Global analysis of nuclear parton distribution functions

Michael Klasen[a,*]

[a]*Institut für Theoretische Physik, Universität Münster,*
Wilhelm-Klemm-Straße 9, 48149 Münster, Germany

E-mail: michael.klasen@uni-muenster.de

We review the theoretical foundations, methodological approaches and current status of the determination of nuclear parton distribution functions (PDFs). A large variety of measurements in fixed-target and collider experiments provide increasingly precise constraints on various aspects of nuclear PDFs, including shadowing, antishadowing, the EMC effect, Fermi motion, flavour separation, deuteron binding, target-mass and other higher-twist effects. We give particular emphasis to measurements carried out in proton-lead collisions at the Large Hadron Collider, which have revolutionised the global analysis during the past decade. These measurements include electroweak boson, isolated photon, single inclusive hadron, jet, and heavy-flavour observables.

31st International Workshop on Deep Inelastic Scattering (DIS2024)
8–12 April 2024
Grenoble, France

---

[*]Speaker

1. Introduction

Nuclear parton distribution functions (PDFs) are an important and dynamic current research topic, as they encode our fundamental understanding of the quark and gluon dynamics in protons (p) and neutrons (n) bound in nuclei (A). They also impact on many other fields in particle and nuclear physics such as global analyses of free proton PDFs, which contain a sizable fraction of nuclear data, the three-dimensional tomography of atomic nuclei, for which they represent important boundary conditions, and the initial-state phase transition to the deconfined quark-gluon plasma, which is believed to be related to gluon saturation and classicalisation. While the evolution of nuclear PDFs with the hard scale Q is calculable in perturbative QCD at next-to-leading order (NLO) and beyond, their dependence on the longitudinal parton momentum fraction x must be fitted to experimental data. For the universality of PDFs in different collision processes, one relies on QCD factorisation theorems. Nuclear modifications such as shadowing at $x \lesssim 0.05$, antishadowing at $0.05 \lesssim x \lesssim 0.3$, the famous EMC effect at $0.3 \lesssim x \lesssim 0.7$, and Fermi motion at $0.7 \lesssim x$ have been known for many years and can be parameterised and fitted, but they still remain to be fundamentally understood.

The key process in any PDF analysis is deep-inelastic scattering (DIS) of charged leptons, which for nuclei has so far only been performed in fixed-target (FT) experiments at relatively small energy and which is at LO only sensitive to quarks. Neutrino scattering provides additional constraints on their flavour separation, in particular on the strange quark and in processes with an identified final-state charm meson. The gluon density is directly accessible in hadronic collisions not only in FT experiments, but also at colliders such as BNL's RHIC and CERN's LHC, and the LHC experiments have collected an enormous amount of data in pPb collisions during the last decade. Since higher-twist (HT) effects, e.g. from final-state rescattering, are enhanced in nuclear collisions, a sufficiently large scale is required to ensure a leading-twist (LT) interpretation of the data. The Fermi motion of individual nucleons seems to be physically reasonably clear, but for shadowing both LT and HT explanations have been proposed, for the EMC effect both partonic and hadronic mechanisms, and antishadowing is mainly justified with the momentum sum rule. More details can be found in our recent review of nuclear PDFs and the references cited therein [1]. Here we present a concise summary of this review, but provide also some complementary information.

2. Methodology

In deep-inelastic scattering, the nuclear structure function

$$F_2^A(x, Q^2) \;=\; \sum_i f_i^{(A,Z)}(x, Q^2) \otimes C_{2,i}(x, Q^2) \tag{1}$$

factorises into perturbatively calculable Wilson coefficients $C_{2,i}$, that are currently known at next-to-next-to-leading order (NNLO) and partially beyond [2], and nuclear PDFs

$$f_i^{(A,Z)}(x, Q^2) \;=\; \frac{Z}{A} f_i^{p/A}(x, Q^2) + \frac{A-Z}{A} f_i^{n/A}(x, Q^2), \tag{2}$$

which can be technically split into (unphysical) bound nucleon PDFs. Here, A and the Z denote the nuclear mass and charge. Evolution of the PDFs with Q^2 is governed by the DGLAP equations

$$\frac{\partial f_i(x, Q^2)}{\partial \log Q^2} \;=\; \int_x^1 \frac{dz}{z} P_{ij}\left(\frac{x}{z}, \alpha_s(Q^2)\right) f_j(z, Q^2) \tag{3}$$

