Revisiting constraints on proton PDFs from HERA DIS, Drell-Yan, W and Z boson production, and projected EIC measurements

Majid Azizi,[1,*] Maryam Soleymaninia,[1,†] Hadi Hashamipour,[2,‡] Maral Salajegheh,[3,§] Hamzeh Khanpour,[4,5,1,‖] and Ulf-G. Meißner[3,6,7,¶]

[1]*School of Particles and Accelerators, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5531, Tehran, Iran*

[2]*Istituto Nazionale di Fisica Nucleare, Gruppo collegato di Cosenza, I-87036 Arcavacata di Rende, Cosenza, Italy*

[3]*Helmholtz-Institut für Strahlen-und Kernphysik and Bethe Center for Theoretical Physics, Universität Bonn, D-53115 Bonn, Germany*

[4]*AGH University, Faculty of Physics and Applied Computer Science, Al. Mickiewicza 30, 30-055 Krakow, Poland*

[5]*Department of Physics, University of Science and Technology of Mazandaran, P.O. Box 48518-78195, Behshahr, Iran*

[6]*Institute for Advanced Simulation (IAS-4), Forschungszentrum Jülich, D-52425 Jülich, Germany*

[7]*Peng Huanwu Collaborative Center for Research and Education, Beihang University, Beijing 100191, China*

(Received 6 November 2024; accepted 21 January 2025; published 14 February 2025)

We present new parton distribution functions (PDFs) at next-to-leading order and next-to-next-to-leading order in perturbative QCD, derived from a comprehensive QCD analysis of high-precision datasets from combined HERA deep-inelastic scattering, the Tevatron, and the LHC. To improve constraints on quark flavor separation, we incorporate Drell-Yan pair production data, which provides critical sensitivity to the quark distributions. In addition, we include the latest W and Z boson production data from the CDF, D0, ATLAS, and CMS collaborations, further refining both quark and gluon distributions. Our nominal QCD fit integrates these datasets and examines the resulting impact on the PDFs and their associated uncertainties. Uncertainties in the PDFs are quantified using the Hessian method, ensuring robust error estimates. Furthermore, we explore the sensitivity of the strong coupling constant, $\alpha_s(M_Z^2)$, and proton PDFs in light of the projected measurements from the Electron-Ion Collider, where improvements in precision are expected. The analysis also investigates the effects of inclusive jet and dijet production data, which provide enhanced constraints on the gluon PDF and $\alpha_s(M_Z^2)$.

DOI: 10.1103/PhysRevD.111.034023

I. INTRODUCTION

Deep inelastic scattering (DIS) is one of the most thoroughly studied processes in perturbative QCD, as reviewed in Refs. [1–4]. DIS plays a fundamental role in probing the internal structure of hadrons, particularly in determining parton distribution functions (PDFs), which describe the momentum distribution of quarks and gluons within a hadron [5]. Parton density functions, $f_a(x, Q)$, are essential for accurate predictions in high-energy physics, especially at colliders such as the CERN Large Hadron Collider (LHC), and future facilities such as the Electron-Ion Collider (EIC) [6,7], the Large Hadron Electron Collider (LHeC) [8], and the Future Circular Collider (FCC) [9–11].

The upcoming EIC and LHeC are anticipated to enhance PDF determinations significantly by expanding the kinematic range accessible to current experiments. These facilities will enable a more precise extraction of PDFs by providing complementary insights into the proton structure. The EIC is expected to offer critical data in the large-x region, which HERA was unable to fully explore, while the LHeC will achieve high precision in probing the small-x region, thereby reducing uncertainties in the gluon and sea quark distributions. However, achieving such precision required for modern experiments presents two primary

*Contact author: Ma.Azizi@ipm.ir
†Contact author: Maryam_Soleymaninia@ipm.ir
‡Contact author: Hadi.Hashamipour@lnf.infn.it
§Contact author: Maral@hiskp.uni-bonn.de
‖Contact author: Hamzeh.Khanpour@cern.ch
¶Contact author: Meissner@hiskp.uni-bonn.de

challenges: missing higher-order QCD corrections in theoretical calculations and inherent uncertainties in the PDFs themselves.

Several global QCD analyses have been carried out by multiple collaborations, including CTEQ-TEA [12,13], NNPDF4.0 [14,15], MSHT20 [16], and other groups [17–19], to address these challenges and improve the PDF determinations. In particular, the CTEQ-TEA Collaboration has developed a new generation of general-purpose PDFs to supersede the CT18 set [12], aimed at a wide range of applications, including precision studies of electroweak processes, Higgs production, and searches for physics beyond the Standard Model.

Recent advancements in PDF determinations have been driven by the inclusion of high-precision data from the LHC. For instance, the NNPDF3.1 set [20] was the first to extensively incorporate LHC data, achieving a precision level of 3–5% in PDF uncertainties. The latest NNPDF4.0 set [14] builds on this by incorporating LHC Run II data at $\sqrt{s} = 13$ TeV and introducing machine learning techniques to optimize the PDF fitting process. This approach, along with rigorous validation through closure tests, represents a significant leap forward in PDF precision.

The MSHT Collaboration introduced the MSHT20 set [16], determined through global analyses of available hard-scattering data up to next-to-next-to-leading-order (NNLO) accuracy. MSHT20 builds on the MMHT14 framework [21] with extended parametrizations and the inclusion of new datasets, ranging from the final HERA combined data on total and heavy-flavor structure functions to recent LHC measurements of vector boson production, inclusive jets, and top quark production at 7 and 8 TeV. Notable updates include improvements in the $u - d$ valence quark difference and the strange-quark PDFs, attributed to new data and updated parametrizations. Additionally, MSHT20 incorporates NNLO corrections for dimuon production in neutrino DIS, leading to reduced uncertainties in key processes such as Higgs and W/Z boson production at the LHC. A recent update from the MSHT Collaboration is presented in Ref. [22], where the authors discuss a global closure test of the fixed parametrization approach to PDF fitting.

In this work, we present new PDFs determined at next-to-leading-order (NLO) and NNLO accuracy through a comprehensive analysis of high-precision data from the LHC, Tevatron, and the combined HERA DIS datasets. The focus of this study is to maximize the sensitivity of the PDFs by incorporating data from Drell-Yan pair production and W/Z boson production at the Tevatron and LHC, with particular emphasis on the strange-quark density, which has been a topic of interest in recent QCD analyses [16,23,24]. A combined PDF fit integrates all these datasets, providing a thorough examination of the resulting PDFs and their associated uncertainties. Uncertainty estimates are calculated using the Hessian method, ensuring robust uncertainty quantification.

Furthermore, we explore the impact of simulated inclusive DIS data from the EIC [6,7] on the determination of proton PDFs. The EIC is expected to provide complementary information in the large Bjorken-x region, which will be crucial for improving PDF precision at both NLO and NNLO accuracy [25,26]. We also estimate the expected experimental uncertainty in the strong coupling constant, $\alpha_s(M_Z^2)$, when incorporating simulated EIC inclusive data into the QCD analysis.

Finally, we examine the significant role that jet and dijet production data play in constraining PDFs, particularly the gluon distribution. Jet production data are directly sensitive to the gluon content in the proton, and their inclusion is expected to reduce uncertainties in the gluon PDF and improve the determination of $\alpha_s(M_Z^2)$. The impact of such datasets on the precision of α_s is also examined and discussed in this work.

The remainder of this paper is organized as follows: In Sec. II, we outline the theoretical framework used in our analysis, focusing on the key aspects of perturbative QCD and the role of proton PDFs. The datasets employed in this analysis, including those from HERA DIS, Tevatron, and the LHC, are discussed comprehensively in Sec. III. Section IV describes the methodology utilized in the global PDF fit, highlighting the statistical tools and techniques used to quantify uncertainties. In Sec. V, we present our main findings, offering an in-depth examination of the impact of different datasets on the resulting PDFs and providing a comparison with previous determinations. This section also explores the influence of simulated inclusive DIS data from the future EIC experiments on the determination of proton PDFs and the precision of $\alpha_s(M_Z^2)$. The role of simulated EIC DIS data in proton PDF determination is discussed in Sec. VI. Section VII covers the impact of jet and dijet production data on proton PDF determination. A detailed examination of the strange-quark density is provided in Sec. VIII. The effect of simulated EIC and jet production data on the determination of the strong coupling constant, α_s, is discussed in Sec. IX. Finally, Sec. XI offers concluding remarks and suggests future research directions for improving PDF precision and reducing theoretical uncertainties.

II. QCD ANALYSIS FRAMEWORK

In this section, we outline the theoretical framework used in our QCD analysis to extract the proton PDFs from high-energy scattering data. Our analysis relies on QCD factorization theorems, which allow the separation of short-distance perturbative interactions, calculable within perturbative QCD, from long-distance nonperturbative effects encoded in the PDFs [27]. This separation ensures that PDFs are universal across different hard-scattering processes, making them indispensable for predicting cross sections at high-energy colliders.

A. Factorization in DIS

DIS is a cornerstone in the determination of PDFs. The DIS cross section is expressed in terms of hadronic structure functions $F_2(x, Q^2)$ and $F_L(x, Q^2)$, which can be written as a convolution of PDFs and perturbatively calculable Wilson coefficients, $C_i(x, \alpha_s(Q^2))$, at a given scale Q^2. Specifically, the structure functions take the form [5,28,29]

$$F_i(x, Q^2) = \sum_{k=1}^{n_f} C_{i,k}(x, \alpha_s(Q^2)) \otimes x q_k^+(x, Q^2) + C_{i,g}(x, \alpha_s(Q^2)) \otimes x g(x, Q^2), \quad i = \{2, L\}, \tag{1}$$

where $q_k^+(x, Q^2)$ are the quark and antiquark distributions and $g(x, Q^2)$ is the gluon distribution. The Wilson coefficients $C_{i,k}(x, \alpha_s(Q^2))$ are known up to NNLO for massless quarks [30], while treatments for massive quarks are implemented through the Variable Flavor Number Scheme (VFNS), allowing for a smooth transition across heavy quark mass thresholds [28,31–34].

B. Drell-Yan process

The Drell-Yan process, in which a quark-antiquark pair annihilates into a virtual photon or Z boson that subsequently decays into a lepton pair, is another critical process for constraining PDFs. The cross section for the Drell-Yan process is factorized as [5,35,36]

$$\sigma_{\mathrm{DY}} = \sum_{q,\bar{q}} \int dx_1 dx_2 f_q(x_1, Q^2) f_{\bar{q}}(x_2, Q^2) \hat{\sigma}_{q\bar{q}\to\ell^+\ell^-}(s), \tag{2}$$

where $f_q(x_1, Q^2)$ and $f_{\bar{q}}(x_2, Q^2)$ are the PDFs for quarks and antiquarks and $\hat{\sigma}_{q\bar{q}\to\ell^+\ell^-}$ is the hard-scattering cross section for the pertinent subprocess. The Drell-Yan process is essential for constraining the sea quark and antiquark distributions, particularly at medium-to-large x.

C. Electroweak boson production

The production of electroweak gauge bosons (W and Z) in hadronic collisions provides powerful constraints on the flavor decomposition of the quark PDFs. The inclusive cross section for W and Z boson production can be expressed as [5,37]

$$\sigma_{W/Z} = \sum_{q,\bar{q}} \int dx_1 dx_2 f_q(x_1, Q^2) f_{\bar{q}}(x_2, Q^2) \hat{\sigma}_{q\bar{q}\to W^\pm/Z}(s). \tag{3}$$

These processes are sensitive to both quark and antiquark distributions, offering precise constraints on quark flavor separations. The inclusion of high-precision measurements from LHC experiments (e.g., ATLAS and CMS) allows for a more refined determination of the quark PDFs, particularly for the up, down, and strange quarks at low to moderate x values.

D. Jet production and gluon PDF constraints

Jet production in hadronic collisions provides crucial information about the gluon content of the proton. The cross section for inclusive jet production can be factorized as [38,39]

$$\sigma_{\mathrm{jet}} = \sum_{a,b=q,\bar{q},g} \int dx_1 dx_2 f_a(x_1, Q^2) f_b(x_2, Q^2) \hat{\sigma}_{ab\to\mathrm{jets}}(s), \tag{4}$$

where the partonic subprocess cross sections $\hat{\sigma}_{ab\to\mathrm{jets}}$ include contributions from quark-quark, quark-gluon, and gluon-gluon scattering. Jet data from HERA and the LHC, particularly at high energies, provide important constraints on the gluon distribution. The recent inclusion of NNLO corrections to jet production processes has further reduced theoretical uncertainties and refined the determination of the gluon PDF, especially at low values of x [13].

E. Higher-order corrections in QCD

The inclusion of higher-order QCD corrections, particularly at NNLO accuracy, is crucial for achieving precise determinations of PDFs. These corrections help to reduce the dependence of cross sections on factorization and renormalization scales, thus providing more reliable theoretical predictions. At NLO, the gluon PDF receives significant corrections from processes such as $qg \to q$ and $gg \to g$. Moving to NNLO, additional contributions from gluon-gluon and quark-antiquark interactions become relevant. NNLO corrections have been shown to significantly improve the agreement with experimental data, especially for processes dominated by gluons, such as Higgs production via gluon fusion [40] and inclusive jet production. The calculation of NNLO splitting functions [33,41–43] and Wilson coefficients allows for more precise Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution of PDFs from the initial scale Q_0^2 to higher scales relevant for collider processes [44]. These corrections are particularly impactful in the small-x regime, where the gluon density rises steeply due to QCD evolution. Incorporating NNLO accuracy thus reduces theoretical uncertainties, especially in regions of phase space where gluon interactions dominate, contributing to improved predictions for hadron collider experiments.

F. Heavy quark treatments in QCD analysis

Heavy quark contributions, particularly from charm and bottom quarks, are treated using the VFNS [28,29,45].

In this scheme, heavy quarks are produced perturbatively at energies above their mass thresholds. This approach allows for smooth matching of PDFs across these thresholds, ensuring that heavy quark PDFs are consistently included in both DIS and hadron collider cross sections. Recent advancements, such as the inclusion of photon-gluon fusion for charm production, have provided valuable constraints on the gluon PDF at small x, thereby improving the precision of the charm PDF. This is especially important for precision studies of the Higgs and electroweak sectors at the LHC.

The treatment of heavy quarks in perturbative QCD is crucial for determining PDFs in processes involving DIS and hadron collisions. Several schemes have been developed to properly account for heavy quark mass effects in different kinematic regions: The fixed-order plus next-to-leading-log (FONLL) approach [46,47] combines fixed-order calculations with next-to-leading log resummations, enabling a smooth transition between the massive and massless cases. The FONLL scheme also allows for the simultaneous treatment of charm and bottom quarks, improving the flexibility and accuracy in PDF determinations, especially at scales where both heavy quarks contribute. The ACOT scheme [48], along with its variants like S-ACOT and ACOT-χ [49], retains mass dependence in the Wilson coefficients, ensuring an accurate treatment of heavy quarks near their production thresholds where $Q^2 \sim m^2$. This enables a precise description of heavy quark contributions in kinematic regions close to the threshold. The ZMVFN scheme [50] neglects heavy quark mass effects at high energy scales, simplifying calculations. However, it can be inaccurate near threshold regions where $Q^2 \sim m^2$, making it less suitable for precise studies involving heavy quark production near their mass thresholds.

The Thorne-Roberts (TR) method [51] is employed in the current analysis to ensure smooth transitions between the massive and massless regimes for heavy quark contributions to PDFs. This method provides a consistent framework for incorporating heavy quark effects across different kinematic regions, effectively bridging the transitions while maintaining the continuity of the PDFs. The TR scheme provided the proper handling of charm and bottom quarks, which is particularly important for accurate predictions at the LHC, including B-meson and Higgs production in association with heavy quarks. The use of the TR scheme for heavy quark treatment also enables us to achieve consistent higher-order predictions and improved constraints on the gluon PDF, which is particularly significant for electroweak studies at the LHC.

III. EXPERIMENTAL MEASUREMENTS

In this section, we provide a detailed overview of the datasets employed in our QCD analysis. We outline the sources, selection criteria, and preprocessing steps applied to ensure the robustness and reliability of our results. In global PDF analyses, DIS datasets, which span a wide range of Bjorken x and momentum transfer Q^2, traditionally serve as the primary constraint. However, DIS data alone cannot sufficiently determine the gluon and sea quark distributions, particularly when it comes to distinguishing between individual quark flavors. In addition to the data from HERA DIS, we utilize a broader array of hard-scattering cross sections, primarily from proton-proton (pp) collisions, to improve the precision of the PDF extractions. Recent studies have demonstrated that incorporating measurements from processes such as heavy flavor production, inclusive jet production, and top quark cross sections significantly enhances the constraints on PDFs. These datasets provide improved coverage across a broad range of x, Q^2, and flavor combinations, ultimately leading to more accurate determinations of the parton distributions. The list of datasets analyzed in this work is presented in Table I, and will be discussed in the next sections.

A. DIS data

The HERA dataset, collected by the H1 and ZEUS collaborations between 1994 and 2007, represents a comprehensive set of DIS measurements of electrons and positrons scattered off protons. This dataset includes over 1 fb^{-1} of integrated luminosity, with proton beam energies ranging from 460 to 920 GeV and an electron beam energy fixed at 27.5 GeV [52]. The data span 6 orders of magnitude in Bjorken x and the negative 4-momentum-transfer squared, Q^2 (from 0.045 to $50,000$ GeV^2), making it the most precise dataset available for ep scattering. These measurements enabled the development of the HERAPDF2.0 PDFs, which were refined through QCD analyses at leading order, NLO, and NNLO accuracy. Additionally, variants such as HERAPDF2.0Jets, which incorporate jet production data, facilitated the simultaneous determination of PDFs and the strong coupling constant α_s, providing valuable insights into proton structure and fundamental interactions [52]. The final combined HERA I + II dataset significantly improved constraints on small-x sea quarks and gluon PDFs. This comprehensive dataset, covering both neutral current (NC) and charged current (CC) interactions at various proton beam energies (460, 575, 820, and 920 GeV), provided a consistent set of cross section measurements across a wide kinematic range. The electron beam energy was fixed at 27.5 GeV, resulting in center-of-mass energies of approximately 225, 251, 300, and 320 GeV. The data are available as functions of Q^2 and x, covering the range $0.045 \leq Q^2 \leq 50,000$ GeV^2 and $6 \times 10^{-7} \leq x \leq 0.65$ for neutral current interactions and $200 \leq Q^2 \leq 50,000$ GeV^2 and $1.3 \times 10^{-2} \leq x \leq 0.40$ for charged current interactions. For the HERA I + II datasets included in our analysis, we follow the procedure in Ref. [53] and apply a kinematic cut at $Q^2 > 10$ GeV^2 to effectively remove higher-twist contributions.

TABLE I. The list of all datasets: DIS HERA I + II, Drell-Yan, and W/Z production data used in the present analysis. For each dataset, we indicate process, measurement, reference and the ranges of their kinematic cuts such as x, y, Q^2 (GeV^2), p_T (GeV), E_T (GeV), $\sqrt{s}$ (TeV), and $\mathcal{L}$ (fb^{-1}).

Dataset	Process	Experiment	Reference	Kinematic ranges and details	
HERA I + II					
	$e^\pm p \to \overset{(-)}{\nu} X$	HERA I + II CC e^+p	[52]	$3\times10^2 \le Q^2 \le 3\times10^4$,	$8.0\times10^{-3} \le x \le 0.4$
		HERA I + II CC e^-p	[52]	$3\times10^2 \le Q^2 \le 3\times10^4$,	$8.0\times10^{-3} \le x \le 0.65$
	$e^\pm p \to e^\pm X$	HERA I + II NC e^-p	[52]	$60 \le Q^2 \le 5\times10^4$,	$8.0\times10^{-4} \le x \le 0.65$
DIS σ		HERA I + II NC e^-p 460	[52]	$6.5 \le Q^2 \le 8\times10^2$,	$3.48\times10^{-5} \le x \le 0.65$
		HERA I + II NC e^-p 575	[52]	$6.5 \le Q^2 \le 8\times10^2$,	$3.48\times10^{-5} \le x \le 0.65$
		HERA I + II NC e^+p 820	[52]	$6.5 \le Q^2 \le 3\times10^4$,	$6.21\times10^{-7} \le x \le 0.4$
		HERA I + II NC e^+p 920	[52]	$6.5 \le Q^2 \le 3\times10^4$,	$5.02\times10^{-6} \le x \le 0.65$
Drell-Yan					
ATLAS	$Z/\gamma^* \to \ell\ell(\ell = e,\mu)$	ATLAS low-mass DY 7 TeV	[54]	$\lvert\eta_\ell\rvert \le 2.4$,	$12 \le m_{\ell\ell} \le 66$
	$Z/\gamma^* \to e^\pm$	ATLAS high-mass DY 7 TeV	[55]	$\lvert\eta_\ell\rvert \le 2.1$,	$116 \le m_{\ell\ell} \le 1500$
FNAL E866/NuSea	$pp(deut) \to \mu^\pm X$	E866 $\sigma^d/2\sigma^p$ (NuSea)	[35]	$0.015 \le x \le 0.35$?,	$4.40 \le m_{\ell\ell} \le 12.9$
W/Z production					
ATLAS	$W^+ \to e^+\nu$	ATLAS W 7 TeV ($\mathcal{L} = 4.6$ pb^{-1})	[53]	$\lvert\eta_\ell\rvert \le 2.5$,	$\lvert y_Z\rvert \le 3.6$
	$W^+ \to \mu^+\nu$	ATLAS W 7 TeV ($\mathcal{L} = 4.6$ pb^{-1})	[53]	$\lvert\eta_\ell\rvert \le 2.5$,	$\lvert y_Z\rvert \le 3.6$
	$Z \to e^+e^-$	ATLAS Z 7 TeV ($\mathcal{L} = 4.6$ pb^{-1})	[53]	$\lvert\eta_\ell\rvert \le 2.5$,	$\lvert y_Z\rvert \le 3.6$
	$Z \to \mu^+\mu^-$	ATLAS Z 7 TeV ($\mathcal{L} = 4.6$ pb^{-1})	[53]	$\lvert\eta_\ell\rvert \le 2.5$,	$\lvert y_Z\rvert \le 3.6$
	$W \to \ell\nu$	ATLAS W 7 TeV ($\mathcal{L} = 35$ pb^{-1})	[58]	$\lvert\eta_\ell\rvert \le 2.4$,	$\lvert y_Z\rvert \le 3.6$
	$Z \to \ell\ell$	ATLAS Z 7 TeV ($\mathcal{L} = 35$ pb^{-1})	[58]	$\lvert\eta_\ell\rvert \le 2.5$,	$\lvert y_Z\rvert \le 3.6$
CMS	$W \to e\nu$	CMS W electron asymmetry 7 TeV	[59]	$\lvert\eta_\ell\rvert \le 2.4$,	$P_T > 35$
	$W \to \mu\nu$	CMS W muon asymmetry 7 TeV	[60]	$\lvert\eta_\ell\rvert \le 2.4$,	$P_T > 25$
	$W \to \mu\nu$	CMS W muon asymmetry 8 TeV	[61]	$\lvert\eta_\ell\rvert \le 2.4$,	$P_T > 25$
	$Z \to e^\pm(\mu^\pm)$	CMS Z muon and electron 7 TeV	[62]	$\lvert y\rvert < 3.5$,	$60 \le m_{\ell\ell} \le 120$
CDF	$W \to e\nu$	CDF W charge asymmetry 1.96 TeV	[63]	$2.1 \le \lvert\eta_\ell\rvert \le 3.5$,	$\lvert y_W\rvert \le 3.0$
	$Z \to e^+e^-$	CDF Z differential 1.96 TeV	[64]	$0.0 \le y_{\ell\ell} \le 2.9$, $66 \le m_{\ell\ell} \le 116$	
D0	$W \to e\nu$	D0 W electron asymmetry 1.96 TeV	[65]	$\lvert\eta^e\rvert \le 2.9$,	$E_T^e > 25$
	$W \to \mu\nu$	D0 W muon asymmetry 1.96 TeV	[66]	$\lvert\eta^\mu\rvert \le 2.0$,	$P_T^\mu > 25$
	$Z \to e^+e^-$	D0 Z differential 1.96 TeV	[67]	$0.05 \le y_{ee} \le 2.75$, $71 \le M_{ee} \le 111$	

B. Drell-Yan datasets

The precise measurement of inclusive W^+, W^-, and Z/γ^* production in pp collisions at the LHC provides a stringent test of perturbative QCD. The rapidity dependence of boson production in the Drell-Yan process offers critical constraints on the PDFs of the proton, as the boson rapidity is strongly correlated with the momentum fractions x_1 and x_2 carried by the partons participating in the hard-scattering subprocess. The weak and electromagnetic components of the NC process, $Z/\gamma^* \to \ell\ell$, along with the weak CC reactions, $W^+ \to \ell^+\nu$ and $W^- \to \ell^-\bar{\nu}$, probe the quark flavor structure of the proton, complementing the information obtained from DIS. In this section, we present the Drell-Yan datasets included in our analysis.

