



# Unveiling saturation effects from nuclear structure function measurements at the EIC



Cyrille Marquet<sup>a</sup>, Manoel R. Moldes<sup>b,\*</sup>, Pía Zurita<sup>b,c</sup>

<sup>a</sup> Centre de Physique Théorique, École Polytechnique, CNRS, Université Paris-Saclay, 91128 Palaiseau, France

<sup>b</sup> Departamento de Física de Partículas and IGFAE, Universidade de Santiago de Compostela, 15706 Santiago de Compostela, Galicia, Spain

<sup>c</sup> Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA

## ARTICLE INFO

### Article history:

Received 9 February 2017

Received in revised form 6 July 2017

Accepted 17 July 2017

Available online 21 July 2017

Editor: J.-P. Blaizot

## ABSTRACT

We analyze the possibility of extracting a clear signal of non-linear parton saturation effects from future measurements of nuclear structure functions at the Electron–Ion Collider (EIC), in the small- $x$  region. Our approach consists in generating pseudodata for electron–gold collisions, using the running-coupling Balitsky–Kovchegov evolution equation, and in assessing the compatibility of these *saturated* pseudodata with existing sets of nuclear parton distribution functions (nPDFs), extrapolated if necessary. The level of disagreement between the two is quantified by applying a Bayesian reweighting technique. This allows to infer the parton distributions needed in order to describe the pseudodata, which we find quite different from the actual distributions, especially for sea quarks and gluons. This tension suggests that, should saturation effects impact the future nuclear structure function data as predicted, a successful refitting of the nPDFs may not be achievable, which would unambiguously signal the presence of non-linear effects.

© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Funded by SCOAP<sup>3</sup>.

## 1. Introduction

It is foreseen that the next high-energy nuclear physics facility will be an Electron–Ion Collider (EIC) [1]. One of the main physics goals of this future QCD laboratory will be to unambiguously unveil the onset of the so-called gluon saturation regime of QCD. This regime of hadronic and nuclear wave functions at small longitudinal momentum fraction  $x$  is characterized by a transverse momentum scale, the saturation scale  $Q_s(x)$ , at which non-linearities become of comparable importance to linear evolution [2]. The emergence of this non-linear regime is a fundamental consequence of QCD dynamics, and it has been subject to steady theoretical progress for the past twenty years. However the general consensus is that a clear experimental discovery remains elusive.

Saturation phenomenology has been successful in all the (admittedly small) phase-space corners where that physics is expected to be relevant (i.e., for every collider process that involves small- $x$  partons and transverse momenta of the order of the saturation scale), for a broad range of observables and various collision systems, including  $d + Au$  collisions at RHIC and in  $e + p$  collisions at HERA [3,4]. The strongest hint in favor of saturation so far is

arguably the suppression of the away-side peak of the di-hadron correlation function at forward rapidities in central  $p + A$  collisions versus  $p + p$  collisions [5,6], discovered at RHIC after it had been predicted [7]. Nevertheless, the global belief in the community is that a direct, unquestionable evidence is still missing, and that the EIC may provide it.

Given this past success, there is little doubt that saturation effects will be at play in  $e + Au$  collisions at the EIC. The question is which is the best way to uncover them, which are the golden observables and smoking-gun measurements. Measurements from which one can undoubtedly conclude that non-linear effects are clearly needed on top of standard linear perturbative QCD dynamics, but also from which one can thoroughly verify that those non-linear effects are under theoretical control, i.e., they are described by non-linear QCD equations as opposed to being fully of non-perturbative origin. In this work, we shall focus on the  $F_2^{\text{Au}}$  and  $F_L^{\text{Au}}$  structure functions at small  $x$ , whose combined measurement is often mentioned as the primary candidate to provide a smoking gun in favor of parton saturation [8]. They provide access to the distributions of quarks and gluons, respectively, and can be extracted from measuring the total  $e + A \rightarrow e + X$  cross-section at different collision energies.

However, even if saturation effects would play an important role in the kinematical range of the future EIC data, there is a possibility that such effects would go unnoticed: standard QCD fits of

\* Corresponding author.

E-mail address: [manoel.rodriguez-moldes@usc.es](mailto:manoel.rodriguez-moldes@usc.es) (M.R. Moldes).

parton distribution functions (PDFs), which are based on the linear DGLAP evolution equation and therefore do not take into account saturation effects, could still be able to describe structure function data inside the saturation regime [9]. This is due to the large amount of freedom allowed in this approach for the initial conditions to the QCD evolution: those can be artificially suppressed in fits, in order to compensate for an evolution that would be too fast, as would be the case if the DGLAP equation was used into the saturation regime without the proper non-linear corrections.

In other words, up to a certain point, it is possible to obtain a successful fit of  $F_2$  and  $F_L$  data (and a resulting PDF set) with an incomplete evolution equation, which can unfortunately lead to wrong conclusions and an inaccurate description of the parton content. The question we set out to answer in this letter is whether or not saturation effects, as well as the lever arm in  $x$ , will be large enough in order to ensure that this will not happen with gold nuclei at the EIC, and that the combined measurement of the  $F_2$  and  $F_L$  structure functions is indeed a golden observable. We note that the possibility of detecting non-linear effects with inclusive measurements has been investigated in earlier works [10–12]; the present work utilizes of a novel approach.

Our strategy is the following. We first generate  $F_2^{\text{Au}}$  and  $F_L^{\text{Au}}$  pseudodata, with saturation effects setting in according to the latest predictions based on the running-coupling Balitsky–Kovchegov (rcBK) evolution [13–18]. Then we answer the question whether various nuclear PDF (nPDF) fits on the market (namely<sup>1</sup> EPS09 [19] and DSSZ [20]), which are well constrained by large- $x$  large- $Q^2$  existing data, can accommodate our EIC pseudodata. This is done using a Bayesian reweighing procedure which addresses the compatibility of new data (and the quantitative modifications they might induce) to an existing set of PDFs [21–27]. If the answer is positive, then saturation effects at the EIC would stay concealed; if the answer is negative, then non-linear QCD dynamics would be needed to describe the data. In the former case, measuring deep inelastic scattering (DIS) off heavy nuclei would still be crucial in order to extract nuclear modifications of quark and gluon distributions at generic values of  $x$ , but the structure functions would not form smoking-gun measurements for parton saturation.