with perturbative splitting functions P_{ij}. In addition, the PDFs have to satisfy number and momentum rules. Furthermore, for nuclear PDFs isospin symmetry

$$f_{d,u}^{n/A}(x, Q^2) \;\; = \;\; f_{u,d}^{p/A}(x, Q^2) \tag{4}$$

is usually assumed. Prior to BNL's RHIC and CERN's LHC, global nuclear PDF analyses relied on neutral-current (NC) and charged-current (CC) DIS and Drell-Yan (DY) FT data and still do so for nuclei other than Au (RHIC) and Pb (LHC). This is, however, about to change with Xe and O in the LHC and in particular with BNL's EIC [3].

Theoretically, the global nuclear PDF analyses nCTEQ15HQ [4], EPPS21 [5] and nNNPDF3.0 [6] have so far been performed only at NLO due to the unavailability of fast NNLO codes for processes other than DIS, DY and W- and Z-boson production, to which the KSASG20 [7] and TUJU21 [8] NNLO analyses were therefore restricted. In DGLAP evolution codes, heavy-quark (HQ) masses are usually taken into account with general-mass variable flavour number schemes such as FONLL, but extensions to the SACOT-χ scheme at NNLO including charged-current DIS are currently in progress [9]. Fitting of the experimental data is performed by optimising the χ^2 figure-of-merit function and propagation of the experimental uncertainties into the PDFs with the Hessian or Monte Carlo method. The validity of the Hessian approximation and the interpretation of the Monte Carlo uncertainty receive currently a lot of interest (cf. the contributions by M. Costantini, N. Derakhshanian, T. Giani and P. Risse at this workshop) [10].

Nuclear data taken with deuterons (D) in FT experiments and at RHIC must be corrected for Fermi motion, weak binding and off-shell effects, in particular at intermediate and large values of x (cf. the contributions by W. Henry and M. Cerutti) [11]. The dominant target mass corrections (TMCs) can be taken into account with the Nachtmann scaling variable $\xi = 2x/(1 + r)$, where $r = \sqrt{1 + 4x^2 M^2/Q^2}$, and they depend only weakly on A. Subleading TMCs contribute at HT (cf. the contribution by R. Ruiz) [12]. At high x and low Q values, HT corrections can generally become important and can be parameterised and fitted additively or multiplicatively (cf. the contributions by M. Cerutti and R. Petti). Since many neutrino DIS data points lie in this kinematic region, all of these corrections have to be taken into account there [13]. Unfortunately, they cannot fully resolve the long-standing tension of the CC with the NC DIS data, so that in practice only a subset of neutrino data is used for the important separation of quark flavours (cf. the contributions by J. Rojo and S. Yrjänheikki) [14]. Much hope lies therefore on future neutrino DIS data from the FASERν and SND@LHC experiments at the LHC Forward Physics Facility (cf. the contributions by J. Atkinson and O. Durhan). With electroweak boson, isolated photon, hadron and jet as well as heavy-quark and quarkonium data, the four main LHC experiments have already extended the kinematic plane by orders of magnitude and down to values of $x < 10^{-5}$ and up to $Q^2 > 10^5$ GeV^2.

