



Can we fit nuclear PDFs with the high- x CLAS data?

Hannu Paukkunen^{1,2,a}, Pia Zurita^{3,b}

¹ Department of Physics, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland

² Helsinki Institute of Physics, University of Helsinki, P.O. Box 64, 00014 Helsinki, Finland

³ Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany

Received: 9 March 2020 / Accepted: 24 April 2020

© The Author(s) 2020

Abstract Nuclear parton distribution functions (nuclear PDFs) are non-perturbative objects that encode the partonic behaviour of bound nucleons. To avoid potential higher-twist contributions, the data probing the high- x end of nuclear PDFs are sometimes left out from the global extractions despite their potential to constrain the fit parameters. In the present work we focus on the kinematic corner covered by the new high- x data measured by the CLAS/JLab collaboration. By using the Hessian re-weighting technique, we are able to quantitatively test the compatibility of these data with globally analyzed nuclear PDFs and explore the expected impact on the valence-quark distributions at high x . We find that the data are in a good agreement with the EPPS16 and nCTEQ15 nuclear PDFs whereas they disagree with TuJu19. The implications on flavour separation, higher-twist contributions and models of EMC effect are discussed.

1 Introduction

The nuclear parton distribution functions (nuclear PDFs) [1,2] quantifying the structure of quarks and gluons in bound nucleons constitute an indispensable ingredient in precision calculations for processes at high interaction scales $Q^2 \gg \Lambda_{\text{QCD}}^2$ in high-energy colliders like the Large Hadron Collider (LHC). Based on the collinear factorization theorem [3], nuclear PDFs are believed to be process independent and the scale dependence to follow the usual linear DGLAP evolution [4–10]. These assumptions have been observed to be consistent with experimental data ranging from deeply inelastic scattering (DIS) to heavy-ion collisions. For example, although in high-energy lead-lead collisions there is evidence for the formation of a state that effectively behaves as a strongly-interacting liquid, the electroweak observables [11–14] are consistent with the nuclear PDF predictions. More-

over, the linear DGLAP evolution in proton-lead collisions has been verified down to $x \sim 10^{-5}$ at low Q^2 through heavy-quark production [15], with no evidence of a breakdown.

It is well known that the PDFs are best constrained through DIS experiments. Indeed, thanks to the HERA data [16], the free-proton PDFs are quite well determined in a wide kinematic window. The regimes where one has to rely on extrapolations are limited to the very small x ($x < 10^{-5}$) and the high- x regions. The former, due to not having been explored in electron-proton experiments; the latter due to the imposition of kinematic cuts on Q^2 and the final-state invariant mass W to avoid potentially large higher-twist contributions such as target mass corrections.¹ However, there has not yet been an experiment equivalent to HERA with nuclear beams – only fixed-target DIS data spanning a rather limited region of the kinematic space (though covering a variety of nuclei) are available. With no “nuclear HERA data” the nuclear PDFs still suffer from large uncertainties and e.g. the flavour separation is only poorly known. Given the fact that high-energy nuclear DIS at the Electron-Ion Collider (EIC) [18] or at the planned LHeC/FCC-eh [19] are at least a decade away, the community has generally sought to improve the situation by using the LHC proton-lead data as new constraints in the global analyses.

An additional possibility is to aim at a more complete use of the already available high- x DIS data. Imitating the typical free-proton fits some of the nuclear-PDF analyses also set stringent cuts on Q^2 and W . Given the low center-of-mass energies of the available fixed-target data, a significant fraction of the data get easily cut away. Lowering the minimum value of Q^2 one reaches lower values in x , while lowering the cut in the final-state invariant mass W the high- x low- Q^2 data enter the fits. In the present paper we concentrate on this latter regime by studying the compatibility and impact of the very precise DIS data measured recently by the CLAS

^a e-mail: hannu.paukkunen@jyu.fi (corresponding author)

^b e-mail: maria.zurita@ur.de

¹ These cuts are routinely applied in proton PDF fits, though their relaxation has been explored, e.g. by the CTEQ-JLab collaboration [17].

collaboration [20] by using recent sets of nuclear PDFs at a next-to-leading order (NLO) accuracy. The data were taken in the high- x region ($0.2 < x < 0.6$) where the so-called EMC effect [21, 22] occurs. On one hand, the approach based on nuclear PDFs is phenomenological in the sense that one does not address the underlying microscopic dynamics of the nuclear effects. Based solely on the framework of QCD and collinear factorization, the predictions from nuclear PDFs aim to be model independent.² On the other hand, in the same x range but at higher Q^2 there are other lepton-nucleus DIS data e.g. from SLAC/NMC collaborations [23–25] and also neutrino-nucleus DIS data from e.g. the CHORUS collaboration [26]. In addition, recent CMS dijet data [27] have been found to be sensitive to the valence-quark EMC effect. Since the Q^2 dependence of the EMC effect in global analyses of nuclear PDFs is fully dynamical, dictated by the DGLAP evolution, an ability (or disability) to describe all these data with an universal initial condition for the Q^2 dependence will (1) quantitatively address the importance of possible higher-twist $\sim Q^{-2n}$ contributions and (2) place restrictions on the possible origin of the EMC effect. In particular, the current nuclear PDF analyses all assume that the strength of the nuclear effects scale as a function of nuclear mass number A , although e.g. models for the EMC effect based on short-range correlations [28] would suggest also an isospin dependence. The current global fits of nuclear PDFs, however, have not found support for an existence of such a component. This only makes the new JLab/CLAS data more welcome and constitute an interesting test bench for the nuclear PDFs. A reliable understanding of the nuclear effects would eventually facilitate unfolding the isospin asymmetry of nucleons using nuclear data [29].

The rest of the document is organised as follows. In Sect. 2 we introduce the framework of our study, including the nuclear PDF sets, details regarding the calculation, target mass corrections and the method employed. In Sect. 3 we discuss our results including the potential of the data to further improve our knowledge of the valence distributions at high $-x$. Finally we summarise our results in Sect. 4.

2 Framework

2.1 Nuclear PDFs at high x

In our present study, we will utilize three modern sets of nuclear PDFs: EPPS16 [30], nCTEQ15 [31] and TuJu19 [32]. All these three sets involve the valence-quark flavour separation and thereby better reflect the prevailing uncer-

tainties at large x . We refrain here from using sets with no flavour separation (e.g. EPS09 [33], DSSZ12 [34], KA15 [35], nNNPDF1.0 [36]). As it is customary in the case of nuclear PDFs, we will discuss the behaviour of the nuclear valence distributions in terms of certain ratios which better reflect the relevant features of nuclear PDFs. We define here the valence-quark nuclear modifications $R_{u_V}^A(x, Q^2)$ and $R_{d_V}^A(x, Q^2)$, as the total up/down valence distribution in a nucleus A with Z protons and N neutrons, divided by the same distribution but with no nuclear effects in the PDFs,

$$R_{u_V}^{\text{Pb}}(x, Q^2) \equiv \frac{u_V^A(x, Q^2)}{Zu_V^{\text{p}}(x, Q^2) + Nd_V^{\text{p}}(x, Q^2)}, \quad (1)$$

$$R_{d_V}^{\text{Pb}}(x, Q^2) \equiv \frac{d_V^A(x, Q^2)}{Zd_V^{\text{p}}(x, Q^2) + Nu_V^{\text{p}}(x, Q^2)}. \quad (2)$$

