

Strangeness in the proton from W + charm production and SIDIS dataTrey Anderson^{1,2}, W. Melnitchouk², and N. Sato²

JAM Collaboration (PDF Analysis Group)

¹*Department of Physics, William & Mary, Williamsburg, Virginia 23185, USA*²*Jefferson Lab, Newport News, Virginia 23606, USA* (Received 11 February 2025; accepted 14 October 2025; published 7 November 2025)

We perform a global QCD analysis of unpolarized parton distribution functions (PDFs) in the proton, including new W + charm production data from pp collisions at the LHC and semi-inclusive pion and kaon production data in lepton-nucleon deep-inelastic scattering, both of which have been suggested for constraining the strange quark PDF. Compared with a baseline global fit that does not include these datasets, the new analysis reduces the uncertainty on the strange quark distribution over the range $0.01 < x < 0.3$, and provides a consistent description of processes sensitive to strangeness in the proton. Including the new datasets, the ratio of strange to nonstrange sea quark distributions is $R_s = (s + \bar{s})/(\bar{u} + \bar{d}) = \{0.72_{-0.34}^{+0.52}, 0.46_{-0.20}^{+0.30}, 0.32_{-0.15}^{+0.23}\}$ for $x = \{0.01, 0.04, 0.1\}$ at $Q^2 = 4 \text{ GeV}^2$. The data place more stringent constraints on the strange asymmetry $s - \bar{s}$, which is found to be consistent with zero in this range.

DOI: [10.1103/7h25-3lg7](https://doi.org/10.1103/7h25-3lg7)**I. INTRODUCTION**

The simple picture of matter that has been built up over the past few decades of probing the femtometer scale structure of the proton attributes the bulk of its properties, such as baryon number, charge, or magnetic moment, to the irreducible core of its valence u and d quarks. In momentum space the distributions of these quarks have been mapped out in considerable detail from deep-inelastic scattering (DIS) and other high-energy scattering observables, over a large range of light-cone momentum fractions x . On top of this structure, we now understand, lies a teeming sea of virtual quark-antiquark pairs. These pairs do not alter the global quantum numbers of the proton, but can affect some of its properties, such as the spin and magnetic moment, and yield nontrivial structures, such as an excess of \bar{d} antiquarks over \bar{u} [1–4] in the proton, and an analogous excess of \bar{u} in the neutron.

While expected to be suppressed because of its larger mass, the creation of strange-antistrange quark pairs is also expected to play a role in the femtoscopic structure of the proton. In ordinary matter, the creation of virtual $u\bar{u}$ and $d\bar{d}$ pairs is screened by the presence of the valence u and d

quarks. Strange quarks, on the other hand, can directly reveal the properties of $q\bar{q}$ pairs in the proton, although the practical realization of this has been severely hampered by the difficulty of obtaining reliable empirical information on the s and \bar{s} distributions in the proton. Although it is anticipated on general grounds that an asymmetry between s and \bar{s} quarks would be a unique window into the nonperturbative dynamics of quarks in the proton [5–7] (for example, in relation to the spontaneous breaking of chiral SU(3) symmetry), in practice even the magnitude of the sum $s + \bar{s}$ is poorly known.

Historically, the shape of the strange quark parton distribution function (PDF) was first studied in neutrino-induced DIS from nuclear targets, particularly in the semi-inclusive production of charmed mesons in charged current reactions from the CCFR [8] and NuTeV [9] Collaborations at the Tevatron, as well as from the CHORUS [10] and NOMAD [11] experiments at CERN. An important complication in this process is the modeling of nuclear corrections in neutrino-nucleus DIS, which are currently poorly understood [12], and in the treatment of charm quark energy loss and charmed D meson-nucleon rescattering inside the nucleus [13]. An alternative avenue that avoids nuclear corrections is semi-inclusive DIS production of kaons, which depends on the strange quark PDFs and strange quark to kaon fragmentation functions (FFs) as an additional nonperturbative input. Attempts have been made to extract the $s + \bar{s}$ distribution from kaon SIDIS data by the HERMES Collaboration [14,15] (see also Refs. [16–18]).

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

More recently, inclusive W and Z boson production has been found to have sensitivity to the strange quark PDF, although mixed results have been obtained from measurements at ATLAS and CMS at the LHC [19–24]. Inclusive $W + \text{charm}$ production in principle could have even greater sensitivity to the s and \bar{s} PDFs [25–27], in analogy with the neutrino DIS measurements, but still without the complications of nuclear targets.

Aside from the intrinsic value of understanding the structure of the proton sea, the precise determination of strange quark PDFs is key for extracting Standard Model parameters, such as the Cabibbo-Kobayashi-Maskawa matrix element V_{cs} and the weak mixing angle, $\sin^2 \theta_w$, as well as precision measurements on the mass of the W -boson that depend on precise knowledge of the strange quark PDF [28,29]. In this paper, we revisit the question of the strange and antistrange quark distributions in the proton with a new analysis that includes all of the above datasets involving proton beams and targets. For the first time we combine the inclusive and semi-inclusive DIS (SIDIS) structure functions and multiplicities, together with cross section ratios and charge asymmetries in W/Z and $W + \text{charm}$ production in pp collisions at the LHC, simultaneously fitting PDFs and FFs to nearly 6 000 data points.

We contrast the pulls of different datasets on the s and \bar{s} distributions, finding that both SIDIS and $W + \text{charm}$ data reduce uncertainties relative to the baseline that does not include these datasets, although each of the scenarios is consistent within uncertainties. Furthermore, we find that SIDIS data alone allow for slightly smaller strangeness at $x \lesssim 0.1$, while the $W + \text{charm}$ data alone generally allow for enhanced strangeness for all $x \lesssim 0.3$. On the other hand, combining both the SIDIS and $W + \text{charm}$ data reduces the overall uncertainty compared to the results with the individual datasets and gives a suppressed strange to nonstrange quark PDF ratio $R_s = (s + \bar{s})/(\bar{d} + \bar{u})$ for $0.02 \lesssim x \lesssim 0.2$ relative to some previous analyses of the ATLAS $W + \text{charm}$ data [26].

We note that our simultaneous reconstruction of PDFs and FFs from all the available datasets leverages the quark flavor sensitivity of the SIDIS data, uniformly propagating uncertainties across all the involved distributions. Some previous analyses, specifically of polarized PDFs [30,31], have included SIDIS data with fixed FFs via a reweighting procedure, which typically requires inflating the experimental errors to include the uncertainties from FFs. Because the FFs are fixed, this hinders simultaneously improving knowledge of the PDFs and FFs from data. Our analysis eliminates this limitation, allowing us to constrain all the nonperturbative functions needed to describe the observational data.

We begin in Sec. II with a review of the theoretical framework used in this analysis, focusing on the new

$W + \text{charm}$ and SIDIS observables, and discussing the nonperturbative modeling for the PDFs and FFs. Section III provides a concise summary of the datasets used in the fit, including DIS, Drell-Yan lepton-pair production in pp and pd scattering, W^\pm , Z , $W + \text{charm}$, and jet production in pp or $p\bar{p}$ reactions. The datasets also include pion, kaon, and charged hadron production in SIDIS, as well as single-inclusive e^+e^- annihilation (SIA). In Sec. IV, we discuss in detail our methodology for implementing the regression problem, including model calibration and the kinematical cuts applied to the data. We then present the results of the data-versus-theory comparisons and the extracted PDFs and FFs. Finally, conclusions are drawn in Sec. V, where we discuss future theoretical and experimental steps that could provide further insights into the strangeness content of the proton.

II. THEORETICAL FRAMEWORK

The theoretical basis for this analysis is collinear QCD factorization, to fixed order in the QCD coupling α_s , for various high-energy scattering processes which involve PDFs or FFs. These include DIS, Drell-Yan lepton-pair production, weak boson and jet production, and $W + \text{charm}$ production, which directly constrain proton PDFs; π , K and unidentified charged hadron production in SIA, which give information on FFs; and SIDIS, which depends on both PDFs and FFs. Since the focus of this study is specifically on the role of strangeness in the proton, we will discuss in greater detail the theoretical framework for processes most sensitive to the s and \bar{s} PDFs in the proton, namely, inclusive $W + \text{charm}$ production in pp collisions and lepton-deuteron SIDIS. We also describe the parametrizations employed for our PDFs and FFs.

A. Physical processes and factorization

The associated production of a W boson and a charm quark in pp collisions,

$$p + p \rightarrow W + c + X, \quad (1)$$

where X represents all other particles in the final state, is expected to be sensitive to the strange content of the proton. Here the charge state $W^- + c$ originates from the scattering off a strange quark, while the $W^+ + \bar{c}$ tags an antistrange quark. The c and \bar{c} quarks are identified within a jet by a muon produced from its semileptonic decay. Within the collinear factorization framework the differential cross section can be written as a convolution of the perturbatively calculated partonic cross section $\hat{\sigma}_{a,b}^{W+c}$ and the nonperturbative PDFs f_a and f_b of partons a and b in the colliding protons,

$$\frac{d\sigma^{W+c}}{d|\eta|} = \sum_{a,b} \int_{x_a} \int_{x_b} d\hat{x}_a d\hat{x}_b f_a(\hat{x}_a, \mu_F) f_b(\hat{x}_b, \mu_F) \times \hat{\sigma}_{a,b}^{W+c} \left(\frac{\hat{x}_a}{x_a}, \frac{\hat{x}_b}{x_b}, \mu_F, \mu_R \right), \quad (2)$$

where η is the pseudorapidity of the lepton from the W -boson decay. The PDFs are functions of the partonic momentum fraction variables $\hat{x}_{a,b}$, with the sum in Eq. (2) running over parton flavors a and b for the contributing partonic channels. At leading order in the strong coupling $\alpha_s(\mu_R)$, the variables $x_{a,b}$ are related to the pseudorapidity by $x_{a,b} = (M_W/\sqrt{s})e^{\pm\eta}$. The partonic cross sections $\hat{\sigma}_{a,b}^{W+c}$ is calculated at next-to-leading order (NLO) in $\alpha_s(\mu_R)$ using the Monte Carlo program MCFM [32], with the factorization scale μ_F and renormalization scale μ_R set to $\mu_F = \mu_R = M_W$.

For the SIDIS of a lepton ℓ from a nucleon N , producing charged pions π^\pm , kaons K^\pm or unidentified hadrons,

$$\ell + N \rightarrow \ell + h^\pm + X, \quad (3)$$

the differential cross section can be written as the double convolution of the partonic cross section $\hat{\sigma}_{a,b}^h$ with the PDF f_a and the parton b to hadron h FF D_b^h ,

$$\frac{d\sigma^h}{dx_B dz_h dQ^2} = \sum_{a,b} \int_{z_h} \int_{x_B} d\hat{x} d\hat{z} f_a(\hat{x}, \mu_F) D_b^h(\hat{z}, \mu_F) \times \hat{\sigma}_{a,b}^h \left(\frac{\hat{x}}{x_B}, \frac{\hat{z}}{z_h}; \mu_F, \mu_R, Q \right). \quad (4)$$

Here $Q^2 \equiv -q^2$ is the squared four-momentum transfer to the nucleon, $x_B = Q^2/2p \cdot q$ is the Bjorken scaling variable, with p and q the target nucleon and virtual photon four-momenta, respectively, and $z_h = p \cdot p_h / p \cdot q$ is the fraction of the virtual photon's energy carried by the fragmenting hadron h with four-momentum p_h . The invariant mass squared of the unmeasured hadronic final state is given by $W_{\text{SIDIS}}^2 = (p + q - p_h)^2$. The partonic cross section $\hat{\sigma}_{a,b}^h$ is evaluated perturbatively to NLO accuracy, and the factorization and renormalization scales are set as $\mu_F = \mu_R = Q$.

For all processes in our analysis we use the $\overline{\text{MS}}$ scheme for the renormalization group equations, with the strong coupling α_s evolved numerically using the QCD β -functions with the boundary condition $\alpha_s(M_Z) = 0.118$ at the Z -boson mass, $M_Z = 91.18$ GeV. The PDFs and FFs are evolved to next-to-leading logarithmic accuracy using the DGLAP evolution equations [33–35] in the zero-mass variable flavor scheme, setting the input scale to the charm quark mass, $\mu = m_c$, for both PDFs and FFs. The heavy quark mass thresholds are taken from PDG to be $m_c = 1.28$ GeV and $m_b = 4.18$ GeV [36].

B. Nonperturbative modeling

We parametrize all PDFs at the input scale $\mu^2 = m_c^2$ using the standard, phenomenologically successful template function

$$f(x, \mu; \mathbf{a}) = \frac{N}{\mathcal{M}} x^\alpha (1-x)^\beta (1 + \gamma\sqrt{x} + \delta x), \quad (5)$$

where the set of parameters to be fitted, $\mathbf{a} = \{N, \alpha, \beta, \gamma, \delta\}$, includes the normalization coefficient N and shape parameters α, β, γ and δ . To ensure that the normalization coefficient N is maximally decorrelated from the shape parameters, we normalize the function using $\mathcal{M} = B[\alpha + 2, \beta + 1] + \gamma B[\alpha + \frac{5}{2}, \beta + 1] + \delta B[\alpha + 3, \beta + 1]$, where B is the beta function (Euler integral of the first kind).