1. LHC Drell-Yan data

The ATLAS Collaboration provides a detailed measurement of the differential cross section for the process $Z/\gamma^* \to \ell\ell$ (with $\ell = e, \mu$) as a function of the dilepton invariant mass in proton-proton collisions at $\sqrt{s} = 7$ TeV, using the ATLAS detector at the LHC [54]. This analysis, based on data collected in 2011, corresponds to an integrated luminosity of 1.6 fb^{-1}, covering invariant masses between 26 and 66 GeV in both the electron and muon channels. Additionally, data from 2010, with 35 pb^{-1} of integrated luminosity, extend the measurement down to 12 GeV in the muon channel. These low-mass Drell-Yan measurements provide crucial constraints on PDFs in the low-x region, complementing higher-mass analyses, and have been included in our analysis.

Building on this, ATLAS also analyzed a dataset focused on the high-mass region of Drell-Yan production, offering precise measurements of the differential cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV [55]. This analysis, based on an integrated luminosity of 4.9 fb^{-1}, examines the $Z/\gamma^* \to e^+e^-$ channel. The differential cross section is measured as a function of the invariant mass m_{ee} in the range $116\ \text{GeV} < m_{ee} < 1500\ \text{GeV}$, within a fiducial region where both the electron and positron have transverse momentum $p_T > 25$ GeV and pseudorapidity $|\eta| < 2.5$. These measurements extend the sensitivity of the analysis to higher x values, complementing the lower-mass data and have been applied in our QCD fit.

2. E866/NuSea (Fermilab)

In addition to the LHC measurements, the E866/NuSea experiment at Fermilab, which is included in this analysis, has made significant contributions to our understanding of flavor asymmetry in the nucleon sea. This experiment measured the ratio of Drell-Yan yields from an 800 GeV/c roton beam incident on liquid hydrogen and deuterium targets [35]. The extensive dataset, comprising approximately 360,000 Drell-Yan muon pairs, enabled the determination of the ratio of $\bar{d}$ to $\bar{u}$ quark distributions in the proton sea across a broad range of Bjorken-x. The results revealed a sharp downturn in the $\bar{d}(x)/\bar{u}(x)$ ratio at large x, leading to substantial revisions in global parametrizations of the nucleon sea and providing tighter constraints on valence PDFs. These findings have inspired further interest in extending measurements to higher x values using the 120 GeV/c proton beam from the Fermilab Main Injector.

It is important to note that while our analysis includes a wide variety of DIS and hadron collider datasets, certain Drell-Yan datasets, particularly from Relativistic Heavy Ion Collider (RHIC) and additional LHC measurements, are not included due to the lack of availability or validation within the xFitter framework. These omitted datasets include Drell-Yan data from LHCb and ATLAS at energies beyond 7 TeV [56,57]. The inclusion of such datasets in future analyses could further enhance the precision of the PDFs, especially in regions of the phase space that remain less constrained. To ensure the integrity and reliability of our results, we have focused exclusively on Drell-Yan datasets that have been rigorously tested and validated for use within xFitter.

C. Precision W/Z production collider data

In addition to DIS and Drell-Yan processes, recent collider data from the LHC and other sources provide complementary sensitivity for improving PDFs. In our QCD analysis, we incorporate W and Z boson production data from both the LHC (ATLAS and CMS collaborations) and the Tevatron (CDF and D0 collaborations), which offer crucial constraints on the quark flavor separation in PDFs, particularly for sea quarks.

The ATLAS Collaboration has provided two key datasets for this analysis:

- *Inclusive W/Z Production (ATLAS, 7 TeV)*: High-precision measurements of inclusive $W^+ \to \ell^+\nu$, $W^- \to \ell^-\bar{\nu}$, and $Z/\gamma^* \to \ell\ell$ ($\ell = e, \mu$) Drell-Yan production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV were published by ATLAS [53]. These measurements are based on data corresponding to an integrated luminosity of 4.6 fb^{-1}. The differential cross sections for W^+ and W^- are measured within the lepton pseudorapidity range $|\eta_\ell| < 2.5$, while the differential Z/γ^* cross sections are measured as a function of the absolute dilepton rapidity $|y_{\ell\ell}| < 3.6$ across three dilepton mass intervals ($46\ \text{GeV} < m_{\ell\ell} < 150\ \text{GeV}$). These measurements, combined across the electron and muon channels, provide stringent constraints on the PDFs of the proton when compared to theoretical predictions using modern PDF sets.
- *W/Z production and strange-quark density*: Another ATLAS dataset, focused on inclusive $W^\pm$ and Z boson production at $\sqrt{s} = 7$ TeV, has been critical in improving the understanding of the light quark sea, particularly the strange-quark density [58]. This analysis includes a NNLO perturbative QCD fit and

finds that the ratio of strange-to-down sea quark densities, r_s, is approximately 1.00 at $x=0.023$ and $Q_0^2 = 1.9$ GeV^2, suggesting a flavor-symmetric light quark sea at low x. These data play a crucial role in refining the strange-quark contribution to the proton sea quark content.

The CMS Collaboration has provided four key datasets that are included in our analysis:

Electron charge asymmetry (7 TeV): A precise measurement of the electron charge asymmetry in $pp \to W + X \to e\nu + X$ production at $\sqrt{s} = 7$ TeV, based on 840 pb^{-1} of data, was performed by CMS [59]. This asymmetry measurement, as a function of electron pseudorapidity $|\eta|$, imposes stringent constraints on PDFs by probing the differences in W^+ and W^- production.

Muon charge asymmetry (7 TeV): CMS also measured the muon charge asymmetry in inclusive $pp \to W + X$ production, based on 4.7 fb^{-1} of data [60]. The improved precision of this measurement, in 11 bins of muon pseudorapidity with different p_T thresholds, provides valuable input for determining valence and strange-quark distributions.

W boson production (8 TeV): A comprehensive measurement of the differential cross section and charge asymmetry for inclusive $pp \to W^{\pm} + X$ production at $\sqrt{s} = 8$ TeV, based on 18.8 fb^{-1} of data, was carried out by CMS [61]. These results, including differential cross sections with respect to p_T and lepton pseudorapidity, are essential for constraining both the valence and sea quark distributions.

Z boson production (7 TeV): CMS also measured the rapidity and transverse momentum distributions of Z bosons in pp collisions at $\sqrt{s} = 7$ TeV, using 36 pb^{-1} of data [62]. These measurements, with rapidity up to $|y| < 3.5$ and transverse momentum up to 350 GeV, provide critical insights into the production dynamics of Z bosons and serve as an important input for refining the gluon and sea quark PDFs.

Z boson production (7 TeV): CMS also measured the rapidity and transverse momentum distributions of Z bosons in pp collisions at $\sqrt{s} = 7$ TeV, using 36 pb^{-1} of data [62]. These measurements, with rapidity up to $|y| < 3.5$ and transverse momentum up to 350 GeV, provide critical insights into the production dynamics of Z bosons and serve as an important input for refining the gluon and sea quark PDFs.

Z boson production (7 TeV): CMS also measured the rapidity and transverse momentum distributions of Z bosons in pp collisions at $\sqrt{s} = 7$ TeV, using 36 pb^{-1} of data [62]. These measurements, with rapidity up to $|y| < 3.5$ and transverse momentum up to 350 GeV, provide critical insights into the production dynamics of Z bosons and serve as an important input for refining the gluon and sea quark PDFs.

Finally, the Tevatron datasets from CDF and D0 collaborations in which added to our data samples include:

CDF W boson production charge asymmetry: The CDF Collaboration measured the charge asymmetry in W boson production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV [63]. This dataset provides key insights into the asymmetry between W^+ and W^- production, helping to refine quark PDFs at intermediate x.

CDF Drell-Yan production: CDF also measured the differential cross section $d\sigma/dy$ for Drell-Yan e^+e^- pairs in the Z boson mass region at $\sqrt{s} = 1.96$ TeV [64], providing constraints on the quark-antiquark PDFs at intermediate x.

D0 electron charge asymmetry: D0 also measured the electron charge asymmetry in $W \to e\nu$ events, providing complementary constraints to the muon charge asymmetry data [65].

D0 muon charge asymmetry: D0 measured the muon charge asymmetry in $W \to \mu\nu$ events at $\sqrt{s} = 1.96$ TeV [66], offering critical constraints on the u and d quark PDFs.

Z boson rapidity distribution: Finally, D0 measured the rapidity distribution of $Z/\gamma^* \to e^+e^-$ in $p\bar{p}$ collisions [67], providing detailed tests of QCD and electroweak theory through the shape of the rapidity distribution.

IV. COMPUTATIONAL SETTING AND PDF PARAMETRIZATION

In this section, we describe the computational framework and the parametrization strategy adopted for the determination of PDFs. The accurate extraction of the PDFs requires a robust computational environment that integrates various aspects of the QCD analysis, including the evolution of parton densities, the computation of physical observables, and the fitting of theoretical predictions to experimental data. Our analysis is implemented within the xFitter framework [68,69], a versatile tool set designed for the global analysis of PDFs. This framework facilitates the numerical solution of the DGLAP evolution equations [70–72], ensuring consistency with perturbative QCD predictions up to NNLO accuracy. Moreover, xFitter allows for the incorporation of various sources of experimental uncertainties, including statistical, systematic, and correlated uncertainties, into the PDF extraction process. The PDF parametrization at the initial scale is a critical component of a QCD analysis, as it forms the basis for subsequent QCD evolution. We adopt a flexible functional form that captures the essential features of parton distributions while maintaining a balance between the number of free parameters and the stability of the fit. The optimization of the parametrization involves fitting the experimental data to the theoretical predictions, which are computed by evolving the PDFs from the initial scale to the relevant

experimental scales. This process is performed using the MINUIT minimization package [73,74], which provides both the central values of the parameters and their associated uncertainties.

A. PDF parametrization

In our QCD analysis, a flexible yet controlled functional form is employed to describe the wide range of experimental data without overfitting the PDFs. The initial scale is chosen to be $Q_0^2 = 1.9\ \mathrm{GeV}^2$, which lies below the charm mass threshold. This ensures that heavy-quark distributions are dynamically generated above their respective mass thresholds, which are set at $m_c = 1.43$ GeV for the charm quark and $m_b = 4.5$ GeV for the bottom quark. These thresholds are determined from the most recent heavy-quark differential cross section measurements conducted at HERA. As we mentioned before, our analysis follows the Thorne-Roberts (RT) VFNS [45,51,75], where heavy quarks are dynamically included in the PDFs once the energy scale surpasses their respective mass thresholds [28,29,45].

We use a HERAPDF-like parametrization for the valence and sea quark distributions, which has been well validated in previous HERA global QCD analyses [76]. The parametrizations for valence quarks x_{uv} and x_{dv} are constrained by the quark number sum rules, while the sea quark distributions are refined iteratively as new experimental datasets are included.

The gluon distribution poses significant challenges in the PDF parametrization, particularly at low x. Previous studies, such as those in the MSTW08 analysis [28,77], observed that when the PDF evolution begins from a low scale $Q_0^2 \sim 1\ \mathrm{GeV}^2$ the input gluon distribution can turn negative at very small values of x, outside the kinematic region of the fit, causing issues for the PDFs at very low x. To address this problem, and following the HERAPDF QCD analysis and similar work by MSTW08, we introduce an additional term in the gluon parametrization: $A_{g'}x^{B_{g'}}(1-x)^{C_{g'}}$. This term allows for greater flexibility in describing the gluon distribution, particularly in the small-x region. The inclusion of this correction is crucial for preventing unphysical negative gluon densities and stabilizing the fit across a wide range of x values.

The use of a two-component gluon distribution, represented by $xg(x) = xg_1(x) + xg_2(x) \sim A_g x^{B_g} + A_{g'}x^{B_{g'}}$, also offers enhanced accuracy in describing both the small- and large-x behavior of the gluon. By capturing the steep rise at low x and the expected falloff at high x, this parametrization improves the overall precision of the proton PDFs. It also ensures that the gluon density fits smoothly with the constraints provided by experimental data, particularly in regions where data are sparse, such as for very small x. This additional term ensures that the parametrization is sufficiently flexible at very low x [76].

A key aspect of understanding the proton structure is the determination of the strange-quark density. In our QCD analysis, we define the strange-quark fraction r_s as the ratio of the strange sea-quark distribution to the down sea-quark distribution [53]:

$$r_s = \frac{s+\bar{s}}{2\bar{d}}. \tag{5}$$

This parameter provides valuable insight into the relative contribution of strange quarks compared to down quarks in the proton sea. Recent ATLAS studies [53] have shown that r_s plays a crucial role in improving the precision of strange-quark distributions, which are critical for predictions of processes like W and Z boson production. Including r_s as a fit parameter allows us to extract precise information on the strange-quark content and its impact on other quark distributions. Our analysis refines this parameter to improve predictions for high-energy processes involving strange quarks.

In addition to the PDFs, the strong coupling constant α_s at the Z boson mass (M_Z) is treated as a free parameter in our QCD fit. This ensures that the coupling evolves consistently with the energy scale, allowing us to account for uncertainties in both α_s and the PDFs. The value of $\alpha_s(M_Z)$ is critical for accurately predicting cross sections in processes like jet and top quark production at the LHC, where gluon interactions dominate.

Finally, we adopt the following functional forms for the parton distributions at the initial scale Q_0^2:

$$\begin{aligned}
xu_v(x) &= A_{uv}x^{B_{uv}}(1-x)^{C_{uv}}(1+E_{uv}x^2),\\
xd_v(x) &= A_{dv}x^{B_{dv}}(1-x)^{C_{dv}},\\
x\bar{u}(x) &= A_{\bar{u}}x^{B_{\bar{u}}}(1-x)^{C_{\bar{u}}},\\
x\bar{d}(x) &= A_{\bar{d}}x^{B_{\bar{d}}}(1-x)^{C_{\bar{d}}},\\
xg(x) &= A_g x^{B_g}(1-x)^{C_g} - A_{g'}x^{B_{g'}}(1-x)^{C_{g'}},\\
xs(x) &= A_{\bar{s}}r_s x^{B_{\bar{s}}}(1-x)^{C_s}.
\end{aligned} \tag{6}$$

The above functional forms for xu_v, xd_v, $x\bar{u}$, $x\bar{d}$, and $x\bar{s}$ are tailored to ensure accurate predictions at both low and high values of x. In this analysis, we assume similar behavior of the up and down sea-quark distributions, setting $A_{\bar{d}} = A_{\bar{u}}$ and $B_{\bar{d}} = B_{\bar{u}}$. These assumptions are consistent with the HERAPDF-like [76] and ATLAS-epWZ16 [53] parametrizations for the sea quark distributions as well. For the strange distributions, we consider $xs(x) = x\bar{s}(x)$, and also assume $A_{\bar{s}} = A_{\bar{d}}$ and $B_{\bar{s}} = B_{\bar{d}}$. Given the enhanced sensitivity to the strange-quark distribution through the CMS and ATLAS W/Z datasets, the C_s appear as a free parameter in the fit. The above assumptions also simplify the parameter space, especially given the limited experimental data constraints on these distributions. Following the HERAPDF analysis, on which our current work is closely based, as well as the MSTW2008 analysis, we have fixed the parameter C_g' to

$C_g' = 25$. As discussed in Refs. [28,76], both analyses have demonstrated that the QCD fits are generally not sensitive to the exact value of C_g', provided that $C_g' \gg C_g$, ensuring that the additional term does not contribute significantly at large x.

B. Definition and minimization of χ^2 function

In this section, we focus on two essential aspects of our QCD analysis: the optimization procedure for determining the PDFs and the incorporation of the experimental uncertainties. The free parameters of the PDFs are estimated using experimental data through a likelihood maximization method. Assuming that the data points are normally distributed, the likelihood maximization reduces to minimizing the χ^2 function, which quantifies the agreement between theory and experiment.

The χ^2 function used in our analysis follows the formalism of the xFitter framework [68,69] and is expressed as

$$\chi^2(\{\zeta_k\}) = \sum_i \frac{[\mathcal{E}_i - \mathcal{T}_i(\{\zeta\})(1-\sum_j \gamma_j^i b_j)]^2}{\delta_{i,\mathrm{unc}}^2 \mathcal{T}_i^2(\{\zeta\}) + \delta_{i,\mathrm{stat}}^2 \mathcal{E}_i \mathcal{T}_i(\{\zeta\})(1-\sum_j \gamma_j^i b_j)} + \sum_i \ln \frac{\delta_{i,\mathrm{unc}}^2 \mathcal{T}_i^2(\{\zeta\}) + \delta_{i,\mathrm{stat}}^2 \mathcal{E}_i \mathcal{T}_i(\{\zeta\})}{\delta_{i,\mathrm{unc}}^2 \mathcal{E}_i^2 + \delta_{i,\mathrm{stat}}^2 \mathcal{E}_i^2} + \sum_j b_j^2, \tag{7}$$

where $\mathcal{E}_i$ represents the experimental measurements and $\mathcal{T}(\{\zeta_k\})$ are the theoretical predictions that depend on a set of parameters $\{\zeta_k\}$. The uncertainties in each measurement are separated into two components: $\delta_{i,\mathrm{stat}}$, the statistical uncertainty, and $\delta_{i,\mathrm{unc}}$, the uncorrelated systematic uncertainty, while γ_j^i denotes the correlated systematic uncertainties. The nuisance parameters b_j quantify the strength of the correlated systematic uncertainties across the dataset.

This formulation allows us to correctly incorporate both statistical and systematic uncertainties into the minimization process. The nuisance parameters b_α are introduced to control deviations in the correlated systematic uncertainties, with a penalty term $\sum_\alpha b_\alpha^2$ added to the χ^2 function to constrain these deviations. The minimization of this function is performed using the MINUIT [73,74] package via the MIGRAD algorithm [73], as implemented in xFitter. Once the minimum is found, the uncertainties in the fitted parameters are evaluated using the Hessian matrix, which is central to uncertainty propagation in the context of global QCD fits.

In a limited number of experimental studies, an alternative approach is employed wherein the covariance matrix, cov_{ij}, is used in place of Eq. (7) to account for correlations between data points,

$$\chi_E^2(a,\lambda) = \sum_{k=1}^{N_{\mathrm{pt}}} \frac{(D_k - T_k(a) - \sum_\alpha \lambda_\alpha \beta_{k\alpha})^2}{s_k^2} + \sum_\alpha \lambda_\alpha^2, \tag{8}$$

where D_k represents the experimental data, $T_k(a)$ represents the theoretical predictions for the kth data point, and λ_α are nuisance parameters that account for the correlated systematic uncertainties from source α. The total uncertainty s_k is the combination of statistical and uncorrelated systematic uncertainties. The λ_α parameters are optimized analytically during the minimization process. In this formulation, the covariance matrix cov_{ij} captures both uncorrelated and correlated uncertainties and is given by

$$\mathrm{cov}_{ij} = s_i^2 \delta_{ij} + \sum_\alpha \beta_{i\alpha} \beta_{j\alpha}, \tag{9}$$

where $\beta_{i\alpha}$ characterizes the sensitivity of the ith and jth data points to the correlated systematic uncertainty α. This decomposition allows for a proper treatment of experimental uncertainties, ensuring that correlated systematic effects are propagated consistently.

The minimization of the χ^2 function is followed by the propagation of uncertainties from the fit parameters to the physical observables. This is accomplished by generating a set of eigenvector PDF sets, which represent variations of the PDF parameters along the directions defined by the eigenvectors of the Hessian matrix. For N free parameters, a total of $2N + 1$ PDF sets are generated: one central fit and two variations along each eigenvector direction. Finally, the calculated proton PDF sets are made available in the standard LHAPDF format [78].

C. Determination of PDF uncertainties

The Hessian formalism is widely used in the global analysis of PDFs, such as in the CT18 framework, to study the uncertainties associated with the fitted PDFs and their physical predictions [79,80]. The method is based on a quadratic approximation to the χ^2 function around its minimum. The eigenvectors of the Hessian matrix characterize the directions in the parameter space in which the χ^2 increases most rapidly, corresponding to directions where the fit is most sensitive to the data. This allows us to generate a set of orthonormal basis PDFs, known as eigenvector PDF sets, which can be used to calculate uncertainties in physical observables by varying the fit parameters along each eigenvector direction [28].

In the Hessian formalism, the χ^2 function near its minimum can be expressed as a quadratic form,

$$\Delta\chi^2 = \chi^2(\boldsymbol{\zeta}) - \chi^2(\boldsymbol{\zeta}_0) \approx \sum_{i,j} \frac{\partial^2 \chi^2}{\partial \zeta_i \partial \zeta_j} (\zeta_i - \zeta_i^0)(\zeta_j - \zeta_j^0), \tag{10}$$

where $\boldsymbol{\zeta}_0$ represents the best-fit parameter vector, $\boldsymbol{\zeta}$ are the fit parameters, and $\frac{\partial^2 \chi^2}{\partial \zeta_i \partial \zeta_j}$ represents the second-order derivatives of χ^2 with respect to the parameters ζ_i. This second-order derivative matrix is called the Hessian matrix, denoted as H_{ij}:

$$H_{ij} = \frac{\partial^2 \chi^2}{\partial \zeta_i \partial \zeta_j}. \tag{11}$$

The Hessian matrix captures the shape and curvature of the χ^2 surface near its minimum, reflecting how sensitive the fit is to variations in the parameters. Diagonalizing this matrix yields the eigenvectors $\mathbf{v}_k$ and corresponding eigenvalues λ_k, which provide a natural basis for exploring the parameter space. The eigenvectors define directions in the parameter space along which the χ^2 increases most rapidly, indicating where the fit is most sensitive to variations in the parameters.

The parameter variations along the eigenvector directions are determined by solving the following eigenvalue problem:

$$H\mathbf{v}_k = \lambda_k \mathbf{v}_k. \tag{12}$$

In practice, these eigenvectors are used to construct a set of orthonormal basis PDFs, known as eigenvector PDF sets. These eigenvector sets allow for the systematic variation of PDF parameters and hence the estimation of uncertainties on physical observables. Once the eigenvectors of the Hessian matrix are known, the uncertainty in any physical observable $\mathcal{O}$ that depends on the PDFs can be calculated [28].

The eigenvector basis is constructed by transforming the parameters of the fit into a new set of coordinates where the Hessian matrix is diagonal. In this new coordinate system, the parameters are normalized such that displacements from the minimum in each direction correspond to fixed increases in χ^2. The uncertainty in any observable $\mathcal{O}$ is then calculated by evaluating the variation of $\mathcal{O}$ along each eigenvector direction. This method provides a robust framework for propagating uncertainties in the PDFs to predictions for physical observables.

The uncertainty in any observable $\mathcal{O}$ that depends on the PDFs is calculated as [28,79]

$$\Delta\mathcal{O} = \sqrt{\sum_{i=1}^{N} \left(\frac{\mathcal{O}_i^+ - \mathcal{O}_i^-}{2}\right)^2}, \tag{13}$$

where $\mathcal{O}_i^+$ and $\mathcal{O}_i^-$ are the observable values computed using the positive and negative variations of the ith eigenvector. This method assumes that the variations of the observables $\mathcal{O}$ can be approximated as linear within the region where the χ^2 function behaves quadratically. Such an assumption holds well near the minimum, ensuring reliable propagation of uncertainties from the PDFs to the final physical observables.

