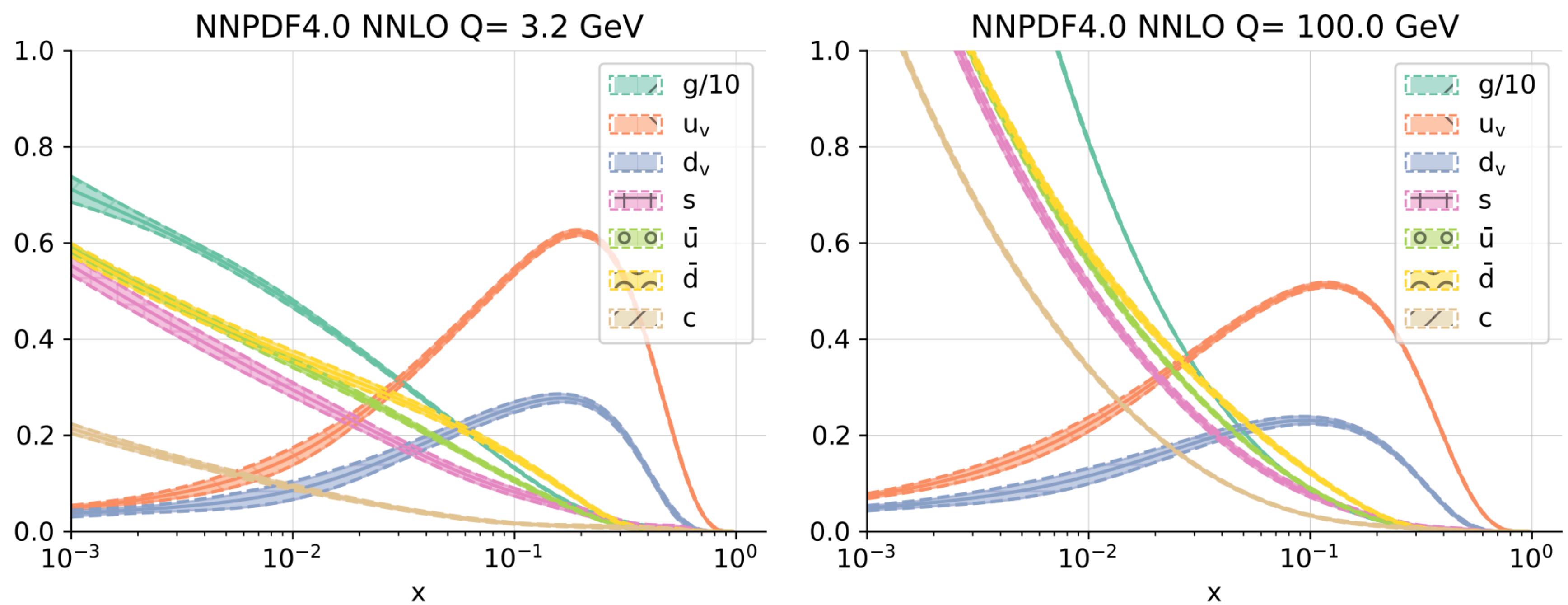
The path to proton structure at 1% accuracy

Juan Rojo, VU Amsterdam & Nikhef

RTG Colloquium - RWTH Aachen University

23rd November 2021



Why Nucleon Structure?

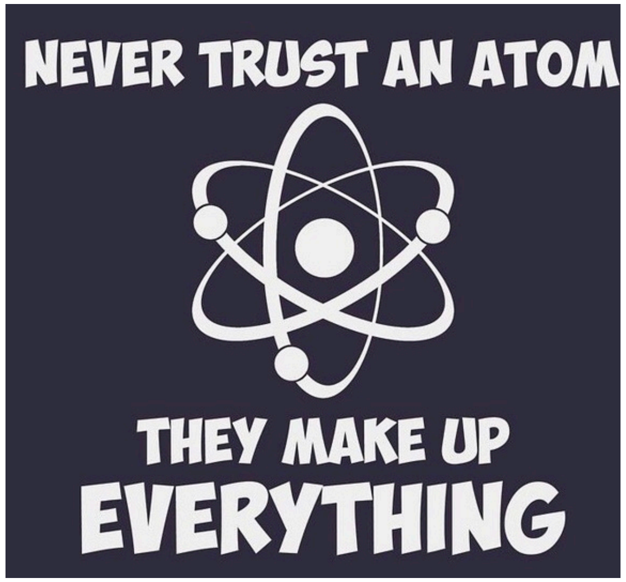
Why Nucleon Structure?

One may claim that the **nucleon is a rather ``boring" particle**, surely after one century of studying it, we know everything about the proton?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!

Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!



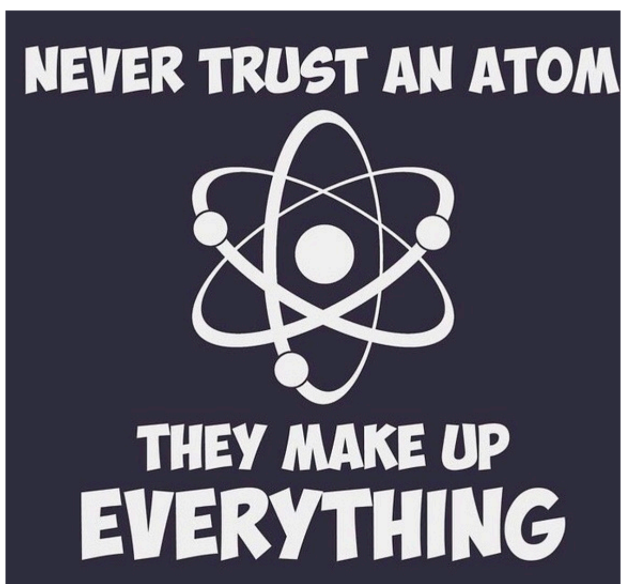
long distances / low energies

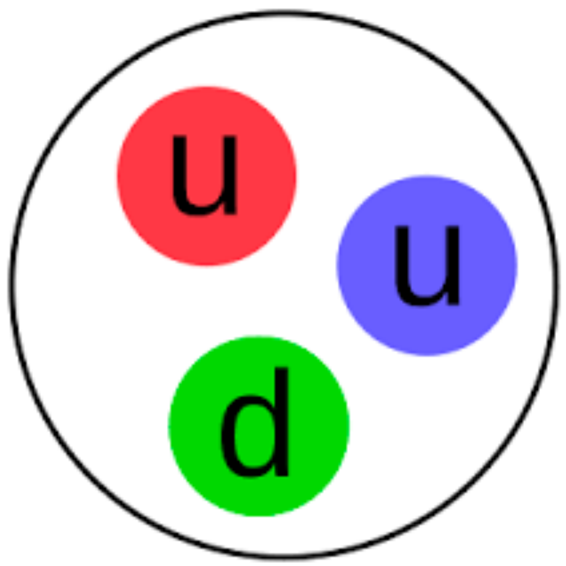
short distances / high energies

a point particle

Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!





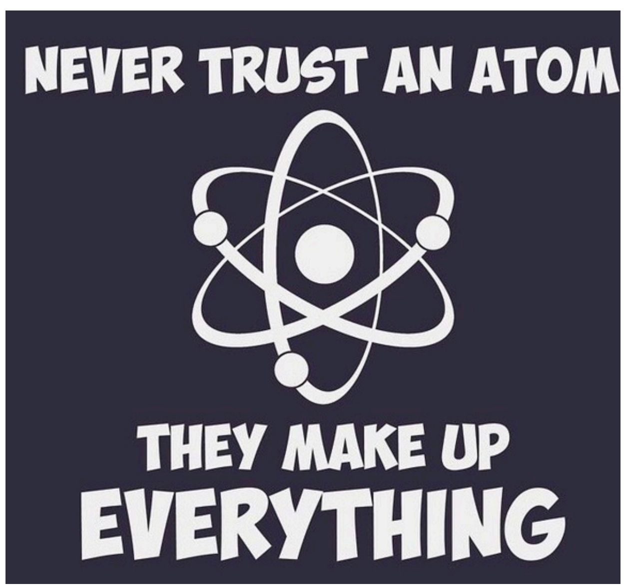
long distances / low energies → *short distances / high energies*

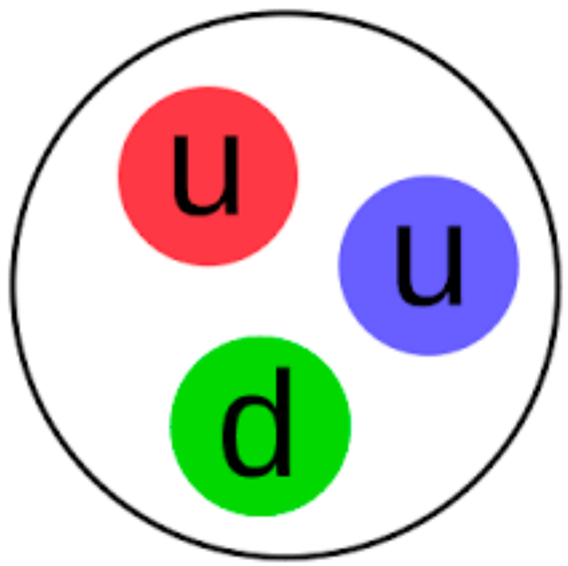
a point particle

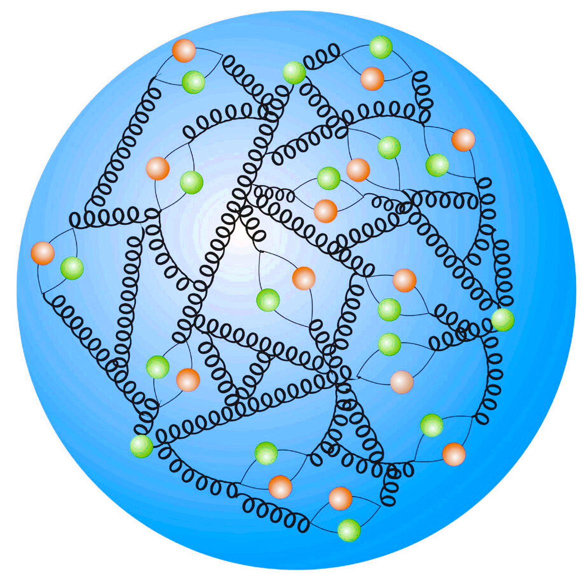
3 valence quarks

Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!







long distances / low energies *short distances / high energies*

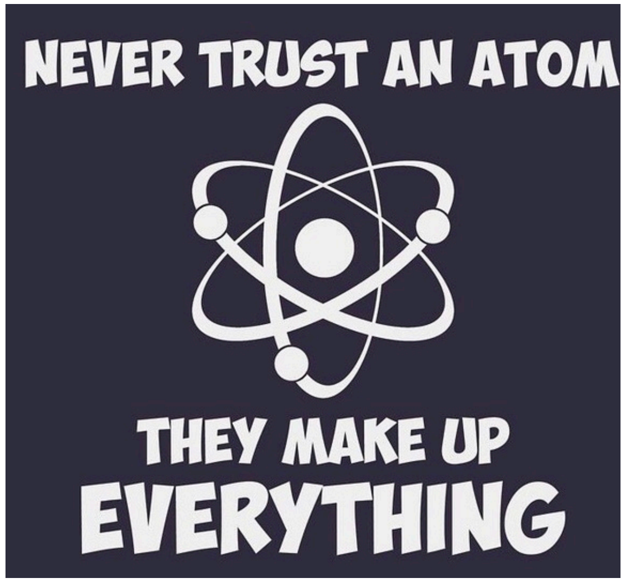
a point particle

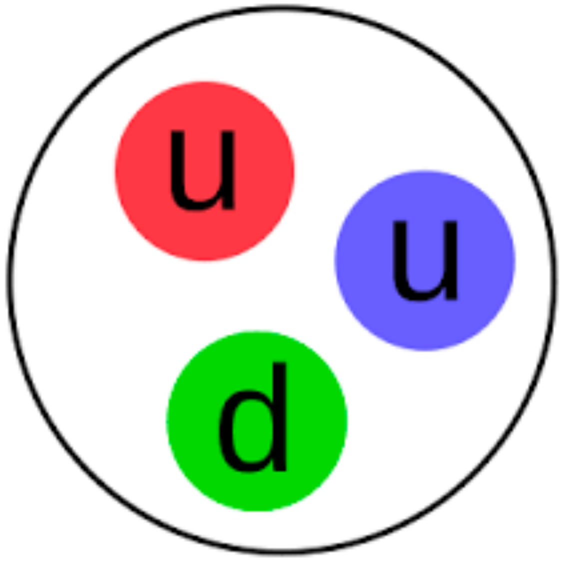
3 valence quarks

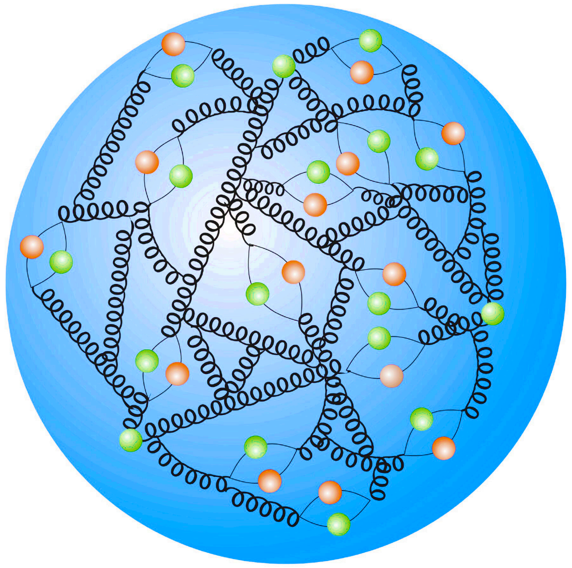
sea quarks, gluons

Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!







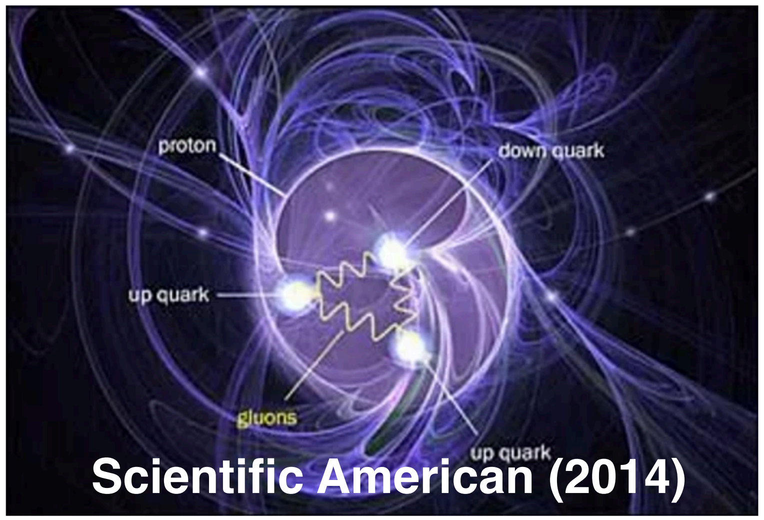
long distances / low energies → *short distances / high energies*

a point particle | 3 valence quarks | sea quarks, gluons | heavy quarks, photons, leptons, gauge and Higgs bosons …

A gateway to unravelling QCD

THE SCIENCES

Proton Spin Mystery Gains a New Clue



Non-zero gluon polarisation

Intrinsic Charm

The proton keeps surprising us as an endless source of **fundamental discoveries!**

QUANTUM PHYSICS

Decades-Long Quest Reveals Details of the Proton's Inner Antimatter

27

Twenty years ago, physicists set out to investigate a mysterious asymmetry in the proton's interior. Their results, published today, show how antimatter helps stabilize every atom's core.

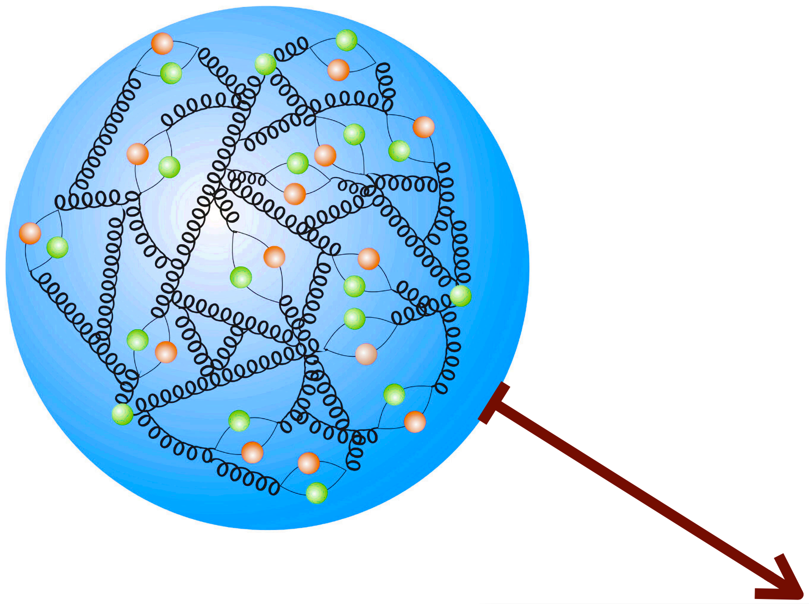
Antimatter asymmetry

After 40 years of studying the strong nuclear force, a revelation

BFKL dynamics

This was the year that analysis of data finally backed up a prediction, made in the mid 1970s, of a surprising emergent behaviour in the strong nuclear force

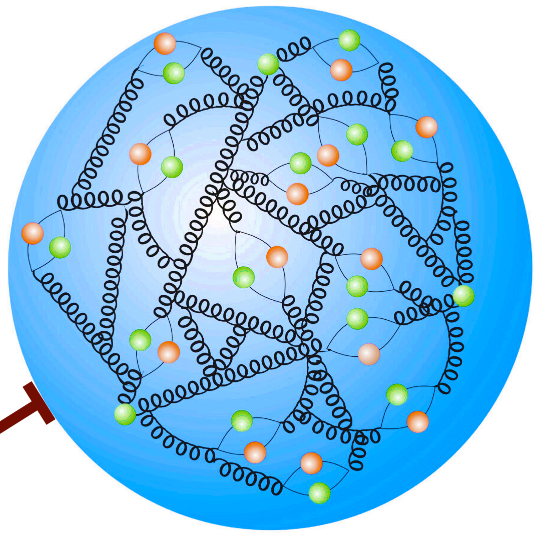
Why Nucleon Structure?



Address fundamental questions about Quantum Chromodynamics

- origin of mass & spin
- **heavy quark** & **antimatter content**
- 3D imaging
- gluon-dominated matter
- **nuclear modifications**
- Interplay with BSM e.g. via ``**SMEFT PDFs**”

Why Nucleon Structure?



Key component of predictions for particle, nuclear, and astro-particle experiments

- **pp: ATLAS, CMS, LHCb, ALICE**
- ep: fixed target DIS, HERA
- neutrinos: IceCube, KM3NET, Forward Physics Facility @ LHC
- heavy ions: LHC Pb, LHC O, RHIC
- *pp (future): HL-LHC, FCC, SppS*
- ep (future): **Electron-Ion Collider,** LHeC, FCC-eh

LHC master formula

hard cross-section

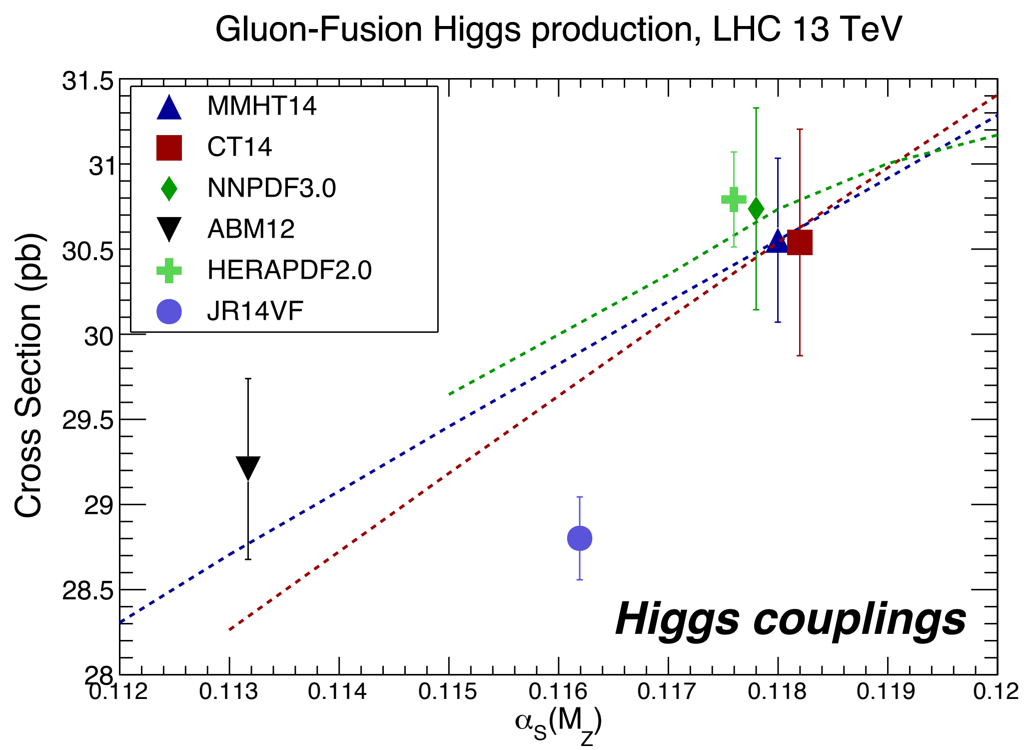
$$\sigma(M,s) \propto \sum_{ij=u,d,g,\ldots} \int_{M^2}^{s} d\hat{s}\ \mathscr{L}_{ij}(\hat{s},s)\,\tilde{\sigma}_{ij}(\hat{s},\alpha_s(M))$$

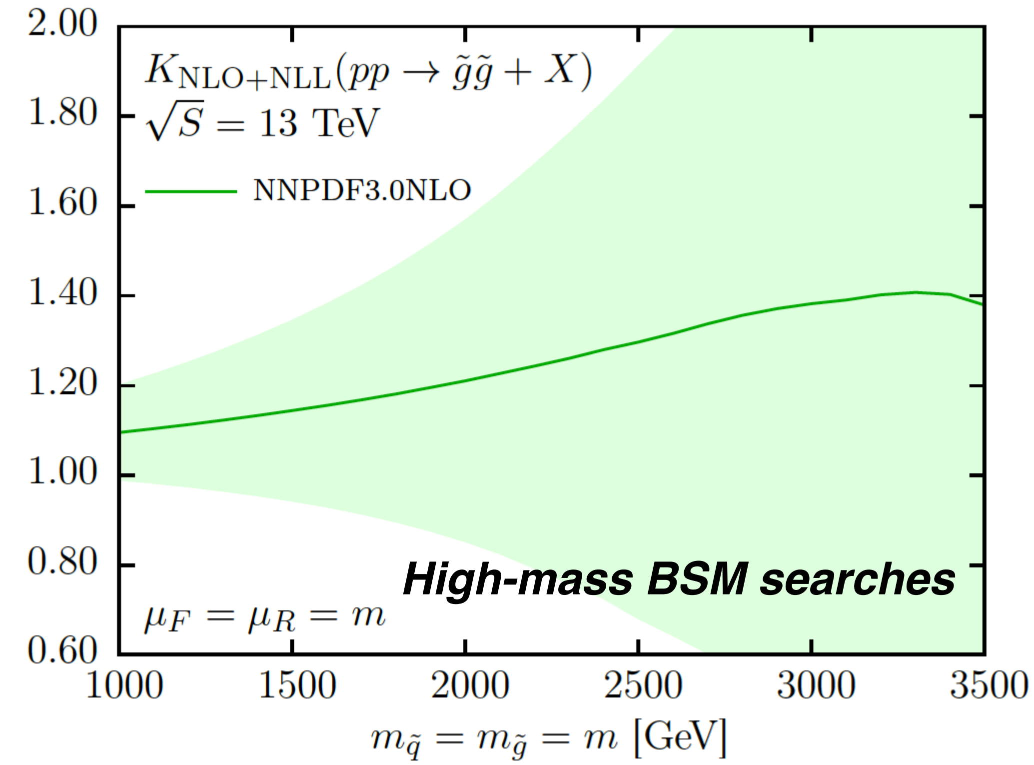
partonic luminosities — *parton distributions*

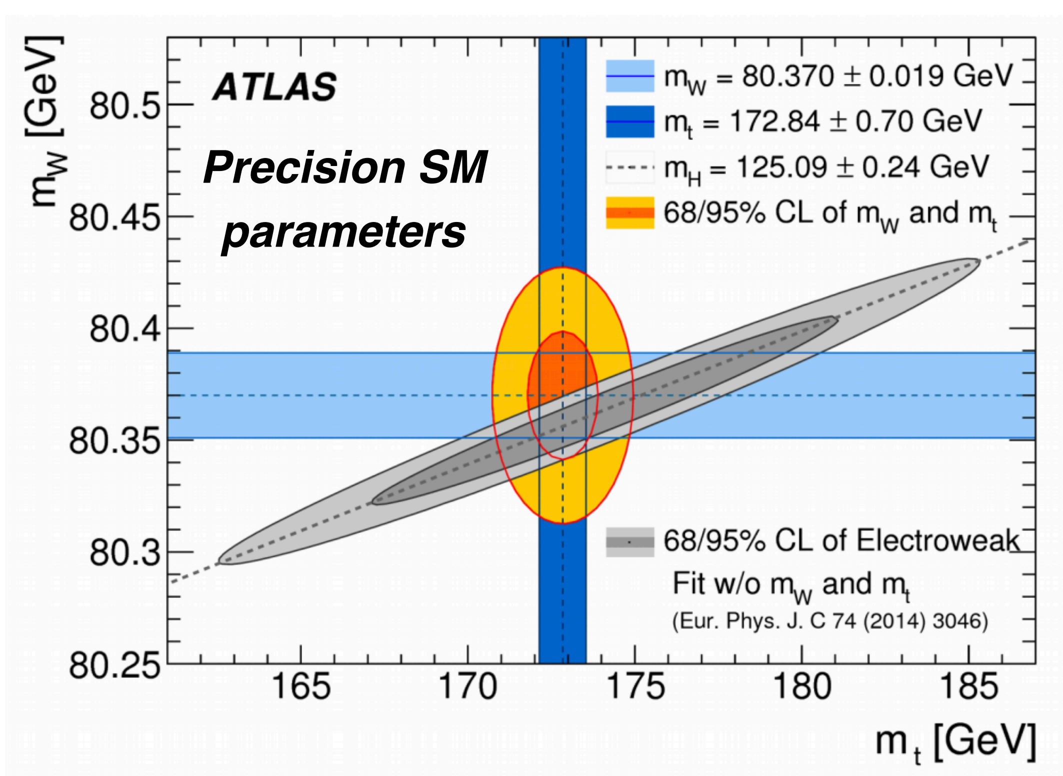
$$\mathscr{L}_{ij}(Q,s) = \frac{1}{s}\int_{Q^2/s}^{1} \frac{dx}{x} f_i\left(\frac{Q^2}{sx},Q\right) f_j\,(x,Q)$$

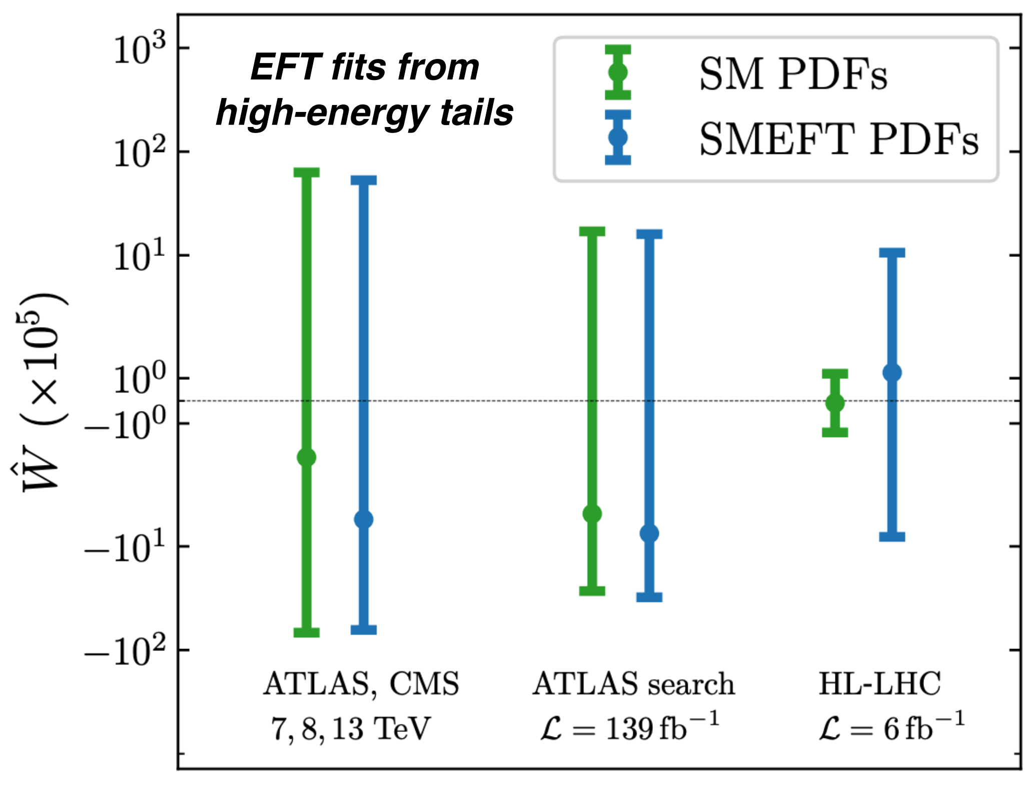
Realising precision physics @ LHC

proton structure uncertainties: limiting factor in **theory interpretation** of LHC analyses