As in most global QCD analyses, we assume isospin symmetry for the PDFs and FFs, so that a u -quark PDF in the proton is equivalent to a d -quark PDF in the neutron, for instance. For the PDFs, we parametrize the valence u and d quark distributions directly,

$$u_v \equiv u - \bar{u}, \quad d_v \equiv d - \bar{d}, \quad (6)$$

along with the gluon distribution, g , via Eq. (5). For the sea quark and antiquark distributions we use the *ansatz*,

$$\begin{aligned} \bar{u} &= S_1 + \delta\bar{u}, & \bar{d} &= S_1 + \delta\bar{d}, \\ s &= S_2 + \delta s, & \bar{s} &= S_2 + \delta\bar{s}, \end{aligned} \quad (7)$$

where S_1 and S_2 contain the bulk of the flavor-independent sea distribution for the light sea and strange sectors, respectively, and δf ($f = \bar{u}, \bar{d}, s, \bar{s}$) are additional flavor-dependent distortions away from $S_{1,2}$ that are required to describe the observational data. Our modeling allows us to explore a range of possibilities for the sea distributions, including the symmetric sea scenario ($S_1 = S_2$ and $\delta f = 0$), light sea quark asymmetry, light and strange sea quark asymmetry, as well as the $s - \bar{s}$ asymmetry. For the parametrizations of S_1, S_2 , and δf we use the same template function as in Eq. (5).

As usual, the normalization parameters for the u_v, d_v , and δs distributions are set using valence number sum rules,

$$\int_0^1 dx u_v = 2, \quad \int_0^1 dx d_v = 1, \quad \int_0^1 dx (s - \bar{s}) = 0, \quad (8)$$

while the normalization parameter for the gluon PDF is set by the momentum sum rule,

$$\int_0^1 dx x \left(\sum_q (q + \bar{q}) + g \right) = 1. \quad (9)$$

The parametrization template in Eq. (5) is also used for FFs, but with x replaced by z , the momentum fraction of the parton carried by the produced hadron. The FF model used in this analysis allows for maximum parametrization flexibility, while ensuring that the number of independent functions does not exceed that which can be reliably constrained by the available SIA and SIDIS observables (see Ref. [37] for further details). For each hadron species, we have 4 independent functions from SIA and 2 independent functions from SIDIS. We allow the g , c , and b FFs to be independent, leaving 3 functions to parametrize the light quarks. Assuming charge symmetry, for the $q \rightarrow \pi^+$ FFs we take

$$\begin{aligned} D_u^{\pi^+} &= D_d^{\pi^+}, \\ D_d^{\pi^+} &= D_u^{\pi^+}, \\ D_q^{\pi^+} &= D_{\bar{q}}^{\pi^+}, \quad q = s, c, b. \end{aligned} \quad (10)$$

For the $q \rightarrow K^+$ FFs we allow the favored $D_u^{K^+}$ and $D_s^{K^+}$ FFs to be independent, but set the FFs for the unfavored flavors to be equal,

$$\begin{aligned} D_d^{K^+} &= D_u^{K^+} = D_{\bar{d}}^{K^+} = D_s^{K^+}, \\ D_q^{K^+} &= D_{\bar{q}}^{K^+}, \quad q = c, b, \end{aligned} \quad (11)$$

with the heavier charm and bottom quark and antiquark FFs also set equal to each other. For the unidentified charged hadron FFs, we follow the previous JAM FF analysis [38] and use a residual term, $D_q^{\text{res}^+}$, to parametrize the difference between the total charged hadron $D_q^{h^+}$ and the $D_q^{\pi^+}$ and $D_q^{K^+}$ FFs,

$$D_q^{h^+} = D_q^{\pi^+} + D_q^{K^+} + D_q^{\text{res}^+}. \quad (12)$$

Since the residual hadrons are dominated by protons, we allow $D_u^{\text{res}^+}$ and $D_d^{\text{res}^+}$ to be independent, but equate the light sea quark FFs,

$$\begin{aligned} D_{\bar{u}}^{\text{res}^+} &= D_{\bar{d}}^{\text{res}^+} = D_s^{\text{res}^+} = D_{\bar{s}}^{\text{res}^+}, \\ D_q^{\text{res}^+} &= D_{\bar{q}}^{\text{res}^+}, \quad q = c, b. \end{aligned} \quad (13)$$

Finally, we use charge symmetry to relate the FFs for positively and negatively charged hadrons, $D_q^{h^-} = D_q^{h^+}$, for all flavors q .

III. DATASETS

In this section we summarize the datasets that are used in the current global analysis. These include the standard baseline sets of DIS, Drell-Yan lepton-pair production, inclusive weak boson and jet production, as well as recent $W + \text{charm}$ production data from the LHC. To facilitate better flavor separation of the PDFs, we utilize in addition SIDIS data from muon-deuterium scattering at COMPASS. Since the SIDIS process involves also FFs, we simultaneously fit these datasets together with e^+e^- SIA data, constraining the FF and PDF parameters self-consistently. The list of reactions and observables, and their associated connection to PDFs and FFs, is summarized in Table I.

For inclusive DIS experiments, we use fixed-target data on proton and deuteron targets from SLAC [39], BCDMS [40], and NMC [41,42], along with reduced neutral current and charged current proton cross sections for the combined H1 and ZEUS analysis of HERA collider data [43]. For the SIDIS measurements, we include data on pion, kaon, and unidentified charged hadron production on deuteron targets from the COMPASS Collaboration [44,45]. We also note that in previous JAM analyses [38,73] of the SIDIS data, one of the sources of systematic uncertainty was interpreted

TABLE I. Summary of processes, observables, and their relation to PDFs and FFs. The symbols †, ★, and ◊ indicate the z_h cut used for high-energy SIA data, low-energy SIA data, and *BABAR* kaon data, respectively.

Process	Observables	Cuts	PDFs/FFs
$\ell + (p, d) \rightarrow \ell' + X$	F_2, σ_{red} [39–43]	$Q^2 > m_c^2$ $W^2 > 10 \text{ GeV}^2$	$f_{i/p}$
$\ell + d \rightarrow \ell' + (\pi^\pm, K^\pm, h^\pm) + X$	dM^h/dz_h [44,45]	$Q^2 > m_c^2$	$f_{i/p}$ $D_i^{(\pi^\pm, K^\pm, h^\pm)}$
$p + (p, d) \rightarrow \ell \bar{\ell} + X$	$d\sigma^{\rho/d}/dx_F dQ^2$ [2,3]		$f_{i/p}$
$p + (p, \bar{p}) \rightarrow W + X$	$A_W, A_\ell, \sigma^{W^+}/\sigma^{W^-}$ [19–21,23,46–50]		$f_{i/p}$
$p + \bar{p} \rightarrow Z/\gamma^* + X$	$d\sigma/dy$ [51,52]		$f_{i/p}$
$p + (p, \bar{p}) \rightarrow \text{jet} + X$	$d\sigma/d\eta dp_T$ [53–55]	$p_T > 8 \text{ GeV}$	$f_{i/p}$
$p + p \rightarrow W + c + X$	$d\sigma/d\eta$ [25–27]		$f_{i/p}$
$\ell + \bar{\ell} \rightarrow (\pi^\pm, K^\pm, h^\pm) + X$	$d\sigma/dz_h$ [56–72]	$0.02 < z_h < 0.9$ † $0.15 < z_h < 0.9$ ★ $0.20 < z_h < 0.9$ ◊	$D_i^{(\pi^\pm, K^\pm, h^\pm)}$

as a normalization uncertainty. This interpretation required a large fitted normalization to adequately describe the data. In the present analysis this uncertainty is instead interpreted as a point-to-point correlated systematic uncertainty, allowing the data to be well described without the need of a large fitted normalization for each dataset.

Drell-Yan lepton-pair production data from the E866 (NuSea) [2] and E906 (SeaQuest) [3] pp and pD scattering experiments at Fermilab are used to constrain mostly the \bar{d}/\bar{u} PDF ratio in the proton at small and intermediate values of x . Data for weak vector boson mediated processes including W -lepton asymmetries from the CMS [19–21,23] and LHCb [46,47] Collaborations at the LHC, and the STAR [50] Collaboration at RHIC, along with W^\pm charge asymmetries and Z/γ^* rapidity distributions from CDF and D0 at the Tevatron [48,49,51,52], are used to

further constrain flavor separation including at high x . Jet production data from CDF and D0 [53,54] and STAR [55] are also included, which are important for constraining the gluon PDF at large x . For W + charm quark production, which have been argued could be more sensitive to the strange quark PDFs, we use cross sections from the ATLAS [26] and CMS [25,27] Collaborations at the LHC. SIA data involving pion, kaon, and unidentified charged hadron from the ARGUS [56], BABAR [57], Belle [58], TASSO [59–61,69,70], TPC [62,63], TOPAZ [64], ALEPH [65,71], DELPHI [66], OPAL [67,72], and SLD [68] Collaborations are included to constrain the FFs.

The kinematic range covered by the data for each type of dataset and experiment is illustrated in Fig. 1. Here the scale Q^2 is shown versus the Bjorken scaling variable x_B for DIS and SIDIS data, and versus $x_{1,2}$ variables for

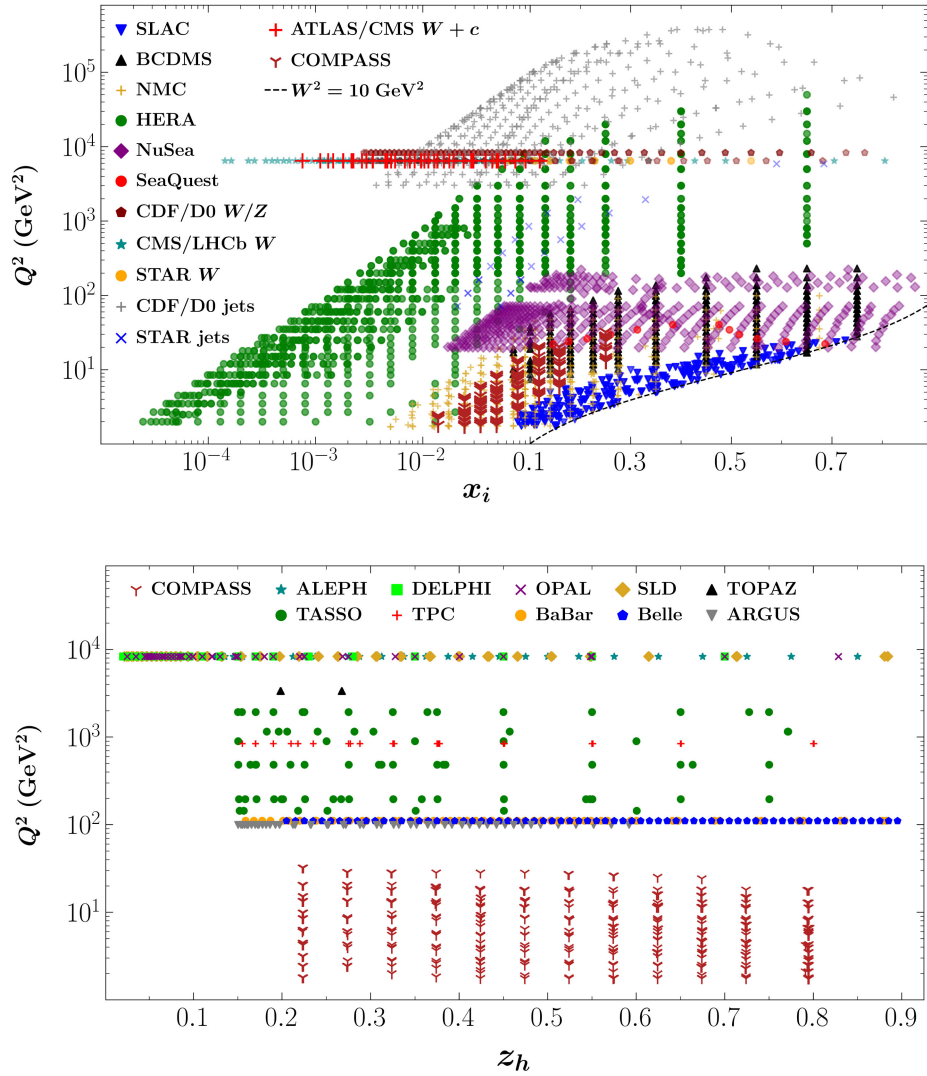


FIG. 1. Kinematic coverage of datasets used in this analysis, with the scale Q^2 versus the Bjorken scaling variable $x_i = x_B$ for DIS/SIDIS, and $x_i = x_1$ or x_2 for Drell-Yan/weak boson/jet production data (upper panel), and the fragmentation variable $z_h = p \cdot p_h / p \cdot q$ for SIDIS and $z_h = 2p_h \cdot q / Q^2$ for SIA data (lower panel).

hadron-induced scattering reactions. In addition, for SIDIS and SIA data, the scale Q^2 is shown versus the fragmentation variables $z_h = p \cdot p_h / p \cdot q$ and $z_h = 2p_h \cdot q / Q^2$, respectively. The SIDIS data are critical, as these are the only datasets that have sensitivity to both the PDFs and FFs.

