



NSTAR 2026

**Juan M. Cruz
Martinez**
jacruz@us.es

Global analyses of PDFs
by **NNPDF**



**Sevilla -
September 2026**



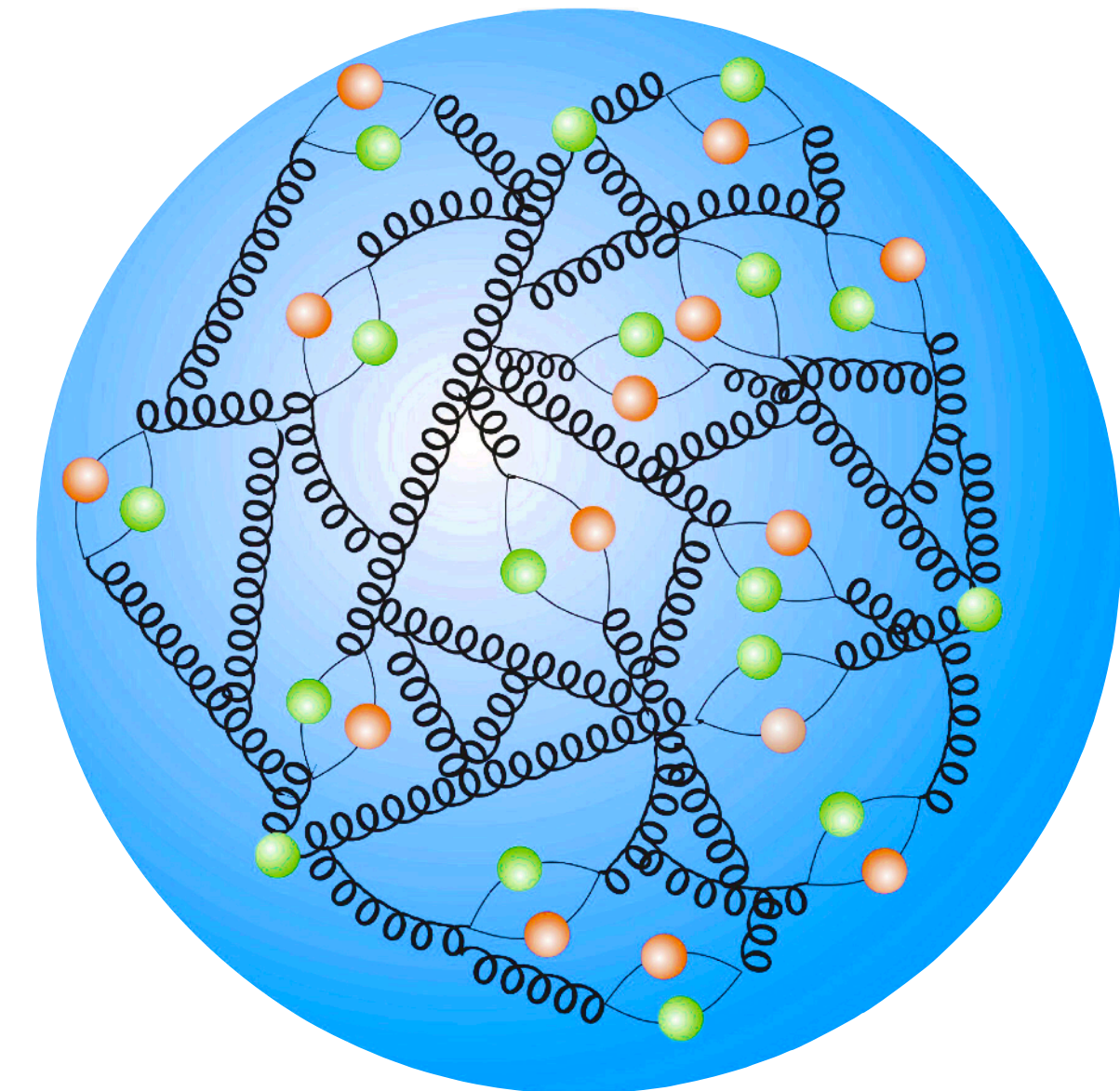
Introduction



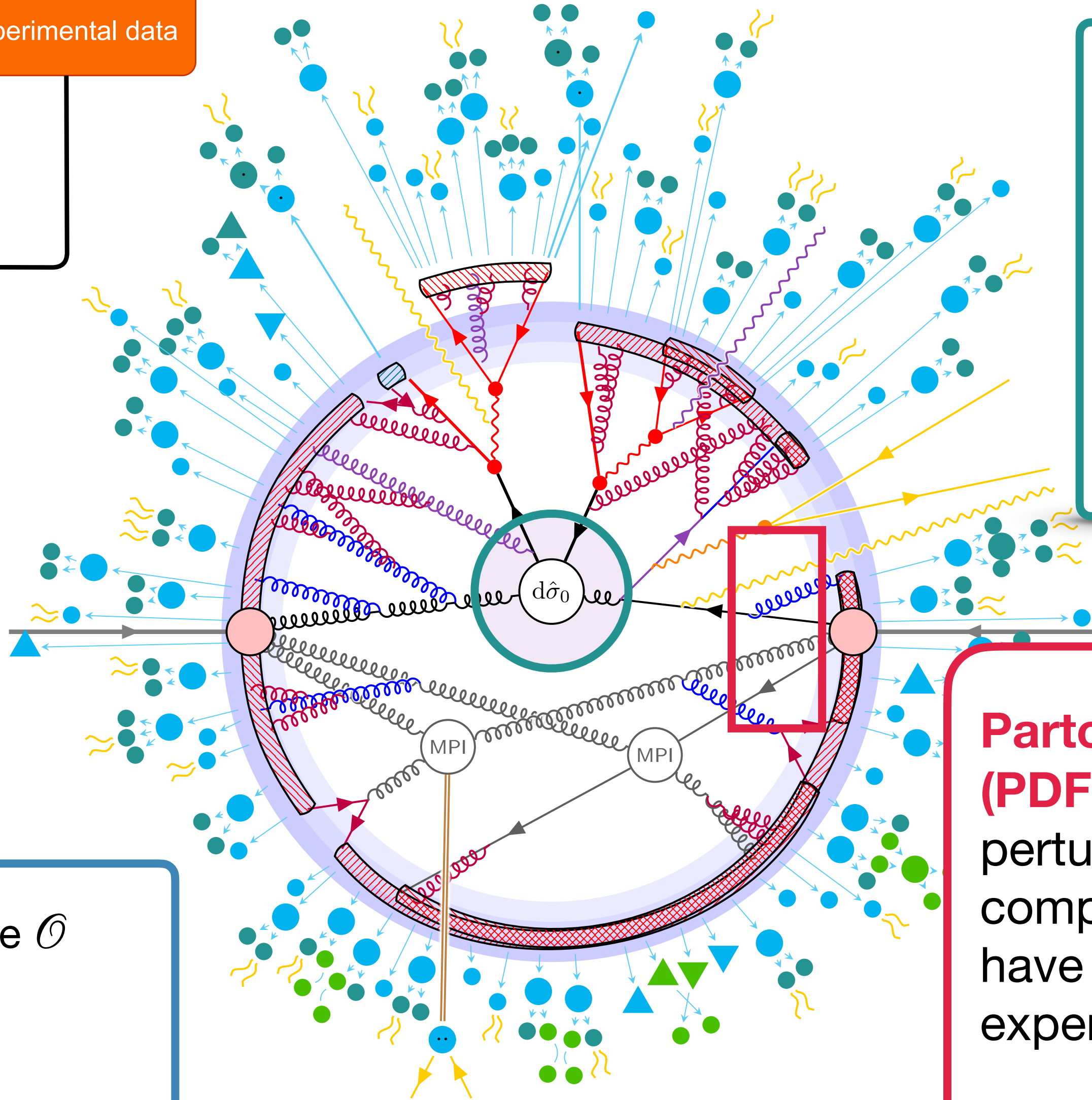
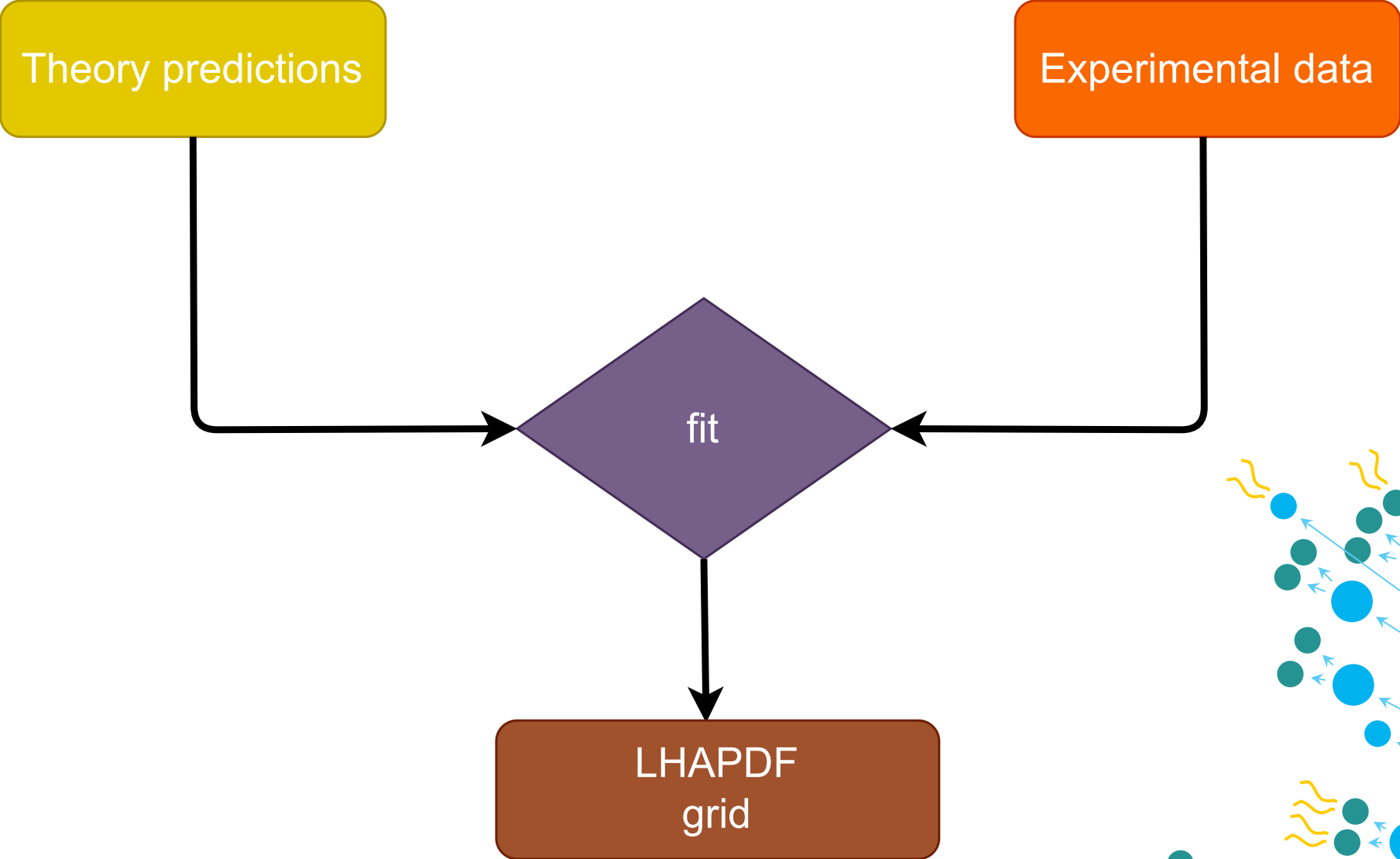
A tale of shrinking uncertainties

The NNPDF methodologies: neural networks
and ensemble regression

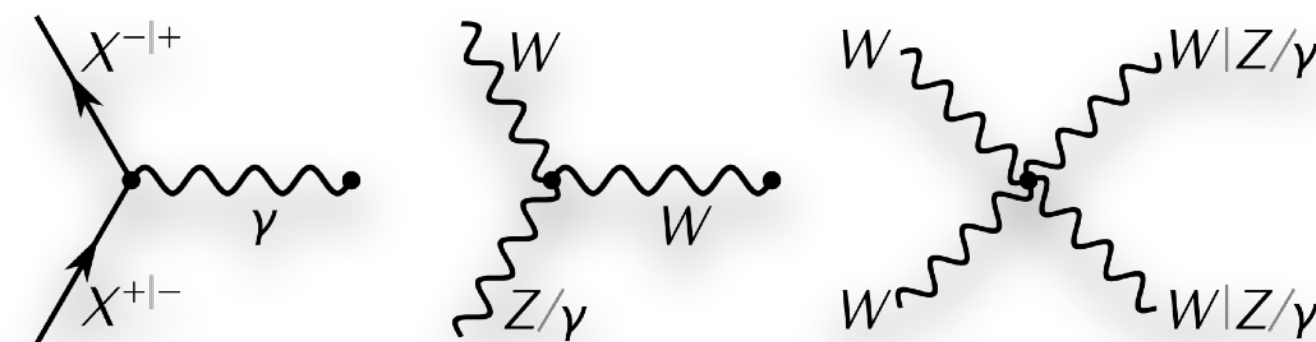
Results & Conclusions



Ingredients for collider predictions in High Energy Physics



Hard scattering $\hat{\sigma}_0$: encodes short-range interactions; computed from first principles.



Experiments measure an Observable $\mathcal{O}$ (cross-section, decay rates, etc.):

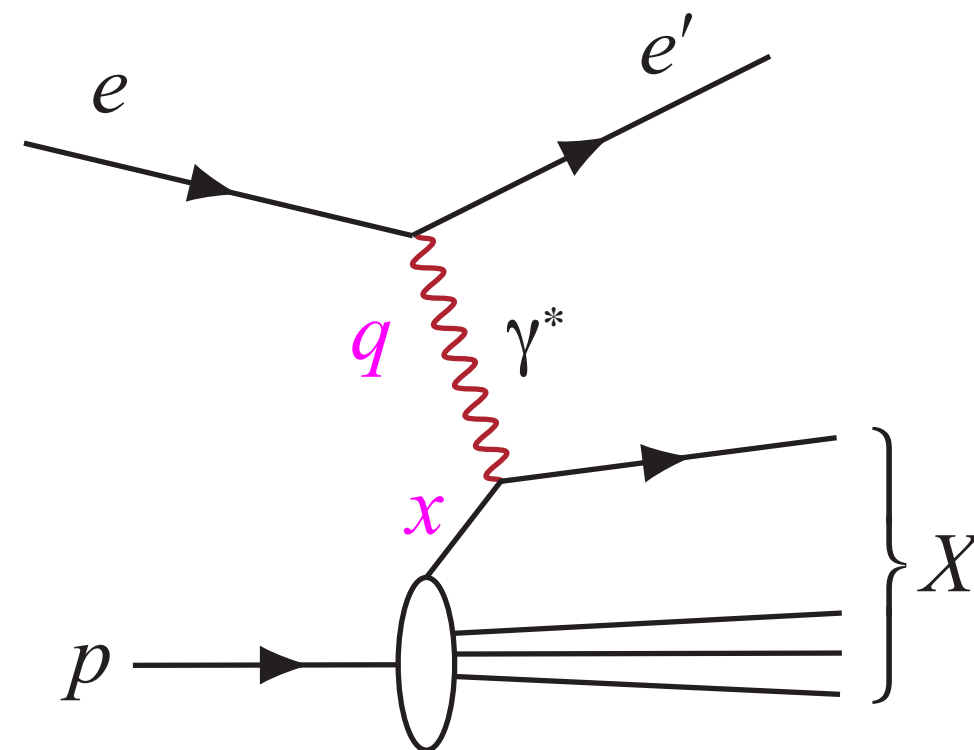
$$\mathcal{O} = \sum \hat{\sigma}_{ij} \otimes \text{PDF}_{ij}$$

Parton Distribution Functions (PDFs): encodes long-range non-perturbative interactions; cannot be computed from first principle and have to be determined from experimental Data.

Important: PDFs are Universal*

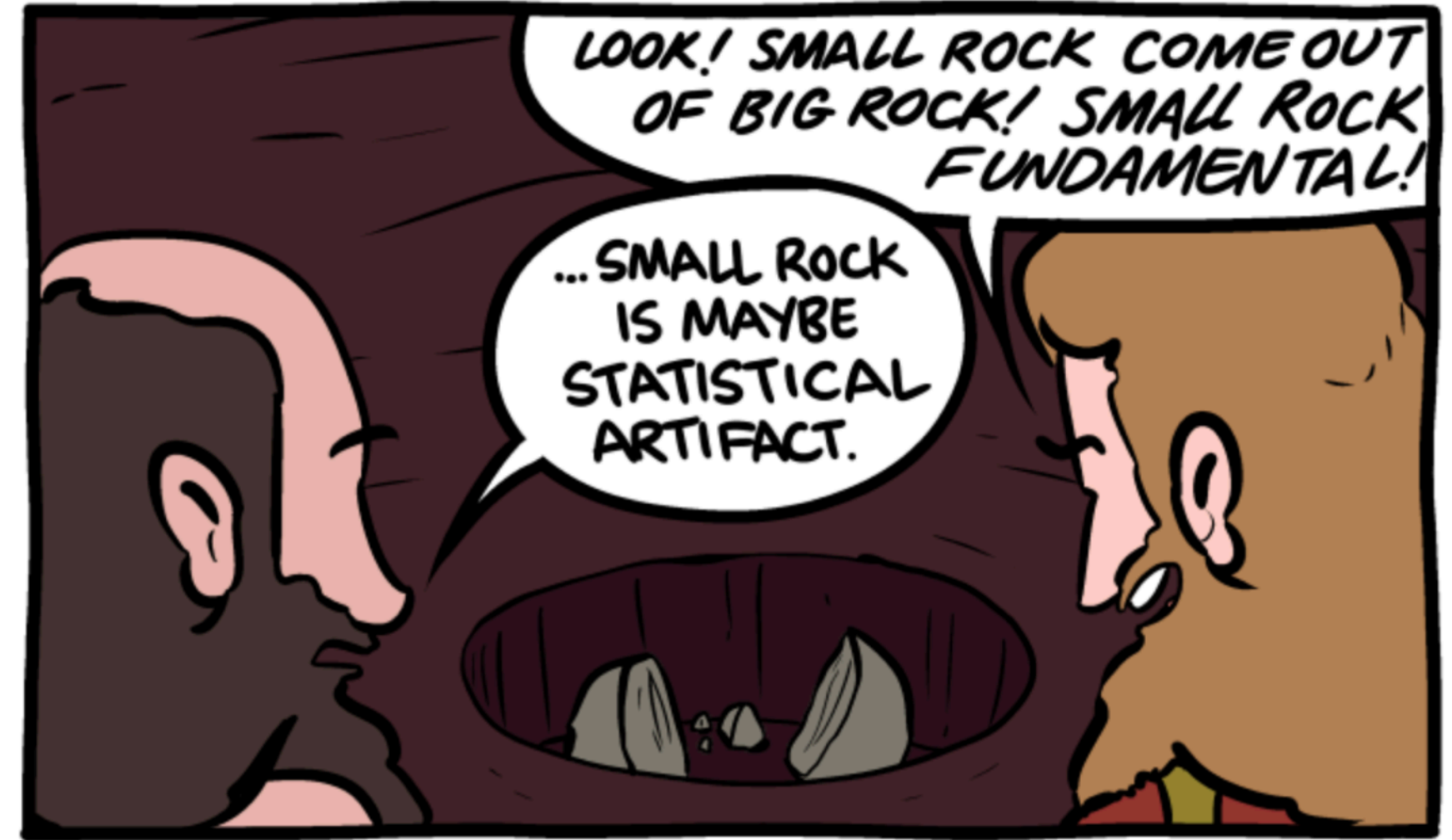
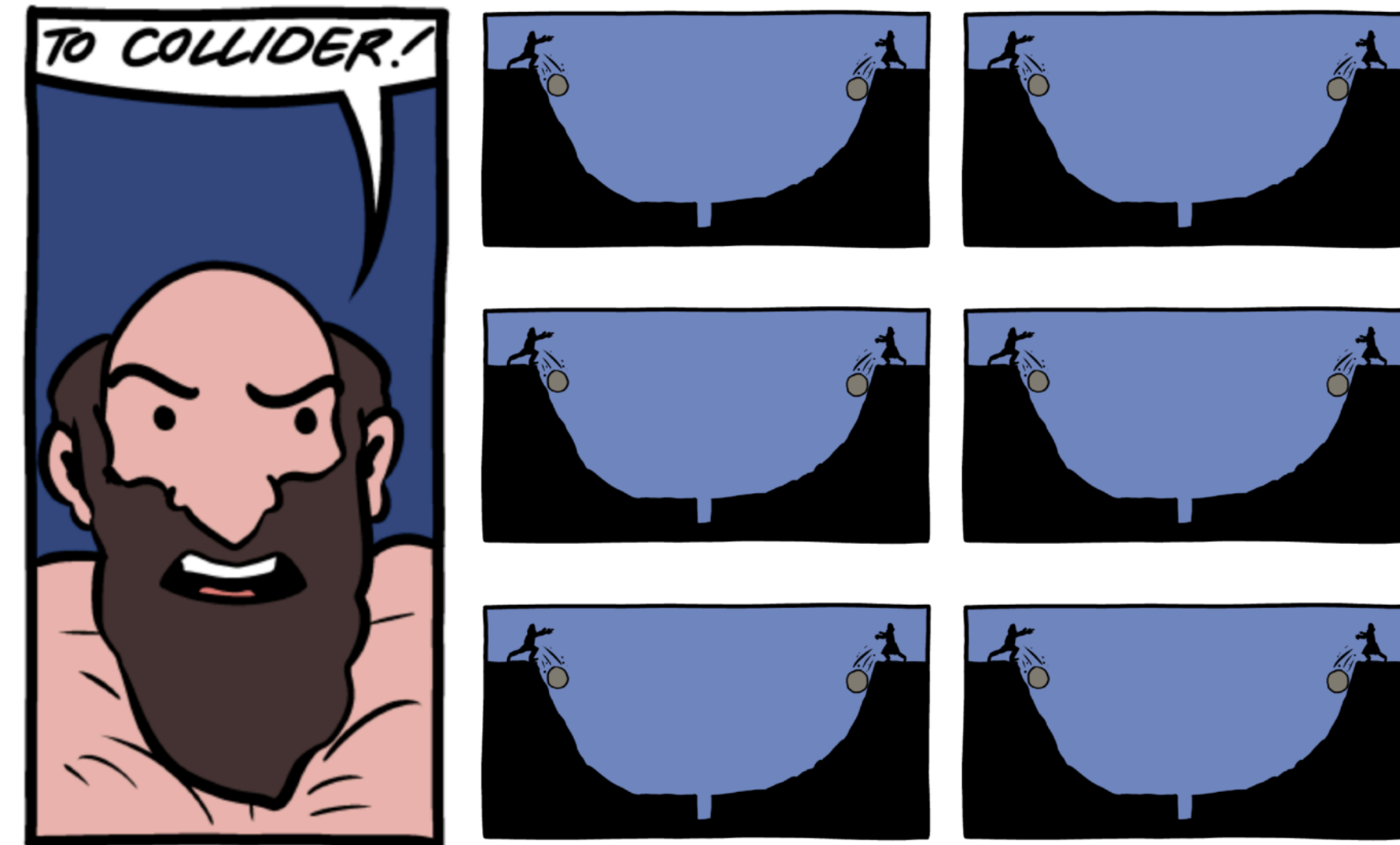
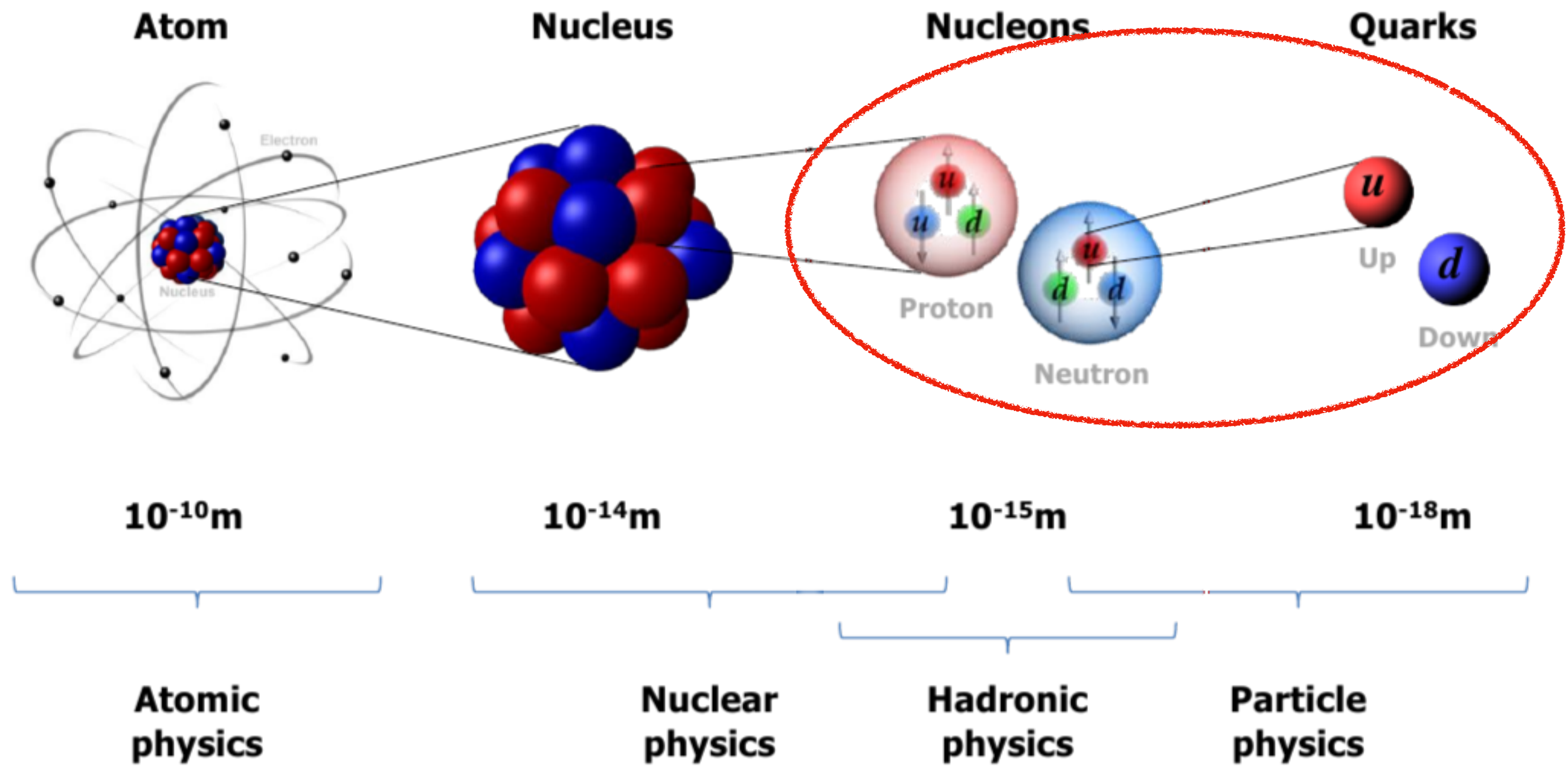
*terms & conditions apply

The internal structure of the proton



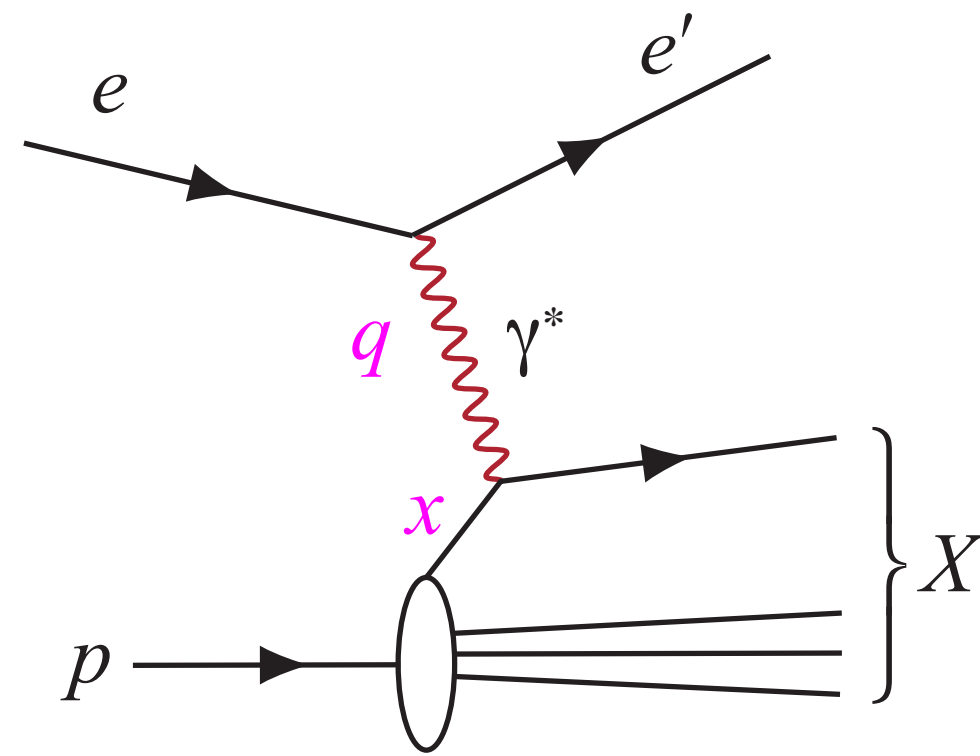
We cannot compute from first principles the internal dynamics of the proton.

With enough data in which protons are collided at high energies, we can get a very good picture of its interior.



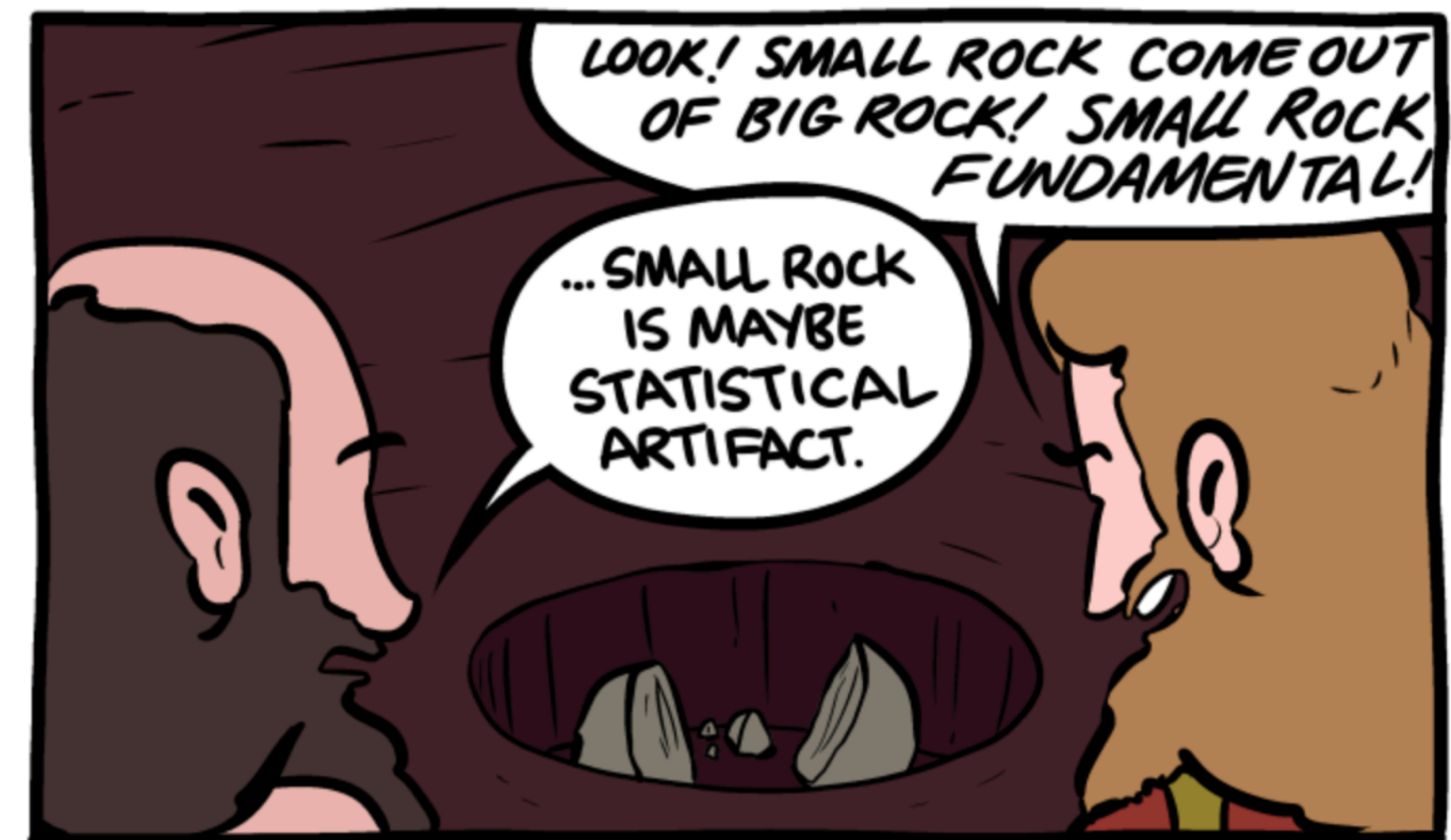
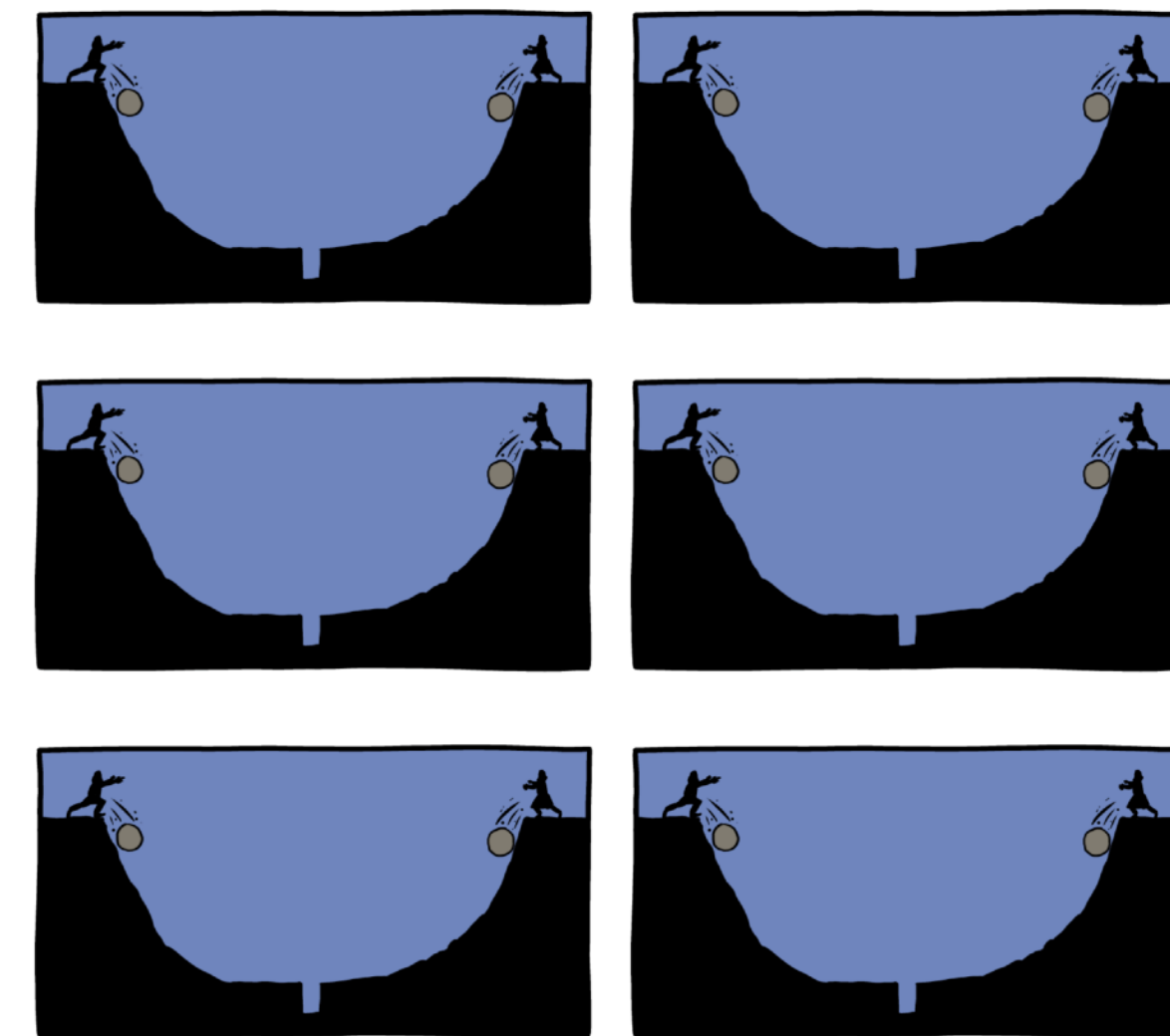
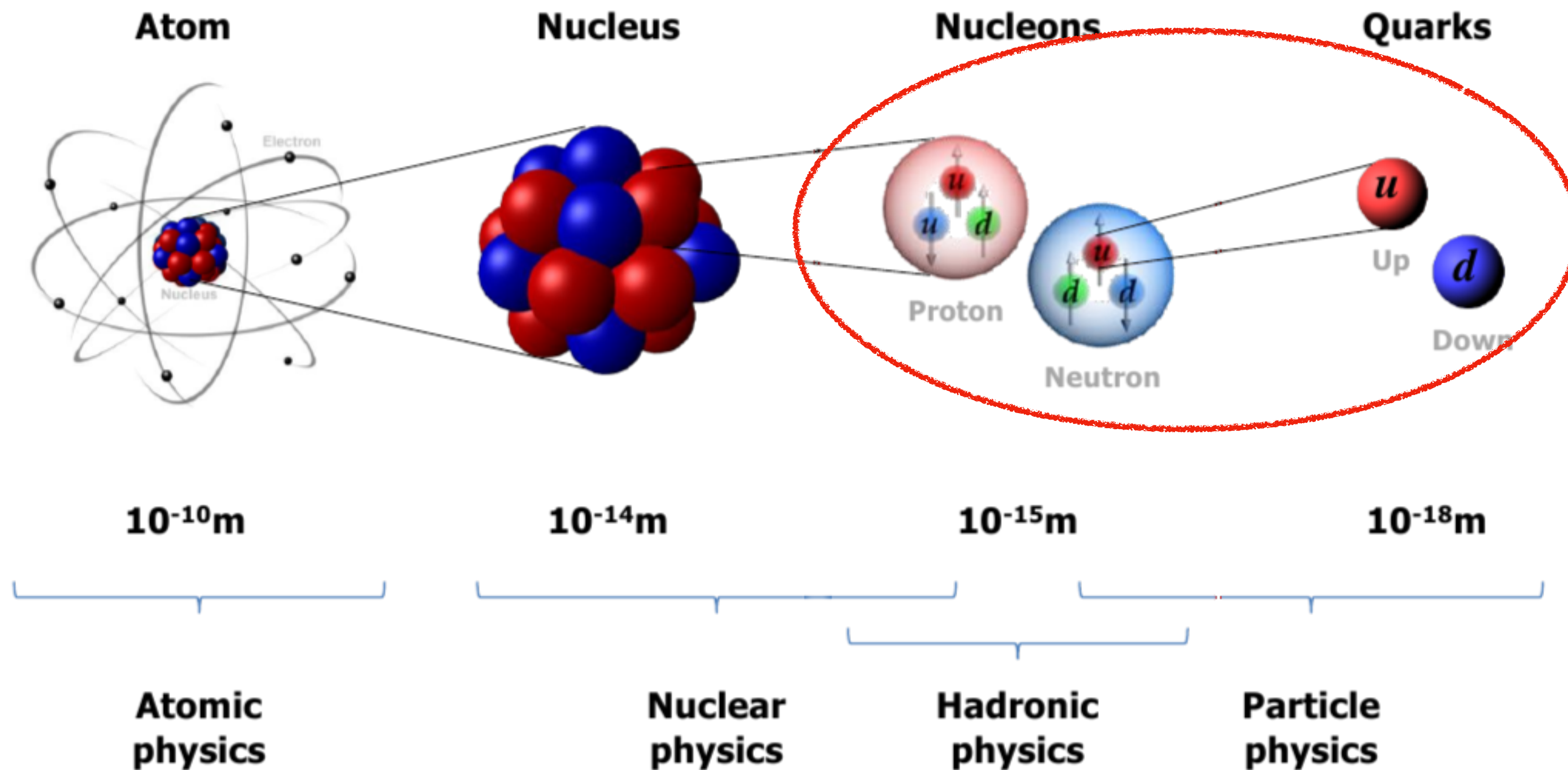
The internal structure of the proton

unpolarized, collinear



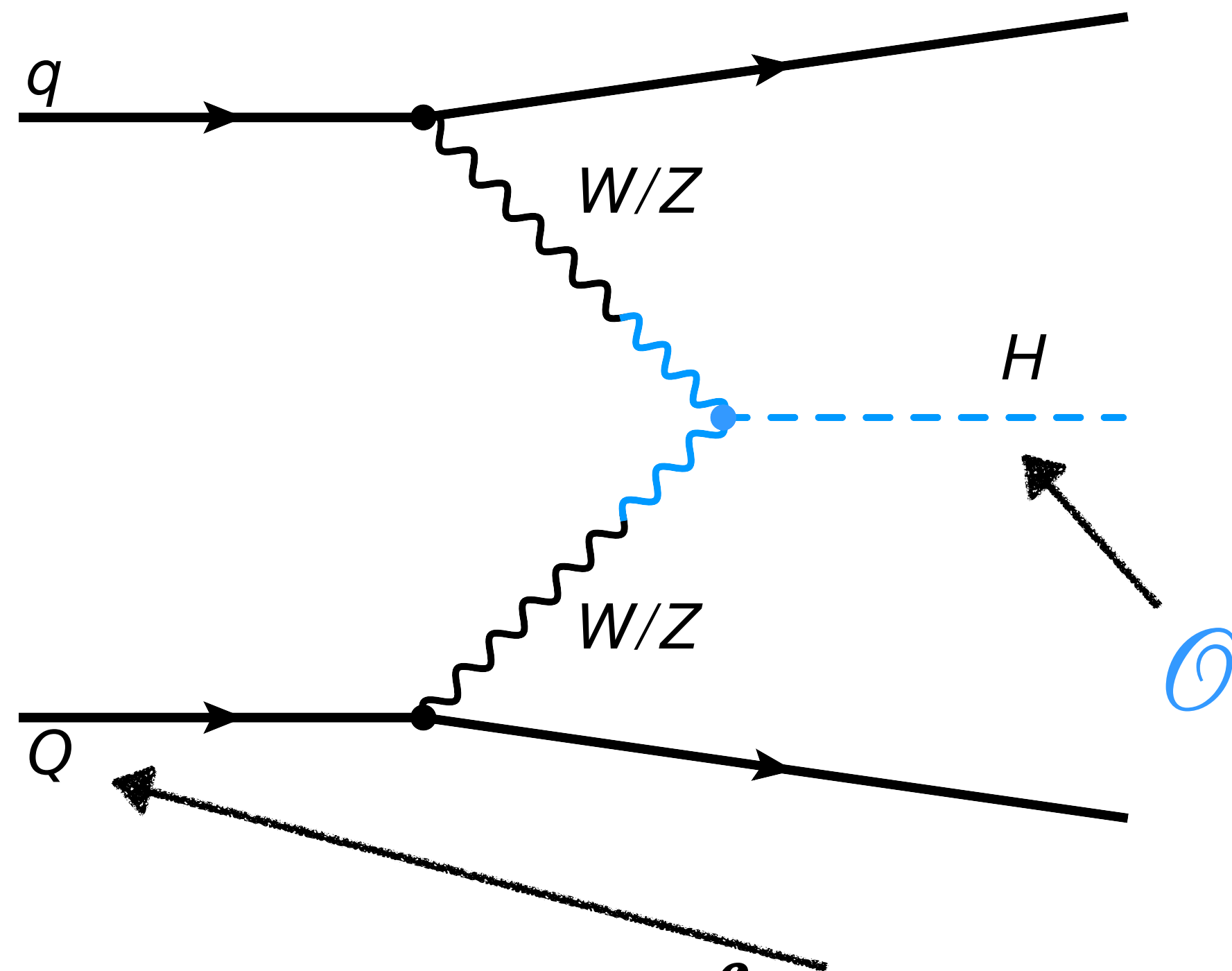
comparing to what we can do with today's vast array of data

We cannot *practically* compute from first principles the internal dynamics of the proton. With enough data in which protons are collided at high energies, we can get a very good picture of its interior.



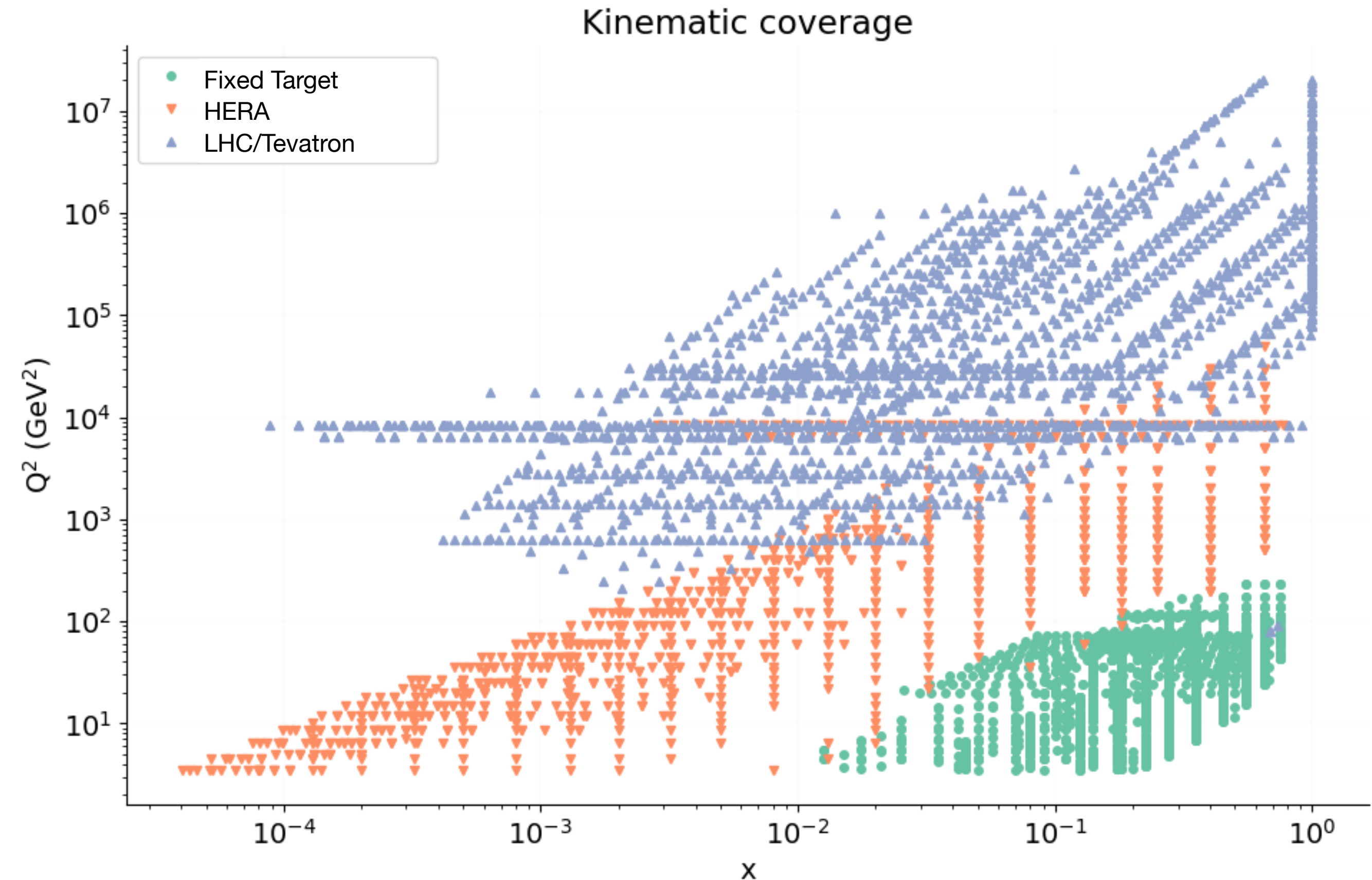
PDF determination ingredients

Theory + Data + Methodology (+ thoughts and prayers)



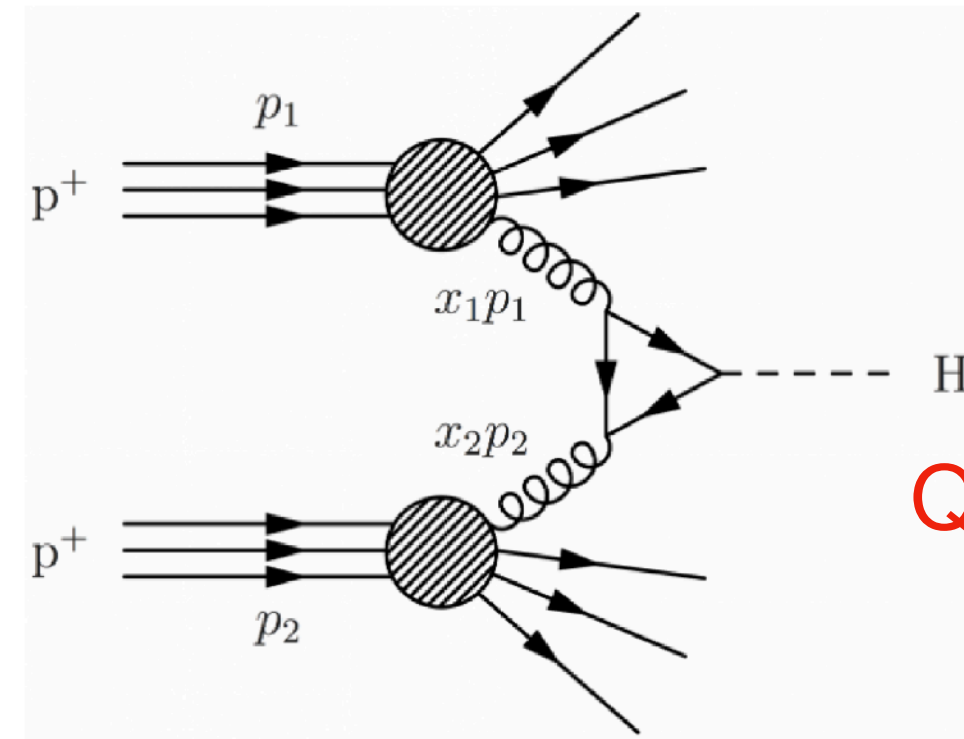
$$\mathcal{O} = \sum_{ij} \int dx_1 dx_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \hat{\sigma}_{ij}(x_1, x_2, \mu_R, \mu_F)$$

Input: quark type and fraction of the energy participating in the collision



PDF determination: theory

Resummation?



