Extending nuclear PDF analyses into the high-x, low-Q^2 region

E. P. Segarra,[1,*] T. Ježo,[2,†] A. Accardi,[3,4] P. Duwentäster,[5] O. Hen,[1] T. J. Hobbs,[6,4,7] C. Keppel,[4] M. Klasen,[5] K. Kovařík,[5] A. Kusina,[8] J. G. Morfín,[9] K. F. Muzakka,[5] F. I. Olness,[6,‡] I. Schienbein,[10] and J. Y. Yu.[10]

[1]*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Institute for Theoretical Physics, KIT, D-76131 Karlsruhe, Germany*
[3]*Hampton University, Hampton, Virginia 23668, USA*
[4]*Jefferson Lab, Newport News, Virginia 23606, USA*
[5]*Institut für Theoretische Physik, Westfälische Wilhelms-Universität Münster, Wilhelm-Klemm-Straße 9, D-48149 Münster, Germany*
[6]*Department of Physics, Southern Methodist University, Dallas, Texas 75275-0175, USA*
[7]*Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616, USA*
[8]*Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 Krakow, Poland*
[9]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[10]*Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, 53 avenue des Martyrs, 38026 Grenoble, France*

(Received 6 January 2021; accepted 27 April 2021; published 15 June 2021)

We use the nCTEQ analysis framework to investigate nuclear parton distribution functions (nPDFs) in the region of large x and intermediate-to-low Q, with special attention to recent JLab deep inelastic scattering data on nuclear targets. This data lies in a region which is often excluded by W and Q cuts in global nPDF analyses. As we relax these cuts, we enter a new kinematic region, which introduces new phenomenology. In particular, we study the impact of (i) target mass corrections, (ii) higher twist corrections, (iii) deuteron corrections, and (iv) the shape of the nuclear PDF parametrization at large-x close to one. Using the above tools, we produce a new nPDF set (named nCTEQ15HIX) which yields a good description of the new JLab data in this challenging kinematic region, and displays reduced uncertainties at large x, in particular for up and down quark flavors.

DOI: 10.1103/PhysRevD.103.114015

I. INTRODUCTION

With the EIC and LHeC/FCC on the horizon, science is now entering a new era of precision in the investigation of hadronic structure enabled by a flood of data from JLab, RHIC and the LHC. Describing one of the four fundamental forces of nature, quantum chromodynamics (QCD)—the theory of the strong interaction—remains deeply complex and enigmatic, although the parton distribution function (PDF) framework has proven remarkably successful in describing processes with hadronic initial states [1–26].

While the study of proton PDFs has grown exceedingly precise, the need to extend this precision to the nuclear sector, involving fits with explicit nuclear degrees of freedom, has become more urgent in recent years in order to enhance the accuracy of experimental analyses involving nuclear targets. Progress in studying QCD dynamics within nuclei has been demonstrated across a number of recent nuclear PDF (nPDF) analyses [1–11]. A significant challenge in the determination of nPDFs has been the acquisition of empirical data from a sufficiently wide variety of experiments as to provide complementary constraints, and, e.g., specify the A dependence of the resulting nPDFs. For this reason, there is a continual need for new datasets to broaden global analyses. In the present work, we build upon the recent nCTEQ15 analysis by including recent JLab data covering an expanded kinematic range. As we shall demonstrate, this data has the potential to furnish an improved understanding of hadronic and nuclear structure and interactions, and, in turn, new insights into QCD.

A. JLab kinematic reach

The recent facility upgrades of the Continuous Electron Beam Accelerator Facility at the Thomas Jefferson National Accelerator Facility (JLab) have enabled

*segarrae@mit.edu
†tomas.jezo@kit.edu
‡olness@smu.edu

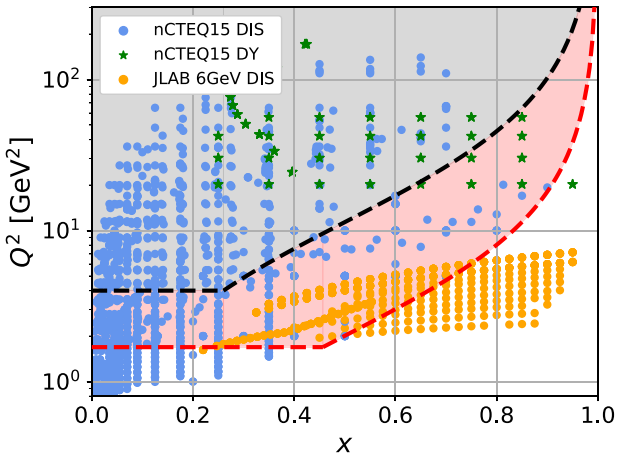

FIG. 1. We display DIS and DY data entering our analysis in the $\{x, Q^2\}$ space indicating the relevant kinematic cuts, where x and Q^2 are the usual DIS variables, and Q^2 for DY is the dilepton mass squared. The more restrictive cuts of $Q = 2$ GeV and $W = 3.5$ GeV (black dashed line) are the cuts used in the original nCTEQ15 analysis. In the present work, we will relax the cuts to $Q = 1.3$ GeV and $W = 1.7$ GeV (red dashed line). This greatly expands the kinematic reach in the high-x region where much of the new JLab data is located.

the measurement of high precision electron-nucleus scattering events in an extended kinematic regime. In particular, the JLab experiments provide a wealth of data in the relatively unexplored kinematic region of large Bjorken x and intermediate to low photon virtuality Q^2. This mostly unexplored kinematic region is often referred to as the "transition" region from resonance dominated production to deep-inelastic scattering, and is of considerable interest to both the charged lepton and neutrino scattering communities.

In Fig. 1 we display a collection of DIS and DY data from selected experiments along with recent data from the JLab DIS experiments. We have also indicated typical cuts in Q^2 and $W^2 = Q^2(1-x)/x + M_N^2$, with M_N the nucleon mass, which are often implemented in many global analyses. The objective of these $\{W, Q\}$ cuts is to remove those data which might have significant nonperturbative or higher twist contributions; unfortunately, these cuts exclude a large subset of the data, as indicated in Table I.

In this investigation we will relax the kinematic cuts to study whether we can describe the broader subset of the JLab data with nuclear PDFs using the nCTEQ global analysis framework [1]. As we expand this analysis into new kinematic regions, we will analyze whether we might be sensitive to new effects including (i) target mass corrections, (ii) higher twist corrections, (iii) deuteron structure function modifications, and (iv) large-x nuclear corrections to allow the nuclear Bjorken variable $x_A > 1$. We will address each of these issues in turn, and then use the optimal combination to extract a set of nPDFs which extends into this new kinematic region.

TABLE I. The table shows the number of remaining data points after the $\{Q^2, W\}$ kinematic cuts, where x and Q^2 are the usual DIS variables, and Q^2 for DY is the dilepton mass squared. The units of Q and W are both in GeV, and Q^2 in GeV^2. For reference, nCTEQ15 used cuts of $Q = 2$ GeV and $W = 3.5$ GeV, while the current nCTEQ15HIX set uses cuts of $Q = 1.3$ GeV and $W = 1.7$ GeV.

		$W_{\rm cut}$	$W_{\rm cut}$	$W_{\rm cut}$	$W_{\rm cut}$	$W_{\rm cut}$
$Q^2_{\rm cut}$	$Q_{\rm cut}$	No Cut	1.3	1.7	2.2	3.5
1.3	$\sqrt{1.3}$	1906	1839	1697	1430	1109
1.69	1.3	1773	1706	1564	1307	1024
2	$\sqrt{2}$	1606	1539	1402	1161	943
4	2	1088	1042	952	817	708

We note here a recent study [9] where PDF reweighting methods were used to investigate compatibility of the current nuclear PDFs with the CLAS data.

B. The EMC effect

Extending the nPDFs into this new kinematic region can also provide new information on x-dependent nuclear effects observed in many DIS experiments. Starting in the early 1980s, the European Muon Collaboration (EMC) [27] found that in the DIS kinematic region the per-nucleon structure functions F_2 for iron and deuterium were not only different, but that this difference also changed as a function of x.

This evidence for nuclear effects in DIS charged-lepton nucleus scattering can be summarized as in Fig. 2, which

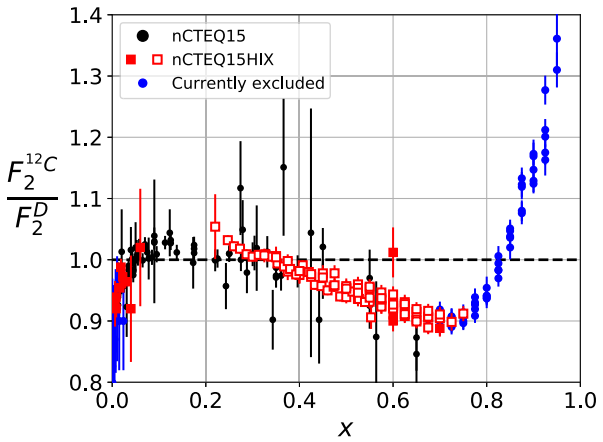

FIG. 2. We display the classic F_2^A/F_2^D ratio for carbon illustrating the nuclear correction factor across the various x regions. The black points indicate the data used in the original nCTEQ15 fit, and the red points with the solid squares represent the additional data from this original set which are now included due to the relaxed Q and W cuts. The red open squares are the new JLab DIS data included in this analysis, and the blue points are those JLab DIS data which are excluded by the current kinematic cuts.

displays the F_2^C/F_2^D per-nucleon structure function ratio as a function of x; this behavior is consistent with measurements by both the SLAC eA [28–30] and the BCDMS μA [31,32] experiments. The observed behavior of the ratio $R = F_2^A/F_2^D$, can be divided into four x regions:

(i) the shadowing region: $R \leq 1$ for $x \lesssim 0.1$,
(ii) the antishadowing region: $R \geq 1$ for $0.1 \lesssim x \lesssim 0.3$,
(iii) the EMC-effect region: $R \leq 1$ for $0.3 \lesssim x \lesssim 0.7$,
(iv) and the Fermi motion region: for $x \gtrsim 0.7$.

For a review of the data and theoretical interpretations see Ref. [33].

The current study concentrates on the higher x region so the effects of shadowing and antishadowing are not of direct impact for this specific study. However, the modifications at medium-to-higher-x in the so-called "EMC-effect region" will be especially important for these new JLab datasets. Since its discovery, there is still no universally accepted explanation for the EMC effect. Recent studies have observed a correlation between the magnitude of the EMC effect and the relative amount of short-range correlated (SRC) nucleon pairs in different nuclei, suggesting that the EMC effect is driven by the modification of nucleons in SRC pairs [34–38]. Forthcoming experiments at JLab will explore the relationship between the off-shellness of nucleons and the modification of the structure function [39,40]. Additionally, it has been demonstrated [41] that the EMC effect persists to lower values of W extending into the resonance region. This is an intriguing result that the nuclear structure functions in the resonance region exhibit similar behavior as in the DIS region. In Sec. V we will explore implications of removing the W cut on the EMC effect.

C. Outline

In Sec. II we present an overview of the various corrections we will apply to better describe the large x region. In Sec. III we introduce the new JLab data used in this study and summarize the various fits we will investigate. In Sec. IV we show the results of the fits including the comparison between data and theory, and present the impact on the PDFs. In Sec. V we explore the implications of extending the nPDFs into the low W kinematic region. In Sec. VI we provide a comparison of the new nCTEQ15HIX nPDFs with other results from the literature. In Sec. VII we summarize our conclusions. Additionally, in the Appendix we tabulate the datasets used in the fit and provide the impact of the kinematic cuts.

II. STRUCTURE FUNCTION MODIFICATIONS

The goal of this study is to extend our global nPDF fit into new kinematic regions at large x and low Q. Many global analyses impose stringent cuts in both Q and W. For example, the nCTEQ15 fit was performed with $Q > 2$ GeV and $W > 3.5$ GeV. The motivation for these cuts is that the Q cut largely eliminates a variety of nonperturbative and/or power-suppressed corrections with the potential to complicate the extraction of leading-twist PDFs. As Q decreases, $\alpha_s(Q^2)$ becomes large, and our perturbation expansion breaks down. Correspondingly, the W cut removes events in both the low Q^2 and large x region where contributions from non-factorizable higher-twist terms become large.

Imposing stringent cuts allows us to avoid kinematic regions that may be difficult to compute. However, the trade-off is that this can significantly reduce the dataset excluding kinematic regions important for both the charged lepton and neutrino communities. In Table I we display the total number of data points which satisfy various kinematic cuts. For example, with the loosest set of cuts $\{Q > 1.14\ \text{GeV}, W > 1.3\ \text{GeV}\}$, the number of data points is 1839; this is in contrast to the very conservative cuts used in the nCTEQ15 fit, $\{Q > 2\ \text{GeV}, W > 3.5\ \text{GeV}\}$, which reduces the number to 708—less than half.

Clearly, it is advantageous to reduce the $\{Q, W\}$ cuts as much as possible, but this region is challenging to compute. For example, if we simply take the nCTEQ15 fit and extend this into the low $\{Q, W\}$ region without accommodating any additional phenomenological effects, the quality of the resulting theory to data comparison is no longer acceptable.

As we mentioned in the introduction, there are a variety of possible effects that may enter our analysis in this region, and we will examine them systematically to study the impact of each one to obtain the best description of the data.

In the remainder of this section, we will review the different effects, and outline how they are included in the global fit.

A. Target mass corrections

We begin by examining the operator product expansion (OPE) to determine what possible corrections we might encounter as we extend the kinematic region to high-x and low Q. The expressions for the structure functions can be derived in the OPE, and a detailed review is provided in Ref. [42]. Using this formalism, the leading, twist-2, structure function expressions will have corrections in powers of (M/Q) where M is the mass of the proton[1]:

$$F_2^{\text{TMC}}(x,Q) = \frac{x^2}{\xi^2 r^3} F_2^{(0)}(\xi,Q) + \ldots \tag{1}$$

Here, $\xi = 2x/(1+r)$ is the Nachtmann variable with $r = \sqrt{1+4x^2M^2/Q^2}$, and $F_2^{(0)}(x,Q)$ is the structure function in the limit where the proton mass M is set to zero. The additional terms, indicated by the dots, are given by convolution integrals for which explicit expressions are given in Eqs. (22)–(24) of Ref. [42]. The magnitude of

[1]See Eq. (23) of Ref. [42] for details. We have implemented an alternative definition derived in the light-cone frame [43,44] which coincides with Eq. (1) to $\lesssim 1\%$.

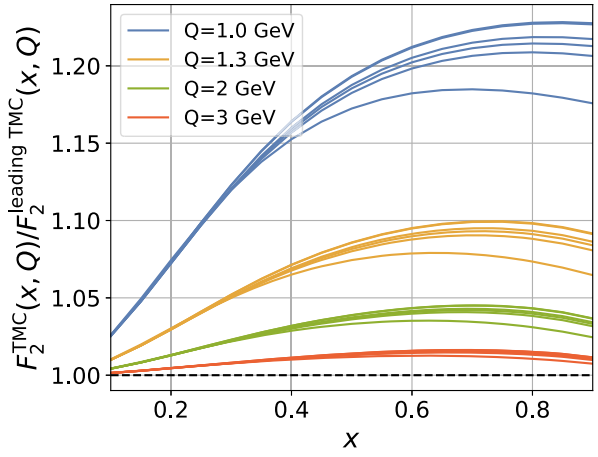

FIG. 3. We display the ratio $F_2^{\rm TMC}(x,Q)/F_2^{\rm leading\ TMC}(x,Q)$ for $Q = \{1, 1.3, 2, 3\}$ GeV (blue, yellow, green, red) for neutral current DIS; the lowest Q values yield the largest corrections. Each colored band contains results for {D, C, Al, Fe, Au, Pb} nuclei, where D is the lowest and Pb is the highest within each band.

these corrections can be sizable, especially for lower scales, $Q \sim M$, as we illustrate in Fig. 3 where we show the x dependence of the ratio $F_2^{\rm TMC}(x,Q)/F_2^{\rm leading\ TMC}(x,Q)$ where $F_2^{\rm leading\ TMC}(x,Q) = x^2/(\xi^2 r^3)F_2^{(0)}(\xi,Q)$. As can be seen, for F_2, the leading TMC gets modified by the additional TMC by up to a $\sim$25% at intermediate to large x values.

