



The NNPDF4.0 global analysis of proton structure

Juan Rojo, VU Amsterdam & Nikhef

ATLAS Standard Model Plenary Meeting

7th October 2021



NNPDF4.0: Setup & Validation

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 <i>t</i> -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×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!

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

Tevatron



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)



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)

 $\begin{array}{l} \text{Loss (``average'')}\\ \text{ML model}\\ \text{hyperparams} \ \boldsymbol{\hat{\theta}} = \operatorname*{arg min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(\frac{1}{n_{\mathrm{fold}}} \sum_{k=1}^{n_{\mathrm{fold}}} \chi_k^2(\boldsymbol{\theta}) \right) \end{array}$

Loss (``max")

$$L = \max\left(\chi_1^2, \chi_2^2, \chi_3^2, \dots, \chi_{n_{ ext{fold}}}^2
ight)$$





Stability wrt hyperopt loss function

Improved fitting methodology

epoch 3



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



Future tests

Fit data restricted to specific kinematic regions,

then verify succesful extrapolation



	70 F = = = = = = = = = = = = = = = = = = =	<i>7</i> C F = = = = F	70
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



$$xV(x, Q_0) \propto NN_V(x)$$

 $xT_3(x, Q_0) \propto NN_{T_3}(x)$

flavour basis PDF parametrisation:

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

first time ever!

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

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

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



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

* The NNPDF collaboration

Search docs

Getting started

Buildmaster

Theory

Servers

Tutorials

External codes

Fitting code: n3fit

Code for data: validphys Handling experimental data:

Storage of data and theory predictions

Continuous integration and deployment

Adding to the Documentation

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!

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)

Comparison with NNPDF3.1



Antimatter asymmetry



Mark SeaQuest measurement claims evidence for quark sea (``proton antimatter'') asymmetry

$$\frac{\sigma_{\rm DY,deuterium}}{\sigma_{\rm DY,hydrogen}} \approx 1 + \frac{\bar{d}_p(x_t)}{\bar{u}_p(x_t)} \qquad \text{with many caveats!}$$

Actually, SeaQuest further confirms the global fit prediction, which agrees with it even when not included

Already well described by NNPDF3.1 within uncertainties

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)

Mathematical States and States a

✓ As opposed to previous studies, impact of the EMC charm measurements mild now. Information provided by EMC F₂^c consistent with latest collider data

Intrinsic charm



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

 $R_{s} = 0.5$

 $R_S \equiv \frac{s + \bar{s}}{\bar{u} + \bar{d}}$

 $R_s = 1$

NNPDF4.0 (w. NOMAD)

NNPDF4.0

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

1.15

 $1.10 \cdot$

1.05

1.00

0.95

0.90

0.85

Ratio to NNPDF4.0



Impact of decorrelation models



 \checkmark 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 x²

Dataset	$N_{ m dat}$	$Z_{ m orig}$	$\chi^2_{ m orig}$	$\chi^2_{ m reg}$
ATLAS W, Z 7 TeV CC ($\mathcal{L} = 4.6 \text{ fb}^{-1}$)	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

