nNNPDF updates

XXXII Cracow Epiphany Conference on the recent results from Heavy Ion Physics

Emanuele R. Nocera

Università degli Studi di Torino and INFN — Torino

Kraków, Poland - 13 January 2026

The NNPDF methodology in a nutshell

Assumption:
fundamental interactions are the same in the vacuum and in the medium,
but PDFs are different, *i.e.* nuclear effects are reabsorbed into nPDFs

Generate $N_{\rm rep}$ Monte Carlo replicas of the data
assuming a multi-Gaussian distribution
and taking into account experimental correlations

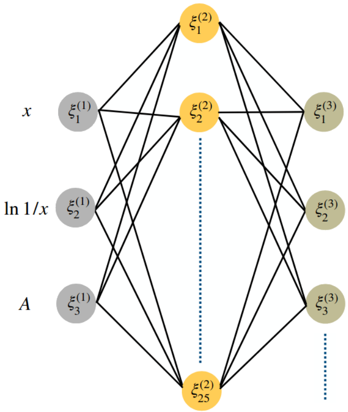
$$D_i^{(k)} = D_i^{(0)} + r_i^{(k)}\sigma_i^{(\rm unc)} + r^{(k)}\sigma_i^{(\rm cor)}$$

Fit a set of nuclear PDFs parametrised at a scale Q_0
with a Neural Network to each replica
impose theoretical constraints
(sum rules, positivity, proton baseline)

$$xf^{p/A}(x,Q_0) = B_f x^{\alpha_f}(1-x)^{\beta_f}\mathrm{NN}_f(x,A)$$

Compute expectation values and variances as

$$E[\mathcal{O}] \approx \frac{1}{N}\sum_k \mathcal{O}(f_k)$$
$$V[\mathcal{O}] \approx \frac{1}{N}\sum_k [\mathcal{O}(f_k) - E[\mathcal{O}]]^2$$

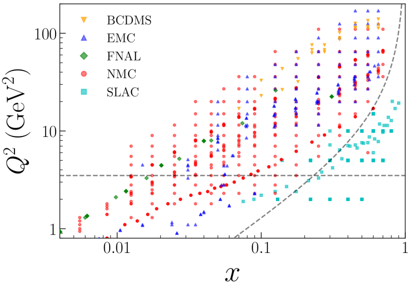


nNNPDF1.0 and nNNPDF2.0

nNNPDF1.0 [EPJ C79 (2019) 6]
Neutral Current DIS $\ell + A/\ell + A'$
SLAC, BCDMS, NMC, EMC, FNAL

2 PDF combinations: Σ, g
no positivity imposed
free proton boundary condition from NNPDF3.1
NLO and NNLO

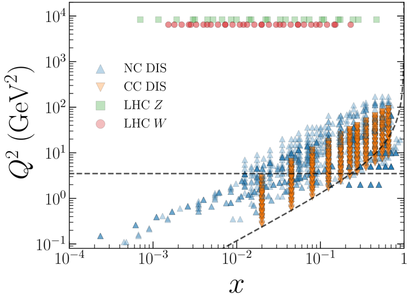
single feed-forward NN, A dependence as input



nNNPDF2.0 [JHEP 09 (2020) 183]
Charged Current DIS $\nu + A$
Chorus, NuTeV
LHC W, Z production $p + Pb$
CMS 5 & 8 TeV, ATLAS 5 TeV

6 PDF combinations: Σ, g, V, V_3, T_3, T_8
positivity imposed
free proton boundary condition from NNPDF3.1
NLO only

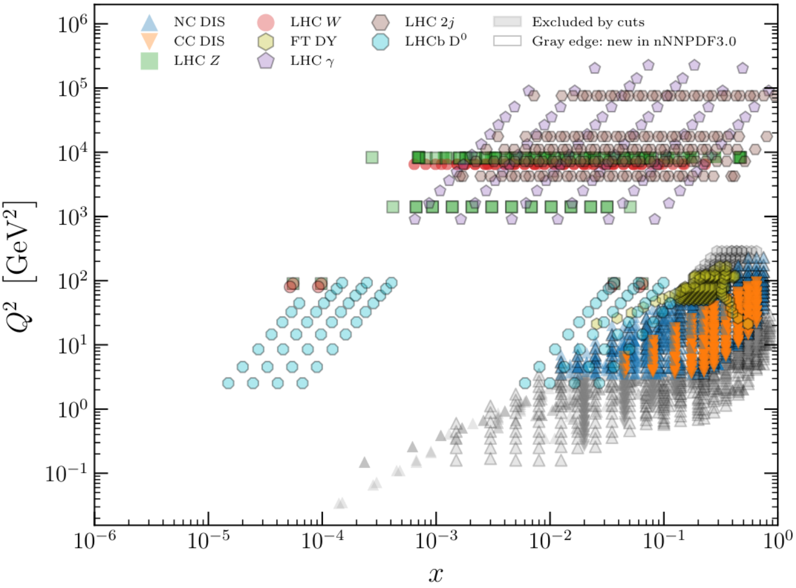
single feed-forward NN, A dependence as input



