Analysis of proton-lead data via re-weighting

P Zurita

Departamento de Física de Partículas and IGFAE, Universidade de Santiago de Compostela, E-15706 Santiago de Compostela, Galicia, Spain.

E-mail: pia.zurita@usc.es

Abstract.

The recent proton-lead run at the LHC shall provide new information on the partonic behaviour within the nuclear medium. At LHC energies the dominant contribution comes from gluon-initiated processes that is, the least well constrained parton density. Therefore it is important to profit from any information that new data can provide us. A time-saving alternative to performing a global fit is the use of Bayesian inference, a powerful tool to realize the impact of data into a set of PDFs independently of the original fitters. In this work we apply the Bayesian re-weighting technique to analyze pseudo data for LHC kinematics in Drell-Yan and hadro-production processes. A set of Monte Carlo replicas for EPS09 is released in a public code for general use.

1. Introduction

The precise determination of nuclear PDFs (nPDFs) has become an important asset in the light of the heavy-ion collisions at the LHC. However, due to the scarce amount of data from previous experiments and their limited kinematical coverage, the parton densities inside a nucleus are far from being well known. In particular, the nuclear gluon distribution has no real constrains. In this respect, the measurements from the LHC proton-lead run have the potential [1, 2, 3, 4, 5, 6, 7] to pinpoint the gluon behaviour in heavy nuclei.

The usual way to obtain information on the PDFs from a new experiment is to include the new results in a global fit, together with all previously analyzed data. Several choices must be done and variations of the fitting strategy have to be explored, turning it into a time-consuming process. One alternative to avoid performing a new fit was developed for proton PDFs based on the Monte Carlo method [8, 9]. By means of Bayesian inference one can quantitatively study the impact of new data on the central values and uncertainties of the PDFs.

In this contribution we summarize the results presented in [10]. There we apply the Bayesian re-weighting technique to explore the constraining potential of the LHC proton-lead run data on EPS09 [11] nuclear PDFs, taking two processes as example: Drell-Yan production and charged hadron inclusive production. The analyses are performed using pseudo-data based on the collinear DGLAP framework and also in the Color Glass Condensate (CGC) framework [12] for the case of hadro-production. The set of Monte Carlo replicas of EPS09 generated for this study is released as a public computer code.

In section 2 we present the main features of the re-weighting procedure and construct a Monte Carlo version of the EPS09 set. The impact of the new data on EPS09 nPDFS for Drell-Yan production and charged hadron production is studied in section 3. In the latter case, we also explore the discriminating power between DGLAP and CGC scenarios. Section 4 summarizes our results.

2. Bayesian re-weighting

Nowadays two methods are used in order to obtain the PDF uncertainties: the Hessian approach (used also for the nuclear PDFs) and the Monte Carlo approach (by the NNPDF Collaboration [13, 14]). In the latter the underlying PDF probability density $\mathcal{P}$ is sampled by generating, through a Monte Carlo procedure, an ensemble of N_{rep} PDFs replicas f_k, $k = 1, ..., N_{rep}$, each fitted to a replica of the experimental data. Then any quantity $\mathcal{O}[f]$ depending on the PDFs can be evaluated by

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

If we now want to include the information of a new measurement consisting of n points with covariance matrix cov_{ij}, $y = y_1, y_2, ..., y_n$, we can use Bayesian inference [8, 9] to update the original probability distribution $\mathcal{P}_{old}(f)$ to a new probability distribution $\mathcal{P}_{new}(f)$. This is achieved by computing a weight w_k for each individual replica f_k, which quantifies its agreement with the new data. In terms of the χ^2_k between the original theory predictions for the $k-th$ replica and the new experimental measurement, the weights are

$$w_k = \frac{(\chi^2_k)^{\frac{1}{2}(n-1)} e^{-\chi^2_k/2}}{\frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} (\chi^2_k)^{\frac{1}{2}(n-1)} e^{-\chi^2_k/2}}\,, \tag{2}$$

So that now the analog of (eq.1) is

$$\langle \mathcal{O} \rangle_{\rm new} = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} w_k \mathcal{O}[f_k]\,. \tag{3}$$

