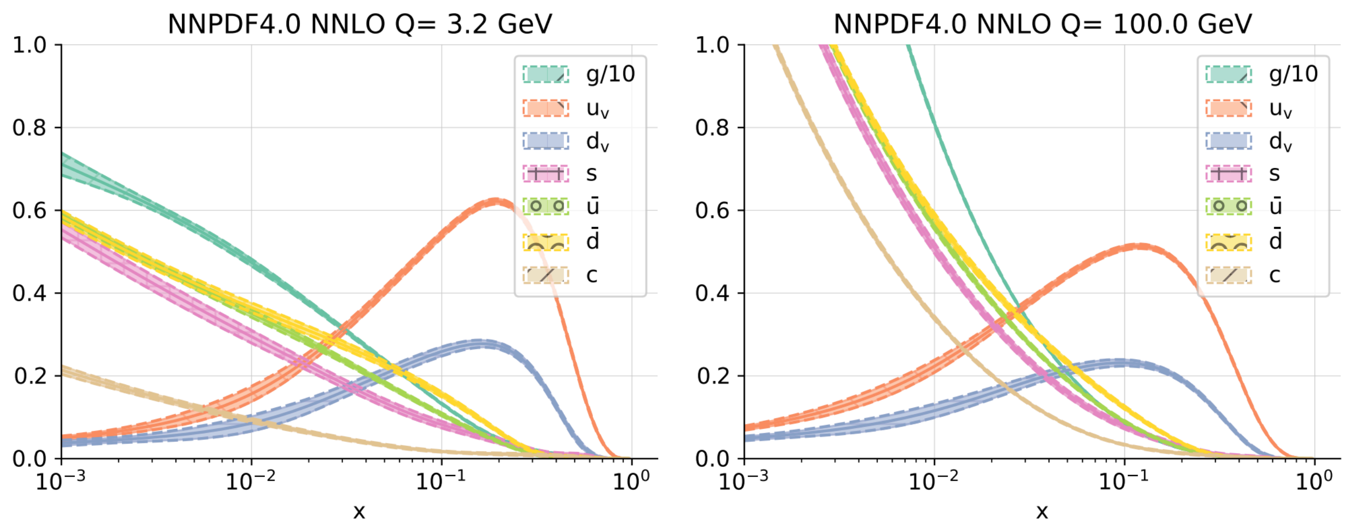
Parton Distributions for the LHC Run III and beyond

Juan Rojo, VU Amsterdam & Nikhef

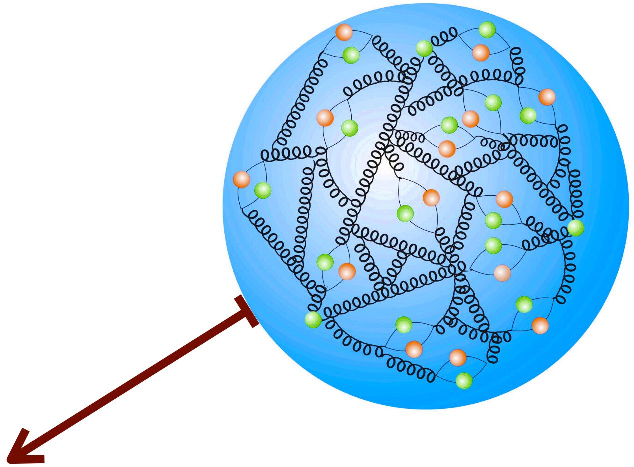
ATLAS Collaboration Week Meeting

13th October 2021



Why Parton Distributions?

Why Parton Distributions?



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

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

LHC master formula

hard cross-section

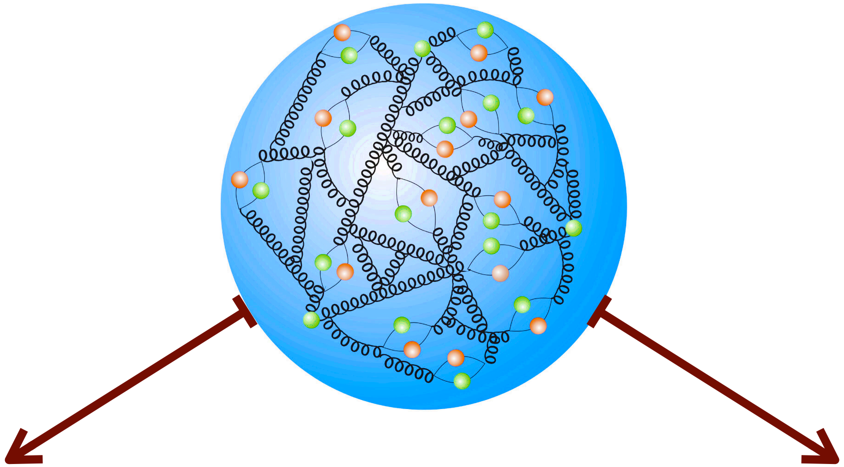
$$\sigma(M,s) \propto \sum_{ij=u,d,g,\ldots} \int_{M^2}^{s} d\hat{s}\; \mathscr{L}_{ij}(\hat{s},s)\, \widetilde{\sigma}_{ij}(\hat{s},\alpha_s(M))$$

partonic luminosities

parton distributions

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

Why Parton Distributions?



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

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

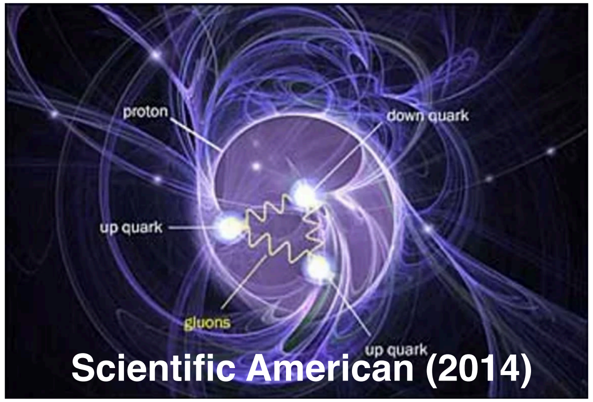
Address fundamental questions about Quantum Chromodynamics

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

PDFs: 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

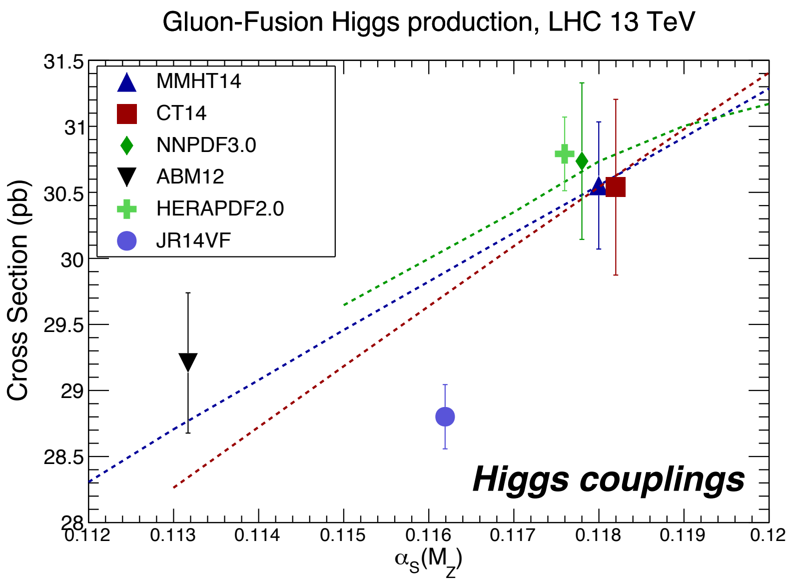
After 40 years of studying the strong nuclear force, a revelation

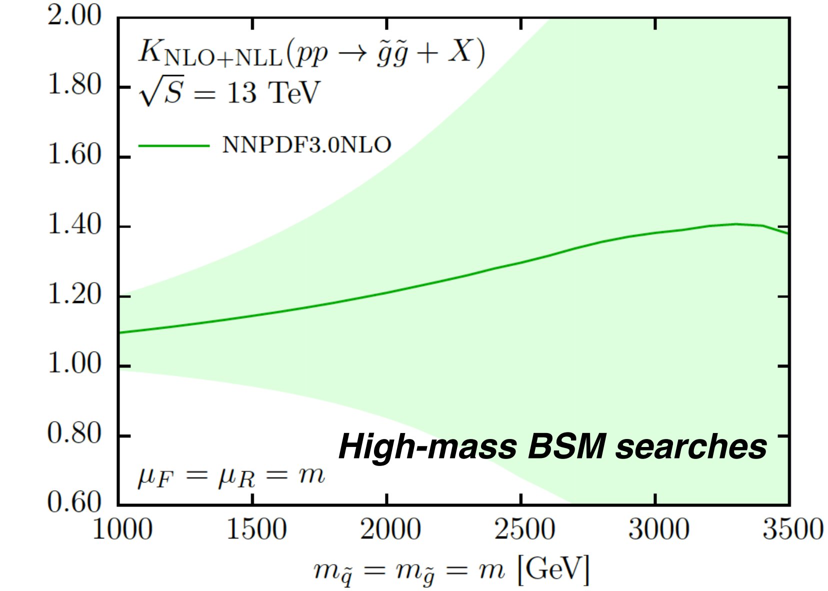
BFKL dynamics

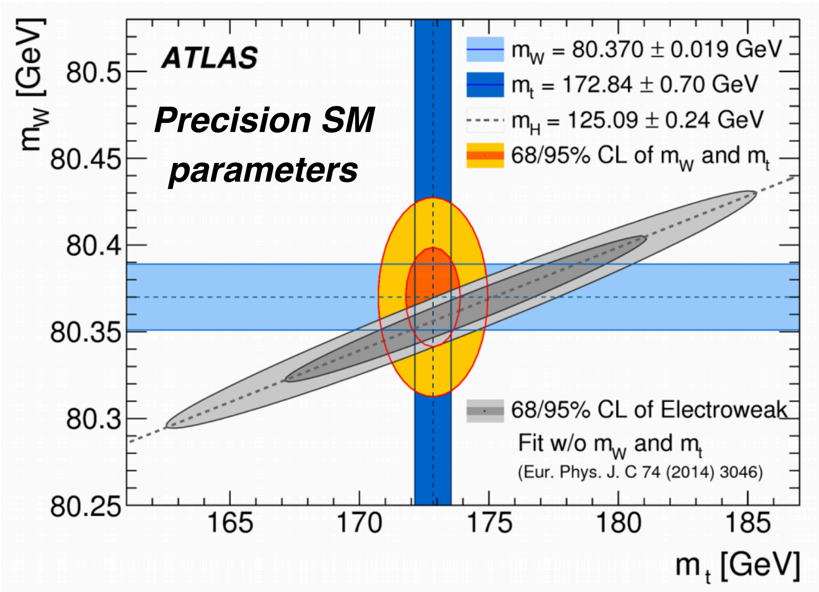
This was the year that analysis of data finally backed up a prediction, made in the mid 1970s, of a surprising emergent behaviour in the strong nuclear force

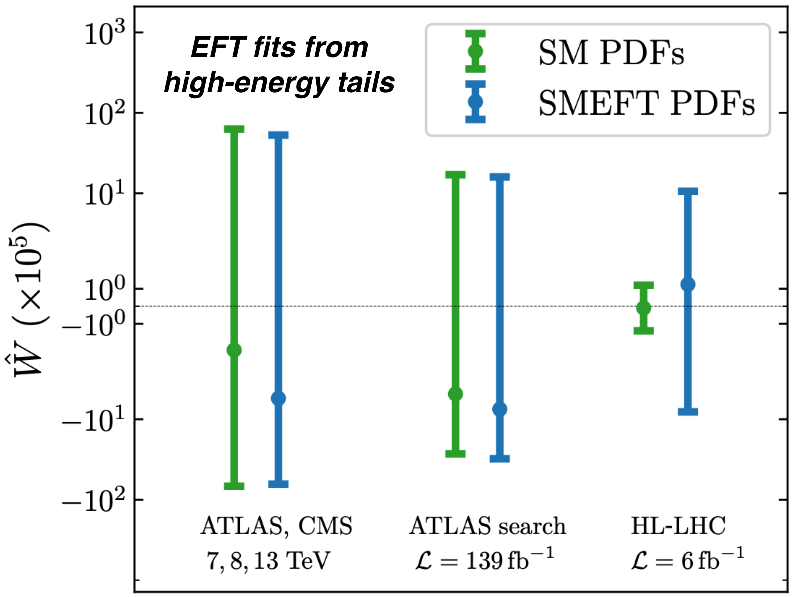
PDFs: realising precision physics @ LHC

PDF uncertainties are limiting factor in **theoretical interpretation** for many LHC analysis









Progress in global PDF analyses

Novel experimental constraints

Improved theory

Methodological developments

Parton Distributions at the Dawn of Run III

	NNPDF4.0	MSHT20	CT18	ABMP16
Released	*Sept 2021*: LHAPDF grids + **fitting code**	*Dec 2020*: LHAPDF grids	*Dec 2019*: LHAPDF grids	*Jan 2017*: LHAPDF grids
Parametrisation	Neural networks (**hyperoptimised**)	Functional form + Chebyshev	Functional form + Bernstein	Functional form
Error estimate	Monte Carlo (**closure + future tested**)	Hessian (dynamic tolerance)	Hessian (dynamic tolerance) + Lagrange mult.	Hessian (no tolerance)
Theory settings	NNLO QCD, GM-VFN (+ NLO electroweak)	NNLO QCD GM-VFN	NNLO QCD GM-VFN	NNLO QCD, FFN

+ PDF4LHC15 (Combination of CT14, MMHT14, NNPDF3.0), released Oct 2015

Dataset comparison for global fits

$$N_{\text{data}}(\text{CT18}) \lesssim N_{\text{data}}(\text{MSHT20}) \lesssim N_{\text{data}}(\text{NNPDF4.0})$$

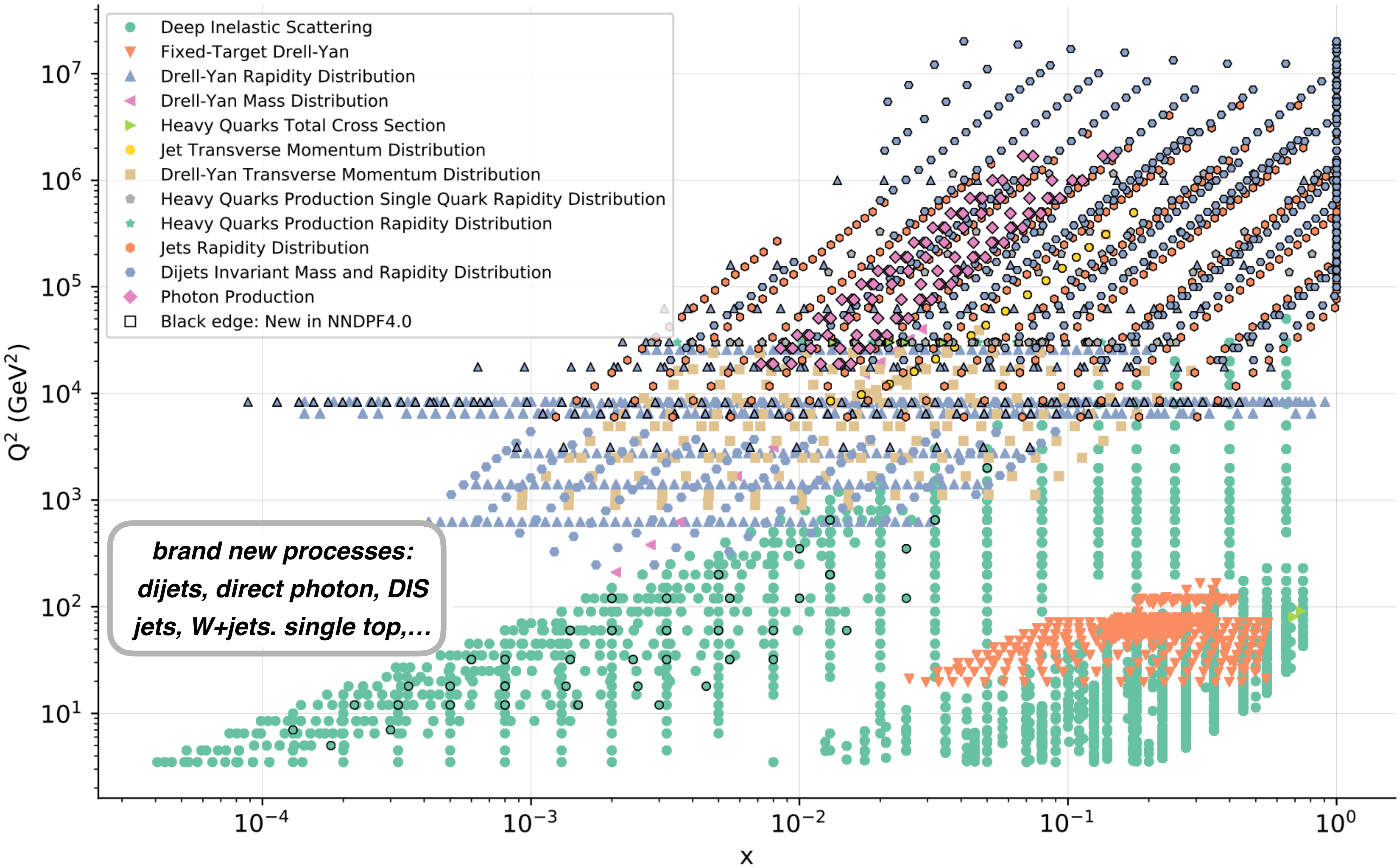
Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
ATLAS W,Z 7 TeV ($\mathcal{L}=35$ pb^{-1})	[51]	✓	✓	✓	✓	✓
ATLAS W,Z 7 TeV ($\mathcal{L}=4.6$ fb^{-1})	[52]	✓	✓	✗	(✓)	✓
ATLAS low-mass DY 7 TeV	[53]	✓	✓	✗	(✓)	✗
ATLAS high-mass DY 7 TeV	[54]	✓	✓	✗	(✓)	✓
ATLAS W 8 TeV	[79]	✗	(✓)	✗	✗	✓
ATLAS DY 2D 8 TeV	[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 σ_{tt}^{tot} 7, 8 TeV	[65]	✓	✓	✓	✗	✗
ATLAS σ_{tt}^{tot} 7, 8 TeV	[260–265]	✗	✗	✓	✗	✗
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L}=3.2$ fb^{-1})	[66]	✓	✗	✓	✗	✗
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L}=139$ fb^{-1})	[134]	✗	✓	✗	✗	✗
ATLAS σ_{tt}^{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 σ_{tt}^{tot} 5 TeV	[88]	✗	✓	✗	✗	✗
CMS σ_{tt}^{tot} 7, 8 TeV	[146]	✓	✓	✗	✗	✗
CMS σ_{tt}^{tot} 8 TeV	[271]	✗	✗	✗	✗	✓
CMS σ_{tt}^{tot} 5, 7, 8, 13 TeV	[68, 272–280]	✗	✗	✓	✗	✗
CMS σ_{tt}^{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

