The path to proton structure at 1% accuracy

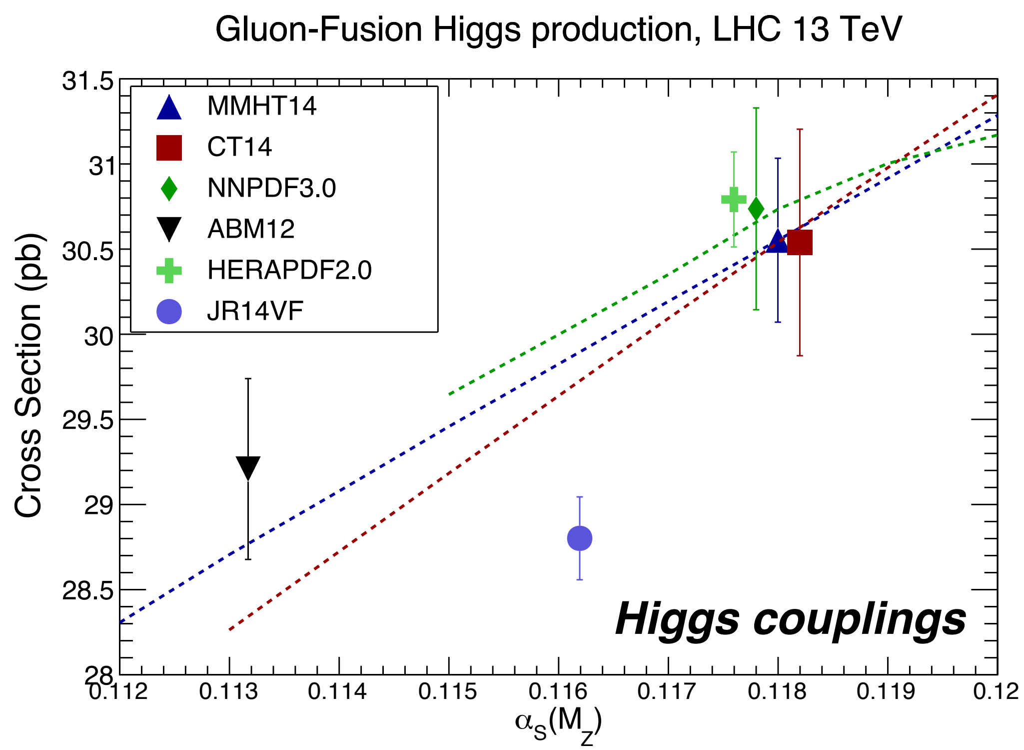
Juan Rojo, VU Amsterdam &, Nikhef

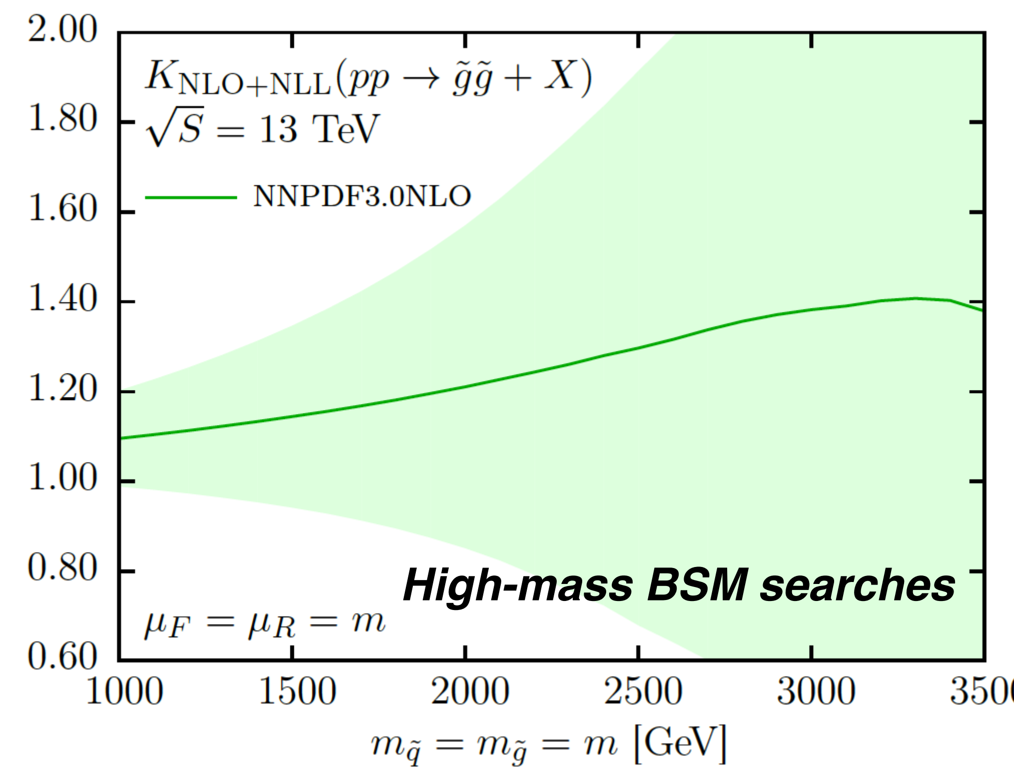
Standard Model Physics at the TeV scale

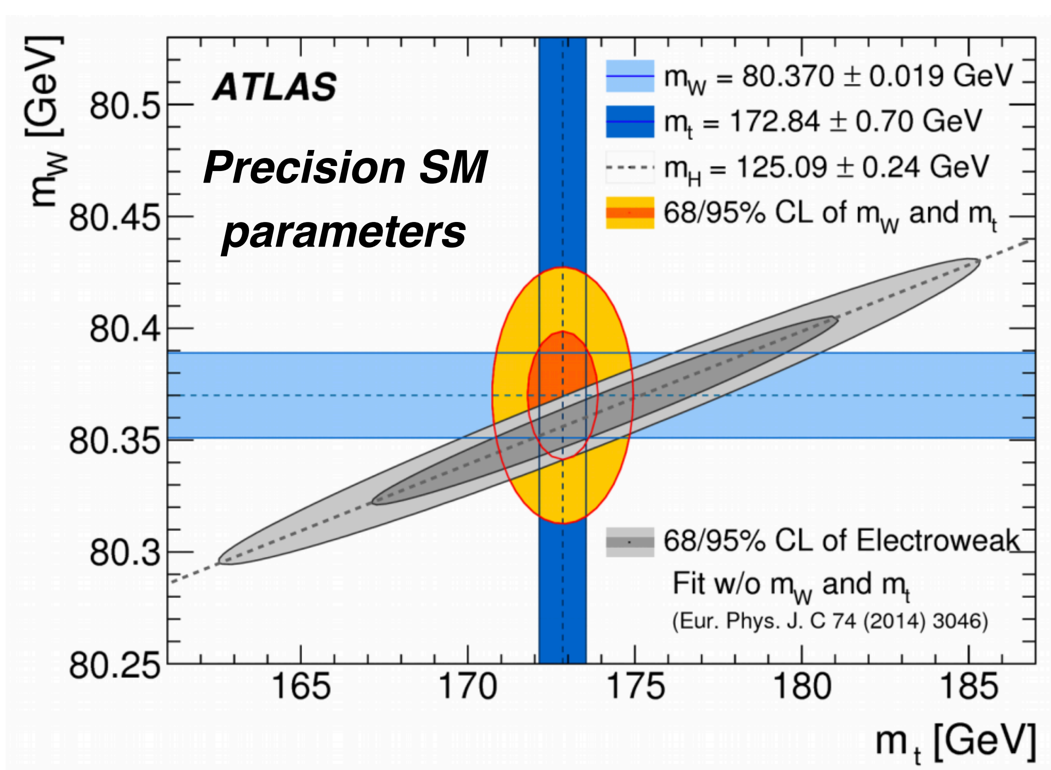
PANIC 2021, 5th September 2021

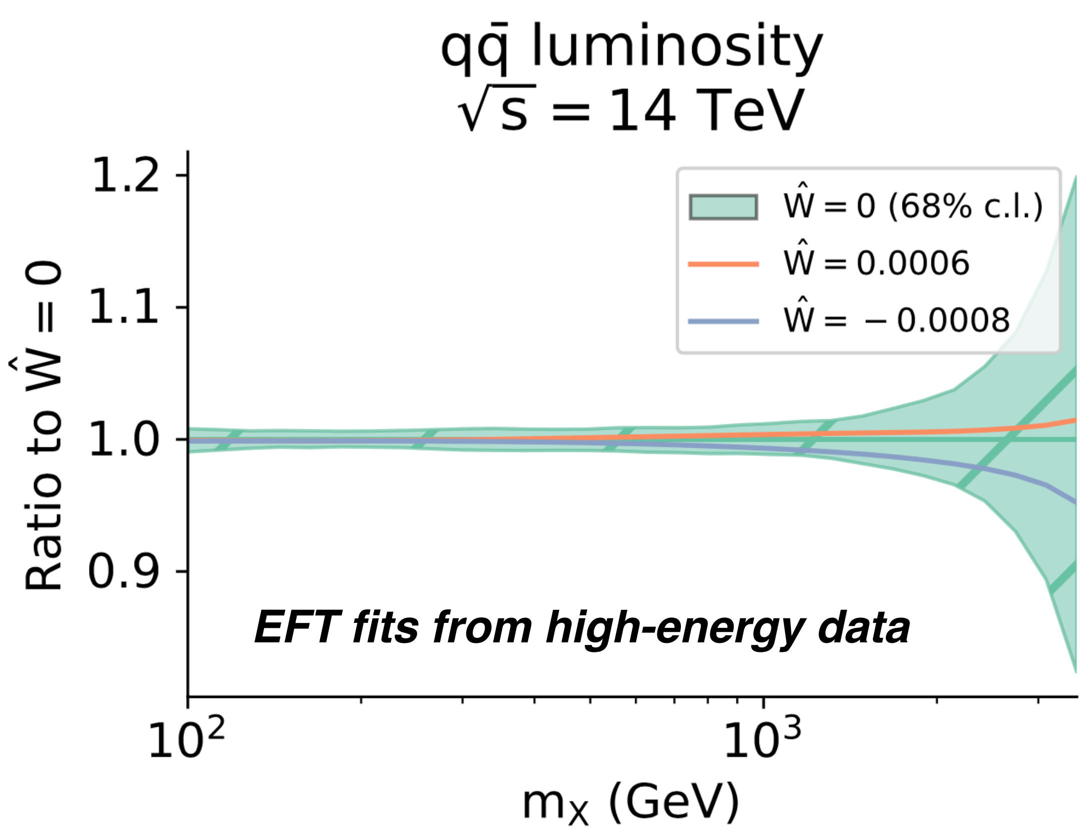
Why we need better PDFs?

