

# Impact of Inclusive Electron Ion Collider Data on Collinear Parton Distributions

Néstor Armesto<sup>1</sup>, Thomas Cridge<sup>2†</sup>, Francesco Giuli<sup>3</sup>, Lucian Harland-Lang<sup>4</sup>, Paul Newman<sup>5</sup>, Barak Schmookler<sup>6</sup>, Robert Thorne<sup>4</sup>, Katarzyna Wichmann<sup>2</sup>

<sup>1</sup> *Departamento de Física de Partículas and IGFAE, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Galicia, Spain*

<sup>2</sup> *Deutsches Elektronen-Synchrotron DESY, Germany*

<sup>3</sup> *CERN, CH-1211 Geneva, Switzerland*

<sup>4</sup> *Department of Physics & Astronomy, University College, London, WC1E 6BT, UK*

<sup>5</sup> *School of Physics & Astronomy, University of Birmingham, B15 2TT, UK*

<sup>6</sup> *University of California, Riverside, Department of Physics & Astronomy, CA 92521, USA*

## Abstract

A study is presented of the impact of simulated inclusive Electron Ion Collider Deep Inelastic Scattering data on the determination of the proton and nuclear parton distribution functions (PDFs) at next-to-next-to-leading and next-to-leading order in QCD, respectively. The influence on the proton PDFs is evaluated relative to the HERAPDF2.0 set, which uses inclusive HERA data only, and also relative to the global fitting approach of the MSHT20 PDFs. The impact on nuclear PDFs is assessed relative to the EPPS21 global fit and is presented in terms of nuclear modification ratios. For all cases studied, significant improvements in the PDF uncertainties are observed for several parton species. The most striking impact occurs for the nuclear PDFs in general and for the region of high Bjorken  $x$  in the proton PDFs, particularly for the valence quark distributions.

---

<sup>†</sup>thomas.cridge@desy.de

# 1 Introduction

Parton distribution functions (PDFs) are essential tools for understanding experimental data from collider experiments, and a necessary part of any theory prediction for hard hadronic interactions. Since the PDF dependence on Bjorken  $x$  cannot be calculated perturbatively, it has to be extracted from experimental measurements. Deep Inelastic Scattering (DIS) data from the HERA  $ep$  collider are at the core of current proton PDF extractions. In different approaches, they are used either as the sole ingredients [1] or else in global fits [2–7] together with fixed target DIS data and measurements of PDF-sensitive processes at the Tevatron and LHC hadron colliders. Nuclear parton distribution functions (nPDFs) describe the behavior of bound partons in the nucleus. Since there has yet to be an electron-nucleus collider, the standard approach is to combine fixed target DIS and Drell-Yan data with hard process measurements ( $\pi^0$ ,  $D$ , electro-weak boson and jet production) from proton-nucleus collisions at RHIC and the LHC, in global analyses [8–12], see [13] for a recent review.

The Electron Ion Collider (EIC) is expected to begin operations around 2030 at Brookhaven National Laboratory in the USA, with a broad scientific programme of enhancing our knowledge of nucleon and nuclear structure and dynamics. It will provide very precise  $ep$  DIS data in a complementary kinematic region to that of HERA. The prospect of EIC data at large  $x$  and intermediate  $Q^2$  is of particular interest, since existing large  $x$  DIS data from HERA are limited in their experimental precision, whilst fixed target data are subject to nuclear corrections and other theoretical complications. The impact of the EIC on proton PDFs has been studied previously in [14–18].

As the world’s first  $eA$  collider, the EIC will explore nuclear structure to an unprecedented level of detail up to the heaviest nuclei, in particular providing the first sensitivity from a DIS experiment in the low  $x$  kinematic regime. The impact of the EIC on nPDFs has previously been considered in [14, 18–20].

In this work we study the impact of the EIC on proton and nuclear PDFs using simulated EIC inclusive cross sections for neutral current (NC) and charged current (CC) DIS, with kinematic range, resolutions and uncertainties derived from a realistic detector design [21]. The uncertainties obtained in fits with parton evolution governed by the DGLAP equations [22–26] at next-to-next-to-leading order (NNLO) [27–29] are compared between baselines given by existing PDF sets and projections with the additional inclusion of simulated EIC data. A closely related study that assesses the EIC impact on the strong coupling  $\alpha_S$  can be found in [30].

## 2 Simulated EIC data

The detailed experimental apparatus to be deployed at the EIC is currently under intense development. However, the broad requirements to realise the science programme are well established [14], including the kinematic range and level of performance expected for inclusive DIS. For the purpose of this work, the simulated EIC data are taken from the studies performed in the framework of the ATHENA detector proposal [21]. Since the proposal publication, ATHENA concepts have been combined with those from the ECCE proposal [31] in the framework of a new ePIC collaboration, while discussions about a second complementary detector are

also ongoing. The results here are not expected to differ significantly for any general purpose EIC detector.

Simulated neutral current (NC) EIC  $ep$  measurements (‘pseudo-data’), are produced with integrated luminosities corresponding to expectations for one year of data taking with each of the five different currently planned beam configurations. The details are outlined in Tab. 1. Charged current (CC) pseudo-data are also available at the highest  $\sqrt{s}$ . For the NC pseudo-data, a grid is produced with five logarithmically-spaced  $x$  and  $Q^2$  values per decade over the range<sup>1</sup>  $0.001 < y < 0.95$ , which is well-justified by the expected resolutions.

The central values of the pseudo-data are taken from predictions using the HERAPDF2.0NNLO [1] and MSHT20 NLO [3] PDF sets (for Sections 3.1 and 3.2, respectively). In the latter case, the NLO set (and matching NLO theory) is taken in order to inject some pseudo-data/theory disagreement, as discussed further in Section 3.2. The pseudo-data points are randomly smeared based on Gaussian distributions with standard deviations given by the projected uncertainties as estimated by the ATHENA collaboration [21,32]. The simulated systematic errors are based on experience from HERA and further considerations in [14] and are rather conservative in the context of the modern detector technologies and large data sets expected at the EIC. Most data points have a point-to-point uncorrelated systematic uncertainty of 1.9%, extending to 2.75% for  $y < 0.01$ . A normalisation uncertainty of 3.4% is also ascribed; it is taken as fully correlated between data at each  $\sqrt{s}$ , and fully uncorrelated between data sets with different  $\sqrt{s}$ . For the purposes of the QCD fits, the point-to-point systematic uncertainties are added in quadrature with the statistical uncertainties and the normalisation uncertainties are treated as nuisance parameters, as in [1].

$e$ -beam energy (GeV)	$p$ -beam energy (GeV)	$\sqrt{s}$ (GeV)	Integrated lumi (fb <sup>-1</sup> )
18	275	141	15.4
10	275	105	100
10	100	63	79.0
5	100	45	61.0
5	41	29	4.4

Table 1: Beam energies, centre-of-mass energies and annual integrated luminosities of the different configurations considered for the EIC.

The locations in the  $(x, Q^2)$  kinematic plane of the HERA data and EIC pseudo-data used in this analysis are shown in Fig. 1. The EIC pseudo-data overlap in their coverage with the HERA data and extend the kinematic reach to the highest  $x$  values in the intermediate  $Q^2$  region.

The large  $x$  HERA data are relatively imprecise due to the kinematic correlation with large  $Q^2$ , the  $1/Q^4$  photon propagator term in the cross section and the modest integrated luminosity ( $\sim 0.5$  fb<sup>-1</sup> per experiment). The large  $x$  region in global fits is therefore constrained to a large extent by measurements from fixed target experiments, e.g. BCDMS and NMC [33, 34]. However, there are uncertainties in the theoretical description of the fixed target data due to their low hadronic final

<sup>1</sup>Here,  $y$  is the usual inelasticity variable,  $y = Q^2/(sx)$ .

