










# Impact of inclusive hadron production data on nuclear gluon PDFs

P. Duwentäster <sup>1,\*</sup> L. A. Husová,<sup>2</sup> T. Ježo <sup>3</sup> M. Klasen <sup>1</sup> K. Kovařík <sup>1</sup>  
A. Kusina <sup>4</sup> K. F. Muzakka <sup>1</sup> F. I. Olness <sup>5</sup> I. Schienbein <sup>6</sup> and J. Y. Yu <sup>5</sup>

<sup>1</sup>*Institut für Theoretische Physik, Westfälische Wilhelms-Universität Münster,  
Wilhelm-Klemm-Straße 9, D-48149 Münster, Germany*

<sup>2</sup>*Institut für Kernphysik, Westfälische Wilhelms-Universität Münster,  
Wilhelm-Klemm-Straße 9, D-48149 Münster, Germany*

<sup>3</sup>*Institute for Theoretical Physics, KIT, D-76131 Karlsruhe, Germany*

<sup>4</sup>*Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 Krakow, Poland*

<sup>5</sup>*Southern Methodist University, Dallas, TX 75275, USA*

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

(Dated: May 21, 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 Large Hadron Collider (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 data sets, 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 pre-computed grids. The extension of the nCTEQ15 analysis to include the SIH data represents an important step toward the next generation of PDFs.

## CONTENTS

I. Introduction	2	C. Quality of the fits	12
A. The gluon PDF	2	D. Comparisons with other FFs	12
B. The nCTEQ++ framework	3	E. Comparison of data and theory	12
C. SIH data sets	4	F. Correlation between data and PDFs	13
II. Fragmentation functions	4	G. Impact of the eta meson	16
A. Available fragmentation functions	4	IV. Conclusion	16
B. Comparison in proton-proton collisions	5	Acknowledgements	18
C. Scale uncertainties	6	A. Fitting data normalizations	18
D. Comparison in dAu and pPb collisions	7	B. Uncertainties of other FFs	18
E. Uncertainties of fragmentation functions	9	References	20
F. Fragmentation kinematics	9		
III. Impact of SIH data on PDF fits	10		
A. Data selection	10		
B. Main PDF fits	10		

\* pit.duw@uni-muenster.de

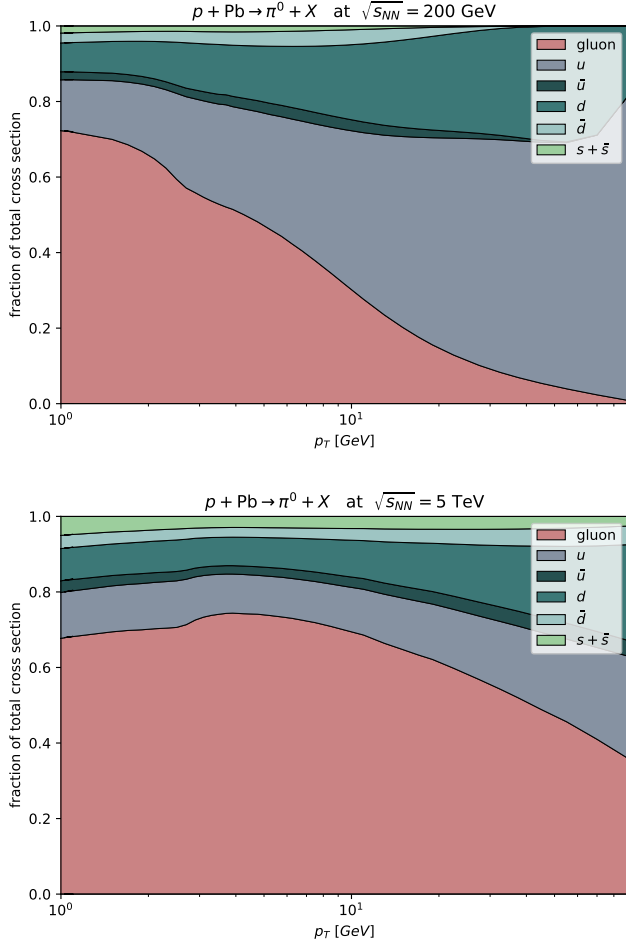


FIG. 1. Fractional contributions of the total  $p+\text{Pb} \rightarrow \pi^0 + X$  cross section initiated by each PDF flavor  $f_i^{\text{Pb}}(x, Q)$  of the lead nucleus at  $\sqrt{s_{NN}} = 200$  GeV (upper panel) and 5 TeV (lower panel) for  $i \in \{g, u, d, \bar{u}, \bar{d}, s + \bar{s}\}$ .

## 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, RHIC and LHC. This theoretical framework will also be essential for both the physics programme 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.

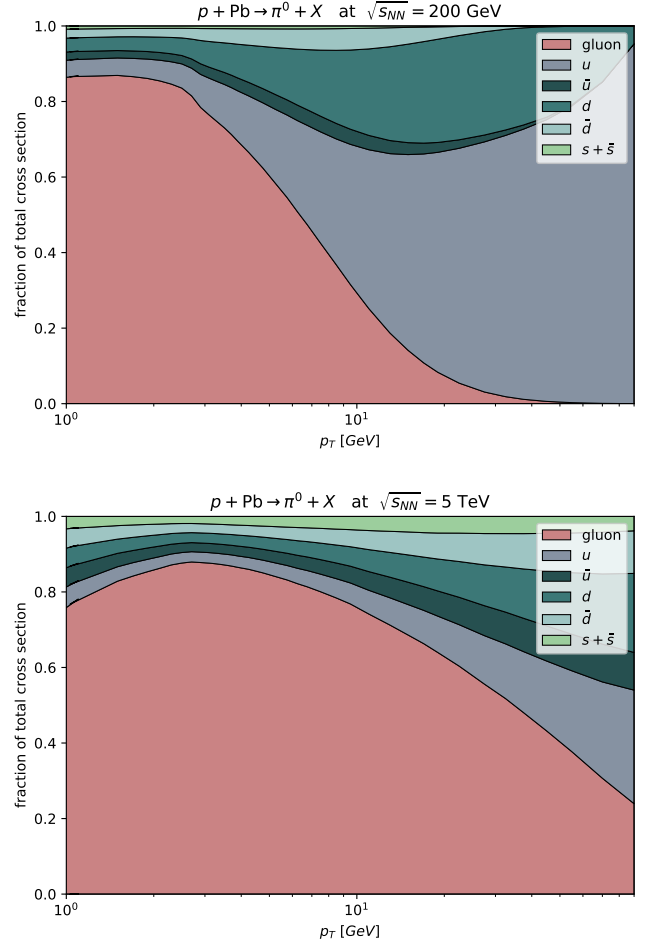


FIG. 2. Fractional contribution of the total  $p+\text{Pb} \rightarrow \pi^0 + X$  cross section contributed by each fragmentation function,  $D_i^{\pi^0}(z, Q)$ , at  $\sqrt{s_{NN}} = 200$  GeV (upper panel) and 5 TeV (lower panel) for  $i \in \{g, u, d, \bar{u}, \bar{d}, s + \bar{s}\}$ .

### A. The gluon PDF

Single Inclusive Hadron (SIH) production data has 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 a gluon, 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 PDF flavors (Fig. 1), but with slightly larger contributions from the antiquarks. Both figures are computed with nCTEQ15WZ PDFs [23] and DSS FFs [24], 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 data sets. 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. The 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. [25]), the nuclear data was 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 data sets 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 [23] and nCTEQ15HIX [26], are performed with a new C++ based code nCTEQ++. This allows us to easily interface external programs such as HOPPET [27], APPLgrid [28], and INCNLO [29].

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 3 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

$Q_0 = 1.3 \text{ GeV}$  so that it had the same form as the other sea quarks.

Our PDFs are parameterized at the initial scale  $Q_0 = 1.3 \text{ GeV}$  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

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

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  $\Lambda/Q$  where  $\Lambda$  is a hadronic scale and  $Q$  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 [30, 31] 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.<sup>1</sup> A detailed overview is given in Ref. [34]. 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

<sup>1</sup> For additional details regarding target mass corrections, see Refs. [32, 33].

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

Data set	Ref.	ID	$\sqrt{s_{NN}}$ [GeV]	Observ.	No. points
PHENIX $\pi^0$	[35]	4003	200	$R_{dAu}$	21
PHENIX $\eta$	[35]	4403	200	$R_{dAu}$	12
PHENIX $\pi^\pm$	[36]	4103	200	$R_{dAu}$	20
PHENIX $K^\pm$	[36]	4203	200	$R_{dAu}$	15
STAR $\pi^0$	[37]	4002	200	$R_{dAu}$	13
STAR $\eta$	[37]	4402	200	$R_{dAu}$	7
STAR $\pi^\pm$	[38]	4102	200	$R_{dAu}$	23
ALICE 5 TeV $\pi^0$	[39]	4001	5020	$R_{pPb}$	31
ALICE 5 TeV $\eta$	[39]	4401	5020	$R_{pPb}$	16
ALICE 5 TeV $\pi^\pm$	[40]	4101	5020	$R_{pPb}$	58
ALICE 5 TeV $K^\pm$	[40]	4201	5020	$R_{pPb}$	58
ALICE 8 TeV $\pi^0$	[41]	4004	8160	$R_{pPb}$	44
ALICE 8 TeV $\eta$	[41]	4404	8160	$R_{pPb}$	14

the fitting process. In order to perform the corresponding calculations and produce such grids we have modified the INCNLO [29] program. The obtained grids have been validated to reproduce the full calculation within a margin significantly smaller than the data uncertainty.

