Bayesian reweighting of nuclear PDFs and constraints from proton-lead collisions at the LHC

Nestor Armesto,[a,b] **Juan Rojo,**[b] **Carlos A. Salgado**[a] **and Pia Zurita**[a]

[a]*Departamento de Física de Partículas and IGFAE, Universidade de Santiago de Compostela,*
E-15706 Santiago de Compostela, Galicia, Spain

[b]*PH Department, TH Unit, CERN,*
CH-1211 Geneva 23, Switzerland

E-mail: nestor.armesto@usc.es, juan.rojo@cern.ch,
carlos.salgado@usc.es, pia.zurita@usc.es

ABSTRACT: New hard-scattering measurements from the LHC proton-lead run have the potential to provide important constraints on the nuclear parton distributions and thus contributing to a better understanding of the initial state in heavy ion collisions. In order to quantify these constraints, as well as to assess the compatibility with available nuclear data from fixed target experiments and from RHIC, the traditional strategy is to perform a global fit of nuclear PDFs. This procedure is however time consuming and technically challenging, and moreover can only be performed by the PDF fitters themselves. In the case of proton PDFs, an alternative approach has been suggested that uses Bayesian inference to propagate the effects of new data into the PDFs without the need of refitting. In this work, we apply this reweighting procedure to study the impact on nuclear PDFs of low-mass Drell-Yan and single-inclusive hadroproduction pseudo-data from proton-lead collisions at the LHC as representative examples. In the hadroproduction case, in addition we assess the possibility of discriminating between the DGLAP and CGC production frameworks. We find that LHC proton-lead data could lead to a substantial reduction of the uncertainties on nuclear PDFs, in particular for the small-x gluon PDFs where uncertainties could decrease by up to a factor two. The Monte Carlo replicas of EPS09 used in the analysis are released as a public code for general use. It can be directly used, in particular, by the experimental collaborations to check, in a straightforward manner, the degree of compatibility of the new data with the global nPDF analyses.

KEYWORDS: Heavy Ion Phenomenology, QCD Phenomenology

ARXIV EPRINT: 1309.5371

Contents

1 Introduction

The knowledge of the parton distribution functions (PDFs) of the proton has substantially improved in the last years [1–5], thanks to the increased coverage and variety of experimental data included, theoretical improvements in higher order computations, as well as from methodological developements. Despite recent developements [6–10], the determination of nuclear PDFs (nPDFs) has not reached the same level of accuracy, due to the scarce amount of available data and the limited kinematical coverage of the measurements done up to date — see e.g. the review [11]). In this respect, measurements from the recent LHC proton-lead run have the potential [12–18] to be of great relevance as they will not only extend the probed kinematical range but also provide information on heavier nuclei where data is currently very limited. In turn, more accurate nPDFs allow for a more reliable characterization of the initial state in heavy ion collisions.

In order to extract information out of a novel measurement, a new determination of PDFs must be done, by including the new data in a global fit with all preceding data. Though straightforward in principle, performing a new fit is a cumbersome and time-consuming process. In addition, exploring formerly unknown regions (such as small-x) might required substantial modifications on the theoretical input and the fitting parameterization strategy. That is, it is not clear just by looking at the new data whether it is compatible or not with previous results. This is particularly true in the case of experiments involving nuclei, as the nuclear medium might present a variety of more complex phenomena than a simple modification of the parton distributions.

As an alternative to repeating the nPDF global fit, it is possible to reweight an existing PDF set with the information contained in the new measurement using Bayesian inference, applying the techniques originally developed for proton PDFs [19, 20]. This method allows to study quantitatively both the compatibility of new data with that used in the original

PDFs determination and to determine its impact on the central values and uncertainties of the PDFs. While the original derivation [19, 20] applied only to PDF sets based on the Monte Carlo method, it was later shown [21] how the same method can be extended to Hessian PDFs, which is the framework adopted for most nuclear PDF sets available.

The aim of this work is thus to perform an exploratory quantitative study of the constraining potential for nuclear PDFs of the LHC proton-lead run data. We have selected two representative processes: low-mass Drell-Yan production and charged hadron inclusive production. Since no data from pPb collisions in the hard scattering regime is yet available (except for charged particle production in the pilot run from ALICE [22]), we will simulate pseudo-data based on a known underlying theory, namely the collinear DGLAP framework and the EPS09 nuclear PDF set. In the case of hadroproduction, in addition pseudo-data has been also simulated in the Color Glass Condensate (CGC) framework (see e.g. the review [23]): this allows to quantify to which extent non-linear effects in charged hadron production can be absorbed in a global nPDF fit based on the DGLAP framework, While our analysis is based on the EPS09 nuclear PDF set, the qualitative results should be valid for all other nPDF sets.

As an extra, we also release the set of Monte Carlo replicas of EPS09 as a public computer code.[1] These replicas can be used directly, in particular by the experimental collaborations, to check the compatibility of the new data with the nuclear parton distributions, as well as to pindown the corresponding constraints for each parton flavor. The procedure to do that is the same as applied here.

This article is organized as follows: first of all, in section 2 we present a brief summary of the main features of the reweighting procedure and construct a Monte Carlo version of the EPS09 set. Then we study the impact of Drell-Yan production in EPS09 in section 3, before moving to section 4 where we study charged hadron production. In this latter case, we explore both the constraints on nuclear PDFs and the discrimination power between DGLAP and CGC scenarios. Section 5 summarizes our results and discusses the prospects for other relevant measurements.

2 Bayesian reweighting of nuclear PDFs

PDF uncertainties can be determined using basically two methods: the Hessian approach (with and without tolerance), upon which all nuclear PDF sets are based, and the Monte Carlo approach. In a Monte Carlo PDF set, such as those of the NNPDF Collaboration [24, 25], the underlying PDF probability density $\mathcal{P}$ is sampled by generating, through a Monte Carlo procedure, an ensemble of $N_{\rm rep}$ PDFs replicas f_k, $k = 1, \ldots, N_{\rm rep}$, each fitted to a replica of the experimental data. Then any quantity $\mathcal{O}[f]$ depending on the PDFs can be evaluated by computing $\mathcal{O}[f_k]$ with $k = 1, \ldots, N_{\rm rep}$ and averaging over the results for individual replicas,

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{O}[f_k] \,. \tag{2.1}$$

---

[1]The code can be downloaded from http://igfae.usc.es/hotlhc/index.php/software.

Consider now a new measurement consisting of n points with covariance matrix cov_{ij}, not included in the orginal determination of $\mathcal{P}$,

$$y = \{y_1, y_2, \ldots, y_n\}. \tag{2.2}$$

Using Bayesian inference [19, 20], it is possible to update the original probability distribution $\mathcal{P}_{\mathrm{old}}(f)$ to a new probability distribution $\mathcal{P}_{\mathrm{new}}(f)$ that accounts for the information contained in the new measurement. This can be achieved by computing the new weight w_k for each individual replica f_k, which measures its agreement with the new data. It should be noted that, as shown explicitly in ref. [19], Bayesian reweighting is fully equivalent to a full new fit, provided the new data is not too constraining so that the effective number of replicas is still large enough (see below).

Following [19, 20], these weights turn out to be

$$w_k = \frac{(\chi_k^2)^{\frac{1}{2}(n-1)} e^{-\chi_k^2/2}}{\frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} (\chi_k^2)^{\frac{1}{2}(n-1)} e^{-\chi_k^2/2}} \, , \tag{2.3}$$

in terms of the χ^2 for each replica between the original theory predictions for the k-th replica and the new experimental measurement,

$$\chi_k^2(y, f_k) = \sum_{i,j=1}^{n} (y_i - y_i[f_k]) \mathrm{cov}_{ij}^{-1} (y_j - y_j[f_k]) \, , \tag{2.4}$$

so that now the analog of eq. (2.1) reads

$$\langle \mathcal{O} \rangle_{\mathrm{new}} = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} w_k \mathcal{O}[f_k] \, . \tag{2.5}$$

The only feature that distinguishes between a full new fit from the Bayesian reweighting is the statistical efficiency of the latter. While a new fit would give the best representation of the underlying density for a given N_{rep}, this is not the case for the PDF reweighting. The replicas with very small weights will become almost irrelevant when computing averages and the accuracy of the representation of the underlying distribution $\mathcal{P}_{\mathrm{new}}(f)$ will decrease. To quantify this efficiency loss, the Shannon entropy can be used to compute N_{eff}, the effective number of replicas after reweighting:

$$N_{\mathrm{eff}} \equiv \exp \left\{ \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} w_k \log(N_{\mathrm{rep}}/w_k) \right\} . \tag{2.6}$$

The above equation determines that the accuracy of the reweighted fit is the same that would be obtained if a new fit with N_{eff} replicas were to be performed. When $N_{\mathrm{eff}} \ll N_{\mathrm{rep}}$ the reweighting method becomes unreliable, and a full refit is mandatory. This scenario might arise either because the new data is inconsistent with the original one within the given theoretical framework, or when the new data contains substantial new information on the PDFs as compared to that in the original determination.

