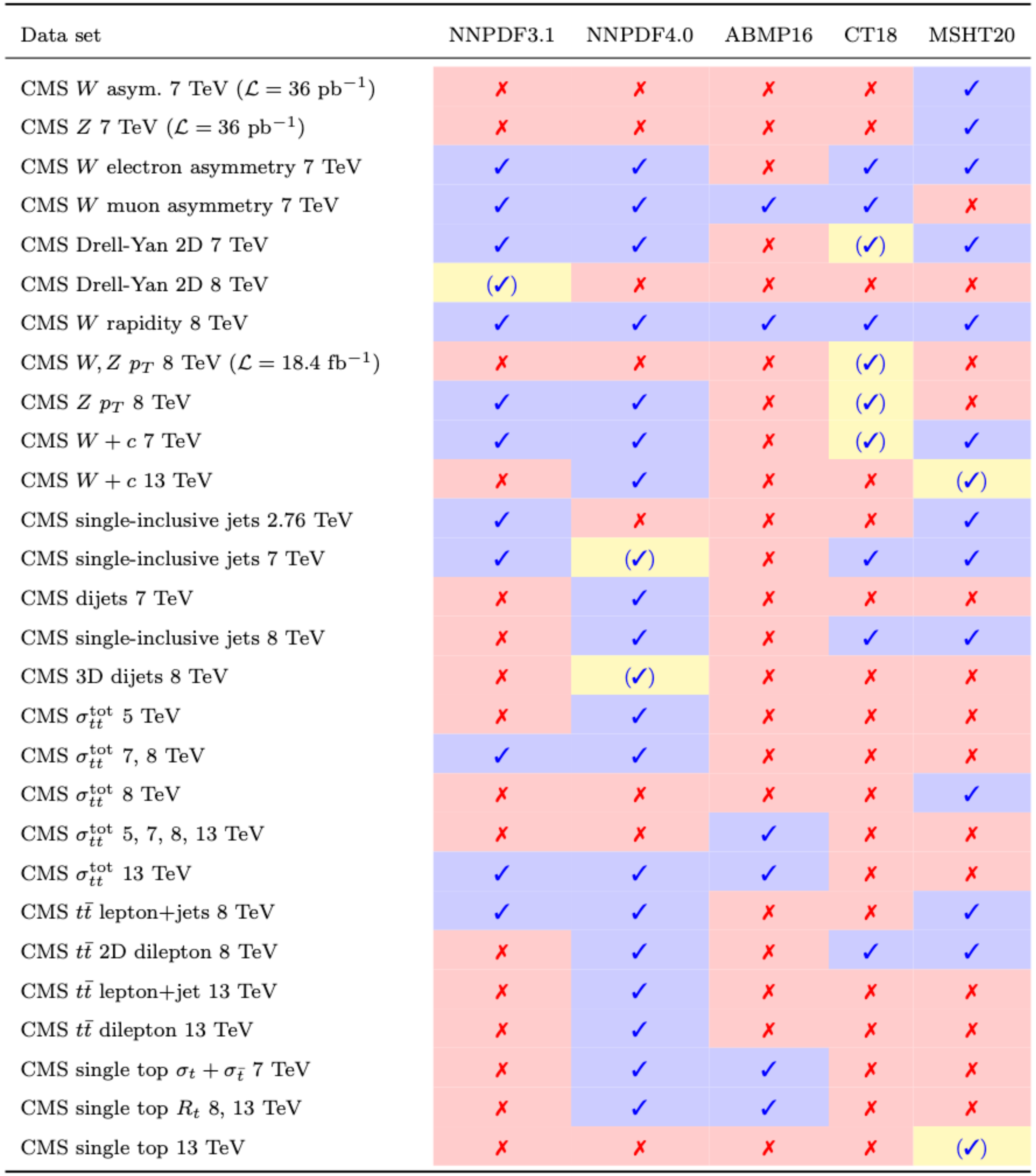
- NNPDF31’ (changes to m_c and dataset)
- CT18’ (changes to m_c)
- MSHT20