3. Electroweak bosons

The analysis of *W*- and *Z*-boson production in pPb collisions at LHC Runs I and II has been led by CMS and ALICE and is now seconded by LHCb and ATLAS. The importance of these data lies in their potential sensitivity to flavour separation and the strange quark, which in practice is, however, smaller than naively expected due to the radiative generation of strange quarks from gluon splittings [15]. DY and W-/Z-boson production thus also impose constraints on the gluon
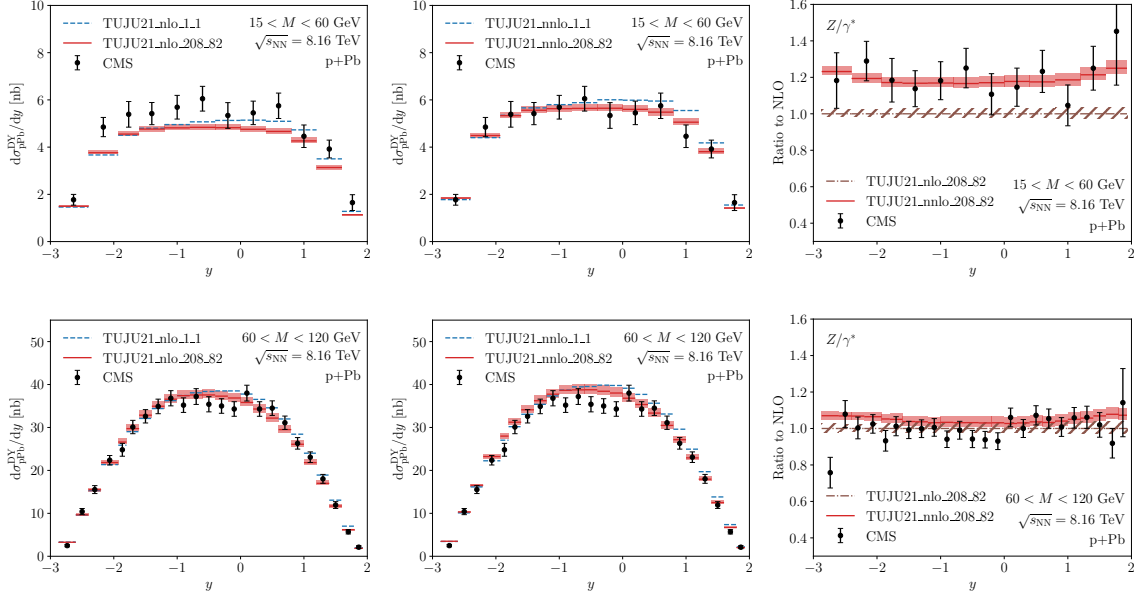
Figure 1: Comparison of DY production in pPb collisions at $\sqrt{s}$ = 8.16 TeV at NLO (left) and NNLO (center) results with (solid with uncertainty band) and without (dashed) nuclear PDF modifications in two invariant mass bins, $15 < M < 60$ GeV (upper panels) and $60 < M < 120$ GeV (lower panels) to CMS data [20]. In the right part we plot the ratios of the NNLO (red with uncertainty) and NLO (dot-dashed brown with hatched uncertainty) together with the data. Taken from Ref. [8].

density [16–19]. Particularly precise are the Run-II Z-boson data from CMS shown in Fig. 1 for two invariant-mass windows and as a function of the rapidity y of the dilepton pair [20]. A comparison of these CMS data with NLO and NNLO calculations using TUJU21 PDFs shows that the NNLO corrections are rather mild (around 5%) for the high-mass bin, but become significant for the low-mass bin, reaching 20% at the largest (absolute) rapidities [8]. Unfortunately, large fluctuations at mid-rapidity make it difficult to have acceptable χ^2-values even at NNLO. In contrast, the CMS Run-II W-boson data [21] have been used in TUJU21 and all three global NLO analyses, with nCTEQ15HQ and nNNPDF3.0 fitting absolute cross sections and EPPS21 ratios of cross sections.

