THE EUROPEAN
PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

The impact of inclusive electron ion collider data on the strong coupling determination in a global PDF fit

L. A. Harland-Lang[1,a] (iD), T. Cridge[2,3] (iD), P. Newman[4] (iD), R. S. Thorne[1] (iD), K. Wichmann[5]

[1] Department of Physics and Astronomy, University College London, London WC1E 6BT, UK
[2] Elementary Particle Physics, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium
[3] Department of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK
[4] School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK
[5] Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

Abstract We present a study of the impact of data from the upcoming Electron Ion Collider (EIC) on the determination of the strong coupling within the context of the global MSHT fitting framework. To achieve this, we generate EIC electron-proton scattering pseudodata according to both conservative and optimistic experimental uncertainty projections and perform a simultaneous fit to obtain the proton PDFs and the value of the strong coupling. In the conservative case the impact is found to be moderate, but non-negligible, while in the optimistic case it is observed to be rather significant. These results therefore underline the promising potential for the EIC in the determination of the strong coupling. We in addition explore the impact of any potential tensions between the EIC data and the rest of the data in the global fit by injecting explicit inconsistencies into the pseudodata generation. This can lead to a noticeable bias in the extracted value of the strong coupling, highlighting the importance of accounting for all sources of theoretical uncertainty in the fit as well as the relevance of an enlarged, conservative, error definition in the determination of the strong coupling.

1 Introduction

Parton distribution functions (PDFs) are an essential part of any theory prediction for hard hadronic interactions, and are therefore key to our understanding of data from collider experiments. There has been significant progress in extracting PDFs as accurately and precisely as possible from different groups, either via global PDF fits [1–4] or more dedicated extractions using data from HERA and the LHC [5,6]. Due to the high precision requirements of such fits, next-to-next-to leading order (NNLO) in the QCD perturbative expansion is now the default, but both MSHT [7] and NNPDF [8] have presented fits at approximate N^3LO (aN^3LO).

Given the inherently QCD nature of the PDFs, and the interactions underlying the processes that enter these fits, as well as being sensitive to the PDFs, they are also in general sensitive the value of the strong coupling. As such, multiple groups have presented extractions of the value of the strong coupling, most recently in [3,4,9,10]. It is of particular note that the quoted uncertainty in these studies is of the same order as, if generally slightly larger than, the PDG average [11], see also [12] for a review.

The Electron-Ion Collider [13] (EIC), which is expected to begin science operations at Brookhaven National Laboratory in the mid 2030 s, is the next particle accelerator that will be able to explore the internal structure of hadrons. Intended to operate at high luminosities and with a wide range of beam energy configurations, it promises to revolutionise our understanding of the QCD. The impact of the EIC on proton PDFs has been studied previously in [14–22], while the potential for a precise determination of the value of the strong coupling has been studied in [23,24]. However, as of yet there has been no analysis of the EIC sensitivity to the strong coupling within the context of a global PDF fit. This arguably provides the most appropriate framework to assess the potential impact of EIC data in the determination of the strong coupling, and in particular its role in the context of the other constraints that enter a global fit, both on the PDFs and strong coupling. In this paper we present such a study, specifically in the context of the MSHT global PDF fitting framework, at both NNLO and aN^3LO in QCD. We in particular evaluate the expected impact from inclusive DIS data from the EIC, by generating

[a] e-mail: l.harland-lang@ucl.ac.uk (corresponding author)

Springer

pseudodata with both optimistic and conservative scenarios for the projected experimental uncertainties.

Studies of this type usually make the implicit assumption of perfect consistency between all of the datasets included. However, in practice we typically observe multiple tensions between datasets in global PDF fits [1,3,9,25–27], which can arise not only due to inconsistencies in newly included data, but also due to theoretical effects or simply issues in previously included data in the global fit as whole. With this in mind, tensions may be expected to exist between the new EIC data when they become available and the rest of the data included in a global fit at that time, without this necessarily reflecting any incompleteness in the experimental determination of the EIC data. These tensions can moderate or alter the impact of the new data in constraining the PDFs and the strong coupling, and so result in important effects that ought to be accounted for. In this work we therefore also analyse the potential impact of tensions between new EIC data and the rest of the data entering the global fit. We achieve this by considering EIC pseudodata where explicit inconsistencies with respect to the baseline fit are built in.

In more detail, we consider three different scenarios. In the first case the pseudodata are generated using aN^{3}LO precision, but the fit is performed at NNLO, which effectively models the impact of missing higher order corrections on the result, and hence acts as a measure of the theoretical uncertainty at NNLO. In the second case, we perform both the EIC pseudodata production and the fit at aN^{3}LO, but generate the pseudodata with target mass corrections included, whilst excluding the corrections from the fit. In the final case, we generate the pseudodata with the NNPDF4.0aN^{3}LO [8] set, which exhibits some tension with the result of the MSHT20 fit to the non-EIC data. The second and third cases emulate the possibility of some tension, due to missing theory ingredients in the calculation or some other unspecified source between the EIC data and the rest of the data in the fit. Such tensions motivate the use of more conservative, enlarged error definitions, similarly to the situation in existing PDF fits, for example in the previous determinations of the strong coupling by the MSHT group, most recently in [9].

The outline of this paper is as follows. In Sect. 2 we present an overview of the pseudodata generation and PDF fitting framework; in Sect. 3 we present our results at both NNLO and aN^{3}LO in Sects. 3.1 and 3.3, respectively, while in Sects. 3.2 and 3.4 we assess the impact of including explicit inconsistencies between the EIC pseudodata and the baseline PDF fit, at NNLO and aN^{3}LO, respectively; finally, in Sect. 4, we conclude.