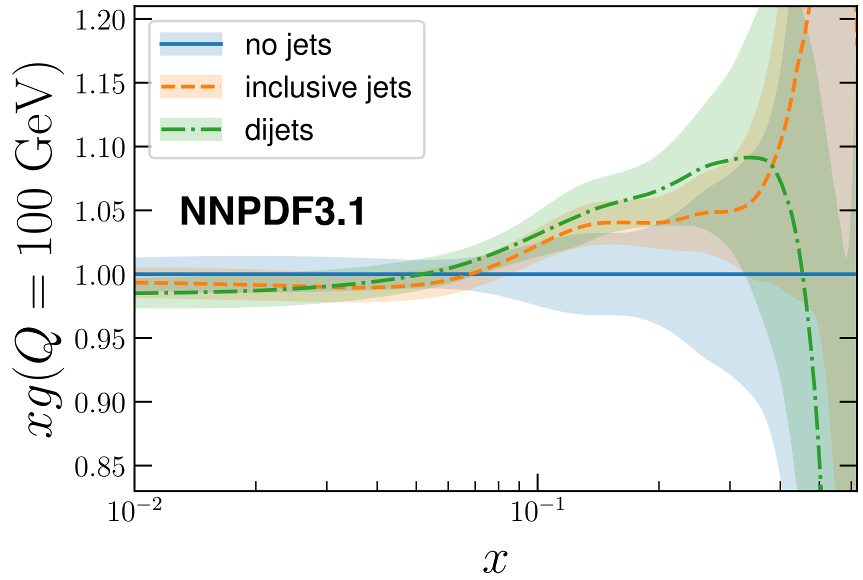
The NNPDF4.0 dataset



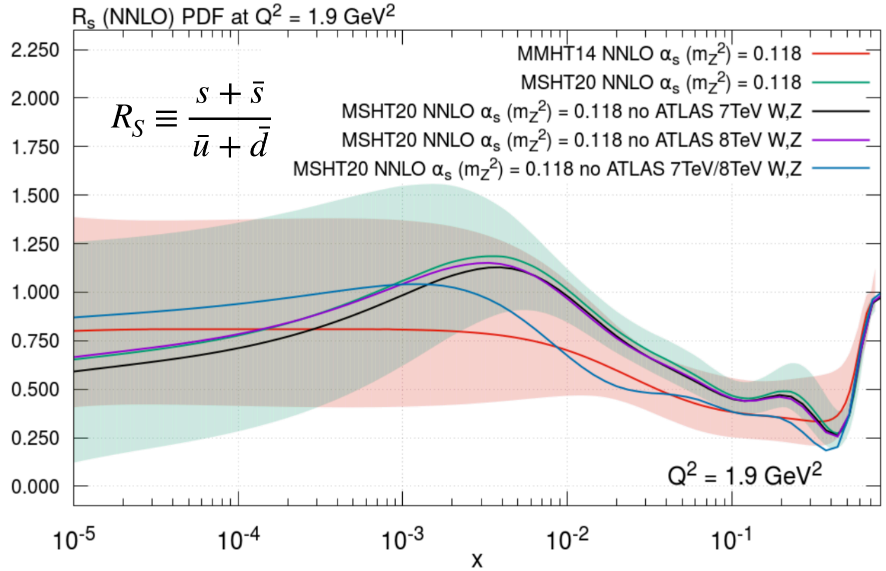
$\mathcal{O}(50)$ data sets investigated; $\mathcal{O}(400)$ data points more in NNPDF4.0 than in NNPDF3.1

Novel experimental constraints

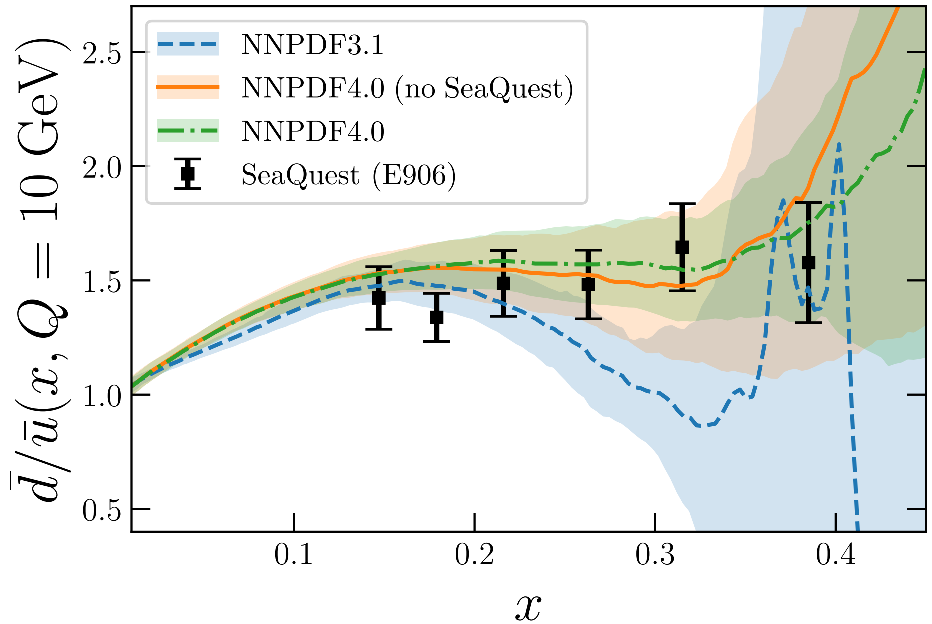
Dijet production



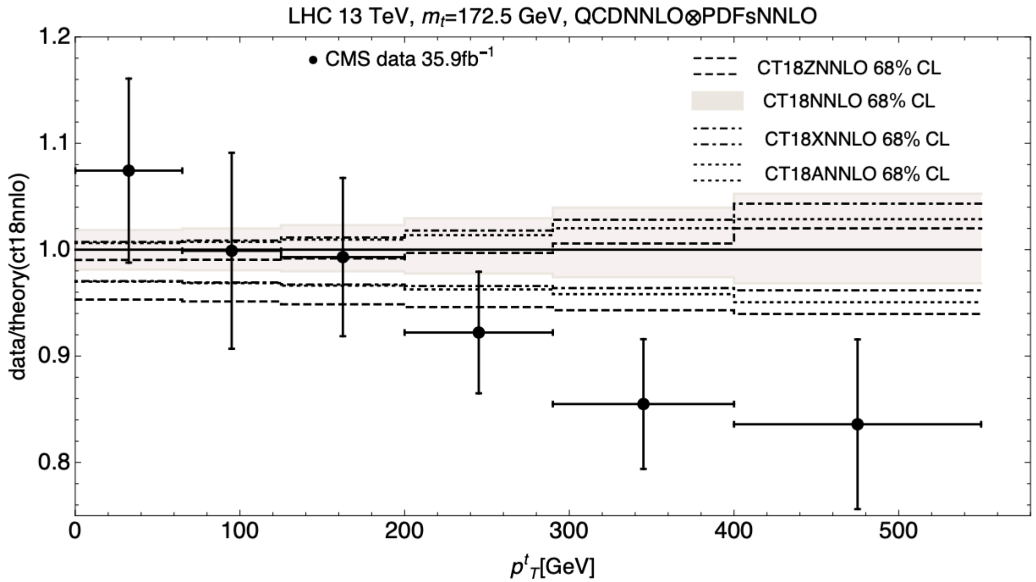
Collider Drell-Yan



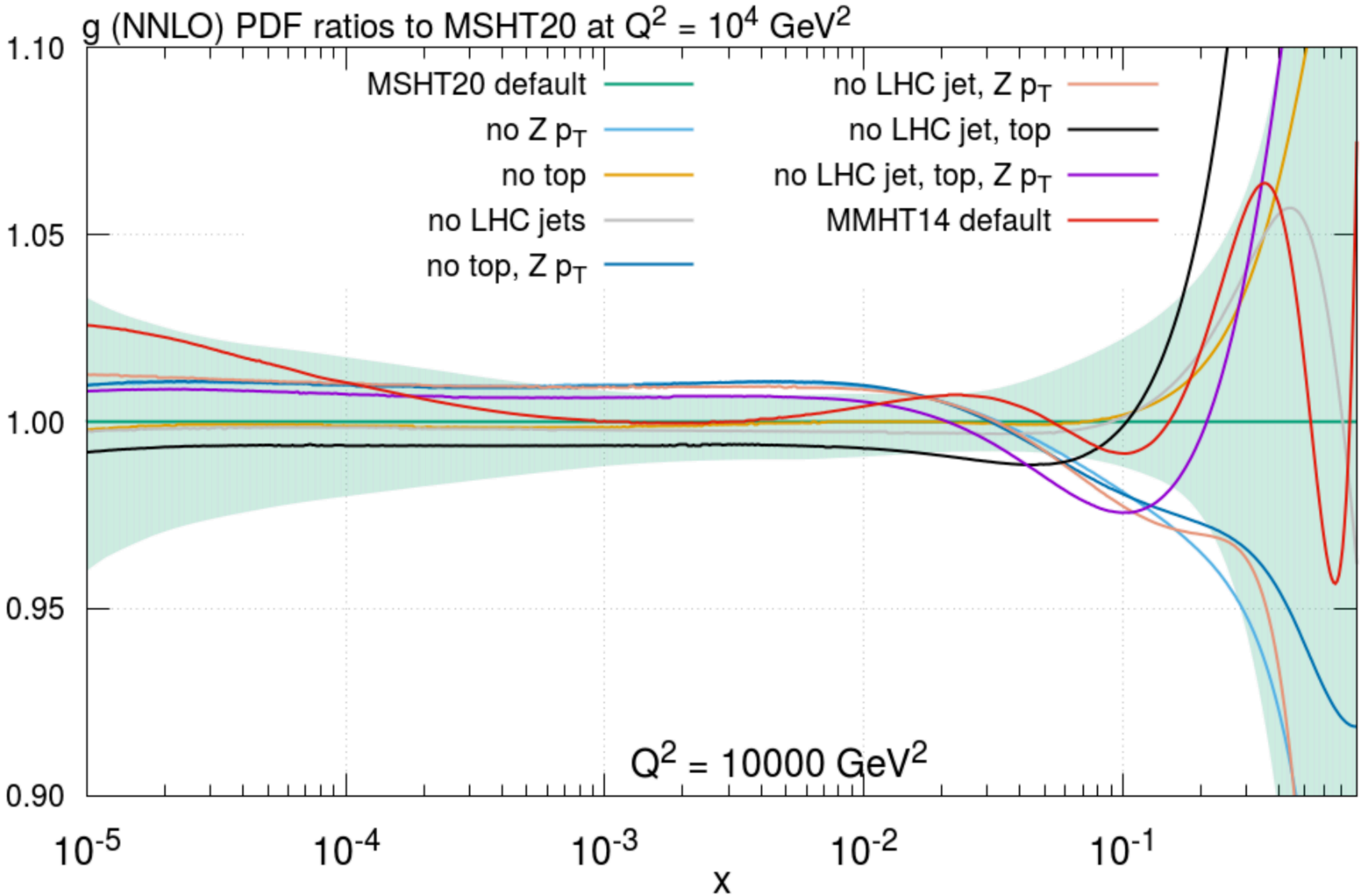
SeaQuest Drell-Yan



differential top quark pair



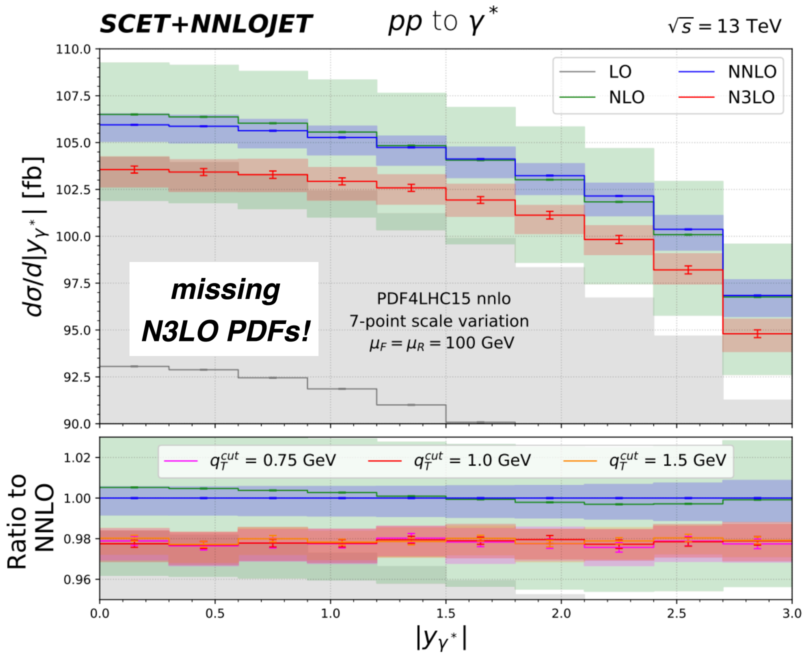
Novel experimental constraints



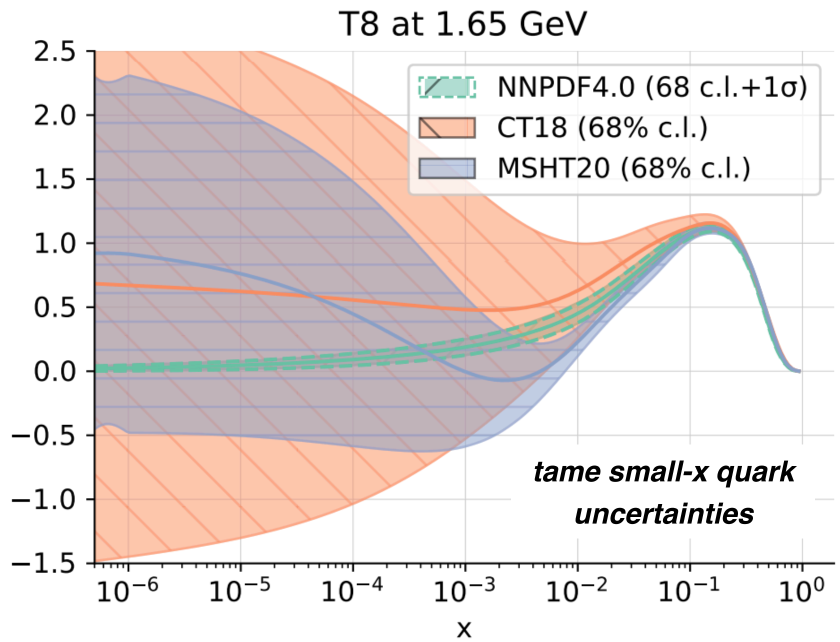
Global fits benefit from **redundancy:** a given PDF combination is constrained by many processes