4. Photons, hadrons and jets

Prompt photon production is well-known to be sensitive to the gluon density through the QCD Compton process [22]. Prior to the LHC, prompt photon data have been taken by FNAL's FT experiment E706 in pBe and by the PHENIX and STAR experiments at RHIC in DAu collisions. The only published LHC pPb cross sections, taken differentially in the photon transverse energy (E_T^γ) and for three photon pseudo-rapidity (η_{CM}) bins at $\sqrt{s}$ = 8.16 TeV, come from ATLAS [23]. Their ratios to ATLAS pp cross sections, taken with the same isolation criterion and extrapolated from $\sqrt{s}$ = 8 TeV [24], are shown in Fig. 2. The data are compared with NLO predictions based on nNNPDF3.0 PDFs fitted with and without LHCb D-meson data. The bands in the theory predictions indicate the PDF uncertainty. For pPb/pp ratios, a good description of the data can be achieved at NLO, and they have therefore been included in the nNNPDF3.0 fit [6]. However, the absolute cross
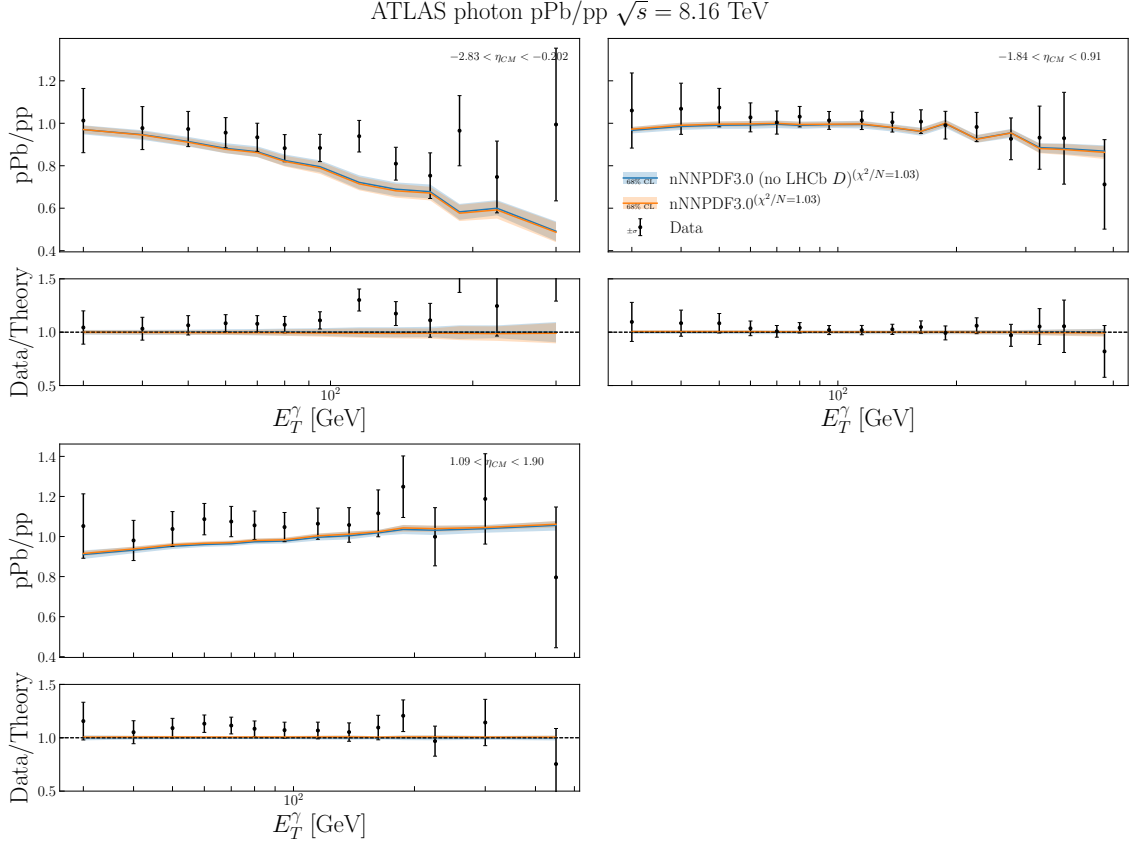
Figure 2: Comparison between ATLAS pPb/pp prompt photon cross section ratios [23] and theoretical predictions based on nNNPDF3.0 PDFs fitted with and without LHCb D-meson data, differential in the photon transverse energy (E_T^γ) and for three photon pseudo-rapidity (η_{CM}) bins. The bands in the theory predictions indicate the PDF uncertainty. The bottom panels display the ratio to the central theory. The values of χ^2 to this dataset for both nNNPDF3.0 fits are indicated in the legend. Taken from Ref. [6].

sections are underestimated at NLO, which may indicate a need for NNLO precision. Preliminary ALICE data [25] are consistent with the ATLAS measurements, but extend them to lower E_T^γ and thus lower x values in the nuclear PDFs, so that their publication is eagerly awaited.

Single inclusive hadron (SIH, specifically π^0) production at RHIC was one of the first processes used to impose constraints on the gluon density. In the meantime, many more data sets, also on charged pions, kaons, and η mesons, have become available from PHENIX and STAR at RHIC and in particular ALICE at the LHC and have been used to extend the original nCTEQ15 analysis [26] to nCTEQ15SIH [27]. As the SIH calculation depends on the fragmentation of the final parton into the observed hadron, a variety of fragmentation functions available in the literature was used to properly estimate this source of uncertainty. This was not possible for η mesons, which, however, also had only a minor impact on the nCTEQ15SIH analysis.