### C. SIH data sets

In this analysis we include the same deep inelastic scattering (DIS) and Drell-Yan (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 is taken at center of mass energies per nucleon of 200 GeV (PHENIX, STAR), 5020 GeV and 8160 GeV (ALICE). Table I gives an overview of the available data sets, 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 reliable for momentum fractions  $z_{\min} < z < 1$  with some minimal  $z_{\min} \sim 0.1$ . This issue

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

FF	Ref.	Year	Available particles
BKK	[42]	1994	$\pi_0, \pi^\pm, K^\pm$
KKP	[43]	2000	$\pi_0, \pi^\pm, K^\pm$
KRETZER	[44]	2000	$\pi_0, \pi^\pm, K^\pm$
HKNS07	[45]	2007	$\pi_0, \pi^\pm, K^\pm$
AKK	[46]	2008	$\pi_0, \pi^\pm, K^\pm$
NNFF	[47]	2017	$\pi_0, \pi^\pm, K^\pm$
JAM20	[48]	2021	$\pi_0, \pi^\pm, K^\pm$
DSS14	[24]	2014	$\pi_0, \pi^\pm$
DSS17	[49]	2017	$K^\pm$
AESSS	[50]	2011	$\eta$

will be further discussed in Sec. III A.

All the single inclusive hadron production data is 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 this data has 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 [26] analysis.

## II. FRAGMENTATION FUNCTIONS

In this analysis we investigate a total of ten different fragmentation functions (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 an final-state hadron. Both the PDFs and FFs are non-perturbative 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 [51, 52].

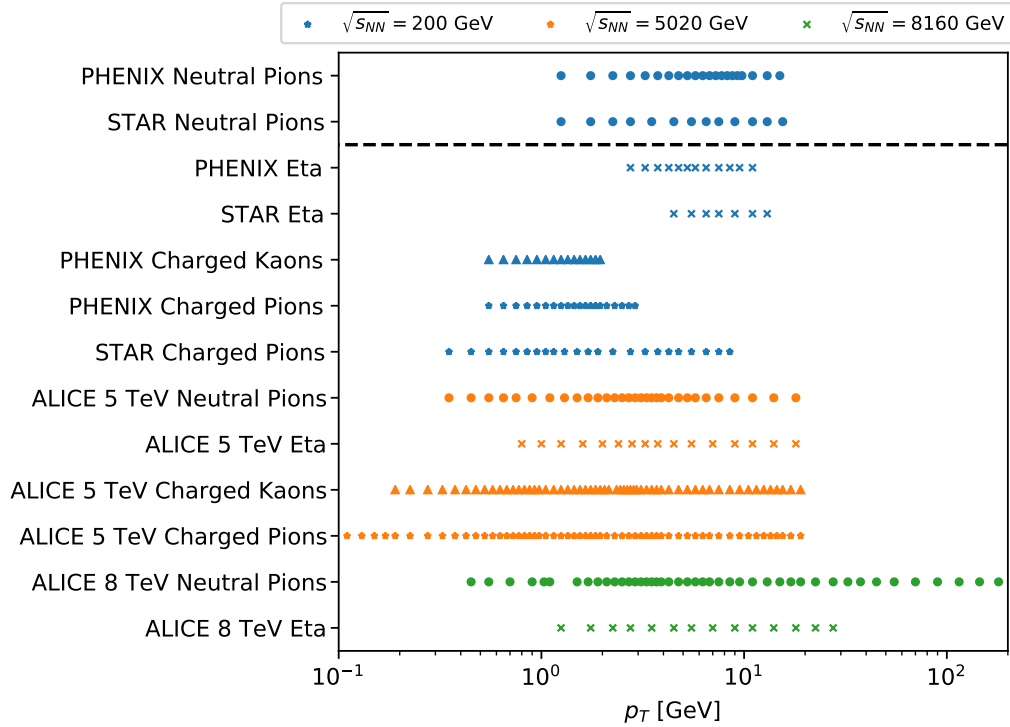


FIG. 3.  $p_T$  values of all data points, separated by set and colored by  $\sqrt{s_{NN}}$ . The data sets above the dashed line were included in the nCTEQ15 analysis.

A selection of FFs for various mesons<sup>2</sup> 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 [53] 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 [54], NNFF1.1h [55]). Some of the aforementioned FFs (AKK, BKK, HKNS, KKP, 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 [56, 57], 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  $\sqrt{s_{NN}} = 200$  GeV 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  $\chi^2/N_{dof} < 1$  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 upwards slope towards higher  $p_T$  for all fragmentation functions, but it is well within the data uncertainties, considering the allowed normalization shift.

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

<sup>2</sup> We 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  $\pi^+$  and  $\pi^-$  FFs.

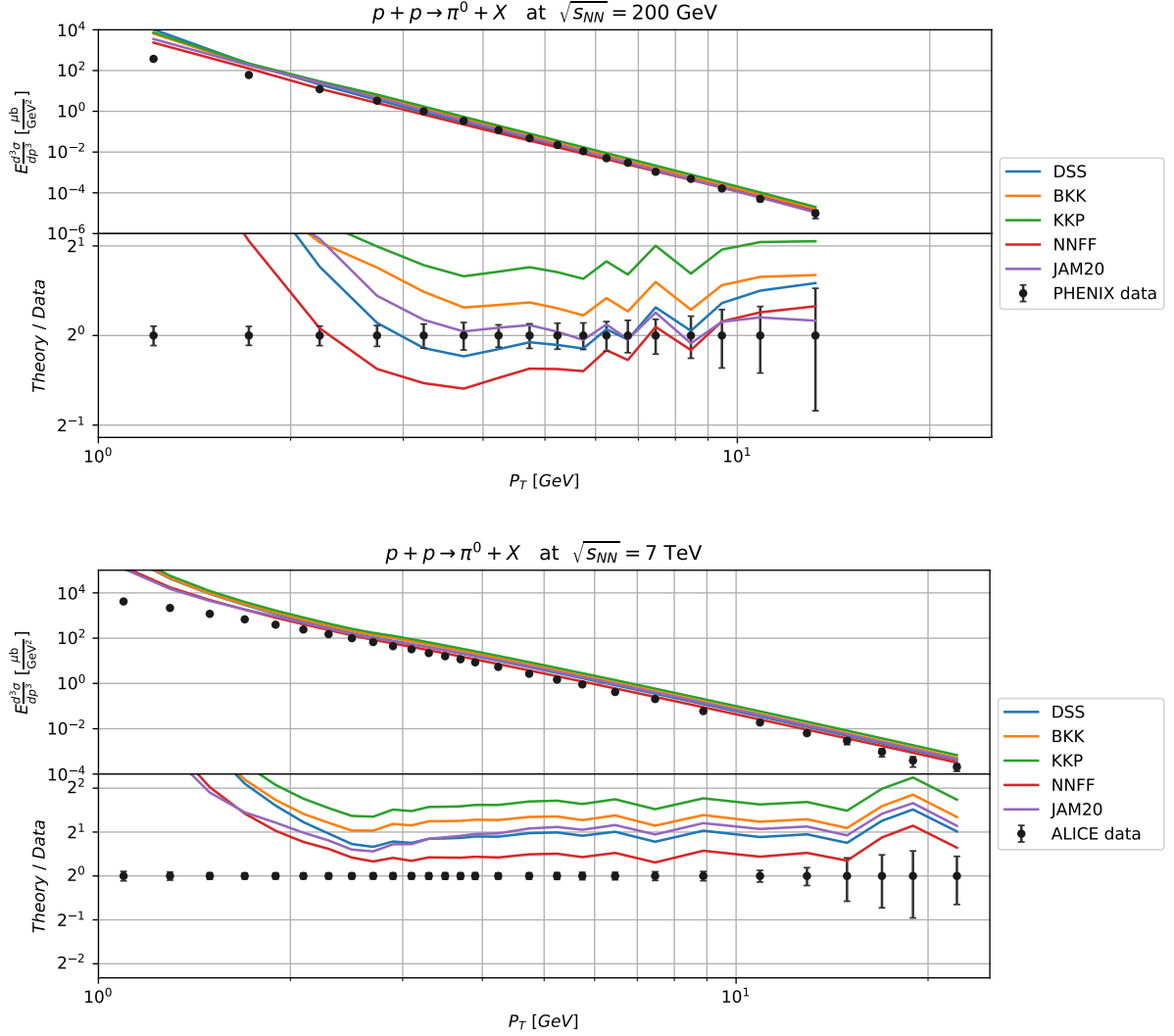


FIG. 4. Comparison of predictions made with different fragmentation functions for  $p + p \rightarrow \pi^0 + X$ . The calculations are done using nCTEQ15 proton PDFs. Both panels show data for neutral pions, with PHENIX data [58] in the upper and ALICE data [59] in the lower one.

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.<sup>3</sup> 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.

### C. Scale uncertainties

The prediction for the SIH production cross section depends on three scale choices: the initial state factorization scale  $\mu_i$ , the final state factorization scale  $\mu_f$  and the renormalization scale  $\mu_r$ . 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  $1/2$  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,  $c = \{1/2, 1\}$ . The case where all scales are equal to  $1/2 p_T$  (bold, grey) gives the best description of the 200 GeV data; additionally, this is also the only scale choice that yields  $\chi^2/N_{dof} < 1$  for the 7 TeV data with  $p_T > 3$  GeV, if the normalization is

<sup>3</sup> In 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 [60] PDFs. Since our focus is the impact on the nuclear PDFs however, we use our proton baseline instead.

