JLAB-THY-22-xxxx

How well do we know the gluon polarization in the proton?

Y. Zhou,[1,2,3,4] N. Sato,[4] and W. Melnitchouk[4]

[1] *Guangdong Provincial Key Laboratory of Nuclear Science, Institute of Quantum Matter, and Guangdong-Hong Kong Joint Laboratory of Quantum Matter, Southern Nuclear Science Computing Center, South China Normal University, Guangzhou 510006, China*

[2] *Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA*

[3] *Department of Physics, William and Mary, Williamsburg, Virginia 23187, USA*

[4] *Jefferson Lab, Newport News, Virginia 23606, USA*

Jefferson Lab Angular Momentum (JAM) Collaboration

(Dated: January 1, 2022)

Abstract

We perform the first simultaneous global QCD analysis of spin-averaged and spin-dependent parton distribution functions (PDFs), including single jet production data from unpolarized and polarized hadron collisions. We critically assess the impact of SU(3) flavor symmetry and PDF positivity assumptions on the quark and gluon helicity PDFs, and find strong bias from these particularly on the gluon polarization. The simultaneous analysis allows for the first time extraction of individual helicity aligned and antialigned PDFs with a consistent treatment of uncertainties.

I. INTRODUCTION

Ever since the discovery by the European Muon Collaboration that only a small fraction ($\lesssim 10\% - 30\%$) of the proton's spin is derived from quarks [1], understanding the decomposition of the proton spin into its fundamental components has challenged the hadron physics community for over 3 decades. Some initial explanations focused on potentially large cancellations from gluonic contributions via the axial anomaly [2, 3], or from large negative strange quark polarization, in violation of the Ellis-Jaffe sum rule [4], or from large higher twist effects that could obscure a simple partonic interpretation [5].

Subsequent experiments at CERN, SLAC, DESY and Jefferson Lab involving polarized inclusive or semi-inclusive deep-inelastic lepton-nucleon scattering over a broad range of kinematics, as well as polarized proton-proton collisions at RHIC, have provided a substantive body of data that have largely confirmed the original conclusion of a small total quark polarization [6–8]. Moreover, recent global QCD analyses, especially ones which do not impose theoretical constraints from SU(3) flavor symmetry [9], have typically favored a fairly small strange quark polarization, consistent with zero, effectively sidelining polarized strangeness from playing a significant role in the proton spin puzzle.

The question of where are the missing pieces of the proton spin has inspired studies of other possible sources, such as gluon helicity or quark and gluon orbital angular momentum [10, 11]. The latter can be related to moments of generalized parton distributions (GPDs) [12–14], the determination of which has motivated experimental programs of high-energy exclusive reaction measurements, such as in deeply-virtual Compton scattering. While progress has been made on the theoretical side with improvements in lattice QCD calculations of GPD moments and Compton form factors [15], the phenomenological information about parton orbital angular momentum is relatively sparse, with experimental programs still largely in their infancy [16].

In contrast, the extraction of gluon helicity has matured to a somewhat more advanced stage, with jet production data available from polarized pp collisions at RHIC [17–24]. Inclusive jet cross sections offer direct sensitivity to the gluon momentum fraction and helicity distributions, without complications associated with final state hadronization [25]. In a seminal 2014 analysis, de Florian *et al.* (DSSV) [26] used the RHIC jet production data to extract the first clearly nonzero signal for a polarized gluon distribution in the proton for

gluon momentum fractions x between $x \approx 0.05$ and ≈ 0.2. While the determination of the total gluon helicity in the proton is still subject to large extrapolation uncertainties in the unmeasured small-x region, the establishment of a positive gluon polarization was a major milestone in the developing story of the proton spin decomposition.

In parallel developments, the Jefferson Lab Angular Momentum (JAM) Collaboration has recently pioneered advances in global QCD analysis with simultaneous Bayesian Monte Carlo determination of different types of co-dependent distributions, such as PDFs and fragmentation functions [9, 27, 28], as well as polarized and unpolarized PDFs [29]. The studies found important correlations between the shapes of the inferred unpolarized [27, 28] and polarized [9] strange quark distributions and inputs assumed for fragmentation functions, along with theoretical biases imposed on the analysis.

In this paper we use the JAM global QCD framework to study the gluon helicity distribution, and in particular the robustness of the extracted signal in view of theoretical assumptions made in previous global analyses [26, 30]. Most common of these is the imposition of SU(3) flavor symmetry in relating the octet and singlet axial charges, as well as positivity constraints on the x dependence of the distributions [31, 32]. In addition, we simultaneously determine both the spin-averaged and spin-dependent PDFs by fitting to unpolarized and polarized lepton and hadron scattering data, including jet production cross sections from unpolarized and polarized pp collisions (and $p\bar{p}$ collisions for spin-averaged scattering). The simultaneous analysis allows for the first time extraction of individual helicity-aligned and antialigned PDFs with a consistent treatment of uncertainties.

We perform a careful study of various scenarios employing different theoretical assumptions, and find that indeed the sea quark and gluon helicity distributions can depend strongly on the constraints imposed. In particular, without restricting PDFs to be positive and assuming SU(3) flavor symmetry for the axial vector charges, existing polarized data allow solutions containing *negative* gluon polarization, in addition to the standard positive gluon solutions found in previous analyses, giving equally acceptable descriptions of the data. Interestingly, a similar double solution was also found earlier by the COMPASS Collaboration in an extraction of spin-dependent PDFs from their proton measurements combined with world inclusive DIS data [33]. We conclude that further input from higher-precision measurements of existing or possibly new observables over a range of kinematics is needed in order to draw firmer conclusions about gluon polarization and its contribution to the proton

spin budget.

The organization of this paper is as follows. In Sec. II we briefly summarize the theoretical foundations on which this work is based, including a discussion of collinear factorization, Mellin space techniques, and PDF parametrizations. The data analysis framework is presented in Sec. III, with details about Bayesian inference, Monte Carlo sampling, and the JAM multi-step strategy. Also included in Sec. III is a summary of the inputs used in the analysis, surveying the experimental data sets fitted and the theoretical scenarios explored in this work. The results of the global QCD fits are presented in Section IV, where we discuss in detail the shapes of the spin-averaged and spin-dependent PDFs, focusing in particular on the determination of the gluon helicity distribution. We also infer for the first time the PDFs in the helicity-basis from the combined unpolarized and polarized PDF analysis with a consistent treatment of PDF uncertainties. Finally, Sec. V summarizes the results and discusses the implications of our analysis. In Appendix A we provide a brief explanation about the statistical tool using the area under the curve of the receiver operating characteristic used to visualize the results in Sec. IV.

II. THEORETICAL FRAMEWORK

In this section we outline the main elements of the theoretical framework on which our analysis is based, with a summary of the essential results from collinear QCD factorization, the use of Mellin space techniques for the scale evolution, and the choice of parametrization for the PDFs.

A. QCD factorization

For our QCD global analysis of spin-averaged and spin-dependent PDFs we consider data on physical observables available from processes for which QCD factorization theorems exist in leading power collinear factorization. These include unpolarized and polarized inclusive deep-inelatic scattering (DIS) from protons (p) and deuterons (D) (and ^{3}He for polarized), Drell-Yan lepton-pair production in unpolarized pp and pD scattering, and inclusive jet production in unpolarized pp and $p\bar{p}$ collisions, and in polarized pp collisions. These are summarized in Table I, where we also indicate the relevant factorization between

TABLE I. Processes studied in this global QCD analysis of spin-averaged (f_i) and spin-dependent (Δf_i) PDFs, including relevant variables and schematic factorization representation. Here N represents a proton p or neutron n, with n extracted from either deuterium D or ^{3}He data.

Reaction	Variables	Factorization
spin-averaged		
$\ell + N \ \to \ \ell' + X$	$x_{_B}$, Q^2	$\mathcal{H}_i^{\rm DIS} \otimes f_i$
inclusive DIS		
$p + N \ \to \ \ell^+ + \ell^- + X$	$x_{_F}$, Q^2	$\mathcal{H}_{ij}^{\rm DY} \otimes f_i \otimes f_j$
Drell-Yan lepton-pair production		
$p + p(\bar{p}) \ \to \ {\rm jet} + X$	$y_{\rm jet}$, $p_{\rm jet}^T$	$\mathcal{H}_{ij}^{\rm jet} \otimes f_i \otimes f_j$
inclusive jet production		
spin-dependent		
$\vec{\ell} + \vec{N} \to \ \ell' + X$	$x_{_B}$, Q^2	$\Delta\mathcal{H}_i^{\rm DIS} \otimes \Delta f_i$
polarized inclusive DIS		
$\vec{p} + \vec{p} \ \to \ {\rm jet} + X$	$y_{\rm jet}$, $p_{\rm jet}^T$	$\Delta\mathcal{H}_{ij}^{\rm jet} \otimes \Delta f_i \otimes \Delta f_j$
polarized inclusive jet production		

the short-distance partonic cross sections $(\Delta)\mathcal{H}$ describing the hard scattering in perturbative QCD and the corresponding nonperturbative PDFs and kinematic variables involved. The repeated indices i, j are summed over all parton flavors.

For DIS, the cross sections are usually given in terms of the Bjorken scaling variable $x_{_B} = Q^2/2P\cdot(\ell - \ell')$, where P is the four-momenta of the nucleon and ℓ and ℓ' the four-momenta of the incident and scattered leptons, and the four-momentum transfer squared $Q^2 \equiv -(\ell - \ell')^2 \approx 2\,\ell\cdot\ell' > 0$. For the Drell-Yan process, the cross sections are functions of the Feynman scaling variable, defined in terms of the longitudinal components of the lepton pair, $x_{_F} = 2(\ell_L^+ + \ell_L^-)/\sqrt{s}$, where s is the invariant mass squared of the hadronic collision, and $Q^2 \equiv 2\ell\cdot\ell'$. For inclusive jet observables, the variables are the rapidity $y_{\rm jet}$ and transverse momentum of the jet $p_{\rm jet}^T$ in the hadronic center of mass frame.

For polarized observables the short distance cross sections $\Delta\mathcal{H}$ represent the difference between helicity dependent cross sections, $\Delta\mathcal{H} = \mathcal{H}^{++} - \mathcal{H}^{+-}$, where the "+" and "−" denote

configurations with the parton and hadron helicities aligned and antialigned, respectively. For polarized inclusive DIS, the labels "++", "+−" represent the helicty configurations of the initial state lepton and hadron, while for inclusive jet production these corresponds to the the helicty configurations of the incident protons.

In our analysis we use the $\overline{\text{MS}}$ scheme for the renormalization group equations, with the strong coupling $\alpha_s(\mu)$ solved numerically using the beta-function at two loops with $\alpha_s(M_Z) = 0.118$ at the Z-boson mass scale $\mu = M_Z$. The spin-averaged and spin-dependent PDFs are evolved at next-to-leading logarithmic accuracy using the DGLAP evolution equations [34–36] by parametrizing only the light quark and gluon distributions at input scale, $\mu = \mu_0 = 1.27$ GeV. The heavy quark PDFs are generated perturbatively via the evolution equations, and for the physical observables we use the zero-mass variable flavor scheme with a charm quark mass $m_c = 1.28$ GeV and bottom quark mass $m_b = 4.18$ GeV. All short-distance partonic cross sections are evaluated at next-to-leading order (NLO) accuracy in perturbative QCD.

For the jet cross sections we utilize the NLO partonic cross sections from Jäger *et al.* [25], along with the corrected expression from Refs. [37, 38] based on the small cone approximation with the appropriate settings for the jet reconstruction algorithm. In practice we use the cone [39] and k_T-type algorithms [40, 41] to match with the corresponding experimental data sets.

B. Mellin space techniques

To perform our numerical analysis we in practice require a fast evaluation of the observables and an efficient procedure to solve the evolution equations for the PDFs. For the latter, we perform the evolution in Mellin space, which admits simple analytical solutions [42]. For inclusive DIS, the cross section is a one-dimensional convolution of the short-distance partonic cross sections and PDFs, which can be rendered as a Mellin integral, $[A \otimes B](x) = \int_x^1 (\mathrm{d}z/z)\, A(z)\, B(x/z)$, with analytic expressions for the hard functions $\mathcal{H}_i^{\text{DIS}}$ and $\Delta\mathcal{H}_i^{\text{DIS}}$ in Mellin space available in the literature [43].