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



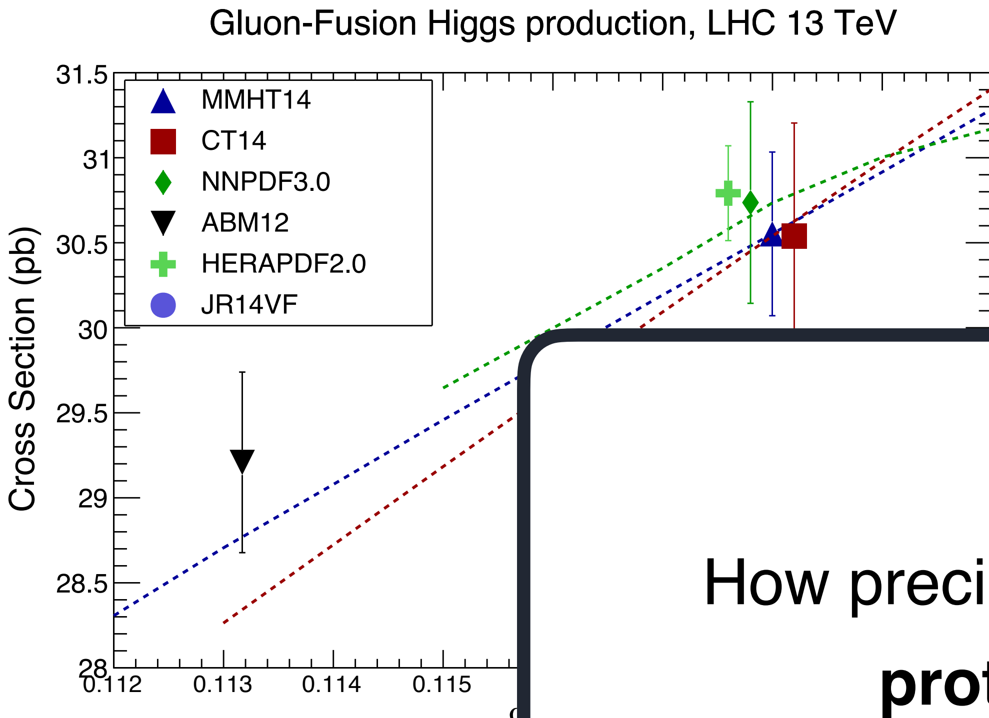


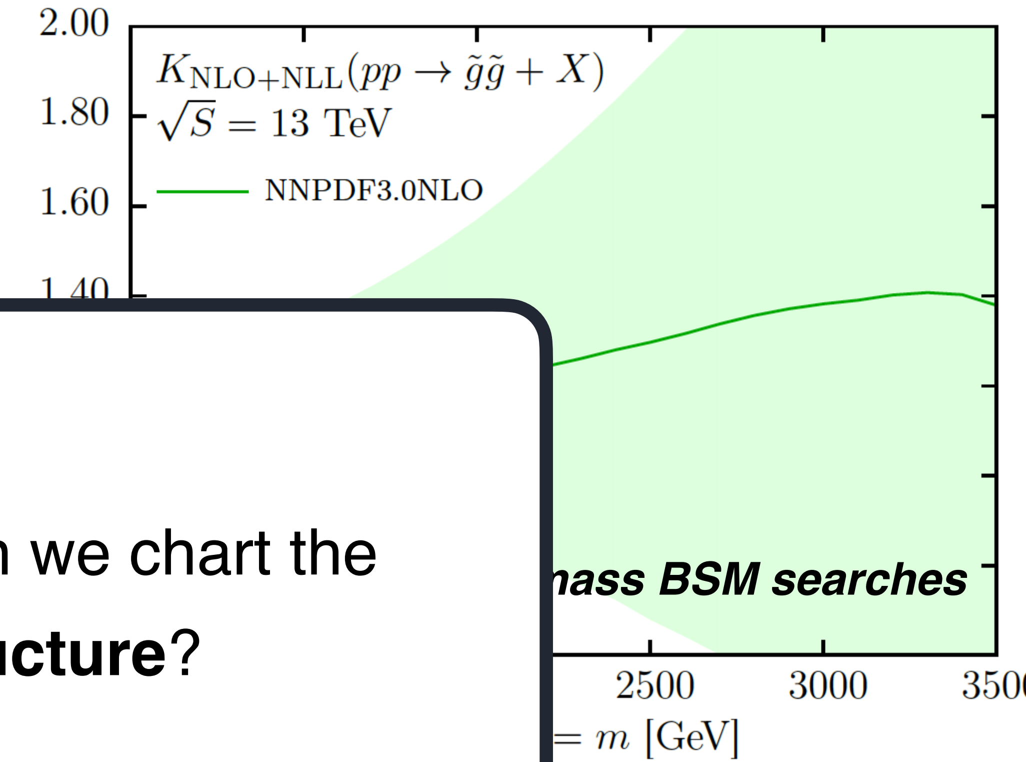


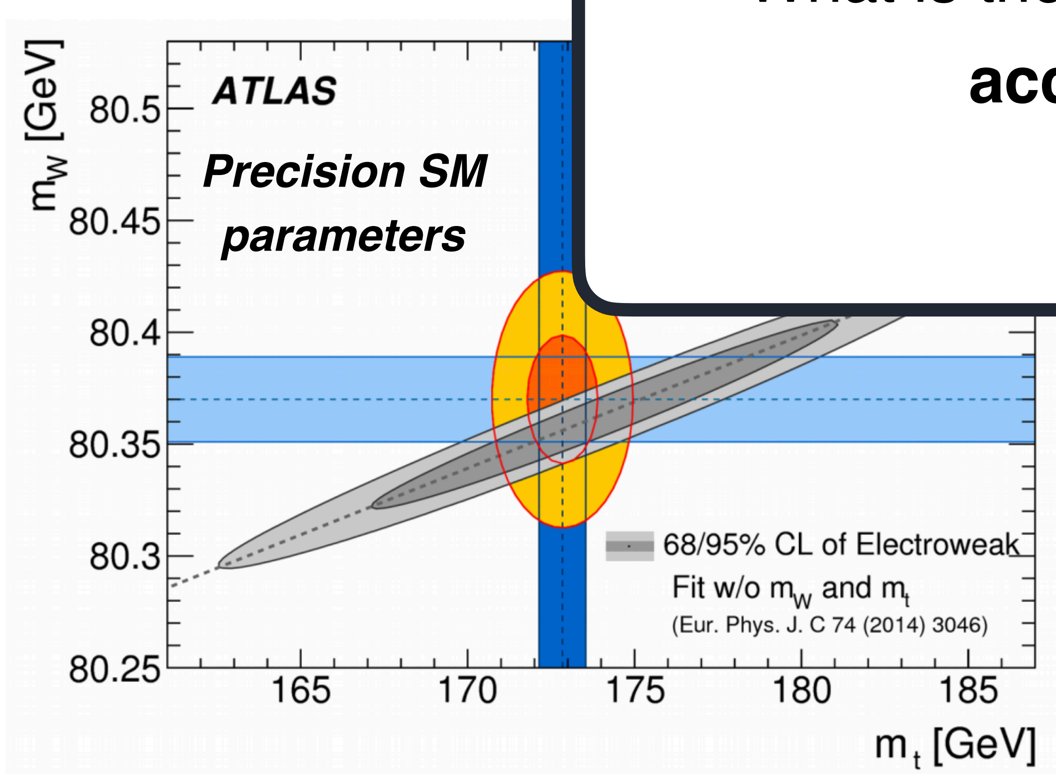


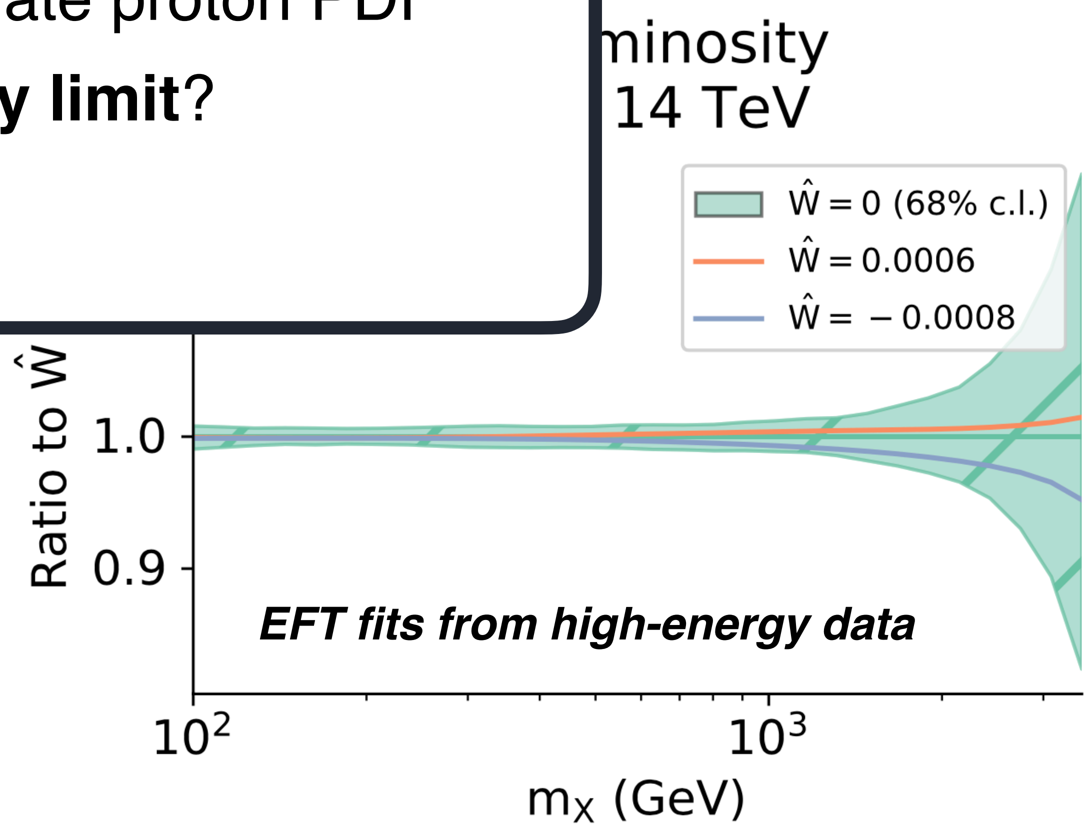
Why we need better PDFs?

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





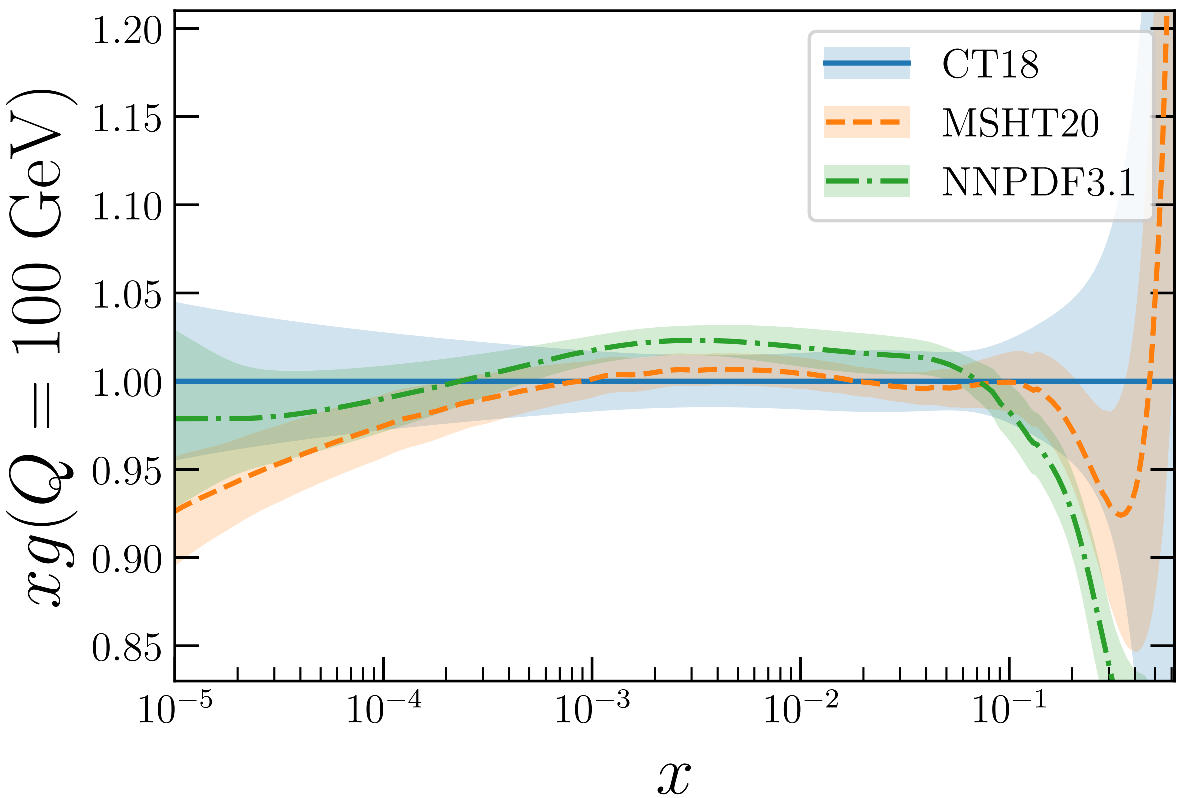




How precisely can we chart the **proton structure**?

What is the ultimate proton PDF **accuracy limit**?