Hadronic jets are less sensitive to details of the fragmentation process than single inclusive hadrons. In nuclear collisions one must, however, pay particular attention to the background from the so-called underlying event, which is much larger than in pp collisions and has been estimated to involve on average 7 ± 5 pN interactions [28]. Multiparton scatterings are therefore usually modelled
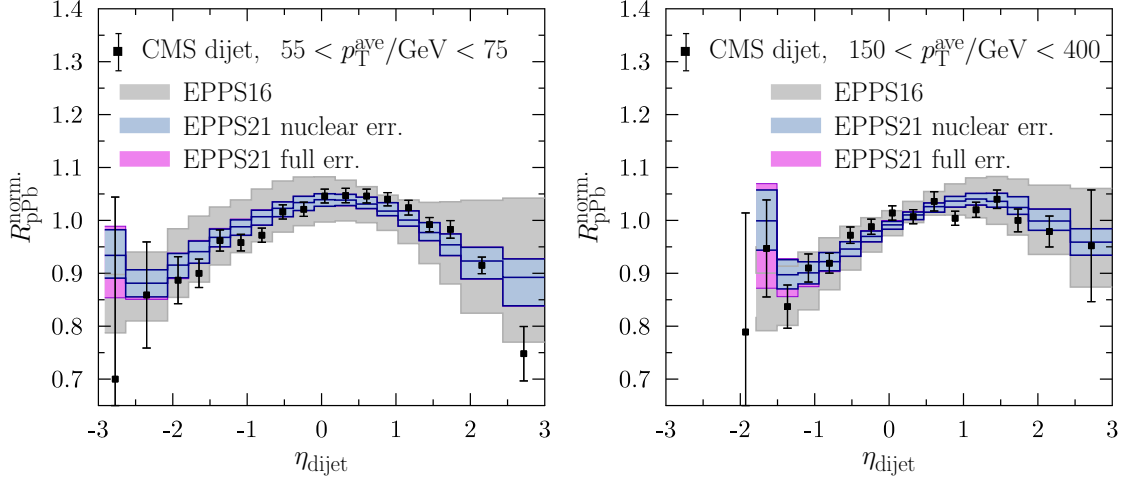
Figure 3: The CMS dijet data [29] compared with the EPPS21 analysis. The solid blue lines show the central results, inner blue bands the nuclear uncertainties, and the purple bands the total uncertainty. The grey bands correspond to the EPPS16 results [30]. Taken from Ref. [5].

by Monte Carlo simulations and subtracted from the experimental data. For further reduction of the associated uncertainty, one resorts to sufficiently large transverse momenta and small jet cones. Currently, the most constraining data have been taken in LHC Run I by CMS on dijets differential in the average p_T (p_T^{ave}) and rapidity (η_{dijet}) of the two jets [29]. They supersede earlier dijet data, which were included in the EPPS16 analysis [30]. The cross sections are normalised to the rapidity-integrated cross section, so that most of the systematic uncertainties cancel. The resulting spectra in pp are then so precise that they challenge the theoretical description, the NLO perturbative QCD calculations being in tension with the data [31]. The impact of the NNLO corrections [32] is still unclear. Despite this tension, ratios of normalised cross sections between pPb and pp collisions, $R_{\mathrm{pPb}}^{\mathrm{norm.}}$, are broadly consistent with nuclear PDFs. They have therefore been included in the EPPS21 [5] and nNNPDF3.0 [6] global fits, where they have a large impact on the gluon. Fig. 3 shows the large reduction of the nuclear uncertainty in the forward region from EPPS16 (grey) to EPPS21 (blue), while in the backward region the gluon uncertainty in the underlying CT18A proton PDFs [33] adds to the full error (purple). However, the most forward data points at the edge of the detector acceptance indicate a suppression, which cannot be fitted.