In contrast, the convolution structure for the Drell-Yan and inclusive jet production cross sections involves a double convolution of the hard cross section and two nonperturbative functions, which in general cannot be rendered as a true Mellin convolution. Instead, we

utilize the Mellin grid approach developed by Stratmann and Vogelsang [44], in which the double convolution is written in the form

$$\mathcal{H}_{ij} \otimes f_i \otimes f_j = \int_{x_1^{\min}}^1 \mathrm{d}x_1 \int_{x_2^{\min}}^1 \mathrm{d}x_2 \; \mathcal{H}_{ij}(x_1, x_2)\, f_i(x_1)\, f_j(x_2)$$
$$= \frac{1}{(2\pi i)^2} \int \mathrm{d}N_1 \int \mathrm{d}N_2 \; F_i(N_1)\, F_j(N_2) \left[\int_{x_1^{\min}}^1 \mathrm{d}x_1 \int_{x_2^{\min}}^1 \mathrm{d}x_2 \; \mathcal{H}_{ij}(x_1, x_2)\, x_1^{-N_1}\, x_2^{-N_2} \right], \quad (1)$$

where x_1 and x_2 denote the partonic fractions in hadrons 1 and 2, respectively. In the second line of Eq. (1) we have replaced the x-space PDFs by their Mellin space representations, $f(x) = 1/(2\pi i) \int \mathrm{d}N \; x^{-N} F(N)$. The factors inside the brackets are independent of the PDFs, and can be calculated for all Mellin moments needed for the double inverse Mellin transform. They contains all of the kinematic dependence of the process inside the hard function $\mathcal{H}_{ij}$ and the limits for the parton momentum fractions $x_{1,2}^{\min}$. The factors can then be the precalculated and the grids stored as lookup tables when evaluating the Drell-Yan lepton-pair and jet production observables.

C. PDF modeling

For the shape of the spin-averaged and spin-dependent PDFs at the input scale μ_0, we use a generic template function defined by

$$\mathrm{T}(x; \boldsymbol{a}, n) = \frac{a_0\, x^{a_1} (1-x)^{a_2} \left(1 + a_3 \sqrt{x} + a_4 x\right)}{\int_0^1 \mathrm{d}x\, x^{n+a_1-1} (1-x)^{a_2} \left(1 + a_3 \sqrt{x} + a_4 x\right)}\,, \quad (2)$$

where $\boldsymbol{a} = \{a_0, \ldots, a_4\}$ is the set of free parameters for each PDF flavor. The template function is normalized with respect to the n-th moment in order to numerically decorrelate the overall normalization parameter a_0 from the shape parameters $a_1, \ldots, a_4$. The spin-averaged PDFs $f_i = f_i(x, \mu_0^2)$ at the input scale are constructed according to

$$u = u_v + 2\bar{u}, \qquad d = d_v + 2\bar{d}, \quad (3a)$$
$$\bar{u} = S + \bar{u}_0, \qquad \bar{d} = S + \bar{d}_0, \quad (3b)$$
$$s = S + s_0, \qquad \bar{s} = S + \bar{s}_0, \quad (3c)$$

where each of the functions u_v, d_v, S, $\bar{u}_0$, $\bar{d}_0$, s_0 and $\bar{s}_0$, along with the gluon PDF $g(x, \mu_0^2)$, is modeled in terms of one template function in Eq. (2) for each distribution. The valence PDFs $u_v = u - \overline{u}$ and $d_v = d - \overline{d}$ are isolated from the sea distributions in order to impose the

quark number sum rules. To ensure integrability of the first moment of each valence PDF, the corresponding a_1 parameters are restricted to be in the range $a_1 > -1$. The parameters a_3 and a_4 are taken to be zero for s_0, $\bar{s}_0$ and S, but are free to vary for all other distributions. The strangeness number sum rule $\int_0^1 \mathrm{d}x\,(s + \bar{s}) = 0$ imposes an additional constraint, which we use to fix the normalization of s_0.

The light sea quark distributions $\bar{u}$, $\bar{d}$, s and $\bar{s}$ are modeled as combinations of a template function S that dominates in the very small-x region, with a_1 in the range $-2 < a_1 < -1$, and a template function for each of $\bar{u}_0$, $\bar{d}_0$, s_0 or $\bar{s}_0$ with $-1 < a_1 < 1$, controlling the shape at intermediate x values. The momentum sum rule is satisfied by adjusting the a_0 parameter of the gluon distribution.

For all template functions we restrict the $(1 - x)$ exponent to be positive, $a_2 > 0$, to ensure that PDFs vanish in the limit $x \to 1$. We choose the value $n = 2$ to normalize the x-dependent factor of the shape functions, since the momentum sum rule requires the existence of the second moment for each PDF flavor. Other choices with $n > 2$ are also possible, and would be compensated by changes in the a_0 parameters.

For the spin-dependent PDFs $\Delta f_i = \Delta f_i(x, \mu_0^2)$ at the input scale we follow a similar strategy, but with a symmetric sea *ansatz* due to the paucity of empirical constrains on the polarized sea quark distributions,

$$\Delta u = \Delta u_v + 2\Delta\bar{u}, \tag{4a}$$

$$\Delta d = \Delta d_v + 2\Delta\bar{d}, \tag{4b}$$

$$\Delta\bar{u} = \Delta\bar{d} = \Delta s = \Delta\bar{s} \equiv \Delta\bar{q}, \tag{4c}$$

where the functions Δu_v and Δd_v are modeled in terms of a template function in Eq. (2). The sea quark helicity distribution $\Delta\bar{q}$ is described by a combination of a template function that dominates at very small x, with the a_1 parameter in the range $a_1 > -1$, and a second template function that controls the shape at intermediate x, with $a_1 > -0.5$, as in the unpolarized case. For modelling the gluon helicity distribution we also adopt two template functions to allow sufficient flexibility for the fits. In contrast to the spin-averaged case, we normalize the x-dependent factors for the spin-dependent template functions with $n = 1$, since all helicity PDFs are required to have finite contributions to the nucleon spin sum. While the normalizations a_0 are in principle free parameters, as we discuss in Sec. III C below, we consider several different scenarios with either SU(2) or SU(3) flavor symmetry im-

posed, which provides additional constraints on the parameters. For all the spin-dependent distributions, we fix the parameters a_3 and a_4 to zero.

In the next section, we present our data analysis framework for constraining the parameters of the spin-averaged and spin-dependent PDFs. An important consequence of our simultaneous analysis will be the possibility to estimate consistently the uncertainty quantification for the individual helicty-basis PDFs, defined in terms of the spin-averaged and spin-dependent distributions as

$$f^{\uparrow} = \frac{1}{2}\big(f + \Delta f\big), \tag{5a}$$

$$f^{\downarrow} = \frac{1}{2}\big(f - \Delta f\big), \tag{5b}$$

for the helicity-aligned and antialigned distributions, respectively.

III. ANALYSIS FRAMEWORK

The analysis framework adopted in this paper is based on the Bayesian Monte Carlo methodology developed in previous JAM analyses [27, 28, 45, 46]. We also discuss additional constraints on the parameters arising from choices for the moments of the helicity PDFs respecting SU(2) or SU(3) flavor symmetry, as well as from positivity constraints on the PDFs (namely, both $f^{\uparrow}$ and $f^{\downarrow}$ being non-negative). Before embarking on those discussion, however, we first review the experimental data sets that will be fitted in this analysis.

A. Experimental data sets

In the following we summarize the types and sources of experimental data for the observables listed in Table I, along with the kinematic cuts imposed in our analysis.

1. Inclusive DIS

The unpolarized fixed target inclusive DIS data included in our global fit are the reconstructed F_2 structure function data from BCDMS [47], SLAC [48], and NMC [49, 50], with cuts $W^2 = M^2 + Q^2(1 - x_{_B})/x_{_B} > 10$ GeV^2 and $Q^2 > m_c^2$, which are chosen to avoid power corrections and nuclear off-shell effects that are known to be more important at large values

of x_B. With the same cuts, we also include the reduced proton neutral current and charged current cross sections from the combined H1 and ZEUS analysis of HERA collider data [51].

The spin-dependent DIS datasets include all fixed target experiments from the EMC [52], SMC [53, 54], COMPASS [33, 55, 56], SLAC [57–62] and HERMES [63, 64] Collaborations, with identical cuts on W^2 and Q^2 as for the unpolarized DIS data [9, 65]. Wherever possible, we use directly the experimental cross section asymmetries $A_{LL} = (\sigma^{++} - \sigma^{+-})/(\sigma^{++} + \sigma^{+-})$ rather than the reconstructed g_1 structure function. This allows the consistent propagation of uncertainties on spin-dependent PDFs stemming from uncertainties in the unpolarized sector that enter through the denominators of the polarization asymmetries. A total of 2680 unpolarized and 365 polarized DIS data points are used in the analysis.

2. *Drell-Yan lepton-pair production*

In addition to inclusive DIS data, we also fit Drell-Yan inclusive lepton-pair production data from pp and pD collisions, available from the Fermilab E866 experiment [66]. The data are selected to exclude regions of invariant mass $M_{\ell^+\ell^-}$ of the lepton-pair in the vicinity of the $q\bar{q}$ resonances, such as the J/ψ, ψ' and Υ states. Furthermore, following Alekhin *et al.* [67], we also apply a cut of $M_{\ell^+\ell^-} > 6$ GeV to avoid tension with DIS data [67], which leaves 250 data points available to be fitted.

3. *Inclusive jet production*

Inclusive jet production data are known for providing unique sensitivity to the gluonic content of hadrons. For unpolarized beams, we use existing $p\overline{p}$ data from the D0 [68] and CDF [69] Collaborations at the Tevatron. In addition, we include for the first time in a global fit the jet production data in pp collisions from the STAR Collaboration at RHIC [17]. For the latter, we include data from the 2003 and 2004 runs, and find that they can be described well for jet transverse momenta $p_T > 8$ GeV.

With polarized proton beams, we use double spin asymmetries A_{LL} from the STAR [17, 19–23] and PHENIX [24] Collaborations at RHIC. We restrict the data to the same p_T range as for the unpolarized jet cross sections in order to guarantee a faithful description of the denominator in the asymmetries. For the renormalization and factorization scales μ_R

and μ_F, respectively, we generally find best agreement for both unpolarized and polarized collisions with a scale $p_T/2$.

The treatment of systematic uncertainties of the jet data used in our analysis requires particular attention. For the CDF data, uncertainties are provided that are only correlated within each rapidity bin. These are treated separately, allowing an overall normalization to move the entire data set uniformly. Furthermore, both D0 [68] and CDF [69] provide parton to hadron correction factors obtained from Monte Carlo simulations [69], which translate the parton level calculation to the hadron level, and these are incorporated in our analysis.

For the STAR jet data, an uncertainty in the relative luminosity measurement usually results in a shift of the A_{LL} data by an additive constant (fully correlated systemic uncertainty) [19], while an uncertainty in the measurement of the polarization magnitude scales the A_{LL} data (normalization uncertainty). We implement these uncertainties in our analysis as described in Sec. III B below. To avoid underestimating uncertainties, we treat weakly correlated uncertainties as uncorrelated, in particular, for the STAR jet A_{LL} data from the 2005 [19], 2009 [20] and 2013 [23] and 2015 [22] runs.

The global kinematic coverage of the unpolarized and polarized datasets is shown in Fig. 1. Clearly, the range of kinematics spanned by the unpolarized data is significantly greater than that of the polarized (by some 2 orders of magnitude at small x and 2 orders of magnitude at high Q^2). Importantly, however, the range covered by the polarized data overlaps with the unpolarized, so that in this range a consistent analysis of spin-dependent and spin-averaged PDFs can be achieved.

