

## EPS09 — A new generation of NLO and LO nuclear parton distribution functions

---

K.J. Eskola,<sup>a,b</sup> H. Paukkunen<sup>a,b</sup> and C.A. Salgado<sup>c</sup>

<sup>a</sup>*Department of Physics, University of Jyväskylä,  
P.O. Box 35, FI-40014, Finland*

<sup>b</sup>*Helsinki Institute of Physics, University of Helsinki,  
P.O. Box 64, FI-00014, Finland*

<sup>c</sup>*Departamento de Física de Partículas and IGFAE,  
Universidad de Santiago de Compostela, Spain*

*E-mail:* [kari.eskola@phys.jyu.fi](mailto:kari.eskola@phys.jyu.fi), [hannu.paukkunen@phys.jyu.fi](mailto:hannu.paukkunen@phys.jyu.fi),  
[carlos.salgado@usc.es](mailto:carlos.salgado@usc.es)

**ABSTRACT:** We present a next-to-leading order (NLO) global DGLAP analysis of nuclear parton distribution functions (nPDFs) and their uncertainties. Carrying out an NLO nPDF analysis for the first time with three different types of experimental input — deep inelastic  $\ell$ +A scattering, Drell-Yan dilepton production in p+A collisions, and inclusive pion production in d+Au and p+p collisions at RHIC — we find that these data can well be described in a conventional collinear factorization framework. Although the pion production has not been traditionally included in the global analyses, we find that the shape of the nuclear modification factor  $R_{dAu}$  of the pion  $p_T$ -spectrum at midrapidity retains sensitivity to the gluon distributions, providing evidence for shadowing and EMC-effect in the nuclear gluons. We use the Hessian method to quantify the nPDF uncertainties which originate from the uncertainties in the data. In this method the sensitivity of  $\chi^2$  to the variations of the fitting parameters is mapped out to orthogonal error sets which provide a user-friendly way to calculate how the nPDF uncertainties propagate to any factorizable nuclear cross-section. The obtained NLO and LO nPDFs and the corresponding error sets are collected in our new release called EPS09. These results should find applications in precision analyses of the signatures and properties of QCD matter at the LHC and RHIC.

**KEYWORDS:** Deep Inelastic Scattering, Hadronic Colliders, Parton Model, QCD

ARXIV EPRINT: [0902.4154](https://arxiv.org/abs/0902.4154)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Framework and analysis method</b>	<b>2</b>
2.1	Definition of nPDFs	2
2.2	Fitting functions and parameters	3
2.3	Experimental input and cross-sections	4
2.4	Definition of the best fit	8
2.5	nPDF errors	9
<b>3</b>	<b>Results and discussion</b>	<b>10</b>
3.1	Final parametrization	10
3.2	Comparison with the data	13
3.3	Comparison with earlier global analyses	17
3.4	Leading-order analysis	18
<b>4</b>	<b>Application</b>	<b>19</b>
<b>5</b>	<b>Summary and outlook</b>	<b>22</b>
<b>A</b>	<b>Error analysis</b>	<b>23</b>
A.1	Hessian matrix	23
A.2	Error propagation	24
A.3	Choice of $\Delta\chi^2$ ?	26
A.4	Calculation of the Hessian Matrix	29
<b>B</b>	<b>How to speed up computation of pion production</b>	<b>30</b>

---

## 1 Introduction

In the advent of the LHC becoming online, the knowledge of free and bound nucleon parton distribution functions (PDFs) together with their uncertainties, is perhaps more topical than ever. Reliable PDF estimates play a significant role in extracting the dynamics of the underlying partonic processes from the plethora of detected events in a hadron collider like Tevatron or LHC.

Building on our experience obtained in a series of leading order (LO) analyses [1–4] of the nuclear PDFs (nPDFs), we here extend our work to the next-to-leading order (NLO) accuracy in perturbative QCD (pQCD). The earlier studies on this topic [5, 6] have already demonstrated that this non-trivial extension is successful, which strongly supports

the collinear-factorization approach in describing the hard nuclear collisions. In addition to confirming this conclusion, our analysis is an extension in two important ways:

First, in contrast to the earlier NLO nPDF analyses that were restricted to include only data from deep inelastic lepton-nucleus scattering (DIS) and Drell-Yan (DY) dilepton production, our analysis contains also inclusive pion production measured at RHIC. This additional input improves especially the determination of the gluon densities.

The second extension concerns the nPDF uncertainty analysis. Although this issue has been pursued earlier [3, 5, 7], it has been very difficult for a general user to actually compute the propagation of the nPDF uncertainties to a cross-section of his/her interest. The necessary tools for doing such estimates have only been available for the groups who did the nPDF analysis and in practice one has hitherto been forced to only compare the best-fit predictions given by independent nPDF fits, see e.g. ref. [8]. To improve upon this issue we follow the ideas originally developed by the CTEQ collaboration [9, 10], and construct a collection of nPDF error sets meant to serve as a practical way for estimating the nPDF-originating uncertainties in the hard process cross-sections in nuclear collisions. The package named EPS09, consisting of our best NLO and LO fits together with 30 error sets for both, will be available in [11].

In what follows, we will first introduce the general framework and the method of the present analysis in section 2. The obtained results are presented in section 3, which also contains comparisons with the data and with earlier global analyses. In section 4, we present a concrete application of the obtained nPDFs and show how to best estimate the PDF-related uncertainties for nuclear cross-sections. In section 5 we give a summary and a brief outlook. The technical details of our error analysis are explained in appendix A, and in appendix B we show how to speed up the numerical calculation for pion production.

## 2 Framework and analysis method

### 2.1 Definition of nPDFs

Following the framework of our earlier analyses, we define the bound proton NLO PDFs  $f_i^A(x, Q^2)$  for each parton flavor  $i$  by

$$f_i^A(x, Q^2) \equiv R_i^A(x, Q^2) f_i^{\text{CTEQ6.1M}}(x, Q^2), \tag{2.1}$$

where  $R_i^A(x, Q^2)$  denotes the nuclear modification to the free proton PDF  $f_i^{\text{CTEQ6.1M}}$  from the CTEQ6.1M set [12] in the  $\overline{MS}$  scheme. Although more recent sets of free proton PDFs exist, this is the latest CTEQ set in the zero-mass variable flavour number scheme (ZM-VFNS) which we also use here — a prescription that treats the heavy quarks as massless particles, each one active only when the factorization scale  $Q^2$  exceeds its mass threshold  $Q^2 \geq m_q^2$ . We assume that isospin symmetry for bound protons and neutrons holds and take the average up and down quark PDF in a nucleus  $A$  with  $Z$  protons to be

$$\begin{aligned} u_A(x, Q^2) &= \frac{Z}{A} f_u^A(x, Q^2) + \frac{A-Z}{A} f_d^A(x, Q^2) \\ d_A(x, Q^2) &= \frac{Z}{A} f_d^A(x, Q^2) + \frac{A-Z}{A} f_u^A(x, Q^2), \end{aligned} \tag{2.2}$$

like in our earlier works [1–4]. We neglect the nuclear modifications in Deuterium  $A = 2$ . Such effects were studied e.g. in ref. [5] where an order of 1...2% effects were suggested. However, since only a minor improvement in  $\chi^2$  was found, we expect that the error we make by neglecting the nuclear effects in Deuterium should be negligible compared to typical uncertainty originating from the experimental errors.

The parametrization of the nuclear modifications  $R_i^A(x, Q^2)$  is performed at the charm quark mass threshold  $Q_0^2 \equiv m_c^2 = 1.69 \text{ GeV}^2$  imposing the momentum and baryon number sum rules

$$\sum_{i=q,\bar{q},g} \int_0^1 dx x f_i^A(x, Q_0^2) = 1, \quad \int_0^1 dx [f_{uV}^A(x, Q_0^2) + f_{dV}^A(x, Q_0^2)] = 3, \quad (2.3)$$

for each nucleus  $A$  separately. At higher scales  $Q^2 > Q_0^2$ , the nPDFs are obtained by solving the DGLAP evolution equations [14] with NLO  $\overline{MS}$  splitting functions [15–17].

We solve these equations numerically for both the nuclear and the free proton PDFs, following an efficient method suggested in [18]: The PDFs are approximated by 3th order polynomials in sufficiently small  $x$ -intervals for which the convolution integrals involving the splitting kernels can be evaluated analytically. The evolution in  $Q^2$  is built as a Taylor expansion around the initial scale  $Q_0^2$ . We have tested the accuracy of our code against the benchmark PDFs [19], and observed that 9th order Taylor expansion with  $\sim 200$   $x$ -points reaches an accuracy well below 1% at  $Q^2 = 10^4 \text{ GeV}^2$  in all  $x$  except the largest- $x$  ( $x \geq 0.9$ ) sea-quark sector. In practice, for the global fit procedure we keep the 9th order Taylor expansion but coarsen the  $x$ -grid to speed up the computation but ensure that the evolution works properly in the required kinematical domain  $Q^2 < 200 \text{ GeV}^2$  (see figure 2 ahead).

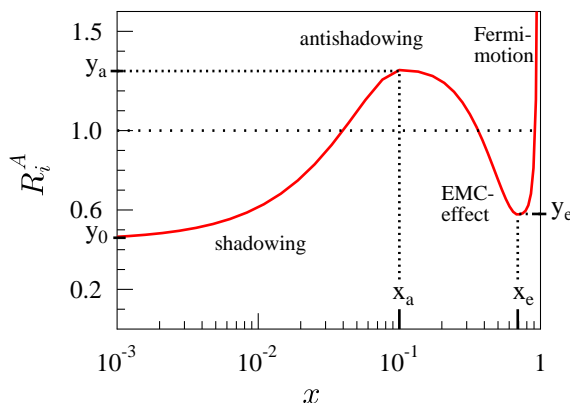
The relevant parameters that specify the evolution are the charm and bottom quark masses  $m_c = 1.3 \text{ GeV}$  and  $m_b = 4.5 \text{ GeV}$ , and the 2-loop strong coupling constant fixed at  $Z$ -pole to  $\alpha_s(M_Z^2) = 0.118$ . The top quark is not considered in the present analysis. These choices coincide with the ones adopted in CTEQ6.1M. In the LO analysis we take the free proton baseline from the CTEQ6L1 [13] set, and solve the LO DGLAP equations with 1-loop  $\alpha_s$ .

## 2.2 Fitting functions and parameters

The diversity of the presently available data is not broad enough to allow the nuclear modifications for each parton flavor to be independently determined. Thus, following our previous works [1–4] we define only three different nuclear corrections at the initial scale  $Q_0^2$ :  $R_V^A$  for both valence quark distributions,  $R_S^A$  for all sea quarks, and  $R_G^A$  for gluons. We parametrize these by piecewise functions

$$R_i^A(x) = \begin{cases} a_0 + (a_1 + a_2x)[\exp(-x) - \exp(-x_a)] & x \leq x_a \\ b_0 + b_1x + b_2x^2 + b_3x^3 & x_a \leq x \leq x_e \\ c_0 + (c_1 - c_2x)(1-x)^{-\beta} & x_e \leq x \leq 1, \end{cases} \quad (2.4)$$

where the parameters  $a_i$ ,  $b_i$ ,  $c_i$ ,  $\beta$ ,  $x_a$  and  $x_e$  are  $A$ -dependent. Joining the three pieces together to give a continuous function with vanishing first derivatives at matching points



**Figure 1.** An illustration of the fit function  $R_i^A(x)$  and the role of the parameters  $x_a$ ,  $x_e$ ,  $y_0$ ,  $y_a$ , and  $y_e$ .

$x_a$  and  $x_e$ , eliminates 6 out of the 13 parameters. The remaining ones are expressed in terms of the following 6 parameters with obvious interpretations:

$y_0$	Height to which shadowing levels as $x \rightarrow 0$
$x_a, y_a$	Position and height of the antishadowing maximum
$x_e, y_e$	Position and height of the EMC minimum
$\beta$	Slope factor in the Fermi-motion part,

the remaining parameter  $c_0$  is fixed to  $c_0 = 2y_e$ . The roles of these parameters are illustrated in figure 1 which also roughly indicates which  $x$ -regions are meant by the commonly used terms: shadowing, antishadowing, EMC-effect, and Fermi-motion.

The  $A$ -dependence of the fit parameters is assumed to follow a power law

$$d_i^A = d_i^{A_{\text{ref}}} \left( \frac{A}{A_{\text{ref}}} \right)^{p_{d_i}}, \quad (2.5)$$

where  $d_i = x_a, y_a \dots$ , and where the reference nucleus is Carbon,  $A_{\text{ref}} = 12$ .

The baryon number and momentum sum rules eliminate  $y_0$  and  $p_{y_0}$  for valence quarks and gluons, leaving us with 32 free parameters. This is still way too large number of parameters to be determined only by the data — further assumptions (based on prior experience) are needed to decide which parameters can truly be deduced from the data and which can be taken as fixed.

### 2.3 Experimental input and cross-sections

The main body of the data in our analysis consists of  $\ell + A$  DIS measurements. We also utilize the DY dilepton production data from fixed target p+A collisions at Fermilab and inclusive neutral-pion production data measured in d+Au and p+p collisions at RHIC.<sup>1</sup> Table 1 lists the sets included in our analysis and figure 2 displays their kinematical reach

<sup>1</sup>In contrast to our previous analysis [4], we do not include the BRAHMS forward rapidity charged hadron d+Au data here. These data will be separately discussed in section 4.