Two consistency tests can be performed in order to clarify this issue. The first one is the examination of the χ^2 profile of the new data,

$$\mathcal{P}(\chi^2) = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} w_k \,, \qquad (2.7)$$

with k all replicas such that $\chi_k^2 \in [\chi^2, \chi^2 + d\chi^2]$. Alternatively, it is possible to estimate if the agreement between data and theory improves if we rescale the experimental uncertainties by a factor α, then the probability density for the rescaling parameter will be

$$\mathcal{P}(\alpha) \propto \frac{1}{\alpha} \sum_{k=1}^{N_{\mathrm{rep}}} w_k(\alpha) \,, \qquad (2.8)$$

where $w_k(\alpha)$ are the weights of eq. (2.3) evaluated by replacing χ_k^2 by χ_k^2/α^2 (proportional to the probability of f_k given a data set with rescaled errors). When the $\mathcal{P}(\alpha)$ is peaked around one, the new data are consistent with the initial theory distribution. If $\mathcal{P}(\alpha)$ is peaked at a value larger than one, this might suggest either that experimental uncertainties have been underestimated or that the theory framework that is used is not the right one to describe this measurement.

While the derivation above that leads to the weights for each replica eq. (2.3) applies to PDF sets based on the Monte Carlo method, the goal of this paper is to study the impact on nuclear PDFs, for which all sets with uncertainty bands available are based on the Hessian method. As discussed in ref. [21], it is possible to generate Monte Carlo replicas starting from Hessian PDF error sets, and then apply the standard formulae such as eq. (2.5). We have thus constructed a Monte Carlo version of the EPS09 [6] nuclear PDF set, as follows. For each parton flavor we took the central PDF set (f_0) and shifted it using the Hessian error sets according to

$$f_k(x, Q^2) = f_0(x, Q^2) + \sum_{i}^{N_{\mathrm{eig}}} \left(f_i^{\pm}(x, Q^2) - f_0(x, Q^2) \right) |r_{k,i}| \,, \qquad (k = 1, \ldots, N_{\mathrm{rep}}) \,, \qquad (2.9)$$

where N_{eig} is the number of pairs of Hessian eigenvectors, $N_{\mathrm{eig}} = 15$ in the particular case of EPS09, and $r_{k,i}$ are random numbers from a Gaussian distribution of mean zero and variance one, and $f_i^{\pm}$ is the nPDF corresponding to the eigenvector $S_i^{\pm}$, and we use the positive sign for $r_{k,i} > 0$ and the negative sign for $r_{k,i} < 0$. Note that a symmetric version of eq. (2.9) can be obtained by averaging over each pair of eigenvectors [21].

Following eq. (2.9), we have generated $N_{\mathrm{rep}} = 1000$ Monte Carlo replicas for EPS09. We have checked the consistency of the procedure for all nuclear PDFs, by comparing central values and 1-sigma uncertainties in the Hessian and MC versions of EPS09, and finding reasonable agreement within the statistical accuracy expected. As an illustration, in figure 1 we show the nuclear ratios for lead ($A = 208$), defined as the ratio of nuclear PDFs for nuclear number A divided by the corresponding proton PDFs,

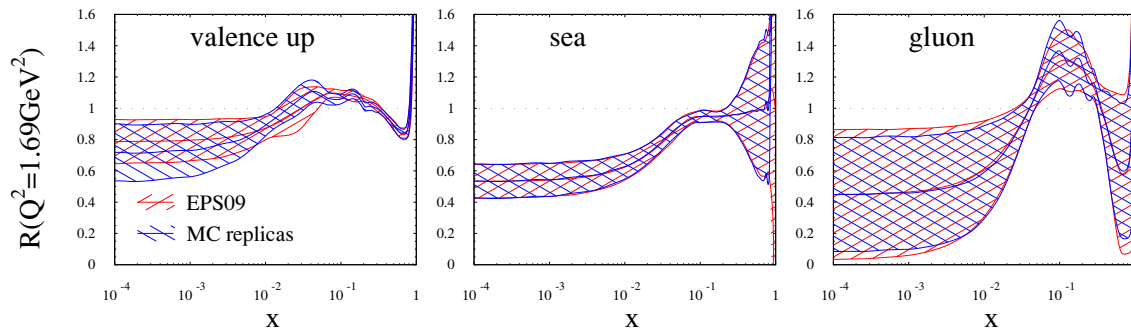$$R_i^A(x, Q^2) = \frac{f_i^A(x, Q^2)}{f_i^p(x, Q^2)} \qquad (2.10)$$

Figure 1. Comparison of the nuclear ratios $R_i^A(x, Q^2)$ for lead ($A = 208$) for three flavours at $Q^2 = 1.69\,\mathrm{GeV}^2$ in EPS09, comparing the original Hessian PDFs with their Monte Carlo counterparts. From left to right: up valence, quark sea, and gluon nuclear ratios.

at $Q^2 = 1.69\,\mathrm{GeV}^2$ for the up valence (left), quark sea (middle) and gluon (right) nuclear PDFs.[2] As we can see, both mean values and 1-sigma uncertainties are in reasonable agreement between the Hessian and Monte Carlo versions of EPS09. The remaining differences can be explained since we are neglecting (small) corrections beyond the linear approximation which is always assumed in Hessian PDF sets [21].

Let us also mention that ref. [26] proposes an alternative approach for including new data into a nuclear PDF fit similar to the reweighting method just discussed.

3 Constraints from Drell-Yan production

After describing our framework, we start by studying the impact on nuclear PDFs of Drell-Yan production in proton-lead collisions at the LHC. In proton-proton collisions, neutral current Drell-Yan production provides important constraints on the proton PDFs, in particular for quarks and antiquarks, but also for gluons, and has been measured at the LHC by ATLAS, CMS and LHCb [27–29]. On the other hand, the amount of Drell-Yan data so far included in nuclear PDFs fits is fairly small and affected by substantial experimental uncertainties, so the constraints derived from it are not as tight as the ones on nuclear valence quark distributions from deep-inelastic scattering (DIS) data.

In this respect, Drell-Yan data from proton-lead collisions at the LHC should provide crucial information to further constrain the gluon density at small and medium x as well as the sea quark distributions. We are in particular interested in low-mass Drell-Yan production, since on the one hand the small scale implies enhanced sensitivity to nuclear effects and in addition the (off-peak) cross-section rises when the invariant mass of the Drell-Yan pair m_{ll} decreases. On the other hand, the small m_{ll} regime is also interesting since some studies [30] predict that non-linear effects might show up as a substantial difference as compared to the linear DGLAP framework in which nuclear PDF fits are based.

---

[2]Note that the wiggles that can be observed in the Monte Carlo version of EPS09 for the gluon and the up valence quark near $x \sim 0.1$ in figure 1 arise from the interpolation in the EPS09 driver routines. These wiggles disappear if the symmetrised version of eq. (2.9) is used. They have a negligible impact on our results, since the LHC pPb data that we consider affects smaller values of x.
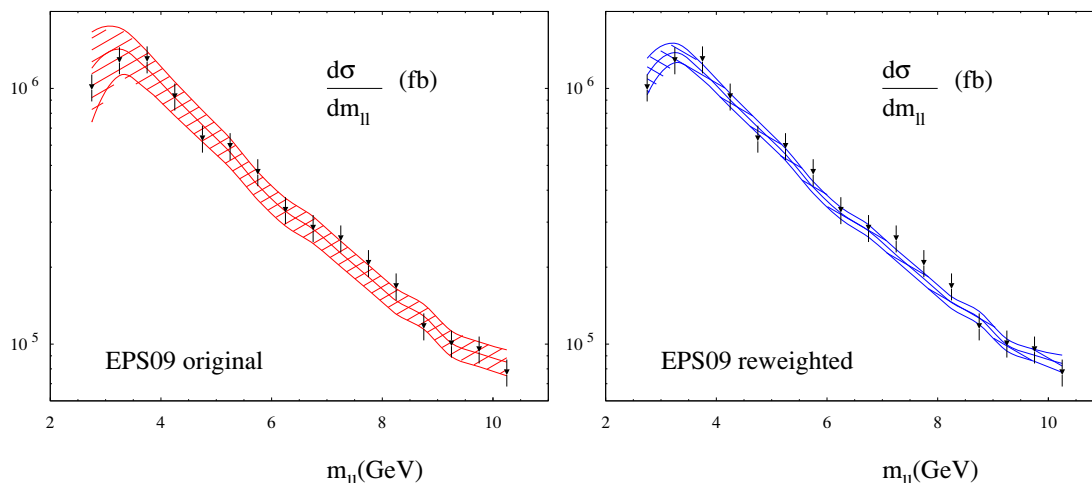
Figure 2. The predictions for the Drell-Yan differential cross sections with `EPS09` compared to pseudo-data as a function of the invariant lepton pair mass m_{ll} at central rapidity before (left) and after (right) PDF reweighting. The error band corresponds the 1-sigma uncertainty in `EPS09`.

Numerical simulations for DY production in pPb collisions applying the same kinematics cuts that were used in the pp case show that, given the integrated luminosity of the present run, around $\mathcal{L}_{\text{int}} \sim 30\,\text{nb}^{-1}$, very few events would be obtained in the forward region, as well as in the central region if the same cuts on lepton $p_{PT,l}$ were to be imposed. Therefore, for the purposes of this exercise we choose to consider only Drell-Yan cross-sections for $|\eta| < 4$, without any kinematical cut on $p_{T,l}$. Of course, the actual cuts will be different in real analysis, but for instance the cut in $p_{T,l}$ can be reduced as compared to pp since there are no problems with triggering most events.