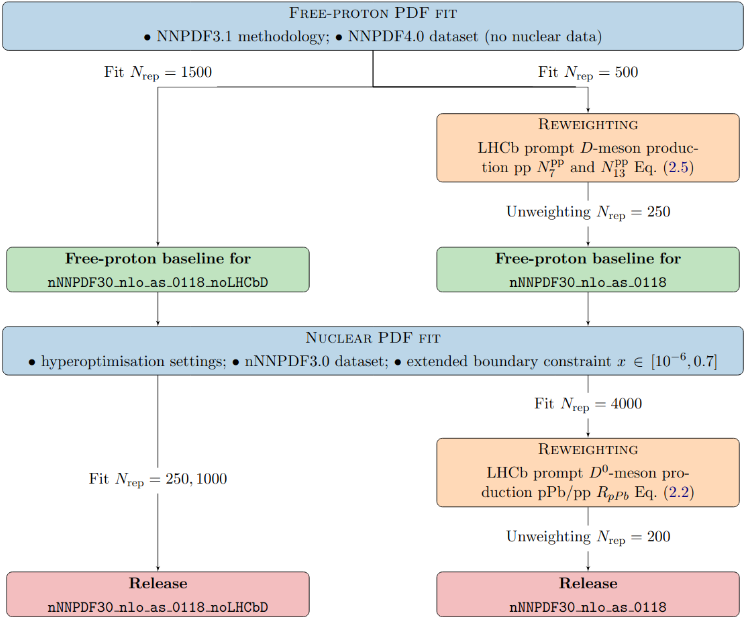
1. The nNNPDF3.0 parton set

[Eur. Phys. J. **C82** (2022) 507]

Input data

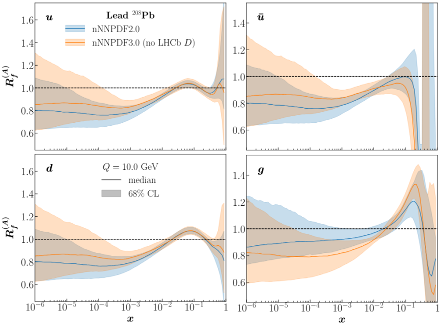

NC DIS
CC DIS
LHC Z
LHC W
FT DY
LHC γ
LHC $2j$
LHCb D^0
Excluded by cuts
Gray edge: new in nNNPDF3.0
Q^2 [GeV^2]
x
10^6
10^5
10^4
10^3
10^2
10^1
10^0
10^{-1}
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^0

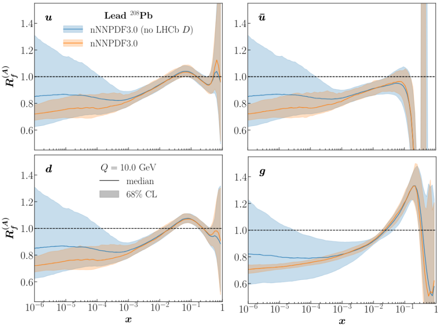
Strategy


Free-proton PDF fit
• NNPDF3.1 methodology; • NNPDF4.0 dataset (no nuclear data)
Fit $N_{\rm rep} = 1500$
Fit $N_{\rm rep} = 500$
Reweighting
LHCb prompt D-meson production pp $N_7^{\rm pp}$ and $N_{13}^{\rm pp}$ Eq. (2.5)
Unweighting $N_{\rm rep} = 250$
Free-proton baseline for
nNNPDF30_nlo_as_0118_noLHCbD
Free-proton baseline for
nNNPDF30_nlo_as_0118
Nuclear PDF fit
• hyperoptimisation settings; • nNNPDF3.0 dataset; • extended boundary constraint $x \in [10^{-6}, 0.7]$
Fit $N_{\rm rep} = 4000$
Reweighting
LHCb prompt D^0-meson production pPb/pp R_{pPb} Eq. (2.2)
Fit $N_{\rm rep} = 250, 1000$
Unweighting $N_{\rm rep} = 200$
Release
nNNPDF30_nlo_as_0118_noLHCbD
Release
nNNPDF30_nlo_as_0118