B. Bayesian inference

Our analysis framework is based on sampling the Bayesian posterior distribution using the data resampling approach. The posterior distribution is given by a product of the likelihood function $\mathcal{L}$ and a prior distribution π,

$$\rho(\boldsymbol{a}, \boldsymbol{\nu} \,|\, \text{data}) \propto \mathcal{L}(\boldsymbol{a}, \boldsymbol{\nu} \,|\, \text{data})\, \pi(\boldsymbol{a}, \boldsymbol{\nu}), \tag{6}$$

where we distinguish the PDF parameters $\boldsymbol{a}$ from the additional nuisance parameters $\boldsymbol{\nu}$ (see below). For the likelihood function we use a Gaussian of the form

$$\mathcal{L}(\boldsymbol{a}, \boldsymbol{\nu} \,|\, \text{data}) = \exp\left[-\frac{1}{2} \sum_{e,i} \left(\frac{d_{e,i} - t_{e,i}(\boldsymbol{a}, \boldsymbol{\nu})}{\alpha_{e,i}} \right)^2 \right], \tag{7}$$

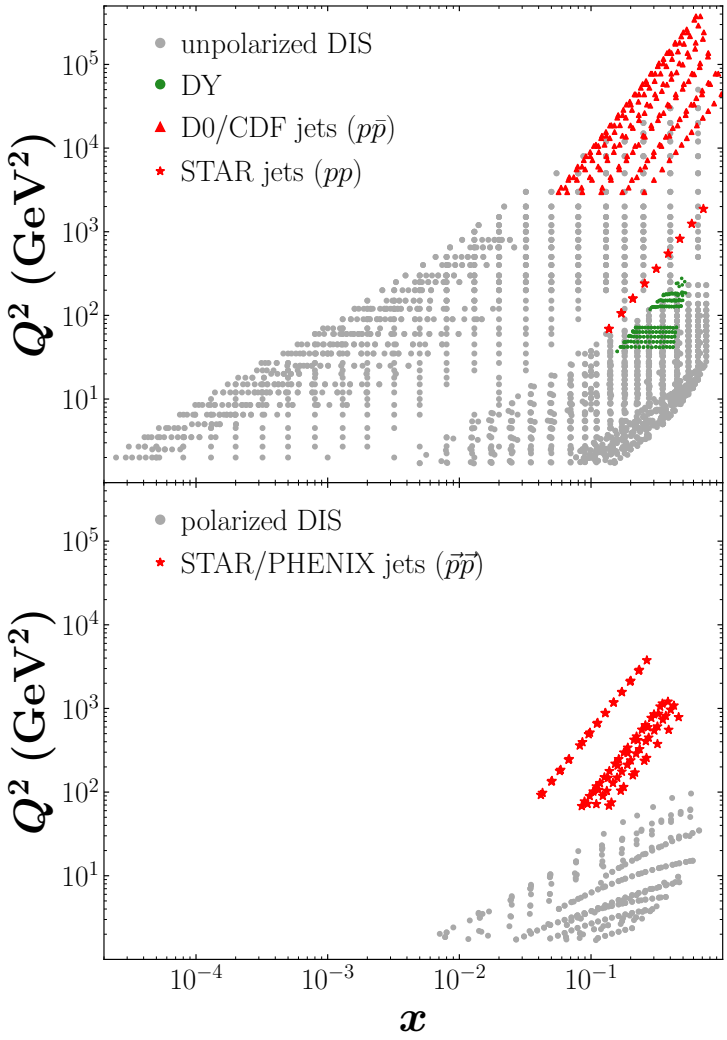

FIG. 1. Kinematic coverage of the unpolarized (upper panel) and polarized (lower panel) datasets used in this analysis. The unpolarized datasets include fixed-target and HERA collider DIS data (gray solid circles), Drell-Yan from pp and pD collisions at Fermilab (green solid circles), and jet production from unpolarized $p\overline{p}$ scattering at Tevatron (red upward triangles) and pp scattering at STAR (red stars). The polarized datasets include spin-dependent fixed-target DIS data (gray solid circles), and jet production in polarized pp scattering at STAR and PHENIX from RHIC (red stars). The variable x denotes the Bjorken scaling variable $x_{_B}$ for DIS and the Feynman variable $x_{_F}$ for Drell-Yan and jet production, while the scale Q^2 represents the four-momentum transfer squared for DIS and DY, and transverse momentum squared for jets.

where $d_{e,i}$ is the i-th data point from experiment e with uncorrelated uncertainty $\alpha_{e,i}$, and $t_{e,i}$ is the corresponding theoretical value. Since the experimental data are distorted by detector effects, we model such distortion with a multiplicative normalization parameter and additive shifts within the quoted systematic uncertainties. Specifically, we compute each theoretical value $t_{e,i}$ according to

$$t_{e,i}(\boldsymbol{a}, \boldsymbol{\nu}) \;=\; \sum_k r_{e,k}\, \beta_{e,k,i} \;+\; \frac{1}{N_e}\, t^0_{e,i}(\boldsymbol{a}), \tag{8}$$

with nuisance parameters $\boldsymbol{\nu} = \{r_{e,k}; N_e\}$, where k labels different types of systematic uncertainties for the i-th data point of the experiment e. The theoretical calculation of the observable $t^0_{e,i}$ is obtained by convoluting the PDFs with the short-distance cross sections, which is then multiplied by a nuisance parameter N_e. The additive shifts are controlled by the nuisance parameters $r_{e,k}$, and weighted by a set of quoted point-by-point correlated systematic uncertainties $\beta_{e,k,i}$.

The prior distribution $\pi(\boldsymbol{a}, \boldsymbol{\nu})$ includes flat priors for the PDF parameters $a_i \in [a_i^{\min}, a_i^{\max}]$ and δ functions to impose the valence quark number and momentum sum rules in the case of unpolarized PDFs. The priors for the nuisance parameters are included as Gaussian penalties. For the normalization parameters, the corresponding priors are modeled in a Gaussian form using the quoted experimental normalization uncertainty, δN_e. The full prior distribution can then be explicitly written as

$$\begin{aligned}\pi(\boldsymbol{a}, \boldsymbol{\nu}) = {} & \delta\big(a_0^{u_v} - \mathrm{SR}(\boldsymbol{a}^{u_v}, u_v)\big)\; \delta\big(a_0^{d_v} - \mathrm{SR}(\boldsymbol{a}^{d_v}, d_v)\big)\; \delta\big(a_0^{g} - \mathrm{SR}(\boldsymbol{a}^{g}, g)\big) \\ & \times \prod_i \theta\big(a_i^{\min} < a_i < a_i^{\max}\big)\; \prod_e \prod_k \exp\Big(-\frac{1}{2} r_{e,k}^2\Big)\; \exp\Big(-\frac{1 - N_e}{2\,\delta N_e}\Big),\end{aligned} \tag{9}$$

where "SR" represents instructions to adjust the a_0 parameters in order to satisfy the appropriate sum rules [70], and the δ functions are implemented analytically.

In adopting the data resampling approach to the construction of Monte Carlo samples for the posterior distribution, we add Gaussian noise to each experimental data point, $d_{e,i} \to d_{e,i} + R_{e,i}\, \alpha_{e,i}$, with statistically-independent normally distributed random numbers $R_{e,i}$. With a given set of distorted data, an optimization is performed in parameter space that maximizes the posterior distribution $\rho(\boldsymbol{a}, \boldsymbol{\nu} \,|\, \mathrm{data} + \mathrm{noise})$. The resulting parameters are added to the list of Monte Carlo samples and the process is repeated $O(1,000)$ times in order

to accumulate sufficient samples to compute statistical estimators for a given observable $\mathcal{O}$

$$\mathrm{E}[\mathcal{O}] = \frac{1}{N}\sum_k \mathcal{O}(\boldsymbol{a}_k), \tag{10a}$$

$$\mathrm{V}[\mathcal{O}] = \frac{1}{N}\sum_k \Big[\mathcal{O}(\boldsymbol{a}_k) - \mathrm{E}[\mathcal{O}]\Big]^2, \tag{10b}$$

where N is the number of parameter samples drawn from the posterior distribution.

While in principle each optimization in the sampling procedure can be carried out with all parameters open, in practice exploring a highly dimensional parameter space (which in our case is ≈ 80, including 30 parameters for spin-averaged PDFs, 18 for spin-dependent PDFs, and 33 normalizations) becomes inefficient due to local minima or vanishing gradients, and only a handful of samples can be collected. This can be overcome by using the multi-step methodology developed in Ref. [27], where the initial parameters for the optimization are pre-tuned by using a restricted set of data for the posterior distribution. Fewer datasets decreases the numerical expense of drawing parameter samples from the posterior via the data resampling, and the procedure can be repeated in several steps with different datasets added sequentially. The resulting samples from a given step become the input samples for the next step, where additional datasets are added to the collection of datasets.

In addition, since the spin-dependent PDFs do not enter in unpolarized physical observables, our sequence of pre-tuning the PDFs parameters can be partitioned into three stages. In the first stage only the unpolarized PDF parameters are tuned. In the second stage the Monte Carlo parameters samples of the unpolarized PDFs are frozen and used to tune the spin-dependent PDF parameters. The final stage uses the collection of pre-optmized input parameters for all PDFs, and a final posterior sampling is performed with all parameters free and all data sets included in the posterior distribution. A sketch of the multi-step strategy for the parameter sampling is illustrated in Fig. 2. Finally, for our diagnostic metric for the global agreement between data and theory, we use the reduced χ^2, defined as

$$\chi^2_{\mathrm{red}} \equiv \frac{\chi^2}{N_{\mathrm{dat}}} = \frac{1}{N_{\mathrm{dat}}}\sum_{e,i}\left(\frac{d_{e,i} - \mathrm{E}[t_{e,i}]}{\alpha_{e,i}}\right)^2, \tag{11}$$

where the expectation value $\mathrm{E}[t_{e,i}]$ for the theoretical quantity $t_{e,i}$ is estimated via Eq. (10a).

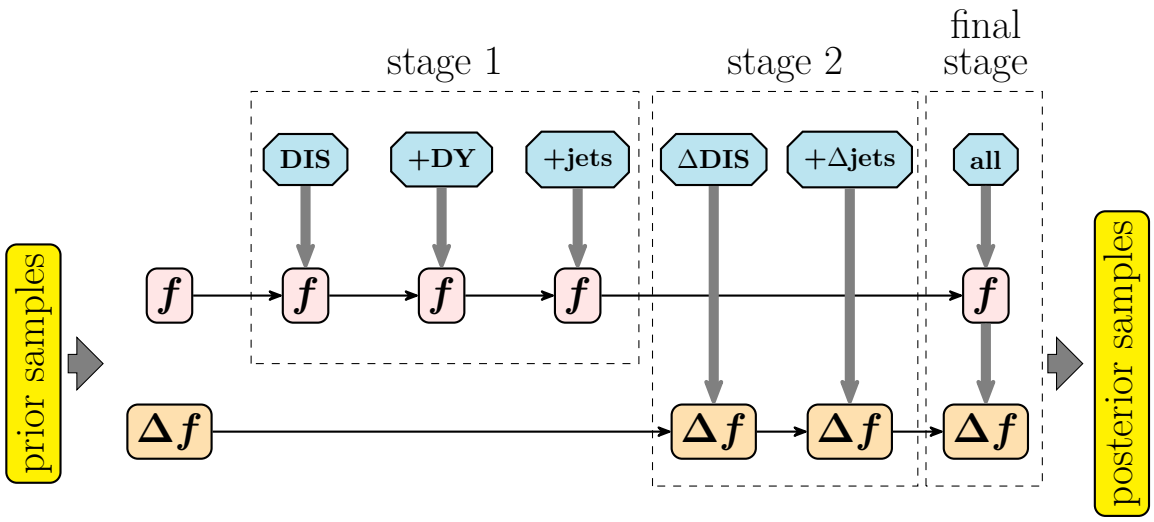

FIG. 2. Schematic overview of the multi-step strategy for parameter sampling employed in the current JAM analysis of spin-averaged (f) and spin-dependent (Δf) PDFs. Unpolarized data include DIS, Drell-Yan (DY), and single jet production in pp and $p\bar{p}$ collisions, while polarized data include polarized DIS (ΔDIS) and jet production in polarized pp collisions (Δjets).

C. Additional constraints

In contrast to the spin-averaged PDFs, where the valence number and momentum sum rules can be imposed based on fundamental physical properties such as baryon number and momentum conservation, no correspondings constraints are available for spin-dependent PDFs. On the other hand, approximate sum rules for spin-dependent PDFs exist involving the triplet and octet axial-vector charges, g_A and a_8, respectively,

$$\int_0^1 \mathrm{d}x \left[\Delta u^+ - \Delta d^+\right](x, Q^2) = g_A, \tag{12a}$$

$$\int_0^1 \mathrm{d}x \left[\Delta u^+ + \Delta d^+ - 2\Delta s^+\right](x, Q^2) = a_8, \tag{12b}$$

where $\Delta q^+ = \Delta q + \Delta \bar{q}$. These relations can be imposed as constraints if the values of g_A and a_8 are known sufficiently accurately. In practice, however, the charges are inferred from neutron beta-decays using SU(2) isospin symmetry, which gives $g_A = 1.269(3)$, and from hyperon beta-decays, which under the assumption of SU(3) flavor symmetry gives $a_8 = 0.586(31)$ [9]. While these constraints are typically imposed in global analyses of spin-dependent PDFs [26, 30, 65], in our analysis we consider several scenarios where one or both

of these are imposed or relaxed. Imposing Eqs. (12a) or (12b) amounts to extending the priors in Eq. (9) according to

$$\begin{aligned} \pi(\boldsymbol{a},\boldsymbol{\nu}) \;\to\; \pi(\boldsymbol{a},\boldsymbol{\nu}) &\times \exp\left(-\frac{a_3^{\rm exp}-a_3^{\rm thy}}{2\,\delta a_3^{\rm exp}}\right) && \text{[SU(2)]} \\ &\times \exp\left(-\frac{a_8^{\rm exp}-a_8^{\rm thy}}{2\,\delta a_8^{\rm exp}}\right), && \text{[SU(3)]} \end{aligned} \tag{13}$$

where the theoretical values $a_{3,8}^{\rm thy}$ are computed using Eqs. (12) and the experimental values $a_{3,8}^{\rm exp}$ and $\delta a_{3,8}^{\rm exp}$ are taken from Ref. [9].