Here $u_V^{\text{p}}(x, Q^2)$ and $d_V^{\text{p}}(x, Q^2)$ denote the free-proton valence-quark PDFs. When forming the ratios the proton PDF used is always the one taken as baseline in the corresponding nuclear-PDFs analysis (e.g. CT14NLO for EPPS16). These ratios and their nominal uncertainty bands (90% confidence level for EPPS16 and nCTEQ15) are plotted in Fig. 1 for the lead (Pb) nucleus. In general, there seems to be a fair agreement between different parametrizations, though the shapes and widths of the uncertainty bands differ from each other as a consequence of different data inputs, PDF parametrizations and error tolerances (see Table 1 ahead). In particular, the nCTEQ15 uncertainty for $R_{u_V}^{\text{Pb}}(x, Q^2)$ is clearly larger than those of EPPS16 and TuJu19. This is presumably due to the facts that the nCTEQ15 analysis (1) used isoscalar DIS data which skews the flavour separation and (2) did not include neutrino DIS data which e.g. in the EPPS16 analysis clearly improved the flavour separation—presumably so also in TuJu19. Indeed, the combination of PDFs probed in neutral-current DIS is of the form (suppressing the x and Q^2 arguments),

$$\begin{aligned} 4u_V^A + d_V^A &= 4 \left[ZR_{u_V}^{p/A} u_V^{\text{p}} + NR_{d_V}^{p/A} d_V^{\text{p}} \right] \\ &\quad + \left[ZR_{d_V}^{p/A} d_V^{\text{p}} + NR_{u_V}^{p/A} u_V^{\text{p}} \right] \\ &\propto R_{u_V}^{p/A} + R_{d_V}^{p/A} \times \frac{d_V^{\text{p}}}{u_V^{\text{p}}} \frac{Z + 4N}{4Z + N} \\ &\xrightarrow{N=Z} R_{u_V}^{p/A} + R_{d_V}^{p/A} \times \frac{d_V^{\text{p}}}{u_V^{\text{p}}}, \end{aligned} \quad (3)$$

where $R_{u_V}^{p/A}$ and $R_{d_V}^{p/A}$ are nuclear modifications of the bound protons (these are what EPPS16 parametrized and also effectively nCTEQ15 and TuJu19 by assuming a fixed proton baseline PDF). Since the valence up is roughly twice-thrice the valence down at high- x , we have $d_V^{\text{p}}/u_V^{\text{p}} \ll 1$ at large x , and it is clear that the relative weight of $R_{u_V}^{p/A}$ in the equation

² In practice, fitting PDFs with a finite amount of free parameters induces a parametrization bias which can easily become the dominant uncertainty in regions with no data constraints.

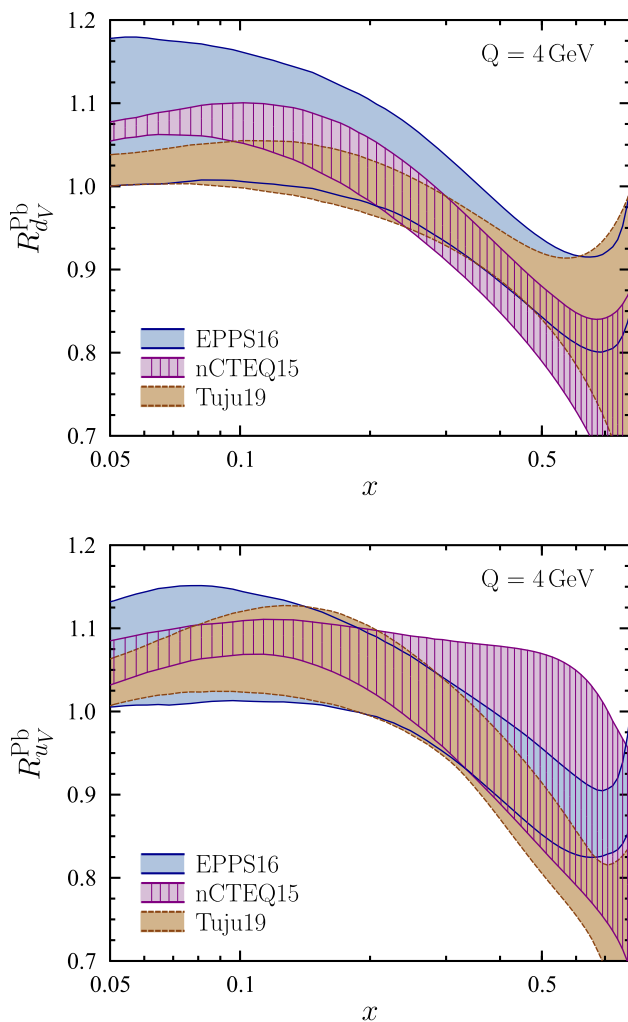


Fig. 1 Comparison of the valence-quark nuclear modifications as encoded in the EPPS16 (blue bands), nCTEQ15 (purple bands with hatching) and TuJu19 (brown bands) parametrizations at $Q = 4$ GeV

Table 1 The values of χ^2/N_{data} for the original central PDF set and for the central set after re-weighting analysis. The induced penalties and original tolerance criteria $\Delta\chi^2$ are indicated as well

PDF set	$\frac{\chi^2_{\text{orig.}}}{N_{\text{data}}}$	$\frac{\chi^2_{\text{rew.}}}{N_{\text{data}}}$	PDF penalty	$\Delta\chi^2$
EPPS16	0.93	0.78	5.9	52
nCTEQ15	0.98	0.72	3.9	35
TuJu19	4.4	1.5	72	50

above is larger and therefore better constrained by the fit. Since the ratio d_V^p/u_V^p is not constant as a function of x , the linear combination of $R_{uV}^{p/A}$ and $R_{dV}^{p/A}$ does not remain constant and through the assumed form of the parametrization one can constrain them separately to some extent even with isoscalar nuclei ($N = Z = A/2$). On the other hand we can

write Eqs. (1) and (2) also as

$$R_{uV}^{\text{Pb}} = \frac{Z R_{uV}^{p/\text{Pb}} u_V^p + N R_{dV}^{p/\text{Pb}} d_V^p}{Z u_V^p + N d_V^p}, \quad (4)$$

$$R_{dV}^{\text{Pb}} = \frac{Z R_{dV}^{p/\text{Pb}} d_V^p + N R_{uV}^{p/\text{Pb}} u_V^p}{Z d_V^p + N u_V^p}, \quad (5)$$

and because the better constrained component $R_{uV}^{p/\text{Pb}}$ has a larger weight in R_{dV}^{Pb} ($N = 126$ and $Z = 82$ for ^{208}Pb) it follows that R_{dV}^{Pb} is better determined than R_{uV}^{Pb} if only isoscalar neutral-current data is used as a constraint. This is clearly what we see in Fig. 1 for nCTEQ15. The valence-quark flavour separation in EPPS16 and TuJu19 is better constrained for using non-isoscalar data and neutrino-nucleus DIS. In Sect. 3 we will discuss how the features seen here are reflected in the predictions in physical DIS cross sections.

