



The quest towards high-precision global parton distributions

Juan Rojo

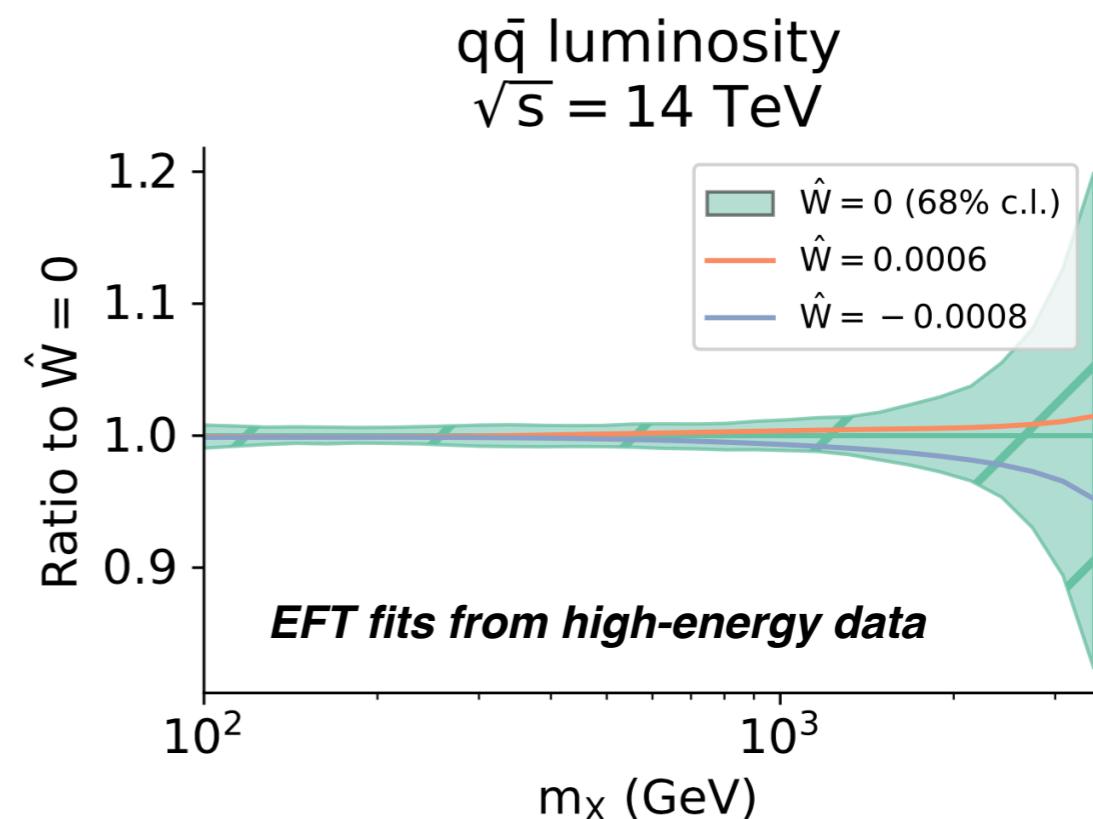
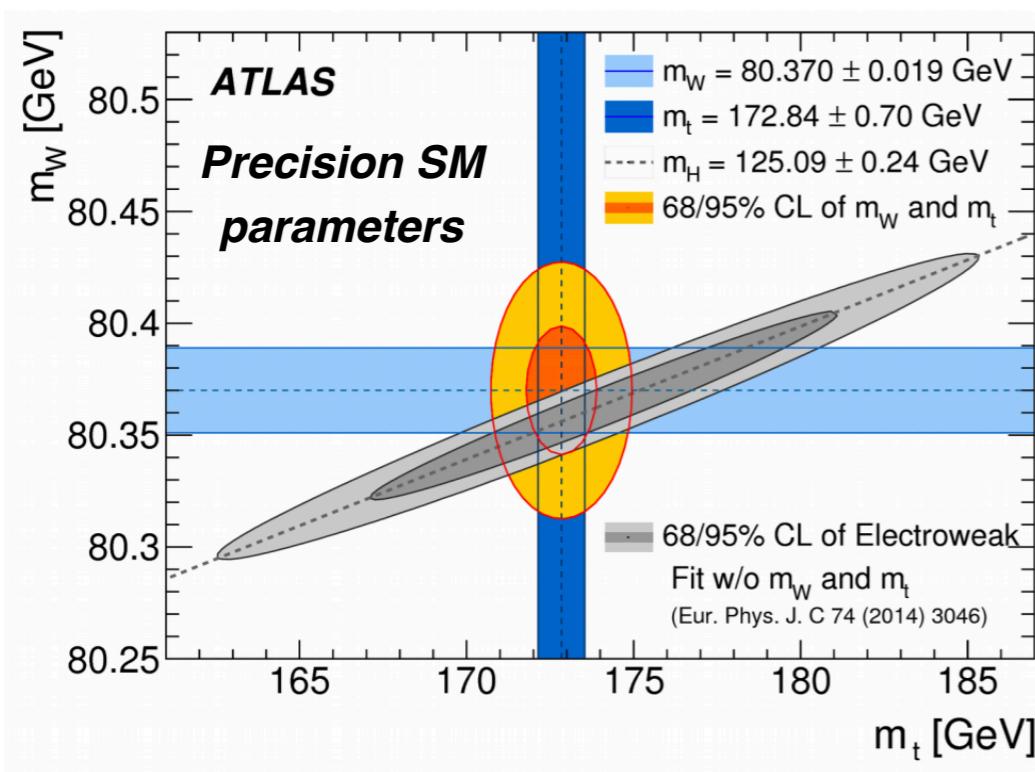
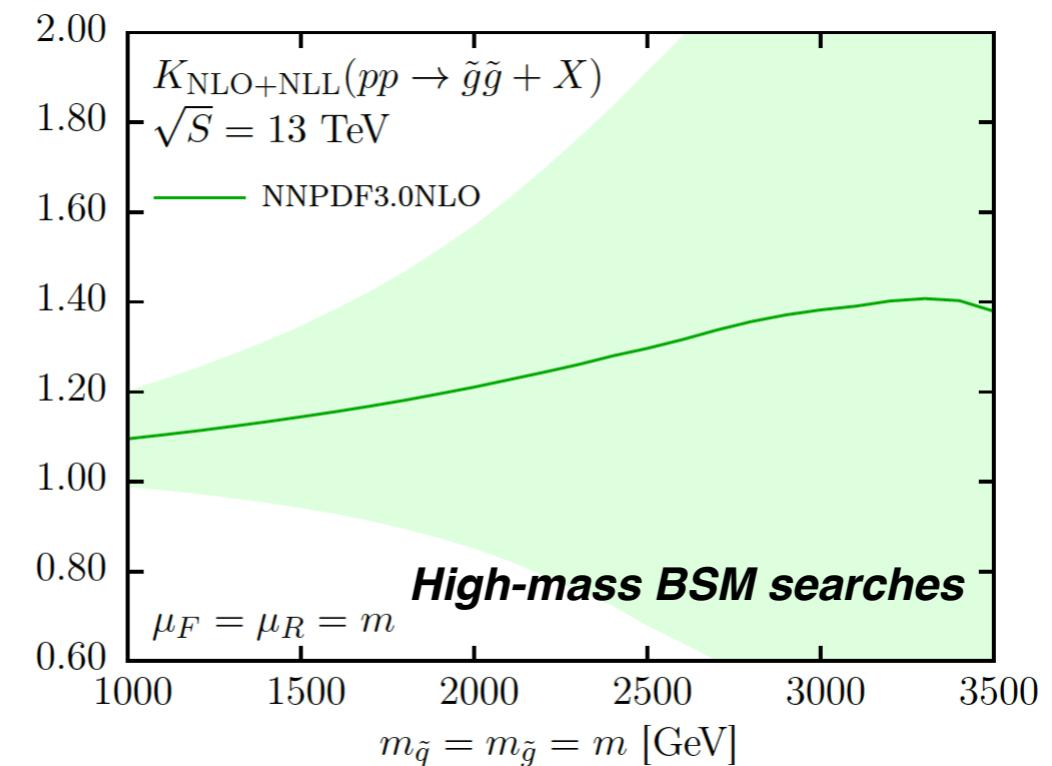
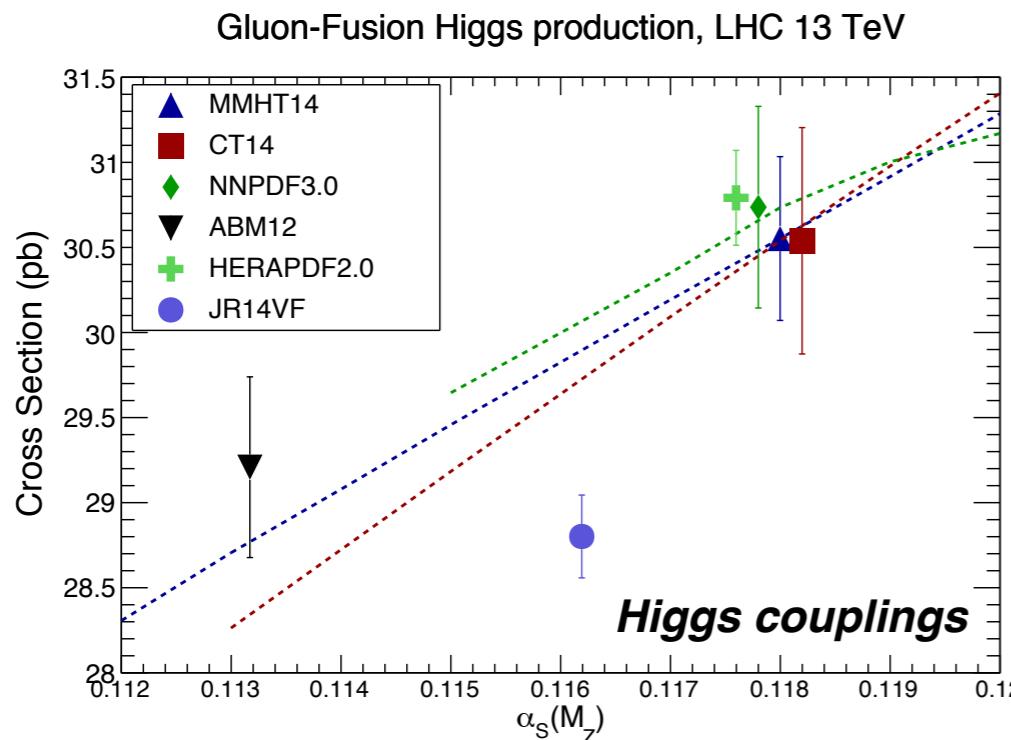
VU Amsterdam & Theory group, Nikhef)