In addition to the SU(2) and SU(3) flavor symmetry constraints on the axial-vector charges, many global analyses also impose phenomenological positivity constraints on the PDFs [26, 30], which requires that

$$\left|\Delta q(x,Q^2)\right| \overset{?}{\leqslant} q(x,Q^2), \qquad \text{[positivity]} \tag{14}$$

for each PDF flavor q, or equivalently that all the helicity-basis PDFs remain positive, $f^{\uparrow/\downarrow}(x,Q^2) \geqslant 0$. The positivity constraints typically affect PDFs in the large-x region, where the absolute magnitudes of the spin-dependent and spin-averaged PDFs become comparable, and can be implemented by selecting the posterior samples that satisfy such criteria. On the other hand, it has been recently argued [32], in contrast to earlier claims in the literature [31], that there is no fundamental requirement for PDFs in the $\overline{\rm MS}$ scheme to remain positive definite.

Whether formally justified or not, it is important to assess the degree to which the constraints (12) and (14) may bias the inference on the PDFs. To this effect, we consider three scenarios, in which we impose either

(i) the **SU(2)** constraint in Eq. (12a);

(ii) the **SU(3)** constraint in Eq. (12b) in addition to (12a); or

(iii) the **SU(3) + positivity** constraints in Eq. (14) in addition to Eqs. (12a) and (12b).

In the next section we analyze in detail the effect on the global QCD analysis of each of these assumptions.

IV. SIMULTANEOUS ANALYSIS RESULTS

Using the theoretical framework and methodology outlined in Secs. II and III, in this section we present the results of our global Monte Carlo analysis, simultaneously extracting both the spin-averaged and spin-dependent PDFs. We begin with the discussion of the spin-averaged PDFs, and in particular the impact of jet data from unpolarized pp collisions at RHIC, obtained at similar kinematics as the polarized pp jet data that are expected to constrain the spin-dependent gluon PDF, Δg. We focus in particular on the determination of Δg in the context of the three scenarios discussed in the previous section, and critically examine the discrimination between the two helicity-basis PDFs under these scenarios. A summary of the fit results, including the $\chi^2_{\rm red}$ values for each type of dataset and for each of the scenarios, is given in Table II. The $\chi^2_{\rm red}$ are computed from Eq. (11) using the average of theory predictions from all replicas in the simultaneous fit.

TABLE II. Results for the simultaneous fits to unpolarized and polarized scattering data, with the reduced $\chi^2_{\rm red} = \chi^2/N_{\rm dat}$ for $N_{\rm dat}$ points for the SU(2), SU(3), and SU(3)+positivity scenarios.

Data	$N_{\rm dat}$	$\chi^2_{\rm red}$		
		SU(2)	**SU(3)**	**SU(3)+pos**
Unpolarized DIS [47–51]	2680	1.20	1.20	1.20
Drell-Yan (pp, pD) [66]	250	1.06	1.06	1.10
Jets				
D0 ($p\overline{p}$) [68]	110	0.89	0.89	0.88
CDF ($p\overline{p}$) [69]	76	1.11	1.11	1.11
STAR 2003 (pp) [17]	3	0.04	0.04	0.04
STAR 2004 (pp) [17]	9	1.06	1.06	1.06
Polarized DIS [33, 52–64]	365	0.92	0.92	0.96
Jets in polarized $\vec{p}\,\vec{p}$				
STAR [17, 19–23]	81	0.82	0.81	0.84
PHENIX [24]	2	0.38	0.38	0.38
Total	**3563**	**1.14**	**1.14**	**1.15**

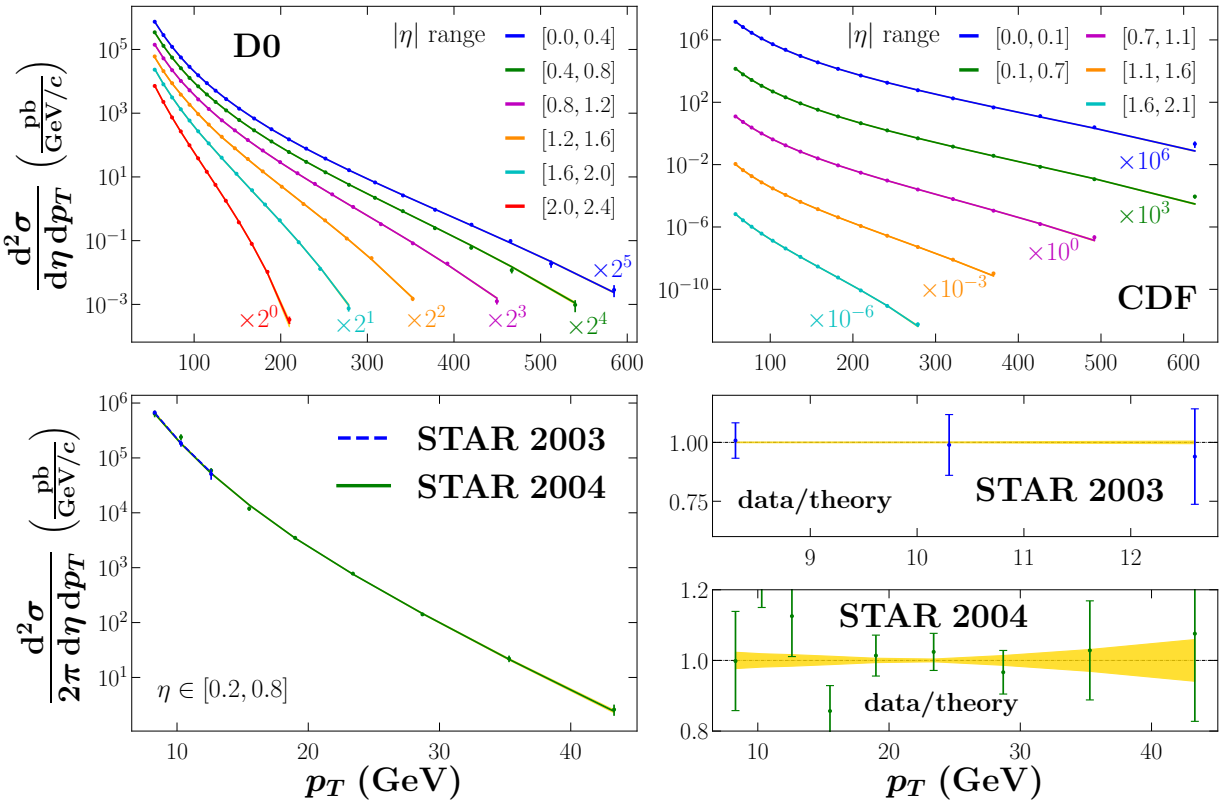

FIG. 3. Comparison with single jet production cross sections in $p\bar{p}$ collisions from D0 [68] and CDF [69], and in pp scattering from STAR [17]. Different pseudorapidity η bins are marked by colors and scaled by factors for D0 and CDF for clarity. Note the extra factor 2π in the STAR cross section data. The data (filled circles) are compared with fits (solid lines) obtained from the average of all Monte Carlo replicas, with 1σ uncertainties (yellow bands). For STAR 2003 and 2004 data, ratios of data to average theory (filled circles) are shown in the bottom right panel.

A. Unpolarized PDFs

As indicated in Table II, good overall agreement is found between our fits and the unpolarized DIS (fixed target and HERA collider), Drell-Yan, and inclusive jet production data. Since the focus of this work is primarily on jet observables (the full fit results for the inclusive DIS and Drell-Yan data comparisons can be found in Ref. [70]), in Fig. 3 we show the differential jet production cross sections, $\mathrm{d}^2\sigma/\,\mathrm{d}\eta\,\mathrm{d}p_T$, for $p\bar{p}$ scattering from D0 and CDF at the Tevatron and pp scattering from STAR at RHIC versus the jet transverse momentum p_T, in specific bins of the pseudorapidity η. The η bins are obtained using their absolute

values for CDF and D0 data, and actual values for STAR. An excellent description of all the jet data is obtained, with $\chi^2_{\rm red} \approx 0.9 - 1.1$ for the D0 and CDF $p\bar{p}$ data, and the STAR 2004 pp data. Since only 3 data points are available for the STAR 2003 run, a $\chi^2_{\rm red} \approx 0$ was obtained for these points. The $\chi^2_{\rm red}$ values for the unpolarized data are, as may be expected, almost independent of the scenario chosen for the spin PDF constraints, Eqs. (12) and (14).

To more graphically illustrate the comparison between data and theory for the STAR pp data, which have not been used in any previous global QCD analysis, we also show in Fig. 3 the data/theory ratios, which our fit describes well within the $\sim 10\% - 20\%$ experimental uncertainties. Note that the cross sections vary by some 7 orders of magnitude for the D0 and CDF data over the range of p_T covered ($p_T \lesssim 600$ GeV), and over 5 orders of magnitude for the STAR data, which span a smaller range of p_T values ($p_T \lesssim 40$ GeV). In principle, the data are available down to rather low p_T values, $p_T \sim$ few GeV. However, some tensions were found when attempting to fit the STAR data with $p_T < 8$ GeV and those with $p_T > 8$ GeV, so that a cut of $p_T > 8$ GeV is made for this analysis. This will have some consequence for the corresponding cut chosen for the polarized scattering data, as we discuss below.

The unpolarized PDFs extracted from the present analysis are shown in Fig. 4 for the valence quark u_v and d_v, light antiquark $\bar{d} + \bar{u}$, and strange $s + \bar{s}$ distributions, as well as the gluon PDF, g, at a scale $Q^2 = 10$ GeV^2. Compared with PDFs from several other global analyses [27, 74–76], the variation between the different PDF sets is relatively small for the valence, light antiquark, and gluon PDFs, while a larger spread is observed for the strange quark distributions. In particular, the magnitude of the strange $s + \bar{s}$ PDFs is slightly larger than the earlier JAM20 [28] (and also JAM19 [27], not shown in Fig. 4) analysis, which found a stronger strange quark suppression due to the inclusion of semi-inclusive DIS and single-inclusive e^+e^- annihilation data, especially for kaon production. We expect that inclusion of the semi-inclusive DIS and e^+e^- annihilation data into the present analysis will produce additional suppression of our strange quark PDF. Also, at higher x values ($x \gtrsim 0.5$) the upward shift in the strangeness PDF is an indirect effect associated with the sensitivity of the jet data to the gluon PDF via the momentum sum rule. However, the detailed structure of the strange quark PDFs does not affect the main goal of our analysis, which is the determination of the spin-dependent gluon distribution.

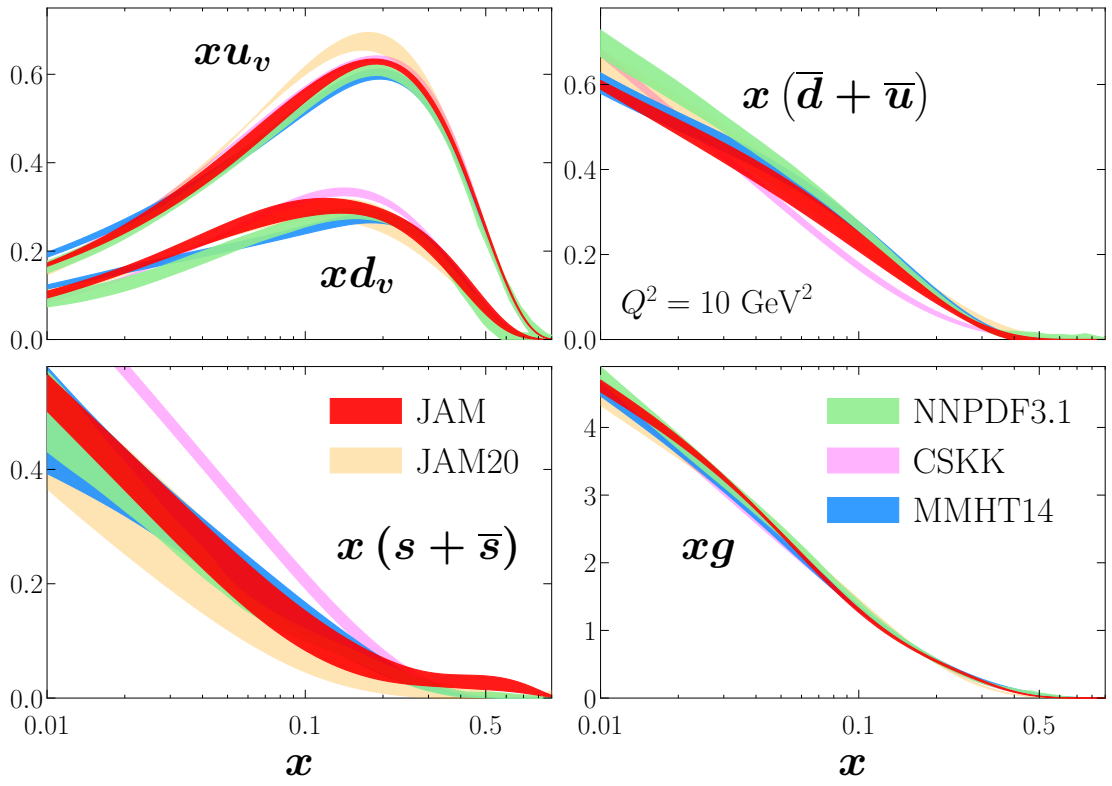

FIG. 4. Comparison of spin-averaged PDFs from the present JAM analysis with other PDF sets from the previous JAM20 analysis [27], and from the NNPDF3.1 [74], CSKK [75], and MMHT14 [76] parametrizations, for the u_v, d_v, $\bar{d}+\bar{u}$, and $s+\bar{s}$ quark and gluon g flavors at $Q^2 = 10$ GeV^2. Note that x times the PDF is shown.