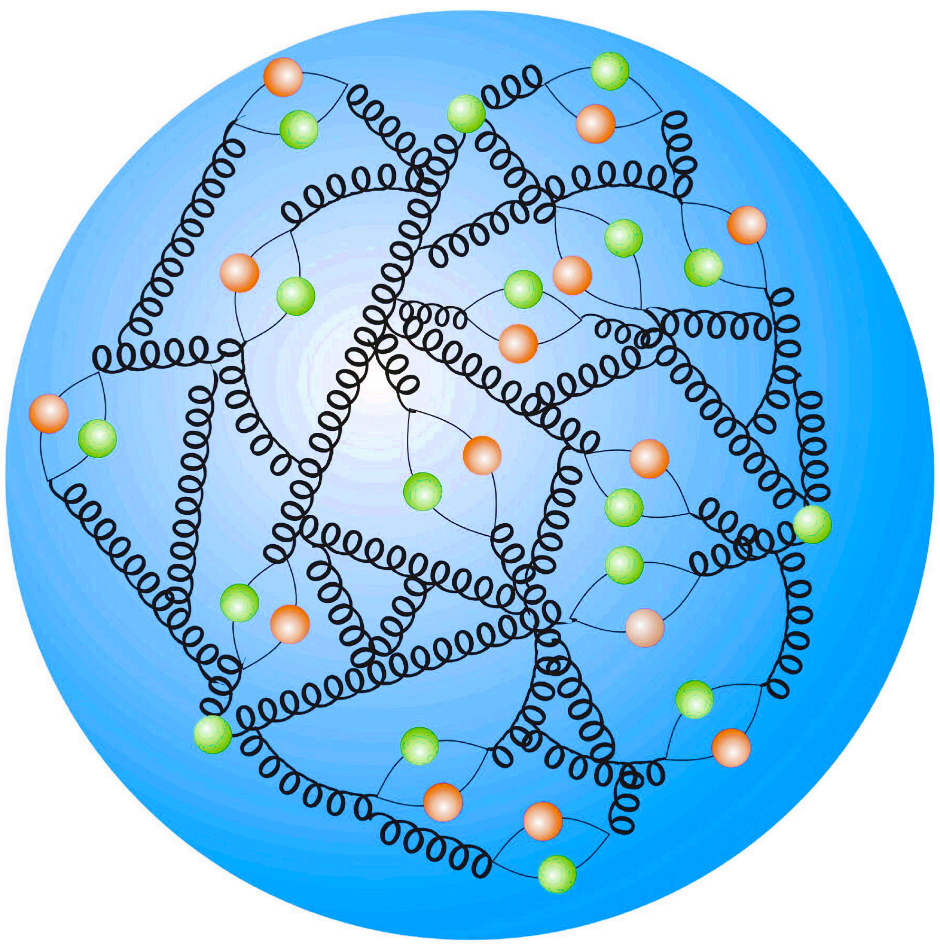






Parton Distributions

Proton energy divided among constituents: **quarks** and **gluons**



Parton Distribution Functions (PDFs)

↓

Determine from **data:**

Global QCD analysis

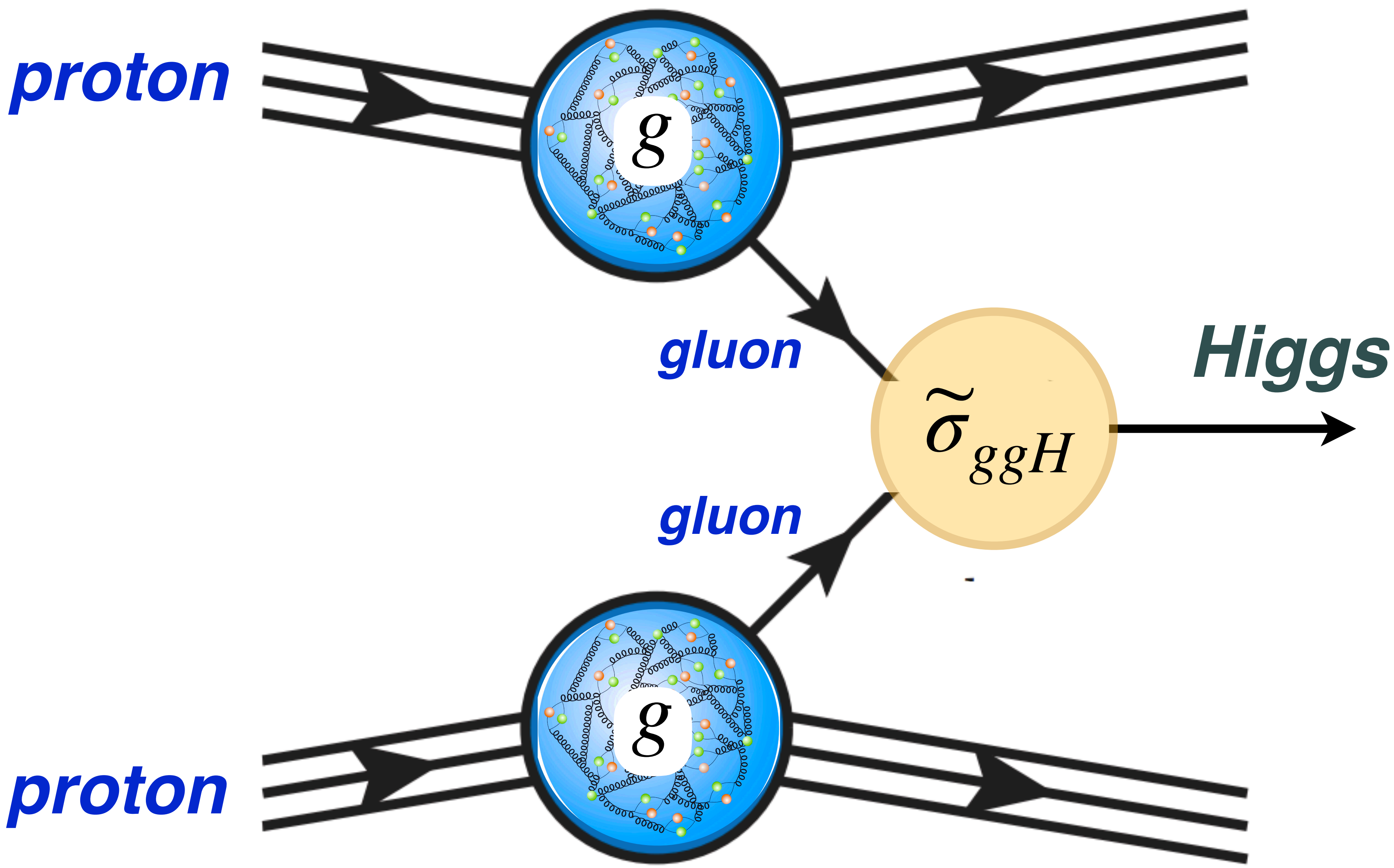
Also important recent results from lattice QCD

What do we need to extract PDFs from data?

Parton Distributions

$$N_{\mathrm{LHC}}(H) \sim g \otimes g \otimes \widetilde{\sigma}_{ggH}$$

Parton Distributions



All-order structure: **QCD factorisation theorems**

Parton Distributions

$$g(x, Q)$$

Probability of **finding a gluon inside a proton**, carrying a fraction ***x*** of the proton momentum, when probed with energy ***Q***

***x*:** fraction of proton momentum carried by gluon

Energy of hard-scattering reaction**:** inverse of resolution length

Dependence on *x* fixed by **non-perturbative QCD dynamics**: extract from experimental data

- **Energy conservation**: momentum sum rule

$$\int_0^1 dx\, x \left(\sum_{i=1}^{n_f} \left[q_i((x, Q^2) + \bar{q}_i(x, Q^2) \right] + g(x, Q^2) \right) = 1$$

- **Quark number conservation**: valence sum rules

$$\int_0^1 dx\, \left(u(x, Q^2) + \bar{u}(x, Q^2) \right) = 2$$

Parton Distributions

$$g(x, Q)$$

Probability of **finding a gluon inside a proton**, carrying a fraction ***x*** of the proton momentum, when probed with energy ***Q***

x: fraction of proton momentum carried by gluon

Energy of hard-scattering reaction**:** inverse of resolution length

Dependence on ***Q*** fixed by **perturbative QCD dynamics**: computed up to $\mathcal{O}\left(\alpha_s^4\right)$

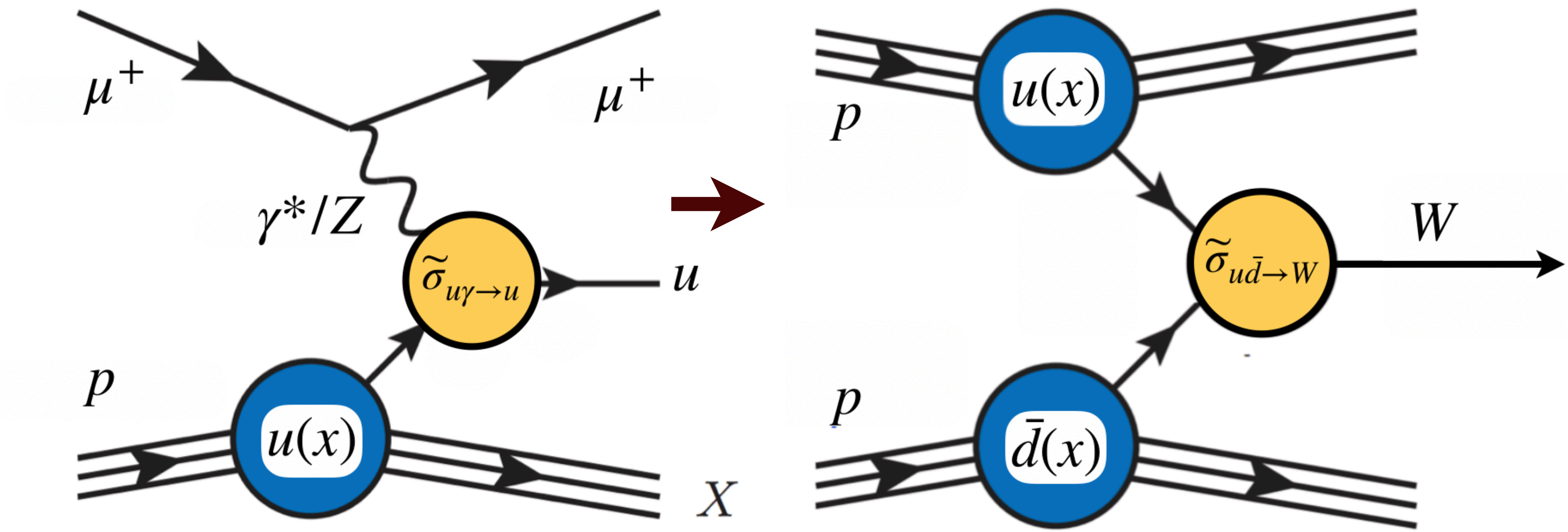
$$\frac{\partial}{\partial \ln Q^2} q_i(x, Q^2) = \int_x^1 \frac{dz}{z} P_{ij}\left(\frac{x}{z}, \alpha_s(Q^2)\right) q_j(z, Q^2)$$

***DGLAP** parton evolution equations*

The global QCD analysis paradigm

QCD factorisation theorems: **PDF universality**

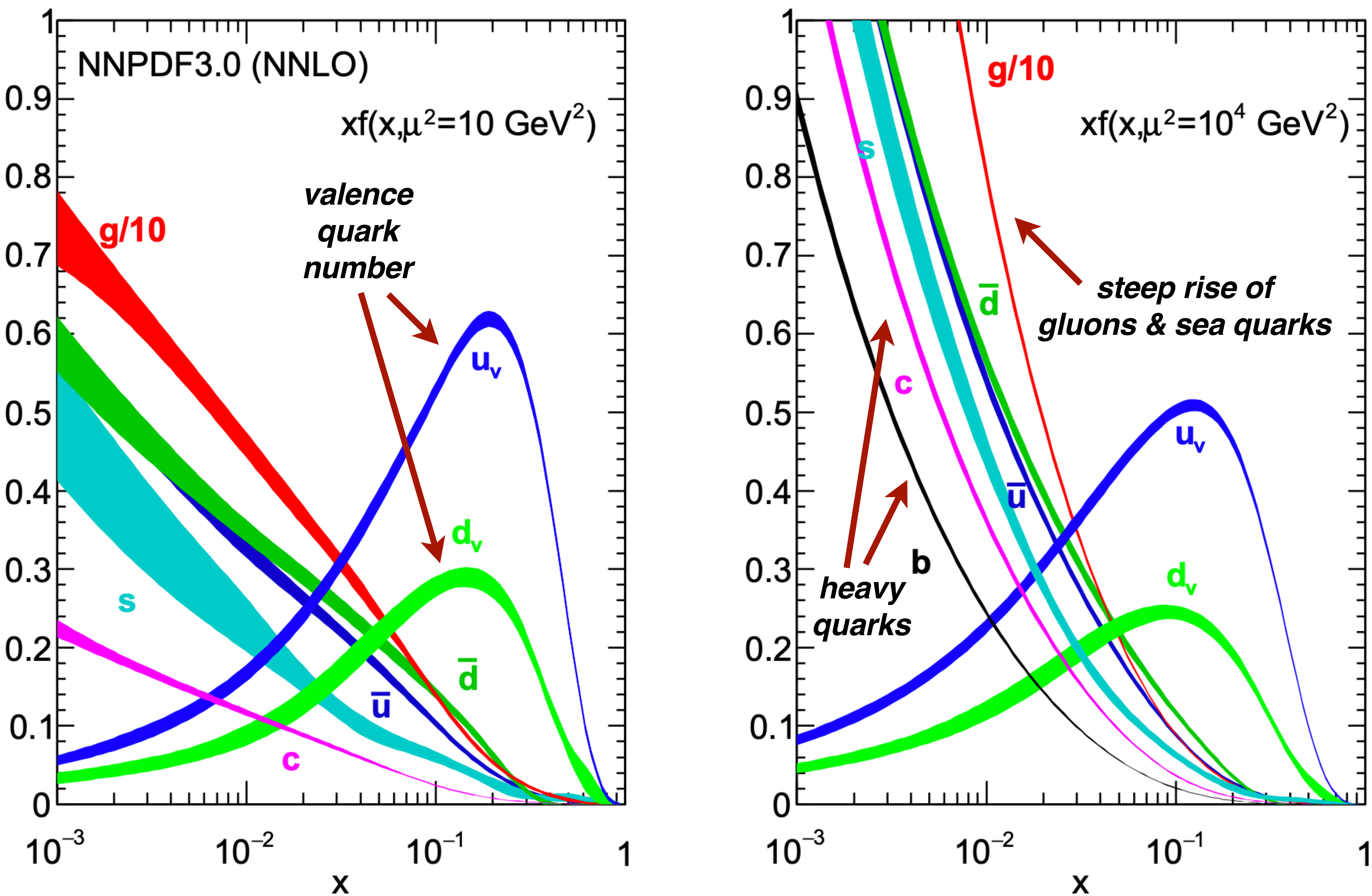
$$\sigma_{l\,p\to\mu\,X} = \widetilde{\sigma}_{u\gamma\to u} \otimes u(x) \quad \rightarrow \quad \sigma_{p\,p\to W} = \widetilde{\sigma}_{u\bar{d}\to W} \otimes u(x) \otimes \bar{d}(x)$$



Determine PDFs from ***deep-inelastic scattering****…*

… and use them to compute predictions for ***proton-proton collisions***

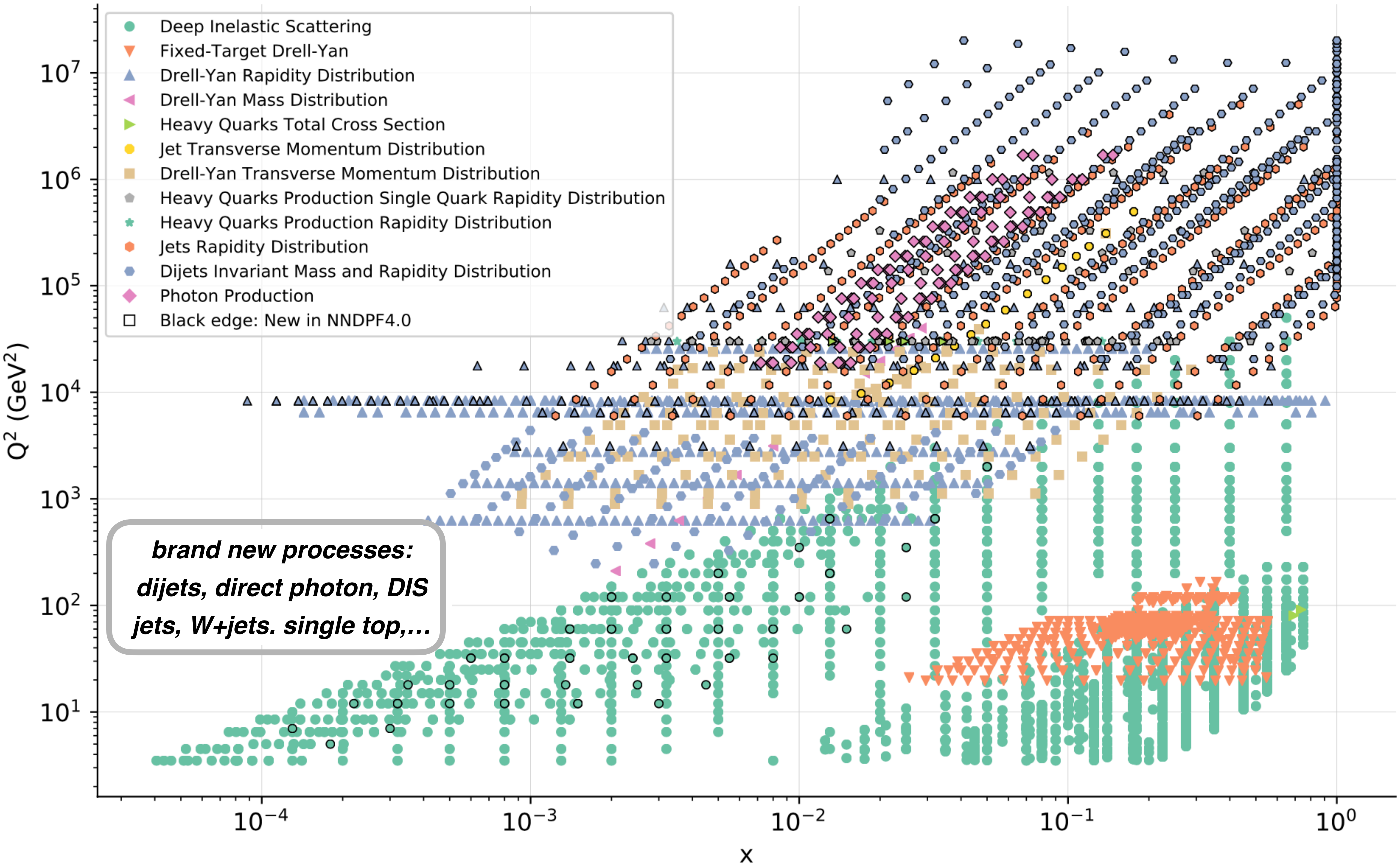
A proton structure snapshop



NNPDF4.0:

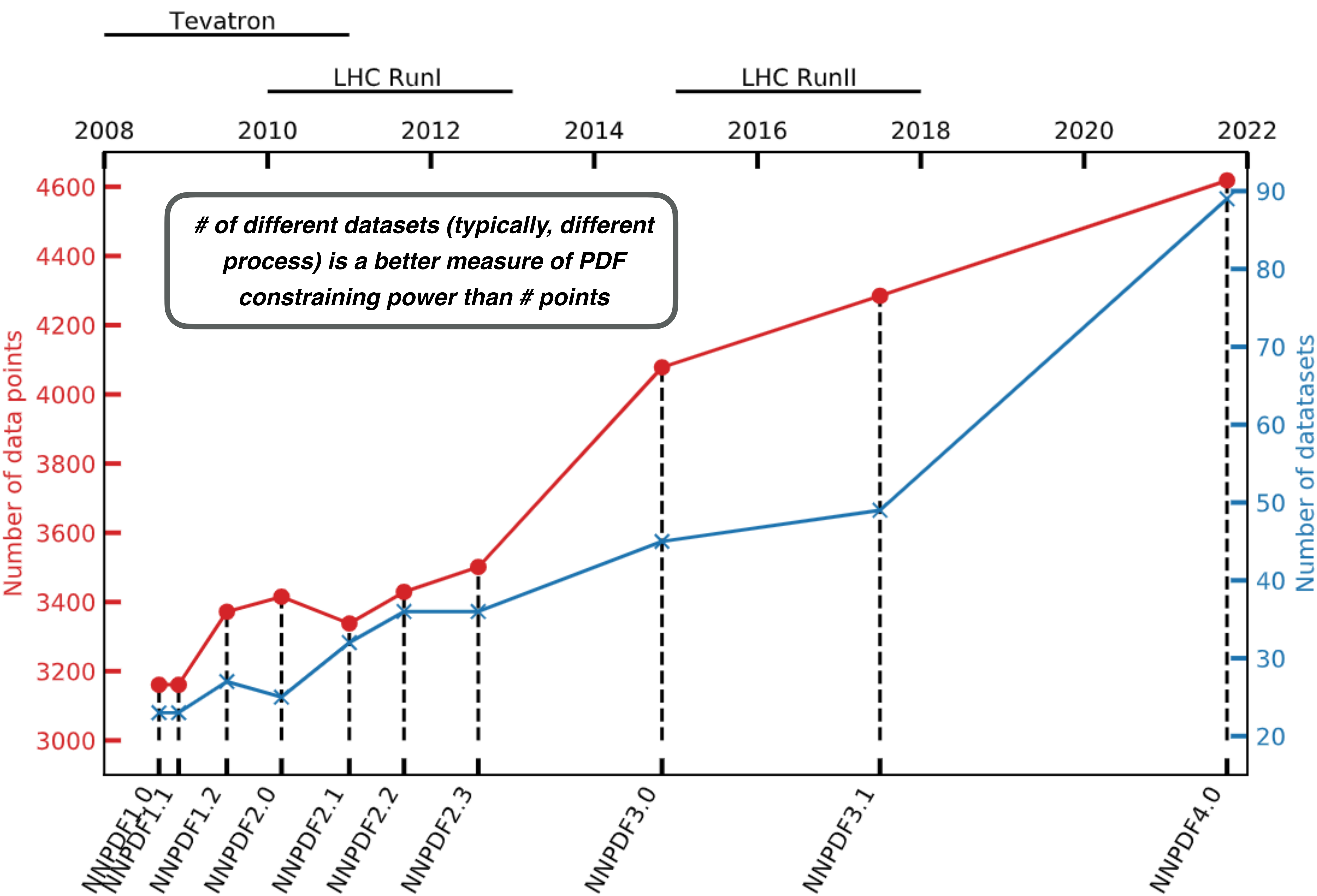
Setup & Validation

The NNPDF4.0 dataset



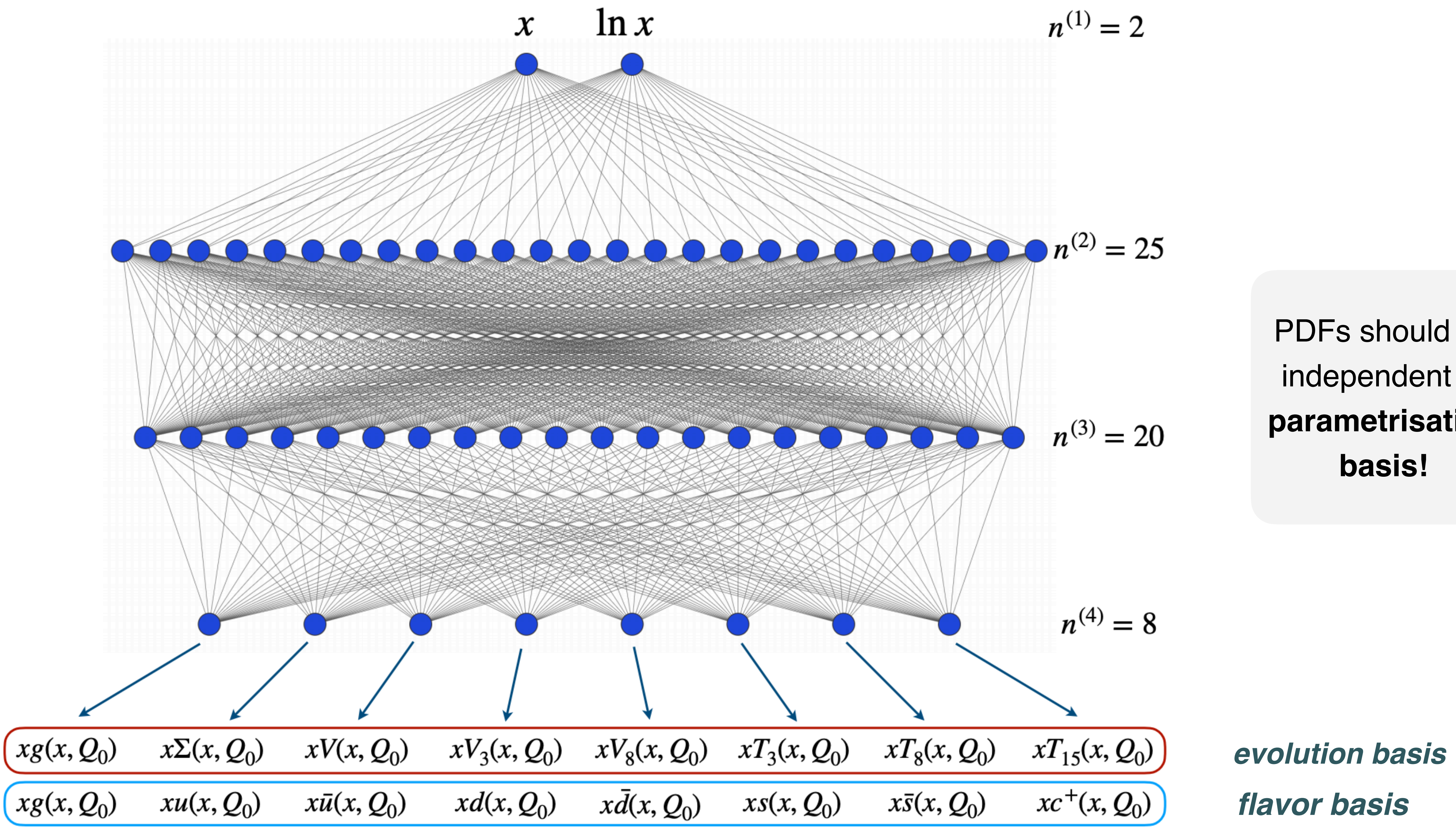
$\mathcal{O}(50)$ data sets investigated; $\mathcal{O}(400)$ data points more in NNPDF4.0 than in NNPDF3.1

From NNPDF1.0 to NNPDF4.0



Improved fitting methodology

- ☑ **Stochastic Gradient Descent** via TensorFlow for neural network training
- ☑ Automated model **hyperparameter optimisation:** NN architecture, minimiser, learning rates …
- ☑ Validation with **future tests** (forecasting new datasets) and **closure tests** (data based on known PDFs)



PDFs should be independent of **parametrisation basis!**

Improved fitting methodology

- ☑ **Stochastic Gradient Descent** via TensorFlow for neural network training
- ☑ Automated model **hyperparameter optimisation:** NN architecture, minimiser, learning rates …
- ☑ Validation with **future tests** (forecasting new datasets) and **closure tests** (data based on known PDFs)