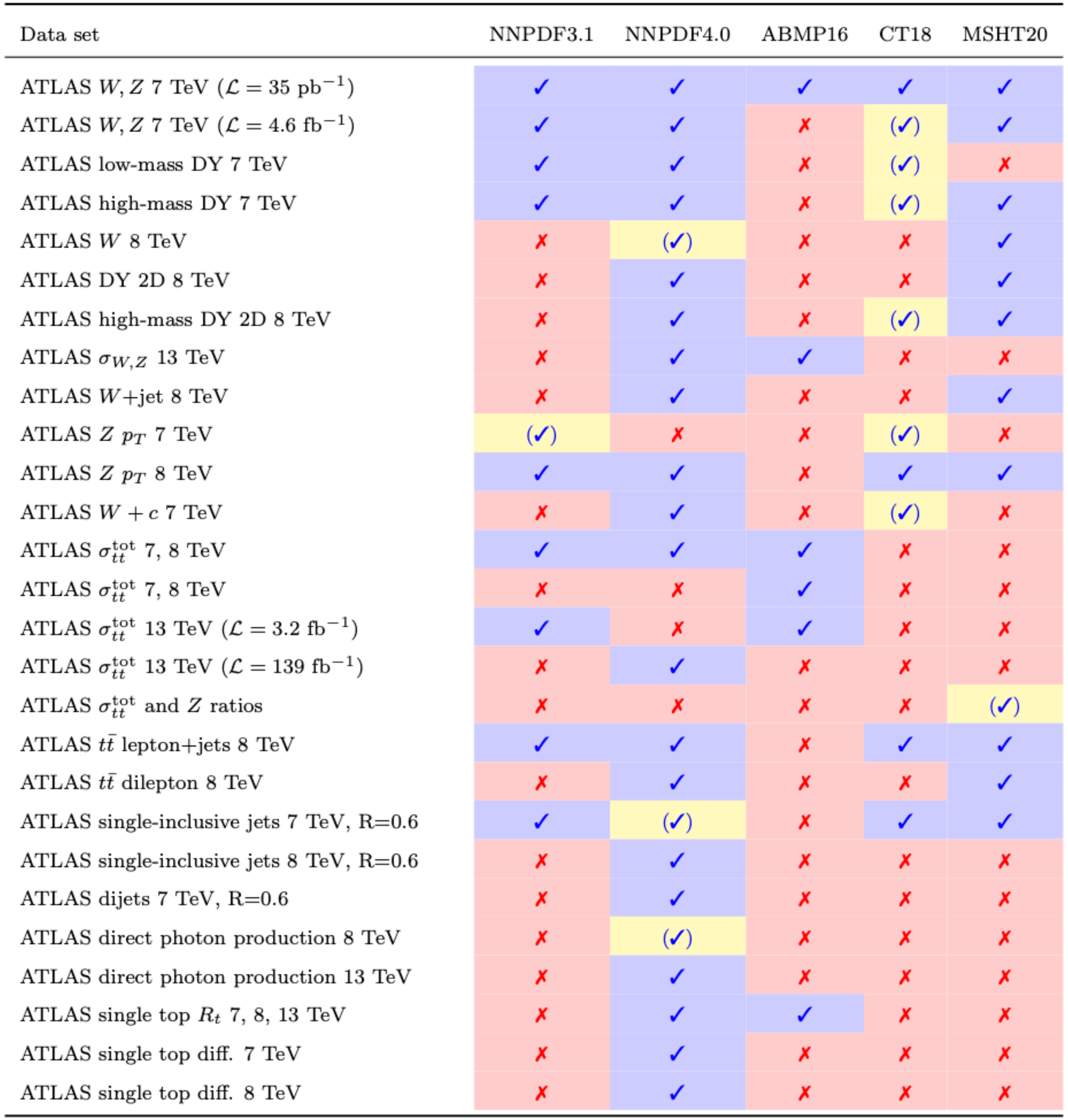


Using NNPDF4.0 dataset as a reference. Full list of datasets in this plot can be checked in Appendix B of the NNPDF4.0 paper: link

List of datasets



Data set	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
CMS W asym. 7 TeV ($\mathcal{L}=36$ pb^{-1})	✗	✗	✗	✗	✓
CMS Z 7 TeV ($\mathcal{L}=36$ pb^{-1})	✗	✗	✗	✗	✓
CMS W electron asymmetry 7 TeV	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	✓	✓	✗	(✓)	✓
CMS Drell-Yan 2D 8 TeV	(✓)	✗	✗	✗	✗
CMS W rapidity 8 TeV	✓	✓	✓	✓	✓
CMS W, Z p_T 8 TeV ($\mathcal{L}=18.4$ fb^{-1})	✗	✗	✗	(✓)	✗
CMS Z p_T 8 TeV	✓	✓	✗	(✓)	✗
CMS $W+c$ 7 TeV	✓	✓	✗	(✓)	✓
CMS $W+c$ 13 TeV	✗	✓	✗	✗	(✓)
CMS single-inclusive jets 2.76 TeV	✓	✗	✗	✗	✓
CMS single-inclusive jets 7 TeV	✓	(✓)	✗	✓	✓
CMS dijets 7 TeV	✗	✓	✗	✗	✗
CMS single-inclusive jets 8 TeV	✗	✓	✗	✓	✓
CMS 3D dijets 8 TeV	✗	(✓)	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 5 TeV	✗	✓	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 8 TeV	✗	✗	✗	✗	✓
CMS $\sigma_{tt}^{\rm tot}$ 5, 7, 8, 13 TeV	✗	✗	✓	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	✗	✓	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	✗	✓	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	✗	✓	✗	✗	✗
CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	✗	✓	✓	✗	✗
CMS single top R_t 8, 13 TeV	✗	✓	✓	✗	✗
CMS single top 13 TeV	✗	✗	✗	✗	(✓)