To generate both the pseudo-data and the predictions for DY production based on the $N_{\text{rep}} = 1000$ replicas of the MC version of `EPS09`, we calculate the cross-sections in the low-mass range $m_{ll} < 12\,\text{GeV}$ by use of the `MCFM` code [31], modified in order to account for the fact that one of the initial particles is a lead nucleus. The pseudo-data was computed from the central values of `EPS09` and adding the corresponding statistical fluctuations. The statistical uncertainties were computed from the number of expected events in each bin, and a total uncorrelated 8% systematic uncertainty was also assumed. This value is a conservative estimate, based on the result that in the proton-proton case the systematic uncertainty is about 4% for the $15\,\text{GeV} < m_{ll} < 20\,\text{GeV}$ invariant mass bin [27]. For the proton PDFs we used MSTW08 [2], though results were essentially unchanged if some other proton PDF set was used. In the following, PDF uncertainties arise only from the `EPS09` nuclear PDFs, neglecting the proton PDFs.

To begin with, in figure 2 we show the pseudodata and the predictions before (left) and after (right) reweighting. The error band in the prediction accounts only for the uncertainties of the `EPS09` nuclear PDFs. It is clear that once the new data is included into the nuclear PDFs, the uncertainty band of the theory prediction is reduced, without affecting the central value, as expected for perfectly consistent data. In table 1 we provide

	χ^2/n	$\langle\chi^2\rangle/n$	N_{eff}
Original	0.64	2.68	–
Reweighted	0.59	0.96	539

Table 1. χ^2/n and $\langle\chi^2\rangle/n$ values before and after the reweighting of EPS with the Drell-Yan pPb pseudo-data, with $n = 16$ points. The effective number of replicas N_{eff} is also provided for the reweighting case.

the effective number of replicas N_{eff}, together with the values for the mean χ^2 and the average over replicas $\langle\chi^2\rangle$ per data point before and after the reweighting.

From table 1 we see that the χ^2/n is $\mathcal{O}(1)$, as expected by the use of consistent pseudo-data, while $\langle\chi^2\rangle/n \sim 2.7$, indicating that some replicas are clearly disfavoured by the pseudo-data. After the reweighting the weight for replicas far from the pseudo-data is suppressed, leadind to a substantial reduction in $\langle\chi^2\rangle/n$: the reweighted sample includes only replicas that agree with the pseudo-data. The number of effective replicas for this pseudo-data set is $N_{\text{eff}} = 539$, while the original sample had N_{eff}=1000, showing that is about half of the replicas are strongly disfavoured after the inclusion of the Drell-Yan pseudo-data.

Then we apply the consistency tests as described in section 2 and present the χ^2, w_k and $\mathcal{P}(\alpha)$ density distributions in figure 3. The χ^2 density distribution before the reweighting (upper left) is peaked around 0.8 but with a tail towards higher χ^2 values. Then, as we already knew, the pseudo-data is compatible with the one used in EPS09 fit and its inclusion in a refit should have a moderate impact. We confirm this by looking the χ^2 after the reweighting (lower left): the peak shifts towards 1 and the tail is significantly reduced. One last check in that regard is the $\mathcal{P}(\alpha)$ distribution in the lower right plot. The most probable value for the error rescaling parameter α is almost 1, so our error estimation was quite good.

We can now assess the impact of the DY pseudo-data on the nuclear PDFs. In this case while for the up and down valence distributions there seems to be no significant change (plot not shown), the sea distributions present a reduction on the respective uncertainty bands in the $x < 10^{-2}$ region accompanied of a very slight but negligible decrease of the central value. This is illustrated in the upper plots of figure 4, where we show the total sea quark nuclear ratio. The same behaviour is found for all sea distributions. The one parton that varies distinctly is however the gluon, as can be seen in the lower plots of figure 4. When x is below 10^{-2} there is a displacement of the central curve towards lower values accompanied by a reduction of the uncertainty of about 50%: the pseudo-data seems to favor more suppressed gluons in that region.

To summarize, the analysis of Drell-Yan pseudo-data from pPb collisions shows that (assuming full compatibility with the collinear framework) this measurement has a strong potential and would provide useful information to improve our knowledge of the gluon density in lead nuclei, which is of course a crucial ingredient to characterize the initial state in nuclear collisions at the LHC.
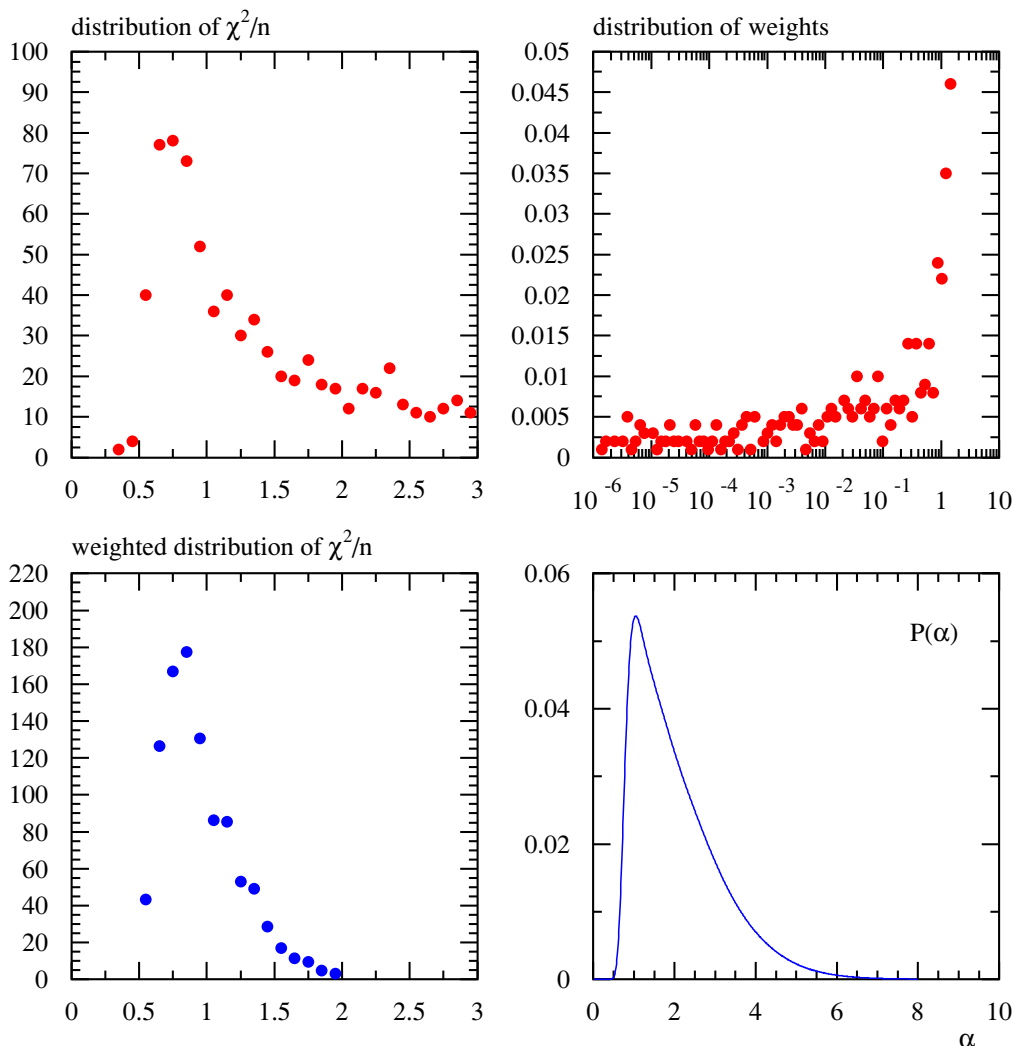
Figure 3. Left plots: distribution of the χ_k^2/n (χ^2 per data point) before (upper) and after (lower) the reweighting of EPS09 with pseudo-data for Drell-Yan production in pPb collisions. Upper right plot: distribution of the weights w_k. Lower right plot: $\mathcal{P}(\alpha)$.

4 Constraints from inclusive hadron production

Now we turn to the study of potential constraints on nPDFs arising from charged hadron single inclusive production, defined as the sum over all final state charged mesons and baryons. In this case our observable will be the cross section in proton-lead collisions divided by the same quantity in proton-proton collisions. This is the way in which experimental collaborations typically present their results, since the ratio cancels several experimental uncertainties and on top removes part of the proton PDF dependence.