Improved theory calculations

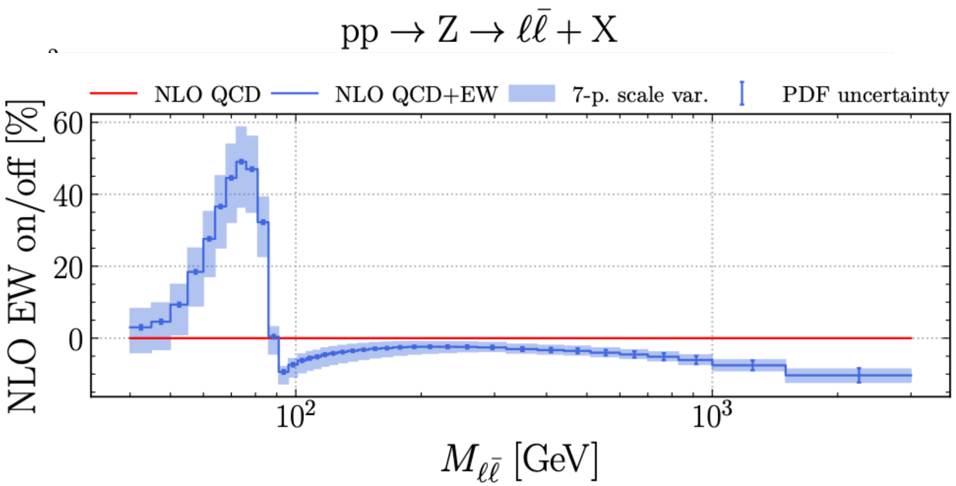
NNLO & N3LO QCD



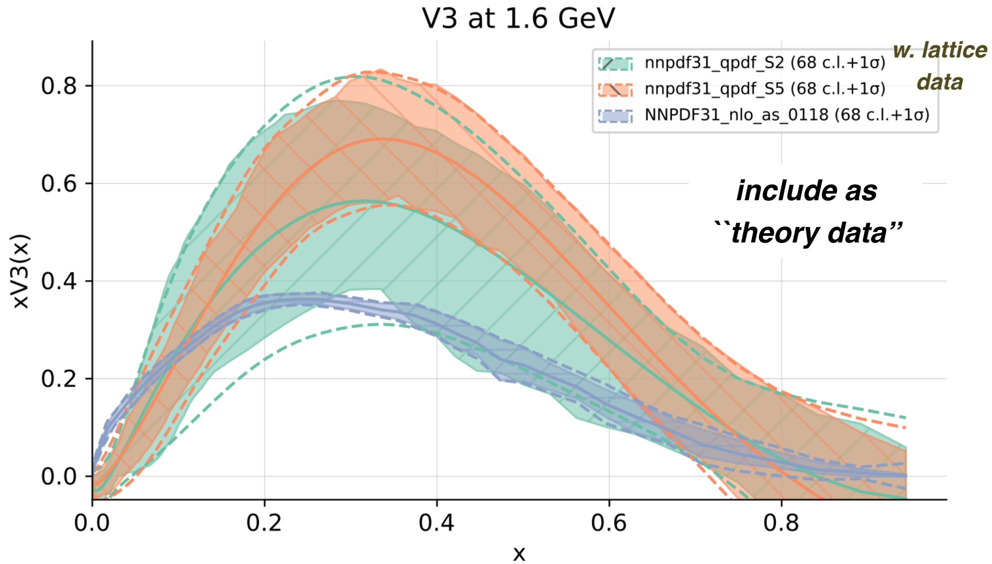
Positivity & integrability of PDFs



NLO EWK corrections

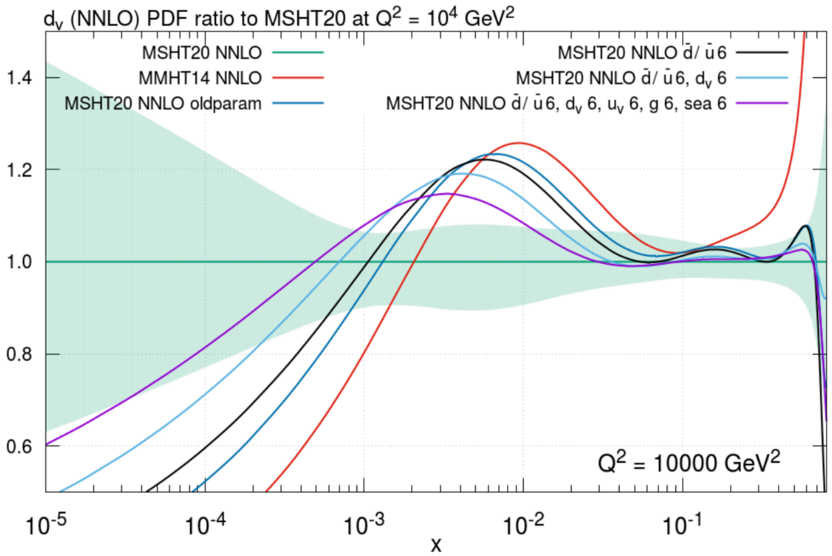


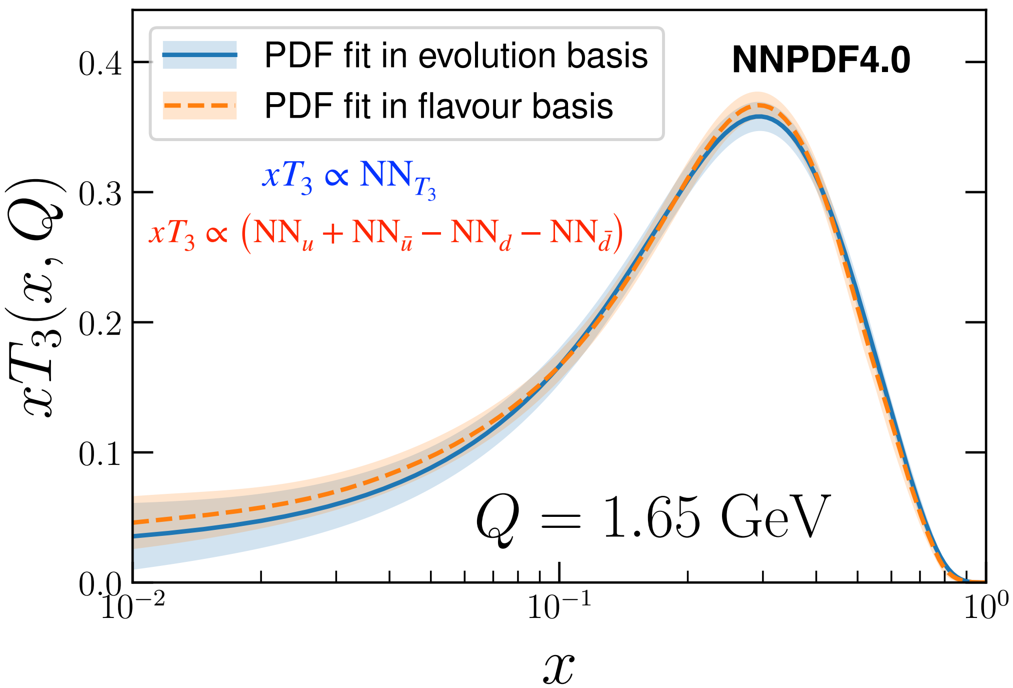
Lattice QCD constraints



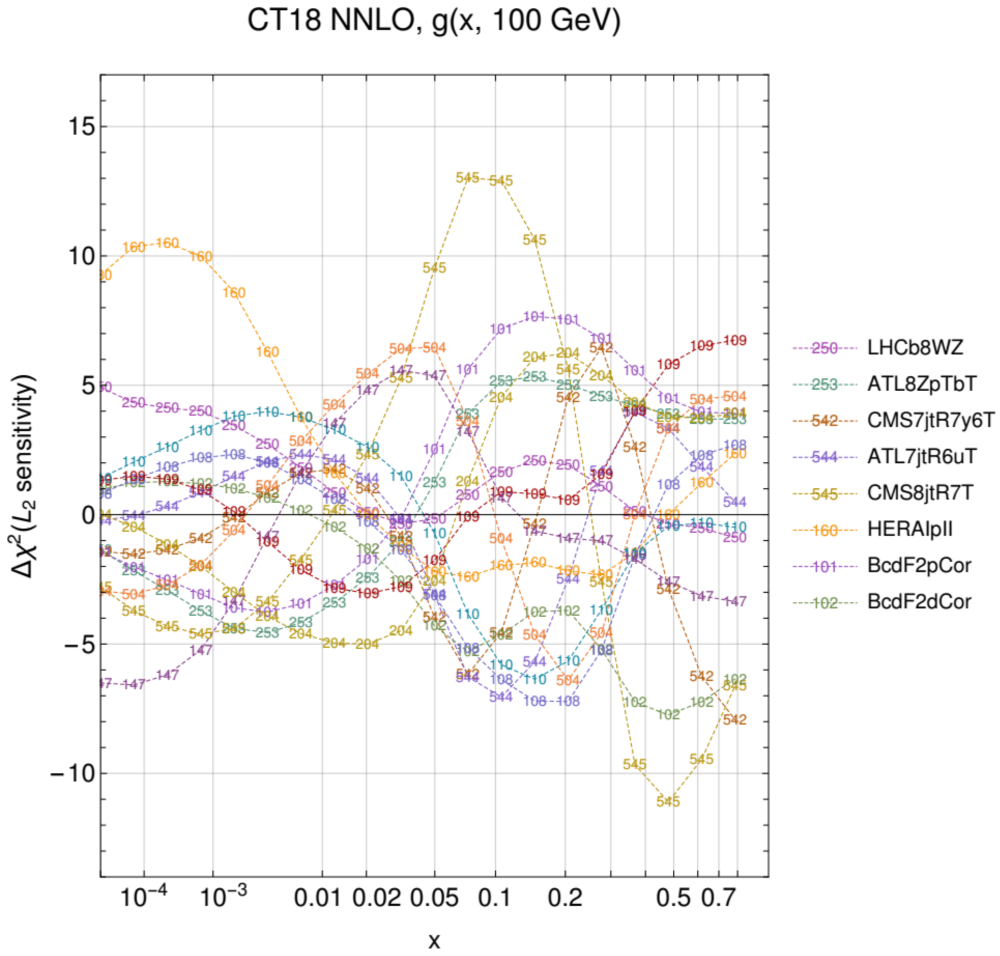
Methodological developments

more flexible PDF parameterisations





new measures of PDF sensitivity



parameterisation-basis independent PDF determinations

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:

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

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

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

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?

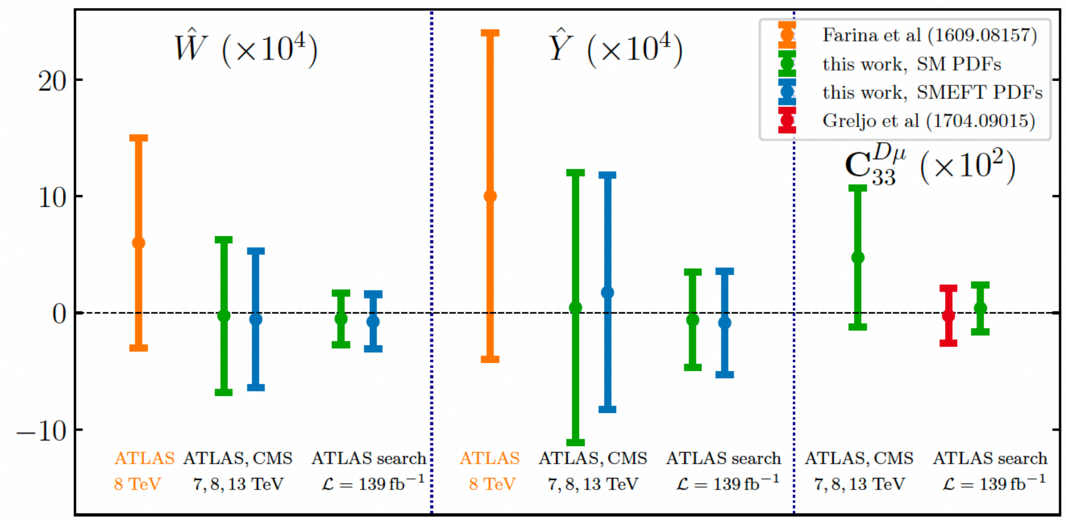
Exp.	$\sqrt{s}$ (TeV)	Ref.	$\mathcal{L}$ (fb^{-1})	Channel	1D/2D	$n_{\rm dat}$	$m_{\ell\ell}^{\rm max}$ (TeV)
ATLAS	7	[120]	4.9	e^-e^+	1D	13	[1.0, 1.5]
ATLAS (*)	8	[86]	20.3	$\ell^-\ell^+$	2D	46	[0.5, 1.5]
CMS	7	[121]	9.3	$\mu^-\mu^+$	2D	127	[0.2, 1.5]
CMS (*)	8	[87]	19.7	$\ell^-\ell^+$	1D	41	[1.5, 2.0]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

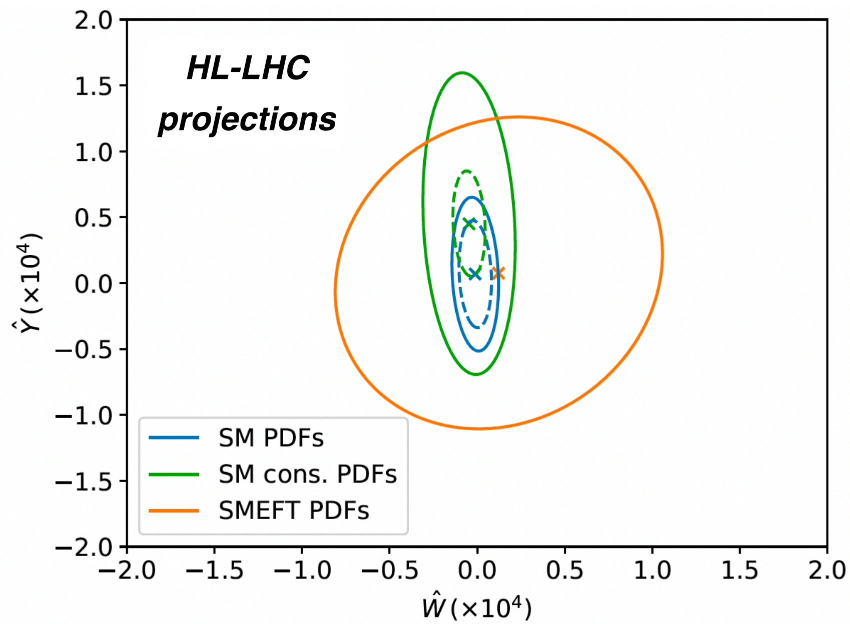
$$d\sigma_{\rm SMEFT} = d\sigma_{\rm SM} \times K_{\rm EFT}$$

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

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



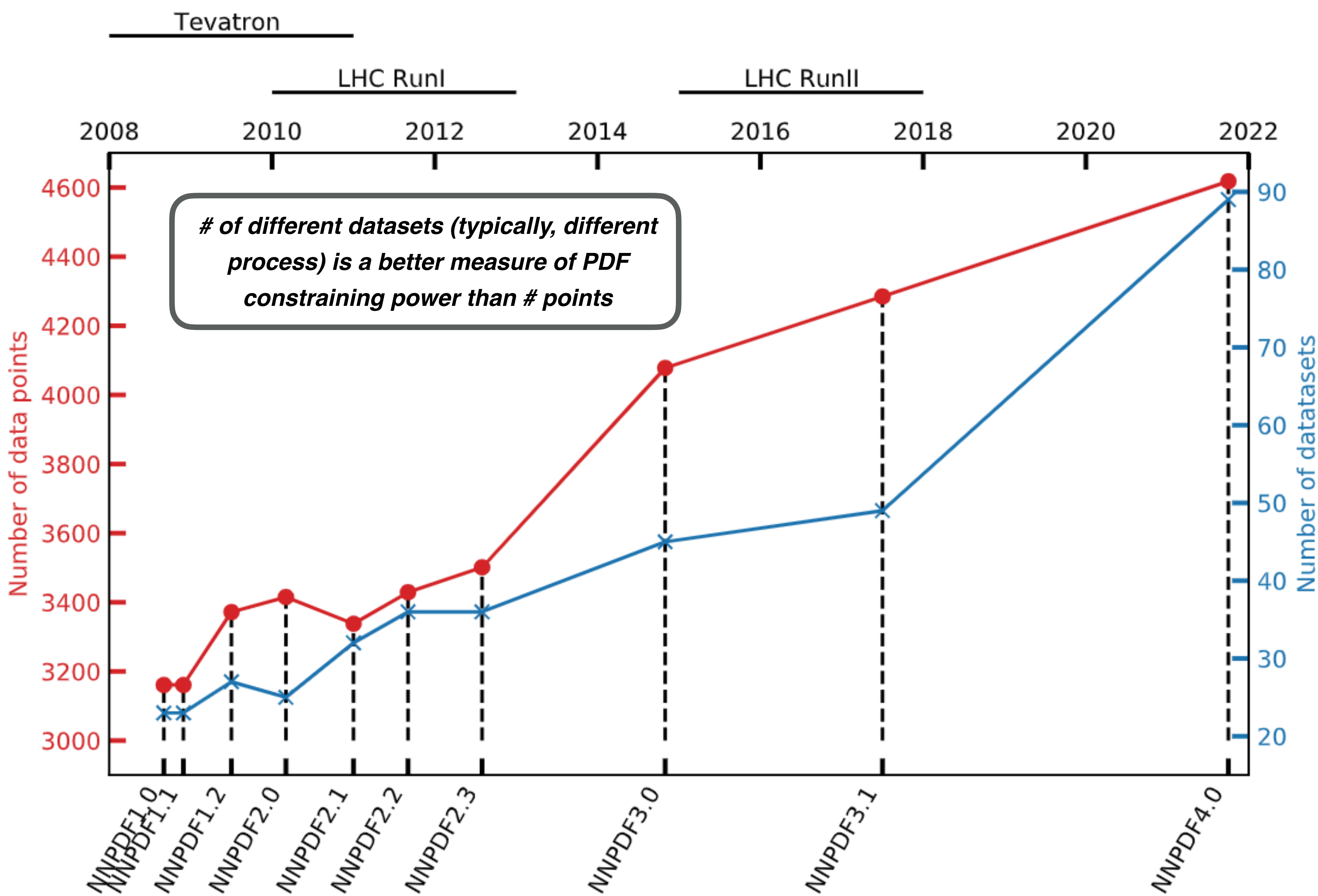
HL-LHC: EFT effects, if present, would be **reabsorbed into PDFs**



Carrazza et al 19, Greljo et al 21

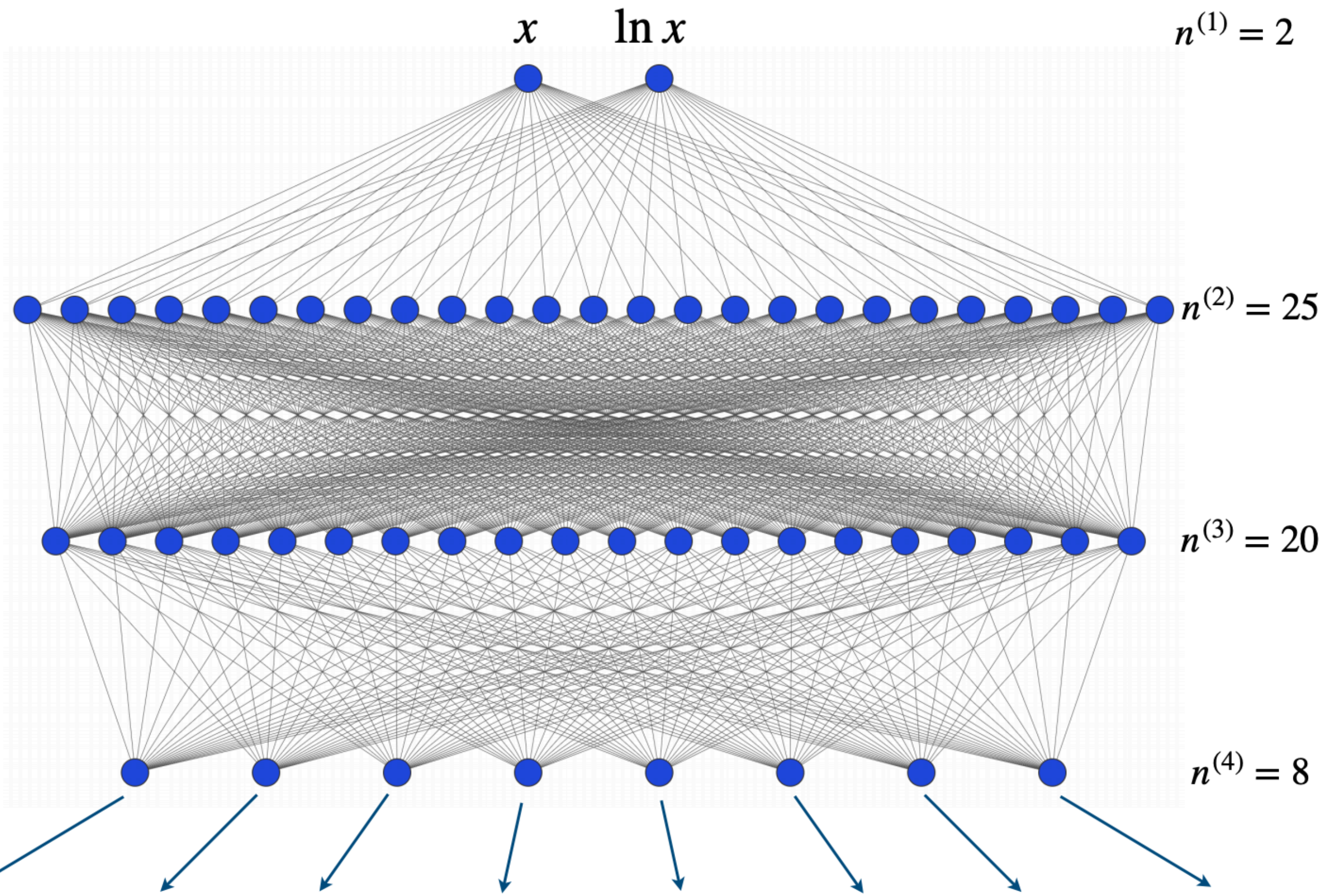
NNPDF4.0

From NNPDF1.0 to NNPDF4.0



Improved fitting methodology

- ☑ **Stochastic Gradient Descent** via TensorFlow for NN 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)



$xg(x,Q_0)$	$x\Sigma(x,Q_0)$	$xV(x,Q_0)$	$xV_3(x,Q_0)$	$xV_8(x,Q_0)$	$xT_3(x,Q_0)$	$xT_8(x,Q_0)$	$xT_{15}(x,Q_0)$	***evolution basis***
$xg(x,Q_0)$	$xu(x,Q_0)$	$x\bar{u}(x,Q_0)$	$xd(x,Q_0)$	$x\bar{d}(x,Q_0)$	$xs(x,Q_0)$	$x\bar{s}(x,Q_0)$	$xc^+(x,Q_0)$	***flavor basis***

PDFs independent of **parametrisation basis!**

Improved fitting methodology

- ☑ **Stochastic Gradient Descent** via TensorFlow for NN 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)

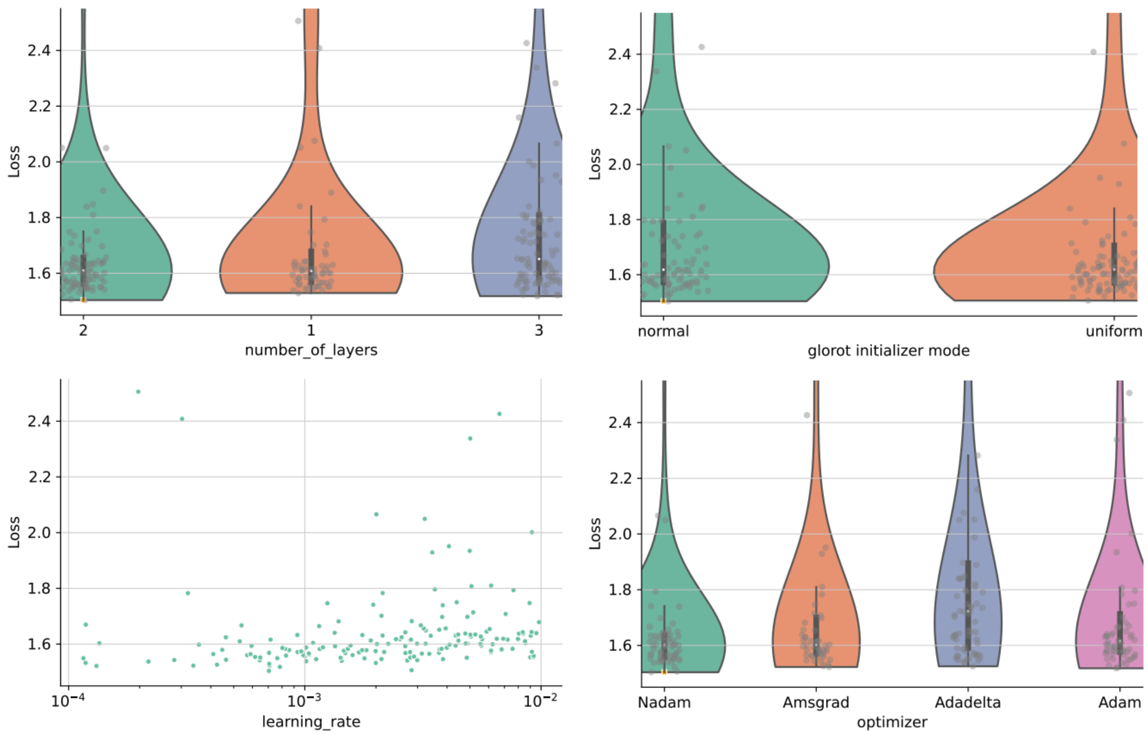
Loss (``average”)