50th International Symposium on Multiparticle Dynamics (ISMD2021)

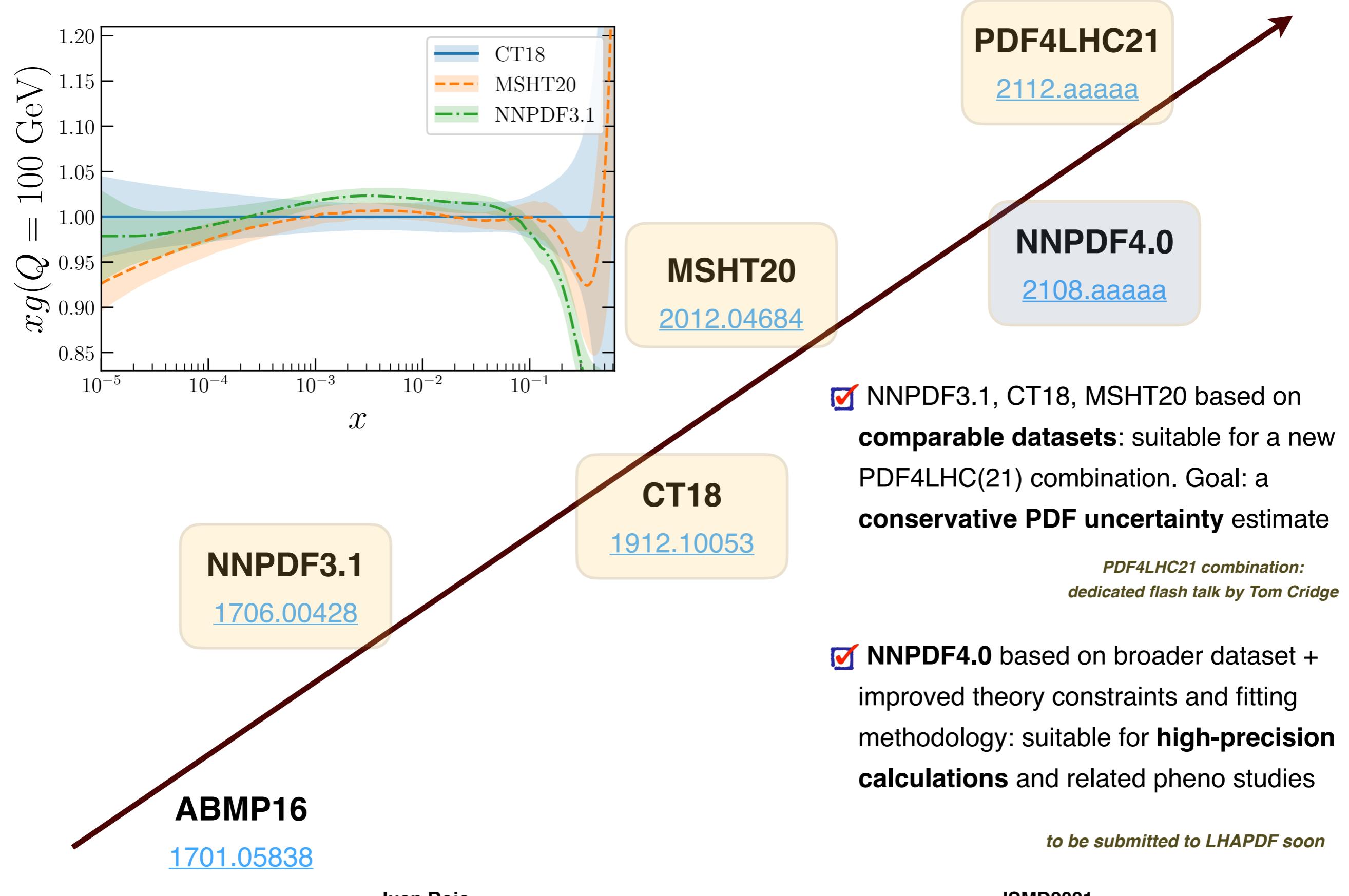


Why we need better PDFs?

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



Recent progress in global fits



Dataset comparison in global PDF fits

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
NMC F_2^d/F_2^p	[30]	✓	✓	✓	✓	✓
NMC $\sigma^{NC,p}$	[31]	✓	✓	✗	✗	✓
SLAC F_2^p, F_2^d	[32]	✓	✓	✗	✓	✓
BCDMS F_2^p, F_2^d	[33]	✓	✓	✓	✓	✓
CHORUS $\sigma_{CC}^\nu, \sigma_{CC}^{\bar{\nu}}$	[34]	✓	✓	✓	✗	✓
NuTeV $\sigma_{CC}^\nu, \sigma_{CC}^{\bar{\nu}}$	[35]	✓	✓	✓	✓	✓
EMC F_2^c	[41]	✗	✓	✗	✗	✗
HERA I+II $\sigma_{NC,CC}^p$	[37]	✓	✓	✓	✓	✓
HERA I+II $\sigma_{c\bar{c}}^{\text{red}}, \sigma_{b\bar{b}}^{\text{red}}$		✗	✓	✗	✓	✓
NOMAD	[114]	✗	✓	✓	✗	✗
CCFR	[35]	✗	✗	✓	✓	✓
DY E866 $\sigma_{DY}^d/\sigma_{DY}^p$ (NuSea)	[44]	✓	✓	✓	✓	✓
DY E866 σ_{DY}^p	[43]	✓	✓	✓	✓	✓
DY E605 σ_{DY}^p	[42]	✓	✓	✓	✓	✗
DY E906 $\sigma_{DY}^d/\sigma_{DY}^p$ (SeaQuest)	[115]	✗	✓	✗	✗	✗
CDF Z rapidity	[207]	✓	✓	✗	✓	✓
CDF k_t single-inclusive jets	[49]	✓	✓	✗	✓	✓
D0 Z rapidity	[47]	✓	✓	✗	✓	✓
D0 $W \rightarrow ee$ asymmetry	[48]	✓	✓	✓	✓	✓
D0 $W \rightarrow \mu\nu$ asymmetry	[47]	✓	✓	✓	✓	✓