Overall fit quality and impact

Process	Data sets	N_{dat}	no LHCb D^0	nNNPDF3.0	
NC DIS	SLAC, NMC, BCDMS, EMC (various nuclei)	862	1.05	1.03	APFEL
CC DIS	NuTeV Fe, CHORUS Pb	922	1.04	1.01	AFPEL
DY NC and CC	FNAL Cu, d, ATLAS, CMS, LHCb Pb 5.02 and 8.16 TeV	240	1.08	1.11	MCFM
Di-jets	CMS pPb/pp 5.02 TeV	84	1.81	1.75	NLOjet++
Prompt photon	ATLAS pPb/pp 8.16 TeV	43	1.03	1.04	MCFM
Total		2151	1.11	1.09	
Forward D^0	LHCb pPb/pp 5.02 TeV	37	—	0.66	POWHEG

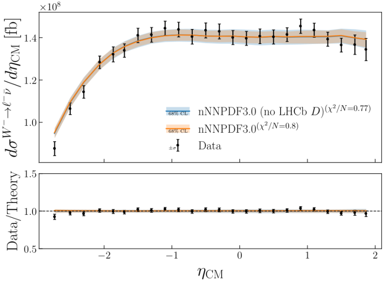




Impact of new data

W bosons in pPb at 8.16 TeV

CMS, Run II [PLB 800 (2020) 135048]



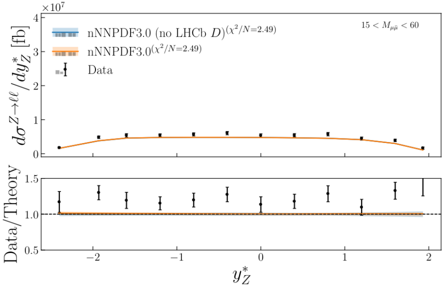
sensitivity to quark flavour separation and to the gluon through quark-gluon correlations at small x and high Q

nCTEQ15WZSIH, TUJU21 and nNNPDF3.0 fit absolute cross section, EPPS21 fit nuclear modification ratios

All find a good description of the data set

Z bosons in pPb at 8.16 TeV

CMS, Run II [JHEP 05 (2021) 182]



nNNPDF3.0 include both low-mass and on-peak data pp/pPb studied in EPPS21 but not included

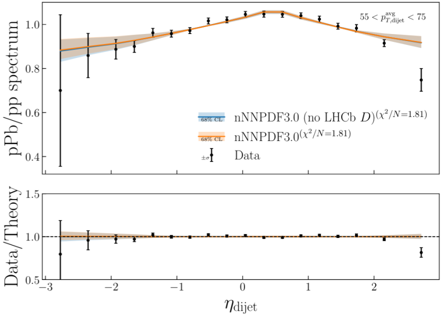
Both EPPS21 and nNNPDF3.0 observe data/theory tensions: shift over rapidity

NNLO to cure for the low-mass data?

Impact of new data

Dijets in pPb/pp at 5.02 TeV

CMS [PRL 121 (2018) 062002]

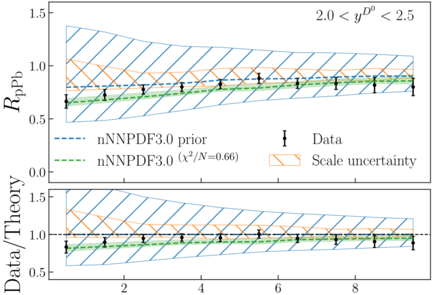


Drastic reduction of nPDF uncertainties
Important constraint for the gluon nPDF

Both EPPS21 and nNNPDF3.0
do not fit the most forward data points
missing data correlations?
NNLO? Nonperturbative effects?

