Published for SISSA by Springer

Received: July 3, 2020
Accepted: September 4, 2020
Published: September 28, 2020

nNNPDF2.0: quark flavor separation in nuclei from LHC data

Rabah Abdul Khalek,[a,b] **Jacob J. Ethier,**[a,b] **Juan Rojo**[a,b] **and Gijs van Weelden**[b]

[a] *Department of Physics and Astronomy, VU Amsterdam, 1081HV Amsterdam, The Netherlands*
[b] *Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*

E-mail: rabah.khalek@gmail.com, j.j.ethier@vu.nl, j.rojo@vu.nl, gijsvanweelden@hotmail.com

Abstract: We present a model-independent determination of the nuclear parton distribution functions (nPDFs) using machine learning methods and Monte Carlo techniques based on the NNPDF framework. The neutral-current deep-inelastic nuclear structure functions used in our previous analysis, nNNPDF1.0, are complemented by inclusive and charm-tagged cross-sections from charged-current scattering. Furthermore, we include all available measurements of W and Z leptonic rapidity distributions in proton-lead collisions from ATLAS and CMS at $\sqrt{s} = 5.02\,\text{TeV}$ and $8.16\,\text{TeV}$. The resulting nPDF determination, nNNPDF2.0, achieves a good description of all datasets. In addition to quantifying the nuclear modifications affecting individual quarks and antiquarks, we examine the implications for strangeness, assess the role that the momentum and valence sum rules play in nPDF extractions, and present predictions for representative phenomenological applications. Our results, made available via the LHAPDF library, highlight the potential of high-energy collider measurements to probe nuclear dynamics in a robust manner.

Keywords: Deep Inelastic Scattering (Phenomenology), QCD Phenomenology

ArXiv ePrint: 2006.14629

Open Access, © The Authors.
Article funded by SCOAP[3].

https://doi.org/10.1007/JHEP09(2020)183

JHEP09(2020)183

Contents

1 Introduction **1**

2 Experimental data and theory calculations **3**
- 2.1 Input dataset 3
- 2.2 Theoretical calculations 7

3 Fitting methodology **8**
- 3.1 Notation and conventions 8
- 3.2 nPDF parameterization 10
- 3.3 Cross-section positivity 18

4 Results **20**
- 4.1 The nNNPDF2.0 determination 21
- 4.2 Comparison with nNNPDF1.0 36
- 4.3 The momentum and valence integrals in nuclei 40
- 4.4 The positivity of physical cross-sections 43

5 Implications for photon and hadron production in nuclear collisions **44**
- 5.1 Isolated photon production in pA collisions with ATLAS 45
- 5.2 Isolated photon production in pA collisions with FoCal 47
- 5.3 Inclusive hadron production in pA collisions 49

6 Summary and outlook **50**

A PDF sensitivity of input cross-sections **53**

JHEP09(2020)183

1 Introduction

Decades of experimental investigations have plainly revealed the inability to describe, within the framework of perturbative QCD, high-energy scattering processes involving heavy nuclei using a free-nucleon formalism. The parton distribution functions (PDFs) of nucleons bound within nuclei, commonly known as nuclear PDFs (nPDFs) [1, 2], can therefore be significantly modified with respect to their free-nucleon counterpart [3] as a result of non-perturbative dynamics. While a first-principles understanding of the theoretical mechanisms that generate such QCD dynamics remains an open challenge, phenomenological determinations of nPDFs have been able to provide vital information about parton behavior in the cold nuclear medium.

Precise extractions of nPDFs are not only crucial to study the strong interaction in the high-density regime, but are also necessary to model the initial state of heavy ion collisions which aim to characterize the Quark-Gluon Plasma (QGP) [4, 5] using hard probes. Furthermore, nPDFs also contribute to global QCD analyses of the proton structure [6–9] via the inclusion of neutrino structure function data collected in reactions involving heavy nuclear targets. These measurements on nuclear targets provide important information on the quark flavor separation and strangeness in the proton [10].

Several groups have recently presented determinations of the nuclear PDFs using different input datasets, theoretical assumptions, and methodological settings [11–15]. While nPDF analyses are based on a significantly reduced dataset compared to the free-nucleon case, the situation has improved in recent years with the availability of hard-scattering cross-section data from proton-lead collisions at the LHC for processes such as jet, W and Z, and heavy quark production [16–33]. These collider measurements can clarify several poorly understood aspects of nuclear PDFs, such as the quark flavor dependence of nuclear effects and the nuclear modifications of the gluon distribution. Several studies have indeed demonstrated the valuable constraints that can be provided for the nuclear PDFs from proton-lead collisions at the LHC, see e.g. refs. [11, 34–37].

This work focuses on the determination of the quarks and anti-quark nuclear PDFs, with emphasis on their flavor separation. Since measurements of neutral-current (NC) deep-inelastic scattering (DIS) nuclear structure functions on isoscalar targets are only sensitive to a single quark PDF combination, one needs to rely on the information provided by independent processes to disentangle quark and antiquarks of different flavors. The main options that are available to accomplish this are neutrino-induced charged current (CC) DIS cross-sections on heavy nuclear targets, sensitive to different quark combinations than the NC case, and electroweak gauge boson production at the LHC.

From the methodological point of view, there exist two primary limitations that affect the separation between quark and antiquark flavors in nPDF extractions. The first one is the reliance on *ad-hoc* theoretical assumptions required to model the dependence of the nuclear modifications on both the parton momentum fraction x and atomic mass number A, where in some cases the expected behavior is hard-coded in the nPDF parameterization. The second is the lack of consistency between the nuclear PDF determination and that of the corresponding proton baseline, to which the former should reduce to in the $A \to 1$ limit in terms of central values and uncertainties. This consistency is particularly important given that the precision LHC data impose stringent constraints on the quark flavor separation for the proton PDFs, for example via measurements of inclusive W and Z production characterized by per-mille level uncertainties. Ensuring that the LHC constrains on the proton PDF baseline are appropriately propagated to the nPDF determination for $A > 2$ is therefore critical.

In this study we present a model-independent determination of nuclear PDFs using machine learning methods and Monte Carlo techniques based on the NNPDF framework [38–47]. We complement our previous nNNPDF1.0 analysis of NC DIS nuclear structure functions with CC inclusive and charm-tagged measurements from fixed-target neutrino experiments as well as with inclusive W and Z production cross-sections in proton-lead collisions

JHEP09(2020)183

from ATLAS and CMS at $\sqrt{s} = 5.02$ TeV (Run I) and 8.16 TeV (Run II). The $A = 1$ proton PDF baseline used in the present analysis is defined to be a variant of the NNPDF3.1 fit which excludes heavy nuclear target data. This choice allows us to indirectly incorporate the constraints on quark flavor separation provided by the pp measurements from the LHC.

The nNNPDF2.0 results allow us to tackle several important issues concerning nuclear effects among various quark flavors. First, we assess the compatibility of the LHC W and Z leptonic rapidity distributions from proton-lead collisions with the constraints coming from DIS structure functions, and demonstrate that the former allow for a marked improvement in the quark PDF uncertainties. We also study the nuclear effects on the total strange content of heavy nuclei, highlighting the interplay between the information provided by DIS and hadronic data. This interplay is also interesting from the proton PDF point of view, where the pull on strangeness provided by the ATLAS W, Z distributions [48] is the opposite from that of neutrino data and other LHC processes such as the W+c cross-sections.

We then analyze the impact that the momentum and total valence sum rule constraints have in the global nPDF determination, and demonstrate that the corresponding integrals agree with QCD predictions within uncertainties even when the sum rules are not explicitly imposed. We conclude the paper by providing theoretical predictions based on nNNPDF2.0 for representative processes of phenomenological interest in proton-ion collisions: isolated photon production in the central and forward rapidity regions and inclusive pion production.

This paper is organized in the following manner. In section 2, we provide the input experimental observables used in this analysis and detail the corresponding theoretical calculations. We define a set of conventions and notation used in this work and describe new aspects of our fitting methodology in section 3. The nNNPDF2.0 nuclear parton distributions are then presented in section 4, followed by a discussion of phenomenological implications in section 5. Lastly, in section 6 we conclude with a summary and highlight future directions of study.

2 Experimental data and theory calculations

In this section, we provide details of the experimental measurements used as input for the nNNPDF2.0 determination. An emphasis is made in particular on the new datasets that are added with respect to those that were present in nNNPDF1.0. We then discuss the theoretical calculations corresponding to these datasets and their numerical implementation in our fitting framework.

2.1 Input dataset

Common to the previous nNNPDF1.0 analysis are the nuclear NC DIS measurements listed in table 1. For each dataset, the target nuclei A_1 and A_2 used by each experiment are indicated together with their atomic mass numbers. We also list the number of data points after the DIS kinematical cuts and provide the corresponding publication reference. The DIS kinematic cuts are the same as in our previous study, i.e. $Q^2 = 3.5\,\mathrm{GeV}^2$ and

JHEP09(2020)183

Experiment	A_1/A_2	N_{dat}	Reference
SLAC E-139	^{4}He/^{2}D	3	[49]
NMC 95, re.	^{4}He/^{2}D	13	[50]
NMC 95	^{6}Li/^{2}D	12	[51]
SLAC E-139	^{9}Be/^{2}D	3	[49]
NMC 96	^{9}Be/^{12}C	14	[52]
EMC 88, EMC 90	^{12}C/^{2}D	12	[53, 54]
SLAC E-139	^{12}C/^{2}D	2	[49]
NMC 95, NMC 95, re.	^{12}C/^{2}D	26	[50, 51]
FNAL E665	^{12}C/^{2}D	3	[55]
NMC 95, re.	^{12}C/^{6}Li	9	[50]
BCDMS 85	^{14}N/^{2}D	9	[56]
SLAC E-139	^{27}Al/^{2}D	3	[49]
NMC 96	^{27}Al/^{12}C	14	[52]
SLAC E-139	^{40}Ca/^{2}D	2	[49]
NMC 95, re.	^{40}Ca/^{2}D	12	[50]
EMC 90	^{40}Ca/^{2}D	3	[54]
FNAL E665	^{40}Ca/^{2}D	3	[55]
NMC 95, re.	^{40}Ca/^{6}Li	9	[50]
NMC 96	^{40}Ca/^{12}C	23	[52]
EMC 87	^{56}Fe/^{2}D	58	[57]
SLAC E-139	^{56}Fe/^{2}D	8	[49]
NMC 96	^{56}Fe/^{12}C	14	[52]
BCDMS 85, BCDMS 87	^{56}Fe/^{2}D	16	[56, 58]
EMC 88, EMC 93	^{64}Cu/^{2}D	27	[53, 59]
SLAC E-139	^{108}Ag/^{2}D	2	[49]
EMC 88	^{119}Sn/^{2}D	8	[53]
NMC 96, Q^2 dependence	^{119}Sn/^{12}C	119	[60]
FNAL E665	^{131}Xe/^{2}D	4	[61]
SLAC E-139	^{197}Au/^{2}D	3	[49]
FNAL E665	^{208}Pb/^{2}D	3	[55]
NMC 96	^{208}Pb/^{12}C	14	[52]
Total NC DIS		**451**	

Table 1. The neutral-current nuclear deep-inelastic input datasets included in nNNPDF2.0. For each dataset, we indicate the nuclei A_1 and A_2 involved, the number of data points that satisfy the baseline kinematical cuts, and the publication reference.

$W^2 = 12.5\,\mathrm{GeV}^2$, consistent with the NNPDF3.1 proton PDF baseline used to satisfy our boundary condition.

Note that all NC DIS measurements listed in table 1 are provided in terms of ratios of structure functions between two different nuclei. In most cases the denominator is given by the deuterium structure function, but ratios to carbon and lithium are also provided. As we will discuss in section 3, our fitting approach parameterizes the PDFs entering the absolute structure functions for each value of A, after which their ratios are constructed.

JHEP09(2020)183

Experiment	A	N_{dat}	Reference
CHORUS ν	208	423	[62]
CHORUS $\bar{\nu}$	208	423	[62]
NuTeV ν	56	39	[63]
NuTeV $\bar{\nu}$	56	37	[63]
Total CC DIS		**922**	
CMS $W^{\pm}$ $\sqrt{s} = 8.16\,\text{TeV}$	208	48	[64]
CMS $W^{\pm}$ $\sqrt{s} = 5.02\,\text{TeV}$	208	20	[25]
CMS Z $\sqrt{s} = 5.02\,\text{TeV}$	208	12	[24]
ATLAS Z $\sqrt{s} = 5.02\,\text{TeV}$	208	14	[23]
Total LHC		**94**	
Total		**1467**	

Table 2. Same as table 1 for the new datasets that have been added to nNNPDF2.0. As opposed to the NC structure function measurements, these datasets are presented as absolute distributions rather than as cross-sections ratios. We also indicate the total number of data points in the fit, combining the NC and CC structure functions with the LHC data.

The remaining input data which is newly added to our nNNPDF2.0 analysis is presented in terms of absolute cross-sections, without normalizing to any baseline nucleus. We list these data in table 2, divided into two categories: CC neutrino DIS reduced cross-sections on nuclear targets and leptonic rapidity distributions in electroweak gauge boson production from proton-lead collisions at the LHC. The neutrino and anti-neutrino reduced cross-sections are further separated into inclusive cross-sections from CHORUS [62] and charm-tagged cross-sections from NuTeV [63]. The LHC measurements are divided into data from ATLAS and from CMS from the Run I and Run II data-taking periods. In this table we also indicate the total number of data points included in the fit, combining the NC and CC cross-sections measurements with the LHC data. In total, the nNNPDF2.0 global fit contains $n_{\text{dat}} = 1467$ data points.

Starting with the CC measurements from CHORUS, we fit the inclusive neutrino and anti-neutrino double-differential cross-sections, $d^2\sigma^{\nu N}/dxdQ^2$. After imposing kinematic cuts, the dataset consists of $n_{\text{dat}} = 846$ data points equally distributed between neutrino and anti-neutrino beams. The fitted cross-sections are not corrected for non-isoscalarity of the lead target, and therefore the corresponding theory calculations take into account effects related to the difference between $Z = 82$ and $Z = A/2 = 104$. The situation is therefore different from the treatment of NC nuclear structure functions, where the experimental results are presented with non-isoscalar effects already subtracted, as discussed in [13].

In addition to the CHORUS reduced cross-sections, nNNPDF2.0 also includes the NuTeV di-muon cross-sections from neutrino-iron scattering. Dimuon events in neutrino DIS are associated with the $\text{W}^{\pm} + s\ (d) \to c$ scattering process, where the charm quark hadronizes into a charmed meson and then decays into a final state containing a muon. This process is dominated by the strange-initiated contributions since other initial states are CKM-suppressed, thus providing direct sensitivity to the strange quark nuclear PDF.

JHEP09(2020)183

In fact, the NuTeV di-muon data are known to play an important role in studies of proton strangeness in global QCD analyses. After kinematic cuts, we end up with $n_{\rm dat} = 39$ and 37 data points for the neutrino and the anti-neutrino cross-sections, respectively. Together with the CHORUS cross section data, the CC measurements comprise a majority of the input dataset with a total of $n_{\rm dat} = 922$ data points.

Moving now to the LHC electroweak gauge boson cross-sections, we consider in this work the four datasets that are listed in table 2. Three of the datasets come from the Run I data-taking period, corresponding to a per-nucleon center-of-mass energy of $\sqrt{s} = 5.02\,\text{TeV}$. These are the ATLAS Z rapidity distributions [23] and the CMS $\text{W}^\pm$ [25] and Z [24] rapidity distributions, which contain $n_{\rm dat} = 14$, 20, and 12 data points respectively. Note that ATLAS does not have a published measurement of the $\text{W}^\pm$ rapidity distributions from Run I and that only preliminary results have been presented [22].

In the same way as the CC reduced cross-sections, the LHC measurements of electroweak gauge boson production are provided as absolute distributions. In this case, however, it is possible to construct new observables with the LHC $\text{W}^\pm$ and Z production data that might be beneficial for nPDF determinations. For example, the EPPS16 analysis composed and analyzed the forward-to-backward ratio, where cross sections at positive lepton rapidities are divided by the ones at negative rapidities. Nevertheless, in this work we choose only to work with the absolute rapidity distributions presented in the experimental publications.

In addition to the three Run I results, we add also for the first time in an nPDF analysis measurements from Run II corresponding to a per-nucleon center-of-mass energy of $\sqrt{s} = 8.16\,\text{TeV}$. More specifically, the measurements correspond to W^+ and W^- leptonic rapidity distributions [64] from CMS, which provide an additional $n_{\rm dat} = 48$ data points. The fact that the amount of data is more than doubled compared to the corresponding Run I measurements is a consequence of the increase in the CoM energy as well as the higher integrated luminosity. In particular, the Run II measurements are based on $\mathcal{L} = 173\ \text{nb}^{-1}$ compared to $\mathcal{L} = 34.6\ \text{nb}^{-1}$ available from Run I. The CMS Run II data are therefore expected to provide important constraints on the quark flavor separation of the nuclear PDFs.

As opposed to the situation in proton-proton collisions, the LHC gauge production measurements do not provide information on the correlation between experimental systematic uncertainties. For this reason, we construct the total experimental error by adding the various sources of uncertainty in quadrature. The only source of systematic error which is kept as fully correlated among all the data bins of a given dataset is the overall normalization uncertainty. Note that this normalization uncertainty is correlated within a single experiment and LHC data-taking period, but elsewhere is uncorrelated between different experiments.

In order to illustrate the coverage of the experimental data that is included in nNNPDF2.0 and summarized in tables 1 and 2, we display in figure 1 their kinematical range in the (x, Q^2) plane. Here the horizontal dashed and curved dashed lines correspond to the kinematic cuts of $Q^2 = 3.5\,\text{GeV}^2$ and $W^2 = 12.5\,\text{GeV}^2$, respectively. In addition, we show for each hadronic data point the two values of x corresponding to the parton momentum fractions of the incoming proton and lead beams, computed to leading order.

JHEP09(2020)183

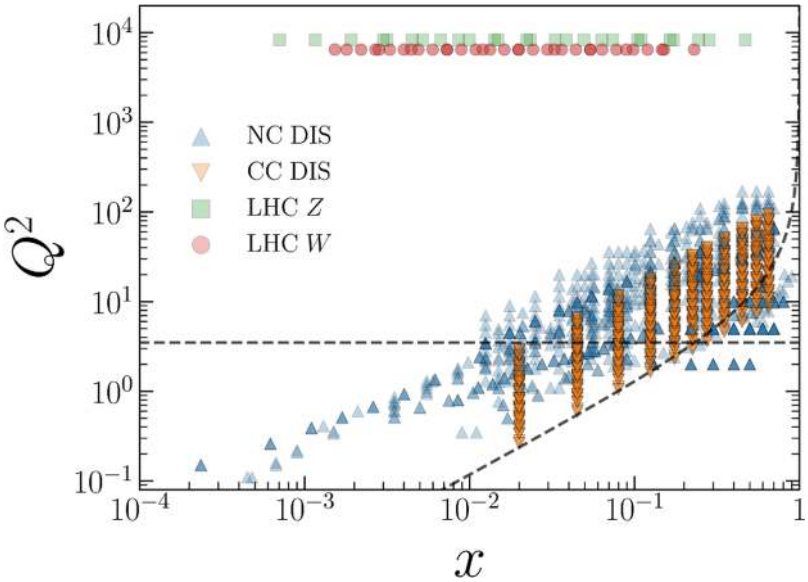

Figure 1. Kinematical coverage in x and Q^2 of the data points included in the nNNPDF2.0 determination. The horizontal dashed and curved dashed lines correspond to $Q^2 = 3.5\,\text{GeV}^2$ and $W^2 = 12.5$ GeV^2, respectively, which are the kinematic cuts imposed in this analysis. For each LHC measurement, there are two values of x associated with leading order kinematics of proton-lead scattering being displayed.

There are several interesting observations that one can make regarding figure 1. First of all, the LHC proton-lead measurements significantly extend the kinematic coverage of the fixed-target DIS reduced cross-sections, both in terms of x and Q^2. In particular, the LHC data reside at Q^2 values that are orders of magnitude larger while the coverage in partonic momentum fraction is extended down to $x \simeq 10^{-3}$. Secondly, the CC reduced cross-sections have a similar coverage compared to the NC ones, providing sensitivity to different quark and antiquark combinations across the shared medium- to large-x region. Finally, the kinematics of the LHC W and Z measurements largely overlap. The ability to describe them simultaneously can therefore demonstrate the compatibility between the experimental data and theoretical calculations.

2.2 Theoretical calculations

DIS structure functions. For the NC DIS structure functions we use the same theoretical settings as in nNNPDF1.0, i.e. the structure functions are evaluated at NLO using `APFEL` [65] in the FONLL-B general-mass variable flavor number scheme [66]. The value of the strong coupling constant is taken to be the same as in the NNPDF3.1 proton PDF fit, $\alpha_S(m_Z) = 0.118$, as well as the charm and bottom mass thresholds $m_c = 1.51\,\text{GeV}$ and $m_b = 4.92\,\text{GeV}$, respectively. The charm PDF is generated perturbatively by the DGLAP evolution equations and is thus absent from the $n_f = 3$ scheme. Lastly, the structure functions are processed by the `APFELgrid` [67] fast interpolation tables which allow for efficient evaluations during the PDF fit.

JHEP09(2020)183

Concerning the CC neutrino reduced cross-sections, most of the theory settings are shared with their NC counterparts. The main difference is that the heavy quark contributions in the CC predictions at NLO are accounted for in the FONLL-A scheme instead to maintain consistency with the proton baseline. Massive $\mathcal{O}\left(\alpha_s^2\right)$ corrections to charm production in CC DIS have been presented in ref. [68], and subsequently used to study their impact in the determination of the strange content of the nucleon in ref. [69]. Further details about the implementation of heavy quark mass corrections in the NNPDF framework for charged-current scattering can be found in ref. [46].

Hadronic cross-sections. The rapidity distributions from W and Z boson production in proton-lead collisions are evaluated at NLO using `MCFM` [70] v6.8 interfaced with `APPLgrid` [71]. We have ensured that the numerical integration uncertainties in the `MCFM` calculations are always much smaller than the corresponding experimental errors. Furthermore, our calculations are benchmarked with the reference theoretical values whenever provided by the corresponding experimental publications. To illustrate this benchmarking, we display in figure 2 the muon rapidity distributions for W^- boson production at $\sqrt{s} = 8.16\,\mathrm{TeV}$ in the center-of-mass frame. Here we compare our `MCFM`-based calculation with the theory predictions provided in ref. [64] at the level of absolute cross-sections (upper panel) and also as ratios to the central experimental values (lower panel). In both cases, the CT14 NLO proton PDF set is adopted as input and nuclear corrections are neglected. As can been seen by the figure, there is good agreement at the $\sim$1% level between our calculations and the reference results provided in the CMS paper, with residual differences likely to be traced back to a different choice of electroweak scheme. Similar agreement is obtained for the rest of hadronic datasets included in the present analysis.

Since the fast interpolation grids are computed in the center-of-mass frame of the proton-lead collision, rapidity bins that are given in the laboratory frame η_{lab} are shifted to the center-of-mass frame η_{CM} when required. This shift is given by $\eta_{\mathrm{lab}} = \eta_{\mathrm{CM}} + 0.465$ both at $\sqrt{s} = 5.02$ and 8.16 center-of-mass energies. Lastly, we note that the same theoretical settings were used for the evaluation of W and Z production in proton-proton collisions for the baseline NNPDF3.1 fit.

3 Fitting methodology

In this section, we describe the fitting methodology that was adopted for the nNNPDF2.0 determination, focusing in particular on the differences and improvements with respect to the nNNPDF1.0 analysis. We begin by establishing the PDF notation and conventions that will be used throughout this work. We then detail our strategy to parameterize the nuclear parton distributions, including the treatment of sum rules, preprocessing factors, and the proton boundary condition. Lastly, we outline the implementation of the cross-section positivity constraint.

3.1 Notation and conventions

Parton distributions can be parametrized in a number of different bases, all of which are related by linear transformations. Two popular ones are the flavor basis, corresponding

JHEP09(2020)183

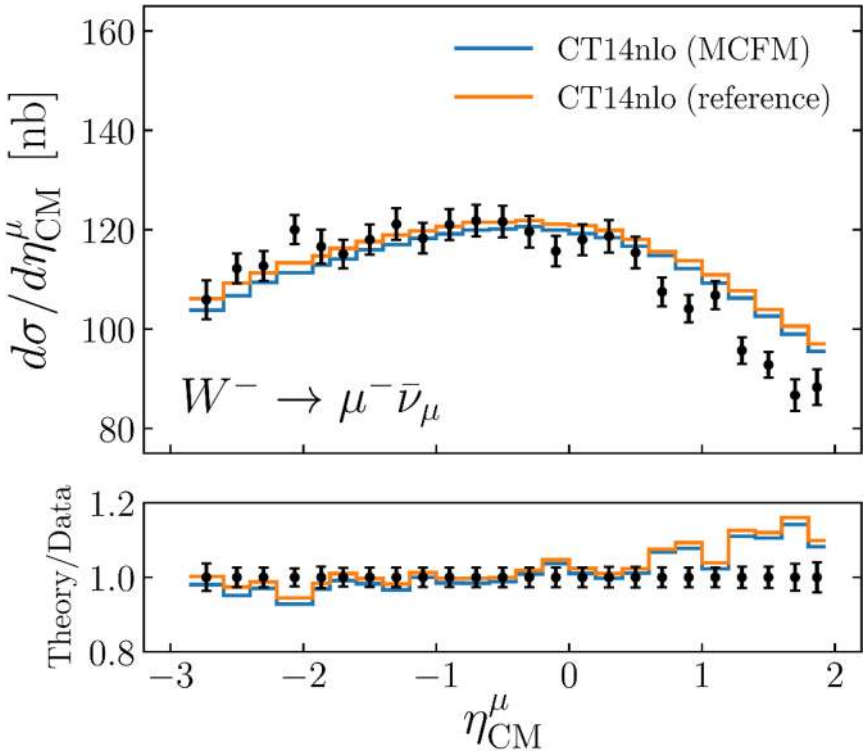

Figure 2. The leptonic rapidity distributions for W^- boson production at $\sqrt{s} = 8.16\,\text{TeV}$ in the center-of-mass frame. Our `MCFM`-based calculation is compared with the theory predictions provided in [64] both as absolute cross-sections (upper) and as ratios to the experimental data (lower panel). In both cases the CT14 NLO proton PDF set is adopted and nuclear corrections are neglected.

to the individual quark and anti-quark PDFs, and the evolution basis, given by the eigenvectors of the DGLAP evolution equations [6]. For illustrative purposes we consider here three active quarks, a vanishing strangeness asymmetry, and heavy quark PDFs that are generated via perturbative evolution. At the parameterization scale $Q_0 < m_c$, the flavor basis is composed of the u, $\bar{u}$, d, $\bar{d}$, s, and g PDFs, with $s = \bar{s}$, while the corresponding evolution basis is given by Σ, T_3, T_8, V, V_3, and g. Expressed in terms of the elements of the flavor basis, the evolution basis distributions are given by

$$\begin{aligned}
\Sigma(x,Q_0) &= \left(u^+ + d^+ + s^+\right)(x,Q_0)\,,\\
T_3(x,Q_0) &= \left(u^+ - d^+\right)(x,Q_0)\,,\\
T_8(x,Q_0) &= \left(u^+ + d^+ - 2s^+\right)(x,Q_0)\,,\\
V(x,Q_0) &= \left(u^- + d^-\right)(x,Q_0)\,,\\
V_3(x,Q_0) &= \left(u^- - d^-\right)(x,Q_0)\,,
\end{aligned} \tag{3.1}$$

where $q^{\pm} = q \pm \bar{q}$. Although the results of an nPDF analysis should be independent of the basis choice for the parameterization, some bases offer practical advantages, for example in the implementation of the sum rules which are discussed later.

In this work, we define $f^{(N/A)}$ to be the PDF for the flavor f associated to the average nucleon N bound in a nucleus with atomic number Z and mass number A. This object

JHEP09(2020)183

can be written as

$$f^{(N/A)}(x,Q_0) = \frac{Z}{A} f^{(p/A)}(x,Q_0) + \left(1 - \frac{Z}{A}\right) f^{(n/A)}(x,Q_0)\,, \tag{3.2}$$

where $f^{(p/A)}$ and $f^{(n/A)}$ represent the PDFs of a proton and a neutron, respectively, bound in the same nucleus of mass number A. Assuming isospin symmetry, the PDFs of the neutron in eq. (3.2) can be expressed in terms of the proton PDFs via

$$\begin{aligned} u^{(n/A)}(x,Q_0) &= d^{(p/A)}(x,Q_0)\,, & \bar{u}^{(n/A)}(x,Q_0) &= \bar{d}^{(p/A)}(x,Q_0) \\ d^{(n/A)}(x,Q_0) &= u^{(p/A)}(x,Q_0)\,, & \bar{d}^{(n/A)}(x,Q_0) &= \bar{u}^{(p/A)}(x,Q_0) \\ s^{(n/A)}(x,Q_0) &= s^{(p/A)}(x,Q_0)\,, & g^{(n/A)}(x,Q_0) &= g^{(p/A)}(x,Q_0)\,. \end{aligned} \tag{3.3}$$

Using the relations above, the strange and gluon distributions of the average bound nucleon ($f^{(N/A)}$) and bound proton ($f^{(p/A)}$) become equivalent, while the up and down flavored distributions of the average bound nucleon are instead linear combinations of the bound proton PDFs with coefficients depending on the values of A and Z.