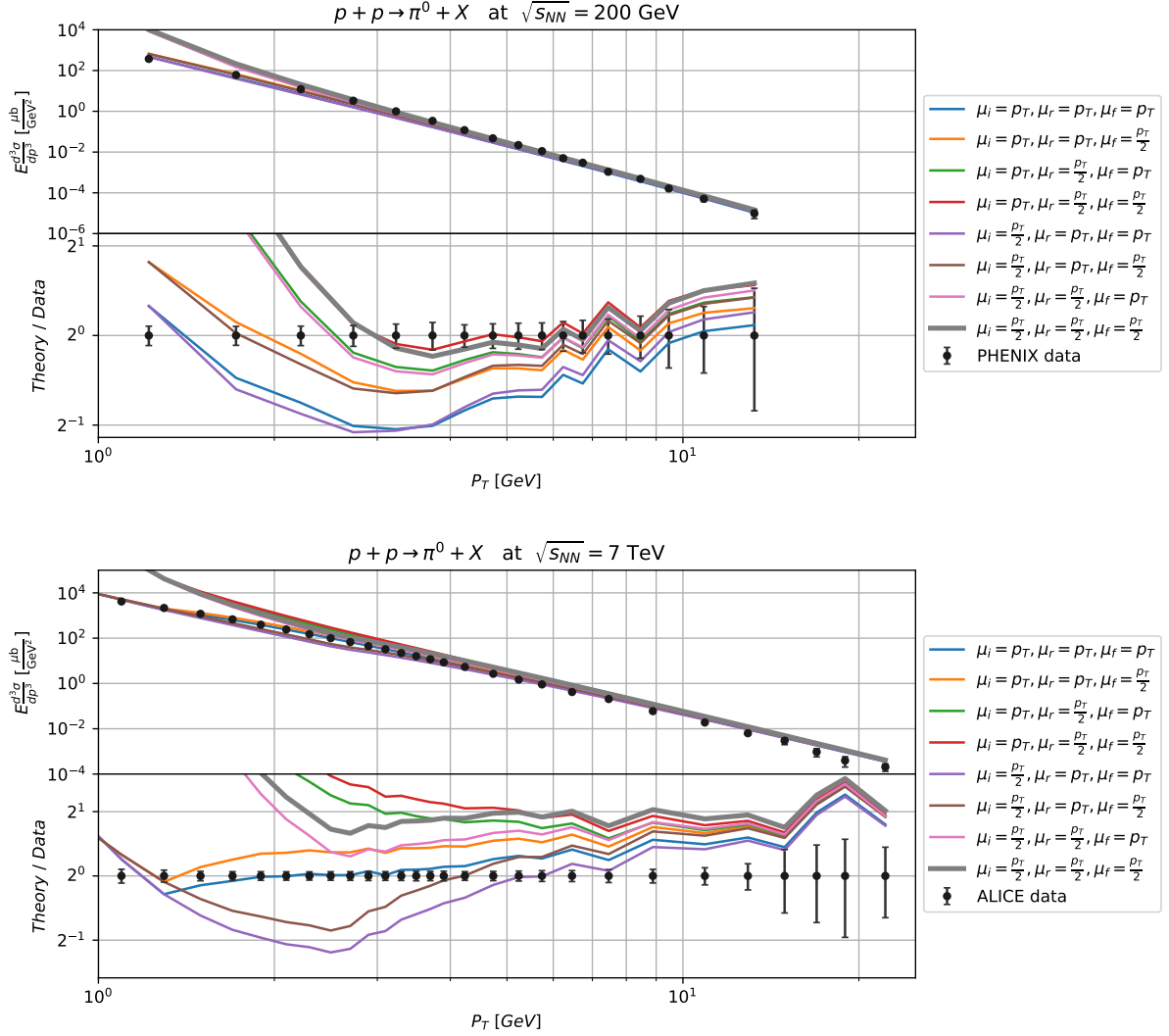


FIG. 5. Comparison of predictions made with different scale choices for  $p + p \rightarrow \pi^0 + X$ . The calculations are performed using nCTEQ15 proton PDFs with DSS fragmentation functions. Both panels show data for neutral pions, with PHENIX data in the upper panel and ALICE data in the lower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

chosen freely. Therefore, in the following comparisons we make the choice  $c = 1/2$  going forward. This means that we need to freeze the initial state factorization scale ( $\mu_i$ ) 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 data sets 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 data sets with predictions using the nCTEQ15WZ nuclear PDFs for each set of FFs. Data taken at  $p_T < 1$  GeV is 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

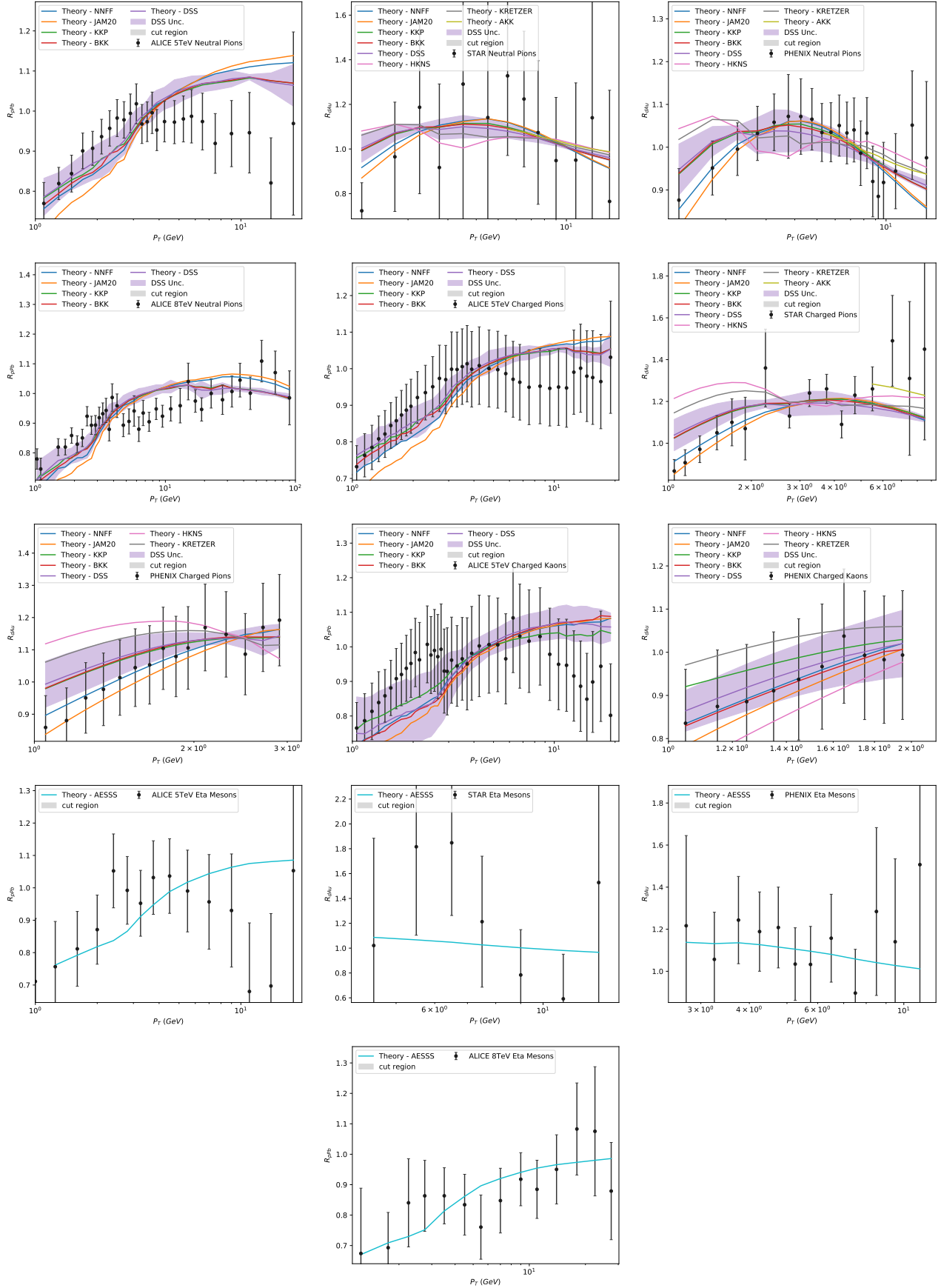


FIG. 6. Comparison of data for nuclear ratios  $R_{AA'}$  for pion, kaon, and eta production with theoretical predictions at NLO QCD using nCTEQ15WZ nuclear PDFs and different FFs. The predictions are scaled by the inverse of their fitted normalization. The uncertainties of the DSS fragmentation functions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



each other across all data sets, 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 is 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 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 FFs 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 parameterization-free neural network instead of a “traditional” parameterization, and a slightly smaller data set. 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 two reasons. Firstly, 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 exists is heavily dependent on the specific parameterization of the FF. Even more important are low  $p_T$  corrections. As we move to lower values of  $p_T$ , perturbation theory begins to breaking down as contributions from non-perturbative sources increase. Thus, for our predictions in the lower  $p_T$  range, we may

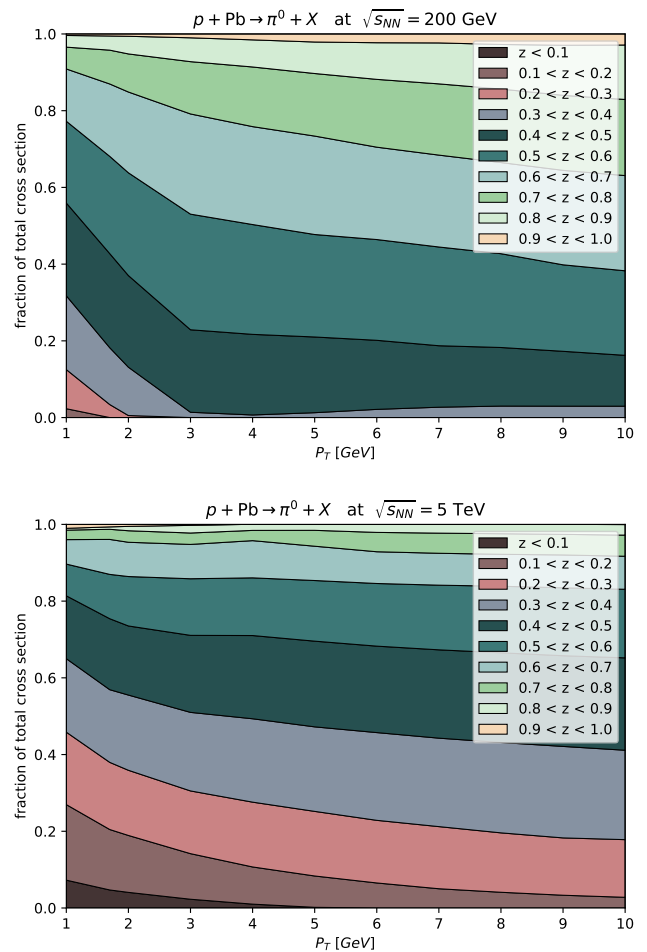


FIG. 7. Contribution of different  $z$  regions of the fragmentation functions,  $D^{\pi^0}(z, Q)$ , to the total  $p+\text{Pb} \rightarrow \pi^0+X$  cross section at  $\sqrt{s_{NN}} = 200$  GeV (top) and 5 TeV (bottom).

reach the transition region between stable perturbative predictions and unreliable non-perturbative predictions.