The leading TMC effects due to the Nachtmann variable ξ and the prefactor $x^2/(\xi^2 r^3)$ are accounted for in all our calculations. However, we have not included the effect of the convolution terms. Since we are fitting only nuclear ratios such as F_2^A/F_2^D and σ_A/σ_D, the nuclear A dependence of these additive convolution terms is very minor.[2]

One important question which arises is: do the TMCs scale as the mass of the nucleus M_A/Q, or the mass of the nucleon M/Q where $M_A = AM$? This issue is especially important for the heavy nuclear targets such as lead ($A = 208$) where a M_A/Q scaling would dramatically impact the low energy data. The answer can be obtained by tracing the mass terms in the derivation of the master formula (e.g., Eqs. (22–24) of Ref. [42]) from the OPE. For a nucleon, the OPE expansion is in terms of Mx/Q while for a full nucleus this becomes $M_A x_A/Q$. Here, x_A is the Bjorken x for the nucleus defined as $x_A = Q^2/(2P_A \cdot Q)$, where P_A is the full nucleus momentum and $x_A \in [0,1]$. Thus, we find $M_A x_A/Q = (AM)(x/A)/Q \equiv Mx/Q$, (where $x = Ax_A$ and $x \in [0, A]$) so even for heavy nuclei the TMC terms are suppressed by powers of M/Q where $M \sim 1$ GeV. Consequently, the master formula of Ref. [42] holds for both nucleons and nuclei. There are certainly some subtle steps in extending the OPE and DGLAP formalism from the case of a proton to a heavy nuclear target, and a separate paper expanding the work of Ref. [42] with detailed derivations is in progress [45].

Figure 3 displays a band of curves for each fixed Q value which show the variation between D (the lowest curve) and Pb (the highest curve). The spread of these bands can be as large as 5% for $Q = 1.0$ GeV, and decreases quickly for larger Q values. For the current investigation, we will focus on cuts of $Q > 1.3$ GeV and $W > 1.7$ GeV. For example, from Fig. 1 we see that for $Q = 1.3$ GeV, the $W > 1.7$ GeV cut includes points out to $x \lesssim 0.46$, and here the spread of this band (yellow) is on the order of $\sim$1%. This represents the maximum deviation between a deuteron (D) target and a lead (Pb) target; lighter nuclei such as carbon (C) would be less. Therefore, we see that within our kinematic cuts, the TMC correction provides a uniform shift, within about 1% or less, of all the nuclei.

Since these additional TMC corrections shift both the A and D results by nearly the same factor, the F_2^A/F_2^D ratio will not be affected by the non-leading TMCs:

$$\frac{F_2^{A,\rm TMC}(x,Q)}{F_2^{D,\rm TMC}(x,Q)} \simeq \frac{F_2^{A,\,\rm leading\ TMC},(x,Q)}{F_2^{D,\,\rm leading\ TMC},(x,Q)} = \frac{F_2^{A,(0)}(\xi,Q)}{F_2^{D,(0)}(\xi,Q)}. \quad (2)$$

Therefore, for our present study these leading-twist TMCs do not impact the results. However, if we were to examine the absolute structure functions (instead of ratios) or the absolute cross-sections, these TMCs must be incorporated in the high x, low Q region.

B. Higher twist corrections

At high x and low Q values, the subleading $1/Q^2$ higher twist (HT) and residual power corrections can be important. To explore these effects, we will use a phenomenological x-dependent function taken from the CJ15 study [46] with the following form:

$$F_2^A(x,Q) \to F_2^A(x,Q)\left[1 + \frac{C_{\rm HT}(x,A)}{Q^2}\right], \quad (3)$$

where the higher-twist coefficient C_{HT} depends on x and the nuclear A:

$$C_{\rm HT}(x,A) = h_0 x^{h_1}(1 + h_2 x)A^\tau, \quad (4)$$

with $\{h_0, h_1, h_2\} = \{-3.3\ {\rm GeV}^2, 1.9, -2.1\}$; h_0 carries units of GeV^2, and the $h_{1,2}$ are dimensionless.

Note that the A^τ scaling was not part of the original CJ15 formulation and was included here to explore a possible nuclear dependence [47]. We have varied the τ exponent in our fits and the optimal values is indistinguishable from zero ($\tau \sim 0 \pm 0.005$); hence, we find no significant A dependence for this correction. As we are primarily focused

[2]The NMC F_2^D dataset ID = 5160 is the one case which is not a ratio. This set extends out to $x \lesssim 0.48$, and we have checked that these TMC corrections are less than the data uncertainty.

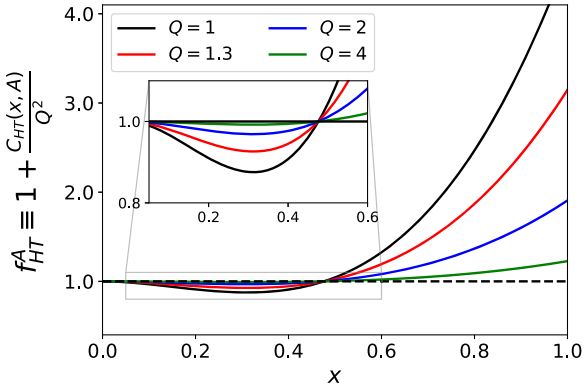

FIG. 4. We display the higher twist corrections as a function of x for $Q = \{1.0, 1.3, 2, 4\}$ GeV. The lower Q values yield larger corrections since they scale as $1/Q^2$.

on lighter nuclei {He, Be, C} in this study, it may be useful to revisit the A^τ dependence for the case of heavy nuclei.

Also note that since the h_0 term in Eq. (4) will generally be scaled by the target mass M, the separation between the higher-twist and the target mass corrections (TMCs) discussed above is not unambiguous.

In Fig. 4 we display the higher twist correction as a function of x for a selection of Q values. The HT correction gives a slight reduction in the intermediate x region ($x \sim 0.3$), and then becomes large and positive at large x, depending on the Q value. For example, at $x = 0.7$, the correction factor is ~1.5 for the $Q = 1.3$ GeV, whereas at $Q = 2$ GeV this is already reduced to ~1.2. In Fig. 4 we also provide an inset plot to show the details in the intermediate x region.[3]

C. Deuteron nuclear structure

A large fraction of the nuclear DIS data incorporated in global nPDF fits, including the present one, are expressed as a F_2^A/F_2^D ratio of structure functions on a heavy nuclear target and on a deuteron target, see Tables III–VI in the Appendix. A careful treatment of the denominator is therefore called for in order to analyze the effects of the nuclear dynamics on the nuclear quark and gluons, as most recently highlighted in Ref. [53].

The deuteron is the lightest and least bound of all compound nuclei. Therefore it is typically considered as composed of a free proton and a free neutron, and the deuteron structure function is computed as the sum of the

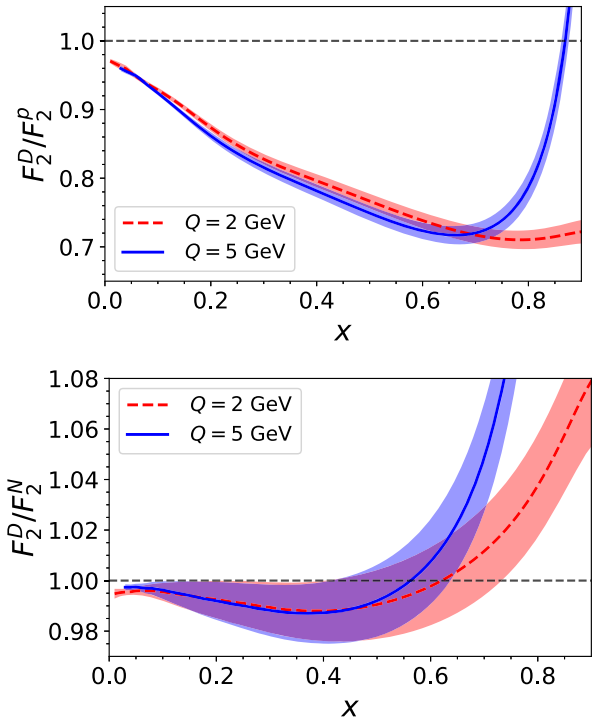

FIG. 5. We display the CJ15 calculation (Ref. [46]) of the structure function ratio of the deuteron F_2^D to the free proton F_2^p as a function of x for selected values of Q. Note that these curves are effectively fits to the experimental data. Additionally we display the ratio of F_2^D to the isoscalar combination $F_2^N = F_2^p + F_2^n$ to demonstrate the nuclear dynamics in the deuteron beyond isospin symmetry effects. The determination of the neutron structure function in the denominator depends on the details of the nuclear correction model fitted in Ref. [46]. The colored bands are 90% confidence level uncertainty arising from the global CJ15 fit parameters.

structure functions of its components. Extensive studies, however, indicate that the deuteron structure is far more complex than a simple combination of proton and neutron PDFs might suggest [33,46,54–58]. Yet, the deuteron differs from the heavier targets considered in this paper because its binding energy is far smaller, making this a loosely bound system of two nucleons rather than a tightly bound and noticeably denser system of many nucleons interacting with each other [59]. Thus its nPDFs may not follow the A-scaling parametrization utilized in this work.

For these reasons—and to avoid double counting with the proton PDF fits that we use to anchor the nuclear PDFs parametrization used in this paper which already include deuteron target data—we choose not to fit the deuteron nPDF. Instead, we build upon the comprehensive studies available in the literature, and calculate the deuteron structure function, using the PDFs and nuclear dynamics simultaneously extracted in the recent CJ15 global QCD analysis [46]. At large x, this nuclear dynamics calculation is based on two main components: (i) a baryonic smearing function evaluated on the basis of the AV18 nuclear wave

[3]Note for the neutrino community: existing early experimental results [48–50] and theoretical considerations [51,52] of higher twist (HT) neutrino-nucleon scattering indicate that the HT contribution does not become positive at higher-x as in Fig. 4, but remains negative in the high x, lower Q region. This neutrino HT behavior needs further exploration, however it suggests consideration of this possibility in applying the results of this study to neutrino scattering in the relevant $x - Q$ region.

function [60,61] that accounts for Fermi motion and nuclear binding effects, and (ii) a fitted parametric function quantifying the effect of the off-shellness of the bound nucleon. At smaller x, shadowing effects are calculated according to Ref. [62].

In Fig. 5 we display the ratio of F_2^D to the proton's F_2^p, which we use in the present nPDF analysis. For comparison, we also display the ratio to the isoscalar combination, $F_2^N = F_2^p + F_2^n$, which illustrates the size and x dependence of the nuclear effects in the deuteron system, after removing the underlying isospin symmetry effects. We can see that the deuteron-to-proton ratio, F_2^D/F_2^p, dips to ~70% at $x \sim 0.7$ before dramatically increasing at larger x. Comparing this with the ratio to the isoscalar structure function, F_2^D/F_2^N, we see much of the previous effect was in fact due to the differing charge factors weighting the quark PDFs in the isoscalar combination, but the residual dynamical nuclear effects are not negligible, particularly at large-x values.

Note that the correct treatment of the deuteron structure function will impact, in principle, a large fraction of the DIS data, see Tables III–V. The nuclear dynamics in the deuteron may reasonably be neglected in typical nPDF fits because, after the usual kinematic cut on $W \gtrsim 3.5$ GeV, most of the DIS data lies in a small-to-medium x region where the corrections are small. However, if we want to extend our fit into the large x region, Fig. 5 demonstrates that it is essential to go beyond pure isospin symmetry considerations.

Operationally, we incorporate the deuteron's nuclear dynamics by rescaling the F_2^A/F_2^D ratio data by a F_2^D/F_2^p factor calculated using the CJ15 PDFs and deuteron correction model [Fig. 5(a)]. Thus, the transformation is:

$$\frac{F_2^A}{F_2^p} \equiv \frac{F_2^A}{F_2^D} \cdot \left(\frac{F_2^D}{F_2^p}\right)_{\mathrm{CJ}}, \tag{5}$$

and this effectively produces a nucleus-to-proton ratio dataset that is then passed to the fit program. We provide the data and associated corrections in our supplementary material.

Note that the $(F_2^D/F_2^p)_{\mathrm{CJ}}$ rescaling factor is computed using the structure function results of the CJ15 global analysis [46,59]. CJ15 includes data from a variety of processes including DIS, Drell-Yan and jets from hadron-hadron collisions at the Tevatron; this combination covers a broad kinematic range which contributes to improved precision. Additionally CJ15 includes data from the BONuS experiment [63,64] which measured the neutron to deuteron F_2 structure function ratio using an innovative spectator tagging technique. The BONuS data provides unique information on the neutron structure function, and this allows substantive improvement of the up and down quark flavor separation in the CJ15 analysis.

In principle we could also use CJ15 to correct to an isoscalar quantity (e.g., F_2^D/F_2^N) as illustrated in Fig. 5-b). The difference between this approach and Eq. (5) depend on the details of the CJ15 PDFs, including the specific F_2 neutron structure function extracted. For this reason, we have chosen to implement Eq. (5); but, we have verified that this choice does not substantively impact our conclusions for the valence distributions, which is the focus of this study.

As in the DIS case, the $p + D$ Drell-Yan cross section is modified by the nuclear dynamics in the deuteron target compared to calculations that treat the deuteron as composed of a free proton and a free neutron. The nuclear smearing model discussed above can also be applied to this case [65]. However, we have verified that the ensuing deuteron corrections amount to less than 1% in the kinematics covered by the available DY data, and can be neglected compared to the experimental uncertainties.

III. FITS TO JLAB DIS DATASETS

Now that we have surveyed the various phenomena which enter the fit in the large x and low Q region, we will implement these effects into our fit so we can compare the relative effects.

In this section, we will first review the experimental data that enters the fit, and then provide an overview of the exploratory fits we will use to investigate the effects discussed in the previous section.

A. The experimental data

Historically many experiments have included isoscalar corrections (accounting a for different number of protons and neutrons in nuclei) in their data, e.g., structure function ratio data [28,66]. Since these corrections are model dependent, for the purpose of the current analysis, we have removed all isoscalar corrections from the used datasets following the specific prescriptions used by the collaborations. We provide supplemental materials with the tabulated data used in our analysis along with these corrections. This is available in a separate text file.

1. Data used in nCTEQ15

For our fits, we will use the full set of DIS and DY data from the nCTEQ15 analysis [1] with isoscalar corrections removed. These data are listed in Tables IV–VI in Appendix.

Additionally, the nCTEQ15 analysis includes RHIC inclusive pion data from PHENIX [67] and STAR [68] collaborations for nuclear modification R^{π}_{DAu}. The pion data helps to constrain the nuclear gluon PDF, especially in the lower x regions. As our focus will be on the high-x region, we have not included the pion data in the current fits which simplifies our analysis as we do not need to consider fragmentation functions. However, as a cross check, we separately compute the χ^2 values for the pion data to ensure our results are compatible with these datasets. These comparisons are reported in Sec. IV C 4 and show very good agreement with these data.

2. The JLab data

In this analysis, we include new DIS data from the Jefferson Lab CLAS [38] and Hall C [69] experiments taken during the 6 GeV electron beam operation [70].

These datasets, in addition to spanning a wide range of A, provide high-precision constraints for nuclear PDFs at high-x and low-Q^2. Nuclei included are ^{3}He, ^{4}He, ^{9}Be, ^{12}C, ^{27}Al, ^{56}Fe, ^{64}Cu, ^{197}Au, ^{208}Pb, spanning x from 0.2 to 0.95. Full details of these datasets are summarized in Appendix as well as in Refs. [38,69].

B. The parameters

As before, our nCTEQ PDFs are parametrized as [1]:

$$xf_i^{p/A}(x,Q_0) = c_0 x^{c_1}(1-x)^{c_2} e^{c_3 x}(1+e^{c_4}x)^{c_5}, \quad (6)$$

where the nuclear A dependence is encoded in the c_k coefficients as[4]:

$$c_k \to c_k(A) \equiv p_k + a_k(1-A^{-b_k}), \quad (7)$$

where $k = \{1,\ldots,5\}$. This parametrization has the advantage that in the limit $A \to 1$, $c_k(A)$ reduces to proton value p_k. The a_k parameters control the magnitude of the modification to the p_k proton "boundary condition," and the b_k parameters control the power of the nuclear-A term. Specifically, within the nCTEQ15 framework we parametrize the combinations $\{u_v, d_v, (\bar{u}+\bar{d}), (\bar{d}/\bar{u}), s, g\}$, and we impose the boundary condition[5] of $s = \bar{s} = \kappa(\bar{u}+\bar{d})/2$ with $\kappa = 0.5$ at $A = 1$.

