

Impact of inclusive hadron production data on nuclear gluon PDFs

P. Duwentäster^{1,*}, L. A. Husová,² T. Ježo,³ M. Klasen,¹ K. Kovařík,¹ A. Kusina,⁴ K. F. Muzakka,¹
F. I. Olness,⁵ I. Schienbein,⁶ and J. Y. Yu⁵

¹*Institut für Theoretische Physik, Westfälische Wilhelms-Universität Münster,
Wilhelm-Klemm-Strasse 9, D-48149 Münster, Germany*

²*Institut für Kernphysik, Westfälische Wilhelms-Universität Münster,
Wilhelm-Klemm-Strasse 9, D-48149 Münster, Germany*

³*Institute for Theoretical Physics, KIT, D-76131 Karlsruhe, Germany*

⁴*Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 Krakow, Poland*

⁵*Southern Methodist University, Dallas, Texas 75275, USA*

⁶*Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes,
CNRS/IN2P3, 53 avenue des Martyrs, 38026 Grenoble, France*



(Received 17 June 2021; accepted 30 September 2021; published 5 November 2021)

A precise knowledge of nuclear parton distribution functions (nPDFs) is—among other things—important for the unambiguous interpretation of hard process data taken in pA and AA collisions at the Relativistic Heavy Ion Collider (RHIC) and the LHC. The available fixed target data for deep inelastic scattering (DIS) and Drell-Yan (DY) lepton pair production mainly constrain the light quark distributions. It is hence crucial to include more and more collider data in global analyses of nPDFs in order to better pin down the different parton flavors, in particular the gluon distribution at small x . To help constrain the nuclear gluon PDF, we extend the nCTEQ15 analysis by including single inclusive hadron (SIH) production data from RHIC (PHENIX and STAR) and LHC (ALICE). In addition to the DIS, DY, and SIH datasets, we will also include LHC W/Z production data. As the SIH calculation is dependent on hadronic fragmentation functions (FFs), we use a variety of FFs available in the literature to properly estimate this source of uncertainty. We study the impact of these data on the PDFs and compare with both the nCTEQ15 and nCTEQ15WZ sets. The calculations are performed using a new implementation of the nCTEQ code (nCTEQ++) including a modified version of INCNLO, which allows faster calculations using precomputed grids. The extension of the nCTEQ15 analysis to include the SIH data represents an important step toward the next generation of PDFs.

DOI: [10.1103/PhysRevD.104.094005](https://doi.org/10.1103/PhysRevD.104.094005)

I. INTRODUCTION

Parton distribution functions (PDFs) are fundamental quantities required to calculate predictions for any process involving hadrons in the initial state. The QCD parton model has been used successfully to make predictions for a variety of experiments at SLAC, HERA, Tevatron, Relativistic Heavy Ion Collider (RHIC) and LHC. This theoretical framework will also be essential for both the physics program of the EIC and proposed future experiments such as the FCC. While precise constraints have been imposed on the proton PDFs, for the case of nuclear PDFs (nPDFs), there is still much room for improvement of the

uncertainties [1–22]. The gluon PDFs are particularly problematic because the cross sections for the deep inelastic scattering (DIS) and the Drell-Yan (DY) processes, which represent the bulk of the precision data in nPDF fits like nCTEQ15 [3], are not directly sensitive to the gluon PDF at leading order.

While many different microscopic models for nuclear effects on PDFs exist, no unambiguous picture has yet emerged for either the shadowing region [23–26], anti-shadowing region [26–28], or the EMC effect [26,29–33]. A particularly promising unified approach is provided by the color glass condensate [34,35]. On the other hand, unbiased fits to the experimental data provide important global constraints on these theoretical ideas and are an indispensable ingredient for many current and future experimental (i.e., at the LHC, but also RHIC and EIC) and theoretical analyses (e.g., for the very successful Statistical Hadronization Model describing the freeze-out of the quark gluon plasma [36]). This is the approach we take in the following. Note that there are currently ongoing

*pit.duw@uni-muenster.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

studies at the LHC of final-state effects caused by the medium created in pA and even pp collisions [37,38]. In our analysis below, we will demonstrate that our results are largely independent of the final-state hadron fragmentation and thus that our interpretation of the nuclear effects as modifications of a cold initial state is currently totally consistent with the available experimental data.

A. Gluon PDF

Single inclusive hadron (SIH) production data have the potential to yield new constraints on the gluon PDF because the gluon contributes a significant part to the overall cross section of this process. The importance of the gluon contribution can be seen in Fig. 1, which shows the fractional contribution to the process $p + \text{Pb} \rightarrow \pi^0 + X$ as a function of the transverse momentum p_T for the various subprocesses initiated by gluons, up, down, and strange partons inside a lead nucleus. In particular, the red shaded area shows the fraction where a parton from the

proton interacts with a gluon from the lead nucleus to produce a neutral pion. The gluon contribution dominates in the low- to mid- p_T region at a center-of-mass energy per nucleon of $\sqrt{s_{NN}} = 200$ GeV. At 5 TeV, the gluon is the dominant contribution even in the mid- to high- p_T region. The remaining contribution is shared roughly evenly between the up and down quarks, while the antiquarks (including up and down) contribute a minor fraction. Charm, bottom, and top are omitted in this figure due to their negligible contributions, but they are fully incorporated in the calculation. The partonic fractions for kaons and eta mesons are similar to those of pions, so we do not present a separate figure.

Figure 2 shows the relative contributions to the cross section of $p + \text{Pb} \rightarrow \pi^0 + X$ of each parton's fragmentation function (FF). For instance, the red area shows the contribution from processes where the initial scattering event produces a gluon, which then fragments into a neutral pion. These contributions are very similar to those of the

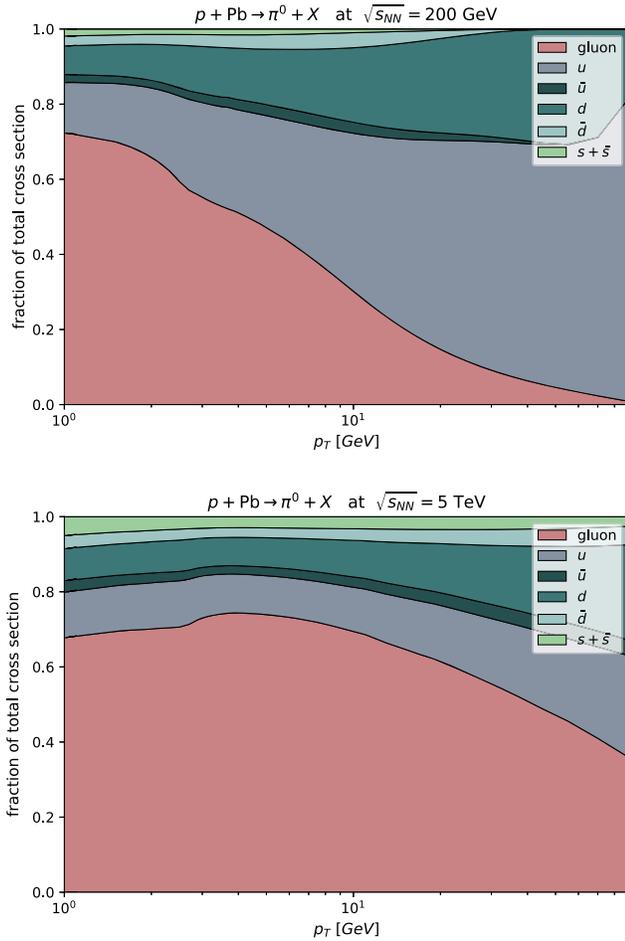


FIG. 1. Fractional contributions of the total $p + \text{Pb} \rightarrow \pi^0 + X$ cross section initiated by each PDF flavor $f_i^{\text{Pb}}(x, Q)$ of the lead nucleus at $\sqrt{s_{NN}} = 200$ GeV (upper panel) and 5 TeV (lower panel) for $i \in \{g, u, d, \bar{u}, \bar{d}, s + \bar{s}\}$.

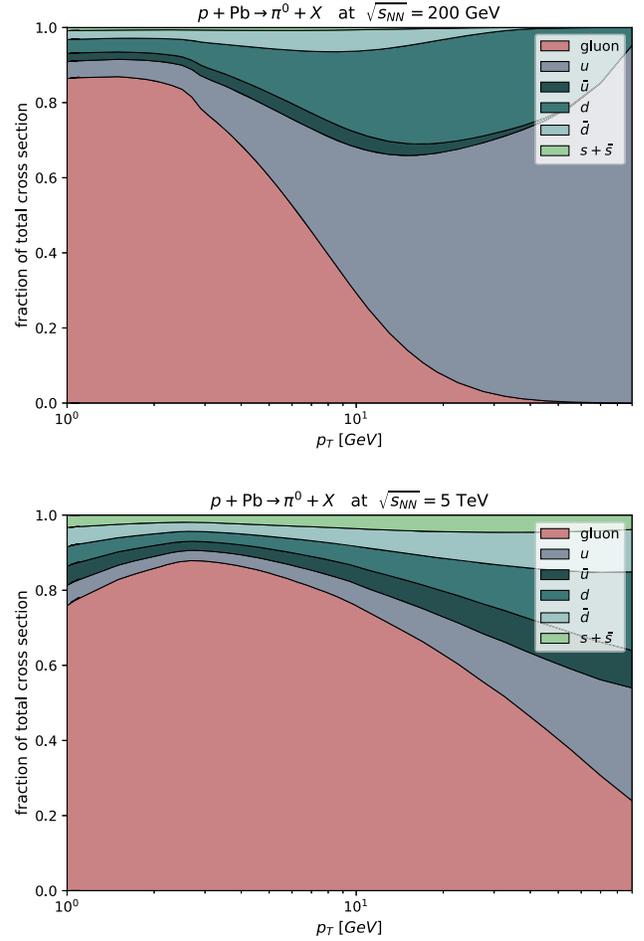


FIG. 2. Fractional contribution of the total $p + \text{Pb} \rightarrow \pi^0 + X$ cross section contributed by each fragmentation function, $D_i^{\pi^0}(z, Q)$, at $\sqrt{s_{NN}} = 200$ GeV (upper panel) and 5 TeV (lower panel) for $i \in \{g, u, d, \bar{u}, \bar{d}, s + \bar{s}\}$.

PDF flavors (Fig. 1), but with slightly larger contributions from the antiquarks. Both figures are computed with nCTEQ15WZ PDFs [39] and DSS FFs [40], but there are no qualitative differences when other nPDFs or FFs are used.

In this investigation, we will study single inclusive hadron production in proton-lead and deuterium-gold collisions. The focus will be to incorporate this process into the global analysis, including the dependence of the fragmentation function, and to determine the resulting impact on the nuclear gluon PDF. The remainder of this section provides an overview of the nCTEQ framework and the available datasets. In Sec. II, we investigate the fragmentation function dependence, along with other theory considerations like the scale dependence. In Sec. III, we present the fits obtained using the SIH data and compare with the theoretical predictions. The main conclusions are summarized in Sec. IV.

B. nCTEQ++ framework

The nCTEQ project expands upon the foundation of the proton PDF global fitting analysis by including the nuclear dimension. In early proton PDF analyses (e.g., Ref. [41]), the nuclear data were used to calculate correction factors, which were then applied to the proton PDF fit without any uncertainties. In contrast, the nCTEQ framework enables full communication between nuclear and proton data, which means that observed tensions between datasets can be investigated through the lens of nuclear corrections.

The details of the nCTEQ15 nPDFs are presented in Ref. [3]. The current analysis, along with the other recent nCTEQ analyses, such as nCTEQ15WZ [39] and nCTEQ15HIX [42], are performed with a new C++ based code nCTEQ++. This allows us to easily interface external programs such as HOPPET [43], APPLgrid[44], and INCNLO [45]. In particular, we work at leading twist and next-to-leading order (NLO) of QCD for both the PDF and FF evolution equations as well as the hard scattering coefficients.

For the fits in this investigation, we use the same 19 parameters as for the nCTEQ15WZ set. These 19 parameters include the 16 free parameters of the nCTEQ15 analysis, with an additional three open parameters for the strange distribution. Recall that for the nCTEQ15 set the strange PDF was constrained by the relation $s = \bar{s} = (\kappa/2)(\bar{u} + \bar{d})$ at the initial scale $Q_0 = 1.3$ GeV so that it had the same form as the other sea quarks.

Our PDFs are parametrized at the initial scale $Q_0 = 1.3$ GeV as

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

and the nuclear A dependence is encoded in the coefficients as

$$c_k \rightarrow c_k(A) \equiv p_k + a_k(1 - A^{-b_k}), \quad (2)$$

where $k = \{1, \dots, 5\}$. The 16 free parameters used for the nCTEQ15 set describe the x -dependence of the $\{g, u_v, d_v, \bar{d} + \bar{u}\}$ PDF combinations, and we do not vary the \bar{d}/\bar{u} parameters; see Ref. [3] for details. As in the nCTEQ15WZ analysis, we have added three strange PDF parameters: $\{a_0^{s+\bar{s}}, a_1^{s+\bar{s}}, a_2^{s+\bar{s}}\}$; these parameters correspond to the nuclear modification of the overall normalization, the low x exponent and the large- x exponent of the strange distribution, respectively. In total, the 19 open parameters are

$$\{a_1^{u_v}, a_2^{u_v}, a_4^{u_v}, a_5^{u_v}, a_1^{d_v}, a_2^{d_v}, a_5^{d_v}, a_1^{\bar{u}+\bar{d}}, a_5^{\bar{u}+\bar{d}}, a_1^g, a_4^g, a_5^g, b_0^g, b_1^g, b_4^g, b_5^g, a_0^{s+\bar{s}}, a_1^{s+\bar{s}}, a_2^{s+\bar{s}}\}.$$

To obtain the cross section for single inclusive hadron production, the PDFs of the two initial-state particles are convoluted with the cross section of the partonic subprocess and the final-state fragmentation function,

$$\sigma_{p+\text{Pb} \rightarrow h+X} = f_a^p \otimes f_b^{\text{Pb}} \otimes \hat{\sigma}_{ab \rightarrow c} \otimes D_c^h, \quad (3)$$

where h is the produced light hadron and a sum over all possible subprocesses $ab \rightarrow c + X$ is understood. The twist-2 factorization formula has an error which is suppressed by a power of the ratio Λ/Q where Λ is a hadronic scale and Q is the hard scale of the process (for example, the p_T of the light hadron). This factorization formula is the result of a rigorous factorization theorem (see Refs. [46,47] and references therein) originally devised for pp collisions. It is supposed to hold true also for pA collisions; however, the error (higher twist terms) is possibly enhanced by the nuclear A , and one has to assess phenomenologically which minimum value for the hard scale is necessary for the twist-2 factorization formula to be a good approximation.¹ A detailed overview is given in Ref. [50]. Performing all convolutions for each data point in each iteration of a fit is, however, too computationally expensive. A solution is to perform the convolution of the proton PDF (or deuteron PDF in case of RHIC data) and the pion FF ahead of time and store the results to a grid; thus, the cross section evaluation can be reduced to a single convolution during the fitting process. In order to perform the corresponding calculations and produce such grids, we have modified the INCNLO [45] program. The obtained grids have been validated to reproduce the full calculation within a margin significantly smaller than the data uncertainty.