The plan of the letter is as follows. In section 2, we detail the saturation framework used to generate our structure-function pseudodata for gold nuclei at the EIC. In section 3, we explain the most relevant features of the Bayesian reweighing procedure used to check the compatibility of our pseudodata with the nPDFs sets. The obtained results are the topic of Sec. 4. Finally we summarize our findings in Sec. 5.

## 2. Nuclear structure function pseudodata from the rcBK saturation model

It is customary to write the DIS structure functions  $F_2$  and  $F_L$  as follows:

$$\begin{aligned} F_2(x, Q^2) &= \frac{4\pi^2\alpha_{em}}{Q^2} \left[ \sigma_T^{\gamma^*A}(x, Q^2) + \sigma_L^{\gamma^*A}(x, Q^2) \right], \\ F_L(x, Q^2) &= \frac{4\pi^2\alpha_{em}}{Q^2} \sigma_L^{\gamma^*A}(x, Q^2), \end{aligned} \quad (1)$$

<sup>1</sup> The choice of these two sets obey to practical reasons. Though nowadays there are several sets of nPDFs available, EPS09 presents the most extreme effect on the gluon density, while DSSZ shows almost no difference between the proton and nuclear gluon. Thus, by considering these two sets, we cover all possible spectrum of the known nPDFs.

where  $\sigma_{L,T}^{\gamma^*A}$  represents the  $\gamma^* + A$  total cross-section, for a transversely (T) or longitudinally (L) polarized virtual photon. In turn, at small  $x$ , these are obtained in the following way:

$$\sigma_{L,T}^{\gamma^*A}(x, Q^2) = \int d^2r \int_0^1 dz \left| \Psi_{L,T}^{\gamma^*}(z, \mathbf{r}; Q^2) \right|^2 \int d^2b \frac{d\sigma_{\text{dip}}^A}{d^2b}(\mathbf{r}, \mathbf{b}; x), \quad (2)$$

where  $\left| \Psi_{L,T}^{\gamma^*}(z, \mathbf{r}; Q^2) \right|^2$  represents the probability for a photon of virtuality  $Q^2$  to produce a  $q\bar{q}$  pair of transverse size  $\mathbf{r}$  and  $z$  denotes the fraction of longitudinal momentum carried by the quark.  $d\sigma_{\text{dip}}^A/d^2b$  stands for the total cross-section for this  $q\bar{q}$  pair to scatter off the target nucleus at an impact parameter  $\mathbf{b}$  and contains the QCD dynamics. By contrast,  $\Psi_{L,T}^{\gamma^*}$  is well known from QED.

To compute the  $q\bar{q}$  dipole-nucleus cross-section, we follow the approach of [28,29]. Using the optical theorem, the total cross-section can be written  $d\sigma_{\text{dip}}^A/d^2b = 2(1 - S_A)$  where the  $S$ -matrix element  $S_A$  represents the probability amplitude for the dipole to *not* interact with the target nucleus. Then, the essence of the model is to write that the dipole scatters independently off the different nucleons. Introducing the coordinates of the individual nucleons  $\{\mathbf{b}_i\}$ , this means  $S_A(\mathbf{r}, \mathbf{b}; x) = \prod_{i=1}^A S_p(\mathbf{r}, \mathbf{b} - \mathbf{b}_i; x)$ , where  $S_p$  refers to the  $q\bar{q}$  dipole-nucleon  $S$ -matrix element.

We further assume that the positions of the nucleons  $\{\mathbf{b}_i\}$  are distributed according to the Woods-Saxon distribution  $T_A(\mathbf{b})$ :

$$T_A(\mathbf{b}) = \int dz \frac{C}{1 + \exp\left[\left(\sqrt{\mathbf{b}^2 + z^2} - R_A\right)/d\right]}, \quad (3)$$

which is normalized to unity  $\int d^2b T_A(\mathbf{b}) = 1$ . The nuclear radius  $R_A$  and surface diffuseness  $d$  are measured from the electric charge distribution, their values can be found in Ref. [30].

In this model, the  $q\bar{q}$  dipole-nucleus cross-section can therefore be written

$$\frac{d\sigma_{\text{dip}}^A}{d^2b}(\mathbf{r}, \mathbf{b}; x) = \int \prod_{i=1}^A \left\{ d^2b_i T_A(\mathbf{b}_i) \right\} 2 \left[ 1 - \prod_{i=1}^A S_p(\mathbf{r}, \mathbf{b} - \mathbf{b}_i; x) \right], \quad (4)$$

which is well approximated by [29]

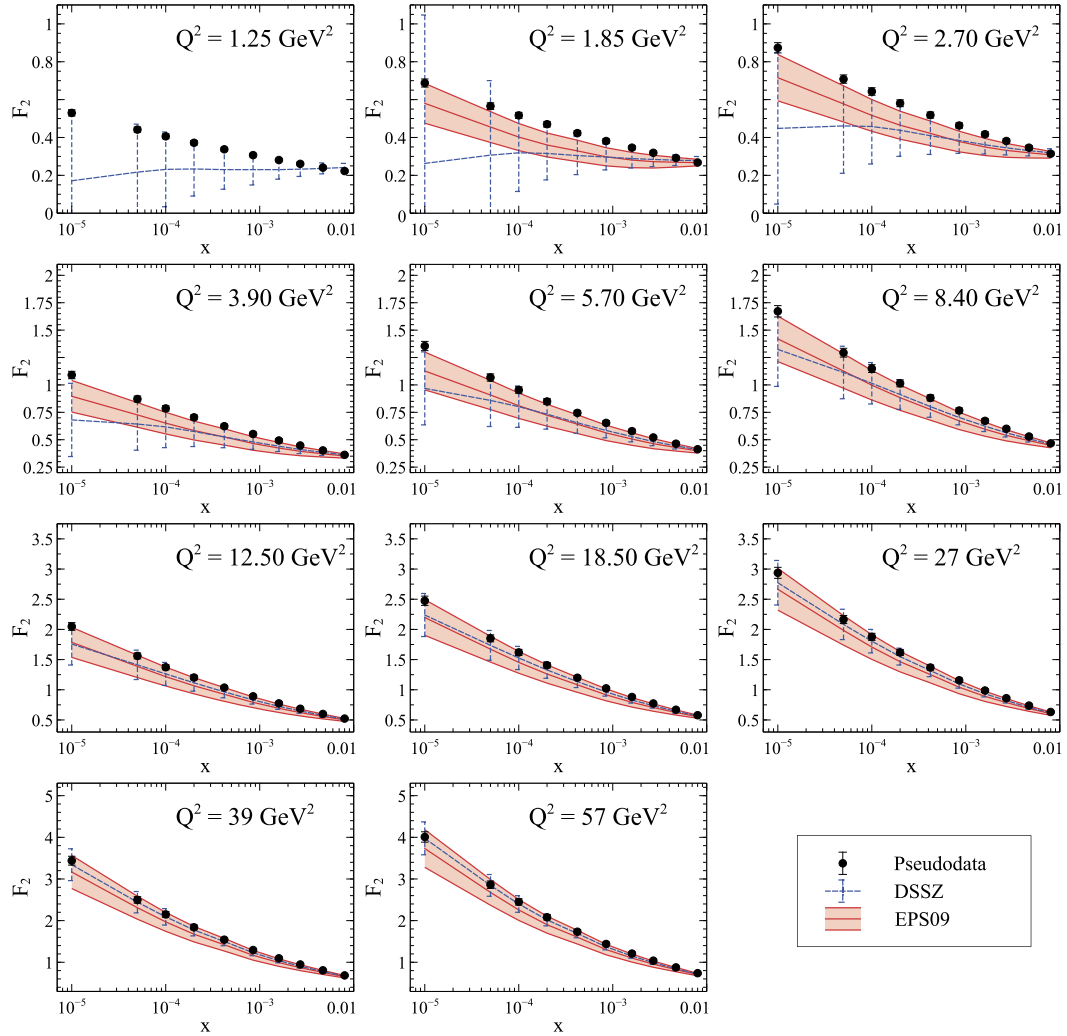
$$\frac{d\sigma_{\text{dip}}^A}{d^2b} \approx 2 \left[ 1 - \left( 1 - \frac{T_A(\mathbf{b})}{2} \sigma_{\text{dip}}^p \right)^A \right]. \quad (5)$$

