PHYSICAL REVIEW D **93,** 085037 (2016)

nCTEQ15: Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework

K. Kovařík,[1] A. Kusina,[2] T. Ježo,[3] D. B. Clark,[4] C. Keppel,[5] F. Lyonnet,[4] J. G. Morfín,[6] F. I. Olness,[4] J. F. Owens,[7] I. Schienbein,[2] and J. Y. Yu[4]

[1]*Institut für Theoretische Physik, Westfälische Wilhelms-Universität Münster, Wilhelm-Klemm-Straße 9, D-48149 Münster, Germany*
[2]*Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, 53 avenue des Martyrs, 38026 Grenoble, France*
[3]*Università di Milano-Bicocca and INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*
[4]*Southern Methodist University, Dallas, Texas 75275, USA*
[5]*Thomas Jefferson National Accelerator Facility, Newport News, Virginia 23606, USA*
[6]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[7]*Department of Physics, Florida State University, Tallahassee, Florida 32306-4350, USA*

(Received 4 September 2015; revised manuscript received 26 February 2016; published 28 April 2016)

We present the new nCTEQ15 set of nuclear parton distribution functions (PDFs) with uncertainties. This fit extends the CTEQ proton PDFs to include the nuclear dependence using data on nuclei all the way up to ^{208}Pb. The uncertainties are determined using the Hessian method with an optimal rescaling of the eigenvectors to accurately represent the uncertainties for the chosen tolerance criteria. In addition to the deep inelastic scattering and Drell-Yan processes, we also include inclusive pion production data from the Relativistic Heavy Ion Collider to help constrain the nuclear gluon PDF. Furthermore, we investigate the correlation of the data sets with specific nuclear PDF flavor components and asses the impact of individual experiments. We also provide comparisons of the nCTEQ15 set with recent fits from other groups.

DOI: 10.1103/PhysRevD.93.085037

I. INTRODUCTION

In the last 30 years, an impressive array of discoveries in particle physics has come from high energy hadron experiments. These discoveries, along with many other key measurements, rely on our understanding of nucleon structure. A nucleon can be described using the language of parton distribution functions (PDFs) which is based on QCD factorization theorems [1–3]. PDFs are determined in global analyses of a variety of different hard scattering processes such as deep inelastic scattering (DIS), Drell-Yan (DY) lepton pair production, vector boson production, and the inclusive jet production. The backbone of any global analysis is the very precise DIS structure function data from HERA which cover a wide kinematic range in (x, Q^2). Several global analyses, based on an ever growing set of precise experimental data and on next-to-next-to-leading order (NNLO) theoretical predictions, are regularly updated and maintained [4–9].

Over the years, a series of global analysis studies have been performed within a single framework, or comparing different frameworks. For example, detailed studies of PDF uncertainties have been compared using Hessian, Lagrangian, and Monte Carlo methods. Furthermore, the precision of experimental data and theoretical predictions in the proton case allows one to perform studies of smaller effects such as the difference between the treatment of heavy quarks in different analyses or the exact treatment of target-mass corrections and higher twist effects. As a consequence, the nucleon structure is quite well known over a wide kinematic range.

Similarly, the theoretical description of hard scattering processes in lepton-nucleus and proton-nucleus reactions requires the knowledge of parton distribution functions inside nuclei characterized by the atomic number A and the charge Z. It has been known since the discovery of the European Muon Collaboration (EMC) effect [10] more than 30 years ago that the nucleus cannot be considered as an ensemble of Z free protons and $(A - Z)$ free neutrons. Consequently, the nuclear PDFs (nPDFs) will differ from the naive additive combination of free-proton and neutron PDFs. As in the proton case, nPDFs have been determined in the literature by global fits to experimental data for hard scale processes including deep inelastic scattering on nuclei and nuclear collision experiments [11–14]. However, compared to the proton, our knowledge of nuclear PDFs is much less advanced. There are several reasons.

On the theoretical side, the description of nuclear induced hard processes is more challenging due to the complex nuclear environment. Still, all global nuclear PDF analyses rely on the assertion that the QCD factorization

theorems remain valid for ℓA and pA hard scattering processes; see, e.g., Refs. [15,16]. In fact, it is only in this context that the universal parton distributions ($f_i^A(x,Q)$) are defined; they are given as matrix elements of the same local twist-2 operators as in the proton case but on nuclear states. The nuclear PDFs then account for nuclear effects (in particular EMC suppression, shadowing, antishadowing) at the twist-2 level in a universal manner, and the entire formalism becomes predictive. However, higher twist contributions are expected to be enhanced in a nucleus ($\propto A^{1/3}$) [15,16]. Here, final-state rescattering corrections due to the propagation of the outgoing partons through the nuclear medium, which are higher twist, should be power suppressed but may be substantial and so must be either included in the analysis or eliminated by suitable kinematic cuts.[1] In addition, other effects like a different propagation of the hadronic fluctuations of the exchange bosons in the nuclear medium,[2] gluon saturation, and deviations from Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution at small x may play a more prominent role in the nuclear case; see, e.g., Refs. [18,19] and references therein.

Ultimately, the validity of the twist-2 factorization formalism will be tested phenomenologically by how well our approach based on the factorization assumption describes the data. The existing global analyses generally lead to a good description of the data confirming this picture; however, it may be challenged by future precision data from the LHC and an electron-ion collider covering an extended kinematic plane. It is notable that tensions between νA DIS data and ℓA DIS data have been reported [20,21], which might be due to higher twist contributions or indicate a breaking of twist-2 factorization. These tensions largely disappear if the correlations between the NuTeV data points are discarded [22].

The other reason why nuclear PDFs lag behind the proton analyses can be traced back to the lack of precise experimental data. For example, the constraints on the nPDFs for any single nucleus are (so far) too scarce so that experimental data from scattering on multiple nuclei must be considered. Since the nuclear effects are clearly dependent on the number of nucleons, this requires modeling of the nontrivial nuclear A dependence of the parton distributions. Even after combining the data sets for different nuclei, the precision of the nuclear PDFs is not yet comparable to the proton PDFs where quark distributions for most flavors together with the gluon distribution are reliably determined over a broad kinematic range, due to the smaller number and hence smaller kinematic coverage of the current relevant nuclear data. As a consequence, the nuclear PDFs in every analysis have large uncertainties as the parton distributions are not fully constrained by the available data. The nuclear PDFs still largely depend on assumptions inherent in every analysis. The dependence on assumptions, such as for example the parametrization form, leads to predictions where different analyses differ by more than the estimated uncertainties. It follows therefore that in order to assess the true uncertainty, all available results and their uncertainties should be considered and combined.

In this paper, we present a new analysis of nuclear PDFs in the CTEQ global PDF fitting framework. We use theoretical predictions at the next-to-leading order to fit all available data from charged lepton DIS and Drell-Yan dilepton production as in our previous analysis [13]. In addition, we have added inclusive pion production data from the Relativistic Heavy Ion Collider (RHIC) and have performed a careful analysis of the uncertainties using the Hessian method. Our framework differs considerably from other global analyses of nuclear PDFs with which we compare our results.

The remainder of this paper is organized as follows. In Sec. II, we introduce in detail the framework including the parametrization of the nPDFs at the input scale together with a review of the Hessian method which we use to estimate the uncertainties on the nPDFs. In Sec. III, we review the experimental data included in the fit. In Sec. IV, we present the results of our fit, compare with recent results from the literature, and examine the correlations between individual PDF flavors and the various experiments. Finally, in Sec. V, we summarize the obtained results. Additionally, we include two Appendixes. In Appendix A, we provide details on the Hessian rescaling method, and in Appendix B, we comment on the usage and availability of our nPDFs.

II. NPDF FRAMEWORK

In this section, we describe in detail the framework of the nCTEQ global analysis. For the purpose of fitting nuclear parton distributions, we will parametrize $f_i^{p/A}(x,Q_0)$, the PDFs of a proton bound in a nucleus A, then construct the full distributions of partons in the nucleus using isospin symmetry and in the end perform a fit just like in the case of the free proton. Indeed, isospin symmetry is used to construct the PDFs of a bound neutron, $f_i^{n/A}(x,Q)$, from those of the proton by exchanging up- and down-quark distributions. Afterward, the parton distributions of the nucleus are constructed as

$$f_i^{(A,Z)}(x,Q) = \frac{Z}{A} f_i^{p/A}(x,Q) + \frac{A-Z}{A} f_i^{n/A}(x,Q), \qquad (2.1)$$

[1]Needless to say that the final-state interactions do not concern the fully inclusive DIS structure functions but may be relevant for less inclusive observables (single pion production, dimuon production in νA DIS, …). On the other hand, power suppressed initial state interactions are expected to be numerically small.

[2]There could be modifications of charged-current neutrino scattering that are different than those for neutral current charged lepton scattering for instance due to the exchange of a charged massive vector boson [17].

where Z is the number of protons and A the number of protons and neutrons in the nucleus.[3]

The theoretical calculations in our global analysis make use of parton distributions of a particular nucleus $f_i^{(A,Z)}$ to determine the DIS structure functions, Drell-Yan cross sections, or the cross section for an inclusive pion production,

$$F_2^A(x,Q^2) = \sum_i f_i^{(A,Z)}(x,Q^2) \otimes C_{2,i}(x,Q^2), \tag{2.2}$$

$$d\sigma_{AB\to l\bar{l}X} = \sum_{ij} f_i^{(A_1,Z_1)} \otimes f_j^{(A_2,Z_2)} \otimes d\hat{\sigma}^{ij\to l\bar{l}X}, \tag{2.3}$$

$$d\sigma_{dA\to \pi X} = \sum_{ijk} f_i^{d} \otimes f_j^{(A,Z)} \otimes d\hat{\sigma}^{ij\to kX} \otimes D_k^{\pi}, \tag{2.4}$$

where $\otimes$ stands for a convolution integral over the momentum fraction. The DIS structure functions calculations are carried out using the ACOT variable flavor number scheme [3,23–25] at next-to-leading order in QCD [26].[4] We take into account only the dominant target-mass effects which are included in the structure function expressions in the ACOT scheme [23]. Full treatment of the target-mass corrections [30] is not necessary in our analysis because they are relevant mostly at large x and low Q^2, a region of phase space which we exclude by kinematic cuts. Moreover, the target-mass corrections are expected to be of lesser importance in the ratios of structure functions.

In all theory calculations, we identify the renormalization and factorization scales: $\mu = \mu_R = \mu_F$. The scale is set differently for different processes: in deep-inelastic scattering, it is set to the virtuality of the exchanged vector boson $\mu^2 = Q^2$; in Drell-Yan production processes, it is set to the invariant mass of the produced lepton pair $\mu^2 = M^2$; and in inclusive pion production, the common scale is set equal to the final-state fragmentation scale as $\mu = \mu_F' = 0.5 p_T$ where p_T is the transverse momentum of the produced π^0. To speed up the evaluation of next-to-leading order (NLO) cross sections in the fit, we have the ability to use K-factors; however, for the final fitting, the full next-to-leading order calculations are used. In the case of inclusive pion production, we use the results of Ref. [31,32] and speed up the calculation by using precomputed grids already including convolutions with one PDF and fragmentation function and leaving only one convolution (with the nuclear PDFs) to be calculated during the fitting procedure.

[3]Note that the PDFs of the nucleus, $f_i^{(A,Z)}(x,Q)$, are the objects of interest which are constrained by the experimental data, whereas the $f_i^{p/A}(x,Q)$ and $f_i^{n/A}(x,Q)$ are just *effective* quantities used internally to decompose the nuclear PDFs. They should not be interpreted literally as matrix elements of local operators where the free nucleon states have been replaced by bound nucleon states in a nuclear medium since they also include effects from multinucleon states. The notion of "effective bound nucleon PDFs" is also used in the literature discussing the factorization in the case of pA interactions [15].

[4]For recent extensions of the ACOT scheme to higher orders, required for global analyses at next-to-next-to-leading order, see Refs. [27,28]; the massless limits have been validated with the help of QCDNUM [29].

A. Parametrization

The starting point of any determination of parton distribution functions is the parametrization of individual distributions at the input scale Q_0. The parametrization of the presented nCTEQ nuclear PDFs is the same as in our previous analyses [13,21,33]. It mimics the parametrization used in the free-proton CTEQ fits [34–36] and takes the following form:

$$\begin{aligned} x f_i^{p/A}(x,Q_0) &= c_0 x^{c_1}(1-x)^{c_2} e^{c_3 x}(1+e^{c_4}x)^{c_5}, \\ &\text{for } i = u_v, d_v, g, \bar{u}+\bar{d}, s+\bar{s}, s-\bar{s}, \\ \frac{\bar{d}(x,Q_0)}{\bar{u}(x,Q_0)} &= c_0 x^{c_1}(1-x)^{c_2} + (1+c_3 x)(1-x)^{c_4}. \end{aligned} \tag{2.5}$$

The input scale is chosen to be the same as for the free-proton fits [34,36], namely $Q_0 = 1.3$ GeV.

However, this parametrization needs to be appropriately modified to accommodate the additional nuclear degrees of freedom. As in other available nuclear PDFs [11,12,14], nuclear targets are characterized only by their atomic mass number A. However, in contrast to those nPDFs where the nuclear effects are added on top of the free-proton PDFs in the form of ratios, in our analysis we introduce the additional A dependence directly to the c-coefficients of Eq. (2.5):

$$c_k \to c_k(A) \equiv c_{k,0} + c_{k,1}(1 - A^{-c_{k,2}}), \quad k = \{1,\ldots,5\}. \tag{2.6}$$

This parametrization is designed in such a way that for $A = 1$ one recovers the underlying PDFs of a free proton. The free-proton PDFs are described by the coefficients $c_{k,0}$ which in our analysis are fixed to values of the fit of Ref. [34] which is close to CTEQ6.1 [36] but has the advantage of having minimal influence from nuclear data.

Although in principle this framework can be used to determine the strange quark content of the bound nucleon, there is not sufficient data available to reliably do that. Therefore, we assume that at the initial scale Q_0,

$$s^{p/A}(x,Q_0) = \bar{s}^{p/A}(x,Q_0) = \frac{\kappa(A)}{2}(\bar{u}^{p/A} + \bar{d}^{p/A}), \tag{2.7}$$

where $\kappa(A)$ is an A-dependent normalization factor parametrized as $\kappa(A) = (c_{0,0}^{s+\bar{s}} + c_{0,1}^{s+\bar{s}}(1 - A^{-c_{0,2}^{s+\bar{s}}}))$.[5]

The normalization coefficients c_0 in Eq. (2.5) are different than the other parameters. They are also dependent on the atomic number, but not all of them are free parameters that can be fitted. Most of them are constrained by sum rules. The normalization coefficients for the valence quark PDFs are constrained for each atomic number A by requiring that they obey the number sum rules

$$\int_0^1 dx f_{u_v}^{p/A}(x, Q_0) = 2, \qquad \int_0^1 dx f_{d_v}^{p/A}(x, Q_0) = 1. \tag{2.8}$$

The remaining normalization coefficients are constrained by the momentum sum rule

$$\int_0^1 dx \sum_i x f_i^{p/A}(x, Q_0) = 1, \tag{2.9}$$

which, however, can only determine one of them. The rest of the normalization parameters are either considered as free parameters in the fit or are fixed using additional assumptions to simplify the analysis [e.g., like Eq. (2.7)]. We choose to introduce free parameters for the momentum fraction of the gluon and for the momentum fraction of $s + \bar{s}$ to be determined during the global fit together with the parameters from Eqs. (2.5) and (2.6). The A-dependent momentum fraction of gluon is parametrized as

$$\int_0^1 dx x g^{p/A}(x, Q_0) = M_g \exp\left[c_{0,0}^g + c_{0,1}^g (1 - A^{-c_{0,2}^g})\right], \tag{2.10}$$

which modifies the momentum fraction of the gluon in a free proton (described by coefficients M_g and $c_{0,0}^g$).

The momentum fraction of the $s + \bar{s}$ combination is then given by

$$\int_0^1 dx x (s^{p/A}(x, Q_0) + \bar{s}^{p/A}(x, Q_0)) = \frac{\kappa}{(2+\kappa)} \left(1 - \int_0^1 dx \sum_i x f_i^{p/A}\right) [c_{0,0}^{s+\bar{s}} + c_{0,1}^{s+\bar{s}} (1 - A^{-c_{0,2}^{s+\bar{s}}})], \tag{2.11}$$

where the sum runs through $i = u_v$, d_v, g. The remaining normalization parameters are taken care of by the momentum sum rule and do not introduce additional free parameters.

[5]This is a straightforward generalization of the approach employed in the underlying proton analysis which also assumes that at the initial scale Q_0 the strange quark PDFs are constrained by $s = \bar{s} = \frac{\kappa}{2}(\bar{u} + \bar{d})$.

The parametrization of Eq. (2.5) together with the whole nCTEQ nuclear PDF framework has been designed in analogy to the free-proton PDFs where parton momentum x is restricted to be in the range (0,1). However, in the nuclear case, x represents the parton fractional momentum with respect to the average momentum carried by a nucleon. Since a particular nucleon can have a momentum bigger than an average nucleon, x can extend up to A in a nucleus with an atomic number A. If one were to take this into account, one would have to modify the sum rules in Eqs. (2.8) and (2.9) together with the DGLAP evolution. However, the structure functions at $x > 1$ fall off rapidly, and the contribution to the moments of the structure functions from the region of $x > 1$ is very small [37,38]. Therefore, all currently available nuclear PDFs have been obtained neglecting the $x > 1$ region, and we follow the same path.[6]

B. Finding the optimal PDFs

The fitting procedure used to find PDFs that describe the considered data best is based on minimizing the appropriate χ^2 function, as described in Ref. [35]. The simplest definition of the χ^2 function for n experiments is

$$\chi^2(\{a_j\}) = \sum_i \frac{[D_i - T_i(\{a_j\})]^2}{\sigma_i^2}, \tag{2.12}$$

where D_i are the measured experimental values, T_i are the corresponding theoretical predictions, and σ_i^2 are the systematic and statistical experimental errors added in quadrature. The parameters $\{a_j\}$ are a set of free parameters which define the PDFs at the input scale [see Eq. (2.5)] and are varied in order to find the minimum of the χ^2 function.

This simple χ^2 definition, with slight modifications allowing for the inclusion of overall changes to data normalization, is used by most of the groups performing nuclear global analyses. However, in the current analysis, as in the previous nCTEQ fits [13,20,21,33], this simple definition is modified to account for correlations in the experimental uncertainties. We follow here the prescription suggested in Ref. [35]. The total χ^2 for n experiments with parameters $\{a_j\}$ is defined to be

$$\chi^2(\{a_j\}) = \sum_n w_n \chi_n^2(\{a_j\}), \tag{2.13}$$

where w_n is the weight for experiment n; for our fits, all weights are set to 1. The χ_n^2 is a contribution from one individual experiment n, and this is given by

[6]In fact, the first next-to-leading order nuclear PDF analysis [39] used a framework which at least in principle allows one to accommodate the case of $x > 1$.

$$\chi^2_n(\{a_j\}) = \sum_i \frac{[D_i - T_i(\{a_j\})]^2}{\alpha_i^2} - \sum_{k,k'} B_k A^{-1}_{kk'} B_{k'}, \quad (2.14)$$

where i runs over data points and k, k' run over sources of the correlated uncertainties. For each experimental data point, we sum the statistical error σ_i together with the uncorrelated systematic error u_i in quadrature to obtain $\alpha_i^2 = \sigma_i^2 + u_i^2$. The components of the correlated uncertainties are given by [35]

$$B_k(\{a_j\}) = \sum_i \frac{\beta_{ik}[D_i - T_i(\{a_j\})]}{\alpha_i^2},$$
$$A_{kk'} = \delta_{kk'} + \sum_i \frac{\beta_{ik}\beta_{ik'}}{\alpha_i^2}, \quad (2.15)$$

where β_{ik} are the sources of correlated systematic errors.

We stress that in this procedure only the experimental uncertainties are accounted for; all theoretical and model uncertainties (e.g., missing higher order corrections, parametrization choice, etc.) are not taken into account.

Having defined the appropriate χ^2 function, it needs to be minimized with respect to the fitting parameters $\{a_j\}$ that define the bound proton PDFs at the initial scale Q_0. We perform the minimization using the pyMinuit package [40] which is a python interface to "SEAL-Minuit" [41]—a C++ rewrite of the original Fortran Minuit package [42].

C. Estimating uncertainties of PDFs

In Sec. II B, we described how we obtain our best estimate (the central value) of the nCTEQ nuclear PDFs as the minimum of the χ^2 function defined in Eq. (2.12). Now, we want to probe the vicinity of this minimum to be able to estimate uncertainties on our prediction. This is done using the Hessian method [43,44], which will be briefly described in the following. We follow the notation of Ref. [43] and refer the reader to this publication for more details on the Hessian formalism.

1. Determination of the Hessian matrix

The basic assumption of the Hessian method is that near its minimum the χ^2 function can be approximated by a quadratic form of the fitting parameters $\{a_i\}$. Therefore, it can be written as

$$\chi^2 = \chi^2_0 + \sum_{i,j} H_{ij} y_i y_j, \quad (2.16)$$

where $y_i = a_i - a_i^0$ are the parameter shifts from the minimum given by the a_i^0 parameters, $\chi^2_0 \equiv \chi^2(\{a_i^0\})$ is the value of the χ^2 function in the minimum, and H_{ij} is the Hessian matrix defined as

$$H_{ij} = \frac{1}{2}\left(\frac{\partial^2 \chi^2}{\partial y_i \partial y_j}\right)_{a_i = a_i^0}. \quad (2.17)$$

Since the Hessian H_{ij} is a symmetric $n \times n$ matrix (where n is the number of free parameters a_i), it has n orthogonal eigenvectors forming a basis in the $\{y_i\}$-space. The characteristic equation can be written as

$$\sum_j H_{ij} V_j^{(k)} = \lambda_k V_i^{(k)}. \quad (2.18)$$

The eigenvectors $V_i^{(k)}$ that we use can be normalized so that

$$\sum_i V_i^{(j)} V_i^{(k)} = \delta_{jk}. \quad (2.19)$$

For our later convenience, we also introduce eigenvectors normalized to the corresponding eigenvalues:

$$\tilde{V}_i^{(k)} = \frac{1}{\sqrt{\lambda_k}} V_i^{(k)}. \quad (2.20)$$

The eigenvectors can be used to disentangle the original PDF parameters and define a new basis $\mathbf{z} \equiv \{z_i\}$ where the Hessian is diagonal[7]:

$$\sum_{i,j} H_{ij} y_i y_j = \sum_{i,j} H^D_{ij} z_i z_j = \mathbf{z}^T.D^T.H.D.\mathbf{z}$$
$$= \mathbf{z}^T.\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}.\mathbf{z}. \quad (2.21)$$

The new coordinates are defined using a matrix D as

$$\mathbf{z} = D^{-1}\mathbf{y}, \quad (2.22)$$

where D is a matrix composed of eigenvectors:

$$D = (V^{(1)}, V^{(2)}, \dots, V^{(n)})$$
$$\equiv \begin{pmatrix} V_1^{(1)} & V_1^{(2)} & \dots & V_1^{(n)} \\ V_2^{(1)} & V_2^{(2)} & \dots & V_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ V_n^{(1)} & V_n^{(2)} & \dots & V_n^{(n)} \end{pmatrix}. \quad (2.23)$$

[7]In the basis defined using the rescaled eigenvectors $\tilde{V}_i^{(k)}$, the Hessian is represented by a unit matrix.

