Eur. Phys. J. C (2023) 83:1011
https://doi.org/10.1140/epjc/s10052-023-12176-5

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Extraction of the strong coupling with HERA and EIC inclusive data

Salim Cerci[1,a], **Zuhal Seyma Demiroglu**[2,3], **Abhay Deshpande**[2,3,4], **Paul R. Newman**[5], **Barak Schmookler**[6], **Deniz Sunar Cerci**[1], **Katarzyna Wichmann**[7]

[1] Department of Physics, Faculty of Arts and Sciences, Adiyaman University, Adiyaman, Turkey
[2] Center for Frontiers in Nuclear Science, Stony Brook University, Stony Brook, NY 11764, USA
[3] Stony Brook University, Stony Brook, NY 11794-3800, USA
[4] Brookhaven National Laboratory, Upton, NY 11973-5000, USA
[5] School of Physics and Astronomy, University of Birmingham, Birmingham, UK
[6] Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA
[7] Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607, Hamburg, Germany

Received: 16 July 2023 / Accepted: 21 October 2023

Abstract Sensitivity to the strong coupling $\alpha_S(M_Z^2)$ is investigated using existing Deep Inelastic Scattering data from HERA in combination with projected future measurements from the Electron Ion Collider (EIC) in a next-to-next-to-leading order QCD analysis. A potentially world-leading level of precision is achievable when combining simulated inclusive neutral current EIC data with inclusive charged and neutral current measurements from HERA, with or without the addition of HERA inclusive jet and dijet data. The result can be obtained with substantially less than one year of projected EIC data at the lower end of the EIC centre-of-mass energy range. Some questions remain over the magnitude of uncertainties due to missing higher orders in the theoretical framework.

1 Introduction

Of the coupling strengths of the fundamental forces, the strong coupling α_s is by far the least well constrained. At the same time, it is an essential ingredient of Standard Model cross section calculations, as well as constraints on new physics and grand unification scenarios [1,2]. It has previously been measured in a wide range of processes [3,4]. In Deep Inelastic Scattering (DIS), recent studies from HERA have shown limited sensitivity when using only inclusive data [5], but much more competitive precision when additionally including jet production cross sections [5–7]. In recent years, the advances in QCD theory from next-to-leading order (NLO) to next-to-next-to-leading order (NNLO) have resulted in a substantial reduction in the uncertainties on α_s extractions due to missing higher order corrections, usually expressed in terms of a QCD scale uncertainty, though they remain by far the largest single source of uncertainty in the best HERA extractions.

The Electron Ion Collider (EIC) [8], currently under preparation at Brookhaven National Laboratory in partnership with the Thomas Jefferson National Accelerator Facility is expected to begin taking data around 2030. The EIC will collide highly polarised electrons with highly polarised protons and light/heavy nuclei. In ep mode, the expected luminosity is of order $10^{33} - 10^{34}$ cm^{-2} s^{-1} and the centre-of-mass energy $\sqrt{s}$ will range from 29 GeV to 141 GeV. As part of the extensive program of EIC physics [9], inclusive DIS cross sections will be measured to high precision in a phase space region that will be complementary to HERA, in particular improving the sensitivity to the large Bjorken-x kinematic region. In this work, the expected experimental uncertainty on the strong coupling at the scale of the Z-pole mass $\alpha_s(M_Z^2)$ is estimated when adding simulated EIC inclusive data to analyses very similar to those performed on HERA data. An earlier study of the impact of inclusive EIC data on α_s precision can be found in [9].

[a] e-mail: Salim.Cerci@cern.ch (corresponding author)

Published online: 07 November 2023

2 Analysis method

2.1 Data samples

The HERA data used in this analysis are the final combined H1 and ZEUS inclusive DIS neutral current (NC) and charged current (CC) cross sections [5] and, where appropriate, the H1 and ZEUS inclusive and dijet measurements used in a recent study of parton distribution function (PDF) sensitivity at NNLO, as summarised in Table 1 of [6]. The HERA cross sections correspond to unpolarised beam configurations at proton beam energies of 920, 820, 575 and 460 GeV and an electron beam energy of 27.5 GeV. The data correspond to an integrated luminosity of about 1 fb^{-1} and span six orders of magnitude in the modules of the four-momentum-transfer squared, Q^2, and in Bjorken x.

The detailed experimental apparatus configurations for the EIC are currently under intense development. However, the broad requirements are well established, as documented for example in [9]. In this paper, the simulated EIC data are taken from the studies performed in the framework for ATHENA detector proposal [10]. The ATHENA configuration has since been combined with ECCE [11] in the framework of a new and fast-evolving ePIC design. Whilst the details of the apparatus are different, the overall kinematic range and achievable precision are expected to be very similar.

As summarised in Table 1, neutral current EIC simulated measurements ('pseudodata') are produced with integrated luminosities corresponding to expectations for one year of data taking with each of the five different beam configurations expected at the EIC. Charged current pseudodata are also available at the highest $\sqrt{s}$. The neutral current pseudodata are produced in a grid of five logarithmically-spaced x and Q^2 values per decade over the range[1] $0.001 < y < 0.95$, which is well-justified by the expected resolutions. The central values are taken from predictions using HERAPDF2.0NNLO [5],[2] randomly smeared based on Gaussian distributions with standard deviations given by the projected uncertainties as estimated by the ATHENA collaboration and as previously used to study collinear PDF sensitivities in [10,12]. The systematic precision is based on experience from HERA and further considerations in [9] and is rather conservative in the context of the more modern detector technologies and larger data sets at the EIC. Most data points have a point-to-point uncorrelated systematic uncertainty of 1.9%, extending to 2.75% at the lowest y values. An additional normalisation uncertainty of 3.4% is ascribed, which is taken to be fully correlated between data at each $\sqrt{s}$, and fully uncorrelated between data sets with different $\sqrt{s}$.

[1] Here, y is the usual inelasticity variable, $y = Q^2/(sx)$.

[2] The variant with $\alpha_S(M_Z^2)$ set to 0.116 is taken, most-closely matching the recent HERA NNLO estimation of 0.1156 [6].

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

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

For the purposes of the QCD fits (Sect. 3), 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 [5].