¹For additional details regarding target mass corrections, see Refs. [48,49].

TABLE I. Overview of the available datasets, including their center-of-mass energy, observable, and number of data points.

Dataset	Ref.	ID	$\sqrt{s_{NN}}$ [GeV]	Observable	Number of points
PHENIX π^0	[51]	4003	200	R_{dAu}	21
PHENIX η	[51]	4403	200	R_{dAu}	12
PHENIX π^\pm	[52]	4103	200	R_{dAu}	20
PHENIX K^\pm	[52]	4203	200	R_{dAu}	15
STAR π^0	[53]	4002	200	R_{dAu}	13
STAR η	[53]	4402	200	R_{dAu}	7
STAR π^\pm	[54]	4102	200	R_{dAu}	23
ALICE 5 TeV π^0	[55]	4001	5020	R_{pPb}	31
ALICE 5 TeV η	[55]	4401	5020	R_{pPb}	16
ALICE 5 TeV π^\pm	[56]	4101	5020	R_{pPb}	58
ALICE 5 TeV K^\pm	[56]	4201	5020	R_{pPb}	58
ALICE 8 TeV π^0	[57]	4004	8160	R_{pPb}	30
ALICE 8 TeV η	[57]	4404	8160	R_{pPb}	14

C. SIH datasets

In this analysis, we include the same DIS and DY lepton pair production data as in the nCTEQ15 analysis. The original nCTEQ15 analysis used also the RHIC pion data allowing to provide constraints on the gluon PDF. We now extend this analysis to include additional hadrons from the RHIC data, as well as new data from recent ALICE analyses. We will study four types of hadrons: neutral pions, charged pions, charged kaons, and eta mesons. The charged mesons always appear as the average of their positively and negatively charged version, which is also how the neutral pions are calculated for their fragmentation

functions. The data are taken at center-of-mass energies per nucleon of 200 (PHENIX, STAR), 5020, and 8160 GeV (ALICE). Table I gives an overview of the available datasets, while Fig. 3 shows the p_T distribution of the available data points for each set.

As in other types of experiments, kinematic cuts are applied to remove data that cannot be adequately described by the theory. For example, in the very low- p_T region, the SIH process becomes non-perturbative, so we will impose a lower p_T cut on the data. Additional restrictions may come from the FFs $D_i^h(z, Q^2)$ required to compute the cross sections. The available sets of FFs are typically only

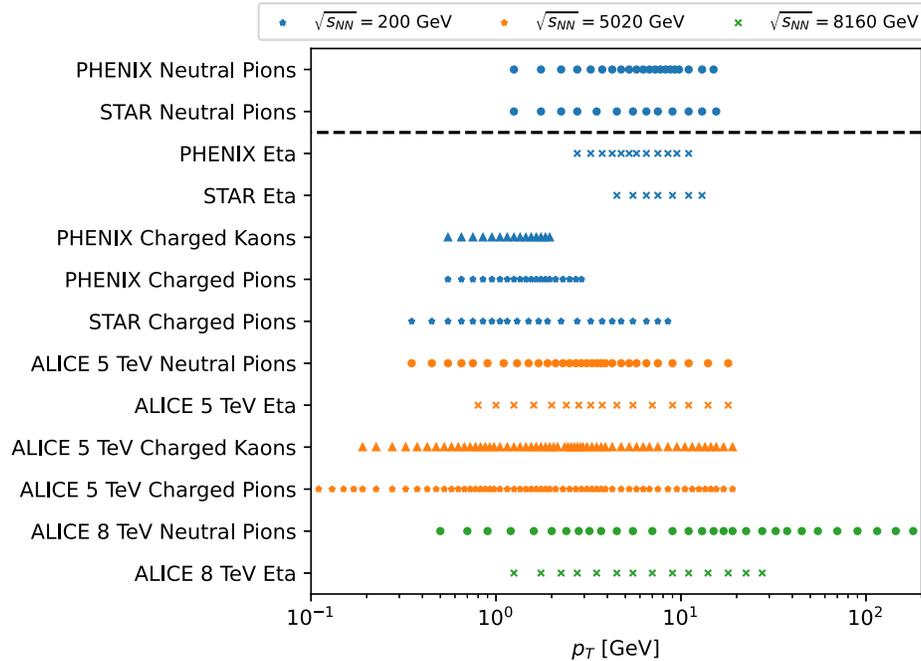


FIG. 3. p_T values of all data points, separated by set and colored by $\sqrt{s_{NN}}$. The datasets above the dashed line were included in the nCTEQ15 analysis.

reliable for momentum fractions $z_{\min} < z < 1$ with some minimal $z_{\min} \sim 0.1$. This issue will be further discussed in Sec. III A.

All the single inclusive hadron production data are given in terms of ratios

$$R_{dAu}^h = \frac{1}{A_d A_{Au}} \frac{\sigma_{d+Au \rightarrow h+X}}{\sigma_{p+p \rightarrow h+X}} \quad (\text{RHIC}) \quad (4)$$

$$R_{pPb}^h = \frac{1}{A_{Pb}} \frac{\sigma_{p+Pb \rightarrow h+X}}{\sigma_{p+p \rightarrow h+X}} \quad (\text{ALICE}). \quad (5)$$

All fits are also repeated including the W^\pm and Z production data from our recent nCTEQ15WZ [3] analysis because these data have a significant impact on the gluon PDF. Since the impact of SIH production on the high- x region is negligible, we do not include a further comparison with the nCTEQ15HIX [42] analysis.

Since all SIH datasets have considerable normalization uncertainties, we need to account for those to avoid datasets pulling the fit in unphysical directions due to their normalization errors. This is done by including a normalization factor for each set as a parameter in the fit. By adding a χ^2 penalty that increases as normalizations stray further from unity, we ensure that they are kept at reasonable values. This is done according to the prescription outlined in Appendix A to avoid the D’Agostini bias [58].

II. FRAGMENTATION FUNCTIONS

In this analysis, we investigate a total of ten different FFs, as listed in Table II. We will give a brief overview of their properties and then compare them both in terms of predictions for proton-proton and proton-nucleus collisions.

A. Available fragmentation functions

In a manner complementary to the PDFs, the FFs describe the hadronization of a partonic constituent into a final-state hadron. Both the PDFs and FFs are

TABLE II. Overview of the available sets of FFs $D_i^h(z, Q)$ and their available particles.

FF	Ref.	Year	Available particles
BKK	[59]	1994	π_0, π^\pm, K^\pm
KKP	[60]	2000	π_0, π^\pm, K^\pm
KRETZER	[61]	2000	π_0, π^\pm, K^\pm
HKNS07	[62]	2007	π_0, π^\pm, K^\pm
AKK	[63]	2008	π_0, π^\pm, K^\pm
NNFF	[64]	2017	π_0, π^\pm, K^\pm
JAM20	[65]	2021	π_0, π^\pm, K^\pm
DSS14	[40]	2014	π_0, π^\pm
DSS17	[66]	2017	K^\pm
AESSS	[67]	2011	η

nonperturbative objects and hence must be obtained by fitting to data. The pioneering fits of FFs used only single-inclusive hadron production in electron-positron annihilation. In more recent analyses, groups have added data from semi-inclusive deep inelastic lepton-nucleon scattering and other processes to improve the accuracy and kinematic range of their fits [68,69].

A selection of FFs for various mesons² is shown in Table II. The fragmentation function sets listed in bold provide uncertainties. We will henceforth denote the combination of DSS14 for pions and DSS17 for kaons simply as “DSS.” Note that HKNS16 [70] exists as an updated version of HKNS07, but no code is available to use these updated fragmentation functions.

Additional fragmentation functions exist for other final states like protons, antiprotons, and unidentified charged hadrons (SGK18 [71] and NNFF1.1h [72]). Some of the aforementioned FFs (AKK, BKK, HKNS, KKP, and KRETZER) also include those, but we exclude those from the analysis due to the comparatively large uncertainties both on the data and the fragmentation functions. There have also been studies on the effect of the nuclear medium on the fragmentation [73,74], but we exclude the fragmentation functions obtained there from our analysis in order to avoid double counting of the shared data points. Also, any possible medium modifications of the FFs are small compared to the uncertainties of the FFs.

B. Comparison in proton-proton collisions

Before we examine the dAu and pPb cases, let us first look at the $p + p \rightarrow h + X$ baseline process to help understand the limitations of the theory prediction due to the uncertainties of the FFs.

Figure 4 shows a comparison of predictions from various FF sets with data taken by the PHENIX and ALICE experiments at $\sqrt{s_{NN}} = 200$ and 7000 GeV, respectively. The nCTEQ15 proton PDF is used in these calculations. The fragmentation functions displayed are BKK, KKP, DSS, NNFF, and JAM20. The KRETZER and HKNS FFs are not shown as their more strict kinematic restrictions preclude comparison with the ALICE data.

At 200 GeV, all fragmentation functions are able to describe the data for $p_T \geq 3$ GeV with a satisfactory $\chi^2/N_{\text{dof}} < 1$ if one allows for a normalization shift. Below p_T of 3 GeV, all the curves display a significant upward slope in Fig. 4, which points to a qualitative disagreement. There is also a slight upward slope toward higher p_T for all fragmentation functions, but it is well within the data uncertainties, considering the allowed normalization shift.

²We include in our analysis only data for the inclusive production of pions, eta mesons, and kaons. Note that the neutral pion FFs are always calculated as the average of π^+ and π^- FFs.

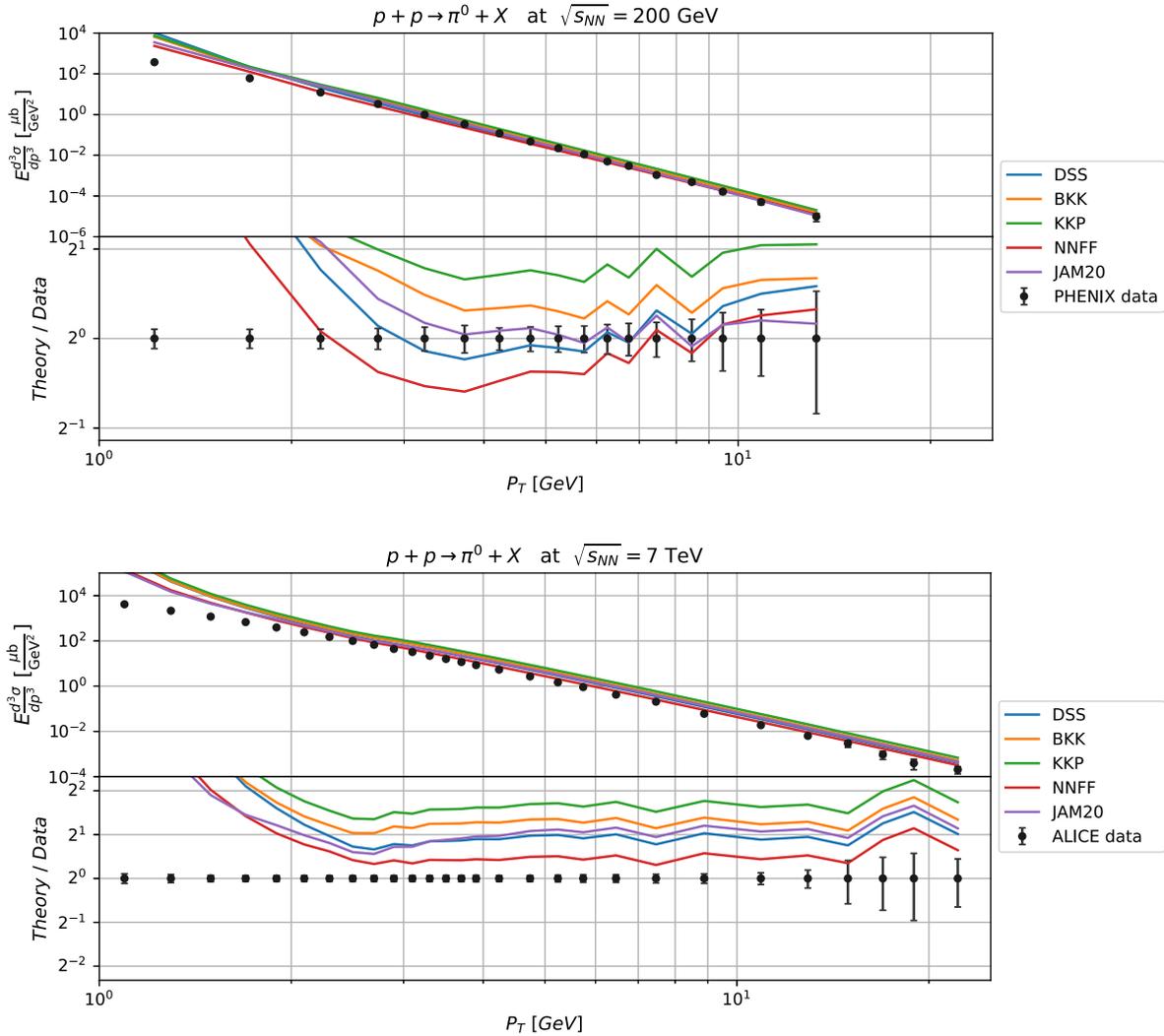


FIG. 4. Comparison of predictions made with different fragmentation functions for $p + p \rightarrow \pi^0 + X$. The calculations are done using nCTEQ15 proton PDFs. Both panels show data for neutral pions, with PHENIX data [75] in the upper and ALICE data [76] in the lower one.