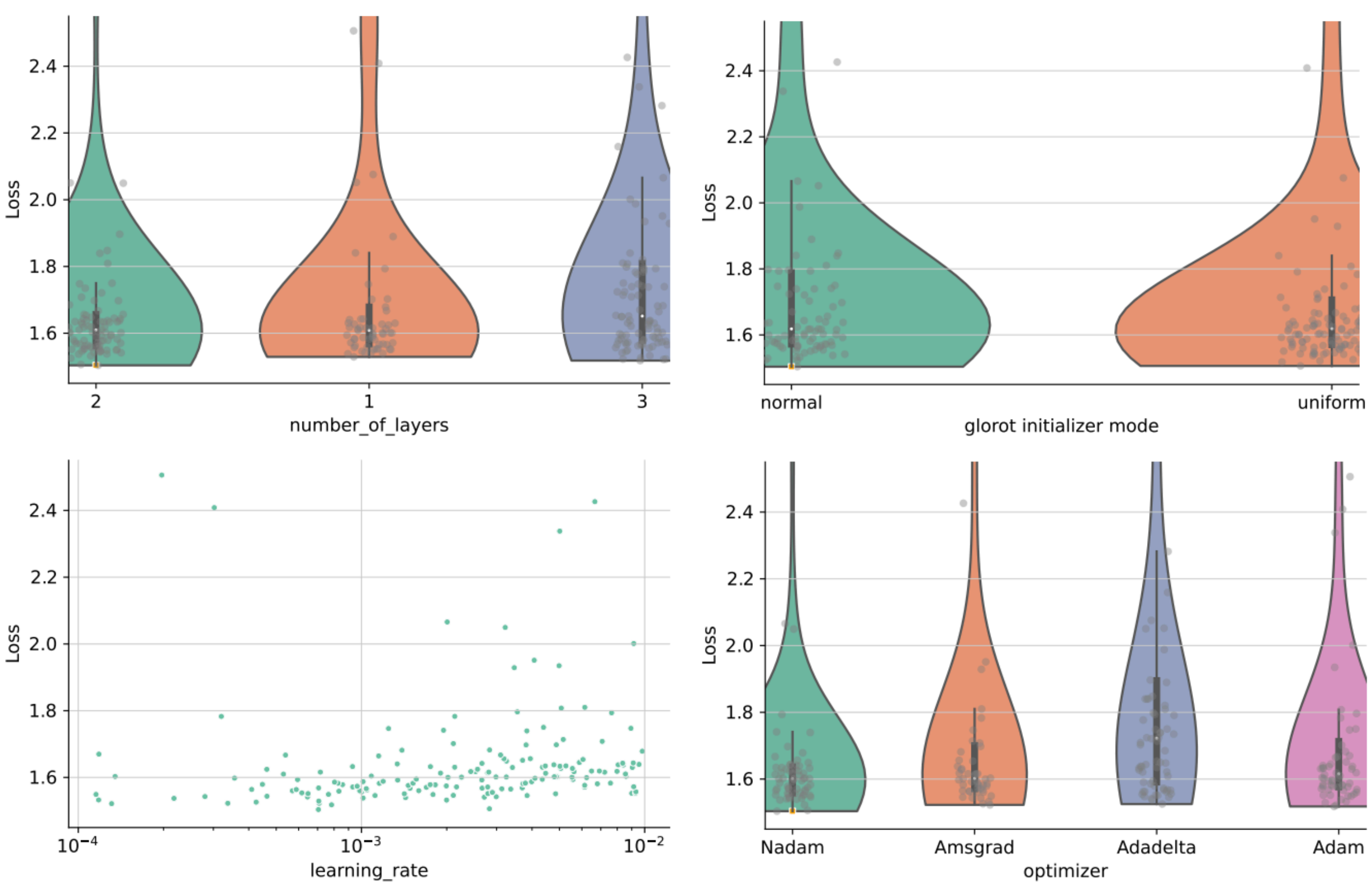
Loss (``average")

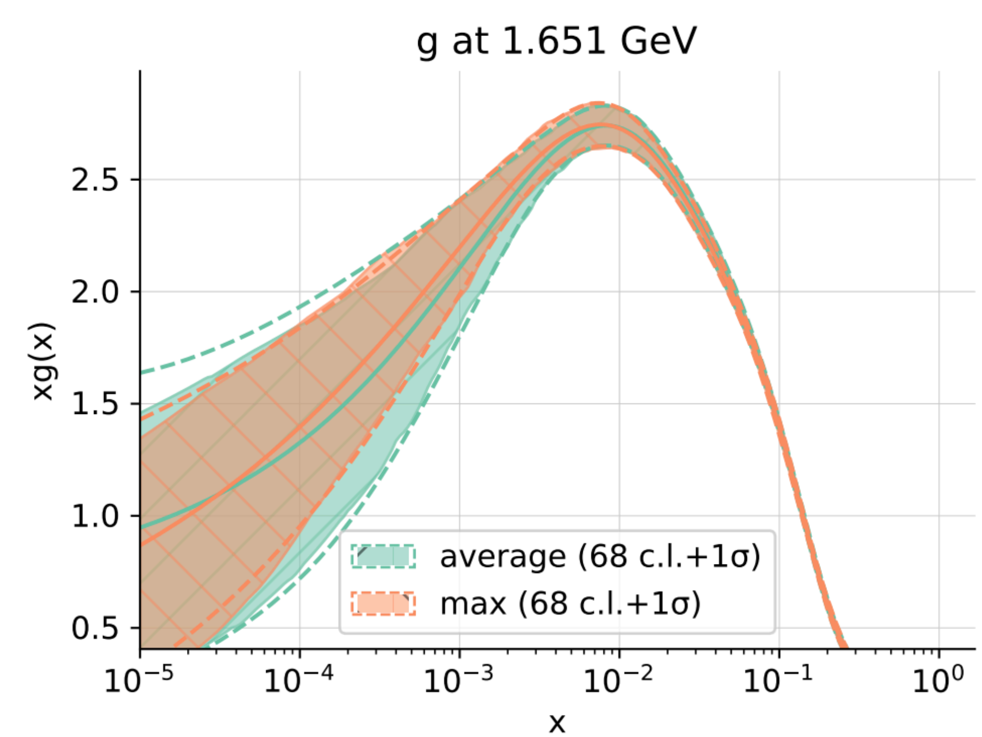
ML model hyperparams

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\left(\frac{1}{n_{\rm fold}}\sum_{k=1}^{n_{\rm fold}}\chi^2_k(\boldsymbol{\theta})\right)$$

Loss (``max")

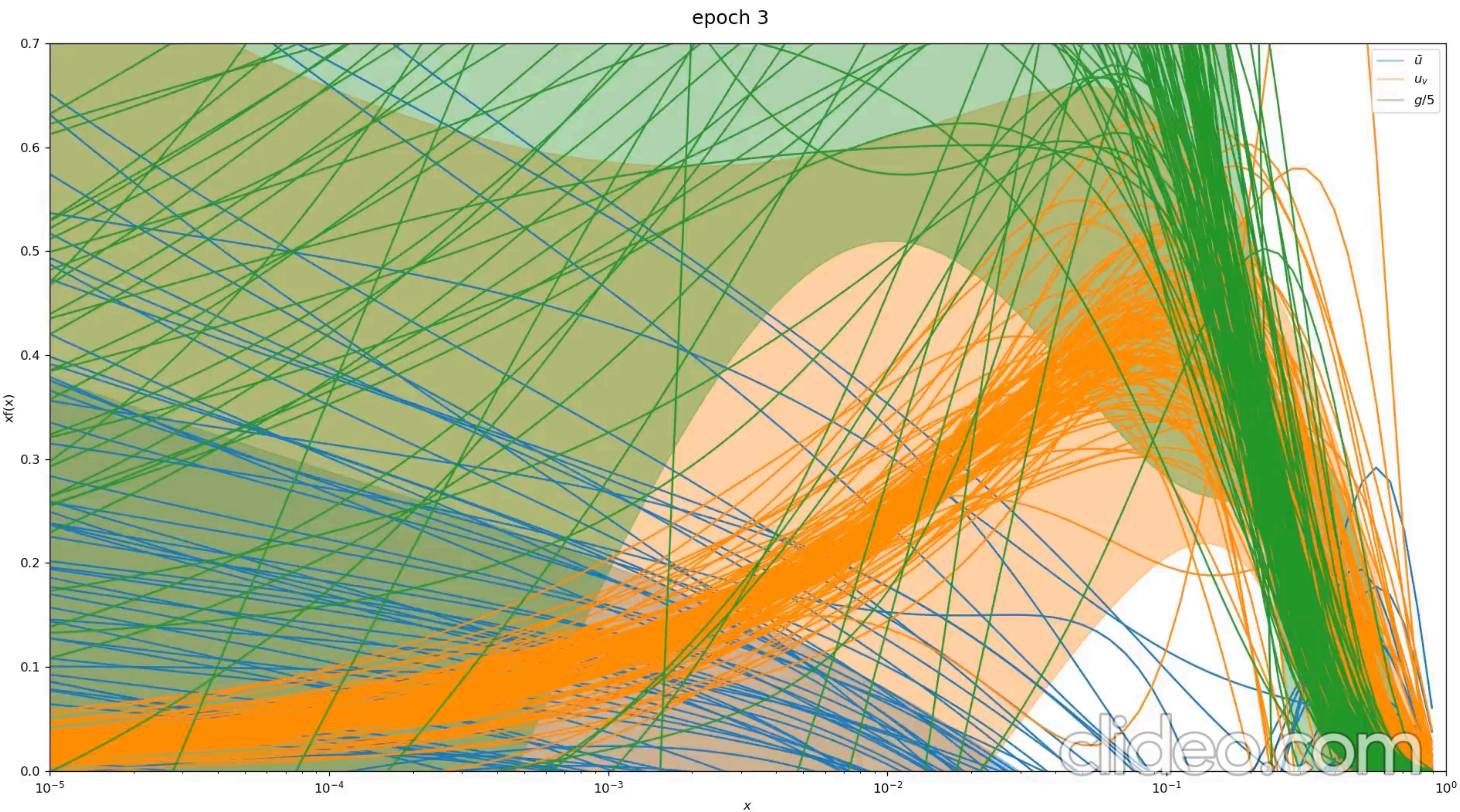
$$L = \max\left(\chi^2_1, \chi^2_2, \chi^2_3, \ldots, \chi^2_{n_{\rm fold}}\right)$$





Stability wrt hyperopt loss function

Improved fitting methodology



Illustrating the outcome of **SGD minimisation** (band: standard deviation over the MC replicas)

Closure and future tests

Closure tests

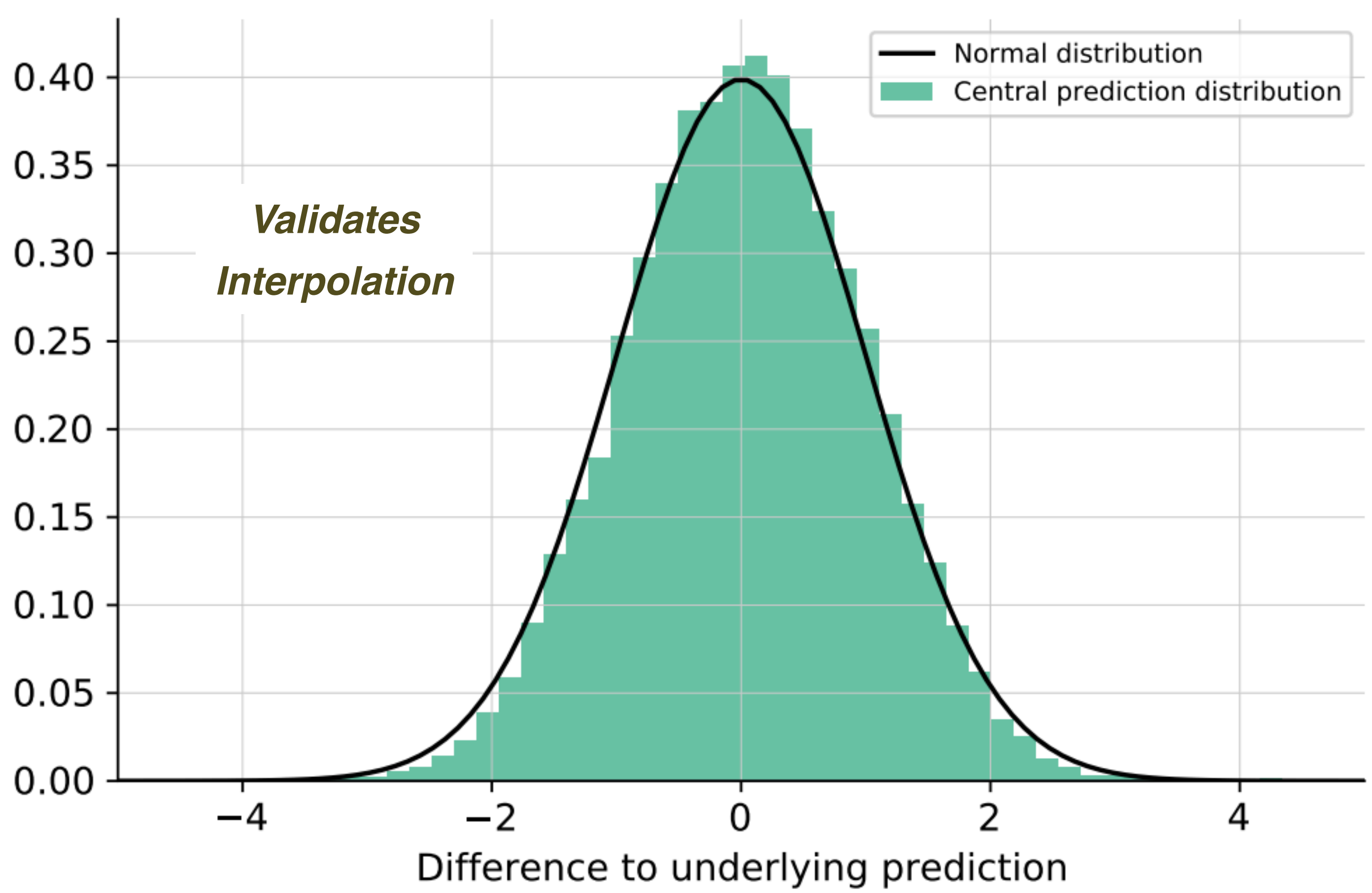
Generate **toy data** based on some known PDF, check *a posteriori* that the **true underlying law is reproduced** within errors

mean NN prediction *true central value*

generate many toys

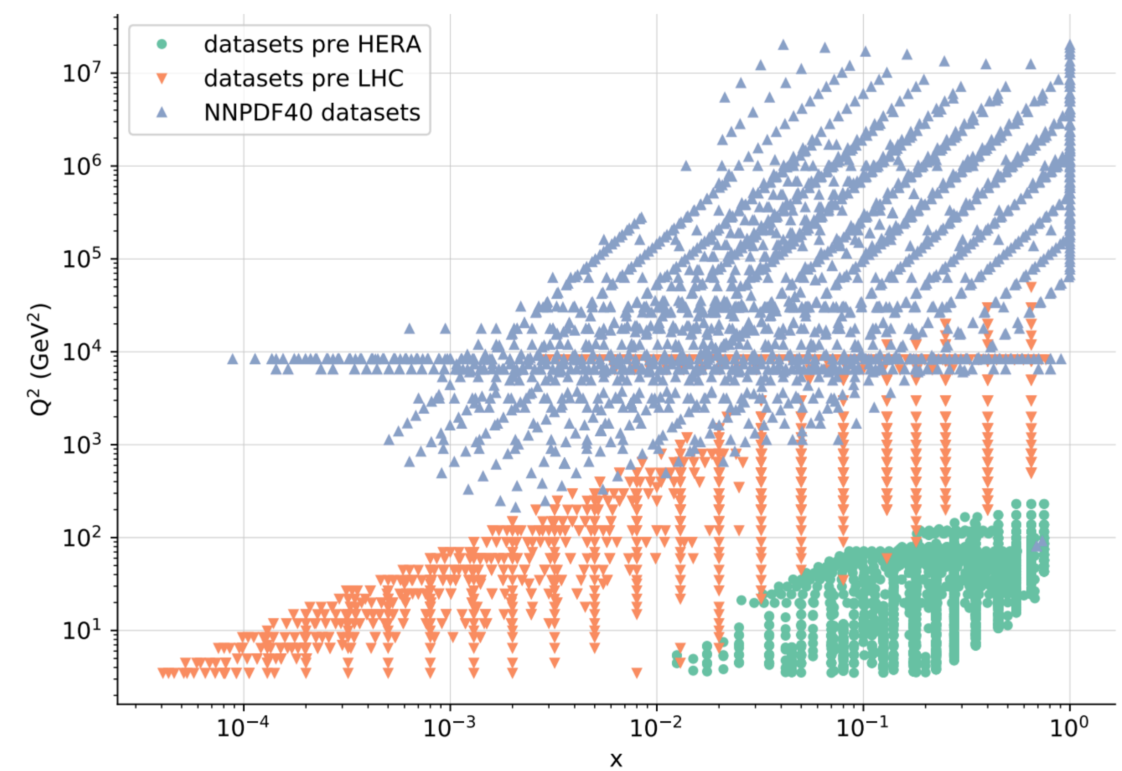
$$\delta_i^{(l)} \equiv \frac{\left(\mathbf{E}_\epsilon\left[g_i\right]^{(l)} - f_i\right)}{\sigma_i^{(l)}}$$

data index *PDF uncertainty*



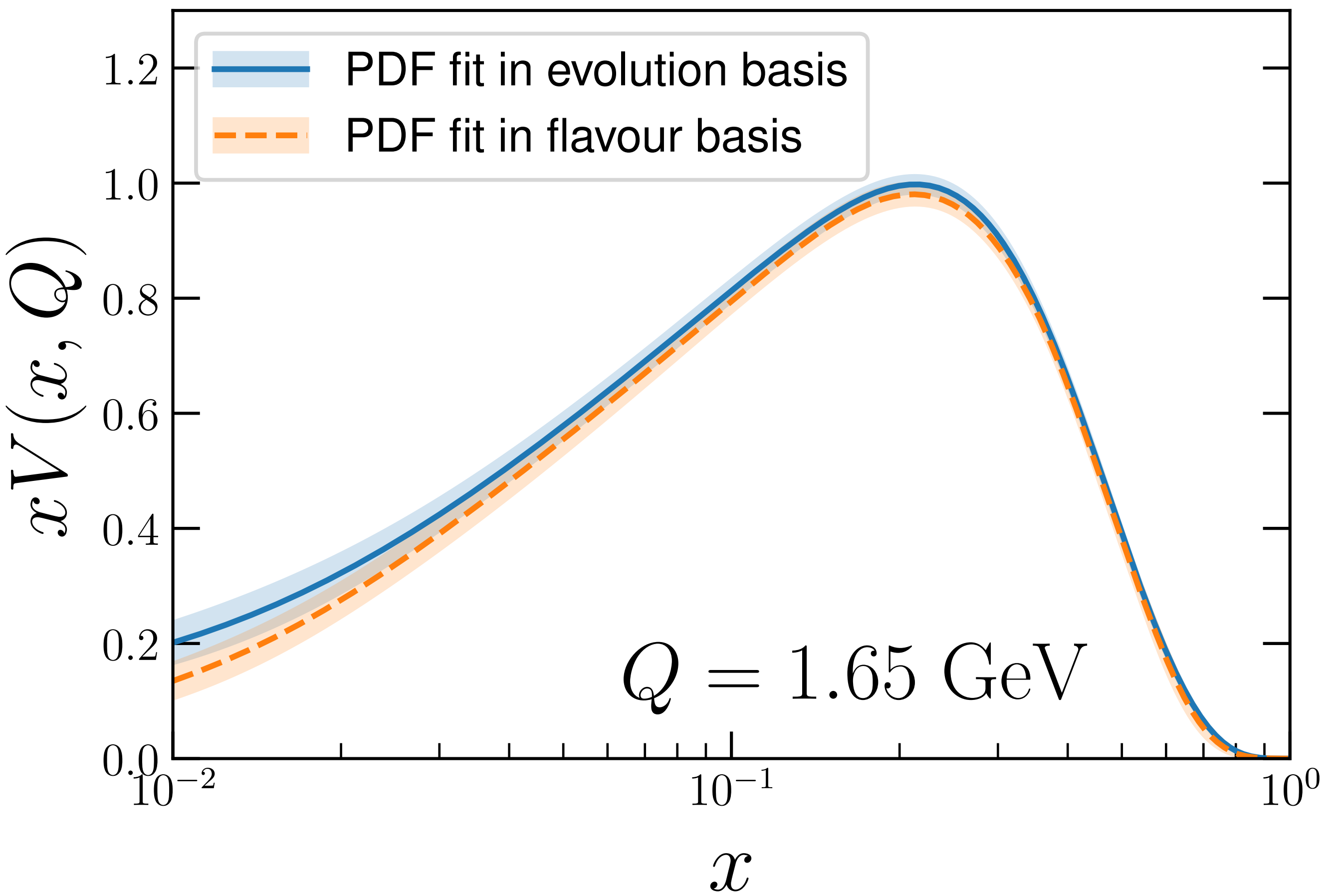
Future tests

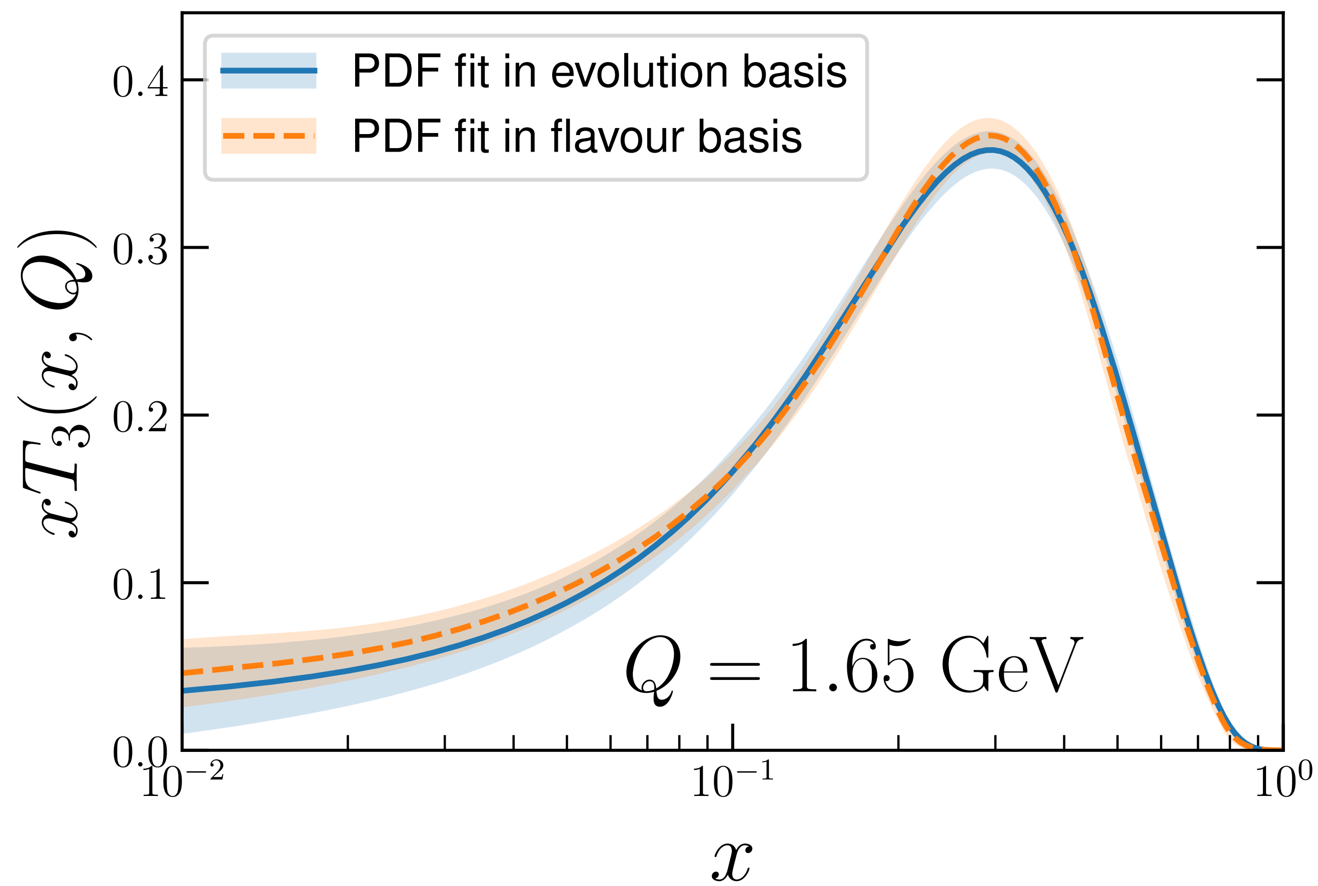
Fit data restricted to specific kinematic regions, then verify **succesful extrapolation**



Process	χ^2 pre-HERA	χ^2 pre-LHC	χ^2 Global
Fixed target NC DIS	1.05	1.18	1.23
Fixed target CC DIS	0.80	0.85	0.87
Fixed target Drell-Yan	0.92	1.27	1.59
HERA	*27.20* (**1.23**)	1.22	1.20
Collider Drell-Yan (Tevatron)	*5.52* (**1.02**)	0.99	1.11
Collider Drell-Yan (LHC)	*18.91* (**1.31**)	*2.63* (**1.58**)	1.53
Top quark production	*20.01* (**1.06**)	*1.30* (**0.87**)	1.01
Jet production	*2.69* (**0.98**)	*2.12* (**1.10**)	1.26

Parametrisation basis independence





evolution basis PDF parametrisation:

$$xV(x, Q_0) \propto \mathrm{NN}_V(x)$$

$$xT_3(x, Q_0) \propto \mathrm{NN}_{T_3}(x)$$

Radically different strategies to parametrize the **quark PDF flavour combinations** lead to identical results: ultimate test of **parametrisation independence**

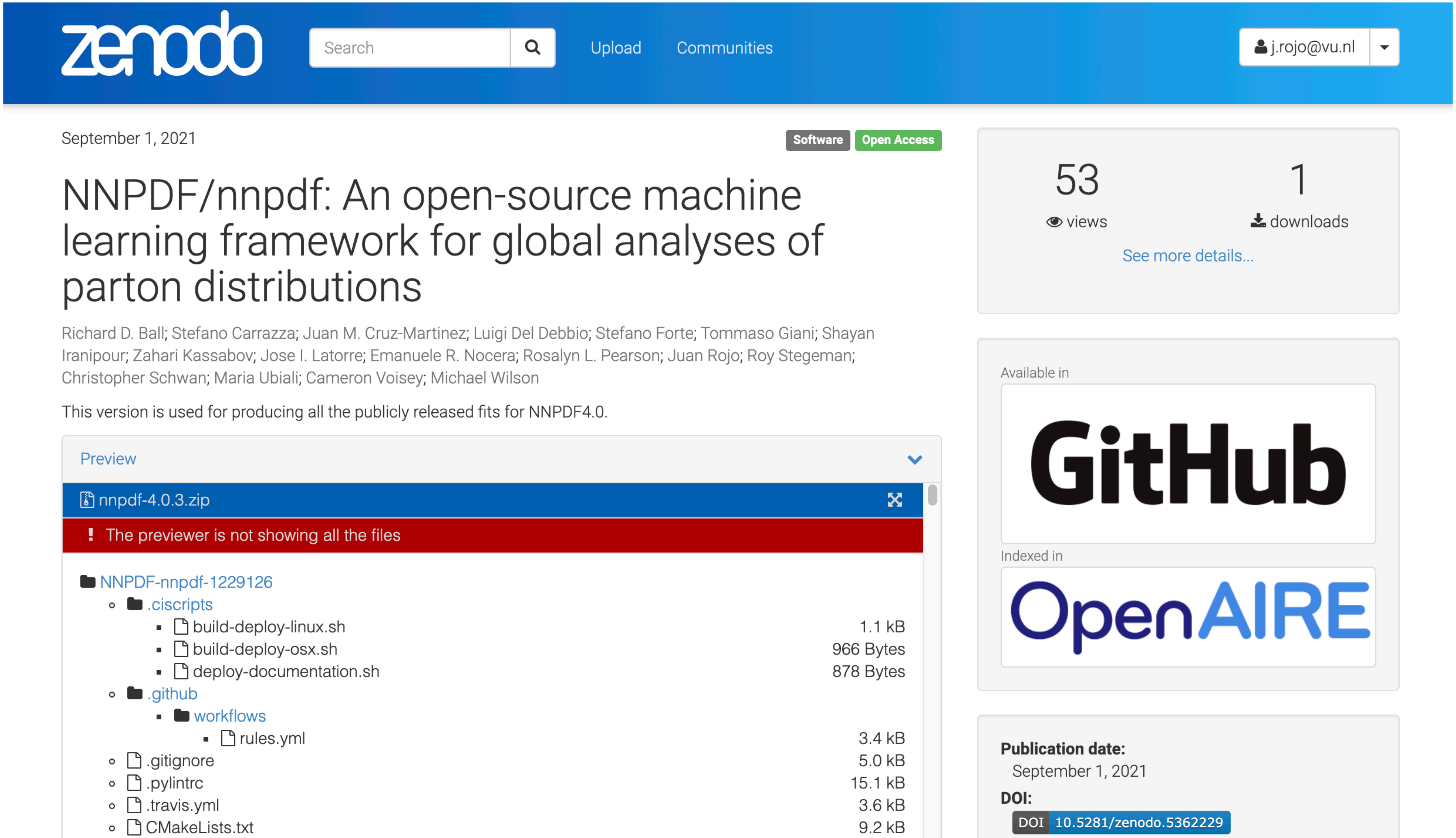
first time ever!