As long as the new data is not too constraining, the Bayesian re-weighting is equivalent to a full new fit, though the former loses some statistical efficiency with respect to the latter. The replicas with very small weights will become irrelevant when averaging and the accuracy of the representation of the underlying $\mathcal{P}_{new}(f)$ will decrease. We quantify the efficiency loss by computing N_{eff}, the effective number of replicas after re-weighting:

$$N_{\rm eff} \equiv \exp\left\{ \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} w_k \log(N_{\rm rep}/w_k) \right\}\,. \tag{4}$$

The accuracy of the re-weighted fit is the same that would be obtained if a new fit with N_{eff} replicas were to be performed. When $N_{eff} \ll N_{rep}$ the re-weighting method becomes unreliable and a full refit is mandatory.

The original re-weighting was developed for PDF sets based on the Monte Carlo approach. However when using a nuclear PDFs set (Hessian uncertainties), we must first generate the Monte Carlo replicas [15]. For this we took the central value of each EPS09 [11] parton (f_0) and shifted it according to

$$f_k(x,Q^2) = f_0(x,Q^2) + \sum_{i}^{N_{\rm eig}} \{ f_i^{\pm}(x,Q^2) - f_0(x,Q^2) \} |r_{k,i}|\,, \quad (k = 1, \ldots, N_{\rm rep})\,, \tag{5}$$

with N_{eig} the number of pairs of Hessian eigenvectors, $r_{k,i}$ random numbers from a Gaussian distribution of mean zero and variance one, and $f_{\pm i}$ the nPDF corresponding to the eigenvector $S_{\pm i}$. A symmetric version can be obtained by averaging over each pair of eigenvectors.

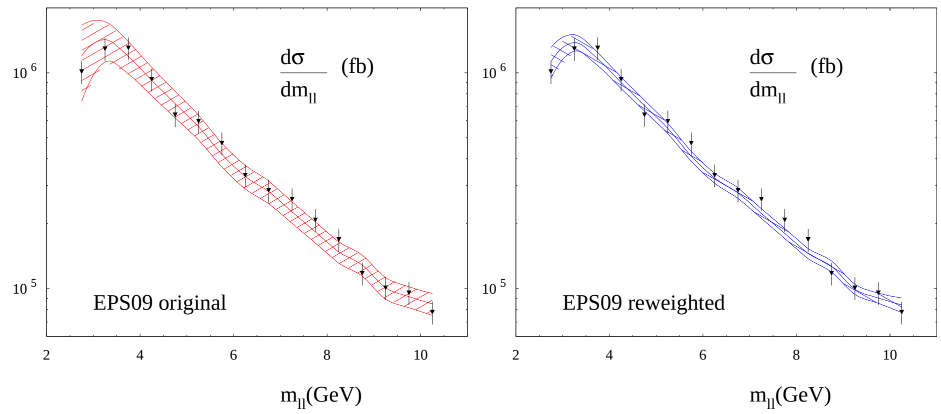

Figure 1. The predictions for the Drell-Yan differential cross sections with EPS09, compared to pseudo-data as a function of the invariant lepton pair mass m_{ll} at central rapidity before (left) and after (right) PDF re-weighting.

3. Results

3.1. Constraints from Drell-Yan production

first we analyze the impact on nuclear PDFs of Drell-Yan production in proton-lead collisions at the LHC, which should provide crucial information to further constrain the gluon density at small and medium x as well as the sea quark distributions. We are particularly interested in low-mass Drell-Yan production due to the enhanced sensitivity to nuclear effects [16].

In our exercise we consider only Drell-Yan cross sections for $|\eta| < 4$, without kinematical cuts on $p_{T,l}$. We generate both the pseudo-data and the predictions for DY production based on the $N_{rep} = 1000$ replicas of the MC version of EPS09 for $m_{ll} < 12GeV$ by use of the MCFM code [17], modified to take into account 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 and a total uncorrelated 8% systematic uncertainty. In what follows we only consider the PDF uncertainties coming from EPS09.