2.2 DIS cross sections and mass scheme

The theoretical predictions for the DIS cross-sections were computed at NLO accuracy using the simplified Aivazis-Collins-Olness-Tung (SACOT) variant of the general-mass variable-flavour-number scheme with the so-called χ rescaling [37,38]. This coincides with the scheme used in the EPPS16, nCTEQ15 and TuJu19 analyses.³ The choice of scheme is not particularly critical, however, given that the heavy-quark production does not play a significant role in the inclusive cross sections at $x \geq 0.2$, and that we are here mainly interested in ratios of cross sections,

$$\frac{d^2\sigma(A)}{dx dQ^2} \bigg/ \frac{d^2\sigma(D)}{dx dQ^2}, \quad (6)$$

where D refers to deuteron. We have verified the scheme independence of our results by comparing our calculations to the ones in the Thorne-Roberts scheme [40]. In principle, the ratios of Eq. (6) can also carry some dependence on the proton PDF as the linear combination of u_V^p and d_V^p are different for different A , see Eq. (3). However, we have explicitly checked that the proton PDF uncertainties largely cancel in these ratios (as long as the same free-proton PDF is used in the numerator and denominator). Thus, also in this respect the predictions are theoretically robust. For consistency, we neglect the nuclear effects in deuteron ($A = 2$) in the case of EPPS16 and nCTEQ15. The TuJu19 parametrization, however, extends down to $A = 2$ and we are able to address the role of deuteron nuclear effects.

³ According to Ref. [39], the FONLL-A scheme used in TuJu19 is equivalent with SACOT.

2.3 Target-mass corrections

When approaching the large- x and low- Q^2 limit, the DIS cross sections will eventually become sensitive to $1/Q^{2n}$ -type power corrections originating from beyond-leading-twist contributions (not determined by PDFs) and finite nucleon mass. When W is low, one also has to eventually consider effects such as nucleon resonances and in the case of nuclei, the fact that bound nucleons can carry more momentum than the average momentum per nucleon (i.e. $0 < x_{\text{nucleon}} < A$). In our calculations we account for the dominant part of target-mass corrections (TMCs) – an effect that is particularly relevant at low Q^2 and high x . In the DIS limit [41], $Q^2, P \cdot q \rightarrow \infty$ (\equiv massless quarks and nucleons), the usual Bjorken variable $x \equiv Q^2/2P \cdot q$ gives the fraction of light-cone momentum of the target nucleon (P) carried by the hit parton.⁴ However at low/moderate virtualities this identification is no longer necessarily accurate. Instead, the parton light-cone momentum fraction is given by the so-called Nachtmann variable ξ :

$$\xi = \frac{2x}{1 + \sqrt{1 + 4x^2 M^2/Q^2}}, \quad (7)$$

where M is the nucleon mass. The difference between x and ξ has to be considered in the calculation of the structure functions. In the present work we use the prescription of Ref. [42],

$$F_2^{\text{TMC}}(x, Q^2) = \frac{x^2}{\xi^2(1 + 4x^2 M^2/Q^2)^{3/2}} F_2^{\text{LT}}(\xi, Q^2), \quad (8)$$

$$F_L^{\text{TMC}}(x, Q^2) = \frac{x^2}{\xi^2(1 + 4x^2 M^2/Q^2)^{1/2}} F_L^{\text{LT}}(\xi, Q^2), \quad (9)$$

where F_i^{LT} refer to leading-twist structure functions i.e. those calculated with PDFs. We neglect the corrections suppressed by additional powers of xM^2/Q^2 whose effect we have found negligible for the cross-section ratios. In practice, since

$$\frac{d^2\sigma(A)}{dx dQ^2} \bigg/ \frac{d^2\sigma(D)}{dx dQ^2} \approx \frac{F_2^A(x, Q^2)}{F_2^D(x, Q^2)}, \quad (10)$$

to a very good approximation, the principal effect of TMCs in the cross section ratios is a shift in the probed value of the momentum fraction. We note that in the prescription used here $F_i^{\text{TMC}}(x = 1, Q^2) \neq 0$ which can be avoided in an alternative approach [43]. However, as now $x < 0.6$ this is not yet an issue. For a review of TMCs corrections we refer the reader to [44]. The possible relevance of the other higher-twist contributions is addressed here by investigating

the compatibility of the data with global fits of nuclear PDF constrained by data at higher Q^2 .

2.4 The CLAS data

In this study we use the high-precision data measured by the CLAS collaboration [20] and assess their potential in constraining the nuclear valence-quark distributions, particularly in the high- x region. These data are ratios of inclusive electron-ion (e^-A) DIS cross-sections with respect to the same observable in electron-deuteron collisions. They cover the kinematic region $0.2 < x < 0.6$, with the average Q^2 spanning the range $1.62 < Q^2/\text{GeV}^2 < 3.37$, and $W \geq 1.8$ GeV, which is just above the resonance region. In typical PDF fits these data would be discarded due to smallness of Q^2 and W (e.g. nCTEQ15 requires $Q^2 > 4 \text{ GeV}^2$ and $W > 3.5 \text{ GeV}$) but e.g. in the EPPS16 analysis no separate cut on W was imposed. There are 26 data points per target and four different nuclear targets: carbon (C), aluminium (Al), iron (Fe) and lead (Pb). In total the number of data points is thus $N_{\text{data}} = 104$. We note that similar JLab data exist also for very light nuclei [45].

2.5 Hessian re-weighting and definition of χ^2

The impact study was done by means of the Hessian re-weighting technique [46–48] in which the sensitivity of the data χ^2 to the PDF error sets is translated into new PDF errors. If the variation remains much smaller than the global tolerance criterion $\Delta\chi^2$ the new data are not bound to have a significant impact, and vice versa. Re-weighting methods have become very popular in recent years to provide fast estimations of the consistency and impact of new data on existing PDFs, and play a key role in studies related to future experiments. For further discussions and validations of the Hessian-PDF re-weighting technique, see e.g. Refs. [49,50].

The underlying idea is to simulate a global analysis by defining a global χ^2 function

$$\chi_{\text{global}}^2 \equiv \Delta\chi^2 \sum_k z_k^2 + \chi_{\text{new}}^2(\mathbf{z}), \quad (11)$$

where the first term approximates all the data included in a given PDF analysis. The Hessian error sets distributed along published PDFs effectively parametrize the PDFs as a function of the coordinates z_k and can be used to approximate the latter term. The coordinates z_k that minimize χ_{global}^2 then define a new set of PDFs. The first term in Eq. (11) at the minimum is what we call “PDF penalty” in what follows. Observing a penalty clearly smaller than the tolerance $\Delta\chi^2$ is a sign that the new data can be included in the global analysis without inconsistencies appearing. A penalty larger than

⁴ The variable q marks the momentum of the virtual photon.