5. Heavy quarks and quarkonia

During the last decade, the four LHC experiments and in particular ALICE and LHCb have taken vast data sets on open heavy quark and quarkonium production. They cover a wide kinematic range and put strong constraints on the nuclear gluon PDF down to $x \leq 10^{-5}$. Theoretical predictions for these data sets can be obtained from a data-driven approach, where pp data are used to determine effective scattering matrix elements. Including an explicit rapidity dependence in the ansatz allows to describe also the very forward LHCb data. This approach has been validated with detailed comparisons to existing NLO calculations in non-relativistic QCD (NRQCD) for quarkonia and in
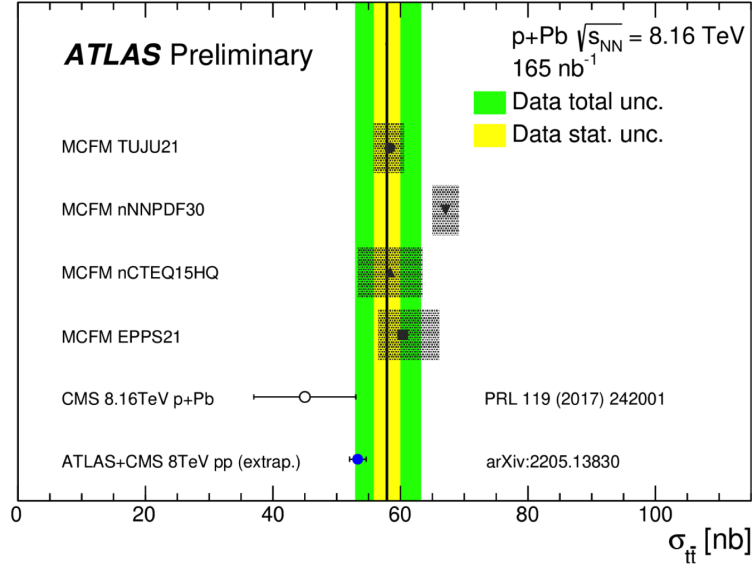
Figure 4: Comparison of the observed top quark cross section in ATLAS [35] with the corresponding CMS measurement [36] at $\sqrt{s}$ = 8.16 TeV. Also shown is the combined ATLAS and CMS measurement in pp collisions at $\sqrt{s}$ = 8 TeV, extrapolated to 8.16 TeV. Theoretical predictions are calculated at NNLO for different nuclear PDF sets with associated PDF uncertainties (grey). The central ATLAS measurement is indicated by the black line, the statistical uncertainty and its combination with the systematic uncertainty by yellow and green shaded areas. Taken from Ref. [35].

the general-mass variable-flavor-number scheme (GM VFNS) for the open heavy-flavoured mesons (cf. the contribution by J. Wissmann). Inclusion of these data has led to an enormous reduction in the gluon uncertainty in the nCTEQ15HQ analysis [4]. In addition, the uncertainties from the data-driven approach have been determined using the Hessian method and accounted for in the PDF fits (cf. the contribution by T. Jezo). One of the most recent LHC measurements in pPb collisions is the observation of top quark production by ATLAS [35] and CMS [36]. The measured cross sections are compared to the PDFs discussed in this contribution in Fig. 4 and show for most of them an impressive agreement, even though these data were not included in the fits.