D^0 mesons in pPb/pp at 5.02 TeV

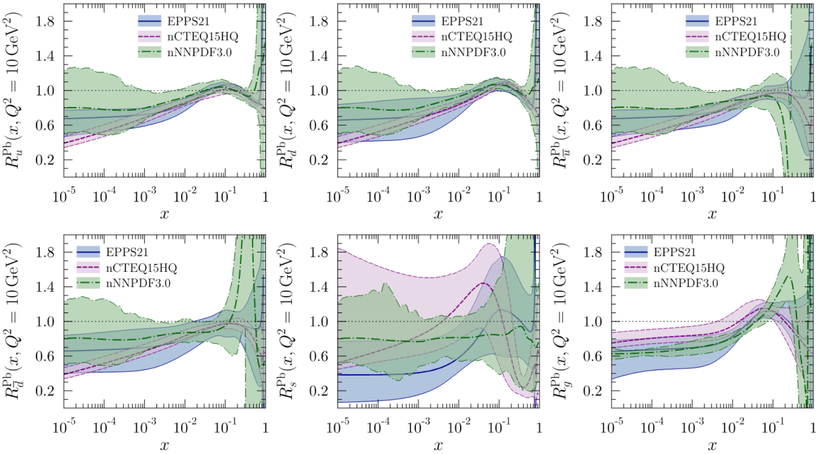
LHCb [JHEP 10 (2017) 090]



Drastic reduction of nPDF uncertainties
Important constraint for the gluon nPDF

nNNPDF3.0: POWHEG+PYTHIA
large scale uncertainty, only forward data
EPPS21: S-ACOT-M_T GM-VFNS
large scale uncertainties not seen

Comparison to other nuclear PDF sets



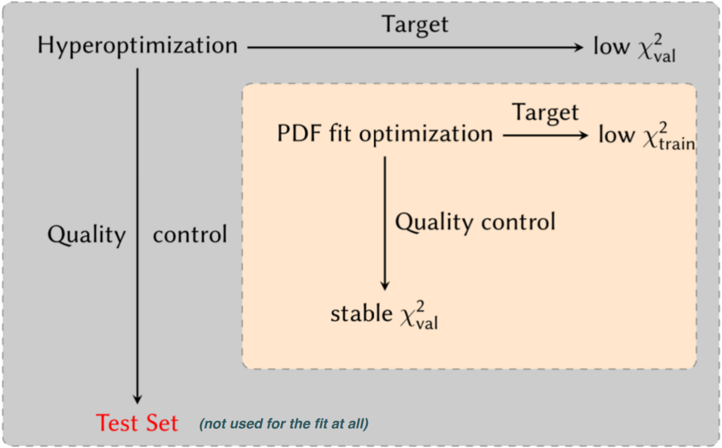
Evidence of shadowing and anti-shadowing in all partons

Evidence of EMC effect in EPPS21 for u and d

Overall qualitative agreement across PDF sets, except for nCTEQ strange

Difference in uncertainties depends on differences in methodologies and in data sets

Hyperoptimisation: fitting the methodology



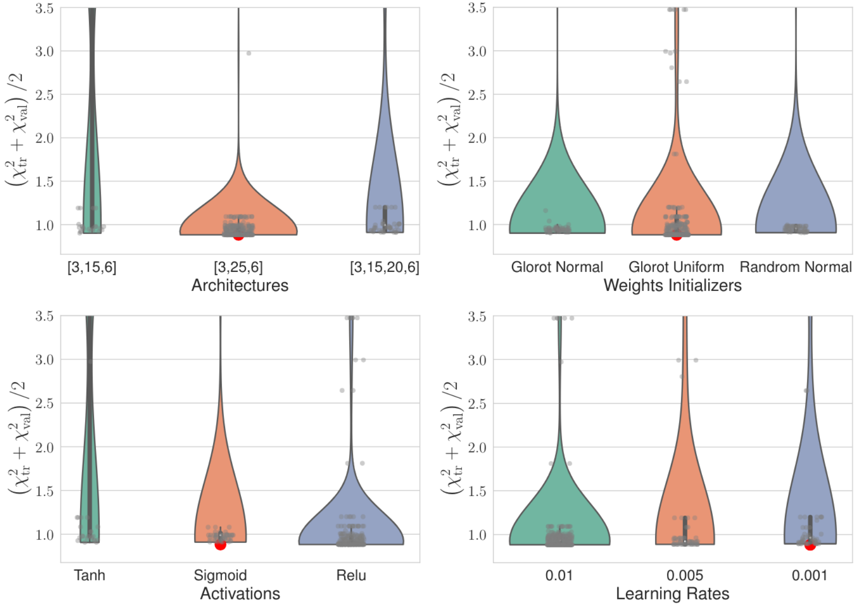
Compare to a Test Set (new set of data previously not used at all)