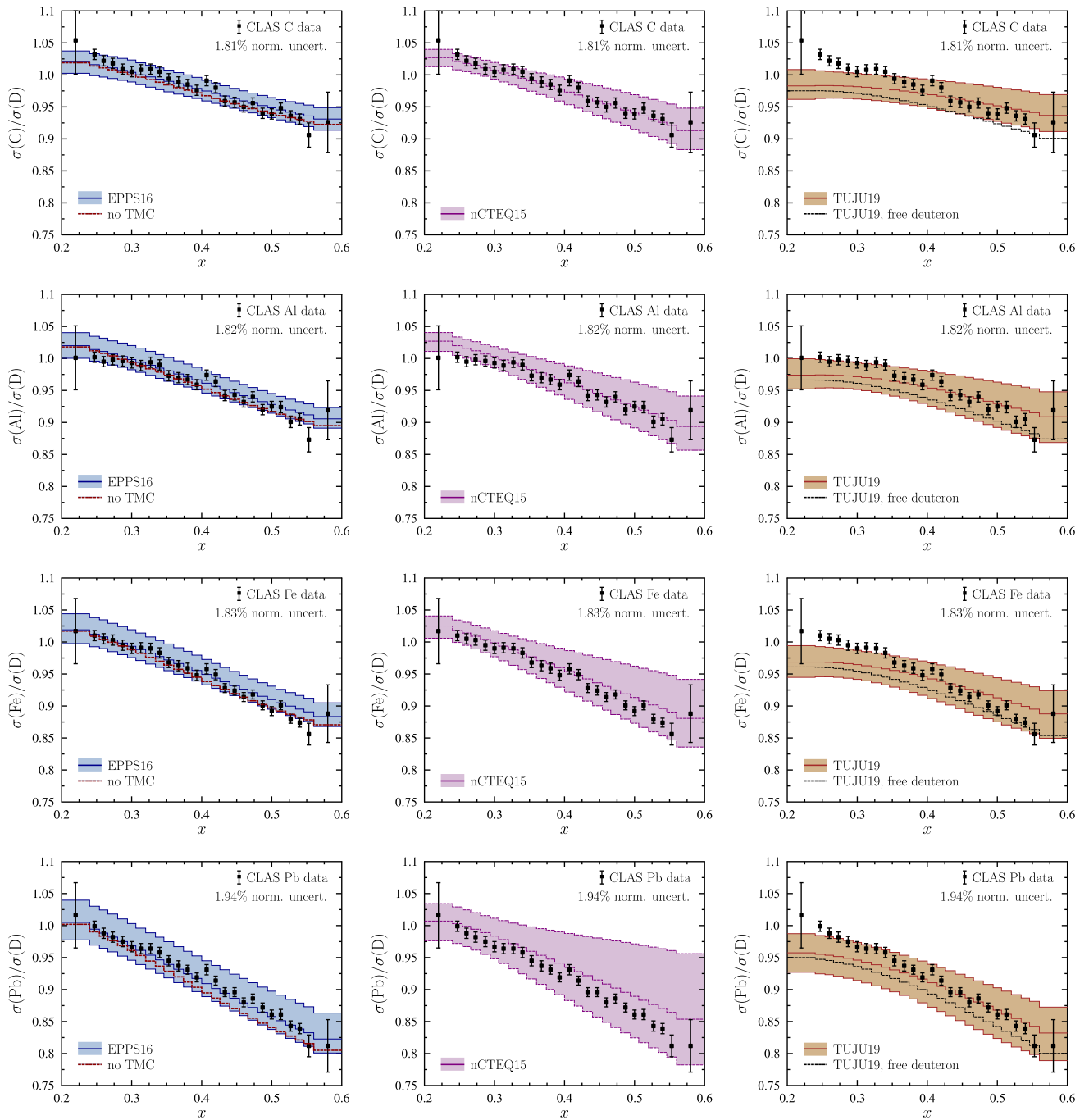


Fig. 2 The CLAS data compared with the nuclear-PDF predictions. Left panels: EPPS16 with (solid line) and without (dashed line) TMCs. Center panels: nCTEQ15. Right panels: TUJU19 with (solid line) and

without (dashed line) nuclear effects in deuteron. The normalization uncertainties are not included in the data error bars

$\Delta\chi^2$, in turn, signifies a tension in the PDF fit between the new data and some other data in the original analysis.

We write our χ_{new}^2 merit-of-figure function as

$$\chi_{\text{new}}^2 \equiv \sum_i \left[\frac{D_i - T_i - \sum_{k=1}^4 s_k \beta_i^k}{\delta_i} \right]^2 + \sum_{k=1}^4 s_k^2, \quad (12)$$

$$\beta_i^k \equiv \delta_{i,k}^{\text{norm.}} T_i \quad (13)$$

where D_i corresponds to central data value and δ_i is the the uncorrelated point-to-point uncertainty. The relative normalization uncertainties $\delta_{i,k}^{\text{norm.}}$ are treated as fully correlated. Note that the systematic shifts $s_k \beta_i^k$ are taken to be proportional to the theory values in order to avoid the D'Agostini

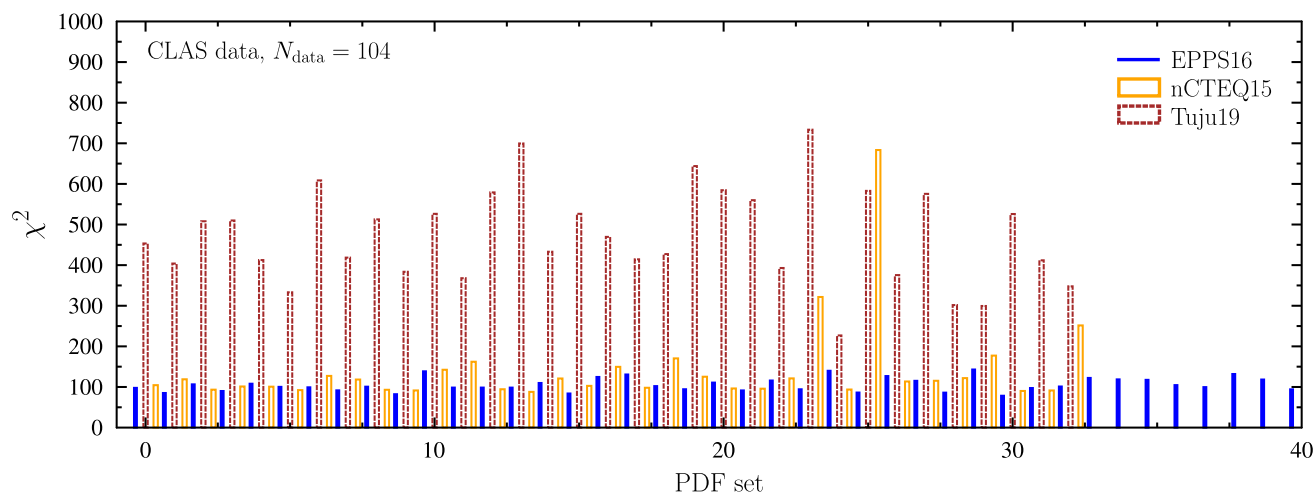


Fig. 3 The χ^2 values corresponding to the central and error sets of the nuclear PDFs

bias [51]. By minimizing the χ^2 with respect to parameters s_k one finds the “optimum shifts” $s_k^{\min} \beta_i^k$ that correspond to a given set of theory predictions T_i .