2 Fit overview

The pseudodata are by default generated exactly as in [21] (see also [23,28]), corresponding to the most up-to-date projections for the EIC beam energy configurations over its lifetime, with expected luminosities for 1 year of running at each. In particular, both charged and neutral current ep scattering data are generated, for a range of beam energy configurations. In addition to the statistical uncertainty, the pseudodata are generated with a point-to-point uncorrelated systematic uncertainty of 1.9%, extending to 2.75% for $y < 0.01$, and a normalization uncertainty of 3.4% is taken. The latter is assumed to be fully correlated between data with the same $\sqrt{s}$ but is uncorrelated between data with different $\sqrt{s}$. However, similarly to the study of [28] we also consider an 'optimistic' scenario, with rather smaller experimental uncertainties. In particular, for this we take a 0.7% normalization uncertainty that is fully correlated across bins within beam energies but not across beam energies and a 0.7% bin-by-bin uncorrelated uncertainty, to be compared with 3.4 and 1.9% respectively in the default case. We note that only the above uncertainty sources are included in the smearing of the pseudodata, with in particular no PDF uncertainty being included, as is standard.

The pseudodata are generated using PDF sets that closely correspond to the result of fitting to the non-EIC data in the fit, unless otherwise stated, and with $\alpha_S(M_Z^2) = 0.118$. It would be interesting to study of varying the value of the strong coupling used here, but this is left to future work. For the aN^{3}LO results, we use the same basic framework as in the original MSHT20aN^{3}LO fit [7], but we now use the more up-to-date splitting functions [29–34] that have become available following the original MSHT20aN^{3}LO fit, as described in [35].

The rest of the datasets in the fit correspond to the real experimental data, as implemented in the latest MSHT fits, see [36]. Following this study, we have updated the treatment of various fixed target datasets. In particular, we now include a more precise account of the correlated systematic uncertainties and separation of the datasets into different beam energies. The updated datasets are the BCDMS [37,38], NMC [39,40] and E665 [41] proton and deuteron data, and NuTeV $F_{2,3}$ structure function data [42].

3 Results

3.1 NNLO QCD fits

We begin by considering fits performed at NNLO in the QCD perturbative expansion. The best fit values of the strong cou-

pling, and an estimate of their uncertainties, for a range of fits that we perform are shown in Fig. 1. The top and bottom plots show the result of the globally and locally preferred values, respectively. The inner (outer) limits correspond to $\Delta\chi^2 = 1$ (9), that result from a quadratic fit about the minimum to the corresponding global profile. These are intended to be indicative of the relative uncertainty on $\alpha_S(M_Z^2)$ in different fits, rather than indicating the actual size of the uncertainty that would result from the dynamic tolerance procedure that is used by the MSHT collaboration, as described in [43]. To evaluate these, and the corresponding best fit value, a scan over fixed values of $\alpha_S(M_Z^2)$, spaced by 0.001, is performed in each case and the three points closest to the minimum are used for a quadratic fit to the $\Delta\chi^2$ profile. Here, the $T = 1$ uncertainty band indicates the size of uncertainty that would result if a textbook uncertainty procedure were used, while $T = 3$ is indicative of the approximate size that would correspond to the result of a dynamic tolerance procedure. Given the sensitivity of the final result to factors such as the central value of the coupling used to generate the pseudodata, this is judged to be a sufficient indiciation for the purposes of these broad comparisons, although we will comment more precisely on the result of the dynamic tolerance procedure below. When showing local profiles, we only consider the neutral current (NC) data, considering this to have a more direct constraint on the strong coupling, although including the charged current data in these makes little difference. For illustration, in Fig. 2 we show the corresponding χ^2 profiles for the conservative with and without EIC pseudodata.

We note that the best fit values for the strong coupling in the no EIC case are not the same as in [9]. This is due to the update to the fixed target data and evolution described in Sect. 2, but also to a corrected treatment of the aN^{3}LO K-factors for the jet and vector boson $p_\perp$ (and + jet) datasets, as discussed further in [36]. Furthermore, in this study we do not include the SeaQuest data [44] in the fit, and for this reason these results are not identical to [36]. This overall difference, while noticeable, remains within the uncertainty due to the dynamic tolerance treatment. Moreover, we emphasise that the aim here is to analyse the impact of the EIC pseudodata with respect to the baseline fit excluding EIC pseudodata, rather than investigate the specific value of the strong coupling.

Comparing the 'No EIC' and 'w EIC' cases we see the impact of including the EIC pseudodata when it is generated consistently at NNLO. Locally the EIC pseudodata prefer $\alpha_S(M_Z^2) \sim 0.118$ as we would expect, given this is used to generate the input. This results in a mild increase in the globally preferred value of the strong coupling, from ~ 0.1171 to 0.1172.

The corresponding fit qualities, i.e. the χ^2 per point, are shown in Table 1. Prior to refitting, the description of the EIC pseudodata is, statistically speaking, relatively poor in the