Experiment	Process	Nuclei	Data points	$\chi^2$ LO	$\chi^2$ NLO	Weight	ref.
SLAC E-139	DIS	He(4)/D	21	6.5	7.3	1	[20]
NMC 95, re.	DIS	He/D	16	14.5	15.6	5	[21]
NMC 95	DIS	Li(6)/D	15	23.6	16.8	1	[22]
NMC 95, $Q^2$ dep.	DIS	Li(6)/D	153	162.2	157.0	1	[22]
SLAC E-139	DIS	Be(9)/D	20	9.6	12.2	1	[20]
NMC 96	DIS	Be(9)/C	15	3.8	3.8	1	[23]
SLAC E-139	DIS	C(12)/D	7	4.1	3.2	1	[20]
NMC 95	DIS	C/D	15	15.0	13.8	1	[22]
NMC 95, $Q^2$ dep.	DIS	C/D	165	141.8	142.0	1	[22]
NMC 95, re.	DIS	C/D	16	19.3	20.5	1	[21]
NMC 95, re.	DIS	C/Li	20	30.3	28.4	1	[21]
FNAL-E772	DY	C/D	9	7.5	8.3	1	[24]
SLAC E-139	DIS	Al(27)/D	20	10.9	12.5	1	[20]
NMC 96	DIS	Al/C	15	6.0	5.8	1	[23]
SLAC E-139	DIS	Ca(40)/D	7	5.0	4.1	1	[20]
FNAL-E772	DY	Ca/D	9	2.9	3.4	15	[24]
NMC 95, re.	DIS	Ca/D	15	25.4	24.7	1	[21]
NMC 95, re.	DIS	Ca/Li	20	23.9	19.6	1	[21]
NMC 96	DIS	Ca/C	15	6.0	6.0	1	[23]
SLAC E-139	DIS	Fe(56)/D	26	19.1	23.9	1	[20]
FNAL-E772	DY	Fe/D	9	2.1	2.2	15	[24]
NMC 96	DIS	Fe/C	15	11.0	10.8	1	[23]
FNAL-E866	DY	Fe/Be	28	20.9	21.7	1	[25]
CERN EMC	DIS	Cu(64)/D	19	13.4	14.8	1	[26]
SLAC E-139	DIS	Ag(108)/D	7	3.8	2.9	1	[20]
NMC 96	DIS	Sn(117)/C	15	9.6	9.1	1	[23]
NMC 96, $Q^2$ dep.	DIS	Sn/C	144	80.2	82.8	10	[27]
						(x=0.0125 only)	
FNAL-E772	DY	W(184)/D	9	7.0	6.7	10	[24]
FNAL-E866	DY	W/Be	28	27.3	24.2	1	[25]
SLAC E-139	DIS	Au(197)/D	21	11.6	13.8	1	[20]
RHIC-PHENIX	$\pi^0$ prod.	dAu/pp	20	7.3	6.3	20	[28]
NMC 96	DIS	Pb/C	15	6.90	7.2	1	[23]
Total			929	738.6	731.3		

**Table 1.** The data used in our analysis. The mass numbers are indicated in parentheses and the number of data points refers to those falling within our kinematical cuts,  $Q^2, M^2 \geq 1.69 \text{ GeV}^2$  for DIS and DY, and  $p_T \geq 1.7 \text{ GeV}$  for hadron production at RHIC. The quoted  $\chi^2$  values correspond to the unweighted contributions of each data set in LO and NLO. Also the weight factors for data sets are shown.

in the  $(x, Q^2)$ -plane. We will use the following notation:

$$\begin{aligned}
 R_{\text{DIS}}^A(x, Q^2) &\equiv \frac{\frac{1}{A} d\sigma_{\text{DIS}}^{IA}/dQ^2 dx}{\frac{1}{2} d\sigma_{\text{DIS}}^{Id}/dQ^2 dx}, & R_{F_2}^A(x, Q^2) &\equiv \frac{F_2^A(x, Q^2)}{F_2^d(x, Q^2)} \\
 R_{\text{DY}}^A(x_{1,2}, M^2) &\equiv \frac{\frac{1}{A} d\sigma_{\text{DY}}^{pA}/dM^2 dx_{1,2}}{\frac{1}{2} d\sigma_{\text{DY}}^{pd}/dM^2 dx_{1,2}} \\
 R_{\text{dAu}}^\pi &\equiv \frac{1}{\langle N_{\text{coll}} \rangle} \frac{d^2 N_\pi^{\text{dAu}}/dp_T dy}{d^2 N_\pi^{\text{pp}}/dp_T dy} \stackrel{\text{min.bias}}{=} \frac{\frac{1}{2A} d^2 \sigma_\pi^{\text{dAu}}/dp_T dy}{d^2 \sigma_\pi^{\text{pp}}/dp_T dy}.
 \end{aligned} \tag{2.6}$$

The kinematical variables in DIS are the Bjorken- $x$  and the virtuality of the photon  $Q^2$ . In DY  $M^2$  denotes the invariant mass of the lepton pair, and  $x_{1,2} \equiv \sqrt{M^2/s} e^{\pm y}$  where  $y$  is the pair rapidity. The inclusive pion production is characterized by the transverse momentum  $p_T$  and rapidity  $y$  of the outgoing pion. The average number of binary nucleon-nucleon collisions (in the centrality class studied) is denoted by  $\langle N_{\text{coll}} \rangle$ . In this analysis we only consider minimum bias data, and do not focus on the transverse coordinate dependence of the nPDFs. The kinematical cuts we impose on the data are  $M^2, Q^2 \geq 1.69 \text{ GeV}^2$  for DIS and DY, and  $p_T \geq 1.7 \text{ GeV}$  for inclusive pion production.

All cross-sections are calculated in the collinear factorization formalism folding the PDFs  $f_i(x, \mu^2)$  with perturbatively calculable pieces  $\hat{\sigma}$ , schematically

$$\begin{aligned}
 \sigma_{\text{DIS}}^{\ell+A \rightarrow \ell+X} &= \sum_{i=q,\bar{q},g} f_i^A(\mu^2) \otimes \hat{\sigma}_{\text{DIS}}^{\ell+i \rightarrow \ell+X}(\mu^2) \\
 \sigma_{\text{DY}}^{p+A \rightarrow l^+l^-+X} &= \sum_{i,j=q,\bar{q},g} f_i^p(\mu^2) \otimes f_j^A(\mu^2) \otimes \hat{\sigma}^{ij \rightarrow l^+l^-+X}(\mu^2) \\
 \sigma^{A+B \rightarrow \pi+X} &= \sum_{i,j,k=q,\bar{q},g} f_i^A(\mu^2) \otimes f_j^B(\mu^2) \otimes \hat{\sigma}^{ij \rightarrow k+X}(\mu^2) \otimes D_{k \rightarrow \pi}(\mu^2).
 \end{aligned}
 \tag{2.7}$$

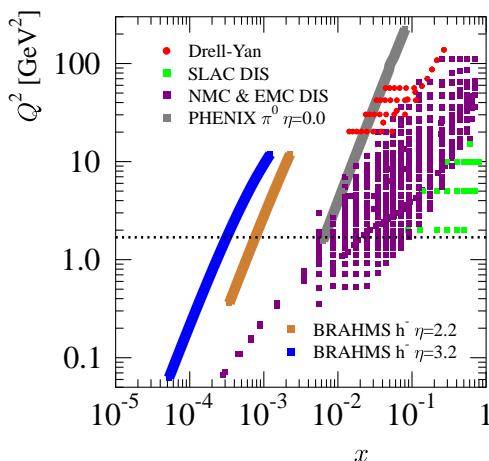
As indicated above, we choose the factorization, renormalization and fragmentation scales to be equal, fixed to a characteristic scale in the process: in DIS to the photon virtuality  $Q^2$ , in DY to the invariant mass of the lepton pair  $M^2$ , and in the pion production to its transverse momentum  $p_T^2$ .

The expressions for double differential DIS cross-sections  $d^2\sigma/dxdQ^2$  involving the NLO coefficients for structure functions  $F_{1,2}$  are well documented e.g. in ref. [29]. In the case of DY we need the rapidity distributions  $d^2\sigma/dM^2dy_R$ , from which  $d^2\sigma/dM^2dx_{1,2}$  are obtained by the relation  $x_{1,2} = \sqrt{M^2/s} e^{\pm y}$ . The relevant partonic cross-sections for DY process can be obtained from [30].

The numerical calculation of NLO DIS cross-sections involve only one-dimensional integrals, posing no problems in evaluating them repeatedly during the fitting process. The calculation of the DY cross-section, however, requires evaluation of double integrals which are already somewhat slower to compute. The small amount of available DY data nevertheless keeps the direct evaluation of the integrals still fast enough, and we do not need to resort to use pre-computed K-factors as is done e.g in [51].

The inclusive pion production embodies the fragmentation functions  $D_{i \rightarrow \pi}(z, \mu^2)$  for parton flavour  $i$  to make a pion carrying a fraction  $z$  of its momentum. In this work we have decided to use the KKP average pion  $\pi^+ + \pi^-$  parametrization [31] for describing the neutral pion  $\pi^0$  production. We have, however, checked that the results obtained with newer ones AKK08 [32] and nDSS [33, 34] practically all coincide in the case of average pions. In practice, we employ the INCNLO [35] program which includes the NLO hadron production code by Aversa et al. [36]. The computation of pion production in NLO involves triple integrals which would be very time consuming to evaluate directly at every round of iteration, but we can speed up the computation significantly by applying the procedure explained in appendix B.

The role played by each data type can be summarized as follows:



**Figure 2.** The kinematical reach of the DIS, DY and pion production data (see table 1) corresponding to the factorization scale choices explained in the text. The points indicate the lowest  $x$  and  $Q^2$  values in which partons are sampled in the cross-section calculation. Also the BRAHMS data [37] for negatively charged hadron production is shown as it will be discussed later in section 4. The dashed horizontal line indicates the kinematical cut imposed on the data.

- **Deep inelastic scattering**

DIS data has an excellent constraining power for quark distributions in the whole range  $0.01 \leq x \leq 1$  spanned by the data. At large  $x$  these data are mainly sensitive to the valence quarks and at small  $x$  to the sea quarks. At moderate  $x$ , however, close to the antishadowing peak near  $x = 0.1$ , such separation of the sea and valence quark contributions is not possible on the basis of this type of data alone. Despite the direct photon-gluon fusion channel contributing to the DIS cross-section at NLO, the main gluon constraint provided by DIS still comes through the scale evolution of sea quarks that is driven by the gluons.

- **Drell-Yan dilepton production**

The DY data, taken together with DIS, can discriminate between valence and sea quarks near  $x = 0.1$ . The DY cross-section retains also some sensitivity to the sea quarks at larger  $x$  but, unfortunately, the precision of the current data is not enough to exploit this constraint in its full potential. The invariant mass  $M^2$  in our data sample is typically large,  $M^2 \gg Q_0^2$ , and consequently there are sizable evolution effects that constrain the gluons also.

- **Inclusive pion production**

This type of data is usually accompanied by a rather large normalization uncertainty stemming, among other sources, from the model-dependent quantity  $\langle N_{\text{coll}} \rangle$ . Apart from the normalization uncertainty, the *shape* of  $R_{\text{dAu}}^\pi$  can nevertheless act as a vital constraint, especially for the nuclear modification for gluons. The slight downward trend seen in the large- $p_T$  part of  $R_{\text{dAu}}^\pi$  at midrapidity [28] indicates the need for a

gluon EMC-effect, and the smallest- $p_T$  part of the  $R_{\text{dAu}}^\pi$  would not be satisfactorily reproduced without shadowing (see figure 9 ahead).

## 2.4 Definition of the best fit

A customary way of finding an optimal set of parameters that fits a large number of experimental data, is to minimize a global  $\chi^2$ -function with the help of a routine such as MINUIT [38]. The requirement for a  $\chi^2$ -function is that its magnitude measures the degree of agreement between the theory predictions and the experimental data, but the exact definition can vary from one analysis to another. In this work we define the global  $\chi^2$  by [4, 9]

$$\chi^2(\{a\}) \equiv \sum_N w_N \chi_N^2(\{a\}) \tag{2.8}$$

$$\chi_N^2(\{a\}) \equiv \left( \frac{1 - f_N}{\sigma_N^{\text{norm}}} \right)^2 + \sum_{i \in N} \left[ \frac{f_N D_i - T_i(\{a\})}{\sigma_i} \right]^2, \tag{2.9}$$

where  $N$  labels the data sets,  $D_i$  denotes an individual data value with  $\sigma_i$  point-to-point uncertainty (statistical and systematic uncertainties added in quadrature), and  $T_i$  is the corresponding theory prediction with a parameter set  $\{a\}$ . If an overall (relative) normalization uncertainty  $\sigma_N^{\text{norm}}$  is given, the normalization factor  $f_N \in [1 - \sigma_N^{\text{norm}}, 1 + \sigma_N^{\text{norm}}]$  is introduced, and determined for each parameter set  $\{a\}$  by minimizing  $\chi_N^2$ . Thus,  $f_N$  is computed in every round of iteration, when searching for the best overall  $\chi^2$ . The final  $f_N$  is thus an output of the analysis.

The overall normalization uncertainty is the simplest example of correlated uncertainties in the data. If the experiment is able to provide a correlation matrix accounting for more complex correlations, the definition of the  $\chi^2$  should be extended further, as done e.g. in [9]. However, the definition (2.9) is adequate for our purposes, since such correlation matrices are presently not available for the nuclear data we use.

The weight factors  $w_N$  (with default value  $w_N = 1$ ) are needed to amplify constraints provided by those data sets whose contribution to  $\chi^2$  would otherwise be too small to be noticed by MINUIT. As the weights apparently induce a piece of subjectivity to the analysis, some further explanation here is in order: Weighting is justified if there is no significant mutual tension between the weighted data set and the others. In this case the weights do not introduce a strong bias in the analysis (such bias was found e.g. in [4]). For example, the  $\pi^0$ -data makes only about 1% of the total  $\chi^2$ , and changes in such a small fraction are practically buried in the numerical noise due to statistical fluctuations in the wealth of other data. Thus, inclusion of these data with a default weight of one would be effectively the same as leaving it completely out from the analysis, which would mean losing all the large- $x$  gluon constraints it can provide. Consequently, the gluon EMC-parameter  $y_e$  would need to be fixed by hand which, we think, would induce clearly more subjectivity to the fit than assigning these data some extra weight and keeping  $y_e$  free. Similarly, weighting the DY data enhances its residual sensitivity to the large- $x$  sea quarks, which enables us to free the large- $x$  sea parameter  $y_e$  in a controlled manner. We see these as clear improvements over the previous nPDF analyses.

In practice, however, finding a reasonable set of weights is largely an iterative procedure:<sup>2</sup> having first found a converging fit with a limited amount of free parameters and no additional weights, it is often easy to notice whether there are features in the data sets not satisfactorily reproduced. These data sets are then assigned some extra weight and, if needed, some previously frozen parameter can be freed and the minimization procedure repeated. In this way we have arrived at the weights listed in table 1.

The gluons at small  $x$  are only weakly constrained by the available data, and we have observed that some parameters easily drift — especially in the error analysis — to an unphysical region where shadowing would get stronger towards smaller nuclei. We cure this unwanted feature by adding a further term

$$1000 [(y_0^G(\text{He}) - y_0^G(\text{Pb})) - (y_0^S(\text{He}) - y_0^S(\text{Pb}))]^2 \quad (2.10)$$

to the  $\chi^2$ , which effectively drives the  $A$ -dependence, but not the magnitude, of  $R_G$  at the smallest- $x$  similar to what there is in the sea quarks. We stress that inclusion of this term to  $\chi^2$  does not affect the agreement with the data in any way, but it secures a systematic deepening of the gluon shadowing when going from lighter nucleus to a heavier one, especially so for the error-sets discussed below.