3.2 nPDF parameterization

The cross-sections for hard scattering processes involving heavy nuclei can be expressed either in terms of $f^{(N/A)}$ or $f^{(p/A)}$. The two options are fully equivalent, as is highlighted by the LO expressions of the observables collected in appendix A. One can therefore choose to parameterize either the PDFs of the average bound nucleon or those of the bound proton in a global nPDF analysis. In this study we choose the latter option for two main reasons. First, the connection with the free-proton boundary condition is more straightforward. In addition, the valence sum rules for non-isoscalar nuclei are independent of A and Z. If instead $f^{(N/A)}$ is parameterized, one of the valence sum rules would depend on the value of the Z/A ratio and thus be different for each nuclei, making it inconvenient from the parameterization point of view.

The relation between $f^{(N/A)}$ and $f^{(p/A)}$ is trivial also for PDF combinations that comprise the evolution basis. Consider for example the total quark singlet, where the flavor combination is the same in the proton and in the neutron, i.e. $\Sigma^{(p/A)} = \Sigma^{(n/A)}$. From eq. (3.2), it simply follows that $\Sigma^{(N/A)} = \Sigma^{(p/A)}$. The same equivalence holds also for V and T_8. However, the distinction is important for T_3 and V_3, for which we have

$$V_3^{(N/A)} = \frac{Z}{A} V_3^{(p/A)} + \left(1 - \frac{Z}{A}\right) V_3^{(n/A)} = \left(\frac{2Z}{A} - 1\right) V_3^{(p/A)}\,, \tag{3.4}$$

$$T_3^{(N/A)} = \frac{Z}{A} T_3^{(p/A)} + \left(1 - \frac{Z}{A}\right) T_3^{(n/A)} = \left(\frac{2Z}{A} - 1\right) T_3^{(p/A)}\,, \tag{3.5}$$

so there is an overall rescaling factor of $(2Z/A - 1)$ between $f^{(N/A)}$ and $f^{(p/A)}$. The main consequence of this relation is highlighted by assuming an isoscalar nucleus, with $Z = A/2$. In this case, $V_3^{(N/A)} = T_3^{(N/A)} = 0$ while their bound proton counterparts are different from zero. Unless otherwise indicated, the nPDFs discussed in this section will always correspond to those of the bound proton.

JHEP09(2020)183

Fitting basis and functional form. In our previous nNNPDF1.0 analysis, we parameterized only three independent evolution basis distributions at the initial scale Q_0, namely the total quark singlet $\Sigma(x, Q_0)$, the quark octet $T_8(x, Q_0)$, and the gluon nPDF $g(x, Q_0)$. From the LO expression of eq. (A.5), it is clear that NC structure functions are sensitive only to a specific combination of Σ and T_8 for isoscalar nuclei, in particular $\Sigma + T_8/4$, while the contribution proportional to T_3 vanishes. In other words, Σ and T_8 are strongly anti-correlated and only the combination $\Sigma + T_8/4$ can be meaningfully determined from the data.

The picture is quite different in the present study, where the availability of charged current DIS data and electroweak gauge boson production cross-sections in proton-lead collisions allow additional elements of the evolution PDF basis to be parameterized (see appendix A). If non-isoscalar effects are neglected, there is only a single distribution to be added to our evolution basis choice, namely the total valence quark combination $V = u^- + d^-$. However, non-isoscalar corrections are necessary for the targets considered in this analysis, particularly for lead. In this case, the quark triplet $T_3 = u^+ - d^+$ and the valence triplet $V_3 = u^- - d^-$ must also be parameterized. Note that since T_3 and V_3 correspond to bound protons, they will be different from zero even for isoscalar nuclei. However, in such cases their contribution to the scattering cross-section vanishes and therefore the data provides no constraint on these combinations.

Putting together these considerations, in this work we parameterize six independent PDF combinations in the evolution basis as follows

$$\begin{aligned}
x\Sigma^{(p/A)}(x, Q_0) &= x^{\alpha_\Sigma}(1-x)^{\beta_\Sigma}\mathrm{NN}_\Sigma(x, A)\,, \\
xT_3^{(p/A)}(x, Q_0) &= x^{\alpha_{T_3}}(1-x)^{\beta_{T_3}}\mathrm{NN}_{T_3}(x, A)\,, \\
xT_8^{(p/A)}(x, Q_0) &= x^{\alpha_{T_8}}(1-x)^{\beta_{T_8}}\mathrm{NN}_{T_8}(x, A)\,, \\
xV^{(p/A)}(x, Q_0) &= B_V x^{\alpha_V}(1-x)^{\beta_V}\mathrm{NN}_V(x, A)\,, \\
xV_3^{(p/A)}(x, Q_0) &= B_{V_3} x^{\alpha_{V_3}}(1-x)^{\beta_{V_3}}\mathrm{NN}_{V_3}(x, A)\,, \\
xg^{(p/A)}(x, Q_0) &= B_g x^{\alpha_g}(1-x)^{\beta_g}\mathrm{NN}_g(x, A)\,.
\end{aligned} \tag{3.6}$$

In these expressions, $\mathrm{NN}_f(x, A)$ stands for the value of the neuron in the output layer of the neural network associated to each specific distribution. As was done in nNNPDF1.0, we use a single artificial neural network consisting of an input layer, one hidden layer with sigmoid activation function, and an output layer with linear activation function. The input layer contains three neurons that take as input the values of the momentum fraction x, $\ln(1/x)$, and atomic mass number A, respectively. Since the hidden layer contains 25 neurons, there are a total of $N_{\rm par} = 256$ free parameters (weights and thresholds) in the neural network used to model our nPDFs.

The neural-net parameterization in eq. (3.6) is then complemented by three normalization coefficients B_g, B_V, and B_{V_3} which are fixed by the sum rules, and by twelve preprocessing exponents α_f and β_f which are fitted simultaneously with the network parameters. Since our proton baseline is a variant of the NNPDF3.1 global NLO fit [72] with perturbative charm, we adopt for consistency the same parameterization scale of $Q_0 = 1\,\mathrm{GeV}$.

JHEP09(2020)183

It is important to emphasize here that the parameterization in eq. (3.6) is valid from $A = 1$ (free-proton) up to $A = 208$ (lead). As a result, the nNNPDF2.0 analysis incorporates an independent determination of the free-proton PDFs, where agreement with the proton PDF baseline is enforced by means of a boundary condition as explained below. This is a relevant distinction, implying that the $A = 1$ PDF can by construction differ slightly from our proton baseline, for example as a result of positivity constraints that are more general in the former case, or by new information contained in the LHC proton-lead cross-sections.

Sum rules. For every nuclei, we assume that the fitted nuclear PDFs satisfy the same valence and momentum sum rules as in the proton case. The sum rules are implemented via an overall normalization factor in the PDF parameterization, which are adjusted each time the neural network parameters are modified in order to ensure that the sum rules remain satisfied. Note that these sum rules need only to be applied at the input scale Q_0, since the properties of DGLAP perturbative evolutions guarantee that they will also be satisfied for other $Q > Q_0$. First, energy conservation leads to the momentum sum rule constraint,

$$\int_0^1 dx\, x \left(\Sigma^{(p/A)}(x, Q_0) + g^{(p/A)}(x, Q_0)\right) = 1\,, \quad \forall\, A\,, \tag{3.7}$$

which is enforced by fixing the gluon normalization to be

$$B_g(A) = \frac{1 - \int_0^1 dx\, x\Sigma^{(p/A)}(x, Q_0)}{\int_0^1 dx\, x g^{(p/A)}(x, Q_0)}\,, \tag{3.8}$$

where the denominator of eq. (3.8) is evaluated using eq. (3.6) and setting $B_g = 1$. Our nuclear PDFs are also constructed to comply with the three valence sum rules that follow from the valence quark quantum numbers of the proton:

$$\int_0^1 dx\, \left(u^{(p/A)}(x, Q_0) - \bar{u}^{(p/A)}(x, Q_0)\right) = 2\,, \quad \forall\, A\,, \tag{3.9}$$

$$\int_0^1 dx\, \left(d^{(p/A)}(x, Q_0) - \bar{d}^{(p/A)}(x, Q_0)\right) = 1\,, \quad \forall\, A\,, \tag{3.10}$$

$$\int_0^1 dx\, \left(s^{(p/A)}(x, Q_0) - \bar{s}^{(p/A)}(x, Q_0)\right) = 0\,, \quad \forall\, A\,, \tag{3.11}$$

where the final relation is trivially satisfied due to our inherent flavor assumption of $s = \bar{s}$.

To implement the former two valence sum rules in our analysis, we first must derive the corresponding constraints in the evolution basis. Adding eqs. (3.9) and (3.10) results in

$$\int_0^1 dx\, \left(u^{(p/A)}(x, Q_0) - \bar{u}^{(p/A)}(x, Q_0) + d^{(p/A)}(x, Q_0) - \bar{d}^{(p/A)}(x, Q_0)\right)$$
$$= \int_0^1 dx\, V^{(p/A)}(x, Q_0) = 3\,, \quad \forall\, A\,. \tag{3.12}$$

This condition can then be implemented in the same way as the momentum sum rule, namely by setting the overall normalization factor of V as

$$B_V(A) = \frac{3}{\int_0^1 dx\, V^{(p/A)}(x, Q_0, A)}\,, \tag{3.13}$$

where the denominator of eq. (3.13) is evaluated using eq. (3.6) and setting $B_V = 1$.

JHEP09(2020)183

The second valence sum rule in the evolution basis is the one related to the quark valence triplet V_3. Subtracting eq. (3.10) from (3.9) gives

$$\int_0^1 dx \, \left(u^{(p/A)}(x,Q_0,A) - \bar{u}^{(p/A)}(x,Q_0,A) - d^{(p/A)}(x,Q_0,A) + \bar{d}^{(p/A)}(x,Q_0,A)\right) = \int_0^1 dx \, V_3^{(p/A)}(x,Q_0,A) = 1\,, \quad \forall\, A\,. \tag{3.14}$$

which again is imposed by setting

$$B_{V_3}(A) = \frac{1}{\int_0^1 dx \, V_3^{(p/A)}(x,Q_0,A)}\,, \tag{3.15}$$

where the denominator of eq. (3.15) is evaluated using eq. (3.6) with $B_{V_3} = 1$.

In this analysis, the normalization pre-factors $B_g(A)$, $B_V(A)$, and $B_{V_3}(A)$ are computed using the trapezoidal rule for numerical integration between $x_{\rm min} = 10^{-9}$ and $x_{\rm max} = 1$ each time the fit parameters are updated by the minimization procedure. With a suitable choice of the ranges for the preprocessing exponents (see discussion below), we guarantee that each quark combination satisfies the corresponding physical integrability conditions. Lastly, we have confirmed that individual replicas satisfy the sum rules with a precision of a few per-mille or better.

An interesting question in the context of nuclear global QCD analyses is the extent to which theoretical constraints such as the sum rules are satisfied by the experimental data when not explicitly imposed. In fact, it was shown in ref. [46] that the momentum sum for the free proton agrees with the QCD expectation within $\sim 1\%$ in this scenario. Here we will revisit this analysis for the nuclear case, and will present in section 4.3 variants of the nNNPDF2.0 fit where either the momentum sum rule, eq. (3.7), or the valence sum rule, eq. (3.12), is not enforced. Interestingly, we will find that the experimental data is in agreement with sum rule expectations, albeit within larger uncertainties, demonstrating the remarkable consistency of the nuclear global QCD analysis.

Preprocessing exponents. The $x^{\alpha_f}(1-x)^{\beta_f}$ polynomial pre-factors appearing in eq. (3.6) are included to increase the efficiency of the parameter optimization, since they approximate the general PDF behavior in the small- and large-x asymptotic limits [73]. Note that the exponents α_f and β_f are A-independent, implying that A dependence of the nPDFs will arise completely from the output of the neural network. As in the case of the nNNPDF1.0 analysis, the values of α_f and β_f are fitted for each Monte Carlo replica on the same footing as the weights and thresholds of the neural network.

The ranges of the preprocessing parameters are determined both by physical considerations and by empirical observations. First of all, the lower limit of the small-x parameter is set so that each PDF combination satisfies various integrability conditions. In particular, the non-singlet combinations xV, xV_3, xT_3, and xT_8 must tend to zero at small-x, else the valence sum rules and other relations such as the Gottfried sum rule [74–76] would be ill-defined. Moreover, the singlet combinations $x\Sigma$ and xg must be exhibit finite integrable behavior as $x \to 0$, otherwise the momentum integral cannot be computed. Concerning

JHEP09(2020)183

the large-x exponents β_f, the lower limits of their ranges ensure that PDFs vanish in the elastic limit, while the upper limit is determined largely from general arguments related to sum rule expectations. In general, however, the upper values of both α_f and β_f are chosen to be sufficiently large to allow flexibility in exploring the parameter space while keeping fit efficiency optimal.

Under these considerations, we restrict the parameter values for the pre-processing factors during the fit to the following intervals,

$$\begin{aligned}
\alpha_\Sigma &\in [-1,5] \quad ([-1,1])\,, & \beta_\Sigma &\in [1,10] \quad ([1,5])\,,\\
\alpha_g &\in [-1,5] \quad ([-1,1])\,, & \beta_g &\in [1,10] \quad ([1,5])\,,\\
\alpha_V &\in [0,5] \quad ([1,2])\,, & \beta_V &\in [1,10] \quad ([1,5])\,,\\
\alpha_{T_8} &\in [-1,5] \quad ([-1,1])\,, & \beta_{T_8} &\in [1,10] \quad ([1,5])\,,\\
\alpha_{V_3} &\in [0,5] \quad ([1,2])\,, & \beta_{V_3} &\in [1,10] \quad ([1,5])\,,\\
\alpha_{T_3} &\in [-1,5] \quad ([-1,1])\,, & \beta_{T_3} &\in [1,10] \quad ([1,5])\,,
\end{aligned} \tag{3.16}$$

where the ranges in parentheses are those used to randomly select the initial values of α_f and β_f at the start of the minimization. We do not impose any specific relation between the small- or large-x exponents of the different quark combinations, so that each are fitted independently. It is also worth emphasizing here that the neural network has the ability to compensate for any deviations in the shape of the preprocessing function, so the dependence on x and A of the nPDFs in the data region will be dominated by the neural network output. This implies that the preprocessing exponents will primarily affect the results in the extrapolation regions.

To illustrate the values of the small and large-x preprocessing exponents preferred by the experimental data, we display in figures 3 and 4 the probability distributions associated with the α_f and β_f exponents, respectively, computed using the $N_{\rm rep} = 1000$ replicas of the nNNPDF2.0 analysis. Note how these exponents are restricted to lie in the interval given by eq. (3.16). For T_3 and T_8, we can see that despite not imposing the strict integrability requirement that $\alpha_f > 0$, it is still being satisfied for a large majority of the replicas, especially for T_3. Interestingly, the gluon seems to prefer a valence-like behavior at small-x. However, such behaviour is only observed at the parameterization scale and as soon as $Q > Q_0$, DGLAP evolution drives it to its expected singlet-like behavior.

Concerning the behavior of the large-x exponents β_f, we find that they are reasonably well constrained for the quark distributions, where the best-fit values are located in a region close to $\beta_f \simeq 3$. The fact that they share similar β_f exponents can be explained by the fact that in the large-x region the quark combinations are dominated by valence components. Interestingly, a best-fit value of $\beta_f \simeq 3$ for the valence quarks is consistent with the expectations from the QCD counting rules [77], as discussed in [73]. Furthermore, the best-fit value for β_g is also consistent with the QCD counting rules prediction of $\beta_g \simeq 5$, although with significant uncertainties. The fact that β_g is found to vary in a wide range is due to the lack of direct constraints on the large-x nuclear gluon PDF in the present analysis.

JHEP09(2020)183

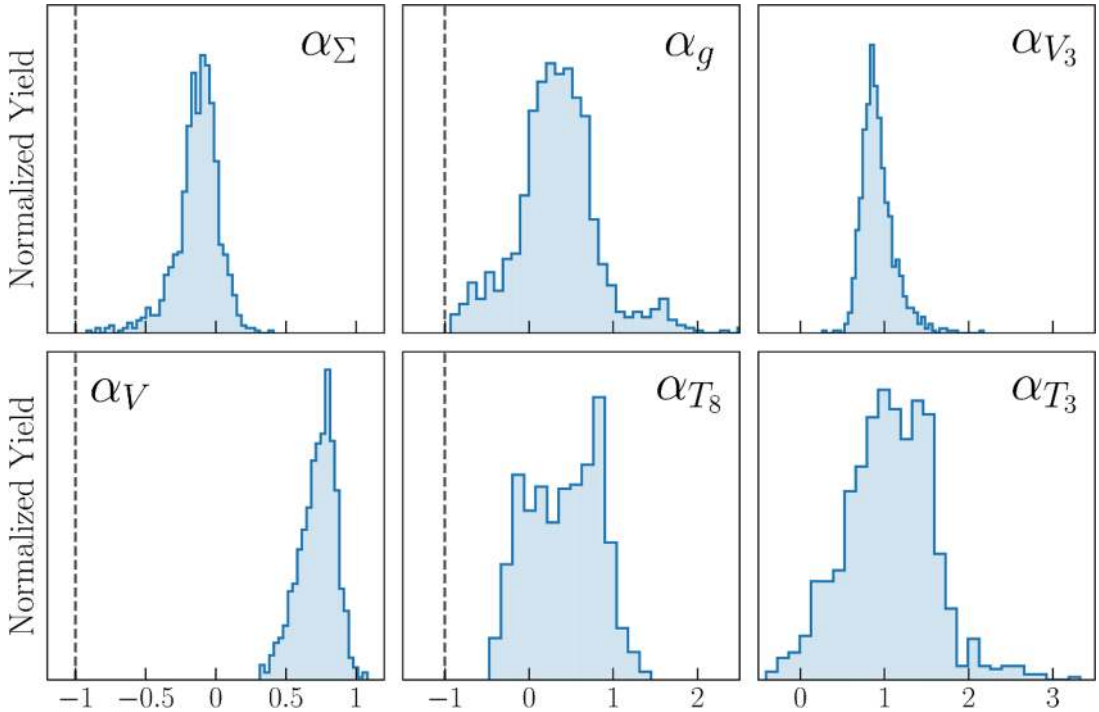

Figure 3. The probability distribution associated to the fitted small-x preprocessing exponents α_f computed with the $N_{\rm rep} = 1000$ replicas of the nNNPDF2.0 NLO set. The ranges for which these exponents are allowed to vary, eq. (3.16), are indicated by horizontal dashed lines. Note the difference in the x-axis range values for each column.

The free-proton boundary condition. As was done in our previous study, we again implement the condition that the proton PDF baseline, obtained with consistent theoretical and methodological choices, is reproduced when $A \to 1$. This condition should be constructed to match the free-proton distributions not only in terms of central values but also at the level of PDF uncertainties. In other words, it should allow a full propagation of the information contained in the proton baseline, which is particularly important to constrain the nPDFs of relatively light nuclei. Note, however, that for the reasons explained above, the nNNPDF2.0 $A = 1$ set will in general not be strictly identical to the corresponding proton baseline.

The proton boundary condition constraint is implemented by adding a quadratic penalty term to the χ^2 of the form

$$\chi^2 \to \chi^2 + \lambda_{\rm BC} \sum_f \sum_{i=1}^{N_x} \left(q_f^{(p/A)}(x_i, Q_0, A=1) - q_f^{(p)}(x_i, Q_0) \right)^2 , \tag{3.17}$$

where the sum over flavors f runs over the six independent elements in the PDF evolution basis. Since the properties of DGLAP evolution ensure that the distributions for $Q > Q_0$ also satisfy the condition, only the PDFs at the parameterization scale Q_0 enter the penalty term. In eq. (3.17), we use a grid with $N_x = 60$ points of which 10 are distributed logarithmically from $x_{\rm min} = 10^{-3}$ to $x_{\rm mid} = 0.1$ and the remaining 50 points are linearly

JHEP09(2020)183

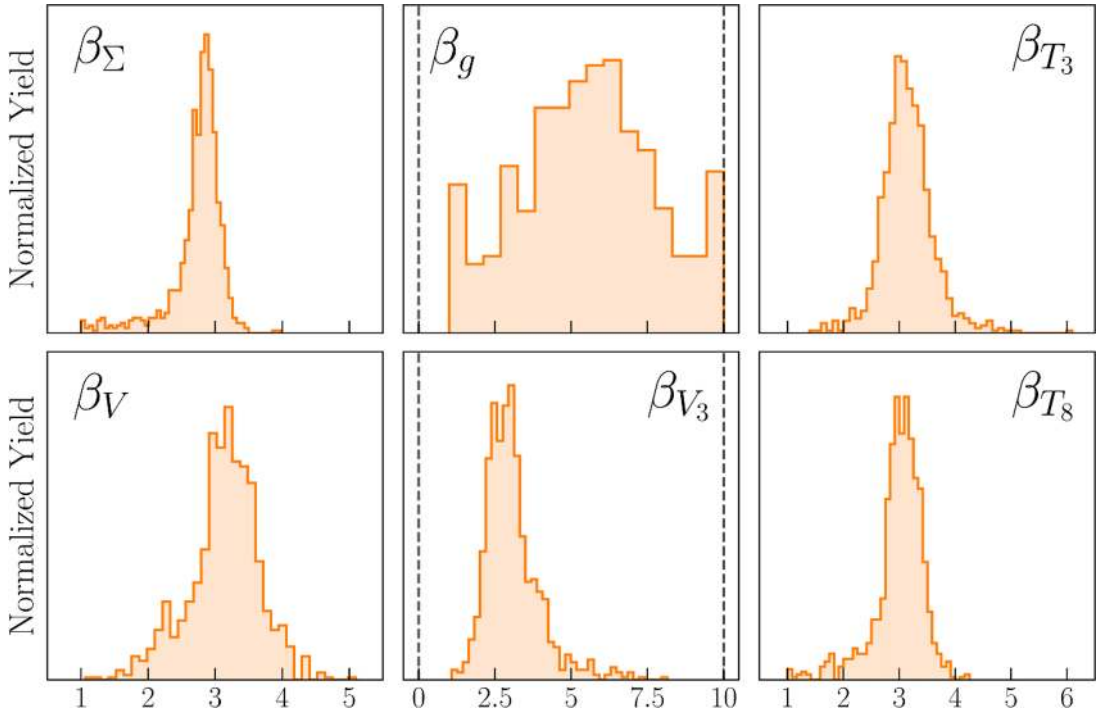

Figure 4. Same as figure 3 for the fitted large-x preprocessing exponents β_f. Note the difference in the x-axis range values for each column.

distributed from $x_{\rm mid} = 0.1$ to $x_{\rm max} = 0.7$. The value of the boundary condition hyper-parameter is fixed to be $\lambda_{\rm BC} = 100$ as was done in the previous nNNPDF1.0 analysis. For such a value, we find that the central values and uncertainties of the proton baseline are reasonably well described (see section 4).

In this analysis the proton baseline, $f^{(p)}(x, Q_0)$, is taken to be a variant of the NNPDF3.1 NLO fit [72] with perturbative charm, where the neutrino DIS cross-sections from NuTeV and CHORUS are removed along with the di-muon production measurements in proton-copper collisions from the E605 experiment [78]. As such, the proton baseline not only avoids double counting of the CC DIS data but also excludes constraints from heavy nuclear target data where nuclear effects are neglected. This choice is different to that used for nNNPDF1.0, where the global NNPDF3.1 fit was used and double-counting of experimental data was not an issue.

In order to illustrate the differences between the free-proton boundary condition used in nNNPDF1.0 and that employed in the present analysis, we compare in figure 5 the NNPDF3.1 NLO global and no heavy nuclear fits at the initial parameterization scale of $Q_0 = 1\,$GeV. Displayed are the gluon, up quark, down sea quark, and total strange PDFs in the range of x with which the proton boundary condition is constrained by eq. (3.17). Here one can see that removing the heavy nuclear data from NNPDF3.1 results in a moderate increase of the PDF uncertainties associated to the quarks as well as an upward shift of the central value of the total strange distribution. The former effect is primarily a consequence of information loss on quark flavor separation with the removal of neutrino scattering

JHEP09(2020)183

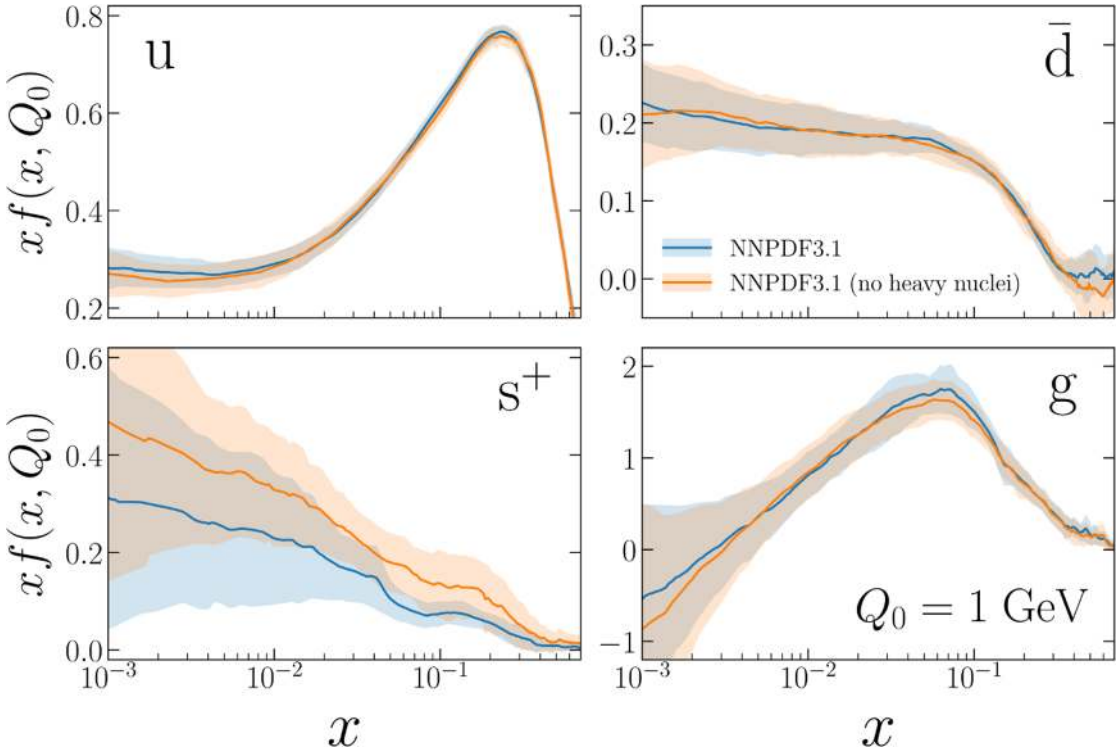

Figure 5. A comparison between the global NNPDF3.1 free-proton analysis with its variant with heavy nuclear data excluded. We show results for the up quark, down antiquark, total strangeness, and the gluon at $Q_0 = 1\,\text{GeV}$. The comparison is presented for the range of x for which the proton boundary condition is implemented in nNNPDF2.0 using eq. (3.17). The PDF uncertainty bands correspond to 90% CL intervals.

measurements. The strangeness feature, on the other hand, arises due to the absence of sensitivity from the NuTeV neutrino dimuon cross-sections, resulting in an upward pull by the ATLAS W, Z 2011 rapidity distributions which are known to produce an enhanced strange with respect to the up and down quark sea. The results of figure 5 highlight the importance of a consistent choice of the free-proton baseline in order to draw solid conclusions on the nuclear modifications, for example those associated to the nucleon's strange content.

In order to ensure that all central values and PDF uncertainties are reproduced, we select a different replica from the NNPDF3.1 proton baseline when constructing eq. (3.17) for each replica of nNNPDF2.0. Since we perform a large N_{rep} number of fits to estimate the uncertainties in nNNPDF2.0, we are able to propagate the necessary information contained in NNPDF3.1 to the resulting nPDFs in a robust manner. Lastly, we note that eq. (3.17) is the only place in the analysis where the free-proton NNPDF3.1 baseline is inserted. In other parts of the fit where a free-nucleon PDF is required, for example in the theoretical predictions of the proton-lead scattering cross-sections, the nNNPDF2.0 set with $A = 1$ is used instead.

JHEP09(2020)183

3.3 Cross-section positivity

While parton distributions are scheme-dependent and thus not necessarily positive-definite beyond leading order in perturbative QCD, physical cross-sections constructed from them are scheme independent and should be positive-definite in the region of validity of the perturbative expansion.[1] In the NNPDF family of proton PDF fits, the requirement that cross-sections remain positive is implemented by adding to the χ^2 a penalty term in the presence of negative cross-section values [6]. The cross-sections that enter this penalty term correspond to theoretical predictions based on pseudo-data generated for representative processes that are directly sensitive to a sufficient number of PDF combinations.