We will then perform fits using a total of 19 free parameters. This includes 5 parameters for each the up and down valence $\{a_k^{u_v}, a_k^{d_v}\}$ where $k = \{1\ldots5\}$, two parameters for the $\bar{u}+\bar{d}$ combination $\{a_1^{\bar{d}+\bar{u}}, a_2^{\bar{d}+\bar{u}}\}$, and seven parameters for the gluon $\{a_1^g, a_4^g, a_5^g, b_0^g, b_1^g, b_4^g, b_5^g\}$. This set of 19 parameters includes the original 16 parameters used in the nCTEQ15 PDF fit, and adds the 3 parameters $\{a_3^{u_v}, a_3^{d_v}, a_4^{d_v}\}$ so that we have all the a_k parameters open for the up and down valence to fully explore the parameter space. While our focus is on the high-x region where the up- and down-quark PDFs dominate, there is some interplay between the sea quarks and the gluon PDFs; hence, this will influence the g and $\bar{d}+\bar{u}$ parameters.

For our analysis we use our new nCTEQ ++ framework which has a modular structure and links to a variety of external tools including a modified version of HOPPET [72] (extended to accommodate grids of multiple nuclei), APPLgrid [73], and MCFM [74]. Additionally, we have used FEWZ [75] for benchmarking our WZ calculations, and xFitter [76] for various cross checks.

[4]Note that the original nCTEQ15 parametrization of Ref. [1] used the notation $\{c_{k,0}, c_{k,1}, c_{k,2}\}$ in place of the current $\{p_k, a_k, b_k\}$ notation of Eq. (7).

[5]Note, that the recent nCTEQ15WZ [71] PDF set allows additional freedom for the strange PDF as compared to the original nCTEQ15 set [1].

TABLE II. We summarize the fits explored in this study. Note the nCTEQ15 and nCTEQ15* results are not re-fits; the first entry, nCTEQ15, lists the fit results from the original analysis, and the second entry, nCTEQ15*, simply computes the χ^2 for the new JLab data with the relaxed $\{Q, W\}$ cuts. The remaining four entries, {BASE, HT, DEUT, nCTEQ15HIX}, are re-fit with the new data, and are detailed in the text.

Fit	χ^2	N_{data}	χ^2/N_{dof}	Q_{cut}	W_{cut}
nCTEQ15	587	740	0.81	2.0	3.5
nCTEQ15*	2664	1564	1.70	1.3	1.7
BASE	1525	1564	0.99	1.3	1.7
HT	1489	1564	0.96	1.3	1.7
DEUT	1331	1564	0.86	1.3	1.7
nCTEQ15HIX	1297	1564	0.84	1.3	1.7

C. The fits

Having specified the required inputs, we now present a brief overview of the fits we will investigate. These are summarized in Table II and are outlined in the text below.

- *nCTEQ15**: This is the original set of nuclear PDFs as computed in Ref. [1]. The new data has not been fit; we have simply computed the χ^2 values.
- *BASE*: This BASE fit will serve as our primary reference fit as it simply extends the nCTEQ15 fit by reducing the kinematic cuts to $Q > 1.3$ GeV and $W > 1.7$ GeV. In particular, this fit does not include any higher-twist corrections, or deuteron modifications.
- *HT*: This fit extends the BASE fit by adding the higher twist (HT) power corrections as described in Eq. (3).
- *DEUT*: This fit extends the BASE fit by including the deuteron modifications as described in Sec. II C.
- *nCTEQ15HIX*: This fit extends the BASE fit by including both the deuteron correction (DEUT) and also the higher twist (HT) corrections.

Note, that all of the above fits include the TMCs originating from the scaling variable ξ and from the prefactor, as explained in the TMC discussion, Sec. II A.

Table II also lists the $\{Q, W\}$ cuts used in each fit. The original nCTEQ15 fit used cuts of $Q > 2.0$ GeV, and $W > 3.5$ GeV. For the new fits, we will relax these cuts to $Q > 1.3$ GeV, and $W > 1.7$ GeV to expand the kinematic range covered.

The $W > 1.7$ GeV cut is chosen to avoid resonance contributions. However, there are indications that it may be possible to further reduce this cut to extend coverage of the shallow inelastic scattering (SIS) region [77], and we examine this briefly in Sec. V.

Now that we have outlined the fits, we present the results in the following section.

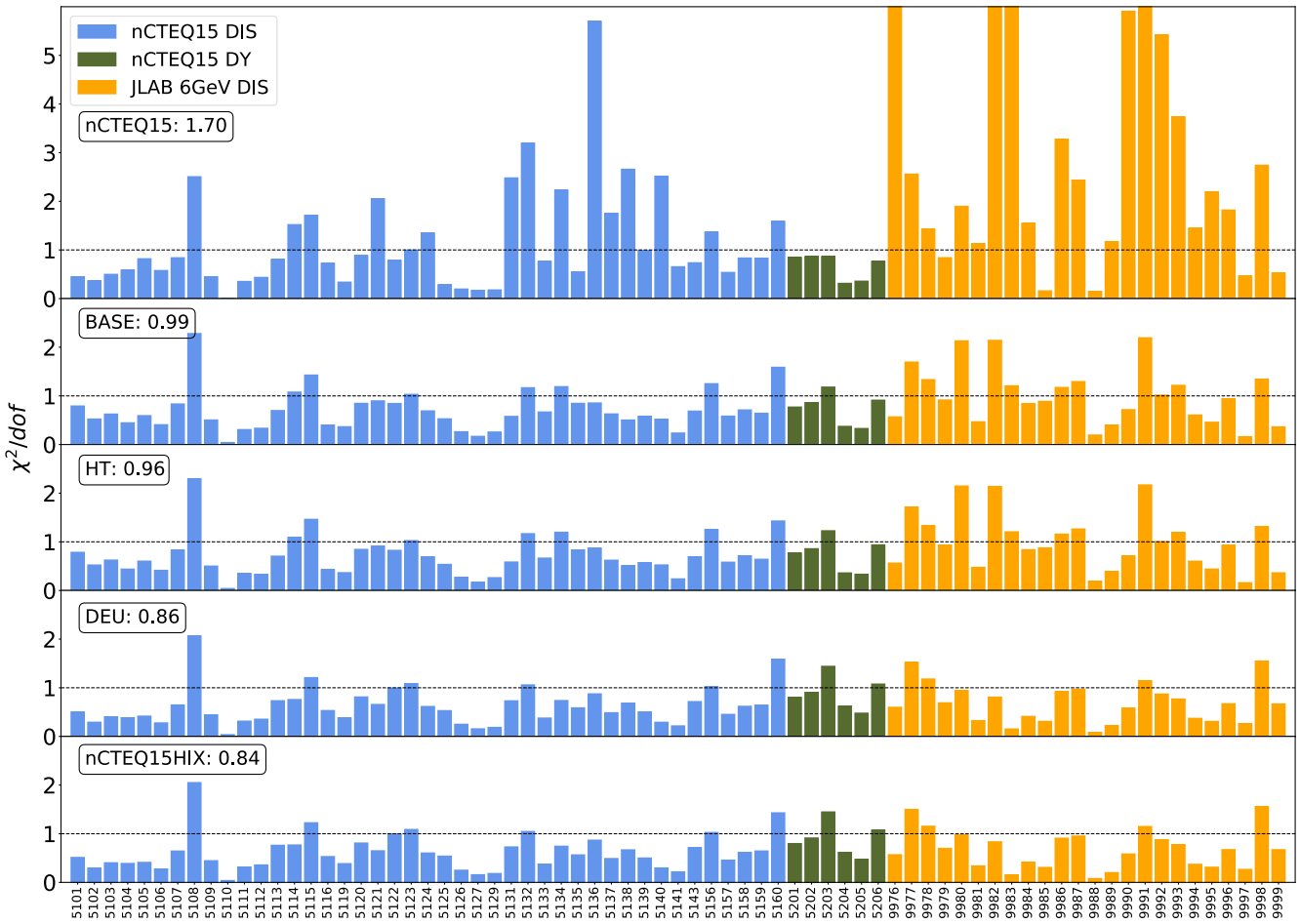

FIG. 6. The $\chi^2/N_{\rm dof}$ for each individual experiment. The experiment ID is listed on the abscissa and are identified in Tables III, IV, V and VI. The original DIS data used in the nCTEQ15 fit are in blue, the DY data in green, and the new JLab data in yellow. We also list the total $\chi^2/N_{\rm dof}$ for each fit inset in the plot.

IV. RESULTS OF THE FITS

A. The χ^2 and $\chi^2/N_{\rm dof}$ values

The $\chi^2/N_{\rm dof}$ values provide a succinct measure of the quality of the fit, and these are summarized in Table II. Additionally, Fig. 6 displays the $\chi^2/N_{\rm dof}$ for each individual experiment for the separate fits, and the experiment IDs are listed in Tables III, IV, V, and VI.

*nCTEQ15**: Starting with nCTEQ15, we first present the values for the original fit which used cuts of $Q > 2$ GeV and $W > 3.5$ GeV and obtained an excellent $\chi^2/N_{\rm dof}$ of 0.81. However, if we take this nPDF and compute the $\chi^2/N_{\rm dof}$ including the new data (and new less stringent cuts)—without fitting—we obtain a very large $\chi^2/N_{\rm dof}$ of 1.70. Inspecting Fig. 6 we see that this large value reflects the fact that these new data, which are beyond the kinematic bounds of the original fit, are not well described.[6] These two entries in Table II provide a useful benchmark so we can see how our new fits improve in comparison.

BASE: The BASE fit will serve as our reference. In this fit, we have simply taken nCTEQ15 as a starting point, and preformed a new fit including the JLab data; no additional corrections are included. Comparing to nCTEQ15*, we see substantial improvement as the χ^2 reduces from 2664 to 1525 to yield a $\chi^2/N_{\rm dof}$ of 0.99. In Fig. 6 we observe that the $\chi^2/N_{\rm dof}$ values for the individual experiments have also improved significantly.

HT: The HT fit applies the higher-twist correction of Sec. II B, and yields moderate improvement compared to the BASE fit (1489 vs 1525), as well as associated improvements in the individual experiments of Fig. 6.

DEUT: The DEUT fit applies the deuteron correction of Sec. II C, and yields significant improvement compared to BASE (1331 vs 1525). Additionally, in Fig. 6 we observe that the fit to the JLab datasets has improved and all the individual $\chi^2/N_{\rm dof}$ values are now below 2.0, in contrast to both the BASE and HT fits.

[6]The effect of removing isoscalar corrections in the original nCTEQ15 dataset is limited.

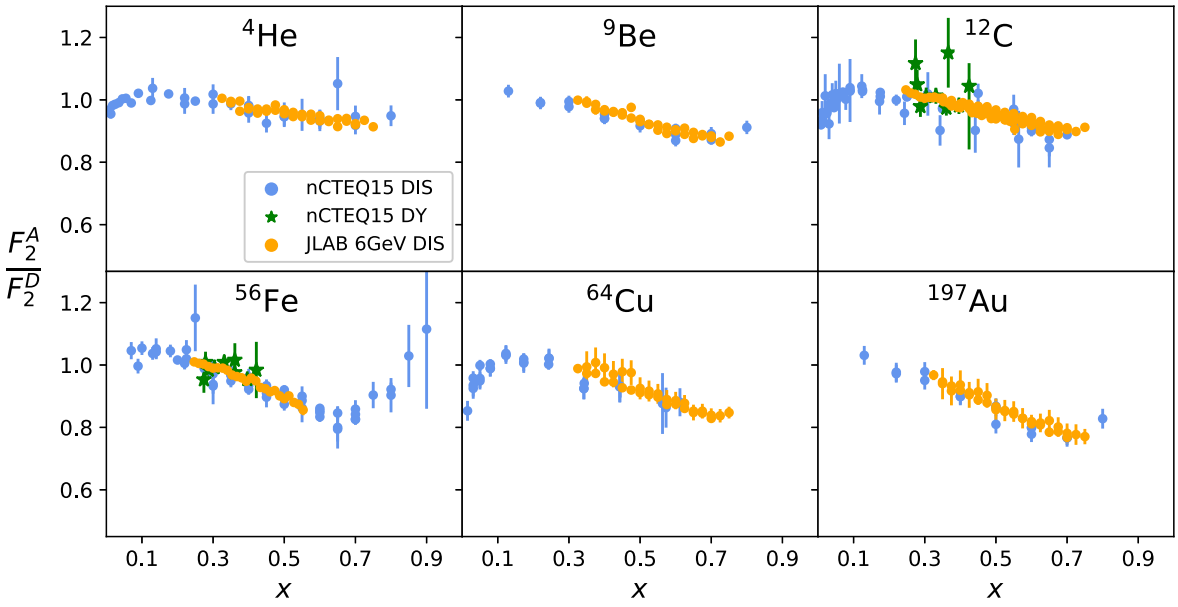

FIG. 7. We display the F_2^A/F_2^D ratio of selected datasets sorted by nuclei. The data from the original nCTEQ15 DIS points are in blue, and the DY in green. The new JLab DIS data are in yellow.

nCTEQ15HIX: Finally, the nCTEQ15HIX fit applies both the deuteron correction and the higher twist power correction. This combination yields the best fit (1297 vs 1525), and reduces the $\chi^2/N_{\rm dof}$ to 0.84. Furthermore, examining the $\chi^2/N_{\rm dof}$ values of Fig. 6 we see each experiment is well fit, with the exception of set 5108 which is a known outlier.[7] Comparing to the BASE fit, we observe that some of the original nCTEQ15 DIS datasets show a reduced $\chi^2/N_{\rm dof}$ indicating the HT and DEUT corrections also improve the fit to these data.

Summary: In summary, we find that the HT modifications provide some improvements ($\sim$3% of $\chi^2/N_{\rm dof}$), the DEUT modifications provide larger improvements ($\sim$10% of $\chi^2/N_{\rm dof}$), and the combination of the HT and DEUT modifications yield the best fit with an improvement of $\sim$15% of $\chi^2/N_{\rm dof}$.

B. Data and theory comparison

Having evaluated the $\chi^2/N_{\rm dof}$ for the separate experiments, we will examine some sample comparisons of our predictions with the data.

In Figure 7 we display the nucleus-to-deuteron F_2^A/F_2^D ratio for the selected datasets sorted by nuclei. The characteristic EMC shape is evident, as well as the A dependence. For example, we see that starting at $x \sim 0.3$ where the ratio is $\sim$1.0 for all A, by $x \sim 0.7$ this ratio dips to $\sim$0.95 for ^{4}He, and to $\sim$0.75 for ^{197}Au.

[7]Note, we find the DIS experiment 5108 (Sn/D EMC-1988) to be an outlier with $\chi^2/N_{\rm dof} > 2$, and this is consistent with other analyses [3,4].

In Fig. 8 we display the nucleus-to-proton $(F_2^A/F_2^D) \cdot (F_2^D/F_2^p)_{\rm CJ}$ ratio, again sorted by nuclei. Here, we have multiplied by the ratio $(F_2^D/F_2^p)_{\rm CJ}$ taken from the CJ15 study [46], shown in Fig. 5, to approximately convert the results of the previous figure to $\sim(F_2^A/F_2^p)$. Note that the introduction of the x-dependent multiplicative $(F_2^D/F_2^p)_{\rm CJ}$ factor visually suppresses the A-dependent change in slope seen in Fig. 7. However, a check of the values of (F_2^A/F_2^p) at $x \sim 0.3$ and $x \sim 0.7$ for ^{4}He and ^{197}Au confirms that the A-dependent change in slope has been maintained.

In Fig. 8 we also display the corresponding theoretical calculations (blue line) obtained with the nCTEQ15HIX PDFs. We can see that they provide a very good description of the fitted data.

C. The PDFs

Finally, we examine the impact of the various corrections on the underlying PDFs. Although the up and down flavors dominate in the high-x region, these PDFs will also influence the gluon and sea-quark distributions indirectly. Thus, we must examine all the flavors to obtain a complete picture of the impact of the JLab data.