Note that because the Hessian is symmetric, $D^{-1} = D^T$. Using the index notation such as $D_{ij} = V_i^{(j)}$, we can write the relation between the original fitting parameters and the new parameters as

$$y_i = \sum_j V_i^{(j)} z_j \equiv \sum_j \tilde{V}_i^{(j)} \tilde{z}_j = \sum_j \frac{1}{\sqrt{\lambda_j}} V_i^{(j)} \tilde{z}_j, \quad (2.24)$$

where we introduced a new basis $\tilde{z}_i$ which corresponds to the rescaled eigenvectors $\tilde{V}_i^{(k)}$. The inverse transformation is given by

$$\begin{aligned} z_i &= \sum_j y_j V_j^{(i)}, \\ \tilde{z}_i &= \lambda_i \sum_j y_j \tilde{V}_j^{(i)} = \sqrt{\lambda_i} \sum_j y_j V_j^{(i)}. \end{aligned} \quad (2.25)$$

In the new coordinates, $\Delta\chi^2 = \chi^2 - \chi_0^2$ has a particularly simple form:

$$\Delta\chi^2 = \sum_i \lambda_i z_i^2 = \sum_i \tilde{z}_i^2. \quad (2.26)$$

Using the Hessian method to analyze the vicinity of the minimum of the χ^2 function seems straightforward in theory, but in practice when applied to a global PDF analysis, one encounters a few problems worth pointing out. As was already mentioned in the discussion of free-proton PDFs [43] and as is the case in our analysis, the eigenvalues of the Hessian span several orders of magnitude. In order to correctly identify all eigenvalues, the precision with which the Hessian matrix is determined needs to be kept under control.

In practice, the Hessian matrix is calculated using finite differences to determine the second derivatives. A careful choice of the step in the finite difference definition of the second derivatives is crucial. If the step is too large, one probes too large a neighborhood of the minimum where the χ^2 function cannot be described by a quadratic approximation anymore. If the step is too small, numerical noise in the χ^2 function prevents a reliable determination of the second derivatives. Moreover, the step size has to be different for each of the parameters as the χ^2 function depends differently on each of them. The relative step sizes Δy_i to each of the parameters are set as

$$\Delta y_i = \sqrt{\frac{\Delta\chi^2}{H_{ii}}}, \quad (2.27)$$

where $\Delta\chi^2 = \chi^2 - \chi_0^2$ defines the small neighborhood from which the derivatives of the χ^2 function are calculated.

It turns out that the numerical noise in the χ^2 function is larger than expected for the case of a global PDF analysis. Contrary to what one would expect, the χ^2 function is not smooth, which influences the determination of the second derivatives for all step sizes. It all comes down to the fact that one evaluation of the χ^2 function requires several hundred evaluations of different next-to-leading order theory calculations which, in their numerical implementations, are not smooth functions of the fit parameters.

To reduce the influence of the noise on the derivatives of the χ^2 function, the standard definition of the derivative using the central differences

$$\frac{df}{dx} = \frac{f_{+1} - f_{-1}}{2h} \quad (2.28)$$

in which $f_k = f(x_0 + kh)$ is replaced by noise reducing derivatives (see Ref. [45]). The central differences approach to derivatives is based on interpolating the χ^2 function by a polynomial which coincides with the χ^2 function in several chosen points; e.g., a quadratic polynomial interpolating the χ^2 function in three points leads to the derivative in Eq. (2.28). If the χ^2 function suffers from numerical noise, the interpolated polynomial suffers as much if not more.

We adopt a different approach, and instead of interpolating N points by a polynomial of the order $N-1$, we allow a polynomial to assume different values in these N points and approximate the χ^2 function by the method of least squares. This approach assumes that the order of the polynomial M has to be strictly less than $N-1$ where N is the number of points. If we use a quadratic polynomial to fit seven symmetrically chosen, equidistant points of the χ^2 function, we obtain the following prescriptions for the seven-point low-noise derivative

$$\frac{df}{dx} = \frac{f_1 - f_{-1} + 2(f_2 - f_{-2}) + 3(f_3 - f_{-3})}{28h}. \quad (2.29)$$

Using these derivatives instead of the standard derivative from Eq. (2.28) and extending this approach to the second derivatives allows us to determine the Hessian with sufficient precision and to eliminate the influence of the numerical noise.

2. Error PDFs

To translate the uncertainties contained in the data to the underlying PDF parameters, we use the fact that the χ^2 function in the diagonalized Hessian approximation is a simple function of the parameters $\tilde{z}_k$. Varying data within their errors corresponds to a change in χ^2 (denoted by $\Delta\chi^2$) which can then in turn be interpreted as a shift in the parameters $\tilde{z}_k$,

$$\tilde{z}_k = \pm\sqrt{\Delta\chi^2},$$
$$z_k = \pm\sqrt{\frac{\Delta\chi^2}{\lambda_k}}, \quad k = 1, 2, \ldots, n. \tag{2.30}$$

A specific change in χ^2 can be obtained by varying the parameters using n independent directions in the parameter space.[8] In the $\tilde{z}_k$ space, all directions are equivalent, so we can choose the n independent directions to coincide with the directions where one single parameter is varied. A change in one direction along one single parameter $\tilde{z}_k$ leads to a simultaneous change in all original parameters a_i,

$$y_i \equiv \Delta a_i = \pm\sqrt{\frac{\Delta\chi^2}{\lambda_k}} V_i^{(k)}. \tag{2.31}$$

The parameter shifts along the direction of the $\tilde{z}_k$ parameter are used to generate $2n$ error PDFs for a specified $\Delta\chi^2$,

$$f_k^{\pm} \equiv f\left(a_i^0 \pm \sqrt{\frac{\Delta\chi^2}{\lambda_k}} V_i^{(k)}\right), \quad \text{for } k = 1, 2, \ldots, n. \tag{2.32}$$

The error PDFs can be used to determine the PDF uncertainty of any observable X which depends on PDFs. This uncertainty, which we denote as ΔX, can be determined in different ways, and in this work we define it by adding errors in quadrature,

$$\Delta X = \frac{1}{2}\sqrt{\sum_k (X(f_k^+) - X(f_k^-))^2}. \tag{2.33}$$

The PDF uncertainty ΔX clearly depends on the exact value chosen for $\Delta\chi^2$. In an ideal case, an increase of χ^2 corresponding to one standard deviation from the central value is $\Delta\chi^2 = 1$. However, in our fit, we combine results from different experiments which are not necessarily uncorrelated or compatible, so the standard argument does not apply, and $\Delta\chi^2$ may be different from 1. To estimate what is the appropriate value for the $\Delta\chi^2$ (often referred to as the tolerance), we use a criterion similar to the one advocated in Refs. [12,46,47], which results in the value $\Delta\chi^2 = 35$. Additionally, since the value of our tolerance is far from 1, the quadratic approximation of the Hessian method becomes less precise. We account for it by introducing an additional procedure of rescaling of the Hessian matrix. Both the rescaling procedure and the criterion for choosing the $\Delta\chi^2$ tolerance are described in detail in Appendix A.

[8]If one allows only positive changes of parameters, there are $2n$ directions.

III. EXPERIMENTAL DATA

In the current analysis, we use DIS data, DY lepton pair production data, and inclusive pion production data from the RHIC (for nuclei with $A > 2$). The details of particular experiments such as the number of data points, measured observables, etc., are summarized in Tables I–IV.

The reason to include data from different processes is that each process helps constrain different combinations of parton distributions. The bulk of our data are from DIS which helps pin down the valence and sea distributions; however, they are not very sensitive to different quark flavors and gluons. The DY data can be used to differentiate

TABLE I. The DIS F_2^A/F_2^D data sets used in the nCTEQ15 fit. The table details values of χ^2 for each experiment, the specific nuclear targets, references, and the number of data points with and without kinematic cuts.

F_2^A/F_2^D					# data after	
Observable	Experiment	ID	Ref.	# data	cuts	χ^2
D	NMC-97	5160	[48]	292	201	247.73
He/D	Hermes	5156	[49]	182	17	13.45
	NMC-95,re	5124	[50]	18	12	9.78
	SLAC-E139	5141	[51]	18	3	1.42
Li/D	NMC-95	5115	[52]	24	11	6.10
Be/D	SLAC-E139	5138	[51]	17	3	1.37
C/D	FNAL-E665-95	5125	[53]	11	3	1.44
	SLAC-E139	5139	[51]	7	2	1.36
	EMC-88	5107	[54]	9	9	7.41
	EMC-90	5110	[55]	9	0	0.00
	NMC-95	5113	[52]	24	12	8.40
	NMC-95,re	5114	[50]	18	12	13.29
N/D	Hermes	5157	[49]	175	19	9.92
	BCDMS-85	5103	[56]	9	9	4.65
Al/D	SLAC-E049	5134	[57]	18	0	0.00
	SLAC-E139	5136	[51]	17	3	1.14
Ca/D	NMC-95,re	5121	[50]	18	12	11.54
	FNAL-E665-95	5126	[53]	11	3	0.94
	SLAC-E139	5140	[51]	7	2	1.63
	EMC-90	5109	[55]	9	0	0.00
Fe/D	SLAC-E049	5131	[58]	14	2	0.78
	SLAC-E139	5132	[51]	23	6	7.76
	SLAC-E140	5133	[59]	10	0	0.00
	BCDMS-87	5101	[60]	10	10	5.77
	BCDMS-85	5102	[56]	6	6	2.56
Cu/D	EMC-93	5104	[61]	10	9	4.71
	EMC-93(chariot)	5105	[61]	9	9	4.88
	EMC-88	5106	[54]	9	9	3.39
Kr/D	Hermes	5158	[49]	167	12	9.79
Ag/D	SLAC-E139	5135	[51]	7	2	1.60
Sn/D	EMC-88	5108	[54]	8	8	17.20
Xe/D	FNAL-E665-92	5127	[62]	10	2	0.72
Au/D	SLAC-E139	5137	[51]	18	3	1.74
Pb/D	FNAL-E665-95	5129	[53]	11	3	1.20
Total:				1205	414	403.70

TABLE II. The DIS $F_2^A/F_2^{A'}$ data sets used in the nCTEQ15 fit. We list the same details for each data set as in Table I.

$F_2^A/F_2^{A'}$ Observable	Experiment	ID	Ref.	# data	# data after cuts	χ^2
C/Li	NMC-95,re	5123	[50]	25	7	5.56
Ca/Li	NMC-95,re	5122	[50]	25	7	1.11
Be/C	NMC-96	5112	[63]	15	14	4.08
Al/C	NMC-96	5111	[63]	15	14	5.39
Ca/C	NMC-95,re	5120	[50]	25	7	4.32
	NMC-96	5119	[63]	15	14	5.43
Fe/C	NMC-96	5143	[63]	15	14	9.78
Sn/C	NMC-96	5159	[64]	146	111	64.44
Pb/C	NMC-96	5116	[63]	15	14	7.74
Total:				296	202	107.85

TABLE III. The Drell-Yan process data sets used in the nCTEQ15 fit. We list the same details for each data set as in Table I.

$\sigma_{DY}^{pA}/\sigma_{DY}^{pA'}$: Observable	Experiment	ID	Ref.	# data	# data after cuts	χ^2
C/H2	FNAL-E772-90	5203	[65]	9	9	7.92
Ca/H2	FNAL-E772-90	5204	[65]	9	9	2.73
Fe/H2	FNAL-E772-90	5205	[65]	9	9	3.17
W/H2	FNAL-E772-90	5206	[65]	9	9	7.28
Fe/Be	FNAL-E886-99	5201	[66]	28	28	23.09
W/Be	FNAL-E886-99	5202	[66]	28	28	23.62
Total:				92	92	67.81

TABLE IV. The pion production data sets used in the nCTEQ15 fit. We list the same details for each data set as in Table I.

$R_{dAu}^{\pi}/R_{pp}^{\pi}$: Observable	Experiment	ID	Ref.	# data	# data after cuts	χ^2
dAu/pp	PHENIX	PHENIX	[67]	21	20	6.63
	STAR-2010	STAR	[68]	13	12	1.41
Total:				34	32	8.04

between u and d quark flavors, and the inclusive pion data have a potential to better constrain the gluon distribution.[9]

We introduce kinematic cuts on the included data which limit possible effects of higher twist contributions and target-mass corrections and at the same time are compatible with the kinematic cuts used in the underlying free-proton analysis. The cuts used in this analysis are:

(i) DIS: $Q > 2$ GeV and $W > 3.5$ GeV,

[9]Note that the inclusive pion production observable is different in the sense that it has an additional dependence on a fragmentation function.

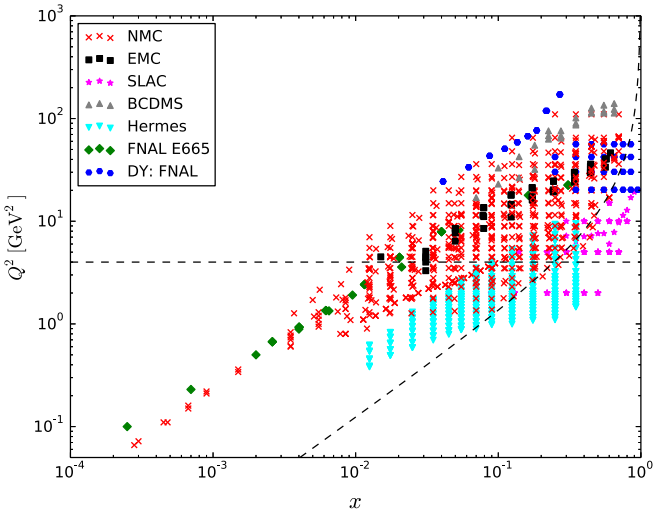


FIG. 1. Kinematic reach of DIS and DY data used in the presented nCTEQ fits. The dashed lines represent the kinematic cuts employed in this analysis ($Q > 2$ GeV, $W > 3.5$ GeV). Only the data points lying above both of these lines are included in the fits.

(ii) DY: $2 < M < 300$ GeV (where M is the invariant mass of the produced lepton pair),

(iii) π^0 production: $p_T > 1.7$ GeV.

After the cuts are applied, 740 data points remain, including 616 DIS, 92 DY, and 32 pion production data points.

Note that the overall number of data points we use is considerably smaller compared to the number of data fitted by other groups (e.g., EPS [12] has 929 data points). One reason is that the other analyses employ less stringent kinematic cuts on Q^2:

(i) EPS [12]: $Q > 1.3$ GeV,
(ii) HKN [14]: $Q > 1$ GeV,
(iii) DSSZ [11]: $Q > 1$ GeV.

In addition, none of the analyses mentioned above employs a cut on W. Whereas the looser cuts allow one to use more data in the fit, there are possible disadvantages connected to this choice. In particular, if one adopts loose cuts, one runs into the danger that the contributions from the target-mass effects or higher twist effects can get enhanced. Especially, the latter effects may be more important in the nuclear case due to the higher density of spectator partons in the nucleus [15,16], and so their effect can be easily underestimated. However, the effect of higher twist and target-mass corrections has been shown to be weakened in ratios of observables [69,70].

The kinematic reach of the DIS and DY data sets used in our fit is summarized in Fig. 1, where individual experimental points are shown in the (x, Q^2) plane. Note that the two dashed lines indicate the kinematic cuts; points lying below these lines are excluded from our analysis.

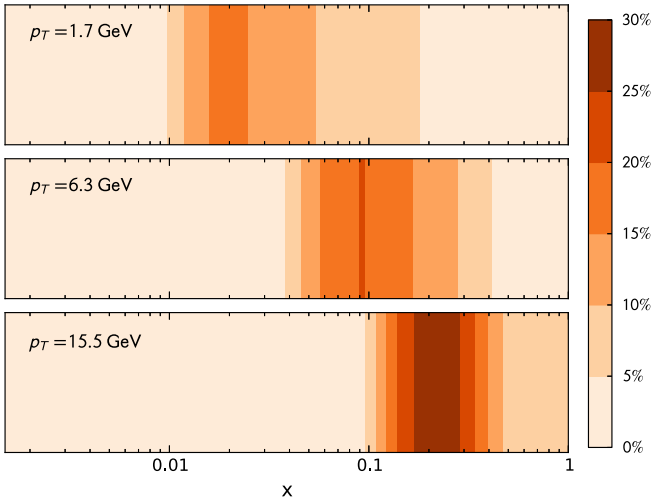


FIG. 2. Approximate x-range for the pion data with the Binnewies-Kniehl-Kramer fragmentation function.

In Fig. 2, we estimate the kinematic impact of the pion data by plotting the cross section for inclusive pion production before convoluting it with gold PDFs; see Eq. (2.4). Figure 2 shows the normalized cross section as a function of the Bjorken-x of a parton inside a nucleon of a gold atom. This is only an estimate which uses the leading order prediction, and it also depends on the fragmentation function (FF) that is used. Nevertheless, it is useful and allows us to see that the x-values probed by the pion data depend quite substantially on the p_T. In particular, for higher p_T, higher x values are probed, e.g., for $p_T \sim 15$ GeV, we are mostly sensitive to $x \in (0.2, 0.3)$, whereas for lower p_T, the probed x values are more diffused, e.g., for $p_T \sim 2$ GeV $x \in (0.01, 0.04)$.

One should mention that there are still experimental data that could have been included in our analysis, but we have decided for different reasons to exclude them from the current work. We comment briefly on the two most important examples.

First, there are neutrino DIS data from CDHSW [71], CHORUS [72], and in particular the NuTeV Collaboration [73]. Since they include a considerable number of data points and probe more flavor combinations than the charged lepton data, they can be used to differentiate individual flavors. However, tensions between the inclusive charged-current νA DIS data from NuTeV and the neutral current $\ell^{\pm}A$ data found in Refs. [13,20,21] indicate that some additional effort is required to understand how these discrepancies can be resolved so that all data could be used in one fit simultaneously. Since these discrepancies appear only if one takes into account the full information contained in the correlated error matrix, neglecting these correlations makes it possible to combine νA and $\ell^{\pm}A$ DIS in one fit [11,22,74]. We plan to revisit the neutrino data in a future publication but decided not to include them in our present PDF release.

Another important set of data which could be included are the already available LHC data. In particular, the cleanest probe of nuclear effects at the LHC comes from the vector boson, $W^{\pm}$, Z, production [75–78]. Results on asymmetries in pPb collisions [78] in particular have a potential to provide valuable input for nuclear PDF analyses. These data are not included in the current release as we first want to provide a baseline analysis without any LHC data.

IV. RESULTS

A key result of the current nCTEQ15 fit compared to the previous nCTEQ releases [13,20] is the inclusion of PDF uncertainties using the Hessian method, cf. Sec. II C.

The second significant addition is the inclusion of a new type of experimental data, namely the pion production data from the PHENIX and STAR collaborations. Since these data have the potential to provide information on the gluon distribution (which otherwise is weakly constrained), it is important to precisely estimate their impact on the resulting PDFs. For this purpose, the nCTEQ15 fit will be compared with a reference fit nCTEQ15-np which is identical except it does *not* include the pion data.

The full set of data we consider is listed in Tables I–IV. Note that we have included QED radiative corrections for the DIS FNAL-E665-95 (Pb/D, Ca/D, C/D) data sets, and this significantly improves the description of these data.[10] In the following, we discuss the results of this nCTEQ15 analysis and compare it with other available sets of nuclear PDFs.

A. nCTEQ15 fit

1. PDF Parametrization

The PDFs in our fit are parametrized at the input scale $Q_0 = 1.3$ GeV according to Eqs. (2.5) and (2.6). This provides considerable flexibility as each of the seven flavor combinations can have ~10 free parameters to describe the x and A dependence.[11] However, the available experimental data are not sufficient to constrain such a flexible parametrization. Therefore, we limit our actual fit to 16 parameters; specifically, we include seven gluon, four u-valence, three d-valence and two $\bar{d} + \bar{u}$ free parameters. The details of the fit are summarized in Table V which shows the best fit values of the free parameters as well as the values of the fixed parameters.

[10]For example, the χ^2 for the FNAL-E665-95 Pb/D data (ID 5129) is reduced from 5.91 to 1.20 (for three data points) when the QED radiative corrections are included.

[11]For each of the five flavor combinations $\{u_v, d_v, g, \bar{u} + \bar{d}, s = \bar{s}\}$ of Eq. (2.5), we have ten parameters $\{c_{k,1}, c_{k,2}\}$ for $k = \{1 \ldots 5\}$ in addition to the normalization parameters c_0 that are partly fixed by the number and momentum sum rules. For $\{\bar{d}/\bar{u}\}$, we have eight parameters at our disposal.

TABLE V. Values of the parameters of the nCTEQ15 fit at the initial scale $Q_0 = 1.3$ GeV. Values in bold represent the free parameters, and values in parentheses are fixed in the fit. The normalization parameters not listed are determined by the momentum and number sum rules as discussed in the text. For completeness, we provide the full set of the free-proton parameters $c_{k,0}$ (first set of rows). The M^i parameters (first row) show the (fixed) momentum fraction carried by different flavors in the case of a free proton.