To ensure that the fit remains well behaved and does not overfit the data, especially in regions where the data are sparse or have large uncertainties, regularization methods are applied. In the xFitter framework, regularization is implemented through the inclusion of length penalty terms in the χ^2 function. These penalty terms discourage overly complex PDF shapes, promoting smoother distributions that are more physically reasonable. Regularization is particularly important in global PDF fits, where the data cover a wide range of kinematic regions, and the parametrization of the PDFs must be flexible enough to accommodate this diversity while avoiding overfitting.

Regularization also plays a crucial role in determining the appropriate degree of flexibility in the parametrization of the PDFs. Too much flexibility can lead to poorly constrained parameters and instabilities in the fit, while too little flexibility may prevent the fit from adequately describing the data. By introducing regularization terms, we can control the complexity of the PDF parametrization and ensure that the fit remains stable and reliable across the full kinematic range.

V. FIT RESULTS

This section presents the main results and findings of this work. As outlined below, different QCD fits have been performed, considering various datasets. We first examine the quality of our QCD fits, then present the extracted PDFs, and finally compare the fit quality across a selection of fitted data for different scenarios.

To explore the impact of Drell-Yan and W/Z boson production data from the CMS, ATLAS, D0, and CDF collaborations on the shape of different parton species and their associated uncertainties, we performed four different QCD fits at NLO and NNLO accuracy. The details of these approaches are described below:

(i) Fit A. In the first fit, we incorporate only the HERA I + II [52] experimental data, which serves as a crucial baseline for any QCD analysis. Utilizing 1016 data points, as per the Q^2 cut mentioned earlier, the results from Fit A provide a robust foundation for assessing the impact of additional experimental measurements on the PDFs.

(ii) Fit B. In this fit, we build on Fit A by including measurements related to the Drell-Yan process from the ATLAS Collaboration [54,55] and the E866 experiment [35]. The configuration of this fit remains identical to Fit A, with the only difference being the inclusion of Drell-Yan data. This fit utilizes 1082 data points, as indicated in Table II, allowing us to evaluate how Drell-Yan measurements affect the shape of different parton species and their associated uncertainties.

(iii) Fit C. Fit C is similar to Fit B, but instead of Drell-Yan data, it includes CMS and ATLAS measurements of W/Z boson cross sections [53,58–62], as well as the corresponding measurements from the D0 and CDF collaborations at Tevatron [64–67]. This fit is performed using 1289 data points,

TABLE II. The extracted numerical values for the correlated χ^2, log penalty χ^2, and the total χ^2 per degree of freedom (d.o.f.) associated with each experimental measurement. The table also lists the input datasets included in the analysis of our PDFs at NNLO accuracy. For each dataset, the name of the experiment and the corresponding reference are also provided.

Observable	Experiment	Reference	Fit A	Fit B	Fit C	Fit D
DIS	HERA1 + 2 CC e^+p	[52]	46/39	42/39	44/39	43/39
	HERA1 + 2 CC e^-p	[52]	55/42	60/42	69/42	68/42
	HERA1 + 2 NC e^-p	[52]	222/159	224/159	220/159	222/159
	HERA1 + 2 NC e^-p 460	[52]	195/177	195/177	195/177	197/177
	HERA1 + 2 NC e^-p 575	[52]	187/221	188/221	187/221	189/221
	HERA1 + 2 NC e^+p 820	[52]	53/61	55/61	53/61	53/61
	HERA1 + 2 NC e^+p 920	[52]	351/317	361/317	359/317	379/317
Drell-Yan	ATLAS high mass	[55]	⋯	13/13	⋯	9.2/13
	ATLAS low mass 1.6 fb^{-1}	[54]	⋯	7.5/8	⋯	8.7/8
	ATLAS low mass extended 35 pb^{-1}	[54]	⋯	6.4/6	⋯	7.4/6
	E866 low mass	[35]	⋯	11/10	⋯	14/10
	E866 mid mass	[35]	⋯	14/14	⋯	13/14
	E866 high mass	[35]	⋯	13/15	⋯	23/15
	ATLAS W^+	[58]	⋯	⋯	15/11	15/11
	ATLAS W^+	[58]	⋯	⋯	15/11	15/11
	ATLAS W^-	[58]	⋯	⋯	8.9/11	8.9/11
	ATLAS W^+	[53]	⋯	⋯	13/11	13/11
	ATLAS W^-	[53]	⋯	⋯	10/11	9.6/11
	CMS W^+	[61]	⋯	⋯	1.6/11	3.3/11
	CMS W^-	[61]	⋯	⋯	1.2/11	2.3/11
	CMS W asymmetry	[61]	⋯	⋯	7.5/11	11/11
	CMS W asymmetry	[59]	⋯	⋯	10.0/11	9.2/11
	CMS W asymmetry	[60]	⋯	⋯	13/11	14/11
W/Z bosons	CDF W asymmetry	[63]	⋯	⋯	38/13	47/13
	D0 $W \to e\nu$ asymmetry	[65]	⋯	⋯	31/13	31/13
	D0 $W \to \mu\nu$ asymmetry	[66]	⋯	⋯	15/10	14/10
	ATLAS high mass CF Z	[53]	⋯	⋯	3.9/6	3.9/6
	ATLAS high mass CC Z	[53]	⋯	⋯	5.6/6	5.5/6
	ATLAS low mass Z	[53]	⋯	⋯	22/6	23/6
	ATLAS peak CF Z	[53]	⋯	⋯	5.5/9	8.4/9
	ATLAS peak CC Z	[53]	⋯	⋯	13/12	15/12
	ATLAS Z	[58]	⋯	⋯	2.9/8	2.5/8
	CMS Z	[62]	⋯	⋯	65/35	67/35
	CDF Z	[64]	⋯	⋯	28/28	28/28
	D0 Z	[67]	⋯	⋯	22/28	22/28
Correlated χ^2			49	61	95	113
Log penalty χ^2			−10.87	−17.10	−5.05	−13.22
Total χ^2/d.o.f.			1149/1000 = 1.149	1236/1066 = 1.159	1551/1273 = 1.218	1678/1339 = 1.253

enabling us to study the specific impact of W/Z boson production data on the extracted PDFs.

(iv) Fit D. The final, or nominal, fit is performed using the combination of all the datasets discussed above, providing the most comprehensive QCD analysis. This fit integrates the HERA I + II, Drell-Yan, and W/Z boson datasets, yielding a complete picture of the PDFs and their associated uncertainties. The

PDFs in this fit are extracted from 1355 experimental data points.

A. Fit quality

The χ^2 values characterizing our NNLO nominal QCD fit to the H1/ZEUS combined datasets as well as the Drell-Yan and W/Z production data from the LHC and Tevatron are listed in Table II. As shown, our NNLO QCD fits provide a good description of the datasets both in terms of individual and total datasets from HERA, LHC, and Tevatron. Several observations can also be made based on the χ^2 values for both the individual and total datasets.

For Fit A, where we utilize only the HERA data, the total χ^2/d.o.f. is 1.149, indicating a good quality QCD fit within the chosen PDF parametrization of Eq. (6). This baseline fit sets the standard against which the inclusion of additional datasets is evaluated.

The inclusion of Drell-Yan data in Fit B slightly increases the total χ^2 value to 1.159. Despite this slight increase, the overall fit quality remains very good, and the description of the individual datasets is still satisfactory. This modest change suggests that the additional Drell-Yan data are largely compatible with the existing HERA data, with minimal tension between the datasets.

When combining the HERA data with the W/Z boson production cross section data in Fit C, we observe a further increase in the total χ^2 to 1.218. While the overall fit quality remains acceptable, some individual datasets show slight discrepancies. Specifically, the CDF W asymmetry data [63] and the ATLAS low-mass Z data [53] show less agreement with our NNLO theory predictions, contributing to the increase in the total χ^2.

In our nominal fit, Fit D, which incorporates all datasets discussed, the total χ^2 increases slightly to 1.253. This fit includes the effects of all combined data, and the observed increase is primarily driven by the tension between certain datasets. In addition to the CDF W asymmetry [63] and ATLAS low-mass Z data [53], a noticeable increase in the χ^2 value is also seen for the HERA I + II NC e^+p 920 dataset [52].

Overall, while there are slight increases in the total χ^2 values with the addition of new datasets, the fits remain well within acceptable ranges, confirming that our approach provides a consistent and comprehensive description across a wide range of experimental data.

We should also highlight here that the overall fit quality is improved at NNLO compared to NLO accuracy. For our nominal fit, Fit D, at NLO accuracy, the total χ^2 value is found to be higher ($\chi^2 = 1.270$), indicating a less precise description of the data. In contrast, the NNLO fit provides much closer agreement with the experimental measurements, as reflected in the lower χ^2 values for both the individual and combined datasets.

This demonstrates the importance of incorporating higher-order QCD corrections to achieve a more accurate determination of the PDFs and their associated uncertainties. The difference between NNLO and NLO is mainly due to the fact that perturbative corrections are generally more significant for hadron-collider processes compared to HERA DIS, emphasizing the necessity of including higher-order QCD corrections, especially for high-precision LHC measurements. Additionally, the quantity and precision of the current LHC data are now sufficient to clearly demonstrate the superiority of NNLO calculations.

B. Parton distributions

We now turn to the examination of our extracted PDFs. First, we present the complete set of four PDFs and compare them to the baseline Fit A results. Following this, we provide a detailed discussion on the impact of different dataset selections on the shapes of various parton species and their associated uncertainties.

In Table III, we list the numerical values of the best-fit parameters and their associated uncertainties, obtained from our QCD fits at the input scale $Q_0^2 = 1.9$ GeV^2. The parameter values are provided for all four QCD fit sets discussed in the text. The strong coupling constant, $\alpha_s(M_Z)$, is treated as a free parameter in our QCD analyses. The value of $\alpha_s(M_Z)$ extracted from our QCD fits is also presented in Table III, reflecting the impact of including different datasets on the determination of the strong coupling constant. Additionally, as discussed earlier, the parameter r_s, defined as the relative strange-to-down sea quark fraction $r_s = \frac{s+\bar{s}}{2\bar{d}}$ is also treated as a free parameter in our QCD fits. The values of r_s extracted from the fits are presented in Table III. This parameter is essential for understanding the relative behavior of the strange and down sea quark distributions, and its inclusion allows for a more flexible parametrization of the PDFs.

Several observations can be made based on the numbers presented in this table. We parametrize six different PDFs: the valence distributions xu_v and xd_v; the gluon density xg; the sea quark densities $x\bar{d}$ and $x\bar{u}$; and finally, a symmetric strange-quark distribution $xs = x\bar{s}$. The parameters A_{u_v}, A_{d_v}, and A_g are determined by the number and momentum sum rules.

As observed, nearly all parameters are well determined from the fit to the data. However, some parameters, such as C_s, $C_{\bar{d}}$, and C_g, exhibit relatively large uncertainties, particularly when only the HERA data are used. This indicates that the HERA combined DIS datasets alone do not sufficiently constrain these parameters. Notably, as shown in Table III, the inclusion of additional datasets, such as Drell-Yan and W-/Z-production data, significantly improves the determination of these parameters, reducing their associated uncertainties.

In the following discussion, we turn our attention to the obtained PDFs and their uncertainties. Detailed comparisons of the NNLO quark and gluon PDFs at the input scale $Q_0^2 = 1.9$ GeV^2 are presented in Figs. 1 and 2. To investigate the impact of including different data selections, we

TABLE III. The numerical values of the best-fit parameters and their associated uncertainties obtained from our QCD fits at the input scale $Q_0^2 = 1.9$ GeV^2. The parameter values are provided for all four different QCD fit sets discussed in the text.

Parameter	Fit A	Fit B	Fit C	Fit D
$A_{\bar{d}}$	0.152 ± 0.017	0.190 ± 0.020	0.1542 ± 0.0077	0.1672 ± 0.0085
$A_{g'}$	0.376 ± 0.089	0.0042 ± 0.0029	0.353 ± 0.056	0.187 ± 0.013
$B_{\bar{d}}$	-0.127 ± 0.015	-0.095 ± 0.018	-0.1167 ± 0.0083	-0.1043 ± 0.0089
B_{dv}	0.903 ± 0.094	1.019 ± 0.095	0.711 ± 0.019	0.775 ± 0.015
B_g	-0.48 ± 0.18	1.08 ± 0.50	-0.789 ± 0.021	-0.854 ± 0.016
$B_{g'}$	-0.57 ± 0.12	-0.839 ± 0.088	-0.817 ± 0.017	-0.877 ± 0.010
B_{uv}	0.752 ± 0.034	0.872 ± 0.039	0.736 ± 0.011	0.7650 ± 0.0085
$C_{\bar{d}}$	8.7 ± 5.7	6.42 ± 0.48	3.66 ± 0.25	5.68 ± 0.32
C_{dv}	4.53 ± 0.43	4.08 ± 0.41	4.239 ± 0.085	4.364 ± 0.083
C_g	3.5 ± 1.2	3.9 ± 1.4	3.8 ± 1.0	0.57 ± 0.20
$C_{g'}$	25 (Fixed)	25 (Fixed)	25 (Fixed)	25 (Fixed)
C_s	8.6 ± 7.5	25.7 ± 7.8	13.1 ± 1.6	16.4 ± 1.5
$C_{\bar{u}}$	5.04 ± 0.82	10.14 ± 0.68	5.83 ± 0.29	8.34 ± 0.46
C_{uv}	4.840 ± 0.097	4.561 ± 0.082	4.954 ± 0.078	4.943 ± 0.067
E_{uv}	11.3 ± 1.5	11.3 ± 2.2	9.51 ± 0.55	9.86 ± 0.59
$\alpha_s(M_Z)$	0.1170 ± 0.0038	0.13005 ± 0.001	0.1075 ± 0.0030	0.1128 ± 0.0014
r_s	0.72 ± 0.47	1.06 ± 0.41	1.025 ± 0.071	1.069 ± 0.053

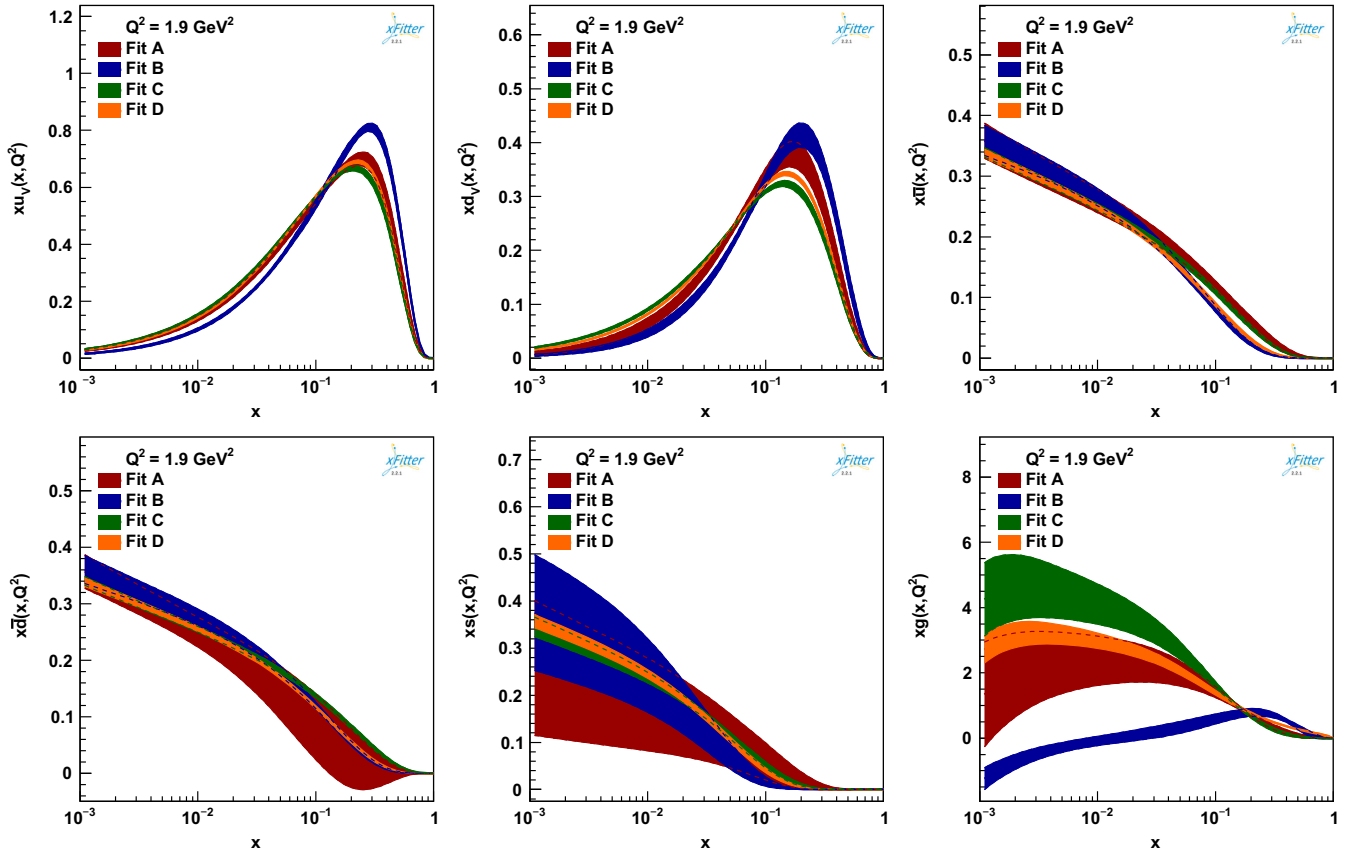

FIG. 1. The parton distribution of xu_v, xd_v, $x\bar{u}$, $x\bar{d}$, xs, and xg for Fit A, Fit B, Fit C, and Fit D as a function of x and at the input scale $Q_0^2 = 1.9$ GeV^2.

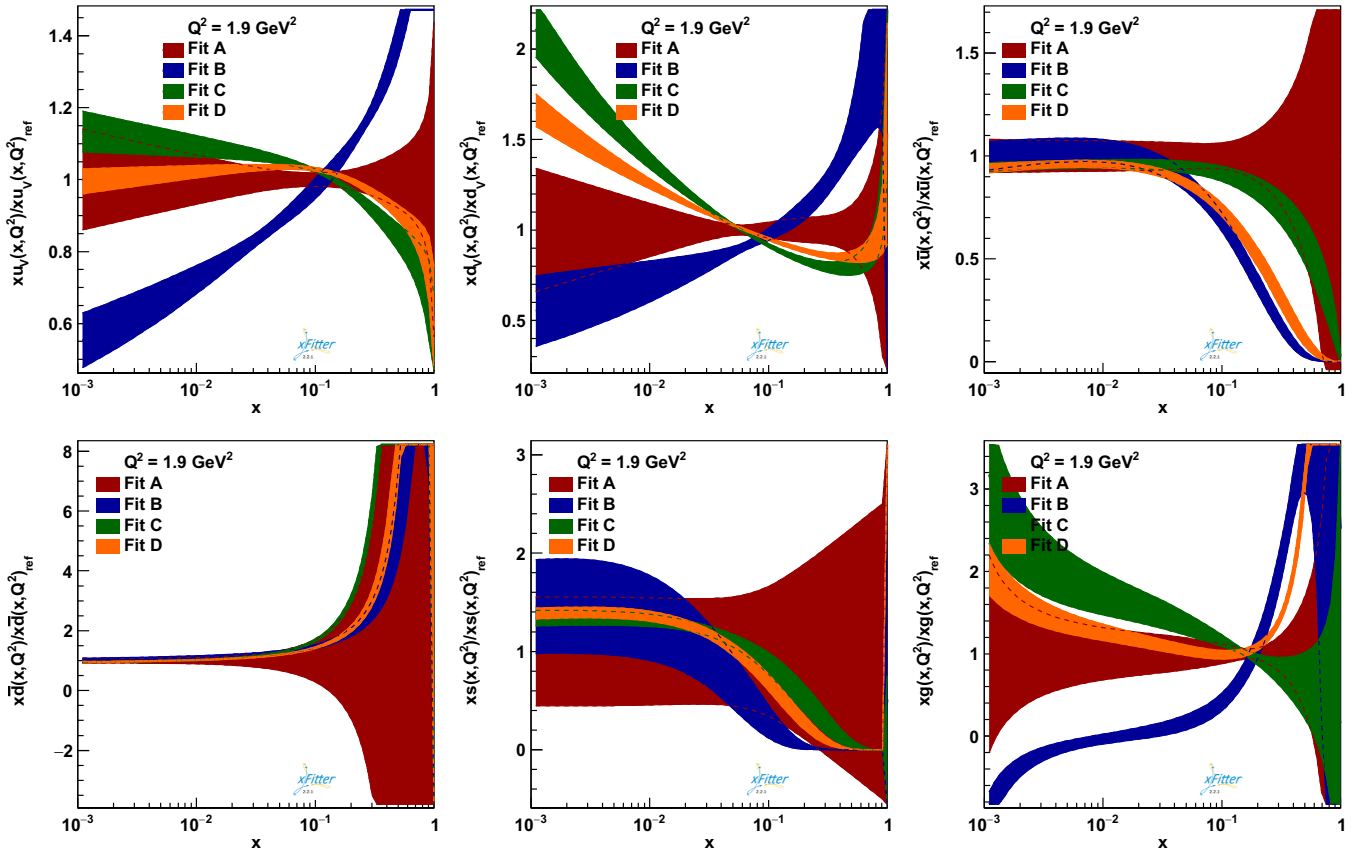

FIG. 2. The ratios of parton distribution $(xf(x,Q^2)/xf(x,Q^2)_{\text{ref}})$ for xu_v, xd_v, $x\bar{u}$, $x\bar{d}$, xs, and xg as a function of x and at the input scale $Q_0^2 = 1.9\ \text{GeV}^2$. The comparisons are displayed as ratios $xf(x,Q^2)/xf(x,Q^2)_{\text{ref}}$ relative to our baseline QCD analysis, Fit A.

compare all four PDF determinations with each other. Specifically, in Fig. 2, the comparisons are displayed as ratios $xf(x,Q^2)/xf(x,Q^2)_{\text{ref}}$ relative to our baseline QCD analysis, Fit A. The uncertainty bands of the PDFs are calculated using the Hessian method, as discussed earlier.

Notable differences in the shapes of the PDFs are observed, especially at small values of x. These variations indicate that different datasets impose different types of constraints on the PDFs. For example, the inclusion of Drell-Yan and W/Z boson production data tends to impact the strange-quark and gluon distributions more prominently at lower x, demonstrating the sensitivity of the PDF shapes to the specific experimental inputs.

As can be seen, the inclusion of the Drell-Yan data in Fit B significantly affects the shape of the valence quark distributions, particularly over medium to large values of x. These differences are even more pronounced for the strange quark and gluon distributions, where notable variations are observed in both the shape and the uncertainty bands across the entire range of x.

Focusing first on the strange-quark PDF, it is evident that there are relatively poor constraints in Fit A, which relies solely on the HERA data. Even with the addition of Drell-Yan datasets in Fit B, the constraints on the strange quark remain limited, as reflected by the large associated uncertainty bands. However, the inclusion of LHC and Tevatron W/Z boson production cross section data in Fit C leads to much better constraints and significantly reduced error bands for the strange-quark PDF across all x values.

For the gluon PDFs, the differences are even more pronounced. The inclusion of additional datasets, particularly those from W/Z boson production, affects the central values; however, it does not lead to a significant reduction in the uncertainty bands. This highlights the critical role of jet production datasets from collider DIS and hadron colliders in providing stronger constraints on the gluon distribution across both small and large values of x.

In our QCD fits discussed above, we chose not to include jet production datasets, as the main motivation of this work is to specifically examine the constraints on individual PDFs from HERA DIS, Drell-Yan, and W/Z boson production datasets. This approach allows us to isolate and better understand the individual contributions and the impact of these particular datasets on the determination of PDFs. The inclusion of jet and dijet production data in our QCD fits will be discussed in Sec. VII, where we will explore their effects on the extracted PDFs and the strong coupling constant.