3 Results

The first thing done was to compare the data with the expected predictions from nuclear PDFs. We present the results for all the three nuclear-PDF sets and all four nuclei in Fig. 2. In the case of EPPS16 we also compare to a calculation without TMCs and in the case of TuJu19 to a calculation which assumes no nuclear effects in deuteron. With EPPS16 and nCTEQ15 the agreement is visually quite good: most data lie within the uncertainty bands and the downward slopes are well reproduced. The nCTEQ15 error bands are generally larger than those of EPPS16 – particularly for Pb – and can be explained by the larger uncertainties in the upvalence distributions as was seen in Fig. 1. This indicates that these data should be able to set significant new constraints especially for nCTEQ15. While the data are also broadly reproduced by the TuJu19 PDFs, the predicted EMC slope appears to be somewhat too flat systematically for all the four nuclei and the predictions tend to underestimate the data around $0.2 < x < 0.35$. While perhaps a bit unexpected, the systematic difference between the EPPS16/nCTEQ15 and TuJu19 predictions is consistent e.g. with Fig. 10 of the original TuJu19 paper [32], where the fit can be seen to somewhat underestimate the NMC data for C/D and Ca/D ratios at $x \gtrsim 0.1$. The EPPS16 values for these same data are somewhat higher, as can be seen from Fig. 13 of the original EPPS16 paper [30], and better agree with the data. Thus, the differences we observe here seem to be consistent.

From the EPPS16 panels of Fig. 2 we see that the effect of TMCs becomes relevant at $x \gtrsim 0.3$ and the TMCs evidently

provoke and upward shift in the predictions. Since the Nachtmann variable ξ is always smaller than Bjorken x , $\xi < x$, by turning on the TMCs one effectively probes the nPDFs at bit lower momentum fraction. As can be seen from Fig. 1, the nuclear effects in PDFs are monotonic in the EMC region so by shifting to a smaller momentum fraction by turning on TMCs the cross-section ratios increase a bit. It appears that a slightly better agreement with the data is obtained with TMCs – we will later on see to what extent this is significant. In any case, the effect of TMCs competes with the uncorrelated data uncertainties so it might become relevant, then, to consider TMCs in future fits of nuclear PDFs.

Out of the three nuclear-PDF fits considered here, the TuJu19 analysis is the only one to consider nuclear effects for deuteron. This was done by extending the parametrization of the A dependence down to $A = 2$ and utilizing deuteron structure-function data as a constraint. The effect of nuclear corrections to deuteron PDFs are indicated in the TuJu19 panels of Fig. 2. The corrections are the largest at the highest values of x , amounting to $\sim 4\%$ at the most. This appears to be in line with e.g. the phenomenological study of Ref. [52]. The estimated effects of deuteron corrections exceed the uncorrelated data uncertainties at $x \gtrsim 0.35$. The EPPS16 and nCTEQ15 analyses do not consider nuclear effects for deuteron basically because the smooth, power-law type parametrization of the A dependence may not be completely reliable for very small nuclei, but some discontinuities could be expected at small A . For example, the HKN07 analysis [53] introduced an extra overall parameter to suppress the otherwise somewhat too strong modifications of the deuteron PDFs. In fact, a possible explanation why TuJu19 fails to reproduce the CLAS data is that the parametrization of the A dependence is too simple to reliably cover all considered nuclei.

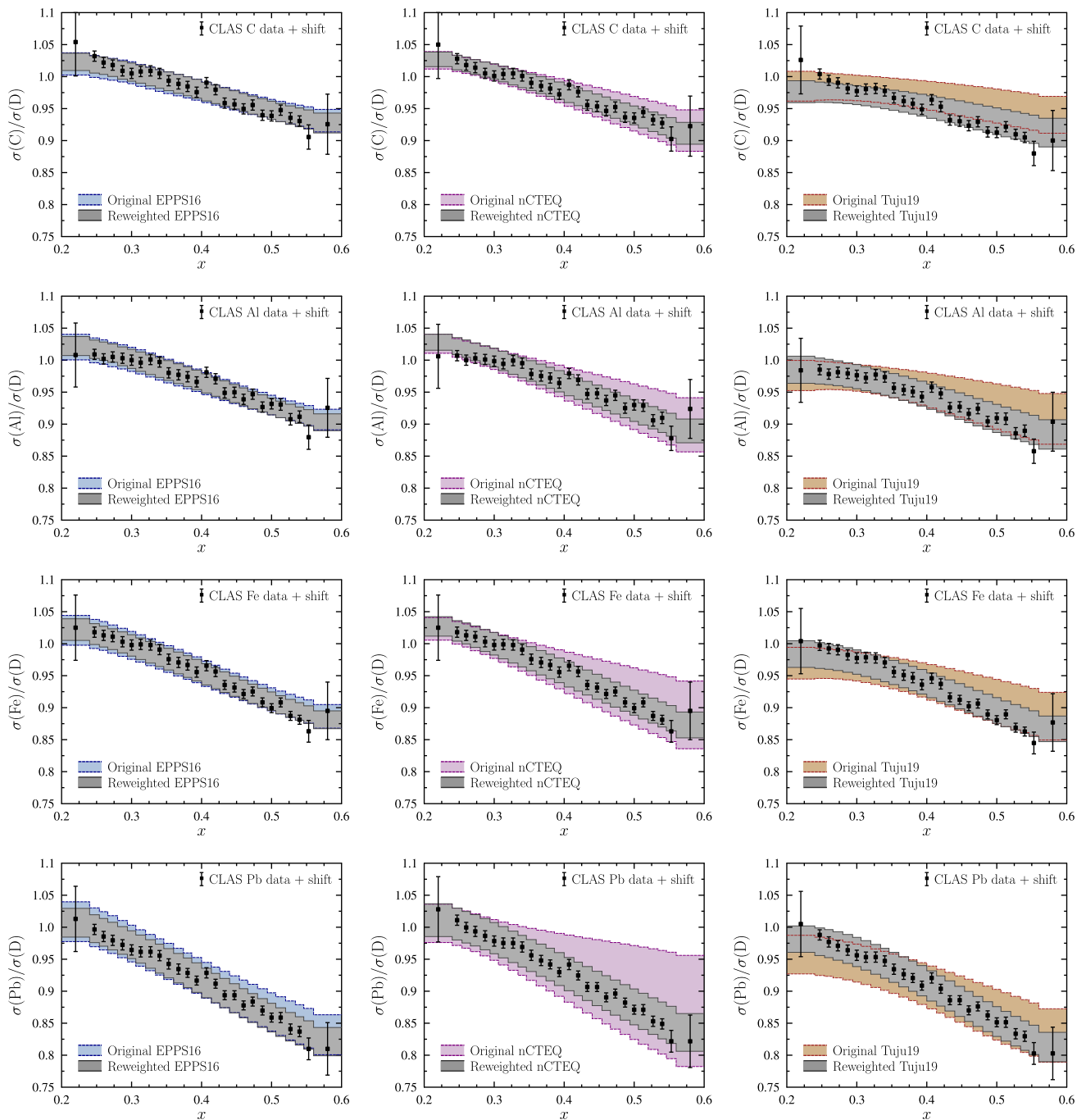


Fig. 4 The CLAS data compared with the re-weighted nuclear-PDF predictions. The optimal shifts that minimize Eq. (12) with the central re-weighted predictions, have been applied to the data points. The normalization uncertainties are not included in the data error bars