Recent progress in global fits



PDF4LHC21
2112.aaaaa

NNPDF4.0
2109.aaaaa

MSHT20
2012.04684

CT18
1912.10053

NNPDF3.1
1706.00428

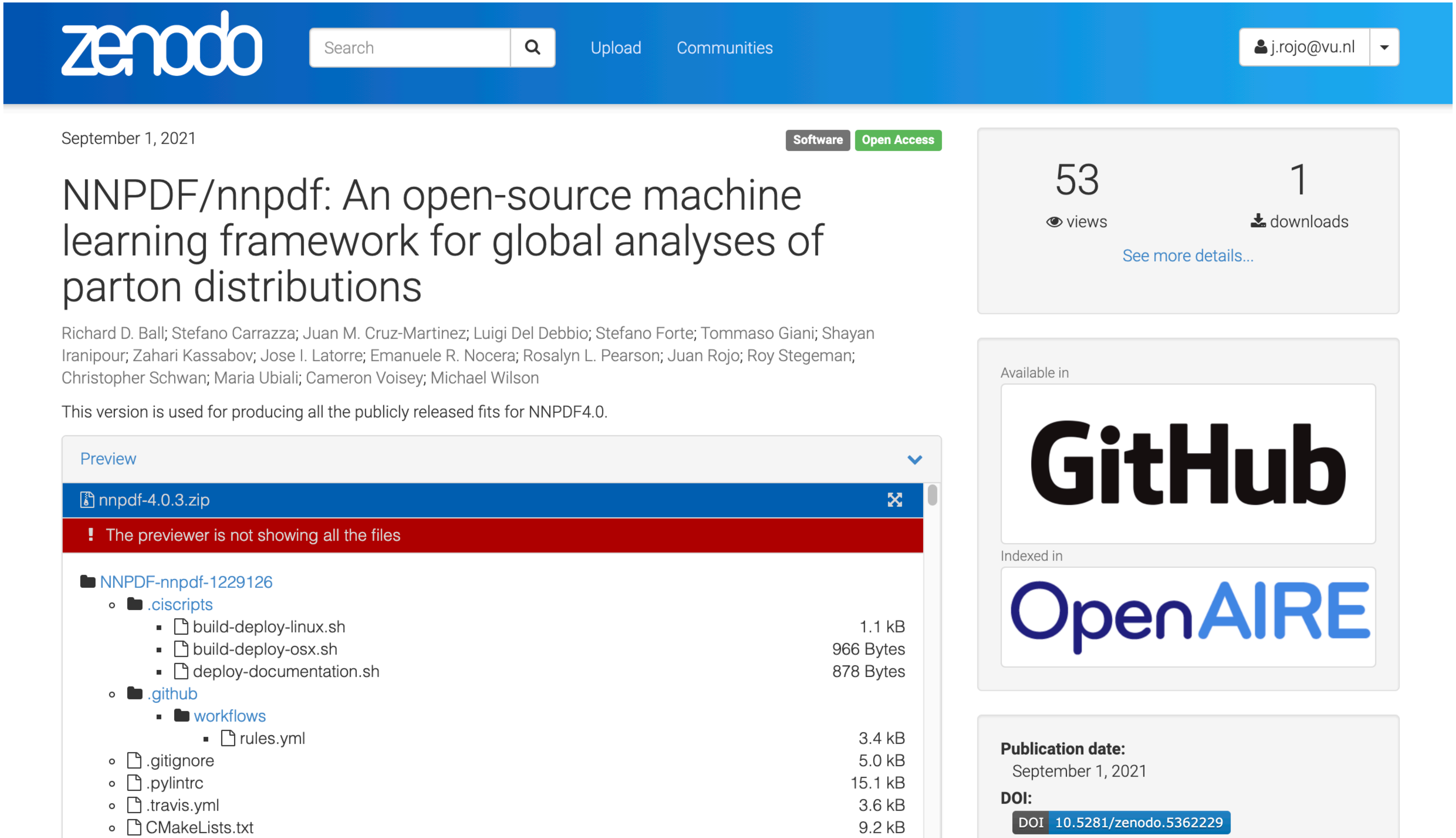
ABMP16
1701.05838

- NNPDF3.1, CT18, MSHT20 based on **comparable datasets**: suitable for a new PDF4LHC(21) combination. Goal: a **conservative PDF uncertainty** estimate

- **NNPDF4.0** based on broader dataset + improved theory constraints and fitting methodology: suitable for **high-precision calculations** and related pheno studies

to be submitted to LHAPDF this week

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

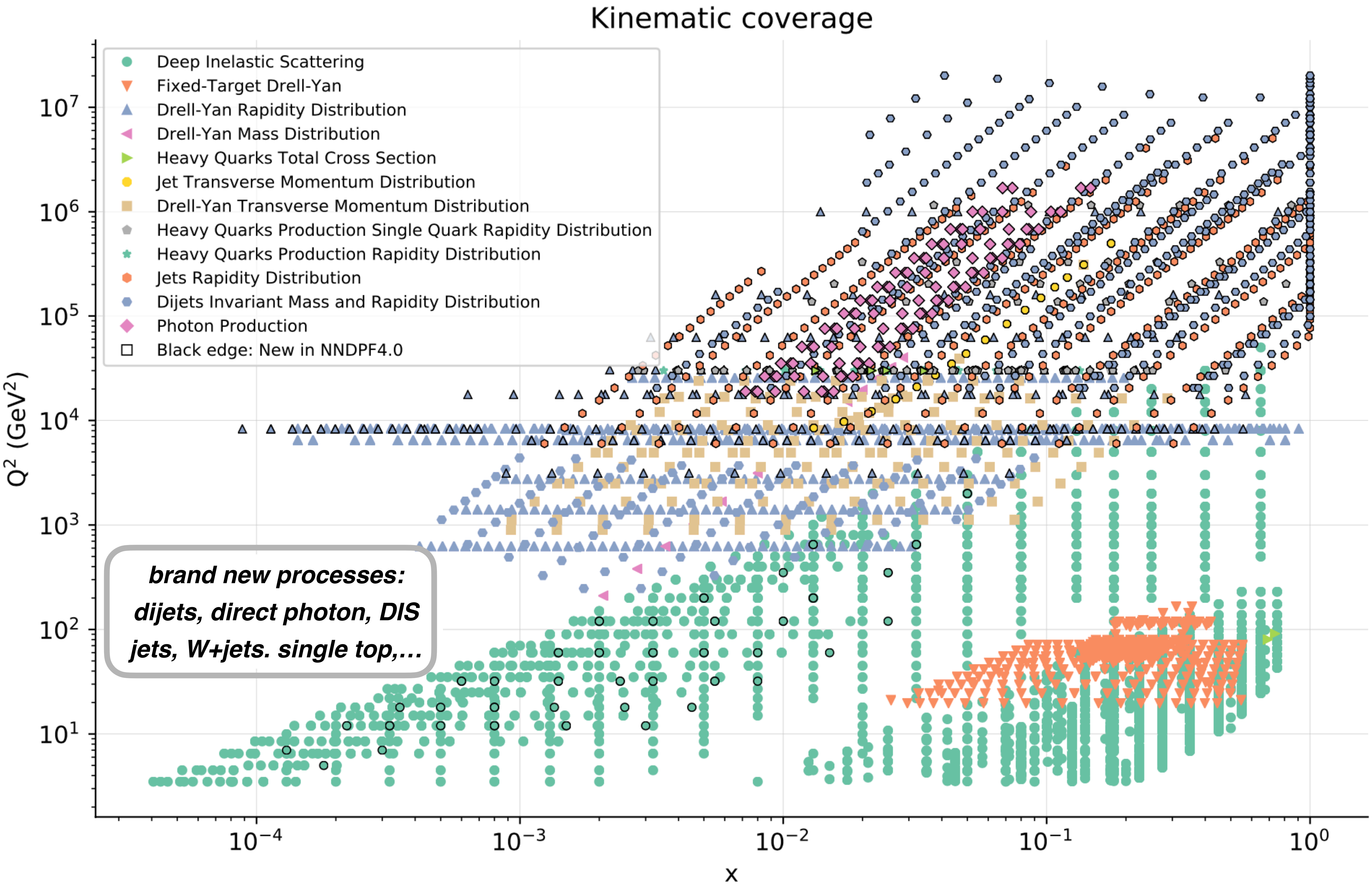
The scientific output of the collaboration is freely available to the publi through the arXiv, journal repositories, and software repositories. Along with this online documentation, we release the NNPDF code used to produce the latest family of PDFs from NNPDF, NNPDF4.0. The code is made available as 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.

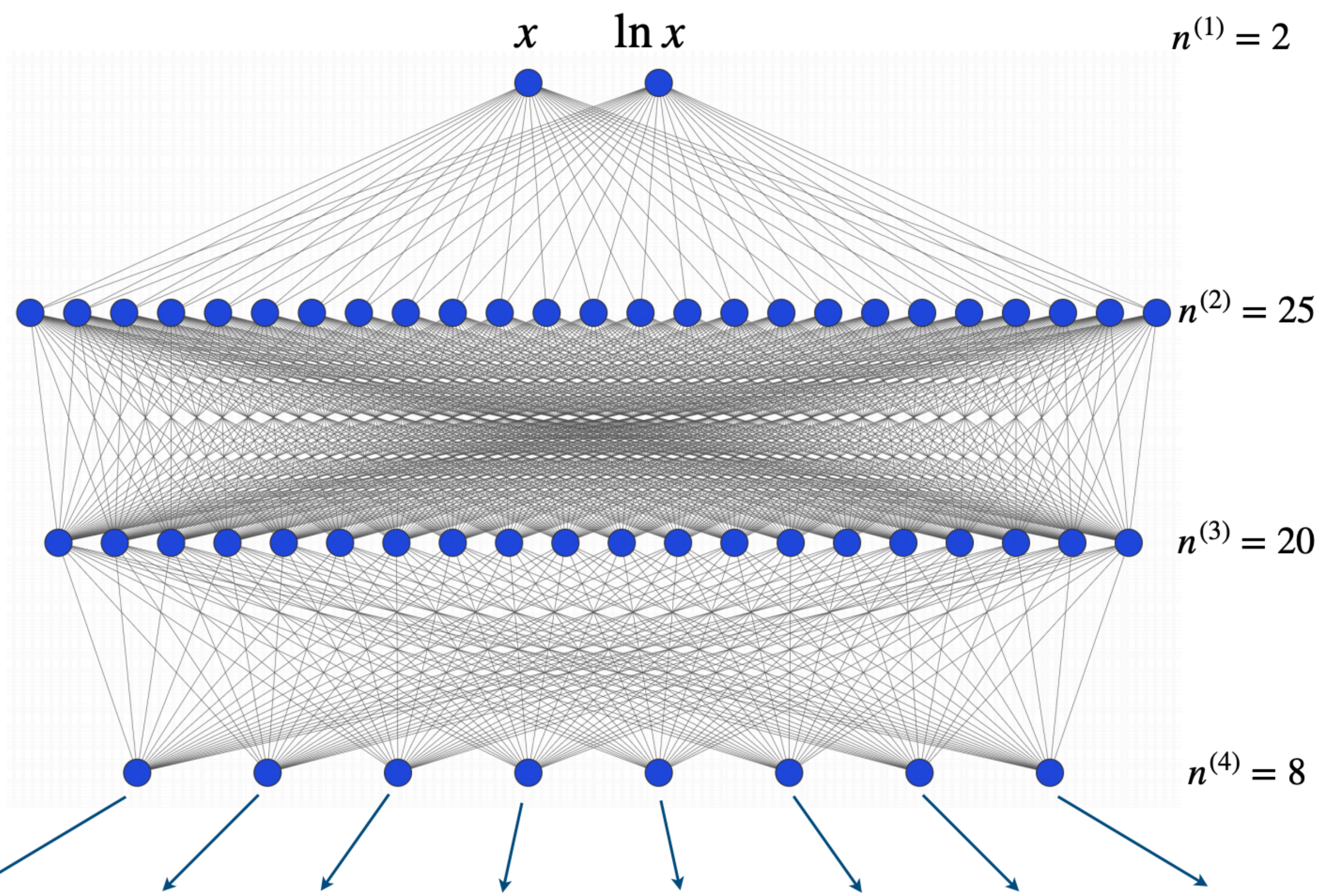
The full **NNPDF machine learning fitting framework** has been publicly released open source, together with extensive documentation and user-friendly examples