flavour basis PDF parametrisation:

$$xV(x, Q_0) \propto \left(\mathrm{NN}_u(x) - \mathrm{NN}_{\bar{u}}(x) + \mathrm{NN}_d(x) - \mathrm{NN}_{\bar{d}}(x) + \mathrm{NN}_s(x) - \mathrm{NN}_{\bar{s}}(x)\right)$$

$$xT_3(x, Q_0) \propto \left(\mathrm{NN}_u(x) + \mathrm{NN}_{\bar{u}}(x) - \mathrm{NN}_d(x) - \mathrm{NN}_{\bar{d}}(x)\right)$$

A ML open-source QCD fitting framework



The full **NNPDF machine learning fitting framework** has been publicly released open source, together with extensive documentation and user-friendly examples

A ML open-source QCD fitting framework

NNPDF

Search docs

- Getting started
- Fitting code: `n3fit`
- Code for data: validphys
- Handling experimental data: Buildmaster
- Storage of data and theory predictions
- Theory
- Continuous integration and deployment
- Servers
- External codes
- Tutorials
- Adding to the Documentation

» The NNPDF collaboration

View page source

The NNPDF collaboration

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

The NNPDF code

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

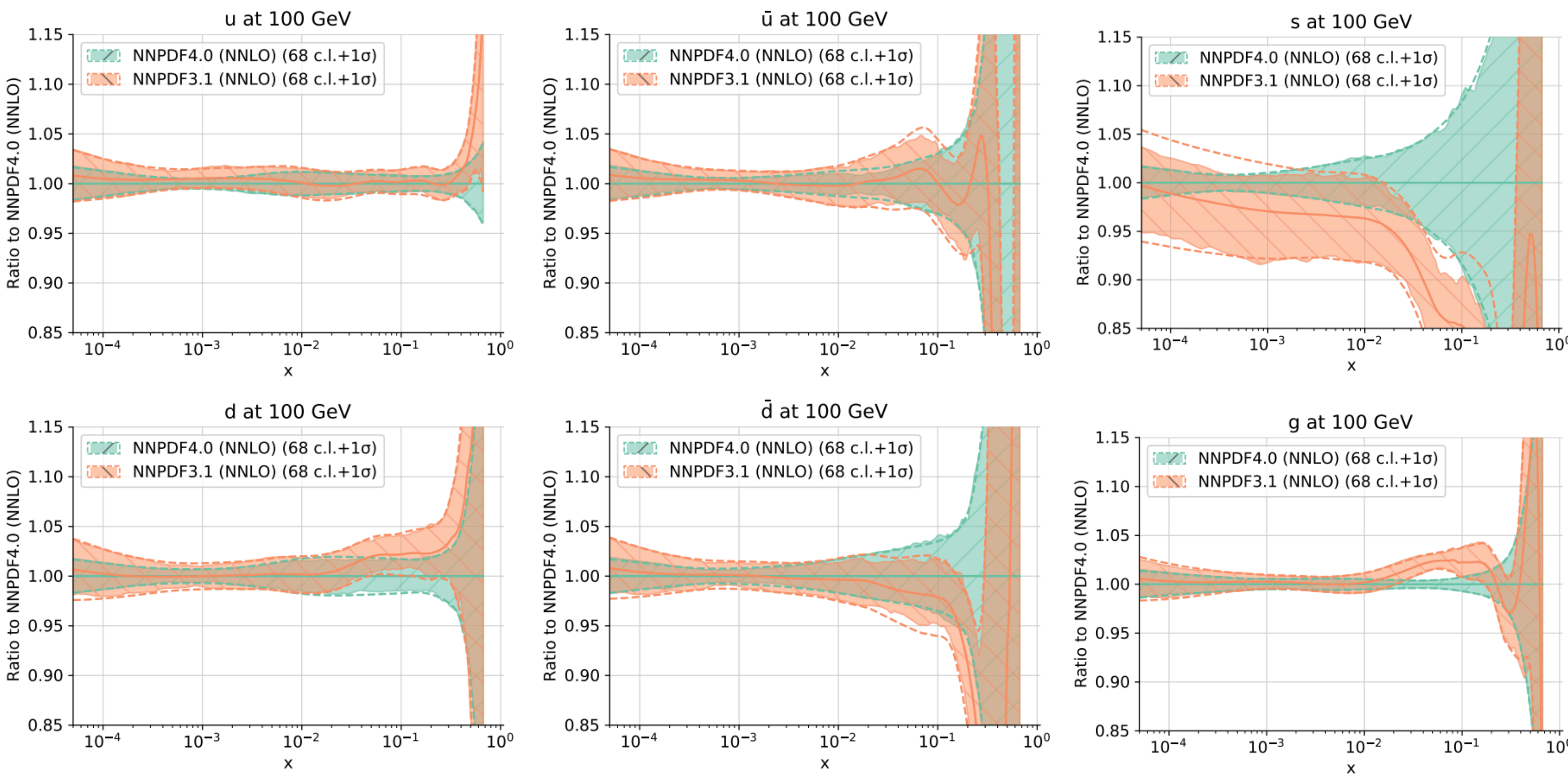
The code can be used to produce the ingredients needed for PDF fits, to run the fits themselves, and to analyse the results. This is the first framework used to produce a global PDF fit made publicly available, enabling for a detailed external validation and reproducibility of the NNPDF4.0 analysis. Moreover, the code enables the user to explore a number of phenomenological applications, such as the assessment of the impact of new experimental data on PDFs, the effect of changes in theory settings on the resulting PDFs and a fast quantitative comparison between theoretical predictions and experimental data over a broad range of observables.

If you are a new user head along to Getting started and check out the Tutorials.

Opportunities for many studies within the **LHC community**: looking forward to suggestions and starting new collaborations!

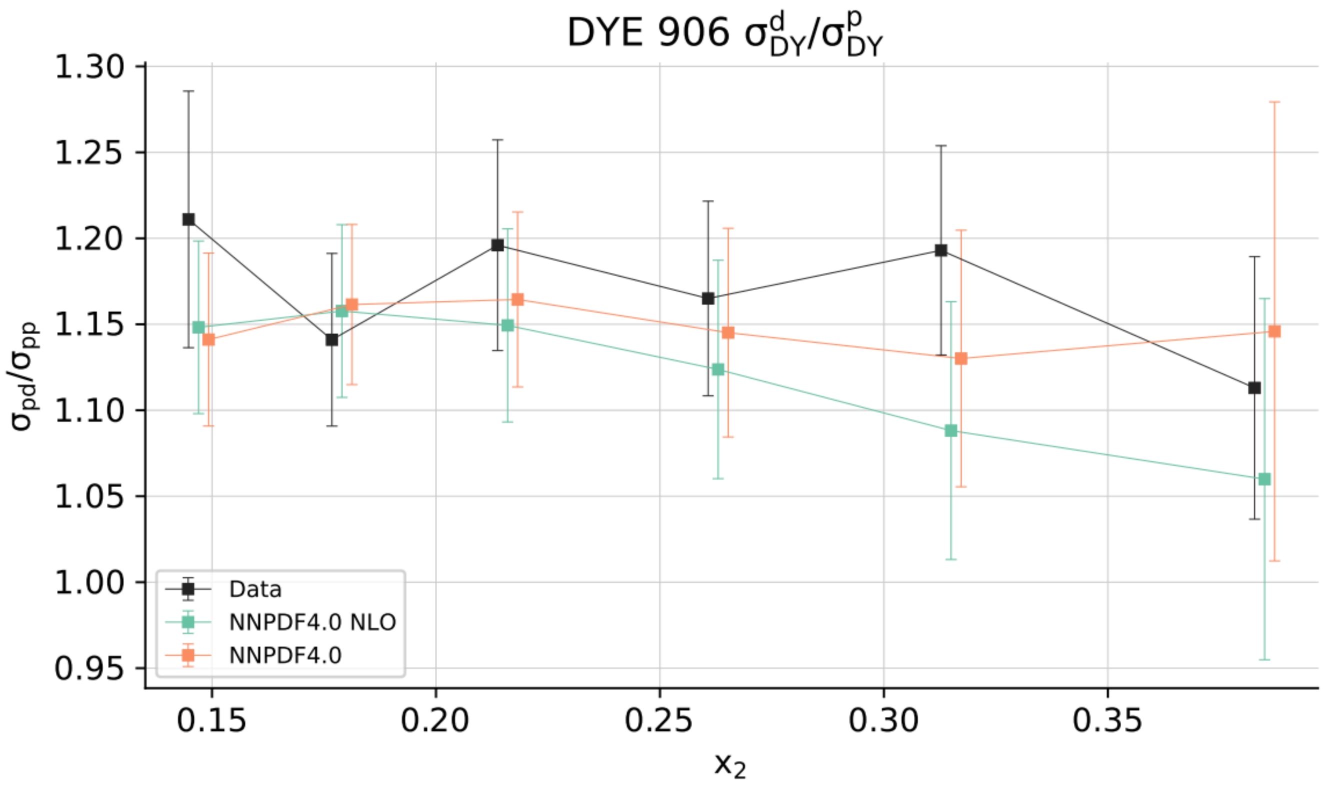
NNPDF4.0: Results

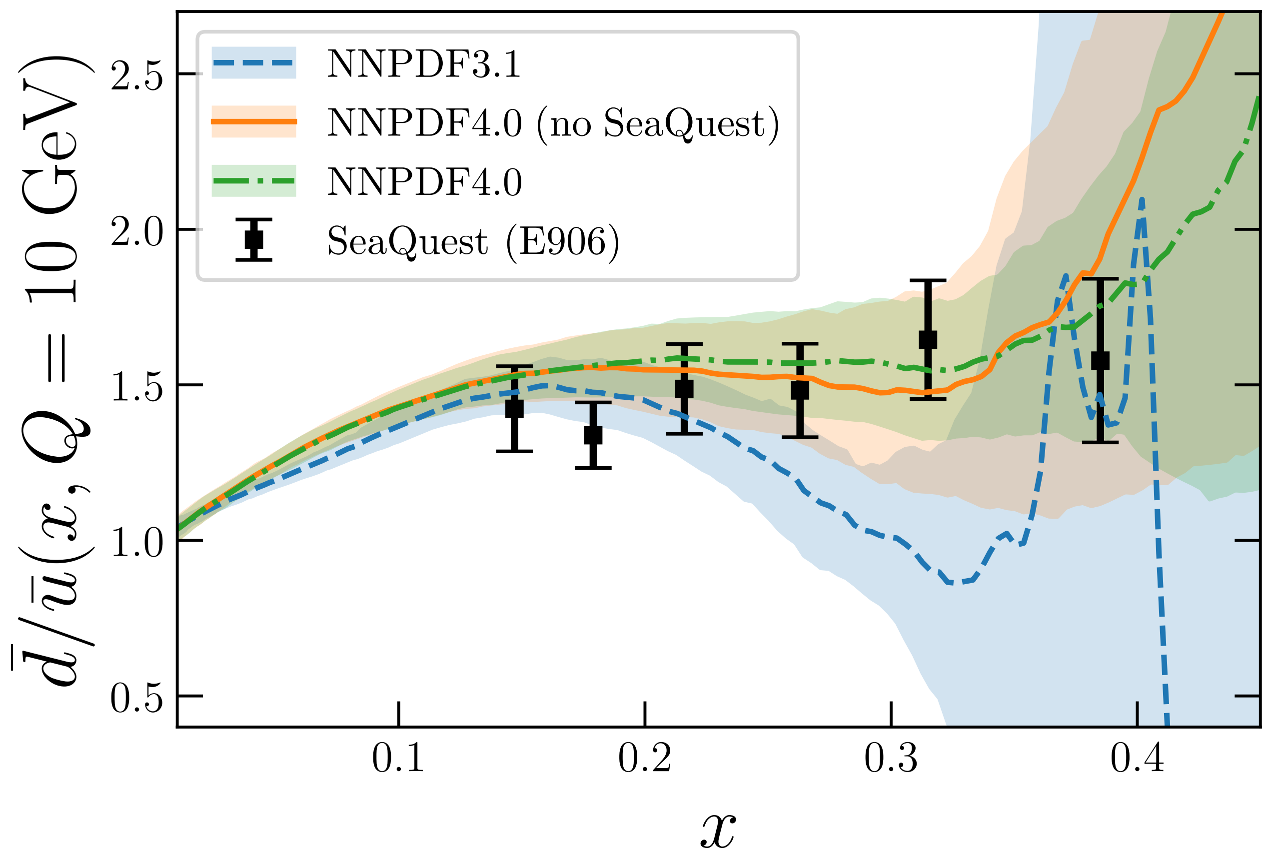
Comparison with NNPDF3.1



- **Good agreement** with NNPDF3.1 within uncertainties, with NNPDF4.0 being more precise
- Differences can be traced back to the **impact of specific datasets** (e.g. dijets for large-x gluon) or improvements in **theory calculations** (e.g. NNLO corrections in dimuon DIS for strangeness)

Antimatter asymmetry





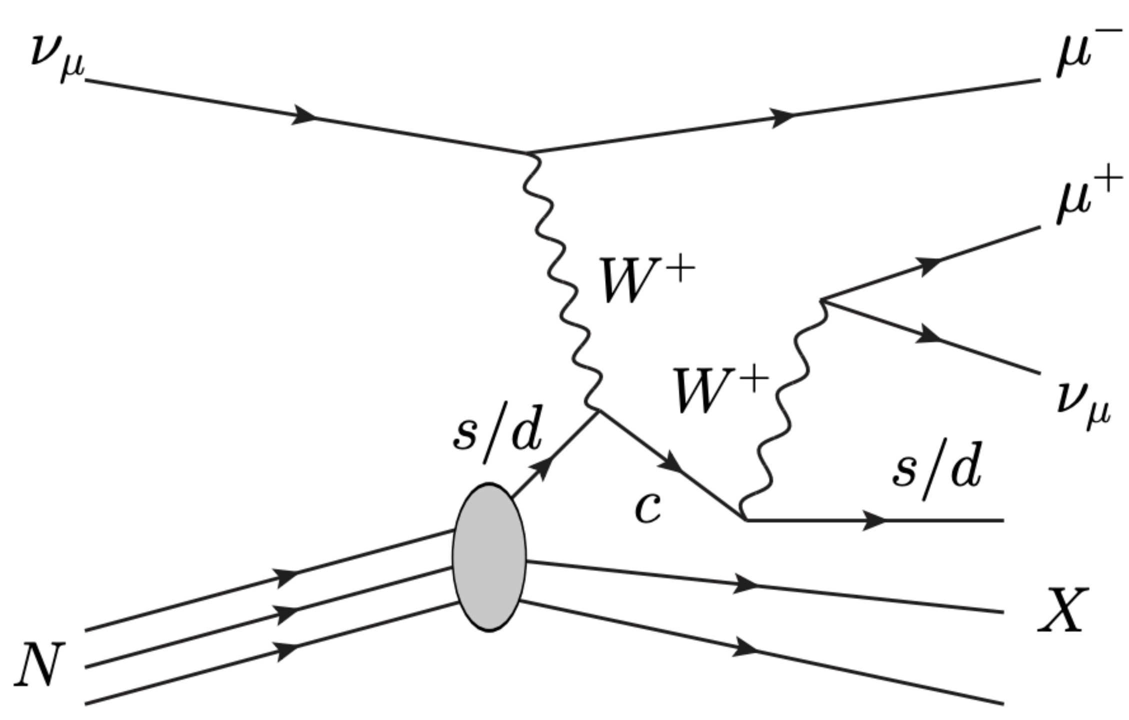
- The recent SeaQuest measurement claims evidence for **quark sea (``proton antimatter") asymmetry**

$$\frac{\sigma_{\mathrm{DY,deuterium}}}{\sigma_{\mathrm{DY,hydrogen}}} \approx 1 + \frac{\bar{d}_p(x_t)}{\bar{u}_p(x_t)}$$

with many caveats!

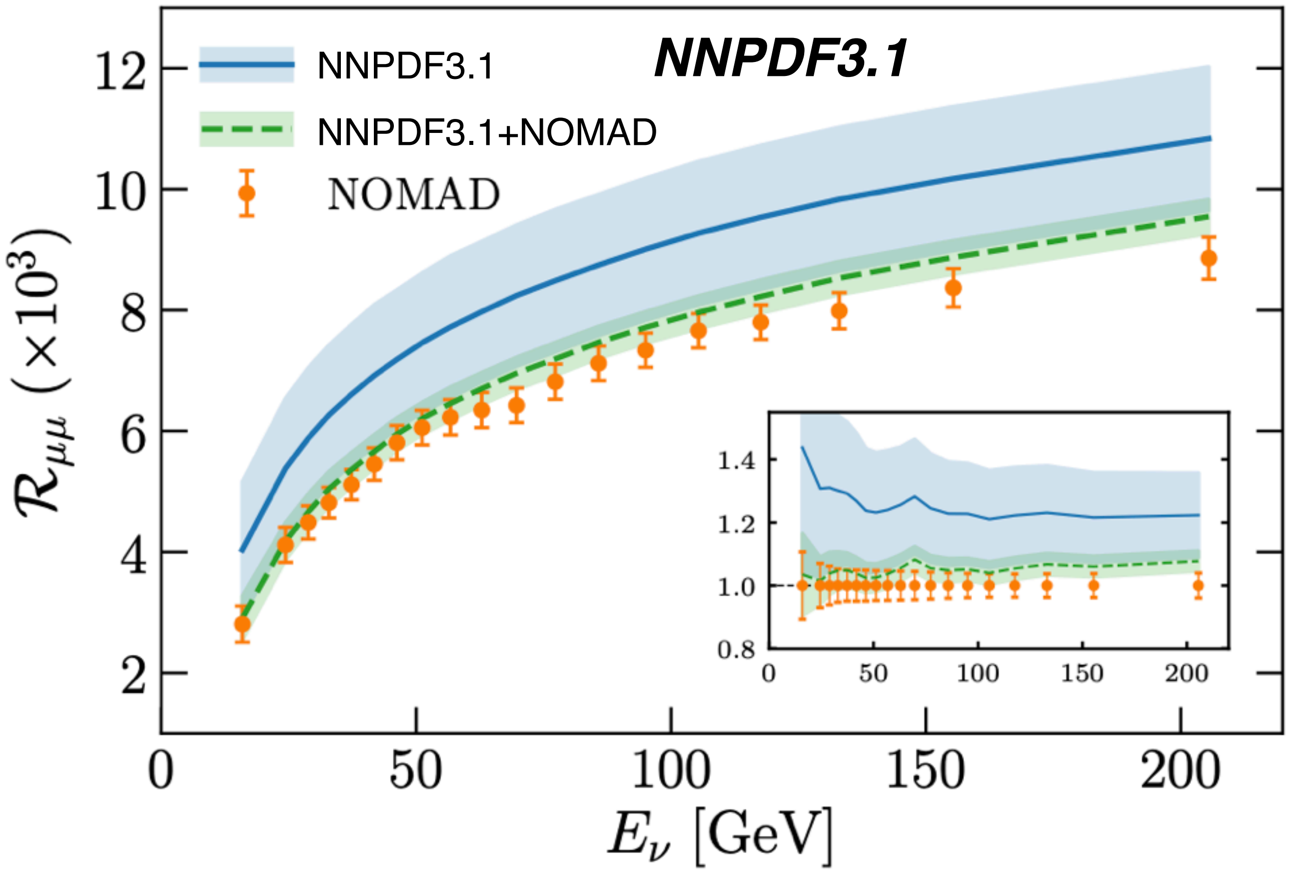
- Actually, SeaQuest further confirms the **global fit prediction**, which agrees with it even when not included
- Already well described by **NNPDF3.1** within uncertainties

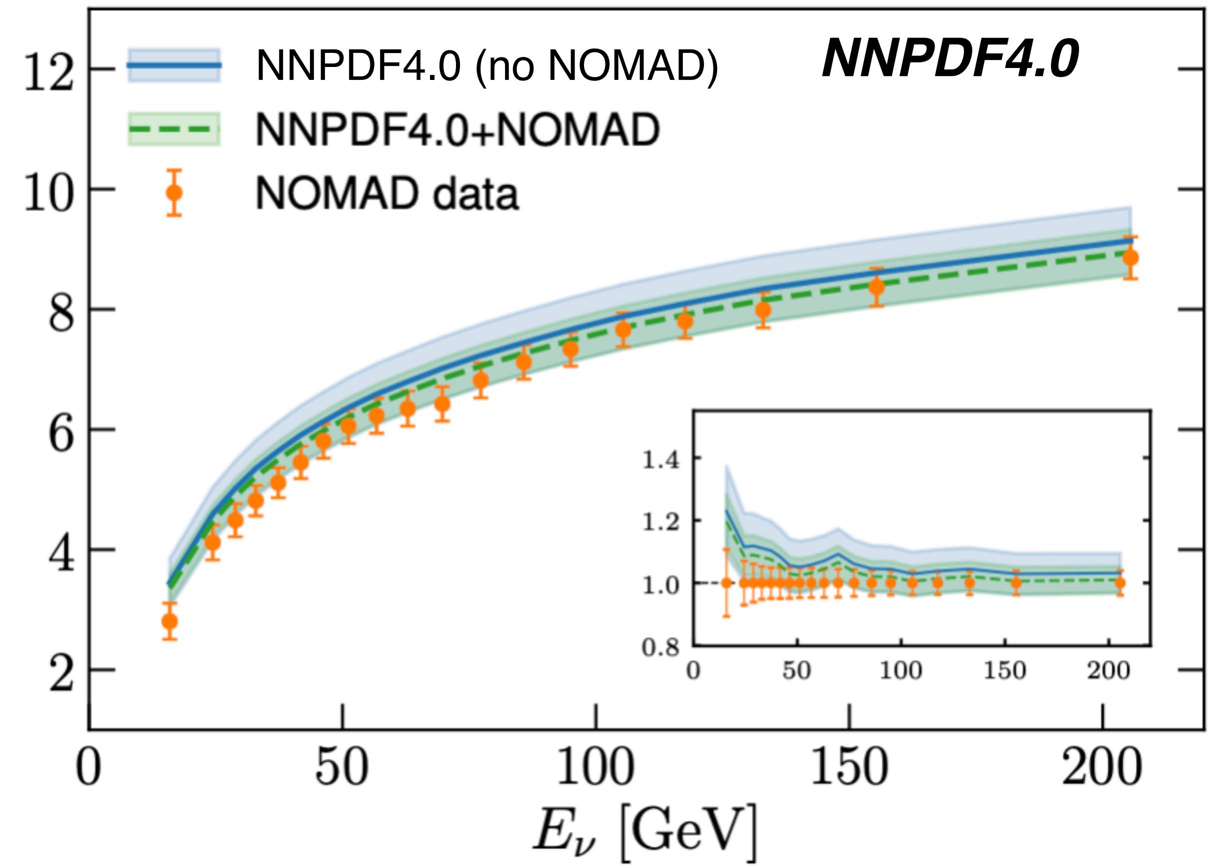
The strangest proton



- ☑ NOMAD dimuon DIS data sensitive to **strangeness** via charged-current scattering
- ☑ Fitting NOMAD had large impact on the strangeness in NNPDF3.1, now in NNPDF4.0 the **no-NOMAD fit** is already spot on the data