A more quantitative estimate of the data-to-theory correspondence can be obtained by looking at the χ^2 values. To this intent, we computed the χ^2 for all the central and error sets. The resulting values are displayed in Fig. 3. The central values are $\chi^2/N_{\text{data}} = 0.93$ for EPPS16, $\chi^2/N_{\text{data}} = 0.98$ for nCTEQ15, and $\chi^2/N_{\text{data}} = 4.4$ for TuJu19. Thus, the central sets of EPPS16 and nCTEQ15 are well compatible with the CLAS data while TuJu19 is not. As can be seen from

Fig. 3, the χ^2 values given by the EPPS16 error sets are all very similar and close to the central value. This insensitivity implies that the CLAS data are well compatible with EPPS16 and that they will not have a very significant effect if included in the analysis. The good χ^2 values are a consequence of the fact that all the EPPS16 sets yield an EMC slope which is more or less compatible with the data, and the overall offsets can be compensated by appropriately shuffling the normal-

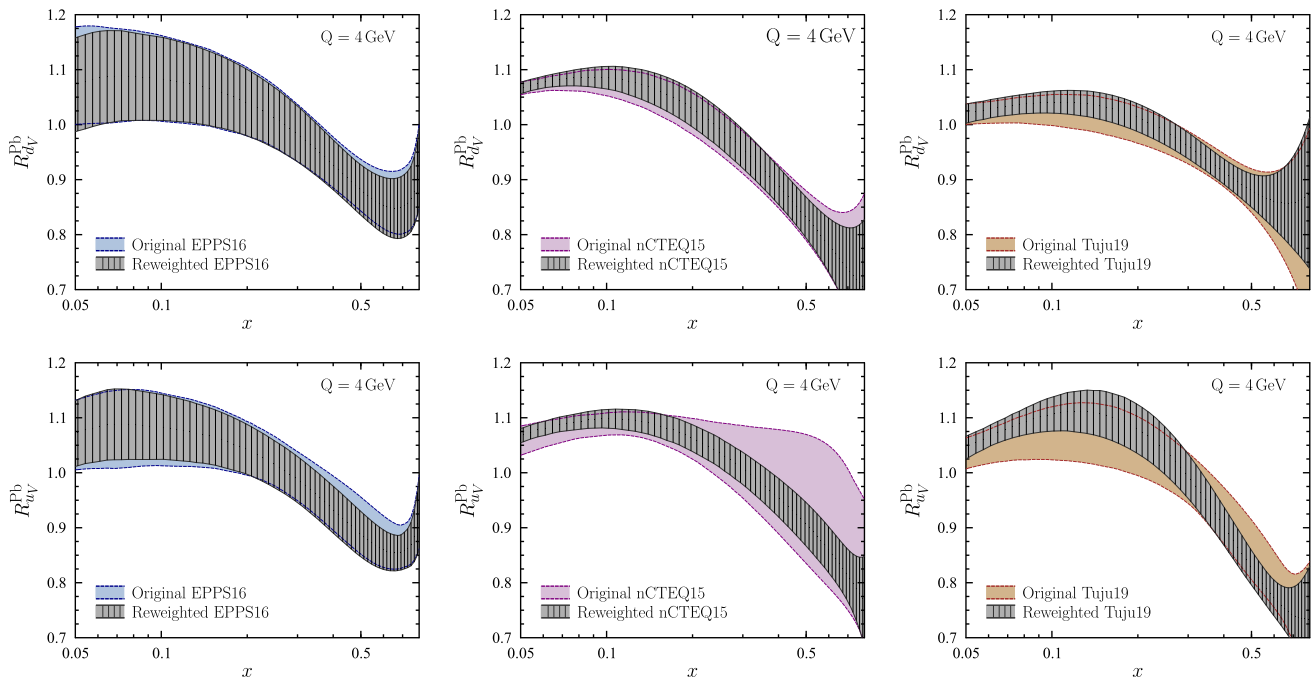


Fig. 5 Effect of re-weighting on nuclear modifications of down (upper row) and up (lower row) valence quarks in Pb at $Q = 4$ GeV. The coloured bands are the original uncertainty bands and the grey bands with hatching are the ones after re-weighting

ization parameters s_k in Eq. (12). In the case of nCTEQ15 there is clearly much more variation from one error set to another. Indeed, the error sets 24, 26, 30 and 33 evidently stick out from the rest. For the high values of χ^2 these error sets correspond to points in the fit-parameter space that are incompatible with the CLAS data. Given that the nCTEQ15 tolerance $\Delta\chi^2 = 35$ is much less than the variation we see in Fig. 3 it can be expected that the CLAS data will have a notable impact on nCTEQ15. There are also quite some variation in the χ^2 values obtained with TuJu19 error sets. While none of the error sets agree with the data, the observed variation implies that it is possible to find combinations of error sets that improve the agreement.

The results of re-weighting are presented in Figs. 4 and 5, with some characteristics given in Table 1. From the numbers in Table 1 we see that the re-weighting has been able to decrease the central value of χ^2 by some tens of units in the case of EPPS16/nCTEQ15, and by some staggering 300 units in the case of TuJu19. The estimated increase in the original minimum χ^2 (PDF penalty) in the EPPS16 and nCTEQ15 analyses is only a few units – clearly less than the tolerances $\Delta\chi^2$. These numbers corroborate the fact that the CLAS data are fully compatible with these two sets of PDFs. In the case of TuJu19 the penalty is ~ 70 units which clearly exceeds the estimated error tolerance $\Delta\chi_{\text{TuJu19}}^2 = 50$. Thus, although the new central value $\chi^2/N_{\text{data}} = 1.5$ is acceptable, some other data in the TuJu19 analysis are no longer satisfactorily reproduced. This means that there is a striking

contradiction between the TuJu19 analysis and the CLAS data.

In Fig. 4 we present a comparison between the original error bands of Fig. 2 and the ones after re-weighting the PDFs with the CLAS data. As anticipated, the re-weighting has induced only modest effects on EPPS16 predictions which are barely visible for other than the two heaviest nuclei. In the case of nCTEQ15 the re-weighted error bands are notably narrower than the original ones – more than a factor of two in some places. For both EPPS16 and nCTEQ15 the optimal shifts in the data due to the normalization uncertainties are not particularly large. In the case of TuJu19 the re-weighting has induced a quite significant change in the EMC slope. The partons have adjusted themselves to clearly steepen the originally too flat EMC slope and also the uncertainties are somewhat reduced. Thus, even if there are now incompatibilities between the original TuJu19 fit and the CLAS data, the uncertainties do not generally grow. The optimal shifts in the central data values are also larger than in the case of EPPS16/nCTEQ15.