6. Conclusion

To summarise, Fig. 5 compares the nuclear modifications of the lead nucleus PDFs at $Q^2 = 10$ GeV^2 from EPPS21 (full, blue) [5], nCTEQ15HQ (dashed, red) [4] and nNNPDF3.0 (dot-dashed, green) [6]. Qualitatively, there is good overall agreement between all three within the 90% CL uncertainty bands (shaded areas). Closer inspection reveals nevertheless still significant differences both among the central values and the widths of the uncertainty bands in several distributions and x regions. The vastly expanding measurements at the LHC and steady theoretical progress, including recent advances in lattice QCD calculations (cf. the contribution by H.W. Lin), will ensure continuous progress in the important dynamic field of nuclear PDFs over the next years, as we eagerly await the commissioning of BNL's EIC.
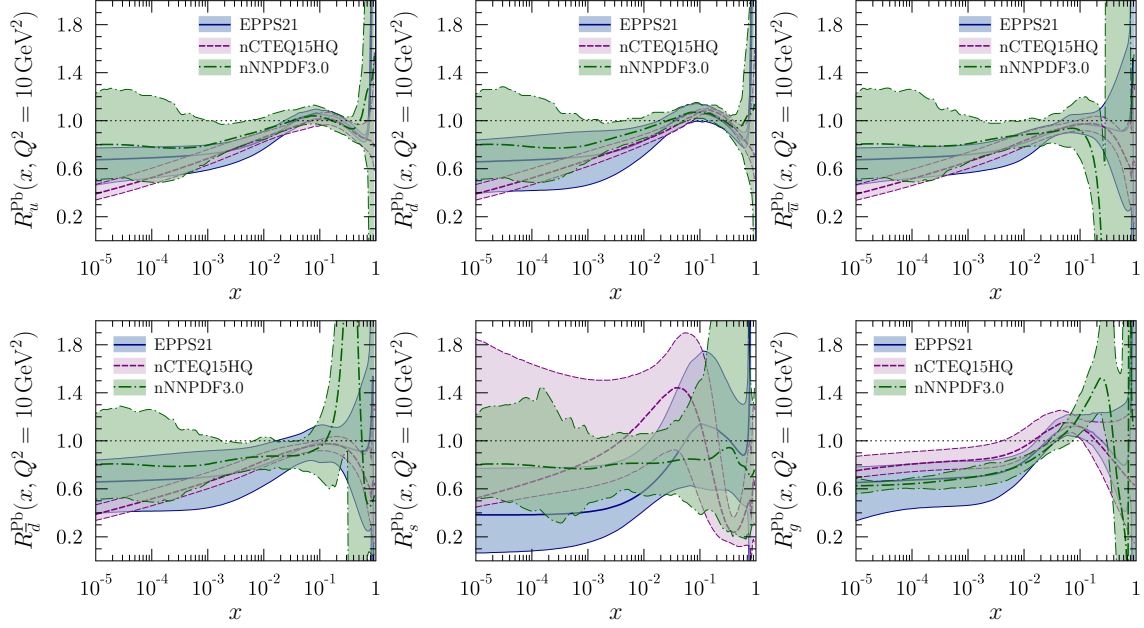
Figure 5: Comparison of the ^{208}Pb nuclear modifications resulting from the EPPS21 (full, blue) [5], nCTEQ15HQ (dashed, red) [4] and nNNPDF3.0 (dot-dashed, green) [6] global analyses of nuclear PDFs, i.e. the PDFs of lead divided by the summed PDFs of 82 free protons and 126 free neutrons. Uncertainty bands correspond to 90% CL. Taken from Ref. [1].

Acknowledgment

References

[1] M. Klasen and H. Paukkunen, *Nuclear PDFs after the first decade of LHC data*, doi:10.1146/annurev-nucl-102122-022747, arXiv:2311.00450 [hep-ph].

[2] G. Falcioni, F. Herzog, S. Moch, A. Pelloni and A. Vogt, *Four-loop splitting functions in QCD - the quark-to-gluon case*, arXiv:2404.09701 [hep-ph].

[3] R. Abdul Khalek *et al.*, *Snowmass 2021 White Paper: Electron Ion Collider for High Energy Physics*, arXiv:2203.13199 [hep-ph].

[4] P. Duwentäster *et al.*, *Impact of heavy quark and quarkonium data on nuclear gluon PDFs*, Phys. Rev. D **105** (2022) 114043.

[5] K. J. Eskola, P. Paakkinen, H. Paukkunen and C. A. Salgado, *EPPS21: A global QCD analysis of nuclear PDFs*, Eur. Phys. J. C **82** (2022) 413.

[6] R. Abdul Khalek *et al.*, *nNNPDF3.0: evidence for a modified partonic structure in heavy nuclei*, Eur. Phys. J. C **82** (2022) 507.

[7] H. Khanpour *et al.*, *Nuclear parton distribution functions with uncertainties in a general mass variable flavor number scheme*, Phys. Rev. D **104** (2021) 034010.

[8] I. Helenius, M. Walt and W. Vogelsang, *NNLO nuclear parton distribution functions with electroweak-boson production data from the LHC*, Phys. Rev. D **105** (2022) 094031.

[9] P. Risse *et al.*, *Fast evaluation of heavy-quark contributions to DIS in APFEL++*, arXiv:2307.08269 [hep-ph].

[10] N. Derakhshanian *et al.*, *Nuclear PDF determination using Markov Chain Monte Carlo methods*, Acta Phys. Polon. Supp. **16** (2023) 7-A33.

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

[12] R. Ruiz *et al.*, *Target mass corrections in lepton-nucleus DIS: Theory and applications to nuclear PDFs*, Prog. Part. Nucl. Phys. **136** (2024) 104096.

[13] E. P. Segarra *et al.*, *Extending nuclear PDF analyses into the high-x , low-Q^2 region*, Phys. Rev. D **103** (2021) 114015.

[14] K. F. Muzakka *et al.*, *Compatibility of neutrino DIS data and its impact on nuclear parton distribution functions*, Phys. Rev. D **106** (2022) 074004.

