JYU DISSERTATIONS 115

Petja Paakkinen

New Constraints for Nuclear Parton Distribution Functions from Hadron–Nucleus Collision Processes

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF MATHEMATICS
AND SCIENCE

JYU DISSERTATIONS 115

Petja Paakkinen

New Constraints for Nuclear Parton Distribution Functions from Hadron–Nucleus Collision Processes

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ylistönrinteen salissa FYS1
syyskuun 6. päivänä 2019 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in Ylistönrinne, auditorium FYS1, on September 6, 2019 at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2019

Editors
Hannu Paukkunen
Department of Physics, University of Jyväskylä
Timo Hautala
Open Science Centre, University of Jyväskylä

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-7828-0

ISBN 978-951-39-7828-0 (PDF)
URN:ISBN:978-951-39-7828-0
ISSN 2489-9003

Abstract

Paakkinen, Petja
New constraints for nuclear parton distribution functions from hadron–nucleus collision processes
Jyväskylä: University of Jyväskylä, 2019
(JYU Dissertations
ISSN 2489-9003, 115)
ISBN 978-951-39-7828-0

This work studies collinearly factorizable nuclear parton distribution functions (nPDFs) in perturbative Quantum Chromodynamics (QCD) at next-to-leading order in the light of hadron–nucleus collision data which have not been included in nPDF analyses previously. The aim is at setting new constraints on the nuclear modifications of the gluon distribution and on the flavour separation of quark nuclear modifications. The introductory part provides an outline of the theoretical framework of QCD collinear factorization and the used statistical methods and relates the work presented here to other similar contemporary analyses.

As a result, a new set of nPDFs, EPPS16, is presented, including for the first time electroweak-boson and dijet production data from CERN-LHC proton–lead collisions and allowing a full flavour separation in the fit. The flavour separation is constrained with Drell–Yan dilepton-production data from fixed target pion–nucleus experiments and neutrino–nucleus deep-inelastic scattering data, which are shown to give evidence for the similarity of the u and d valence-quark nuclear modifications. For studying the gluon degrees of freedom, collider data are essential and in the EPPS16 analysis new constraints are derived from the dijet production at the LHC.

Possible further constraints for the gluons are investigated in terms of the LHC data on nuclear modification ratios of dijet and D-meson production. Using a non-quadratically improved Hessian reweighting method, these measurements are found to put stringent constraints on the gluon modifications in the lead nucleus, reaching smaller values of the nucleon momentum fraction than previously accessible. A study on the future prospects of constraining nPDFs within a multi-observable approach with the BNL-RHIC is also given.

Author Petja Paakkinen
Department of Physics
University of Jyväskylä
Finland

Supervisors Doc. Hannu Paukkunen
Department of Physics
University of Jyväskylä
Finland

Prof. Kari J. Eskola
Department of Physics
University of Jyväskylä
Finland

Reviewers Dr. Urs A. Wiedemann
Theoretical Physics Department
CERN
Switzerland

Prof. Néstor Armesto
Departamento de Física de Partículas and IGFAE
Universidade de Santiago de Compostela
Spain

Opponent Prof. Juan Rojo
Department of Physics and Astronomy
Vrije Universiteit Amsterdam
The Netherlands

Preface

The research presented in this thesis has been carried out at the University of Jyväskylä during the period from November 2015 to August 2019. In addition to the University of Jyväskylä, the work has been funded by the Magnus Ehrnrooth Foundation and during the very final stages of the work also by the Academy of Finland, Project 308301. The Academy of Finland, Project 297058, is also acknowledged for the generous travel funding which has enabled the author to present the results of this work at various scientific meetings around the world.

My foremost thanks go to my supervisors Doc. Hannu Paukkunen and Prof. Kari J. Eskola. It was in Kari's Particle Physics course where I first got introduced to the concept of parton distribution functions. His enthusiasm on the topic was, least to say, contagious, so that when I was asked whether I wanted to work on the subject, I did not need to think twice. I am thankful for his guidance and support ever since. I have come to know Hannu as one of the leading younger researchers in the field. I feel grateful for the encouraging and collaborative manner with which he has overseen my progress. I wish to thank also my collaborators Prof. Carlos A. Salgado, Dr. Ilkka Helenius, Prof. John Lajoie and Dr. Joseph D. Osborn for their valuable contributions to the articles presented in this thesis. I thank Dr. Urs A. Wiedemann and Prof. Néstor Armesto for reviewing the manuscript and Prof. Juan Rojo for promising to act as the opponent at the public examination.

Many colleagues and friends, all of whom I cannot possibly name here, have had a less direct influence on the completion of this thesis. I feel I should thank the whole Jyväskylä QCD theory group, past and present, for the inspiring research environment that we have. The work would have been far less fun without the people at the office YFL 353 a.k.a. Holvi. Special thanks go to the latest generation of young aspiring scientists H. Hänninen, L. Jokiniemi, M. Kuha, T. Löytäinen and P. Pirinen. I also wish to thank Matias, Valtteri, Pekka and Juha, with whom I spent most of my undergraduate years, for the long-lasting friendship.

I thank my father, whose book collection gave me some of the first inspirations to study theoretical physics, my mother, who always encouraged me to study mathematical subjects, and my whole family for their unconditional support on all of my endeavours. Finally, with the utmost gratitude, I wish to thank Sanna for all her love.

Jyväskylä, August 2019 Petja Paakkinen

List of Publications

This thesis consists of an introductory part and of the following publications:

[I] **Applicability of pion–nucleus Drell–Yan data in global analysis of nuclear parton distribution functions**
Petja Paakkinen, Kari J. Eskola and Hannu Paukkunen, *Phys. Lett.* B768 (2017) 7–11, arXiv: 1609.07262 [nucl-th].

[II] **EPPS16: Nuclear parton distributions with LHC data**
Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen and Carlos A. Salgado, *Eur. Phys. J.* C77, no.3 (2017) 163, arXiv: 1612.05741 [hep-ph].

[III] **Non-quadratic improved Hessian PDF reweighting and application to CMS dijet measurements at 5.02 TeV**
Kari J. Eskola, Petja Paakkinen and Hannu Paukkunen, *Eur. Phys. J.* C79, no.6 (2019) 511, arXiv: 1903.09832 [hep-ph].

[IV] **Nuclear gluons at RHIC in a multi-observable approach**
Ilkka Helenius, John Lajoie, Joseph D. Osborn, Petja Paakkinen and Hannu Paukkunen, *Phys. Rev.* D100, no.1 (2019) 014004, arXiv: 1904.09921 [hep-ph].

[V] **A QCD analysis of LHCb D-meson data in p+Pb collisions**
Kari J. Eskola, Ilkka Helenius, Petja Paakkinen and Hannu Paukkunen, submitted to *JHEP*, arXiv: 1906.02512 [hep-ph].

The author performed all the calculations and wrote the first draft for the article **[I]**. For the article **[II]**, the author produced look-up tables for fast calculations of the cross-section ratios discussed in the article **[I]** and participated in writing of the manuscript. The author performed all the calculations presented in the article **[III]** except those for the 7 TeV proton–proton inclusive jet cross sections and wrote the first version of the manuscript and handled the revision process. For the articles **[IV]** and **[V]**, the author performed the reweighting analysis using the tools developed for the article **[III]** and participated in the preparation of the manuscripts.

Contents

Chapter 1

Introduction

The Standard Model of particle physics describes our present-day best knowledge of the fundamental particles of Nature and their electromagnetic, weak and strong interactions. It is a renormalizable quantum field theory with a local $\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)$ gauge symmetry. The $\mathrm{U}(1) \times \mathrm{SU}(2)$ symmetry, spontaneously broken through the Higgs mechanism, gives rise to the electroweak interactions, while the unbroken $\mathrm{SU}(3)$ symmetry dictates, in a theory called Quantum Chromodynamics (QCD), the strong interaction between particles carrying an $\mathrm{SU}(3)$ colour charge. What makes the strong interaction different from electroweak phenomena is the property of confinement: the strong force binds coloured particles, quarks and gluons, into colourless hadrons. We thus never observe freely propagating quarks and gluons, only the hadrons they constitute.

This poses a difficulty in the theoretical description of the QCD phenomena, as the asymptotic states are not the fundamental degrees of freedom of the theory. Fortunately, scattering processes involving a large momentum transfer factorize [1], i.e. the cross sections of these hard processes can be obtained by convoluting the scattering probabilities of the fundamental particles with long-distance functions describing their distributions in the involved hadrons. This makes it possible to study the distributions of partons, particles inside hadrons, by measurements of hard-process cross sections. These long-distance functions are called *parton distribution functions* (PDFs).

The PDFs are universal, independent of the scattering process, and hence distributions extracted from one process can be used to make predictions of another. It is not, however, possible to determine the PDFs of all different parton flavours independently from a single observable and instead large sets of data from different measurements are needed for their reliable extraction. This has led to the development of the field of PDF global analyses with ever-increasing precision in the obtained PDFs [2].

This thesis deals with PDFs of a particular kind, the nuclear PDFs (nPDFs), describing the partonic content of nucleons bound in nuclei. Even 20 years

after the pioneering works [3–6], the nPDFs carry large uncertainties. Until very recently, the possible asymmetry in nuclear modifications of different valence and sea-quark flavours has not been considered in the nPDF global analyses. Also the nuclear modifications of gluons, for which direct constraints have been scarce, have remained poorly known. These open problems are addressed in this thesis. In particular, we will discuss the impact of new constraints from hadron–nucleus collision processes which have not been used in nPDF global analyses previously. These include CERN-LHC proton–lead measurements of electroweak bosons, dijets and inclusive D^0-production, but also older pion–nucleus Drell–Yan measurements. Further, by considering prospects at present and upcoming experiments, we will try to pave the way towards a better understanding of the PDF nuclear modifications in the future. The work presented in this thesis is performed at the level of next-to-leading order perturbative QCD.

The introductory part is organized as follows. Chapter 2 introduces the theoretical framework of the thesis, the collinear factorization of QCD. The discussion here is rather minimal, with the aim at presenting the relevant concepts, but avoiding any unnecessary calculational details. In Chapter 3, the used statistical methods are presented. Emphasis is given to the treatment of correlated uncertainties, which will become increasingly important in the nPDF fits with precision data from the LHC becoming available. The novel physics results are discussed and compared to the results of earlier analyses in Chapter 4 and summarized in Chapter 5.

Chapter 2

Parton distributions in collinear factorization

Quantum Chromodynamics, the theory of strong interactions within the Standard Model, is characterized by its Lagrangian, defined in terms of the quark and gluon fields ψ_i and A^a_μ as [7, 8]

$$\mathcal{L} = \sum_i \bar{\psi}_i (\mathrm{i}\not{D} - m_i)\psi_i - \frac{1}{4} F^a_{\mu\nu} F^{a,\mu\nu}, \tag{2.1}$$

where the sum goes over the quark flavours with masses m_i and

$$D_\mu = \partial_\mu - \mathrm{i} g_\mathrm{s} A^a_\mu t^a, \tag{2.2}$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_\mathrm{s} f^{abc} A^b_\mu A^c_\nu \tag{2.3}$$

are the covariant derivative and the gluon field strength tensor, respectively. Here, t^a are the $\mathrm{SU}(3)$ generators in the fundamental representation and f^{abc} the structure constants. The Lagrangian is invariant in local $\mathrm{SU}(3)$ gauge transformations, which in the case of an infinitesimal shift θ can be written as

$$\psi \to \psi + \mathrm{i}\theta^a t^a \psi, \quad A^a_\mu \to A^a_\mu + \frac{1}{g_\mathrm{s}} \partial_\mu \theta^a + f^{abc} A^b_\mu \theta^c, \tag{2.4}$$

leading to conservation of the $\mathrm{SU}(3)$ colour charge.

The strength of the QCD interactions is set by the coupling g_s. In renormalizing the theory, this bare coupling must be traded with the running coupling, usually expressed in terms of $\alpha_\mathrm{s}(Q^2) = g_\mathrm{s}^2(Q^2)/4\pi$, which depends on the interaction scale Q^2 as

$$Q^2 \frac{\partial \alpha_\mathrm{s}}{\partial Q^2} = \beta(\alpha_\mathrm{s}), \tag{2.5}$$

with a negative beta function, $\beta < 0$. At low energies the coupling is large, permitting colour confinement, but towards higher scales the coupling gets

weaker, asymptotically approaching zero. This phenomenon is called *asymptotic freedom* [9, 10] and it allows one to use perturbation theory to calculate high-momentum-transfer cross sections in QCD.

2.1 Deep inelastic scattering in parton model

The easiest way to study experimentally the inner workings of hadrons is by deep inelastic scattering (DIS). In this process, illustrated in Figure 2.1, a lepton l with high energy E scatters off a hadron h, which then breaks apart into an inclusive final state X with a large invariant mass $W \gg M$, where M is the hadron mass. In the target rest frame (TRF), the square of the four-momentum transfer from the lepton to the hadron $q \overset{\text{TRF}}{=} (E - E', \boldsymbol{k} - \boldsymbol{k}')$, where $\boldsymbol{k}$, $\boldsymbol{k}'$ are the three-momenta of the initial and final state leptons, is given in terms of the energy E' and scattering angle θ of the final state lepton l' by

$$Q^2 := -q^2 \overset{\text{TRF}}{=} 2EE'(1 - \cos\theta). \tag{2.6}$$

The other relevant kinematical quantities for this process, the Bjorken x and the inelasticity y, are defined as

$$x := \frac{Q^2}{2P \cdot q} \overset{\text{TRF}}{=} \frac{Q^2}{2M(E - E')}, \qquad y := \frac{P \cdot q}{P \cdot k} \overset{\text{TRF}}{=} 1 - \frac{E'}{E}, \tag{2.7}$$

where $P \overset{\text{TRF}}{=} (M,0,0,0)$ is the four momentum of the hadron. In these Lorentz invariant variables, the unpolarized double-differential cross section can be expressed as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2 \mathrm{d}x} = \frac{4\pi\alpha_{\text{em}}^2}{Q^4} \frac{y^2}{2Q^2} L_{\mu\nu}(k,k') W^{\mu\nu}(P,q), \tag{2.8}$$

where $\alpha_{\text{em}} = e^2/4\pi$ is the electromagnetic fine-structure constant and $L_{\mu\nu}$, $W^{\mu\nu}$ refer to the leptonic and hadronic tensors, respectively.

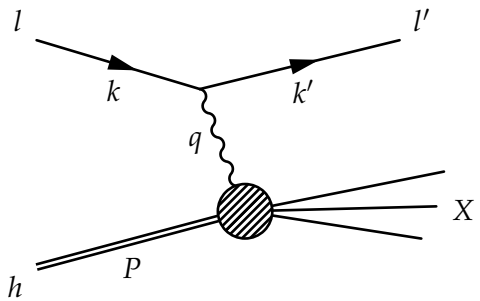

Figure 2.1. Deep inelastic lepton–hadron scattering.

In a neutral-current electromagnetic scattering mediated by a virtual photon, the leptonic tensor is simply

$$L_{\mu\nu}(k,k') = 2\,(k_\mu k'_\nu + k'_\mu k_\nu - k\cdot k' g_{\mu\nu}). \tag{2.9}$$

and, by conservation of current and the leptonic tensor being real and symmetric, the hadronic tensor can be expressed as [11]

$$\begin{aligned} W^{\mu\nu}(P,q) = &-\left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right) F_1(x,Q^2) \\ &+ \frac{1}{P\cdot q}\left(P^\mu - \frac{P\cdot q}{q^2}q^\mu\right)\left(P^\nu - \frac{P\cdot q}{q^2}q^\nu\right) F_2(x,Q^2), \end{aligned} \tag{2.10}$$

where the structure functions $F_{1,2}$ encode our ignorance of the hadron structure. In these terms, the cross section reads

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}x} = \frac{4\pi\alpha_{\mathrm{em}}^2}{Q^4}\frac{1}{x}\left\{xy^2F_1(x,Q^2) + \left(1 - y - x^2y^2\frac{M^2}{Q^2}\right)F_2(x,Q^2)\right\}. \tag{2.11}$$

Parton model

Now, let us consider the DIS in a frame where the hadron is moving very fast, e.g. the Breit frame, where assuming $Q^2 \gg M^2$, we can take

$$P \overset{\text{Breit}}{=} (Q/2x,0,0,Q/2x), \qquad q \overset{\text{Breit}}{=} (0,0,0,-Q), \qquad Q = \sqrt{Q^2}. \tag{2.12}$$

In such a frame the hadron is Lorentz contracted and the interaction times of its constituents are strongly dilated. During the short phase when the collision with the lepton takes place the hadron is thus "frozen" and the lepton can scatter incoherently from the individual partons. This description is the basis of the "naive" parton model [12, 13] giving the leading behaviour of the DIS cross section. In this picture, the partons move collinearly with the parent hadron and we can define

$f_i(\xi)$ = the probability density of finding a parton i within the hadron carrying a fraction ξ of the hadrons momentum.

In more formal terms the PDFs can be defined as operator expectation values, see Refs. [11, 14].

The photon couples only to electrically charged particles and hence at the leading order the hadronic tensor takes the form

$$W^{\mu\nu}(P,q) = \sum_{i=q,\bar{q}}\int\frac{\mathrm{d}\xi}{\xi}f_i(\xi)\,\hat{W}_i^{\mu\nu}(p,q) + \mathcal{O}\left(\frac{1}{Q^2}\right), \tag{2.13}$$

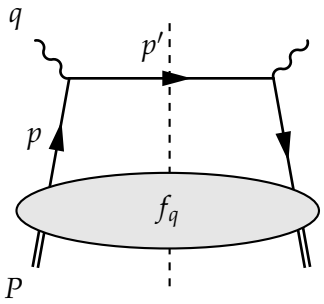

Figure 2.2. Parton-model picture of the hadronic tensor. Labels refer to the four-momenta of the particles.

where $p = \xi P \overset{\text{Breit}}{=} (\xi Q/2x, 0, 0, \xi Q/2x)$ and $\hat{W}_i^{\mu\nu}$ denotes a so-called partonic tensor. This can be pictorically represented as a "handbag" diagram, given in Figure 2.2, where the left-hand side of the cut represents the scattering amplitude and the right-hand side its complex conjugate. The antiquark contribution is obtained simply by changing the direction of the fermion line. At low scales, additional "higher-twist" contributions, essentially originating from multi-parton interactions, to the simple parton model picture, denoted by $\mathcal{O}\left(1/Q^2\right)$ in Eq. (2.13) can become important. At the clearly perturbative scales $Q^2 \gg M^2$ these should be negligible and we do not discuss them further here. In this leading order (LO), or "Born", approximation the quark-initiated partonic tensor is

$$\hat{W}^{\mu\nu}_{q,\text{Born}}(p,q) = \frac{x}{2Q^2}\frac{e_q^2}{2}\,\text{Tr}[\not{p}\gamma^\nu \not{n}\gamma^\mu]\,\delta(\xi - x), \tag{2.14}$$

where

$$n = q + xP \overset{\text{Breit}}{=} (Q/2, 0, 0, -Q/2), \qquad n^2 = 0, \tag{2.15}$$

and e_q^2 is the square of quark fractional charge. The delta function in Eq. (2.14) arises from integrating over the final state quark momentum p' and shows us that, to leading order perturbative accuracy, the Bjorken x measures the momentum fraction of the parton.

Now, using Eqs. (2.13) and (2.14), the differential cross section takes the form

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\text{LO}} = \sum_q e_q^2 f_q(x)\left(\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\text{Born}}, \tag{2.16}$$

where the sum is understood to be over both quarks and antiquarks and

$$\left(\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\text{Born}} = \frac{4\pi\alpha_{\text{em}}^2}{Q^4}\left\{\frac{y^2}{2} + \left(1 - y - x^2y^2\frac{M^2}{Q^2}\right)\right\}, \tag{2.17}$$

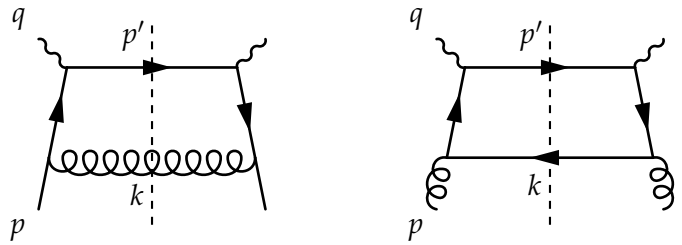

Figure 2.3. Ladder diagrams at order α_s. Left: Real-gluon emission from the initial-state quark. Right: Initial-state gluon splitting into a quark–antiquark pair.

or equivalently, if expressed in terms of the structure functions, Eq. (2.11), we have

$$2xF_1(x) = F_2(x) = x\sum_q e_q^2 f_q(x). \tag{2.18}$$

That is, in the naive parton model, the structure functions depend only on x and not on the scale Q^2. This phenomenon, called *Bjorken scaling*, is however broken by radiative corrections, as we will discuss next.

2.2 DGLAP evolution

The leading-order DIS cross section found in the previous section is subject to various radiative and virtual corrections at higher orders in perturbation theory. For massless partons, these corrections include collinear and soft divergences. The soft and final-state collinear divergences cancel at the level of summing over different contributions, but for the initial-state collinear divergences this cancellation is not complete. One can, however, resum these initial state divergences into the definitions of the PDFs, leading to the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution of the parton densities [15–18]. In the following, we present the general idea of how this is done. For more thorough discussions and calculational details the reader is directed to Refs. [19–23].

For the problem at hand, it is convenient to use the light-cone gauge [24, 25], where a gauge-fixing term $\mathcal{L}_{\text{gauge-fixing}} = -\frac{1}{2\lambda}(n^\mu A_\mu^a)^2$, with the limit $\lambda \to 0$ understood, is added to the Lagrangian in Eq. (2.1) and the gauge vector n is set to be that in Eq. (2.15). In this gauge, the only non-cancelling divergent contributions come from the "ladder"-type diagrams shown in Figure 2.3 (here as well, also the contributions obtained by reversing the fermion line need to be summed). These diagrams contain fermion propagators with denominators $(p-k)^2$, which diverge at the limit where k is collinear to p. Decomposing the momentum k à la Sudakov [26],

$$k^\mu = (1-z)p^\mu + \frac{\boldsymbol{k}_\perp^2}{1-z}\frac{n^\mu}{2p\cdot n} + k_\perp^\mu, \tag{2.19}$$

where $k_\perp \overset{\text{Breit}}{=} (0, \boldsymbol{k}_\perp, 0)$ is the component of momentum k transverse to p and n, we find that the contribution from the ladder diagram in Figure 2.3 (left) to the quark tensor is

$$\hat{W}^{\mu\nu}_{q,\text{Ladder}} = \frac{x}{2Q^2}\frac{e_q^2}{2}\,\text{Tr}[\not{p}\gamma^\nu \not{n}\gamma^\mu]\,\frac{\alpha_\text{s}}{2\pi}\int_x^1 \frac{\text{d}z}{z}\,C_\text{F}\left(\frac{1+z^2}{1-z}\right)\delta\left(\xi - \frac{x}{z}\right)\int_0^{Q^2}\frac{\text{d}\boldsymbol{k}_\perp^2}{\boldsymbol{k}_\perp^2} + \ldots, \tag{2.20}$$

with a colour factor $C_\text{F} = 4/3$ and where "…" refers to terms that do not contain a collinear divergence. The divergence thus appears as a logarithmic $\boldsymbol{k}_\perp$-integral, which we here regulate with a simple cut-off $\boldsymbol{k}_\perp^2 \geq m^2$, whereby the integral becomes

$$\int_{m^2}^{Q^2}\frac{\text{d}\boldsymbol{k}_\perp^2}{\boldsymbol{k}_\perp^2} = \log\left(\frac{Q^2}{m^2}\right). \tag{2.21}$$

Eq. (2.20) still contains a soft divergence at the limit $z \to 1$, but this cancels when we take into account the quark self-energy (SE) correction to the quark field-strength renormalization.

Combining the real and virtual corrections, we find the total divergent $\mathcal{O}(\alpha_\text{s})$ contribution to the hadronic tensor to be

$$W^{\mu\nu}_{\text{quark,Ladder+SE}} \overset{\text{LL}}{=} \frac{x}{4Q^2}\,\text{Tr}[\not{p}\gamma^\nu \not{n}\gamma^\mu]\sum_q e_q^2\,\frac{\alpha_\text{s}}{2\pi}\log\left(\frac{Q^2}{m^2}\right)\int_x^1\frac{\text{d}z}{z}\,P_{qq}(z)\,f_q\left(\frac{x}{z}\right), \tag{2.22}$$

where

$$P_{qq}(z) = C_\text{F}\left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z)\right] \tag{2.23}$$

is the leading-order *Altarelli–Parisi splitting function* for the quark-to-quark transition [18], with the plus distribution defined in terms of an integral equation

$$\int_0^1 \text{d}z\,\frac{f(z)}{(1-z)_+} = \int_0^1 \text{d}z\,\frac{f(z)-f(1)}{(1-z)} \tag{2.24}$$

Similarly, the divergent contribution from initial-state gluons given by the ladder diagram in Figure 2.3 (right) can be expressed as

$$W^{\mu\nu}_{\text{gluon,Ladder}} \overset{\text{LL}}{=} \frac{x}{4Q^2}\,\text{Tr}[\not{p}\gamma^\nu \not{n}\gamma^\mu]\sum_q e_q^2\,\frac{\alpha_\text{s}}{2\pi}\log\left(\frac{Q^2}{m^2}\right)\int_x^1\frac{\text{d}z}{z}\,P_{qg}(z)\,f_g\left(\frac{x}{z}\right), \tag{2.25}$$

where we have the leading-order gluon-to-quark splitting function

$$P_{qg}(z) = T_\text{F}\left((1-z)^2 + z^2\right), \qquad T_\text{F} = \frac{1}{2}. \tag{2.26}$$

In the above equations, we have denoted by "LL" that we are only considering the *leading logarithmic* contributions to the hadronic tensor. There are also further

contributions at order α_s (see e.g. Ref. [27]), but as these are of non-divergent nature, they are not important for the present discussion. Now, summing all the $\mathcal{O}(\alpha_s)$ leading logarithmic terms with the leading order expression, we have

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\mathrm{NLO}} \overset{\mathrm{LL}}{=} \sum_q e_q^2 \left\{ \left[1 + \frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{m^2}\right) P_{qq}\right] \otimes f_q(x) + \frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{m^2}\right) P_{qg} \otimes f_g(x) \right\} \left(\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\mathrm{Born}}, \tag{2.27}$$

where the symbol $\otimes$ above denotes a multiplicative convolution, defined as

$$h \otimes f(x) = \int_x^1 \frac{\mathrm{d}z}{z} h(z) f\left(\frac{x}{z}\right), \qquad 1 \otimes f(x) = \int_x^1 \frac{\mathrm{d}z}{z} \delta(1-z) f\left(\frac{x}{z}\right) = f(x). \tag{2.28}$$

We notice that the result in Eq. (2.27) is nothing but the parton-model cross section in Eq. (2.16) with the parton distribution $f_q(x)$ replaced with the term in curly braces. The collinear divergences thus *factorize* from the partonic process.

At this point, as we have seen that the collinear divergences occur when an internal quark gets on-shell and is thus allowed to propagate a long distance before the scattering with the virtual photon, it appears natural to redefine the PDFs as to include these long-distance effects. But before we do so, we have to note that similar collinear divergences can appear at *all* orders of perturbation theory, thus potentially spoiling this simple picture. The crucial thing here is that these divergent contributions exponentiate, and the DIS cross section can be written, in the leading-logarithm accuracy, as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}x} \overset{\mathrm{LL}}{=} \sum_q e_q^2 \begin{pmatrix}1 & 0\end{pmatrix} \exp\left[\frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{m^2}\right)\begin{pmatrix}P_{qq} & P_{qg}\\ P_{gq} & P_{gg}\end{pmatrix}\right] \otimes \begin{pmatrix}f_q\\ f_g\end{pmatrix}(x) \left(\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\mathrm{Born}}, \tag{2.29}$$

where the exponential convolution should be understood as

$$\exp[\mathcal{P}] \otimes f(x) = \sum_n \frac{1}{n!} \underbrace{\mathcal{P} \otimes \cdots \otimes \mathcal{P}}_{n \text{ times}} \otimes f(x) \tag{2.30}$$

and where we now also have the leading-order quark-to-gluon and gluon-to-gluon splitting functions

$$P_{gq}(z) = C_F\left(\frac{1+(1-z)^2}{z}\right), \tag{2.31}$$

$$P_{gg}(z) = 2C_A\left(\frac{1-z}{z} + \frac{z}{(1-z)_+} + z(1-z)\right) + \left(\frac{11}{6}C_A - \frac{2}{3}n_f T_F\right)\delta(1-z),$$

where $C_A = 3$ and n_f is the number of active quark flavours. Hence it makes sense to define *scale-dependent* parton distribution functions as

$$\begin{pmatrix} f_q(x,Q^2) \\ f_g(x,Q^2) \end{pmatrix} := \exp\left[\frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{m^2}\right)\begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix}\right] \otimes \begin{pmatrix} f_q \\ f_g \end{pmatrix}(x)\,, \tag{2.32}$$

from where, by taking the Q^2 derivative, we find the *Dokshitzer–Gribov–Lipatov–Altarelli–Parisi* (DGLAP) *evolution equations* [15–18],

$$Q^2\frac{\partial}{\partial Q^2}\begin{pmatrix} f_q(x,Q^2) \\ f_g(x,Q^2) \end{pmatrix} \overset{\text{LL}}{=} \frac{\alpha_s}{2\pi}\begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} f_q(Q^2) \\ f_g(Q^2) \end{pmatrix}(x)\,. \tag{2.33}$$

Now, substituting the definition in Eq. (2.32) back to Eq. (2.29), the physical predictions become independent of the collinear regulator and in this *QCD-improved* parton model, the full, *finite*, leading-order + leading-logarithm DIS cross section reads

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\text{LO+LL}} = \sum_q e_q^2 f_q(x,Q^2)\left(\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}Q^2\mathrm{d}x}\right)_{\text{Born}}, \tag{2.34}$$

with the PDF scale evolution governed by Eq. (2.33).

2.3 Factorization schemes and scales

In the discussion above, we have only considered the leading logarithmic contributions. At higher orders, $\alpha_s^{n+1}\log^n(Q^2/m^2)$, etc., both the partonic cross sections after the extraction of divergences (or *coefficient functions*) and the splitting functions get perturbative corrections [27–29]. Moreover, the definition given in Eq. (2.32) is not unique, leading to *scheme dependence* of the PDFs and of the splitting and coefficient functions [30]. Any physical predictions are still independent of the scheme to the perturbative order to which they have been calculated. To elaborate this more, let us write here the full NLO expression of the structure function F_2 as

$$\begin{aligned} F_2^{\text{NLO}} = x\sum_q e_q^2 \Bigg\{ &\left[1 + \frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{m^2}\right)P_{qq} + R_q\right] \otimes f_q(x) \\ &+ \left[\frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{m^2}\right)P_{qg} + R_g\right] \otimes f_g(x)\Bigg\} \end{aligned} \tag{2.35}$$

where $R_{q,g}$ denote the remainder parts including all the non-divergent $\mathcal{O}(\alpha_s)$ terms which we neglected in the discussion leading to Eq. (2.27). While we must

include the large logarithms to the redefinitions of the PDFs, nothing prevents us from including also some of the finite parts. Defining now

$$f_i(x,Q^2) := \sum_j \left[\delta_{ij} + \frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{m^2}\right)P_{ij} + f_{ij}^{\text{scheme}}\right] \otimes f_j(x) + \mathcal{O}\left(\alpha_s^2\right), \tag{2.36}$$

we can write

$$F_2^{\text{NLO}} = x\sum_q e_q^2 \{ \underbrace{\left[1 + R_q - f_{qq}^{\text{scheme}}\right]}_{=:C_q^{\text{scheme}}} \otimes f_q(Q^2)(x) + \underbrace{\left[R_g - f_{qg}^{\text{scheme}}\right]}_{=:C_g^{\text{scheme}}} \otimes f_g(Q^2)(x) \}, \tag{2.37}$$

where $C_{q,g}^{\text{scheme}}$ are now the NLO coefficient functions in the chosen scheme.

In a similar fashion, one can also choose to define the PDFs at some *factorization scale* Q_f different from Q, including the remaining $\log(Q^2/Q_f^2)$ terms in the coefficient functions,

$$\begin{aligned} F_2^{\text{NLO}} = x\sum_q e_q^2 \Bigg\{ & \overbrace{\left[1 + \frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{Q_f^2}\right)P_{qq} + R_q - f_{qq}^{\text{scheme}}\right]}^{=:C_q^{\text{scheme}}(Q^2/Q_f^2)} \otimes f_q(Q_f^2)(x) \\ & + \underbrace{\left[\frac{\alpha_s}{2\pi}\log\left(\frac{Q^2}{Q_f^2}\right)P_{qg} + R_g - f_{qg}^{\text{scheme}}\right]}_{=:C_g^{\text{scheme}}(Q^2/Q_f^2)} \otimes f_g(Q_f^2)(x) \Bigg\}. \end{aligned} \tag{2.38}$$

Again the different scale choices are formally equivalent up to corrections of one order of α_s higher. For this property, it is possible to estimate the uncertainties arising from the termination of the perturbative expansion by calculating so-called scale uncertainties through variations of the indefinite scales of the process.

While for the purpose of demonstrating the appearance of collinear divergences and their resummation in the discussion above it was useful to work in four spacetime dimensions and use cut-off regulators, it is more common in practical calculations to use dimensional regularization [31], which does not break any symmetries of the theory. In the dimensional regularization the spacetime is continued to $4-2\varepsilon$ dimensions and the collinear divergences now appear as poles at $\varepsilon = 0$. The NLO F_2 structure function takes in this case the form

$$\begin{aligned} F_2^{\text{NLO}} = x\sum_q e_q^2 \Bigg\{ & \left[1 + \frac{\alpha_s}{2\pi}\left(-\frac{1}{\hat{\varepsilon}} + \log\left(\frac{Q^2}{\mu^2}\right)\right)P_{qq} + R_q\right] \otimes f_q(x) \\ & + \left[\frac{\alpha_s}{2\pi}\left(-\frac{1}{\hat{\varepsilon}} + \log\left(\frac{Q^2}{\mu^2}\right)\right)P_{qg} + R_g\right] \otimes f_g(x) \Bigg\}, \end{aligned} \tag{2.39}$$

where μ^2 is an arbitrary scale needed in the dimensional regularization in order to keep the coupling dimensionless, and $1/\hat{\varepsilon} = 1/\varepsilon - \gamma_{\mathrm{E}} + \log(4\pi)$ with γ_{E} being the Euler–Mascheroni constant. This suggests the use of

$$f_i(x, Q_f^2) \coloneqq \sum_j \left[\delta_{ij} + \frac{\alpha_{\mathrm{s}}}{2\pi} \left(-\frac{1}{\hat{\varepsilon}} + \log\left(\frac{Q_f^2}{\mu^2} \right) \right) P_{ij} \right] \otimes f_j(x) + \mathcal{O}\left(\alpha_{\mathrm{s}}^2\right), \qquad (2.40)$$

defining the modified minimal subtraction $\overline{\mathrm{MS}}$ scheme [32, 33], which is also the scheme employed in this thesis. The structure function can be expressed as

$$F_2^{\mathrm{NLO}} = x \sum_q e_q^2 \left\{ C_q^{\overline{\mathrm{MS}}}(Q^2/Q_f^2) \otimes f_q(Q_f^2)(x) + C_g^{\overline{\mathrm{MS}}}(Q^2/Q_f^2) \otimes f_g(Q_f^2)(x) \right\}, \qquad (2.41)$$

with the $\overline{\mathrm{MS}}$ coefficient functions $C_{q,g}^{\overline{\mathrm{MS}}}$ available e.g. in Ref. [33]. To be exact, in the global PDF analysis presented in this thesis, we take the DIS and other partonic cross sections to NLO accuracy, and evolve the PDFs according to DGLAP equations using NLO splitting functions [28, 29].

2.4 Heavy-quark PDFs

So far we have treated all partons as massless, but for heavy quarks, particularly charm and bottom, with their masses in the GeV range, this is not always justifiable [34]. When the energy of the process is not high enough to produce heavy quarks, they should simply not contribute to the cross section. Above the mass threshold, the heavy-quark production becomes possible, in DIS through the partonic processes like the one shown in Figure 2.3 (right). Here, the heavy-quark mass m_{H} regulates the $\boldsymbol{k}_\perp$-integrals and thus the cross section remains finite. At very high scales $Q \gg m_{\mathrm{H}}$, however, the resulting logarithms $\log(Q^2/m_{\mathrm{H}}^2)$ become very large and their resummation into heavy-quark PDFs becomes necessary. How to interpolate between the extremes of high ($Q \gg m_{\mathrm{H}}$) and low ($Q \sim m_{\mathrm{H}}$) scales is, again, scheme dependent.

In the simplest *zero-mass variable flavour number scheme* (ZM-VFNS), one treats the heavy quark above the threshold as a massless active parton, using the same massless coefficient functions as for the n_f light quarks with now $n_f + 1$ flavours participating in the DGLAP evolution. This, however, ignores the mass effects important at scales $Q \sim m_{\mathrm{H}}$. To take account of the mass effects, the simplest way is that of a *fixed flavour number scheme* (FFNS), where one keeps the number of flavours in the evolution fixed to n_f and uses massive coefficient functions for the heavy quark, but this approach loses its validity at the high scales $Q \gg m_{\mathrm{H}}$. Combining the above two approaches with validity extended to all scales, in a *general-mass variable flavour number scheme* (GM-VFNS), one switches to the $n_f + 1$

evolution and uses the massive coefficient functions subtracted with terms that prevent double counting. These subtraction terms depend on which mass terms one chooses to include in the heavy-quark coefficient functions and thus there is not one, but many different GM-VFN schemes.

The scheme utilized in the articles of this thesis is that of the simplified Aivazis–Collins–Olness–Tung (SACOT) [35, 36]. In this scheme, one uses the $\overline{\text{MS}}$ coefficient functions together with the heavy-quark PDFs, $C_{q_\text{H}} = C_q^{\overline{\text{MS}}}$. Below a transition scale $Q_t \sim m_\text{H}$ a fixed-flavour prescription is used, with the F_2 structure function expressible as

$$\begin{aligned} F_2^{\text{NLO}} \overset{Q^2<Q_t^2}{=} & \, x\sum_{q_\ell} e_{q_\ell}^2 \left\{ C_q^{\overline{\text{MS}}}(Q^2/Q_f^2) \otimes f_{q_\ell}(Q_f^2)\,(x) + C_g^{\overline{\text{MS}}}(Q^2/Q_f^2) \otimes f_g(Q_f^2)\,(x) \right\} \\ & + x\sum_{q_\text{H}} e_{q_\text{H}}^2 \, C_{g\to q_\text{H}}^{\text{FFNS}}(m_\text{H}^2/Q^2) \otimes f_g(Q_f^2)\,(\chi), \end{aligned} \tag{2.42}$$

where $\chi = x(1+4m_\text{H}^2/Q^2)$ is the rescaling variable which accounts for the energy needed in heavy-quark pair production, and n_f flavours are taken in the evolution. The fixed-flavour coefficient function $C_{g\to q_\text{H}}^{\text{FFNS}}$ can be found in Ref. [37]. Above the transition scale, the structure function can be writen as [38]

$$\begin{aligned} F_2^{\text{NLO}} \overset{Q^2>Q_t^2}{=} & \, x\sum_{q_\ell} e_{q_\ell}^2 \left\{ C_q^{\overline{\text{MS}}}(Q^2/Q_f^2) \otimes f_{q_\ell}(Q_f^2)\,(x) + C_g^{\overline{\text{MS}}}(Q^2/Q_f^2) \otimes f_g(Q_f^2)\,(x) \right\} \\ & + x\sum_{q_\text{H}} e_{q_\text{H}}^2 \left\{ \left[C_{g\to q_\text{H}}^{\text{FFNS}}(m_\text{H}^2/Q^2) - \frac{\alpha_\text{s}}{2\pi}\log(Q_f^2/m_\text{H}^2) P_{qg} \right] \otimes f_g(Q_f^2)\,(\chi) \right. \\ & \left. + \; C_q^{\overline{\text{MS}}}(Q^2/Q_f^2) \otimes f_{q_\text{H}}(Q_f^2)\,(\chi) \right\}, \end{aligned} \tag{2.43}$$

with now $n_f + 1$ flavours in the scale evolution. In the above expressions, the sums should again be understood to go over both quarks and antiquarks, in the first sum for the light (massless) flavours and in the second for the heavy-quark flavour.

2.5 Sum rules and symmetry relations

Due to the conservation of flavour in QCD interactions, we have the following sum rules for the proton PDFs

$$\int_0^1 \mathrm{d}x\, u_\text{V}^p(x,Q^2) = 2, \qquad \int_0^1 \mathrm{d}x\, d_\text{V}^p(x,Q^2) = 1, \tag{2.44}$$

where the valence distributions are defined as $q_\text{V} = q - \bar{q}$ and where we have introduced the shorthand $q(x,Q^2) = f_q(x,Q^2)$, $g(x,Q^2) = f_g(x,Q^2)$. Similarly,

the conservation of momentum requires the sum rule

$$\int_0^1 \mathrm{d}x\, x \sum_i f_i(x, Q^2) = 1. \tag{2.45}$$

Due to the degeneracy in the u and d quark masses, QCD has an approximative symmetry called isospin. This symmetry relates the PDFs of the proton and neutron with

$$u^p \overset{\text{IS}}{=} d^n, \qquad d^p \overset{\text{IS}}{=} u^n, \qquad \bar{u}^p \overset{\text{IS}}{=} \bar{d}^n, \qquad \bar{d}^p \overset{\text{IS}}{=} \bar{u}^n \tag{2.46}$$

and $f_i^p \overset{\text{IS}}{=} f_i^n$ for $i \neq u, d$. One can also use charge conjugation (CC), which is an exact symmetry of QCD, to relate the PDFs of proton and antiproton,

$$q^p \overset{\text{CC}}{=} \bar{q}^{\bar{p}}, \qquad \bar{q}^p \overset{\text{CC}}{=} q^{\bar{p}}, \qquad g^p \overset{\text{CC}}{=} g^{\bar{p}}, \tag{2.47}$$

or, using both symmetries, the PDFs of charged pions,

$$\begin{gathered} u^{\pi^+} \overset{\text{IS}}{=} d^{\pi^-} \overset{\text{CC}}{=} \bar{d}^{\pi^+} \overset{\text{IS}}{=} \bar{u}^{\pi^-}, \qquad d^{\pi^+} \overset{\text{IS}}{=} u^{\pi^-} \overset{\text{CC}}{=} \bar{u}^{\pi^+} \overset{\text{IS}}{=} \bar{d}^{\pi^-}, \\ q^{\pi^+} \overset{\text{IS}}{=} q^{\pi^-} \overset{\text{CC}}{=} \bar{q}^{\pi^+} \overset{\text{IS}}{=} \bar{q}^{\pi^-} \text{ for } q \neq u, d, \qquad g^{\pi^-} \overset{\text{CC}}{=} g^{\pi^+}. \end{gathered} \tag{2.48}$$

The isospin symmetry relations for nucleons, Eq. (2.46), are essential for the discussion in Chapter 4, assumed by practically all nPDF analyses. We also need to employ the charged pion relations, Eq. (2.48), when discussing the results of the article [**I**] in Section 4.2.1.

2.6 Factorization in hadron–hadron collisions

The same perturbative approach which we have discussed in previous sections in the case of DIS also applies to hadron–hadron collision processes. This is stated formally in the factorization theorem which says that, order by order in perturbation theory, the collinear logarithms arising in hard-process calculations can be resummed into scale dependent long-distance functions in such a way that the full cross section becomes finite [1]. Importantly, the structure of collinear divergences is the same in DIS and hadron–hadron processes, leading to *universality* of the PDFs.

The relevant processes for this thesis are illustrated in Figure 2.4. In the work presented in this thesis, various publicly available codes have been used in calculating them at the NLO level. In first of these processes, Drell–Yan (DY) dilepton production, $h + h' \to l^- l^+ + X$, the leading-order process happens through an annihilation of a quark and antiquark originating from the colliding

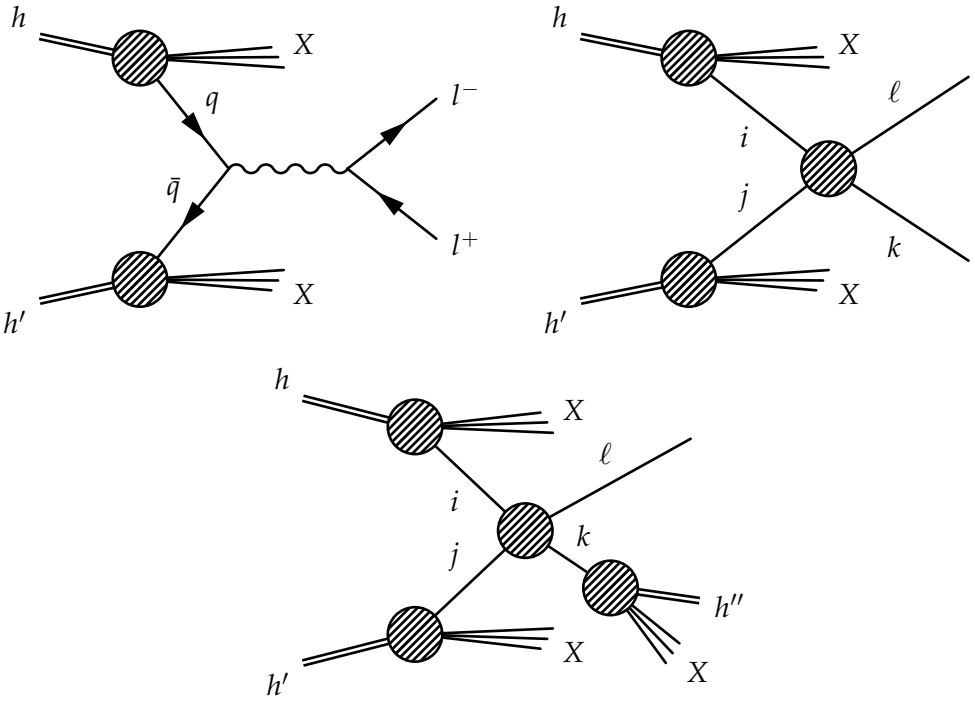

Figure 2.4. Drell–Yan dilepton pair (upper left), dijet (upper right) and inclusive hadron (bottom) production in hadron–hadron collision at leading order of perturbation theory.

hadrons h and h', as shown in Figure 2.4 (upper left). In more general terms, the cross section factorizes, schematically

$$\sigma_{\mathrm{DY}}^{h+h'} = \sum_{i,j=q,\bar{q},g} f_i^h \otimes f_j^{h'} \otimes \hat{\sigma}^{ij\to l^- l^+ + X}, \tag{2.49}$$

where there are now two PDFs, f_i^h and $f_j^{h'}$, convoluted with the perturbatively calculable pieces. The production of massive electroweak (EW) gauge bosons proceeds in a similar way. For practical applications, the MCFM program [39] was used in calculating the NLO pion–nucleus DY cross sections in the articles **[I]** and **[II]**, and for the EW-boson cross sections in the article **[II]**.

It is also possible to consider the production of various hadronic final states, such as production of a dijet system, $h + h' \to \mathrm{jet} + \mathrm{jet} + X$. In this process at leading order, initial-state partons i,j undergo a scattering into final-state partons ℓ,k, which are observed as high-p_{T} hadronic jets in the detector, as shown in Figure 2.4 (upper right). At higher orders, this simple parton-to-jet correspondence is lost, and the jets are defined in terms of jet algorithms. Formally still, the perturbative part of the cross section can be expressed in terms of a

measurement function F_{dijet} that defines the dijet,

$$\sigma^{h+h'}_{\text{dijet}} = \sum_{i,j=q,\bar{q},g} f_i^h \otimes f_j^{h'} \otimes \hat{\sigma}^{ij}[F_{\text{dijet}}]. \tag{2.50}$$

For the calculation of dijet cross sections, the article [**III**] utilized the NLOJet++ code [40], while the MEKS program [41] was used in the jet calculations of the articles [**II**] and [**IV**].

Instead of measuring jets, one can alternatively consider final states inclusive in a hadron species h'',

$$\sigma^{h+h'\to h''+X} = \sum_{i,j,k=q,\bar{q},g} f_i^h \otimes f_j^{h'} \otimes \hat{\sigma}^{ij\to k+X} \otimes D_k^{h''}, \tag{2.51}$$

illustrated in Figure 2.4 (bottom). In such processes also the final state collinear logarithms need to be resummed, this time into *fragmentation functions* $D_k^{h''}(z, Q^2)$, which give the probability for finding a final state hadron h, which has fragmented off from a hard parton i, carrying a fraction z of the partons momentum. The inclusive pion-production cross sections considered in the article [**II**] were calculated with the INCNLO code [42]. The calculations for heavy-flavoured mesons are much more involved [43] with various mass schemes again applicable, similarly to what was discussed in Section 2.4. In the article [**V**], a recently developed variant of the SACOT scheme [44] was used with the zero-mass contributions obtained from the INCNLO [42] and the massive contributions from the MNR [45] codes.

Chapter 3

Global analysis and uncertainty estimation

As discussed in the previous chapter, the PDFs describe long-range physics and cannot be calculated perturbatively from first principles. The common approach for obtaining them is then to use the means of statistical inference: By performing a "global analysis" on multiple observables sensitive to the PDFs, one aims to deduce the partonic structure from the measured hard-process data. This is in principle an infinite-dimensional optimization problem, as there is no a priori knowledge of the functional form. However, we do not have an infinite amount of perfectly precise data from which the PDFs could be obtained by inversion. For this reason, the PDFs need to be parametrized in a way or another, be it some suitably chosen functional form or a neural network [46].

Once the parametrization form is decided upon, one then needs to find the range of parameter values the data would support. For this, one defines a goodness-of-fit function χ^2, the minimum of which corresponds to the best-fit values of the parameters. The various steps needed in the χ^2 minimization are illustrated in Figure 3.1. One begins by setting a suitable first guess for the parameter values, which give the PDFs at a chosen parametrization scale Q_0^2. Using the DGLAP equations, these PDFs are then evolved to higher scales and convoluted with the coefficient functions to obtain theoretical predictions. To reduce the time required by the fitting, fast methods for performing these convolutions are needed [47–49], such as the use of look-up tables as explained in the Section 3.3 of the article [**II**]. Comparing these predictions with the measured values, one calculates the χ^2 figure-of-merit value for the chosen parameters. This procedure is then repeated for different sets of parameter values, until the minimum of χ^2 is reached.

In addition to the functional form, the obtained result depends on various other inputs. The most obvious of these is which data one chooses to use. In principle, one would like to include as much data as possible to have the best

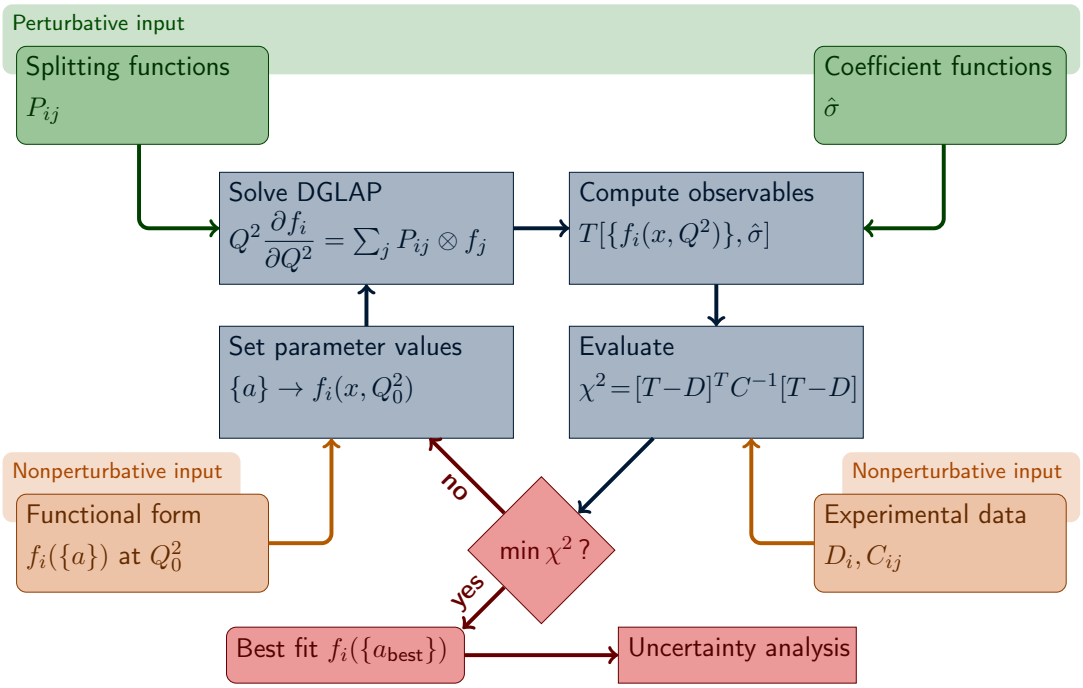

Figure 3.1. Flow of the χ^2 minimization in a PDF global analysis. Figure from Ref. [50].

constraining power, but care must be taken to only include measurements where one can trust the theoretical description of the process to avoid possible biases. For example, one should only include processes which are clearly in the perturbative regime to be able to neglect power corrections, but the exact value of minimum Q^2 to allow is somewhat arbitrary and different cuts are used by different groups, see Table 4.1 for conventions in nPDF fits.

The results of minimization also depend on the level or perturbative accuracy in the used splitting and coefficient functions. It is hard to quantify the size of these theoretical uncertainties and they are usually neglected in reporting PDF errors, although work towards taking these uncertainties into account in global analyses is ongoing [51–53]. Therefore, one usually only propagates the experimental uncertainties into the uncertainties of the PDFs and the subsequent predictions. Section 3.3 discusses how this is done in the Hessian formalism [54] applied in this thesis work.

3.1 Statistical basis of global analysis

In this and the following section, we show how the χ^2-function minimization arises as a maximum-likelihood estimator of the parameters. The viewpoint

taken here is that of frequentist probability theory, for a Bayesian equivalent we refer the reader to Ref. [14].

Due to experimental uncertainties, each measured value D_i of any observable differs from its true value T_i by some error δ_i,

$$D_i = T_i + \delta_i. \tag{3.1}$$

Let us first assume that these errors are uncorrelated between the measurements, $\delta_i = \delta_i^{\text{uncorr.}}$ with $\langle(\delta_i^{\text{uncorr.}} - \langle\delta_i^{\text{uncorr.}}\rangle)(\delta_j^{\text{uncorr.}} - \langle\delta_j^{\text{uncorr.}}\rangle)\rangle = 0$ for $i \neq j$, and follow a Gaussian distribution with a zero mean, $\langle\delta_i^{\text{uncorr.}}\rangle = 0$, and a variance $\langle(\delta_i^{\text{uncorr.}} - \langle\delta_i^{\text{uncorr.}}\rangle)^2\rangle = (\sigma_i^{\text{uncorr.}})^2$. The probability density for each $\delta_i^{\text{uncorr.}}$ thus reads

$$P(\delta_i^{\text{uncorr.}}) = \frac{1}{\sqrt{2\pi}\sigma_i^{\text{uncorr.}}} \mathrm{e}^{-(\delta_i^{\text{uncorr.}})^2/2(\sigma_i^{\text{uncorr.}})^2}. \tag{3.2}$$

Since the errors are independent, the joint probability of a set of errors $\{\delta_i^{\text{uncorr.}}\}$ is simply

$$P(\{\delta_i^{\text{uncorr.}}\}) = \prod_i P(\delta_i^{\text{uncorr.}}). \tag{3.3}$$

By changing variables to D_i according to Eq. (3.1), we can construct the joint probability for obtaining a set of N_{data} mutually independent measurements $\{D_i\}$ for given $\{T_i\}$,

$$\begin{aligned} P(\{D_i\}|\{T_i\}) &= \prod_i \int \mathrm{d}\delta_i^{\text{uncorr.}} \delta(D_i - T_i - \delta_i^{\text{uncorr.}}) P(\delta_i^{\text{uncorr.}}) \\ &= \frac{1}{(2\pi)^{N_{\text{data}}/2} \prod_i \sigma_i^{\text{uncorr.}}} \mathrm{e}^{-\frac{1}{2}\sum_i (D_i - T_i)^2/(\sigma_i^{\text{uncorr.}})^2}. \end{aligned} \tag{3.4}$$

In PDF fits the true values $\{T_i\}$ are of course not known, but neglecting the theoretical uncertainties, one can trade these with the pQCD predictions with PDFs given by a set of parameters $\{a\}$, $T_i = T_i(\{a\})$. The *likelihood* for a certain set of values of $\{a\}$ is then related to the probability of obtaining $\{D_i\}$ for given $\{a\}$ as

$$L(\{a\}) \coloneqq P(\{D_i\}|\{a\}) = \frac{1}{(2\pi)^{N_{\text{data}}/2} \prod_i \sigma_i^{\text{uncorr.}}} \mathrm{e}^{-\frac{1}{2}\sum_i (D_i - T_i(\{a\}))^2/(\sigma_i^{\text{uncorr.}})^2}. \tag{3.5}$$

In the global fit, we wish to find the parameter values which maximize this likelihood function.

The parameter values $\{a\}$ which give the maximal likelihood also minimize the χ^2 function

$$\chi^2(\{a\}) \coloneqq -2\log L(\{a\}) + \text{const.} = \sum_i \left(\frac{D_i - T_i(\{a\})}{\sigma_i^{\text{uncorr.}}} \right)^2, \tag{3.6}$$

which just shows that in the case of Gaussian errors, the maximum-likelihood and least-squares estimators are the same [55]. Note that the χ^2 function defined above essentially compares observed data fluctuations $D_i - T_i$ to expected ones $\sigma_i^{\text{uncorr.}}$ and in the limit of perfect theoretical description of the data we should obtain $\chi^2 \approx N_{\text{d.o.f.}} = N_{\text{data}} - N_{\text{par.}}$, the number of degrees of freedom, where $N_{\text{par.}}$ is the number of free parameters. Thus, on one hand, a value much higher than this would then tell that the fit does not describe the data well and, on the other hand, a significantly smaller value would be a signal of possible overfitting. In this sense, the χ^2 is a goodness-of-fit function. A similar interpretation cannot be given for the value of the likelihood function at its maximum due to the way it is normalized.

In deriving Eq. (3.6) we have assumed that the errors have a Gaussian distribution. This is an assumption that we often make in lack of better knowledge. In fact, the measured quantities are often cross sections, which should not go negative, but with the Gaussian distribution, we are assuming a nonvanishing probability for the measured value to be less than zero. However, when uncertainties are small, any corrections to Eq. (3.6) should be small and its usage perfectly valid.

3.2 Fitting to data with correlated uncertainties

Let us now discuss the treatment of data with correlated uncertainties. We take these to be additive, leaving the treatment of multiplicative uncertainties to Section 3.2.3. The total measurement error can then be decomposed as

$$\delta_i = \delta_i^{\text{uncorr.}} + \delta_i^{\text{corr.}}, \tag{3.7}$$

where $\delta_i^{\text{uncorr.}}$ is the uncorrelated error distributed according to Eq. (3.2) and $\delta_i^{\text{corr.}} = \sum_k \beta_i^k \lambda_k$ sums the errors from independent systematical sources λ_k. Sections 3.2.1 and 3.2.2 discuss two ways of treating the λ_k in formulating the χ^2 function, "marginalization" and "profiling". In the case of additive Gaussian uncertanties these methods give identical results [56].

3.2.1 Covariance matrix from marginalization

We take here the λ_k to be Gaussian distributed random variables with zero mean and normalized such that

$$P(\lambda_k) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\lambda_k^2/2}. \tag{3.8}$$

This way, β_i^k can be interpreted as the response of the ith data point on a one standard deviation shift in the kth experimental systematic source of error. While

the $\delta_i^{\text{corr.}}$ defined this way are correlated amongst themselves, the λ_k are taken to be independent and hence

$$P(\{\delta_i^{\text{uncorr.}}\},\{\lambda_k\}) = \prod_i P(\delta_i^{\text{uncorr.}}) \prod_k P(\lambda_k). \tag{3.9}$$

Again, we can trade the $\delta_i^{\text{uncorr.}}$ with D_i using Eqs. (3.1) and (3.7) to obtain

$$\begin{aligned} &P(\{D_i\},\{\lambda_k\}|\{a\}) \\ &= \frac{1}{(2\pi)^{N_{\text{data}}/2}\prod_i \sigma_i^{\text{uncorr.}} (2\pi)^{N_{\text{syst.}}/2}} \mathrm{e}^{-\frac{1}{2}\sum_i (D_i - T_i(\{a\}) - \sum_k \beta_i^k \lambda_k)^2/(\sigma_i^{\text{uncorr.}})^2 - \frac{1}{2}\sum_k \lambda_k^2}, \end{aligned} \tag{3.10}$$

where $N_{\text{syst.}}$ is the number of systematical sources. We can integrate over the $\{\lambda_k\}$ in Eq. (3.10) to get the *marginal probability distribution* for the data points,

$$\begin{aligned} P(\{D_i\}|\{a\}) &= \int \prod_k \mathrm{d}\lambda_k P(\{D_i\},\{\lambda_k\}|\{a\}) \\ &= \frac{1}{(2\pi)^{N_{\text{data}}/2}\prod_i \sigma_i^{\text{uncorr.}}} \mathrm{e}^{-\frac{1}{2}\sum_i (D_i - T_i)^2/(\sigma_i^{\text{uncorr.}})^2} \\ &\quad \times \frac{1}{(2\pi)^{N_{\text{syst.}}/2}} \int \prod_k \mathrm{d}\lambda_k \mathrm{e}^{-\frac{1}{2}\sum_{k,\ell}\lambda_k \overbrace{\left(\sum_i \frac{\beta_i^k \beta_i^\ell}{(\sigma_i^{\text{uncorr.}})^2} + \delta^{k\ell}\right)}^{=:A^{k\ell}} \lambda_\ell + \sum_k \sum_i (D_i - T_i) \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} \lambda_k}, \end{aligned} \tag{3.11}$$

where we dropped the explicit $\{a\}$ dependence of T_i for simplicity. The matrix A, with components defined above, is symmetric and positive definite, whereby the Gaussian integral in Eq. (3.11) can be performed. This yields

$$\begin{aligned} P(\{D_i\}|\{a\}) = \frac{1}{(2\pi)^{N_{\text{data}}/2}\prod_i \sigma_i^{\text{uncorr.}}} &\mathrm{e}^{-\frac{1}{2}\sum_i (D_i - T_i)^2/(\sigma_i^{\text{uncorr.}})^2} \\ &\times \frac{1}{\sqrt{\det A}} \mathrm{e}^{-\frac{1}{2}\sum_{i,j}(D_i - T_i)\left[\sum_{k,\ell} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{k\ell} \frac{\beta_j^\ell}{(\sigma_j^{\text{uncorr.}})^2}\right](D_j - T_j)} \end{aligned} \tag{3.12}$$

The likelihood function is then defined similarly as with the uncorrelated uncertainties in Section 3.1,

$$L(\{a\}) := P(\{D_i\}|\{a\}) = \frac{1}{(2\pi)^{N_{\text{data}}/2}\prod_i \sigma_i^{\text{uncorr.}}} \frac{1}{\sqrt{\det A}} \mathrm{e}^{-\frac{1}{2}\chi^2(\{a\})}, \tag{3.13}$$

where now

$$\begin{aligned} \chi^2(\{a\}) &= \sum_{i,j}(D_i - T_i) \underbrace{\left[\frac{\delta_{ij}}{(\sigma_i^{\text{uncorr.}})^2} - \sum_{k,h} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{kh} \frac{\beta_j^h}{(\sigma_j^{\text{uncorr.}})^2}\right]}_{=:C_{ij}^{-1}} (D_j - T_j) \\ &= [D - T]^T C^{-1} [D - T]. \end{aligned} \tag{3.14}$$

The matrix C^{-1} defined above is simply the inverse of the covariance matrix of the data, which is given by

$$C_{ij} = \langle (D_i - \langle D_i \rangle)(D_j - \langle D_j \rangle) \rangle = (\sigma_i^{\text{uncorr.}})^2 \delta_{ij} + \sum_k \beta_i^k \beta_j^k, \tag{3.15}$$

as can be easily shown by taking the matrix product

$$\sum_n C_{in} C_{nj}^{-1} = \delta_{ij} + \sum_k \frac{\beta_i^k \beta_j^k}{(\sigma_i^{\text{uncorr.}})^2} - \sum_{h,\ell} \Bigg[\underbrace{\beta_i^h + \sum_{n,k} \frac{\beta_i^k \beta_n^k \beta_n^h}{(\sigma_i^{\text{uncorr.}})^2}}_{= \sum_k \beta_i^k A^{kh}} \Bigg] (A^{-1})^{h\ell} \frac{\beta_j^\ell}{(\sigma_j^{\text{uncorr.}})^2} = \delta_{ij}. \tag{3.16}$$

Eq. (3.14) is the standard covariance-matrix formulation of the χ^2 function. It reduces to the uncorrelated form Eq. (3.6) in the limit where $\beta_i^k \to 0$ for all i, k.

3.2.2 Nuisance parameter profiling

Another way to treat the correlated uncertainties is to take the systematic shifts λ_k to be free parameters of our statistical model. As these are not parameters of primary interest, they are called "nuisance parameters". Since parameters are not allowed to have probabilities in the frequentist approach that we have adopted, Eq. (3.8) does not apply directly here. Rather, we should understand each of the nuisance parameters to be constrained by some systematical statistic $\hat{\lambda}_k$ distributed by

$$P(\hat{\lambda}_k | \lambda_k) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(\hat{\lambda}_k - \lambda_k)^2/2}, \tag{3.17}$$

and having an experimental value $\hat{\lambda}_k = 0$. The likelihood function for the full set of parameters then reads

$$\begin{aligned} L(\{a\}, \{\lambda_k\}) &:= P(\{D_i\}, \{\hat{\lambda}_k = 0\} | \{a\}, \{\lambda_k\}) \\ &= \frac{1}{(2\pi)^{(N_{\text{data}} + N_{\text{syst.}})/2} \prod_i \sigma_i^{\text{uncorr.}}} \mathrm{e}^{-\frac{1}{2}\chi^2(\{a\}, \{\lambda_k\})}, \end{aligned} \tag{3.18}$$

where the χ^2 function in this case is defined as

$$\chi^2(\{a\}, \{\lambda_k\}) = \sum_i \frac{(D_i - T_i(\{a\}) - \sum_k \beta_i^k \lambda_k)^2}{(\sigma_i^{\text{uncorr.}})^2} + \sum_k \lambda_k^2. \tag{3.19}$$

As Eq. (3.19) is quadratic in λ_k we can find the minimum analytically. Setting the first derivatives to zero,

$$\left. \frac{\partial \chi^2}{\partial \lambda_k} \right|_{\{\lambda_\ell\} = \{\lambda_\ell^{\min}\}} = -2\beta_i^k \sum_i \frac{D_i - T_i - \sum_\ell \beta_i^\ell \lambda_\ell^{\min}}{(\sigma_i^{\text{uncorr.}})^2} + 2\lambda_k^{\min} = 0 \quad \text{for all } k, \tag{3.20}$$

we find

$$\sum_\ell \Bigg(\underbrace{\sum_i \frac{\beta_i^k \beta_i^\ell}{(\sigma_i^{\text{uncorr.}})^2} + \delta^{k\ell}}_{=A^{k\ell}} \Bigg) \lambda_\ell^{\min} = \sum_i (D_i - T_i) \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2}, \tag{3.21}$$

where the matrix A is the same which we encountered in Eq. (3.11). Performing a matrix multiplication with its inverse to Eq. (3.21) gives

$$\lambda_h^{\min} = \sum_k (A^{-1})^{hk} \sum_\ell A^{k\ell} \lambda_\ell^{\min} = \sum_i (D_i - T_i) \sum_k (A^{-1})^{hk} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2}. \tag{3.22}$$

The obtained values can be substituted back to Eq. (3.18), giving us a *profile likelihood*, which is a function of $\{a\}$ only. At the minimum of Eq. (3.19) we have

$$\begin{aligned} &\sum_i \frac{(D_i - T_i - \sum_k \beta_i^k \lambda_k^{\min})^2}{(\sigma_i^{\text{uncorr.}})^2} \\ &= \sum_{i,j} (D_i - T_i) \Bigg[\frac{\delta_{ij}}{(\sigma_i^{\text{uncorr.}})^2} - 2 \overbrace{\sum_{k,h} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{kh} \frac{\beta_j^h}{(\sigma_j^{\text{uncorr.}})^2}}^{=:B_{ij}^{(1)}} \\ &\qquad + \underbrace{\sum_{k,\ell,m,h} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{k\ell} \sum_n \frac{\beta_n^\ell \beta_n^m}{(\sigma_n^{\text{uncorr.}})^2} (A^{-1})^{mh} \frac{\beta_j^h}{(\sigma_j^{\text{uncorr.}})^2}}_{=:B_{ij}^{(2)}} \Bigg] (D_j - T_j) \end{aligned} \tag{3.23}$$

and

$$\sum_k (\lambda_k^{\min})^2 = \sum_{i,j} (D_i - T_i) \Bigg[\underbrace{\sum_{k,\ell,h} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{k\ell} (A^{-1})^{\ell h} \frac{\beta_j^h}{(\sigma_j^{\text{uncorr.}})^2}}_{=:B_{ij}^{(3)}} \Bigg] (D_j - T_j). \tag{3.24}$$

Here we notice that

$$\begin{aligned} B_{ij}^{(2)} + B_{ij}^{(3)} &= \sum_{k,\ell,m,h} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{k\ell} \Bigg(\underbrace{\sum_n \frac{\beta_n^\ell \beta_n^m}{(\sigma_n^{\text{uncorr.}})^2} + \delta^{\ell m}}_{=A^{\ell m}} \Bigg) (A^{-1})^{mh} \frac{\beta_j^h}{(\sigma_j^{\text{uncorr.}})^2} \\ &= \sum_{k,h} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{kh} \frac{\beta_j^h}{(\sigma_j^{\text{uncorr.}})^2} = B_{ij}^{(1)} \end{aligned} \tag{3.25}$$

and hence

$$\min_{\{\lambda_k\}} \chi^2 = \sum_{i,j} (D_i - T_i) \underbrace{\left[\frac{\delta_{ij}}{(\sigma_i^{\text{uncorr.}})^2} - \sum_{k,h} \frac{\beta_i^k}{(\sigma_i^{\text{uncorr.}})^2} (A^{-1})^{kh} \frac{\beta_j^h}{(\sigma_j^{\text{uncorr.}})^2} \right]}_{=C_{ij}^{-1}} (D_j - T_j)$$

$$= [D - T]^T C^{-1} [D - T]. \tag{3.26}$$

This shows that the covariance-matrix and nuisance-parameter formulations of the χ^2 function are equivalent and either one can be used to treat the correlated uncertainties.

The nuisance-parameter approach facilitates an easy way for a graphical data-to-theory comparison in situations where simply adding quadratically the correlated and uncorrelated uncertainties would exaggerate the uncertainties. By defining

$$D_i^{\text{shifted}} := D_i - \sum_k \beta_i^k \lambda_k^{\min}, \tag{3.27}$$

we may write

$$\min_{\{\lambda_k\}} \chi^2 = \sum_i \frac{(D_i^{\text{shifted}} - T_i)^2}{(\sigma_i^{\text{uncorr.}})^2} + \sum_k (\lambda_k^{\min})^2. \tag{3.28}$$

That is, if we shift the data according to Eq. (3.27), the remaining differences between data and theory should be from the uncorrelated uncertainties, point by point. This method was used for example in the article **[III]** for presenting the inclusive jet data.

3.2.3 Normalization uncertainties

Until now we have taken the considered uncertainties to be of additive nature, i.e. each of the errors simply adds on the difference between the measured and true value, irrespective of what these values are. However, some uncertainties are known to be multiplicative in the sense that their magnitudes depend on the measured (or true) value. Luminosity uncertainties are good examples of such: the errors they pose on the measured cross sections are proportional to the (expected) number of events. Experiments often give these uncertainties in terms of normalization uncertainties, where each measured data point is subject to a mutual, fully correlated, percentual uncertainty, but also more complicated situations are possible. These uncertainties need to be treated correctly to avoid possible biases, as we will discuss next.

d'Agostini bias

Since the normalization uncertainty is a property of the data, it might appear natural to take it into account in the χ^2 by introducing a normalization factor f_N multiplying the data points and assuming a Gaussian uncertainty $\sigma^{\text{norm.}}$ for it, and therefore write

$$\chi^2(\{a\}, f_N) = \sum_i \left(\frac{f_N D_i - T_i(\{a\})}{\sigma_i^{\text{uncorr.}}} \right)^2 + \left(\frac{f_N - 1}{\sigma^{\text{norm.}}} \right)^2, \tag{3.29}$$

as was done e.g. in Ref. [56] and also in the article [**II**] of this thesis. However, it can be shown that this formulation is subjective to so-called d'Agostini bias [57].

Following the example given in Ref. [57], let us assume that we have taken N_{data} measurements $\{D_i\}$ of a single observable quantity and that these data points share a common normalization uncertainty $\sigma^{\text{norm.}}$. We would then like to find the best estimate for the true value T from which the measured values derive. For simplicity, let us also assume that the data points all have identical uncorrelated statistical uncertainties with variances $(\sigma^{\text{uncorr.}})^2$. The χ^2 function of Eq. (3.29) then becomes

$$\chi^2(T, f_N) = \frac{1}{(\sigma^{\text{uncorr.}})^2} \sum_i (f_N D_i - T)^2 + \left(\frac{f_N - 1}{\sigma^{\text{norm.}}} \right)^2. \tag{3.30}$$

This is easily minimized with respect to both T and f_N. We find

$$T^{\min} = f_N^{\min} \hat{D}, \qquad f_N^{\min} = \frac{1}{1 + N_{\text{data}} \frac{(\sigma^{\text{norm.}})^2}{(\sigma^{\text{uncorr.}})^2} \sigma_D^2}, \tag{3.31}$$

where

$$\hat{D} = \frac{1}{N_{\text{data}}} \sum_i D_i, \qquad \sigma_D^2 = \frac{1}{N_{\text{data}}} \sum_i D_i^2 - \left(\frac{1}{N_{\text{data}}} \sum_i D_i \right)^2 \tag{3.32}$$

are the sample mean and the sample variance of the data, respectively. Now, as we have not introduced a statistical model, but taken the χ^2 function as given, it is not clear how σ_D^2 is related to the uncorrelated error. However, if we assume the *true* normalization to be simply unity, one can then show that

$$\langle \sigma_D^2 \rangle = \frac{N_{\text{data}} - 1}{N_{\text{data}}} (\sigma^{\text{uncorr.}})^2 \tag{3.33}$$

and hence the expected value for the optimal normalization following from Eq. (3.30) is

$$\langle f_N^{\min} \rangle = \frac{1}{1 + (N_{\text{data}} - 1)(\sigma^{\text{norm.}})^2}. \tag{3.34}$$

This is clearly biased, as it tends towards zero as we increase the number of measurements. One can see why this happens also in a more general case by looking at Eq. (3.29). By making both f_N and $\{T_i\}$ smaller, also the difference in the numerator of the first term in Eq. (3.29) diminishes. As there is no similar compensation in the denominator, such a decrease in the normalization is favoured in the fit, whether that be truly statistically motivated or not. This can cause a significant bias in the found $\{T_i\}$ and thus also in the fitted parameters.

In real world PDF fits, such as in the article **[II]**, the bias is typically not as severe as in the above simple case. Here we assumed that the quantity of interest T was completely free in the fit, but in a typical global fit the parameters are constrained by multiple independent data sets and limited also by the sum rules. Still, in article **[IV]** of this thesis we encountered a case where this bias had an effect on the results and an unbiased method was called for.

Unbiased fitting

Let us assume, in a general setting, that each of the measured values D_i deviates from the true value T_i by a common normalization factor f_N plus an individual, uncorrelated error $\delta_i^{\text{uncorr.}}$ such that

$$D_i = f_N T_i + \delta_i^{\text{uncorr.}} \tag{3.35}$$

and, treating f_N as a nuisance parameter, the measured normalization deviates from the true one by $\hat{f}_N = f_N + \delta^{\text{norm.}}$. Taking all uncertainties to be Gaussian distributed and independent, with Eq. (3.2) and

$$P(\delta^{\text{norm.}}) = \frac{1}{\sqrt{2\pi}\sigma^{\text{norm.}}} \mathrm{e}^{-(\delta^{\text{norm.}})^2/2(\sigma^{\text{norm.}})^2}, \tag{3.36}$$

and taking the experimental value $\hat{f}_N = 1$, we have

$$\begin{aligned} &P(\{D_i\}, \hat{f}_N = 1 | \{T_i\}, f_N) \\ &= \frac{1}{(2\pi)^{(N_{\text{data}}+1)/2}\sigma^{\text{norm.}}\prod_i \sigma_i^{\text{uncorr.}}} \mathrm{e}^{-\frac{1}{2}\sum_i (D_i - f_N T_i)^2/(\sigma_i^{\text{uncorr.}})^2 - \frac{1}{2}(f_N-1)^2/(\sigma^{\text{norm.}})^2}. \end{aligned} \tag{3.37}$$

In this case, the likelihood function takes the form

$$L(\{a\}, f_N) = \frac{1}{(2\pi)^{(N_{\text{data}}+1)/2}\sigma^{\text{norm.}}\prod_i \sigma_i^{\text{uncorr.}}} \mathrm{e}^{-\frac{1}{2}\chi^2(\{a\}, f_N)}, \tag{3.38}$$

maximized at the minimum of

$$\chi^2(\{a\}, f_N) = \sum_i \left(\frac{D_i - f_N T_i(\{a\})}{\sigma_i^{\text{uncorr.}}} \right)^2 + \left(\frac{f_N - 1}{\sigma^{\text{norm.}}} \right)^2, \tag{3.39}$$

which only differs from Eq. (3.29) in so that f_N multiplies the theory values, not the data. Note, on the contrary, that minimization of Eq. (3.29) does *not* follow directly as a maximum-likelihood estimator from assuming $f_N D_i = T_i + \delta_i^{\text{uncorr.}}$ as in this case the likelihood function would have f_N in its normalization. Now, in the simple scenario discussed previously, one finds

$$T^{\min} = \hat{D}, \qquad f_N^{\min} = 1, \tag{3.40}$$

as should be the case when the data cannot provide additional information on the normalization. Eq. (3.39) is thus free from the d'Agostini bias. We note that there is also another way to treat the multiplicative uncertainties, called t_0 method, which is free also from a "non-decoupling bias", see Ref. [58].