At ALICE energies, the data can be well described by BKK, KKP, DSS, and NNFF down to p_T values of 3 GeV if a normalization is introduced. The JAM20 result is also relatively constant across the p_T range, but it begins to decrease slightly for lower p_T values in the range of $p_T \approx 5$ GeV and below.³ Again, the theory predictions increasingly overshoot the data the further one goes below 3 GeV. Since this effect is not fragmentation function dependent, it is also independent of the produced final state.

³In principle, for the computation of Fig. 4, the FFs should be combined with their matching PDFs, i.e., JAM20 FFs with JAM20 PDFs, and DSS FFs with MSTW2008 [77] PDFs. Since our focus is the impact on the nuclear PDFs, however, we use our proton baseline instead.

C. Scale uncertainties

The prediction for the SIH production cross section depends on three scale choices: the initial-state factorization scale μ_i , the final-state factorization scale μ_f , and the renormalization scale μ_r . Frequently, they are taken to be $\mu_i = \mu_f = \mu_r = c p_T$, where c is a constant that is commonly chosen as either $\frac{1}{2}$ or 1, but there is no unambiguous prescription for their choice.

Figure 5 shows the prediction for pion production at 200 GeV and 7 TeV with each scale varied independently between the two common choices, $c = \{\frac{1}{2}, 1\}$. The case where all scales are equal to $\frac{1}{2} p_T$ (bold, gray) gives the best description of the 200 GeV data; additionally, this is also the only scale choice that yields $\chi^2/N_{\text{dof}} < 1$ for the 7 TeV data with $p_T > 3$ GeV, if the normalization is chosen freely. Therefore, in the following comparisons, we make the choice $c = \frac{1}{2}$ going forward. This means that we need

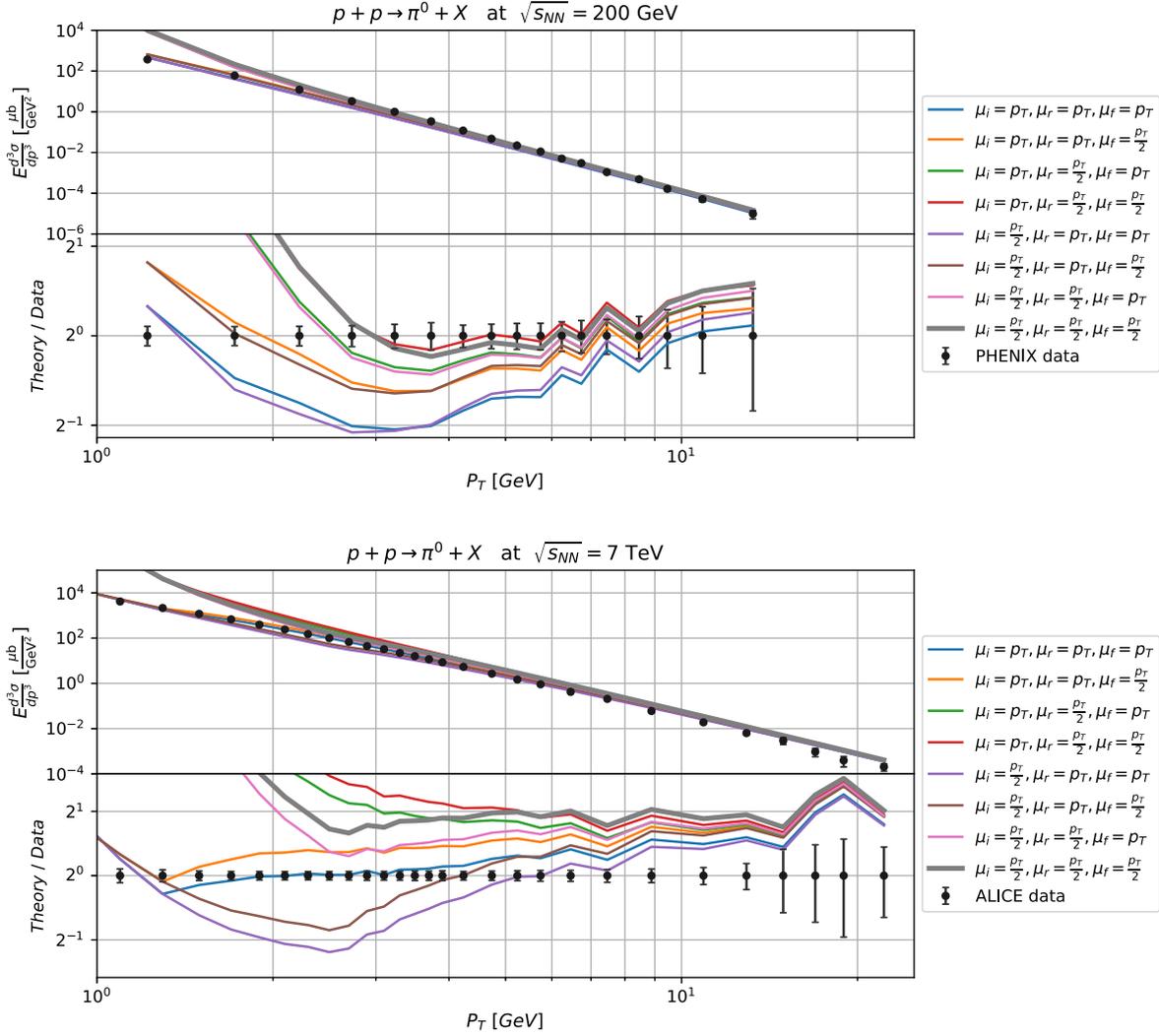


FIG. 5. Comparison of predictions made with different scale choices for $p + p \rightarrow \pi^0 + X$. The calculations are performed using nCTEQ15 proton PDFs with DSS fragmentation functions. Both panels show data for neutral pions, with PHENIX data in the upper panel and ALICE data in the lower one.

to freeze the initial-state factorization scale (μ_i) to the initial scale of our PDF evolution ($Q_0 = 1.3$ GeV) whenever $c p_T < Q_0$, i.e., for $p_T \leq 2.6$ GeV:

$$\mu_i = \begin{cases} 1.3 \text{ GeV} & \text{for } \frac{1}{2} p_T < 1.3 \text{ GeV}, \\ \frac{1}{2} p_T & \text{otherwise.} \end{cases} \quad (6)$$

Otherwise, we would have to interpolate the PDFs to scales below the initial scale, which is technically challenging.

Note that the 200 GeV and 7 TeV datasets shown in Fig. 5 are actually included in the fit of the DSS FFs, where they impose a cut of $p_T \geq 5$ GeV in their analysis, which uses a scale choice of $\mu_i = \mu_f = \mu_r = p_T$ (blue curve).

D. Comparison in dAu and pPb collisions

We now examine the impact of the different FFs on the nuclear ratios $R_{p\text{Pb}}$ and $R_{d\text{Au}}$ for pion, kaon, and eta

production. Figure 6 compares all the datasets with predictions using the nCTEQ15WZ nuclear PDFs for each set of FFs. Data taken at $p_T < 1$ GeV are not shown as the twist-2 formula for the theory is certainly not valid in that region. We also display the uncertainty band for the DSS FFs to gauge the spread of the various FFs as this represents a typical FF uncertainty. We observe that the predictions with BKK, KKP, and DSS show very close agreement. The most notable difference between BKK and KKP is seen in the charged kaon production, where KKP lies a bit lower for high p_T values. NNFF and JAM20 also agree very well with each other across all datasets, and the only instance where they lie outside of the uncertainty given by DSS is for the high- p_T ALICE pion data. Since the data uncertainties in this region are quite large, this should not have any significant impact on our fits. In the kinematic region where AKK allows predictions, they also agree with the previous FFs. The KRETZER FFs show some qualitative

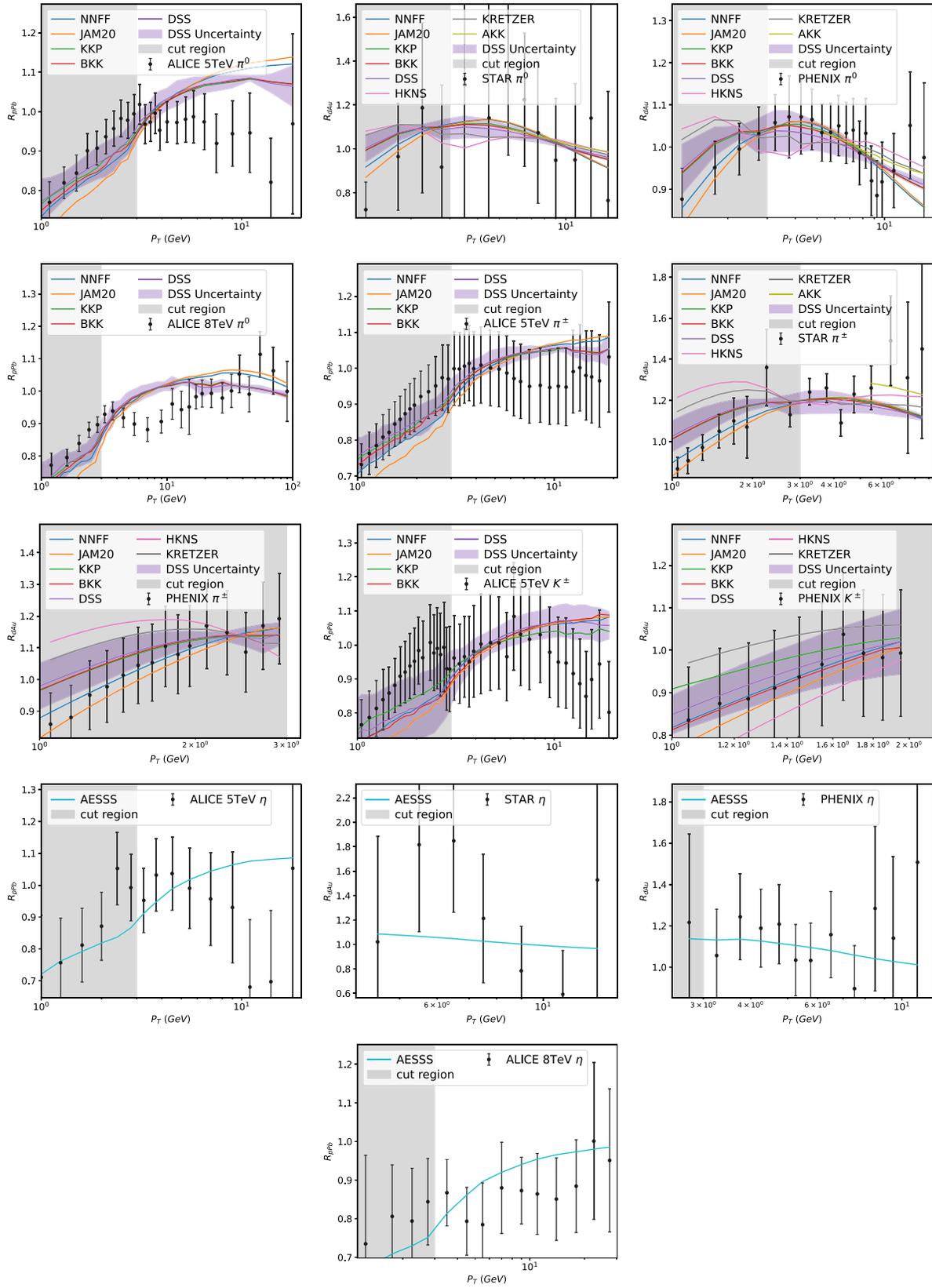


FIG. 6. Comparison of data for nuclear ratios $R_{AA'}$ for pion, kaon, and eta production with theoretical predictions at NLO QCD using nCTEQ15WZ nuclear PDFs and different FFs. The predictions are scaled by the inverse of their fitted normalization. The uncertainties of the DSS fragmentation functions are shown as purple bands. The gray region shows the data that are cut from the fits.

differences in the region just above the cut but lie within the uncertainty of DSS. For HKNS, the disagreement is slightly larger but still well below the data uncertainties. The predictions made with AESSS agree well with the eta meson production data, but since AESSS is the only available fragmentation function for eta mesons, no comparisons can be made.

We calculate fragmentation function uncertainties from the DSS FFs (see below) for each data point and add these as a systematic uncertainty in our fit. Although these uncertainties also depend on the PDF, this dependence is very weak and can be neglected. Note also that the predictions are already quite close to the data values. While this suggests that the data will not significantly change the central value of the PDFs, the data may well reduce their uncertainties.

E. Uncertainties of fragmentation functions

We now consider the FF uncertainties in further detail. Four of the available FF sets include uncertainties; HKNS and DSS provide their uncertainties in terms of Hessian eigenvectors, while JAM20 and NNFF provide Monte Carlo replicas. We show the uncertainties for NNFF and JAM20 in Figs. 16 and 17 of Appendix B. The HKNS FFs yield uncertainties that are larger than the data uncertainty for p_T values below 10 GeV; hence, they will not help constrain the PDFs in this kinematic region. The NNFF fragmentation functions yield slightly larger uncertainties than those of DSS shown in Fig. 6. This may be due, in part, to the use of a parametrization-free neural network instead of a “traditional” parametrization and a slightly smaller dataset. Lastly, the uncertainties of the JAM20 fragmentation are so small across the kinematic region with $p_T > 1$ GeV that they can be neglected when compared with the data uncertainty.