## 2.5 nPDF errors

As mentioned earlier, one of the main goals of the present work is to develop a general-use tool for estimating the propagation of the nPDF uncertainties to any hard process quantity  $X$ . This brings the uncertainty analysis of nPDFs up to a similar level as in the free proton case.

In this paper we will only consider uncertainties originating from the experimental errors in the fitted data. Although the set of parameters  $\{a^0\}$  giving the minimum  $\chi^2$  represents our best estimate for the nPDFs, the presence of experimental uncertainties allows us to move in the neighborhood of  $\{a^0\}$  without immediately deteriorating the agreement with individual data sets. To find this relevant neighborhood and to quantify the relative differences to our central nPDFs properly, we adopt the Hessian method.<sup>3</sup> The details of this method are documented in appendix A, we summarize here only the main points.

The Hessian method rests on a quadratic approximation for a departure of  $\chi^2$  from its minimum  $\chi_0^2$

$$\chi^2 - \chi_0^2 \approx \sum_{ij} \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} (a_i - a_i^0)(a_j - a_j^0) \equiv \sum_{ij} H_{ij} (a_i - a_i^0)(a_j - a_j^0), \quad (2.11)$$

which defines the Hessian matrix  $H$ . In constructing the expansion (2.11), special attention should be paid to multi-parameter correlations. If several parameters control the same parton type in a certain  $x$ -region, there are usually notable correlations between them. This

<sup>2</sup>For a weight determination in a neural network approach, see [39].

<sup>3</sup>For alternative methods see e.g. refs. [9, 40].

may lead to over-estimating the uncertainty in that region when applying the approximation in eq. (2.11). In particular, changes in the valence quark modification  $R_V$  obtained by small variations of the Fermi-motion parameter  $\beta$  can be largely realized by varying  $x_e$  and  $y_e$  instead. In this sense  $\beta$  is redundant in the error analysis and we eliminate it from the expansion (2.11). The same argument applies to  $A$ -dependence of these parameters. The 15 final parameters considered in our error analysis are indicated in table 2.

A practical and numerically more stable way is to work not with the original parameters  $\{a\}$ , but use a basis  $\{z\}$  which diagonalizes the Hessian matrix. Assuming linear propagation of uncertainties, the symmetric extreme deviation in any quantity  $X$  that corresponds to a given rise  $\Delta\chi^2$  in  $\chi^2$ , can be written as

$$(\Delta X)^2 \approx \frac{1}{4} \sum_k (X(S_k^+) - X(S_k^-))^2, \quad (2.12)$$

where  $X(S_k^+)$  and  $X(S_k^-)$  denote the values of the quantity  $X$  computed by the nPDF sets  $S_k^+$  and  $S_k^-$ , which are obtained by displacing the fit parameters in the  $z$ -space direction  $k$  such that  $\chi^2$  grows by a certain fixed amount  $\Delta\chi^2$ . Lower and upper uncertainties  $\Delta X^\pm$  can be computed also separately, using a prescription [41]

$$\begin{aligned} (\Delta X^+)^2 &\approx \sum_k [\max\{X(S_k^+) - X(S^0), X(S_k^-) - X(S^0), 0\}]^2 \\ (\Delta X^-)^2 &\approx \sum_k [\max\{X(S^0) - X(S_k^+), X(S^0) - X(S_k^-), 0\}]^2, \end{aligned} \quad (2.13)$$

where  $S^0$  denotes the central PDF set giving the minimum  $\chi^2$ . All the errors shown in this paper have been computed by the latter method, with  $\Delta\chi^2 = 50$  which corresponds to a “90% confidence criterion” defined in appendix A. As discussed in section 3 the Hessian matrix is computed for 15 fitting parameters in our analysis. According to the procedure explained above, this leads to 30 error-sets in addition to the best-fit central set. All these 31 nPDF sets needed for the evaluation of eq. (2.13) are included in our EPS09 release [11]. This allows the computation of uncertainties in nuclear cross-section ratios, similar to those in eq. (2.6), by any interested user. The corresponding calculation of uncertainty estimates for the absolute cross-sections (instead of ratios), will be explained later in section 4.

In this paper we do not discuss the possible shortcomings concerning our whole framework such as breakdown of pQCD factorization at small  $Q^2$  or small/large- $x$  corrections to the DGLAP splitting functions already at the nucleon-nucleon level collisions — for an extensive study of such issues, we recommend the reader to consult refs. [42, 43]. Also, the presence of additional  $(Q^2)^{-n}$  power corrections are expected to become increasingly important for large nuclei, see e.g. [44–48]. As we will see next, however, the present framework seems to work very well in the kinematical domain considered in this analysis.

### 3 Results and discussion

#### 3.1 Final parametrization

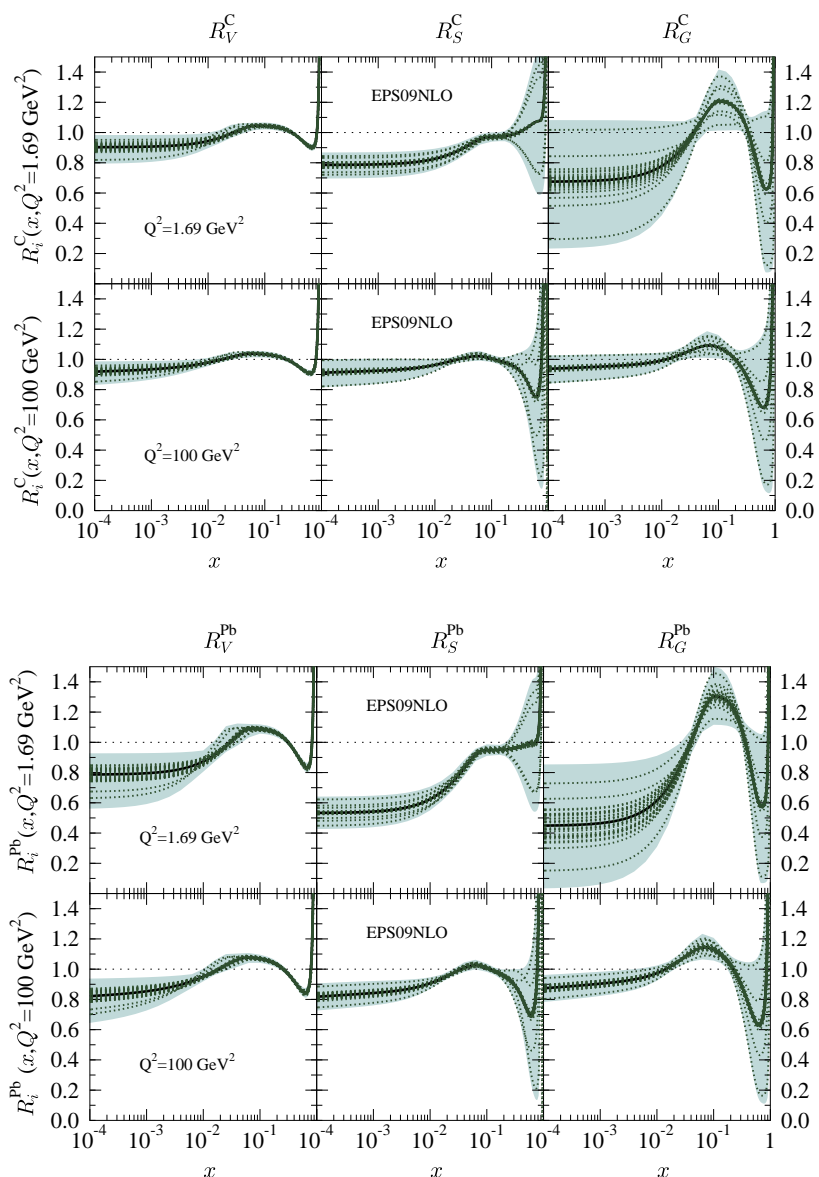
In table 2 we present the values of the 17 parameters resulting from the  $\chi^2$  minimization and those that we fixed. The error analysis was performed for the 15 parameters shown in

NLO	Param.	Valence $R_V^A$	Sea $R_S^A$	Gluon $R_G^A$
1	$y_0$	Baryon sum rule	<b>0.785</b>	Momentum sum rule
2	$p_{y_0}$	—	<b>-0.136</b>	—
3	$x_a$	<b><math>6.56 \times 10^{-2}</math></b>	<b><math>8.74 \times 10^{-2}</math></b>	0.1 fixed
4	$p_{x_a}$	0, fixed	0, fixed	0, fixed
5	$x_e$	<b>0.688</b>	as valence	as valence
6	$p_{x_e}$	0, fixed	0, fixed	0, fixed
7	$y_a$	<b>1.05</b>	<b>0.970</b>	<b>1.207</b>
8	$p_{y_a}$	<b><math>1.47 \times 10^{-2}</math></b>	<b><math>-8.12 \times 10^{-3}</math></b>	<b><math>2.72 \times 10^{-2}</math></b>
9	$y_e$	<b>0.901</b>	<b>1.076</b>	<b>0.625</b>
10	$p_{y_e}$	<b><math>-2.81 \times 10^{-2}</math></b>	as valence	as valence
11	$\beta$	1.31	1.3, fixed	1.3, fixed
12	$p_\beta$	$4.63 \times 10^{-2}$	0, fixed	0, fixed
LO				
1	$y_0$	Baryon sum rule	<b>0.890</b>	Momentum sum rule
2	$p_{y_0}$	—	<b><math>-8.03 \times 10^{-2}</math></b>	—
3	$x_a$	<b><math>6.84 \times 10^{-2}</math></b>	<b>0.127</b>	0.1 fixed
4	$p_{x_a}$	0, fixed	0, fixed	0, fixed
5	$x_e$	<b>0.727</b>	as valence	as valence
6	$p_{x_e}$	0, fixed	0, fixed	0, fixed
7	$y_a$	<b>1.04</b>	<b>0.971</b>	<b>1.096</b>
8	$p_{y_a}$	<b><math>1.53 \times 10^{-2}</math></b>	<b><math>-1.65 \times 10^{-2}</math></b>	<b><math>4.57 \times 10^{-2}</math></b>
9	$y_e$	<b>0.902</b>	<b>1.062</b>	<b>0.901</b>
10	$p_{y_e}$	<b><math>-2.97 \times 10^{-2}</math></b>	as valence	as valence
11	$\beta$	1.34	1.3, fixed	1.3, fixed
12	$p_\beta$	$5.07 \times 10^{-2}$	0, fixed	0, fixed

**Table 2.** List of parameters defining the modifications  $R_V^A$ ,  $R_S^A$  and  $R_G^A$  through eqs. (2.4) and (2.5) at our initial scale  $Q_0^2 = 1.69 \text{ GeV}^2$  in NLO and LO. The final values for the 15 free parameters considered in the minimization procedure and in our error analysis are shown in bold face. The values for the parameters  $\beta$  and  $p_\beta$  for valence quarks were fixed on the basis of a 17 parameter fit.

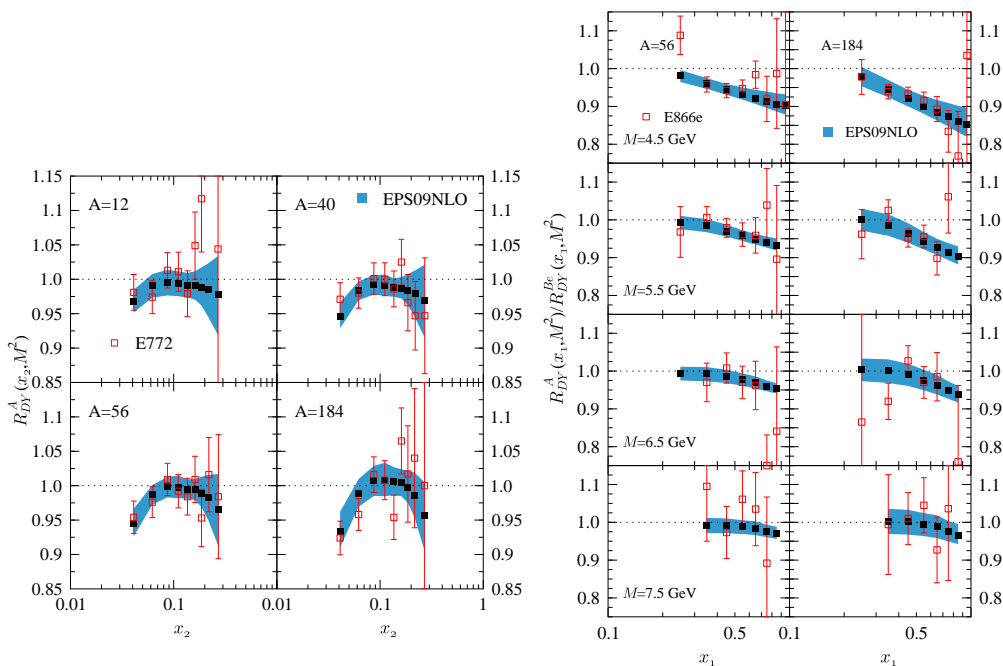
bold face. The corresponding nuclear modifications together with the uncertainty estimates at our initial scale  $Q_0^2 = 1.69 \text{ GeV}^2$  and at a higher scale  $Q^2 = 100 \text{ GeV}^2$  are shown for Carbon and Lead in figure 3. We make the following observations for gluons, sea quarks and valence quarks respectively.

At our parametrization scale  $Q_0^2$  there are large uncertainties in both small- $x$  and large- $x$  gluons. Only at moderate  $x$  the gluons are somewhat better controlled as the precision small- $x$  DIS data — although directly more sensitive to the sea quarks — constrain the gluons at slightly higher  $x$  due to the parton branching encoded into DGLAP evolution. At higher  $Q^2$  the small- $x$  uncertainty rapidly shrinks whereas at large  $x$  a sizable uncertainty band persists.



**Figure 3.** The nuclear modifications  $R_V$ ,  $R_S$ ,  $R_G$  for Carbon (upper group of panels) and Lead (lower group of panels) at our initial scale  $Q_0^2 = 1.69 \text{ GeV}^2$  and at  $Q^2 = 100 \text{ GeV}^2$ . The thick black lines indicate the best-fit results, whereas the dotted green curves denote the error sets. The shaded bands are computed from eq. (2.13).