Fixed Order accuracy

α_s

EW?

$$\mathcal{O} = \int d\Phi_n dx_1 dx_2 f_1(x_1, \mu_F^2) f_2(x_2, \mu_F^2) |M(\{p_n\}, \mu_R)|^2 \mathcal{J}_m^n(\{p_n\})$$

Higher twist effects

Parton Distribution Functions

QED?

Interpolation Grids

And many others: heavy mass flavour scheme, parametrisation, fitting scale, treatment of missing higher order uncertainties, electroweak parameters, quark masses, evolution settings...

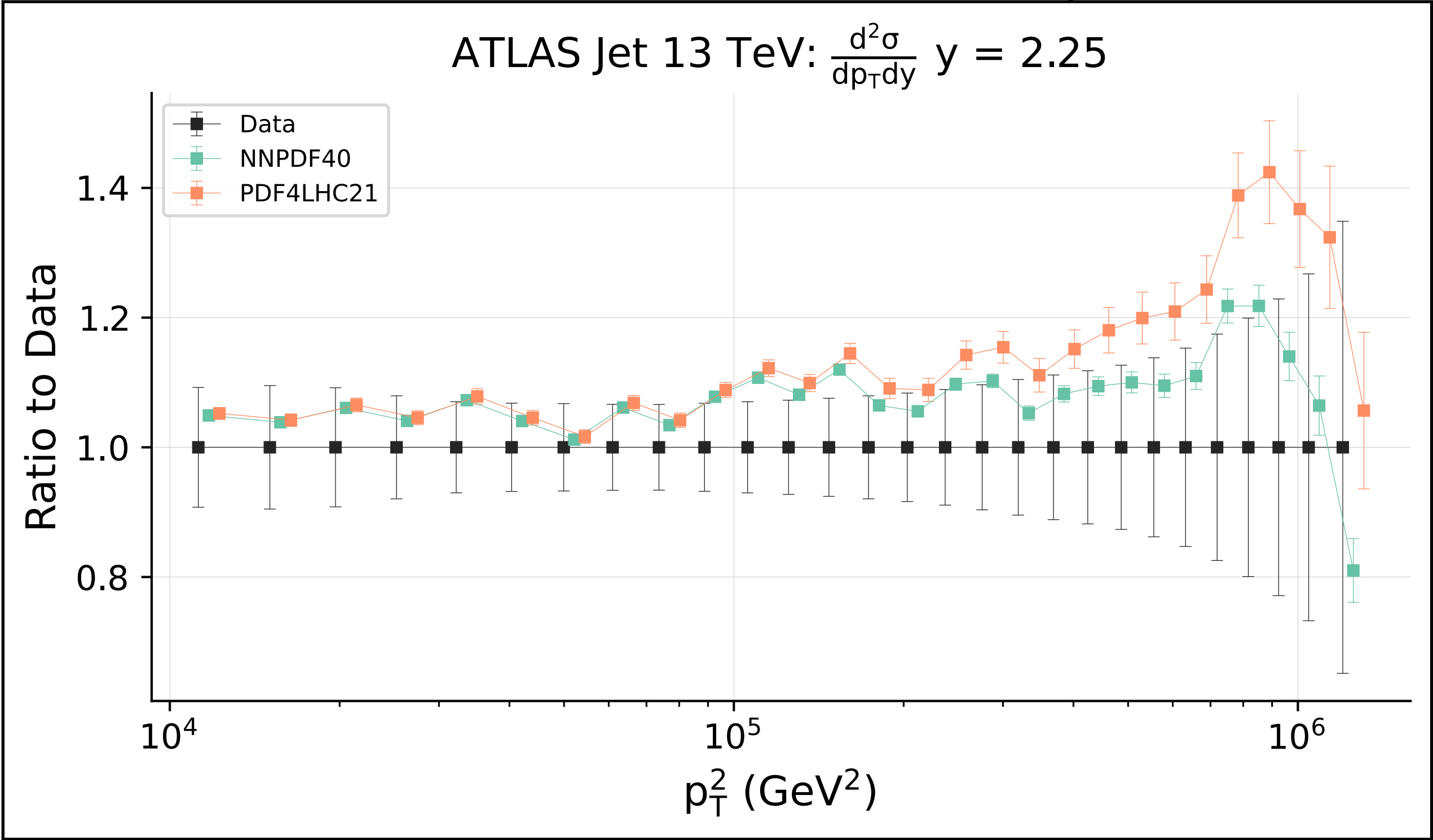
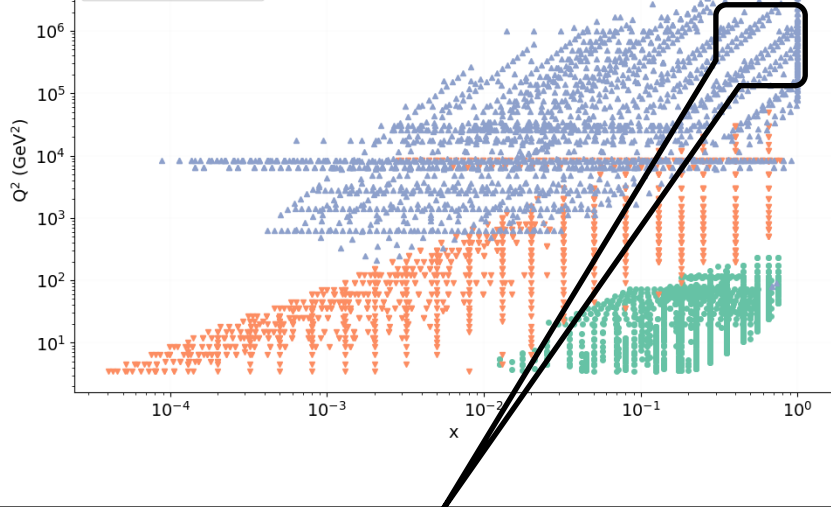
PDF determination: data



Screenshots from Hepdata: the cause of, and solution to, all of PDF fitters' problems!

RE	p p -> jet jet
SQRTS(S)	13000 GeV
$p_{T,jet}$ [GeV]	$d\sigma/dp_{T,jet}d Y $
100 - 116	6.646300e+03 +6.271000e+01 -6.271000e+01 stat +0.000000e+00 -0.000000e+00 syst_JER_CROSS_CALIB_FORWARD +0.000000e+00 -0.000000e+00 syst_JER_NOISE_FORWARD + 334 more errors <a>Show all

-1.127000e-01	
-1.068000e-07	syst_JES_EtaIntercalibration_Stat114
-0.000000e+00	syst_JES_EtaIntercalibration_Stat115
-0.000000e+00	syst_JES_EtaIntercalibration_Stat116
-1.283800e-02	syst_JES_EtaIntercalibration_Stat117
-0.000000e+00	syst_JES_EtaIntercalibration_Stat118
-3.488000e-01	syst_JES_EtaIntercalibration_Stat119
-0.000000e+00	syst_JES_EtaIntercalibration_Stat12
-0.000000e+00	syst_JES_EtaIntercalibration_Stat120
-0.000000e+00	syst_JES_EtaIntercalibration_Stat121
+1.277500e-02	syst_JES_EtaIntercalibration_Stat122
-2.836000e-02	syst_JES_EtaIntercalibration_Stat123
-8.124000e-01	syst_JES_EtaIntercalibration_Stat124
+3.476000e+00	syst_JES_EtaIntercalibration_Stat125
-4.962000e+00	syst_JES_EtaIntercalibration_Stat126
+3.316000e-02	syst_JES_EtaIntercalibration_Stat127
-4.155000e+00	syst_JES_EtaIntercalibration_Stat128
-6.888000e-02	syst_JES_EtaIntercalibration_Stat129
+2.591000e-02	syst_JES_EtaIntercalibration_Stat13
-0.000000e+00	syst_JES_EtaIntercalibration_Stat130
-0.000000e+00	syst_JES_EtaIntercalibration_Stat131
-2.714000e-02	syst_JES_EtaIntercalibration_Stat132
+8.704000e-02	syst_JES_EtaIntercalibration_Stat133
-4.705000e+00	syst_JES_EtaIntercalibration_Stat134
+4.264000e+00	syst_JES_EtaIntercalibration_Stat135
-0.000000e+00	syst_JES_EtaIntercalibration_Stat136
-0.000000e+00	syst_JES_EtaIntercalibration_Stat137
-1.294100e+01	syst_JES_EtaIntercalibration_Stat138
+1.112100e+01	syst_JES_EtaIntercalibration_Stat139
-4.251000e+00	syst_JES_EtaIntercalibration_Stat14
+3.258000e+00	syst_JES_EtaIntercalibration_Stat140
-6.252000e-01	syst_JES_EtaIntercalibration_Stat141
+2.819000e-01	syst_JES_EtaIntercalibration_Stat142
+2.461000e-04	syst_JES_EtaIntercalibration_Stat143
+3.602000e-07	syst_JES_EtaIntercalibration_Stat144
-0.000000e+00	syst_JES_EtaIntercalibration_Stat145
-0.000000e+00	syst_JES_EtaIntercalibration_Stat146
-0.000000e+00	syst_JES_EtaIntercalibration_Stat147
-1.283800e-02	syst_JES_EtaIntercalibration_Stat148
-0.000000e+00	syst_JES_EtaIntercalibration_Stat149
-1.283800e-02	syst_JES_EtaIntercalibration_Stat150
-3.488000e-01	syst_JES_EtaIntercalibration_Stat151
-0.000000e+00	syst_JES_EtaIntercalibration_Stat152
-3.488000e-01	syst_JES_EtaIntercalibration_Stat153
-0.000000e+00	syst_JES_EtaIntercalibration_Stat154
-3.488000e-01	syst_JES_EtaIntercalibration_Stat155
-0.000000e+00	syst_JES_EtaIntercalibration_Stat156
-0.000000e+00	syst_JES_EtaIntercalibration_Stat157
-0.000000e+00	syst_JES_EtaIntercalibration_Stat158
-0.000000e+00	syst_JES_EtaIntercalibration_Stat159
-0.000000e+00	syst_JES_EtaIntercalibration_Stat160
-0.000000e+00	syst_JES_EtaIntercalibration_Stat161
-0.000000e+00	syst_JES_EtaIntercalibration_Stat162
-0.000000e+00	syst_JES_EtaIntercalibration_Stat163
-0.000000e+00	syst_JES_EtaIntercalibration_Stat164
-0.000000e+00	syst_JES_EtaIntercalibration_Stat165
-0.000000e+00	syst_JES_EtaIntercalibration_Stat166
-0.000000e+00	syst_JES_EtaIntercalibration_Stat167
-0.000000e+00	syst_JES_EtaIntercalibration_Stat168
-0.000000e+00	syst_JES_EtaIntercalibration_Stat169
-0.000000e+00	syst_JES_EtaIntercalibration_Stat170
-0.000000e+00	syst_JES_EtaIntercalibration_Stat171
-0.000000e+00	syst_JES_EtaIntercalibration_Stat172
-0.000000e+00	syst_JES_EtaIntercalibration_Stat173
-0.000000e+00	syst_JES_EtaIntercalibration_Stat174
-0.000000e+00	syst_JES_EtaIntercalibration_Stat175
-0.000000e+00	syst_JES_EtaIntercalibration_Stat176
-0.000000e+00	syst_JES_EtaIntercalibration_Stat177
-0.000000e+00	syst_JES_EtaIntercalibration_Stat178
-0.000000e+00	syst_JES_EtaIntercalibration_Stat179
-0.000000e+00	syst_JES_EtaIntercalibration_Stat180
-0.000000e+00	syst_JES_EtaIntercalibration_Stat181
-0.000000e+00	syst_JES_EtaIntercalibration_Stat182
-0.000000e+00	syst_JES_EtaIntercalibration_Stat183
-0.000000e+00	syst_JES_EtaIntercalibration_Stat184
-0.000000e+00	syst_JES_EtaIntercalibration_Stat185
-0.000000e+00	syst_JES_EtaIntercalibration_Stat186
-0.000000e+00	syst_JES_EtaIntercalibration_Stat187
-0.000000e+00	syst_JES_EtaIntercalibration_Stat188
-0.000000e+00	syst_JES_EtaIntercalibration_Stat189
-0.000000e+00	syst_JES_EtaIntercalibration_Stat190
-0.000000e+00	syst_JES_EtaIntercalibration_Stat191
-0.000000e+00	syst_JES_EtaIntercalibration_Stat192
-0.000000e+00	syst_JES_EtaIntercalibration_Stat193
-0.000000e+00	syst_JES_EtaIntercalibration_Stat194
-0.000000e+00	syst_JES_EtaIntercalibration_Stat195
-0.000000e+00	syst_JES_EtaIntercalibration_Stat196
-0.000000e+00	syst_JES_EtaIntercalibration_Stat197
-0.000000e+00	syst_JES_EtaIntercalibration_Stat198
-0.000000e+00	syst_JES_EtaIntercalibration_Stat199
-0.000000e+00	syst_JES_EtaIntercalibration_Stat200



This table is too large to load automatically.

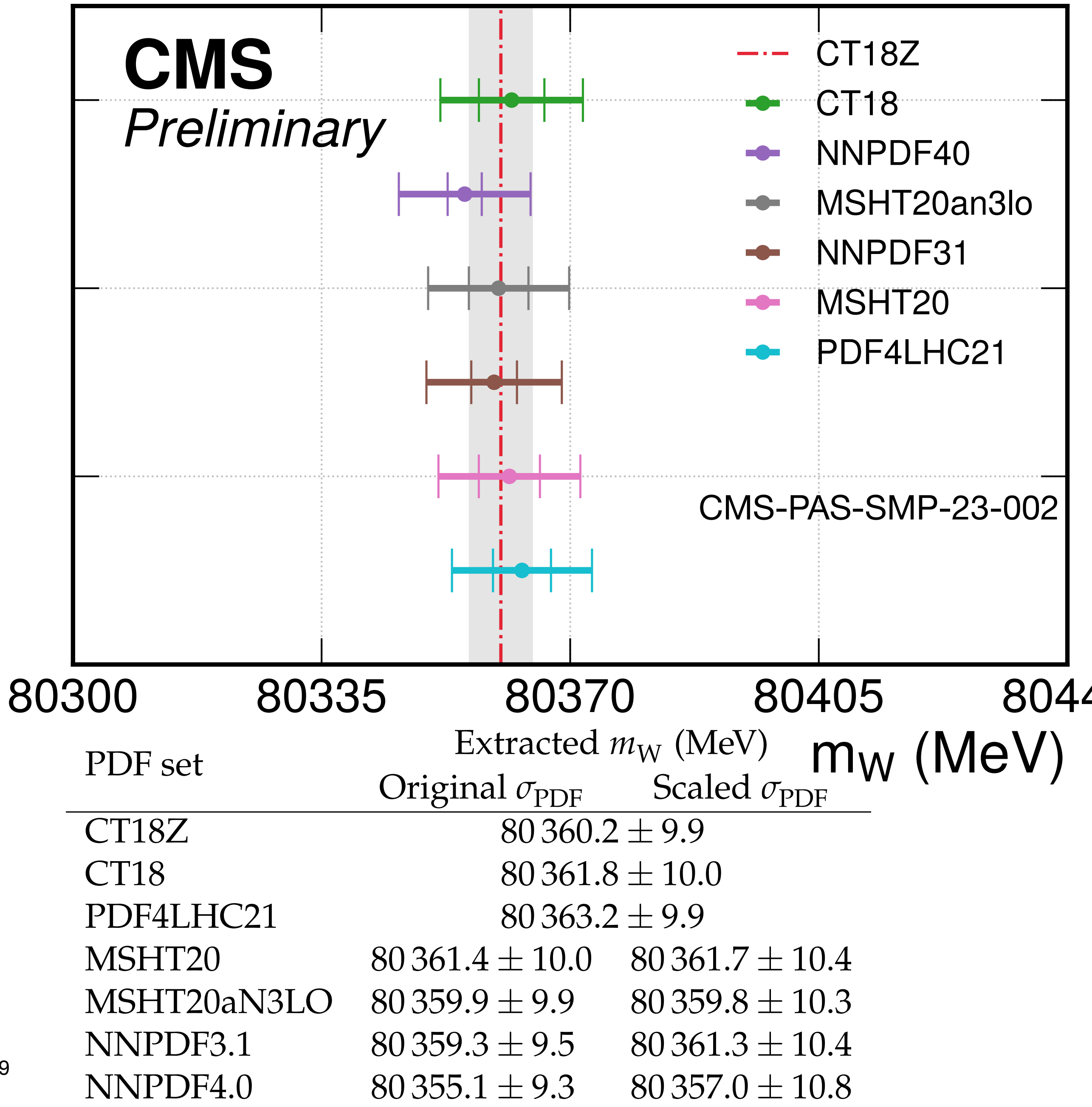
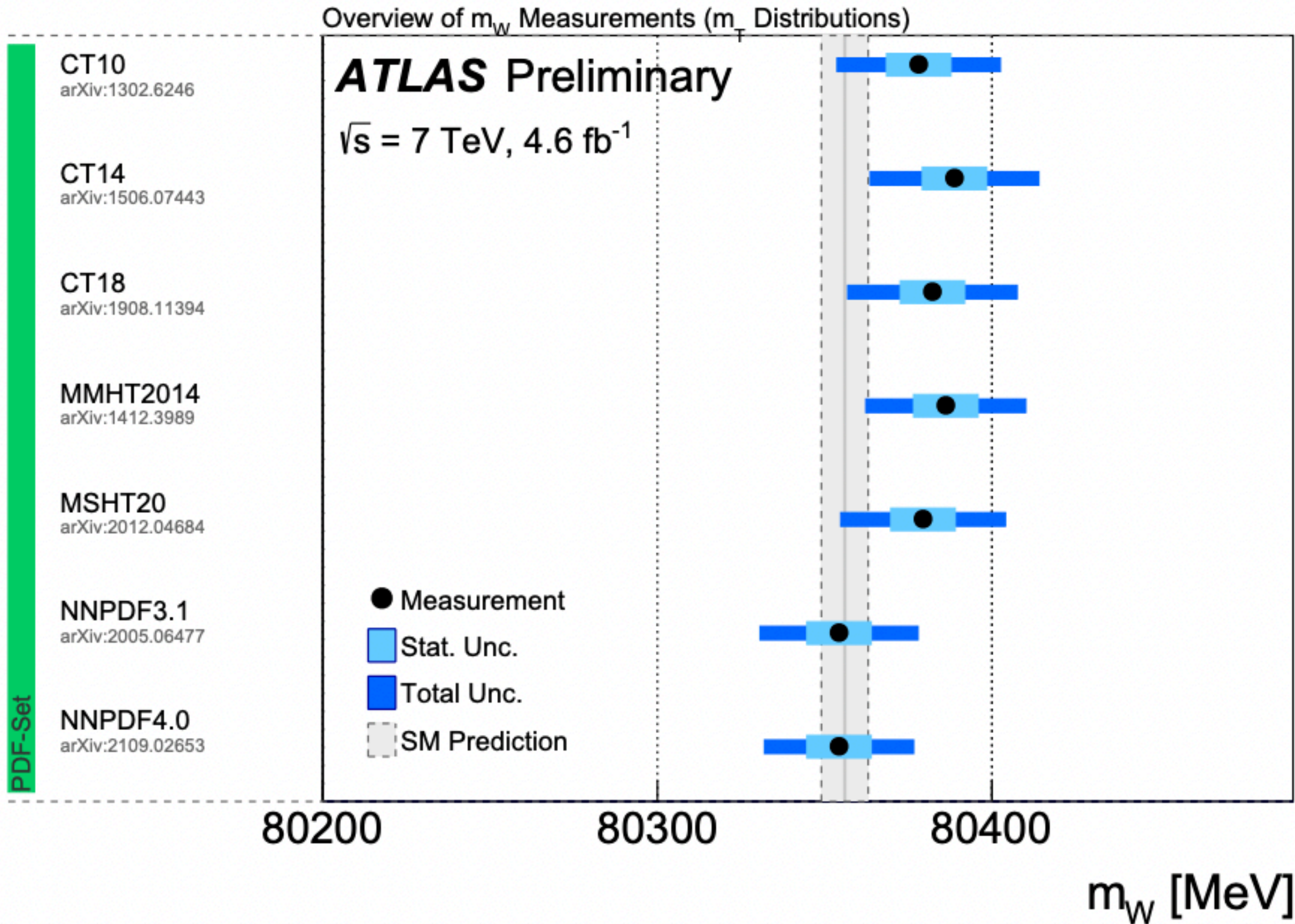
The table size is 1.11 MB, which is greater than our threshold of 1.00 MB.

Load Table

...and these are just the experimental uncertainties!

What is the impact of the PDF determination?

Despite the convergence seen on data, theory, and methodology, the choice of PDF can still have a large phenomenological impact.

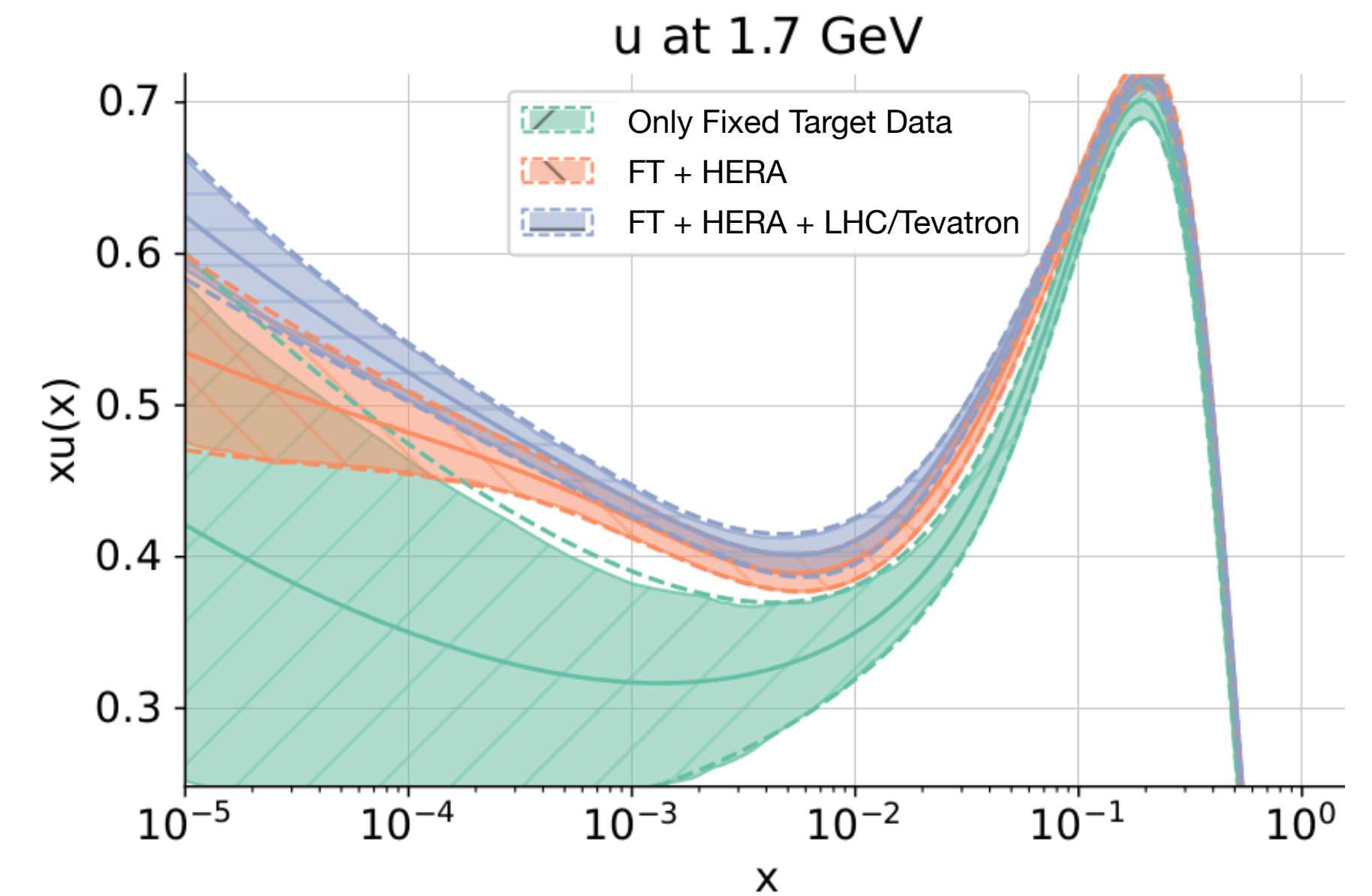


Introduction

A tale of shrinking uncertainties

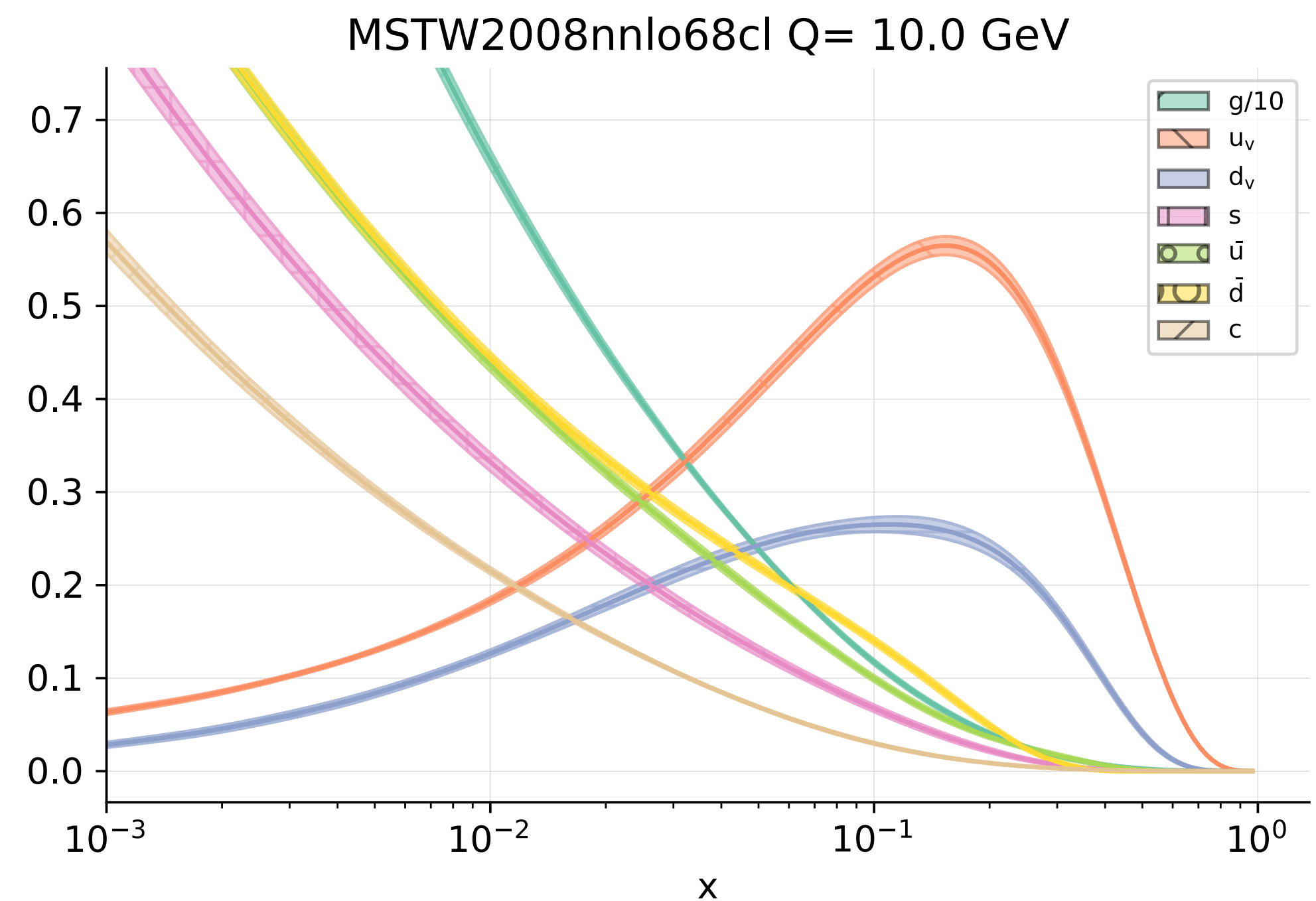
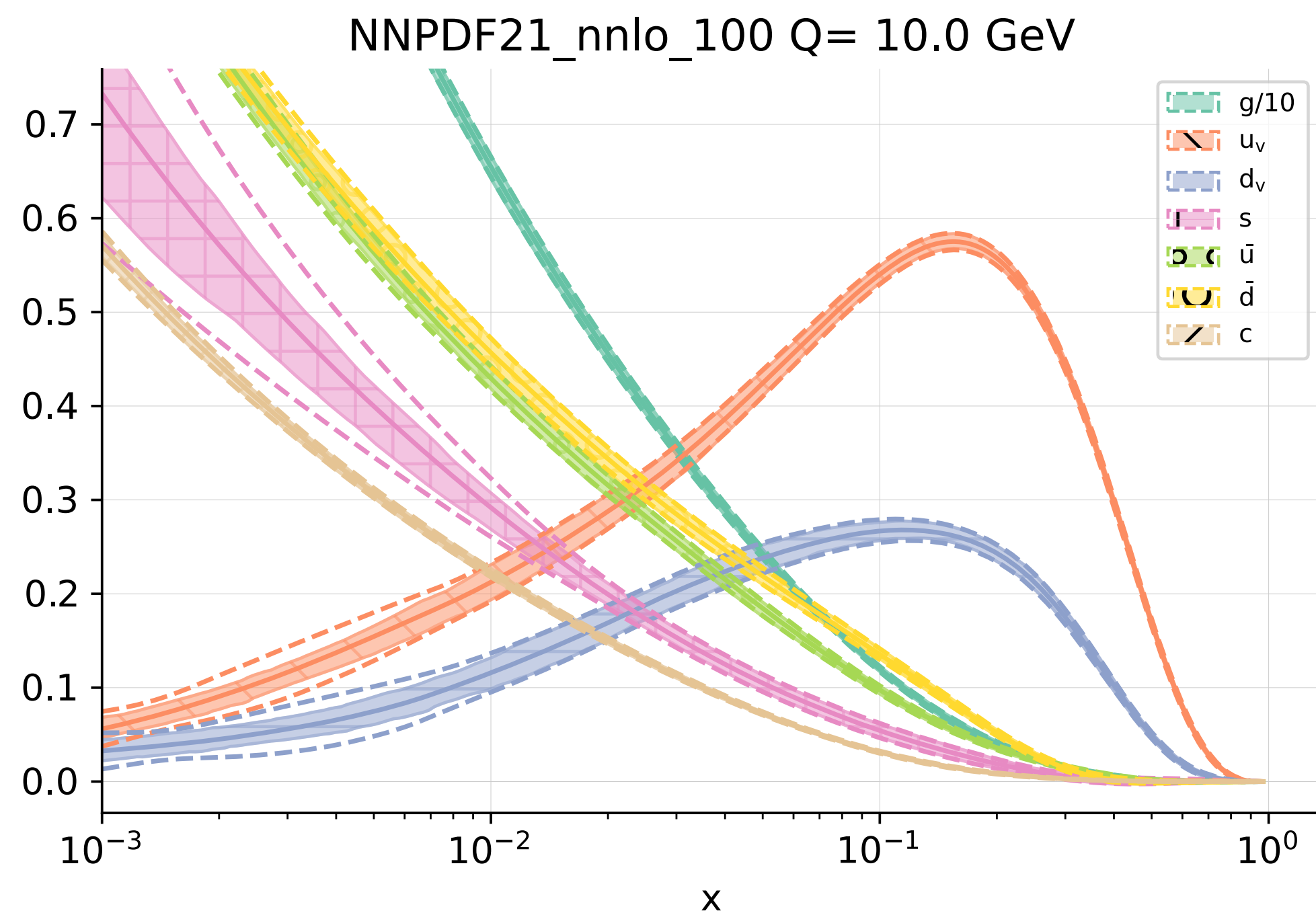
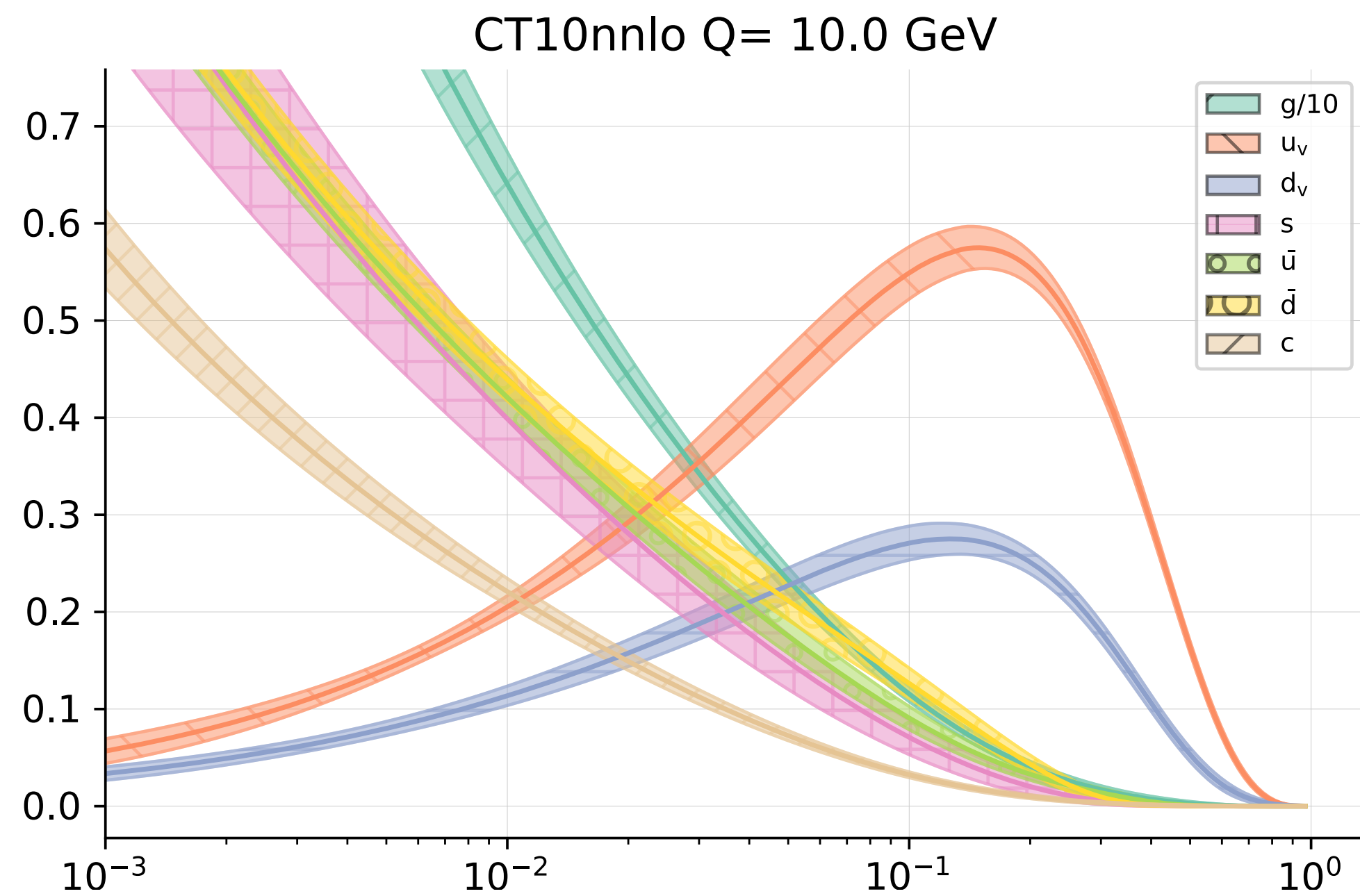
The NNPDF methodologies: neural networks
and ensemble regression

Results & Conclusions



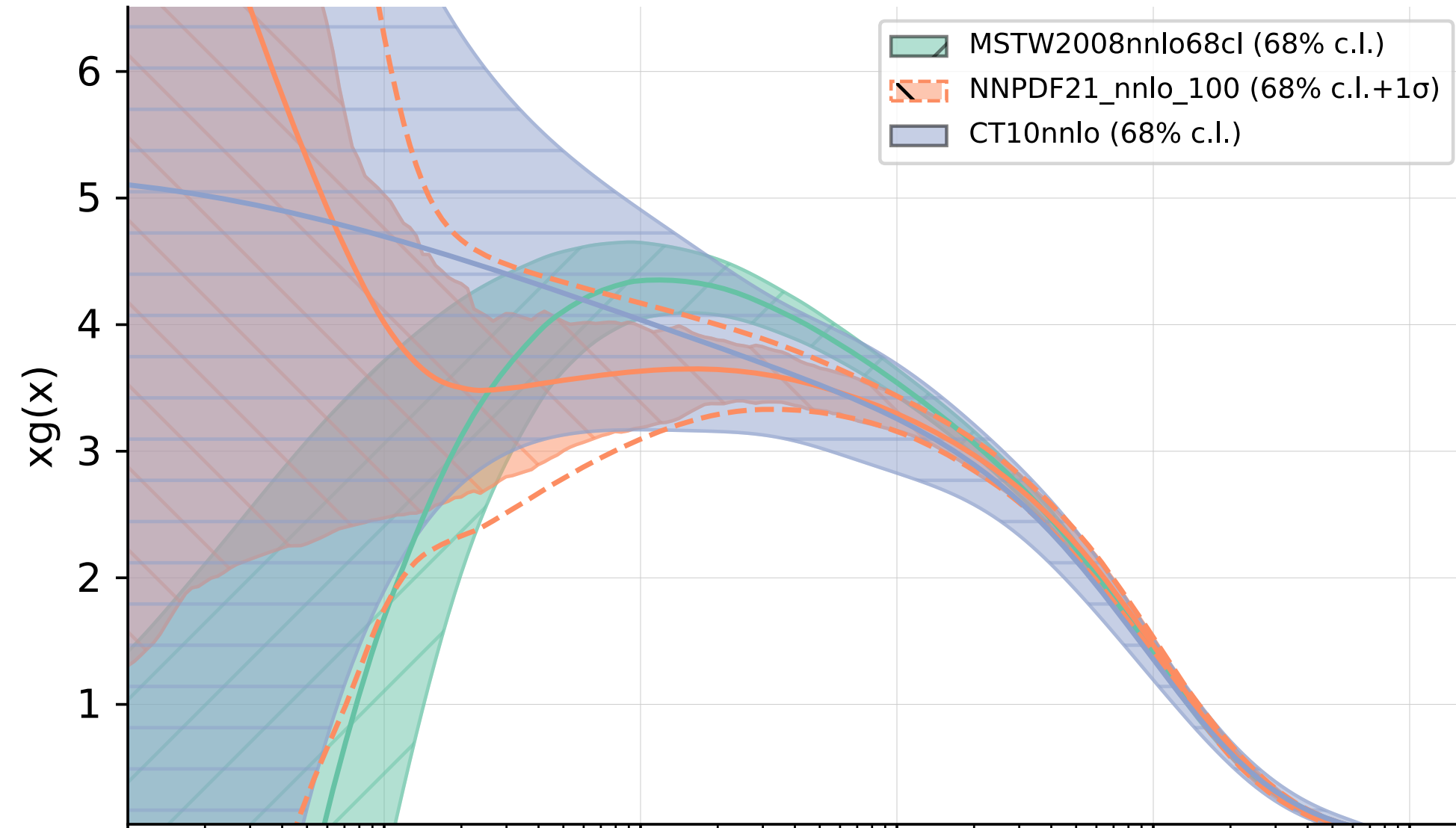
Parton Distribution Functions... 15 years ago

- LHC has just started: no data included in the determination yet
 - Nominally NNLO but that only included DIS data and evolution
 - PDF uncertainties were a relatively new feature!
 - Experimental uncertainties were quite big in certain regions which mean the data was not as constraining as it is today
- (15 years ago is not that long ago!)