The kinematic plane of the inclusive data used in this analysis is shown in Fig. 1. The EIC pseudodata overlap in their coverage with the HERA data and extend the kinematic reach in the high x, moderate Q^2 region. Their impact at large x is significant since the large x HERA data are relatively imprecise due to their kinematic correlation with large Q^2, the $1/Q^4$ photon propagator term in the cross section and the limited integrated luminosity.

2.2 Theoretical framework and Fitting Procedure

The analysis is based on a QCD fit that follows the HERAPDF [5] theoretical framework, PDF parameterisations and model parameter choices. In the fit, the proton PDFs and $\alpha_S(M_Z^2)$ are constrained simultaneously in a χ^2-minimisation procedure in which the Q^2 evolution is performed according to the NNLO DGLAP evolution equations [13–22]. The xFitter framework [23–25] is used, with the light quark coefficient functions calculated to NNLO as implemented in `QCDNUM` [26]. The `MINUIT` program [27] is used for the minimisation.

The general-mass variable-flavor-number scheme [28,29] is used for the contributions of heavy quarks. The renormalisation and factorisation scales are taken to be $\mu_r = \mu_f = \sqrt{Q^2}$ for the inclusive DIS data, while $\mu_r^2 = \mu_f^2 = Q^2 + p_T^2$ is used for inclusive jet data and $\mu_r^2 = \mu_f^2 = Q^2 + < p_T >_2^2$ for dijets, where $< p_T >_2$ is the average of the transverse momenta of the two jets. The charm and beauty quark masses (M_c, M_b) follow the choices in [5]. The minimum Q^2 of the inclusive data included in the fits is $Q^2_{\rm min} = 3.5$ GeV^2. As well as avoiding complications associated with low Q^2, this requirement also reduces the possible influence of $\ln(1/x)$ resummation [30]. An additional cut is applied on the squared hadronic final state invariant mass, $W^2 = Q^2(1-x)/x > 10$ GeV^2, which removes data points with low Q^2 and high x that are most likely to be influenced by power-like higher

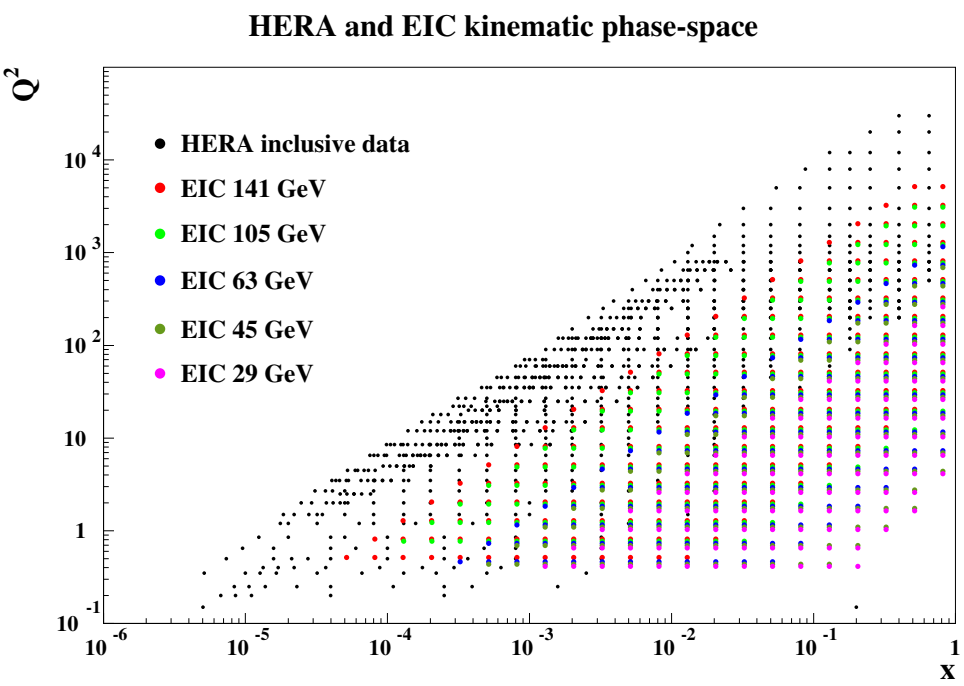

Fig. 1 The locations in the (x, Q^2) plane of the HERA and EIC neutral current inclusive DIS data points included in the analysis

twist or resummation effects. This cut influences the EIC data sets at the lowest $\sqrt{s}$. For the central fit, the PDFs are parameterised at a starting scale for QCD evolution of $\mu_{f0} = 1.9\ \text{GeV}^2$, as in the HERAPDF2.0 fits [5].

The PDFs are parameterised at the starting scale in terms of the gluon distribution (xg), the valence quark distributions (xu_v, xd_v), and the u-type and d-type anti-quark distributions ($x\bar{U}$, $x\bar{D}$), where $x\bar{U} = x\bar{u}$ corresponds to anti-up quarks only and $x\bar{D} = x\bar{d} + x\bar{s}$ is the sum of anti-down and anti-strange quarks. Symmetry is assumed between the sea quarks and antiquarks for each flavour. Strange quarks are suppressed relative to light quarks through a factor $f_s = 0.4$ whereby $x\bar{s} = f_s x\bar{D}$ for all x. The nominal parameterisation is

$$xg(x) = A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{25}; \quad (1)$$

$$xu_v(x) = A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}\left(1 + E_{u_v}x^2\right); \quad (2)$$

$$xd_v(x) = A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}}; \quad (3)$$

$$x\bar{U}(x) = A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}\left(1 + D_{\bar{U}}x\right); \quad (4)$$

$$x\bar{D}(x) = A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}}. \quad (5)$$

The parameters A_{u_v} and A_{d_v} are fixed using the quark counting rules and A_g is fixed using the momentum sum rule. The requirement $x\bar{u} = x\bar{d}$ is imposed as $x \to 0$ through corresponding conditions on $A_{\bar{U}}$, $A_{\bar{D}}$, $B_{\bar{U}}$, $B_{\bar{D}}$ and f_s. There are thus a total of 14 PDF free parameters.