## HERA and EIC kinematic phase-space

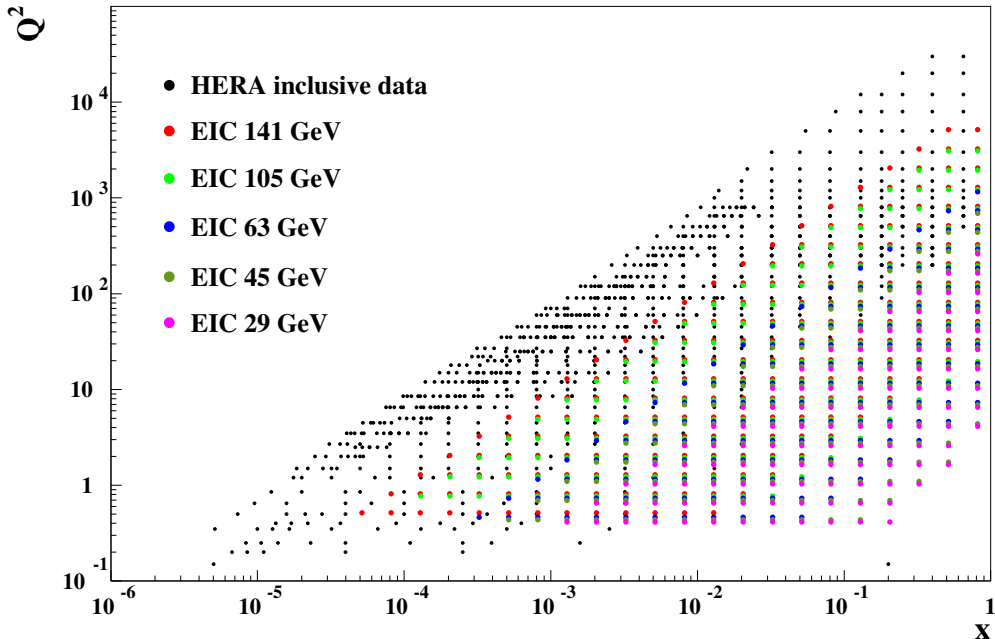


Figure 1: The locations in the  $(x, Q^2)$  plane of the HERA and EIC neutral current inclusive DIS data points included in the analysis. Figure reproduced from [30].

state invariant masses<sup>2</sup>, values where it becomes difficult to disentangle perturbative corrections from power-like effects. The EIC is thus particularly promising in the high  $x$  region, where it is expected to provide data that are both high precision and theoretically clean.

$eA$  pseudodata were produced analogously, considering the nucleus to be Au, and per-nucleon integrated luminosities of  $4.4 \text{ fb}^{-1}$ ,  $79 \text{ fb}^{-1}$  and  $79 \text{ fb}^{-1}$  for  $5 \times 41 \text{ GeV}$ ,  $10 \times 110 \text{ GeV}$  and  $18 \times 110 \text{ GeV}$ , respectively. Note that we are interested in the uncertainties while the central values are irrelevant for this study. Therefore, the same PDF set HERAPDF2.0NNLO [1] used for the proton is employed for  $eA$ , corresponding to a central value of the nuclear modification factor (defined as the ratio of each parton density in a proton bound inside a nucleus to that in a free proton) equal to 1.

## 3 EIC Impact on Proton PDFs

### 3.1 Comparison with HERA-Only PDFs

The results presented in this section are obtained from global QCD fits at NNLO, performed in the HERAPDF2.0 framework [1] using xFitter, an open source QCD fit platform [35]. Fits with identical configurations are performed to HERA data only, corresponding to HERAPDF2.0NNLO in [1], and also with the additional inclusion of the simulated EIC pseudo-data described in Section 2. To avoid regions that may

<sup>2</sup>The hadronic final state invariant mass  $W$  is related to the other standard DIS kinematic variables through  $W^2 = Q^2(1 - x)/x$ .

be strongly affected by higher twist or resummation effects, a cut on the squared hadronic final state invariant mass,  $W^2 = Q^2(1-x)/x > 10 \text{ GeV}^2$  is included for the EIC data. The central values of the PDFs with and without the EIC pseudo-data coincide by construction, so the uncertainties can be compared directly.

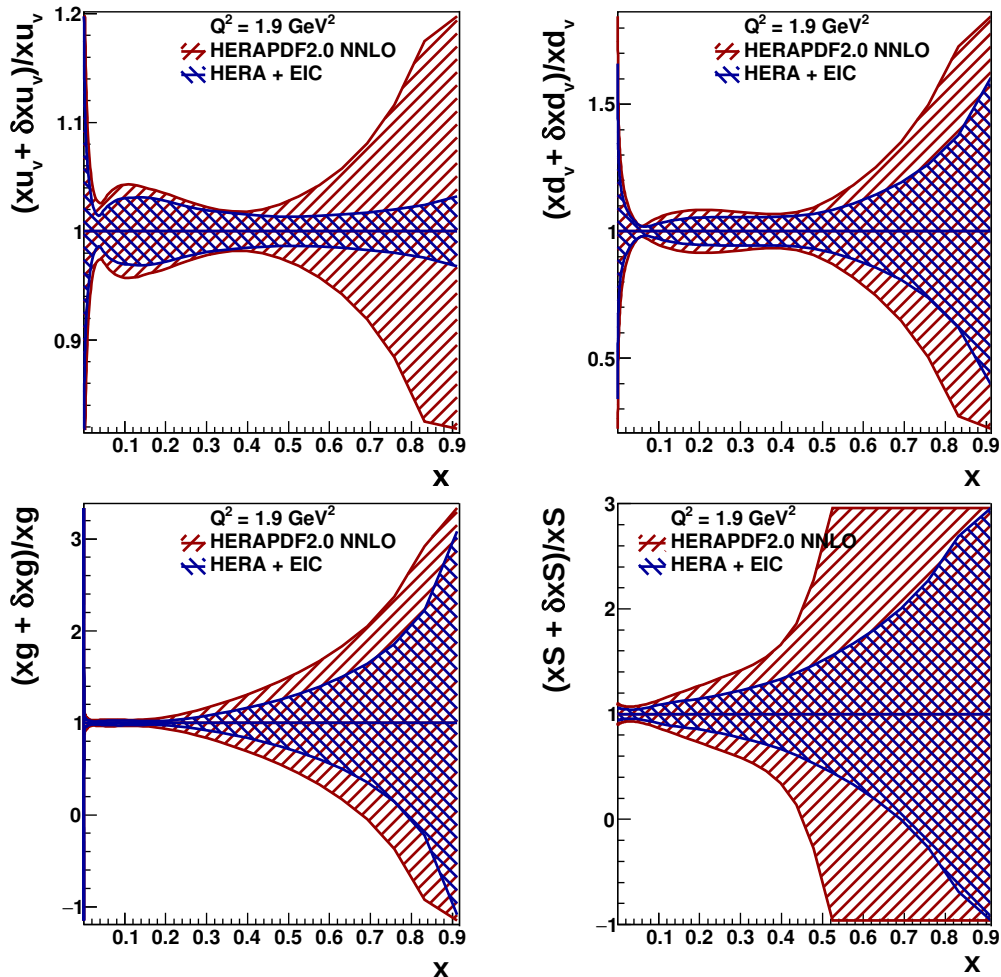


Figure 2: Impact of simulated EIC data on the NNLO collinear parton distributions of the proton. The bands show relative total uncertainties as a function of  $x$  for the up-valence, down-valence, gluon and total sea distributions, for  $Q^2 = 1.9 \text{ GeV}^2$ . The HERAPDF2.0NNLO total uncertainties (using HERA data alone) are compared with results in which simulated EIC data are also included in the HERAPDF2.0NNLO fitting framework.

The impact of the EIC pseudo-data on the experimental uncertainties in the HERAPDF2.0NNLO fits is illustrated in figures 2 and 3. Relative uncertainties are shown for the up-valence, down-valence, gluon and total sea-quark densities. The uncertainty bands show the symmetrised total uncertainty, including experimental, model and parameterisation contributions<sup>3</sup>, as discussed in detail in [1]. The experimental uncertainties are generally dominant in regions where there is data. A

<sup>3</sup>For technical reasons, the number of model and parameterisation uncertainties differs between the HERAPDF2.0 baseline and the version with added EIC pseudodata. However, this is expected to cause only comparatively minor differences.