These two factors are part of 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  $p+\text{Pb} \rightarrow \pi^0+X$  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 towards 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.

### III. IMPACT OF SIH DATA ON PDF FITS

#### A. Data selection

Before performing the fits, we need to decide which data sets to include, and which kinematic cuts to impose. Firstly, we choose not to include the eta meson data in the current analysis as we only have a single data set with comparatively large uncertainties; but, we will examine this data set 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] which both 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 86 (out of 188) 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.

- The included data sets are neutral pions (STAR, PHENIX, ALICE 5 and 8 TeV), charged pions (STAR, ALICE 5 TeV) and charged kaons (ALICE 5 TeV).
- Cuts are applied below  $p_T = 3$  GeV for all data sets.

- Eta mesons are not included in the current fits; we examine this data later in Sec. III G.
- PHENIX charged hadrons are excluded by our  $p_T$  cut.
- Normalizations of all SIH data sets are fitted according to the prescription given in the Appendix A.
- 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.
- 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 since the new data is taken on lead and gold, which is 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' 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 a 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 constrains the region  $x \gtrsim 0.01$ . Including the DSS uncertainty in this case causes no visible difference

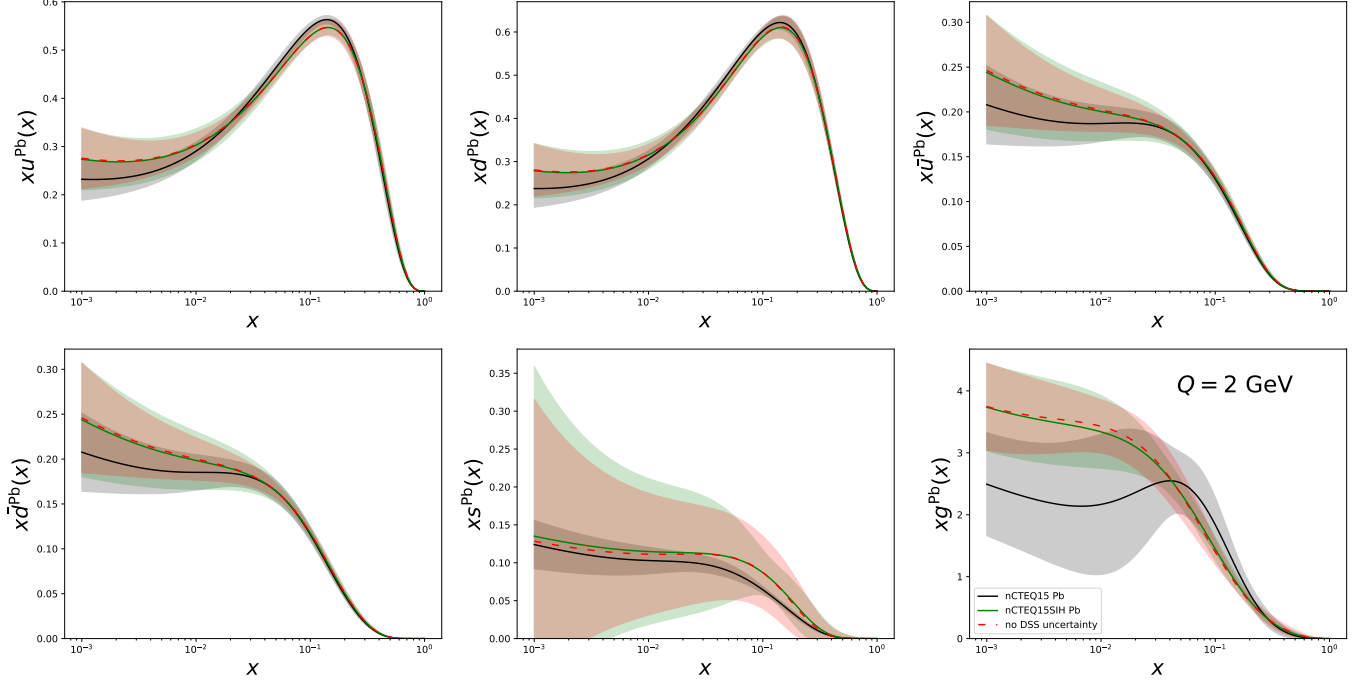


FIG. 8. Lead PDFs from fits to the nCTEQ15 data + SIH data. The baseline nCTEQ15 fit is shown in black, the fit with unmodified data is shown in red and the fit where the uncertainties from the DSS FFs were added as a systematic uncertainty (nCTEQ15SIH) is shown in green.

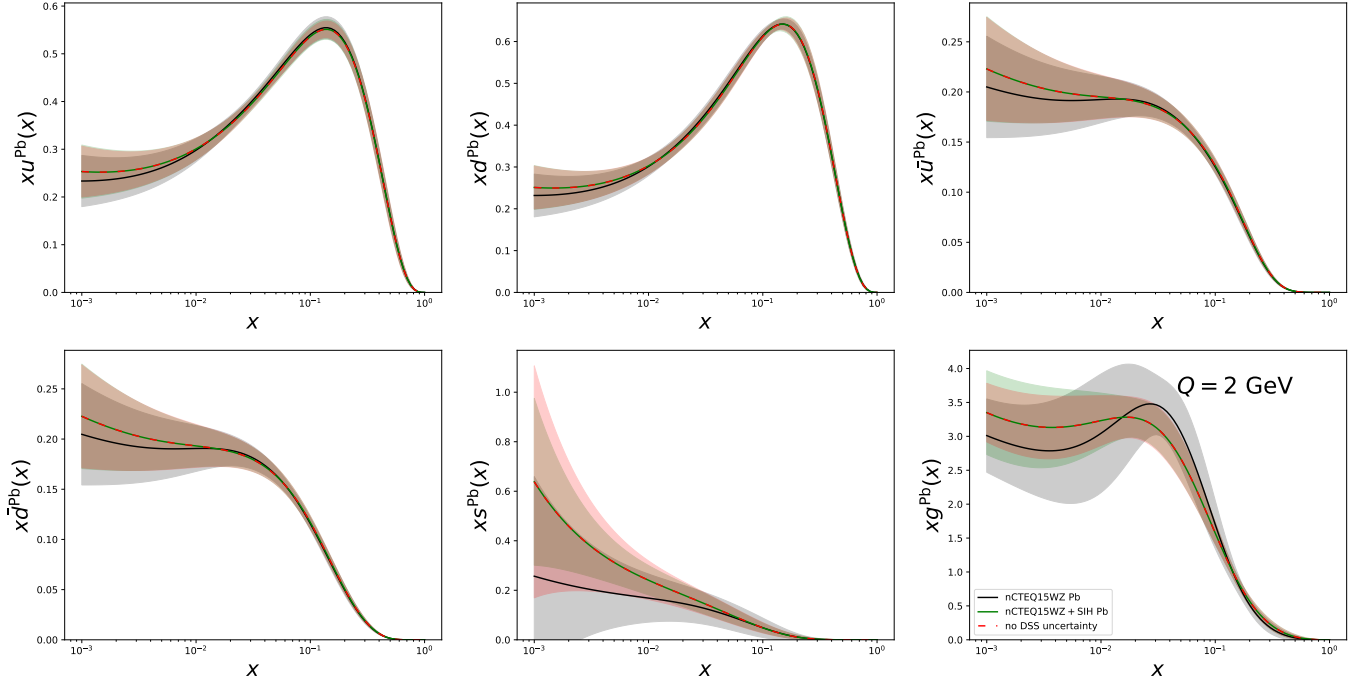


FIG. 9. Lead PDFs from fits to the nCTEQ15WZ data + SIH data. The baseline nCTEQ15WZ fit is shown in black, the fit with unmodified data is shown in red and the fit where the uncertainties from the DSS FFs were added as a systematic uncertainty (nCTEQ15WZ+SIH) is shown in green.

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  $\chi^2$  values. Figure 10 shows the  $\chi^2/N_{dof}$  for each of the fitted data sets of the two main fits, nCTEQ15SIH and nCTEQ15WZ+SIH. We see that the DIS and DY data sets are still well described by the new PDF, and generally satisfy  $\chi^2/N_{dof} < 1$ , with one exception.<sup>4</sup> The  $W/Z$  data also remains well described when including the SIH data, with the exception of data set 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  $\chi^2/N_{dof}$  by experiment type and by data set. 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.

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  $\chi^2/N_{dof}$  values. Adding SIH data to the fit (nCTEQ15SIH) significantly improves the SIH data from  $\chi^2/N_{dof} = 1.23$  to 0.38, as well as the  $W/Z$  data from  $\chi^2/N_{dof} = 3.74$  to 2.32. There is also a slight improvement in the DY data, with a marginal increase in the DIS  $\chi^2$ .

In a similar manner, the nCTEQ15WZ fit yields good  $\chi^2$  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  $\chi^2/N_{dof} = 0.81$  to 0.41. This results in marginal shifts for the DIS and DY data, but does increase the  $W/Z$  data from  $\chi^2/N_{dof} = 0.90$  to 1.02. However, the total  $\chi^2/N_{dof}$  for the combined fit nCTEQ15WZ+SIH is  $\chi^2/N_{dof} = 0.85$  as compared to the nCTEQ15WZ with 0.90.