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
M^g	(0.382)	M^{u_v}	(0.327)	M^{d_v}	(0.136)	$M^{\bar{d}+\bar{u}}$	(0.129)	$M^{s+\bar{s}}$	(0.026)	$M^{\bar{d}/\bar{u}}$	(0.000)
$c^g_{0,0}$	(0.000)	–	–	–	–	–	–	$c^{s+\bar{s}}_{0,0}$	(0.500)	–	–
$c^g_{1,0}$	(0.523)	$c^{u_v}_{1,0}$	(0.630)	$c^{d_v}_{1,0}$	(0.513)	$c^{\bar{d}+\bar{u}}_{1,0}$	(−0.324)	$c^{s+\bar{s}}_{1,0}$	(−0.324)	$c^{\bar{d}/\bar{u}}_{1,0}$	(10.075)
$c^g_{2,0}$	(3.034)	$c^{u_v}_{2,0}$	(2.934)	$c^{d_v}_{2,0}$	(4.211)	$c^{\bar{d}+\bar{u}}_{2,0}$	(8.116)	$c^{s+\bar{s}}_{2,0}$	(8.116)	$c^{\bar{d}/\bar{u}}_{2,0}$	(4.957)
$c^g_{3,0}$	(4.394)	$c^{u_v}_{3,0}$	(−2.369)	$c^{d_v}_{3,0}$	(−2.375)	$c^{\bar{d}+\bar{u}}_{3,0}$	(0.413)	$c^{s+\bar{s}}_{3,0}$	(0.413)	$c^{\bar{d}/\bar{u}}_{3,0}$	(15.167)
$c^g_{4,0}$	(2.359)	$c^{u_v}_{4,0}$	(1.266)	$c^{d_v}_{4,0}$	(0.965)	$c^{\bar{d}+\bar{u}}_{4,0}$	(4.754)	$c^{s+\bar{s}}_{4,0}$	(4.754)	$c^{\bar{d}/\bar{u}}_{4,0}$	(17.000)
$c^g_{5,0}$	(−3.000)	$c^{u_v}_{5,0}$	(1.718)	$c^{d_v}_{5,0}$	(3.000)	$c^{\bar{d}+\bar{u}}_{5,0}$	(0.614)	$c^{s+\bar{s}}_{5,0}$	(0.614)	$c^{\bar{d}/\bar{u}}_{5,0}$	(9.948)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
$c^g_{0,1}$	(−0.256)	–	–	–	–	–	–	$c^{s+\bar{s}}_{0,1}$	(0.167)	–	–
$c^g_{1,1}$	**−0.001**	$c^{u_v}_{1,1}$	**−2.729**	$c^{d_v}_{1,1}$	**0.272**	$c^{\bar{d}+\bar{u}}_{1,1}$	**0.411**	$c^{s+\bar{s}}_{1,1}$	(0.411)	$c^{\bar{d}/\bar{u}}_{1,1}$	(0.000)
$c^g_{2,1}$	(0.000)	$c^{u_v}_{2,1}$	**−0.162**	$c^{d_v}_{2,1}$	**−0.198**	$c^{\bar{d}+\bar{u}}_{2,1}$	(0.415)	$c^{s+\bar{s}}_{2,1}$	(0.415)	$c^{\bar{d}/\bar{u}}_{2,1}$	(0.000)
$c^g_{3,1}$	(0.383)	$c^{u_v}_{3,1}$	(0.018)	$c^{d_v}_{3,1}$	(0.085)	$c^{\bar{d}+\bar{u}}_{3,1}$	(−0.759)	$c^{s+\bar{s}}_{3,1}$	(0.000)	$c^{\bar{d}/\bar{u}}_{3,1}$	(0.000)
$c^g_{4,1}$	**0.055**	$c^{u_v}_{4,1}$	**12.176**	$c^{d_v}_{4,1}$	(3.874)	$c^{\bar{d}+\bar{u}}_{4,1}$	(−0.203)	$c^{s+\bar{s}}_{4,1}$	(0.000)	$c^{\bar{d}/\bar{u}}_{4,1}$	(0.000)
$c^g_{5,1}$	**0.002**	$c^{u_v}_{5,1}$	**−1.141**	$c^{d_v}_{5,1}$	**−0.072**	$c^{\bar{d}+\bar{u}}_{5,1}$	**−0.087**	$c^{s+\bar{s}}_{5,1}$	(0.000)	$c^{\bar{d}/\bar{u}}_{5,1}$	(0.000)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
$c^g_{0,2}$	**−0.037**	–	–	–	–	–	–	$c^{s+\bar{s}}_{0,2}$	(0.104)	–	–
$c^g_{1,2}$	**−1.337**	$c^{u_v}_{1,2}$	(0.006)	$c^{d_v}_{1,2}$	(0.466)	$c^{\bar{d}+\bar{u}}_{1,2}$	(0.172)	$c^{s+\bar{s}}_{1,2}$	(0.172)	$c^{\bar{d}/\bar{u}}_{1,2}$	(0.000)
$c^g_{2,2}$	(0.000)	$c^{u_v}_{2,2}$	(0.524)	$c^{d_v}_{2,2}$	(0.440)	$c^{\bar{d}+\bar{u}}_{2,2}$	(0.290)	$c^{s+\bar{s}}_{2,2}$	(0.290)	$c^{\bar{d}/\bar{u}}_{2,2}$	(0.000)
$c^g_{3,2}$	(0.520)	$c^{u_v}_{3,2}$	(0.073)	$c^{d_v}_{3,2}$	(0.107)	$c^{\bar{d}+\bar{u}}_{3,2}$	(0.298)	$c^{s+\bar{s}}_{3,2}$	(0.000)	$c^{\bar{d}/\bar{u}}_{3,2}$	(0.000)
$c^g_{4,2}$	**−0.514**	$c^{u_v}_{4,2}$	(0.038)	$c^{d_v}_{4,2}$	(−0.018)	$c^{\bar{d}+\bar{u}}_{4,2}$	(0.888)	$c^{s+\bar{s}}_{4,2}$	(0.000)	$c^{\bar{d}/\bar{u}}_{4,2}$	(0.000)
$c^g_{5,2}$	**−1.417**	$c^{u_v}_{5,2}$	(0.615)	$c^{d_v}_{5,2}$	(−0.236)	$c^{\bar{d}+\bar{u}}_{5,2}$	(1.353)	$c^{s+\bar{s}}_{5,2}$	(0.000)	$c^{\bar{d}/\bar{u}}_{5,2}$	(0.000)

For the pion data, we allow for the normalization to vary, and we obtain 1.031 for the PHENIX data [67] and 0.962 for the STAR data [68].[12] Our obtained normalization shifts of ∼4% lie well within the experimental normalization uncertainty.[13]

Our parametrization smoothly interpolates between different nuclei as a function of the nuclear mass number A; the number of protons Z and neutrons $(A - Z)$ enters only through the isospin composition of a nucleus, cf. Eq. (2.1). Figure 3 shows the A dependence of the fitting parameters normalized by the corresponding values of the free-proton baseline parameters $c_{k,0}$. (Note that some of these parameters are fixed, cf. Table V.) Many of the parameters change rapidly in the region of light nuclei $A \lesssim 25$ and are relatively stable for heavy nuclei $A \gtrsim 50$. Also, we observe that the parameters responsible for the small x behavior $\{c_1\}$ typically exhibit a strong A dependence, whereas the large x parameters $\{c_2\}$ are comparably insensitive to the type of nucleus. In particular, the biggest effect occurs for the gluon where the c_1^g parameter describing the low-x gluon PDF and c_5^g parameter (responsible for mid-x) are changing linearly throughout the whole range of A.

2. χ^2 of the fit

We now examine the overall statistical quality of the fit as measured by the χ^2. For the nCTEQ15 fit, we obtain a total χ^2 of 587.4 with 740 data points (after kinematic cuts). With 18 free parameters (including two data normalization parameters), this leads to a $\chi^2/\text{dof} = 0.81$ which indicates a good fit. Furthermore, this χ^2/dof is not too small which could indicate deficiencies of the fit such as overfitting.

To better evaluate the fit quality, in Fig. 4(a) we plot the χ^2/dof for the individual experiments and check that the majority of experiments has a $(\chi^2/\text{dof}) \simeq 1$. While most experiments satisfy this "goodness of fit" criterion, there is one experiment that stands out as having a poor fit: the DIS EMC-88 data for Sn/D (ID 5108). Several previous global

[12]Note that the data normalization parameters do not enter the Hessian analysis of uncertainties.

[13]See Table 1 and Fig. 2 in Ref. [67] for PHENIX and Fig. 25 in Ref. [68] and Table 5 in Ref. [79] for STAR.

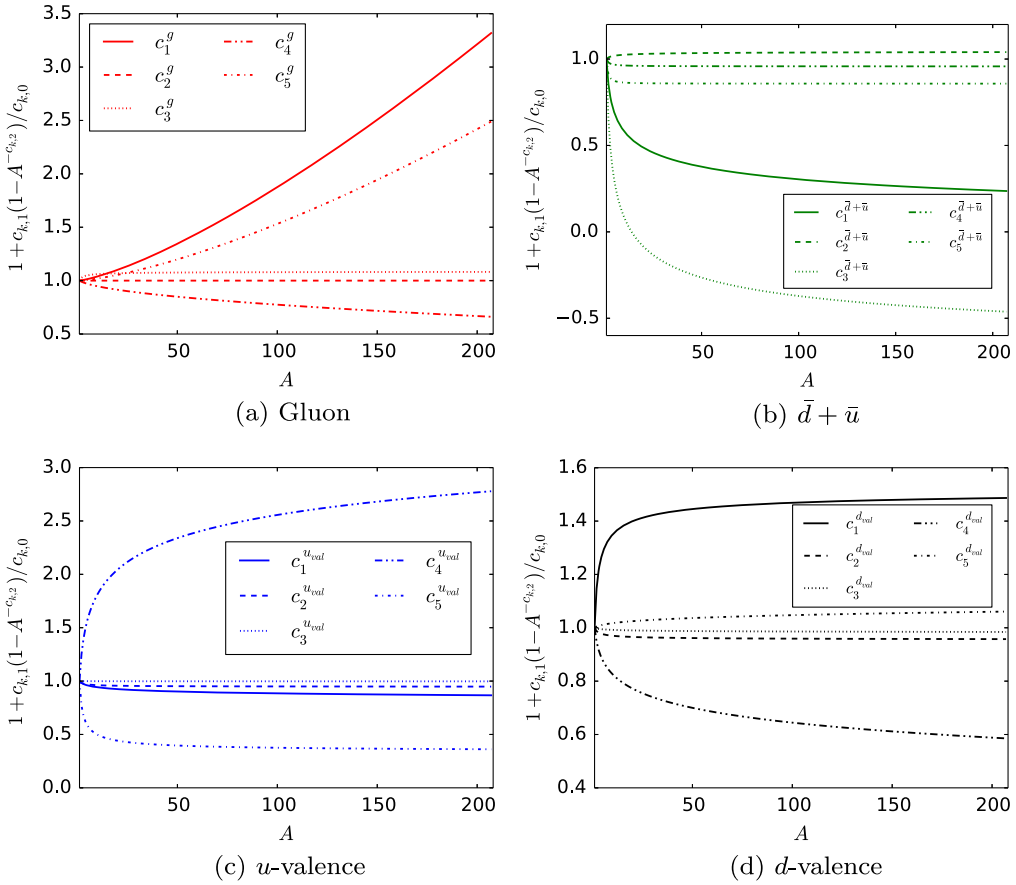


FIG. 3. A dependence of the fit parameters as given in Eq. (2.6). Specifically, we plot $c_k(A) = c_{k,0} + c_{k,1}(1 - A^{-c_{k,2}})$ for each flavor normalized to the corresponding free-proton parameter $c_{k,0}$. The superscripts $\{1, 2, \ldots\}$ in the legend correspond to the parameters $\{c_1, c_2, \ldots\}$ in Eq. (2.5).

analyses have also found it challenging to accommodate the Sn/D data [11,14].

In Fig. 4(b), we show again the χ^2/dof, but this time the experiments are grouped by nuclear target and are sorted by increasing nuclear mass number A. This allows us to see that there are no systematic effects associated with our choice of the A parametrization. With the noted exception of Sn/D, all other nuclear targets from helium up to lead are described very well with a $\chi^2/\text{dof} \simeq 1$.

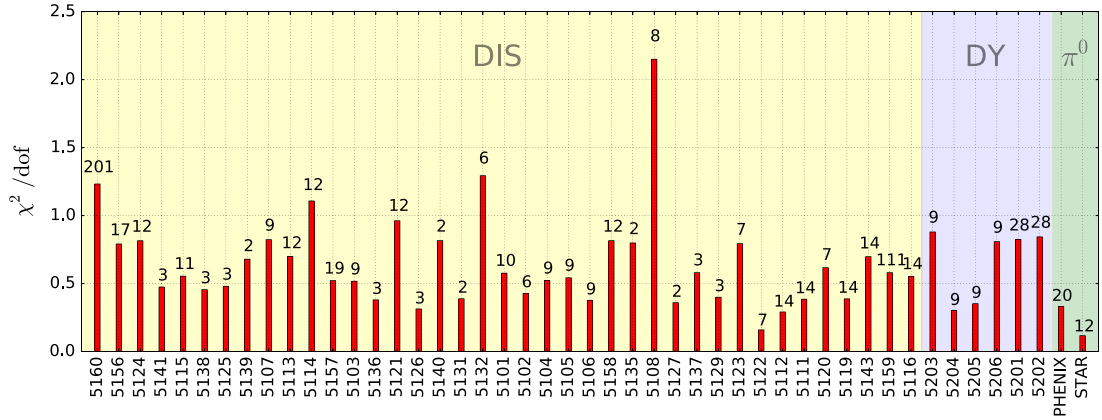
3. Error PDF reliability

Before we examine the actual nCTEQ15 predictions, we first investigate the quality of the Hessian error analysis. This will allow us to judge the reliability of our error estimates and, in turn, the quality of our predictions.[14] There are two factors that need to be assessed:

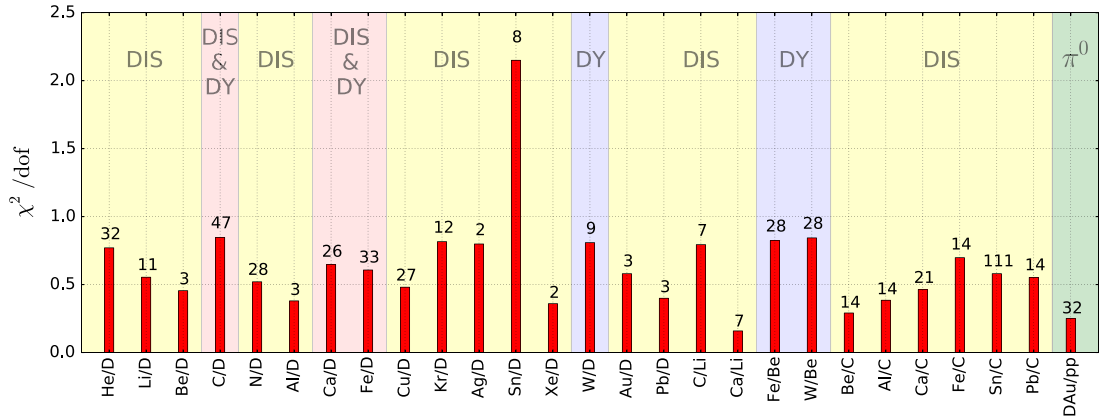
(i) the quality of the quadratic approximation,
(ii) how well the Hessian approximation describes the actual χ^2 function in a region around the minimum given by our tolerance criterion, $\Delta\chi^2 = 35$.

To estimate these factors, we plot the χ^2 function relative to its value at the minimum ($\Delta\chi^2 = \chi^2 - \chi^2_0$) along the 16 error directions in the eigenvector space (see Fig. 5). For comparison, we also display the Hessian approximation given by the quadratic form $\Delta\chi^2 = \tilde{z}_i^2$. The plots are ordered according to the decreasing values of the eigenvalues corresponding to the $\tilde{z}_i$ directions; the largest eigenvalue is of order 10^9, and the smallest is of order 10. For the largest few eigenvalues of Fig. 5, the quadratic approximation works extremely well; however, for the smaller eigenvalues (e.g., numbers 10 and 14), it can deviate from the χ^2 function. Nevertheless, in all the cases, we are able to obtain a good description of the actual χ^2 function for $\tilde{z}_i \sim [-6, 6]$ which corresponds to our tolerance criterion $\sqrt{\Delta\chi^2} = \sqrt{35} \sim 6$. This analysis verifies that the error PDFs defined using the modified Hessian formalism will closely reflect the actual χ^2 function determined by the experimental data and will not be severely affected by the

[14]Note that by construction, the Hessian method can only probe the local minimum connected to the "best fit" (central prediction) and is not sensitive to a landscape with multiple minima. Unfortunately, in case of nPDFs fits, multiple minima are possible as there are not sufficient data to fully constrain the nPDFs.



(a) Value of χ^2/dof for the individual experiments which are identified by the IDs that are listed in Tables I—IV. The `51xx` IDs correspond to the DIS experiments, the `52xx` IDs are the DY data, and the pion data are labeled by the collaboration name. The experiments are sorted left-to-right: {DIS, DY, π^0} and sub-sorted by the nuclear mass number A.



(b) Value of χ^2/dof per nuclear target used in the `nCTEQ15` fit sorted left-to-right by the nuclear mass number A.

FIG. 4. Value of χ^2/dof for (a) individual experiments and (b) per nuclear target used in the nCTEQ15 fit. The numbers on top of the bars represent the number of data points (after kinematic cuts).

imperfections of the quadratic approximation that occurs for directions corresponding to lower eigenvalues.[15]

4. nPDFs vs nuclear A

We now examine the results of the nCTEQ15 fit starting with the A dependence of the various nPDF flavors. In Fig. 6, we display the central fit predictions for a range of nuclear A values from $A = 1$ (proton) to $A = 208$ (lead). When examining the A dependence, we observe that as we move to larger A, the gluon and sea distributions $\{g, \bar{u}, \bar{d}, s\}$ decrease at small x values. This trend is also present for the $\{u, d\}$ PDFs. On the other hand, the A dependence of $\{u_v, d_v\}$ distributions is reduced relative to the other flavor components.

Finally, Figs. 7 and 8 show our nPDFs ($f^{p/Pb}$) for a lead nucleus together with the nuclear correction factors at the input scale $Q = Q_0 = 1.3$ GeV and at $Q = 10$ GeV to show the evolution effects when the PDFs are probed at a typical hard scale. We have chosen to present results for the rather heavy lead nucleus because of its relevance for the heavy ion program at the LHC. In all cases, we display the uncertainty band arising from the error PDF sets based upon our eigenvectors and the tolerance criterion. It should be noted that the uncertainty bands for $x \lesssim 10^{-2}$ and $x \gtrsim 0.7$ are not directly constrained by data but only by the momentum and number sum rules. The uncertainty bands are the result of extrapolating the functional form of our parametrization into these unconstrained regions.

Some comments are in order:

(i) As can be seen from Fig. 7(a), our input gluon is strongly suppressed/shadowed with respect to the free proton in the $x \lesssim 0.04$ region. In fact, it has a valencelike structure [see Fig. 7(b)] which vanishes at small x. Consequently, the steep small x rise of the gluon distribution at $Q = 10$ GeV (see Fig. 8) is entirely due to the QCD evolution. However, we should note that there are no data constraints below

[15]In the modified Hessian approach that we use, the discrepancies at $\Delta\chi^2 = 35$ originate mostly from the nonsymmetric behavior of the χ^2 function; see Sec. II C and Appendix A for details.

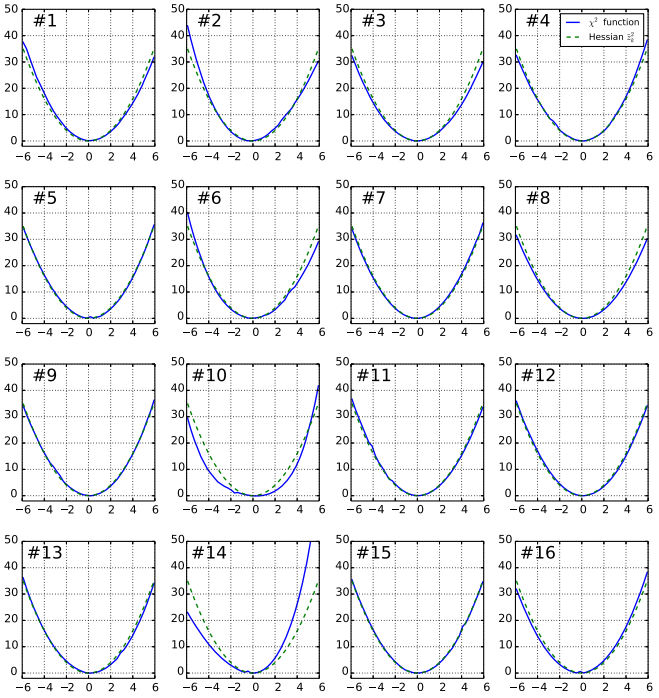


FIG. 5. χ^2 function relative to its value at the minimum, $\Delta\chi^2 = \chi^2 - \chi_0^2$, plotted along the 16 error directions in the eigenvector space, $\tilde{z}_i^2$. We display the true χ^2 function (solid lines) and the quadratic approximation given by Hessian method $\Delta\chi^2 = \tilde{z}_i^2$ (dashed lines). The eigenvector directions are ordered from the largest to the smallest eigenvalue.

$x \sim 0.01$ and the gluon uncertainty in this region is underestimated. In addition, our gluon has an anti-shadowing peak around $x \sim 0.1$ and then exhibits suppression in the EMC region $x \sim 0.5$. However, the large x gluon features a wide uncertainty band reflecting the fact that there are no data constraints.

(ii) In our analysis, we determine the $\bar{u} + \bar{d}$ combination and assume that there is no nuclear modification to the $\bar{d}/\bar{u}$ combination (see Sec. II and Table V). As a result, the $\bar{u}$ and $\bar{d}$ PDFs are very similar, and the small difference between the two comes from the underlying free-proton PDFs.

(iii) In this analysis, we do not fit the strange distribution but relate it to the light quarks sea distribution; see Eq. (2.7). As a result, the strange quark distribution is very similar to the $\bar{u}$ and $\bar{d}$ distributions.

(iv) Contrary to the other existing nPDFs where the nuclear correction factors for the valence distributions are assumed to be the same, we treat u_v and d_v as independent. This leads to an interesting feature of our result where u_v is suppressed and d_v is enhanced in the EMC region. This behavior is not entirely unexpected; there are nuclear models predicting a flavor dependence for the EMC effect [70,80,81].

(v) The above difference for the nuclear correction in u_v and d_v appears at the level of the bound proton PDFs. When we construct a physical combination representing the full nuclear PDF, $f^A = \frac{Z}{A} f^{p/A} + \frac{A-Z}{A} f^{n/A}$, such as lead in Fig. 9, the combination yields net corrections for u_v and d_v which are close to each other and similar to those in the literature. We will discuss this in more detail in Sec. IV E.

Once more data are included, e.g., from the LHC, neutrino DIS experiments, and a future eA collider, it should be possible to relax some of the assumptions.

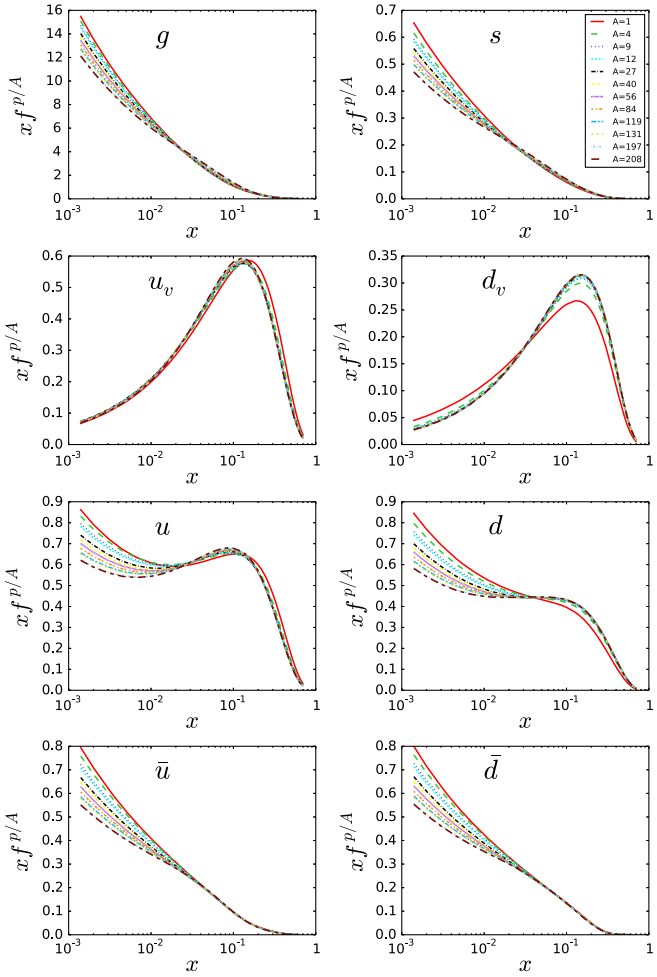


FIG. 6. nCTEQ15 bound proton PDFs at the scale $Q = 10$ GeV for a range of nuclei from the free proton ($A = 1$) to lead ($A = 208$).

In the following section, we will investigate the impact of these nPDFs and the corresponding uncertainty bands on the physical observables.