We will present a sample of three representative nuclei: {^{12}C, ^{56}Fe, ^{208}Pb}

. As the bulk of the new data is at lower A values we will focus our attention on the ^{12}C PDFs, but we will want to examine the heavier nuclei to infer the various trends in A as well as to verify that our extrapolations to Pb are reasonable. For these plots, we have also displayed the uncertainty bands computed with the Hessian method for both nCTEQ15HIX and the original nCTEQ15 PDF sets. Note, that when plotting nuclear PDFs, we always display

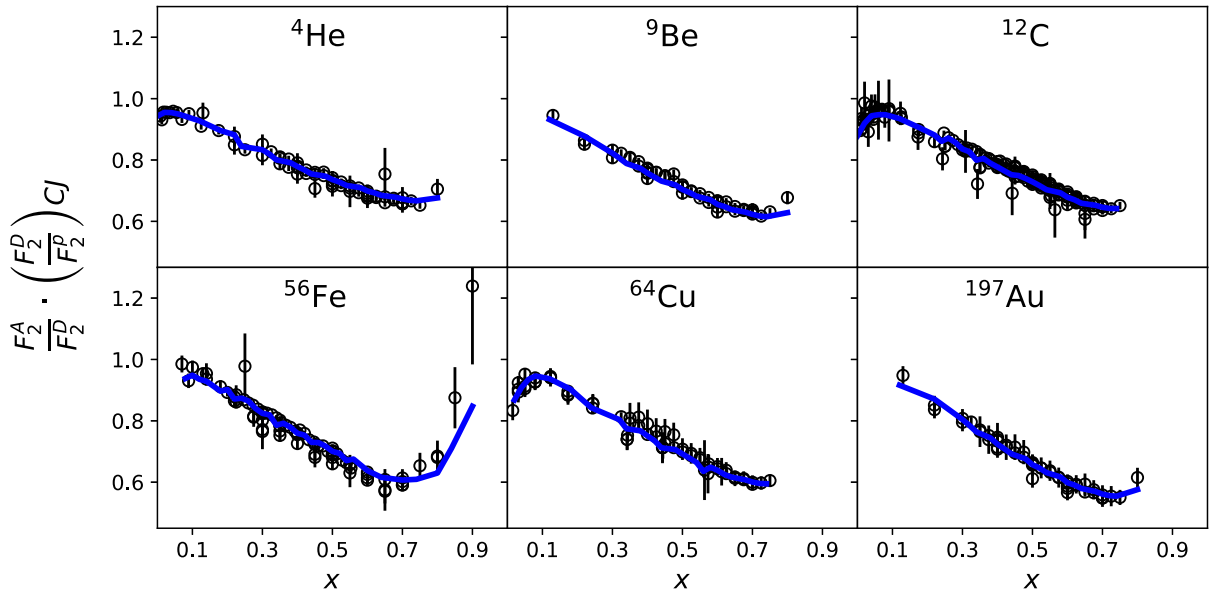

FIG. 8. We display the $F_2^A/F_2^D \cdot (F_2^D/F_2^p)_{\mathrm{CJ}}$ ratio for selected datasets sorted by nuclei. We also overlay the theoretical prediction of nCTEQ15HIX in blue. The theory predictions have been calculated at averaged Q values where datasets overlap.

the full nuclear distribution, f^A, constructed out of the bound proton (neutron) distributions, $f^{p(n)/A}$, as:

$$f^A = \frac{Z}{A} f^{p/A} + \frac{(A-Z)}{A} f^{n/A}, \tag{8}$$

with $f^{n/A}$ constructed using the isospin symmetry.

In Figs. 9 and 10 we present the results for ^{12}C. Figure 9 shows the magnitude of the PDF $xf(x,Q)$ at $Q = 2$ GeV on a log-scale. We observe that the separate fits are generally similar across the full x kinematic range and that there is no abnormal behavior in either the low or high x region. Additionally, this figure underscores the extent to which the PDFs are decreasing as we move to larger x, especially

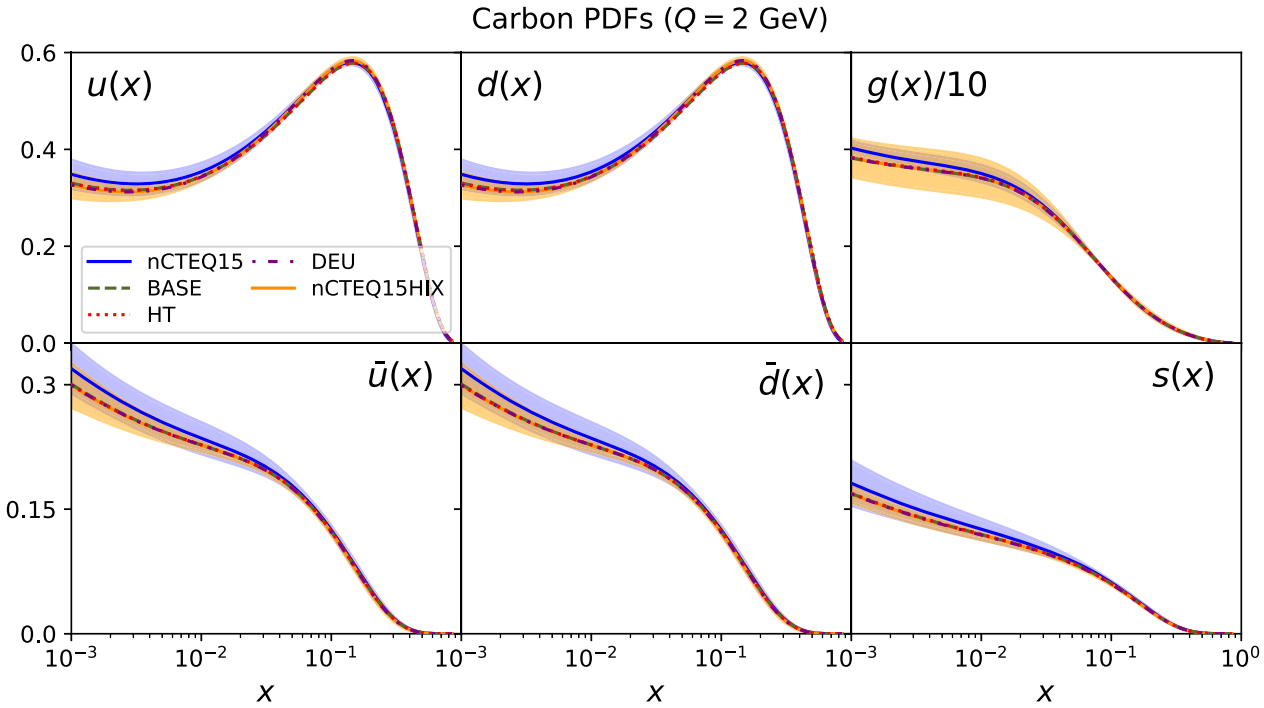

FIG. 9. Carbon (^{12}C) nPDFs $xf^{\mathrm{C}}(x,Q)$ at $Q = 2$ GeV. We show the uncertainty bands for nCTEQ15 (blue) and nCTEQ15HIX (yellow) computed with the Hessian method.

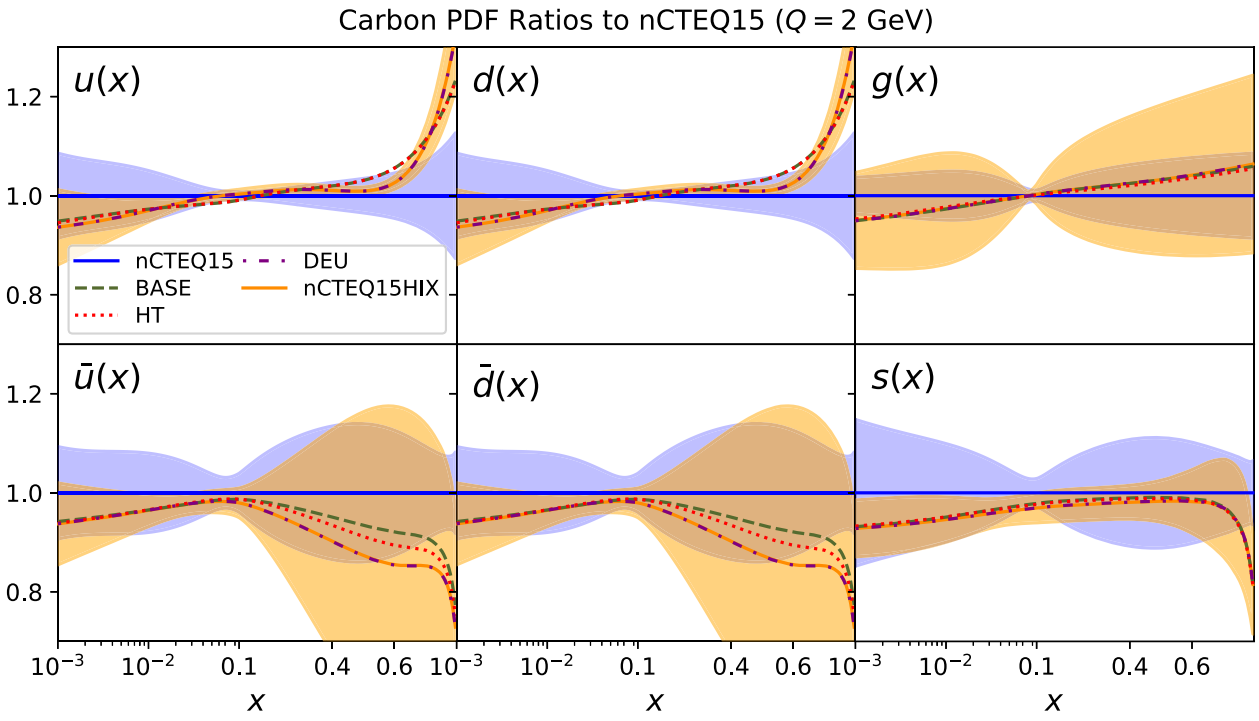

FIG. 10. The corresponding ratio of nPDFs compared to the nCTEQ15 central nPDFs for ^{12}C using a log-linear scale to highlight the large-x region. We show the uncertainty bands for nCTEQ15 (blue) and nCTEQ15HIX (yellow) computed with the Hessian method. Note that while DEUT and nCTEQ15HIX are distinct nPDFs which yield differing $\chi^2/N_{\rm dof}$ values, these differences are imperceptible on the scale of this figure, as well as in Figs. 11–12.

for the gluon and sea-quarks $\{g, \bar{u}, \bar{d}, s = \bar{s}\}$; obtaining large event statistics in the high-x region is challenging.

1. The small x region

In Fig. 10 we display the ratio of the fits compared to the nCTEQ15 nPDFs, we use here a log-linear scale to better illustrate the full x range. Our first observation is that all of our new fits to the JLab data yield very comparable results in the small x region. This is a reasonable result as the different corrections we are applying in the fits (e.g., HT, DEUT) all primarily impact the large x region. We also observe that all these fits are uniformly below the nCTEQ15 PDFs by ∼6%. Thus, it appears the fits are effectively hardening the high-x PDFs of select parton flavors (especially those of the u- and d-quark), while somewhat suppressing PDFs in the small-x region. This behavior suggests increased shadowing and merits further investigation.

2. The large x region

From Fig. 10 it is clear that the fits are hardening the nuclear PDFs at larger x to better describe the new datasets. In particular, we see the up, down and gluon PDFs all are larger than nCTEQ15 while the sea quarks $\{\bar{u}, \bar{d}, s = \bar{s}\}$ are smaller. For example, if we look in the region of $x \sim 0.7$ we see the $\bar{u}$ and $\bar{d}$ PDFs are reduced by as much as ∼10% while the gluon is increased by ∼5% and the up and down PDFs are increased by ∼10%.

Comparing the individual fits, we see a very clear pattern in the $\bar{u}$ and $\bar{d}$ PDFs. As we move from the nCTEQ15 fit, to the BASE, HT, DEUT, and finally the nCTEQHIX fit we see the $\bar{u}$ and $\bar{d}$ PDFs are uniformly decreasing in the large x region. Recall that this ordering of the fits also shows a monotonically decreasing value of $\chi^2/N_{\rm dof}$, cf. Table II.

We also observe that the BASE and the HT fits are quite similar, as the addition of the HT corrections, that are substantially smaller for structure function ratios than for either the numerator or the denominator, see Fig. 4, only show a slight improvement in the $\chi^2/N_{\rm dof}$.

Similarly, the DEUT and the nCTEQ15HIX are also quite similar as these fits only differ by the inclusion of the HT corrections (for the nCTEQ15HIX fit). This difference is barely discernible in the figures as the curves are generally within a percent, but nCTEQ15HIX generates a 3% improvement in $\chi^2/N_{\rm dof}$ compared to DEUT. These two fits also produce the largest downward shift of the $\bar{u}$, $\bar{d}$ and strange PDFs while yielding a smaller increase of the up and down PDFs. Recall that DEUT and nCTEQ15HIX have the lowest $\chi^2/N_{\rm dof}$ values.

3. Large A results

Having investigated the carbon results in detail, we now want to explore the larger A values to see how the above effects scale to heavier nuclei. In Figs. 11 and 12 we display the PDF ratios (compared to nCTEQ15) for iron and lead PDFs on a log-linear scale.

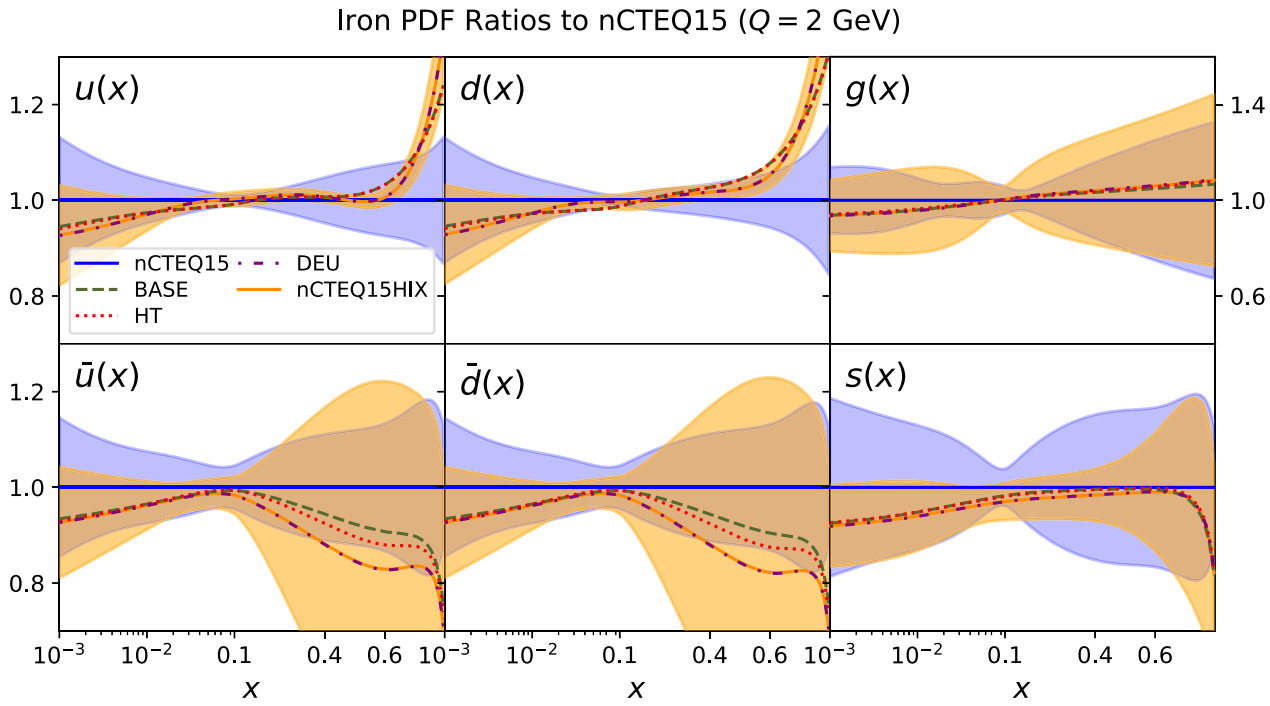

FIG. 11. Iron (^{56}Fe) PDFs ratio compared to nCTEQ15 at $Q = 2$ GeV. We show the uncertainty bands for nCTEQ15 (blue) and nCTEQ15HIX (yellow) computed with the Hessian method.

The small x behavior is quite similar to the carbon PDF result. At $x = 10^{-3}$ all the PDFs are shifted downward by ~6% to ~8%, and the shift of the individual fits is quite uniform.