Table III shows the fitted normalizations of the SIH data sets. All the resulting normalization parameters are 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  $\chi^2/N_{dof}$  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  $\chi^2$  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 modified 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  $\chi^2/N_{dof}$  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 data sets as compared to the other fits. This observation suggests that as we add more data set, 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%.

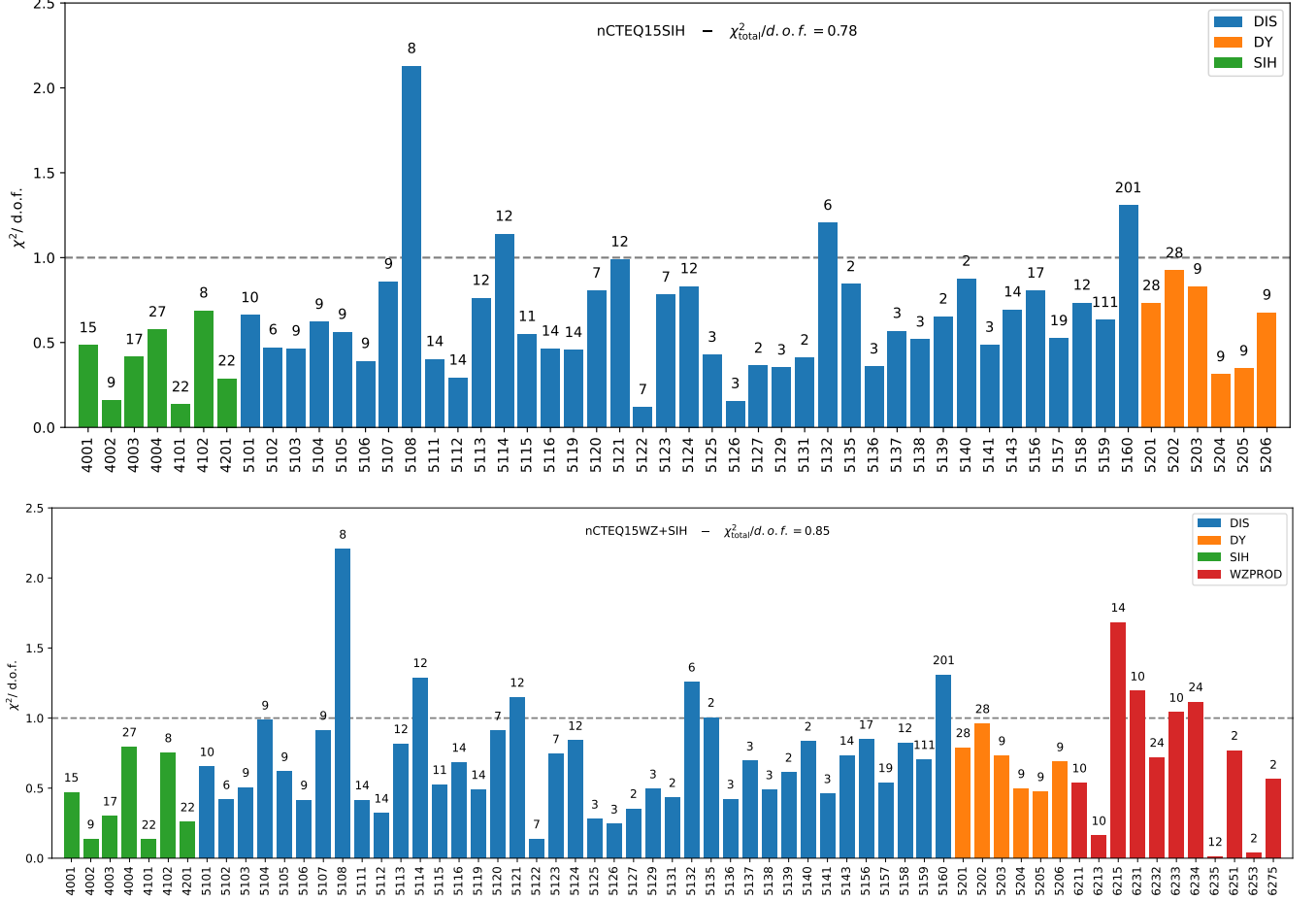
Comparing the nCTEQ15SIH fit and its baseline (nCTEQ15), we see a considerable shift in the  $R_{AA'}$  nuclear ratio when the SIH data is included. In contrast, the nCTEQ15WZ+SIH fit and its baseline (nCTEQ15WZ) show only subtle differences, aside from the normalization, which is not surprising given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

<sup>4</sup> The notable exception with a large  $\chi^2/N_{dof}$  is data set 5108 (Sn/D EMC-1988) with 8 data points. However, other analyses

also found a large  $\chi^2/N_{dof}$  for this data set [4, 61].

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

	STAR		PHENIX	ALICE			
	$\pi^0$	$\pi^\pm$	$\pi^0$	5 TeV $\pi^0$	5 TeV $\pi^\pm$	5 TeV $K^\pm$	8 TeV $\pi^0$
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

FIG. 10. The  $\chi^2/N_{dof}$  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. [23].

### F. Correlation between data and PDFs

By looking at the PDFs alone we cannot judge the impact of each individual new data set on the fit. Therefore we make use of two further methods to study how each data set 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, 62]:

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

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

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 and error size of the experiments because the normalization

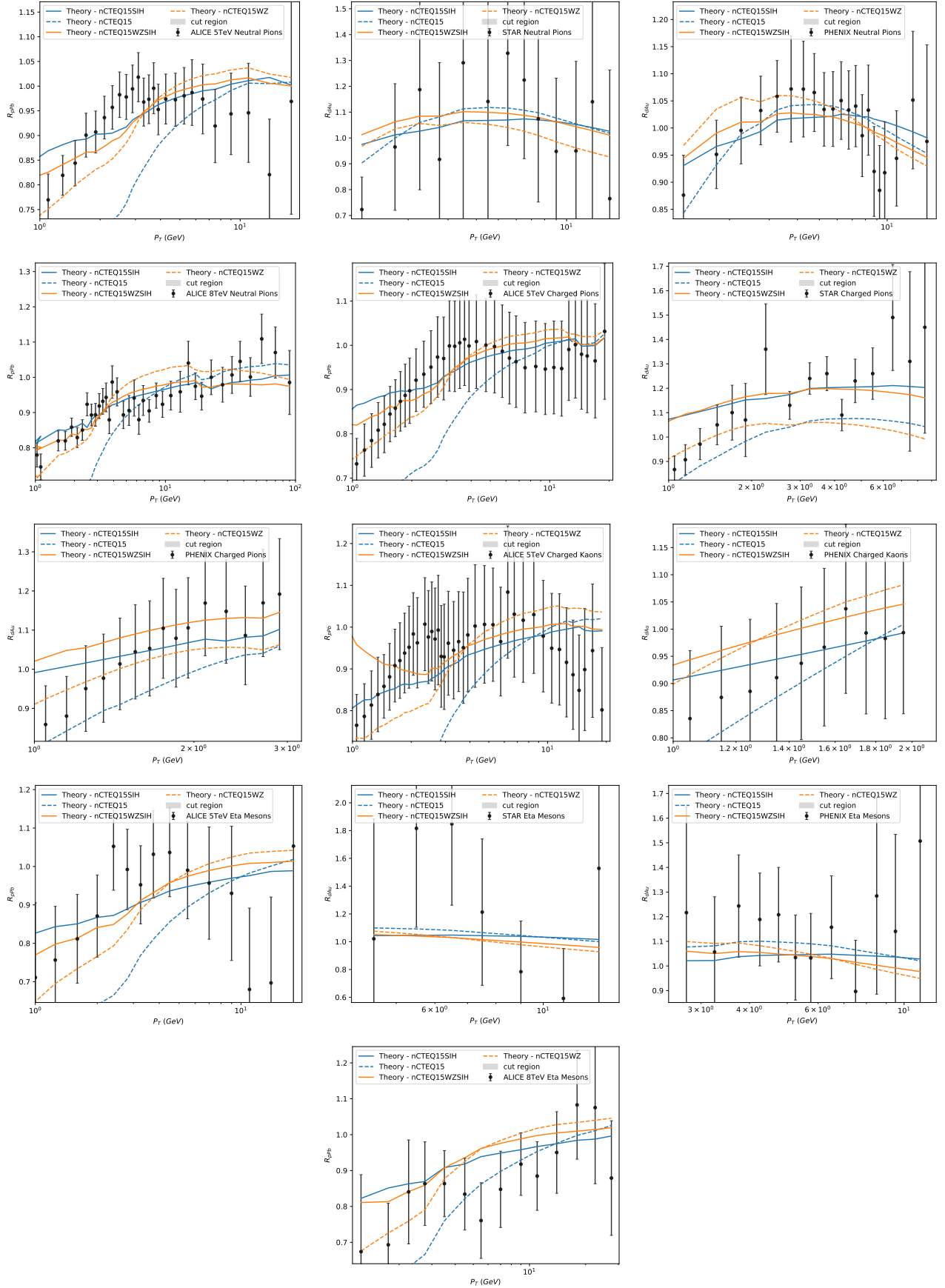


FIG. 11. Theory predictions for the main fits and their respective baselines. Dashed curves indicate the baseline fits, while solid curves show the fits with SIH data included. The blue curves are based on nCTEQ15 and the orange ones on nCTEQ15WZ.