We will consider two scenarios: one in which our pseudo-data is generated from CGC predictions, which include non-linear effects, and another using the collinear DGLAP framework and EPS09, which assume linear QCD evolution. The analysis based on CGC pseudo-
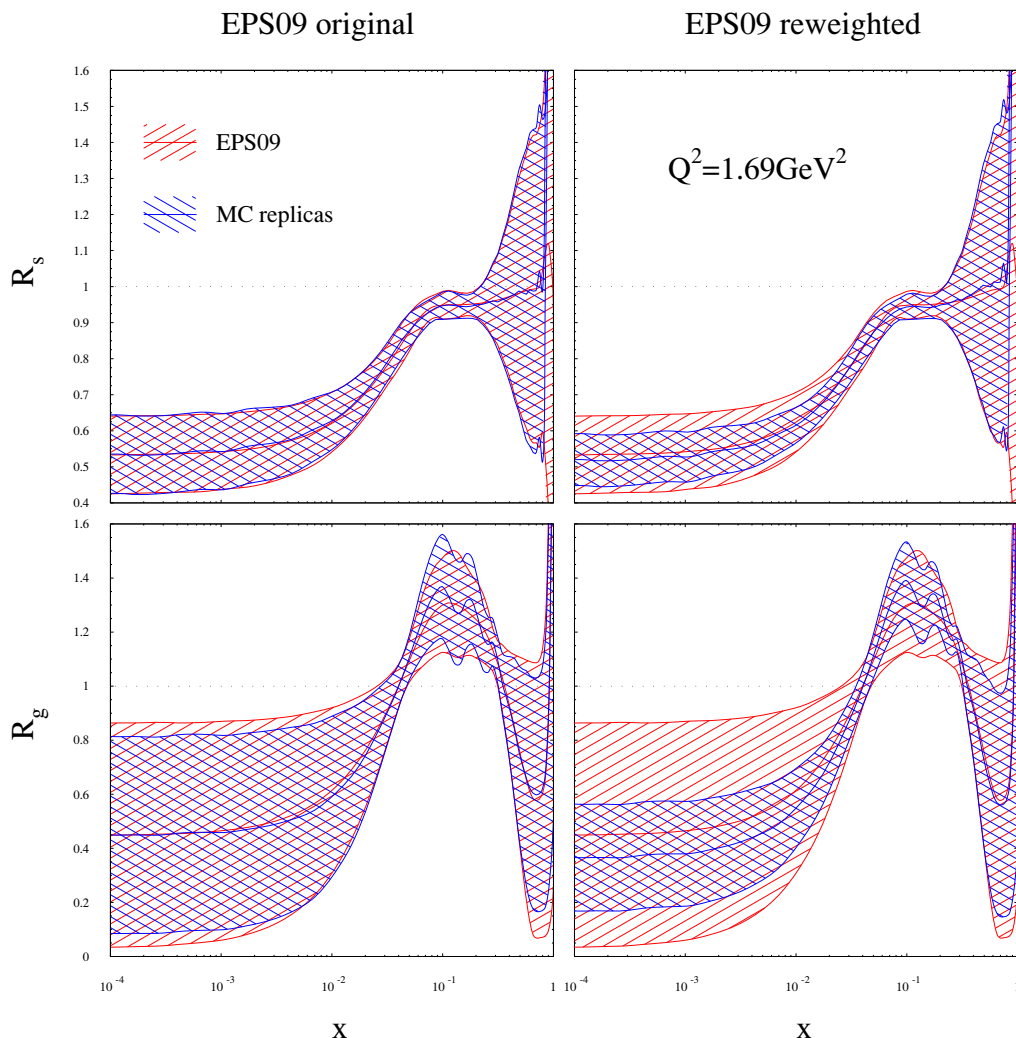
Figure 4. EPS09 nuclear ratios $R_i^A(x, Q^2)$ for $Q^2 = 1.69\,\text{GeV}^2$, both in the Hessian and in the Monte Carlo versions, before (left plots) and after (right plots) the reweighting with LHC pPb Drell-Yan production pseudo-data. We show both the nuclear modifications of sea quarks (upper plots) and of gluons (lower plots).

data is relevant to determine the discriminating power of this measurement regarding saturation dynamics. It should be taken into account that at least part of possible non-linear effects, if present in data, can be absorbed in the DGLAP fit, as discussed in refs. [32, 33] regarding this same problem in electron-proton collisions. On the other hand, when the pseudo-data is generated using EPS09, we can determine the improvement in the accuracy of nuclear PDFs in the case of a consistent underlying theory.

Let us mention that inclusive charged particle production is closely related to inclusive pion production, another process that has been used in the past to constrain nPDFs using dAu collisions from RHIC, although available data from RHIC is scarce and affected by large uncertainties. While it would be more realistic (from the experimental point of view) to include in our analysis pseudo-data for identified pions and kaons rather than

	$\eta = 0$			$\eta = 2$		
	χ^2/n	$\langle\chi^2\rangle/n$	N_{eff}	χ^2/n	$\langle\chi^2\rangle/n$	N_{eff}
Before	1.11	1.75		0.95	1.82	
After	0.84	1.02	624	0.92	1.08	612

Table 2. Same as table 1 for the reweighting of `EPS09` with inclusive charged hadron production data, for central ($\eta = 0$) and forward ($\eta = 2$) rapidities. Pseudo-data has been generated in the DGLAP framework.

sum inclusively over all charged particles, we have done so in order to compare with the available CGC predictions.

4.1   Hadroproduction in the DGLAP framework

Let us begin with the analysis of inclusive charged hadron production using pseudo-data generated with `EPS09` central values in the collinear DGLAP framework, as done in the previous section for the DY process. The statistical uncertainties are determined from the expected number of events in each data bin, and we have assumed 5% and 7% (uncorrelated) systematic and normalization uncertainties respectively, slightly larger than the corresponding proton-proton results [34–36]. Theoretical predictions have been computed using the code for NLO inclusive hadron production of ref. [37], modified to account for nuclear effects as discussed in ref. [38]. No nuclear effects were taken into account for the fragmentation in the nuclear medium, and the DSS [39] fragmentation functions (FFs) were employed.

In figure 5 we present the comparison between the the cross sections for the pseudo-data and for the `EPS09` predictions, for two different hadron rapidities, central ($\eta = 0$) and forward ($\eta = 2$). After including the pseudo-data into the nuclear fit by reweighting, we find a shift of the central values, suppression for $p_T < 8\,\text{GeV}$ and enhancement above, together with a clear narrowing of the uncertainties. Then in table 2 we show the values of χ^2/n, $\langle\chi^2\rangle/n$ and N_{eff}. For both rapidities almost two thirds of the replicas survive the reweighting. The improvement in $\langle\chi^2\rangle/n$ after including the pseudo-data indicates, as for DY, that theoretical predictions inconsistent with the pseudo-data have been effectively removed.

We then consider the consistency tests and in figure 6 show the χ^2, w_k and $\mathcal{P}(\alpha)$ distributions in the central region ($\eta = 0$). Similar information is derived from the analysis of the forward region. Before the reweighting the χ^2-density distribution (upper left) is peaked around 1.25, with a tail towards higher χ^2; afterwards (lower left) the peak moves towards one and the tail drastically reduces.

To conclude we turn our attention to the modification of the nuclear parton densities induced by the constraints from inclusive charged hadron production pseudo-data in `EPS09`. For the sake of clarity, we present in figure 7 the plots for both $\eta = 0$ (central) and $\eta = 2$ (right), with the upper plots corresponding to the sea quark nuclear ratios and the lower ones to the nuclear gluon ratio. The valence quarks turn out to be almost unaffected (since they are already quite constrained by nuclear DIS data). The same effects occur for both
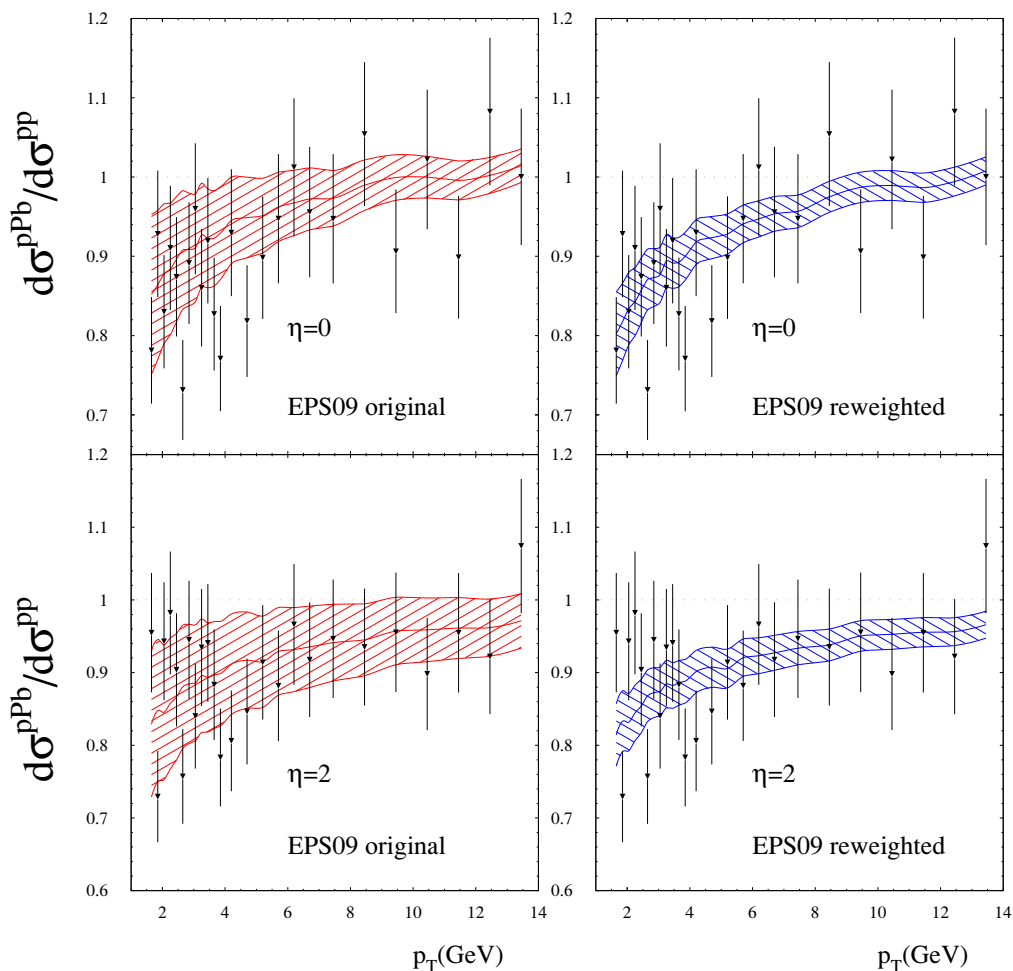
Figure 5. Upper plots: ratio of the charged hadron single inclusive cross section in *p*Pb with respect to *pp* collisions, as a function of the hadron transverse momentum p_T at central rapidity before (left) and after (right) the reweighting, where the pseudo-data has been generated in the collinear DGLAP framework using EPS09. Lower plots: the same for forward rapidities, $\eta = 2$.

rapidities: the sea density slightly decreases and the error band reduces a bit for $x < 10^{-2}$. In the case of the gluon, the central values show an enhancement for $0.07 < x < 0.2$ while for $x < 10^{-2}$ it is suppressed and the error band shrinks around 50%. While it is true that the displacement of the central curves is less pronounced in the forward region, they are nonetheless fully compatible within uncertainties, and compatible also with the variation from figure 4. The substantial error reduction in the small-x nuclear gluon ratio confirm that this observable is potentially very important to be included in nPDF fits once the LHC data becomes available.