In the large x region, the behavior is also similar to the carbon PDFs, but the size of the shifts is slightly larger. For example, at $x \sim 0.7$ the $\bar{u}$ and $\bar{d}$ PDFs for carbon were shifted downward by ~10% compared to nCTEQ15, while

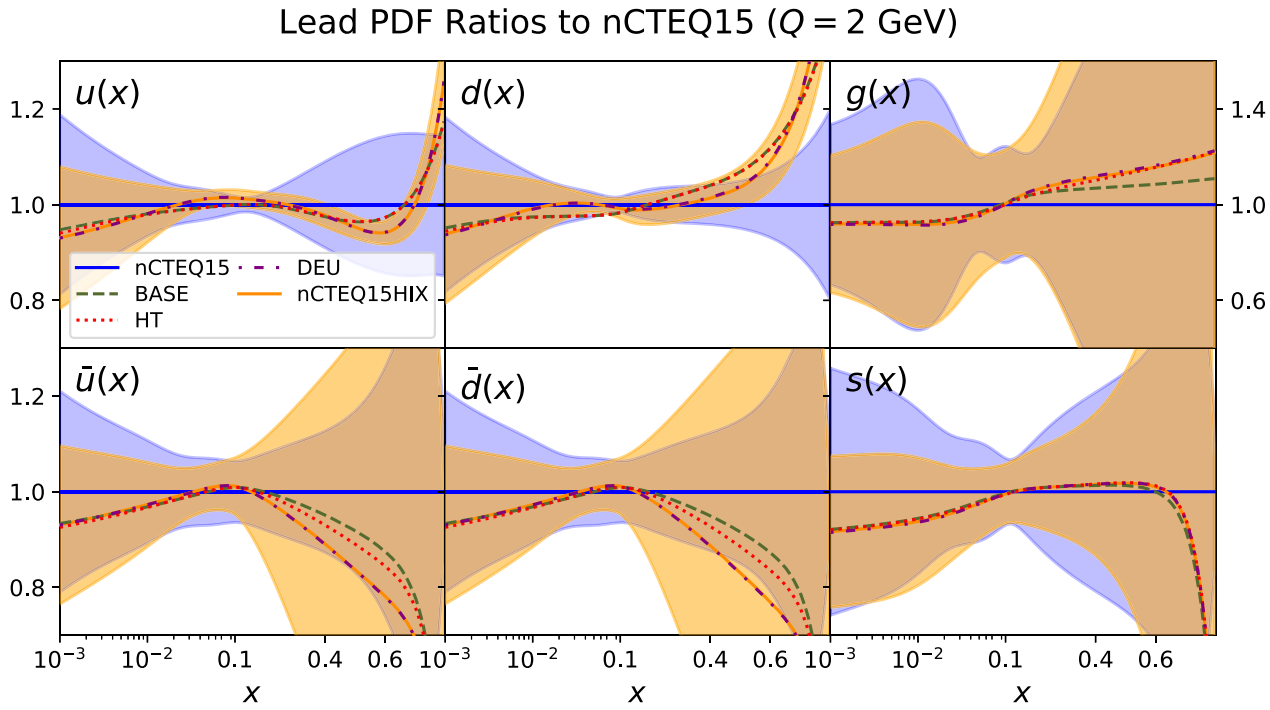

FIG. 12. Lead (^{208}Pb) PDFs ratios compared to nCTEQ15 at $Q = 2$ GeV. We show the uncertainty bands for nCTEQ15 (blue) and nCTEQ15HIX (yellow) computed with the Hessian method.

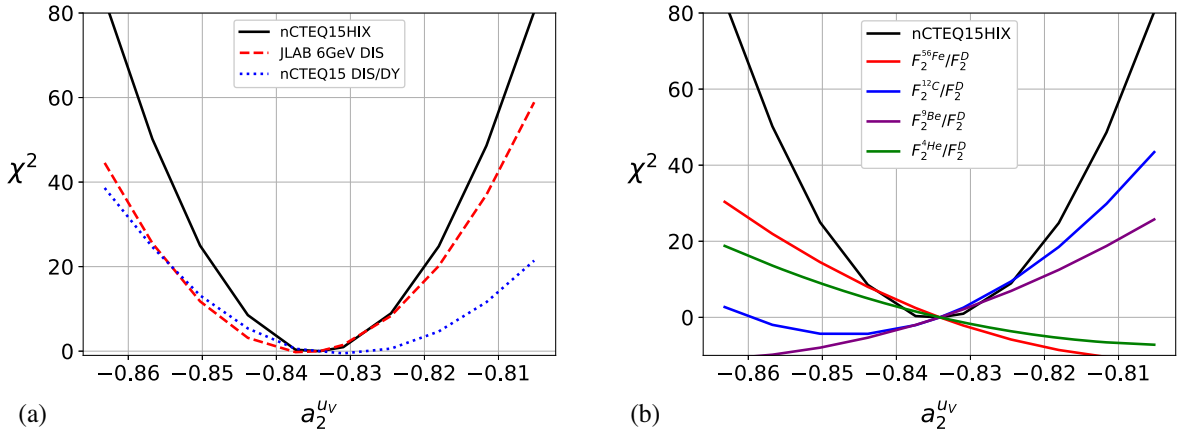

FIG. 13. PDF parameter profiles of the up-valence parameter $a_2^{u_v}$. (a) We display the χ^2 shift due to the original nCTEQ15 dataset (blue dots), the new JLab dataset (red dashes), and the total (black solid). (b) We display the χ^2 shift for the most significant projectile-target combinations constraining the $a_2^{u_v}$ parameter.

the iron and lead are shifted by up to ∼30%. We again see that the gluon, up and down PDFs are increased at large x. Again comparing at $x \sim 0.7$, while the up and down PDFs of carbon are increased by ∼10%, the iron and lead are shifted by up to ∼20%.

Clearly the new data has significant impact on the PDFs in this kinematic region. We also observe that the impact of the individual corrections {HT, DEUT} are rather uniform, and we obtain the best fit using both the HT and DEUT corrections which give us the nCTEQ15HIX fit. Therefore in the following, we will now focus on this fit.

4. PDF uncertainties

Finally, we now compare the uncertainties of the nCTEQ15 and nCTEQ15HIX PDF sets as displayed in Figs. 9–12. These results should be interpreted carefully as we must be attentive to parametrization bias and other imposed constraints. For example, our parametrization ties the strange PDF to the average $(\bar{u} + \bar{d})$ distribution, so this is not a fully independent distribution.[8]

Examining Fig. 9 showing the ^{12}C PDFs, both nCTEQ15 and nCTEQ15HIX have comparably sized uncertainty bands, with the exception of the gluon. The larger uncertainty band for the nCTEQ15HIX gluon may be more representative of the true uncertainties as it can be difficult to fully explore the loosely constrained gluon in the small x region where minimization methods often encounter troublesome flat parameter directions. In this sense, the larger gluon uncertainty band of, for example, the EPPS16 nPDF set [4] may be more characteristic (cf. Fig. 18).

Note that the nCTEQ15 fit includes data on single inclusive pion production from the STAR and PHENIX experiments, which we do not include in the nCTEQ15HIX fit. However, we have verified these experiments are compatible with our nCTEQ15HIX fit, and they yield a $\chi^2/N_{\rm dof} \sim 0.5$. The pion data can help constrain the nuclear gluon PDF, especially in the mid- to lower-x region, and to some extent, the slight increase of the gluon uncertainty of nCTEQ15HIX can thus be attributed to this.

Turning to the PDF ratio plots of Figs. 10–12, again these should be interpreted carefully, especially in the large x region where we have a ratio of small numbers. We observe that both the up and down PDFs appear to have reduced uncertainty bands in the larger x region. This makes sense as the bulk of our new data is constraining these flavors in this kinematic region. The gluon uncertainty is increased compared to nCTEQ15, presumably for the same reasons discussed above. The $\bar{u}$ and $\bar{d}$ uncertainties are roughly comparable between the nCTEQ15 and nCTEQ15HIX, as we expect only indirect constraints on these flavors.

While the uncertainties of the deuteron structure functions displayed in Fig. 5 are not used in the computation of the error bands, their impact can be assessed by comparing the fits without (BASE, HT) and with (DEUT, nCTEQ15HIX) the deuteron corrections. These differences lie within the uncertainty bands in all cases, except for the up and down nPDFs at intermediate x values $(x \sim 0.5)$; hence, our uncertainties in this region may be slightly underestimated.

Finally, the strange PDF also shows a reduced uncertainty band. As mentioned in Sec. III B, this distribution is tied to the $(\bar{u} + \bar{d})$ sea quarks and we have not introduced any free strange parameters. For a more representative uncertainty estimate of the strange PDF, compare with the error bands of nCTEQ15WZ (cf. Ref. [71], Fig. 12).

D. Parameter scans

We have performed parameter scans for the 19 parameters of the nCTEQ15HIX fit, and selected results are

[8]Note, our recently released nCTEQ15WZ nPDF analysis included LHC W/Z production data which allowed to provide additional freedom (and constraints) for the strange PDF [71].

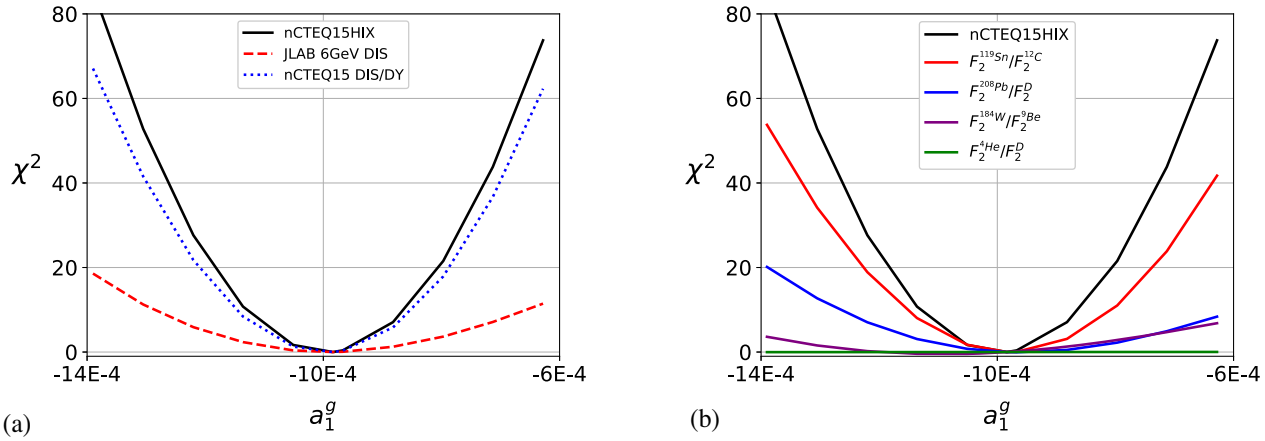

FIG. 14. PDF parameter profiles of the gluon parameter a_1^g. (a) We display the χ^2 shift due to the original nCTEQ15 dataset (blue dots), the new JLab dataset (red dashes), and the total (black solid). (b) We display the χ^2 shift for the most significant projectile-target combinations constraining the a_1^g parameter.

displayed in Figs. 13 and 14. As expected, the JLab data exerts strong constraints on the parameters of the up and down quarks, while the impact on the sea quarks and gluon is reduced. We also note that for the sea quarks (not shown) and gluon, the value of the parameters at the minimum χ^2 is quite similar for both the nCTEQ15 set and the new JLab set. However, for the up and down parameters, where the JLab data has more constraining power, we find in some cases (e.g., $a_2^{u_v}$) the fit minimum is a compromise between the two datasets. This is a common feature in global PDF fits with large datasets. To gain additional insight on which datasets are driving this result we now examine Figs. 13 and 14 in detail.

In Fig. 13 we display the profile for the $a_2^{u_v}$ parameter. The first plot, Fig. 13(a), shows the total χ^2 shift, as well as the individual shifts due to the original nCTEQ15 dataset and the new JLab dataset. Clearly the new JLab data help constrain this parameter compared to the nCTEQ15 data alone. We also observe the original nCTEQ15 data prefers slightly larger values for the $a_2^{u_v}$ parameter as compared to the JLab data; this type of behavior is common for global PDF fits. In Fig. 13(b) we display the χ^2 shift for the most significant datasets constraining the $a_2^{u_v}$ parameter. We note the datasets with the largest impact are the ^{56}Fe/D data which is pulling $a_2^{u_v}$ to higher values, and the ^{12}C/D data which is pulling $a_2^{u_v}$ to lower values.

In Fig. 14 we display the profile for the a_1^g parameter. Fig. 14(a) again shows the total χ^2 shift, as well as the individual shifts due to the original nCTEQ15 dataset and the new JLab dataset. We observe the new JLab data provides a weaker constraint on the gluon as compared to the valence quarks. We also note the preferred minima of both the original nCTEQ15 data and the JLab data coincide for the gluon parameter. Turning to Fig. 14(b) we see the χ^2 shift for the most significant datasets constraining the a_1^g parameter. In contrast to the above case of $a_2^{u_v}$, we note the datasets with the largest impact are ^{119}Sn/^{12}C and ^{208}Pb/D; in this case, the ^{56}Fe/D data has less impact on this parameter.

V. THE LOW-*W* KINEMATIC REGION

The $W > 1.7$ GeV cut on the data fitted in the nCTEQ15HIX analysis discussed above was imposed to restrict the study to the DIS regime, where we can reasonably assume QCD factorization to work, and avoid resonance excitations in electron-scattering events. However, it has been shown that the characteristic x dependence of nuclear structure-function ratios displayed, for example, in Fig. 7 persists into the resonance region at low W values. Specifically, Fig. 3 of Ref. [41] overlays data from the resonance region ($1.1 < W < 1.7$ GeV, $Q \sim 2$ GeV) with $W > 1.7$ GeV DIS data from the SLAC E139 [28], SLAC E87 [30], and BCDMS [32] experiments, and finds that "the agreement of the resonance region data with the DIS measurement [...] is quite striking." This is likely a manifestation of the quark-hadron duality phenomenon, abetted in the case of nuclear scattering by Fermi motion of the bound nucleons, which results in a further, effective averaging of the nucleon structure functions over multiple resonances; together, these dynamics may permit a description of nuclear structure functions in terms of partonic degrees-of-freedom, even in kinematic regions where resonance excitation is the dominant effect. It is therefore interesting to remove the W cut, and explore to what extent the resonance region data can be described in terms of nPDFs.

A. nPDFs at large *x*

An important difference between nuclear PDFs and proton PDFs is that, in principle, the momentum fraction x of a parton with respect to the average nucleon momentum in a nucleus can range from $[0, A]$ whereas

that of a proton is confined to lie in the range $[0, 1]$. Phenomenologically, at the photon energies under consideration, nuclear partons belong to bound nucleons and share a relative fraction $x_N \in [0, 1]$ of their momentum. As the nucleons interact among themselves, they exchange momentum; thus, their partons can receive a boost in the photon's direction, causing their momentum fraction x to shift and exceed 1. Since the nucleon-nucleon interaction is on average soft, the shifts are moderate and the nuclear parton distributions are still predominantly within $x \sim [0, 1]$ with only a small tail exceeding 1.

This effect is currently not captured in the nPDF parametrization adopted in this paper, see Sec. III B, which is confined to $x \in [0, 1]$ and needs to be suitably extended. In fact, a full description of nPDFs in the $x \in [0, A]$ range would also entail a generalization of DGLAP evolution equations and careful consideration of the sum rules [78].[9] In this paper we limit ourselves to an exploratory study and demonstrate, as a proof of principle, that a simple generalization of the adopted parametrization can lead to a reasonable description of electron-nucleon collisions in terms of nuclear PDFs even in the resonance region. We will leave more detailed studies for future work.

1. Convolution kernel

The parton momentum shift caused by nucleon-nucleon interactions can be effectively described by a convolution of the nucleon-level PDF f with a smearing function S_A:

$$f_A(x, Q) = \int_x^A \frac{dy}{y} S_A(y, Q) f(x/y, Q), \tag{9}$$

where S_A represents the probability that the parton f belongs to a nucleon of momentum fraction y compared to the average nucleon momentum. Such representation is, in fact, grounded on the so-called weak binding approximation to the calculation of nuclear structure functions, where S_A can be obtained starting from the nucleon wave function in the corresponding nucleus [79–84]. In the weak binding approximation, the smearing function is in general scale-dependent at finite Q values. Therefore f_A does not satisfy the DGLAP evolution equations except at $Q \to \infty$ where S_A becomes scale independent, or at small x where the smearing effect is marginal. In the context of nPDF fits, however, we rather regard the convolution of Eq. (9) as a generalization of the nPDF parametrization discussed in Sec. III B, and would only apply it at the initial scale Q_0.