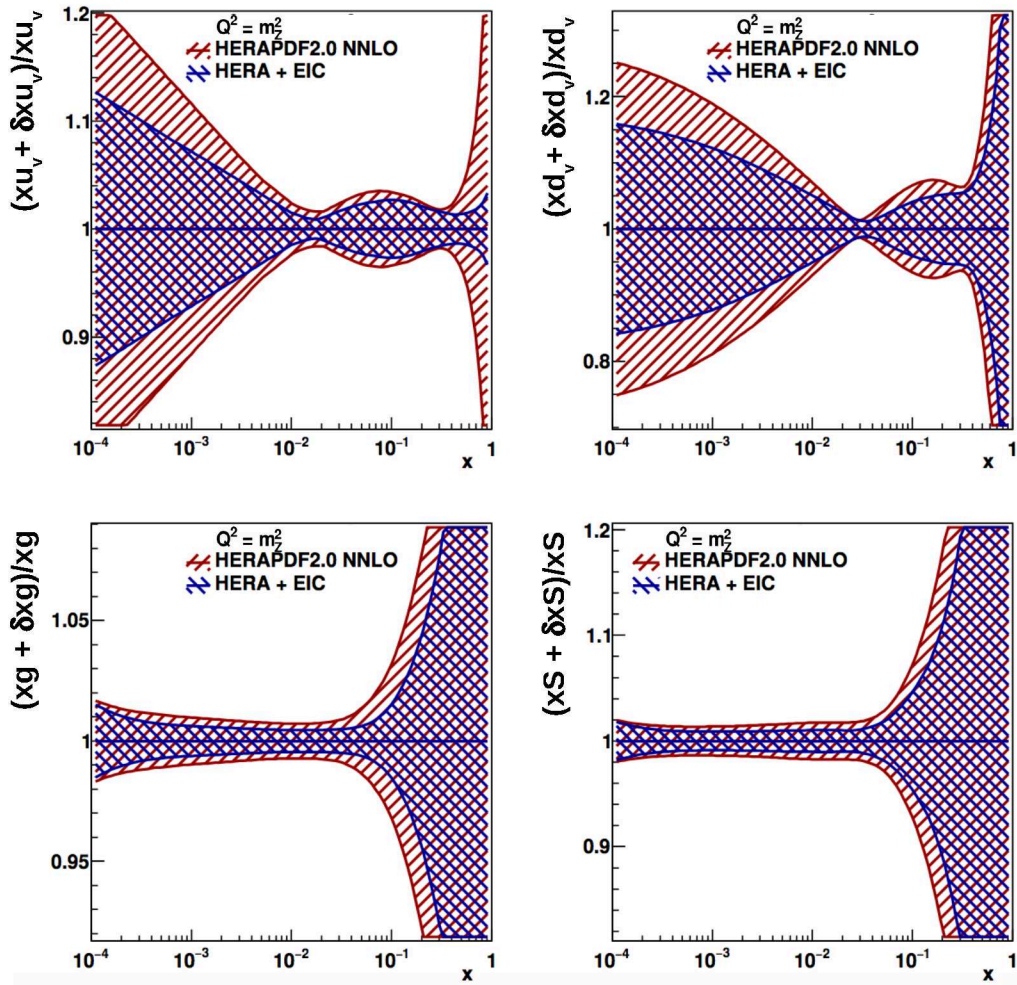


Figure 3: Impact of simulated EIC data on the NNLO collinear parton distributions of the proton shown on a logarithmic  $x$  scale for  $Q^2 = m_Z^2$  GeV<sup>2</sup>. The bands show relative total uncertainties as a function of  $x$  for the up-valence, down-valence, gluon and total sea distributions. The HERAPDF2.0NNLO total uncertainties (using HERA data alone) are compared with results in which simulated EIC data are also included in the HERAPDF2.0NNLO fitting framework.

significant reduction in uncertainties for all parton species is observed when adding the EIC pseudo-data, as discussed in detail in the following.

Since the high  $x$  regime is of particular interest, the potential improvements are most readily visualised on a linear  $x$  scale, as shown in figure 2 at the starting scale for DGLAP evolution,  $Q^2 = 1.9$  GeV<sup>2</sup>. The EIC impact is most striking in the very large  $x$  region  $x \gtrsim 0.7$ , particularly dramatic for the  $u$  valence quark.<sup>4</sup> The improved constraints are traceable to the large integrated luminosity of the EIC pseudo-data and the correspondingly improved data precision in the high- $x$  region compared with HERA. The charge-squared weighting of the photon couplings result in stronger sensitivity to the up than to the down quark density. Figure 3 shows

<sup>4</sup>Near to the starting scale for QCD evolution, the PDFs are dominated by the valence quark densities for typically  $x \gtrsim 0.2$ , with the gluon density becoming dominant at lower  $x$  values.

the same information as that in figure 2, but as a logarithmic function of  $x$  and at the electroweak scale,  $Q^2 = m_Z^2 \text{ GeV}^2$ . At low and intermediate  $x$ , where our current knowledge of proton collinear structure is dominated by HERA data, the EIC pseudo-data continue to have a considerable impact on the quark densities. Whilst this is in part due to the large EIC luminosity and extended phase-space, the very low  $x$  constraints on the valence quarks come primarily from the valence quark number sum rules, and thus derive indirectly from the improved precision at large  $x$ . The gluon density, which is most strongly constrained by the scaling violations of the neutral current cross sections, is not impacted so strongly, but there is still a notable improvement at intermediate  $x$ , where the EIC pseudo-data improve the precision at low and intermediate  $Q^2$ .

The lowest  $x$  and  $Q^2$  HERA data show evidence for  $\ln(1/x)$  (BFKL) resummation corrections [36]. To investigate whether precise EIC data in the low  $x$  region might enhance the overall sensitivity to such effects, modified EIC pseudo-data are generated with central values obtained from the NNPDF31sx\_nnlo\_nllx\_as\_0118 PDF set [36], which includes the  $\ln(1/x)$  resummation effects. Fits at NNLO using the modified pseudo-data are compared with and without the addition of next-to-leading  $\ln(1/x)$  resummation, implemented using the HELL [37,38] code. The overall  $\chi^2$  of the fit does not change significantly between the two cases, indicating that the EIC data are not sensitive to the BFKL effects. The same conclusion is reached when using a more flexible PDF parameterisation [39], designed to give a better description in the small- $x$  region without introducing too many new parameters.

### 3.2 Comparison with Global PDFs

Several groups worldwide are engaged in ‘global’ fits that constrain the proton PDFs using a range of input data from fixed target DIS experiments, HERA and hadron-hadron colliders, most notably the LHC [40]. Within this global dataset, HERA inclusive DIS data remain a key constraint, in particular at low and intermediate  $x$ . However, other datasets also play an important role. For example, fixed target DIS data are important in constraining the quarks in the high  $x$  region, while LHC measurements are playing an increasingly important role, most notably in the quark flavour decomposition and for the gluon at high  $x$  [2,41]. Within such a fit, the impact of any additional data will necessarily be balanced by the pulls of other datasets in the fit. The overall impact is therefore expected to be reduced in comparison to a fit where the EIC pseudo-data are added to a more limited, HERA-only, baseline.

The impact of EIC data on the high  $x$  PDFs has been studied relative to a recent example global fit, MSHT20 [3,42]. The same cuts,  $Q^2 \geq 2 \text{ GeV}^2$  and  $W^2 \geq 15 \text{ GeV}^2$  are applied as in [3]. As described in Section 2 the EIC pseudodata are produced using NLO QCD theory and, consistently, with MSHT20NLO PDFs, while the fit is performed at NNLO. This is in order to effectively inject some inconsistency between theory and pseudo-data, as one might expect to occur in a real comparison between data and theory. The impact of this procedure on the corresponding PDF uncertainties, relative to the case adopted in Section 3.1 where theory and pseudo-data agree by construction, is minimal.

As expected, the EIC impact relative to MSHT20 is substantially reduced compared with that relative to HERAPDF2.0. However, there are still significant effects, as illustrated in figure 4. Due to the charge-squared coupling of the virtual photon in

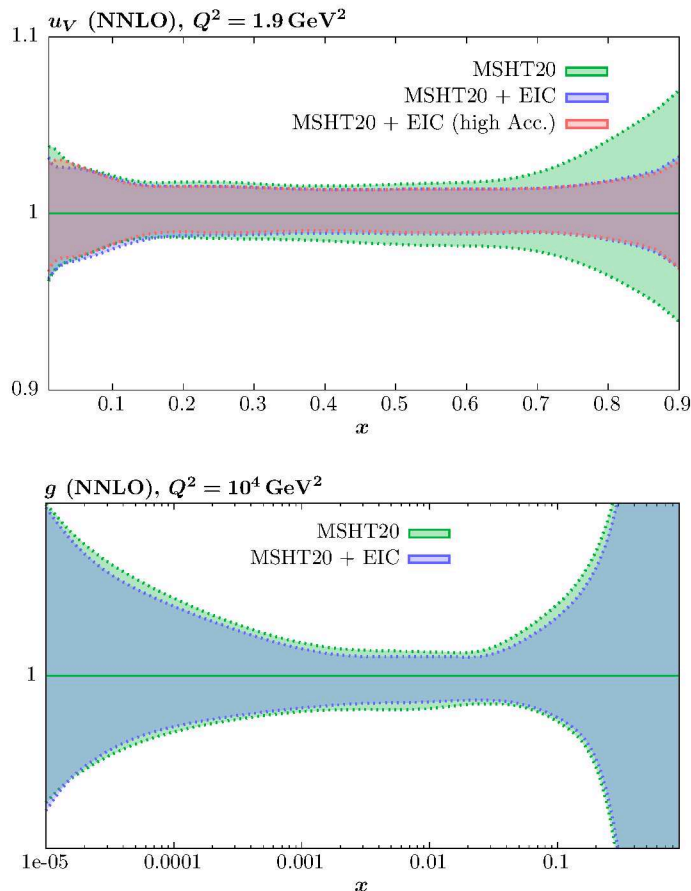


Figure 4: Impact of simulated EIC data on the collinear proton parton distributions relative to the MSHT20 global fits. The bands show relative uncertainties as a function of  $x$ , comparing the MSHT20 baseline with results when additionally including EIC data. Top: up valence density at  $Q^2 = 1.9 \text{ GeV}^2$ , also comparing EIC scenarios with a restriction to  $y > 10^{-2}$  ('MSHT20 + EIC') with the standard requirement  $y > 10^{-3}$  ('MSHT20 + EIC (high Acc.)'). Bottom: gluon density at  $Q^2 = 10^4 \text{ GeV}^2$ .