Detailed comparisons of the NNLO quark and gluon PDFs are presented in Figs. 3–6, this time at higher values

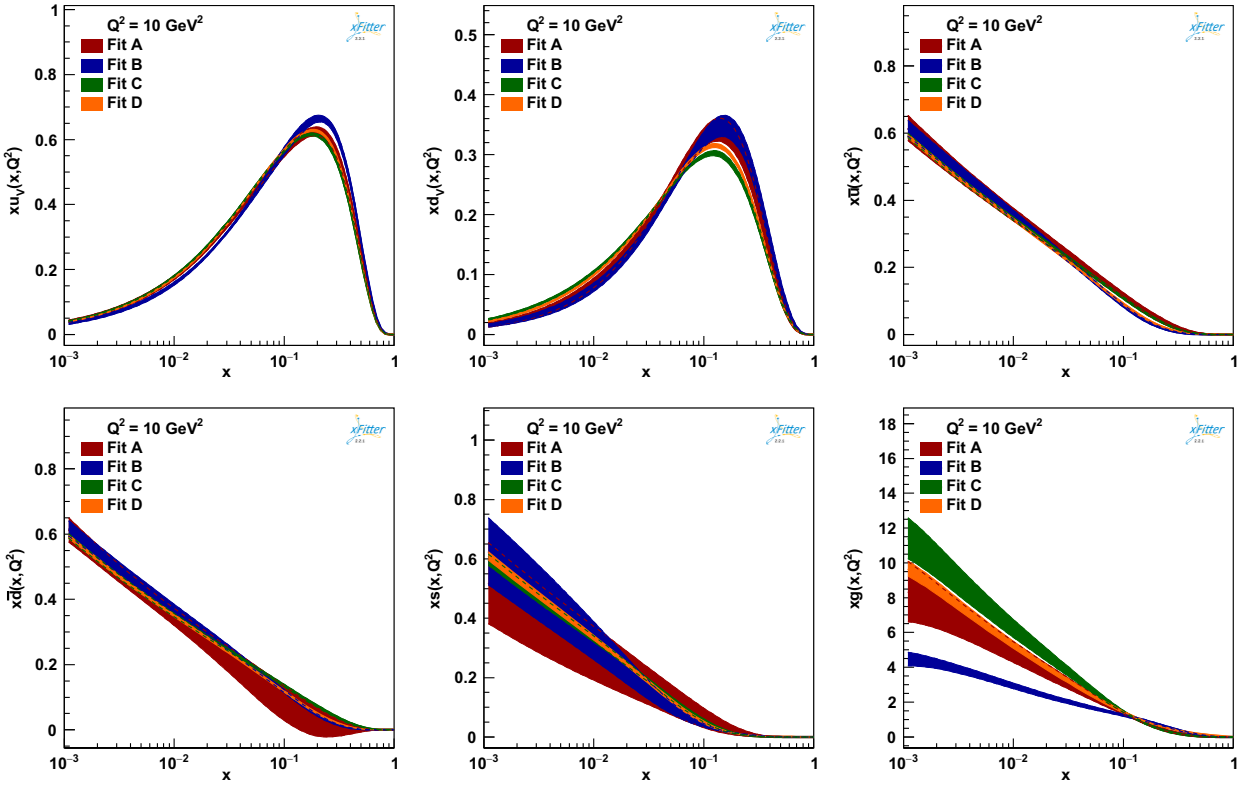

FIG. 3. Same as Fig. 1 but this time for $Q^2 = 10$ GeV^2.

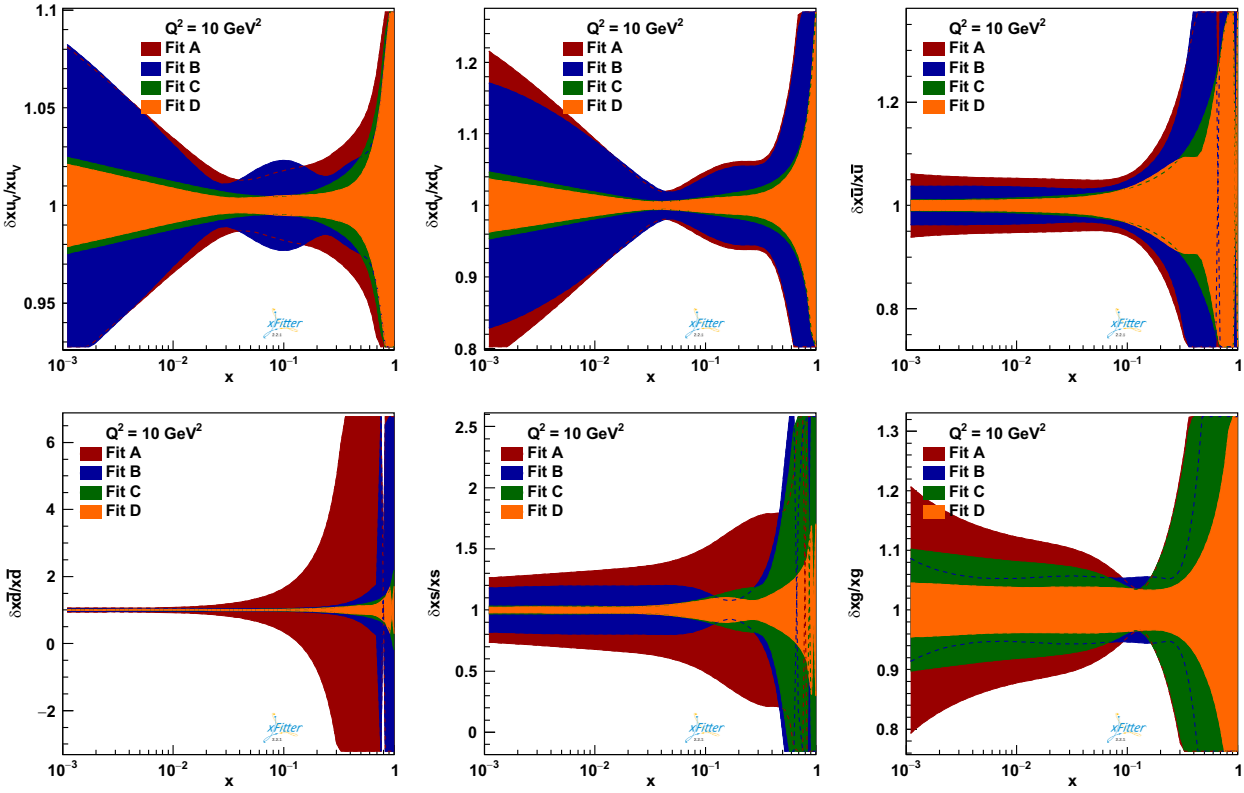

FIG. 4. Same as Fig. 2 but this time for $Q^2 = 10$ GeV^2. The relative uncertainties $\delta xq(x, Q^2)/xq(x, Q^2)$ are also shown.

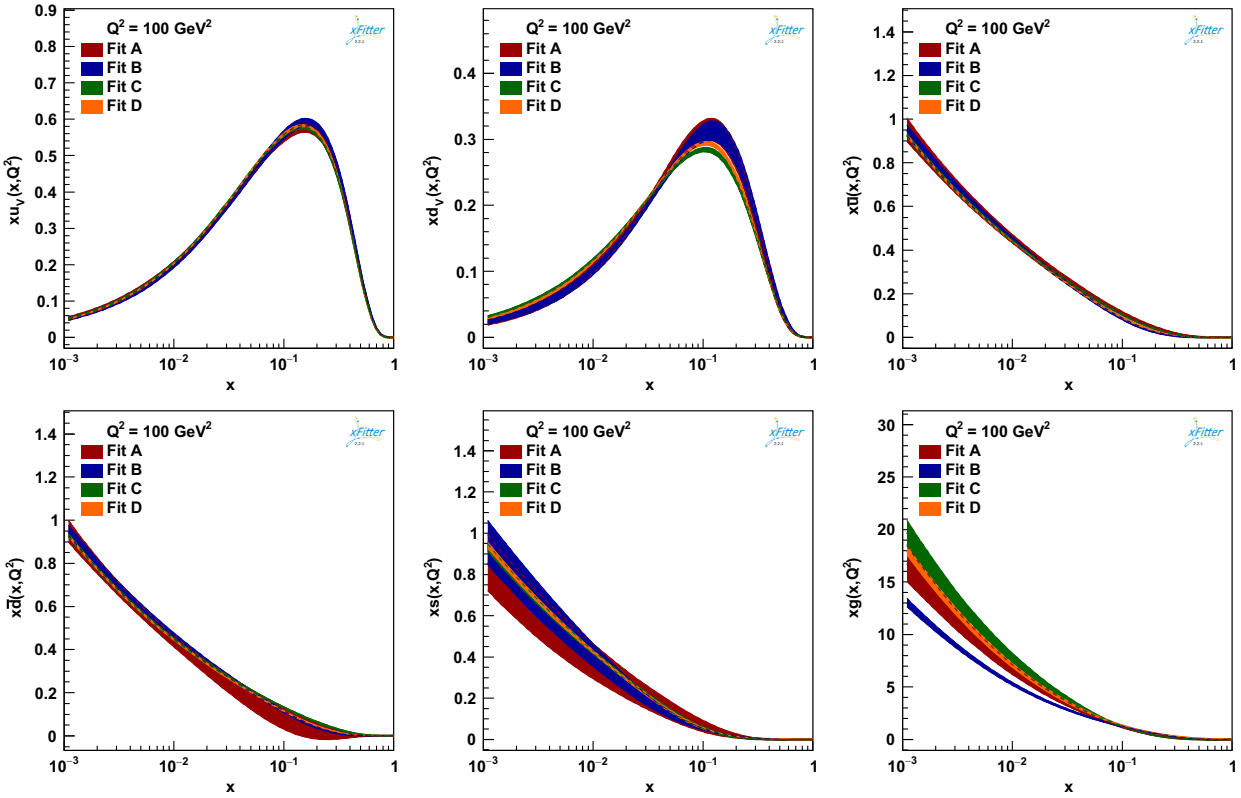

FIG. 5. Same as Fig. 1 but this time for $Q^2 = 100$ GeV^2.

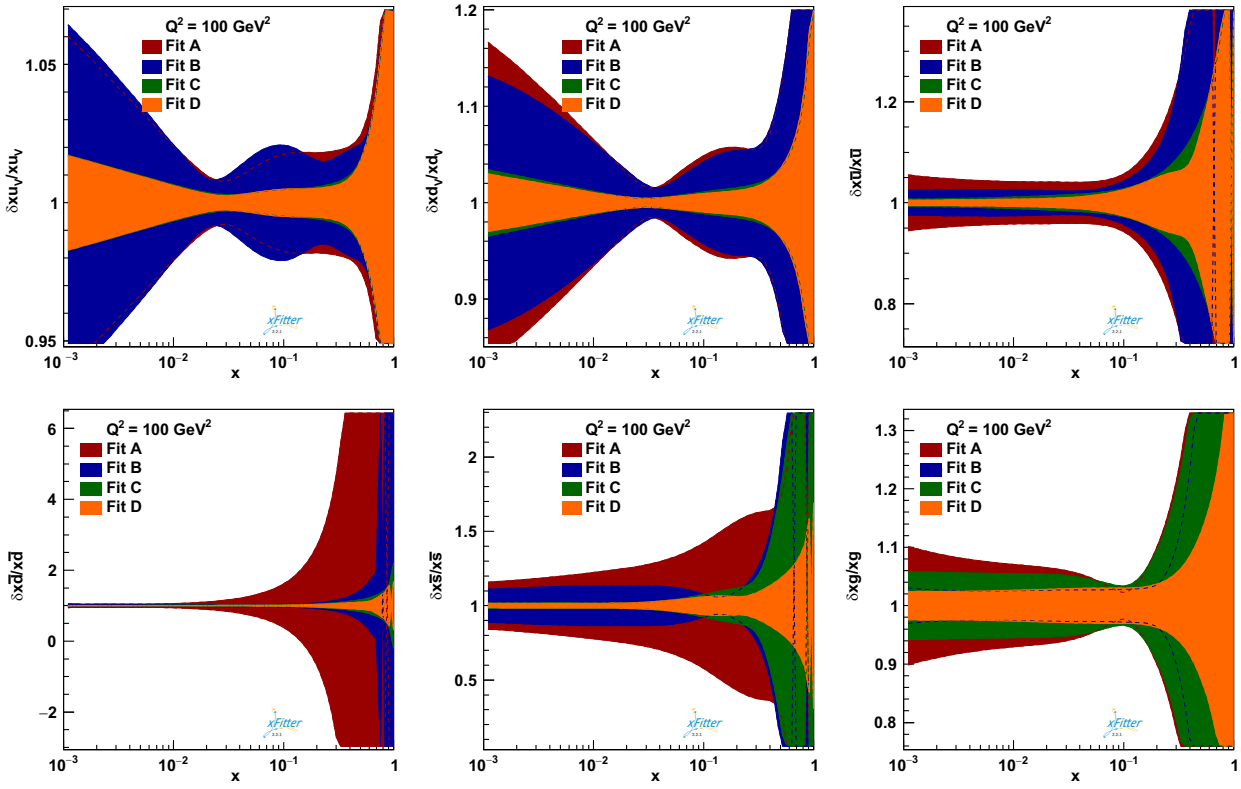

FIG. 6. Same as Fig. 2 but this time for $Q^2 = 100$ GeV^2. The relative uncertainties $\delta xq(x, Q^2)/xq(x, Q^2)$ are also shown.

of $Q^2 = 10$ and $100\ \mathrm{GeV}^2$, respectively. The general observations discussed earlier remain valid for the PDFs at these higher scales as well. However, it is important to highlight a key finding apparent in these plots regarding the combined inclusion of all datasets. As shown, incorporating all the datasets together results in more significant constraints on the PDFs, which in turn leads to noticeably smaller uncertainty bands for all parton species, as illustrated in the ratio plots of Figs. 4 and 6.

C. Impact of higher-order QCD corrections

In this section, we discuss the effects of higher-order QCD corrections on the extracted PDFs and their associated uncertainties. As presented in Table II, the overall fit quality remains excellent for both individual and global datasets for NNLO accuracy, indicating that the NNLO corrections result in a good description of the experimental data across a wide range of observables.

The inclusion of NNLO corrections is vital for achieving high-precision results, especially for processes such as Drell-Yan production, W/Z boson production, and jet production, which are sensitive to higher-order effects. The CT18, MSHT20, and NNPDF4.0 analyses also emphasize the importance of NNLO corrections in reducing theoretical uncertainties and improving the overall agreement between theory and experimental data. For example, the MSHT20 analysis shows improvements in the description of LHC data, particularly for gluon-sensitive processes, while NNPDF4.0 utilizes advanced machine-learning techniques to incorporate NNLO corrections and reduce biases in PDF parametrizations.

In Fig. 7, we show the PDFs at both NNLO and NLO as functions of x for three different Q^2 values: 1.9, 10, and $100\ \mathrm{GeV}^2$. The uncertainty bands, calculated using the Hessian method, illustrate the effect of higher-order corrections on the PDFs. The impact of NNLO QCD corrections on the central values of the PDFs is generally modest relative to the uncertainties; however, differences can be observed in specific parton distributions. For instance, the total singlet distribution is larger at NNLO than at NLO, particularly in the small to medium x range. The gluon PDF shows a slight reduction at NNLO in the small to medium x region, reflecting the sensitivity of gluon-dominated processes to higher-order effects.

Additionally, as Fig. 8 demonstrates, the changes in the central values of the PDFs due to NNLO corrections are most pronounced at lower values of x, where higher-order effects are more substantial. Both the total singlet and gluon distributions exhibit the largest deviations from their NLO counterparts.

While the central values of the PDFs shift slightly due to NNLO corrections, the associated uncertainties remain largely unaffected. This stability is evident for all parton species, as the uncertainty bands at NLO and NNLO are of comparable size. The χ^2 values in Table IV confirm that the NNLO fits provide only a modest improvement in the overall uncertainty reduction. This suggests that the inclusion of NNLO corrections does not drastically reduce uncertainties but primarily leads to shifts in the central values, particularly for the singlet and gluon PDFs.

Table IV also compares the χ^2 values obtained from the NLO and NNLO fits for the different datasets analyzed in this work. As shown, the total χ^2 is slightly lower at NNLO, with a total $\chi^2/\mathrm{d.o.f.}$ of $1678/1339$ compared to $1699/1339$ at NLO. This improvement is consistent with the results from the CT18 and MSHT20 studies, where NNLO corrections lead to better overall fit quality and improved agreement with experimental data.

Notably, the fit to the ATLAS W/Z data shows a more pronounced improvement at NNLO, with a $\chi^2/\mathrm{d.o.f.}$ of $104.8/91$ compared to $108.6/91$ at NLO. This indicates that NNLO corrections are particularly important for precision measurements of electroweak processes, where small corrections can significantly improve the fit. Similarly, the χ^2 for the HERA $\mathrm{I}+\mathrm{II}$ dataset shows a slight improvement at NNLO ($1151/1016$ vs $1155/1016$ at NLO).

Overall, the χ^2 values demonstrate that the inclusion of NNLO corrections leads to modest improvements in fit quality across most datasets, confirming the necessity of higher-order corrections for precise determinations of PDFs, particularly in processes dominated by gluon-gluon or quark-gluon interactions. While NLO fits remain reasonably accurate for many applications, NNLO fits are essential for precision studies at high-energy colliders such as the LHC and future experiments like the EIC, LHeC, and FCC (he, hh).

D. Comparison to other PDF sets

In this section, we compare our nominal NNLO PDFs (Fit D) with other recent global sets, specifically NNPDF4.0 [14], CT18 [12], and MSHT20 [16]. All results are presented at $Q^2 = 1.9$, 10, and $100\ \mathrm{GeV}^2$, normalized to our nominal (Fit D) NNLO PDFs. The PDF uncertainties are consistently shown at the 68% confidence level (CL) for all sets, including those from the comparison groups, ensuring a fair basis for comparison. This consistency in uncertainty presentation is intended to provide a clear understanding of the relative uncertainties across the different PDF sets.

As discussed in Sec. IV A, our analysis includes six independently parametrized distributions: xu_v, xd_v, $x\bar{u}$, $x\bar{d}$, $x\bar{s}$, and xg. Similar to CT18 and MSHT20, we rely on a dynamically generated charm distribution through the VFNS, rather than independently parametrizing the charm quark. In contrast, NNPDF4.0 independently parametrizes eight PDFs, which include both the strange and charm distributions. It is also important to note that there are substantial differences in the underlying datasets used in these QCD analyses.

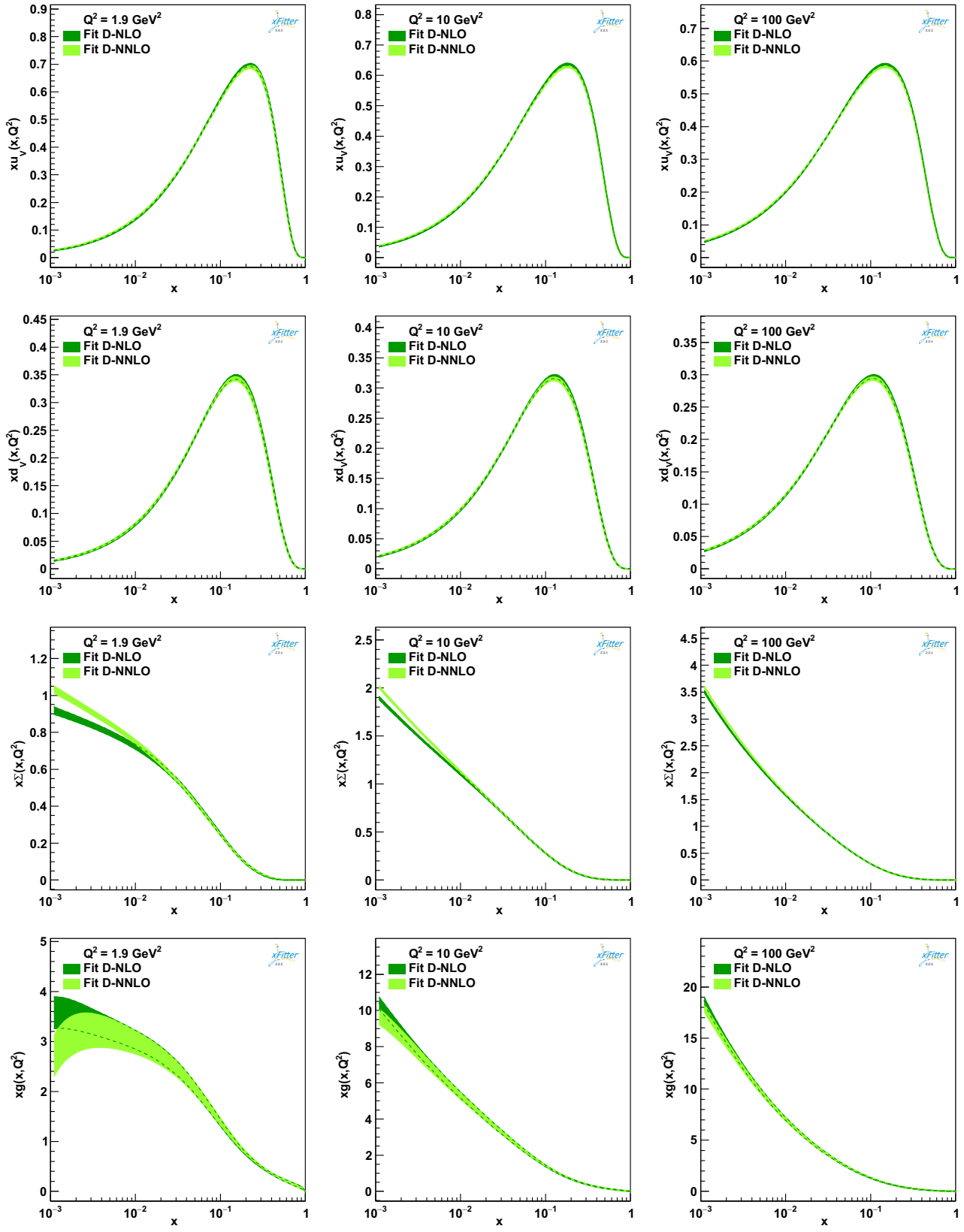

FIG. 7. The parton distribution of xu_v, xd_v, $x\Sigma$, and xg for our NLO and NNLO QCD fits as a function of x and at $Q^2 = 1.9$, 10, and 100 GeV^2.

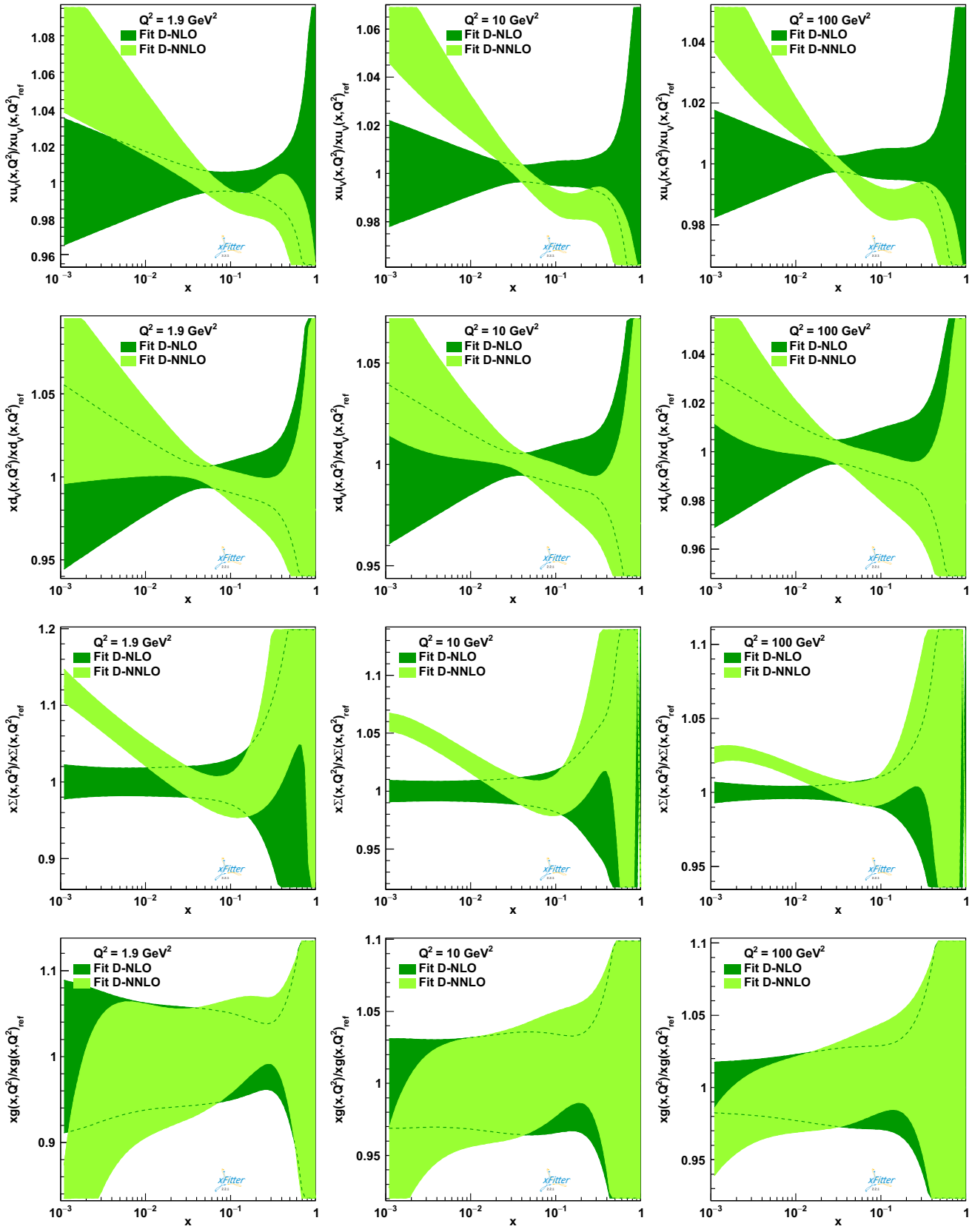

FIG. 8. The ratio of parton distribution $xf(x,Q^2)/xf(x,Q^2)_{\mathrm{ref}}$ as a function of x and at $Q^2 = 1.9$, 10, and 100 GeV^2.