At x values not exceeding 1, the nucleon-nucleon interaction is soft and the smearing function can be approximated by a normalized Gaussian,

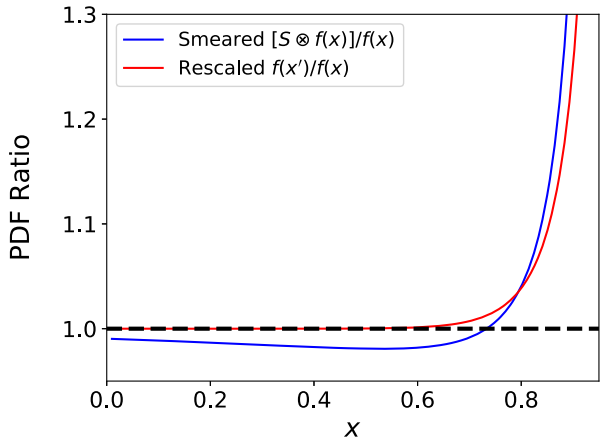

FIG. 15. We display the ratio of the modified (smeared/rescaled) PDF compared to the unmodified PDF for both the Gaussian smearing and the x-rescaling. For the Gaussian smearing (blue), we display $S \otimes f(x)$ over $f(x)$ with $\Delta = 0.05$, and $\delta = -0.007$ (which resembles an $A = 12$ nucleus). For the x-rescaling (red), we display $f(x'_A)/f(x)$ for $\varepsilon = 0.02$, $\kappa = 10$, and $A = 12$. While there is a small relative difference at intermediate to small x values, the rise at large x is quite similar for both methods.

$$S_A(y) = \frac{1}{\sqrt{2\pi}\Delta_A} \exp\left\{\frac{-[y - (1 - \delta_A)]^2}{2\Delta_A^2}\right\}, \tag{10}$$

where the effect of nucleon-nucleon interactions is described by the Gaussian width Δ_A. Furthermore, the nucleon binding causes a (small) shift of the Gaussian peak toward smaller x values, and this is controlled by the δ_A term. Note that the larger the nucleus, the more the nucleons that interact with each other (hence a larger Δ_A) and the larger the binding energy (hence a larger δ_A).

The net result of this convolution is to deplete the partons in the intermediate x region, and enriching the large $x \gtrsim 1$ region. Hence the nuclear ratio F_2^A/F_2^B with $A > B$ will also display a dip below unity at intermediate x values (the EMC region) and a relatively steep rise above unity in the $x \to 1$ Fermi motion region, as seen experimentally. With a suitable choice or fit of the Δ_A and δ_A parameters one can therefore expect to quantitatively describe the observed nuclear effects in the Fermi motion region.[10]

As an example, in Fig. 15 we show the ratio of convoluted to unmodified PDF for a toy model $f(x) = x^{-1.5}(1 - x)^3$, which reproduces the main features of the up and down quark PDFs. The parameters have been chosen to

[9]Note that restricting our formalism to $x \in [0, 1]$, i.e., assuming $f_i^A(x, Q) = 0$ for $x > 1$ for all parton flavors i as done in all current nPDF analyses, is self-consistent under scale-evolution and preserves the parton sum rules.

[10]In principle, by generalizing the Gaussian *ansatz* in Eq. (10) for the smearing function and including, for example, power law tails [85], the convolution formalism of Eq. (9) can also be used to extend the analysis into the $x > 1$ region. We leave this for future work.

approximate the behavior of the carbon structure function data.

2. *x*-rescaling

While the smearing approach has a close connection to the underlying nuclear dynamics and guarantees that the PDF sum rules are satisfied, the convolution is computationally expensive. Hence, in this exploratory study we will try a simple remapping of the parametrization considered so far, by an A-dependent shift of the x variable.

We want this remapping to approximate the rise of F_2^A/F_2^p in the large x region without distorting the phenomenology in the intermediate to small x region. To ensure these conditions are satisfied, the remapping should (i) only impact the very large x region ($x \sim 1$), and (ii) ensure the various momentum sum rules are satisfied within uncertainties.

The specific remapping we shall use is

$$f_A(x, Q) \to f_A(x'_A, Q) \tag{11}$$

where f_A are the PDFs determined in the nCTEQ15HIX fit, and

$$x'_A = x - \varepsilon x^\kappa \log_{10} A. \tag{12}$$

As the PDFs are typically decreasing functions of the parton's fractional momentum, the negative shift ensures that the transformed function is larger than the unmodified one, and nonvanishing as $x \to 1$. The ε parameter controls the overall size of the rescaling effect, and the x^κ factor ensures that only the large-x region is modified. The $\log_{10} A$ term ensures an increasing modification across the full range of nuclear A values from proton ($A = 1$) to lead ($A = 208$).

In Fig. 16, we show the impact of the rescaling on the ratio f_A/f_p for $A = 208$ and a selection of ε and κ values. For simplicity, we have used here the same $f(x) = (1-x)^3/x^{1.5}$ toy PDF considered in the nuclear smearing discussion. Clearly, the remapping does not preserve the PDF normalization within the $x \in [0, 1]$ range, and therefore breaks the PDF sum rules. However, in all cases the breaking amounts to a small increase of the momentum sum by less than 1%, which is well within the uncertainty of the nPDFs.

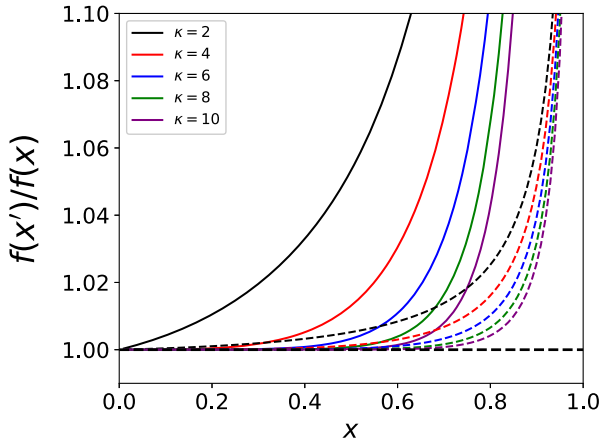

FIG. 16. We show the impact of the x-rescaling of Eq. (12) for varying choices of ε and κ for a "toy" PDF, $f(x) \sim (1-x)^3/x^{1.5}$. The plot shows the ratio $f(x'_A)/f(x)$ for 2 groups of 5 curves for $A = 208$. The largest 5 curves (solid) are for $\varepsilon = 0.01$ with exponent $\kappa = \{2, 4, 6, 8, 10\}$ indicated by the colors {black, red, blue, green, purple}, respectively. The lower 5 curves (dashed) are for $\varepsilon = 0.001$ with the same range of κ values.

B. nPDFs in the $x \to 1$ limit

Overall, the remapping accomplishes our stated goals, and can be used to investigate how well it can describe the data at the highest x, that was excluded from our nCTEQ15HIX fit due the $W > 1.7$ GeV cut.

To simplify the discussion, we fix the small-x suppression parameter to $k = 10$ and will study the effect of varying the ϵ parameter. One can appreciate that this choice is reasonable by looking at Fig. 15, where the $A = 12$ remapped f_A/f_p, plotted in red, is compared with the blue convolution model discussed earlier. By design the remapping only deforms the PDF at large x, and does not reproduce by itself the intermediate-x dip displayed by the convolution model. This however is small and can be easily taken care of in a fit by the standard nPDF parameters discussed in Sec. III B. On the contrary, the large-x steep rise is pretty well approximated by the remapped PDF. We have furthermore verified that $k = 10$ is also a good choice for other nuclear targets, given the assumed logarithmic A dependence of the rescaled x'_A momentum fraction.

We can now investigate the impact of rescaling the nCTEQ15HIX nPDFs and compare these to the high-x data. In Fig. 17 we display a selection of datasets that have data points in the $x \to 1$ region, and compare them to calculations performed with shifted nPDFs for fixed $\kappa = 10$ and a representative choice of ϵ parameters in the [0, 0.1] range. Note that in the ^{56}Fe and ^{64}Cu panels, the empty symbols represent data points at $W < 1.7$ GeV, that were excluded from the nPDF fit. We also display the computed χ^2 values in the plot legend. These are computed both for the points that satisfy the $W > 1.7$ GeV cut (labeled χ^2_{cut}), and for all the points (labeled χ^2_{nocut}). Note that for these χ^2 calculations we have simply applied the x rescaling when computing $F_2(x, Q)$ using the nCTEQ15HIX nPDFs; this is not a separate fit.

If we focus on the top two panels of Fig. 17, ^{12}C and ^{27}Al, we see all these data points satisfy the $W > 1.7$ GeV cut. Turning our attention to the bottom two panels of Fig. 17, ^{56}Fe and ^{64}Cu, some data points do not satisfy the

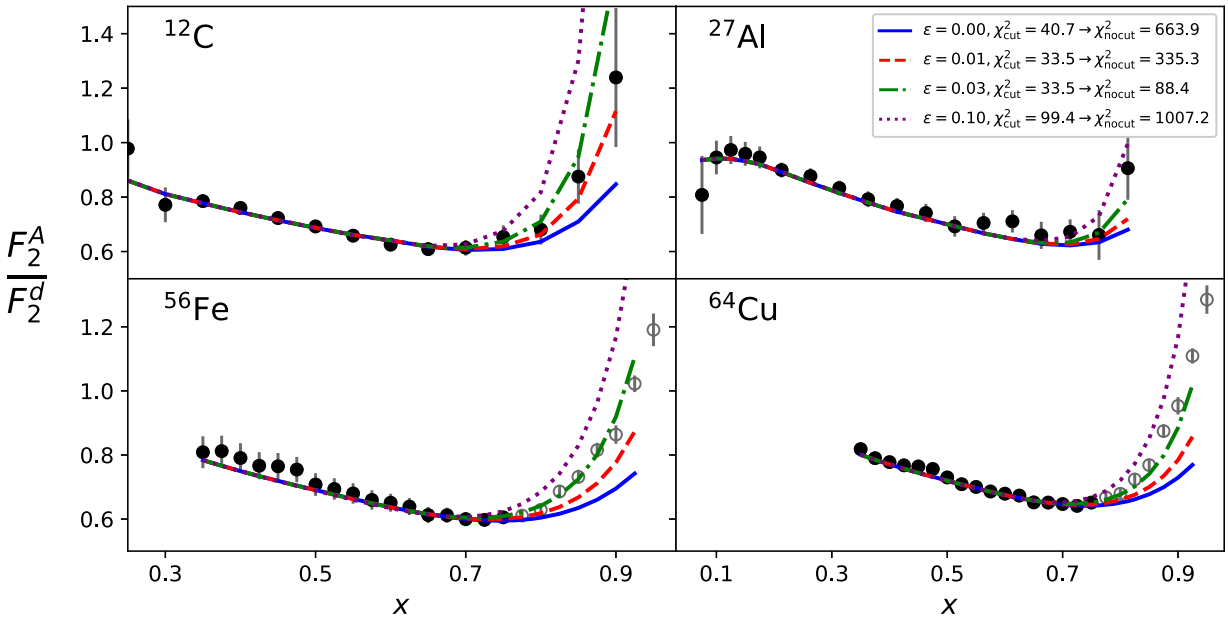

FIG. 17. We show F_2^A/F_2^D for selected datasets in the large x region. The different curves show the impact of the variations of $\epsilon = \{0, 0.01, 0.03, 0.1\}$ (solid blue, dashed red, dot-dashed green, dotted purple) with $\kappa = 10$ in the rescaling of Eq. (12). The combined χ^2 values are displayed in the legend, both with and without the $W > 1.7$ GeV cut. For reference, the bottom curve (blue, solid) in each plot corresponds to $\epsilon = 0$ which is the nCTEQ15HIX result. Points which are excluded from the nCTEQ15HIX fit by the $W > 1.7$ GeV cut are indicated with a hollow (gray) symbol; this is evident for the Fe and Cu datasets. For reference, experiment Fe/D is ID=5131, Al/D is 5134, Cu/D is 9984, and C/D is 9990.

$W > 1.7$ GeV cut, and these points give rise to the differing $\chi^2_{\rm cut}$ and $\chi^2_{\rm nocut}$ values.

Overall, we observe there is a small but significant variation in $\chi^2_{\rm cut}$, and a dramatic variation in $\chi^2_{\rm nocut}$. In both cases, the fit shows preference for $\epsilon \approx 0.03$. With this parameter, we can describe very well the turnover of the structure function ratio at $x \sim 0.7$, and the subsequent large-x rise over the whole A range. Overall, we obtain a very reasonable description of the large-x data in the resonance region.

C. Extending nPDF fits to low W

We have found that the x rescaling provides a simple means of mimicking the behavior of the nuclear PDF in the $x \to 1$ region. While we have used it here *post-facto*, it can also be effectively utilized to generalize the nPDF parametrization discussed in Sec. III B with only 2 additional free parameters. As the x rescaling already increases the F_2^A/F_2^D ratio in the Fermi motion region by about the right amount compared to the baseline nCTEQ15HIX nPDFs, the remaining parameters will more easily adjust to the data, and we can expect an improvement in the fit quality in the high x, low W region.

Furthermore, this analysis suggests that it may be possible to expand the kinematic reach of the fit in the $\{x, Q\}$-space to include data in the low $W < 1.7$ GeV resonance region. This is an interesting possibility, and we reserve the details for a future study.

VI. COMPARISONS WITH OTHER nPDFs

In this section we compare our new nCTEQ15HIX nPDFs with other results from the literature. In particular, in Figs. 18 and 19 we compare with nPDFs from EPPS16 [4], nNNPDF2.0 [6], and TUJU19 [5] analyses. We find the results are generally compatible within the uncertainty bands.

Examining the ratios of Fig. 19 we note some features similar to the comparison between nCTEQ15 and nCTEQ15HIX of Fig. 10. Specifically, for the up, down, and gluon PDFs, nCTEQ15HIX is generally above the other PDFs in the high x region. Correspondingly, for the anti-up and anti-down PDFs, nCTEQ15HIX is generally below the other PDFs in the high x region. This is similar to the pattern observed in Fig. 10, although these difference are well within the uncertainty bands. It will be of interest to follow these differences as the uncertainty bands are reduced in future analyses.

We do not display a comparison with our recent nPDFs nCTEQ15WZ [71] where the nCTEQ15 analysis was supplemented by the available W/Z boson production data from the LHC. This is because here we concentrate on the high-x region of quark distributions, and the W/Z data were mostly relevant for the strange and gluon distributions at lower x values. Hence, they had minimal impact on the high-x behavior of the nPDFs. Therefore, the high x behavior of the nCTEQ15WZ PDFs was quite similar to

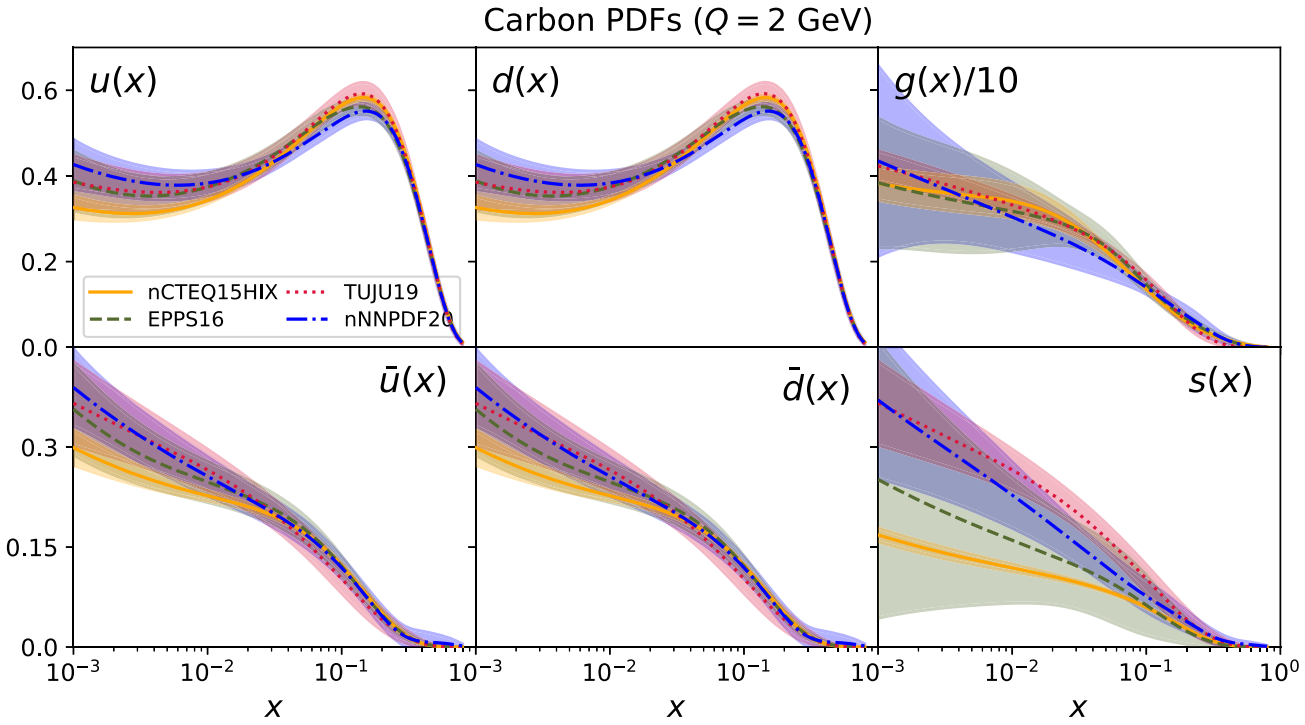

FIG. 18. We compare our nCTEQ15HIX results to other nPDF sets from the literature including EPPS16 [4], nNNPDF2.0 [6], and TUJU19 [5]. We plot $xf(x, Q)$ for carbon ^{12}C at $Q = 2$ GeV on a log scale.