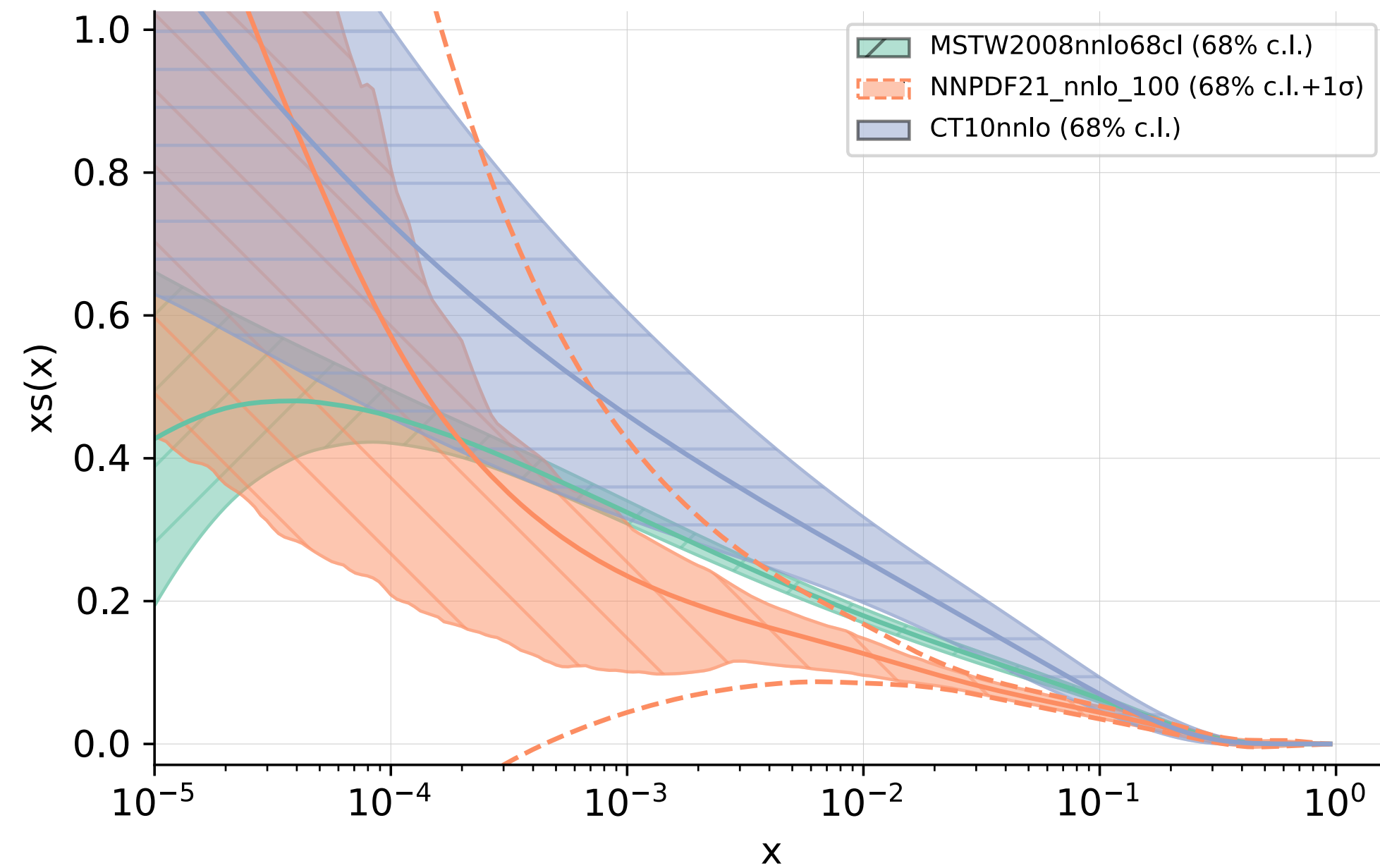


Parton Distribution Functions... 15 years ago

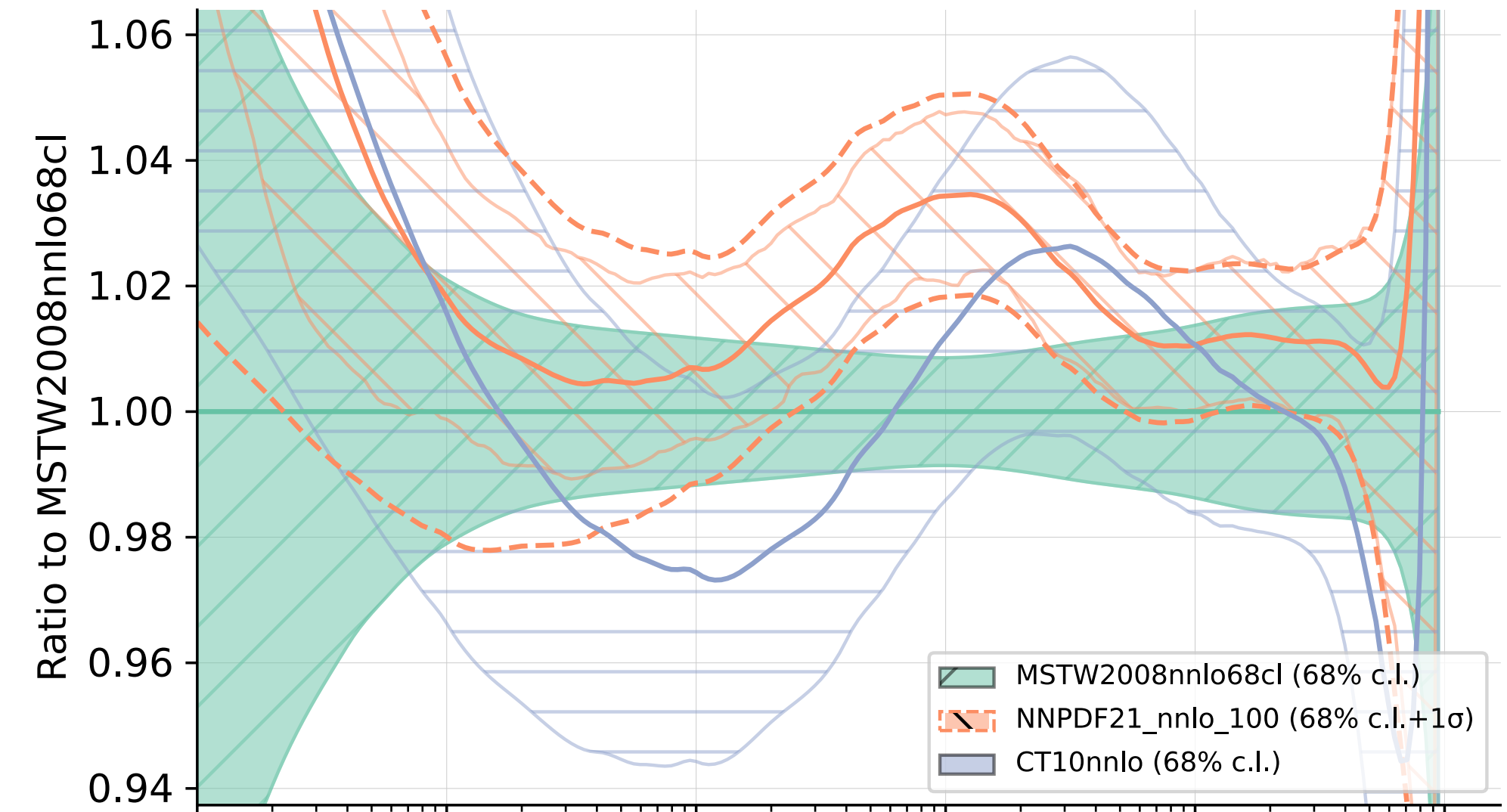
g at 2.0 GeV



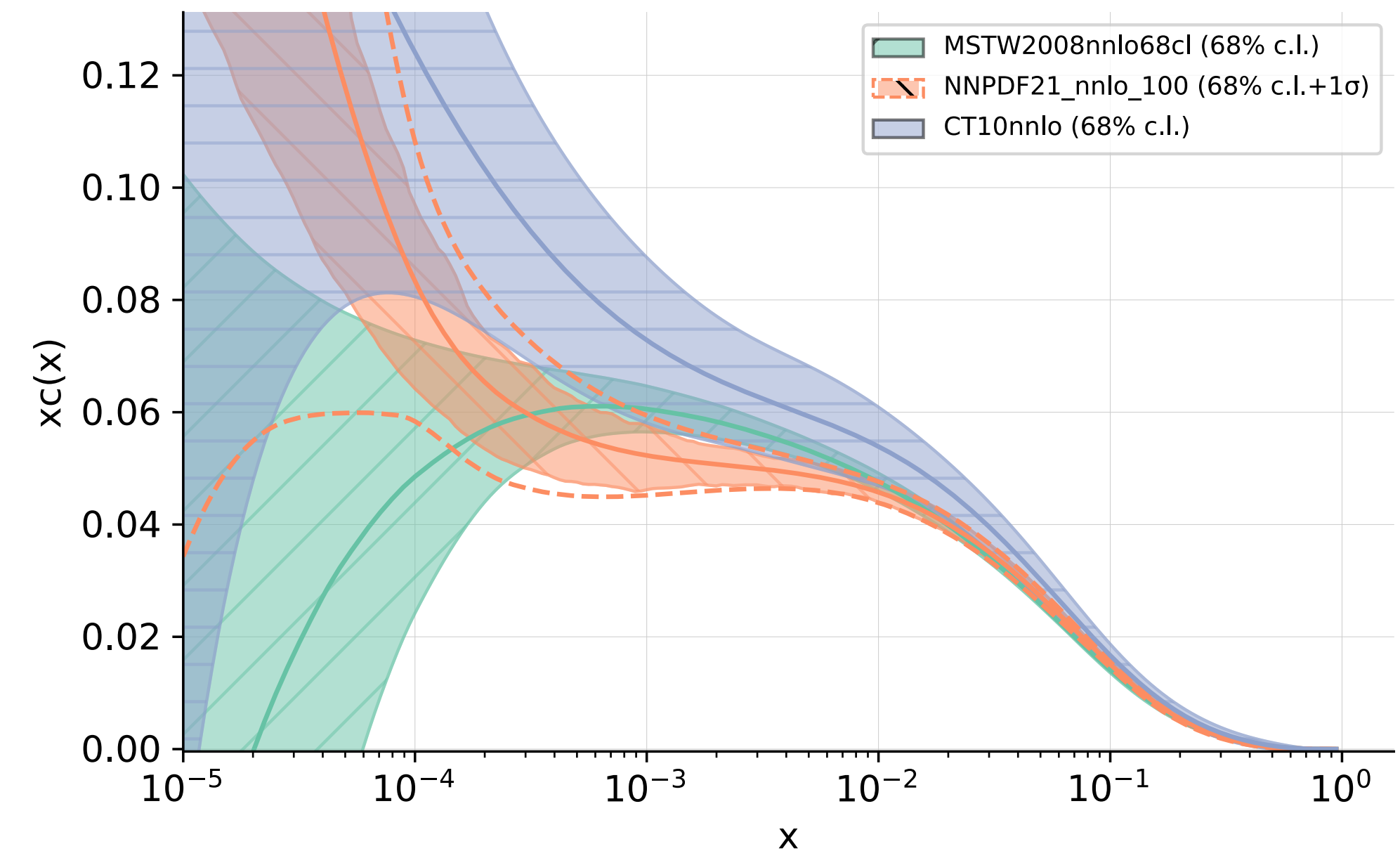
s at 2.0 GeV



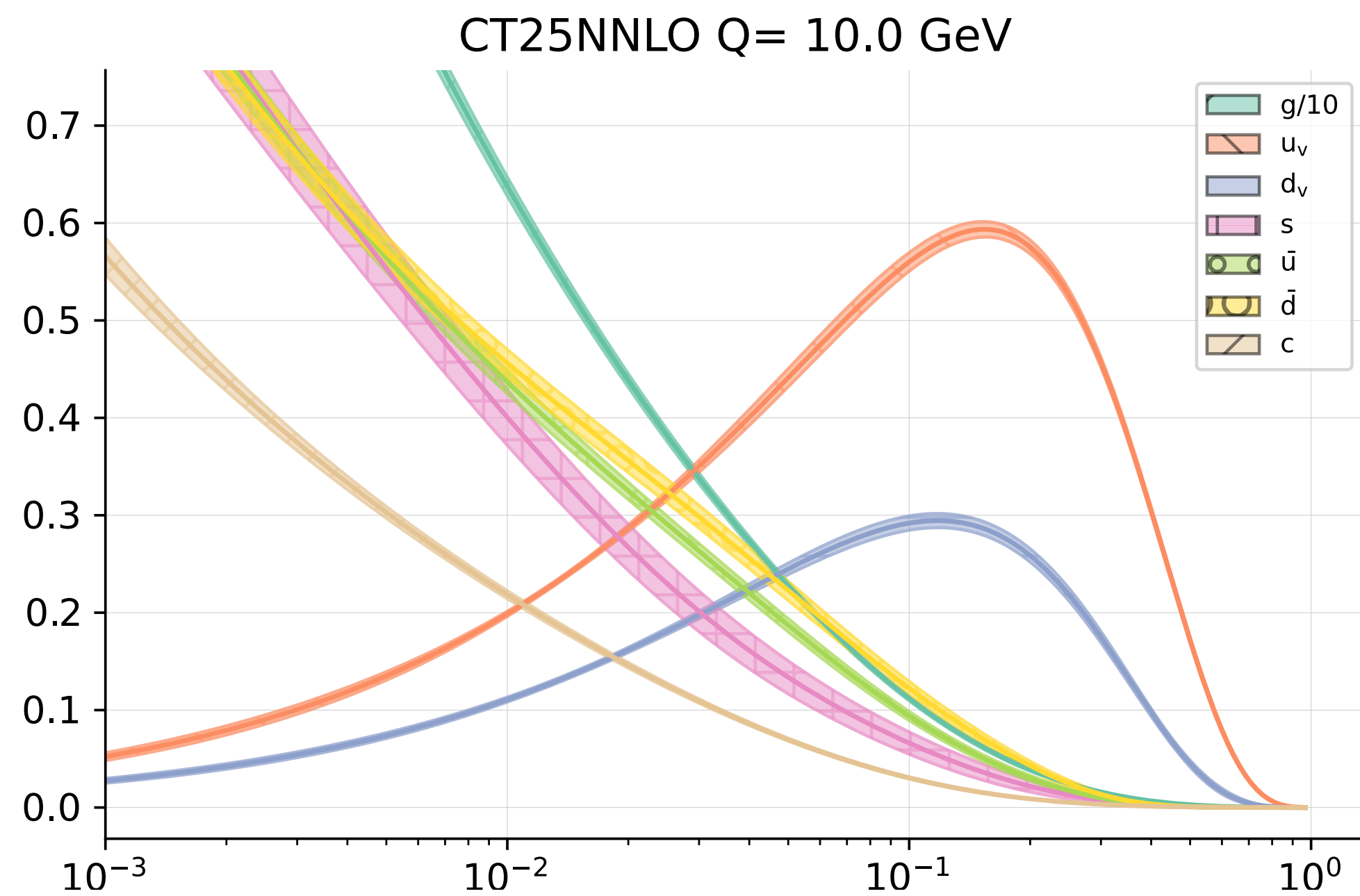
u at 100.0 GeV



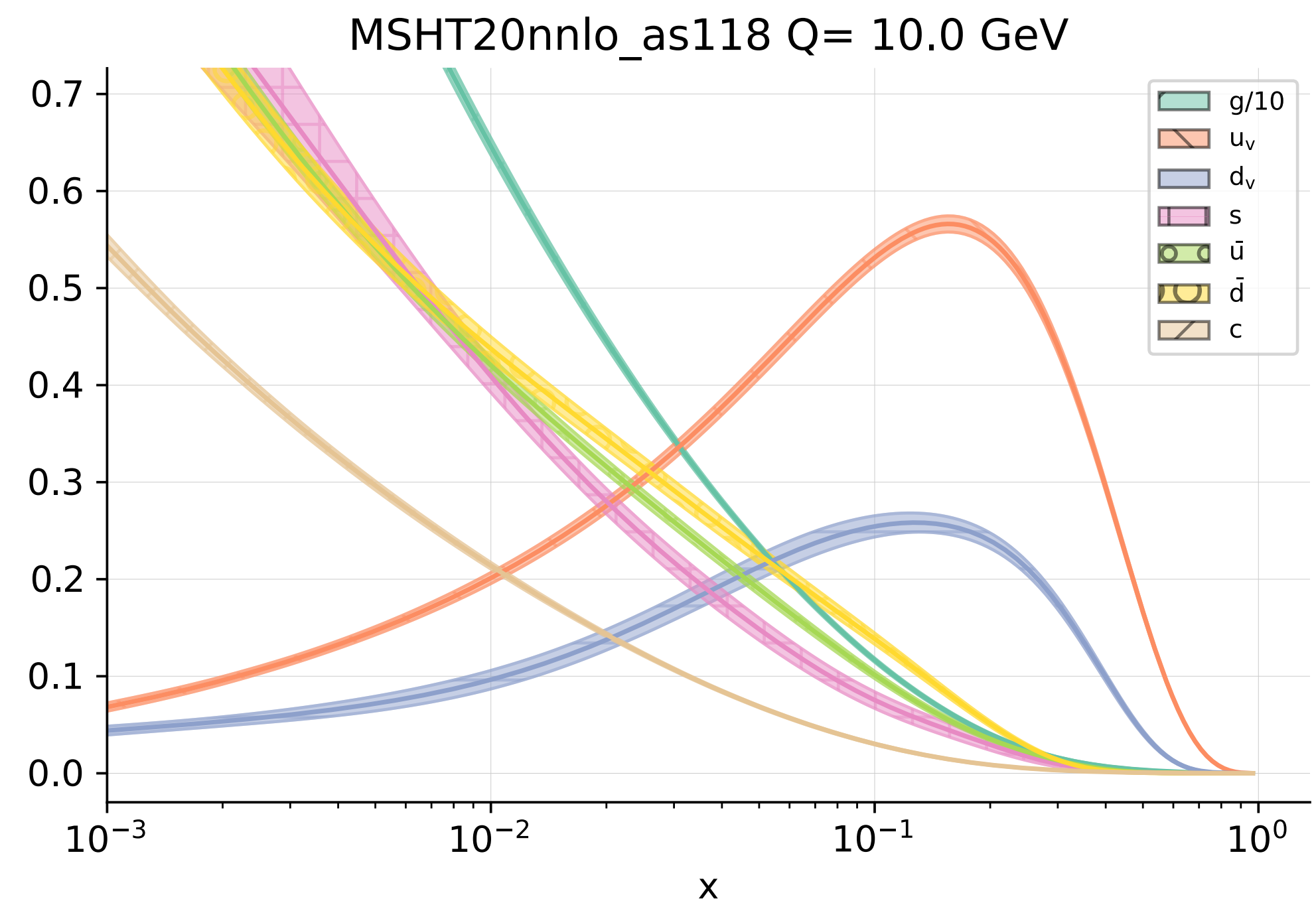
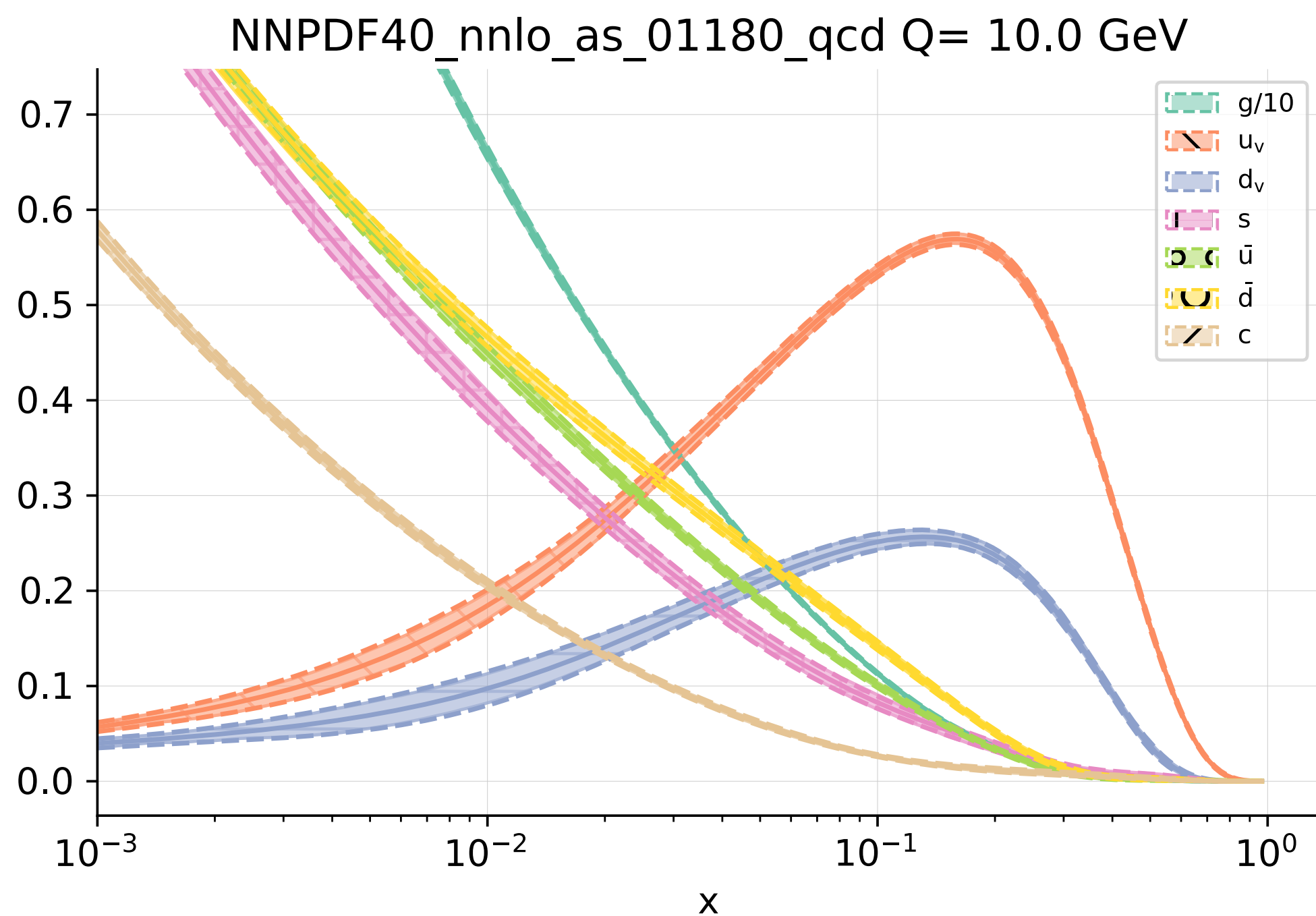
c at 2.0 GeV



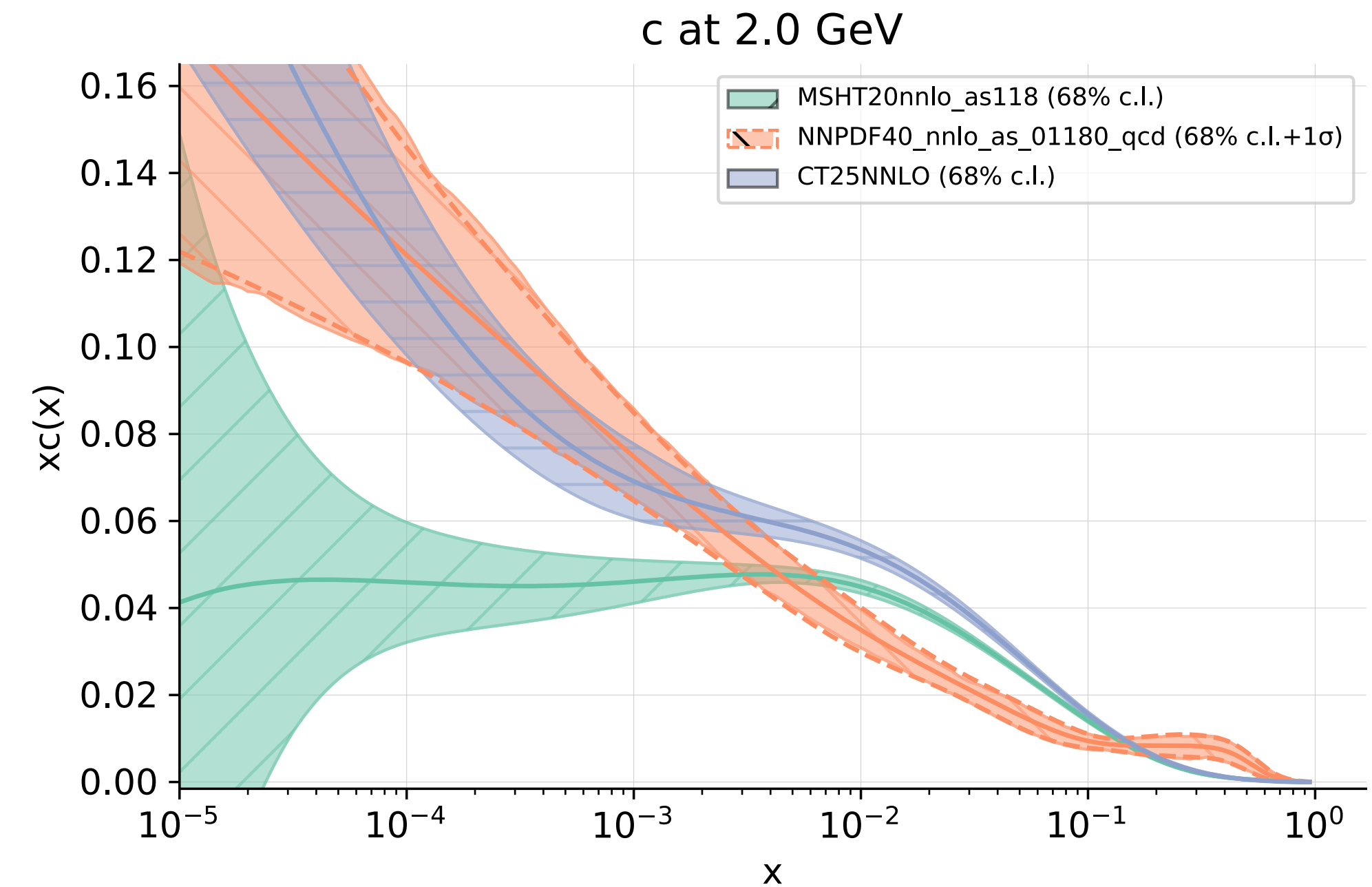
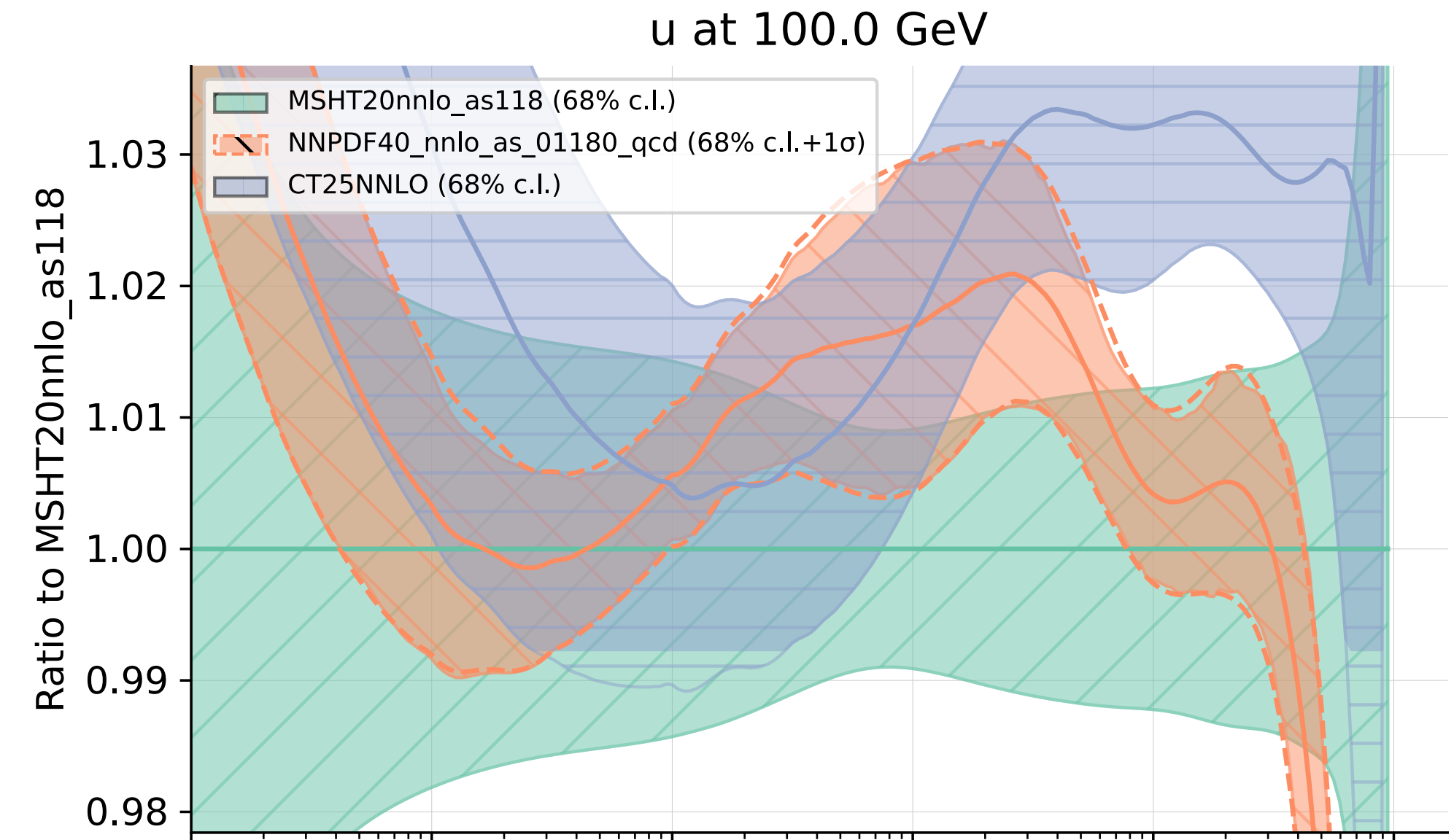
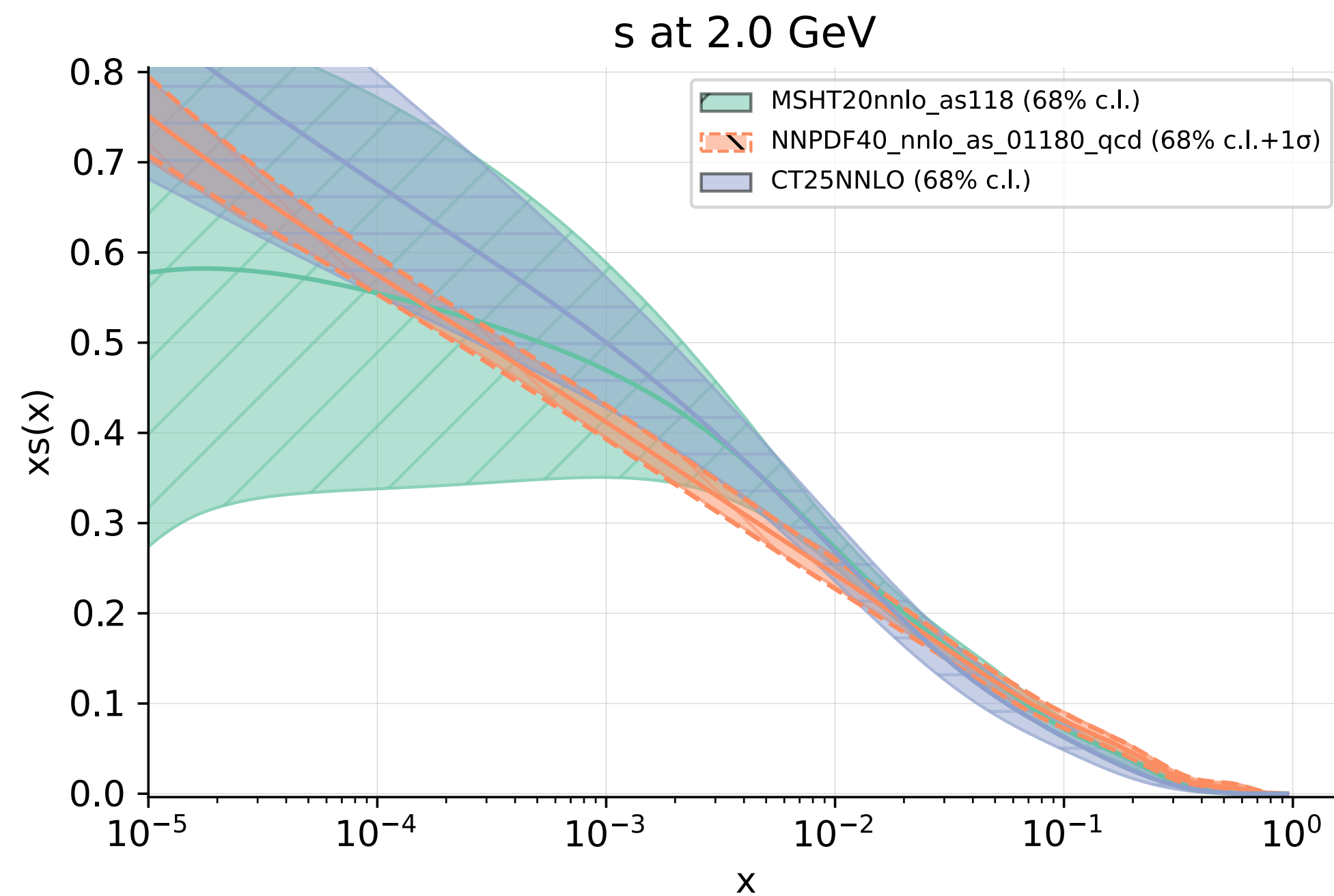
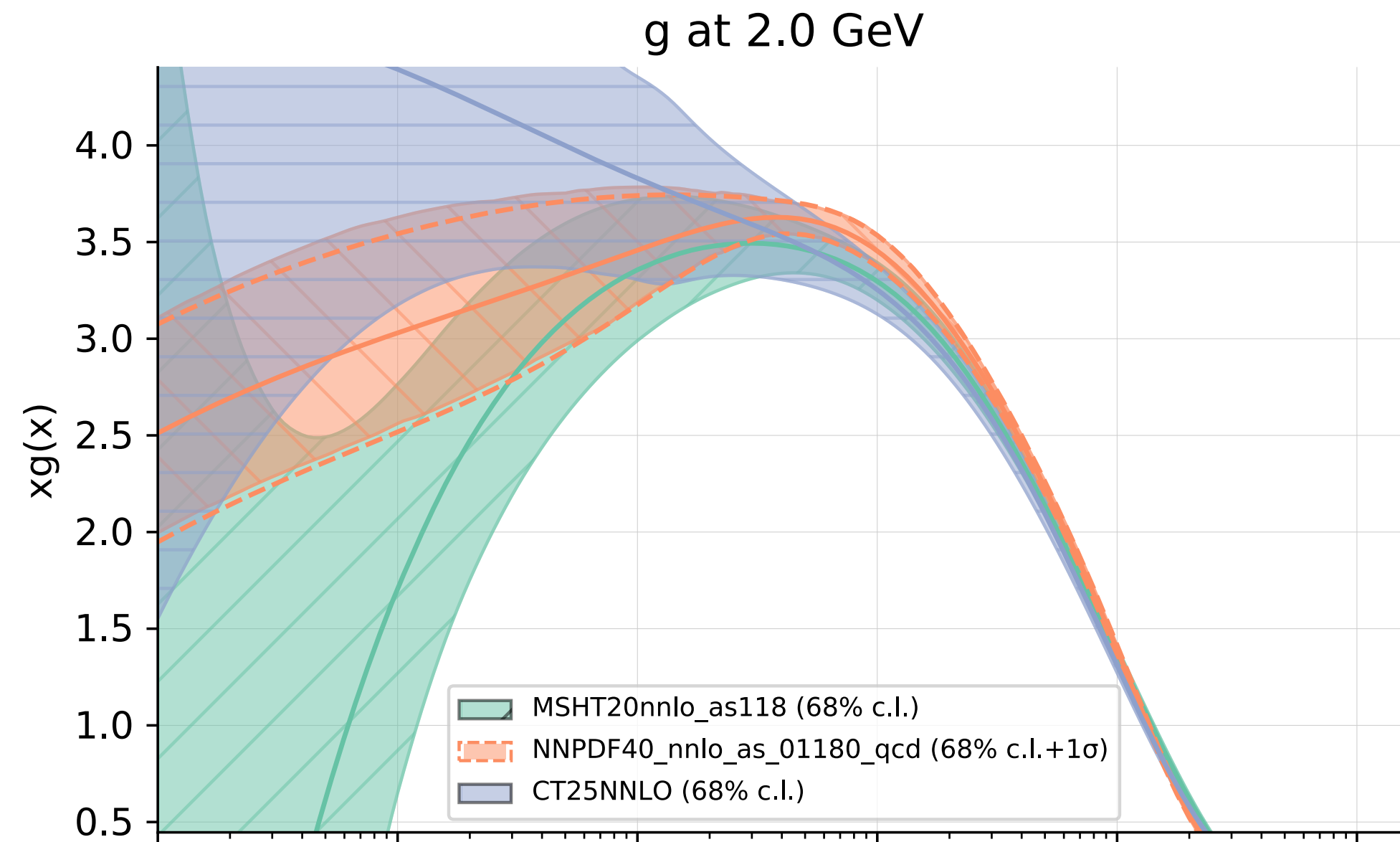
Parton Distribution Functions... today: NNLO



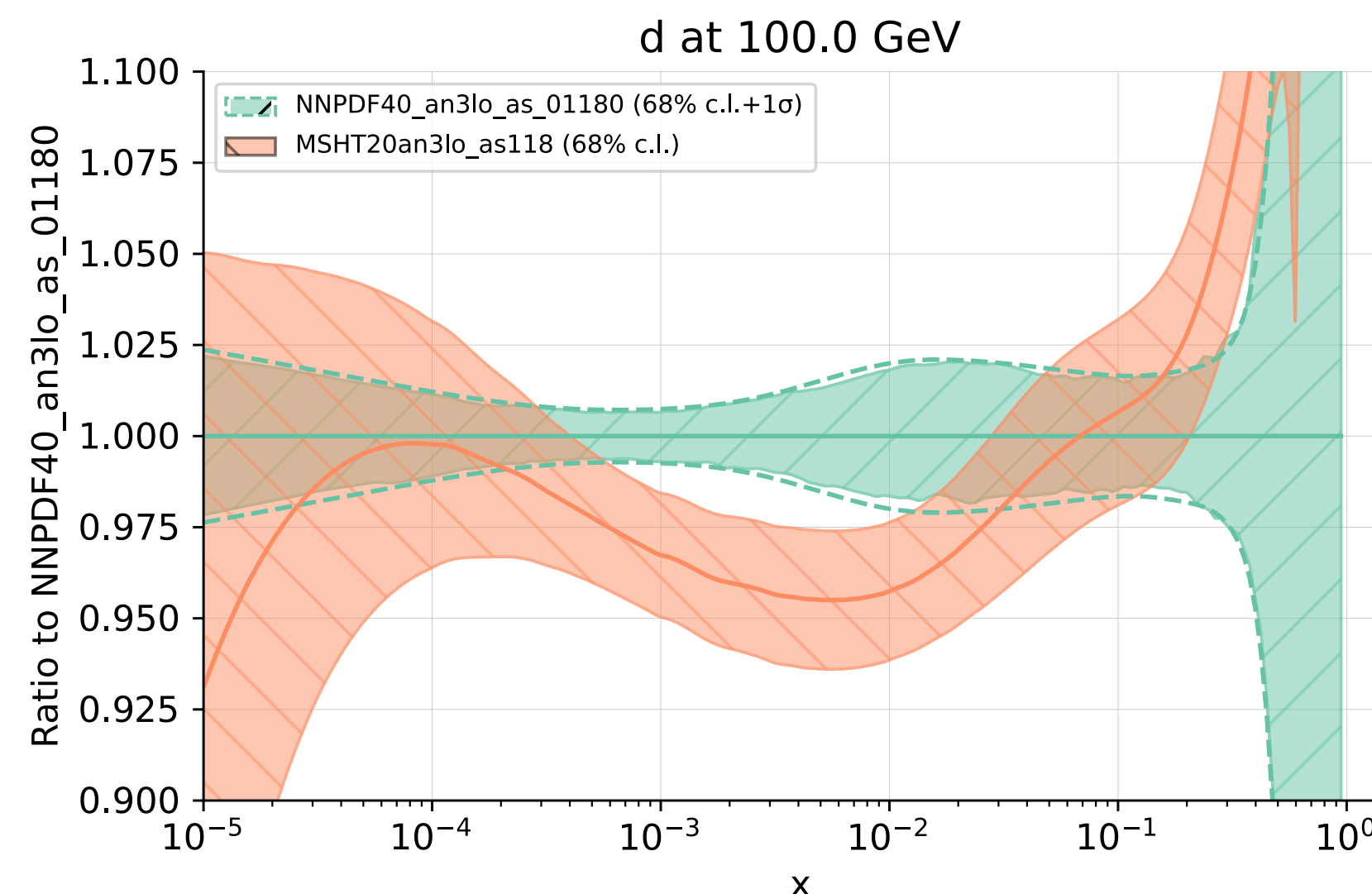
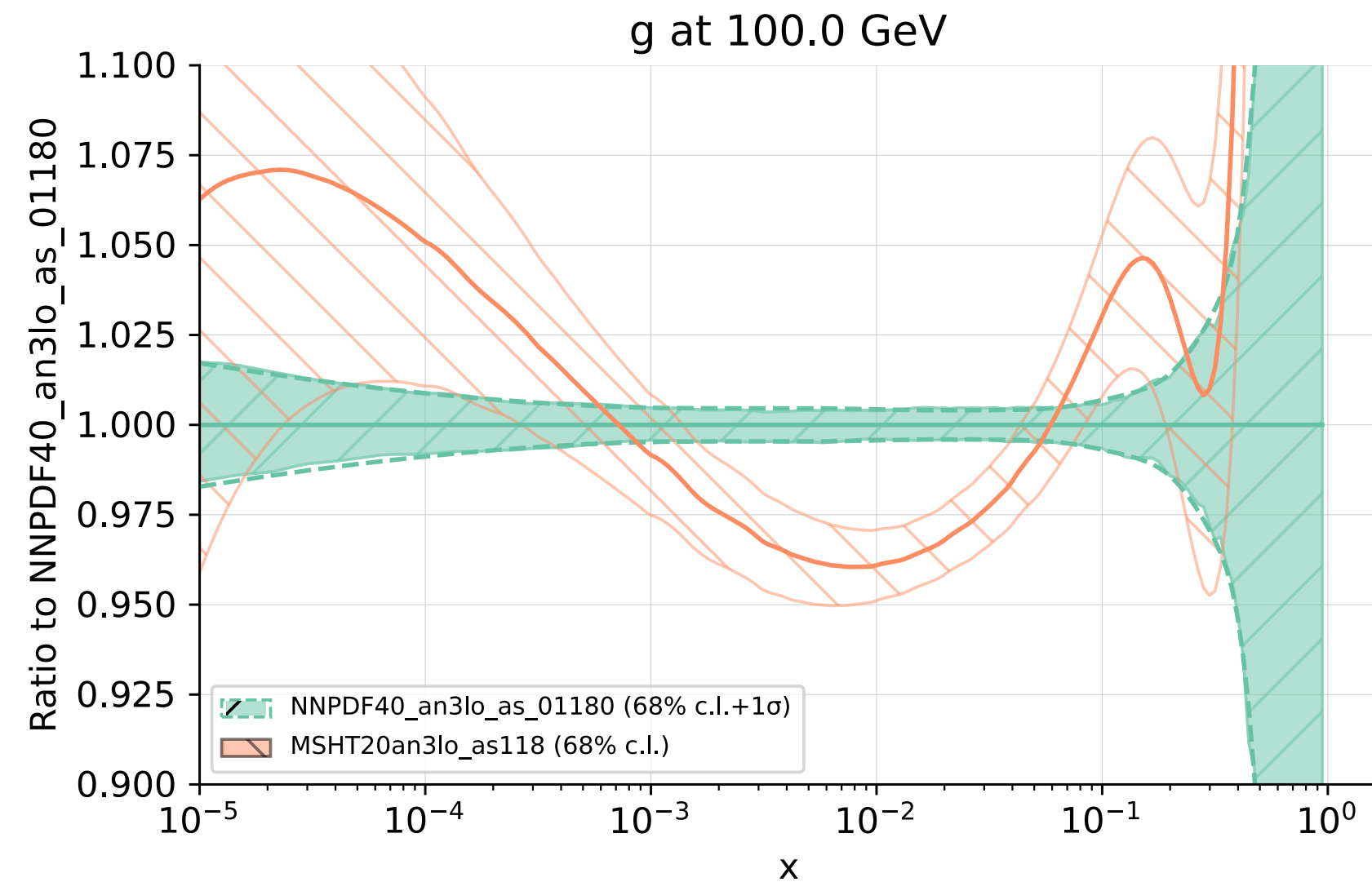
- LHC Data is now commonplace and even dominates
- NNLO corrections included for all data (sometimes in the form of k-factors though)
- Much greater coverage of the entire phase space
- Experimental uncertainties are often so small that tensions arise between datasets
- New sources of uncertainties are now considered explicitly: missing theoretical information, functional uncertainty



Parton Distribution Functions... today: NNLO



Parton Distribution Functions... today: aN3LO



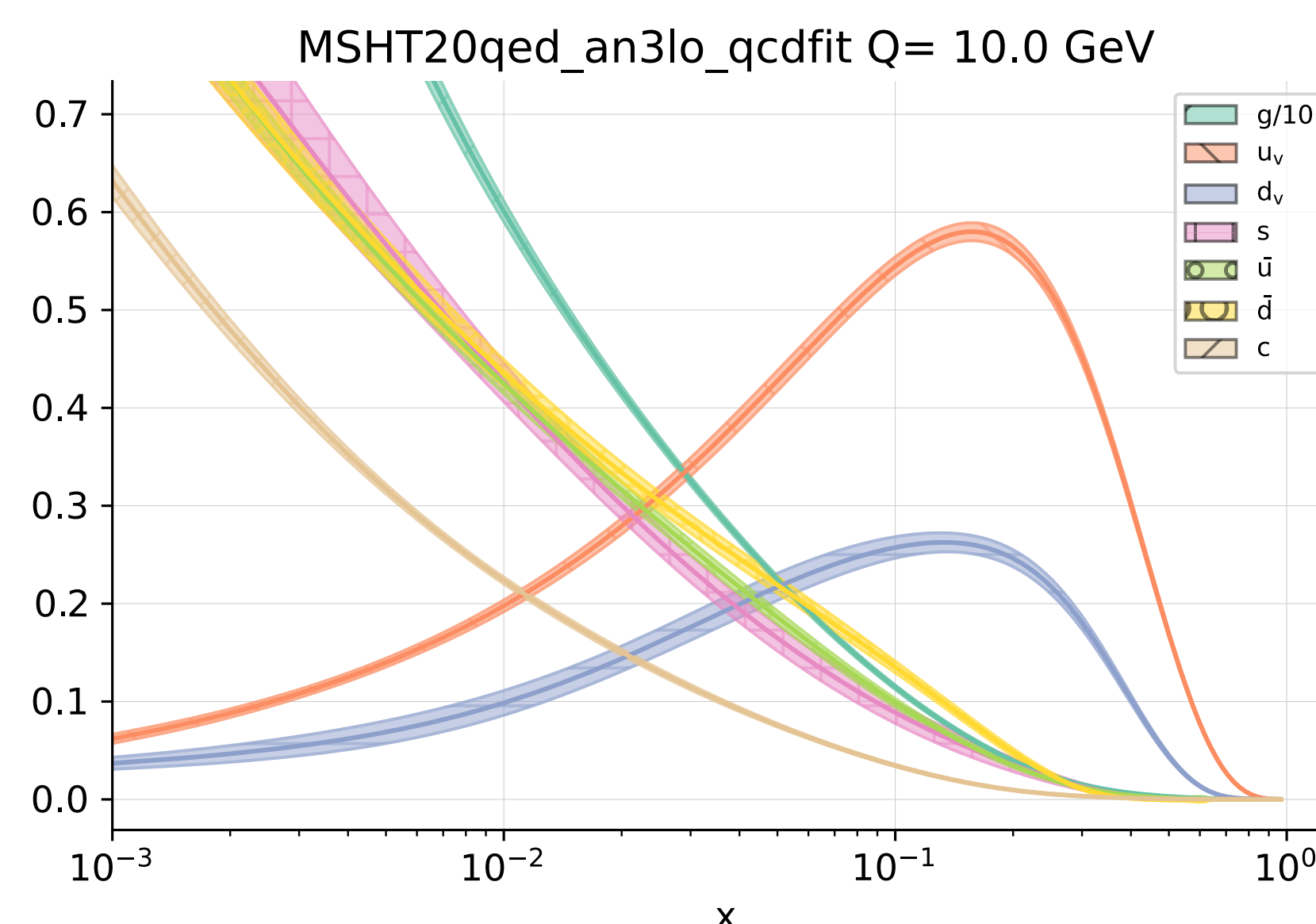
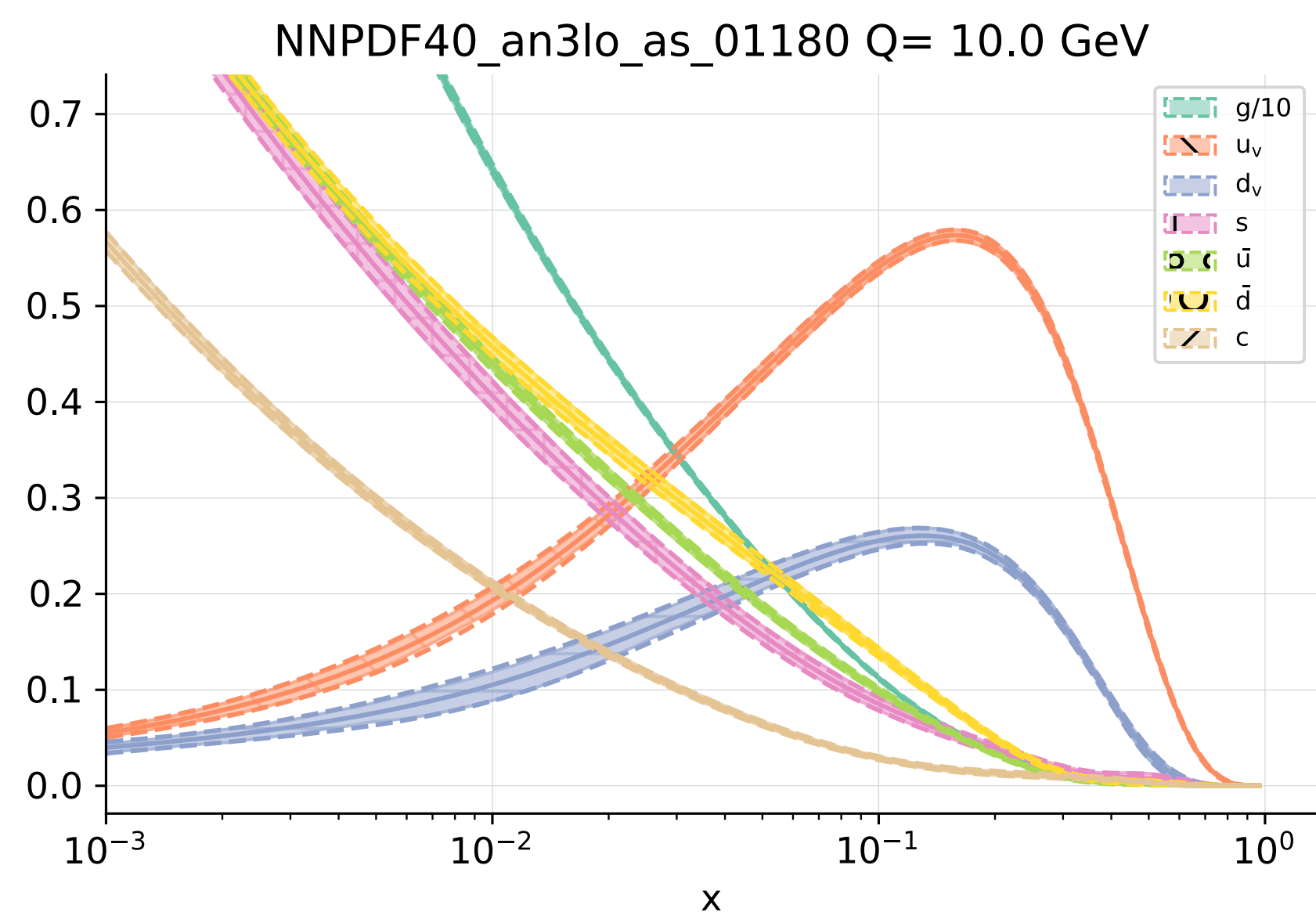
“a” because they are approximated:
the full splitting functions are still not available and N3LO partonic cross sections are only included for DIS data, hadronic is still NNLO

Only MSHT and NNPDF available at the moment, a first combination has been provided in:

Combination of aN3LO PDFs and implications for Higgs production cross-sections at the LHC