3.3 Uncertainty estimation in Hessian method

In a global analysis, one aims at finding the best estimate for the PDFs based on available data and, importantly, determining the uncertainties in the results and communicating these in a way that allows to propagate the uncertainties into predictions made with the obtained PDFs. A common way to do this is the Hessian method [54]. Having found the values $\{a_i^{\min}\}$ which minimize the χ^2, we can approximate the behaviour around the minimum by

$$\chi^2(\{a_i\}) \approx \chi_0^2 + \sum_{i,j} (a_i - a_i^{\min})\, H_{ij}\, (a_j - a_j^{\min}), \tag{3.41}$$

where $\chi_0^2 = \chi^2(\{a_i^{\min}\})$ is the value at the minimum and $H_{ij} = \frac{1}{2}\partial^2\chi^2/\partial a_i \partial a_j|_{\{a_i^{\min}\}}$ are the elements of the Hessian matrix, which is symmetric and must be positive definite, for otherwise we would not be at the minimum. Due to these properties, the Hessian matrix has a complete set of orthonormal eigenvectors $\vec{v}^{(k)}$ with positive eigenvalues ε_k,

$$\sum_j H_{ij} v_j^{(k)} = \varepsilon_k v_i^{(k)}, \tag{3.42}$$

$$\sum_i v_i^{(k)} v_i^{(\ell)} = \delta_{k\ell}, \qquad \sum_k v_i^{(k)} v_j^{(k)} = \delta_{ij}. \tag{3.43}$$

By defining new parameters

$$z_k = \sum_i \sqrt{\varepsilon_k}\, v_i^{(k)}\, (a_i - a_i^{\min}) \tag{3.44}$$

the Hessian matrix can be diagonalized and the Equation (3.41) written as

$$\chi^2 \approx \chi_0^2 + \sum_k z_k^2. \tag{3.45}$$

This facilitates an easy way to propagate the uncertainties. Let us assume that we associate each of the new parameters z_k with an uncertainty Δz_k. Since the parameters in this basis are uncorrelated up to non-quardratic corrections, the related uncertainty in any quantity y_i can be written in this approximation by the standard law of error propagation as

$$(\Delta y_i)^2 = \sum_k \left(\frac{\partial y_i}{\partial z_k} \Delta z_k \right)^2 . \tag{3.46}$$

It then becomes a question of how large variations Δz_k one should allow. These can be related in the quadratic approximation to a global tolerance $\Delta\chi^2$ in the growth of the χ^2 from its minimum simply as $\Delta z_k = \sqrt{\Delta\chi^2}$. In presence of ideal Gaussian statistics one could further derive values of $\Delta\chi^2$ corresponding to exact confidence regions in the parameters [55]. However, for non-quadratic χ^2 functions using such pre-determined $\Delta\chi^2$ values can give only approximate coverage of the true parameter values [59]. Using a $\Delta\chi^2$ larger than some idealized value has also been motivated by conflicts between data sets [54] and parametrization uncertainties [60]. In fact, it has become more common to obtain the error tolerances by requiring that all the data sets remain in agreement within some confidence criterion under variations in each of the parameter directions, either separately [61], or on average as in the article [**II**]. This method is described in detail in the article [**II**] and thus will not be discussed further here.

Figure 3.2 shows the shape of the χ^2 function around the minimum in the EPPS16 analysis [**II**]. The quadratic approximation is typically very good, with only few eigendirections showing clear cubic or quartic components. To take into account such deviations from the ideal behaviour, one defines $\Delta z_k = (\delta z_k^+ - \delta z_k^-)/2$, where $\delta z_k^\pm$ are the values of z_k where χ^2 has grown from its minimum by $\Delta\chi^2$. To simplify the expressions, it is useful to define PDF error sets $S_i^\pm$ obtained with parameter values

$$z_k[S_i^\pm] = \delta_{ki}\, \delta z_i^\pm . \tag{3.47}$$

The derivative in Eq. (3.46) can then be approximated with

$$\frac{\partial y_i}{\partial z_k} = \frac{y_i[S_k^+] - y_i[S_k^-]}{2\,\Delta z_k}, \tag{3.48}$$

whereby the errors in PDFs or related observables can be calculated simply by using

$$(\Delta y_i)^2 = \frac{1}{4} \sum_k \left(y_i[S_k^+] - y_i[S_k^-] \right)^2 . \tag{3.49}$$

It is also possible to extend this expression into an asymmetric error prescription [62]

$$(\delta y_i^\pm)^2 = \sum_k \left[\begin{matrix} \max \\ \min \end{matrix} \left\{ y_i[S_k^+] - y_i[S_0], y_i[S_k^-] - y_i[S_0], 0 \right\} \right]^2 , \tag{3.50}$$

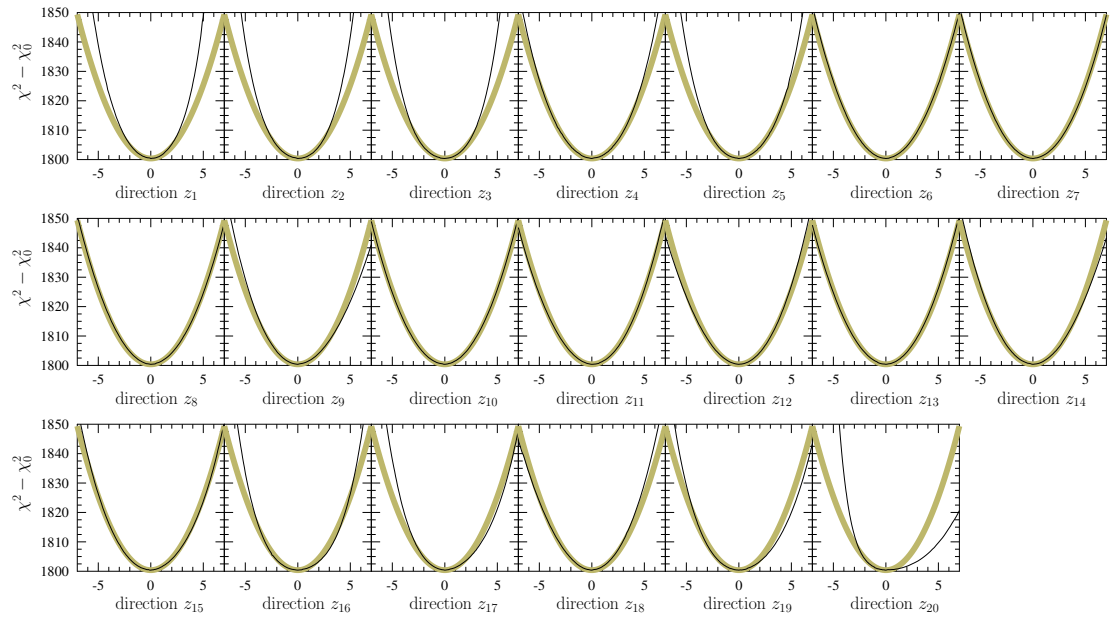

Figure 3.2. Growth of the χ^2 function in each of the eigendirections of the EPPS16 analysis. Black lines show the true dependence, whereas yellow lines represent the ideal quadratic behaviour. Figure from article **[II]**.

where S_0 is the central set with $z_k[S_0] = 0$ for all k.

3.4 Hessian PDF reweighting

Using the Hessian uncertainty estimation, it is also possible to estimate the impact of a new data set on the PDFs [63–65, **III**]. Assume that

$$\chi^2_{\text{old}}(\{z_k\}) \approx \chi^2_0 + \sum_k z_k^2 \tag{3.51}$$

is the χ^2 function of a PDF global analysis. To add a new data set $\{D_i^{\text{new}}\}$ to the analysis, we can simply write

$$\chi^2_{\text{new}} = \chi^2_{\text{old}} + \chi^2_{\text{new data}}, \tag{3.52}$$

where

$$\chi^2_{\text{new data}}(\{z_k\}) = \sum_{i,j} (T_i^{\text{new}}(\{z_k\}) - D_i^{\text{new}})\, C_{ij}^{-1}\, (T_j^{\text{new}}(\{z_k\}) - D_j^{\text{new}}). \tag{3.53}$$

By using Eq. (3.48), where we take $\Delta z_k = \sqrt{\Delta\chi^2}$ in accordance with the quadratic approximation in Eq. (3.51), we can estimate the parameter dependence of

any PDF-dependent quantity with a linear function

$$y_i(\{z_k\}) \approx y_i[S_0] + \sum_k \frac{\partial y_i}{\partial z_k} z_k, \qquad \frac{\partial y_i}{\partial z_k} = \frac{y_i[S_k^+] - y_i[S_k^-]}{2\sqrt{\Delta\chi^2}}. \tag{3.54}$$

Applying this approximation to $\{T_i^{\rm new}\}$, we find that $\chi^2_{\rm new}$ is a quadratic function of $\{z_k\}$ and can be minimized analytically. The new minimum is found at [64]

$$z_k^{\rm min} = \sum_\ell h^{-1}_{k\ell} \left[\sum_{i,j} \frac{\partial T_i^{\rm new}}{\partial z_\ell} C^{-1}_{ij} \left(D_j^{\rm new} - T_j^{\rm new}[S_0] \right) \right], \tag{3.55}$$

where

$$\frac{\partial T_i^{\rm new}}{\partial z_k} = \frac{T_i^{\rm new}[S_k^+] - T_i^{\rm new}[S_k^-]}{2\sqrt{\Delta\chi^2}} \tag{3.56}$$

and

$$h_{k\ell} = \delta_{k\ell} + \sum_{i,j} \frac{\partial T_i^{\rm new}}{\partial z_k} C^{-1}_{ij} \frac{\partial T_j^{\rm new}}{\partial z_\ell} \tag{3.57}$$

is the new Hessian matrix in

$$\chi^2_{\rm new}(\{z_k\}) \approx \chi^2_{\rm new}(\{z^{\rm min}\}) + \sum_{k\ell} (z_k - z_k^{\rm min})\, h_{k\ell}\, (z_\ell - z_\ell^{\rm min}). \tag{3.58}$$

Now, updated central predictions for related quantities can be obtained simply by substituting the found $z_k^{\rm min}$ to the linear approximation in Eq. (3.54). For example, the new best-fit PDFs are a simple weighted sum of the original ones

$$f_i[S_0^{\rm new}] \approx f_i[S_0] + \sum_k \frac{z_k^{\rm min}}{2\sqrt{\Delta\chi^2}} (f_i[S_k^+] - f_i[S_k^-]), \tag{3.59}$$

that is, the PDFs are *reweighted* in the process. Similarly, one can diagonalize the new Hessian matrix in Eq. (3.57) and find in these new eigendirections the parameter values corresponding to the tolerance $\Delta\chi^2$ to obtain the new error sets and then use Eq. (3.54) to propagate the updated uncertainties into the observables. It should be emphasized that the obtained results are only approximative of those of a full global fit, limited by the approximations made and also restricted by all the assumptions that were made in the original analysis, such as the functional forms assumed.

Including higher-order terms

As discussed at length in the article **[III]**, the Hessian reweighting with the quadratic approximation of $\chi^2_{\rm old}$ and a linear approximation in the predictions

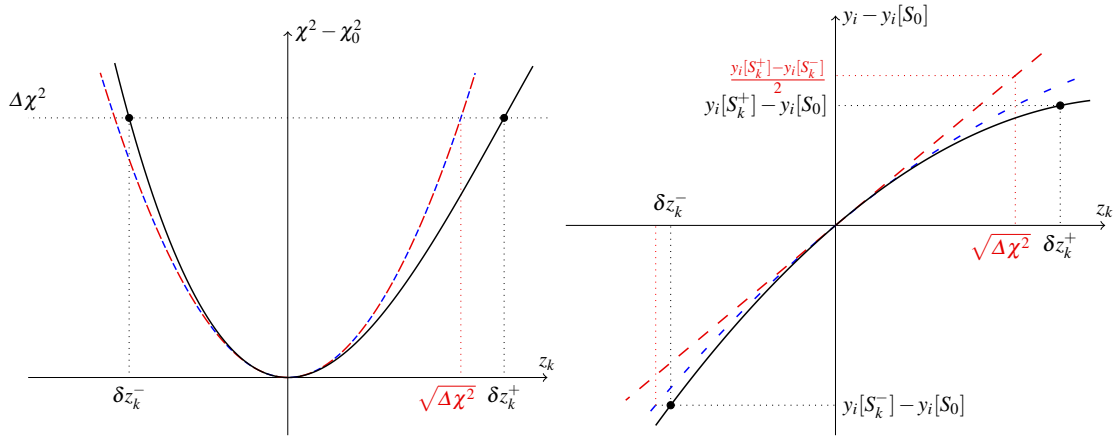

Figure 3.3. Different approximations of the χ^2 function (left) and the PDF-dependent quantities (right) in the reweighting. Figure from article [**III**].

y_i, shown as dashed red lines in Figure 3.3, can be extended to include also higher-order terms. Simply by using only the PDF central and error sets, one can extend Eq. (3.54) to include also quadratic terms, shown with blue dashed lines in Figure 3.3, as derived in Ref. [64]. However, if additional information on the original fit is provided, one can also include cubic terms in the approximation of the original χ^2 function,

$$\chi^2_{\text{old}} \approx \chi^2_0 + \sum_k (a_k z_k^2 + b_k z_k^3), \tag{3.60}$$

with

$$a_k = \frac{\Delta\chi^2}{\delta z_k^+ - \delta z_k^-}\left(\frac{\delta z_k^+}{(\delta z_k^-)^2} - \frac{\delta z_k^-}{(\delta z_k^+)^2}\right), \quad b_k = \frac{\Delta\chi^2}{\delta z_k^+ - \delta z_k^-}\left(\frac{1}{(\delta z_k^+)^2} - \frac{1}{(\delta z_k^-)^2}\right). \tag{3.61}$$

where $\delta z_k^\pm$ are the parameter values determining the error sets in Eq. (3.47). Then, approximating the PDF-dependent quantities with a quadratic function,

$$y_i \approx y_i[S_0] + \sum_k (d_{ik} z_k + e_{ik} z_k^2), \tag{3.62}$$

the coefficients then read

$$\begin{aligned} d_{ik} &= \frac{1}{\delta z_k^+ - \delta z_k^-}\left[-\frac{\delta z_k^-}{\delta z_k^+}\left(y_i[S_k^+] - y_i[S_0]\right) + \frac{\delta z_k^+}{\delta z_k^-}\left(y_i[S_k^-] - y_i[S_0]\right)\right], \\ e_{ik} &= \frac{1}{\delta z_k^+ - \delta z_k^-}\left[\frac{1}{\delta z_k^+}\left(y_i[S_k^+] - y_i[S_0]\right) - \frac{1}{\delta z_k^-}\left(y_i[S_k^-] - y_i[S_0]\right)\right]. \end{aligned} \tag{3.63}$$

This approximation is shown as solid black lines in Figure 3.3. These additions can help improve the accuracy of the method, especially in situations when

uncertainties are large. On a downside, in including these terms the simple quadratic form of χ^2_{new} is lost and the minimization needs to be done numerically.

Chapter 4

Nuclear modifications of partonic structure

As a first approximation, one could think of a nucleus as a loosely bound ensemble of nucleons. There is, however, ample experimental evidence that this simple picture is too crude to explain hard-scattering phenomena and that the partonic structure of the nucleons in nuclei is modified in a nontrivial way. Already from early DIS measurements on deuteron targets it was known that the *Fermi motion* of the bound nucleons increases the probability of finding a parton with a large momentum fraction x_N with respect to the average nucleon momentum. What came as a surprise in DIS experiments with heavy nuclei was that the quark distributions in bound nucleons are suppressed compared to those of a free proton for $0.3 \lesssim x_N \lesssim 0.8$. This phenomenon carries the name *EMC effect* due to its first observation by the European Muon Collaboration (EMC) [66]. Later experiments also revealed an enhancement in the parton content at $0.03 \lesssim x_N \lesssim 0.3$ and a suppression again at $x_N \lesssim 0.03$, nowadays known as *antishadowing* and *shadowing*, respectively.

Over the years, a plethora of models to explain the nuclear effects have appeared, see Refs. [67–71] for reviews. The approach taken in nPDF analyses is, however, rather different. By parametrizing the nPDFs with suitably flexible functions and determining their parameters through a global analysis as described in Chapter 3, one aims to get rid of any model dependence and to obtain a fully data-driven estimate of the nuclear modifications of parton distributions. From these, one can then make model-independent predictions for, e.g., production rates of hard probes of the Quark Gluon Plasma in ultrarelativistic heavy-ion collisions [72] or for ultra-high energy scattering cross sections in neutrino telescopes [73] and importantly also quantify the bias in free-proton PDFs caused by using nuclear data in their fits [74].

The PDFs of different nuclei are, in principle, independent quantities and should be determined from the data nucleus by nucleus, but the present data

are far from sufficient to do so reliably for any single nucleus other than perhaps lead. Therefore, the mass-number dependence is parametrized in the nPDF fits. It is conventional to decompose the PDFs of an average nucleon f_i^A in a nucleus with a mass number A and an atomic number Z as

$$f_i^A(x_N, Q^2) = \frac{Z}{A} f_i^{p/A}(x_N, Q^2) + \frac{A-Z}{A} f_i^{n/A}(x_N, Q^2), \tag{4.1}$$

where $f_i^{p/A}$ is the PDF of a proton bound in a nucleus and $f_i^{n/A}$ the PDF of a bound neutron, with the latter obtained from the first by the approximative isospin symmetry according to Eq. (2.46). With this, one disentangles the isospin effects from other nuclear modifications.

4.1 Nuclear PDF parametrizations

By far the most common way to parametrize the nPDFs is through nuclear modification functions R_i^A, such that at the parametrization scale Q_0^2 the PDFs of a proton bound in a nucleus are defined as

$$f_i^{p/A}(x_N, Q_0^2) = R_i^A(x_N, Q_0^2) f_i^p(x_N, Q_0^2), \tag{4.2}$$

where f_i^p are the PDFs of the free proton. This approach has been adopted by e.g. the EPS09 [75], DSSZ [76], KA15 [77] and EPPS16 [**II**] analyses. An illustration of the R_i^A parametrization of the EPPS16 analysis is given in Figure 4.1. The functional form follows the pattern anticipated by the nuclear effects discussed in the beginning of the chapter, with free parameters controlling the amount of shadowing, antishadowing and EMC-effect and the location of the extremum of the latter two.

The nCTEQ15 [78] and nNNPDF1.0 [79] analyses have taken a different approach and parameterized the bound nucleon PDFs $f_i^{p/A}$ directly, in the case of nCTEQ15 by making the PDF fit function parameters an A-dependent and in nNNPDF1.0 by using a common neural network to parametrize all the nuclei with A as an input to the network. Yet another approach was taken in the nDS analysis [80], where the nPDFs were defined as a convolution

$$f_i^{p/A}(x_N, Q_0^2) = \int_{x_N}^{A} \frac{\mathrm{d}y}{y} W_i^A(y) f_i^p(x_N/y, Q_0^2) \tag{4.3}$$

with suitably parametrized weight functions $W_i^A(y)$. Note that the integration range goes up to A allowing $x_N > 1$. This is perfectly valid, as in the nuclear environment individual partons can borrow momentum from different nucleons. The parton distributions in this region are, however, expected to be very small

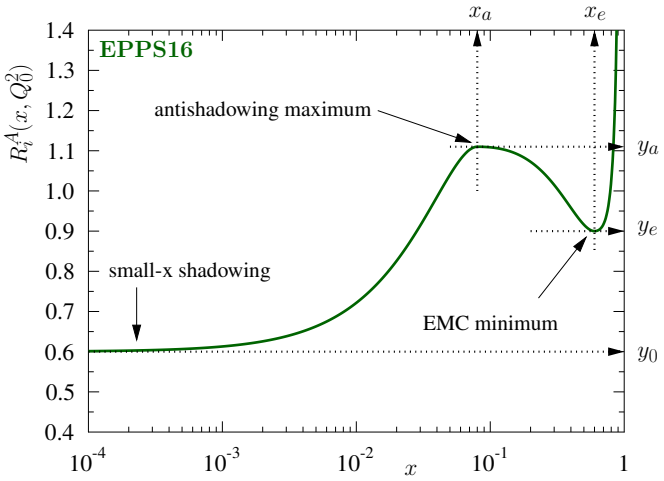

Figure 4.1. The functional form of the PDF nuclear modifications used in the EPPS16 analysis. Figure from article [**II**].

and therefore, and in lack of constraining data, most of the analyses have simply assumed $f_i^{p/A}(x_N > 1) = 0$. With this assumption, $f_i^{p/A}$ follow the same sum rules and evolution equations as the free-proton PDFs and the different nPDF parametrization approaches are practically equivalent.

Table 4.1 summarizes the most recent nPDF global analyses. We list here the perturbative order of the analysis, the included data types, the minimum scale at which data is included and the total number of data points. Further indicated are the number of free parameters, the approach in error analysis, the used free-proton PDFs and heavy-quark mass scheme and the amount of detail in flavour separation. We will discuss the similarities and differences of these analyses further in the following sections.

4.2 Resolving flavour asymmetry

The bulk of the data used in the nPDF global analyses consists of electromagnetic neutral-current DIS measurements. At large x_N, where contributions from sea quarks can be neglected, the per-nucleon structure function F_2^A at leading order reads

$$F_2^A \approx \frac{5}{18} x_N \left[\left(u_\mathrm{V}^{p/A} + d_\mathrm{V}^{p/A} \right) + \frac{3}{5} \left(\frac{2Z}{A} - 1 \right) \left(u_\mathrm{V}^{p/A} - d_\mathrm{V}^{p/A} \right) \right]. \tag{4.4}$$

For isoscalar nuclei the $\frac{2Z}{A} - 1$ prefactoring the valence PDF difference is exactly zero and even for neutron-rich isotopes such as ^{208}Pb it is approximately only

Table 4.1. nPDF releases from the past ten years. Table adapted from Ref. [81].

	EPS09	DSSZ	KA15	nCTEQ15	EPPS16	nNNPDF1.0
Order in α_s	LO & NLO	NLO	NNLO	NLO	NLO	NNLO
lA/ld NC DIS	✓	✓	✓	✓	✓	✓
pA/pd DY	✓	✓	✓	✓	✓	
RHIC dAu/pp π	✓	✓		✓	✓	
νA DIS		✓			✓	
πA DY					✓	
LHC pPb W, Z					✓	
LHC pPb jets					✓	
Q cut in DIS	1.3 GeV	1 GeV	1 GeV	2 GeV	1.3 GeV	1.87 GeV
Data points	929	1579	1479	708	1811	451
Free parameters	15	25	16	16	20	$\sim$183
Error analysis	Hessian	Hessian	Hessian	Hessian	Hessian	Monte Carlo
Error tolerance $\Delta\chi^2$	50	30	not given	35	52	not applicable
Free proton PDFs	CTEQ6.1	MSTW2008	JR09	CTEQ6M-like	CT14	NNPDF3.1
HQ treatment	ZM-VFNS	GM-VFNS	ZM-VFNS	GM-VFNS	GM-VFNS	GM-VFNS
Flavour separation	no	no	no	valence	full	no
Reference	[75]	[76]	[77]	[78]	[**II**]	[79]

-0.2. Hence F_2^A at large x_N is always predominantly sensitive to the sum of the valence quarks making it difficult to constrain the flavour separation. The same happens for the sea quarks at small x_N.

Moreover, the published structure functions are often *isoscalarized*, i.e. reported in terms of $F_2^{\text{isoscalar-}A} = \beta F_2^A$, where the factor

$$\beta = \frac{A}{2}\left(1 + \frac{F_2^n}{F_2^p}\right) \Bigg/ \left(Z + (A-Z)\frac{F_2^n}{F_2^p}\right), \tag{4.5}$$

with the ratio of neutron and proton structure functions F_2^n/F_2^p suitably parametrized, is applied to facilitate an easy comparison with the deuteron structure function, such that $F_2^{\text{isoscalar-}A}/F_2^{\text{D}}$ would be unity if there were no nuclear modifications beyond isospin effects. Unfortunately, this "correction" makes the extraction of flavour separation even more challenging. For the above reasons, most of the nPDF analyses (cf. Table 4.1) have made simplifying assumptions to fix the flavour dependence of the valence quarks, and separately for the sea quarks. The first exception from this rule was the nCTEQ15 analysis, where the valence quarks were parametrized independently. However, in lack of constraining data, the uncertainties on the flavour separation remained large.

4.2.1 Pion–nucleus Drell–Yan as a novel probe

It was suggested in Ref. [82] that by studying pion-induced fixed-target DY data, one would get additional information on the flavour separation in the EMC region. In particular, the cross section ratios

$$R^-_{A_1/A_2}(x_2) = \frac{\frac{1}{A_1}\mathrm{d}\sigma^{\pi^-+A_1}_{\text{DY}}/\mathrm{d}x_2}{\frac{1}{A_2}\mathrm{d}\sigma^{\pi^-+A_2}_{\text{DY}}/\mathrm{d}x_2}, \qquad R^{+/-}_A(x_2) = \frac{\mathrm{d}\sigma^{\pi^++A}_{\text{DY}}/\mathrm{d}x_2}{\mathrm{d}\sigma^{\pi^-+A}_{\text{DY}}/\mathrm{d}x_2} \tag{4.6}$$

were advocated. These are differential in $x_2 = \frac{M}{\sqrt{s}}\mathrm{e}^{-y}$, where $\sqrt{s}$ is the pion–nucleon center-of-mass energy, M is the invariant mass of the lepton pair and y its rapidity in the center-of-mass frame, probing at leading order the momentum fraction of the parton from the nucleus, $x_2 \approx x_N$. Employing the isospin and charge-conjugation relations of Eq. (2.48) and assuming that we are in a kinematic region where the pion sea quarks can be neglected, the leading order expressions for these ratios can be written as

$$R^-_{A_1/A_2}(x_2) \approx \frac{4u^{A_1}(x_2) + \bar{d}^{A_1}(x_2)}{4u^{A_2}(x_2) + \bar{d}^{A_2}(x_2)}, \qquad R^{+/-}_A(x_2) \approx \frac{4\bar{u}^A(x_2) + d^A(x_2)}{4u^A(x_2) + \bar{d}^A(x_2)}. \tag{4.7}$$

This shows that the remaining pion valence PDFs, which are not well known, cancel in these ratios, making them potential probes of the PDF nuclear modifications. Due to having valence anti-quarks in the pions, these ratios probe different

flavour combinations than the DIS structure functions. In particular in the region $x_2 \gtrsim 0.1$, we have

$$R^{-}_{A/\mathrm{D}} \approx \frac{u_{\mathrm{V}}^{p/A} + d_{\mathrm{V}}^{p/A}}{u_{\mathrm{V}}^{p} + d_{\mathrm{V}}^{p}} + \left(\frac{2Z}{A} - 1\right) \frac{u_{\mathrm{V}}^{p/A} - d_{\mathrm{V}}^{p/A}}{u_{\mathrm{V}}^{p} + d_{\mathrm{V}}^{p}}, \tag{4.8}$$

where we notice a factor $5/3$ increase in the sensitivity to the flavour separation compared to that in the F_2^A in Eq. (4.4).

Article [**I**] discusses in detail the applicability and prospects of using the existing measurements of these observables in the nPDF global analysis. An important check was to make sure that the cancellation of pion degrees of freedom in the ratios would work also beyond leading order. This is shown in Figure 4.2, where NLO calculations performed with the public MCFM code [39] using pion PDFs from GRV [83] and SMRS [84] analyses are compared with measurements from the NA3 [85], NA10 [86] and E615 [87] experiments. The NA10 data have been published with a similar isospin correction applied to them as discussed above and this had to be taken into account in the calculations. As can be seen from the figure, the cancellation of the pion PDFs is extremely good, and hence these ratios are insensitive to the rather poorly known pion structure.

Figure 4.3 compares these data with NLO calculations using the nCTEQ15 and EPS09 nuclear PDFs. The NA10 data have a large normalization uncertainty, which was treated by normalizing predictions from each PDF set according to the optimal normalization found with Eq. (3.29) (with the PDF parameters kept fixed, there is no danger of d'Agostini bias in this case). The overall agreement between data and theory is rather good, which shows that these data can be used in a nPDF global fit.

Since valence flavour separation was allowed in the nCTEQ15 analysis, the related uncertainty bands in Figure 4.3 are larger than those in EPS09. To study this in more detail, we show also the results with the nCTEQ15 error sets 25 and 26, which have the largest and smallest flavour asymmetry, respectively. The error set 25 of nCTEQ15 shows a flatter x_2 dependence than that in the NA3 and NA10 data, while predictions with nCTEQ15 error set 26 and EPS09 central set are in perfect agreement with the measurements. This hints towards similarity of valence-quark nuclear modifications, but as the experimental uncertainties are large, more stringent constraints are clearly needed.

4.2.2 Global analysis with full flavour separation

In article [**II**], we provided the first nPDF global analysis with full flavour separation, EPPS16, using the above-mentioned pion–nucleus DY data and other observables to constrain the valence and sea quark asymmetries. A very good fit to the pion DY data was found, shown as an example for NA10 data in Figure 4.4,

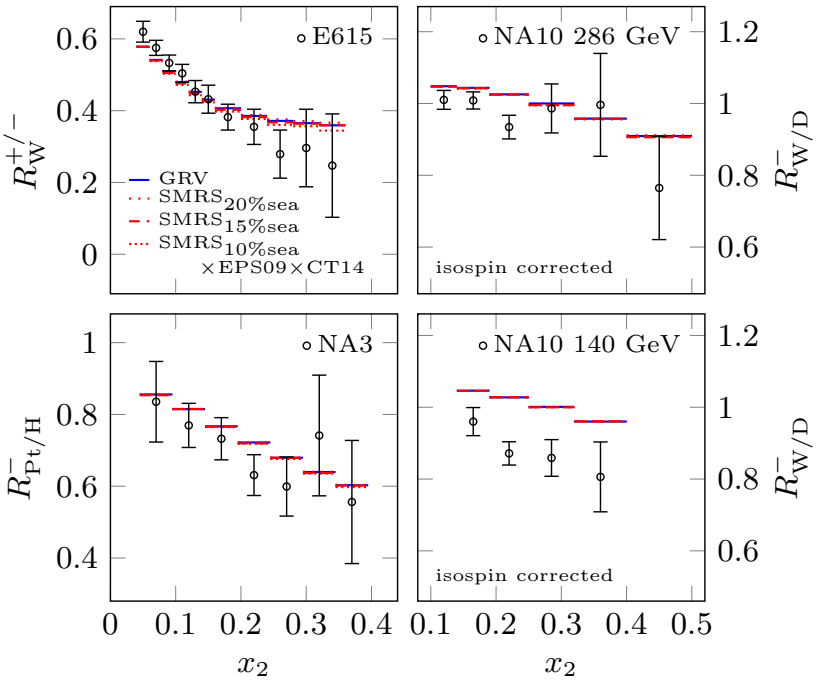

Figure 4.2. The cancellation of pion PDFs in pion-induced Drell–Yan cross-section ratios at next-to-leading order. Figure from article **[I]**.

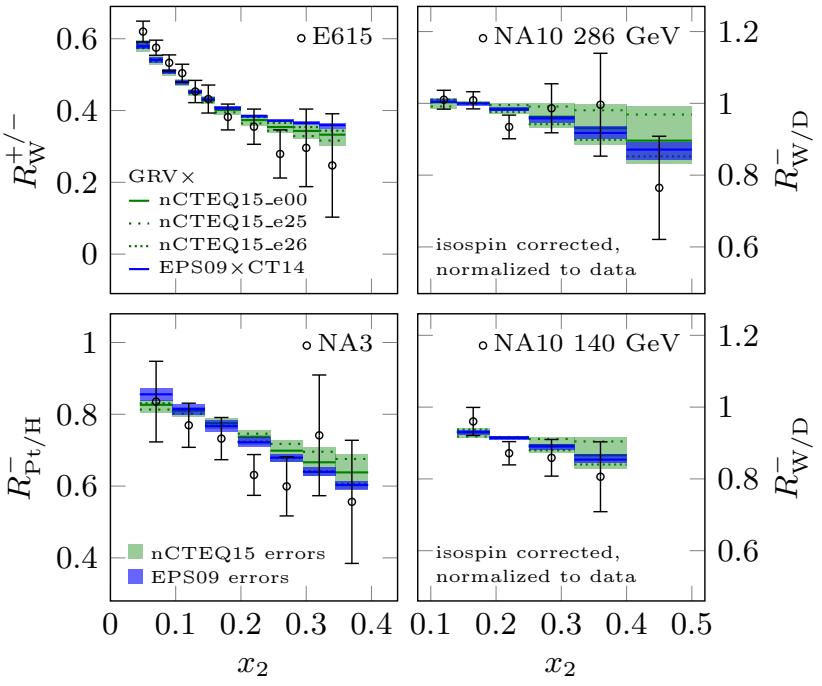

Figure 4.3. A comparison of measured pion-induced Drell–Yan cross-section ratios with predictions using the nCTEQ15 and EPS09 nuclear PDFs. Figure from article **[I]**.

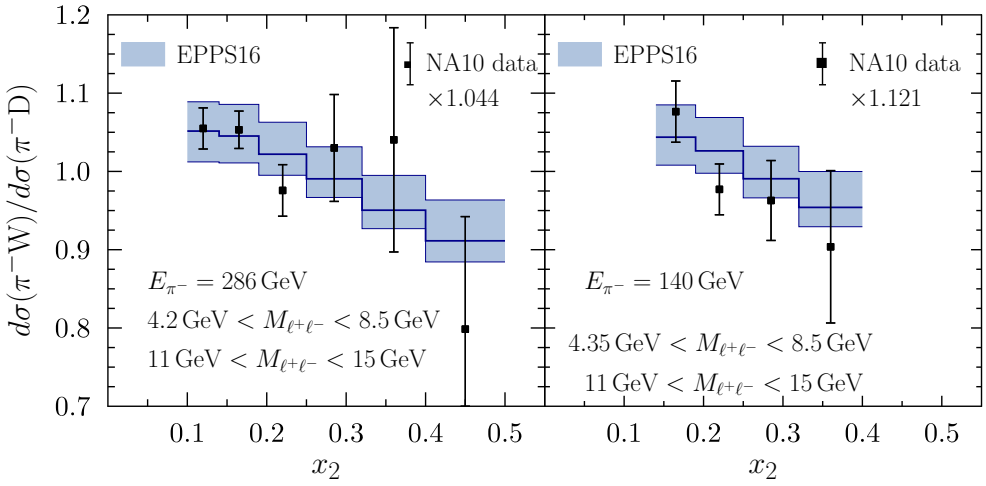

Figure 4.4. NA10 pion–tungsten DY data compared with the EPPS16 fit results. Figure from article **[II]**.

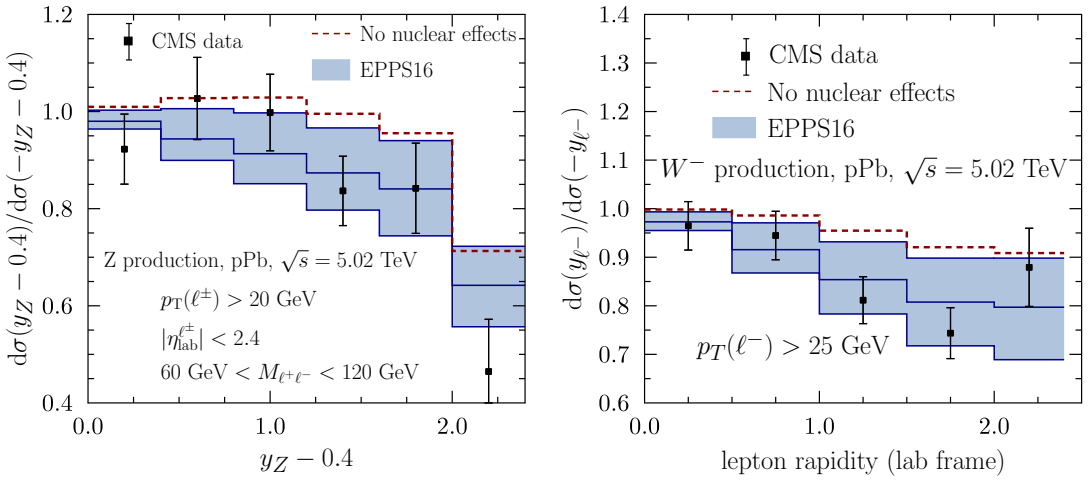

Figure 4.5. CMS proton–lead Z and W^- data compared with the EPPS16 fit results. Figure from article **[II]**.

but due to large data uncertainties the impact in the fit was somewhat limited. Also new in this analysis, by using published isoscalar-correction factors β of the charged-lepton DIS data, the non-isoscalarized ratios F_2^A / F_2^{D} were obtained from the "corrected" ones, gaining enhanced sensitivity to the flavour separation compared to a case where the fit would be simply considered to be done on isoscalar targets.

The EPPS16 analysis was the first to use data from LHC proton–lead collisions,

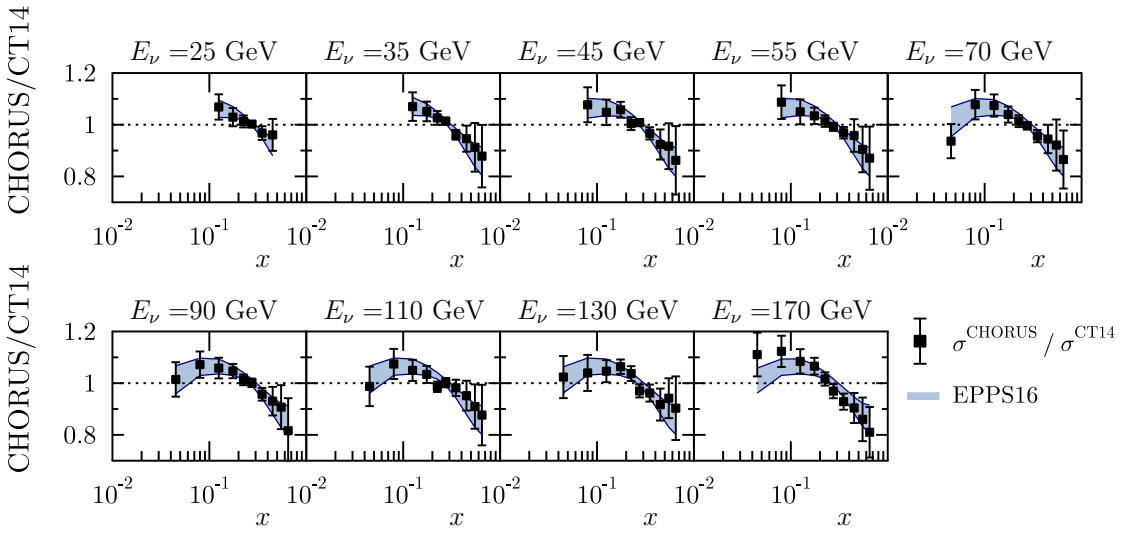

Figure 4.6. Normalized CHORUS neutrino DIS data compared with the results of the EPPS16 analysis. Figure from article **[II]**.

including W and Z production data from CMS and ATLAS experiments [88–90]. In principle, these observables are sensitive to different flavour combinations than the neutral-current DIS and DY data and could help constrain the flavour separation. However, as at the time no proton–proton baseline measurements were available, these data were added in the analysis as forward-to-backwards ratios, where the differential cross sections at forward rapidities $\mathrm{d}\sigma(y)$ are divided with those at backward rapidities $\mathrm{d}\sigma(-y)$ to reduce the uncertainties arising from the applied free-proton PDFs. Figure 4.5 shows these observables for the Z and W^- production with a comparison of the CMS data and EPPS16 fit results. As can be seen from the figure, the agreement is very good and supports nuclear modifications of the PDFs, namely shadowing, in the region $x \lesssim 0.1$ probed by the data. However, because of low statistics, the data did not give strong constraints. Moreover, as these data probe the PDFs at high Q^2, any small differences at the parametrization scale are likely to be hidden under a large flavour-symmetric sea component generated through $g \to q\bar{q}$ splittings in the scale evolution, hindering the potential constraints for flavour separation. This, and also direct contribution from quark–gluon scattering at NLO make these data sensitive to also gluon nuclear modifications, discussed in more detail in the context of the EPPS16 analysis in Section 4.3.

More stringent constraints were obtained from CHORUS neutrino and antineutrino DIS data [91]. These data were included already in the DSSZ analysis [76], but as no flavour separation was allowed in the fit, their constraining potential was not fully utilized. In EPPS16, these data were self-normalized at each beam-energy bin according to a procedure introduced in Ref. [63] to deal

with data normalization uncertainty, propagating the data correlations, given in terms of the systematic shifts discussed in Section 3.2, consistently to the normalized cross sections. Figure 4.6 shows the normalized neutrino-beam data, again in comparison with the EPPS16 results, divided with predictions with no nuclear effects in the PDFs to ease the interpretation. These data had a large impact in the fit, giving the u and d valence quark modifications a similar shape.

Figure 4.7 compares the quark nuclear modifications of the two analyses which allow flavour separation in the fits. The uppermost two panels show the average valence and light sea quark modifications in lead,

$$R^{\mathrm{Pb}}_{u_\mathrm{V}+d_\mathrm{V}} = \frac{u_\mathrm{V}^{p/\mathrm{Pb}} + d_\mathrm{V}^{p/\mathrm{Pb}}}{u_\mathrm{V}^{p} + d_\mathrm{V}^{p}}, \qquad R^{\mathrm{Pb}}_{\bar{u}+\bar{d}+\bar{s}} = \frac{\bar{u}^{p/\mathrm{Pb}} + \bar{d}^{p/\mathrm{Pb}} + \bar{s}^{p/\mathrm{Pb}}}{\bar{u}^{p} + \bar{d}^{p} + \bar{s}^{p}}, \tag{4.9}$$

at the scale $Q^2 = 10\ \mathrm{GeV}^2$. As should be expected, the EPPS16 and nCTEQ15 analyses are well constrained and agree nicely in the region $x \gtrsim 10^{-2}$ where data from fixed target DIS and DY are available. Within the uncertainties, we can clearly state that the valence quarks exhibit antishadowing and EMC effect and that shadowing for both valence and sea quarks seems to be preferred.

When we compare the modifications for individual valence quarks on the second row of Figure 4.7, we find a large difference in the results. For EPPS16 the u_V and d_V are very similar, driven by the CHORUS data and also consistently with the pion–nucleus DY and CERN EW boson data, while for nCTEQ15, where no valence-quark constraints beyond neutral-current DIS and DY were included, the fit shows a large flavour asymmetry. For both EPPS16 and nCTEQ15 we find a narrow throat in the uncertainties, which is likely a fit-function artefact as in EPPS16 this happens at $x \approx 0.03$ where there are no data constraints from CHORUS. Even with the new constraints included, the EPPS16 uncertainties remain much larger for the individual flavours than for the average modification, simply reflecting the fact that the approximate isoscalarity of most nuclei makes it difficult to constrain the asymmetry. The same applies to the sea quarks, shown in the next three panels, where the nuclear modifications for all flavours are qualitatively similar and the EPPS16 and nCTEQ15 fits are in agreement. As flavour separation was allowed in EPPS16, the uncertainties of the individual flavours are larger than for the average sea-quark combination and also larger than in nCTEQ15 where the sea quarks are related to each other in a fixed way.

While the nNNPDF1.0 analysis uses only neutral-current DIS in the fit and is thus not yet in a fully global footing, the used methodology is somewhat different compared to other analyses, and hence it is interesting to compare EPPS16 and nCTEQ15 also with the results from this analysis. The nNNPDF1.0 analysis uses Monte Carlo sampling of PDFs [92, 93], which allows for a more reliable uncertainty estimation than the Hessian method in regions poorly constrained by the data. Since in lack of DY data no discrimination between valence and sea quarks

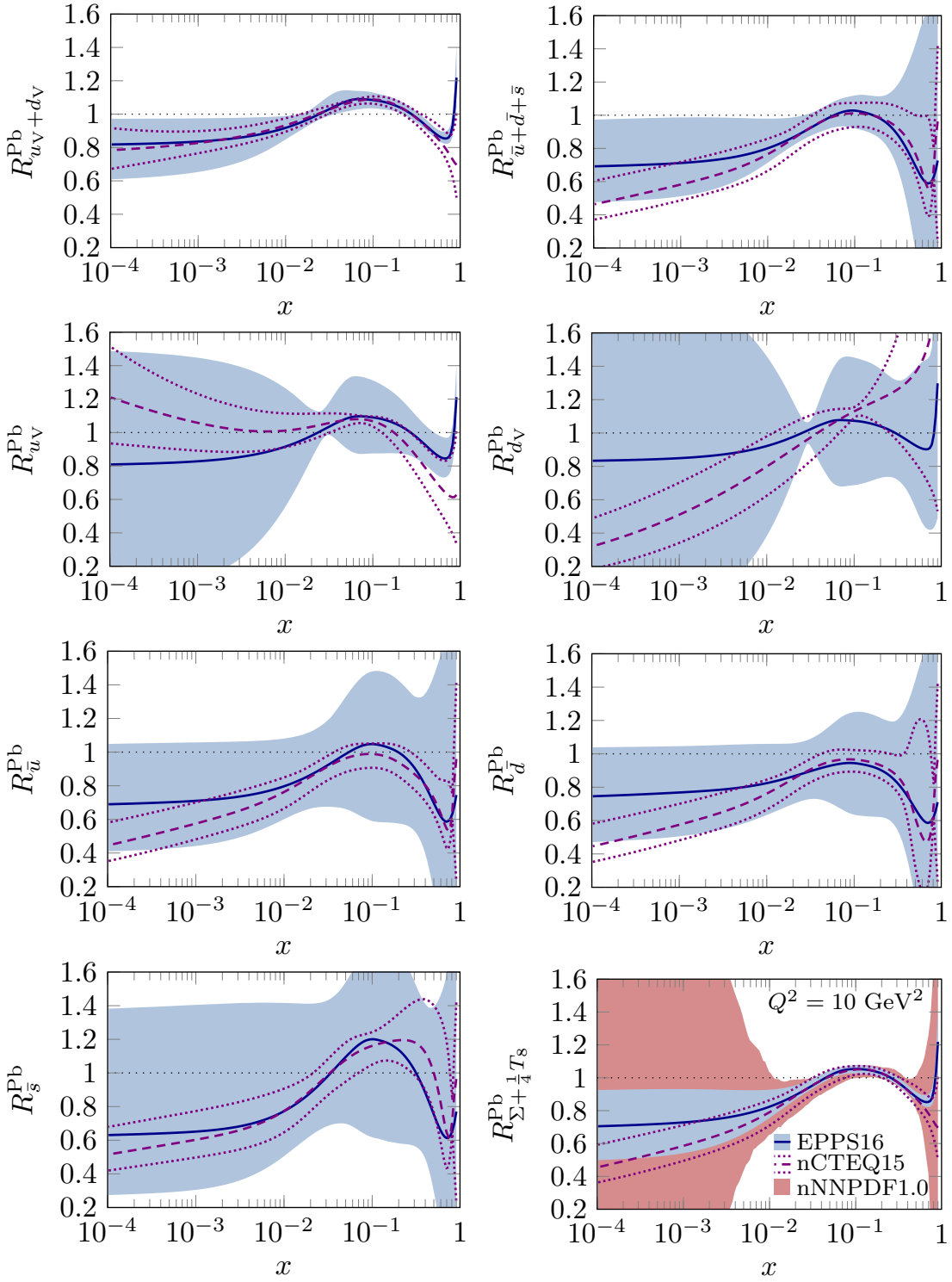

Figure 4.7. Comparison of quark nuclear modifications at the scale $Q^2 = 10\ \mathrm{GeV}^2$ in lead nucleus as found in the EPPS16, nCTEQ15 and nNNPDF1.0 analyses.

was possible in the nNNPDF1.0 fit, the only meaningfully constrained quark combination in their analysis, also bearing to the approximate isoscalarity of the targets, is, with the evolution basis variables $\Sigma = u^{p/A} + \bar{u}^{p/A} + d^{p/A} + \bar{d}^{p/A} + s^{p/A} + \bar{s}^{p/A}$ and $T_8 = u^{p/A} + \bar{u}^{p/A} + d^{p/A} + \bar{d}^{p/A} - 2(s^{p/A} + \bar{s}^{p/A})$ according to the nomenclature in Ref. [79],

$$\Sigma + \frac{1}{4}T_8 = \frac{5}{4}(u_{\mathrm{V}}^{p/A} + d_{\mathrm{V}}^{p/A}) + \frac{5}{2}(\bar{u}^{p/A} + \bar{d}^{p/A}) + \frac{1}{2}(s^{p/A} + \bar{s}^{p/A}). \tag{4.10}$$

Hence, we compare nNNPDF1.0 with EPPS16 and nCTEQ15 only in this combination, shown in the bottom-right panel of Figure 4.7 for the lead nucleus. The three analyses agree nicely in the region constrained by the DIS data, but at small x the nNNPDF1.0 uncertainties are vastly larger than those of EPPS16 and nCTEQ15. The EPPS16 and nCTEQ15 small-x uncertainty bands should be understood as an extrapolation of those at higher x, through the assumed form of the fit functions motivated by low-Q^2 nuclear DIS data [94] as well as requiring consistent A-systematics of nuclear effects. Studies with more flexible parametrizations within the EPPS16 framework, leading to similar inflation in small-x uncertainties as seen with the nNNPDF1.0, can be found in Ref. [95].

4.2.3 New observable for future pion–nucleus experiments

Even though the pion–nucleus DY data were not able to give stringent constraints in the EPPS16 fit, the increased sensitivity to the flavour separation makes these processes a potential probe in future experiments. To this end, we have proposed in Ref. [96] a new observable

$$R^{\Delta}_{A_1/A_2}(x_2) = \frac{\frac{1}{A_1}(\mathrm{d}\sigma_{\mathrm{DY}}^{\pi^- + A_1}/\mathrm{d}x_2 - \mathrm{d}\sigma_{\mathrm{DY}}^{\pi^+ + A_1}/\mathrm{d}x_2)}{\frac{1}{A_2}(\mathrm{d}\sigma_{\mathrm{DY}}^{\pi^- + A_2}/\mathrm{d}x_2 - \mathrm{d}\sigma_{\mathrm{DY}}^{\pi^+ + A_2}/\mathrm{d}x_2)}. \tag{4.11}$$

To leading-order accuracy, the contributions involving pion sea quarks cancel in the differences and we have

$$R^{\Delta}_{A_1/\mathrm{D}} \approx \frac{u_{\mathrm{V}}^{p/A} + d_{\mathrm{V}}^{p/A}}{u_{\mathrm{V}}^{p} + d_{\mathrm{V}}^{p}} + \frac{5}{3}\left(\frac{2Z}{A} - 1\right)\frac{u_{\mathrm{V}}^{p/A} - d_{\mathrm{V}}^{p/A}}{u_{\mathrm{V}}^{p} + d_{\mathrm{V}}^{p}} \tag{4.12}$$

at *all* x_2. Note that there is yet another factor $5/3$ increase in the sensitivity to the flavour separation compared to $R^{-}_{A/\mathrm{D}}$ in Eq. (4.8). Figure 4.8 shows this ratio for Pb/D and W/H computed in NLO. The errors calculated with EPS09, where no flavour separation was allowed, are rather small, whereas the EPPS16 and nCTEQ15 predictions have large uncertainties and somewhat different shape, showing that this observable could discern the differences in the nPDFs. There exist now plans to measure this ratio for W/C in a future experiment at the CERN M2 beam line [97], with projections showing possible discriminating power over the nPDFs in the region $x \gtrsim 0.2$.

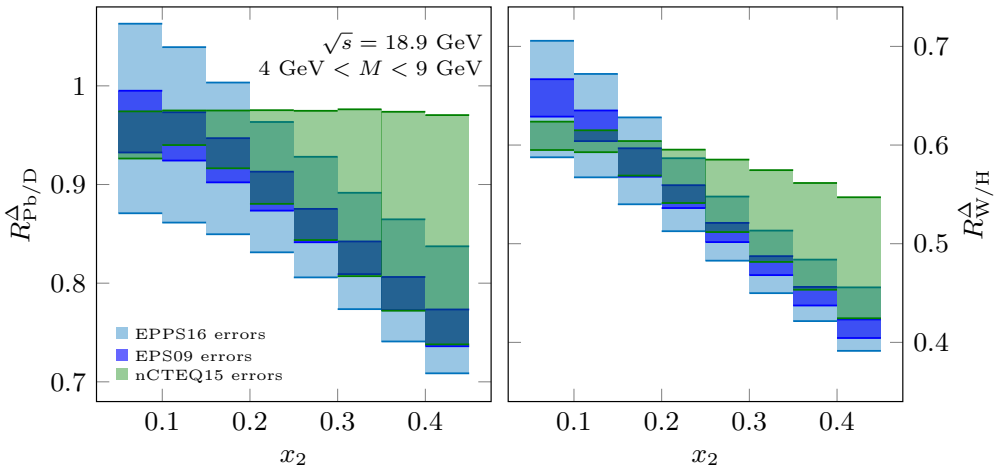

Figure 4.8. New pion–nucleus Drell–Yan observable for constraining the flavour asymmetry in valence quark nuclear modifications. Figure from Ref. [96].

4.3 New constraints for gluon nuclear modifications

Prior to EPPS16, direct gluon constraints were obtained only from RHIC inclusive pion-production data [98, 99] and indirect constraints mainly through the Q^2-dependence of DIS structure functions. The interpretation of the RHIC data is, however, not completely unambiguous. While the EPS08 [100], where these data were used for the first time, and the later EPS09 and nCTEQ15 analyses used these data under the assumption that the observed nuclear effects would only come from the nuclear modifications of the PDFs, the DSSZ analysis employed nuclear modified fragmentation functions. As a result, the DSSZ analysis finds very small gluon PDF nuclear modifications compared to the other analyses. New data were therefore needed to settle the issue.

In the EPPS16 analysis, described in the article **[II]** of this thesis, dijet data from CMS measurement at 5.02 TeV proton–lead collisions [101] were utilized for the first time. The EPPS16 fit results are compared with the data and predictions from nCTEQ15 and DSSZ in the left-hand-side panel of Figure 4.9. As is evident from the figure, the dijet data disagrees with DSSZ, whereas EPPS16 with nuclear modifications in the gluon PDF and no modifications on the fragmentation functions (with the KKP fragmentation functions [102] used in the analysis) finds a good agreement with both the dijet data and the PHENIX pion data shown in the right-hand-side panel.

Figure 4.10 shows again the EPPS16, nCTEQ15 and nNNPDF1.0 nuclear modifications, now for comparison of the gluon PDFs. For the nNNPDF1.0 fit, where no direct gluon constraints were included, the uncertainties are large at all x values. This emphasizes the importance of collider data in constraining

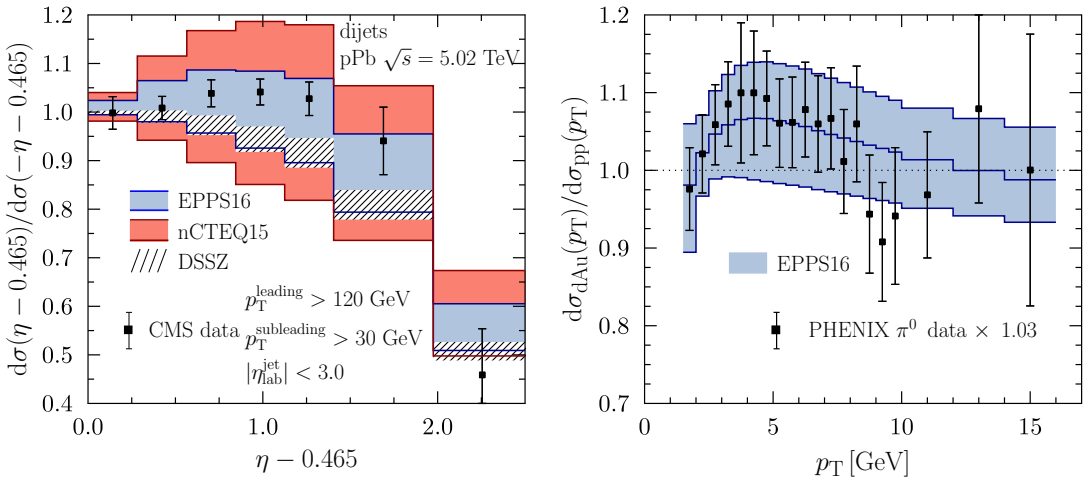

Figure 4.9. Left: Forward-to-backward ratio of the dijet production data measured by CMS compared with the EPPS16 fit results and predictions using the nCTEQ15 and DSSZ nPDFs. Right: PHENIX inclusive pion-production data and the EPPS16 fit. Figures from article [**II**].

the gluons. With the RHIC pion data included, the nCTEQ15 analysis was able to establish an antishadowing pattern, but the gluon modifications remained otherwise largely unconstrained, which also leads to the large uncertainties seen in Figure 4.9. In the EPPS16 analysis gluons are much better constrained due to the inclusion of the CMS dijet data. In particular, these data have a preference for an EMC-type slope for the gluon modification in lead. It should be also noted that the EW-boson data discussed in Section 4.2.2 are, to some extent, sensitive to the gluon PDF and seem to be consistent with the onsetting of small-x gluon shadowing. At very small x, however, the EPPS16 uncertainties should again be understood as an extrapolation, fixed by the momentum sum rule and the assumed fit function form.

4.3.1 Nuclear modification ratio of dijet spectra

In lack of a corresponding proton–proton baseline, the CMS dijet data were included in EPPS16 again as forward-to-backward ratios to reduce the sensitivity to the free-proton PDFs. Thus, again some information were lost and the full potential of a dijet measurement in proton–lead was not fully unleashed. The subsequent 5.02 TeV proton–proton data taking, allowing CMS to provide a measurement of the nuclear modification factor of dijet spectra [103], was thus very fortunate for nPDF analyses. The new CMS dijet data are provided as ratios of self-normalized rapidity distributions differential in η_{dijet} in bins of average

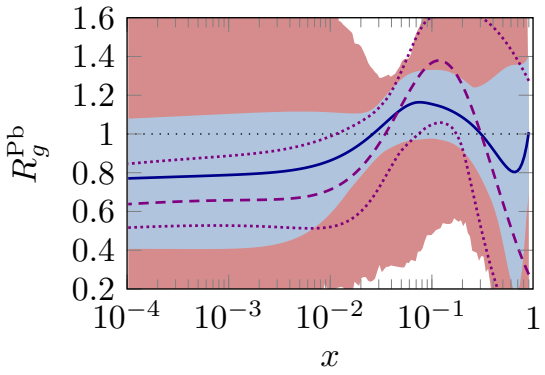

Figure 4.10. Comparison of the gluon nuclear modifications at the scale $Q^2 = 10\ \mathrm{GeV}^2$ in lead nucleus as found in the EPPS16, nCTEQ15 and nNNPDF1.0 analyses. Colours and line explanations are the same as in Figure 4.7.

transverse momentum $p_{\mathrm{T}}^{\mathrm{ave}}$ of the jet pair,

$$R_{\mathrm{pPb}}^{\mathrm{norm.}} = \frac{\frac{1}{\mathrm{d}\sigma^{\mathrm{p+Pb}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}}\,\mathrm{d}^2\sigma^{\mathrm{p+Pb}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}\mathrm{d}\eta_{\mathrm{dijet}}}{\frac{1}{\mathrm{d}\sigma^{\mathrm{p+p}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}}\,\mathrm{d}^2\sigma^{\mathrm{p+p}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}\mathrm{d}\eta_{\mathrm{dijet}}}. \tag{4.13}$$

This is advantageous due to the cancellation of the normalization uncertainty arising from imprecisions in the luminosity determination and also for the cancellation of hadronization effects, separately for proton–proton and proton–lead.

The expected impact of these data on the EPPS16 nPDFs were studied in article **[III]** with the Hessian PDF reweighting outlined in Section 3.4 including the higher order terms. The original NLO predictions for this observable, produced using the NLOJet++ [40] code, are shown in comparison with the data in the upper panels of Figure 4.11. Compared to EPPS16, the data has much smaller uncertainties and strong additional constraints can be expected from including these data in a nPDF global analysis. Further, as shown in the figure, the scale and proton-PDF uncertainties cancel very effectively in the ratio, making this observable an efficient probe of the PDF nuclear modifications. This is fortunate since, as studied extensively in article **[III]**, the proton–lead spectra before taking the ratio with proton–proton baseline have large uncertainties from the proton PDFs, preventing a clean extraction of nPDFs directly from the spectra.

The data are very precise and the uncertainties are systematics dominated, but unfortunately the correlations have not been published, and hence the statistical and systematical uncertainties had to be simply added in quadrature in the reweighting. The impact on the predictions is quantified in lower panels of Figure 4.11, where the results with EPPS16 before and after the reweighting are shown. There is a substantial reduction in the uncertainties, showing that these data can place tight constraints on nPDFs. At forward rapidities, $\eta_{\mathrm{dijet}} \gtrsim 2$, where

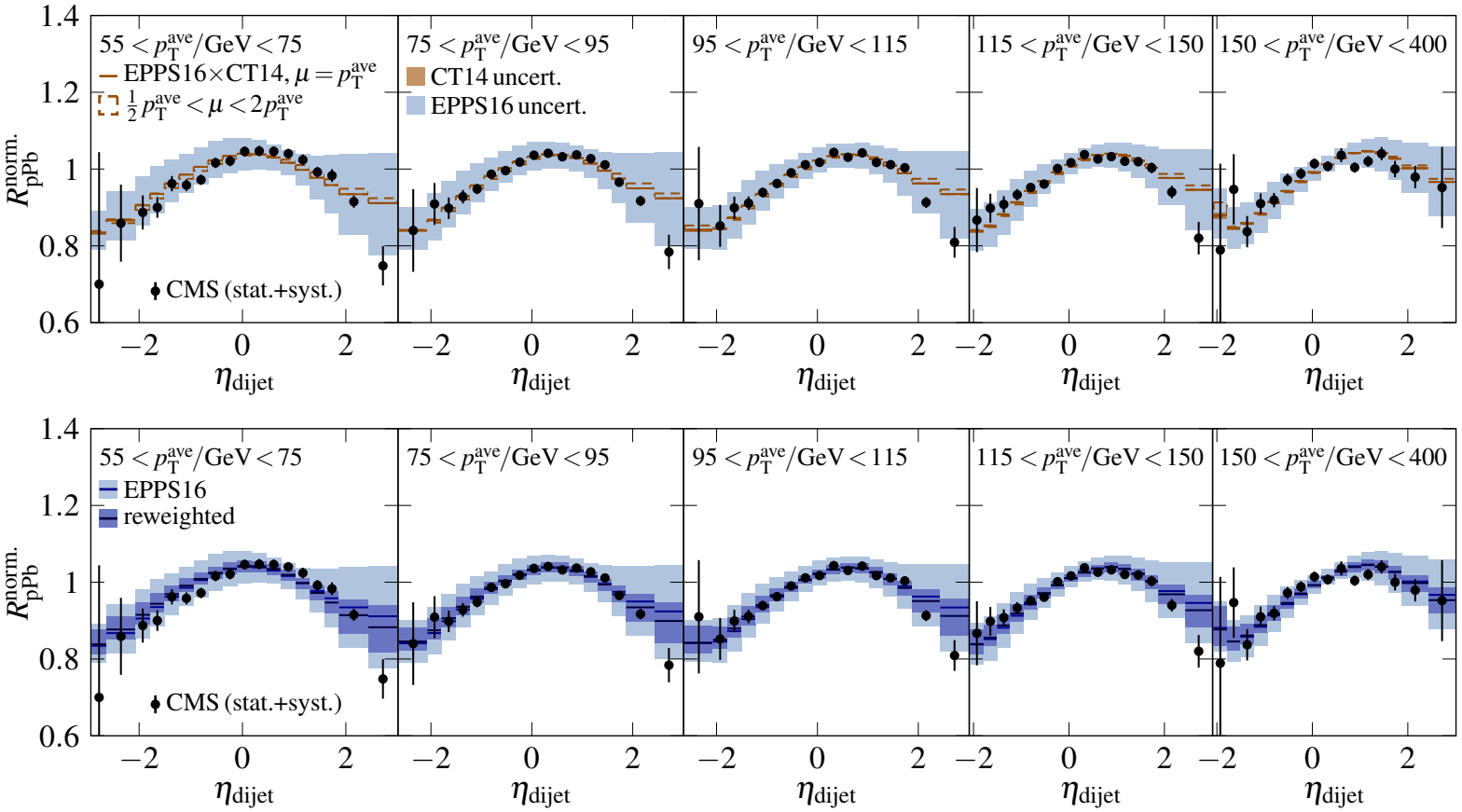

Figure 4.11. Top: Comparison of the CMS data on self-normalized dijet nuclear modification factor with NLO predictions using EPPS16 nPDFs. The hollow dashed boxes indicating the scale uncertainty partially overlap with the central predictions. Bottom: Results of reweighting the EPPS16 nPDFs with the CMS dijet data. Figures from article [**III**].

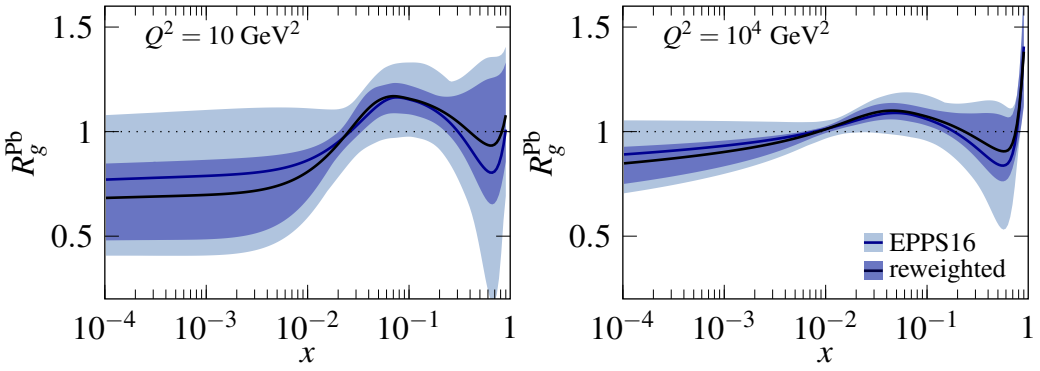

Figure 4.12. Impact on the EPPS16 gluon modifications at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 upon reweighting with the CMS dijet $R_{\mathrm{pPb}}^{\mathrm{norm.}}$ measurement. Figure from article [**III**].

the data points lie systematically below the original EPPS16 central prediction, a downward pull is observed, indicating a preference for deeper gluon shadowing.

The reweighting has a drastic effect on the EPPS16 gluon uncertainties, which shrink by a large factor throughout the probed x range, as shown in Figure 4.12. The most stringent constraints are put on the mid-x region, where the resulting uncertainty is reduced to less than half of its original size. With the reweighted modifications exceeding unity in this region, we seem to be able to confirm the existence of gluon antishadowing. Similarly, at small x the uncertainty band lies below one, supporting gluon shadowing.

Even with an enhanced shadowing in the central set compared to the original EPPS16, the fit has trouble in reproducing the most forward data points. Such a steep decrease as seen in the data in going from the second-to-most-forward data points to the most forward ones can be expected to be hard to come by in any global fit as the gluon modifications probed in this high-Q^2 region are smoothed by the scale evolution, as can be seen in Figure 4.12. It is thus essential to have also other forward data to tell whether the drop in the data is a real physics effect, or perhaps caused by the systematic uncertainties. Further constraints are also needed in the high-x region, where the uncertainties are are reduced compared to EPPS16 before reweighting, but still large enough such that we cannot yet confirm EMC effect for gluons. While we have here studied only the constraints on EPPS16, the impact on nCTEQ15 and nNNPDF1.0, with larger uncertainties to begin with, would be even more dramatic.

4.3.2 Small-x constraints with D-meson production

While the normalized dijet nuclear modification factor discussed in the previous section proved to be an efficient probe of the gluon nuclear modifications, the

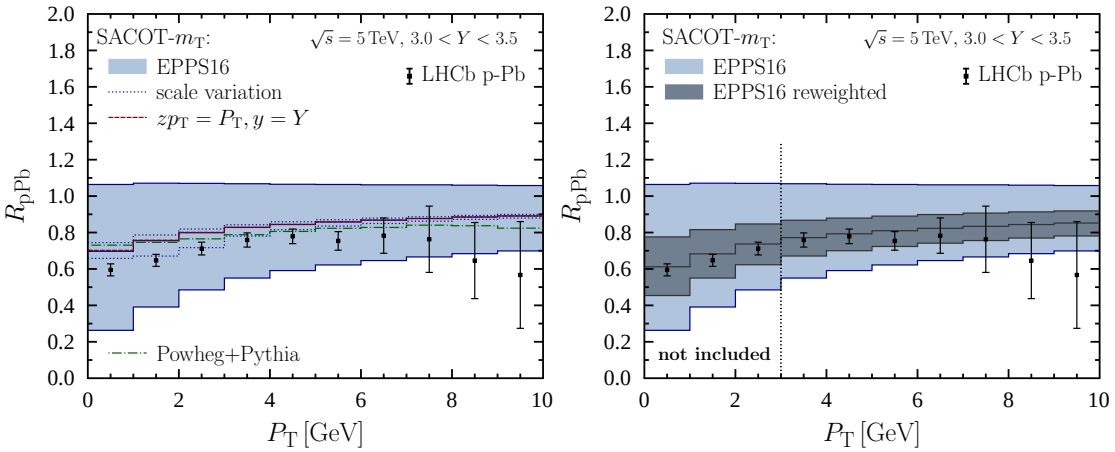

Figure 4.13. Data-to-theory comparison of inclusive D^0-meson production at forward rapidity (left) and the effect of reweighting on EPPS16 predictions in this bin (right). Figures from article [**V**].

x-reach of the CMS measurement goes only down to about $2 \cdot 10^{-3}$ in the lowest p_T^{ave} bin, leaving the region of very small x still unconstrained. To study the low-x region, the use of inclusive forward production of D mesons has been proposed e.g. in Ref. [104]. This process and its measurement at the LHCb experiment [105] are discussed in the light of nPDFs in article [**V**]. While these data have been studied previously [106, 107], either a direct evaluation of the impact on nPDFs has not been given, or if done, then using a less rigorous theoretical framework. Thus, article [**V**] provides the first fully QCD-based estimate of the impact of the LHCb D^0 data on nPDFs. More precisely, the analysis is performed in the SACOT-m_T [44] scheme of GM-VFNS with KKKS08 fragmentation functions [108] and using the Hessian PDF reweighting method.

The LHCb measurements are given in terms of nuclear modification factors,

$$R_{pPb}^{D^0}(P_T, Y) = \frac{\frac{1}{208} d\sigma^{p+Pb \to D^0+X}/dP_T dY}{d\sigma^{p+p \to D^0+X}/dP_T dY}, \tag{4.14}$$

where P_T is the transverse momentum of the measured D^0 and Y its rapidity. The left-hand-side panel of Figure 4.13 shows this ratio in a bin of forward rapidity, $3.0 < Y < 3.5$. Again, the scale uncertainties are found to cancel to a large extent in the ratio. Still, at $P_T < 3$ GeV these uncertainties begin to grow and, due to the minimum scale $Q = 1.3$ GeV in the EPPS16 PDFs, are potentially even underestimated in this region. To avoid possible bias, it is therefore safest not to include the $P_T < 3$ GeV data in a nPDF fit.

Even with this cut in place, the LHCb data are able to constrain nPDFs down to $x \approx 10^{-5}$. To study the possible impact on nPDFs in detail, we have performed

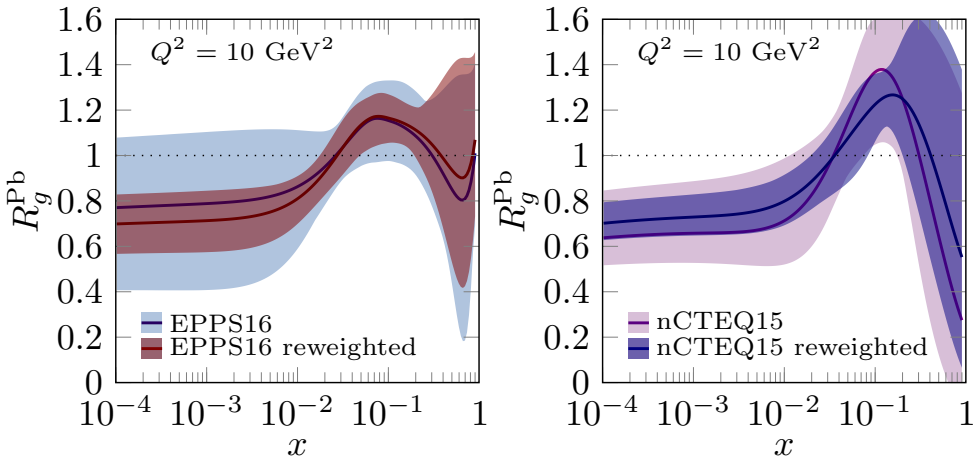

Figure 4.14. Impact of LHCb inclusive D^0-meson data on EPPS16 and nCTEQ15 nPDFs in Hessian reweighting. Figures from article [**V**].

a similar reweighting analysis as was done with the dijet data, here for both EPPS16 and nCTEQ15. The right-hand-side panel of Figure 4.13 shows the resulting change in the EPPS16 predictions in the $3.0 < Y < 3.5$ bin, where a large reduction in the EPPS16 uncertainties are found. At backward rapidities the reduction is not as large, but still significant. Interestingly, throughout the data range, also the data below the 3 GeV cut agree with the reweighted predictions, supporting the validity of collinear factorization down to $P_{\mathrm{T}} = 0$ GeV in this process. No need for including nuclear modifications of fragmentation functions is found here either.

Figure 4.14 shows the impact on the EPPS16 and nCTEQ15 gluon nuclear modifications. The similarity with the results obtained in reweighting EPPS16 with the dijet data is striking (see Figure 4.12), lending further support for the process independence of nuclear PDFs. We also find that the assumed parametrization in EPPS16 is not too restrictive and can describe both data simultaneously. While the mid-x constraints from D^0-meson data are somewhat less restrictive than those from dijets, at small x significant further constraints are obtained, not only because the resulting uncertainty bands are smaller, but also more importantly since the data constraints extend to significantly lower x. The next generation nPDFs with both of these data included in the analysis can thus be expected to have the gluon modifications constrained with previously inaccessible precision.

4.3.3 Multi-observable approach with RHIC

While the BNL-RHIC provided the first direct constraints for the nuclear gluon PDFs, no further measurements have found their way to nPDF fits yet, even though the potential of the collider with its flexible beam line to provide nPDF

constraints is indisputable. In article [**IV**], we have provided a systematic study on the prospects of a simultaneous analysis on multiple observables to constrain the nPDFs, revolving around the potential of the proposed forward upgrades of the STAR and sPHENIX experiments [109, 110].

The left-hand-side panels of Figure 4.15 show the pseudodata for DY nuclear modification factor generated with expected luminosities and efficiencies at the sPHENIX forward-arm upgrade. The data points at low dilepton invariant masses have smaller projected statistical uncertainties, shown as error bars, than the uncertainties from predictions with the EPPS16 nPDFs, which would promise new constraints on the nPDFs. However, on top of the statistical uncertainties, we are expecting a normalization uncertainty of the order of 4 percent, stemming from uncertanties in the luminosity determination. A reweighting performed with these pseudodata, "EPPS16+DY" in Figure 4.15, thus finds barely any impact on the PDFs.

It is worth to note that this is also a situation where the d'Agostini bias can become potentially dangerous. As the EPPS16 nPDFs and the pseudodata generated from them have a rather flat x dependence, any alteration in the predictions could be compensated by a respective change in the data normalization. Then, if the χ^2 function from Eq. (3.29) were used, there would be a bias favoring smaller normalization (and thus enhanced shadowing). For this reason, we have used instead the unbiased χ^2 function in Eq. (3.39). The flatness of the data also prevents using a similar self-normalization trick as was used in Section 4.3.1 to treat the normalization of dijets.

Now, the idea of the multi-observable approach is that the luminosity uncertainty is correlated over all measurements using the same proton–gold and proton–proton run statistics. Thus, if one includes in the fit data from a better constrained region, say, dijets at central rapidity, this would constrain the normalization also in the less-constrained small-x region. This is illustrated in the right-hand-side panels of Figure 4.15, where now the combined "EPPS16+DY+dijet" fit, with normalization uncertainty correlated between the DY and dijet pseudodata, achieves a much larger impact at small x.