B. Spin-dependent PDFs

As for the polarized case, we obtain an excellent description of the spin-dependent observables, including polarized lepton-nucleon DIS and jet production in polarized pp collisions. For the former, we find a total $\chi^2_{\rm red} \approx 0.9$ (see Ref. [70] for the corresponding DIS data to theory comparisons). For the latter, we show in Fig. 5 the inclusive polarization asymmetries, A_{LL}, for the STAR and PHENIX data, for the SU(3) scenario, with $\chi^2_{\rm red} \approx 0.8$ for the polarized pp jet production data. The two sets of bands represent solutions with $\Delta g > 0$ and $\Delta g < 0$, as we discuss below, with each giving very similar descriptions. Only the fits to the STAR 2005 [19] and 2012 [21] data show noticeable deviations, with $\chi^2_{\rm red} \approx 1.5$ for these sets, which is mostly due to the presence of some outliers in these spectra.

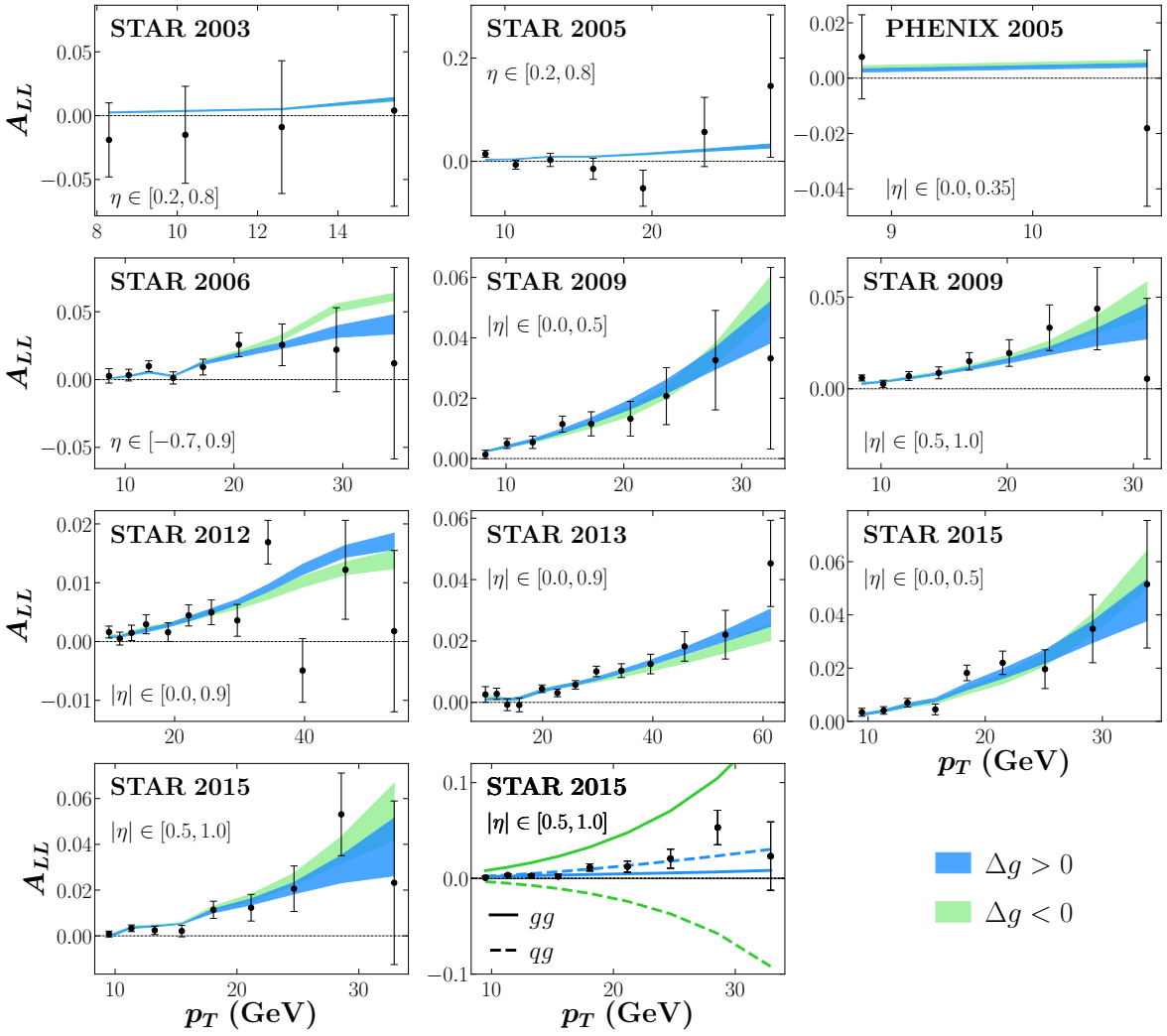

FIG. 5. Double longitudinal spin asymmetries A_{LL} in polarized pp collisions from STAR [17, 19–23] and PHENIX [24] versus jet transverse momentum p_T for bins in pseudorapidity η. The data are compared with the JAM global QCD analysis using the SU(3) scenario in Eq. (12b) for the "positive" gluon solutions $\Delta g > 0$ (blue bands) and "negative" gluon solutions $\Delta g < 0$ (green bands) with 1σ uncertainties. The final panel (with the same data as for the penultimate STAR 2015 panel) shows the contributions from the gg (solid lines) and qg channels (dashed lines).

The fits to the jet A_{LL} asymmetries are similar for the other scenarios, with $\chi^2_{\rm red}$ values almost identical, as listed in Table II. As may be expected, the less restrictive SU(2) scenario produces moderately wider uncertainty bands at the larger p_T values, $p_T \gtrsim 30$ GeV, due to the relatively larger uncertainties on the helicity PDFs in the absence of the SU(3) fla-

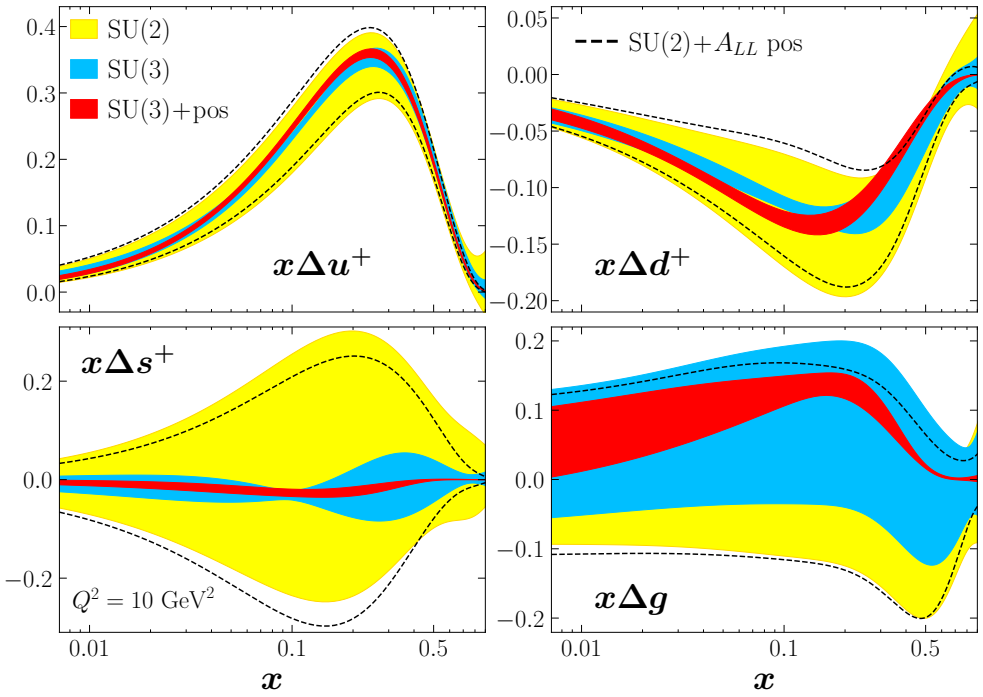

FIG. 6. Expectations values for spin-dependent Δu^+, Δd^+, Δs^+, and Δg PDFs at $Q^2 = 10$ GeV^2 fitted under various theory assumptions according to the SU(2) (yellow 1σ bands), SU(3) (blue 1σ bands) and SU(3)+positivity (red 1σ bands) scenarios, as well as with the SU(2) scenario but filtered to ensure A_{LL} positivity at large x (dashed lines).

vor symmetry constraint. Conversely, the more restrictive SU(3)+positivity scenario yields narrower error bands for $p_T \gtrsim 30$ GeV as a result of the significant suppression of the Δg solution space from the positivity constraints, as we discuss next.

To illustrate more explicitly the influence of theoretical assumptions on the PDFs and their uncertainties, we compare in Fig. 6 the Δu^+, Δd^+, Δs^+ and Δg distributions at $Q^2 = 10$ GeV^2 for the different scenarios with SU(2), SU(3), or SU(3)+positivity constraints. For the least constrained fit with only the SU(2) relation in Eq. (12a) imposed, the Δu^+ and Δd^+ PDFs are reasonably well determined, while the Δs^+ and Δg distributions have very large uncertainties and are consistent with zero. The imposition of the SU(3) relation in Eq. (12b) has a dramatic effect on the polarized quark PDF uncertainties, especially for the Δs^+ distribution, but also on the nonstrange spin PDFs which have reduced uncertainties. Interestingly, the uncertainty on the gluon polarization Δg is not affected significantly and remains large even with the inclusion of the polarized pp jet data and SU(3) constraint.

Imposition of the positivity constraints in Eq. (14) further reduces the uncertainties on the polarized quark PDFs, especially for the strange quark, and augments somewhat the shape of the Δd^+ PDF in particular. The latter effect is induced by assuming a flavor symmetric polarized sea, $\Delta\bar{u} = \Delta\bar{d} = \Delta\bar{s} = \Delta s$, so that changes in Δs propagate to the Δu^+ and Δd^+ distributions. Since the absolute values of Δd^+ are smaller than those of Δu^+, the impact on the polarized d quark is greater. The dependence of the strange helicity distribution on theoretical assumptions, such as SU(3) symmetry and positivity, may be reduced with additional experimental data on semi-inclusive DIS and single inclusive e^+e^- annihilation data, which can provide independent combinations of the quark flavor PDFs [9, 28].

For the gluon helicity distribution, the addition of the positivity constraint has quite a profound effect, reducing the uncertainty significantly and yielding a generally positive shape which more closely resembles that obtained in some previous PDF analyses [26, 30]. Examining the solution space more closely, we observe that the Δg solutions are extremely non-Gaussian for the SU(2) and SU(3) scenarios, as Fig. 7 illustrates for the individual replicas. In particular, our analysis identifies two distinct types of solutions for these scenarios, one with Δg mostly positive and one with Δg mostly negative. Despite their dramatic differences in sign and shape, however, the two Δg solutions can describe the A_{LL} data equally well. This is illustrated in the STAR 2015 panel of Fig. 5, where for the $\Delta g > 0$ solution the asymmetry is given by a sum of (small) positive contributions from the gg and qg channels, while for the $\Delta g < 0$ solution the asymmetry results from a cancellation between large positive gg and large negative qg pieces. As is evident in Fig. 7, our analysis disfavors solutions with $\Delta g = 0$, as well as with small negative Δg values, which would generally produce very small A_{LL} asymmetries, in contradiction with the data in Fig. 5.

Including the positivity constraint reduces the uncertainty on the gluon helicity distribution at high values of x ($x \gtrsim 0.3$) and eliminates the negative Δg solutions observed in Fig. 7. This occurs because in order to satisfy PDF positivity, the shape of the negative Δg solution becomes distorted and consequently fails to describe the jet A_{LL} data, giving rather large $\chi^2_{\rm red}$ values [70].