IV. GLOBAL QCD ANALYSIS

Having outlined the theoretical framework and data selection, in this section we present the results of our global QCD analysis. We will first review the methodological aspects of our analysis, followed by a detailed description of our strategies for constructing posterior distributions for the PDFs and FFs. The quality of the results will be discussed by examining the agreement between data and theory, with a focus on the impact of SIDIS multiplicities for π^\pm , K^\pm , and unidentified hadron h^\pm production data from COMPASS, as well as $W +$ charm production cross section data from the LHC, on the magnitude and shape of the strange quark PDF. We discuss several scenarios, including the baseline results, which exclude SIDIS and $W +$ charm data; two additional scenarios, where the baseline is supplemented with either SIDIS or $W +$ charm data; and the final combined result that incorporates all datasets in the analysis. We label these analyses as “baseline,” “+SIDIS,” “+ W -charm,” and “JAM24,” respectively.

A. Methodology and model calibration

The reconstruction of PDFs and FFs follows the multistep Monte Carlo-based approach for Bayesian inference employed in previous JAM global QCD analyses [4,38,74]. This approach allows us to calibrate our models, quantify their performance across the parameter space, and mitigate overparametrization. The key algorithmic aspects of the methodology include the use of data resampling for optimization, and the multistep strategy developed in Ref. [73].

In the current analysis several notable improvements have been introduced. First, the sea quark PDFs are initially treated as flavor symmetric during the steps prior to incorporating flavor sensitive experimental data. Once such data are added, they allow the separation of specific flavors from the total sea. This incremental approach enables a more reliable determination of PDF uncertainties by gradually introducing flavor asymmetries between different sea quark distributions.

Second, we replace the initial flat prior distributions with Gaussian priors, choosing specific hyperparameters for each type of template parameter. The hyperparameters are chosen to span a large range of solutions to ensure the Gaussian priors are not too informative and bias the space of solutions. We also include corresponding Gaussian penalties of the form $((a_i - \mu_i) / \sigma_i)^2$, where μ_i and σ_i are

the mean and width hyperparameters, respectively, of the distribution for the parameter a_i was sampled. Gaussian penalties are utilized in the optimization procedure across the multisteps to prevent the parameters from growing indefinitely and accumulating at parameter boundaries. In addition, the use of Gaussian penalties ensures the self-consistency of the template modeling (for example, the γ and δ parameters in Eq. (5) should act as correction terms to the leading $x^\alpha(1-x)^\beta$ term of the template function), and enables the specification of priors in extrapolation regions.

Lastly, when computing any physical observables from the MC parameter samples, we use the mean values as the central results, while uncertainties are estimated using Bayesian credible intervals at the 95% confidence level. Specifically, for any observable such as a PDF, FF, cross section or asymmetry, the edges of the confidence intervals are determined using a nonparametric estimate of the inverse cumulative distribution function [75].

With these improvements we first perform model calibration for the baseline scenario, excluding the SIDIS and $W +$ charm data. The multistep optimization is initiated by first considering only DIS data, using a symmetric sea *ansatz*, $\delta\bar{u} = \delta\bar{d} = \delta s = \delta\bar{s} = 0$ and $S_1 = S_2$. The obtained posteriors are then used as priors for the next step in which DY data are added, and $\delta\bar{u}$ and $\delta\bar{d}$ allowed to vary away from zero. The optimization is continued by incrementally extending the dataset to include W/Z and jet production data in subsequent steps. An additional step is performed where S_2 is varied away from S_1 and δs is varied independently away from $\delta\bar{s}$ to account for possible differences between the light and strange quark PDFs. The resulting MC replicas then constitute our baseline results.

During the optimization process of the baseline setup, we found that the γ and δ parameters in Eq. (5) had little or no effect on the δs , $\delta\bar{s}$, and S_2 distributions. Consequently, these parameters were set to zero for these functions. However, they were allowed to vary freely for each of the u_v , d_v , g , $\delta\bar{u}$, $\delta\bar{d}$, and S_1 distributions. This results in a total of 35 free parameters for the PDFs in the baseline analysis, comprising 30 shape parameters and 5 normalization parameters.

The model calibration for FFs is initially performed using only the SIA data. For each hadron, two template shape functions were found to be necessary to describe the data for all quark FFs. Using the FF relations given in Sec. II B, this results in 60 free FF parameters for each hadron species (12 normalization parameters and 48 shape parameters).

As mentioned above, we consider different scenarios as extensions of the baseline fit. These include the addition of SIDIS data, $W +$ charm data, and both SIDIS and $W +$ charm data. Each scenario is implemented by performing additional steps, starting from the baseline results and incorporating the corresponding datasets. Importantly,

the Gaussian penalties applied during the model calibration stages are removed in the final MC runs for each of the scenarios. The combined full analysis utilized 215 parameters to model the input scale PDFs and FFs, along with an additional 43 free normalization parameters for various datasets, bringing the total number of free parameters to be inferred from the data to 258.

B. Figures of merit

As in previous JAM analyses, the optimization of the PDF and FF parameters against the data is performed using nuisance parameters. Specifically, a given experimental data point i from a dataset e with value $d_{i,e}$ is compared against the corresponding factorization-based calculation $T_{i,e}$, with additional systematic distortions of the form

$$T_{i,e} \rightarrow \hat{T}_{i,e} \equiv \sum_k r_e^k \beta_{i,e}^k + \frac{T_{i,e}}{N_e}, \quad (14)$$

where $\beta_{i,e}^k$ represents the k th quoted source of point-to-point correlated systematic uncertainties. The nuisance parameters r_e^k and N_e modify the original theory additively and multiplicatively, respectively, to best describe the data within the quoted systematic uncertainties. To avoid overfitting, the nuisance parameters are regulated with Gaussian penalties $\sum_k (r_e^k)^2$ for the additive shifts and by $(1 - N_e)^2 / (\delta N_e)^2$ for the multiplicative shift, where δN_e is the quoted normalization uncertainty for experiment e .

Following Ref. [76], in previous JAM analyses [4,38] a theory-scaled point-by-point correlated systematic uncertainty was used, with $\beta_{i,e}^k \rightarrow (T_{i,e}/D_{i,e}) \times \beta_{i,e}^k$ in Eq. (14). Using this convention, we found that the SIDIS data could not be well described without introducing large additive shifts. Removing this rescaling significantly decreased systematic shifts needed to describe the SIDIS data, while leaving other dataset descriptions unaffected.

The quantity $\hat{T}_{i,e}$ in Eq. (14) represents a modified theory that accounts for systematic biases present in the data, under the assumption that $T_{i,e}$ is the correct underlying law for the reconstructed observable. In practice, however, $T_{i,e}$ is only computable within the limits dictated by factorization theorems, the perturbative accuracy of the theoretical computation, and the expressivity of the nonperturbative modeling for PDFs and FFs. In this context, the deviations of $\hat{T}_{i,e}$ from $T_{i,e}$ are not solely a measure of experimental biases, but also reflect the computational limitations of $T_{i,e}$. These include constraints from the theoretical framework and modeling assumptions. We therefore aim to minimize these deviations by adjusting our model assumptions, as will be discussed in Sec. IV D below.

To assess the quality of the global analysis, we utilize two figures of merit: the reduced χ^2 for each dataset, defined as

$$\chi_{\text{red}}^2 \equiv \frac{1}{N_{\text{dat}}} \sum_{i,e} \left(\frac{d_{i,e} - E[\hat{T}_{i,e}]}{\alpha_{i,e}} \right)^2, \quad (15)$$

where $E[\dots]$ is the expectation value, and the associated Z-score to evaluate the probabilities of the estimated χ_{red}^2 , taking into account the number of kinematical data points N_{dat} for each dataset. The quantities $\alpha_{i,e}$ are the quadrature sums of all the quoted uncorrelated point-by-point uncertainties. The Z-score is defined in terms of the inverse of the normal cumulative distribution function,

$$Z = \Phi^{-1}(p) \equiv \sqrt{2} \text{erf}^{-1}(2p - 1), \quad (16)$$

where the p -value is computed according to the $\chi^2 \equiv N_{\text{dat}} \chi_{\text{red}}^2$ distribution with the number of data points N_{dat} as the degrees of freedom.

C. Data selection

For the DIS datasets, we apply the kinematic cuts $W^2 > 10 \text{ GeV}^2$ and $Q^2 > m_c^2$ to avoid the nucleon resonance region, and suppress higher twist corrections to the leading twist approximations for the structure functions. For the SIDIS data, we also examined the effects of various cuts on W_{SIDIS}^2 and z_h on the description of the data. We found that the SIDIS multiplicities can be well described using the kinematic cut $Q^2 > m_c^2$, with no additional cuts required on W_{SIDIS}^2 and z_h beyond the kinematic coverage of the COMPASS data, namely, $W_{\text{SIDIS}}^2 > 7 \text{ GeV}^2$ and $0.2 < z_h < 0.8$. Note that the computation of W_{SIDIS}^2 requires specifying the hadron transverse momentum in the Breit frame, p_{hT} [77]. Since the SIDIS cross sections receive their largest contributions from regions with small transverse momentum, we set its value to zero when evaluating W_{SIDIS}^2 .

For other SIDIS data at lower energies, such as those from Jefferson Lab Hall C [78], additional cuts on W_{SIDIS}^2 may be needed to allow one to isolate regions where higher twists and target mass corrections can be neglected and the current and target fragmentation regions can be isolated. For SIA, to avoid large perturbative effects at small and large values of z_h and allow maximal overlap with the SIDIS kinematics of COMPASS, we restrict the data using the upper cut of $z_h < 0.9$ and lower cuts of 0.02 and 0.15 for high-energy and low-energy SIA data, respectively. An exception is the *BABAR* K^\pm dataset, for which the cut $z_h > 0.2$ is needed for a good overall fit, as also observed in Ref. [79].

D. Data versus theory comparison

A summary of the figures of merit is presented in Table II and Fig. 2 for the scenarios discussed in this work. Starting with the baseline scenario, we find that all datasets considered in the analysis are described quite well.

TABLE II. Reduced χ_{red}^2 and Z-score values for each of the four scenarios considered in this analysis (with the total number of data points for each): baseline (4326 data points), +SIDIS (5816), +W-charm (4363), and the full JAM24 analysis (5853). The N_{dat} values listed in the table for each type of observable correspond to the JAM24 fit.

Process	N_{dat}	χ_{red}^2 (Z-score)			
		Baseline	+SIDIS	+W-charm	JAM24
DIS					
fixed target [39–42]	1495	1.07(+1.95)	1.08(+2.23)	1.07(+1.92)	1.08(+2.25)
HERA [43]	1185	1.16(+3.64)	1.18(+4.22)	1.15(+3.57)	1.18(+4.22)
Drell-Yan [2,3]	205	1.15(+1.50)	1.17(+1.69)	1.14(+1.43)	1.17(+1.69)
W-lepton asymmetry [19–21,23,46,47,50]	70	0.88(−0.69)	0.91(−0.50)	0.87(−0.72)	0.91(−0.52)
W charge asymmetry [48,49]	27	0.91(−0.24)	1.00(+0.08)	0.91(−0.25)	1.00(+0.08)
Z rapidity [51,52]	56	1.18(+0.98)	1.22(+1.15)	1.24(+1.25)	1.24(+1.24)
Inclusive jets [53–55]	198	1.02(+0.27)	0.94(−0.56)	1.04(+0.41)	0.94(−0.54)
W + charm [25–27]	37	0.57(−2.10)	0.59(−1.99)
SIDIS					
π^\pm [44]	498	...	0.91(−1.42)	...	0.89(−1.74)
K^\pm [45]	494	...	0.91(−1.50)	...	0.88(−1.96)
h^\pm [44]	498	...	0.90(−1.65)	...	0.86(−2.22)
SIA					
π^\pm [56–62,64–68]	403	0.73(−4.30)	0.82(−2.77)	0.73(−4.30)	0.82(−2.67)
K^\pm [56–61,63–68]	377	0.65(−5.55)	0.78(−3.20)	0.65(−5.55)	0.79(−3.04)
h^\pm [59,61,63,66,68–72]	310	0.76(−3.29)	0.73(−3.73)	0.76(−3.29)	0.73(−3.69)
Total	5853	1.00(+0.11)	1.00(−0.09)	1.00(−0.06)	0.99(−0.51)

An exception is the neutral current $e^\pm p$ HERA data at $\sqrt{s} = 318$ GeV. The reduced χ^2 and Z-score values for this dataset are relatively high, with $\chi_{\text{red}}^2 = 1.35$ and $Z = 4.52$ for $e^+ p$, and $\chi_{\text{red}}^2 = 1.42$ and $Z = 3.37$ for $e^- p$ scattering, indicating potential difficulties in describing the DIS data across the full range of HERA kinematics.

Since the uncertainties are typically larger at higher x , it is likely that the observed tensions occur at small x , where higher-order corrections $\sim \log x$ are known to play a more important role. We have verified this hypothesis by performing the analysis with a larger Q^2 cut, $Q^2 > 3.5$ GeV², on the HERA data to remove data at $x < 3.46 \times 10^{-5}$. The results, shown in Fig. 2 as open circles, demonstrate that the Z-scores can be reduced from 4.22 to as low as 3.19.