[15] A. Kusina *et al.*, *Impact of LHC vector boson production in heavy ion collisions on strange PDFs*, Eur. Phys. J. C **80** (2020) 968.

[16] E. L. Berger, L. E. Gordon and M. Klasen, *Massive lepton pairs as a prompt photon surrogate*, Phys. Rev. D **58** (1998) 074012.

[17] M. Klasen and M. Brandt, *Parton densities from LHC vector boson production at small and large transverse momenta*, Phys. Rev. D **88** (2013) 054002.

[18] M. Brandt, M. Klasen and F. König, *Nuclear parton density modifications from low-mass lepton pair production at the LHC*, Nucl. Phys. A **927** (2014) 78.

[19] A. Andronic, T. Ježo, M. Klasen, C. Klein-Bösing and A. P. Neuwirth, *Di-electron production at the LHC: unravelling virtual-photon and heavy-flavour contributions*, JHEP **05** (2024) 222.

[20] A. M. Sirunyan *et al.* [CMS], *Study of Drell-Yan dimuon production in proton-lead collisions at $\sqrt{s_{\mathrm{NN}}}$ = 8.16 TeV*, JHEP **05** (2021) 182.

[21] A. M. Sirunyan *et al.* [CMS], *Observation of nuclear modifications in $W^\pm$ boson production in pPb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 8.16 TeV*, Phys. Lett. B **800** (2020) 135048.

[22] M. Klasen, C. Klein-Bösing and H. Poppenborg, *Prompt photon production and photon-jet correlations at the LHC*, JHEP **03** (2018) 081.

[23] M. Aaboud *et al.* [ATLAS], *Measurement of prompt photon production in* $\sqrt{s_{NN}}$ = 8.16 *TeV p+Pb collisions with ATLAS*, Phys. Lett. B **796** (2019) 230.

[24] G. Aad *et al.* [ATLAS], *Measurement of the inclusive isolated prompt photon cross section in pp collisions at* $\sqrt{s}$ = 8 *TeV with the ATLAS detector*, JHEP **08** (2016) 005.

[25] F. Jonas, talk presented at Hard Probes 2023, Aschaffenburg (ALI-PREL-538803).

[26] K. Kovarik *et al.*, *nCTEQ15 - global analysis of nuclear parton distributions with uncertainties in the CTEQ framework*, Phys. Rev. D **93** (2016) 085037.

[27] P. Duwentäster *et al.*, *Impact of inclusive hadron production data on nuclear gluon PDFs*, Phys. Rev. D **104** (2021) 094005.

[28] C. Loizides, J. Kamin and D. d'Enterria, *Improved Monte Carlo Glauber predictions at present and future nuclear colliders*, Phys. Rev. C **97** (2018) 054910 [erratum: Phys. Rev. C **99** (2019) 019901].

[29] A. M. Sirunyan *et al.* [CMS], *Constraining gluon distributions in nuclei using dijets in proton-proton and proton-lead collisions at* $\sqrt{s_{NN}}$ = 5.02 *TeV*, Phys. Rev. Lett. **121** (2018) 062002.

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

[31] K. J. Eskola, P. Paakkinen and H. Paukkunen, *Non-quadratic improved Hessian PDF reweighting and application to CMS dijet measurements at 5.02 TeV*, Eur. Phys. J. C **79** (2019) 511.

[32] J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss and J. Pires, *Precise predictions for dijet production at the LHC*, Phys. Rev. Lett. **119** (2017) 152001.

[33] T. J. Hou, J. Gao, T. J. Hobbs, K. Xie, S. Dulat, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin and C. Schmidt, *et al. New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, Phys. Rev. D **103** (2021) 014013.

[34] A. Kusina, J. P. Lansberg, I. Schienbein and H. S. Shao, *Reweighted nuclear PDFs using heavy-flavor production data at the LHC*, Phys. Rev. D **104** (2021) 014010.

[35] M. Aaboud *et al.* [ATLAS], *Observation of* $t\bar{t}$ *production in lepton+jets and dilepton channels in p+Pb collisions at* $\sqrt{s_{NN}}$ = 8.16 TeV *with the ATLAS detector*, ATLAS-CONF-2023-063.

[36] A. M. Sirunyan *et al.* [CMS], *Observation of top quark production in proton-nucleus collisions*, Phys. Rev. Lett. **119** (2017) 242001.