The expression in parenthesis in (5) can be further replaced by  $\exp(-AT_A(\mathbf{b})\sigma_{\text{dip}}^p/2)$  for large  $A$ . The parameters of the model come from the Woods–Saxon distribution (3) and from the AAMQS fit [33] of the  $q\bar{q}$  dipole-nucleon cross-section  $\sigma_{\text{dip}}^p$  to the reduced cross-section data from HERA, which we shall detail now.

In the AAMQS fit, the dipole cross-section is expressed in terms of an impact-parameter independent dipole scattering amplitude  $\mathcal{N}$ , which contains the small- $x$  QCD dynamics:

$$\sigma_{\text{dip}}^p(\mathbf{r}, x) = \sigma_0 \mathcal{N}(\mathbf{r}, x). \quad (6)$$

We note that other models could be used here, in particular models that implement a non-trivial impact parameter dependence already at the level of dipole-nucleon cross section, however, this does not lead to important differences for the inclusive observables considered in this work. It would in the case of diffractive and exclusive observables, but this goes beyond the scope of this paper.



**Fig. 1.**  $F_2^{\text{Au}}$  pseudodata (normalized to the number of nucleons  $A = 197$ ) obtained from the rcBK saturation model, and the corresponding values computed in the collinear factorization framework with the EPS09 and DSSZ nPDFs.

The dipole scattering amplitude  $\mathcal{N}$  obeys the rcBK equation [13–18]:

$$\frac{\partial \mathcal{N}(r, Y)}{\partial Y} = \int d^2 \mathbf{r}_1 K^{\text{run}}(\mathbf{r}, \mathbf{r}_1, \mathbf{r}_2) [\mathcal{N}(r_1, Y) + \mathcal{N}(r_2, Y) - \mathcal{N}(r, Y) - \mathcal{N}(r_1, Y) \mathcal{N}(r_2, Y)], \quad (7)$$

with  $Y = \ln(x_0/x)$ . Using Balitsky's prescription [17], the kernel in (7) reads

$$K^{\text{run}}(\mathbf{r}, \mathbf{r}_1, \mathbf{r}_2) = \frac{N_c \alpha_s(r^2)}{2\pi^2} \left[ \frac{r^2}{r_1^2 r_2^2} + \frac{1}{r_1^2} \left( \frac{\alpha_s(r_1^2)}{\alpha_s(r_2^2)} - 1 \right) + \frac{1}{r_2^2} \left( \frac{\alpha_s(r_2^2)}{\alpha_s(r_1^2)} - 1 \right) \right], \quad (8)$$

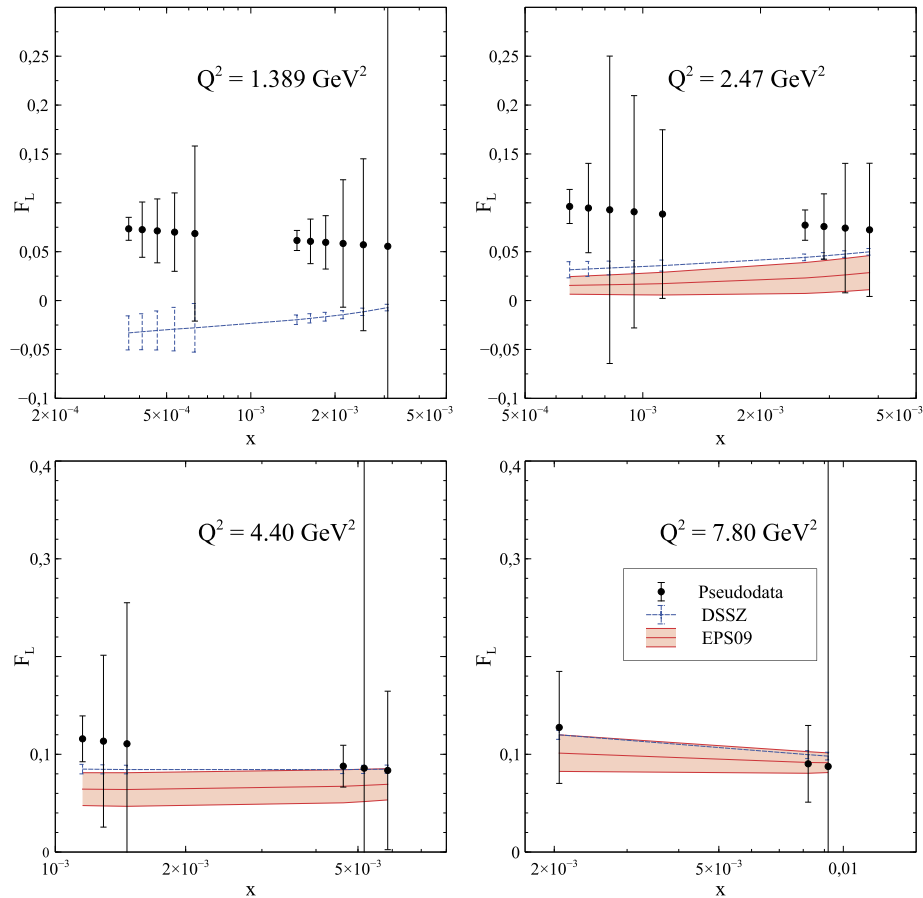
where  $\mathbf{r}_2 = \mathbf{r} - \mathbf{r}_1$  and  $v \equiv |\mathbf{v}|$  for two-dimensional vectors. Following [31], the running coupling is regulated in equations (7) and (8) by freezing it to a constant value  $\alpha_s^{fr} = 0.7$  in the infrared. The initial condition for the evolution is the so-called McLerran-Venugopalan (MV) model [32]:

$$\mathcal{N}_F(r, Y=0) = 1 - \exp \left[ -\frac{r^2 Q_{s0}^2}{4} \ln \left( \frac{1}{\Lambda r} + e \right) \right], \quad (9)$$

where  $Q_{s0}^2$  is the initial saturation scale, and where  $\Lambda = 0.241$  GeV. All the parameters can be found in [33], we shall make use of the “e” fit of that reference. Recently, more advanced fits have appeared where Eq. (7) is also supplemented with collinear resummations [34]; however those are not yet available for public use, but could as easily be used in the future.

With all the ingredients described above, we can generate the central values of our pseudodata for  $F_2^{\text{Au}}$  and  $F_L^{\text{Au}}$ . Concerning the projected error bars, they have been generated for us by the “BNL Task Force for the EIC” [35]. As in many EIC dedicated studies, an integrated luminosity of  $10 \text{ fb}^{-1}$  was assumed, implying that systematic uncertainties dominate. The latter are estimated to be around 3% for  $F_2$ , and for  $F_L$ , three center-of-mass energies ranging from 30 to 90 GeV were assumed for the extraction, yielding bigger errors. In Figs. 1 and 2, we show a comparison of those pseudodata, with calculations obtained from the EPS09 and DSSZ nPDF sets, with  $x$  ranging from  $10^{-5}$  to  $10^{-2}$ , with  $Q^2$ -values within the perturbative QCD range. These theoretical curves are the result of incorporating the nuclear PDFs into the next-to-leading order code for structure functions from Ref. [36], meaning that we are using the same MSTW proton baseline for both our EPS09 and DSSZ nPDF calculations.

In the case of  $F_2$  (Fig. 1), the pseudodata at the highest  $x$ -values are well within the theoretical uncertainties for both nPDF sets



**Fig. 2.**  $F_L^{Au}$  pseudodata (normalized to the number of nucleons  $A = 197$ ) obtained from the rcBK saturation model, and the corresponding values computed in the collinear factorization framework with the EPS09 and DSSZ nPDFs.

considered. In contrast, for  $x \lesssim 10^{-3}$  the rcBK prediction lies in the upper limit of the nPDFs uncertainties or even above them. This is due to the fact that nPDFs fits are, at the moment, not constrained at such low  $x$ -values, and the theoretical predictions shown on the figures rely on extrapolations. Moreover, contrary to what one could have naively expected, the nPDF extrapolations induce  $F_2$  values which are smaller than what the saturation model predicts, not larger. This means that, with the current nPDF sets, one should not expect the *golden* scenario in which the collinear factorization predictions will overshoot the data at small  $x$ , due to the lack of non-linear effects in that framework.

With respect to the longitudinal structure function, there are two main reasons that explain the extreme differences seen in Fig. 2. On the one hand, as before, part of the generated pseudodata lie in an unexplored kinematical region and thus the nPDF extrapolations are not reliable. On the other hand,  $F_L$  is sensitive to the gluon distribution already at the lowest order, unlike  $F_2$ . This gluon density is not well determined, as the bulk of data considered in nPDFs extractions is not sensitive to it, either due to the kinematical range covered or to the observables themselves.

All these reasons come together to explain the level of disagreement between the *saturated* pseudodata and the collinear factorization predictions. And to illustrate that these discrepancies relate to nuclear effects, and not to differences at the level of the proton structure functions, we show in Fig. 3 that indeed the AAMQS and the collinear-factorization calculations are compatible, as they should since these are build upon the same HERA data.

Finally, it is noteworthy that the upper left panel in both Figs. 1 and 2, corresponding to the smallest  $Q^2$ -value, only has a predic-

tion for DSSZ, as it is outside the validity range of EPS09. Given this, the total number of pseudodata included in the analysis below is not the same for each nPDF set. We would also like to point out that the projected error bars for the  $F_L$  measurements at the EIC are much bigger than those in the case of  $F_2$ , and as a consequence including our  $F_L$  pseudodata in the analysis makes little difference.

### 3. Bayesian reweighting of nuclear PDFs

The partonic densities (PDFs) are a necessary piece for the theoretical predictions of physical observables for processes involving at least one hadron in the initial or final state. However, their determination from experimental data is an involved and time consuming procedure. For this reason, updating the PDFs every time a new set of data becomes available is rather vexing, as a priori one can't know if new information would be obtained from it.

Therefore, statistical methods have been developed in order to bypass this obstacle: the reweighting techniques are tools that allow to incorporate information from newly measured data into a set of PDFs without recurring to the standard procedure of global fitting. At least two methods are now available: the Hessian reweighting and the Bayesian reweighting [21–27]. While in this work we chose the latter, it has been shown that they are equivalent for PDF sets with theoretical uncertainties determined by the Hessian method as the ones we use.