To test whether the tensions in the HERA data may be lessened by increasing the perturbative accuracy of our analysis, we compare in Table III the results from the NNLO analyses by NNPDF [80] and CT18 [81]. The comparison indicates that our results are similar to those obtained by these groups. In particular, we find that applying a larger Q^2 cut, as in the NNPDF analysis [80], yields χ_{red}^2 and Z-score values very similar to our results.

Moreover, we have verified that our reconstructed PDFs and FFs are not significantly affected by the value of the DIS cut. However, since our primary objective is to study the reconstruction of the strange quark PDF from SIDIS and W + charm data—and because applying a $Q^2 > 3.5$ GeV² cut would remove a significant amount of

SIDIS data—we retain our nominal $Q^2 > m_c^2$ cut as the final choice for all scenarios.

The inclusion of the SIDIS data in the +SIDIS scenario slightly increases the figures of merit for most PDF-dependent observables, with the most notable cases being the HERA cross sections and W charge asymmetries, which can modify the gluon and light sea quark PDFs, respectively, changing Z-scores within 1σ . Interestingly, while the Z-scores for the SIA pion and kaon data shift by 1.53σ and 2.35σ , respectively, those for the unidentified charged hadron SIA data improve by 0.44σ . However, all SIA χ_{red}^2 and Z-score values remain acceptable with the inclusion of SIDIS data. Moreover, we do not find any significant tensions when combining SIDIS data with the rest of the global dataset considered in our analysis.

Note also that in Ref. [38] the h^- SIDIS dataset was found to have an anomalously low χ_{red}^2 . We have verified that this is in fact to an overestimation of the experimental uncorrelated uncertainties, and the corrected results in the current analysis give a χ_{red}^2 and Z-score comparable to other SIDIS datasets.

The figures of merit in the +W-charm scenario are largely compatible with the baseline results. We find that most dataset descriptions improve, with the exception of the Z rapidity data from CDF and DØ, possibly due to sea quark sensitivity in the low- x region. For the W + charm datasets, both the ATLAS and CMS cross sections are well described, with $\chi_{\text{red}}^2 < 1$, apart from the CMS W + charm ratio measurement, which has a $\chi_{\text{red}}^2 = 1.76$. However, the

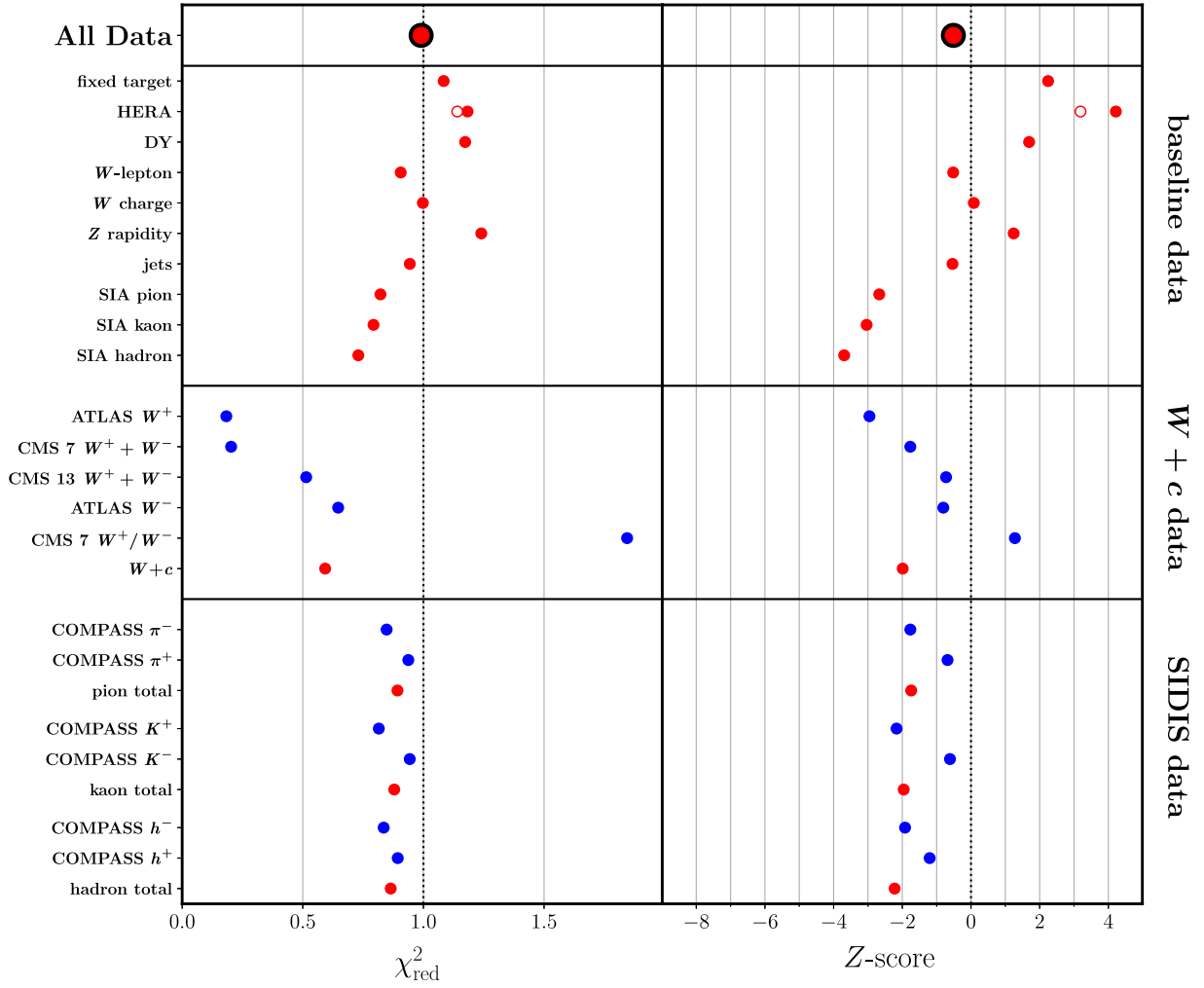


FIG. 2. χ^2_{res} and Z-score values of all datasets used in the current analysis, including for the baseline datasets (top panel), $W + c$ data from CMS and ATLAS (middle panel), and SIDIS π^\pm , K^\pm , and h^\pm production from COMPASS (bottom panel). The reduced χ^2 for individual datasets and averages over datasets are indicated by the blue and red points, respectively, and an additional open circle shows the reduced χ^2 and Z-score of the HERA data using for $Q^2 > 3.5 \text{ GeV}^2$ cut.

relatively low Z-score of 1.19 does not indicate any anomaly.

In the final JAM24 scenario, where both SIDIS and $W + c$ data are fitted along with the baseline datasets, the figures of merit are most similar to the +SIDIS fit. For SIDIS data, we find a decrease in χ^2_{red} compared to the +SIDIS fit, while the χ^2_{red} for the SIA and $W + c$

data increase marginally compared to the +SIDIS and + $W + c$ scenarios, respectively. Focusing on the main datasets of interest for the current analysis, a detailed comparison between theory and data for the ATLAS and CMS $W + c$ production cross sections, differential in the lepton pseudorapidity $|\eta_\ell|$, is shown in Fig. 3 for various values of the lepton transverse momentum lower cutoff, p_T^ℓ . Remarkably, we find a consistent agreement between the description of the $W + c$ data in the +SIDIS and JAM24 fits, indicating no tension between the observables. For all of the CMS $W + c$ summed cross sections, as well as the ATLAS $W^+ + \bar{c}$ data, excellent overall agreement is seen for both the 7 TeV and 13 TeV data. For the ATLAS $W^+ + \bar{c}$ data, the theoretical result slightly underestimates the measurements at $|\eta_\ell| \sim 0.8$, while for the CMS $(W^+ + \bar{c})/(W^- + c)$ ratio the description at $|\eta_\ell| = 1.35$ differs by 2.35σ , leading to a larger χ^2_{red} than for the other datasets.

TABLE III. Comparisons of the reduced χ^2_{red} and Z-score values for HERA inclusive DIS data for the full JAM24 global analysis with the NNPDF4.0 [80] and CT18 [81] fits.

Analysis	pQCD accuracy	Q_{cut}^2 (GeV 2)	N_{dat}	χ^2_{red} (Z-score)
NNPDF4.0 [80]	NNLO	3.49	1145	1.17 (3.89)
CT18 [81]	NNLO	4.0	1120	1.30 (6.51)
JAM24	NLO	3.5	1120	1.14 (3.19)
	NLO	m_c^2	1185	1.18 (4.22)

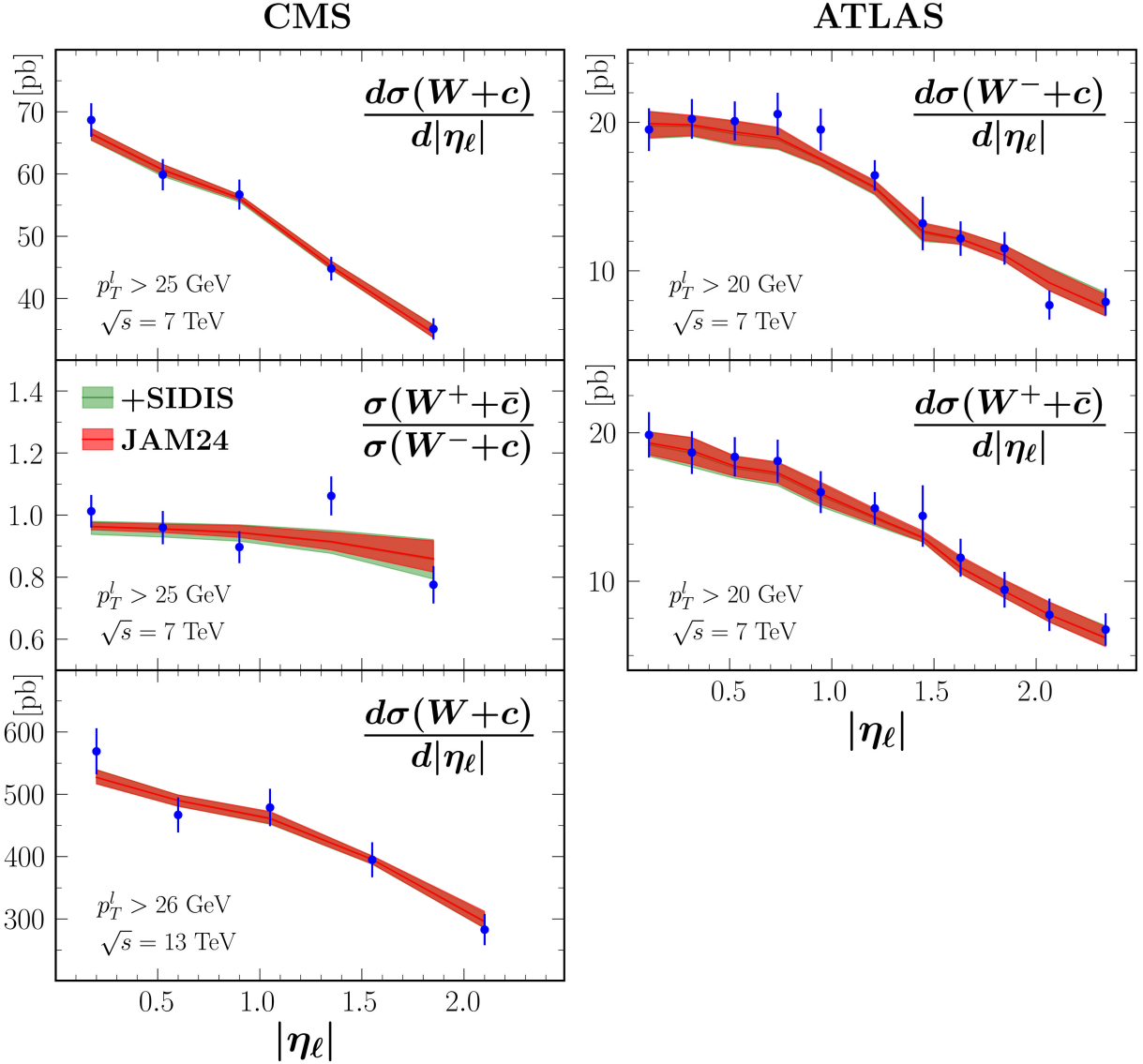


FIG. 3. Comparison of CMS and ATLAS $W + \text{charm}$ data with the JAM24 global analysis as a function of lepton pseudorapidity, $|\eta_\ell|$, with 95% credible interval uncertainty bands. (Left): Sum (top) and ratio (middle) at 7 TeV [25], and sum (bottom) at 13 TeV [27] of $W^+ + \bar{c}$ and $W^- + c$ cross sections from CMS. (Right): Differential cross section of $W^- + c$ (top) and $W^+ + \bar{c}$ (middle) from ATLAS at 7 TeV [26]. The cuts on the transverse momentum of the final state lepton p_T^ℓ are indicated on the panels.