It is important to note that the displayed bands do not reflect the full uncertainty of the theory prediction but rather represent a lower bound for the following reasons: First, the theory predictions for low p_T points may depend on fragmentation functions extrapolated beyond their fitted kinematic region, and the accuracy of the Hessian method outside of the region where data exist is heavily dependent on the specific parametrization of the FF. Even more important are low p_T corrections. As we move to lower values of p_T , perturbation theory begins to break down as contributions from nonperturbative sources increase. This can also make it difficult to disentangle initial- from final-state effects in hadron production processes. Medium suppression due to energy loss can be observed not only in AA but even in pA and pp collisions [78–80]. These higher twist effects, however, are suppressed by powers of the hard scale p_T . Thus, for our predictions in the lower p_T range, we may reach the transition region between stable perturbative predictions and unreliable nonperturbative predictions. These factors are the reason why we need to impose cuts on low p_T values to ensure reliable predictions.

F. Fragmentation kinematics

Finally, it is interesting to investigate the correspondence of the p_T value of the data to the z region of the fragmentation function.

Figure 7 shows the contribution of different z regions to the total $p + Pb \rightarrow \pi^0 + X$ cross section calculated using nCTEQ15WZ PDFs and DSS fragmentation functions at 200 GeV and 5 TeV. We can see that the $z < 0.2$ region does not have a substantial contribution to the cross section at 200 GeV, and even the region $z < 0.4$ hardly contributes above $p_T > 3.0$ GeV.

In the lower panel, we see that the higher energy (5 TeV) shifts the z bands toward lower z values, with the $z < 0.2$ region still contributing a non-negligible amount even at $p_T \approx 10$ GeV. The $z < 0.1$ region starts contributing below $p_T = 4$ GeV, but stays below 10%. Since most fragmentation functions include data at least down to $z = 0.05$ in their fits, this eliminates concerns about FF extrapolation having any significant impact on our results.

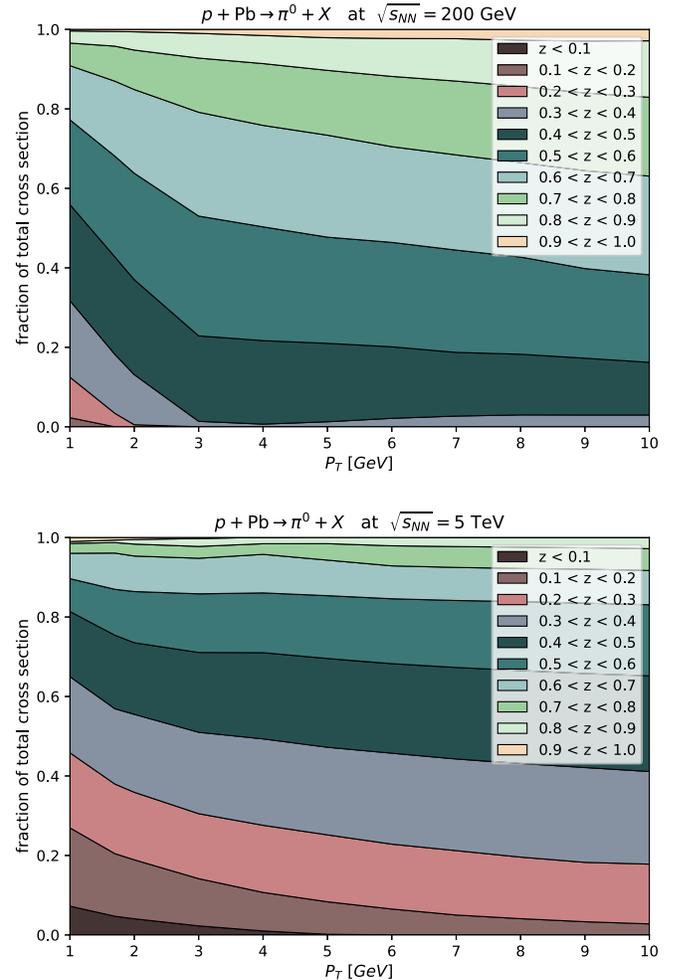


FIG. 7. Contribution of different z regions of the fragmentation functions, $D^{\pi^0}(z, Q)$, to the total $p + Pb \rightarrow \pi^0 + X$ cross section at $\sqrt{s_{NN}} = 200$ GeV (top) and 5 TeV (bottom).

III. IMPACT OF SIH DATA ON PDF FITS

A. Data selection

Before performing the fits, we need to decide which datasets to include and which kinematic cuts to impose. First, we choose not to include the eta meson data in the current analysis as we only have a single FF without known uncertainties, but we will examine this dataset in Sec. III G.

To make sure that we can accurately describe the proton baseline as presented in Fig. 4, we cut all data with $p_T < 3$ GeV. This is a more restrictive cut than in nCTEQ15(WZ) and EPPS16 [4], both of which used RHIC neutral pion data with p_T values down to 1.7 GeV. Our p_T cut is also sufficient to ensure that even at the highest $\sqrt{s_{NN}}$ of the ALICE data the fragmentation functions are used only within their well constrained region. This cut leaves us with 77 (out of 174) ALICE and 32 (out of 77) RHIC data points.

To account for the fragmentation function uncertainties, we take our error estimate from the DSS eigenvectors and add them in quadrature with the systematic uncertainties of the data.

B. Main PDF fits

We now use the single inclusive hadron production data to extend both the nCTEQ15 and nCTEQ15WZ fits. For comparison, we will produce two baseline fits. We produce one baseline with the BKK fragmentation functions as this was the set used in the previous nCTEQ15 and

nCTEQ15WZ analyses. We produce also a second baseline with the DSS fragmentation functions as these come from a more recent analysis and include uncertainties. We will then compare these fits with other available fragmentation functions in Sec. III D.

A short summary of the properties of the main fits are given in the following:

- (i) The included datasets are neutral pions (STAR, PHENIX, ALICE 5, and 8 TeV), charged pions (STAR and ALICE 5 TeV) and charged kaons (ALICE 5 TeV).
- (ii) Cuts are applied below $p_T = 3$ GeV for all datasets.
- (iii) Eta mesons are not included in the current fits; we examine these data later in Sec. III G.
- (iv) PHENIX charged hadrons are excluded by our p_T cut.
- (v) Normalizations of all SIH datasets are fitted according to the prescription given in Appendix A.
- (vi) Fits are performed first with data uncertainties alone, and again with uncertainties from the DSS fragmentation functions added as a systematic uncertainty to the data.
- (vii) Except for those items specified above, all other inputs to the fit are kept equal to the baseline fit.

The resulting fits are shown in Figs. 8 and 9 for the nCTEQ15 and nCTEQ15WZ baseline, respectively. The plots show the baseline fit in black, the fit with regular data uncertainties in red, and the fit with DSS uncertainties added to the data in green. We focus only on the lead PDF

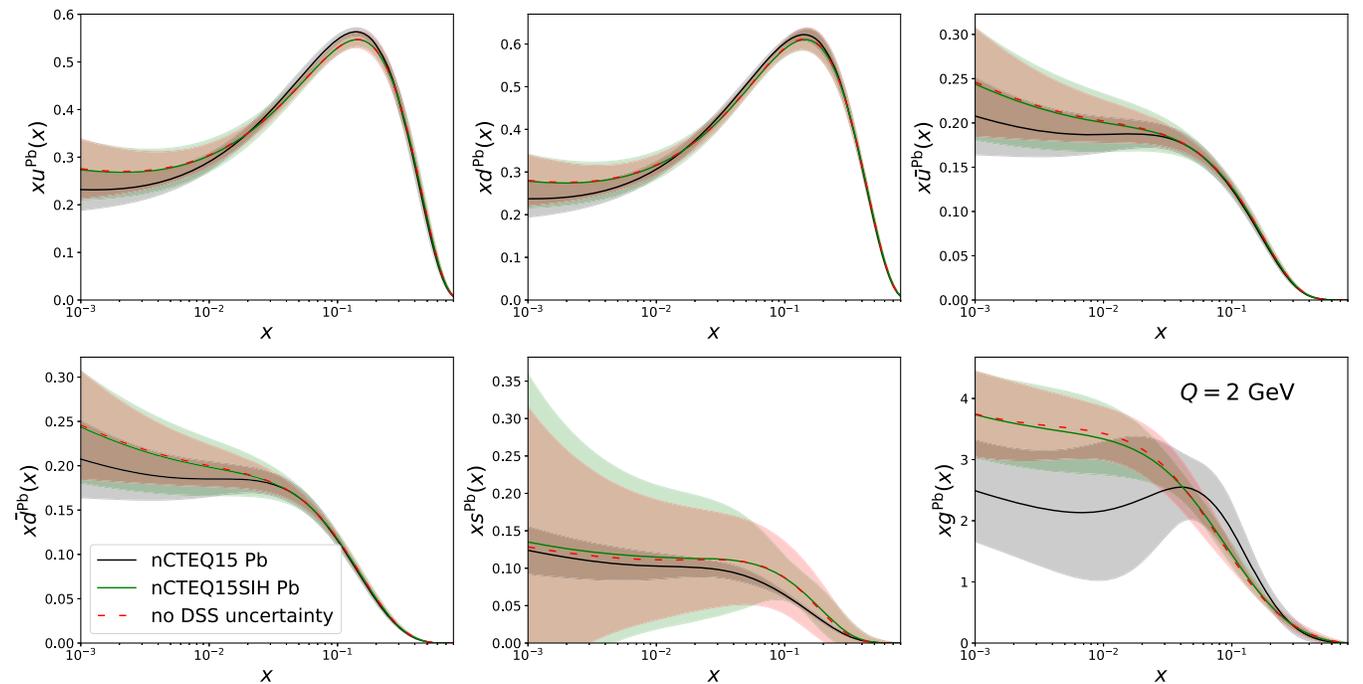


FIG. 8. Lead PDFs from fits to the nCTEQ15 data + SIH data. The baseline nCTEQ15 fit is shown in black, the fit with unmodified data is shown in red, and the fit where the uncertainties from the DSS FFs were added as a systematic uncertainty (nCTEQ15SIH) is shown in green.

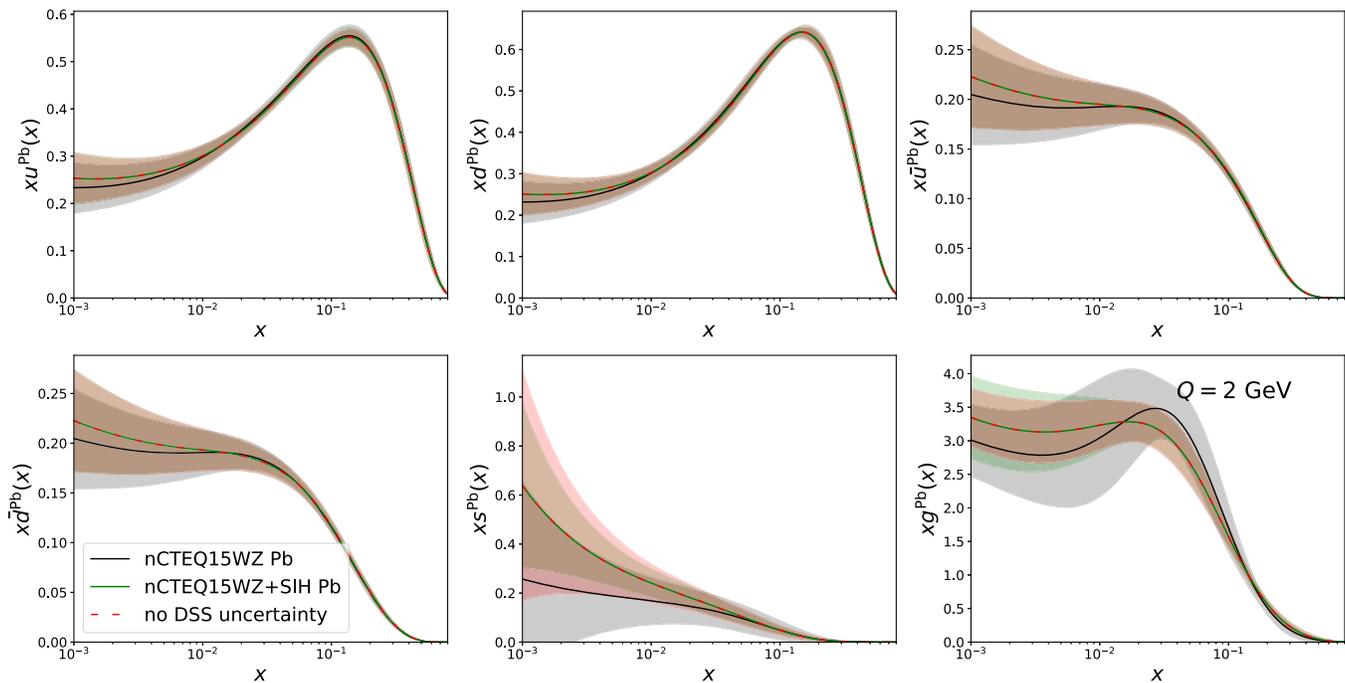


FIG. 9. Lead PDFs from fits to the nCTEQ15WZ data + SIH data. The baseline nCTEQ15WZ fit is shown in black, the fit with unmodified data is shown in red and the fit where the uncertainties from the DSS FFs were added as a systematic uncertainty (nCTEQ15WZ + SIH) is shown in green.

since the new data are taken on lead and gold, which are similarly heavy.

Examining Fig. 8, the most obvious change between the baseline (black) and the new fits (red, green) is the change in the gluon, which is enhanced at $x < 0.05$ and suppressed at $0.05 < x < 0.3$. The central values of the other flavors also exhibit some slight changes as they are of course coupled to the gluons via the DGLAP evolution. The uncertainties for up, down, and strange flavors are larger in the new fits than in the baseline due to the newly opened strange parameters. This is not unexpected and was also seen in the recent nCTEQ15WZ analysis where the same strange parameters were opened up. The inclusion of the DSS uncertainty does not cause any significant change in the central value but does result in an increased PDF error band, which is most noticeable at small x , especially for the strange PDF. Somewhat surprisingly, the region $x \sim 0.1$ sees a slight decrease in uncertainties, likely caused by slight shifts in the Hessian basis's eigenvector directions.