Dataset comparison

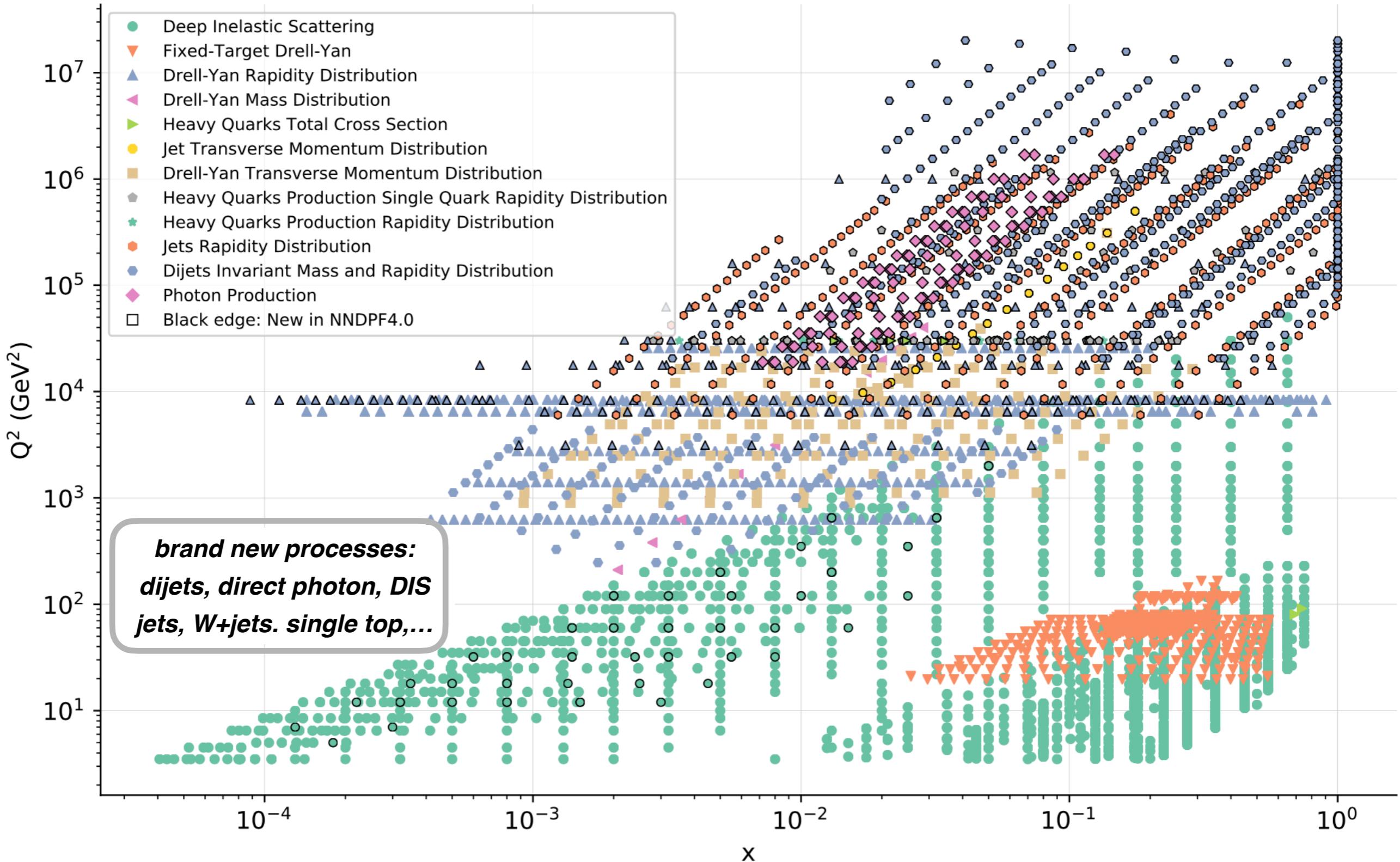
Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV ($\mathcal{L} = 35 \text{ pb}^{-1}$)	[50]	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6 \text{ fb}^{-1}$)	[51]	✓	✓	✗	✗	✓
ATLAS low-mass DY 7 TeV	[52]	✓	✓	✗	✗	✗
ATLAS high-mass DY 7 TeV	[53]	✓	✓	✗	✗	✓
ATLAS W 8 TeV	[79]	✗	✓	✗	✗	✓
ATLAS 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 $Z p_T$ 8 TeV	[62]	✓	✓	✗	✓	✓
ATLAS σ_{tt}^{tot} 7, 8 TeV	[69]	✓	✓	✓	✗	✗
ATLAS σ_{tt}^{tot} 7, 8 TeV	[208–213]	✗	✗	✓	✗	✗
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L} = 3.2 \text{ fb}^{-1}$)	[70]	✓	✗	✓	✗	✗
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L} = 139 \text{ fb}^{-1}$)	[115]	✗	✓	✗	✗	✗
ATLAS $t\bar{t}$ lepton+jets 8 TeV	[71]	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	[89]	✗	✓	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, $R=0.6$	[66]	✓	✗	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, $R=0.6$	[86]	✗	✓	✗	✗	✗
ATLAS dijets 7 TeV, $R=0.6$	[126]	✗	✓	✗	✗	✗
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	[98]	✗	✓	✗	✗	✗

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
CMS W asym. 7 TeV ($\mathcal{L} = 36 \text{ pb}^{-1}$)	[214]	✗	✗	✗	✗	✓
CMS Z 7 TeV ($\mathcal{L} = 36 \text{ pb}^{-1}$)	[215]	✗	✗	✗	✗	✓
CMS W electron asymmetry 7 TeV	[54]	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	[55]	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	[56]	✓	✓	✗	✗	✓
CMS W rapidity 8 TeV	[57]	✓	✓	✓	✓	✓
CMS $Z p_T$ 8 TeV	[63]	✓	✓	✗	✗	✗
CMS $W + c$ 7 TeV	[75]	✓	✓	✗	✗	✓
CMS $W + c$ 13 TeV	[123]	✗	✓	✗	✗	✗
CMS single-inclusive jets 2.76 TeV	[68]	✓	✗	✗	✗	✓
CMS single-inclusive jets 7 TeV	[125]	✓	✗	✗	✓	✓
CMS dijets 7 TeV	[67]	✗	✗	✗	✗	✗
CMS single-inclusive jets 8 TeV	[87]	✗	✓	✗	✓	✓
CMS 3D dijets 8 TeV	[127]	✗	✓	✗	✗	✗
CMS σ_{tt}^{tot} 5 TeV	[88]	✗	✓	✗	✗	✗
CMS σ_{tt}^{tot} 7, 8 TeV	[124]	✓	✓	✗	✗	✗
CMS σ_{tt}^{tot} 8 TeV	[216]	✗	✗	✗	✗	✓
CMS σ_{tt}^{tot} 5, 7, 8, 13 TeV	[72, 217–225]	✗	✗	✓	✗	✗
CMS σ_{tt}^{tot} 13 TeV	[73]	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	[74]	✓	✓	✗	✗	✓
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]	✗	✓	✓	✗	✗

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
LHCb Z 7 TeV ($\mathcal{L} = 940 \text{ pb}^{-1}$)	[58]	✓	✓	✗	✗	✓
LHCb $Z \rightarrow ee$ 8 TeV ($\mathcal{L} = 2 \text{ fb}^{-1}$)	[60]	✓	✓	✓	✓	✓
LHCb W 7 TeV ($\mathcal{L} = 37 \text{ pb}^{-1}$)	[226]	✗	✗	✗	✗	✓
LHCb $W, Z \rightarrow \mu$ 7 TeV	[76]	✓	✓	✓	✓	✓
LHCb $W, Z \rightarrow \mu$ 8 TeV	[61]	✓	✓	✓	✓	✓
LHCb $Z \rightarrow \mu\mu, ee$ 13 TeV	[82]	✗	✓	✗	✗	✗

NNPDF4.0: data set extension

Kinematic coverage



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

From NNPDF3.1 to NNPDF4.0

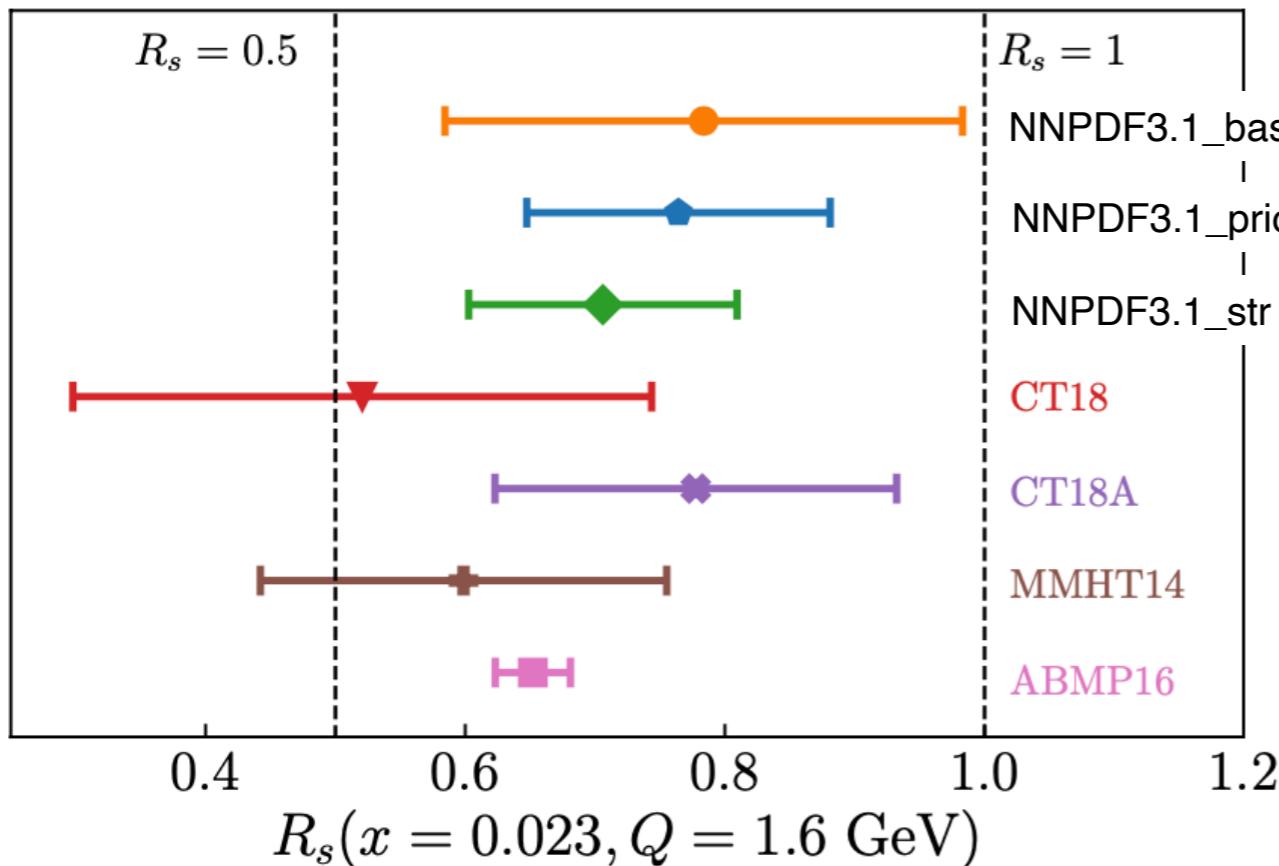
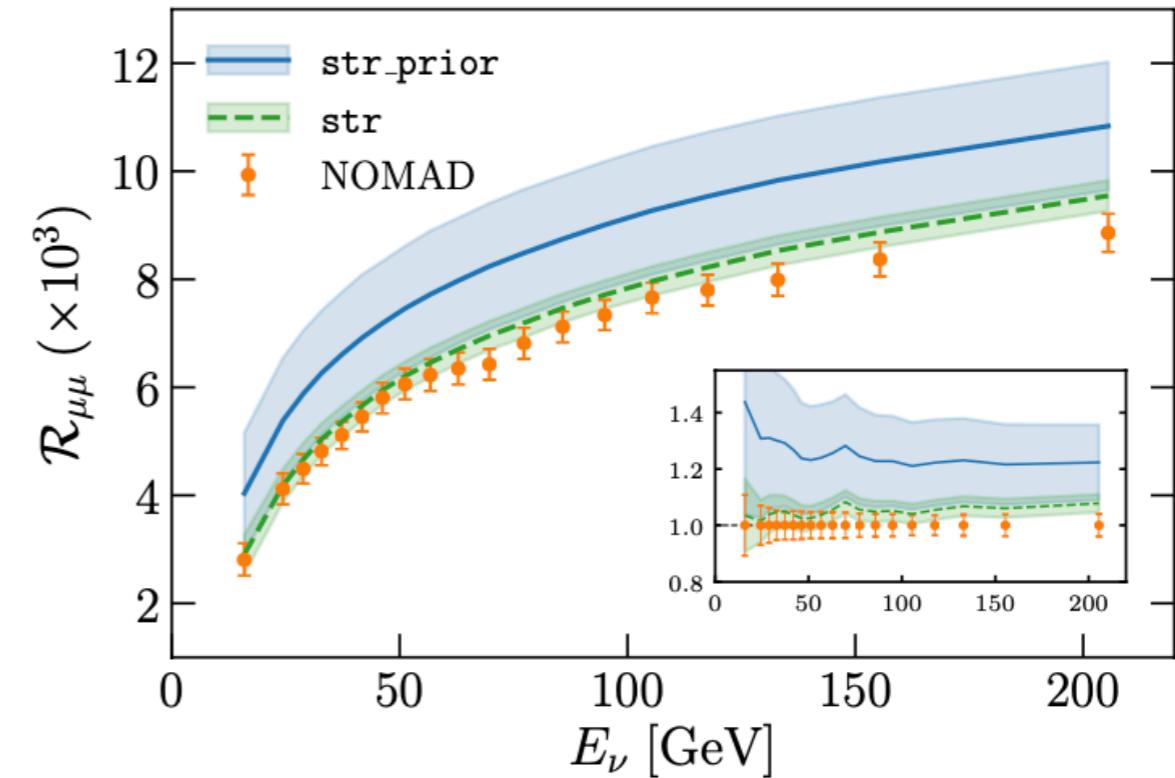
Collaborative progress towards extending **data**, **theory** and **methodology**