Excellent consistency of global dataset



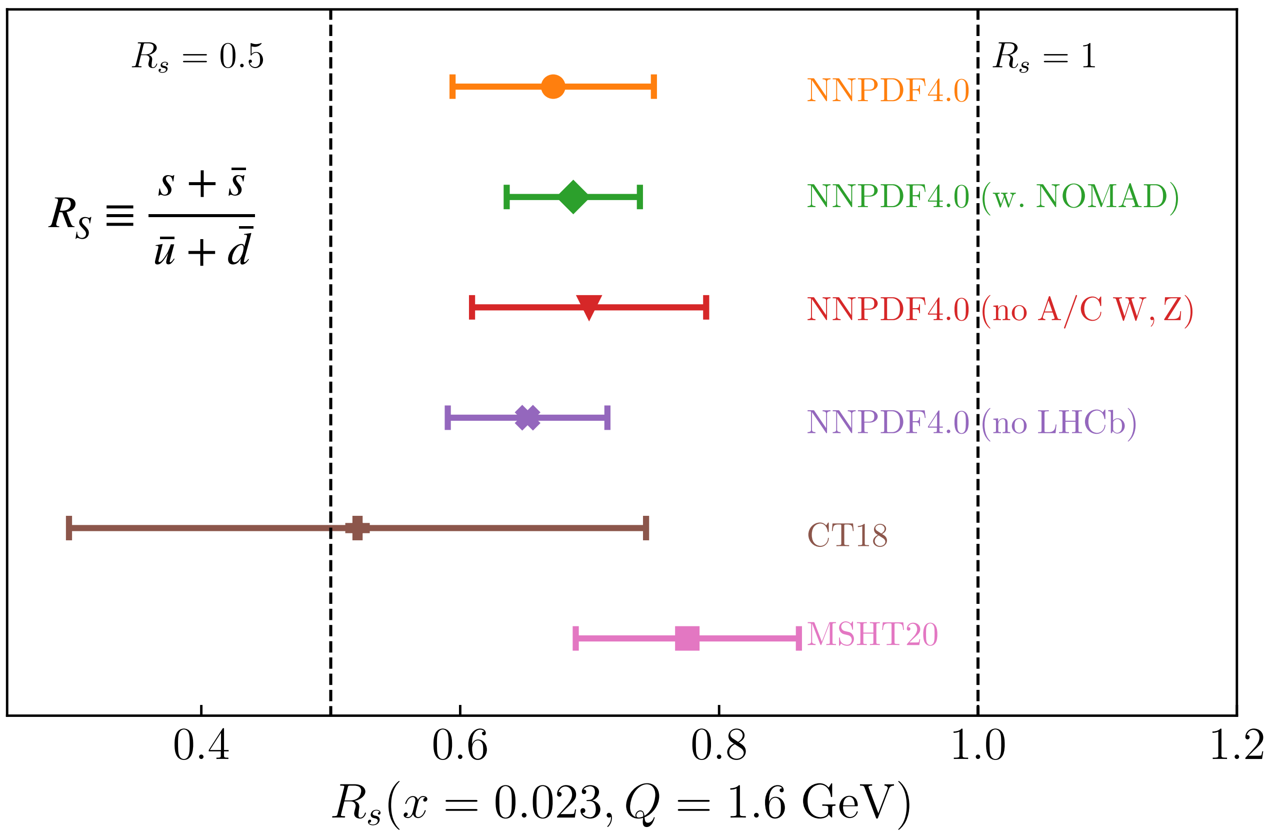


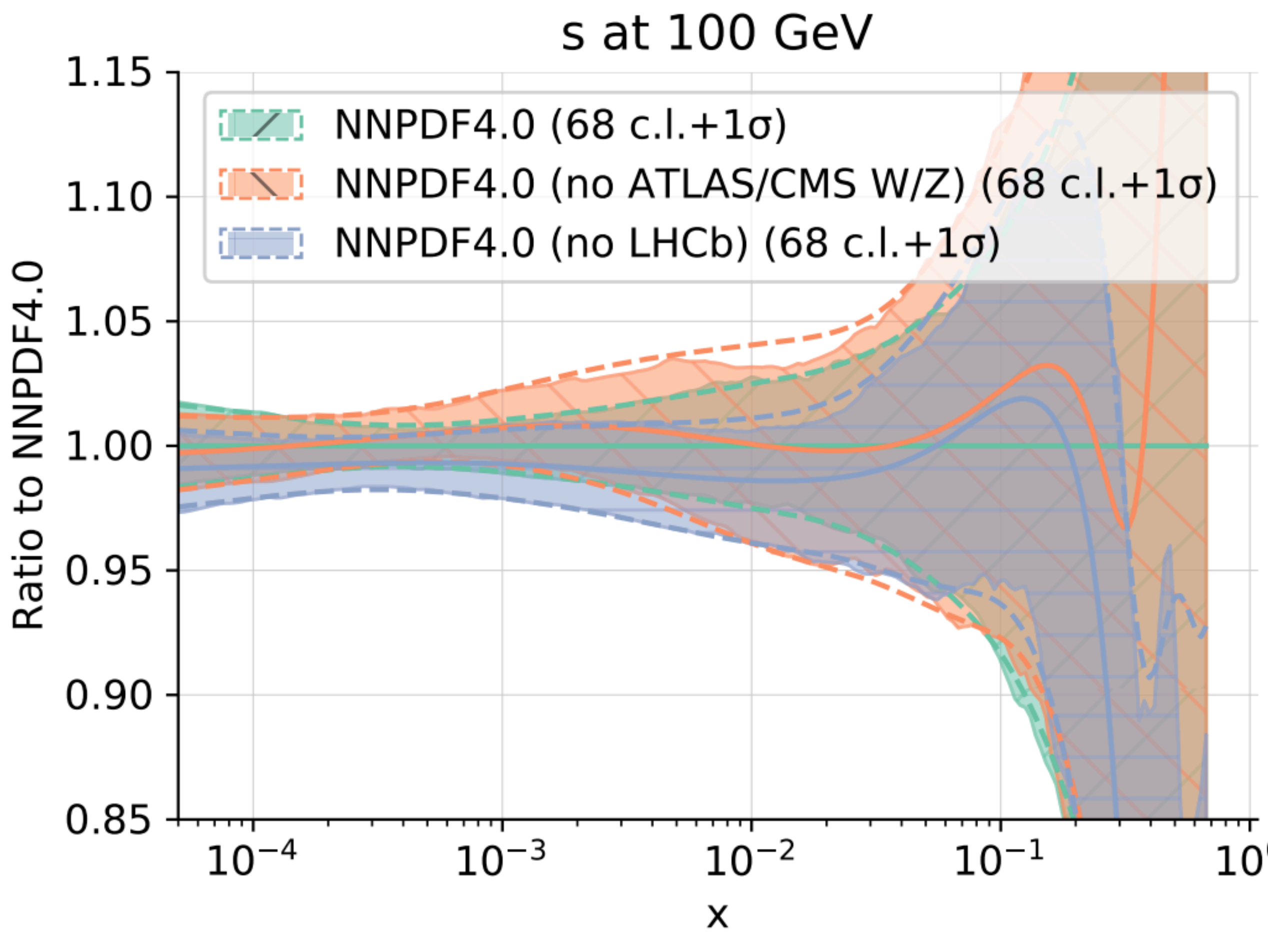
The strangest proton

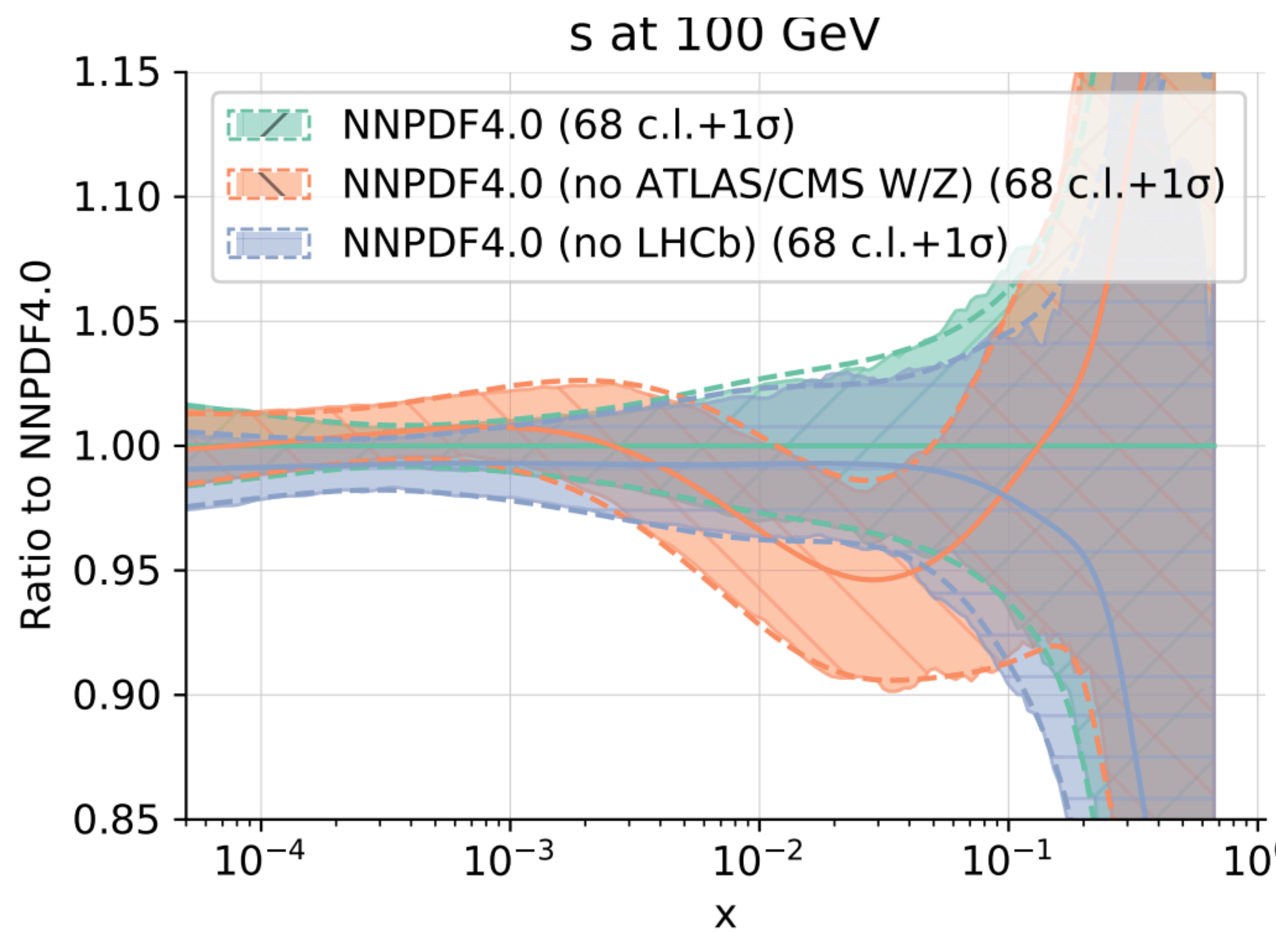
- ☑ The LHC **inclusive *W, Z* production** data are also sensitive probes of the proton strangeness

- ☑ Fit results stable, within uncertainties, when either **ATLAS/CMS** or **LHCb *W, Z*** data are removed

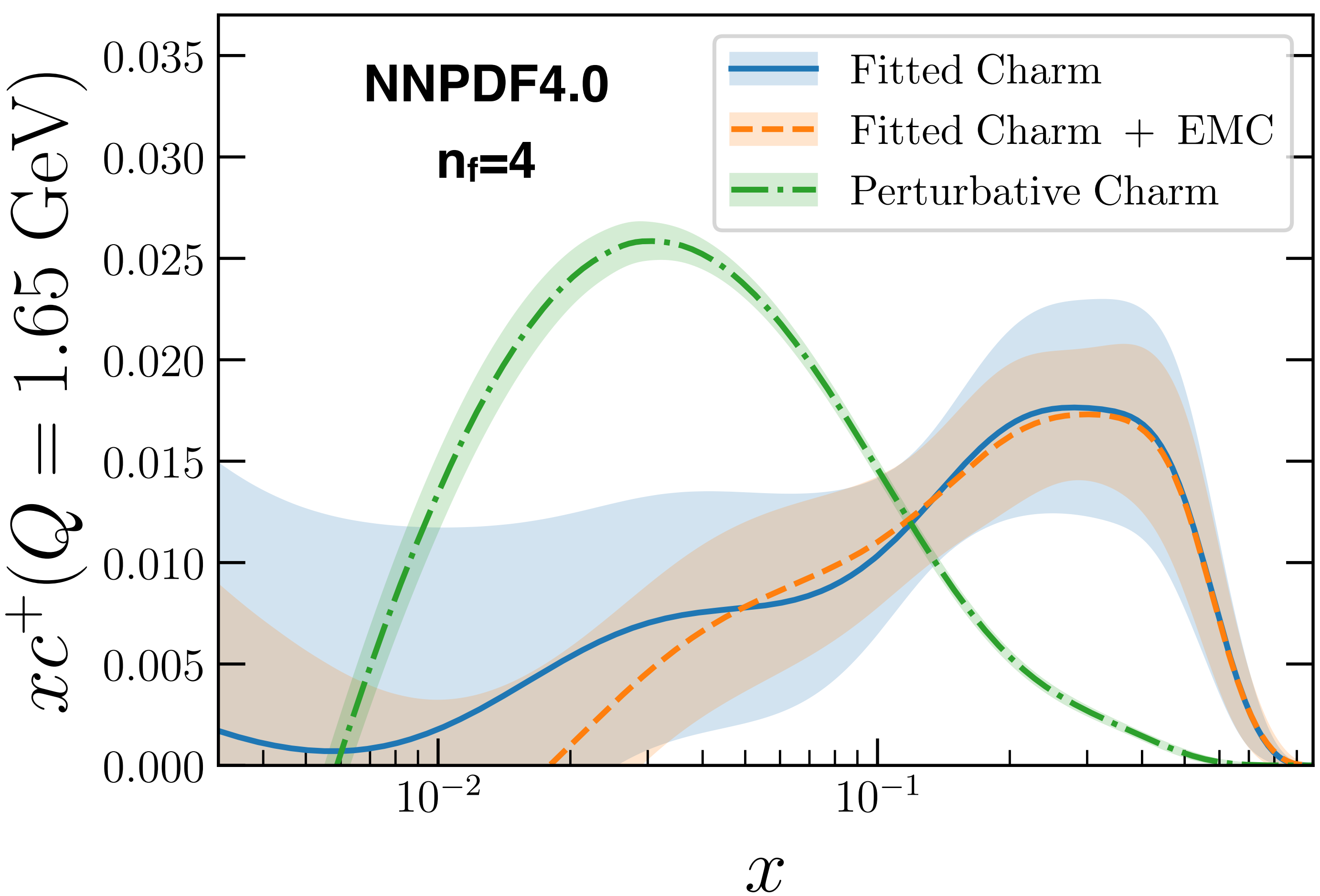
- ☑ **No tension** between LHC and DIS neutrino data observed

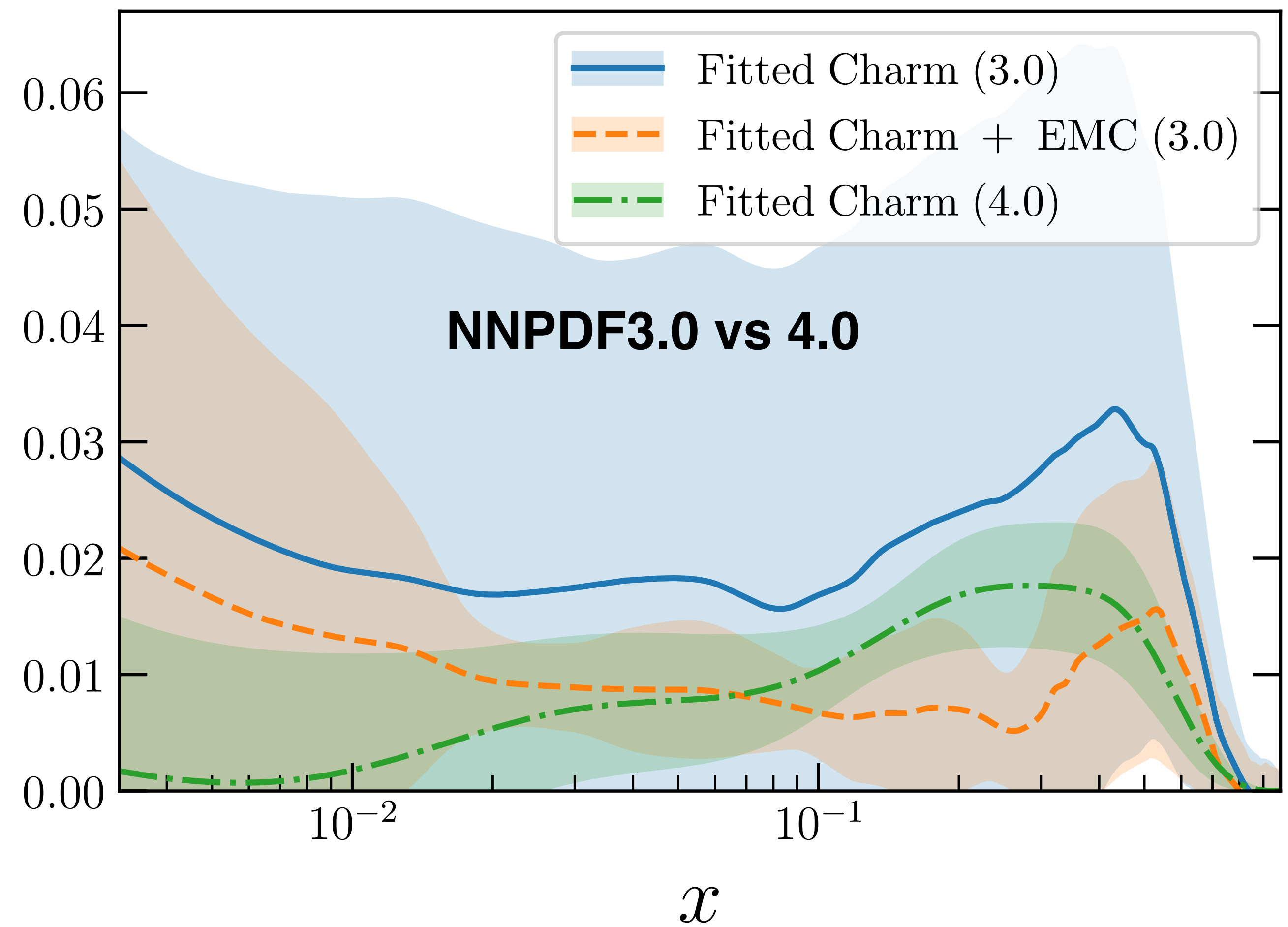






Intrinsic charm





- ☑ Increasing evidence for **non-perturbative charm component** within the proton, robust upon conversion to the **3FNS** via backwards evolution and matching conditions
- ☑ Bulk of constraints provided by new **precision LHC data**, complemented by fixed-target DIS
- ☑ As opposed to previous studies, impact of the **EMC charm measurements** mild now. Information provided by EMC F_2^c consistent with latest collider data

Intrinsic charm

Charm PDF constrained by experimental data

5FNS

$$xc^{(n_f=5)}(x,Q) \quad Q > \mu_b = 4.91\,\mathrm{GeV}$$

$\mu_b = 4.91\,\mathrm{GeV}$

4FNS to 5FNS transformation

$$xc^{(n_f=4)}(x,Q) \quad Q > Q_0 = 1.65$$

4FNS

Charm PDF parametrised with deep learning

$$xc^{(n_f=4)}(x,Q_0) = x^{-\alpha_c}(1-x)^{\beta_c}\mathrm{NN}_c(x)\,, \quad Q_0 = 1.65\ \mathrm{GeV}$$

$\mu_c = 1.51\,\mathrm{GeV}$

4FNS to 3FNS transformation

Intrinsic charm

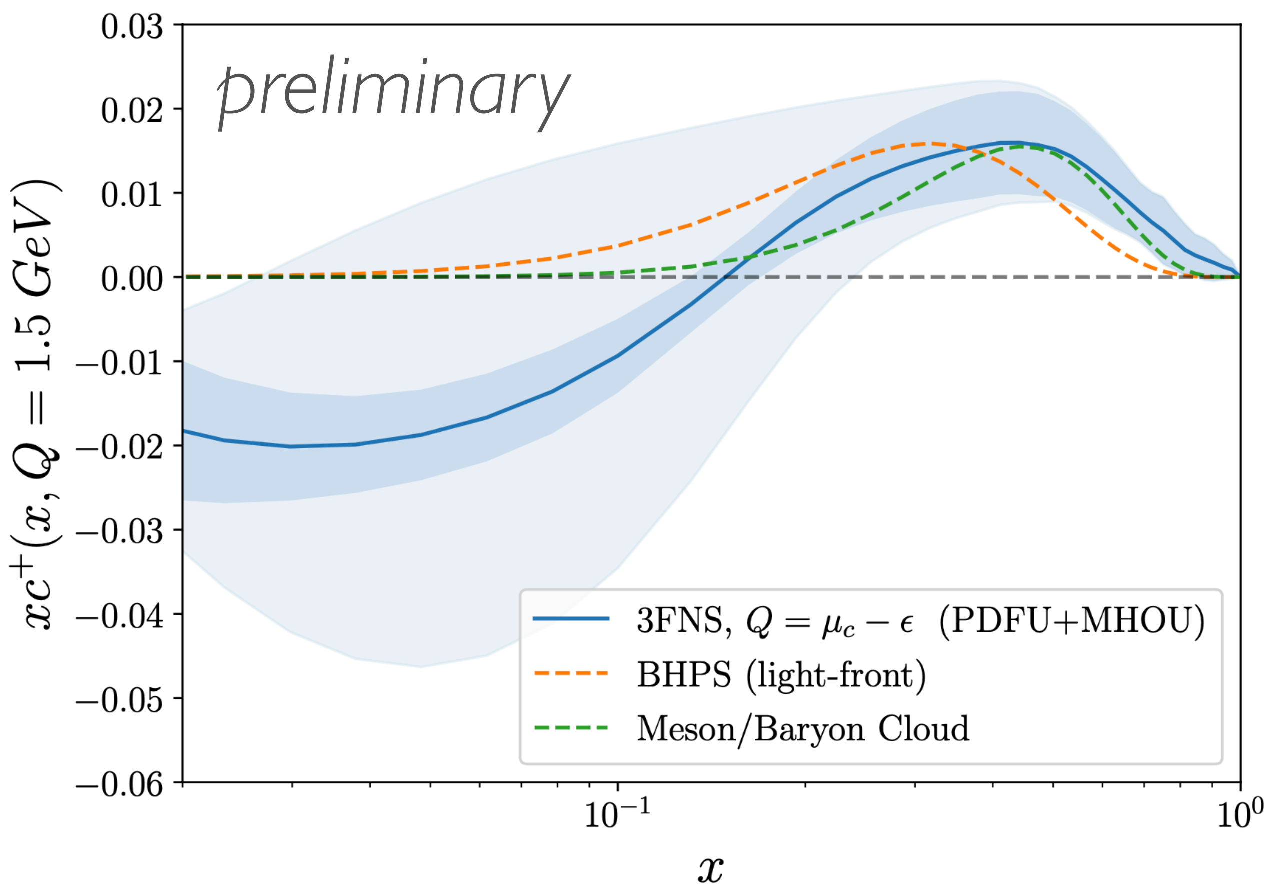
3FNS

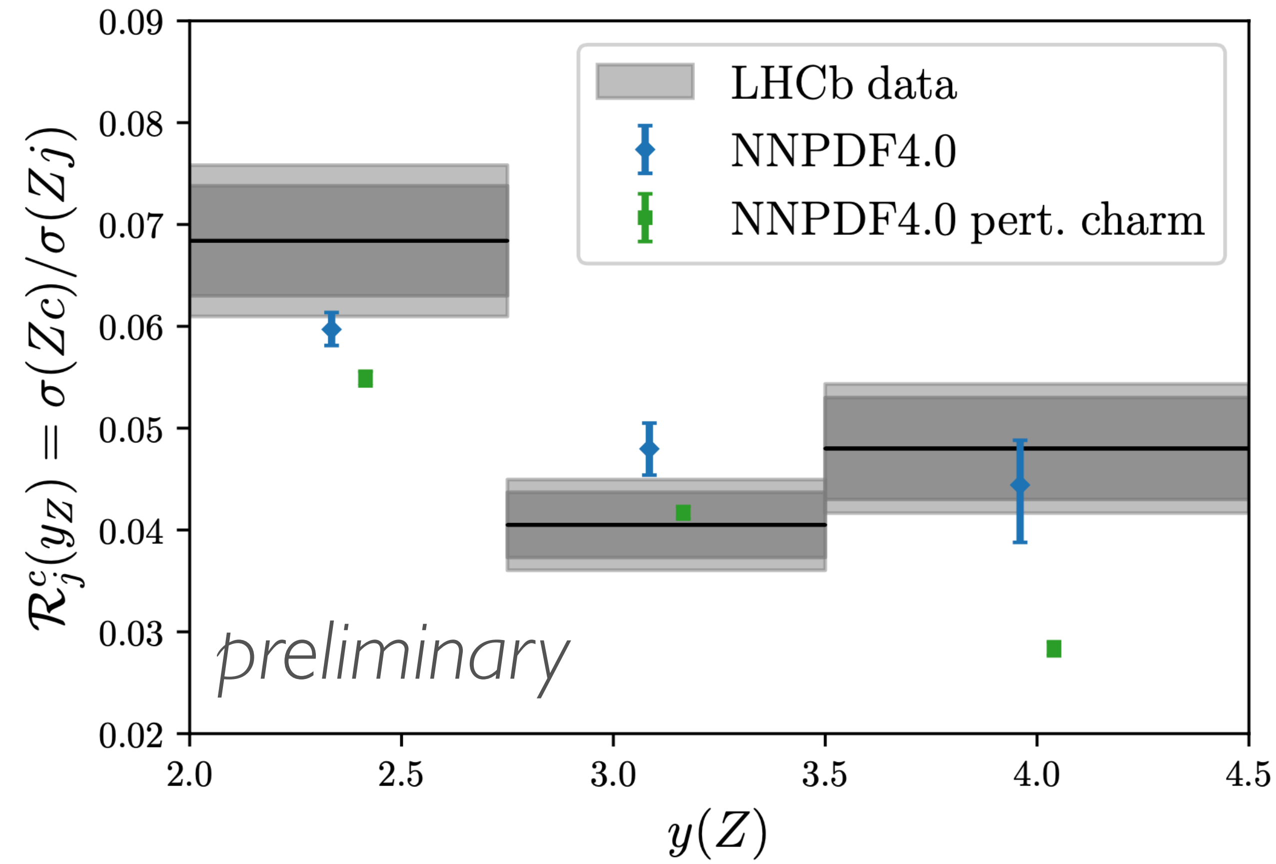
$$xc^{(n_f=3)}(x,Q) = xc_{(\mathrm{int})}(x)\,, \quad Q = 1.5\ \mathrm{GeV}$$

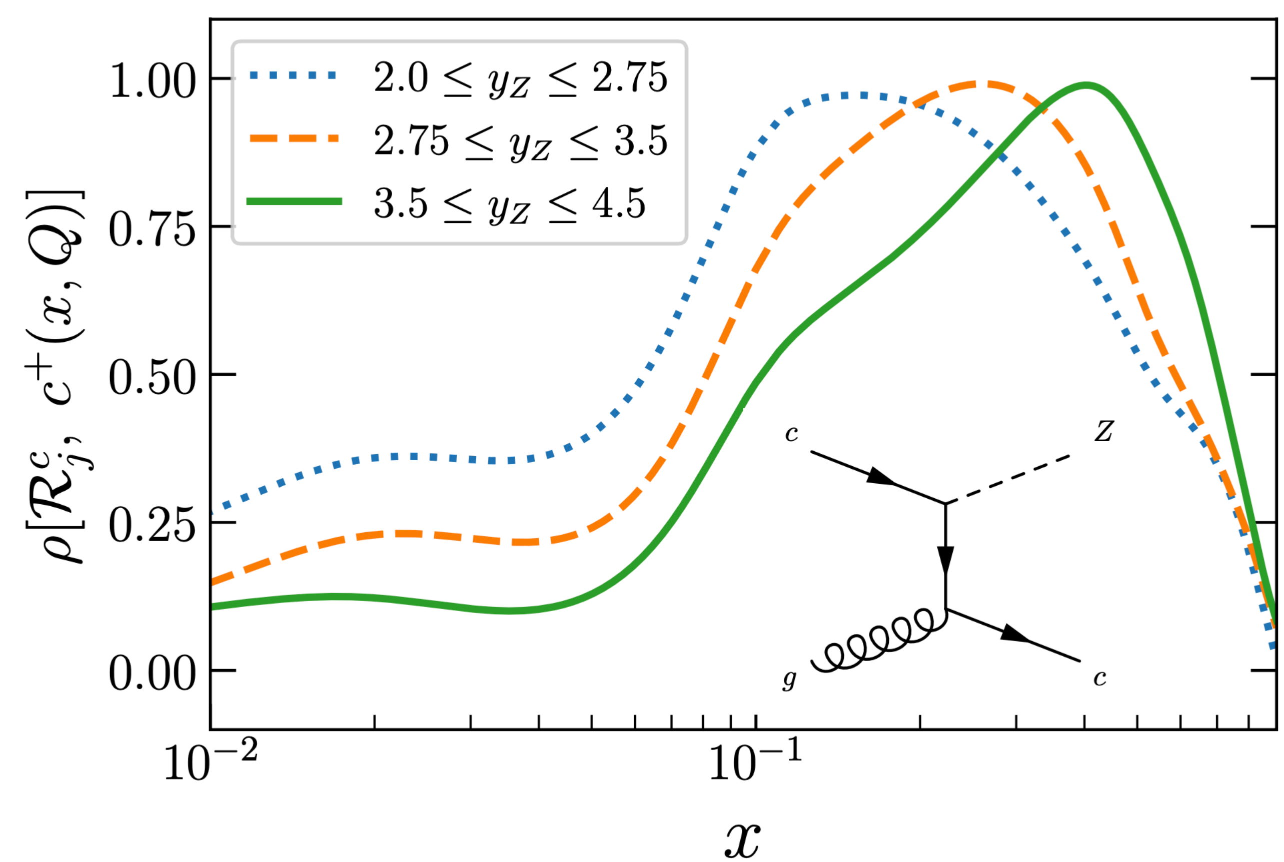
Forward DGLAP evolution

Backwards DGLAP evolution

Intrinsic charm



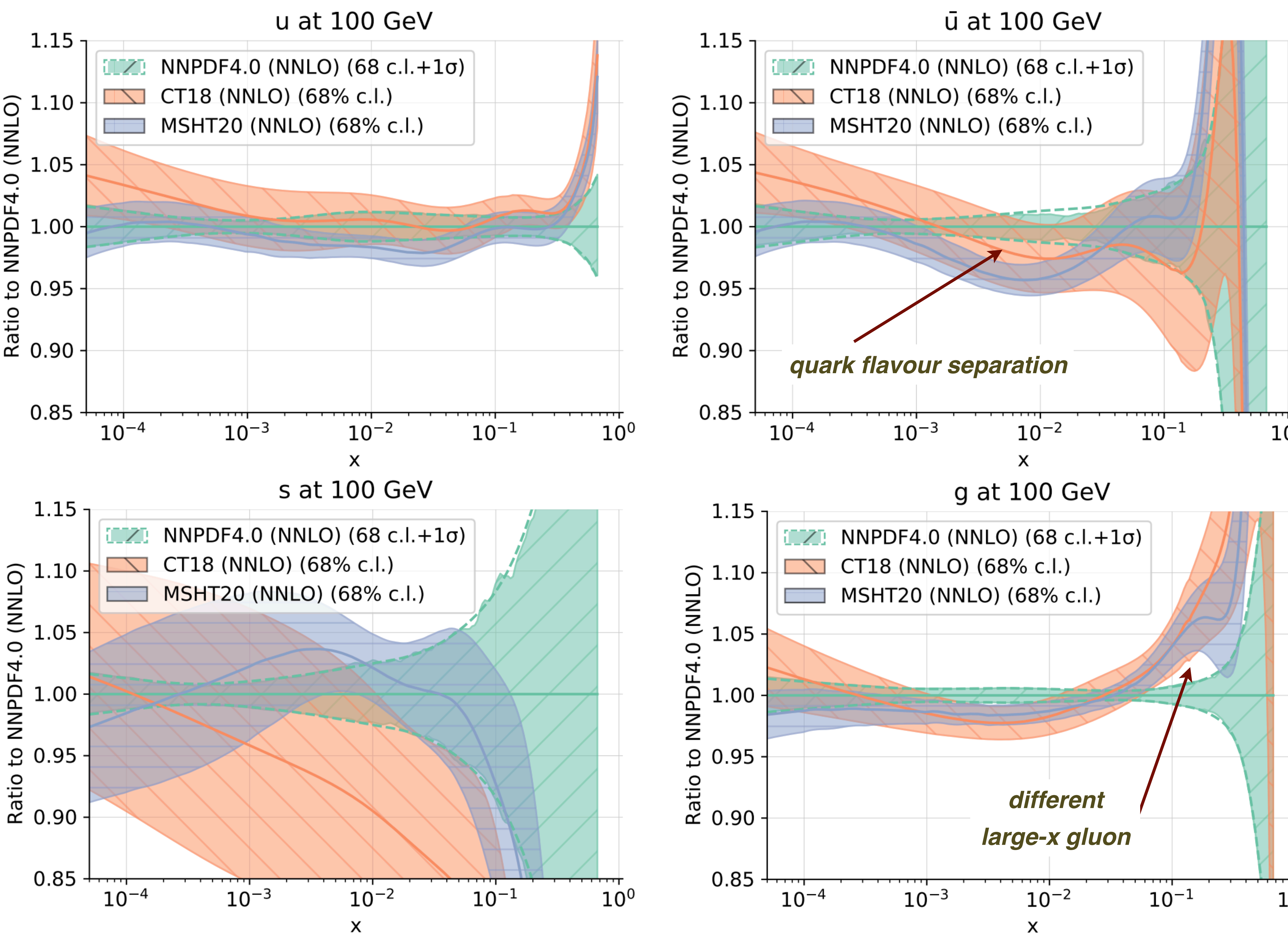




- Good agreement of the 3FNS (intrinsic) charm PDF with **non-perturbative models**
- Independent validation with recent LHCb measurements of **Z+charm**
- Consistency of indirect and direct constraints on intrinsic charm

Comparison between global fits

reasonable agreement with CT18, and MSHT20, different pattern of PDF uncertainties



PDFs in the SMEFT

Can New Physics hide inside the proton?