Who picks the Test Set? Automatic generalisation based on K folds

Divide the data into n representative sets, fit $n-1$ sets and use the n-th set as test set

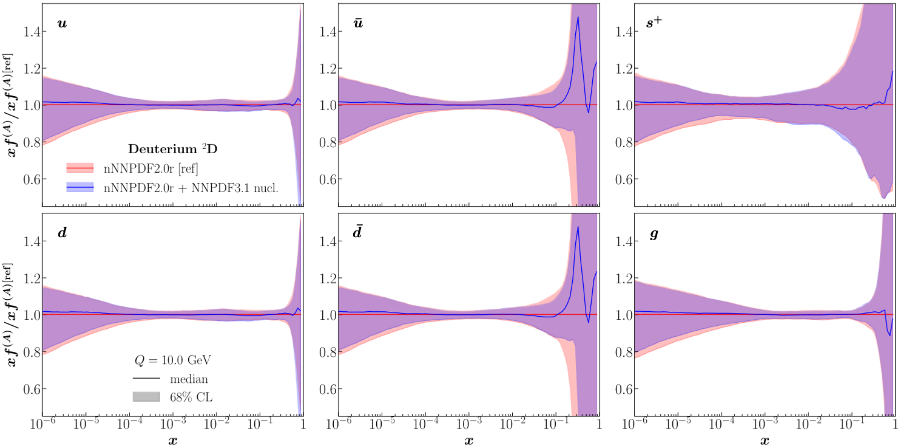
Hyperoptimise on mean and standard deviation of $\chi^2_{\text{test},i}$, over folds $i = 1 \ldots n$

Hyperoptimisation: results



Impact of proton baseline fit

What's the impact of varying the proton baseline fit in the nPDF determination?

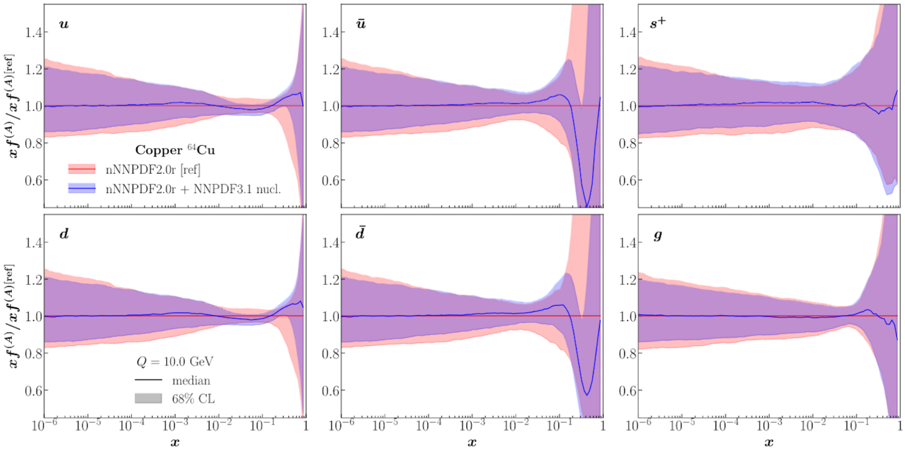


Case 1: remove the deuteron data from the proton baseline

(NMC, SLAC, BCDMS, E866, E906)

Impact of proton baseline fit

What's the impact of varying the proton baseline fit in the nPDF determination?



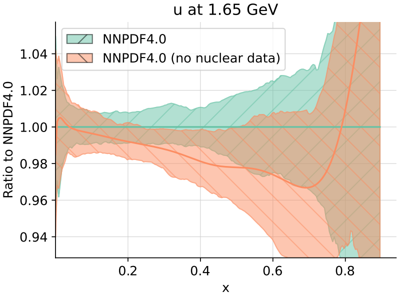
Case 2: remove the heavy nuclei data from the proton baseline

(CHORUS, NuTeV, E605)

2. Beyond nNNPDF3.0

[NNPDF, in preparation]