In figure 1 we show the pseudo-data and the predictions before (left) and after (right) re-weighting. Once the new data is included the uncertainty band of the theory prediction is reduced, without affecting the central value, as expected for consistent data. In table 1 we present the effective number of replicas N_{eff}, and the values for the mean χ^2 and the average over replicas $\langle\chi^2\rangle$ per data point before and after the re-weighting. We see that the χ^2/n is $\mathcal{O}(1)$, as expected by the use of consistent pseudo-data. About half of the replicas are disfavored by the pseudo-data: these are suppressed by re-weighting, reducing the value of $\langle\chi^2\rangle$.

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 $N_{\rm eff}$ is also provided for the re-weighting case.

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

By applying the re-weighting to the parton densities, we can assess the impact of the DY pseudo-data on the nuclear PDFs. While for the valence distributions there is no significant change, the sea distributions present a reduction on uncertainty bands in the $x < 10^{-2}$ region (plots not shown). The one parton that varies distinctly is the gluon, as can be seen in figure 2: the pseudo-data seems to favor more suppressed gluons in the $x < 10^{-2}$ region, and the estimated reduction of the uncertainty is up to 50%.

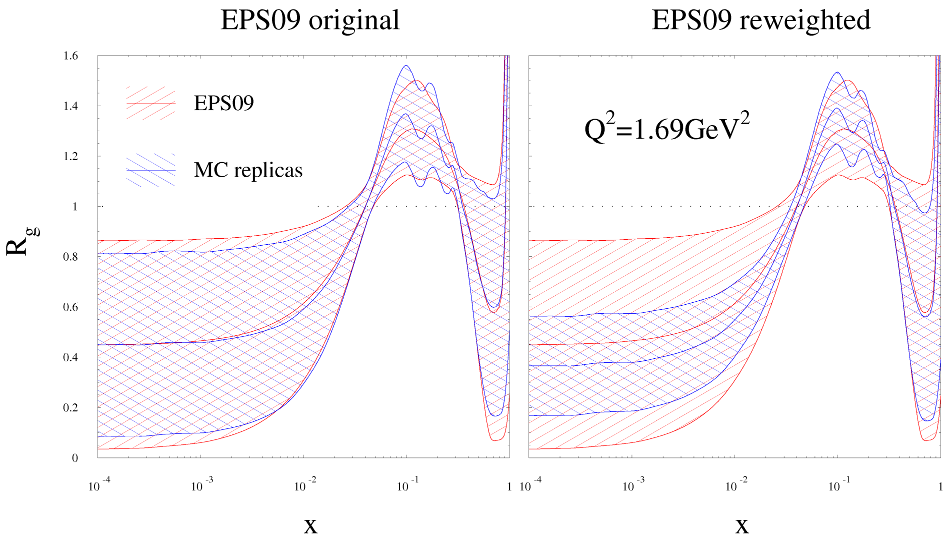

Figure 2. EPS09 nuclear ratio $R_g^A(x, Q^2)$ for $Q^2 = 1.69$ GeV^2, both in the Hessian and in the Monte Carlo versions, before (left) and after (right) the reweighting with LHC pPb Drell-Yan production pseudo-data.

Therefore, assuming full compatibility with the collinear framework, this measurement would provide useful information to improve our knowledge of the gluon density in lead nuclei.

3.2. Constraints from inclusive hadron production

In this case our observable is the cross section of charged hadron single inclusive production in proton-lead collisions divided by the same quantity in proton-proton collisions. Two scenarios were considered: one with pseudo-data generated from the collinear DGLAP framework and EPS09 and another using CGC predictions. The latter is of particular interest to study the discriminating power of this measurement regarding saturation dynamics. In both cases the statistical uncertainty was determined from the expected number of events in each data bin, and we have assumed 5% and 7% (uncorrelated) systematic and normalization uncertainties respectively. Theoretical predictions have been computed using the code for NLO inclusive hadron production of [18], modified to account for nuclear effects as discussed in [19].

In figure 3 we show the comparison between the pseudo-data and the EPS09 predictions, for two rapidities: $\eta = 0$ and $\eta = 2$. The re-weighting produces a shift of the central values (suppression for $p_T < 8$ GeV and enhancement above said value) and a narrowing of the uncertainties. Then in table 2 we present the values of χ^2/n, $\langle\chi^2\rangle/n$ and N_{eff}. Unsurprisingly, as data and theory are compatible by construction, many replicas remain. The behaviour for both rapidities is similar and thus we expect a compatible modification of the nuclear PDFs.