To further study the constraining potential in the multi-observable framework, we have performed reweighting analyses using combined sets of DY, dijet and photon–jet pseudodata. Figure 4.16 shows the total expected impact on gluon modifications in gold in the central-barrel only (CB) and including forward instrumentation (CB+FI) scenarios. The constraints found in the CB scenario are rather modest, especially when acknowledging the fact that the small-x constraints in this case are mostly due to momentum-conservation induced correlations. The inclusion of forward instrumentation significantly increases the constraining power, particularly in the small-x region. We have found that these additional small-x constraints are driven at the present setup by the forward DY data. To leading order, the DY process happens through quark–antiquark

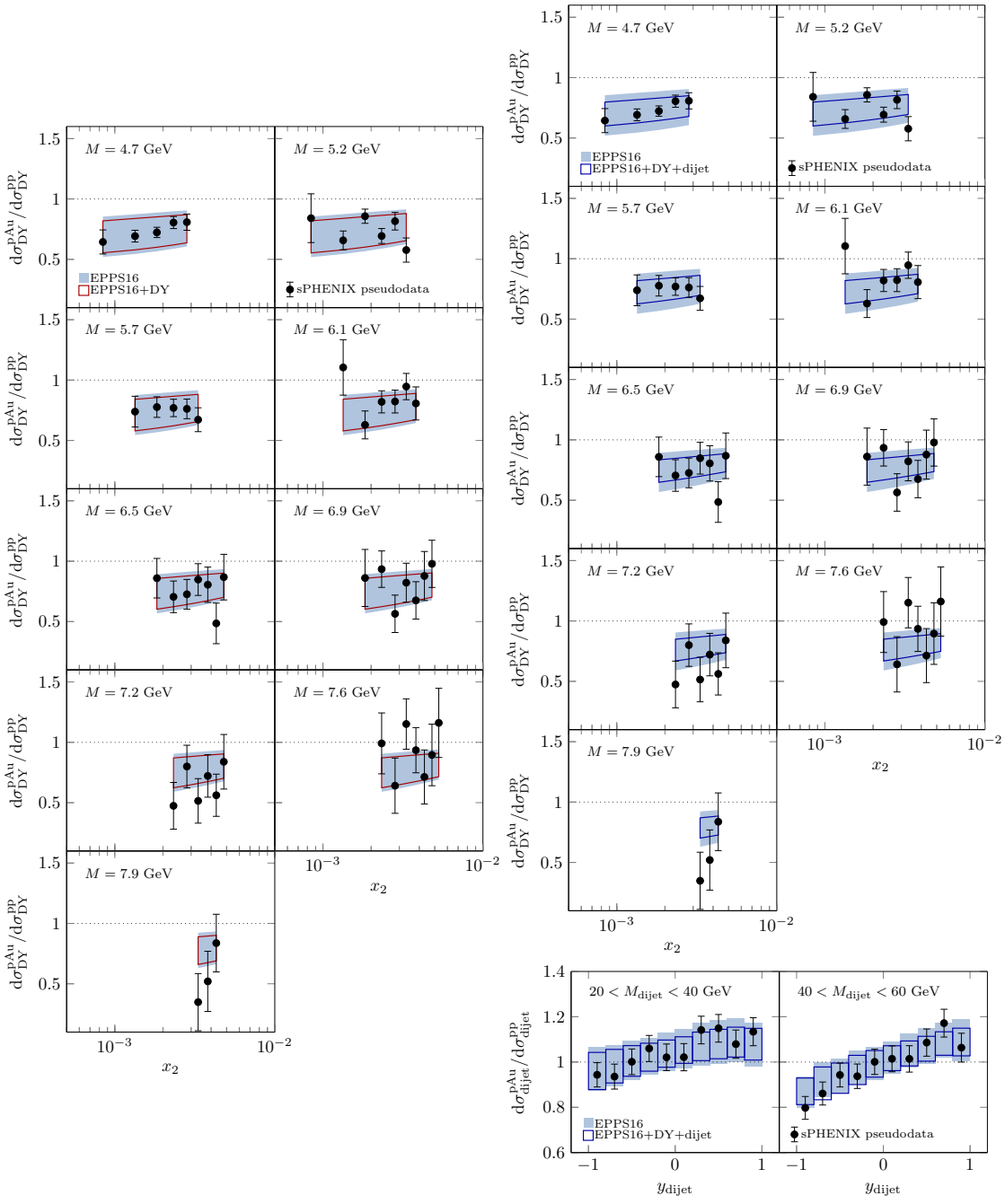

Figure 4.15. The impact of sPHENIX pseudodata on EPPS16 predictions in reweighting with the forward Drell–Yan only (left) or including also the central dijets (right). Figures from article [**IV**].

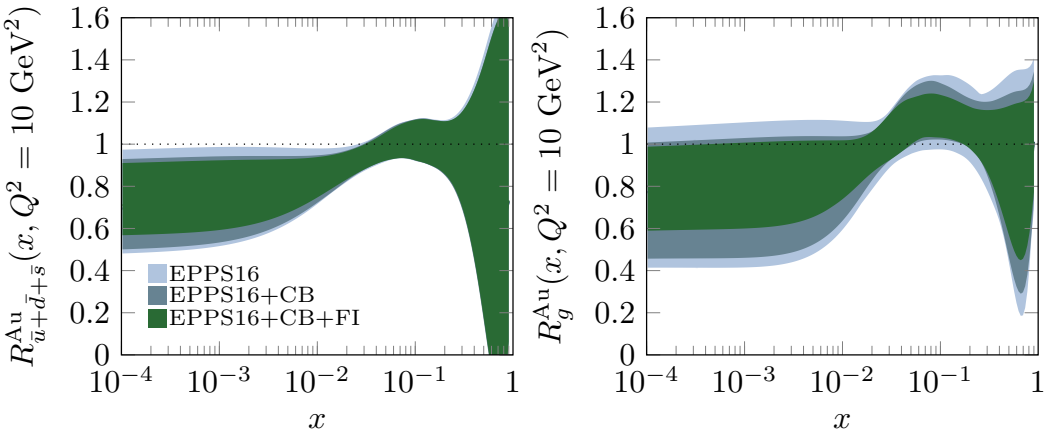

Figure 4.16. Projected impact of sPHENIX measurements of DY, dijet and gamma–jet nuclear modification ratios on EPPS16 in the central-barrel only (CB) and including forward instrumentation (CB+FI) scenarios. Figure from article [**IV**].

annihilation, but since at small x the level of sea quark distributions is set by the evolution from gluons and also the direct NLO contribution from quark–gluon scattering becomes increasingly important, the main impact at the scale shown in Figure 4.16 is on the gluon modifications.

While the constraints appear to be smaller than what was observed with dijet and D^0 production data from the LHC, the proton–gold measurements at RHIC are important in checking that the results obtained for lead are still valid at slightly smaller nucleus and to guide our assumptions on how the nuclear effects will evolve towards smaller nuclei. In fact, while the LHC will keep on providing constraints mostly for the lead nucleus, the flexibility of the RHIC beam line would allow performing a proper A-scan to put constraints also on the mass-number dependence of the gluon modifications.

Chapter 5

Conclusions

In this thesis, we have discussed the extraction of nuclear parton distribution functions (nPDFs), particularly in the light of new constraints derivable from various hadron–nucleus collision data which have not been previously included in nPDF global analyses. As a highlight, the article [**II**] of this thesis, with the EPPS16 nPDF set as output, presents the first nPDF global analysis including LHC data on electroweak gauge boson and dijet production. Summarizing further the main results of this thesis:

We have shown in the article [**I**] that in certain ratios of pion–nucleus Drell–Yan cross sections the pion PDFs and thus also the uncertainties they come with efficiently cancel at the level of NLO cross sections. Due to contributions involving valence antiquarks of the pions, these observables show an enhanced sensitivity to the flavour separation of quark nuclear modifications, which, as discussed in Section 4.2, is hard to constrain. The existing data have somewhat large uncertainties and thus do not yield very strong constraints, but nonetheless indicate that valence quark modifications should not be too asymmetrical. As recognized also by the experimental community, there are interesting prospects in performing such measurements in future experiments.

The tensionless fit found in the article [**II**] gives evidence for the universality of nPDFs across a wide variety of different processes in the kinematical range $Q \geq 1.3$ GeV studied. In this analysis, we have found the most decisive new data to be those from the CHORUS neutrino–nucleus DIS and CMS proton–nucleus measurements, putting new constraints on the flavour separation and gluon nuclear modifications, respectively. This analysis is also the first one to allow for a full flavour separation in the quark nuclear modifications, thus significantly reducing the bias in predictions sensitive to such differences.

Using Hessian PDF reweighting tools, we have quantified in articles [**III**] and [**V**] the potential constraints on nPDFs from CMS dijet and LHCb D^0-production data. The impact on nPDFs is found to be dramatic, with the dijets putting stringent constraints on gluon modifications especially in mid-x region and D^0

mesons respectively for gluons at small x. When used together in a global fit, these data can be expected to constrain the gluon modifications in lead to an unprecedented accuracy.

The field of nPDF analyses is currently evolving quickly, driven mainly by the constantly increasing amount of data constraints from the LHC proton–lead collisions. In addition to the data discussed above, newly-finalized measurements of ATLAS dijet conditional yields at 5.02 TeV [111] and CMS W bosons at 8.16 TeV [112] can shed additional light on the nuclear modifications in lead. In the coming years, with the LHC turning into a high-luminosity mode, the precision of especially electroweak observables is expected to improve significantly [113]. On a further note, additional observables, such as direct photons [114] or photoproduction of dijets [115], can also place new constraints on the gluons and test the universality of nPDFs.

All this makes the constraints for lighter nuclei to lag behind. While a fixed-target programme at the LHC [116] can help the situation significantly, the data from such measurements are bound to give constraints only in the large-x region. The RHIC collider, with a history of successful studies in a wide class of different collision systems and forward upgrades in its experiments coming up, thus offers a unique opportunity to constrain the mass-number dependence of the nuclear modifications. Article [**IV**] discusses the prospects of constraining the gluon nuclear modifications with RHIC. There, we found that a simultaneous analysis on multiple observables can help in reducing luminosity-related normalization uncertainties and thus improve the impact of forward DY measurements. Ultimately still, an electron–ion collider would be needed to truly pin down the PDF nuclear modifications [95, 117–119].

References

[1] John C. Collins, Davison E. Soper and George F. Sterman, Factorization of Hard Processes in QCD, *Adv. Ser. Direct. High Energy Phys.* 5 (1989) 1–91, arXiv: hep-ph/0409313.

[2] Jun Gao, Lucian Harland-Lang and Juan Rojo, The Structure of the Proton in the LHC Precision Era, *Phys. Rept.* 742 (2018) 1–121, arXiv: 1709.04922 [hep-ph].

[3] L. L. Frankfurt, M. I. Strikman and S. Liuti, Evidence for enhancement of gluon and valence quark distributions in nuclei from hard lepton nucleus processes, *Phys. Rev. Lett.* 65 (1990) 1725–1728.

[4] K. J. Eskola, Scale dependence of nuclear gluon structure, *Nucl. Phys.* B400 (1993) 240–266.

[5] K. J. Eskola, V. J. Kolhinen and P. V. Ruuskanen, Scale evolution of nuclear parton distributions, *Nucl. Phys.* B535 (1998) 351–371, arXiv: hep-ph/9802350.

[6] K. J. Eskola, V. J. Kolhinen and C. A. Salgado, The Scale dependent nuclear effects in parton distributions for practical applications, *Eur. Phys. J.* C9 (1999) 61–68, arXiv: hep-ph/9807297.

[7] T. Muta, *Foundations of quantum chromodynamics: An Introduction to perturbative methods in gauge theories*, World Scientific, 1987.

[8] R. Keith Ellis, W. James Stirling and B. R. Webber, *QCD and collider physics*, Cambridge University Press, 1996.

[9] David J. Gross and Frank Wilczek, Ultraviolet Behavior of Nonabelian Gauge Theories, *Phys. Rev. Lett.* 30 (1973) 1343–1346.

[10] H. David Politzer, Asymptotic Freedom: An Approach to Strong Interactions, *Phys. Rept.* 14 (1974) 129.

[11] John Collins, *Foundations of perturbative QCD*, Cambridge University Press, 2011.

[12] Richard P. Feynman, Very high-energy collisions of hadrons, *Phys. Rev. Lett.* 23 (1969) 1415.

[13] J. D. Bjorken and Emmanuel A. Paschos, Inelastic Electron Proton and γ-Proton Scattering, and the Structure of the Nucleon, *Phys. Rev.* 185 (1969) 1975.

[14] Karol Kovarik, Pavel M. Nadolsky and Davison E. Soper, *Hadron structure in high-energy collisions*, arXiv: 1905.06957 [hep-ph].

[15] Yuri L. Dokshitzer, Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics., *Sov. Phys. JETP* 46 (1977), [Zh. Eksp. Teor. Fiz. 73 (1977) 1216] 641.

[16] V. N. Gribov and L. N. Lipatov, Deep inelastic e p scattering in perturbation theory, *Sov. J. Nucl. Phys.* 15 (1972), [Yad. Fiz. 15 (1972) 781] 438.

[17] L. N. Lipatov, The parton model and perturbation theory, *Sov. J. Nucl. Phys.* 20 (1975), [Yad. Fiz. 20 (1974) 181] 94.

[18] Guido Altarelli and G. Parisi, Asymptotic Freedom in Parton Language, *Nucl. Phys.* B126 (1977) 298.

[19] Yuri L. Dokshitzer, Dmitri Diakonov and S. I. Troian, Hard Processes in Quantum Chromodynamics, *Phys. Rept.* 58 (1980) 269.

[20] Guido Altarelli, Partons in Quantum Chromodynamics, *Phys. Rept.* 81 (1982) 1.

[21] Yuri L. Dokshitzer, Valery A. Khoze, Alfred H. Mueller and S. I. Troian, *Basics of perturbative QCD*, Editions Frontieres, 1991.

[22] Hannu Paukkunen, Global analysis of nuclear parton distribution functions at leading and next-to-leading order perturbative QCD, PhD thesis, University of Jyväskylä, 2009, arXiv: 0906.2529 [hep-ph].

[23] Petja Paakkinen, Dokshitzer–Gribov–Lipatov–Altarelli–Parisi evolution equations, Master's thesis, University of Jyväskylä, 2015.

[24] A. Bassetto, M. Dalbosco, I. Lazzizzera and R. Soldati, Yang-Mills Theories in the Light Cone Gauge, *Phys. Rev.* D31 (1985) 2012.

[25] George Leibbrandt, Introduction to Noncovariant Gauges, *Rev. Mod. Phys.* 59 (1987) 1067.

[26] V. V. Sudakov, Vertex parts at very high-energies in quantum electrodynamics, *Sov. Phys. JETP* 3 (1956) 65.

[27] Guido Altarelli, R. Keith Ellis and G. Martinelli, Large Perturbative Corrections to the Drell-Yan Process in QCD, *Nucl. Phys.* B157 (1979) 461.

[28] W. Furmanski and R. Petronzio, Singlet Parton Densities Beyond Leading Order, *Phys. Lett.* B97 (1980) 437.

[29] G. Curci, W. Furmanski and R. Petronzio, Evolution of Parton Densities Beyond Leading Order: The Nonsinglet Case, *Nucl. Phys.* B175 (1980) 27.

[30] Raymond Brock et al., Handbook of perturbative QCD: Version 1.0, *Rev. Mod. Phys.* 67 (1995) 157–248.

[31] Gerard 't Hooft and M. J. G. Veltman, Regularization and Renormalization of Gauge Fields, *Nucl. Phys.* B44 (1972) 189–213.

[32] William A. Bardeen, A. J. Buras, D. W. Duke and T. Muta, Deep Inelastic Scattering Beyond the Leading Order in Asymptotically Free Gauge Theories, *Phys. Rev.* D18 (1978) 3998.

[33] W. Furmanski and R. Petronzio, Lepton–Hadron Processes Beyond Leading Order in Quantum Chromodynamics, *Z. Phys.* C11 (1982) 293.

[34] R. S. Thorne and W. K. Tung, PQCD Formulations with Heavy Quark Masses and Global Analysis, *Proceedings, HERA and the LHC Workshop Series on the implications of HERA for LHC physics: 2006-2008* 332–351, arXiv: 0809.0714 [hep-ph].

[35] John C. Collins, Hard scattering factorization with heavy quarks: A General treatment, *Phys. Rev.* D58 (1998) 094002, arXiv: hep-ph/9806259 [hep-ph].

[36] Michael Krämer, Fredrick I. Olness and Davison E. Soper, Treatment of heavy quarks in deeply inelastic scattering, *Phys. Rev.* D62 (2000) 096007, arXiv: hep-ph/0003035 [hep-ph].

[37] M. Gluck and E. Reya, Deep Inelastic Quantum Chromodynamic Charm Leptoproduction, *Phys. Lett.* 83B (1979) 98–102.

[38] Wu-Ki Tung, Stefan Kretzer and Carl Schmidt, Open heavy flavor production in QCD: Conceptual framework and implementation issues, *J. Phys.* G28 (2002) 983–996, arXiv: hep-ph/0110247 [hep-ph].

[39] John M. Campbell, R. Keith Ellis and Walter T. Giele, A Multi-Threaded Version of MCFM, *Eur. Phys. J.* C75, no.6 (2015) 246, arXiv: 1503.06182 [physics.comp-ph].

[40] Zoltan Nagy, Next-to-leading order calculation of three jet observables in hadron hadron collision, *Phys. Rev.* D68 (2003) 094002, arXiv: hep-ph/0307268 [hep-ph].

[41] Jun Gao et al., MEKS: a program for computation of inclusive jet cross sections at hadron colliders, *Comput. Phys. Commun.* 184 (2013) 1626–1642, arXiv: 1207.0513 [hep-ph].

[42] F. Aversa, P. Chiappetta, Mario Greco and J. P. Guillet, QCD Corrections to Parton-Parton Scattering Processes, *Nucl. Phys.* B327 (1989) 105.

[43] B. A. Kniehl, G. Kramer, I. Schienbein and H. Spiesberger, Collinear subtractions in hadroproduction of heavy quarks, *Eur. Phys. J.* C41 (2005) 199–212, arXiv: hep-ph/0502194 [hep-ph].

[44] Ilkka Helenius and Hannu Paukkunen, Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme, *JHEP* 05 (2018) 196, arXiv: 1804.03557 [hep-ph].

[45] Michelangelo L. Mangano, Paolo Nason and Giovanni Ridolfi, Heavy quark correlations in hadron collisions at next-to-leading order, *Nucl. Phys.* B373 (1992) 295–345.

[46] Stefano Forte, Lluis Garrido, Jose I. Latorre and Andrea Piccione, Neural network parametrization of deep inelastic structure functions, *JHEP* 05 (2002) 062, arXiv: hep-ph/0204232 [hep-ph].

[47] Tancredi Carli et al., A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project, *Eur. Phys. J.* C66 (2010) 503–524, arXiv: 0911.2985 [hep-ph].

[48] M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz and F. Stober, Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes (2011), arXiv: 1109.1310 [hep-ph].

[49] Valerio Bertone, Rikkert Frederix, Stefano Frixione, Juan Rojo and Mark Sutton, aMCfast: automation of fast NLO computations for PDF fits, *JHEP* 08 (2014) 166, arXiv: 1406.7693 [hep-ph].

[50] Petja Paakkinen, Nuclear parton distribution functions, *Proceedings, Old and New Strong Interactions from LHC to Future Colliders (LFC17): Trento, Italy, September 11-15, 2017* 33–40, arXiv: 1802.05927 [hep-ph].

[51] Matteo Cacciari and Nicolas Houdeau, Meaningful characterisation of perturbative theoretical uncertainties, *JHEP* 09 (2011) 039, arXiv: 1105.5152 [hep-ph].

[52] L. A. Harland-Lang and R. S. Thorne, On the Consistent Use of Scale Variations in PDF Fits and Predictions, *Eur. Phys. J.* C79, no.3 (2019) 225, arXiv: 1811.08434 [hep-ph].

[53] Rabah Abdul Khalek et al., *A First Determination of Parton Distributions with Theoretical Uncertainties*, arXiv: 1905.04311 [hep-ph].

[54] J. Pumplin et al., Uncertainties of predictions from parton distribution functions. 2. The Hessian method, *Phys. Rev.* D65 (2001) 014013, arXiv: hep-ph/0101032.

[55] M. Tanabashi et al., Review of Particle Physics, *Phys. Rev.* D98, no.3 (2018) 030001.

[56] D. Stump et al., Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method, *Phys. Rev.* D65 (2001) 014012, arXiv: hep-ph/0101051.

[57] G. D'Agostini, On the use of the covariance matrix to fit correlated data, *Nucl. Instrum. Meth.* A346 (1994) 306–311.

[58] Richard D. Ball et al., Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties, *JHEP* 05 (2010) 075, arXiv: 0912.2276 [hep-ph].

[59] Robert D. Cousins, Why isn't every physicist a Bayesian?, *Am. J. Phys.* 63 (1995) 398.

[60] Jon Pumplin, Parametrization dependence and $\Delta\chi^2$ in parton distribution fitting, *Phys. Rev.* D82 (2010) 114020, arXiv: 0909.5176 [hep-ph].

[61] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, Parton distributions for the LHC, *Eur. Phys. J.* C63 (2009) 189–285, arXiv: 0901.0002 [hep-ph].

[62] Hung-Liang Lai et al., New parton distributions for collider physics, *Phys. Rev.* D82 (2010) 074024, arXiv: 1007.2241 [hep-ph].

[63] Hannu Paukkunen and Carlos A. Salgado, Agreement of Neutrino Deep Inelastic Scattering Data with Global Fits of Parton Distributions, *Phys. Rev. Lett.* 110, no.21 (2013) 212301, arXiv: 1302.2001 [hep-ph].

[64] Hannu Paukkunen and Pia Zurita, PDF reweighting in the Hessian matrix approach, *JHEP* 12 (2014) 100, arXiv: 1402.6623 [hep-ph].

[65] Carl Schmidt, Jon Pumplin, C. P. Yuan and P. Yuan, Updating and optimizing error parton distribution function sets in the Hessian approach, *Phys. Rev.* D98, no.9 (2018) 094005, arXiv: 1806.07950 [hep-ph].

[66] J. J. Aubert et al., The ratio of the nucleon structure functions F_2^N for iron and deuterium, *Phys. Lett.* 123B (1983) 275–278.

[67] Michele Arneodo, Nuclear effects in structure functions, *Phys. Rept.* 240 (1994) 301–393.

[68] Donald F. Geesaman, K. Saito and Anthony William Thomas, The nuclear EMC effect, *Ann. Rev. Nucl. Part. Sci.* 45 (1995) 337–390.

[69] Gunther Piller and Wolfram Weise, Nuclear deep inelastic lepton scattering and coherence phenomena, *Phys. Rept.* 330 (2000) 1–94, arXiv: hep-ph/9908230.

[70] P. R. Norton, The EMC effect, *Rept. Prog. Phys.* 66 (2003) 1253–1297.

[71] Nestor Armesto, Nuclear shadowing, *J. Phys.* G32 (2006) R367–R394, arXiv: hep-ph/0604108 [hep-ph].

[72] Michelangelo L Mangano, Helmut Satz and Urs Achim Wiedemann, editors, *Hard Probes in Heavy Ion Collisions at the LHC*, CERN Yellow Reports: Monographs, CERN, 2004, ISBN: 9290832347.

[73] Valerio Bertone, Rhorry Gauld and Juan Rojo, Neutrino Telescopes as QCD Microscopes, *JHEP* 01 (2019) 217, arXiv: 1808.02034 [hep-ph].

[74] Richard D. Ball, Emanuele R. Nocera and Rosalyn L. Pearson, Nuclear Uncertainties in the Determination of Proton PDFs, *Eur. Phys. J.* C79, no.3 (2019) 282, arXiv: 1812.09074 [hep-ph].

[75] K. J. Eskola, H. Paukkunen and C. A. Salgado, EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions, *JHEP* 04 (2009) 065, arXiv: 0902.4154 [hep-ph].

[76] Daniel de Florian, Rodolfo Sassot, Pia Zurita and Marco Stratmann, Global Analysis of Nuclear Parton Distributions, *Phys. Rev.* D85 (2012) 074028, arXiv: 1112.6324 [hep-ph].

[77] Hamzeh Khanpour and S. Atashbar Tehrani, Global Analysis of Nuclear Parton Distribution Functions and Their Uncertainties at Next-to-Next-to-Leading Order, *Phys. Rev.* D93, no.1 (2016) 014026, arXiv: 1601.00939 [hep-ph].

[78] K. Kovarik et al., nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework, *Phys. Rev.* D93, no.8 (2016) 085037, arXiv: 1509.00792 [hep-ph].

[79] Rabah Abdul Khalek, Jacob J. Ethier and Juan Rojo, Nuclear Parton Distributions from Lepton-Nucleus Scattering and the Impact of an Electron-Ion Collider, *Eur. Phys. J.* C79, no.6 (2019) 471, arXiv: 1904.00018 [hep-ph].

[80] D. de Florian and R. Sassot, Nuclear parton distributions at next-to-leading order, *Phys. Rev.* D69 (2004) 074028, arXiv: hep-ph/0311227.

[81] Hannu Paukkunen, Status of nuclear PDFs after the first LHC p–Pb run, *Nucl. Phys.* A967 (2017) 241–248, arXiv: 1704.04036 [hep-ph].

[82] D. Dutta, J. C. Peng, I. C. Cloet and D. Gaskell, Pion-induced Drell-Yan processes and the flavor-dependent EMC effect, *Phys. Rev.* C83 (2011) 042201, arXiv: 1007.3916 [nucl-ex].

[83] M. Gluck, E. Reya and A. Vogt, Pionic parton distributions, *Z. Phys.* C53 (1992) 651–656.

[84] P. J. Sutton, Alan D. Martin, R. G. Roberts and W. James Stirling, Parton distributions for the pion extracted from Drell-Yan and prompt photon experiments, *Phys. Rev.* D45 (1992) 2349–2359.

[85] J. Badier et al., Test of Nuclear Effects in Hadronic Dimuon Production, *Phys. Lett.* 104B (1981), [,807(1981)] 335.

[86] P. Bordalo et al., Nuclear Effects on the Nucleon Structure Functions in Hadronic High Mass Dimuon Production, *Phys. Lett.* B193 (1987) 368.

[87] J. G. Heinrich et al., Measurement of the Ratio of Sea to Valence Quarks in the Nucleon, *Phys. Rev. Lett.* 63 (1989) 356–359.

[88] Vardan Khachatryan et al., Study of W boson production in pPb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, *Phys. Lett.* B750 (2015) 565–586, arXiv: 1503.05825 [nucl-ex].

[89] Vardan Khachatryan et al., Study of Z boson production in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Phys. Lett.* B759 (2016) 36–57, arXiv: 1512.06461 [hep-ex].

[90] Georges Aad et al., Z boson production in p+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV measured with the ATLAS detector, *Phys. Rev.* C92, no.4 (2015) 044915, arXiv: 1507.06232 [hep-ex].

[91] G. Onengut et al., Measurement of nucleon structure functions in neutrino scattering, *Phys. Lett.* B632 (2006) 65–75.

[92] Walter T. Giele and Stephane Keller, Implications of hadron collider observables on parton distribution function uncertainties, *Phys. Rev.* D58 (1998) 094023, arXiv: hep-ph/9803393 [hep-ph].

[93] Walter T. Giele, Stephane A. Keller and David A. Kosower, *Parton Distribution Function Uncertainties*, arXiv: hep-ph/0104052 [hep-ph].

[94] M. Arneodo et al., The Structure Function ratios F2(li) / F2(D) and F2(C) / F2(D) at small x, *Nucl. Phys.* B441 (1995) 12–30, arXiv: hep-ex/9504002 [hep-ex].

[95] E. C. Aschenauer, S. Fazio, M. A. C. Lamont, H. Paukkunen and Pia Zurita, Nuclear Structure Functions at a Future Electron-Ion Collider, *Phys. Rev.* D96, no.11 (2017) 114005, arXiv: 1708.05654 [nucl-ex].

[96] Petja Paakkinen, Kari J. Eskola and Hannu Paukkunen, Pion–nucleus Drell–Yan data as a novel constraint for nuclear PDFs, *PoS* DIS2017 (2018) 205, arXiv: 1710.05736 [hep-ph].

[97] B. Adams et al., *Letter of Intent: A New QCD facility at the M2 beam line of the CERN SPS (COMPASS++/AMBER)*, arXiv: 1808.00848 [hep-ex].

[98] S. S. Adler et al., Centrality dependence of pi0 and eta production at large transverse momentum in s(NN)**(1/2) = 200-GeV d+Au collisions, *Phys. Rev. Lett.* 98 (2007) 172302, arXiv: nucl-ex/0610036.

[99] B. I. Abelev et al., Inclusive π^0, η, and direct photon production at high transverse momentum in $p + p$ and d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV, *Phys. Rev.* C81 (2010) 064904, arXiv: 0912.3838 [hep-ex].

[100] Kari J. Eskola, Hannu Paukkunen and Carlos A. Salgado, An Improved global analysis of nuclear parton distribution functions including RHIC data, *JHEP* 07 (2008) 102, arXiv: 0802.0139 [hep-ph].

[101] Serguei Chatrchyan et al., Studies of dijet transverse momentum balance and pseudorapidity distributions in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, *Eur. Phys. J.* C74, no.7 (2014) 2951, arXiv: 1401.4433 [nucl-ex].

[102] Bernd A. Kniehl, G. Kramer and B. Potter, Fragmentation functions for pions, kaons, and protons at next-to-leading order, *Nucl. Phys.* B582 (2000) 514–536, arXiv: hep-ph/0010289 [hep-ph].

[103] Albert M Sirunyan et al., Constraining gluon distributions in nuclei using dijets in proton-proton and proton-lead collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, *Phys. Rev. Lett.* 121, no.6 (2018) 062002, arXiv: 1805.04736 [hep-ex].

[104] Rhorry Gauld, Forward D predictions for pPb collisions, and sensitivity to cold nuclear matter effects, *Phys. Rev.* D93, no.1 (2016) 014001, arXiv: 1508.07629 [hep-ph].

[105] Roel Aaij et al., Study of prompt D^0 meson production in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5$ TeV, *JHEP* 10 (2017) 090, arXiv: 1707.02750 [hep-ex].

[106] G. Kramer and H. Spiesberger, Study of heavy meson production in p–Pb collisions at $\sqrt{S}$= 5.02 TeV in the general-mass variable-flavour-number scheme, *Nucl. Phys.* B925 (2017) 415–430, arXiv: 1703.04754 [hep-ph].

[107] Aleksander Kusina, Jean-Philippe Lansberg, Ingo Schienbein and Hua-Sheng Shao, Gluon Shadowing in Heavy-Flavor Production at the LHC, *Phys. Rev. Lett.* 121, no.5 (2018) 052004, arXiv: 1712.07024 [hep-ph].

[108] T. Kneesch, B. A. Kniehl, G. Kramer and I. Schienbein, Charmed-meson fragmentation functions with finite-mass corrections, *Nucl. Phys.* B799 (2008) 34–59, arXiv: 0712.0481 [hep-ph].

[109] Elke-Caroline Aschenauer et al., *The RHIC Cold QCD Plan for 2017 to 2023: A Portal to the EIC*, arXiv: 1602.03922 [nucl-ex].

[110] A. Adare et al., *An Upgrade Proposal from the PHENIX Collaboration*, arXiv: 1501.06197 [nucl-ex].

[111] Morad Aaboud et al., *Dijet azimuthal correlations and conditional yields in pp and p+Pb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 5.02 TeV with the ATLAS detector*, arXiv: 1901.10440 [nucl-ex].

[112] Albert M Sirunyan et al., *Observation of nuclear modifications in $\mathrm{W}^{\pm}$ boson production in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV*, arXiv: 1905.01486 [hep-ex].

[113] Z. Citron et al., Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams, *HL/HE-LHC Workshop: Workshop on the Physics of HL-LHC, and Perspectives at HE-LHC Geneva, Switzerland, June 18-20, 2018*, arXiv: 1812.06772 [hep-ph].

[114] Ilkka Helenius, Kari J. Eskola and Hannu Paukkunen, Probing the small-x nuclear gluon distributions with isolated photons at forward rapidities in p+Pb collisions at the LHC, *JHEP* 09 (2014) 138, arXiv: 1406.1689 [hep-ph].

[115] V. Guzey and M. Klasen, Constraints on nuclear parton distributions from dijet photoproduction at the LHC, *Eur. Phys. J.* C79, no.5 (2019) 396, arXiv: 1902.05126 [hep-ph].

[116] C. Hadjidakis et al., *A Fixed-Target Programme at the LHC: Physics Case and Projected Performances for Heavy-Ion, Hadron, Spin and Astroparticle Studies*, arXiv: 1807.00603 [hep-ex].

[117] A. Accardi et al., Electron Ion Collider: The Next QCD Frontier, *Eur. Phys. J.* A52, no.9 (2016), edited by A. Deshpande, Z. E. Meziani and J. W. Qiu 268, arXiv: 1212.1701 [nucl-ex].

[118] J. L. Abelleira Fernandez et al., A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector, *J. Phys.* G39 (2012) 075001, arXiv: 1206.2913 [physics.acc-ph].

[119] Hannu Paukkunen, An update on nuclear PDFs at the LHeC, *PoS* DIS2017 (2018) 109, arXiv: 1709.08342 [hep-ph].

ORIGINAL PAPERS

I

APPLICABILITY OF PION–NUCLEUS DRELL–YAN DATA IN GLOBAL ANALYSIS OF NUCLEAR PARTON DISTRIBUTION FUNCTIONS

by

Paakkinen, P., Eskola, K. J. & Paukkunen, H., 2017

Phys. Lett. B768 (2017) 7–11

arXiv: 1609.07262 [nucl-th]

Physics Letters B 768 (2017) 7–11

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Applicability of pion–nucleus Drell–Yan data in global analysis of nuclear parton distribution functions

Petja Paakkinen [a,*], Kari J. Eskola [a,b], Hannu Paukkunen [a,b,c]

[a] University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland
[b] Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland
[c] Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, E-15782 Galicia, Spain

ARTICLE INFO

Article history:
Received 3 October 2016
Accepted 7 February 2017
Available online 11 February 2017
Editor: J.-P. Blaizot

Keywords:
Drell–Yan process
Pion–nucleus scattering
Nuclear parton distribution functions

ABSTRACT

Despite the success of modern nuclear parton distribution functions (nPDFs) in describing nuclear hard-process data, they still suffer from large uncertainties. One of the poorly constrained features is the possible asymmetry in nuclear modifications of valence u and d quarks. We study the possibility of using pion–nucleus Drell–Yan dilepton data as a new constraint in the global analysis of nPDFs. We find that the nuclear cross-section ratios from the NA3, NA10 and E615 experiments can be used without imposing significant new theoretical uncertainties and, in particular, that these datasets may have some constraining power on the u/d-asymmetry in nuclei.

1. Introduction

Since the discovery of the EMC effect in 1983 [1] the nuclear effects in bound-hadron partonic structure have been under active study [2,3]. For collinearly factorizable hard processes this phenomenon can be described by nuclear modifications of parton distribution functions (PDFs), the latest global extractions being EPS09 [4], DSSZ [5] and nCTEQ15 [6], see Refs. [7,8] for reviews. Despite the success of nPDFs in describing also nuclear hard-process data from the LHC [9], they still suffer from large uncertainties. One of the shortcomings is the lack of data which would constrain the nuclear effects of all parton flavors simultaneously without any a priori assumptions. For example, it has been customary to assume that nuclear modifications for both valence quarks u and d are the same. While this assumption has been consistent e.g. with the available LHC data [9] and neutrino–nucleus deep inelastic scattering [10], the two are not expected to be exactly the same [11]. It is only recently that an attempt to fit these separately has been carried out [6] but due to the lack of constraining data inconclusive results are obtained. Among other possibilities [12,13] it has been also suggested [14] that Drell–Yan dilepton data from pion–nucleus collision experiments could be used in nPDF global analyses to constrain the u/d-asymmetry. In this Letter, we provide a detailed study of this possibility in terms of the available data and next-to-leading order (NLO) cross-section computations with the EPS09 and nCTEQ15 nPDFs.

2. Dependence on pion PDFs

The NA3 [15], NA10 [16] and E615 [17] experiments provide pion–nucleus ($\pi^{\pm}+A$) Drell–Yan dilepton (l^-l^+) production data in the following per-nucleon cross-section ratios:

$$R_A^{+/-}(x_2) \equiv \frac{\mathrm{d}\sigma(\pi^+ + A \to l^-l^+ + X)/\mathrm{d}x_2}{\mathrm{d}\sigma(\pi^- + A \to l^-l^+ + X)/\mathrm{d}x_2}, \tag{1}$$

$$R^-_{A_1/A_2}(x_2) \equiv \frac{\frac{1}{A_1}\mathrm{d}\sigma(\pi^- + A_1 \to l^-l^+ + X)/\mathrm{d}x_2}{\frac{1}{A_2}\mathrm{d}\sigma(\pi^- + A_2 \to l^-l^+ + X)/\mathrm{d}x_2}. \tag{2}$$

Here, $x_2 \equiv \frac{M}{\sqrt{s}}\mathrm{e}^{-y}$, where M and y are the invariant mass and rapidity of the lepton pair. The pion–nucleon center-of-mass energy is denoted by $\sqrt{s}$. At leading order (LO), the Drell–Yan cross section reads

$$\frac{\mathrm{d}\sigma(\pi^{\pm}+A\to l^-l^+ + X)}{\mathrm{d}x_2} \overset{\mathrm{LO}}{=} \int_{\Delta M} \mathrm{d}M \frac{8\pi\alpha^2}{9sx_2M}\sum_q e_q^2[q_{\pi^{\pm}}(x_1)\bar{q}_A(x_2) + \bar{q}_{\pi^{\pm}}(x_1)q_A(x_2)], \tag{3}$$

* Corresponding author.
E-mail addresses: petja.paakkinen@jyu.fi (P. Paakkinen), kari.eskola@jyu.fi (K.J. Eskola), hannu.paukkunen@jyu.fi (H. Paukkunen).

http://dx.doi.org/10.1016/j.physletb.2017.02.009

where α is the fine-structure constant, $x_1 \equiv \frac{M}{\sqrt{s}}e^y = \frac{M^2}{sx_2}$, and the sum goes over the quark flavors q with e_q being the quark charge. The quark/antiquark distributions in a pion (nucleus) at factorization scale $Q \sim M$ are denoted by $q_{\pi^{\pm}(A)}/\bar{q}_{\pi^{\pm}(A)}$.

The range of the mass integral (ΔM) as well as $\sqrt{s}$ depend on the experiment and are 4.1 GeV $< M <$ 8.5 GeV and $\sqrt{s} =$ 16.8 GeV for NA3. The NA10 experiment provides data at two different beam energies, 286 GeV ($\sqrt{s} = 23.2$ GeV) and 140 GeV ($\sqrt{s} = 16.2$ GeV), with a mass range 4.2 GeV $< M <$ 15 GeV for the higher and 4.35 GeV $< M <$ 15 GeV for the lower energy, but in both cases excluding the Υ peak region 8.5 GeV $< M <$ 11 GeV.[1] In the E615 data the mass range is 4.05 GeV $< M <$ 8.55 GeV at $\sqrt{s} = 21.7$ GeV, but with an additional kinematical cut $x_1 > 0.36$, which was imposed by the experiment to reduce contributions from the pion sea quarks.

Assuming the isospin and charge conjugation symmetry we have $u_{\pi^+} = d_{\pi^-} = \bar{d}_{\pi^+} = \bar{u}_{\pi^-}$ and $d_{\pi^+} = u_{\pi^-} = \bar{u}_{\pi^+} = \bar{d}_{\pi^-}$. Hence, in the limit where the pion sea quarks can be neglected and assuming that the mass integration range is narrow enough so that the scale evolution of the PDFs does not play a role, the LO approximation gives

$$R_A^{+/-}(x_2) \approx \frac{4\bar{u}_A(x_2) + d_A(x_2)}{4u_A(x_2) + \bar{d}_A(x_2)}, \tag{4}$$

$$R^-_{A_1/A_2}(x_2) \approx \frac{4u_{A_1}(x_2) + \bar{d}_{A_1}(x_2)}{4u_{A_2}(x_2) + \bar{d}_{A_2}(x_2)}, \tag{5}$$

where u_A and d_A are the per-nucleon distributions of u and d quarks in a nucleus A with Z protons,

$$u_A \equiv \frac{Z}{A}u_{p/A} + \frac{A-Z}{A}d_{p/A}, \tag{6}$$

$$d_A \equiv \frac{Z}{A}d_{p/A} + \frac{A-Z}{A}u_{p/A}. \tag{7}$$

Here, $u_{p/A}$, $d_{p/A}$ are the parton distribution functions of a bound proton and we have again used the isospin symmetry to write $u_{n/A} = d_{p/A}$, $d_{n/A} = u_{p/A}$. As the dependence on the pion PDFs essentially cancels in $R^-_{A_1/A_2}$ and $R_A^{+/-}$, these quantities promise to be good candidates for global nPDF analyses, where the objective is to probe the nuclear modifications without being significantly sensitive to (possibly poorly known) pion structure. By comparing Equations (4) and (5) we see that while $R^-_{A_1/A_2}$ probes dominantly the valence quarks, $R_A^{+/-}$ carries more sensitivity to sea quarks as well.

The above approximative cancellation of the pion PDFs in cross-section ratios has to be tested explicitly in a NLO calculation to avoid including any biased constraints to nPDF analysis. In Fig. 1, we plot the NA3, NA10 and E615 data along with our NLO results using the GRV [18] and SMRS [19] pion PDFs together with EPS09 nuclear modifications and CT14 [20] free-proton PDFs.[2] For hydrogen and deuterium we use the unmodified CT14 PDFs. In the upper-left panel we have taken into account the kinematical cut $x_1 > 0.36$ and in the right-hand-side panels an isospin correction as described in the next section has been applied. The NLO calculations were done using MCFM 7.0.1 [21]. For the data points only statistical errors are available, but these are in any case expected to be dominant in comparison to the systematical errors (except the normalization error of the NA10 data discussed in the next section).

[1] Dutta et al. [14] used the NA10 data combined from the two different beam energies. We take these as separate datasets.

[2] The NA3 data is originally given as $R^-_{\mathrm{H/Pt}}$ which we have inverted as it is customary to take the ratio with respect to the lighter nucleus.

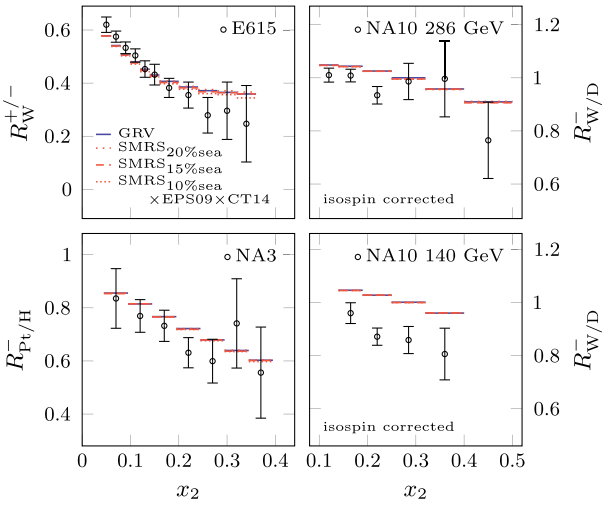

Fig. 1. Comparison of NLO predictions with the E615, NA10 and NA3 data. In all panels, we use the GRV (blue) and SMRS (red) PDFs for the pion, and the EPS09 nuclear modifications with the CT14 proton PDFs for the nuclei. In the upper-left panel we have taken into account the kinematical cut $x_1 > 0.36$ and in the right-hand-side panels an isospin correction as described in Section 3 has been applied to the theory predictions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The SMRS pion PDFs provide three different sets to account for the uncertainty in the fraction of pion momentum carried by the sea quarks. We find that the NLO predictions are largely insensitive to the choice of pion PDFs. Especially the SMRS 15% sea set which is to be considered as their central prediction is almost indistinguishable from the GRV results. A slight separation between the different SMRS sets is observed towards large x_2 in $R_{\mathrm{W}}^{+/-}$, but in comparison to the data uncertainties this is insignificant.

3. Isospin correction and normalization of NA10 datasets

The NA10 collaboration has corrected their data for the isospin effects. The exact form of correction was obtained from a LO Monte Carlo simulation but is not quoted point by point along with the data [16].[3] To mimic these corrections and compare with the data the best we can, we apply an isospin correction by computing the theory predictions as

$$\begin{aligned}(R^-_{\mathrm{W/D}})^{\mathrm{NLO}}_{\text{isospin corrected}} \\ = (R^-_{\text{isoscalar-W/W}})^{\mathrm{LO}}_{\text{no nPDFs}} \times (R^-_{\mathrm{W/D}})^{\mathrm{NLO}},\end{aligned} \tag{8}$$

where "isoscalar-W" is the isospin-symmetrized W nucleus ($Z = A/2$) and where the LO correction factor $(R^-_{\text{isoscalar-W/W}})^{\mathrm{LO}}_{\text{no nPDFs}}$ is evaluated with the central set of CT14 without nuclear modifications in PDFs. This correction has been applied on the right-hand-side panels of Fig. 1 and the effect can be seen in Fig. 2, where we plot both the corrected and uncorrected predictions using GRV pion PDFs. In Fig. 2, we also show the error bands from the CT14 proton PDFs (using the asymmetric prescription [22] to combine the uncertainties from the error sets) which are typically rather small in comparison to the data uncertainties except, perhaps, the E615 data at smallest values of x_2. To some extent, the isospin corrected NA10 data also contain input from the proton PDFs used by the experiment in their Monte Carlo code, but we do not study such a source of uncertainty here further.

[3] We thank P. Bordalo for discussion on this matter.

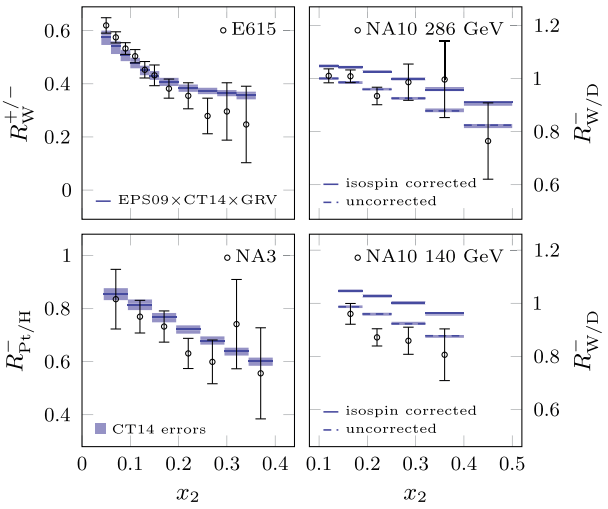

Fig. 2. As Fig. 1, but showing the error estimates from the CT14 PDFs as shaded blue bands for the results obtained with EPS09 and GRV pion PDFs. In the right-hand-side panels we show both the isospin corrected (solid) and uncorrected (dashed) NLO results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

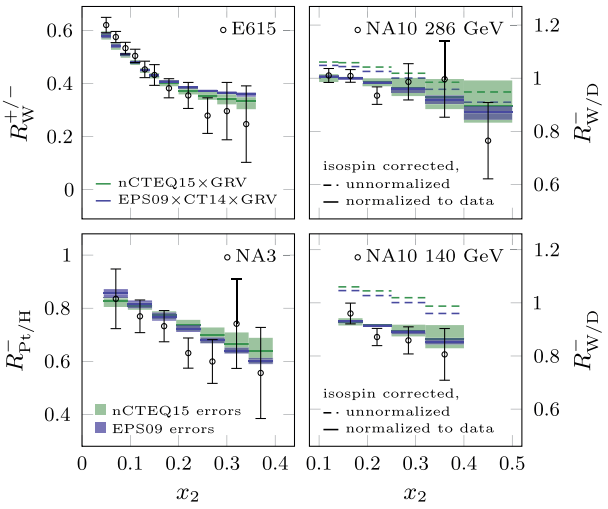

Fig. 3. A comparison of the uncertainty bands obtained using the EPS09 (blue lines and bands) and nCTEQ15 (green lines and bands) nuclear PDFs. In the right-hand-side panels we show both the unnormalized (dashed) and results normalized to the data (solid). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Normalization factors for the NA10 data sets.

nPDF	$\mathcal{N}^{\text{data}}$	
	286 GeV data	140 GeV data
EPS09	1.044	1.125
nCTEQ15	1.058	1.141

We observe that our isospin corrected theory prediction overshoots especially the low-energy NA10 data. This can be accounted for by the systematic overall normalization uncertainty of the data, quoted in [16] to be $\sigma_{\mathcal{N}^{\text{data}}} = 6\%$. To compare the predictions from different nPDFs with the NA10 data in shape and not in overall normalization, we normalize the results as follows: We fix the optimal normalization factor $\mathcal{N}^{\text{data}}$ for each data set and theory prediction separately by minimizing

$$\chi^2(\mathcal{N}^{\text{data}}) = \sum_i \frac{(\mathcal{N}^{\text{data}} R_i^{\text{data}} - R_i^{\text{theory}})^2}{(\sigma_i^{\text{data}})^2} + \frac{(\mathcal{N}^{\text{data}} - 1)^2}{(\sigma_{\mathcal{N}^{\text{data}}})^2} \tag{9}$$

with respect to data normalization $\mathcal{N}^{\text{data}}$ [23]. In the above equation R_i^{data} and R_i^{theory} are the experimental and theoretical values for ith bin in a data set, and σ_i^{data} is the data uncertainty (here statistical). We then obtain the theory predictions normalized to data as

$$(R_i^{\text{theory}})_{\text{normalized}} = \frac{R_i^{\text{theory}}}{\mathcal{N}^{\text{data}}}. \tag{10}$$

The values for $\mathcal{N}^{\text{data}}$ are given in Table 1 and the normalized results as well as the unnormalized ones are presented in Fig. 3 for the EPS09 and nCTEQ15 nuclear PDFs.[4] For predictions with nCTEQ15 PDFs we use their own free proton set for hydrogen and deuterium (and CT14 for EPS09). When calculating the nPDF errors, we have also normalized each error set separately. We observe that the optimal normalization for the NA10 286 GeV dataset is within the given 6% overall normalization uncertainty, but for the 140 GeV dataset it is more than twice the suggested uncertainty limit. Such a large normalization issue is not unheard of: For example, while the carbon-to-deuteron and lead-to-deuteron nuclear ratios in deep inelastic scattering measured by the E665 collaboration [24] are individually largely apart from other measurements, the lead-to-carbon ratio formed from these two agrees well with other experiments [25]. A similar normalization issue may be in question here as well.

[4] Since nCTEQ15 grids for platinum have not been available for us, we have used their grids for gold instead in $R^{-}_{\text{Pt/H}}$. Since the mass numbers are very close, $A_{\text{Pt}} = 195$ and $A_{\text{Au}} = 197$, this should be an excellent approximation.

4. Compatibility with nuclear PDFs

Comparing the results obtained with the EPS09 and nCTEQ15 nuclear PDFs in Fig. 3 we find that both these sets are in a fairly good agreement with the data, but display a large difference in their uncertainty estimates. To understand this, let us study the $R^{-}_{\text{W/D}}$ ratio measured by NA10. For large x_2, only the valence quarks in nuclei contribute and in the LO approximation we have

$$R^{-}_{\text{W/D}} \overset{x_2 \to 1}{\approx} R^{\text{W}}_{\text{V-isoscalar}} + R^{\text{W}}_{\text{V-nonisoscalar}}, \tag{11}$$

where

$$R^{A}_{\text{V-isoscalar}} \equiv \frac{u^{\text{V}}_{p/A} + d^{\text{V}}_{p/A}}{u^{\text{V}}_{p} + d^{\text{V}}_{p}} \tag{12}$$

is the nuclear modification factor for an average valence quark in an isoscalar nucleus and

$$R^{A}_{\text{V-nonisoscalar}} \equiv \left(\frac{2Z}{A} - 1\right) \frac{u^{\text{V}}_{p/A} - d^{\text{V}}_{p/A}}{u^{\text{V}}_{p} + d^{\text{V}}_{p}} \tag{13}$$

the corresponding non-isoscalarity correction. For neutron-rich nuclei this correction is negative and typically small in comparison to the isoscalar contribution.

In Fig. 4, we plot these two components for tungsten along with the nuclear modification factors

$$R^{\text{W}}_{u_{\text{V}}} \equiv \frac{u^{\text{V}}_{p/\text{W}}}{u^{\text{V}}_{p}}, \qquad R^{\text{W}}_{d_{\text{V}}} \equiv \frac{d^{\text{V}}_{p/\text{W}}}{d^{\text{V}}_{p}} \tag{14}$$

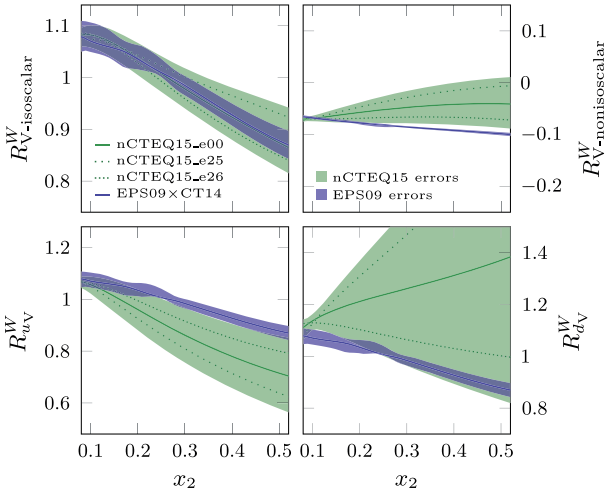

Fig. 4. The different LO valence-quark contributions to $R^{-}_{\mathrm{W/D}}$ (upper panels) and the valence quark nuclear modification factors (lower panels) at factorization scale $Q = 5$ GeV. Solid lines correspond to the EPS09 (blue) and nCTEQ15 (green) central sets and dotted lines indicate the error sets 25 and 26 of the nCTEQ15. The uncertainty bands are shown as green (nCTEQ15) and blue (EPS09) bands. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

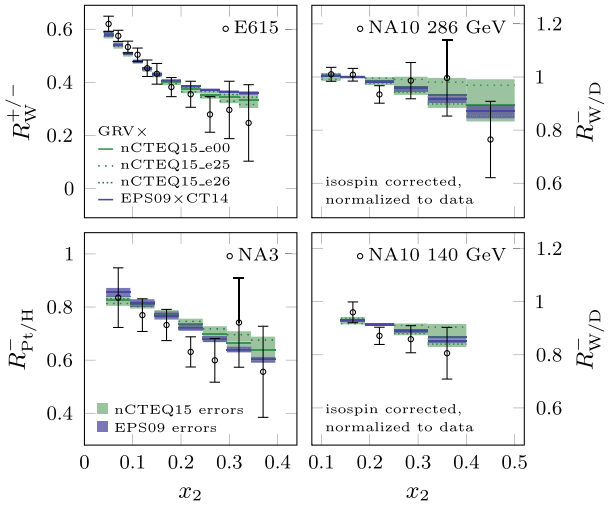

Fig. 5. As Fig. 3, but with only normalized results shown and the nCTEQ15 error sets 25 and 26 (dotted lines) plotted.

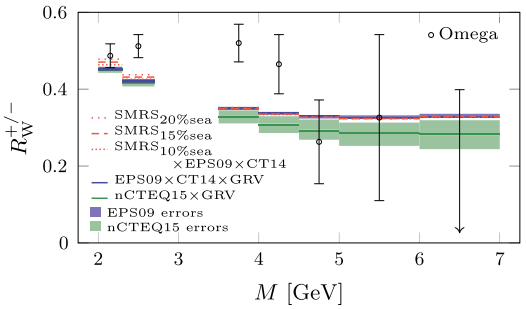

Fig. 6. Comparison of the Omega data with predictions using the GRV (blue) and SMRS (red) pion parton distributions together with the EPS09 nuclear modifications combined to the CT14 proton PDFs and also from using the nCTEQ15 (green) nuclear PDFs with the GRV pion PDFs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

at factorization scale $Q = 5$ GeV. We find that EPS09 and nCTEQ15 agree on $R^{\mathrm{W}}_{\text{V-isoscalar}}$, which is well constrained in both analyses, but there is a slight disagreement on $R^{\mathrm{W}}_{\text{V-nonisoscalar}}$. In addition, we see that nCTEQ15 has significantly larger error bands in both of these components. To study this difference in more detail, we plot in Fig. 4 also the nCTEQ15 error sets 25 and 26, which give the largest deviations from the central-set predictions. We can make two observations: First, from the lower panels in Fig. 4, we see that these two error sets are related to the nuclear modifications of u and d valence quarks with set 25 giving the most extreme difference, and set 26 being closer to uniform modifications. Second, from the upper panels in Fig. 4, we find that the deviations from the central prediction are in the same direction for both $R^{\mathrm{W}}_{\text{V-isoscalar}}$ and $R^{\mathrm{W}}_{\text{V-nonisoscalar}}$ (upwards for set 25, downwards for set 26), and combine additively in Equation (11) thereby explaining the larger error bands seen in Fig. 3.

It is now evident that the studied observables are sensitive to the mutual differences between u and d valence quark nuclear modifications. On one hand, the EPS09 error sets underestimate the true uncertainty because flavor dependence of valence quark nuclear modifications was not allowed in that particular analysis. On the other hand, the nCTEQ15 error bands are large since the flavor dependence was allowed, but not well constrained in their analysis. The size of nCTEQ15 error bands suggest that the pion–nucleus Drell–Yan data can have some constraining power on the difference of valence modifications. Indeed, in Fig. 5 we plot the predictions using the nCTEQ15 error sets 25 and 26, and observe that the most extreme deviation from identical nuclear modifications of u and d quarks given by set 25 is disfavored by NA3 and NA10 data.

In addition to the NA3, NA10 and E615 data we have studied also the results from the Omega experiment [26]. The data at $\sqrt{s} = 8.7$ GeV as a function of the lepton pair invariant mass are shown in Fig. 6 for $x_{\mathrm{F}} \equiv \frac{2p^*_{\mathrm{L}}}{\sqrt{s}} > 0$, where p^*_{L} is the longitudinal momentum of the lepton pair along the beam line in the center-of-mass frame. We find that the data disagree with theory predictions in bins around the J/ψ peak. Furthermore, at low invariant masses the choice of pion PDFs becomes significant and that especially towards larger invariant masses the data are not precise enough to discriminate between the nuclear PDFs. Hence it is not reasonable to include this dataset into a global nPDF analysis.

5. Conclusions

We have studied the prospects of including NA3, NA10, E615 and Omega pion–nucleus Drell–Yan data to global analyses of nuclear parton distribution functions. The NA3, NA10 and E615 data are compatible (modulo NA10 normalization at lower beam energies) with modern nPDFs and can thus be used in a global analysis without causing significant tension. The Omega data is not compatible with the NLO theory predictions and not precise enough to be useful in the nPDF analysis. The cross-section ratios used in the experiments are largely independent of pion parton distributions and hence the inclusion of these data will not impose significant new theoretical uncertainties to the analysis. Some sensitivity to baseline proton PDFs however still persists. When implementing these data to a global analysis, one needs to take into account the isospin correction and normalization uncertainty in the NA10 datasets. This can be done as described above. Motivated by this

study, these pion–nucleus Drell–Yan data have recently been included in the successor of the EPS09 analysis [27].

The considered nuclear ratios are sensitive to the possible u/d-asymmetry of nuclear modification factors but the data are not precise enough to pin down this difference completely. Regarding this matter we seem to reach a somewhat different conclusion than Dutta et al. [14] who claimed that NA3 data would favor flavor-dependent nuclear PDFs. We, in our analysis, find a very good agreement between the data and u/d-symmetric (EPS09) nuclear modifications. Moreover, our analysis suggests that the most extreme differences in u and d quark nuclear modifications as given by particular nCTEQ15 error sets are disfavored by the NA3 and NA10 datasets.

Acknowledgements

This research was supported by the Academy of Finland, Project 297058 of K.J.E., and by the European Research Council grant HotLHC ERC-2011-StG-279579 and by Xunta de Galicia (Conselleria de Educacion) – H.P. is part of the Strategic Unit AGRUP2015/11. P.P. gratefully acknowledges the financial support from the Magnus Ehrnrooth Foundation.

References

[1] J.J. Aubert, et al., The ratio of the nucleon structure functions F_2^N for iron and deuterium, Phys. Lett. B 123 (1983) 275–278, http://dx.doi.org/10.1016/0370-2693(83)90437-9.

[2] M. Arneodo, Nuclear effects in structure functions, Phys. Rep. 240 (1994) 301–393, http://dx.doi.org/10.1016/0370-1573(94)90048-5.

[3] S. Malace, D. Gaskell, D.W. Higinbotham, I. Cloet, The challenge of the EMC effect: existing data and future directions, Int. J. Mod. Phys. E 23 (08) (2014) 1430013, http://dx.doi.org/10.1142/S0218301314300136, arXiv:1405.1270.

[4] K.J. Eskola, H. Paukkunen, C.A. Salgado, EPS09: a new generation of NLO and LO nuclear parton distribution functions, J. High Energy Phys. 04 (2009) 065, http://dx.doi.org/10.1088/1126-6708/2009/04/065, arXiv:0902.4154.

[5] D. de Florian, R. Sassot, P. Zurita, M. Stratmann, Global analysis of nuclear parton distributions, Phys. Rev. D 85 (2012) 074028, http://dx.doi.org/10.1103/PhysRevD.85.074028, arXiv:1112.6324.

[6] K. Kovarik, et al., nCTEQ15 – global analysis of nuclear parton distributions with uncertainties in the CTEQ framework, Phys. Rev. D 93 (8) (2016) 085037, http://dx.doi.org/10.1103/PhysRevD.93.085037, arXiv:1509.00792.

[7] K.J. Eskola, Global analysis of nuclear PDFs – latest developments, Nucl. Phys. A 910–911 (2013) 163–170, http://dx.doi.org/10.1016/j.nuclphysa.2012.12.029, arXiv:1209.1546.

[8] H. Paukkunen, Nuclear PDFs in the beginning of the LHC era, Nucl. Phys. A 926 (2014) 24–33, http://dx.doi.org/10.1016/j.nuclphysa.2014.04.001, arXiv:1401.2345.

[9] N. Armesto, H. Paukkunen, J.M. Penín, C.A. Salgado, P. Zurita, An analysis of the impact of LHC Run I proton–lead data on nuclear parton densities, Eur. Phys. J. C 76 (4) (2016) 218, http://dx.doi.org/10.1140/epjc/s10052-016-4078-9, arXiv:1512.01528.

[10] H. Paukkunen, C.A. Salgado, Agreement of neutrino deep inelastic scattering data with global fits of parton distributions, Phys. Rev. Lett. 110 (21) (2013) 212301, http://dx.doi.org/10.1103/PhysRevLett.110.212301, arXiv:1302.2001.

[11] S.J. Brodsky, I. Schmidt, J.-J. Yang, Nuclear antishadowing in neutrino deep inelastic scattering, Phys. Rev. D 70 (2004) 116003, http://dx.doi.org/10.1103/PhysRevD.70.116003, arXiv:hep-ph/0409279.

[12] W.-C. Chang, I. Cloet, D. Dutta, J.-C. Peng, Probing flavor-dependent EMC effect with W boson production, Phys. Lett. B 720 (2013) 188–191, http://dx.doi.org/10.1016/j.physletb.2013.02.017, arXiv:1109.3108.

[13] I.C. Cloet, W. Bentz, A.W. Thomas, Parity-violating DIS and the flavour dependence of the EMC effect, Phys. Rev. Lett. 109 (2012) 182301, http://dx.doi.org/10.1103/PhysRevLett.109.182301, arXiv:1202.6401.

[14] D. Dutta, J.C. Peng, I.C. Cloet, D. Gaskell, Pion-induced Drell–Yan processes and the flavor-dependent EMC effect, Phys. Rev. C 83 (2011) 042201, http://dx.doi.org/10.1103/PhysRevC.83.042201, arXiv:1007.3916.

[15] J. Badier, et al., Test of nuclear effects in hadronic dimuon production, Phys. Lett. B 104 (1981) 335, http://dx.doi.org/10.1016/0370-2693(81)90137-4. In: Lisbon 1981, Proceedings, High Energy Physics, 1981, p. 807.

[16] P. Bordalo, et al., Nuclear effects on the nucleon structure functions in hadronic high mass dimuon production, Phys. Lett. B 193 (1987) 368, http://dx.doi.org/10.1016/0370-2693(87)91253-6.

[17] J.G. Heinrich, et al., Measurement of the ratio of sea to valence quarks in the nucleon, Phys. Rev. Lett. 63 (1989) 356–359, http://dx.doi.org/10.1103/PhysRevLett.63.356.

[18] M. Gluck, E. Reya, A. Vogt, Pionic parton distributions, Z. Phys. C 53 (1992) 651–656, http://dx.doi.org/10.1007/BF01559743.

[19] P.J. Sutton, A.D. Martin, R.G. Roberts, W.J. Stirling, Parton distributions for the pion extracted from Drell–Yan and prompt photon experiments, Phys. Rev. D 45 (1992) 2349–2359, http://dx.doi.org/10.1103/PhysRevD.45.2349.

[20] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, C.P. Yuan, New parton distribution functions from a global analysis of quantum chromodynamics, Phys. Rev. D 93 (3) (2016) 033006, http://dx.doi.org/10.1103/PhysRevD.93.033006, arXiv:1506.07443.

[21] J.M. Campbell, R.K. Ellis, W.T. Giele, A multi-threaded version of MCFM, Eur. Phys. J. C 75 (6) (2015) 246, http://dx.doi.org/10.1140/epjc/s10052-015-3461-2, arXiv:1503.06182.

[22] P.M. Nadolsky, Z. Sullivan, PDF uncertainties in WH production at Tevatron, eConf C 010630 (2001) P510, arXiv:hep-ph/0110378.

[23] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H.L. Lai, W.K. Tung, Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method, Phys. Rev. D 65 (2001) 014012, http://dx.doi.org/10.1103/PhysRevD.65.014012, arXiv:hep-ph/0101051.

[24] M.R. Adams, et al., Shadowing in inelastic scattering of muons on carbon, calcium and lead at low x(Bj), Z. Phys. C 67 (1995) 403–410, http://dx.doi.org/10.1007/BF01624583, arXiv:hep-ex/9505006.

[25] M. Arneodo, et al., The A dependence of the nuclear structure function ratios, Nucl. Phys. B 481 (1996) 3–22, http://dx.doi.org/10.1016/S0550-3213(96)90117-0.

[26] M. Corden, et al., Production of muon pairs in the continuum region by 39.5-GeV/c $\pi^{\pm}$, $K^{\pm}$, p and $\bar{p}$ beams incident on a tungsten target, Phys. Lett. B 96 (1980) 417–421, http://dx.doi.org/10.1016/0370-2693(80)90800-X.

[27] H. Paukkunen, K.J. Eskola, P. Paakkinen, C.A. Salgado, EPPS16: Nuclear parton distributions with LHC data, arXiv:1612.05741 [hep-ph].

II

EPPS16: NUCLEAR PARTON DISTRIBUTIONS WITH LHC DATA

by

Eskola, K. J., Paakkinen, P., Paukkunen, H. & Salgado C. A., 2017

Eur. Phys. J. C77, no.3 (2017) 163

arXiv: 1612.05741 [hep-ph]

Eur. Phys. J. C (2017) 77:163
DOI 10.1140/epjc/s10052-017-4725-9

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

EPPS16: nuclear parton distributions with LHC data

Kari J. Eskola[1,2,a], **Petja Paakkinen**[1,b], **Hannu Paukkunen**[1,2,3,c], **Carlos A. Salgado**[3,d]

[1] University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland
[2] Helsinki Institute of Physics, P.O. Box 64, FI-40014 University of Helsinki, Finland
[3] Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, 15782 Galicia, Spain

Received: 21 December 2016 / Accepted: 27 February 2017 / Published online: 16 March 2017

Abstract We introduce a global analysis of collinearly factorized nuclear parton distribution functions (PDFs) including, for the first time, data constraints from LHC proton–lead collisions. In comparison to our previous analysis, EPS09, where data only from charged-lepton–nucleus deep inelastic scattering (DIS), Drell–Yan (DY) dilepton production in proton–nucleus collisions and inclusive pion production in deuteron–nucleus collisions were the input, we now increase the variety of data constraints to cover also neutrino–nucleus DIS and low-mass DY production in pion–nucleus collisions. The new LHC data significantly extend the kinematic reach of the data constraints. We now allow much more freedom for the flavor dependence of nuclear effects than in other currently available analyses. As a result, especially the uncertainty estimates are more objective flavor by flavor. The neutrino DIS plays a pivotal role in obtaining a mutually consistent behavior for both up and down valence quarks, and the LHC dijet data clearly constrain gluons at large momentum fraction. Mainly for insufficient statistics, the pion–nucleus DY and heavy-gauge-boson production in proton–lead collisions impose less visible constraints. The outcome – a new set of next-to-leading order nuclear PDFs called EPPS16 – is made available for applications in high-energy nuclear collisions.

1 Introduction

Proton–lead (pPb) and lead–lead (PbPb) collisions at the Large Hadron Collider (LHC) have brought heavy-ion physics to the high-energy realm [1–4]. A more than tenfold increase in the center-of-mass energy with respect to the deuteron–gold (DAu) collisions at the Relativistic Heavy-Ion Collider (RHIC) has made it possible to study novel hard-process observables in a heavy-ion environment. For example, production cross sections of heavy-gauge bosons (Z and $W^{\pm}$) and jets have been measured. Because of the new experimental information from the LHC it is now also timely to update the pre-LHC global analyses of collinearly factorized nuclear parton distribution functions (PDFs) – for reviews, see e.g. Refs. [5,6].

The original idea of having nuclear effects in PDFs was data-driven as the early deep inelastic scattering (DIS) experiments unexpectedly revealed significant nuclear effects in the cross sections [7,8]. It was then demonstrated [9,10] that such effects in DIS and fixed nuclear-target Drell–Yan (DY) cross sections can be consistently described by modifying the free nucleon PDFs at low Q^2 and letting the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution [11–17] take care of the Q^2 dependence. In other words, the data were in line with a concept that the measured nuclear effects are of non-perturbative origin but at sufficiently high Q^2 there is no fundamental difference in the scattering off a nucleon or off a nucleus. These ideas eventually led to the first global fit and the EKS98 set of leading-order nuclear PDFs [18,19]. Since then, several parametrizations based on the DIS and DY data have been released at leading order (EKPS [20], HKM [21], HKN04 [22]), next-to-leading order (nDS [23], HKN07 [24], nCTEQ [25], AT12 [26]), and next-to-next-to-leading order (KA15 [27]) perturbative QCD.[1] For the rather limited kinematic coverage of the fixed-target data and the fact that only two types of data were used in these fits, significant simplifying assumptions had to be made e.g. with respect to the flavor dependence of the nuclear effects. The constraints on the gluon distribution are also weak in these analyses, and it is only along with the RHIC pion data [31] that an observable carrying direct information on the nuclear gluons has been

[a] e-mail: kari.eskola@jyu.fi
[b] e-mail: petja.paakkinen@jyu.fi
[c] e-mail: hannu.paukkunen@jyu.fi
[d] e-mail: carlos.salgado@usc.es

[1] For studies addressing origins of the nuclear effects, see e.g. Refs. [28–30].

added to the global fits – first in EPS08 [32] and EPS09 [33], later in DSSZ [34] and nCTEQ15 [35]. The interpretation of the RHIC pion production data is not, however, entirely unambiguous as the parton-to-pion fragmentation functions (FFs) may as well undergo a nuclear modification [36]. This approach was adopted in the DSSZ fit, and consequently their gluons show clearly weaker nuclear effects than in EPS09 (and nCTEQ15) where the FFs were considered to be free from nuclear modifications. To break the tie, more data and new observables were called for. To this end, the recent LHC dijet measurements [37] from pPb collisions have been most essential as a consistent description of these data is obtained with EPS09 and nCTEQ15 but not with DSSZ [38,39].

Another observable that has caused some controversy and debate during the past years is the neutrino–nucleus DIS. It has been claimed [40] (see also Ref. [41]) that the nuclear PDFs required to correctly describe neutrino data are different from those optimal for the charged-lepton induced DIS measurements. However, it has been demonstrated [42,43] that problems appear only in the case of one single data set and, furthermore, that it seems to be largely a normalization issue (which could e.g. be related to the incident neutrino flux which is model-dependent). The neutrino data were also used in the DSSZ fit without visible difficulties.

New data from the LHC 2013 p-Pb run have gradually become available and their impact on the nuclear PDFs has been studied [39,44] in the context of PDF reweighting [45]. Apart from the aforementioned dijet data [37] which will e.g. require a complete renovation of the DSSZ approach, the available W [46,47] and Z [48,49] data were found to have only a rather mild effect mainly for the limited statistical precision of the data. However, the analysis of Ref. [39] used only nuclear PDFs (EPS09, DSSZ) in which flavor-independent valence and light sea-quark distributions were assumed at the parametrization scale. Thus, it could not reveal the possible constraints that these electroweak observables could have for a particular quark flavor. On the other hand, the analysis of Ref. [44] involves some flavor dependence but the usage of absolute cross sections which are sensitive to the free-proton baseline PDFs complicates the interpretation of the results.

In the present paper, we update the EPS09 analysis by adding a wealth of new data from neutrino DIS [50], pion–nucleus DY process [51–53], and especially LHC pPb dijet [37], Z [48,49] and W [46] production. Thus, we take the global nuclear PDF fits onto a completely new level in the variety of data types. In addition, in comparison to EPS09, a large part of the whole framework is upgraded: we switch to a general-mass formalism for the heavy quarks, relax the assumption of the flavor independent nuclear modifications for quarks at the parametrization scale, undo the isospin corrections that some experiments had applied on their data, and also importantly, we now assign no extra weights to any of the data sets. In this updated analysis, we find no significant tension between the data sets considered, which lends support to the assumption of process-independent nuclear PDFs in the studied kinematical region. The result of the analysis presented in this paper is also published as a new set of next-to-leading order (NLO) nuclear PDFs, which we call EPPS16 and which supersedes our earlier set EPS09. The new EPPS16 set will be available at [54].

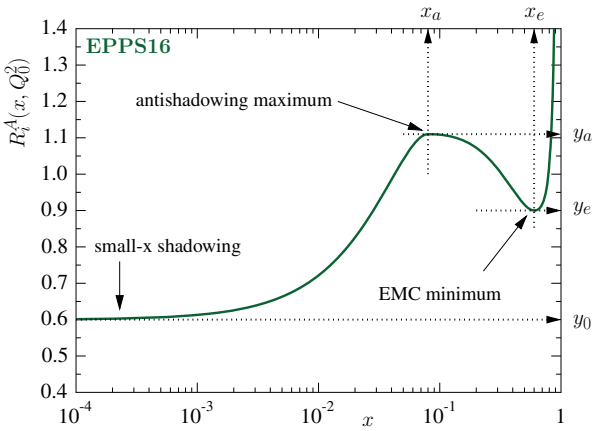

Fig. 1 Illustration of the EPPS16 fit function $R_i^A(x, Q_0^2)$

2 Parametrization of nuclear PDFs

Similarly to our earlier work, the bound proton PDF $f_i^{\mathrm{p}/A}(x, Q^2)$ for mass number A and parton species i is defined relative to the free-proton PDF $f_i^{\mathrm{p}}(x, Q^2)$ as

$$f_i^{\mathrm{p}/A}(x, Q^2) = R_i^A(x, Q^2) f_i^{\mathrm{p}}(x, Q^2), \tag{1}$$

where $R_i^A(x, Q^2)$ is the scale-dependent nuclear modification. Our free-proton baseline is CT14NLO [55]. Consistently with this choice, our analysis here uses the SACOT (simplified Aivazis–Collins–Olness–Tung) general-mass variable flavor number scheme [56–58] for the DIS cross sections. The fit function for the nuclear modifications $R_i^A(x, Q_0^2)$ at the parametrization scale Q_0^2, illustrated in Fig. 1, is also largely inherited from our earlier analyses [18,20,32,33],

$$R_i^A(x, Q_0^2) = \begin{cases} a_0 + a_1(x - x_a)^2 & x \le x_a \\ b_0 + b_1 x^\alpha + b_2 x^{2\alpha} + b_3 x^{3\alpha} & x_a \le x \le x_e \\ c_0 + (c_1 - c_2 x)\,(1 - x)^{-\beta} & x_e \le x \le 1, \end{cases} \tag{2}$$

where $\alpha = 10 x_a$ and the i and A dependencies of the parameters on the r.h.s. are left implicit.[2] The purpose of

[2] See Ref. [59] for a study experimenting with a more flexible fit function at small x.

the exponent α is to avoid the "plateau" that would otherwise (that is, if $\alpha = 1$) develop if $x_a < 0.1$. The coefficients a_i, b_i, c_i are fully determined by the asymptotic small-x limit $y_0 = R_i^A(x \to 0, Q_0^2)$, the antishadowing maximum $y_a = R_i^A(x_a, Q_0^2)$ and the EMC minimum $y_e = R_i^A(x_e, Q_0^2)$, as well as requiring continuity and vanishing first derivatives at the matching points x_a and x_e. The A dependencies of y_0, y_a, y_e are parametrized as

$$y_i(A) = y_i(A_{\rm ref}) \left(\frac{A}{A_{\rm ref}}\right)^{\gamma_i[y_i(A_{\rm ref})-1]}, \tag{3}$$

where $\gamma_i \geq 0$ and $A_{\rm ref} = 12$. By construction, the nuclear effects (deviations from unity) are now larger for heavier nuclei. Without the factor $y_i(A_{\rm ref}) - 1$ in the exponent one can more easily fall into a peculiar situation in which e.g. $y_i(A_{\rm ref}) < 1$, but $y_i(A \gg A_{\rm ref}) > 1$, which seems physically unlikely. For the valence quarks and gluons the values of y_0 are determined by requiring the sum rules

$$\int_0^1 {\rm d}x f_{u_{\rm V}}^{{\rm p}/A}(x, Q_0^2) = 2, \tag{4}$$

$$\int_0^1 {\rm d}x f_{d_{\rm V}}^{{\rm p}/A}(x, Q_0^2) = 1, \tag{5}$$

$$\int_0^1 {\rm d}x x \sum_i f_i^{{\rm p}/A}(x, Q_0^2) = 1, \tag{6}$$

separately for each nucleus and thus the A dependence of these y_0 is not parametrized. All other parameters than y_0, y_a, y_e are A-independent. In our present framework we consider the deuteron ($A = 2$) to be free from nuclear effects though few-percent effects at high x are found e.g. in Ref. [60]. The bound neutron PDFs $f_i^{{\rm n}/A}(x, Q^2)$ are obtained from the bound proton PDFs by assuming isospin symmetry,

$$f_{u,\bar{u}}^{{\rm n}/A}(x, Q^2) = f_{d,\bar{d}}^{{\rm p}/A}(x, Q^2), \tag{7}$$

$$f_{d,\bar{d}}^{{\rm n}/A}(x, Q^2) = f_{u,\bar{u}}^{{\rm p}/A}(x, Q^2), \tag{8}$$

$$f_i^{{\rm n}/A}(x, Q^2) = f_i^{{\rm p}/A}(x, Q^2) \quad \text{for other flavors.} \tag{9}$$

Above the parametrization scale $Q^2 > Q_0^2$ the nuclear PDFs are obtained by solving the DGLAP evolution equations with two-loop splitting functions [61,62]. We use our own DGLAP evolution code which is based on the solution method described in Ref. [63] and also explained and benchmarked in Ref. [64]. Our parametrization scale Q_0^2 is fixed to the charm pole mass $Q_0^2 = m_c^2$ where $m_c = 1.3\,{\rm GeV}$. The bottom quark mass is $m_b = 4.75\,{\rm GeV}$ and the value of the strong coupling constant is set by $\alpha_{\rm s}(M_{\rm Z}) = 0.118$, where $M_{\rm Z}$ is the mass of the Z boson.