DIS, up quarks are more strongly impacted than down quarks, such that the biggest impact is on the up-valence distribution, for which a reduction in uncertainties of up to  $\sim 50\%$  is observed in the highest  $x$  region. The relative impact of data in the low  $y$  kinematic region is investigated by comparing the standard selection with results from a dataset restricted to  $y > 10^{-2}$ . The difference in precision on the up-valence distribution is negligible, which is largely due to the overlapping phase space coverages of the different EIC beam energy configurations, removing the need for difficult measurements at extreme low  $y$  values.

The simulated EIC data bring a small, but nonetheless valuable, improvement in the precision on all of the other parton species at all  $x$  and  $Q^2$  values. The gluon density at the electroweak scale is chosen for illustration here; overall uncertainties are relatively small at such large scales due to the constraints from scaling violations in gluon-initiated DIS.

The results shown here are broadly consistent with previous studies of EIC impact relative to global PDF fits [14–18]. Some of these studies have also noted

other smaller impacts in reducing the uncertainties of the strangeness or  $d/u$  ratio at high  $x$ . However there were differences observed between groups and depending on the study [18]. We do not see notable changes in either the strange or  $d/u$  ratio uncertainty at high  $x$ , though our study has only considered proton data and electron beams. The inclusion of deuteron data and/or positron beams may be expected to allow a greater sensitivity to  $d$ -type quarks and may therefore lead to such improvements.

It is worth noting that the more comprehensive datasets included in global PDF fits bring associated complexities and issues that are not necessarily present in the DIS-only case. As well as the more complex theoretical description of hadron-hadron collisions, and the necessity of including non-perturbative input for example in the nuclear corrections to fixed-target data and hadronisation corrections to jet cross sections, there is observed to be a degree of tension between the different datasets that enter the global fit [3, 41, 43–45]. In the MSHT case, this is accounted for by including an enlarged ‘tolerance’ with respect to textbook statistical expectations when evaluating the corresponding PDF uncertainties [46]. The milder impact of the EIC pseudo-data in the MSHT global fit should be considered in this context. In particular, the addition of EIC data may shed light on tensions in other parts of the global fit, giving them value beyond that indicated by the projections shown here.

### 3.3 Parton Luminosities

Any improvement in PDF uncertainties has consequences for phenomenology at the LHC. The results here have therefore been propagated to investigate the influence on the uncertainties on LHC parton luminosities, which encode the precision with which the initial state conditions are known in  $qq$ ,  $gg$ ,  $qg$  and  $q\bar{q}$  sub-process collisions as a function of the parton-parton invariant mass,  $m_X$ . The parton-parton luminosities are calculated as defined in [40].

The impact of the EIC pseudo-data on the parton luminosity uncertainties is shown relative to HERAPDF2.0 in figure 5. The total uncertainties are shown, incorporating the experimental, model and parameterisation components combined in quadrature, as is usual for HERAPDF2.0<sup>5</sup>. All parton luminosities shown have reduced uncertainties following the addition of the EIC pseudo-data. The largest impacts are at high  $m_X$ , reflecting the constraints placed at high  $x$  on the PDFs. The gluon-gluon luminosity shows the largest reduction in uncertainty, with effects of up to approximately  $\sim 50\%$  at larger invariant masses. Nonetheless the uncertainty reductions also extend across the invariant mass range down to low masses. This is as expected given the impact of the pseudo-data across the whole range of  $x$  on the PDFs. The  $q\bar{q}$  and  $qg$  luminosities are similarly reduced in uncertainty down to  $\sim 20$  GeV, though again with smaller effects as the invariant mass is reduced. On the other hand the  $qq$  luminosity uncertainty is largely affected only at high invariant masses.

The impact of the EIC pseudo-data on the MSHT20 parton luminosity uncertainties (figure 6) is significantly milder than that relative to HERAPDF2.0, consistent with the changes seen in the PDF uncertainties. The largest effect is in the

---

<sup>5</sup>Note however that the model dependence uncertainty eigenvectors are used in pairs here in order to avoid non-smooth features in the luminosities.

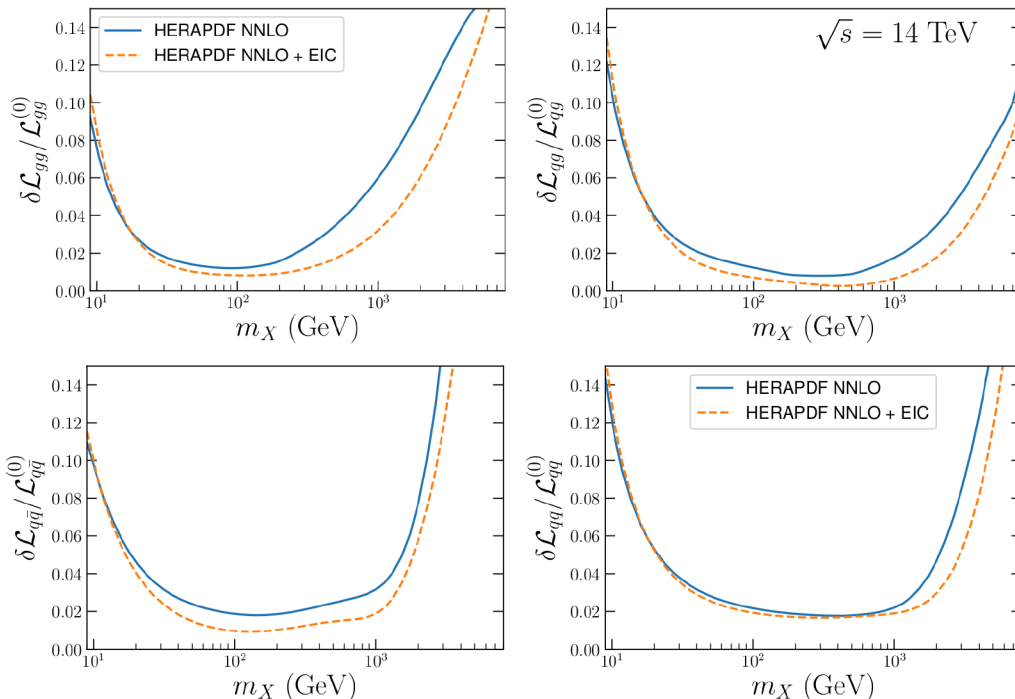


Figure 5: Impact of simulated EIC data on the LHC collinear PDF luminosities relative to the HERAPDF2.0 PDF fit as a function of the parton-parton invariant mass,  $m_X$ . Total uncertainties are included. Upper left: gluon-gluon luminosity. Upper right: quark-gluon luminosity. Lower left: quark-antiquark luminosity. Lower right: quark-quark luminosity.

gluon-gluon luminosity, which shows a small reduction in uncertainty, approximately constant throughout the invariant mass range, reflecting the reduction in the gluon uncertainties across a range of  $x$  values.

A reduction in the gluon-gluon luminosity uncertainty directly affects the precision with which the Higgs production cross-section from gluon-gluon fusion can be predicted [17, 18]. The uncertainty in the  $gg$  luminosity at  $m_H = 125$  GeV reduces from 1.2% to 0.8% with the addition of the EIC pseudodata. The corresponding improvement in precision on the gluon-gluon fusion Higgs production cross section is shown in figure 7. The PDF uncertainty, calculated using `ggHiggs` [47–57] with N<sup>3</sup>LO matrix elements and NNLO PDFs at a central scale of  $\mu_f = \mu_r = m_H/2$  and at  $\sqrt{s} = 13$  TeV, reduces in proportion to the  $gg$  luminosity uncertainty, i.e. also from 1.2% to 0.8%. However, the overall impact on the total uncertainty is much smaller, due to the dominance of the scale uncertainty.