As happened for DY, the valence is almost unaffected and the sea density decreases with the error band slightly reduced for $x < 10^{-2}$ (plots not shown). In figure 4 we present the modification of the gluon density due to the re-weighting. The central values show an enhancement for $0.07 < x < 0.2$ while for $x < 10^{-2}$ they are suppressed and the error band shrinks about 50%. The displacements of the central curves, less pronounced in the forward region, are fully compatible within uncertainties, and compatible also with the variation from

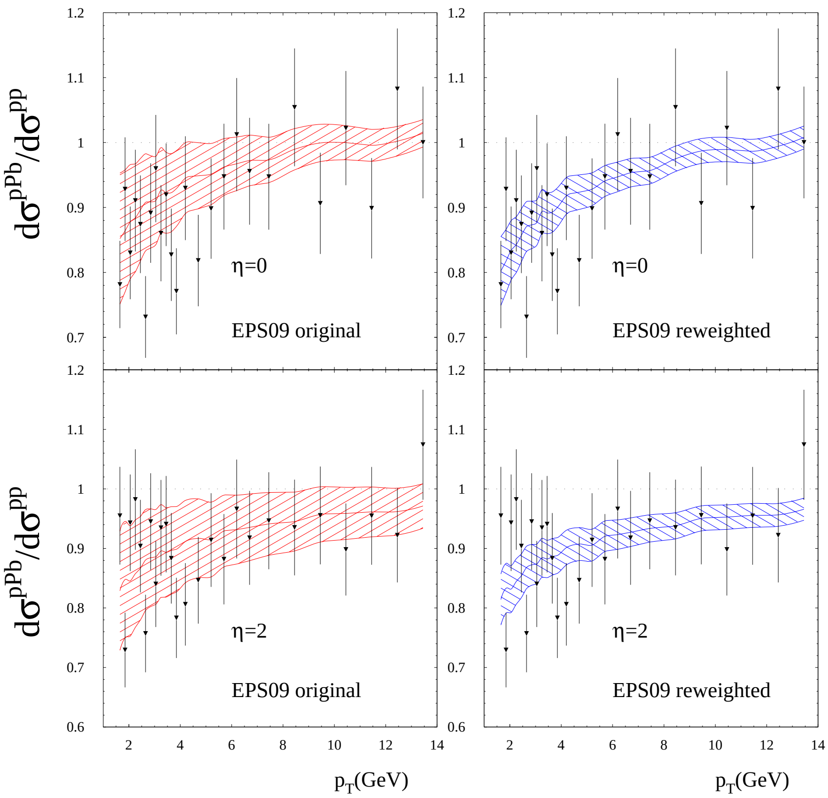

Figure 3. Upper plots: ratio of the charged hadron single inclusive cross section in pPb with respect to pp collisions, as a function of the hadron transverse momentum p_T at central rapidity before (left) and after (right) the reweighting, with pseudo-data generated in the collinear DGLAP framework using EPS09. Lower plots: the same for forward rapidities, $\eta = 2$.

Table 2. Same as Table 1 for the re-weighting 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.

	$\eta = 0$			$\eta = 2$		
	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

figure 2. Such error reduction in the small-x nuclear gluon ratio is desirable and supports the importance of including this observable in the nuclear PDF fits.

Now we turn our attention to pseudo-data that has been generated in the Color Glass Condensate scenario [12, 20]. In this case the underlying theory for pseudo-data and the one used for the EPS09 predictions are independent, allowing us to analyze the potential discriminating power of this observable. As before, we consider the central and forward rapidity regions. The latter is particularly interesting as data from RHIC in the forward region has not been included so far in nuclear PDFs fits.

We present in figure 5 the nuclear cross section ratios corresponding to $\eta = 0$ (upper plots) and $\eta = 2$ (lower plots) before and after the re-weighting, as a function of the transverse momentum p_T of the final hadron. The corresponding values of χ^2/n, $\langle\chi^2\rangle$ and N_{eff} are shown in table 3.