Let us assume that we have the representation of the underlying probability distribution  $\mathcal{P}_{old}(f)$  of the PDFs given by a large ensemble of PDFs  $f_k$ ,  $k = 1 \dots N_{rep}$ . Then the expectation value for any quantity  $\mathcal{O}$  depending on the PDFs can be computed as

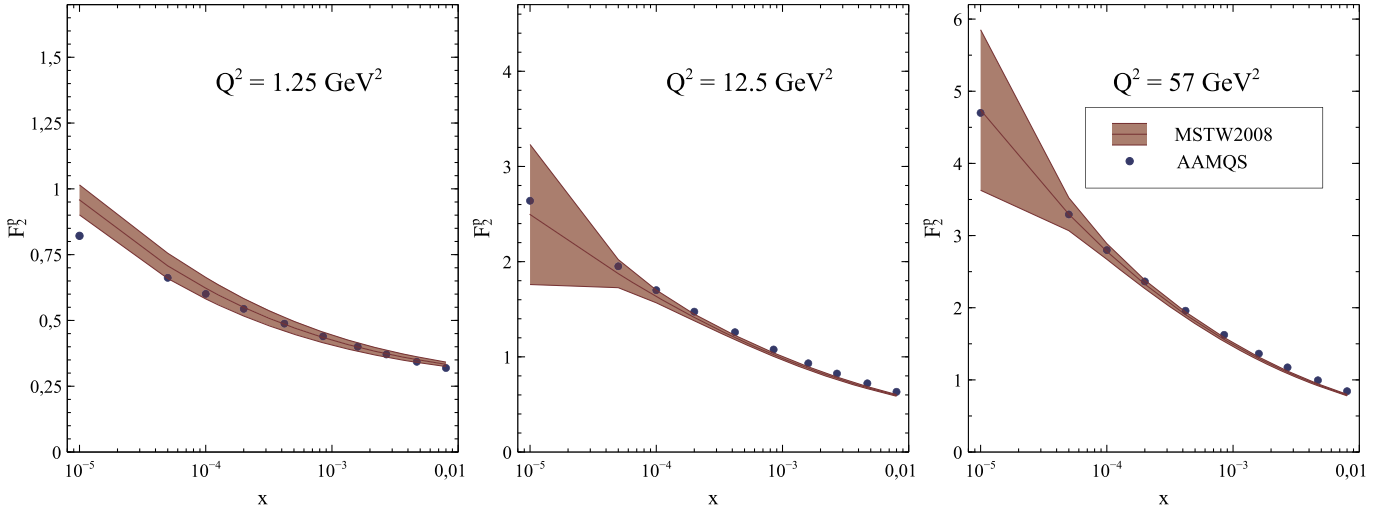


Fig. 3. Comparison of  $F_2^p$  predictions obtained from the rcBK saturation model and the collinear-factorization framework with the MSTW PDFs.

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \mathcal{O}[f_k], \quad (10)$$

with variance

$$\delta(\mathcal{O}) = \sqrt{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} (\mathcal{O}[f_k] - \langle \mathcal{O} \rangle)^2}. \quad (11)$$

If new data  $\vec{y} \equiv \{y_i, i = 1, \dots, N_{\text{data}}\}$  with covariance matrix  $C_{ij}$  are made available, we can update  $\mathcal{P}_{\text{old}}(f)$  incorporating the information contained in the new data by use of the Bayes theorem as

$$\mathcal{P}_{\text{new}}(f) \propto \mathcal{P}(\vec{y}|f) \mathcal{P}_{\text{old}}(f), \quad (12)$$

where  $\mathcal{P}(\vec{y}|f)$  is the likelihood for the new data given the original set of parton densities. With this modification the quantities defined in equations (10) and (11) turn into weighted averages

$$\langle \mathcal{O} \rangle_{\text{new}} = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \omega_k \mathcal{O}[f_k], \quad (13)$$

$$\delta(\mathcal{O})_{\text{new}} = \sqrt{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \omega_k (\mathcal{O}[f_k] - \langle \mathcal{O} \rangle_{\text{new}})^2}, \quad (14)$$

with the weights  $\omega_k$  proportional to the likelihood.

The adequate selection of the likelihood is a delicate matter and several options have been proposed [26,21–23]. In particular, we follow the option of [27], equivalent to a refit for PDF sets with uncertainties based on the Hessian method (with  $N_{\text{eig}}$  eigenvalues and fixed tolerance  $\Delta\chi^2$ ):

$$\omega_k = \frac{\exp[-\chi_k^2/2\Delta\chi^2]}{(1/N_{\text{rep}}) \sum_{k=1}^{N_{\text{rep}}} \exp[-\chi_k^2/2\Delta\chi^2]}, \quad (15)$$

where the theoretical values  $y_i[f]$  are estimated by

$$y_i[f] \approx y_i[S_0] + \sum_k \left. \frac{\partial y_i[S]}{\partial z_k} \right|_{S=S_0} z_k, \quad (16)$$

with the deviation between data and each replica  $k$  computed as

$$\chi_k^2 = \sum_{i,j=1}^{N_{\text{data}}} (y_i[f_k] - y_i) C_{ij}^{-1} (y_j[f_k] - y_j). \quad (17)$$

An interesting feature of the Bayesian method is the existence of a quantitative estimator of the agreement between the data originally considered for the PDF fit and the new one. This estimator, the effective number of replicas  $N_{\text{eff}}$ , is defined as

$$N_{\text{eff}} \equiv \exp \left\{ \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \omega_k \log(N_{\text{rep}}/\omega_k) \right\}. \quad (18)$$

$N_{\text{eff}} \ll N_{\text{rep}}$  points to a large tension between data sets, either due to incompatibility or too much new information in the new data.  $N_{\text{eff}} \approx N_{\text{rep}}$  instead hints at a strong compatibility and the use of the new reweighted PDF set reliable. Nevertheless, the real meaning of  $N_{\text{eff}}$  can be less clear when comparing different PDFs sets, as was pointed in [37].

The most general version of the procedure, applicable to any process with at least one hadron (or nucleus) in the initial or final state, starts by generating the replicas  $f_k$  by

$$f_k \equiv f_{S_0} + \sum_i^{N_{\text{eig}}} \left( \frac{f_{S_i^+} - f_{S_i^-}}{2} \right) R_{ik}, \quad (19)$$

with  $f_{S_0}$  and  $f_{S_i^\pm}$  the PDFs for the central fit and eigenvectors, respectively, and  $R_{ik}$  random numbers with Gaussian distribution centered at zero and with variance one. Once these are calculated for a high enough amount of replicas ( $N_{\text{rep}} > 10^3$ ), they are used to compute the theoretical values of the observable needed for each replica. For most of the studied quantities it involves a large quantity of time consuming convolutional integrals and this goes in detriment of the reweighting positive feature of speed. However, should the PDF set under study enter the computation linearly (as in the present case), it is possible to alter the order of the steps and save time. Let us consider the electron–nucleus collision we are investigating. For the  $k$ -th replica, the observable can be schematically written as

$$\mathcal{O}_k = \hat{\mathcal{O}} \otimes f_k^A, \quad (20)$$

where  $\otimes$  denotes the convolution of the hard part  $\hat{\mathcal{O}}$  with the PDF, and the sum over the partonic species is implicit. Using Eq. (19) to replace  $f_k^A$ , we end up with



**Table 1**

Results of the reweighting process: pseudodata taken into account (first column), nPDF considered (second column),  $\chi^2$  per number of pseudodata points before and after the reweighting (third and fourth column, respectively), and effective number of replicas remaining (fifth column) of the  $N_{\text{rep}} = 10^5$  initial replicas.