There are finite effects for the other parton-parton luminosity cases presented in figure 6. However, the effects are small, reflecting the relatively small impact of EIC pseudo-data on most of the MSHT20 quark density uncertainties and the integration over rapidities and parton species that must be performed in calculating parton luminosities. For example in the quark-quark case, the notable improvements in the up-valence uncertainties at the highest  $x$  values are diluted when combined with the down and strange densities, leading to a relatively small impact even at large invariant masses.

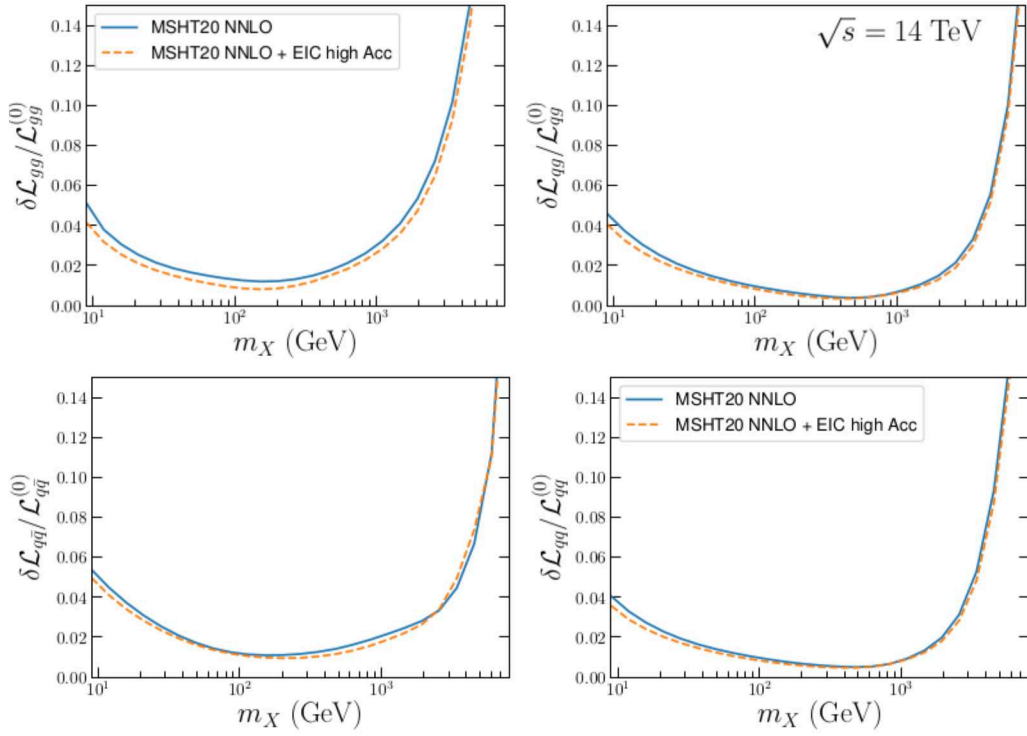


Figure 6: Impact of simulated EIC data on the LHC collinear PDF luminosities relative to the MSHT20 global fits as a function of the parton-parton invariant mass,  $m_X$ . Upper left: gluon-gluon luminosity. Upper right: quark-gluon luminosity. Lower left: quark-antiquark luminosity. Lower right: quark-quark luminosity.

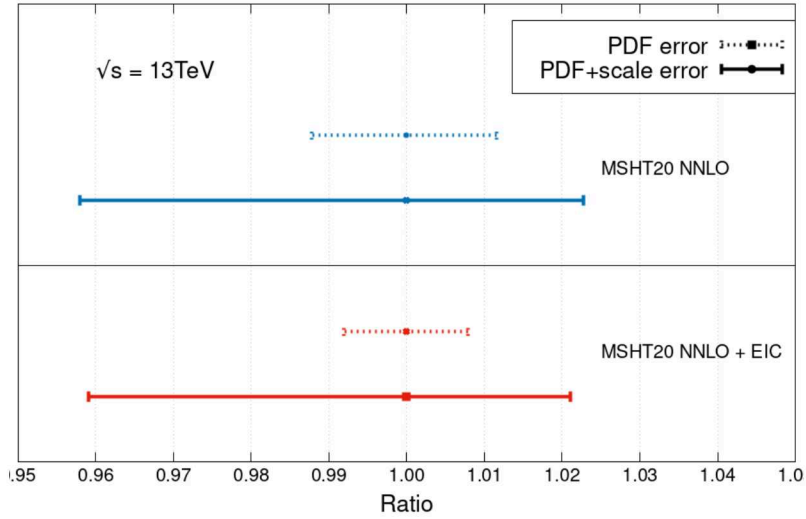


Figure 7: Impact of simulated EIC data on the Higgs production cross section results via gluon fusion (with  $\sqrt{s} = 13$  TeV) at the central scale  $\mu = m_H/2$ . The cross-section is calculated using  $N^3$ LO matrix elements but only NNLO PDFs. The dotted lines indicate the PDF only uncertainties, the solid lines are the PDF+scale uncertainties combined in quadrature, with the scale uncertainties determined by varying  $\mu$  by a factor of 2 following the 9-point prescription.

## 4 Nuclear PDFs

As the world’s first  $eA$  collider, the EIC will explore partonic nuclear structure at an unprecedented level of detail. In particular, it opens up a new region at low  $x$  that has not been constrained previously in DIS or Drell-Yan data. The partonic structure of nuclei is commonly discussed in terms of nuclear PDFs (nPDFs), or nuclear modification ratios, which encode the deviations of nPDFs from simple scaling of free nucleon PDFs with atomic mass  $A$  after appropriately accounting for varying proton-to-neutron ratios using isospin symmetry. The deviations from this scaling with  $A$  may be due to binding effects or, at low  $x$ , to new parton dynamics [58] (‘saturation’ phenomena) associated with the denser systems of gluons found in heavy nuclei than in nucleons.

Present DIS data feeding into nPDFs are limited to fixed target measurements at large  $x$  and relatively low  $Q^2$ . Data from fixed target and colliding mode hadron-nucleus experiments can be used to extend the sensitivity, but with similar associated theoretical difficulties to those discussed in the proton context in Section 3. Since the uncertainties in the nuclear modification factors are large in the low  $x$  region that will be newly explored in DIS, the EIC is expected to have an impact with relatively modest amounts of  $eA$  data.

The potential impact on nuclear PDFs of simulated EIC data has been studied in the xFitter framework [35]. Data from EIC only are used as input to fits in which the PDFs evolve according to the next-to-leading order (NLO) DGLAP equations, with a minimum  $Q^2$  of 3.5 GeV<sup>2</sup> using a parameterisation at the starting scale taken from the HERAPDF2.0 studies<sup>6</sup>. Figures 8, 9 and 10 show the results for the gluon density, the sea up quark density and the sea valence quark density, respectively. The relative precision is shown separately for the proton and for gold nuclei, as well as for their ratio, i.e., the nuclear modification factor<sup>7</sup>. The EIC-only projections for the nuclear modification ratios are compared with the precision of a representative current global fit, EPPS21 [10]<sup>8</sup>, which includes data from fixed target DIS and Drell-Yan experiments, hard processes in  $pA$  collisions at the LHC and  $\pi^0$  data from PHENIX.

The behaviour of gluons at high densities is a cornerstone of the EIC physics programme. Since it is most clearly addressed through nuclear dynamics at low  $x$ , the case of the gluon density (figure 8) is pivotal. The precision obtainable using EIC data only is typically at the 5% level for the proton and at the 10% level for gold. Large improvements over the EPPS21 baseline are observed for all  $x$  values. The precision in the region  $x \sim 0.1$ , where there are plentiful input data in the EPPS21 case, is improved by around a factor of 2 when using EIC data. This is partly due to the use of a  $\Delta\chi^2 = 1$  condition in defining the uncertainty bands in the EIC case, as would be standard when fitting data from a single experiment, compared with  $\Delta\chi^2 = 33$  as applied in the EPPS21 case to account for tensions between the different input data sets.

The largest improvements in the gluon nuclear modification ratio appear in the

---

<sup>6</sup>This lower cut in  $Q^2$  leads to minimum  $x$  values for the pseudodata of 0.0005 and 0.00125 in  $ep$  and  $eA$ , respectively, which are indicated in Figures 8, 9 and 10 by vertical dotted lines.

<sup>7</sup>Regions in which the gluon contribution becomes negative, which only appear at low  $Q^2$  and small  $x$  and are of very little relevance in the EIC kinematics, are removed.

<sup>8</sup>The EPPS21 bands have been computed including both nuclear and proton baseline uncertainties.