``How can you be sure you are not reabsorbing BSM physics into your PDFs?”

perhaps most frequent question I am asked in talks!

Assuming the **SM**, the theory calculations that enter a global PDF fit are:

$$\sigma_{\rm LHC}(\boldsymbol{\theta}) \propto \sum_{ij=u,d,g,\ldots} \int_{M^2}^{s} d\hat{s}\ \mathscr{L}_{ij}(\hat{s}, s, \boldsymbol{\theta})\, \widetilde{\sigma}_{{\rm SM},ij}(\hat{s}, \alpha_s(M))$$

SM PDFs

However in the case of BSM physics, here parametrised by the **SMEFT,** the correct expression is:

SMEFT PDFs

$$\sigma_{\rm LHC}\left(\boldsymbol{c}, \Lambda, \boldsymbol{\theta}\right) \simeq \left(\int_{M^2}^{s} d\hat{s}\ \mathscr{L}_{ij}(\hat{s}, s, \boldsymbol{\theta})\, \widetilde{\sigma}_{{\rm SM},ij}(\hat{s}, \alpha_s(M)) \right) \times \left(1 + \sum_{m=1}^{N_6} c_m \frac{\kappa_m}{\Lambda^2} + \sum_{m,n=1}^{N_6} c_m c_n \frac{\kappa_{mn}}{\Lambda^4} \right),$$

PDF parameters

SMEFT coefficients

How different are ``SM PDFs” & ``SMEFT PDFs”? Can we quantify the risk of **fitting away BSM** in PDFs?

Can New Physics hide inside the proton?

Exp.	$\sqrt{s}$ (TeV)	Ref.	$\mathcal{L}$ (fb^{-1})	Channel	1D/2D	$n_{\rm dat}$	$m_{\ell\ell}^{\rm max}$ (TeV)
ATLAS	7	[120]	4.9	e^-e^+	1D	13	[1.0, 1.5]
ATLAS (*)	8	[86]	20.3	$\ell^-\ell^+$	2D	46	[0.5, 1.5]
CMS	7	[121]	9.3	$\mu^-\mu^+$	2D	127	[0.2, 1.5]
CMS (*)	8	[87]	19.7	$\ell^-\ell^+$	1D	41	[1.5, 2.0]
CMS (*)	13	[122]	5.1	e^-e^+, $\mu^-\mu^+$ $\ell^-\ell^+$	1D	43, 43 43	[1.5, 3.0]
Total						**270 (313)**	

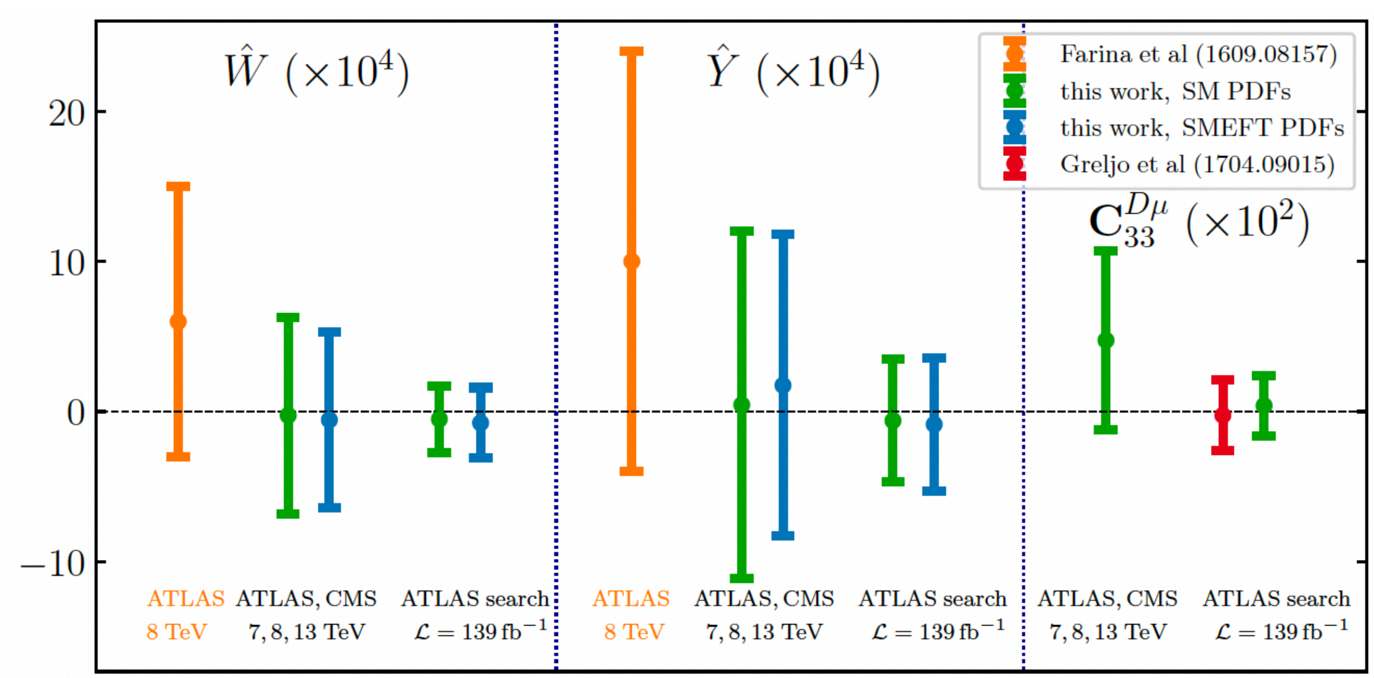
Extract PDFs from global fit where **high-mass DY cross-sections** account for EFT effects in two benchmark scenarios

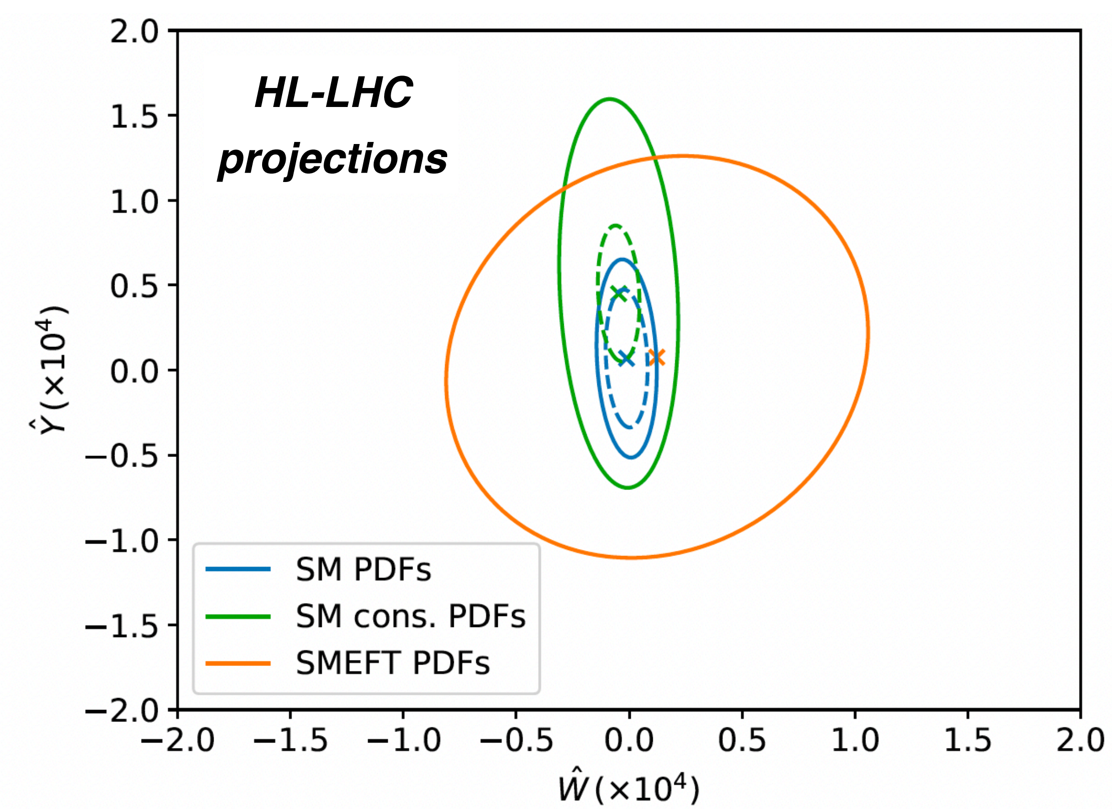
$$d\sigma_{\rm SMEFT} = d\sigma_{\rm SM} \times K_{\rm EFT}$$

$$K_{\rm EFT} = 1 + \sum_{n=1}^{n_{\rm op}} c_n R^{(n)}_{\rm SMEFT} + \sum_{n,m=1}^{n_{\rm op}} c_n c_m R^{(n,m)}_{\rm SMEFT}$$

Available data: limited interplay between PDF and EFT fits, best constraints from **searches**

HL-LHC: EFT effects, if present, would be **reabsorbed into PDFs**

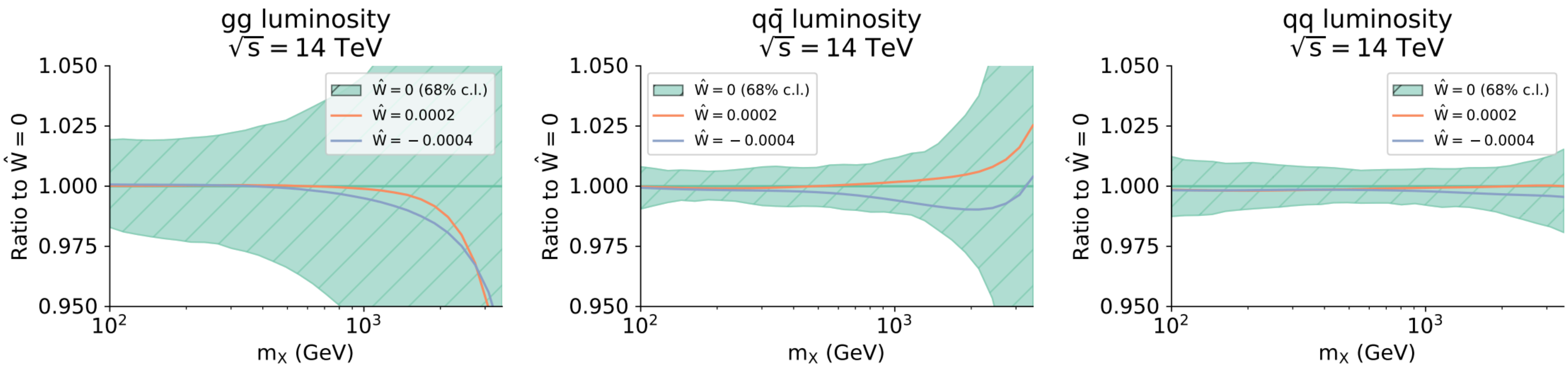




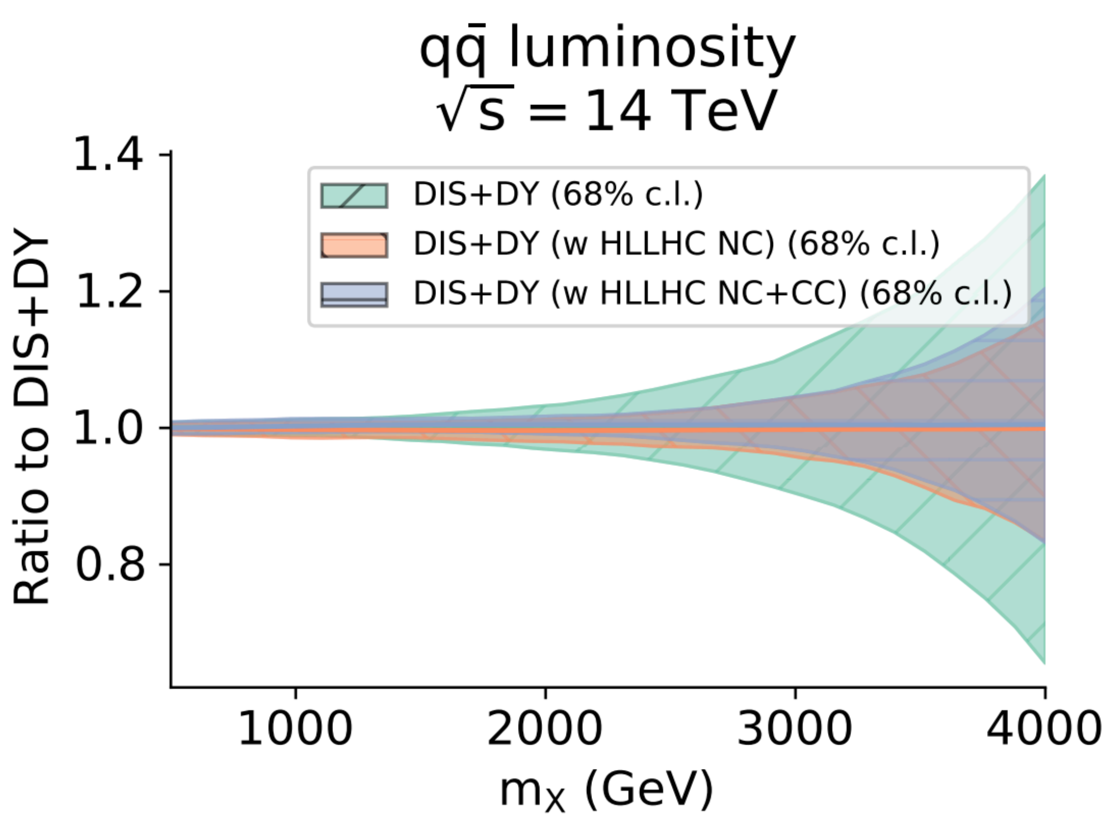
Carrazza et al 19, Greljo et al 21

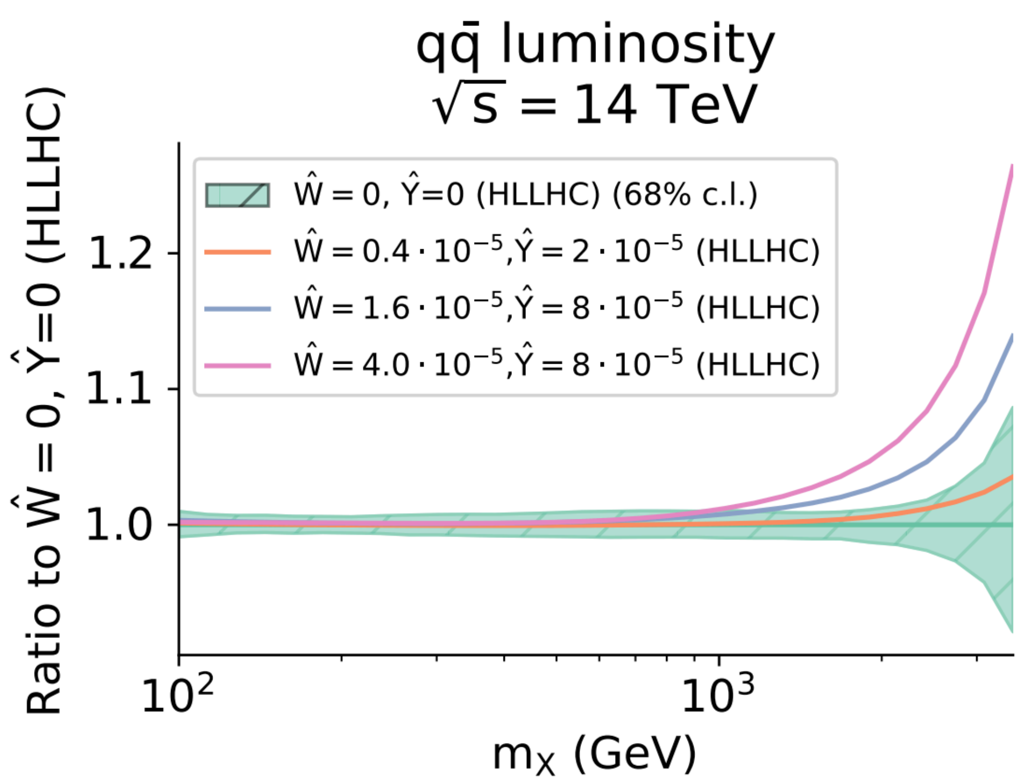
Can New Physics hide inside the proton?

with current (published) DY data, interplay between PDF and EFT effects **moderate ….**



… while at the HL-LHC EFT effects may be **reabsorbed into the PDFs**: careful separation instrumental



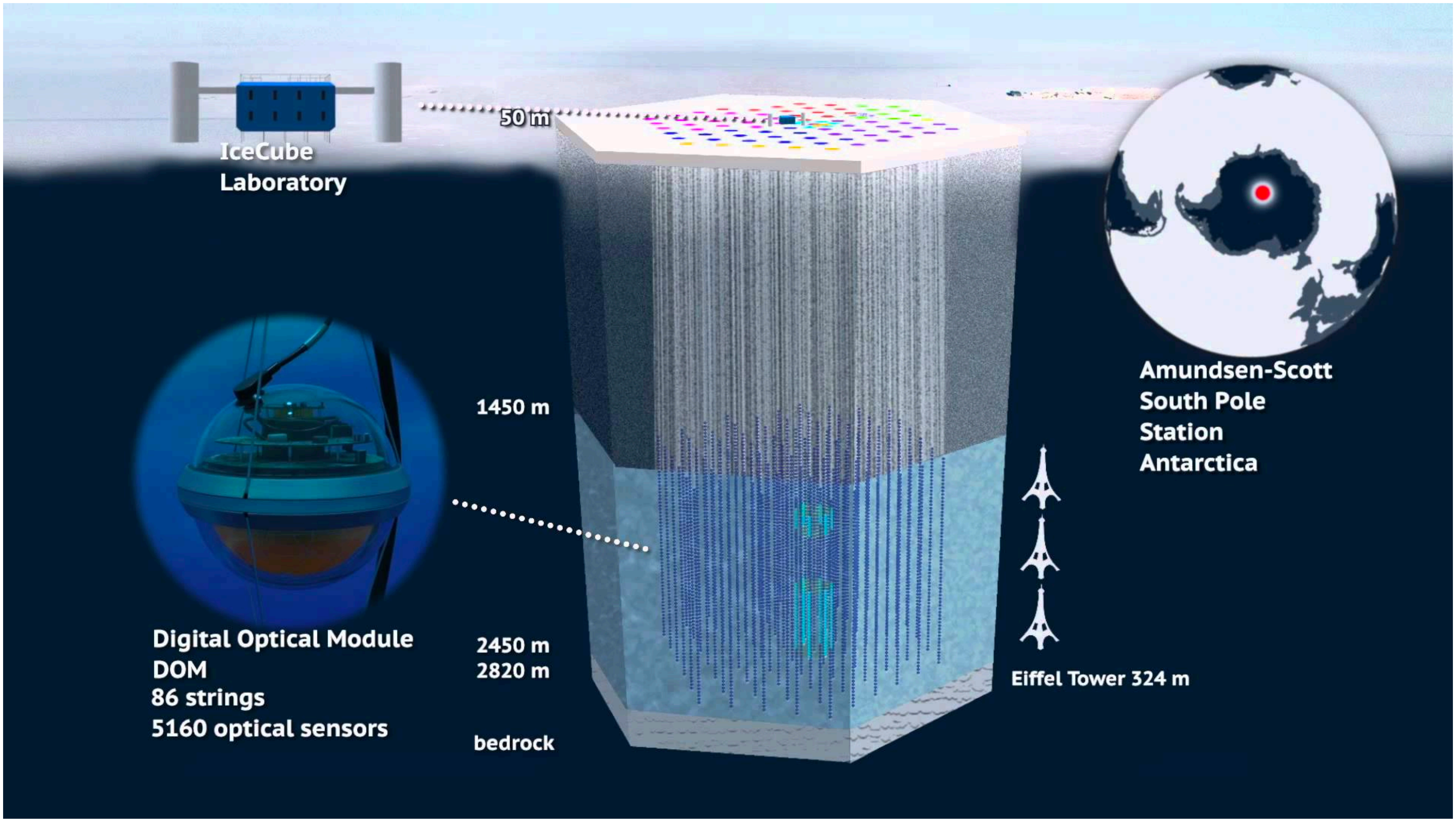


PDFs &

Neutrino Physics

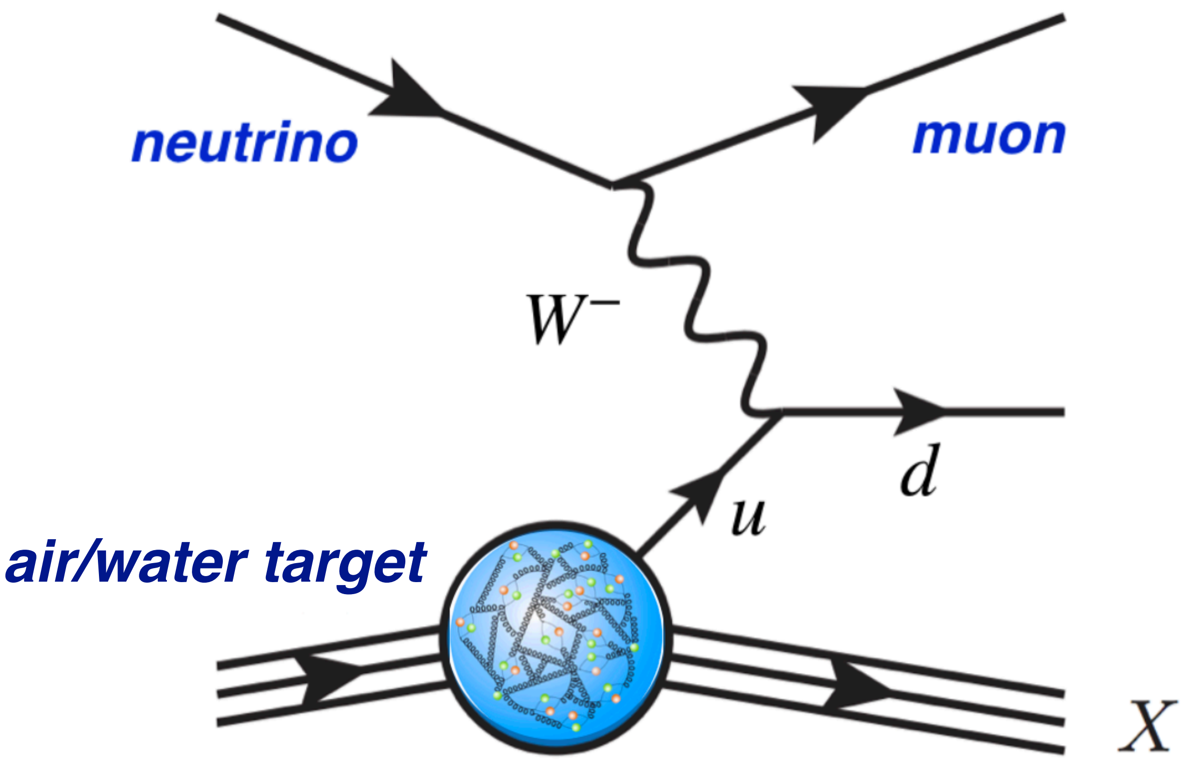
Neutrino telescopes

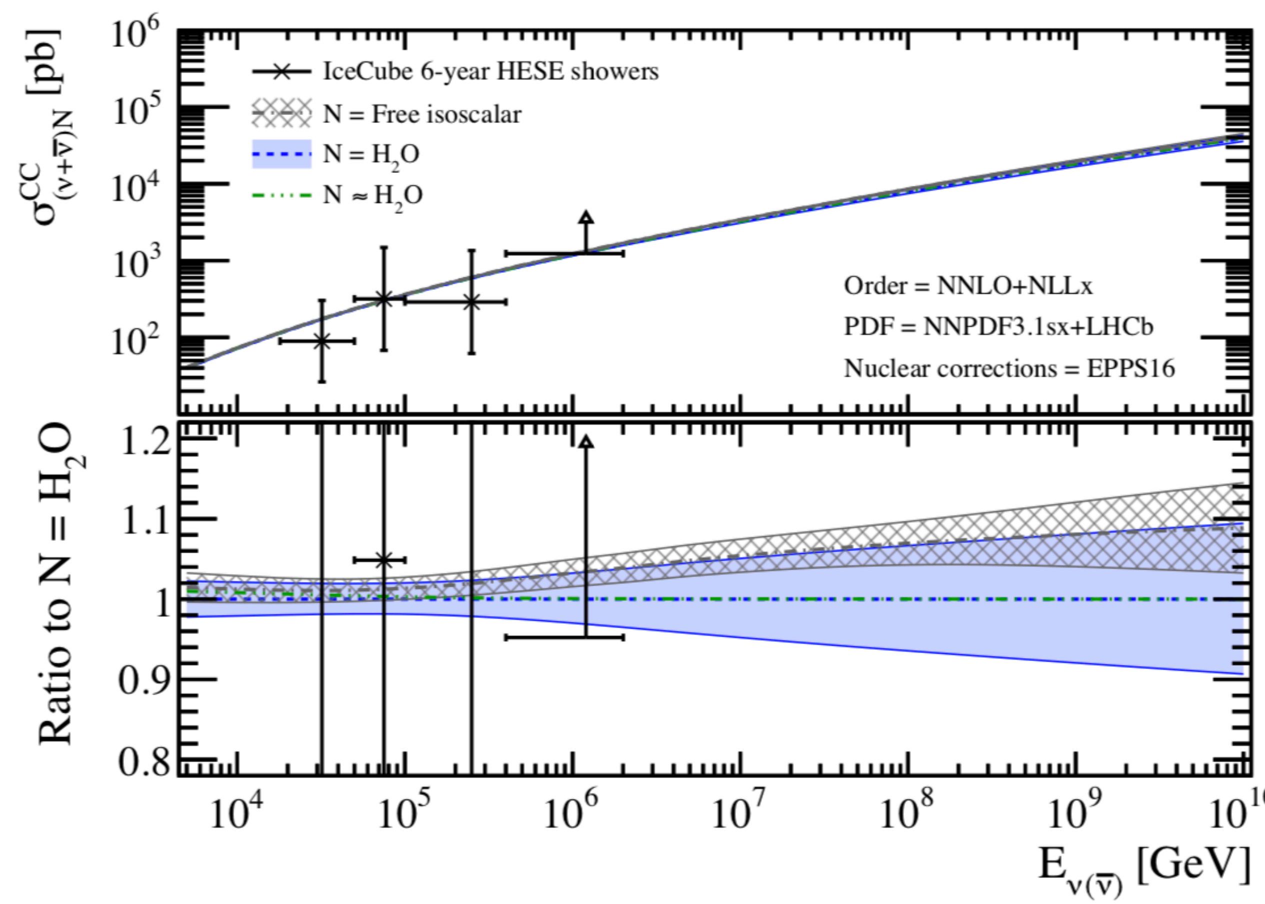
Ultra-high energy (UHE) neutrinos: novel window to the **extreme Universe**