Both the ATLAS and CMS data indicate a slightly larger $W^- + c$ cross section compared with that for $W^+ + \bar{c}$. In principle, this could suggest a positive strange-antistrange asymmetry, $s - \bar{s}$, arising from a small preference of the $gs \rightarrow W^- + c$ channel versus the $g\bar{s} \rightarrow W^+ + \bar{c}$ channel. However, this conclusion is still somewhat limited by the current experimental uncertainties, and additional, higher precision data would be needed to make more a definitive statement.

Turning now to the SIDIS data from COMPASS, in Figs. 4, 5, and 6 we show the z_h dependence of the π^\pm , K^\pm , and h^\pm multiplicities, respectively, defined as ratios of cross sections for SIDIS and inclusive DIS at the same values of x_B and Q^2 ,

$$\frac{dM^h}{dz_h} = \frac{d\sigma^h/dx_B dz_h dQ^2}{d\sigma^{\text{DIS}}/dx_B dQ^2}. \quad (17)$$

For all pions and unidentified hadrons, better agreement between theory and experimental data is found for negatively charged hadrons than for positively charged hadrons, while the opposite is found for kaon production. This can be observed in both the high- x_B and low- Q^2 (low- y) regions, where the differences between the positively and negatively charged hadron multiplicities become larger, especially for kaons. In general, a good description of the COMPASS data is obtained across most kinematic regions and bins, with poorer agreement in the high- z_h region and the low- Q^2 region, particularly for the high- x_B

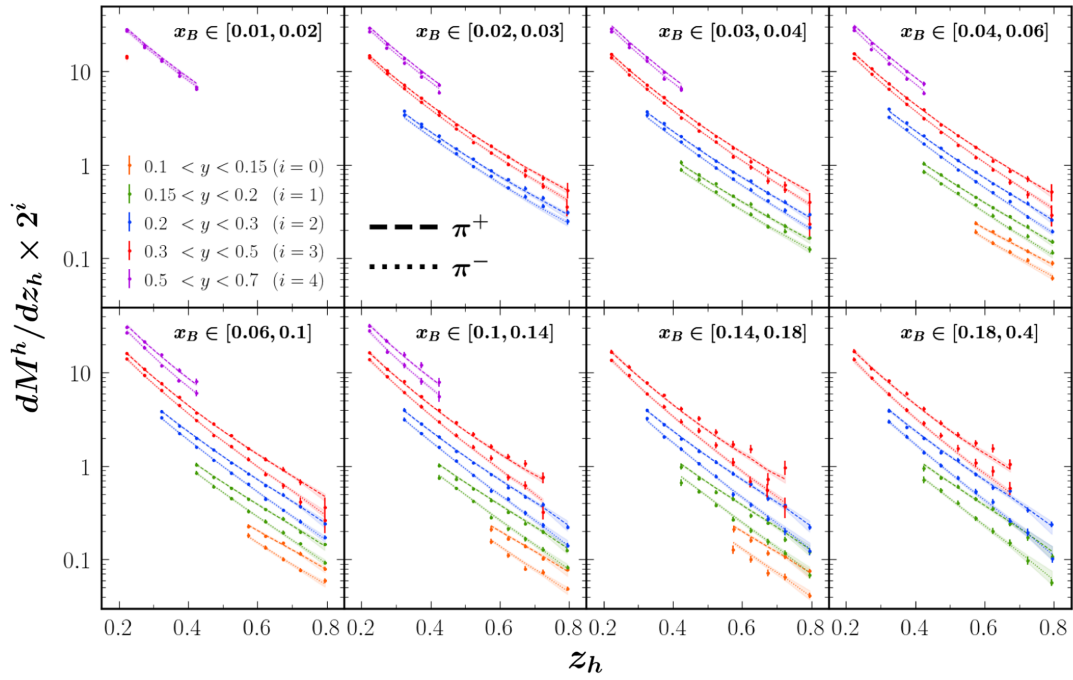


FIG. 4. SIDIS multiplicities dM^h/dz_h from COMPASS [44] as a function of z_h for $h = \pi^+$ (dashed lines) and $h = \pi^-$ (dotted lines) compared with the JAM24 fit result. The data and curves are scaled by a factor of 2^i ($i = 0, \dots, 4$) to more clearly separate and isolate the various x_B and y bins.

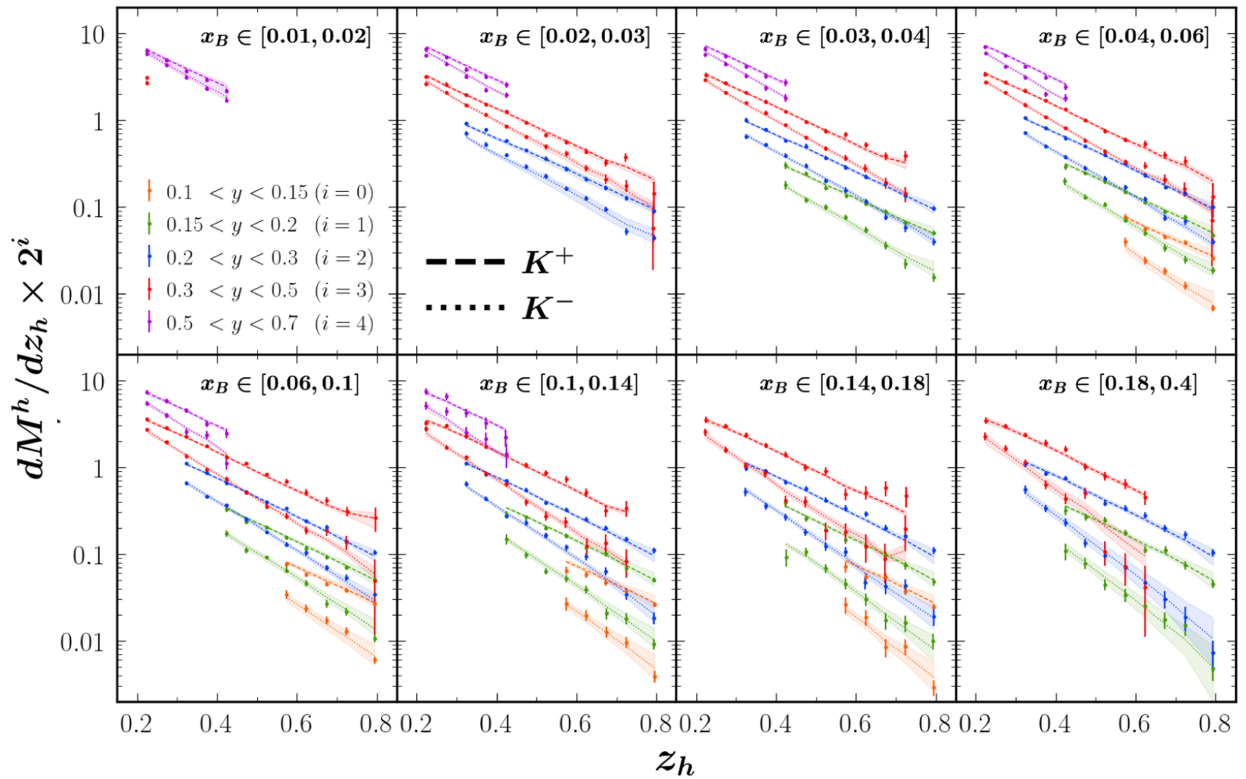
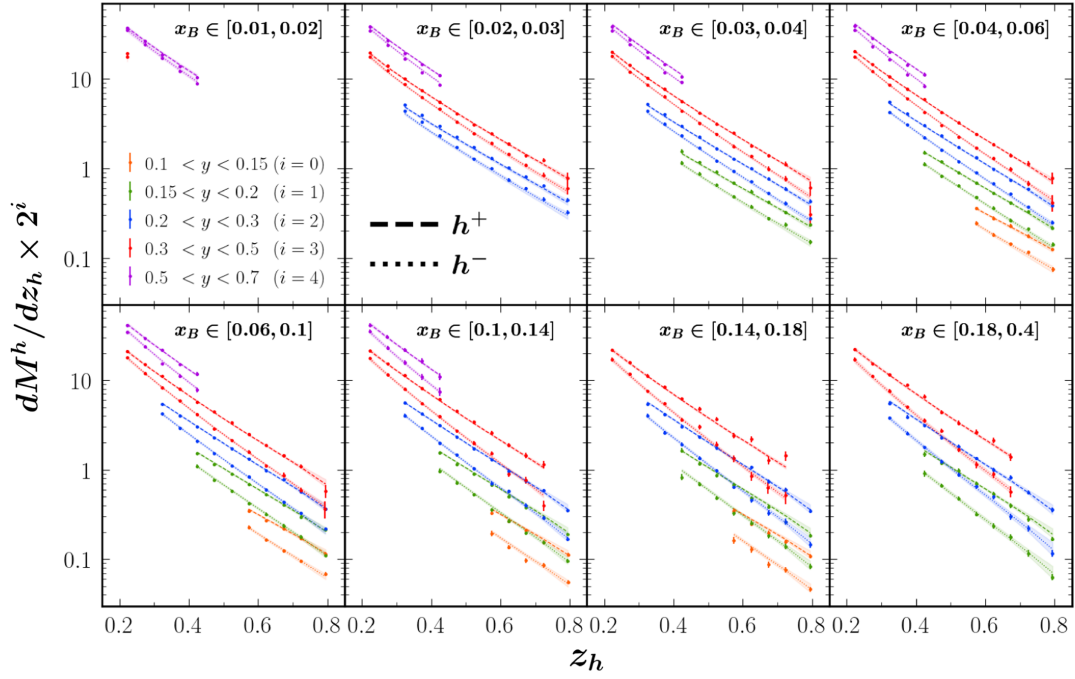


FIG. 5. As in Fig. 4, but for K^\pm data.

FIG. 6. As in Fig. 4, but for unidentified hadrons h^\pm .

bins and some of the lower- x_B bins. Better agreement may be obtained by including hadron mass and other power corrections [82–84], which are known to be more important at large x_B and large z_h , as one approaches the exclusive limit.

For completeness, we also present in Fig. 7 the data and theory comparisons for DY datasets from the NuSea and SeaQuest experiments, as well as the W -lepton cross section ratios from STAR, demonstrating an excellent description of these data within the JAM24 scenario. While these datasets are not directly sensitive to the strange quark PDF, they do provide constraints on the light sea

sector, particularly the $\bar{d} - \bar{u}$ asymmetry. Changes in the light sea quark sector can, however, induce modifications in the strange quark PDF through QCD evolution and the overall description of the global dataset.

Lastly, we address the impact of adopting different small- x asymptotic behaviors for the light sea and strange quark PDFs, modeled using the two independent functions, S_1 and S_2 , in Eq. (7). During the model calibration, we observed differences in the data by contrasting the results for $S_1 \neq S_2$ and $S_1 = S_2$. While both cases provide good descriptions of the data, we found that $S_1 = S_2$ gives a slightly poorer description of the data requiring the need to

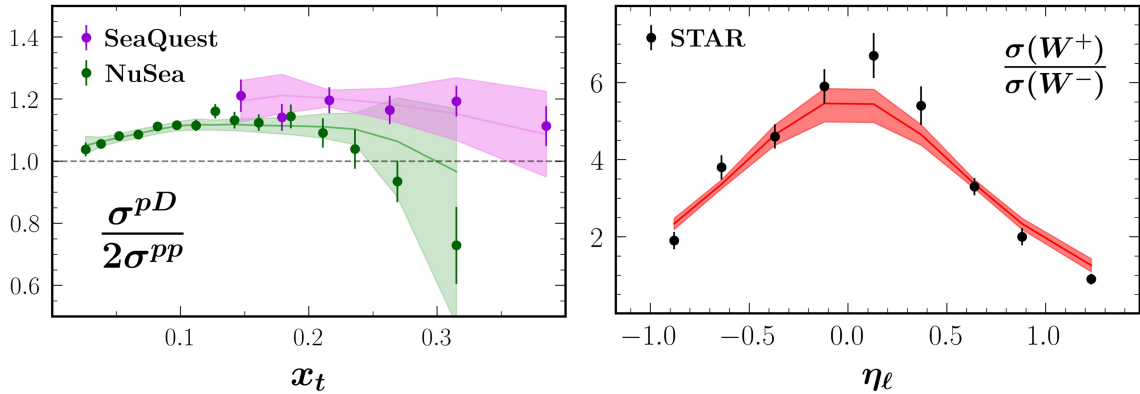


FIG. 7. Left: ratio of pD to pp Drell-Yan cross sections versus the target momentum variable x_t , from SeaQuest (purple circles) for $0.48 < x_b < 0.69$ and NuSea (green circles) for $0.32 < x_b < 0.56$, compared to the JAM24 fit result (colored bands). Right: W -lepton ratio cross sections for pp collisions from STAR at $\sqrt{s} = 510$ GeV and $p_T^\ell > 25$ GeV as a function of pseudorapidity η_ℓ compared to the JAM24 fit (red band).