NNPDF4.0: data set extension



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

Improved fitting methodology in NNPDF4.0

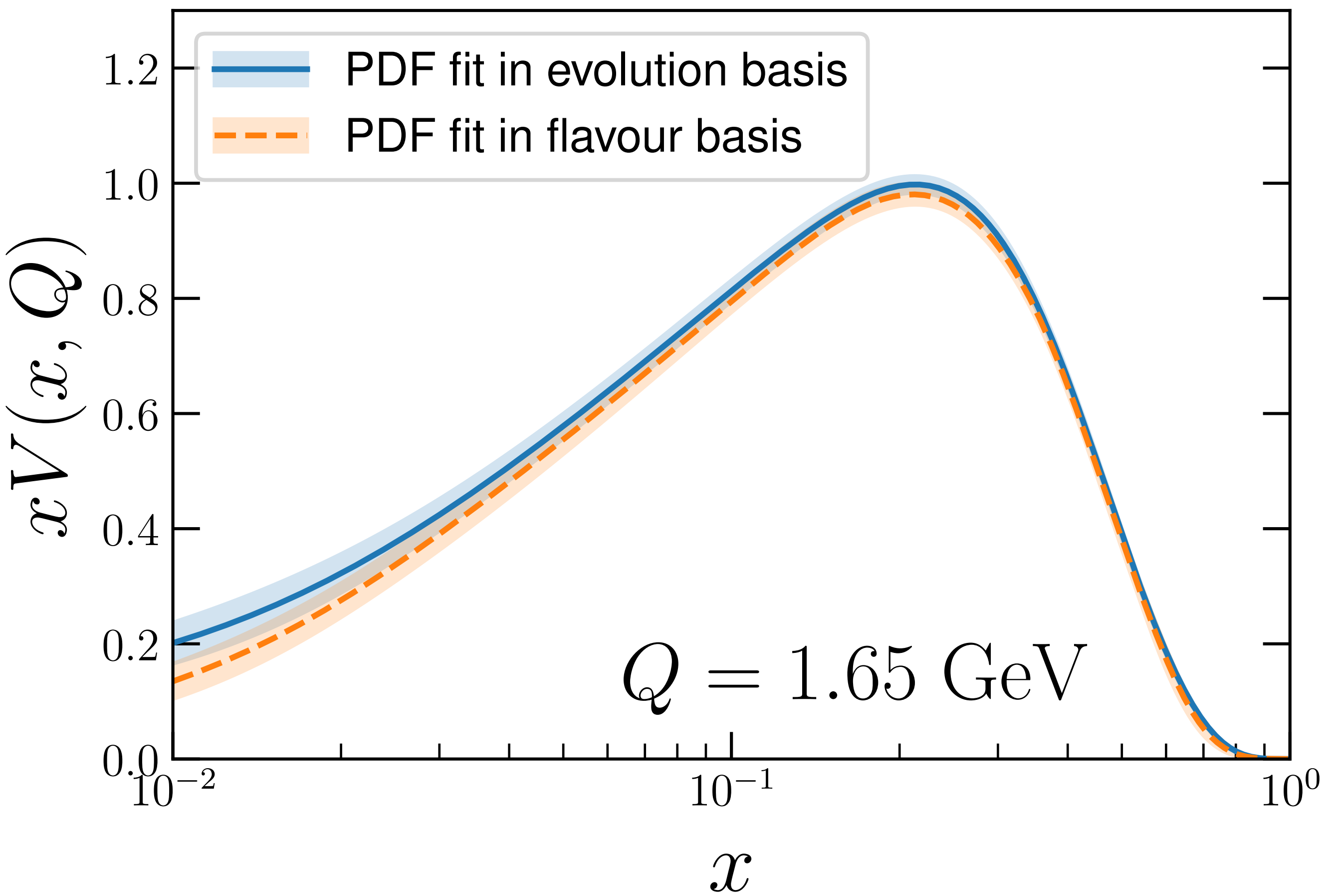


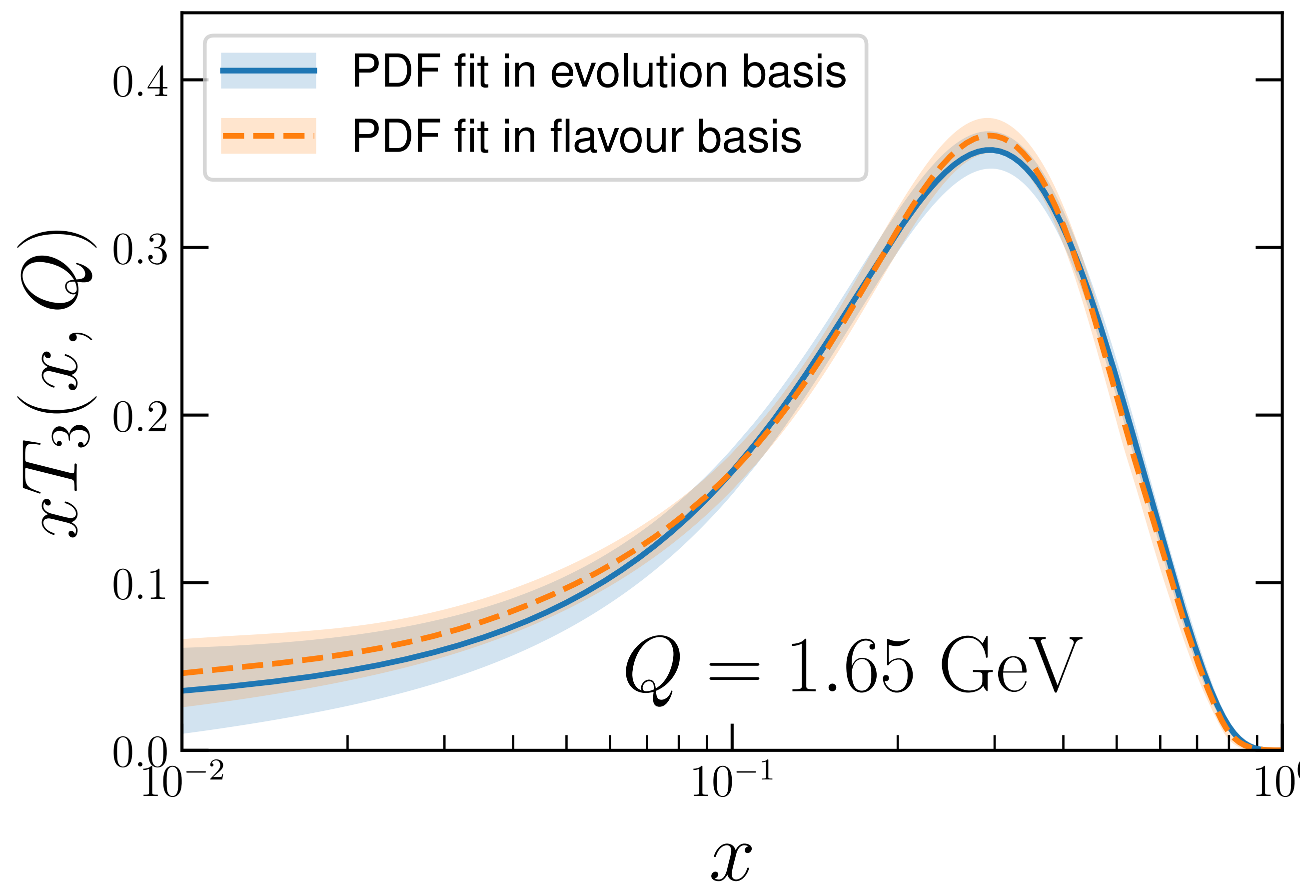
$xg(x, Q_0)$	$x\Sigma(x, Q_0)$	$xV(x, Q_0)$	$xV_3(x, Q_0)$	$xV_8(x, Q_0)$	$xT_3(x, Q_0)$	$xT_8(x, Q_0)$	$xT_{15}(x, Q_0)$	***evolution basis***
$xg(x, Q_0)$	$xu(x, Q_0)$	$x\bar{u}(x, Q_0)$	$xd(x, Q_0)$	$x\bar{d}(x, Q_0)$	$xs(x, Q_0)$	$x\bar{s}(x, Q_0)$	$xc^+(x, Q_0)$	***flavor basis***

$$f_i(x, Q_0) = x^{-\alpha_i}(1 - x)^{\beta_i}\mathrm{NN}_i(x)$$

- ☑ **Stochastic Gradient Descent** (via TensorFlow) for NN training
- ☑ Automated model **hyperparameter optimisation**
- ☑ Validation with **future tests** (forecasting new datasets) and **closure tests** (pseudo-data based on known PDFs)
- ☑ **Generative Adversarial Networks** for replica compression

Parametrisation basis independence





evolution basis PDF parametrisation:

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

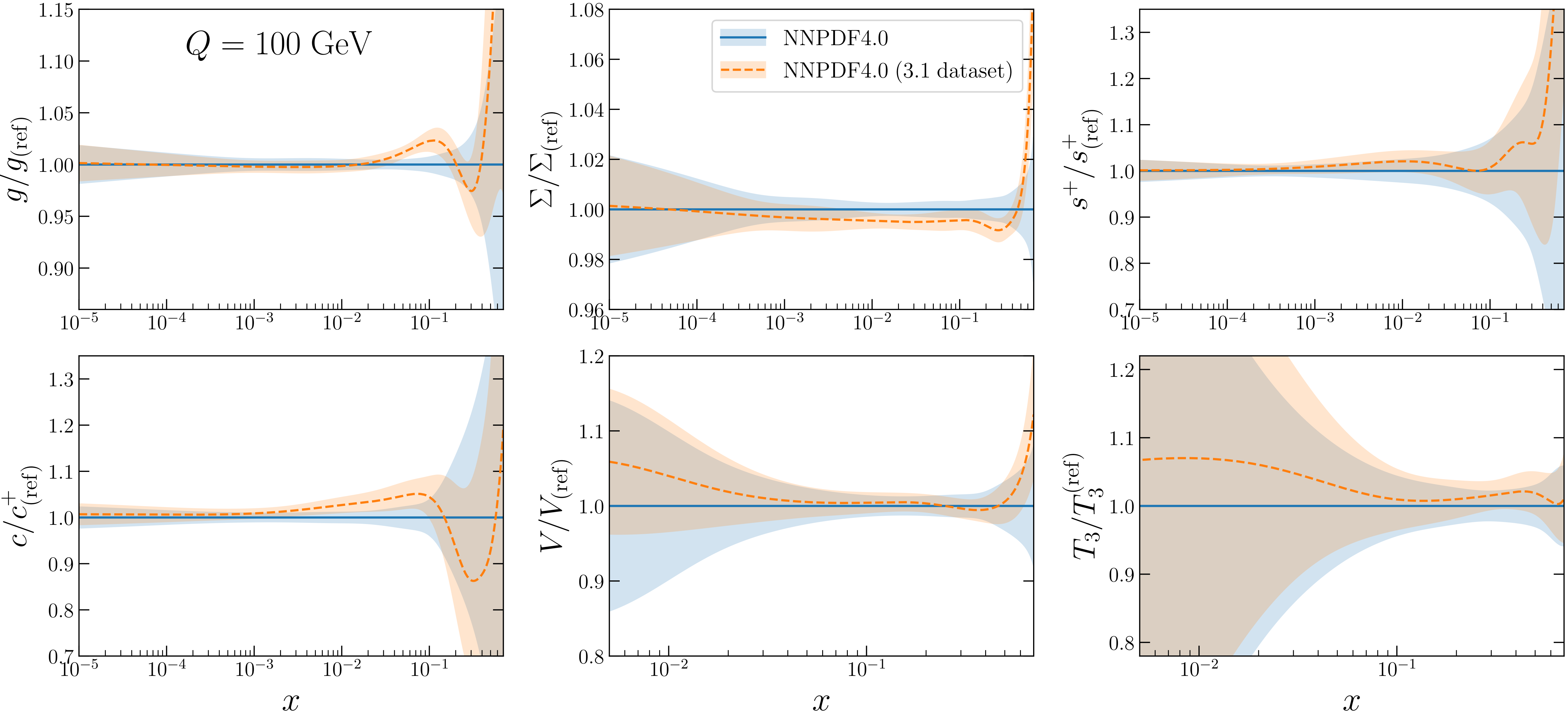
Radically different strategies to parametrize the **quark PDF flavour combinations** lead to identical results: ultimate test of **parametrisation independence**

flavour basis PDF parametrisation:

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

Impact of new data in NNPDF4.0

Good consistency with NNPDF3.1-like dataset

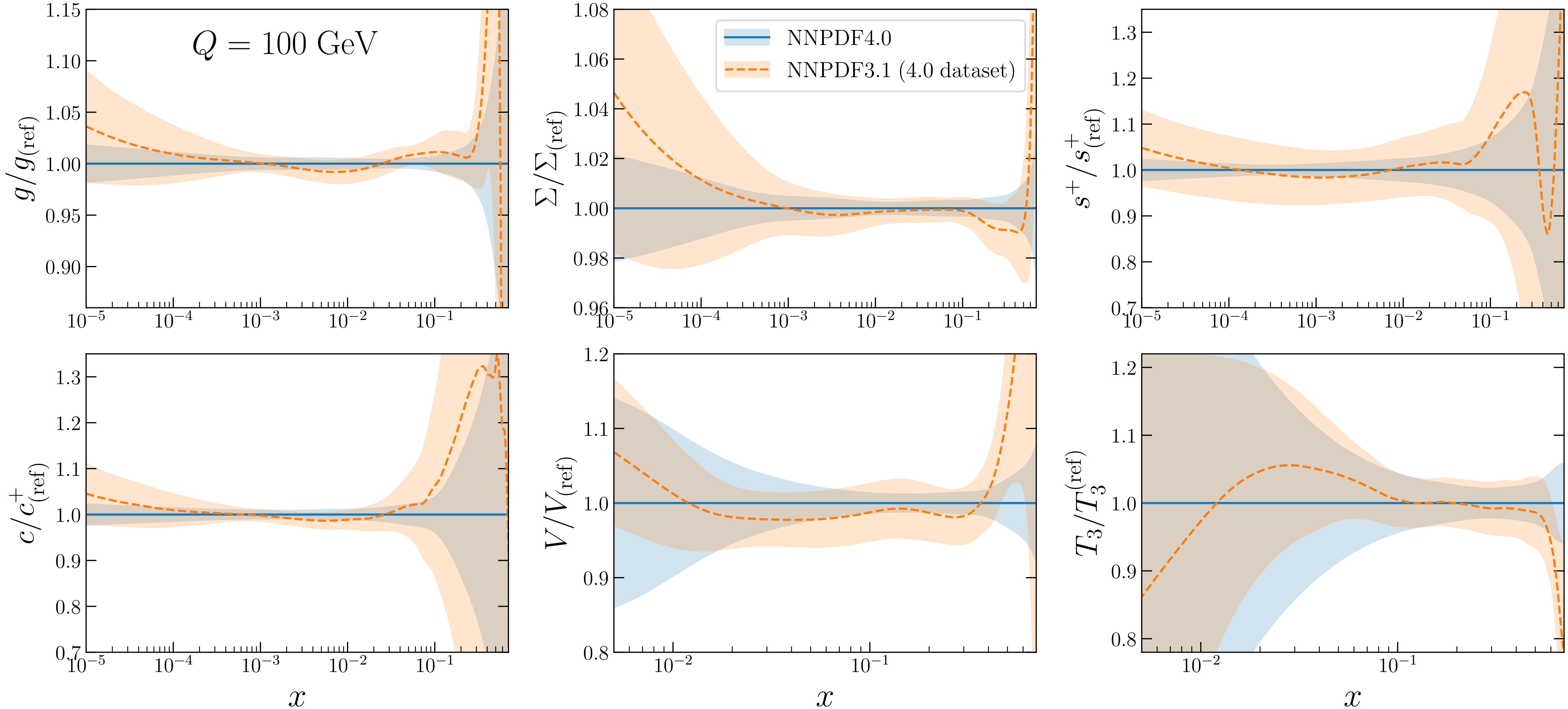


Moderate reduction of PDF uncertainties, shifts in central values at the one-sigma level

PDF uncertainties in the quark sector constrained by the integrability of T_3 and T_8 (Gottfried sum rule)

Impact of new fitting methodology

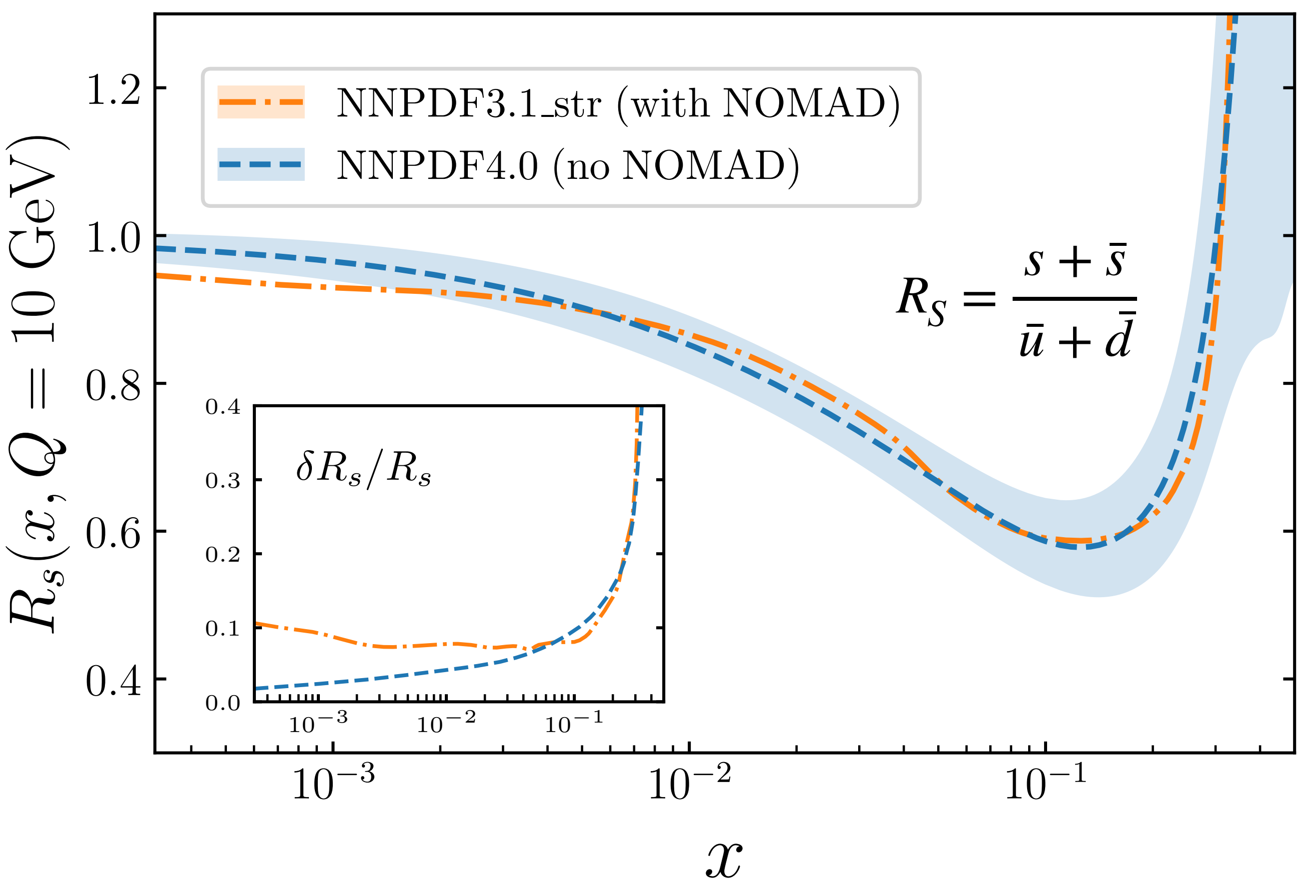
Compare fits based on the **same NNPDF4.0 dataset** but either with the **previous 3.1** or the **new 4.0 methodology:** significant reduction of uncertainties + consistency

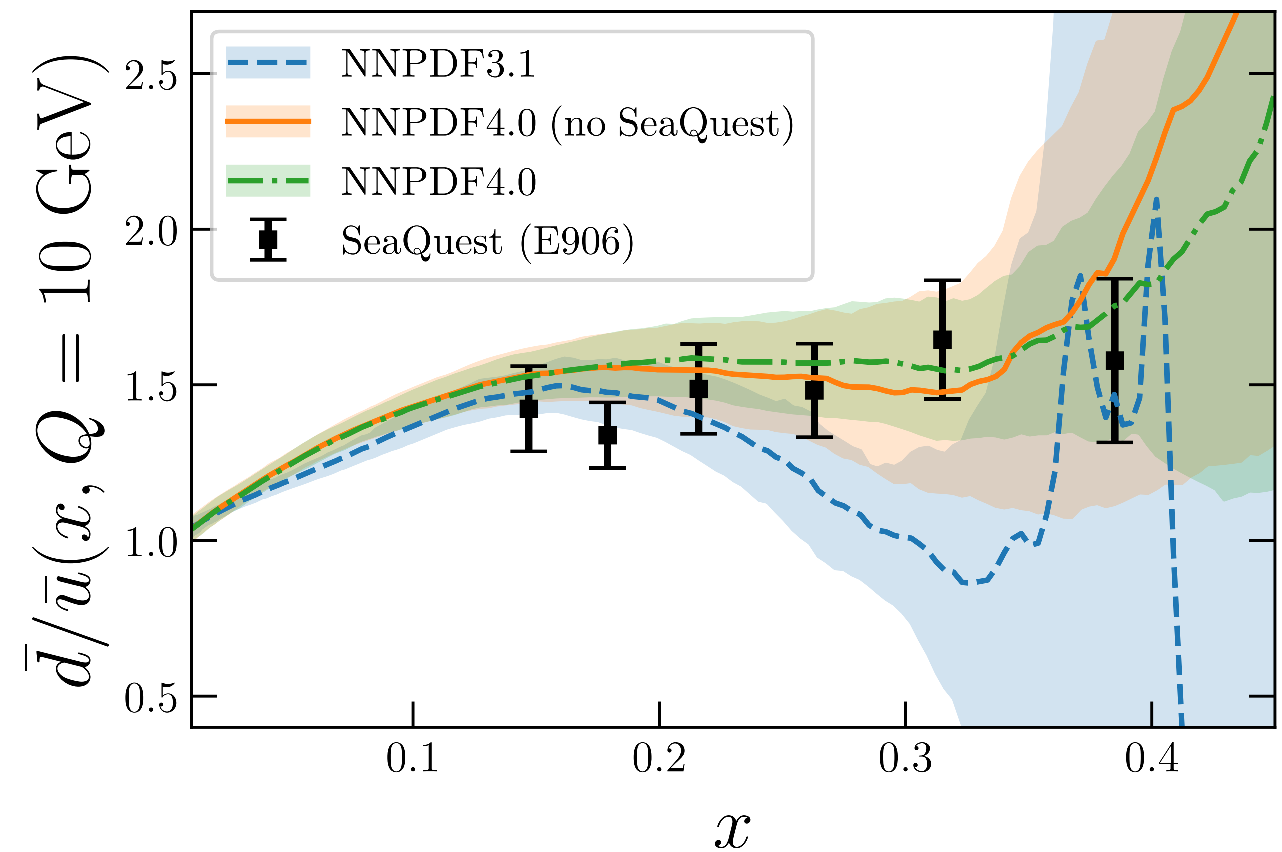


PDF uncertainties validated by means of **closure tests, future tests,** and **basis independence tests**

Validation tests succesful both with the NNPDF4.0 and NNPDF3.1 methodologies

From antimatter asymmetry to intrinsic charm

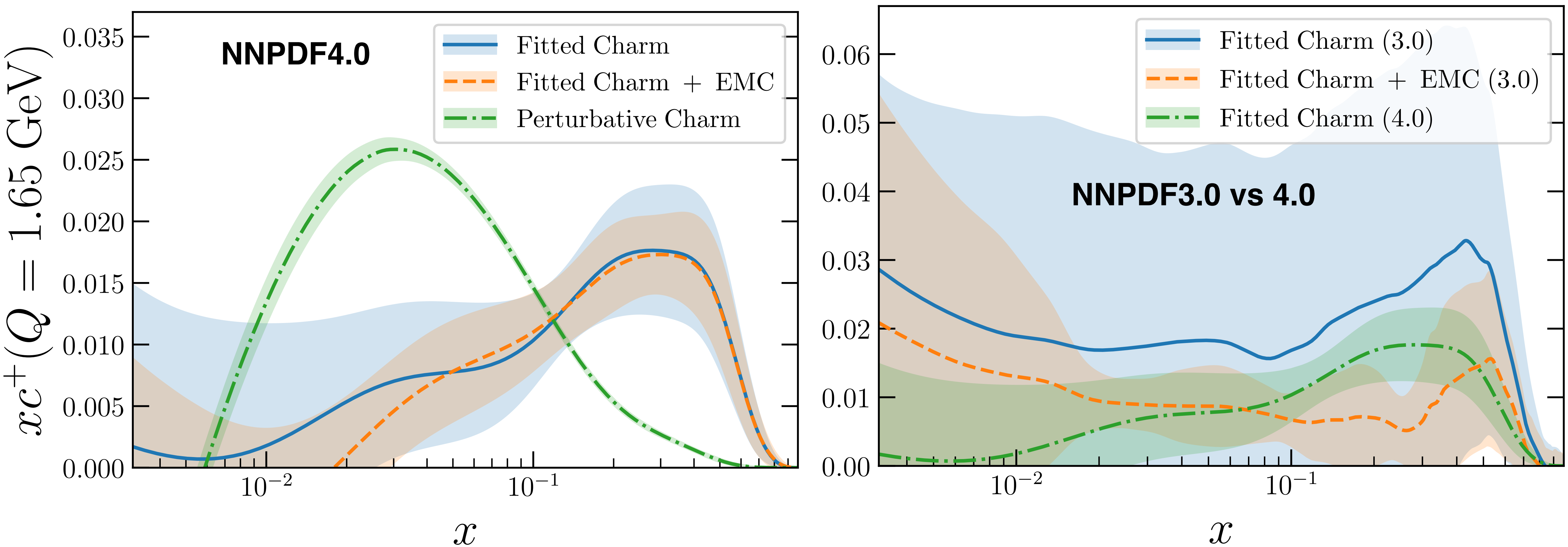




High-resolution mapping of proton structure:

- Confirmation of **strangeness suppression factor**, consistency with NOMAD constraints
- Coherent pull of global fit dataset on strangeness
- Small-x behaviour constrained by **T_3 and T_8 integrability**
- NNPDF4.0 fits the very recent **SeaQuest data**
- Improved determination of **large-x antiquarks** (relevant for many BSM searches at LHC)
- NNPDF3.1 and **NNPDF4.0(no SeaQuest)** already agree with SeaQuest data within errors

From antimatter asymmetry to intrinsic charm

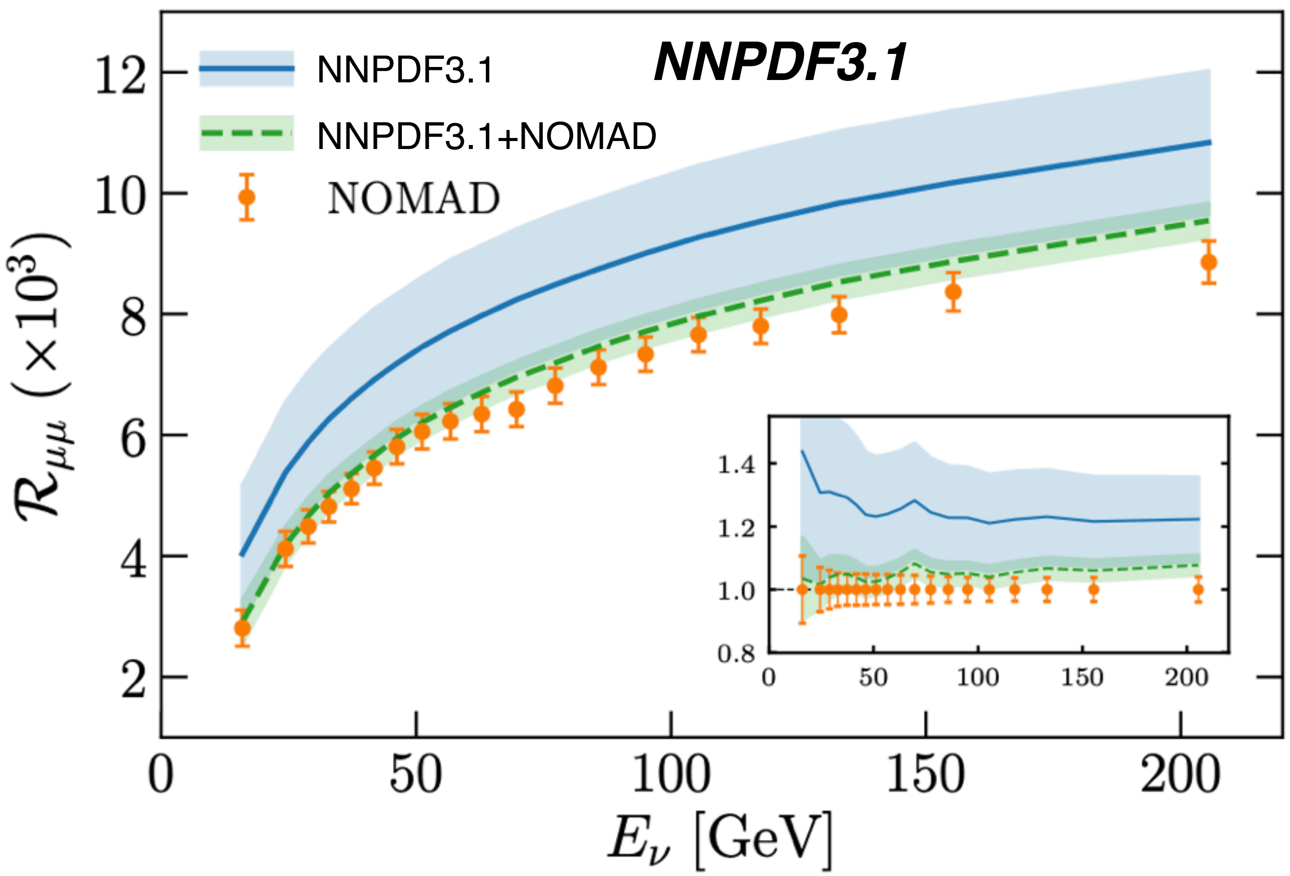


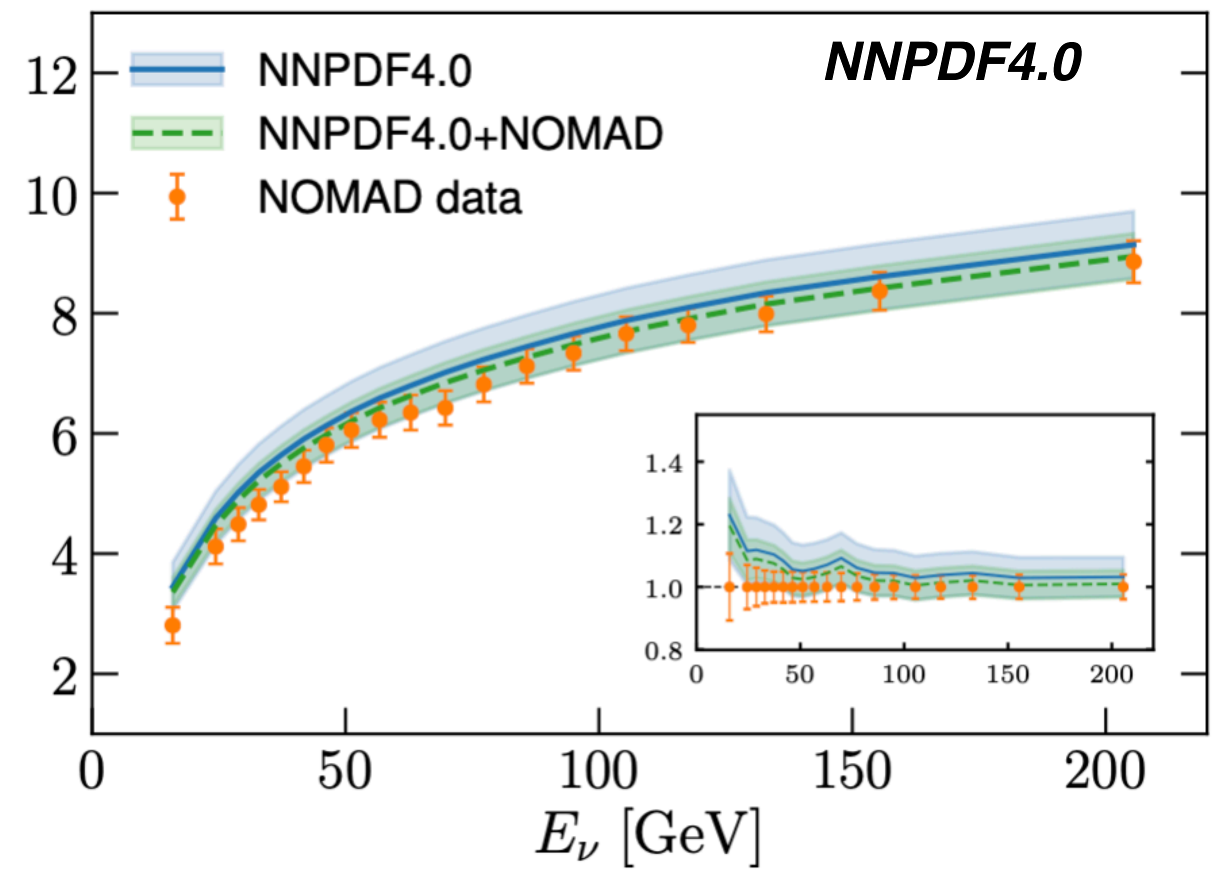
- Increased evidence for **non-perturbative charm component** within the proton, robust upon conversion to the **3FNS** via backwards evolution and matching conditions

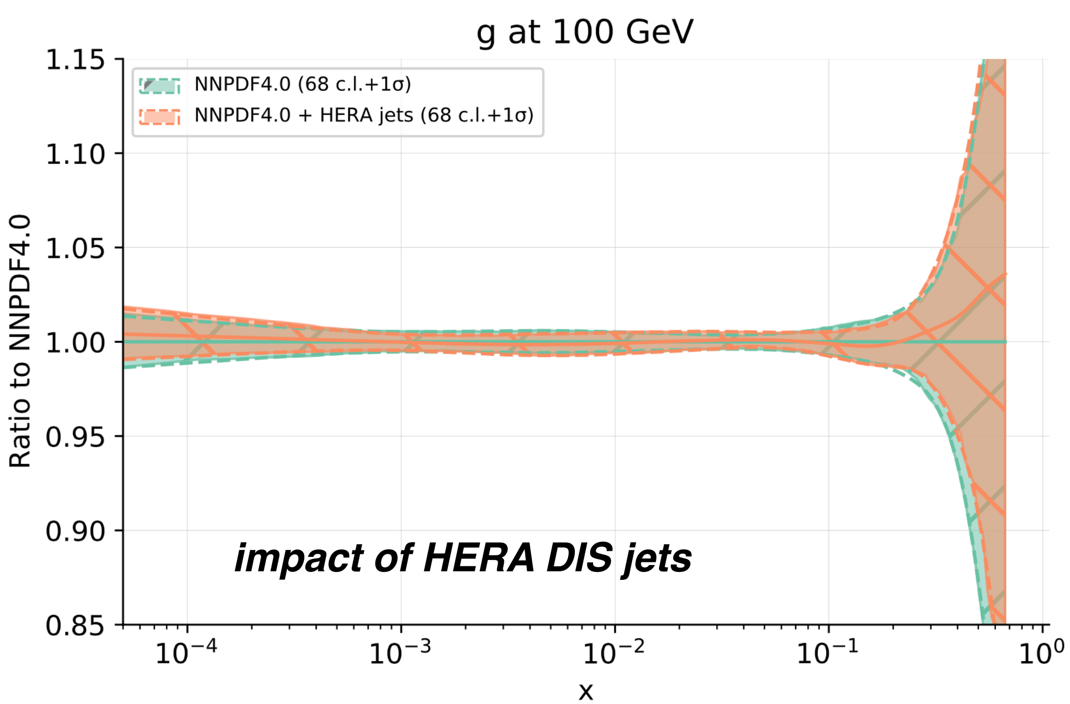
- Bulk of constraints provided by **precision LHC data**, complemented by fixed-target DIS

- As opposed to previous studies, impact of the **EMC charm measurements** mild now. Information provided by EMC F_2^c (early 80s!) consistent with latest collider data