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 $\overline{\text{MS}}$ PDFs	[JHEP 11 (2020) 129]
08/2020	PineAPPL : fast evaluation of EW×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**

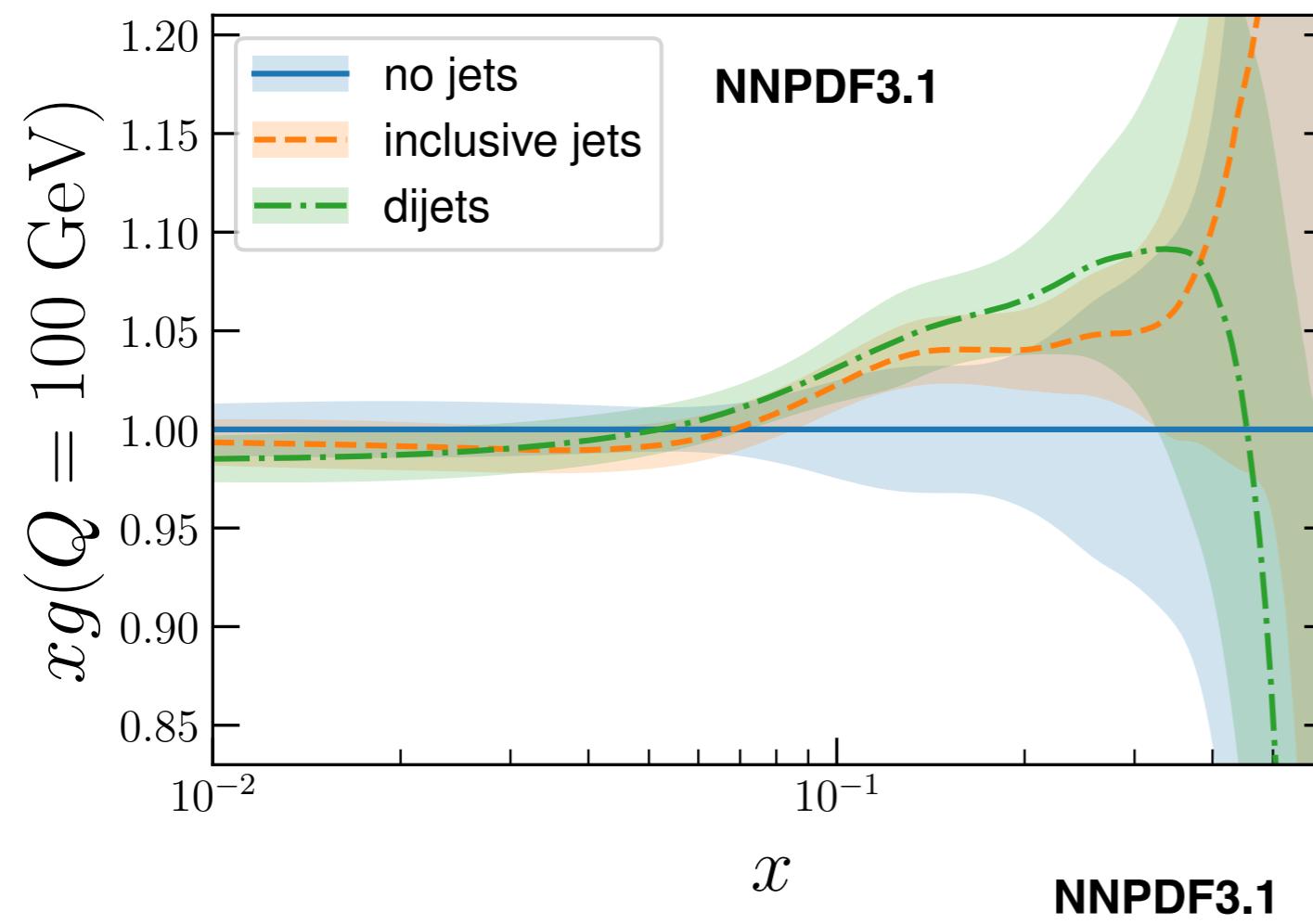
Process	Dataset	n_{dat}	χ^2_{base}	χ^2_{pr}	χ^2_{str}
ν DIS ($\mu\mu$)	NuTeV [9]	76/76/95/91/95	0.76	0.71	0.53
	NOMAD [10]	—/—/19/15/19	[9.3]	[8.8]	0.55
W, Z (incl.)	ATLAS [12]	391/418/418/418/418	1.45	1.40	1.40
		34/61/61/61/61	1.96	1.65	1.67
$W+c$	CMS [17, 18]	—/37/37/37/37	[0.73]	0.68	0.60
	ATLAS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

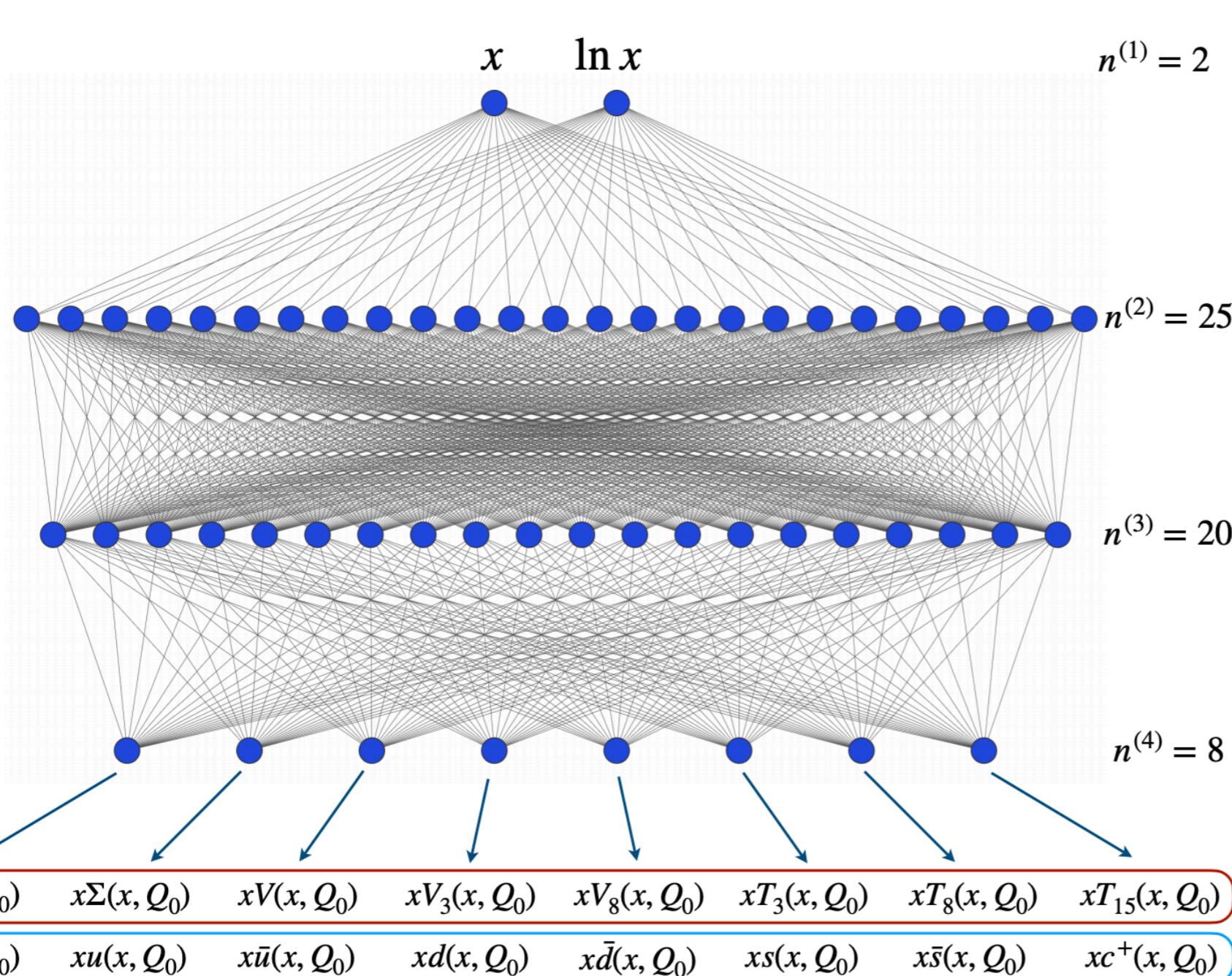
- Constrains on the gluon **qualitatively consistent** between dijet and inclusive jet observables