Data set	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV ($\mathcal{L}=35$ pb^{-1})	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV ($\mathcal{L}=4.6$ fb^{-1})	✓	✓	✗	(✓)	✓
ATLAS low-mass DY 7 TeV	✓	✓	✗	(✓)	✗
ATLAS high-mass DY 7 TeV	✓	✓	✗	(✓)	✓
ATLAS W 8 TeV	✗	(✓)	✗	✗	✓
ATLAS DY 2D 8 TeV	✗	✓	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	✗	✓	✗	(✓)	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	✗	✓	✓	✗	✗
ATLAS W+jet 8 TeV	✗	✓	✗	✗	✓
ATLAS Z p_T 7 TeV	(✓)	✗	✗	(✓)	✗
ATLAS Z p_T 8 TeV	✓	✓	✗	✓	✓
ATLAS $W+c$ 7 TeV	✗	✓	✗	(✓)	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✗	✗	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV ($\mathcal{L}=3.2$ fb^{-1})	✓	✗	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV ($\mathcal{L}=139$ fb^{-1})	✗	✓	✗	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ and Z ratios	✗	✗	✗	✗	(✓)
ATLAS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	✗	✓	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	✓	(✓)	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	✗	✓	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	✗	✓	✗	✗	✗
ATLAS direct photon production 8 TeV	✗	(✓)	✗	✗	✗
ATLAS direct photon production 13 TeV	✗	✓	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	✗	✓	✓	✗	✗
ATLAS single top diff. 7 TeV	✗	✓	✗	✗	✗
ATLAS single top diff. 8 TeV	✗	✓	✗	✗	✗

Data set	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
CDF Z rapidity	✓	✓	✗	✓	✓
CDF $W \to \ell\nu$ asymmetry (1.8 TeV)	✗	✗	✗	✓	✗
CDF $W \to e\nu$ asymmetry ($\mathcal{L} = 170$ pb^{-1})	✗	✗	✗	✓	✗
CDF $W \to e\nu$ asymmetry ($\mathcal{L} = 1$ fb^{-1})	✗	✗	✗	✗	✓
CDF k_t inclusive jets	✓	✗	✗	✗	✓
CDF cone-based inclusive jets	✗	✗	✗	✓	✗
D0 Z rapidity	✓	✓	✗	✓	✓
D0 $W \to e\nu$ asymmetry ($\mathcal{L} = 0.75$ fb^{-1})	✗	✗	✗	✗	✓
D0 $W \to e\nu$ (prod.) asymmetry ($\mathcal{L} = 9.7$ fb^{-1})	✗	✗	(✓)	✗	✓
D0 $W \to e\nu$ (prod. and decay) asymmetry ($\mathcal{L} = 9.7$ fb^{-1})	✓	(✓)	✓	✓	✗
D0 $W \to \mu\nu$ asymmetry ($\mathcal{L} = 0.3$ fb^{-1})	✗	✗	✗	✓	✗
D0 $W \to \mu\nu$ asymmetry ($\mathcal{L} = 7.3$ fb^{-1})	✓	✓	✓	✗	✓
D0 cone-based inclusive jets	✗	✗	✗	✓	✓
CDF and D0 top-pair production	✗	✗	(✓)	✗	✓
CDF and D0 single-top production	✗	✗	✓	✗	✗

Data set	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
DY E866 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$ (NuSea)	✓	✓	✓	✓	✓
DY E866 $\sigma^p_{\rm DY}$	✓	✓	✗	✓	✓
DY E605 $\sigma^p_{\rm DY}$	✓	✓	✓	✓	✗
DY E906 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$ (SeaQuest)	✗	✓	✗	✗	✗
LHCb Z 7 TeV ($\mathcal{L} = 940$ pb^{-1})	✓	✓	✗	✗	✓
LHCb $Z \to ee$ 8 TeV ($\mathcal{L} = 2$ fb^{-1})	✓	✓	✓	✓	✓
LHCb W 7 TeV ($\mathcal{L} = 37$ pb^{-1})	✗	✗	✗	✗	✓
LHCb $W, Z \to \mu$ 7 TeV	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 8 TeV	✓	✓	✓	✓	✓
LHCb $W \to e$ 8 TeV	✗	(✓)	✗	✗	✗
LHCb $Z \to \mu\mu, ee$ 13 TeV	✗	✓	✗	✗	✗