In addition to the scenarios discussed above, we also note that some replicas give unphysical values for the polarized DIS asymmetry at kinematics $x \approx 0.8$ and momentum transfer $Q^2 > 50$ GeV^2 that are outside the currently measured region, but which could be probed at a future Electron-Ion Collider [71]. After removing these replicas, the result shown in

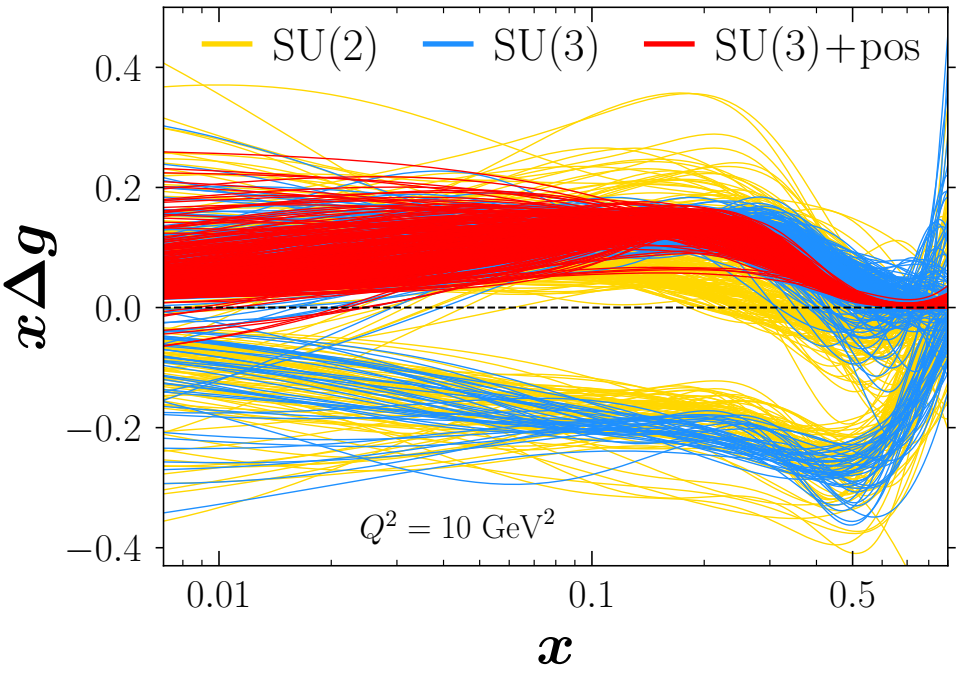

FIG. 7. Monte Carlo replicas for the spin-dependent gluon PDF $x\Delta g$ at $Q^2 = 10$ GeV^2 fitted under various theory assumptions according to the SU(2) (yellow lines), SU(3) (blue lines) and SU(3)+positivity (red lines) scenarios, with 300 replicas randomly selected from the total of 723, 647 and 639 for the three scenarios, respectively.

Fig. 6 for the SU(2) scenario indicates the main effect is observed at high x for the quark distributions, while the effect on Δg is negligible. Similarly for the other two scenarios, the impact of imposing the observable positivity on A_{LL} outside measured regions is only marginal.

To further explore the robustness of our findings for Δg, we also considered the scenario whereby the SU(2) constraint is imposed together with PDF positivity. In this case the resulting spin-dependent PDFs are found to be very similar to those from the SU(3) + positivity scenario, indicating that the order of imposing the SU(3) flavor symmetry and the positivity constraints does not change the observation of the two types of solutions for Δg in Fig. 7.

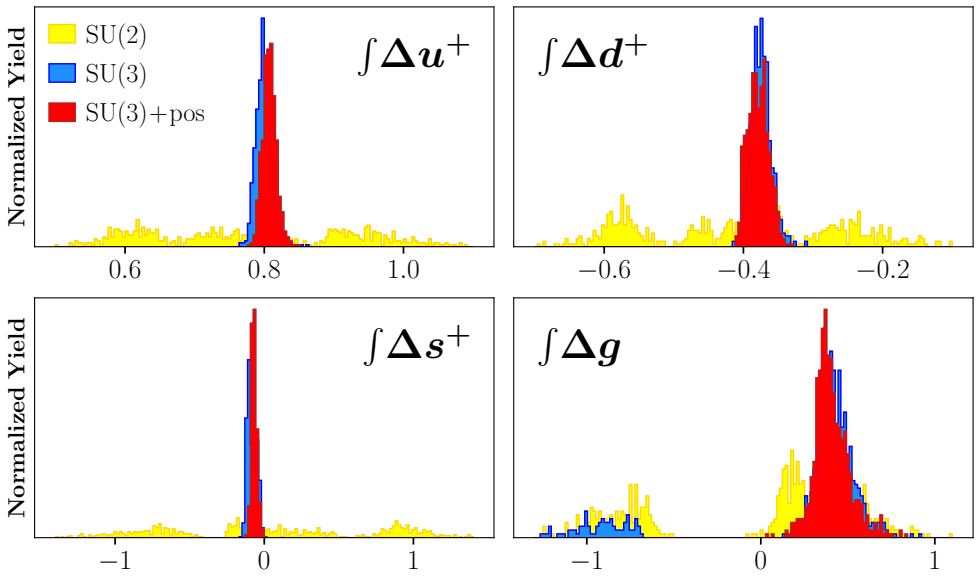

FIG. 8. Distribution of truncated moments of spin-dependent quark and gluon PDFs, integrated from $x_{\rm min} = 0.0071$ to 1, extracted from our global analysis under the different theoretical scenarios of SU(2) (yellow), SU(3) (blue) and SU(3)+positivity (red historgrams) at a scale $Q^2 = 10$ GeV^2.

C. Truncated moments

Along with visualizing the x dependence of the spin-dependent PDFs, a complementary way to assess the impact of the theoretical inputs on the quark and gluon helicities is to consider the truncated moments of the PDFs, defined as

$$\int \Delta q^+ \equiv \int_{x_{\rm min}}^1 {\rm d}x\, \Delta q^+, \qquad \int \Delta g \equiv \int_{x_{\rm min}}^1 {\rm d}x\, \Delta g, \tag{15}$$

where $x_{\rm min}$ is the lower limit of the integral. We choose the lower limit to be the smallest x value to which polarized data have sensitivity, $x_{\rm min} = 0.0071$ (see Fig. 1). In Fig. 8 we show the distribution of the quark and gluon truncated moments at a scale $Q^2 = 10$ GeV^2. The distributions of the quark truncated moments for the SU(2) scenario are rather broad (the SU(2) nonsinglet $\int \Delta u^+ - \int \Delta d^+$ is much more constrained though), and the gluon truncated moment displays the clear gap between the positive and negative solutions seen in Fig. 7. The reduction of the uncertainties with the imposition of SU(3) symmetry is quite striking for the quark moments, but does not qualitatively alter the distribution of gluon truncated moments, other than a slightly stronger peak in the positive Δg solution space.

With the positivity constraints imposed, on the other hand, the negative Δg solution is eliminated, with a prominent single peak around $\int \Delta g \approx 0.4$.

The central values and uncertainties of the truncated moments of all the quark and gluon flavors are shown in Table III for the three different scenarios. For the gluon in the SU(2) and SU(3) constraints scenarios we also include the individual contributions to the truncated moment from the positive and negative Δg solutions, along with the total. While the central values of the polarized quark moments do not vary much across the scenarios, the values of the gluon moments depend strongly on the theoretical assumptions. For the positive Δg solutions, the truncated moments remain at ~ 0.4, but together with the negative Δg solutions for the SU(2) and SU(3) scenarios, which yield ≈ -0.9, the combined moment ranges from ≈ 0 to 0.25, with large uncertainty ≈ 0.5.

Compared with the results from the DSSV14 analysis [26], which gave $\int \Delta g = 0.20(5)$ for $x_{\min} = 0.05$, our positive Δg solutions give values comparable to the DSSV14 result for all scenarios, ranging from 0.20(13) for the least restrictive SU(2) case to 0.25(3) for the most restrictive SU(3)+positivity case. Combining positive and negative Δg solutions, however, the gluon truncated moments over this range are 0.0(4) and 0.1(3) for the SU(2) and SU(3) scenarios, in clear contrast to the DSSV14 result. This strong dependence on the theoretical assumptions used in the analysis suggests that additional data with greater sensitivity to the shape and sign of Δg are needed before definitive, experiment-driven conclusions about gluon polarization can be reached.

TABLE III. Truncated moments and uncertainties of the quark and gluon PDFs integrated from $x_{\min} = 0.0071$ to 1 at $Q^2 = 10$ GeV^2 for the SU(2), SU(3), and SU(3)+positivity scenarios. For the SU(2) and SU(3) cases, the positive and negative $\int \Delta g$ contributions are also shown.

$\int \Delta f$	**SU(2)**		**SU(3)**		**SU(3)+pos**
Δu^+	0.8(1)		0.80(1)		0.81(1)
Δd^+	−0.4(1)		−0.37(1)		−0.38(2)
Δs^+	0.1(7)		−0.08(3)		−0.07(2)
Δg	0.0(6)		0.3(5)		0.39(9)
	$\Delta g > 0$	$\Delta g < 0$	$\Delta g > 0$	$\Delta g < 0$	
	0.4(2)	−0.8(2)	0.4(1)	−0.9(2)	

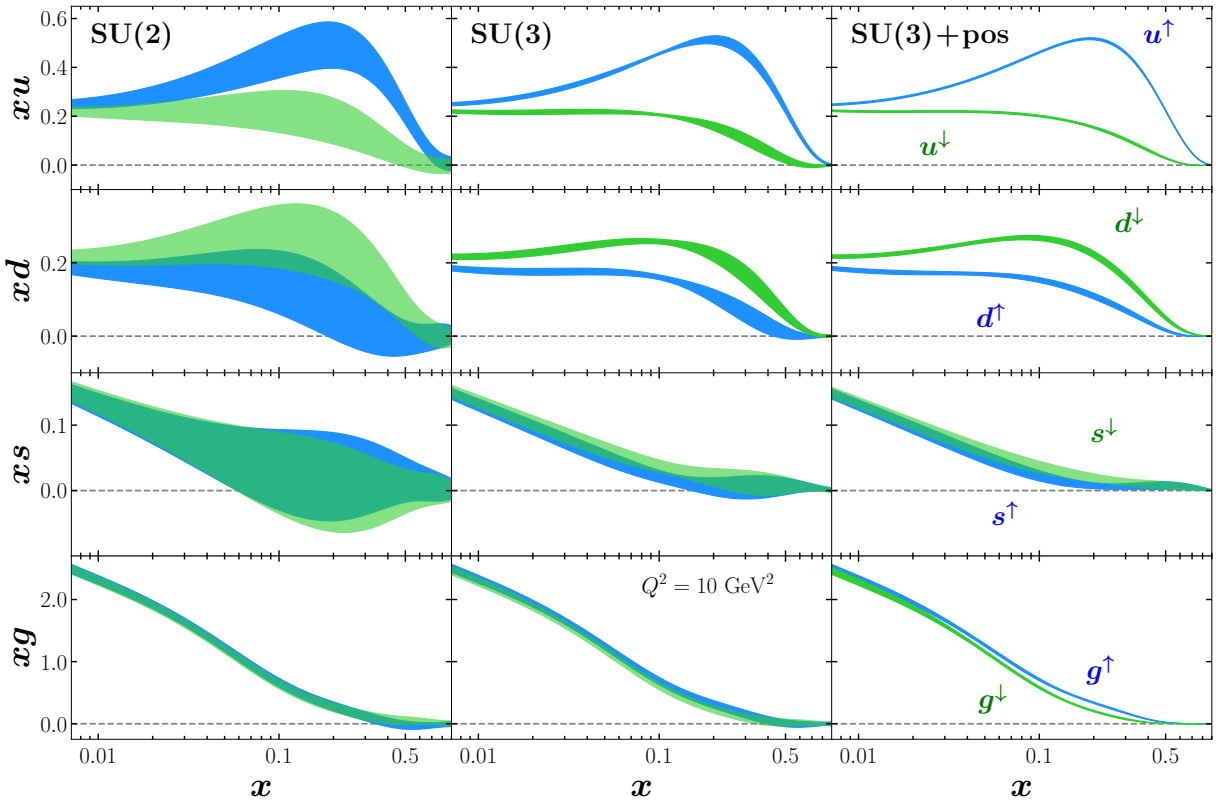

FIG. 9. Helicity-basis PDFs for helicity-aligned $f^\uparrow$ (blue bands) and antialigned $f^\downarrow$ (green bands) distributions for $f = u$, d, s and g at $Q^2 = 10$ GeV^2, in the SU(2), SU(3), and SU(3)+positivity scenarios.

D. Helicity-basis PDFs

Having obtained good agreement with both the unpolarized and polarized world data sets (see Table II), we can now analyze the simultaneously extracted spin-averaged and spin-dependent PDFs from the combined analysis. Using Eqs. (5), the distributions with spins parallel ($f^\uparrow$) and antiparallel ($f^\downarrow$) to the proton spin can be extracted for the first time with a consistent treatment of PDF uncertainties. In Fig. 9 we show the helicity-basis PDFs for all the u, d, s and g flavors at $Q^2 = 10$ GeV^2 in the SU(2), SU(3), and SU(3)+positivity scenarios.