In Fig. 9, the same fits are shown with W/Z data included. Here, we see that the new fits for nCTEQ15WZ+SIH are generally (with the exception of the strange PDFs) more similar to the baseline fit than was the case for the nCTEQ15SIH fits shown in Fig. 8. For the gluon, we see similar behavior as in Fig. 8, but slightly less pronounced due to the additional constraints from the W/Z data. A somewhat surprising feature of this fit is the enhancement of the strange quark at low x . As Fig. 1 shows no particular strange sensitivity of the SIH data, presumably this enhanced strange PDF is being driven, in part, by the influence of the W/Z

data. While the resulting strange PDF in the new fits is substantially larger than the baseline at low x values, it is important to recall that the LHC heavy ion data primarily constrain the region $x \gtrsim 0.01$. Including the DSS uncertainty in this case causes no visible difference in central values but yields slightly larger uncertainty bands on the gluon. The shifted eigenvector basis results in slightly decreased strange quark uncertainties in the low- x region.

C. Quality of the fits

To judge the quality of the fits, we first take a detailed look at the resulting χ^2 values. Figure 10 shows the χ^2/N_{dof} for each of the fitted datasets of the two main fits, nCTEQ15SIH and nCTEQ15WZ + SIH. We see that the DIS and DY datasets are still well described by the new PDF and generally satisfy $\chi^2/N_{\text{dof}} < 1$, with one exception.⁴ The W/Z data also remain well described when including the SIH data, with the exception of dataset 6215 (ATLAS Run I, Z production); this behavior was also observed in the nCTEQ15WZ analysis.

More quantitative insights regarding the fit results can be obtained from Table IV, which shows a breakdown of the χ^2/N_{dof} by experiment type and by dataset. Note that there is a small difference for the STAR and PHENIX pion results reported here (with $p_T > 3$ GeV) and in the nCTEQ15WZ analysis, which used a $p_T > 1.7$ GeV cut.

⁴The notable exception with a large χ^2/N_{dof} is dataset 5108 (Sn/D EMC-1988) with eight data points. However, other analyses also found a large χ^2/N_{dof} for this dataset [4,81].

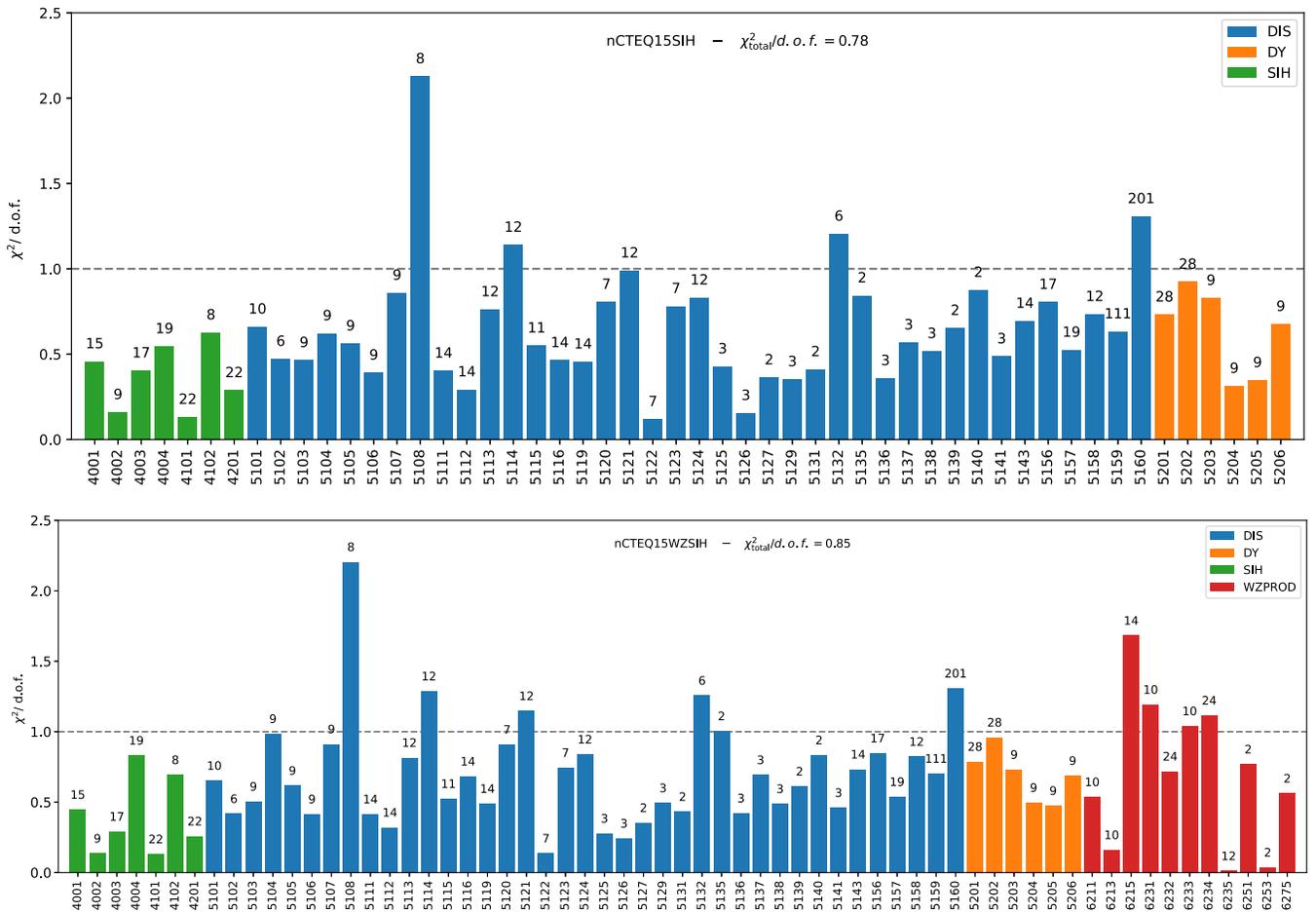


FIG. 10. The χ^2/N_{dof} of the individual experiments for the nCTEQ15SIH fit (top panel) and the nCTEQ15WZ + SIH fit (bottom panel). The number of data points is indicated at the top of each bar. The ID numbers for the SIH data are listed in Table I, and the other processes are listed in Ref. [39].

Beginning with the nCTEQ15 fit, we see that the DIS and DY data are well described. In contrast, the W/Z and SIH data (which were *not* fitted) yield large χ^2/N_{dof} values. Adding SIH data to the fit (nCTEQ15SIH) significantly improves the SIH data from $\chi^2/N_{\text{dof}} = 1.23$ to 0.38, as well as the W/Z data from $\chi^2/N_{\text{dof}} = 3.74$ to 2.32. There is also a slight improvement in the DY data, with a marginal increase in the DIS χ^2 .

In a similar manner, the nCTEQ15WZ fit yields good χ^2 values for the DIS, DY, and W/Z data, but the fit to the SIH data is not optimal. Including the SIH data in the fit, we find an improvement from $\chi^2/N_{\text{dof}} = 0.81$ to 0.41. This results in marginal shifts for the DIS and DY data but does increase the

W/Z data from $\chi^2/N_{\text{dof}} = 0.90$ to 1.02. However, the total χ^2/N_{dof} for the combined fit nCTEQ15WZ+SIH is $\chi^2/N_{\text{dof}} = 0.85$ as compared to the nCTEQ15WZ with 0.90.

Adding either inclusive hadron data or colorless weak boson production data also improves the fit for the other dataset when compared to the nCTEQ15 baseline. This supports the consistency of our interpretation of the nuclear modification in the inclusive hadron data as cold nuclear effects. A less ambiguous signal for medium effects would therefore certainly be correlations of several particles as observed, e.g., in the production of two hadrons [82,83].

Table III shows the fitted normalizations of the SIH datasets. All the resulting normalization parameters are

TABLE III. Normalization uncertainties and fitted normalizations of the SIH datasets in the nCTEQ15WZ + SIH fit.

	STAR		PHENIX	ALICE			
	π^0	π^\pm	π^0	5 TeV π^0	5 TeV π^\pm	5 TeV K^\pm	8 TeV π^0
Normalization uncertainty	17%	17%	10%	6%	6%	6%	3.4%
Fitted normalization	0.942	0.866	1.010	0.995	0.994	1.021	1.021

TABLE IV. We present the χ^2/N_{dof} for the individual SIH datasets; the individual processes DIS, DY, SIH, and WZ; and the total. The shown χ^2 is the sum of regular χ^2 and normalization penalty. Excluded processes are shown in parentheses. Note that both nCTEQ15 and nCTEQ15WZ included the neutral pions from STAR and PHENIX.

	χ^2/N_{dof} for selected experiments and processes											Total
	STAR		PHENIX	ALICE				DIS	DY	WZ	SIH	
	π^0	π^\pm	π^0	5 TeV π^0	5 TeV π^\pm	5 TeV K^\pm	8 TeV π^0					
nCTEQ15	0.13	2.68	0.30	2.53	0.62	0.71	1.96	0.86	0.78	(3.74)	(1.23)	1.28
nCTEQ15SIH	0.16	0.69	0.41	0.48	0.13	0.29	0.58	0.87	0.72	(2.32)	0.38	1.00
nCTEQ15WZ	0.17	3.24	0.23	0.67	0.21	0.41	1.58	0.90	0.78	0.90	(0.81)	0.90
nCTEQ15WZ + SIH	0.14	0.75	0.30	0.47	0.13	0.26	0.79	0.91	0.77	1.02	0.41	0.85

TABLE V. The χ^2/N_{dof} values of the SIH data obtained with different fragmentation functions and PDF parameters taken from the nCTEQ15WZ + SIH fit. We show the DSS result both with (modified data) and without (unmodified data) the added systematics arising from the fragmentation function uncertainties.

DSS unmodified data	DSS modified data	KKP	BKK	NNFF	JAM20
0.461	0.412	0.401	0.420	0.456	0.553

consistent with unity within the normalization uncertainty. Therefore, no significant normalization penalties are applied.

D. Comparisons with other FFs

To investigate the influence of the choice of fragmentation function on the quality of our fit to the SIH data, in Table V, we compute the χ^2/N_{dof} for the collection of fragmentation functions listed in Table II using the parameters from our nCTEQ15WZ + SIH fit. Additionally, we show the result using DSS both with and without the added systematics arising from the fragmentation function uncertainties. For these two DSS results, it is clear that including the additional uncertainties yields a lower χ^2 value. The results shown with the other fragmentation functions are computed using the modified data including fragmentation function uncertainties.

In Table V, we find the results from KKP and BKK are quite comparable to the DSS result (with modified data), and the NNFF is just slightly higher. Nevertheless, all these results are below the DSS result with unmodified data, and this suggests that the inclusion of the extra uncertainties taken from the DSS error bands provides our fit with a reasonable estimate of the impact of the fragmentation function choice. The JAM20 fragmentation functions yield a higher χ^2/N_{dof} than the others, and this reflects the observations noted in Sec. II B and Fig. 4, which displayed the comparisons with the $p + p \rightarrow \pi^0 + X$ data.

E. Comparison of data and theory

We now present a detailed comparison between our new fits and the SIH data in Fig. 11, which displays the nuclear ratios $R_{AA'}$ as a function of p_T . Although our fits imposed a

3 GeV p_T cut on the data, we extrapolate to lower p_T values in the shaded regions of the figure.

Examining the nCTEQ15 PDF curves in Fig. 11, we notice these have a significant positive slope for most of the datasets as compared to the other fits. This observation suggests that as we add more datasets the final predictions exhibit a reduced slope, and if we focus on the fitted region with $p_T > 3$ GeV, the curves approach unity within approximately 10%.

Comparing the nCTEQ15SIH fit and its baseline (nCTEQ15), we see a considerable shift in the $R_{AA'}$ nuclear ratio when the SIH data are included. In contrast, the nCTEQ15WZ + SIH fit and its baseline (nCTEQ15WZ) show only subtle differences, aside from the normalization, which is not surprising given that Table IV indicated that the W/Z data pull in the same direction as the SIH data. Finally, the nCTEQ15SIH and nCTEQ15WZ + SIH fits are quite comparable, certainly given the uncertainty of the data.

F. Correlation between data and PDFs

By looking at the PDFs alone, we cannot judge the impact of each individual new dataset on the fit. Therefore, we make use of two further methods to study how each dataset impacts the gluon specifically. The first quantity we want to analyze is the cosine of the correlation angle between two observables X and Y , as used in Refs. [12,84],

$$\cos(\phi[X, Y]) = \frac{\sum_i (X_i^{(+)} - X_i^{(-)}) (Y_i^{(+)} - Y_i^{(-)})}{\sqrt{\sum_{i'} (X_{i'}^{(+)} - X_{i'}^{(-)})^2} \sqrt{\sum_{i''} (Y_{i''}^{(+)} - Y_{i''}^{(-)})^2}}, \quad (7)$$

where the index of each sum runs over the 19 eigenvector directions.