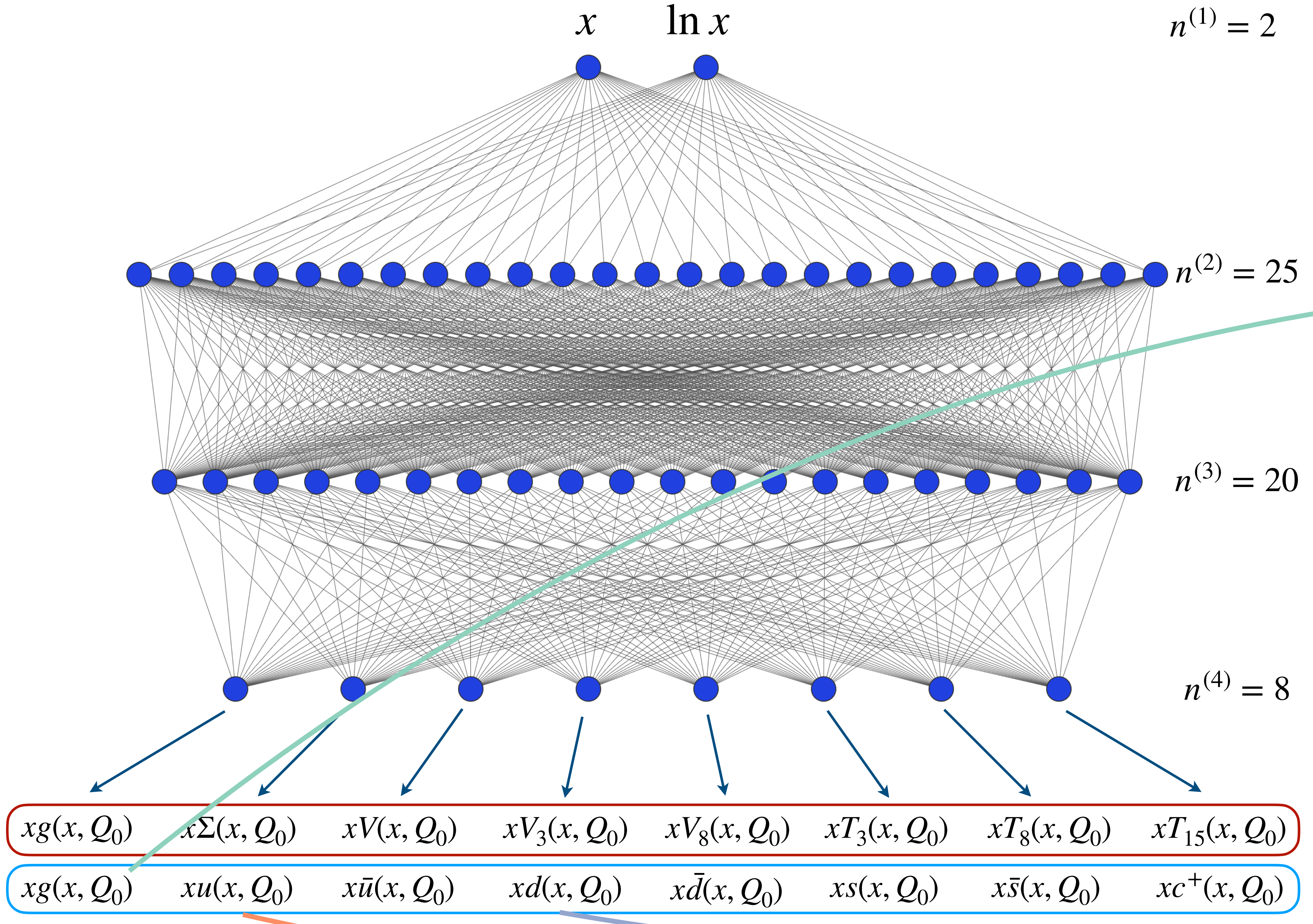
Data set	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
NMC F_2^d/F_2^p	✓	✓	✗	✗	✓
NMC $\sigma^{{\rm NC},p}$	✓	✓	✗	✓	✓
SLAC F_2^p, F_2^d	✓	✓	✓	✗	✓
BCDMS F_2^p	✓	✓	✓	✓	✓
BCDMS F_2^d	✓	✓	✗	✓	✓
BCDMS, NMC, SLAC F	✗	✗	✗	✗	✓
CHORUS $\sigma^{\nu}_{CC}, \sigma^{\bar{\nu}}_{CC}$	✓	✓	✗	✗	✓
CHORUS	✗	✗	✓	✗	✗
NuTeV F_2, F_3	✗	✗	✗	✗	✓
NuTeV/CCFR $\sigma^{\nu}_{CC}, \sigma^{\bar{\nu}}_{CC}$	✓	✓	✓	✓	✓
EMC F_2^c	(✓)	(✓)	✗	✗	✗
NOMAD	✗	(✓)	✓	✗	✗
CCFR xF_3^p	✗	✗	✗	✓	✗
CCFR F_2^p	✗	✗	✗	✓	✗
CDSHW F_2^p, xF_3^p	✗	✗	✗	✓	✗
E665 F_2^p, F_2^d	✗	✗	✗	✗	✓
HERA NC, CC	✗	✗	✗	✗	✓
HERA I+II $\sigma^p_{\rm NC,CC}$	✓	✓	✓	✓	✗
HERA I+II $\sigma^{\rm red}_{c\bar{c}}$	✗	✓	✗	(✓)	✓
HERA I+II $\sigma^{\rm red}_{b\bar{b}}$	✗	✓	✗	(✓)	✗
HERA I+II $\sigma^{\rm red}_{c\bar{c}}$	✓	✗	✓	✓	✗
H1 $F_2^{c\bar{c}}$	✗	✗	✗	✓	✗
H1 $F_2^{b\bar{b}}$	✓	✗	✓	✗	✗
ZEUS $\sigma^{\rm red}_{b\bar{b}}$	✓	✗	✓	✗	✗
H1 $F_{\rm L}$	✗	✗	✗	✓	✓
H1 and ZEUS $F_{\rm L}$	✗	✗	✗	✗	✓
ZEUS 820 (HQ) (1j)	✗	(✓)	✗	✗	✗
ZEUS 920 (HQ) (1j)	✗	(✓)	✗	✗	✗
H1 (LQ) (1j-2j)	✗	(✓)	✗	✗	✗
H1 (HQ) (1j-2j)	✗	(✓)	✗	✗	✗
ZEUS 920 (HQ) (2j)	✗	(✓)	✗	✗	✗