Pseudodata	nPDF	$\chi^2/n _{\text{before}}$	$\chi^2/n _{\text{after}}$	$N_{\text{eff}}$
$F_2$	DSSZ	84.22	2.24	1877
	EPS09	26.51	1.38	15010
$F_L$	DSSZ	197.63	162.96	74335
	EPS09	42.03	39.05	98045
$F_2 + F_L$	DSSZ	109.06	38.15	1865
	EPS09	29.66	5.67	14910

$$\mathcal{O}_k = \hat{\mathcal{O}} \otimes \left[ f_{S_0}^A + \sum_i^{N_{\text{eig}}} \left( \frac{f_{S_i^+}^A - f_{S_i^-}^A}{2} \right) R_{ik} \right]. \quad (21)$$

By distributing the terms we have

$$\mathcal{O}_k = \mathcal{O}_{S_0} + \sum_i^{N_{\text{eig}}} \frac{R_{ik}}{2} [\mathcal{O}_{S_i^+} - \mathcal{O}_{S_i^-}], \quad (22)$$

where  $\mathcal{O}_{S_0}$  is the observable obtained with the central set, and  $\mathcal{O}_{S_i^\pm}$  the ones corresponding to the eigenvectors. Thus we avoid computing the observable  $N_{\text{rep}}$  times and only do so  $2N_{\text{eig}} + 1$  times.

#### 4. Results

Given the pseudodata of Figs. 1 and 2 for  $e + \text{Au}$  collisions ( $A = 197$ ), we perform a Bayesian reweighting using an initial number of replicas  $N_{\text{rep}} = 10^5$ . For the sake of making better sense of the results, we perform the analysis three times, considering only  $F_2$ , only  $F_L$  and both  $F_2$  and  $F_L$  pseudodata together. The results are shown on Table 1, which summarizes the quantitative estimators of the adequacy of the nPDF description.

If we look at the central column of Table 1, we see that the  $\chi^2$  is always larger for DSSZ than for EPS09. This is because, as mentioned before, the pseudodata in the upper left panel of Fig. 1 and 2 is only included in the analysis with the DSSZ set. This fact is specially relevant in the case of  $F_L$ , since the discrepancy between pseudodata and theoretical curve is larger than for  $F_2$ .

Regarding  $F_2$ , the deviation of the pseudodata from the central predictions combined with the small uncertainties give huge

contributions to the  $\chi^2$  (central column). As for  $F_L$ , we have to distinguish between nPDFs, in spite of the description not being good in either case. The error bars are much larger than for  $F_2$ , so this yields a total  $\chi^2$  per number of points lower for EPS09. Nevertheless, for DSSZ, the inclusion of the upper left panel of Fig. 2 means a towering increase of  $\chi^2$ , despite the larger error bars.

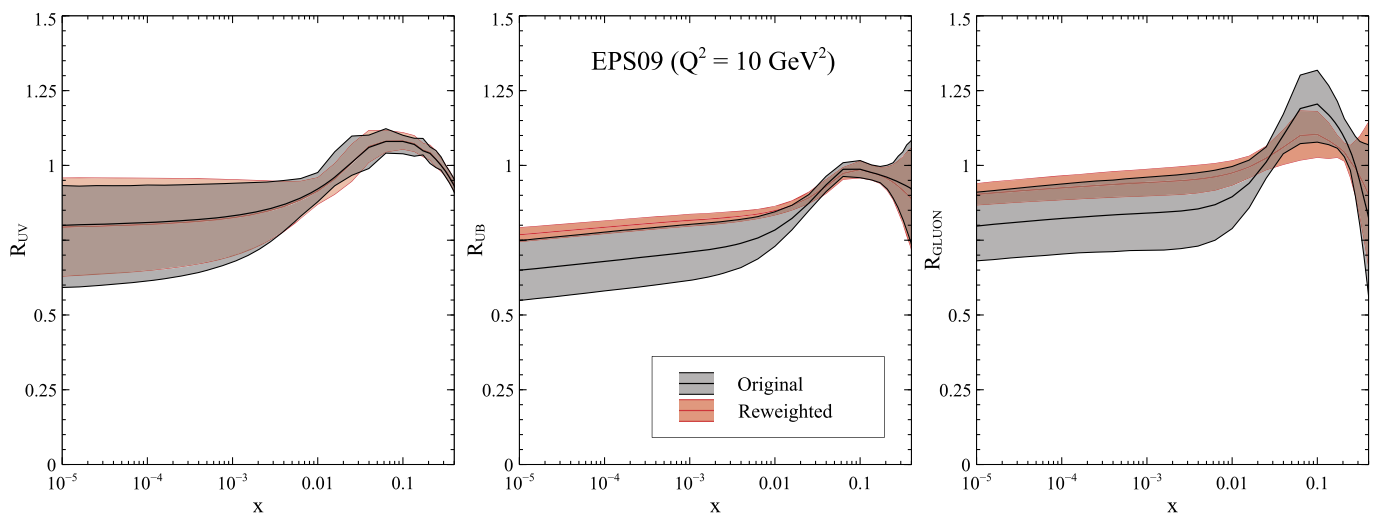
These  $\chi^2$  determine the weights according to equation (15) which in turn give us the total number of meaningful replicas ( $N_{\text{eff}}$ ) that we see in the far right column of Table 1. If we compare  $N_{\text{eff}}$  with the total number of replicas  $N_{\text{rep}}$ , we can see, on one hand, that for  $F_L$  most of the replicas survive (for EPS09, the practical totality of the replicas survive), noting that this pseudodata is compatible with the current nPDF fits. This does not come as a surprise, given the huge error bars of the  $F_L$  pseudodata. On the other hand, considering both  $F_2$  and  $F_L$ , we can see that less than 15% of the replicas survive the reweighting in the case of EPS09. This number shrinks to less than 2% if we consider the DSSZ set. This means that there is a big tension between the pseudodata and the theoretical predictions, and a new fit should be mandatory.