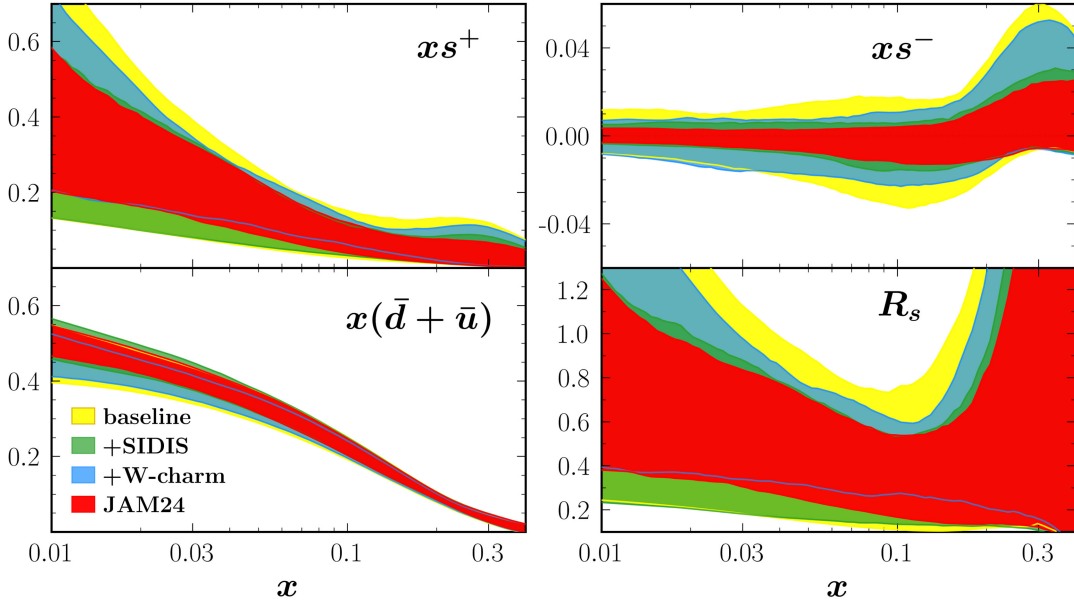


FIG. 8. Impact of the individual SIDIS (green bands) and $W + \text{charm}$ (blue bands) data on the PDFs xs^+ , xs^- , $x(\bar{d} + \bar{u})$ and ratio R_s compared with the baseline fit (yellow bands) and the full JAM24 fit (red bands) that includes all of the data. Uncertainty bands represent a 95% credible interval at the scale $Q^2 = 4 \text{ GeV}^2$.

include larger systematic shifts to the theory. We therefore choose the more general $S_1 \neq S_2$ condition in all the scenarios explored in this analysis.

E. Reconstructed PDFs and FFs

Having established the agreement between our fits and the data, we now present the reconstructed PDFs and FFs in the various scenarios, focusing in particular on understanding the impact of the SIDIS and $W + \text{charm}$ data on both the magnitude and constraints of the strange quark PDF. The effects of the additional datasets beyond the baseline on the strange and light antiquark distributions are illustrated in Fig. 8, where results for the initial baseline, +SIDIS, + W -charm, and the final JAM24 fit that incorporates both the SIDIS and $W + \text{charm}$ datasets, are compared at a common scale of $Q^2 = 4 \text{ GeV}^2$.

The inclusion of either the SIDIS or $W + \text{charm}$ datasets is clearly observed to decrease the uncertainties in the $s^+ = s + \bar{s}$ distribution for all x . Both datasets decrease the upper limit of uncertainty in s^+ compared to the baseline analysis, with SIDIS data constraining the upper limit more than $W + \text{charm}$ data; however, the $W + \text{charm}$ data decrease the lower uncertainty, whereas SIDIS data give results similar to those for the baseline fit. The net strange quark content between all scenarios is within uncertainties, but with SIDIS data having a slight preference for lower strangeness, and $W + \text{charm}$ data preferring larger strangeness. For the light sea quark sector, there is agreement between scenarios for all x values within uncertainties.

The ratio of strange to nonstrange sea quark PDFs follows a similar behavior as the s^+ PDF, with the

$W + \text{charm}$ data raising the lower uncertainty bound and expanding the upper uncertainty bound, while the SIDIS data align with the baseline fit at the lower uncertainty bound but reduce the upper uncertainty compared to the $W + \text{charm}$ data. All scenarios produce a monotonically decreasing R_s in the range $0.01 < x < 0.3$, while outside of this range the uncertainties become too large to draw clear conclusions. This suggests compatibility with an SU(3) symmetric sea for $x \lesssim 0.02$, albeit within sizeable uncertainties, but a suppressed strange quark sea at higher x , in contrast to the earlier finding by the ATLAS Collaboration [26]. This discrepancy may arise due to dataset limitations or insufficient parametrization flexibility in that analysis. Interestingly, the apparent rise in R_s at $x \gtrsim 0.2$ is similar to that observed in Ref. [85], which included also neutrino-nucleus cross sections but not SIDIS data. The final results of the full JAM24 fit favor a small strange quark suppression consistent with the +SIDIS scenario, which reflects the larger quantity of SIDIS data (1490 points) compared to the $W + \text{charm}$ data (37 points).

While both the SIDIS and $W + \text{charm}$ data reduce the uncertainty on s^+ , particularly at $x \lesssim 0.1$, future data with greater sensitivity to the strange quark will be needed to further reduce the uncertainty in this region. Such data could come from experiments at Jefferson Lab, involving parity-violating inclusive DIS measurements and high-precision SIDIS multiplicity data from Halls B and C. For the $s^- = s - \bar{s}$ difference, we find a notable consistency for all scenarios, which show a magnitude consistent with zero for all x , but a slight rise at $x \sim 0.3$. The datasets with sensitivity to strangeness show a reduction of uncertainty, with SIDIS data constraining the upper limit more

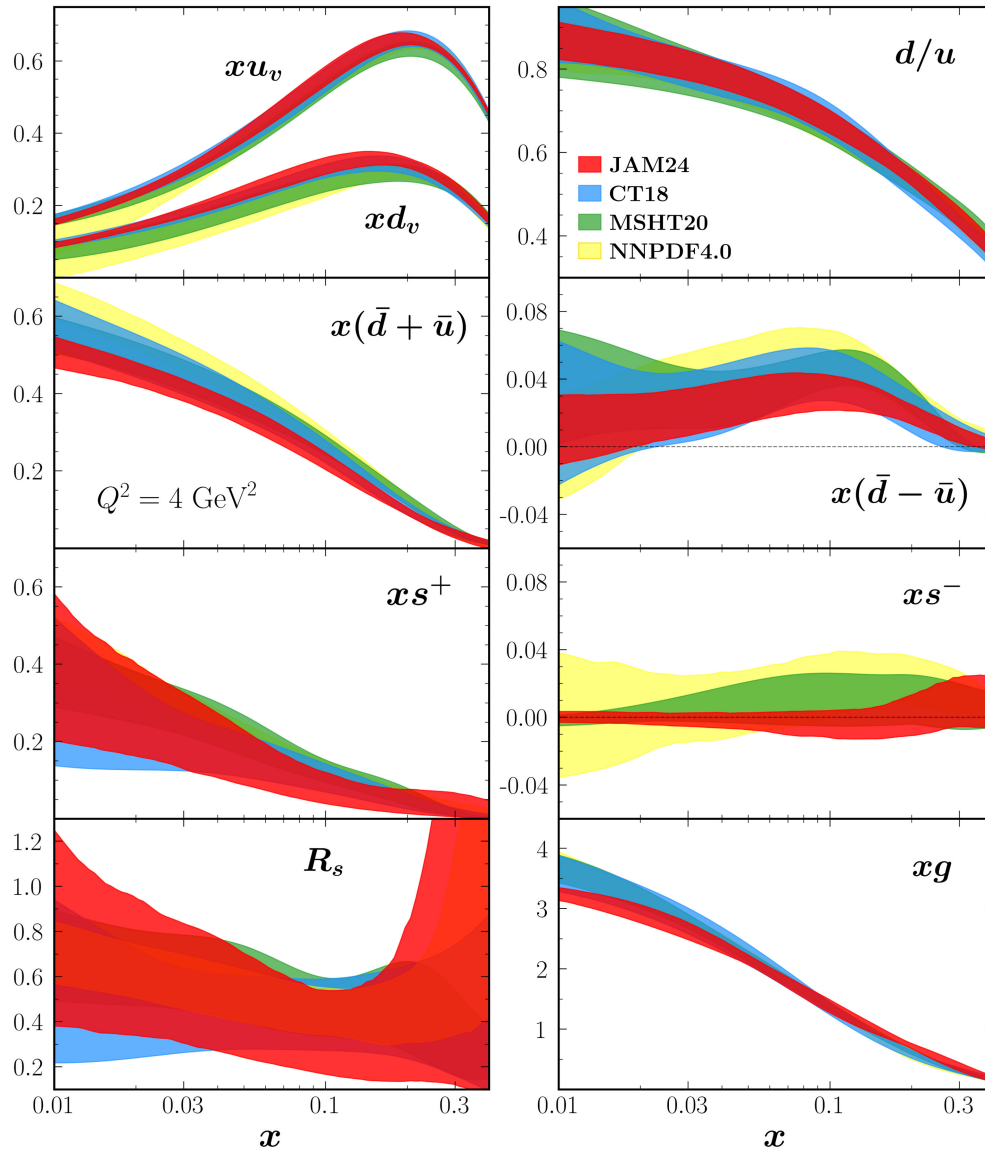


FIG. 9. Comparison of various PDFs from the present JAM24 global analysis (red bands) with results from the CT18 [81] (blue), MSHT20 [86] (green), and NNPDF4.0 [80] (yellow) NLO parametrizations at the scale $Q^2 = 4 \text{ GeV}^2$. Bands represent a 95% credible interval for JAM24 and NNPDF4.0, a 90% CL for CT18, and a 68% CL for MSHT20.

than the $W + \text{charm}$ data at high x . Although QCD does not require the strange and antistrange PDFs to be identical, the current datasets provide no clear indication of a nonzero s^- asymmetry within the present uncertainties.

The results of our full global analysis for the PDFs are illustrated in Fig. 9, where we show each of the fitted PDFs at $Q^2 = 4 \text{ GeV}^2$. For comparison, we also show the results from other NLO parametrizations, including the CT18 [81], MSHT20 [86] and NNPDF4.0 [80], with uncertainty bands representing 90% confidence level (CL), 68% CL, and 95% credible interval, respectively. In these comparisons we focus on the kinematic region of parton momentum fractions $0.01 \lesssim x \lesssim 0.4$ where the $W + \text{charm}$ and SIDIS datasets have the greatest impact on the strange quark PDF.

For the valence u - and d -quark distributions, as well as the d/u PDF ratio, our results are in overall agreement with those from other groups. The total light antiquark distribution $\bar{d} + \bar{u}$ is slightly below the NNPDF4.0 result, but is in agreement with other fits for $x \lesssim 0.03$, and is consistent for higher x , while the $\bar{d} - \bar{u}$ difference sits at the lower end of the other PDF sets in the intermediate- x range. The total strange distribution s^+ is in general agreement with other fits at low x , but is slightly lower compared to NNPDF4.0 and MSHT20 in the $0.1 \lesssim x \lesssim 0.2$ range. This is also seen with the strange to nonstrange sea quark ratio, R_s , which is $0.72^{+0.52}_{-0.34}$ at $x = 0.01$, but falls to $0.34^{+0.47}_{-0.21}$ at $x = 0.2$. The observed differences in uncertainty bands are likely attributed to methodological choices in uncertainty estimation rather than differences in data selection.

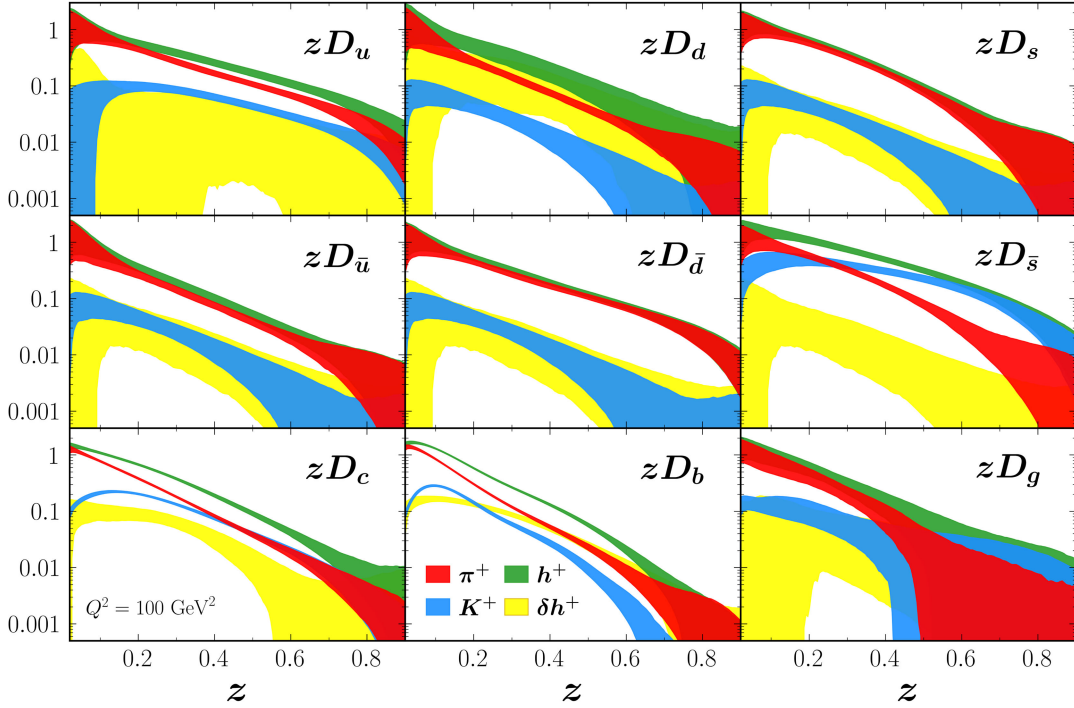


FIG. 10. Parton to hadron fragmentation functions zD_f versus z for the production of charged pions π^+ (red bands), kaons K^+ (blue), unidentified hadrons h^+ (green), and residual hadrons δh^+ (yellow) for 95% credible interval at $Q^2 = 100 \text{ GeV}^2$.