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

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

TABLE V. The  $\chi^2/N_{dof}$  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

does not cancel these factors out. For an experiment  $E_j$  and an observable  $X$ , it is defined as:

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

To investigate the impact of individual experimental data sets  $E_j$  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  $\chi^2$  difference  $\Delta\chi_{\text{eff}}^2[g(x, Q), E_j]$ . Since neither of these quantities display 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 data sets ( $\pi^0, \pi^\pm, K^\pm$ ) display a strong anti-correlation ( $\cos \phi \sim -0.9$ ) 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 has 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 anti-correlation ( $\cos \phi \sim -0.9$ ), which then becomes strong and positive ( $\cos \phi \sim +0.8$ ) for larger  $x$ . The 8 TeV ALICE neutral pion data show a correlation behaviour similar to the NMC96 SnC data set (which is the dominant DIS set due to its large size and  $Q$  coverage), and somewhat opposite to the STAR and PHENIX neutral pion data. Examining the larger  $x$  region ( $x > 0.1$ ), the influence of the various data sets is more mixed with with the STAR and PHENIX  $\pi^0$  data yielding a large positive correlation and ATLAS 8 TeV  $\pi^0$  and STAR  $\pi^\pm$  yielding a large negative correlation,

with the result that the high  $x$  gluon remains mostly unchanged in Fig. 9.

Turning to the  $\chi_{\text{eff}}^2$  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 data sets also providing constraints.

Among the SIH data sets, the 8 TeV neutral pion data has the largest  $\chi_{\text{eff}}^2$ , followed by the 5 TeV neutral pion data. However, they generally do not reach values as high as the previously mentioned DIS and WZ 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.

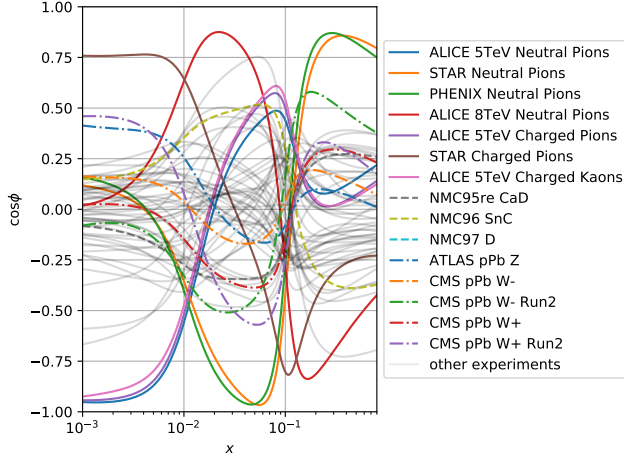


FIG. 12. Cosine of the correlation angle  $\cos(\phi[g(x, Q), \chi^2(E_j)])$  between gluon PDF and  $\chi^2$  of each experimental data set ( $E_j$ ) for the nCTEQ15WZ+SIH fit at  $Q = 10$  GeV.

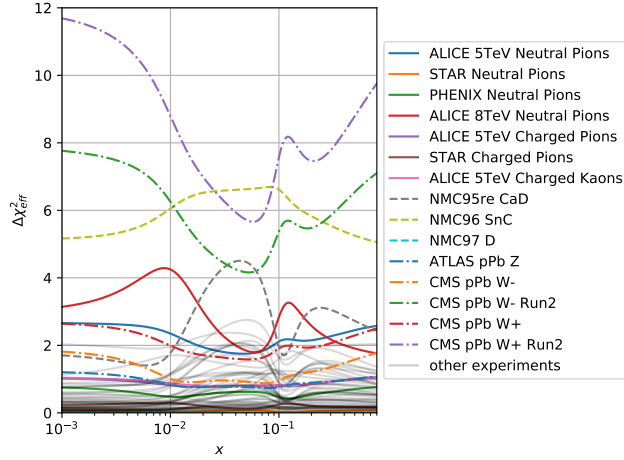


FIG. 13.  $\Delta\chi^2_{\text{eff}}[g(x, Q), E_j]$  for the gluon PDF with each experimental data set ( $E_j$ ) for the nCTEQ15WZ+SIH fit at  $Q = 10$  GeV.

### 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 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 provides 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 upwards 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 nCTEQWZ+SIH fit with and without the eta meson data are virtually identical. Regarding the PDF uncertainties, the error bands of 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 (SIH) 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  $\chi^2/N_{\text{dof}}$  for all the data sets, 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.



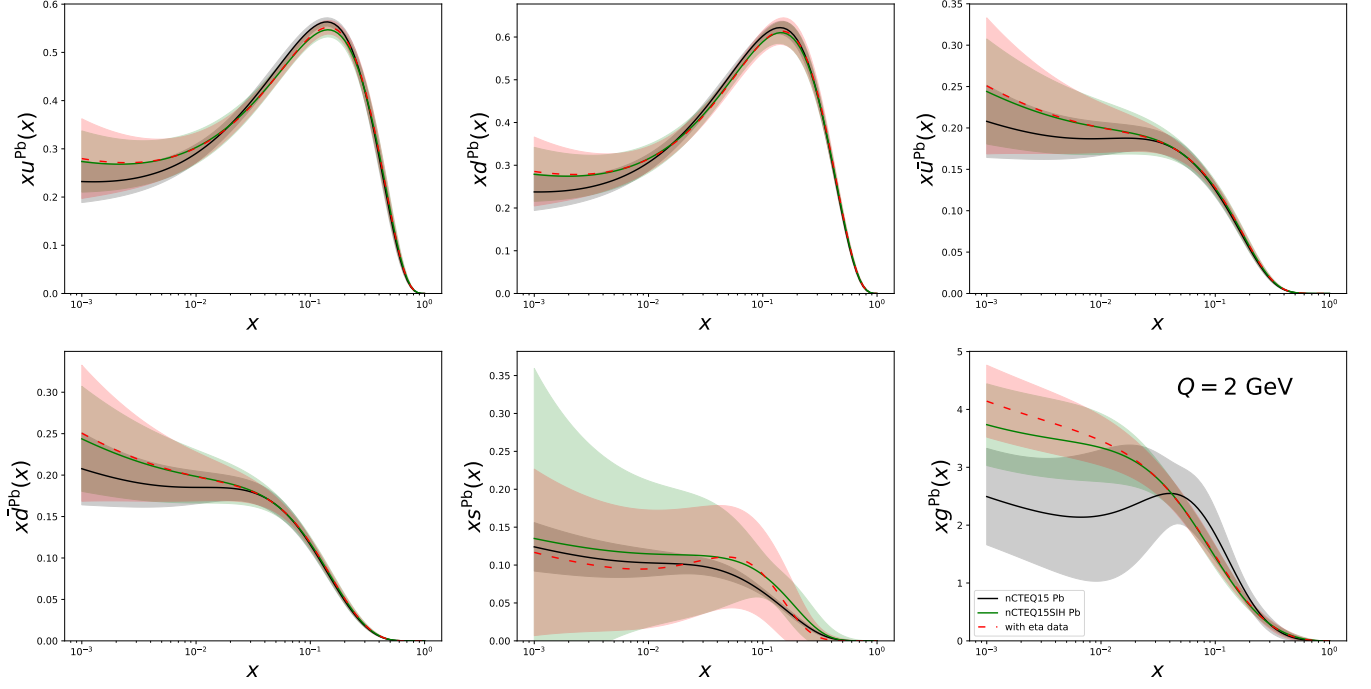


FIG. 14. Lead PDFs from fits to the nCTEQ15 data including the SIH eta data. The baseline nCTEQ15 fit is shown in black, the fit with eta meson data is shown in red and the corresponding main fit is shown in green. The nCTEQ15 (black) and nCTEQ15SIH (green) are also displayed in Fig. 8.

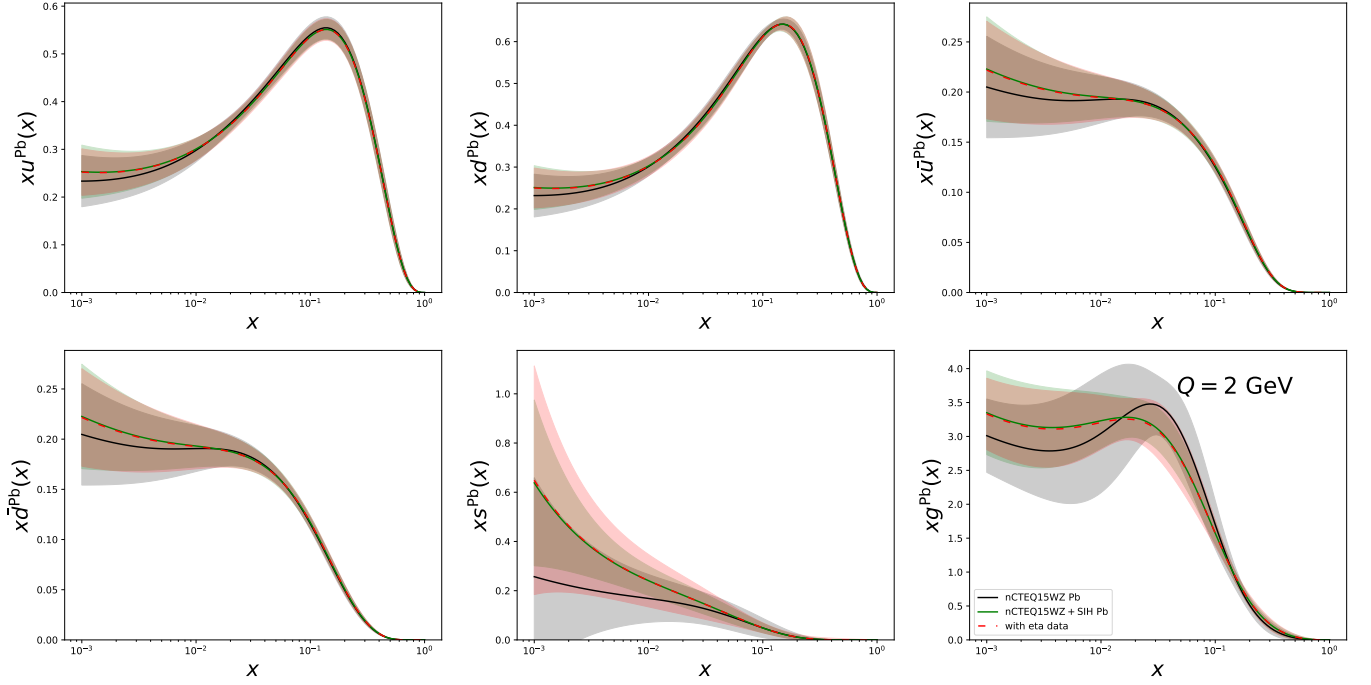


FIG. 15. Lead PDFs from fits to the nCTEQ15WZ data including the SIH eta data. The baseline nCTEQ15WZ fit is shown in black, the fit with eta meson data is shown in red and the corresponding main fit is shown in green. The nCTEQ15 (black) and nCTEQ15WZ+SIH (green) are also displayed in Fig. 9.