As is well known, at NLO and beyond the PDFs do not need to be positive definite and we do not impose such a restriction either. In fact, doing so would be artificial since the parametrization scale is, in principle, arbitrary, and positive definite PDFs, say, at $Q_0^2 = m_c^2$ may easily correspond to negative small-x PDFs at a scale just slightly below Q_0^2. As we could have equally well parametrized the PDFs at such a lower value of Q_0^2, we see that restricting the PDFs to be always positive would be an unphysical requirement.

3 Experimental data

All the $\ell^- A$ DIS, pA DY and RHIC DAu pion data sets we use in the present analysis are the same as in the EPS09 fit. The only modification on this part is that we now remove the isoscalar corrections of the EMC, NMC and SLAC data (see the next subsection), which is important as we have freed the flavor dependence of the quark nuclear modifications. The $\ell^- A$ DIS data (cross sections or structure functions F_2) are always normalized by the ℓ^-D measurements and, as in EPS09, the only kinematic cut on these data is $Q^2 > m_{\rm c}^2$. This is somewhat lower than in typical free-proton fits and the implicit assumption is (also in not setting a cut in the mass of the hadronic final state) that the possible higher-twist effects will cancel in ratios of structure functions/cross sections. While potential signs of $1/Q^2$ effects have been seen in the HERA data [65] already around $Q^2 = 10\,{\rm GeV}^2$, these effects occur at significantly smaller x than what is the reach of the $\ell^- A$ DIS data.

From the older measurements, also pion–nucleus DY data from the NA3 [51], NA10 [52], and E615 [53] collaborations are now included. These data have been shown [66,67] to carry some sensitivity to the flavor-dependent EMC effect. However, more stringent flavor-dependence constraints at large x are provided by the CHORUS (anti)neutrino–Pb DIS data [50], whose treatment in the fit is detailedly explained in Sect. 3.2.

The present analysis is the first one to directly include LHC data. To this end, we use the currently published pPb data for heavy-gauge boson [46,48,49] and dijet production [37]. These observables have already been discussed in the literature [39,44,68–71] in the context of nuclear PDFs. Importantly, we include the LHC pPb data always as forward-to-backward ratios in which the cross sections at positive (pseudo)rapidities $\eta > 0$ are divided by the ones at negative rapidities $\eta < 0$. This is to reduce the sensitivity to the chosen free-proton baseline PDFs as well as to cancel the experimental luminosity uncertainty. However, upon taking the ratio part of the information is also lost as, for example, the points near $\eta = 0$ are, by construction, always close to unity and carry essentially no information. In addition, since the correlations on the systematic errors are not available, all the experimental uncertainties are added in quadrature when forming these ratios (except for the CMS W measure-

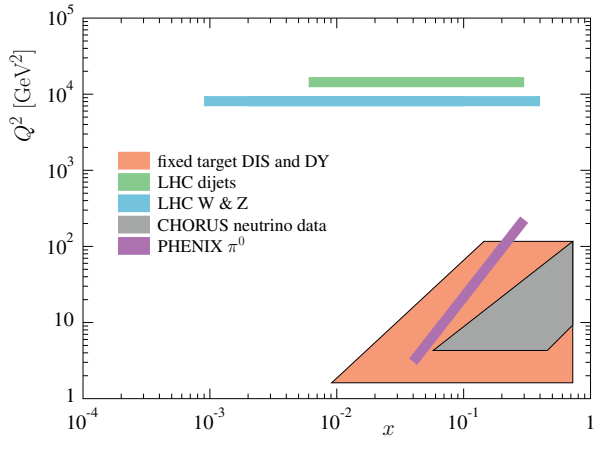

Fig. 2 The approximate regions in the (x, Q^2) plane at which different data in the EPPS16 fit probe the nuclear PDFs

ment [46] which is taken directly from the publication) which partly undermines the constraining power of these data. The baseline pp measurements performed at the same $\sqrt{s}$ as the pPb runs may, in the future, also facilitate a direct usage of the nuclear modification factors $\mathrm{d}\sigma^{\mathrm{pPb}}/\mathrm{d}\sigma^{\mathrm{pp}}$. The technicalities of how the LHC data are included in our analysis are discussed in Sect. 3.3.

In Fig. 2 we illustrate the predominant x and Q^2 regions probed by the data. Clearly, the LHC data probe the nuclear PDFs at much higher in Q^2 than the earlier DIS and DY data. For the wide rapidity coverage of the LHC detectors the new measurements also reach lower values of x than the old data, but for the limited statistical precision the constraints for the small-x end still remain rather weak. All the exploited data sets including the number of data points, their χ^2 contribution and references are listed in Table 1. We note that approximately half of the data are now for the ^{208}Pb nucleus while in the EPS09 analysis only 15 Pb data points (NMC 96) were included. Most of this change is caused by the inclusion of the CHORUS neutrino data.

3.1 Isoscalar corrections

Part of the charged-lepton DIS data that have been used in the earlier global nPDF fits had been "corrected", in the original publications, for the isospin effects. That is, the experimental collaborations had tried to eliminate the effects emerging from the unequal number of protons and neutrons when making the comparison with the deuteron data. In this way the ratios F_2^A/F_2^{D} could be directly interpreted in terms of nuclear effects in the PDFs. However, this is clearly an unnecessary operation from the viewpoint of global fits, which has previously caused some confusion regarding the nuclear valence-quark modifications: the particularly mild effects found in the nDS [23] and DSSZ [34] analyses (see Fig. 27) most likely originate from neglecting such a correction.

The structure function of a nucleus A with Z protons and N neutrons can be written as

$$F_2^A = \frac{Z}{A}F_2^{\mathrm{p},A} + \frac{N}{A}F_2^{\mathrm{n},A}, \tag{10}$$

where $F_2^{\mathrm{p},A}$ and $F_2^{\mathrm{n},A}$ are the structure functions of the bound protons and neutrons. The corresponding isoscalar structure function is defined as the one containing an equal number of protons and neutrons,

$$\hat{F}_2^A = \frac{1}{2}F_2^{\mathrm{p},A} + \frac{1}{2}F_2^{\mathrm{n},A}. \tag{11}$$

Using Eq. (10), the isoscalar structure function reads

$$\hat{F}_2^A = \beta F_2^A, \tag{12}$$

where

$$\beta = \frac{A}{2}\left(1 + \frac{F_2^{\mathrm{n},A}}{F_2^{\mathrm{p},A}}\right) \Big/ \left(Z + N\frac{F_2^{\mathrm{n},A}}{F_2^{\mathrm{p},A}}\right). \tag{13}$$

Usually, it has been assumed that the ratio $F_2^{\mathrm{n},A}/F_2^{\mathrm{p},A}$ is free from nuclear effects,

$$\frac{F_2^{\mathrm{n},A}}{F_2^{\mathrm{p},A}} = \frac{F_2^{\mathrm{n}}}{F_2^{\mathrm{p}}}, \tag{14}$$

and parametrized according to the DIS data from proton and deuteron targets. Different experiments have used different versions:

– EMC parametrization [78]:

$$\frac{F_2^{\mathrm{n}}}{F_2^{\mathrm{p}}} = 0.92 - 0.86x,$$

– SLAC parametrization [72]:

$$\frac{F_2^{\mathrm{n}}}{F_2^{\mathrm{p}}} = 1 - 0.8x,$$

– NMC parametrization [80]:

$$\frac{F_2^{\mathrm{n}}}{F_2^{\mathrm{p}}} = A(x)\left(\frac{Q^2}{20}\right)^{B(x)}\left(1 + \frac{x^2}{Q^2}\right)$$
$$A(x) = 0.979 - 1.692x + 2.797x^2 - 4.313x^3 + 3.075x^4$$
$$B(x) = -0.171x^2 + 0.244x^3.$$

Table 1 The data sets used in the EPPS16 analysis, listed in the order of growing nuclear mass number. The number of data points and their contribution to χ^2 counts only those data points that fall within the kinematic cuts explained in the text. The new data with respect to the EPS09 analysis are marked with a superscript a

Experiment	Observable	Collisions	Data points	χ^2	References
SLAC E139	DIS	e^-He(4), e^-D	21	12.2	[72]
CERN NMC 95, re	DIS	μ^-He(4), μ^-D	16	18.0	[73]
CERN NMC 95	DIS	μ^-Li(6), μ^-D	15	18.4	[74]
CERN NMC 95, Q^2 dep	DIS	μ^-Li(6), μ^-D	153	161.2	[74]
SLAC E139	DIS	e^-Be(9), e^-D	20	12.9	[72]
CERN NMC 96	DIS	μ^-Be(9), μ^-C	15	4.4	[75]
SLAC E139	DIS	e^-C(12), e^-D	7	6.4	[72]
CERN NMC 95	DIS	μ^-C(12), μ^-D	15	9.0	[74]
CERN NMC 95, Q^2 dep	DIS	μ^-C(12), μ^-D	165	133.6	[74]
CERN NMC 95, re	DIS	μ^-C(12), μ^-D	16	16.7	[73]
CERN NMC 95, re	DIS	μ^-C(12), μ^-Li(6)	20	27.9	[73]
FNAL E772	DY	pC(12), pD	9	11.3	[76]
SLAC E139	DIS	e^-Al(27), e^-D	20	13.7	[72]
CERN NMC 96	DIS	μ^-Al(27), μ^-C(12)	15	5.6	[75]
SLAC E139	DIS	e^-Ca(40), e^-D	7	4.8	[72]
FNAL E772	DY	pCa(40), pD	9	3.33	[76]
CERN NMC 95, re	DIS	μ^-Ca(40), μ^-D	15	27.6	[73]
CERN NMC 95, re	DIS	μ^-Ca(40), μ^-Li(6)	20	19.5	[73]
CERN NMC 96	DIS	μ^-Ca(40), μ^-C(12)	15	6.4	[75]
SLAC E139	DIS	e^-Fe(56), e^-D	26	22.6	[72]
FNAL E772	DY	e^-Fe(56), e^-D	9	3.0	[76]
CERN NMC 96	DIS	μ^-Fe(56), μ^-C(12)	15	10.8	[75]
FNAL E866	DY	pFe(56), pBe(9)	28	20.1	[77]
CERN EMC	DIS	μ^-Cu(64), μ^-D	19	15.4	[78]
SLAC E139	DIS	e^-Ag(108), e^-D	7	8.0	[72]
CERN NMC 96	DIS	μ^-Sn(117), μ^-C(12)	15	12.5	[75]
CERN NMC 96, Q^2 dep	DIS	μ^-Sn(117), μ^-C(12)	144	87.6	[79]
FNAL E772	DY	pW(184), pD	9	7.2	[76]
FNAL E866	DY	pW(184), pBe(9)	28	26.1	[77]
CERN NA10[a]	DY	π^-W(184), π^-D	10	11.6	[52]
FNAL E615[a]	DY	π^+W(184), π^-W(184)	11	10.2	[53]
CERN NA3[a]	DY	π^-Pt(195), π^-H	7	4.6	[51]
SLAC E139	DIS	e^-Au(197), e^-D	21	8.4	[72]
RHIC PHENIX	π^0	dAu(197), pp	20	6.9	[31]
CERN NMC 96	DIS	μ^-Pb(207), μ^-C(12)	15	4.1	[75]
CERN CMS[a]	$W^\pm$	pPb(208)	10	8.8	[46]
CERN CMS[a]	Z	pPb(208)	6	5.8	[48]
CERN ATLAS[a]	Z	pPb(208)	7	9.6	[49]
CERN CMS[a]	dijet	pPb(208)	7	5.5	[37]
CERN CHORUS[a]	DIS	νPb(208), $\overline{\nu}$Pb(208)	824	998.6	[50]
Total			1811	1789	

Using these functions we calculate the correction factors β thereby obtaining the ratios F_2^A/F_2^{D}, to be used in the fit, from the isoscalar versions $\hat{F}_2^A/F_2^{\mathrm{D}}$ reported by the experiments.

As discussed in Ref. [67], also the $\pi^- A$ DY data from the NA10 collaboration [52] have been balanced for the neutron excess. The correction was done by utilizing the leading-

order DY cross section. Here, we account for this with the isospin correction factor given in Eq. (8) of Ref. [67].

3.2 Treatment of neutrino DIS data

In the present work we make use of the CHORUS neutrino and antineutrino DIS data [50]. Similar measurements are available also from the CDHSW [81] and NuTeV [82] collaborations, but only for the CHORUS data the correlations of the systematic uncertainties are directly available in the form we need.[3] Moreover, the ^{208}Pb target has a larger neutron excess than the iron targets of CDHSW and NuTeV, thereby carrying more information on the flavor separation. The data are reported as double differential cross sections $d\sigma^{\nu,\overline{\nu}}_{i,\mathrm{exp}}/\mathrm{d}x\mathrm{d}y$ in the standard DIS variables and, guided by our free-proton baseline fit CT14NLO [55], the kinematic cuts we set on these data are $Q^2 > 4\,\mathrm{GeV}^2$ and $W^2 > 12.25\,\mathrm{GeV}^2$.[4] In the computation of these NLO neutrino DIS cross sections, we apply the dominant electroweak [83] and target-mass [84] corrections as in Refs. [42,43], together with the SACOT quark-mass scheme.

In order to suppress the theoretical uncertainties related to the free-proton PDFs, as well as experimental systematic uncertainties, we treat the data following the normalization prescription laid out in Ref. [43]. For each (anti)neutrino beam energy E, we compute the total cross section as

$$\sigma^{\nu,\overline{\nu}}_{\mathrm{exp}}(E) = \sum_i \frac{\mathrm{d}\sigma^{\nu,\overline{\nu}}_{i,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y}\Delta^{xy}_i \delta_{E,E_i}, \tag{15}$$

where E_i is the beam energy corresponding to the ith data point. By Δ^{xy}_i we mean the size of the (x, y) bin (rectangles) to which the ith data point belongs. The original data are then normalized by the estimated total cross sections of Eq. (15) as

$$\frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{i,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y} \equiv \frac{\mathrm{d}\sigma^{\nu,\overline{\nu}}_{i,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y} \Big/ \sigma^{\nu,\overline{\nu}}_{\mathrm{exp}}(E = E_i). \tag{16}$$

As discussed e.g. in [45,85], the χ^2 contribution of data with correlated uncertainties is obtained in terms of the covariance matrix C as

$$\chi^2 = \sum_{i,j}\left(\frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{i,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y} - \frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{i,\mathrm{th}}}{\mathrm{d}x\mathrm{d}y}\right) C^{-1}_{ij}\left(\frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{j,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y} - \frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{j,\mathrm{th}}}{\mathrm{d}x\mathrm{d}y}\right), \tag{17}$$

where now the theory values $\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{j,\mathrm{th}}/\mathrm{d}x\mathrm{d}y$ are the computed differential cross sections normalized by the corresponding integrated cross section (similarly to Eq. (16)). The elements of the covariance matrix are in our case defined as

$$C_{ij} \equiv (\tilde{\delta}^{\mathrm{stat}}_i)^2\delta_{ij} + \sum_k \tilde{\beta}^k_i\tilde{\beta}^k_j, \tag{18}$$

where the statistical uncertainty $\tilde{\delta}^{\mathrm{stat}}_i$ on $\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{i,\mathrm{exp}}/\mathrm{d}x\mathrm{d}y$ is computed from the original statistical uncertainties δ^{stat}_i by

$$\tilde{\delta}^{\mathrm{stat}}_i \equiv \delta^{\mathrm{stat}}_i/\sigma^{\nu,\overline{\nu}}_{\mathrm{exp}}(E_i). \tag{19}$$

Here we neglect the statistical uncertainty of $\sigma^{\nu,\overline{\nu}}(E)$ as for this integrated quantity it is always clearly smaller than that of the individual data points. The point-to-point correlated systematic uncertainties $\tilde{\beta}^k_i$ for the normalized data points we form as

$$\tilde{\beta}^k_i \equiv \left(\frac{\mathrm{d}\sigma^{\nu,\overline{\nu}}_{i,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y} + \beta^k_i\right) \Big/ \sigma^{\nu,\overline{\nu}}_k(E_i) - \frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{i,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y}, \tag{20}$$

where

$$\sigma^{\nu,\overline{\nu}}_k(E) = \sum_i \left(\frac{\mathrm{d}\sigma^{\nu,\overline{\nu}}_{i,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y} + \beta^k_i\right)\Delta^{xy}_i\delta_{E,E_i}. \tag{21}$$

Above, the index k labels the parameters controlling the experimental systematic uncertainties and β^k_i are the cross-section shifts corresponding to a one standard deviation change in the kth parameter. We note that $\tilde{\beta}^k_i$ in Eq. (20) for the relative cross sections in Eq. (16) are constructed such that if the β^k_i correspond only to the same relative normalization shift for all points, then $\tilde{\beta}^k_i$ are just zero. We also note that in Eq. (18) we have assumed that the response of $\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{i,\mathrm{exp}}/\mathrm{d}x\mathrm{d}y$ to the systematic uncertainty parameters is linear.

As shown in e.g. Ref. [42], the Q^2 dependence of nuclear effects in neutrino DIS data is weak. Hence, for a concise graphical presentation of the data as a function of x, we integrate over the y variable by

$$\frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{\mathrm{exp}}}{\mathrm{d}x}(E) = \sum_j \frac{\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{j,\mathrm{exp}}}{\mathrm{d}x\mathrm{d}y}\Delta^y_j\delta_{x,x_j}\delta_{E,E_j}, \tag{22}$$

where Δ^y_j is the size of the y bin to which the jth data point belongs, and x_j the corresponding value of the x variable. The overall statistical uncertainty to the relative cross section in Eq. (22) is computed as

$$\delta^{\mathrm{stat}}(E, x) = \sqrt{\sum_j \left(\tilde{\delta}^{\mathrm{stat}}_j\Delta^y_j\right)^2\delta_{x,x_j}\delta_{E,E_j}}, \tag{23}$$

[3] http://choruswww.cern.ch/Publications/DIS-data/.

[4] The cuts are more stringent here than for other DIS data as only absolute cross sections are available (instead of those relative to a lighter nucleus).

and the total systematic uncertainty is given by

$$\delta^{\rm sys}(E,x) = \sqrt{\sum_k \left[\delta_k^{\rm sys}(E,x)\right]^2}, \tag{24}$$

where

$$\delta_k^{\rm sys}(E,x) = \sum_j \tilde{\beta}_j^k \Delta_j^y \delta_{x,x_j} \delta_{E,E_j}. \tag{25}$$

In the plots for $\mathrm{d}\tilde{\sigma}^{\nu,\overline{\nu}}_{\rm exp}/\mathrm{d}x$ presented in Sect. 5 (Figs. 20, 21), the statistical and total systematic uncertainties have been added in quadrature. We also divide by the theory values obtained by using the CT14NLO free-proton PDFs (but still with the correct amount of protons and neutrons). We stress that Eqs. (22)–(25) are used only for a simple graphical presentation of the data but not for the actual fit.

3.3 Look-up tables for LHC observables and others

In order to efficiently include the LHC observables in our fit at the NLO level, a fast method to evaluate the cross sections is essential. We have adopted the following pragmatic approach: For a given observable, a hard-process cross section $\sigma^{\rm pPb}$ in pPb collisions, we set up a grid in the x variable of the Pb nucleus, $x_0, \ldots, x_N = 1$, and evaluate, for each x bin k and parton flavor j

$$\sigma^{\rm pPb}_{j,k} = \sum_i f_i^{\rm p} \otimes \hat{\sigma}_{ij} \otimes f^{\rm Pb}_{j,k}, \tag{26}$$

where $\hat{\sigma}_{ij}$ are the coefficient functions appropriate for a given process and $f^{\rm Pb}_{j,k}$ involve only proton PDFs with no nuclear modifications,

$$f^{\rm Pb}_{j,k}(x) \equiv \sum_\ell \left[Z f_\ell^{\rm p,Pb}(x) + N f_\ell^{\rm n,Pb}(x) \right]\Big|_{R_j^{\rm Pb}=1, R_{i\neq j}^{\rm Pb}=0} \times \theta\left(x - x_{k-1}\right)\theta\left(x_k - x\right). \tag{27}$$

Thus, the functions $f^{\rm Pb}_{j,k}$ pick up the partonic weight of the nuclear modification $R_j^{\rm Pb}$ in a given interval $x_{k-1} < x < x_k$. Since the nuclear modification factors R_i^A are relatively slowly varying functions in x (e.g. in comparison to the absolute PDFs), the observable $\sigma^{\rm pPb}$ can be computed as a sum of $\sigma^{\rm pPb}_{j,k}$ weighted by the appropriate nuclear modification,

$$\sigma^{\rm pPb} = \sum_{j,k} \sigma^{\rm pPb}_{j,k} R_j^{\rm Pb}(x_{k-1} < x < x_k). \tag{28}$$

As an illustration, in Fig. 3, we show the histograms of $\sigma^{\rm pPb}_{j,k}$ corresponding to W^+ production measured by CMS in the bin $1 < \eta_{\rm lab} < 1.5$. For the electroweak LHC observables

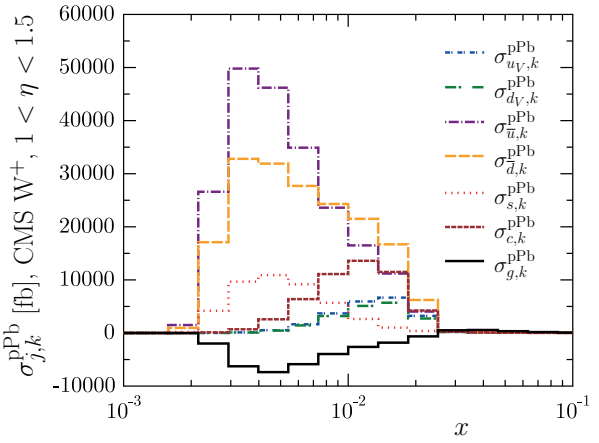

Fig. 3 An example of the $\sigma^{\rm pPb}_{j,k}$ histograms used in evaluating the LHC pPb cross sections in Eq. (28). The cross section $\sigma^{\rm pPb}$ is computed as a sum of all the bins weighted by the appropriate nuclear modification factors. The sum of all the bins gives the cross section with no nuclear modifications ($R_i^{\rm Pb} = 1$)

we have used the MCFM code [86] to compute the grids, and for dijet production the modified EKS code [87–89].

We set up similar grids also for inclusive pion production in DAu collisions at RHIC using the INCNLO [90] code with KKP FFs [91], and for the DY process in πA collisions using MCFM with the GRV pion PDFs [92]. In all cases, we have checked that the grids reproduce a direct evaluation of the observables within 1% accuracy in the case of EPS09 nuclear PDFs.

4 Analysis procedure

The standard statistical procedure for comparing experimental data to theory is to inspect the behavior of the overall χ^2 function, defined as

$$\chi^2(\mathbf{a}) \equiv \sum_k \chi_k^2(\mathbf{a}), \tag{29}$$

where $\mathbf{a}$ is a set of theory parameters and $\chi_k^2(\mathbf{a})$ denotes the contribution of each independent data set k,

$$\chi_k^2(\mathbf{a}) \equiv \sum_{i,j} \left[T_i(\mathbf{a}) - D_i\right] C_{ij}^{-1} \left[T_j(\mathbf{a}) - D_j\right]. \tag{30}$$

Here, $T_i(\mathbf{a})$ denote the theoretical values of the observables in the data set k, D_i are the corresponding experimental values, and C_{ij} is the covariance matrix. In most cases, only the total uncertainty is known, and in this case $C_{ij} = (\delta_i^{\rm uncorr.})^2 \delta_{ij}$, where $\delta_i^{\rm uncorr.}$ is the point-to-point uncorrelated data uncertainty. In the case that the only correlated uncertainty is the overall normalization $\delta^{\rm norm.}$, we can

also write

$$\chi_k^2(\mathbf{a}) = \left(\frac{1-f_N}{\delta^{\text{norm.}}}\right)^2 + \sum_i \left[\frac{T_i(\mathbf{a}) - f_N D_i}{\delta_i^{\text{uncorr.}}}\right]^2, \tag{31}$$

which is to be minimized with respect to f_N. All the uncertainties are considered additive (e.g. the possible D'Agostini bias [93] or equivalent is neglected). The central fit is then defined to correspond to the minimum value of the global χ^2 obtainable with a given set of free parameters,

$$\chi^2\left(\mathbf{a}^0\right) \equiv \min[\chi^2(\mathbf{a})]. \tag{32}$$

In practice, we minimize the χ^2 function using the Levenberg–Marquardt method [94–96].

In our previous EPS09 analysis, additional weight factors were included in Eq. (29) to increase the importance of some hand-picked data sets. We emphasize that in the present EPPS16 study we have abandoned this practice due to the subjectiveness it entails. In the EPS09 analysis the use of such data weights was also partially related to technical difficulties in finding a stable minimum of $\chi^2(\mathbf{a})$ when using the `MINUIT` [97] library. In the EPS09 analysis an additional penalty term was also introduced to the $\chi^2(\mathbf{a})$ function to avoid unphysical A dependence at small x (i.e. to have larger nuclear effects for larger nuclei). Here, such a term is not required because of the improved functional form discussed in Sect. 2.

As the nuclear PDFs are here allowed to go negative it is also possible to drift to a situation in which the longitudinal structure function F_{L}^A becomes negative. To avoid this, we include penalty terms in $\chi^2(\mathbf{a})$ at small x that grow quickly if $F_{\text{L}}^A < 0$. We observe, however, that the final results in EPPS16 are not sensitive to such a positiveness requirement.

4.1 Uncertainty analysis

As in our earlier analysis EPS09, we use the Hessian-matrix-based approach to estimate the PDF uncertainties [98]. The dominant behavior of the global χ^2 about the fitted minimum can be written as

$$\chi^2(\mathbf{a}) \approx \chi_0^2 + \sum_{ij} \delta a_i H_{ij} \delta a_j, \tag{33}$$

where $\delta a_j \equiv a_j - a_j^0$ are differences from the best-fit values and $\chi_0^2 \equiv \chi^2(\mathbf{a}^0)$ is the lowest attainable χ^2 of Eq. (32). The Hessian matrix H_{ij} can be diagonalized by defining a new set of parameters by

$$z_k \equiv \sum_j D_{kj} \delta a_j, \tag{34}$$

with

$$D_{kj} \equiv \sqrt{\epsilon_k} v_j^{(k)}, \tag{35}$$

where ϵ_k are the eigenvalues and $v_j^{(k)}$ are the components of the corresponding orthonormal eigenvectors of the Hessian matrix,

$$H_{ij} v_j^{(k)} = \epsilon_k v_i^{(k)}, \tag{36}$$

$$\sum_i v_i^{(k)} v_i^{(\ell)} = \sum_i v_k^{(i)} v_\ell^{(i)} = \delta_{k\ell}. \tag{37}$$

In these new coordinates,

$$\chi^2(\mathbf{z}) \approx \chi_0^2 + \sum_i z_i^2. \tag{38}$$

In comparison to Eq. (33), here in Eq. (38) all the correlations among the original parameters a_i are hidden in the definition Eq. (34), which facilitates a very simple error propagation [98]. Indeed, since the directions z_i are uncorrelated, the upward/downward-symmetric uncertainty for any PDF-dependent quantity $\mathcal{O}$ can be written as

$$\Delta\mathcal{O} = \sqrt{\sum_i (\Delta z_i)^2 \left(\frac{\partial \mathcal{O}}{\partial z_i}\right)^2}, \tag{39}$$

with an uncertainty interval $\Delta z_i = (t_i^+ + t_i^-)/2$ where $t_i^\pm$ are z_i-interval limits which depend on the chosen tolerance criterion. The partial derivatives in Eq. (39) are evaluated with the aid of PDF error sets $S_i^\pm$ defined in the space of z_i coordinates in terms of $t_i^\pm$ as

$$\begin{aligned} \mathbf{z}(S_1^\pm) &= \pm t_1^\pm (1, 0, \ldots, 0), \\ &\vdots \\ \mathbf{z}(S_N^\pm) &= \pm t_N^\pm (0, 0, \ldots, 1), \end{aligned} \tag{40}$$

where N is the number of the original parameters a_i. It then follows that

$$\Delta\mathcal{O} = \frac{1}{2}\sqrt{\sum_i \left[\mathcal{O}\left(S_i^+\right) - \mathcal{O}\left(S_i^-\right)\right]^2}. \tag{41}$$

Although simple on paper, in practice it is a non-trivial task to obtain a sufficiently accurate Hessian matrix in a multivariate fit such that Eq. (38) would be accurate. One possibility, used e.g. in Ref. [99], is to use the linearized Hessian matrix obtained from Eq. (30)

$$H_{ij}^{\text{linearized}} \equiv \sum_{k,\ell} \frac{\partial T_k}{\partial a_i} C_{k\ell}^{-1} \frac{\partial T_\ell}{\partial a_j}, \tag{42}$$

where the partial derivatives are evaluated by finite differences. The advantage is that by this definition, the Hessian matrix is always positive definite and thereby has automatically positive eigenvalues and e.g. Eq. (34) is always well defined.

Another possibility, which is the option chosen in the present study, is to scan the neighborhood of the minimum χ^2 and fit it with an ansatz

$$\chi^2(\mathbf{a}) = \chi_0^2 + \sum_{i,j} \delta a_i h_{ij} \delta a_j, \tag{43}$$

whose parameters h_{ij} then correspond to the components of the Hessian matrix. While this gives more accurate results than the linearized method (where some information is thrown away), the eigenvalues of the Hessian become easily negative for the presence of third- and higher-order components in the true χ^2 profile. Hence, to arrive at positive-definite eigenvalues, some manual labour is typically required e.g. in tuning the parameter intervals used when scanning the global χ^2. Yet, the resulting uncertainties always depend somewhat on the chosen parameter intervals, especially when the uncertainties are large. To improve the precision, we have adopted an iterative procedure similar to the one in Ref. [100]: After having obtained the first estimate for the Hessian matrix and the z coordinates, we recompute the Hessian matrix in the z space by re-scanning the vicinity of $\mathbf{z} = 0$ and fitting it with a polynomial

$$\chi^2(\mathbf{z}) = \chi_0^2 + \sum_{i,j} z_i \hat{h}_{ij} z_j, \tag{44}$$

where $\hat{h}_{ij}$ is an estimate for the Hessian matrix in the z space. We then re-define the z coordinates by

$$z_k \to \sum_{\ell} \hat{D}_{k\ell} \delta a_\ell, \tag{45}$$

where

$$\hat{D}_{k\ell} \equiv \sum_j \sqrt{\hat{\epsilon}_k} \hat{v}_j^{(k)} D_{j\ell}, \tag{46}$$

and $\hat{\epsilon}_k$ and $\hat{\mathbf{v}}^{(k)}$ are now the eigenvalues and eigenvectors of the matrix $\hat{h}_{ij}$. Then we repeat the iteration a few times, using the $\hat{D}_{ij}$ of the previous round as D_{ij} in Eq. (46). Ideally, one should find that the eigenvalues $\hat{\epsilon}_k$ converge to unity during the iteration but in practice, some deviations will always persist for the presence of non-quadratic components in the true χ^2 profile. We have also noticed that, despite the iteration, the resulting uncertainty bands still depend somewhat on the finite step sizes and grids used in the χ^2-profile scanning especially in the regions where the uncertainties are large. In such regions the Hessian method starts to be unreliable and the uncertainties found represent only the lower limits for the true uncertainties.

The global χ^2 profiles as a function of the final eigenvector directions, which we arrive at in the present EPPS16 analysis, are shown in Fig. 4. In obtaining these, during the iteration, the finite step sizes (z_i in Eq. (44)) along each provisional eigenvector direction were adjusted such that the total χ^2 increased by 5 units from the minimum. As seen in the figure, in most cases, the quadratic approximation gives a very good description of the true behavior of χ^2, but in some cases higher-order (e.g. cubic and quartic) components are evidently present. The effects of higher-order components can be partly compensated by using larger step sizes during the iteration such that the quadratic polynomial approximates the true χ^2 better up to larger deviations from the minimum (but is less accurate near the minimum). However, we have noticed that with increasing step sizes the resulting PDF uncertainties get eventually smaller, which indicates that some corners of the parameter space are not covered as completely as with the now considered 5-unit increase in χ^2.

The basic idea in the determination of the PDF uncertainty sets in the present work is similar to that in the EPS09 analysis. As in EPS09, for each data set k with N_k data points we determine a 90% confidence limit $\chi^2_{k,\max}$ by solving

$$\int_0^{M_k} \frac{d\chi^2}{2\Gamma(N_k/2)} \left(\frac{\chi^2}{2}\right)^{N_k/2-1} \exp\left(-\chi^2/2\right) = 0.90, \tag{47}$$

where

$$M_k = \chi^2_{k,\max} \times \frac{N_k - 2}{\chi^2_{k,0}}, \tag{48}$$

and in which $\chi^2_{k,0}$ is the value of χ^2 for kth data set at the global minimum. The integrand in Eq. (47) is the usual χ^2 distribution – the probability density to obtain a given value of χ^2 when the data are distributed in a Gaussian way around the known truth. The effect of Eq. (48) is, as sketched in Fig. 5, to scale the χ^2 distribution such that its maximum occurs at the central value of the fit $\chi^2_{0,k}$, against which the confidence limit is defined. In other words, we assume that if the experiment would be repeated several times the outcome would follow the scaled distribution (the blue curve in Fig. 5) and not the ideal one (the green curve in Fig. 5). This procedure allows one to define confidence limits also for data sets which happen to give very large χ^2_k/N_k for e.g. underestimated uncertainties or particularly large fluctuations [101].

For each eigenvector direction z_i and data set k we find the interval $[z^k_{i,\min}, z^k_{i,\max}]$ for which $\chi^2_k < \chi^2_{k,\max}$. Looping over all the data sets k we then find the intersection of the intervals $[z^k_{i,\min}, z^k_{i,\max}]$ for each i. In other words, we require all the individual data sets to remain within the defined 90% limit,

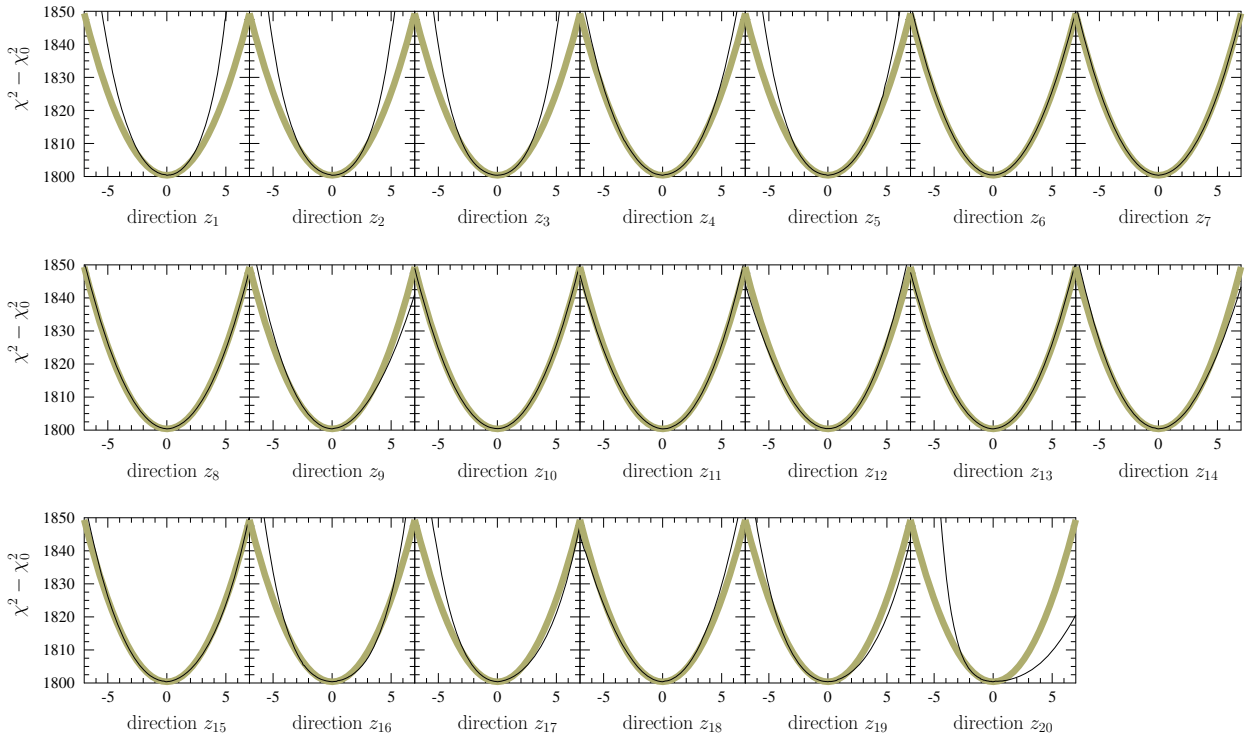

Fig. 4 The χ^2 profiles (*black curves*) as a function of final eigenvector directions z_i compared to ideal behavior $\chi^2_0 + z_i^2$ (*thicker colored curves*)

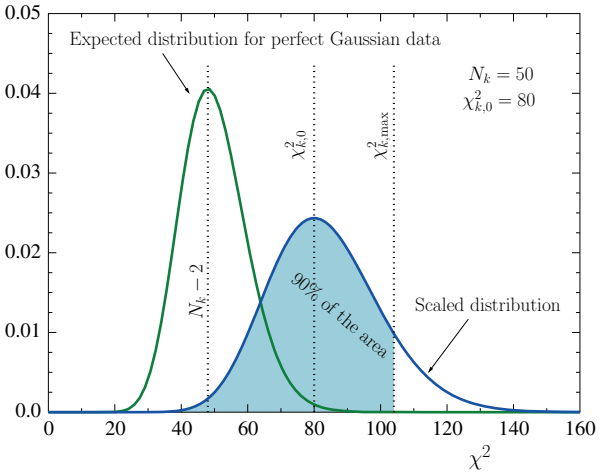

Fig. 5 Determination of 90% confidence limit for an individual data set with $N_k = 50$ data points and for which the global minimum corresponds to $\chi^2_{k,0} = 80$

$$z_{i,\min} \equiv \max\{z^k_{i,\min}\},$$
$$z_{i,\max} \equiv \min\{z^k_{i,\max}\}. \tag{49}$$

The outcome of this process is shown in Figs. 6 and 7 for all eigendirections. The individual limits $[z^k_{i,\min}, z^k_{i,\max}]$ are shown as solid lines (with bars or arrows) and the intersection $[z_{i,\min}, z_{i,\max}]$ as a gray band. This procedure is repeated for all eigendirections i. We note that we have here grouped together all the data (summing the χ^2 contributions) from a given experiment and thus, in Figs. 6 and 7 there are less labels than individual contributions in Table 1. Motivation for such a grouping is that even if an experiment gives data for various nuclei (e.g. SLAC E139) these are not unrelated e.g. for the baseline measurement and detector systematics. Furthermore, it may also happen (e.g. direction 8, lower limit; in Fig. 6) that none of the individual experiments (with grouped data) places stringent uncertainty limits, i.e. the intervals $[z_{i,\min}, z_{i,\max}]$ become rather wide and the total χ^2grows substantially above χ^2_0. In such a case, the data from various experiments together may provide a better constraint than an individual experiment. To take this into account, we treat the aggregate of all the data as a single additional "experiment" (the first rows in the panels of Figs. 6 and 7.

We study two options to define the PDF uncertainty sets $S_i^{\pm}$. In the first one, we set $t_i^+ = z_{i,\max}$ and $t_i^- = -z_{i,\min}$ in Eq. (40), i.e.,

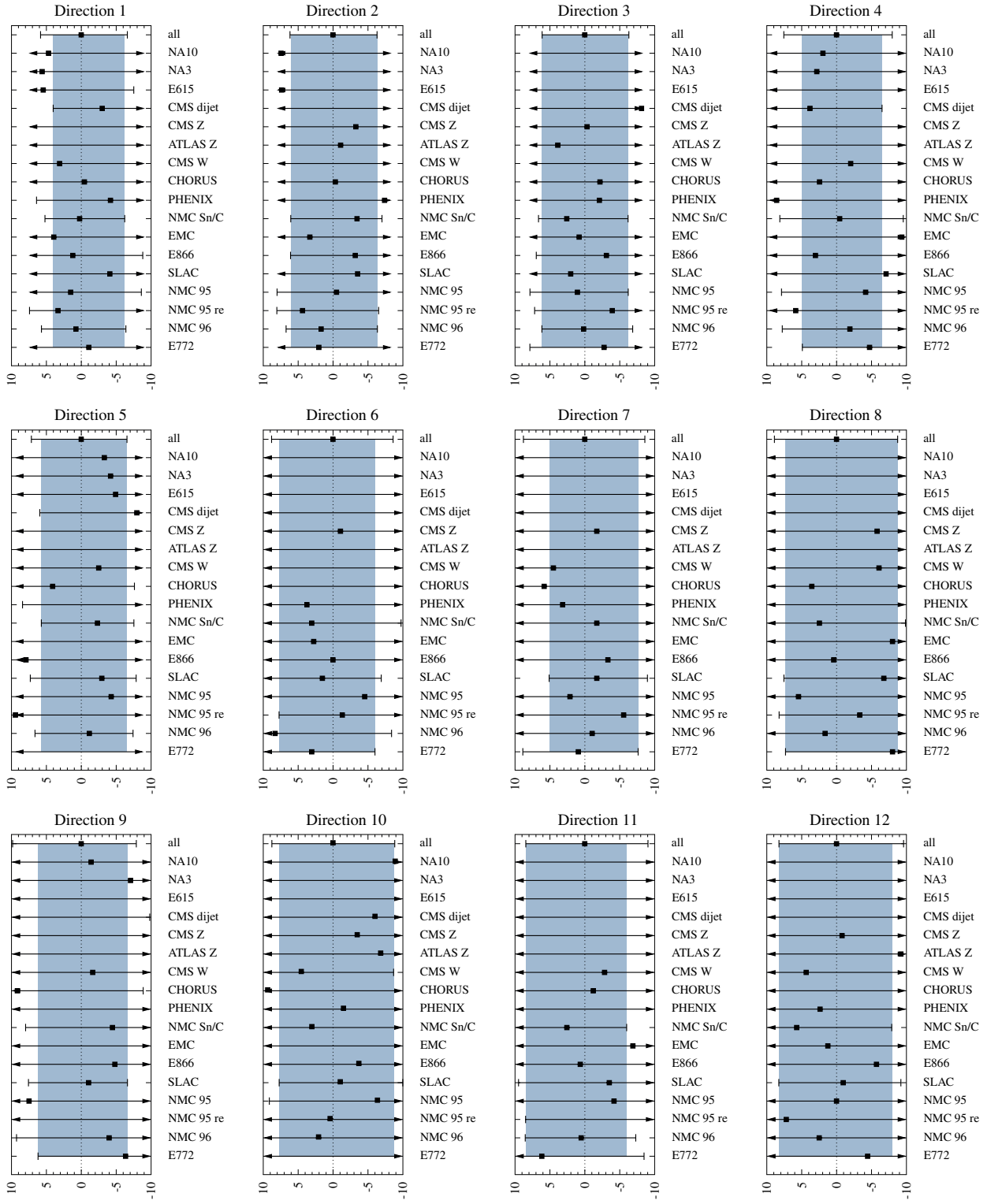

Fig. 6 Determination of the confidence limits for the eigendirections 1 to 12. The *bars* show the limits $z^k_{i,\min}$, $z^k_{i,\max}$ for each individual (or grouped) data set k and the marker in between indicates where the minimum $\chi^2_{k,0}$ of that data set is reached. The set "all" refers to all data combined. An *arrow* signifies that the confidence limit has not yet been reached in the scanned interval. The *gray bands* are the intersection intervals $\left[z_{i,\min}, z_{i,\max}\right]$ explained in the text

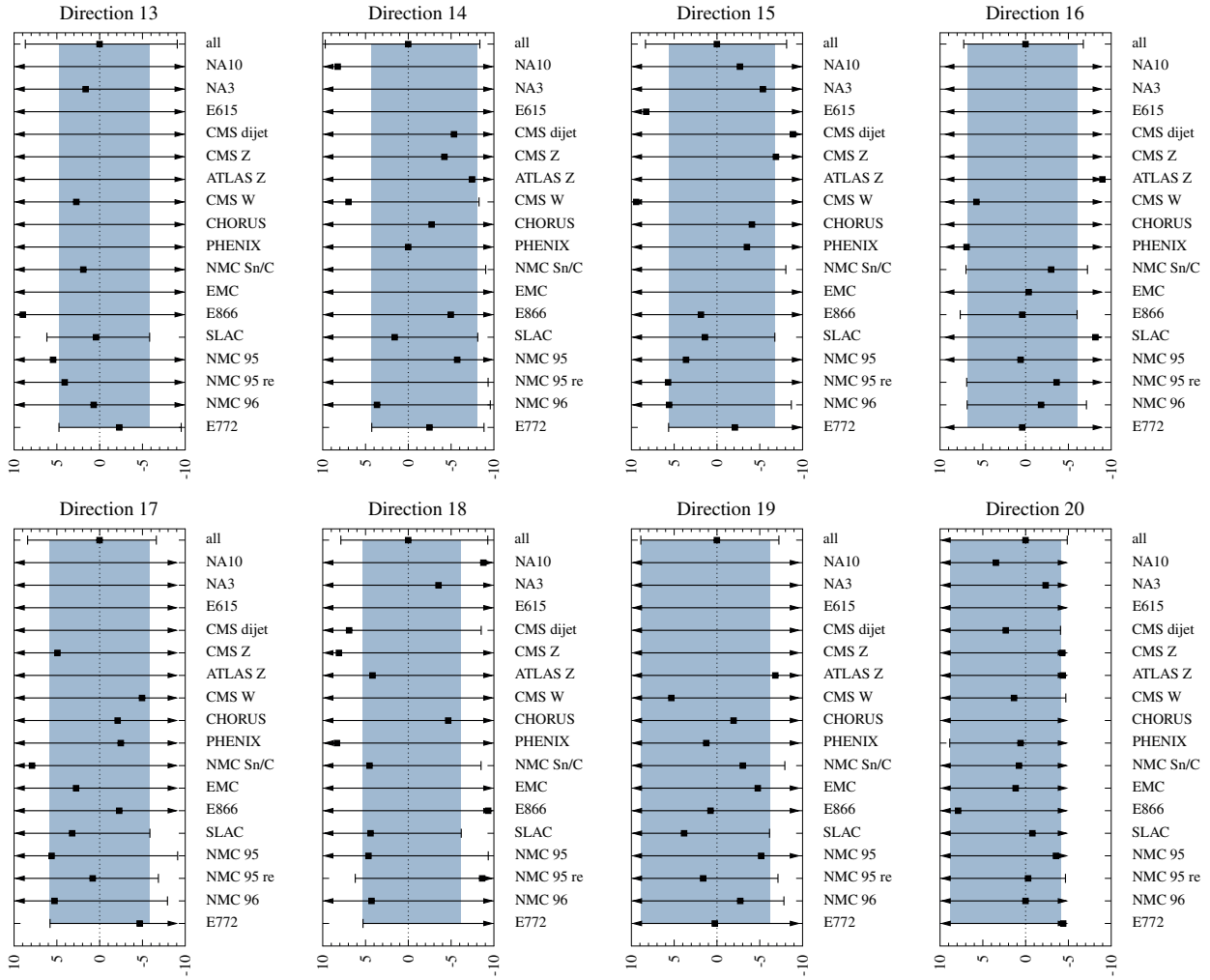

Fig. 7 As Fig. 6 but for eigendirections 13 to 20

$$
\begin{aligned}
\mathbf{z}(S_1^+[\mathrm{dyn}]) &= z_{1,\max}\,(1,0,\ldots,0)\\
\mathbf{z}(S_1^-[\mathrm{dyn}]) &= z_{1,\min}\,(1,0,\ldots,0)\\
&\vdots\\
\mathbf{z}(S_N^+[\mathrm{dyn}]) &= z_{N,\max}\,(0,0,\ldots,1)\\
\mathbf{z}(S_N^-[\mathrm{dyn}]) &= z_{N,\min}\,(0,0,...,1)\,,
\end{aligned}
\tag{50}
$$

where the numbers $z_{i,\min/\max}$ are obtained as described above. This is sometimes referred to as dynamic tolerance determination [99]. For the second option, we specify an average tolerance $\Delta\chi^2$ as

$$
\Delta\chi^2 \equiv \frac{1}{N}\sum_i \frac{\chi^2\left(S_i^-[\mathrm{dyn}]\right)+\chi^2\left(S_i^+[\mathrm{dyn}]\right)-2\chi_0^2}{2}, \tag{51}
$$

where $\chi^2(S_i^\pm[\mathrm{dyn}])$ are the values of χ^2 that correspond to the error sets $S_i^\pm\left[\mathrm{dyn}\right]$ defined above. For the present fit with all the data, we find $\Delta\chi^2 \approx 52$. This averaging process is illustrated in Fig. 8 which shows the individual differences $\chi^2(S_i^-[\mathrm{dyn}]) - \chi_0^2$ and $\chi^2(S_i^+[\mathrm{dyn}]) - \chi_0^2$ as bars together with the found average. In this case the PDF uncertainty sets $S_i^\pm\left[\Delta\chi^2\right]$ are defined by imposing a fixed global tolerance $\Delta\chi^2 = 52$,

$$
\begin{aligned}
\mathbf{z}(S_1^\pm[\Delta\chi^2]) &= \delta z_1^\pm\,(1,0,\ldots,0)\\
&\vdots\\
\mathbf{z}(S_N^\pm[\Delta\chi^2]) &= \delta z_N^\pm\,(0,0,...,1)
\end{aligned}
\tag{52}
$$

where the numbers $\delta z_i^\pm$ are the deviations in positive and negative direction chosen such that the χ^2 grows by 52. The obtained values for $\delta z_i^\pm$ are listed in Table 2.

As expected, Fig. 8 shows rather significant variations in $\chi^2(S_i^\pm[\mathrm{dyn}]) - \chi_0^2$ depending on which eigendirection one looks at. However, the corresponding variations in

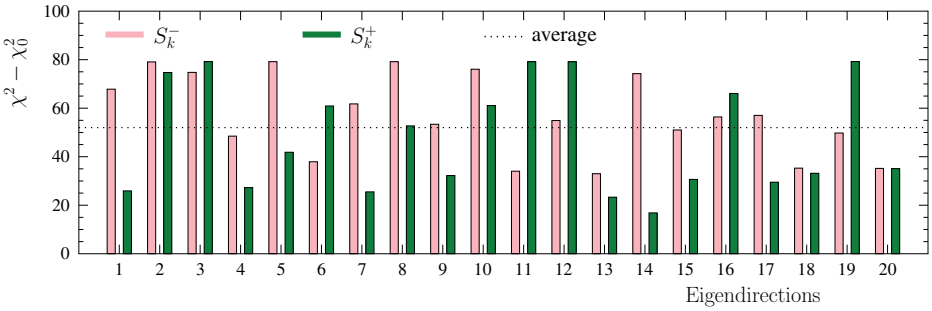

Fig. 8 The individual values of $\chi^2(S_k^\pm) - \chi_0^2$ compared with the average $\Delta\chi^2 = 52$

Table 2 The parameter deviations $\delta z_i^\pm$ defining the EPPS16 error sets in Eq. (52)

δz_i^-	Value	δz_i^+	Value
δz_1^-	−5.620	δz_1^+	5.121
δz_2^-	−5.489	δz_2^+	5.395
δz_3^-	−5.496	δz_3^+	5.344
δz_4^-	−6.705	δz_4^+	6.412
δz_5^-	−5.631	δz_5^+	6.194
δz_6^-	−7.013	δz_6^+	7.148
δz_7^-	−7.021	δz_7^+	7.219
δz_8^-	−7.092	δz_8^+	7.268
δz_9^-	−6.532	δz_9^+	7.935
δz_{10}^-	−7.231	δz_{10}^+	7.133
δz_{11}^-	−7.396	δz_{11}^+	6.968
δz_{12}^-	−7.674	δz_{12}^+	6.814
δz_{13}^-	−7.343	δz_{13}^+	7.065
δz_{14}^-	−6.863	δz_{14}^+	7.749
δz_{15}^-	−6.810	δz_{15}^+	7.080
δz_{16}^-	−5.847	δz_{16}^+	6.327
δz_{17}^-	−5.669	δz_{17}^+	7.238
δz_{18}^-	−7.531	δz_{18}^+	6.510
δz_{19}^-	−6.240	δz_{19}^+	7.576
δz_{20}^-	−4.485	δz_{20}^+	10.53

$z_{i,\mathrm{min/max}} \sim \sqrt{\chi^2(S_i^\pm[\mathrm{dyn}]) - \chi_0^2}$ which determine the error sets are much milder. Hence, it can be expected that the two error-set options, $S_i^\pm[\mathrm{dyn}]$ and $S_i^\pm[\Delta\chi^2]$, will eventually lead to rather similar uncertainty estimates. In the following (see Fig. 11), we will verify that this indeed is the case. Hence, and also to enable PDF reweighting [45], we choose the $S_i^\pm[\Delta\chi^2]$ with the single global tolerance $\Delta\chi^2$ as the final EPPS16 error sets.

As in EPS09, the propagation of PDF uncertainties into an observable $\mathcal{O}$ will be here computed separately for the upward and downward directions,

$$\left(\delta\mathcal{O}^\pm\right)^2 = \sum_i \left[\begin{matrix}\max\\ \min\end{matrix}\{\mathcal{O}(S_i^+) - \mathcal{O}(S_0), \mathcal{O}(S_i^-) - \mathcal{O}(S_0), 0\}\right]^2, \quad (53)$$

where $\mathcal{O}(S_0)$ denotes the prediction with the central set and $\mathcal{O}\left(S_i^\pm\right)$ are the values computed with the error sets [102].

5 Results

5.1 Parametrization and its uncertainties

The parameter values that define the fit functions, the nuclear modifications R_i^A in Eq. (2) at the initial scale Q_0^2 are listed in Table 3 where we also indicate the parameters that were fixed to those of other parton species or assumed to have some particular value. The fixed value of $\beta = 1.3$ for all flavors as well as setting $\gamma_{y_a} = 0$ for sea quarks are motivated

Table 3 List of parameters defining the central set of EPPS16 at the initial scale $Q_0^2 = 1.69\,\mathrm{GeV}^2$. The numbers in bold indicate the 20 parameters that were free in the fit

Parameter	u_V	d_V	$\overline{u}$
$y_0(A_\mathrm{ref})$	Sum Rule	Sum rule	**0.844**
γ_{y_0}	Sum Rule	Sum rule	**0.731**
x_a	**0.0717**	As u_V	**0.104**
x_e	**0.693**	As u_V	As u_V
$y_a(A_\mathrm{ref})$	**1.06**	**1.05**	**1.03**
γ_{y_a}	**0.278**	As u_V	0, fixed
$y_e(A_\mathrm{ref})$	**0.908**	**0.943**	**0.725**
γ_{y_e}	**0.288**	As u_V	As u_V
β	1.3, fixed	1.3, fixed	1.3, fixed
Parameter	$\overline{d}$	s	g
$y_0(A_\mathrm{ref})$	**0.889**	**0.723**	Sum rule
γ_{y_0}	as $\overline{u}$	as $\overline{u}$	Sum rule
x_a	As $\overline{u}$	As $\overline{u}$	**0.0820**
x_e	As u_V	As u_V	As u_V
$y_a(A_\mathrm{ref})$	**0.919**	**1.24**	**1.12**
γ_{y_a}	0, fixed	0, fixed	As u_V
$y_e(A_\mathrm{ref})$	As $\overline{u}$	As $\overline{u}$	**0.874**
γ_{y_e}	As u_V	As u_V	As u_V
β	1.3, fixed	1.3, fixed	1.3, fixed

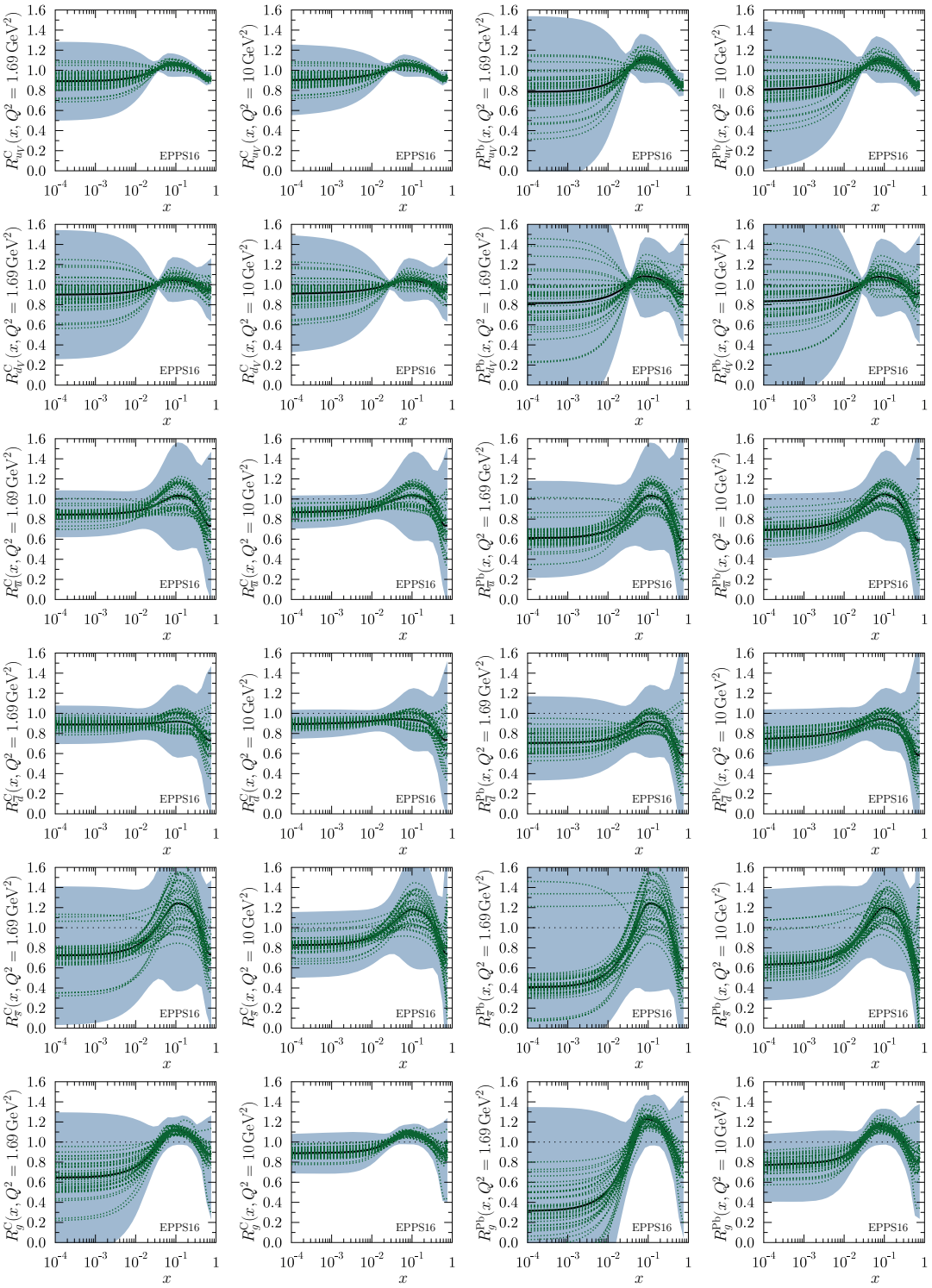

Fig. 9 The EPPS16 nuclear modifications for carbon (*leftmost columns*) and lead (*rightmost columns*) at the parametrization scale $Q^2 = 1.69\,\mathrm{GeV}^2$ and at $Q^2 = 10\,\mathrm{GeV}^2$. The *thick black curves* correspond to the central fit S_0 and the *dotted curves* to the individual error sets $S_i^{\pm}\left[\Delta\chi^2\right]$ of Eq. (52). The total uncertainties are shown as *blue bands*

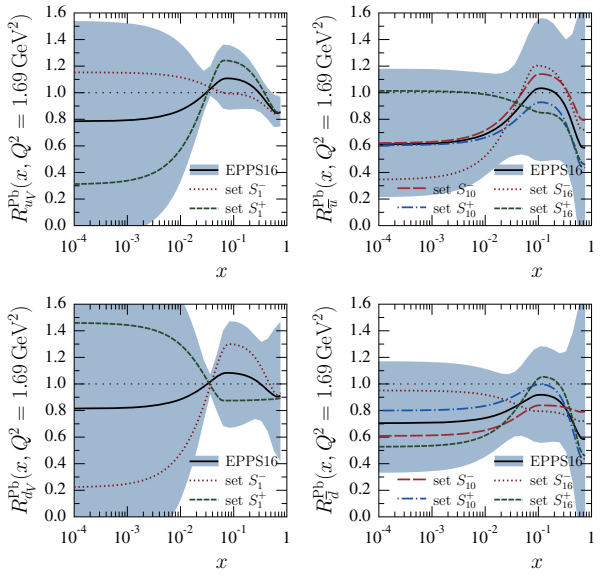

Fig. 10 The EPPS16 nuclear modifications for valence and sea u & d quarks for lead at the parametrization scale $Q^2 = 1.69\,\mathrm{GeV}^2$. The *solid black curves* correspond to the central result and the *dotted/dashed curves* to the specific error sets as indicated. The total uncertainties are shown as *blue bands*

by the EPS09 analysis. Freeing the latter easily leads to an unphysical case ($\gamma_{y_a} < 0$) and thus we have decided to keep it fixed at this stage.

The R_i^A functions themselves with error sets of Eq. (52) and uncertainty bands of Eq. (53) are plotted in Fig. 9 for carbon and lead nuclei at $Q^2 = Q_0^2$ and $Q^2 = 10\,\mathrm{GeV}^2$. Regarding these results, we make the following observations:

First, the obtained valence modifications $R_{u_\mathrm{V}}^A$ and $R_{d_\mathrm{V}}^A$ are very similar in the central set S_0, and strongly anticorrelated: as the average valence modification is fairly well constrained (see Fig. 27) an error set whose, say, $R_{u_\mathrm{V}}^A$ is clearly below the central value has to have an $R_{d_\mathrm{V}^A}$ which is correspondingly above the central value, and vice versa. This is further demonstrated in Fig. 10 where only the errors sets $S_1^\pm$ are shown for valence. The large error bands for $R_{u_\mathrm{V}}^A$ and $R_{d_\mathrm{V}}^A$ at small x in Fig. 9 reflect the fact that the flavor separation is not stringently constrained in the antishadowing region: the finite uncertainties there induce (via the sum rules) larger uncertainties in the shadowing region; see Fig. 10.

Second, interestingly also the u and d sea-quark modifications are very similar in the central set S_0, and anticorrelated (except in the large-x region where they were assumed to be the same at Q_0^2), though not as strongly as the valence quarks because also the strange-quark distribution plays some role. An example is shown in Fig. 10 where the errors sets $S_{10}^\pm$ and $S_{16}^\pm$ have been plotted. In contrast to the valence quarks, individual sets are not always anticorrelated throughout all the x values, but sets that are anticorrelated e.g. near x_a can be very similar toward $x \to 0$.

Third, the central value of the strange-quark nuclear modification indicates stronger nuclear effects than for the other light sea quarks. On the other hand, the uncertainty is also significant and even a large enhancement at small x appears possible. While such an effect is theoretically unlikely (we would expect shadowing), it is consistent with the utilized data whose uncertainties our uncertainty bands represent. It should also be borne in mind that the determination of the strange quark in CT14 (our baseline PDF) may suffer from uncertainties (e.g. related to treatment of dimuon process in neutrino–nucleus DIS) and can, to some extent, affect the nuclear modifications we obtain. Thus, building a "hard wall" e.g. prohibiting an enhancement at small x is not justified either. Nevertheless, the found central values of the strange-quark nuclear modifications are clearly in a sensible ballpark.

Fourth, for gluon distributions the uncertainties are large at small x at Q_0^2 but quickly diminish as the scale is increased. The gluon distributions in some error sets also go negative at small x at low Q^2 but since F_L remains positive, this is allowed.

Fifth, on average, the nuclear effects of lead tend to be stronger than those of carbon and also the uncertainties on lead are larger than those on carbon. Given that most of the data are for heavier nuclei than carbon, especially the smaller errors for carbon may appear a bit puzzling. The reason is in the new way of parametrizing the A dependence of the nuclear effects, see Eq. (3), that favours larger nuclei to exhibit larger nuclear effects.

Sixth, the parametrization bias that our fit function entails is particularly well visible in the valence-quark panels where a narrow "throat" at $x \approx 0.02$ can be seen. This is an artifact of not allowing for more freedom at small x while requiring the sum rules in Eq. (4) and Eq. (5): to satisfy the sum rule, an enhancement around $x = 0.1$ must be accompanied by a depletion at small x (or vice versa), and since x_a for valence is fairly well determined the fit function always crosses unity near $x \approx 0.02$.

In Sect. 4.1 we mentioned that the two error-determination options, the dynamical tolerance and fixed global tolerance, lead to similar uncertainty estimates. To demonstrate this, we plot in Fig. 11 the error bands of the nuclear effects R_i^Pb at $Q^2 = 10\ \mathrm{GeV}^2$ obtained correspondingly from the error sets $S_i^\pm[\mathrm{dyn}]$ and $S_i^\pm[\Delta\chi^2]$. Indeed, we find no significant differences between the two options.

5.2 Comparison with data

Figures 12, 13, 14, 15, 16, 17, 18, 19, 20 and 21 present a comparison of the EPPS16 fit with the experimental data of Table 1, computing the PDF error propagation accord-

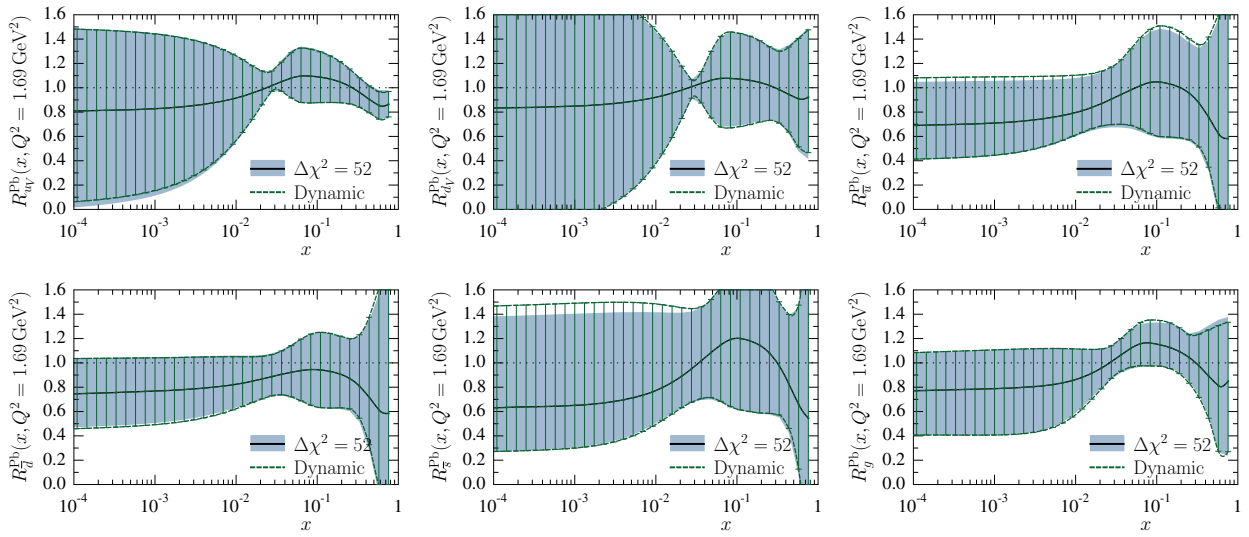

Fig. 11 The *error bands* of nuclear modifications at $Q^2 = 10\,\mathrm{GeV}^2$ from the global tolerance $\Delta\chi^2 = 52$ used in the final EPPS16 fit (*black central line* and *light-blue bands*) compared to the *error bands* from the dynamical tolerance determination (*hatching*) explained in Sect. 4.1

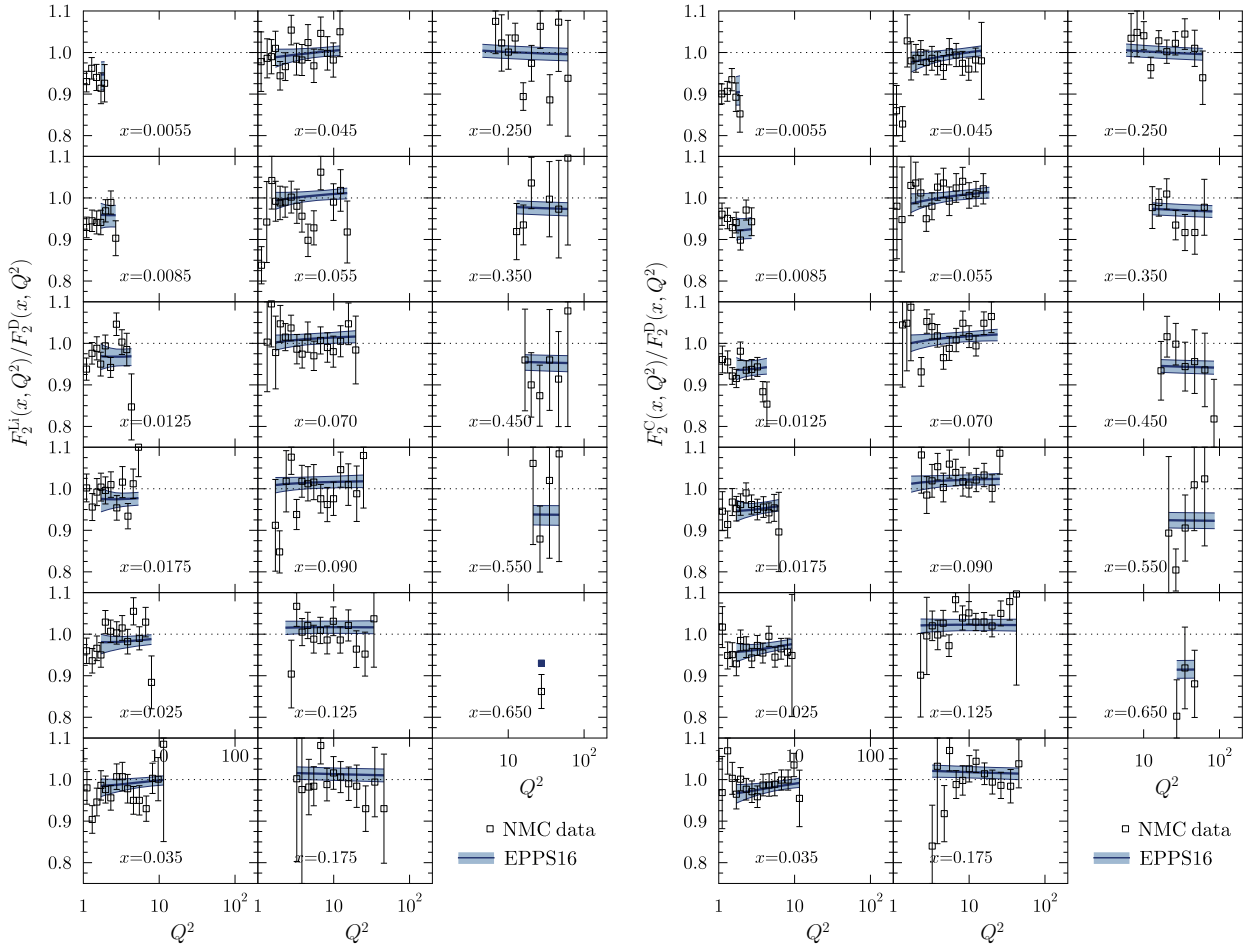

Fig. 12 The Q^2 dependence of structure-function ratios as measured by the NMC collaboration [74], compared with the EPPS16 fit. *Solid lines* show our central set results, and *error bands* are computed from Eq. (53)

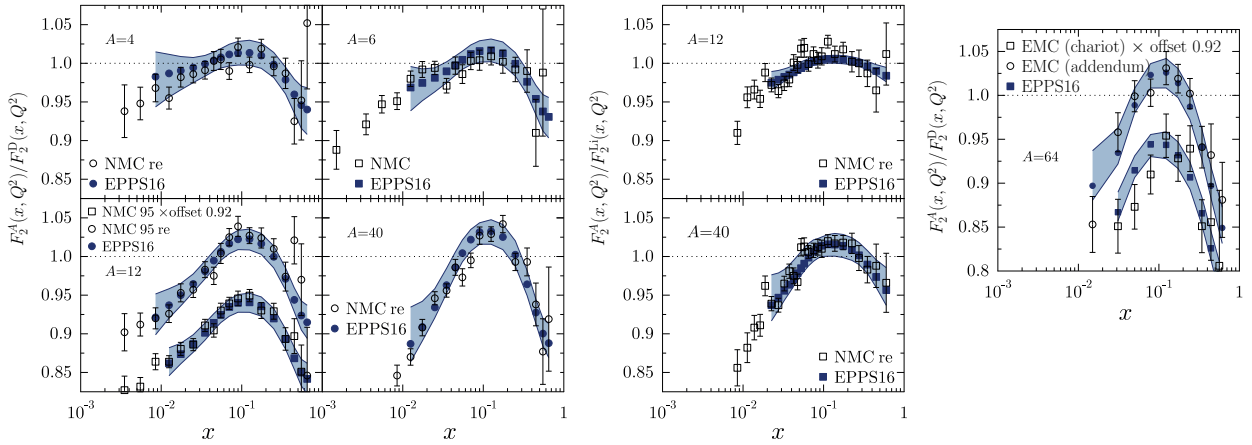

Fig. 13 Ratios of structure functions for various nuclei as measured by the NMC [73,74] and EMC [78] collaborations, compared with the EPPS16 fit. In the *rightmost panel* the labels "addendum" and "chariot" refer to the two different experimental setups in Ref. [78]. For a better visibility, some data sets have been offset by a factor of 0.92 as indicated

ing to Eq. (53). The error bars shown on the experimental data correspond to the statistical and systematic errors added in quadrature. The charged-lepton DIS data are shown in Figs. 12, 13, 14 and 15. We note that, for undoing the isoscalar corrections as explained in Sect. 3.1, the data appear somewhat different from those e.g. in the EPS09 paper. On average, the data are well reproduced by the fit. In some cases the uncertainty bands are rather asymmetric (see e.g. the NMC data panel in Fig. 15) which was the case in the EPS09 fit as well. This is likely to come from the fact that the A dependence is parametrized only at few values of x (small-x limit, x_a, x_e) and in between these points the A dependence appears to be somewhat lopsided in some cases. The Q^2 dependence of the data visible in Figs. 12 and 14 is also nicely consistent with EPPS16.