For the FFs extracted in our analysis, in Fig. 10 we show the z dependence for positively charged π^+ , K^+ , unidentified hadrons h^+ , and residual hadrons δh^+ defined in Eq. (12) at a scale $Q^2 = 100 \text{ GeV}^2$, for a 95% credible interval. Since the pion is the lightest hadron, as expected we find the magnitude of the π^+ FFs is generally larger than those for kaons and other hadrons for most quark flavors. An exception to this is the FF for $\bar{s} \rightarrow K^+$, where the \bar{s} flavor is favored by K^+ but disfavored by π^+ . Additionally, the π^+ FFs are comparable to the $c \rightarrow K^+$ and the $b \rightarrow \delta h^+$ FFs in the intermediate- z region, and $g \rightarrow \pi^+$ comparable to other hadrons in the high- z sector. These findings are similar to those found in previous JAM analyses [38,73] and by other groups [87].

A summary of our findings for the strange to nonstrange ratio R_s is shown in Fig. 11 for $x = 0.01, 0.04, \text{ and } 0.1$ at $Q^2 = 4 \text{ GeV}^2$. Here we compare the R_s values from the all fits considered in this analysis, namely, the baseline +SIDIS, + W -charm, and the full JAM24 analysis, with results from the CT18 [81], ATLAS21 [88], and NNPDF4.0 [80] PDF analyses at the same x and Q^2 , and with the HERMES extraction at $Q^2 = 2.5 \text{ GeV}^2$ [14]. For the final JAM24 result, we find a strange to nonstrange sea quark ratio of $R_s = 0.72^{+0.52}_{-0.34}$ at $x = 0.01$ to be compatible with other fits in this analysis and results from other groups. In the intermediate- x region, we find a more suppressed strange-quark PDF, with $R_s = 0.46^{+0.30}_{-0.20}$ at $x = 0.04$ and $R_s = 0.32^{+0.23}_{-0.15}$ at a higher $x = 0.1$. This is again compatible

with most other fits at these kinematics and is most similar in magnitude and uncertainty to the +SIDIS fit.

V. CONCLUSIONS

How “strange” is the proton has been a question that has perplexed nuclear and particle physicists for decades, attracting considerable experimental and theoretical attention in the quest to understand the detailed structure of the proton’s sea quarks, and in particular the size of the strange sea relative to the light quark sea. In this study we have for the first time performed a comprehensive global QCD analysis of the effect on the strange quark PDF in the proton from simultaneously including W + charm production data in pp collisions at the LHC and hadron production data in semi-inclusive muon-deuteron scattering from COMPASS, both of which are expected to have sensitivity to the s and \bar{s} quark distributions. The SIDIS data in particular require a simultaneous fit to both unpolarized PDFs and parton to hadron FFs.

An excellent overall reduced χ^2 value of $\chi^2_{\text{red}} = 0.99$ is obtained for the global fit, which includes datasets from inclusive DIS, Drell-Yan lepton-pair production, weak boson and jet production, SIDIS, and e^+e^- SIA reactions, comprising nearly 6 000 data points. It is the first time that such a large body of data, constraining both unpolarized PDFs and FFs, has been successfully described within a collinear QCD factorization framework. We find $\chi^2_{\text{red}} \lesssim 1$ for both the LHC W + charm datasets, and the COMPASS

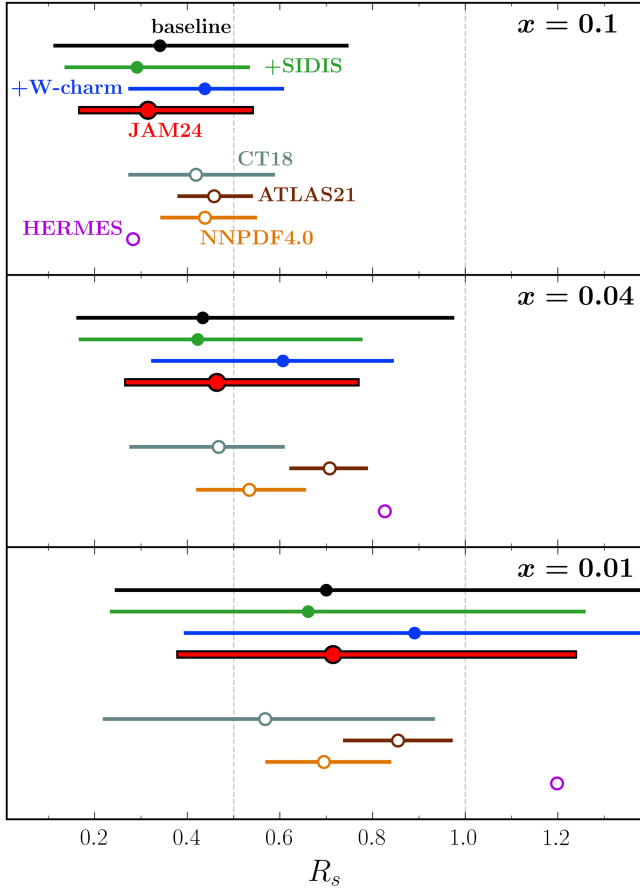


FIG. 11. Strange to nonstrange quark PDF ratio R_s at $x = 0.1$ (top), $x = 0.04$ (middle), and $x = 0.01$ (bottom) at $Q^2 = 2.5 \text{ GeV}^2$ for HERMES data [14] and $Q^2 = 4 \text{ GeV}^2$ for all other fits [80,81,88]. The main JAM24 results (red circles) are compared with various other scenarios and analyses. Uncertainties shown represent a 95% credible interval for baseline, +SIDIS, +W-charm, JAM24, and NNPDF4.0, a 90% confidence level for CT18, and a tolerance $T = 1$ for ATLAS21.

SIDIS data on π^\pm , K^\pm , and unidentified h^\pm lepton production. In comparison to a baseline global fit that does not include these datasets, we find that the combined SIDIS and $W + \text{charm}$ datasets are consistent in magnitude but reduce the uncertainty bounds for all x . Without the $W + \text{charm}$ data, a broader range of suppressed s^+ PDFs are allowed for $x \lesssim 0.1$, while without SIDIS data a more enhanced range of s^+ PDFs is permitted for all x .

For the combined JAM24 fit, the SIDIS data have a stronger pull on s^+ at high x , while both the SIDIS and $W + \text{charm}$ data have similar pulls at lower x values. Overall, the strange to nonstrange sea quark distribution ratio at $Q^2 = 4 \text{ GeV}^2$ is determined to be $R_s =$

$\{0.72^{+0.52}_{-0.34}, 0.46^{+0.30}_{-0.20}, 0.32^{+0.22}_{-0.15}\}$ for $x = \{0.01, 0.04, 0.1\}$. We conclude therefore that at $x \approx 0.01$, the global dataset is compatible with an SU(3) flavor symmetric sea, but indicates strong SU(3) symmetry breaking in the sea for $x \gtrsim 0.02$. Our analysis also places more stringent constraints on the strange–antistrange PDF asymmetry, $s - \bar{s}$, with a significant reduction relative to the baseline fit, but still compatible with zero over the entire x range studied.

In the future our analysis can be extended in several ways. First, we plan to include inclusive $pp \rightarrow hX$ data to better constrain the FFs [37], supplementing the constraints from SIA and SIDIS. We will also explore the possibility of using new lattice QCD simulations of PDF moments and pseudo-Ioffe time distributions to provide complementary constraints on the s and \bar{s} distributions. Upcoming data that will help with the reconstruction of the s and \bar{s} PDFs will include parity-violating DIS, which gives access to a new combination of u^+ , d^+ and s^+ PDFs, as well as SIDIS for pion and kaon production at Jefferson Lab, with a 12 GeV and possibly a 22 GeV electron beam, as well as the Electron-Ion Collider. Finally, a definitive analysis will also include neutrino-nucleus DIS data, which historically been used to constrain the strange-quark PDF, with a systematic treatment of nuclear and hadronization uncertainties.

The PDF and FF replicas for JAM24 produced in this analysis will be available on LHAPDF [89].

ACKNOWLEDGMENTS

We thank Rabah Abdul Khalek, Patrick Barry, Chris Cocuzza, Amanda Cooper-Sarkar, Emanuele Nocera, and Juan Rojo for helpful discussions and communications. This work was supported by the DOE Contract No. DE-AC05-06OR23177, under which Jefferson Science Associates, LLC operates Jefferson Lab. T. A. acknowledges support from JSA/JLab Graduate Fellowship Program. The work of N. S. was supported by the DOE, Office of Science, Office of Nuclear Physics in the Early Career Program.

DATA AVAILABILITY

The data that support the findings of this article are openly available [89].

APPENDIX: FIT SUMMARY

In Tables IV–XIV, we summarize the figures of merit for all datasets used in the JAM24 analysis. The values presented are determined from the expectation values of the modified theory given in Eq. (14).

TABLE IV. DIS datasets used in the present JAM24 analysis. Along with each dataset is the corresponding observable measured, number of data points after/before kinematic cuts, number of correlated systematic uncertainties, reduced χ^2 and Z-score, correlated χ^2 , normalized χ^2 , and the fitted normalization.

Dataset	Observable	N_{dat}	N_{cor}	χ^2_{red} (Z-score)	χ^2_{cor}	χ^2_{norm}	Fitted norm.
SLAC [39]	F_2^p	218/661	...	0.87(-1.38)	...	0.55	0.984(9)
SLAC [39]	F_2^d	228/692	...	0.65(-4.25)	...	2.04	0.976(9)
BCDMS [40]	F_2^p	348/351	5	1.14(+1.74)	13.44	1.02	0.970(6)
BCDMS [40]	F_2^d	254/254	5	1.07(+0.84)	7.99	0.21	0.986(8)
NMC [41]	F_2^p	273/292	11	1.67(+6.62)	8.35	0.71	1.017(10)
NMC [42]	F_2^d/F_2^p	174/260	5	0.90(-0.90)	3.36	0.08	0.997(6)
HERA NC e^+ (1) [43]	σ_{red}^p	402/485	169	1.42(+5.27)	120.65
HERA NC e^+ (2) [43]	σ_{red}^p	75/112	169	1.03(+0.22)	13.77
HERA NC e^+ (3) [43]	σ_{red}^p	259/260	169	0.89(-1.25)	14.13
HERA NC e^+ (4) [43]	σ_{red}^p	209/209	169	0.98(-0.19)	15.74
HERA NC e^- [43]	σ_{red}^p	159/159	169	1.45(+3.58)	25.58
HERA CC e^+ [43]	σ_{red}^p	39/39	169	1.17(+0.80)	2.78
HERA CC e^- [43]	σ_{red}^p	42/42	169	1.02(+0.15)	5.91

TABLE V. As in Table IV, but for Drell-Yan data.

Dataset	Observable	N_{dat}	N_{cor}	χ^2_{red} (Z-score)	χ^2_{cor}	χ^2_{norm}	Fitted norm.
E866 (NuSea) [2]	$M^3 d^2\sigma/dM dx_F$	184/184	...	1.20(+1.88)	...	0.31	0.961(28)
E866 (NuSea) [2]	$d\sigma^{p/d}/2d\sigma^{pp}$	15/15	...	1.00(+0.12)	...	1.07	0.990(13)
E906 (SeaQuest) [3]	$d\sigma^{p/d}/2d\sigma^{pp}$	6/6	1	0.65(-0.51)	0.07	0.36	1.012(10)

TABLE VI. As in Table IV, but for W-lepton asymmetry data.

Dataset	Observable	N_{dat}	N_{cor}	χ^2_{red} (Z-score)	χ^2_{cor}
CMS 8 TeV [23]	A_μ	11/11	6	0.36(-1.91)	0.22
CMS 7 TeV [21]	A_μ	11/11	12	1.44(+1.05)	0.39
CMS 7 TeV [20]	A_e	11/11	2	0.83(-0.27)	0.32
CMS 7 TeV [19]	A_e	6/6	2	0.79(-0.19)	0.36
CMS 7 TeV [19]	A_μ	6/6	2	0.05(-3.34)	0.43
LHCb 7 TeV [46]	A_μ	8/8	...	0.48(-1.12)	...
LHCb 8 TeV [47]	A_μ	8/8	1	0.39(-1.43)	0.03
STAR [50]	$\sigma^{W^+}/\sigma^{W^-}$	9/9	...	2.49(+2.42)	...