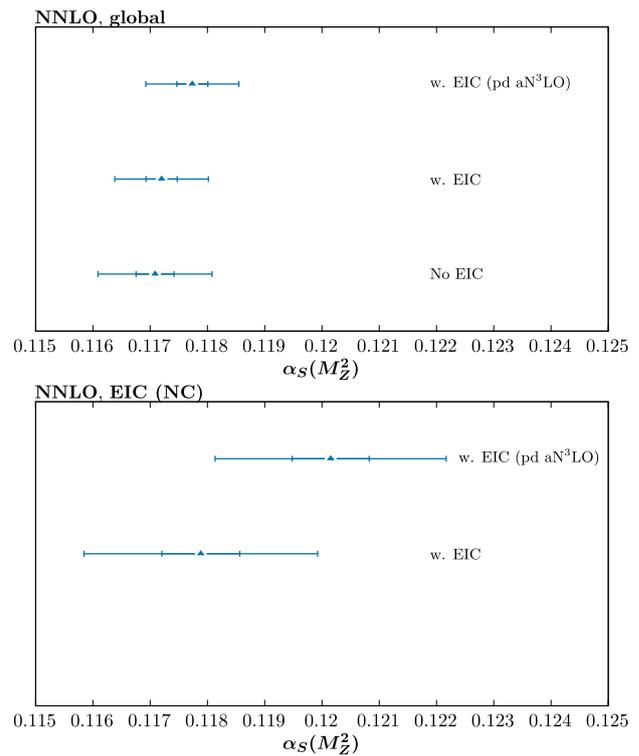**Fig. 1** Best fit value of the strong coupling $\alpha_S(M_Z^2)$ for a range of NNLO fits, including ('w. EIC') or excluding EIC ('no EIC') pseudodata ('pd'), in the conservative scenario. The inner (outer) limits correspond to $\Delta\chi^2 = 1(9)$, and are intended to be indicative of the relative uncertainty on $\alpha_S(M_Z^2)$ in different fits, rather than indicating the actual size of the uncertainty that would result from the dynamic tolerance procedure. The top (bottom) plots show the result of the globally and locally (EIC NC) preferred values

comparison. This is permissible, given that while the underlying theory used to generate the pseudodata is broadly consistent with the fit, it is not identical; that is, the underlying MSHT20NNLO set used to generate the pseudodata, as well as the input value of $\alpha_S(M_Z^2) = 0.118$ are not identical to the result of this, more recent, fit excluding EIC pseudodata. There is in particular no account of PDF uncertainties in the fit qualities shown here. On the other hand, after refitting, we can see that the fit quality is indeed close to unity with only a rather mild deterioration in the description of the non-EIC data, and so the pseudodata can be safely accommodated.

3.2 Impact of explicitly inconsistent EIC pseudodata: NNLO

We next investigate the impact of intentionally injecting some inconsistency between the generated pseudodata and the existing best fit PDFs arising from the global fit to existing data. In practice such an inconsistency could arise due to the EIC data being more sensitive to certain theoretical corrections than the existing data, or due to issues in the experi-
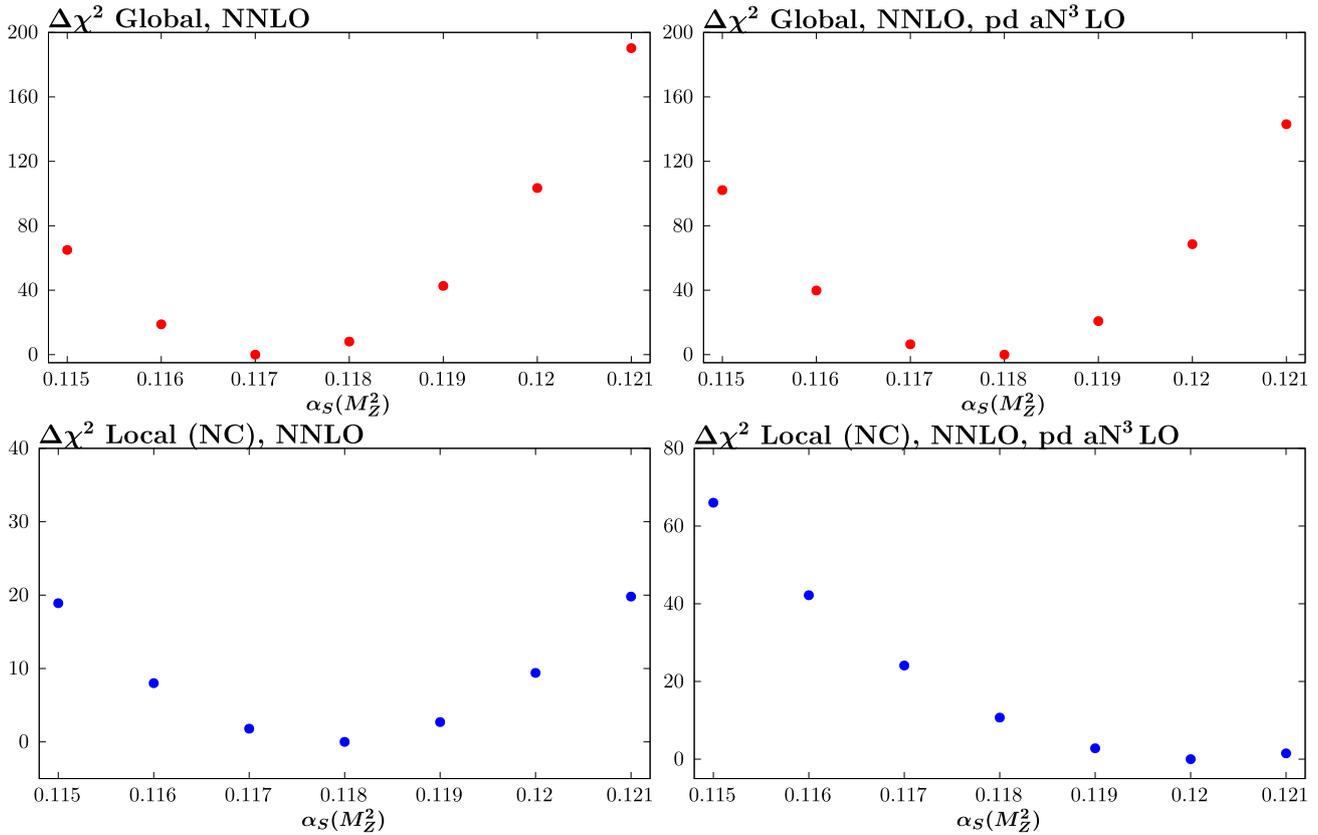
Fig. 2 χ^2 profiles corresponding to the 'w. EIC' and 'w. EIC (pd aN^{3}LO)' cases in Fig. 1