Neutrino telescopes as QCD microscopes

Ultra-high energy (cosmic) neutrino - nucleus scattering:
unique probe of **small-x PDFs and QCD**





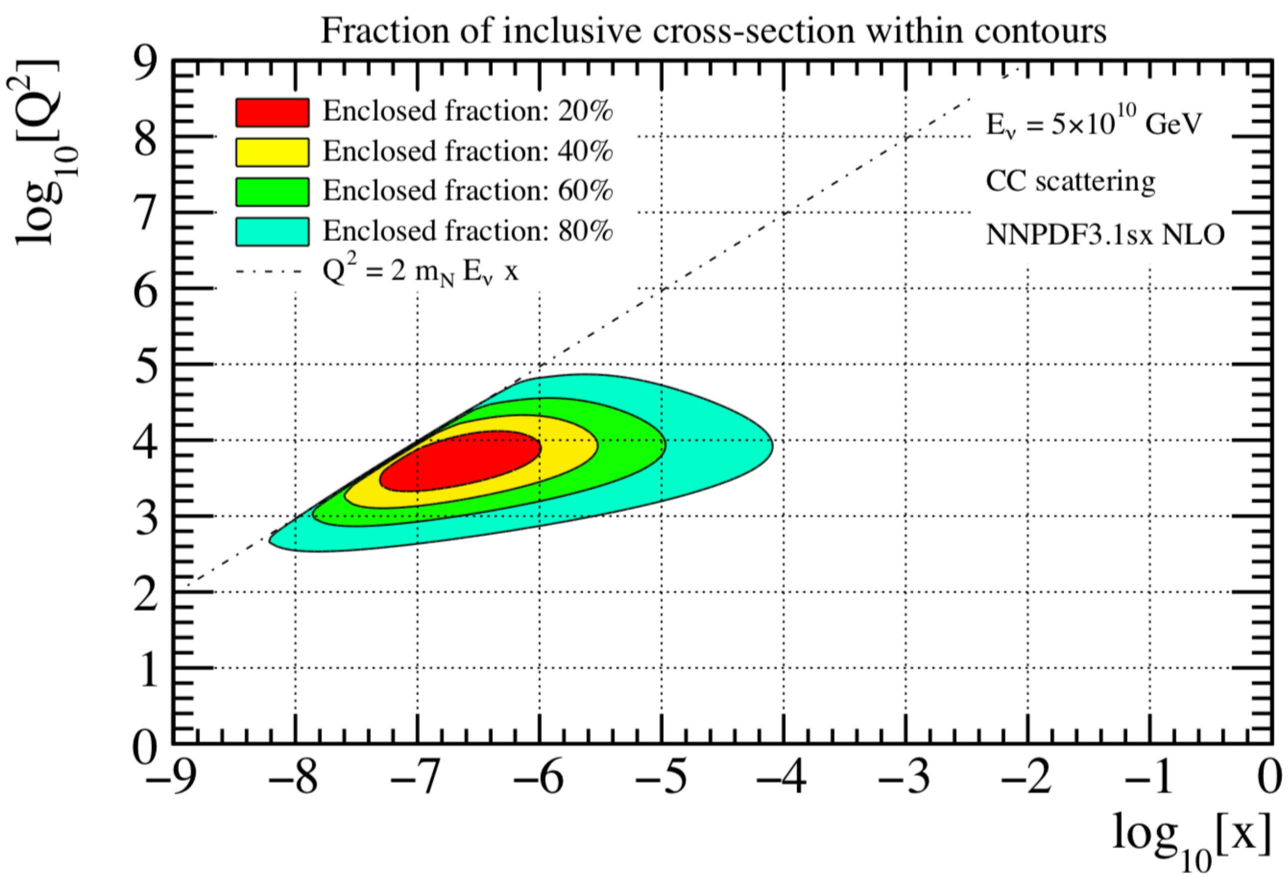
Bertone, Gauld, JR 18

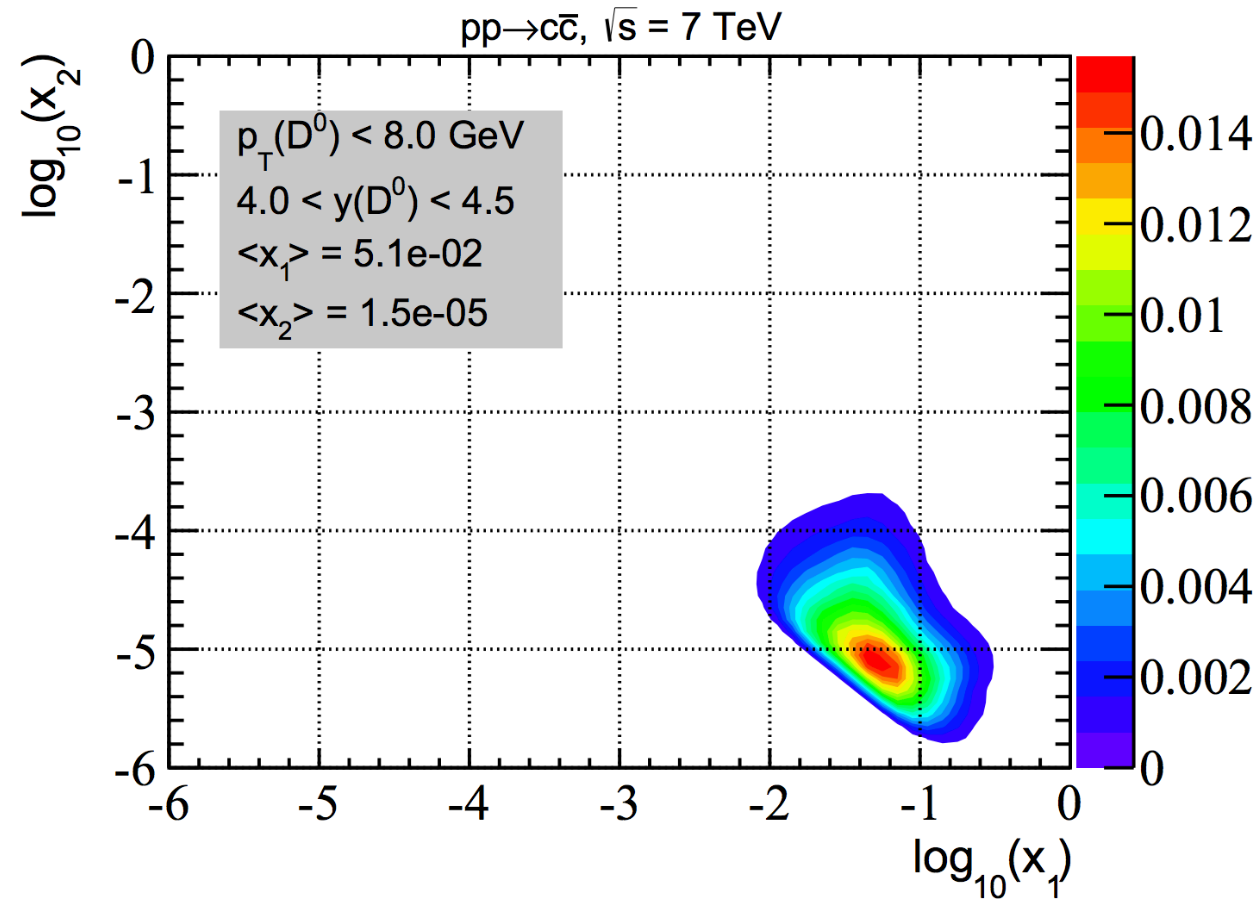
Sensitive to **small-*x* quarks** (and gluons via evolution)
down to $\mathbf{x \simeq 10^{-8}}$ at $\mathbf{Q \simeq M_W}$

Neutrino telescopes as QCD microscopes

***signal**: cosmic neutrino - nucleus scattering*

***background**: prompt charm production*

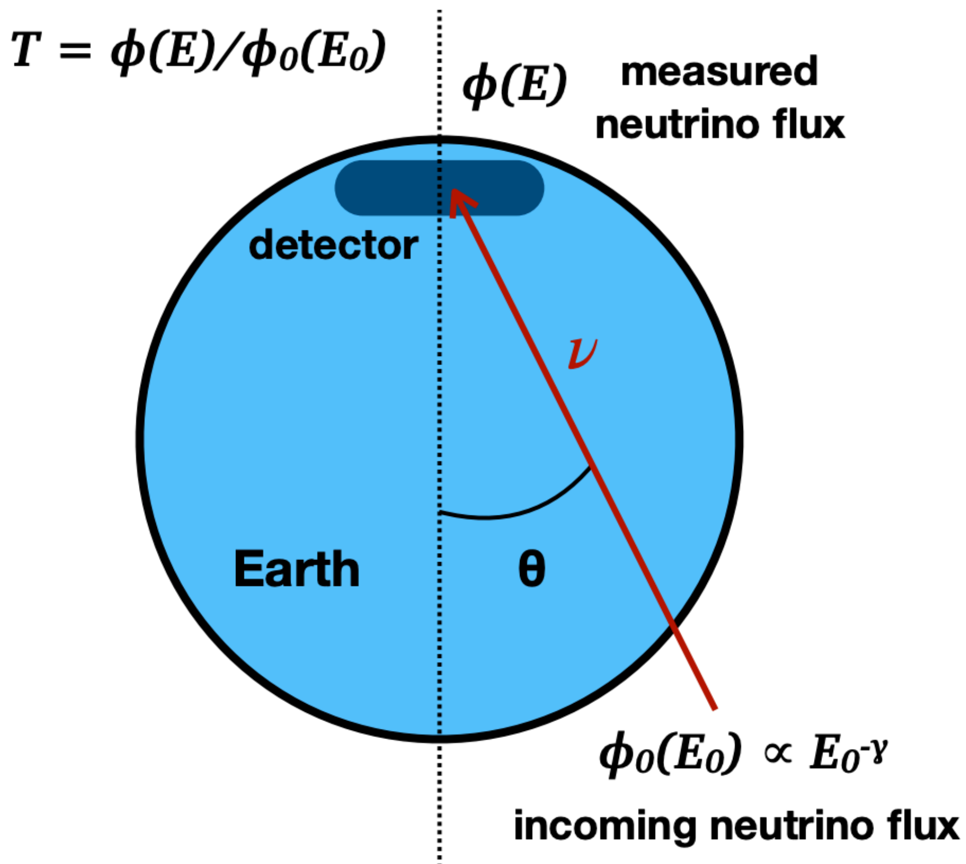


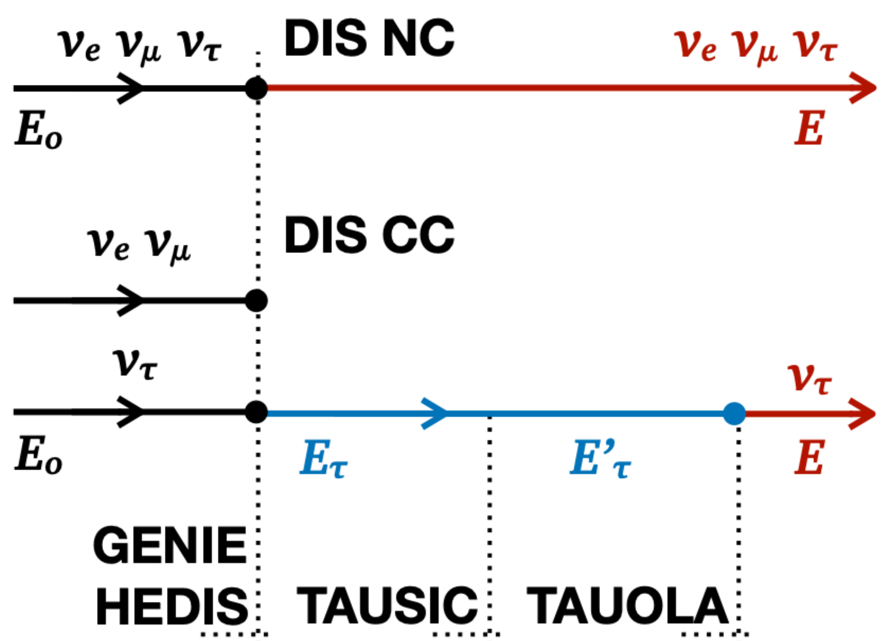


Sensitive to **small-*x* quarks** (and thus gluons via evolution) down to **x** ≃ **10^{-8}** and **Q** ≃ **M_W**

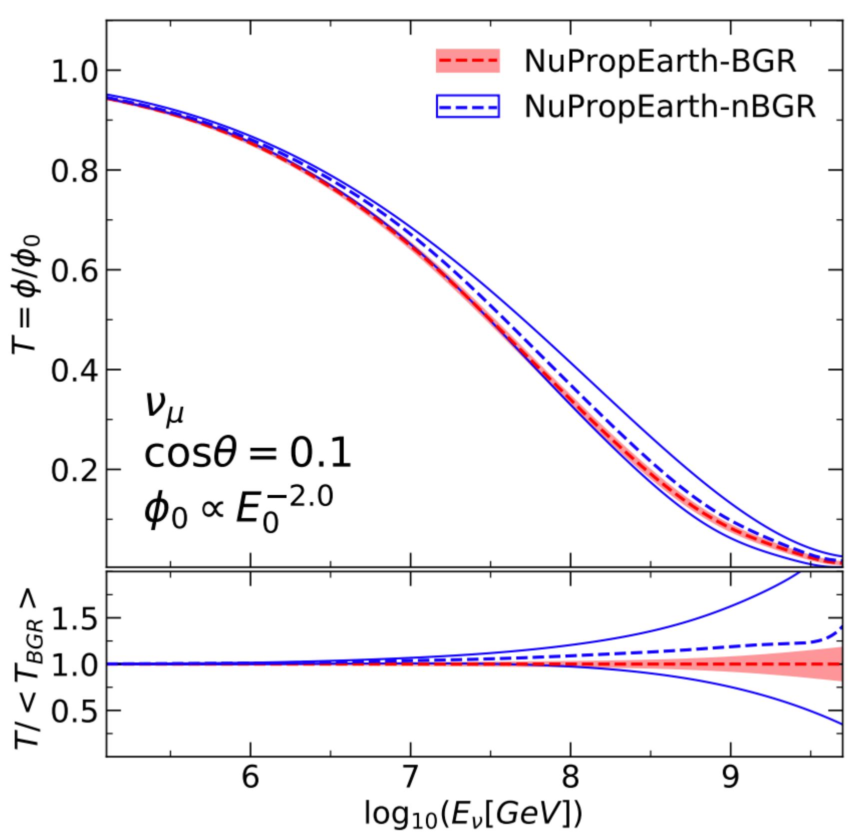
Sensitive to **small-*x* gluons** down to **x** ≃ **10^{-6}** and **Q** ≃ **M_{charm}** in the **centre-of-mass frame**

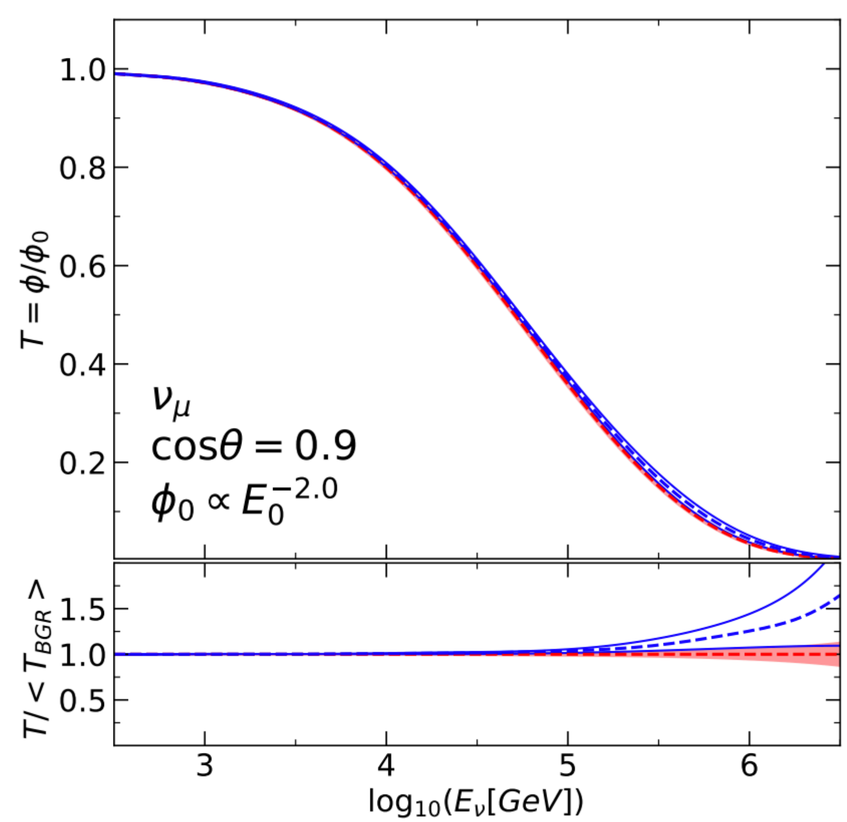
Neutrino fluxes attenuation





- As UHE neutrinos cross Earth on their way to the detector, they loss energy by interactions with Earth matter





- Precise predictions of these **attenuation rates** require a good understanding of proton and nuclear PDFs at small-x

Garcia, Gauld, Heijboer, JR 20

Summary and outlook

- The global NNPDF4.0 fit achieves **high accuracy** in an unprecedentedly broad kinematic range, thanks so its **extensive dataset** combined with **deep-learning optimisation models**

- Its faithfulness in representing PDF uncertainties is validated by **closure tests**, **future tests**, and **parametrisation basis independence**

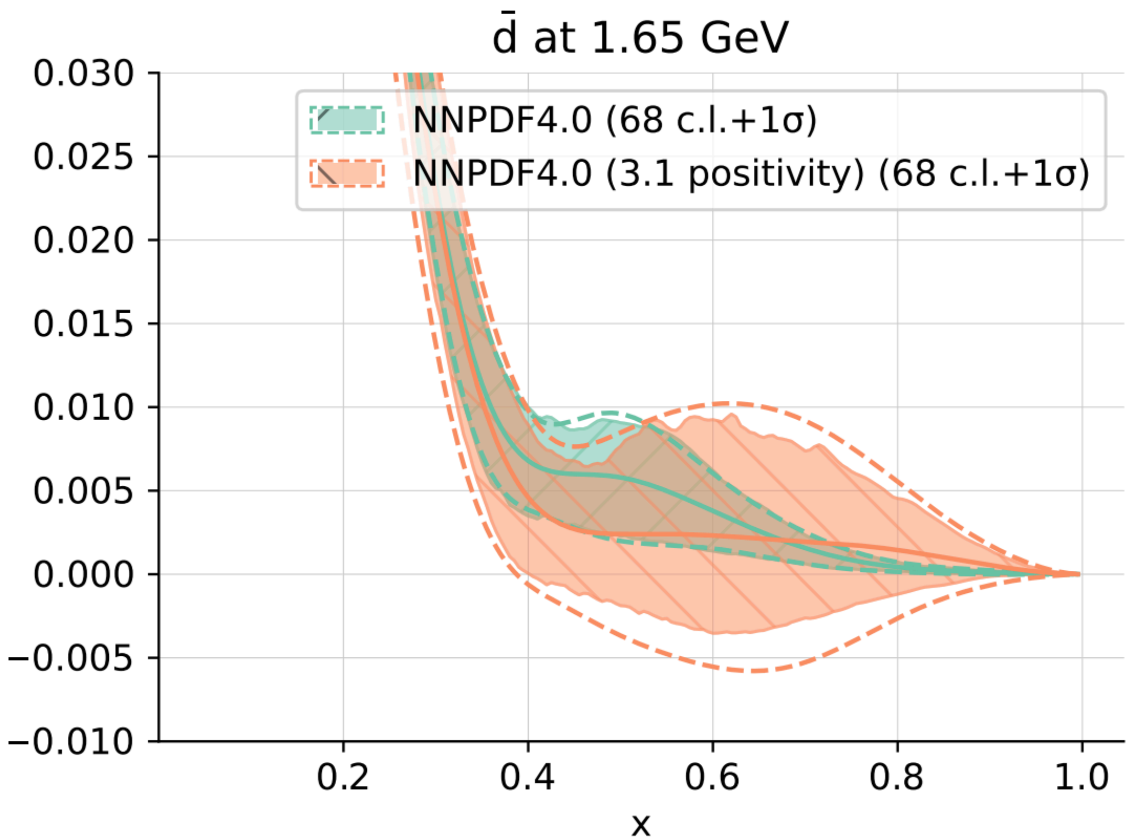
- In addition to implications for **LHC precision physics**, NNPDF4.0 sheds light on aspects of proton structure from **light antiquark asymmetries** to **strangeness** and **intrinsic charm**

- The current level of PDF uncertainties challenges the accuracy of theoretical predictions and demand an increased effort towards the systematic inclusion in the fit of **theoretical uncertainties** (nuclear, higher orders, SM parameters, . . .) and **higher-order QCD** (including N3LO) **and EW corrections**

- Full NNPDF software framework is now **open source** and welcoming contributions!

Extra Material

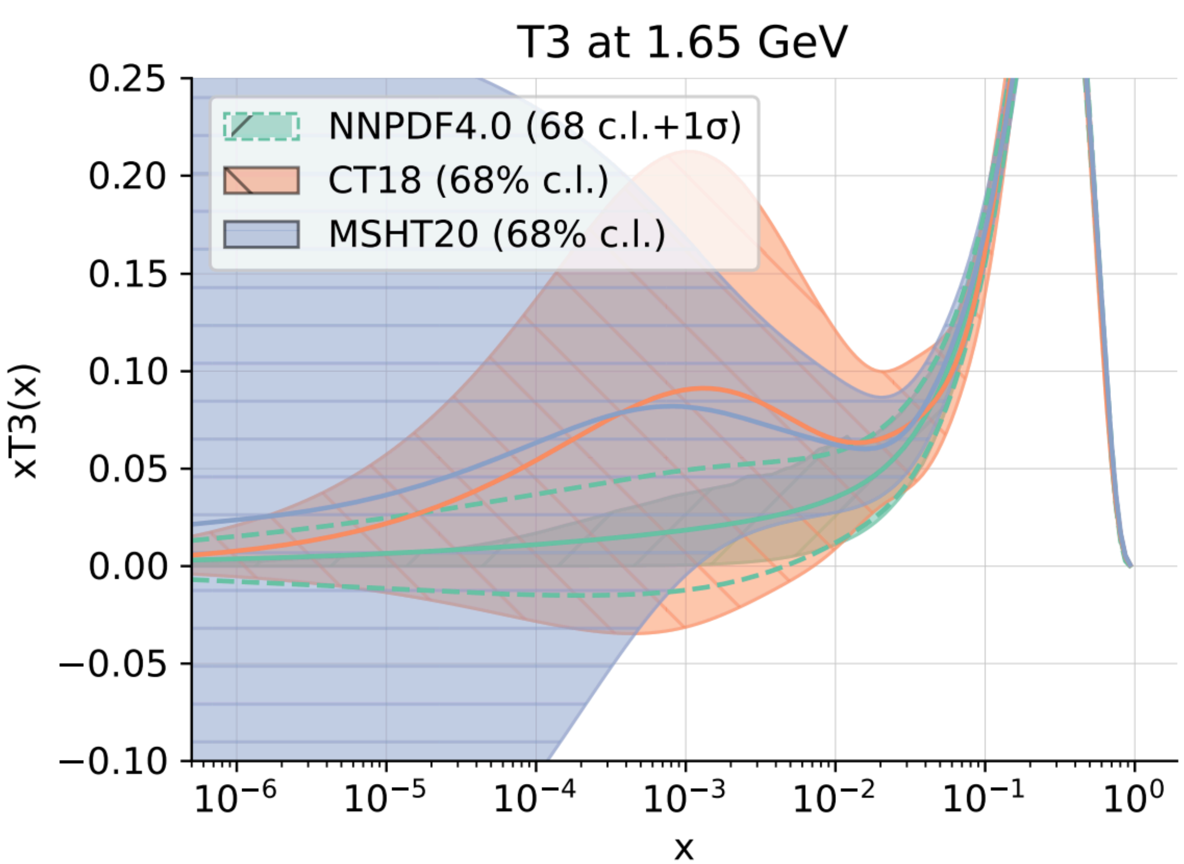
Positivity and integrability

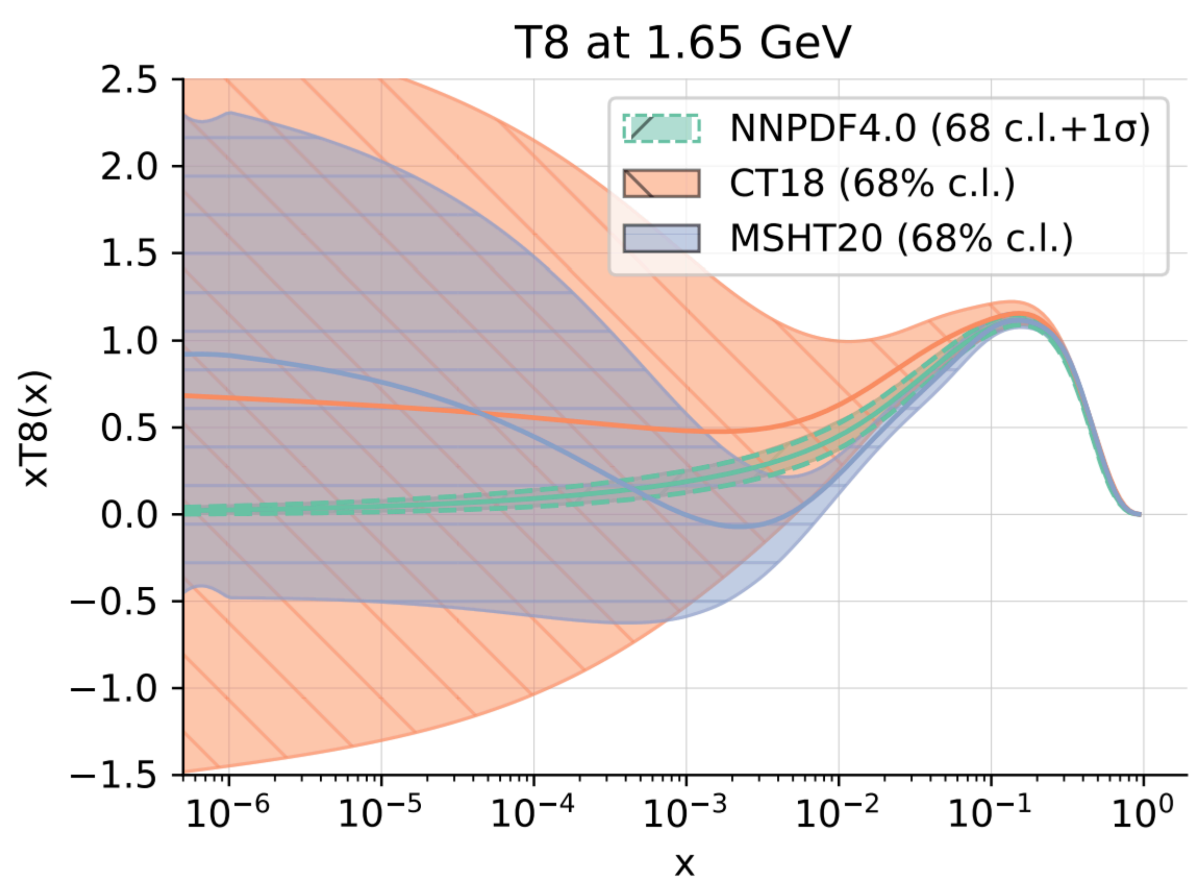


- MSbar PDFs have been shown to satisfy **positivity** requirements at all orders: **reduce large-*x* uncertainties**

- The non-singlet quark triplet and octet should be *integrable* (e.g. Gottfried sum rule): **reduce small-*x* uncertainties**