The pA vs. pD Drell–Yan data are shown in Figs. 16 and 17. In the calculation of the corresponding differential NLO cross sections $\mathrm{d}\sigma^{\mathrm{DY}}/\mathrm{d}x\mathrm{d}M$ we define $x_{1,2} \equiv (M/\sqrt{s})e^{\pm y}$ where M is the invariant mass and y the rapidity of the dilepton. The scale choice in the PDFs is $Q = M$. While these data are well reproduced, the scatter of the data from one nucleus to another is the main reason we are unable to pin down any systematic A dependence for the sea quarks at x_a (some A dependence develops via DGLAP evolution, however). For example, as is well visible in Fig. 17, it is not clear from the data whether there is a suppression or an enhancement for $x \gtrsim 0.1$.

The pion–A DY data are presented in Fig. 18. As is evident from the figure, these data set into the EPPS16 fit without causing a significant tension. Overall, however, the statistical weight of these data is not enough to set stringent additional constraints to nuclear PDFs. Similarly to the findings of Ref. [67], the optimal data normalization of the lower-energy NA10 data (the lower right panel) is rather large ($f_N = 1.121$), but the x_2 dependence of the data is well in line with the fit.

The collider data, i.e. new LHC pPb data as well as the PHENIX DAu data, are shown in Fig. 19. To ease the interpretation of the LHC data (forward-to-backward ratios), the baseline with no nuclear effects in PDFs is always indicated as well. The baseline deviates from unity for isospin effects (unequal amount of protons and neutrons in Pb) as well as for experimental acceptances. For the electroweak observables, the nuclear effects cause suppression in the computed forward-to-backward ratios (with respect to the baseline with no nuclear effects) as one is predominantly probing the region below $x \sim 0.1$ where the net nuclear effect of sea quarks has a downward slope toward small x. Very roughly, the probed nuclear x-regions can be estimated by $x \approx (M_{\mathrm{W,Z}}/\sqrt{s})e^{-y}$ and thus, toward more forward rapidities ($y > 0$) one probes smaller x than in the backward direction ($y < 0$). The suppression comes about as smaller-x quark distributions are divided by larger-x (less-shadowed or antishadowed) quarks. In the case of dijets, the nuclear PDFs are sampled at higher x and, in contrast to the electroweak bosons, an enhancement is observed. In our calculations, this follows essentially from antishadowed gluons becoming divided by EMC-suppressed gluon distributions; see Ref. [70] for more detailed discussions. The PHENIX pion data [31] is also well consistent with EPPS16, though, for the more precise CMS dijet data, its role is no longer as essential as in the EPS09 analysis.

Finally, comparisons with the CHORUS neutrino and antineutrino data are shown in Figs. 20 and 21. The data exhibit a rather typical pattern of antishadowing followed by an EMC effect at large x. The incident beam energies are not high enough to reach the small-x region where a shadowing

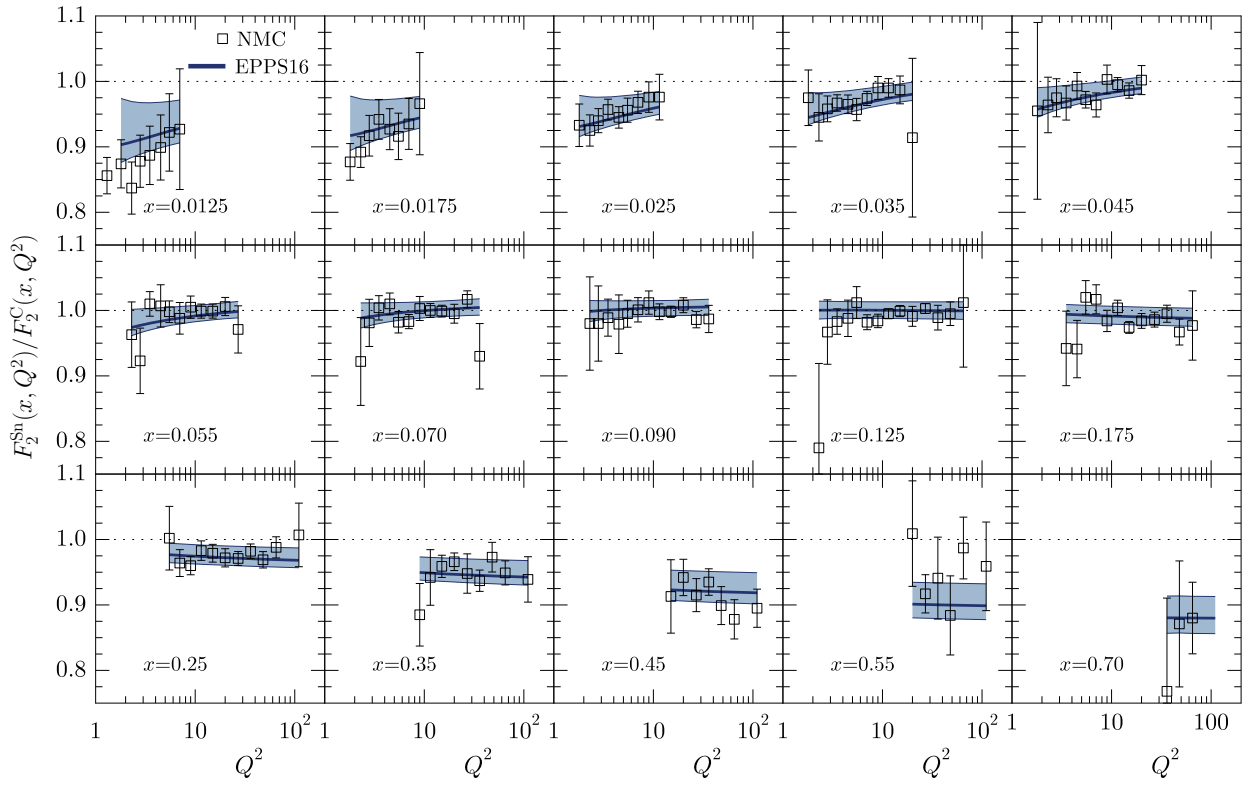

Fig. 14 The Q^2 dependence of the ratio $F_2^{\mathrm{Sn}}/F_2^{\mathrm{C}}$ for various values of x as measured by NMC [79], compared with EPPS16

effect would be expected. Toward small x, however, the data do appear to show a slight downward bend, a possible onset of shadowing.

5.3 Comparison with baseline

To appreciate the effects induced by the new data (pion–A DY, neutrino DIS and LHC data) in the EPPS16 fit, we have performed another fit excluding these data sets but still correcting the DIS data for the isospin effects. This fit is referred to as "baseline" in the following. The resulting nuclear modifications for Pb at $Q^2 = 10\,\mathrm{GeV}^2$ with a comparison to the EPPS16 results are shown in Fig. 22. For the baseline fit here, the global tolerance is $\Delta\chi^2_{\mathrm{baseline}} = 35$. As seen in the figure, it is not always the case that the uncertainties of EPPS16 would be smaller than those of the baseline. This originates from the mutually different global tolerances of the two fits and from the differences of the χ^2 behavior around the minima. In any case, the uncertainty bands always overlap and both of these enclose the central values both from the baseline fit and the full analysis. Thus, the two are consistent. Qualitatively, the most notable changes are that, in comparison to the baseline, the EPPS16 central values of both valence-quark flavors as well as that of gluons exhibit a very similar antishadowing effect followed by an EMC pit. We have observed that this difference is mostly caused by the addition of neutrino DIS data (valence quarks) and the CMS dijet data (gluons). This is also illustrated in Fig. 23 where the left-hand panel shows the χ^2 contribution of the CHORUS data as a function of $y_a^{u_V} - y_a^{d_V}$ (the antishadowing peak heights for A_{ref} as in Table 3) and the right-hand panel the χ^2 contribution of the CMS dijet data as a function of $y_a^g - y_e^g$. The individual points correspond to the EPPS16 and baseline-fit error sets. From these panels we learn that in order to optimally reproduce the CHORUS data we need $y_a^{u_V} \sim y_a^{d_V}$, and agreement with the CMS dijet data requires $y_a^g > y_e^g$ (EMC effect). The other new data (pion–A DY, LHC electroweak data) do not generate such a strong pull away from the central set of the baseline fit. Also the PHENIX data prefers a solution with a gluon EMC effect, but the contribution of these data in the total χ^2 budget is so small that such a tendency is practically lost in the noise (in the EPS09 analysis this was compensated by giving these data an additional weight). The inclusion of the dijet data has also decreased the gluon uncertainties at large x, excluding the solutions with no antishadowing. In the case of u and d sea quarks there are no significant differences between the baseline fit and EPPS16. It appears that the s-quark uncertainty at small x has somewhat reduced by the inclusion of the new data, but the uncertainty is in any case large.

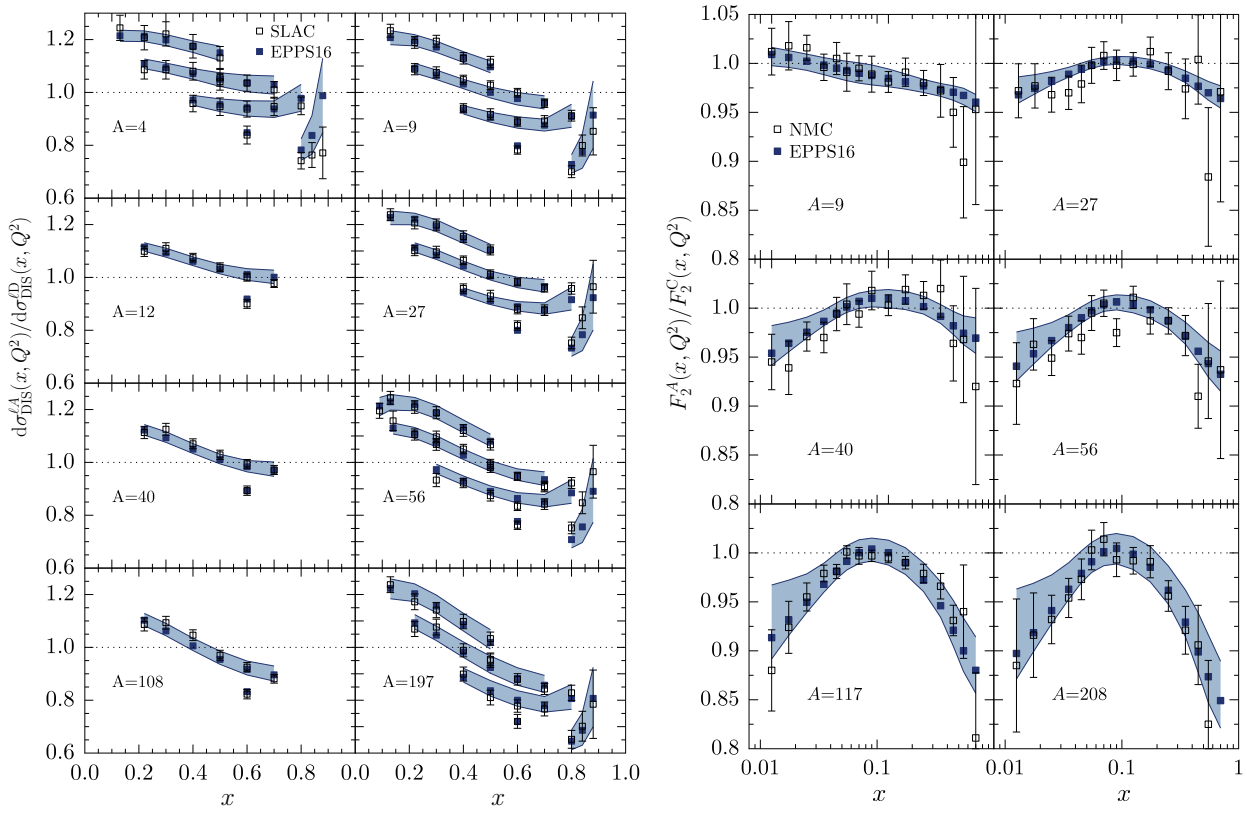

Fig. 15 The SLAC [72] and NMC [75] data for DIS cross-section and structure-function ratios compared with the EPPS16 fit. For a better visibility, the SLAC data have been multiplied by 1.2, 1.1, 1.0, 0.9 for $Q^2 = 2\,\mathrm{GeV}^2$, $Q^2 = 5\,\mathrm{GeV}^2$, $Q^2 = 10\,\mathrm{GeV}^2$, $Q^2 = 15\,\mathrm{GeV}^2$, and the largest-x set by 0.8

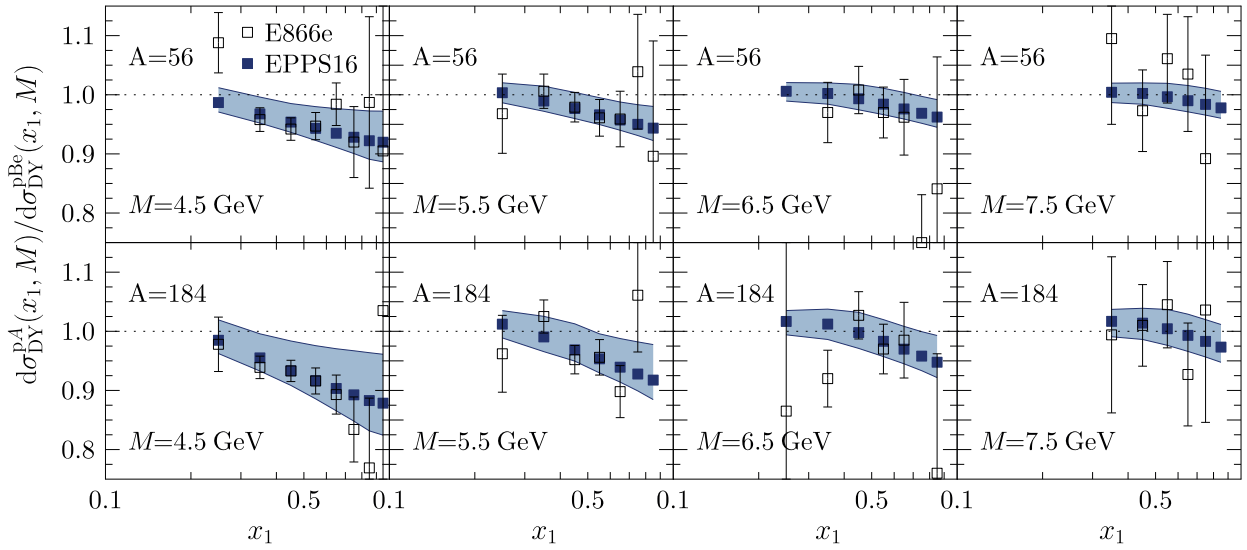

Fig. 16 Ratios of Drell–Yan dilepton cross sections $d\sigma^{pA}/d\sigma^{p\mathrm{Be}}$ as a function of x_1 at various values of fixed M as measured by E866 [77], compared with EPPS16

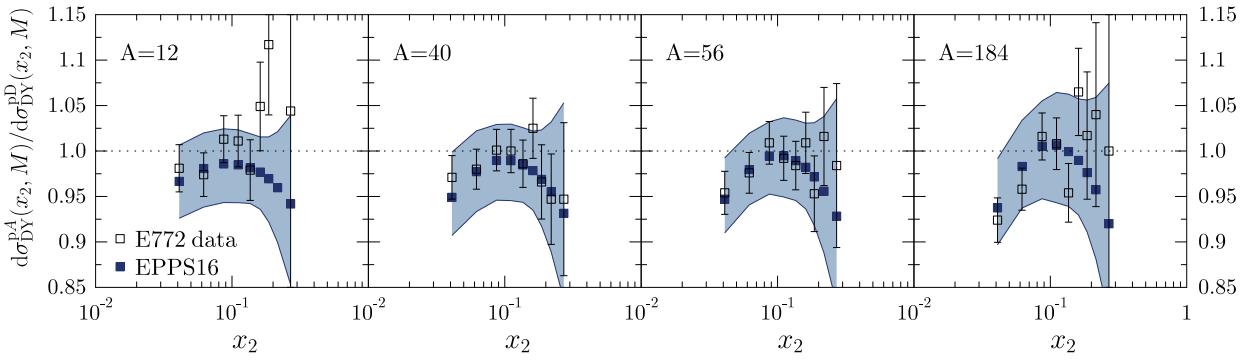

Fig. 17 Ratios of Drell–Yan cross sections measured by E772 as a function of x_2 at fixed values of M, compared with the EPPS16 fit

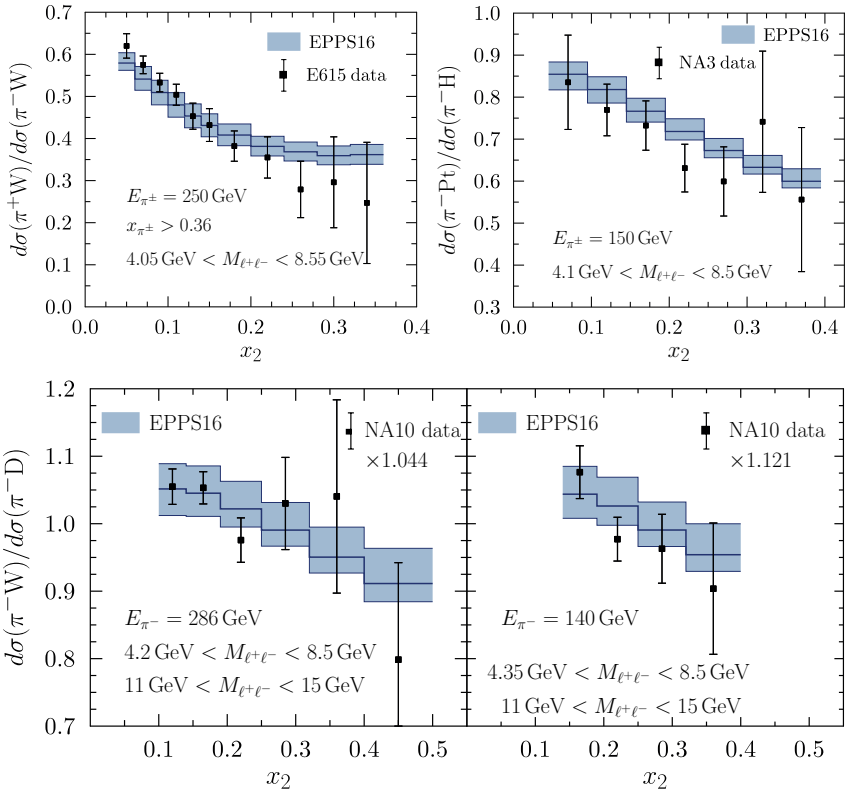

Fig. 18 The $\pi^{\pm}$–A Drell–Yan data from E615 [53], NA3 [51] and NA10 [52], compared with the EPPS16 fit. The NA10 data have been multiplied by the optimized normalization factor f_N from Eq. (31)

The values of χ^2/N_{data} for individual data sets are shown in Fig. 24. For the CMS dijet data the baseline fit gives a very large value but this disagreement disappears when these data are included in the fit. However, upon including the new data no obvious conflicts with the other data sets show up and thus the new data appear consistent with the old. While it is true that on average χ^2/N_{data} for the old data grows when including the new data (and this is mathematically inevitable) no disagreement ($\chi^2/N_{\mathrm{data}} \gg 1$) occurs. For the NMC Ca/D data χ^2/N_{data} is somewhat large but, as can be clearly seen from Fig. 13, there appears to be large fluctuations in the data (see the two data points below the EPPS16 error band). While the improvement in χ^2/N_{data} for the CHORUS data looks smallish in Fig. 24, for the large amount of data points (824) the absolute decrease in χ^2 amounts to 106 units and is therefore significant.

5.4 Comparison with other nuclear PDFs

In Fig. 25 we compare our EPPS16 results at the scale $Q^2 = 10\,\mathrm{GeV}^2$ with those of the nCTEQ15 analysis [35]. The nCTEQ15 uncertainties are defined by a fixed tolerance $\Delta\chi^2 = 35$, which is similar to our average value $\Delta\chi^2 = 52$ and in this sense one would expect uncertainty bands of comparable size. The quark PDFs were allowed to be partly flavor dependent in the nCTEQ15 analysis (although to a much

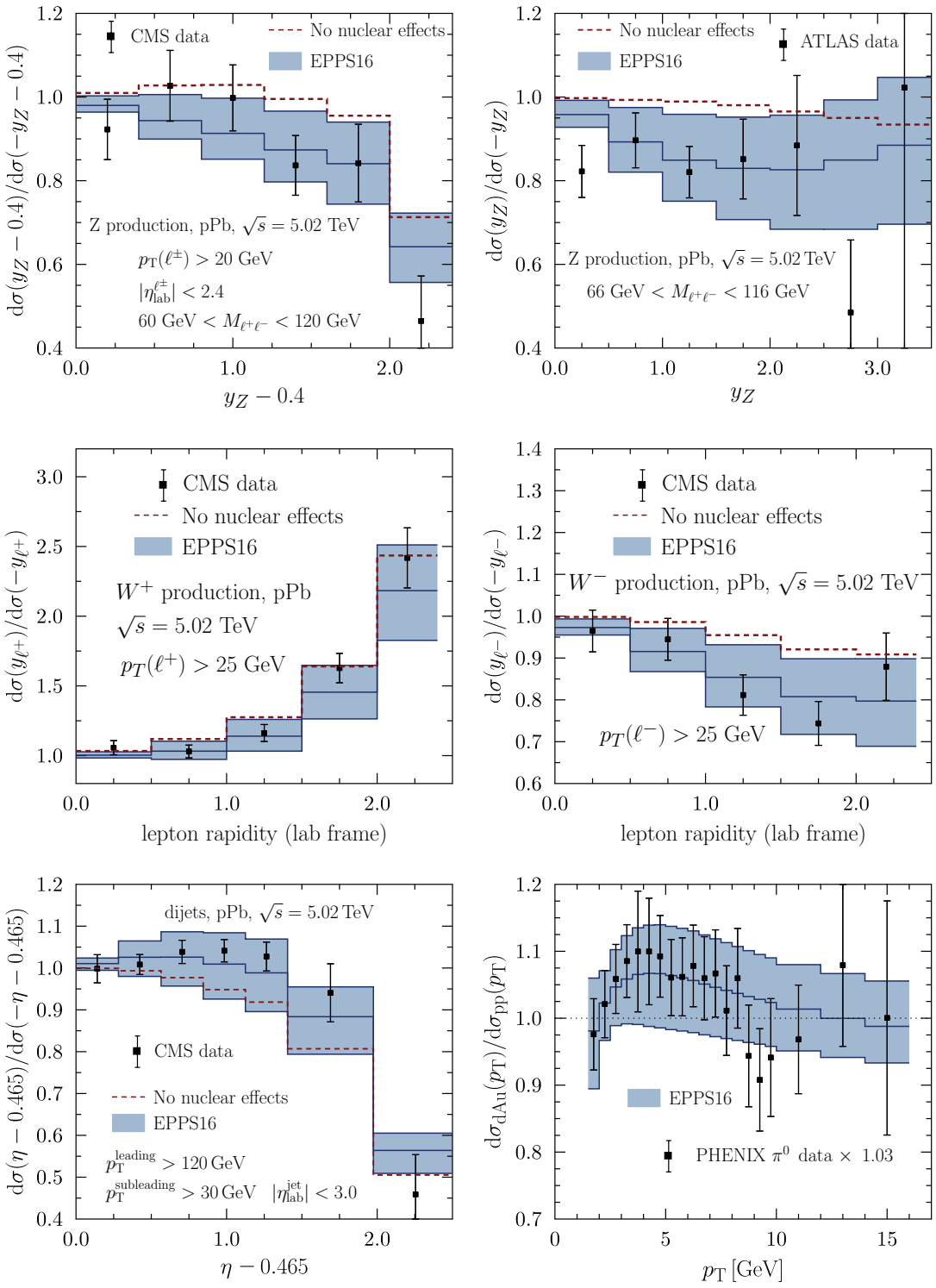

Fig. 19 The LHC pPb data from CMS [37,46,48] and ATLAS [49] for Z (*upper panels*) $W^\pm$ (*middle panels*), and dijet production (*lower left panel*) compared with the EPPS16 fit. The *dashed lines* indicate the results with no nuclear modifications in the PDFs. The PHENIX DAu data [31] for inclusive pion production (*lower right panel*) are shown as well and have been multiplied by the optimal normalization factor $f_N = 1.03$ computed by Eq. (31)

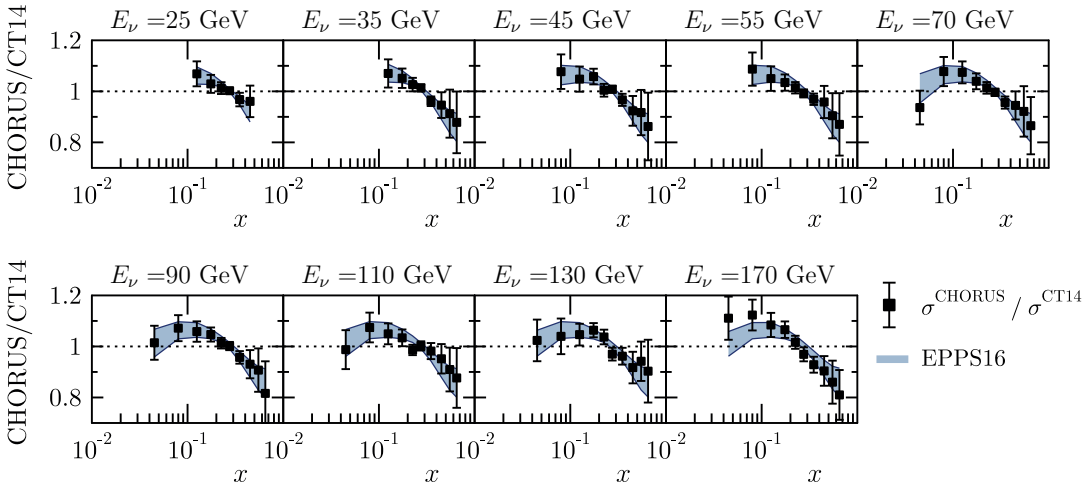

Fig. 20 The neutrino–nucleus DIS data based on CHORUS [50] measurements, compared with the EPPS16 fit. The data as well as the theory curves have been obtained as described in Sect. 3.2

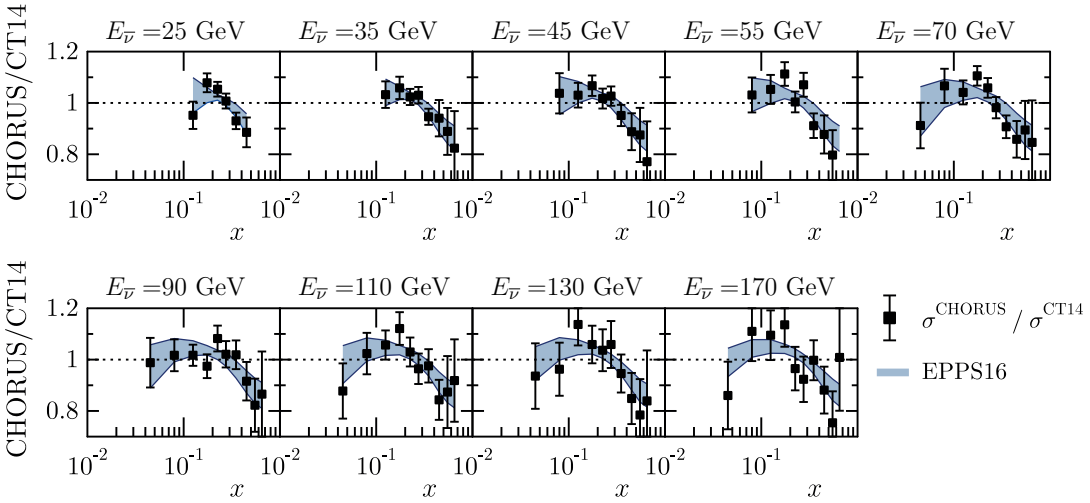

Fig. 21 As Fig. 20 but for antineutrino beam

lesser extent than in EPPS16), hence we show the comparison for all parametrized parton species. The two fits (as well as nCTEQ15 and our baseline fit in Fig. 22) can be considered compatible since the uncertainty bands always overlap. For all the sea quarks the nCTEQ15 uncertainties appear clearly smaller than those of EPPS16 though less data was used in nCTEQ15. This follows from the more restrictive assumptions made in the nCTEQ15 analysis regarding the sea-quark fit functions: nCTEQ15 has only 2 free parameters for all sea quarks together, while EPSS16 has 9. Specifically, the nCTEQ15 analysis constrains only the sum of nuclear $\bar{u}+\bar{d}$ with an assumption that the nuclear s quarks are obtained from $\bar{u}+\bar{d}$ in a fixed way. In contrast, EPPS16 has freedom for all sea-quark flavors separately, and hence also larger, but less biased, error bars. For the valence quarks, the nCTEQ15 uncertainties are somewhat larger than the EPPS16 errors around the x-region of the EMC effect which is most likely related to the extra constraints the EPPS16 analysis has obtained from the neutrino DIS data. Especially the central value for d_V is rather different from that of EPPS16. The very small nCTEQ15 uncertainty at $x \sim 0.1$ is presumably a similar fit-function artifact as what we have for EPPS16 at slightly smaller x. Such a small uncertainty is supposedly also the reason why nCTEQ15 arrives at smaller uncertainties in the shadowing region than EPPS16. For the gluons the nCTEQ15 uncertainties are clearly larger than those of EPPS16, except in the small-x region. While, in part, the larger uncertainties are related to the LHC dijet data that are included in EPPS16 but not in nCTEQ15, this is not the complete explanation as around $x \sim 0.1$ the nCTEQ15

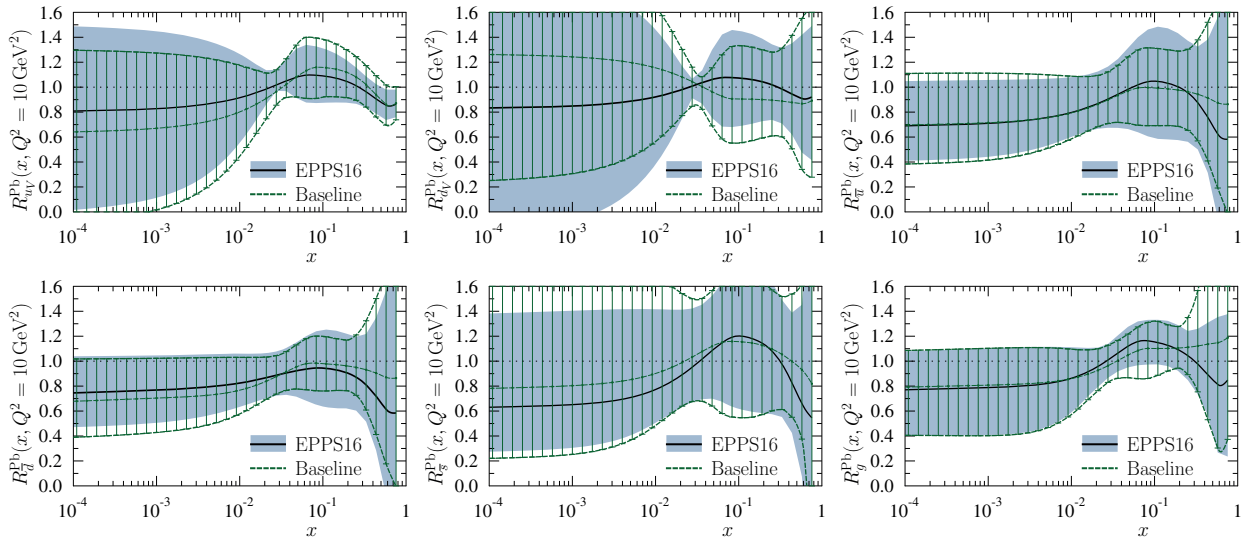

Fig. 22 The nuclear modifications at $Q^2 = 10\,\mathrm{GeV}^2$ from the EPPS16 fit (*black central line* and *light-blue bands*) compared with the baseline fit (*green curves with hatching*) which uses only the data included in the EPS09 fit

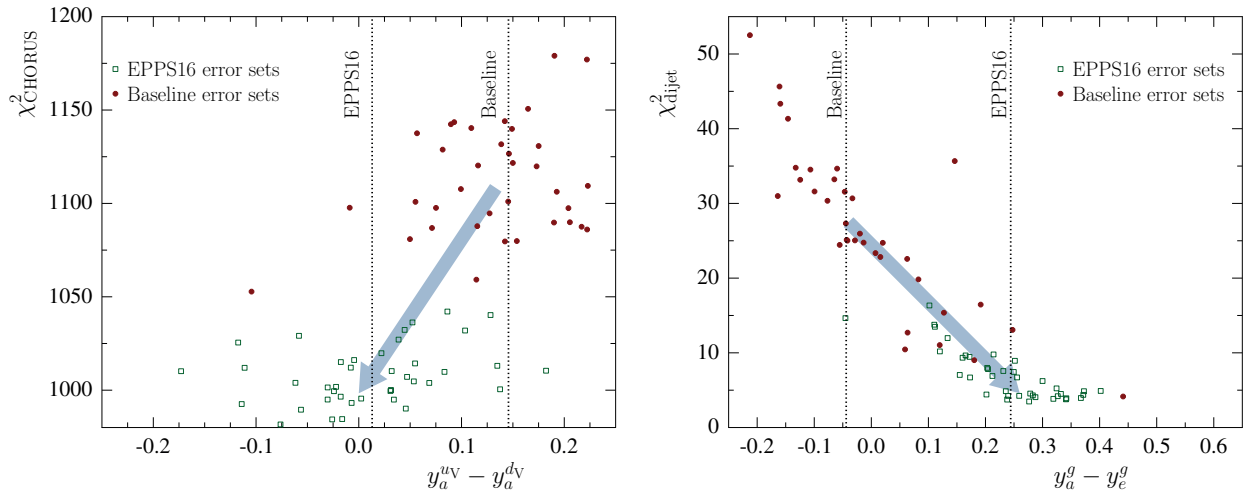

Fig. 23 The contribution of the CHORUS data [50] to the total χ^2 as a function of $y_a^{u_V} - y_a^{d_V}$ (*left*) and the contribution of the CMS dijet data [37] to the total χ^2 as a function of $y_a^g - y_e^g$ (*right*). The *green squares* correspond to the EPPS16 error sets and the *red circles* to the error sets from the baseline fit. The *arrow* indicates the direction of change induced by inclusion of these data into the analysis

uncertainties also largely exceed the uncertainties from our baseline fit (see Fig. 22). Since the data constraints for gluons in both analyses are essentially the same, the reason must lie in the more stringent Q^2 cut ($Q^2 > 4\,\mathrm{GeV}^2$) used in the nCTEQ15 analysis, which cuts out low-Q^2 data points where the indirect effects of gluon distributions via parton evolution are the strongest. The inclusion of the dijet data into the nCTEQ15 analysis would clearly have a dramatic impact. This can be understood from Fig. 26 where we compare the CMS dijet data also with the nCTEQ15 prediction (here, we have formed the nCTEQ15 nuclear modifications from their absolute distributions and used the same dijet grid as in the EPPS16 analysis).

A comparison of EPPS16 with EPS09 [33] and DSSZ [34] is presented in Fig. 27. In the EPS09 and DSSZ analyses the nuclear modifications of valence and sea quarks were flavor independent at the parametrization scale and, to make a fair comparison we plot, in addition to the gluons, the average nuclear modifications for the valence quarks and light sea quarks,

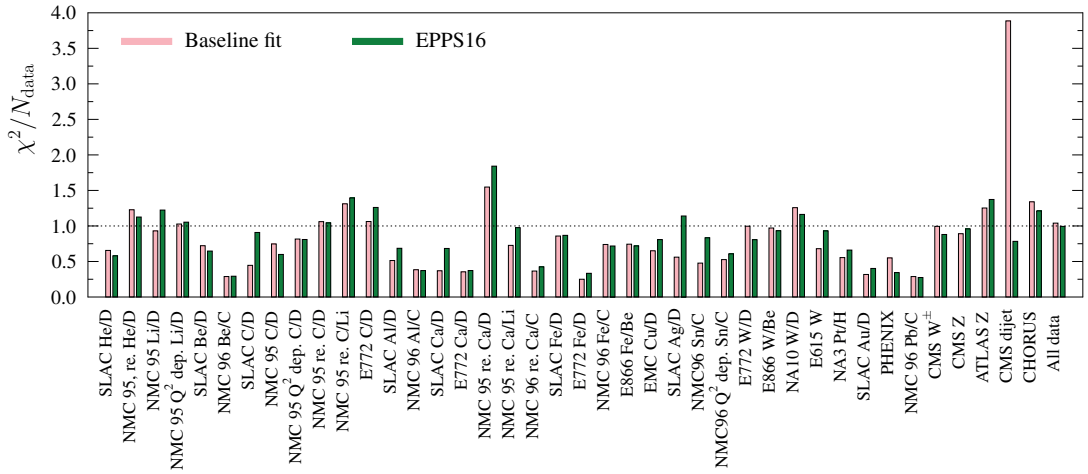

Fig. 24 The values of $\chi^2/N_{\rm data}$ from the baseline fit (*red bars*) and EPPS16 (*green bars*) for data in Table 1

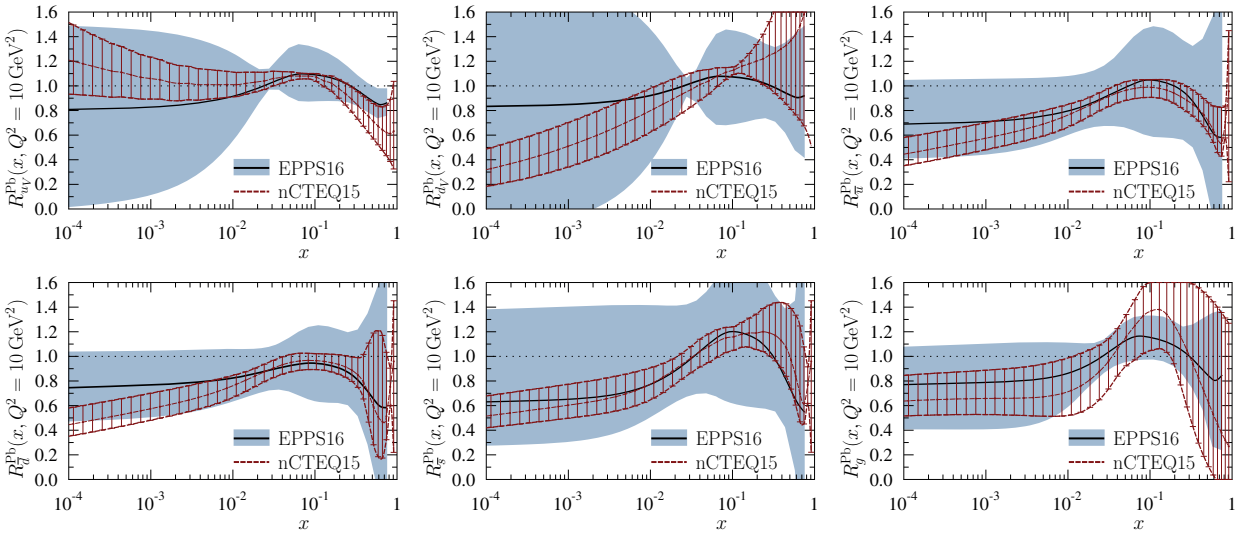

Fig. 25 Comparison of the EPPS16 nuclear modifications (*black central curve with shaded uncertainty bands*) with those from the nCTEC15 analysis [35] (*red curves with hatching*) at $Q^2 = 10\,\mathrm{GeV}^2$

$$R_{\rm V}^{\rm Pb} \equiv \frac{u_{\rm V}^{\rm p/Pb} + d_{\rm V}^{\rm p/Pb}}{u_{\rm V}^{\rm p} + d_{\rm V}^{\rm p}}, \tag{54}$$

$$R_{\rm S}^{\rm Pb} \equiv \frac{\overline{u}^{\rm p/Pb} + \overline{d}^{\rm p/Pb} + \overline{s}^{\rm p/Pb}}{\overline{u}^{\rm p} + \overline{d}^{\rm p} + \overline{s}^{\rm p}}, \tag{55}$$

instead of individual flavors. For the valence sector, all parametrizations give very similar results except DSSZ in the EMC-effect region. As noted earlier in Sect. 3.1 and in Ref. [6] this is likely to originate from ignoring the isospin corrections in the DSSZ fit. The sea-quark modifications look also mutually rather alike, the EPPS16 uncertainties being somewhat larger than the others as, being flavor dependent, the sea quarks in EPPS16 have more degrees of freedom. As has been understood already some while ago [5,6], the DSSZ parametrization has almost no nuclear effects in gluons as nuclear effects were included in the FFs [36] when computing inclusive pion production at RHIC. As a result, DSSZ does not reproduce the new CMS dijet measurements as shown here in Fig. 26. Between EPS09 and EPPS16, the gluon uncertainties are larger in EPPS16. While EPPS16 includes more constraints for the gluons (especially the CMS dijet data), in EPS09 the PHENIX data was assigned an additional weight factor of 20. This in effect increased the importance of these data, making the uncertainties smaller than what they would have been without such a weight (the baseline-fit gluons in Fig. 22 serve as a representative of an unweighted case). In addition, in EPPS16 one more gluon parameter is left free (x_a) which also increases the uncertainties in comparison to EPS09.

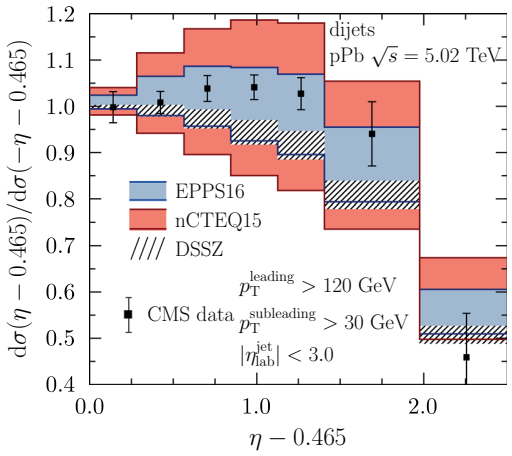

Fig. 26 The CMS dijet data [37] compared with the results obtained with the EPPS16 (*blue bands*), nCTEQ15 [35] (*red bands*) and DSSZ [34] (*hatched bands*) nuclear PDFs

6 Application: W charge asymmetry

The W charge-asymmetry measurement by CMS in pPb collisions [46] revealed some deviations from the NLO calculations in the backward direction and it was suggested that this difference could be due to flavor-dependent PDF nuclear modifications. While it was shown in Ref. [103] that such a difference does not appear in the ATLAS PbPb data [104] at the same probed values of x, the situation still remains unclear. To see how large variations the new EPPS16 can accommodate, we compare in Fig. 28 the CMS data with the EPPS16 and EPS09 predictions using the CT14NLO proton PDFs. As discussed in the original EPS09 paper [33], the total uncertainty should be computed by adding in quadrature the uncertainties stemming separately from EPPS16 and from the free-proton baseline PDFs,

$$(\delta\mathcal{O}_{\text{total}})^2 = (\delta\mathcal{O}_{\text{EPPS16}})^2 + (\delta\mathcal{O}_{\text{baseline}})^2, \tag{56}$$

where $\delta\mathcal{O}_{\text{EPPS16}}$ is evaluated by Eq. (53) using the uncertainty sets of EPPS16 with the central set of free-proton PDFs, and $\delta\mathcal{O}_{\text{baseline}}$ by the same equation but using the free-proton error sets with the central set of EPPS16. The same has been done in the case of EPS09 results. While the differences between the central predictions of EPPS16 and EPS09 are tiny, it can be seen that the uncertainty bands of EPPS16 are clearly wider and, within the uncertainties, the data and EPPS16 are in a fair agreement. As this observable is mostly sensitive to the free-proton baseline (to first approximation the nuclear effects in PDFs cancel) we do not use these asymmetry data as a constraint in the actual fit in which we aim to expose the nuclear effects in PDFs.

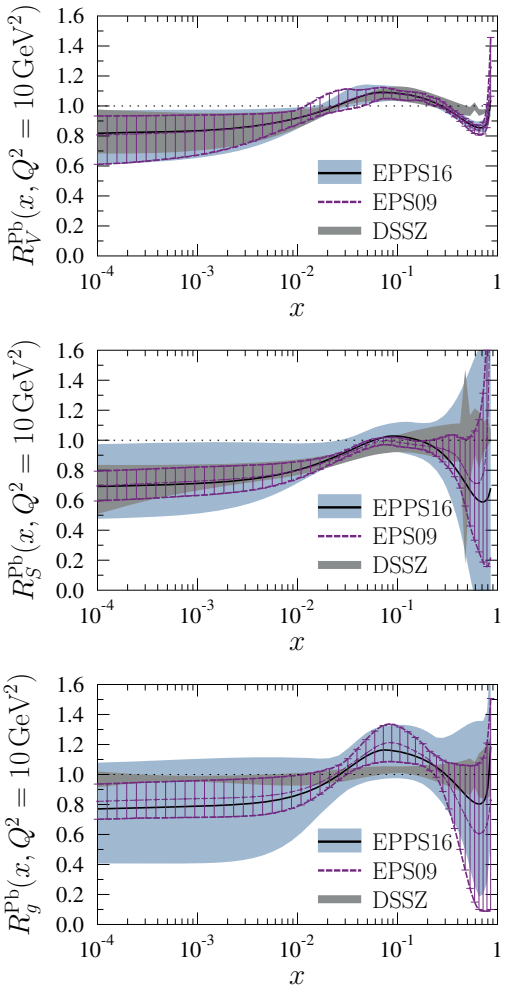

Fig. 27 Comparison of the EPPS16 nuclear modifications (*black central curve with light-blue uncertainty bands*) to those from the EPS09 analysis (*purple curves with hatching*) and DSSZ [34] (*gray bands*) at $Q^2 = 10\,\text{GeV}^2$. The *upper panels* correspond to the average valence and sea-quark modifications of Eqs. (54) and (55), the *bottom panel* is for gluons

7 Summary and outlook

We have introduced a significantly updated global analysis of NLO nuclear PDFs – EPPS16 – with less biased, flavor-dependent fit functions and a larger variety of data constraints than in other concurrent analyses. In particular, new LHC data from the 2013 pPb run are for the first time directly included. Another important addition here is the neutrino–nucleus DIS data. Also the older pion–nucleus DY data are now for the first time part of the analysis. From the new data, the most significant role is played by the neutrino DIS data and the LHC dijet measurements whose addition leads to a consistent picture of qualitatively similar nuclear modifications for all partonic species. Remarkably, the addition of

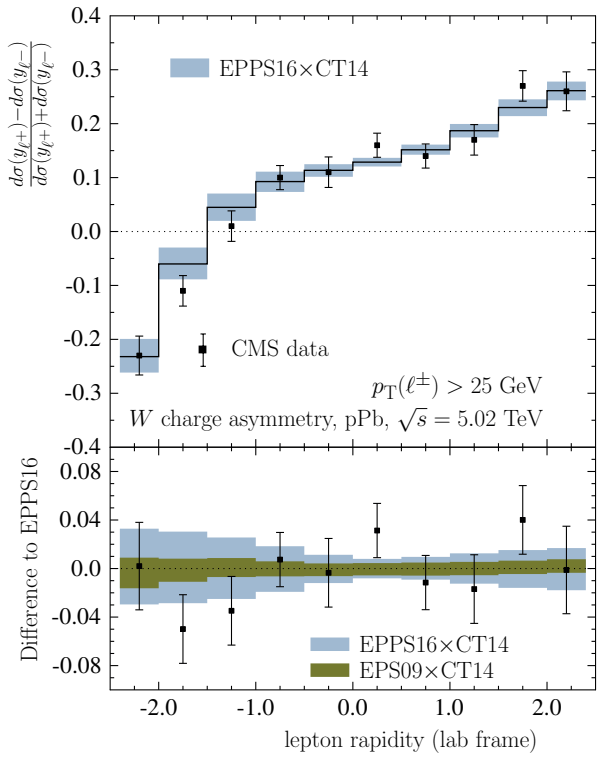

Fig. 28 The CMS W charge-asymmetry measurement [46] compared with the predictions using EPPS16 nuclear modifications and CT14NLO proton PDFs. In both panels the *blue bands* correspond to the combined EPPS16+CT14 uncertainty and in the *lower panel the green band* to the combined EPS09+CT14 uncertainty

new data types into the global fit does not generate notable tensions with the previously considered data sets. This lends support to the validity of collinear factorization and process-independent nuclear PDFs in the kinematical x, Q^2 region we have considered.

However, the uncertainties are still significant for all components and, clearly, more data is therefore required. In this respect, the prospects for rapid developments of nuclear PDFs are very good: It can be expected that new data from the LHC will be available soon. For example, from the 2013 pPb data taking, a more differential dijet analysis by the CMS collaboration [105] as well as W data by ATLAS [106] are still being prepared. In November–December 2016, the LHC has recorded pPb collisions at the highest energy ever, $\sqrt{s} = 8.16$ TeV, with more than six times more statistics than that from the 2013 pPb run at $\sqrt{s} = 5.02$ TeV.[5] The new data from this run will provide further constraints to the nuclear PDFs in the near future. As in the case of free-proton PDFs [107,108] heavy-flavor production at forward direction [109] may offer novel small-x input. An interesting opportunity is also the possibility of the LHCb experiment to operate in a fixed-target mode and measure e.g. pNe (and other noble gases) collisions [110]. From other experiments, new fixed-target proton-induced Drell–Yan data from the Fermilab E-906/SeaQuest experiment [115] should also provide better constraints e.g. for the A dependence of the sea-quark nuclear modifications.

Further in the future, the planned Electron–Ion Collider [111] (and LHeC [112] if materialized) will provide high-precision DIS constraints for all nuclear parton flavors. In addition, the possible realization of a new forward calorimeter (FOCAL) at the ALICE experiment [113] would, in turn, give a possibility to measure isolated photons in the region sensitive to low x gluons [114].

On the theoretical side, there is ample room for improvements as well. For example, similarly to the free-proton fits, an upgrade to next-to-NLO or inclusion of photon distributions and mixed QCD–QED parton evolution are obvious further developments. In a longer run, to avoid biases due to specific baseline proton PDFs, especially regarding the s quark sector, fitting the proton PDFs and nuclear PDFs in one single analysis is ultimately needed.

Acknowledgements This research was supported by the Academy of Finland, Project 297058 of K.J.E.; by the European Research Council Grant HotLHC ERC-2011-StG-279579; by Ministerio de Ciencia e Innovación of Spain under project FPA2014-58293-C2-1-P; and by Xunta de Galicia (Conselleria de Educacion) – H.P. and C.A.S. are part of the Strategic Unit AGRUP2015/11. P.P. acknowledges the financial support from the Magnus Ehrnrooth Foundation. Part of the computing has been done in T. Lappi's project at CSC, the IT Center for Science in Espoo, Finland.

Funded by SCOAP[3].

References

1. C.A. Salgado et al., J. Phys. G **39**, 015010 (2012). doi:10.1088/0954-3899/39/1/015010. arXiv:1105.3919 [hep-ph]
2. C.A. Salgado, J.P. Wessels, Ann. Rev. Nucl. Part. Sci. **66**, 449 (2016)
3. N. Armesto, E. Scomparin, Eur. Phys. J. Plus **131**(3), 52 (2016). doi:10.1140/epjp/i2016-16052-4. arXiv:1511.02151 [nucl-ex]
4. P. Foka, M.A. Janik, Rev. Phys. **1**, 172 (2016). doi:10.1016/j.revip.2016.11.001
5. K.J. Eskola, Nucl. Phys. A **910–911**, 163 (2013). doi:10.1016/j.nuclphysa.2012.12.029. arXiv:1209.1546 [hep-ph]
6. H. Paukkunen, Nucl. Phys. A **926**, 24 (2014). doi:10.1016/j.nuclphysa.2014.04.001. arXiv:1401.2345 [hep-ph]
7. J.J. Aubert et al., European Muon Collaboration. Phys. Lett. B **123**, 275 (1983). doi:10.1016/0370-2693(83)90437-9
8. M. Arneodo, Phys. Rep. **240**, 301 (1994). doi:10.1016/0370-1573(94)90048-5

[5] https://lpc.web.cern.ch/lumiplots_2016_PbPb.htm.

9. L.L. Frankfurt, M.I. Strikman, S. Liuti, Phys. Rev. Lett. **65**, 1725 (1990). doi:10.1103/PhysRevLett.65.1725
10. K.J. Eskola, Nucl. Phys. B **400**, 240 (1993). doi:10.1016/0550-3213(93)90406-F
11. Y.L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977)
12. Y.L. Dokshitzer, Zh. Eksp. Teor. Fiz. **73**, 1216 (1977)
13. V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972)
14. V.N. Gribov, L.N. Lipatov, Yad. Fiz. **15**, 781 (1972)
15. V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 675 (1972)
16. V.N. Gribov, L.N. Lipatov, Yad. Fiz. **15**, 1218 (1972)
17. G. Altarelli, G. Parisi, Nucl. Phys. B **126**, 298 (1977). doi:10.1016/0550-3213(77)90384-4
18. K.J. Eskola, V.J. Kolhinen, P.V. Ruuskanen, Nucl. Phys. B **535**, 351 (1998). doi:10.1016/S0550-3213(98)00589-6. arXiv:hep-ph/9802350
19. K.J. Eskola, V.J. Kolhinen, C.A. Salgado, Eur. Phys. J. C **9**, 61 (1999). doi:10.1007/s100520050513, 10.1007/s100529900005. arXiv:hep-ph/9807297
20. K.J. Eskola, V.J. Kolhinen, H. Paukkunen, C.A. Salgado, JHEP **0705**, 002 (2007). doi:10.1088/1126-6708/2007/05/002. arXiv:hep-ph/0703104
21. M. Hirai, S. Kumano, M. Miyama, Phys. Rev. D **64**, 034003 (2001). doi:10.1103/PhysRevD.64.034003. arXiv:hep-ph/0103208
22. M. Hirai, S. Kumano, T.-H. Nagai, Phys. Rev. C **70**, 044905 (2004). doi:10.1103/PhysRevC.70.044905. arXiv:hep-ph/0404093
23. D. de Florian, R. Sassot, Phys. Rev. D **69**, 074028 (2004). doi:10.1103/PhysRevD.69.074028. arXiv:hep-ph/0311227
24. M. Hirai, S. Kumano, T.-H. Nagai, Phys. Rev. C **76**, 065207 (2007). doi:10.1103/PhysRevC.76.065207. arXiv:0709.3038 [hep-ph]
25. I. Schienbein, J.Y. Yu, K. Kovarik, C. Keppel, J.G. Morfin, F. Olness, J.F. Owens, Phys. Rev. D **80**, 094004 (2009). doi:10.1103/PhysRevD.80.094004. arXiv:0907.2357 [hep-ph]
26. S. Atashbar, Tehrani. Phys. Rev. C **86**, 064301 (2012). doi:10.1103/PhysRevC.86.064301
27. H. Khanpour, S. Atashbar Tehrani, Phys. Rev. D **93**(1), 014026 (2016). doi:10.1103/PhysRevD.93.014026. arXiv:1601.00939 [hep-ph]
28. L. Frankfurt, V. Guzey, M. Strikman, Phys. Rep. **512**, 255 (2012). doi:10.1016/j.physrep.2011.12.002. arXiv:1106.2091 [hep-ph]
29. N. Armesto, J. Phys. G **32**, R367 (2006). doi:10.1088/0954-3899/32/11/R01. arXiv:hep-ph/0604108
30. S.A. Kulagin, R. Petti, Nucl. Phys. A **765**, 126 (2006). doi:10.1016/j.nuclphysa.2005.10.011. arXiv:hep-ph/0412425
31. S.S. Adler et al., PHENIX Collaboration, Phys. Rev. Lett. **98**, 172302 (2007). doi:10.1103/PhysRevLett.98.172302. arXiv:nucl-ex/0610036
32. K.J. Eskola, H. Paukkunen, C.A. Salgado, JHEP **0807**, 102 (2008). doi:10.1088/1126-6708/2008/07/102. arXiv:0802.0139 [hep-ph]
33. K.J. Eskola, H. Paukkunen, C.A. Salgado, JHEP **0904**, 065 (2009). doi:10.1088/1126-6708/2009/04/065. arXiv:0902.4154 [hep-ph]
34. D. de Florian, R. Sassot, P. Zurita, M. Stratmann, Phys. Rev. D **85**, 074028 (2012). doi:10.1103/PhysRevD.85.074028. arXiv:1112.6324 [hep-ph]
35. K. Kovarik et al., Phys. Rev. D **93**(8), 085037 (2016). doi:10.1103/PhysRevD.93.085037. arXiv:1509.00792 [hep-ph]
36. R. Sassot, M. Stratmann, P. Zurita, Phys. Rev. D **81**, 054001 (2010). doi:10.1103/PhysRevD.81.054001. arXiv:0912.1311 [hep-ph]
37. S. Chatrchyan et al. [CMS Collaboration], Eur. Phys. J. C **74**(7), 2951 (2014). doi:10.1140/epjc/s10052-014-2951-y. arXiv:1401.4433 [nucl-ex]
38. H. Paukkunen, K.J. Eskola, C. Salgado, Nucl. Phys. A **931**, 331 (2014). doi:10.1016/j.nuclphysa.2014.07.012. arXiv:1408.4563 [hep-ph]
39. N. Armesto, H. Paukkunen, J.M. Penín, C.A. Salgado, P. Zurita, Eur. Phys. J. C **76**(4), 218 (2016). doi:10.1140/epjc/s10052-016-4078-9. arXiv:1512.01528 [hep-ph]
40. K. Kovarik, I. Schienbein, F.I. Olness, J.Y. Yu, C. Keppel, J.G. Morfin, J.F. Owens, T. Stavreva, Phys. Rev. Lett. **106**, 122301 (2011). doi:10.1103/PhysRevLett.106.122301. arXiv:1012.0286 [hep-ph]
41. M. Hirai, arXiv:1603.07854 [hep-ph]
42. H. Paukkunen, C.A. Salgado, JHEP **1007**, 032 (2010). doi:10.1007/JHEP07(2010)032. arXiv:1004.3140 [hep-ph]
43. H. Paukkunen, C.A. Salgado, Phys. Rev. Lett. **110**(21), 212301 (2013). doi:10.1103/PhysRevLett.110.212301. arXiv:1302.2001 [hep-ph]
44. A. Kusina et al., arXiv:1610.02925 [nucl-th]
45. H. Paukkunen, P. Zurita, JHEP **1412**, 100 (2014). doi:10.1007/JHEP12(2014)100. arXiv:1402.6623 [hep-ph]
46. V. Khachatryan et al., CMS Collaboration, Phys. Lett. B **750**, 565 (2015). doi:10.1016/j.physletb.2015.09.057. arXiv:1503.05825 [nucl-ex]
47. J. Adam et al. [ALICE Collaboration], arXiv:1611.03002 [nucl-ex]
48. V. Khachatryan et al., CMS Collaboration, Phys. Lett. B **759**, 36 (2016). doi:10.1016/j.physletb.2016.05.044. arXiv:1512.06461 [hep-ex]
49. G. Aad *et al.* [ATLAS Collaboration], Phys. Rev. C **92**(4), 044915 (2015). doi:10.1103/PhysRevC.92.044915. arXiv:1507.06232 [hep-ex]
50. G. Onengut et al., CHORUS Collaboration, Phys. Lett. B **632**, 65 (2006). doi:10.1016/j.physletb.2005.10.062
51. J. Badier et al., NA3 Collaboration, Phys. Lett. B **104**, 335 (1981). doi:10.1016/0370-2693(81)90137-4
52. P. Bordalo et al., NA10 Collaboration, Phys. Lett. B **193**, 368 (1987). doi:10.1016/0370-2693(87)91253-6
53. J.G. Heinrich et al., Phys. Rev. Lett. **63**, 356 (1989). doi:10.1103/PhysRevLett.63.356
54. https://www.jyu.fi/fysiikka/en/research/highenergy/urhic/nPDFs
55. S. Dulat et al., Phys. Rev. D **93**(3), 033006 (2016). doi:10.1103/PhysRevD.93.033006. arXiv:1506.07443 [hep-ph]
56. M. Kramer, F.I. Olness, D.E. Soper, Phys. Rev. D **62**, 096007 (2000). doi:10.1103/PhysRevD.62.096007. arXiv:hep-ph/0003035
57. J.C. Collins, Phys. Rev. D **58**, 094002 (1998). doi:10.1103/PhysRevD.58.094002. arXiv:hep-ph/9806259
58. R.S. Thorne, W.K. Tung, arXiv:0809.0714 [hep-ph]
59. I. Helenius, H. Paukkunen, N. Armesto, arXiv:1606.09003 [hep-ph]
60. A.D. Martin, A.J.T.M. Mathijssen, W.J. Stirling, R.S. Thorne, B.J.A. Watt, G. Watt, Eur. Phys. J. C **73**(2), 2318 (2013). doi:10.1140/epjc/s10052-013-2318-9. arXiv:1211.1215 [hep-ph]
61. W. Furmanski, R. Petronzio, Phys. Lett. B **97**, 437 (1980). doi:10.1016/0370-2693(80)90636-X
62. G. Curci, W. Furmanski, R. Petronzio, Nucl. Phys. B **175**, 27 (1980). doi:10.1016/0550-3213(80)90003-6
63. P. Santorelli, E. Scrimieri, Phys. Lett. B **459**, 599 (1999). doi:10.1016/S0370-2693(99)00698-X. arXiv:hep-ph/9807572
64. P. Santorelli, E. Scrimieri, Phys. Lett. B **459**, 599 (1999). doi:10.1016/S0370-2693(99)00698-X. arXiv:hep-ph/9807572
65. I. Abt, A.M. Cooper-Sarkar, B. Foster, V. Myronenko, K. Wichmann, M. Wing, arXiv:1604.02299 [hep-ph]
66. D. Dutta, J.C. Peng, I.C. Cloet, D. Gaskell, Phys. Rev. C **83**, 042201 (2011). doi:10.1103/PhysRevC.83.042201. arXiv:1007.3916 [nucl-ex]

67. P. Paakkinen, K.J. Eskola, H. Paukkunen, Phys.Lett. B **768**, 7–11 (2017). doi:10.1016/j.physletb.2017.02.009
68. H. Paukkunen, C.A. Salgado, JHEP **1103**, 071 (2011). doi:10.1007/JHEP03(2011)071. arXiv:1010.5392 [hep-ph]
69. P. Ru, B.W. Zhang, L. Cheng, E. Wang, W.N. Zhang, J. Phys. G **42**(8), 085104 (2015). doi:10.1088/0954-3899/42/8/085104. arXiv:1412.2930 [nucl-th]
70. K.J. Eskola, H. Paukkunen, C. A. Salgado, JHEP **1310**, 213 (2013). doi:10.1007/JHEP10(2013)213. arXiv:1308.6733 [hep-ph]
71. P. Ru, S.A. Kulagin, R. Petti, B.W. Zhang, arXiv:1608.06835 [nucl-th]
72. J. Gomez et al., Phys. Rev. D **49**, 4348 (1994)
73. P. Amaudruz et al., New Muon Collaboration, Nucl. Phys. B **441**, 3 (1995). doi:10.1016/0550-3213(94)00023-9. arXiv:hep-ph/9503291
74. M. Arneodo et al., New Muon Collaboration, Nucl. Phys. B **441**, 12 (1995). doi:10.1016/0550-3213(95)00023-2. arXiv:hep-ex/9504002
75. M. Arneodo et al., New Muon Collaboration, Nucl. Phys. B **481**, 3 (1996). doi:10.1016/S0550-3213(96)90117-0
76. D.M. Alde et al., Phys. Rev. Lett. **64**, 2479 (1990). doi:10.1103/PhysRevLett.64.2479
77. M.A. Vasilev et al., NuSea Collaboration. Phys. Rev. Lett. **83**, 2304 (1999). doi:10.1103/PhysRevLett.83.2304. arXiv:hep-ex/9906010
78. J. Ashman et al., European Muon Collaboration, Z. Phys. C **57**, 211 (1993)
79. M. Arneodo et al., New Muon Collaboration, Nucl. Phys. B **481**, 23 (1996). doi:10.1016/S0550-3213(96)90119-4
80. P. Amaudruz et al., New Muon Collaboration, Nucl. Phys. B **371**, 3 (1992)
81. J.P. Berge et al., Z. Phys. C **49**, 187 (1991). doi:10.1007/BF01555493
82. M. Tzanov et al., NuTeV Collaboration, Phys. Rev. D **74**, 012008 (2006). doi:10.1103/PhysRevD.74.012008, arXiv:hep-ex/0509010
83. A.B. Arbuzov, D.Y. Bardin, L.V. Kalinovskaya, JHEP **0506**, 078 (2005). doi:10.1088/1126-6708/2005/06/078, arXiv:hep-ph/0407203
84. A. Accardi, J.W. Qiu, JHEP **0807**, 090 (2008). doi:10.1088/1126-6708/2008/07/090. arXiv:0805.1496 [hep-ph]
85. J. Gao et al., Phys. Rev. D **89**(3), 033009 (2014). doi:10.1103/PhysRevD.89.033009. arXiv:1302.6246 [hep-ph]
86. J.M. Campbell, R.K. Ellis, W.T. Giele, Eur. Phys. J. C **75**(6), 246 (2015). doi:10.1140/epjc/s10052-015-3461-2. arXiv:1503.06182 [physics.comp-ph]
87. J. Gao, Z. Liang, D.E. Soper, H.L. Lai, P.M. Nadolsky, C.-P. Yuan, Comput. Phys. Commun. **184**, 1626 (2013). doi:10.1016/j.cpc.2013.01.022. arXiv:1207.0513 [hep-ph]
88. Z. Kunszt, D.E. Soper, Phys. Rev. D **46**, 192 (1992). doi:10.1103/PhysRevD.46.192
89. S.D. Ellis, Z. Kunszt, D.E. Soper, Phys. Rev. Lett. **69**, 1496 (1992). doi:10.1103/PhysRevLett.69.1496
90. F. Aversa, P. Chiappetta, M. Greco, J.P. Guillet, Nucl. Phys. B **327**, 105 (1989). doi:10.1016/0550-3213(89)90288-5
91. B.A. Kniehl, G. Kramer, B. Potter, Nucl. Phys. B **582**, 514 (2000). doi:10.1016/S0550-3213(00)00303-5. arXiv:hep-ph/0010289
92. M. Gluck, E. Reya, A. Vogt, Z. Phys. C **53**, 651 (1992). doi:10.1007/BF01559743
93. G. D'Agostini, Nucl. Instrum. Methods A **346**, 306 (1994). doi:10.1016/0168-9002(94)90719-6
94. K. Levenberg, Q. Appl. Math. **2**, 164 (1944)
95. D.W. Marquardt, J. Soc. Ind. Appl. Math. **11**(2), 431 (1963)
96. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (ISBN-9780521430647)
97. F. James, CERN-D-506, CERN-D506
98. J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.L. Lai, W.K. Tung, Phys. Rev. D **65**, 014013 (2001). doi:10.1103/PhysRevD.65.014013. arXiv:hep-ph/0101032
99. A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189 (2009). doi:10.1140/epjc/s10052-009-1072-5. arXiv:0901.0002 [hep-ph]
100. J. Pumplin, D.R. Stump, W.K. Tung, Phys. Rev. D **65**, 014011 (2001). doi:10.1103/PhysRevD.65.014011. arXiv:hep-ph/0008191
101. J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, P.M. Nadolsky, W.K. Tung, JHEP **0207**, 012 (2002). doi:10.1088/1126-6708/2002/07/012. arXiv:hep-ph/0201195
102. P.M. Nadolsky, Z. Sullivan, eConf C **010630**, P510 (2001). arXiv:hep-ph/0110378
103. F. Arleo, E. Chapon, H. Paukkunen, Eur. Phys. J. C **76**(4), 214 (2016). doi:10.1140/epjc/s10052-016-4049-1. arXiv:1509.03993 [hep-ph]
104. G. Aad et al. [ATLAS Collaboration], Eur. Phys. J. C **75**(1), 23 (2015). doi:10.1140/epjc/s10052-014-3231-6. arXiv:1408.4674 [hep-ex]
105. CMS Collaboration [CMS Collaboration], CMS-PAS-HIN-16-003
106. The ATLAS collaboration, ATLAS-CONF-2015-056
107. O. Zenaiev et al. [PROSA Collaboration], Eur. Phys. J. C **75**(8), 396 (2015). doi:10.1140/epjc/s10052-015-3618-z. arXiv:1503.04581 [hep-ph]
108. R. Gauld, J. Rojo, arXiv:1610.09373 [hep-ph]
109. The LHCb Collaboration [LHCb Collaboration], LHCb-CONF-2016-003, CERN-LHCb-CONF-2016-003
110. Y. Zhang [LHCb Collaboration], arXiv:1605.07509 [hep-ex]
111. A. Accardi et al., Eur. Phys. J. A **52**(9), 268 (2016). doi:10.1140/epja/i2016-16268-9. arXiv:1212.1701 [nucl-ex]
112. J.L. Abelleira Fernandez et al., LHeC Study Group Collaboration, J. Phys. G **39**, 075001 (2012). doi:10.1088/0954-3899/39/7/075001. arXiv:1206.2913 [physics.acc-ph]
113. T. Peitzmann [ALICE FoCal Collaboration], PoS DIS **2016**, 273 (2016). arXiv:1607.01673 [hep-ex]
114. I. Helenius, K.J. Eskola, H. Paukkunen, JHEP **1409**, 138 (2014). doi:10.1007/JHEP09(2014)138. arXiv:1406.1689 [hep-ph]
115. B.P. Dannowitz, FERMILAB-THESIS-2016-13

III

NON-QUADRATIC IMPROVED HESSIAN PDF REWEIGHTING AND APPLICATION TO CMS DIJET MEASUREMENTS AT 5.02 TEV

by

Eskola, K. J., Paakkinen, P. & Paukkunen, H., 2019

Eur. Phys. J. C79, no.6 (2019) 511

arXiv: 1903.09832 [hep-ph]

Eur. Phys. J. C (2019) 79:511
https://doi.org/10.1140/epjc/s10052-019-6982-2

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

Non-quadratic improved Hessian PDF reweighting and application to CMS dijet measurements at 5.02 TeV

Kari J. Eskola[1,2,a], **Petja Paakkinen**[1,2,b], **Hannu Paukkunen**[1,2,c]

[1] Department of Physics, University of Jyvaskyla, P.O. Box 35, Fl-40014 Jyvaskyla, Finland
[2] Helsinki Institute of Physics, University of Helsinki, P.O. Box 64, Fl-00014 Helsinki, Finland

Received: 27 March 2019 / Accepted: 23 May 2019

Abstract Hessian PDF reweighting, or "profiling", has become a widely used way to study the impact of a new data set on parton distribution functions (PDFs) with Hessian error sets. The available implementations of this method have resorted to a perfectly quadratic approximation of the initial χ^2 function before inclusion of the new data. We demonstrate how one can take into account the first non-quadratic components of the original fit in the reweighting, provided that the necessary information is available. We then apply this method to the CMS measurement of dijet pseudorapidity spectra in proton–proton (pp) and proton–lead (pPb) collisions at 5.02 TeV. The measured pp dijet spectra disagree with next-to-leading order (NLO) theory calculations using the CT14 NLO PDFs, but upon reweighting the CT14 PDFs, these can be brought to a much better agreement. We show that the needed proton-PDF modifications also have a significant impact on the predictions for the pPb dijet distributions. Taking the ratio of the individual spectra, the proton-PDF uncertainties effectively cancel, giving a clean probe of the PDF nuclear modifications. We show that these data can be used to further constrain the EPPS16 nuclear PDFs and strongly support gluon nuclear shadowing at small x and antishadowing at around $x \approx 0.1$.

1 Introduction

The proton structure at high momentum-transfer, as encoded in the collinearly factorized parton distribution functions (PDFs), is not only an interesting subject in its own right, but plays a pivotal role in many applications, such as precision electroweak and Higgs physics, searches for new physics, etc. [1]. Likewise, their counterparts for nucleons bound in nuclei, the nuclear PDFs (nPDFs), are essential in e.g. studying the production of hard probes of the Quark Gluon Plasma [2]. In practice, despite the ongoing effort in lattice methods [3], the PDFs are obtained by the well-established means of global analysis using hard-process data. As such, the PDFs have uncertainties which derive from those in the available data and also from the lack of data constraints in certain phase-space regions. It is then often the case that when new data are published or a future experiment is being planned, one would like to study the impact that the measurement could have on the PDFs. A good example of such a case is the recent CMS measurement of dijet pseudorapidity spectra in proton–proton (pp) and proton–lead (pPb) collisions at 5.02 TeV [4], where, on one hand, the measured pp spectra seem to be in a disagreement with next-to-leading order (NLO) perturbative QCD (pQCD) calculations using CT14 [5] and MMHT14 [6] PDFs (see the Supplemental Material of Ref. [4]), while, on the other hand, the nuclear-modification ratio of the pPb and pp spectra appear to have much smaller uncertainties than predictions with various nPDFs. One should therefore study the impact these data could have on both the free-proton PDFs and their nuclear modifications.

As producing a full global fit remains rather involved, even with publicly available tools like the xFitter [7] (built upon the former HERAFitter [8]) coming available, it is in most cases impractical for a general user to try to learn about the constraining power of a data set in this way. For this purpose, approximative methods have been developed, first in the formalism of Bayesian reweighting of Monte Carlo PDF ensembles [9–13] and later in a framework using Hessian error sets [14–16]. These methods have their limitations, as the new PDFs rely on all the theoretical assumptions of the original PDF analysis, such as the parametrization form, the value of α_s and the used heavy-quark scheme. There are also limitations related to how well the methods approximate the true parameter likelihood in the region constrained by the new data. In particular, the applications of Hessian PDF reweight-

[a] e-mail: kari.eskola@jyu.fi
[b] e-mail: petja.paakkinen@jyu.fi
[c] e-mail: hannu.paukkunen@jyu.fi

Published online: 15 June 2019

ing have resorted to a perfectly quadratic approximation of the χ^2 goodness-of-fit function before inclusion of the new data and to linear or up to quadratic terms for responses in the new observables. This applies to the implementation in the xFitter package, where the method is referred to as "Hessian profiling", as well as to the new software package which has appeared under the name ePump [16]. It is not, however, uncommon that non-quadratic terms in the χ^2 function are large (see e.g. Figure 6 of Ref. [17]), and thus it would be beneficial to have a way to take these into account.

The purpose of this article is twofold: first, in Sect. 2, we describe how one can include into Hessian PDF reweighting the first non-quadratic terms in the χ^2 function consistently with the original fit, provided that the needed information is available. Second, in Sect. 3, we apply the Hessian PDF reweighting to the aforementioned CMS dijet measurements at 5.02 TeV [4]. We show that the strong disagreement between the pp measurement and next-to-leading order (NLO) calculations using CT14 NLO PDFs [5] can be brought to a much better agreement upon reweighting the CT14 PDFs, but that this requires rather strong modifications for high-x gluons. We demonstrate that such changes in the proton PDFs have also an important impact on predictions for dijet production in pPb. Finally, we then reweight the EPPS16 nPDFs [18] with the nuclear modification ratio of the measured pPb and pp dijet spectra using the non-quadratic approximation developed in Sect. 2 and present a discussion on the importance of these higher-order terms in the reweighting. Preliminary work on this topic can be found in Refs. [19,20].

2 PDF uncertainties and reweighting in Hessian method

In this section, we first recapitulate the uncertainty determination in the Hessian approach [21], assuming the use of a global tolerance criterion. We then describe how one can perform a reweighting upon such determined error sets, taking into account the first non-quadratic terms in the χ^2 function. We end the section with a discussion on the applicability of this method in the case of non-global tolerances.

2.1 Hessian uncertainties with global tolerance criterion

In PDF global analyses, the goodness-of-fit of a parameter vector $\mathbf{a}$ is dictated by the χ^2 function

$$\chi^2(\mathbf{a}) = \sum_{ij} (y_i(\mathbf{a}) - y_i^{\text{data}})\, C_{ij}^{-1}\, (y_j(\mathbf{a}) - y_j^{\text{data}}), \tag{1}$$

where $y_i(\mathbf{a})$ are theory predictions for the observables included in the analysis, y_i^{data} the corresponding measured values and C_{ij}^{-1} the elements of the inverse covariance matrix for these data. In the Hessian method for uncertainty estimation, one takes the parameter values $\mathbf{a}^{\min}$ which minimize Eq. (1), $\chi^2(\mathbf{a}^{\min}) \equiv \min \chi^2(\mathbf{a}) \equiv \chi_0^2$, as the central, best-fit values and studies the behaviour of the χ^2 function around this minimum to determine the uncertainty in these parameters.