The NNPDF4.0 methodology

Neural Network

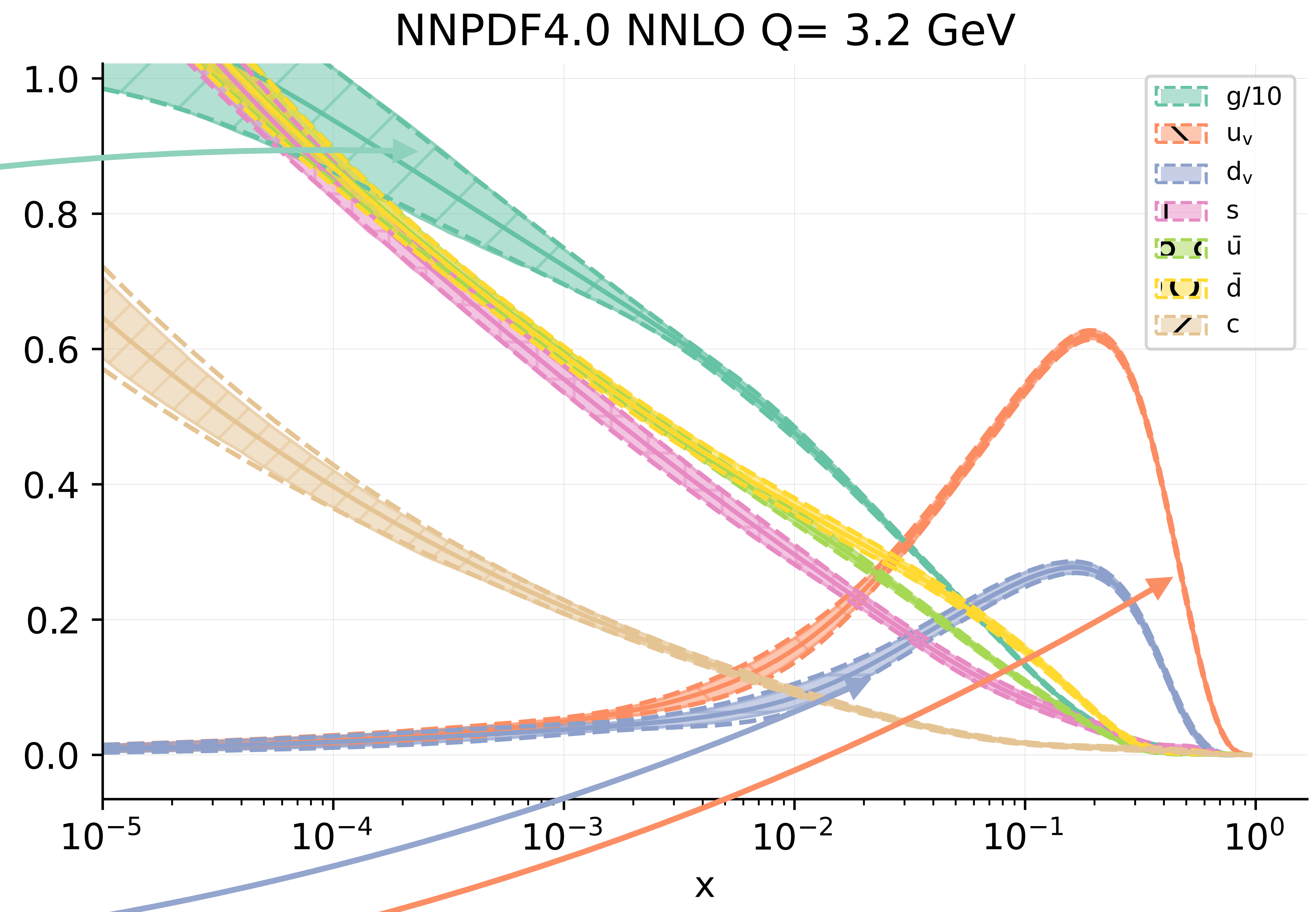
Parton Distribution Function

- Charm fitted $c+=c+\bar{c}\neq 0$, so not limited to perturbative evolution.
- Errors propagated via Monte Carlo replicas instead of hessian procedure.
- Central result as average of all replicas (instead of minimum of χ^2)