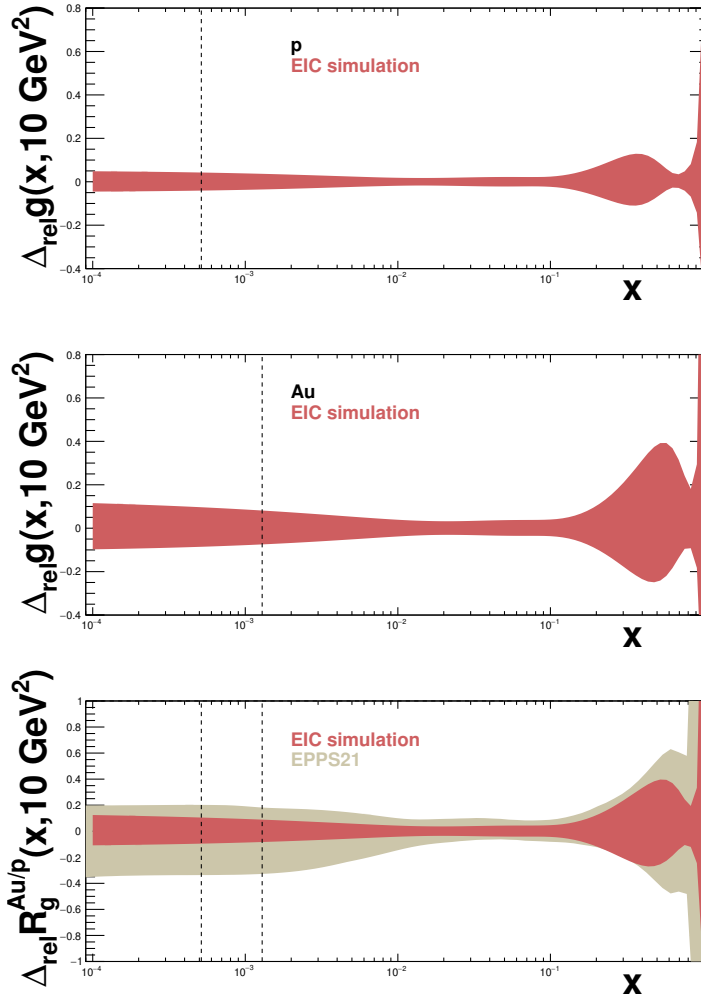


Figure 8: Impact of EIC data on the understanding of nuclear effects in the collinear gluon distribution, as obtained from DGLAP-based QCD fits. Top: projected relative uncertainty on the gluon density of the proton as a function of  $x$  for  $Q^2 = 10 \text{ GeV}^2$ , using only EIC input data. Middle: projected relative uncertainty on the gluon density of a proton in the gold nucleus as a function of  $x$  for  $Q^2 = 10 \text{ GeV}^2$ , using only EIC input data. Bottom: Nuclear modification factors formed from the ratio of projected gluon densities in gold and in the proton. The results obtained using only EIC data are compared with those from a global fit (EPPS21 [10]). Vertical dotted lines indicate the lowest values of  $x$  for pseudo-data used in the fit, see the text.

previously poorly constrained region below  $x \sim 10^{-2}$ . Although the EPPS21 fit accesses  $x$  values as low as  $10^{-5}$ , all information below a few times  $10^{-3}$  comes from D-meson production in  $pPb$  collisions at the LHC, and the uncertainties remain relatively large. The precise and theoretically clean EIC pseudodata extend to approximately  $x = 0.001$ , opening up an order of magnitude of previously unexplored low  $x$  physics in DIS in which novel dynamics may be observed. The gluon nuclear modification ratio is constrained with a precision of approximately 10% in this region. As illustrated in figures 9 and 10, even larger, profound, improvements are

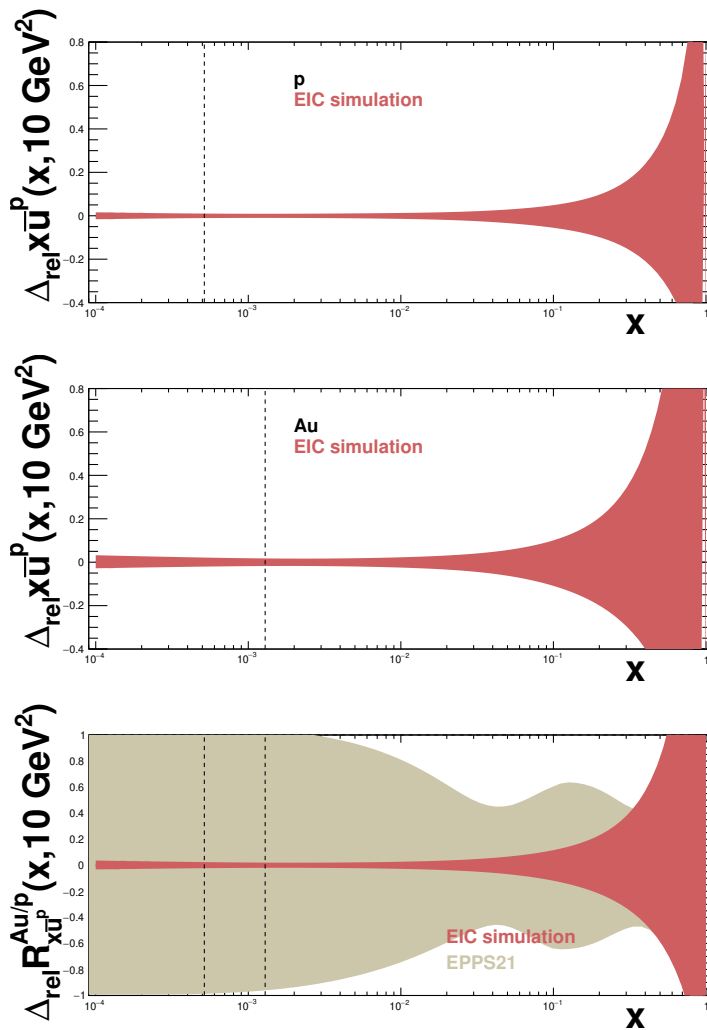


Figure 9: As for figure 8, but for the sea up quark density.

expected for all other parton species.

In addition to EPPS21, new LHC  $pPb$  measurements have also been included in other recent nPDF analyses [9, 11–13]. When compared to a previous analysis such as EPPS16 [59], EPPS21 results in a reduction of the uncertainty in the gluon nuclear modification factor by around a factor 2 at small  $x$ , whereas the reduction is larger in nNNPDF3.0. The variability of the impact of these data between the different analyses suggests that the associated additional theoretical uncertainties and tensions with other data sets have yet to be fully determined. The use of hadron-hadron data in constraining nPDFs also relies on the assumption of collinear factorisation [60] for proton-nucleus collisions, which remains to be fully established. In contrast, the EIC will provide a homogeneous and theoretically straightforward data set from a single experiment and for a single nucleus, with small uncertainties and the possibility of using  $\Delta\chi^2 = 1$  in Hessian analyses to quantify the uncertainties.

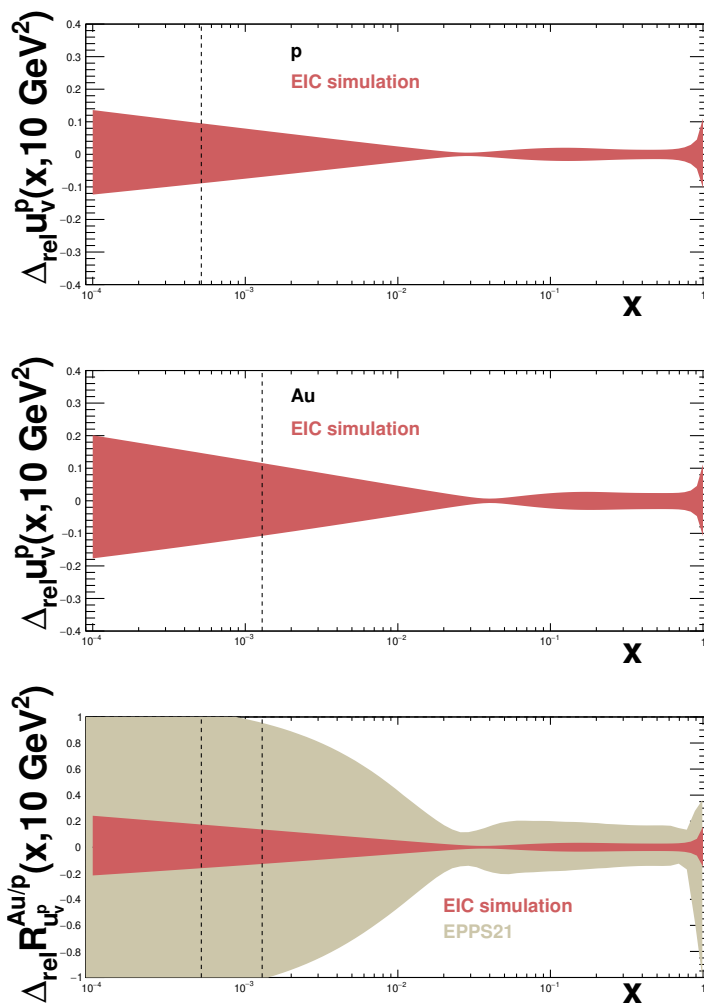


Figure 10: As for figure 8, but for the valence up quark density.