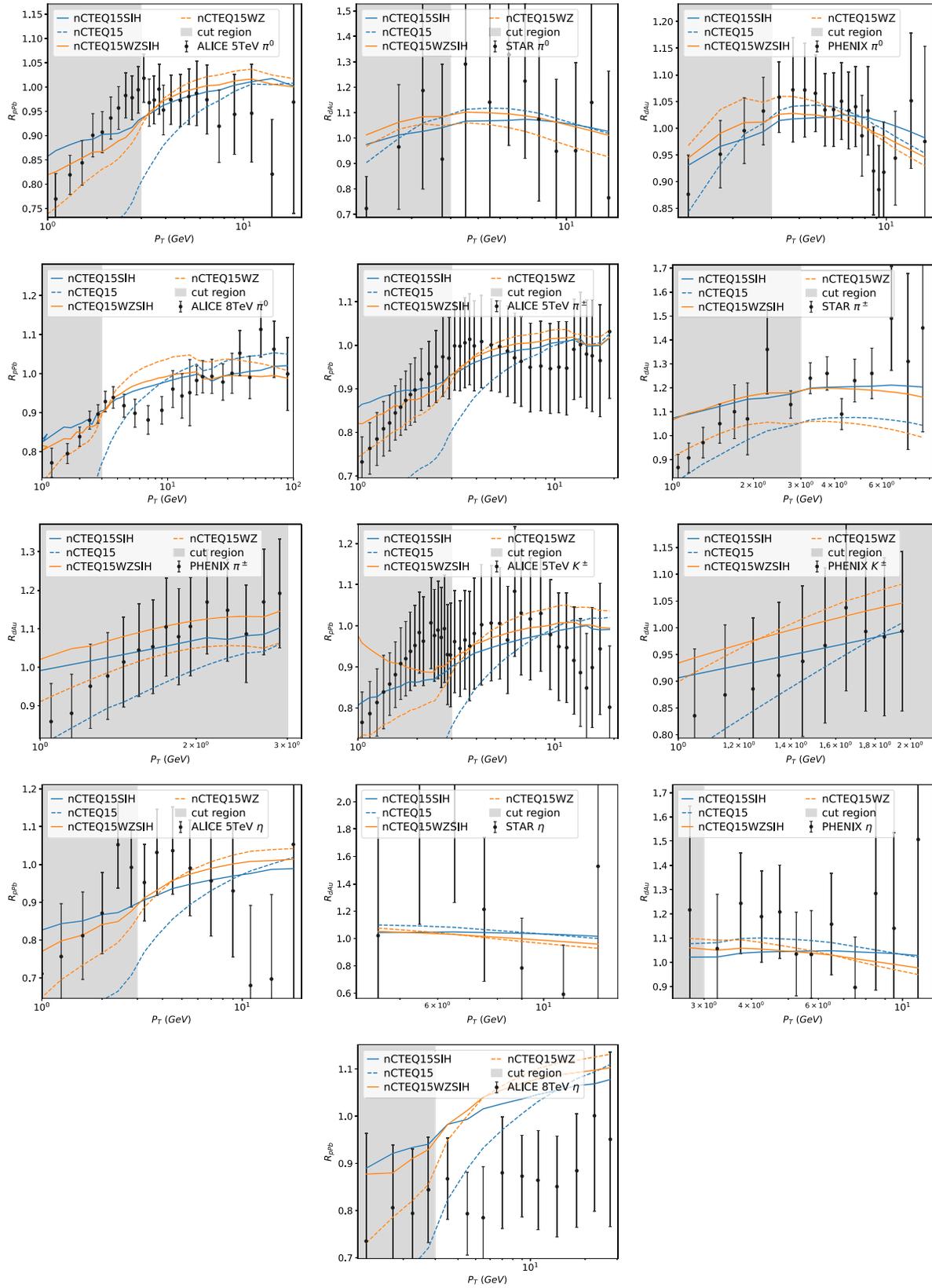


FIG. 11. Theory predictions for the main fits and their respective baselines. Dashed curves indicate the baseline fits, while solid curves show the fits with SIH data included. The blue curves are based on nCTEQ15, and the orange ones are based on nCTEQ15WZ.

Another useful quantity is the effective $\Delta\chi_{\text{eff}}^2$ as introduced in Ref. [3]. In contrast to the cosine of the correlation angle, $\Delta\chi_{\text{eff}}^2$ is more sensitive to the number of data points

$$\Delta\chi_{\text{eff}}^2[X, E_j] = \sum_i \frac{1}{2} \{ |\chi_i^{2(+)}(E_j) - \chi_i^{2(0)}(E_j)| + |\chi_i^{2(-)}(E_j) - \chi_i^{2(0)}(E_j)| \} \left(\frac{X_i^{(+)} - X_i^{(-)}}{\sqrt{\sum_{i'} (X_{i'}^{(+)} - X_{i'}^{(-)})^2}} \right)^2. \quad (8)$$

To investigate the impact of individual experimental datasets E_j on the gluon PDF $g(x, Q)$, we look at the cosine of the correlation angle $\cos(\phi[g(x, Q), \chi^2(E_j)])$ and the effective χ^2 difference $\Delta\chi_{\text{eff}}^2[g(x, Q), E_j]$. Since neither of these quantities displays a strong Q dependence, we show them only for the value of $Q = 10$ GeV in Figs. 12 and 13. We also limit ourselves to the gluon in lead, as the focus of the SIH data is on the heavy elements; the results for gold are similar to lead.

In Fig. 12, we see how the 5 TeV SIH ALICE datasets (π^0, π^\pm, K^\pm) display a strong anticorrelation ($\cos\phi \sim -0.9$) with the low x gluon ($x \sim 10^{-3}$) that is not seen in any of the remaining data, including the 8 TeV ALICE neutral pions. This observation suggests that the 5 TeV SIH ALICE data have a significant impact on the resulting gluon in the small x region. Interestingly, the correlation angle of the STAR and PHENIX neutral pion data are quite similar to each other, and in the region $x \sim 5 \times 10^{-2}$, they also exhibit a strong anticorrelation ($\cos\phi \sim -0.9$), which then becomes strong and positive ($\cos\phi \sim +0.8$) for larger x . The 8 TeV ALICE neutral pion data show a correlation behavior similar to the NMC96 SnC dataset (which is the dominant DIS set due to its large size and Q coverage), and somewhat

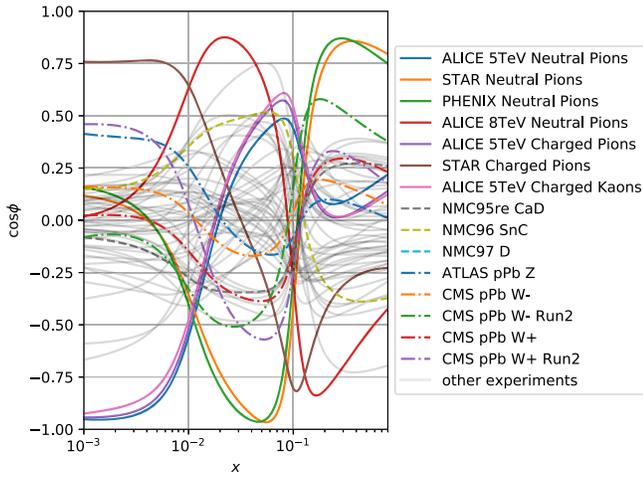


FIG. 12. Cosine of the correlation angle $\cos(\phi[g(x, Q), \chi^2(E_j)])$ between gluon PDF and χ^2 of each experimental dataset (E_j) for the nCTEQ15WZ + SIH fit at $Q = 10$ GeV.

and error size of the experiments because the normalization does not cancel these factors out. For an experiment E_j and an observable X , it is defined as

opposite to the STAR and PHENIX neutral pion data. Examining the larger x region ($x > 0.1$), the influence of the various datasets is more mixed with the STAR and PHENIX π^0 data yielding a large positive correlation and ATLAS 8 TeV π^0 and STAR π^\pm yielding a large negative correlation, with the result that the high- x gluon remains mostly unchanged in Fig. 9.

Turning to the χ_{eff}^2 in Fig. 13, we can see that the CMS Run II W^\pm and NMC96 SnC data remain the main forces determining the gluon, with the ALICE neutral pion and NMC95re CaD datasets also providing constraints.

Among the SIH datasets, the 8 TeV neutral pion data have the largest χ_{eff}^2 , followed by the 5 TeV neutral pion data. However, they generally do not reach values as high as the previously mentioned DIS and W and Z Boson production data. It is unfortunate that we must impose the $p_T > 3$ GeV cut on the SIH data due to limitations of our perturbative theoretical calculations; this removes a large amount of precision SIH data from our analysis. Improved theoretical techniques such as resummation may allow us to extend our analysis to smaller p_T values in the future so that a larger amount of the SIH data can be included in the PDF determination.

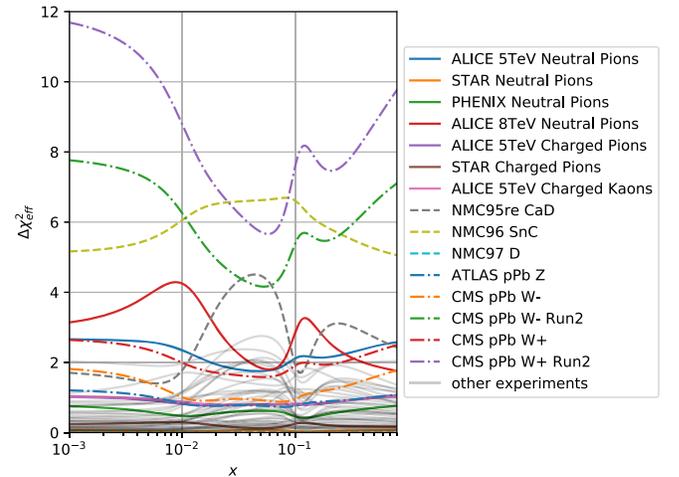


FIG. 13. $\Delta\chi_{\text{eff}}^2[g(x, Q), E_j]$ for the gluon PDF with each experimental dataset (E_j) for the nCTEQ15WZ + SIH fit at $Q = 10$ GeV.

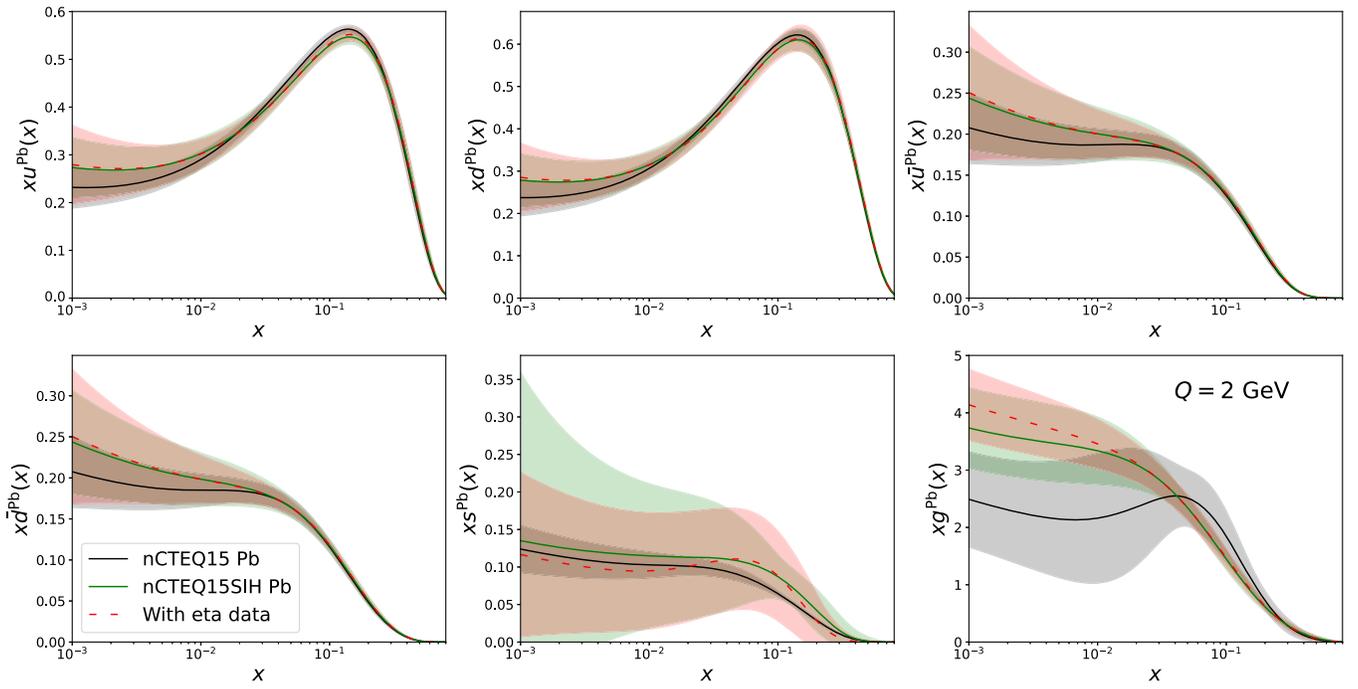


FIG. 14. Lead PDFs from fits to the nCTEQ15 data including the SIH eta data. The baseline nCTEQ15 fit is shown in black, the fit with eta meson data is shown in red, and the corresponding main fit is shown in green. The nCTEQ15 (black) and nCTEQ15SIH (green) are also displayed in Fig. 8.

G. Impact of the eta meson

We now investigate the impact of including the eta meson data. We will compute our fit with the DSS

fragmentation functions for the pions and kaons and use the AESSS fragmentation functions for the eta mesons. The AESSS FFs do not provide any uncertainties, so we will not

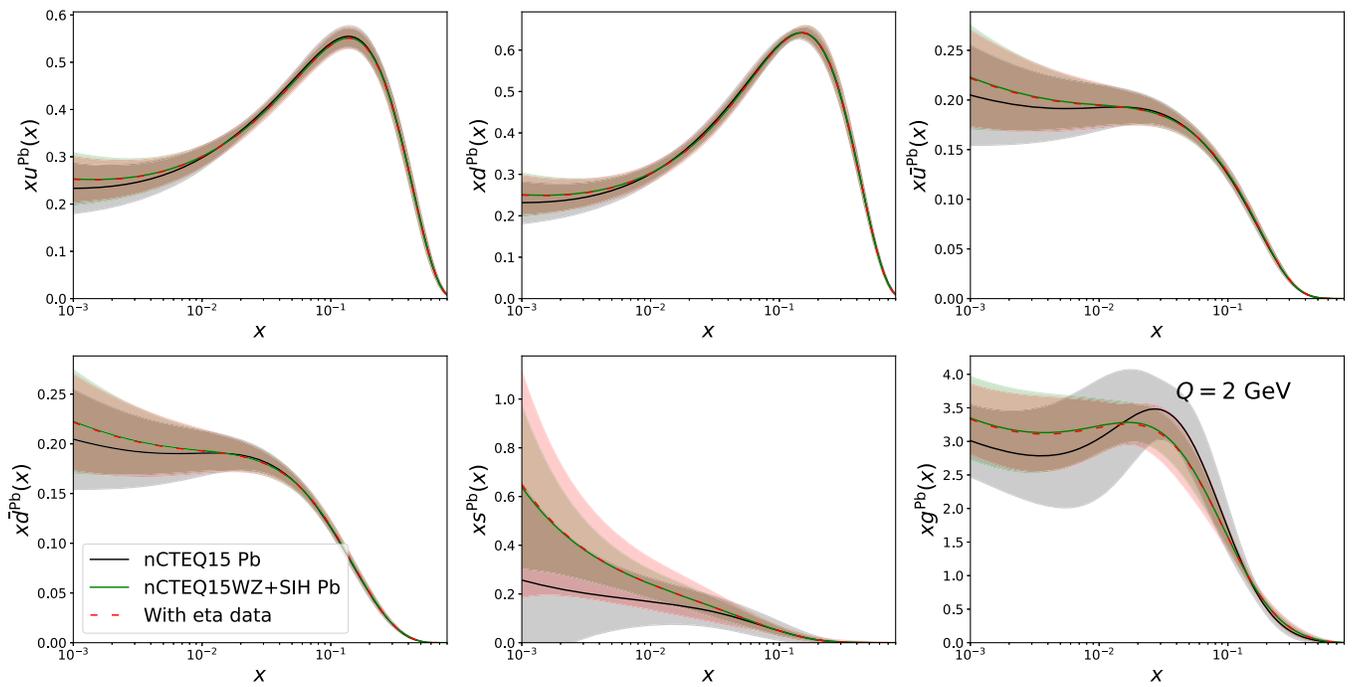


FIG. 15. Lead PDFs from fits to the nCTEQ15WZ data including the SIH eta data. The baseline nCTEQ15WZ fit is shown in black, the fit with eta meson data is shown in red, and the corresponding main fit is shown in green. The nCTEQ15 (black) and nCTEQ15WZ + SIH (green) are also displayed in Fig. 9.

include any for the eta meson data. Using the same $p_T \geq 3$ cut as before for the pions and kaons, the eta meson data now provide an additional 18 data points from RHIC and 19 from ALICE.