B. Comparison with data

While the χ^2/dof is one measure of the quality of the fit, this alone obviously does not capture all the relevant characteristics. To investigate the nCTEQ15 result in more

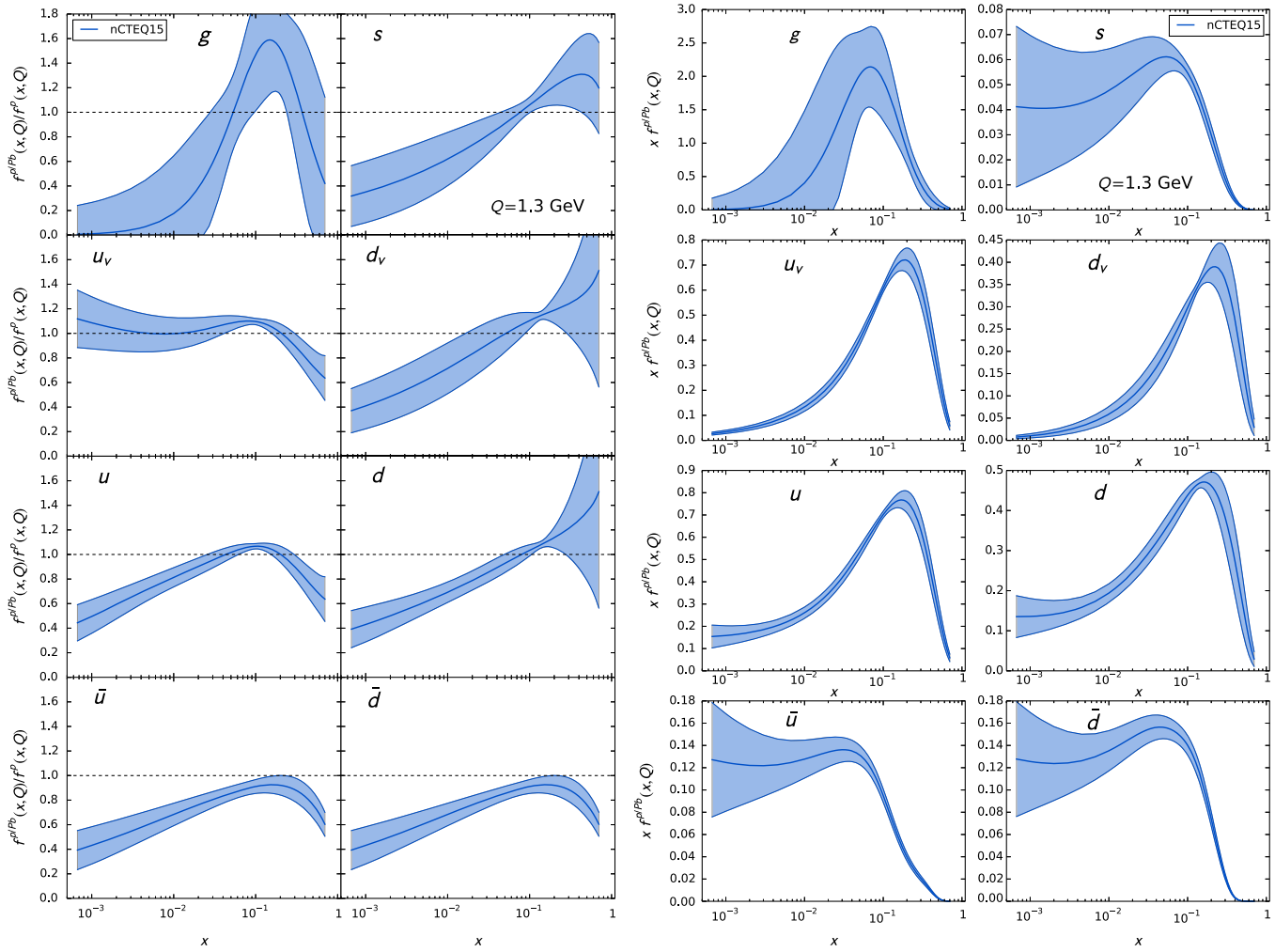


FIG. 7. Results of the nCTEQ15 fit. On the left, we show nuclear modification factors defined as ratios of proton PDFs bound in lead to the corresponding free-proton PDFs, and on the right we show the actual bound proton PDFs for lead. In both cases, the scale is equal to $Q = 1.3$ GeV.

detail, we compare it to the most important and constraining data sets and consider strengths and limitations of both the fit and the available data sets.

1. DIS data sets

The data from the deep inelastic scattering experiments are by far the most numerous and provide the dominant contribution to the total χ^2. These experiments are performed on a variety of nuclei which allow us to constrain the A dependence of our parameters. Most of the data are extracted as a ratio of F_2 structure functions $R = F_2^{A_1}/F_2^{A_2}$ for two different targets A_1 and A_2. Note that in the present study we do not fit data from the very high-x region $x \gtrsim 0.7$ since they do not pass our kinematic cuts. As already mentioned, the high-x region is theoretically challenging due to a host of effects (higher twist, target-mass corrections, large x resummation, deuteron wave function, and nuclear off-shell effects). Some of these effects in the large-x and low-Q^2 area have been investigated extensively in the proton case by the CTEQ-CJ Collaboration [82,83]. The nuclear case is even more challenging due to enhanced higher twist and Fermi motion effects which lead to a steep rise of the structure function ratios in the limit $x \to 1$. For these reasons, we avoid fitting the high-x region for the time being.

The comparison of our fit to the DIS F_2 ratio data is shown in Figs. 10 and 11 as a function of x. Note that in these figures, the data for different Q^2 are combined into a single plot as the scaling violations (discussed later) occur on a logarithmic scale and largely cancel out in the ratios.

Figure 10 shows the ratio $F_2^A(x, Q^2)/F_2^D(x, Q^2)$ for a variety of experiments. The overall agreement of the fit with the data is excellent for a majority of the nuclei. The discrepancy which can be seen for the EMC data taken on tin (Sn/D) is the same discrepancy we have pointed out in Sec. IV A 2 when we investigated the χ^2 of the individual experiments. As already mentioned, this problem has been

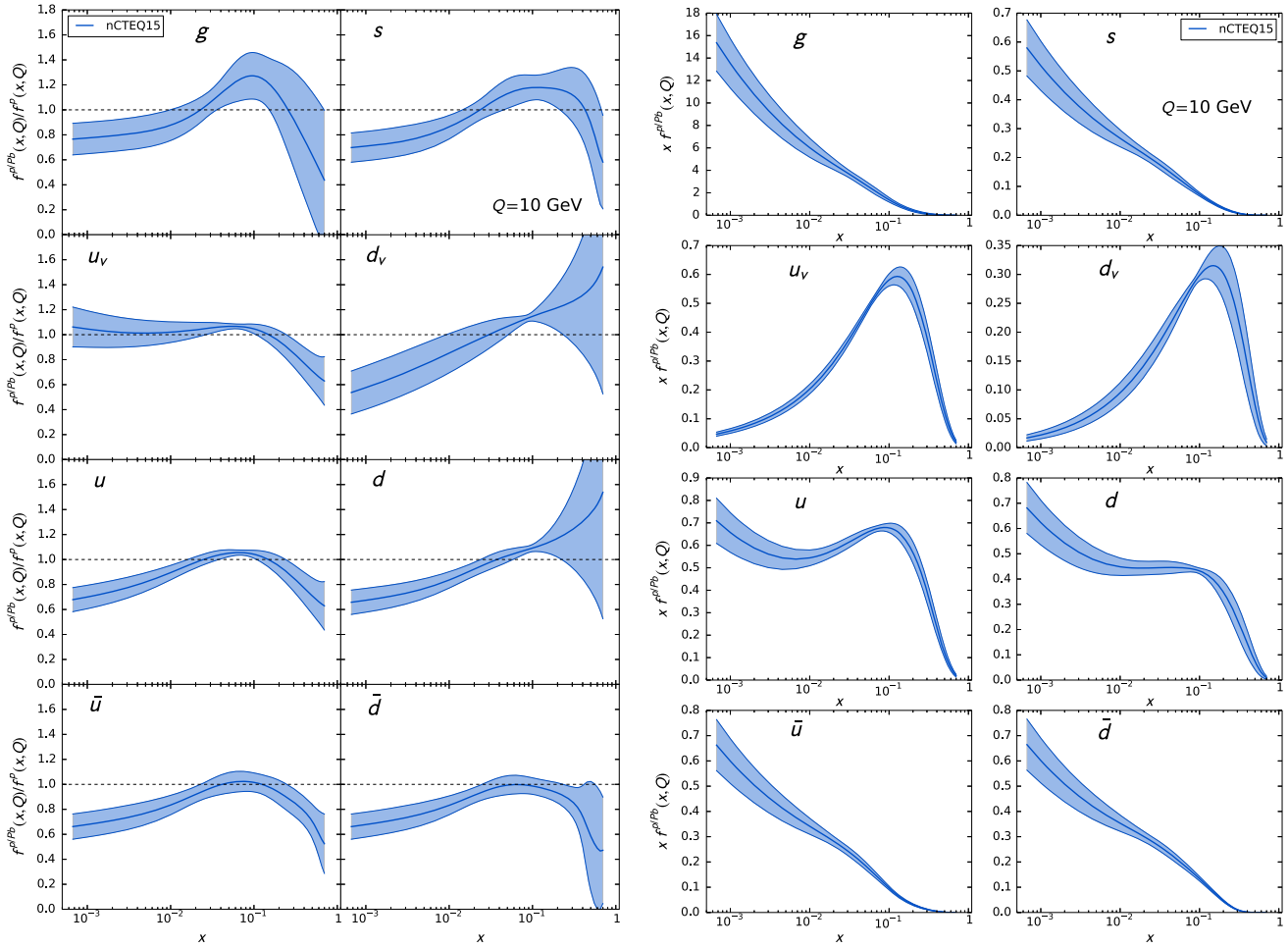


FIG. 8. Results of the nCTEQ15 fit. On the left, we show nuclear modification factors defined as ratios of proton PDFs bound in lead to the corresponding free-proton PDFs, and on the right we show the actual bound proton PDFs for lead. In both cases, the scale is equal to $Q = 10$ GeV.

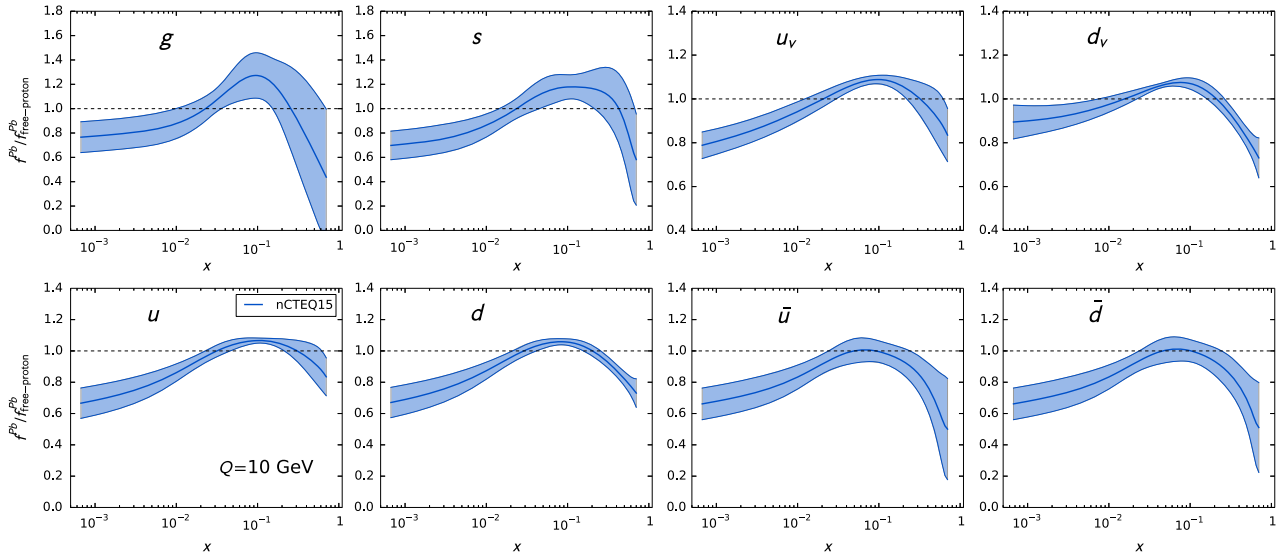


FIG. 9. We show nuclear modification factors defined as ratios of lead PDFs compared to a lead PDF constructed from free-proton PDFs. The PDFs are constructed using $f^A = \frac{Z}{A} f^{p/A} + \frac{A-Z}{A} f^{n/A}$ for ^{207}Pb and the free proton, with a scale of $Q = 10$ GeV.

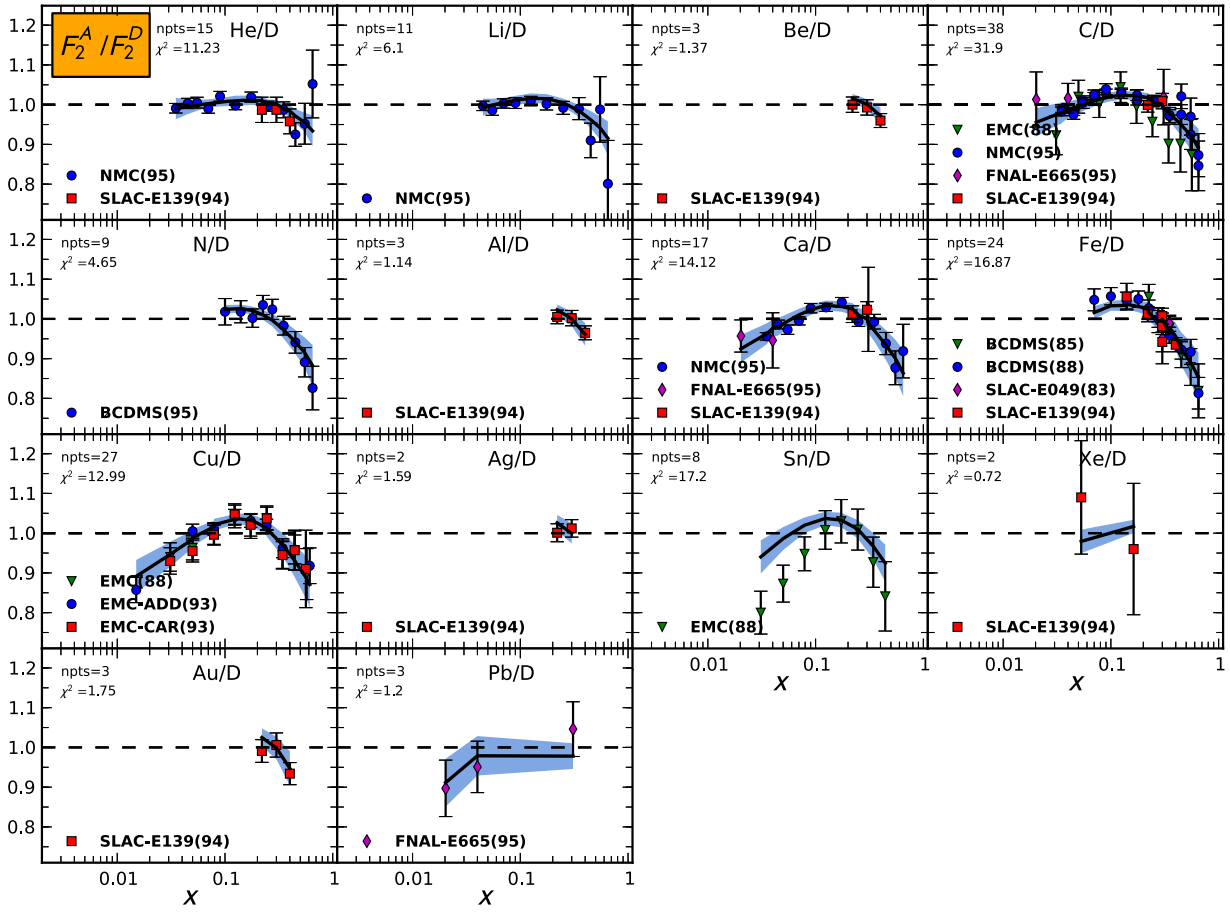


FIG. 10. Comparison of the nCTEQ15 NLO theory predictions for $R = F_2^A(x, Q^2)/F_2^D(x, Q^2)$ as a function of x with nuclear target data. The theory predictions have been calculated at the Q^2 values of the corresponding data points. The bands show the uncertainty from the nuclear PDFs.

also encountered in previous analyses [11,14], and we are unable to reconcile it with our fit.

Similarly, Fig. 11 shows the structure function ratio $F_2^A(x, Q^2)/F_2^{A'}(x, Q^2)$ in comparison to NMC data for a variety of nuclear targets. These high-statistics data are also well described by the results of the nCTEQ15 fit.

The NMC data taken on tin and carbon ($R = F_2^{Sn}/F_2^C$) cover a wider range in Q^2, and we display these in Fig. 12 as a function of Q^2 binned in x. As is well known, the logarithmic Q^2 scaling violations of the structure functions provide constraints on the low-x gluon distribution. Of course, compared to the very precise HERA data on the proton F_2 structure function which extends over a very wide range of Q^2 values, the NMC data have a much smaller Q^2 lever arm. As a consequence, the NMC data provide relatively weaker constraints on the nuclear gluon PDF in the x range of (0.05,0.1). We will discuss data constraints on the gluon in more detail in Sec. IV D.

In Fig. 13, we plot the nuclear correction $R = F_2^{Fe}/F_2^D$ for iron vs x for two Q^2 values and compare the results with experimental data and with results from different nPDF groups. Comparing these two figures, we again see that there is a rather weak Q^2 dependence of the structure function ratio between $Q^2 = 5$ GeV^2 and $Q^2 = 20$ GeV^2. As discussed above, due to our strict kinematic cuts, we do not extend our predictions to the high-x region ($x \gtrsim 0.7$).

Taking into account both the nPDF uncertainty (represented by the error bands) and the experimental error bars, the data are generally compatible with the nCTEQ15 fit. In addition to comparing with data, we compare our predictions with those of HKN [14] and EPS [12] and find a good agreement within the errors of our analysis.

2. Drell-Yan data sets

We now turn to the Drell-Yan muon pair production process $p + A \to \mu^+ + \mu^- + X$. In Fig. 14(a), we display the differential cross section ratio, $R = (d\sigma_{\rm DY}^{pA}/dx_2 dM)/(d\sigma_{\rm DY}^{pD}/dx_2 dM)$, measured by the Fermilab experiment E772, where x_2 is the momentum fraction of the parton inside the nucleus and the invariant mass of the produced muon pair, M, covers the range $\sim(4.5, 13)$ GeV (excluding the charmonium and bottonium resonances). These data

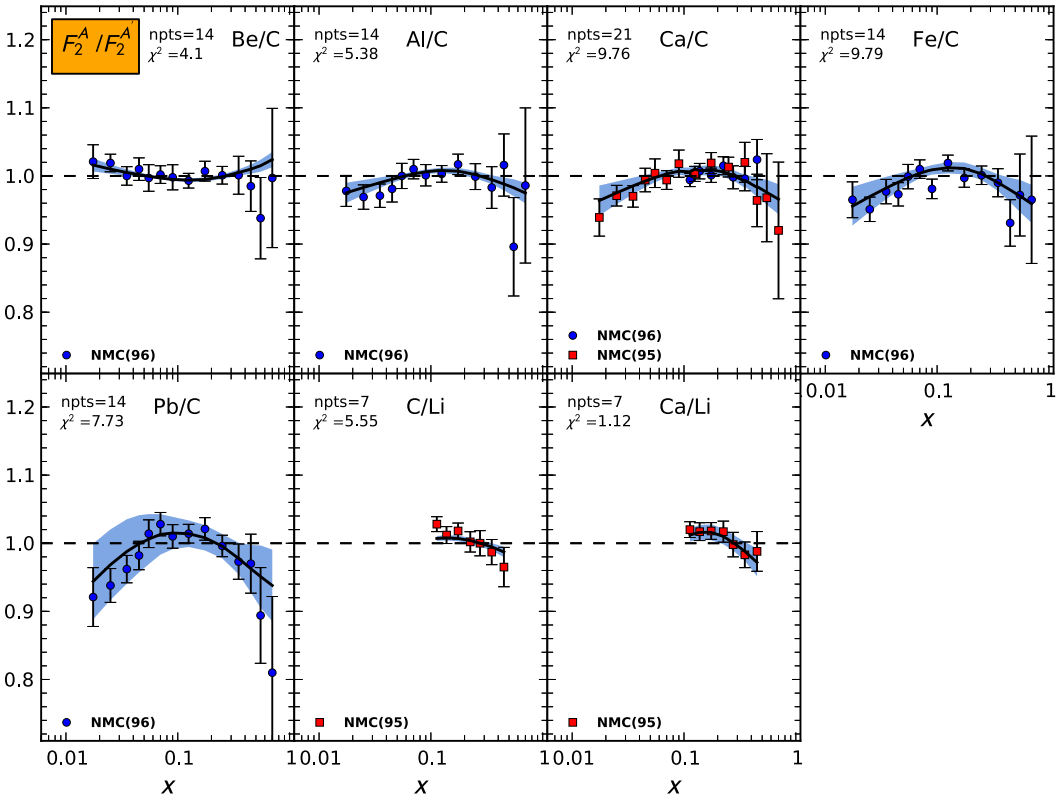


FIG. 11. Same as in Fig. 10 for $R = F_2^A(x, Q^2)/F_2^{A'}(x, Q^2)$.

have been taken for large Feynman $x_F \sim x_1 - x_2$ corresponding to smallish x_2 values.

Similarly, in Fig. 14(b), we present a comparison of our predictions with large x_F data from the E866 experiment for the ratio $R = (d\sigma_{\rm DY}^{pA}/dx_1 dM)/(d\sigma_{\rm DY}^{pD}/dx_1 dM)$. The data are arranged in four bins of the invariant mass ($M = \{4.5, 5.5, 6.5, 7.5\}$ GeV) and are presented as a function of the proton momentum fraction x_1.

As can be seen, the theory predictions describe the data quite well, except for some isolated points (generally those with large error bars).

3. Pion production data sets

The newest addition to the current analysis as compared to Ref. [13] is the ratios of double differential cross sections for single inclusive pion data from the STAR and PHENIX experiments at the RHIC. Specifically, we fit the ratio

$$R_{\rm dAu}^{\pi} = \frac{\frac{1}{2A} d^2\sigma_{\pi}^{\rm dAu}/dp_T dy}{d^2\sigma_{\pi}^{\rm pp}/dp_T dy}, \tag{4.1}$$

and we include only the data measured at central rapidity to exclude potential final-state effects (this criterion excludes any data from BRAHMS). Additionally, we fit the normalizations of the RHIC data and obtain 1.031 and 0.962 for PHENIX and STAR, respectively. These values are within the experimental uncertainty.[16] Fitting the single inclusive pion production has the added complication that it depends on the FFs. As mentioned in Sec. II, precomputed grids of convolutions with the free deuterium PDFs and a set of FFs are used to speed up the NLO calculation.

In Fig. 15(a), PHENIX and STAR data are compared with predictions from the nCTEQ15 fit using the Binnewies-Kniehl-Kramer (BKK) fragmentation functions [84]. As the PHENIX data are more precise than the STAR data, the former will have a correspondingly larger impact on the resulting fit.

The EPS09 analysis [12] also used this data, and we compare with their result in Fig. 15(b). Our central prediction for $R_{\rm dAu}^{\pi}$ differs from EPS09 but lies within their uncertainty band; however, our estimate of the PDF uncertainties differs substantially from EPS09.[17] The main

[16]We note that the EPS09 analysis obtained similar normalizations.

[17]The EPS09 analysis uses a different asymmetric definition of uncertainties given by

$$(\Delta X^+)^2 = \sum_k [\max\{X(S_k^+) - X(S_k^0), X(S_k^-) - X(S_k^0), 0\}]^2,$$

$$(\Delta X^-)^2 = \sum_k [\max\{X(S_k^0) - X(S_k^+), X(S_k^0) - X(S_k^-), 0\}]^2.$$

To make this comparison consistent, we adopt the same definition when comparing with the EPS09 prediction.