The experimental, model, and parameterisation uncertainties on $\alpha_s(M_Z^2)$ are evaluated as described in [6]. The modelling uncertainties are obtained through variations of Q^2_{min}, f_s, M_c and M_b as shown in Table 2; the parameters are altered independently and the changes relative to the central value

Table 2 Central values of model input parameters and their one-sigma variations. It was not possible to implement the variations marked * because $\mu_{f0} < M_c$ is required, see Ref. [6]. In these cases, the uncertainty on the PDF obtained from the other variation was symmetrised

Parameter	Central value	Downwards variation	Upwards variation
Q^2_{min} (GeV^2)	3.5	2.5	5.0
f_s	0.4	0.3	0.5
M_c (GeV)	1.41	1.37*	1.45
M_b (GeV)	4.20	4.10	4.30

of $\alpha_s(M_Z^2)$ are added in quadrature. For the PDF parameterisation uncertainties, the procedure of [6] is followed, based on variations resulting from the addition of further D and E parameters to the expressions in Eqs. 1–5 and changes in the starting scale μ^2_{f0} by $\pm 0.3\ \text{GeV}^2$. The fits are repeated with each of these variations and the largest difference relative to the nominal $\alpha_s(M_Z^2)$ is taken to be the uncertainty. The model and parameterisation uncertainties are added in quadrature in quoting the final results. For the jet data, the uncertainties in the hadronisation corrections are treated in the same manner as the experimental correlated systematic uncertainties.

The influence on the extracted $\alpha_s(M_Z^2)$ of missing orders in the perturbation series beyond NNLO is estimated via a scale uncertainty, in which the the renormalisation μ_r and factorisation μ_f scales are varied up and down by a factor of two. Combinations are considered in which μ_r and μ_f are changed together or separately and the largest resulting positive and negative deviations on $\alpha_s(M_Z^2)$ (with the exclusion of the two extreme combinations of the scales) is taken

as the scale uncertainty. As is currently customary in global QCD fits,[3] no scale variations are made in the treatment of the inclusive data. This topic is further discussed in Sect. 3.4.2.

3 Results

3.1 Fits with EIC inclusive data and HERA inclusive and jet data

A simultaneous NNLO fit to extract the PDFs and $\alpha_s(M_Z^2)$ from HERA inclusive and jet data and EIC simulated inclusive data at all five $\sqrt{s}$ values is performed as described in Sect. 2. The result is

$$\alpha_s(M_Z^2) = 0.1160 \pm 0.0004\ (\text{exp})\ ^{+0.0003}_{-0.0002} \\ (\text{model} + \text{parameterisation}) \pm 0.0005\ (\text{scale}).$$

By construction of the EIC simulated data, $\alpha_s(M_Z^2)$ must be close to 0.116. As expected, the PDF parameters obtained in the fits are also fully compatible with those from the HERAPDF2.0 set. The uncertainties from the joint fit to HERA and EIC data can be compared with those from the HERAPDF2.0Jets NNLO result [6]:

$$\alpha_s(M_Z^2) = 0.1156 \pm 0.0011\ (\text{exp})\ ^{+0.0001}_{-0.0002} \\ (\text{model} + \text{parameterisation}) \pm 0.0029\ (\text{scale}).$$

The results and uncertainties with and without the inclusion of EIC data are shown in the form of a χ^2 scan as a function of $\alpha_s(M_Z^2)$ in Fig. 2. Each point in the figure corresponds to a full QCD fit, with all 14 PDF parameters free and a fixed strong coupling value. The result without EIC data corresponds exactly to the most recent HERA result [6].

Adding the simulated inclusive EIC data leads to a remarkable improvement in the estimated experimental and scale uncertainties. The source of the improvement in experimental precision is discussed in Sect. 3.4.1. The scale uncertainty is reduced to a similar level to the combined model and parameterisation uncertainties and becomes smaller than the experimental uncertainty. This is a consequence of the reduced dependence of the fit on the jet data. The scale uncertainty is not yet evaluated for the inclusive data, as further discussed in Sect. 3.4.2.

3.2 Fits with EIC and HERA inclusive data only

The very significant impact of the EIC inclusive data on the $\alpha_s(M_Z^2)$ precision naturally raises the question of whether a similar result can be obtained without the HERA jet data, i.e. using only inclusive DIS measurements. A further question of interest is how important a role is played by the multiple $\sqrt{s}$ values available at the EIC. Correspondingly, further fits are performed to the following inclusive data sets with the fit procedures otherwise unchanged:

- HERA inclusive data only, as already published in the HERAPDF2 paper [5];
- HERA inclusive data and the EIC simulated inclusive data described in Sect. 2.1, including all five different $\sqrt{s}$ values in Table 1;
- HERA inclusive data and the EIC simulated inclusive data, separately for each of the five $\sqrt{s}$ values.

Figure 3 shows the results of this investigation. The fits to HERA data alone show only a limited dependence of the fit χ^2 on the strong coupling $\alpha_s(M_Z^2)$, corresponding to a relatively poor constraint [5]. In contrast, the χ^2 minimum around $\alpha_s(M_Z^2) = 0.116$ is very well pronounced for fits that additionally include EIC data. Although the best result is obtained when including all EIC $\sqrt{s}$ values together, the precision degrades only slightly when restricting the EIC data to only one EIC $\sqrt{s}$ value. In the latter case, the precision improves as the $\sqrt{s}$ value of the chosen EIC data decreases. The second lowest $\sqrt{s}$ value, corresponding to $E_e \times E_p = 5 \times 100$ GeV, is shown in Fig. 3.

The strong coupling extracted from the simultaneous fit for the PDFs and $\alpha_s(M_Z^2)$, using the full set of EIC pseudodata together with the HERA inclusive data, is

$$\alpha_s(M_Z^2) = 0.1159 \pm 0.0004\ (\text{exp})\ ^{+0.0002}_{-0.0001} \\ (\text{model} + \text{parameterisation}), \quad (6)$$

corresponding to a total precision of better than 0.4%. As discussed in Sect. 2.2, no scale uncertainty is quoted here. It is expected to be significantly reduced in a fit to inclusive data only relative to the result quoted in Sect. 3.1. Section 3.4.2 contains a discussion of possible ways of estimating the scale uncertainties in this case.