Table 1 χ^2 values for MSHT fits performed at NNLO, including and excluding conservative EIC pseudodata in different forms, as indicated. The global fit quality as well as that for various subsets is shown. For the fit excluding EIC pd, the predicted χ^2 values are shown (not in bold) for the EIC pd. The χ^2 values are indicated, with the χ^2 per point given in brackets

	No EIC	w. EIC	w EIC (pd aN^{3}LO)
EIC	690.0 (1.25)	**522.5 (0.95)**	**583.5 (1.06)**
Fixed target	**2168.0 (1.09)**	**2197.0 (1.11)**	**2178.0 (1.10)**
HERA	**1623.5 (1.28)**	**1612.3 (1.28)**	**1625.6 (1.29)**
Hadron Collider	**2513.6 (1.40)**	**2513.4 (1.40)**	**2520.0 (1.41)**
Global (no EIC)	**6305.2 (1.25)**	**6322.6 (1.26)**	**6323.7 (1.26)**
Global	6995.1 (1.25)	**6845.0 (1.22)**	**6907.2 (1.24)**

mental or theoretical uncertainty estimations in older DIS or hadron collider data that constrain similar partons to the EIC data. Introducing the explicit inconsistency thus serves as a proxy for a wide range of possible sources of tension in the fits. Ideally such tensions would be addressed by identifying the underlying reason for the inconsistency and introducing explicit additional sources of uncertainty accordingly. However, where the reasons are not identifiable, such tensions require caution in the interpretation of the fit results.

In Fig. 1 we show the 'w. EIC (pd aN^{3}LO)' result where the pseudodata are instead generated at aN^{3}LO, while maintaining a NNLO fit; in other words, such that the pseudodata and the fit itself are by construction inconsistent. The pseudodata now locally prefer a higher value of $\alpha_S(M_Z^2) \sim 0.120$, despite the fact that these are generated at $\alpha_S(M_Z^2) = 0.118$. That is, the inconsistency between the aN^{3}LO generated pseudodata and the NNLO fit results in a bias in the preferred value of the strong coupling for the EIC pseudodata, with the larger value effectively compensating for the missing higher order between the NNLO and aN^{3}LO description. Globally, this is seen to pull the preferred value of $\alpha_S(M_Z^2)$ up from ~ 0.1171 to ~ 0.1176. These differences are on the edge of the textbook $\Delta\chi^2 = T^2 = 1$ uncertainty, but well within the enlarged $T = 3$ tolerance more typical of the

Table 2 χ^2 values for MSHT fits performed at aN^{3}LO, including and excluding EIC pseudodata in different forms, as indicated. The global fit quality as well as that for various subsets is shown. For the fit exclud-ing EIC pd, the predicted χ^2 values are shown (not in bold) for the EIC pd. The 'TMC' and 'NNPDF4.0' cases correspond to generating the pseudodata with explicit inconsistencies, as described in the text

	No EIC	w. EIC	w EIC (TMC)	w EIC (NNPDF4.0)
EIC	572.3 (1.10)	**509.2 (0.98)**	**532.0 (1.02)**	**584.9 (1.12)**
Fixed target	**2153.6 (1.09)**	**2155.3 (1.09)**	2149.8 (1.08)	2215.9 (1.29)
HERA	**1626.3 (1.29)**	**1629.3 (1.29)**	1631.1 (1.29)	1633.9 (1.29)
Hadron Collider	**2388.1 (1.33)**	**2390.3 (1.34)**	2388.7 (1.33)	2432.4 (1.36)
Global (no EIC)	**6168.0 (1.23)**	**6175.0 (1.23)**	6169.6 (1.23)	6282.2 (1.25)
Global	6740.3 (1.21)	**6684.2 (1.20)**	**6701.6 (1.21)**	**6787.0 (1.24)**

MSHT dynamic tolerance treatment. At the level of the fit quality, however, we can see from Table 1 that the description of the EIC pseudodata remains completely acceptable.[1] This therefore indicates the need for caution when assessing the level of incompatability between datasets from the overall fit quality alone. In particular, while this is not at all evident in the χ^2 deterioration, the impact on the extracted value of the strong coupling, if a naive $T^2 = 1$ approach is taken, is not negligible. We discuss this point further in Sect. 3.4, in the context of the aN^{3}LO fits.

3.3 aN^{3}LO QCD fits

In this section we consider the aN^{3}LO fit. The fit quality is shown in Table 2. Before refitting, the description of the EIC pseudodata is good, being less than 2σ above unity. While an acceptable level of deviation due to purely statistical scatter alone, we note that as discussed above in the case of the NNLO fit, while the underlying PDFs used to generate the pseudodata are broadly consistent with those extracted from the baseline fit, they are not identical, and hence some deviation may be expected. After refitting, on the other hand, we can see that the χ^2 per point is close to unity with only a rather mild deterioration in the description of the non-EIC data, and so the pseudodata can again be safely accommodated.