The fits including the eta meson data are shown in Figs. 14 and 15, and are compared with the baseline fit (nCTEQ15WZ) in black, the corresponding main fit (nCTEQ15SIH) in green, and the fit with eta in red (nCTEQ15SIH+eta).

Examining the results of Figs. 14, the impact of the eta data yields a slight upward shift of the gluon in the low- x region and a downward modification of the strange quark both at small x and larger $x \sim 0.2$. The uncertainty of the gluon shrinks by a small amount, the strange uncertainty is reduced in the low x region, while it increases at medium x , and the other flavors show a slight increase at very small x .

Examining the results of the second case based on the nCTEQ15WZ fit as shown in Figs. 15, we observe that the central values of the nCTEQ15WZ + SIH fit with and without the eta meson data are virtually identical. Regarding the PDF uncertainties, the error bands of the gluon and down quark are reduced only by a negligible margin, while the strange quark uncertainty grows very slightly. Since the uncertainty of the eta fragmentation function is expected to be larger than that of pions and kaons, the net effect of including the eta data into the fit would most likely be inconsequential if these additional uncertainties were included in the analysis.

IV. CONCLUSION

Using the nCTEQ++ framework, we incorporated new data on single inclusive hadron production from ALICE and RHIC into our PDF analysis. We investigated the choice of scales and fragmentation functions (including their uncertainties) and identified a p_T region where reliable perturbative predictions can be applied to help constrain the PDFs.

We obtained a good χ^2/N_{dof} for all the datasets and found that the new SIH meson data had a noticeable impact on the gluon PDF at low to medium x . Compared to the nCTEQ15WZ PDF set, the gluon flattens out in the region around $x = 0.05$, and the uncertainties in this region shrink.

The necessary $p_T > 3$ GeV cut limits our ability to constrain the PDFs in the low- x region. If it were possible to expand the theoretical predictions to lower p_T values with improved calculation techniques, then we could use the very precise ALICE data in this region to further improve the determination of the gluon PDF.

Nevertheless, even with the current limitations on the kinematics, the SIH data provide useful constraints on the nuclear gluon distribution, which is still one of the least constrained nPDFs. As such, we believe the presented analysis and the obtained PDFs provide an important step on the way to more precise knowledge of the nuclear

structure. Also on the practical level, the reduced gluon uncertainties are important for many applications. The PDFs of the nCTEQ15WZ + SIH fit for a selection of nuclei will be provided through the LHAPDF website, and those of other fits can be obtained upon request.

ACKNOWLEDGMENTS

The work of P. D., M. K., and K. K. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 273811115 SFB 1225. L. A. H., M. K., K. K. and K. F. M. also acknowledge support of the DFG through the Research Training Group GRK 2149. A. K. acknowledges the support of Narodowe Centrum Nauki under Grant No. 2019/34/E/ST2/00186. The work of T. J. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 396021762—TRR 257. F. O. acknowledges support through U.S. Department of Energy Grant No. DE-SC0010129 and the National Science Foundation under Grant No. NSF PHY-1748958. The work of I. S. was supported by the French CNRS via the IN2P3 Project No. GLUE@NLO.

APPENDIX A: FITTING DATA NORMALIZATIONS

We use the χ^2 prescription given in Ref. [58] to fit the normalizations of the SIH and W/Z production data. For a dataset D with N data points and S correlated systematic errors, the χ^2 of the dataset is given by

$$\chi_D^2 = \sum_{i,j} \left(D_i - \frac{T_i}{N_{\text{norm}}} \right) (C^{-1})_{ij} \left(D_j - \frac{T_j}{N_{\text{norm}}} \right) + \left(\frac{1 - N_{\text{norm}}}{\sigma_{\text{norm}}} \right)^2, \quad (\text{A1})$$

where σ_{norm} is the normalization uncertainty and T_i is the theoretical prediction for point i . The last term of Eq. (A1) is called the normalization penalty, and it vanishes when the fitted normalization is equal to unity. The penalty is scaled by the normalization uncertainty σ_{norm} , which is around 0.03 for W/Z production and ranges from 0.034 to 0.17 for SIH production. The covariance matrix C_{ij} is defined as

$$C_{ij} = \sigma_i^2 \delta_{ij} + \sum_{\alpha} \bar{\sigma}_{i\alpha} \bar{\sigma}_{j\alpha}, \quad (\text{A2})$$

where σ_i is the total uncorrelated uncertainty (added in quadrature) for data point i and $\bar{\sigma}_{i\alpha}$ is the correlated systematic uncertainty for data point i from source α . We use the analytical formula for the inverse of the correlation matrix found in Ref. [85] to obtain

$$\chi_D^2 = \sum_i \left(\frac{D_i - T_i/N_{\text{norm}}}{\sigma_i} \right)^2 - B^T A^{-1} B + \left(\frac{1 - N_{\text{norm}}}{\sigma_{\text{norm}}} \right)^2 \quad (\text{A3})$$

with

$$A_{\alpha\gamma} = \delta_{\alpha\gamma} + \sum_i \frac{\bar{\sigma}_{i\alpha} \bar{\sigma}_{i\gamma}}{\sigma_i^2} \quad (\text{A4})$$

and

$$B_\alpha = \sum_i \frac{\bar{\sigma}_{i\alpha} (D_i - T_i/N_{\text{norm}})}{\sigma_i^2}. \quad (\text{A5})$$

APPENDIX B: UNCERTAINTIES OF OTHER FFs

We compare the data with our theoretical predictions with nCTEQ15WZ PDFs using the uncertainties taken from the NNFF and JAM20 fragmentation functions in Figs. 16 and 17, respectively.

The NNFF fragmentation functions yield slightly larger uncertainties than those of DSS shown in Fig. 6. This may be due, in part, to the use of a parametrization-free neural network instead of a “traditional” parametrization and a slightly smaller dataset. The uncertainties of the JAM20 fragmentation are so small across the kinematic region with $p_T > 1$ GeV that they can be neglected when compared with the data uncertainty.

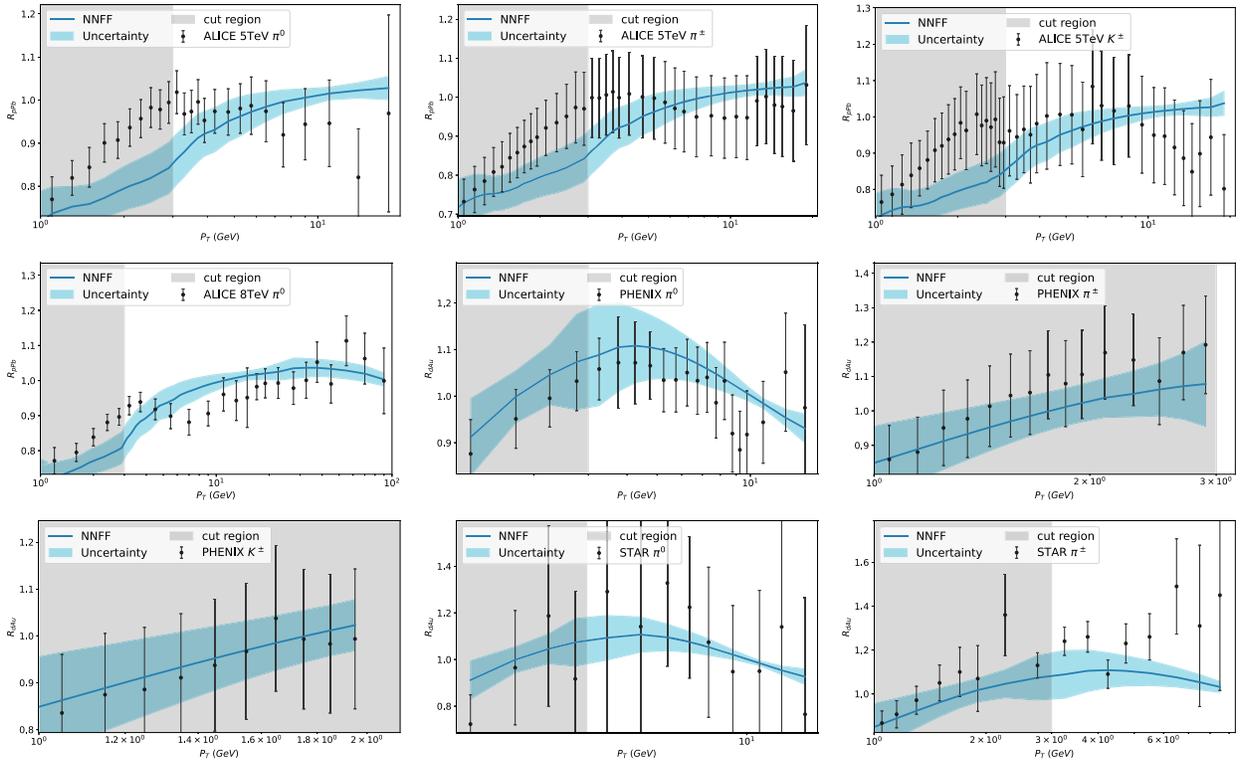


FIG. 16. Uncertainties calculated from the NNFF replicas using nCTEQ15WZ PDFs. The computed uncertainties use 83 of the 101 provided replicas because the remaining 18 lead to unphysical behavior such as large jumps from one p_T value to another or negative cross sections due to numerical problems.

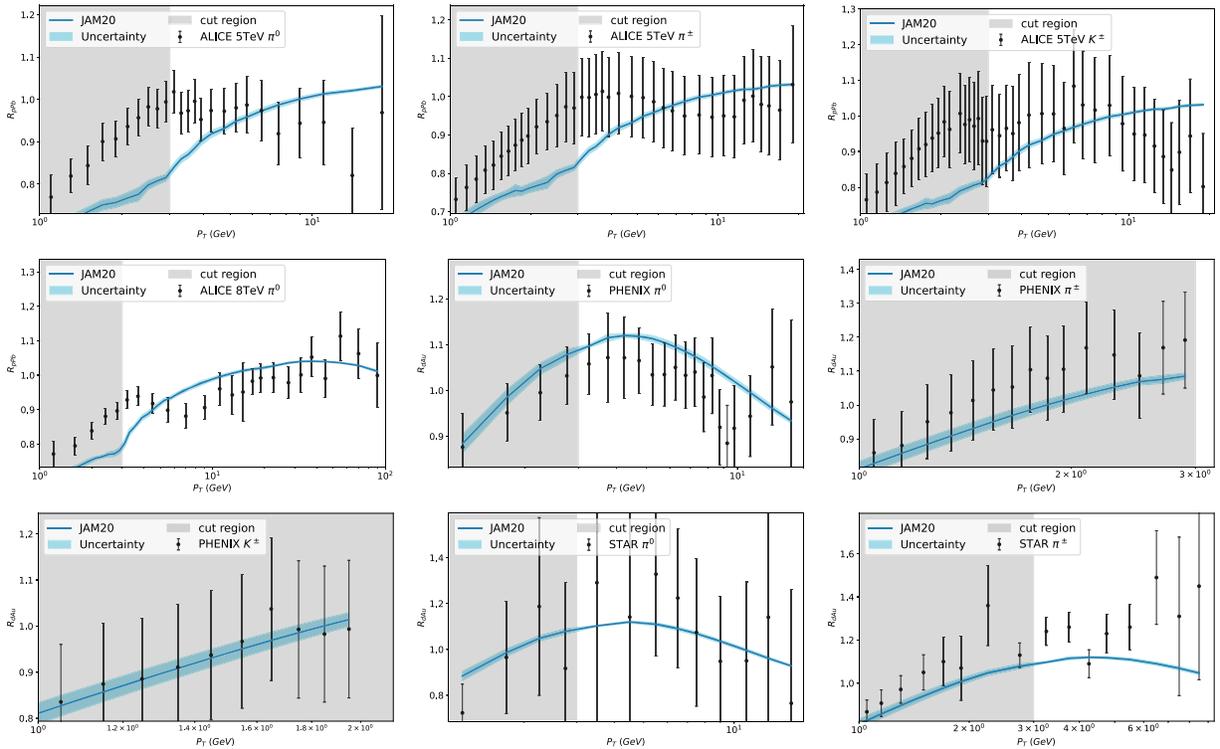


FIG. 17. Uncertainties calculated from the JAM20 replicas using nCTEQ15WZ PDFs.