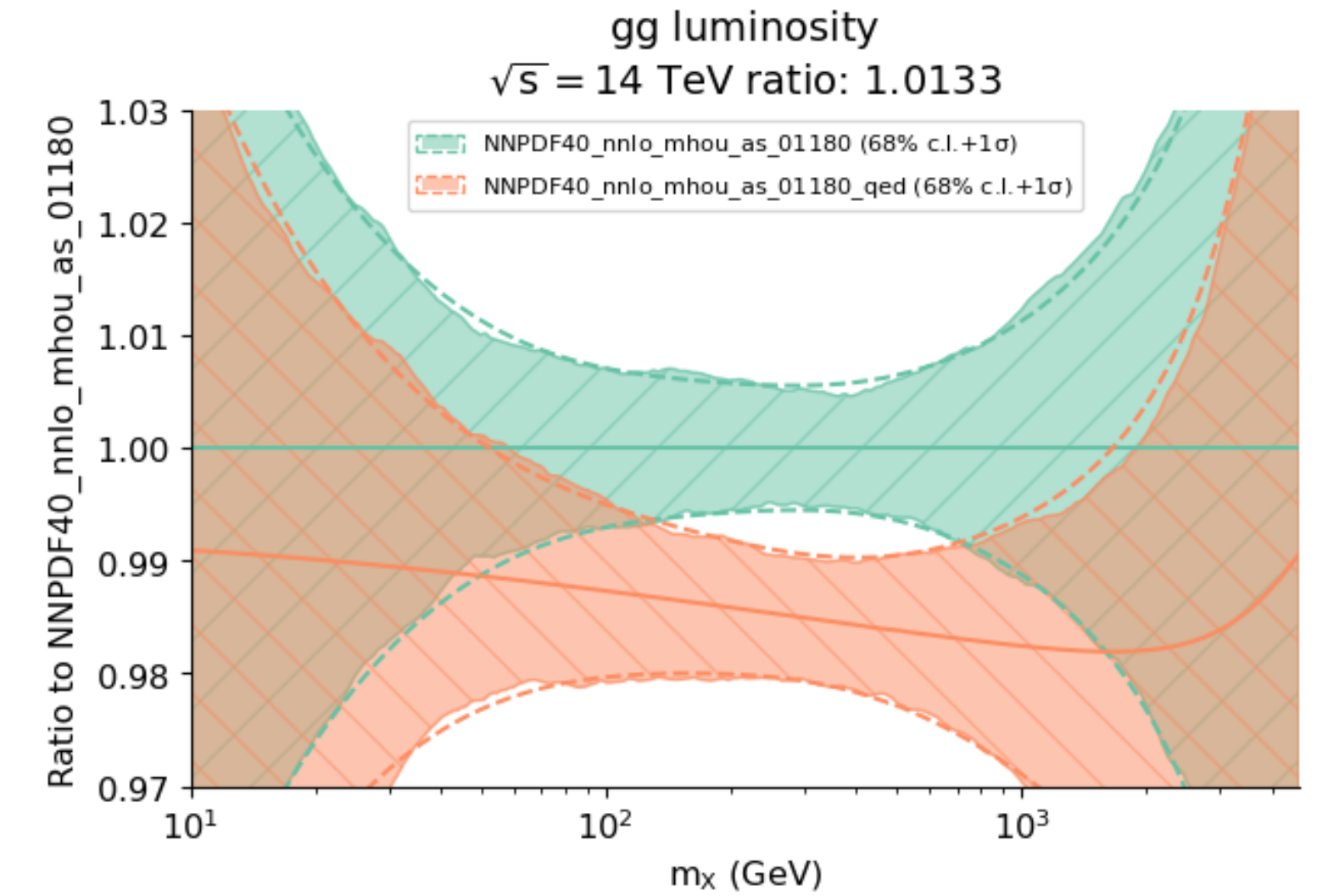
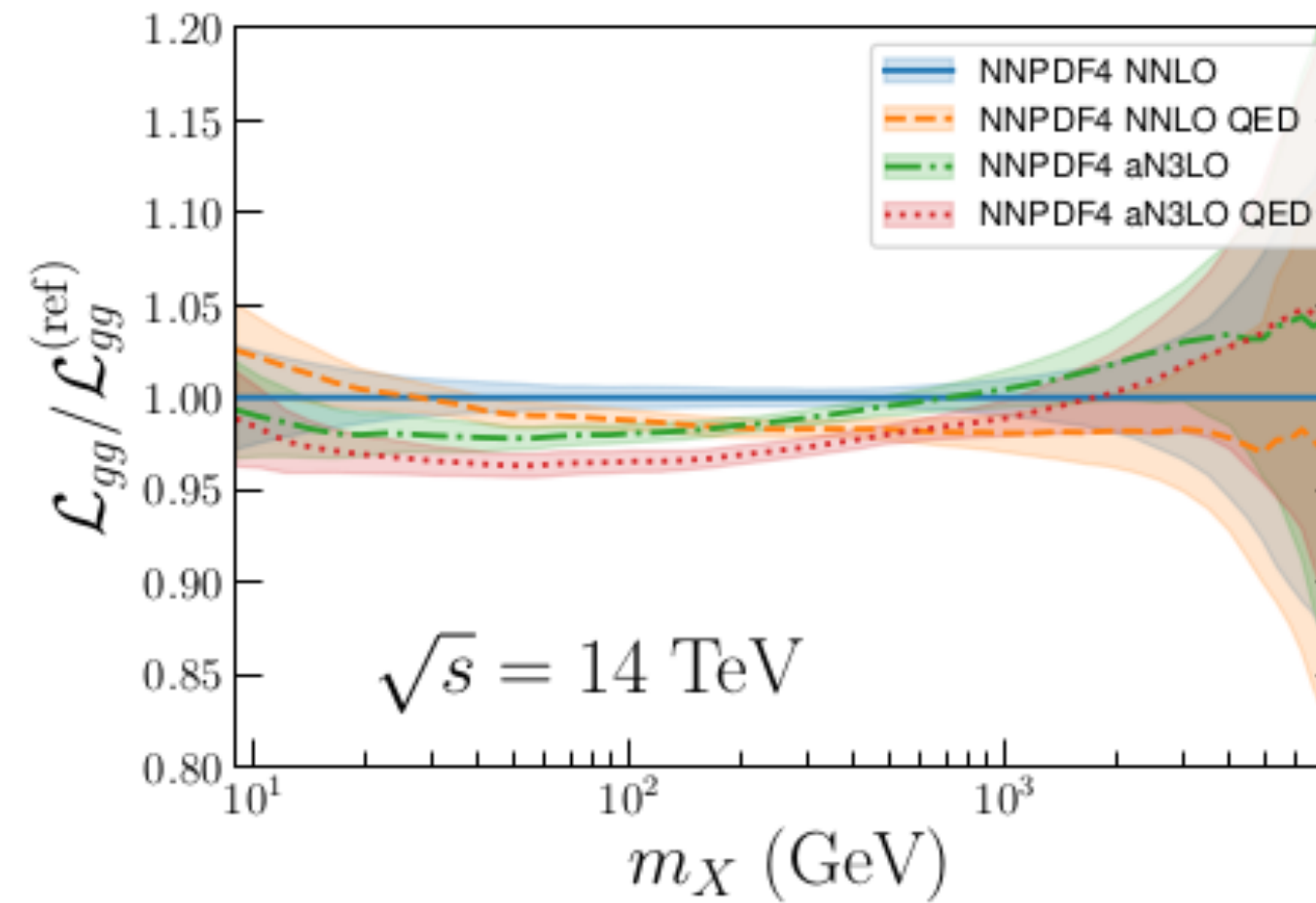
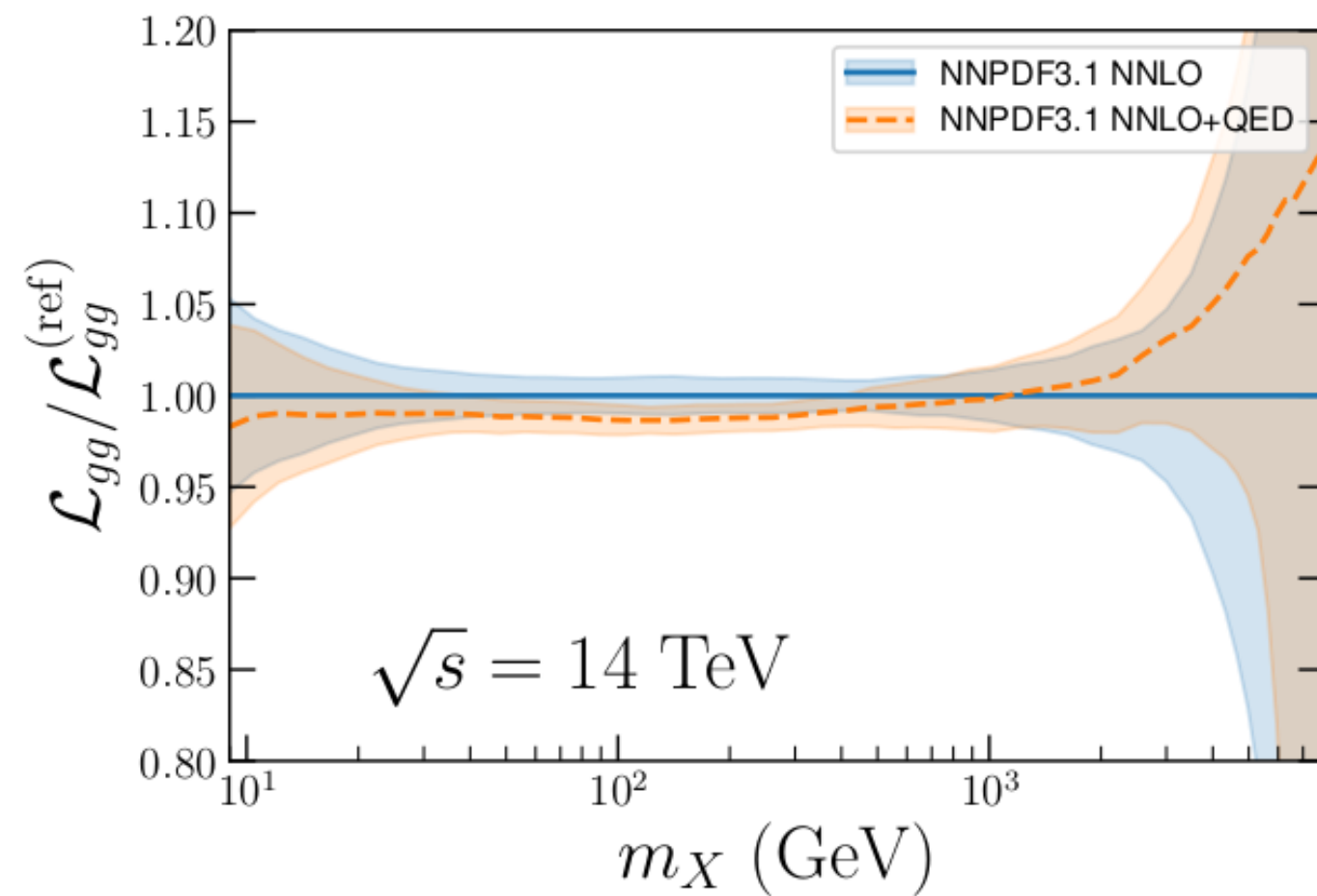
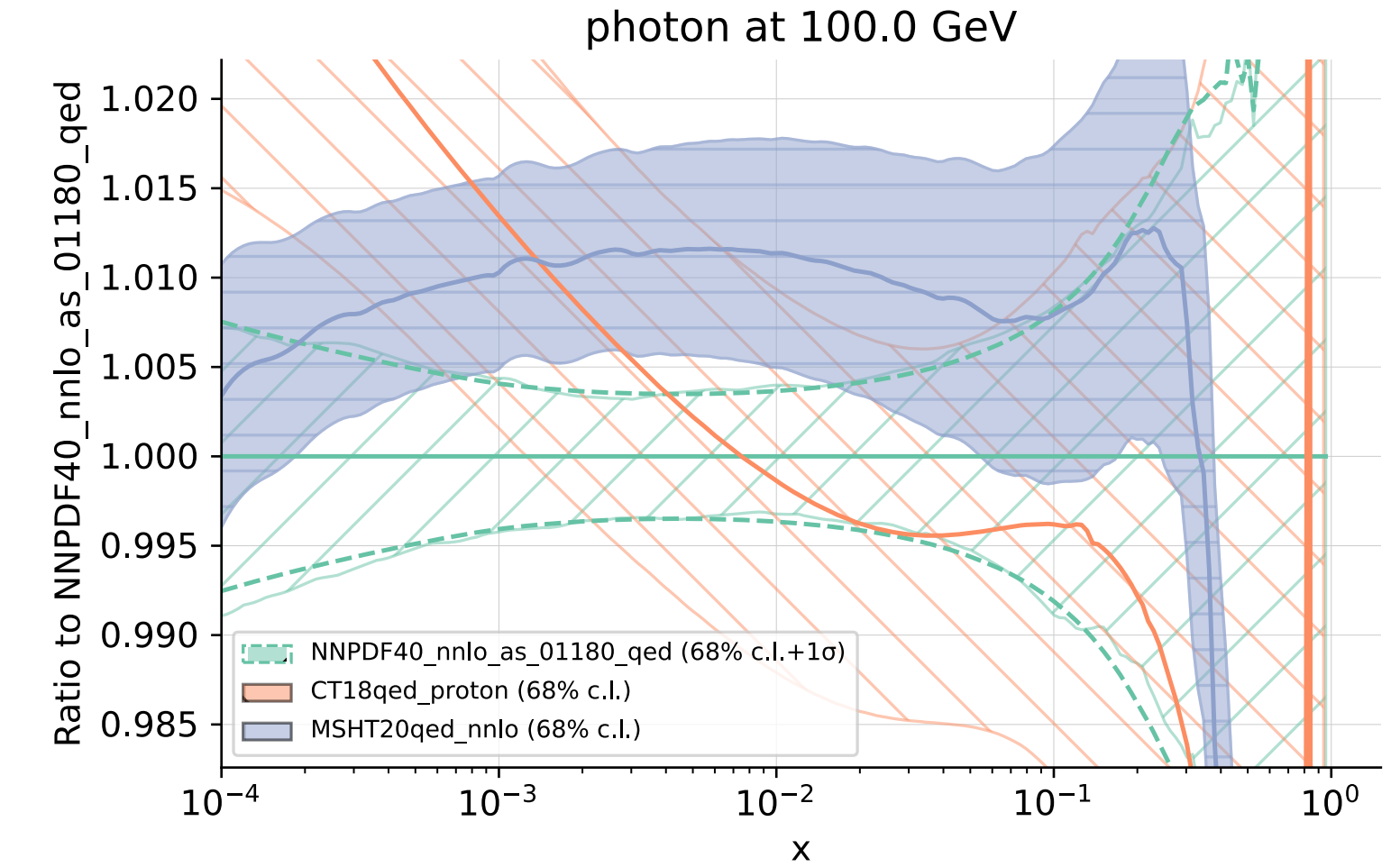
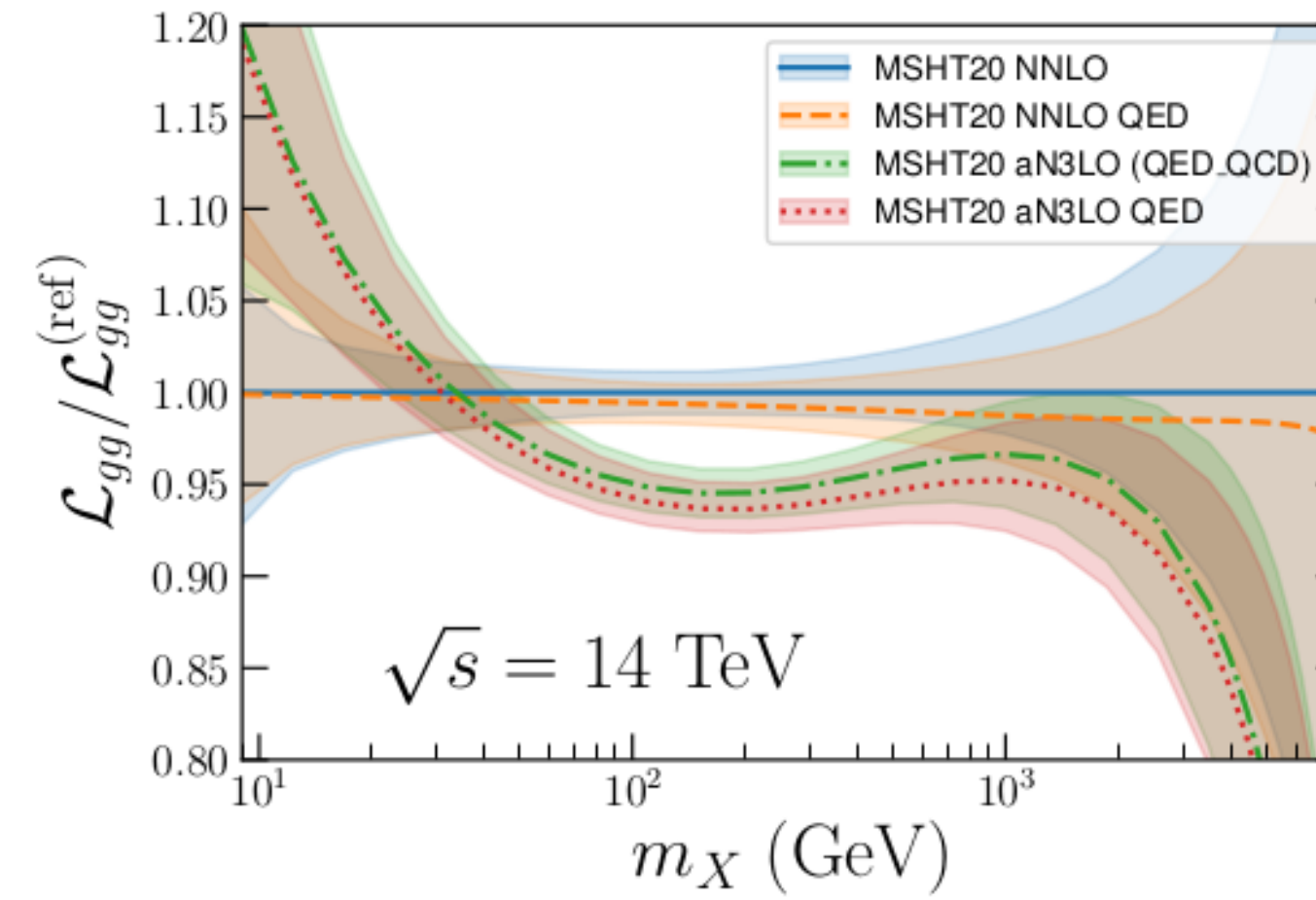
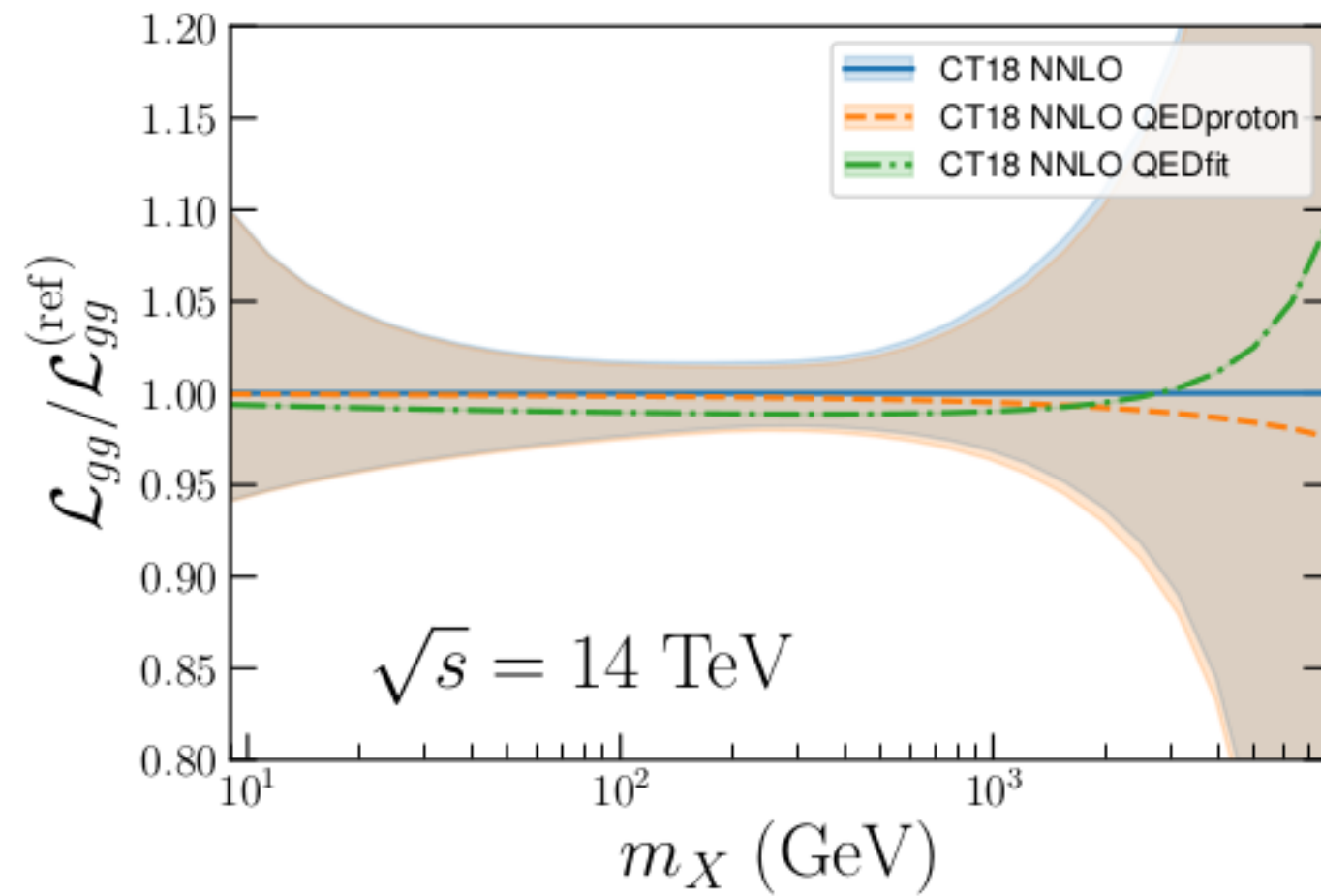
MSHT & NNPDF collaborations

[hep-ph/2411.05373](https://arxiv.org/abs/hep-ph/2411.05373)



An update by CTEQ + updates from all groups with the latest information available (making that “a” smaller with each passing day!) and a full benchmark will come in the near future!

Parton Distribution Functions... today: QED

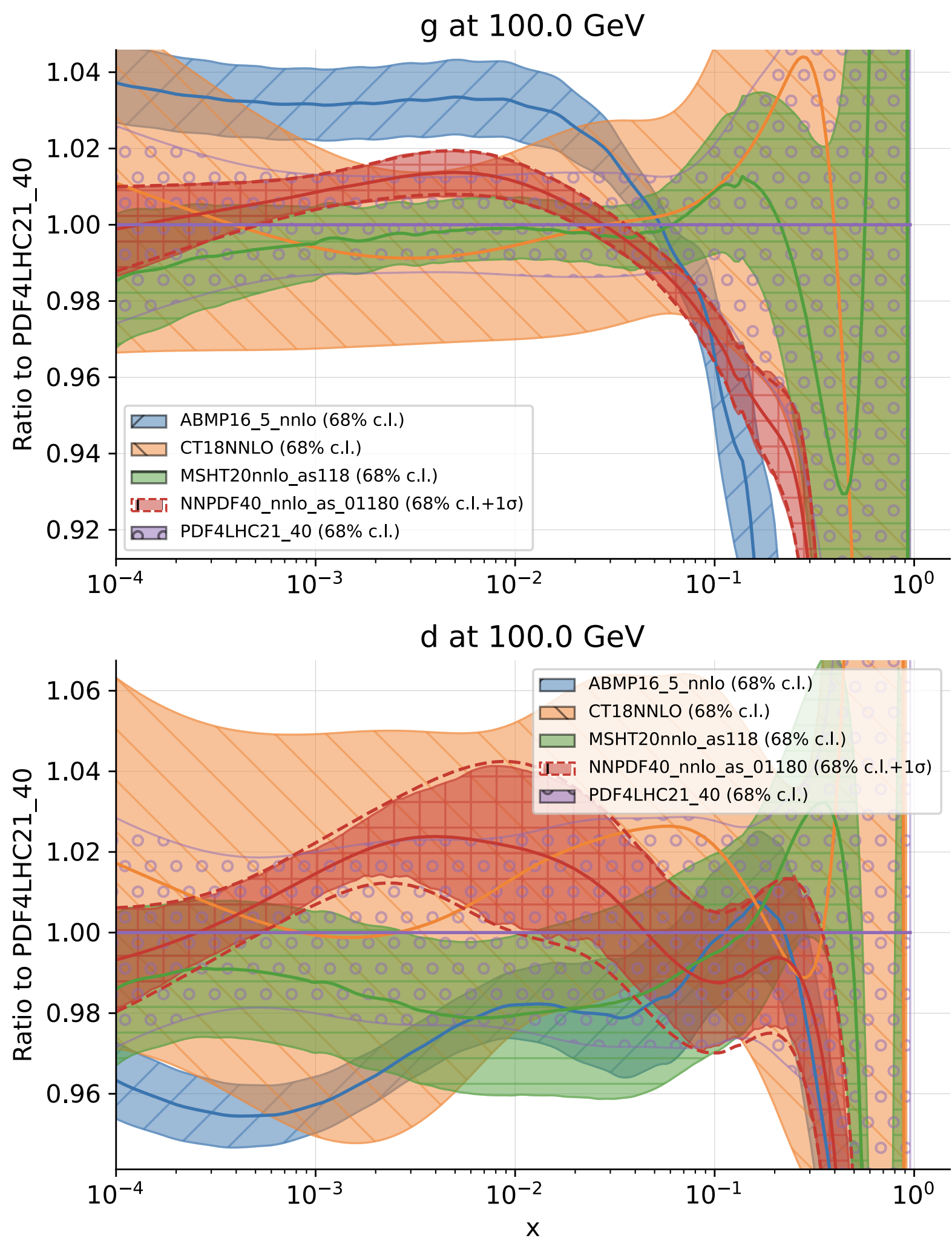
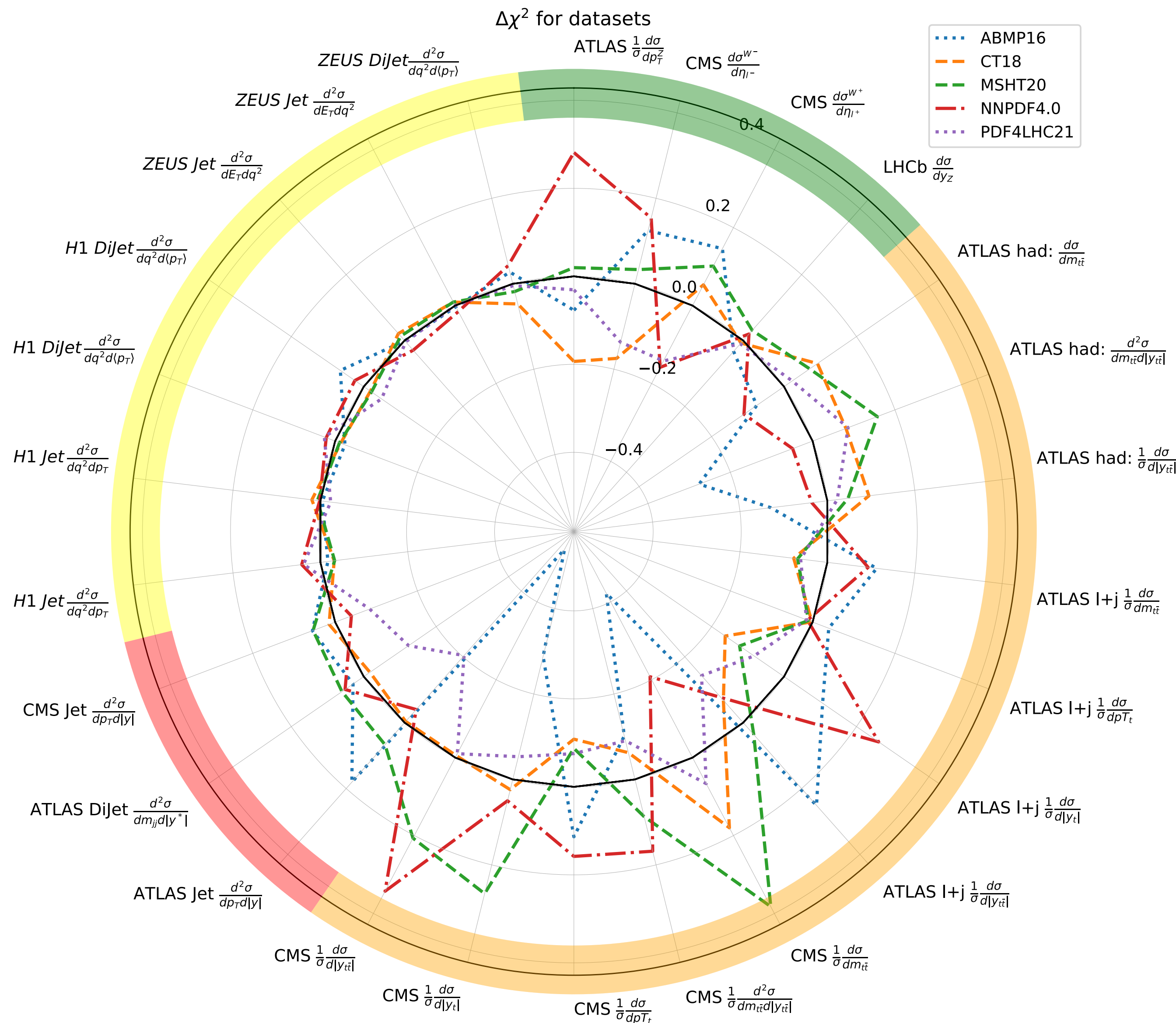


A. Cooper-Sarkar, T. Cridge, T. Hobbs, J. Huston, P. Nadolsky, M. Ponce-Chavez, K. Xie - hep-ph/2508.06603

T. Cridge, JCM, J. Huston - hep-ph/2602.06908

Similar fit quality when presented with unseen data

A. Chiefa, M. Costantini, JCM, E. Nocera, T. Rabemananjara, J. Rojo, R. Stegeman, T. Sharma, M. Ubiali - [hep-ph/2501.10359](https://arxiv.org/abs/hep-ph/2501.10359)



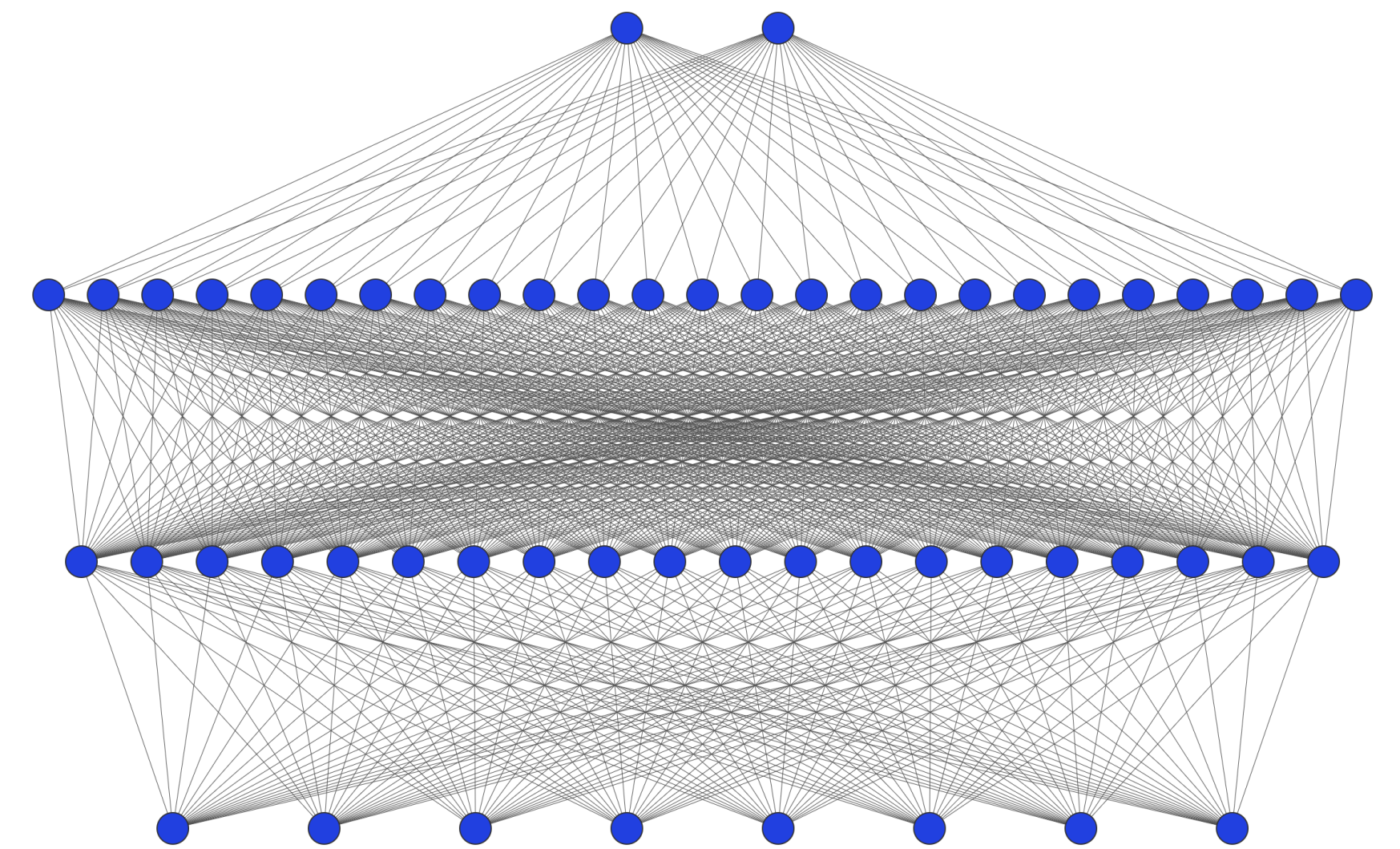
Introduction

A tale of shrinking uncertainties

The NNPDF methodologies: neural networks
and ensemble regression



Results & Conclusions

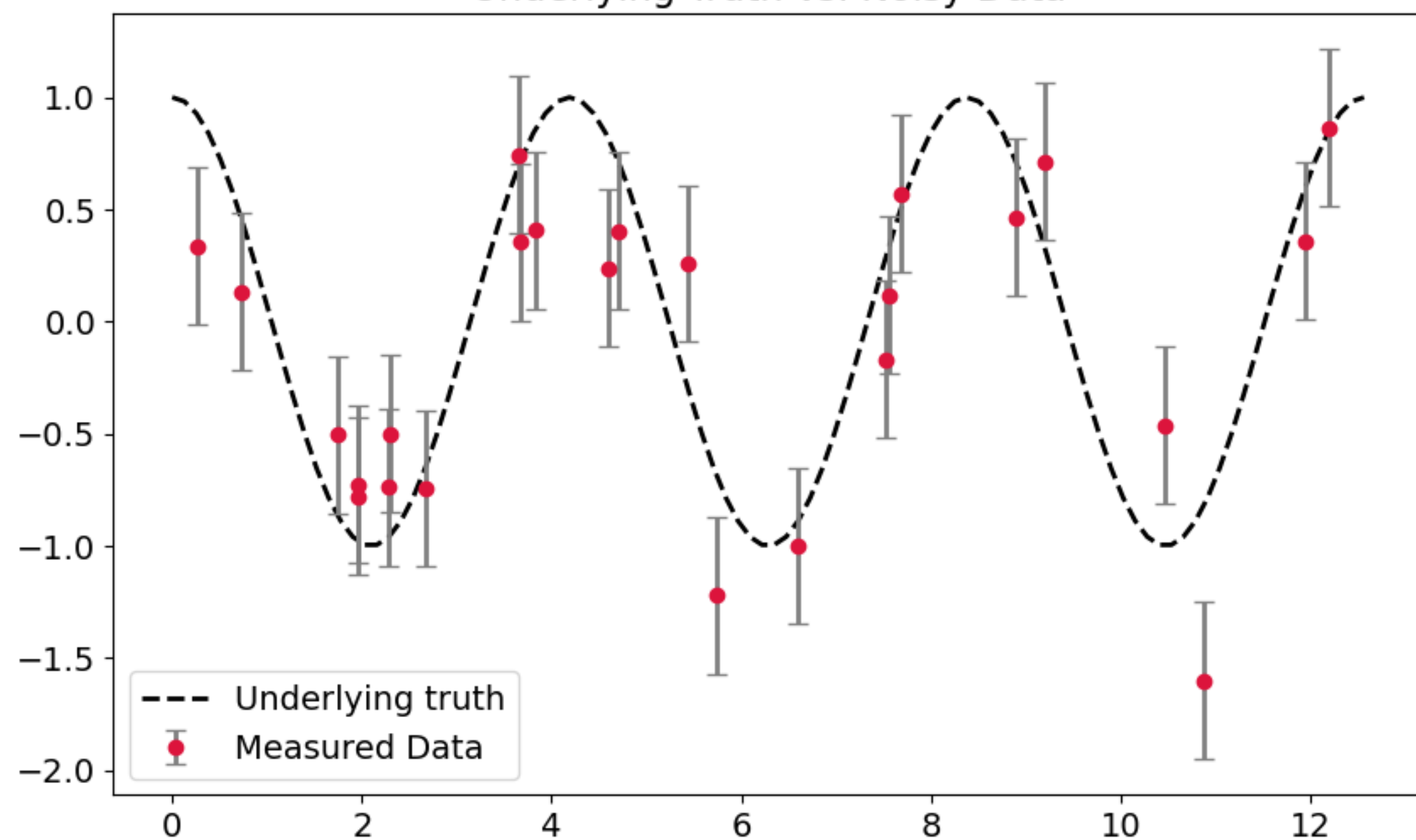


Monte Carlo Uncertainties 101

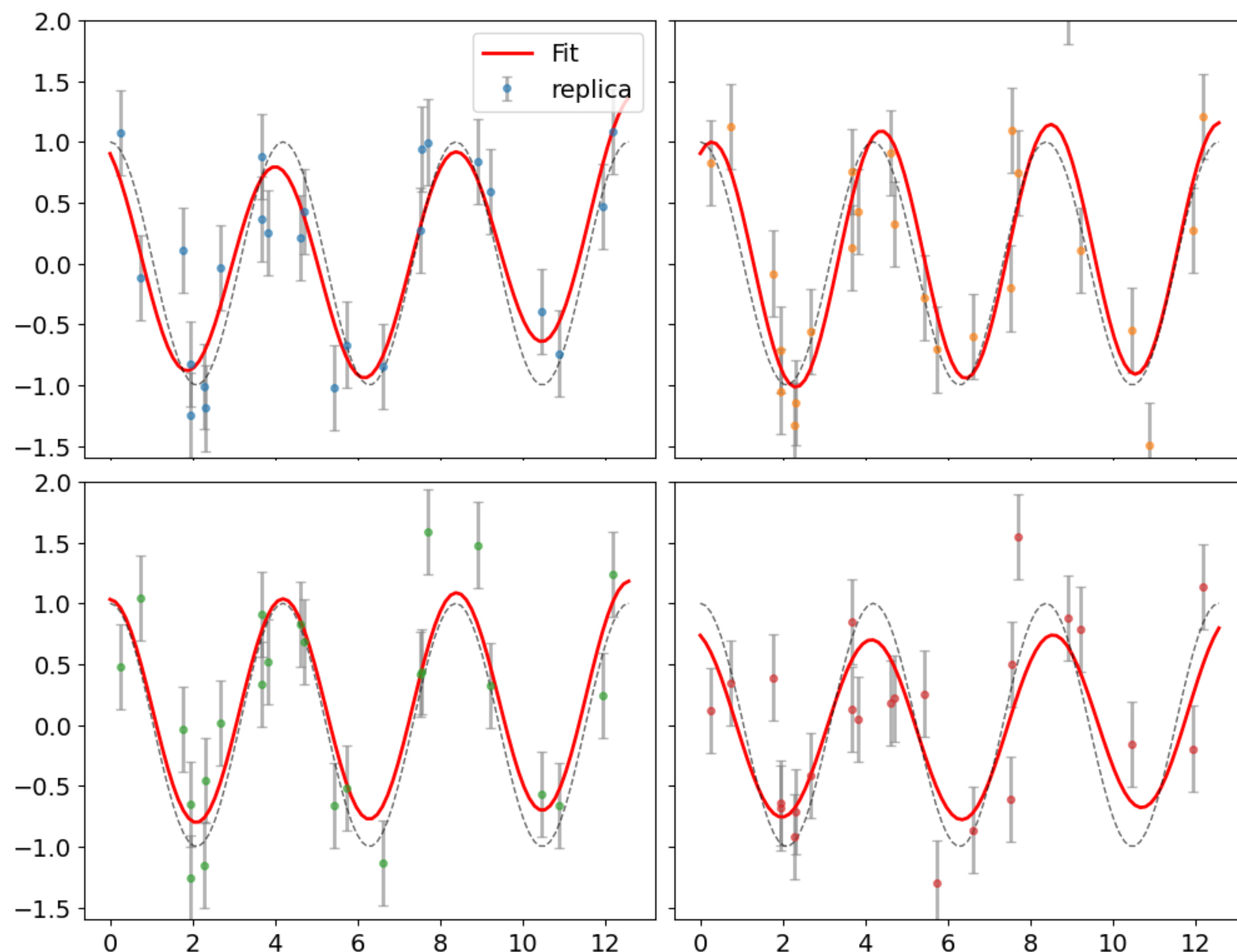
Truth: $y = \cos(3x/2)$

Fit: $y = A \sin Bx + C$

Underlying Truth vs. Noisy Data



Ensemble of independent fits

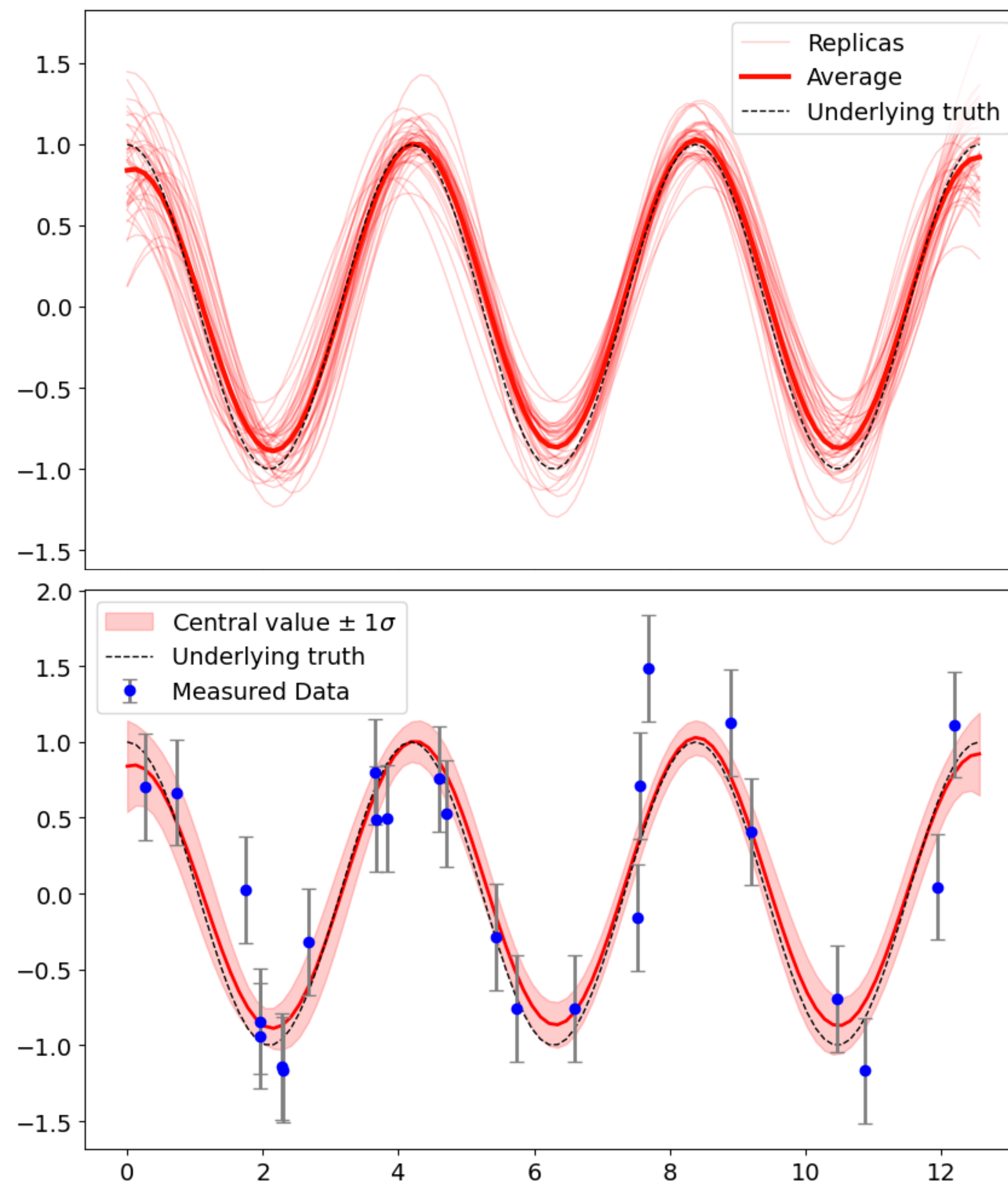
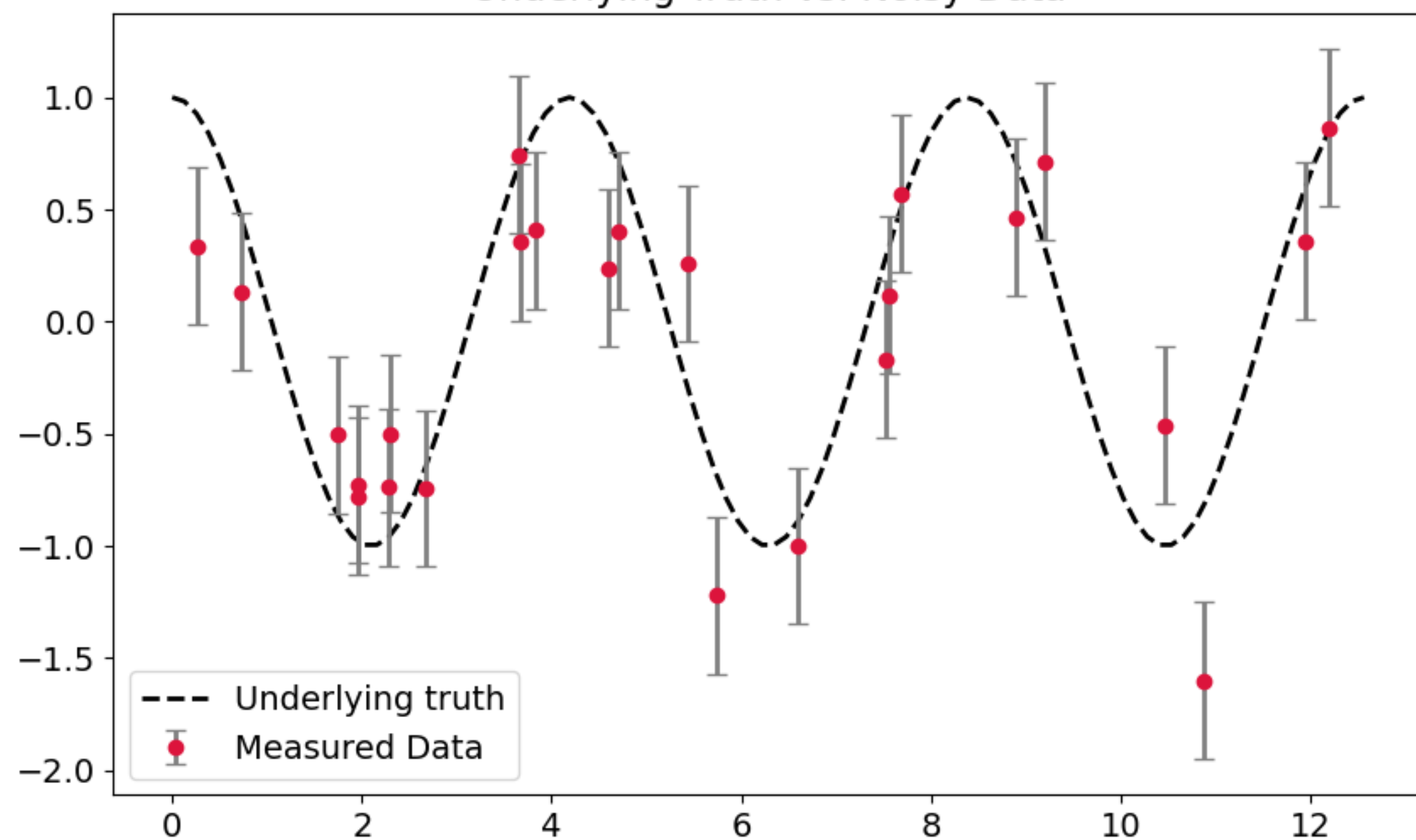


Monte Carlo Uncertainties 101

Truth: $y = \cos(3x/2)$

Fit: $y = A \sin Bx + C$

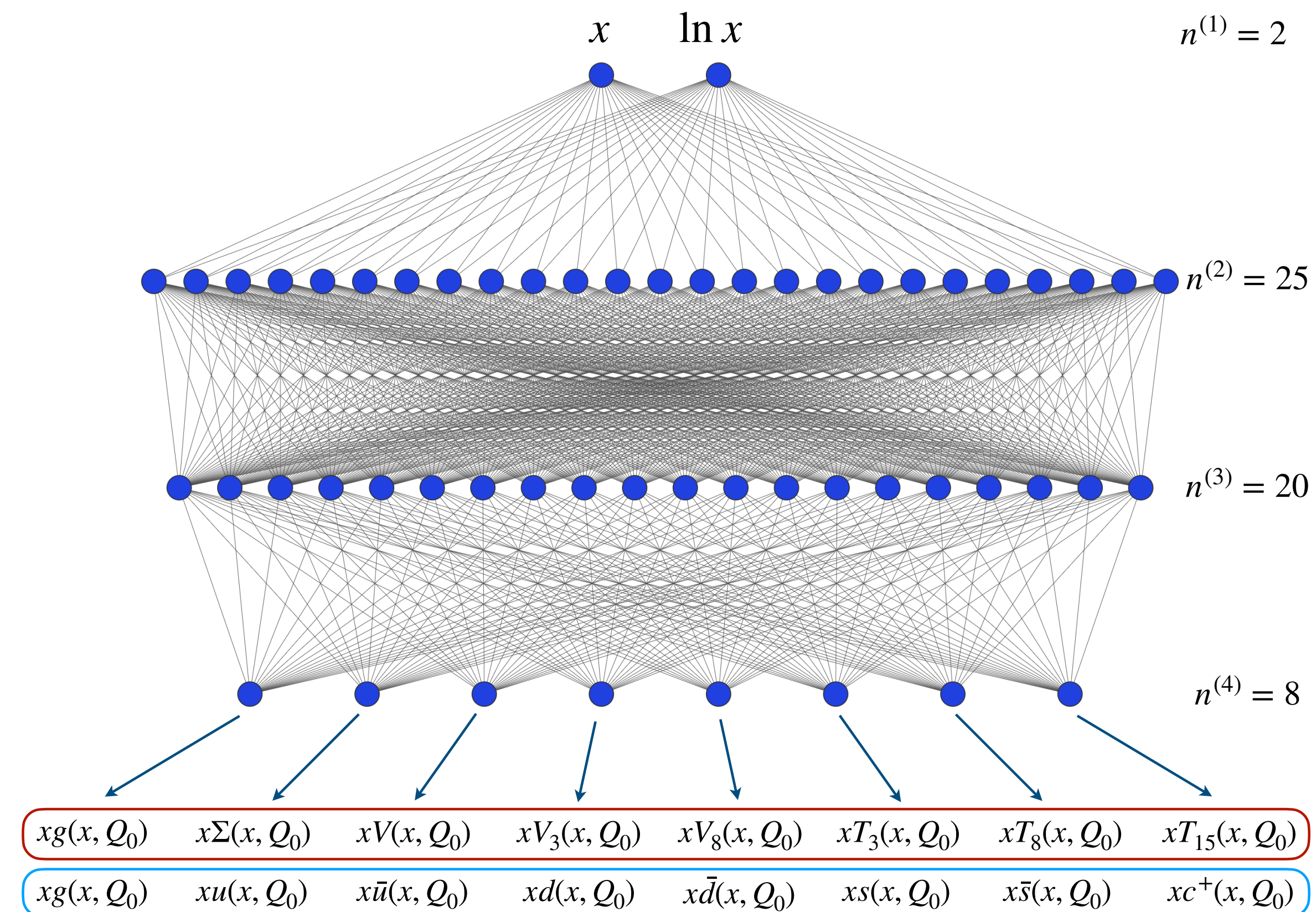
Underlying Truth vs. Noisy Data



But what if I know nothing about the functional form?

Neural Networks can act as universal functional approximation.