The fit using inclusive data only is further extended to investigate the influence of the integrated luminosity of the EIC data on the $\alpha_s(M_Z^2)$ precision. The statistical uncertainties of the EIC data are scaled such that the pseudodata samples at each beam energy correspond to 1 fb^{-1}, approximately matching the integrated luminosity of the HERA data. This results in only a small change compared with the results shown in Fig. 3.

The relative importance of the different EIC beam energy configurations has also been investigated. When including only a single EIC data set with $\sqrt{s} = 45$ GeV, the experimental uncertainty is approximately ± 0.0010, only slightly more

[3] Scale variations are typically applied to all hadronic final state observables, including jet data from *ep* collisions.

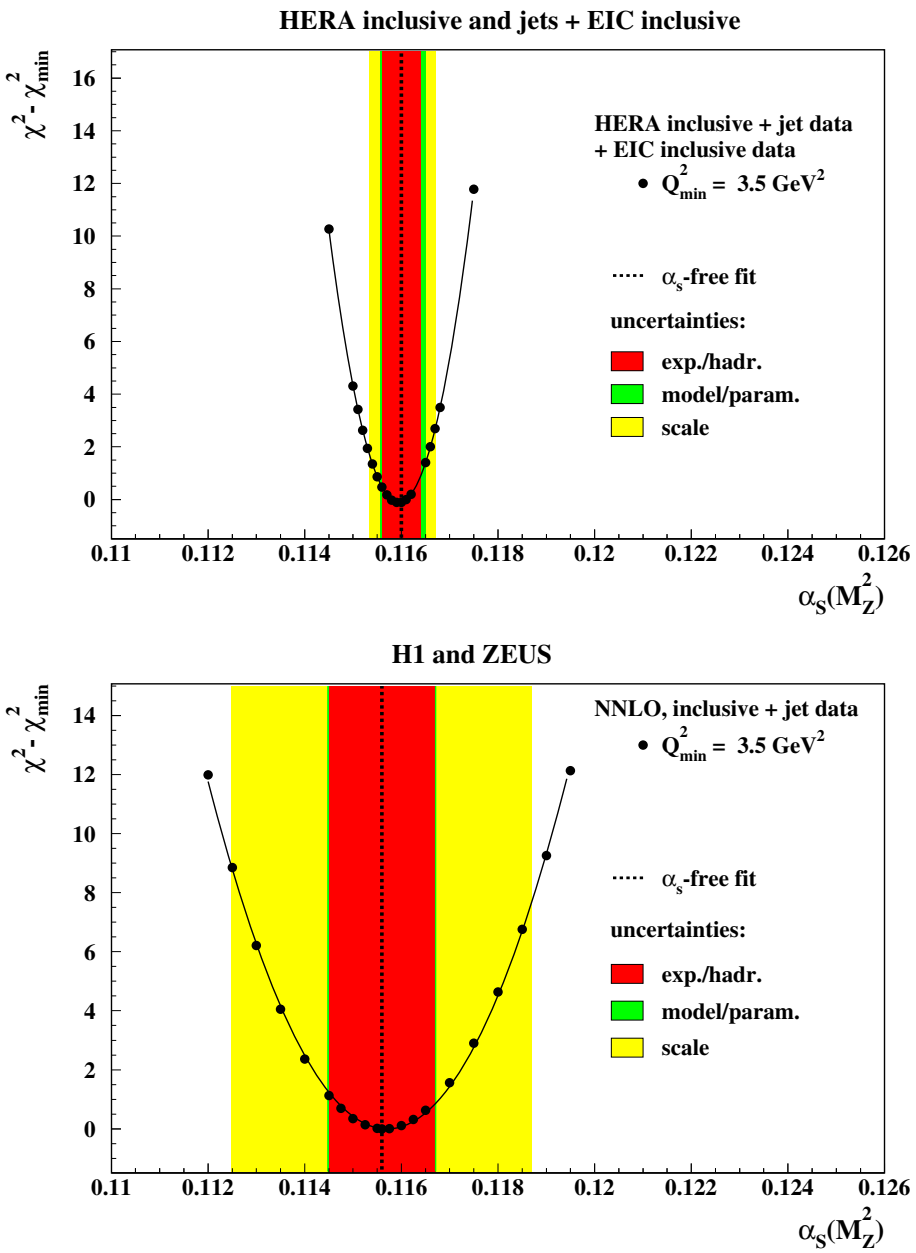

Fig. 2 $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ vs. $\alpha_s(M_Z^2)$ for the NNLO fits to HERA inclusive and jets data in addition to the simulated EIC inclusive data (top) and without the EIC data as published in [6] (bottom). The experimental, model, parameterisation, and scale uncertainties are displayed

than a factor of two larger than that obtained when including all EIC $\sqrt{s}$ values and significantly better than current results from DIS data. Given that the earliest EIC data are expected to be at low $\sqrt{s}$, this result might be obtainable after significantly less than one year of EIC data taking. If only the very lowest energy EIC dataset ($\sqrt{s} = 29$ GeV) is used, the uncertainty grows considerably, due to the influence of the W^2 cut.

To further test the influence of the EIC systematic uncertainty assumptions, the fit is repeated with the correlated systematic uncertainties increased by a factor of two. The uncertainty on the extracted $\alpha_s(M_Z^2)$ is barely influenced. Conversely, if the uncorrelated systematic uncertainties are increased by a factor of two, the uncertainty on $\alpha_s(M_Z^2)$ increases to around 1.7%. The uncorrelated systematic uncertainties are thus the most closely correlated with the precision on $\alpha_s(M_Z^2)$.

3.3 Variations in analysis procedure

The robustness of the extracted $\alpha_s(M_Z^2)$ and PDF results and their uncertainties is tested by varying the details of the fits in a number of ways. The relative sensitivity to α_s of different kinematic regions within the simulated EIC data is also investigated.

To check for a possible bias from the data simulation procedure, the HERA data were replaced with pseudodata obtained using the same method as for the EIC samples.