4.2 Hadroproduction in the CGC framework

Now we consider the scenario in which pseudo-data has been generated in the Color Glass Condensate scenario (se e.g. the review [23]), following the approach of ref. [40]. Therefore,
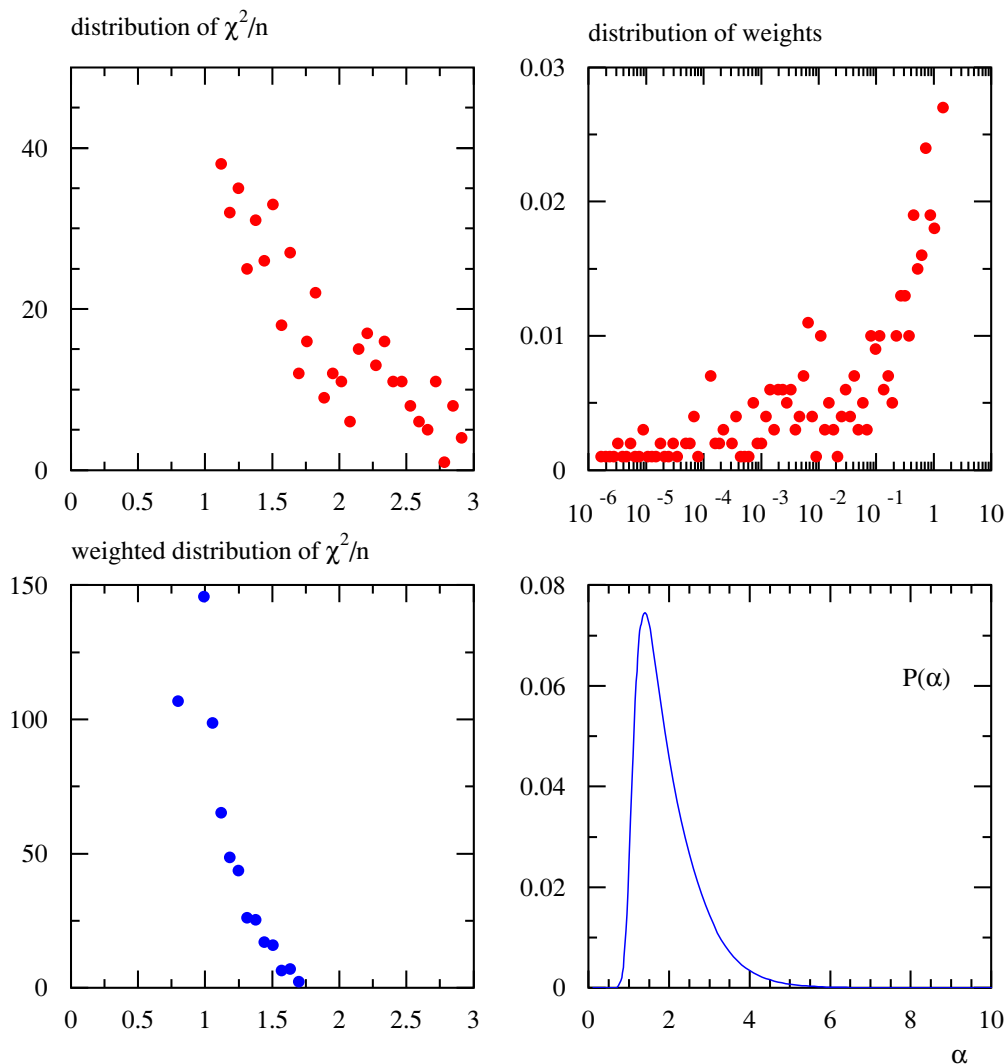
Figure 6. Same as figure 3 for inclusive charged hadron central production $\eta = 0$, with pseudo-data generated in the DGLAP framework with EPS09.

as opposed to the previous case, the underlying theory for pseudo-data is independent from the one used to compute the EPS09 predictions, which is always the perturbative DGLAP framework. We shall consider again the central and forward rapidity regions, with $\eta = 2$, noting that CGC and DGLAP predictions are known to differ more in the latter case. While identified meson production data at central rapidity from RHIC [41–44] has already been included in nuclear PDF fits [6, 8], data from the forward region has not been included as it introduced a rather strong tension with DIS measurements [45], which might arise from non-linear effects. It is then interesting to include both regions in our study.

To begin with, we present in figure 8 the nuclear cross section ratios corresponding to $\eta = 0$ (upper plots) and $\eta = 2$ (lower plots) before and after the reweighting, as a function of the transverse momentum p_T of the final hadron. In table 3 we present the values for χ^2/n and $\langle \chi^2 \rangle/n$ (with $n = 25$) for both rapidities and the corresponding N_{eff}.
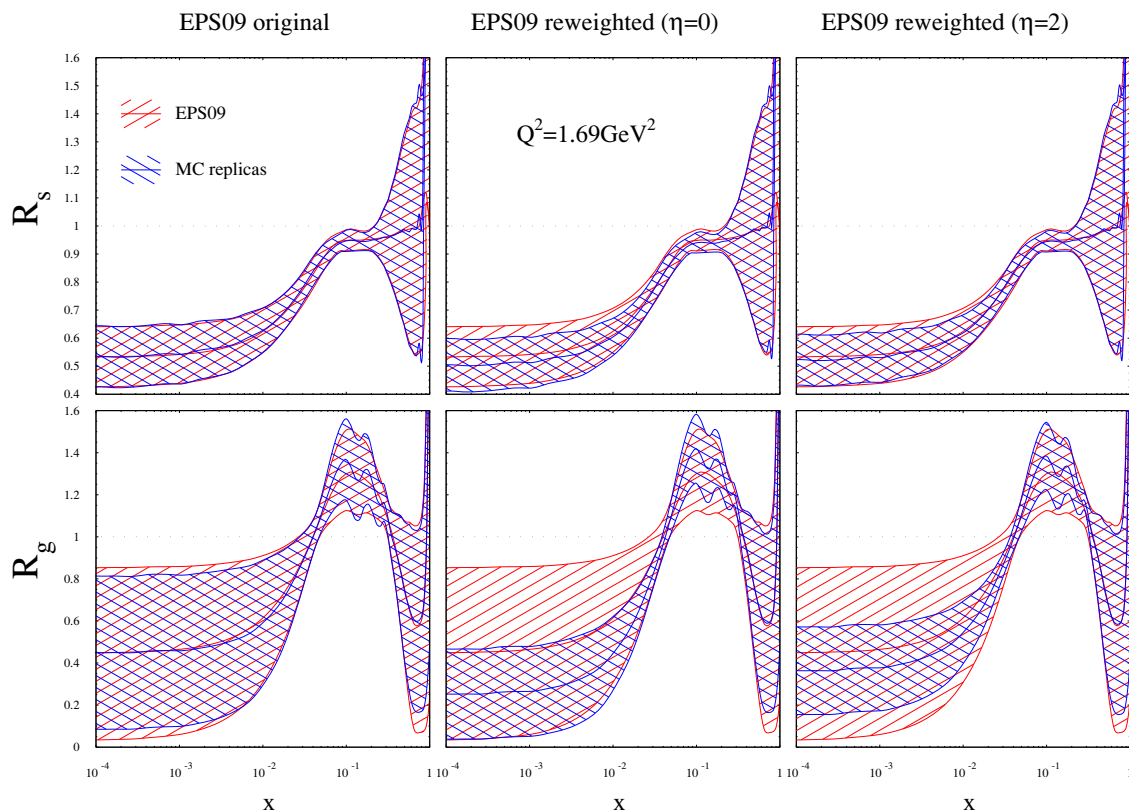
Figure 7. Same as figure 4, now for the nuclear modifications of sea quarks (upper plots) and gluons (lower plots) when pPb LHC pseudo-data for inclusive charged hadron production at central ($\eta = 0$) and forward ($\eta = 2$) rapidities is included in EPS09. The pseudo-data has been generated in the collinear DGLAP framework.

	$\eta = 0$			$\eta = 2$		
	χ^2/n	$\langle\chi^2\rangle/n$	$N_{\rm eff}$	χ^2/n	$\langle\chi^2\rangle/n$	$N_{\rm eff}$
Before	2.25	2.76		36.43	38.62	
After	1.50	1.58	229	1.85	1.85	1

Table 3. Same as table 2 for the case in which pseudo-data has been generated in the CGC framework.