The narrow throat in the sea quark uncertainty band at  $x \sim 10^{-2} \dots 10^{-1}$  reflects the good constraining power of the precision DIS data. Towards higher  $x$ , the uncertainty grows as the accuracy of the DY data is not enough to decisively nail down nuclear modification for the sea quarks there. Note, however, that unlike in our earlier works, the parameter  $y_e$  was free. Towards small  $x$  the errors are perhaps surprisingly small given that there are no direct data constraints. This is an artefact of the chosen form of the fit function as the tight constraints at  $x \sim 10^{-2} \dots 10^{-1}$  fix also the smaller- $x$  behaviour leading to an



**Figure 4.** The computed NLO  $R_{DY}^A(x, M^2)$  (filled squares and error bands) as a function of  $x_2$  (left) and  $x_1$  (right) compared with the E772 [24] and E866 [25] data (open squares).

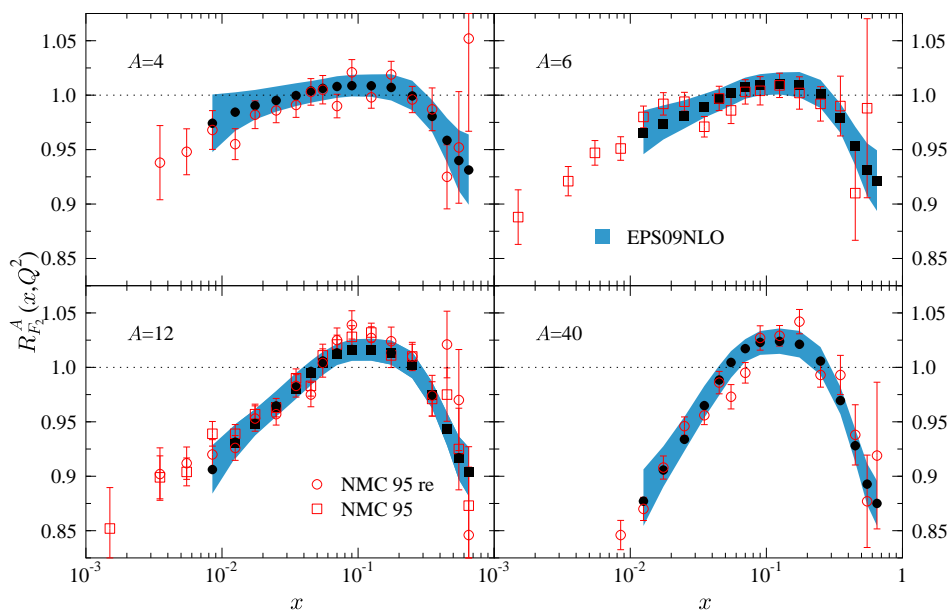
unrealistically small uncertainty band there. This should be kept in mind when computing hard cross-sections involving very-small- $x$  gluons.

The valence quark modifications are well under control at  $x \gtrsim 0.1$  — thanks to the DIS data probing the valence quarks at this region. At small- $x$ , the baryon number sum rule together with the chosen form of the fit function halts the uncertainty band from growing, although there are no data that would directly constrain the valence quarks there.

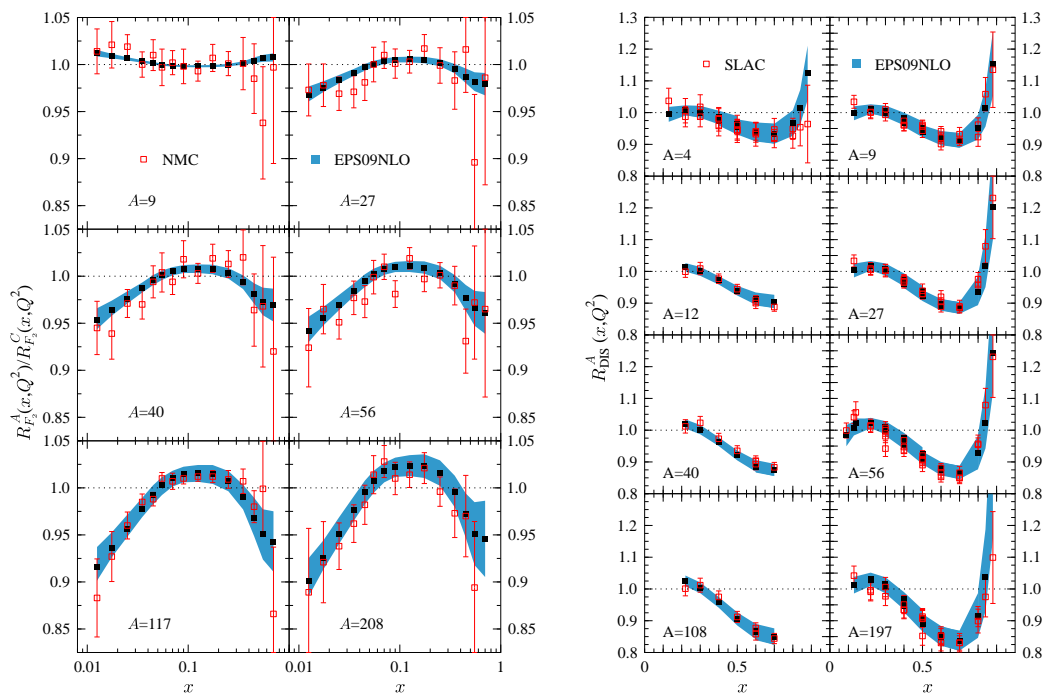
### 3.2 Comparison with the data

In this section we demonstrate how the obtained NLO parametrization reproduces the data and how the nPDF uncertainties in figure 3 relate to the error bars of the data. In all figures we present, the open symbols with error bars denote the experimental data and the filled symbols or thicker lines are the corresponding calculated values. The shaded light blue bands indicate the uncertainty estimates computed using our 30 error sets in eq. (2.13). Unless otherwise stated, the point-to-point systematic and statistical errors have been added in quadrature.

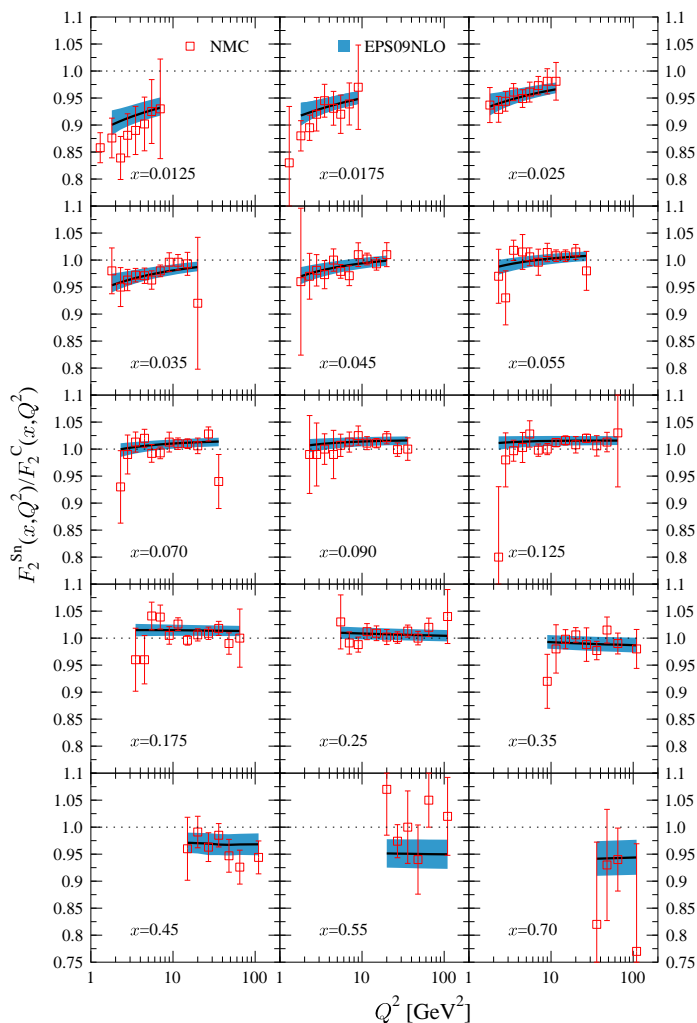
In figure 4 we show the Drell-Yan data. As mentioned earlier, we have found that there is some dependence on the large- $x$  sea quarks in the E772 data shown in the left panels. However, the apparent large scatter from one nucleus to another makes it difficult to well exploit this sensitivity and completely impossible to extract the  $A$ -dependence of the large- $x$   $R_S$  independently. The growing uncertainty band towards larger  $x$  is nicely in line with the growing large- $x$  uncertainty for  $R_S$  in figure 3. We would like to emphasize two



**Figure 5.** The calculated NLO  $R_{F_2}^A(x, Q^2)$  and (filled symbols and error bands) compared with the NMC 95 [22] and the reanalysed NMC 95 [21] data (open symbols).



**Figure 6.** The computed NLO ratios  $R_{F_2}^A(x, Q^2)$  vs.  $R_{F_2}^C(x, Q^2)$  and  $R_{DIS}^A(x, Q^2)$  (filled squares and error bands) compared with the NMC [23] (left) and SLAC [20] (right) data (open squares).



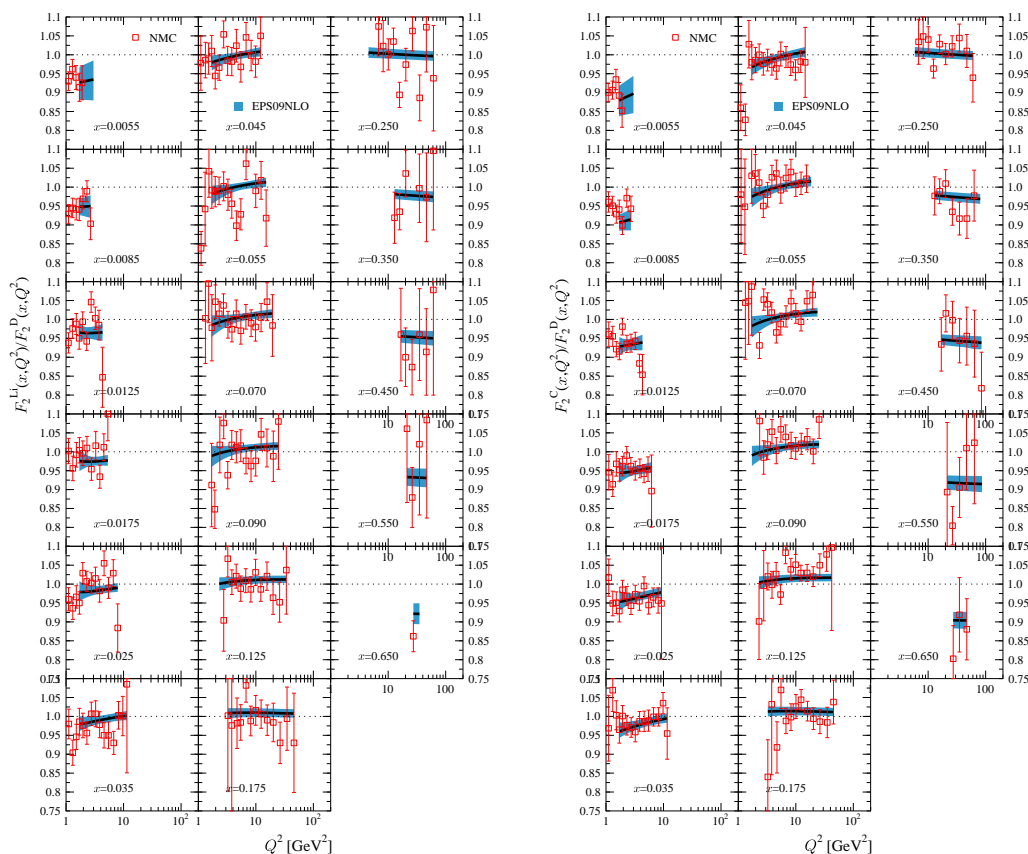
**Figure 7.** The calculated NLO scale evolution (solid black lines and error bands) of the ratio  $F_2^{\text{Sn}}/F_2^{\text{C}}$  compared with the NMC data [27] for several fixed values of  $x$ .

systematic features present in the data, which are well caught by our global analysis: The downward trend (shading) of the smallest- $x$  points in the E772 data, and the diminishing nuclear effects towards larger invariant mass  $M$  in the E886 data.

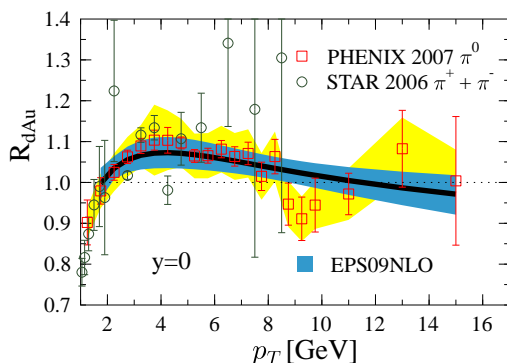
The DIS data in figures 5-8 comprise the bulk of our experimental constraints. Since a large part of the DIS data points lies at relatively low  $Q^2$  (see figure 2), the good agreement between the data and the theory also verifies that our fit function is flexible enough. At higher  $Q^2$ , the evolution effects become sizable and the fit function dependence is no longer as critical.

The  $A$ -dependence of the nuclear effects is encoded in the parametrization through eq. (2.5), and its determination is mainly possible due to the existence of DIS data for several nuclei. This dependence and its correct reproduction is obvious e.g. from figures 5 and 6 in which data from several nuclei are gathered together for an easy comparison.

In addition to the DY data, the scaling violation effects are clearly visible in the NMC



**Figure 8.** The calculated NLO scale evolution (solid black lines and error bands) of the ratios  $F_2^{Li}/F_2^D$  and  $F_2^C/F_2^D$  compared with the NMC data [27] for several fixed values of  $x$ . One percent systematic error has been added in quadrature with the statistical errors.



**Figure 9.** The computed  $R_{dAu}$  (thick black line and blue error band) at  $y = 0$  for inclusive pion production compared with the PHENIX [28] data (open squares). The error bars are the statistical uncertainties, and the yellow band indicate the point-to-point systematic errors. The additional 10% overall normalization uncertainty in the data is not shown. The data have been multiplied by the optimized normalization factor  $f_N = 1.03$ , which is an output of our analysis. Also the STAR data [50] (open circles) multiplied by a normalization factor  $f_N = 0.90$  are shown for comparison.

data figures 7 and 8, which display the  $Q^2$ -evolution of the DIS structure function  $F_2$  ratios. Thus, the DGLAP evolution plays a significant role in a successful description of these data.