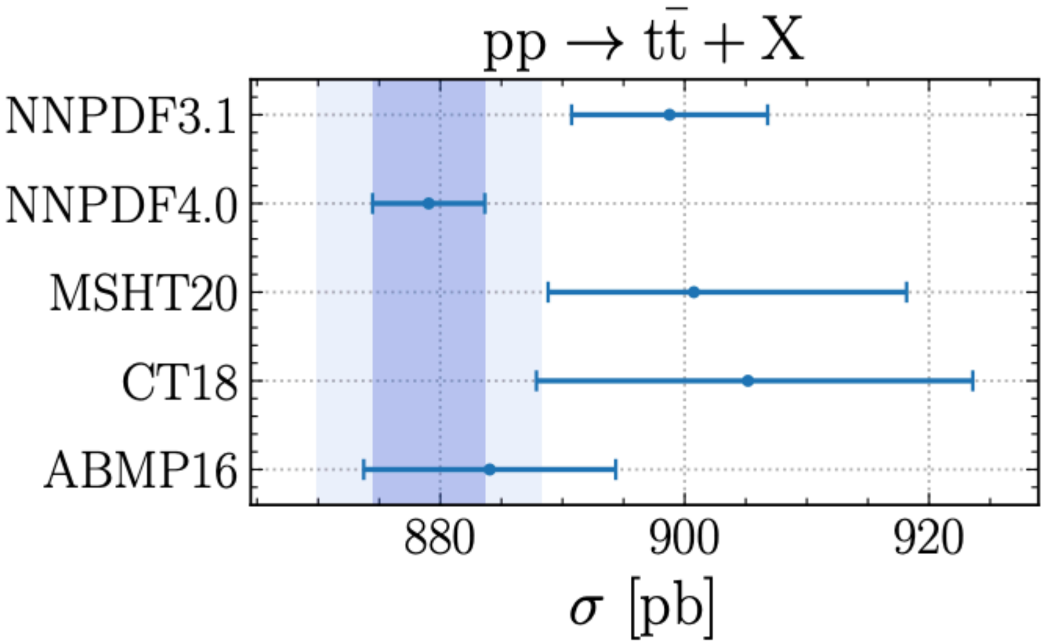
$$T_8 = (u + \bar{u}) + \left(d + \bar{d}\right) - 2\,(s + \bar{s})$$

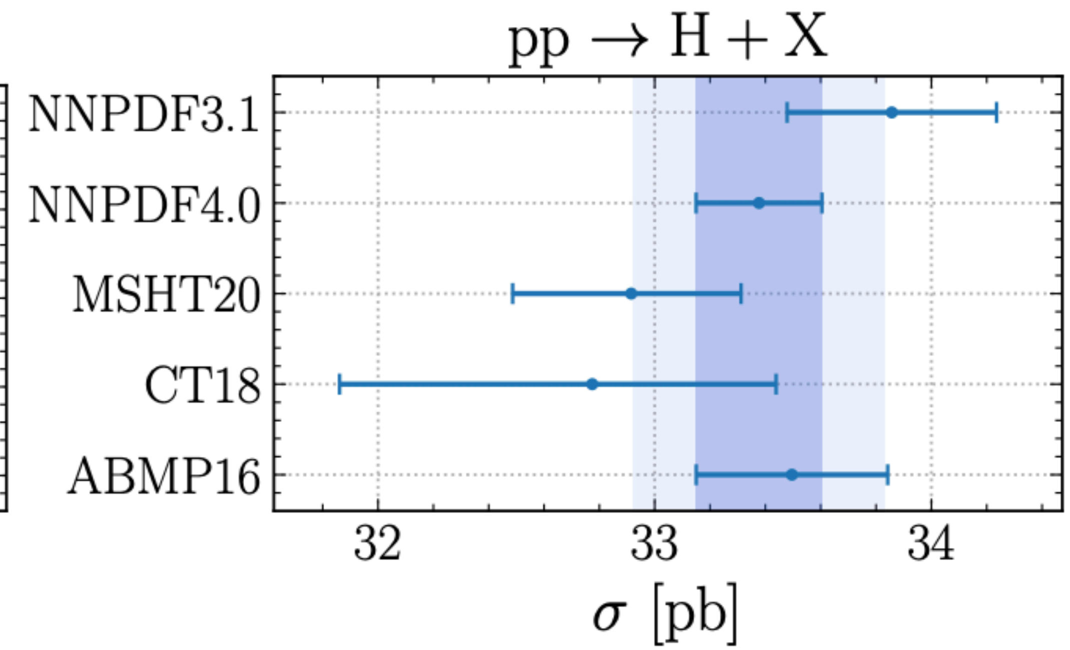


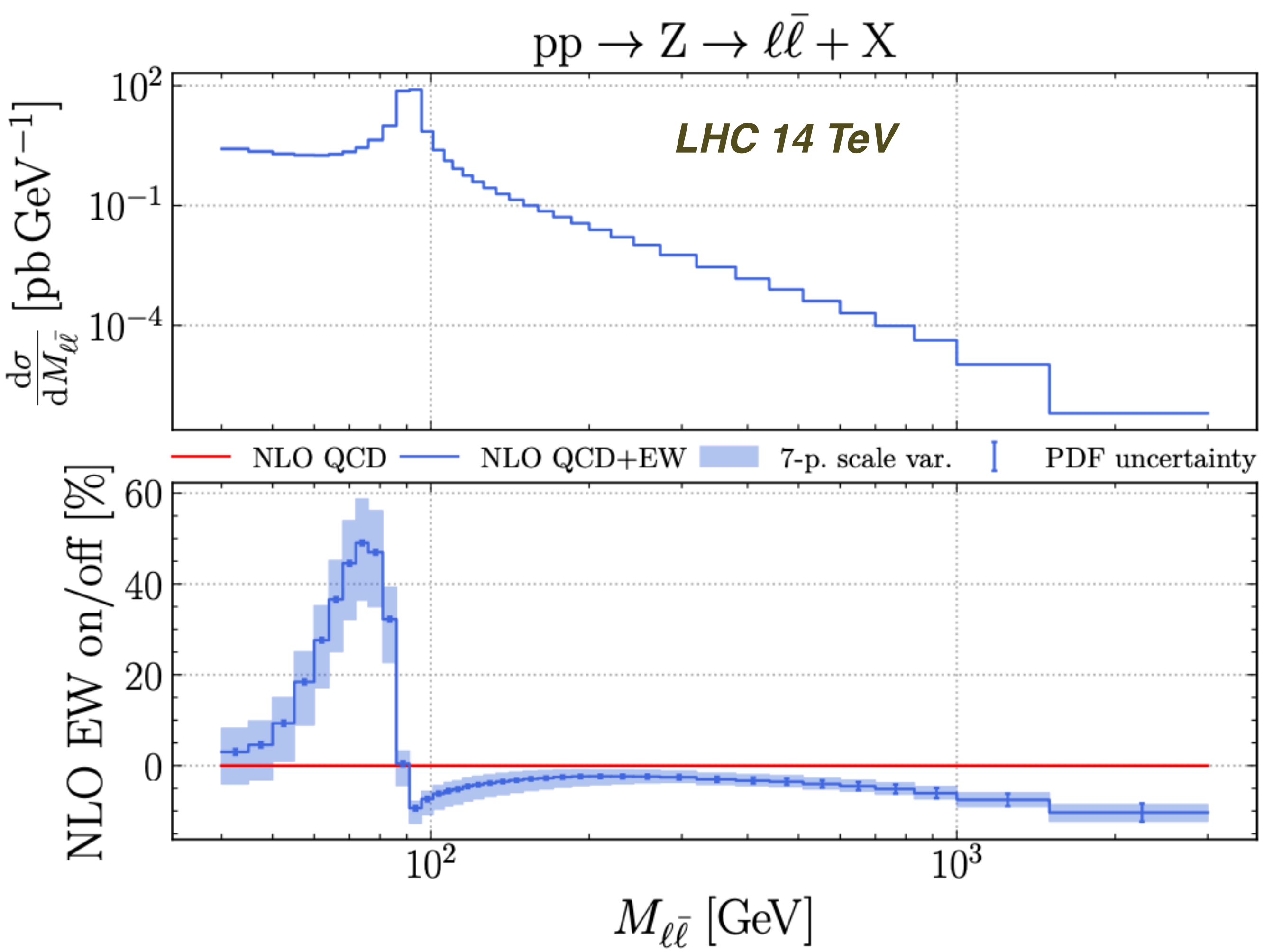


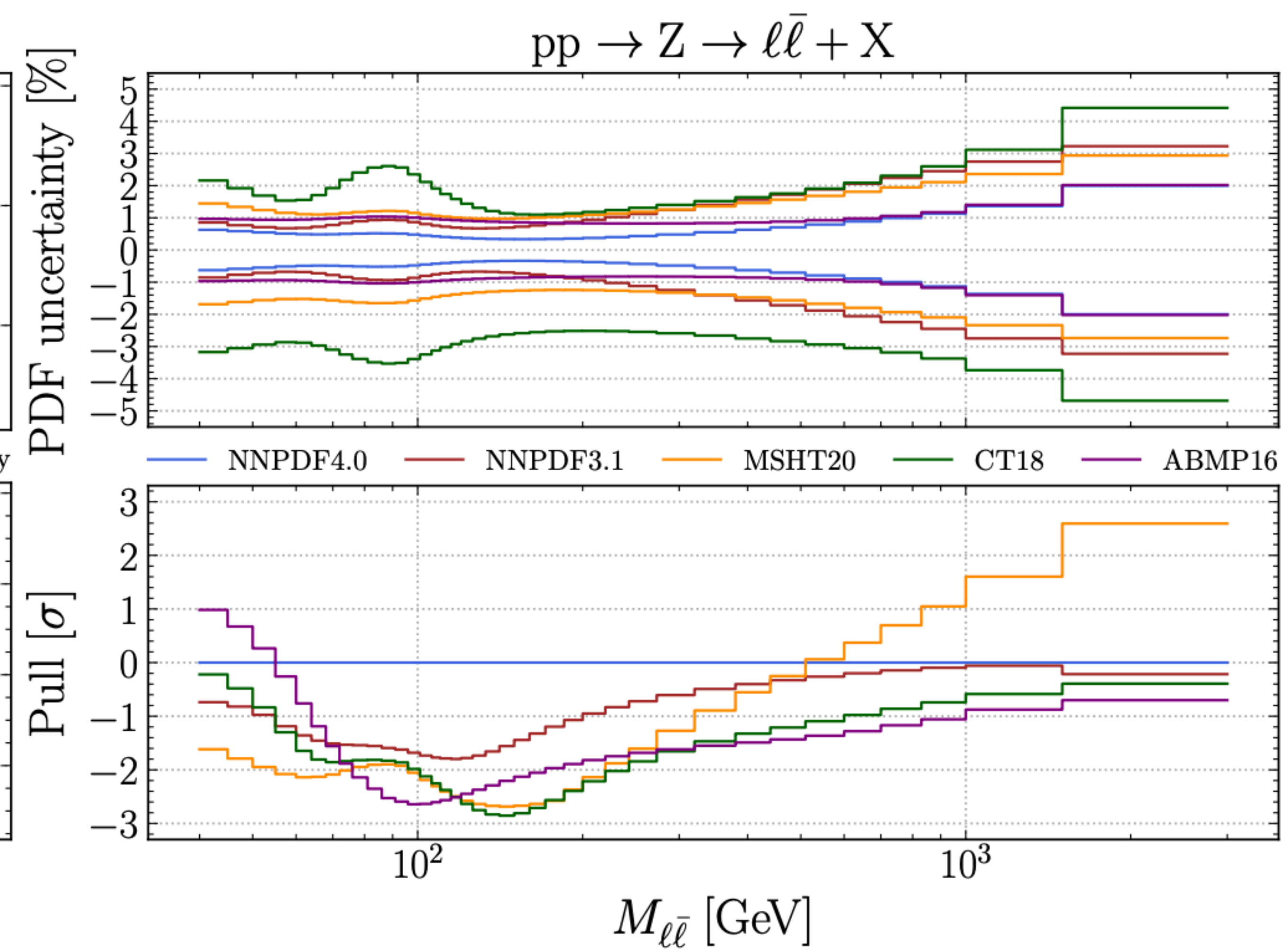
LHC phenomenology

extensive comparisons between global PDF fits for **inclusive** and **differential LHC cross-sections**

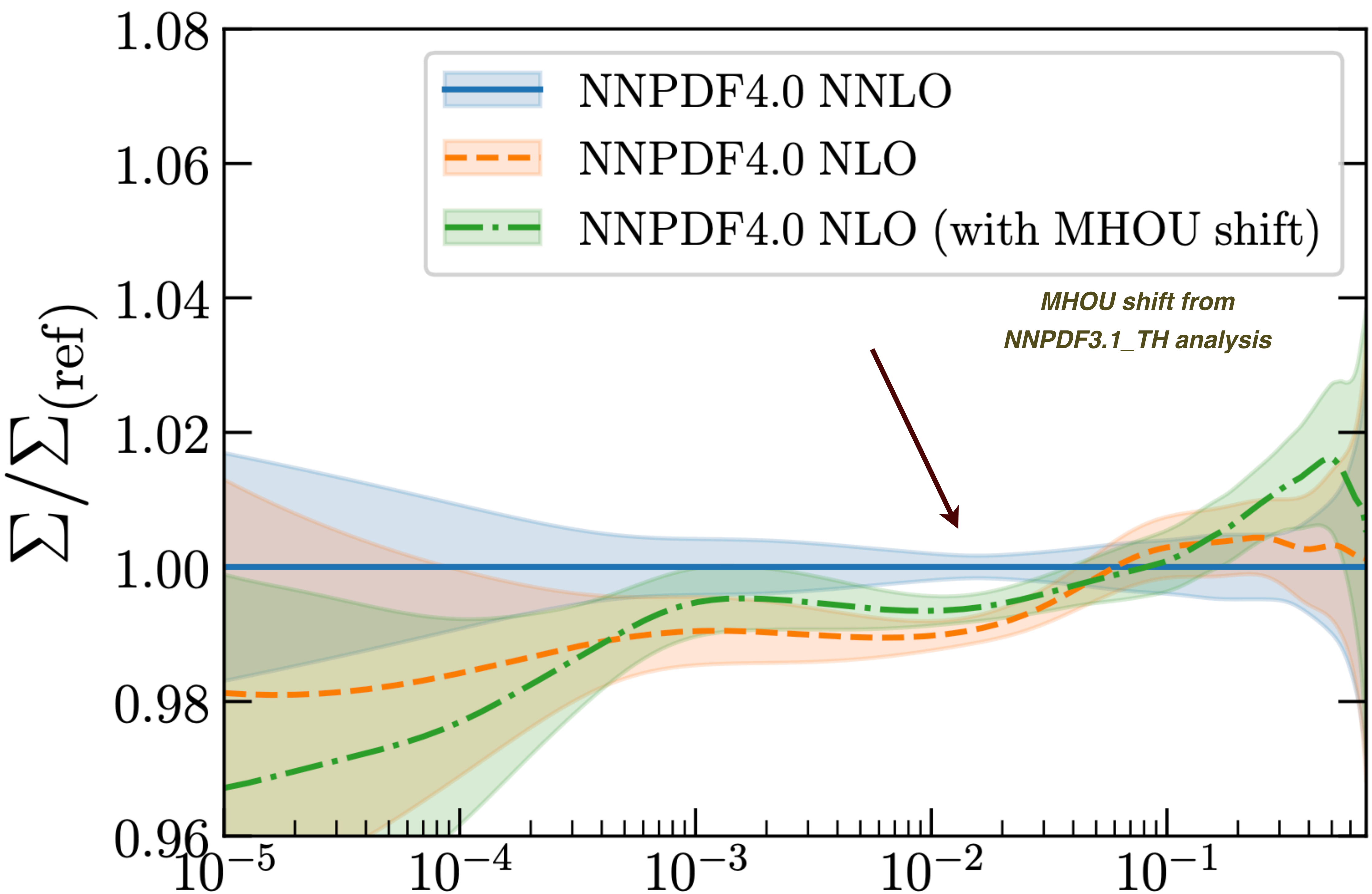








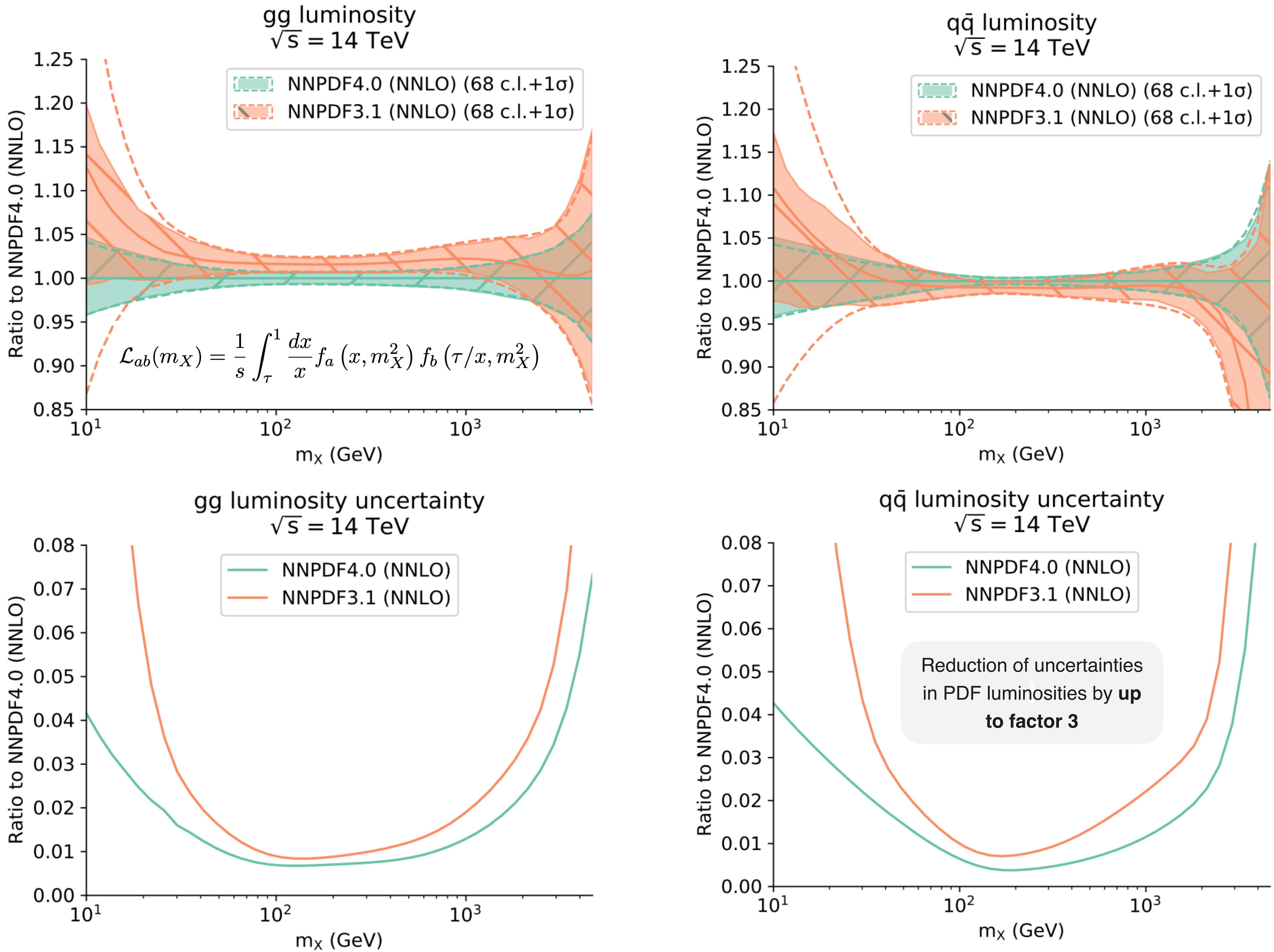
Missing higher order QCD uncertainties



Certainly NLO, but also **likely NNLO PDFs**, underestimate uncertainties without MHOUs

State-of-the-art LHC pheno demands both **NNLO PDFs with MHOUs** and **N3LO PDFs**: WIP!

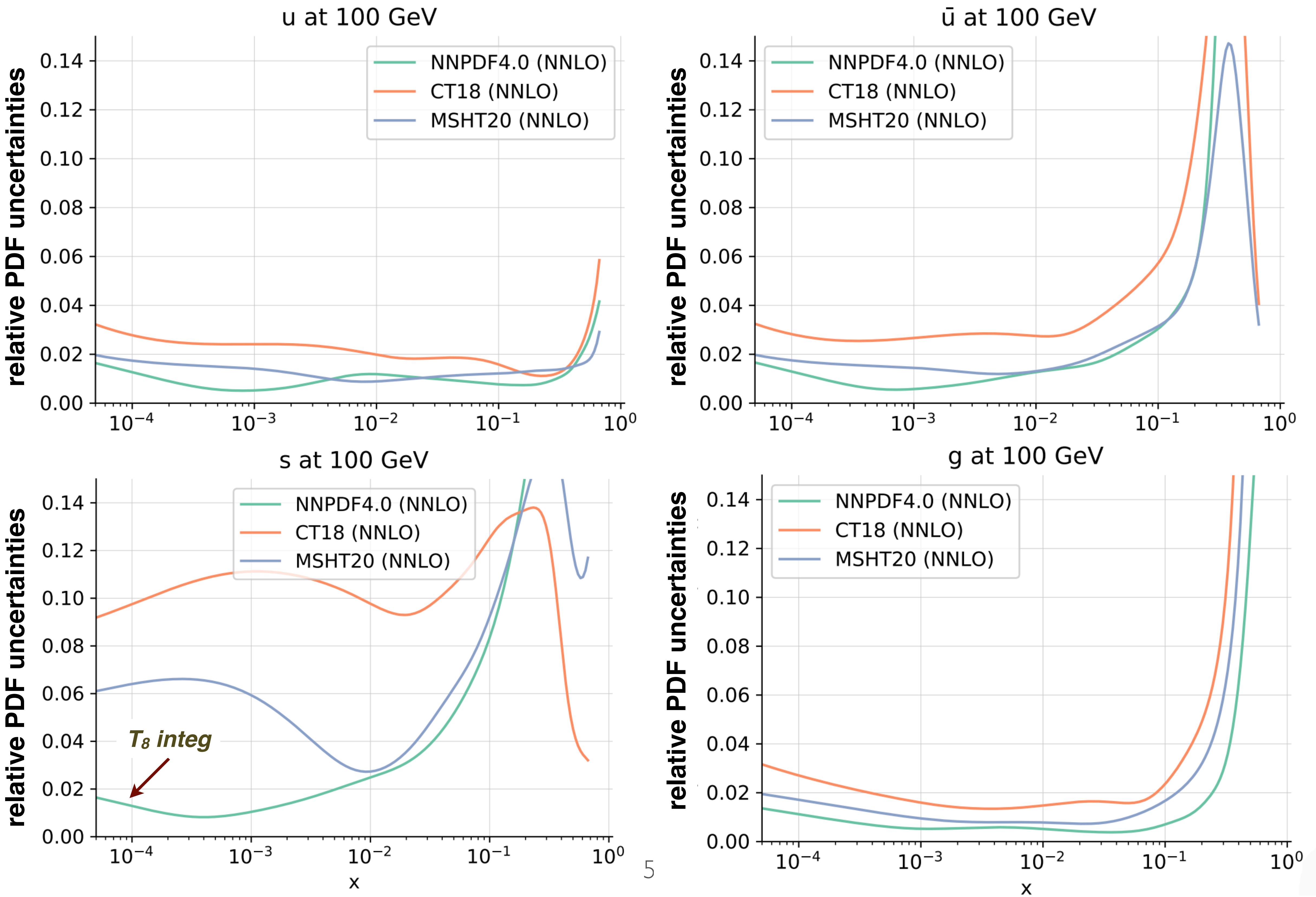
Comparison with NNPDF3.1



Reduction of uncertainties in PDF luminosities by **up to factor 3**

Comparison between global fits

different **pattern of PDF uncertainties** … $\delta_{\rm PDF}({\rm CT}) \gtrsim \delta_{\rm PDF}({\rm MSHT}) \gtrsim \delta_{\rm PDF}({\rm NNPDF})$



Comparison between global fits

… follows pattern of **input datasets**

$\delta_{\rm PDF}({\rm CT}) \gtrsim \delta_{\rm PDF}({\rm MSHT}) \gtrsim \delta_{\rm PDF}({\rm NNPDF})$

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
ATLAS W,Z 7 TeV ($\mathcal{L}=35$ pb^{-1})	[51]	✓	✓	✓	✓	✓
ATLAS W,Z 7 TeV ($\mathcal{L}=4.6$ fb^{-1})	[52]	✓	✓	✗	(✓)	✓
ATLAS low-mass DY 7 TeV	[53]	✓	✓	✗	(✓)	✗
ATLAS high-mass DY 7 TeV	[54]	✓	✓	✗	(✓)	✓
ATLAS W 8 TeV	[79]	✗	(✓)	✗	✗	✓
ATLAS DY 2D 8 TeV	[78]	✗	✓	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	[77]	✗	✓	✗	(✓)	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	[81]	✗	✓	✓	✗	✗
ATLAS W+jet 8 TeV	[93]	✗	✓	✗	✗	✓
ATLAS Z p_T 7 TeV	[259]	(✓)	✗	✗	(✓)	✗
ATLAS Z p_T 8 TeV	[63]	✓	✓	✗	✓	✓
ATLAS $W+c$ 7 TeV	[83]	✗	✓	✗	(✓)	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	[65]	✓	✓	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	[260–265]	✗	✗	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV ($\mathcal{L}=3.2$ fb^{-1})	[66]	✓	✗	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV ($\mathcal{L}=139$ fb^{-1})	[134]	✗	✓	✗	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ and Z ratios	[266]	✗	✗	✗	✗	(✓)
ATLAS $t\bar{t}$ lepton+jets 8 TeV	[67]	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	[89]	✗	✓	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	[73]	✓	(✓)	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	[86]	✗	✓	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	[148]	✗	✓	✗	✗	✗
ATLAS direct photon production 8 TeV	[100]	✗	(✓)	✗	✗	✗
ATLAS direct photon production 13 TeV	[101]	✗	✓	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	[94, 96, 98]	✗	✓	✓	✗	✗
ATLAS single top diff. 7 TeV	[94]	✗	✓	✗	✗	✗
ATLAS single top diff. 8 TeV	[96]	✗	✓	✗	✗	✗

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
CMS W asym. 7 TeV ($\mathcal{L}=36$ pb^{-1})	[267]	✗	✗	✗	✗	✓
CMS Z 7 TeV ($\mathcal{L}=36$ pb^{-1})	[268]	✗	✗	✗	✗	✓
CMS W electron asymmetry 7 TeV	[55]	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	[56]	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	[57]	✓	✓	✗	(✓)	✓
CMS Drell-Yan 2D 8 TeV	[269]	(✓)	✗	✗	✗	✗
CMS W rapidity 8 TeV	[58]	✓	✓	✓	✓	✓
CMS W,Z p_T 8 TeV ($\mathcal{L}=18.4$ fb^{-1})	[270]	✗	✗	✗	(✓)	✗
CMS Z p_T 8 TeV	[64]	✓	✓	✗	(✓)	✗
CMS $W+c$ 7 TeV	[76]	✓	✓	✗	(✓)	✓
CMS $W+c$ 13 TeV	[84]	✗	✓	✗	✗	(✓)
CMS single-inclusive jets 2.76 TeV	[75]	✓	✗	✗	✗	✓
CMS single-inclusive jets 7 TeV	[147]	✓	(✓)	✗	✓	✓
CMS dijets 7 TeV	[74]	✗	✓	✗	✗	✗
CMS single-inclusive jets 8 TeV	[87]	✗	✓	✗	✓	✓
CMS 3D dijets 8 TeV	[149]	✗	(✓)	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 5 TeV	[88]	✗	✓	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	[146]	✓	✓	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 8 TeV	[271]	✗	✗	✗	✗	✓
CMS $\sigma_{tt}^{\rm tot}$ 5, 7, 8, 13 TeV	[68, 272–280]	✗	✗	✓	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 13 TeV	[69]	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	[70]	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	[90]	✗	✓	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	[91]	✗	✓	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	[92]	✗	✓	✗	✗	✗
CMS single top $\sigma_t+\sigma_{\bar{t}}$ 7 TeV	[95]	✗	✓	✓	✗	✗
CMS single top R_t 8, 13 TeV	[97, 99]	✗	✓	✓	✗	✗
CMS single top 13 TeV	[281, 282]	✗	✗	✗	✗	(✓)

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
LHCb Z 7 TeV ($\mathcal{L}=940$ pb^{-1})	[59]	✓	✓	✗	✗	✓
LHCb $Z\to ee$ 8 TeV ($\mathcal{L}=2$ fb^{-1})	[61]	✓	✓	✓	✓	✓
LHCb W 7 TeV ($\mathcal{L}=37$ pb^{-1})	[283]	✗	✗	✗	✗	✓
LHCb $W,Z\to\mu$ 7 TeV	[60]	✓	✓	✓	✓	✓
LHCb $W,Z\to\mu$ 8 TeV	[62]	✓	✓	✓	✓	✓
LHCb $W\to e$ 8 TeV	[80]	✗	(✓)	✗	✗	✗
LHCb $Z\to\mu\mu, ee$ 13 TeV	[82]	✗	✓	✗	✗	✗

✓	in baseline dataset
✗	not considered
(✓)	impact assessed but excluded from baseline