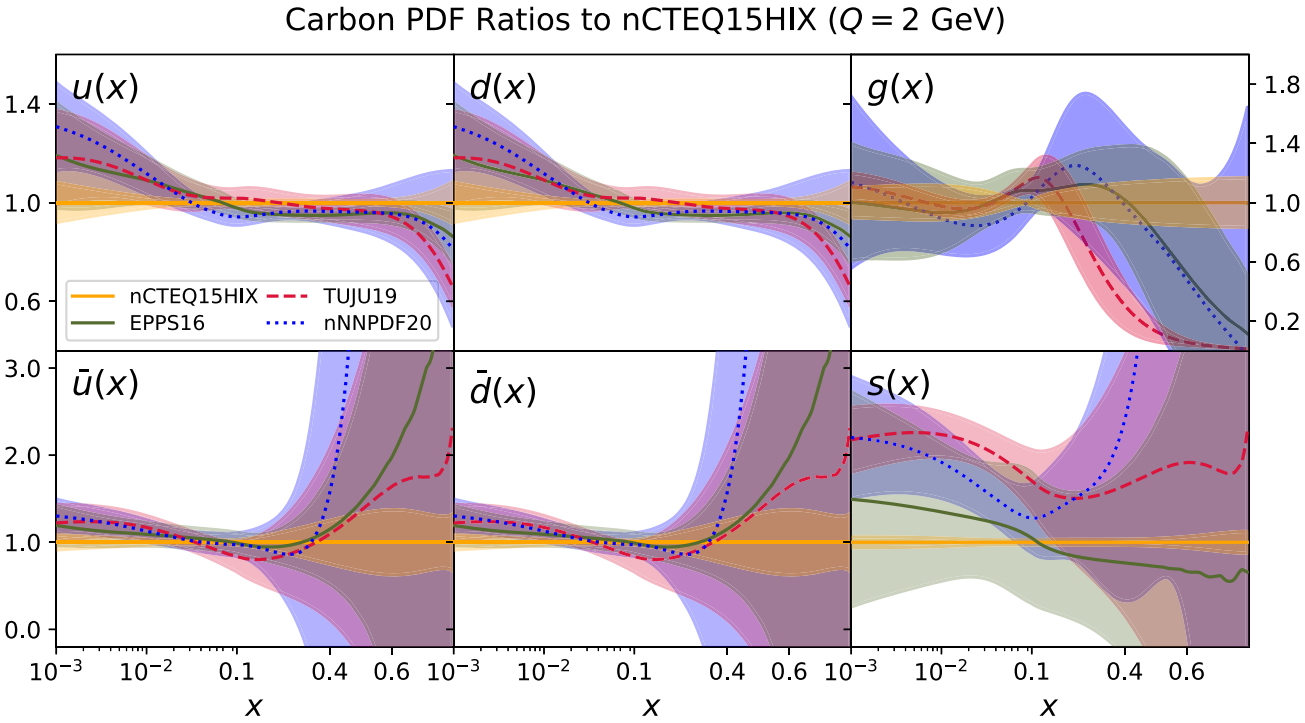

FIG. 19. We compare our nCTEQ15HIX results to other nPDF sets from the literature including EPPS16 [4], nNNPDF2.0 [6], and TUJU19 [5]. We plot the nPDF ratio for carbon ^{12}C at $Q = 2$ GeV compared to nCTEQ15HIX on a log-linear scale.

the original nCTEQ15 nPDFs, and this was already presented in Figs. 9–12.

VII. CONCLUSION

PDFs encode the dynamics of the strong interaction and provide a crucial link between experimental measurements and theoretical models. Thus, the updated nCTEQ15HIX nPDFs presented here will contribute to increased precision of experimental analyses, which can thereby yield further insights into the QCD. The novel aspects of this study consisted of several elements.

The newly included JLab data provide important additional constraints on the nuclear PDFs in the high-x and low-Q regime. We have explored this region with the nCTEQ15HIX set by relaxing our kinematic cuts to $Q > 1.3$ GeV and $W > 1.7$ GeV.

In the high-x region, there are various theoretical corrections which must be taken into account, especially for the $Q \sim$ few GeV region now included with our relaxed kinematic cuts. Among these, the higher-twist (HT) modifications discussed in Sec. II B reduce $\chi^2/N_{\rm dof}$ by ∼3%, while the Deuteron (DEUT) modifications of Sec. II C reduce $\chi^2/N_{\rm dof}$ by ∼10%; ultimately, the combination of both (which was used in nCTEQ15HIX) reduces $\chi^2/N_{\rm dof}$ by ∼15%.

We also investigated the low-W kinematic region including the impact of extending the nuclear PDFs into the region of very large x. We used a x-rescaling technique that mimics the parton momentum shift to higher x induced by the nucleon's Fermi motion. We find that the x-rescaling provides a simple (albeit approximate) means to extend the nPDF parametrization to larger x values than currently included in our main fit, and to address data falling within the resonance region. Calculations based on a simple two-parameter x rescaling were indeed able to improve the $\chi^2/N_{\rm dof}$ for data at large-x not only inside our current cuts, but—crucially and more markedly—also for those data that are currently outside and do not contribute to the nPDF fits. This preliminary analysis suggests it may be possible to further reduce our W cut, extend the fits into the resonance region, and further improve the precision of the nPDFs at large momentum fraction x.

The structure function ratio $(F_2^A/F_2^D) \cdot (F_2^D/F_2^p)_{\rm CJ}$ (Fig. 8) displays the characteristic EMC shape, and the qualitative x dependence of the nuclear ratio in the EMC Region was found to be consistent for the range of A considered in this analysis.

Our parameter scans (Figs. 13–14) show that while the new data provide substantive constraints on the up-valence and down-valence distributions, there are nonetheless some tensions between individual datasets. This behavior is common for global PDF fits, but some of the details, such as the tension between the Fe and C datasets (Fig. 13) may warrant additional investigation.

A general feature of the new nCTEQ15HIX fit is a relative hardening of the nuclear PDFs involving shifts of select distributions toward higher x, an effect realized in the enhancement of the up-valence, down-valence, and gluon distributions, and corresponding depletion of the sea-quark PDFs, $\{\bar{u}, \bar{d}, \bar{s}\}$. This pattern is apparent when comparing nCTEQ15HIX to nCTEQ15 (Figs. 9–12). Additionally, we find similar behavior when comparing nCTEQ15HIX to other nPDFs from the literature including EPPS16, nNNPDF2.0, and TUJU19 (Fig. 19). While these variations are within the nPDF uncertainty bands, it will be interesting to see to what extent this behavior persists with improved precision.

As PDF global analyses represent a primary computational tool available to describe hadronic interactions and structure in the context of QCD, the nCTEQ15HIX nPDFs will enable a new level of precision in the exploration of nuclear dynamics and particle phenomenology. In a broader sense, these tools can validate features of the standard model to next-generation precision, aid in the search for discrepancies which may signal undiscovered phenomena, and thereby yield deeper insights into the QCD theory and the structure of hadronic matter.

ACKNOWLEDGMENTS

We are grateful to C. Leger, S. Li, W. Melnitchouk, P. Nadolsky, J. Owens, R. Ruiz, and N. Sato for valuable discussion. This work was supported by the U.S. Department of Energy under Grant No. DE-SC0010129. T. H. acknowledges support from an EIC Center Fellowship. The work of T. J. was supported by the DFG under Grant No. 396021762–TRR257. Work in Münster is supported by the DFG through the Research Training Network 2149 "Strong and weak interactions—from hadrons to dark matter" and through Project-ID 273811115–SFB 1225 "ISOQUANT." A. K. is grateful for the support of Narodowe Centrum Nauki under Grant No. 2019/34/E/ST2/00186. The work of A. A. and C. K. E. was supported by the U.S. Department of Energy Contract No. DE-AC05-06OR23177, under which Jefferson Science Associates LLC manages and operates Jefferson Lab. A. A. was furthermore supported by DOE Contract No. DE-SC0008791. J. G. M. has been supported by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The work of I. S. was supported by the French CNRS via the IN2P3 project GLUE@NLO.

APPENDIX: DATASETS USED IN FIT

Table III lists the JLab DIS data of the form F_2^A/F_2^D, Table IV lists the nCTEQ15 DIS data of the form F_2^A/F_2^D, Table V lists the nCTEQ15 DIS data of the form $F_2^A/F_2^{A'}$, and Table VI lists the nCTEQ15 DY data of the form $\sigma_{\rm DY}^{pA}/\sigma_{\rm DY}^{pA'}$.

TABLE III. This table summarizes the new JLab data used in the fit including the CLAS data from Ref. [38] and the Hall C data from Ref. [69]. The last column shows the number of data points remaining after the nCTEQ15HIX cuts $Q > 1.3$ GeV and $W > 1.7$ GeV are applied.

$\mathbf{F_2^A/F_2^D}$:				# data	
Observable	Experiment	ID	Ref.	# data	After cuts
^{208}Pb/D	CLAS	9976	[38]	25	24
^{56}Fe/D	CLAS	9977	[38]	25	24
^{27}Al/D	CLAS	9978	[38]	25	24
^{12}C/D	CLAS	9979	[38]	25	24
^{4}He/D	Hall C	9980	[69]	25	17
		9981	[69]	26	16
^{3}He/D	Hall C	9982	[69]	25	17
		9983	[69]	26	16
^{64}Cu/D	Hall C	9984	[69]	25	17
		9985	[69]	26	16
^{9}Be/D	Hall C	9986	[69]	25	17
		9987	[69]	26	16
^{197}Au/D	Hall C	9988	[69]	24	17
		9989	[69]	26	16
^{12}C/D	Hall C	9990	[69]	25	17
		9991	[69]	17	7
		9992	[69]	26	16
		9993	[69]	18	6
		9994	[69]	17	7
		9995	[69]	15	2
		9996	[69]	19	7
		9997	[69]	16	2
		9998	[69]	21	8
		9999	[69]	18	3
Total				546	336

TABLE IV. The DIS F_2^A/F_2^D datasets used in this fit. The table details the specific nuclear targets, the references, and the total number of data points in the set. The last column shows first the number of data points remaining with the nCTEQ15 cuts of $Q > 2$ GeV and $W > 3.5$ GeV, and second the number of data points remaining with the nCTEQ15HIX cuts of $Q > 1.3$ GeV and $W > 1.7$ GeV. The NMC dataset ID = 5160 is F_2^D; this is the one case which is not a ratio.

$\mathbf{F_2^A/F_2^D}$:				# data	
Observable	Experiment	ID	Ref.	# data	After cuts
D	NMC-97	5160	[86]	292	201/275
^{3}He/D	Hermes	5156	[87]	182	17/92
^{4}He/D	NMC-95,re	5124	[88]	18	12/16
	SLAC-E139	5141	[28]	18	3/17
Li/D	NMC-95	5115	[89]	24	11/15
Be /D	SLAC-E139	5138	[28]	17	3/16

(Table continued)

TABLE IV. *(Continued)*

$\mathbf{F_2^A/F_2^D}$:				# data	
Observable	Experiment	ID	Ref.	# data	After cuts
C/D	FNAL-E665-95	5125	[90]	11	3/4
	SLAC-E139	5139	[28]	7	2/7
	EMC-88	5107	[91]	9	9/9
	EMC-90	5110	[92]	9	0/2
	NMC-95	5113	[89]	24	12/15
	NMC-95,re	5114	[88]	18	12/16
N/D	Hermes	5157	[87]	175	19/92
	BCDMS-85	5103	[31]	9	9/9
Al/D	SLAC-E049	5134	[93]	18	0/18
	SLAC-E139	5136	[28]	17	3/16
Ca/D	NMC-95,re	5121	[88]	18	12/15
	FNAL-E665-95	5126	[90]	11	3/4
	SLAC-E139	5140	[28]	7	2/7
	EMC-90	5109	[92]	9	0/2
Fe/D	SLAC-E049	5131	[30]	14	2/14
	SLAC-E139	5132	[28]	23	6/22
	SLAC-E140	5133	[29]	10	0/6
	BCDMS-87	5101	[32]	10	10/10
	BCDMS-85	5102	[31]	6	6/6
Cu/D	EMC-93	5104	[66]	10	9/10
	EMC-93(chariot)	5105	[66]	9	9/9
	EMC-88	5106	[91]	9	9/9
Kr/D	Hermes	5158	[87]	167	12/84
Ag /D	SLAC-E139	5135	[28]	7	2/7
Sn /D	EMC-88	5108	[91]	8	8/8
Xe /D	FNAL-E665-92	5127	[94]	10	2/4
Au /D	SLAC-E139	5137	[28]	18	3/17
Pb /D	FNAL-E665-95	5129	[90]	11	3/4
Total				1205	414/857

TABLE V. The DIS $F_2^A/F_2^{A'}$ datasets used in this fit. We list the same details for each dataset as in Tab. IV.

$\mathbf{F_2^A/F_2^{A'}}$:				# data	
Observable	Experiment	ID	Ref.	# data	After cuts
C/Li	NMC-95,re	5123	[88]	25	7/20
Ca/Li	NMC-95,re	5122	[88]	25	7/20
Be/C	NMC-96	5112	[95]	15	14/15
Al/C	NMC-96	5111	[95]	15	14/15
Ca/C	NMC-95,re	5120	[88]	25	7/20
	NMC-96	5119	[95]	15	14/15
Fe/C	NMC-96	5143	[95]	15	14/15
Sn/C	NMC-96	5159	[96]	146	111/144
Pb/C	NMC-96	5116	[95]	15	14/15
Total				296	202/279

TABLE VI. The Drell-Yan process datasets used in this fit. We list the same details for each dataset as in Tab. IV.

$\sigma^{pA}_{DY}/\sigma^{pA'}_{DY}$:					# data
Observable	Experiment	ID	Ref.	# data	After cuts
C/D	FNAL-E772-90	5203	[97]	9	9/9
Ca/D	FNAL-E772-90	5204	[97]	9	9/9
Fe/D	FNAL-E772-90	5205	[97]	9	9/9
W/D	FNAL-E772-90	5206	[97]	9	9/9
Fe/Be	FNAL-E886-99	5201	[98]	28	28/28
W/Be	FNAL-E886-99	5202	[98]	28	28/28
Total				92	92/92

In Table III we indicate the total number of data in the sets, and the remaining number after the nCTEQ15HIX cuts. In Tables IV, V and VI we indicate the total number of data in the sets, and the remaining number after the nCTEQ15 kinematic cuts (first) as well as the nCTEQ15HIX cuts (second).

NMC F_2^D: The NMC F_2^D dataset ID = 5160 of Table IV is the one case which is not a ratio of structure functions. It serves as a cross-check of how well we describe absolute structure functions in the deuteron for which we apply a dedicated correction, instead of fitting as is done for heavier nuclei. Recall from Sec. III B that our proton parameters p_k are fixed, and for the special case of the deuteron $(A = 2)$ we construct this using isospin symmetry. We then apply the deuteron structure function modifications described in Sec. II C for the DEUT and nCTEQ15HIX fits, and find this improves $\chi^2/N_{\rm dof}$ for F_2^D by $\sim$1%.

Data files: We have provided a full list of the data used in this fit in the text file "fitted_data.txt" which is included in the Supplementary Material [99].

[1] K. Kovarik *et al.* (nCTEQ Collaboration), nCTEQ15—Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework, Phys. Rev. D **93**, 085037 (2016).

[2] M. Hirai, S. Kumano, and T. H. Nagai, Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order, Phys. Rev. C **76**, 065207 (2007).

[3] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, Global analysis of nuclear parton distributions, Phys. Rev. D **85**, 074028 (2012).

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

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

[6] R. A. Khalek, J. J. Ethier, and J. Rojo (NNPDF Collaboration), Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider, Eur. Phys. J. C **79**, 471 (2019).

[7] R. A. Khalek, J. J. Ethier, J. Rojo, and G. van Weelden, nNNPDF2.0: Quark flavor separation in nuclei from LHC data, J. High Energy Phys. 09 (2020) 183.