Let us begin by discussing the impact of CGC pseudo-data from the central rapidity region, $\eta = 0$. Despite the fact that pseudo-data and theory predictions are generated with different underlying theories, the agreement after reweighting is still reasonable (though clearly not as good as in the case of DGLAP pseudo-data). The main impact of adding the pseudo-data is to drag the central value upwards, as well as reducing the PDF uncertainties. The effective number of replicas $N_{\rm eff}$t that we obtain is 229, and the final χ^2/n is 1.5, indicating some tension between the CGC prediction and the DGLAP theory predictions that cannot be accomodated by a change in shape or normalization of the nuclear PDFs.
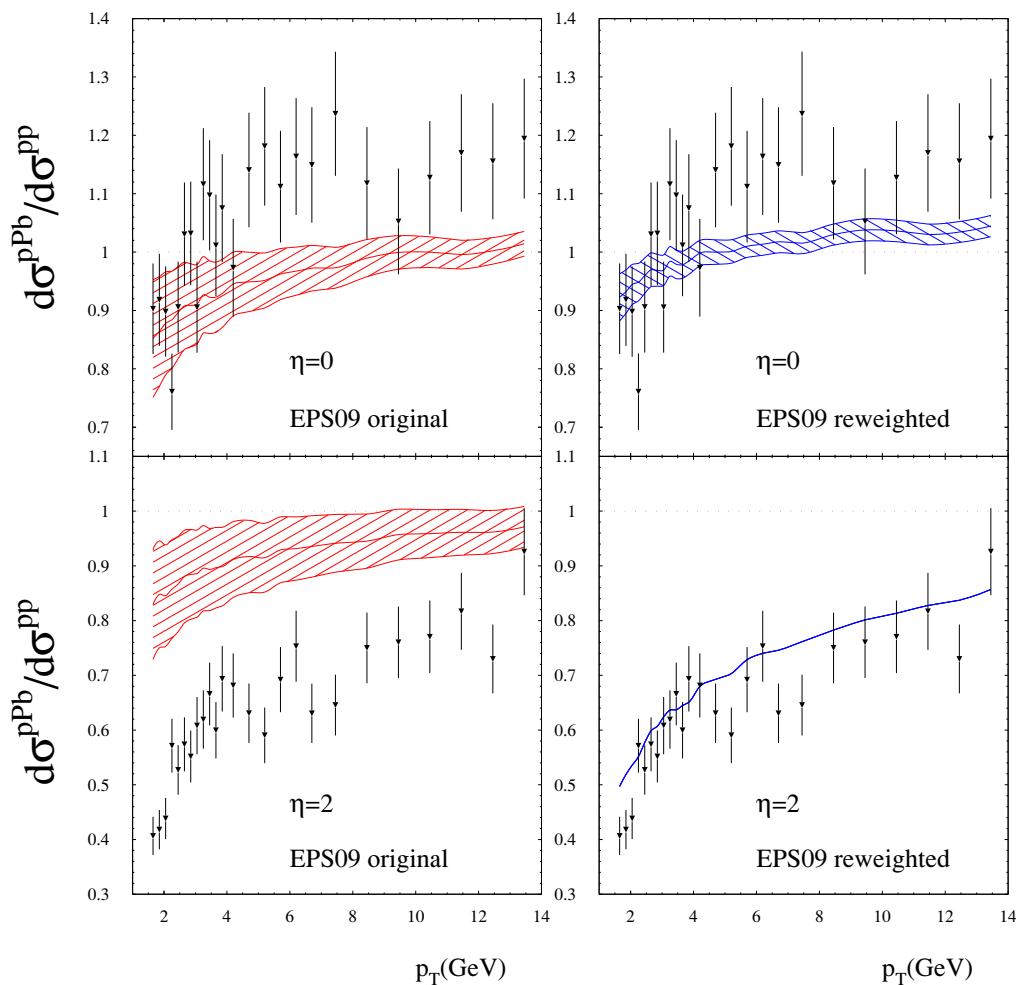
Figure 8. Upper plots: ratio of the charged hadron single inclusive cross section in pPb with respect to pp *at* central rapidity, as a function of the final hadron transverse momentum p_T before (left) and after (right) the reweighting. Lower plots: the same for $\eta = 2$. Note that the pseudo-data has been generated from the CGC predictions, while `EPS09` predictions are based on the collinear DGLAP framework.

If we move to the consistency tests, shown in figure 9, we find that the distribution of the χ^2/n is peaked around 2 with a flat tail towards higher values, while after the reweighting the peak moves as expected towards lower values but remaining slightly above 1. From the α distribution (lower right) we clearly see the inconsistency between data and theory predictions. It implies that a satisfactory agreement would also be obtained at the expenses of increasing experimental uncertainties by a factor two. Conversely, the more precise the experimental data is, the more sensitive to the differences between DGLAP and CGC predictions it will be, even in the central region where these differences are moderate.

As for the impact of the $\eta = 0$ charged hadron production CGC pseudo-data in the nuclear PDF ratios, the valence distributions are affected, presenting the non-negligible
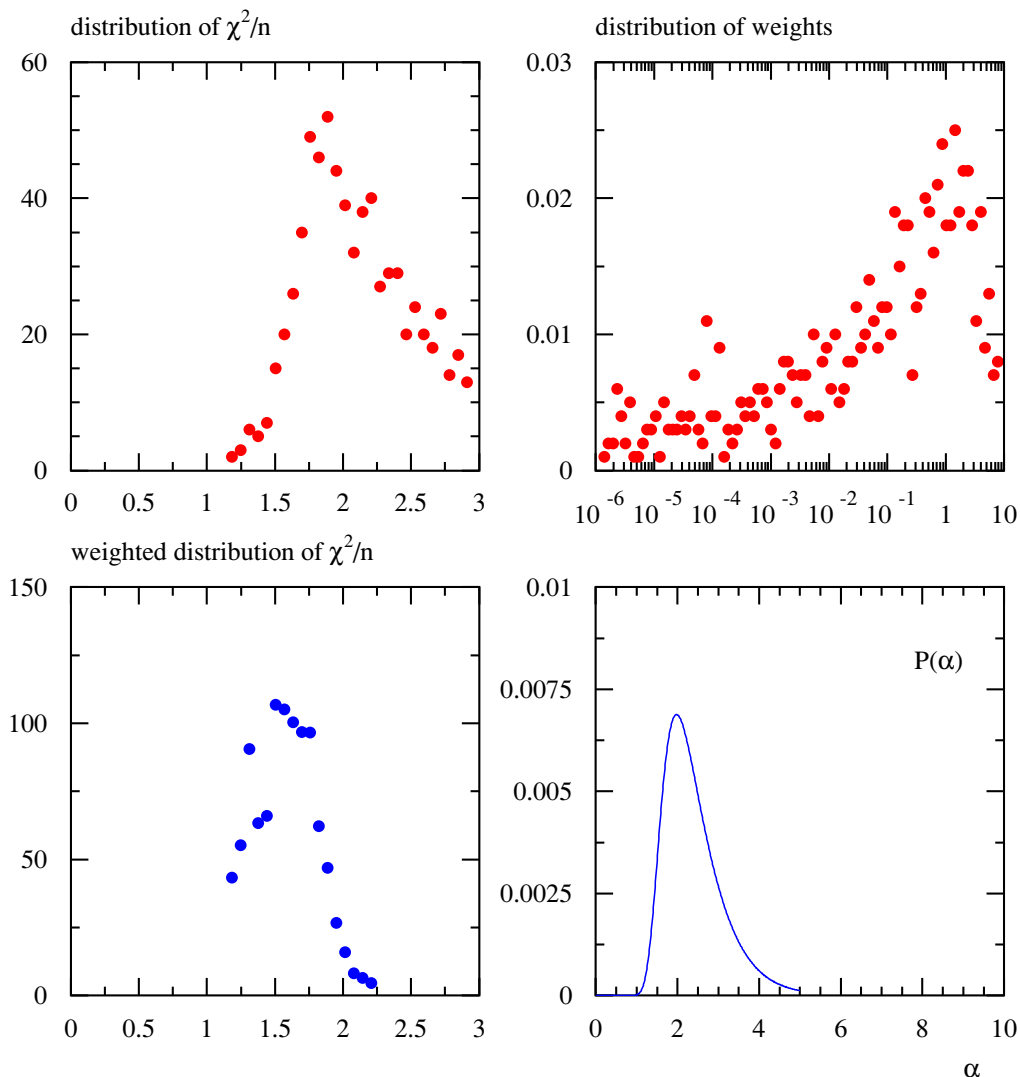
Figure 9. Same as figure 3 for inclusive charged hadron central production $\eta = 0$, with pseudo-data generated in the CGC framework.

decrease of the central value shown in the upper right plot of figure 10 for the up quark; no noticeable change is seen for the sea densities. For the gluons instead (lower plots of figure 10) the central value moves upwards and the uncertainty shrinks around 30% for $x < 10^{-2}$. Note that while the shriking of the gluon nPDF uncertainties is qualitatively similar as that seen in figure 7, where pseudo-data was generated with DGLAP, the trend of the central value is the opposite: in the CGC case, we get a harder small-x gluon, with a softer one in the case of DGLAP pseudo-data. This different trend can be explained because for central charged hadron production, the CGC prediction leads to an enhancement as compared to the DGLAP one.