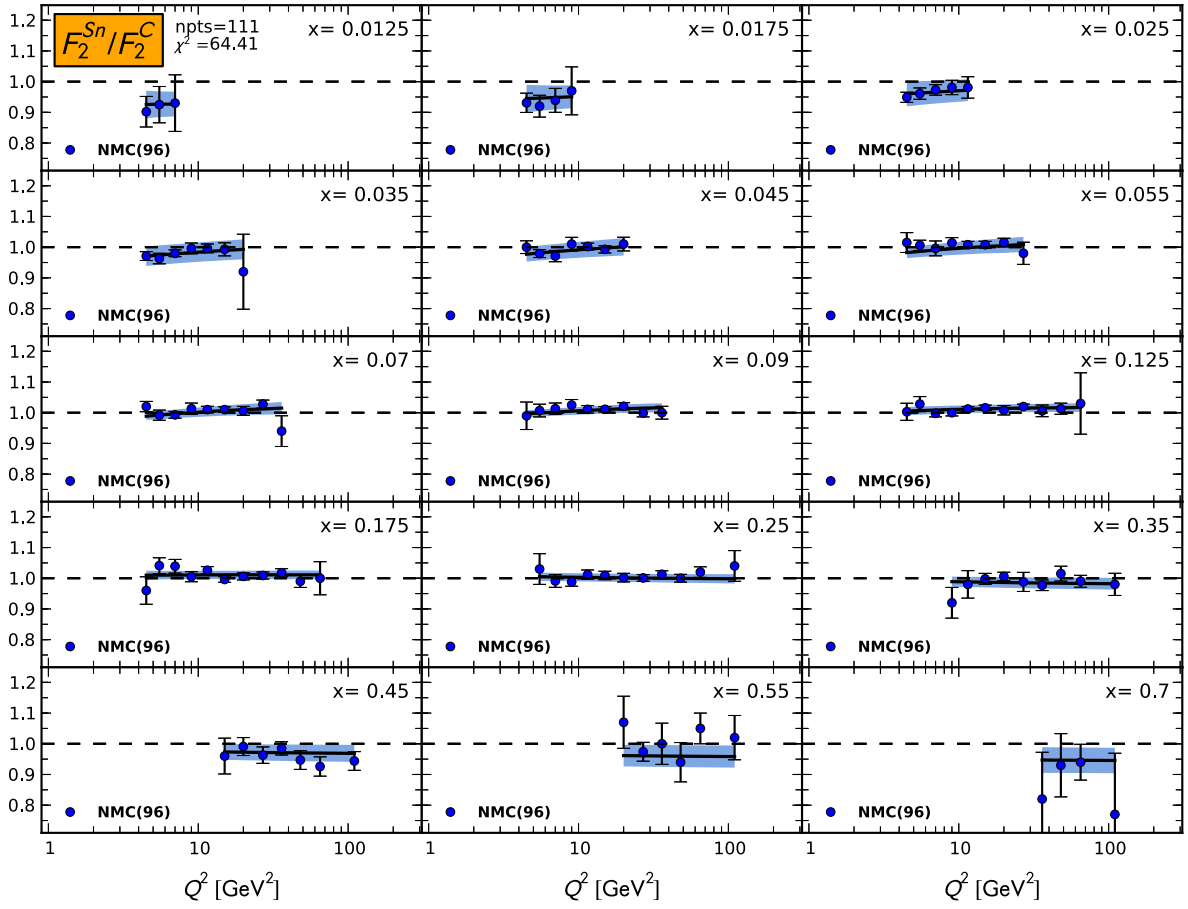


FIG. 12. Comparison of the nCTEQ15 NLO theory predictions for $R = F_2^{Sn}/F_2^C$ as a function of Q^2 with nuclear target data from the NMC Collaboration. The bands show the uncertainty from the nuclear PDFs.

reason for this difference is the fact that EPS09 chooses to include the single inclusive pion data with a large weight (×20) to enhance its importance, and this choice leads to the suppression of the corresponding uncertainties.

Another source of difference can arise from the choice of the fragmentation functions. The EPS09 analysis uses the Kniehl-Kramer-Pötter (KKP) fragmentation functions [85], whereas the nCTEQ15 fit is based on the BKK FFs. To

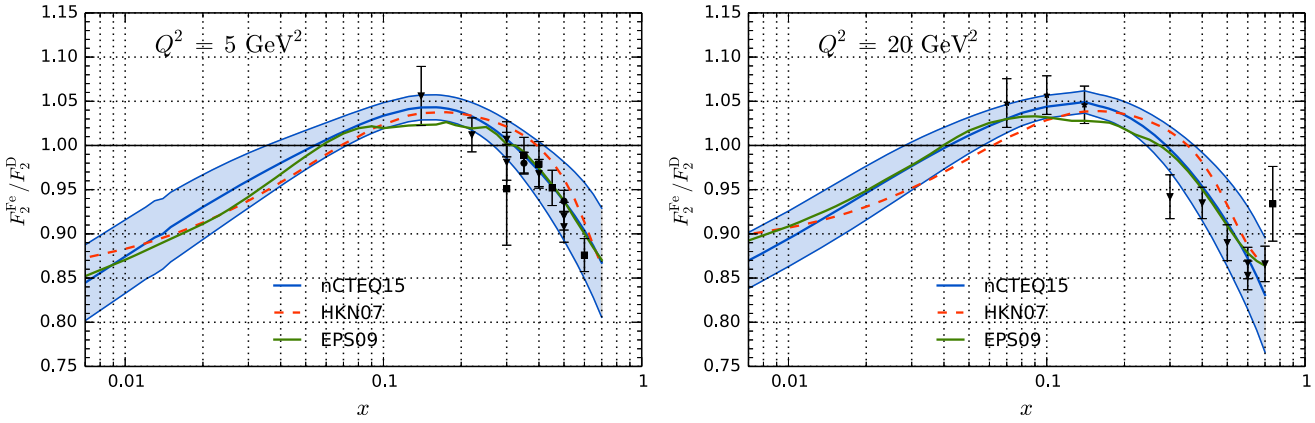


FIG. 13. Ratio of the F_2 structure functions for iron and deuteron calculated with the nCTEQ15 fit at (a) $Q^2 = 5$ GeV^2 and (b) $Q^2 = 20$ GeV^2. This is compared with the fitted data from SLAC-E049 [57], SLAC-E139 [51], SLAC-E140 [59], BCDMS-85 [56], and BCDMS-87 [60] experiments and results from EPS09 and HKN07. (The data points shown are within 50% of the nominal Q^2 value.)

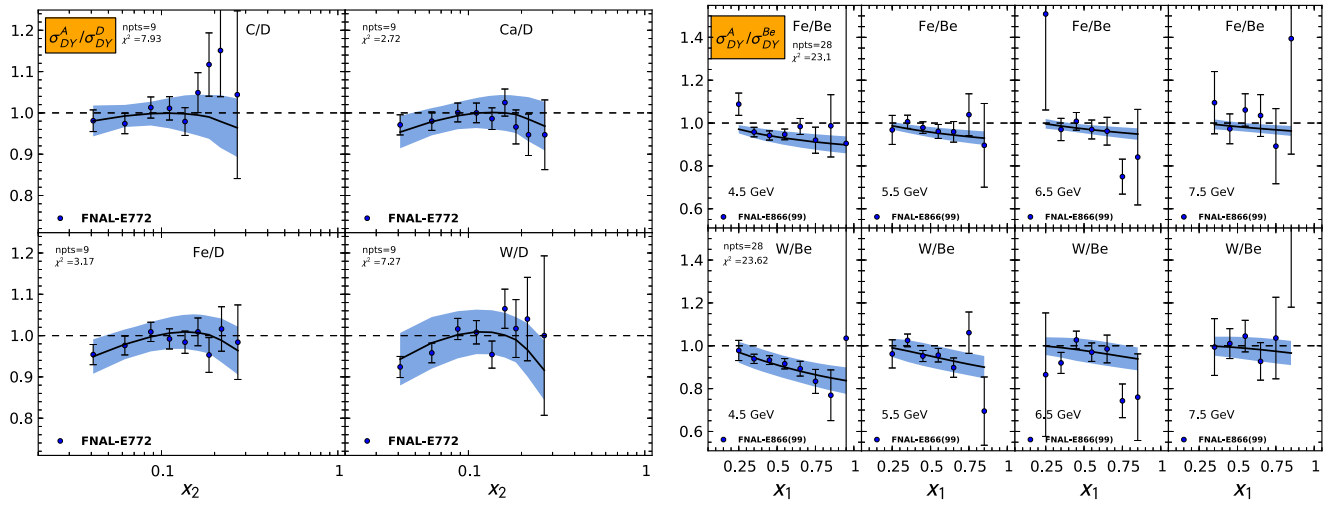
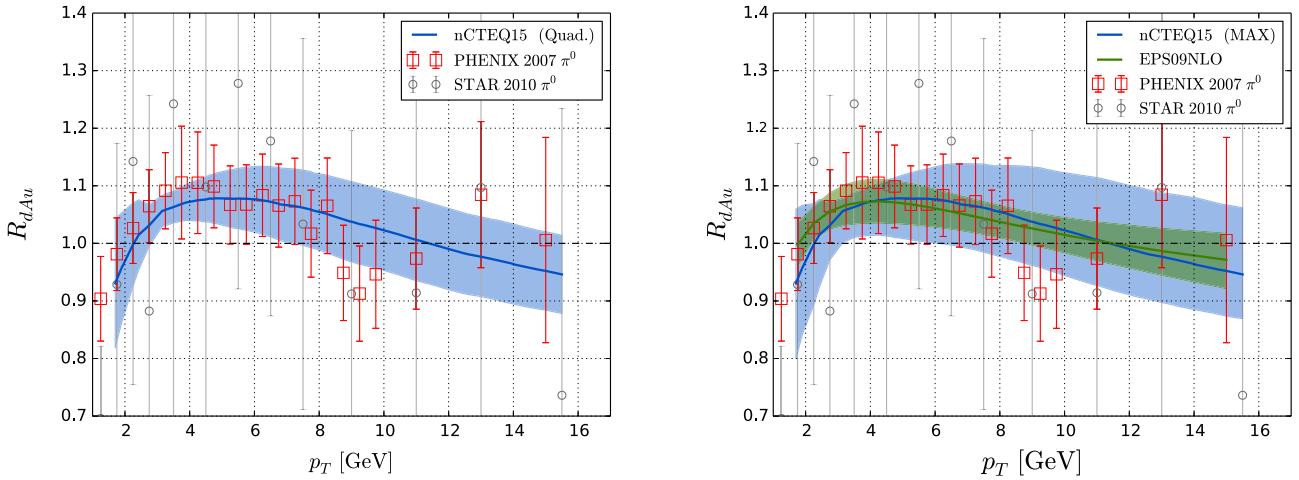


FIG. 14. Comparison of the nCTEQ15 NLO theory predictions for $R = \sigma^A_{\rm DY}/\sigma^{A'}_{\rm DY}$ with data for several nuclear targets from the Fermilab experiments E772 (left) and E866 (right). The error bands show the uncertainty from the nuclear PDFs.



(a) Comparison of the `nCTEQ15` fit with the data. The error bands are computed by adding the uncertainties in quadrature.

(b) Comparison of the `nCTEQ15` and EPS09 fits with the data. The `nCTEQ15` error bands are computed using asymmetric uncertainties (MAX) to match EPS09.

FIG. 15. We display the comparison of the nCTEQ15 and EPS09 fits with the PHENIX [67] and STAR [68] data for the ratio $R^{\pi}_{\rm dAu}$. The plotted PHENIX and STAR data are shifted by our fitted normalization.

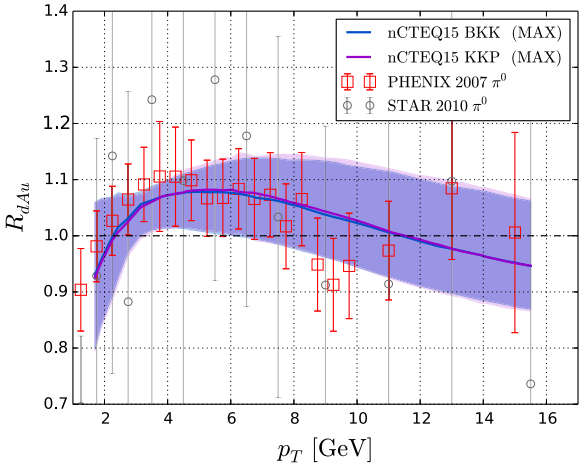
investigate the effect of different fragmentation functions, we have calculated $R^{\pi}_{\rm dAu}$ using the KKP FFs *but* still using the nCTEQ15 nPDFs obtained employing the BKK FFs [see Fig. 16(a)]. As can be seen, the choice of different fragmentation functions yields only minor differences.

In a second step, we have also performed a complete reanalysis of the nuclear PDFs using the KKP fragmentation functions in both the fit and also for the calculation of $R^{\pi}_{\rm dAu}$, and this is shown in Fig. 16(b). The use of the KKP FFs does not change the central prediction for $R^{\pi}_{\rm dAu}$ but slightly changes the nPDF uncertainties in the high-p_T region.

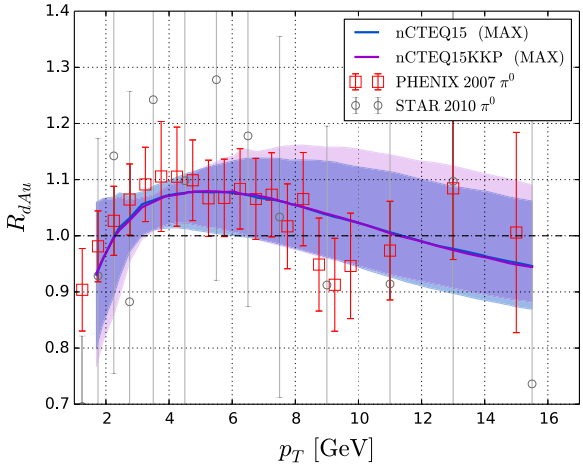
In summary, the use of two different sets of fragmentation functions, BKK and KKP, has only a minor effect on the resulting nPDFs. This does not exclude a possibility that a larger effect on nPDFs is possible if other fragmentation functions are used [86].

C. Fit without inclusive pion data (nCTEQ-np)

To further analyze the impact of the newly added inclusive pion data and because the pion data introduce an unwanted dependence on fragmentation functions, we performed an alternative analysis which *does not* include the RHIC inclusive pion data (nCTEQ15-np).



(a) Comparison of the `nCTEQ15` fit using the default BKK (blue) and the KKP fragmentation (violet) functions for the calculation of R^{π}_{dAu}.



(b) Same as previous figure, but with a full re-analysis using the BKK (blue) and the KKP fragmentation (violet) functions throughout the fitting procedure.

FIG. 16. We compare the impact of different fragmentation functions on the observable R^{π}_{dAu}. The nCTEQ15 error bands are computed using asymmetric uncertainties to match EPS09.

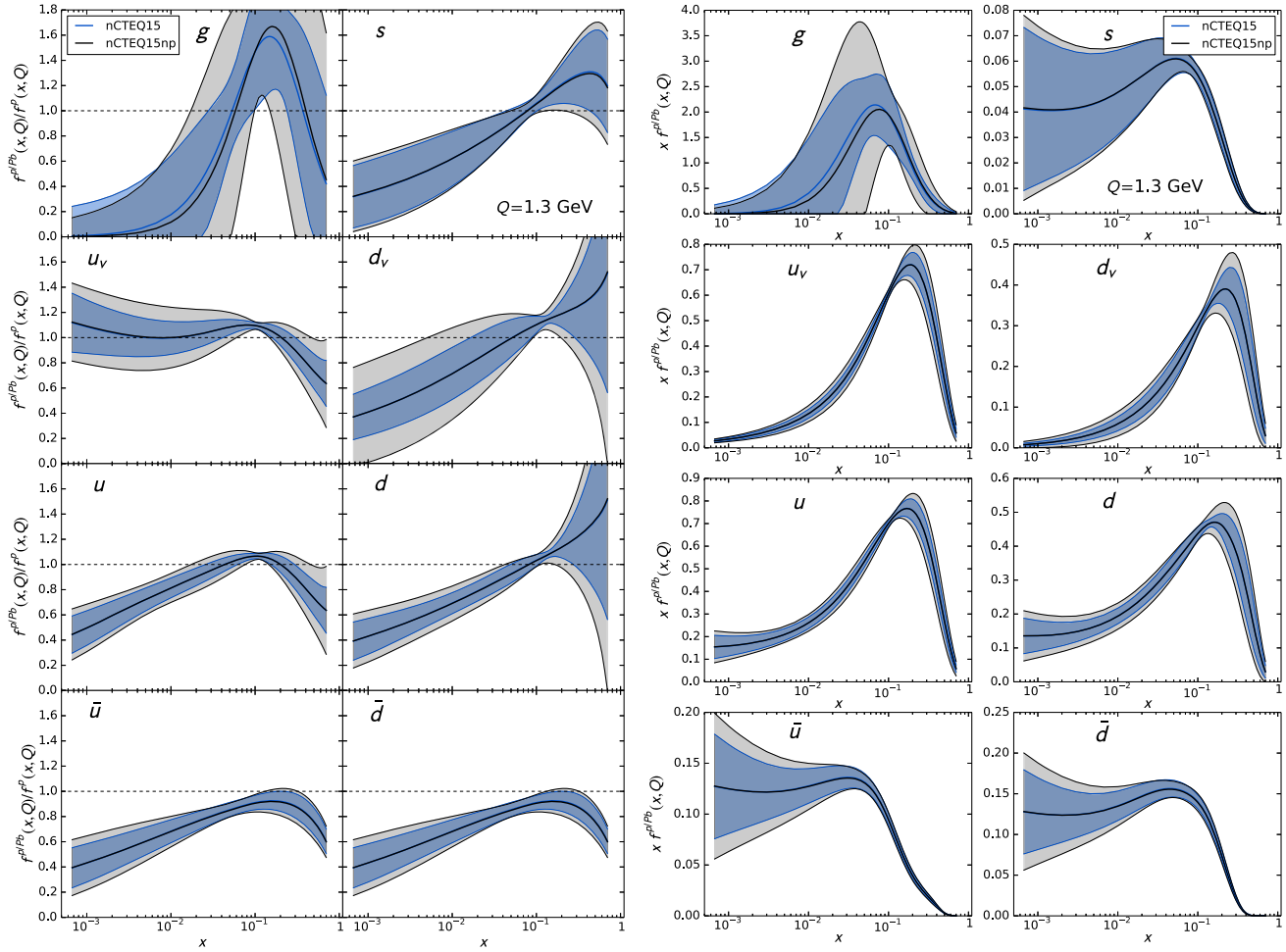


FIG. 17. Comparison of the nCTEQ15 fit (blue) with the nCTEQ15-np fit without pion data (gray). On the left, we show nuclear modification factors defined as ratios of proton PDFs bound in lead to the corresponding free-proton PDFs, and on the right we show the actual bound proton PDFs for lead. In both cases, the scale is equal to $Q = 1.3$ GeV.

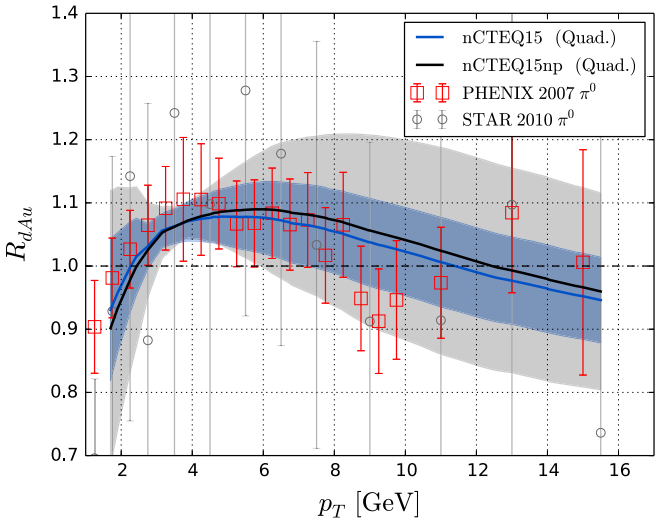


FIG. 18. Comparison of the predictions of the nCTEQ15 (solid blue) and nCTEQ15-np (dashed gray) fits to inclusive pion production data from PHENIX and STAR demonstrating the effect of including these data sets. Note that the dark blue area is the overlap between the blue and gray bands.

In Fig. 17, we compare the results of the nCTEQ15 fit with the ones of the alternative analysis nCTEQ15-np. When examining the nuclear correction factors (left panels), we see the pion data have an impact on the gluon PDF and to a lesser extent on the valence and sea quark distributions. For the central prediction, the inclusion of the pion data decreases the lead gluon PDF at large x and increases it for smaller x; the two gluon distributions cross each other at $x \sim 0.08$. Throughout most of the x range, the error bands are reduced with the exception of $x \sim 0.1$ (and very small x values) where they stay more or less unchanged. This is precisely the range that is sensitive to the DIS Sn/C (and DY) data. For most of the other PDF flavors, the change in the central value is minimal (except for a few cases at high x where the magnitudes of the PDFs are small). For these other PDFs, the inclusion of the pion data generally decreases the size of the error band.

In Fig. 18, the predictions of the nCTEQ15 and nCTEQ15-np fits are compared to the RHIC pion production data. The effect of the pion data is to increase R^{π}_{dAu} for small p_T and decrease it at larger p_T by up to 5%. The two central predictions cross each other at $p_T \sim 4$ GeV. This can be connected to the crossing of the gluon distributions in Fig. 17 (at $x \sim 0.08$) which is in line with the kinematic mapping in Fig. 2.

D. Constraining the PDF flavors with data

Global analyses of PDFs necessarily include data from a wide variety of experiments which are differently sensitive to various PDF flavors. Examining the leading order expressions for DIS, DY, or π production provides a simple estimate of which observable can constrain which PDF flavor combination. Additionally, we have to take into account the number of data points and their statistical and systematic uncertainties. All of these factors contribute to the χ^2 function; hence, we start with this measure to evaluate the impact of different experiments upon the PDF flavors.

1. χ^2 vs the gluon parameters

In Fig. 19, we compare the change of the global χ^2 and the contributions from individual experiments to this change as a function of the shift of selected gluon parameters $\{c^g_{1,1}, c^g_{4,1}, c^g_{0,2}\}$ from the respective best fit values. Recall that the parameters $\{c^g_{1,1}, c^g_{4,1}\}$ control the shape of the gluon PDF whereas $\{c^g_{0,2}\}$ controls the A dependence of the normalization. The remaining gluon parameters behave in a similar manner as $c^g_{1,1}$ and $c^g_{4,1}$.

One feature that is immediately apparent is that the nCTEQ15 minimum is not necessarily a minimum for all the experiments individually. For example, we see that the PHENIX experiment would prefer to shift $c^g_{1,1}$ to larger values ($\sim$0.002), while some of the DIS experiments (e.g., ID = 5116, NMC-96 Pb/C) prefer a lower value for $c^g_{1,1}$ (~ -0.002). Therefore, the obtained fit is a compromise that depends on the relative weight of the various data sets. This observation is part of the reason we consider a $\Delta\chi^2 = 1$ tolerance criterion impractical and choose $\Delta\chi^2 = 35$ (see Appendix A 1). Moreover, for some experiments, there may not even be a local minimum in the vicinity of the nCTEQ15 solution. Thus, these figures highlight some of the tensions between the individual data sets that the global fit must accommodate.

On top of that, Fig. 19 shows which experiments are most sensitive to the change of the underlying gluon parameters. In turn, the same experiments are the ones which have the largest impact when constraining the gluon PDF. Perhaps in contrast with expectations, the parameters analyzed in Fig. 19 are mostly constrained by the NMC Sn/C data and data from several other DIS experiments. We also see that the inclusive pion production from PHENIX is sensitive to the gluon shape parameters ($c^g_{1,1}$, $c^g_{4,1}$) but not to its normalization ($c^g_{0,2}$).

2. Correlations between data sets and PDFs

Looking at the dependence of the χ^2 function on only three gluon PDF parameters cannot give a complete picture, and neither would inspecting the behavior for all gluon parameters because the momentum sum rule connects in fact all PDF flavors together. Therefore, in the following, we use different methods to study the impact of individual experiments on different PDF flavors.