Phenomenology

Comparison between global fits

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



Comparison between global fits



Comparison between global fits

... follows pattern of input datasets

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

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20	Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV ($\mathcal{L} = 35 \text{ pb}^{-1}$)	[51]	1	1	1	1	1	CMS W asym. 7 TeV ($\mathcal{L} = 36 \text{ pb}^{-1}$)	[267]	×	×	×	×	1
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6 \text{ fb}^{-1}$)	[52]	1	1	×	(🗸)	1	CMS Z 7 TeV ($\mathcal{L} = 36 \text{ pb}^{-1}$)	[268]	×	×	×	×	1
ATLAS low-mass DY 7 TeV	[53]	1	1	×	(🗸)	×	CMS W electron asymmetry 7 TeV	[55]	1	1	×	1	1
ATLAS high-mass DY 7 TeV	[54]	1	1	×	(🗸)	1	CMS W muon asymmetry 7 TeV	[56]	1	1	1	1	×
ATLAS W 8 TeV	[79]	×	(✔)	×	×	1	CMS Drell-Yan 2D 7 TeV	[57]	1	1	×	(✔)	1
ATLAS DY 2D 8 TeV	[78]	×	1	×	×	1	CMS Drell-Yan 2D 8 TeV	[269]	(✔)	×	×	×	×
ATLAS high-mass DY 2D 8 TeV	[77]	×	1	×	(🗸)	1	CMS W rapidity 8 TeV	[58]	1	1	1	1	1
ATLAS $\sigma_{W,Z}$ 13 TeV	[81]	×	1	1	×	×	CMS $W, Z \ p_T$ 8 TeV ($\mathcal{L} = 18.4 \ \text{fb}^{-1}$)	[270]	×	×	×	(✔)	×
ATLAS W +jet 8 TeV	[93]	×	1	×	×	1	CMS $Z p_T$ 8 TeV	[64]	1	1	×	(✔)	×
ATLAS $Z p_T$ 7 TeV	[259]	(✔)	×	×	(🗸)	×	CMS $W + c$ 7 TeV	[76]	1	1	×	(✔)	1
ATLAS $Z p_T 8$ TeV	[63]	1	1	×	1	1	$\mathrm{CMS}\ W + c\ 13\ \mathrm{TeV}$	[84]	×	1	×	×	(✔)
ATLAS $W + c$ 7 TeV	[83]	×	1	×	(🗸)	×	CMS single-inclusive jets 2.76 TeV	[75]		×	×	×	1
ATLAS σ_{tt}^{tot} 7, 8 TeV	[65]	1	1	1	×	×	CMS single-inclusive jets 7 TeV	[147]	1	(•)	×	1	×
ATLAS σ_{tt}^{tot} 7, 8 TeV	[260-265]	×	×	1	×	×	CMS dijets 7 TeV	[74]	×		×	×	×
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L} = 3.2 \text{ fb}^{-1}$)	[66]	1	×	1	×	×	CMS single-inclusive jets 8 TeV	[87]	×		×	~	v
ATLAS σ_{tt}^{tot} 13 TeV ($\mathcal{L} = 139 \text{ fb}^{-1}$)	[134]	×	1	×	×	×	CMS 3D dijets 8 TeV	[149]	×	(•)	×	×	×
ATLAS σ_{tot}^{tot} and Z ratios	[266]	×	×	×	×	(🗸)	CMS σ_{ii}^{tot} 5 rev	[00]	<u>^</u>		<u></u>	Û.	<u></u>
ATLAS $t\bar{t}$ lepton+jets 8 TeV	[67]	1	1	×	1	1	$CMS = tot + T_{A}$	[140]	~	~	<u></u>	Û	<u></u>
ATLAS $t\bar{t}$ dilepton 8 TeV	[89]	×	1	×	×	1	$CMS \sigma_{tt}^{tot} 5 7 8 13 \text{ TeV}$	[271] [68-272_280]	<u>,</u>	Ŷ.		<u> </u>	×
ATLAS single-inclusive jets 7 TeV, R=0.6	[73]	1	(🗸)	×	1	1	$CMS \sigma_{tt}^{tot} 13 \text{ TeV}$	[60]				Ŷ	x
ATLAS single-inclusive jets 8 TeV, R=0.6	[86]	×	1	×	×	×	$CMS t\bar{t}$ lepton+jets 8 TeV	[70]		,	×	x	
ATLAS dijets 7 TeV, R=0.6	[148]	×	1	×	×	×	$CMS \ t\bar{t} \ 2D \ dilepton \ 8 \ TeV$	[90]	×	1	×	1	1
ATLAS direct photon production 8 TeV	[100]	×	(🗸)	×	×	×	$CMS t\bar{t}$ lepton+iet 13 TeV	[91]	×	1	×	×	×
ATLAS direct photon production 13 TeV	[101]	×		×	×	×	CMS $t\bar{t}$ dilepton 13 TeV	[92]	×	1	×	×	×
ATLAS single top R_t 7, 8, 13 TeV	[94, 96, 98]	×	1	1	×	×	CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	[95]	×	1	1	×	×
ATLAS single top diff. 7 TeV	[94]	×	1	×	×	×	CMS single top R_t 8, 13 TeV	[97, 99]	×	1	1	×	×
ATLAS single top diff. 8 TeV	[<mark>96</mark>]	×	1	×	×	×	CMS single top 13 TeV	[281, 282]	×	×	×	×	(•

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
LHCb Z 7 TeV ($\mathcal{L} = 940 \text{ pb}^{-1}$)	[59]	1	1	×	×	1
LHCb $Z \rightarrow ee \ 8 \ \text{TeV} \ (\mathcal{L} = 2 \ \text{fb}^{-1})$	[<mark>61</mark>]	1	1	1	1	1
LHCb W 7 TeV ($\mathcal{L} = 37 \text{ pb}^{-1}$)	[283]	×	×	×	×	1
LHC b $W,Z \to \mu$ 7 TeV	[<mark>60</mark>]	 Image: A second s	1	1	1	1
LHC b $W,Z \to \mu$ 8 TeV	[<mark>62</mark>]	1	1	1	1	1
LHCb $W \to e \ 8 \ {\rm TeV}$	[80]	×	(✔)	×	×	×
LHC b $Z \to \mu \mu, ee$ 13 TeV	[82]	×	1	×	×	×



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 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

Why Nucleon Structure?



Address fundamental questions about the strong nuclear force

Provide predictions for particle, nuclear, and astro-particle experiments

origin of mass and spin in QCD, heavy quark content, 3D imaging, gluon-dominated matter, nuclear modifications

LHC, Electron Ion Collider, RHIC, IceCube, AMBER, COMPASS, Auger, ...

Why Nucleon Structure?

THE SCIENCE

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 @ SeaQuest





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 for precision LHC physics

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





PDFs for precision LHC physics

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



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!