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

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	$d^3\sigma/dp_{T,\text{avg}} dy_b dy^*$	122

*dijet study revisited
in NNPDF4.0 paper*

Improved fitting methodology in NNPDF4.0



Stochastic Gradient Descent
(via TensorFlow) for NN training

Automated model
hyperparameter optimisation

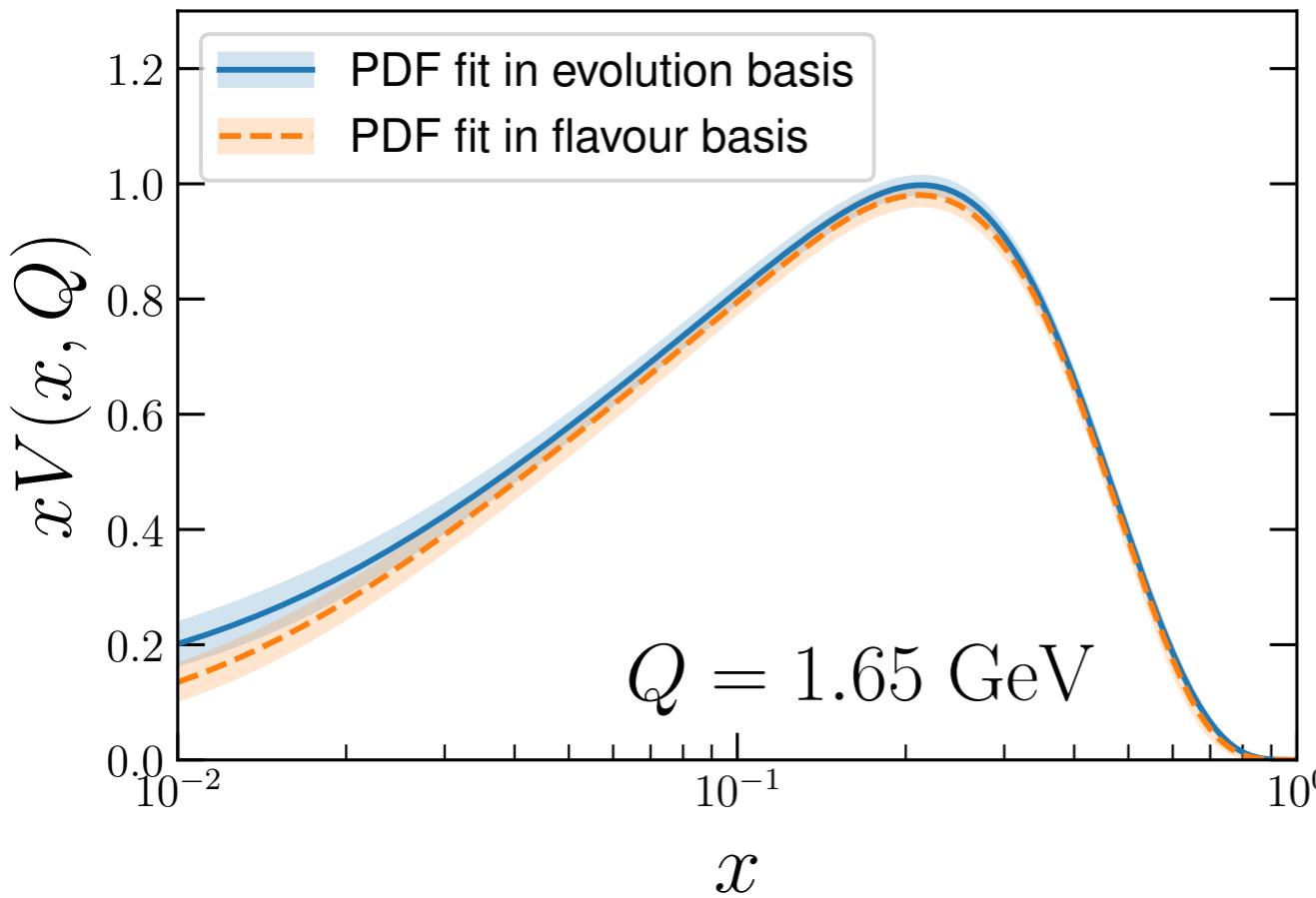
Validation with **future tests**
(forecasting new datasets) and
closure tests (pseudo-data
based on known PDFs)

Generative Adversarial
Networks for replica
compression

evolution basis
flavor basis

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

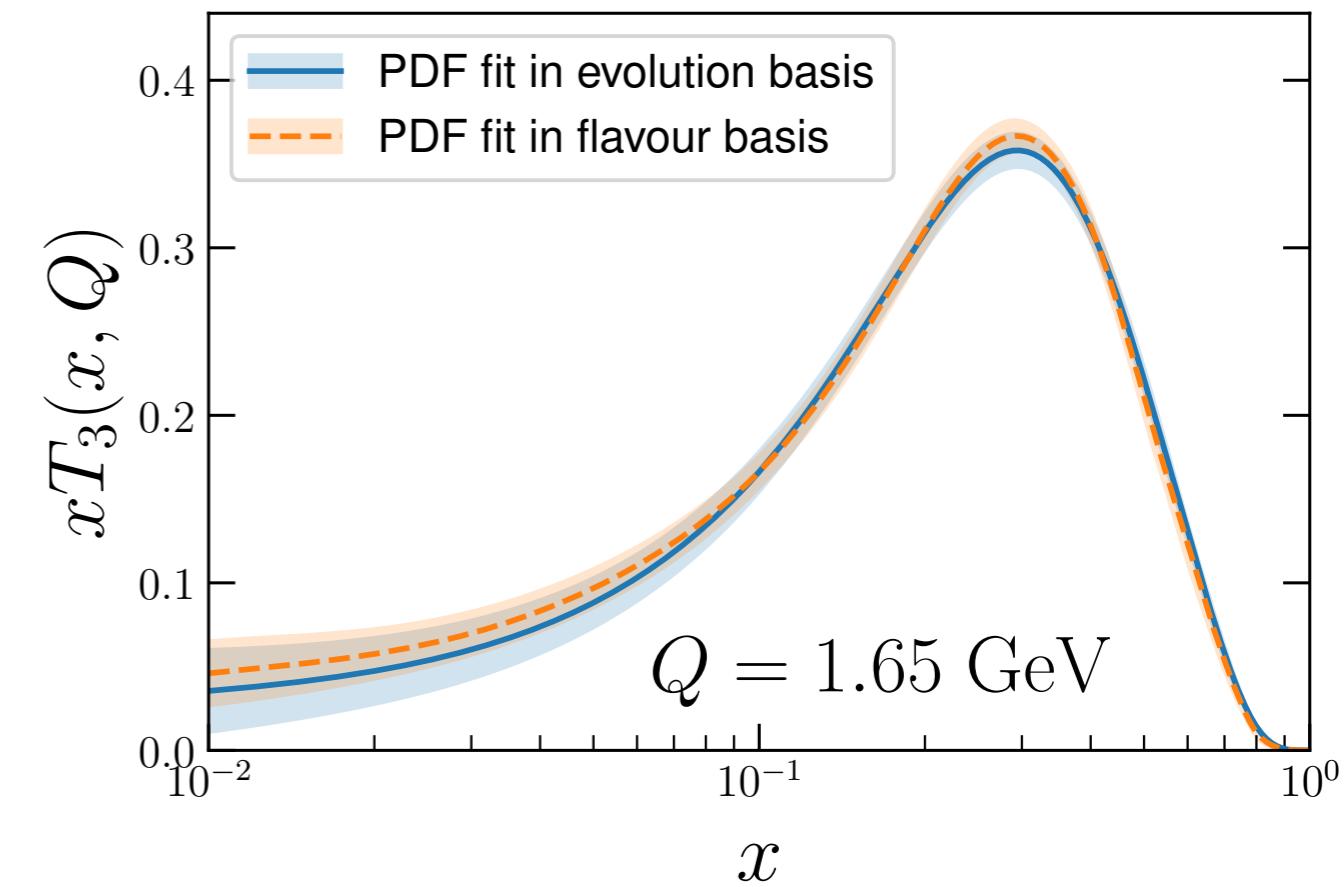
Parametrisation basis independence



evolution basis PDF parametrisation:

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

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



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

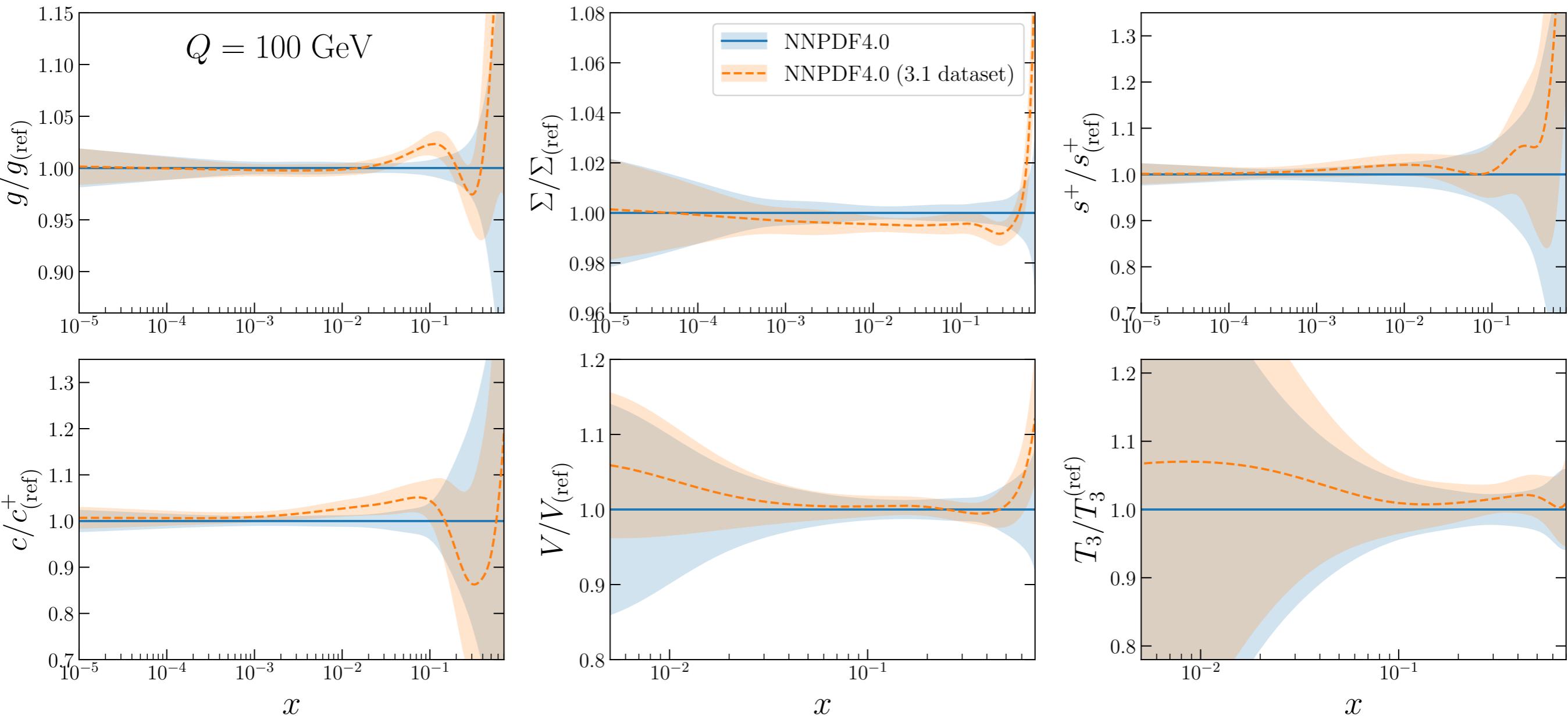
flavour basis PDF parametrisation:

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

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

Impact of new data in NNPDF4.0

Good consistency with NNPDF3.1-like dataset

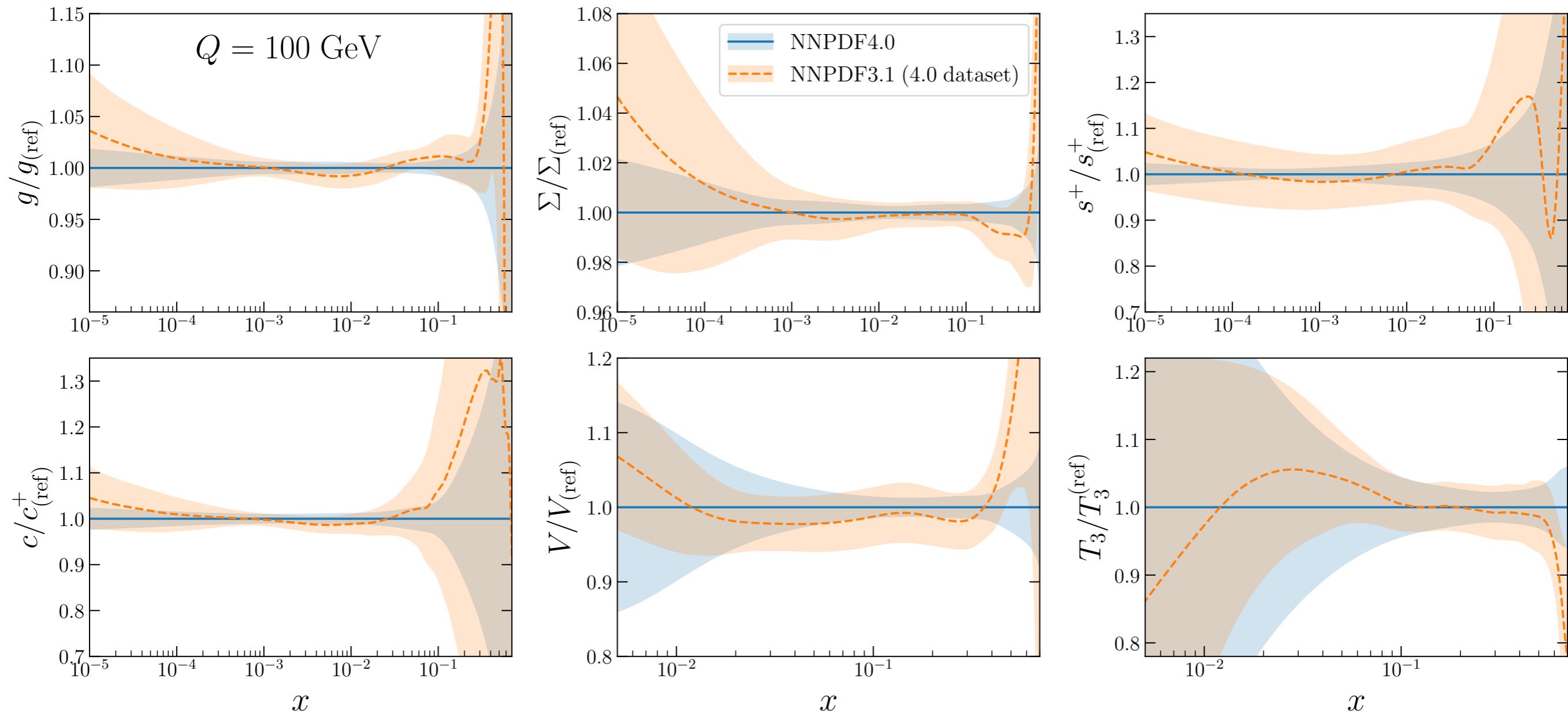


Moderate reduction of PDF uncertainties, shifts in central values at the one-sigma level