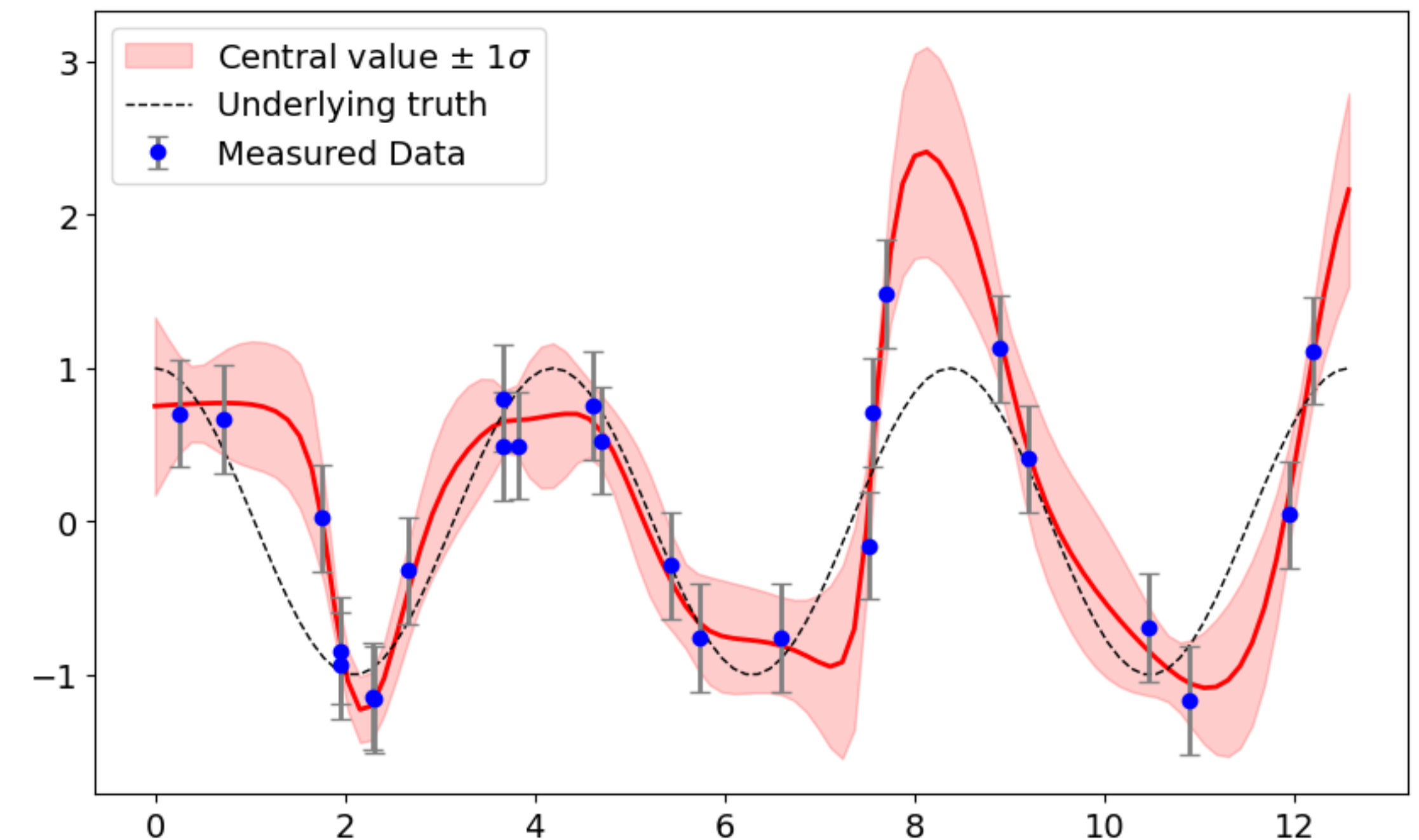
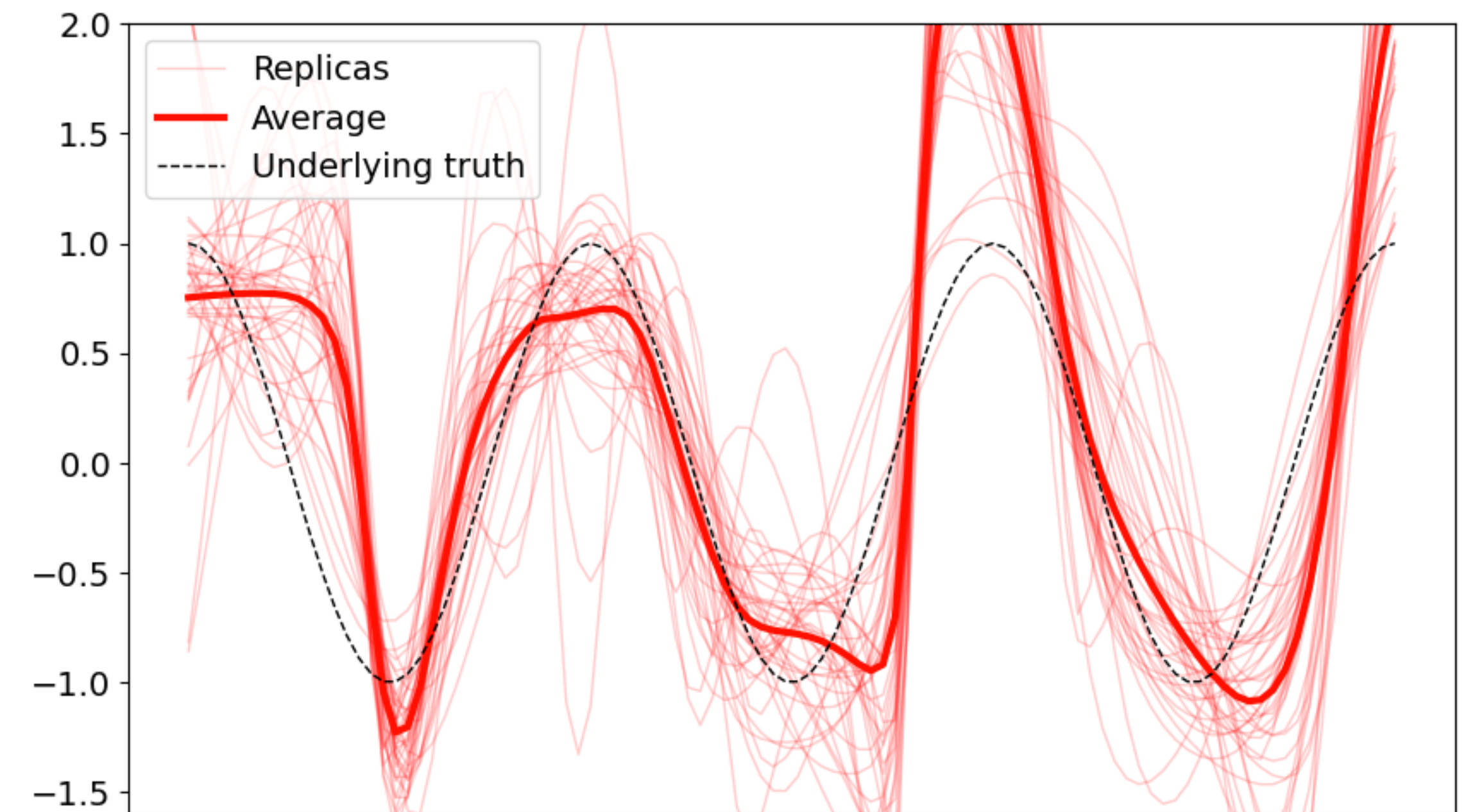
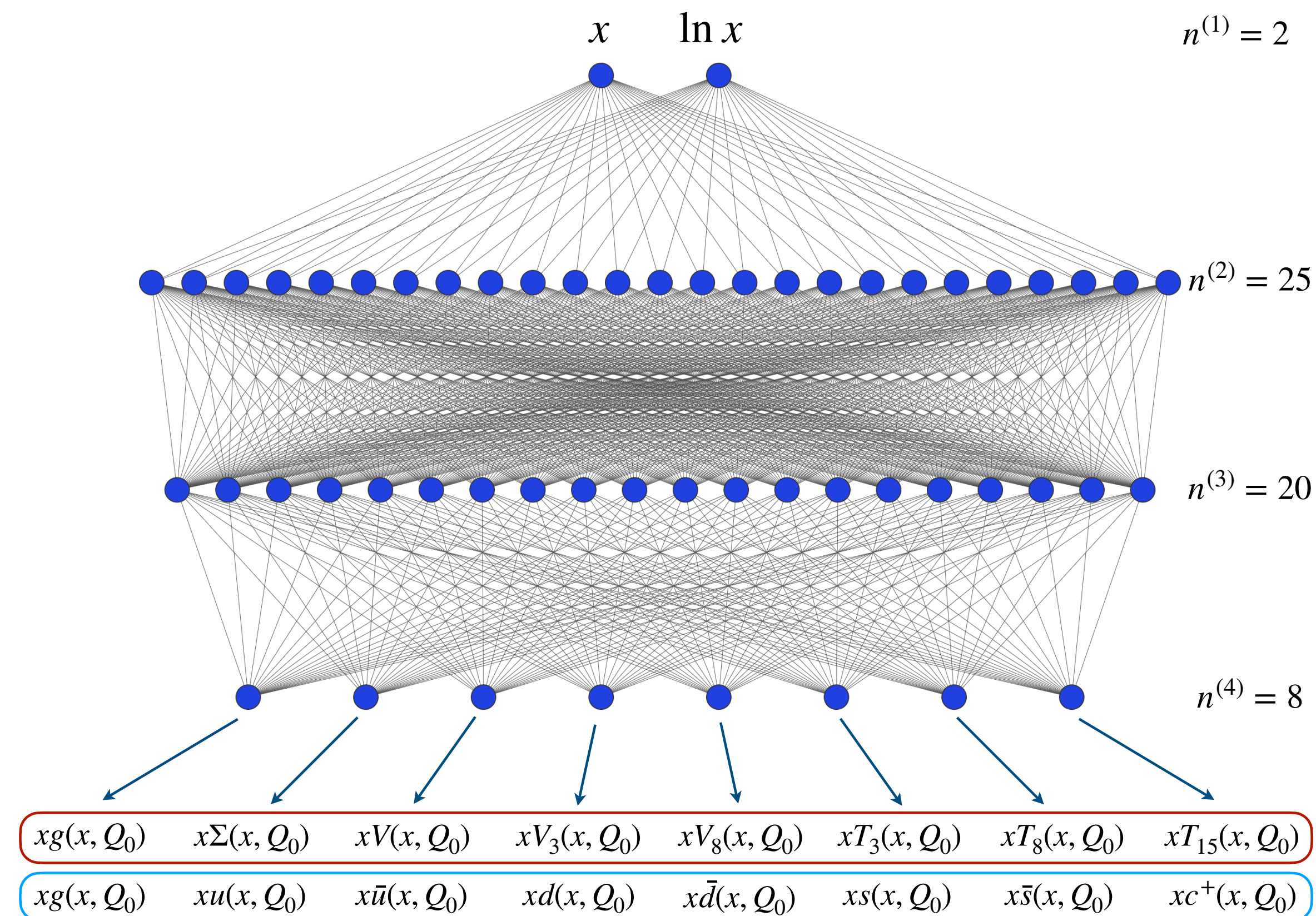
Just give them enough data (or physical constraints) to help them out



But what if I know nothing about the functional form?

Neural Networks can act as universal functional approximation.

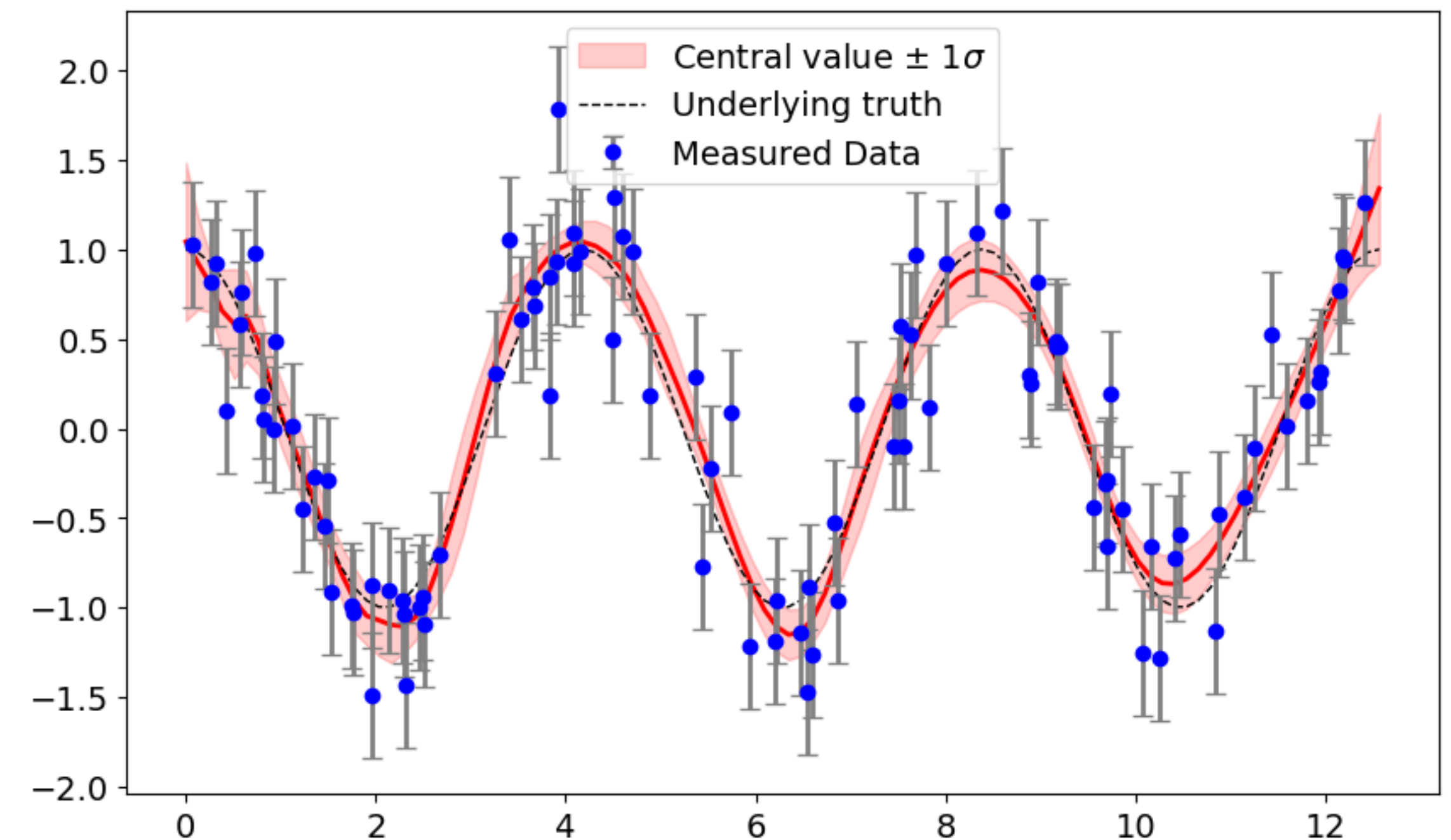
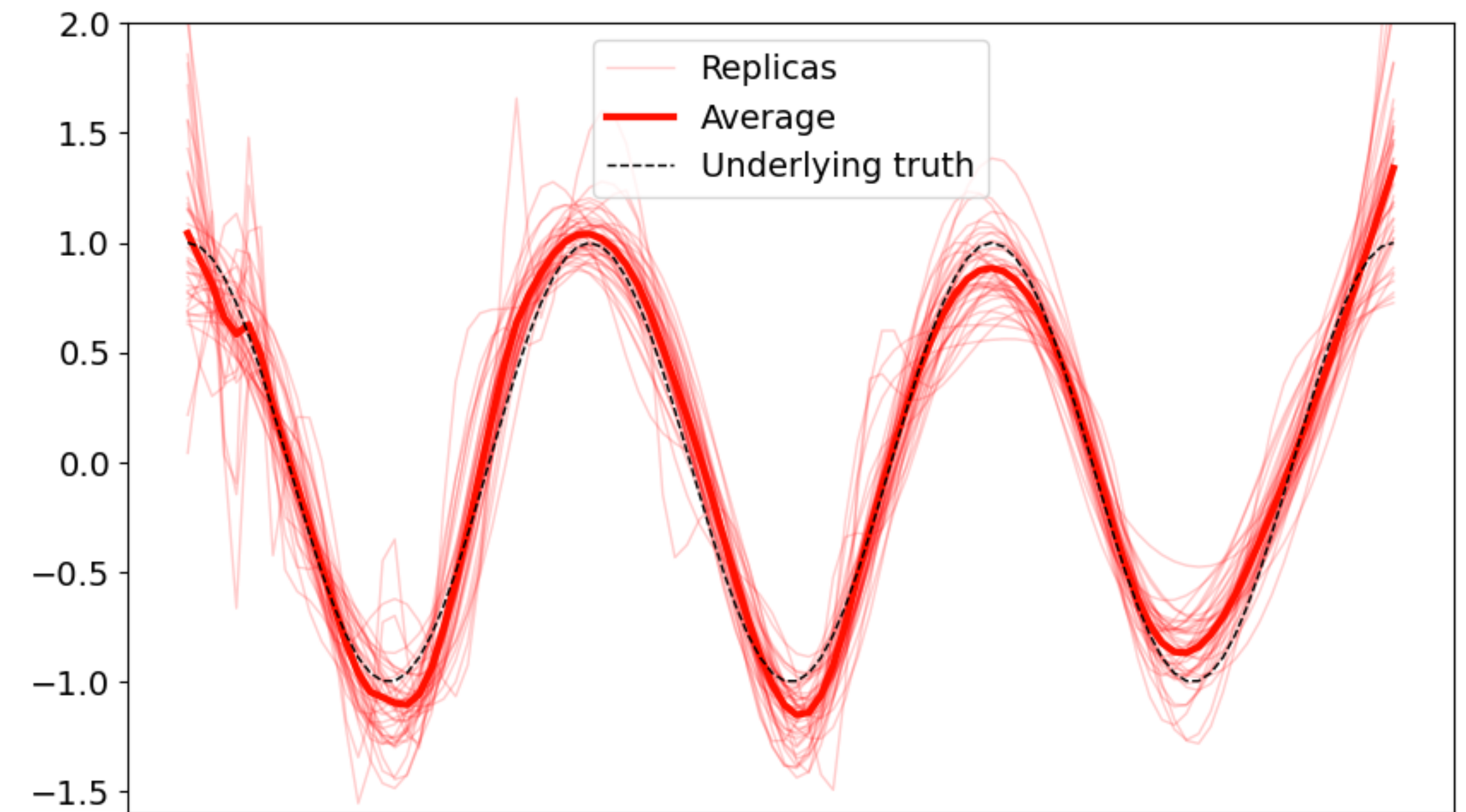
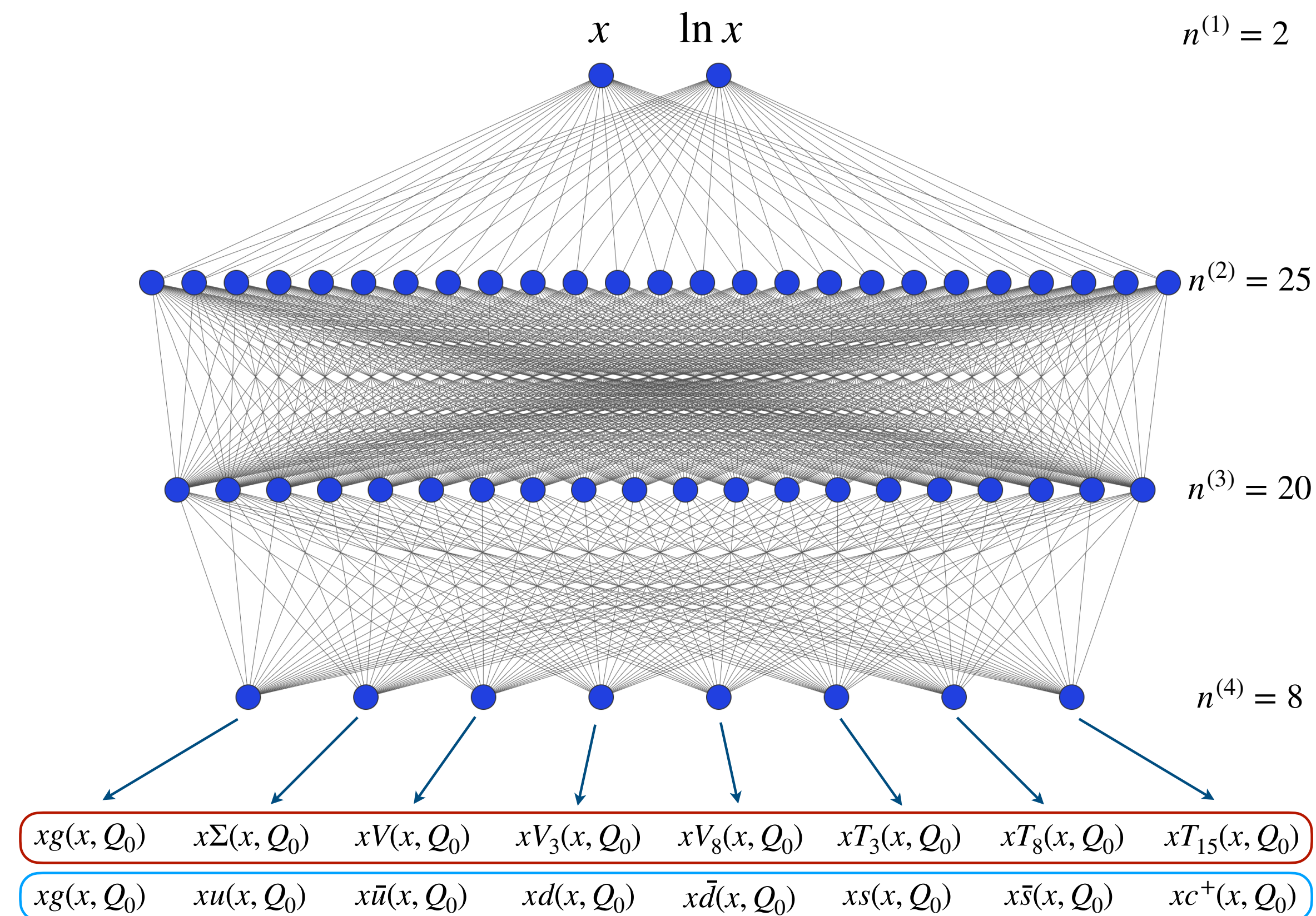
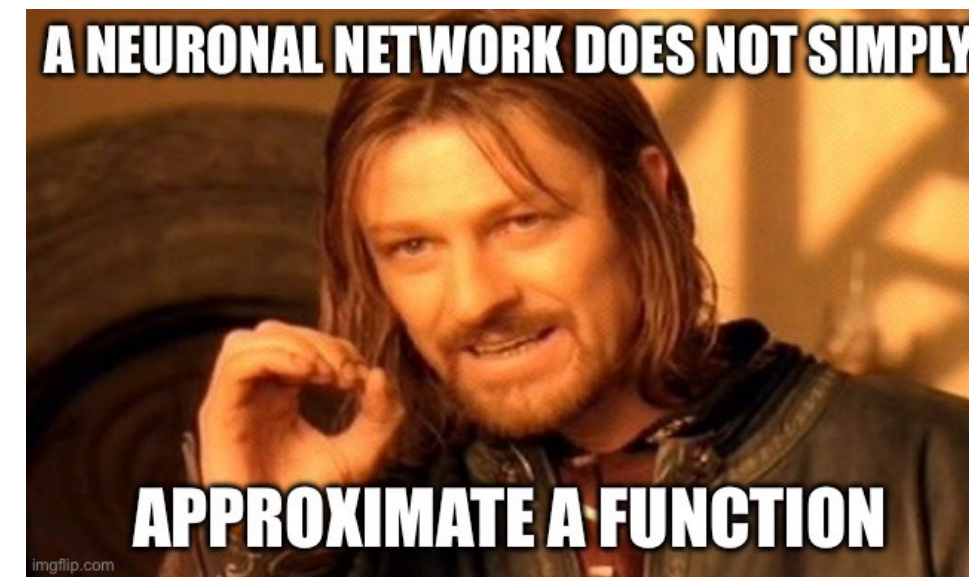
Just give them enough data (or physical constraints) to help them out



But what if I know nothing about the functional form?

Neural Networks can act as universal functional approximation.

Just give them enough data (or physical constraints) to help them out



Accelerated Fits using GPUs

PDF Fits are not good candidates for vectorization outside of the typical ML GPU-usage. However, PDF fits applications are!

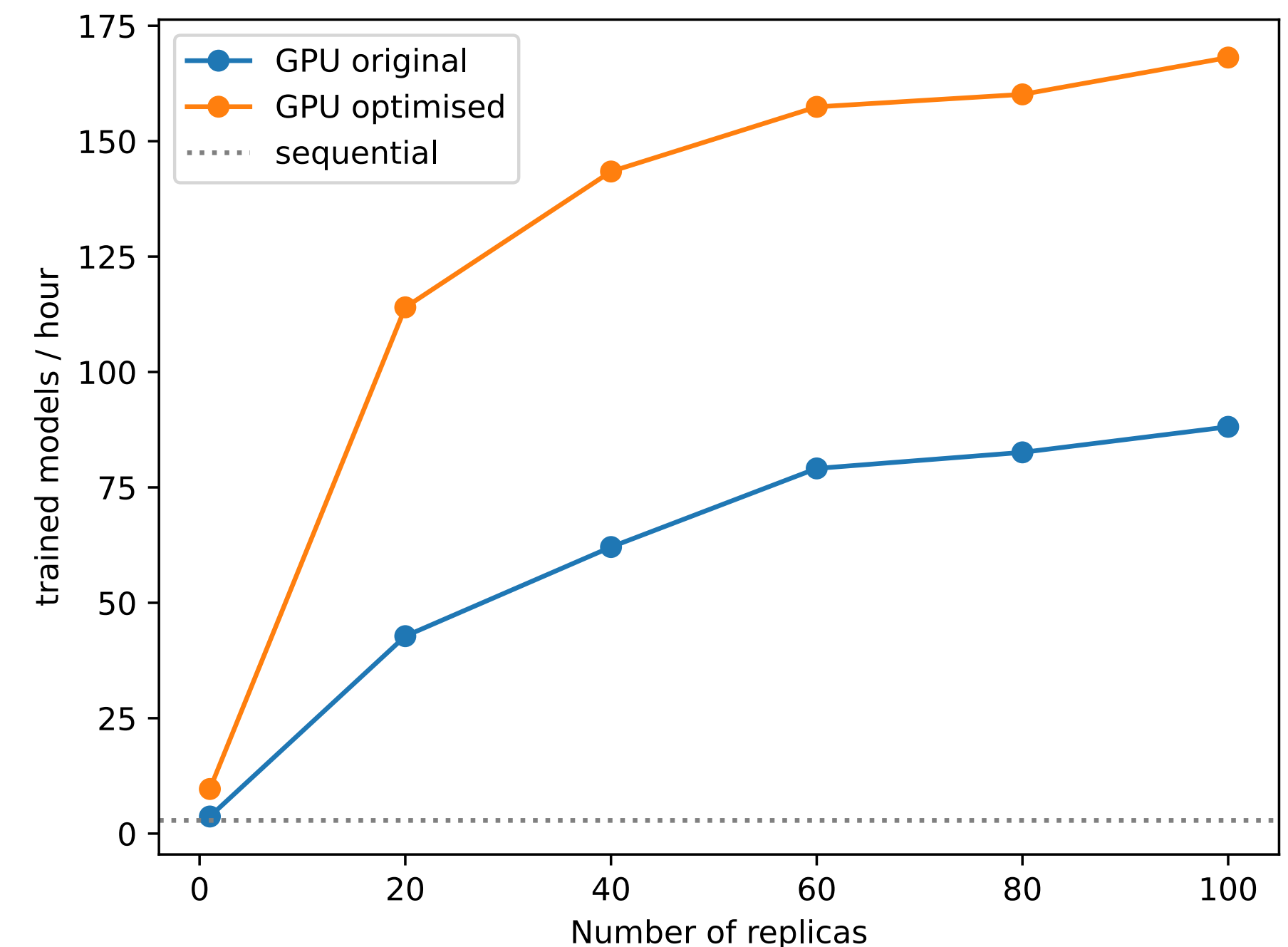
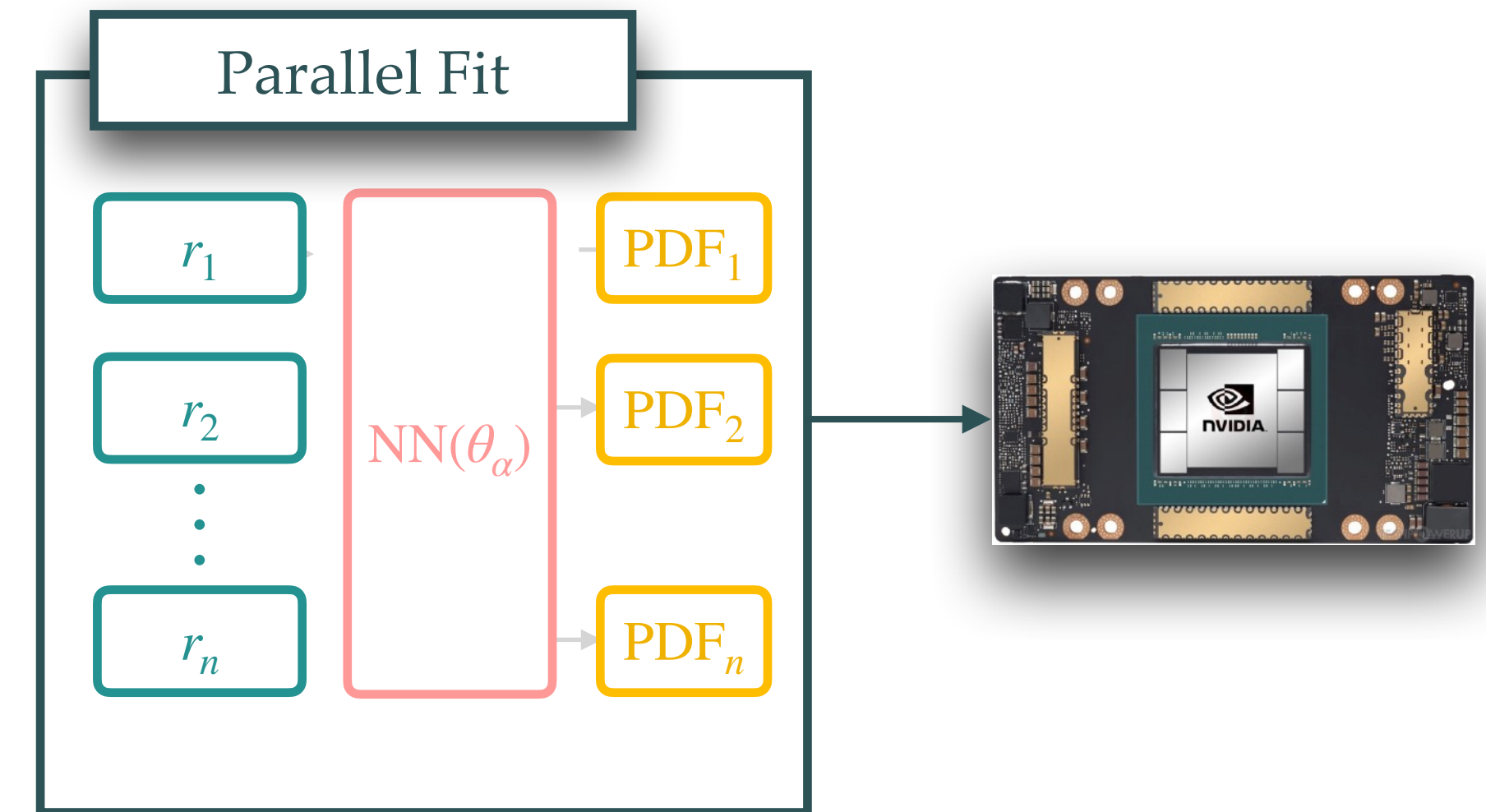
- **Simultaneous fit of multiple replicas:**
 - The code is hardware agnostic: same codebase can be used for CPU and GPU
 - Required redesigning parts of the framework in order for memory-heavy object (GBs of data and predictions!) to be shared among replicas and models
- **⇒ We can now run hundreds of replicas and models at once on a GPU in the time it would take to run a single replica**
- **Distributed asynchronous scans**
 - Different fits can share a single database for a scan of parameters (e.g., simultaneous fits of SM parameters)
 - First example implementation for a hyperparameter scan

Hyperparameter Optimisation in Deep Learning Models with Ensemble Methods and Application to Proton Structure Fits - [hep-ph] [2410.16248](#)

JCM, A. Jansen, G. van Oord, T. Rabemanjara, C. Rocha, J. Rojo, R. Stegeman

Hyeroptimisation algorithm for the next generation of PDF determinations: ensemble regression with an unbiased selection model

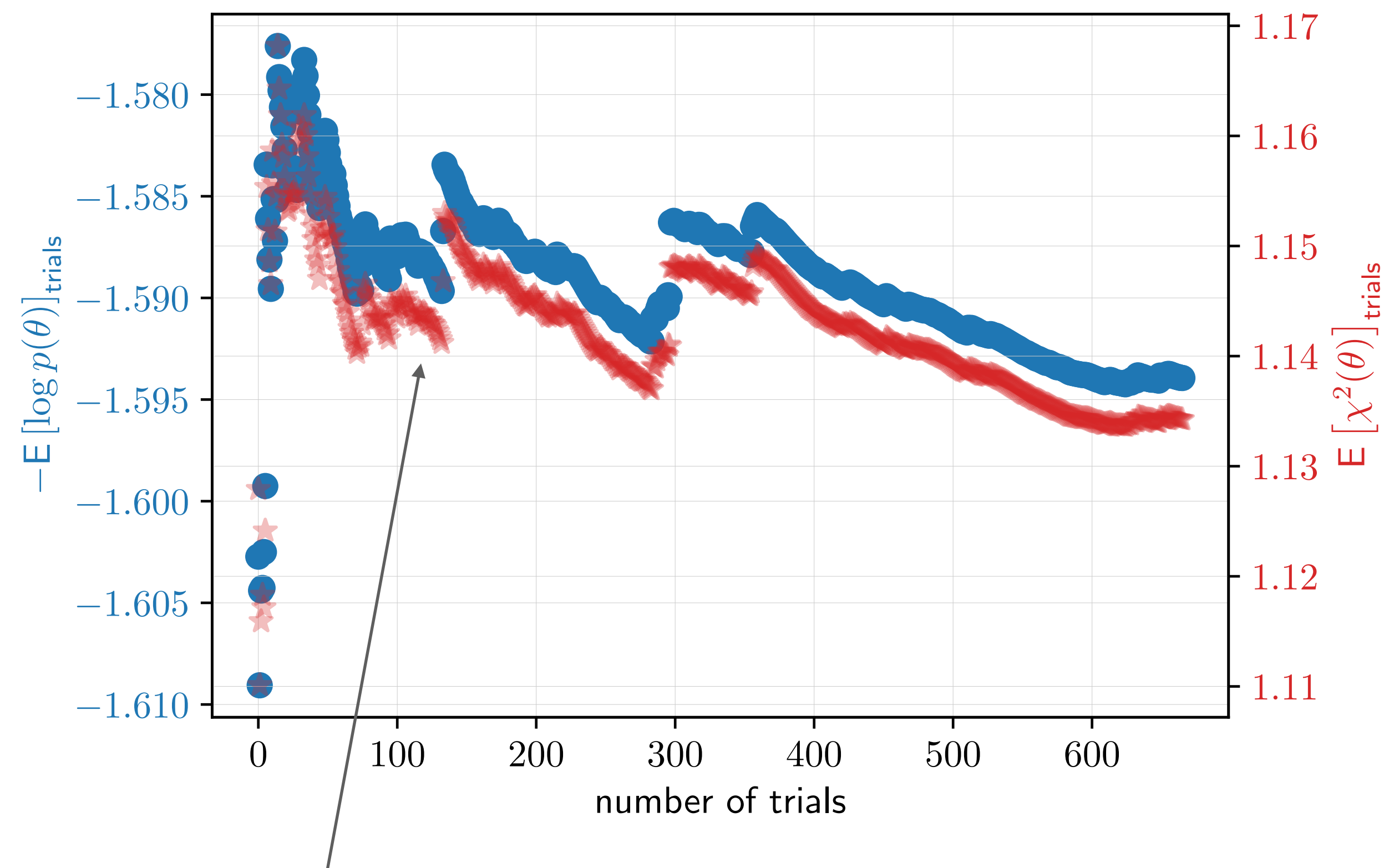
[hep-ph] [2605.31482](#) - **JCM**, T. Giani, T. Rabemanjara



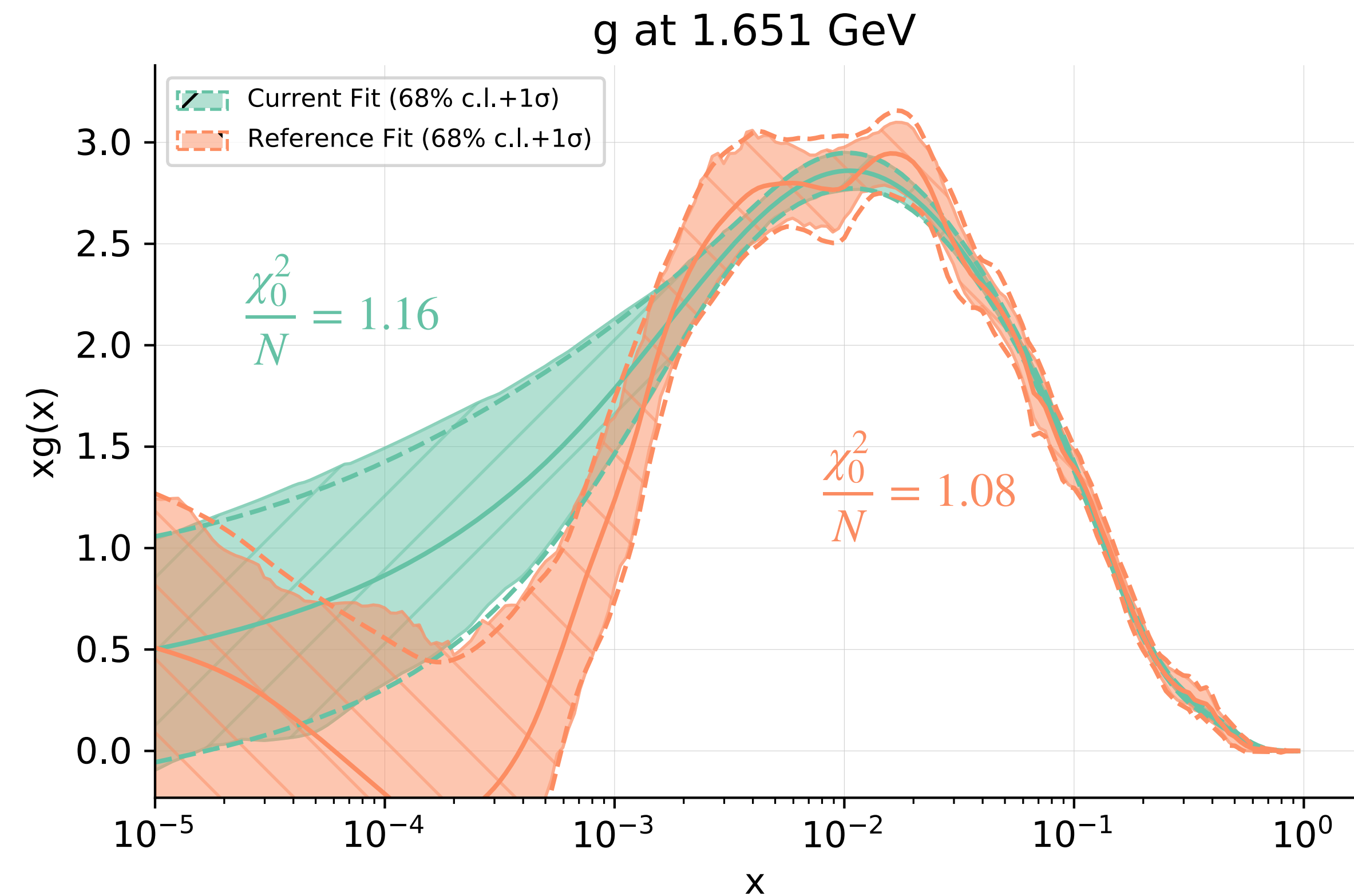
With great power comes great responsibility

Getting a χ^2 many units below the nominal NNPDF4.0 (~ 1.16) is relatively “easy”, but that doesn’t mean it is a good fit.

A thermalisation phase is necessary to avoid falling into overfitting configurations.



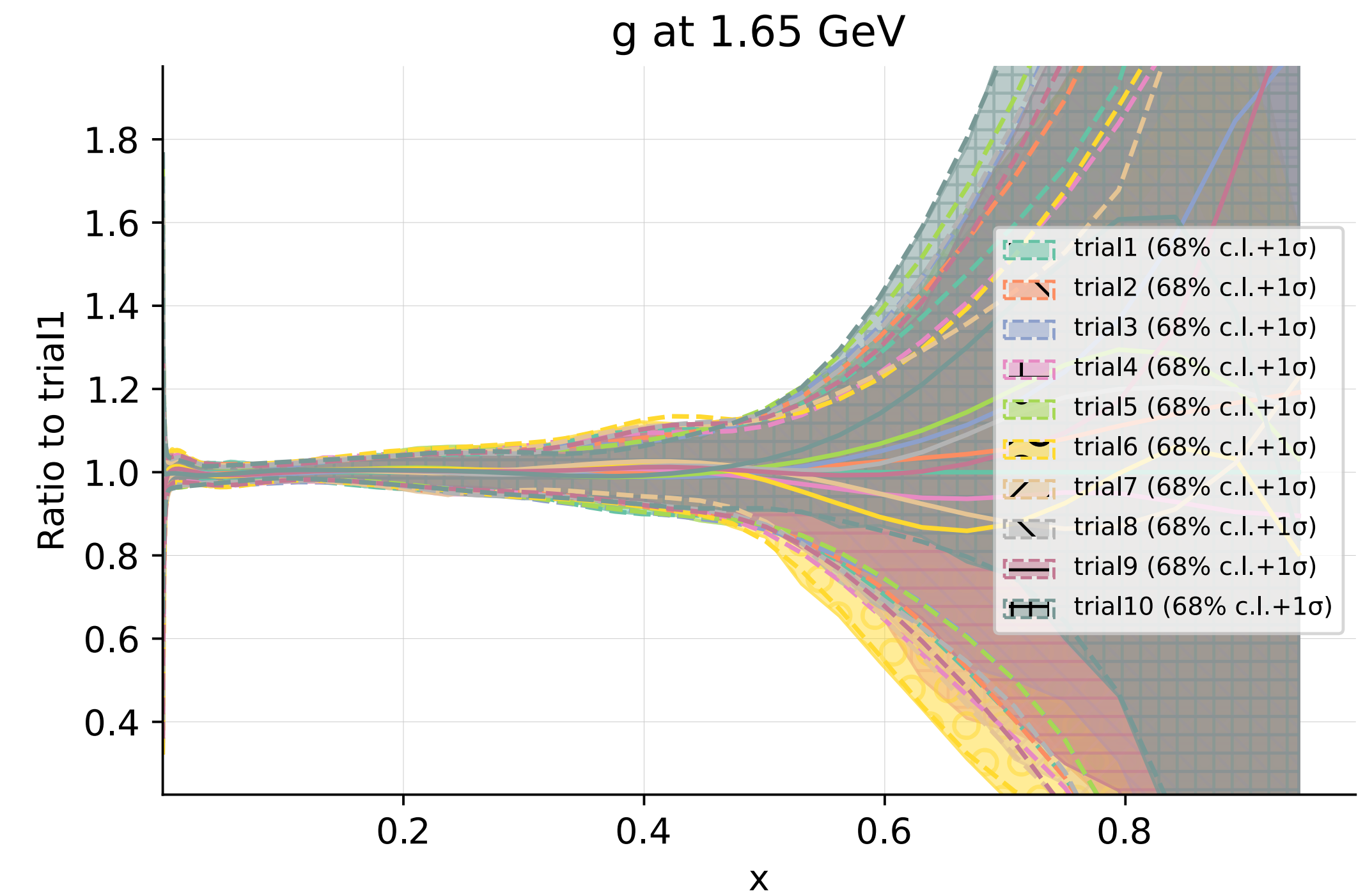
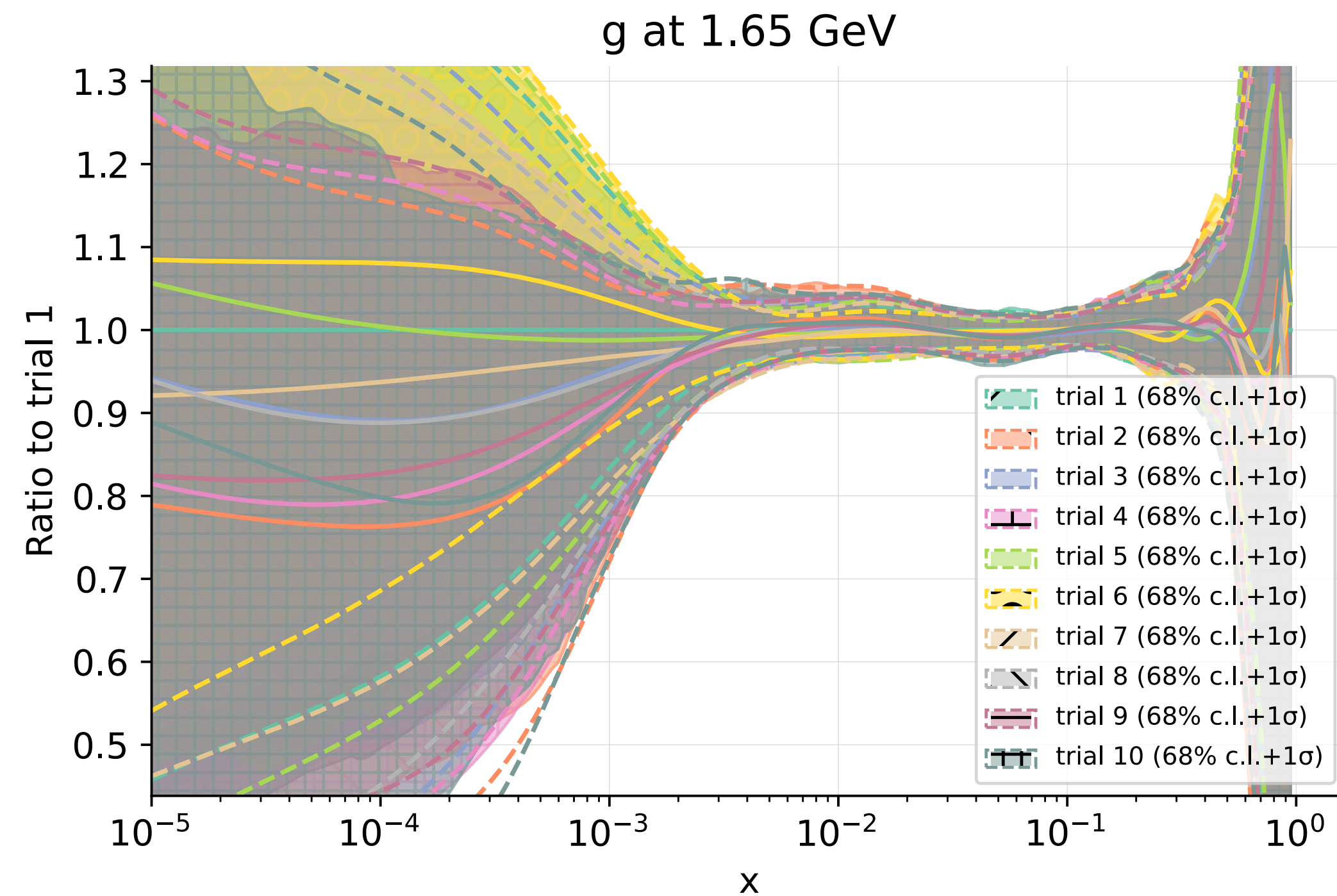
Each of these points ~200 replicas



The precision follows the data

When we put everything together we find that:

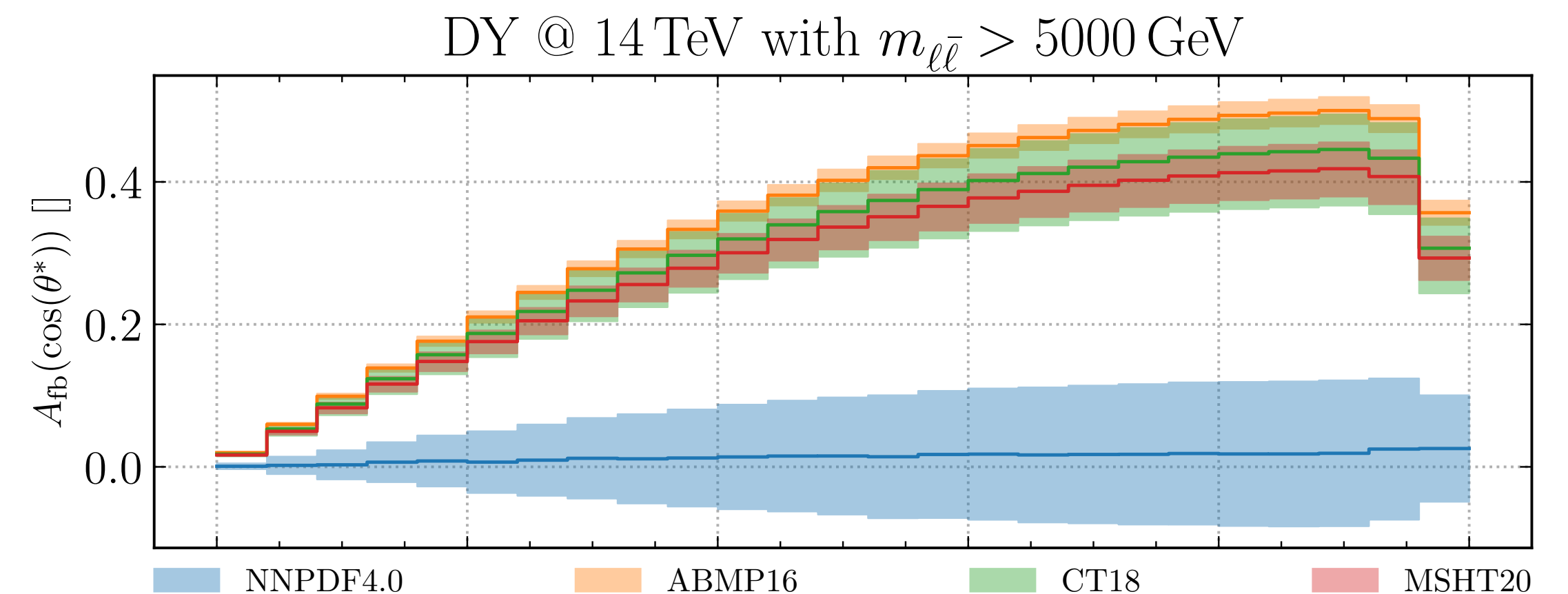
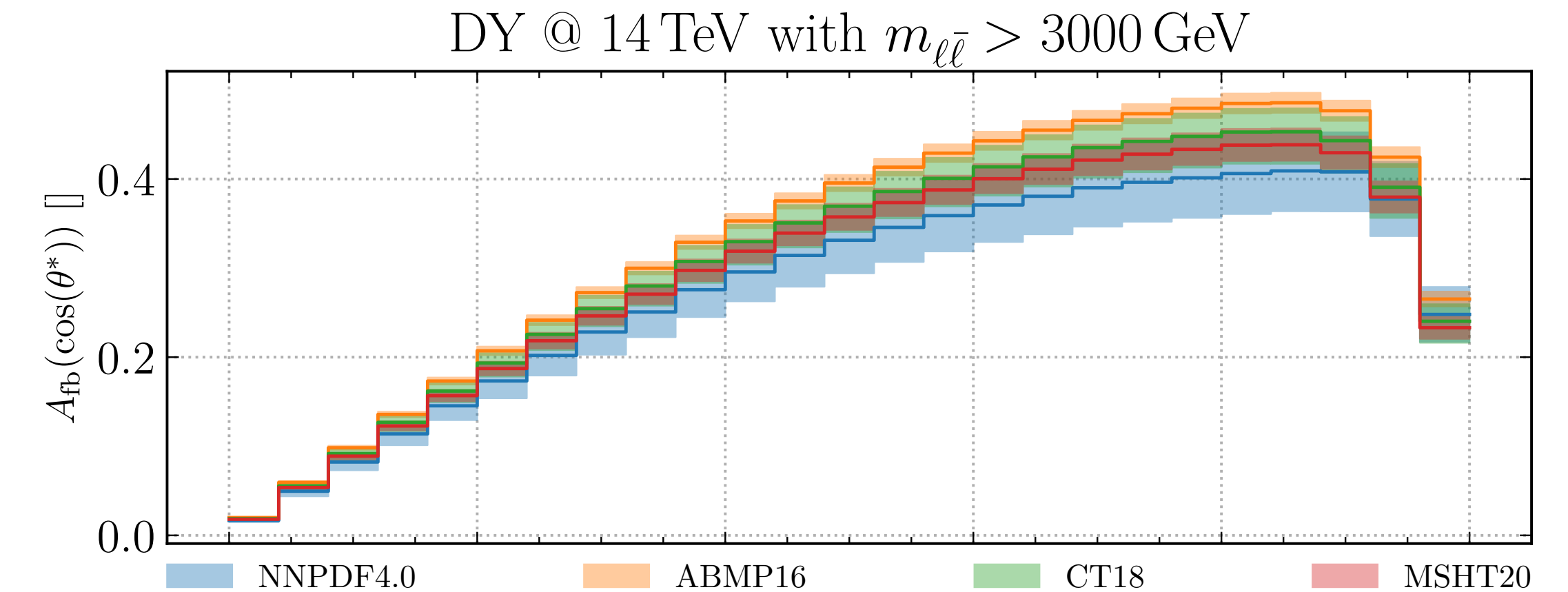
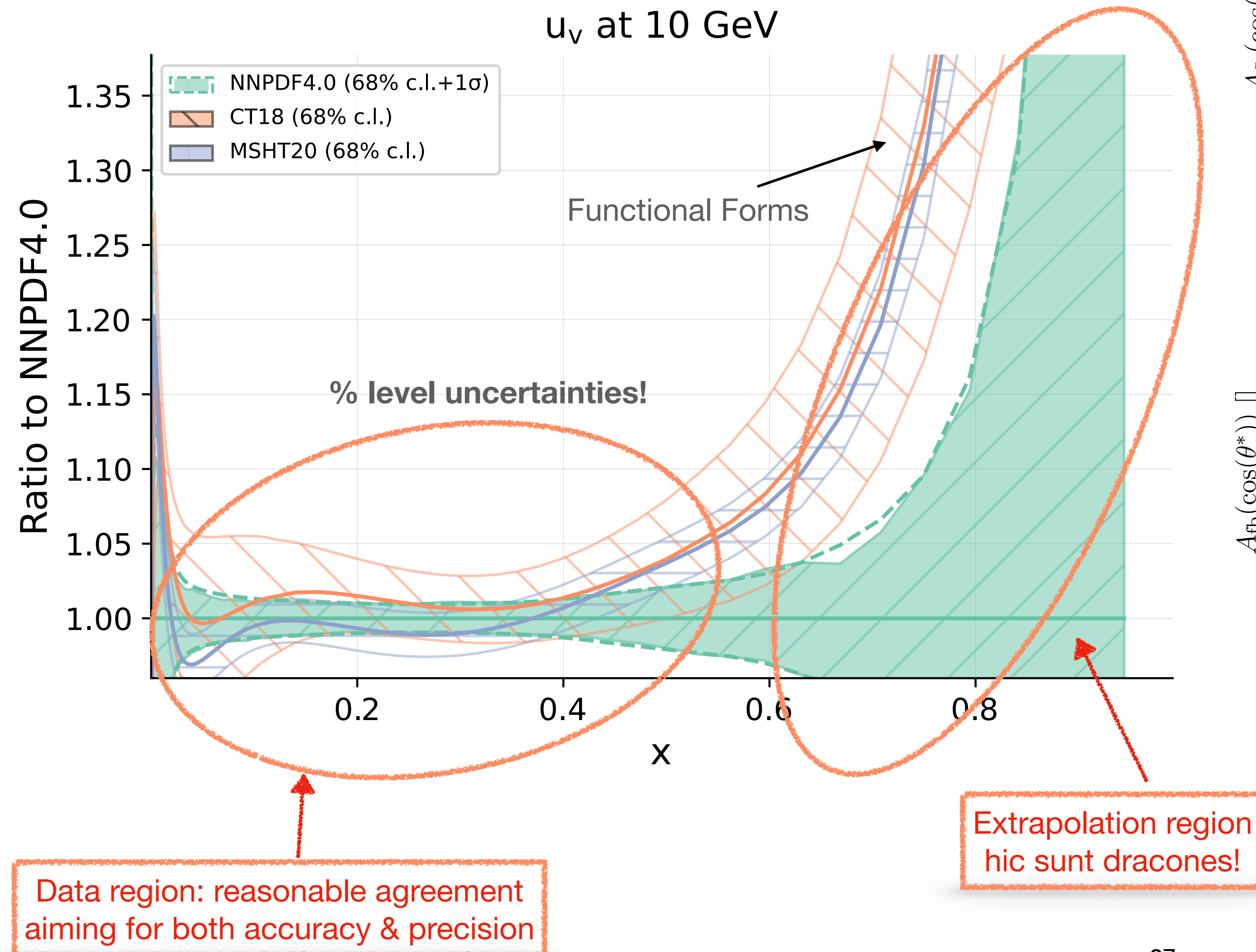
- ✓ The regions in which data is abundant, all models collapse to provide similar answers.
- ✓ Where data is scarce, sampling in the space of models provides extra components to the PDF uncertainty.
- ✓ We did not restrict the choice of model or even the optimization parameters, everything is data-driven.