Nevertheless, we can study what happens with the few remaining replicas and what nuclear partonic behavior is favored by the pseudodata. The reweighting process affects the nPDFs in the following way:

- In the case of EPS09 (Fig. 4), the reweighting suppresses minimally the central value of the valence at low  $x$  (left panel), slightly increasing the shadowing in that region while the uncertainty is shifted up just a little bit. For the sea distribution (central panel) the central value increases and the shadowing suppression gets smaller, with the uncertainties shrinking dramatically. Finally the gluon density (right panel) flattens and the strong shadowing/anti-shadowing that characterizes this fit smooths so much that leaves a curve almost compatible with unity.

- In the case of DSSZ (Fig. 5), the reweighting returns a valence distribution with stronger shadowing and anti-shadowing regions, a sea distribution with an enhancement of the central value in the shadowing area and a deeper EMC-effect. The reweighted gluon distribution behaves the same way as the reweighted sea, but the shadowing region for the gluon goes over unity, becoming, in fact, anti-shadowing. All these distributions get narrower uncertainties after the reweighting.

The curious behavior at low  $x$  produced by the pseudodata originates from the ansatz in the initial parameterizations. In EPS09, the limit of  $R$  for  $x \rightarrow 0$  is a parameter, while for DSSZ there is no such constrain. Once obtained a fit, any prediction outside the



**Fig. 4.**  $u$  valence quark (left),  $\bar{u}$  sea quark (center) and gluon (right) distributions for the EPS09 PDF set before and after the reweighting.

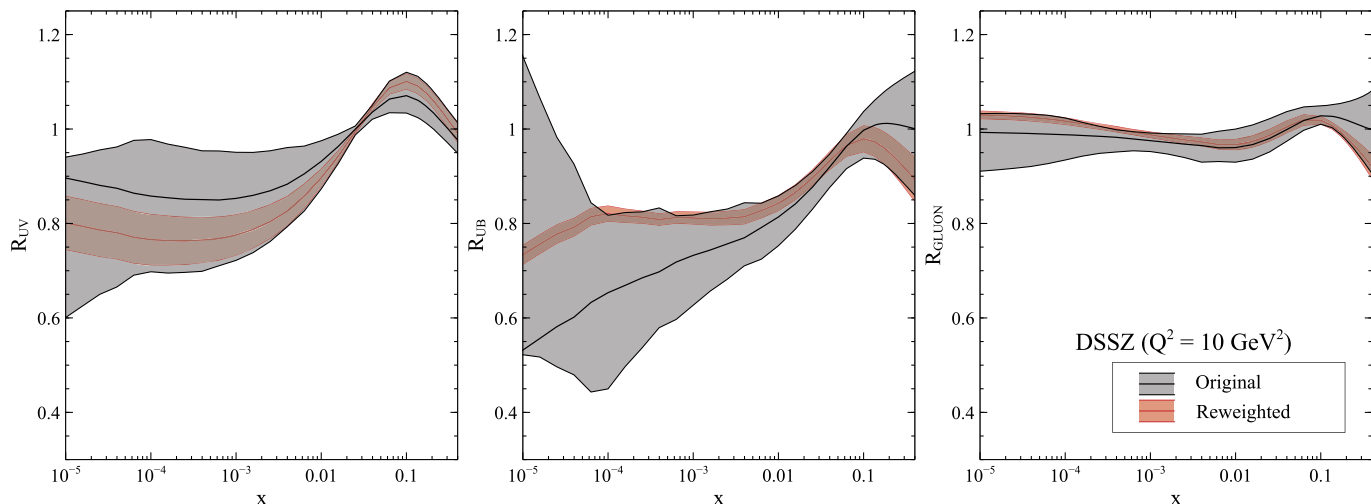


Fig. 5.  $u$  valence quark (left),  $\bar{u}$  sea quark (center) and gluon (right) distributions for the DSSZ PDF set before and after the reweighting.

kinematical range probed comes from an extrapolation, thus allowing for the puzzling and unrealistic curves at small  $x$  of Fig. 5 to occur.

## 5. Summary

On the search for gluon saturation we have analyzed the impact of future nuclear structure function low- $x$  data at the EIC. Our aim was to assess whether or not the collinear factorization approach with nuclear PDFs will be able to fit the data, if saturation sets in according to current expectations. As it is customary for nuclear PDFs, in this study we used the structure function  $F_2$ . In addition, we have also considered the longitudinal structure function  $F_L$ , as it is much more sensitive to the gluon density and thus potentially capable of providing interesting information, although with hindsight we found it not to be the case due to the large projected errors.

We generated *saturated* pseudodata for  $F_2^{\text{Au}}$  and  $F_L^{\text{Au}}$ , with central values obtained from the rcBK predictions and error bars estimated taking into account the design parameters the future detectors, and compared them with the corresponding collinear factorization predictions. The latter, that rely on extrapolations as most of the points lie outside the kinematical range explored in the original fits, lead to smaller structure functions than what the saturation model predicts at small  $x$ . This is particularly true for  $F_L$ , as the extrapolation for the nuclear gluon distributions at low  $x$  are very unreliable. The quantitative impact of the pseudodata on the nPDFs was obtained by means of a Bayesian reweighting technique. The results look quite different from the original distributions, especially for sea quarks and gluons. This strong tension is confirmed by the numbers in Table 1. The effect is bigger in the case of DSSZ for which the gluon parameterization lacks flexibility, but numerical estimators confirm the existence of a non trifling tension also in the case of EPS09.

Our results suggest that, should the EIC provide data compatible with the expected theoretical description from the saturation model studied (or a similar one), a successful refitting of the nPDFs may not be achievable, which would unambiguously signal the presence of non-linear effects. However, in order to be fully conclusive and determine whether or not genuine saturation effects can be unveiled from nuclear structure function measurements, performing a new global nPDF fit will be necessary. At the moment we can not exclude the possibility of a successful refitting of the nuclear PDFs because the nuclear gluon distribution is currently essentially unconstrained at small- $x$ . In that case, one would

have to resort to diffractive observables in order to pin down saturation effects at the EIC (see e.g. [38–40]).

## Acknowledgements

M.R.M. thanks the hospitality of the CPTH at École Polytechnique, where part of this work was performed. This research was supported by the European Research Council grant HotLHC ERC-2011-StG-279579; Ministerio de Ciencia e Innovación of Spain under project FPA2014-58293-C2-1-P; Xunta de Galicia (Consellería de Educación) – the group is part of the Strategic Unit AGRUP2015/11. P.Z. was supported in part by the Office of Nuclear Physics within the U.S. DOE Office of Science. M.R.M. was supported by Spanish INEM (SEPE).