We introduce two quantities which will help us analyze the impact individual experiments have on constraining

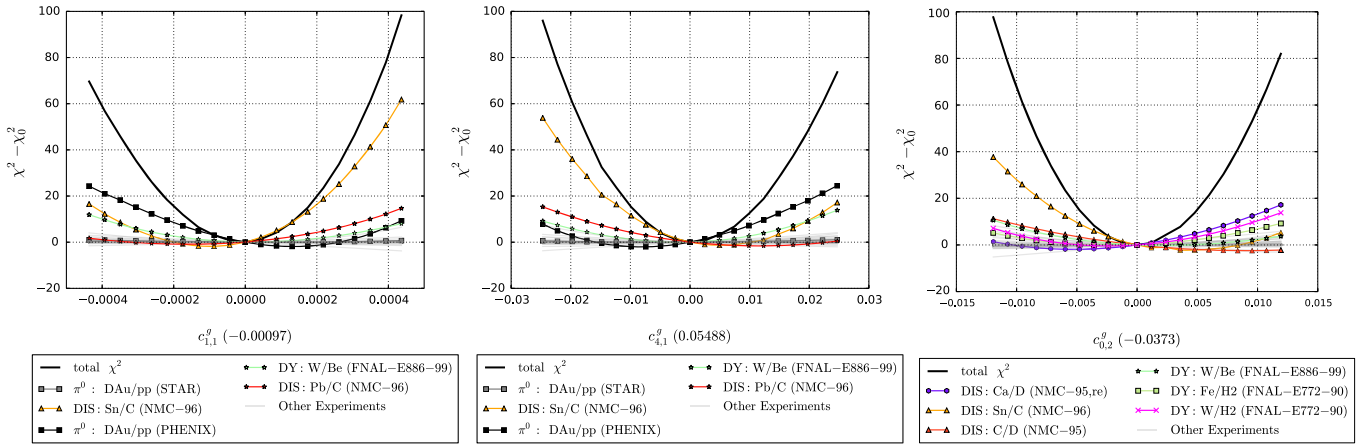


FIG. 19. Contribution of different experiments to the total $\Delta\chi^2 = \chi^2 - \chi^2_0$ function (solid black line) for a selection of gluon parameters (a) $c^g_{1,1}$, (b) $c^g_{4,1}$, (c) $c^g_{0,2}$. On the x axis, we show the shift from the best fit value (indicated in the parenthesis, cf. Table V).

given PDF flavors. The first quantity is the cosine of the correlation angle between two observables X and Y which was used in Refs. [44,87] and can be defined as

$$\cos\phi[X,Y] = \frac{\sum_{i_{\rm pdf}} (X^{(+)}_{i_{\rm pdf}} - X^{(-)}_{i_{\rm pdf}})(Y^{(+)}_{i_{\rm pdf}} - Y^{(-)}_{i_{\rm pdf}})}{\sqrt{\sum_{i'_{\rm pdf}} (X^{(+)}_{i'_{\rm pdf}} - X^{(-)}_{i'_{\rm pdf}})^2}\sqrt{\sum_{i''_{\rm pdf}} (Y^{(+)}_{i''_{\rm pdf}} - Y^{(-)}_{i''_{\rm pdf}})^2}}, \tag{4.2}$$

where the indices $i_{\rm pdf}$ run over the 16 $z_{i_{\rm pdf}}$ eigenvector directions.

In the following, we will use the cosine of the correlation angle to investigate the correlations between the χ^2 functions of the individual experiments and a single PDF. For example, in the case of the gluon PDF, the cosine of the correlation angle has the form $\cos\phi[g(x,Q),\chi^2(j_{\rm exp})]$. This correlation cosine depends on x and Q through the gluon PDF, $g(x,Q)$, and on the particular experiment through $\chi^2(j_{\rm exp})$.

Even though the cosine of the correlation angle is a useful quantity, it does not highlight the experiments with more data or smaller errors. It turns out that the normalization factors in Eq. (4.2) strongly reduce any sensitivity to the number of data points or to the size of the errors of an experimental data set. Therefore, we introduce an alternate measure, the effective χ^2 for an experiment $j_{\rm exp}$, defined as

$$\Delta\chi^2_{\rm eff}(j_{\rm exp},X) = \sum_{i_{\rm pdf}} \frac{1}{2}\left(|\chi^{2(+)}_{i_{\rm pdf}}(j_{\rm exp}) - \chi^{2(0)}_{i_{\rm pdf}}(j_{\rm exp})| + |\chi^{2(-)}_{i_{\rm pdf}}(j_{\rm exp}) - \chi^{2(0)}_{i_{\rm pdf}}(j_{\rm exp})|\right)\left(\frac{X^{(+)}_{i_{\rm pdf}} - X^{(-)}_{i_{\rm pdf}}}{\sqrt{\sum_{i'_{\rm pdf}} (X^{(+)}_{i'_{\rm pdf}} - X^{(-)}_{i'_{\rm pdf}})^2}}\right)^2. \tag{4.3}$$

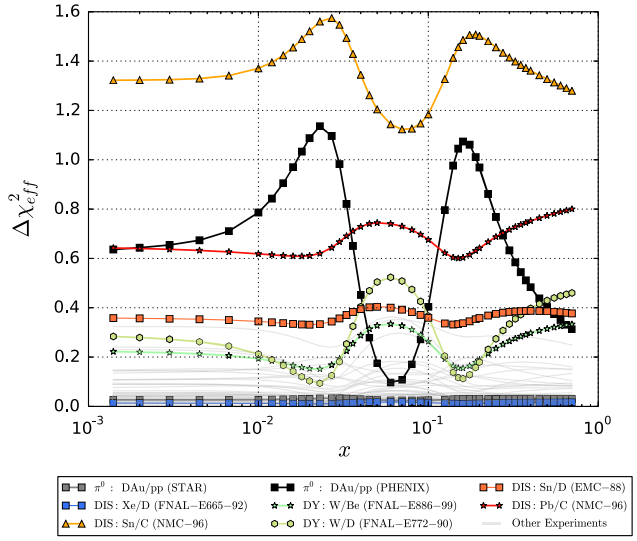
As before, the index $i_{\rm pdf}$ runs over the 16 z_i eigenvector directions.

$\Delta\chi^2_{\rm eff}$ is positive definite, and comparing the definitions (4.2) and (4.3), it is missing the normalization factor for the χ^2 function which allows it to be more sensitive to experiments with more data or smaller errors, i.e., experiments which have a larger impact in constraining single PDF flavors.

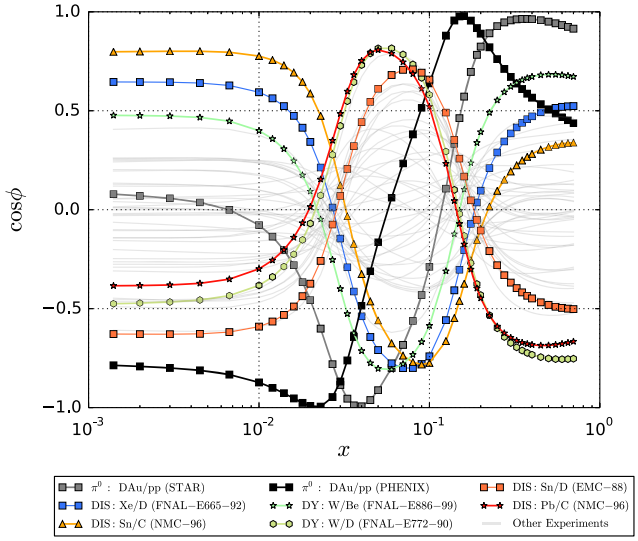
In Figs. 20 and 21, we display both the $\Delta\chi^2_{\rm eff}$ and correlation cosine as a function of x. These plots do not exhibit a strong Q dependence, so we only display them for one value of $Q = 10$ GeV.

We now examine the $\Delta\chi^2_{\rm eff}$ results for the gluon PDF in lead ($A = 207$) as shown in Fig. 20(a). For readability, we primarily show the data sets which have the largest impact on $\Delta\chi^2_{\rm eff}$; these are generally the data sets which involve the heaviest targets. The strong influence of the DIS Sn/C set reflects a combination of the large Q^2 coverage of the data and the small errors. The DIS Pb/C data, and to a lesser extent the DIS Sn/D data, also provide constraints for the gluon PDF in lead. The PHENIX pion production data contributes strongly in the central x region; conversely, the effect of the STAR data is negligible due to the larger uncertainties. Additionally, the DY data on heavy targets (W tungsten with Be and D) also play a role in determining the gluon lead PDF; this is due to the fact that the DY data cover a range $\sim(20, 170)$ GeV^2 in the invariant mass of the muon pair, which creates some sensitivity to the gluon PDF via scale evolution.

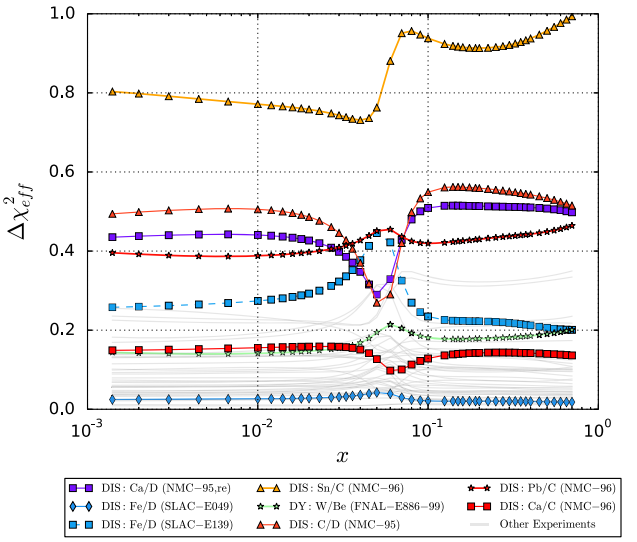
In Fig. 20(b), we show the correlation cosine for the gluon PDF in lead. The DIS Sn/C and DY W/Be data sets



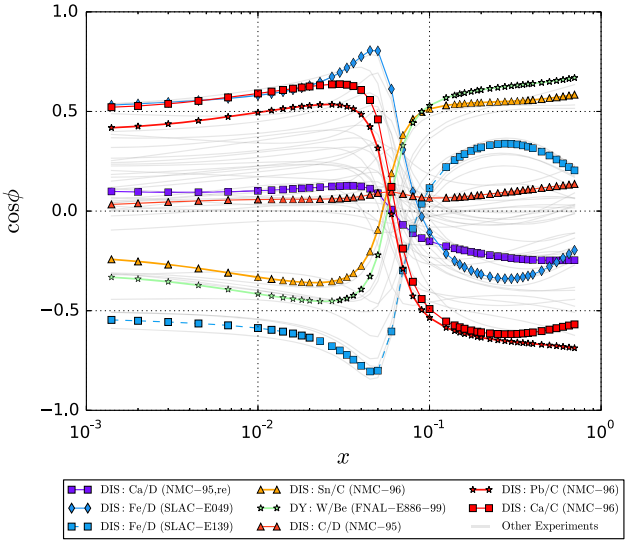
(a) **gluon**: $\Delta\chi^2_{\text{eff}}(g)$ at $Q = 10$ GeV for lead



(b) **gluon**: $\cos\phi(g, \chi^2)$ at $Q = 10$ GeV for lead



(c) **gluon**: $\Delta\chi^2_{\text{eff}}(g)$ at $Q = 10$ GeV for carbon



(d) **gluon**: $\cos\phi(g, \chi^2)$ at $Q = 10$ GeV for carbon

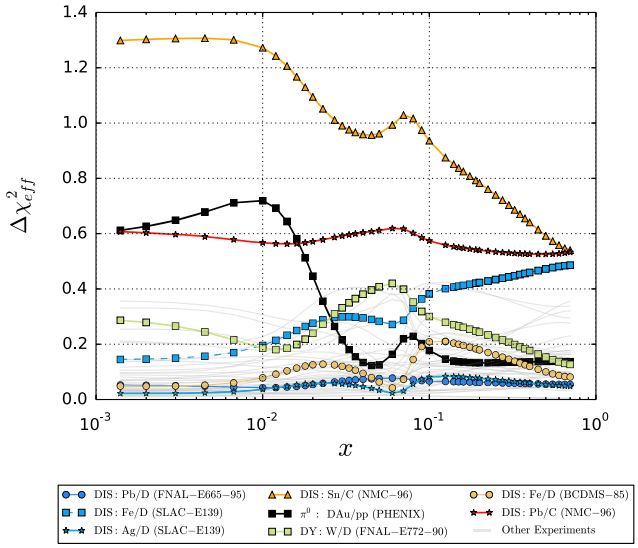
FIG. 20. Correlation measures for lead and carbon at $Q = 10$ GeV for the gluon of the nCTEQ15 fit. The left panels display the effective χ^2, and the right panels display the correlation cosine as a function of x.

have positive correlations at large and small x and a negative dip in the middle. Contrary to this, the DIS Pb/C, Sn/D, and DY W/D data sets have the opposite behavior. Hence, these data sets are anticorrelated which indicates that they pull against each other in the fit. This is precisely what we have observed in Fig. 19 for the gluon parameters. Also, the PHENIX data have a separate x dependence (arising from a separate production mechanism), and this will further help us separate the PDF flavor components.

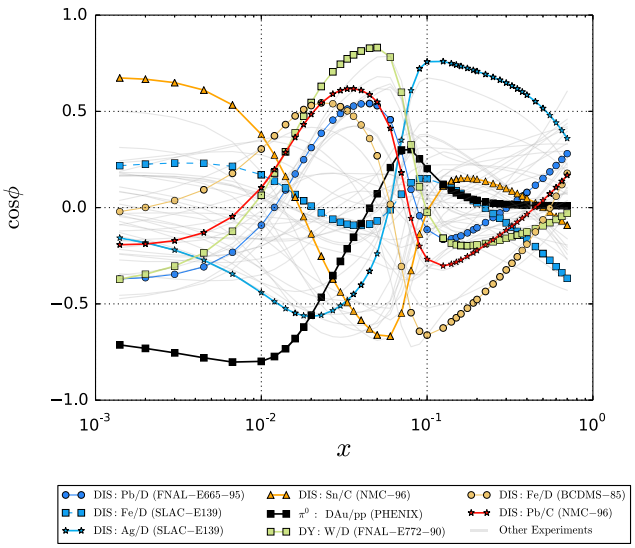
Finally, there are two data sets (STAR and DIS Xe/D) that have relatively large correlation cosines but do not have a large influence on the $\Delta\chi^2_{\text{eff}}$ of Fig. 20(a); thus, we need to take care when interpreting the results of the correlation cosine plots and use this in combination with $\Delta\chi^2_{\text{eff}}$.

We now consider the gluon PDF in carbon ($A = 12$) to see if the general observations above apply in the case of a lighter nuclei. In Fig. 20(c), we see the primary data sets constraining $\Delta\chi^2_{\text{eff}}$ are the DIS sets involving ratios of carbon (Sn/C, C/D, Pb/C) or other comparable nuclei (Ca/D). Note the DY data on heavy tungsten (W) and the pion production data on gold (Au) are not shown as they do not contribute significantly to $\Delta\chi^2_{\text{eff}}$ for carbon.

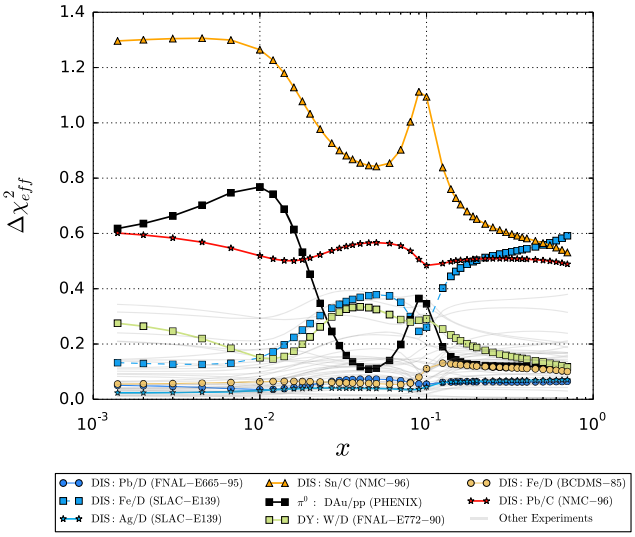
The correlation cosines for the gluon PDF in carbon are shown in Figs. 20(d). We see the DIS Pb/C data have a positive correlation cosine at small x and a negative correlation cosine at large x. The DIS Sn/C data show the opposite behavior; hence, these data sets will pull against each other in the fit. The DIS C/D and Ca/D data generally have a small correlation cosine throughout the x



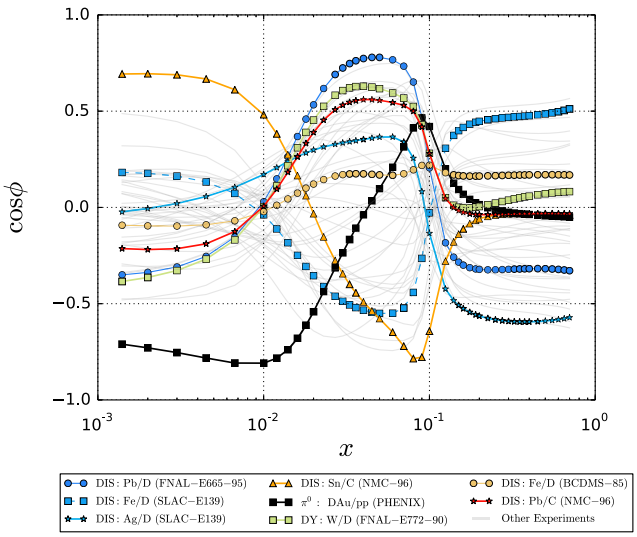
(a) ***u*-quark**: $\Delta\chi^2_{\text{eff}}(u)$ at $Q = 10$ GeV for lead



(b) ***u*-quark**: $\cos\phi(u, \chi^2)$ at $Q = 10$ GeV for lead



(c) ***d*-quark**: $\Delta\chi^2_{\text{eff}}(d)$ at $Q = 10$ GeV for lead



(d) ***d*-quark**: $\cos\phi(d, \chi^2)$ at $Q = 10$ GeV for lead

FIG. 21. Correlation measures for lead at $Q = 10$ GeV for the u-quark and d-quark distributions of the nCTEQ15 fit. The left panels display the effective χ^2, and the right panels display the correlation cosine as a function of x.

range. As in the case of the gluon in lead, we see there are a number of data sets (such as DIS Fe/D) that have a large correlation cosine but yield a small contribution to the $\Delta\chi^2_{\text{eff}}$; thus, we need to use both the $\Delta\chi^2_{\text{eff}}$ and $\cos\phi$ information together when drawing our conclusions.

We now turn our attention to $u^{p/Pb}$ and $d^{p/Pb}$ distributions for lead at $Q = 10$ GeV as shown in Fig. 21. For these PDFs, not only is the Q dependence rather mild, but the differences between heavy and light nuclei are also not as pronounced as in the gluon case. The $\Delta\chi^2_{\text{eff}}$ for the u and d PDFs depends almost exclusively on the DIS data from heavy targets (Sn/C, Pb/C, Fe/D), with some contributions from PHENIX pion production data at small x and a minimal contribution from the DY W/D data.

Turning to the $\cos\phi$ plots, we see the DIS Sn/C and the Fe/D data start with a positive correlation cosine at small x and move negative for increasing x, while the DIS Pb/C and the DY W/D data do the opposite; hence, these sets are anticorrelated in this region. For small to medium x values, the general pattern is similar between the u and d correlation plots, but they differ some at large x where we see, for example, the DIS Fe/D data has a positive correlation cosine for u but a negative one for d; this will be useful in differentiating u and d PDFs at large x. As with the gluon correlation plots, there are a number of data sets

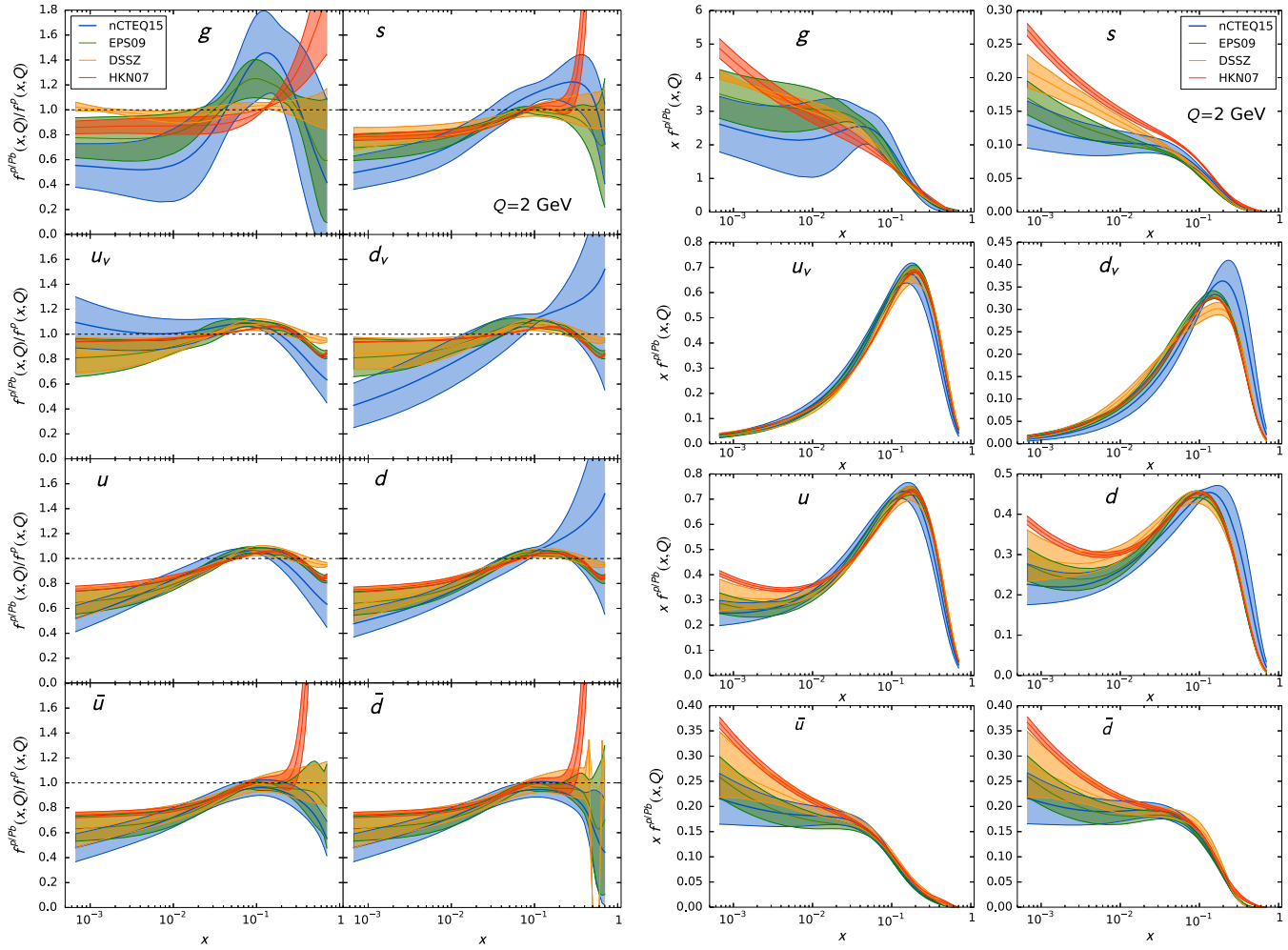


FIG. 22. Comparison of the nCTEQ15 fit (blue) with results from other groups: EPS09 [12] (green), DSSZ [11] (orange), and HKN07 [14] (red). The left panel shows nuclear modification factors for lead, and the right panel shows the actual PDFs of a proton bound in *lead*. The scale is $Q = 2$ GeV. The wide spread of the ratios at large x is an unphysical artifact due to the vanishing of the PDFs in this region.

(such as the DIS Ag/D) which have large correlation cosines but small contributions to $\Delta\chi^2_{\text{eff}}$; thus, they have minimal effect constraining the PDFs.

E. Comparison with different global analyses

We now compare our nCTEQ15 PDFs with other recent nuclear parton distributions in the literature. Specifically, we will consider DSSZ [11], EPS09 [12], and HKN07 [14].[18] Our data set selection and technical aspects of our analysis are closest to that of EPS09. In Figs. 22 and 23, we plot nuclear modifications for the PDFs of a proton bound in lead, $f^{p/Pb}/f^p$ (left), as well as the bound proton PDFs themselves, $f^{p/Pb}$ (right), for different flavors for a selection of Q scales.

For the $\bar{u}$ and $\bar{d}$ PDFs at $Q = 2$ GeV, nCTEQ15 has significant overlap with the other sets through much of the x range with a stronger shadowing at small x. Our results at $x < 10^{-2}$ are extrapolated since they are not constrained by data due to the cut $Q > 2$ GeV which was imposed in order to reduce higher twist contributions. Therefore, it is likely that the uncertainty band at $x < 10^{-2}$ underestimates the true PDF uncertainties. While this trend repeats itself for the strange quark PDF, the spread at small x is slightly increased.[19] In fact, at $Q = 2$ GeV, the small x behavior of

[18]Note that there is also a very recent global nPDF analysis performed at NNLO level [88].