In the nNNPDF1.0 analysis, cross-section positivity was not imposed and led to some observables, such as the longitudinal structure function $F_L(x, Q^2)$, becoming negative at small-x values outside the data region. To bypass this problem, and also to improve the methodological consistency with the free-proton baseline, in nNNPDF2.0 we impose the positivity of physical cross-sections for all nuclei used in the fit by adding a suitable penalty to the figure of merit. In this case, the positivity penalty is expressed as

$$\chi^2 \to \chi^2 + \lambda_{\rm pos} \sum_{l=1}^{N_{\rm pos}} \sum_{j=1}^{N_A} \sum_{i_l=1}^{N_{\rm dat}^{(l)}} \max\left(0, -\mathcal{F}_{i_l}^{(l)}(A_j)\right) , \tag{3.18}$$

for $N_{\rm pos}$ positivity observables $\mathcal{F}^{(l)}$. Each of the observables contain $N_{\rm dat}^{(l)}$ kinematic points that are chosen to cover an adequately large region of phase space relevant to various PDF combinations, as we discuss in more detail below. The computed observables are summed over all N_A nuclei for which we have experimental data, as listed in tables 1 and 2, as well as for the free-proton at $A = 1$. Finally, the value of the hyper-parameter $\lambda_{\rm pos} = 1000$ is determined by manual inspection of the optimization process and is chosen so that positivity is satisfied without distorting the training on the real experimental data.[2]

In table 3 we list the $\mathcal{F}^{(l)}$ processes used in this analysis for which the positivity of physical cross-sections is imposed using eq. (3.18). For each observable, the LO expressions in terms of the average bound nucleon PDFs and bound proton distributions are given together with the number of pseudo-data points $N_{\rm dat}$ and the corresponding kinematic coverage. Note that the LO expressions in table 3 are shown for illustration purposes only, and in our analysis these observables are computed using the full NLO formalism.

Here we consider two types of positivity observables. The first type are the DIS structure functions F_2^u, F_2^d, F_2^s, and F_L. The former three quantities, which contain only u, d, and s contributions, respectively, are constructed to be positive-definite since there exists consistent physical theories where the photon couples only to up-, down-, or strange-type quarks. To further clarify this point, consider such observables from an EFT perspective. For example, in the Standard Model Effective Field Theory (SMEFT) [81] there are four-

[1] A recent study [79] suggests, however, that from a practical point of view PDFs in the $\overline{\rm MS}$-scheme should also satisfy positivity beyond the LO approximation in the perturbative region.

[2] In future work it might be advantageous to determine dynamically the fit hyper-parameters such as $\lambda_{\rm BC}$ and $\lambda_{\rm pos}$ using the hyper-optimization method presented in ref. [80].

JHEP09(2020)183

Observable	LO expression	$N_{\rm dat}$	Kinematic coverage
$F_2^u(x,Q^2,A)$	$\propto \left(u^{N/A}+\bar{u}^{N/A}\right)$ $\propto \left[(Z/A)\left(u^{p/A}+\bar{u}^{p/A}\right)+(1-Z/A)\left(d^{p/A}+\bar{d}^{p/A}\right)\right]$	20	$Q^2=5\,\text{GeV}$ $5\times10^{-7}\leq x\leq 0.9$
$F_2^d(x,Q^2,A)$	$\propto \left(d^{N/A}+\bar{d}^{N/A}\right)$ $\propto \left[(Z/A)\left(d^{p/A}+\bar{d}^{p/A}\right)+(1-Z/A)\left(u^{p/A}+\bar{u}^{p/A}\right)\right]$	20	$Q^2=5\,\text{GeV}$ $5\times10^{-7}\leq x\leq 0.9$
$F_2^s(x,Q^2,A)$	$\propto \left(s^{N/A}+\bar{s}^{N/A}\right)$ $\propto \left(s^{p/A}+\bar{s}^{p/A}\right)$	20	$Q^2=5\,\text{GeV}$ $5\times10^{-7}\leq x\leq 0.7$
$F_L(x,Q^2,A)$	sensitive to $xg(x,Q^2)$ (see text)	20	$Q^2=5\,\text{GeV}$ $5\times10^{-7}\leq x\leq 0.9$
$\sigma^{DY}_{u\bar{u}}(y,M^2,A)$	$\propto \left(u^p(x_1)\times\bar{u}^{N/A}(x_2)\right)$ $\propto \left(u^p(x_1)\times\left(Z\bar{u}^{p/A}(x_2)+(A-Z)\bar{d}^{p/A}(x_2)\right)\right)$	20	$Q^2=5\,\text{GeV}$ $10^{-2}\leq x\leq 0.9$
$\sigma^{DY}_{d\bar{d}}(y,M^2,A)$	$\propto \left(d^p(x_1)\times\bar{d}^{N/A}(x_2)\right)$ $\propto \left(d^p(x_1)\times\left(Z\bar{d}^{p/A}(x_2)+(A-Z)\bar{u}^{p/A}(x_2)\right)\right)$	20	$Q^2=5\,\text{GeV}$ $10^{-2}\leq x\leq 0.9$
$\sigma^{DY}_{s\bar{s}}(y,M^2,A)$	$\propto \left(s^p(x_1)\times\bar{s}^{N/A}(x_2)\right)$ $\propto \left(s^p(x_1)\times\bar{s}^{p/A}(x_2)\right)$	20	$Q^2=5\,\text{GeV}$ $10^{-2}\leq x\leq 0.9$
$\sigma^{DY}_{\bar{u}d}(y,M^2,A)$	$\propto \left(\bar{u}^p(x_1)\times d^{N/A}(x_2)\right)$ $\propto \left(\bar{u}^p(x_1)\times\left(Zd^{p/A}(x_2)+(A-Z)u^{p/A}(x_2)\right)\right)$	20	$Q^2=5\,\text{GeV}$ $10^{-2}\leq x\leq 0.9$
$\sigma^{DY}_{\bar{d}u}(y,M^2,A)$	$\propto \left(\bar{d}^p(x_1)\times u^{N/A}(x_2)\right)$ $\propto \left(\bar{d}^p(x_1)\times\left(Zu^{p/A}(x_2)+(A-Z)d^{p/A}(x_2)\right)\right)$	20	$Q^2=5\,\text{GeV}$ $10^{-2}\leq x\leq 0.9$
$\sigma^{DY}_{u\bar{s}}(y,M^2,A)$	$\propto \left(u^p(x_1)\times\bar{s}^{N/A}(x_2)\right)$ $\propto \left(u^p(x_1)\times\bar{s}^{p/A}(x_2)\right)$	20	$Q^2=5\,\text{GeV}$ $10^{-2}\leq x\leq 0.9$
$\sigma^{DY}_{\bar{u}s}(y,M^2,A)$	$\propto \left(\bar{u}^p(x_1)\times s^{N/A}(x_2)\right)$ $\propto \left(\bar{u}^p(x_1)\times s^{p/A}(x_2)\right)$	20	$Q^2=5\,\text{GeV}$ $10^{-2}\leq x\leq 0.9$

Table 3. The processes used to impose the positivity of physical cross-sections by means of the constraint of eq. (3.18). For each process we indicate the corresponding LO expressions, the number of data points $N_{\rm dat}$, and the kinematic coverage spanned by the pseudo-data.

fermion operators at dimension 6 of the form

$$\mathcal{L}_{\rm SMEFT}\supset\frac{a_{q_il_j}}{\Lambda^2}\left(\bar{q}_i\Gamma_\mu q_i\right)\left(\bar{l}_j\Gamma^\mu l_j\right), \tag{3.19}$$

which describes the 4-point interaction of a quark of flavor i with a lepton of type j. Here, Λ corresponds to the energy scale in which the EFT expansion becomes invalid, and Γ^μ represents the general Lorentz structure for the interaction. Since any theory with a single non-zero coupling $a_{q_il_j}$ is theoretically consistent [82], any cross-sections computed with it must be positive definite, including the ones that lead to our positivity observables. Then one can take the $\Lambda\to\infty$ limit, in which case the cross section will be vanishingly small, yet still remain positive, as we demand in our fit via the positivity constraint.

JHEP09(2020)183

While the positivity of the quark structure functions impact the various quark PDFs, the longitudinal structure function F_L largely impacts the nuclear gluon PDF since F_L enters only at NLO and is dominated by the gluon contribution. Together, we evaluate each of the structure functions on a grid of $N = 20$ pseudo-data points between $x = 10^{-7}$ and $x = 0.9$ at $Q = \sqrt{5}\,\text{GeV}$, which is slightly above the input parameterization scale $Q_0 = 1\,\text{GeV}$ to ensure perturbative stability.

The second type of observable for which the cross-section positivity is imposed is the double-differential Drell-Yan cross-section. In particular, we enforce the positivity of both neutral- and charged-current Drell-Yan cross-sections in pA scattering for specific combinations of quark-antiquark annihilation listed in table 3. At leading order, the Drell-Yan cross-section can be written schematically as

$$\frac{d^2\sigma^{DY}_{q_{f_1}\bar{q}_{f_2}}}{dydQ^2} \propto \left(f_1^{(p)}(x_1,Q^2)\bar{f}_2^{\,(p/A)}(x_2,Q^2)\right) , \tag{3.20}$$

where the momentum fractions x_1 and x_2 are related to the rapidity y and invariant mass of the final state Q at leading order by $x_{1,2} = (Q/\sqrt{s})\,e^{\pm y}$. Here we set $Q^2 = 5\,\text{GeV}^2$ and adjust the rapidity range and center-of-mass energy $\sqrt{s}$ so that the LO kinematic range for x_1 and x_2 correspond to $10^{-2} \leq x_{1,2} \leq 0.9$.

Note that positivity of eq. (3.20) will affect also the fitted nNNPDF2.0 $A = 1$ distribution which enters as the free-proton PDF. While most of the positivity observables coincide with those included in the free-proton baseline from which the $A = 1$ distribution is derived, the $u\bar{d}$, $\bar{u}d$, $u\bar{s}$, and $\bar{u}s$ combinations of table 3 are new in the nNNPDF2.0 determination. Consequently, we impose these new observables only for proton-iron and proton-lead collisions, the two nuclei for which experimental data from charged-current DIS and Drell-Yan reactions are analyzed to study quark flavor separation.

In section 4.4 we will demonstrate the positivity of cross-sections for all relevant processes in the entire kinematical range. We have verified that, in the absence of these constraints, the DIS structure functions and the DY cross-section will in general not satisfy positivity.

4 Results

In this section we present the main results of this work, namely the nNNPDF2.0 determination of nuclear PDFs. We first study the features of the nNNPDF2.0 fit by assessing the quality of its agreement with experimental data, focusing largely on the LHC weak boson production cross-sections, and by studying the behavior of nuclear modification ratios across different nuclei. Subsequently, we contrast this new nPDF determination with its predecessor, nNNPDF1.0, and trace the origin of observed differences via a series of fits with systematic changes. We then study the role that the valence and momentum sum rules play in the global nPDF determination by presenting two variants of the nNNPDF2.0 fit in which one of the two sum rules is not imposed. Finally, we demonstrate that the nNNPDF2.0 fit satisfies the positivity of physical cross-sections in the kinematic range where experimental data is available.

JHEP09(2020)183

4.1 The nNNPDF2.0 determination

We begin by discussing the fit quality which is assessed across the various datasets and quantified by the χ^2 figure of merit. A comparison is then made using the nNNPDF2.0 predictions with the LHC weak gauge boson production measurements. Following this, we take a closer look on the nNNPDF2.0 parton distributions and the corresponding ratios to the free-nucleon case. Lastly, we investigate the sensitivity of the nuclear modification factors with respect to the atomic mass A, in particular on the sea quarks and strangeness, and compare our results with those from the EPPS16 analysis.

Fit quality. In tables 4 and 5 we collect the values of the χ^2 per data point for all the datasets included in the nNNPDF2.0 analysis, i.e. the neutral and charged current DIS structure functions as well as gauge boson production measurements at the LHC. We compare the nNNPDF2.0 results with a variant fit where we exclude all LHC datasets (DIS only) and with EPPS16.[3] Values in italics indicate predictions for datasets not included in the corresponding fits. The numbers presented in tables 4 and 5 contain only the contribution to the χ^2 associated with the experimental data, and do not include penalty from the proton boundary condition or positivity constraint (the latter of which vanishes for all final nNNPDF2.0 replicas anyway). Moreover, while we use the t_0 prescription [43] during the optimization to avoid the D'Agostini bias, the quoted numbers correspond to the experimental definition of the χ^2 instead [83], in which the central experimental value is used to compute the correlated multiplicative uncertainties.

From the results of table 4 and 5, one can see that the nNNPDF2.0 determination achieves a satisfactory description of all datasets included in this analysis. A good χ^2 is obtained in particular for the charged-current DIS cross-sections and LHC gauge boson production distributions. For instance, for the precise W boson rapidity distributions at $\sqrt{s} = 8.16\,\text{TeV}$ from CMS one finds $\chi^2/n_{\rm dat} = 0.74$ for $n_{\rm dat} = 48$ data points. The corresponding predictions using EPPS16 (which do not include this dataset) also lead to a good agreement with $\chi^2/n_{\rm dat} = 0.88$. The description of most neutral current DIS datasets is comparable to that of nNNPDF1.0. Some datasets, such as the SLAC iron structure functions, are somewhat deteriorated with respect to nNNPDF1.0, possibly due to some mild tension with the CC cross-sections. Further, we find that our resulting fit quality to the CC deep-inelastic structure functions is similar to that obtained in the corresponding proton PDF analysis [72].

Overall, the resulting $\chi^2_{\rm tot}/n_{\rm dat} = 0.976$ for the $n_{\rm dat} = 1467$ data points included in the fit highlights the remarkable consistency of the experimental data on nuclear targets and the corresponding theory predictions based on the QCD factorization framework. A similar total χ^2 is obtained for the theoretical predictions computed using EPPS16 as input. Note that while our parameterization is more flexible than that used in the EPPS analysis, we also account for a number of constraints such as the positivity constraint and boundary condition, which restricts the range of functional forms available. In any case, the fact

[3] For the EPPS16 calculation we use CT14nlo as the free-proton PDF set for consistency.

JHEP09(2020)183

		nNNPDF2.0 (DIS)	nNNPDF2.0	EPPS16nlo
Dataset	$n_{\rm dat}$	$\chi^2/n_{\rm dat}$	$\chi^2/n_{\rm dat}$	$\chi^2/n_{\rm dat}$
NMC (He/D)	13	1.11	1.129	0.829
SLAC (He/D)	3	0.623	0.638	0.152
NMC (Li/D)	12	1.083	1.166	0.74
SLAC (Be/D)	3	1.579	1.719	0.098
EMC (C/D)	12	1.292	1.321	1.174
FNAL (C/D)	3	0.932	0.838	0.985
NMC (C/D)	26	2.002	2.171	0.872
SLAC (C/D)	2	0.286	0.251	1.075
BCDMS (N/D)	9	2.439	2.635	n/a
SLAC (Al/D)	3	0.606	0.864	0.326
EMC (Ca/D)	3	1.72	1.722	1.82
FNAL (Ca/D)	3	1.253	1.194	1.354
NMC (Ca/D)	12	1.503	1.747	1.772
SLAC (Ca/D)	2	0.82	0.771	1.642
BCDMS (Fe/D)	16	2.244	2.743	*0.765*
EMC (Fe/D)	58	0.827	0.875	0.445
SLAC (Fe/D)	8	2.171	2.455	1.06
EMC (Cu/D)	27	0.523	0.572	0.714
SLAC (Ag/D)	2	0.667	0.691	1.595
EMC (Sn/D)	8	2.197	2.248	2.265
FNAL (Xe/D)	4	0.414	0.384	n/a
SLAC (Au/D)	3	1.216	1.353	1.916
FNAL (Pb/D)	3	2.243	2.168	2.044
NMC (Be/C)	14	0.268	0.269	0.27
NMC (C/Li)	9	1.063	1.117	0.9
NMC (Al/C)	14	0.345	0.354	0.396
NMC (Ca/C)	23	0.468	0.44	0.564
NMC (Fe/C)	14	0.663	0.667	0.751
NMC (Sn/C)	119	0.607	0.638	0.626
NMC (Ca/Li)	9	0.259	0.276	0.15

Table 4. The values of the χ^2 per data point for the DIS neutral current structure function datasets included in nNNPDF2.0. We compare the nNNPDF2.0 global and DIS-only results with those obtained using EPPS16 as input for the theory predictions.

that both global fits lead to comparable χ^2 values can be explained by the corresponding similarities at the nPDF level, as will be shown below.

Comparison with experimental data. To facilitate our discussion regarding the comparison between data and theory calculations, we first introduce here the conventions that we use to define the nuclear modification factors. Following the notations of section 3.1, the Drell-Yan rapidity distributions in proton-nucleus collision can be expressed as

$$\frac{d\sigma_{\rm DY}(y)}{dy} \equiv A\frac{d\sigma_{\rm DY}^{(N/A)}(y)}{dy} = Z\frac{d\sigma_{\rm DY}^{(p/A)}(y)}{dy} + (A-Z)\,\frac{d\sigma_{\rm DY}^{(n/A)}(y)}{dy}\,, \tag{4.1}$$

JHEP09(2020)183

		nNNPDF2.0 (DIS)	nNNPDF2.0	EPPS16nlo
Dataset	$n_{\rm dat}$	$\chi^2/n_{\rm dat}$	$\chi^2/n_{\rm dat}$	$\chi^2/n_{\rm dat}$
NuTeV ($\bar{\nu}$Fe)	37	0.946	1.094	*0.639*
NuTeV (νFe)	39	0.287	0.264	*0.381*
CHORUS ($\bar{\nu}$Pb)	423	0.938	0.97	1.107
CHORUS (νPb)	423	1.007	1.015	1.024
ATLAS$^{5\rm TEV}$ Z	14	*1.469*	1.134	1.12
CMS$^{5\rm TeV}$ W$^-$	10	*1.688*	1.078	0.857
CMS$^{8\rm TeV}$ W$^-$	24	*1.453*	0.72	*0.825*
CMS$^{5\rm TeV}$ W$^+$	10	*2.32*	1.125	1.211
CMS$^{8\rm TeV}$ W$^+$	24	*3.622*	0.772	*0.951*
CMS$^{5\rm TeV}$ Z	12	*0.58*	0.52	0.639
Total	**1467**	**1.013**	**0.976**	

Table 5. Same as table 4 now for the new datasets included in nNNPDF2.0: charged current DIS structure functions and gauge boson production at the LHC. We also provide the values of $\chi^2/n_{\rm dat}$ for the total dataset. Values in italics indicate predictions for datasets not included in the corresponding fit.

where the superindices N/A, p/A, and n/A indicate respectively the collision between a proton with an average nucleon, a proton, or a neutron bound within a nucleus of atomic number Z and mass number A. As in the case of the PDFs, the bound proton and nucleon cross-sections $\sigma_{\rm DY}^{(p/A)}$ and $\sigma_{\rm DY}^{(n/A)}$ are related to each other via isospin symmetry.

The expression in eq. (4.1) helps in emphasizing the two reasons why the Drell-Yan cross-sections will be different between pp and pA collisions. The first is due to the modifications of the bound proton PDFs in nuclei, namely the fact that $\sigma_{\rm DY}^{(p/A)} \neq \sigma_{\rm DY}^{(p)}$. Secondly, the predictions of pA collisions using non-isoscalar nuclei would still differ from those of pp reactions in the absence of nuclear modifications, i.e. $\sigma_{\rm DY}^{(p/A)} = \sigma_{\rm DY}^{(p)}$, as a consequence of the unequal amount of protons and neutrons in the target, resulting in $\sigma_{\rm DY}^{(N/A)} \neq \sigma_{\rm DY}^{(p)}$.

Taking these considerations into account, one should define the nuclear modification factor in Drell-Yan proton-nuclear collisions as

$$R_A^{\rm DY}(y) \equiv \left(\frac{d\sigma_{\rm DY}^{(p/A)}(y)}{dy} + \left(\frac{A}{Z} - 1\right) \frac{d\sigma_{\rm DY}^{(n/A)}(y)}{dy} \right) \Bigg/ \left(\frac{d\sigma_{\rm DY}^{(p)}(y)}{dy} + \left(\frac{A}{Z} - 1\right) \frac{d\sigma_{\rm DY}^{(n)}(y)}{dy} \right) . \tag{4.2}$$

With the above definition, one should find that $R_A^{\rm DY}(y) \neq 1$ only in the presence of genuine nuclear corrections, that is, when $\sigma_{\rm DY}^{(p/A)} \neq \sigma_{\rm DY}^{(p)}$. The definition of eq. (4.2) differs from that of an observable frequently measured in proton-lead collisions, where the proton-nucleus cross-section is normalized to a proton-proton baseline,

$$\begin{aligned} R_{A,\rm exp}^{\rm DY}(y) &\equiv \frac{d\sigma_{\rm DY}^{(N/A)}(y)}{dy} \Bigg/ \frac{d\sigma_{\rm DY}^{(p)}(y)}{dy} \\ &= \left(\frac{Z}{A} \frac{d\sigma_{\rm DY}^{(p/A)}(y)}{dy} + \left(1 - \frac{Z}{A}\right) \frac{d\sigma_{\rm DY}^{(n/A)}(y)}{dy} \right) \Bigg/ \left(\frac{d\sigma_{\rm DY}^{(p)}(y)}{dy} \right) . \end{aligned} \tag{4.3}$$

JHEP09(2020)183

As explained above, in proton-nuclear collisions one will find $R^{\rm DY}_{A,\rm exp}(y) \neq 1$ for non-isoscalar targets even when $\sigma^{(p/A)}_{\rm DY} = \sigma^{(p)}_{\rm DY}$ due to the imbalance between the number of protons and neutrons. In this section, we will exclusively use the definition of eq. (4.2) when evaluating the theoretical predictions of nuclear modification ratios in Drell-Yan distributions.

In figure 6 we display the comparison between the ATLAS and CMS measurements of Z boson production in proton-lead collisions at $\sqrt{s} = 5.02\,$TeV with the theoretical predictions using nNNPDF2.0. The theory calculations were computed using eq. (4.1) with the nNNPDF2.0 $A = 1$ and $A = 208$ distributions, in the latter case also including the corresponding 90% CL uncertainty band. Note that for the theory cross-section obtained with the $A = 1$ PDFs, nuclear effects are absent since it corresponds to the free proton distributions. From top to bottom, the three panels display the absolute cross-sections as a function of the Z rapidity y, the ratio between data and theory, and the nuclear modification factor ratio $R_A(y)$ defined by eq. (4.2).

Here the ATLAS and CMS measurements of the Z rapidity distributions are both provided in the center-of-mass reference frame of the proton-lead collision. The CMS absolute cross-sections are lower than ATLAS due to the different kinematical selection cuts. We also note that for these datasets, as well as for the rest of LHC measurements, forward rapidities correspond to the direction of the incoming proton. From the comparisons in figure 6 we can see that the theory predictions are in good agreement with the experimental data. Interestingly, the $R_A(y)$ ratios exhibit a strong preference for nuclear modifications, especially at forward rapidities which correspond to small values in x for the bound nucleons. As will be discussed below, this behavior can be explained at the level of the nuclear PDFs by a notable shadowing effect at small-x for up and down quarks and antiquarks.

The corresponding comparisons for the CMS muon rapidity distributions in W^- and W^+ production at $\sqrt{s} = 5.02\,$TeV and 8.16 TeV are displayed in figures 7 and 8 respectively. The results are presented as functions of the rapidity of the charged lepton from the W boson decay in the laboratory frame. In all cases the theoretical predictions based on nNNPDF2.0 provide a satisfactory description of the experimental data. It is interesting to note that for the high-statistics CMS measurement at 8 TeV, the $A = 208$ prediction is remarkably better than the free-proton one, particularly at forward rapidities where one is sensitive to the small-x region of the bound nucleons. This feature highlights how the CMS 8 TeV W production data provides direct evidence for the nuclear modifications of valence and sea quark distributions.

Nuclear parton distributions. In figure 9 we display the nNNPDF2.0 set of nuclear PDFs for three different nuclei, ^{12}C, ^{56}Fe, and ^{208}Pb, constructed using eq. (3.2) at a scale of $Q = 10\,$GeV. Specifically, we display the gluon, the up and down valence quarks, and the down, strange, and charm sea quark distributions. For isoscalar nuclei such as ^{12}C, the up and down valence distributions are equivalent, $u_v^{(N/A)} = d_v^{(N/A)}$, as well as the up and down sea PDFs, $\bar{u}^{(N/A)} = \bar{d}^{(N/A)}$, as a result of eq. (3.2).

From the comparisons in figure 9, we can see that the nuclear PDFs exhibit a moderate dependence on the atomic number A. The resulting pattern of PDF uncertainties can partly be explained by the input data. For example, nPDF uncertainties on strangeness

JHEP09(2020)183

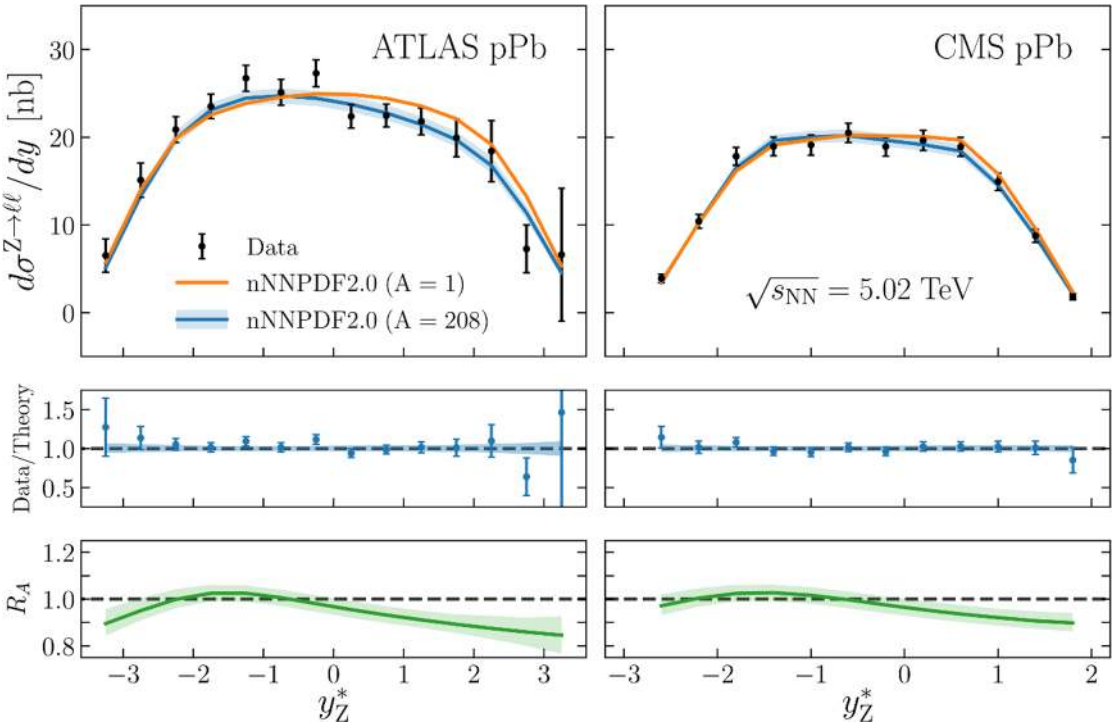

Figure 6. Comparison between the ATLAS (left) and CMS (right panel) measurements of Z boson production in proton-lead collisions at $\sqrt{s} = 5.02$ TeV with the theoretical predictions using nNNPDF2.0 as input. We show the predictions obtained both for $A = 1$ and $A = 208$, in the later case including the 90% confidence level band. From top to bottom, the three panels display the absolute cross-sections, the ratio between data and theory, and the nuclear modification $R_A(y)$.

are smaller in ^{12}C and ^{56}Fe compared to ^{208}Pb, due to the impact of the proton boundary condition and the NuTeV dimuon data, respectively. We also observe that the PDF uncertainties on the gluon (and correspondingly on the dynamically generated charm PDF) at medium and small-x are larger in iron than in carbon and lead. While the gluon uncertainties for carbon are largely determined by the impact of the free-proton boundary condition, those on lead nuclei can likely be attributed to the LHC measurements of W and Z production and the large amount of charged-current DIS data, which indirectly provide constraints via DGLAP evolution.

Figure 9 also shows that in the case of ^{208}Pb, there is a clear difference between the d_v and u_v distributions due to the isoscalar nature of the nucleus, where d_v is larger due to the significant neutron excess in lead. The fact that u_v and d_v do not overlap within the 90% CL bands in a wide range of x highlights how a careful treatment of the quark and antiquark flavor separation is essential in order to describe the precise data available on lead targets, especially the weak boson production measurements in proton-lead collisions at the LHC.

To further illustrate the features of the nNNPDF2.0 determination, it is useful to study them in terms of ratios with respect to the corresponding free-nucleon baseline. In the following we will define the nuclear modification ratios of PDFs for a nucleus with mass

JHEP09(2020)183

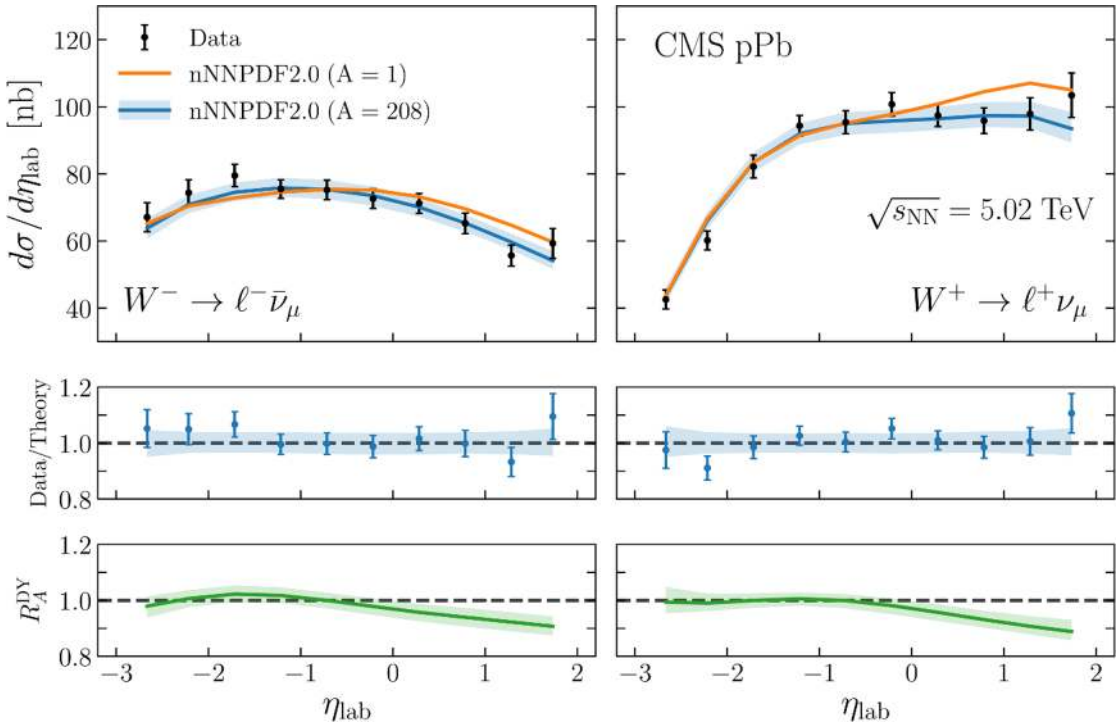

Figure 7. Same as figure 6 now for the CMS muon rapidity distributions in W^- and W^+ production at $\sqrt{s} = 5.02$ TeV.