## References

- [1] A. Accardi, et al., Eur. Phys. J. A 52 (9) (2016) 268, arXiv:1212.1701 [nucl-ex].
- [2] F. Gelis, E. Iancu, J. Jalilian-Marian, R. Venugopalan, Annu. Rev. Nucl. Part. Sci. 60 (2010) 463, arXiv:1002.0333 [hep-ph].
- [3] T. Lappi, Int. J. Mod. Phys. E 20 (2011) 1, arXiv:1003.1852 [hep-ph].
- [4] J.L. Albacete, C. Marquet, Prog. Part. Nucl. Phys. 76 (2014) 1, arXiv:1401.4866 [hep-ph].
- [5] A. Adare, et al., PHENIX Collaboration, Phys. Rev. Lett. 107 (2011) 172301, arXiv:1105.5112 [nucl-ex].
- [6] E. Braidot, STAR Collaboration, arXiv:1005.2378 [hep-ph].
- [7] C. Marquet, Nucl. Phys. A 796 (2007) 41, arXiv:0708.0231 [hep-ph].
- [8] D. Boer, et al., arXiv:1108.1713 [nucl-th].
- [9] J.L. Albacete, J.G. Milhano, P. Quiroga-Arias, J. Rojo, Eur. Phys. J. C 72 (2012) 2131, arXiv:1203.1043 [hep-ph].
- [10] K.J. Eskola, H. Honkanen, V.J. Kolhinen, J.w. Qiu, C.A. Salgado, Nucl. Phys. B 660 (2003) 211, arXiv:hep-ph/0211239.
- [11] E.R. Cazaroto, F. Carvalho, V.P. Goncalves, F.S. Navarra, Phys. Lett. B 671 (2009) 233, arXiv:0805.1255 [hep-ph].
- [12] A. Accardi, V. Guzey, J. Rojo, arXiv:1106.3839 [hep-ph].
- [13] I. Balitsky, Nucl. Phys. B 463 (1996) 99, arXiv:hep-ph/9509348.
- [14] Y.V. Kovchegov, Phys. Rev. D 60 (1999) 034008, arXiv:hep-ph/9901281.
- [15] E. Gardi, J. Kuokkanen, K. Rummukainen, H. Weigert, Nucl. Phys. A 784 (2007) 282, arXiv:hep-ph/0609087.
- [16] Y.V. Kovchegov, H. Weigert, Nucl. Phys. A 784 (2007) 188, arXiv:hep-ph/0609090.
- [17] I. Balitsky, Phys. Rev. D 75 (2007) 014001, arXiv:hep-ph/0609105.
- [18] I. Balitsky, G.A. Chirilli, Phys. Rev. D 77 (2008) 014019, arXiv:0710.4330 [hep-ph].
- [19] K.J. Eskola, H. Paukkunen, C.A. Salgado, J. High Energy Phys. 0904 (2009) 065, arXiv:0902.4154 [hep-ph].
- [20] D. de Florian, R. Sassot, P. Zurita, M. Stratmann, Phys. Rev. D 85 (2012) 074028, arXiv:1112.6324 [hep-ph].
- [21] W.T. Giele, S. Keller, Phys. Rev. D 58 (1998) 094023, arXiv:hep-ph/9803393.

- [22] R.D. Ball, et al., NNPDF Collaboration, Nucl. Phys. B 849 (2011) 112; R.D. Ball, et al., NNPDF Collaboration, Nucl. Phys. B 854 (2012) 926 (Erratum); R.D. Ball, et al., NNPDF Collaboration, Nucl. Phys. B 855 (2012) 927 (Erratum), arXiv:1012.0836 [hep-ph].
- [23] R.D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, N.P. Hartland, J.J. Latorre, et al., Nucl. Phys. B 855 (2012) 608, arXiv:1108.1758 [hep-ph].
- [24] G. Watt, R.S. Thorne, J. High Energy Phys. 1208 (2012) 052, arXiv:1205.4024 [hep-ph].
- [25] B.J.A. Watt, P. Motylinski, R.S. Thorne, arXiv:1311.5703 [hep-ph].
- [26] N. Sato, J.F. Owens, H. Prosper, arXiv:1310.1089 [hep-ph].
- [27] H. Paukkunen, P. Zurita, J. High Energy Phys. 1412 (2014) 100, arXiv:1402.6623 [hep-ph].
- [28] N. Armesto, Eur. Phys. J. C 26 (2002) 35, arXiv:hep-ph/0206017.
- [29] H. Kowalski, D. Teaney, Phys. Rev. D 68 (2003) 114005, arXiv:hep-ph/0304189.
- [30] H. De Vries, C.W. De Jager, C. De Vries, At. Data Nucl. Data Tables 36 (1987) 495.
- [31] J.L. Albacete, N. Armesto, J.G. Milhano, C.A. Salgado, Phys. Rev. D 80 (2009) 034031, arXiv:0902.1112 [hep-ph].
- [32] L.D. McLerran, R. Venugopalan, Phys. Lett. B 424 (1998) 15, arXiv:nucl-th/9705055.
- [33] J.L. Albacete, N. Armesto, J.G. Milhano, P. Quiroga-Arias, C.A. Salgado, Eur. Phys. J. C 71 (2011) 1705, arXiv:1012.4408 [hep-ph].
- [34] J.L. Albacete, Nucl. Phys. A 957 (2017) 71, arXiv:1507.07120 [hep-ph].
- [35] We are especially grateful to Matt Lamont for providing the projected error bars to us
- [36] A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C 63 (2009) 189, arXiv:0901.0002 [hep-ph].
- [37] N. Armesto, H. Paukkunen, J.M. Penin, C.A. Salgado, P. Zurita, Eur. Phys. J. C 76 (4) (2016) 218, arXiv:1512.01528 [hep-ph].
- [38] M.S. Kugeratski, V.P. Goncalves, F.S. Navarra, Eur. Phys. J. C 46 (2006) 413, arXiv:hep-ph/0511224.
- [39] H. Kowalski, T. Lappi, R. Venugopalan, Phys. Rev. Lett. 100 (2008) 022303, arXiv:0705.3047 [hep-ph].
- [40] H. Kowalski, T. Lappi, C. Marquet, R. Venugopalan, Phys. Rev. C 78 (2008) 045201, arXiv:0805.4071 [hep-ph].