In terms of the impact on the strong coupling determination, shown in Fig. 3, the results for the conservative scenario are rather similar to the NNLO case. We show the corresponding χ^2 profiles with and without EIC pseudodata (conservative scenario) in Fig. 4. The locally preferred value of the EIC pseudodata remains at 0.118, and this increases the globally preferred value of the strong coupling somewhat. In more detail, we observe that the global impact of the EIC pseudodata on the best fit value of the strong coupling is more significant than in the NNLO case, increasing it from ~ 0.1162 to ~ 0.1167.
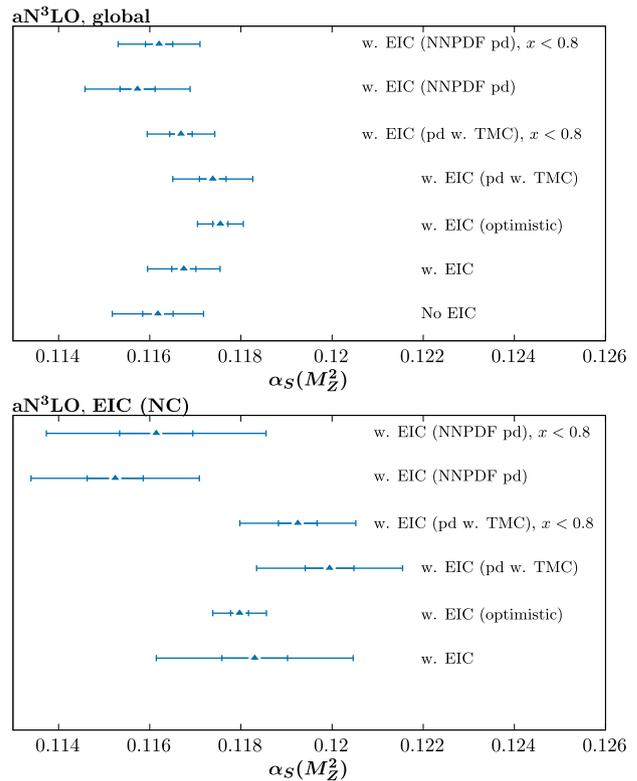

Fig. 3 Best fit value of the strong coupling $\alpha_S(M_Z^2)$ for a range of aN^{3}LO fits, including or excluding EIC pseudodata. The inner (outer) limits correspond to $\Delta\chi^2 = 1$ (9) that result from a quadratic fit about the minimum to the corresponding global profile, and are intended to be indicative of the relative uncertainty on $\alpha_S(M_Z^2)$ in different fits, rather than indicating the actual size of the uncertainty that would result from the dynamic tolerance procedure. The left (right) plots show the result of the globally and locally (EIC NC) preferred values. Unless otherwise specified, the pseudodata correspond to the conservative scenario

The impact on the uncertainty of the strong coupling can be seen from Fig. 4 to be relatively moderate, in particular reducing the uncertainty with a global tolerance by $\sim 20\%$. We show this here with global $T = 1$ and $T = 3$ as the inner and outer error bars. To be more precise, this can be evaluated following the standard dynamic tolerance criterion discussed in e.g. [1], which we discuss now, although we do not show
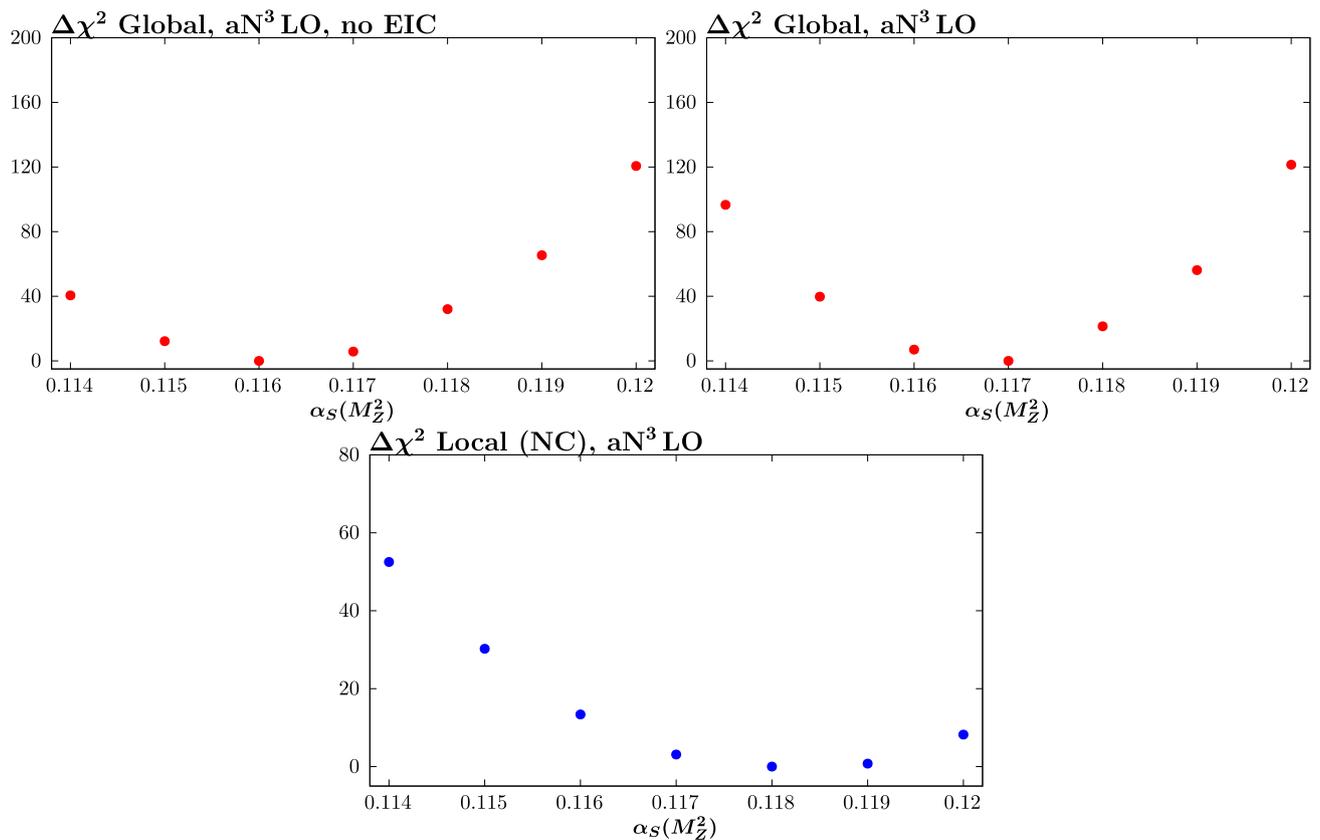
---

[1] A 1σ variation in the χ^2/N for the number of points is $\sqrt{2/N} \sim 0.06$, and hence this is only $\sim 1\sigma$ above unity.