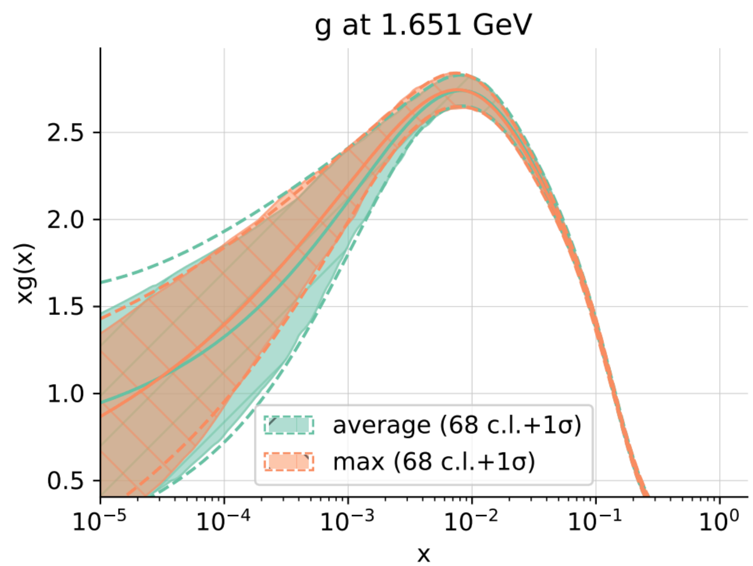
ML model hyperparams

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

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

Closure and future tests

Closure tests

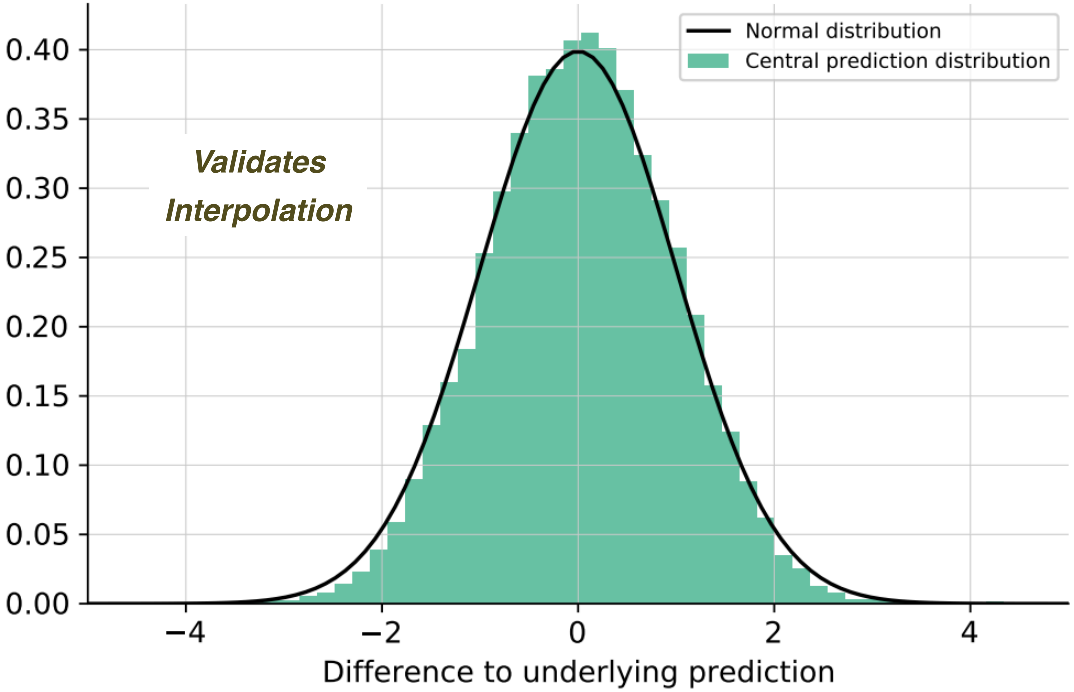
Generate **toy data** based on some known PDF, check *a posteriori* that the **true underlying law is reproduced** within errors

mean NN prediction *true central value*

generate many toys

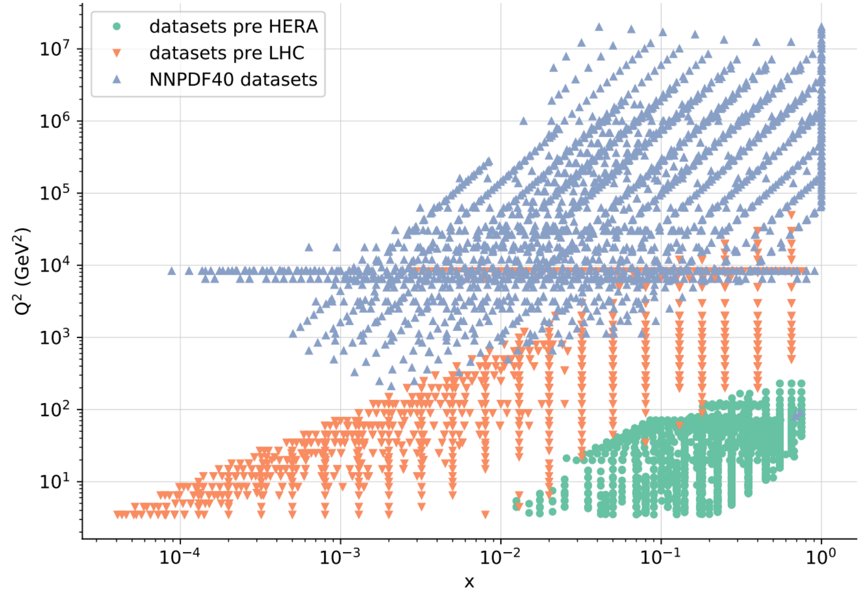
$$\delta_i^{(l)} \equiv \frac{\left(\mathbf{E}_\epsilon\left[g_i\right]^{(l)} - f_i\right)}{\sigma_i^{(l)}}$$

data index *PDF uncertainty*



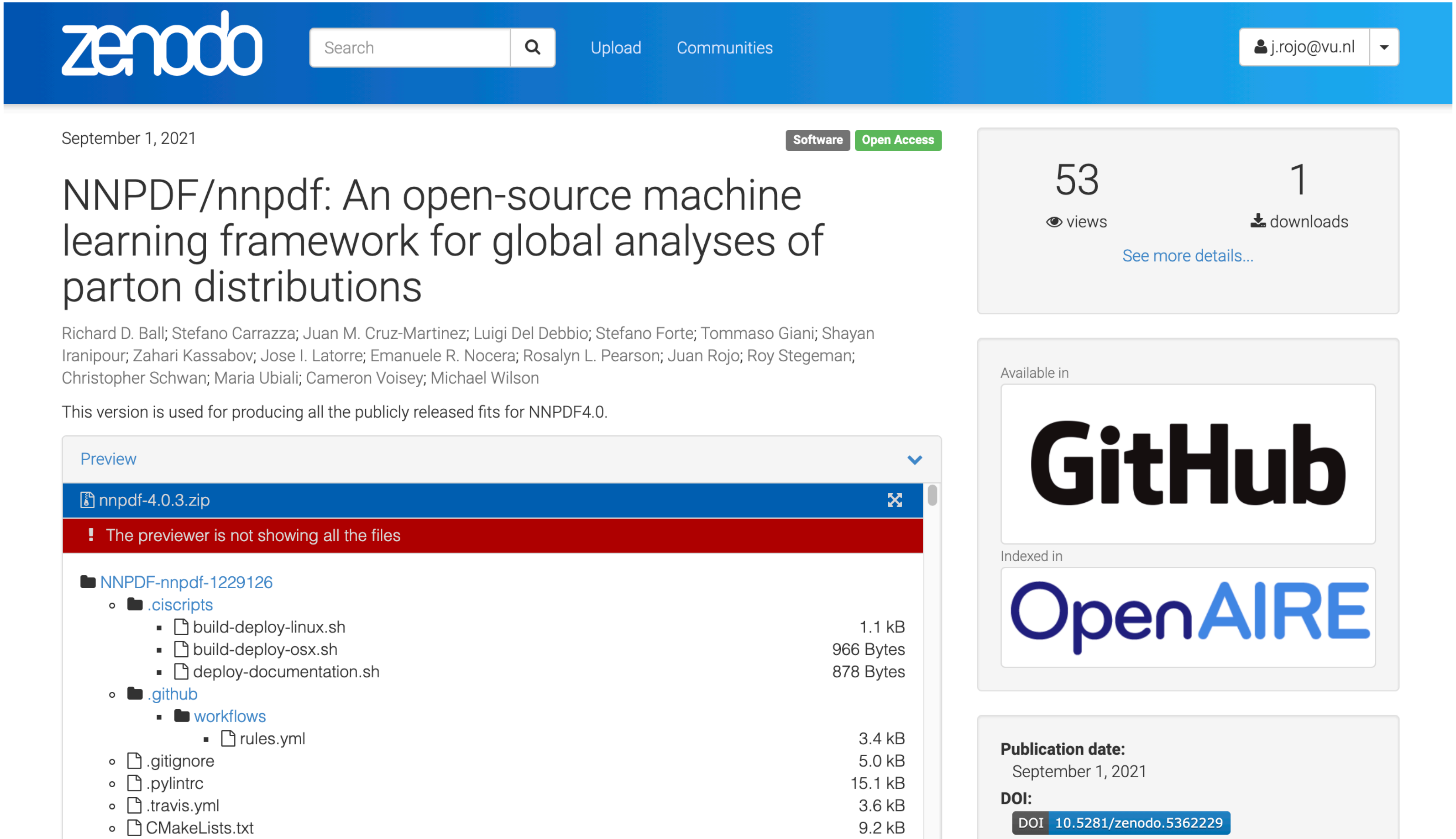
Future tests

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



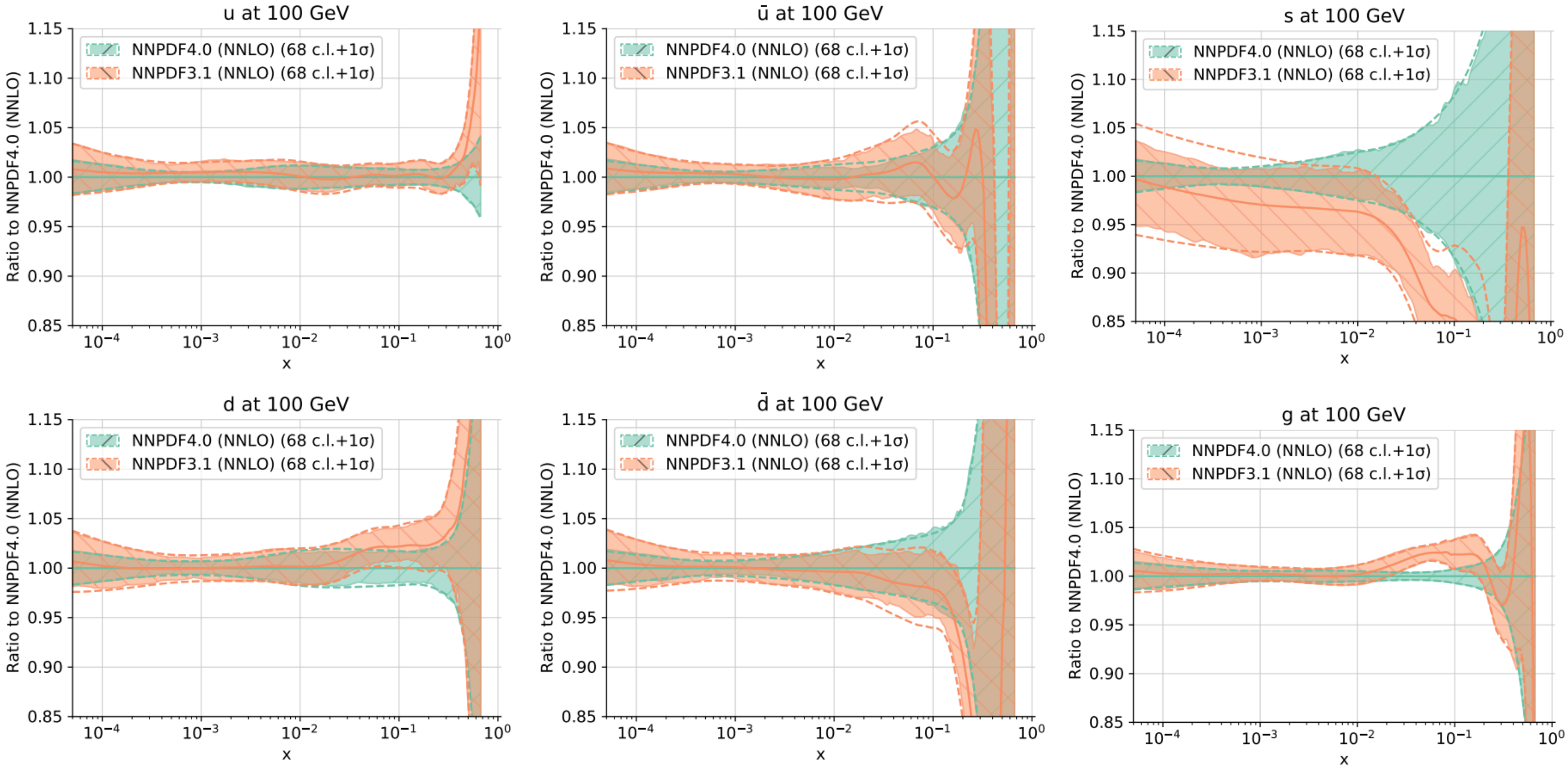
Process	χ^2 pre-HERA	χ^2 pre-LHC	χ^2 Global
Fixed target NC DIS	1.05	1.18	1.23
Fixed target CC DIS	0.80	0.85	0.87
Fixed target Drell-Yan	0.92	1.27	1.59
HERA	*27.20* (**1.23**)	1.22	1.20
Collider Drell-Yan (Tevatron)	*5.52* (**1.02**)	0.99	1.11
Collider Drell-Yan (LHC)	*18.91* (**1.31**)	*2.63* (**1.58**)	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

A ML open-source QCD fitting framework



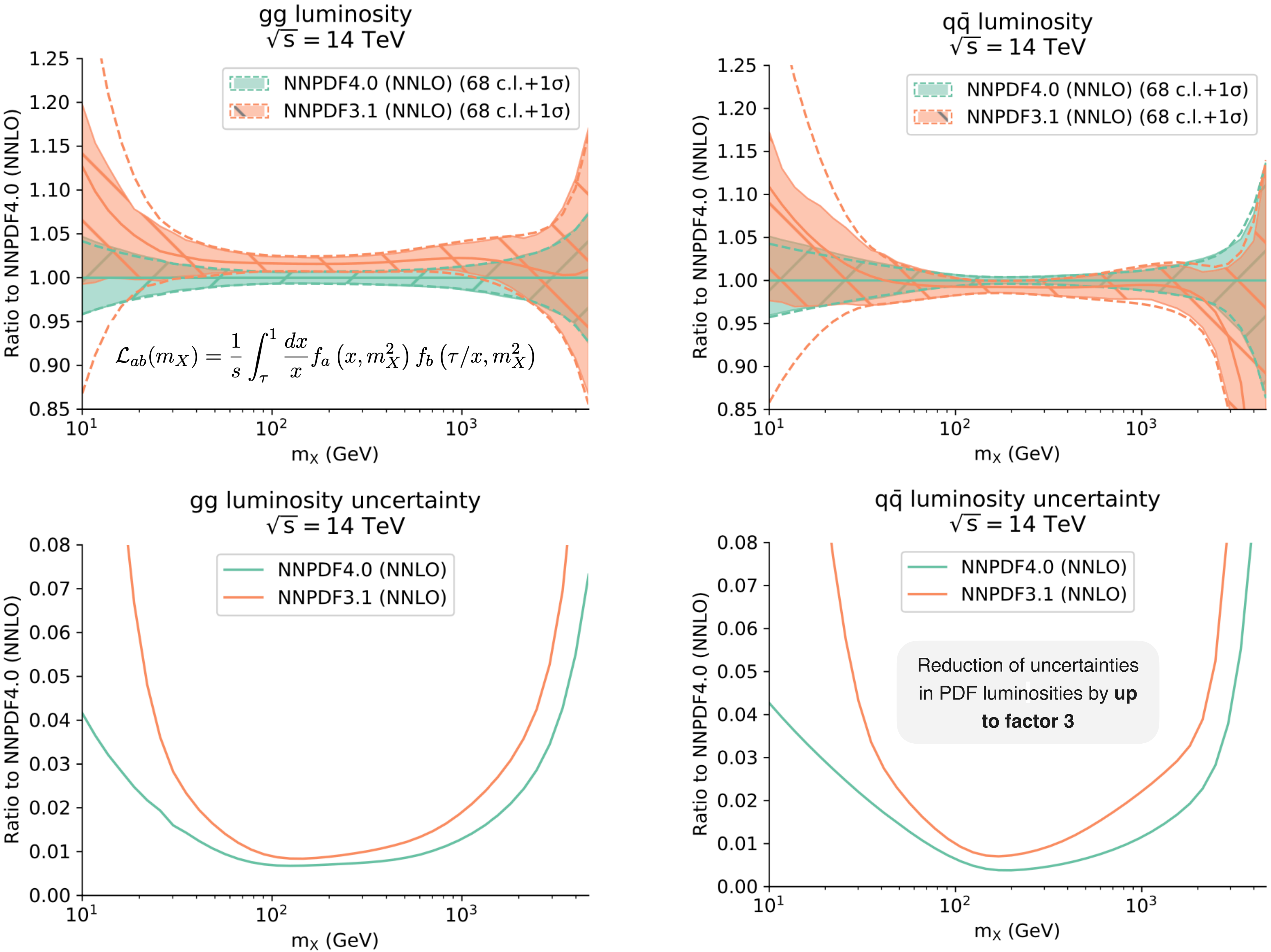
The **NNPDF machine learning fitting framework** has been publicly released open source, together with extensive documentation and user-friendly examples. Many opportunities for many studies within the **LHC experimental community,** looking forward to new collaborations!

Comparison with NNPDF3.1

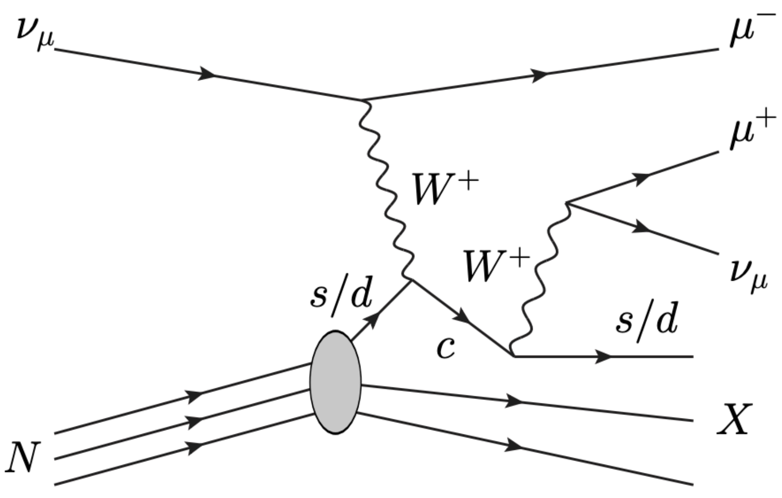


- **Good agreement** with NNPDF3.1 within uncertainties, with NNPDF4.0 being more precise
- Differences can be traced back to the i**mpact of specific datasets** (e.g. dijets for large-*x* gluon) or improvements in **theory calculations** (e.g. NNLO corrections in dimuon DIS for strangeness)