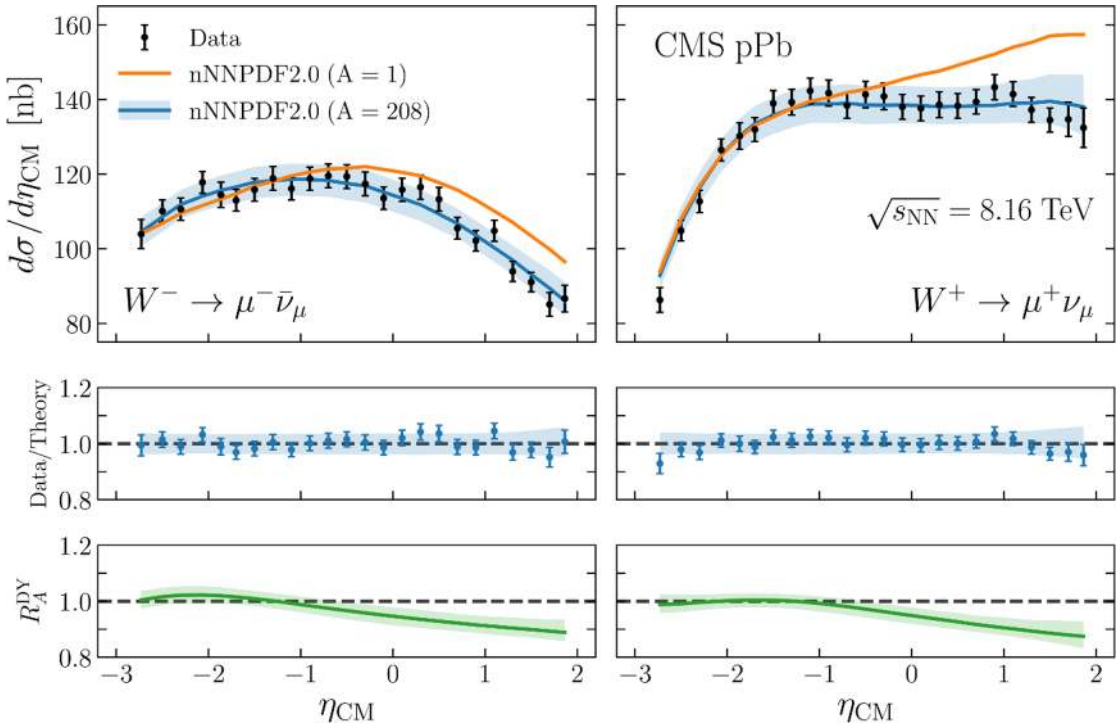

Figure 8. Same as figure 6 now for the CMS muon rapidity distributions in W^- and W^+ production at $\sqrt{s} = 8.16$ TeV.

JHEP09(2020)183

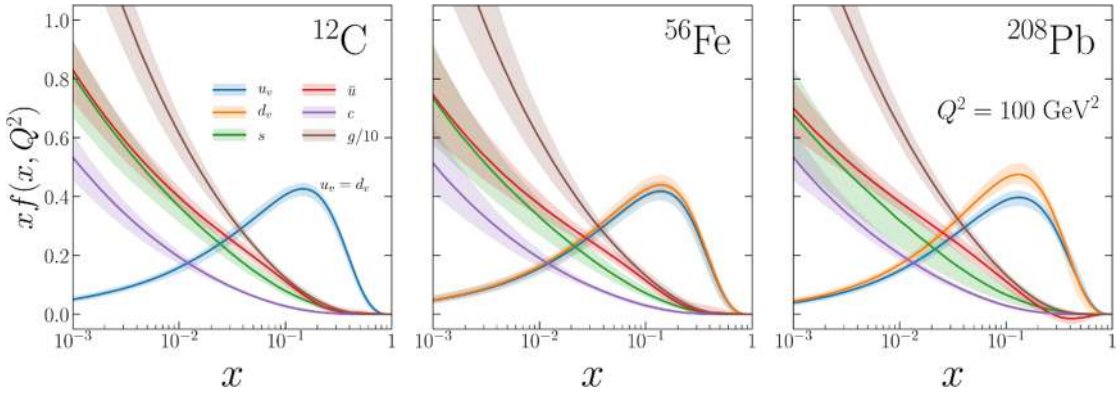

Figure 9. The nNNPDF2.0 set of nuclear PDFs for ^{12}C, ^{56}Fe, and ^{208}Pb at the scale $Q = 10\,\text{GeV}$. We display the gluon, the up and down valence quarks (which coincide for isoscalar nuclei), as well as the down, strange, and charm sea quark distributions. The bands indicate the 90% CL uncertainty range.

number A as,

$$R_f^A(x, Q^2) \equiv \frac{Zq_f^{(p/A)}(x, Q^2) + (A - Z)q_f^{(n/A)}(x, Q^2)}{Zq_f^{(p)}(x, Q^2) + (A - Z)q_f^{(n)}(x, Q^2)} \,. \tag{4.4}$$

When evaluating eq. (4.4), it is important to account both for the uncertainties associated to the nuclear and free-proton PDFs. In the case of a Monte Carlo set such as nNNPDF2.0, this entails evaluating R_f^A for each of the N_{rep} replicas and then determining the resulting median and 90% CL interval.

In figure 10 we show the nNNPDF2.0 distributions for $A = 1$ that enter the denominator of eq. (4.4). They are compared with the NNPDF3.1 proton baseline used for implementation the boundary condition via eq. (3.17) at the input parametrisation scale $Q_0 = 1\,\text{GeV}$. Overall, there is very good agreement between our $A = 1$ result and the proton boundary condition, particularly in the region of x where the constraint is being imposed, $10^{-3} < x < 0.7$. It is important to emphasize that due to the nNNPDF2.0 $A = 1$ set being determined not only by the boundary condition but also by the positivity constraints and the LHC cross-sections, one expects some moderate differences with the NNPDF3.1 proton baseline. The most notable differences indeed are found in the $\bar{u}$ and $\bar{d}$ PDFs at medium to large x, where the newly added DY positivity observables for $\bar{u}d$ and $u\bar{d}$ quark combinations, as well as the LHC proton-lead data, play a significant role. Nevertheless, the level of agreement reported in figure 10 is quite remarkable and highlights how the nNNPDF2.0 determination manages to take into account the extensive information provided by the global analysis of free-proton structure.

In figure 11 we display the nuclear PDF ratios, defined by eq. (4.4), for the same parton flavors as in figure 10. Here the ratios for ^{12}C, ^{56}Fe, and ^{208}Pb nuclei are compared at $Q = 10\,\text{GeV}$. The shaded bands indicating the 90% confidence level intervals take into account also the correlations with the proton baseline. Overall, the comparison highlights the dependence on the central value and uncertainties of the nuclear ratios R_f^A as the value of A is varied from lighter to heavier nuclei.

JHEP09(2020)183

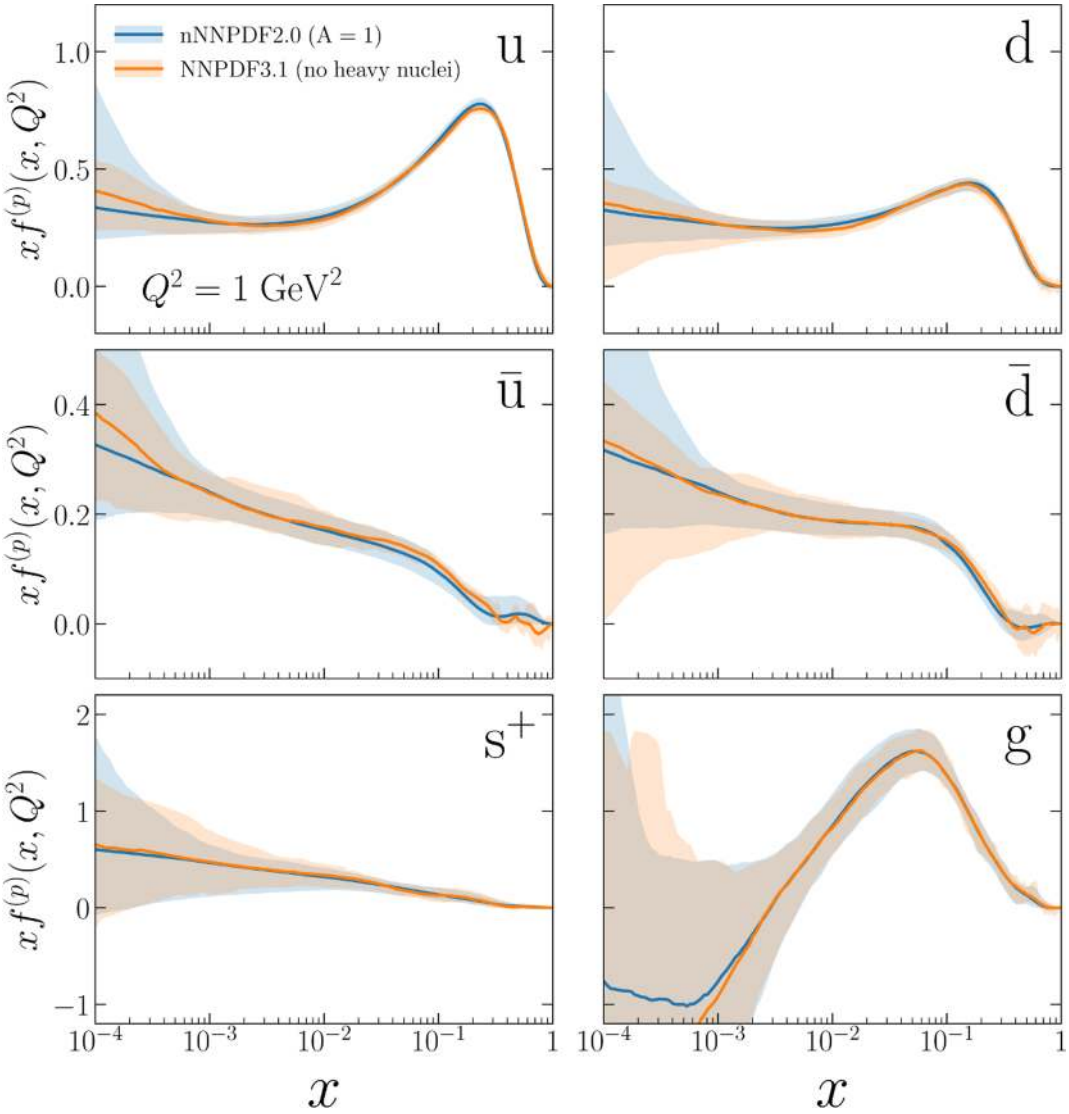

Figure 10. Comparison of the nNNPDF2.0 parton distributions for $A = 1$ with the NNPDF3.1 baseline used for the boundary condition in eq. (3.17) at the parametrisation scale $Q_0 = 1\,\text{GeV}$.

For the up and down quark nPDFs in figure 11, we can see that the shadowing effects become more prominent at small-x as A increases, with the central value reaching $R_f^A \simeq 0.75$ at $x = 10^{-4}$ for the lead ratios. Interestingly, the nPDF uncertainties on the quarks for $x \lesssim 10^{-2}$ are reduced in lead as compared to the lighter nuclei. This is a consequence of the constraints provided by the LHC data, as will be shown in section 4.2. In the large-x region, deviations from the $R_f^A = 1$ scenario (no nuclear corrections) appear more prominent for the quarks and antiquarks of heavier nuclei.

Turning now to the valence quarks, one can distinguish the shadowing and anti-shadowing regions for all values of A, though nuclear effects in carbon are quite small. While one generally finds a suppression $R_f^A < 1$ at large x that is consistent with the EMC-effect expectation, the position of the so-called "EMC minimum" is not universal

JHEP09(2020)183

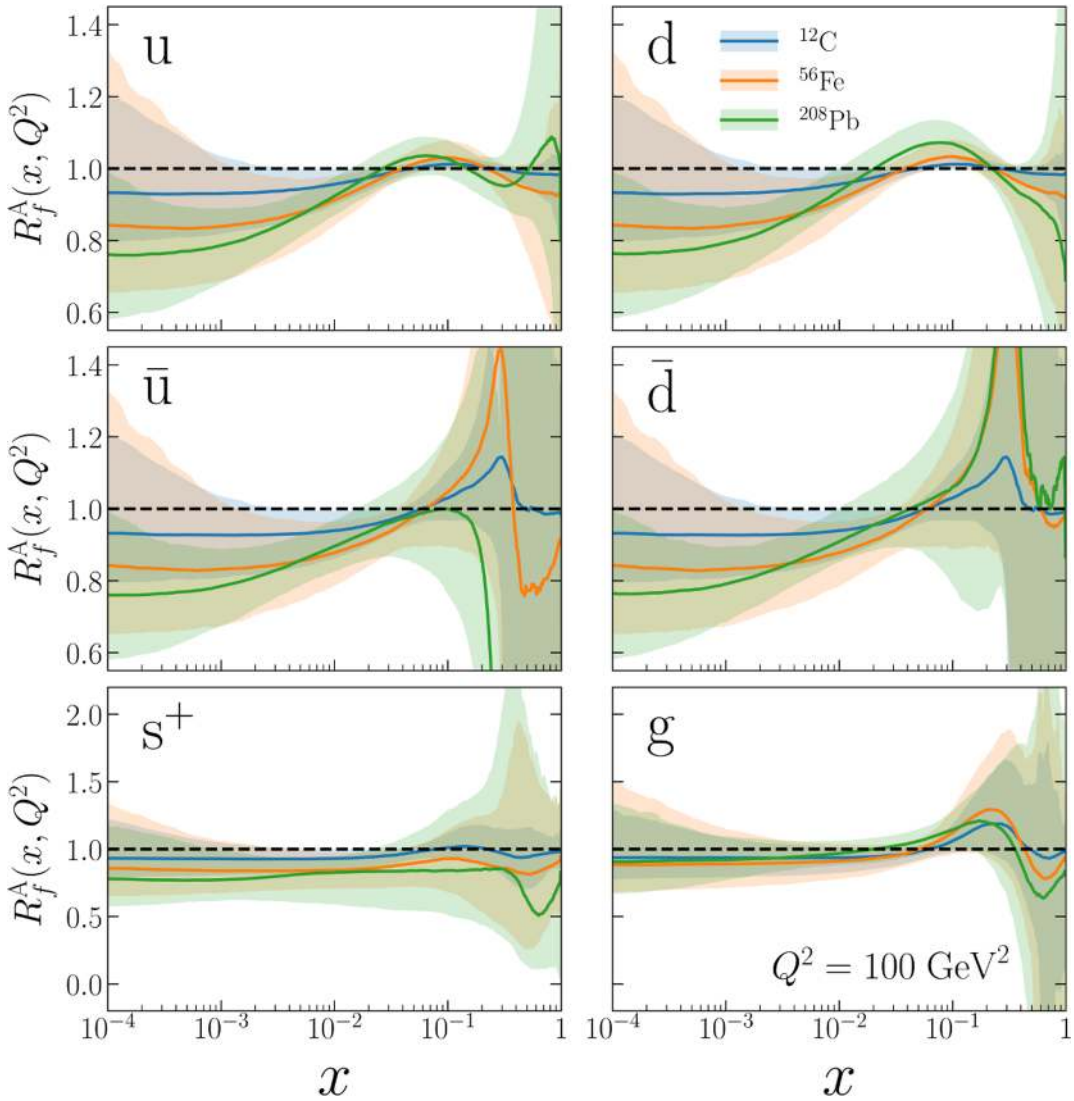

Figure 11. Comparison of the nuclear PDF ratios, eq. (4.4), for three different nuclei, ^{12}C, ^{56}Fe, and ^{208}Pb, at $Q = 10\,\text{GeV}$. From top to bottom we show the up and down quarks, the corresponding antiquarks, the total strangeness, and the gluon. The bands indicate the 90% confidence level intervals and take into account the correlations with the proton baseline used for the normalization.

or even guaranteed at the nPDF level. For the anti-quarks, the only region where a well-defined qualitative behavior is observed is the small-x shadowing region, while at large-x the nPDF uncertainties are too large to draw any solid conclusion. Finally, we observe that the nuclear modifications on the gluon nPDF are rather stable as A is varied.

To illustrate further the dependence of nuclear modifications on the atomic mass A, we show in figure 12 the ratios R_f as a function of A for three values of x at $Q = 10\,\text{GeV}$. Overall, the ratios are stable and exhibit smooth dependence for all parton flavors across the A values that are included in the fit. For the up and down quark ratios, the transition from anti-shadowing to shadowing is clearly visible as x is decreased from 0.1 to 0.001. We

JHEP09(2020)183

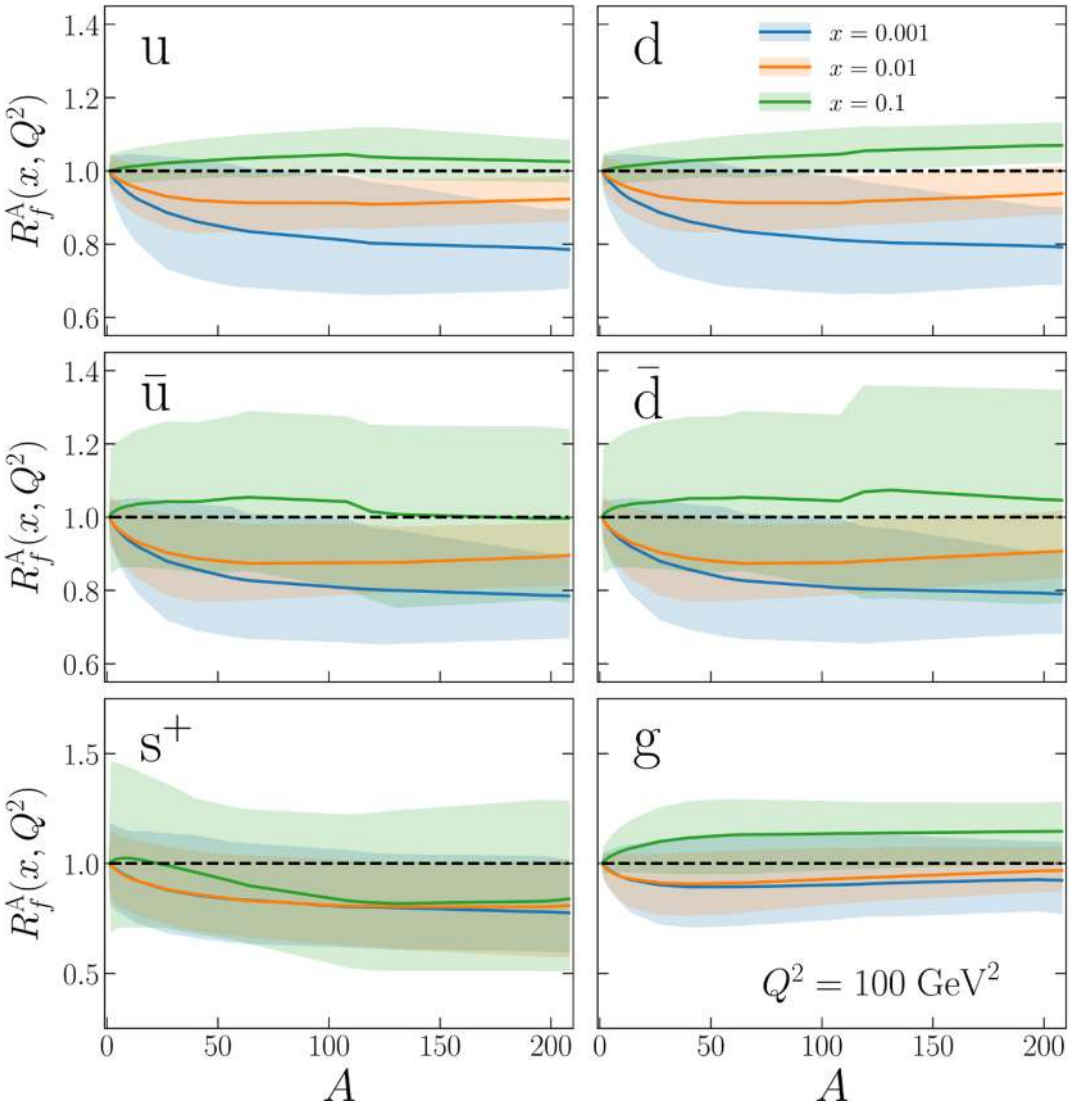

Figure 12. Comparison of the nuclear PDF ratios, eq. (4.4) as a function of atomic mass A for three different x values, $x = 0.001$, $x = 0.01$, and $x = 0.1$, at $Q = 10\,\text{GeV}$. From top to bottom we show the up and down quarks, the corresponding antiquarks, the total strangeness, and the gluon. The bands indicate the 90% confidence level intervals and take into account the correlations with the proton baseline used for the normalization.

can also see a notable improvement in PDF uncertainties as A approaches 208 (lead) at low x as result of the constraints from LHC measurements. Similar behavior is seen also for the anti-quarks, although with larger uncertainties. On the other hand, the strange quark modifications are stable across both x and A, with significantly larger errors at $x = 0.1$. Lastly, the gluon distribution also displays smooth dependence on A for the modification ratios. Here the notable shift from anti-shadowing to shadowing is also seen, although the ratios all agree with unity within the 90% CL range.

JHEP09(2020)183

Comparison with EPPS16 and nCTEQ15. In figures 13 and 14, we display the nuclear PDF modification ratios at $Q = 10\,\text{GeV}$ for iron, $R_f^{\text{Fe}}(x, Q^2)$, and lead, $R_f^{\text{Pb}}(x, Q^2)$, for nNNPDF2.0, EPPS16, and nCTEQ15, each normalized to the corresponding free-proton baseline. As in figure 11, we show the up and down quarks, the corresponding antiquarks, the total strangeness, and the gluon. The bands again correspond to the 90% CL uncertainties constructed using the appropriate prescription for each set. This means that for the other analyses, the error is computed by adding in quadrature the differences in value between the N_{eg} eigenvectors of the Hessian set and the best fit result.[4]

Beginning with the nPDF comparison for iron nuclei, we find that there is good agreement between the results of nNNPDF2.0 and EPPS16 both in terms of central values and of the nPDF uncertainties in the range of x for which experimental data is available. Our result also agrees with the nCTEQ parameterization within uncertainties, although the errors for nCTEQ are much smaller in the low-x region, displaying more significant shadowing effects. Such differences in uncertainties between nCTEQ and the other analyses may be attributed to the choice of tolerance criteria in Hessian error propagation and/or the absence of proton PDF uncertainties. Note also that the nCTEQ15 analysis does not include any charged-current DIS or LHC data. The quark flavor separation is instead obtained from Drell-Yan di-lepton data at the Tevatron and inclusive pion production data at RHIC.

Our result displays more visible differences with the other analyses in the small- and large-x extrapolation regions, where our uncertainties are larger as a result of a more flexible parameterization. The Fermi-motion-like growth of R_u^{Fe} and R_d^{Fe} at very large x, which is built into the EPPS16 parameterization, is absent in the nNNPDF2.0 and nCTEQ15 results. There instead one finds a suppression compared to the free-proton case, especially in the case of R_d^{Fe}. However, our result remains fully consistent with Fermi-motion behavior within the uncertainties. As expected, the observed pattern of nuclear modifications is very similar between up and down quarks and between the corresponding antiquarks due to iron being nearly isoscalar.

Considering the behavior of the sea quarks, nNNPDF2.0 and EPPS16 agree well in terms of central values and uncertainties in the shadowing region, $x \lesssim 0.05$. Here the nCTEQ result again shows more prominent shadowing effects with smaller uncertainties. There are more significant differences at large-x, where the qualitative behavior between the nPDF sets is the opposite: nNNPDF2.0 favors an enhancement compared to the free-proton baseline, while EPPS16 and nCTEQ15 prefer instead a suppression. In any case, the differences are well within the large uncertainty bands, and additional data is needed to be able to ascertain the correct behavior in this region. Note that at large-x the free-proton baseline antiquarks are also affected by large errors, complicating the interpretation of $R_{\bar{u}}^{\text{Fe}}$ and $R_{\bar{d}}^{\text{Fe}}$.

Turning to the nuclear modification of the total strangeness, in nNNPDF2.0 we find a suppression of $\sim 20\%$ compared to the proton baseline in the relevant range of x. This is consistent with studies of the interplay between the NuTeV dimuon and the ATLAS W,Z

[4]When computing PDF ratios with EPPS16 and nCTEQ15 we neglect proton PDF uncertainties, since adding these errors in quadrature is likely to represent an overestimate given the missing mutual correlations.

JHEP09(2020)183

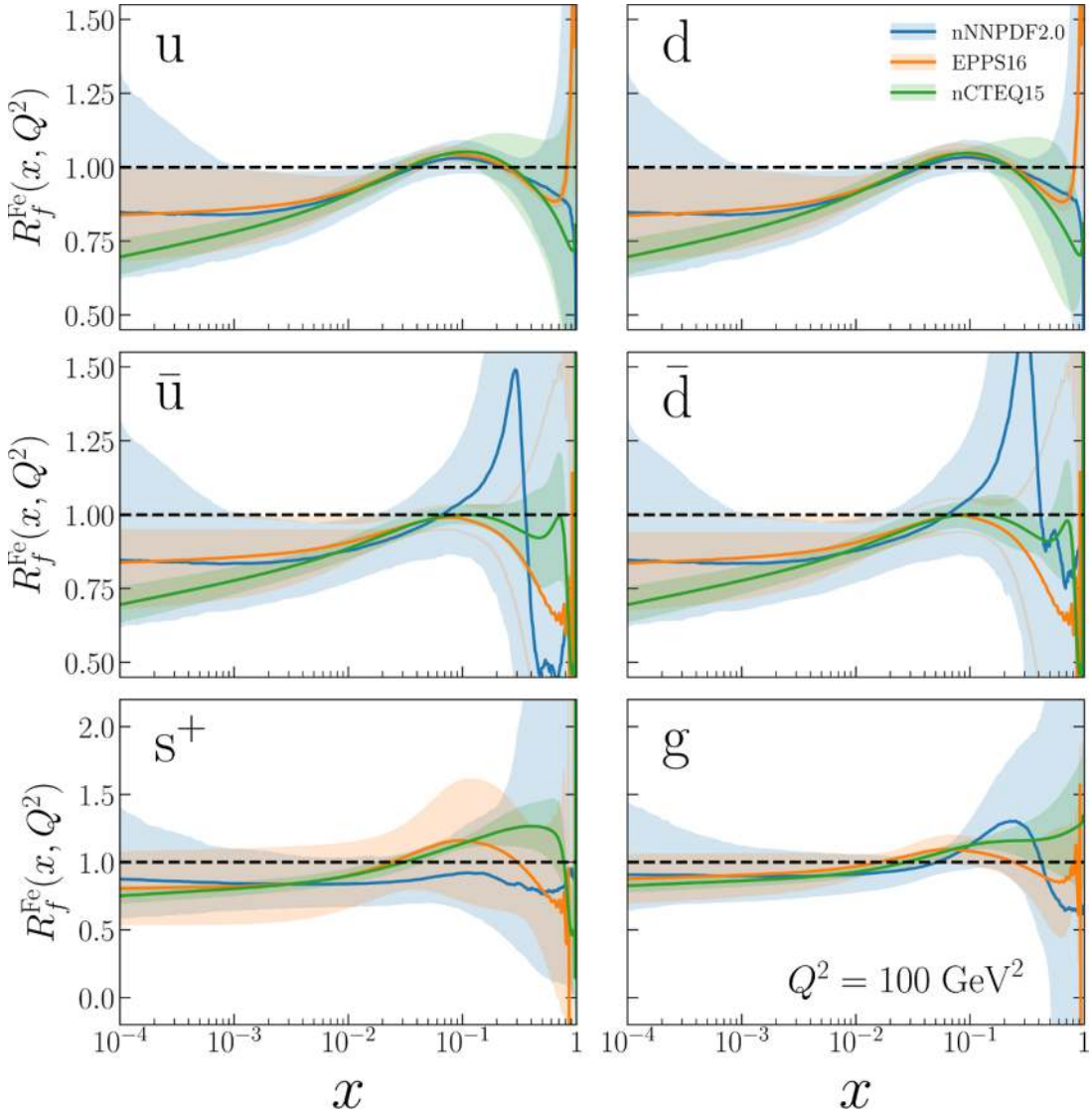

Figure 13. The nuclear PDF modification ratios for iron, $R_{\rm Fe}(x, Q^2)$, as a function of x at $Q = 10\,$GeV for nNNPDF2.0, EPPS16, and nCTEQ15. From top to bottom we show the up and down quarks, the corresponding antiquarks, the total strangeness, and the gluon. The bands correspond to the 90% CL uncertainties, and each nPDF set is normalised to its corresponding free-proton baseline as indicated by eq. (4.4).