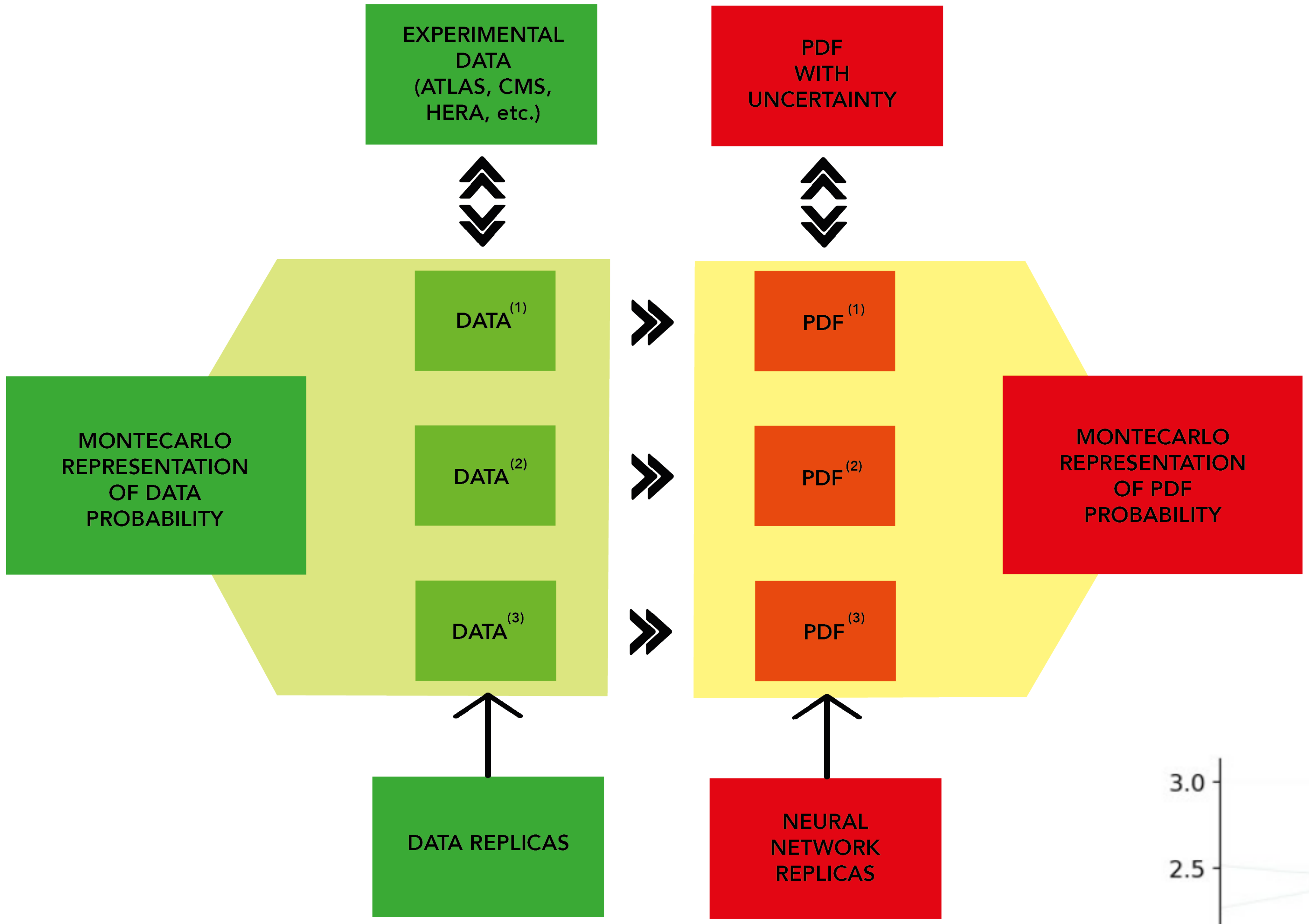


$xg(x,Q_0)$ $x\Sigma(x,Q_0)$ $xV(x,Q_0)$ $xV_3(x,Q_0)$ $xV_8(x,Q_0)$ $xT_3(x,Q_0)$ $xT_8(x,Q_0)$ $xT_{15}(x,Q_0)$

$xg(x,Q_0)$ $xu(x,Q_0)$ $x\bar{u}(x,Q_0)$ $xd(x,Q_0)$ $x\bar{d}(x,Q_0)$ $xs(x,Q_0)$ $x\bar{s}(x,Q_0)$ $xc^+(x,Q_0)$


NNPDF4.0 NNLO Q= 3.2 GeV

Monte Carlo Uncertainties, from data to PDF



EXPERIMENTAL DATA (ATLAS, CMS, HERA, etc.)