Comparison with NNPDF3.1

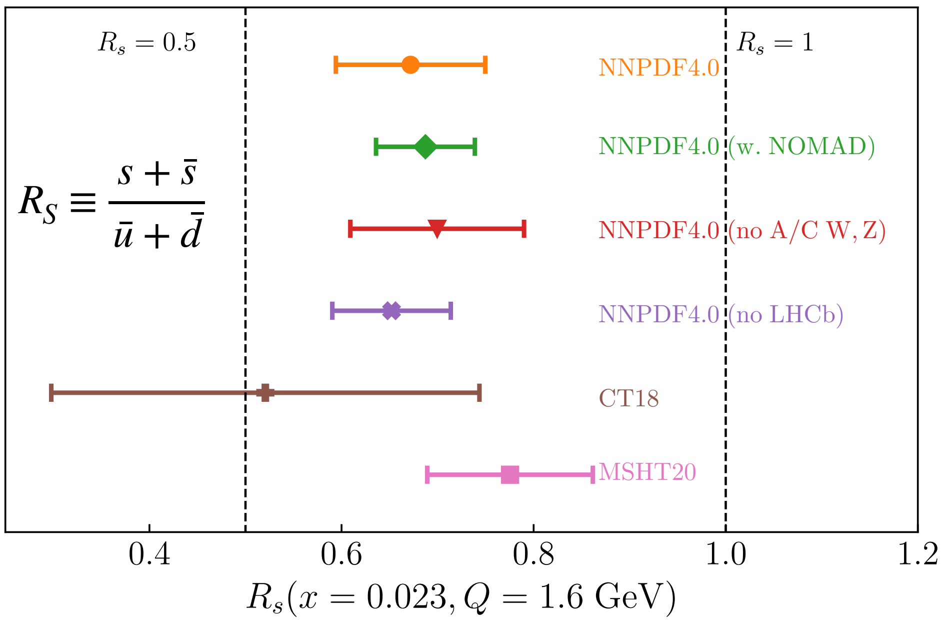


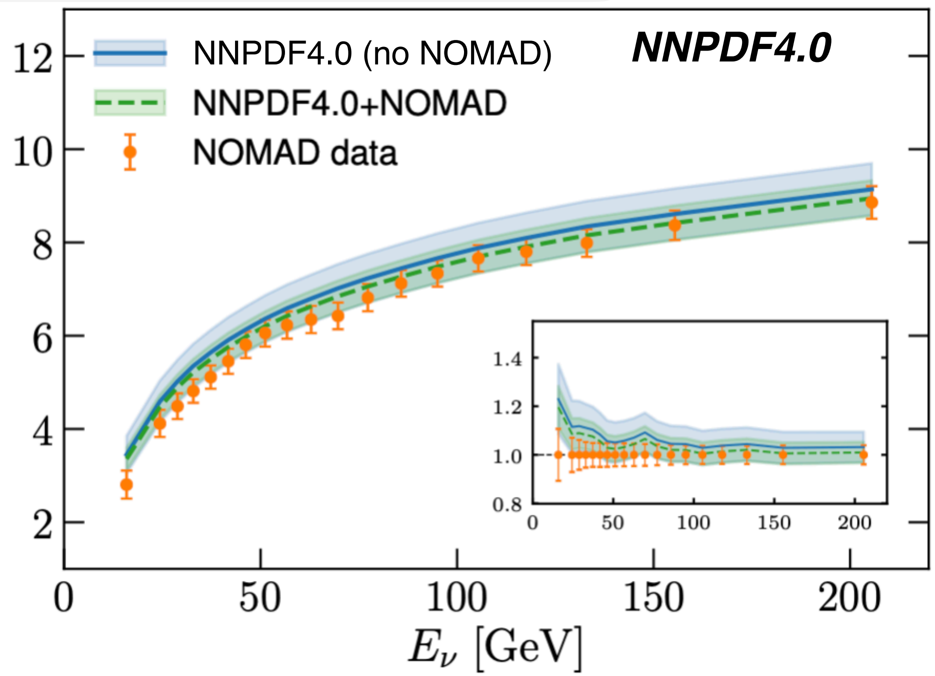
The strangest proton



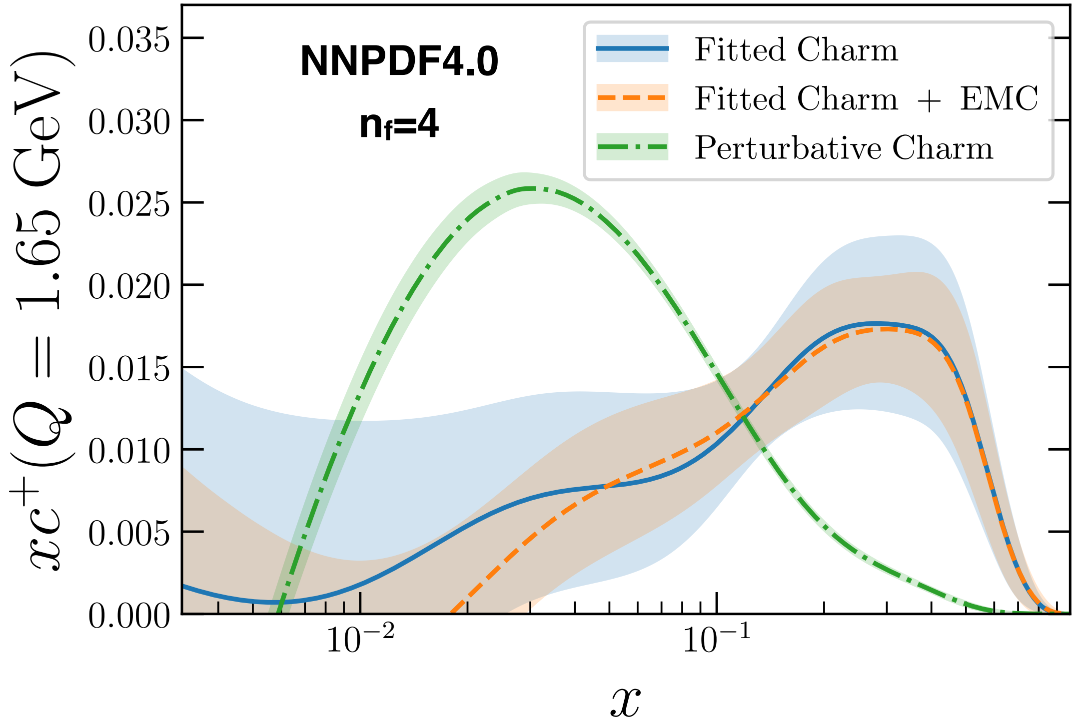
- Dimuon DIS data sensitive to **strangeness** via charged-current scattering
- NNPDF4.0 predicts the **NOMAD dimuon data,** which constraints stringently strangeness
- NNPDF4.0 strangeness stable when either **ATLAS/CMS** or **LHCb *W, Z*** data are removed
- **No tension** between LHC and DIS neutrino data!

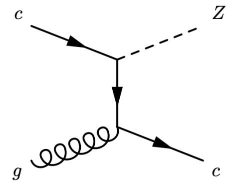
Excellent consistency of strangeness constraints from global dataset

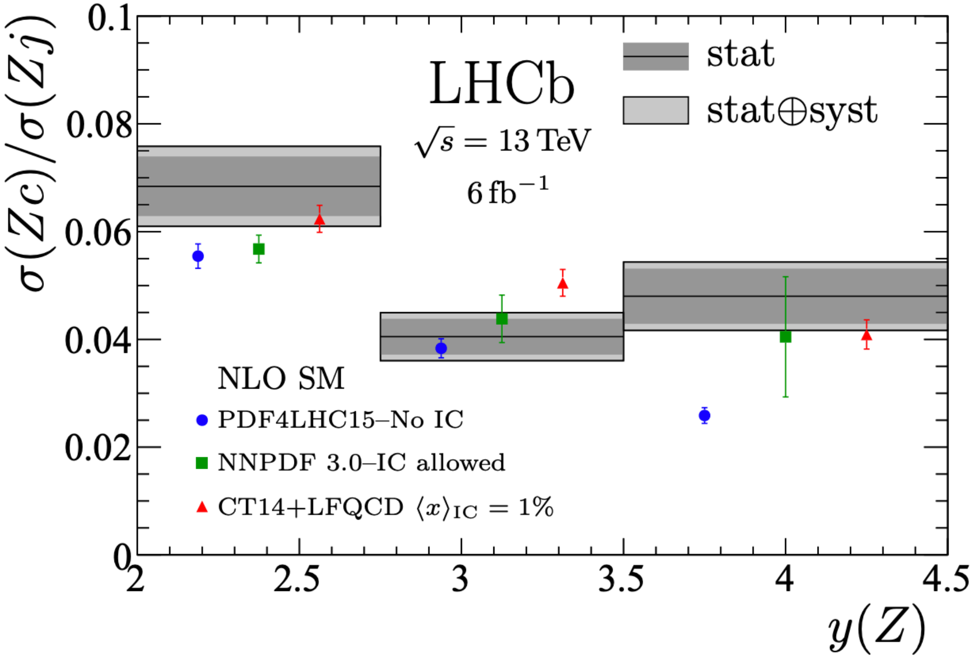




Intrinsic charm?







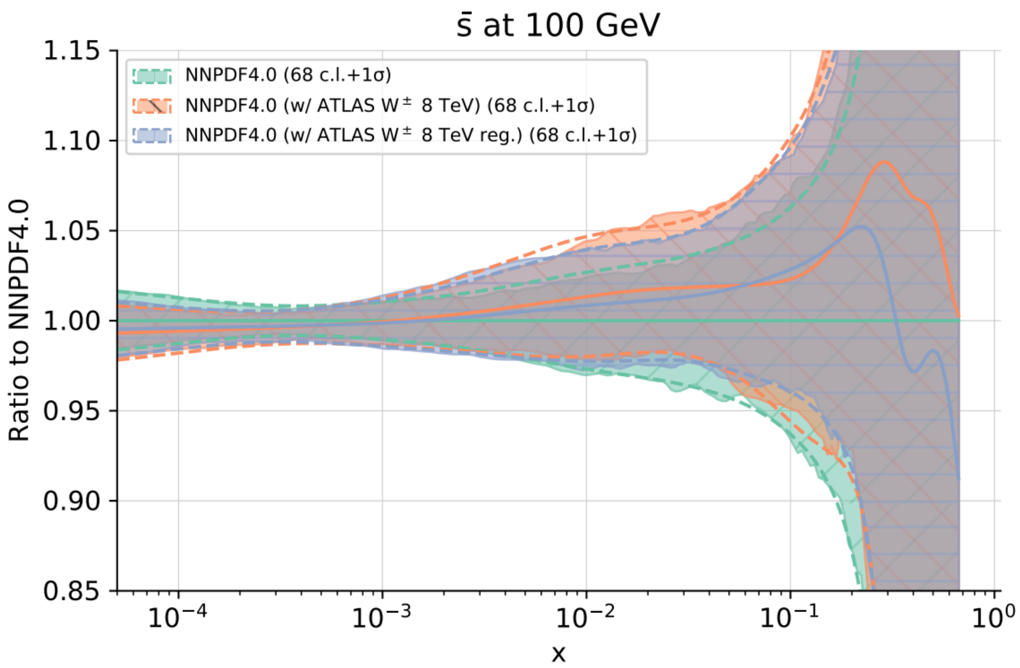
- Increasing evidence for **non-perturbative charm component** within the proton, robust upon conversion to the **3FNS** via backwards evolution and matching conditions

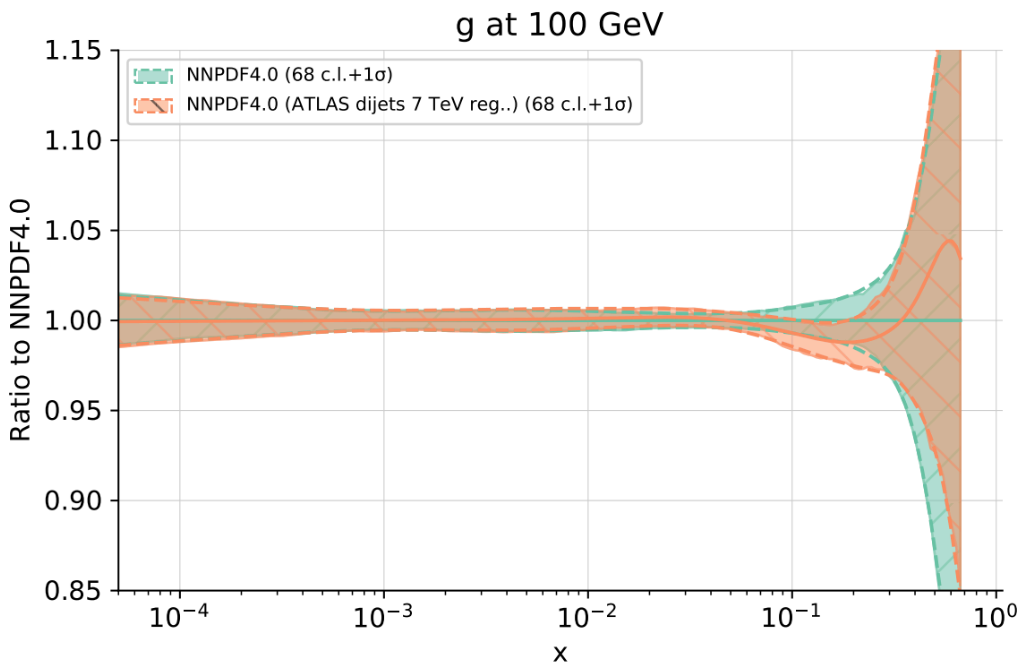
- Bulk of constraints provided by new **precision LHC data**, complemented by fixed-target DIS

- Consistent with recent LHCb measurement of **forward *Z+D* production**, directly sensitive to the (large-*x*) charm content of the nucleon

Towards evidence for intrinsic charm?

Impact of decorrelation models





- Certain datasets exhibit covmats where **small changes in correlations** lead to **large impact in χ^2**

- Assess impact in fit by transforming the original covmat into a matrix with the same eigenvectors but with **clipped eigenvalues** below some cut-off: **stable PDFs** with much lower χ^2

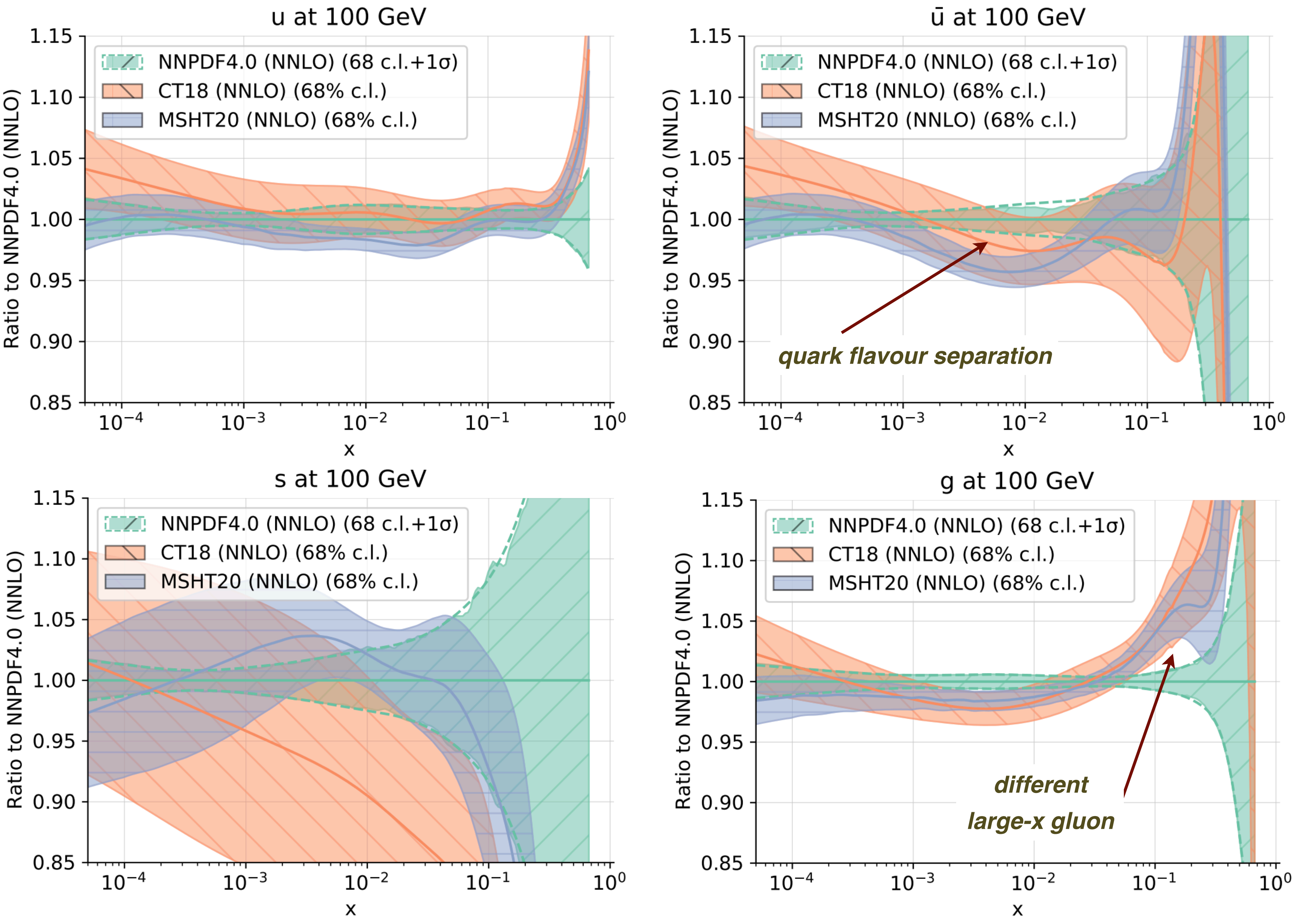
Dataset	$N_{\rm dat}$	$Z_{\rm orig}$	$\chi^2_{\rm orig}$	$\chi^2_{\rm reg}$
ATLAS W, Z 7 TeV CC ($\mathcal{L} = 4.6$ fb^{-1})	46	9.01	1.89	0.93
ATLAS W 8 TeV (*)	22	11.28	3.50	1.15
CMS dijets 7 TeV	54	4.70	1.81	1.73
ATLAS dijets 7 TeV	90	9.93	2.14	0.92
CMS 3D dijets 8 TeV (*)	122	4.47	1.50	0.92

Baseline NNPDF4.0 fits based on official correlation models provided by experiments!

Comparison with global fits and LHC pheno

Comparison between global fits

reasonable agreement with CT18 and MSHT20 (with some exceptions)