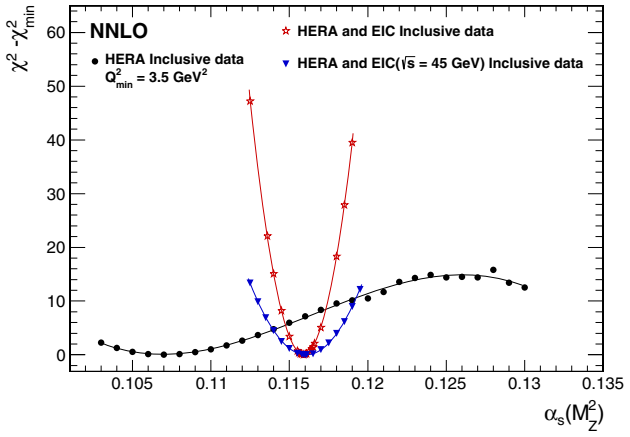

Fig. 3 $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ vs. $\alpha_s(M_Z^2)$ for the NNLO fits to HERA data on inclusive ep scattering only (black), and also with the addition of simulated EIC inclusive data for all five $\sqrt{s}$ values together (red) or for only $\sqrt{s} = 45$ GeV (blue). The black full points are taken from [5]

The $\alpha_s(M_Z^2)$ scan using the HERA pseudodata alone (Fig. 3) closely follows that of the real HERA data, with no distinct minimum observed.

The established technique for including correlated systematic uncertainties in global QCD fits treats each source of correlated uncertainty separately, whereas the EIC estimate is in terms of only a single normalisation uncertainty for each $\sqrt{s}$, corresponding to the sum of all such sources. Studies are therefore conducted in which the correlated EIC systematic uncertainty is decomposed into the separate sources, following Table 10.5 in the EIC Yellow Report [9]. The changes to the results are negligible.

The lowest Q^2 data are most likely to be influenced by missing higher orders, higher twist effects and $\ln(1/x)$ resummation effects [30]. To check that the precision is not dramatically altered by excluding these data, the analysis is repeated with the Q^2_{min} cut increased from 3.5 GeV^2 to 10 GeV^2 or 20 GeV^2. The distinct minima shown in Figs. 2 and 3 are still observed, with only a small dependence (up to 0.2%) on Q^2_{min}. Excluding the lowest x EIC and HERA data such that the analysis is restricted to $x > 0.001$ only increases the uncertainty on the extracted α_s to 0.0005, although precision is lost in the PDF determinations. If all data below $x = 0.01$ are excluded, the precision on α_s remains at a similar level, though the PDF determination becomes over-parameterised, leading to instabilities and biases.

The restriction to $W^2 > 10$ GeV^2 applied here is necessary to avoid theoretical complications associated with higher twists or resummations. It removes data points with the highest x values at low Q^2 for the EIC data sets with the lowest $\sqrt{s}$ values, and has no influence on the largest $\sqrt{s}$ EIC data or the HERA data. When only the lowest energy EIC data $\sqrt{s} = 29$ GeV^2 are included in the fit, a systematic dependence on the W^2 cut is observed, which is diluted when the higher $\sqrt{s}$ data are also included. Nonetheless, in the fit to the full HERA and EIC inclusive data, the experimental uncertainty increases from 0.34 to 0.52% when the restriction is altered to $W^2 > 15$ GeV^2. This kinematic region is therefore observed to be important to the EIC α_s sensitivity, motivating a full understanding of the range of validity of the theoretical framework as W becomes small.

3.4 Discussion

The precision on $\alpha_s(M_Z^2)$ obtained in the fits using only inclusive HERA and EIC data, and also additionally using HERA jet data, are compared in Fig. 4 with results from previous DIS studies and with extractions using a wide range of other processes. The world average of experimental measurements according to the Particle Data Group (PDG) [3] and an average from lattice QCD calculations [31] are also shown. The projected results from the current analyses show a level of precision that is significantly better than both the world average and the lattice average. This very encouraging result is subject to the caveat that no uncertainty has been included due to missing higher orders beyond NNLO in the QCD analysis.

3.4.1 Origin of the EIC sensitivity

The variations in the kinematic range of the fit described in Sect. 3.3 show that the improvement in experimental precision is attributable to the addition of precise EIC pseudodata in the large x, moderate Q^2 region, complementing the kinematic coverage of the HERA data. This additional phase space coverage leads to improved precision on the Q^2 dependence of the inclusive cross section, corresponding to the logarithmic derivative of the inclusive structure function $\mathrm{d}F_2/\mathrm{d}\ln Q^2$. At the highest x values, this quantity is driven primarily by the $q \to qg$ splitting, and therefore samples the product of α_s and the large x quark densities. Since F_2 is itself a measure of the quark densities and other components of the splitting functions are known exactly at a given order, the logarithmic Q^2 derivative at large x essentially depends only on α_s. This contrasts with the scaling violations at lower x, as well as the longitudinal structure function F_L [38,39] and DIS jet data, all of which have leading contributions that are proportional to the product of the gluon parton distribution and α_s [40–42]. The improvement in precision can thus be traced to the decoupling of α_s from the gluon density, enabled by the high x simulated EIC data. This interpretation is supported by the correlation coefficients between α_s and the other free parameters in the fit.

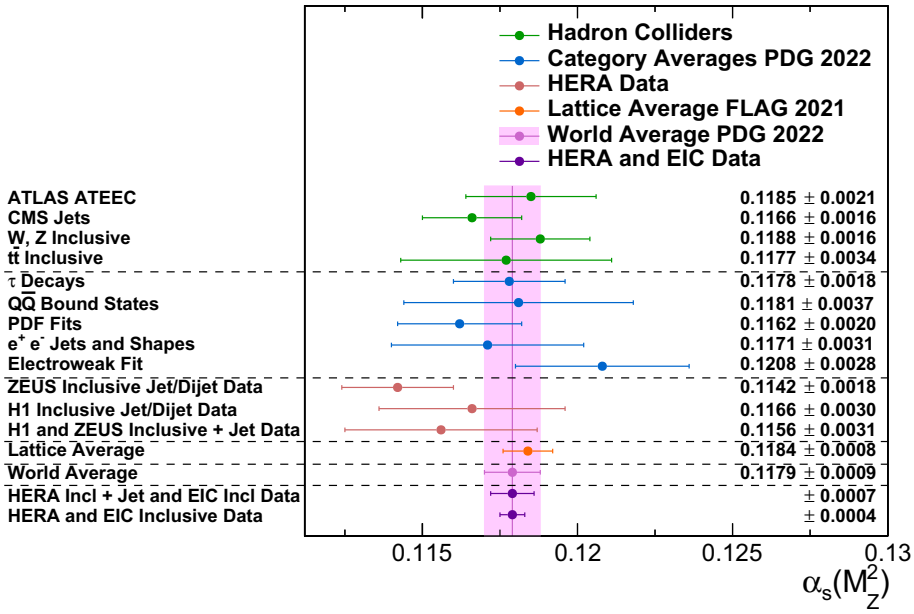

Fig. 4 Projected total uncertainties on the strong coupling constant $\alpha_s(M_Z^2)$ estimated using HERA and simulated EIC data, compared with extractions using other data sets and methods [6,7,32–37], with the world average according to the PDG [3] and with an average from lattice QCD calculations [31]. Scale uncertainties are not yet included in the treatment of inclusive DIS data for any of the results shown. The plotting style follows [32]