The leading deviations from the minimum value χ_0^2 are given by the quadratic approximation

$$\chi^2 \approx \chi_0^2 + \sum_{ij} (a_i - a_i^{\min})\, H_{ij}\, (a_j - a_j^{\min}), \tag{2}$$

where $H_{ij} = \frac{1}{2}\partial^2\chi^2/\partial a_i \partial a_j|_{\mathbf{a}=\mathbf{a}^{\min}}$ are the elements of the Hessian matrix. In practice, these elements need to be obtained numerically. Since the Hessian matrix is symmetric, it has a complete set of orthonormal eigenvectors $\mathbf{v}^{(k)}$ such that

$$\sum_j H_{ij}\, v_j^{(k)} = \varepsilon_k\, v_i^{(k)}, \tag{3}$$

$$\sum_i v_i^{(k)}\, v_i^{(\ell)} = \delta_{k\ell}, \qquad \sum_k v_i^{(k)}\, v_j^{(k)} = \delta_{ij}, \tag{4}$$

where ε_k are the eigenvalues of the Hessian matrix. With this eigendecomposition we can define new parameters

$$z_k = \sum_i \sqrt{\varepsilon_k}\, v_i^{(k)}\, (a_i - a_i^{\min}) \tag{5}$$

such that Eq. (2) becomes

$$\chi^2 \approx \chi_0^2 + \sum_k z_k^2. \tag{6}$$

Since the new parameters z_k are uncorrelated in the quadratic approximation, one can use the standard law of error propagation to translate the uncertainties in the parameters z_k to the uncertainty of any PDF-dependent quantity X as [21]

$$\Delta X = \sqrt{\sum_k \left(\frac{\partial X}{\partial z_k}\Delta z_k\right)^2}. \tag{7}$$

Given a well justified global tolerance $\Delta\chi^2$ for the allowed growth of χ^2 from its minimum, one can determine the allowed parameter variations Δz_k.[1] If the χ^2 function were perfectly quadratic, the uncertainty of the parameter z_k corresponding to the tolerance $\Delta\chi^2$ would be simply $\Delta z_k = \sqrt{\Delta\chi^2}$. As this is generally not true, one instead finds $\delta z_k^{\pm}$, the positive and negative values of z_k corresponding to the

[1] The intricacies of choosing an appropriate value for $\Delta\chi^2$ are outside the scope of this article, see Refs. [17,21,22] for discussion.

$\Delta\chi^2$ increase, and assigns $\Delta z_k = (\delta z_k^+ - \delta z_k^-)/2$. It is convenient to define error sets $S_i^\pm$ corresponding to parameter values

$$z_k[S_i^\pm] = \begin{array}{ll} \delta z_i^\pm, & k = i \\ 0, & k \neq i \end{array}, \tag{8}$$

along with the central set S_0, where $z_k[S_0] = 0$ for all k. Estimating

$$\frac{\partial X}{\partial z_k} = \frac{X[S_k^+] - X[S_k^-]}{2\,\Delta z_k}, \tag{9}$$

where $X[S_k^\pm]$ stands for the quantity X calculated with the parameter set of Eq. (8), yields then a simple form

$$\Delta X = \frac{1}{2}\ \overline{\sum_k \left(X[S_k^+] - X[S_k^-]\right)^2}. \tag{10}$$

As the response in X to the upward and downward parameter shifts can be uneven, one can alternatively specify an upward–downward asymmetric error prescription e.g. with [23]

$$\delta X^\pm = \ \overline{\sum_k \left[\begin{matrix}\max\\ \min\end{matrix}\left\{X[S_k^+] - X[S_0],\, X[S_k^-] - X[S_0],\, 0\right\}\right]^2}. \tag{11}$$

2.2 Non-quadratic reweighting

In the presence of a new data set, the total χ^2 can be written as

$$\chi^2_{\rm new}(\mathbf{z}) = \chi^2_{\rm old}(\mathbf{z}) + \sum_{ij} \left(y_i(\mathbf{z}) - y_i^{\rm data}\right) C_{ij}^{-1} \left(y_j(\mathbf{z}) - y_j^{\rm data}\right), \tag{12}$$

where y_i ($y_i^{\rm data}$) now correspond to the new theoretical (measured) values and $\chi^2_{\rm old}$ incorporates our knowledge of the original global analysis. Now, as we do not wish to produce a full global analysis with $\chi^2_{\rm new}$, we need to make suitable approximations. The simplest choice is to use the quadratic approximation in Eq. (6), according to the method introduced in Ref. [15], but if the parameter variations $\delta z_k^\pm$ and the global tolerance $\Delta\chi^2$ of this fit are known (as is the case with EPPS16 nPDFs, see Table 2 in Ref. [18]), then $\chi^2_{\rm old}$ can be approximated with a third order polynomial in each of the eigendirections,

$$\chi^2_{\rm old} \approx \chi^2_0 + \sum_k (a_k z_k^2 + b_k z_k^3), \tag{13}$$

where the coefficients are obtained with

$$a_k = \frac{\Delta\chi^2}{\delta z_k^+ - \delta z_k^-}\left(\frac{\delta z_k^+}{(\delta z_k^-)^2} - \frac{\delta z_k^-}{(\delta z_k^+)^2}\right), \tag{14}$$

$$b_k = \frac{\Delta\chi^2}{\delta z_k^+ - \delta z_k^-}\left(\frac{1}{(\delta z_k^+)^2} - \frac{1}{(\delta z_k^-)^2}\right). \tag{15}$$

This is illustrated in Fig. 1 (upper diagram), where we show an example of a situation where the χ^2 grows asymmetrically with respect to z_k. The quadratic approximation fails to acknowledge this fact and a third order polynomial is needed to reproduce the $\Delta\chi^2$ growth at δz_k^- and δz_k^+. Similarly, as illustrated in Fig. 1 (lower diagram), the y_i can be expanded in terms of z_k as

$$y_i(\mathbf{z}) \approx y_i[S_0] + \sum_k (d_{ik} z_k + e_{ik} z_k^2), \tag{16}$$

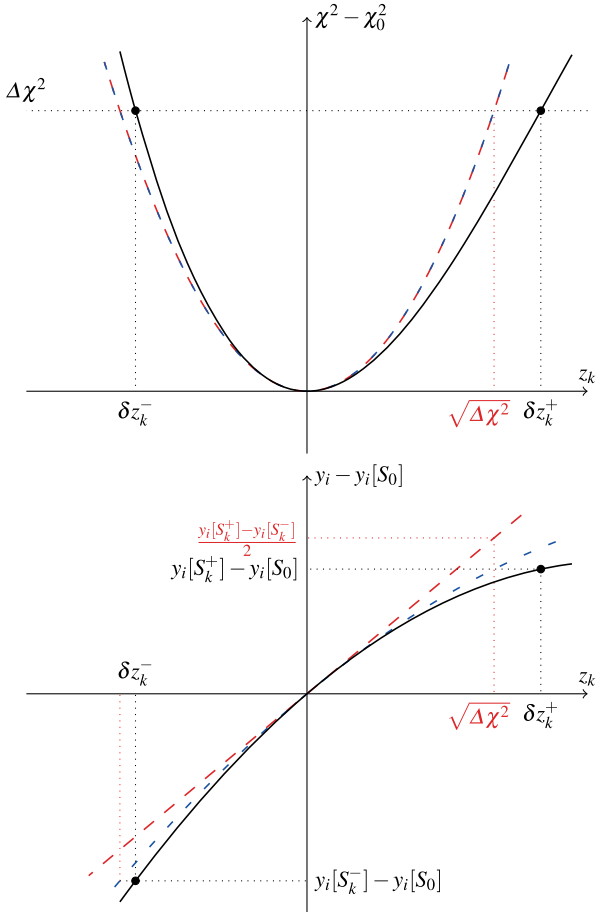

Fig. 1 An illustration for the response of χ^2 (top) and y_i (bottom) with respect to a change of parameter z_k in quadratic–linear (red, long dashed), quadratic–quadratic (blue, short dashed) and cubic–quadratic (black, solid) approximations

where

$$d_{ik} = \frac{1}{\delta z_k^+ - \delta z_k^-}\Bigg[-\frac{\delta z_k^-}{\delta z_k^+}\left(y_i[S_k^+] - y_i[S_0]\right) + \frac{\delta z_k^+}{\delta z_k^-}\left(y_i[S_k^-] - y_i[S_0]\right)\Bigg], \tag{17}$$

$$e_{ik} = \frac{1}{\delta z_k^+ - \delta z_k^-}\Bigg[\frac{1}{\delta z_k^+}\left(y_i[S_k^+] - y_i[S_0]\right) - \frac{1}{\delta z_k^-}\left(y_i[S_k^-] - y_i[S_0]\right)\Bigg]. \tag{18}$$

One should note that the above approximations do not yield a full Taylor expansion to cubic and quadratic order in χ^2_{old} and $y_i(\mathbf{z})$, respectively, as we have neglected off-diagonal terms proportional to $z_l z_k^2$ and $z_l z_k$ for $l \neq k$. Even so, we will refer to reweighting with these approximations as a *cubic–quadratic* one.

Changing variables to $w_k = 2z_k/(\delta z_k^+ - \delta z_k^-)$ and defining $r_k = -\delta z_k^+/\delta z_k^-$, we may alternatively write

$$\chi^2_{\text{new}}(\mathbf{w}) - \chi^2_0 \approx \sum_k (A_k w_k^2 + B_k w_k^3) + \sum_{ij} (y_i(\mathbf{w}) - y_i^{\text{data}})\, C_{ij}^{-1}\, (y_j(\mathbf{w}) - y_j^{\text{data}}), \tag{19}$$

where

$$A_k = \frac{\Delta\chi^2}{4}\left(\frac{1}{r_k^2} + \frac{1}{r_k} + r_k + r_k^2\right), \tag{20}$$

$$B_k = \frac{\Delta\chi^2}{8}\left(\frac{1}{r_k^2} + \frac{2}{r_k} - 2r_k - r_k^2\right), \tag{21}$$

and

$$y_i(\mathbf{w}) \approx y_i[S_0] + \sum_k (D_{ik} w_k + E_{ik} w_k^2), \tag{22}$$

$$D_{ik} = \frac{1}{2}\Bigg[\frac{1}{r_k}\left(y_i[S_k^+] - y_i[S_0]\right) - r_k\left(y_i[S_k^-] - y_i[S_0]\right)\Bigg], \tag{23}$$

$$E_{ik} = \frac{1}{4}\Bigg[\left(1 + \frac{1}{r_k}\right)\left(y_i[S_k^+] - y_i[S_0]\right) + (1 + r_k)\left(y_i[S_k^-] - y_i[S_0]\right)\Bigg]. \tag{24}$$

Now, it is a simple numerical task to minimize Eq. (19) with respect to $\mathbf{w}$. We use MINUIT [24] for the practical applications in the following sections. The found minimum should correspond to that of a full global fit, provided that the approximations (19) and (22) are good enough. This is not trivially true, but we should expect the approximations work better the closer we are to the original minimum. Thus it makes sense to define a "penalty term"

$$P = \sum_k (A_k (w_k^{\min})^2 + B_k (w_k^{\min})^3) \approx \chi^2_{\text{old}}(\mathbf{w}^{\min}) - \chi^2_0, \tag{25}$$

which essentially counts how much χ^2_{old} has grown from its minimum value, $w_k^{\min}$ being the values of w_k at the minimum of $\chi^2_{\text{new}}(\mathbf{w})$. If $P \ll \Delta\chi^2$, the approximations (19) and (22) should work well and the reweighted results can be viewed as a proxy for those of a full global fit. Once P grows close to or above $\Delta\chi^2$, the results of reweighting become more sensitive on the made assumptions and one should be cautious on the interpretations. Moreover, a large P signals a tension between the original fit and the new data, which might be due to incompatibilities of some data sets, but can also be caused by an inflexible PDF parametrization, or other limitations of theory description, such as missing higher-order corrections.

The beauty of the reweighting method lies in the fact that the reweighted result for *any* quantity can be obtained simply by using Eq. (22). For example, the new, reweighted, PDFs are obtained by replacing y_i with f_i. One should note that while this expression is quadratic in w_k, the new PDFs retain a linear dependence on the old ones and thus satisfy the PDF sum rules and DGLAP evolution equations.[2] This applies also to the new error sets, which can be obtained essentially by following the same procedure as in Sect. 2.1, with the exception that the new Hessian matrix $\hat{H}_{kl} = \frac{1}{2}\partial^2\chi^2_{\text{new}}/\partial w_k \partial w_l|_{\mathbf{w}=\mathbf{w}^{\min}}$ in

$$\chi^2_{\text{new}}(\mathbf{w}) \approx \chi^2_{\text{new}}(\mathbf{w}^{\min}) + \sum_{kl} (w_k - w_k^{\min})\, \hat{H}_{kl}\, (w_l - w_l^{\min}) \tag{26}$$

can be put to an explicit form

$$\hat{H}_{kl} = (A_k + 3B_k w_k^{\min})\delta_{kl} + \sum_{ij}(D_{ik} + 2E_{ik} w_k^{\min}) C_{ij}^{-1} (D_{jl} + 2E_{jl} w_l^{\min}) + \sum_{ij}(2E_{ik}\delta_{kl}) C_{ij}^{-1} (y_j(\mathbf{w}^{\min}) - y_j^{\text{data}}), \tag{27}$$

by taking second derivatives of Eq. (19). Diagonalizing $\hat{H}$ and finding the deviations in the new eigenvector directions corresponding to $\Delta\chi^2$ growth from $\chi^2_{\text{new}}(\mathbf{w}^{\min})$, one obtains the parameter values for the new error sets, using which the

[2] The fact that the new PDFs are linear combinations of the original ones, with a certain weight factor applied to each of them, also justifies the usage of term "reweighting" in this context.

uncertainties of any quantity can again be obtained according to Eq. (22).

The cubic–quadratic approximation considered above is not applicable to all cases, as it requires the knowledge of the $\delta z_k^{\pm}$. Lower-order approximations, initially introduced in Ref. [15], can be obtained from the above results by taking appropriate limits. Taking $r_k \to 1$ one finds $A_k = \Delta\chi^2$, $B_k = 0$, thus recovering the quadratic approximation for χ^2_{old}. In this limit also the definition of the penalty term in Eq. (25) reduces to that of Ref. [15]. As y_i retains its quadratic parameter dependence in this limit, we call this a *quadratic–quadratic* approximation. In many cases this is the best option one can resort to, as it only requires access to the PDF error sets and the value of $\Delta\chi^2$. Even simpler, *quadratic–linear*, approximation can be achieved by taking also $E_{ik} \to 0$. This version is very easy to implement, as finding the new central and error sets in this approximation involves only solving a system of linear equations [15].

2.3 Comment on non-global tolerances

The reweighting method can also be extended to non-global tolerances [17], simply by setting

$$A_k = \frac{1}{4}\left((T_k^+)^2\left(\frac{1}{r_k^2}+\frac{1}{r_k}\right)+(T_k^-)^2\left(r_k+r_k^2\right)\right), \tag{28}$$

$$B_k = \frac{1}{8}\left((T_k^+)^2\left(\frac{1}{r_k^2}+\frac{2}{r_k}+1\right)-(T_k^-)^2\left(1+2\,r_k+r_k^2\right)\right), \tag{29}$$

where $(T_k^{\pm})^2 = \chi^2_{\text{old}}(\delta z_k^{\pm}) - \chi_0^2$ are the tolerances of the individual error sets, determined by requiring acceptable values of χ^2 for each individual data set in the original analysis [17]. While the new, reweighted central PDF set can be obtained uniquely in this way, the determination of the new error sets involves additional arbitrariness. As the new eigenvector directions obtained by diagonalizing the Hessian matrix in Eq. (27) are not parallel to the original ones, it is not directly obvious how large tolerances should be allowed in each of these new parameter directions. It was argued in Ref. [16] that if the new eigendirections are not significantly rotated away from the original ones, it would be sufficient to use the original tolerances $(T_k^{\pm})^2$ also for obtaining the new error sets. While this can work in some cases, it would be advisable to have a measure on the amount of parameter rotations in the reweighting to test whether the limits of this assumption are met. Another possibility would be to use a global tolerance for the reweighted PDFs, e.g. by taking the average over the $(T_k^{\pm})^2$, but this also would lead to changing the error definition from the original one, thus reducing the comparability of the new and old uncertainties. In general, setting the new non-global tolerances reliably would require a complete refit.

3 CMS 5.02 TeV dijets and their impact on PDFs

The CMS dijet data [4] consist of distributions of dijet pseudorapidity

$$\eta_{\text{dijet}} = \frac{1}{2}(\eta^{\text{leading}} + \eta^{\text{subleading}}) \tag{30}$$

in bins of average transverse momentum of the jet pair

$$p_{\text{T}}^{\text{ave}} = \frac{1}{2}(p_{\text{T}}^{\text{leading}} + p_{\text{T}}^{\text{subleading}}). \tag{31}$$

Here, $\eta^{\text{(sub)leading}}$ and $p_{\text{T}}^{\text{(sub)leading}}$ refer to the pseudorapidity and transverse momentum of the jet with (second to) largest transverse momentum of the event. Jets are defined with the anti-k_{T} algorithm [25] using a distance parameter $R = 0.3$. The events used in the analysis are required to have a leading jet with transverse momentum $p_{\text{T}}^{\text{leading}} > 30$ GeV and a subleading jet with $p_{\text{T}}^{\text{subleading}} > 20$ GeV and the two jets are required to have an azimuthal angle separation $\Delta\phi > 2\pi/3$. In pPb collisions the two jets are required to be in a rapidity interval $-3 < \eta_{\text{jet}}^{\text{lab}} < 3$ in the laboratory frame. Due to unequal beam energies, $E_{\text{p}} = 4$ TeV and $E_{\text{Pb}} = \frac{82}{208}E_{\text{p}}$, the nucleon–nucleon center-of-mass system is boosted in this frame. To attain corresponding coverages in the center-of-mass frames, CMS measured the pp spectra in the interval $-3.465 < \eta_{\text{jet}}^{\text{lab}} < 2.535$. Here, as in the CMS publication, the pp data are shifted in pseudorapidity by $+0.465$, so that the measured dijets cover a pseudorapidity range $-3 < \eta_{\text{dijet}} < 3$ in both pp and pPb.

The CMS data are self-normalized in each bin of $p_{\text{T}}^{\text{ave}}$, i.e. given in the form

$$\frac{1}{\mathrm{d}\sigma/\mathrm{d}\,p_{\text{T}}^{\text{ave}}}\,\mathrm{d}^2\sigma/\mathrm{d}\,p_{\text{T}}^{\text{ave}}\mathrm{d}\eta_{\text{dijet}}. \tag{32}$$

This is advantageous due to a partial cancellation of correlated experimental (including luminosity-) uncertainties and theoretical hadronization corrections.[3] Accordingly, we do not apply nonperturbative corrections to our predictions. We work at NLO as the NNLO calculations of Ref. [27] are not publicly available at this moment. Our theory calculations are performed with NLOJet++ [28] using the anti-k_{T} algorithm through FastJet package [29]. We fix the factorization and renormalization scales to be the same, $\mu_{\text{F}} = \mu_{\text{R}} = \mu$,

[3] For a demonstration of cancellation of the hadronization effects in the normalization, see Ref. [26]. With the relatively small $R = 0.3$, the contribution from underlying event should be small in the first place.

and use $\mu = p_{\mathrm{T}}^{\mathrm{ave}}$ as our central scale choice to keep consistency with the CT14 and EPPS16 fits, but study also variations around this central scale choice to approximate the magnitude of missing higher-order uncertainties (MHOUs).[4] In all figures, PDF uncertainties are presented with the asymmetric prescription of Eq. (11). As the data correlations are not available, we simply add the statistical and systematical uncertainties in quadrature.

3.1 Proton–proton dijet spectra and CT14 reweighting

The self-normalized pp dijet spectra measured by CMS are shown in Fig. 2 along with theory calculations using the CT14 NLO PDFs. While the predictions describe well the $p_{\mathrm{T}}^{\mathrm{ave}}$ systematics of the data, we see that the predicted pseudorapidity spectra are systematically wider than the measured distributions, with the discrepancy between the data and CT14 central prediction being much larger than the experimental uncertainties, yielding a very poor figure of merit, $\chi^2/N_{\mathrm{data}} = 7.5$. To study the possible source of this discrepancy, we show in Fig. 2 both the uncertainties from CT14 PDFs, as well as factor of two scale variations around the central scale choice $\mu = p_{\mathrm{T}}^{\mathrm{ave}}$ and results from a leading order (LO) calculation at the central scale.

We see that in most bins, especially towards high $p_{\mathrm{T}}^{\mathrm{ave}}$, the discrepancy between the data and theory is larger than the associated scale uncertainty. As the factor of two scale variations often underestimate the true size of higher order corrections (see e.g. Ref. [27]), not much can be learned from this fact alone. However, as the LO-to-NLO corrections shown in Fig. 2 (lower panels) are of the same size as the scale uncertainties, we should not expect the NLO-to-NNLO corrections to be any larger than these. Hence the discrepancy is unlikely to be just due to missing NNLO terms, which in turn points into the direction that the CT14 PDFs need to be modified for a better description of the data. Towards smaller $p_{\mathrm{T}}^{\mathrm{ave}}$ the scale variation effects become more important, leaving room for improvement with NNLO corrections. Another possible scale choice would be the invariant mass of the dijet, $\mu = M_{\mathrm{dijet}}$, a choice which was found in Ref. [27] to yield a better perturbative convergence up to leading-color NNLO precision. We have tested this option, shown also in Fig. 2 (lower panels), and report that here at the NLO level it tends to give smaller scale-uncertainty bands especially at low $p_{\mathrm{T}}^{\mathrm{ave}}$ and that the results do not differ much from the central $\mu = p_{\mathrm{T}}^{\mathrm{ave}}$ predictions. This points again towards smallness of the NNLO corrections. With even slightly wider predictions, $\mu = M_{\mathrm{dijet}}$ gives a worse data description than the $\mu = p_{\mathrm{T}}^{\mathrm{ave}}$ scale choice, and thus we work with the latter in what follows.

To see the modifications on the CT14 PDFs the CMS dijet data would indicate, we have performed a reweighting study with these data. As most of the data points lie outside the CT14 uncertainties, we could expect the needed modifications to be rather strong. Nominally, the CT14 uncertainties correspond to a global tolerance $\Delta\chi^2 = 100$, but to enforce a 90% confidence level agreement individually with each data set used in the analysis, CT14 uses in addition so called "Tier-2 penalties". Hence, the parameter variations $\delta z_k^{\pm}$ in CT14 do not exactly match with $\pm\sqrt{100}$, but can be somewhat smaller. As no detailed information is available on how large these deviations are, the best we can do is to assume χ^2 to be perfectly quadratic and use $\Delta\chi^2 = 100$. For this reason, we perform the CT14 reweighting in the quadratic–quadratic approximation, noting that the reweighted uncertainties might not be directly comparable with the original ones, and that the new central set underestimates the true impact on CT14, as the use of $\Delta\chi^2 = 100$ overestimates the growth of χ^2_{old} in varying the PDF parameters.

The resulting reweighted PDFs are compared with the original CT14 NLO PDFs in Fig. 3. For all quark flavours, the found modifications are modest compared to the size of PDF uncertainties. Only at very large x we can see a clear downward bend in the central valence-quark PDFs, caused by the fit trying to adapt to the data at large rapidities, where gluon–valence-quark scattering dominates the cross sections. There is a similar, but even more pronounced, large-x depletion for the gluons. In addition, we find an enhancement for gluons at $x \sim 0.1$, compensating for the excess in data at midrapidity. Such modifications to gluon PDF are not totally unexpected. The MMHT14 gluon PDF [6], which closely resembles that of CT14, acquires rather similar modifications when confronted with the 7 TeV high-luminosity inclusive jet data [30]. Also, attributed to including 8 TeV differential top-quark data, the NNPDF3.1 fit has large-x gluons suppressed compared to CT14 and MMHT14 [31]. In addition, a recent reweighting study using multiple top-production data sets found very similar CT14 modifications as we do here [32]. Thus, we have evidence that the CT14 gluon distribution is simply too hard to be able to fully describe jet and top-quark measurements.

Figure 4 shows the reweighted dijet spectra in comparison to data and original CT14 predictions. The reweighting clearly improves compatibility with the data, especially in the midrapidity region, where the data and theory are now in agreement within the associated uncertainties. At $\eta_{\mathrm{dijet}} \lesssim -1$, the data still deviates from the reweighted results. This is also reflected in the figure of merit, $\chi^2/N_{\mathrm{data}} = 2.0$, which is still quite high, but vastly better than before the reweighting. For a comparison, we have calculated the dijet

[4] The CT14 analysis uses the individual-jet p_{T} as the scale for the inclusive-jet cross sections. To LO, $p_{\mathrm{T}}^{\mathrm{leading}} = p_{\mathrm{T}}^{\mathrm{subleading}}$, and thus using $\mu = p_{\mathrm{T}}^{\mathrm{ave}}$ for dijets recovers the CT14 scale definition in the $2 \to 2$ limit.

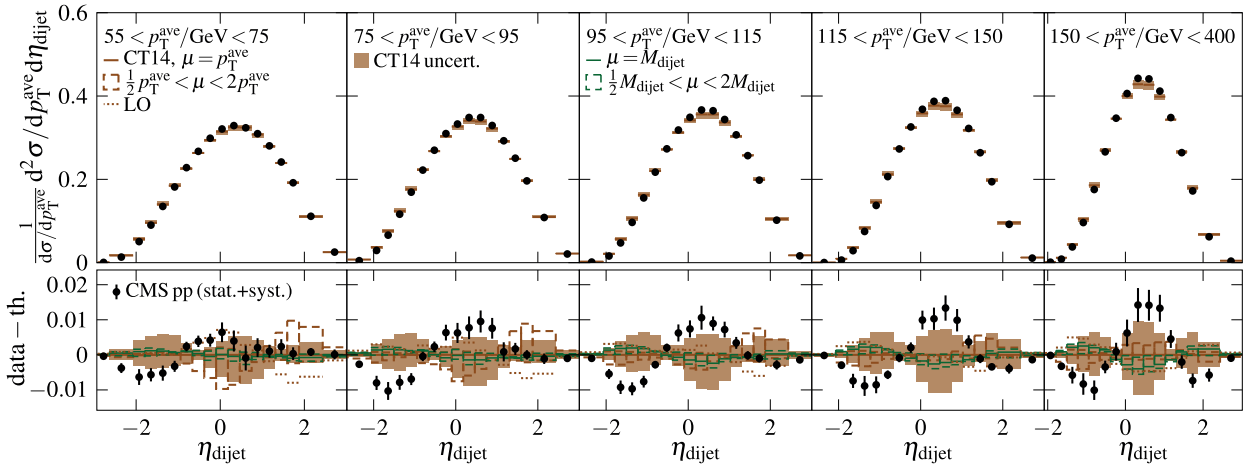

Fig. 2 Upper panels: distributions of dijets in 5.02 TeV proton–proton collisions against η_{dijet} and normalized to unity in each bin of $p_{\text{T}}^{\text{ave}}$. The imposed kinematic cuts are discussed in text. Black markers show the data from the CMS measurement [4] with vertical bars showing the statistical and systematical uncertainties added in quadrature. Solid orange lines represent the results from the NLO pQCD calculation using the central set of the CT14 NLO PDFs [5] with $\mu = p_{\text{T}}^{\text{ave}}$ scale choice, light orange boxes the associated PDF uncertainties from the CT14 NLO error sets. Lower panels: difference to the central CT14 result. Dashed hollow boxes show the dependence of NLO predictions on factor two upward and downward variations of the scale choice. Dotted lines represent the results from the respective LO pQCD calculation. The results with $\mu = M_{\text{dijet}}$ scale choice and its factor two variations are indicated in green

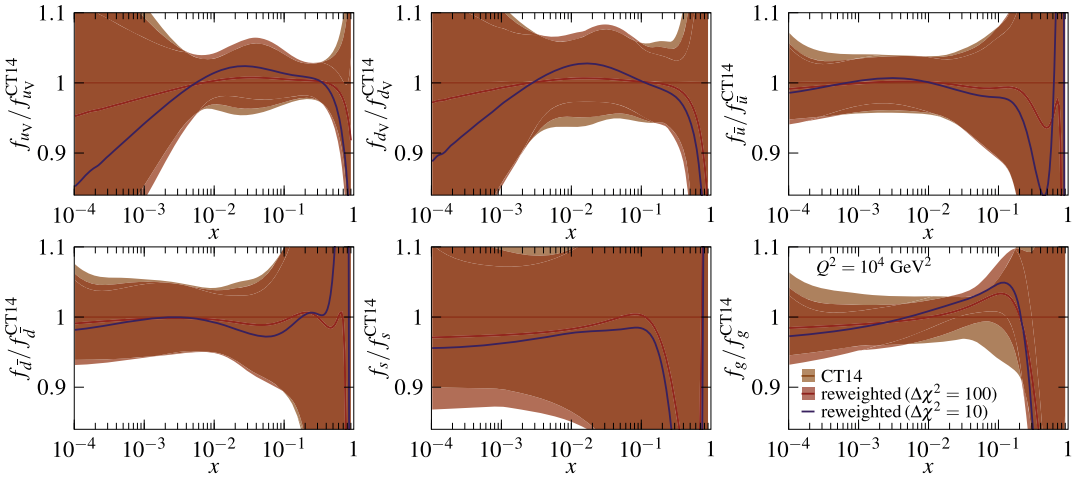

Fig. 3 The impact of reweighting on CT14 NLO PDFs at $Q^2 = 10^4$ GeV^2. The original CT14 PDFs are shown in orange, with the solid line representing the central set PDFs, the ratio to which is shown in each panel. The corresponding PDFs obtained with quadratic–quadratic reweighting using $\Delta\chi^2 = 100$ are shown in red and the central set of the reweighting with $\Delta\chi^2 = 10$ is presented with a solid purple line

spectra also using the MMHT14 [6], NNPDF3.1 [31] and 5-flavour ABMP16 [33] NLO PDFs. These yield χ^2/N_{data} goodness-of-fit values 4.7, 4.0 and 2.7, respectively, showing that less than perfect agreement with the data is not only a problem with CT14. However, the very strong disagreement between data and CT14 before reweighting appears to be a rather extreme case. In Fig. 5 the gluon PDFs of MMHT14, NNPDF3.1 and ABMP16 are compared with the CT14 before and after the reweighting. The reweighting brings the CT14 gluon distribution to a closer agreement with the other PDFs, particularly at small x to the MMHT14 and NNPDF3.1 and, more importantly, at large x to the NNPDF3.1 and ABMP16. Clearly a reduction in high-x gluons compared to CT14 similar to those in the NNPDF3.1 and ABMP16 fits is preferred by the data.

The penalty term for the reweighted CT14 fit is rather high, with $P/\Delta\chi^2 = 1.17$, clearly indicating that we are reaching the limits of the applicability of the reweighting

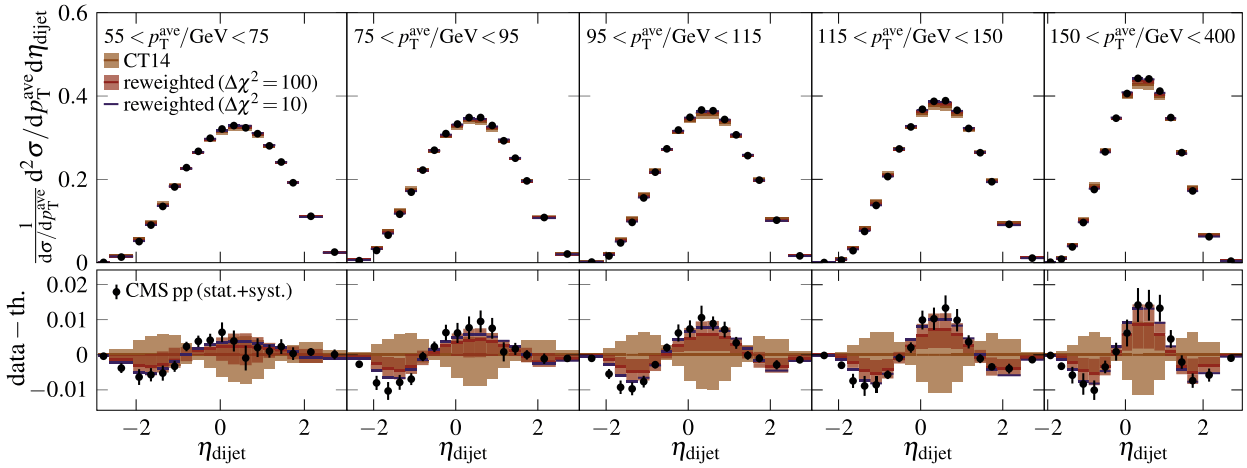

Fig. 4 Upper panels: the impact of reweighting on CT14 predictions of pp dijet spectra. The original predictions are shown in orange and the results obtained with quadratic–quadratic reweighting using $\Delta\chi^2 = 100$ are shown in red. In both cases the solid lines corresponding to the central set and the shaded boxes showing the PDF uncertainty. In addition, resulting spectra from reweighting with $\Delta\chi^2 = 10$ are shown as purple lines. Lower panels show again the difference to the original central CT14 results

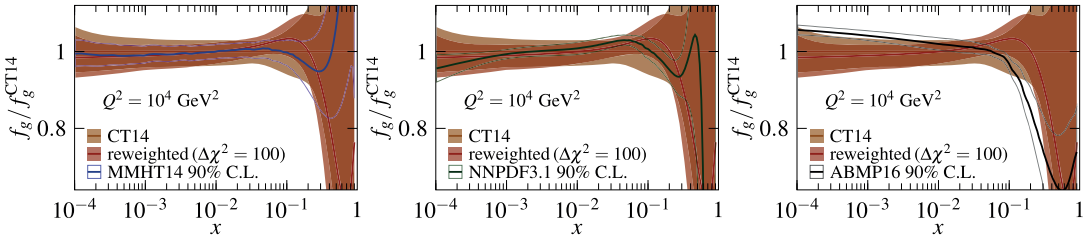

Fig. 5 Comparison of the NLO gluon PDFs of the original and reweighted CT14 sets with those from the MMHT14, NNPDF3.1 and 5-flavour ABMP16 analyses. The uncertainty bands of the latter have been scaled with a factor 1.64 to nominally match with the 90% confidence level definition of the CT14 analysis

method. This can be interpreted either as a tension between the dijet data and some datasets used in the CT14 analysis, or as an inflexibility of the CT14 fit form in the high-x region which is probed by the dijets at large rapidities, where the data were not well reproduced and where the data would support even stronger suppression in the PDFs. To test if the CT14 parametrization could adapt to the dijet data, we have performed a reweighting also with an artificially low $\Delta\chi^2 = 10$. In a global fit, this would translate to putting an additional tenfold weight on the new data. The results for the new central PDF set are shown as purple lines in Figs. 3 and 4. With stronger low- and high-x suppression and mid-x enhancement for gluons, this fit achieves a much more reasonable goodness-of-fit $\chi^2/N_{data} = 0.9$ for these data. For this, substantial help from valence quarks, which get strong modifications in this case, is also needed. Still, the data at $\eta_{dijet} \lesssim -1$ are not perfectly reproduced, which might be a signal of a parametrization issue, as the relative contribution from the original fit to the total χ^2 is decreased with the lowered $\Delta\chi^2$. With $P/\Delta\chi^2 = 3.61$, this fit is in a clear tension with the original CT14 analysis. Of course, once the correlations in the dijet data are made available, one should study whether a shift in some of the systematic parameters could improve the fit at $\eta_{dijet} \lesssim -1$. It is also conceivable that the residual disagreement is due to the NNLO corrections.

A comprehensive study of possibly conflicting datasets within CT14 is outside the scope of this article, but as a cross check we have tested the compatibility of the reweighted PDFs with the CMS 7 TeV inclusive jet measurements [34] which are included in the CT14 analysis. For these calculations we use the pre-computed fastNLO grids [35], setting the renormalization and factorization scales equal to the transverse momentum p_T of the individual jet as in the CT14 analysis. Figure 6 shows the data-to-theory ratio for the NLO predictions with the CT14 PDFs reweighted with the dijet data using $\Delta\chi^2 = 100$. Also the ratios of the original CT14 central predictions with the reweighted ones are indicated. The data-to-theory agreement happens to be even slightly

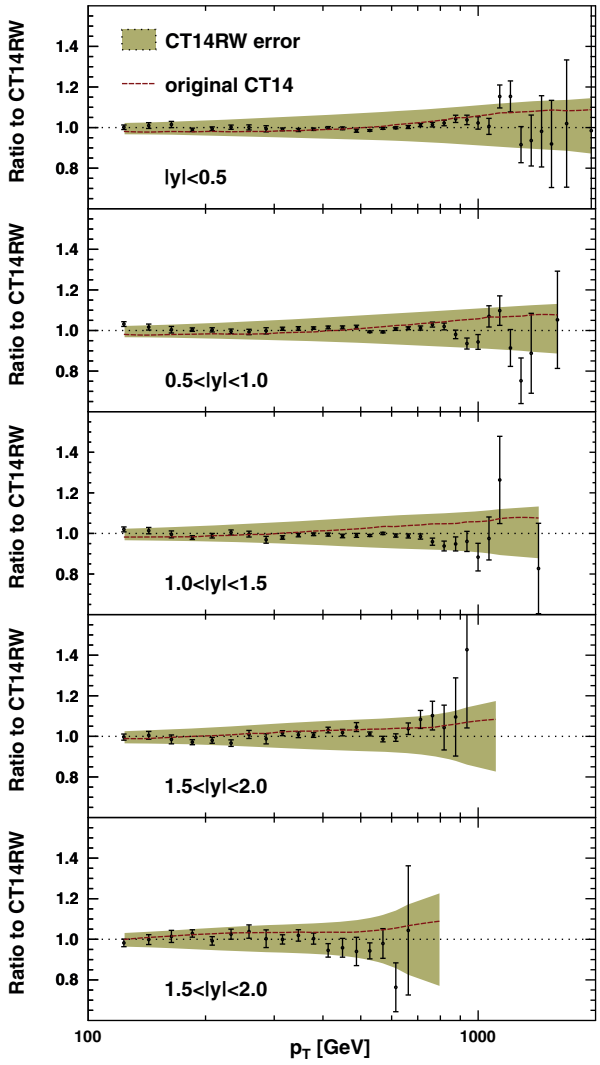

Fig. 6 Comparison of CMS 7 TeV inclusive jet measurements [34] and NLO predictions obtained using the CT14 NLO PDFs [5] reweighted with the 5.02 TeV dijet data [4]. The optimal systematic shifts in the correlated experimental uncertainties are applied to the data points (similarly as in Ref. [15]) and only statistical uncertainties are shown. Dashed red lines show the ratio of predictions with the original CT14 PDFs to those with the reweighted PDFs

better for the reweighted PDFs, with $\chi^2/N_{\text{data}} = 1.2$, than for the original set, for which $\chi^2/N_{\text{data}} = 1.3$. Thus we find that, in the light of reweighting, the CMS measurements of inclusive jets at 7 TeV and dijets at 5.02 TeV are mutually compatible.

3.2 Significance of proton PDF uncertainties in proton–lead dijet spectra

The pPb dijet spectra, shown in Fig. 7, have a rather similar data-to-theory systematics as we had in the pp case. Here, we use the EPPS16 nuclear modifications along with the CT14 NLO proton PDFs in the predictions, i.e. the PDF of a flavour i in a proton bound in lead at scale Q^2 is obtained with

$$f_i^{\text{p/Pb}}(x, Q^2) = R_i^{\text{Pb}}(x, Q^2) f_i^{\text{p}}(x, Q^2), \tag{33}$$

where R_i^{Pb} is the nuclear modification from the EPPS16 analysis and f_i^{p} the corresponding CT14 PDF of the free proton. The total PDF uncertainties in the cross sections are calculated with

$$\delta X^{\pm}_{\text{total}} = \sqrt{\left(\delta X^{\pm}_{\text{EPPS16}}\right)^2 + \left(\delta X^{\pm}_{\text{CT14}}\right)^2}, \tag{34}$$

where $\delta X^{\pm}_{\text{EPPS16}}$ are the upward and downward uncertainties obtained with Eq. (11) using the EPPS16 error sets and keeping the CT14 central set fixed, and $\delta X^{\pm}_{\text{CT14}}$, respectively, the uncertainties from the CT14 error sets keeping the EPPS16 central set fixed.

Again, these predictions give wider distributions than seen in the CMS data, resulting with $\chi^2/N_{\text{data}} = 6.9$. While in this case the data points are mostly within the combined nuclear and free-proton PDF uncertainty bands, we can expect that the modifications to the CT14 PDFs, which were found necessary to improve the description of the pp data, play a role also here. Indeed, in Fig. 8 we show results with the PDFs obtained by reweighting CT14 with the pp data, observing a clear improvement in the data to theory agreement. We obtain $\chi^2/N_{\text{data}} = 2.8$ for the predictions with CT14 reweighted using $\Delta\chi^2 = 100$ and $\chi^2/N_{\text{data}} = 1.6$ when using $\Delta\chi^2 = 10$. These numbers are somewhat higher than what we obtained in the pp case, reflecting the fact that also the EPPS16 nuclear modifications need to be adjusted for optimal description of the data. This can also be seen by comparing the data-to-theory agreement in pPb at $\eta_{\text{dijet}} \gtrsim 2$ to that in pp: while the CT14 predictions reweighted using $\Delta\chi^2 = 100$ describe well the pp data in these rapidities, the pPb data points lie systematically below the predictions, which hints a preference for deeper nuclear shadowing – the suppression in the gluon PDF, $R_g^{\text{Pb}} < 1$, at small x – than that in the EPPS16 central set. We will verify this claim in the next section.

An important thing to notice here is that most of the deviations from central theory predictions actually originate from the issues with the free-proton PDFs instead of the nuclear modifications. This large free-proton PDF bias prevents a clean extraction of the PDF nuclear modifications from the pPb spectra. The dijet spectra are certainly not the only pPb observable sensitive to such a free-proton PDF dependence, but the refined proton PDFs found here could also have an effect for example on the predictions for inclusive $t\bar{t}$ production at 8.16 TeV pPb collisions where calculations with

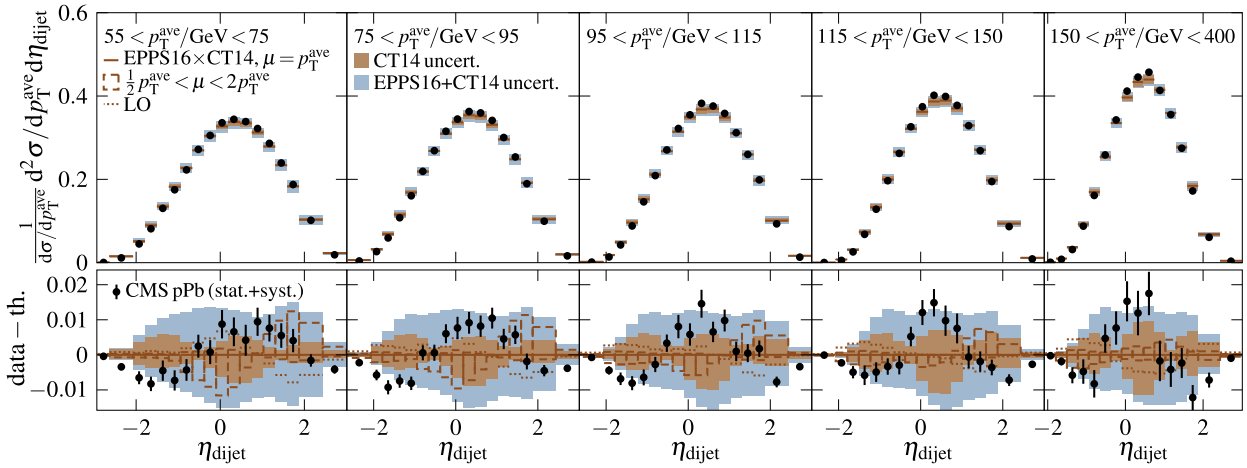

Fig. 7 As Fig. 2, but now with pPb data and predictions with EPPS16 nuclear modifications imposed on the CT14 NLO proton PDFs and omitting the results with $\mu = M_{dijet}$ for clarity. Light blue boxes show the combined uncertainty from the CT14 and EPPS16 PDFs

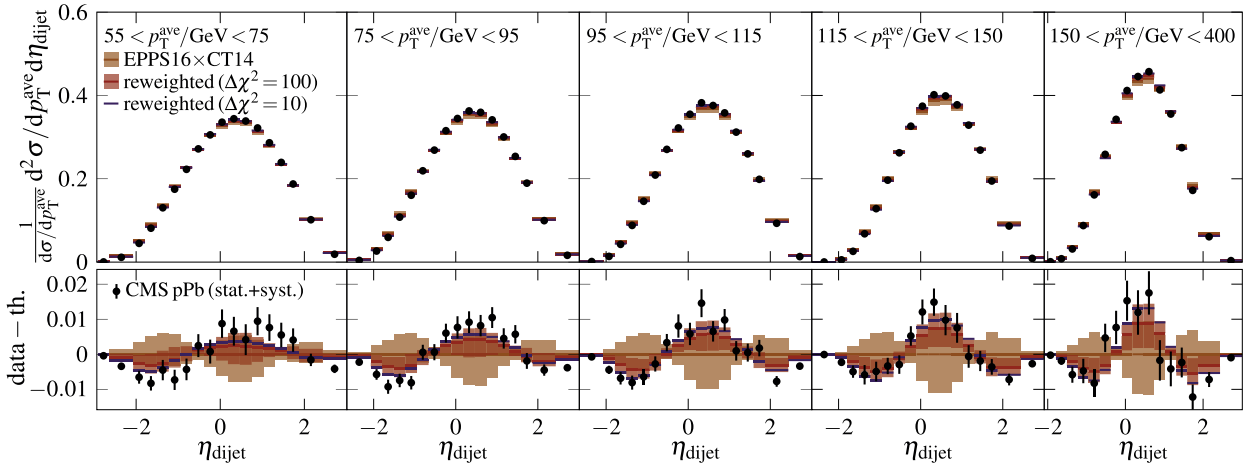

Fig. 8 As Fig. 4, but now with pPb data and with EPPS16 nuclear modifications imposed on the original and reweighted CT14 PDFs. Only uncertainties from the free-proton PDFs are shown

CT14+EPPS16 overshoot, but are still compatible with the data [36].

3.3 Nuclear modification ratio and EPPS16 reweighting

Let us now consider the nuclear modification ratio of the normalized dijet spectra discussed above, defined as

$$R_{\mathrm{pPb}}^{\mathrm{norm.}} = \frac{\frac{1}{\mathrm{d}\sigma^{\mathrm{pPb}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}}\,\mathrm{d}^2\sigma^{\mathrm{pPb}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}\mathrm{d}\eta_{\mathrm{dijet}}}{\frac{1}{\mathrm{d}\sigma^{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}}\,\mathrm{d}^2\sigma^{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ave}}\mathrm{d}\eta_{\mathrm{dijet}}}. \tag{35}$$

As we have seen that the dijet rapidity distributions in pp and pPb have very similar dependence on the free proton PDFs, we can expect this dependence to efficiently cancel in the ratio. This statement is verified in Fig. 9, where we observe the uncertainty band given by CT14 PDFs to be vanishingly small. Also the scale uncertainties, while being larger than the CT14 uncertainties, are small in this observable, implying that MHOUs can be expected to be small as well. This leaves the nuclear modifications as the dominant source of theory uncertainty.

We observe that the CMS data and EPPS16 predictions are in good agreement within the uncertainties. This does not come as a surprise, as part of these data, namely the high-p_T^{ave} part of the pPb cross section [37], were used in the EPPS16 fit. Still, this agreement is not trivial as with the new pp baseline and being a more differential measurement, these $R_{\mathrm{pPb}}^{\mathrm{norm.}}$ data contain plenty of new information compared to the 7 data points of forward-to-backward ratios included in the EPPS16 analysis. As was anticipated above, the data points at forward rapidities deviate from the cen-

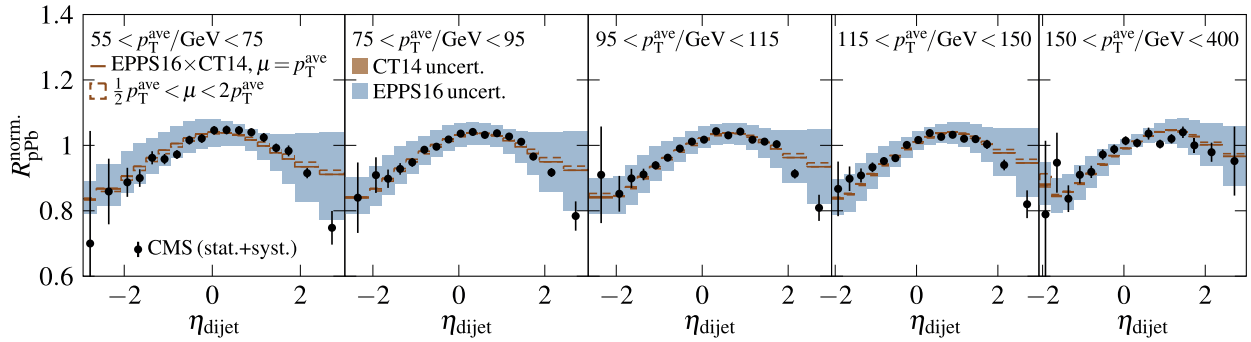

Fig. 9 The nuclear modification ratio of normalized pPb and pp differential cross sections. Black markers show the data from CMS measurement [4] with vertical bars showing the statistical and systematical uncertainties added in quadrature. Solid orange lines represent the NLO pQCD calculation with $\mu = p_T^{ave}$ scale choice using the central set of the CT14 NLO PDFs [5] with EPPS16 [18] nuclear modifications

tral EPPS16 prediction, indicating a preference for a deeper shadowing in the nPDFs.

Compared to the data, the EPPS16 predictions have much larger uncertainties, which promises a good constraining power when fitting to these data. To study the impact these data would have had in the EPPS16 fit, we have performed a reweighting in the cubic–quadratic approximation introduced in Sect. 2.2, using $\Delta\chi^2 = 52$ and taking the values of $\delta z_k^{\pm}$ from Table 2 of Ref. [18]. The results for $R_{pPb}^{norm.}$ are shown in Fig. 10. Most notably, there is a vast reduction in the EPPS16 uncertainties. Also, at forward rapidities the central prediction comes down a bit, as is expected from the low-lying data points in this region. In the backward direction a slight enhancement in the central prediction can be observed, but this is far less prominent than the suppression in the forward bins. In total, we obtain an improvement in the goodness of fit from $\chi^2/N_{data} = 1.7$ to 1.4 with a penalty $P/\Delta\chi^2 = 0.14$.

The corresponding effects on the EPPS16 nuclear modifications in lead at the parametrization scale $Q^2 = 1.69\ \text{GeV}^2$ are presented in Fig. 11. There is a striking impact on gluon modification uncertainties, which are reduced across all x. In the best-constrained mid-x region, the uncertainties are reduced to less than half of their original size. As the uncertainty band lies clearly above unity in this region, we find strong evidence for gluon antishadowing in lead. At small x, the reweighted uncertainty band goes respectively below unity, giving evidence for gluon shadowing. These findings are in accordance with those of Ref. [38], where inclusive heavy-flavour production data from measurements at the LHC were used to study the gluon PDF modifications in nuclei. As expected from inspecting the ratio of the dijet spectra, the new central set seems to support stronger shadowing than in the original EPPS16 central fit.

Even with the increased gluon shadowing, the most forward bins of $R_{pPb}^{norm.}$ are not well reproduced by the reweighted results, which is also the reason why the χ^2/N_{data} remained somewhat high even after the reweighting. To be consistent with these forward data points, a very deep shadowing for the gluons would be required. Moreover, the probed x region

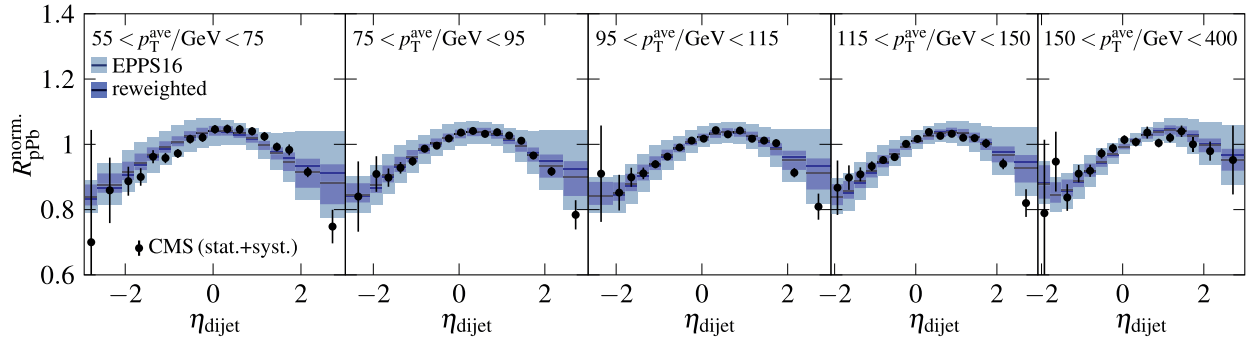

Fig. 10 The impact of reweighting on EPPS16 predictions of the nuclear modification ratio of the dijet spectra. The original predictions are shown with solid blue lines and light blue boxes representing the central predictions and the nPDF uncertainties, respectively. The corresponding results after the reweighting are shown with solid black lines and purple boxes

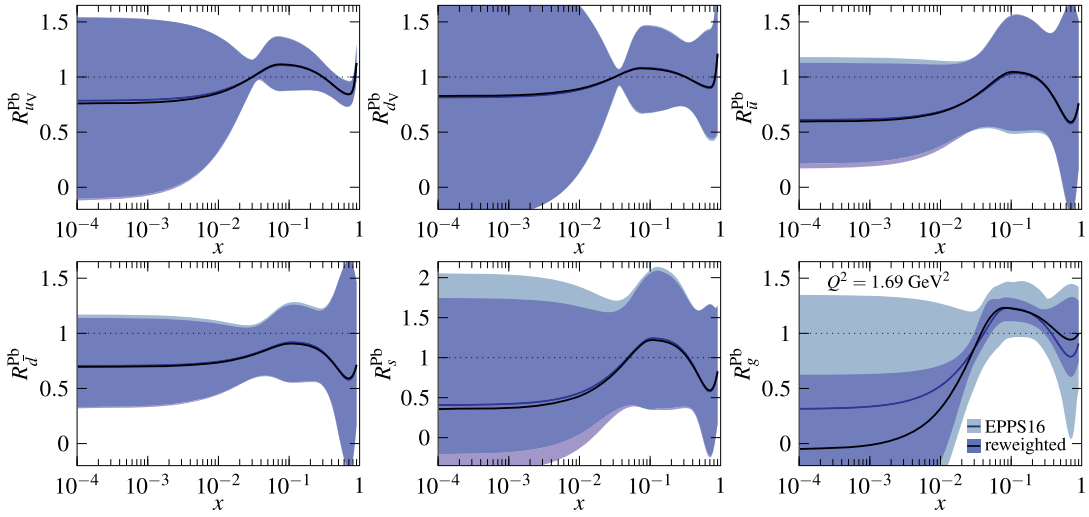

Fig. 11 The impact of reweighting the EPPS16 nPDFs with the data on the nuclear modification ratio of the dijet spectra. The original and reweighted EPPS16 nuclear modifications for the lead nucleus are presented at the parametrization scale $Q^2 = 1.69\ \mathrm{GeV}^2$. For better visibility, the s-quark modifications are presented with a different vertical axis scaling

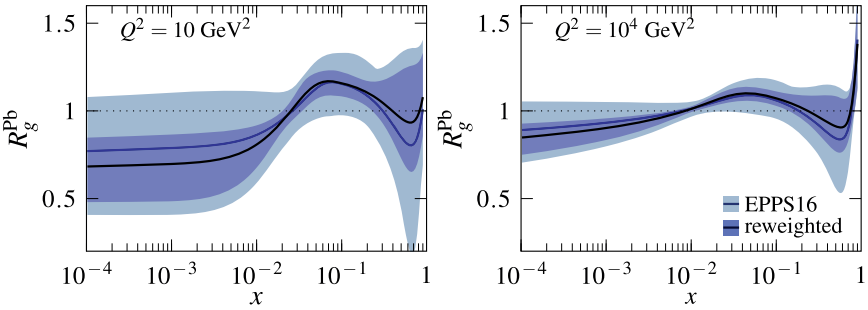

Fig. 12 The EPPS16 gluon nuclear modifications in Pb at the scales $Q^2 = 10\ \mathrm{GeV}^2$ and $Q^2 = 10^4\ \mathrm{GeV}^2$ before and after reweighting with the dijet data

changes very little between the last and second-to-last η_{dijet} data point, and thus such a steep drop as that suggested by the data is difficult to attain. This is because the DGLAP evolution efficiently smooths out even steep structures in the gluon nuclear modification, as can be seen in Fig. 12 where we show the gluon nuclear modifications evolved to higher scales. We also note that the systematic uncertainty dominates in the last η_{dijet} bins, and thus taking into account the data correlations, once available, could improve the fit quality. These findings should, in the future, be contrasted also with the recent ATLAS conditional yield measurement, where an order of 10–20% nuclear suppression for dijets was found in the most forward configuration [39].

Also at large x, the reweighted gluon modifications are better constrained than in the original EPPS16 analysis. The new central set has R^{Pb}_g closer to unity at x around 0.7. This is partly enforced by momentum sum rule in combination with the stiffness of the EPPS16 fit function and the deepened small-x shadowing. In any case, the uncertainty remains large, and either an enhancement or a suppression for gluons is possible in this region. On this basis, the conclusion made in Ref. [4], that the dijet data would give evidence of large-x gluon suppression, seems premature. This claim was based on comparison of the data with EPS09 [40] and DSSZ [41] nPDFs, where the former, with gluon suppression at large x, agreed well with the data at backward rapidities, but the latter, having the nuclear gluons unmodified, did not. However, going towards backward rapidities, and thus larger x from the Pb side, the contribution of nuclear quarks to the dijet cross section grows rapidly. Hence the difference in predictions with EPS09 and DSSZ in this region has a large contribution from different valence quark modifications. As DSSZ has much smaller large-x suppression for valence quarks than EPS09 (see e.g. Ref. [42]), this also partly explains the difference in the dijet predictions of Ref. [4].

On these grounds, it might appear surprising that the dijet data are not able to constrain the valence quark modifications at all, as can be seen from the first two panels in Fig. 11. The reason for this is that due to smallness of isospin corrections [43], the backward dijet data mainly probe the average valence modifications,

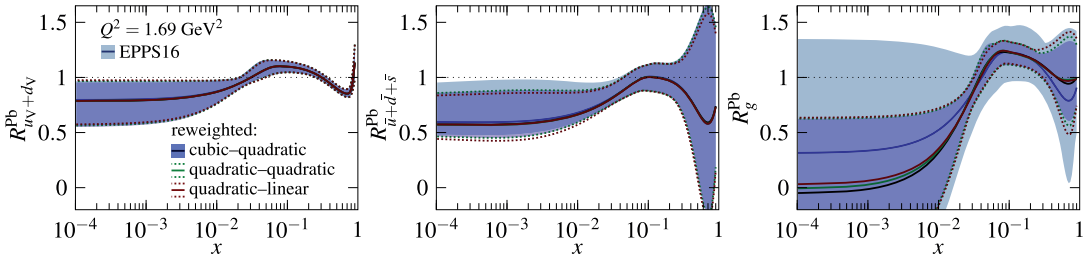

Fig. 13 The impact on the average valence and sea quark and gluon modifications under different approximations in the reweighting

$$R^{\mathrm{Pb}}_{u_\mathrm{V}+d_\mathrm{V}} = \frac{u_\mathrm{V}^{\mathrm{p/Pb}} + d_\mathrm{V}^{\mathrm{p/Pb}}}{u_\mathrm{V}^{\mathrm{p}} + d_\mathrm{V}^{\mathrm{p}}}, \tag{36}$$

shown in Fig. 13. This combination is much better constrained than the individual flavours shown in Fig. 11 and has vastly smaller uncertainties at large x than the gluon modifications. Thus, while large-x valence quarks dominate the dijet cross section at backward rapidities, the uncertainty in the EPPS16 predictions in this region comes dominantly from the less-constrained gluons, and hence it is the gluon modifications which are constrained in the reweighting. Figure 13 shows also the average sea quark modification

$$R^{\mathrm{Pb}}_{\bar{u}+\bar{d}+\bar{s}} = \frac{\bar{u}^{\mathrm{p/Pb}} + \bar{d}^{\mathrm{p/Pb}} + \bar{s}^{\mathrm{p/Pb}}}{\bar{u}^{\mathrm{p}} + \bar{d}^{\mathrm{p}} + \bar{s}^{\mathrm{p}}}, \tag{37}$$

which is the dominant quark combination constrained at forward rapidities. We observe a modest reduction in the small-x uncertainty, much smaller than that for the gluons. At the level of individual flavours, shown in Fig. 11, these constraints affect mostly the s-quark modifications, which were poorly constrained in EPPS16.

3.4 Importance of non-quadratic and non-linear terms in reweighting

We may now ask whether the inclusion of higher-order (non-quadratic and non-linear) components in the reweighting had a sizable effect on our results. Figure 13 shows the impact of the dijet data on the EPPS16 nuclear modifications in all three approximations discussed in Sect. 2.2. While, for simplicity of presentation, we show only the average valence and light-sea-quark modifications in addition to those for gluons, the conclusions below apply to individual flavours as well. We find that the cubic–quadratic and quadratic–quadratic approximations give almost identical results. This is rather easy to understand: The new data are precise enough to dominate the shape of the total χ^2 function in the parameter directions that it constrains (mainly those related to gluon degrees of freedom), making the non-quadratic components sub-dominant in the reweighting. Moreover, as the new central set does not divert far from the original, we are working in a region where the quadratic approximation for χ^2_{old} is rather good. Under different circumstances this might not be the case and the cubic terms could alter the reweighting results significantly.

Next, we consider the reweighting results in the quadratic–linear approximation. Here, we use the linear approximation for the cross sections, but decide to keep the quadratic dependence in the PDFs for better comparability.[5] Again, the differences to the results of the cubic–quadratic approximation are rather modest, though for the high-x gluons the quadratic–linear approximation seems to suggest slightly less stringent constraints. The similarity of results in the different approximations can also be seen as a reassuring fact: the results of reweighting do not seem to depend on minute details of our method and we seem to be able to make reliable conclusions based on rather limited information about the original global analysis, at least in this particular case. The obtained results are thus not likely to change if even higher-order contributions are added.

4 Summary and conclusions

In this work, we have presented a non-quadratic extension of the Hessian PDF reweighting introduced in Ref. [15] and applied the method in the context of CMS dijet measurements at 5.02 TeV. This improved method makes use of the knowledge of parameter variations at which the error sets of the original PDFs are defined, to solve for cubic components of the χ^2 function before inclusion of new data. Similarly, quadratic components in the responses of observables to parameter variations were taken into account. The additional information needed in this cubic–quadratic approximation prevented us from using it when reweighting the CT14 NLO PDFs with the pp dijet distributions, where we had to resort to a simpler quadratic–quadratic approximation, but we were able to apply it to reweight the EPPS16 nPDFs, for

[5] Note that using a linear parameter dependence for the PDFs would render the PDF uncertainties to be perfectly symmetric, so that the comparison with cubic–quadratic and quadratic–quadratic approximations would be meaningful only under the symmetric prescription of Eq. (10).

which the needed information is available, with the nuclear modification ratio of the dijet spectra. While no large differences were found in the results of reweighting EPPS16 in the cubic–quadratic or quadratic–quadratic approximation, this observation was limited to one specific case, and under different circumstances the cubic terms could become more important. We thus encourage PDF fitters to publish the details of their analysis to a sufficient accuracy, such that the reweighting including the higher-order terms becomes possible. This can be done by publishing the numerical values of the $\delta z_k^{\pm}$ parameters as defined in Sect. 2 in addition to the tolerance $\Delta\chi^2$. Care must be taken in communicating which error set corresponds to each of these values, so that there is no chance of misinterpretation e.g. in what is called a "plus" and what a "minus" direction. A neat way to do this with LHAPDF [44] would be to set in each PDF grid file a custom flag such as "`ParamVal`" to hold the value $\delta z_k^{\pm}$. These parameter values could then be retrieved by using the method `info().get_entry("ParamVal")` for each of the PDF error sets.

Comparing the measured pp dijet pseudorapidity spectra with theory calculations using the CT14 NLO PDFs revealed a large discrepancy. We showed that at high $p_{\mathrm{T}}^{\mathrm{ave}}$ this difference is larger than the associated scale uncertainties and exceeds the size of the NLO corrections, thus being unlikely due to missing NNLO terms alone. This suggested the need for modifying the CT14 PDFs to reach a better agreement with the data. In reweighting CT14 with the dijet data, the gluon PDF acquired significant modifications, especially at large x, where a substantial reduction was observed. We discussed also evidence from other studies pointing into the same direction. After reweighting, a much more reasonable χ^2 value for the dijet data was found, but this came with a price of a rather high penalty term, i.e. the new central set had diverted quite far from the original minimum. The reason for this apparent discrepancy between CT14 and the dijet data remains elusive. We tested the reweighted PDFs against CMS 7 TeV inclusive jet measurements finding good agreement, and thus no conflict between the considered dijet and inclusive jet data. By performing a reweighting with an artificially low $\Delta\chi^2$, we showed that the CT14 PDFs still had trouble in reproducing the data at $\eta_{\mathrm{dijet}} \lesssim -1$, signaling a possible parametrization issue, although NNLO corrections and correlated systematics can also play a role here. Solving this issue is beyond the reach of the reweighting tools and should be studied in the context of a global analysis.

Similar discrepancy as seen with the pp spectra is observed also in the case of pPb. We showed that applying the same CT14 modifications as found in the reweighting with pp data substantially improves the data-to-theory agreement also in pPb. As the pPb dijet distributions contain a substantial free-proton PDF dependence, a clean extraction of their nuclear modifications is not possible from these data directly. Taking the ratio of the pPb and pp spectra, however, leads to a very efficient cancellation of not only the free-proton uncertainties but also of the scale uncertainties, thus giving an excellent probe of the nPDFs. We showed that the measured nuclear-modification ratio of dijet spectra is in a good agreement with the NLO predictions using the EPPS16 nPDFs. Some deviation from the EPPS16 central prediction was observed at $\eta_{\mathrm{dijet}} \gtrsim 2$, supporting a stronger shadowing for gluons than present in the EPPS16 central set. As a whole, these data give compelling evidence of small-x gluon nuclear shadowing and mid-x antishadowing, as was revealed in reweighting the EPPS16 nPDFs. We obtained significant new constraints on the EPPS16 gluon modifications in lead throughout the probed range, reducing the uncertainties even to less than half of their original size.

Acknowledgements We thank Yen-Jie Lee for discussions. We have received funding from the Academy of Finland, Project 297058 of K.J.E. and 308301 of H.P.; P.P. acknowledges the financial support from the Magnus Ehrnrooth Foundation. We thank the Finnish IT Center for Science (CSC) for the computational resources allocated under the Project jyy2580.

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

Funded by SCOAP[3].

References

1. J. Gao, L. Harland-Lang, J. Rojo, Phys. Rep. **742**, 1 (2018). arXiv:1709.04922 [hep-ph]
2. M. Mangano, H. Satz, U. Wiedemann (eds.), Hard Probes in Heavy-Ion Collisions at the LHC. CERN Yellow Report, CERN-2004-009
3. H.W. Lin et al., Prog. Part. Nucl. Phys. **100**, 107 (2018). arXiv:1711.07916 [hep-ph]
4. A.M. Sirunyan et al. [CMS Collaboration], Phys. Rev. Lett. **121**(6), 062002 (2018). arXiv:1805.04736 [hep-ex]
5. S. Dulat et al., Phys. Rev. D **93**(3), 033006 (2016). arXiv:1506.07443 [hep-ph]
6. L.A. Harland-Lang, A.D. Martin, P. Motylinski, R.S. Thorne, Eur. Phys. J. C **75**(5), 204 (2015). arXiv:1412.3989 [hep-ph]
7. V. Bertone et al., xFitter Developers' Team, PoS DIS **2017**, 203 (2018). arXiv:1709.01151 [hep-ph]
8. S. Alekhin et al., Eur. Phys. J. C **75**(7), 304 (2015). arXiv:1410.4412 [hep-ph]
9. W.T. Giele, S. Keller, Phys. Rev. D **58**, 094023 (1998). arXiv:hep-ph/9803393
10. R.D. Ball et al. [NNPDF Collaboration], Nucl. Phys. B **849**, 112 (2011). Erratum: [Nucl. Phys. B **854**, 926 (2012)]. Erratum: [Nucl. Phys. B **855**, 927 (2012)]. arXiv:1012.0836 [hep-ph]

11. R.D. Ball et al., Nucl. Phys. B **855**, 608 (2012). arXiv:1108.1758 [hep-ph]
12. G. Watt, R.S. Thorne, JHEP **1208**, 052 (2012). arXiv:1205.4024 [hep-ph]
13. N. Sato, J.F. Owens, H. Prosper, Phys. Rev. D **89**(11), 114020 (2014). arXiv:1310.1089 [hep-ph]
14. H. Paukkunen, C.A. Salgado, Phys. Rev. Lett. **110**(21), 212301 (2013). arXiv:1302.2001 [hep-ph]
15. H. Paukkunen, P. Zurita, JHEP **1412**, 100 (2014). arXiv:1402.6623 [hep-ph]
16. C. Schmidt, J. Pumplin, C.P. Yuan, P. Yuan, Phys. Rev. D **98**(9), 094005 (2018). arXiv:1806.07950 [hep-ph]
17. A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189 (2009). arXiv:0901.0002 [hep-ph]
18. K.J. Eskola, P. Paakkinen, H. Paukkunen, C.A. Salgado, Eur. Phys. J. C **77**(3), 163 (2017). arXiv:1612.05741 [hep-ph]
19. K.J. Eskola, P. Paakkinen, H. Paukkunen, PoS DIS **2018**, 014 (2018). arXiv:1806.08208 [hep-ph]
20. K.J. Eskola, P. Paakkinen, H. Paukkunen, PoS HardProbes **2018**, 124 (2019). arXiv:1812.05438 [hep-ph]
21. J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.L. Lai, W.K. Tung, Phys. Rev. D **65**, 014013 (2001). arXiv:hep-ph/0101032
22. J. Pumplin, Phys. Rev. D **82**, 114020 (2010). arXiv:0909.5176 [hep-ph]
23. H.L. Lai, M. Guzzi, J. Huston, Z. Li, P.M. Nadolsky, J. Pumplin, C.-P. Yuan, Phys. Rev. D **82**, 074024 (2010). arXiv:1007.2241 [hep-ph]
24. F. James, M. Roos, Comput. Phys. Commun. **10**, 343 (1975)
25. M. Cacciari, G.P. Salam, G. Soyez, JHEP **0804**, 063 (2008). arXiv:0802.1189 [hep-ph]
26. D. Gulhan, talk at HI Jet Workshop, Ecole Polytechnique, Paris, July (2016). https://indico.cern.ch/event/507670/contributions/2218034/attachments/1314800/1969379/2016_07_24_HIJetWorkshop.pdf. Accessed 22 Mar 2019
27. J. Currie, A. Ridder, T. Gehrmann-De Gehrmann, E.W.N. Glover, A. Huss, J. Pires, Phys. Rev. Lett. **119**(15), 152001 (2017). arXiv:1705.10271 [hep-ph]
28. Z. Nagy, Phys. Rev. D **68**, 094002 (2003). arXiv:hep-ph/0307268
29. M. Cacciari, G.P. Salam, G. Soyez, Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097 [hep-ph]
30. L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Eur. Phys. J. C **78**(3), 248 (2018). arXiv:1711.05757 [hep-ph]
31. R.D. Ball et al. [NNPDF Collaboration], Eur. Phys. J. C **77**(10), 663 (2017). arXiv:1706.00428 [hep-ph]
32. M. Azizi, A. Khorramian, H. Abdolmaleki, S. Paktinat Mehdiabadi, Int. J. Mod. Phys. A **33**(24), 1850142 (2018). arXiv:1810.12063 [hep-ph]
33. S. Alekhin, J. Blümlein, S. Moch, Eur. Phys. J. C **78**(6), 477 (2018). arXiv:1803.07537 [hep-ph]
34. S. Chatrchyan et al. [CMS Collaboration], Phys. Rev. D **87**(11), 112002 (2013). Erratum: [Phys. Rev. D **87**(11), 119902 (2013)]. arXiv:1212.6660 [hep-ex]
35. M. Wobisch et al. [fastNLO Collaboration]. Report number: FERMILAB-PUB-11-418-PPD, arXiv:1109.1310 [hep-ph]
36. A.M. Sirunyan et al. [CMS Collaboration], Phys. Rev. Lett. **119**(24), 242001 (2017). arXiv:1709.07411 [nucl-ex]
37. S. Chatrchyan et al. [CMS Collaboration], Eur. Phys. J. C **74**(7), 2951 (2014). arXiv:1401.4433 [nucl-ex]
38. A. Kusina, J.P. Lansberg, I. Schienbein, H.S. Shao, Phys. Rev. Lett. **121**(5), 052004 (2018). arXiv:1712.07024 [hep-ph]
39. M. Aaboud et al. [ATLAS Collaboration]. Report number: CERN-EP-2018-324, arXiv:1901.10440 [nucl-ex]
40. K.J. Eskola, H. Paukkunen, C.A. Salgado, JHEP **0904**, 065 (2009). arXiv:0902.4154 [hep-ph]
41. D. de Florian, R. Sassot, P. Zurita, M. Stratmann, Phys. Rev. D **85**, 074028 (2012). arXiv:1112.6324 [hep-ph]
42. H. Paukkunen, Nucl. Phys. A **926**, 24 (2014). arXiv:1401.2345 [hep-ph]
43. K.J. Eskola, H. Paukkunen, C.A. Salgado, JHEP **1310**, 213 (2013). arXiv:1308.6733 [hep-ph]
44. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, G. Watt, Eur. Phys. J. C **75**, 132 (2015). arXiv:1412.7420 [hep-ph]

IV

NUCLEAR GLUONS AT RHIC IN A MULTI-OBSERVABLE APPROACH

by

Helenius, I., Lajoie, J., Osborn, J. D., Paakkinen, P. &
Paukkunen, H. 2019

Phys. Rev. D100, no.1 (2019) 014004

arXiv: 1904.09921 [hep-ph]

Nuclear gluons at RHIC in a multiobservable approach

Ilkka Helenius,[1,2,*] John Lajoie,[3,†] Joseph D. Osborn,[4,‡] Petja Paakkinen,[1,2,§] and Hannu Paukkunen[1,2,‖]

[1]*University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland*
[2]*Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland*
[3]*Iowa State University, Department of Physics, Ames, Iowa 50011, USA*
[4]*University of Michigan, Department of Physics, Ann Arbor, Michigan 48109, USA*

(Received 23 April 2019; published 8 July 2019)

We explore the possibility of measuring nuclear gluon distributions at the Relativistic Heavy-Ion Collider (RHIC) with $\sqrt{s} = 200$ GeV proton-nucleus collisions. In addition to measurements at central rapidity, we consider also observables at forward rapidity, consistent with proposed upgrades to the experimental capabilities of STAR and sPHENIX. The processes we consider consist of Drell-Yan dilepton, dijet, and direct photon-jet production. The Drell-Yan process is found to be an efficient probe of gluons at small momentum fractions. In order to fully utilize the potential of Drell-Yan measurements we demonstrate how the overall normalization uncertainty present in the experimental data can be fixed using other experimental observables. An asset of the RHIC collider is its flexibility to run with different ion beams, and we outline how this ability could be taken advantage of to measure the A dependence of gluon distributions for which the current constraints are scarce.

DOI: 10.1103/PhysRevD.100.014004

I. INTRODUCTION

Good control over the partonic structure of protons and heavier nuclei has become an indispensable ingredient in modern particle, heavy-ion, and astroparticle physics. For processes involving a large momentum transfer, $Q \gg \Lambda_{\mathrm{QCD}} \sim 200$ MeV, the nucleon's relevant degrees of freedom can be described by parton distribution functions (PDFs). Despite the progress in theoretical first-principle methods [1], the PDFs are still most reliably determined through a statistical analysis of a global set of experimental data. Along with the precise data from the Large Hadron Collider (LHC), the list of data types that are included in state-of-the-art PDF analyses has grown, now ranging from traditional inclusive deeply inelastic scattering to jet, top-quark and heavy gauge-boson production [2,3]. At this moment, global fits of proton PDFs do not use any data from the Relativistic Heavy-Ion Collider (RHIC), and nuclear-PDF fits [4,5] use only inclusive pion data from RHIC [6,7]. The advantage of the lower center-of-mass (c.m.) energies of RHIC, $\sqrt{s} = 200$ and 500 GeV, is that the underlying event is not as large as it is at the LHC, and thus e.g., jets can be better resolved at lower transverse momenta (p_{T}) [8,9]. These jet measurements are compatible with next-to-leading-order (NLO) perturbative QCD calculations [9–11] down to $p_{\mathrm{T}} \sim 10$ GeV (which is the minimum p_{T} of the measurements), so nothing really forbids using them in PDF analysis. Similarly, low-mass Drell-Yan events can be better resolved from the decays of heavy flavor. In $p + p$ collisions at the higher c.m. energy of $\sqrt{s} = 500$ GeV, measurements of $W^{\pm}$ bosons also become feasible [12]. These measurements provide complementary constraints on the fixed-target measurements [13] of the $\bar{u}/\bar{d}$ ratio.