In practice we are sampling each replica in the space of: data uncertainties, model hyperparameter and scale variations. Other options available: theory parameters or models, dataset choices, etc.

Does the precision really follow the data?

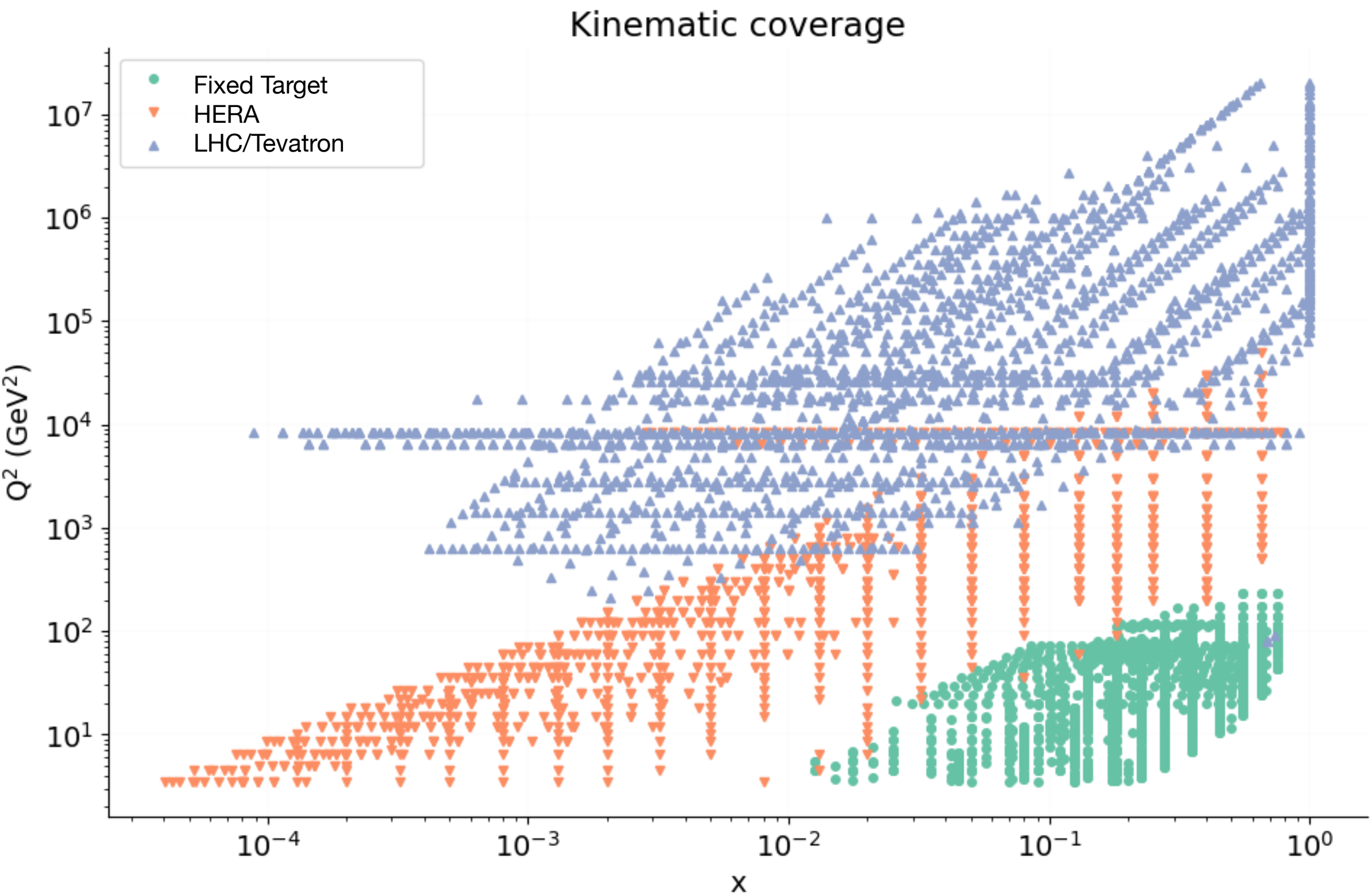
Not all regions are equally well determined, for PDFs the “data region” ends at around $x \sim 0.5$



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!

Validation and testing

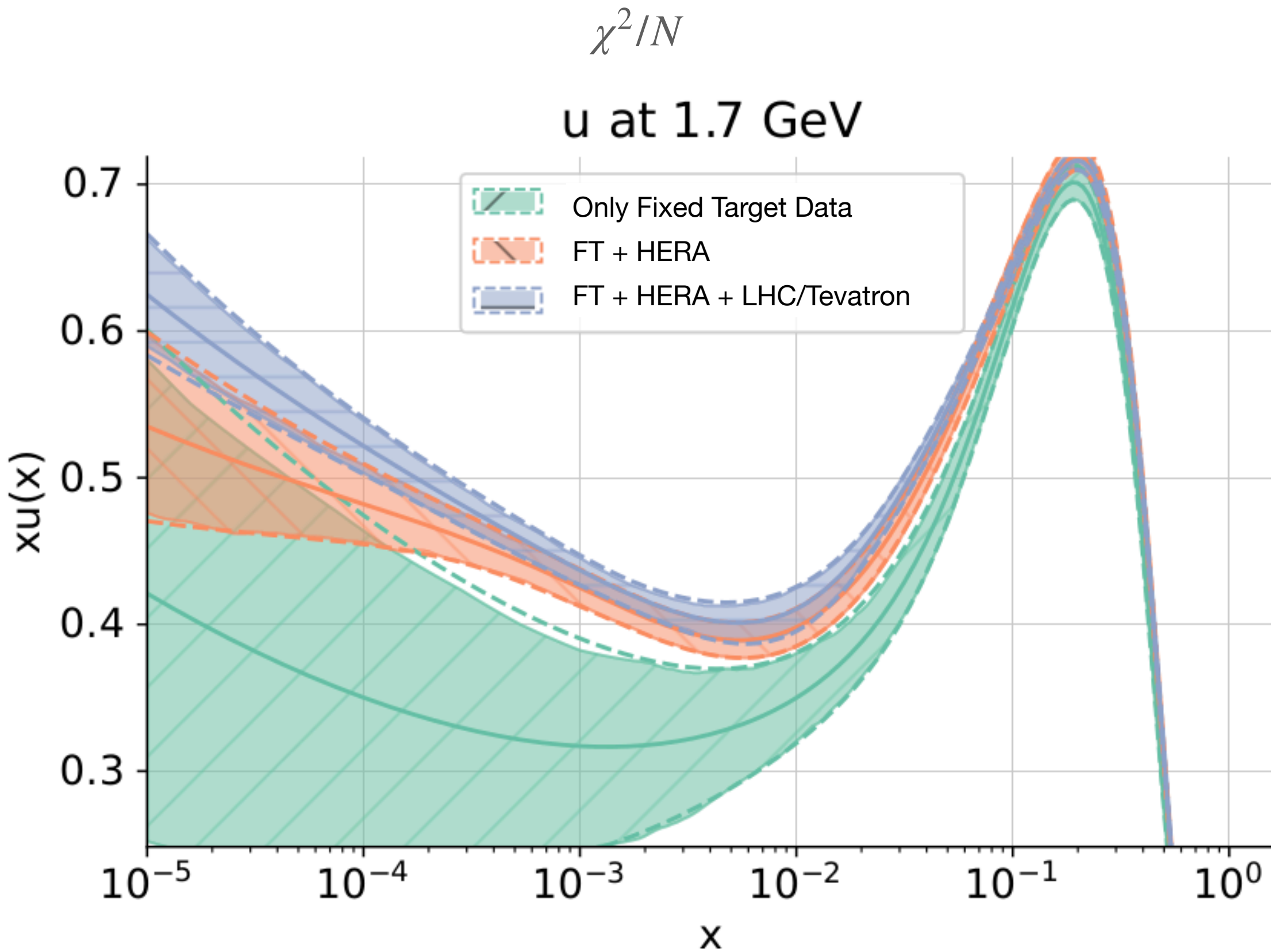
Future tests



Future tests of parton distributions
JCM, E. Nocera, S. Forte - [hep-ph] 2103.08606

Ignoring the PDF uncertainties. Looking only at the central value.

dataset \ fit	Full	FT + HERA	Fixed Target
Fixed Target	1,06	1,01	0,91
FT + HERA	1,20	1,21	26,1
Full	1,29	2,15	22,57

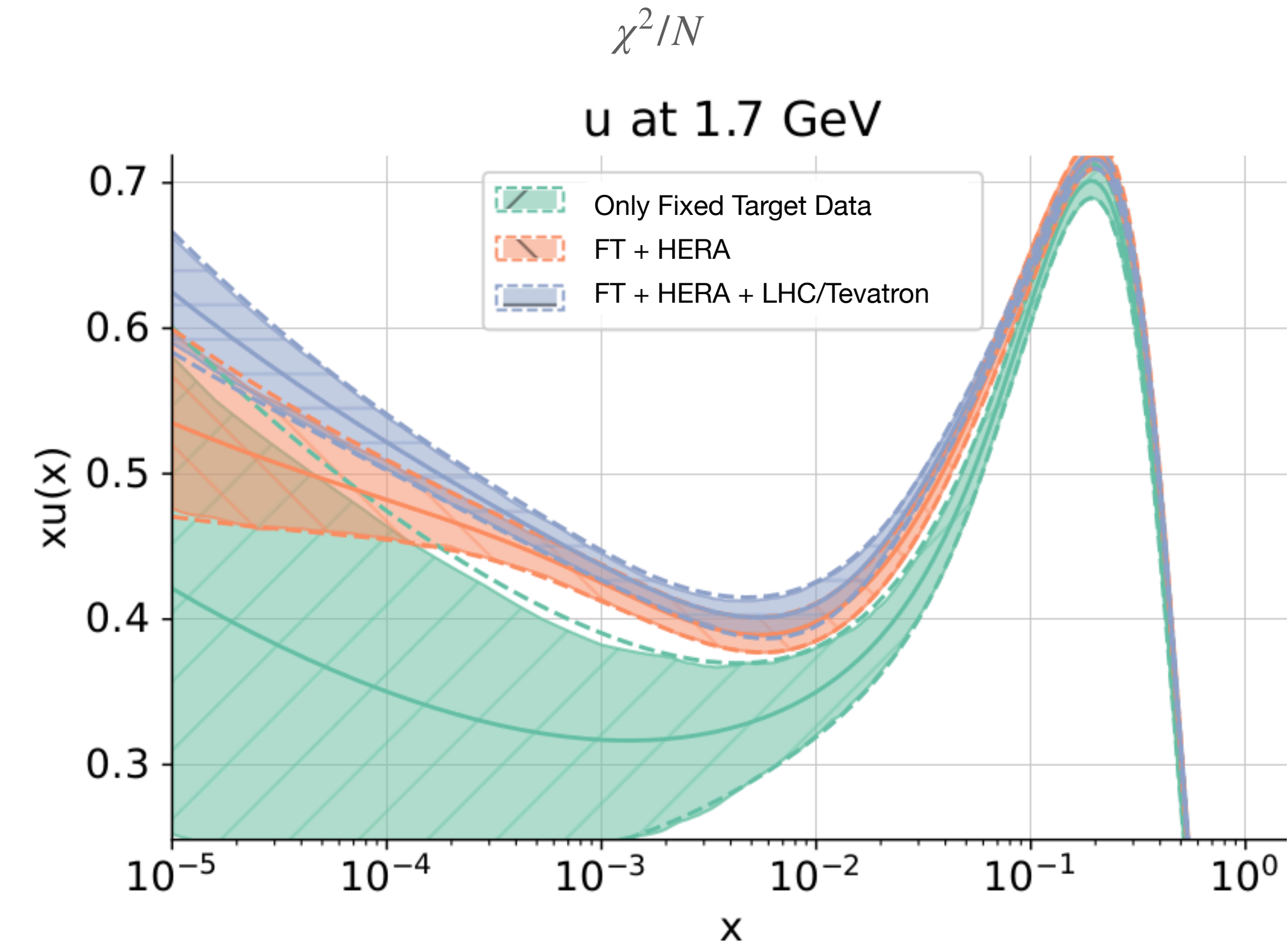
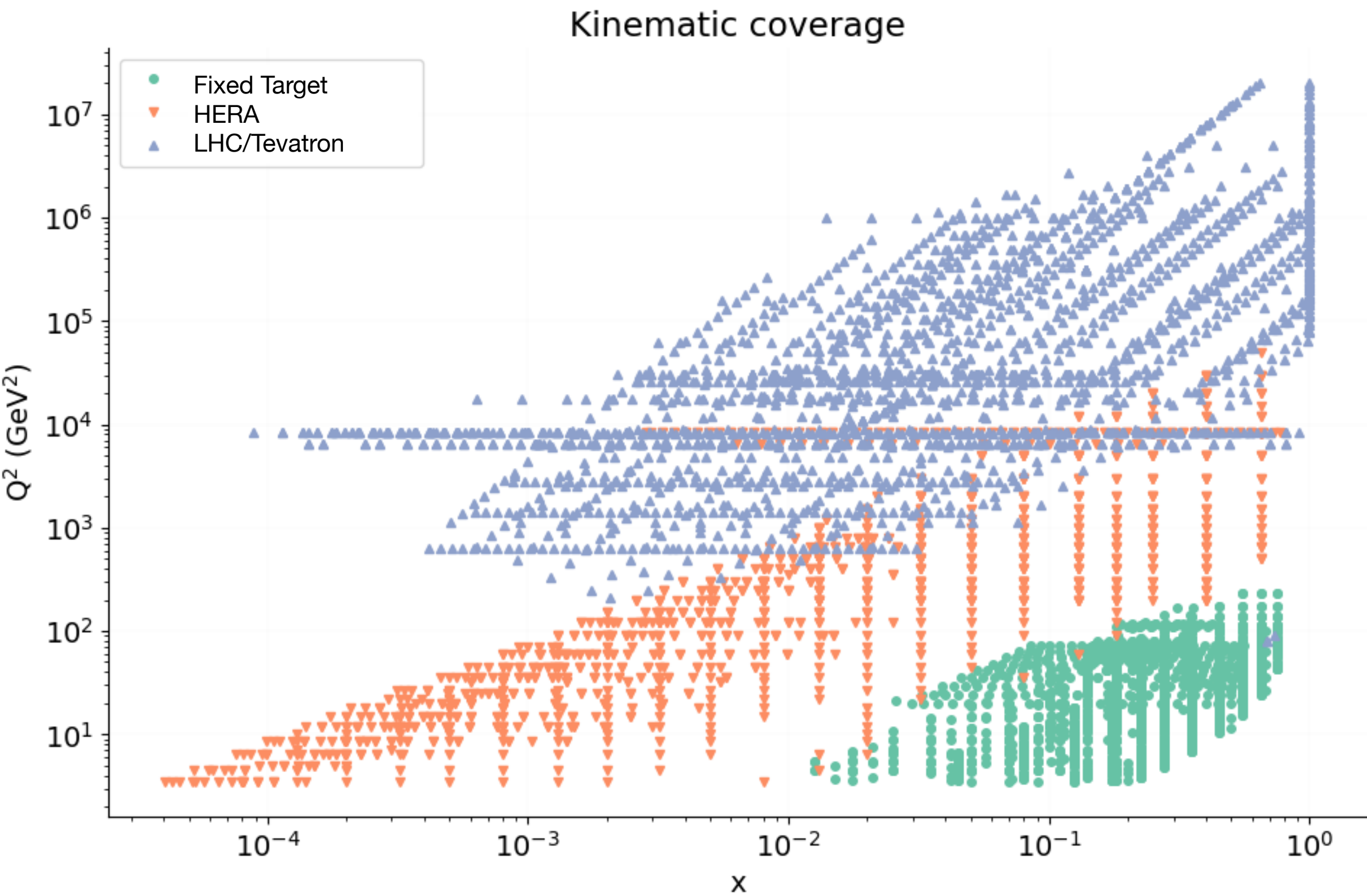


Validation and testing

Future tests

Considering the PDF uncertainties.

dataset \ fit	Full	FT + HERA	Fixed Target
Fixed Target			0,87
FT + HERA		1,18	1,22
Full	1,12	1,30	1,38



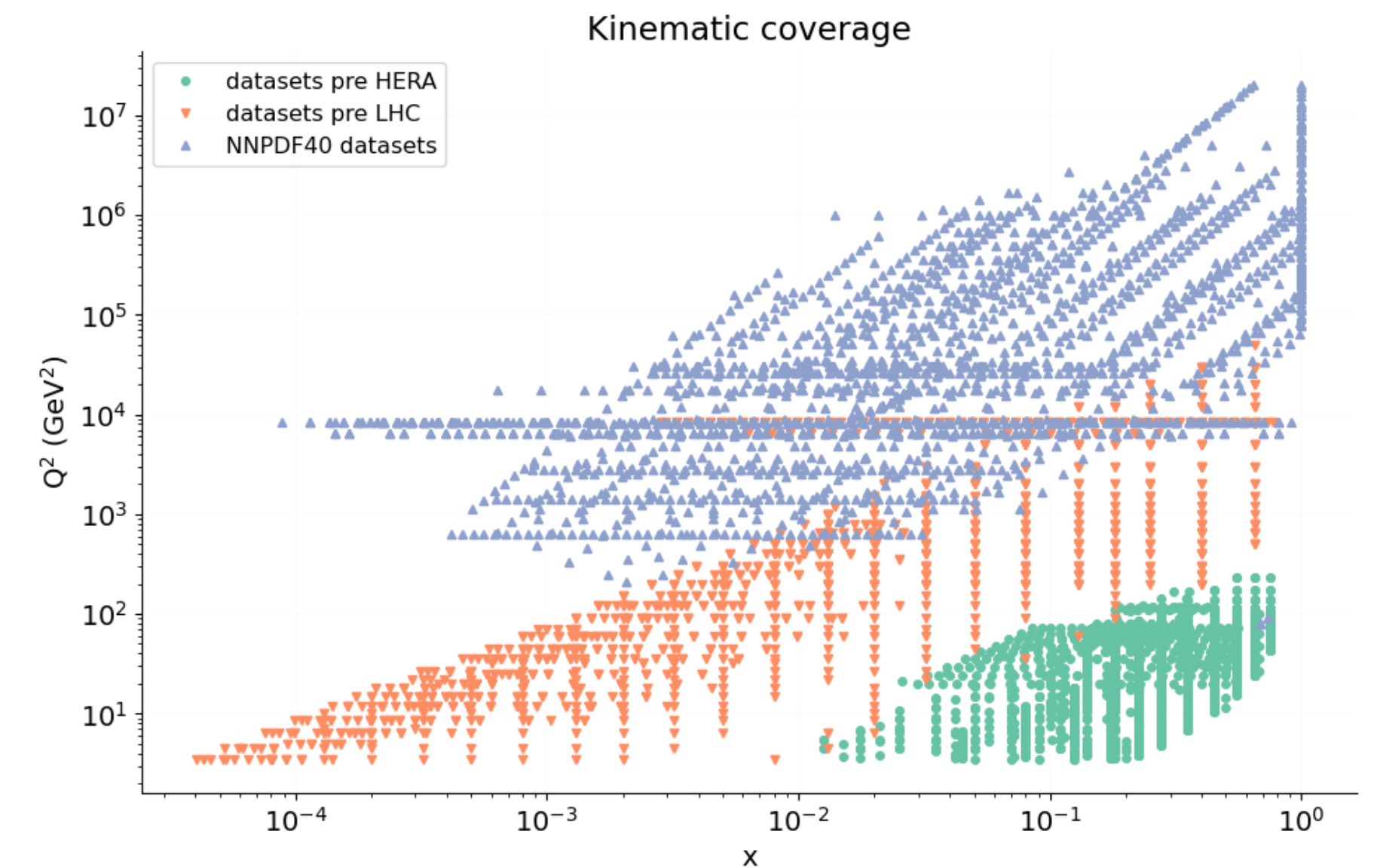
Future tests of parton distributions
JCM, E. Nocera, S. Forte - [hep-ph] 2103.08606

Introduction

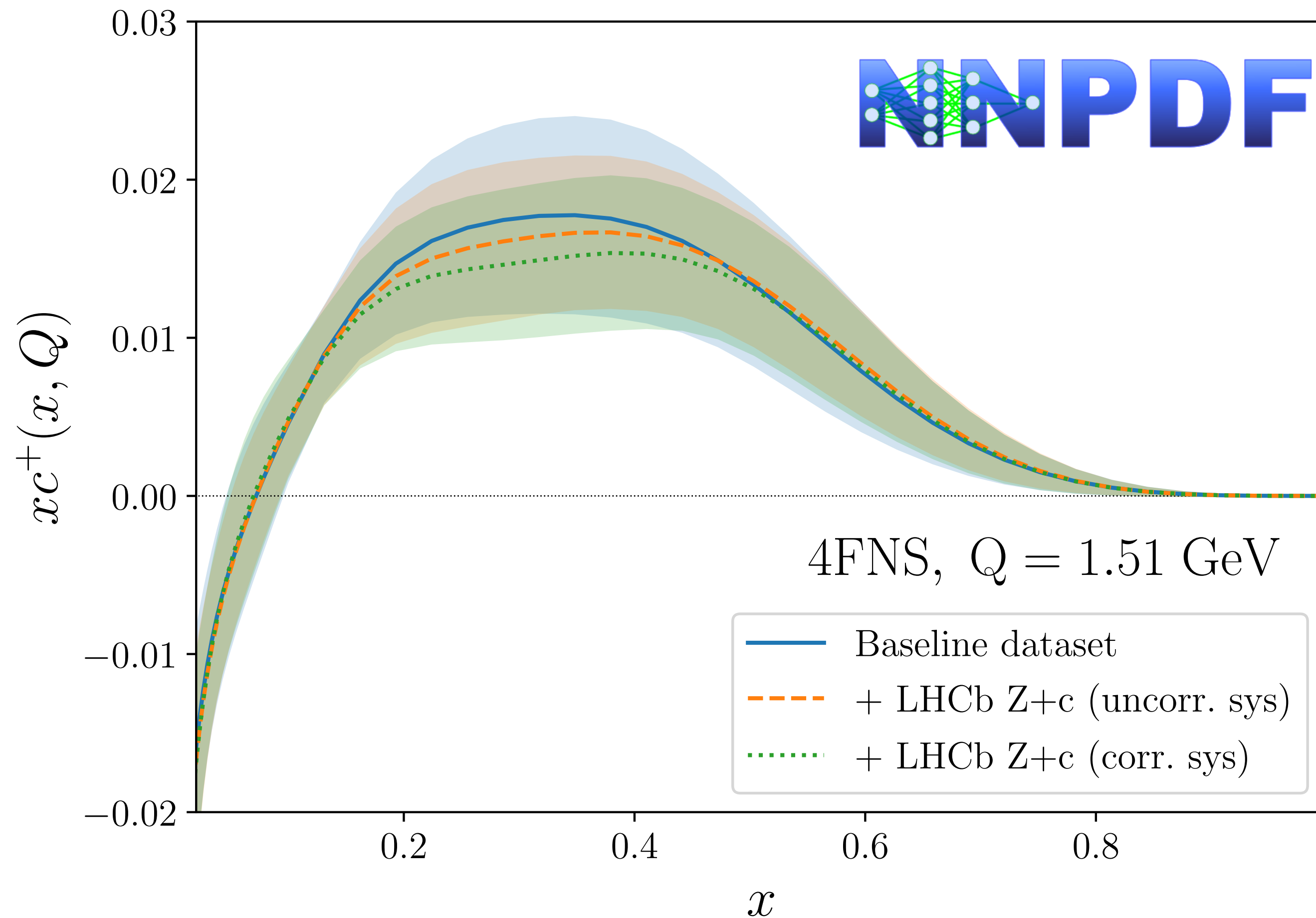
A tale of shrinking uncertainties

The NNPDF methodologies: neural networks
and ensemble regression

Results & Conclusions



Intrinsically charming



Evidence for intrinsic charm quarks in the proton
NNPDF Collaboration — [hep-ph] [2208.08372](#)

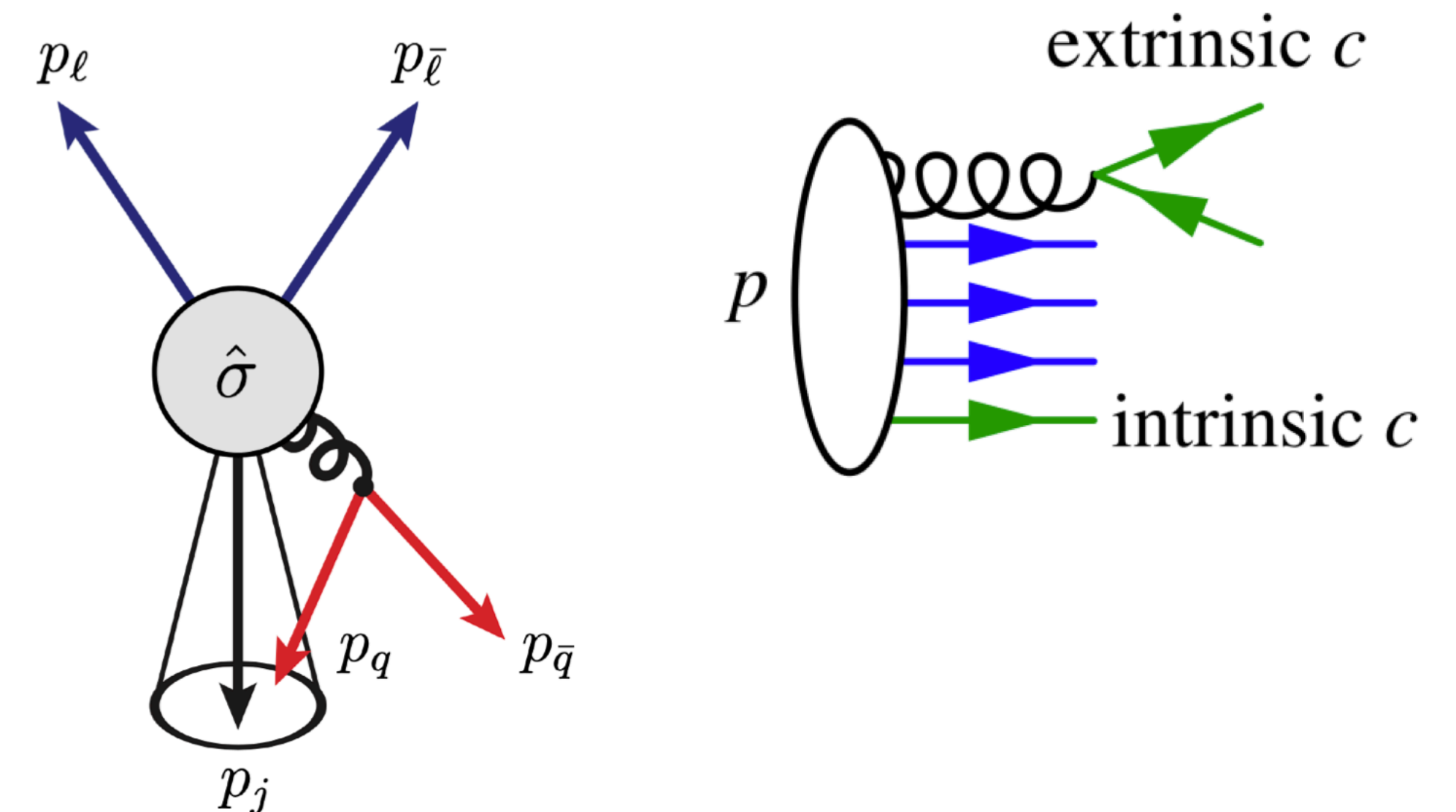
NNPDF is the only collaboration which fits charm by default, i.e., $c^+ = c + \bar{c} \neq 0$ at the fitting scale which means the contribution is not limited to DGLAP evolution

Open challenges:

- Better grasp of MHOU
- Improved jet algorithms

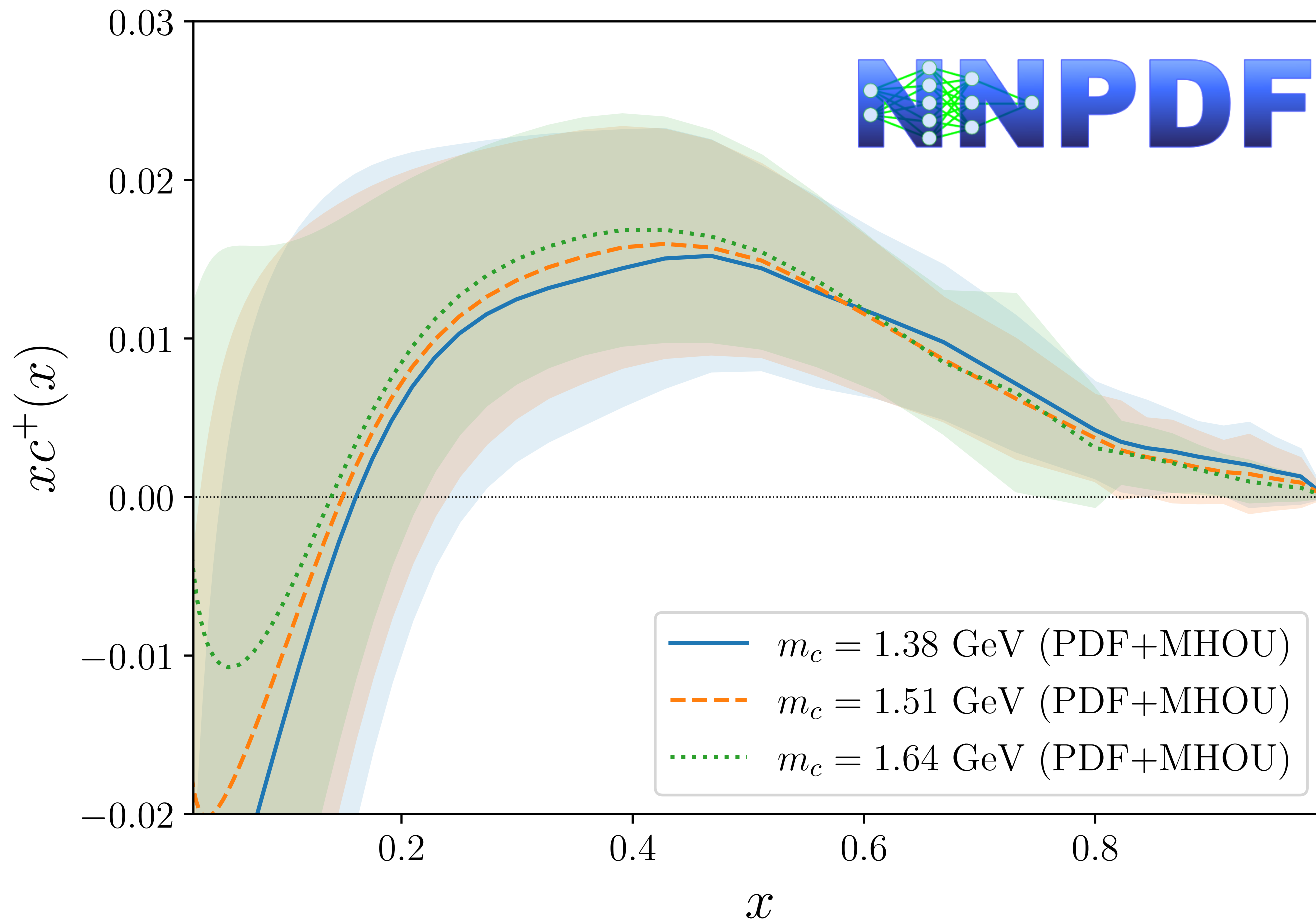
in order to match data and predictions...

collinear-safe jet algorithms need to be used



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

Intrinsically charming



Evidence for intrinsic charm quarks in the proton
NNPDF Collaboration — [hep-ph] [2208.08372](#)

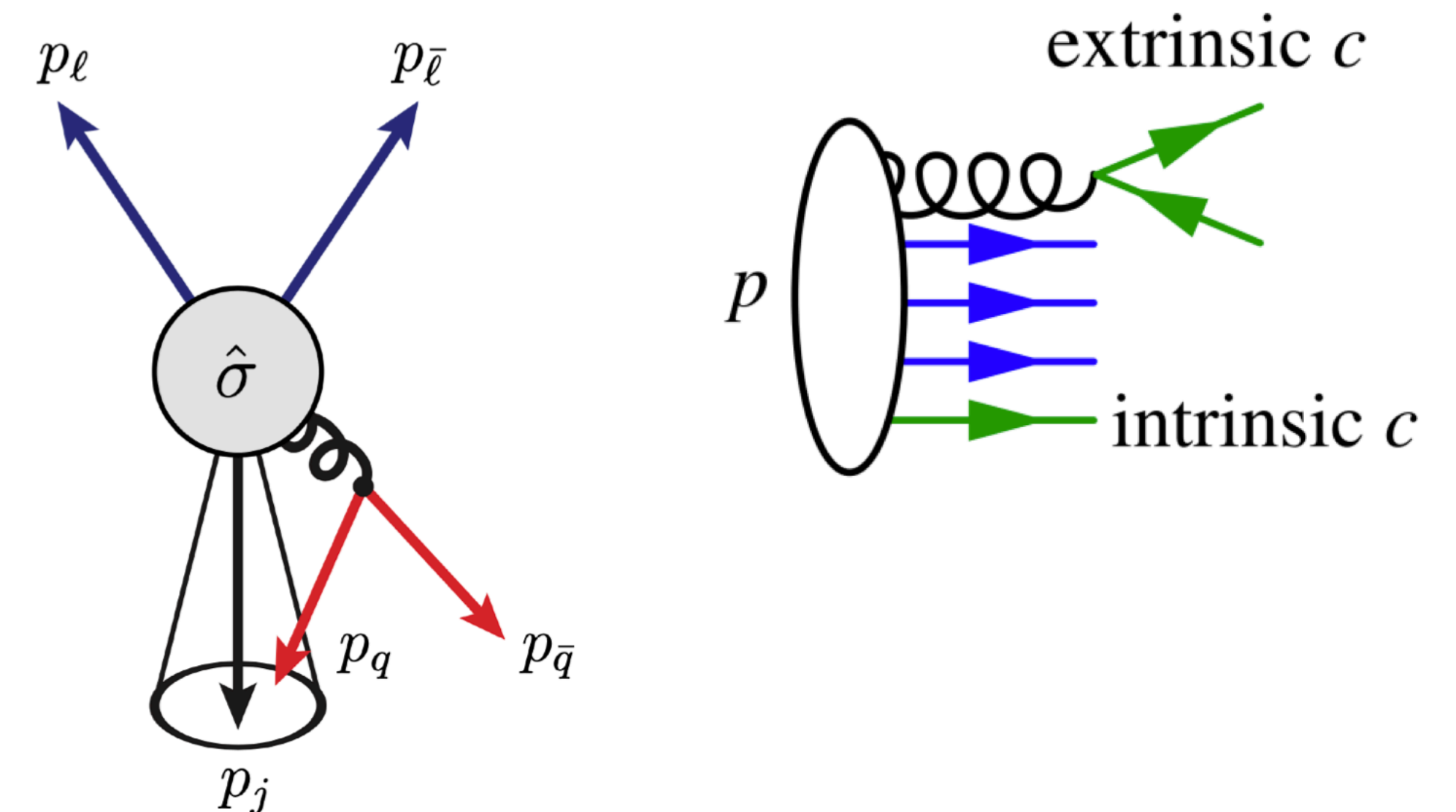
NNPDF is the only collaboration which fits charm by default, i.e., $c^+ = c + \bar{c} \neq 0$ at the fitting scale which means the contribution is not limited to DGLAP evolution

Open challenges:

- Better grasp of MHOU
- Improved jet algorithms

in order to match data and predictions...

collinear-safe jet algorithms need to be used



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

Missing Higher Orders and towards N3LO

The first step towards a full N3LO determination is to be able to faithfully account for missing ingredients in the theoretical calculations: missing higher order corrections, unknown splitting functions, etc.

If we propagate uncertainties only from the data we just have half of the story, as fixed-order predictions also contain uncertainties, e.g.,

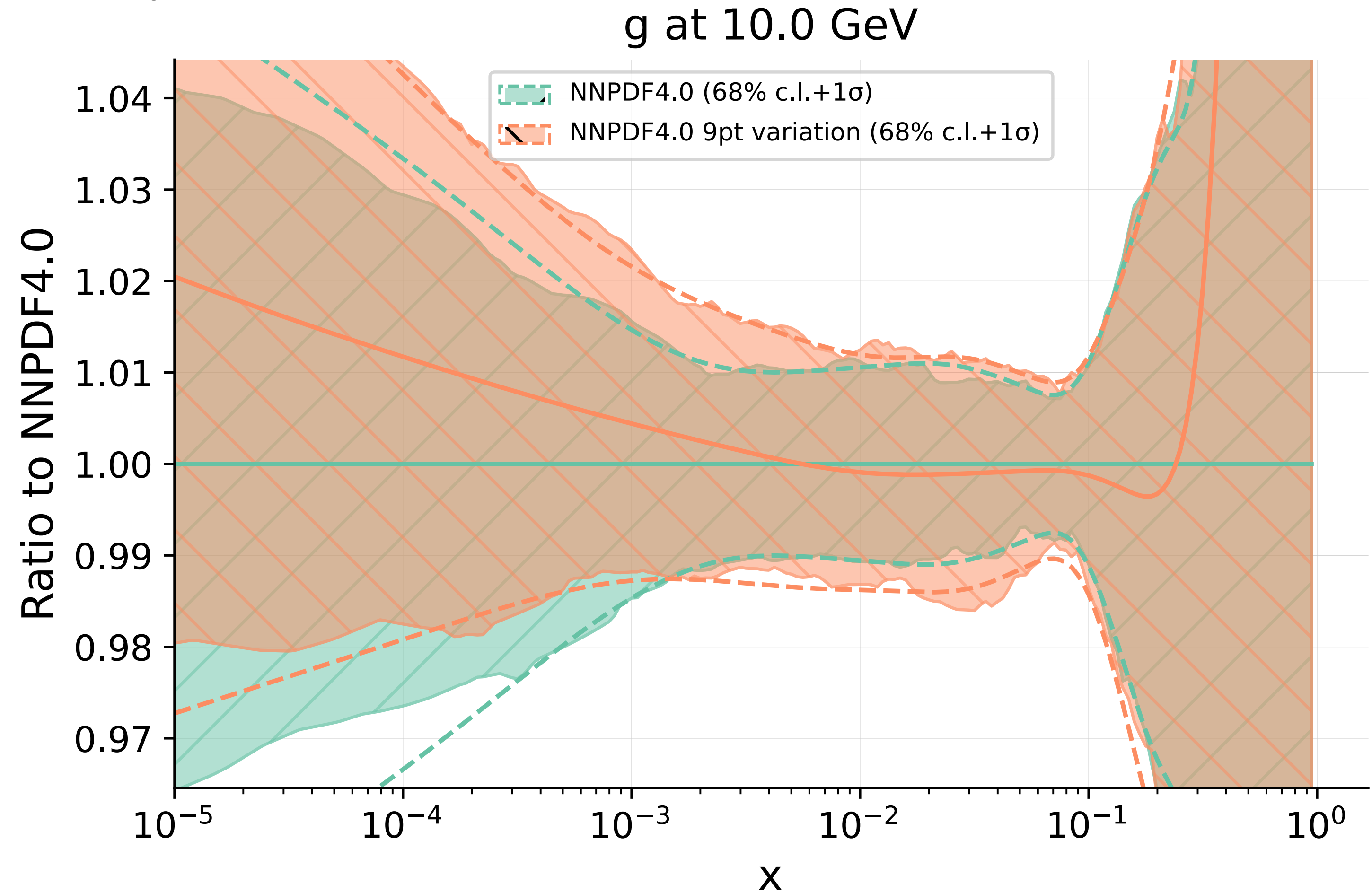
$$\sigma_{NNLO} = \sigma_0 + \alpha_s \sigma_1 + \alpha_s^2 \sigma_2 + \mathcal{O}(\alpha_s^3)$$

We exploit the spurious dependence on unphysical scales (renormalization, factorization) which is kept when we truncate the perturbative series to estimate missing higher orders.

Theory uncertainties are included in the fit by constructing a “theory covariance matrix” which plays along well with experimental uncertainties:

$$cov_{ij} = cov_{ij}^{\text{exp}} + cov_{ij}^{\text{th}}$$

Determination of the theory uncertainties from missing higher orders on NNLO parton distributions with percent accuracy
NNPDF collaboration - [hep-ph/2401.10319](https://arxiv.org/abs/hep-ph/2401.10319)



N3LO: the next frontier

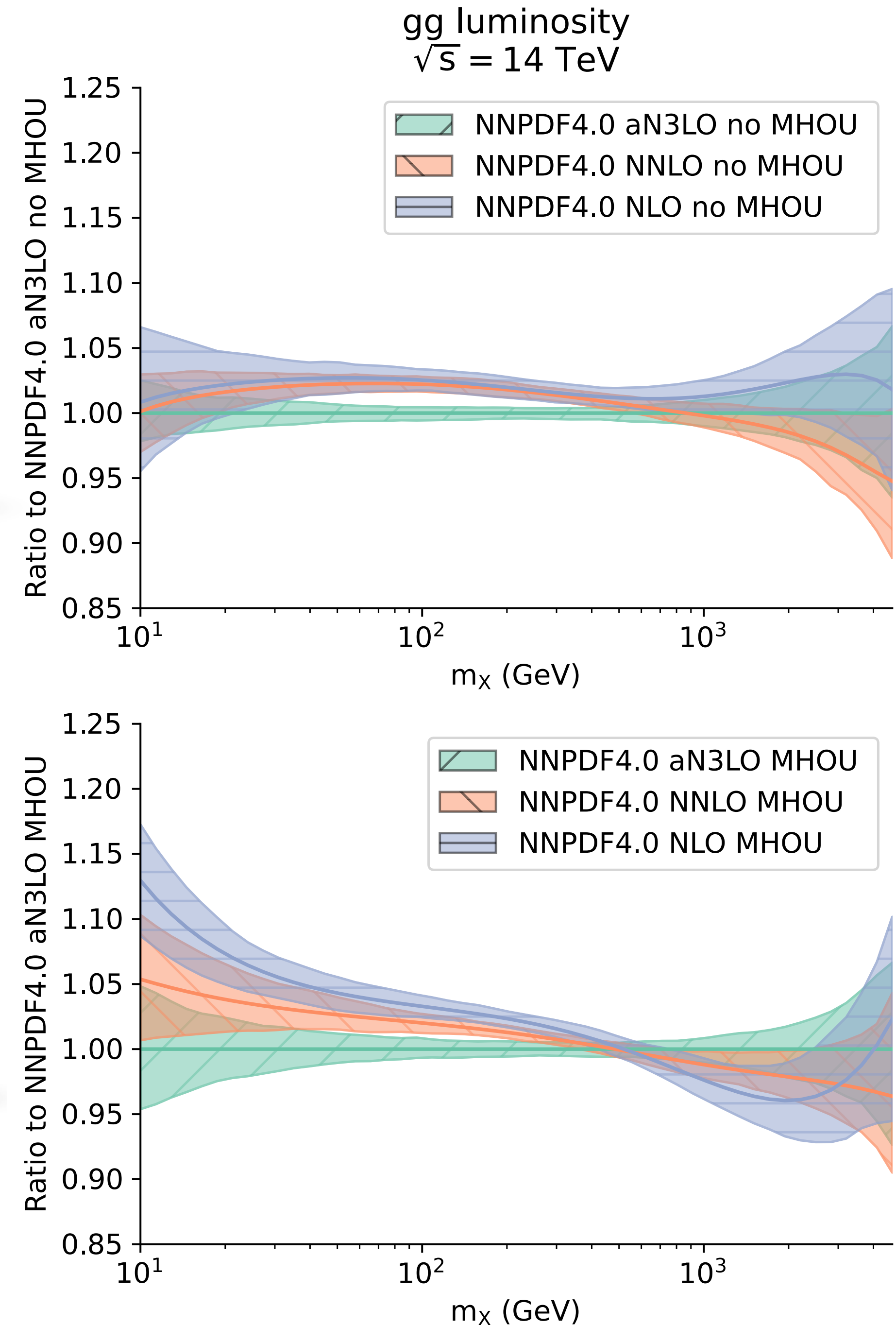
Approximated N3LO results are only available at the moment by the NNPDF and MSHT collaborations, with CTEQ in preparation.