[19]In this analysis, the s-quark nuclear effects are completely determined by the $\bar{u}$ and $\bar{d}$ nuclear PDFs and by the gluon nuclear PDF through evolution. Due to these constraints, the error of the s-quark nuclear PDF is underestimated. A comprehensive analysis would require including the charged-current ν-DIS data as in Ref. [21] along with using a proton PDF baseline where the strange distribution was determined from different data such as the $W + c$ production at the LHC.

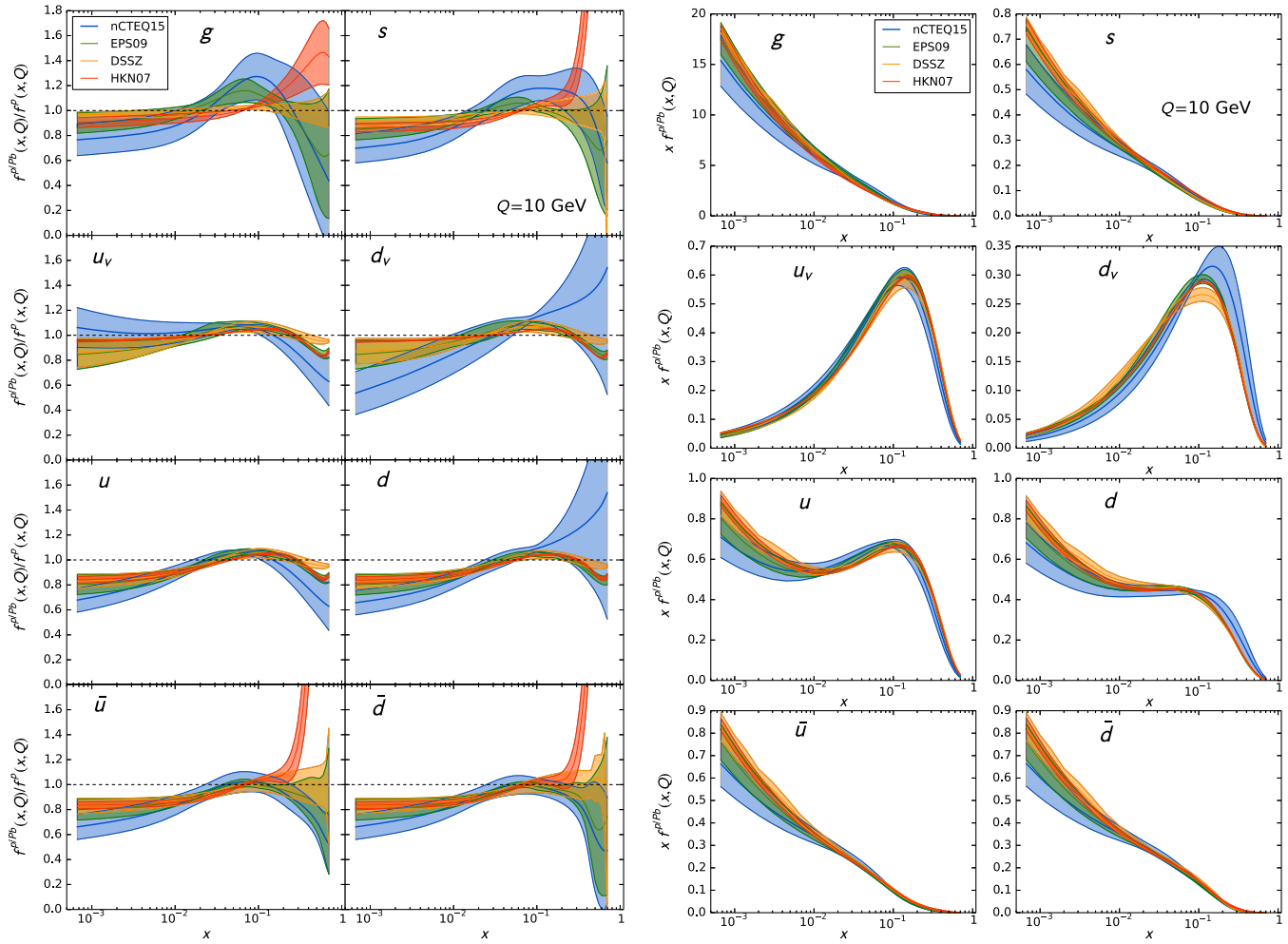


FIG. 23. Same as Fig. 22, with $Q = 10$ GeV.

the strange PDF of all four fits is quite distinct with little overlap between the uncertainty bands [see Fig. 22(b)]. As we move to higher Q values, the DGLAP evolution tends to bring the various PDF sets into closer agreement, particularly at small x values. For example, already at $Q = 10$ GeV, the nCTEQ15 bands overlap the other PDFs across a much broader x range than at low-Q values.

In the case of the gluon, there is considerable variation among the different PDF sets at Q close to the initial scale. Again, the nCTEQ15 exhibits a stronger shadowing suppression along with a larger enhancement in the antishadowing region ($x \sim 0.1$). In addition, the uncertainty band for $x \gtrsim 0.02$ is considerably larger than the uncertainty bands of the other groups. The nCTEQ15 result is largely compatible with the result of EPS09 even though the shape of the central prediction is more suppressed in the shadowing region and enhanced in the antishadowing region. We have less overlap with the HKN07 and DSSZ bands, in part due to their smaller uncertainty bands. Moving to larger Q values, the DGLAP evolution again causes the different PDFs to converge.

Note that the ratio plots of Figs. 22 and 23 have quite a wide spread at large x values. This unphysical behavior is an artifact due to the vanishing of the PDFs in this region. The spread is largest for those PDFs with minimal support at large x—specifically g, s, $\bar{u}$, $\bar{d}$. Also, these effects are reduced when we construct the full nuclear lead distribution as shown in Fig. 24.

Examining the u- and d-valence distributions, one can see that PDF sets $\{\mathrm{HKN07}, \mathrm{EPS09}, \mathrm{DSSZ}\}$ agree quite closely with each other throughout the x range. While the nCTEQ15 fit uncertainty bands generally overlap the other sets, we see on average the u_v distribution is softer and the d_v distribution is harder. These differences reflect the fact that the HKN07, EPS09, and DSSZ fits assume that the nuclear corrections R_{u_v} and R_{d_v} are the same, while the nCTEQ15 fit allows them to vary independently. Clearly, there is no physical reason to assume that u_v and d_v must have a universal nuclear correction factor, and there exist models in the literature [70,80,81] which indeed predict nonuniversal modifications.

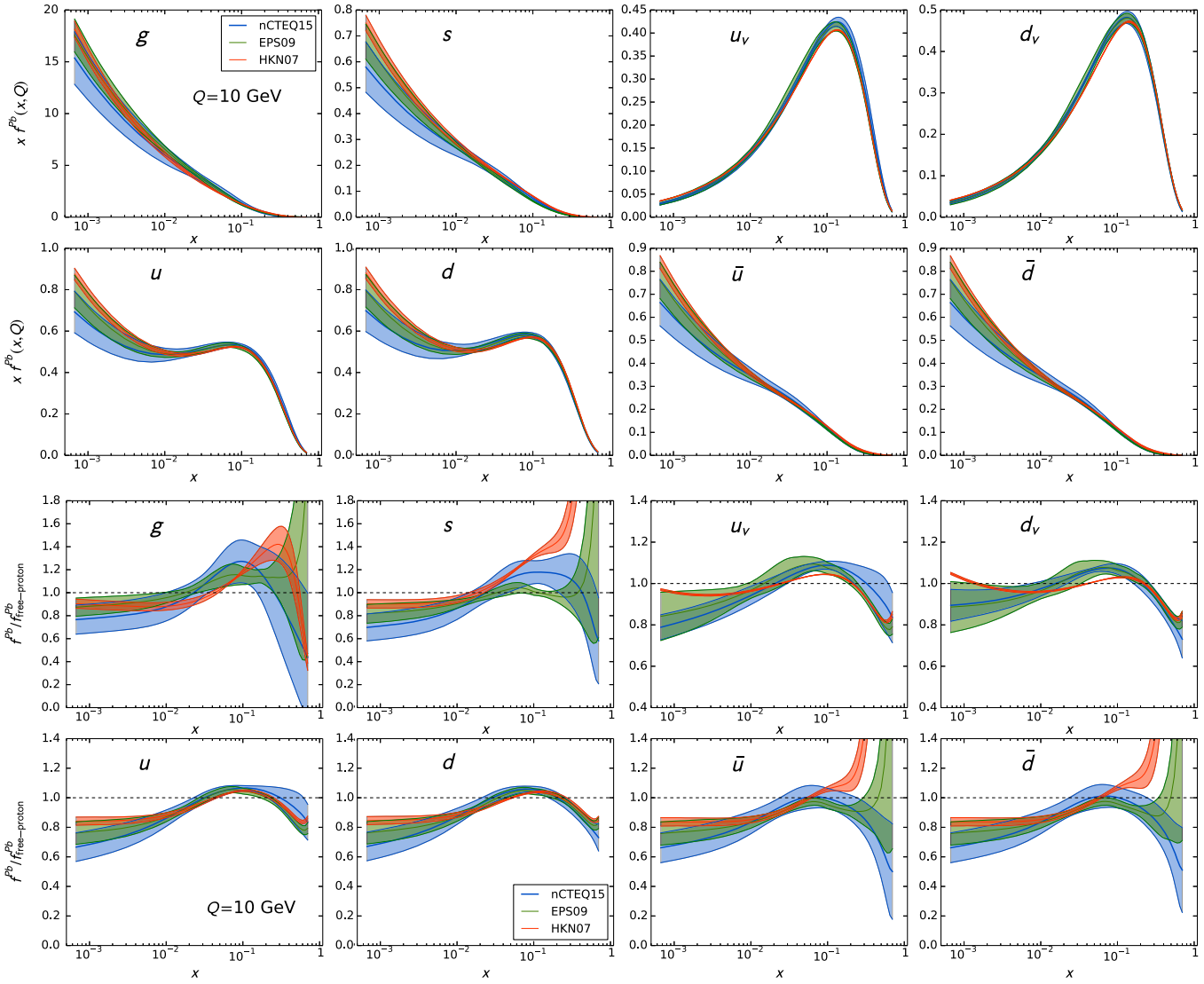


FIG. 24. (upper panel) Comparison of the full nuclear lead distributions, $f^{Pb} = \frac{82}{207} f^{p/Pb} + \frac{207-82}{207} f^{n/Pb}$, for nCTEQ15 (blue), EPS09 (green), and HKN07 (red) at $Q = 10$ GeV. The lower panel shows the same distributions compared to the lead PDF, f^{Pb}, constructed of free-proton distributions. The wide spread of the ratios at large x is an unphysical artifact due to the vanishing of the PDFs in this region.

The obvious question is whether the additional freedom to decouple the R_{u_v} and R_{d_v} nuclear corrections yields a substantial improvement in the fit. To shed more light on this issue, we have generated a modified fit where we have forced the u_v and d_v nuclear corrections to be similar to the EPS09 PDF set.[20] We find that the χ^2/dof for this modified fit is comparable ($\Delta\chi^2 \lesssim 5$) to our original nCTEQ15 at a level well below our tolerance criteria of $\Delta\chi^2 = 35$. Therefore, we conclude that the current data sets are not sufficiently sensitive to distinguish the u_v and d_v nuclear corrections to a good degree. Hence, the advantage of independent R_{u_v} and R_{d_v} correction factors is currently limited, which, however, will change with more data (e.g., from the LHC).[21]

To better understand this result, we observe in Figs. 22 and 23 that the u_v and d_v ratios exhibit opposite x dependence as compared with the {HKN07,EPS09,DSSZ} sets.

[20]As we are fitting directly the nuclear PDFs $f^{p/A}(x,Q)$ and not the ratios $f^{p/A}(x,Q)/f^{p}(x,Q)$, it is nontrivial to force the nuclear corrections to be exactly the same if the underlying proton PDFs differ. We are able to find an approximate solution by equating the u_v and d_v coefficients $c_{i,j}$ for $\{ij\} = \{11,12,21,22,31,32,51,52\}$ and refitting the PDFs.

[21]In an earlier study, we did find an *apparent* difference due to independent R_{u_v} and R_{d_v} nuclear corrections. The present updated analysis additionally includes (i) an improved treatment of the $\{A,Z\}$ isoscalar corrections and (ii) QED radiative corrections for DIS data sets, (iii) use of full theory (instead of K-factors) to obtain the final minimum, and (iv) improved numerical precision for the DY process. With these improvements, the χ^2 of the modified fit is now comparable to nCTEQ15.

That is, the u_v ratio is below the other sets at large x and above at small x; the d_v ratio does the opposite. The nuclear data sets probe a linear combination of u_v and d_v, which raises the question as to whether the above differences might cancel when combined.

In Fig. 24, we now compare the full nuclear lead PDFs from the different groups. The upper panel shows the PDFs themselves, and the lower one shows their ratio compared to the nuclear combination constructed out of the free proton—the full nuclear correction. From this comparison, we can clearly see that the large differences in the effective bound proton distributions of valence quarks (Figs. 22 and 23) translate into much smaller differences in the full nuclear PDFs that actually enter the calculation of observables.[22] In particular, we see that u_v and d_v distributions of the nCTEQ15 fit are in very good agreement with the EPS09 results and have substantial (but not complete) overlap with HKN07.[23]

Of course, as the data can only constrain the full nuclear PDF in the combination $f^A = \frac{Z}{A} f^{p/A} + \frac{A-Z}{A} f^{n/A}$, we conclude that better separation of u_v and d_v distributions requires more data on nonisoscalar targets. We also note that the currently available DIS data use a number of nonisoscalar targets and would have the potential to partially distinguish u_v and d_v distributions; unfortunately, many of these data sets have been corrected for the neutron excess and in turn lost this ability.

V. SUMMARY AND CONCLUSIONS

In this paper, we have presented the first complete analysis of nuclear PDFs with errors in the CTEQ framework. The resulting fit, nCTEQ15, uses the available charged lepton DIS, DY, and inclusive pion data taken on a variety of nuclear targets. The uncertainty of this analysis is presented in the form of error PDFs which are constructed using an adapted Hessian method.

Within our framework, we are able to obtain a good fit to all data. The output of the nCTEQ15 analysis is a complete set of nuclear PDFs with uncertainties for any $A = \{1, \ldots, 208\}$. A selection of nuclear PDFs for the most common nuclei are made publicly available,[24] but custom nPDFs can be generated for any $\{A, Z\}$ combination.

In comparison to our previous analysis [13], we have included the data from the inclusive pion production at the RHIC. The new data provide additional constraints mostly for the nuclear gluon PDF, but the description of the data relies on the fragmentation functions. Therefore, we also provide an alternative conservative result nCTEQ15-np which does not include the inclusive pion data and is hence fragmentation function independent.

Compared to other global analyses (HKN07, EPS09, and DSSZ), there are a number of important differences:

(i) In contrast to the other analyses, we parametrize the nuclear PDFs directly instead of the nuclear corrections factors.
(ii) In addition, our u- and d-valence distributions are parametrized independently.
(iii) Other differences arise from the selection of data points used in the fit. In particular, we impose more conservative kinematic cuts in order to minimize effects from higher twists and target-mass corrections.

Overall, our results are compatible with the other nPDFs, but after a detailed look, we see distinct differences (see Fig. 24):

(i) The nCTEQ15 nuclear gluon PDF has a larger shadowing at small x than the other global analyses. Our result is compatible with the result of EPS09 as the error bands are overlapping throughout the entire x range. The overlap in the case of HKN07 and (especially) DSSZ is limited especially in the small-x region where no data constraints are present (and uncertainties of HKN07 and DSSZ are very small). This highlights the fact that nPDF uncertainties, in particular for gluon, are underestimated and different gluon solutions are possible [33].
(ii) Our valence distributions for a bound proton in lead differ as we allow separate nuclear corrections for u_v and d_v. Compared to the other groups, our d-valence PDF is harder, and our u-valence PDF is softer. However, when the full lead nucleus is constructed, these differences are substantially reduced, and we observe a good agreement between all groups.
(iii) The nCTEQ15 light sea quark distributions are in very good agreement with the ones from the other groups for $x \gtrsim 10^{-2}$. At smaller x where there are no data constraints, the individual error bands clearly underestimate the uncertainty.
(iv) It should be also mentioned that strange distributions are currently not fitted in any of the nPDF analyses and are fixed by imposing additional assumptions; this leads to quite significant differences between different groups.

All in all, we find relatively good agreement between different nPDFs. Most of the noticeable differences between them occur in regions without any constraints from data, and so they can be attributed to different assumptions such as parametrization of the nuclear effects.

In view of the differences, the true nPDF uncertainties should be obtained by combining the results of all analyses and their uncertainties. In particular, this is true for the gluon distribution where the small-x behavior is basically

[22]Note that only up and down distributions differ between the full nuclear PDFs and the PDFs of the bound proton; the gluon and strange distributions are the same.

[23]The DSSZ set (not show) is similar to HKN07 in that it has substantial (but not complete) overlap.

[24]The nPDF sets for the current nCTEQ15 analysis as well as for the alternative nCTEQ15-np analysis are available for download at http://ncteq.hepforge.org as well as on the LHAPDF website.

unconstrained, and every single nPDF analysis substantially underestimates it (see our earlier study [33]).

The nCTEQ framework used for the nCTEQ15 fit can combine data from both proton and nuclear targets into a single coherent analysis. Using the nCTEQ15 fit as a reference, it will be interesting to include the upcoming LHC data as we continue to investigate the relations between the proton and the nuclear PDFs.

ACKNOWLEDGMENTS

The authors would like to thank J. Gao for proposing the use of $\Delta\chi^2_{\rm eff}$ as an alternative to the correlation cosine and for other useful discussions. Furthermore, we would like to thank M. Botje, E. Godat, S. Kumano, P. Nadolsky, and V. Radescu for valuable discussions. We acknowledge the hospitality of CERN, DESY, and Fermilab where a portion of this work was performed. The work of J. F. O. was supported by the DOE under Grant No. DE-FG02-13ER41942. T. J. was partly supported by the Research Executive Agency of the European Union under Grant No. PITN-GA-2010-264564 (LHCPhenoNet). This work was also partially supported by the U.S. Department of Energy under Grant No. DE-FG02-13ER41996 and by Projet International de Cooperation Scientifique PICS05854 between France, and the U.S. Fermilab is operated by Fermi Research Alliance, LLC, under Contract No. DE-AC02-07CH11359 with the United States Department of Energy.

APPENDIX A: DETERMINATION OF $\Delta\chi^2$ AND HESSIAN RESCALING

1. Determination of $\Delta\chi^2$

In this Appendix, we discuss the details of the determination of $\Delta\chi^2$ which is motivated by the treatment presented in Refs. [12,46,47]. We investigate how the global fit describes each experiment by examining χ^2_k which is the individual χ^2 contribution of experiment k with N_k data points. We can then see how χ^2_k changes when varying PDF parameters along each eigenvector direction $\tilde{z}_i$ of Eq. (2.25).

The probability distribution for the χ^2_k given that the fit has N_k degrees of freedom is

$$P(\chi^2_k, N_k) = \frac{(\chi^2_k)^{N_k/2-1} e^{-\chi^2_k/2}}{2^{N_k/2}\Gamma(N_k/2)}. \tag{A1}$$

This allows us to define the percentiles ξ_p via

$$\int_0^{\xi_p} P(\chi^2, N) d\chi^2 = p\% \quad \text{where } p = \{50, 90, 99\}. \tag{A2}$$

Here, ξ_{50} serves as an estimate of the mean of the χ^2 distribution, and ξ_{90}, for example, gives us the value where there is only a 10% probability that a fit with $\chi^2 > \xi_{90}$ genuinely describes the given set of data.

Due to fluctuations in the data and possible incompatibilities between experiments, the global χ^2 minimum does not necessarily coincide with χ^2-minima of individual experiments. Moreover, for the same reason, the minimum χ^2 for each experiment, $\chi^2_{k,0}$, can be far away from the expected minimum given by ξ_{50}. In order to use the percentiles defined in Eq. (A2) to define the 90% C.L., we rescale the ξ_{90} percentile to take into account the position of the minimum as

$$\tilde{\xi}_{90} \to \xi_{90}\left(\frac{\chi^2_{k,0}}{\xi_{50}}\right). \tag{A3}$$

For each eigenvector direction given by a variation of the parameter $\tilde{z}_i$ and every experiment, we define an interval

$$z_i^{(k)-} \le \tilde{z}_i \le z_i^{(k)+}, \tag{A4}$$

where the χ^2_k stays within the 90% C.L. limit (i.e., $\chi^2_k < \tilde{\xi}_{90}$). For each eigenvector direction, we then construct an interval (z_i^-, z_i^+) where all experiments stay within the 90% C.L. limit as

$$(z_i^-, z_i^+) \equiv \bigcap_k (z_i^{(k)-}, z_i^{(k)+}). \tag{A5}$$

These intervals can obviously be different for each eigenvector, depending on how well the experiments constrain the variations in this eigenvector direction. For n free parameters, we obtain $2n$ parameters $\{z_1^-, z_1^+, z_2^-, z_2^+, \ldots, z_n^-, z_n^+\}$ which we can use to define the global tolerance as

$$\Delta\chi^2 \equiv \sum_i \frac{(z_i^+)^2 + (z_i^-)^2}{2n}. \tag{A6}$$

Having performed the procedure described in this section, we have arrived at $\Delta\chi^2 = 35$. One can compare how this choice of global tolerance (the same for every eigenvector direction) agrees with the rescaled 90% C.L. for each experiment in every direction. In Fig. 25, we show this comparison for only one single eigenvector direction as all the others are rather similar.