TABLE IV. The values of χ^2 obtained from our NLO and NNLO fits for the different datasets analyzed in this work.

Experiment	NLO (Fit D)	NNLO (Fit D)
HERA I + II	1155/1016	1151/1016
CMS W/Z	106.4/90	106.8/90
ATLAS W/Z	108.6/91	104.8/91
ATLAS Drell-Yan	26.3/27	25.3/27
D0W/Z	68/51	67/51
CDF W/Z	73/41	75/41
E866 Drell-Yan	50/39	50/39
Correlated χ^2	114	113
Log penalty χ^2	−1.26	−13.22
Total χ^2/d.o.f.	1699/1339	1678/1339

Overall, the four parton sets shown in these plots are generally in good agreement within their respective uncertainties, though some differences in shape are observed. These differences are more pronounced at the input scale, $Q_0^2 = 1.9$ GeV^2, particularly for the total singlet $x\Sigma$ and gluon xg distributions. For the valence densities, xu_v and xd_v, NNPDF4.0 is slightly suppressed at intermediate values of x compared to the others. Notably, our gluon density is larger than the others, especially at small values of x. Our valence and total singlet densities are in good agreement and are consistently within the uncertainty envelopes of CT18 and MSHT20.

More pronounced differences are observed for the gluon distribution. However, it is important to note that all other PDF sets include jet data, which provide additional constraints on the gluon PDF. In contrast, our Fit D does not include any jet production datasets, as our primary goal is to investigate the impact of Drell-Yan and W/Z boson production data on the shape of different parton species and their associated uncertainties. The gluon distribution in our analysis is in fair agreement in the small-x region, which is relevant for dominant Higgs boson production at the LHC.

More interesting findings emerge from Fig. 9 when comparing PDF uncertainties. Our uncertainties for the valence and total singlet distributions are smaller than those of all other sets across the entire x region. However, the associated uncertainty for our gluon PDF is generally larger compared to the uncertainties of NNPDF4.0 and MSHT20 for $x < 0.1$. The CT18 analysis, on the other hand, exhibits generally larger uncertainties for all parton densities shown in Fig. 9.

The same findings hold for higher values of $Q^2 = 10$ and 100 GeV^2, as shown in Figs. 10 and 11. As can be seen, the gluon PDF from all other groups is slightly suppressed in the smaller x region ($x < 0.01$) compared to our Fit D. Because of the lack of data directly constraining the gluon PDF, our uncertainty is larger than those of NNPDF4.0 and MSHT20. Once again, CT18 generally exhibits larger uncertainties for all parton densities across the entire x region, as shown in Figs. 10 and 11.

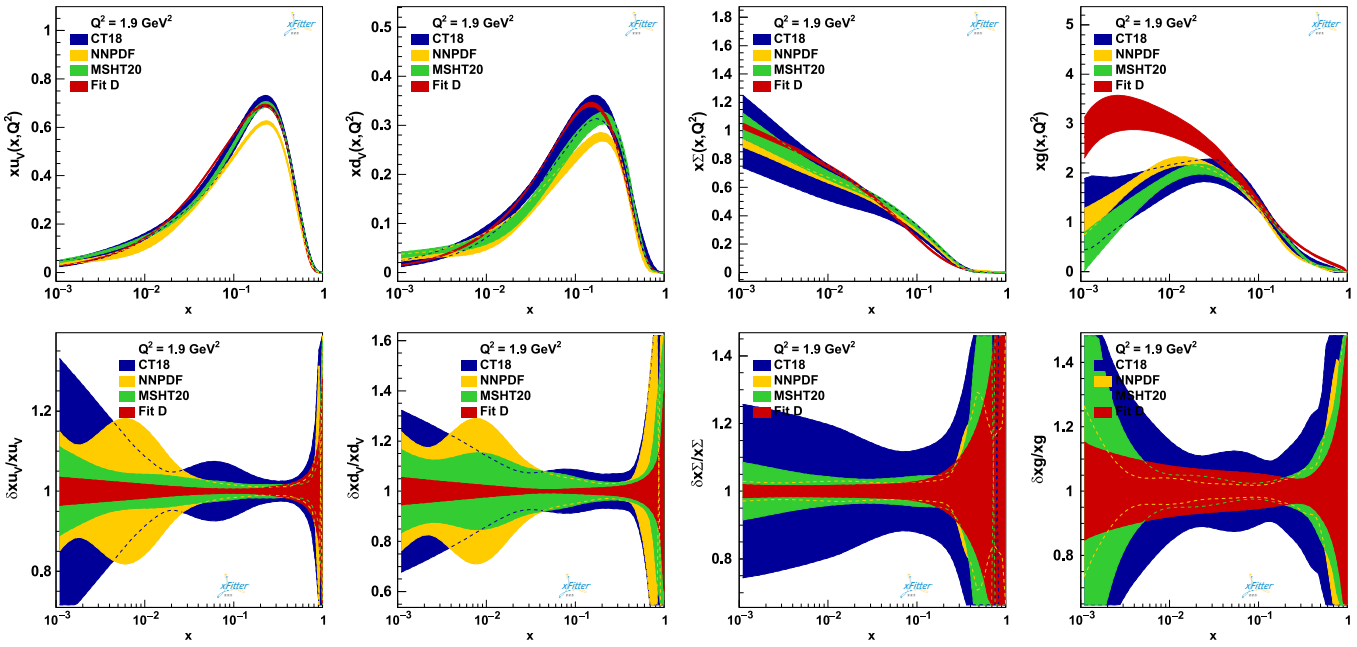

FIG. 9. Comparison of our nominal fit (Fit D) at NNLO with other very recent determinations of PDFs, including NNPDF4.0 [14], CT18 [12], and MSHT20 [16] for xu_v, xd_v, $x\Sigma$, and xg distribution as a function of x and at $Q^2 = 1.9$ GeV^2. The relative uncertainties $\delta xq(x, Q^2)/xq(x, Q^2)$ are also shown.

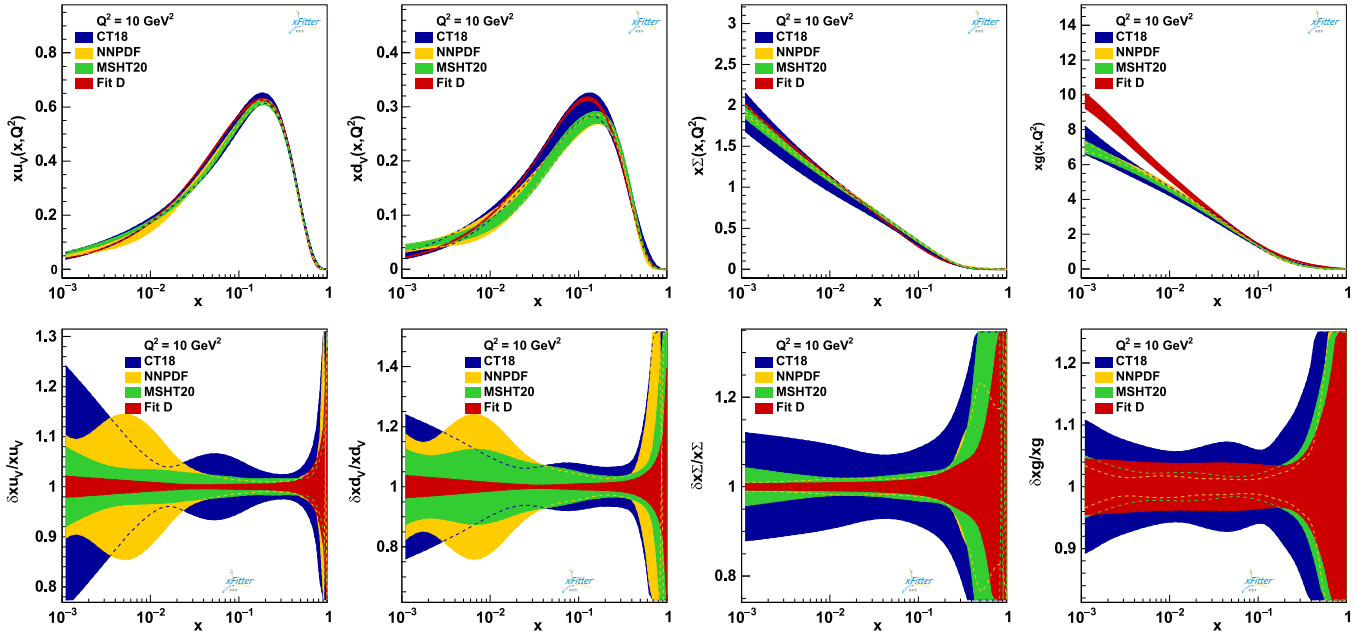

FIG. 10. Same as Fig. 9 but this time for higher Q^2 value of 10 GeV^2.

E. Comparison to experimental data

Here, we illustrate the ability of our NNLO QCD fit to describe the individual experiments included in our QCD analysis, with particular attention paid to the Drell-Yan data and W/Z boson production. We organize our data/theory discussions according to the specific physical processes.

In Figs. 12–14, we present a comparison between a selection of data included in our QCD fits and the corresponding NNLO best-fit results. This comparison aims to visually assess the fit quality and the relative size of data and PDF uncertainties. The data shown are representative of the global dataset, starting with the H1/ZEUS combined datasets, as shown in Fig. 12. Specifically, we show the results for both NC (top panel) and CC (bottom panel) HERA combined DIS data, where the bulk of the sensitivity in our fit still arises from HERA data. The data error bars shown in this figure correspond to the sum in quadrature of all uncertainties. Quantitative assessments of

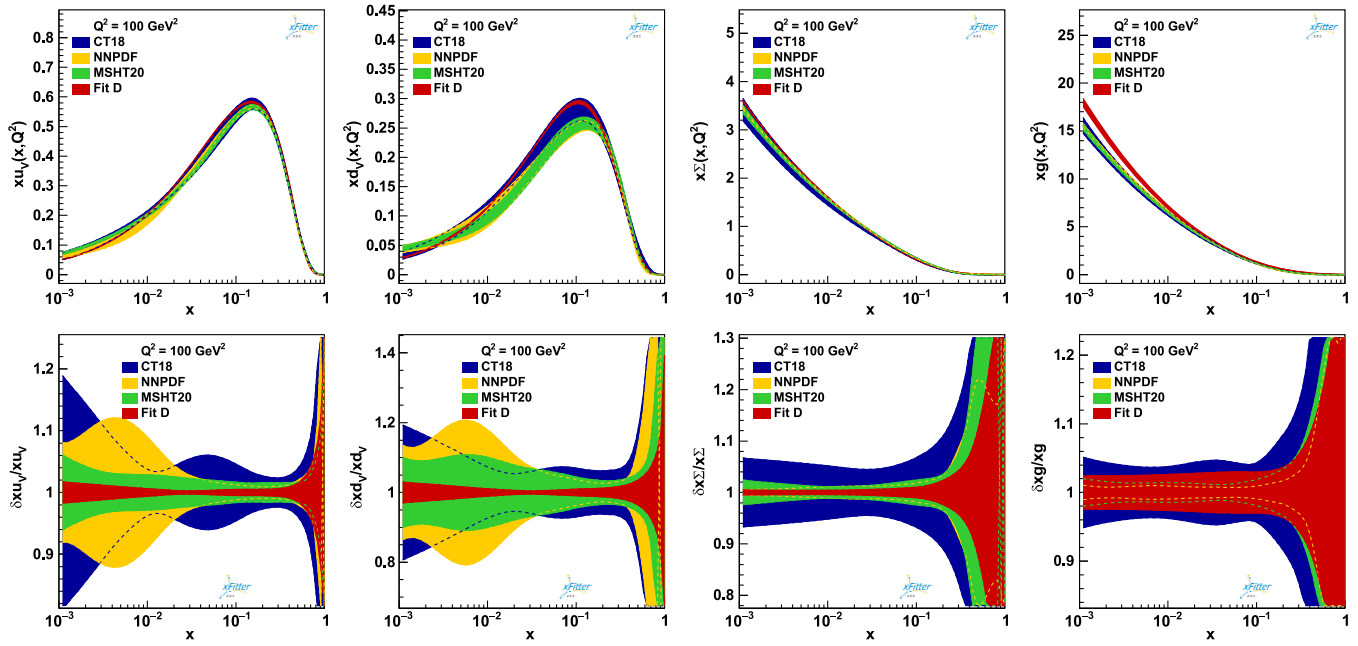

FIG. 11. Same as Fig. 9 but this time for higher Q^2 value of 100 GeV^2.

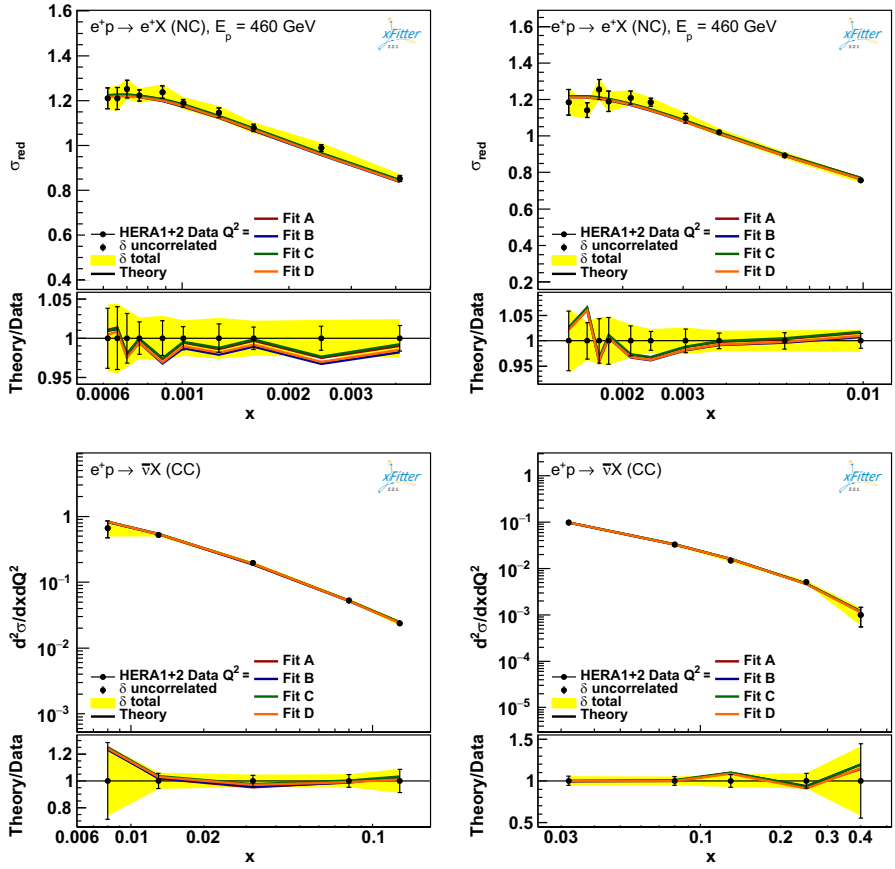

FIG. 12. Comparison between experimental data points and the theory prediction based on our NNLO best-fit results for a selection of fitted data points from H1/ZEUS combined DIS datasets (see text). The experimental uncertainty shown in the plots is the sum in quadrature of all statistical and systematic uncertainties.

these comparisons are provided by the χ^2 values presented in Table II.

Since the HERA combined DIS datasets serve as our base dataset selection, we present the comparison of our NNLO theory predictions for all four sets of PDFs discussed in Sec. V. The main finding from Fig. 12 is the very good agreement of our NNLO theory predictions with the HERA data for the kinematics shown.

Next, we discuss the comparisons between our NNLO theory predictions and selected Drell-Yan datasets. In Fig. 13, we show comparisons for the E866 Drell-Yan data and the ATLAS low-mass Drell-Yan data. As previously discussed, Drell-Yan cross section measurements have a significant impact on our PDFs compared to other datasets. The inclusion of these measurements allows for a clearer separation of the sea quark distributions, resulting in a notable reduction in the error bands as well.

Finally, in Fig. 14, we present a detailed comparison of our NNLO theory predictions with selected datasets from W/Z boson production measurements by CMS, ATLAS, D0, and CDF. Overall, the fit quality is satisfactory, and our nominal Fit D provides a generally good description of the data. However, some disagreements are observed in specific kinematic regions of certain datasets. As indicated by the χ^2 values in Table II, there are evident tensions between specific datasets. These tensions are particularly notable for the CDF W asymmetry data [63], the ATLAS low-mass Z data [53], and the HERA1 + 2 NC e^+p 920 dataset [52]. Such inconsistencies suggest that these datasets impose conflicting constraints on the PDFs, potentially affecting the overall quality of the fit.

VI. IMPACT OF SIMULATED EIC DIS DATA ON PROTON PDF DETERMINATION

In this section, we explore the impact of simulated inclusive DIS data from the future EIC on the determination of proton PDFs at NLO and NNLO accuracy in QCD. Comparisons are made relative to our nominal QCD fit, Fit D, as discussed in the previous sections. We also estimate the expected experimental uncertainty in the strong coupling constant, $\alpha_s(M_Z^2)$, when simulated EIC inclusive data are incorporated into analyses similar to those performed using HERA data.

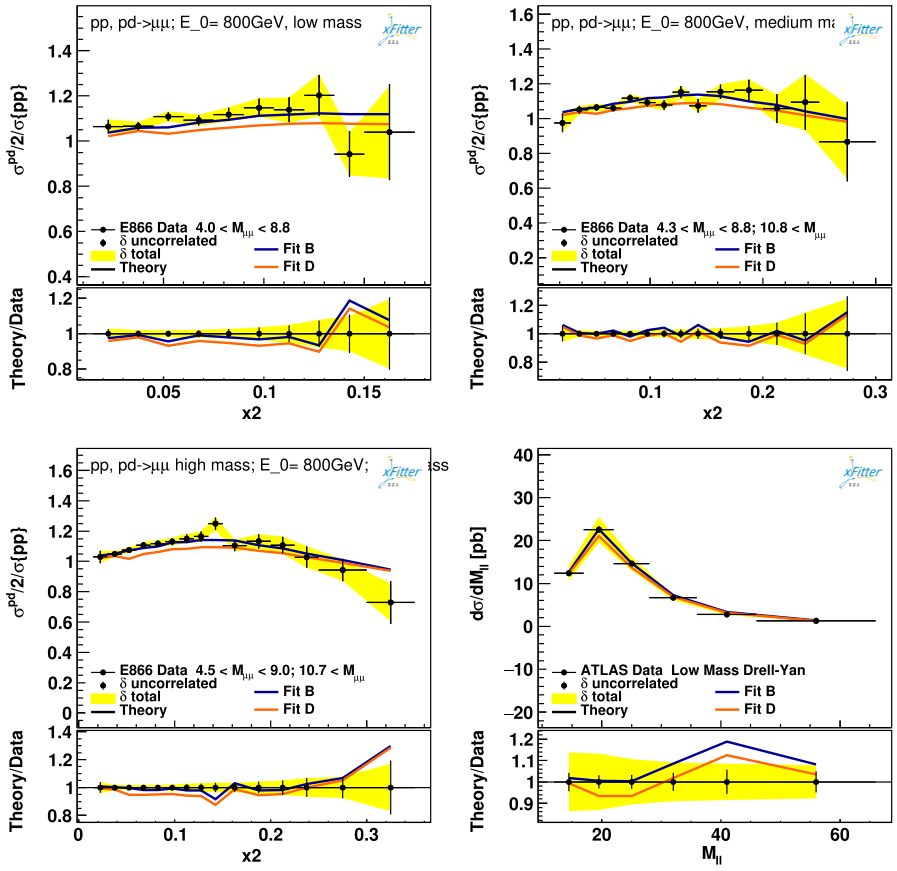

FIG. 13. Comparison between experimental data points and the theory prediction based on our NNLO best-fit results for a selection of fitted data points for Drell-Yan datasets (see text).

The EIC, currently being developed at Brookhaven National Laboratory in collaboration with the Thomas Jefferson National Accelerator facility, is expected to start collecting data around 2030 [6,81,82]. The EIC will collide highly polarized electrons with highly polarized protons, as well as light or heavy nuclei. In the electron-proton mode, the anticipated luminosity will range from 10^{33}–10^{34} cm^{-2} s^{-1}, with a center-of-mass energy $\sqrt{s}$ spanning from 29 to 141 GeV. The broad physics program of the EIC includes high-precision measurements of inclusive DIS cross sections, with a particular focus on the large Bjorken-x kinematic region, i.e., $x \gtrsim 0.7$, complementing the measurements from H1 and ZEUS at HERA [25]. Earlier investigations of the impact of inclusive EIC data on α_s precision and proton PDFs can be found in Refs. [25,26].

Table V shows the different beam energy configurations and their corresponding center-of-mass energies. The EIC will operate with different beam configurations, involving both electron and proton beams at a range of energies. The main datasets we include correspond to NC DIS pseudo-data generated for five different beam energy combinations, with center-of-mass energies ranging from 29 to 141 GeV. The most important electron-proton beam configurations included in this study are 10×275, 18×275, 5×41, 5×100, and 10×100 GeV [6,81,82]. Each dataset represents an integrated luminosity anticipated for one year of data collection at the EIC, with simulated measurements performed across a range of x and Q^2 values, logarithmically spaced across 6 orders of magnitude. The kinematic coverage of the EIC is expected to complement that of HERA, particularly by filling the high-x gap that HERA could not reach. Although there will be overlap in the kinematic regions covered, the EIC will provide much higher precision, offering new opportunities to probe the proton structure in this poorly constrained region. The key advantage of the EIC lies in its ability to provide precise data at large x and relatively low Q^2, which remained relatively underconstrained by previous HERA DIS experiments [25].