3.4.2 Missing higher order uncertainty

Analyses that are sensitive to strong interactions commonly include estimates of the missing higher order uncertainty (MHOU) in the perturbative QCD framework through the variation of the renormalisation and factorisation scales (often referred to as a 'scale uncertainty'). As in many other dedicated and global analyses, the approach used for the jet data included here is to obtain a scale uncertainty by varying the scales by factors of two. However, in the global QCD fits to extract PDFs [43–46], MHOUs have routinely not been included in the treatment of inclusive DIS data, since they are expected to be relatively small in comparison with other PDF uncertainties. Since the present analysis adopts a perturbative QCD treatment at NNLO, it is reasonable to assume that it also has relatively small MHOUs, and the situation is expected to improve further as the state-of-the-art moves towards N^{3}LO [47]. Nevertheless, the MHOU associated with inclusive data must clearly be finite and cannot be ignored completely at the very high level of precision suggested in the present analysis. First studies of their influence on PDFs have been performed by the NNPDF collaboration [30], though the impact on the strong coupling was not included. There is as yet no consensus how to estimate MHOUs for inclusive DIS. Some discussion of possible methods is supplied in the following.

In a previous analysis at NLO accuracy [48], the H1 collaboration made a combined fit of inclusive-only DIS data from HERA and from the fixed target BCDMS experiment. The strong coupling was found to be well-constrained, with the BCDMS data playing a similar role to the EIC pseudodata here. A MHOU was obtained by varying the factorisation and renormalisation scales in the usual way, resulting in a large uncertainty at the level of 4%. However, as shown in the context of PDF uncertainties in [49] (Appendix B), applying this method in an NLO analysis results in an estimate of the MHOU that is larger than the difference between NLO and NNLO results by a factor as large as 20–50.

A similarly conservative approach might be to fit pseudo-data simulated using QCD evolution at NNLO using an NLO framework and vice versa, taking the MHOU on $\alpha_s(M_Z^2)$ to be the deviation of the extracted $\alpha_s(M_Z^2)$ from the input. Applying this method to the present analysis also results in an uncertainty of around 4%, but is also likely to be a very significant over-estimate.

A potentially promising approach is suggested by the NNPDF group [49]. First a theory covariance matrix is computed, typically using scale variations to include missing higher order uncertainties. Including the covariance matrix explicitly in the PDF fit ensures that the theory uncertainties propagate properly, including those associated with $\alpha_s(M_Z^2)$ if it is a fit parameter. However, until a consensus around a well-developed method for including inclusive data such as this emerges, the MHOU in the present $\alpha_s(M_Z^2)$ extraction remains to be evaluated.

4 Conclusions and outlook

This work shows that the strong coupling can be determined with potentially world-leading precision in a simultaneous fit of PDFs and $\alpha_s(M_Z^2)$ at NNLO in perturbative QCD, using only inclusive DIS data from HERA and simulated data from the EIC. The estimated uncertainty on the strong coupling when including one year's data at each of the five expected EIC $\sqrt{s}$ values is better than 0.4%, substantially improving

on the precision of the present world experimental and lattice averages. If the EIC pseudodata are restricted to a small fraction of a standard expected year of running at a centre-of-mass energy of 45 GeV, as expected in an early phase of operation, the estimated total uncertainty is at the level of 0.9%. The improvement in precision is traceable to the large x, intermediate Q^2 region that was not accessed at HERA, but is well covered by the EIC. The constraint arises primarily from the evolution of the quark densities in this region and is largely decoupled from the uncertainty on the gluon density. It still remains to assign a meaningful uncertainty due to missing higher order contributions beyond NNLO in the theory.

Further improvements of the $\alpha_s(M_Z^2)$ precision may be obtainable by adding inclusive jet and dijet EIC data to the QCD analysis, for example using theory grids for the EIC energies in the fastNLO framework [50]. Other observables carrying information on the strong coupling that may be measured at the EIC include event shapes, jet substructure and jet radius parameters. As well as a DIS-only approach, it would also be interesting to investigate the impact of EIC data on α_s determinations in global QCD fits that also include data from the LHC and elsewhere [43–46].

In the time before the start of the EIC, it is hoped that new light will be shed on the issue of higher order uncertainties, leading to a consensus on how they should be treated in $\alpha_s(M_Z^2)$ determinations relying on EIC data.

Acknowledgements We are grateful to many colleagues in the EIC experimental community who have worked on all aspects of the project over many years and are now developing detector concepts towards the acquisition of real data similar to those simulated here. We thank numerous theoretical physics colleagues for very valuable discussions about the theory uncertainties: Néstor Armesto, Andrea Barontini, Thomas Cridge, Stefano Forte, Lucian Harland-Lang, Anna M. Staśto and Robert S. Thorne, as well as Valerio Bertone and Francesco Giuli for their help with the `APFEL` program and Christopher Schwan for his help with the `PineAPPL` tool. This work of Z.S.D and A.D. was supported in part by the U.S. Department of Energy and the Simons Foundation.

Data availability This manuscript has no associated data or the data will not be deposited. [Authors' comment: In the manuscript, the data from H1 and ZEUS collaborations in Table 1 of Ref. [6] were used. https://doi.org/10.1140/epjc/s10052-015-3710-4 and https://doi.org/10.1140/epjc/s10052-022-10083-9 were used. These two articles were cited in the Ref. [5] and [6]. Furthermore, the data from ATHENA with https://doi.org/10.1088/1748-0221/17/10/P10019 were also used.]