The original up- and down-valence distributions of Fig. 1 are compared with the re-weighted ones in Fig. 5 for EPPS16 (left panels), nCTEQ15 (middle panels) and TuJu19 (right panels). The upper row corresponds to the valence down-quark distributions which seem to remain rather stable upon performing the re-weighting. The lower panels correspond to the valence up-quark distributions. Again, EPPS16 remains nearly unchanged while there are now significant differences

in nCTEQ15 and TuJu19. In the case of nCTEQ15 the theoretical uncertainties for the up-valence distribution reduce quite dramatically in the region spanned by the data. Through the assumed form of the fit functions these improvements are also reflected at smaller x . The reason why the CLAS data has restricted particularly the up-valence distributions can be understood on the basis of non-isoscalarity of the heaviest CLAS nucleus. Indeed, e.g. for Pb nucleus,

$$\frac{Z + 4N}{4Z + N} \approx 1.3, \quad (14)$$

so that the CLAS data are sensitive to several different linear combinations of $R_{uv}^{p/A}$ and $R_{dv}^{p/A}$ – not just the one indicated in the last row of Eq. (3) when only isoscalar nuclei are used in the fit. As a result, one can better unfold both $R_{uv}^{p/A}$ and $R_{dv}^{p/A}$ separately. Since $R_{uv}^{p/A}$ is better constrained already before the re-weighting most of the new constraints go to $R_{dv}^{p/A}$. From Eqs. (4) and (5) we in turn see, that when $R_{dv}^{p/A}$ gets better constrained the impact is stronger in R_{uv}^{Pb} . This explains the hierarchy seen in Fig. 5 for nCTEQ15. For TuJu19 the main effect is that the up-valence distribution has become steeper from its original shape. This increased steepness is in agreement with the steeper cross sections observed in Fig. 4.

As a final exercise we have investigated the role of TMCs when performing the re-weighting. In Fig. 6 we plot the EPPS16 up- and down-valence distributions also in the case the TMCs are not applied (in our default results the TMCs are always incorporated). We see that the differences are very moderate between the TMC and no-TMC cases. This is presumably related to the overall normalization uncertainties which can hide the differences seen in Fig. 2 if the TMCs are not applied, by appropriately reshuffling the systematic parameters s_k in the χ^2 function.

4 Summary

In the present work we have scrutinized the recent high- x neutral-current DIS data measured by the CLAS collaboration. In particular, we have investigated whether these data are in agreement with the modern nuclear PDFs and whether they could provide additional constraints. We have found that the data agree nicely with the EPPS16 and nCTEQ15 global analyses of nuclear PDFs while they disagree with TuJu19. As a feasible explanation for the clash with TuJu19 we entertained the possibility that extending the parametrization of the A dependence down to deuteron may bias the predictions at large A . In any case, from the good agreement with EPPS16 and nCTEQ15 we can conclude that these data are compatible with the other world data used in global fits of nuclear PDFs.

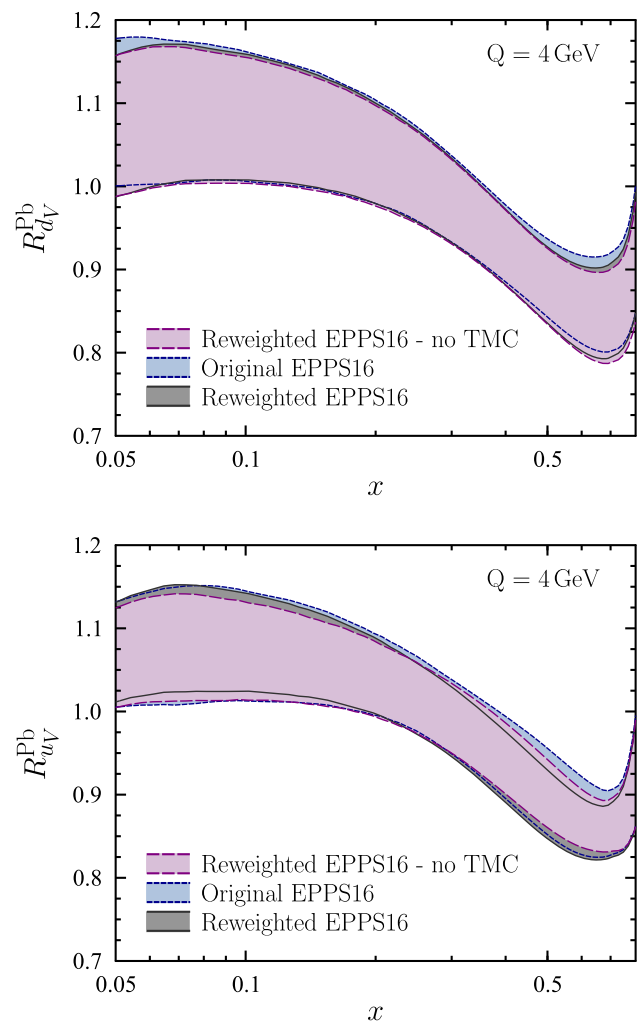


Fig. 6 Effect of re-weighting on EPPS16 with and without the TMCs

What is also interesting here is that the CLAS data are situated at lower Q^2 than the other large- x data in the global fits. The agreement we find indicates that there are no significant additional higher-twist contributions present. Although the target-mass effects are of the same size as the uncorrelated CLAS data uncertainties, their impact in the global analysis is predicted to be small due to the normalization uncertainties that can partly shroud these effects. Including TMCs in future fits of nuclear PDFs would then be a recommendable but not a crucial practice. In addition, the nuclear PDFs do not encode non-trivial nuclear effects that would depend on the isospin. We thus find no evidence of short-range correlations or equivalent phenomena that would depend on the relative number of protons and neutrons in the nuclei. This is in line with the results of e.g. Ref. [54]. Our findings allow us then to give an affirmative answer to the question raised in the title.

Acknowledgements We thank E. Segarra for providing us the CLAS data. P.Z. was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Research Unit FOR 2926, grant number 409651613. H.P. wishes to acknowledge the funding from the Academy of Finland project 308301.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: All relevant analysis output data are presented in figures or given in text and the numerical values are available from the authors upon request.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Funded by SCOAP³.