In figure 9 we finally show the nuclear modification for inclusive pion yield in d+Au collisions at midrapidity. We plot both PHENIX  $\pi^0$  [28] and STAR  $\pi^+ + \pi^-$  [50] data.<sup>4</sup> We are, at the moment, the only group that includes this type of data in the global DGLAP analysis of nPDFs. We observe, however, that the shape of the pionic  $R_{dAu}$  as a function of  $p_T$  is very well reproduced without any additional effects, indicating that the universality of the nPDFs works well. We thus find no reason to exclude these data from the global fit analyses.

Typically, the error bands in most of the figures are roughly of the same size as (especially never badly overshooting) the error bars in the data, verifying that we have successfully propagated them to our nPDFs. In some cases, however, the size of the nPDF errors is apparently below the experimental uncertainties. This can happen for two reasons: First, in cases like figure 8, the small error bands result simply because there are other, more precise data sets probing the same  $x$ -region and, by construction, the size of our error bands are controlled by the most stringent ones. Second, in cases like the panels of figure 6 which compare Beryllium or Aluminium to Carbon, where cross-section ratio is taken between two nuclei very close in  $A$ , the errors in other parameters than those from the powers  $p_{d_i}$  in eq. (2.5) tend to cancel.

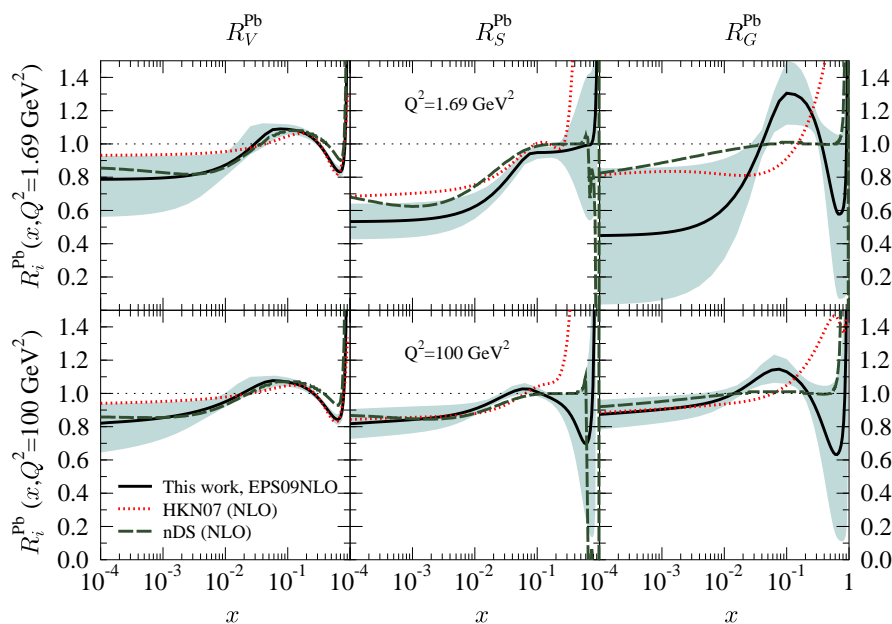
### 3.3 Comparison with earlier global analyses

In figure 10 we compare the results obtained in the present analysis EPS09 to the earlier NLO global analyses nDS [6] and HKN07 [5]. The comparison is shown at two scales,  $Q^2 = 1.69 \text{ GeV}^2$  and  $Q^2 = 100 \text{ GeV}^2$ , overlaid on our error bands. Evidently, the main differences in the central fits are in the gluon modifications  $R_G$ , especially at small  $x$  and low  $Q^2$ , but going to higher  $Q^2$  the curves tend to gather closer together. The exception is the large- $x$  region where significant differences persist. This comparison illustrates the large leftover freedom especially in nuclear gluon sector if only DIS and DY data are used as constraints.

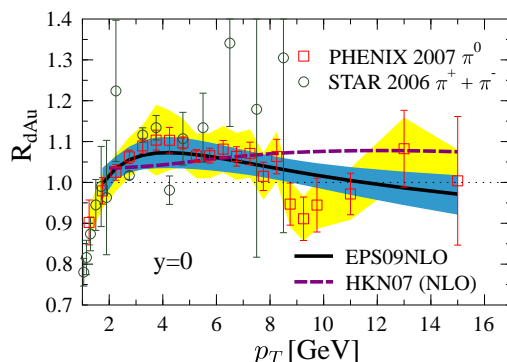
The majority of the discrepancies in figure 10 originate from the adopted forms of the fit functions and assumptions made about the parameters in those  $x$ -regions which are not well constrained by the DIS or DY data. In particular, HKN07 and nDS did not include pion d+Au data which favors the presence of an EMC suppression for gluons. In order to demonstrate the importance of these data in constraining the large- $x$  gluons, figure 11 displays the pion  $R_{dAu}$  including also the prediction obtained by using the HKN07 NLO nPDFs. Although the HKN07 prediction remains largely within the experimental errors, the shape of the EPS09 curve is clearly more consistent with the data. The dominant partonic channel for inclusive pion production is this  $p_T$  range the gluon-initiated one. Thus, the qualitatively contradicting behaviour between HKN07 and EPS09 in figure 11 is inevitably linked to the differences in their nuclear effects for gluon PDFs.

---

<sup>4</sup>The STAR data was, however, not included in the fit due to poorer statistics and more restricted



**Figure 10.** Comparison of the average valence and sea quark, and gluon modifications at  $Q^2 = 1.69 \text{ GeV}^2$  and  $Q^2 = 100 \text{ GeV}^2$  for Pb nucleus from the NLO global DGLAP analyses HKN07 [5], nDS [6] and this work, EPS09NLO.



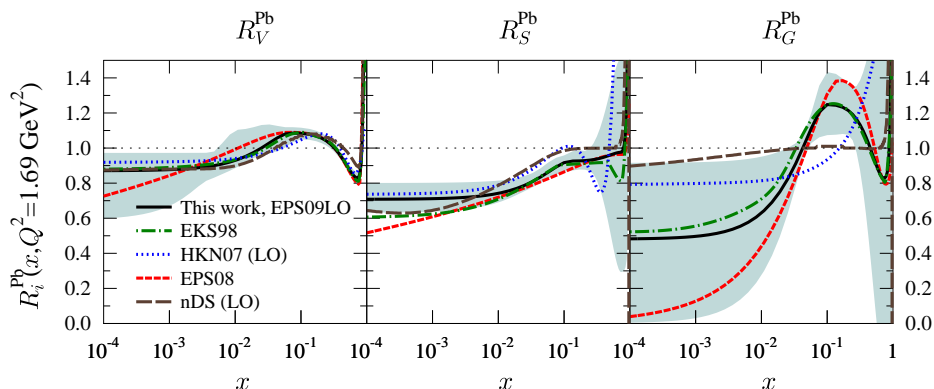
**Figure 11.** As figure 9 but also the prediction from HKN07 (NLO) is shown. The difference between EPS09 and HKN07 here demonstrates the constraining power of these data in pinning down the nuclear gluon PDFs in the mid- $x$  and large- $x$  regions.

### 3.4 Leading-order analysis

Although the NLO analysis is the main objective in the present paper, we have also performed a new LO analysis to provide the tools for computing uncertainty estimates also in this widely-used framework. The LO framework is basically the same as in NLO, but the partonic cross-sections and DGLAP splitting functions are one power lower in  $\alpha_s$ , and we

---

kinematical reach.



**Figure 12.** Comparison of the average valence and sea quark, and gluon modifications at  $Q^2 = 1.69 \text{ GeV}^2$  for Pb nucleus from LO global DGLAP analyses EKS98 [1, 2], EKPS [3], nDS [6], HKN07 [5], and this work EPS09LO.

use the CTEQ6L1 [13] free proton PDF set as the baseline. The parameters specifying the LO fit are listed in table 2.

Due to the qualitatively similar behaviour of the LO and NLO fits, we do not present a detailed discussion of the LO results but only show, in figure 12, the resulting nuclear modifications for Lead at the initial scale  $Q_0^2$  with their  $\Delta\chi^2 = 50$  uncertainty bands. The same figure also presents the comparison to other LO analyses. We find that our error bands can accommodate most of the earlier analyses. So, in practice, using the EPS09LO error sets all these different nPDF parametrizations are effectively covered.

As can be read off from table 1 there is practically no difference between the LO and NLO fits at the level of the values of  $\chi^2$  obtained. Intuitively, however, the NLO fit should be better constrained than the LO one as there are more partonic channels open in the NLO DIS and DY cross-sections. Indeed this expectation is confirmed by figures 3 and 12. Especially the large- $x$  gluons are better constrained at NLO.

## 4 Application

In this section we apply the obtained EPS09NLO parametrization — the central set and 30 error sets — to a cross-section that was not included in the fit. Through this example we also want to demonstrate how our parametrization should be applied in practice.

We consider here inclusive negative hadron  $h^-$  production at forward (pseudo) rapidities  $\eta = 2.2$  and  $\eta = 3.2$ , in p+p and d+Au collisions, measured by the BRAHMS collaboration [37] at RHIC. In our previous article [4] we discussed how the suppression observed in the nuclear modification  $R_{dAu}$  obtained from these data would strongly favour very deep gluon shadowing, and we searched for the strongest possible one that would still not contradict the available DIS and DY data. The analysis [4] was performed in a LO framework and we were forced to use fragmentation functions for average hadrons  $h^+ + h^-$

instead of charge-separated ones for  $h^-$  only.<sup>5</sup> In the current NLO setup we relax such simplification and employ the charge-separated NLO fragmentation functions by Sassot et al. [33].

We first investigate how well the NLO pQCD calculation can reproduce the shape and magnitude of the differential  $h^-$  yields measured by BRAHMS in p+p and d+Au collisions from which the nuclear modification  $R_{dAu}$  is computed. The inclusive yields are linked to the cross-sections by

$$\frac{d^2 N^{pp}}{dp_T dy} \stackrel{\text{min.bias}}{=} \frac{1}{\sigma_{NN}^{\text{inelastic}}} \frac{d^2 \sigma^{pp}}{dp_T dy} \quad ; \quad \frac{d^2 N^{dAu}}{dp_T dy} \stackrel{\text{min.bias}}{=} \frac{\langle N_{\text{coll}} \rangle}{\sigma_{NN}^{\text{inelastic}}} \frac{1}{2A} d^2 \sigma^{dAu}, \quad (4.1)$$

where  $\sigma_{NN}^{\text{inelastic}}$  is the total inelastic nucleon-nucleon cross-section and  $\langle N_{\text{coll}} \rangle$  is the average number of binary nucleon-nucleon collisions. Here we take  $\sigma_{NN}^{\text{inelastic}} = 40$  mb [49], and  $\langle N_{\text{coll}} \rangle = 7.2$  [37], understanding that there is an inherent uncertainty in both:  $\sigma_{NN}^{\text{inelastic}}$  has not been measured for  $\sqrt{s} = 200$  GeV and determining its value requires extrapolation. In turn,  $\langle N_{\text{coll}} \rangle$  is a model-dependent number resulting from a Glauber-model simulation performed by the experiment. However, both of these quantities are only multiplicative overall factors and do not affect the shape of the obtained distribution.

Whenever comparing theoretical predictions to experimental data, we emphasize the importance of presenting also the theoretical uncertainties deriving from the PDF errors. For absolute cross-sections in p+p collisions these are obtained from eq. (2.13) using the 40 error sets provided by CTEQ [12]. In the d+Au case there is an additional contribution from the uncertainties of the nuclear effects which we have quantified in the present analysis by the 30 error sets with the CTEQ central set as a baseline. In this case, we compute the d+Au cross-sections separately with 1+40+30=71 parton sets defined by:

$$\begin{aligned} R_{S^0}^A &\times f_{\text{Central set}}^{\text{CTEQ6.1M}} \\ R_{S^0}^A &\times f_{\text{Error set } \pm 1}^{\text{CTEQ6.1M}} \\ R_{S^0}^A &\times f_{\text{Error set } \pm 2}^{\text{CTEQ6.1M}} \\ &\vdots \\ R_{S_1^\pm}^A &\times f_{\text{Central set}}^{\text{CTEQ6.1M}} \\ R_{S_2^\pm}^A &\times f_{\text{Central set}}^{\text{CTEQ6.1M}} \\ &\vdots, \end{aligned}$$

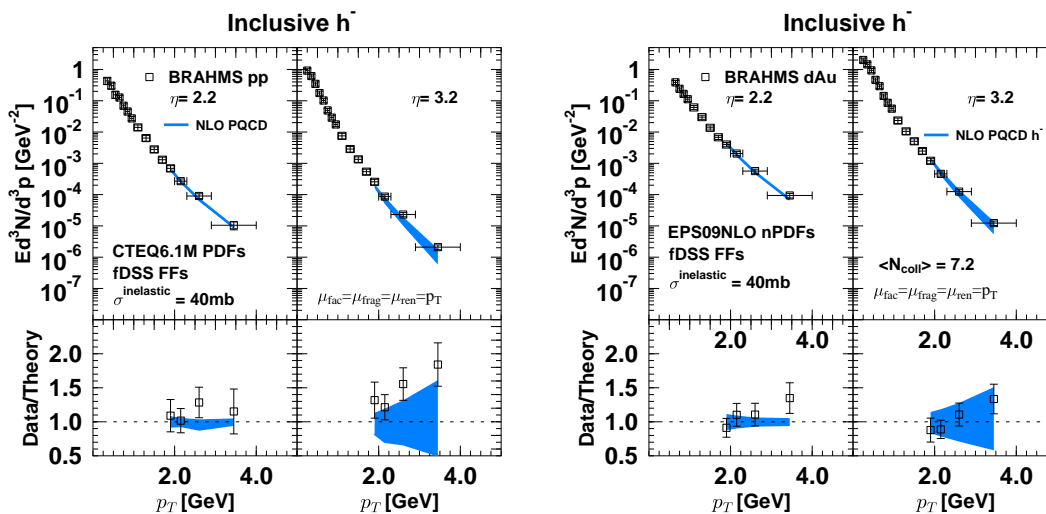
where the first one gives the central prediction, and others<sup>6</sup> are pairwise contributing to the size of the upper and lower errors via eq. (2.13).

We stress that this way of computing the uncertainty for d+Au cross-section is a simplification due to parametrizing the nuclear modifications on top of fixed proton baseline PDFs. For example, when the CTEQ error sets are multiplied by our central-set, the

---

<sup>5</sup>The extraction of the charge separated fragmentation functions from p+p data is reliable only at NLO due to significant perturbative  $\mathcal{O}(\alpha_s^2)$  corrections.