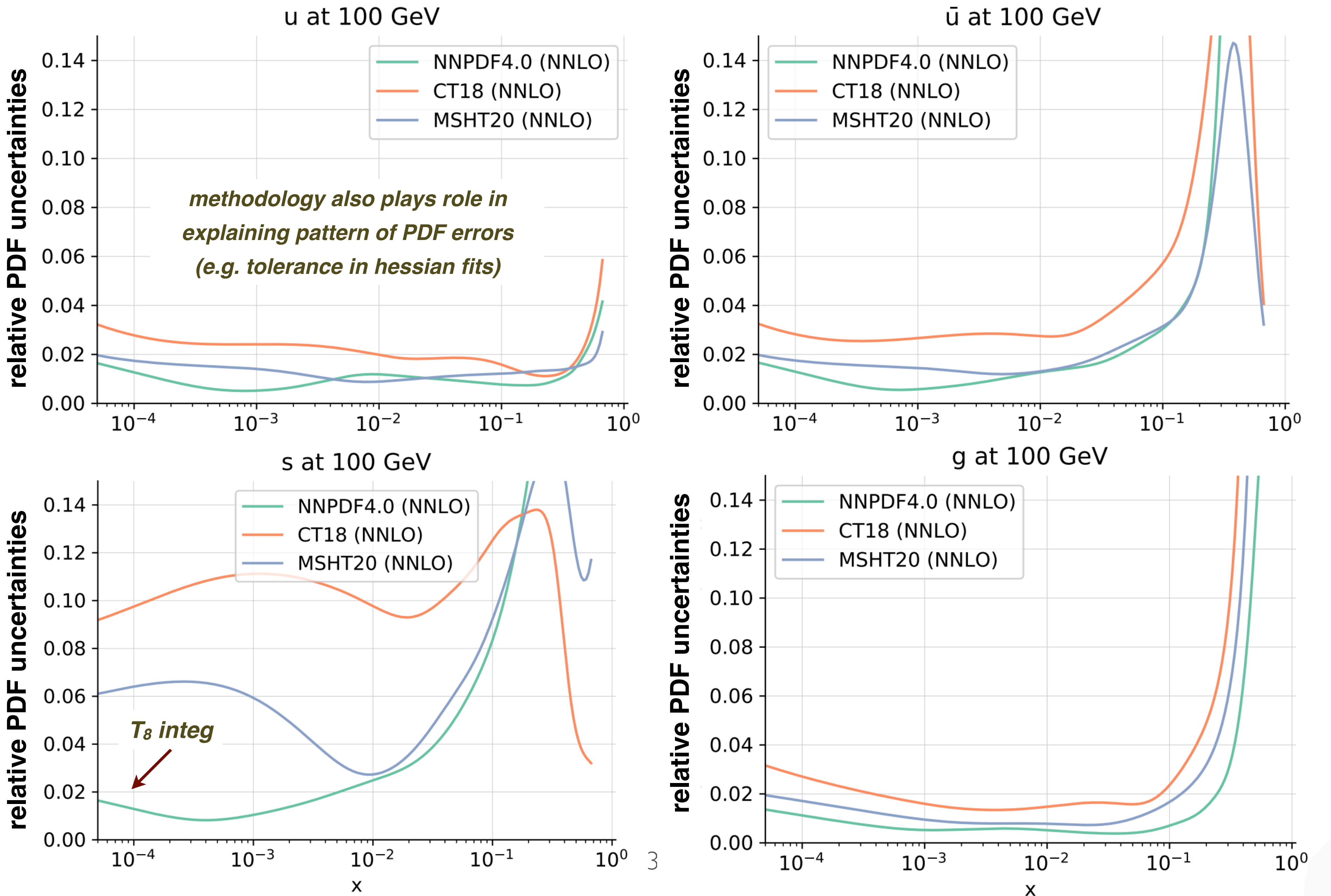


Comparison between global fits

pattern of PDF uncertainties follows (qualitatively) the **dataset size**

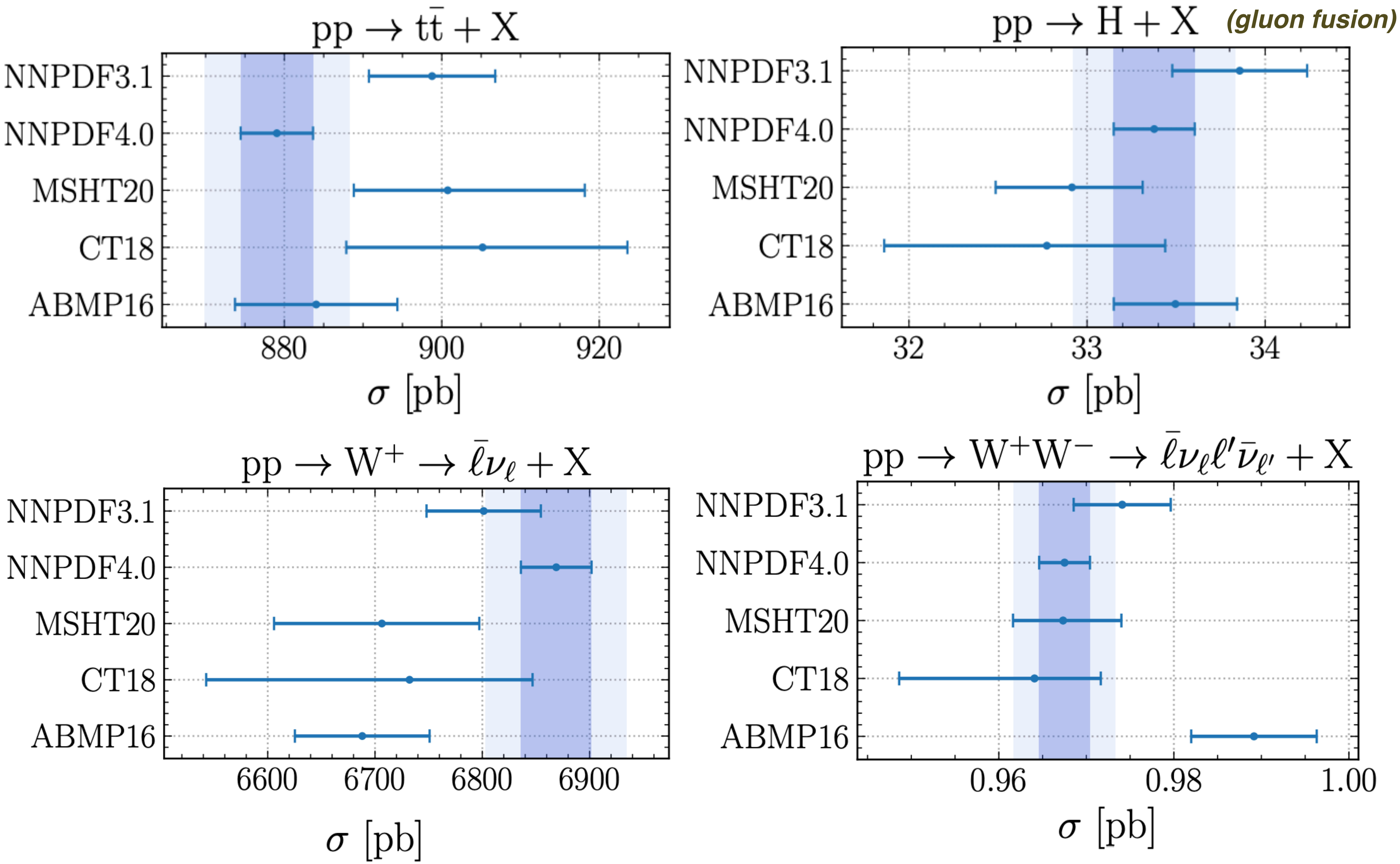
$$\delta_{\rm PDF}({\rm CT18}) \gtrsim \delta_{\rm PDF}({\rm MSHT20}) \gtrsim \delta_{\rm PDF}({\rm NNPDF4.0})$$

$$N_{\rm data}({\rm CT18}) \lesssim N_{\rm data}({\rm MSHT20}) \lesssim N_{\rm data}({\rm NNPDF4.0})$$



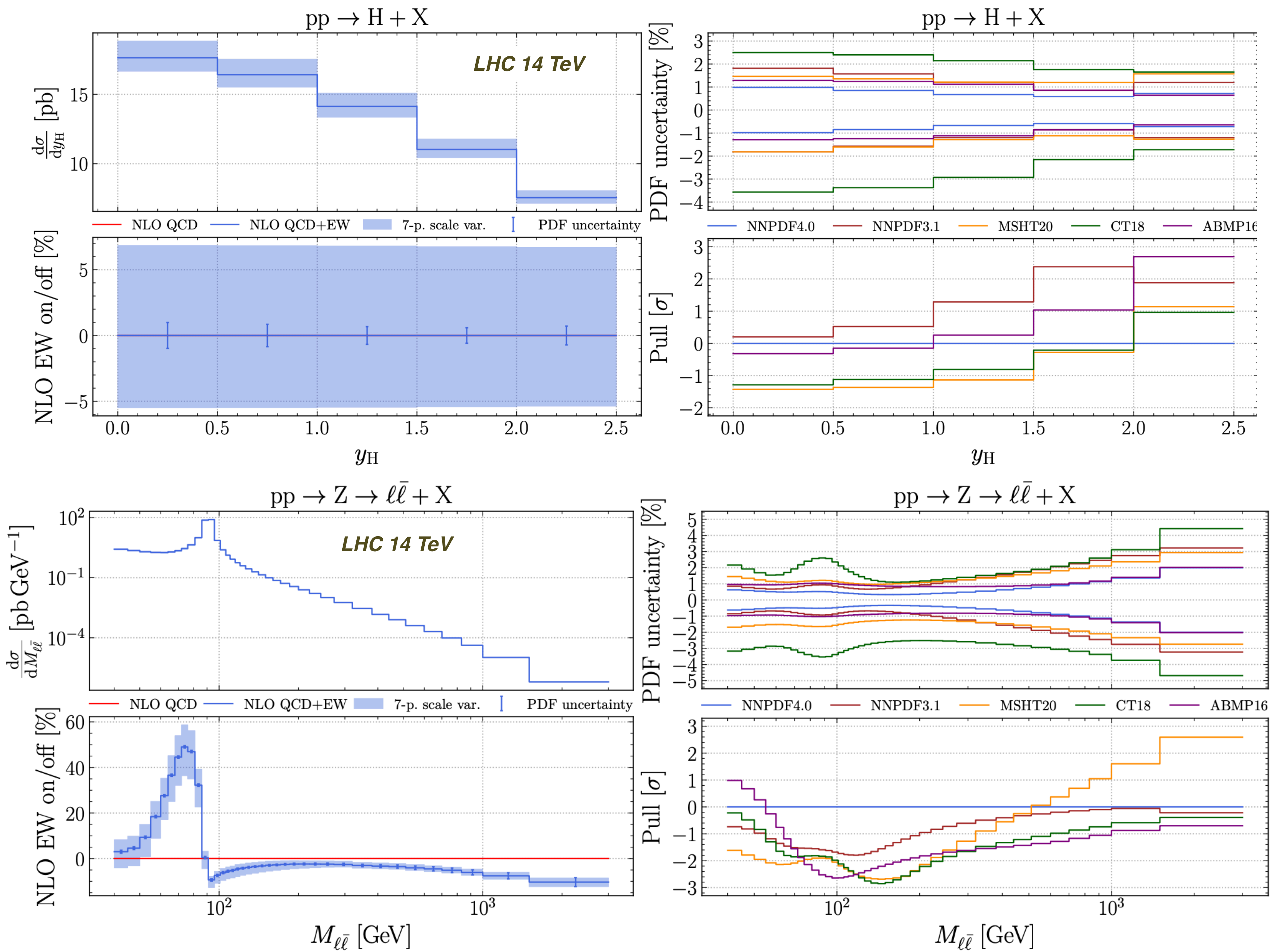
LHC phenomenology

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



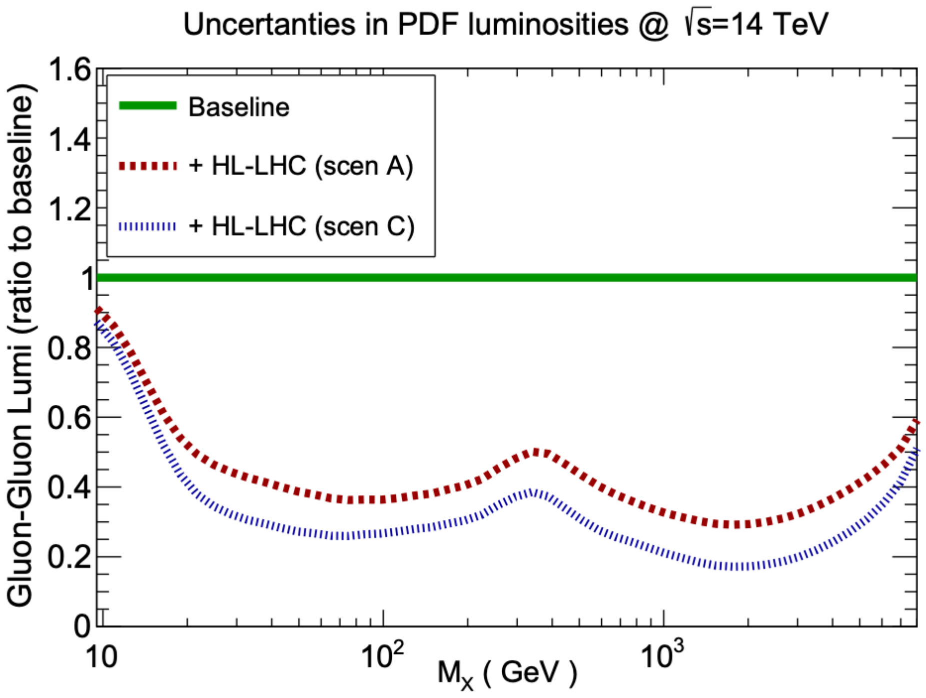
agreement of NNPDF4.0 with CT18 & MSHT20 **at two-sigma level**, differences traced back to large-*x* gluon or quark flavour separation

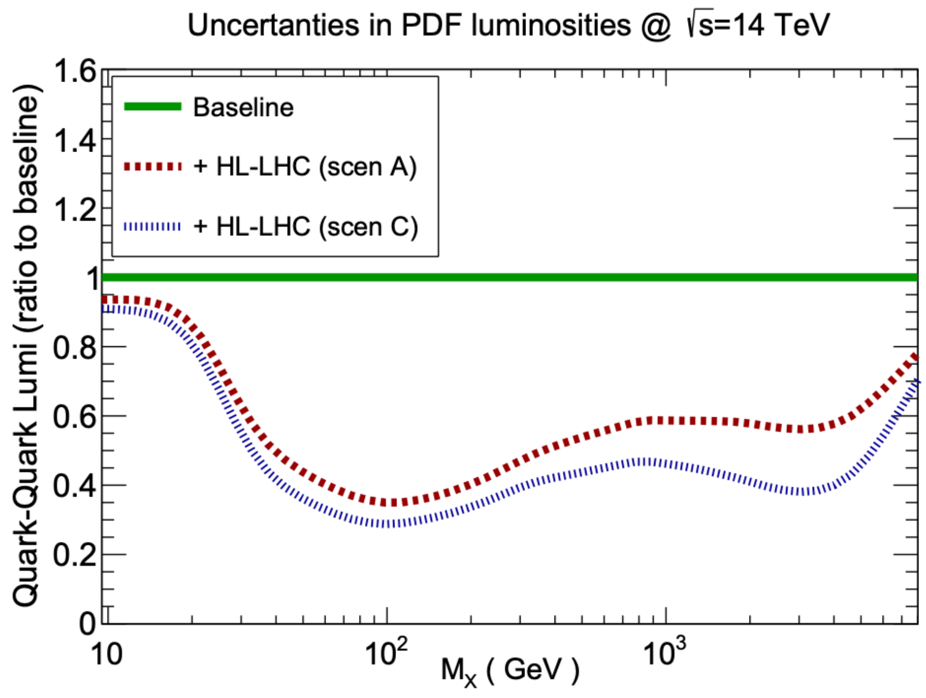
NLO QCD+EW cross-sections with NNLO PDFs

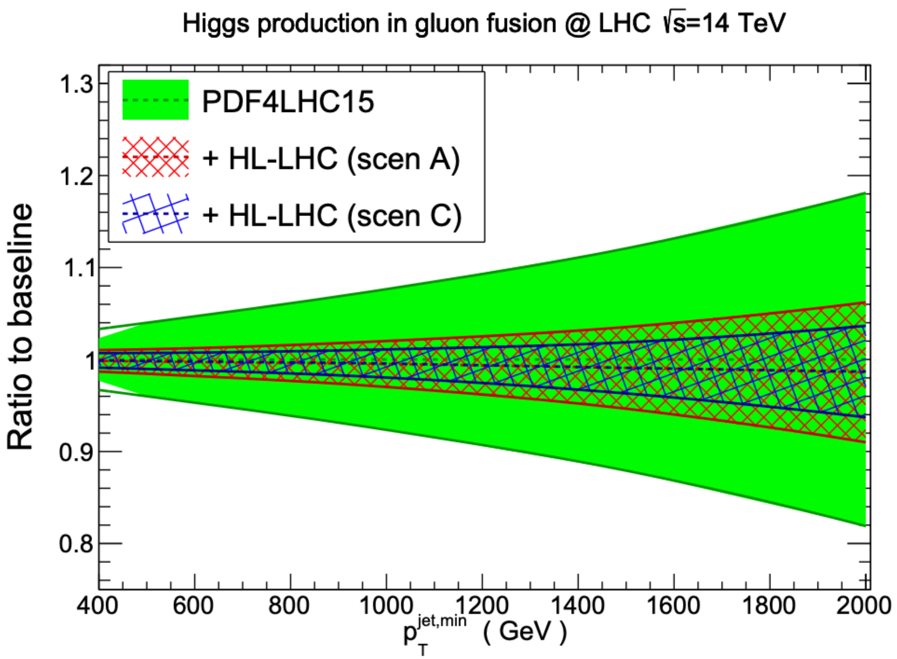

pp → H + X
LHC 14 TeV
$\frac{d\sigma}{dy_H}$ [pb]
NLO QCD
NLO QCD+EW
7-p. scale var.
PDF uncertainty
NLO EW on/off [%]
y_H
PDF uncertainty [%]
NNPDF4.0
NNPDF3.1
MSHT20
CT18
ABMP16
Pull [σ]
pp → Z → $\ell\bar{\ell}$ + X
$\frac{d\sigma}{dM_{\ell\bar{\ell}}}$ [pb GeV−1]
$M_{\ell\bar{\ell}}$ [GeV]

What to expect from Run III and beyond

Towards ultimate PDFs at the HL-LHC







- Dedicated projections for the **HL-LHC Yellow Reports** highlight plenty of room for PDF improvements from future LHC data
- These projections account only for ``standard" processes for PDF constraints (e.g. DY, top, jets): one could do better by thinking outside the box!
- Crucial to account for **theory uncertainties**, *e.g.* MHOUs, when interpreting future LHC measurements

What we need from the LHC @ Run III

How can we ensure to fully exploit the PDF constraining potential of Run III data?

- Focus on measurements that are now **limited by statistics:** large-p_T tails, rare processes, ….

- Focus on processes sensitive to PDF regions where **global PDF fits disagree**, inclusive jets & dijets & top quark to constrain the **large-*x* gluon**

- Release **full statistical models** (relevant e.g. for Poisson statistics or two-point systematics)

- Progress towards **unbinned measurements** (ideal binning/distribution evolves with time!)

- Assess **uncertainties in correlation models**, or provide a range of sensible correlation models

- Provide **correlations** between different distributions for the same process and for **different processes** (e.g. top and jets), and eventually between ATLAS & CMS

Most of these considerations apply to general interpretations of LHC data, eg. in the context of **EFT global fits**

What we need from the LHC @ Run III

How can we ensure to fully exploit the PDF constraining potential of Run III data?

Design **new tailored observables** (*eg unbinned*) with **complete statistical model** (*eg full likelihood*)

arXiv:2109.13243v1 [hep-ph] 27 Sep 2021

Presenting Unbinned Differential Cross Section Results

Miguel Arratia,[a,b] Anja Butter,[c] Mario Campanelli,[d] Vincent Croft,[e] Dag Gillberg,[f] Kristin Lohwasser,[g] Bogdan Malaescu,[h] Vinicius Mikuni,[i] Benjamin Nachman,[j,k] Juan Rojo,[l,m] Jesse Thaler,[n,o] Ramon Winterhalder[p]

[a] *Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
[b] *Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*
[c] *Institut für Theoretische Physik, Universität Heidelberg, Heidelberg, Germany*
[d] *University College London, London WC1E 6BT, UK*
[e] *Department of Physics and Astronomy, Tufts University, Boston, MA 02155, USA*
[f] *Department of Physics, Carleton University, Ottawa ON K1S 5B6, Canada*
[g] *University of Sheffield, Sheffield, S10 2TN, UK*
[h] *LPNHE, Sorbonne Université, Université de Paris, CNRS/IN2P3, Paris, France*
[i] *National Energy Research Scientific Computing Center, Berkeley, CA 94720, USA*
[j] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[k] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[l] *Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[m] *Department of Physics and Astronomy, Vrije Universiteit Amsterdam, NL-1081 HV Amsterdam, The Netherlands*
[n] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[o] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[p] *Centre for Cosmology, Particle Physics and Phenomenology (CP3), Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium*

E-mail: bpnachman@lbl.gov

Abstract: Machine learning tools have empowered a qualitatively new way to perform differential cross section measurements whereby the data are unbinned, possibly in many dimensions. Unbinned measurements can enable, improve, or at least simplify comparisons between experiments and with theoretical predictions. Furthermore, many-dimensional measurements can be used to define observables after the measurement instead of before. There is currently no community standard for publishing unbinned data. While there are

SciPost Physics Submission

arXiv:2109.04981v1 [hep-ph] 10 Sep 2021

Publishing statistical models: Getting the most out of particle physics experiments

Kyle Cranmer[1*], Sabine Kraml[2‡], Harrison B. Prosper[3§] (editors), Philip Bechtle[4], Florian U. Bernlochner[4], Itay M. Bloch[5], Enzo Canonero[6], Marcin Chrzaszcz[7], Andrea Coccaro[8], Jan Conrad[9], Glen Cowan[10], Matthew Feickert[11], Nahuel Ferreiro Iachellini[12,13] Andrew Fowlie[14], Lukas Heinrich[15], Alexander Held[1], Thomas Kuhr[13,16], Anders Kvellestad[17], Maeve Madigan[18], Farvah Mahmoudi[15,19], Knut Dundas Morå[20], Mark S. Neubauer[11], Maurizio Pierini[15], Juan Rojo[8], Sezen Sekmen[22], Luca Silvestrini[23], Veronica Sanz[24,25], Giordon Stark[26], Riccardo Torre[8], Robert Thorne[27], Wolfgang Waltenberger[28], Nicholas Wardle[29], Jonas Wittbrodt[30]

1 New York University, USA **2** LPSC Grenoble, France **3** Florida State University, USA **4** University of Bonn, Germany **5** School of Physics and Astronomy, Tel-Aviv University, Israel **6** University of Genova, Italy **7** Institute of Nuclear Physics, Polish Academy of Sciences, Krakow, Poland **8** INFN, Sezione di Genova, Italy **9** Oskar Klein Centre, Stockholm University, Sweden **10** Royal Holloway, University of London, UK **11** University of Illinois at Urbana-Champaign, USA **12** Max Planck Institute for Physics, Munich, Germany **13** Exzellenzcluster ORIGINS, Garching, Germany **14** Nanjing Normal University, Nanjing, PRC **15** CERN, Switzerland **16** Ludwig-Maximilians-Universität München, Germany **17** University of Oslo, Norway **18** DAMTP, University of Cambridge, UK **19** Lyon University, France **20** Columbia University 10027, USA **21** VU Amsterdam and Nikhef, The Netherlands **22** Kyungpook National University, Daegu, Korea **23** INFN, Sezione di Roma, Italy **24** University of Sussex, UK **25** IFIC, Universidad de Valencia-CSIC, Spain **26** SCIPP, UC Santa Cruz, CA, USA **27** University College London, UK **28** HEPHY and University of Vienna, Austria **29** Imperial College London, UK **30** Lund University, Sweden

* kyle.cranmer@nyu.edu, ‡ sabine.kraml@lpsc.in2p3.fr, § harry@hep.fsu.edu
(mailing list: open-likelihoods@cern.ch)