Funded by SCOAP^3. SCOAP^3 supports the goals of the International Year of Basic Sciences for Sustainable Development.

References

1. H. Georgi, S.L. Glashow, Unity of all elementary particle forces. Phys. Rev. Lett. **32**, 438 (1974). https://doi.org/10.1103/PhysRevLett.32.438
2. S. Dimopoulos, S. Raby, F. Wilczek, Supersymmetry and the scale of unification. Phys. Rev. D **24**, 1681 (1981). https://doi.org/10.1103/PhysRevD.24.1681
3. Particle Data Group collaboration, Review of Particle Physics. PTEP **2022**, 083C01 (2022). https://doi.org/10.1093/ptep/ptac097
4. D. d'Enterria et al., The strong coupling constant: state of the art and the decade ahead (Snowmass 2021 White Paper). arXiv:2203.08271
5. H1, ZEUS collaboration, Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C **75**, 580 (2015). https://doi.org/10.1140/epjc/s10052-015-3710-4. arXiv:1506.06042
6. H1, ZEUS collaboration, Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions. Eur. Phys. J. C **82**, 243 (2022). https://doi.org/10.1140/epjc/s10052-022-10083-9. arXiv:2112.01120
7. H1 collaboration, Determination of the strong coupling constant $\alpha_s(m_Z)$ in next-to-next-to-leading order QCD using H1 jet cross section measurements. Eur. Phys. J. C **77**, 791 (2017). https://doi.org/10.1140/epjc/s10052-017-5314-7. arXiv:1709.07251 [Erratum: Eur. Phys. J .C **81**, 738 (2021)]
8. A. Accardi et al., Electron Ion Collider: the Next QCD Frontier: Understanding the glue that binds us all. Eur. Phys. J. A **52**, 268 (2016). https://doi.org/10.1140/epja/i2016-16268-9. arXiv:1212.1701
9. R. Abdul Khalek et al., Science requirements and detector concepts for the electron-ion collider: EIC Yellow Report. Nucl. Phys. A **1026**, 122447 (2022). https://doi.org/10.1016/j.nuclphysa.2022.122447. arXiv:2103.05419
10. ATHENA collaboration, ATHENA detector proposal—a totally hermetic electron nucleus apparatus proposed for IP6 at the Electron-Ion Collider. JINST **17**, P10019 (2022). https://doi.org/10.1088/1748-0221/17/10/P10019. arXiv:2210.09048
11. J.K. Adkins et al., Design of the ECCE detector for the Electron Ion Collider. arXiv:2209.02580 [physics.ins-det]
12. P. Newman et al., Proton and Nuclear Collinear Parton Densities at the Electron-Ion Collider using simulated ATHENA data, Talk at DIS22, Santiago de Compostella, (May 2022). http://epweb2.ph.bham.ac.uk/user/newman/diffraction/DIS22.PDFs.pdf
13. V.N. Gribov, L.N. Lipatov, Deep inelastic e p scattering in perturbation theory. Sov. J. Nucl. Phys. **15**, 438 (1972)
14. V.N. Gribov, L.N. Lipatov, e+ e- pair annihilation and deep inelastic e p scattering in perturbation theory. Sov. J. Nucl. Phys. **15**, 675 (1972)
15. L.N. Lipatov, The parton model and perturbation theory. Yad. Fiz. **20**, 181 (1974)
16. Y.L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and e+ e- annihilation by perturbation theory in quantum chromodynamics. Sov. Phys. JETP **46**, 641 (1977)
17. G. Altarelli, G. Parisi, Asymptotic freedom in Parton language. Nucl. Phys. B **126**, 298 (1977). https://doi.org/10.1016/0550-3213(77)90384-4