Fig. 4 χ^2 profiles corresponding to the 'No EIC' and 'w. EIC' cases in Fig. 3

these limits on the corresponding figures. As the locally preferred value is rather higher than the global best fit value, we expect the lower limit to be more stringent. This is indeed the case, with a lower limit corresponding to ~ -0.0015, i.e. comparable to existing lower limits in the most recent MSHT study [9], which come from lower energy fixed-target data. We do not perform a full reanalysis here, as the precise uncertainty on the coupling will depend on factors such as how the pseudodata are generated; if the input value of the coupling is more in line with the globally preferred one, for example, this will change the precise limits. Nonetheless, we broadly find that the expected limit is of the order of existing ones. Given the potentially better understanding of the detector and more refined treatment of experimental uncertainties that we may expect at the EIC, this is encouraging. The pull on the global value of the coupling is rather larger in comparison to the NNLO case since the preferred value at aN^{3}LO is somewhat lower than at NNLO, and lower than in the recent study [9], as discussed above. That is, the central value for the baseline without EIC pseudodata is rather further away than the value of 0.118 that we have used to generate the pseudodata than at NNLO, leading to a larger expected shift.

We next consider the impact of the EIC pseudodata generated in the optimistic scenario described in the Sect. 2. The

difference in the local profile is significant, decreasing the fixed T uncertainty by a factor of $\sim 3 - 4$. Globally, this is washed out to some extent, as we would expect, but nonetheless the fixed tolerance uncertainty is reduced by $\sim 50\%$. This therefore demonstrates the clear potential for a large impact on the strong coupling determination by EIC data. Applying the dynamic tolerance procedure, we find that the lower limit on the strong coupling is greatly reduced with respect to previous MSHT results, being a limit of ~ -0.0005 (i.e. rather in line with the fixed $T^2 = 9$ result shown in the figure), while the upper limit is ~ 0.0015. As always the precise values will depend on the details of the real data, but this again indicates the strong potential constraining power of such data.

3.4 Impact of explicitly inconsistent EIC pseudodata: aN^{3}LO

We next, as in Sect. 3.2 but now at aN^{3}LO, investigate the impact of intentionally injecting some inconsistency between the generated pseudodata and the existing best fit PDFs arising from the global fit to existing data.

First, the 'EIC (pd w. TMC)' result shows the impact of generating the pseudodata with theoretical predictions that include target mass corrections (TMCs) in F_2, as outlined

in e.g. [45], see Eq. (61) of this reference, and considered recently in [36] . For the fit, however, we omit these; while in principle these can (and should) be included, where relevant this can provide some handle on any missing theoretical ingredients in the calculation of this sort. Interestingly, this is seen to result in a significant upwards shift in the locally preferred value of the strong coupling, and due to this a further increase in the globally preferred value, from ~ 0.1162 to ~ 0.1174 in comparison to the result without EIC pseudodata, and from ~ 0.1167 to ~ 0.1174 in comparison to the result with EIC pseudodata in the consistent case. This is despite the fact that the overall fit quality, shown in Table 2, does not noticeably deteriorate. This is not a contradiction, given the difference between the parameter fitting criterion one in principle uses to evaluate the best fit value of the strong coupling and its uncertainty, in comparison to the hypothesis testing criterion used to evaluate the overall fit quality. Nonetheless, this clearly indicates the danger of naively assuming an e.g. $\Delta \chi^2 = 1$ criterion can be used, based on observation of the overall fit quality being reasonable. In particular, the global $\Delta \chi^2 = 1$ uncertainty in the current case is $\sim \pm 0.0003$, and hence the above shift would naively correspond to a 4σ shift. For the arguably more relevant comparison to the preferred value when the EIC pseudodata are consistently included, for which the preferred value of the strong coupling is ~ 0.1167, this corresponds to a $\sim 2\sigma$ shift with respect to this.

However, the reason for this shift lies almost entirely in the $x = 0.815$ bins that exist in the pseudodata, which are at sufficiently high x that even for the reasonably high values of Q^2 ($\sim 100 \, \text{GeV}^2$ or more) probed, TMCs are relevant. Indeed, by simply removing these bins we end up with a result that is again consistent with the result from the fit with consistent pseudodata, i.e. the biasing effect is largely removed in the global fit (although it remains somewhat visible in the local profile).

To investigate another possible source of inconsistency, we instead generate the pseudodata with the NNPDF4.0aN^3 LO [8] set. As shown in Fig. 5 for the relevant case of the charge weighted quark singlet, there are some regions of x where these are in tension, even beyond the quoted uncertainty bands. This therefore provides a method to model the impact of a situation where the eventual EIC data prefer a somewhat different shape to the underlying PDFs, while being agnostic as to the underlying reason.

The results are shown in Fig. 3 as the 'w. EIC (NNPDF pd)' cases. The local profile is now shifted towards lower values of the strong coupling, albeit with a relatively shallow minimum. The reason for this is most likely due to the fact that for the smaller NNPDF4.0 charge weighted quark singlet, the corresponding EIC pseudodata are in general lower and hence a smaller coupling is required to reduce the theoretical prediction and match these pseudodata. This leads to a
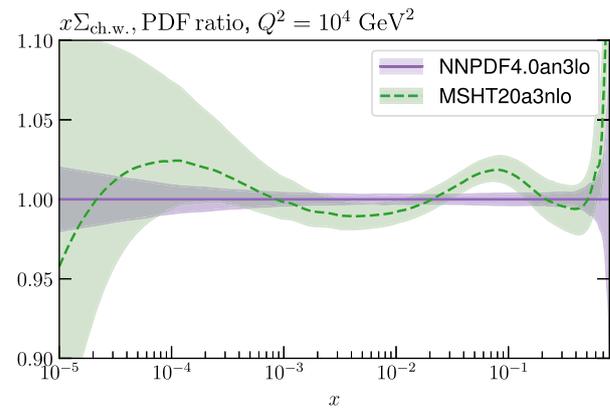