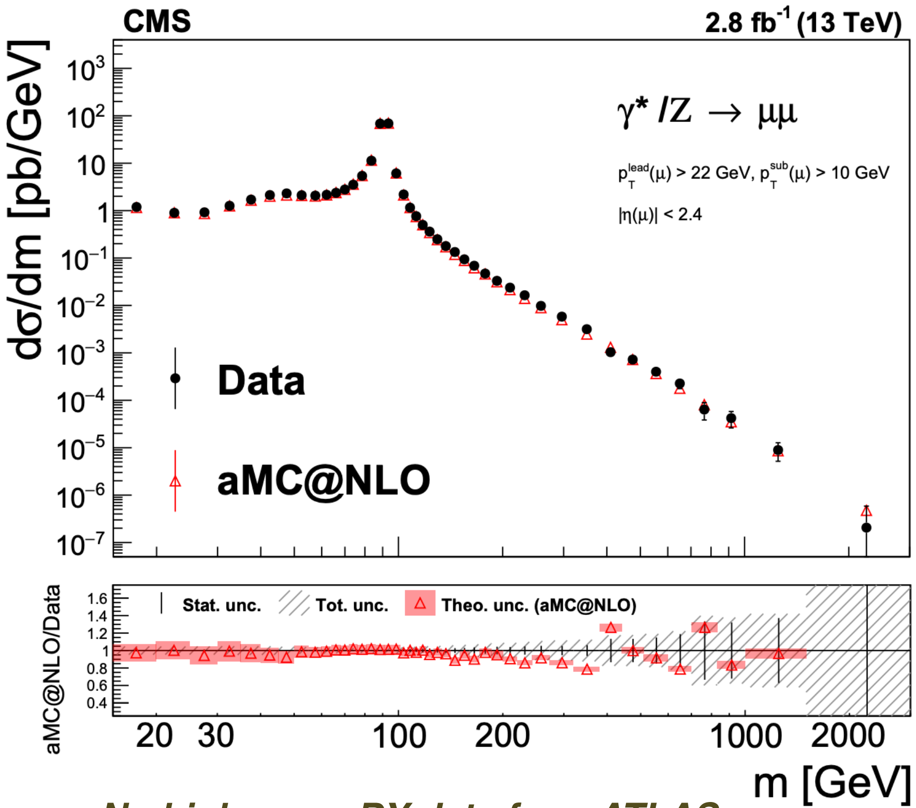
September 9, 2021

Abstract

What we need from the LHC @ Run III

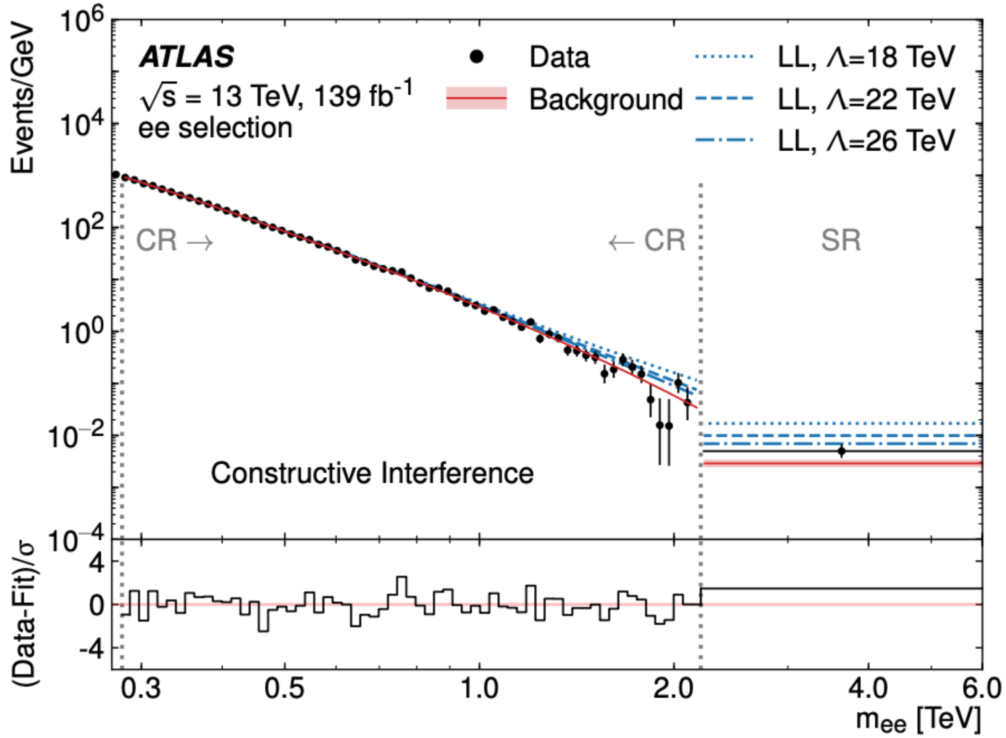
How can we ensure to fully exploit the PDF constraining potential of Run III data?

Most updated SM measurement



No high-mass DY data from ATLAS

One of the recent BSM searches



ATLAS non-resonant dilepton search

SM measurements often lag behind **searches**, but on the longer term their impact is bigger.

Furthermore, ``measurements" are also crucial for BSM searches *e.g.* **with EFT interpretations**

Prioritise SM measurements!

Summary and outlook

- Recent progress in PDF fits enables the **precision physics program of the LHC** as well as makes possible unravelling open questions in Quantum Chromodynamics *e.g.* antimatter and heavy quark content of the nucleon

- 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.** The full NNPDF software framework is now **open source** and welcoming contributions!

- Pushing forward the PDF precision frontier requires **N3LO PDFs** and **PDFs with electroweak corrections and photon PDF:** WIP

- 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:** missing higher orders, nuclear corrections, SM parameters, …

- From experimental side, progress in PDFs requires focusing on key measurements where **impact on PDFs** is expected to be the largest (*e.g.* related to large-*x* gluon), as well as dedicated efforts towards **tailored measurements** with **statistical models available**

Summary and outlook

- Recent progress in PDF fits enables the **precision physics program of the LHC** as well as makes possible unravelling open questions in Quantum Chromodynamics *e.g.* antimatter and heavy quark content of the nucleon

- The global NNPDF4.0 fit achieves **high accuracy** in an unprecedentedl… …ange, thanks so its **extensive dataset** combined with **deep-learnin…** …ll NNPDF software framework is now **open source** an…

- Pushing forward the PDF precisi… **…Fs with electroweak corrections and phot…**

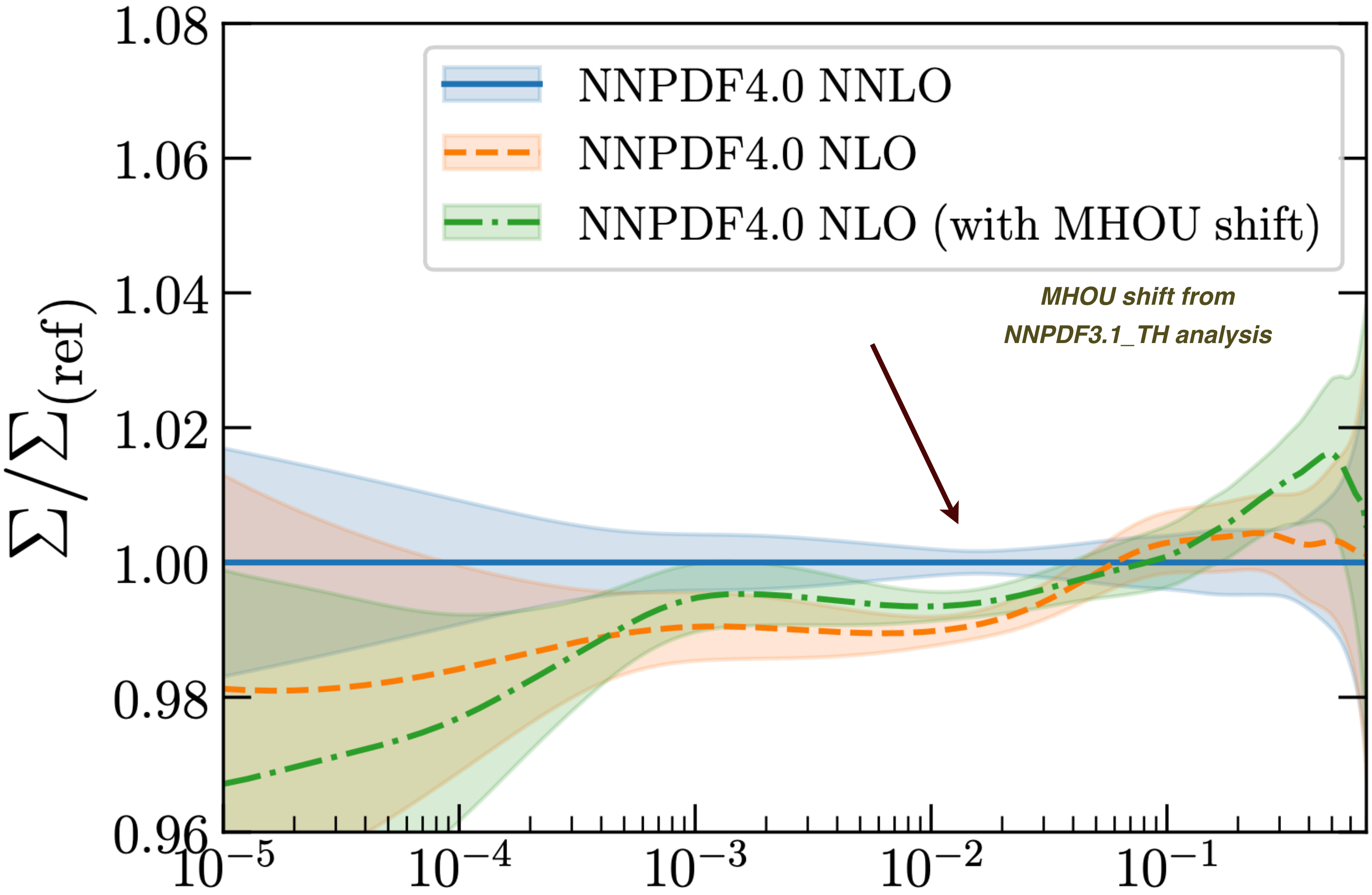
- The… …allenges the accuracy of **theoretical predictions** and dema… …ards the systematic inclusion in the fit of **theoretical uncertainties:** missin… …, nuclear corrections, SM parameters, …

Thanks for your attention!

- From experimental side, progress in PDFs requires focusing on key measurements where **impact on PDFs** is expected to be the largest (*e.g.* related to large-x gluon), as well as dedicated efforts towards **tailored measurements** with **statistical models available**

Extra Material

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!

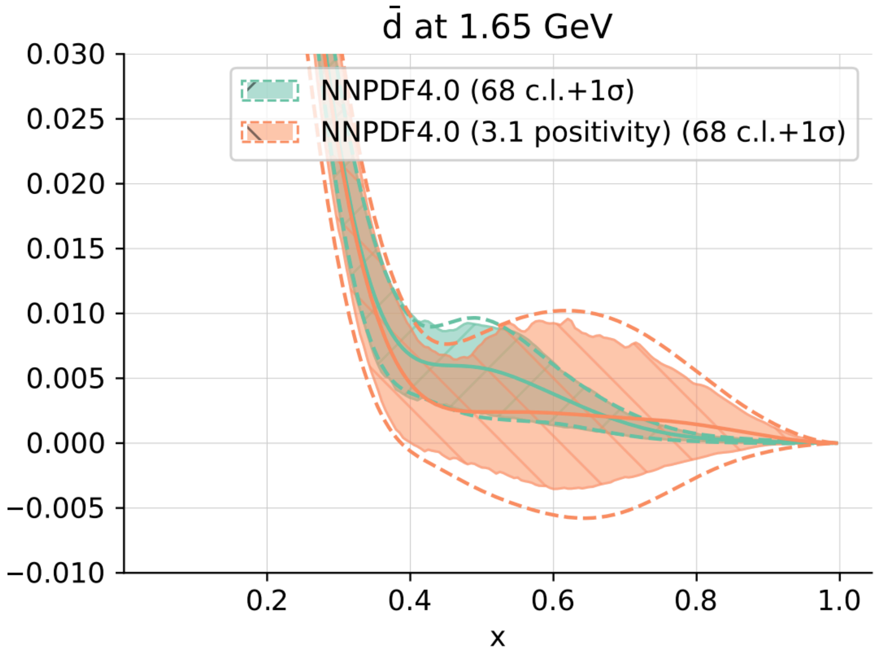
The path to NNPDF4.0

Collaborative progress towards extending data, theory and methodology

06/2017	**NNPDF3.1**	[EPJ C77 (2017) 663]
10/2017	NNPDF3.1sx: PDFs with small-x resummation	[EPJ C78 (2018) 321]
12/2017	NNPDF3.1luxQED: consistent photon PDF à la luxQED	[SciPost Phys. 5 (2018) 008]
02/2018	NNPDF3.1+ATLASphoton: inclusion of direct photon data	[EPJ C78 (2018) 470]
12/2018	NNPDF3.1alphas: α_s from a correlated-replica method	[EPJ C78 (2018) 408]
12/2018	NNPDF3.1nuc: heavy ion nuclear uncertainties in a fit	[EPJ C79 (2019) 282]
05/2019	NNPDF3.1th: missing higher-order uncertainties in a fit	[EPJ C79 (2019) 838; ibid. 931]
07/2019	Gradient descent and hyperoptimisation in PDF fits	[EPJ C79 (2019) 676]
12/2019	NNPDF3.1singletop: inclusion of single top t-channel data	[JHEP 05 (2020) 067]
05/2020	NNPDF3.1dijets: comparative study of single- and di-jets	[EPJ C80 (2020) 797]
06/2020	Positivity of $\overline{\mathrm{MS}}$ PDFs	[JHEP 11 (2020) 129]
08/2020	PineAPPL: fast evaluation of EW$\times$QCD corrections	[JHEP 12 (2020) 108]
08/2020	NNPDF3.1strangeness: assessment of strange-sensitive data	[EPJ C80 (2020) 1168]
11/2020	NNPDF3.1deu: deuteron uncertainties in a fit	[EPJ C81 (2021) 37]
03/2021	Future tests	[arXiv:2103.08606]
2021	**NNPDF4.0**	**September 2021**

Culmination of extensive efforts from the last four years!

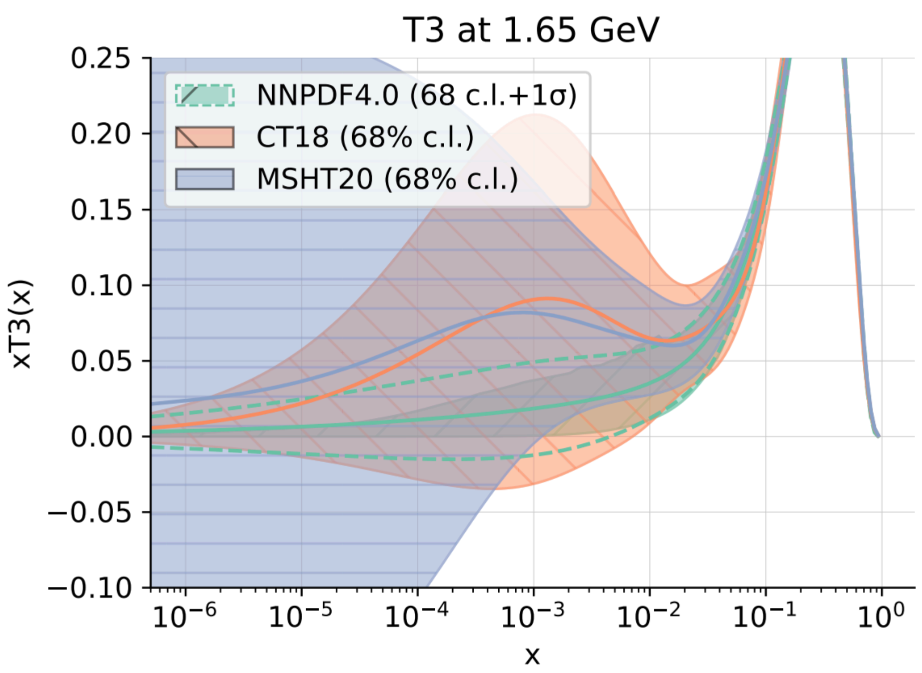
Positivity and integrability

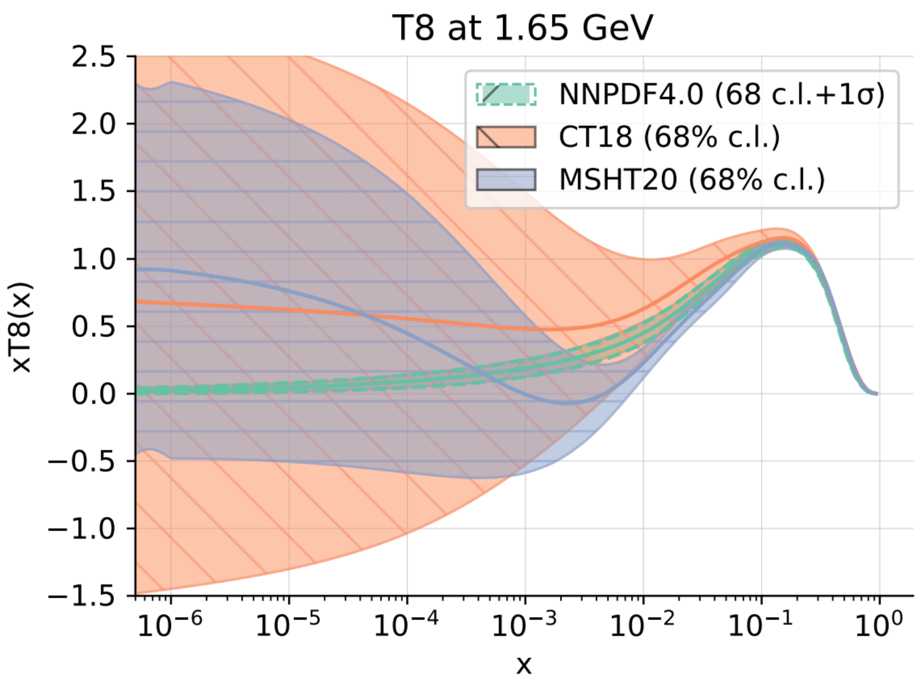


- MSbar PDFs have been shown to satisfy **positivity** requirements at all orders: **reduce large-*x* uncertainties**

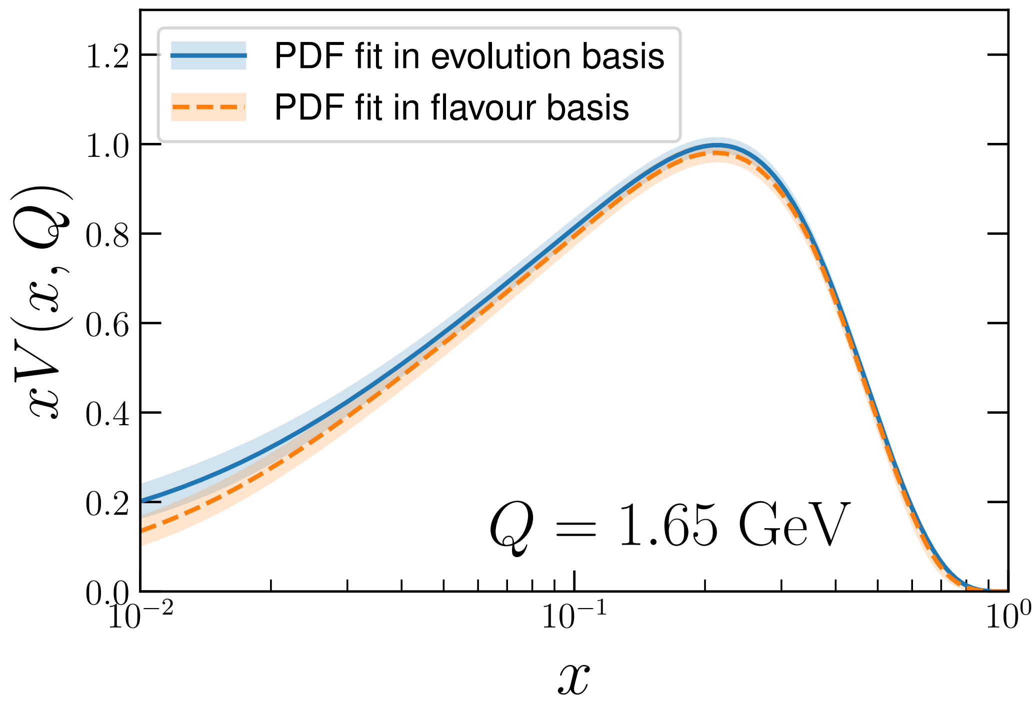
- The non-singlet quark triplet and octet should be *integrable* (e.g. Gottfried sum rule): **reduce small-*x* uncertainties**