18. S. Moch, J.A.M. Vermaseren, A. Vogt, The three loop splitting functions in QCD: the Nonsinglet case. Nucl. Phys. B **688**, 101 (2004). https://doi.org/10.1016/j.nuclphysb.2004.03.030. arXiv:hep-ph/0403192
19. A. Vogt, S. Moch, J.A.M. Vermaseren, The Three-loop splitting functions in QCD: the singlet case. Nucl. Phys. B **691**, 129 (2004). https://doi.org/10.1016/j.nuclphysb.2004.04.024. arXiv:hep-ph/0404111
20. A.A. Almasy, S. Moch, A. Vogt, On the next-to-next-to-leading order evolution of flavour-singlet fragmentation functions. Nucl. Phys. B **854**, 133 (2012). https://doi.org/10.1016/j.nuclphysb.2011.08.028. arXiv:1107.2263
21. A. Mitov, S. Moch, A. Vogt, Next-to-next-to-leading order evolution of non-singlet fragmentation functions. Phys. Lett. B **638**, 61 (2006). https://doi.org/10.1016/j.physletb.2006.05.005. arXiv:hep-ph/0604053
22. J. Blümlein, P. Marquard, C. Schneider, K. Schönwald, The three-loop unpolarized and polarized non-singlet anomalous dimensions from off shell operator matrix elements. Nucl. Phys. B **971**, 115542 (2021). https://doi.org/10.1016/j.nuclphysb.2021.115542. arXiv:2107.06267
23. S. Alekhin et al., HERAFitter. Eur. Phys. J. C **75**, 304 (2015). https://doi.org/10.1140/epjc/s10052-015-3480-z. arXiv:1410.4412
24. H1, ZEUS collaboration, Combined measurement and QCD analysis of the inclusive e+- p scattering cross sections at HERA. JHEP **01**, 109 (2010). https://doi.org/10.1007/JHEP01(2010)109. arXiv:0911.0884
25. H1 collaboration, A precision measurement of the inclusive ep scattering cross section at HERA. Eur. Phys. J. C **64**, 561 (2009). https://doi.org/10.1140/epjc/s10052-009-1169-x. arXiv:0904.3513
26. M. Botje, QCDNUM: fast QCD evolution and convolution. Comput. Phys. Commun. **182**, 490 (2011). https://doi.org/10.1016/j.cpc.2010.10.020. arXiv:1005.1481
27. F. James, M. Roos, Minuit: a system for function minimization and analysis of the parameter errors and correlations. Comput. Phys. Commun. **10**, 343 (1975). https://doi.org/10.1016/0010-4655(75)90039-9
28. R.S. Thorne, R.G. Roberts, An Ordered analysis of heavy flavor production in deep inelastic scattering. Phys. Rev. D **57**, 6871 (1998). https://doi.org/10.1103/PhysRevD.57.6871. arXiv:hep-ph/9709442
29. R.S. Thorne, A variable-flavor number scheme for NNLO. Phys. Rev. D **73**, 054019 (2006). https://doi.org/10.1103/PhysRevD.73.054019. arXiv:hep-ph/0601245
30. R.D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo, L. Rottoli, Parton distributions with small-x resummation: evidence for BFKL dynamics in HERA data. Eur. Phys. J. C **78**, 321 (2018). https://doi.org/10.1140/epjc/s10052-018-5774-4. arXiv:1710.05935
31. Flavour Lattice Averaging Group (FLAG) collaboration, FLAG review 2021. Eur. Phys. J. C **82**, 869 (2022). https://doi.org/10.1140/epjc/s10052-022-10536-1. arXiv:2111.09849
32. ATLAS collaboration, A precise determination of the strong-coupling constant from the recoil of Z bosons with the ATLAS experiment at $\sqrt{s} = 8$ TeV, in ATLAS CONF Note (2023). https://cds.cern.ch/record/2855244
33. CMS collaboration, Measurement and QCD analysis of double-differential inclusive jet cross sections in proton–proton collisions at $\sqrt{s}$ = 13 TeV. JHEP **02**, 142 (2022). https://doi.org/10.1007/JHEP02(2022)142. arXiv:2111.10431
34. ATLAS collaboration, Determination of the strong coupling constant from transverse energy−energy correlations in multijet events at $\sqrt{s} = 13$ TeV with the ATLAS detector. JHEP **07**, 085 (2023). https://doi.org/10.1007/JHEP07(2023)085. arXiv:2301.09351
35. T. Klijnsma, S. Bethke, G. Dissertori, G.P. Salam, Determination of the strong coupling constant $\alpha_s(m_Z)$ from measurements of the total cross section for top-antitop quark production. Eur. Phys. J. C **77**, 778 (2017). https://doi.org/10.1140/epjc/s10052-017-5340-5. arXiv:1708.07495
36. D. d'Enterria, A. Poldaru, Extraction of the strong coupling $\alpha_s(\mathrm{m}_Z)$ from a combined NNLO analysis of inclusive electroweak boson cross sections at hadron colliders. JHEP **06**, 016 (2020). https://doi.org/10.1007/JHEP06(2020)016. arXiv:1912.11733
37. ZEUS collaboration, Measurement of jet production in deep inelastic scattering and NNLO determination of the strong coupling at ZEUS. arXiv:2309.02889, accepted by EPJC
38. H1 collaboration, Measurement of inclusive ep cross sections at high Q^2 at $\sqrt{s} = 225$ and 252 GeV and of the longitudinal proton structure function F_L at HERA. Eur. Phys. J. C **74**, 2814 (2014). https://doi.org/10.1140/epjc/s10052-014-2814-6. arXiv:1312.4821
39. ZEUS collaboration, Deep inelastic cross-section measurements at large y with the ZEUS detector at HERA. Phys. Rev. D **90**, 072002 (2014). https://doi.org/10.1103/PhysRevD.90.072002. arXiv:1404.6376
40. A.M. Cooper-Sarkar, G. Ingelman, K.R. Long, R.G. Roberts, D.H. Saxon, Measurement of the longitudinal structure function and the small X gluon density of the proton. Z. Phys. C **39**, 281 (1988). https://doi.org/10.1007/BF01551005
41. E.B. Zijlstra, W.L. van Neerven, Order alpha-s**2 QCD corrections to the deep inelastic proton structure functions F2 and F(L). Nucl. Phys. B **383**, 525 (1992). https://doi.org/10.1016/0550-3213(92)90087-R
42. G.R. Boroun, B. Rezaei, Analysis of the proton longitudinal structure function from the gluon distribution function. Eur. Phys. J. C **72**, 2221 (2012). https://doi.org/10.1140/epjc/s10052-012-2221-9. arXiv:1401.7804
43. NNPDF collaboration, The path to proton structure at 1% accuracy. Eur. Phys. J. C **82**, 428 (2022). https://doi.org/10.1140/epjc/s10052-022-10328-7. arXiv:2109.02653
44. S. Bailey, T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs. Eur. Phys. J. C **81**, 341 (2021). https://doi.org/10.1140/epjc/s10052-021-09057-0. arXiv:2012.04684
45. T.-J. Hou et al., New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC. Phys. Rev. D **103**, 014013 (2021). https://doi.org/10.1103/PhysRevD.103.014013. arXiv:1912.10053
46. S. Alekhin, J. Blümlein, S. Moch, R. Placakyte, Parton distribution functions, α_s, and heavy-quark masses for LHC Run II. Phys. Rev. D **96**, 014011 (2017). https://doi.org/10.1103/PhysRevD.96.014011. arXiv:1701.05838
47. J. McGowan, T. Cridge, L.A. Harland-Lang, R.S. Thorne, Approximate $\mathrm{N^3LO}$ parton distribution functions with theoretical uncertainties: $\mathrm{MSHT20aN^3LO}$ PDFs. Eur. Phys. J. C **83**, 185 (2023). https://doi.org/10.1140/epjc/s10052-023-11236-0. arXiv:2207.04739
48. H1 collaboration, Deep inelastic inclusive e p scattering at low x and a determination of alpha(s). Eur. Phys. J. C **21**, 33 (2001). https://doi.org/10.1007/s100520100720. arXiv:hep-ex/0012053
49. NNPDF collaboration, Parton Distributions with Theory Uncertainties: general formalism and first phenomenological studies. Eur. Phys. J. C **79**, 931 (2019). https://doi.org/10.1140/epjc/s10052-019-7401-4. arXiv:1906.10698
50. T. Kluge, K. Rabbertz, M. Wobisch, FastNLO: fast pQCD calculations for PDF fits, in 14th International Workshop on Deep Inelastic Scattering, 9 (2006), p. 483–486. https://doi.org/10.1142/9789812706706_0110. arXiv:hep-ph/0609285