References

- Paukkunen Hannu, Nuclear PDFs Today. *PoS, HardProbes* **2018**,014 (2018)
- Hannu Paukkunen, Status of nuclear PDFs after the first LHC p-Pb run. *Nucl. Phys. A* **967**, 241–248 (2017)
- John C. Collins, Davison E. Soper, George F. Sterman, Factorization of hard processes in QCD. *Adv. Ser. Direct. High Energy Phys.* **5**, 1–91 (1989)
- Yuri L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics. *Sov. Phys. JETP* **46**, 641–653 (1977)
- Yuri L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics. *Zh. Eksp. Teor. Fiz.* **73**, 1216 (1977)
- V.N. Gribov, L.N. Lipatov, Deep inelastic e^+p scattering in perturbation theory. *Sov. J. Nucl. Phys.* **15**, 438–450 (1972)
- V.N. Gribov, L.N. Lipatov, Deep inelastic e^+p scattering in perturbation theory. *Yad. Fiz.* **15**, 781 (1972)
- V.N. Gribov, L.N. Lipatov, e^+e^- pair annihilation and deep inelastic e^+p scattering in perturbation theory. *Sov. J. Nucl. Phys.* **15**, 675–684 (1972)
- V.N. Gribov, L.N. Lipatov, e^+e^- pair annihilation and deep inelastic e^+p scattering in perturbation theory. *Yad. Fiz.* **15**, 1218 (1972)
- Guido Altarelli, G. Parisi, Asymptotic freedom in parton language. *Nucl. Phys.* **B126**, 298–318 (1977)
- Georges Aad et al., Centrality, rapidity and transverse momentum dependence of isolated prompt photon production in lead-lead collisions at $\sqrt{s_{NN}} = 2.76$ TeV measured with the ATLAS detector. *Phys. Rev.* **C93**(3), 034914 (2016)
- Georges Aad et al., Measurement of the production and lepton charge asymmetry of W bosons in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector. *Eur. Phys. J.* **C75**(1), 23 (2015)
- S. Acharya et al., Measurement of Z^0 -boson production at large rapidities in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett* **B780**, 372–383 (2018)
- Serguei Chatrchyan et al., Study of Z production in PbPb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV in the dimuon and dielectron decay channels. *JHEP* **03**, 022 (2015)
- Kari J. Eskola, Ilkka Helenius, Petja Paakkinen, and Hannu Paukkunen, A QCD analysis of LHCb D-meson data in p+Pb collisions. (2019)
- H. Abramowicz et al., Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data. *Eur. Phys. J. C* **75**(12), 580 (2015)
- A. Accardi, L.T. Brady, W. Melnitchouk, J.F. Owens, N. Sato, Constraints on large- x parton distributions from new weak boson production and deep-inelastic scattering data. *Phys. Rev. D* **93**(11), 114017 (2016)
- A. Accardi et al., Electron Ion Collider: The Next QCD Frontier. *Eur. Phys. J. A* **52**(9), 268 (2016)
- J. L. Abelleira Fernandez et al., A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector. *J. Phys.* **G39**, 075001 (2012)
- B. Schmookler et al., Modified structure of protons and neutrons in correlated pairs. *Nature* **566**(7744), 354–358 (2019)
- Michele Arneodo, Nuclear effects in structure functions. *Phys. Rept.* **240**, 301–393 (1994)
- Simona Malace, David Gaskell, Douglas W. Higinbotham, Ian Cloet, The Challenge of the EMC Effect: existing data and future directions. *Int. J. Mod. Phys. E* **23**(08), 1430013 (2014)
- J. Gomez et al., Measurement of the A-dependence of deep inelastic electron scattering. *Phys. Rev. D* **49**, 4348–4372 (1994)
- M. Arneodo et al., The A dependence of the nuclear structure function ratios. *Nucl. Phys. B* **481**, 3–22 (1996)
- P. Amaudruz et al., A Reevaluation of the nuclear structure function ratios for D, He, Li-6, C and Ca. *Nucl. Phys.* **B441**, 3–11 (1995)
- G. Onengut et al., Measurement of nucleon structure functions in neutrino scattering. *Phys. Lett. B* **632**, 65–75 (2006)
- Albert M Sirunyan et al., Constraining gluon distributions in nuclei using dijets in proton-proton and proton-lead collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. Lett.* **121**(6), 062002 (2018)
- O. Hen, G.A. Miller, E. Piasetzky, L.B. Weinstein, Nucleon-Nucleon Correlations, Short-lived Excitations, and the Quarks Within. *Rev. Mod. Phys.* **89**(4), 045002 (2017)
- S. I. Alekhin, A. L. Kataev, Sergey A Kulagin, M. V. Osipenko, Evaluation of the isospin asymmetry of the nucleon structure functions with CLAS++. *Nucl. Phys.* **A755**, 345–349 (2005)
- Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen, Carlos A. Salgado, EPPS16: Nuclear parton distributions with LHC data. *Eur. Phys. J. C* **77**(3), 163 (2017)
- K. Kovarik et al., nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework. *Phys. Rev. D* **93**(8), 085037 (2016)
- Marina Walt, Ilkka Helenius, Werner Vogelsang, Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO. *Phys. Rev. D* **100**(9), 096015 (2019)
- K.J. Eskola, H. Paukkunen, C.A. Salgado, EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions. *JHEP* **04**, 065 (2009)
- Daniel de Florian, Rodolfo Sassot, Pia Zurita, Marco Stratmann, Global Analysis of Nuclear Parton Distributions. *Phys. Rev. D* **85**, 074028 (2012)
- Hamzeh Khanpour, S. Atashbar Tehrani, Global Analysis of Nuclear Parton Distribution Functions and Their Uncertainties at Next-to-Next-to-Leading Order. *Phys. Rev* **D93**(1), 014026 (2016)
- Rabah Abdul Khalek, Jacob J Ethier, Juan Rojo, Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider. *Eur. Phys.* **C79**(6), 471 (2019)
- Michael Krämer, Fredrick I. Olness, Davison E. Soper, Treatment of heavy quarks in deeply inelastic scattering. *Phys. Rev. D* **62**, 096007 (2000)

38. Wu-Ki Tung, Stefan Kretzer, Carl Schmidt, Open heavy flavor production in QCD: Conceptual framework and implementation issues. *J. Phys.* **G28**, 983–996 (2002)
39. Stefano Forte, Eric Laenen, Paolo Nason, Juan Rojo, Heavy quarks in deep-inelastic scattering. *Nucl. Phys. B* **834**, 116–162 (2010)
40. R.S. Thorne, Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections. *Phys. Rev. D* **86**, 074017 (2012)
41. R.G. Roberts, *The Structure of the proton: Deep inelastic scattering* (Cambridge University Press, Cambridge Monographs on Mathematical Physics, 1994)
42. Howard Georgi, H David Politzer, Freedom at Moderate Energies: Masses in Color Dynamics. *Phys. Rev.* **D14**, 1829 (1976)
43. Alberto Accardi, Jian-Wei Qiu, Collinear factorization for deep inelastic scattering structure functions at large Bjorken $x(B)$. *JHEP* **07**, 090 (2008)
44. Ingo Schienbein et al., A Review of Target Mass Corrections. *J. Phys.* **G35**, 053101 (2008)
45. J. Seely et al., New measurements of the EMC effect in very light nuclei. *Phys. Rev. Lett.* **103**, 202301 (2009)
46. Hannu Paukkunen, Carlos A. Salgado, Agreement of Neutrino Deep Inelastic Scattering Data with Global Fits of Parton Distributions. *Phys. Rev. Lett.* **110**(21), 212301 (2013)
47. Hannu Paukkunen, Pia Zurita, PDF reweighting in the Hessian matrix approach. *JHEP* **12**, 100 (2014)
48. Kari J Eskola, Petja Paakkinen, Hannu Paukkunen, Non-quadratic improved Hessian PDF reweighting and application to CMS dijet measurements at 5.02 TeV. *Eur. Phys. J* **C79**(6), 511 (2019)
49. Carl Schmidt, Jon Pumplin, C.P. Yuan, P. Yuan, Updating and optimizing error parton distribution function sets in the Hessian approach. *Phys. Rev. D* **98**(9), 094005 (2018)
50. Tie-Jiun Hou, Zhite Yu, Sayipjamal Dulat, Carl Schmidt, C.-P. Yuan, Updating and optimizing error parton distribution function sets in the Hessian approach. II. *Phys. Rev. D* **100**(11), 114024 (2019)
51. Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, Maria Ubiali, Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties. *JHEP* **05**, 075 (2010)
52. A.D. Martin, AJThM Mathijssen, W.J. Stirling, R.S. Thorne, B.J.A. Watt, G. Watt, Extended Parameterisations for MSTW PDFs and their effect on Lepton Charge Asymmetry from W Decays. *Eur. Phys. J. C* **73**(2), 2318 (2013)
53. M. Hirai, S. Kumano, T.H. Nagai, Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order. *Phys. Rev. C* **76**, 065207 (2007)
54. J. Arrington, N. Fomin, Searching for flavor dependence in nuclear quark behavior. *Phys. Rev. Lett.* **123**(4), 042501 (2019)