- Order-by-order converge improved w.r.t. NNLO
- Good agreement between collaborations' results, despite using very different strategies for the approximation.
- Necessary to match the order of theoretical calculations.

NB: results are not *exact* N3LO but rather **aN3LO** (for approximated, although, I'd argue, for *almost*)

- Splitting functions not fully known, but known in enough limits to build a reasonable approximation. Details and benchmarks in 2406.16188 and references therein.
- Fiducial double-hadronic predictions only accessible at NNLO.
- The effect of missing contributions improves convergence. It is estimated through scale variations for NNPDF and nuisance parameters for MSHT.

The path to N3LO parton distributions
NNPDF collaboration - [hep-ph/2402.18635](https://arxiv.org/abs/hep-ph/2402.18635)



A tasted beyond, polarized PDFs by NNPDF

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?

NNPDFpol2.0: a global determination of polarised PDFs and their uncertainties at next-to-next-to-leading order
JCM, T. Hasenack, F. Hekhorn, G. Magni, E. Nocera, T. Rabemananjara et al - [hep-ph] [2503.11814](#)

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

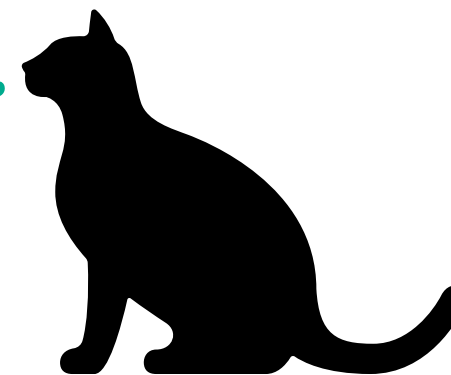
$$S_P \equiv \frac{1}{2} = \frac{1}{2} \sum_q \eta_q(Q^2) + \eta_g(Q^2) + \langle \sum_q \mathcal{L}_q + \mathcal{L}_g \rangle(Q^2)$$

$$\eta_k = \int_0^1 dx \Delta f_k(x)$$

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

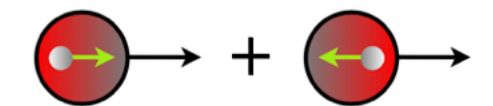
Not well constrained and compatible with 0

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



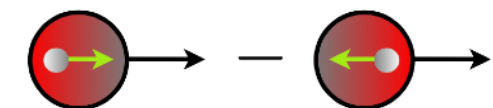
Spin averaged (usual) PDFs

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



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

What's new about 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 for access to the code from [hep-ph] 0404057

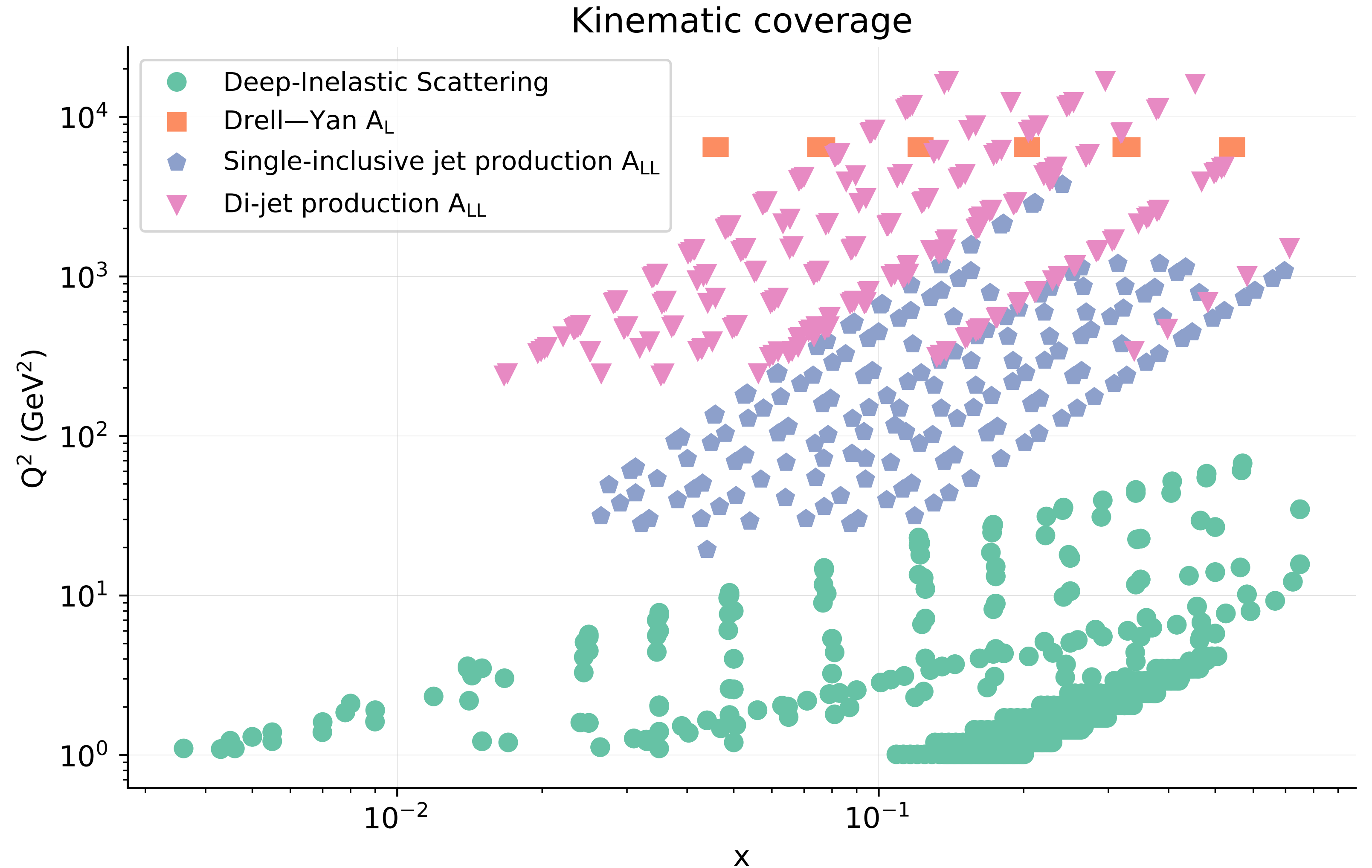
4. Renewed fitting methodology: more faithful uncertainty estimation (future tested methodology!), optimization of the hyperparameters of the Machine Learning methodology, GPU computing, open-source framework.

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:

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

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

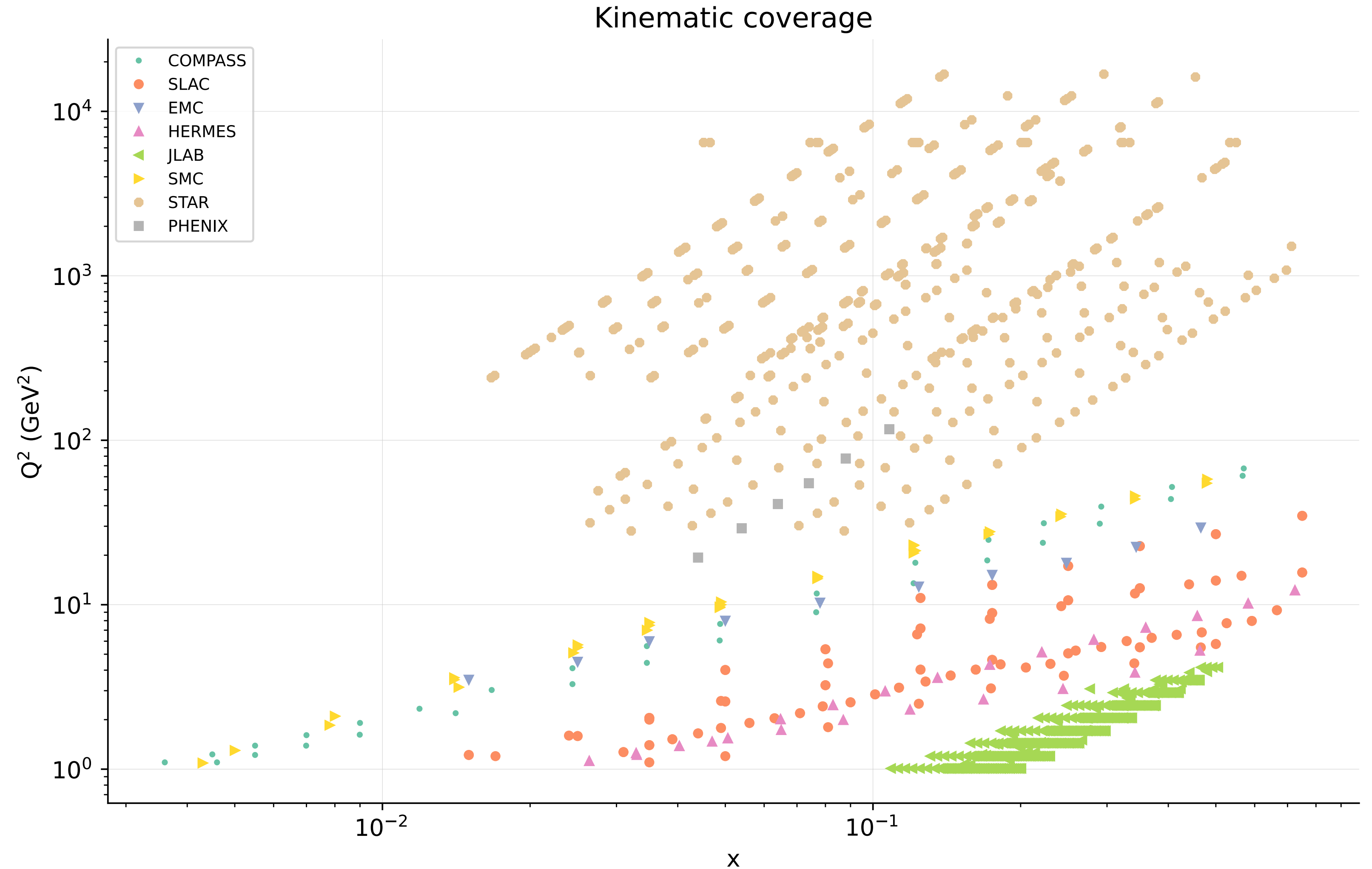


Kinematic Coverage of NNPDFpol 2.0

- 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:

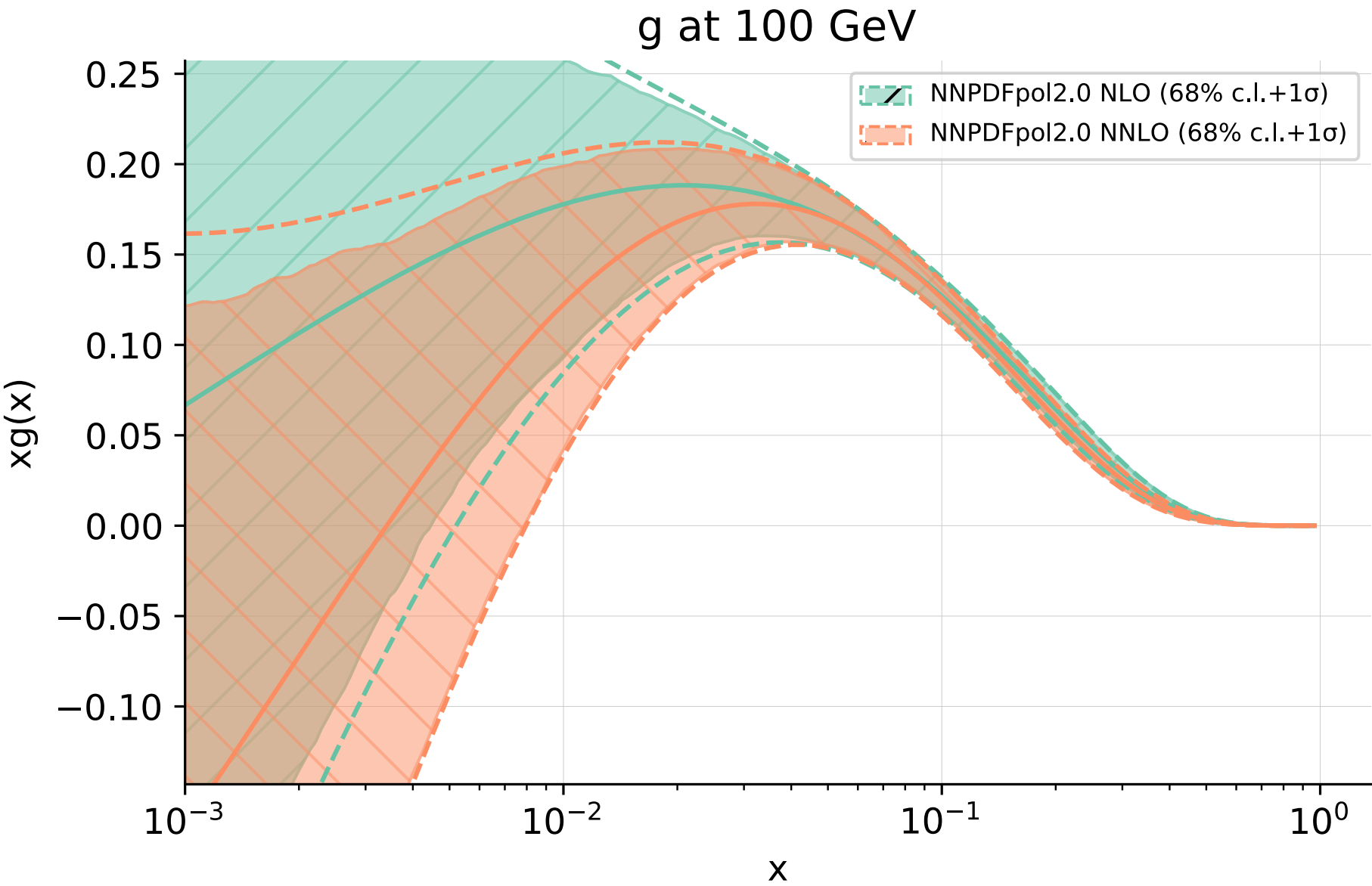
$$Q^2 \geq 1 \text{ GeV}^2$$

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

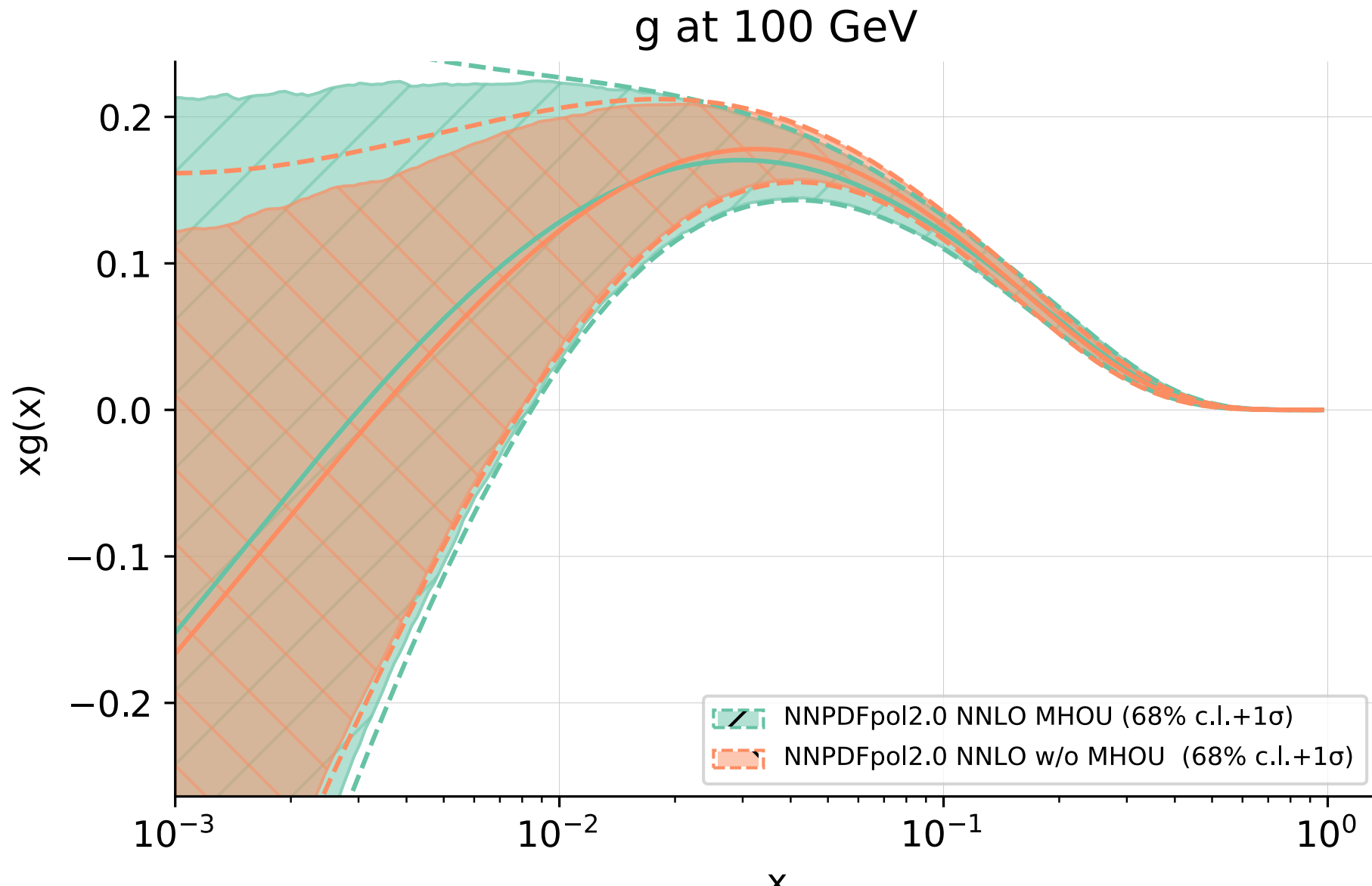


NNPDFpol 2.0: the perturbative impact

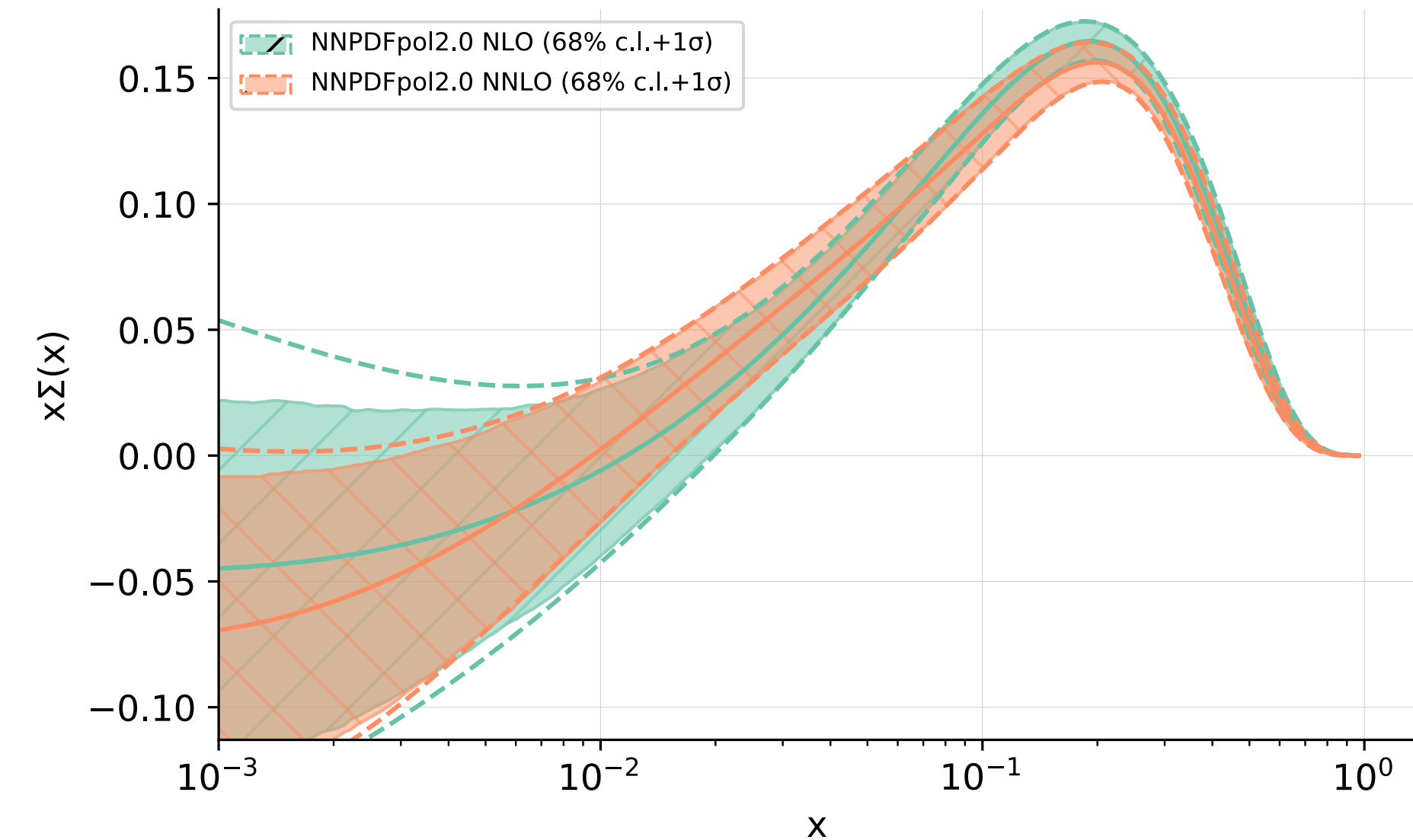
NLO vs NNLO



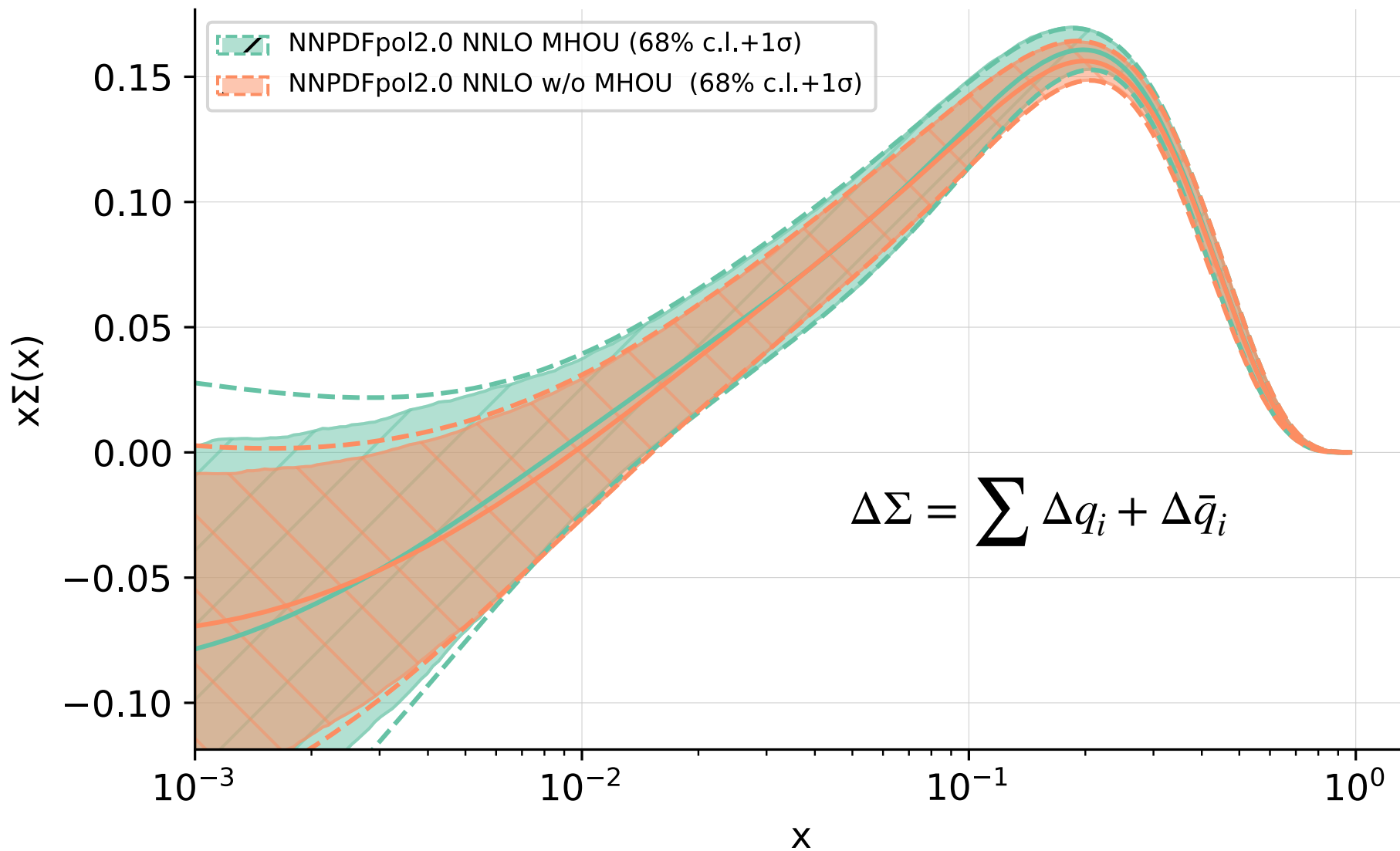
NNLO, with and w/o MHOU



Σ at 100 GeV



Σ at 100 GeV



Conclusions

Thanks to the latest developments in machine learning, their applications to PDF determination, and theoretical advances in the field: **we are entering the PDF precision era**

- The determination of Parton Distribution Function is one of the most important ingredients of predictions for highly energetic hadron collisions.
- The Monte Carlo replica method, coupled with Neural Networks and ensemble regression, allows us to avoid choosing a rigid functional form and allows us to incorporate easily correlations across datasets both theoretical and experimental.
- Data-driven methods must be used to choose the right architecture and settings for the problem to avoid biasing the extraction.
- Applied successfully to collinear polarized and unpolarized PDFs. The NNPDF tools and framework can also be extended beyond.

Backup



NNPDF

Open source

The public framework include:

1. Fitting Code
2. Analysis code
3. Documentation, examples and tutorials
4. All data (some from hepdata, some painfully extracted from papers or theses)
5. All theory predictions (some from plougshare, most painfully generated by us)

The whole NNPDF fitting framework is open source: code, data and theory predictions, is documented and available.

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

<https://github.com/NNPDF/nnpdf>

<https://docs.nnpdf.science/>

While the focus is on collinear and unpolarized PDFs on high energy collision, it is modular and extendable.

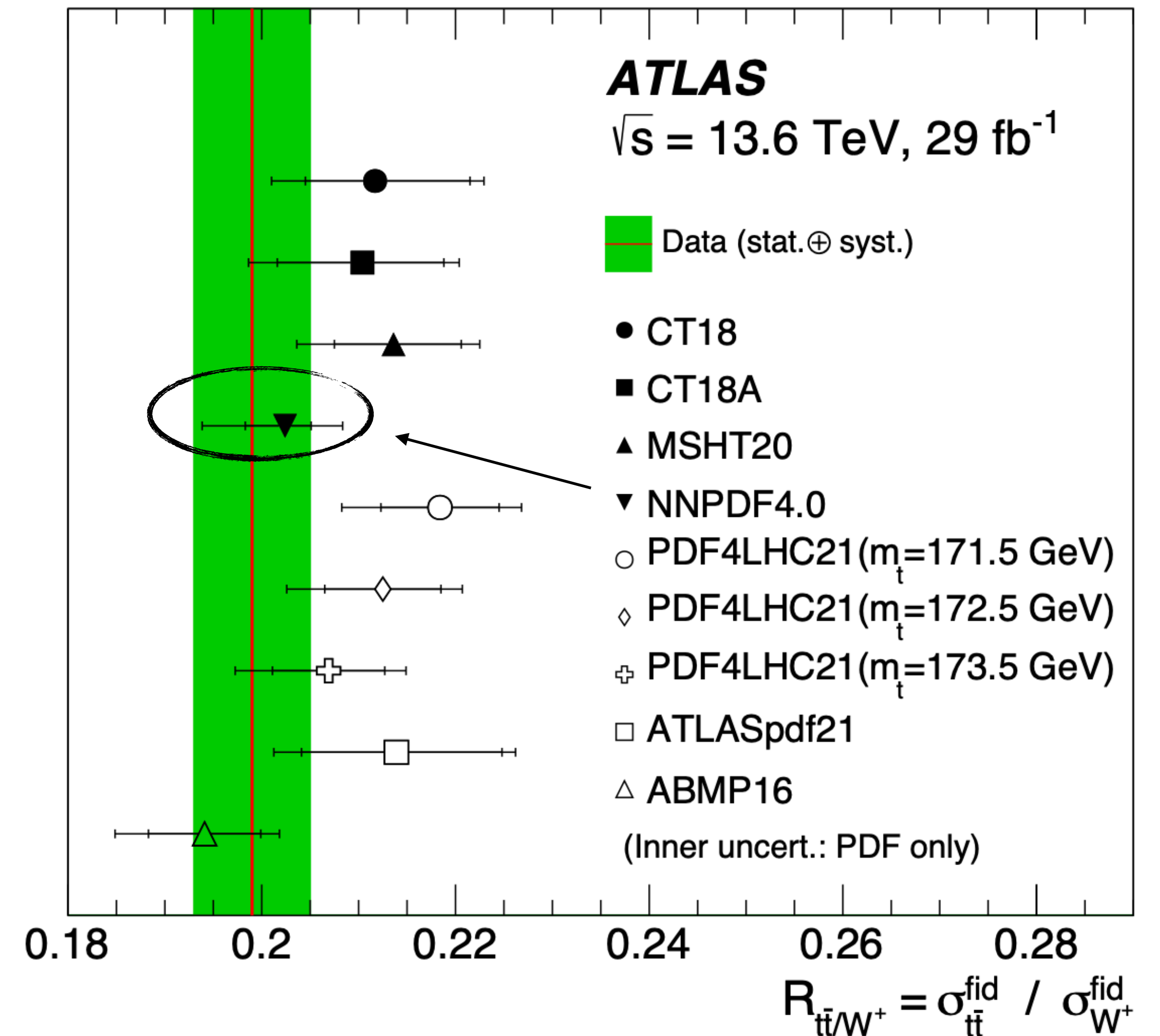
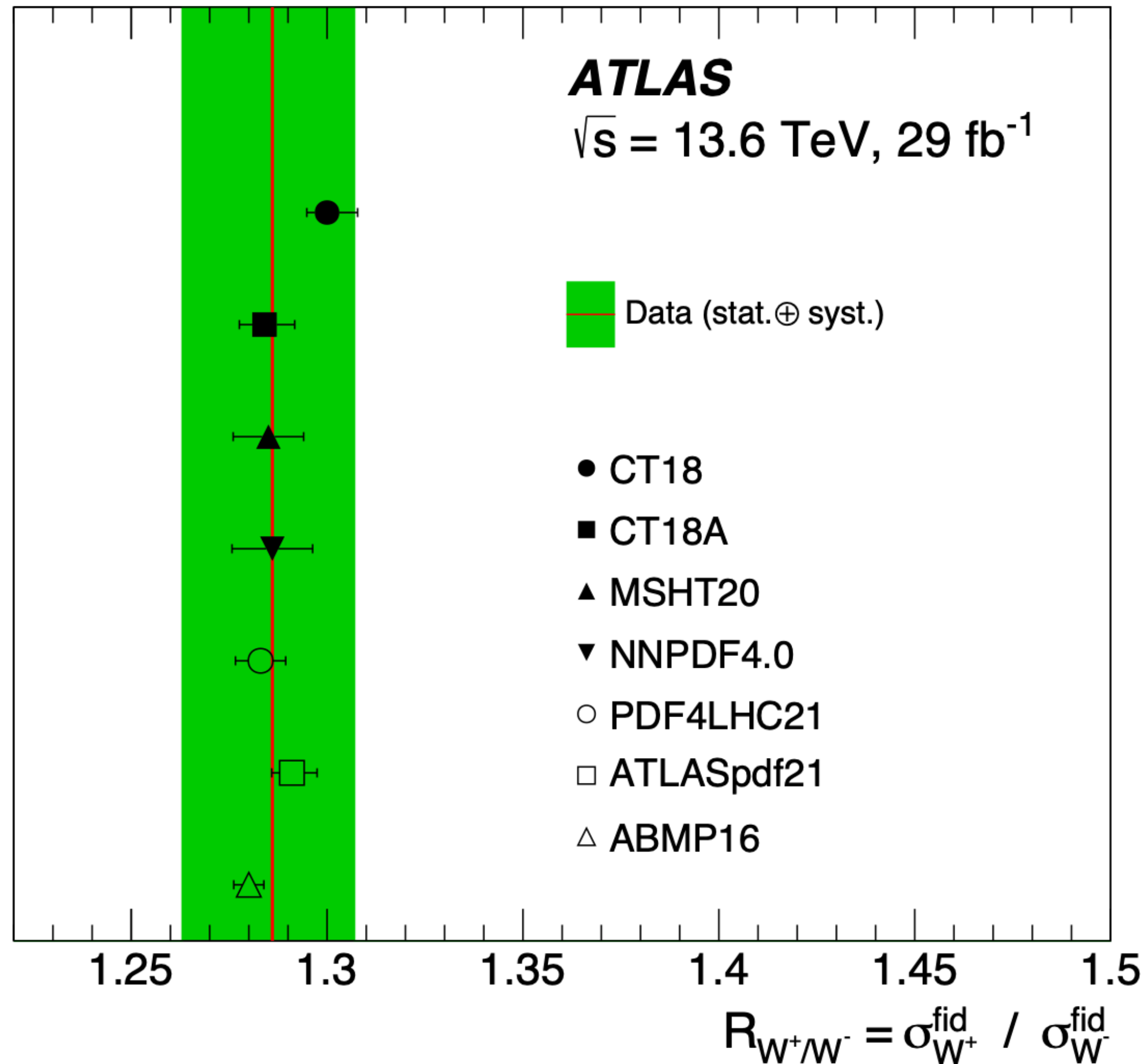
At leading order, any reaction in which the cross section can be recasted as a convolution could be in principle interfaced with our tools.

At higher orders, many specific techniques can surely be also extended

What is the impact of the PDF determination?

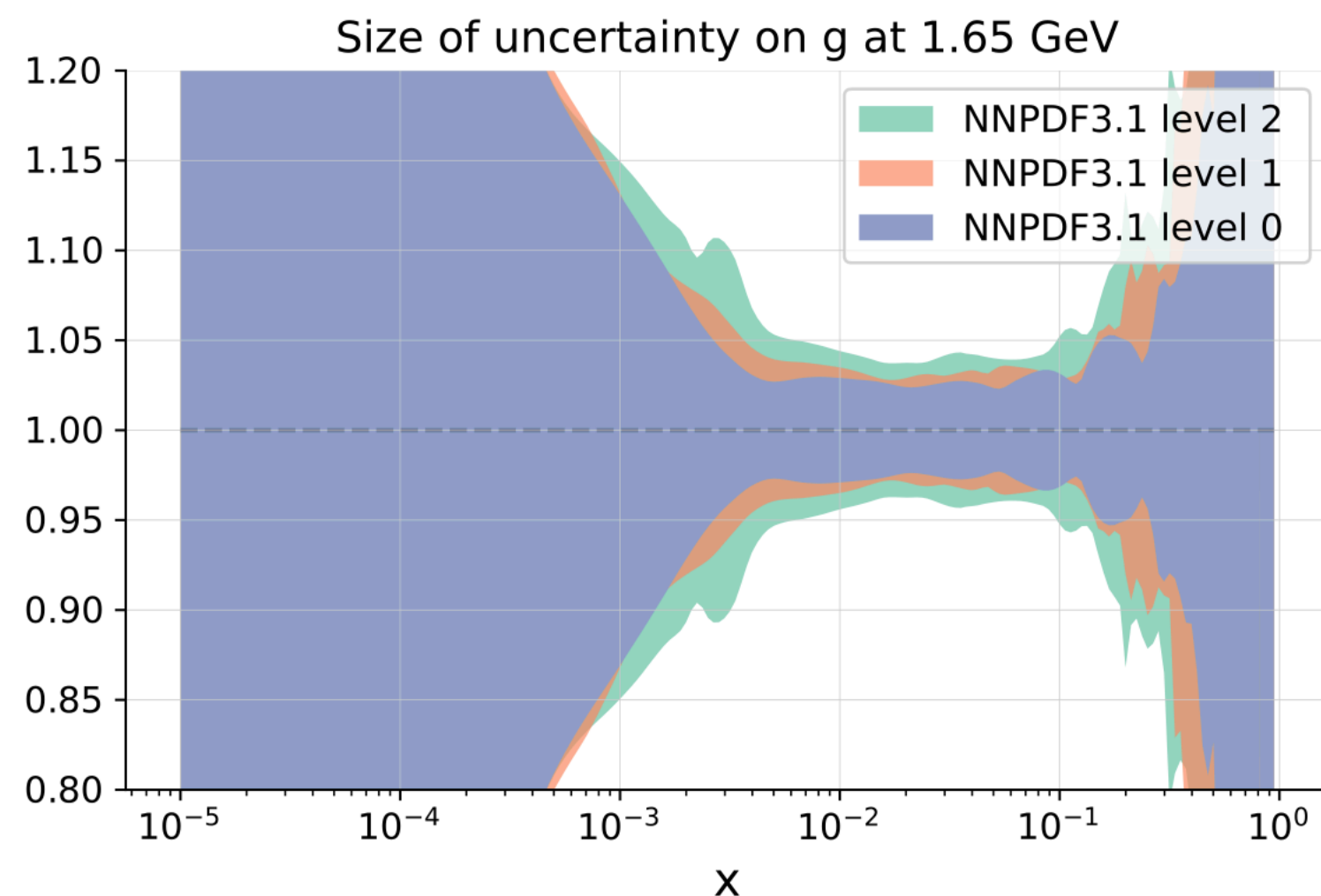
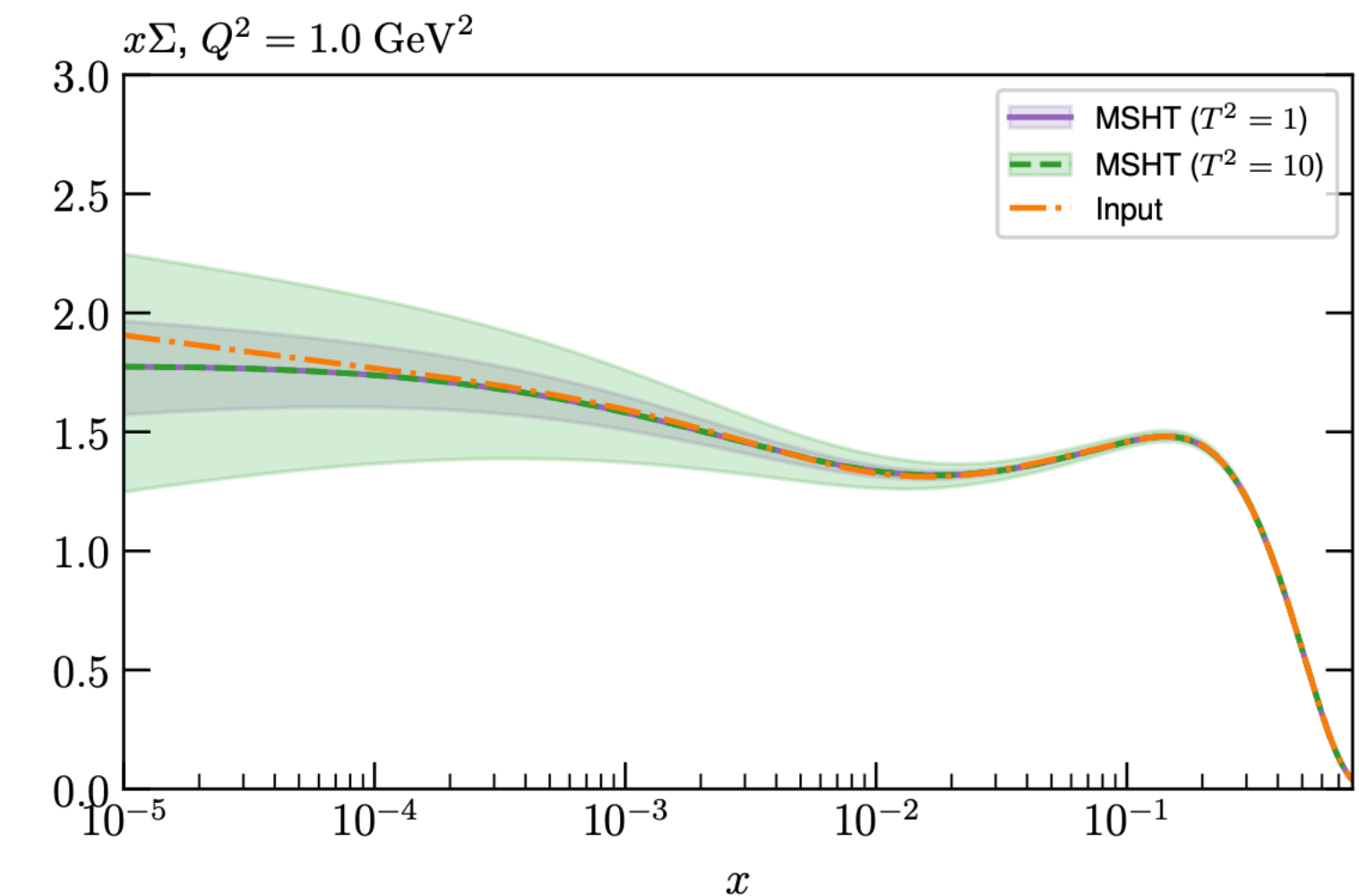
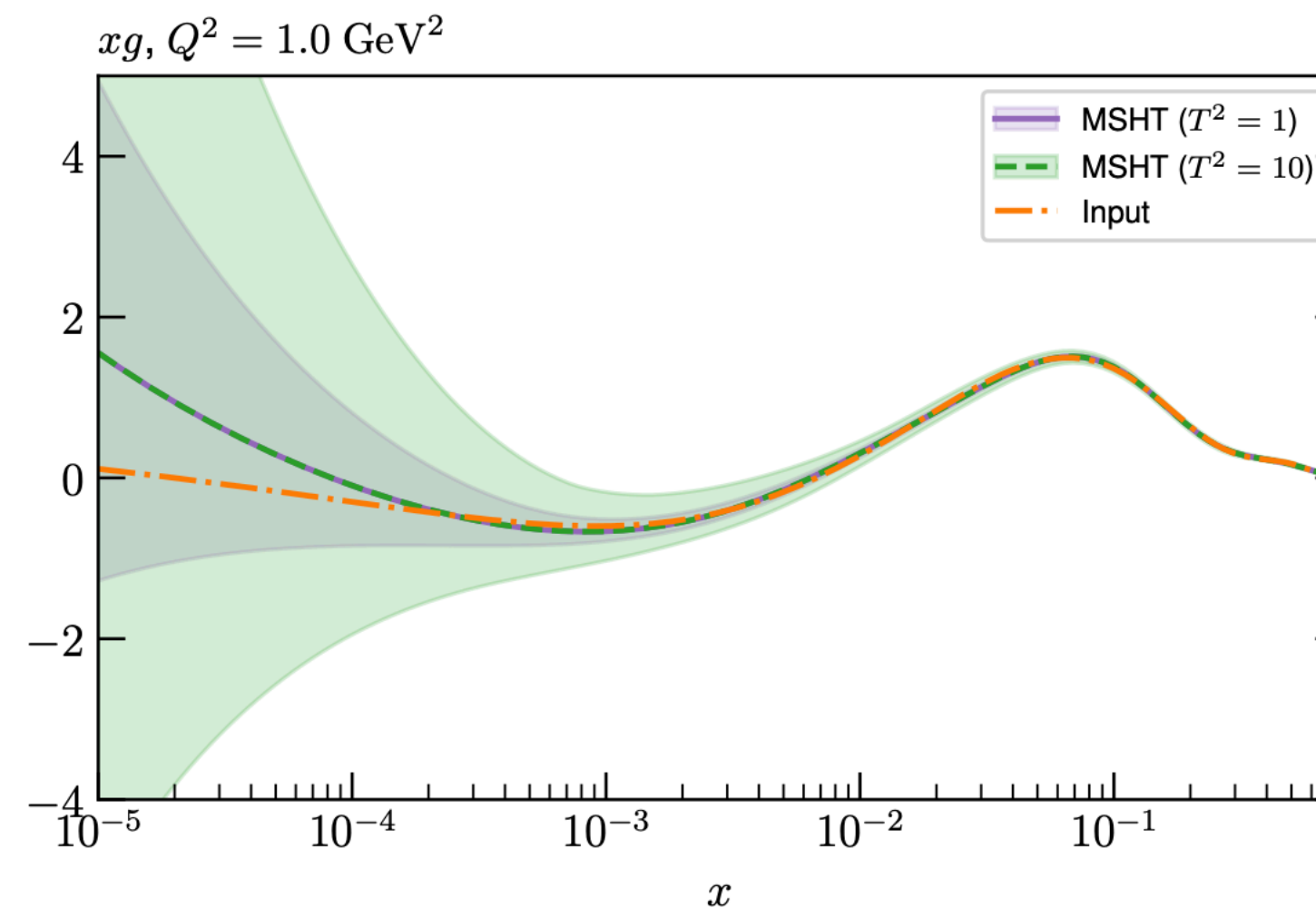
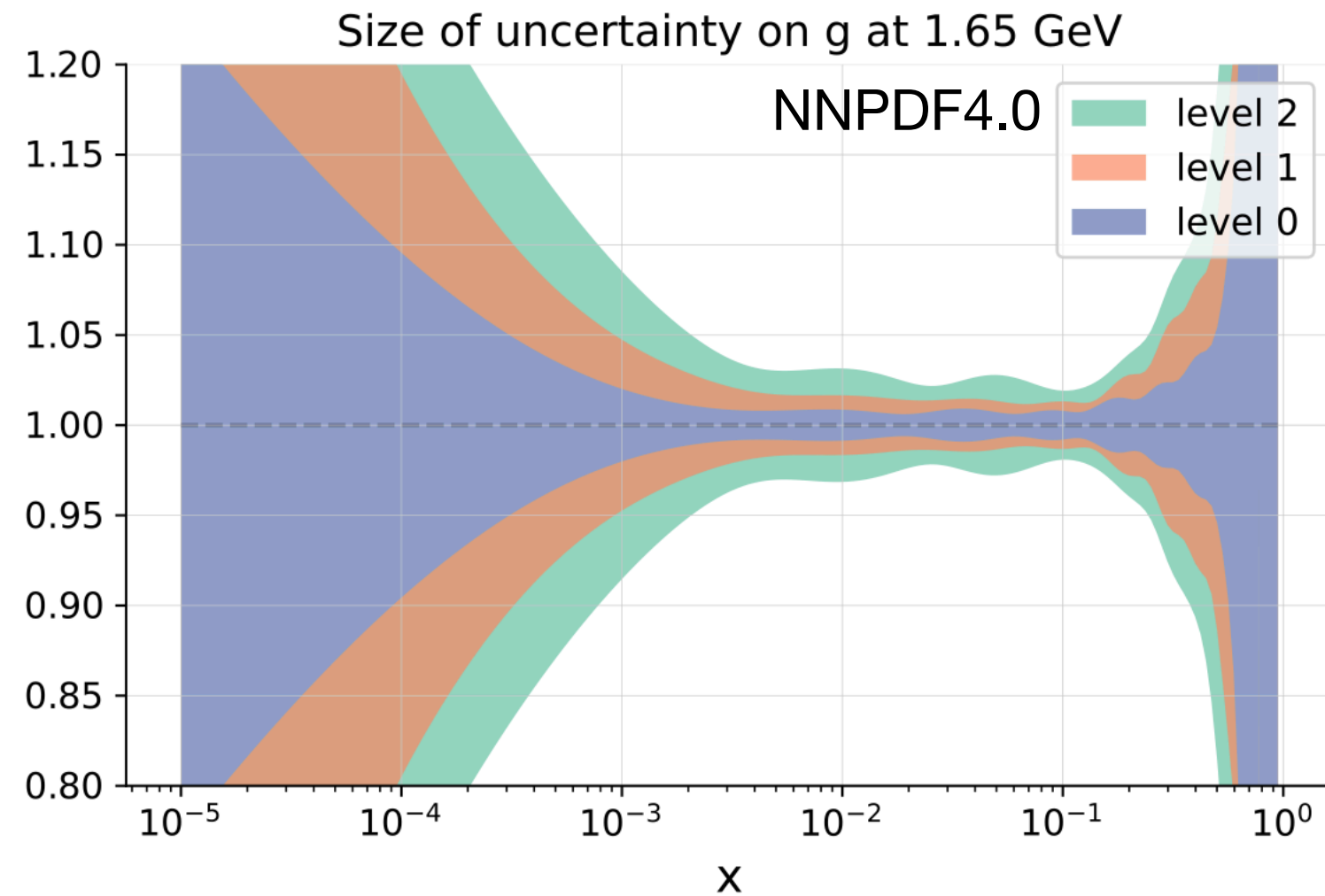
Example: Measurement of vector boson production cross section and their ratios
at $\sqrt{s} = 13.6$ with the ATLAS detector

[2403.12902](#)



Closure tests

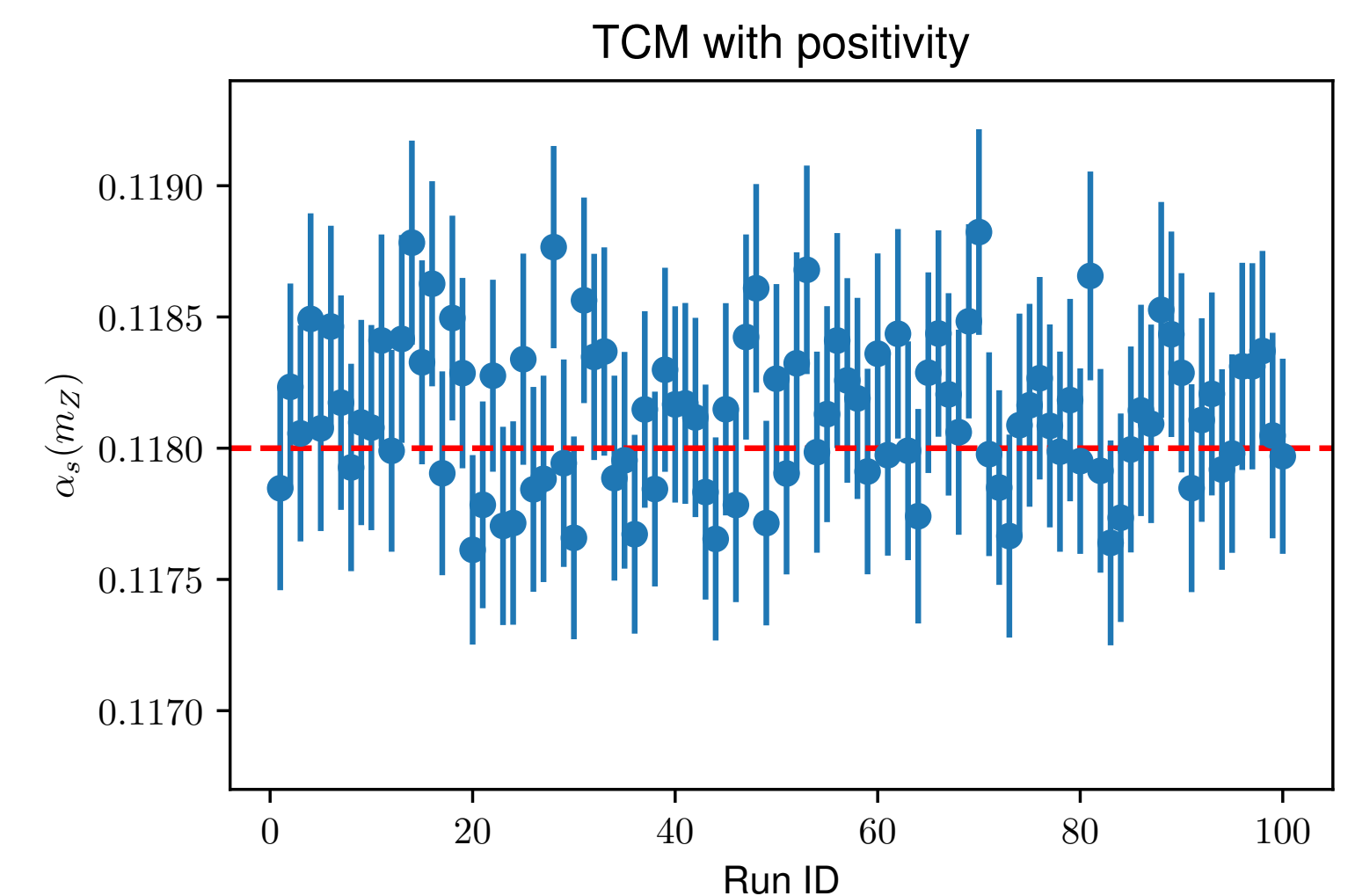
A stress test of global PDF fits: closure testing the MSHT PDFs and a first direct comparison to the neural net approach
MSHT collaboration - hep-ph/2407.07944



A Determination of $\alpha_s(m_Z)$ at $\text{aN}^3\text{LO}_{\text{QCD}} \otimes \text{NLO}_{\text{QED}}$ Accuracy from a Global PDF Analysis
NNPDF collaboration - hep-ph/2506.13871

Bayesian Approach to Inverse Problems: an Application to NNPDF Closure Testing
L. Del Debbio, T. Giani, M. Wilson - hep-ph/2111.05787

A simple but powerful idea:
Given a known input, make sure the methodology can reproduce it accurately with the proper propagation of input uncertainties.