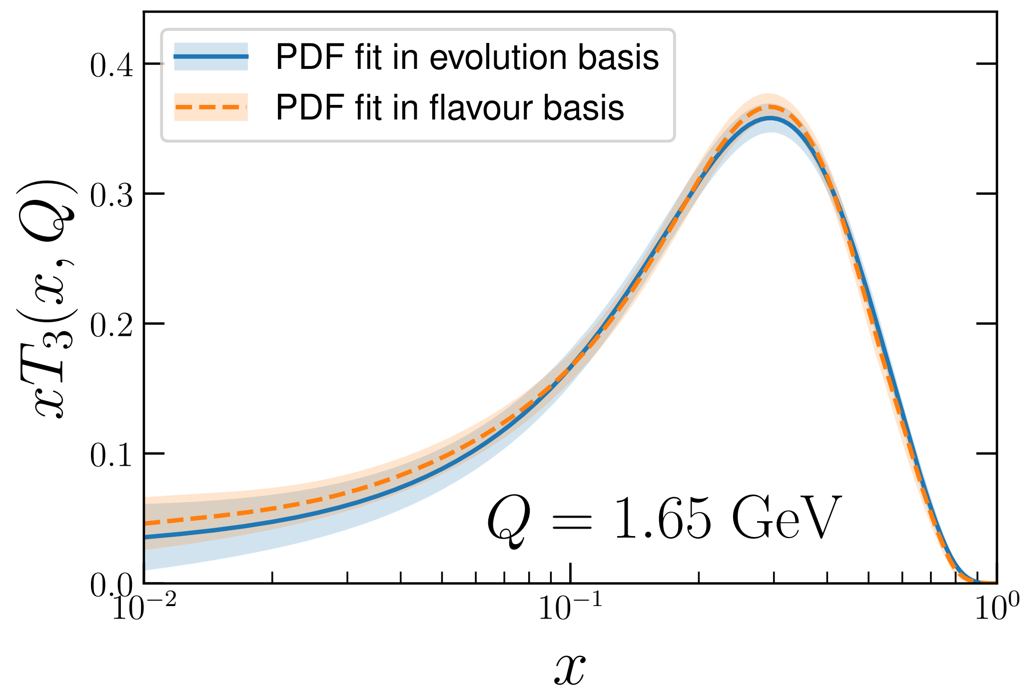
$$T_8 = (u + \bar{u}) + \left(d + \bar{d}\right) - 2\,(s + \bar{s})$$





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

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

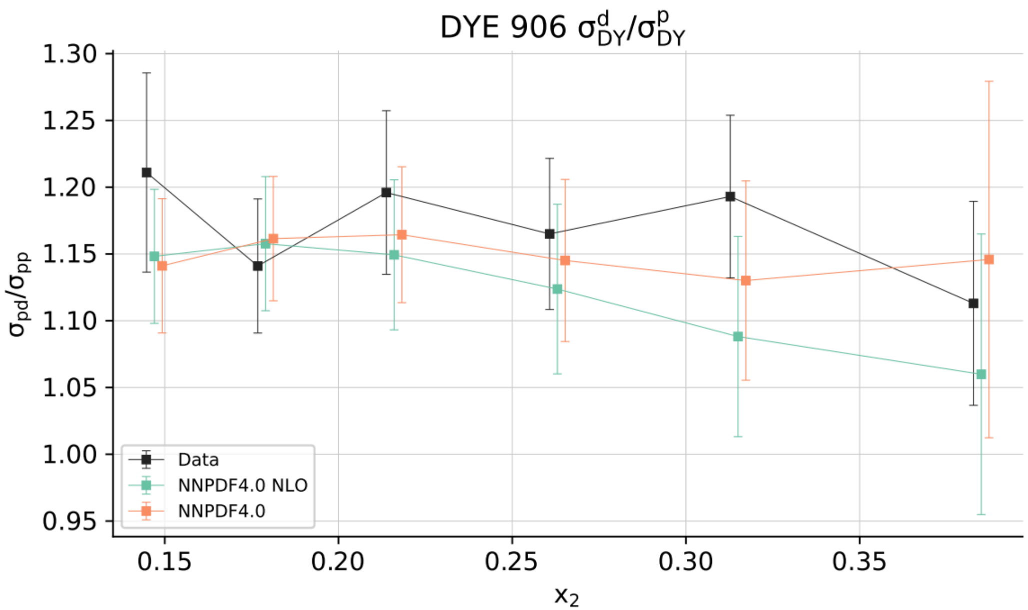
The scientific output of the collaboration is freely available to the publi through the arXiv, journal repositories, and software repositories. Along with this online documentation, we release the NNPDF code used to produce the latest family of PDFs from NNPDF, NNPDF4.0. The code is made available as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

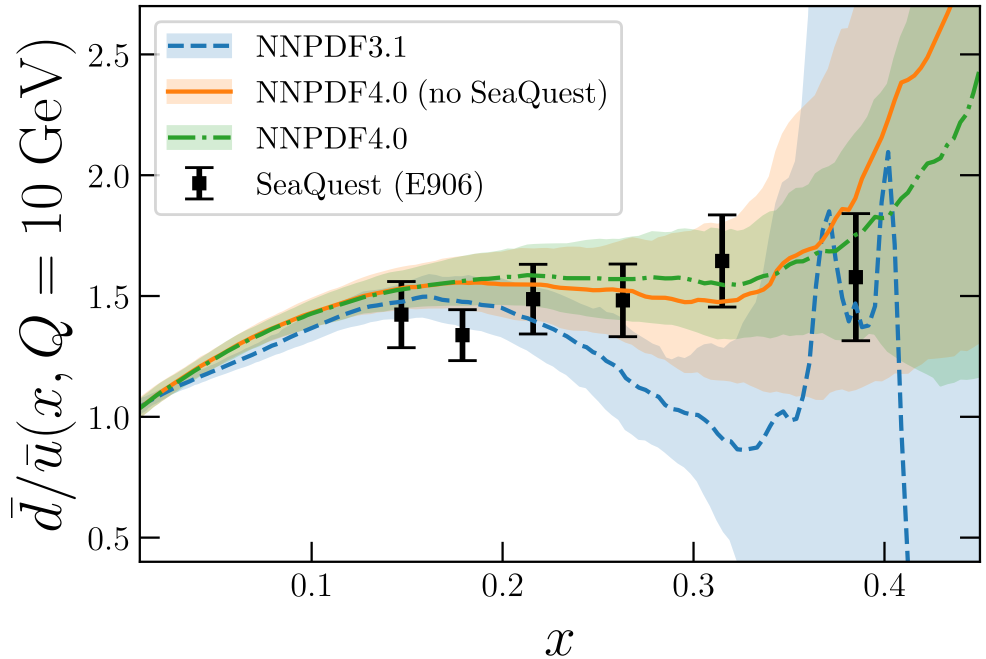
The code can be used to produce the ingredients needed for PDF fits, to run the fits themselves, and to analyse the results. This is the first framework used to produce a global PDF fit made publicly available, enabling for a detailed external validation and reproducibility of the NNPDF4.0 analysis. 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 experimental community**: looking forward to suggestions and starting new collaborations!

Antimatter asymmetry





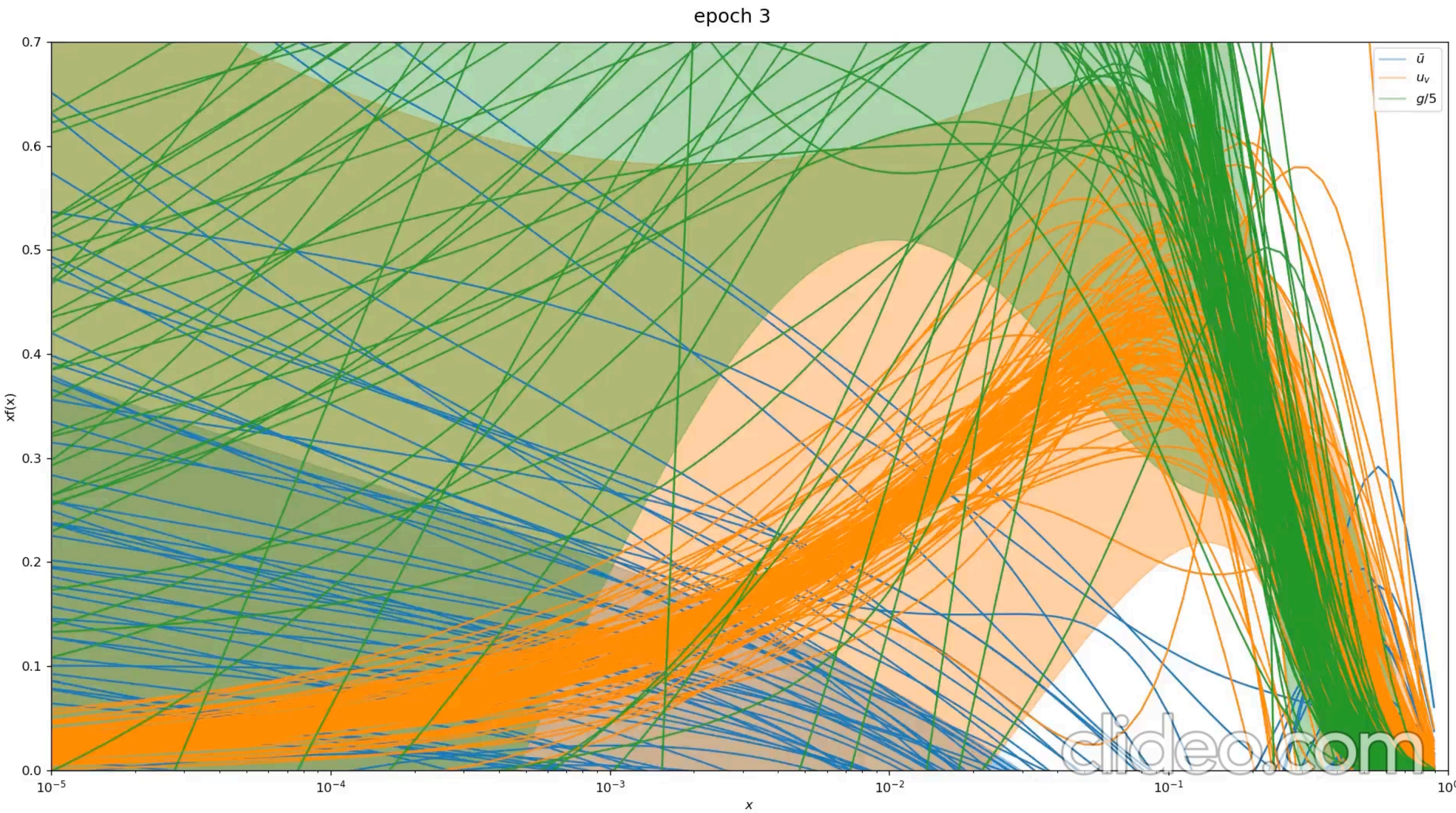
- ☑ The recent SeaQuest measurement claims evidence for **quark sea (``proton antimatter") asymmetry**

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

with many caveats!

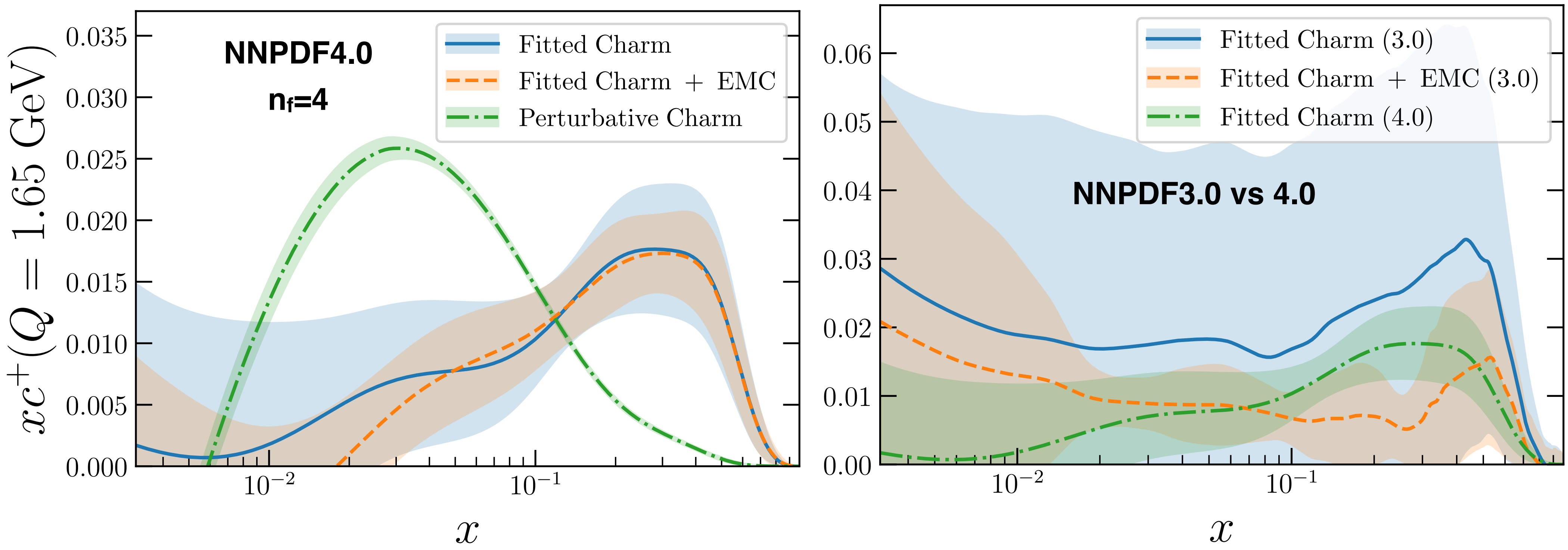
- ☑ Actually, SeaQuest further confirms the **global fit prediction**, which agrees with it even when not included
- ☑ Already well described by **NNPDF3.1** within uncertainties

Improved fitting methodology



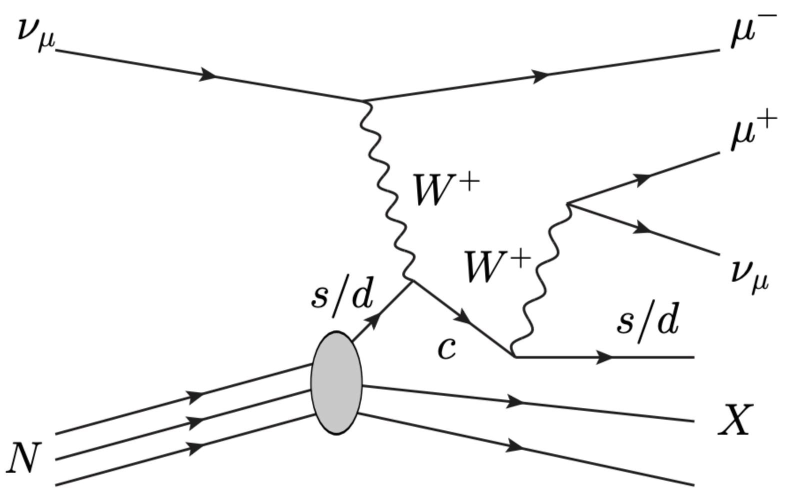
Illustrating the outcome of **SGD minimisation** (band: standard deviation over the MC replicas)

Intrinsic charm



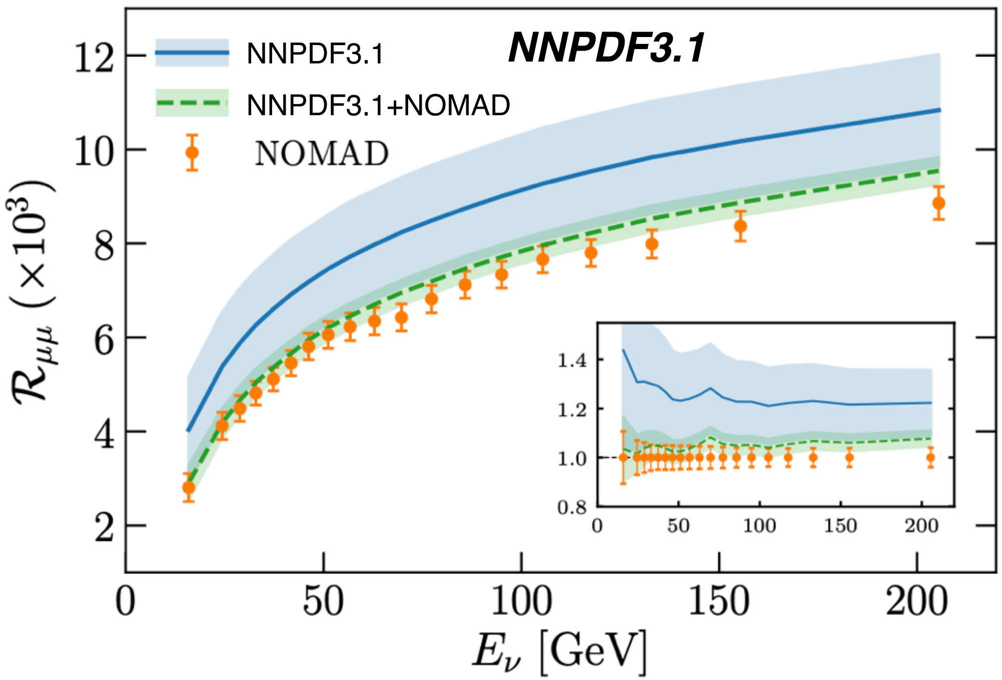
- Increasing evidence for **non-perturbative charm component** within the proton, robust upon conversion to the **3FNS** via backwards evolution and matching conditions (WIP)
- Bulk of constraints provided by new **precision LHC data**, complemented by fixed-target DIS
- As opposed to previous studies,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

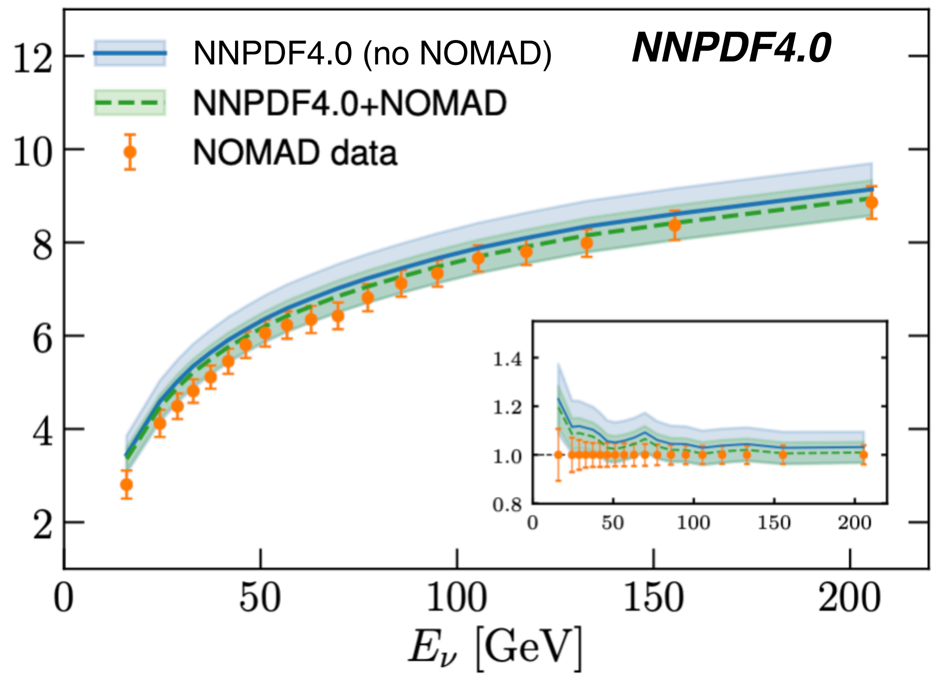
The strangest proton



- ☑ 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