Therefore we conclude that the data from the central region cannot really discriminate between the two production scenarios, DGLAP and CGC.
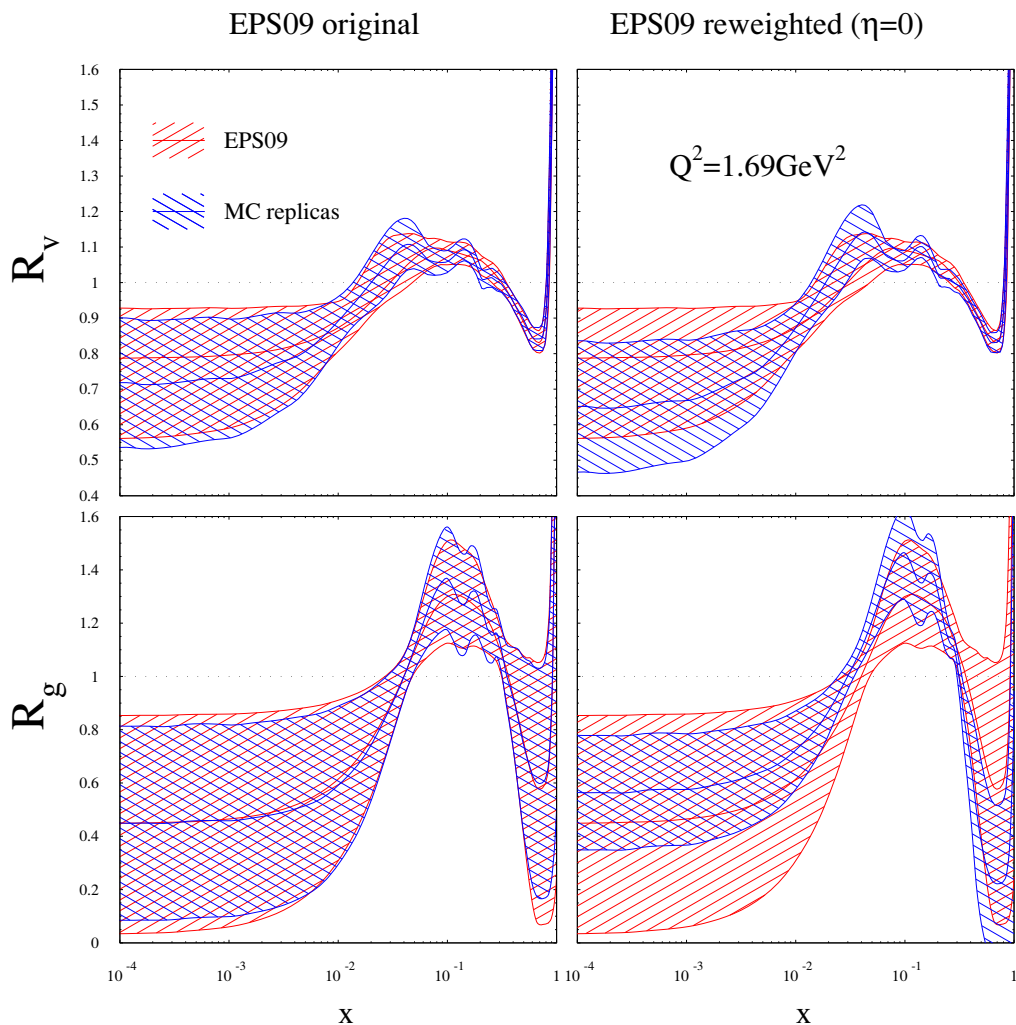
Figure 10. Same as figure 4, now for the nuclear modifications of up valence quarks (upper plots) and gluons (lower plots) when pPb LHC pseudo-data for inclusive charged hadron production at central ($\eta = 0$) rapidities is included in EPS09. The pseudo-data has been generated in the CGC framework.

Let us consider now the case in which CGC pseudo-data for charged hadron production from the the forward region is included in the nuclear PDF fit. From the reweighted cross section, shown in the lower right plot in figure 8, we see that including the (inconsistent) pseudo-data into the fit is not impossible, since the inital χ^2/n of 36.4 is reduced down to around 1.8 (see table 3). On the other hand, the effective number of replicas is only $N_{\text{eff}} = 1$, implying that all MC replicas have been discarded except the one that gives the best agreement with the CGC pseudo-data. Under these situations of extreme incompatibilty, the PDF reweighting method breaks down and becomes unreliable.

What this analysis suggests is that the differences in forward charged hadron production in pPb collisions at the LHC between the DGLAP and CGC frameworks is so substantial that it is not possible to absorb it by an update of the global nuclear PDF

analysis, and therefore that the potential for the discrimination between the two scenarios is very good.[3] Of course, measuring this very same processes at even forward rapidities would make the differences even more striking.

5   Summary and outlook

The accurate determination of how parton distributions are modified in nuclei is an essential input for our understanding of heavy ion collisions and of the physics of the quark-gluon plasma. Current analyses of nuclear PDFs include all available experimental information on the partonic structure of the nucleus, with however the bulk of the data restricted to deep-inelastic scattering at medium and large x. However, DIS is mostly sensitive to quark valence distribution, but fails to constrain sea and gluon densities, which are most relevant at small-x. Drell-Yan data has been included since the first analyses to provide constraints to the sea quarks. These fixed target data is again limited to the intermediate region of x and due to the large uncertainties is almost irrelevant e.g. for gluons. Recently, inclusive hadroproduction data from RHIC were also included in the fits to provide further constraints for the gluons. In any case, the amount of data is still quite limited both in kinematical coverage and in accuracy.

Therefore, the recent proton-lead run at the LHC has the potential to provide, before electron-ion colliders [46, 47] become eventually available, very important information on the nuclear modifications of PDFs, in particular for the small-x gluon and sea quarks which are now very badly constrained, which in turn would improve our theoretical predictions for lead-lead collisions. In addition, proton-lead data offers the possibility of uncovering new regimes of QCD, in particular, non-linear effects such as those incorporated in the Color Glass Condensate scenario could be important enough to be disentangled from standard collinear factorisation. Of course, the only way of quantifying the tension between the DGLAP and CGC frameworks is to perform a new global nuclear fit and verify if the new data can be accommodated or not by modifications of the nuclear PDFs in the collinear scenario.

With this motivation in mind, and since essentially no data in the hard-scattering regime is still available, in this paper we have presented a first study of the potential of LHC proton-lead measurements to constrain nuclear PDFs, based on simulated data. From the methodological point of view, instead of performing new versions of the EPS09 fit, we have applied the technique of Bayesian PDF reweighting, which is now routinely used in the case of proton PDFs. We have considered two representative processes: Drell-Yan production and inclusive charged particle production, both of which are sensitive to nuclear modifications of both gluons and sea quarks. For the case of Drell-Yan, we found that under conservative assumptions, available data has the potential to reduce the PDF uncertainties on the small-x nuclear sea quarks and specially in the medium and small-x nuclear gluon, where uncertainties can decrease by up to a factor two.

---

[3]Still, variations of the initial parameterisations of the nPDFs should be explored for this conclusion to be definitive.

Then we turned our attention to single inclusive hadroproduction. In this case two sets of pseudo-data were studied, one generated using the CGC scenario and other one generated using the same collinear DGLAP framework as the one used to produce the theory predictions. In the latter case, we find a similar impact as in the case of Drell-Yan pseudo-data, namely reduction of nuclear PDF uncertainties in the sea quarks but specially on the gluon. When pseudo-data is generated using CGC predictions, we find that the global nuclear fit is able to absorb the non-linear effects only for pseudo-data in the central region, while pseudo-data in the forward region was manifestly incompatible with the DGLAP predictions. Therefore, this process is interesting whatever the underlying behavior in real data turns out to be: if non linear effects are small, very useful constraints on virtually unknown nuclear PDFs will be derived; if on the other hand they are large, it is likely that experimental accuracy is enough to clearly identify the onset of the saturation dynamics.

Our study provides a first quantitative estimate of the potential of the proton-lead data to constrain PDFs, and confirms that such experimental results should be an essential ingredient of nuclear global PDF fits in the coming years, and thus become crucial for improving our understanding of heavy ion collisions. Of course, we have used some simplified assumptions, and crude estimates of the experimental uncertainties, so at the end only when the actual LHC measurements become available we will be able to quantify the impact of the data on the nuclear PDFs.

From the methodological point of view, the availability of a Monte Carlo version of EPS09 implies that the experimental groups themselves can study the impact of their data on nuclear PDF by means of the reweighting method discussed here, without the need to wait for an updated fit. With this motivation, EPS09MC has been made publicly available. The driver code, data files and documentation to use the EPS09 Monte Carlo sets (both the symmetric and asymmetric cases are provided) can be obtained from

<div align="center">http://igfae.usc.es/hotlhc/index.php/software.</div>

Acknowledgments

References

[1] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B 867** (2013) 244 [arXiv:1207.1303] [ɪɴSPIRE].

[2] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C 63** (2009) 189 [arXiv:0901.0002] [ɪɴSPIRE].

[3] J. Gao et al., *The CT10 NNLO global analysis of QCD*, arXiv:1302.6246 [ɪɴSPIRE].

[4] R.D. Ball et al., *Parton distribution benchmarking with LHC data*, *JHEP* **04** (2013) 125 [arXiv:1211.5142] [ɪɴSPIRE].

[5] S. Forte and G. Watt, *Progress in the determination of the partonic structure of the proton*, *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 291 [arXiv:1301.6754] [ɪɴSPIRE].

[6] K.J. Eskola, H. Paukkunen and C.A. Salgado, *EPS09: a new generation of NLO and LO nuclear parton distribution functions*, *JHEP* **04** (2009) 065 [arXiv:0902.4154] [ɪɴSPIRE].