2011 data in proton global analyses, where the latter data set largely suppress strangeness in contrast to the former. Such behavior is not reported by EPPS16, which exhibits much larger nPDF uncertainties that are likely due to the absence of the strange-sensitive NuTeV cross-sections in their analysis. Furthermore, the ATLAS W,Z 2011 distributions are missing from the CT14 proton baseline used by EPPS16 (although these data have been accounted for in the recent CT18 release [7]).

Interestingly, the nCTEQ15 result also shows strange quark suppression at low-x with significantly smaller uncertainties, despite not including any strange-sensitive observables.

JHEP09(2020)183

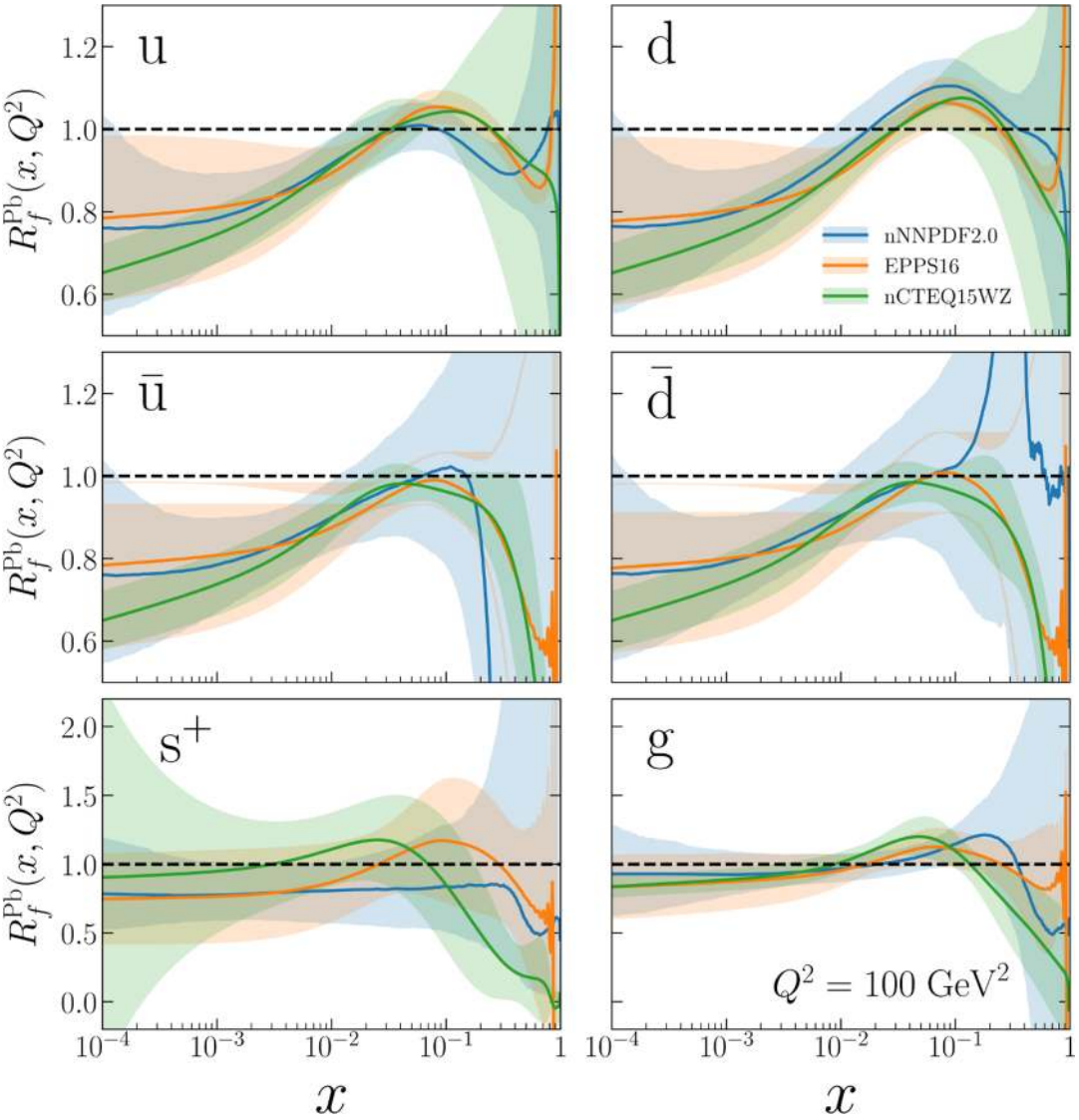

Figure 14. Same as figure 13 in the case of lead nuclei, $R_{\rm Pb}(x, Q^2)$.

Furthermore, their analysis exhibits a clear anti-shadowing behavior in the intermediate x region, in better agreement with EPPS16 central values. In any case, all distributions remain in agreement within uncertainties and additional strange-sensitive observables are needed to clarify the nuclear modifications of the strange.

Finally, concerning the gluon PDF we find from this comparison that in the nNNPDF2.0 fit there is little evidence for nuclear shadowing, with $R_g^{\rm Fe} \simeq 1$ in the region $x \leq 0.05$. We also find that the nPDF uncertainties on the gluon are larger compared to EPPS16 by roughly a factor of two. At larger values of x, the uncertainties increase significantly and nNNPDF2.0 prefers a suppressed central value, unlike the nCTEQ15 result. We note that none of the analyses include direct constraints on the large-x nuclear gluons, hence the sizeable nPDF uncertainties, although available data on dijet and photon production could improve this situation.

JHEP09(2020)183

In the corresponding comparison for lead nuclei, displayed in figure 14, one observes a number of similarities and differences with respect to the nPDFs of iron. Concerning the up and down quarks, we find our nNNPDF2.0 result provides significant evidence for shadowing at small-x. For instance, at $x \simeq 5 \times 10^{-3}$ we obtain $R_u^{\rm Pb} \neq 1$ at the four-sigma level or higher. Interestingly, nPDF uncertainties in the shadowing region are up to a factor of two smaller in nNNPDF2.0 than in EPPS16. Here we illustrate the ratio comparison with the updated nCTEQ15WZ analysis [84], which included constraints from W and Z boson production from proton-lead collisions at the LHC. In this case, the uncertainties are more comparable to the EPPS16 result, but still smaller than nNNPDF2.0 at low x.

While in all cases anti-shadowing at $x \simeq 0.1$ is observed, the larger x qualitative behavior is different between the analyses, with EPPS16 finding the (built-in) EMC suppression followed by Fermi-motion rise while in nNNPDF2.0 and nCTEQ15WZ the pattern of nuclear modifications is different. In any case, the agreement between the central values of nNNPDF2.0 and the other analyses for $R_u^{\rm Pb}$ and $R_d^{\rm Pb}$ in the region of $x \lesssim 0.3$ is quite remarkable given the very different methodologies employed in each study.

Concerning the nuclear modifications of the sea quarks in lead nuclei, one finds a similar qualitative behavior as in the case of iron. For $x \lesssim 0.1$ there is good agreement between the central values of $R_{\bar{u}}^{\rm Pb}$ and $R_{\bar{d}}^{\rm Pb}$ between EPPS16 and nNNPDF2.0, with the latter exhibiting smaller uncertainties. The nCTEQ15WZ result again displays more significant sea quark modifications at low-x, but resides in agreement with our result within the 90% error. The sets are also notably different for $x \gtrsim 0.1$, where EPPS16 and nCTEQ15WZ predict a EMC-like suppression common to $\bar{u}$ and $\bar{d}$ while nNNPDF2.0 favours a large suppression for $\bar{u}$ but an enhancement for $\bar{d}$. However, the large nPDF uncertainties in this region prevent any definitive conclusions, as the two fits are fully consistent within the 90% CL bands. One possible source for the differences could be in the respective free-proton counterparts, where large-x antiquarks are poorly known. For the total strangeness, the behavior of $R_{s^+}^{\rm Pb}$ is similar to that of iron, where nNNPDF2.0 predicts a suppression more or less independent of x, with rather larger uncertainties for EPPS16 compared to our nNNPDF2.0 result due to the missing constraints from the NuTeV dimuon cross-sections. On the other hand, the nCTEQ15WZ distribution displays much different behavior than the ratio for iron nuclei constructed with the preceding nCTEQ results, since the former analysis contains some strange sensitivity from the W/Z LHC data. In this case, the strange uncertainties are comparable in size to our result but shows much less suppression. Moreover, the nCTEQ15WZ shows a strong EMC-like effect in the region of $x \sim 0.2 - 0.3$.

Finally, regarding the nuclear modifications of the lead gluon PDF illustrated in the bottom right panel of figure 14, we again find that $R_g^{\rm Pb}$ agrees with unity across all relevant x. Here the initial-scale differences are washed out partially by DGLAP evolution, but clearly the shadowing in the nuclear gluons is less apparent than for the quarks, aside from the nCTEQ15 result which clearly shows shadowing behavior. Although the nPDF uncertainties increase at large x due to the lack of direct constraints, the qualitative behavior of $R_g^{\rm Pb}$ differs between the various PDF determinations.

JHEP09(2020)183

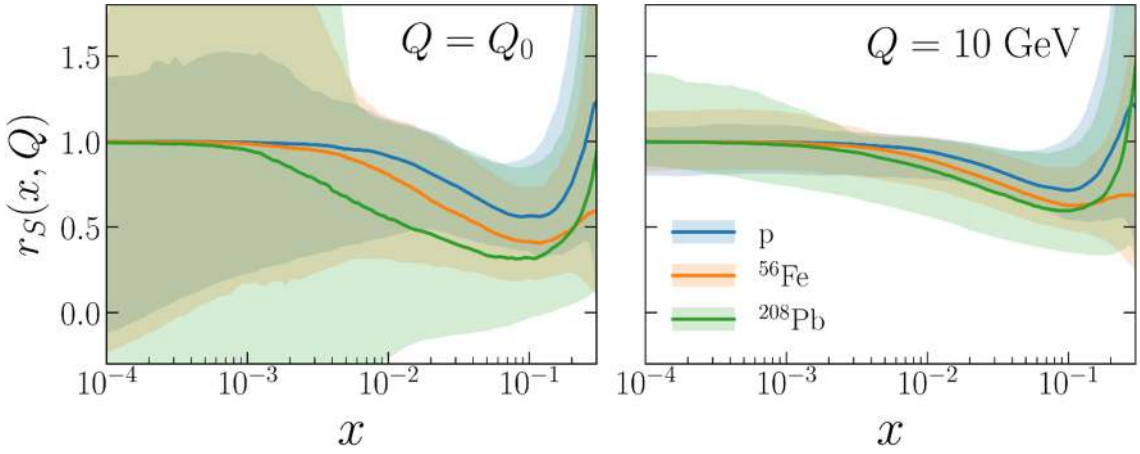

Figure 15. The strangeness ratio $r_s(x, Q^2)$, defined in eq. (4.5), in nNNPDF2.0 comparing the results for the free-proton baseline, ^{56}Fe, and ^{208}Pb at both the input parametrisation scale $Q_0 = 1\,\text{GeV}$ (left) and at a higher scale $Q = 10\,\text{GeV}$ (right plot).

Nuclear strangeness. The strangeness content of the proton in unpolarized PDF fits has attracted a lot of attention recently. Traditionally, the determination of $s(x, Q^2)$ in global proton PDF fits has been dominated by the constraints provided by charm production in neutrino DIS [85–89]. These measurements suggest that the strange sea is suppressed compared to its up and down quark counterparts, favoring values of around $r_s \simeq 0.5$ when expressed in terms of the strangeness ratio defined as

$$r_s(x, Q^2) \equiv \frac{s(x, Q^2) + \bar{s}(x, Q^2)}{\bar{u}(x, Q^2) + \bar{d}(x, Q^2)}\,. \tag{4.5}$$

Other strange-sensitive processes agree qualitatively with the constraints on r_s provided by the neutrino DIS data, such as W production in association with charm quarks [90] from CMS [91–93] and ATLAS 7 TeV [94], and semi-inclusive deep-inelastic scattering (SIDIS) [95–97]. However, the ATLAS measurements of the leptonic rapidity distributions in inclusive W and Z production at 7 TeV [48, 98] exhibit instead a strong preference for a symmetric strange sea with $r_s \simeq 1$. One should point out that general considerations based on perturbative DGLAP evolution imply that $r_s \to 1$ at large Q and small-x, but at low Q and medium/large-x the value of r_s is dictated by non-perturbative dynamics.

As was motivated in ref. [10], it is important to carefully assess the nuclear uncertainties associated to the nuclear strangeness, given that these will potentially affect the determination of the proton strangeness from global fits based on neutrino data. We display in figure 15 the strangeness ratio $r_s(x, Q^2)$ defined by eq. (4.5) obtained with our nNNPDF2.0 result for ^{1}p, ^{56}Fe, and ^{208}Pb at both the input parameterization scale $Q_0 = 1\,\text{GeV}$ and at a higher scale of $Q = 10$ GeV.

From the comparison in figure 15 we find that at the input parameterization scale r_s is particularly suppressed in the case of lead, where the central value of nNNPDF2.0 satisfies $r_s < 0.5$ for $5 \cdot 10^{-3} \lesssim x \lesssim 0.2$. A similar preference for a suppressed strange sea, albeit less pronounced, can be seen in iron nuclei. In any case, the nPDF uncertainties affecting this

JHEP09(2020)183

ratio are rather large, in particular for the heavier nuclei. The fact that for $x \lesssim 10^{-3}$ one obtains $r_s \simeq 1$ for all three nuclei is a consequence of the parameterization preprocessing, whose ranges are chosen to ensure that in the small-x extrapolation region all quark and antiquark flavors behave in the same way (see figures 4 and 3). Once DGLAP evolution takes place, r_s tends to become closer to unity across a wider range in x, but even at the higher scale a suppressed strangeness for $x \gtrsim 0.01$ is preferred for both iron and lead.

The results in figure 15 suggest that including neutrino CC structure functions such as CHORUS and NuTeV in proton PDF fits without accounting for nuclear uncertainties might not be a justified approximation, given the current precision that modern fits achieve. It will be interesting nonetheless to determine the impact on the global NNPDF proton PDF fits when nNNPDF2.0 is used to account for nuclear uncertainties using the procedure outlined in ref. [10].

4.2 Comparison with nNNPDF1.0

We now turn to study the differences between the nNNPDF1.0 and nNNPDF2.0 determinations by tracing back the impact of the various improvements in the latter with a series of comparisons. The goal of this exercise is to assess which of these differences can be identified with specific methodological improvements, such as the cross-section positivity constraint, and which ones are related to the impact of the new experimental information, either the DIS charged current structure functions or the LHC gauge boson production measurements.

The starting point for this study will be a fit denoted nNNPDF1.0r, which has been obtained with the code used to produce nNNPDF2.0 but using the same theory, methodology settings, and input dataset as in the nNNPDF1.0 analysis. The only differences at this level are related to optimizations and improvements implemented in the code to speed up its performance. We have verified that nNNPDF1.0 and nNNPDF1.0r are statistically indistinguishable, thus we can safely adopt the latter as baseline for the comparisons in what follows.

We have then produced several variants of this nNNPDF1.0r baseline, each time adding one extra feature or dataset. The first of these two variants is a fit where the proton boundary condition has been updated to the no-nuclear NNPDF3.1 fit shown in figure 5. The second is a fit where, in addition to the updated boundary condition, the positivity of cross-sections has been imposed following the procedure described in section 3.3. We display in figure 16 the comparison between nNNPDF1.0r and these two fit variants. Since the isoscalar neutral-current DIS structure functions used in nNNPDF1.0 are primarily sensitive to the specific quark combination $\Sigma + T_8/4$, we plot this together with the gluon distribution as a function of x at $Q^2 = 10\,\mathrm{GeV}^2$ for carbon, iron, and lead.

First, one can see from figure 16 that the impact of the new proton boundary condition in the nuclear fit is generally moderate concerning the size of the uncertainty band. There are some differences at the central value level for the small-x quarks and for the nuclear gluon PDF of lead, but in both cases the shifts are much smaller than the associated uncertainties. This does not imply that using the updated proton boundary condition is irrelevant for nNNPDF2.0, but rather that this choice is not particularly impactful for the

JHEP09(2020)183

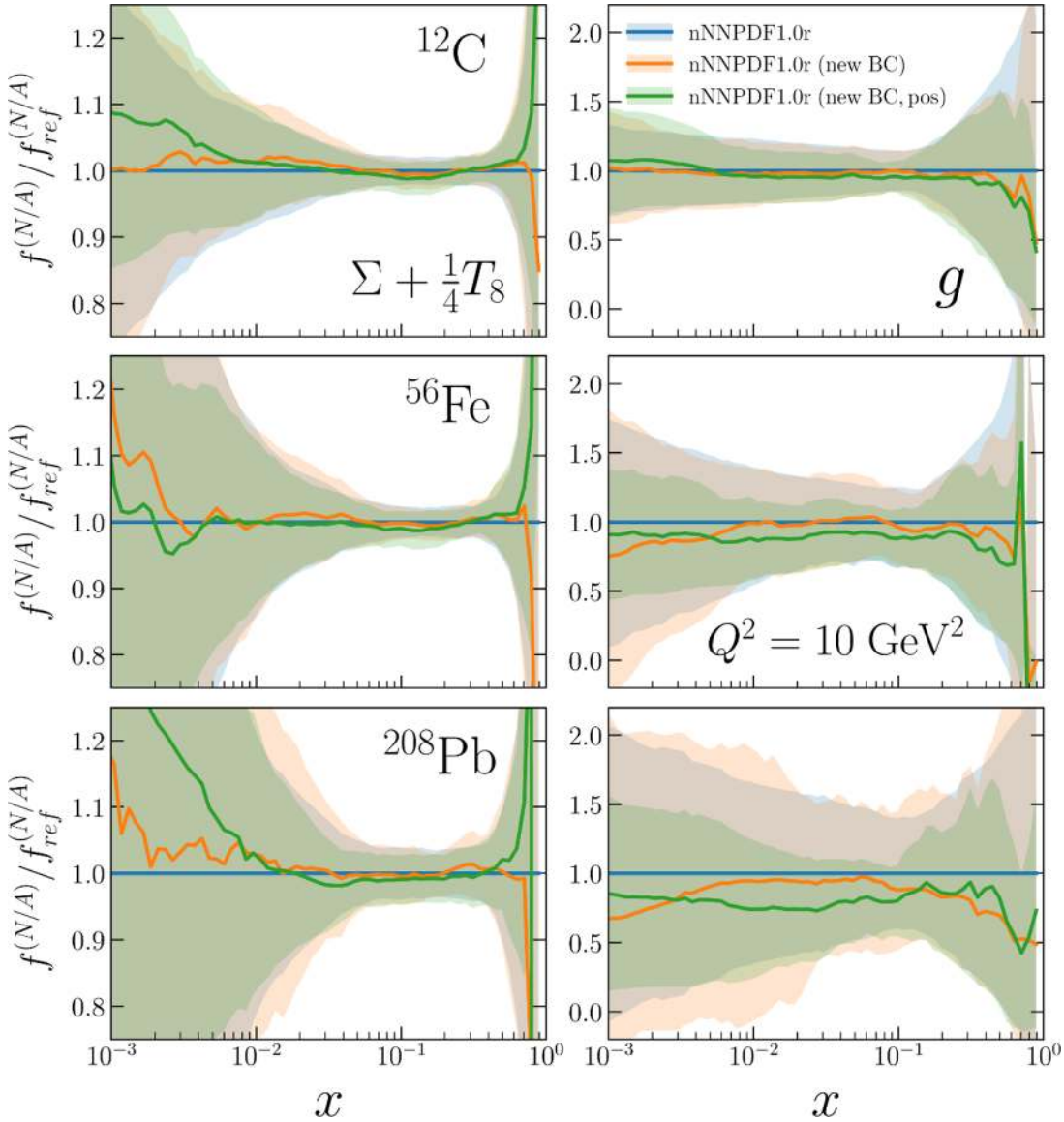

Figure 16. Comparison between the nNNPDF1.0r baseline and two fit variants based on the same input dataset, one where the proton boundary condition has been updated and another where in addition the positivity of physical cross-sections has been imposed. We show the $\Sigma + T_8/4$ quark combination (left) and gluon (right) at $Q^2 = 10\,\mathrm{GeV}^2$ for three values of A.

specific PDF combinations that can be constrained by the nNNPDF1.0 dataset. As shown in figure 5, the differences between the two variants of the proton boundary conditions are more distinguished for the total strangeness compared to the other quark flavors.

On the other hand, imposing the positivity of the cross-sections leads to more important differences. This is not completely unexpected, since it is well known that in general a model-independent (n)PDF analysis will lead to some cross-sections being negative unless their positivity is explicitly imposed. In our case, one finds that there is not much difference in the quarks, but there are clear changes for the nuclear gluons in iron and lead,

JHEP09(2020)183

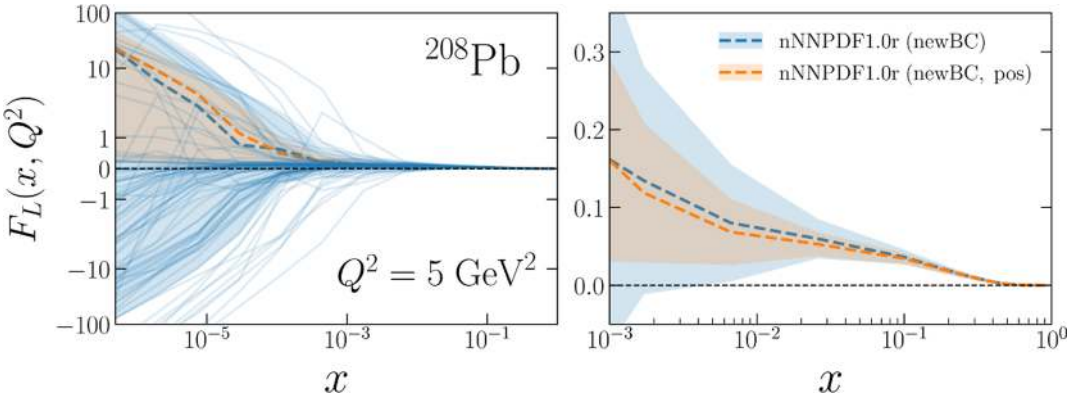

Figure 17. The longitudinal structure function $F_L(x, Q^2)$ at the positivity scale $Q^2 = 5\,\text{GeV}$. We compare the predictions of the nNNPDF1.0r(newBC) fits with and without the cross-section positivity constraint imposed. We show the extrapolation (left) and the data (right) regions, in the former case displaying also the predictions from the individual replicas in the nNNPDF1.0r(newBC) fit that do not satisfy the positivity constraints.

especially in the latter. Here we see that imposing the positivity of cross-sections leads to a significant reduction of the nPDF uncertainty band, which in the case of lead can be up to a factor of two.

To illustrate the impact of the cross-section positivity constraint, we display in figure 17 the longitudinal structure function $F_L(x, Q^2)$ at the positivity scale $Q^2 = 5\,\text{GeV}$. We compare the predictions of the nNNPDF1.0r fits including the updated proton boundary condition with and without the cross-section positivity constraint imposed in both the extrapolation and the data regions. In the left panel, we display also the predictions from the individual replicas of the nNNPDF1.0r(newBC) fit that do not satisfy the cross-section positivity constraints. Indeed, one can observe that many F_L replicas become negative in some region of x unless this constraint is explicitly imposed, and that removing them leads to a significant reduction of the nPDF uncertainties, particularly in the small-x region. Interestingly, at medium-x it is largely the upper (rather than the lower) 90% CL limit which is reduced by the positivity constraint: this can be explained by the fact that the very negative F_L replicas at small-x were actually higher than the median value at medium-x in order to satisfy the momentum sum rule.

Concerning the impact of the new datasets, a direct comparison of the nNNPDF1.0r-like fits with those including CC DIS and LHC data is not possible since as discussed in section 3.2 the input parameterization basis and the flavor assumptions are different. However, we are still able to assess the relative contribution of the CC structure functions and the LHC gauge boson cross-sections in determining the nNNPDF2.0 results. In figure 18 we display the nuclear modification ratios for the nPDFs in lead, as was shown in figure 14, but now comparing the nNNPDF2.0 baseline results with those of a fit that is restricted to DIS structure functions, including charged-current scattering, and that uses identical theoretical and methodological settings.

One of the most remarkable features of this comparison is the sizeable impact that LHC measurements have in reducing the uncertainties of the nuclear PDFs. This effect is

JHEP09(2020)183

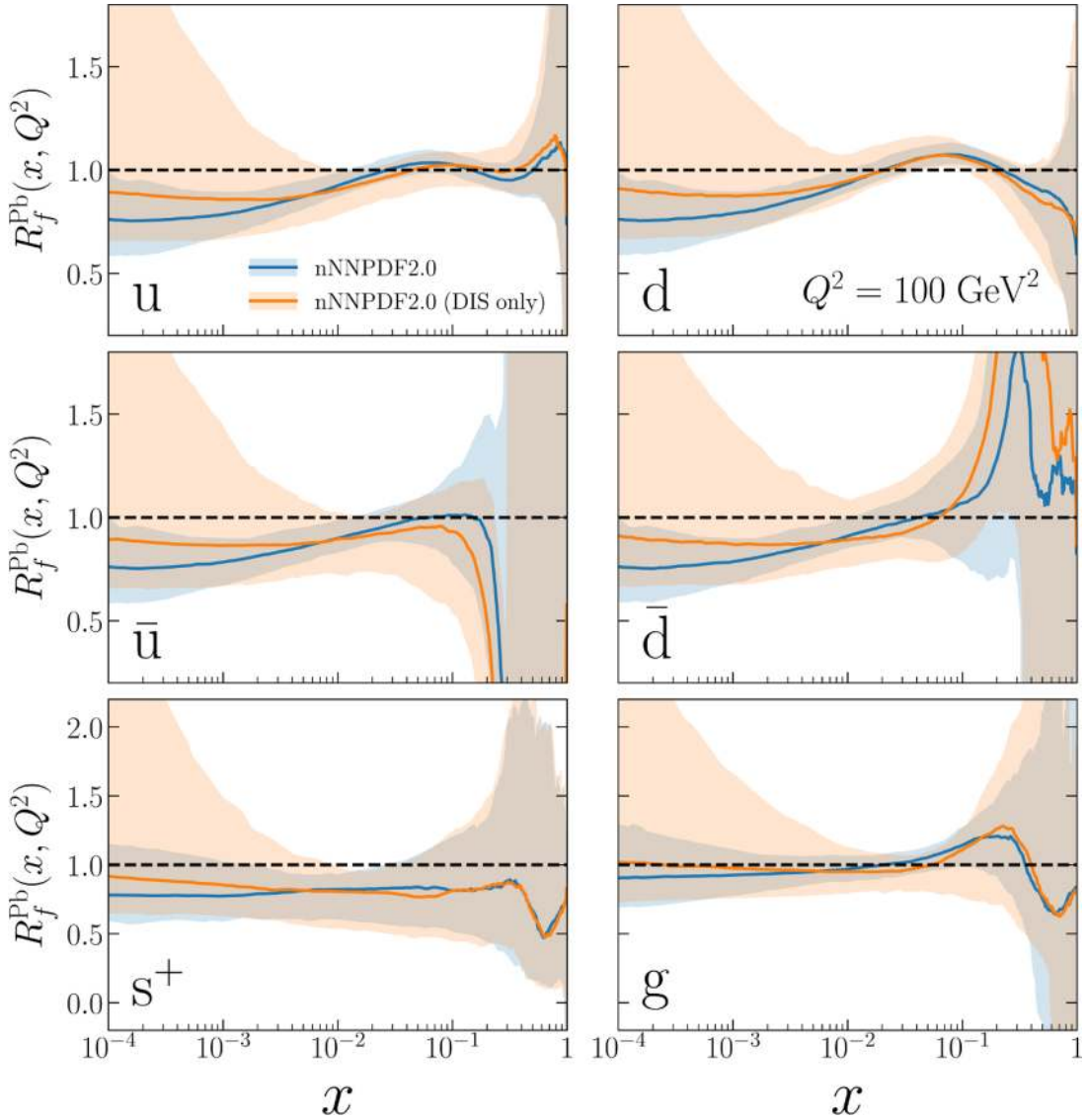

Figure 18. Same as figure 14 now comparing the nNNPDF2.0 baseline results with those of a fit based on identical settings but restricted to a DIS-only dataset.

particularly significant for the gluon and for all quark flavors at $x \lesssim 0.1$. On one hand, the LHC data clearly reveals the presence of nuclear shadowing at small-x for both the valence and sea quarks, something which is not accessible in a DIS-only fit. This result is consistent with the nuclear modification ratios for the LHC Drell-Yan distributions reported e.g. in figure 8. On the other hand, the impact of the LHC data on the central values and uncertainties of nNNPDF2.0 at $x \gtrsim 0.1$ is less prominent, although in that region one also observes a reduction of the uncertainties. The fact that $R_{\bar{u}} \ll 1$ and $R_{\bar{d}} \gg 1$ for lead nuclei at large-x is already present at the level of DIS-only fits implies that this trend is favored by the CHORUS and NuTeV charged-current structure functions.