Comparison between aN3LO results with different methodologies, data and theory

E.g., the recent combination of MSHT and NNPDF: [hep-ph/2411.05373](https://arxiv.org/abs/hep-ph/2411.05373)

In the combination of MSHT and NNPDF aN3LO PDFs many different effects are confounded:

1. Different methodologies
2. Slightly different choices of datasets
3. Different theory settings (most notably charm fitted or not)

Can we separate the methodological effects from the physics choices?

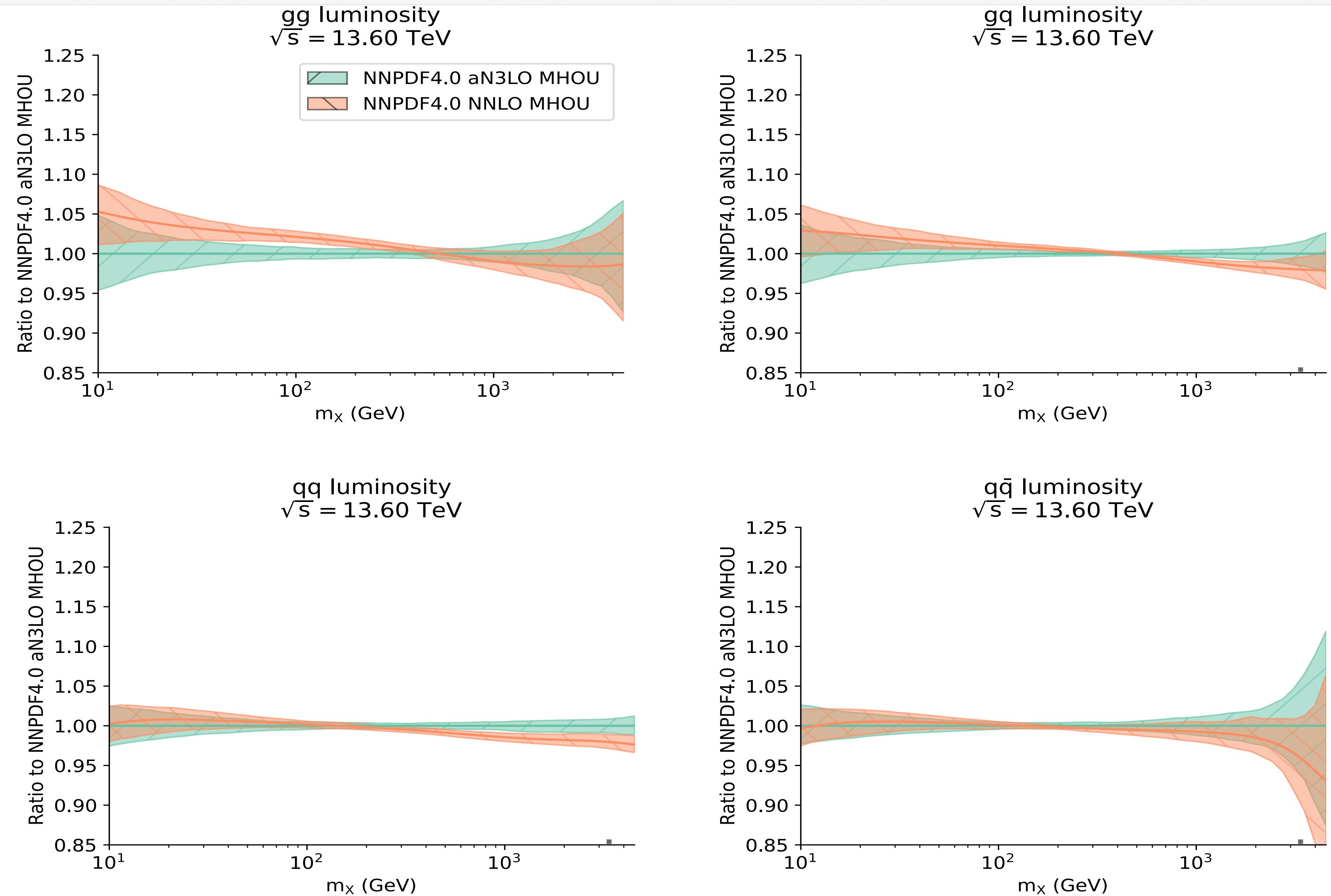


Figure 4 from [hep-ph/2411.05373](https://arxiv.org/abs/hep-ph/2411.05373)

Comparison between aN3LO results: same data and theory

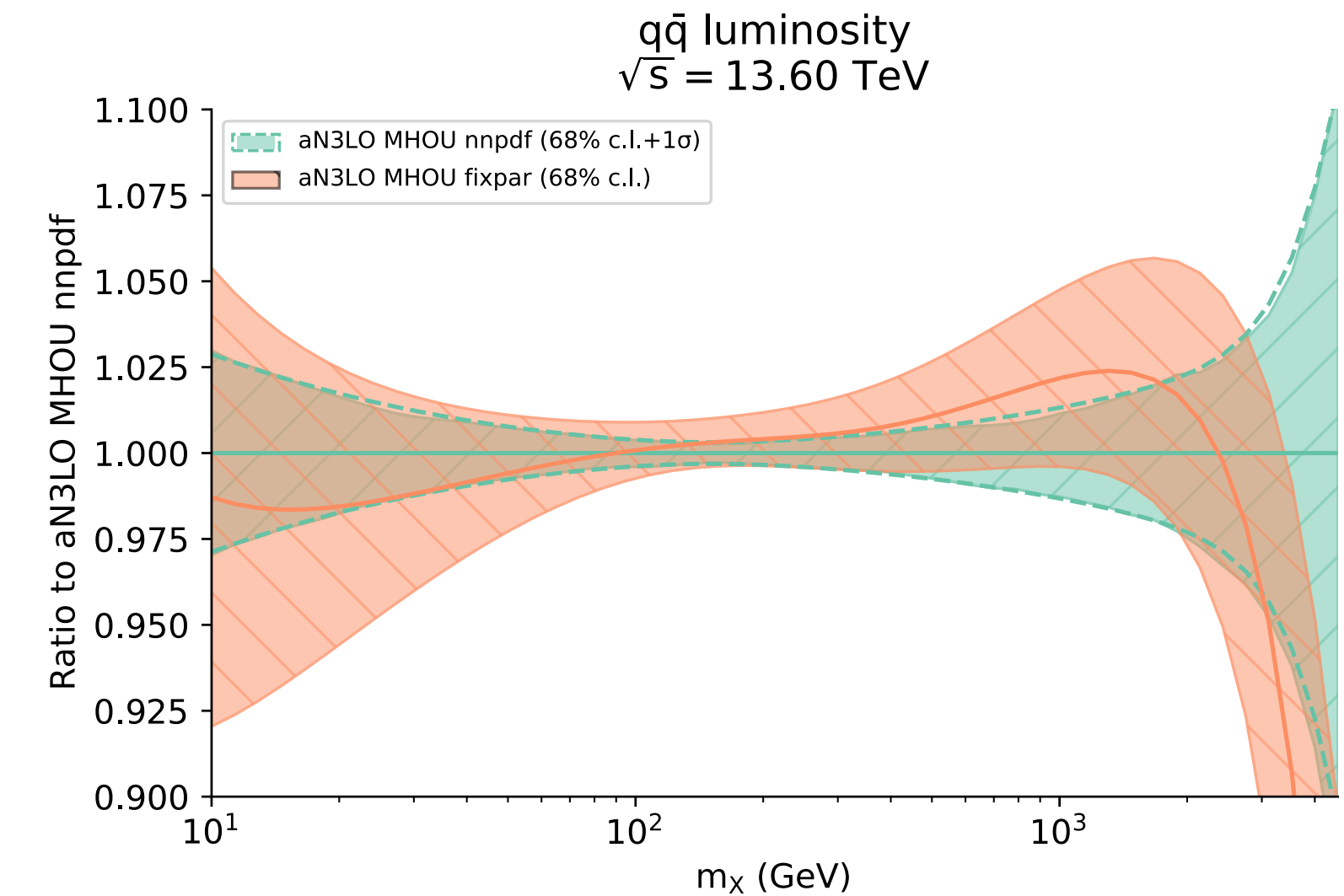
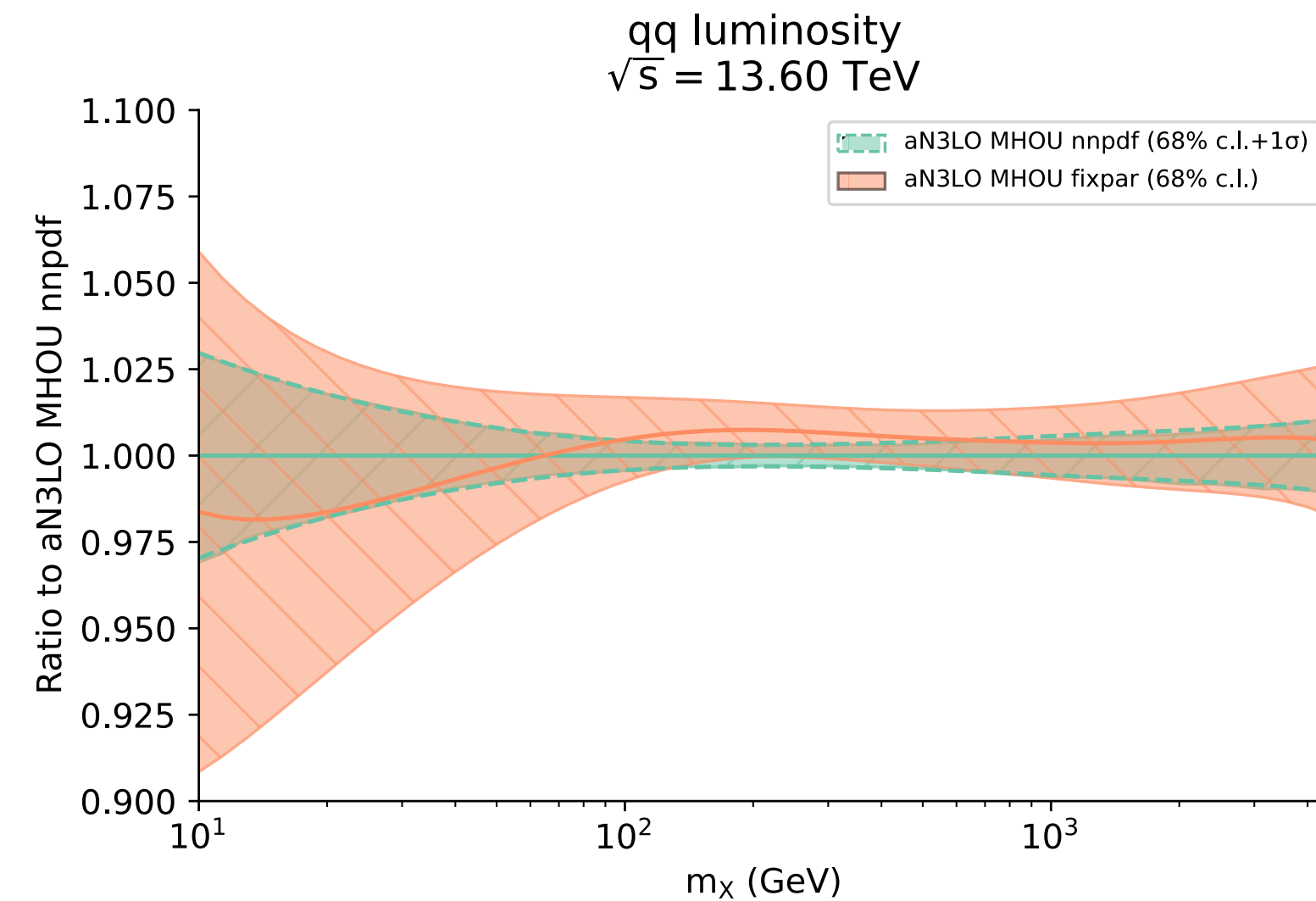
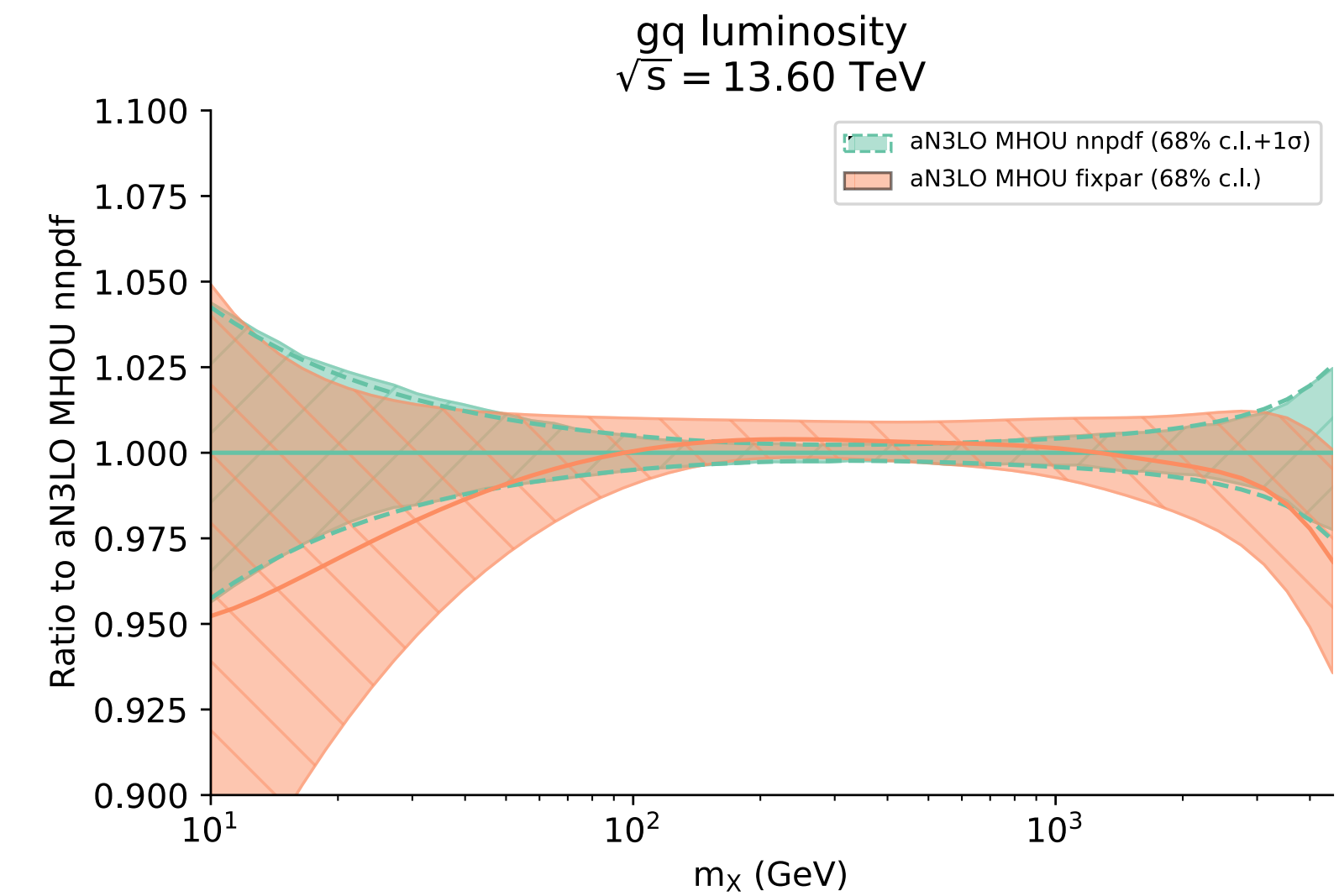
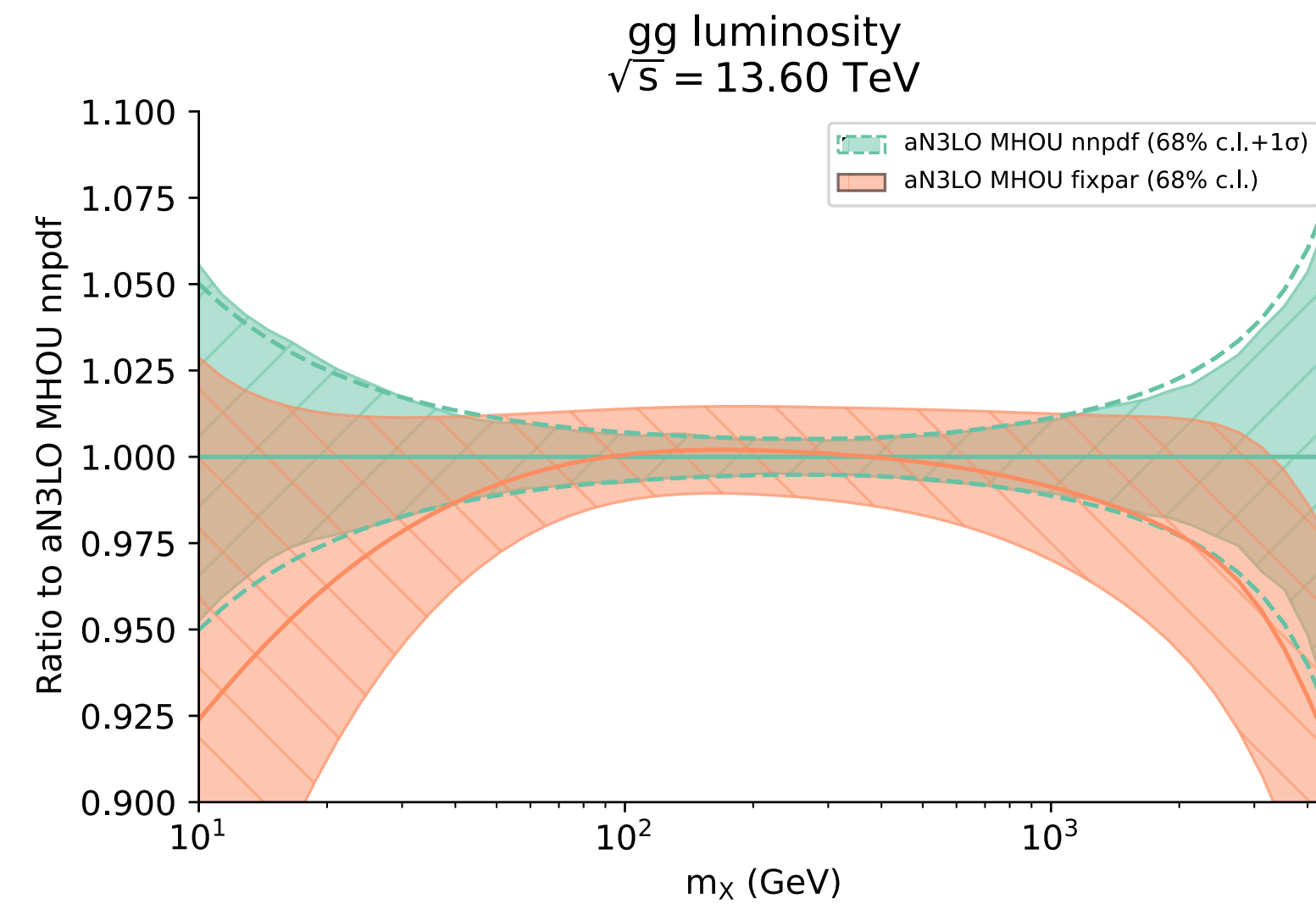
Of course we can :)

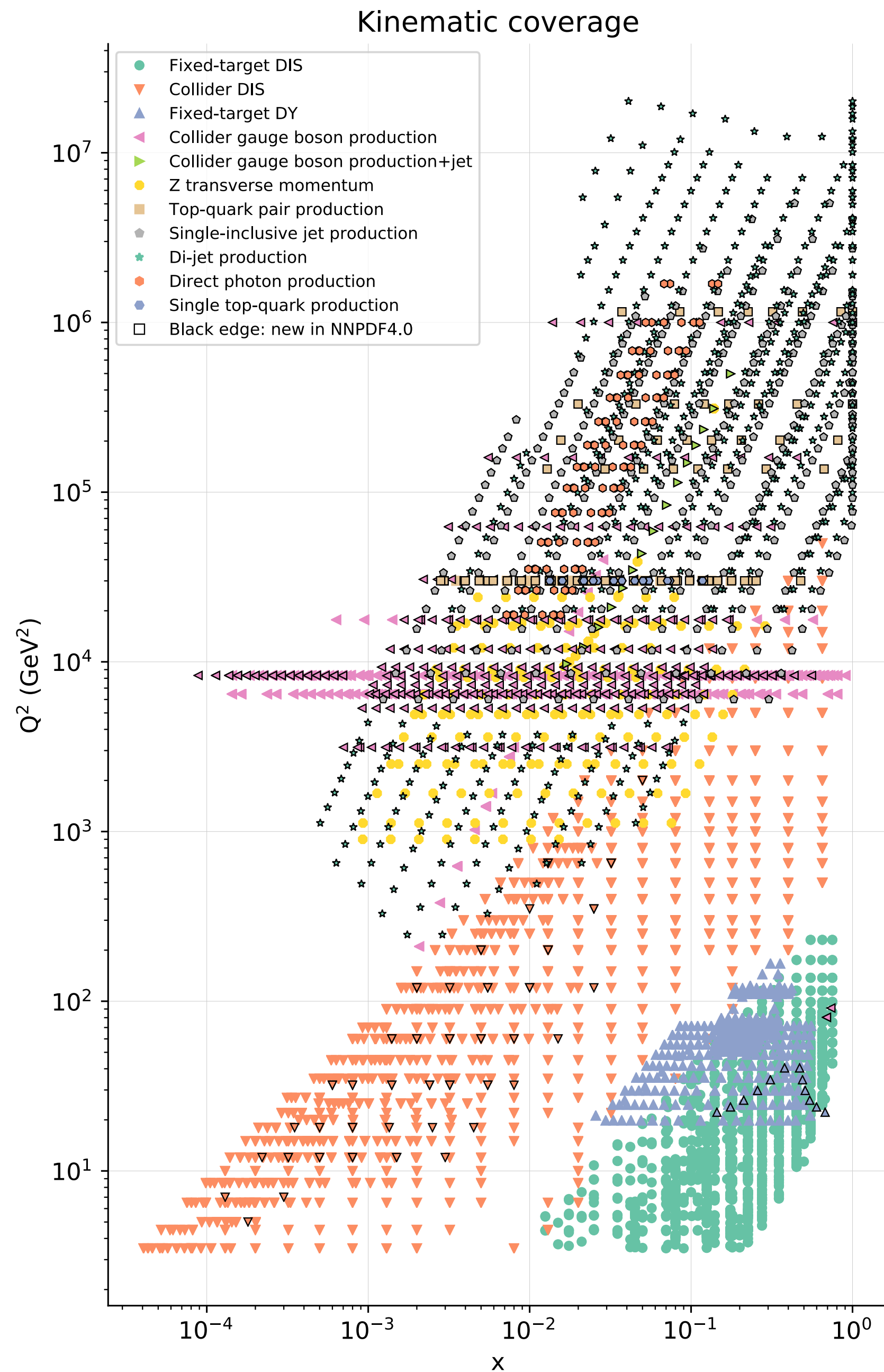
~~FP
NNPDF~~

With the code presented in [hep-ph/2602.07118](https://arxiv.org/abs/hep-ph/2602.07118) we can disentangle these effects.

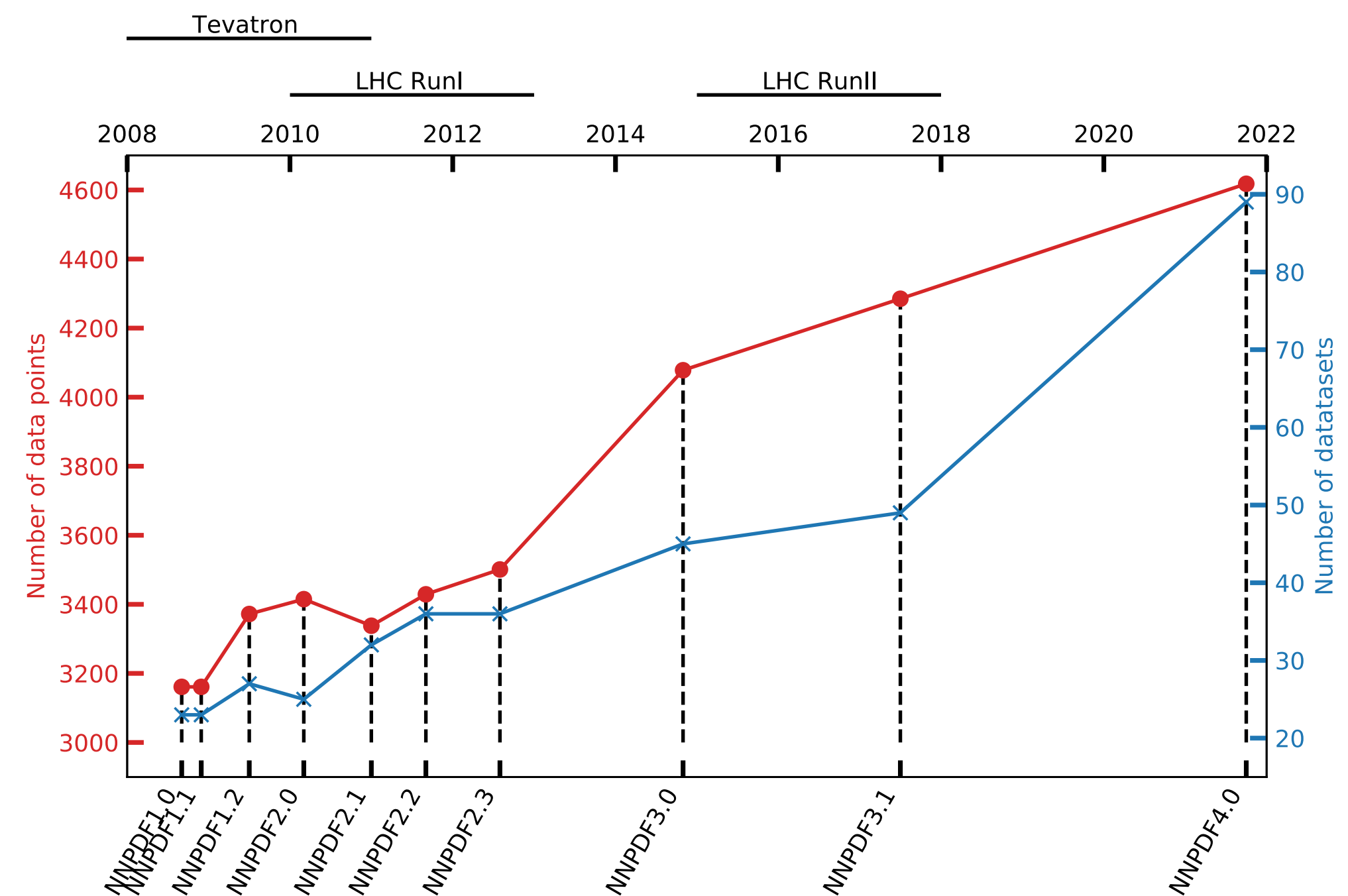
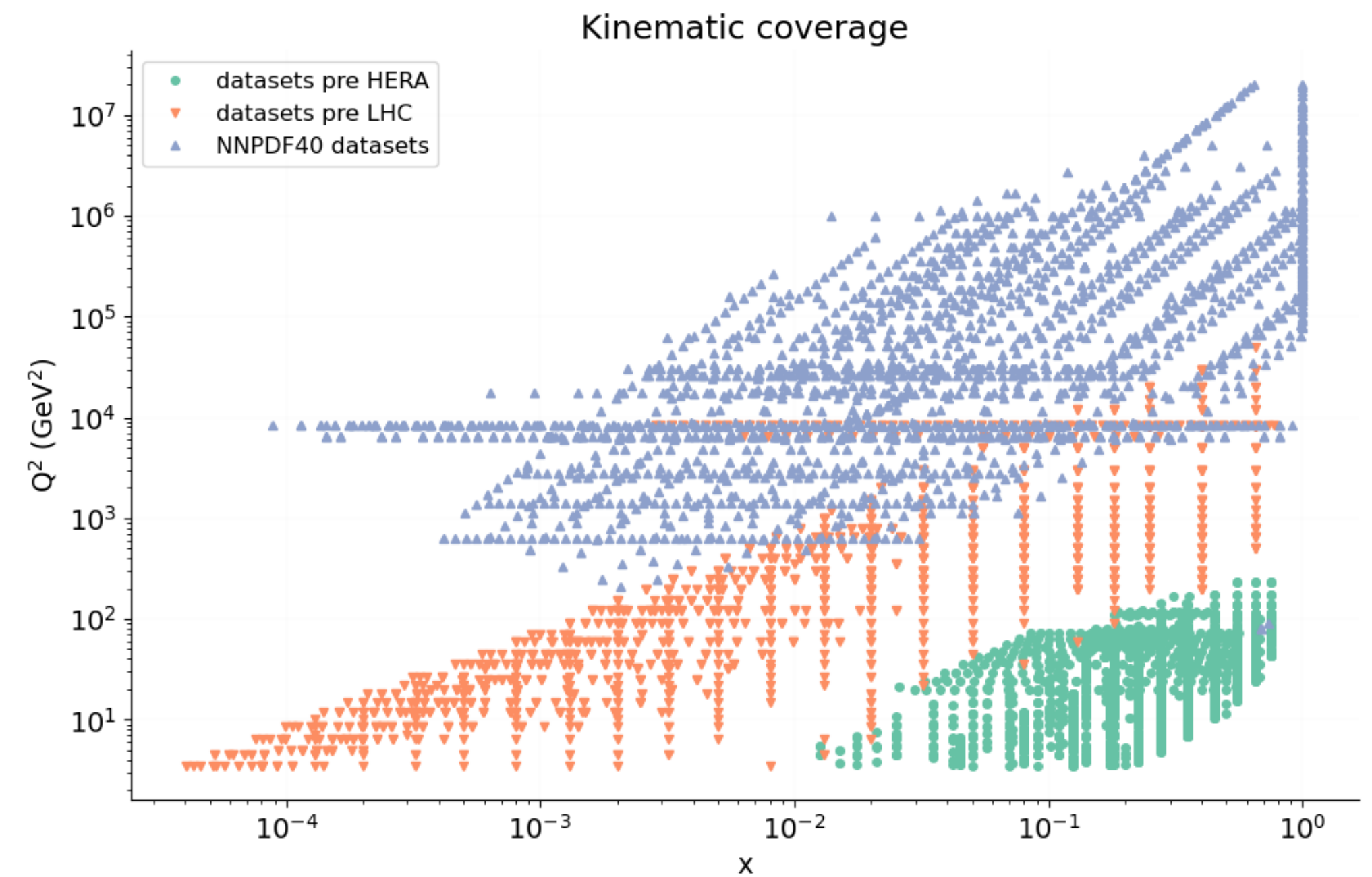
Indeed, it seems that the **bulk of the differences is driven by data and theory, and not methodologies!**

Some differences remain, although within uncertainties, so still room for more studies with FPDF 🙄🙄





The full list of datasets used 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 ($\mathcal{L} = 36 \text{ pb}^{-1}$)	✗	✗	✗	✗	✓
CMS Z 7 TeV ($\mathcal{L} = 36 \text{ pb}^{-1}$)	✗	✗	✗	✗	✓
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 ($\mathcal{L} = 18.4 \text{ fb}^{-1}$)	✗	✗	✗	(✓)	✗
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 σ_{tt}^{tot} 5 TeV	✗	✓	✗	✗	✗
CMS σ_{tt}^{tot} 7, 8 TeV	✓	✓	✗	✗	✗
CMS σ_{tt}^{tot} 8 TeV	✗	✗	✗	✗	✓
CMS σ_{tt}^{tot} 5, 7, 8, 13 TeV	✗	✗	✓	✗	✗
CMS σ_{tt}^{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 ($\mathcal{L} = 35 \text{ pb}^{-1}$)	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6 \text{ fb}^{-1}$)	✓	✓	✗	(✓)	✓
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 σ_{tt}^{tot} 7, 8 TeV	✓	✓	✓	✗	✗
ATLAS σ_{tt}^{tot} 7, 8 TeV	✗	✗	✓	✗	✗
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L} = 3.2 \text{ fb}^{-1}$)	✓	✗	✓	✗	✗
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L} = 139 \text{ fb}^{-1}$)	✗	✓	✗	✗	✗
ATLAS σ_{tt}^{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 $W \rightarrow \ell\nu$ asymmetry (1.8 TeV)	✗	✗	✗	✓	✗
CDF $W \rightarrow e\nu$ asymmetry ($\mathcal{L} = 170 \text{ pb}^{-1}$)	✗	✗	✗	✓	✗
CDF $W \rightarrow e\nu$ asymmetry ($\mathcal{L} = 1 \text{ fb}^{-1}$)	✗	✗	✗	✗	✓
CDF k_t inclusive jets	✓	✗	✗	✗	✓
CDF cone-based inclusive jets	✗	✗	✗	✓	✗
D0 Z rapidity	✓	✓	✗	✓	✓
D0 $W \rightarrow e\nu$ asymmetry ($\mathcal{L} = 0.75 \text{ fb}^{-1}$)	✗	✗	✗	✗	✓
D0 $W \rightarrow e\nu$ (prod.) asymmetry ($\mathcal{L} = 9.7 \text{ fb}^{-1}$)	✗	✗	(✓)	✗	✓
D0 $W \rightarrow e\nu$ (prod. and decay) asymmetry ($\mathcal{L} = 9.7 \text{ fb}^{-1}$)	✓	(✓)	✓	✓	✗
D0 $W \rightarrow \mu\nu$ asymmetry ($\mathcal{L} = 0.3 \text{ fb}^{-1}$)	✗	✗	✗	✓	✗
D0 $W \rightarrow \mu\nu$ asymmetry ($\mathcal{L} = 7.3 \text{ fb}^{-1}$)	✓	✓	✓	✗	✓
D0 cone-based inclusive jets	✗	✗	✗	✓	✓
CDF and D0 top-pair production	✗	✗	(✓)	✗	✓
CDF and D0 single-top production	✗	✗	✓	✗	✗

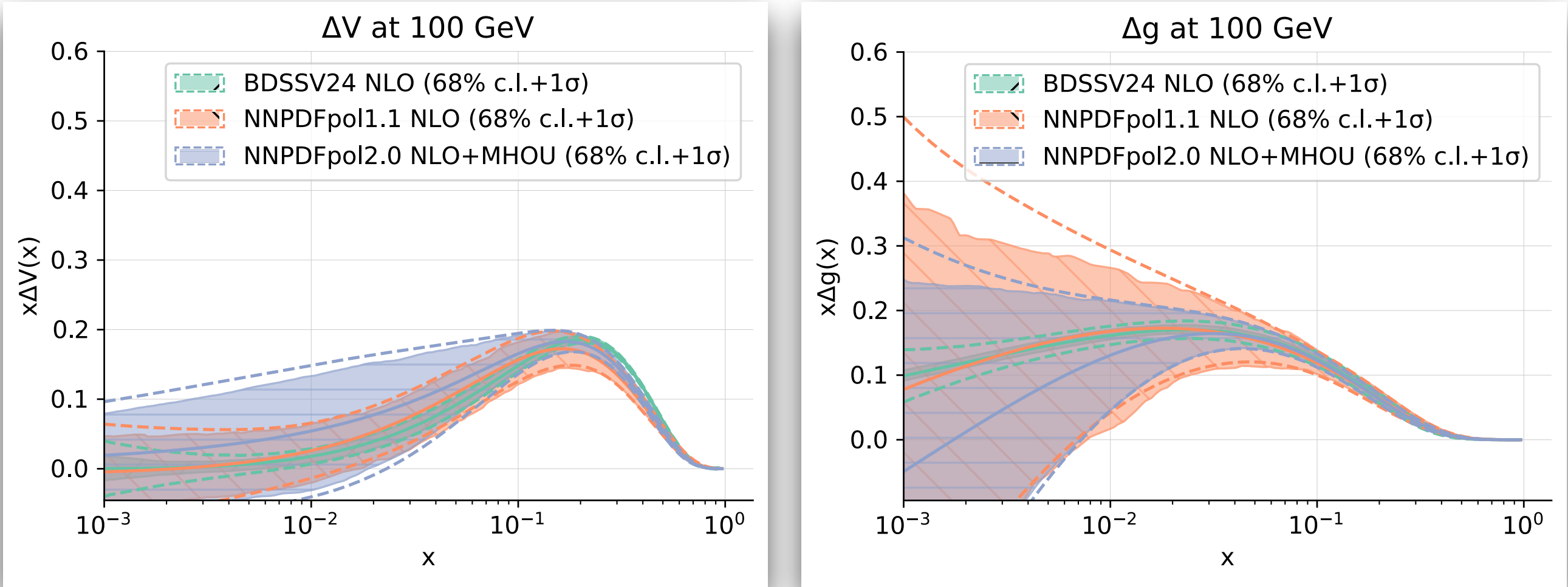
Data set	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
DY E866 $\sigma_{\text{DY}}^d/\sigma_{\text{DY}}^p$ (NuSea)	✓	✓	✓	✓	✓
DY E866 σ_{DY}^p	✓	✓	✗	✓	✓
DY E605 σ_{DY}^p	✓	✓	✓	✓	✗
DY E906 $\sigma_{\text{DY}}^d/\sigma_{\text{DY}}^p$ (SeaQuest)	✗	✓	✗	✗	✗

LHCb Z 7 TeV ($\mathcal{L} = 940 \text{ pb}^{-1}$)	✓	✓	✗	✗	✓
LHCb $Z \rightarrow ee$ 8 TeV ($\mathcal{L} = 2 \text{ fb}^{-1}$)	✓	✓	✓	✓	✓
LHCb W 7 TeV ($\mathcal{L} = 37 \text{ pb}^{-1}$)	✗	✗	✗	✗	✓
LHCb $W, Z \rightarrow \mu$ 7 TeV	✓	✓	✓	✓	✓
LHCb $W, Z \rightarrow \mu$ 8 TeV	✓	✓	✓	✓	✓
LHCb $W \rightarrow e$ 8 TeV	✗	(✓)	✗	✗	✗
LHCb $Z \rightarrow \mu\mu, ee$ 13 TeV	✗	✓	✗	✗	✗

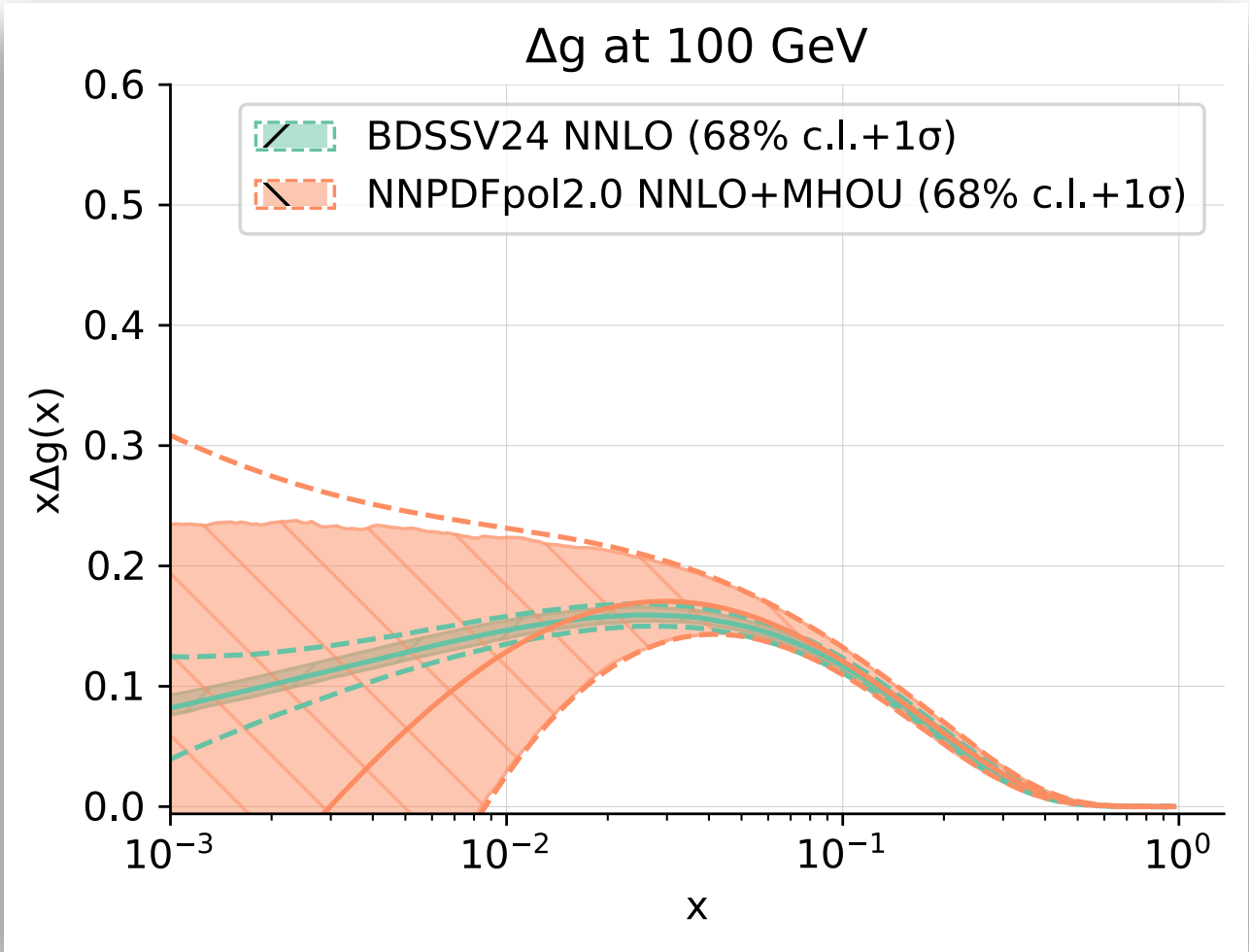
Data set	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
NMC F_2^d/F_2^p	✓	✓	✗	✗	✓
NMC $\sigma^{\text{NC},p}$	✓	✓	✗	✓	✓
SLAC F_2^p, F_2^d	✓	✓	✓	✗	✓
BCDMS F_2^p	✓	✓	✓	✓	✓
BCDMS F_2^d	✓	✓	✗	✓	✓
BCDMS, NMC, SLAC F_2	✗	✗	✗	✗	✓
CHORUS $\sigma_{CC}^\nu, \sigma_{CC}^{\bar{\nu}}$	✓	✓	✗	✗	✓
CHORUS	✗	✗	✓	✗	✗
NuTeV F_2, F_3	✗	✗	✗	✗	✓
NuTeV/CCFR $\sigma_{CC}^\nu, \sigma_{CC}^{\bar{\nu}}$	✓	✓	✓	✓	✓
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 $\sigma_{\text{NC,CC}}^p$	✓	✓	✓	✓	✗
HERA I+II σ_{cc}^{red}	✗	✓	✗	(✓)	✓
HERA I+II σ_{bb}^{red}	✗	✓	✗	(✓)	✗
HERA I+II σ_{cc}^{red}	✓	✗	✓	✓	✗
H1 $F_2^{c\bar{c}}$	✗	✗	✗	✓	✗
H1 $F_2^{b\bar{b}}$	✓	✗	✓	✗	✗
ZEUS σ_{bb}^{red}	✓	✗	✓	✗	✗
H1 F_L	✗	✗	✗	✓	✓
H1 and ZEUS F_L	✗	✗	✗	✗	✓
ZEUS 820 (HQ) (1j)	✗	(✓)	✗	✗	✗
ZEUS 920 (HQ) (1j)	✗	(✓)	✗	✗	✗
H1 (LQ) (1j-2j)	✗	(✓)	✗	✗	✗
H1 (HQ) (1j-2j)	✗	(✓)	✗	✗	✗
ZEUS 920 (HQ) (2j)	✗	(✓)	✗	✗	✗

Comparison to other studies

NLO



NNLO



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.

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.

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