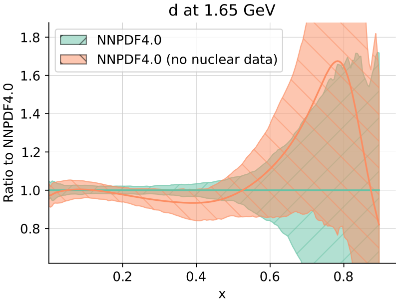
Nuclear data in proton PDF fits

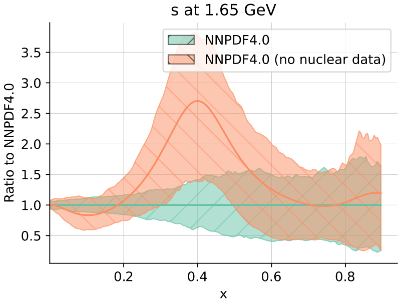


Nuclear data sets have a significant impact on proton PDFs

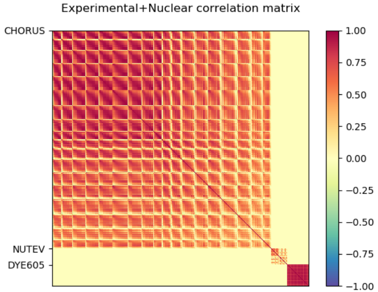
This despite nuclear data sets are a (limited) subset of the global data set (in NNDPF4.0, 1417 out of 4618 data points)

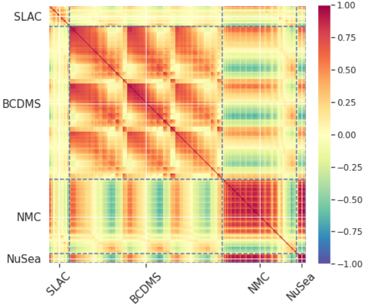
on deuteron (NC DIS): SLAC, BCDMS, NMC
on heavy nuclei (CC DIS): CHORUS, NuTeV
on both (FT DY): E866, E906, E605





Nuclear uncertainty in PDF determination

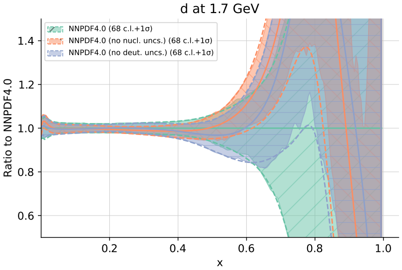




Effect of nuclear uncertainties relevant at large x
to reconcile FT DIS with LHC DY data

$\chi^2_{\rm tot} = 1.17 \to \chi^2_{\rm tot} = 1.26$ (no nucl. uncs.)
$\chi^2_{\rm LHCb} = 1.54 \to \chi^2_{\rm tot} = 1.76$ (no nucl. uncs.)

The bulk of the effect is due to nuclear uncertainties for heavy nuclei
deuteron uncertainties have a comparatively smaller effect at intermediate values of x



[EPJ C79 (2019) 282; EPJ C81 (2021) 37]

Towards *integrated* proton and nuclear PDF fits

Store PDF-independent predictions in interpolation grids

Idea not new (APPLgrid, fastNLO)

Interpolate PDFs with polynomials

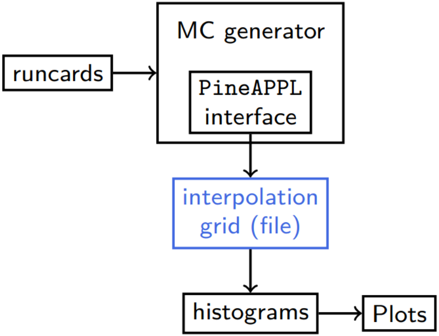
$$f = \sum_i f^i L_i$$

Convolution turns into sums

$$\frac{d\sigma}{d\mathcal{O}} = \sum_f \int dx \frac{d\hat{\sigma}}{d\mathcal{O}} = \sum_{f,i} f^i G_i$$

with interpolation grid

$$G_i = \int dx L_i \hat{\sigma}$$



PineAPPL originally developed to handle QCDxEW corrections [JHEP 12 (2020) 108]

Now interfaced to a wide range of MC generators (MadGraph, Matrix, MCFM, NNLOjet, SHERPA)

Increasingly used in a variety of HEP projects (NNPDF, xFitter, nCTEQ, . . .)