To improve our understanding of the impact of simulated EIC data on proton PDF determination, we have used NC and CC DIS cross sections at various center-of-mass energies, as anticipated for the EIC. These simulated datasets were generated based on predictions using HERAPDF2.0NLO and HERAPDF2.0NNLO as the baselines for NLO and NNLO accuracy, respectively [25,26]. For other inputs, such as heavy quark masses and the value of the

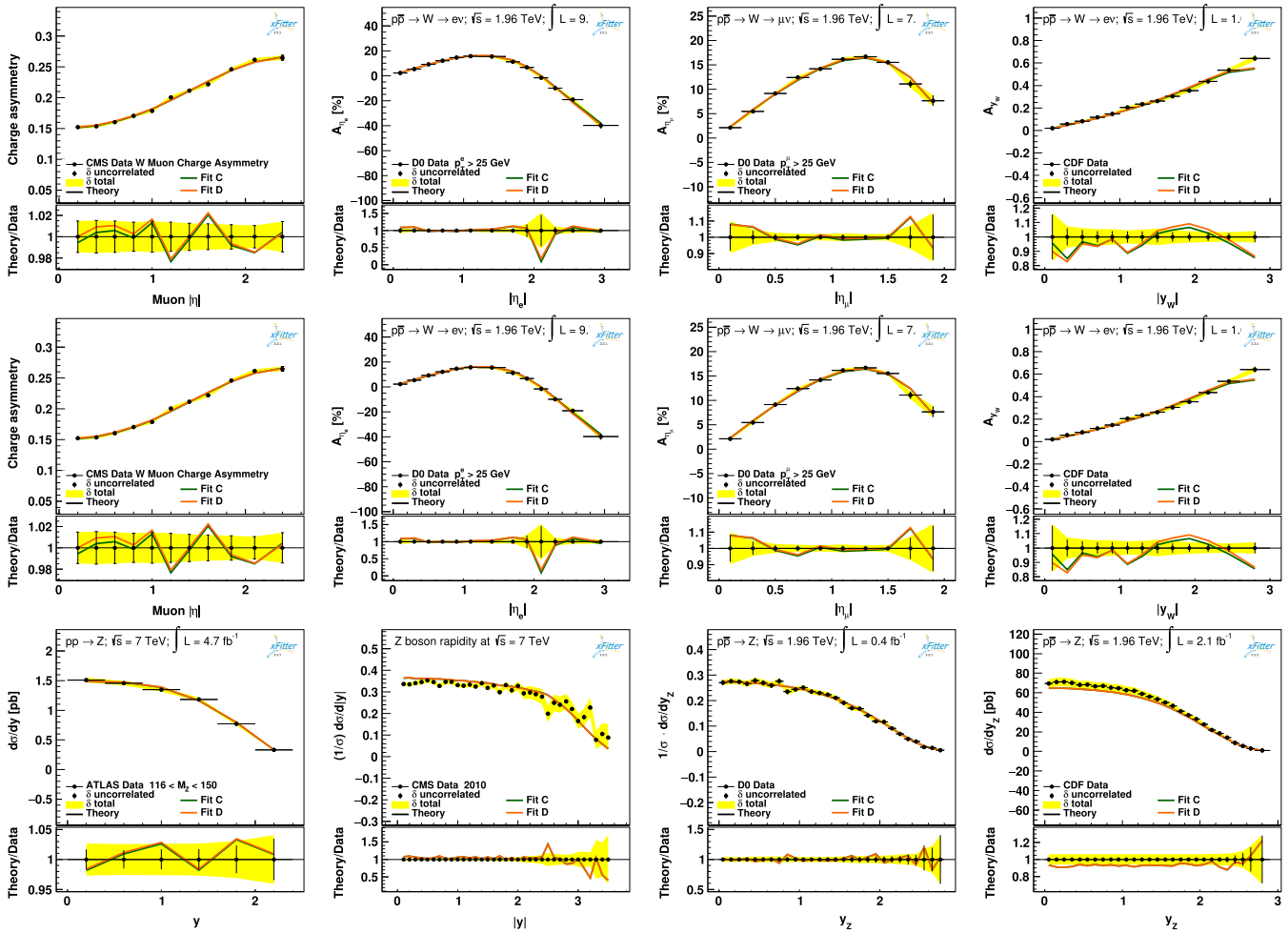

FIG. 14. Comparison between experimental data points and the theory prediction based on our NNLO best-fit results for a selection of fitted data points for W/Z boson production (see text).

strong coupling constant, we used the nominal values from HERAPDF2.0.

The pseudodata include statistical uncertainties and systematic errors modeled based on experience from HERA, conservatively adjusted to reflect the potential precision of the EIC. Specifically, an uncorrelated systematic uncertainty of 1.9% is applied across most data points, extending to 2.75% for the lowest inelasticity values (i.e., $y < 0.01$), along with a normalization uncertainty of 3.4%, fully correlated for each energy configuration. These uncertainties encompass both statistical and systematic components, designed to capture the challenges expected in high-precision EIC measurements.

TABLE V. EIC beam energy configurations, center-of-mass energies, and integrated luminosity.

Electron beam energy (GeV)	Proton beam energy (GeV)	Center-of-mass energy $\sqrt{s}$ (GeV)	Integrated luminosity (fb^{-1})
18	275	141	15.4
10	275	105	100
10	100	63	79
5	100	45	61
5	41	29	4.4

The inclusive DIS EIC pseudodata employed in our QCD analysis are generated following the methodology described in Ref. [25], which uses binning schemes based on those in the ATHENA detector proposal [83]. This approach allows us to simulate realistic EIC datasets capable of significantly constraining the proton PDFs, particularly in the region of x where existing data from HERA and other sources provide limited coverage. For the purposes of our QCD fits, the point-to-point systematic uncertainties were combined in quadrature with the statistical uncertainties, while the normalization uncertainties were treated as nuisance parameters, following the approach used in Ref. [76].

Figure 15 shows the locations of the HERA DIS datasets and EIC simulated NC and CC inclusive DIS data points in the (x, Q^2) kinematic plane used in this QCD analysis.

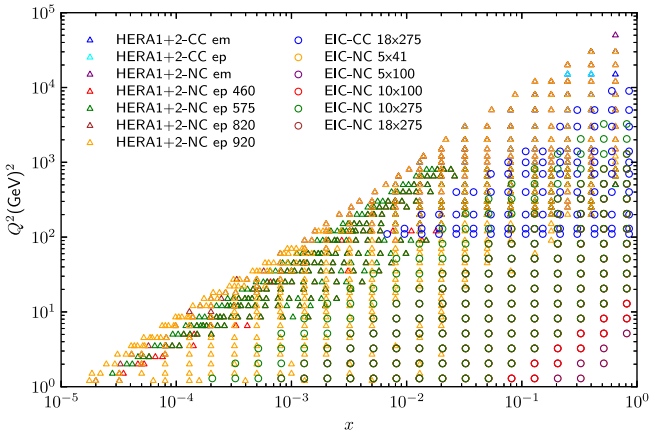

FIG. 15. The kinematic coverage in the (x, Q^2) plane for the HERA DIS data and the EIC simulated NC and CC inclusive DIS pseudodata used in this analysis. The same kinematic cuts applied to the HERA DIS datasets have also been used for the EIC pseudodata.

As illustrated in the figure, the EIC pseudodata overlap with the HERA DIS data in some regions while also extending the kinematic reach to higher values of x, in the intermediate Q^2 range. It includes, for example, $x > 0.1$ for $Q^2 < 10^2$ GeV^2. As can be seen, the EIC data provide coverage at very low x values as well, extending to $x < 10^{-4}$. Additionally, the same cuts imposed on the HERA datasets were also applied to the EIC pseudodata, ensuring consistency in the treatment of kinematic constraints and thereby avoiding the need for additional higher-twist corrections in our analysis.

The inclusion of the simulated EIC pseudodata in our QCD fits provides substantial improvements in constraining the gluon and sea quark distributions, especially at medium to large values of x, where the precision added by the EIC data is unprecedented. As shown in Figs. 16 and 17, incorporating these datasets in our QCD fit (Fit D + EIC) reduces the uncertainties of gluon PDFs at $x \leq 0.2$ for moderate-to-high Q^2. Furthermore, the impact of EIC on the determination of the strong coupling constant $\alpha_s(M_Z^2)$ is also significant, which we explore in detail in the next section.

In Table VI, we present the extracted χ^2 values from our QCD fits, comparing the baseline analysis (Fit D) with the extended analysis that includes the simulated EIC data (Fit D + EIC). These χ^2 values, corresponding to different experimental datasets, provide a quantitative measure of the fit quality and illustrate the impact of incorporating EIC pseudodata in the QCD analysis.

The addition of the EIC data results in a significant improvement in the overall fit quality, with the total χ^2 per degree of freedom decreasing from 1.25 in Fit D to 1.20 in Fit D + EIC. The fit quality for HERA I + II slightly improves with the addition of EIC data, as indicated by the reduction in χ^2 from 1151/1016 in Fit D to 1130/1016 in Fit D + EIC. This suggests that the EIC data are consistent with the HERA measurements and help refine the global fit. However, the χ^2 values for the W/Z production data from CMS, ATLAS, D0, and CDF exhibit mixed results, with minimal changes in the χ^2 for the ATLAS and E866 Drell-Yan datasets when EIC data are added.

The simulated EIC datasets show good agreement with theoretical predictions, with reasonable χ^2 values across all beam energy configurations. The NC and CC EIC datasets exhibit χ^2/d.o.f. values close to 1, indicating a good fit within the global QCD analysis. These datasets considerably improve the constraints on PDFs, particularly in regions of medium to large x and moderate Q^2, where previous data provided limited constraints.

Our findings clearly demonstrate that the inclusion of EIC pseudodata in the global QCD fit significantly enhances the fit quality, as evidenced by the improved χ^2/d.o.f. The EIC data introduce crucial new constraints, resulting in a more precise determination of the proton PDFs. Despite minor tensions with certain datasets, the overall improvement underscores the importance of the EIC for future QCD analyses [25,26].

VII. IMPACT OF JET AND DIJET PRODUCTION DATA ON PROTON PDF DETERMINATION

This section discusses the significant role that jet and dijet production data play in determining PDFs and reducing their associated uncertainties, particularly for the gluon distribution. We also explore the effect of these data on the precision of the determination of the strong coupling constant, α_s. Historically, inclusive jet and dijet production measurements have been crucial for constraining the gluon density, $g(x, Q^2)$, due to the high sensitivity of these processes to the gluon PDF, especially at high-energy colliders. Early jet data from HERA and Tevatron Run-II have had a noticeable impact on global PDF fits, as demonstrated in numerous QCD analyses.

The HERA jet data used in our work include measurements of inclusive jet production in DIS at high Q^2 values, ranging from 150 to 15,000 GeV^2, recorded by the H1 Collaboration [84]. Additionally, we include data from H1 on jet production in the lower Q^2 range, below 100 GeV^2, based on an integrated luminosity of 43.5 pb^{-1} [85]. The ZEUS Collaboration inclusive jet differential cross section measurements in DIS for $Q^2 > 125$ GeV^2 with 38.6 pb^{-1} [86] and an expanded dataset with 82 pb^{-1} [87] are also included. Furthermore, inclusive dijet cross sections in DIS measured by ZEUS with an integrated luminosity of 374 pb^{-1} [88] are considered as well.

In Fig. 18, we compare our nominal fit (Fit D) at NNLO accuracy with the extended fit (Fit D + jet/dijet),

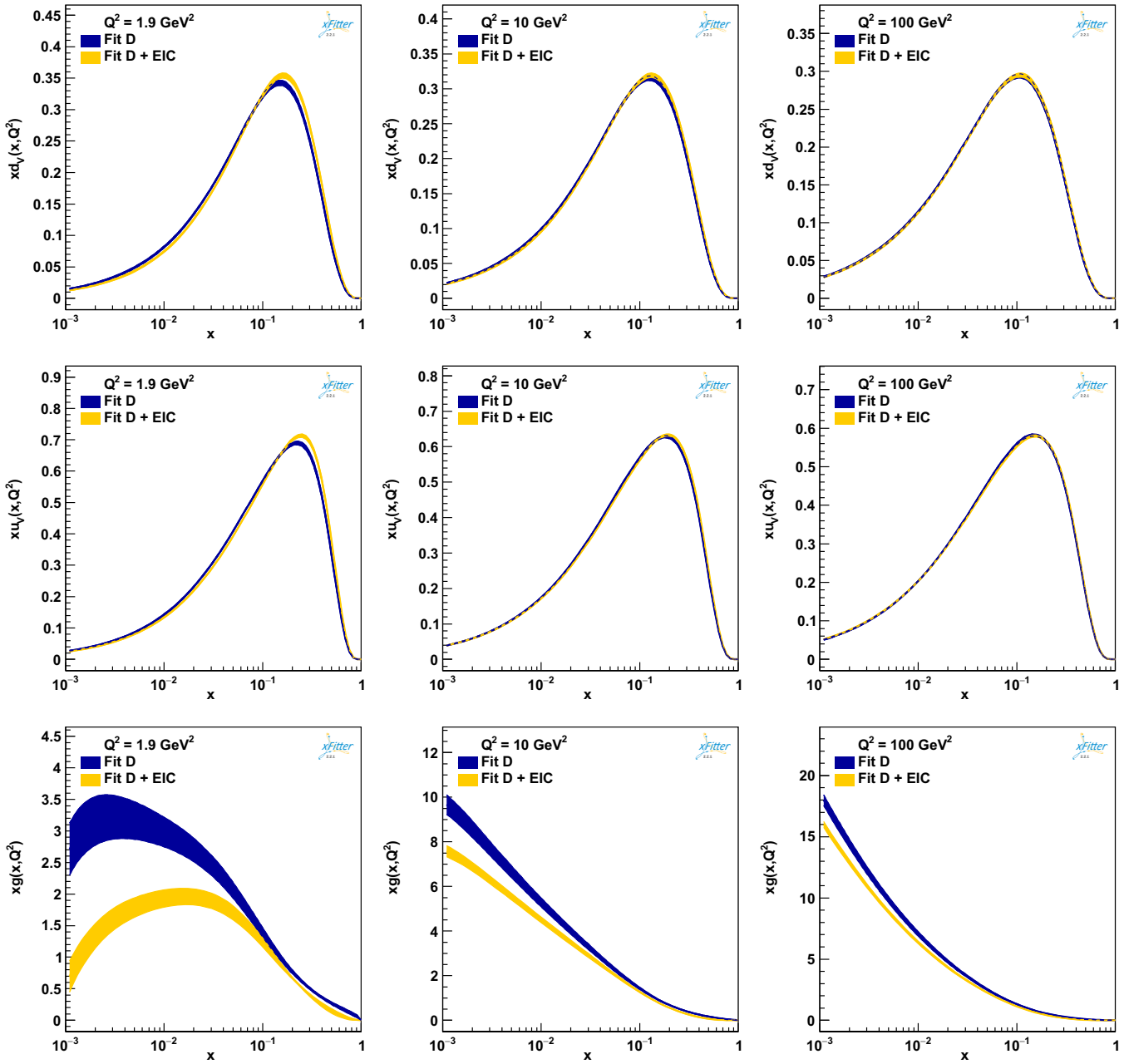

FIG. 16. Impact of EIC pseudodata on xu_v, xd_v, and xg distribution functions at $Q^2 = 1.9$, 10, 100 GeV^2.

which includes the aforementioned HERA jet and dijet production data. The comparisons are presented for the relative uncertainties of the gluon PDFs, $\delta xg(x, Q^2)/xg(x, Q^2)$, at $Q^2 = M_W^2$ and M_Z^2. As can clearly be seen, the inclusion of the gluon-sensitive jet and dijet data significantly reduces the error bands, especially at moderate to very small values of x. This reduction indicates the crucial role of jet and dijet production data in constraining the gluon PDF, which is inherently difficult to determine with precision in the absence of such data. The enhanced constraints provided by the HERA jet and dijet datasets are particularly impactful in reducing uncertainties in the low x region, where the gluon contribution is less well determined. In terms of the overall quality of the fit, as summarized in Table VII, the χ^2/d.o.f. value improves from 1.25 in the nominal fit (Fit D) to 1.22 when the jet and dijet data are included (Fit D + jet/dijet). The reduction in χ^2/d.o.f. demonstrates the importance of these datasets in achieving more precise proton PDF determinations.

In Figs. 19 and 20, we compare the results of our nominal fit (Fit D) at NNLO accuracy with fits that include

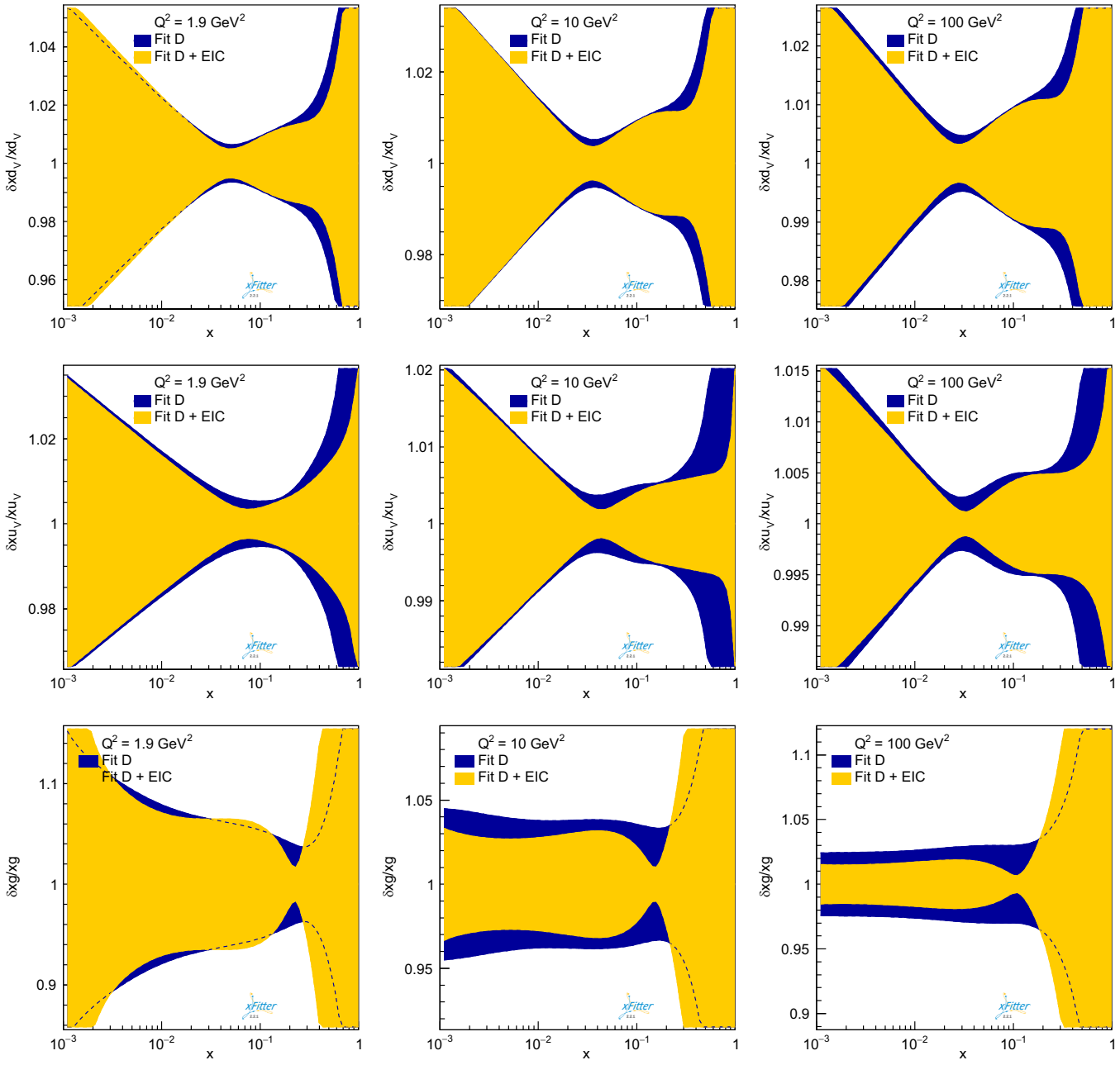

FIG. 17. Same as Fig. 16 but this time for the relative uncertainties $\delta xq(x, Q^2)/xq(x, Q^2)$.

TABLE VI. The values of extracted χ^2 from the QCD fits with and without EIC data.

Experiment	Fit D	Fit D + EIC
HERA I + II	1151/1016	1130/1016
CMS W/Z	106.8/90	101.8/90
ATLAS W/Z	104.8/91	111.3/91
ATLAS Drell-Yan	25.3/27	24.3/27
D0W/Z	67/51	77/51
CDF W/Z	75/41	69/41
E866 Drell-Yan	50/39	70/39
EIC NCep 10x275	⋯	117/103
EIC NCep 18x275	⋯	113/116
EIC NCep 5x41	⋯	31/36
EIC CCep 18x275	⋯	112/88
EIC NCep 5x100	⋯	33/55
EIC NCep 10x100	⋯	64/66
Correlated χ^2	113	149
Log penalty χ^2	−13.22	−40.97
Total χ^2/d.o.f.	1678/1339 = 1.25	2160/1803 = 1.20

(Table continued)

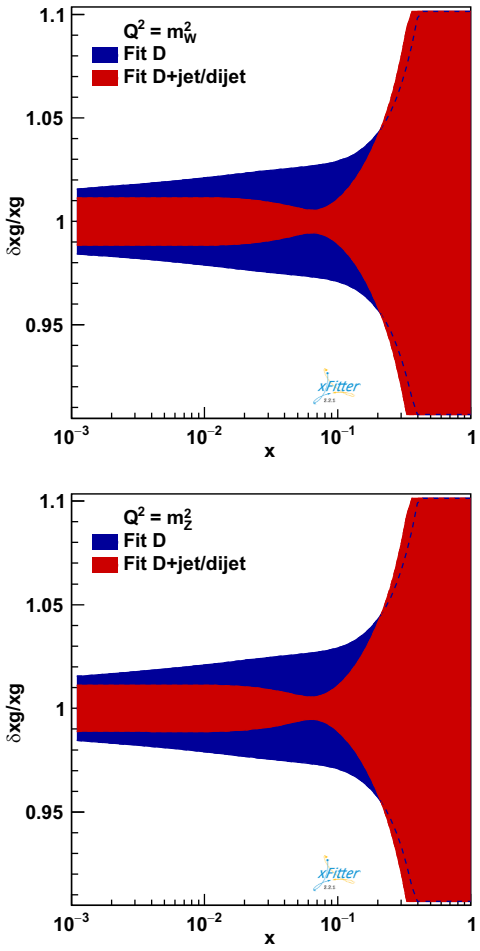

FIG. 18. Comparison of our nominal fit (Fit D) at NNLO accuracy with the extended fit (Fit D + jet/dijet) that includes the HERA jet and dijet production data. The comparisons are shown for the relative uncertainties of the gluon PDFs, $\delta xg(x, Q^2)/xg(x, Q^2)$.

simulated EIC data (Fit D + EIC) and those that incorporate both EIC and HERA jet/dijet data (Fit D + EIC +jet/dijet). The comparisons are presented for the distributions xu_v, xd_v, and xg as functions of x, at $Q^2 = 1.9$, 10, and 100 GeV^2.

As observed in Figs. 19 and 20, the inclusion of HERA jet and dijet data significantly constrains the gluon distribution, resulting in reduced uncertainties, particularly at medium and very small values of x. This is because jet production cross sections in DIS are directly sensitive to the gluon density.

Moreover, jet production data are instrumental in determining the strong coupling constant, $\alpha_s(M_Z^2)$. Because of the sensitivity of jet production to higher-order QCD corrections, incorporating jet and dijet datasets allows for a more precise extraction of α_s, thereby enhancing the overall fit quality of global PDF analyses.

In Table VII, we present the χ^2 values for the QCD fit both with and without the inclusion of jet and dijet production data. The χ^2 for the HERA I + II dataset remains stable, with a slight improvement from 1151/1016 to 1130/1016. A significant improvement is observed in the χ^2 values for both CMS and ATLAS W-/Z-production datasets. The χ^2 values for the ATLAS and E866 Drell-Yan datasets show minor changes, with the ATLAS χ^2 improving slightly.

Finally, the addition of jet and dijet production data from HERA provides critical information about the gluon distribution. The χ^2 values for these datasets reflect the strong constraints that these data provide, particularly in the high Q^2 region, where gluon-initiated processes dominate. Including these data leads to a significant reduction in the overall uncertainty of the gluon PDF, which in turn improves the precision of predictions for processes sensitive to gluons. The improved χ^2/d.o.f.

TABLE VII. The χ^2 values from QCD fits comparing results with and without the inclusion of jet and dijet production data.