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

Impact of new fitting methodology

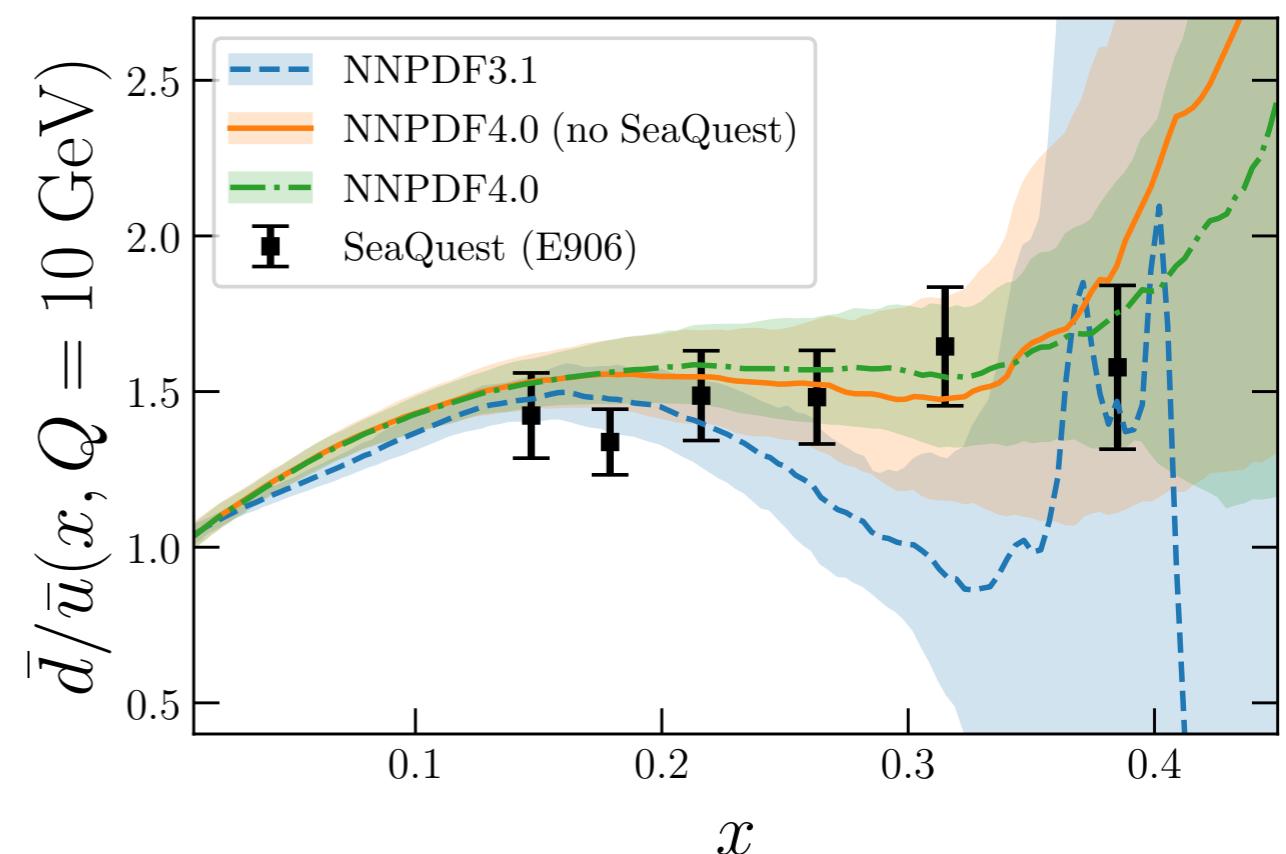
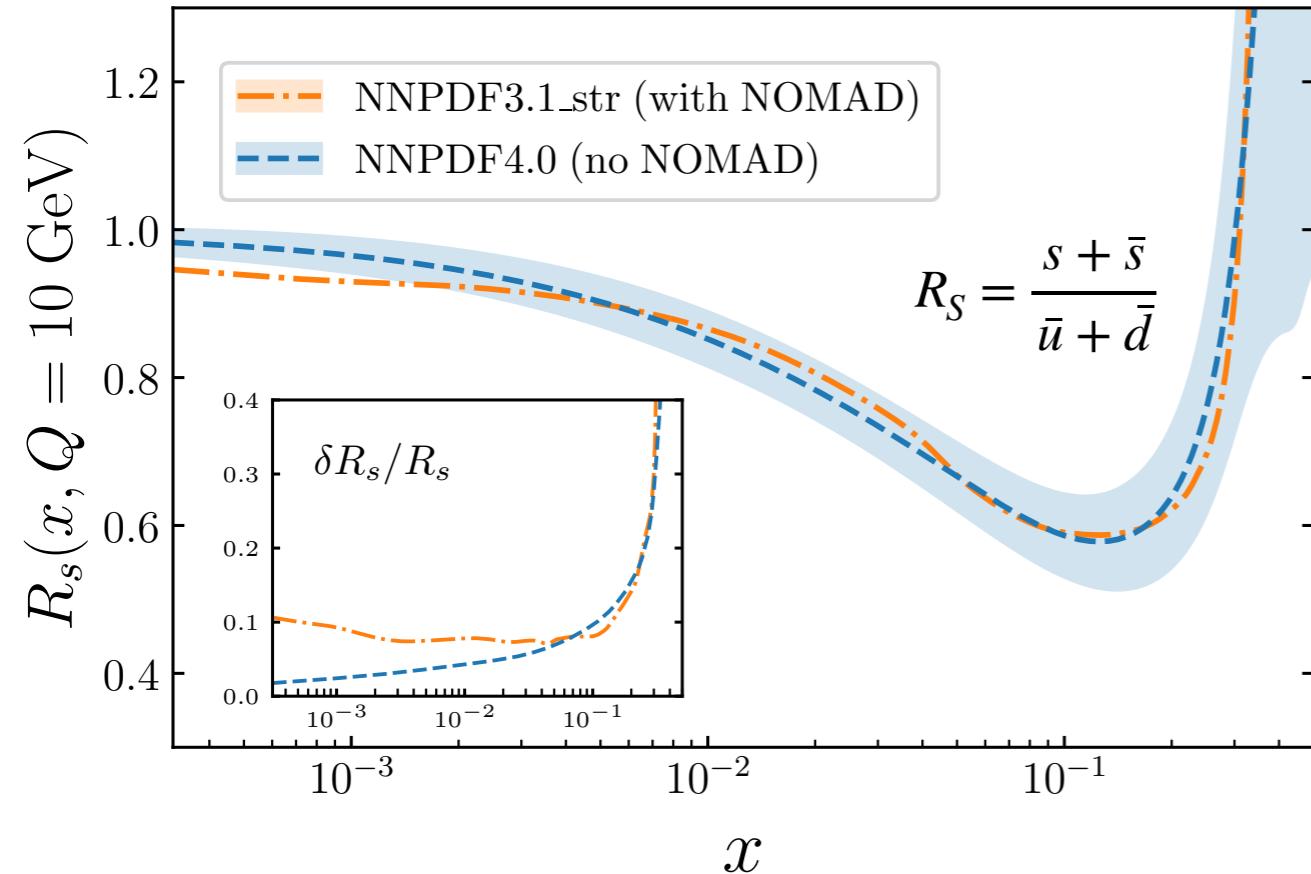
Compare fits based on the **same NNPDF4.0 dataset** but either with the **previous 3.1** or the **new 4.0 methodology**: significant reduction of uncertainties + consistency



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

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

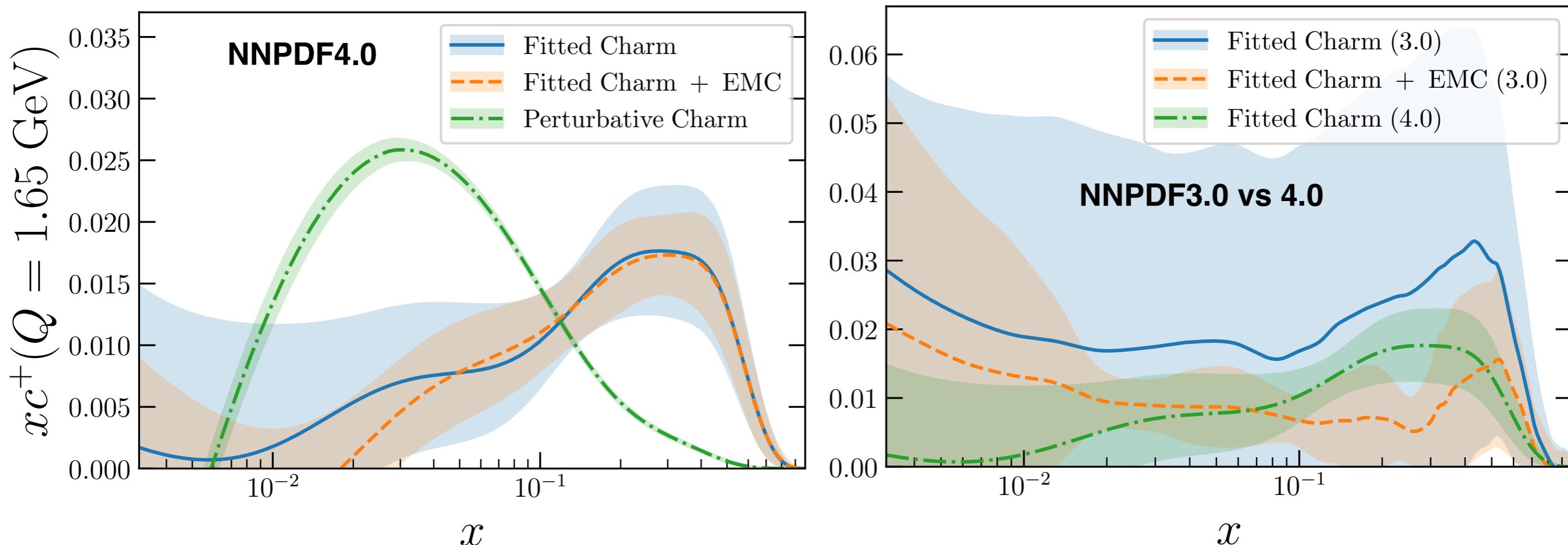
From antimatter asymmetry to intrinsic charm



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 **T₃ and T₈ 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

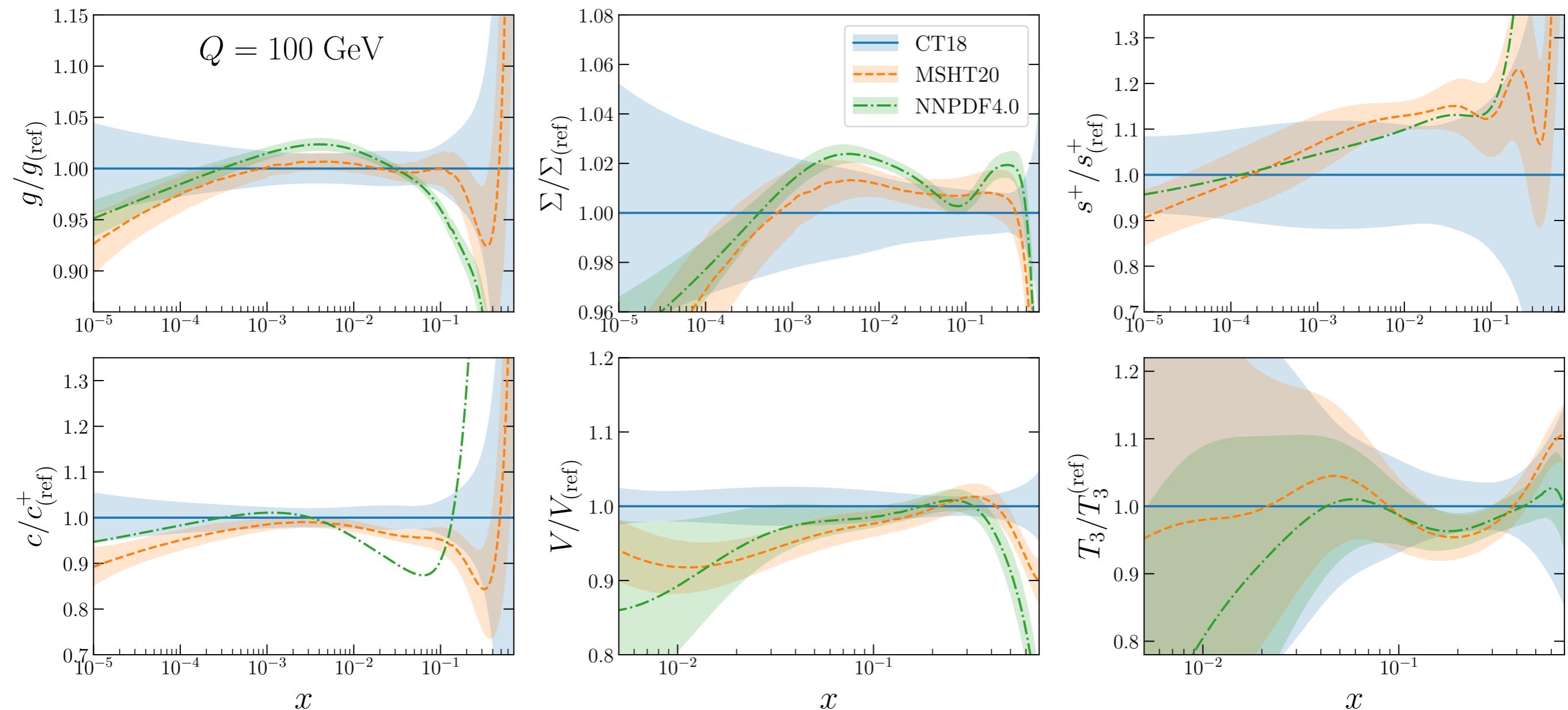
From antimatter asymmetry to intrinsic charm