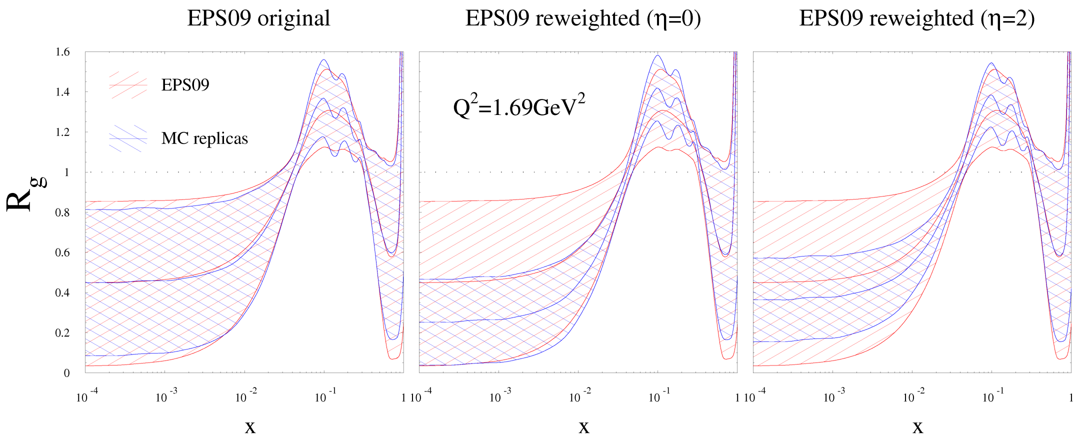

Figure 4. Same as fig. 2, when pPb LHC pseudo-data for inclusive charged hadron production at central ($\eta = 0$) and forward ($\eta = 2$) rapidities are included in EPS09. The pseudo-data has been generated in the collinear DGLAP framework.

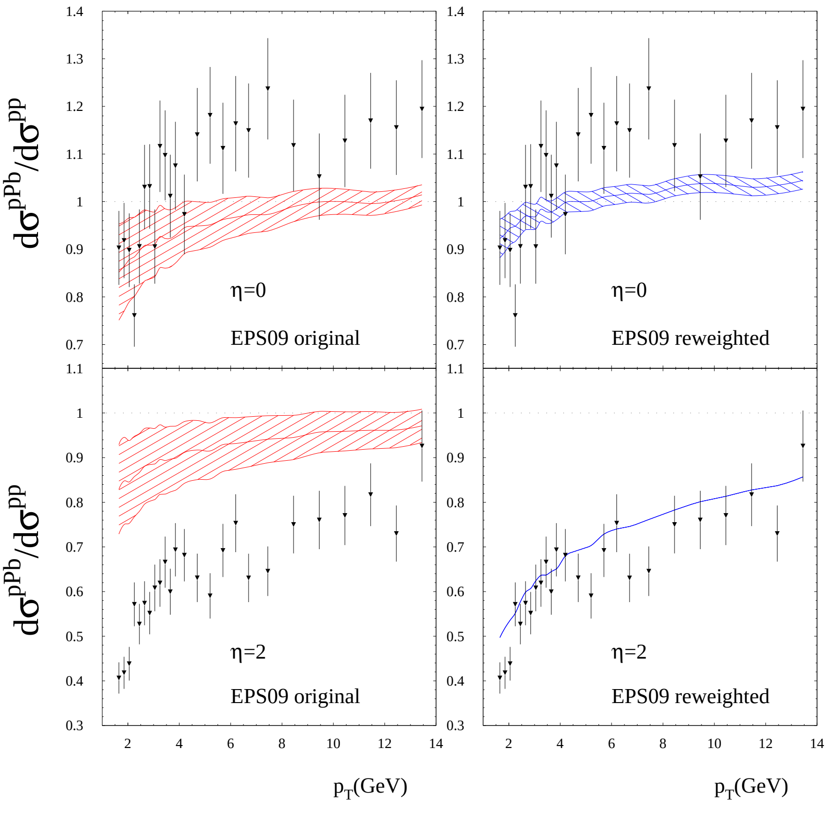

Figure 5. Upper plots: ratio of the charged hadron single inclusive cross section in pPb with respect to pp at central rapidity, as a function of the final hadron transverse momentum p_T before (left) and after (right) the reweighting. Lower plots: the same for $\eta = 2$. Note that the pseudo-data has been generated from the CGC predictions, while EPS09 predictions are based on the collinear DGLAP framework.