Consistency checks





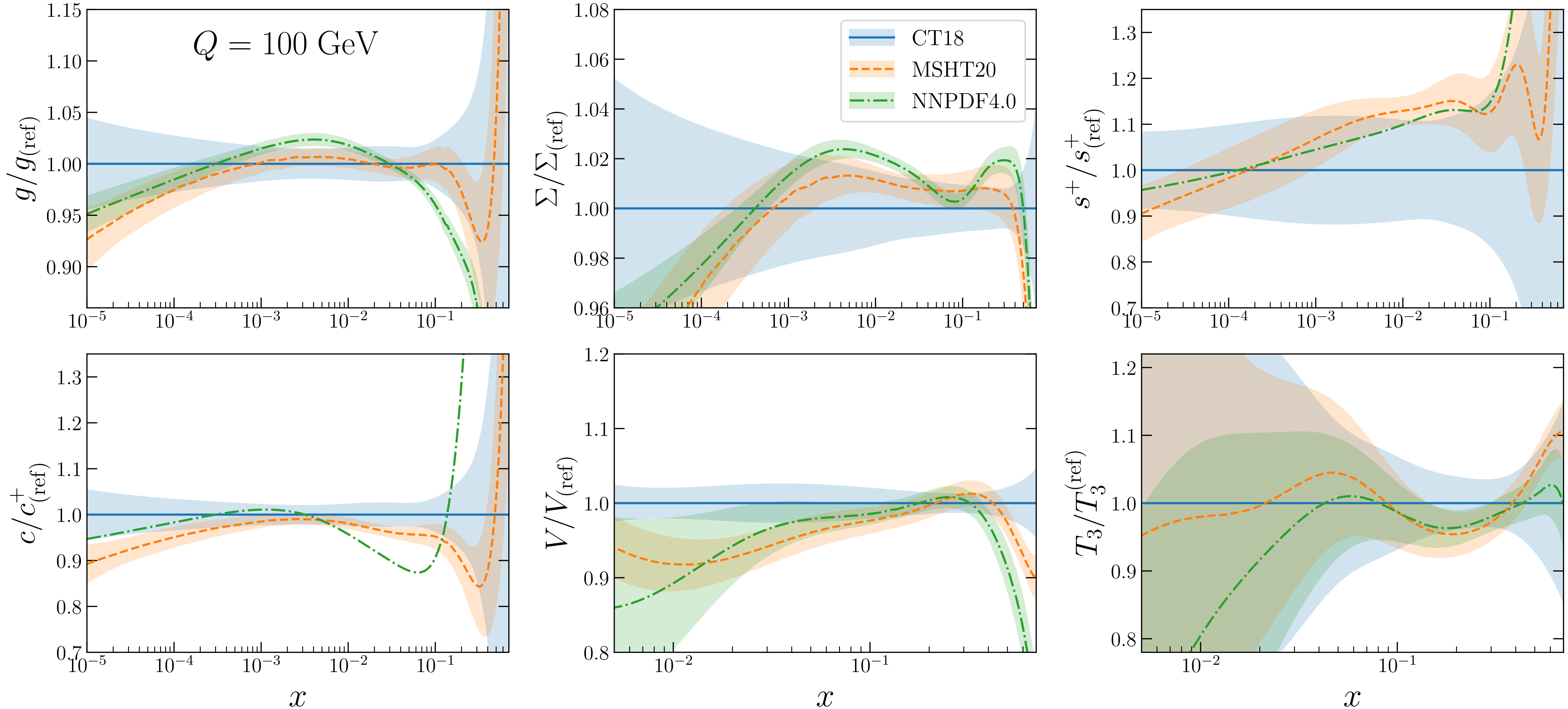


- ☑ Fitting NOMAD had a large impact on the proton strangeness in NNPDF3.1, now in NNPDF4.0 the **no-NOMAD fit** is already spot on
- ☑ Fitting the HERA DIS jets (300 data points) leads to a **gluon** in perfect agreement with NNPDF4.0

Excellent consistency of global dataset

Comparison between global fits

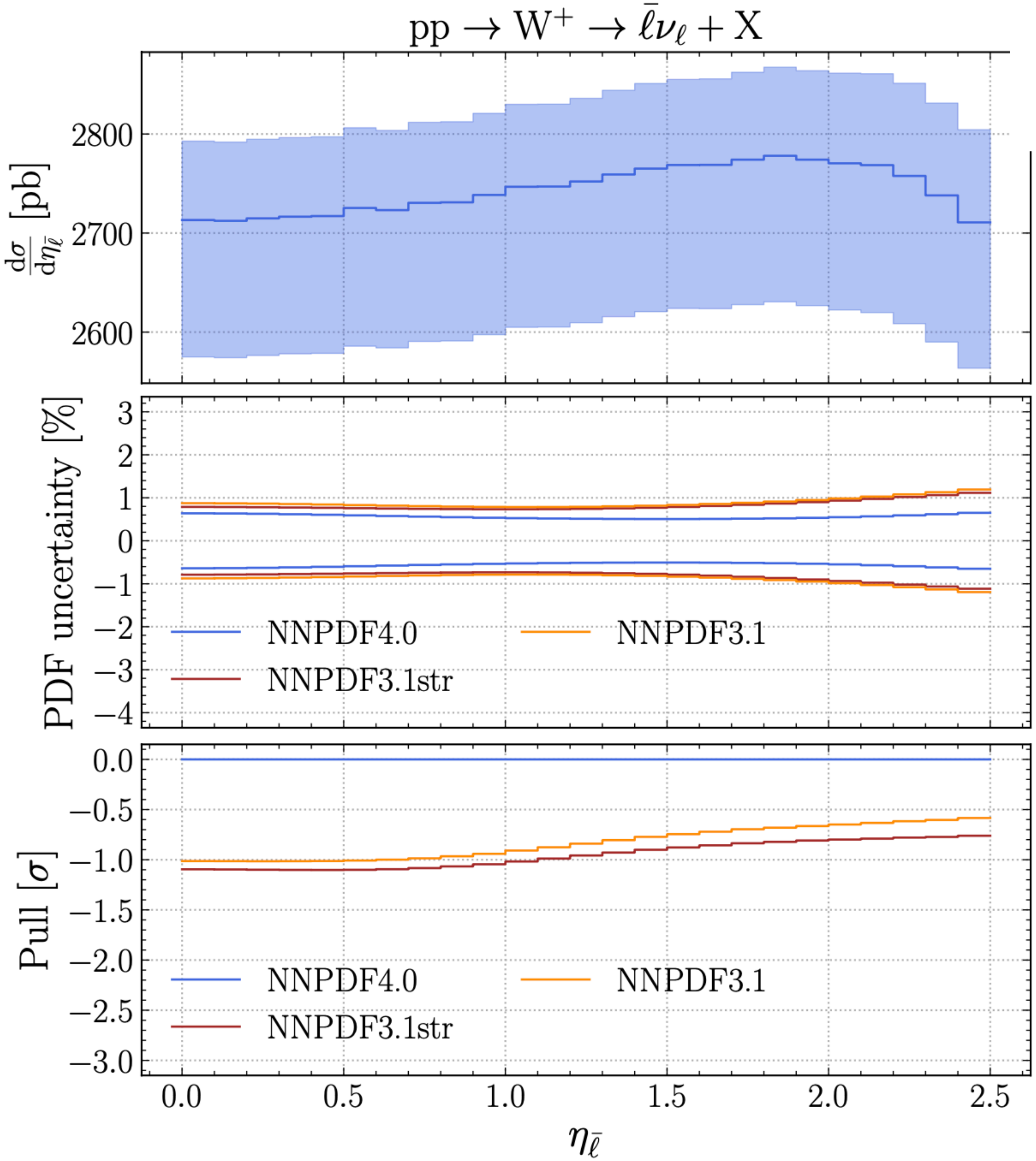
In general good agreement between NNPDF4.0, CT18, and MSHT20, though rather **different pattern of PDF uncertainties**

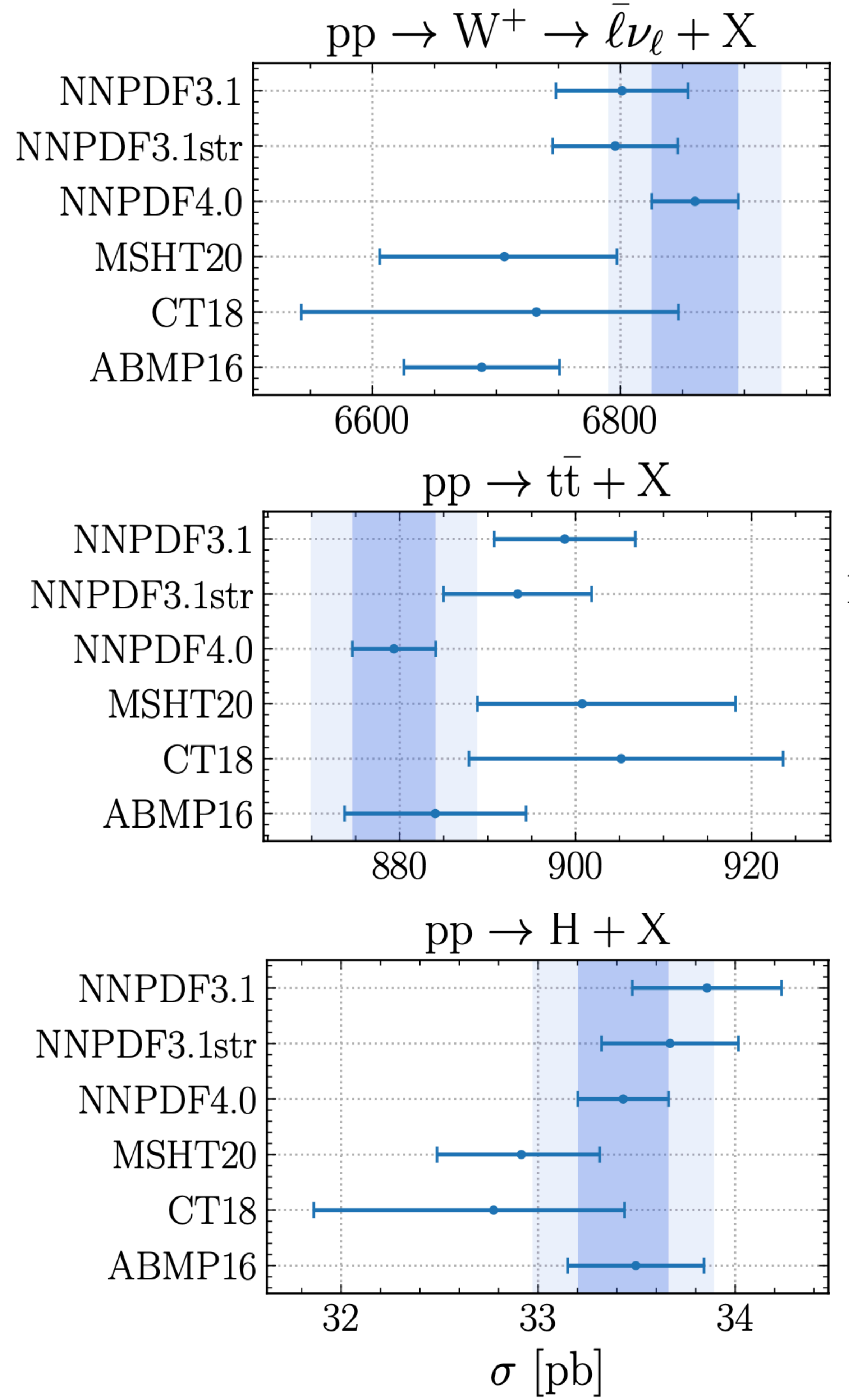


some differences in large-*x* gluon and quark flavour decomposition

LHC phenomenology

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





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 completely validated by **closure tests**, **future tests**, and **parametrisation basis independence**

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

- The current level of PDF uncertainties challenge 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**