<sup>6</sup>The uncertainty sets in CTEQ6.1M and EPS09 correspond to similar 90%-confidence criterion.



**Figure 13.** Inclusive yield of negatively charged hadrons  $h^-$  in pp and dAu collisions. The experimental data shown by open squares is from [37] with statistical and systematical errors added in quadrature, and the horizontal bars indicate the  $p_T$  bin. The blue band indicates the 90% confidence range derived from free proton and nPDF uncertainties. The calculated cross-sections have been averaged over the  $p_T$ -bin width.

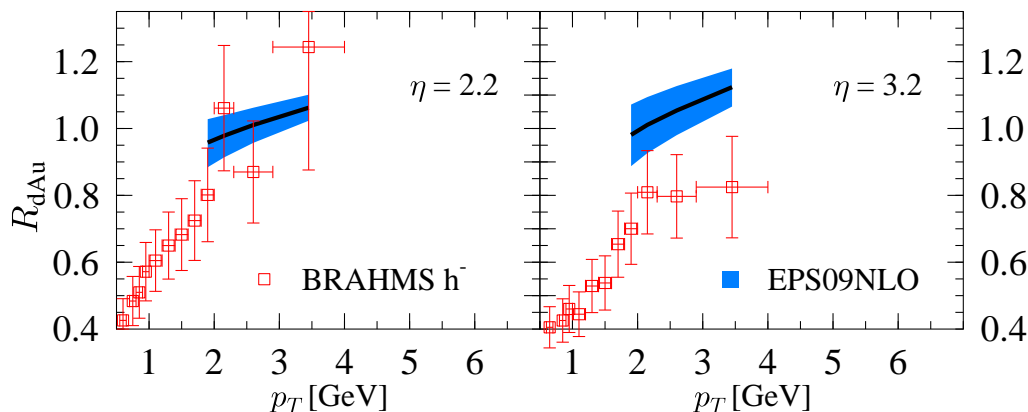
momentum is not strictly conserved. In other words, although the central prediction of our nPDFs is well-defined, the free and bound proton PDF analyses cannot be completely separated in the uncertainty analysis. The only way of obtaining completely self-consistent uncertainty predictions for d+Au collisions would require putting the free and bound proton PDF analyses together into a one single analysis in which the artificial separation to free proton uncertainties and those stemming from the nuclear modifications would not be required. In the absence of such master analysis, we recommend to calculate the uncertainties for absolute cross-sections as explained above.

The results and comparison with the BRAHMS data are shown in figure 13, in which the blue bands indicate the PDF uncertainty range. The lower panels show the data/theory ratios. Looking first at the  $\eta = 2.2$  panels, the flatness of the data/theory ratio shows that the shapes of the measured distributions for both p+p and d+Au cases are well reproduced by the pQCD calculation. As these ratios are also very close to one, we conclude that the normalization factors  $\sigma_{NN}^{\text{inelastic}}$  and  $\langle N_{\text{coll}} \rangle$  are well estimated and that fixing the factorization scale to hadronic  $p_T$  is also a valid choice.

In the  $\eta = 3.2$  panels, one first observes that in the case of p+p collision the agreement between the theory and the data is clearly worse and the experimental and PDF errors are only barely overlapping. Such mismatch is not, however, observed in the d+Au cross-sections. Thus, we conclude that if the nuclear modification

$$R_{\text{dAu}} \equiv \frac{1}{\langle N_{\text{coll}} \rangle} \frac{d^2 N^{\text{dAu}}/dp_T dy}{d^2 N^{\text{pp}}/dp_T dy} \stackrel{\text{min.bias}}{=} \frac{\frac{1}{2A} d^2 \sigma^{\text{dAu}}/dp_T dy}{d^2 \sigma^{\text{pp}}/dp_T dy}. \quad (4.2)$$

is formed from these data, the result with  $\eta = 3.2$  will inevitably lie below the pQCD prediction. This is demonstrated in figure 14 where we show the comparison of the BRAHMS



**Figure 14.** The computed nuclear modification ratio  $R_{\text{dAu}}$  at forward rapidities (filled squares) for negatively-charged hadron production, compared with the BRAHMS data [37] (open squares). The error bars are the statistical and systematic uncertainties added in quadrature. The additional overall normalization uncertainty of 5% is not shown. The blue bands are computed using the EPS09NLO sets.

$R_{\text{dAu}}$  data with the pQCD prediction from our global analysis. We stress that this happens not because we would incorrectly reproduce the d+Au cross-section but because the NLO pQCD calculation seems to undershoot the p+p data. This is the reason why we do not include these data to the current global nPDF analysis. However, as figure 13 shows, there is no reason why one should abandon the factorization for  $h^-$  in d+Au collisions. Presumably, more statistics would be needed to resolve the discrepancy at  $\eta = 3.2$ .

## 5 Summary and outlook

We have presented a global analysis of the nuclear corrections to the free proton PDFs at NLO accuracy in pQCD. Since the quality of the obtained fit is excellent and since there seems not be any strong tensions between the fitted data sets,<sup>7</sup> we argue that our analysis supports the validity of collinear factorization in describing hard nuclear collisions such as DIS, DY and inclusive high- $p_T$  pion production. The qualities of both NLO and LO fits are almost identical,  $\chi^2/N \approx 0.79$ , but the NLO analysis somewhat reduces the inherent uncertainties. This is contrast to the free proton analyses where the quality of the NLO fit is significantly better (see e.g. [51]). In the near future, with new data for inclusive hadron and direct photon production from RHIC and LHC experiments, factorization will be tested further.

For the first time in the nuclear case, we have now also quantified the uncertainties originating from the experimental errors in a manner that we can release a collection of nPDF error sets. Using these 1+30 sets, the nPDF uncertainties can be propagated to any hard process of interest. In lieu of a still missing master analysis which would consistently combine both the free and bound proton PDFs, we recommend to estimate the uncertainties

<sup>7</sup>e.g. in EPS08 [4] there was a significant tension between the DIS and forward BRAHMS data.

as follows: If an absolute cross-section  $\sigma$  is considered the PDF and nPDF uncertainties should be combined independently, schematically as follows

$$(\Delta\sigma)_{\text{Total}}^2 = (\Delta\sigma)_{\text{Proton}}^2 + (\Delta\sigma)_{\text{EPS09}}^2.$$

For the nuclear cross-section ratios  $R$ , — like those in eq. (2.6) — it should suffice to consider only the nPDF errors,

$$(\Delta R)_{\text{Total}}^2 = (\Delta R)_{\text{EPS09}}^2$$

(see sections 2.5 and 4 for the details). In the latter case we do not recommend to include  $(\Delta R)_{\text{Proton}}^2$ , as this would lead to an overestimation of the errors. Moreover, for consistency EPS09 should be used in conjunction with the same free proton PDF set that was used in the analysis, namely CTEQ6.1M at NLO and CTEQ6L1 at LO.

The most up-to-date free proton PDF analyses (e.g. [51, 52]) implements a somewhat better organized method for treating the heavy quarks than the ZM-VFNS employed in this analysis. Extension of the nPDF analysis to such general-mass scheme remains as a future work. Recently, the CTEQ collaboration has looked at the nuclear PDFs from a slightly different point of view by fitting the cross-section data from neutrino-Iron scattering [53]. Interestingly, the central fit points to a somewhat different nuclear modifications than what are obtained from electromagnetic DIS. Before making any strong conclusions about the possible difference, these data should be included in a global analysis and the possible tension between these and the other data sets carefully investigated. Also this is left as a future task.

## Acknowledgments

KJE and HP thank the Academy of Finland, Project 115262 and GRASPANP for financial support. CAS is supported by Ministerio de Ciencia e Innovación of Spain under projects FPA2005-01963 and FPA2008-01177; by Xunta de Galicia (Programa Incite); and by the European Commission grant PERG02-GA-2007-224770. CAS is a Ramón y Cajal researcher.

## A Error analysis

In this appendix we explain the error analysis in more detail. The basic formalism we present here was introduced in the original CTEQ articles [9, 10], but we have taken guidelines also subsequent CTEQ and MRST papers e.g. [13, 54].

### A.1 Hessian matrix

The Hessian approach is based on a quadratic expansion of  $\chi^2 \equiv \chi^2(\{a\})$  near its minimum  $\chi_0^2 \equiv \chi^2\{a^0\}$ ,

$$\chi^2 \approx \chi_0^2 + \sum_{ij} \delta a_i H_{ij} \delta a_j, \tag{A.1}$$

where  $\delta a_i \equiv a_i - a_i^0$ , and  $H_{ij}$  is the Hessian matrix defined here by

$$H_{ij} \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} \Big|_{a=a^0}. \quad (\text{A.2})$$

Being symmetric, the Hessian matrix has a set of normalized eigenvectors  $\mathbf{v}^{(k)}$  and eigenvalues  $\epsilon_k$  such that

$$\sum_j H_{ij} v_j^{(k)} = \epsilon_k v_i^{(k)}, \quad \sum_j v_j^{(k)} v_j^{(\ell)} = \delta_{k\ell}. \quad (\text{A.3})$$

Introducing a new set of coordinates  $\{z\}$ , as

$$\delta a_i = \sum_j v_i^{(j)} \sqrt{\frac{1}{\epsilon_j}} z_j, \quad (\text{A.4})$$

one finds that

$$z_k = \sqrt{\epsilon_k} \sum_j v_j^{(k)} \delta a_j, \quad (\text{A.5})$$

and

$$\chi^2 \approx \chi_0^2 + \sum_i z_i^2, \quad (\text{A.6})$$

which implies that in the region where the quadratic approximation (A.1) is valid, the new coordinates  $\{z\}$  are independent — the  $\chi^2$  increases uniformly in an arbitrary direction in the  $z$ -space.

## A.2 Error propagation

Let us consider any quantity  $X$  that depends on the PDFs, that is  $X \equiv X(\{z\})$ . In a linear approximation,  $X$  may be expanded in the vicinity of its central value  $X_0 \equiv X(\{z=0\})$  as  $X = X_0 + \Delta X$ , where

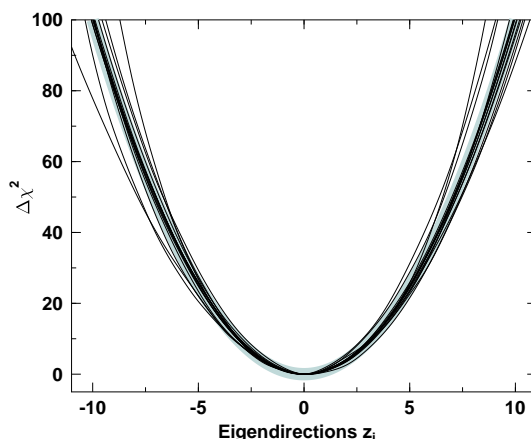
$$\Delta X \approx \sum_j \left( \frac{\partial X}{\partial z_j} \right) \delta z_j. \quad (\text{A.7})$$

Since  $\chi^2$  grows uniformly in all  $z$ -space directions, the  $z$ -space vector that extremizes  $\Delta X$  for a given  $\Delta \chi^2$ , is in the direction of the gradient of  $X$  and has a length of  $\sqrt{\Delta \chi^2}$ . The components of this vector are

$$\delta z_i = \sqrt{\Delta \chi^2} \left( \frac{\partial X}{\partial z_j} \right) \left( \sum_j \left( \frac{\partial X}{\partial z_j} \right)^2 \right)^{-1/2}, \quad (\text{A.8})$$

giving

$$(\Delta X)_{\text{extremum}}^2 \approx \Delta \chi^2 \sum_j \left( \frac{\partial X}{\partial z_j} \right)^2. \quad (\text{A.9})$$



**Figure 15.** This figure illustrates how the  $\Delta\chi^2$  in our NLO analysis grows when the fit parameters are scanned along the eigenvector directions. The shaded light blue band indicates the ideal quadratic behaviour  $\Delta\chi^2 = z_i^2$ . We can see that the quadratic approximation is good but not perfect.

In order to facilitate the computation of the derivatives  $\partial X/\partial z_k$  in eq. (A.9), we define auxiliary PDFs in several  $z$ -space coordinates:

$$\begin{aligned}
 S_0 &= (0, 0, 0, \dots, 0) \\
 S_1^\pm &= \pm\sqrt{\Delta\chi^2} (1, 0, 0, \dots, 0) \\
 S_2^\pm &= \pm\sqrt{\Delta\chi^2} (0, 1, 0, \dots, 0) \\
 &\vdots \\
 S_{N-1}^\pm &= \pm\sqrt{\Delta\chi^2} (0, 0, \dots, 1, 0) \\
 S_N^\pm &= \pm\sqrt{\Delta\chi^2} (0, 0, \dots, 0, 1).
 \end{aligned}
 \tag{A.10}$$

The first set  $S_0$  is the central set giving the minimum  $\chi^2$ , and in the other ones the fit parameters are changed by a fixed amount in each  $z$ -space direction separately. Using these PDF sets, the derivatives  $\partial X/\partial z_k$  can be approximated by a finite difference

$$\frac{\partial X}{\partial z_k} \approx \frac{X(S_k^\pm) - X(S_0)}{\pm\sqrt{\Delta\chi^2}} \approx \frac{X(S_k^+) - X(S_k^-)}{2\sqrt{\Delta\chi^2}},
 \tag{A.11}$$

where  $X(S_k^\pm)$  denotes the value of  $X$  computed with the PDF set  $S_k^\pm$ . Upon insertion to eq. (A.9) we get

$$(\Delta X)_{\text{extremum}}^2 \approx \sum_k (X(S_k^\pm) - X(S_0))^2 \approx \frac{1}{4} \sum_k (X(S_k^+) - X(S_k^-))^2.
 \tag{A.12}$$

These are the simplest equations by which the PDF uncertainties are propagated to any quantity  $X$ .