## 5 Summary and conclusions

An analysis studying the impact of simulated EIC inclusive DIS data on the uncertainty of the proton and nuclear PDFs at NNLO and NLO, respectively, is presented.

In the context of a pure-DIS fit for proton PDFs of the HERAPDF2.0 type, the EIC pseudo-data offer orders of magnitude larger integrated luminosities than were available at HERA and cover a lower, but still comfortably perturbative,  $Q^2$  range at large  $x$ . They also optimise the kinematic coverage through variations in the beam energies, avoiding the need for measurements at very low  $y$ . The EIC pseudo-data are thus very well suited to improving the precision at large  $x$  in a pure DIS fit for the proton PDFs. Their most striking impact occurs for the up-valence quark. Precision PDFs in this large  $x$  region are of enormous importance to any hadron-hadron or lepton-hadron experiment with an interest in physics near the  $x \rightarrow 1$  kinematic limit, notably the energy frontier discovery programme at the CERN LHC. These PDF uncertainty reductions at high  $x$  propagate into reductions in the parton-parton luminosity uncertainties, with largest effects at large invariant masses.

As expected, smaller improvements in precision are observed when adding EIC pseudo-data as an additional component of the global MSHT20 PDF fit. Nonetheless, there are visible improvements for several parton species, the up valence quark once again being the most prominent case. There are also consequent impacts on the parton-parton luminosities, though again notably milder than observed for the HERAPDF2.0 part of the study, with the gluon-gluon luminosity uncertainty reduced relatively uniformly across much of the invariant mass range. The latter then enables a reduction in the PDF uncertainty in gluon-gluon fusion Higgs production, albeit the effect on the combined PDF+scale uncertainty band is smaller. EIC data are also expected to influence fits of this type in a wider manner, enabling a theoretically straightforward approach that may help to unravel tensions between other data sets either due to experimental discrepancies or theoretical complications.

The potential EIC impact on the nuclear PDFs is investigated in terms of nuclear modification ratios. For the gluon densities large improvements over an EPPS21 baseline are observed for all  $x$  values, the largest improvements occurring in the region below  $x \sim 10^{-2}$ , which has not previously been constrained by DIS data. Even larger improvements are observed for all other parton species.

Altogether, it is clear that the EIC will significantly improve our understanding of the collinear partonic structure of protons and nuclei, in a manner that is achievable using DIS data alone and thus complements the approach of global fits.

## Acknowledgements

NA has received financial support from Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2019-2022, ref. ED421G-2019/05), by European Union ERDF, by the "María de Maeztu" Units of Excellence program MDM2016-0692, and by the Spanish Research State Agency under project PID2020-119632GB-I00. This work has been performed in the framework of the European Research Council project ERC-2018-ADG-835105 YoctoLHC and the MSCA RISE 823947 "Heavy ion collisions: collectivity and precision in saturation physics" (HIEIC), and has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824093. TC acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE). L. H.-L. and R.S.T. thank STFC for support via grant award ST/T000856/1.

## References

- [1] H. Abramowicz *et al.* [H1 and ZEUS], Eur. Phys. J. C **75**, no.12, 580 (2015) doi:10.1140/epjc/s10052-015-3710-4 [arXiv:1506.06042 [hep-ex]].
- [2] S. Amoroso, A. Apyan, N. Armesto, R. D. Ball, V. Bertone, C. Bissolotti, J. Bluemlein, R. Boughezal, G. Bozzi and D. Britzger, *et al.* Acta Phys. Polon. B **53** (2022) no.12, 1 doi:10.5506/APhysPolB.53.12-A1 [arXiv:2203.13923 [hep-ph]].

- [3] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin and R. S. Thorne, *Eur. Phys. J. C* **81** (2021) no.4, 341 doi:10.1140/epjc/s10052-021-09057-0 [arXiv:2012.04684 [hep-ph]].
- [4] T. J. Hou, J. Gao, T. J. Hobbs, K. Xie, S. Dulat, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin and C. Schmidt, *et al.* *Phys. Rev. D* **103** (2021) no.1, 014013 doi:10.1103/PhysRevD.103.014013 [arXiv:1912.10053 [hep-ph]].
- [5] R. D. Ball *et al.* [NNPDF], *Eur. Phys. J. C* **82** (2022) no.5, 428 doi:10.1140/epjc/s10052-022-10328-7 [arXiv:2109.02653 [hep-ph]].
- [6] S. Alekhin, J. Blümlein, S. Moch and R. Placakyte, *Phys. Rev. D* **96** (2017) no.1, 014011 doi:10.1103/PhysRevD.96.014011 [arXiv:1701.05838 [hep-ph]].
- [7] R. D. Ball *et al.* [PDF4LHC Working Group], *J. Phys. G* **49** (2022) no.8, 080501 doi:10.1088/1361-6471/ac7216 [arXiv:2203.05506 [hep-ph]].
- [8] H. Khanpour, M. Soleymaninia, S. Atashbar Tehrani, H. Spiesberger and V. Guzey, *Phys. Rev. D* **104** (2021) no.3, 034010 doi:10.1103/PhysRevD.104.034010 [arXiv:2010.00555 [hep-ph]].
- [9] I. Helenius, M. Walt and W. Vogelsang, *Phys. Rev. D* **105** (2022) no.9, 9 doi:10.1103/PhysRevD.105.094031 [arXiv:2112.11904 [hep-ph]].
- [10] K. J. Eskola, P. Paakkinen, H. Paukkunen and C. A. Salgado, *Eur. Phys. J. C* **82** (2022) no.5, 413 doi:10.1140/epjc/s10052-022-10359-0 [arXiv:2112.12462 [hep-ph]].
- [11] R. Abdul Khalek, R. Gauld, T. Giani, E. R. Nocera, T. R. Rabemananjara and J. Rojo, *Eur. Phys. J. C* **82** (2022) no.6, 507 doi:10.1140/epjc/s10052-022-10417-7 [arXiv:2201.12363 [hep-ph]].
- [12] P. Duventäster, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K. F. Muzakka, F. I. Olness, R. Ruiz, I. Schienbein and J. Y. Yu, *Phys. Rev. D* **105** (2022) no.11, 114043 doi:10.1103/PhysRevD.105.114043 [arXiv:2204.09982 [hep-ph]].
- [13] P. K. M. Paakkinen, *PoS LHCP2022* (2023), 137 doi:10.22323/1.422.0137 [arXiv:2211.08906 [hep-ph]].
- [14] R. Abdul Khalek *et al.*, *Nucl. Phys. A* **1026**, 122447 (2022) doi:10.1016/j.nuclphysa.2022.122447 [arXiv:2103.05419 [hep-exp]].
- [15] R. A. Khalek, J. J. Ethier, E. R. Nocera and J. Rojo, *Phys. Rev. D* **103** (2021) no.9, 096005 doi:10.1103/PhysRevD.103.096005 [arXiv:2102.00018 [hep-ph]].
- [16] T. J. Hobbs, B. T. Wang, P. M. Nadolsky and F. I. Olness, *Phys. Rev. D* **100** (2019) no.9, 094040 doi:10.1103/PhysRevD.100.094040 [arXiv:1904.00022 [hep-ph]].
- [17] T. J. Hobbs, B. T. Wang, P. M. Nadolsky and F. I. Olness, *PoS DIS2019* (2019), 247 doi:10.22323/1.352.0247 [arXiv:1907.00988 [hep-ph]].