## ACKNOWLEDGEMENTS

The work of P.D., M.K. and K.K. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project-id 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 396021762 - TRR 257. F.O. acknowledges support through US DOE 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 GLUE@NLO.

## Appendix A: Fitting data normalizations

We use the  $\chi^2$  prescription given in Ref. [63] to fit the normalizations of the SIH and  $W/Z$  production data. For a data set  $D$  with  $N$  data points and  $S$  correlated systematic errors, the  $\chi^2$  of the data set is given by:

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

where  $\sigma_{\text{norm}}$  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  $\sigma_{\text{norm}}$ , which is around 0.03 for  $W/Z$  production and ranges from 0.06 to 0.17 for SIH production. The covariance matrix  $C_{ij}$  is defined as:

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

where  $\sigma_i$  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  $\alpha$ . We use the analytical formula for the inverse of the correlation matrix found in Ref. [64] 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 parameterization-free neural network instead of a “traditional” parameterization, and a slightly smaller data set. 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.

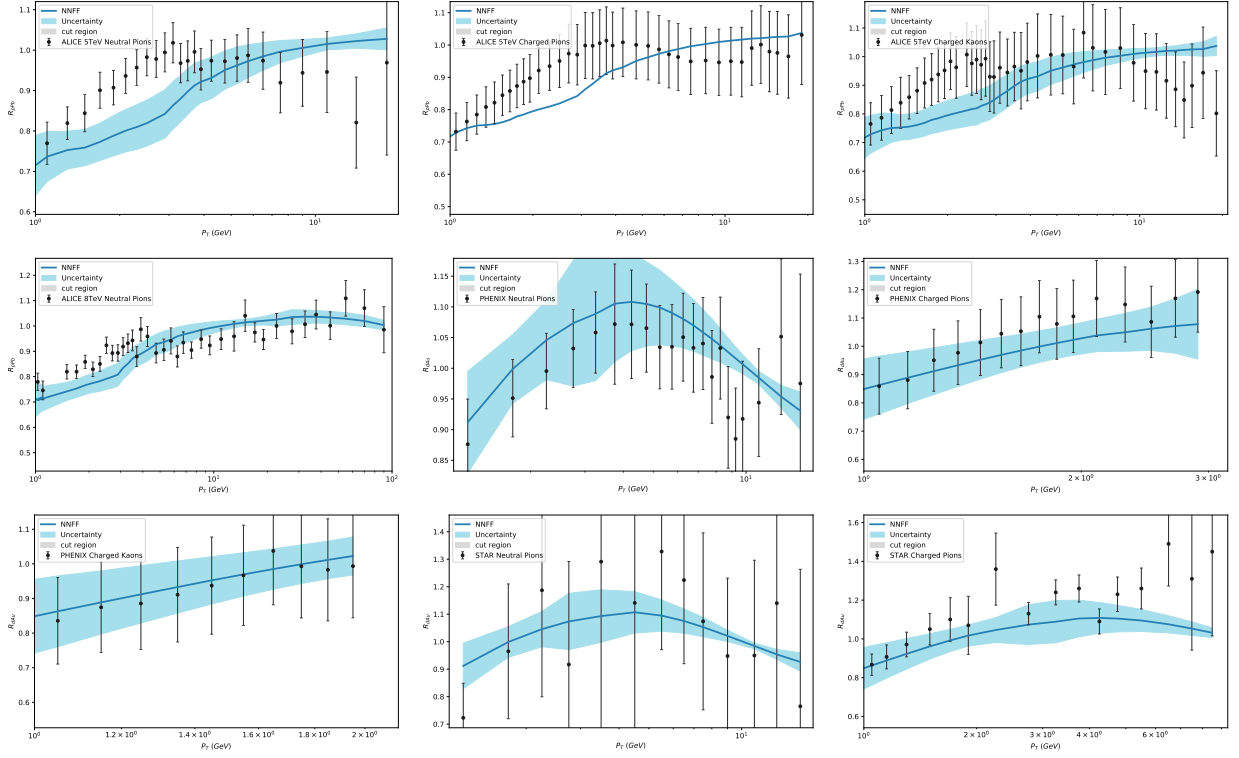


FIG. 16. Uncertainties calculated from the NNFF replicas using nCTEQ15WZ PDFs. The computed uncertainties use 83 of the 101 provided replicas because the remaining 18 lead to unphysical behaviour such as large jumps from one  $p_T$  value to another or negative cross sections due to numerical problems.

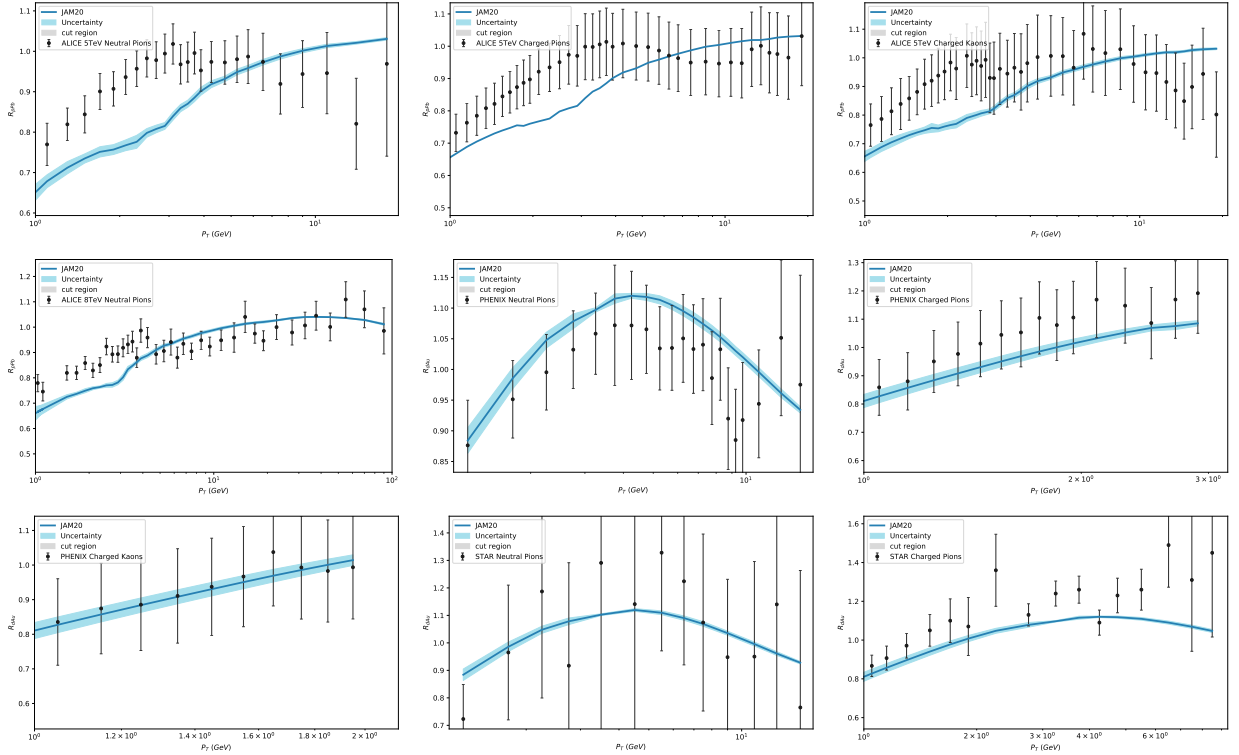


FIG. 17. Uncertainties calculated from the JAM20 replicas using nCTEQ15WZ PDFs.