More and more PineAPPL grids are being uploaded to Ploughshare

PineAPPLv1 [T. Jezo, E.R. Nocera, T. Rabemananjara, C. Schwan, T. Sharma, J. Wissmann, in preparation]

Supports an arbitrary number of convolutions

exclusive particle production in proton-proton and proton-nucleus collisions (4 convolutions)

Towards *integrated* proton and nuclear PDF fits

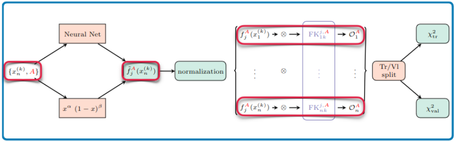
Idea: deploy the NNPDF4.0 methodology
to a simultaneous determination of proton and nuclear PDFs

Proton PDF ($A = 1$ boundary) automatically reflected in nuclear PDFs

Nuclear DIS data contribute consistently to both proton and nuclear PDFs

Take advantage of automated generation of theoretical observables [CPC 297 (2024) 109061]

Playground to optimise GPU parallelisation



Methodology Improvements NNPDF4.0

- **Stochastic Gradient Descent** (SGD) for NN training using Tensorflow (easily changeable)
- **Automated Optimisation of Hyperparameters** (scanning of the parameter space)
- Methodology Validation using **Closure Tests**, **Future Tests**, and **Parametrisation Basis Independence**
- Proton, Deuteron, and heavy Nuclear PDFs are **fitted simultaneously** in a Self-Consistent Way ⇔ **Fitting Smoothly the *A*-dependence**

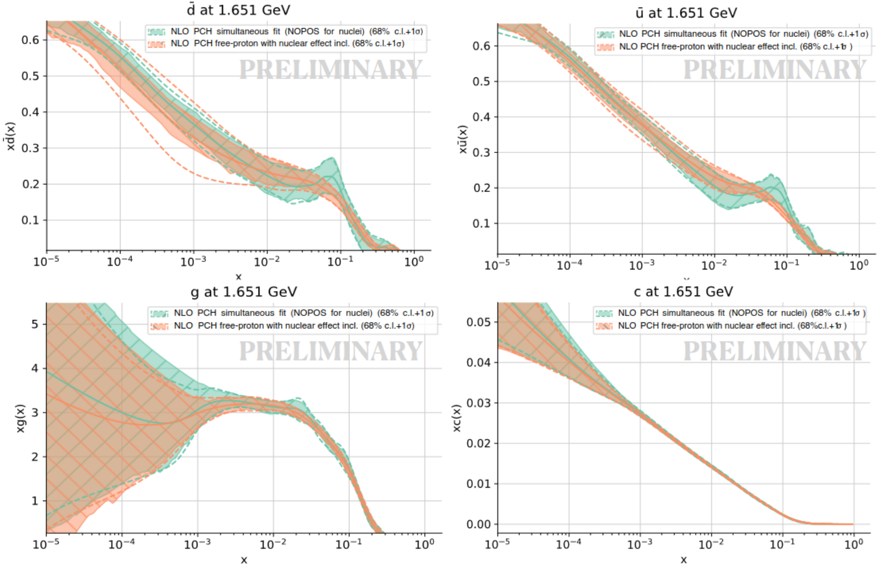
$$\chi^2_{\text{tot}} = \sum_A \chi^2_{t_0}(A)$$

Dataset	NNPDF4.0	Integrated
NMC	1.63	1.62
HERACOMB	1.18	1.16
CDF	1.29	1.24
ATLAS	1.57	1.38
CMS	1.44	1.53
LHCb	1.54	1.46

Dataset	nNNPDF3.0	Integrated
NMC	0.88	1.61
SLAC	1.09	1.38
EMC	0.78	1.80
FNAL	1.01	0.50
BCDMS	0.82	1.35
LHC	1.09	1.15