However, often the quadratic approximation is not perfect, revealed by non-ideal behaviour of  $\chi^2$  in the  $z$ -space. From figure 15 such an behaviour can be concretely seen to

some extent happen also in our NLO analysis. In such a case the growth of  $\chi^2$  induced by  $\sqrt{\Delta\chi^2}$  deviations in the eigenvector directions can be significantly different from the ideal  $\Delta\chi^2$ . Moreover, the eigendirections may exhibit very asymmetric behaviour, and the required deviations for  $\chi^2$  to grow by  $\Delta\chi^2$  may be significantly different on either side of the origin. This observation necessitates to modify the definition of the auxiliary PDFs to

$$\begin{aligned} S_1^\pm &= \pm\delta z_1^\pm (1, 0, 0, \dots, 0) \\ S_2^\pm &= \pm\delta z_2^\pm (0, 1, 0, \dots, 0) \\ &\vdots \end{aligned} \tag{A.13}$$

where  $\delta z_i^+$  and  $\delta z_i^-$  are the deviations in positive and negative side of  $i$ th eigendirection respectively, that cause  $\chi^2$  to grow by  $\Delta\chi^2$ . Replacing the eq. (A.12) by the following prescription [41]

$$\begin{aligned} (\Delta X^+)^2 &\approx \sum_k [\max\{X(S_k^+) - X(S^0), X(S_k^-) - X(S^0), 0\}]^2 \\ (\Delta X^-)^2 &\approx \sum_k [\max\{X(S^0) - X(S_k^+), X(S^0) - X(S_k^-), 0\}]^2, \end{aligned} \tag{A.14}$$

serves as a way to obtain the upper and lower uncertainties separately. If the quadratic approximation for  $\chi^2$  and the linear approximation for  $X$  were both perfect the eq. (A.12) and eq. (A.14) would lead to identical results.

### A.3 Choice of $\Delta\chi^2$ ?

In the derivation above, we assumed that there is a fixed  $\Delta\chi^2$  which sets a  $\chi^2$  boundary beyond which the global fit quickly deteriorates. In a statistically ideal case, when the uncertainties of the data are Gaussian-type and the theory is known to be “right enough”, the one standard deviation corresponds to  $\Delta\chi^2 = 1$  [49]. Such requirements can usually be met only if data from a very limited amount of sources are accepted — like PDF fits including only HERA data [43], where the ideal choice  $\Delta\chi^2 = 1$  is indeed justified.

However, in a *truly global* PDF analysis which, by definition, aims to combine data from as many independent experiments as possible, the situation becomes quite far from being ideal in the statistical sense: There are often sizable, unexplainable fluctuations within individual data sets and in practice all data sets are not mutually fully compatible. Thus, taking  $\Delta\chi^2 = 1$  in such analysis becomes practically meaningless.

This is most obviously true also in the present analysis — combining all data from the lightest nuclei up to the heaviest ones — which poses an additional difficulty which is not met in the free-proton analysis: In lieu of large amount of data for each nucleus separately, we model the  $A$ -dependence of the fit parameters by a power law (2.5). In this way we can evidently catch the correct general trend, but there will inevitably be some small pull between individual nuclei in various places. Thus, the choice for  $\Delta\chi^2$  should account also for our inability to treat the individual nuclei separately.

Below, we first introduce somewhat extended statistical approach which is sometimes used in determining  $\Delta\chi^2$ , and continue with a physically better motivated one, which we use in our error analysis.

**Statistically motivated method:** let us assume that the probability distribution for the global  $\chi^2$ -function is a simple Gaussian in the  $N$ -dimensional  $z$ -space

$$P_D(|\mathbf{z}|, N) = \frac{1}{(2\pi)^{N/2}} e^{-\mathbf{z}^2/2}. \quad (\text{A.15})$$

The probability  $P(\Delta\chi^2, N)$  for  $\chi^2$  increasing at most  $\Delta\chi^2$  is then obtained from

$$P(\Delta\chi^2, N) = \int_{\mathbf{z}^2 \leq \Delta\chi^2} \left( \prod_{i=1}^N dz_i \right) P_D(|\mathbf{z}|, N) = \int_0^{\Delta\chi^2} dS P_D^{\chi^2}(N, S) \quad (\text{A.16})$$

$$P_D^{\chi^2}(N, S) = \frac{1}{2\Gamma(N/2)} \left(\frac{S}{2}\right)^{N/2-1} e^{-S/2}. \quad (\text{A.17})$$

For usual “one standard deviation”  $P \approx 68.3\%$  and for  $N = 15$  parameters, this approach would suggest  $\Delta\chi^2 \approx 17$ . This a way of choosing  $\Delta\chi^2$  has been routinely employed e.g. by Kumano et al. [5] and previously also by us [3].

**Physically motivated method:** an alternative way for finding  $\Delta\chi^2$  is to examine how the  $\chi^2$ -contributions  $\chi_k^2$  of individual data sets with  $N_k$  data points, behave in the  $z$ -space (see e.g. [13]). Loosely speaking, by determining acceptable intervals for each of the  $z$ -space directions, and taking a suitable average, we can arrive at a physically better motivated value for  $\Delta\chi^2$  than what is obtained from the idealized statistical way described above.

Restricting the range how much individual  $\chi^2$ -contributions  $\chi_k^2$  are allowed to grow above the central set value  $\chi_{0,k}^2$ , is complicated by fluctuations in the data which can make  $\chi_{0,k}^2/N_k \gg 1$ , even if the overall agreement was visually good. In order to compensate for such effects and to treat all data sets equally, we follow the procedure suggested by the CTEQ and define normalized variables  $\chi_{k,n}^2$  by

$$\chi_k^2 \rightarrow \chi_{k,n}^2 \equiv \chi_k^2 \left( \frac{N_k - 2}{\chi_{0,k}^2} \right). \quad (\text{A.18})$$

Assuming that each  $\chi_{k,n}^2$  follows the probability distribution of eq. (A.17) with  $N = N_k$ , the most probable value of  $\chi_{k,n}^2$  is attained by the central set  $S_0$ . For each experiment we denote the largest acceptable  $\chi_{k,n}^2$  by  $M(P, \chi_k^2)$  and define it as a solution to equation

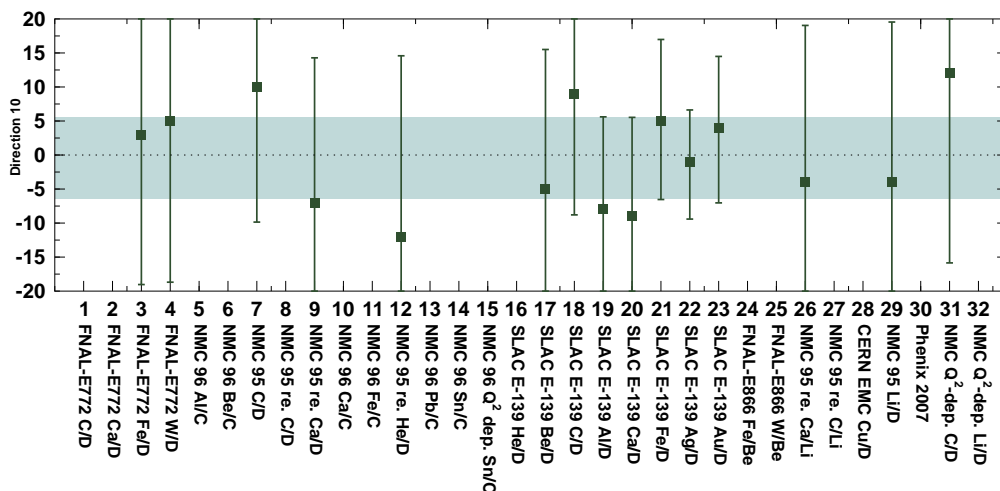
$$\int_0^{M(P, \chi_k^2)} dS P_D^{\chi^2}(N_k, S) = P, \quad (\text{A.19})$$

for fixed confidence level  $P$ . Sliding each parameter  $z_i$  across the parameter space and simultaneously monitoring the size of  $\chi_{k,n}^2$ , we determine the range

$$z_{i,-}^{(k)} \leq z_i \leq z_{i,+}^{(k)} \quad (\text{A.20})$$

that fulfills the criterion

$$\chi_{k,n}^2 \leq M(\chi_k^2). \quad (\text{A.21})$$



**Figure 16.** The 90% confidence limits for different data sets in the eigendirection direction  $z_{10}$ . The solid boxes indicate the locations of the minima for each  $\chi_k^2$ . No line is shown if the limits exceed the shown range. The global minimum is at  $z = 0$ .

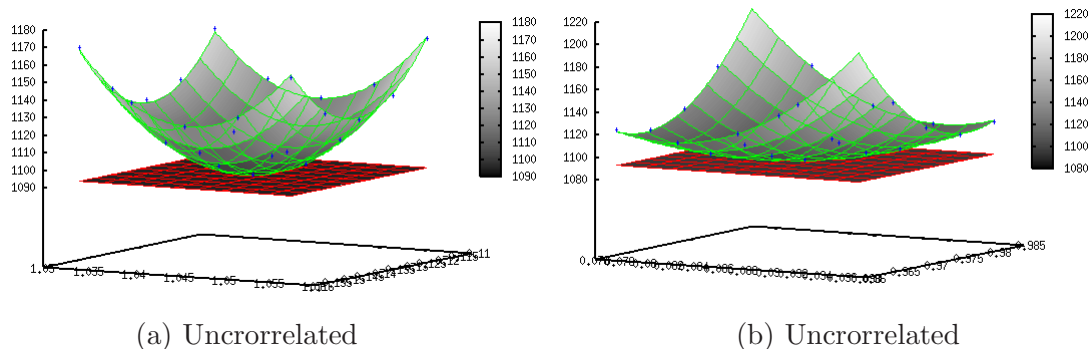
In figure 16, we display the  $P = 0.90$  limits for eigendirection  $z_{10}$  obtained by this procedure. The intersection  $\cap_k [z_{i,-}^{(k)}, z_{i,+}^{(k)}]$  gives us lower and upper bounds  $z_i^-$  and  $z_i^+$ , for which none of the individual  $\chi^2$ -contributions exceed their 90% confidence limits. This intersection is the shaded band in the figure 16.

Finally, denoting the increase of  $\chi^2$  in the boundaries by  $\Delta\chi^2(z_i^\pm)$ , the following average is taken as an overall estimate for  $\Delta\chi^2$ :

$$\Delta\chi^2 \equiv \sum_i \frac{\Delta\chi^2(z_i^+) + \Delta\chi^2(z_i^-)}{2N} \approx \sum_i \frac{(z_i^+)^2 + (z_i^-)^2}{2N}, \quad (\text{A.22})$$

where the latter relation would be exact in an ideal case with perfectly quadratic  $\chi^2$ . Performing such a procedure we find  $\Delta\chi^2 \approx 50$  — three times the value that would be indicated by the idealized statistical method.

We note that it is also possible to take  $\delta z_i^\pm = z_i^\pm$  in eq. (A.13). This choice is employed e.g. in a recent MSTW free proton analysis [51]. The benefit is that one circumvents the need for a single, “master”  $\Delta\chi^2$  — each direction  $z_i$  is treated separately. We have tested also this approach and, in comparison to the method described above, observed somewhat larger uncertainty predictions in those  $x$ -regions, where there are no stringent data constraints. This is caused by larger parameter intervals allowed in those directions which do not quickly meet the 90% confidence criterion, although the reproduction of several data sets gets worse, indicated by the approximate growth  $\Delta\chi^2 \approx (\delta z_i)^2$ . In other words,  $\Delta\chi^2$  for the individual error sets would exhibit large differences among themselves. For this reason, we choose not to follow this prescription, but construct the error sets in eq. (A.13) such that correspond to a fixed increase of  $\Delta\chi^2 = 50$ .



**Figure 17.** Two illustrative examples how our  $\chi^2$ -function may look in a two-parameter plane. The individual blue points are the exact  $\chi^2$  values and the shaded surface results from the fit of eq. (A.23). In the figure (a), the two parameters are fairly uncorrelated and the  $\chi^2$  increases almost uniformly in all directions. In the other case (b) there is a significant correlation between the two parameters revealed by the shallow “valley” going from a corner to another.

#### A.4 Calculation of the Hessian Matrix

A practical problem in applying the Hessian method is how to obtain a reliable approximation for the Hessian matrix. Ideally, if the  $\chi^2$ -function was smooth at the minimum and the  $\Delta\chi^2$  very small, we would obtain a useful Hessian matrix as a side product of the MINUIT minimization. In reality, our  $\chi^2$ -function is a complicated functional of the fit parameters and due to the finite precision of the DGLAP solver and multiple numerical integrations, it generally becomes non-continuous [55] in small parameter intervals. Such numerical noise makes it impossible to obtain a reliable Hessian matrix from a general purpose fitting program like MINUIT.

In order to obtain an estimate for the Hessian matrix, we adopt the following algorithm:

1. Determine the parameter step sizes  $\pm\delta a_i$  from their central values  $a_i^0$  that give an approximately fixed increase in  $\chi^2$ .
2. Compute the  $\chi^2$  values for each parameter pair in a two dimensional grid spanning the range  $(a_i, a_j) \in [a_i^0 - \delta a_i, a_i^0 + \delta a_i] \times [a_j^0 - \delta a_j, a_j^0 + \delta a_j]$ .
3. Fit the coefficients with quadratic function of the form

$$f(a_i, a_j) = \chi_0^2 + c_i(a_i - a_i^0)^2 + c_j(a_j - a_j^0)^2 + c_{ij}(a_i - a_i^0)(a_j - a_j^0) \quad (\text{A.23})$$

to the computed  $\chi^2$ -values at the grid points, and read off the Hessian matrix entries:  $H_{ij} = c_{ij}/2$ . The diagonal elements  $H_{ii}$  are obtained by a similar fit but in one dimension.

In figure 17 we show two examples of how the  $\chi^2$  typically looks as a function of two parameters. The shaded surface shows what is obtained from the fitted function of eq. (A.23). From figures like this, we have verified that the quadratic function (A.23) fits the neighborhood of the  $\chi_0^2$  well. These figures clearly reveal the two-parameter correlations too.

Obtaining a trustworthy approximation for the Hessian matrix from the above fitting method, requires knowledge of the desired  $\Delta\chi^2$  beforehand for adjusting the parameter intervals to be physically relevant: If too small or large step sizes are used to construct the Hessian matrix, the resulting quadratic approximation (A.1) may be bad for the  $\Delta\chi^2$  which will eventually be required by eq. (A.22). Therefore also this procedure is necessary somewhat iterative: To begin with, the Hessian matrix is computed by fixing the parameter step sizes to cause a fixed increase in  $\chi^2$ , say 25. Following the procedure outlined in earlier section A.3, an appropriate  $\Delta\chi^2$  is then determined. Inspecting the behaviour of  $\chi^2$  as a function of the deviations in each eigendirections, as shown in figure 15, one may judge whether the achieved quadratic approximation is valid or not near the obtained  $\Delta\chi^2$ . If needed, one may then try to adjust the parameter intervals in the computation of the Hessian accordingly, in order to improve the quadratic approximation.