For the SU(2) scenario, the $u^\uparrow$ and $u^\downarrow$ distributions, and to some degree the $d^\uparrow$ and $d^\downarrow$, are distinguishable, while the strange and gluon helicity-basis distributions for the most part cannot be distinguished. With the imposition of additional constraints, the uncertainties

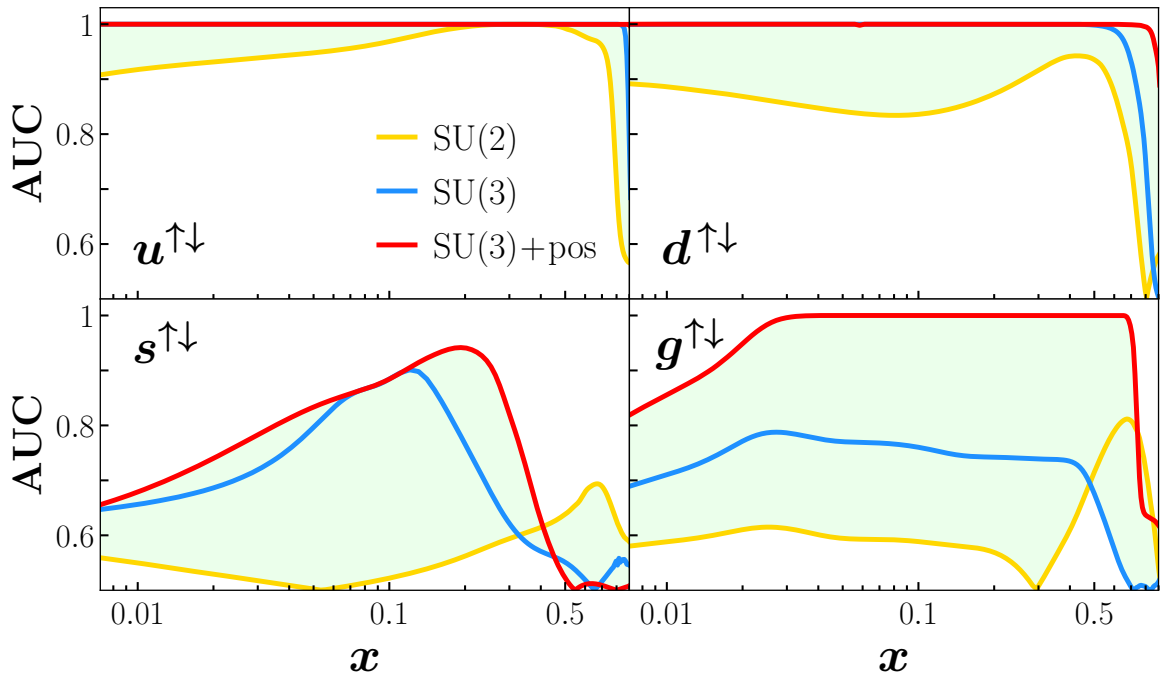

FIG. 10. The area under the curve (AUC) of the receiver operating characteristic (ROC) for the helicity-basis PDFs $f^{\uparrow}$ and $f^{\downarrow}$ for $f = u$, d, s, and g at $Q^2 = 10$ GeV^2, for the SU(2) (yellow lines), SU(3) (blue lines), and SU(3)+positivity (red lines) constraint scenarios. The discrimination between $f^{\uparrow}$ and $f^{\downarrow}$ improves as AUC $\to 1$. For clarity the region between the lines has been shaded (light green).

decrease and the helicity-basis PDFs $f^{\uparrow}$ and $f^{\downarrow}$ for each flavor become more clearly separated. Imposing the SU(3) constraint reduces the uncertainties of the helicity-basis PDFs for all the light quark flavors as a result of the extra constraint provided by a_8 in Eq. (12b). Moreover, the addition of positivity constraints restricts each helicity-basis PDF to remain positive, and suppresses uncertainties in the high-x region.

The degree to which the different helicity-basis PDFs in Fig. 9 can be delineated can be more accurately quantified by considering an "area under the curve" (AUC) plot. The AUC is defined as the area under a receiver operating characteristic (ROC) curve [77, 78] (see Appendix A for details), and is often used to visualize the discrimination power in binary classification problems. The closer the AUC approaches 1, the better the discrimination between the $f^{\uparrow}$ and $f^{\downarrow}$ distributions, and, conversely, the closer the AUC approaches 1/2, the more difficult it is to discriminate.

In Fig. 10 we show the AUC plot for the $f^{\uparrow}$ and $f^{\downarrow}$ helicity-basis PDFs for $f = u$, d, s, and g at $Q^2 = 10$ GeV^2, for the SU(2), SU(3), and SU(3)+positivity scenarios. Firstly, we note that the discrimination between the $u^{\uparrow}$ and $u^{\downarrow}$ PDFs is not affected significantly by the theory inputs, which reflects that both the spin-averaged and spin-dependent u-quark distributions are already well constrained empirically. The discrimination between $d^{\uparrow}$ and $d^{\downarrow}$, on the other hand, receives discernible improvement from the SU(3) flavor symmetry constraint, which produces a reduction of the uncertainty in the polarized Δd PDF (see Fig. 6). For the s-quark helicity-basis PDFs, imposing SU(3) symmetry makes a large improvement to their discrimination in the data-sensitive region $0.01 \lesssim x \lesssim 0.5$, driven by the significant reduction of the uncertainty for the Δs^+ PDF.

The positivity constraints make almost no improvement to the discrimination between the $u^{\uparrow}$ and $u^{\downarrow}$ or $d^{\uparrow}$ and $d^{\downarrow}$ distributions, given that the helicity-basis PDFs of both flavors are already well separated, and the positivity constraints only reduce their uncertainties. For the s-quark helicity PDFs, on the other hand, one may have expected that, given the significant reduction of the Δs^+ uncertainty and the less well discriminated $s^{\uparrow}$ and $s^{\downarrow}$ PDFs in the SU(3) scenario, the positivity constraints should discriminate between $s^{\uparrow}$ and $s^{\downarrow}$ more effectively. However, because the spin-averaged strange distributions are less well constrained at high x, the reduction of the Δs^+ uncertainty does not result in better discrimination between $s^{\uparrow}$ and $s^{\downarrow}$.

While the effects of the theory assumptions on the gluon distributions have been discussed extensively in the previous sections, we can obtain further information on their impact through the AUC representation in Fig. 10. Starting with the SU(2) scenario, without additional constraints the $g^{\uparrow}$ and $g^{\downarrow}$ distributions are barely distinguishable. After imposing the SU(3) constraint the reduction of the Δg uncertainties (see Fig. 7) discernibly improves the discrimination power. Finally, with the positivity constraints added, the negative Δg solution is eliminated, and the individual helicity-basis PDFs can be clearly distinguished in the experimentally measured region, $0.01 \lesssim x \lesssim 0.5$. This further illustrates our main conclusion, which is that theoretical inputs, especially SU(3) flavor symmetry and PDF positivity, can introduce significant bias into the extraction of PDFs that are not well constrained by experimental data, such as the Δs and Δg spin distributions.

V. CONCLUSION

We have performed the first simultaneous global QCD analysis of spin-averaged and spin-dependent PDFs in the nucleon within the JAM multi-step Monte Carlo framework, focusing in particular on the extraction of the gluon polarization. Good fits to the world unpolarized and polarized DIS data, as well as Drell-Yan data and inclusive jet production cross sections and asymmetries in hadronic collisions, were achieved, with a global reduced $\chi^2_{\rm red} \approx 1.1$. Our study was the first time that unpolarized RHIC pp jet cross sections were included together with the polarized pp jet data, along with the previous jet measurements in $p\bar{p}$ collisions at the Tevatron. While the direct impact of the unpolarized data on the spin-dependent PDFs is not significant, a simultaneous description of both observables is needed in order to delineate the kinematic domain of applicability of the collinear factorization framework.

Our study critically assessed the impact of theoretical assumptions on the determination of the gluon polarization Δg, including scenarios in which SU(2) or SU(3) flavor symmetry is assumed for axial charges determined from neutron and hyperon beta-decays, as well as the imposition of positivity on PDFs, which has been debated recently in the literature. The least biased scenario involving only the SU(2) constraint produced relatively large PDF uncertainties, especially for the polarized strange and gluon distributions. In particular, we found two distinct types of solutions for the gluon polarization, of opposite sign, each giving almost identical descriptions of experimental spin-dependent observables, including the double polarization asymmetry A_{LL} in inclusive jet production.

With the SU(3) flavor symmetry constraint, the uncertainties on Δs were reduced significantly, although the Δg distribution was largely unaffected. It was only with the further addition of PDF positivity that the negative Δg solution could be eliminated and results resembling those found in earlier literature recovered. We identify this as a bias introduced in the extraction of spin-dependent PDFs, and in the absence of a clear theoretical requirement for PDFs to be positive at all values of x, we conclude that data-driven analysis alone does not constrain the gluon polarization to be uniquely positive.

Such a conclusion naturally has profound implications for our understanding of the proton spin decomposition, and provides greater urgency to identifying other means to constrain the gluon helicity, as well as the quark and gluon orbital angular momentum. Further data are needed to obtain clarity on this issue, including new observables that are linearly sensitive

to gluon helicity distributions. An example of this may be polarized semi-inclusive DIS with the production of large transverse momentum hadrons, in which the gluon polarization enters at the same order as the polarized quark contribution [79].

A novel new feature of our simultaneous analysis of polarized and unpolarized observables is the ability to systematically extract the individual helicity-aligned and antialigned PDFs with a consistent treatment of uncertainties. Our demonstration of the current empirical situation using the AUC plot illustrates the impact of various theoretical assumptions on our ability to discriminate between different helicity-basis PDFs. It confirms that, while imposing SU(3) flavor symmetry improves the discrimination between the $s^{\uparrow}$ and $s^{\downarrow}$ PDFs, the ability to separate clearly the $g^{\uparrow}$ and $g^{\downarrow}$ distributions requires additional assumptions about PDF positivity. We anticipate that future high-precision data from existing and planned facilities, including the Electron-Ion Collider [71], will elucidate the question of the gluon polarization and the proton spin decomposition more definitely.

ACKNOWLEDGEMENTS

We thank C. Andres, P. Barry, C. Cocuzza, J. Owens, T. Rogers and F. Ringer for helpful discussions, and W. Vogelsang for providing the code for the calculation of the jet observables. This work was supported by the US Department of Energy (DOE) contract No. DE-AC05-06OR23177, under which Jefferson Science Associates, LLC operates Jefferson Lab. The work of N.S. was supported by the DOE, Office of Science, Office of Nuclear Physics in the Early Career Program.

Appendix A: Area Under the ROC Curve

The receiver operating characteristic (ROC) and the corresponding area under the ROC curve (AUC) are often used in visualizing the discrimination in binary classification problems [77, 78]. In this appendix we will use a simple example to illustrate the application of ROC and AUC plots for PDF discrimination.

Defining two normalized Gaussian distributions $\mathcal{N}_1$ and $\mathcal{N}_2$ in a common variable ω, with central values μ_1 and μ_2 and widths $\sigma_1 = \sigma_2$, in Fig. 11 we examine how well these can be discriminated from each other for varying $\mu_1 - \mu_2$. We consider the cumulative integration values of $\mathcal{N}_2$ versus those of $\mathcal{N}_1$, defined as

$$\int_{-\infty}^{\overline{\omega}} \mathrm{d}\omega\, \mathcal{N}_i(\omega) = \int_{-\infty}^{\overline{\omega}} \frac{\mathrm{d}\omega}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(\omega-\mu_i)^2}{2\sigma_i^2}\right) \quad \text{for} \;\; \overline{\omega} \in \big(-\infty, \infty\big), \tag{A1}$$

for $i = 1, 2$. When the two distributions $\mathcal{N}_1$ and $\mathcal{N}_2$ overlap entirely with each other, $\mu_1 = \mu_2$, their cumulative integration values are of course identical, and these produce a diagonal line in the ROC plot of $\int_{-\infty}^{\overline{\omega}} \mathrm{d}\omega\, \mathcal{N}_2$ versus $\int_{-\infty}^{\overline{\omega}} \mathrm{d}\omega\, \mathcal{N}_1$ in Fig. 11 (top row). In this case the AUC value is exactly 1/2.

When the mean values of $\mathcal{N}_1$ and $\mathcal{N}_2$ differ, but not dramatically, $\mu_1 \lesssim \mu_2$ (middle row), the distributions start to deviate from the each other, and the ROC bends away from the diagonal since $\int_{-\infty}^{\overline{\omega}} \mathrm{d}\omega\, \mathcal{N}_2$ reaches its maximum value slower than $\int_{-\infty}^{\overline{\omega}} \mathrm{d}\omega\, \mathcal{N}_1$. The AUC value here increases away from the minimum of 1/2.