[7] I. Helenius, K.J. Eskola, H. Honkanen and C.A. Salgado, *Impact-parameter dependent nuclear parton distribution functions: EPS09s and EKS98s and their applications in nuclear hard processes*, *JHEP* **07** (2012) 073 [arXiv:1205.5359] [ɪɴSPIRE].

[8] D. de Florian, R. Sassot, P. Zurita and M. Stratmann, *Global analysis of nuclear parton distributions*, *Phys. Rev.* **D 85** (2012) 074028 [arXiv:1112.6324] [ɪɴSPIRE].

[9] M. Hirai, S. Kumano and T.-H. Nagai, *Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order*, *Phys. Rev.* **C 76** (2007) 065207 [arXiv:0709.3038] [ɪɴSPIRE].

[10] I. Schienbein et al., *PDF nuclear corrections for charged and neutral current processes*, *Phys. Rev.* **D 80** (2009) 094004 [arXiv:0907.2357] [ɪɴSPIRE].

[11] N. Armesto, *Nuclear shadowing*, *J. Phys.* **G 32** (2006) R367 [hep-ph/0604108] [ɪɴSPIRE].

[12] C.A. Salgado et al., *Proton-nucleus collisions at the LHC: scientific opportunities and requirements*, *J. Phys.* **G 39** (2012) 015010 [arXiv:1105.3919] [ɪɴSPIRE].

[13] J.L. Albacete et al., *Predictions for $p+Pb$ collisions at $\sqrt{s_{NN}} = 5\,TeV$*, *Int. J. Mod. Phys.* **E 22** (2013) 1330007 [arXiv:1301.3395] [ɪɴSPIRE].

[14] H. Paukkunen and C.A. Salgado, *Constraints for the nuclear parton distributions from Z and W production at the LHC*, *JHEP* **03** (2011) 071 [arXiv:1010.5392] [ɪɴSPIRE].

[15] T. Stavreva et al., *Probing gluon and heavy-quark nuclear PDFs with $\gamma + Q$ production in pA collisions*, *JHEP* **01** (2011) 152 [arXiv:1012.1178] [ɪɴSPIRE].

[16] F. Arleo, K.J. Eskola, H. Paukkunen and C.A. Salgado, *Inclusive prompt photon production in nuclear collisions at RHIC and LHC*, *JHEP* **04** (2011) 055 [arXiv:1103.1471] [ɪɴSPIRE].

[17] Z.-B. Kang and J.-W. Qiu, *Nuclear modification of vector boson production in proton-lead collisions at the LHC*, *Phys. Lett.* **B 721** (2013) 277 [arXiv:1212.6541] [ɪɴSPIRE].

[18] K.J. Eskola, H. Paukkunen and C.A. Salgado, *A perturbative QCD study of dijets in $p+Pb$ collisions at the LHC*, arXiv:1308.6733 [ɪɴSPIRE].

[19] NNPDF collaboration, R.D. Ball et al., *Reweighting NNPDFs: the W lepton asymmetry*, *Nucl. Phys.* **B 849** (2011) 112 [*Erratum ibid.* **B 854** (2012) 926] [*Erratum ibid.* **B 855** (2012) 927] [arXiv:1012.0836] [ɪɴSPIRE].

[20] R.D. Ball et al., *Reweighting and unweighting of parton distributions and the LHC W lepton asymmetry data*, *Nucl. Phys.* **B 855** (2012) 608 [arXiv:1108.1758] [INSPIRE].

[21] G. Watt and R.S. Thorne, *Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs*, *JHEP* **08** (2012) 052 [arXiv:1205.4024] [INSPIRE].

[22] ALICE collaboration, *Transverse momentum distribution and nuclear modification factor of charged particles in p-Pb collisions at $\sqrt{s_{NN}} = 5.02\,TeV$*, *Phys. Rev. Lett.* **110** (2013) 082302 [arXiv:1210.4520] [INSPIRE].

[23] J.L. Albacete, A. Dumitru and C. Marquet, *The initial state of heavy-ion collisions*, *Int. J. Mod. Phys.* **A 28** (2013) 1340010 [arXiv:1302.6433] [INSPIRE].

[24] NNPDF collaboration, R.D. Ball et al., *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* **B 809** (2009) 1 [*Erratum ibid.* **B 816** (2009) 293] [arXiv:0808.1231] [INSPIRE].

[25] R.D. Ball et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, *Nucl. Phys.* **B 838** (2010) 136 [arXiv:1002.4407] [INSPIRE].

[26] H. Paukkunen and C.A. Salgado, *Agreement of neutrino deep inelastic scattering data with global fits of parton distributions*, *Phys. Rev. Lett.* **110** (2013) 212301 [arXiv:1302.2001] [INSPIRE].

[27] CMS collaboration, *Measurement of the Drell-Yan cross section in pp collisions at $\sqrt{s} = 7\,TeV$*, *JHEP* **10** (2011) 007 [arXiv:1108.0566] [INSPIRE].

[28] ATLAS collaboration, *Measurement of the high-mass Drell-Yan differential cross-section in pp collisions at $\sqrt{s} = 7\,TeV$ with the ATLAS detector*, *Phys. Lett.* **B 725** (2013) 223 [arXiv:1305.4192] [INSPIRE].

[29] LHCb collaboration, *Inclusive low mass Drell-Yan production in the forward region at $\sqrt{s} = 7\,TeV$*, LHCb-CONF-2012-013 (2012).

[30] K. Golec-Biernat, E. Lewandowska and A.M. Stasto, *Drell-Yan process at forward rapidity at the LHC*, *Phys. Rev.* **D 82** (2010) 094010 [arXiv:1008.2652] [INSPIRE].

[31] J.M. Campbell and R.K. Ellis, *Radiative corrections to $Zb\bar{b}$ production*, *Phys. Rev.* **D 62** (2000) 114012 [hep-ph/0006304] [INSPIRE].

[32] F. Caola, S. Forte and J. Rojo, *Deviations from NLO QCD evolution in inclusive HERA data*, *Phys. Lett.* **B 686** (2010) 127 [arXiv:0910.3143] [INSPIRE].

[33] J.L. Albacete, J.G. Milhano, P. Quiroga-Arias and J. Rojo, *Linear vs non-linear QCD evolution: from HERA data to LHC phenomenology*, *Eur. Phys. J.* **C 72** (2012) 2131 [arXiv:1203.1043] [INSPIRE].

[34] CMS collaboration, *Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7\,TeV$*, *Phys. Rev. Lett.* **105** (2010) 022002 [arXiv:1005.3299] [INSPIRE].

[35] ATLAS collaboration, *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, *New J. Phys.* **13** (2011) 053033 [arXiv:1012.5104] [INSPIRE].

[36] ALICE collaboration, *Energy dependence of the transverse momentum distributions of charged particles in pp collisions measured by ALICE*, arXiv:1307.1093 [INSPIRE].

[37] B. Jager, A. Schafer, M. Stratmann and W. Vogelsang, *Next-to-leading order QCD corrections to high p_T pion production in longitudinally polarized pp collisions*, *Phys. Rev.* **D 67** (2003) 054005 [hep-ph/0211007] [INSPIRE].

[38] R. Sassot, P. Zurita and M. Stratmann, *Inclusive hadron production in the CERN-LHC era*, *Phys. Rev.* **D 82** (2010) 074011 [arXiv:1008.0540] [INSPIRE].

[39] 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] [INSPIRE].

[40] J.L. Albacete, A. Dumitru, H. Fujii and Y. Nara, *CGC predictions for p+Pb collisions at the LHC*, *Nucl. Phys.* **A 897** (2013) 1 [arXiv:1209.2001] [INSPIRE].

[41] PHENIX collaboration, S.S. Adler et al., *Centrality dependence of π^0 and η production at large transverse momentum in $\sqrt{s_{NN}} = 200\,GeV\,d+Au$ collisions*, *Phys. Rev. Lett.* **98** (2007) 172302 [nucl-ex/0610036] [INSPIRE].

[42] STAR collaboration, J. Adams et al., *Pion, kaon, proton and anti-proton transverse momentum distributions from $p+p$ and $d+Au$ collisions at $\sqrt{s_{NN}} = 200\,GeV$*, *Phys. Lett.* **B 616** (2005) 8 [nucl-ex/0309012] [INSPIRE].

[43] STAR collaboration, J. Adams et al., *Identified hadron spectra at large transverse momentum in $p+p$ and $d+Au$ collisions at $\sqrt{s_{NN}} = 200\,GeV$*, *Phys. Lett.* **B 637** (2006) 161 [nucl-ex/0601033] [INSPIRE].

[44] STAR collaboration, B.I. Abelev et al., *Inclusive π^0, η and direct photon production at high transverse momentum in $p+p$ and $d+Au$ collisions at $\sqrt{s_{NN}} = 200\,GeV$*, *Phys. Rev.* **C 81** (2010) 064904 [arXiv:0912.3838] [INSPIRE].

[45] 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] [INSPIRE].

[46] A. Accardi et al., *Electron ion collider: the next QCD frontier — understanding the glue that binds us all*, arXiv:1212.1701 [INSPIRE].

[47] LHeC Study Group collaboration, J.L. Abelleira Fernandez et al., *A large hadron electron collider at CERN: report on the physics and design concepts for machine and detector*, *J. Phys.* **G 39** (2012) 075001 [arXiv:1206.2913] [INSPIRE].