Fig. 5 Ratio of the MSHT20aN^{3}LO to NNPDF4.0aN^{3}LO charge weighted quark singlet PDFs at $Q^2 = 10^4 \, \text{GeV}^2$. Results in both cases correspond to the published PDF sets

shift downwards in the globally preferred value of the strong coupling, from ~ 0.1162, the value without EIC pseudodata included, to ~ 0.1157. This shift is $\sim 1-2\sigma$ if a textbook $\Delta \chi^2 = 1$ uncertainty is applied, corresponding to a ~ 0.0003 variation in the strong coupling. In fact, we recall from above that if the pseudodata are applied without any explicit inconsistency then the preferred value of the strong coupling is shifted upward to ~ 0.1167 (see Sect. 3.3), i.e. an increase by $\sim 1-2\sigma$ towards the value of 0.118 used to generate the pseudodata. In other words, this change in the preferred value of the strong coupling is of the same order of shift as the case with consistently generated pseudodata. However, due to the tension now present, this shift is in the opposite direction to the consistent case, that is *away* from the value of the strong coupling of 0.118 that is used to generate the pseudodata.

These results indicate the manner in which tensions between datasets can lead to differences in the preferred value of the strong coupling. We note that the overall fit quality, shown in Table 2, deteriorates by $\sim 2\sigma$, which is therefore noticeable but not unusual in the context of global PDF fits. If the dynamic tolerance criterion is applied we find that, due in part to the flatter local profile but also the precise position of the preferred value of the coupling, the constraints due to the EIC pseudodata are looser than in the consistent case, with the (smaller) upper limit being $\sim +0.002$.

If we remove the $x = 0.815$ data points, as above, then we may expect from Fig. 5 that this will moderate the tension between the EIC pseudodata generated with NNPDF4.0 and the rest of the global fit, given the largest difference between the NNPDF4.0 and the MSHT20 set for the charge weighted quark singlet is in this very high x region. The result of doing this is shown in Fig. 3, and indeed there is a tendency to reduce the shift towards lower values of the strong coupling in the local profile; consequently, globally the central value of the

strong coupling is ~ 0.1162, i.e. essentially unchanged with respect to the case without EIC pseudodata.

However, both of these cases should also be considered in light of the result for the optimistic scenario. We can see in Fig. 3 that the result of generating the pseudodata with the NNPDF4.0 set is in significant tension with the EIC (optimistic) result, when using the $T^2 = 1$ uncertainty. This is however rather reduced (though still somewhat present) if an enlarged tolerance is used. For the TMC case, the upwards shift in fact accidentally moves the result close to the optimistic case. However the size of the shift from this effect is noticeable with respect to the $T^2 = 1$ uncertainty. Clearly if the inconsistent pseudodata exercises were performed within the optimistic scenario, the level of tension would increase further.

4 Summary and outlook

In this paper we have presented a study of the impact that EIC measurements of inclusive DIS have on the determination of the strong coupling, within the context of the MSHT global PDF fitting framework, at both NNLO and aN^{3}LO in QCD. We have supplemented the existing MSHT global dataset with EIC pseudodata and performed a simultaneous PDF and α_S determination. We have performed this study with both a conservative and more optimistic set of assumptions about the eventual experimental uncertainties on the EIC data. We have found their impact in the conservative case to be moderate, but non-negligible, while in the optimistic scenario EIC data lead to a significant reduction in the projected uncertainty.

We have in addition extended the standard approach to pseudodata projection studies to allow for the possibility of some tension between the EIC data and other data in the fit, which could occur due to a wide range of theoretical or experimental effects that are not fully treated in the uncertainty determinations, even under the assumption that the EIC data are perfectly correct within their uncertainties. This is arguably more representative of the genuine situation in a global PDF fit, where such tensions often occur. To this end, we have considered a range of approaches to inject a source of tension into the pseudodata, namely: by generating the pseudodata at aN^{3}LO order, but fitting at NNLO; by including target mass corrections in the pseudodata but not in the fit; and by generating the pseudodata with the NNPDF4.0aN^{3}LO set, which exhibits some difference relative to MSHT.

We have found that for the inconsistent fits considered above, the overall fit quality to the EIC pseudodata can be perfectly acceptable, but even in the conservative uncertainty scenario lead to shifts of $\sim 1 - 2\sigma$ in the value of the strong coupling, if a textbook $\Delta\chi^2 = 1$ criterion is used. In the optimistic scenario the relative impact would be larger still.

This provides further evidence in support of the need of a complete account of theoretical uncertainties, and of a more conservative approach to error determination, as is advocated by MSHT, in the context of a fit where departures from textbook statistics are evident.

More broadly, however, our results underline the promising and exciting potential for the EIC in the determination of the strong coupling, and QCD analyses more generally. This is particularly evident under more optimistic assumptions about the eventual experimental uncertainties, subject to all other sources of theoretical uncertainty being under control.

Data Availability Statement Data will be made available on reasonable request. [Authors' comment: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.]

Code Availability Statement Code/software will be made available on reasonable request. [Authors' comment: The code/software generated during and/or analysed during the current study is available from the corresponding author on reasonable request.]