Let us begin by discussing the central rapidity region, $\eta = 0$. After re-weighting the agreement between theory and pseudo-data is reasonable. The central value is shifted upwards and the nPDF uncertainties are reduced, leaving 229 effective replicas. The final χ^2/n is above one, pointing the existence of some tension between the CGC and the DGLAP predictions.

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

Regarding the impact on the nuclear PDF ratios, the behaviour is different from the one seen with DGLAP pseudo-data: the valence distributions are affected, presenting a non-negligible decrease of the central value but no noticeable change is seen for the sea densities (plots not shown). As for the gluons (figure 6) the central value moves upwards and the uncertainty shrinks around 30% for $x < 10^{-2}$. This trend is opposite to the one seen in figure 4, but they are not incompatible and thus data from the central region is not likely to have the power to discriminate between the two production scenarios.

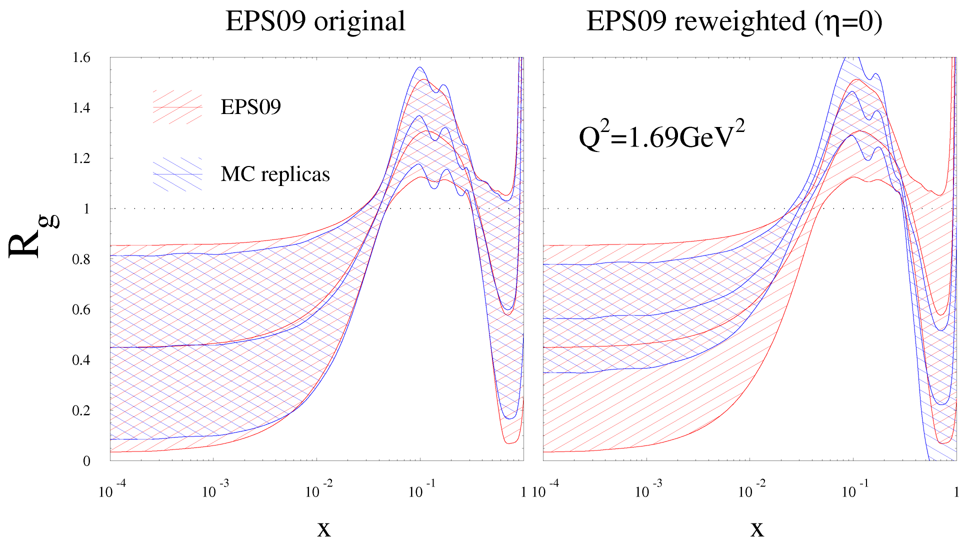

Figure 6. Same as fig. 2, now for pseudo-data generated in the CGC framework.

Now we move to the forward region. The inclusion of the, in principle, inconsistent data is not impossible (lower right plot in figure 5), since the re-weighting brings down the original χ^2/n from 36.4 to 1.8 (see table 3). However, the remaining number of replicas is only $N_{eff} = 1$, implying that almost all MC replicas have been discarded, and just retaining the one that gives the best agreement with the CGC pseudo-data. This is a situation of extreme incompatibility and the re-weighting method breaks down, becoming unreliable. In turn, our result suggests that the differences in forward charged hadron production between the DGLAP and CGC frameworks are substantial and that the LHC has the potential for the discrimination between the two scenarios.

4. Summary

We have presented a study of the potential of LHC proton-lead measurements to constrain nuclear PDFs, based on simulated data. We applied the technique of Bayesian PDF re-weighting and considered two representative processes sensitive to gluons and sea quarks: Drell-Yan production and inclusive charged particle production. In the first case we found that the 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 (up to a factor two). In the second case we analyzed one set of pseudo-data generated using the CGC scenario and another one generated using the collinear DGLAP framework. In the latter we found a similar behaviour as for Drell-Yan pseudo-data. However, when pseudo-data comes from CGC predictions the global nuclear fit can be adapted to absorb non-linear effects only in the central rapidity region. In the forward rapidity region, the manifest incompatibility between data and theory breaks down the re-weighting. Regardless the behaviour that real data might have, this process is interesting: constraints on virtually unknown nuclear PDFs can be derived if the non-linear effects are small and, if they are large, experimental accuracy could be enough to identify the onset of the saturation dynamics.