[8] J. J. Ethier and E. R. Nocera, Parton distributions in nucleons and nuclei, Annu. Rev. Nucl. Part. Sci. **70**, 43 (2020).

[9] H. Paukkunen and P. Zurita, Can we fit nuclear PDFs with the high-x CLAS data?, Eur. Phys. J. C **80**, 381 (2020).

[10] A. Kusina, F. Lyonnet, D. B. Clark, E. Godat, T. Jezo, K. Kovarik, F. I. Olness, I. Schienbein, and J. Y. Yu, Vector boson production in pPb and PbPb collisions at the LHC and its impact on nCTEQ15 PDFs, Eur. Phys. J. C **77**, 488 (2017).

[11] H. Khanpour, M. Soleymaninia, S. A. Tehrani, H. Spiesberger, and V. Guzey, Nuclear parton distribution functions with uncertainties in the general mass variable flavor number scheme, arXiv:2010.00555 [Phys. Rev. D (to be published)].

[12] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan, Implications of CTEQ global analysis for collider observables, Phys. Rev. D **78**, 013004 (2008).

[13] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali (NNPDF Collaboration), Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering, Nucl. Phys. **B823**, 195 (2009).

[14] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions from high-precision collider data, Eur. Phys. J. C **77**, 663 (2017).

[15] S. Alekhin, J. Blümlein, and S. Moch, Strange sea determination from collider data, Phys. Lett. B **777**, 134 (2018).

[16] H.-W. Lin *et al.*, Parton distributions and lattice QCD calculations: A community white paper, Prog. Part. Nucl. Phys. **100**, 107 (2018).

[17] J. Gao, L. Harland-Lang, and J. Rojo, The structure of the proton in the LHC precision era, Phys. Rep. **742**, 1 (2018).

[18] R. A. Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, Towards ultimate parton distributions at the high-luminosity LHC, Eur. Phys. J. C **78**, 962 (2018).

[19] K. Kovařík, P. M. Nadolsky, and D. E. Soper, Hadronic structure in high-energy collisions, Rev. Mod. Phys. **92**, 045003 (2020).

[20] T.-J. Hou *et al.*, New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC, Phys. Rev. D **103**, 014013 (2021).

[21] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk (JAM Collaboration), Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions, Phys. Rev. D **101**, 074020 (2020).

[22] 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**, 204 (2015).

[23] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. Nathvani, Updates of PDFs using the MMHT framework, *Proc. Sci.*, DIS2019 (2019) 036 [arXiv:1907.08147].

[24] M. Constantinou *et al.*, Parton distributions and lattice QCD calculations: Toward 3D structure, arXiv:2006.08636.

[25] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo, and M. Ubiali, The strangest proton?, Eur. Phys. J. C **80**, 1168 (2020).

[26] R. A. Khalek *et al.*, Science requirements and detector concepts for the electron-ion collider: EIC yellow report, arXiv:2103.05419.

[27] J. J. Aubert *et al.* (European Muon Collaboration), The ratio of the nucleon structure functions $F2_n$ for iron and deuterium, Phys. Lett. **123B**, 275 (1983).

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

[29] S. Dasu *et al.*, Measurement of kinematic and nuclear dependence of R = sigma-L / sigma-t in deep inelastic electron scattering, Phys. Rev. D **49**, 5641 (1994).

[30] A. Bodek *et al.*, Electron Scattering from Nuclear Targets and Quark Distributions in Nuclei, Phys. Rev. Lett. **50**, 1431 (1983).

[31] G. Bari *et al.* (BCDMS Collaboration), A measurement of nuclear effects in deep inelastic muon scattering on deuterium, nitrogen and iron targets, Phys. Lett. **163B**, 282 (1985).

[32] A. C. Benvenuti *et al.* (BCDMS Collaboration), Nuclear effects in deep inelastic muon scattering on deuterium and iron targets, Phys. Lett. B **189**, 483 (1987).

[33] S. Malace, D. Gaskell, D. W. Higinbotham, and I. Cloet, The challenge of the EMC effect: Existing data and future directions, Int. J. Mod. Phys. E **23**, 1430013 (2014).

[34] E. P. Segarra, A. Schmidt, T. Kutz, D. W. Higinbotham, E. Piasetzky, M. Strikman, L. B. Weinstein, and O. Hen, Neutron Valence Structure from Nuclear Deep Inelastic Scattering, Phys. Rev. Lett. **124**, 092002 (2020).

[35] O. Hen, G. A. Miller, E. Piasetzky, and L. B. Weinstein, Nucleon-nucleon correlations, short-lived excitations, and the quarks within, Rev. Mod. Phys. **89**, 045002 (2017).

[36] O. Hen, D. W. Higinbotham, G. A. Miller, E. Piasetzky, and L. B. Weinstein, The EMC effect and high momentum nucleons in nuclei, Int. J. Mod. Phys. E **22**, 1330017 (2013).

[37] L. B. Weinstein, E. Piasetzky, D. W. Higinbotham, J. Gomez, O. Hen, and R. Shneor, Short Range Correlations and the EMC Effect, Phys. Rev. Lett. **106**, 052301 (2011).

[38] B. Schmookler *et al.* (CLAS Collaboration), Modified structure of protons and neutrons in correlated pairs, Nature (London) **566**, 354 (2019).

[39] O. Hen *et al.*, In medium proton structure functions, SRC, and the EMC effect, (JLab PAC Proposal, 2015), https://misportal.jlab.org/pacProposals/proposals/1159/attachments/93263/E12-11-003A.pdf.

[40] O. Hen *et al.*, Jefferson Lab Experiments, Reports No. E12-11-107 and No. E12-11-003A.

[41] J. Arrington, R. Ent, C. E. Keppel, J. Mammei, and I. Niculescu, Low Q scaling, duality, and the EMC effect, Phys. Rev. C **73**, 035205 (2006).

[42] I. Schienbein *et al.*, A review of target mass corrections, J. Phys. G **35**, 053101 (2008).

[43] M. A. G. Aivazis, F. I. Olness, and W.-K. Tung, Leptoproduction of heavy quarks. 1. General formalism and kinematics of charged current and neutral current production processes, Phys. Rev. D **50**, 3085 (1994).

[44] M. A. G. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, Leptoproduction of heavy quarks. 2. A Unified QCD formulation of charged and neutral current processes from fixed target to collider energies, Phys. Rev. D **50**, 3102 (1994).

[45] I. Schienbein *et al.* (nCTEQ Collaboration), Target mass corrections in lepton-nucleus DIS revisited, Report No. SMU-HEP-21-01 (to be published).

[46] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato, Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data, Phys. Rev. D **93**, 114017 (2016).

[47] J.-W. Qiu and G. F. Sterman, QCD and rescattering in nuclear targets, Int. J. Mod. Phys. E **12**, 149 (2003).

[48] J. G. Morfin, H. Weerts, A. G. Frodesen, D. Bertrand, G. Bertrand-Coremans, M. Dewit, P. Vilain, R. Blaes, B. Escoubes, P. Petitjean, *et al.*, Interpretation of the nucleon structure functions measured in the gargamelle SPS neutrino/anti-neutrino experiment, Phys. Lett. B **107**, 450 (1981).

[49] P. C. Bosetti *et al.*, Analysis of nucleon structure functions in CERN bubble chamber neutrino experiments, Nucl. Phys. **B142**, 1 (1978).

[50] K. Varvell *et al.*, Measurement of the structure functions F2 and Xf3 and comparison with QCD predictions including kinematical and dynamical higher twist effects, Z. Phys. C **36**, 1 (1987).

[51] S. P. Luttrell, S. Wada, and B. R. Webber, The wilson coefficient functions of four quark operators and the four quark process in deep inelastic scattering, Nucl. Phys **B188**, 232 (1981).

[52] E. V. Shuryak and A. I. Vainshtein, Theory of power corrections to deep inelastic scattering in quantum chromodynamics. 1. Q^{**2} effects, Nucl. Phys. **B199**, 451 (1982).

[53] H. Szumila-Vance, C. Keppel, S. Escalante, and N. Kalantarians, Examining the EMC effect using the F_2^n neutron structure function, Phys. Rev. C **103**, 015201 (2021).

[54] J. F. Owens, J. Huston, C. E. Keppel, S. Kuhlmann, J. G. Morfin, F. Olness, J. Pumplin, and D. Stump, The impact of

new neutrino DIS and Drell-Yan data on large-x parton distributions, Phys. Rev. D **75**, 054030 (2007).
[55] J. F. Owens, A. Accardi, and W. Melnitchouk, Global parton distributions with nuclear and finite-Q^2 corrections, Phys. Rev. D **87**, 094012 (2013).
[56] S. I. Alekhin, S. A. Kulagin, and R. Petti, Nuclear effects in the deuteron and constraints on the d/u ratio, Phys. Rev. D **96**, 054005 (2017).
[57] R. D. Ball, E. R. Nocera, and R. L. Pearson, Deuteron uncertainties in the determination of proton PDFs, Eur. Phys. J. C **81**, 37 (2021).
[58] A. Accardi, T. J. Hobbs, X. Jing, and P. M. Nadolsky, Deuterium scattering experiments in CTEQ global QCD analyses: a comparative investigation, arXiv:2102.01107
[59] A. Accardi, PDFs from nucleons to nuclei, *Proc. Sci.*, DIS2015 (2015) 001 [arXiv:1602.02035].
[60] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, An accurate nucleon-nucleon potential with charge independence breaking, Phys. Rev. C **51**, 38 (1995).
[61] S. Veerasamy and W. N. Polyzou, A momentum-space Argonne V18 interaction, Phys. Rev. C **84**, 034003 (2011).
[62] W. Melnitchouk and A. W. Thomas, Shadowing in deuterium, Phys. Rev. D **47**, 3783 (1993).
[63] S. Tkachenko *et al.* (CLAS Collaboration), Measurement of the structure function of the nearly free neutron using spectator tagging in inelastic ^{2}H(e, e'p)X scattering with CLAS, Phys. Rev. C **89**, 045206 (2014); Addendum Phys. Rev. C **90**, 059901(A) (2014).
[64] N. Baillie *et al.* (CLAS Collaboration), Measurement of the Neutron F2 Structure Function via Spectator Tagging with CLAS, Phys. Rev. Lett. **108**, 142001 (2012); Erratum, Phys. Rev. Lett. **108**, 199902 (2012).
[65] P. J. Ehlers, A. Accardi, L. T. Brady, and W. Melnitchouk, Nuclear effects in the proton-deuteron Drell-Yan process, Phys. Rev. D **90**, 014010 (2014).
[66] J. Ashman *et al.* (European Muon Collaboration), A measurement of the ratio of the nucleon structure function in copper and deuterium, Z. Phys. C **57**, 211 (1993).
[67] S. Adler *et al.* (PHENIX Collaboration), Centrality Dependence of π^0 and η Production at Large Transverse Momentum in $\sqrt{s_{NN}} = 200$-GeV d + Au Collisions, Phys. Rev. Lett. **98**, 172302 (2007).
[68] B. I. Abelev *et al.* (STAR Collaboration), Inclusive π^0, η, and direct photon production at high transverse momentum in $p + p$ and d + Au collisions at $\sqrt{s_{NN}} = 200$ GeV, Phys. Rev. C **81**, 064904 (2010).
[69] J. Seely *et al.*, New Measurements of the EMC Effect in Very Light Nuclei, Phys. Rev. Lett. **103**, 202301 (2009).
[70] D. Higinbotham, W. Melnitchouk, and A. Thomas, New insights into the structure of matter: The first decade of science at Jefferson Lab, J. Phys. Conf. Ser. **299**, 012010 (2011).
[71] A. Kusina *et al.* (nCTEQ Collaboration), Impact of LHC vector boson production in heavy ion collisions on strange PDFs, Eur. Phys. J. C **80**, 968 (2020).
[72] G. P. Salam and J. Rojo, A higher order perturbative parton evolution toolkit (HOPPET), Comput. Phys. Commun. **180**, 120 (2009).
[73] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID project, Eur. Phys. J. C **66**, 503 (2010).
[74] J. M. Campbell, R. K. Ellis, and W. T. Giele, A multi-threaded version of MCFM, Eur. Phys. J. C **75**, 246 (2015).
[75] Y. Li and F. Petriello, Combining QCD and electroweak corrections to dilepton production in FEWZ, Phys. Rev. D **86**, 094034 (2012).
[76] S. Alekhin *et al.*, HERAFitter, Eur. Phys. J. C **75**, 304 (2015).
[77] M. S. Athar and J. G. Morfín, Neutrino(antineutrino)–nucleus interactions in the shallow- and deep-inelastic scattering regions, J. Phys. G **48**, 034001 (2021).
[78] A. J. Freese, W. Cosyn, and M. M. Sargsian, QCD evolution of superfast quarks, Phys. Rev. D **99**, 114019 (2019).
[79] S. A. Kulagin, Deep inelastic scattering on nuclei: Impulse approximation and mesonic corrections, Nucl. Phys. **A500**, 653 (1989).
[80] S. A. Kulagin, G. Piller, and W. Weise, Shadowing, binding and off-shell effects in nuclear deep inelastic scattering, Phys. Rev. C **50**, 1154 (1994).
[81] S. A. Kulagin, Nuclear effects in F(3) structure function for finite and asymptotic Q**2, Nucl. Phys. **A640**, 435 (1998).
[82] S. A. Kulagin and R. Petti, Global study of nuclear structure functions, Nucl. Phys. **A765**, 126 (2006).
[83] S. A. Kulagin and R. Petti, Neutrino inelastic scattering off nuclei, Phys. Rev. D **76**, 094023 (2007).
[84] S. A. Kulagin and R. Petti, Structure functions for light nuclei, Phys. Rev. C **82**, 054614 (2010).
[85] C. C. Degli Atti and S. Simula, Realistic model of the nucleon spectral function in few and many nucleon systems, Phys. Rev. C **53**, 1689 (1996).
[86] M. Arneodo *et al.* (New Muon Collaboration), Measurement of the proton and deuteron structure functions, F2(p) and F2(d), and of the ratio sigma-L/sigma-T, Nucl. Phys. **B483**, 3 (1997).
[87] A. Airapetian *et al.* (HERMES Collaboration), Measurement of $R = \sigma(L)/\sigma(T)$ in deep inelastic scattering on nuclei, arXiv:hep-ex/0210068.
[88] P. Amaudruz *et al.* (New Muon Collaboration), A reevaluation of the nuclear structure function ratios for D, He, Li-6, C and Ca, Nucl. Phys. **B441**, 3 (1995).
[89] M. Arneodo *et al.* (New Muon Collaboration), The structure function ratios F2(Li)/F2(D) and F2(C)/F2(D) at small x, Nucl. Phys. **B441**, 12 (1995).
[90] M. R. Adams *et al.* (E665 Collaboration), Shadowing in inelastic scattering of muons on carbon, calcium and lead at low x(Bj), Z. Phys. C **67**, 403 (1995).
[91] J. Ashman *et al.* (European Muon Collaboration), Measurement of the ratios of deep inelastic muon—nucleus cross-sections on various nuclei compared to deuterium, Phys. Lett. B **202**, 603 (1988).
[92] M. Arneodo *et al.* (European Muon Collaboration), Measurements of the nucleon structure function in the range $0.002 < x < 0.17$ and $0.2 < Q^2 < 8$ GeV^2 in deuterium, carbon and calcium, Nucl. Phys. **B333**, 1 (1990).
[93] A. Bodek *et al.*, A Comparison of the Deep Inelastic Structure Functions of Deuterium and Aluminum Nuclei, Phys. Rev. Lett. **51**, 534 (1983).

[94] M. R. Adams *et al.* (E665 Collaboration), Saturation of Shadowing at Very Low x_{BJ}, Phys. Rev. Lett. **68**, 3266 (1992).
[95] M. Arneodo *et al.* (New Muon Collaboration), The A dependence of the nuclear structure function ratios, Nucl. Phys. **B481**, 3 (1996).
[96] M. Arneodo *et al.* (New Muon Collaboration), The Q^2 dependence of the structure function ratio F_2^{Sn}/F_2^C and the difference $R^{Sn} - R^C$ in deep inelastic muon scattering, Nucl. Phys. **B481**, 23 (1996).
[97] D. M. Alde *et al.*, Nuclear Dependence of Dimuon Production at 800-GeV. FNAL-772 Experiment, Phys. Rev. Lett. **64**, 2479 (1990).
[98] M. A. Vasilev *et al.* (NuSea Collaboration), Parton Energy Loss Limits and Shadowing in Drell-Yan Dimuon Production, Phys. Rev. Lett. **83**, 2304 (1999).
[99] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevD.103.114015 for an ASCII file containing the data used in this fit.