- 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

Comparison between global fits

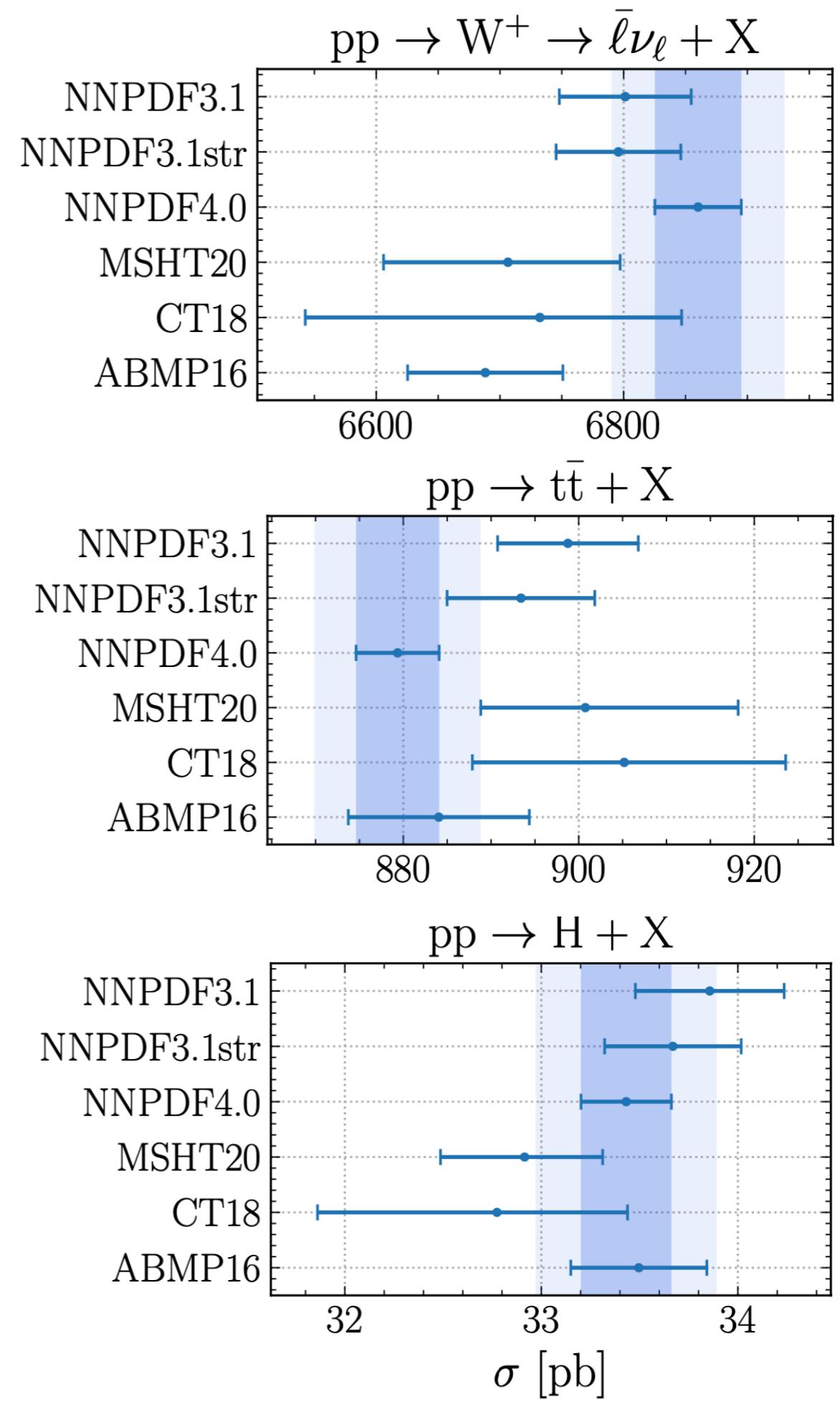
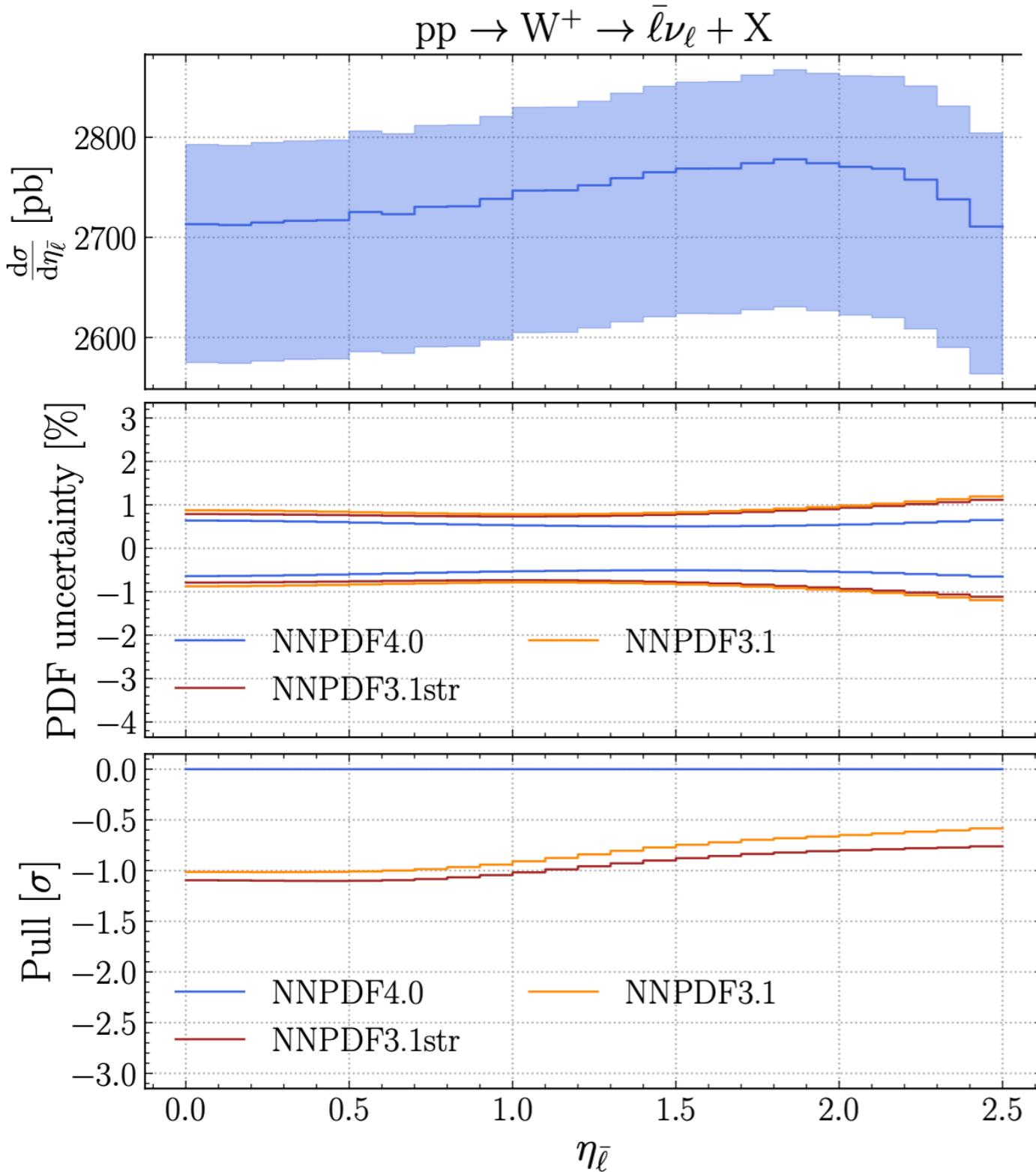
In general good agreement between NNPDF4.0, CT18, and MSHT20, though rather **different pattern of PDF uncertainties**



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**

Summary and outlook



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

[Home](#) » 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 public 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 **NNPDF machine learning fitting framework** will be publicly released **open source**, together with extensive documentation and user-friendly examples, at same time as upcoming paper