- [18] R. Abdul Khalek, U. D’Alesio, M. Arratia, A. Bacchetta, M. Battaglieri, M. Begel, M. Boggione, R. Boughezal, R. Boussarie and G. Bozzi, *et al.* [arXiv:2203.13199 [hep-ph]].
- [19] E. C. Aschenauer, S. Fazio, M. A. C. Lamont, H. Paukkunen and P. Zurita, *Phys. Rev. D* **96** (2017) no.11, 114005 doi:10.1103/PhysRevD.96.114005 [arXiv:1708.05654 [nucl-ex]].
- [20] R. Abdul Khalek *et al.* [NNPDF], *Eur. Phys. J. C* **79** (2019) no.6, 471 doi:10.1140/epjc/s10052-019-6983-1 [arXiv:1904.00018 [hep-ph]].
- [21] J. Adam *et al.*, *JINST* **17**, 10 (2022) P10019 doi:10.1088/1748-0221/17/10/P10019 [arXiv:2210.09048 [hep-exp]].
- [22] V. N. Gribov and L. N. Lipatov, *Sov. J. Nucl. Phys.* **15** (1972), 438-450 IPTI-381-71.
- [23] V. N. Gribov and L. N. Lipatov, *Sov. J. Nucl. Phys.* **15** (1972), 675-684
- [24] L. N. Lipatov, *Yad. Fiz.* **20** (1974), 181-198
- [25] Y. L. Dokshitzer, *Sov. Phys. JETP* **46** (1977), 641-653
- [26] G. Altarelli and G. Parisi, *Nucl. Phys. B* **126** (1977), 298-318 doi:10.1016/0550-3213(77)90384-4
- [27] S. Moch, J. A. M. Vermaseren and A. Vogt, *Nucl. Phys. B* **688** (2004), 101-134 doi:10.1016/j.nuclphysb.2004.03.030 [arXiv:hep-ph/0403192 [hep-ph]].
- [28] A. Vogt, S. Moch and J. A. M. Vermaseren, *Nucl. Phys. B* **691** (2004), 129-181 doi:10.1016/j.nuclphysb.2004.04.024 [arXiv:hep-ph/0404111 [hep-ph]].
- [29] J. Blümlein, P. Marquard, C. Schneider and K. Schönwald, *Nucl. Phys. B* **971**, 115542 (2021) doi:10.1016/j.nuclphysb.2021.115542 [arXiv:2107.06267 [hep-ph]].
- [30] S. Cerci, Z. S. Demiroglu, A. Deshpande, P. R. Newman, B. Schmookler, D. Sunar Cerci and K. Wichmann, [arXiv:2307.01183 [hep-ph]].
- [31] J.K. Adkins *et al.*, [arXiv:2209.02580 [hep-exp]].
- [32] P.R. Newman Talk at DIS22, Santiago de Compostella, May 2022, <http://epweb2.ph.bham.ac.uk/user/newman/diffraction/DIS22.PDFs.pdf>
- [33] A. C. Benvenuti *et al.* [BCDMS], *Phys. Lett. B* **223** (1989), 485-489 doi:10.1016/0370-2693(89)91637-7
- [34] M. Arneodo *et al.* [New Muon], *Nucl. Phys. B* **483** (1997), 3-43 doi:10.1016/S0550-3213(96)00538-X [arXiv:hep-ph/9610231 [hep-ph]].
- [35] S. Alekhin, O. Behnke, P. Belov, S. Borroni, M. Botje, D. Britzger, S. Camarda, A. M. Cooper-Sarkar, K. Daum and C. Diaconu, *et al.* *Eur. Phys. J. C* **75**, no.7, 304 (2015) doi:10.1140/epjc/s10052-015-3480-z [arXiv:1410.4412 [hep-ph]].

- [36] R. D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo and L. Rottoli, *Eur. Phys. J. C* **78** (2018) no.4, 321 doi:10.1140/epjc/s10052-018-5774-4 [arXiv:1710.05935 [hep-ph]].
- [37] M. Bonvini, S. Marzani and T. Peraro, *Eur. Phys. J. C* **76** (2016) no.11, 597 doi:10.1140/epjc/s10052-016-4445-6 [arXiv:1607.02153 [hep-ph]].
- [38] M. Bonvini, S. Marzani and C. Muselli, *JHEP* **12** (2017), 117 doi:10.1007/JHEP12(2017)117 [arXiv:1708.07510 [hep-ph]].
- [39] M. Bonvini and F. Giuli, *Eur. Phys. J. Plus* **134** (2019) no.10, 531 doi:10.1140/epjp/i2019-12872-x [arXiv:1902.11125 [hep-ph]].
- [40] J. Gao, L. Harland-Lang and J. Rojo, *Phys. Rept.* **742** (2018), 1-121 doi:10.1016/j.physrep.2018.03.002 [arXiv:1709.04922 [hep-ph]].
- [41] X. Jing, A. Cooper-Sarkar, A. Courtoy, T. Cridge, F. Giuli, L. Harland-Lang, T. J. Hobbs, J. Huston, P. Nadolsky and R. S. Thorne, *et al.* [arXiv:2306.03918 [hep-ph]].
- [42] T. Cridge, L. A. Harland-Lang, A. D. Martin and R. S. Thorne, *Eur.Phys.J.C* **82** (2022) no.1, 90 doi:10.1140/epjc/s10052-022-10028-2 [arXiv:2111.05357 [hep-ph]].
- [43] T. Cridge [PDF4LHC21 combination group], *SciPost Phys. Proc.* **8** (2022), 101 doi:10.21468/SciPostPhysProc.8.101 [arXiv:2108.09099 [hep-ph]].
- [44] R. Thorne, S. Bailey, T. Cridge, L. Harland-Lang and A. D. Martin, *SciPost Phys. Proc.* **8** (2022), 018 doi:10.21468/SciPostPhysProc.8.018
- [45] J. McGowan, T. Cridge, L. A. Harland-Lang and R. S. Thorne, *Eur. Phys. J. C* **83** (2023) no.3, 185 [erratum: *Eur. Phys. J. C* **83** (2023) no.4, 302] doi:10.1140/epjc/s10052-023-11236-0 [arXiv:2207.04739 [hep-ph]].
- [46] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Eur. Phys. J. C* **63** (2009), 189-285 doi:10.1140/epjc/s10052-009-1072-5 [arXiv:0901.0002 [hep-ph]].
- [47] R. D. Ball, M. Bonvini, S. Forte, S. Marzani and G. Ridolfi, *Nucl. Phys. B* **874** (2013), 746-772 doi:10.1016/j.nuclphysb.2013.06.012 [arXiv:1303.3590 [hep-ph]].
- [48] M. Bonvini, R. D. Ball, S. Forte, S. Marzani and G. Ridolfi, *J. Phys. G* **41** (2014), 095002 doi:10.1088/0954-3899/41/9/095002 [arXiv:1404.3204 [hep-ph]].
- [49] M. Bonvini, S. Marzani, C. Muselli and L. Rottoli, *JHEP* **08** (2016), 105 doi:10.1007/JHEP08(2016)105 [arXiv:1603.08000 [hep-ph]].
- [50] T. Ahmed, M. C. Kumar, P. Mathews, N. Rana and V. Ravindran, *Eur. Phys. J. C* **76** (2016) no.6, 355 doi:10.1140/epjc/s10052-016-4199-1 [arXiv:1510.02235 [hep-ph]].
- [51] M. Bonvini and S. Marzani, *Phys. Rev. Lett.* **120** (2018) no.20, 202003 doi:10.1103/PhysRevLett.120.202003 [arXiv:1802.07758 [hep-ph]].

- [52] M. Bonvini, Eur. Phys. J. C **78** (2018) no.10, 834 doi:10.1140/epjc/s10052-018-6315-x [arXiv:1805.08785 [hep-ph]].
- [53] M. Bonvini, EPJ Web Conf. **60** (2013), 12008 doi:10.1051/epjconf/20136012008 [arXiv:1306.6633 [hep-ph]].
- [54] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog and B. Mistlberger, JHEP **03** (2015), 091 doi:10.1007/JHEP03(2015)091 [arXiv:1411.3584 [hep-ph]].
- [55] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos and B. Mistlberger, JHEP **05** (2016), 058 doi:10.1007/JHEP05(2016)058 [arXiv:1602.00695 [hep-ph]].
- [56] B. Mistlberger, JHEP **05** (2018), 028 doi:10.1007/JHEP05(2018)028 [arXiv:1802.00833 [hep-ph]].
- [57] M. Bonvini, <https://www.ge.infn.it/~bonvini/higgs/>, version 4.1 (2020).
- [58] N. Armesto, J. Phys. G **32** (2006), R367-R394 doi:10.1088/0954-3899/32/11/R01 [arXiv:hep-ph/0604108 [hep-ph]].
- [59] K. J. Eskola, P. Paakkinen, H. Paukkunen and C. A. Salgado, Eur. Phys. J. C **77** (2017) no.3, 163 doi:10.1140/epjc/s10052-017-4725-9 [arXiv:1612.05741 [hep-ph]].
- [60] J. C. Collins, D. E. Soper and G. F. Sterman, Adv. Ser. Direct. High Energy Phys. **5** (1989), 1-91 doi:10.1142/9789814503266\_0001 [arXiv:hep-ph/0409313 [hep-ph]].