References

1. S. Bailey, T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Eur. Phys. J. C **81**, 341 (2021). arXiv:2012.04684
2. NNPDF, R.D. Ball et al., Eur. Phys. J. C **82**, 428 (2022). arXiv:2109.02653
3. T.-J. Hou et al., Phys. Rev. D **103**, 014013 (2021). arXiv:1912.10053
4. S. Alekhin, J. Blümlein, S. Moch, R. Placakyte, Phys. Rev. D **96**, 014011 (2017). arXiv:1701.05838
5. H1, ZEUS, H. Abramowicz et al., Eur. Phys. J. C **75**, 580 (2015). arXiv:1506.06042
6. ATLAS, G. Aad et al., Eur. Phys. J. C **82**, 438 (2022). arXiv:2112.11266
7. J. McGowan, T. Cridge, L.A. Harland-Lang, R.S. Thorne, Eur. Phys. J. C **83**, 185 (2023). arXiv:2207.04739

8. NNPDF, R.D. Ball et al., (2024). arXiv:2402.18635
9. T. Cridge, L.A. Harland-Lang, R.S. Thorne, Eur. Phys. J. C **84**, 1009 (2024). arXiv:2404.02964
10. NNPDF, R.D. Ball et al., Eur. Phys. J. C **85**, 1001 (2025). arXiv:2506.13871
11. Particle Data Group, S. Navas et al., Phys. Rev. D **110**, 030001 (2024)
12. D. d'Enterria et al., J. Phys. G **51**, 090501 (2024). arXiv:2203.08271
13. A. Accardi et al., Eur. Phys. J. A **52**, 268 (2016). arXiv:1212.1701
14. R. Abdul Khalek et al., Nucl. Phys. A **1026**, 122447 (2022). arXiv:2103.05419
15. R.A. Khalek, J.J. Ethier, E.R. Nocera, J. Rojo, Phys. Rev. D **103**, 096005 (2021). arXiv:2102.00018
16. T.J. Hobbs, B.-T. Wang, P.M. Nadolsky, F.I. Olness, Phys. Rev. D **100**, 094040 (2019). arXiv:1904.00022
17. T.J. Hobbs, B.-T. Wang, P.M. Nadolsky, F.I. Olness, PoS **DIS2019**, 247 (2019). arXiv:1907.00988
18. M. Arratia, Y. Furletova, T.J. Hobbs, F. Olness, S.J. Sekula, Phys. Rev. D **103**, 074023 (2021). arXiv:2006.12520
19. R. Abdul Khalek et al., arXiv:2203.13199 (2022)
20. S. Amoroso et al., Acta Phys. Polon. B **53**, 12 (2022). arXiv:2203.13923
21. N. Armesto et al., Phys. Rev. D **109**, 054019 (2024). arXiv:2309.11269
22. NNPDF, R.D. Ball et al., Phys. Rev. D **109**, L091501 (2024). arXiv:2311.00743
23. S. Cerci et al., Eur. Phys. J. C **83**, 1011 (2023). arXiv:2307.01183
24. M. Azizi, M. Soleymaninia, H. Hashamipour, M. Salajegheh, H. Khanpour, U.G. Meißner, Phys. Rev. D **111**(3), 034023 (2025). https://doi.org/10.1103/PhysRevD.111.034023
25. G. Watt, R.S. Thorne, JHEP **08**, 052 (2012). arXiv:1205.4024
26. K. Kovařík, P.M. Nadolsky, D.E. Soper, Rev. Mod. Phys. **92**, 045003 (2020). arXiv:1905.06957
27. X. Jing et al., Phys. Rev. D **108**, 034029 (2023). arXiv:2306.03918
28. J. Jiménez-López, P.R. Newman, K. Wichmann, Phys. Rev. D **111**, 056014 (2025). arXiv:2412.16123
29. G. Falcioni, F. Herzog, S. Moch, A. Vogt, Phys. Lett. B **842**, 137944 (2023). arXiv:2302.07593
30. G. Falcioni, F. Herzog, S. Moch, A. Vogt, Phys. Lett. B **846**, 138215 (2023). arXiv:2307.04158
31. G. Falcioni, F. Herzog, S. Moch, J. Vermaseren, A. Vogt, Phys. Lett. B **848**, 138351 (2024). arXiv:2310.01245
32. S. Moch, B. Ruijl, T. Ueda, J. Vermaseren, A. Vogt, Phys. Lett. B **849**, 138468 (2024). arXiv:2310.05744
33. G. Falcioni, F. Herzog, S. Moch, A. Pelloni, A. Vogt, Phys. Lett. B **856**, 138906 (2024). arXiv:2404.09701
34. G. Falcioni, F. Herzog, S. Moch, A. Pelloni, A. Vogt, Phys. Lett. B **860**, 139194 (2025). arXiv:2410.08089
35. T. Cridge, L.A. Harland-Lang, R.S. Thorne, MSHT approximate N3LO PDFs: updates and consequences for phenomenology. in *32nd International Workshop on Deep-Inelastic Scattering and Related Subjects* (2025). arXiv:2510.09321
36. L.A. Harland-Lang, T. Cridge, M. Reader, R.S. Thorne, arXiv:2510.03753 (2025)
37. BCDMS, A.C. Benvenuti et al., Phys. Lett. B **223**, 485 (1989)
38. BCDMS, A.C. Benvenuti et al., Phys. Lett. B **237**, 592 (1990)
39. New Muon Collaboration, M. Arneodo et al., Nucl. Phys. B **483**, 3 (1997). arXiv:hep-ph/9610231
40. New Muon Collaboration, M. Arneodo et al., Nucl. Phys. B **487**, 3 (1997). arXiv:hep-ex/9611022
41. E665, M.R. Adams et al., Phys. Rev. D **54**, 3006 (1996)
42. NuTeV, M. Tzanov et al., Phys. Rev. D **74**, 012008 (2006). arXiv:hep-ex/0509010
43. A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189 (2009). arXiv:0901.0002
44. SeaQuest, J. Dove et al., Nature **590**, 561 (2021). arXiv:2103.04024. [Erratum: Nature 604, E26 (2022)]
45. I. Schienbein et al., J. Phys. G **35**, 053101 (2008). arXiv:0709.1775