Experiment	Fit D	Fit D + jet/dijet	Fit D + EIC	Fit D + EIC + jet/dijet
HERA I + II	1151/1016	1128/1016	1130/1016	1127/1016
CMS W/Z	106.8/90	103/91	101.8/90	99.9/90
ATLAS W/Z	104.8/91	112/91	111.3/91	111/91
ATLAS Drell-Yan	25.3/27	24/27	24.3/27	24.4/27
D0 W/Z	67/51	77/51	77/51	68/51
CDF W/Z	75/41	66/41	69/41	78/41
E866 Drell-Yan	50/39	72/39	70/39	73/39
H1 jet	⋯	30/52	⋯	31/52
ZEUS jet	⋯	54/60	⋯	54/60
ZEUS dijet	⋯	18/22	⋯	17/22
EIC NCep 5x41	⋯	⋯	31/36	31/36
EIC NCep 10x275	⋯	⋯	117/103	117/103
EIC NCep 18x275	⋯	⋯	113/116	113/116
EIC CCep 18x275	⋯	⋯	112/88	112/88
EIC NCep 5x100	⋯	⋯	33/55	32/55

(Table continued)

TABLE VII. *(Continued)*

Experiment	Fit D	Fit D + jet/dijet	Fit D + EIC	Fit D + EIC + jet/dijet
EIC NCep 10x100	⋯	⋯	64/66	64/66
Correlated χ^2	113	142	149	156
Log penalty χ^2	−13.22	−27	−40.97	−48.72
Total χ^2/d.o.f.	1678/1339 = 1.25	1799/1473 = 1.22	2160/1803 = 1.20	2259/1937 = 1.17

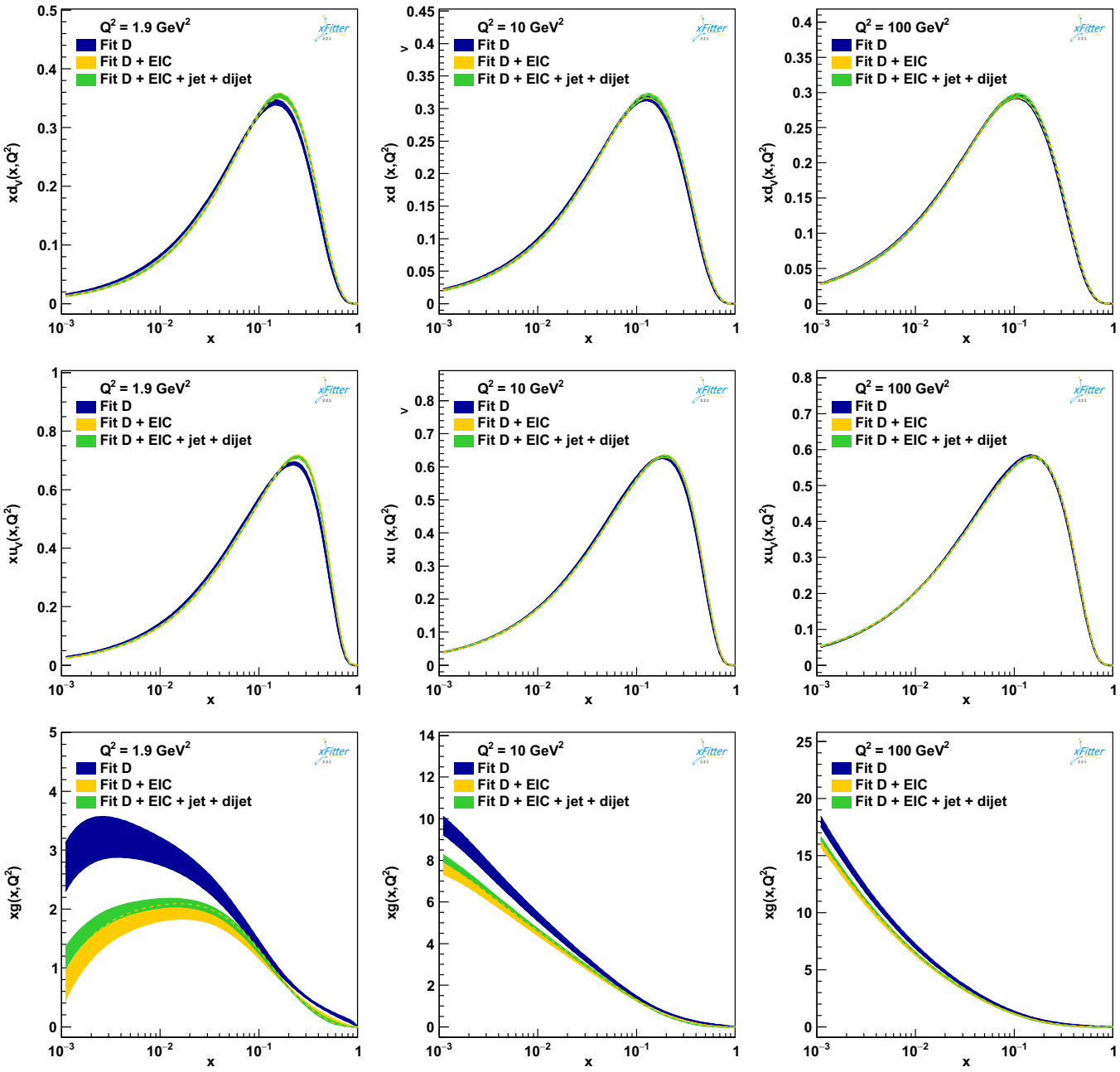

FIG. 19. Comparison of our nominal fit (Fit D) at NNLO accuracy with (Fit D + EIC) in which include the simulated EIC data and (Fit D + EIC + jet/dijet) which include the HERA jet data. The comparisons are shown for xu_v, xd_v, and xg distribution as a function of x and at $Q^2 = 1.9$, 10, and 100 GeV^2.

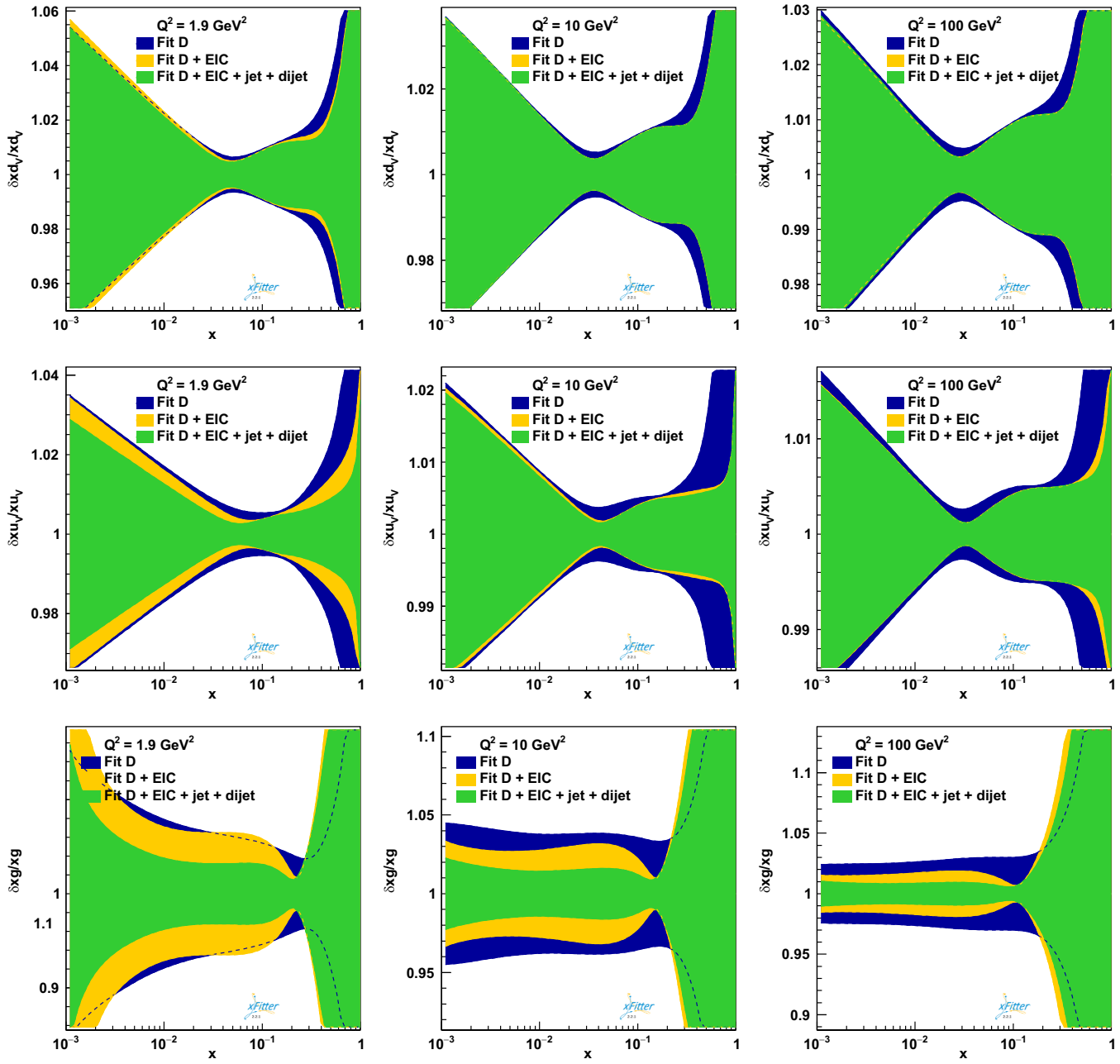

FIG. 20. Same as Fig, 19, but this time the comparisons are shown for the relative uncertainties $\delta xf(x, Q^2)/xf(x, Q^2)$.

reflects enhanced consistency between theoretical predictions and experimental measurements, further supporting the inclusion of jet data from the LHC in global QCD fits to improve the precision of proton PDFs and α_s determinations.

Further improvements in PDF uncertainties, particularly for the gluon distribution, are anticipated when inclusive jet and dijet data from the EIC are added to the QCD analysis. These data are also expected to significantly contribute to the precision of α_s. Future studies could employ theory grids for various EIC energies using the fastNLO framework [89], although such an analysis is beyond the scope of this work. Additionally, incorporating a broader range of hadron collider jet and dijet datasets could further constrain the gluon PDFs and improve the precision of α_s.

VIII. STRANGE-QUARK DENSITY

The strange-quark distribution in the proton plays a crucial role in understanding the quark structure of the proton. One of the key quantities characterizing the strange-quark fraction is the ratio r_s, which is defined as the ratio of the strange-quark distribution to the down sea-quark distribution:

TABLE VIII. The numerical values of $\alpha_s(M_Z)$ and r_s extracted from our Fit D, Fit D + jet/dijet, Fit D + EIC, and Fit D + EIC + jet/dijet QCD analyses.

Parameter	Fit D	Fit D + jet/dijet	Fit D + EIC	Fit D + EIC + jet/dijet
$\alpha_s(M_Z)$	0.1128 ± 0.0014	0.1192 ± 0.0012	0.1188 ± 0.0008	0.1183 ± 0.0003
r_s	1.069 ± 0.053	1.148 ± 0.053	1.164 ± 0.051	1.138 ± 0.048

$$r_s = \frac{s + \bar{s}}{2\bar{d}}. \tag{14}$$

This ratio offers insight into the balance between strange and down quarks in the proton sea and is important for various QCD and electroweak processes, particularly in high-energy collider experiments.

In this analysis, we determine r_s at $Q^2 = 1.9$ GeV^2 and $x = 0.023$, a value selected because it lies within the region of maximum sensitivity in the ATLAS data for central rapidity at $\sqrt{s} = 7$ TeV [53].

We explore the strange-quark fraction further by examining the values of r_s obtained under different fit scenarios. These are summarized in Table VIII, which shows the extracted values of r_s from various fits.

The baseline fit (Fit D) yields $r_s = 1.069 \pm 0.053$, indicating a nearly equal strange-to-down sea-quark density, consistent with previous findings of an unsuppressed strange-quark content at small x. For our Fit D + EIC when projected data from the EIC are included in the fit, r_s increases to 1.164 ± 0.051, suggesting that the EIC will significantly enhance the precision of strange-quark density. Finally, in Fit D + EIC + jet/dijet in which the jet and dijet production data included to the fit, r_s is slightly reduced to 1.138 ± 0.048, while the uncertainty is reduced even further.

The ATLAS data, combined with future measurements from the EIC and precision jet experiments, will continue to play a crucial role in reducing the uncertainties associated with the strange-quark distribution. The extracted values of r_s in this analysis suggest that strange quarks contribute significantly to the proton sea. These findings will have a broad impact on QCD phenomenology and future high-energy physics experiments, particularly in processes involving electroweak boson production and DIS.

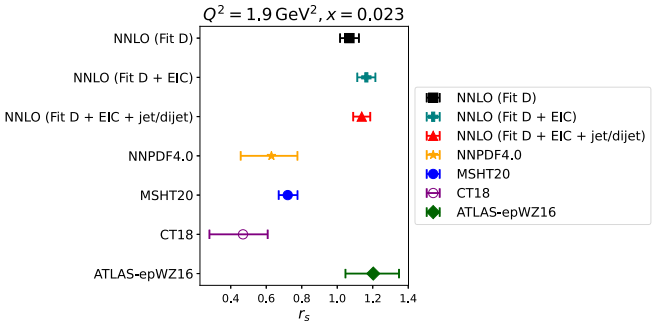

FIG. 21. Determination of the relative strange-to-down sea quark fraction r_s, calculated at the initial scale $Q_0^2 = 1.9$ GeV^2 and at $x = 0.023$, corresponding to the point of greatest sensitivity to the ATLAS W/Z data. The results are compared with other global QCD analyses, specifically NNPDF4.0 [14], MSHT20 [16], CT18 [12], and ATLAS-epWZ16 [53].

Our determinations of r_s for the three different QCD analyses mentioned above are illustrated in Fig. 21. The measurements are presented with both experimental and PDF-fit-related uncertainties. The results are also compared with recent global QCD analyses, specifically NNPDF4.0 [14], MSHT20 [16], CT18 [12], and the ATLAS-epWZ16 [53] analysis. As shown, the r_s predictions from NNPDF4.0, MSHT20, and CT18 global fits are significantly lower than unity, with values ranging between approximately 0.45 and 0.75. In contrast, our result from the Fit D + EIC analysis aligns more closely with the ATLAS-epWZ16 analysis, likely due to both analyses relying on the same ATLAS W/Z dataset. This consistency highlights the impact of incorporating high-precision data in constraining the strange-to-down sea quark fraction.

IX. STRONG COUPLING CONSTANT

The precise determination of the strong coupling constant, α_s, relies on various experimental datasets that are sensitive to different aspects of the proton PDFs. In this section, we examine how the EIC and jet/dijet production data contribute to constraining α_s, focusing on the complementary information these datasets provide about quark and gluon interactions. This sensitivity arises from the unique kinematic coverage and specific processes probed by the EIC and jet production measurements, making them highly valuable for reducing uncertainties in α_s and enhancing the precision of global QCD fits. The EIC will significantly enhance the determination of the strong coupling constant, $\alpha_s(M_Z^2)$, through its extensive kinematic reach, particularly in regions where current datasets, such as those from HERA, are less effective.

The strong coupling constant, α_s, governs the evolution of quark densities as the energy scale Q^2 increases. In the large-x regime, quark distributions are particularly sensitive to α_s because parton-parton interactions dominate in this region. By probing large x, the EIC directly measures the evolution of quark PDFs, thereby providing precise constraints on the behavior of α_s in this regime.

The sensitivity of EIC to α_s also arises from the specific processes it probes and the unique kinematic overlap it offers with previous data. In DIS processes, the interaction

between quarks and gluons is directly affected by α_s. The structure functions measured in EIC experiments, such as $F_2(x, Q^2)$, depend on quark-gluon interactions, which are controlled by α_s. The quantity $dF_2/d\ln Q^2$, which measures scaling violations, is particularly sensitive to α_s, especially at large x. In this region, the evolution is dominated by the $q \to qg$ splitting, which involves the product of α_s and the large-x quark densities. The high precision of EIC measurements leads to more accurate constraints on the quark-gluon coupling and, consequently, on α_s, with a reduced dependence on the gluon distribution compared to lower x values.

This decoupling reduces the correlation between α_s and the gluon distribution, resulting in a more precise extraction of $\alpha_s(M_Z^2)$. The inclusion of simulated EIC data in the fits provides additional coverage of the phase space, particularly in regions where HERA data alone were limited. This unique contribution from the EIC highlights its role in providing world-leading precision for the determination of α_s, supporting future QCD analyses and experimental research [25,26].

As illustrated in Table VIII, the inclusion of EIC pseudodata significantly reduces the uncertainty on $\alpha_s(M_Z^2)$ from 0.0014 (in the baseline Fit D) to 0.0008. This reduction is primarily due to the enhanced sensitivity to quark-gluon interactions and the broader kinematic coverage provided by the EIC. A summary of the $\alpha_s(M_Z)$ measurements at different collaborations, including the H1, ZEUS, ATLAS, and CMS. [76,90–103], as well as the world average [104], is also presented in Fig. 22. Our analyses performed at NNLO are indicated separately for comparison.

Jet and dijet production measurements offer complementary sensitivity to α_s, primarily through their ability to probe the gluon distribution in the proton. Jet production is particularly sensitive to gluon-gluon and quark-gluon interactions, both of which are controlled by α_s. The strong coupling constant directly enters the cross sections for jet production, making these measurements a powerful tool for constraining α_s. Consequently, the inclusion of jet and dijet data in the QCD fit provides direct constraints on the gluon distribution and α_s, particularly at high energy scales.

Jet and dijet production data, particularly from the LHC and HERA, cover a broad range of high Q^2 values. In this regime, gluon interactions become increasingly important, enhancing the sensitivity to α_s. The precise measurement of jet cross sections at large Q^2 significantly reduces the uncertainty in the determination of α_s, as evidenced by the improved results in Fit D + EIC + jet/dijet, where the uncertainty on $\alpha_s(M_Z^2)$ decreases to 0.0003.

While the EIC data provide essential constraints on quark distributions at lower energy scales, jet and dijet production measurements contribute by refining our understanding of gluon distributions at higher energy scales. Together, these datasets offer complementary sensitivity to

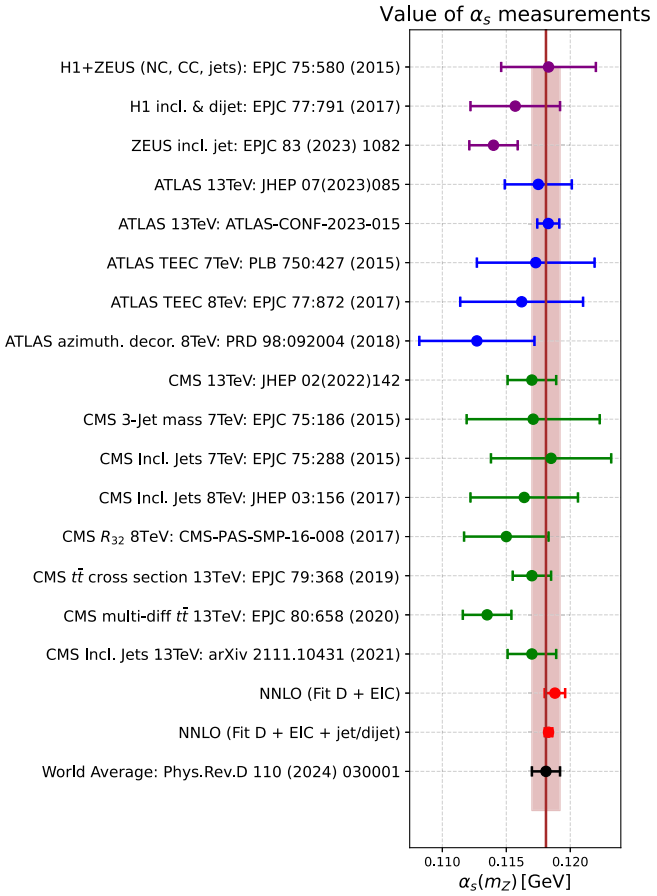

FIG. 22. Values of $\alpha_s(M_Z)$ resulting from different measurements, including H1, ZEUS, ATLAS, and CMS. [76,90–103]. Our analyses performed at NNLO are also indicated separately for comparisons.

α_s, ensuring that both quark and gluon interactions are accurately described in the global fit.

The combination of EIC and jet/dijet data leads to a significant improvement in the precision of $\alpha_s(M_Z^2)$. As seen in Table VIII and Fig. 22, the inclusion of both datasets in the global fit reduces the uncertainty on $\alpha_s(M_Z^2)$ from 0.0008 (Fit D + EIC) to 0.0003 (Fit D + EIC +jet/dijet).

These results suggest that further improvements in the precision of $\alpha_s(M_Z^2)$ could be achieved by incorporating inclusive jet and dijet data from the EIC into the QCD analysis. This could be facilitated by using theory grids for EIC energies within the fastNLO framework. Additionally, other observables measurable at the EIC could provide further opportunities for constraining the strong coupling constant.

Beyond a DIS-only approach, it would be valuable to explore the impact of EIC data on α_s determinations in global QCD fits that integrate data from LHC measurements not included in this work, as well as other experimental sources. Before the EIC becomes operational, progress in understanding higher-order uncertainties will

be crucial. Reaching a consensus on addressing these uncertainties in $\alpha_s(M_Z^2)$ determinations based on EIC data will be essential for maximizing the precision of future measurements.

X. SENSITIVITY OF VARIOUS DATASETS TO THE STRONG COUPLING CONSTANT

In every proton PDF analysis, the sensitivity of each dataset whether individually or collectively to the variation of the strong coupling constant $\alpha_s(M_Z^2)$ depends on the chosen settings, such as the parametrization and the values of applied cuts. One of the most effective approaches for assessing this sensitivity is the χ^2-scan method. This method allows us to evaluate how different experimental datasets influence the fit for $\alpha_s(M_Z^2)$. The following equation is used for the χ^2-scan:

$$\Delta\chi_e^2(\pi) = \chi_e^2(\pi) - \chi_e^2(\pi_{\min}). \tag{15}$$

In this equation, $\Delta\chi_e^2(\pi)$ is a function of the fit parameters π, which, in this case, is the variation of the QCD coupling constant ($\pi = \alpha_s(M_Z^2)$). The term $\chi_e^2(\pi_{\min})$ represents the minimum value obtained from the final fit, specifically from Fit D + EIC +jet/dijet. This value corresponds to the optimal parameter set, denoted as $\pi_{\min}$, which yields the lowest χ^2. In contrast, $\chi_e^2(\pi)$ refers to the χ^2 value obtained for each variation of $\alpha_s(M_Z^2)$ for a specific experimental dataset (e) or for all datasets together.

In the χ^2-scan procedure, the value of the parameter under study, $\alpha_s(M_Z^2)$, is fixed at different specific values, while all other fit parameters are allowed to vary freely, as described in Sec. IV. The difference between the values obtained from these two types of calculations [$\chi_e^2(\pi)$ and $\chi_e^2(\pi_{\min})$] provides the $\Delta\chi_e^2(\pi)$ value. This value effectively captures the change in the goodness of fit for different fixed values of $\alpha_s(M_Z^2)$.

Figure 23 illustrates the dependence of $\Delta\chi^2$ for different experimental datasets and for the full set of data on variations of $\alpha_s(M_Z^2)$, ranging from 0.116 to 0.120. From the comparison of the different curves in the figure, it is evident that $\Delta\chi_{EIC}^2(\alpha_s(M_Z^2))$ shows a strong sensitivity to the variation of $\alpha_s(M_Z^2)$, especially when $\alpha_s(M_Z^2)$ is less than 0.11834. On the other hand, other datasets, such as HERA I + II, W/Z, Drell-Yan, and jet/dijet production data, exhibit lower sensitivity to the variation of the strong coupling constant. This suggests that EIC data could play a crucial role in constraining $\alpha_s(M_Z^2)$ with higher precision compared to the other datasets.