PDF WITH UNCERTAINTY

MONTECARLO REPRESENTATION OF DATA PROBABILITY

DATA$^{(1)}$

DATA$^{(2)}$

DATA$^{(3)}$

PDF$^{(1)}$

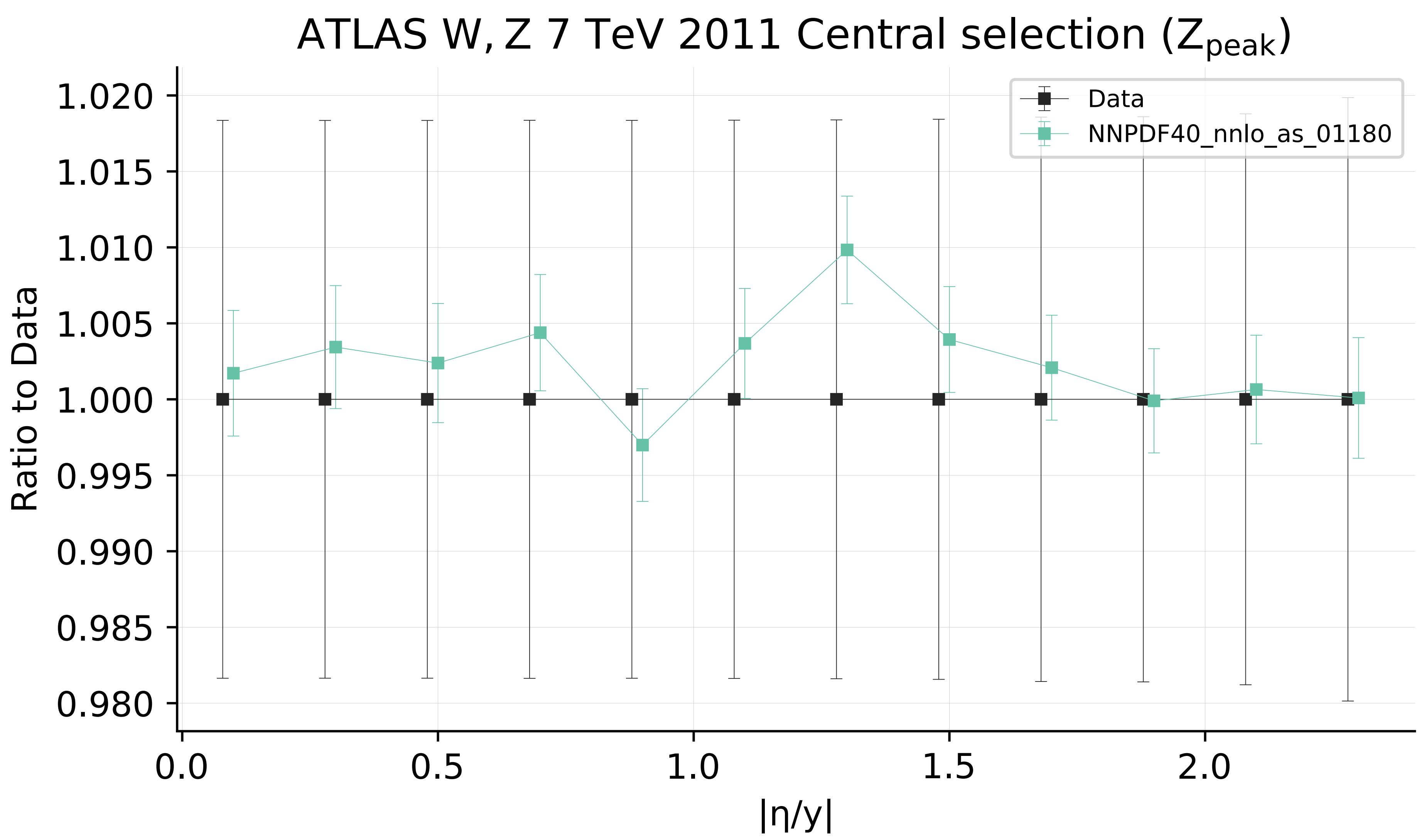
PDF$^{(2)}$

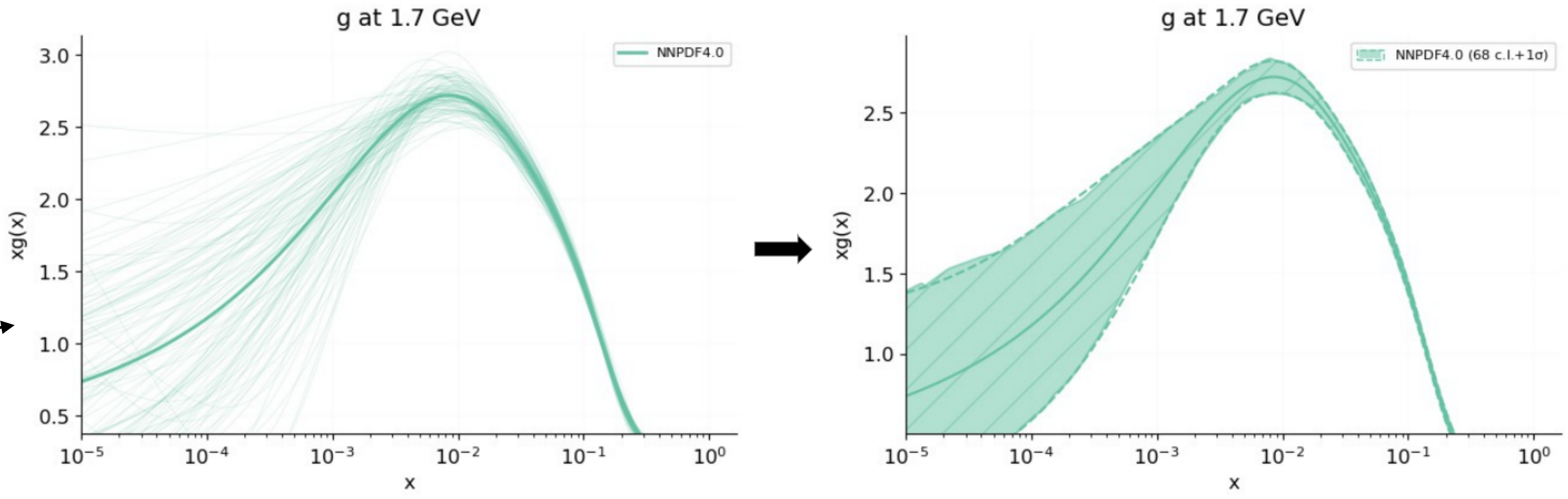
PDF$^{(3)}$

MONTECARLO REPRESENTATION OF PDF PROBABILITY

DATA REPLICAS

NEURAL NETWORK REPLICAS



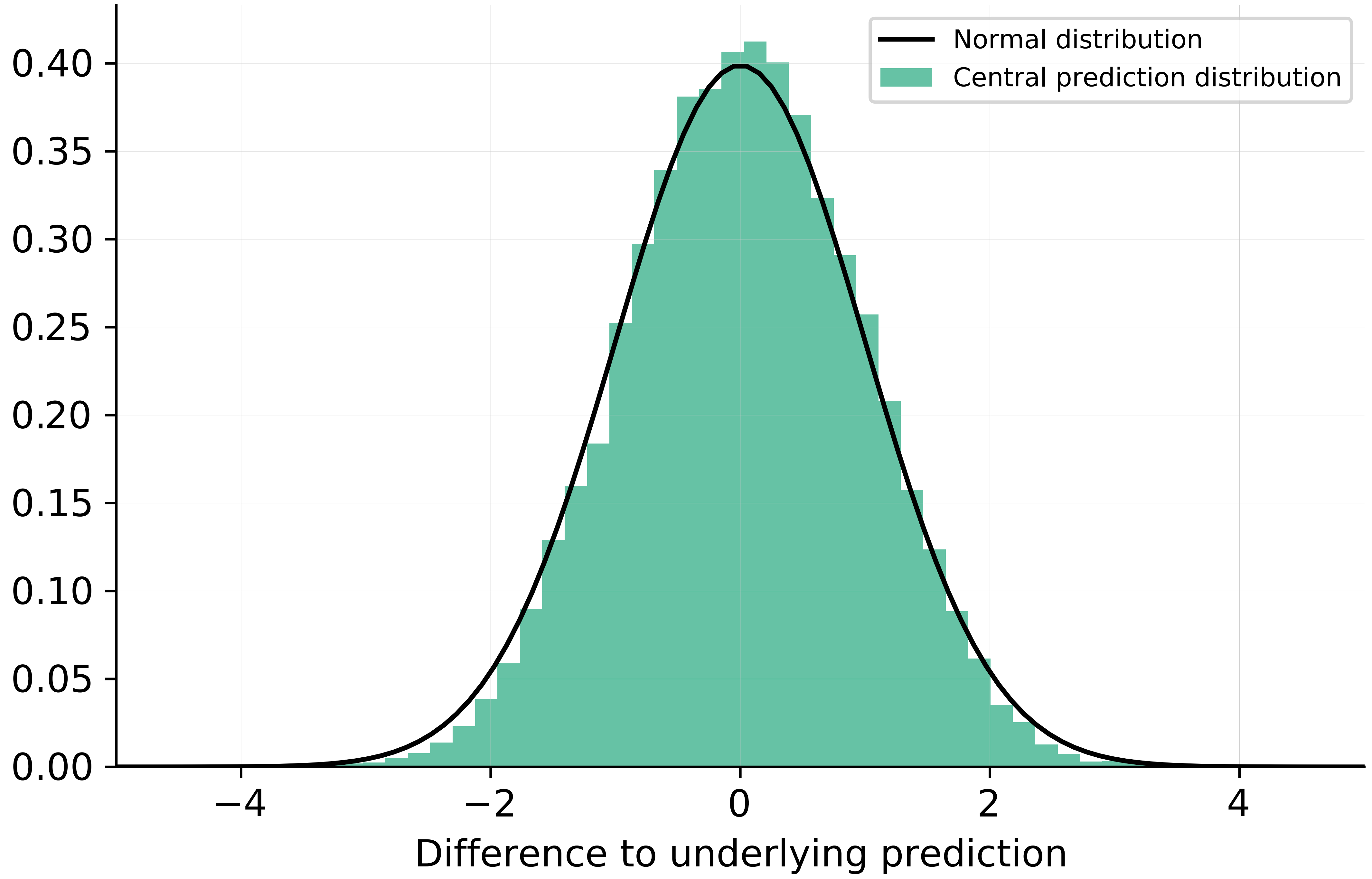


Perform thousand of fits, each to an “new” measure of the experimental data available.

Closure test

A powerful tool to test the reliability of a methodology.

1. Select some other PDF as the truth
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 known unknowns: missing higher order corrections, systematics, inconsistencies, etc…)

Missing Higher Orders at NLO

Uncertainties beyond the data

PDF uncertainties are propagate mainly from the data but this is just half of the story, fixed-order prediction also contain uncertainties:

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

A spurious dependence on unphysical scales (renormalization, factorization) is kept. This is exploited to generate a theory uncertainty. Two possible approaches:

- Modifying the covariance matrix 1906.10698
- Monte Carlo sampling of scales 2207.07616