2. Hessian rescaling

Choosing a larger tolerance $\Delta\chi^2 = 35$ as argued above might pose a problem for the Hessian approach as it requires using information from a larger neighborhood of the global minimum which is not necessarily well described in the quadratic approximation. Figure 26 confirms that this is the case for the nCTEQ15 fit. Both in the original parameter space, Fig. 26(a), and in the eigenvector basis, Fig. 26(b), we can see directions where the χ^2 function deviates substantially from the quadratic approximation when $\Delta\chi^2 \sim 35$. This is a problem because in the

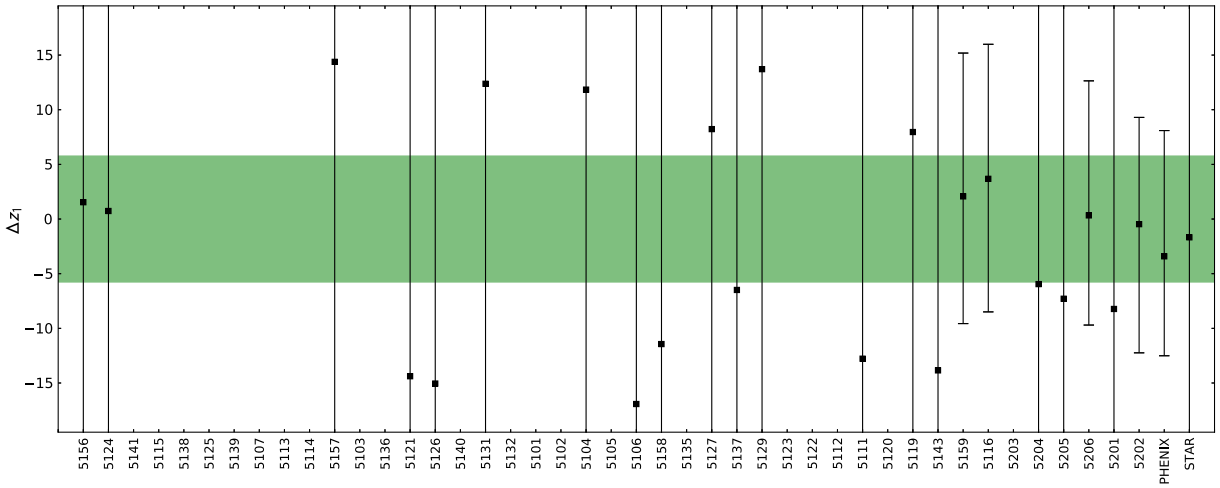
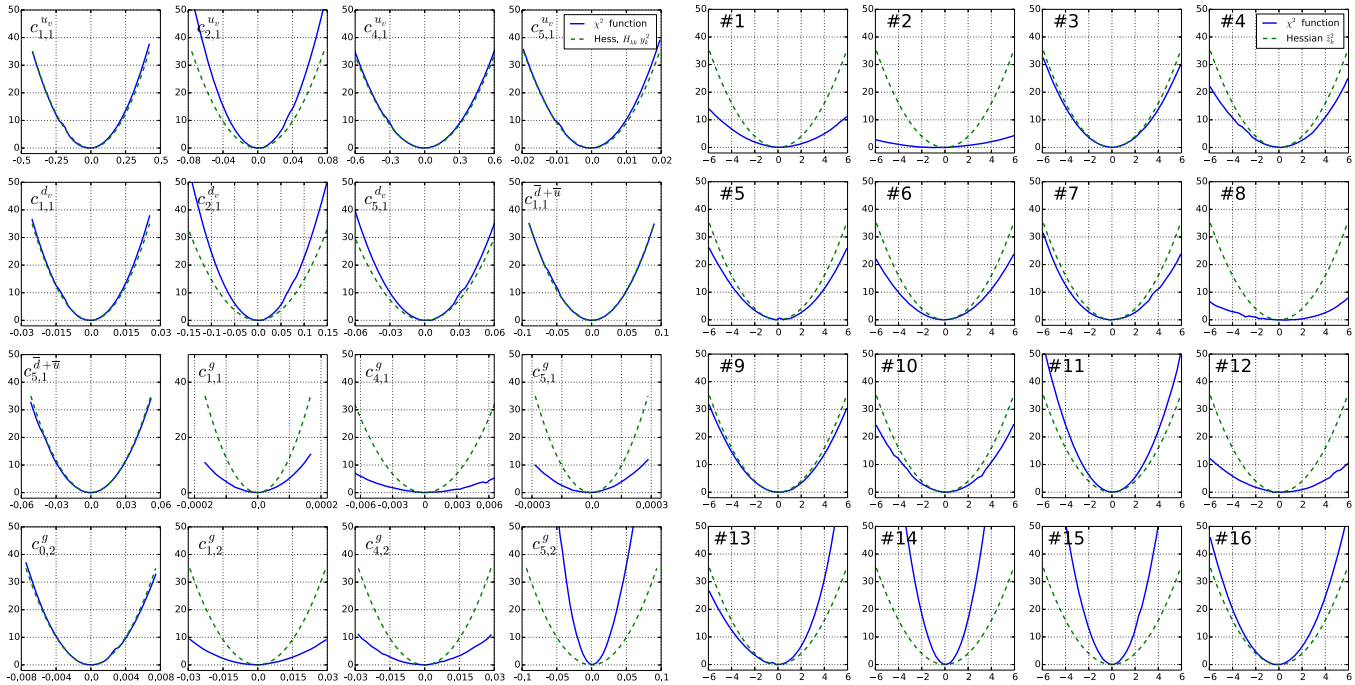


FIG. 25. The 90% confidence level limits from different data sets in the eigenvector direction $\tilde{z}_1$. The χ^2-minimum for each experiment is denoted by a black square, and the green band demonstrates the interval of the eigenvector parameter corresponding to the final $\Delta\chi^2$.

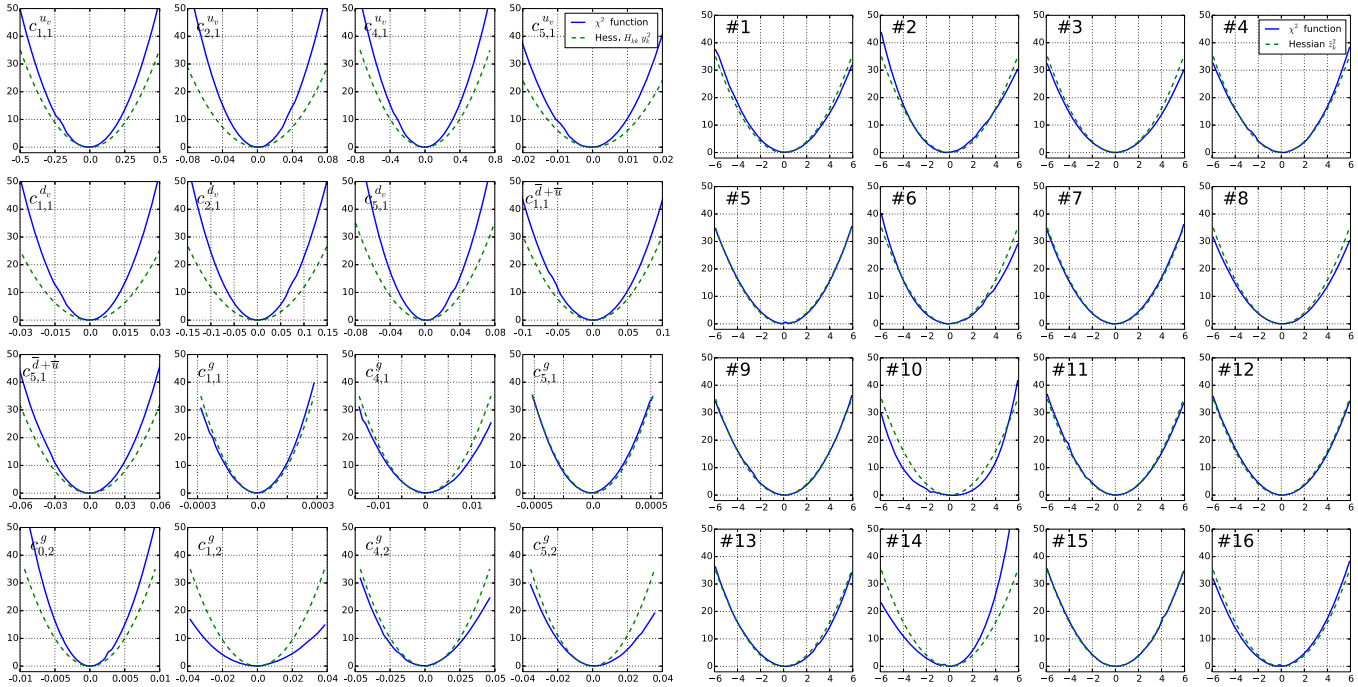
Hessian approach we use the eigenvector basis to determine the ranges of the normalized parameters $\tilde{z}_i$. Figure 26 shows that if we take $\Delta\tilde{z}_i = \sqrt{\Delta\chi^2} \approx \sqrt{35}$, then depending on the specific eigendirection, we would largely overestimate or underestimate the error on our parameters (see, e.g., plots 1, 2, and 14 in Fig. 26(b)).

To improve the constraints provided by the χ^2 function, we redefine the Hessian which we use to determine the error PDFs using the formalism described in Sec. II C. We keep the eigenvector information intact but rescale the eigenvalues of the original Hessian (which corresponds to rescaling the parameters $\tilde{z}_i$) so that the modified Hessian better describes the χ^2 function not only in the minimum ($\Delta\chi^2 = 0$) but also at $\Delta\chi^2 = 35$. For each eigenvector direction, we identify the parameter values $\tilde{z}_i^{\pm}$, where



(a) χ^2 function in parameter space. (b) χ^2 function in eigenvector space.

FIG. 26. These plots display the Hessian *before* the "rescaling" procedure. The χ^2 function relative to its value at the minimum, $\Delta\chi^2 = \chi^2 - \chi_0^2$, plotted along the 16 fitting parameters of the original space (left) and along the z_i directions in the eigenvector space (right). The actual χ^2 function is plotted with solid lines, and the Hessian approximation $\Delta\chi^2 = \tilde{z}_i^2$ is shown with dashed lines.



(a) χ^2 function in parameter space. (b) χ^2 function in eigenvector space.

FIG. 27. These plots display the Hessian *after* the rescaling procedure. The χ^2 function relative to its value at the minimum, $\Delta\chi^2 = \chi^2 - \chi_0^2$, plotted along the 16 fitting parameters of the original space (left) and along the z_i directions in the eigenvector space (right). The actual χ^2 function is plotted with solid lines, and the Hessian approximation $\Delta\chi^2 = \tilde{z}_i^2$ is shown with dashed lines.

$$\Delta\chi^2(\tilde{z}_i^{\pm}) \equiv \chi^2(\tilde{z}_i^{\pm}) - \chi_0^2 = 35, \tag{A7}$$

where χ_0^2 is the minimum of the χ^2. Using the $\tilde{z}_i^{\pm}$, we rescale the corresponding eigenvalue as

$$\lambda_i \mapsto \lambda_i' = \frac{|\tilde{z}_i^{+}|^2 + |\tilde{z}_i^{-}|^2}{2\sqrt{\Delta\chi^2}}\lambda_i. \tag{A8}$$

The impact of the rescaling of the Hessian can be seen on Fig. 27(b), where one notices that the description of the χ^2 function in the eigenvector basis is highly improved, especially in the region where $\Delta\chi^2 = 35$. The description of the χ^2 function in the original parameter space [Fig. 27(a)] is also improved but to a lesser extent. However, this is a secondary feature as we are working in the eigenvector space when defining the error PDFs.

APPENDIX B: USAGE OF nCTEQ PDFs

We provide a set of PDF tables for the nCTEQ15 and nCTEQ15-np fits at the nCTEQ Hepforge website [89]. We provide the tables in the older CTEQ PDS format together with a dedicated interface as well as in the new LHAPDF6 format [90]. In the future, the LHAPDF6 grids will be also available at the LHAPDF website [91].

We provide tables for both bound proton PDFs $f^{p/A}(x, Q)$ as well as grids for the resulting full nuclear PDFs $f^A = Z/Af^{p/A} + (A-Z)/Af^{n/A}$. The bound proton PDFs allow a direct comparison of the nPDFs for different A values as displayed in Fig. 6. On the other hand, the full nuclear PDFs can be used directly to calculate cross sections for the nuclear collisions.

At the moment, we distribute grids for a selection of nuclei that are commonly used in the high energy/nuclear experiments. In particular, we provide girds for $\{\mathrm{He}, \mathrm{Li}, \mathrm{Be}, \mathrm{C}, \mathrm{N}, \mathrm{Al}, \mathrm{Ca}, \mathrm{Fe}, \mathrm{Cu}, \mathrm{Kr}, \mathrm{Ag}, \mathrm{Sn}, \mathrm{Xe}, \mathrm{W}, \mathrm{Au}, \mathrm{Pb}\}$. Since our parametrization is continuous in A and Z, it allows us to generate PDFs for any nuclei or isotopes. In case users are interested in having the nCTEQ15 distributions for a nucleus that is not included in our standard selection, we can generate the PDFs upon request.

[1] J. C. Collins, D. E. Soper, and G. F. Sterman, Factorization for short distance hadron—Hadron scattering, Nucl. Phys. **B261**, 104 (1985).

[2] G. T. Bodwin, Factorization of the Drell-Yan cross-section in perturbation theory, Phys. Rev. D **31**, 2616 (1985).

[3] J. C. Collins, Hard scattering factorization with heavy quarks: A General treatment, Phys. Rev. D **58**, 094002 (1998).

[4] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, CT10 next-to-next-to-leading order global analysis of QCD, Phys. Rev. D **89**, 033009 (2014).

[5] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions for the LHC Run II, J. High Energy Phys. 04 (2015) 040.

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

[7] S. Alekhin, J. Bluemlein, and S. Moch, The ABM parton distributions tuned to LHC data, Phys. Rev. D **89**, 054028 (2014).

[8] J. Owens, A. Accardi, and W. Melnitchouk, Global parton distributions with nuclear and finite-Q^2 corrections, Phys. Rev. D **87**, 094012 (2013).

[9] P. Jimenez-Delgado and E. Reya, Delineating parton distributions and the strong coupling, Phys. Rev. D **89**, 074049 (2014).

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

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

[12] K. Eskola, H. Paukkunen, and C. Salgado, EPS09: A new generation of NLO and LO nuclear parton distribution functions, J. High Energy Phys. 04 (2009) 065.

[13] I. Schienbein, J. Y. Yu, K. Kovařík, C. Keppel, J. G. Morfín, F. I. Olness, and J. F. Owens, PDF nuclear corrections for charged and neutral current processes, Phys. Rev. D **80**, 094004 (2009).

[14] 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).

[15] J.-W. Qiu, QCD Factorization and Re-scattering in Proton-Nucleus Collisions, arXiv:hep-ph/0305161.

[16] A. Accardi *et al.*, Hard probes in heavy ion collisions at the LHC: PDFs, shadowing and pA collisions, arXiv:hep-ph/0308248.

[17] J. G. Morfin, J. Nieves, and J. T. Sobczyk, Recent developments in neutrino/antineutrino-nucleus interactions, Adv. High Energy Phys. **2012**, 1 (2012).

[18] N. Armesto, Nuclear shadowing, J. Phys. G **32**, R367 (2006).

[19] B. Z. Kopeliovich, J. G. Morfin, and I. Schmidt, Nuclear shadowing in electro-weak interactions, Prog. Part. Nucl. Phys. **68**, 314 (2013).

[20] I. Schienbein, J. Y. Yu, C. Keppel, J. G. Morfín, F. Olness, and J. F. Owens, Nuclear parton distribution functions from neutrino deep inelastic scattering, Phys. Rev. D **77**, 054013 (2008).

[21] K. Kovařík, I. Schienbein, F. I. Olness, J. Y. Yu, C. Keppel, J. G. Morfín, J. F. Owens, and T. Stavreva, Nuclear Corrections in Neutrino-Nucleus DIS and Their Compatibility with Global NPDF Analyses, Phys. Rev. Lett. **106**, 122301 (2011).

[22] H. Paukkunen and C. A. Salgado, Agreement of Neutrino Deep Inelastic Scattering Data with Global Fits of Parton Distributions, Phys. Rev. Lett. **110**, 212301 (2013).

[23] M. 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).

[24] 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).

[25] Kramer, Michael, F. I. Olness, and D. E. Soper, Treatment of heavy quarks in deeply inelastic scattering, Phys. Rev. D **62**, 096007 (2000).

[26] S. Kretzer and I. Schienbein, Heavy quark initiated contributions to deep inelastic structure functions, Phys. Rev. D **58** (1998).

[27] M. Guzzi, P. M. Nadolsky, H.-L. Lai, and C.-P. Yuan, General-mass treatment for deep inelastic scattering at two-loop accuracy, Phys. Rev. D **86**, 053005 (2012).

[28] T. Stavreva, F. I. Olness, I. Schienbein, T. Ježo, A. Kusina, K. Kovařík, and J. Y. Yu, Heavy quark production in the ACOT scheme at NNLO and N3LO, Phys. Rev. D **85**, 114014 (2012).

[29] M. Botje, QCDNUM: Fast QCD evolution and convolution, Comput. Phys. Commun. **182**, 490 (2011).

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

[31] P. Aurenche, M. Fontannaz, J. Guillet, B. A. Kniehl, and M. Werlen, Large p_T inclusive π^0 cross-sections and next-to-leading-order QCD predictions, Eur. Phys. J. C **13**, 347 (2000).

[32] Incnlo version 1.4., http://lapth.cnrs.fr/PHOX_FAMILY/readme_inc.html.

[33] T. Stavreva, I. Schienbein, F. Arleo, K. Kovařík, F. Olness, J. Y. Yu, and J. F. Owens, Probing gluon and heavy-quark nuclear PDFs with gamma + Q production in pA collisions, J. High Energy Phys. 01 (2011) 152.

[34] J. Owens, J. Huston, C. E. Keppel, S. Kuhlmann, J. G. Morfín, 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).

[35] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, New generation of parton distributions with uncertainties from global qcd analysis, J. High Energy Phys. 07 (2002) 012.

[36] D. Stump, J. Huston, J. Pumplin, W.-K. Tung, H.-L. Lai, S. Kuhlmann, and J. F. Owens, Inclusive jet production, parton distributions, and the search for new physics, J. High Energy Phys. 10 (2003) 046.

[37] I. Niculescu, J. Arrington, R. Ent, and C. Keppel, Moments of nuclear and nucleon structure functions at low Q**2 and the momentum sum rule, Phys. Rev. C **73**, 045206 (2006).

[38] N. Fomin *et al.*, Scaling of the F_2 Structure Function in Nuclei and Quark Distributions at $x > 1$, Phys. Rev. Lett. **105**, 212502 (2010).

[39] D. de Florian and R. Sassot, Nuclear parton distributions at next-to-leading order, Phys. Rev. D **69**, 074028 (2004).

[40] pyMinuit package, https://code.google.com/p/pyminuit/.

[41] SEAL-Minuit, http://seal.web.cern.ch/seal/work-packages/mathlibs/minuit/release/download.html.

[42] F. James, MINUIT Reference Manual, CERN Program Library Writeup D506, https://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html.

[43] J. Pumplin, D. Stump, and W. Tung, Multivariate fitting and the error matrix in global analysis of data, Phys. Rev. D **65**, 014011 (2001).

[44] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Uncertainties of predictions from parton distribution functions. 2. The Hessian method, Phys. Rev. D **65**, 014013 (2001).

[45] Low-noise lanczos differentiators, http://www.holoborodko.com/pavel/numerical-methods/numerical-derivative/lanczos-low-noise-differentiators/.

[46] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method, Phys. Rev. D **65**, 014012 (2001).

[47] A. Martin, W. Stirling, R. Thorne, and G. Watt, Parton distributions for the LHC, Eur. Phys. J. C **63**, 189 (2009).

[48] M. Arneodo *et al.* (New Muon Collaboration), Measurement of the proton and deuteron structure functions, $F_2{}^p$ and $F_2{}^d$, and of the ratio σ_L/σ_T, Nucl. Phys. **B483**, 3 (1997).

[49] A. Airapetian *et al.* (HERMES Collaboration), Measurement of $R = \sigma_L/\sigma_T$ in Deep-Inelastic Scattering on Nuclei, arXiv:hep-ex/0210068.

[50] 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).

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

[52] M. Arneodo *et al.* (New Muon Collaboration), The structure function ratios F_2^{Li}/F_2^D and F_2^C/F_2^D at small x, Nucl. Phys. **B441**, 12 (1995).

[53] 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).

[54] 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).

[55] M. Arneodo *et al.* (European Muon Collaboration), Measurements of the nucleon structure function in the range $0.002 < x < 0.17$ and $0.2\ \mathrm{GeV}^2 < Q^2 < 8\ \mathrm{GeV}^2$ in deuterium, carbon and calcium, Nucl. Phys. **B333**, 1 (1990).

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

[57] A. Bodek *et al.*, A Comparison of the Deep Inelastic Structure Functions of Deuterium and Aluminum Nuclei, Phys. Rev. Lett. **51**, 534 (1983).

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

[59] 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).

[60] 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).

[61] 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).

[62] M. R. Adams *et al.* (E665 Collaboration), Saturation of Shadowing at Very Low x_{Bj}, Phys. Rev. Lett. **68**, 3266 (1992).

[63] M. Arneodo *et al.* (New Muon Collaboration), The A dependence of the nuclear structure function ratios, Nucl. Phys. **B481**, 3 (1996).

[64] 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).

[65] D. M. Alde *et al.*, Nuclear Dependence of Dimuon Production at 800 GeV. FNAL-772 Experiment, Phys. Rev. Lett. **64**, 2479 (1990).

[66] M. A. Vasilev *et al.* (FNAL E866 Collaboration), Parton Energy Loss Limits and Shadowing in Drell-Yan Dimuon Production, Phys. Rev. Lett. **83**, 2304 (1999).

[67] S. Adler *et al.* (PHENIX Collaboration), Centrality Dependence of pi0 and eta Production at Large Transverse Momentum in $s(\mathrm{NN}) * *(1/2) = 200$-GeV d + Au Collisions, Phys. Rev. Lett. **98**, 172302 (2007).

[68] B. Abelev *et al.* (STAR Collaboration), Inclusive π^0, η, and direct photon production at high transverse momentum in $p + p$ and $d + \mathrm{Au}$ collisions at $\sqrt{s_{NN}} = 200$ GeV, Phys. Rev. C **81**, 064904 (2010).

[69] 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).

[70] 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).

[71] J. Berge *et al.*, A measurement of differential cross-sections and nucleon structure functions in charged current neutrino interactions on iron, Z. Phys. C **49**, 187 (1991).

[72] G. Onengut *et al.* (CHORUS Collaboration), Measurement of nucleon structure functions in neutrino scattering, Phys. Lett. B **632**, 65 (2006).

[73] M. Tzanov *et al.* (NuTeV Collaboration), Precise measurement of neutrino and anti-neutrino differential cross sections, Phys. Rev. D **74**, 012008 (2006).

[74] H. Paukkunen and C. A. Salgado, Compatibility of neutrino DIS data and global analyses of parton distribution functions, J. High Energy Phys. 07 (2010) 032.

[75] G. Aad *et al.* (ATLAS Collaboration), Measurement of the production and lepton charge asymmetry of W bosons in $\mathrm{Pb} + \mathrm{Pb}$ collisions at $\sqrt{\mathbf{s_{NN}}} = \mathbf{2.76\ TeV}$ with the ATLAS detector, Eur. Phys. J. C **75**, 23 (2015).

[76] G. Aad *et al.* (ATLAS Collaboration), Z boson production in $p + \mathrm{Pb}$ collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV measured with the ATLAS detector, Phys. Rev. C **92**, 044915 2015.

[77] S. Chatrchyan *et al.* (CMS Collaboration), Study of Z production in PbPb and pp collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV

in the dimuon and dielectron decay channels, J. High Energy Phys. 03 (2015) 022.

[78] V. Khachatryan *et al.* (CMS Collaboration), Study of W boson production in pPb collisions at sqrt(s[NN]) = 5.02 TeV, Phys. Lett. B **750**, 565 (2015).

[79] 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).

[80] I. C. Cloet, W. Bentz, and A. W. Thomas, Isovector EMC Effect Explains the NuTeV Anomaly, Phys. Rev. Lett. **102**, 252301 (2009).

[81] D. Dutta, J. C. Peng, I. C. Cloet, and D. Gaskell, Pion-induced Drell-Yan processes and the flavor-dependent EMC effect, Phys. Rev. C **83**, 042201 (2011).

[82] A. Accardi, M. E. Christy, C. E. Keppel, W. Melnitchouk, P. Monaghan, J. G. Morfín, and J. F. Owens, New parton distributions from large-x and low-Q^2 data, Phys. Rev. D **81**, 034016 (2010).

[83] A. Accardi, W. Melnitchouk, J. F. Owens, M. E. Christy, C. E. Keppel, L. Zhu, and J. G. Morfín, Uncertainties in determining parton distributions at large x, Phys. Rev. D **84**, 014008 (2011).

[84] J. Binnewies, B. A. Kniehl, and G. Kramer, Next-to-leading order fragmentation functions for pions and kaons, Z. Phys. C **65**, 471 (1995).

[85] B. A. Kniehl, G. Kramer, and B. Potter, Fragmentation functions for pions, kaons, and protons at next-to-leading order, Nucl. Phys. **B582**, 514 (2000).

[86] D. d'Enterria, K. J. Eskola, I. Helenius, and H. Paukkunen, Confronting current NLO parton fragmentation functions with inclusive charged-particle spectra at hadron colliders, Nucl. Phys. **B883**, 615 (2014).

[87] 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).

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

[89] nCTEQ website, http://ncteq.hepforge.org/.

[90] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, LHAPDF6: Parton density access in the LHC precision era, Eur. Phys. J. C **75**, 132 (2015).

[91] LHAPDF website, https://lhapdf.hepforge.org/.