The χ^2 scan method used here is similar to the approach employed in the CT18 analysis, which also investigated the sensitivity of different datasets, including DIS and jet production data, to $\alpha_s(M_Z^2)$. Similar to the CT18 results, the $\Delta\chi^2$ curves for individual datasets generally exhibit a parabolic shape, consistent with the central limit theorem,

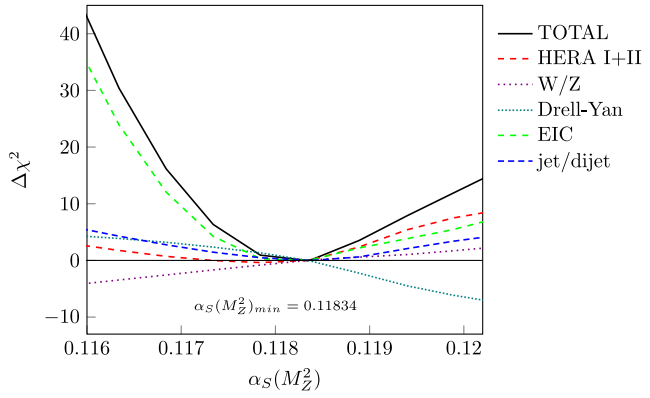

FIG. 23. The curves of $\Delta\chi^2$ as a function of the strong coupling constant $\alpha_s(M_Z^2)$ obtained from Fit D + EIC + jet/dijet for all types of datasets. The $\Delta\chi^2_{\text{TOTAL}}$ curve represents the sensitivity when all experimental data are considered simultaneously.

indicating their respective pulls on the preferred value of $\alpha_s(M_Z^2)$. This helps identify any tensions or consistencies across different datasets, providing insights into how well each dataset constrains α_s.

XI. CONCLUSIONS

In this work, we have presented new proton PDFs at both NLO and NNLO accuracy, derived from a comprehensive analysis of high-precision data from the LHC, combined HERA DIS dataset, and additional crucial inputs. We have explored the impact of including data from Drell-Yan pair production, as well as W and Z boson production, on the PDFs and their associated uncertainties. Our findings demonstrate that the inclusion of these datasets significantly enhances the precision and reliability of the proton PDFs, reducing uncertainties across a wide range of kinematic regions. Specifically, the Drell-Yan, W-, and Z-production data provided important sensitivity to quark distributions at moderate to high values of Bjorken x, thereby complementing the information obtained from DIS data. As has been shown, the inclusion of jet and dijet production data further constrained the gluon distribution, reducing the gluon uncertainties for $x \leq 0.2$. The resulting PDFs are accompanied by well-defined error estimates calculated using the Hessian method, ensuring a rigorous quantification of uncertainties. This approach guarantees that the uncertainties are propagated appropriately, making these PDFs highly suitable for precision physics studies. Our results highlight the importance of continuously updating PDF determinations with the latest experimental data to achieve percent-level accuracy in theoretical predictions for high-energy processes. This level of precision is crucial for interpreting measurements at the LHC, particularly in processes involving electroweak boson production, jet production, and beyond, where a precise knowledge of the partonic structure is necessary. The simulated data from the future Electron-Ion Collider also

play a significant role in improving our understanding of the partonic structure of the proton. We have shown that the inclusion of EIC simulated data, in combination with jet and dijet production data from HERA, has the potential to not only improve the determination of the gluon distribution but also enhance the precision of the determination of the strong coupling constant $\alpha_s(M_Z^2)$. These improvements underscore the critical role of future facilities like the EIC in global QCD analyses, particularly for better constraining gluon and sea-quark distributions in previously unexplored kinematic regions. Furthermore, the sensitivity study conducted on $\alpha_s(M_Z^2)$ demonstrates the complementary nature of different datasets, such as DIS, Drell-Yan, W/Z boson, and jet production data, in constraining the strong coupling constant.

XII. AVAILABILITY OF PROTON PDFS SETS

Our proton PDFs, determined at NLO and NNLO in perturbative QCD, have been made available in the LHAPDF library to facilitate their use in a wide variety of high-energy physics analyses. The LHAPDF format allows easy integration of these PDFs with various Monte Carlo event generators, QCD analysis tools, and simulation software used in collider experiments. These PDF sets, corresponding to different QCD fit scenarios (e.g., Fit D, Fit D + EIC, and Fit D + EIC + jet/dijet), are intended to provide flexibility for experimental analyses requiring precise PDF information across different kinematic regions. The uncertainties in the PDFs are quantified using the Hessian method, and each PDF set includes eigenvector variations that can be used to assess the uncertainties in observables. The Hessian PDF sets are also available in the LHAPDF library, allowing users to perform detailed uncertainty propagation in their analyses. The PDF sets are formatted in compliance with the LHAPDF standard (versions 6 and above) to ensure maximum compatibility with modern QCD analysis frameworks and simulation-based analyses. One can access our PDF sets via email upon request, where they are categorized based on the fit scenario (e.g., Fit D and Fit D + EIC).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the helpful discussions and insightful comments provided by, Oleksandr Zenaiev, and Andrei Kataev, which significantly elevated and enriched the quality of this paper. The authors also thank the School of Particles and Accelerators at the Institute for Research in Fundamental Sciences for their financial support of this project. Hamzeh Khanpour appreciates the financial support from NAWA under Grant No. BPN/ULM/2023/1/00160, as well as from the IDUB program at AGH University. The work of U.-G. M. was also supported in part by the Chinese Academy of Sciences President's International Fellowship Initiative (Grant No. 2025PD0022).

DATA AVAILABILITY

The data that support the findings of this article are openly available [105].

[1] D. M. South and M. Turcato, Review of searches for rare processes and physics beyond the standard model at HERA, Eur. Phys. J. C **76**, 336 (2016).
[2] F. Gross, E. Klempt, S. J. Brodsky, A. J. Buras, V. D. Burkert, G. Heinrich, K. Jakobs, C. A. Meyer, K. Orginos, M. Strickland *et al.*, 50 years of quantum chromodynamics, Eur. Phys. J. C **83**, 1125 (2023).
[3] J. Blümlein, Deep-inelastic scattering: What do we know?, Int. J. Mod. Phys. A **39**, 2441004 (2024).
[4] J. Blumlein, The theory of deeply inelastic scattering, Prog. Part. Nucl. Phys. **69**, 28 (2013).
[5] J. Gao, L. Harland-Lang, and J. Rojo, The structure of the proton in the LHC precision era, Phys. Rep. **742**, 1 (2018).
[6] R. Abdul Khalek, A. Accardi, J. Adam, D. Adamiak, W. Akers, M. Albaladejo, A. Al-bataineh, M. G. Alexeev, F. Ameli, P. Antonioli *et al.*, Science requirements and detector concepts for the electron-ion collider: EIC yellow report, Nucl. Phys. **A1026**, 122447 (2022).
[7] R. Abdul Khalek, U. D'Alesio, M. Arratia, A. Bacchetta, M. Battaglieri, M. Begel, M. Boglione, R. Boughezal, R. Boussarie, G. Bozzi *et al.*, Snowmass 2021 white paper: Electron ion collider for high energy physics, arXiv: 2203.13199.
[8] P. Agostini *et al.* (LHeC and FCC-he Study Groups), The large hadron–electron collider at the HL-LHC, J. Phys. G **48**, 110501 (2021).
[9] A. Abada *et al.* (FCC Collaboration), FCC physics opportunities: Future circular collider conceptual design report volume 1, Eur. Phys. J. C **79**, 474 (2019).
[10] A. Abada *et al.* (FCC Collaboration), FCC-hh: The hadron collider: Future circular collider conceptual design report volume 3, Eur. Phys. J. Special Topics **228**, 755 (2019).
[11] A. Abada *et al.* (FCC Collaboration), HE-LHC: The high-energy large hadron collider: Future circular collider conceptual design report volume 4, Eur. Phys. J. Special Topics **228**, 1109 (2019).
[12] T. J. Hou, J. Gao, T. J. Hobbs, K. Xie, S. Dulat, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt *et al.*, New CTEQ global analysis of quantum chromodynamics with

high-precision data from the LHC, Phys. Rev. D **103**, 014013 (2021).
[13] A. Ablat, A. Courtoy, S. Dulat, M. Guzzi, T. J. Hobbs, T. J. Hou, J. Huston, K. Mohan, H. W. Lin, P. Nadolsky *et al.*, New results in the CTEQ-TEA global analysis of parton distributions in the nucleon, Eur. Phys. J. Plus **139**, 1063 (2024).
[14] R. D. Ball *et al.* (NNPDF Collaboration), The path to proton structure at 1% accuracy, Eur. Phys. J. C **82**, 428 (2022).
[15] J. Cruz-Martinez, S. Forte, N. Laurenti, T. R. Rabemananjara, and J. Rojo, LO, NLO, and NNLO parton distributions for LHC event generators, J. High Energy Phys. 09 (2024) 088.
[16] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs, Eur. Phys. J. C **81**, 341 (2021).
[17] I. Abt *et al.* (H1 and ZEUS Collaborations), 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).
[18] G. Aad *et al.* (ATLAS Collaboration), Determination of the parton distribution functions of the proton using diverse ATLAS data from pp collisions at $\sqrt{s} = 7$, 8 and 13 TeV, Eur. Phys. J. C **82**, 438 (2022).
[19] S. Alekhin, M. V. Garzelli, S. O. Moch, and O. Zenaiev, NNLO PDFs driven by top-quark data, arXiv:2407.00545.
[20] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions from high-precision collider data, Eur. Phys. J. C **77**, 663 (2017).
[21] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, Parton distributions in the LHC era: MMHT 2014 PDFs, Eur. Phys. J. C **75**, 204 (2015).
[22] L. A. Harland-Lang, T. Cridge, and R. S. Thorne, A stress test of global PDF fits: Closure testing the MSHT PDFs and a first direct comparison to the neural net approach, arXiv:2407.07944.
[23] A. M. Cooper-Sarkar and K. Wichmann, QCD analysis of the ATLAS and CMS $W^{\pm}$ and Z cross-section measurements and implications for the strange sea density, Phys. Rev. D **98**, 014027 (2018).
[24] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk (JAM Collaboration), Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions, Phys. Rev. D **101**, 074020 (2020).
[25] N. Armesto, T. Cridge, F. Giuli, L. Harland-Lang, P. Newman, B. Schmookler, R. Thorne, and K. Wichmann, Impact of inclusive electron ion collider data on collinear parton distributions, Phys. Rev. D **109**, 054019 (2024).
[26] S. Cerci, Z. S. Demiroglu, A. Deshpande, P. R. Newman, B. Schmookler, D. Sunar Cerci, and K. Wichmann, Extraction of the strong coupling with HERA and EIC inclusive data, Eur. Phys. J. C **83**, 1011 (2023).
[27] J. C. Collins, D. E. Soper, and G. F. Sterman, Factorization of hard processes in QCD, Adv. Ser. Dir. High Energy Phys. **5**, 1 (1989).
[28] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Parton distributions for the LHC, Eur. Phys. J. C **63**, 189 (2009).
[29] S. Alekhin, J. Blumlein, and S. Moch, Parton distribution functions and benchmark cross sections at NNLO, Phys. Rev. D **86**, 054009 (2012).
[30] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald, The massless three-loop Wilson coefficients for the deep-inelastic structure functions F_2, F_L, xF_3 and g_1, J. High Energy Phys. 11 (2022) 156.
[31] J. Blümlein, A. De Freitas, C. G. Raab, and K. Schönwald, The unpolarized two-loop massive pure singlet Wilson coefficients for deep-inelastic scattering, Nucl. Phys. **B945**, 114659 (2019).
[32] J. Blümlein, J. Ablinger, A. Behring, A. De Freitas, A. von Manteuffel, C. Schneider, and C. Schneider, Heavy flavor Wilson coefficients in deep-inelastic scattering: Recent results, Proc. Sci. QCDEV2017 (**2017**) 031 [arXiv:1711.07957].
[33] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider, The 3-loop pure singlet heavy flavor contributions to the structure function $F_2(x, Q^2)$ and the anomalous dimension, Nucl. Phys. **B890**, 48 (2014).
[34] J. Ablinger, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, C. Schneider, and F. Wissbrock, The transition matrix element $A_{gq}(N)$ of the variable flavor number scheme at $O(\alpha_s^3)$, Nucl. Phys. **B882**, 263 (2014).
[35] R. S. Towell *et al.* (NuSea Collaboration), Improved measurement of the anti-d / anti-u asymmetry in the nucleon sea, Phys. Rev. D **64**, 052002 (2001).
[36] R. Hamberg, W. L. van Neerven, and T. Matsuura, A complete calculation of the order $\alpha - s^2$ correction to the Drell-Yan K factor, Nucl. Phys. **B359**, 343 (1991); **B644**, 403(E) (2002).
[37] P. M. Nadolsky, Theory of W and Z boson production, AIP Conf. Proc. **753**, 158 (2005).
[38] S. Sapeta, QCD and jets at hadron colliders, Prog. Part. Nucl. Phys. **89**, 1 (2016).
[39] S. Catani and M. H. Seymour, A general algorithm for calculating jet cross-sections in NLO QCD, Nucl. Phys. **B485**, 291 (1997); **B510**, 503(E) (1998).
[40] R. V. Harlander and W. B. Kilgore, Next-to-next-to-leading order Higgs production at hadron colliders, Phys. Rev. Lett. **88**, 201801 (2002).
[41] S. Moch, J. A. M. Vermaseren, and A. Vogt, The three loop splitting functions in QCD: The Nonsinglet case, Nucl. Phys. **B688**, 101 (2004).
[42] A. Vogt, S. Moch, and J. A. M. Vermaseren, The three-loop splitting functions in QCD: The singlet case, Nucl. Phys. **B691**, 129 (2004).
[43] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider, The three-loop splitting functions $P_{qg}^{(2)}$ and $P_{gg}^{(2,N_F)}$, Nucl. Phys. **B922**, 1 (2017).
[44] V. Bertone and A. Karlberg, Benchmark of deep-inelastic-scattering structure functions at $\mathcal{O}(\alpha_s^3)$, Eur. Phys. J. C **84**, 774 (2024).

[45] R. S. Thorne, A variable-flavor number scheme for NNLO, Phys. Rev. D **73**, 054019 (2006).

[46] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, Neural network parametrization of deep inelastic structure functions, J. High Energy Phys. 05 (2002) 062.

[47] A. Barontini, A. Candido, F. Hekhorn, G. Magni, and R. Stegeman, An FONLL prescription with coexisting flavor number PDFs, J. High Energy Phys. 10 (2024) 004.

[48] M. A. G. Aivazis, F. I. Olness, and W. K. Tung, Leptoproduction of heavy quarks. 1. General formalism and kinematics of charged current and neutral current production processes, Phys. Rev. D **50**, 3085 (1994).

[49] J. C. Collins, Hard scattering factorization with heavy quarks: A general treatment, Phys. Rev. D **58**, 094002 (1998).

[50] M. Buza, Y. Matiounine, J. Smith, and W. L. van Neerven, Charm electroproduction viewed in the variable flavor number scheme versus fixed order perturbation theory, Eur. Phys. J. C **1**, 301 (1998).

[51] R. S. Thorne and R. G. Roberts, An ordered analysis of heavy flavor production in deep inelastic scattering, Phys. Rev. D **57**, 6871 (1998).

[52] H. Abramowicz, 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).

[53] M. Aaboud *et al.* (ATLAS Collaboration), Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector, Eur. Phys. J. C **77**, 367 (2017).

[54] G. Aad *et al.* (ATLAS Collaboration), Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s} =$ 7 TeV using the ATLAS detector, J. High Energy Phys. 06 (2014) 112.

[55] G. Aad *et al.* (ATLAS Collaboration), Measurement of the high-mass Drell–Yan differential section in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector, Phys. Lett. B **725**, 223 (2013).

[56] M. Aaboud *et al.* (ATLAS Collaboration), Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s} = 8$ TeV, J. High Energy Phys. 12 (2017) 059.

[57] G. Aad *et al.* (ATLAS Collaboration), Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, J. High Energy Phys. 08 (2016) 009.

[58] G. Aad *et al.* (ATLAS Collaboration), Determination of the strange quark density of the proton from ATLAS measurements of the $W \to \ell\nu$ and $Z \to \ell\ell$ cross sections, Phys. Rev. Lett. **109**, 012001 (2012).

[59] S. Chatrchyan *et al.* (CMS Collaboration), Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV, Phys. Rev. Lett. **109**, 111806 (2012).

[60] S. Chatrchyan *et al.* (CMS Collaboration), Measurement of the muon charge asymmetry in inclusive $pp \to W + X$ production at $\sqrt{s} = 7$ TeV and an improved determination of light parton distribution functions, Phys. Rev. D **90**, 032004 (2014).

[61] V. Khachatryan *et al.* (CMS Collaboration), Measurement of the differential cross section and charge asymmetry for inclusive pp $\to W^{\pm} + X$ production at $\sqrt{s} = 8$ TeV, Eur. Phys. J. C **76**, 469 (2016).

[62] S. Chatrchyan *et al.* (CMS Collaboration), Measurement of the rapidity and transverse momentum distributions of Z bosons in pp collisions at $\sqrt{s} = 7$ TeV, Phys. Rev. D **85**, 032002 (2012).

[63] T. Aaltonen *et al.* (CDF Collaboration), Direct measurement of the W production charge asymmetry in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. Lett. **102**, 181801 (2009).

[64] T. A. Aaltonen *et al.* (CDF Collaboration), Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z mass region from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, Phys. Lett. B **692**, 232 (2010).

[65] V. M. Abazov *et al.* (D0 Collaboration), Measurement of the electron charge asymmetry in $p\bar{p} \to W + X \to e\nu + X$ decays in $\boldsymbol{p\bar{p}}$ collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D **91**, 032007 (2015); **91**, 079901(E) (2015).

[66] V. M. Abazov *et al.* (D0 Collaboration), Measurement of the muon charge asymmetry in $p\bar{p} \to \mathrm{W} + \mathrm{X} \to \mu\nu + X$ events at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D **88**, 091102 (2013).

[67] V. M. Abazov *et al.* (D0 Collaboration), Measurement of the shape of the boson rapidity distribution for $p\bar{p} \to Z/\gamma^* \to e^+e^- + X$ events produced at $\sqrt{s}$ of 1.96-TeV, Phys. Rev. D **76**, 012003 (2007).

[68] H. Abdolmaleki *et al.* (xFitter Collaboration), xFitter: An open source QCD analysis framework. A resource and reference document for the Snowmass study, arXiv:2206.12465.

[69] S. Alekhin, O. Behnke, P. Belov, S. Borroni, M. Botje, D. Britzger, S. Camarda, A. M. Cooper-Sarkar, K. Daum, C. Diaconu *et al.*, HERAFitter, Eur. Phys. J. C **75**, 304 (2015).

[70] G. Altarelli and G. Parisi, Asymptotic freedom in Parton language, Nucl. Phys. **B126**, 298 (1977).

[71] V. N. Gribov and L. N. Lipatov, Deep inelastic e p scattering in perturbation theory, Sov. J. Nucl. Phys. **15**, 438 (1972).

[72] 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).

[73] F. James and M. Roos, MINUIT: A system for function minimization and analysis of the parameter errors and correlations, Comput. Phys. Commun. **10**, 343 (1975).

[74] F. James, MINUIT function minimization and error analysis: Reference manual version 94.1, Report No. CERN-D-506.

[75] R. S. Thorne, Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections, Phys. Rev. D **86**, 074017 (2012).

[76] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), 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).

[77] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, MRST2001: Partons and α_s from precise deep inelastic scattering and Tevatron jet data, Eur. Phys. J. C **23**, 73 (2002).

[78] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, LHAPDF6:

Parton density access in the LHC precision era, Eur. Phys. J. C **75**, 132 (2015).
[79] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Uncertainties of predictions from parton distribution functions. 2. The Hessian method, Phys. Rev. D **65**, 014013 (2001).
[80] W. T. Giele, S. A. Keller, and D. A. Kosower, Parton distribution function uncertainties, arXiv:hep-ph/0104052.
[81] V. D. Burkert, L. Elouadrhiri, A. Afanasev, J. Arrington, M. Contalbrigo, W. Cosyn, A. Deshpande, D. I. Glazier, X. Ji, S. Liuti *et al.*, Precision studies of QCD in the low energy domain of the EIC, Prog. Part. Nucl. Phys. **131**, 104032 (2023).
[82] A. Accardi, J. L. Albacete, M. Anselmino, N. Armesto, E. C. Aschenauer, A. Bacchetta, D. Boer, W. K. Brooks, T. Burton, N. B. Chang *et al.*, Electron ion collider: The next QCD frontier: Understanding the glue that binds us all, Eur. Phys. J. A **52**, 268 (2016).
[83] J. Adam *et al.* (ATHENA Collaboration), ATHENA detector proposal—A totally hermetic electron nucleus apparatus proposed for IP6 at the electron-ion collider, J. Instrum. **17**, P10019 (2022).
[84] A. Aktas *et al.* (H1 Collaboration), Measurement of inclusive jet production in deep-inelastic scattering at high Q^2 and determination of the strong coupling, Phys. Lett. B **653**, 134 (2007).
[85] F. D. Aaron *et al.* (H1 Collaboration), Jet production in ep collisions at Low Q^2 and determination of α_s, Eur. Phys. J. C **67**, 1 (2010).
[86] S. Chekanov *et al.* (ZEUS Collaboration), Inclusive jet cross-sections in the Breit frame in neutral current deep inelastic scattering at HERA and determination of α_s, Phys. Lett. B **547**, 164 (2002).
[87] S. Chekanov *et al.* (ZEUS Collaboration), Inclusive-jet and dijet cross-sections in deep inelastic scattering at HERA, Nucl. Phys. **B765**, 1 (2007).
[88] H. Abramowicz *et al.* (ZEUS Collaboration), Inclusive dijet cross sections in neutral current deep inelastic scattering at HERA, Eur. Phys. J. C **70**, 965 (2010).
[89] T. Kluge, K. Rabbertz, and M. Wobisch, FastNLO: Fast pQCD calculations for PDF fits, in *Deep Inelastic Scattering DIS 2006* (World Scientific, Singapore, 2007).
[90] I. Abt *et al.* (ZEUS Collaboration), Measurement of jet production in deep inelastic scattering and NNLO determination of the strong coupling at ZEUS, Eur. Phys. J. C **83**, 1082 (2023).
[91] G. Aad *et al.* (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, J. High Energy Phys. 07 (2023) 085.
[92] 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, Report No. ATLAS-CONF-2023-015.
[93] A. Tumasyan *et al.* (CMS Collaboration), Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 02 (2022) 142.
[94] V. Andreev *et al.* (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); **81**, 738(E) (2021).
[95] G. Aad *et al.* (ATLAS Collaboration), Measurement of transverse energy-energy correlations in multi-jet events in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector and determination of the strong coupling constant $\alpha_s(m_Z)$, Phys. Lett. B **750**, 427 (2015).
[96] M. Aaboud *et al.* (ATLAS Collaboration), Determination of the strong coupling constant α_s from transverse energy–energy correlations in multijet events at $\sqrt{s} = 8$ TeV using the ATLAS detector, Eur. Phys. J. C **77**, 872 (2017).
[97] M. Aaboud *et al.* (ATLAS Collaboration), Measurement of dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and determination of the strong coupling, Phys. Rev. D **98**, 092004 (2018).
[98] V. Khachatryan *et al.* (CMS Collaboration), Measurement of the inclusive 3-jet production differential cross section in proton–proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range, Eur. Phys. J. C **75**, 186 (2015).
[99] V. Khachatryan *et al.* (CMS Collaboration), Constraints on parton distribution functions and extraction of the strong coupling constant from the inclusive jet cross section in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C **75**, 288 (2015).
[100] V. Khachatryan *et al.* (CMS Collaboration), Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV, J. High Energy Phys. 03 (2017) 156.
[101] CMS Collaboration, Determination of the strong coupling constant from the measurement of inclusive multijet event cross sections in pp collisions at $\sqrt{s} = 8$ TeV, Report No. CMS-PAS-SMP-16-008.
[102] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of the $\mathrm{t\bar{t}}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C **79**, 368 (2019).
[103] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of $\mathrm{t\bar{t}}$ normalised multi-differential cross sections in pp collisions at $\sqrt{s} = 13$ TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions, Eur. Phys. J. C **80**, 658 (2020).
[104] S. Navas *et al.* (Particle Data Group), Review of particle physics, Phys. Rev. D **110**, 030001 (2024).
[105] https://www.hepdata.net/