- [1] T.-J. Hou *et al.*, New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC, *Phys. Rev. D* **103**, 014013 (2021).
- [2] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions from high-precision collider data, *Eur. Phys. J. C* **77**, 663 (2017).
- [3] K. Kovarik *et al.*, nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework, *Phys. Rev. D* **93**, 085037 (2016).
- [4] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, EPPS16: Nuclear parton distributions with LHC data, *Eur. Phys. J. C* **77**, 163 (2017).
- [5] R. Abdul Khalek, J. J. Ethier, and J. Rojo (NNPDF Collaboration), Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider, *Eur. Phys. J. C* **79**, 471 (2019).
- [6] R. Abdul Khalek, J. J. Ethier, J. Rojo, and G. van Weelden, nNNPDF2.0: Quark flavor separation in nuclei from LHC data, *J. High Energy Phys.* **09** (2020) 183.
- [7] J. J. Ethier and E. R. Nocera, Parton distributions in nucleons and nuclei, *Annu. Rev. Nucl. Part. Sci.* **70**, 43 (2020).
- [8] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, Towards ultimate parton distributions at the high-luminosity LHC, *Eur. Phys. J. C* **78**, 962 (2018).
- [9] J. Gao, L. Harland-Lang, and J. Rojo, The structure of the proton in the LHC precision era, *Phys. Rep.* **742**, 1 (2018).
- [10] K. Kovarik, P. M. Nadolsky, and D. E. Soper, Hadron structure in high-energy collisions, *Rev. Mod. Phys.* **92**, 045003 (2020).
- [11] S. Alekhin, J. Blmlein, and S. Moch, Strange sea determination from collider data, *Phys. Lett. B* **777**, 134 (2018).
- [12] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan, Implications of CTEQ global analysis for collider observables, *Phys. Rev. D* **78**, 013004 (2008).
- [13] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk (JAM Collaboration), Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions, *Phys. Rev. D* **101**, 074020 (2020).
- [14] 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).
- [15] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. Martin, and R. Nathvani, Updates of PDFs using the MMHT framework, *Proc. Sci.*, DIS2019 (2019) 036 [arXiv:1907.08147].
- [16] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali (NNPDF

- Collaboration), Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering, *Nucl. Phys.* **B823**, 195 (2009).
- [17] H.-W. Lin *et al.*, Parton distributions and lattice QCD calculations: A community white paper, *Prog. Part. Nucl. Phys.* **100**, 107 (2018).
- [18] H.-W. Lin *et al.*, Parton distributions and lattice QCD calculations: Toward 3D structure, *Prog. Part. Nucl. Phys.* **121**, 103908 (2021).
- [19] V. Guzey and M. Klasen, Constraints on nuclear parton distributions from dijet photoproduction at the LHC, *Eur. Phys. J. C* **79**, 396 (2019).
- [20] M. Klasen, K. Kovarik, and J. Potthoff, Nuclear parton density functions from jet production in DIS at an EIC, *Phys. Rev. D* **95**, 094013 (2017).
- [21] M. Klasen and K. Kovarik, Nuclear parton density functions from dijet photoproduction at the EIC, *Phys. Rev. D* **97**, 114013 (2018).
- [22] N. Armesto, H. Paukkunen, J. M. Penín, C. A. Salgado, and P. Zurita, An analysis of the impact of LHC Run I proton–lead data on nuclear parton densities, *Eur. Phys. J. C* **76**, 218 (2016).
- [23] N. Armesto, Nuclear shadowing, *J. Phys. G* **32**, R367 (2006).
- [24] L. Frankfurt, V. Guzey, and M. Strikman, Leading twist nuclear shadowing phenomena in hard processes with nuclei, *Phys. Rep.* **512**, 255 (2012).
- [25] B. Z. Kopeliovich, J. G. Morfin, and I. Schmidt, Nuclear shadowing in electro-weak interactions, *Prog. Part. Nucl. Phys.* **68**, 314 (2013).
- [26] S. A. Kulagin and R. Petti, Global study of nuclear structure functions, *Nucl. Phys.* **A765**, 126 (2006).
- [27] S. J. Brodsky and H. J. Lu, Shadowing and Antishadowing of Nuclear Structure Functions, *Phys. Rev. Lett.* **64**, 1342 (1990).
- [28] S. J. Brodsky, I. Schmidt, and J.-J. Yang, Nuclear antishadowing in neutrino deep inelastic scattering, *Phys. Rev. D* **70**, 116003 (2004).
- [29] D. F. Geesaman, K. Saito, and A. W. Thomas, The nuclear EMC effect, *Annu. Rev. Nucl. Part. Sci.* **45**, 337 (1995).
- [30] P. R. Norton, The EMC effect, *Rep. Prog. Phys.* **66**, 1253 (2003).
- [31] O. Hen, D. W. Higinbotham, G. A. Miller, E. Piasetzky, and L. B. Weinstein, The EMC effect and high momentum nucleons in nuclei, *Int. J. Mod. Phys. E* **22**, 1330017 (2013).
- [32] 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).
- [33] O. Hen, G. A. Miller, E. Piasetzky, and L. B. Weinstein, Nucleon-nucleon correlations, short-lived excitations, and the quarks within, *Rev. Mod. Phys.* **89**, 045002 (2017).
- [34] E. Iancu, A. Leonidov, and L. D. McLerran, Nonlinear gluon evolution in the color glass condensate. 1., *Nucl. Phys.* **A692**, 583 (2001).
- [35] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, The color glass condensate, *Annu. Rev. Nucl. Part. Sci.* **60**, 463 (2010).
- [36] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, Decoding the phase structure of QCD via particle production at high energy, *Nature (London)* **561**, 321 (2018).
- [37] J. Adam *et al.* (ALICE Collaboration), Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions, *Nat. Phys.* **13**, 535 (2017).
- [38] A. Ortiz (ALICE, ATLAS, CMS, LHCb Collaborations), Particle production and flow-like effects in small systems, *Proc. Sci., LHCP2019* (2019) 091 [arXiv:1909.03937].
- [39] A. Kusina *et al.*, Impact of LHC vector boson production in heavy ion collisions on strange PDFs, *Eur. Phys. J. C* **80**, 968 (2020).
- [40] D. de Florian, R. Sassot, M. Epele, R. J. Hernández-Pinto, and M. Stratmann, Parton-to-pion fragmentation reloaded, *Phys. Rev. D* **91**, 014035 (2015).
- [41] F. Olness, J. Pumplin, D. Stump, J. Huston, P. M. Nadolsky, H. L. Lai, S. Kretzer, J. F. Owens, and W. K. Tung, Neutrino dimuon production and the strangeness asymmetry of the nucleon, *Eur. Phys. J. C* **40**, 145 (2005).
- [42] E. P. Segarra *et al.*, nCTEQ15HIX – Extending nPDF analyses into the high- x , low Q^2 region, *Phys. Rev. D* **103**, 114015 (2021).
- [43] G. P. Salam and J. Rojo, A higher order perturbative parton evolution toolkit (HOPPET), *Comput. Phys. Commun.* **180**, 120 (2009).
- [44] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID project, *Eur. Phys. J. C* **66**, 503 (2010).
- [45] M. Werlen, INCNLO-direct photon and inclusive hadron production code website, Version 1.4. http://laph.cnrs.fr/PHOX_FAMILY.
- [46] J. C. Collins, D. E. Soper, and G. F. Sterman, Factorization of hard processes in QCD, *Adv. Ser. Dir. High Energy Phys.* **5**, 1 (1989).
- [47] S. Albino, The hadronization of partons, *Rev. Mod. Phys.* **82**, 2489 (2010).
- [48] I. Schienbein *et al.*, A review of target mass corrections, *J. Phys. G* **35**, 053101 (2008).
- [49] I. Schienbein *et al.* (nCTEQ Collaboration), Target mass corrections in lepton-nucleus DIS revisited, Report No. SMU-HEP-21-01 (to be published).
- [50] P. Aurenche, M. Fontannaz, J. P. 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).
- [51] S. S. Adler *et al.* (PHENIX Collaboration), Centrality Dependence of π^0 and η Production at Large Transverse Momentum in $\sqrt{s_{NN}} = 200$ GeV d + Au Collisions, *Phys. Rev. Lett.* **98**, 172302 (2007).
- [52] A. Adare *et al.* (PHENIX Collaboration), Spectra and ratios of identified particles in Au + Au and d + Au collisions at $\sqrt{s_{NN}} = 200$ GeV, *Phys. Rev. C* **88**, 024906 (2013).
- [53] B. I. Abelev *et al.* (STAR Collaboration), Inclusive π^0 , η , and direct photon production at high transverse momentum in p + p and d + Au collisions at $\sqrt{s_{NN}} = 200$ GeV, *Phys. Rev. C* **81**, 064904 (2010).
- [54] J. Adams *et al.* (STAR Collaboration), Identified hadron spectra at large transverse momentum in p + p and d + Au

- collisions at $\sqrt{s_{NN}} = 200$ GeV, *Phys. Lett. B* **637**, 161 (2006).
- [55] S. Acharya *et al.* (ALICE Collaboration), Neutral pion and η meson production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Eur. Phys. J. C* **78**, 624 (2018).
- [56] J. Adam *et al.* (ALICE Collaboration), Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Phys. Lett. B* **760**, 720 (2016).
- [57] S. Acharya *et al.* (ALICE Collaboration), Nuclear modification factor of light neutral-meson spectra up to high transverse momentum in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV, [arXiv:2104.03116](https://arxiv.org/abs/2104.03116).
- [58] G. D'Agostini, On the use of the covariance matrix to fit correlated data, *Nucl. Instrum. Methods Phys. Res., Sect. A* **346**, 306 (1994).
- [59] J. Binnewies, B. A. Kniehl, and G. Kramer, Next-to-leading order fragmentation functions for pions and kaons, *Z. Phys. C* **65**, 471 (1995).
- [60] 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).
- [61] S. Kretzer, Fragmentation functions from flavor inclusive and flavor tagged e^+e^- annihilations, *Phys. Rev. D* **62**, 054001 (2000).
- [62] M. Hirai, S. Kumano, T. H. Nagai, and K. Sudoh, Determination of fragmentation functions and their uncertainties, *Phys. Rev. D* **75**, 094009 (2007).
- [63] S. Albino, B. A. Kniehl, and G. Kramer, AKK update: Improvements from new theoretical input and experimental data, *Nucl. Phys.* **B803**, 42 (2008).
- [64] V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo (NNPDF Collaboration), A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties, *Eur. Phys. J. C* **77**, 516 (2017).
- [65] E. Moffat, W. Melnitchouk, T. Rogers, and N. Sato, Simultaneous Monte Carlo analysis of parton densities and fragmentation functions, *Phys. Rev. D* **104**, 016015 (2021).
- [66] D. de Florian, M. Epele, R. J. Hernandez-Pinto, R. Sassot, and M. Stratmann, Parton-to-kaon fragmentation revisited, *Phys. Rev. D* **95**, 094019 (2017).
- [67] C. A. Aidala, F. Ellinghaus, R. Sassot, J. P. Seele, and M. Stratmann, Global analysis of fragmentation functions for eta mesons, *Phys. Rev. D* **83**, 034002 (2011).
- [68] 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).
- [69] A. Metz and A. Vossen, Parton fragmentation functions, *Prog. Part. Nucl. Phys.* **91**, 136 (2016).
- [70] M. Hirai, H. Kawamura, S. Kumano, and K. Saito, Impacts of B-factory measurements on determination of fragmentation functions from electron-positron annihilation data, *Prog. Theor. Exp. Phys.* (2016), 113B04.
- [71] M. Soleymaninia, M. Goharipour, and H. Khanpour, First QCD analysis of charged hadron fragmentation functions and their uncertainties at next-to-next-to-leading order, *Phys. Rev. D* **98**, 074002 (2018).
- [72] V. Bertone, N. P. Hartland, E. R. Nocera, J. Rojo, and L. Rottoli (NNPDF Collaboration), Charged hadron fragmentation functions from collider data, *Eur. Phys. J. C* **78**, 651 (2018).
- [73] R. Sassot, M. Stratmann, and P. Zurita, Fragmentation functions in nuclear media, *Phys. Rev. D* **81**, 054001 (2010).
- [74] M. D. Sievert, I. Vitev, and B. Yoon, A complete set of in-medium splitting functions to any order in opacity, *Phys. Lett. B* **795**, 502 (2019).
- [75] S. S. Adler *et al.* (PHENIX Collaboration), Mid-Rapidity Neutral Pion Production in Proton Proton Collisions at $\sqrt{s} = 200$ -GeV, *Phys. Rev. Lett.* **91**, 241803 (2003).
- [76] B. Abelev *et al.* (ALICE Collaboration), Neutral pion and η meson production in proton-proton collisions at $\sqrt{s} = 0.9$ TeV and $\sqrt{s} = 7$ TeV, *Phys. Lett. B* **717**, 162 (2012).
- [77] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Parton distributions for the LHC, *Eur. Phys. J. C* **63**, 189 (2009).
- [78] G. Aad *et al.* (ATLAS Collaboration), Transverse momentum, rapidity, and centrality dependence of inclusive charged-particle production in $\sqrt{s_{NN}} = 5.02$ TeV $p + \text{Pb}$ collisions measured by the ATLAS experiment, *Phys. Lett. B* **763**, 313 (2016).
- [79] V. Khachatryan *et al.* (CMS Collaboration), Charged-particle nuclear modification factors in PbPb and pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *J. High Energy Phys.* **04** (2017) 039.
- [80] S. Acharya *et al.* (ALICE Collaboration), Transverse momentum spectra and nuclear modification factors of charged particles in pp, p-Pb and Pb-Pb collisions at the LHC, *J. High Energy Phys.* **11** (2018) 013.
- [81] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, Global analysis of nuclear parton distributions, *Phys. Rev. D* **85**, 074028 (2012).
- [82] S. Acharya *et al.* (ALICE Collaboration), Jet fragmentation transverse momentum measurements from di-hadron correlations in $\sqrt{s} = 7$ TeV pp and $\sqrt{s_{NN}} = 5.02$ TeV p-Pb collisions, *J. High Energy Phys.* **03** (2019) 169.
- [83] S. Acharya *et al.* (ALICE Collaboration), One-dimensional charged kaon femtoscopy in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Phys. Rev. C* **100**, 024002 (2019).
- [84] 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).
- [85] 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).