JHEP09(2020)183

4.3 The momentum and valence integrals in nuclei

As was discussed in section 3.2, we impose three sum rules in the nNNPDF2.0 determination, namely the momentum sum rule, eq. (3.7), and the two valence sum rules, eqs. (3.12) and (3.14). These constraints are satisfied by adjusting the overall prefactors B_f in eq. (3.6) for the gluon g, the total valence V, and the valence triplet V_3 distributions, respectively. Furthermore, they are independently imposed for each value of A for which there is available experimental data.

Here we investigate the role played by these sum rules in the global nPDF determination. In particular, we address whether or not the physical requirements of energy and valence quark number conservation are satisfied by the phenomenological fit to experimental data (within uncertainties) when the sum rules are not explicitly imposed. Recently, theoretical arguments have been put forward that the momentum sum rule for nucleons in nuclei might not hold [99]. Motivated in this respect, we have carried out a similar study to the one presented in ref. [46], where global proton PDF fits without imposing the momentum sum rule were performed. In the proton case, while the LO prediction for the momentum integral was expected to be far from the QCD expectation, both the NLO and NNLO fits exhibited remarkable agreement at the $\simeq 1\%$ level [46].

We have therefore produced two variants of the nNNPDF2.0 analysis, each based on $N_{\rm rep} = 250$ replicas, where either the momentum sum rule or the total valence sum rule is not imposed. Afterwards, we evaluate in each case the corresponding momentum and total valence integrals, defined as

$$I_{\rm M}(A) \equiv \int_0^1 dx\, x \left(\Sigma^{(p/A)}(x,Q_0) + g^{(p/A)}(x,Q_0)\right) , \tag{4.6}$$

$$I_{\rm V}(A) \equiv \int_0^1 dx\; V^{(p/A)}(x,Q_0) , \tag{4.7}$$

and assess whether or not they are in agreement with the QCD expectations, namely $I_{\rm M}(A) = 1$ and $I_V(A) = 3$ respectively. One should note that the momentum and valence sum rules are already satisfied at the level of the proton boundary condition, and thus some constraints are expected to be propagated to the lighter nuclei. However, the analysis of ref. [46] demonstrates that results would be largely unchanged if the momentum and valence sum rules would have been excluded also from the free-proton baseline.

In figure 19 we display the distribution of the momentum and valence integrals, eqs. (4.6) and (4.7), respectively, in the variants of the nNNPDF2.0 fit where the corresponding sum rules are not being explicitly imposed. We show the relative frequency of the momentum and valence integral for three representative nuclei: ^{12}C, ^{56}Fe, and ^{208}Pb. The dashed vertical line in figure 19 indicates the QCD expectations for $I_{\rm M}(A)$ and $I_{\rm V}(A)$. The corresponding values for the 90% confidence level intervals for each of these two integrals for the relevant values of A, as well as for the free-proton baseline $A = 1$, can be found in table 6.

From the results presented in figure 19 and table 6 one finds that the momentum integral is in agreement with the QCD expectation, $I_{\rm M}(A) = 1$, within uncertainties for all nuclei. In the case of ^{12}C for example, one finds that $0.97 \lesssim I_{\rm M} \lesssim 1.10$ at the 90%

JHEP09(2020)183

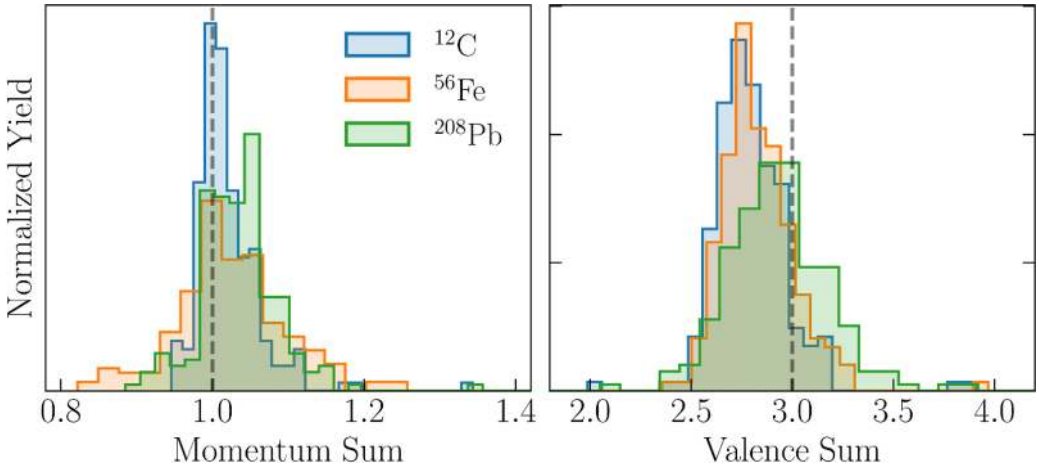

Figure 19. The distribution of the momentum (left) and total valence (right panel) integrals, eqs. (4.6) and (4.7) respectively, in the variants of the nNNPDF2.0 determination where the corresponding sum rules have not been explicitly imposed. We show the relative frequency of the momentum and valence integral for three representative nuclei: ^{12}C, ^{56}Fe, and ^{208}Pb. The dashed vertical line indicates the corresponding the QCD expectations, $I_{\rm M}(A) = 1$ and $I_{\rm V}(A) = 3$. The associated 90% CL ranges are reported in table 6.

A	$I_{\rm M}(A)$	$I_{\rm V}(A)$
1	$[0.99, 1.06]$	$[2.53, 3.12]$
12	$[0.97, 1.10]$	$[2.56, 3.11]$
56	$[0.90, 1.16]$	$[2.58, 3.16]$
208	$[0.94, 1.12]$	$[2.54, 3.34]$

Table 6. The 90% CL ranges for the momentum and valence integrals, eqs. (4.6) and (4.7), in the variants of the nNNPDF2.0 fits whether either one or the other sum rule is not imposed.

confidence level, with somewhat larger uncertainties for the heavier nuclei. Even for lead, where the proton boundary condition has little effect, the median of the distribution is reasonably close to the QCD expectation. The uncertainties on $I_{\rm M}$ are larger in the nuclear PDF analysis than the $\simeq 1\%$ error found in the proton case [46], as expected since the experimental data for nuclear collisions is far less abundant and further distributed between different nuclei. Nevertheless, the overall consistency with the QCD expectations is quite compelling. Note also that here the proton boundary condition is imposed only for $x \geq 10^{-3}$, and therefore our prediction for $I_{\rm M}(A = 1)$ is expected to be less accurate compared to the proton global analysis case.

The result that the momentum integral agrees with the theoretical predictions for all nuclei is a non-trivial validation of the global nuclear PDF analysis framework based on the QCD factorization hypothesis. It further demonstrates the robustness of our fitting methodology, in that the resulting nPDFs are reasonably stable regardless of whether or not the momentum sum rule is imposed during the fit. To illustrate better this latter point, in figure 20 we provide a comparison between the baseline nNNPDF2.0 fit at $Q_0 = 1\,$GeV

JHEP09(2020)183

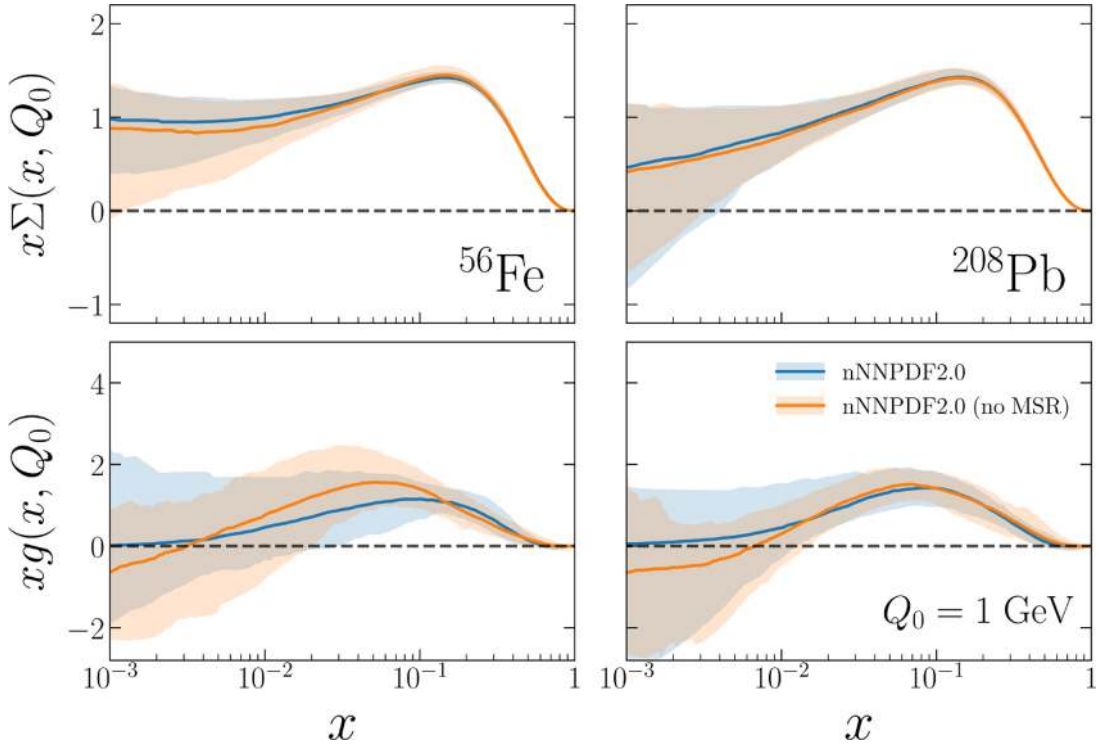

Figure 20. Comparison of the baseline nNNPDF2.0 fit at $Q_0 = 1\,\text{GeV}$ with the variant in which the momentum sum rule is not being imposed. We show the total quark singlet (left) and the gluon (right) for ^{56}Fe (upper) and ^{208}Pb (lower panels).

with the variant in which the momentum sum rule is not being imposed. We show the total quark singlet and the gluon for both ^{56}Fe and ^{208}Pb. Recall that the momentum sum rule is used to fix the overall gluon normalization in eq. (3.6). In the case of lead, where the experimental constraints are relatively abundant, we find that both the singlet and the gluon are reasonably similar irrespective of whether or not the momentum sum rule is imposed. The momentum sum rule plays a larger role in iron, especially in reducing the gluon nPDF uncertainties, but interestingly the central value of the all distributions is quite stable when comparing the two fits. This stability is consistent with the results reported in figure 19.

The main conclusions are qualitatively similar for a fit in which the total valence sum rule has not been imposed. Results of this fit are displayed in the right panel of figure 19, where the normalized frequency of the total valence integral are shown for the same three nuclei discussed previously. The corresponding 90% CL intervals are also reported in table 6. Similar to the momentum sum results, we find that for the valence integral the fit results agree with the QCD expectations within uncertainties. The preferred value of the valence integral (median) turns out to be $I_{\rm V} \simeq 2.8$ irrespective of A. This implies that even when eq. (4.7) is not imposed explicitly, the experimental measurements favor the QCD prediction within 5% for all values of A relevant for the present study. The value for the χ^2 per degree of freedom is somewhat smaller (0.953) than in the baseline (0.976). We

JHEP09(2020)183

have also verified that, in a similar way as in figure 20, the resulting nPDFs are reasonably stable regardless of whether or not the valence sum rule is imposed.

Putting together the results of these two exercises, one can conclude that the fit results are relatively stable in the nNNPDF framework even in the absence of the sum rules, consistent with the fact that experimental data and the QCD expectations based on the factorization theorem are in agreement with each other for hard-scattering collisions involving heavy nuclei.

4.4 The positivity of physical cross-sections

As was discussed in section 3.3, we impose the requirement that the cross-sections of arbitrary physical processes are positive-definite quantities. This constraint is implemented by means of an additive penalty term in the figure of merit, eq. (3.18). Moreover, the penalty is constructed from the pseudo-data summarized in table 3, which corresponds to lepton-nuclear scattering structure functions and Drell-Yan cross-sections in proton-nucleus collisions. Recall that the kinematics of the positivity pseudo-data were chosen to cover those of the actual data used in the fit, see figure 1.

Here we want to demonstrate that the nNNPDF2.0 determination indeed satisfies these various positivity constraints. In figure 21 we display a representative selection of the positivity observables imposed in nNNPDF2.0. In particular, we show the DIS structure functions $F_2^s(x, Q^2)$ and $F_L(x, Q^2)$, as well as the Drell-Yan rapidity distributions $\sigma_{u\bar{u}}^{\rm DY}(y)$ and $\sigma_{\bar{u}d}^{\rm DY}(y)$, where the bands indicates the 90% confidence level uncertainty interval. We use a scale of $Q^2 = 5\,{\rm GeV}^2$, which corresponds to the same scale in which eq. (3.18) is imposed. Furthermore, we provide the positivity predictions for both iron and lead nuclei. Note that since the Drell-Yan cross-sections are not normalized by the value of A, the absolute magnitude of the two nuclei are different. Of course, the overall normalization is not relevant for the implementation of the positivity constraint.

The selection of positivity observables in figure 21 is representative since it contains one of the quark structure functions (F_2^s) constraining a q^+ combination, F_L that is sensitive to the gluon positivity, and a diagonal and off-diagonal DY cross-section which are relevant for different aspects of quark flavor separation (confer also the LO expressions in appendix A). Here the Drell-Yan cross-sections are represented as a function of x_2, which corresponds to the momentum fraction of the nuclear projectile obtained using the LO kinematics of eq. (A.15). While the positivity constraint was only implemented for $x_2 \gtrsim 10^{-2}$ with a per-nucleon center-of-mass energy of $\sqrt{s} = 23.5\,{\rm GeV}$, we illustrate instead the positivity for a choice of kinematics that allow a reach to $x_2 \sim 10^{-3}$, with a per-nucleon center-of-mass energy of $\sqrt{s} = 74.5\,{\rm GeV}$. As can be seen from figure 21, the nNNPDF2.0 determination satisfies the positivity of physical cross-sections in the entire kinematic range. Here the nPDF uncertainty bands become larger near the kinematic endpoints ($x = 1$ for DIS and $x_2 \simeq 10^{-3}$ for Drell-Yan), since these correspond to regions of the phase space where experimental constraints are scarce. Recall that by virtue of DGLAP evolution properties, these results ensure the cross-sections involving higher momentum transfers, $Q^2 > 5\,{\rm GeV}^2$, will also be positive provided one maintains the initial coverage in x_2. Therefore, we conclude that while we have not explicitly imposed the positivity at the level of the nuclear

JHEP09(2020)183

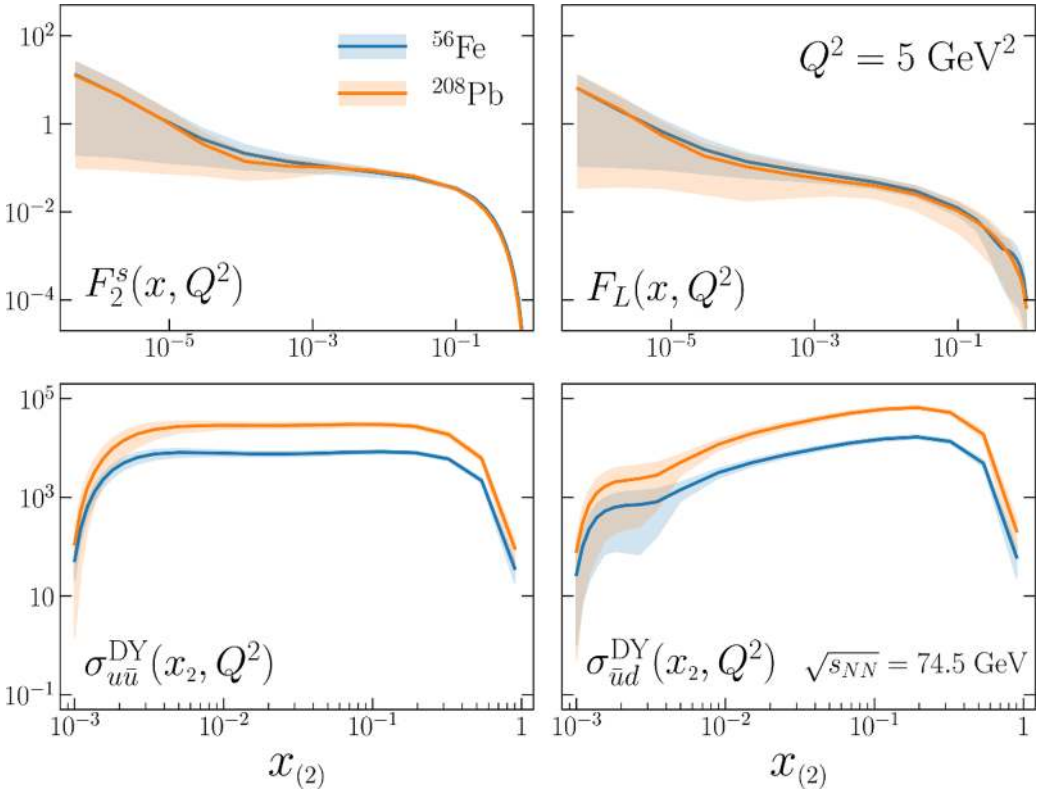

Figure 21. A representative selection of the positivity observables used in nNNPDF2.0. From top to bottom and from left to right, we show the DIS structure functions F_2^s and F_L and the Drell-Yan rapidity distributions $\sigma_{u\bar{u}}^{\rm DY}$ and $\sigma_{\bar{u}d}^{\rm DY}$. The bands indicate the 90% confidence level interval.

PDFs, physical observables constructed from nNNPDF2.0 are guaranteed to satisfy the positivity requirement.

5 Implications for photon and hadron production in nuclear collisions

In this section we discuss some phenomenological applications of the nNNPDF2.0 determination. Theoretical predictions for isolated photon production in proton-lead collisions at the LHC are first compared with recent measurements from the ATLAS collaboration taken at $\sqrt{s} = 8.16$ TeV. We then revisit the potential of the FoCal upgrade to the ALICE detector in constraining the small-x gluon nuclear PDF using measurements of direct photon production in the forward region. Finally, we provide predictions based on nNNPDF2.0 for inclusive hadron production in proton-nuclear collisions, a process that can constrain both the quark and gluon nuclear PDFs as well as the corresponding fragmentation functions in vacuum and in medium.

Several additional applications of our nNNPDF2.0 result are expected to be of phenomenological interest. In particular, our nPDFs could be used to study the constraining power of inclusive and heavy quark structure function measurements at the recently approved Electron Ion Collider (EIC) [100] and the proposed Large Hadron electron Collider

JHEP09(2020)183

(LHeC) [101]. Initial studies based on nNNPDF1.0 and EIC neutral-current structure function pseudo-data were presented in ref. [13]. However, updated projections for EIC pseudo-data are now being finalized based on more realistic accelerator and detector settings. We will therefore defer an update to our nNNPDF1.0 study of EIC pseudo-data to an upcoming Conceptual Design Report where the more accurate EIC specifications will be presented together with impact studies from various nuclear PDF analysis groups.

5.1 Isolated photon production in pA collisions with ATLAS

Production of isolated photons in proton-proton collisions is primarily sensitive to the gluon content of the proton via the QCD Compton scattering process [102, 103]. However, several complications associated with the measurement of photon production complicate a clean interpretation in terms of hard-scattering cross-sections, such as the need for subtracting the fragmentation component and the removal of photons coming from pion decays. Although early PDF fits used photon production in fixed-target scattering to constrain the gluon data, the data were eventually discarded in favor of the cleaner and more abundant data on jet production at the TeVatron [3]. However, photon production measurements from ATLAS and CMS were later revisited using NLO QCD theory in ref. [104] and again at NNLO in ref. [105]. These studies demonstrated the consistency of collider-based isolated photon production measurements with QCD predictions and with the rest of the datasets in the global analysis. Moreover, they help to reduce the uncertainties on the gluon PDF at $x \simeq 10^{-2}$, a region that is particularly relevant for theoretical predictions of Higgs production in gluon fusion.

Photon production is also a highly relevant process in the context of heavy ion collisions. Being a QCD-neutral probe, it traverses the quark-gluon plasma without modifications and thus represents a robust baseline to study the hot and dense medium properties [106]. In order to disentangle hot from cold nuclear matter effects, photon production has been measured in proton-lead collisions at the LHC, providing a new channel to constrain the nuclear modifications of the gluon PDF. Here we focus on the recent ATLAS measurements from Run II at $\sqrt{s} = 8.16\,\text{TeV}$ based on an integrated luminosity of $\mathcal{L} = 165$ nb^{-1} [107]. This analysis provides the inclusive production rates of isolated prompt photons in three different rapidity regions as a function of E_T^γ, the photon transverse energy, in the range 20 GeV to 550 GeV.

To compute the corresponding theoretical predictions, we use NLO QCD theory with the same settings as in ref. [105] by adopting a modified version of `MCFM` v6.8 interfaced to `APPLgrid`. The settings of the calculation have been adjusted to map the experimental isolation conditions and thus bypass the need to explicitly account for the fragmentation component. We have benchmarked the results of this `MCFM`-based calculation with the theory predictions presented in ref. [107] based on the `JETPHOX` program [108], finding a reasonable agreement but also some differences at large rapidities.

In figure 22, we display the comparison between the ATLAS measurements of the photon transverse energy (E_T^γ) distributions in three rapidity bins with the corresponding NLO QCD theory calculations based on `MCFM` with nNNPDF2.0 and EPPS16 as input. For each rapidity bin, the upper panels display the absolute distributions and the lower

JHEP09(2020)183

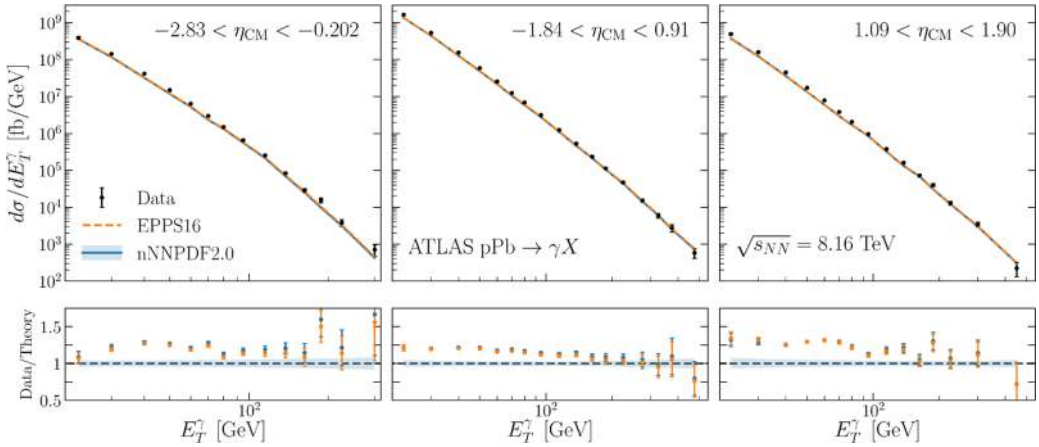

Figure 22. Comparison between the ATLAS measurements of the photon E_T^γ distributions in three rapidity bins in the center of mass frame with the corresponding NLO QCD theory calculations based on `MCFM` with nNNPDF2.0 and EPPS16 as input. For each rapidity bin, the upper panels display the absolute distributions and the lower panels the corresponding ratio between the theory calculations and the central value of the experimental data.

panels the corresponding ratio between the central value of the experimental data and the theory calculations. The error bands in the experimental measurements indicate the sum in quadrature of the statistical and systematic experimental uncertainties, while in the theory calculations the error bands correspond to the 90% CL ranges.

From this comparison one finds that the theory calculations appears to undershoot the experimental data by roughly 25% in the three rapidity bins for most of the E_T^γ range, both for nNNPDF2.0 and EPPS16. This discrepancy cannot be accommodated within the stated experimental and theoretical uncertainties. A similar qualitative behavior was reported in the original ATLAS publication based on the `JETPHOX` predictions. On the other hand, the NNPDF3.1 proton PDFs are known to describe well the corresponding isolated photon measurements from proton collisions at both $\sqrt{s} = 8$ and 13 TeV [105].

As expected, the disagreement between the experimental data and the theory calculations found in figure 22 is translated into poor χ^2 values. Using nNNPDF2.0, one obtains that $\chi^2/n_{\rm dat} = 9.1, 10.5$, and 8.5 in the forward, central, and backwards rapidity bin. Similar numbers are obtained in the case of the theory predictions based on EPPS16. The situation does not improve by much if the ATLAS photon data is added to the nNNPDF2.0 global analysis. In such a case the agreement between the theory calculations and the experimental data improves somewhat, with $\chi^2/n_{\rm dat} = 6.1, 7.5$, and 5.7 for the three rapidity bins, but it remains far from satisfactory.

Until the origin of this disagreement between theory and data is better understood, it will not be possible to include the ATLAS prompt photon production measurements in a global nPDF fit. Alternatively, one could instead consider fitting related observables that are presented in the same ATLAS publication. The first of these is the nuclear modifications ratio $R_{p\rm Pb}$, where the absolute E_T^γ distributions in pPb collisions are normalized to their pp counterparts, the latter being derived from a simulation-derived extrapolation from

JHEP09(2020)183

data in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$. The second is the ratio between different rapidity bins, such as between forward and backward rapidities, as a function of E_T^γ. The advantages of such ratios is that many experimental systematic uncertainties partially cancel out, thus facilitating the comparison with theoretical predictions. On the other hand, these observables might also exhibit a reduced nPDF sensitivity, in particular for the ratio between different rapidity bins. Future studies should shed more light on the usefulness of the prompt photon measurements to constrain nuclear PDFs within a global analysis.

5.2 Isolated photon production in pA collisions with FoCal

Current measurements of direct photon production at the LHC, such as those discussed above from the ATLAS collaboration [107] as well as related measurements from CMS and ALICE [109], are restricted to the central rapidity region. The reason is that this is the only region instrumented with electromagnetic calorimeters and thus suitable to identify photons. A measurement of isolated photon production in the forward region, however, is also highly interesting for nPDF studies. Not only would such measurements provide direct access to the poorly-known gluon nuclear modifications at small-x, but it would also allow testing for the possible onset of QCD non-linear dynamics [110].

With this motivation, a new forward calorimeter extension of the ALICE detector, dubbed FoCal [111, 112], has been proposed. Both the acceptance and instrumentation of this detector have been optimized to provide access to the nuclear PDFs at low scales and small momentum fractions via the measurement of isolated photon production at low transverse momenta and forward rapidities in proton-ion collisions. The FoCal is proposed for installation during the Long Shutdown 3 (2025-2026) phase of the LHC.

The impact of future FoCal measurements on the small-x nuclear PDFs was first studied in ref. [113]. In that analysis, pseudo-data based on the expected kinematical reach and experimental uncertainties for FoCal was generated and used to constrain the nNNPDF1.0 determination by means of the Bayesian reweighting method [114, 115]. It was found that the FoCal measurement would constrain the nuclear gluon modifications down to $x \simeq 10^{-5}$, leading to an uncertainty reduction by up to an order of magnitude as compared to the baseline fit. These results indicated a comparable or superior constraining power on the small-x nPDFs when compared to related projections from future facilities, such as the Electron Ion Collider [116].

Motivated by the new and improved projections for the FoCal pseudodata that have recently became available, we revisit their impact on nuclear PDFs using the present nPDF determination. In this case, the nNNPDF2.0 PDFs represent a more realistic baseline since they provide a robust quark flavor separation with a better handle on the gluon. Moreover, the positivity of physical cross-sections is guaranteed, a constraint that helps to reduce the small-x nuclear PDF uncertainties.

For this study we have adopted the same settings as in ref. [113] and computed NLO QCD predictions with a modified version of `INCNLO` that benefits from improved numerical stability at forward rapidities [117]. Theoretical predictions for FoCal cross-sections have been computed with $N_{\text{rep}} = 400$ replicas of nNNPDF2.0, which are subsequently used to

JHEP09(2020)183

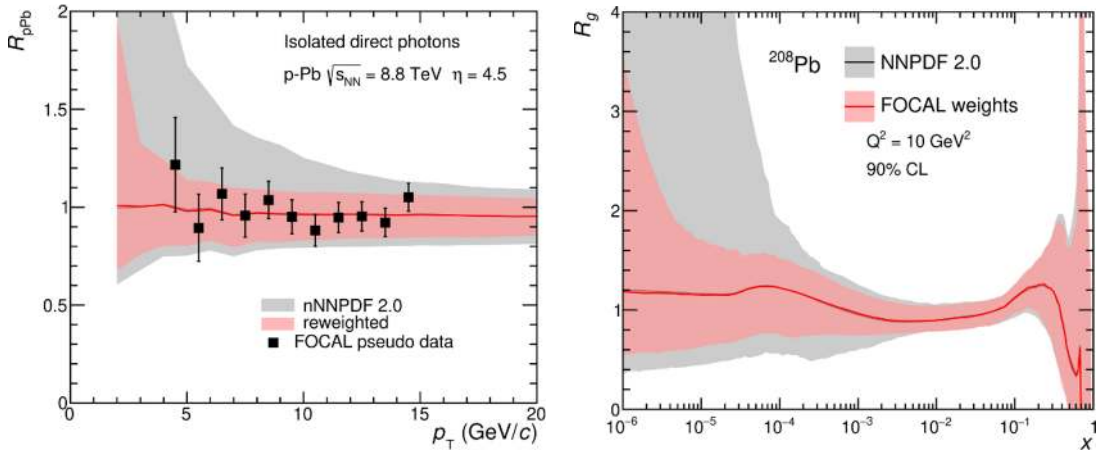

Figure 23. Left: the nuclear modification factor $R_{\rm pPb}(p_T^\gamma)$ for direct photon production in pPb collisions at $\sqrt{s} = 8.8\,$TeV for a rapidity of $\eta_\gamma = 4.5$ as a function of the photon transverse momentum p_T^γ. The theoretical predictions are compared with the FoCal pseudo-data for two sets of input nPDFs: the original nNNPDF2.0 set, and the variant that has been reweighted with FoCal projections. Here the FoCal pseudo data assumes the central value of the nNNPDF2.0 prediction. Right: the gluon nuclear modification factor $R_g(x, Q)$ for $Q^2 = 10$ GeV^2 for both the original and the reweighted nNNPDF2.0 fits. The nPDF uncertainties correspond in both cases to the 90% confidence level intervals.

account for the impact of the FoCal pseudo-data by means of Bayesian reweighting.[5] Figure 23 displays the nuclear modification factor $R_{\rm pPb}(p_T^\gamma)$ for direct photon production in pPb collisions at $\sqrt{s} = 8.8\,$TeV for a rapidity of $\eta_\gamma = 4.5$ as a function of the photon's transverse momentum p_T^γ. The theoretical predictions based on NLO QCD theory are compared with the FoCal pseudo-data for two sets of input nPDFs: the original nNNPDF2.0 set, and the variant that has been reweighted with the FoCal projections. Here the central value of the FoCal pseudo-data has been chosen to be the same as that of the nNNPDF2.0 prediction. In the right panel of figure 23 we show the gluon nuclear modification factor $R_g(x, Q)$ for $Q^2 = 10\,$GeV^2 for both the original and the reweighted nNNPDF2.0 fits. In all cases, the nPDF uncertainty bands correspond to the 90% confidence level intervals.