- 
- [1] T.-J. Hou et al., “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC,” [1912.10053](#).
- [2] **NNPDF** Collaboration, R. D. Ball et al., “Parton distributions from high-precision collider data,” *Eur. Phys. J.* **C77** (2017), no. 10, 663, [1706.00428](#).
- [3] K. Kovarik et al., “nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework,” *Phys. Rev. D* **93** (2016), no. 8, 085037, [1509.00792](#).
- [4] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, “EPPS16: Nuclear parton distributions with LHC data,” *Eur. Phys. J. C* **77** (2017), no. 3, 163, [1612.05741](#).
- [5] **NNPDF** Collaboration, R. Abdul Khalek, J. J. Ethier, and J. Rojo, “Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider,” *Eur. Phys. J.* **C79** (2019), no. 6, 471, [1904.00018](#).
- [6] R. Abdul Khalek, J. J. Ethier, J. Rojo, and G. van Weelden, “nNNPDF2.0: Quark Flavor Separation in Nuclei from LHC Data,” [2006.14629](#).
- [7] J. J. Ethier and E. R. Nocera, “Parton Distributions in Nucleons and Nuclei,” *Ann. Rev. Nucl. Part. Sci.* (2020), no. 70, 1–34, [2001.07722](#).
- [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** (2018), no. 11, 962, [1810.03639](#).
- [9] J. Gao, L. Harland-Lang, and J. Rojo, “The Structure of the Proton in the LHC Precision Era,” *Phys. Rept.* **742** (2018) 1–121, [1709.04922](#).
- [10] K. Kovarik, P. M. Nadolsky, and D. E. Soper, “Hadron structure in high-energy collisions,” *Rev. Mod. Phys.* **92** (2020) 045003, [1905.06957](#).
- [11] S. Alekhin, J. Blümlein, and S. Moch, “Strange sea determination from collider data,” *Phys. Lett. B* **777** (2018) 134–140, [1708.01067](#).
- [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** (2008) 013004, [0802.0007](#).
- [13] **JAM** Collaboration, N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk, “Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions,” *Phys. Rev. D* **101** (2020), no. 7, 074020, [1905.03788](#).
- [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** (2015), no. 5, 204, [1412.3989](#).
- [15] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. Martin, and R. Nathvani, “Updates of PDFs using the MMHT framework,” *PoS DIS2019* (2019) 036, [1907.08147](#).
- [16] **NNPDF** Collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali, “Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering,” *Nucl. Phys. B* **823** (2009) 195–233, [0906.1958](#).
- [17] H.-W. Lin et al., “Parton distributions and lattice QCD calculations: a community white paper,” *Prog. Part. Nucl. Phys.* **100** (2018) 107–160, [1711.07916](#).
- [18] H.-W. Lin et al., “Parton distributions and lattice QCD calculations: toward 3D structure,” [2006.08636](#).
- [19] V. Guzey and M. Klasen, “Constraints on nuclear parton distributions from dijet photoproduction at the LHC,” *Eur. Phys. J. C* **79** (2019), no. 5, 396, [1902.05126](#).
- [20] M. Klasen, K. Kovarik, and J. Potthoff, “Nuclear parton density functions from jet production in DIS at an EIC,” *Phys. Rev. D* **95** (2017), no. 9, 094013, [1703.02864](#).
- [21] M. Klasen and K. Kovarik, “Nuclear parton density functions from dijet photoproduction at the EIC,” *Phys. Rev. D* **97** (2018), no. 11, 114013, [1803.10985](#).
- [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** (2016), no. 4, 218, [1512.01528](#).
- [23] A. Kusina et al., “Impact of LHC vector boson production in heavy ion collisions on strange PDFs,” *Eur. Phys. J. C* **80** (2020), no. 10, 968, [2007.09100](#).
- [24] D. de Florian, R. Sassot, M. Epele, R. J. Hernández-Pinto, and M. Stratmann, “Parton-to-Pion Fragmentation Reloaded,” *Phys. Rev. D* **91** (2015), no. 1, 014035, [1410.6027](#).
- [25] 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** (2005) 145–156, [hep-ph/0312323](#).
- [26] E. P. Segarra et al., “nCTEQ15HIX – Extending nPDF Analyses into the High- $x$ , Low  $Q^2$  Region,” [2012.11566](#).
- [27] G. P. Salam and J. Rojo, “A Higher Order Perturbative Parton Evolution Toolkit (HOPPET),” *Comput. Phys. Commun.* **180** (2009) 120–156, [0804.3755](#).
- [28] 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** (2010) 503–524, [0911.2985](#).
- [29] M. Werlen, “INCNLO-direct photon and inclusive hadron production code website.” Version 1.4. [http://lapth.cnrs.fr/PHOX\\_FAMILY](http://lapth.cnrs.fr/PHOX_FAMILY).
- [30] J. C. Collins, D. E. Soper, and G. F. Sterman, “Factorization of Hard Processes in QCD,” *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, [hep-ph/0409313](#).
- [31] S. Albino, “The Hadronization of partons,” *Rev. Mod. Phys.* **82** (2010) 2489–2556, [0810.4255](#).
- [32] I. Schienbein et al., “A Review of Target Mass Corrections,” *J. Phys. G* **35** (2008) 053101, [0709.1775](#).
- [33] **nCTEQ** Collaboration, I. Schienbein et al., “Target mass corrections in lepton-nucleus DIS revisited.” SMU-HEP-21-01 (In preparation).
- [34] P. Aurenche, M. Fontannaz, J. P. Guillet, B. A. Kniehl, and M. Werlen, “Large p(T) inclusive  $\pi^0$  cross-sections and next-to-leading-order QCD predictions,” *Eur. Phys.*

- J. C **13** (2000) 347–355, [hep-ph/9910252](#).
- [35] **PHENIX** Collaboration, S. S. Adler et al., “Centrality dependence of  $\pi^0$  and  $\eta$  production at large transverse momentum in  $\sqrt{s_{NN}} = 200$  GeV d+Au collisions,” *Phys. Rev. Lett.* **98** (2007) 172302, [nucl-ex/0610036](#).
- [36] **PHENIX** Collaboration, A. Adare et al., “Spectra and ratios of identified particles in Au+Au and d+Au collisions at  $\sqrt{s_{NN}} = 200$  GeV,” *Phys. Rev. C* **88** (2013), no. 2, 024906, [1304.3410](#).
- [37] **STAR** Collaboration, B. I. Abelev et al., “Inclusive  $\pi^0$ ,  $\eta$ , 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** (2010) 064904, [0912.3838](#).
- [38] **STAR** Collaboration, J. Adams et al., “Identified hadron spectra at large transverse momentum in p+p and d+Au collisions at  $\sqrt{s_{NN}} = 200$  GeV,” *Phys. Lett. B* **637** (2006) 161–169, [nucl-ex/0601033](#).
- [39] **ALICE** Collaboration, S. Acharya et al., “Neutral pion and  $\eta$  meson production in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *Eur. Phys. J. C* **78** (2018), no. 8, 624, [1801.07051](#).
- [40] **ALICE** Collaboration, J. Adam et al., “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** (2016) 720–735, [1601.03658](#).
- [41] **ALICE** Collaboration, S. Acharya et al., “Nuclear modification factor of light neutral-meson spectra up to high transverse momentum in p-Pb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV,” [2104.03116](#).
- [42] J. Binnewies, B. A. Kniehl, and G. Kramer, “Next-to-leading order fragmentation functions for pions and kaons,” *Z. Phys. C* **65** (1995) 471–480, [hep-ph/9407347](#).
- [43] B. A. Kniehl, G. Kramer, and B. Potter, “Fragmentation functions for pions, kaons, and protons at next-to-leading order,” *Nucl. Phys. B* **582** (2000) 514–536, [hep-ph/0010289](#).
- [44] S. Kretzer, “Fragmentation functions from flavor inclusive and flavor tagged  $e^+e^-$  annihilations,” *Phys. Rev. D* **62** (2000) 054001, [hep-ph/0003177](#).
- [45] M. Hirai, S. Kumano, T. H. Nagai, and K. Sudoh, “Determination of fragmentation functions and their uncertainties,” *Phys. Rev. D* **75** (2007) 094009, [hep-ph/0702250](#).
- [46] S. Albino, B. A. Kniehl, and G. Kramer, “AKK Update: Improvements from New Theoretical Input and Experimental Data,” *Nucl. Phys. B* **803** (2008) 42–104, [0803.2768](#).
- [47] **NNPDF** Collaboration, V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo, “A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties,” *Eur. Phys. J. C* **77** (2017), no. 8, 516, [1706.07049](#).
- [48] E. Moffat, W. Melnitchouk, T. Rogers, and N. Sato, “Simultaneous Monte Carlo analysis of parton densities and fragmentation functions,” [2101.04664](#).
- [49] D. de Florian, M. Epele, R. J. Hernandez-Pinto, R. Sassot, and M. Stratmann, “Parton-to-Kaon Fragmentation Revisited,” *Phys. Rev. D* **95** (2017), no. 9, 094019, [1702.06353](#).
- [50] 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** (2011) 034002, [1009.6145](#).
- [51] D. d’Enterria, K. J. Eskola, I. Helenius, and H. Paukkunen, “Confronting current NLO parton fragmentation functions with inclusive charged-particle spectra at hadron colliders,” *Nucl. Phys. B* **883** (2014) 615–628, [1311.1415](#).
- [52] A. Metz and A. Vossen, “Parton Fragmentation Functions,” *Prog. Part. Nucl. Phys.* **91** (2016) 136–202, [1607.02521](#).
- [53] M. Hirai, H. Kawamura, S. Kumano, and K. Saito, “Impacts of B-factory measurements on determination of fragmentation functions from electron-positron annihilation data,” *PTEP* **2016** (2016), no. 11, 113B04, [1608.04067](#).
- [54] 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** (2018), no. 7, 074002, [1805.04847](#).
- [55] **NNPDF** Collaboration, V. Bertone, N. P. Hartland, E. R. Nocera, J. Rojo, and L. Rottoli, “Charged hadron fragmentation functions from collider data,” *Eur. Phys. J. C* **78** (2018), no. 8, 651, [1807.03310](#).
- [56] R. Sassot, M. Stratmann, and P. Zurita, “Fragmentations Functions in Nuclear Media,” *Phys. Rev. D* **81** (2010) 054001, [0912.1311](#).
- [57] 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** (2019) 502–510, [1903.06170](#).
- [58] **PHENIX** Collaboration, S. S. Adler et al., “Mid-rapidity neutral pion production in proton proton collisions at  $\sqrt{s} = 200$ -GeV,” *Phys. Rev. Lett.* **91** (2003) 241803, [hep-ex/0304038](#).
- [59] **ALICE** Collaboration, B. Abelev et al., “Neutral pion and  $\eta$  meson production in proton-proton collisions at  $\sqrt{s} = 0.9$  TeV and  $\sqrt{s} = 7$  TeV,” *Phys. Lett. B* **717** (2012) 162–172, [1205.5724](#).
- [60] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC,” *Eur. Phys. J. C* **63** (2009) 189–285, [0901.0002](#).
- [61] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, “Global Analysis of Nuclear Parton Distributions,” *Phys. Rev. D* **85** (2012) 074028, [1112.6324](#).
- [62] 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** (2001) 014013, [hep-ph/0101032](#).
- [63] G. D’Agostini, “On the use of the covariance matrix to fit correlated data,” *Nucl. Instrum. Meth. A* **346** (1994) 306–311.
- [64] 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** (2001) 014012, [hep-ph/0101051](#).