Extra Material

From NNPDF3.1 to NNPDF4.0

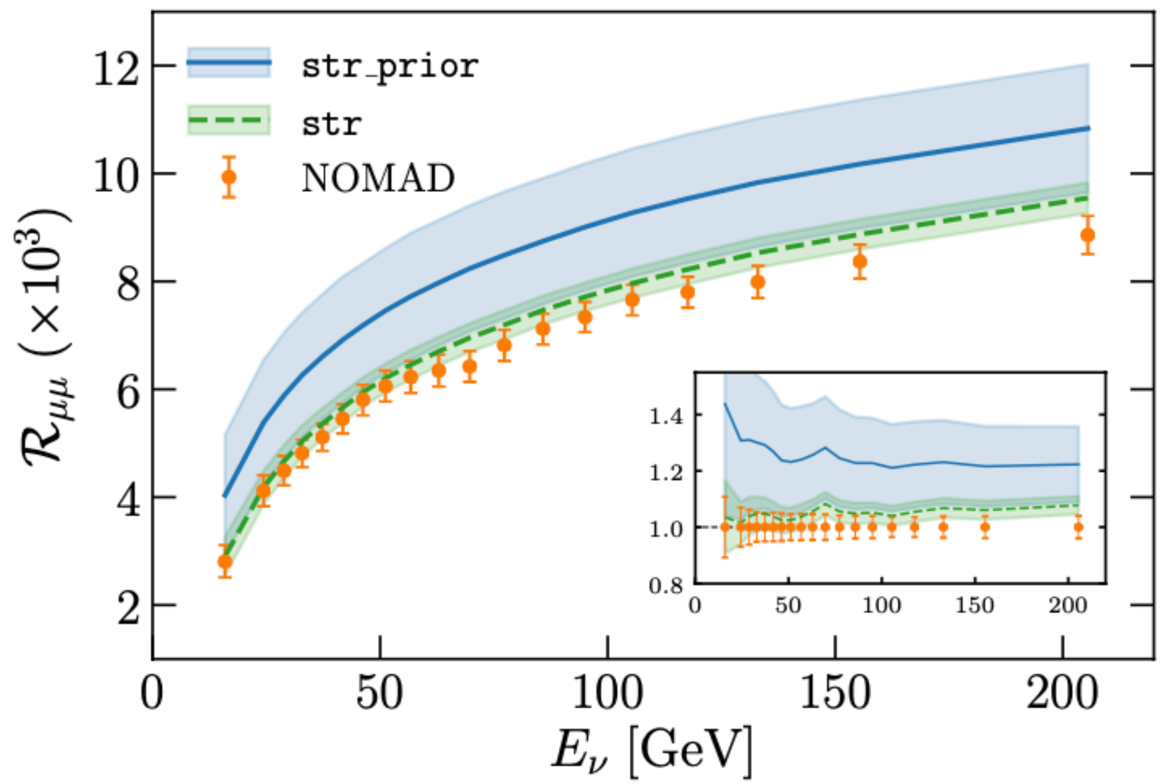
Collaborative progress towards extending data, theory and methodology

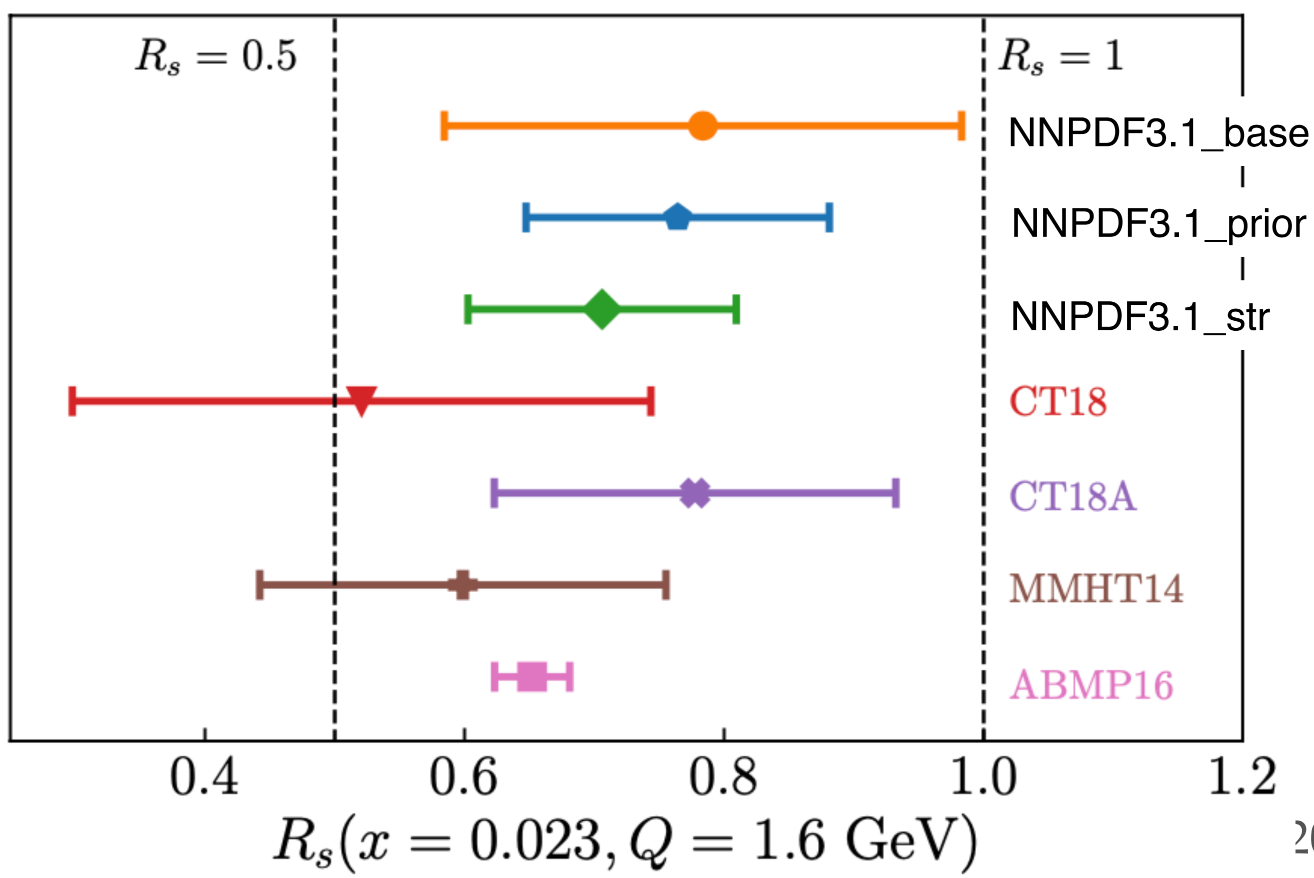
06/2017	**NNPDF3.1**	[EPJ C77 (2017) 663]
10/2017	NNPDF3.1sx: PDFs with small-x resummation	[EPJ C78 (2018) 321]
12/2017	NNPDF3.1luxQED: consistent photon PDF à la luxQED	[SciPost Phys. 5 (2018) 008]
02/2018	NNPDF3.1+ATLASphoton: inclusion of direct photon data	[EPJ C78 (2018) 470]
12/2018	NNPDF3.1alphas: 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**	[to appear]

The strangest proton?

Reappraisal of the proton strangeness based combination of **all relevant experimental inputs**

Process	Dataset	$n_{\rm dat}$	$\chi^2_{\rm base}$	$\chi^2_{\rm pr}$	$\chi^2_{\rm str}$
νDIS ($\mu\mu$)		76/76/95/91/95	0.76	0.71	0.53
	NuTeV [9]	76/76/76/76/76	0.76	0.71	0.53
	NOMAD [10]	—/—/19/15/19	[9.3]	[8.8]	0.55
W, Z (incl.)		391/418/418/418/418	1.45	1.40	1.40
	ATLAS [12]	34/61/61/61/61	1.96	1.65	1.67
W+c		—/37/37/37/37	[0.73]	0.68	0.60
	CMS [17, 18]	—/15/15/15/15	[1.04]	0.98	0.96
	ATLAS [16]	—/22/22/22/22	[0.52]	0.48	0.42
W+jets	ATLAS [15]	—/32/32/32/32	[1.58]	1.18	1.18
Total		3981/4077/4096/4092/4096	1.18	1.17	1.17

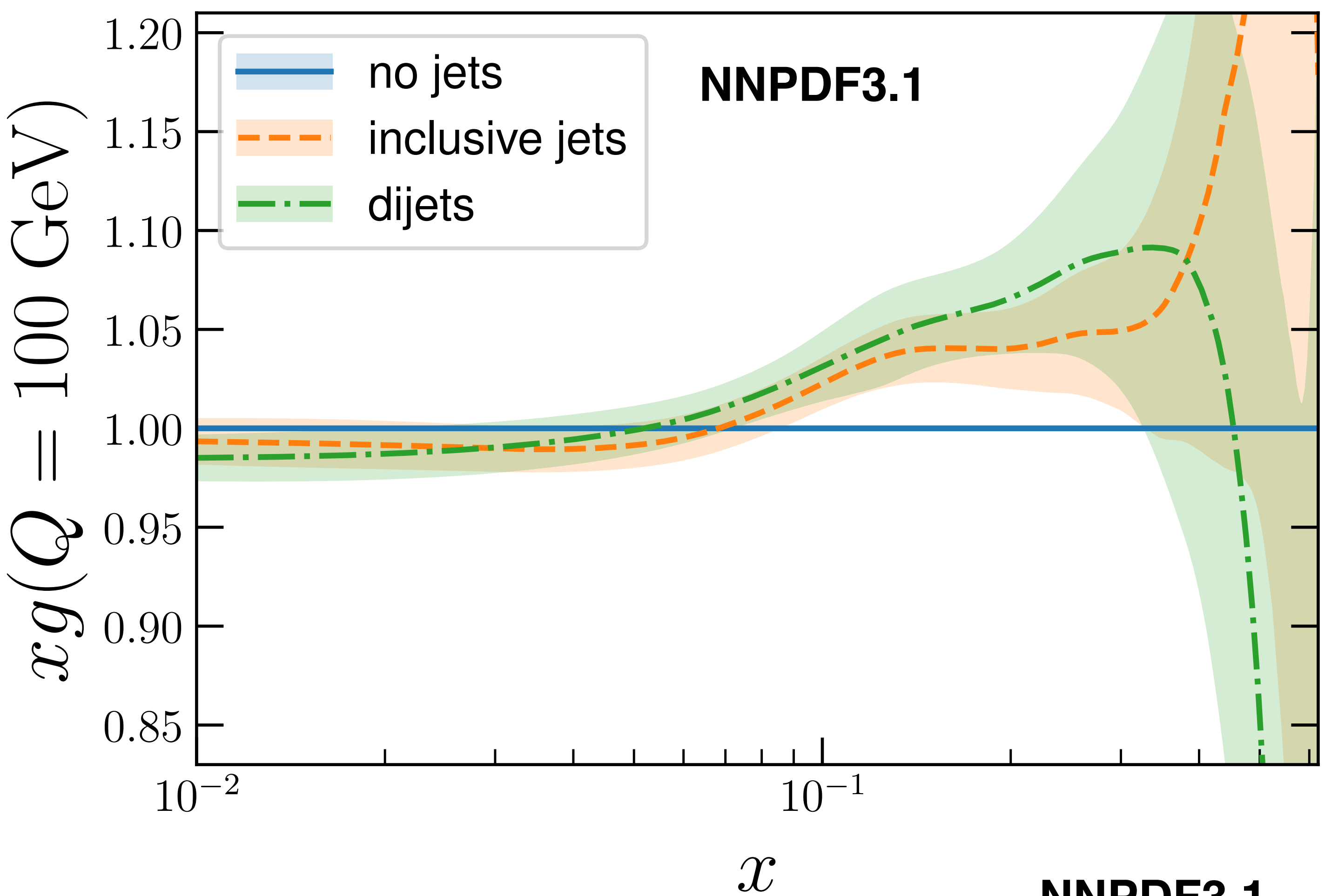




- ☑ Satisfactory simultaneous description of all datasets
- ☑ No **evidence for tension** between datasets or groups of processes
- ☑ Sizeable constraints from **NOMAD neutrino DIS** data, consistent with collider data
- ☑ Preference for a **moderately suppressed strangeness**

$$R_S(x = 0.023,\ Q = 1.6\ \text{GeV}) = 0.71 \pm 0.10$$

LHC dijets to map the gluon PDF



- With NNLO QCD (+ NLO EW) calculations, **all available dijet cross-sections** are successfully included in global fit, good agreement with data

first use of dijets in PDF fits!

- Constrains on the gluon **qualitatively consistent** between dijet and inclusive jet observables

- Inclusive jets requires introducing **tailored decorrelation** models to achieve **good χ^2**

NNPDF3.1

Experiment	Measurement	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [fb^{-1}]	R	Distribution	n_{dat}
ATLAS	Inclusive jets	7	4.5	0.6	$d^2\sigma/dp_T d\|y\|$	140
CMS	Inclusive jets	7	4.5	0.7	$d^2\sigma/dp_T d\|y\|$	133
ATLAS	Inclusive jets	8	20.2	0.6	$d^2\sigma/dp_T d\|y\|$	171
CMS	Inclusive jets	8	19.7	0.7	$d^2\sigma/dp_T d\|y\|$	185
ATLAS	Dijets	7	4.5	0.6	$d^2\sigma/dm_{jj} d\|y^*\|$	90
CMS	Dijets	7	4.5	0.7	$d^2\sigma/dm_{jj} d\|y_{\max}\|$	54
CMS	Dijets	8	19.7	0.7	$d^3\sigma/dp_{T,\mathrm{avg}} dy_b dy^*$	122

dijet study revisited in NNPDF4.0 paper