The current status of the global determination of nuclear PDFs has been recently reviewed e.g., in Refs. [4,5], and the field is rapidly developing. This is mainly driven by the $p + \mathrm{Pb}$ measurements at the LHC, but is also motivated by theoretical advances in upgrading global analyses to the next-to-NLO (NNLO) QCD level. Currently, the most recent global NLO fits are EPPS16 [14], nCTEQ15 [15] and DSSZ12 [16]. Out of these, EPPS16 has the widest data coverage and is currently the only one to use LHC measurements. In the current NNLO-level fits [17,18], the experimental input is restricted to fixed-target data only. In this paper, we report our studies on the future prospects for constraining nuclear PDFs at RHIC, particularly with measurements at central and forward rapidities where forward acceptance corresponds to that proposed for the STAR [19] and sPHENIX experiments [20]. Projections of

*ilkka.m.helenius@jyu.fi
†lajoie@iastate.edu
‡jdosbo@umich.edu
§petja.paakkinen@jyu.fi
‖hannu.paukkunen@jyu.fi

forward direct photon and Drell-Yan measurements at RHIC on nuclear PDFs have been separately considered e.g., in Refs. [19,21,22]. Here, we aim for a more systematic approach by combining multiple observables into a simultaneous analysis and more carefully assessing the experimental normalization uncertainty. We will base our study mainly on the EPPS16 analysis.

II. EXPERIMENTAL DATA PROJECTIONS

Several processes are expected to have an impact on nuclear PDFs at RHIC c.m. energies. Here, we will focus on such double differential measurements which, to leading order, probe PDFs at fixed momentum fractions. In particular, we construct pseudodata projections for the Drell-Yan dilepton, dijet, and direct photon-jet processes, differential in the invariant mass M and rapidity y of the produced pair. While the Drell-Yan process (on fixed target) has been used as a constraint for nuclear PDFs already in the pioneering EKS98 analysis [23], the use of dijets [24,25] has been realized only in the recent EPPS16 fit [14]. Currently, there are no available photon-jet data to be included in the global analyses though the potential of the process has been discussed [26,27]. In principle, these processes individually constrain different quark-gluon combinations of PDFs, and can also be used together to limit the effect of normalization uncertainties as will be described later.

To generate our projections, we first impose fiducial acceptance requirements on a barrel detector with forward instrumentation. Uncertainty projections were generated for a barrel covering the full azimuth of $0 < \phi < 2\pi$ and pseudorapidity acceptance of $|\eta| < 1$, where the detector is assumed to have full tracking as well as electromagnetic and hadronic calorimetry such that jets can be robustly measured. In conjunction with the barrel central rapidity detector, a forward spectrometer, incorporating tracking and electromagnetic and hadronic calorimetry with pseudorapidity acceptance of $1.4 < \eta < 4$ and full azimuthal coverage, is also considered.

Projections were determined by taking the cross sections as predicted in the PYTHIA 6 event generator [28,29] and multiplying them by total integrated luminosity projections at RHIC. Assumed luminosities were 197 pb^{-1} for $p+p$ collisions and 0.33 pb^{-1} for $p+\text{Au}$ collisions, corresponding to the anticipated sPHENIX run plan for the second and third years of operation in the early 2020s. Estimates of experimental efficiencies are also applied, as described below for each process. The total expected yields were converted to per event yields by dividing by the total $p+p$ cross section times the expected luminosity. Thus, the ratio of the $p+\text{Au}$ to $p+p$ yields is always unity and the statistical uncertainties of the ratio are indicative of the actual statistical uncertainties on a measurement of $R_{\text{p}A}$, where $R_{\text{p}A}$ is defined as

TABLE I. A summary table showing the different combinations of pseudorapidity measurements for each channel generated in this study.

Central-central	Forward-central	Forward-forward
Drell-Yan		Drell-Yan
Dijets	Dijets	
Photon-jet		

$$R_{\text{p}A} \equiv \frac{1}{A}\frac{\text{d}\sigma_{\text{p}A}/\text{d}y\text{d}M^2}{\text{d}\sigma_{\text{pp}}/\text{d}y\text{d}M^2}. \quad (1)$$

Since the detector has both central and forward instrumentation, there are a number of rapidity regions that each observable can probe. Ideally measurements should be made in each region, as different values of x will be probed in both the proton and the nucleus when measuring at central and/or forward rapidities. Thus, we consider several combinations of observables, for which a summary table is shown in Table I.

Drell-Yan data are generated in both the central barrel and the forward arm independently from one another. In each case, the Drell-Yan dilepton pair is measured in the invariant mass range of $4.5 < M_{\ell^+\ell^-} < 9$ GeV/c^2. Experimental efficiencies for the reconstruction of Drell-Yan pairs were estimated from a full GEANT4 [30] simulation of the sPHENIX detector (including forward instrumentation). Tight cuts on the simulated data were used to reduce backgrounds so that the simulated measurement is dominated by Drell-Yan pairs and backgrounds from decays, conversions, etc., were minimal. Overall pair efficiencies vary between 10–15% over the invariant mass range considered for both central and forward measurements. The dijet data are considered in both regions to probe a variety of x values. Jets were determined from final-state PYTHIA particles with the anti-k_{T} algorithm with $R = 0.4$ [31]. Two dijet pseudodata samples are constructed, one in which both jets are measured in the central barrel and another where one jet is measured at central rapidity and the other is measured at forward rapidity. Since the p_T reach of jets becomes smaller at forward pseudorapidities, the leading jet at central rapidity is required to have $p_{\text{T}} > 12$ GeV/c and the subleading jet at forward rapidity is required to have $p_{\text{T}} > 8$ GeV/c. Experimental efficiencies for the reconstruction of dijet pairs were estimated in a similar fashion as the Drell-Yan data. Jets were reconstructed with both hadronic and electromagnetic calorimeter deposits and efficiencies were found to be approximately 80%, where these efficiencies become smaller towards the edge of the detector acceptance when some fraction of the jet cone lies outside of the detector. At forward pseudorapidity, the efficiencies were generally smaller, varying between 40 and 70% depending on the pseudorapidity of the jet. The direct photon-jet channel is

expected to have a high impact on the gluon nuclear PDF at small x when the process is measured in the forward direction. However, the photon-jet channel is difficult to measure at forward pseudorapidities due to large backgrounds from $\pi^0 \to \gamma\gamma$ decays. Thus, we only generate photon-jet pseudodata where both are measured in the central barrel, where previous direct photon measurements at RHIC have been made and where future RHIC experimental upgrades are expected to be able to measure this process. We note that if new forward instrumentation at RHIC were available to separate direct photons from backgrounds at forward pseudorapidities this would add a powerful additional observable to constrain the nuclear gluon PDF at low x, complementary to Drell-Yan [19]. The photon-jet cross sections were generated with $p_{\mathrm{T}}^{\gamma} > 10$ GeV$/c$ and $p_{\mathrm{T}}^{\mathrm{jet}} > 8$ GeV$/c$. Photon-jet reconstruction efficiencies were evaluated similarly to the dijet and Drell-Yan data, where the efficiency was found to be approximately 70% integrated across the central rapidity of the barrel detector.

A. Generation of pseudodata

From the PYTHIA simulations for $p+p$ and $p+\mathrm{Au}$ collisions we keep the relative statistical error, but construct the pseudodata points for the expected nuclear modification R_{pA} as

$$R_{\mathrm{pA}} = R_{\mathrm{pA}}^{\mathrm{EPPS16}} \times [1 + r\delta^{\mathrm{uncorr}}], \tag{2}$$

where δ^{uncorr} signifies the total uncorrelated data uncertainty and r is a Gaussian random variable. To obtain δ^{uncorr}, we add in quadrature the statistical uncertainty in the anticipated yield in $p+p$ and $p+\mathrm{Au}$ collisions. A 4% normalization uncertainty is assumed to account for the model dependence in determining $\langle N_{\mathrm{coll}} \rangle$ used in determining the R_{pA} ratio. The overall scale of this uncertainty is unimportant, however, assuming it is common to all measurements, as we will detail later. In addition, for dijet (photon-jet) measurement, another 5% (4%) uncorrelated bin-to-bin systematic uncertainty is added, corresponding to the residual experimental systematic error that does not cancel in the ratio. For the Drell-Yan case the statistical uncertainty dominates and no additional systematic uncertainty is added. A systematic uncertainty of the order of 5% is clearly smaller than what one can expect to be present in measurements for the absolute cross sections. However, if the $p+p$ and $p+\mathrm{Au}$ runs are made soon after each other (so that the detector configuration and calibration remains unaltered), much of the systematic uncertainty can be expected to cancel. We note that recent dijet measurements by the CMS Collaboration [32] quote a systematic uncertainty even less than 5%.

The central values for R_{pA} in Eq. (2) were obtained by NLO-level calculations using the CT14NLO [33] free-proton PDFs and EPPS16 [14] nuclear modifications. For dijets we used MEKS (v1.0) [34], with the anti-k_{T} algorithm, taking a jet cone $R = 0.4$, and fixing the QCD scales to the average of the two highest-p_{T} jets. The leading jet was required to have $p_{\mathrm{T}} > 12$ GeV$/c$ and the subleading jet $p_{\mathrm{T}} > 8$ GeV$/c$. These unequal cuts are necessary to avoid sensitivity to soft-gluon resummation. For photon-jet we used JETPHOX (v1.3.1) [35,36] where the jet was defined by a k_{T} algorithm with $R = 0.4$ and the QCD scales were fixed to the p_{T} of the photon. No isolation criteria for the photons were imposed. The NLO Drell-Yan cross sections are standard, and were calculated with a private code based on Ref. [37], fixing the QCD scales to the invariant mass of the dilepton pair. Besides the photon-jet process for which there are no data in the EPPS16 analysis, the scale choices are the same as those made in the EPPS16 fit.

III. IMPACT ON EPPS16

A. The Hessian reweighting technique in a nutshell

We estimate the impact of the projected data on the EPPS16 nuclear PDFs by the PDF reweighting (also called PDF profiling) method [25,38–41]. In this method, one studies the function

$$\chi^2(\vec{z}) = \sum_i (a_i z_i^2 + b_i z_i^3) + \chi^2_{\text{new data}}(\vec{z}), \tag{3}$$

where the first term describes the behavior of the original global χ^2 in the EPPS16 analysis, and the second term is the contribution of the new data to the overall χ^2 budget. The central fit of EPPS16 corresponds to $\vec{z} = \vec{0}$, and the error sets $S_i^{\pm}$ are defined in the z space by

$$\begin{aligned} S_1^{\pm} &= (\delta_1^{\pm}, 0, \ldots, 0), \\ S_2^{\pm} &= (0, \delta_2^{\pm}, \ldots, 0), \\ &\vdots \\ S_N^{\pm} &= (0, 0, \ldots, \delta_N^{\pm}), \end{aligned} \tag{4}$$

and they are known to increase the original χ^2 function by $T = 52$ units. The values for $\delta_i^{\pm}$ are given in the EPPS16 paper [14], from which the a_i and b_i coefficients in Eq. (3) can be solved. The contribution from the "new" data is defined as

$$\chi^2_{\text{new data}}(\vec{z}) = \sum_i \left[\frac{D_i - f_N T_i(\vec{z})}{E_i}\right]^2 + \left[\frac{f_N - 1}{E^{\mathrm{norm}}}\right]^2, \tag{5}$$

where D_i and E_i denote the ith data point and its error. The overall normalization uncertainty is marked by E^{norm}. We write the theoretical prediction $T_i(\vec{z})$ as

$$T_i(\vec{z}) = T_i(\vec{z} = \vec{0}) + \sum_k [\beta_{ik} z_k + \gamma_{ik} z_k^2], \tag{6}$$

where the coefficient for β_{ik} and γ_{ik} can be obtained by computing the predictions for T_i with all the PDF error sets. As a result, the total χ^2 in Eq. (3) becomes an analytic function of $\vec{z}$ which we numerically minimize with respect to $\vec{z}$ and f_N. We note that to avoid D'Agostini bias [42], the normalization factor in Eq. (5) f_N multiplies the theoretical prediction T_i, and not the data value D_i. Since the pseudodata are based on EPPS16, the new minimum is, by construction, always very close to $\vec{z} = \vec{0}$. After finding the parameters $\vec{z}_{\min}$ that correspond to the minimum of Eq. (3), we expand

$$\Delta\chi^2(\vec{z}) \equiv \chi^2(\vec{z}) - \chi^2(\vec{z}_{\min}) \approx (\vec{z} - \vec{z}_{\min})^{\mathrm{T}} H (\vec{z} - \vec{z}_{\min}), \tag{7}$$

where H is the second-derivative (Hessian) matrix. By diagonalizing the matrix H this becomes

$$\Delta\chi^2(\vec{z}) = \Delta\chi^2(\vec{v}) \approx (\vec{v})^2, \tag{8}$$

where $\vec{v} = P(\vec{z} - \vec{z}_{\min})$, where P is the orthogonal matrix that diagonalizes H, i.e., $P^T H P = 1$. The new error sets are then defined as in Eq. (4) assuming that the original tolerance is not altered, i.e., that each new error set $\hat{S}_i^{\pm}$ still corresponds to $\Delta\chi^2(\hat{S}_i^{\pm}) = 52$.

B. Correlating the overall normalization

The normalization uncertainty in Eq. (5) we discuss here is that of the luminosity determination of the minimum-bias data sample. At the LHC, the $p+A$ luminosities are determined by Van der Meer scans [43]. Alternatively, the measured per-event yields $\mathrm{d}N^{\mathrm{p}A}/N_{\mathrm{events}}$ are converted to cross sections $\mathrm{d}\sigma^{\mathrm{p}A}$ by

$$\mathrm{d}\sigma^{\mathrm{p}A} = \frac{\sigma_{\mathrm{pn}}^{\mathrm{inelastic}}}{\langle N_{\mathrm{coll}} \rangle} \frac{\mathrm{d}N^{\mathrm{p}A}}{N_{\mathrm{events}}}, \tag{9}$$

where the average number of binary nucleon-nucleon collisions $\langle N_{\mathrm{coll}} \rangle$ is estimated from a Glauber-type model [44]. This leads to a model-dependent normalization uncertainty which is difficult to determine. Furthermore, the inelastic proton-nucleon cross section $\sigma_{\mathrm{pn}}^{\mathrm{inelastic}}$ appearing in Eq. (9) is very sensitive to the physics at low scales $Q^2 \sim \Lambda_{\mathrm{QCD}}^2$, and it is presumably lower than the values measured in proton-proton collisions due to shadowing/saturation effects. The overall normalization is thus problematic in this approach. However, the normalization issue can be overcome by simultaneously measuring several observables from the same minimum-bias data sample. The reason is that there is only one single normalization uncertainty and in Eq. (5) the index i runs through all data points, not just those belonging to one single observable. Including data that probe PDFs in a relatively better constrained region thus serves to "calibrate" the overall normalization.

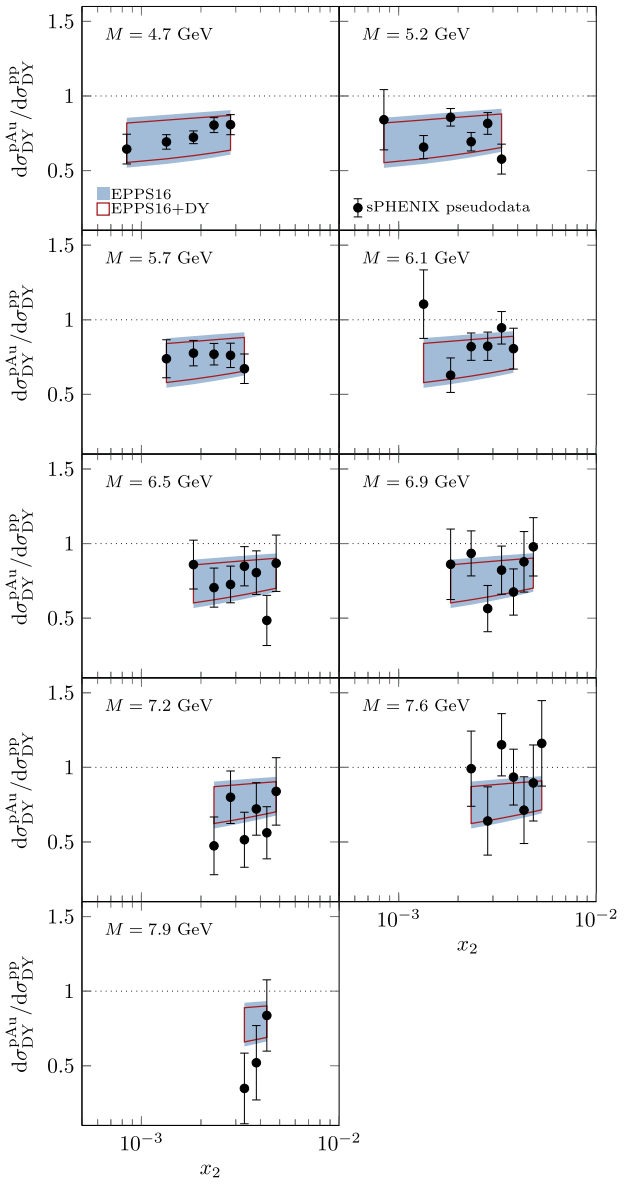

FIG. 1. Effect of PDF reweighting when only the forward Drell-Yan data (shown in the plot) are used in the analysis. The light-blue bands denote the original EPPS16 uncertainties, and the red lines indicate the new upper and lower uncertainty limits after reweighting.

To demonstrate how this works we have performed a PDF-profiling analysis first using only the forward Drell-Yan data, and then supplementing these data with the central-barrel dijet data. When only the Drell-Yan data are used, the constraints appear very weak. This is shown in Fig. 1 where the original EPPS16 uncertainties on the predictions are barely affected by the reweighting.

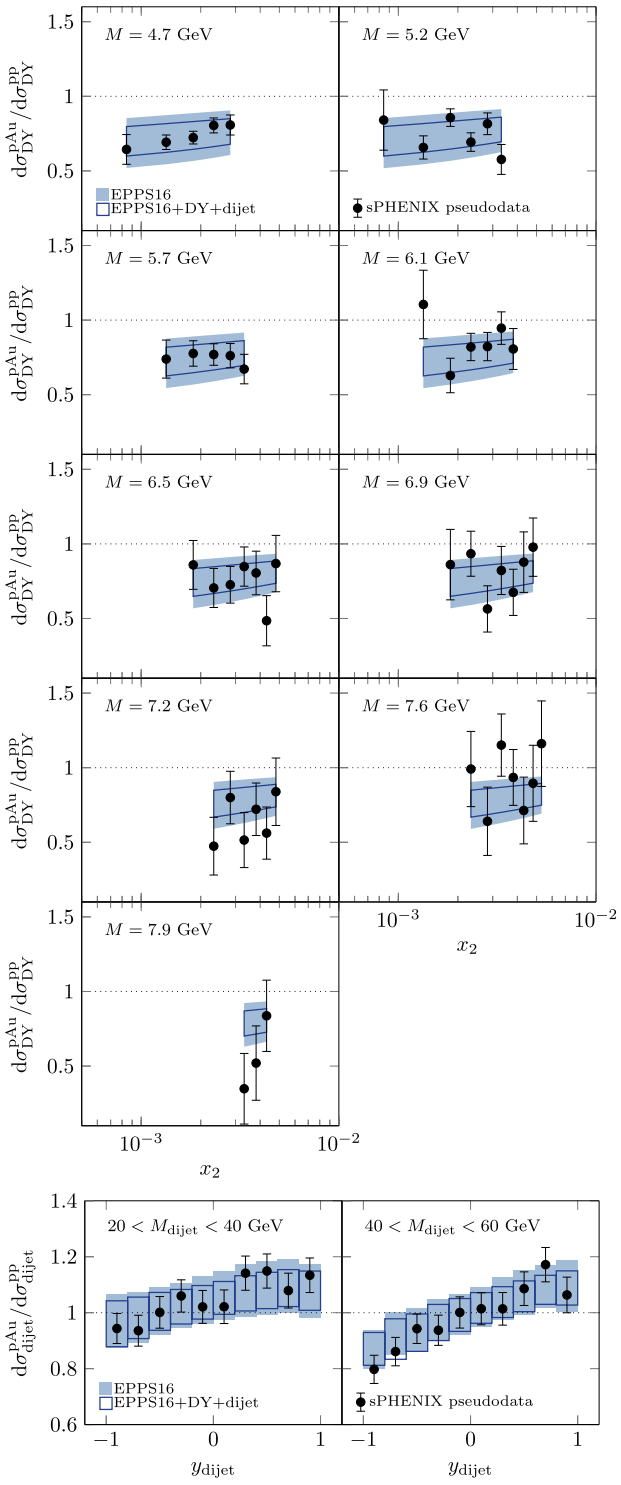

FIG. 2. Effect of PDF profiling using both the forward Drell-Yan (upper panels) and central-barrel dijet data (lower panels) with common normalization. The dark-blue lines indicate the new upper and lower uncertainty limits after the PDF profiling.

The reason for the inability of these data to provide constraints is that the nuclear modification is predicted to be rather flat at small x and the variations in PDFs around the original central value can be compensated by suitably tuning the normalization f_N. The flatness of the predicted Drell-Yan nuclear modification originates, to some extent, from the fit functions used in the EPPS16 analysis, but also from the scale evolution of PDFs which tends to flatten out the nuclear modifications in sea quarks. Here, we took the normalization uncertainty to be 4%, but if a larger number would have been used (e.g., 10%) even fewer constraints would have been obtained.

The situation changes when the central-barrel dijet projections are also included. These data probe the nuclear PDFs at much higher x than the Drell-Yan data and carry significant sensitivity also to the rather-well constrained sum of valence quarks $u^A_{\text{valence}} + d^A_{\text{valence}}$. The nuclear modifications for the dijets are expected to exhibit some excess (antishadowing) around $y_{\text{dijet}} \sim 1$ which turns into a suppression (EMC effect) for $y_{\text{dijet}} \sim -1$. Such a pattern cannot be mimicked by the overall normalization and leaves thus less room for f_N variation. Since the normalization is now common for the dijet and Drell-Yan data, the Drell-Yan data have a much larger impact. This is shown in Fig. 2, which should be compared to Fig. 1. While the uncertainties for the central-barrel dijet data are only slightly reduced from their original EPPS16 values, the inclusion of these data is crucial in fixing the overall normalization. We have also observed that our results do not significantly depend on the exact value we pick for the normalization uncertainty.

C. Simultaneous analysis of Drell-Yan, dijet and photon-jet pseudodata

Following the observation made in the previous subsection, our strategy is to simultaneously analyze several observables that share the common normalization uncertainty. To separate the effect of forward-arm measurements, we first present the results using only the central-barrel data, and then include the data simulated with the forward-arm acceptance.

In Fig. 3 we summarize the Drell-Yan, dijet and photon-jet pseudodata within the central-barrel acceptance $-1 < \eta < 1$. The light-blue bands ("EPPS16") show the original EPPS16 predictions, and the darker bands ("EPPS16+CB") show the error bands obtained after the reweighting analysis. We observe a modest improvement in the uncertainty bands for dijet and γ-jet cases. The precision of the Drell-Yan measurements is not expected to be high enough to set constraints as the sea quarks at $10^{-2} < x < 10^{-1}$ are already rather well constrained by the fixed-target DIS data. In Fig. 4, in turn, we show the combined

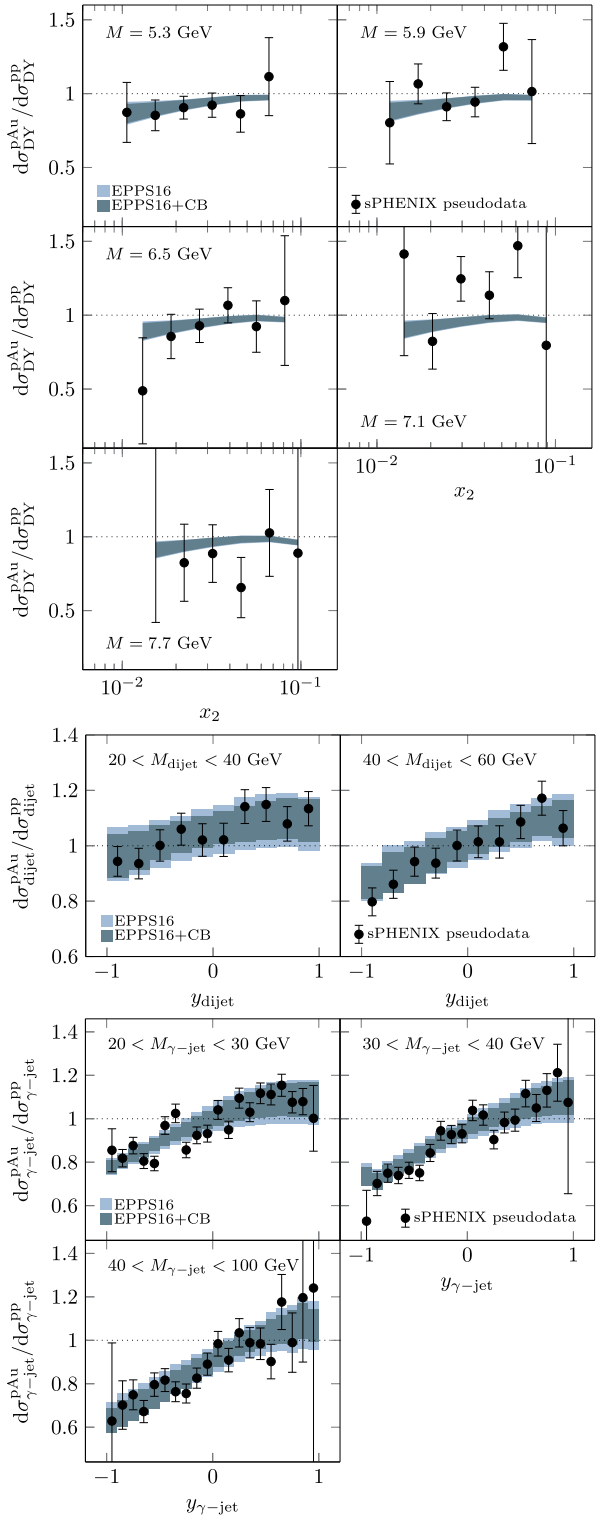

FIG. 3. Effect of PDF profiling using simultaneously the central-barrel Drell-Yan (upmost panels), dijet (middle panels), and photon-jet (bottom panels) pseudodata.

pseudodata within the full central-barrel and forward-instrumentation acceptance [45]. The light blue bands are again the original EPPS16 predictions, and the green bands ("EPPS16 + CB + FI") are the uncertainties obtained in the reweighting exercise. The reduction in the PDF uncertainties is now more significant than in the central-barrel-only case shown in Fig. 3.

The impact of both "EPPS16 + CB" and "EPPS16 + CB + FI" analyses on EPPS16 is shown in Fig. 5 where we plot the average sea-quark modification for Au,

$$R^{\mathrm{Au}}_{\bar{u}+\bar{d}+\bar{s}}(x,Q^2) \equiv \frac{f^{\mathrm{p/Au}}_{\bar{u}}(x,Q^2)+f^{\mathrm{p/Au}}_{\bar{d}}(x,Q^2)+f^{\mathrm{p/Au}}_{\bar{s}}(x,Q^2)}{f^{\mathrm{p}}_{\bar{u}}(x,Q^2)+f^{\mathrm{p}}_{\bar{d}}(x,Q^2)+f^{\mathrm{p}}_{\bar{s}}(x,Q^2)}, \tag{10}$$

together with the gluon nuclear modification,

$$R^{\mathrm{Au}}_{g}(x,Q^2) \equiv \frac{f^{\mathrm{p/Au}}_{g}(x,Q^2)}{f^{\mathrm{p}}_{g}(x,Q^2)}. \tag{11}$$

Here, $f^{\mathrm{p}/A}_i(x,Q^2)$ denotes the parton density in a bound proton and $f^{\mathrm{p}}_i(x,Q^2)$ is the free-proton PDF. We omit here the valence quarks as we found no effects there. The improvement we find in $R^{A}_{\bar{u}+\bar{d}+\bar{s}}$ is rather weak in both cases. In the central-barrel analysis, there is a modest improvement in the gluons across all values of x, though the small-x improvement is merely a consequence of the better constrained antishadowing regime which is transmitted to small x via momentum conservation and correlations in the EPPS16 fit function. The improvement for $R^{A}_{g}(x,Q^2)$ in the full analysis is clearly larger. Thanks to the wider acceptance, the full pseudodata sample is able to provide better direct constraints also at lower x. In particular, the gluon distribution gets significantly better constrained, the level of improvement being of the order of 50% in places.

Here we have found that the most decisive factor for constraining the small-x gluons is the forward-arm Drell-Yan data sample. At leading order, the Drell-Yan production occurs only via $q\bar{q}$ annihilation, but at small x the behavior of sea quarks is still strongly driven by the gluons. At NLO and beyond there is, in addition, a direct gluon contribution from the quark-gluon scattering. To further illustrate the sensitivity of the Drell-Yan process to the gluon PDF, Fig. 6 shows examples of the correlation cosine [46] between the gluon PDF and the Drell-Yan cross sections at fixed forward kinematics. Using the notation of Sec. III A, the correlation cosine of two quantities X and Y is defined as

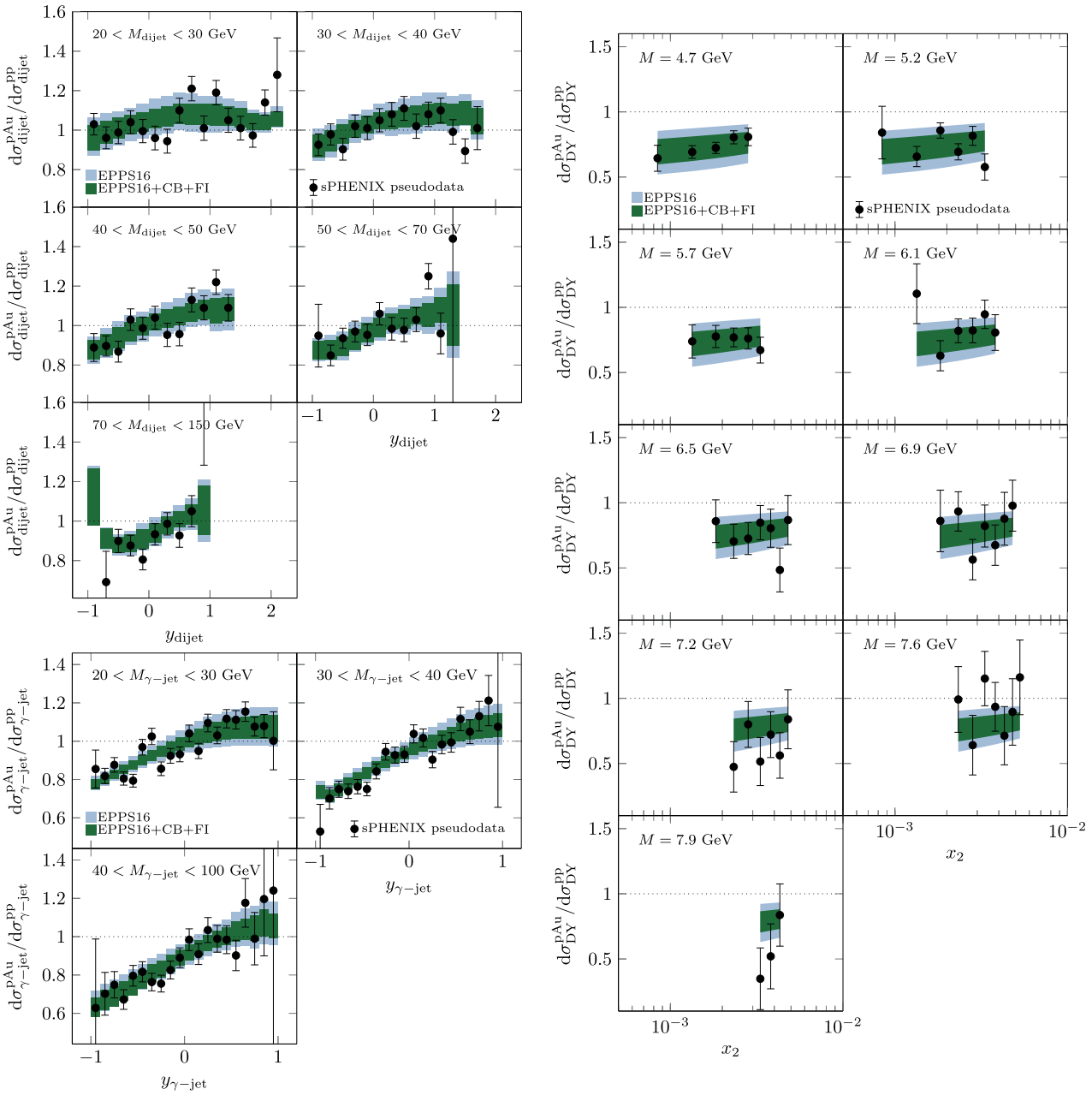

FIG. 4. Profiling analysis using simultaneously the full dijet (upper left panels), photon-jet (lower left panels), and Drell-Yan (rightmost panels) pseudodata. Central-barrel Drell-Yan pseudodata are omitted from the figure.

$$\cos(X,Y) \equiv \frac{\sum_k \Delta X_k \Delta Y_k}{(\sum_i \Delta X_i^2)(\sum_k \Delta Y_k^2)}, \tag{12}$$

$$\Delta X_k \equiv X(S_k^+) - X(S_k^-), \tag{13}$$

$$\Delta Y_k \equiv Y(S_k^+) - Y(S_k^-). \tag{14}$$

We take $X = f_g(x, Q^2)$ and $Y = d\sigma_{\rm pAu}/dydM^2$. If $\cos(X,Y) \sim (-)1$, the two quantities X and Y are strongly (anti)correlated whereas if $\cos(X,Y) \sim 0$, the two are independent. In computing the correlation cosine, we have kept the proton PDF $f_i^{\rm p}$ fixed to the CT14NLO central set, and used the CT14NLO error sets to vary $f_i^{\rm Au}$. In other words, we compute the cross sections using $f_i^p = f_i^p(S_0)$ for the proton and $f_i^{\rm p/Au}(S_k^\pm) = f_i^{\rm p}(S_k^\pm) R_i^{\rm Au,EPPS16}(S_0)$ to form Eq. (12). The point in using the CT14NLO error sets is that the CT14NLO fit function is somewhat more flexible at small x than the EPPS16 ansatz, so this should give a

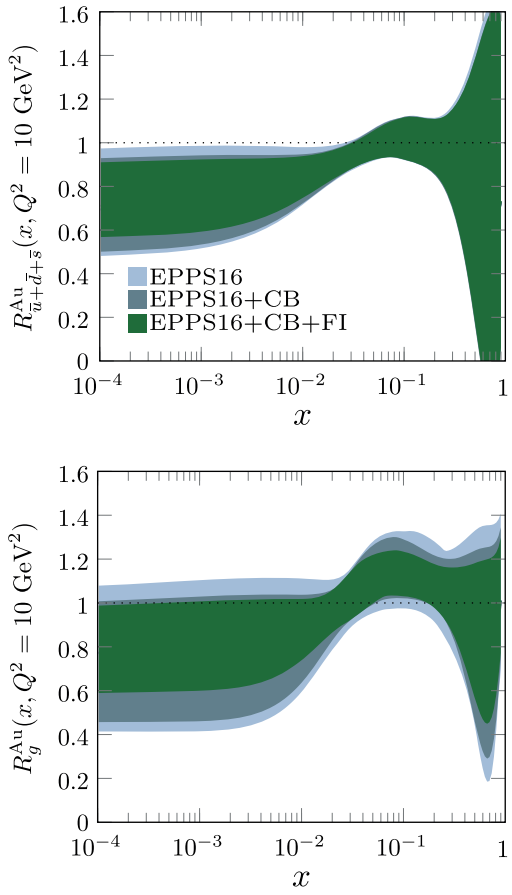

FIG. 5. Effects of PDF profiling for EPPS16. The light-blue bands ('EPPS16") are the original EPPS16 errors and the darker bands ("EPPS16 + CB") are those after profiling with the central-barrel pseudodata. The results of adding also the forward-instrumentation data are shown as green bands ("EPPS16 + CB + FI").

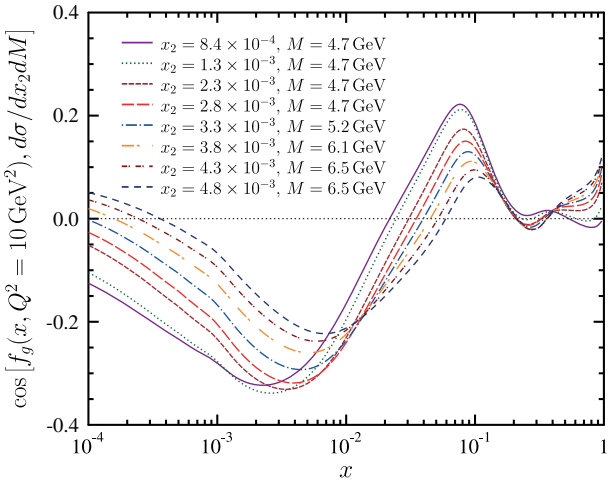

FIG. 6. Correlation cosine between the gluon PDF at $Q^2 = 10\ \mathrm{GeV}^2$ and small-x_2 Drell-Yan cross sections.

better estimate of the true correlations. From Fig. 6 we see that the gluon distribution at small x is anticorrelated with the forward Drell-Yan cross sections and at larger x we see a positive correlation. The main reason for the small-x anticorrelation is the direct contribution from the quark-gluon scattering, present at NLO and beyond, which is negative and clearly non-negligible. In our case, this amounts to ~15–45% of the cross sections with all partonic channels included. This contribution becomes increasingly important towards low x_2 and higher M. The large-x positive correlation persists also in a leading-order calculation so it is due to the indirect constraints from the scale evolution and momentum sum rule. Because the $q\bar{q}$ channel dominates the cross sections, the correlation with the gluon PDF is moderate but can reach almost up to ~40% at small x. Below $x \sim 10^{-3}$ the correlation begins to decrease as this region is beyond the kinematic reach of the projected experimental acceptance. In part, the residual nonzero correlations $x \lesssim 10^{-3}$ are due to the assumed form of the small-x fit function, but the momentum conservation and evolution effects also place indirect constraints. All in all, we can conclude that the Drell-Yan production at forward kinematics is indeed sensitive to the small-x gluon PDFs.

We note that the dijet and photon-jet pseudodata probe the mid- and high-x part of the nuclear PDFs. The uncertainties for these two observables are dominated by the assumed 5% uncorrelated bin-to-bin systematic error and the obtained improvements in nuclear PDFs are dictated by this assumption. If systematic uncertainties like those achieved in $p + \mathrm{Pb}$ collisions at the LHC [32] could be reached, the impact would be clearly larger. In addition, the systematic uncertainty of the LHC measurements is almost always of a correlated nature, but such correlation is difficult to estimate in advance. All in all, assuming a 5% uncorrelated systematic uncertainty appears thus a reasonable test scenario which should not overstate the constraining power.

IV. CONSTRAINING THE *A* DEPENDENCE OF NUCLEAR PDFs WITH LIGHTER IONS

The mass-number (A) dependence of the current nuclear PDFs is not well known: direct constraints exist only for large-x valence quarks and intermediate-x sea quarks. On the one hand, e.g., in the EPPS16 analysis, the guideline has been that the nuclear effect should be larger for larger nuclei at the parametrization scale $Q = m_{\mathrm{charm}}$ which then tends to lead to physically sound A systematics also at larger Q. On the other hand, in the recent nuclear-PDF analysis by the NNPDF Collaboration [18] there is less direct control over the A dependence and thus the nuclear effects from one nucleus to another can fluctuate significantly. Due to the $p + \mathrm{Pb}$ and $\mathrm{Pb} + \mathrm{Pb}$ collisions program at the LHC, the near-future improvements on nuclear PDFs are bound to be driven by the Pb nucleus. For example, the

dijet [32], D-meson [47] and $W^{\pm}$ [48] measurements efficiently constrain [25,49] the gluons in the Pb nucleus, perhaps providing even stronger constraints for Pb than what we have found in the present study for Au. The LHCb fixed-target mode facilitates measurements on lighter noble-gas targets [50], but only the very-high x regime of nuclear PDFs is accessible. However, e.g., in astrophysical applications the relevant nuclei (e.g., oxygen and nitrogen) are much lighter and thus collider measurements involving lighter nuclei would be very useful [51]. In addition, the study of the onset of jet quenching and saturation phenomena with lighter ions will require nuclear PDFs for nuclei other than Au or Pb. Interest in light-ion beams at the LHC has been expressed [52] but since the main focus of the LHC is still on $p+p$ collisions, it is not clear whether and when this would materialize. Here, the flexibility of RHIC to run with different ions is a clear asset. Indeed, at least p, d, Al, Cu, Ru, Zr and U ions have already been used in physics runs which demonstrates that a proper "*A scan*" is, in principle, possible. The same multi-ion option would also be available if RHIC is eventually turned into an electron-ion collider [53], where the possibilities to constrain nuclear PDFs are undisputed [54,55]. To highlight the present uncertainties for light ions, Fig. 7 shows the nuclear effects for $A=40$ (Ca, Ar) from the EPPS16 and nCTEQ15 [15] global fits of nuclear PDFs. While the uncertainty bands overlap, the shapes at intermediate and large x are quite different: while the nuclear effects in nCTEQ15 monotonically rise towards high x, the EPPS16 error band more closely resembles the typical pattern of shadowing, antishadowing and EMC effect. Figure 8 shows how this different behavior would be reflected in dijet production. In the backward direction ($y_{\rm dijet}<0$) one is sensitive to the large-x part of nuclear PDFs and the nCTEQ15 prediction tends to be above the EPPS16 one, consistently with Fig. 7. The difference in Fig. 8 is not as marked as in Fig. 7 as towards large x the valence quarks also play an increasingly important role. Towards $y_{\rm dijet}\gg 0$ the probed x gets lower and, in line with Fig. 7, the nCTEQ15 prediction tends to be at the lower limit of

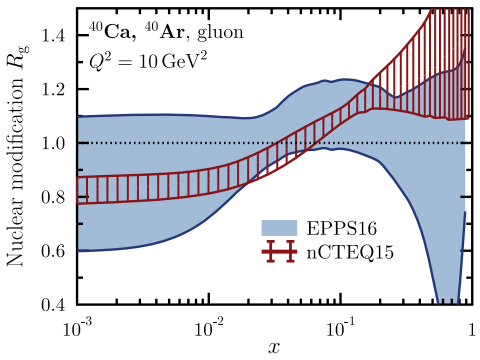

FIG. 7. Nuclear modifications of gluon PDFs for $A=40$ nuclei from EPPS16 and nCTEQ15.

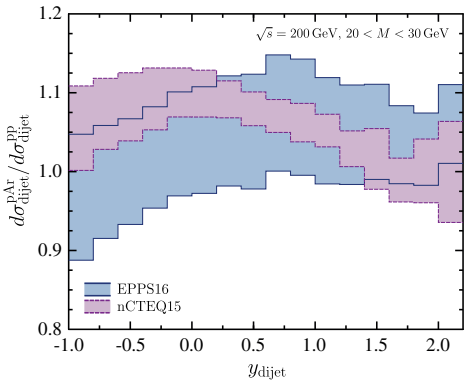

FIG. 8. Dijet nuclear modification in p-Ar scattering as predicted by EPPS16 and nCTEQ15 for an invariant mass $20<M<30$.

EPPS16. Assuming a similar ∼50% reduction in the gluon PDF uncertainties as found for Au in Fig. 5, it appears reasonable that the measurements would be able to resolve between nCTEQ15 and EPPS16. In an approach like that of the NNPDF Collaboration [18], where more freedom for the A dependence is allowed than in nCTEQ15 or EPPS16, the benefit would be even more pronounced.

An additional interesting possibility we would like to point out would be to study $p+A_i$ collisions of two isobaric nuclei A_1 and A_2 (e.g., $A_1={}^{96}_{44}$Ru vs $A_2={}^{96}_{40}$Zr collisions) with constant A but differences in proton and neutron content. Precision measurements of e.g., $(p+\mathrm{Ru})/(p+\mathrm{Zr})$ ratios for hard processes (like those discussed in this paper) would allow a study of the assumptions made in the present global fits of nuclear PDFs. Indeed, it is currently assumed that the nuclear effects depend only on the mass number A, and not on the mutual balance of neutrons and protons. In addition, the isospin symmetry (i.e., $u^{\rm proton/A}=d^{\rm neutron/A}$) is assumed to be exact. Thus, $(p+\mathrm{Ru})/(p+\mathrm{Zr})$ ratios, or other similar constant-A combinations, would test the assumptions made in global analyses at a deeper level and also test other theoretical approaches, e.g., the importance of short-range nucleon-nucleon correlations [56], or the lack of them [57]. In principle, in an optimal situation the neutron-to-proton mixture in the two nuclei should be as different as possible, with (at least nearly) constant A. Such a measurement makes optimal use of the flexibility of the RHIC facility.

V. SUMMARY

Using the Au nucleus as a test case, we have examined the prospects for constraining nuclear gluon PDFs at RHIC with new measurements that assume detector acceptances similar to those proposed for STAR and sPHENIX with forward upgrades. We have found that the Drell-Yan process at low invariant mass is able to significantly constrain the low-x gluon distribution with up to 50%

reduction in the current EPPS16 uncertainty. The constraints at higher x depend considerably on the assumed systematic uncertainty, which is expected to dominate over the statistical uncertainty for dijet and photon-jet processes. Assuming an order of 5% bin-to-bin independent systematic uncertainty leads to modest constraints in the mid- and high-x regions. Even so, we find the inclusion of additional observables along with the Drell-Yan data of utmost importance to overcome the overall normalization uncertainty in the $R_{\mathrm{p}A}$ ratio. Without supplementing the Drell-Yan pseudodata with other observables (here either dijets, photon-jet, or both), we find that the power of the measurement of Drell-Yan to constrain the small-x behavior of the gluon is lost. It is possible that even stronger constraints could be obtained if measurements of forward direct photons could be added to this suite of observables.

While the focus of our analysis was on the Au nucleus, similar constraints can be expected to be obtained for any other nucleus. In this respect we briefly discussed the A dependence of nuclear PDFs and highlighted the significant opportunity for improvements that could be attained with a proper A *scan*—measuring the same observables with several nuclear beams—for which the RHIC collider provides a unique opportunity.

ACKNOWLEDGMENTS

I. H. and H. P. acknowledge the support from the Academy of Finland, Project No. 308301. The work of P. P. has been supported by the Magnus Ehrnrooth Foundation. The Finnish IT Center for Science (CSC) is acknowledged for super-computing resources under the Project No. jyy2580. J. O. acknowledges support from the Margaret and Herman Sokol Faculty Awards of University of Michigan and the Office of Nuclear Physics in the Office of Science of the Department of Energy under Grant No. DE-SC0013393. J. L. acknowledges support from the Office of Nuclear Physics in the Office of Science of the Department of Energy under Grants No. DE-FG02-10ER41719 and DE-FG02-92ER40962.

[1] H.-W. Lin *et al.*, Prog. Part. Nucl. Phys. **100**, 107 (2018).
[2] S. Forte and G. Watt, Annu. Rev. Nucl. Part. Sci. **63**, 291 (2013).
[3] J. Gao, L. Harland-Lang, and J. Rojo, Phys. Rep. **742**, 1 (2018).
[4] H. Paukkunen, *Proc. Sci.* HardProbes2018 (2019) 014 [arXiv:1811.01976].
[5] H. Paukkunen, Nucl. Phys. **A967**, 241 (2017).
[6] S. S. Adler *et al.* (PHENIX Collaboration), Phys. Rev. Lett. **98**, 172302 (2007).
[7] B. I. Abelev *et al.* (STAR Collaboration), Phys. Rev. C **81**, 064904 (2010).
[8] X. Li, Proc. Sci., DIS2015 (**2015**) 203 [arXiv:1506.06314].
[9] A. Adare *et al.* (PHENIX Collaboration), Phys. Rev. Lett. **116**, 122301 (2016).
[10] S. D. Ellis, Z. Kunszt, and D. E. Soper, Phys. Rev. Lett. **64**, 2121 (1990).
[11] Z. Nagy, Phys. Rev. D **68**, 094002 (2003).
[12] M. Posik (STAR Collaboration), Proc. Sci. DIS2017 (**2018**) 214.
[13] R. S. Towell *et al.* (NuSea Collaboration), Phys. Rev. D **64**, 052002 (2001).
[14] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, Eur. Phys. J. C **77**, 163 (2017).
[15] K. Kovarik *et al.*, Phys. Rev. D **93**, 085037 (2016).
[16] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, Phys. Rev. D **85**, 074028 (2012).
[17] H. Khanpour and S. A. Tehrani, Phys. Rev. D **93**, 014026 (2016).
[18] R. A. Khalek, J. J. Ethier, and J. Rojo, arXiv:1904.00018.
[19] E.-C. Aschenauer *et al.*, arXiv:1602.03922.
[20] A. Adare *et al.* (PHENIX Collaboration), arXiv:1501.06197.
[21] F. Arleo and T. Gousset, Phys. Lett. B **660**, 181 (2008).
[22] F. Arleo, K. J. Eskola, H. Paukkunen, and C. A. Salgado, J. High Energy Phys. 04 (2011) 055.
[23] K. J. Eskola, V. J. Kolhinen, and C. A. Salgado, Eur. Phys. J. C **9**, 61 (1999).
[24] K. J. Eskola, H. Paukkunen, and C. A. Salgado, J. High Energy Phys. 10 (2013) 213.
[25] K. J. Eskola, P. Paakkinen, and H. Paukkunen, arXiv: 1903.09832.
[26] T. Stavreva, I. Schienbein, F. Arleo, K. Kovarik, F. Olness, J. Y. Yu, and J. F. Owens, J. High Energy Phys. 01 (2011) 152.
[27] M. Klasen, C. Klein-Bösing, and H. Poppenborg, J. High Energy Phys. 03 (2018) 081.
[28] T. Sjostrand, S. Mrenna, and P. Z. Skands, J. High Energy Phys. 05 (2006) 026.
[29] T. Sjostrand, S. Mrenna, and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008).
[30] S. Agostinelli *et al.* (GEANT4 Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).
[31] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012).
[32] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **121**, 062002 (2018).
[33] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, Phys. Rev. D **93**, 033006 (2016).
[34] J. Gao, Z. Liang, D. E. Soper, H.-L. Lai, P. M. Nadolsky, and C. P. Yuan, Comput. Phys. Commun. **184**, 1626 (2013).
[35] S. Catani, M. Fontannaz, J. P. Guillet, and E. Pilon, J. High Energy Phys. 05 (2002) 028.
[36] Z. Belghobsi, M. Fontannaz, J. P. Guillet, G. Heinrich, E. Pilon, and M. Werlen, Phys. Rev. D **79**, 114024 (2009).

[37] H. Paukkunen, Global analysis of nuclear parton distribution functions at leading and next-to-leading order perturbative QCD, Ph.D. thesis, Jyvaskyla University, 2009.
[38] H. Paukkunen and C. A. Salgado, Phys. Rev. Lett. **110**, 212301 (2013).
[39] H. Paukkunen and P. Zurita, J. High Energy Phys. 12 (2014) 100.
[40] S. Camarda *et al.* (HERAFitter Developers' Team), Eur. Phys. J. C **75**, 458 (2015).
[41] C. Schmidt, J. Pumplin, C. P. Yuan, and P. Yuan, Phys. Rev. D **98**, 094005 (2018).
[42] G. D'Agostini, Nucl. Instrum. Methods Phys. Res., Sect. A **346**, 306 (1994).
[43] P. Grafström and W. Kozanecki, Prog. Part. Nucl. Phys. **81**, 97 (2015).
[44] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, Annu. Rev. Nucl. Part. Sci. **57**, 205 (2007).
[45] We note that it would still be possible to increase the acceptance towards large nuclear x by swapping the beam directions (p-A $\leftrightarrow$ A-p) as has been done at the LHC.
[46] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Phys. Rev. D **65**, 014013 (2001).
[47] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 10 (2017) 090.
[48] CMS Collaboration, arXiv:1905.01486 [Phys. Lett. B (to be published)].
[49] A. Kusina, J.-P. Lansberg, I. Schienbein, and H.-S. Shao, Phys. Rev. Lett. **121**, 052004 (2018).
[50] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **122**, 132002 (2019).
[51] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **121**, 222001 (2018).
[52] Z. Citron *et al.*, in *HL/HE-LHC Workshop: Workshop on the Physics of HL-LHC, and Perspectives at HE-LHC Geneva, Switzerland, 2018* (2018) [arXiv:1812.06772].
[53] A. Accardi *et al.*, Eur. Phys. J. A **52**, 268 (2016).
[54] E. C. Aschenauer, S. Fazio, M. A. C. Lamont, H. Paukkunen, and P. Zurita, Phys. Rev. D **96**, 114005 (2017).
[55] E. C. Aschenauer, S. Fazio, J. H. Lee, H. Mantysaari, B. S. Page, B. Schenke, T. Ullrich, R. Venugopalan, and P. Zurita, Rep. Prog. Phys. **82**, 024301 (2019).
[56] O. Hen, E. Piasetzky, and L. B. Weinstein, Phys. Rev. C **85**, 047301 (2012).
[57] J. Arrington and N. Fomin, arXiv:1903.12535.

V

A QCD ANALYSIS OF LHCB D-MESON DATA IN p+Pb COLLISIONS

by

Eskola, K. J., Helenius, I., Paakkinen, P. & Paukkunen, H., 2019

submitted to JHEP

arXiv: 1906.02512 [hep-ph]

Prepared for submission to JHEP

A QCD analysis of LHCb D-meson data in p+Pb collisions

Kari J. Eskola, Ilkka Helenius, Petja Paakkinen and Hannu Paukkunen

University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland

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

E-mail: kari.eskola@jyu.fi, ilkka.m.helenius@jyu.fi, petja.paakkinen@jyu.fi, hannu.t.paukkunen@jyu.fi

Abstract: We scrutinize the recent LHCb data for D^0-meson production in p+Pb collisions within a next-to-leading order QCD framework. Our calculations are performed in the SACOT-m_T variant of the general-mass variable-flavour-number scheme (GM-VFNS), which has previously been shown to provide a realistic description of the LHC p+p data. Using the EPPS16 and nCTEQ15 nuclear parton distribution functions (PDFs) we show that a very good agreement is obtained also in the p+Pb case both for cross sections and nuclear-modification ratios in the wide rapidity range covered by the LHCb data. Encouraged by the good correspondence, we quantify the impact of these data on the nuclear PDFs by the Hessian reweighting technique. We find compelling direct evidence of gluon shadowing at small momentum fractions x, with no signs of parton dynamics beyond the collinear factorization. We also compare our theoretical framework to other approaches and are led to conclude that a full GM-VFNS calculation is most essential in constraining general-purpose PDFs with D-meson data.

Contents

1 Introduction

In the collinear-factorization approach to describe scattering of protons and heavier nuclei in Quantum Chromodynamics (QCD), the non-perturbative structure of the hadrons — parton distribution functions (PDFs) — is factorized from the perturbatively calculable coefficient functions [1, 2]. The PDFs are typically extracted from experimental data via global analysis and their accurate determination has been a long-standing effort in the community [2, 3]. For the free proton PDF fits there are plenty of accurate data available and the most recent global analyses [4–8] result with PDFs that are reasonably well constrained within the typical kinematics probed at the Large Hadron Collider (LHC).

For PDFs in heavier nuclei, nuclear PDFs (nPDFs), the available data have been rather sparse until very lately [9]. Indeed, even some recent analyses still rely only on older fixed-target deep inelastic scattering (DIS) and Drell-Yan (DY) data [10, 11]. Due to the relatively low center-of-mass (c.m.) energy $\sqrt{s}$, these data provide constraints only for momentum fractions $x \gtrsim 0.01$, and the gluons are constrained only indirectly via scale-evolution effects and momentum sum rule [12]. To obtain better gluon constraints, the potential of inclusive pion production in d+Au collisions at RHIC [13–16] was first discussed in ref. [17] and eventually the data were incorporated into the global fits [18–21]. The x reach was still, however, rather similar to the available DIS data. The currently most comprehensive nPDF analysis, EPPS16 [22], includes also LHC Run-I data for electroweak-boson ($\mathrm{W}^{\pm}$ and Z^0) [23–25] and dijet production [26] in p+Pb collisions. Because of the large masses of the $\mathrm{W}^{\pm}$ and Z^0 bosons, the interaction scale is high and a significant sensitivity to gluons via

evolution effects will eventually set constraints on gluons, as has been shown in ref. [27] (sect. 10.4.2). However, the Run-I $W^{\pm}$ and Z^0 data have still a rather limited impact due to the low statistics. The dijet production, on the other hand, probes the gluon density much more directly and already the Run-I data clearly helps to narrow down the gluons in the $x \gtrsim 0.002$ region [28]. All this still leaves the small-x region only weakly constrained. To probe gluons at small x, almost any conceivable observable at lowish interaction scales and forward rapidity $y \gg 0$ would do. Good candidates at hadron colliders include e.g. low-mass Drell-Yan dilepton and isolated-photon production at low transverse momentum p_{T} [29–35]. Isolated photons in p+Pb collisions have already been measured at central rapidities [36], and the large-y measurements appear to be within the capabilities of the LHCb collaboration [37]. In further future, measurements of isolated-photon production would be a central goal of the ALICE FoCal upgrade [38].

Another promising observable for gluon constraints is the inclusive D- and B-meson production where the heavy-quark mass provides the hard scale even at zero p_{T}. In fact, the LHCb collaboration has published low-p_{T} data on D-meson production at forward kinematics in p+p collisions at different $\sqrt{s}$ [39–41], and recently also in the p+Pb case at $\sqrt{s} = 5$ TeV [42]. The use of these D-meson data as a free proton and nuclear PDF constraint has been advocated e.g. in refs. [43–48] and studied otherwise [49], but for the moment the default sets of globally fitted general-purpose PDFs [4–8, 21, 22] do not include any D-meson data. Here, our purpose is to provide a first estimate of the impact the recent LHCb p+Pb data have on globally fitted nPDFs within a rigorous next-to-leading order (NLO) perturbative-QCD framework. We will focus only on the LHCb measurements [42], as the the central-rapidity ALICE [50] data are not as precise and as the ATLAS central-rapidity data [51] are only preliminary. The cross sections are calculated in the SACOT-m_{T} general-mass variable-flavour-number scheme (GM-VFNS) presented in ref. [52]. Our framework takes fully into account the D mesons produced by gluon fragmentation — something that has been overlooked in the above-mentioned works [43–48] — and thus provides a realistic estimate of the data impact. To quantify the impact on the EPPS16 [22] and nCTEQ15 [21] nPDFs, we will use the Hessian reweighting technique [28, 53–55] that facilitates an estimate of the data impact without re-doing the complete global analysis.

The paper will now continue as follows: In section 2, we introduce our theoretical setup, including the GM-VFNS framework and the applied reweighting machinery. Then, in section 3, we compare the resulting cross sections and nuclear modification ratios with the LHCb data, demonstrate the impact these data have on nPDFs, and discuss their sensitivity to small-x gluons. We summarize our findings in section 4.

2 Theoretical framework

2.1 SACOT-m_{T} scheme for heavy-quark production

The general idea of D-meson hadroproduction in the GM-VFNS approach [52, 56] is to reproduce the results of (3-flavour) fixed flavour-number scheme (FFNS) at the small p_{T} limit and match to the massless calculation at high values of p_{T}. Let us first discuss the FFNS limit, in which the cross section for inclusive production of a heavy-flavoured hadron

h_3 at a given transverse momentum $P_{\rm T}$ and rapidity Y in a collision of two hadrons, h_1 and h_2, can be written as

$$\begin{aligned}\frac{\mathrm{d}\sigma^{h_1+h_2\to h_3+X}}{\mathrm{d}P_{\rm T}\mathrm{d}Y}\Big|_{\rm FFNS} &= \sum_{ij}\int_{z^{\rm min}}^1\frac{\mathrm{d}z}{z}\int_{x_1^{\rm min}}^1\mathrm{d}x_1\int_{x_2^{\rm min}}^1\mathrm{d}x_2 \\ &\times D_{Q\to h_3}(z)\,f_i^{h_1}(x_1,\mu^2_{\rm fact})f_j^{h_2}(x_2,\mu^2_{\rm fact})\frac{\mathrm{d}\hat{\sigma}^{ij\to Q+X}}{\mathrm{d}p_{\rm T}\mathrm{d}y}(\tau_1,\tau_2,m,\mu^2_{\rm ren},\mu^2_{\rm fact})\,.\end{aligned} \tag{2.1}$$

In this expression, $f_{i,j}^{h_{1,2}}$ are the PDFs (in 3-flavour scheme) for partons i and j in hadrons h_1 and h_2 with momentum fractions x_1 and x_2, and $\mathrm{d}\hat{\sigma}^{ij\to Q+X}/\mathrm{d}p_{\rm T}\mathrm{d}y$ denote the perturbatively calculable coefficient functions for inclusive heavy-quark Q (here charm) production [57] with fixed rapidity y and transverse momentum $p_{\rm T}$ of Q. The renormalization and factorization scales are denoted by $\mu^2_{\rm ren},\mu^2_{\rm fact}$ and m is the heavy-quark (here charm) mass. The fragmentation of a heavy-quark to hadron h_3 is described by a scale-independent fragmentation function (FF) $D_{Q\to h_3}$ (such as in ref. [58]). The invariants τ_i can be calculated from the partonic transverse mass $m_{\rm T}=\sqrt{p_{\rm T}^2+m^2}$ and rapidity y as

$$\tau_1\equiv\frac{p_1\cdot p_3}{p_1\cdot p_2}=\frac{m_{\rm T}\mathrm{e}^{-y}}{x_2\sqrt{s}}\quad\text{and}\quad\tau_2\equiv\frac{p_2\cdot p_3}{p_1\cdot p_2}=\frac{m_{\rm T}\mathrm{e}^{y}}{x_1\sqrt{s}}. \tag{2.2}$$

where p_1 and p_2 are the momenta of the incoming massless partons, and p_3 is the final-state heavy-quark momentum. When masses are neglected, the relation between partonic and hadronic variables is simply $y=Y$ and $P_{\rm T}=zp_{\rm T}$. However, when the masses of the heavy quark and the final-state hadron are taken into account, the definition of z becomes ambiguous [59]. Adopting the choice made in [52],

$$z\equiv\frac{P_3\cdot(P_1+P_2)}{p_3\cdot(P_1+P_2)}, \tag{2.3}$$

where P_i is the momentum of hadron h_i, the z variable can be interpreted as the fraction of partonic energy carried by the outgoing hadron in the c.m. frame of the initial-state hadrons h_1 and h_2. The relations between partonic and hadronic variables become somewhat more involved, but eq. (2.1) stays intact.

When the transverse momentum of the produced hadron h_3 is large, $P_{\rm T}\gg m$, the heavy-quark mass can be neglected and thus the zero-mass description becomes the most relevant. In this limit, the cross section can be written as [60],

$$\begin{aligned}\frac{\mathrm{d}\sigma^{h_1+h_2\to h_3+X}}{\mathrm{d}P_{\rm T}\mathrm{d}Y}\Big|_{\rm ZM} &= \sum_{ijk}\int_{z^{\rm min}}^1\frac{\mathrm{d}z}{z}\int_{x_1^{\rm min}}^1\mathrm{d}x_1\int_{x_2^{\rm min}}^1\mathrm{d}x_2 \\ &\times D_{k\to h_3}(z,\mu^2_{\rm frag})\,f_i^{h_1}(x_1,\mu^2_{\rm fact})f_j^{h_2}(x_2,\mu^2_{\rm fact})\frac{\mathrm{d}\hat{\sigma}^{ij\to k+X}}{\mathrm{d}p_{\rm T}\mathrm{d}y}(\tau_1^0,\tau_2^0,\mu^2_{\rm ren},\mu^2_{\rm fact},\mu^2_{\rm frag})\,.\end{aligned} \tag{2.4}$$

The formal difference with respect to eq. (2.1) is that now the FFs are fragmentation-scale $\mu^2_{\rm frag}$ dependent, and a summation over all partonic channels is included. For massless partons the invariants τ_i^0 are obtained as

$$\tau_1^0=\lim_{m\to0}\tau_1=\frac{p_{\rm T}\mathrm{e}^{-y}}{x_2\sqrt{s}}\quad\text{and}\quad\tau_2^0=\lim_{m\to0}\tau_2=\frac{p_{\rm T}\mathrm{e}^{y}}{x_1\sqrt{s}}. \tag{2.5}$$

The GM-VFNS technique [52, 56] provides a general framework to match the two extremes of eq. (2.1) and eq. (2.4) in a way that is consistent with collinear factorization. If we start from the FFNS description and increase $P_{\rm T}$, the cross sections will quickly be dominated by $\log(p_{\rm T}/m)$ terms whose origin is in the initial- and final-state partons' collinear splittings into $Q\overline{Q}$ pairs. In GM-VFNS these logarithms are resummed to the scale-dependent heavy-quark PDFs and scale-dependent FFs. Because the FFNS expressions already contain the first of the resummed logarithmic terms, subtraction terms are needed to avoid double counting and ensure the correct zero-mass limit of eq. (2.4). For example, the inclusion of the gluon production channel $gg \to gg$,

$$\begin{aligned}\frac{\mathrm{d}\sigma^{h_1+h_2\to h_3+X}}{\mathrm{d}P_{\rm T}\mathrm{d}Y}\Big|_{gg\to gg} &= \int_{z^{\rm min}}^1 \frac{\mathrm{d}z}{z}\int_{x_1^{\rm min}}^1 \mathrm{d}x_1 \int_{x_2^{\rm min}}^1 \mathrm{d}x_2 \\ &\times D_{g\to h_3}(z,\mu^2_{\rm frag})\, f_g^{h_1}(x_1,\mu^2_{\rm fact}) f_g^{h_2}(x_2,\mu^2_{\rm fact}) \frac{\mathrm{d}\hat{\sigma}^{gg\to g+X}}{\mathrm{d}p_{\rm T}\mathrm{d}y}(\tilde{\tau}_1,\tilde{\tau}_2,\mu^2_{\rm ren},\mu^2_{\rm fact},\mu^2_{\rm frag})\end{aligned} \tag{2.6}$$

on top of eq. (2.1), must be accompanied by a subtraction term which has otherwise the same expression as eq. (2.6) but where the gluon-to-h_3 FF is replaced by

$$\begin{aligned}D_{g\to h_3}(x,\mu^2_{\rm frag}) &= \frac{\alpha_{\rm s}}{2\pi}\log\left(\frac{\mu^2_{\rm frag}}{m^2}\right)\int_x^1 \frac{\mathrm{d}z}{z} P_{qg}(x/z) D_{Q\to h_3}(z) \\ &= \frac{\alpha_{\rm s}}{2\pi}\log\left(\frac{\mu^2_{\rm frag}}{m^2}\right)\int_x^1 \frac{\mathrm{d}z}{z} P_{qg}(x/z) D_{Q\to h_3}(z,\mu^2_{\rm frag}) + \mathcal{O}(\alpha_s^2)\,,\end{aligned} \tag{2.7}$$

which is the first term in the definition of scale-dependent FFs with massive quarks. In an NLO-accurate $\mathcal{O}(\alpha_s^3)$ calculation, only the leading-order part of $\mathrm{d}\hat{\sigma}^{gg\to g+X}$ is included in the subtraction term. However, the exact form of $\mathrm{d}\hat{\sigma}^{gg\to g+X}$ in the equation above is not fixed by this construction. The only condition is that we recover the standard zero-mass $\overline{\rm MS}$ expression at $p_{\rm T} \to \infty$ to meet eq. (2.4). This means that we can include mass-dependent terms in $\mathrm{d}\hat{\sigma}^{gg\to g+X}$ as we like, and a specific choice defines a scheme. The difference between the added and subtracted contributions discussed above is formally of order $\mathcal{O}(\alpha_s^4)$, so that different schemes are formally equivalent up to $\mathcal{O}(\alpha_s^3)$. Here we adopt the so-called SACOT-$m_{\rm T}$ scheme [52]. It is rooted in a simple observation that in order to make a heavy-flavoured hadron in QCD, a $Q\bar{Q}$ pair must be first produced. That is, the relevant invariants to describe the process are the massive ones, $\tilde{\tau}_{1,2} = \tau_{1,2}$, even for seemingly massless partonic contribution (like the $gg \to gg$ channel). Importantly, the mass then prevents the partonic cross sections from diverging towards small $p_{\rm T}$ exactly in the same way as the FFNS cross section are finite at $p_{\rm T} = 0$. In the previous GM-VFNS approach [56] such a physical behaviour is obtained only by a particular choice of QCD scales [61, 62].

The final differential cross sections are then calculated by using the FFNS expressions for the explicit $Q\overline{Q}$ production, and for all other channels zero-mass expressions with the mentioned massive kinematics. The subtraction terms discussed above are included to avoid double counting and to ensure proper matching between α_s and PDFs in 3- and 4-flavour schemes. The switch from 3- to 4-flavour scheme is done at the charm-mass threshold. The bottom decays to D^0 are an order of magnitude smaller [63] than the "direct" charm

fragmentation to D^0. Thus, the treatment of the bottom mass is not as critical, and in our present setup we switch from 4- to 5-flavour scheme at the bottom-mass threshold with no matching conditions and ignoring the bottom mass. For the numerical implementation of the described SACOT-m_T scheme the massless NLO matrix elements are obtained from the INCNLO [60] code and the FFNS part with explicit heavy-quark production is obtained from the MNR code [64]. As presented in refs. [52, 63], this framework is in a very good agreement with the ALICE [50, 63] and LHCb [39–41] data for inclusive D-meson production in p+p collisions in a broad rapidity range.

2.2 Powheg+Pythia approach

We will also contrast our results in the SACOT-m_T framework with a Monte-Carlo based NLO computation that is often applied to heavy-meson phenomenology at the LHC in the context of PDFs [44, 45, 65]. This approach is based on the POWHEG method [66] to combine NLO matrix elements with a parton shower and hadronization from a general-purpose Monte-Carlo event generator. The underlying idea is to generate the partonic $2 \to 2$ and $2 \to 3$ events with the NLO-correct matrix elements. These events are then passed to any parton shower generator that provides the rest of the partonic branchings, accounting for the fact that the first one may already have occurred. The parton shower can be considered as being analogous to the scale evolution of FFs and PDFs as the splitting probabilities are based on the DGLAP evolution equations in both cases.

We generate the partonic events with the heavy-quark pair production (HVQ) scenario [67] of the POWHEG BOX framework [68] which we pass on to PYTHIA 8 [69] for showering and hadronization. As POWHEG generates only events where the heavy-quark pair is produced in the Born-level process or in the first (hardest) splitting, it ignores the component where the $Q\overline{Q}$ would be created only later on in the shower e.g. starting from a hard $gg \to gg$ process. Such contributions are, however, effectively included in any GM-VFNS framework via the scale-dependent PDFs and FFs. Since charm quarks are abundantly produced in parton showers at the LHC energies [70], truncating the resummation of the splittings to the first one will underestimate the charmed-meson cross section as pointed out in ref. [52]. Similarly the results in ref. [44] show that the cross sections obtained with this method are below the D-meson data measured by LHCb and a compatibility can be concluded only due to large scale uncertainties. Furthermore, this is bound to result as an overestimate of the sensitivity to low-x PDFs as the neglected contributions with several emissions would always require a higher value of x to produce a heavy meson at a fixed P_T and Y. Therefore these comparisons should be taken as an estimation for the effect of truncating the chain of partonic splittings.

2.3 Reweighting machinery

We will quantify the impact of the single inclusive D^0-meson production data in p+Pb collisions on nuclear PDFs by the Hessian reweighting method [28, 53–55]. The method has recently been discussed at length e.g. in ref. [28] so here we only outline the basic underlying idea. Let us consider a global PDF analysis whose fit parameters a_i are tuned

to minimize a global χ^2 function, $\chi^2_0 = \min \chi^2 = \chi^2\{a_i = a_i^0\}$. The χ^2 function is expanded around the best fit as

$$\chi^2\{a\} \approx \chi^2_0 + \sum_{ij}(a_i - a_i^0)H_{ij}(a_j - a_j^0) = \chi^2_0 + \sum_i z_i^2\,, \tag{2.8}$$

where H_{ij} is the Hessian matrix, $H_{ij} = \frac{1}{2}\partial^2\chi^2/(\partial a_i \partial a_j)$. Denoting by O the orthogonal matrix that diagonalizes the Hessian matrix, $OHO^{\mathrm{T}} = I$, the z_i variables are linear combinations $z_i \equiv O_{ij}(a_j - a_j^0)$. We refer to the best-fit as S_0, and it corresponds to the point $z = 0$. The Hessian error sets $S_k^{\pm}$ can then be defined by $z_i(S_k^{\pm}) = \pm\sqrt{\Delta\chi^2}\delta_{ik}$, where $\Delta\chi^2$ is the estimated tolerance. It follows [71] that for any PDF-dependent quantity X there are unique points in the z space that extremize its positive and negative deviations from the central value $X(S_0)$. These deviations, $\Delta X^{\pm}$, are given by

$$\Delta X^{\pm} = \pm\frac{1}{2}\sqrt{\sum_k \left[X(S_k^+) - X(S_k^-)\right]^2}\,. \tag{2.9}$$

This, or its asymmetric version (see later), is normally quoted as the uncertainty in Hessian PDF fits. In a global analysis, the χ^2 contributions of individual data sets are simply summed in the overall χ^2. Thus, if we wish to include a new set of data into our global fit, we just add its contribution to eq. (2.8),

$$\chi^2_{\mathrm{new}} \equiv \chi^2_0 + \sum_k z_k^2 + \sum_{i,j}\left(y_i\{z\} - y_i^{\mathrm{data}}\right)C_{ij}^{-1}\left(y_j\{z\} - y_j^{\mathrm{data}}\right), \tag{2.10}$$

where y_i^{data} denote the new data points with a covariance matrix C_{ij}. The PDF-dependent values $y_i\{z\}$ can now be approximated linearly as

$$y_i\{z\} \approx y_i\left[S_0\right] + \sum_k \frac{\partial y_i[S]}{\partial z_k}\Big|_{S=S_0} z_k \approx y_i\left[S_0\right] + \sum_k \frac{y_i[S_k^+] - y_i[S_k^-]}{2}\frac{z_k}{\sqrt{\Delta\chi^2}}\,, \tag{2.11}$$

and by substituting this into eq. (2.10), we see that χ^2_{new} is still quadratic in variables z_k and has therefore a unique minimum which we denote by $z_k = z_k^{\mathrm{min}}$. Note that we do not need to know the value of χ^2_0. The PDFs $f_i^{\mathrm{new}}(x, Q^2)$ that correspond to this new minimum are obtained by replacing y_i in eq. (2.11) by PDFs,

$$f_i^{\mathrm{new}}(x,Q^2) \approx f_i^{S_0}(x,Q^2) + \sum_k \frac{f_i^{S_k^+}(x,Q^2) - f_i^{S_k^-}(x,Q^2)}{2}\frac{z_k^{\mathrm{min}}}{\sqrt{\Delta\chi^2}}. \tag{2.12}$$

Since we now know χ^2_{new} analytically, we can repeat the original treatment by computing the new Hessian matrix and diagonalizing it exactly the same way as outlined above. As a result, we have an approximation of how a new set of data has affected a set of PDFs and its errors. In comparison to a full global analysis, the advantage of the reweighting technique is that it avoids the time-consuming fitting procedure which, in practice, is only available to the people that performed the PDF analysis itself. In addition, and also importantly, there is no need to implement a potentially CPU-expensive cross-section

computation as a part of the fitting framework or to compute partial cross sections to form three dimensional (x_1,x_2,$\mu^2_{\rm fact}$) grids to facilitate a rapid cross-section evaluation. The downside is that since the reweighting method relies completely on the assumptions made in the prior PDF analysis, including e.g. a specific parametrization which may artificially overestimate the impact in a kinematic region beyond the reach of a given observable.

The Hessian reweighting method sketched above relied on a linear approximation for the PDFs and observables in the z space, and on a quadratic expansion of the original χ^2 function. These are not always good approximations and, as described in ref. [28], the results can be refined by taking into account higher order terms in z. The results presented in this paper (section 3.3) have been obtained using a quadratic extension of the approximation made in eq. (2.11). In the case of EPPS16 we also take into account cubic terms in the original χ^2 profile, eq. (2.8). See ref. [28] for further technical details.

3 Results

Throughout this section, we will use two recent globally-fitted nPDF sets, EPPS16 [22] and nCTEQ15 [21] in our calculations. In the case of EPPS16 we use CT14NLO [5] as the free proton PDF set and with nCTEQ15 we use its own proton PDF (with no uncertainties on it). As a default setup for the GM-VFNS calculation we adopt the KKKS08 [72] parton-to-hadron FFs and set the renormalization and factorization scales as $\mu_{\rm ren} = \mu_{\rm fact} = \sqrt{P_{\rm T}^2 + m_{\rm c}^2}$ with $m_{\rm c} = 1.3$ GeV for the charm quark mass. For the fragmentation scale we set $\mu_{\rm frag} = \sqrt{P_{\rm T}^2 + (1.5\ {\rm GeV})^2}$ as the KKKS08 analysis assumed this slightly higher value for the charm-quark mass. In the matrix elements we always use $m_{\rm c} = 1.3$ GeV. For the $\rm D^0$ mass, relevant for transforming the partonic kinematics to hadronic ones, we adopt the value $M_{\rm D^0} = 1.87$ GeV [73]. With the POWHEG approach, we use the same nuclear and proton PDFs and the same value for the charm mass but the renormalization and factorization scales are fixed to transverse mass of the produced charm quark, $\sqrt{p_{\rm T}^2 + m_{\rm c}^2}$. At the time of generating the partonic events with POWHEG it is not yet known which $P_{\rm T}$ the D meson will have (if formed at all), so relating the scales to the partonic variables is the only reasonable option. The parton shower and hadronization for the POWHEG events are generated with the PYTHIA version 8.235 [69] using parameters from the default MONASH tune [74].