Then our study, based on simplified assumptions, gives a first estimate of the potential of the proton-lead data to constrain nPDFs. Of course when the actual LHC measurements become available we will be able to quantify the impact of the data on the nuclear PDFs. We also release a public version of the Monte Carlo EPS09 replicas, that allows the experimental groups to study the impact of their data on nuclear PDFs, without the need to wait for an updated fit. The driver code, data files and documentation can be obtained from **http://igfae.usc.es/hotlhc/index.php/software**.

Acknowledgments

I thank my collaborators on the work described here: N. Armesto, J. Rojo and C. A. Salgado. J. R. was partially supported by a Marie Curie IntraEuropean fellowship of the European Community′s 7th framework Programme under contract number PIEF-GA-2010-272515. This work is supported by European Research Council grant HotLHC ERC-2011-StG-279579; by Ministerio de Ciencia e Innovación of Spain under projects FPA2009-11951 and FPA2011-22776, and the Consolider-Ingenio 2010 Programme CPAN (CSD2007-00042); by Xunta de Galicia (GRC2013-024); and by FEDER. I thank Kari Eskola and Hannu Paukkunen for useful discussions, and Javier Albacete for providing us with the CGC predictions. I am also grateful for the hospitality of the CERN TH Unit where part of this work was performed.

References

[1] Salgado C A, Alvarez-Muniz J, Arleo F, Armesto N, Botje M, Cacciari M, Campbell J and Carli C *et al.* 2012 *J. Phys. G* **39** 015010

[2] Albacete J, Armesto N, Baier R, Barnafoldi G G, Barrette J, De S, Deng W-T and Dumitru A *et al.* 2013 *Int. J. Mod. Phys.* E **22** 1330007

[3] Paukkunen H and Salgado C A 2011 *JHEP* **1103** 071

[4] Stavreva T, Schienbein I, Arleo F, Kovarik K, Olness F, Yu J Y and Owens J F 2011 *JHEP* **1101** 152

[5] Arleo F, Eskola K J, Paukkunen H and Salgado C A 2011 *JHEP* **1104** 055

[6] Kang Z B and Qiu J W 2013 *Phys. Lett.* B **721** 277

[7] Eskola K J, Paukkunen H and Salgado C A 2013 A perturbative QCD study of dijets in p+Pb collisions at the LHC (*Preprint* hep-ph/1308.6733)

[8] Ball R D *et al.* [NNPDF Collaboration] 2011 *Nucl. Phys.* B **849** 112, 2012 [Erratum-ibid. B **854** (2012) 926], [Erratum-ibid. B **855** (2012) 927]

[9] Ball R D, Bertone V, Cerutti F, Del Debbio L, Forte S, Guffanti A, Hartland N P and Latorre J I *et al.* [NNPDF Collaboration] 2012 *Nucl. Phys.* B **855** 608

[10] Armesto N, Rojo J, Salgado C A and Zurita P 2013 *JHEP* **1311** 015

[11] Eskola K J, Paukkunen H and Salgado C A 2009 *JHEP* **0904** 065

[12] Armesto N 2006 *J. Phys.* G **32** R367

[13] Ball R D *et al.* [NNPDF Collaboration] 2009 *Nucl. Phys.* B **809** 1, [Erratum-ibid. B **816** 293]

[14] Ball R D, Del Debbio L, Forte S, Guffanti A, Latorre J I, Rojo J and Ubiali M [NNPDF Collaboration] 2010 *Nucl. Phys.* B **838** 136

[15] Watt G and Thorne R S 2012 *JHEP* **1208** 052

[16] Golec-Biernat K, Lewandowska E and Stasto A M 2010 *Phys. Rev.* D **82** 094010

[17] Campbell J M and Ellis R K 2000 *Phys. Rev.* D **62** 114012

[18] Jager B, Schafer A, Stratmann M and Vogelsang W 2003 *Phys. Rev.* D **67** 054005

[19] Sassot R, Stratmann M and Zurita P 2010 *Phys. Rev.* D **82** 074011

[20] Albacete J, Dumitru A, Fujii H and Nara Y 2013 *Nucl. Phys.* A **897** 1