From the results of figure 23, one finds that the FoCal measurements would still impact the uncertainties of the nuclear gluon modifications at small-x, especially in the upper limit of the uncertainty band. The effective number of replicas in this case is $N_{\rm eff} = 345$. Note that nNNPDF2.0 exhibits a preference for $R_{\rm pPb} \simeq 1$, and thus shadowing is not favored in the gluon sector, consistent with the results reported in figure 11. On the other hand, nNNPDF2.0 does not contain any dataset with particular sensitivity to the nuclear gluon modifications, implying that the projections for the impact of FoCal in the global nPDF analysis could be somewhat over-optimistic (see also the discussion in section 6).

Crucially, however, we have assumed in this exercise that the central value of the FoCal measurement would be unchanged compared to the initial baseline prediction. In figure 24 we display instead the results of the reweighting for a scenario in which the FoCal pseudodata have a value of $R_{\rm pPb} \simeq 0.6$. In this case, the effective number of

[5] We are grateful to Marco van Leeuwen for providing us with the results presented here.

JHEP09(2020)183

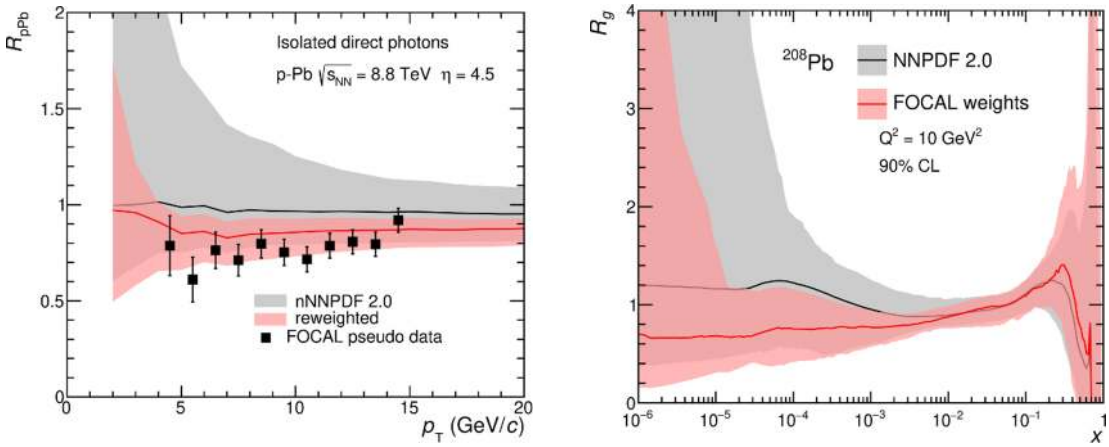

Figure 24. Same as figure 23 now for the case where the FoCal pseudo data has been generated under the assumption that $R_{\rm pPb} \simeq 0.6$ rather than based on the nNNPDF2.0 central value.

replicas is much smaller, $N_{\rm eff} = 117$, indicating that the FoCal data are adding a significant amount of new information to the global fit. Here the resulting value for the gluon nuclear modification ratio at small-x would be $R_g \simeq 0.7$.[6] Therefore, this analysis indicates that FoCal measurements could be sensitive either to the gluon shadowing effects or to possible non-linear QCD dynamics. To disentangle one from the other, a dedicated analysis of the χ^2 and nPDF behavior in the small-x region would be required, following the approach developed in ref. [118].

5.3 Inclusive hadron production in pA collisions

The inclusive production of pions and kaons in hadronic collisions provides information not only on the initial state (parton distribution functions) but also on the final-state hadronization mechanism of partons into hadrons. The latter is described by the fragmentation functions (FFs), which are extracted from experimental data by means of a global analysis akin to that of the PDFs [97, 119–124]. Likewise, in proton-nuclear collisions the production of identified hadrons can provide information on the initial state nuclear PDFs as well the parton-to-hadron hadronization in the presence of cold nuclear matter effects.

In figure 25, we display the nuclear modification ratio $R^{\pi^0}_{\rm Pb}$ for the production of neutral pions in proton-lead collisions as a function of the pion transverse momentum p_T. The theoretical calculations are based on NLO QCD and use the DSS14 hadron fragmentation functions [119] for both the nNNPDF2.0 and EPPS16 predictions.[7] Moreover, the central values and 90% CL uncertainties are provided for RHIC kinematics, corresponding to $\sqrt{s} = 200\,$GeV, and for LHC kinematics, where $\sqrt{s} = 8.16$ TeV. In both cases, the pions are assumed to be measured at central rapidities, $y_{\pi^0} = 0$. See refs. [120, 125] for additional details regarding the theoretical calculation of inclusive pion production in hadronic collisions.

[6]Note that the reweighting technique may lead to unreliable uncertainty bands when using data values that fall outside the predictions produced by the prior.

[7]We are grateful to Ilkka Helenius for providing us with the results of this calculation.

JHEP09(2020)183

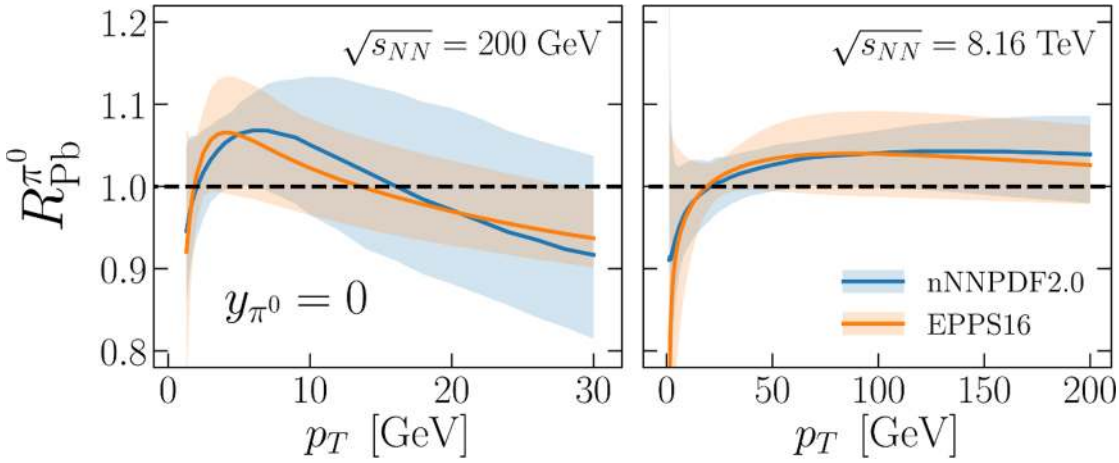

Figure 25. The nuclear modification ratio $R^{\pi^0}_{\rm Pb}$ for the production of neutral pions in proton-lead collisions as a function of the pion transverse momentum p_T. We provide theoretical predictions based on NLO QCD and the DSS14 hadron fragmentation functions both for nNNPDF2.0 and EPPS16, with the corresponding nPDF uncertainties in each case. Results are provided for the RHIC kinematics (left), corresponding to $\sqrt{s} = 200\,\text{GeV}$, and for the LHC kinematics (right), where $\sqrt{s} = 8.16\,\text{TeV}$, and in both cases pions are produced centrally, $|y_{\pi^0}| = 0$.

From figure 25 we can see that the nNNPDF2.0 prediction for $R^{\pi^0}_{\rm Pb}$ is consistent with unity within uncertainties for all values of p_T both at RHIC and LHC kinematics. At RHIC kinematics, we find that the ratio is less than one at the smallest p_T values, becomes $R > 1$ between $p_T = 3$ and $17\,\text{GeV}$, and then goes back to $R < 1$. Since inclusive hadron production is dominated by quark-gluon scattering, in particular the scattering of valence quarks for neutral pion production, this behavior is consistent with the results shown in figure 14. From low to high p_T, one moves from the shadowing region to the anti-shadowing enhancement, and ends in the region sensitive to EMC suppression. A similar explanation can be made for the trends in $R^{\pi^0}_{\rm Pb}$ at the LHC kinematics. However, here the ratio $p_T/\sqrt{s}$ does not become large enough to reach the EMC region, and thus the ratio remains larger than one for most of the p_T range as a result of anti-shadowing effects. Lastly, the EPPS16 predictions agree with the nNNPDF2.0 result well within uncertainties, reflecting the underlying consistency at the nPDF level.

Overall, the results of figure 25 confirm that inclusive hadron production in proton-nucleus collisions can provide a handle on the nuclear PDF modifications at medium and large-x, although an optimal interpretation of the experimental data can only be achieved by the simultaneous determination of the nPDFs together with the hadron fragmentation functions.

6 Summary and outlook

In this work we have presented a model-independent global determination of nuclear parton distributions by incorporating the constraints from nuclear DIS structure functions and gauge boson production in proton-lead collisions. We have demonstrated that a satisfactory description of all the fitted data sets can be achieved, highlighting the reliability of the QCD

JHEP09(2020)183

factorization paradigm in the heavy nuclear sector. Our results demonstrate significant nuclear effects among the quark flavors in nuclei, in particular a shadowing of the up and down quark distributions in heavy nuclei such as lead. Nuclear modifications are found also in the strangeness of heavier nuclei, displaying a suppression with respect to the free proton across a large region of x. In addition, we have shown that upon releasing the momentum and valence sum rule constraints, the data prefer integral values that agree with QCD expectations for all values of A.

We have also explored some phenomenological implications of the nNNPDF2.0 determination. We first compared nNNPDF2.0 theoretical predictions with ATLAS measurements of isolated photon production at the LHC, an important hard probe in proton-lead collisions. We then studied the impact on the small-x nuclear gluon PDF from the forward isolated photons production at the FoCal upgrade of the ALICE detector. Lastly, we analyzed our theory predictions for neutral pion production in proton-lead collisions at RHIC and LHC center-of-mass energies. Apart from these applications, the nNNPDF2.0 PDF set can be used as input to theoretical predictions for a range of other hard processes in pPb and PbPb collisions, in particular for heavy ion collisions involving lighter nuclei, in comparisons with non-perturbative nuclear models, and with QCD calculations at small-x involving dense nuclear matter. We also expect the nNNPDF2.0 release to be used in future proton global PDF fits to estimate the theory uncertainties associated with neutrino scattering data [10], and also in high-energy astroparticle physics processes that involve hard scattering on nuclei [126, 127].

While the input data set used in this work allowed for a state-of-the-art determination of the nuclear quarks and anti-quarks, it only provided loose constraints on the nuclear gluon PDF, especially for heavier nuclei where uncertainties are relatively large. To bypass this limitation, the next step in the nNNPDF family of nuclear PDF fits will be to include additional datasets that provide direct information on the nuclear gluon modifications. In addition to the isolated photon production measurements discussed in section 5.1, perhaps the most attractive candidate in this respect is dijet production in pPb collisions. Measurements of dijet production from Run I in pp collisions have been recently analyzed in the framework of NNLO QCD theory in ref. [128], demonstrating a good compatibility with the global dataset and a marked constraining power on the large-x gluon. In the corresponding pPb case, an EPPS16-based profiling analysis [37] of CMS dijet data at $\sqrt{s} = 5.02\,\text{TeV}$ [129] revealed a significant pull of this measurement on the nuclear gluon modifications.

Another process that is known to provide important information on the nuclear gluon PDF is charmed meson production, in particular from the LHCb measurements in the forward region [32]. This process offers unique sensitivity to the small-x (n)PDFs down to $x \simeq 10^{-6}$, as was demonstrated by proton [127, 130, 131] and nuclear studies [35, 132]. Fully exploiting the constraints provided by these measurements requires, as for the rest of hard probes in nuclear collisions, a consistent theoretical and methodological treatment of charm production in both proton and nuclear global QCD analyses.

On a longer timescale, one might aim to achieve a determination of the proton and nuclear PDFs simultaneously from a universal analysis, thus bypassing the need to in-

JHEP09(2020)183

clude proton information by means of the proton boundary condition penalty. In the same spirit of the QCD analyses of proton PDFs and fragmentation functions presented in refs. [97, 133], such an integrated fit of proton and nuclear PDFs would ensure the ultimate theoretical and methodological consistency of the determination of the nuclear modifications of the free-nucleon quark and gluon structure.

The nNNPDF2.0 determination is available in the `LHAPDF6` library [134] for all relevant nuclei from $A = 1$ to $A = 208$. The nNNPDF2.0 sets are available both for the nPDFs of bound protons, $f^{(p/A)}(x, Q^2)$, and those of bound nucleons, $f^{(N/A)}(x, Q^2)$, following the conventions in section 3.1. Each of these sets is composed by $N_{\rm rep} = 250$ correlated replicas, see section 5 of [13] for their usage prescriptions. The naming convention used for the sets is the following:

$f^{(N/A)}(x, Q^2)$	$f^{(p/A)}(x, Q^2)$
nNNPDF20_nlo_as_0118_N1	nNNPDF20_nlo_as_0118_N1
nNNPDF20_nlo_as_0118_D2	nNNPDF20_nlo_as_0118_p_A2_Z1
nNNPDF20_nlo_as_0118_He4	nNNPDF20_nlo_as_0118_p_A4_Z2
nNNPDF20_nlo_as_0118_Li6	nNNPDF20_nlo_as_0118_p_A6_Z3
nNNPDF20_nlo_as_0118_Be9	nNNPDF20_nlo_as_0118_p_A9_Z4
nNNPDF20_nlo_as_0118_C12	nNNPDF20_nlo_as_0118_p_A12_Z6
nNNPDF20_nlo_as_0118_N14	nNNPDF20_nlo_as_0118_p_A14_Z7
nNNPDF20_nlo_as_0118_Al27	nNNPDF20_nlo_as_0118_p_A27_Z13
nNNPDF20_nlo_as_0118_Ca40	nNNPDF20_nlo_as_0118_p_A40_Z20
nNNPDF20_nlo_as_0118_Fe56	nNNPDF20_nlo_as_0118_p_A56_Z26
nNNPDF20_nlo_as_0118_Cu64	nNNPDF20_nlo_as_0118_p_A64_Z29
nNNPDF20_nlo_as_0118_Ag108	nNNPDF20_nlo_as_0118_p_A108_Z47
nNNPDF20_nlo_as_0118_Sn119	nNNPDF20_nlo_as_0118_p_A119_Z50
nNNPDF20_nlo_as_0118_Xe131	nNNPDF20_nlo_as_0118_p_A131_Z54
nNNPDF20_nlo_as_0118_Au197	nNNPDF20_nlo_as_0118_p_A197_Z79
nNNPDF20_nlo_as_0118_Pb208	nNNPDF20_nlo_as_0118_p_A208_Z82

Supplementary sets such as:

nNNPDF20_nlo_as_0118_O16	nNNPDF20_nlo_as_0118_p_A16_Z8
nNNPDF20_nlo_as_0118_W184	nNNPDF20_nlo_as_0118_p_A184_Z74

and additional variants of the nNNPDF2.0 NLO fit present in this work, such as the fits without the momentum and valence sum rules and a $N_{\rm rep} = 1000$ replica set for lead, are available on the NNPDF collaboration website:

http://nnpdf.mi.infn.it/for-users/nnnpdf2-0/

Acknowledgments

We are grateful to our colleagues of the NNPDF collaboration for support and stimulating discussions, and in particular to Emanuele R. Nocera for his continuous assistance during

JHEP09(2020)183

this work. We thank Kari Eskola, Petja Paakkinen, and Hannu Paukkunen for discussions about the EPPS16 analysis. We thank Marco van Leeuwen for providing the predictions for FoCal and Ilkka Helenius for those of pion production in heavy ion collisions. R. A. K., J. E., and J. R. are supported by the Netherlands Organization for Scientific Research (NWO). We acknowledge grants of computer capacity from the Finnish Grid and Cloud Infrastructure (persistent identifier urn:nbn:fi:research-infras-2016072533).

A PDF sensitivity of input cross-sections

In this appendix we indicate the PDF sensitivity of all processes used as input in the nNNPDF2.0 determination by providing their explicit cross-section expressions at leading order (LO) in perturbative QCD. We provide these expressions for a general nuclear target with Z protons and $(A-Z)$ neutrons in terms of bound proton distributions of the physical and evolution bases, and in terms of the average bound nucleon PDFs. In what follows, we further simplify the conventions adopted in section 3.1 to

$$f \equiv f^{(p/A)}\,, \tag{A.1}$$

$$f_A \equiv f^{(N/A)}\,. \tag{A.2}$$

1. Lepton-nucleus scattering. The double differential cross-section for the DIS of a charged lepton off a nucleus with mass number A is given by

$$\frac{d^2\sigma^{i,l^\pm}}{dxdQ^2}(x,Q^2,A) = \frac{2\pi\alpha^2}{xQ^4}\eta^i\left[Y_+F_2^i(x,Q^2,A) \mp Y_-xF_3^i(x,Q^2,A) - y^2F_L^i(x,Q^2,A)\right]\,, \tag{A.3}$$

with i = NC, CC, $\eta^{\rm NC} = 1$, $\eta^{\rm CC} = (1\pm\lambda)^2\eta_W$, where η_W denotes the squared ratio of the W-boson couplings and propagator with respect to those of the photon, and λ is the helicity of the incoming lepton. In the case of NC DIS we restrict ourselves to photon-mediated processes, for which $F_3^{\rm NC} = 0$. The usual DIS kinematic variables are defined as $Y_\pm = 1 \pm (1-y)^2$ and

$$x = \frac{Q^2}{2P\cdot q}, \quad Q^2 = -q^2, \quad y = \frac{q\cdot P}{k\cdot P}\,. \tag{A.4}$$

A1. Neutral-current DIS. For NC DIS, the underlying process is lepton-nucleus scattering mediated by a virtual photon exchange, $l^\pm + A \xrightarrow{\gamma} l^\pm + X$. Since the available data is at $Q^2 \ll M_Z$, the contributions from Z boson exchange can be neglected. The double-differential cross-section for the scattering of a charged lepton off a nucleus with atomic and mass numbers Z and A is proportional at leading order to the F_2 structure function which can be expressed as

$$\begin{aligned} F_2^{\rm NC} &= \frac{x}{9A}\Big[(4A-3Z)d^+ + As^+ + (A+3Z)u^+\Big] = \frac{x}{9}\Big[4u_A^+ + d_A^+ + s_A^+\Big] \\ &= \frac{x}{18A}\Big[4A\Sigma + AT_8 + 3(2Z-A)T_3\Big]\,, \end{aligned} \tag{A.5}$$

JHEP09(2020)183

Here we define the structure function per nucleon, rather than the structure function of the nucleus as a whole. We can see from the above expression that for an isoscalar nucleus $A = 2Z$ the contribution proportional to T_3 vanishes and the nuclear structure functions depend only on Σ and T_8 via the $\Sigma + (1/4)T_8$ combination. At LO only eq. (A.5) is relevant for the description of neutral-current DIS since the longitudinal structure function vanishes due to the Callan-Gross relation, $F_L^{\rm NC} = F_2^{\rm NC} - 2xF_1^{\rm NC} = 0$.

B1. Charged-current DIS. For CC DIS, the underlying process is neutrino scattering off of a nucleus via the exchange of a W boson, $\nu + A \xrightarrow{W^-} l^- + X$ and $\bar{\nu} + A \xrightarrow{W^+} +l^+ + X$. Due to the fact that W^+ and W^- bosons couple to different quark flavors, the difference between neutrino and anti-neutrino structure functions provides a handle on quark flavor separation. Here we provide the expressions for $\nu(\bar{\nu})A$ scattering, the expressions for the conjugate process involving the CC scattering of charged leptons is the same. The expressions for the inclusive CC structure functions via W^--boson exchange, $\bar{\nu} + A \xrightarrow{W^-} +l^+X$, are the following:

$$\begin{aligned}
F_2^{\bar{\nu}A \xrightarrow{W^-} l^+X} &= \frac{2x}{A}\Big[|V_{ud}|^2\Big(Z(u+\bar{d}) + (A-Z)(d+\bar{u})\Big) \\
&\qquad + |V_{us}|^2\Big(Zu + (A-Z)d + A\bar{s}\Big)\Big] \\
&= 2x\Big[|V_{ud}|^2\Big(u_A + \bar{d}_A\Big) + |V_{us}|^2\Big(u_A + \bar{s_A}\Big)\Big] \\
&= \frac{x}{6A}\Big[|V_{ud}|^2\Big(2A(2\Sigma + T_8) + 6(2Z-A)V_3\Big) \\
&\qquad + |V_{us}|^2\Big(3(2Z-A)T_3 + 3(2Z-A)V_3 + A(4\Sigma - T_8 + 3V_8)\Big)\Big]
\end{aligned} \tag{A.6}$$

$$\begin{aligned}
F_3^{\bar{\nu}A \xrightarrow{W^-} l^+X} &= \frac{2}{A}\Big[|V_{ud}|^2\Big(Z(u-\bar{d}) + (A-Z)(d-\bar{u})\Big) + |V_{us}|^2\Big(Zu + (A-Z)d - A\bar{s}\Big)\Big] \\
&= 2\Big[|V_{ud}|^2\Big(u_A - \bar{d}_A\Big) + |V_{us}|^2\Big(u_A - \bar{s_A}\Big)\Big] \\
&= \frac{-1}{6A}\Big[|V_{ud}|^2\Big(6(A-2Z)T_3 - 2A(2V + V_8)\Big) \\
&\qquad + |V_{us}|^2\Big(+ 3(A-2Z)T_3 + 3(A-2Z)V_3 + A(-3T_8 - 4V + V_8)\Big)\Big]
\end{aligned} \tag{A.7}$$

Here we consider only the contribution from up-, down-, and strange-initiated processes for simplicity. The generalization to heavy-quark initiated processes is straightforward. An interesting observation from these expressions is that F_2 and F_3 provide complementary information on quark flavor separation.

The corresponding expressions for the inclusive charged-current structure functions in the case of neutrino scattering via W^+-boson exchange, $\nu + A \xrightarrow{W^+} +l^-X$, are given by

$$\begin{aligned}
F_2^{\nu A \xrightarrow{W^+} l^-X} &= \frac{2x}{A}\Big[|V_{ud}|^2\Big(Z(\bar{u}+d) + (A-Z)(u+\bar{d})\Big) \\
&\qquad + |V_{us}|^2\Big(Z\bar{u} + (A-Z)\bar{d} + As\Big)\Big]
\end{aligned}$$

JHEP09(2020)183

$$
\begin{aligned}
&= 2x\Big[|V_{ud}|^2\Big(\bar{u}_A + d_A\Big) + |V_{us}|^2\Big(\bar{u}_A + s_A\Big)\Big] \\
&= \frac{x}{6A}\Big[|V_{ud}|^2\Big(2A(2\Sigma + T_8) + 6(A-2Z)V_3\Big) \\
&\qquad + |V_{us}|^2\Big(A(4\Sigma - T_8 - 3V_8) + 3(2Z-A)T_3 + 3(A-2Z)V_3\Big)\Big]
\end{aligned} \tag{A.8}
$$

$$
\begin{aligned}
F_3^{\nu A \xrightarrow{W^+} l^- X} &= \frac{2}{A}\Big[|V_{ud}|^2\Big(Z(d-\bar{u}) + (A-Z)(u-\bar{d})\Big) \\
&\qquad + |V_{us}|^2\Big(-Z\bar{u} - (A-Z)\bar{d} + As\Big)\Big] \\
&= 2\Big[|V_{ud}|^2\Big(-\bar{u}_A + d_A\Big) + |V_{us}|^2\Big(-\bar{u}_A + s_A\Big)\Big] \\
&= \frac{1}{6A}\Big[|V_{ud}|^2\Big(2A(2V + V_8) + 6(A-2Z)T_3\Big) \\
&\qquad + |V_{us}|^2\Big(A(-3T_8 + 4V - V_8) + 3(A-2Z)T_3 + 3(2Z-A)V_3\Big)\Big]
\end{aligned} \tag{A.9}
$$

We now turn to the exclusive CC charm production structure functions, required for the description of the NuTeV cross-sections. Following the FONLL treatment of heavy quark structure functions, the charm contribution to the inclusive structure function is defined by all terms where the charm quark couples to the W boson. Note that this definition is somewhat different from the experimental definition, where charm structure functions are identified by requesting the presence of charm in the final state, while the theory definition includes charm-initiated contributions as well. The charm production structure functions in the case of antineutrino- and neutrino-initiated scattering reads for W^- as

$$
\begin{aligned}
F_2^{\bar{\nu} A \xrightarrow{W^-} l^+ cX} &= \frac{2x}{A}\Big[|V_{cd}|^2\Big(Z\bar{d} + (A-Z)\bar{u}\Big) + |V_{cs}|^2 A\bar{s}\Big] \\
&= 2x\Big[|V_{cd}|^2\bar{d}_A + |V_{cs}|^2\bar{s}_A\Big] \\
&= \frac{x}{6A}\Big[|V_{cd}|^2\Big(A(2\Sigma + T_8 - 2V - V_8) + 3(A-2Z)T_3 + 3(2Z-A)V_3\Big) \\
&\qquad + |V_{cs}|^2 2A\Big(\Sigma - T_8 - V + V_8\Big)\Big]
\end{aligned} \tag{A.10}
$$

$$
\begin{aligned}
F_3^{\bar{\nu} A \xrightarrow{W^-} l^+ cX} &= \frac{2}{A}\Big[|V_{cd}|^2\Big(-Z\bar{d} - (A-Z)\bar{u}\Big) - |V_{cs}|^2 A\bar{s}\Big] \\
&= -2\Big[|V_{cd}|^2\bar{d}_A + |V_{cs}|^2\bar{s}_A\Big] \\
&= \frac{-1}{6A}\Big[|V_{cd}|^2\Big(3(A-2Z)T_3 + 3(2Z-A)V_3 + A(2\Sigma + AT_8 - 2AV - AV_8)\Big) \\
&\qquad + |V_{cs}|^2 2A\Big(+\Sigma - T_8 - V + V_8\Big)\Big]
\end{aligned} \tag{A.11}
$$

and for W^+:

$$
\begin{aligned}
F_2^{\nu A \xrightarrow{W^+} l^- cX} &= \frac{2x}{A}\Big[|V_{cd}|^2\Big(Zd + (A-Z)u\Big) + |V_{cs}|^2 As\Big] \\
&= 2x\Big[|V_{cd}|^2 d_A + |V_{cs}|^2 s_A\Big]
\end{aligned}
$$

JHEP09(2020)183

$$= \frac{x}{6A}\Big[|V_{cd}|^2\Big(A(2\Sigma + T_8 + 2V + V_8) + 3(A-2Z)T_3 + 3(A-2Z)V_3\Big)$$
$$+ |V_{cs}|^2 2A\Big(+ \Sigma - T_8 + V - V_8\Big)\Big] \tag{A.12}$$

$$F_3^{\nu A \xrightarrow{W^+} l^- cX} = \frac{-2}{A}\Big[|V_{cd}|^2\Big(+ Zd + (A-Z)u\Big) + |V_{cs}|^2 As\Big]$$
$$= -2\Big[|V_{cd}|^2 d_A + |V_{cs}|^2 s_A\Big]$$
$$= \frac{-1}{6A}\Big[|V_{cd}|^2\Big(A(2\Sigma + T_8 + 2V + V_8) + 3(A-2Z)T_3 + 3(A-2Z)V_3\Big)$$
$$+ |V_{cs}|^2 2A\Big(+ \Sigma - T_8 + V - V_8\Big)\Big] \tag{A.13}$$

2. Weak boson production. Here we provide the corresponding leading order expressions for weak gauge boson production in proton-nucleus collisions. Experimental measurements for NC observables are provided in terms of the invariant mass M and rapidity of the dilepton final state, while CC measurements are binned in terms of the charged lepton rapidity. Here we show the LO expressions without accounting for the gauge boson decay, so in the case of $W^\pm$ production the connection with the experimental data is somewhat less direct than for Z boson production.