## B How to speed up computation of pion production

In this appendix we sketch the method to speed up the numerical calculation of inclusive pion production cross-section in d + Au collisions

$$d(K_1) + Au(K_2) \rightarrow \pi(K_3) + X,$$

where  $K_1$  and  $K_2$  denote the incoming momenta, and  $K_3$  the momentum of the observed pion. The cross-section can be written as a convolution of the PDFs  $f$ , the fragmentation function  $D_{l \rightarrow \pi}$  for parton  $l$  to make a pion, and the partonic cross-section  $\hat{\sigma}$  [36]

$$E_3 \frac{d\sigma(d + Au \rightarrow \pi + X)}{d^3K_3} = \sum_{ijl} \int dx_1 \int dx_2 \int \frac{dx_3}{x_3^2} f_i^d(x_1, \mu_{\text{fact}}^2) f_j^{\text{Au}}(x_2, \mu_{\text{fact}}^2) \quad (\text{B.1})$$

$$D_{l \rightarrow \pi}(x_3, \mu_{\text{frag}}^2) p_3^0 \frac{d\hat{\sigma}(p_1^i + p_2^j \rightarrow p_3^l, \mu_{\text{ren}}^2)}{d^3p_3} \left| \begin{array}{l} p_1 = x_1 K_1 \\ p_2 = x_2 K_2 \\ p_3 = K_3 / x_3 \end{array} \right.$$

Here one faces a practical problem of significantly extended computing time consumed in evaluating the triple integrals, apparently making the fitting procedure — involving tens of thousands of calls — very slow. However, if the factorization scale  $\mu_{\text{fact}}^2$  is chosen not to depend on the partonic momenta, e.g. fixing all scales to a common scale  $\mu^2$  proportional to the transverse momentum of the pion, one can define a set of weighting functions

$$F_j(x_2, \mu^2) = \sum_{il} \int dx_1 \int \frac{dx_3}{x_3^2} f_i^d(x_1, \mu^2) D_{l \rightarrow h}(x_3, \mu^2) p_3^0 \frac{d\hat{\sigma}(p_1^i + p_2^j \rightarrow p_3^l, \mu^2)}{d^3p_3}, \quad (\text{B.2})$$

which can be evaluated and tabulated in advance as  $F_j(x_2)$  do not depend on the nuclear PDFs. After such preparations, the eq. (B.1) can be evaluated in advance up a to single integral

$$E_3 \frac{d\sigma}{d^3K_3} = \sum_j \int dx_2 F_j(x_2, \mu^2) f_j^{\text{Au}}(x_2, \mu^2). \quad (\text{B.3})$$

With the above trick we overcome the problem of slowness and make the inclusion of hadron production data feasible also in NLO. We should also note that this method is possible as we take the free proton (and deuterium) PDFs as fully known, i.e these are not subject to iteration here. If this were not the case, only one integral could be predone in eq. (B.1).

## References

- [1] 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](#)] [[SPIRES](#)].
- [2] K.J. Eskola, V.J. Kolhinen and C.A. Salgado, *The scale dependent nuclear effects in parton distributions for practical applications*, *Eur. Phys. J. C* **9** (1999) 61 [[hep-ph/9807297](#)] [[SPIRES](#)].
- [3] K.J. Eskola, V.J. Kolhinen, H. Paukkunen and C.A. Salgado, *A global reanalysis of nuclear parton distribution functions*, *JHEP* **05** (2007) 002 [[hep-ph/0703104](#)] [[SPIRES](#)].
- [4] 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](#)] [[SPIRES](#)].
- [5] M. Hirai, S. Kumano and T.H. Nagai, *Determination of nuclear parton distribution functions and their uncertainties at next-to-leading order*, *Phys. Rev. C* **76** (2007) 065207 [[arXiv:0709.3038](#)] [[SPIRES](#)].
- [6] D. de Florian and R. Sassot, *Nuclear parton distributions at next to leading order*, *Phys. Rev. D* **69** (2004) 074028 [[hep-ph/0311227](#)] [[SPIRES](#)].
- [7] M. Hirai, S. Kumano and T.H. Nagai, *Nuclear parton distribution functions and their uncertainties*, *Phys. Rev. C* **70** (2004) 044905 [[hep-ph/0404093](#)] [[SPIRES](#)].
- [8] E.R. Cazaroto, F. Carvalho, V.P. Goncalves and F.S. Navarra, *Constraining the nuclear gluon distribution in  $eA$  processes at RHIC*, *Phys. Lett. B* **669** (2008) 331 [[arXiv:0804.2507](#)] [[SPIRES](#)].
- [9] D. Stump et al., *Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method*, *Phys. Rev. D* **65** (2002) 014012 [[hep-ph/0101051](#)] [[SPIRES](#)].
- [10] J. Pumplin et al., *Uncertainties of predictions from parton distribution functions. 2. The hessian method*, *Phys. Rev. D* **65** (2001) 014013 [[hep-ph/0101032](#)] [[SPIRES](#)].
- [11] See <https://www.jyu.fi/fysiikka/en/research/highenergy/urhic/nPDFs>.
- [12] D. Stump et al., *Inclusive jet production, parton distributions and the search for new physics*, *JHEP* **10** (2003) 046 [[hep-ph/0303013](#)] [[SPIRES](#)].
- [13] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012 [[hep-ph/0201195](#)] [[SPIRES](#)].
- [14] Y.L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and  $e^+e^-$  Annihilation by Perturbation Theory in Quantum Chromodynamics. (In Russian)*, *Sov. Phys. JETP* **46** (1977) 641 [*Zh. Eksp. Teor. Fiz.* **73** (1977) 1216] [[SPIRES](#)]; V.N. Gribov and L.N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438 [*Yad. Fiz.* **15** (1972) 781] [[SPIRES](#)];  *$e^+e^-$  pair annihilation and deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 675 [*Sov. J. Nucl. Phys.* **15** (1972) 675] [[SPIRES](#)];

- G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nucl. Phys. B* **126** (1977) 298 [SPIRES].
- [15] G. Curci, W. Furmanski and R. Petronzio, *Evolution of parton densities beyond leading order: the nonsinglet case*, *Nucl. Phys. B* **175** (1980) 27 [SPIRES].
- [16] W. Furmanski and R. Petronzio, *Singlet parton densities beyond leading order*, *Phys. Lett. B* **97** (1980) 437 [SPIRES].
- [17] R.K. Ellis, W.J. Stirling and B.R. Webber, *QCD and collider physics*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1 [SPIRES].
- [18] P. Santorelli and E. Scrimieri, *A semianalytical method to evolve parton distributions*, *Phys. Lett. B* **459** (1999) 599 [hep-ph/9807572] [SPIRES].
- [19] W. Giele et al., *The QCD/SM working group: summary report*, hep-ph/0204316 [SPIRES].
- [20] J. Gomez et al., *Measurement of the A-dependence of deep inelastic electron scattering*, *Phys. Rev. D* **49** (1994) 4348 [SPIRES].
- [21] NEW MUON collaboration, P. Amaudruz et al., *A Reevaluation of the nuclear structure function ratios for D, He, Li-6, C and Ca*, *Nucl. Phys. B* **441** (1995) 3 [hep-ph/9503291] [SPIRES].
- [22] NEW MUON. collaboration, M. Arneodo et al., *The structure function ratios  $F_2(\text{Li})/F_2(\text{D})$  and  $F_2(\text{C})/F_2(\text{D})$  at small x*, *Nucl. Phys. B* **441** (1995) 12 [hep-ex/9504002] [SPIRES].
- [23] NEW MUON collaboration, M. Arneodo et al., *The A dependence of the nuclear structure function ratios*, *Nucl. Phys. B* **481** (1996) 3 [SPIRES].
- [24] D.M. Alde et al., *Nuclear dependence of dimuon production at 800 GeV. FNAL-772 experiment*, *Phys. Rev. Lett.* **64** (1990) 2479 [SPIRES].
- [25] FNAL E866 collaboration, M.A. Vasilev et al., *Parton energy loss limits and shadowing in Drell-Yan dimuon production*, *Phys. Rev. Lett.* **83** (1999) 2304 [hep-ex/9906010] [SPIRES].
- [26] EUROPEAN MUON collaboration, J. Ashman et al., *A measurement of the ratio of the nucleon structure function in copper and deuterium*, *Z. Phys. C* **57** (1993) 211 [SPIRES].
- [27] NEW MUON collaboration, M. Arneodo et al., *The  $Q^2$  dependence of the structure function ratio  $F_2 \text{ Sn}/F_2 \text{ C}$  and the difference  $R \text{ Sn}-R \text{ C}$  in deep inelastic muon scattering*, *Nucl. Phys. B* **481** (1996) 23 [SPIRES].
- [28] PHENIX collaboration, S.S. Adler et al., *Centrality dependence of  $\pi^0$  and eta production at large transverse momentum in  $s(\text{NN})^{1/2} = 200 \text{ GeV } d + \text{Au}$  collisions*, *Phys. Rev. Lett.* **98** (2007) 172302 [nucl-ex/0610036] [SPIRES].
- [29] W. Furmanski and R. Petronzio, *Lepton-hadron processes beyond leading order in quantum chromodynamics*, *Zeit. Phys. C* **11** (1982) 293 [SPIRES].
- [30] C. Anastasiou, L.J. Dixon, K. Melnikov and F. Petriello, *Dilepton rapidity distribution in the Drell-Yan process at NNLO in QCD*, *Phys. Rev. Lett.* **91** (2003) 182002 [hep-ph/0306192] [SPIRES].
- [31] B.A. Kniehl, G. Kramer and B. Potter, *Fragmentation functions for pions, kaons and protons at next-to-leading order*, *Nucl. Phys. B* **582** (2000) 514 [hep-ph/0010289] [SPIRES].
- [32] 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] [SPIRES].

- [33] D. de Florian, R. Sassot and M. Stratmann, *Global analysis of fragmentation functions for protons and charged hadrons*, *Phys. Rev. D* **76** (2007) 074033 [[arXiv:0707.1506](#)] [[SPIRES](#)].
- [34] D. de Florian, R. Sassot and M. Stratmann, *Global analysis of fragmentation functions for pions and kaons and their uncertainties*, *Phys. Rev. D* **75** (2007) 114010 [[hep-ph/0703242](#)] [[SPIRES](#)].
- [35] See [http://wwwlapp.in2p3.fr/lapth/PHOX\\_FAMILY/readme\\_inc.html](http://wwwlapp.in2p3.fr/lapth/PHOX_FAMILY/readme_inc.html).
- [36] F. Aversa, P. Chiappetta, M. Greco and J.P. Guillet, *QCD corrections to parton-parton scattering processes*, *Nucl. Phys. B* **327** (1989) 105 [[SPIRES](#)].
- [37] BRAHMS collaboration, I. Arsene et al., *On the evolution of the nuclear modification factors with rapidity and centrality in  $d + Au$  collisions at  $s(NN)^{1/2} = 200$  GeV*, *Phys. Rev. Lett.* **93** (2004) 242303 [[nucl-ex/0403005](#)] [[SPIRES](#)].
- [38] F. James and M. Roos, *Minuit: a system for function minimization and analysis of the parameter errors and correlations*, *Comput. Phys. Commun.* **10** (1975) 343 [[SPIRES](#)].
- [39] NNPDF collaboration, L. Del Debbio et al., *Neural network determination of parton distributions: the nonsinglet case*, *JHEP* **03** (2007) 039 [[hep-ph/0701127](#)] [[SPIRES](#)].
- [40] NNPDF collaboration, *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys. B* **809** (2009) 1 [[arXiv:0808.1231](#)] [[SPIRES](#)].
- [41] P.M. Nadolsky and Z. Sullivan, *PDF uncertainties in  $W H$  production at Tevatron*, in the proceedings of the *APS/DPF/DPB Summer Study on the Future of Particle Physics (Snowmass2001)*, June 30–July 21, Snowmass, Colorado, U.S.A. (2001), [hep-ph/0110378](#) [[SPIRES](#)].
- [42] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, *Uncertainties of predictions from parton distributions. I: theoretical errors*, *Eur. Phys. J. C* **35** (2004) 325 [[hep-ph/0308087](#)] [[SPIRES](#)].
- [43] M. Dittmar et al., *Parton Distributions*, [arXiv:0901.2504](#) [[SPIRES](#)].
- [44] J.-W. Qiu and I. Vitev, *Resummed QCD power corrections to nuclear shadowing*, *Phys. Rev. Lett.* **93** (2004) 262301 [[hep-ph/0309094](#)] [[SPIRES](#)].
- [45] L.V. Gribov, E.M. Levin and M.G. Ryskin, *Semihard processes in QCD*, *Phys. Rept.* **100** (1983) 1 [[SPIRES](#)].
- [46] A.H. Mueller and J.-W. Qiu, *Gluon recombination and shadowing at small values of  $x$* , *Nucl. Phys. B* **268** (1986) 427 [[SPIRES](#)].
- [47] Z.-B. Kang and J.-W. Qiu, *Nuclear modification to parton distribution functions and parton saturation*, *J. Phys. G* **34** (2007) S607 [[hep-ph/0702040](#)] [[SPIRES](#)].
- [48] S.A. Kulagin and R. Petti, *Global study of nuclear structure functions*, *Nucl. Phys. A* **765** (2006) 126 [[hep-ph/0412425](#)] [[SPIRES](#)].
- [49] PARTICLE DATA GROUP collaboration, C. Amsler et al., *Review of particle physics*, *Phys. Lett. B* **667** (2008) 1 [[SPIRES](#)].
- [50] STAR collaboration, J. Adams et al., *Identified hadron spectra at large transverse momentum in  $p + p$  and  $d + Au$  collisions at  $s(NN)^{1/2} = 200$  GeV*, *Phys. Lett. B* **637** (2006) 161 [[nucl-ex/0601033](#)] [[SPIRES](#)].

- [51] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Parton distributions for the LHC*, [arXiv:0901.0002](#) [[SPIRES](#)].
- [52] P.M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, *Phys. Rev. D* **78** (2008) 013004 [[arXiv:0802.0007](#)] [[SPIRES](#)].
- [53] I. Schienbein et al., *Nuclear PDFs from neutrino deep inelastic scattering*, *Phys. Rev. D* **77** (2008) 054013 [[arXiv:0710.4897](#)] [[SPIRES](#)].
- [54] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, *Uncertainties of predictions from parton distributions. 1: experimental errors*, *Eur. Phys. J. C* **28** (2003) 455 [[hep-ph/0211080](#)] [[SPIRES](#)].
- [55] J. Pumplin, D.R. Stump and W.K. Tung, *Multivariate fitting and the error matrix in global analysis of data*, *Phys. Rev. D* **65** (2002) 014011 [[hep-ph/0008191](#)] [[SPIRES](#)].