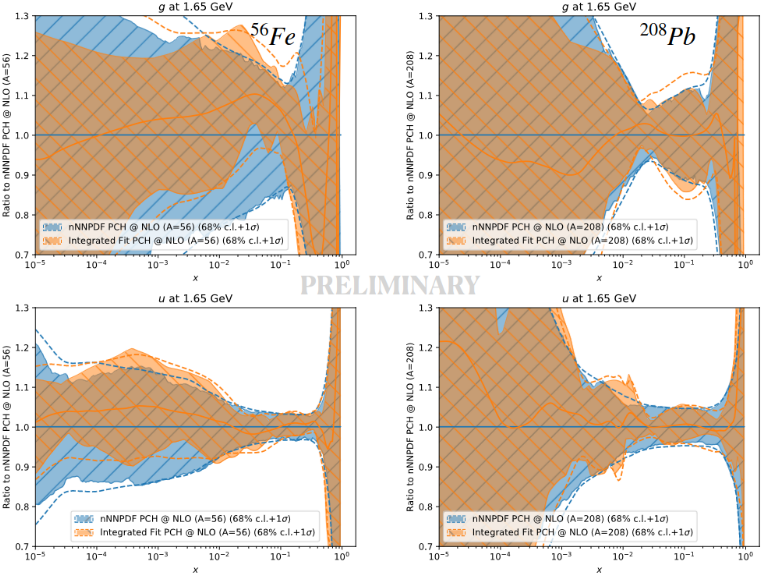
[NNPDF, in preparation; slide by courtesy of T.R. Rabemananjara]

Towards *integrated* proton and nuclear PDF fits



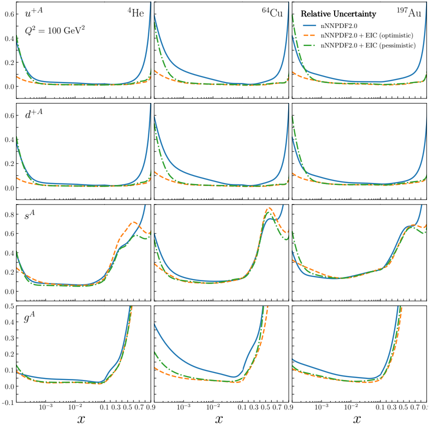
$A = 1$: generally good agreement, although further optimisation is still needed

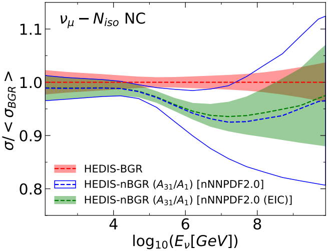
Towards *integrated* proton and nuclear PDF fits

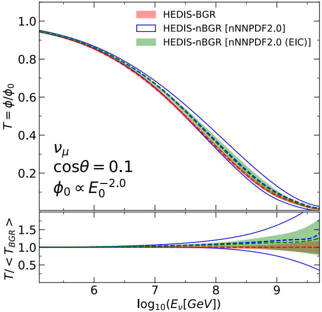


$A > 1$: generally good agreement, although better for Lead than for Iron

Nuclear PDFs at the EIC (nNNPDF2.0)







Impact from pseudodata [PRD 103 (2021) 096005]

e^-N CC DIS, N=^{4}He, ^{12}C, ^{40}Ca, ^{64}Cu, ^{197}Au

$E_\ell \times E_p$ [GeV]: 18× 100; 10× 110; 5× 41

$\mathcal{L} = 10$ fb^{-1}; $\sigma_u = 1.5/2.3\%$; $\sigma_c = 2.5/4.3\%$

+ other future facilities, such as the Forward Physics Facility

3. To conclude

Summary and outlook

Progress in incorporating new data in nNNPDF fits
LHCb pPb/pp D^0-meson production data puts unprecedented constraints on the gluon nPDF
which shows evidence of shadowing and antishadowing

Advancements in various methodological aspects of the nNNPDF framework
hyperoptimisation, dependence on the proton baseline,
interpolation grids for fast PDF convolution

Ongoing work to understand correlations between proton and nuclear PDF analyses
promising integrated QCD fits are ongoing

The future will possibly bring a wealth of precise new measurements
LHC Run III, sPHENIX, SMOG@LHCb, FoCal@ALICE, EIC, FPF

Summary and outlook

Progress in incorporating new data in nNNPDF fits
LHCb pPb/pp D^0-meson production data puts unprecedented constraints on the gluon nPDF
which shows evidence of shadowing and antishadowing

Advancements in various methodological aspects of the nNNPDF framework
hyperoptimisation, dependence on the proton baseline,
interpolation grids for fast PDF convolution

Ongoing work to understand correlations between proton and nuclear PDF analyses
promising integrated QCD fits are ongoing

The future will possibly bring a wealth of precise new measurements
LHC Run III, sPHENIX, SMOG@LHCb, FoCal@ALICE, EIC, FPF

Thank you