Finally, when the mean values are clearly separated from each other, $\mu_1 \ll \mu_2$ (bottom row of Fig. 11), the ROC deviates significantly from the diagonal and the AUC value approaches 1. Of course, if one had $\mu_1 \gtrsim \mu_2$ or $\mu_1 \gg \mu_2$, the ROC would be curved downwards and the AUC would deviate from 1/2 while approaching 0; in this case the $1 - \text{AUC}$ would usually be used in order to keep the figures intuitive.

For the binary classification problem in Sec. IV D, the goal is to evaluate the degree to which the helicity-basis PDFs $f^{\uparrow}$ and $f^{\downarrow}$ can be discriminated from each other at every point in the parton momentum fraction x. The AUC plot in Fig. 10 then represents the quality of discrimination as a function of x for the different theoretical scenarios assuming SU(2) symmetry, SU(3) symmetry, or SU(3) + PDF positivity.

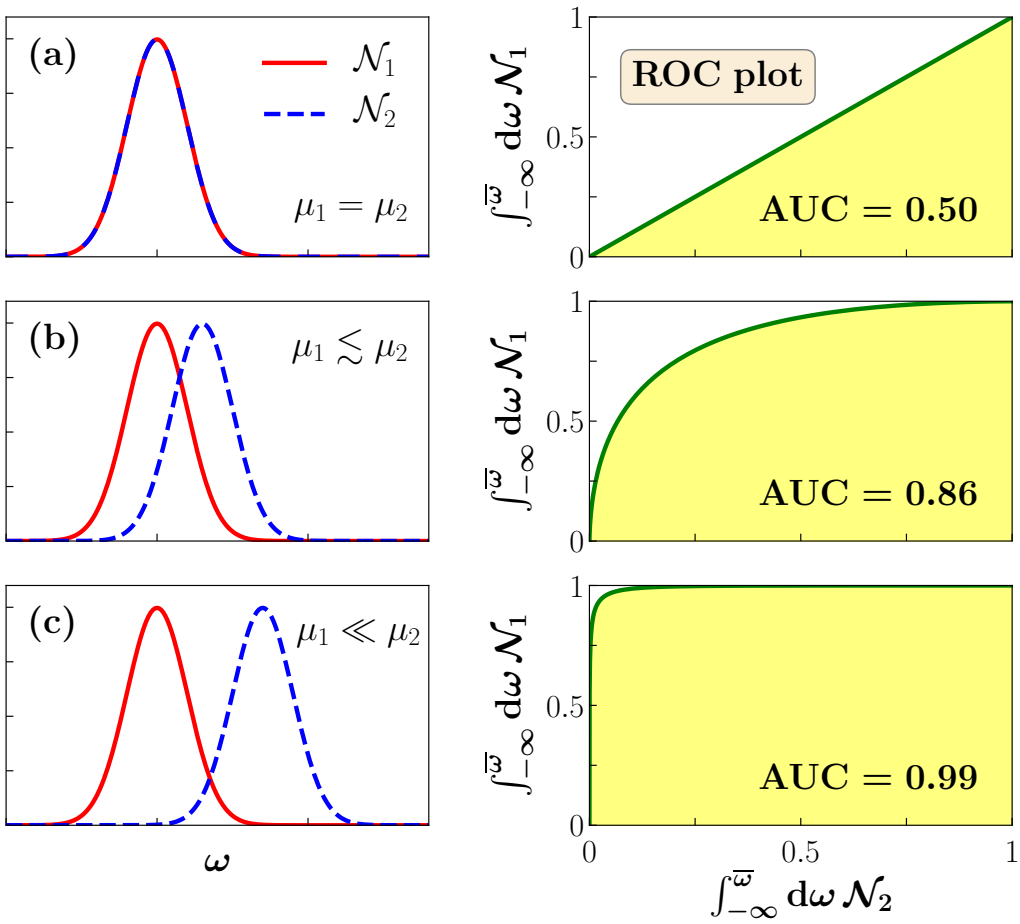

FIG. 11. Illustration of ROC and AUC plots for Gaussian distributions $\mathcal{N}_1$ and $\mathcal{N}_2$ (red solid and blue dashed lines) to be discriminated from each other (left panels), and the corresponding ROC curves (green lines) and AUC values (right panels) for different mean values μ_1 and μ_2: **(a)** $\mu_1 = \mu_2$, **(b)** $\mu_1 \lesssim \mu_2$, **(c)** $\mu_1 \ll \mu_2$.

[1] J. Ashman *et al.*, Phys. Lett. B **206**, 364 (1988).

[2] G. Altarelli and G. G. Ross, Phys. Lett. B **212**, 391 (1988).

[3] R. D. Carlitz, J. C. Collins and A. H. Mueller, Phys. Lett. B **214**, 229 (1988).

[4] J. R. Ellis and R. L. Jaffe, Phys. Rev. D **9**, 1444 (1974) [Erratum: Phys. Rev. D **10**, 1669 (1974).

[5] F. E. Close and R. G. Roberts, Phys. Lett. B **316**, 165 (1993).

[6] M. Anselmino, A. Efremov and E. Leader, Phys. Rep. **261**, 1 (1995). [Erratum: Phys. Rep. **281**, 399 (1997)].

[7] B. Lampe and E. Reya, Phys. Rep. **332**, 1 (2000).

[8] C. A. Aidala, S. D. Bass, D. Hasch and G. K. Mallot, Rev. Mod. Phys. **85**, 655 (2013).

[9] J. J. Ethier, N. Sato and W. Melnitchouk, Phys. Rev. Lett. **119**, 132001 (2017).

[10] M. Diehl, Phys. Rep. **388**, 41 (2003).

[11] E. Leader and C. Lorcé, Phys. Rep. **541**, 163 (2014).

[12] D. Müller, D. Robaschik, B. Geyer, F. M. Dittes and J. Hořejši, Fortsch. Phys. **42**, 101 (1994).

[13] X. Ji, Phys. Rev. D **55**, 7114 (1997).

[14] A. V. Radyushkin, Phys. Lett. B **380**, 417 (1996).

[15] C. Alexandrou *et al.*, Phys. Rev. Lett. **119**, 142002 (2017).

[16] K. Kumericki, S. Liuti and H. Moutarde, Eur. Phys. J. A **52**, 157 (2016).

[17] B. I. Abelev *et al.*, Phys. Rev. Lett. **97**, 252001 (2006).

[18] B. I. Abelev *et al.*, Phys. Rev. Lett. **100**, 232003 (2008).

[19] L. Adamczyk *et al.*, Phys. Rev. D **86**, 032006 (2012).

[20] L. Adamczyk *et al.*, Phys. Rev. Lett. **115**, 092002 (2015).

[21] J. Adam *et al.*, Phys. Rev. D **100**, 052005 (2019).

[22] M. Abdallah *et al.*, Phys. Rev. D **103**, L091103 (2021).

[23] M. S. Abdallah *et al.*, arXiv:2110.11020 [hep-ex].

[24] A. Adare *et al.*, Phys. Rev. D **84**, 012006 (2011).

[25] B. Jäger *et al.*, Phys. Rev. D **70**, 034010 (2004).

[26] D. de Florian, R. Sassot, M. Stratmann and W. Vogelsang, Phys. Rev. Lett. **113**, 012001 (2014).

[27] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk, Phys. Rev. D **101**, 074020 (2020).

[28] E. Moffat, W. Melnitchouk, T. C. Rogers and N. Sato, Phys. Rev. D **104**, 016015 (2021).

[29] C. Cocuzza, W. Melnitchouk, A. Metz and N. Sato, in preparation (2021).

[30] E. R. Nocera *et al.*, Nucl. Phys. **B887**, 276 (2014).

[31] A. Candido, S. Forte and F. Hekhorn, J. High Energy Phys. 11 (2020) 129.

[32] J. C. Collins, T. C. Rogers and N. Sato, arXiv:2111.01170 [hep-ph].

[33] C. Adolph *et al.*, Phys. Lett. B **753**, 18 (2016).

[34] Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).

[35] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).

[36] G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).

[37] Z. B. Kang, F. Ringer and W. J. Waalewijn, J. High Energy Phys. 07 (2017) 064.

[38] W. Vogelsang, private communication (2021).

[39] G. C. Blazey *et al.*, arXiv:hep-ex/0005012.

[40] S. D. Ellis and D. E. Soper, Phys. Rev. D **48**, 3160 (1993).

[41] M. Cacciari, G. P. Salam and G. Soyez, J. High Energy Phys. 04 (2008) 063.

[42] A. Vogt, Comput. Phys. Commun. **170**, 65 (2005).

[43] E. G. Floratos, C. Kounnas and R. Lacaze, Nucl. Phys. **B192**, 417 (1981).

[44] M. Stratmann and W. Vogelsang, Phys. Rev. D **64**, 114007 (2001).

[45] C. Cocuzza, W. Melnitchouk, A. Metz and N. Sato, Phys. Rev. D **104**, 074031 (2021).

[46] C. Cocuzza, C. E. Keppel, H. Liu, W. Melnitchouk, A. Metz, N. Sato and A.W. Thomas, Phys. Rev. Lett. **127**, 242001 (2021).

[47] A. C. Benvenuti *et al.*, Phys. Lett. B **223**, 485 (1989).

[48] L. W. Whitlow *et al.*, Phys. Lett. B **282**, 475 (1992).

[49] M. Arneodo *et al.*, Nucl. Phys. **B483**, 3 (1997).

[50] M. Arneodo *et al.*, Nucl. Phys. **B487**, 3 (1997)

[51] H. Abramowicz *et al.*, Eur. Phys. J. C **75**, 580 (2015).

[52] J. Ashman *et al.*, Nucl. Phys. **B328**, 1 (1989).

[53] B. Adeva *et al.*, Phys. Rev. D **58**, 112001 (1998).

[54] B. Adeva *et al.*, Phys. Rev. D **60**, 072004 (1999) [Erratum: Phys. Rev. D **62**, 079902 (2000)].

[55] M. G. Alekseev *et al.*, Phys. Lett. B **690**, 466 (2010).

[56] V. Y. Alexakhin *et al.*, Phys. Lett. B **647**, 8 (2007).

[57] G. Baum *et al.* Phys. Rev. Lett. **51**, 1135 (1983).

[58] P. L. Anthony *et al.*, Phys. Rev. D **54**, 6620 (1996).

[59] K. Abe *et al.*, Phys. Rev. D **58**, 112003 (1998).

[60] K. Abe *et al.*, Phys. Rev. Lett. **79**, 26 (1997).

[61] P. L. Anthony *et al.*, Phys. Lett. B **493**, 19 (2000).

[62] P. L. Anthony *et al.*, Phys. Lett. B **463**, 339 (1999).

[63] K. Ackerstaff *et al.*, Phys. Lett. B **404**, 383 (1997).

[64] A. Airapetian *et al.*, Phys. Rev. D **75**, 012007 (2007).

[65] N. Sato, W. Melnitchouk, S. E. Kuhn, J. J. Ethier, and A. Accardi, Phys. Rev. D **93**, 074005 (2016).

[66] E. A. Hawker *et al.*, Phys. Rev. Lett. **80**, 3715 (1998).

[67] S. Alekhin, K. Melnikov and F. Petriello, Phys. Rev. D **74**, 054033 (2006).

[68] V. M. Abazov *et al.*, Phys. Rev. Lett. **101**, 062001 (2008).

[69] A. Abulencia *et al.*, Phys. Rev. D **75**, 092006 (2007) [Erratum: Phys. Rev. D **75**, 119901 (2007)].

[70] Y. Zhou, Dissertations, Theses, and Masters Projects, William & Mary, Paper 1627407586.

[71] R. Abdul Khalek *et al.*, arXiv:2103.05419 [physics.ins-det].

[72] P. Jimenez-Delgado, A. Accardi and W. Melnitchouk, Phys. Rev. D **89**, 034025 (2014).

[73] N. Sato, J. J. Ethier, W. Melnitchouk, M. Hirai, S. Kumano, and A. Accardi, Phys. Rev. D **94**, 114004 (2016).

[74] R. D. Ball *et al.*, Eur. Phys. J. C **77**, 663 (2017).

[75] A. M. Cooper-Sarkar and K. Wichmann, Phys. Rev. D **98**, 014027 (2018).

[76] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, Eur. Phys. J. C **75**, 204 (2015).

[77] J. P. Egan, *Signal detection theory and ROC analysis*, Series in Cognition and Perception (Academic Press, New York, 1975).

[78] T. Fawcett, Pattern Recogn. Lett. **27**, 861 (2006).

[79] R. M. Whitehill, Y. Zhou, N. Sato and W. Melnitchouk, in preparation (2022).