A2. Neutral current DY. We start by discussing NC Drell-Yan (DY), namely Z-boson production in proton-nucleus collisions, $p + A \xrightarrow{Z} l^+l^-$. The leading order expressions are given by

$$\frac{d\sigma_Z^{pA}}{dM^2dy} \propto A\Big[a_u\Big(u_1\bar{u}_2 + \bar{u}_1u_2\Big) + a_d\Big(d_1\bar{d}_2 + \bar{d}_1d_2 + s_1\bar{s}_2 + \bar{s}_1s_2\Big)\Big] \tag{A.14}$$

where $q_1 = q(x_1, M^2)$ and $q_2 = q(x_2, M^2)$ and the Drell-Yan kinematics at LO are:

$$x_{1,2} = \frac{M}{\sqrt{s}}e^{\pm y}\,, \qquad M^2 = x_1x_2s\,, \tag{A.15}$$

with $\sqrt{s}$ being the center-of-mass energy of the collision, M being the invariant mass of the final state, and y the rapidity of the final state system. The effective weak couplings are given by

$$a_q = (\bar{g}_V^q + \bar{g}_A^q), \qquad \bar{g}_V^f = (t_3^{(f)} - 2Q_f\sin^2\theta_W), \qquad \bar{g}_A^f = t_3^{(f)}$$

where $t_3^{(f)}$ is the weak isospin of fermion f, Q_f is the electric charge and θ_W is Weinberg's angle.

B2. Charged current DY. The corresponding LO expressions in the case of W^- and W^+ boson production in pA collisions are given by

$$\frac{d\sigma_{W^-}^{pA,}}{dM^2dy} \propto |V_{ud}|^2\Big(Z(d_1\bar{u}_2 + \bar{u}_1d_2) + (A-Z)(\bar{d}_1u_2 + u_1\bar{d}_2)\Big)$$
$$+ |V_{us}|^2\Big(Z(d_1\bar{s}_2 + \bar{s}_1d_2) + (A-Z)(u_1\bar{s}_2 + \bar{s}_1u_2)\Big) \tag{A.16}$$

JHEP09(2020)183

and:

$$\frac{d\sigma^{pA,}_{W^+}}{dM^2dy} \propto |V_{ud}|^2\Big(Z(u_1\bar{d}_2+\bar{d}_1u_2)+(A-Z)(d_1\bar{u}_2+\bar{u}_1d_2)\Big) \\ +|V_{us}|^2\Big(Z(u_1\bar{s}_2+\bar{s}_1u_2)+(A-Z)(d_1\bar{s}_2+\bar{s}_1d_2)\Big) \tag{A.17}$$

where in this case y stands for the W boson rapidity, which in general is different from the pseudo-rapidity of the charged lepton that is measured by experiments. However, in either case the PDF combinations that enter at LO remain the same. From the comparison between the LO expressions of the DIS structure functions and the W and Z production cross-sections in pA collisions, it is clear that each group of process is sensitive to a different combination of quark and antiquark PDFs. Therefore, their combination (due to the PDF universality in QCD) into a global QCD analysis provides a unique handle for a robust separation between quark and antiquark flavors in nuclei.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] J. Rojo, *The partonic content of nucleons and nuclei*, arXiv:1910.03408 [INSPIRE].

[2] J.J. Ethier and E.R. Nocera, *Parton distributions in nucleons and nuclei*, *Ann. Rev. Nucl. Part. Sci.* (2020) 1 [arXiv:2001.07722] [INSPIRE].

[3] J. Gao, L. Harland-Lang and J. Rojo, *The structure of the proton in the LHC precision era*, *Phys. Rept.* **742** (2018) 1 [arXiv:1709.04922] [INSPIRE].

[4] N. Armesto et al., *Heavy ion collisions at the LHC - Last Call for predictions*. *J. Phys. G* **35** (2008) 054001 [INSPIRE].

[5] STAR collaboration, *Experimental and theoretical challenges in the search for the quark gluon plasma: the STAR collaboration's critical assessment of the evidence from RHIC collisions*, *Nucl. Phys. A* **757** (2005) 102 [nucl-ex/0501009] [INSPIRE].

[6] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [arXiv:1410.8849] [INSPIRE].

[7] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, arXiv:1912.10053 [INSPIRE].

[8] L.A. Harland-Lang, A.D. Martin, P. Motylinski and R.S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) 204 [arXiv:1412.3989] [INSPIRE].

[9] S. Alekhin, J. Blümlein, S. Moch and R. Placakyte, *Parton distribution functions, α_s, and heavy-quark masses for LHC Run II*, *Phys. Rev. D* **96** (2017) 014011 [arXiv:1701.05838] [INSPIRE].

[10] NNPDF collaboration, *Nuclear uncertainties in the determination of proton PDFs*, *Eur. Phys. J. C* **79** (2019) 282 [arXiv:1812.09074] [INSPIRE].

JHEP09(2020)183

[11] K.J. Eskola, P. Paakkinen, H. Paukkunen and C.A. Salgado, *EPPS16: nuclear parton distributions with LHC data*, Eur. Phys. J. C **77** (2017) 163 [arXiv:1612.05741] [INSPIRE].

[12] K. Kovarik et al., *nCTEQ15 — Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework*, Phys. Rev. D **93** (2016) 085037 [arXiv:1509.00792] [INSPIRE].

[13] NNPDF collaboration, *Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider*, Eur. Phys. J. C **79** (2019) 471 [arXiv:1904.00018] [INSPIRE].

[14] M. Walt, I. Helenius and W. Vogelsang, *Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO*, Phys. Rev. D **100** (2019) 096015 [arXiv:1908.03355] [INSPIRE].

[15] H. Khanpour and S. Atashbar Tehrani, *Global analysis of nuclear parton distribution functions and their uncertainties at next-to-next-to-leading order*, Phys. Rev. D **93** (2016) 014026 [arXiv:1601.00939] [INSPIRE].

[16] ALICE collaboration, *Measurement of charged jet production cross sections and nuclear modification in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Phys. Lett. B **749** (2015) 68 [arXiv:1503.00681] [INSPIRE].

[17] ALICE collaboration, *Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ *TeV*, Phys. Lett. B **760** (2016) 720 [arXiv:1601.03658] [INSPIRE].

[18] ALICE collaboration, *Measurement of dijet* k_T *in p-Pb collisions at* $\sqrt{s}_{NN} = 5.02$ *TeV*, Phys. Lett. B **746** (2015) 385 [arXiv:1503.03050] [INSPIRE].

[19] ATLAS collaboration, *Transverse momentum, rapidity, and centrality dependence of inclusive charged-particle production in* $\sqrt{s_{NN}} = 5.02$ *TeV p + Pb collisions measured by the ATLAS experiment*, Phys. Lett. B **763** (2016) 313 [arXiv:1605.06436] [INSPIRE].

[20] CMS collaboration, *Studies of dijet transverse momentum balance and pseudorapidity distributions in pPb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Eur. Phys. J. C **74** (2014) 2951 [arXiv:1401.4433] [INSPIRE].

[21] ALICE collaboration, *Measurement of W-boson production in p–pb collisions at* $\sqrt{s_{NN}} =$ *5.02 tev with alice at the LHC*, J. Phys. Conf. Ser. **612** (2015) 012009.

[22] *Measurement of* $W \to \mu\nu$ *production in p+Pb collision at* $\sqrt{s_{NN}} = 5.02$ *TeV with ATLAS detector at the LHC*, ATLAS-CONF-2015-056 (2015).

[23] ATLAS collaboration, *Z boson production in p+Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV measured with the ATLAS detector*, Phys. Rev. C **92** (2015) 044915 [arXiv:1507.06232] [INSPIRE].

[24] CMS collaboration, *Study of Z boson production in pPb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Phys. Lett. B **759** (2016) 36 [arXiv:1512.06461] [INSPIRE].

[25] CMS collaboration, *Study of W boson production in pPb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Phys. Lett. B **750** (2015) 565 [arXiv:1503.05825] [INSPIRE].

[26] CMS collaboration, *Charm-tagged jet production in pPb collisions at* 5.02 *TeV and pp collisions at* 2.76 *TeV*, CMS-PAS-HIN-15-012 (2015).

JHEP09(2020)183

[27] ALICE collaboration, *Measurement of D-meson production versus multiplicity in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *JHEP* **08** (2016) 078 [arXiv:1602.07240] [INSPIRE].

[28] ALICE collaboration, *Measurement of electrons from heavy-flavour hadron decays in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *Phys. Lett. B* **754** (2016) 81 [arXiv:1509.07491] [INSPIRE].

[29] ALICE collaboration, *Measurement of prompt D-meson production in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *Phys. Rev. Lett.* **113** (2014) 232301 [arXiv:1405.3452] [INSPIRE].

[30] CMS collaboration, *Transverse momentum spectra of inclusive b jets in pPb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, *Phys. Lett. B* **754** (2016) 59 [arXiv:1510.03373] [INSPIRE].

[31] CMS collaboration, *Study of B meson production in p+Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV using exclusive hadronic decays*, *Phys. Rev. Lett.* **116** (2016) 032301 [arXiv:1508.06678] [INSPIRE].

[32] LHCb collaboration, *Study of prompt* D^0 *meson production in p-Pb collisions at* $\sqrt{s_{NN}} = 5$ *TeV*, *JHEP* **10** (2017) 090 [arXiv:1707.02750] [INSPIRE].

[33] LHCb collaboration, *Measurement of* B^+, B^0 *and* Λ_b^0 *production in p-Pb collisions at* $\sqrt{s_{NN}} = 8.16$ *TeV*, *Phys. Rev. D* **99** (2019) 052011 [arXiv:1902.05599] [INSPIRE].

[34] A. Kusina et al., *Vector boson production in pPb and PbPb collisions at the LHC and its impact on nCTEQ15 PDFs*, *Eur. Phys. J. C* **77** (2017) 488 [arXiv:1610.02925] [INSPIRE].

[35] A. Kusina, J.-P. Lansberg, I. Schienbein and H.-S. Shao, *Gluon shadowing in heavy-flavor production at the LHC*, *Phys. Rev. Lett.* **121** (2018) 052004 [arXiv:1712.07024] [INSPIRE].

[36] N. Armesto, H. Paukkunen, J.M. Penín, C.A. Salgado and P. Zurita, *An analysis of the impact of LHC Run I proton-lead data on nuclear parton densities*, *Eur. Phys. J. C* **76** (2016) 218 [arXiv:1512.01528] [INSPIRE].

[37] K.J. Eskola, P. Paakkinen and H. Paukkunen, *Non-quadratic improved Hessian PDF reweighting and application to CMS dijet measurements at 5.02 TeV*, *Eur. Phys. J. C* **79** (2019) 511 [arXiv:1903.09832] [INSPIRE].

[38] S. Forte, L. Garrido, J.I. Latorre and A. Piccione, *Neural network parametrization of deep inelastic structure functions*, *JHEP* **05** (2002) 062 [hep-ph/0204232] [INSPIRE].

[39] NNPDF collaboration, *Unbiased determination of the proton structure function F(2)**p with faithful uncertainty estimation*, *JHEP* **03** (2005) 080 [hep-ph/0501067] [INSPIRE].

[40] NNPDF collaboration, *Neural network determination of parton distributions: the nonsinglet case*, *JHEP* **03** (2007) 039 [hep-ph/0701127] [INSPIRE].

[41] NNPDF collaboration, *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys. B* **809** (2009) 1 [*Erratum ibid.* **816** (2009) 293] [arXiv:0808.1231] [INSPIRE].

[42] NNPDF collaboration, *Update on neural network parton distributions: NNPDF1.1*, arXiv:0811.2288 [INSPIRE].

[43] NNPDF collaboration, *Fitting parton distribution data with multiplicative normalization uncertainties*, *JHEP* **05** (2010) 075 [arXiv:0912.2276] [INSPIRE].

[44] R.D. Ball et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, *Nucl. Phys. B* **838** (2010) 136 [arXiv:1002.4407] [INSPIRE].

JHEP09(2020)183

[45] R.D. Ball et al., *Impact of heavy quark masses on parton distributions and LHC phenomenology*, Nucl. Phys. B **849** (2011) 296 [arXiv:1101.1300] [INSPIRE].

[46] NNPDF collaboration, *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, Nucl. Phys. B **855** (2012) 153 [arXiv:1107.2652] [INSPIRE].

[47] R.D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244 [arXiv:1207.1303] [INSPIRE].

[48] ATLAS collaboration, *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 367 [arXiv:1612.03016] [INSPIRE].

[49] J. Gomez et al., *Measurement of the A-dependence of deep inelastic electron scattering*, Phys. Rev. D **49** (1994) 4348 [INSPIRE].

[50] New Muon collaboration, *A reevaluation of the nuclear structure function ratios for D, He, Li-6, C and Ca*, Nucl. Phys. B **441** (1995) 3 [hep-ph/9503291] [INSPIRE].

[51] New Muon collaboration, *The structure function ratios $F2(li)/F2(D)$ and $F2(C)/F2(D)$ at small x*, Nucl. Phys. B **441** (1995) 12 [hep-ex/9504002] [INSPIRE].

[52] New Muon collaboration, *The A dependence of the nuclear structure function ratios*, Nucl. Phys. B **481** (1996) 3 [INSPIRE].

[53] European Muon collaboration, *Measurement of the ratios of deep inelastic muon — Nucleus cross-sections on various nuclei compared to deuterium*, Phys. Lett. B **202** (1988) 603 [INSPIRE].

[54] European Muon collaboration, *Measurements of the nucleon structure function in the range $0.002\ GeV^2 < x < 0.17\ GeV^2$ and $0.2\ GeV^2 < q^2 < 8\ GeV^2$ in deuterium, carbon and calcium*, Nucl. Phys. B **333** (1990) 1 [INSPIRE].

[55] E665 collaboration, *Shadowing in inelastic scattering of muons on carbon, calcium and lead at low $x(Bj)$*, Z. Phys. C **67** (1995) 403 [hep-ex/9505006] [INSPIRE].

[56] D.M. Alde et al., *Nuclear dependence of dimuon production at 800 GeV. FNAL-772 experiment*, Phys. Rev. Lett. **64** (1990) 2479 [INSPIRE].

[57] European Muon collaboration, *Measurements of the nucleon structure functions $F2_n$ in deep inelastic muon scattering from deuterium and comparison with those from hydrogen and iron*, Nucl. Phys. B **293** (1987) 740 [INSPIRE].

[58] BCDMS collaboration, *Nuclear effects in deep inelastic muon scattering on deuterium and iron targets*, Phys. Lett. B **189** (1987) 483 [INSPIRE].

[59] European Muon collaboration, *A measurement of the ratio of the nucleon structure function in copper and deuterium*, Z. Phys. C **57** (1993) 211 [INSPIRE].

[60] New Muon collaboration, *The Q^2 dependence of the structure function ratio F_2^{Sn}/F_2^C in deep inelastic muon scattering*, Nucl. Phys. B **481** (1996) 23 [INSPIRE].

[61] Fermilab E665 collaboration, *Shadowing in the muon xenon inelastic scattering cross-section at 490 GeV*, Phys. Lett. B **287** (1992) 375 [INSPIRE].

[62] CHORUS collaboration, *Measurement of nucleon structure functions in neutrino scattering*, Phys. Lett. B **632** (2006) 65 [INSPIRE].

JHEP09(2020)183

[63] NuTeV collaboration, *Precise measurement of dimuon production cross-sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe deep inelastic scattering at the Tevatron*, Phys. Rev. D **64** (2001) 112006 [hep-ex/0102049] [INSPIRE].

[64] CMS collaboration, *Observation of nuclear modifications in $W^\pm$ boson production in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV*, Phys. Lett. B **800** (2020) 135048 [arXiv:1905.01486] [INSPIRE].

[65] V. Bertone, S. Carrazza and J. Rojo, *APFEL: a PDF evolution library with QED corrections*, Comput. Phys. Commun. **185** (2014) 1647 [arXiv:1310.1394] [INSPIRE].

[66] S. Forte, E. Laenen, P. Nason and J. Rojo, *Heavy quarks in deep-inelastic scattering*, Nucl. Phys. B **834** (2010) 116 [arXiv:1001.2312] [INSPIRE].

[67] V. Bertone, S. Carrazza and N.P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, Comput. Phys. Commun. **212** (2017) 205 [arXiv:1605.02070] [INSPIRE].

[68] E.L. Berger, J. Gao, C.S. Li, Z.L. Liu and H.X. Zhu, *Charm-quark production in deep-inelastic neutrino scattering at next-to-next-to-leading order in QCD*, Phys. Rev. Lett. **116** (2016) 212002 [arXiv:1601.05430] [INSPIRE].

[69] J. Gao, *Massive charged-current coefficient functions in deep-inelastic scattering at NNLO and impact on strange-quark distributions*, JHEP **02** (2018) 026 [arXiv:1710.04258] [INSPIRE].

[70] R. Boughezal et al., *Color singlet production at NNLO in MCFM*, Eur. Phys. J. C **77** (2017) 7 [arXiv:1605.08011] [INSPIRE].

[71] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, Eur. Phys. J. C **66** (2010) 503 [arXiv:0911.2985] [INSPIRE].

[72] NNPDF collaboration, *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77** (2017) 663 [arXiv:1706.00428] [INSPIRE].

[73] R.D. Ball, E.R. Nocera and J. Rojo, *The asymptotic behaviour of parton distributions at small and large x*, Eur. Phys. J. C **76** (2016) 383 [arXiv:1604.00024] [INSPIRE].

[74] K. Gottfried, *Sum rule for high-energy electron-proton scattering*, Phys. Rev. Lett. **18** (1967) 1174 [INSPIRE].

[75] S. Forte, *The Gottfried sum rule and the light flavor content of the nucleon*, Phys. Rev. D **47** (1993) 1842 [INSPIRE].

[76] R. Abbate and S. Forte, *Re-evaluation of the Gottfried sum using neural networks*, Phys. Rev. D **72** (2005) 117503 [hep-ph/0511231] [INSPIRE].

[77] S.J. Brodsky and G.R. Farrar, *Scaling laws at large transverse momentum*, Phys. Rev. Lett. **31** (1973) 1153 [INSPIRE].

[78] G. Moreno et al., *Dimuon production in proton-copper collisions at $\sqrt{s} = 38.8$ GeV*, Phys. Rev. D **43** (1991) 2815 [INSPIRE].

[79] A. Candido, S. Forte and F. Hekhorn, *Can $\overline{MS}$ parton distributions be negative?*, arXiv:2006.07377 [INSPIRE].

[80] S. Carrazza and J. Cruz-Martinez, *Towards a new generation of parton densities with deep learning models*, Eur. Phys. J. C **79** (2019) 676 [arXiv:1907.05075] [INSPIRE].

JHEP09(2020)183

[81] I. Brivio and M. Trott, *The standard model as an effective field theory*, *Phys. Rept.* **793** (2019) 1 [arXiv:1706.08945] [INSPIRE].

[82] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo and M. Ubiali, *Can new physics hide inside the proton?*, *Phys. Rev. Lett.* **123** (2019) 132001 [arXiv:1905.05215] [INSPIRE].

[83] R.D. Ball et al., *Parton distribution benchmarking with LHC data*, *JHEP* **04** (2013) 125 [arXiv:1211.5142] [INSPIRE].

[84] A. Kusina et al., *Impact of LHC vector boson production in heavy ion collisions on strange PDFs*, arXiv:2007.09100 [INSPIRE].

[85] CCFR collaboration, *Determination of the strange quark content of the nucleon from a next-to-leading order QCD analysis of neutrino charm production*, *Z. Phys. C* **65** (1995) 189 [hep-ex/9406007] [INSPIRE].

[86] CHORUS collaboration, *Leading order analysis of neutrino induced dimuon events in the CHORUS experiment*, *Nucl. Phys. B* **798** (2008) 1 [arXiv:0804.1869] [INSPIRE].

[87] D.A. Mason, *Measurement of the strange-antistrange asymmetry at NLO in QCD from NuTeV dimuon data*, FERMILAB-THESIS-2006-01 (2006).

[88] NuTeV collaboration, *Measurement of the nucleon strange-antistrange asymmetry at next-to-leading order in QCD from NuTeV dimuon data*, *Phys. Rev. Lett.* **99** (2007) 192001 [INSPIRE].

[89] NOMAD collaboration, *A precision measurement of charm dimuon production in neutrino interactions from the NOMAD experiment*, *Nucl. Phys. B* **876** (2013) 339 [arXiv:1308.4750] [INSPIRE].

[90] W.J. Stirling and E. Vryonidou, *Charm production in association with an electroweak gauge boson at the LHC*, *Phys. Rev. Lett.* **109** (2012) 082002 [arXiv:1203.6781] [INSPIRE].

[91] CMS collaboration, *Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2014) 013 [arXiv:1310.1138] [INSPIRE].

[92] CMS collaboration, *Measurement of the associated production of a W boson and a charm quark at $\sqrt{s} = 8$ TeV*, CMS-PAS-SMP-18-013 (2019).

[93] CMS collaboration, *Measurement of associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019) 269 [arXiv:1811.10021] [INSPIRE].

[94] ATLAS collaboration, *Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **05** (2014) 068 [arXiv:1402.6263] [INSPIRE].

[95] HERMES collaboration, *Reevaluation of the parton distribution of strange quarks in the nucleon*, *Phys. Rev. D* **89** (2014) 097101 [arXiv:1312.7028] [INSPIRE].

[96] I. Borsa, R. Sassot and M. Stratmann, *Probing the sea quark content of the proton with one-particle-inclusive processes*, *Phys. Rev. D* **96** (2017) 094020 [arXiv:1708.01630] [INSPIRE].

[97] JAM collaboration, *Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions*, *Phys. Rev. D* **101** (2020) 074020 [arXiv:1905.03788] [INSPIRE].

JHEP09(2020)183

[98] ATLAS collaboration, *Determination of the strange quark density of the proton from ATLAS measurements of the $W \to \ell\nu$ and $Z \to \ell\ell$ cross sections*, Phys. Rev. Lett. **109** (2012) 012001 [arXiv:1203.4051] [INSPIRE].

[99] S.J. Brodsky, I. Schmidt and S. Liuti, *Is the momentum sum rule valid for nuclear structure functions?*, arXiv:1908.06317 [INSPIRE].

[100] D. Boer et al., *Gluons and the quark sea at high energies: distributions, polarization, tomography*, arXiv:1108.1713 [INSPIRE].

[101] LHeC Study Group collaboration, *A large hadron electron collider at CERN: report on the physics and design concepts for machine and detector*, J. Phys. G **39** (2012) 075001 [arXiv:1206.2913] [INSPIRE].

[102] W. Vogelsang and A. Vogt, *Constraints on the proton's gluon distribution from prompt photon production*, Nucl. Phys. B **453** (1995) 334 [hep-ph/9505404] [INSPIRE].

[103] R. Ichou and D. d'Enterria, *Sensitivity of isolated photon production at TeV hadron colliders to the gluon distribution in the proton*, Phys. Rev. D **82** (2010) 014015 [arXiv:1005.4529] [INSPIRE].

[104] D. d'Enterria and J. Rojo, *Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data*, Nucl. Phys. B **860** (2012) 311 [arXiv:1202.1762] [INSPIRE].

[105] J.M. Campbell, J. Rojo, E. Slade and C. Williams, *Direct photon production and PDF fits reloaded*, Eur. Phys. J. C **78** (2018) 470 [arXiv:1802.03021] [INSPIRE].

[106] I. Vitev and B.-W. Zhang, *A systematic study of direct photon production in heavy ion collisions*, Phys. Lett. B **669** (2008) 337 [arXiv:0804.3805] [INSPIRE].

[107] ATLAS collaboration, *Measurement of prompt photon production in $\sqrt{s_{NN}} = 8.16$ TeV p+Pb collisions with ATLAS*, Phys. Lett. B **796** (2019) 230 [arXiv:1903.02209] [INSPIRE].

[108] S. Catani, M. Fontannaz, J.P. Guillet and E. Pilon, *Cross-section of isolated prompt photons in hadron hadron collisions*, JHEP **05** (2002) 028 [hep-ph/0204023] [INSPIRE].

[109] ALICE collaboration, *Measurement of isolated photon-hadron correlations in $\sqrt{s_{NN}} = 5.02$ TeV pp and p-Pb collisions*, arXiv:2005.14637 [INSPIRE].

[110] S. Benic, K. Fukushima, O. Garcia-Montero and R. Venugopalan, *Probing gluon saturation with next-to-leading order photon production at central rapidities in proton-nucleus collisions*, JHEP **01** (2017) 115 [arXiv:1609.09424] [INSPIRE].

[111] ALICE FoCal collaboration, *FoCal: a highly granular digital calorimeter*, Nucl. Instrum. Meth. A **958** (2020) 162059.

[112] C. ALICE Collaboration, *Letter of intent: a Forward Calorimeter (FoCal) in the ALICE experiment*, CERN-LHCC-2020-009 (2020).

[113] M. van Leeuwen, *Constraining nuclear parton density functions with forward photon production at the LHC*, arXiv:1909.05338 [INSPIRE].

[114] NNPDF collaboration, *Reweighting NNPDFs: the W lepton asymmetry*, Nucl. Phys. B **849** (2011) 112 [*Erratum ibid.* **854** (2012) 926] [*Erratum ibid.* **855** (2012) 927] [arXiv:1012.0836] [INSPIRE].

[115] R.D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti et al., *Reweighting*

JHEP09(2020)183

and Unweighting of Parton Distributions and the LHC W lepton asymmetry data, Nucl. Phys. B **855** (2012) 608 [arXiv:1108.1758] [INSPIRE].

[116] A. Accardi et al., *Electron ion collider: the next QCD frontier*, Eur. Phys. J. A **52** (2016) 268 [arXiv:1212.1701] [INSPIRE].

[117] I. Helenius, K.J. Eskola and H. Paukkunen, *Probing the small-x nuclear gluon distributions with isolated photons at forward rapidities in p+Pb collisions at the LHC*, JHEP **09** (2014) 138 [arXiv:1406.1689] [INSPIRE].

[118] R.D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo and L. Rottoli, *Parton distributions with small-x resummation: evidence for BFKL dynamics in HERA data*, Eur. Phys. J. C **78** (2018) 321 [arXiv:1710.05935] [INSPIRE].

[119] D. de Florian, R. Sassot, M. Epele, R.J. Hernández-Pinto and M. Stratmann, *Parton-to-pion fragmentation reloaded*, Phys. Rev. D **91** (2015) 014035 [arXiv:1410.6027] [INSPIRE].

[120] D. d'Enterria, K.J. Eskola, I. Helenius and H. Paukkunen, *Confronting current NLO parton fragmentation functions with inclusive charged-particle spectra at hadron colliders*, Nucl. Phys. B **883** (2014) 615 [arXiv:1311.1415] [INSPIRE].

[121] NNPDF collaboration, *Charged hadron fragmentation functions from collider data*, Eur. Phys. J. C **78** (2018) 651 [arXiv:1807.03310] [INSPIRE].

[122] NNPDF collaboration, *A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties*, Eur. Phys. J. C **77** (2017) 516 [arXiv:1706.07049] [INSPIRE].

[123] S. Albino, B.A. Kniehl and G. Kramer, *Fragmentation functions for light charged hadrons with complete quark flavor separation*, Nucl. Phys. B **725** (2005) 181 [hep-ph/0502188] [INSPIRE].

[124] M. Hirai, S. Kumano, T.-H. Nagai and K. Sudoh, *Determination of fragmentation functions and their uncertainties*, Phys. Rev. D **75** (2007) 094009 [hep-ph/0702250] [INSPIRE].

[125] J.L. Albacete et al., *Predictions for cold nuclear matter effects in p+Pb collisions at* $\sqrt{s_{NN}} = 8.16$ *TeV*, Nucl. Phys. A **972** (2018) 18 [arXiv:1707.09973] [INSPIRE].

[126] A. Garcia, R. Gauld, A. Heijboer and J. Rojo, *Complete predictions for high-energy neutrino propagation in matter*, JCAP **09** (2020) 025 [arXiv:2004.04756] [INSPIRE].

[127] V. Bertone, R. Gauld and J. Rojo, *Neutrino telescopes as QCD microscopes*, JHEP **01** (2019) 217 [arXiv:1808.02034] [INSPIRE].

[128] R. Abdul Khalek et al., *Phenomenology of NNLO jet production at the LHC and its impact on parton distributions*, Eur. Phys. J. C **80** (2020) 797 [arXiv:2005.11327] [INSPIRE].

[129] CMS collaboration, *Constraining gluon distributions in nuclei using dijets in proton-proton and proton-lead collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Phys. Rev. Lett. **121** (2018) 062002 [arXiv:1805.04736] [INSPIRE].

[130] PROSA collaboration, *Impact of heavy-flavour production cross sections measured by the LHCb experiment on parton distribution functions at low x*, Eur. Phys. J. C **75** (2015) 396 [arXiv:1503.04581] [INSPIRE].

[131] R. Gauld and J. Rojo, *Precision determination of the small-x gluon from charm production at LHCb*, Phys. Rev. Lett. **118** (2017) 072001 [arXiv:1610.09373] [INSPIRE].

JHEP09(2020)183

[132] K.J. Eskola, I. Helenius, P. Paakkinen and H. Paukkunen, *A QCD analysis of LHCb D-meson data in p+Pb collisions*, *JHEP* **05** (2020) 037 [arXiv:1906.02512] [INSPIRE].

[133] J.J. Ethier, N. Sato and W. Melnitchouk, *First simultaneous extraction of spin-dependent parton distributions and fragmentation functions from a global QCD analysis*, *Phys. Rev. Lett.* **119** (2017) 132001 [arXiv:1705.05889] [INSPIRE].

[134] A. Buckleys et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132 [arXiv:1412.7420] [INSPIRE].

JHEP09(2020)183