3.1 Double-differential cross section for $\rm D^0$ production in p+Pb collisions

To benchmark our GM-VFNS framework in p+Pb collisions we first compare our calculations with the double-differential single-inclusive $\rm D^0$ production cross section measured by LHCb [42]. This comparison is important since a good agreement with the measured cross sections would indicate that the framework includes e.g. all the relevant partonic processes. In this way we ensure that the framework is realistic.

In figure 1 we compare the calculated cross sections with the LHCb data at backward rapidities (Pb-going direction) in five different rapidity bins spanning $-5.0 < Y < -2.5$ in the nucleon-nucleon (NN) c.m. frame. The resulting cross sections with the default setup

are shown for both the EPPS16 and nCTEQ15 nPDFs, whereas the theoretical uncertainties are quantified with EPPS16 only. These include now scale variations and PDF uncertainties. The former are calculated by varying the three QCD scales independently by a factor of two around the default choice. In addition, ratios $\mu_{\mathrm{fact}}/\mu_{\mathrm{ren}}$ and $\mu_{\mathrm{frag}}/\mu_{\mathrm{ren}}$ are required to stay within $[0.5, 2]$ and the mass of the charm quark is used as a lower limit for all scales. For the PDF uncertainties the error bands from proton and nuclear PDFs are added in quadrature as they are approximately independent in the EPPS16 global analysis. Here, we use the asymmetric error prescription

$$\Delta X^{+} = \sqrt{\sum_{k} \max\left[X(S_k^{+}) - X(S^0), X(S_k^{-}) - X(S_k^0), 0\right]^2}, \tag{3.1}$$

$$\Delta X^{-} = \sqrt{\sum_{k} \min\left[X(S_k^{+}) - X(S^0), X(S_k^{-}) - X(S_k^0), 0\right]^2}, \tag{3.2}$$

where the sum now runs over both the EPPS16 and CT14NLO error sets. Uncertainties due to the mentioned ambiguity in defining the fragmentation variable z, FFs, or e.g. variation in charm-quark mass are not considered. In addition to the GM-VFNS results, comparison with the POWHEG+PYTHIA setup is shown. The correspondence between the data and the GM-VFNS calculation with both EPPS16 and nCTEQ15 is found to be very good, though the theoretical uncertainties become large at $P_{\mathrm{T}} < 3$ GeV. Interestingly the PDF uncertainty at small P_{T} is large above the central result but small below it. This can be traced back to the parametrization applied in the CT14 analysis where the requirement for positive-definite PDFs limits the small-x behaviour as already the central set for gluons near the initial scale Q_0^2 at small x is close to zero. Since similar positivity restriction was not applied in NNPDF3.1 [7], the PDF uncertainties shown in ref. [52] behave in a different manner at small values of P_{T}. As in the p+p case [52], the cross sections obtained from the POWHEG+PYTHIA setup fall below the data. As discussed in ref. [52] and mentioned in the preceding section, this likely follows from truncating the collinear splittings producing $Q\bar{Q}$ pairs after the hardest one. The corresponding cross sections at forward rapidities (p-going direction) are shown in figure 2. Here the five rapidity bins cover the range $1.5 < Y < 4.0$. The conclusions are very similar as at backwards rapidities, the agreement between the GM-VFNS calculation and the data being very good, particularly at $P_{\mathrm{T}} \gtrsim 3$ GeV where the theoretical uncertainties are in control. The comparisons with the absolute cross sections lead us to conclude that the SACOT-m_{T} framework [52] works very well also for p+Pb collisions and can be faithfully applied to study the nPDF constraints — at least for $P_{\mathrm{T}} \gtrsim 3$ GeV.

3.2 Nuclear modification ratio for D^0 production in p+Pb collisions

To constrain nPDFs with D mesons, it is useful to consider an observable in which theoretical uncertainties related to scale variations, free proton PDFs, and FFs cancel out to a large extent. In the case of single-inclusive hadron production a suitable observable is the nuclear modification factor $R_{AB}^{h_3}$, defined for inclusive D^0 meson production in p+Pb collisions at

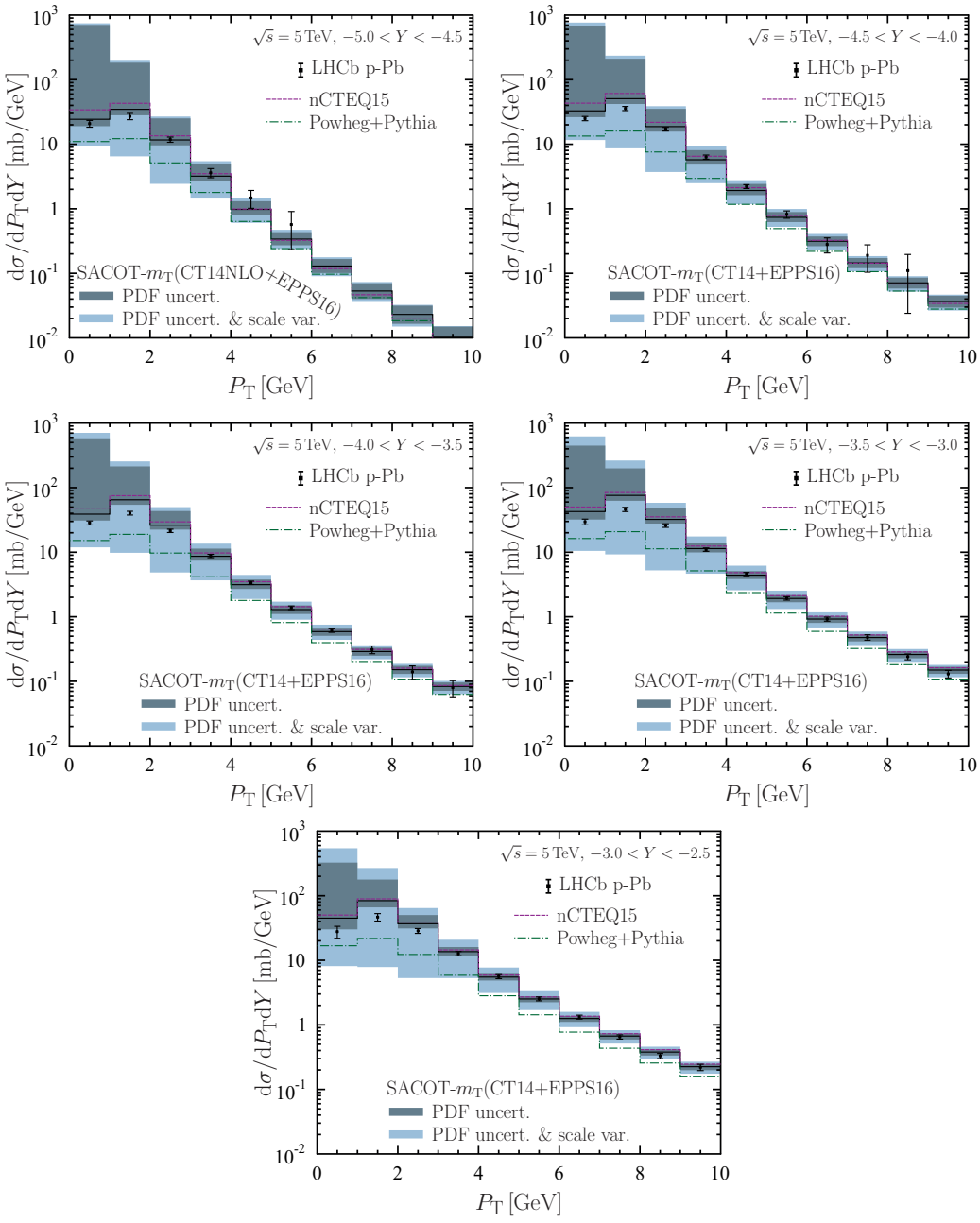

Figure 1. Double-differential cross section for D^0 production in p+Pb collisions at $\sqrt{s_{\text{NN}}} = 5.0$ TeV in five different Y bins at backward rapidities. Data from LHCb [42] are compared to the GM-VFNS calculations with EPPS16 (solid black) and nCTEQ15 (dashed purple) nPDFs, and to a POWHEG+PYTHIA setup with EPPS16 nPDFs (dot-dashed green). The theoretical uncertainties related to the PDFs are shown with dark grey and the combination of the scale variations and PDF uncertainties with light blue.

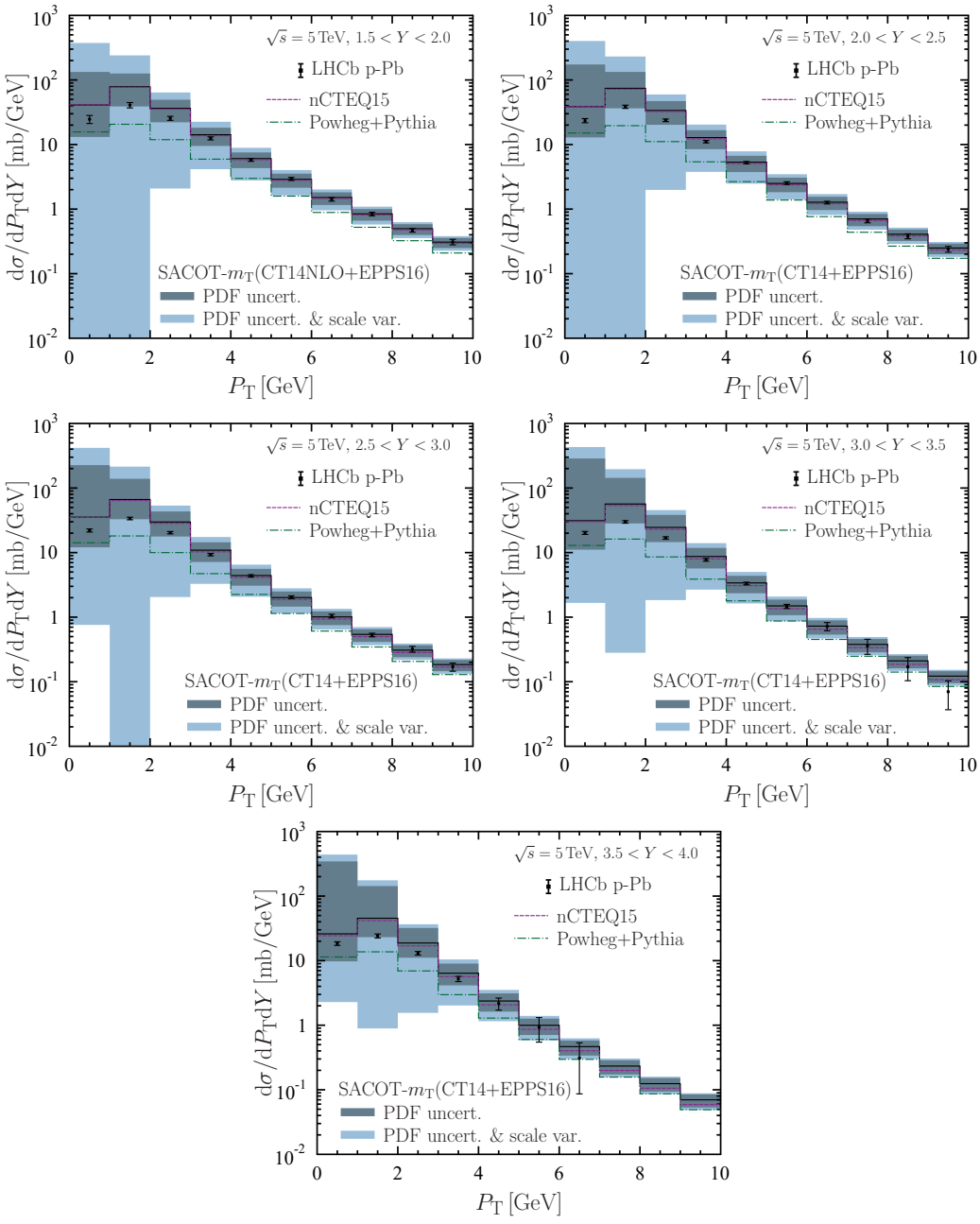

Figure 2. Same as figure 1 but at forward rapidities.

the LHC as

$$R^{\mathrm{D}^0}_{\mathrm{pPb}}(P_{\mathrm{T}}, Y) \equiv \frac{1}{208} \frac{\mathrm{d}\sigma^{\mathrm{p+Pb}\to\mathrm{D}^0+X}}{\mathrm{d}P_{\mathrm{T}}\mathrm{d}Y} \Big/ \frac{\mathrm{d}\sigma^{\mathrm{p+p}\to\mathrm{D}^0+X}}{\mathrm{d}P_{\mathrm{T}}\mathrm{d}Y}. \tag{3.3}$$

We compare our calculations with the measured $R^{\mathrm{D}^0}_{\mathrm{pPb}}$ in figures 3 and 4 at backward and forward rapidities, respectively. The LHCb data span over four Y bins in a range $-4.5 < Y < -2.5$ at backward rapidities and $2.0 < Y < 4.0$ at forward rapidities. Comparisons with the EPPS16 and nCTEQ15 nPDFs using the GM-VFNS framework and POWHEG+PYTHIA setup are separately shown in each panel, and the uncertainty bands correspond to the nPDF errors calculated in the GM-VFNS approach. Furthermore, also the GM-VFNS result using the zero-mass definition for the fragmentation variable, and the scale variation band, are shown in each kinematic bin.

First observation is that the data uncertainties are in most of the cases smaller than the nPDF-originating ones with both nPDF sets considered. Especially at forward rapidities the EPPS16 nPDF uncertainty bands are much larger than the experimental uncertainties due to the poorly-constrained small-x nuclear gluon distributions. This demonstrates the potential of these data to significantly constrain the current nPDFs at small-x where no other data currently exist. Also, the good overall agreement with the calculated and measured $R^{\mathrm{D}^0}_{\mathrm{pPb}}$ over the wide rapidity range provides a strong indication of the applicability of factorization-based approach in this previously unconstrained kinematic region. The large uncertainties from scale variations observed for the differential cross sections largely cancel out in the nuclear modification ratio. However, at $P_{\mathrm{T}} < 3$ GeV they start to grow and the downward uncertainty is limited by the minimum scale $Q = 1.3$ GeV of EPPS16 and nCTEQ15. If the PDF parametrizations would extend to lower values, the downward uncertainty would probably be much larger. Similarly, the use of massless definition for the fragmentation variable z — taken here as an indicator of the associated uncertainty — can lead to a significant variation in the calculated $R^{\mathrm{D}^0}_{\mathrm{pPb}}$ at small values of P_{T} at backward rapidities. The reason is that the definition of z provides the link between hadronic and partonic kinematics and therefore the probed x regions are slightly different from one definition to another. In backward direction we are sensitive to the mid-x region where the slope in both EPPS16 and nCTEQ15 nuclear gluon modifications is somewhat steepish (see figures 9 and 11 ahead), and changes in the probed x regions matter. To make sure that we stay in a region where these theoretical uncertainties are in control, it seems sufficient to discard the data points below $P_{\mathrm{T}} = 3$ GeV.

Since many theoretical uncertainties get suppressed in $R^{\mathrm{D}^0}_{\mathrm{pPb}}$, we might expect that the POWHEG+PYTHIA results would be very close to GM-VFNS ones. While the two are indeed very similar, we find that the POWHEG+PYTHIA results tend to lie systematically below the GM-VFNS calculations. In part, the differences can be explained by the different scale choices (p_{T} instead of P_{T}) but since the differences persist even at the largest P_{T} bins, this cannot be the full explanation. Indeed, the main factor seems to be, as argued also in ref. [52], that POWHEG+PYTHIA framework misses the contributions in which the $\mathrm{c\bar{c}}$ pair would be produced only at later stages of the shower and therefore biases the kinematics to lower values of x_2 compared to the GM-VFNS calculation. Thus, the nuclear effects in the POWHEG+PYTHIA predictions at a given P_{T} come from smaller x_2 than in

GM-VFNS. This explains why, when compared to the GM-VFNS results, the nuclear effects in POWHEG+PYTHIA predictions are seemingly shifted towards higher values of P_{T} in all rapidity bins, apart from the very lowest P_{T} bins where the impact of the scale choice becomes important. We again emphasize that the difference between the two frameworks should not be taken as an additional theoretical uncertainty but as a measure of the effect arising from truncating the series of collinear partonic splittings.

3.3 Impact of the LHCb data on nPDFs

The observed consistency between the measured and calculated $R_{\mathrm{pPb}}^{\mathrm{D}^0}$ indicates that these data could be used in a global nPDF analysis. As a preparation for this, we now estimate the impact of the LHCb data for $R_{\mathrm{pPb}}^{\mathrm{D}^0}$ on the EPPS16 and nCTEQ15 nPDFs by applying the reweighting method outlined in section 2.3. By excluding the data points at $P_{\mathrm{T}} < 3$ GeV we are left with $N_{\mathrm{data}} = 48$ data points. The level of agreement is quantified by calculating the standard figure-of-merit χ^2 before and after reweighting. The numbers are presented in table 1. Before the reweighting, the central nCTEQ15 value is somewhat high, but upon performing the reweighting both the EPPS16 and nCTEQ15 values are close to unity, indicating a good agreement with the data. To further study the statistical properties of our

Table 1. Values of χ^2/N_{data} for the EPPS16 and nCTEQ15 nPDFs before and after reweighting.

χ^2/N_{data}	EPPS16	nCTEQ15
before reweighting	1.56	2.09
after reweighting	1.02	1.12

results, histograms of the data residuals are shown in figure 5. The residuals are calculated (for uncorrelated errors) as a difference between the theory value T_i and corresponding data point D_i normalised with the experimental uncertainty δ_i. Ideally the distribution of the residuals should follow a Gaussian distribution with standard deviation of one and zero mean to which the calculated values are compared to. In addition, Gaussian fits are performed for the residuals obtained after reweighting to ease the comparison with the ideal distributions. With the original central EPPS16 and nCTEQ15 results the distributions show a behaviour diverting from the ideal Gaussian, but after reweighting a closer resemblance to that is obtained. With both nPDF sets the resulting distributions are slightly narrower than the ideal distribution but the mean is close to zero, confirming a reasonable statistical behaviour.

The results for $R_{\mathrm{pPb}}^{\mathrm{D}^0}$ after reweighting, compared with the data and original predictions, are shown in figures 6 and 7. As expected, the reweighted results are in an excellent agreement with the data across the wide rapidity range covered by the data, the only exception being the most backward bin where the data show a stronger enhancement than the reweighted PDF predictions. The new nPDF uncertainties computed from the reweighted nPDFs are significantly reduced in comparison to the original error bands. This holds especially at forward rapidities where the small-x region with no previous data

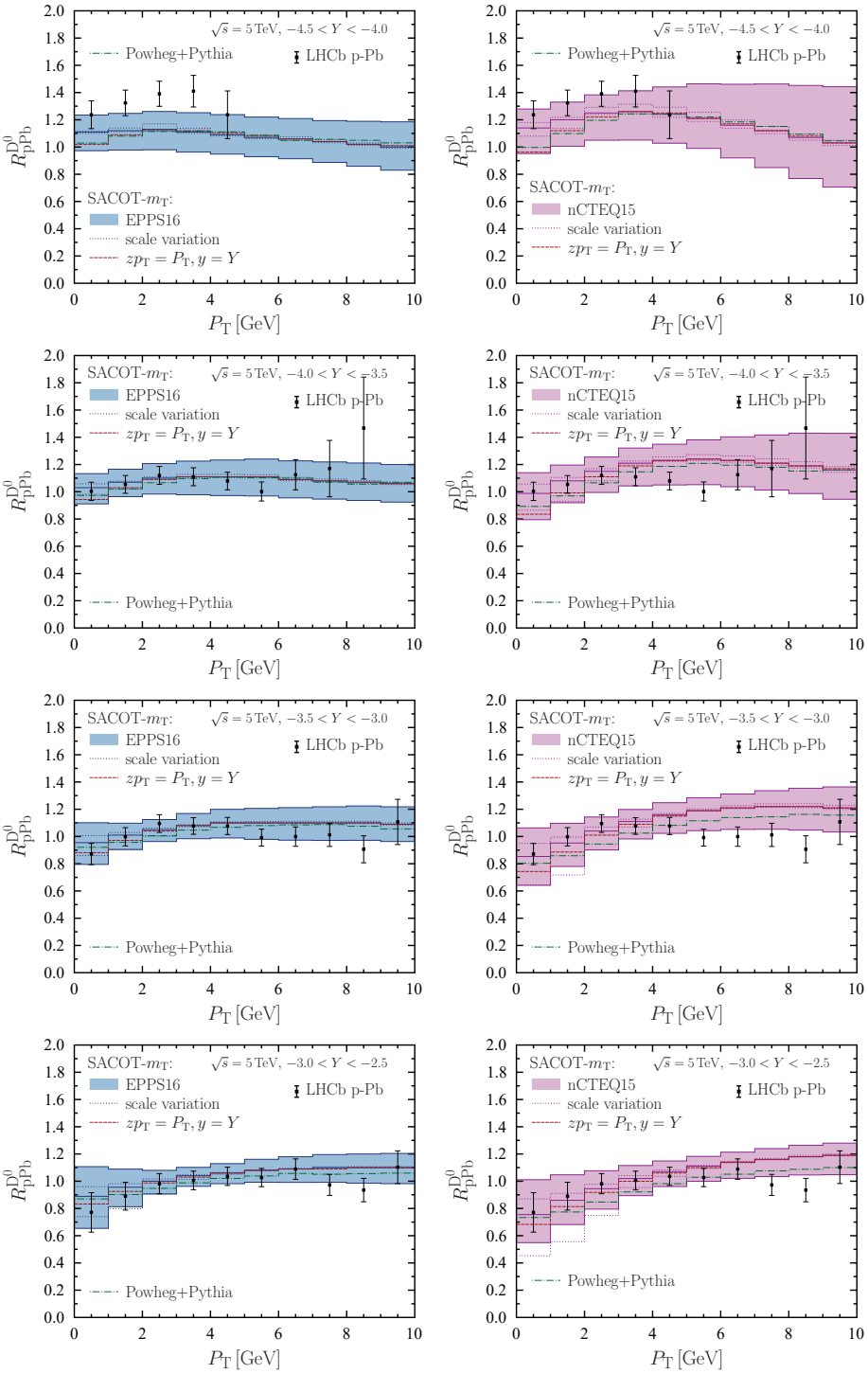

Figure 3. Nuclear modification ratio for D^0 production in p+Pb collisions in different backward-rapidity bins from the LHCb measurement [42] (black points with error bars) and the SACOT-m_T calculation with the EPPS16 (left) and nCTEQ15 (right) nPDFs. In addition to the central result (solid) and the nPDF-originating uncertainties (coloured bands), the scale variations (dotted band) and the result with massless definition of the fragmentation variable (dashed) are shown, as well as the Powheg+Pythia predictions (dot-dashed).

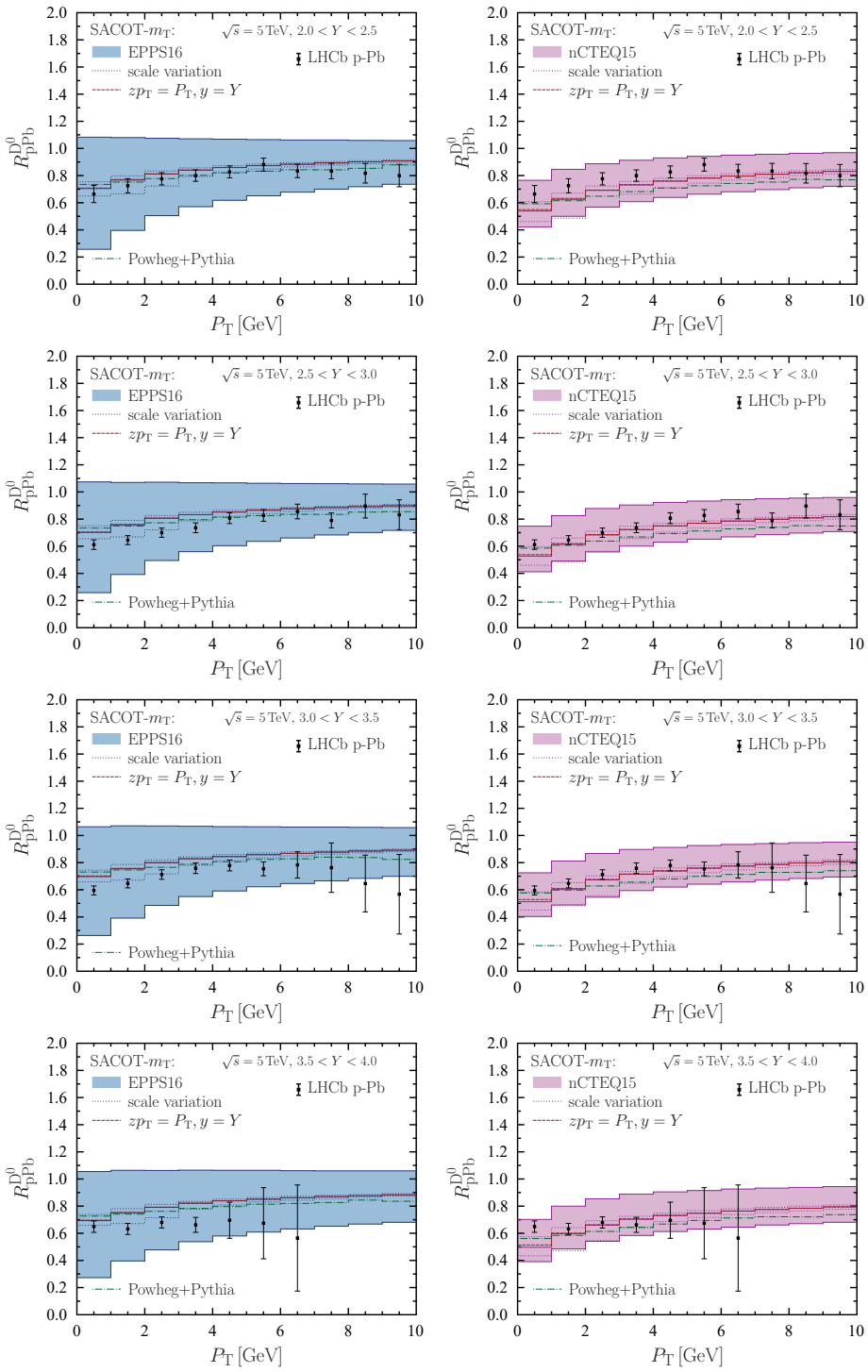

Figure 4. Same as figure 3 but at forward rapidities.

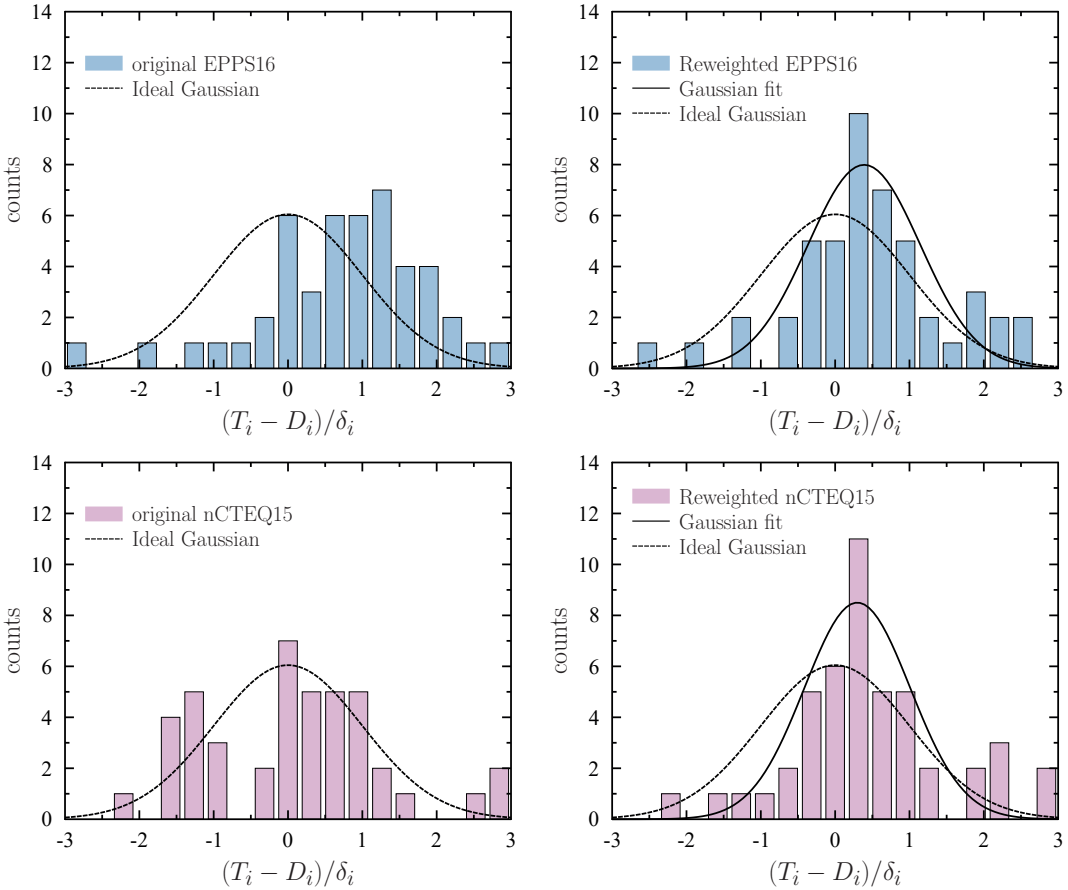

Figure 5. The histograms of the $R_{\mathrm{pPb}}^{\mathrm{D}^0}$-data residuals obtained before (left) and after (right) the reweighting with EPPS16 (top) and nCTEQ15 (bottom). Ideal Gaussian distributions (dashed) are compared to fitted ones (solid) in the reweighted case.

constraints, is probed. For the EPPS16 nPDFs an improvement of a factor of three is observed whereas for nCTEQ15 the improvement is somewhat more modest. This difference follows from a bit more rigid functional form of the nCTEQ15 parametrization which leads to smaller errors to begin with. Interestingly, even though the lowest-P_{T} bins were not included in the analysis, the agreement remains very good also with the data points in the $P_{\mathrm{T}} < 3$ GeV region. We can thus conclude that to describe these data, no physics outside collinear factorization is needed.

In figures 8 – 11 we finally compare the EPPS16 and nCTEQ15 nuclear modifications in bound protons, $R_i^{\mathrm{p/Pb}}(x, Q^2) = f_i^{\mathrm{Pb}}(x, Q^2)/f_i^{\mathrm{p}}(x, Q^2)$, before and after reweighting. We present the results at two different scales: the initial scale of the original analyses, $Q^2 = 1.69$ GeV^2, and a somewhat higher scale $Q^2 = 10$ GeV^2 directly probed by the considered observable when reweighting to the $P_{\mathrm{T}} > 3$ GeV subset of data. The valence and sea quark distributions are shown separately for each partonic flavour. For the EPPS16 analysis these are plotted in figures 8 and 9. The central values remain unchanged for all

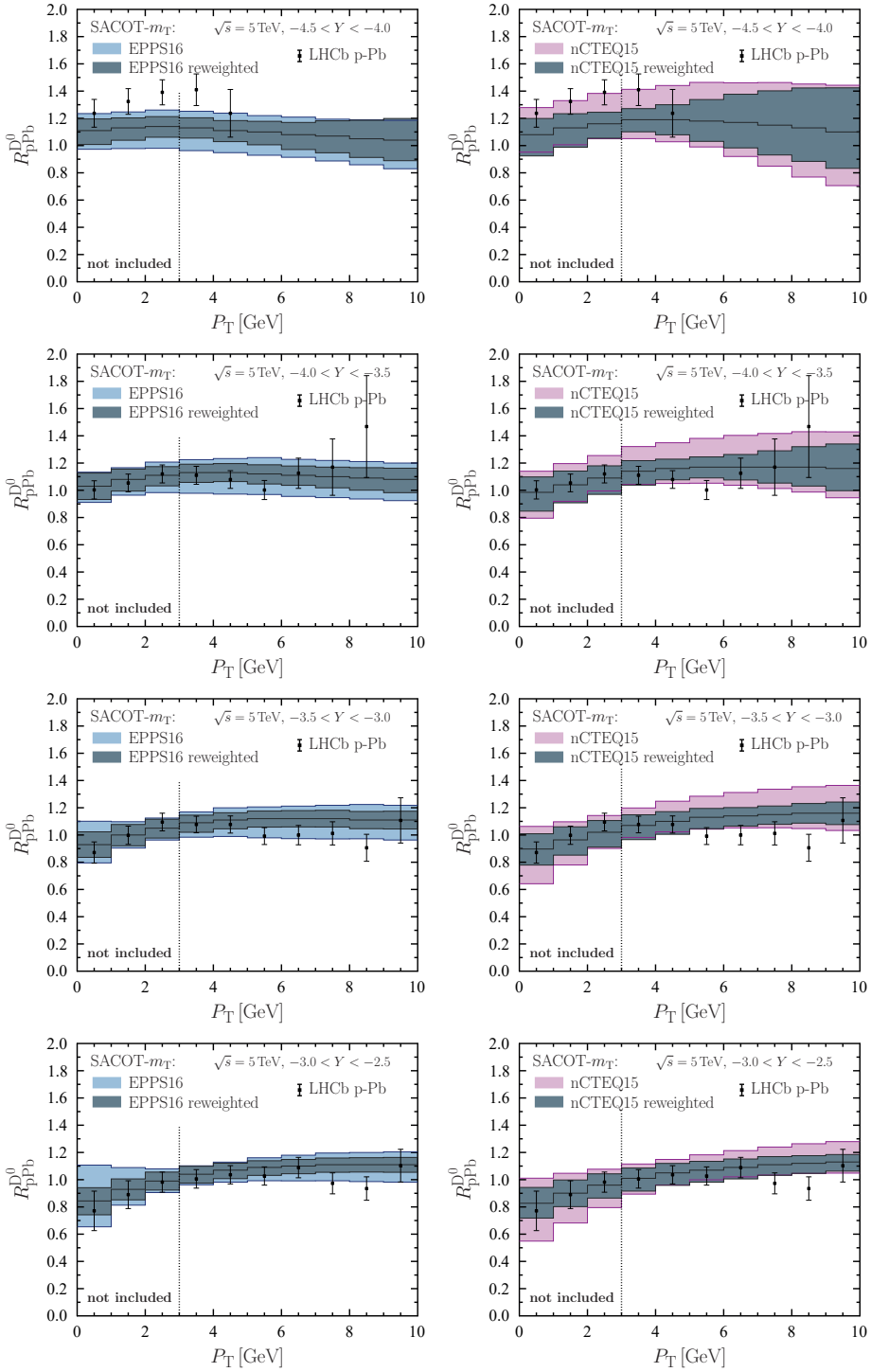

Figure 6. Nuclear modification ratio for D^0 production at backward rapidities in p+Pb collisions at $\sqrt{s_{NN}} = 5.0$ TeV from the LHCb measurement [42] (black points with error bars) compared with the SACOT-m_T calculation using the EPPS16 (left) and nCTEQ15 (right) nPDFs with uncertainties before (light-coloured bands) and after reweighting (dark-grey bands) including the central result from the reweighted nPDFs (solid).

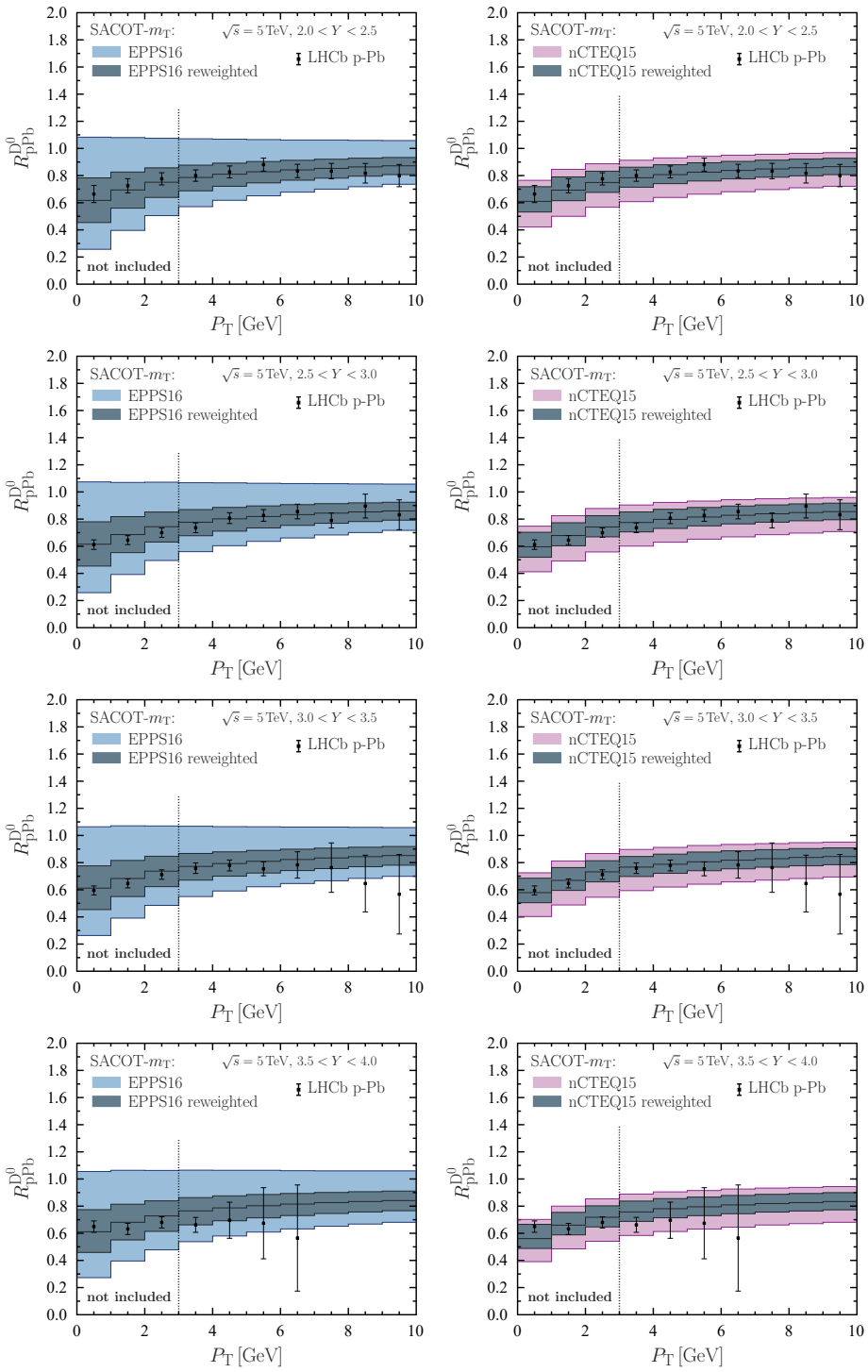

Figure 7. Same as in figure 6 but at forward rapidities.

quark flavours but for gluons a somewhat stronger shadowing and slightly weaker EMC suppression are preferred by the data. At the parametrization scale $Q^2 = 1.69$ GeV^2 the uncertainty bands remain practically unchanged for quarks but a drastic reduction is observed for small-x gluons. At $Q^2 = 10$ GeV^2 also the sea-quark uncertainties are slightly reduced due to the DGLAP evolution which correlates sea quarks with gluons. For gluons the strong shadowing at the initial scale is reduced to around 0.7 at $x \lesssim 0.01$ due to the evolution effects. Incidentally, the changes in the EPPS16 gluon PDFs are remarkably similar as found in ref. [28] based on the recent CMS dijet data [75]. In addition, since the central values are only slightly modified, the good agreement with the recent $\mathrm{W}^{\pm}$ data at $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV [76] is expected to persist. We should also mention that the gluon errors at $Q^2 = 1.69$ GeV^2 dropping negative is of no concern. Indeed, a backward evolution by the DGLAP equations will make any gluon PDF negative at sufficiently low scales, and demanding a positive-definite gluon distribution at any arbitrary scale would be an unphysical requirement. At a deeper level, the resummation of $\log(1/x)$ terms in the DGLAP splitting functions [77] may slow down the evolution speed particularly at low Q^2 and thereby better retain the gluons positive.

For nCTEQ15 the original and D-meson updated nuclear modifications are plotted in figures 10 and 11. As was the case with EPPS16, the quark nuclear modifications remain more or less the same after reweighting with the LHCb data. The originally strong shadowing for small-x gluons becomes slightly weaker after reweighting and is now rather similar to the gluon shadowing obtained with the reweighted EPPS16. The resulting uncertainties for the gluon shadowing are also on the same ballpark with with EPPS16. In addition, the reweighted nCTEQ15 nuclear modifications for gluons tend to have somewhat less anti-shadowing (the bump around $x \sim 0.1$) than in the original analysis and the uncertainties are significantly reduced also in this regime.

3.4 Impact without the lower cut on P_{T}

The agreement between the measured and calculated $R_{\mathrm{pPb}}^{\mathrm{D}^0}$ was found to be very good also at $P_{\mathrm{T}} < 3$ GeV which we excluded from the reweighting due to theoretical concerns. To check how much potential constraints we threw away, we have repeated the reweighting procedure this time including all the LHCb data. The resulting gluon nPDFs at $Q^2 = 1.69$ GeV^2 and $Q^2 = 10$ GeV^2 are shown in figure 12 for EPPS16 and nCTEQ15. Effect for quark nPDFs was found negligible at $Q^2 = 1.69$ GeV^2. In both cases the reweighted central results remain practically unchanged but the uncertainties are further reduced at small x in the case of EPPS16 and also at larger x in the case of nCTEQ15. However, the bulk part of the uncertainty reduction still comes from the data in the "safe region" $P_{\mathrm{T}} > 3$ GeV such that inclusion of the $P_{\mathrm{T}} < 3$ GeV data is not critical. As we will argue next, including the lower P_{T} data would not even increase the sensitivity to the small x region significantly.

3.5 Sensitivity to small-x region

The x values probed by a given P_{T} and Y are often in the literature estimated with simplified leading-order kinematics, see e.g. ref. [46]. To get a more complete understanding on the small-x sensitivity of D^0 production at forward rapidities we show the contributions from

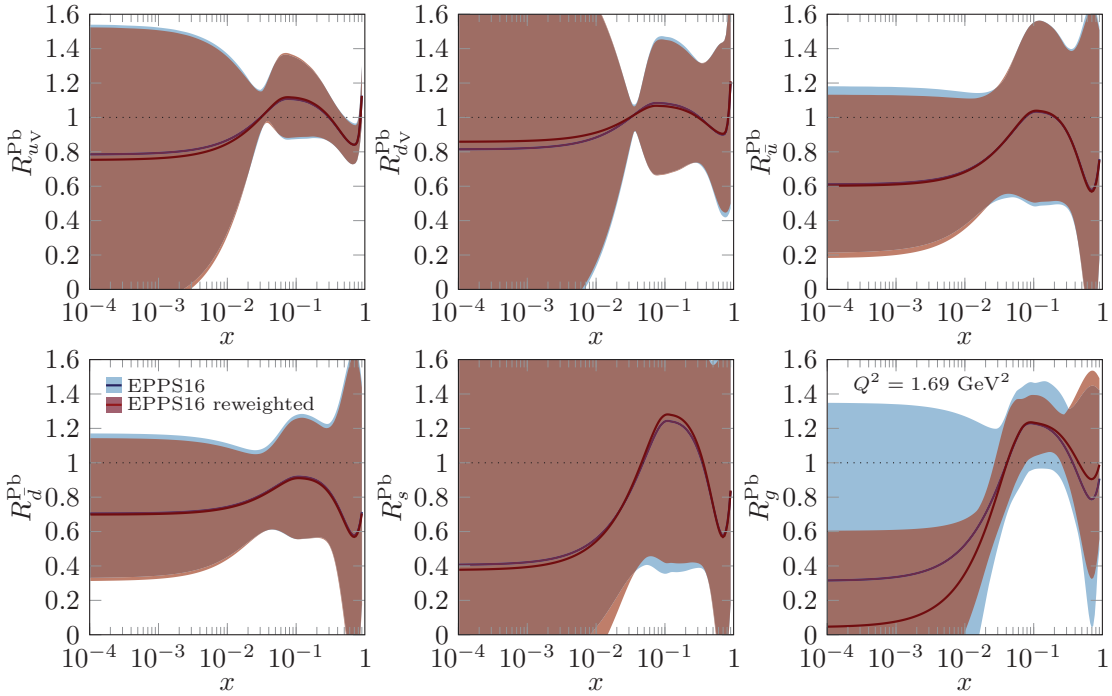

Figure 8. The EPPS16 nuclear modifications for bound-proton PDFs in Pb nucleus before (blue) and after (red) reweighting with the LHCb data. The scale is $Q^2 = 1.69$ GeV^2.

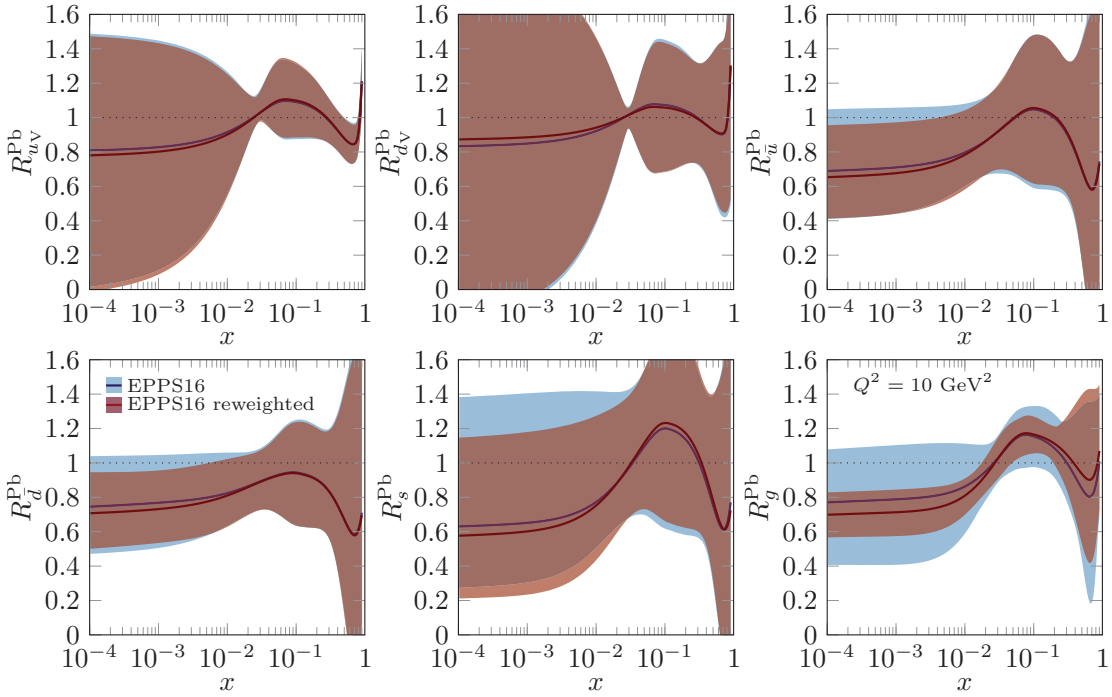

Figure 9. The EPPS16 nuclear modifications for bound-proton PDFs in Pb nucleus before (blue) and after (red) reweighting with the LHCb data. The scale is $Q^2 = 10$ GeV^2.

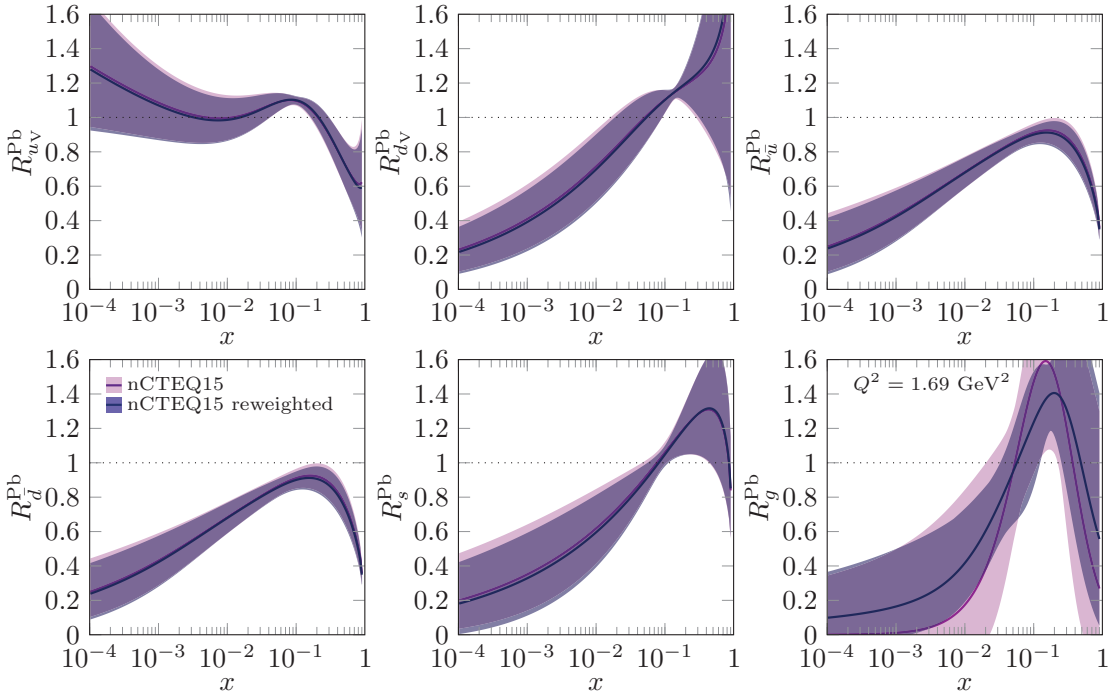

Figure 10. The nCTEQ15 nuclear modifications for bound-proton PDFs in Pb nucleus before (purple) and after (blue) reweighting with the LHCb data. The scale is $Q^2 = 1.69$ GeV^2.

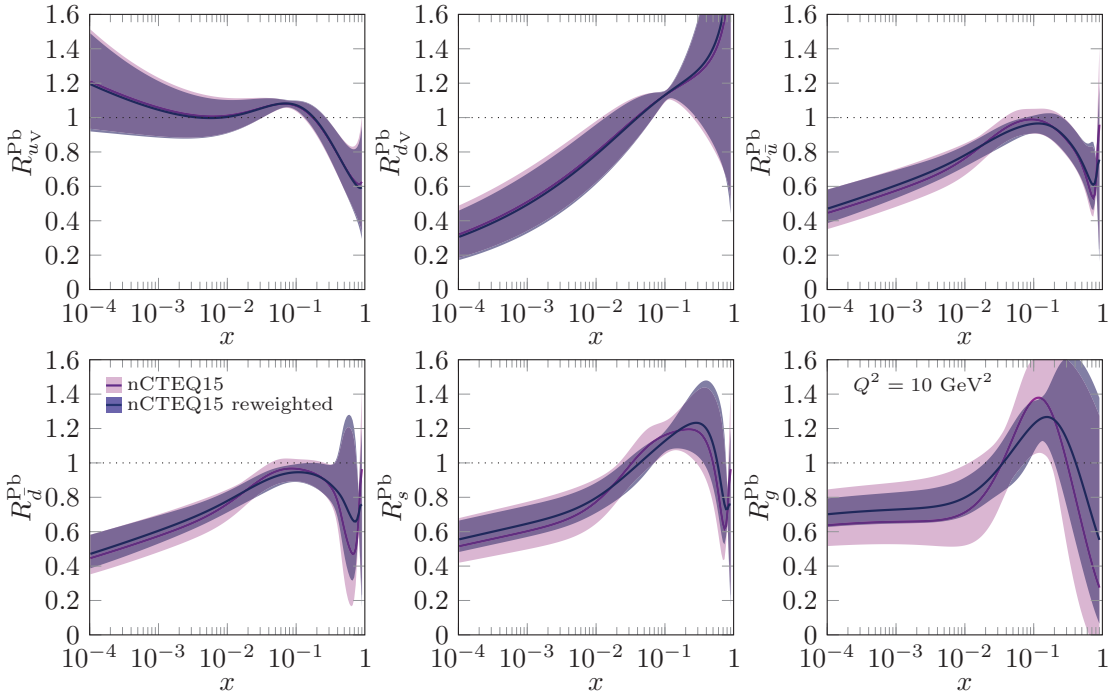

Figure 11. The nCTEQ15 nuclear modifications for bound-proton PDFs in Pb nucleus before (purple) and after (blue) reweighting with the LHCb data. The scale is $Q^2 = 10$ GeV^2.

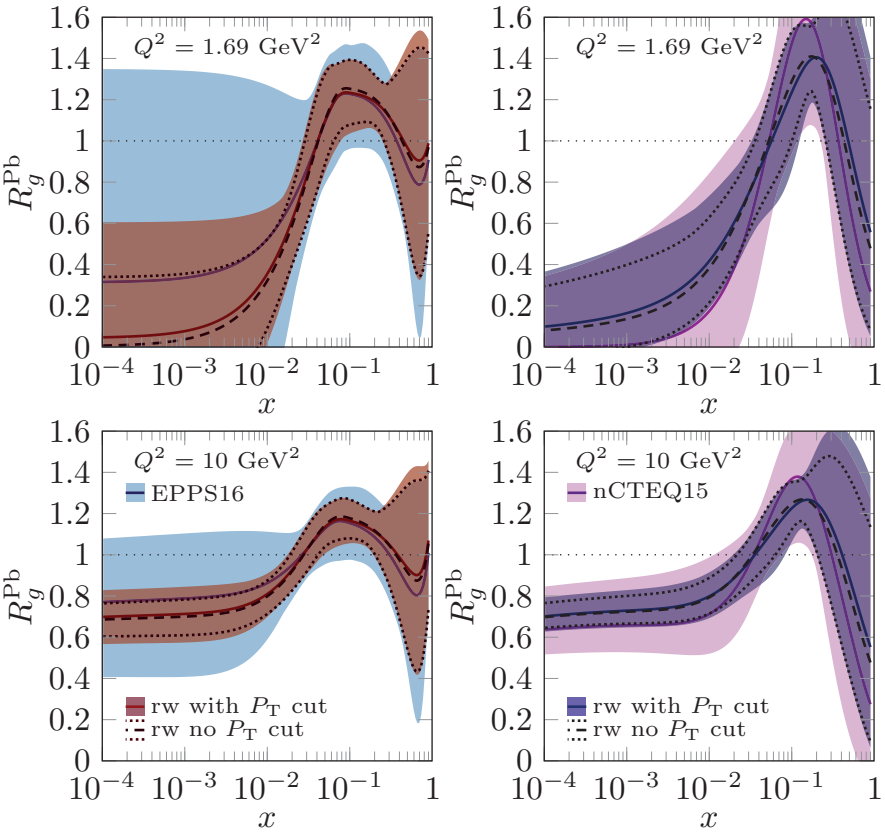

Figure 12. The EPPS16 (left) and nCTEQ15 (right) nuclear modifications for bound-proton PDFs in Pb nucleus before (EPPS16 blue, nCTEQ15 purple), after reweighting with the LHCb data with $P_T > 3$ GeV (EPPS16 red, nCTEQ15 blue), and including all data points (dotted curves). The results are shown at $Q^2 = 1.69$ GeV^2 (upper panels) and at $Q^2 = 10$ GeV^2 (lower panels).

different values of x_2 (momentum fraction in nucleus) to the D^0 cross section in figure 13. These distributions are based on full NLO GM-VFNS calculation with EPPS16 including the convolution with fragmentation functions. The results are compared to distributions from a "matrix-element fitting" approach similar to the one introduced in ref. [78] and applied in ref. [48] to study the impact of the LHCb data on nPDFs. In the latter method the squared matrix element $|\mathcal{M}|^2$ for D-meson production is parametrized and the parameters are fitted to data from p+p collisions assuming that the only contribution is gluon-gluon initiated $2 \to 2$ scattering. The parameters used for the result in figure 13 are obtained from ref. [78] but the correspondence is not guaranteed to be exact since the details of the applied two-body phase space are not explicitly defined in the reference. However, the main point here is that the assumed $x_{1,2}$ dependence which, together with PDFs, dictates the shape of the x distributions is rather trivial, of the form $|\mathcal{M}|^2 \propto x_1 x_2$. The x distributions from the full NLO GM-VFNS calculation are shown for P_T-integrated case with and without the lower cut of $P_T > 3$ GeV. As expected, the D^0 meson production at forward rapidities is indeed sensitive to small-x region reaching down to 10^{-5} in the

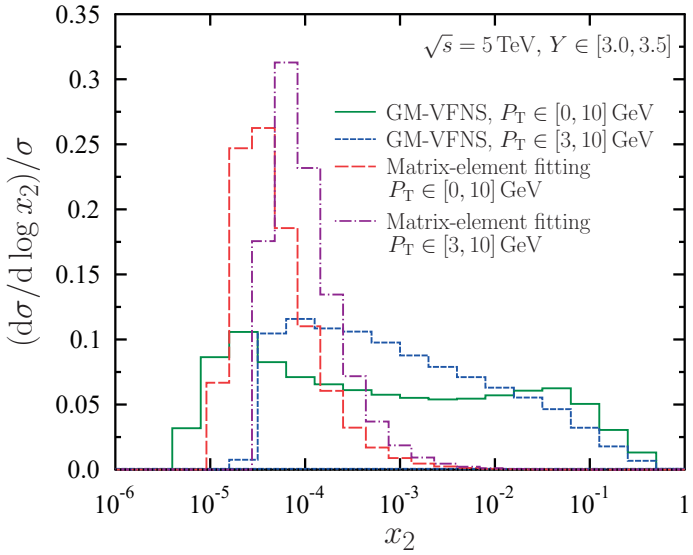

Figure 13. Contributions to differential D^0 cross section from different values of x_2 at $3.0 < Y < 3.5$ from the GM-VFNS in P_T ranges of $[0, 10]$ GeV (solid green) and $[3, 10]$ GeV (short-dashed blue) and from matrix-element fitting approach for same P_T ranges (long-dashed red and dot-dashed purple).

considered $3.0 < Y < 3.5$ bin. However, there is still a significant contribution from larger x. These large-x tails mainly arise from the convolutions with the fragmentation functions which smears the connection between partonic and hadronic kinematics. Also the NLO corrections contribute to the tail as discussed in ref. [52]. Maybe a bit surprisingly, the tail extends to higher values of x when no lower cut on P_T is applied. A very similar behaviour has been seen in the case of inclusive photon production [33]. In part, this can be explained by the valence-like gluons at low scales which shift the cross section to higher x region. In addition, the nuclear effects in EPPS16 are most pronounced at low scales and the shadowing further suppresses the contributions from small x, whereas anti-shadowing tends to increase the larger-x tail. All this dilutes the extra small-x constraints that could be obtained by releasing the $P_\text{T} > 3$ GeV cut. Thus, a significant part of the reduced small-x uncertainties in figure 12 can be explained just by the increased statistics (24 data points more) rather than pushing to smaller x. These long large-x tails are not visible in the distributions obtained with the matrix-element fitting approach as it assumes leading-order partonic kinematics and, in particular, a naive $|\mathcal{M}|^2 \propto x_1 x_2$ behaviour of the coefficient function. Thus the matrix-element fitting approach would overestimate the sensitivity of the LHCb data on the small-x PDFs and would lead to an overly optimistic impact at small x if used in a global analysis. This underlines the importance of using a proper calculation in order to realistically estimate the impact of D-meson data on PDFs.

4 Summary

We have presented the first direct QCD analysis of the recent LHCb data [42] for D^0 meson production in p+Pb collisions and their impact on nuclear PDFs. To accomplish this we have used the Hessian reweighting method and the cross sections calculated within GM-VFNS using the recently introduced SACOT-m_{T} scheme at NLO [52]. The advantage of the new scheme over the previous GM-VFNS implementations is that by explicitly including the heavy-quark masses in the kinematics also for processes where the $Q\overline{Q}$ pair is produced from light-flavour fragmentation, a sensible behaviour in the $P_{\mathrm{T}} \to 0$ limit is always obtained. The resulting cross sections are in a very good agreement with the single-inclusive D-meson P_{T} spectra in the wide rapidity range covered by the LHCb measurement. We also computed predictions by a frequently used POWHEG approach in which the heavy quarks are first produced in the partonic $2 \to 2$ and $2 \to 3$ scattering events, and then showered and hadronized with PYTHIA. This approach undershoots the absolute differential cross sections roughly by a factor 2. We attribute this to the omission of contributions in which the heavy-quark is produced in $2 \to 4$ processes and beyond. These are resummed in GM-VFNS.

A very good agreement with the $R_{\mathrm{pPb}}^{\mathrm{D}^0}$ data is found with both of the considered nPDF analyses, EPPS16 and nCTEQ15, and the data are accurate enough to set significant further constraints. For quark PDFs the modifications in the central values are weak but for gluons a somewhat stronger (weaker) small-x shadowing than originally in EPPS16 (nCTEQ15) is preferred by the data. The reweighting also brings the gluon shadowing in these two nPDF sets into a better mutual agreement. The main impact of the data is, however, the substantial reduction of the uncertainties for gluon nuclear modifications at $x < 0.01$. In fact, these are the first data directly sensitive to small-x gluons in heavy nuclei at clearly perturbative scales, and therefore provide the first unambiguous direct evidence for nuclear gluon shadowing in the context of a global analysis. The backward data seem to confirm the presence of a moderate gluon antishadowing at large x. We note that the effect of these data on EPPS16 are remarkably similar as recently found from dijet data at significantly higher interaction scales, though there the region $x < 0.002$ is not directly probed [28].

By studying how the cross section builds up from different values of nuclear x we have shown that the LHCb D^0 data constrain nPDFs down to $x \sim 10^{-5}$ but, due to the convolution with FFs, there is still a notable contribution from the high-x region. The importance of using a full QCD calculation to quantify the impact of D-meson data was also underlined. Indeed, a simplified framework can lead to an apparent increase in the sensitivity to the small-x region and would therefore not provide a realistic estimation of the constraints. The good agreement between the nPDF calculation and the data down to $P_{\mathrm{T}} = 0$ GeV — even when rejecting data points at $P_{\mathrm{T}} < 3$ GeV from the fit — implies that the pure collinear-factorization approach is valid also in the small-x region. All in all, we conclude that the LHCb D-meson data can be included in future updates of global nPDF analyses without causing conflicts with the other existing data. To more deeply test the factorization and the universality of nPDFs, data with similar x-reach but for a different observable would be crucial.

Acknowledgments

We wish to thank Michael Winn and Yanxi Zhang for discussions and help concerning the LHCb data. The Academy of Finland projects 297058 (K.J.E.) and 308301 (H.P.), and the Magnus Ehrnrooth Foundation (P.P.) are acknowledged for financial support. The Finnish IT Center for Science (CSC) is acknowledged for the computing time within the project jyy2580.

References

[1] 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 [hep-ph/0409313].

[2] K. Kovarik, P. M. Nadolsky and D. E. Soper, *Hadron structure in high-energy collisions*, arXiv:1905.06957.

[3] J. Gao, L. Harland-Lang and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, *Phys. Rept.* **742** (2018) 1 [arXiv:1709.04922].

[4] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J.* **C 75** (2015) 204 [arXiv:1412.3989].

[5] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.* **D 93** (2016) 033006 [arXiv:1506.07443].

[6] H1, ZEUS collaboration, *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J.* **C 75** (2015) 580 [arXiv:1506.06042].

[7] NNPDF collaboration, *Parton distributions from high-precision collider data*, *Eur. Phys. J.* **C 77** (2017) 663 [arXiv:1706.00428].

[8] S. Alekhin, J. Blümlein and S. Moch, *NLO PDFs from the ABMP16 fit*, *Eur. Phys. J.* **C 78** (2018) 477 [arXiv:1803.07537].

[9] H. Paukkunen, *Nuclear PDFs Today*, in *Proceedings of the 9th International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions: Hard Probes 2018 (HP2018): Aix-Les-Bains, France, October 1-5, 2018*, PoS HardProbes2018 014, 2018, arXiv:1811.01976, DOI.

[10] R. Abdul Khalek, J. J. Ethier and J. Rojo, *Nuclear Parton Distributions from Lepton-Nucleus Scattering and the Impact of an Electron-Ion Collider*, arXiv:1904.00018.

[11] H. Khanpour and S. Atashbar Tehrani, *Global Analysis of Nuclear Parton Distribution Functions and Their Uncertainties at Next-to-Next-to-Leading Order*, *Phys. Rev.* **D 93** (2016) 014026 [arXiv:1601.00939].

[12] K. J. Eskola, V. J. Kolhinen and P. V. Ruuskanen, *Scale evolution of nuclear parton distributions*, *Nucl. Phys.* **B 535** (1998) 351 [hep-ph/9802350].

[13] PHENIX collaboration, *Centrality dependence of π^0 and η production at large transverse momentum in $\sqrt{s_{\rm NN}} = 200$ GeV d+Au collisions*, *Phys. Rev. Lett.* **98** (2007) 172302 [nucl-ex/0610036].

[14] STAR collaboration, *Identified hadron spectra at large transverse momentum in p+p and d+Au collisions at* $\sqrt{s_{\rm NN}} = 200$ GeV, *Phys. Lett.* **B 637** (2006) 161 [nucl-ex/0601033].

[15] STAR collaboration, *Inclusive* π^0*,* η*, and direct photon production at high transverse momentum in* $p + p$ *and d+Au collisions at* $\sqrt{s_{\rm NN}} = 200$ GeV, *Phys. Rev.* **C 81** (2010) 064904 [arXiv:0912.3838].

[16] PHENIX collaboration, *Spectra and ratios of identified particles in Au+Au and d+Au collisions at* $\sqrt{s_{\rm NN}} = 200$ GeV, *Phys. Rev.* **C 88** (2013) 024906 [arXiv:1304.3410].

[17] D. de Florian and R. Sassot, *Nuclear parton distributions at next-to-leading order*, *Phys. Rev.* **D 69** (2004) 074028 [hep-ph/0311227].

[18] K. J. Eskola, H. Paukkunen and C. A. Salgado, *An Improved global analysis of nuclear parton distribution functions including RHIC data*, *JHEP* **07** (2008) 102 [arXiv:0802.0139].

[19] K. J. Eskola, H. Paukkunen and C. A. Salgado, *EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions*, *JHEP* **04** (2009) 065 [arXiv:0902.4154].

[20] D. de Florian, R. Sassot, P. Zurita and M. Stratmann, *Global Analysis of Nuclear Parton Distributions*, *Phys. Rev.* **D 85** (2012) 074028 [arXiv:1112.6324].

[21] K. Kovarik et al., *nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework*, *Phys. Rev.* **D 93** (2016) 085037 [arXiv:1509.00792].

[22] 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 [arXiv:1612.05741].

[23] CMS collaboration, *Study of Z boson production in pPb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ TeV, *Phys. Lett.* **B 759** (2016) 36 [arXiv:1512.06461].

[24] ATLAS collaboration, *Z boson production in p+Pb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ TeV *measured with the ATLAS detector*, *Phys. Rev.* **C 92** (2015) 044915 [arXiv:1507.06232].

[25] CMS collaboration, *Study of W boson production in pPb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ TeV, *Phys. Lett.* **B 750** (2015) 565 [arXiv:1503.05825].

[26] CMS collaboration, *Studies of dijet transverse momentum balance and pseudorapidity distributions in pPb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ TeV, *Eur. Phys. J.* **C 74** (2014) 2951 [arXiv:1401.4433].

[27] Z. Citron et al., *Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams*, in *HL/HE-LHC Workshop: Workshop on the Physics of HL-LHC, and Perspectives at HE-LHC Geneva, Switzerland, June 18-20, 2018*, 2018, arXiv:1812.06772.

[28] K. J. Eskola, P. Paakkinen and H. Paukkunen, *Non-quadratic improved Hessian PDF reweighting and application to CMS dijet measurements at 5.02 TeV*, arXiv:1903.09832.

[29] F. Arleo and T. Gousset, *Measuring gluon shadowing with prompt photons at RHIC and LHC*, *Phys. Lett.* **B 660** (2008) 181 [arXiv:0707.2944].

[30] T. Stavreva, I. Schienbein, F. Arleo, K. Kovarik, F. Olness, J. Y. Yu et al., *Probing gluon and heavy-quark nuclear PDFs with gamma + Q production in pA collisions*, *JHEP* **01** (2011) 152 [arXiv:1012.1178].

[31] F. Arleo, K. J. Eskola, H. Paukkunen and C. A. Salgado, *Inclusive prompt photon production in nuclear collisions at RHIC and LHC*, *JHEP* **04** (2011) 055 [arXiv:1103.1471].

[32] 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 [arXiv:1401.6817].

[33] I. Helenius, K. J. Eskola and H. Paukkunen, *Probing the small-x nuclear gluon distributions with isolated photons at forward rapidities in p+Pb collisions at the LHC*, *JHEP* **09** (2014) 138 [arXiv:1406.1689].

[34] M. Goharipour and S. Rostami, *Probing nuclear modifications of parton distribution functions through the isolated prompt photon production at the LHC*, *Phys. Rev.* **C 99** (2019) 055206 [arXiv:1808.05639].

[35] I. Helenius, J. Lajoie, J. D. Osborn, P. Paakkinen and H. Paukkunen, *Nuclear gluons at RHIC in a multi-observable approach*, arXiv:1904.09921.

[36] ATLAS collaboration, *Measurement of prompt photon production in* $\sqrt{s_{\rm NN}} = 8.16$ TeV *p+Pb collisions with ATLAS*, *Submitted to: Phys. Lett.* (2019) [arXiv:1903.02209].

[37] LHCb collaboration, *Direct photon production at LHCb*, *Nucl. Phys.* **A 982** (2019) 251.

[38] ALICE collaboration, *Upgrade of the ALICE Experiment*, *EPJ Web Conf.* **71** (2014) 00106.

[39] LHCb collaboration, *Prompt charm production in pp collisions at* $\sqrt{s} = 7$ TeV, *Nucl. Phys.* **B 871** (2013) 1 [arXiv:1302.2864].

[40] LHCb collaboration, *Measurements of prompt charm production cross-sections in pp collisions at* $\sqrt{s} = 13$ TeV, *JHEP* **03** (2016) 159 [arXiv:1510.01707].

[41] LHCb collaboration, *Measurements of prompt charm production cross-sections in pp collisions at* $\sqrt{s} = 5$ TeV, *JHEP* **06** (2017) 147 [arXiv:1610.02230].

[42] LHCb collaboration, *Study of prompt* D^0 *meson production in pPb collisions at* $\sqrt{s_{\rm NN}} = 5$ TeV, *JHEP* **10** (2017) 090 [arXiv:1707.02750].

[43] M. Cacciari, M. L. Mangano and P. Nason, *Gluon PDF constraints from the ratio of forward heavy-quark production at the LHC at* $\sqrt{s} = 7$ *and* 13 TeV, *Eur. Phys. J.* **C 75** (2015) 610 [arXiv:1507.06197].

[44] R. Gauld, J. Rojo, L. Rottoli and J. Talbert, *Charm production in the forward region: constraints on the small-x gluon and backgrounds for neutrino astronomy*, *JHEP* **11** (2015) 009 [arXiv:1506.08025].

[45] R. Gauld and J. Rojo, *Precision determination of the small-x gluon from charm production at LHCb*, *Phys. Rev. Lett.* **118** (2017) 072001 [arXiv:1610.09373].

[46] PROSA collaboration, *Impact of heavy-flavour production cross sections measured by the LHCb experiment on parton distribution functions at low x*, *Eur. Phys. J.* **C 75** (2015) 396 [arXiv:1503.04581].

[47] R. Gauld, *Forward D predictions for* pPb *collisions, and sensitivity to cold nuclear matter effects*, *Phys. Rev.* **D 93** (2016) 014001 [arXiv:1508.07629].

[48] A. Kusina, J.-P. Lansberg, I. Schienbein and H.-S. Shao, *Gluon Shadowing in Heavy-Flavor Production at the LHC*, *Phys. Rev. Lett.* **121** (2018) 052004 [arXiv:1712.07024].

[49] G. Kramer and H. Spiesberger, *Study of heavy meson production in p-Pb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ TeV *in the general-mass variable-flavour-number scheme*, *Nucl. Phys.* **B 925** (2017) 415 [arXiv:1703.04754].

[50] ALICE collaboration, *D-meson production in p-Pb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ TeV *and in pp collisions at* $\sqrt{s} = 7$ TeV, *Phys. Rev.* **C 94** (2016) 054908 [arXiv:1605.07569].

[51] ALICE collaboration, *Measurement of the production of charm jets tagged with* D^0 *mesons in pp collisions at* $\sqrt{s} = 7$ TeV, arXiv:1905.02510.

[52] I. Helenius and H. Paukkunen, *Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme*, *JHEP* **05** (2018) 196 [arXiv:1804.03557].

[53] H. Paukkunen and C. A. Salgado, *Agreement of Neutrino Deep Inelastic Scattering Data with Global Fits of Parton Distributions*, *Phys. Rev. Lett.* **110** (2013) 212301 [arXiv:1302.2001].

[54] H. Paukkunen and P. Zurita, *PDF reweighting in the Hessian matrix approach*, *JHEP* **12** (2014) 100 [arXiv:1402.6623].

[55] C. Schmidt, J. Pumplin, C. P. Yuan and P. Yuan, *Updating and optimizing error parton distribution function sets in the Hessian approach*, *Phys. Rev.* **D 98** (2018) 094005 [arXiv:1806.07950].

[56] B. A. Kniehl, G. Kramer, I. Schienbein and H. Spiesberger, *Inclusive D*+- production in p anti-p collisions with massive charm quarks*, *Phys. Rev.* **D 71** (2005) 014018 [hep-ph/0410289].

[57] P. Nason, S. Dawson and R. K. Ellis, *The One Particle Inclusive Differential Cross-Section for Heavy Quark Production in Hadronic Collisions*, *Nucl. Phys.* **B 327** (1989) 49.

[58] C. Peterson, D. Schlatter, I. Schmitt and P. M. Zerwas, *Scaling Violations in Inclusive e+ e- Annihilation Spectra*, *Phys. Rev.* **D 27** (1983) 105.

[59] S. Albino, B. A. Kniehl and G. Kramer, *AKK Update: Improvements from New Theoretical Input and Experimental Data*, *Nucl. Phys.* **B 803** (2008) 42 [arXiv:0803.2768].

[60] F. Aversa, P. Chiappetta, M. Greco and J. P. Guillet, *QCD Corrections to Parton-Parton Scattering Processes*, *Nucl. Phys.* **B 327** (1989) 105.

[61] B. A. Kniehl, G. Kramer, I. Schienbein and H. Spiesberger, *Inclusive Charmed-Meson Production at the CERN LHC*, *Eur. Phys. J.* **C 72** (2012) 2082 [arXiv:1202.0439].

[62] B. A. Kniehl, G. Kramer, I. Schienbein and H. Spiesberger, *Inclusive B-meson production at small p_T in the general-mass variable-flavor-number scheme*, *Eur. Phys. J.* **C 75** (2015) 140 [arXiv:1502.01001].

[63] ALICE collaboration, *Measurement of* D^0, D^+, D^{*+} *and* D_s^+ *production in pp collisions at* $\sqrt{s} = 5.02$ TeV *with ALICE*, arXiv:1901.07979.

[64] M. L. Mangano, P. Nason and G. Ridolfi, *Heavy quark correlations in hadron collisions at next-to-leading order*, *Nucl. Phys.* **B 373** (1992) 295.

[65] PROSA collaboration, *Prompt neutrino fluxes in the atmosphere with PROSA parton distribution functions*, *JHEP* **05** (2017) 004 [arXiv:1611.03815].

[66] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [arXiv:0709.2092].

[67] S. Frixione, P. Nason and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126 [arXiv:0707.3088].

[68] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [arXiv:1002.2581].

[69] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012].

[70] E. Norrbin and T. Sjostrand, *Production and hadronization of heavy quarks*, *Eur. Phys. J.* **C 17** (2000) 137 [hep-ph/0005110].

[71] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk et al., *Uncertainties of predictions from parton distribution functions. 2. The Hessian method*, *Phys. Rev.* **D 65** (2001) 014013 [hep-ph/0101032].

[72] T. Kneesch, B. A. Kniehl, G. Kramer and I. Schienbein, *Charmed-meson fragmentation functions with finite-mass corrections*, *Nucl. Phys.* **B 799** (2008) 34 [arXiv:0712.0481].

[73] Particle Data Group collaboration, *Review of Particle Physics*, *Phys. Rev.* **D 98** (2018) 030001.

[74] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur. Phys. J.* **C 74** (2014) 3024 [arXiv:1404.5630].

[75] CMS collaboration, *Constraining gluon distributions in nuclei using dijets in proton-proton and proton-lead collisions at* $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, *Phys. Rev. Lett.* **121** (2018) 062002 [arXiv:1805.04736].

[76] CMS collaboration, *Observation of nuclear modifications in* $W^{\pm}$ *boson production in pPb collisions at* $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV, arXiv:1905.01486.

[77] M. Bonvini, S. Marzani and T. Peraro, *Small-x resummation from HELL*, *Eur. Phys. J.* **C 76** (2016) 597 [arXiv:1607.02153].

[78] J.-P. Lansberg and H.-S. Shao, *Towards an automated tool to evaluate the impact of the nuclear modification of the gluon density on quarkonium, D and B meson production in proton–nucleus collisions*, *Eur. Phys. J.* **C 77** (2017) 1 [arXiv:1610.05382].