TABLE VII. As in Table IV, but for W charge asymmetry data.

Dataset	Observable	N_{dat}	N_{cor}	χ^2_{red} (Z-score)	χ^2_{cor}
CDF [49]	A_W	13/13	5	1.25(+0.73)	2.80
DØ [48]	A_W	14/14	...	0.76(-0.57)	...

TABLE VIII. As in Table IV, but for Z rapidity data.

Dataset	Observable	N_{dat}	N_{cor}	χ^2_{red} (Z-score)	χ^2_{cor}	χ^2_{norm}	Fitted norm.
CDF [52]	$d\sigma/dy$	28/28	1	1.35(+1.28)	2.32	1.11	0.937(12)
DØ [51]	$\sigma^{-1}d\sigma/dy$	28/28	1	1.13(+0.54)	0.41

TABLE IX. As in Table IV, but for inclusive jet production data.

Dataset	Observable	N_{dat}	N_{cor}	χ_{red}^2 (Z-score)	χ_{cor}^2	χ_{norm}^2	Fitted norm.
DØ [53]	$d^2\sigma/d\eta dp_T$	110/110	22	0.91(-0.62)	17.71	4.46	1.137(34)
CDF [54]	$d^2\sigma/d\eta dp_T$	76/76	20	0.91(-0.54)	15.37	0.93	1.056(37)
STAR [55]	$d^2\sigma/2\pi d\eta dp_T$	3/5	1	0.11(-1.66)	0.42	0.03	1.013(3)
STAR [55]	$d^2\sigma/2\pi d\eta dp_T$	9/9	1	1.91(+1.68)	0.23	1.61	1.101(16)

TABLE X. As in Table IV, but for W + charm production data.

Dataset	Observable	N_{dat}	N_{cor}	χ_{red}^2 (Z-score)	χ_{cor}^2	χ_{norm}^2	Fitted norm.
ATLAS 7 TeV [26]	$d\sigma^{W^+c}/d \eta $	11/11	86	0.65(-0.81)	6.88	0.08	0.995(55)
ATLAS 7 TeV [26]	$d\sigma^{W^+c}/d \eta $	11/11	86	0.18(-2.96)	4.21	0.10	0.994(54)
CMS 7 TeV [25]	$d\sigma^{W^+c}/d \eta $	5/5	5	0.20(-1.77)	3.14
CMS 7 TeV [25]	$\sigma^{W^+c}/\sigma^{W^+c}$	5/5	...	1.84(+1.28)
CMS 13 TeV [27]	$d\sigma^{W^+c}/d \eta $	5/5	12	0.51(-0.73)	0.93	0.00	1.000(58)

TABLE XI. As in Table IV, but for SIDIS data.

Dataset	Observable	N_{dat}	N_{cor}	χ_{red}^2 (Z-score)	χ_{cor}^2
COMPASS [44]	dM^{π^+}/dz_h	249/311	1	0.94(-0.68)	2.41
COMPASS [44]	dM^{π^-}/dz_h	249/311	1	0.85(-1.77)	4.97
COMPASS [45]	dM^{K^+}/dz_h	247/309	1	0.82(-2.17)	4.21
COMPASS [45]	dM^{K^-}/dz_h	247/309	1	0.94(-0.61)	6.50
COMPASS [44]	dM^{h^+}/dz_h	249/311	1	0.89(-1.20)	0.74
COMPASS [44]	dM^{h^-}/dz_h	249/311	1	0.84(-1.92)	2.49

TABLE XII. As in Table IV, but for e^+e^- SIA data with π^\pm production. Note that an asterisk (*) indicates that a normalization is fitted without penalty for datasets not providing normalized measurements, a dagger (†) indicates datasets with shared normalizations, and a star (★) indicates that a covariance matrix was provided for systematic uncertainties.

Dataset	Observable	N_{dat}	N_{cor}	χ_{red}^2 (Z-score)	χ_{cor}^2	χ_{norm}^2	Fitted norm.
Belle [58]	$d\sigma/dz$	70/78	...	0.19(-7.71)	0.624(9)*
TASSO 12 GeV [59]	$d\sigma/dz$	1/5	...	0.15(-0.51)	2.7(3.1)*
TASSO 14 GeV [60]	$d\sigma/dz$	5/11	...	2.58(+1.98)	0.956(49)*
TASSO 22 GeV [60]	$d\sigma/dz$	5/13	...	0.57(-0.59)	1.075(72)*
TASSO 34 GeV [61]	$\sigma^{-1}d\sigma/dz_p$	6/16	...	1.60(+1.07)	...	0.65	0.952(24)
TASSO 44 GeV [61]	$\sigma^{-1}d\sigma/dz_p$	4/12	...	1.87(+1.22)	...	0.52	0.957(28)
TPC(c) [62]	$(\sigma\beta)^{-1}d\sigma/dz$ (c)	4/15	...	0.59(-0.43)
TPC(b) [62]	$(\sigma\beta)^{-1}d\sigma/dz$ (b)	4/15	...	1.85(+1.19)
TOPAZ [64]	$\sigma^{-1}d\sigma/d\xi$	2/17	...	0.54(-0.21)
SLD [68]	$\sigma^{-1}d\sigma/dz_p$	34/40	...	0.83(-0.65)	...	0.07	1.003(3)†
SLD(uds) [68]	$\sigma^{-1}d\sigma/dz_p$ (uds)	34/40	...	0.77(-0.94)	1.003(3)†
SLD(c) [68]	$\sigma^{-1}d\sigma/dz_p$ (c)	34/40	...	1.07(+0.36)	1.003(3)†
SLD(b) [68]	$\sigma^{-1}d\sigma/dz_p$ (b)	34/40	...	0.60(-1.84)	1.003(3)†
ALEPH [65]	$\sigma^{-1}d\sigma/dz_p$	23/39	...	0.54(-1.82)	...	0.11	0.990(3)
OPAL [67]	$\sigma^{-1}d\sigma/dp$	24/51	...	1.73(+2.18)
DELPHI [66]	$\sigma^{-1}d\sigma/dz_p$	21/23	...	0.99(+0.07)

(Table continued)

TABLE XII. (Continued)

Dataset	Observable	N_{dat}	N_{cor}	χ_{red}^2 (Z-score)	χ_{cor}^2	χ_{norm}^2	Fitted norm.
DELPHI(uds) [66]	$\sigma^{-1}d\sigma/dz_p$ (uds)	21/23	...	0.89(-0.25)
DELPHI(b) [66]	$\sigma^{-1}d\sigma/dz_p$ (b)	21/23	...	0.77(-0.71)
BABAR (prompt) [57]	$\sigma^{-1}d\sigma/dp$	31/45	1*	0.34(-3.54)	50.71	40.10	0.938(5)
ARGUS [56]	$\sigma^{-1}d\sigma/dp$	25/52	...	1.72(+2.19)	...	11.50	0.939(10)

TABLE XIII. As in Table XII, but for e^+e^- SIA data with K^\pm production.

Dataset	Observable	N_{dat}	N_{cor}	χ_{red}^2 (Z-score)	χ_{cor}^2	χ_{norm}^2	Fitted norm.
Belle [58]	$d\sigma/dz$	70/78	...	0.06(-11.54)	0.852(10)*
TASSO 12 GeV [59]	$d\sigma/dz$	1/3	...	0.26(-0.28)	0.8(2.6)*
TASSO 14 GeV [60]	$d\sigma/dz$	3/9	...	1.00(+0.28)	0.860(131)*
TASSO 22 GeV [60]	$d\sigma/dz$	3/10	...	0.12(-1.61)	1.180(227)*
TASSO 34 GeV [61]	$\sigma^{-1}d\sigma/dz_p$	3/11	...	0.34(-0.84)	...	0.17	0.975(22)
TPC [63]	$\sigma^{-1}d\sigma/dz_p$	7/21	...	2.15(+1.81)
TOPAZ [64]	$\sigma^{-1}d\sigma/d\xi$	2/12	...	0.12(-1.22)
SLD [68]	$\sigma^{-1}d\sigma/dz_p$	35/36	...	0.45(-2.90)	...	0.30	0.995(6)†
SLD(uds) [68]	$\sigma^{-1}d\sigma/dz_p$ (uds)	35/36	...	1.80(+2.81)	0.995(6)†
SLD(c) [68]	$\sigma^{-1}d\sigma/dz_p$ (c)	35/36	...	1.43(+1.68)	0.995(6)†
SLD(b) [68]	$\sigma^{-1}d\sigma/dz_p$ (b)	35/36	...	1.15(+0.68)	0.995(6)†
ALEPH [65]	$\sigma^{-1}d\sigma/dz_p$	19/29	...	0.92(-0.13)	...	0.51	1.021(12)
OPAL [67]	$\sigma^{-1}d\sigma/dp$	10/33	...	0.35(-1.83)
DELPHI [66]	$\sigma^{-1}d\sigma/dz_p$	22/23	...	0.99(+0.06)
DELPHI(uds) [66]	$\sigma^{-1}d\sigma/dz_p$ (uds)	22/23	...	1.15(+0.58)
DELPHI(b) [66]	$\sigma^{-1}d\sigma/dz_p$ (b)	22/23	...	0.61(-1.41)
BABAR (prompt) [57]	$\sigma^{-1}d\sigma/dp$	28/45	1*	0.12(-5.74)	53.77	27.81	0.948(4)
ARGUS [56]	$\sigma^{-1}d\sigma/dp$	25/42	...	0.84(-0.52)	...	1.10	0.981(7)

TABLE XIV. As in Table XII, but for e^+e^- SIA data with unidentified h^\pm hadron production.

Dataset	Observable	N_{dat}	χ_{red}^2 (Z-score)	χ_{norm}^2	Fitted norm.
TASSO 12 GeV [59]	$\sigma^{-1}d\sigma/dz_p$	5/7	0.81(-0.11)	2.75	0.925(22)
TASSO 14 GeV [69]	$\sigma^{-1}d\sigma/dz_p$	11/20	0.92(-0.05)	2.88	0.924(17)
TASSO 22 GeV [69]	$\sigma^{-1}d\sigma/dz_p$	11/20	1.41(+0.99)	2.36	0.931(17)
TASSO 30 GeV [70]	$\sigma^{-1}d\sigma/dz_p$	5/7	1.86(+1.29)	0.05	1.010(18)
TASSO 44 GeV [61]	$\sigma^{-1}d\sigma/dz_p$	11/20	1.59(+1.31)	2.32	0.932(15)
TPC [63]	$\sigma^{-1}d\sigma/dz_p$	12/34	1.65(+1.47)
SLD [68]	$\sigma^{-1}d\sigma/dz_p$	34/40	0.68(-1.39)	1.24	1.011(2)†
SLD(uds) [68]	$\sigma^{-1}d\sigma/dz_p$ (uds)	34/40	0.68(-1.44)	...	1.011(2)†
SLD(c) [68]	$\sigma^{-1}d\sigma/dz_p$ (c)	34/40	0.86(-0.51)	...	1.011(2)†
SLD(b) [68]	$\sigma^{-1}d\sigma/dz_p$ (b)	34/40	0.67(-1.46)	...	1.011(2)†
ALEPH [71]	$\sigma^{-1}d\sigma/dz_p$	37/46	0.44(-3.04)	2.01	1.014(3)
OPAL [72]	$\sigma^{-1}d\sigma/dz$	19/22	0.73(-0.82)
DELPHI [66]	$\sigma^{-1}d\sigma/dz_p$	21/27	0.28(-3.32)
DELPHI(uds) [66]	$\sigma^{-1}d\sigma/dp$	21/27	0.34(-2.82)
DELPHI(b) [66]	$\sigma^{-1}d\sigma/dz_p$	21/27	0.42(-2.37)

- [1] A. W. Thomas, *Phys. Lett.* **126B**, 97 (1983).
- [2] R. S. Towell *et al.*, *Phys. Rev. D* **64**, 052002 (2001).
- [3] J. Dove *et al.*, *Nature (London)* **590**, 561 (2021); **604**, E26 (2022).
- [4] C. Cocuzza, W. Melnitchouk, A. Metz, and N. Sato, *Phys. Rev. D* **104**, 074031 (2021).
- [5] A. I. Signal and A. W. Thomas, *Phys. Lett. B* **191**, 205 (1987).
- [6] X. Ji and J. Tang, *Phys. Lett. B* **362**, 182 (1995).
- [7] W. Melnitchouk and M. Malheiro, *Phys. Rev. C* **55**, 431 (1997).
- [8] A. O. Bazarko *et al.*, *Z. Phys. C* **65**, 189 (1995).
- [9] D. Mason *et al.*, *Phys. Rev. Lett.* **99**, 192001 (2007).
- [10] A. Kayis-Topaksu *et al.*, *New J. Phys.* **13**, 093002 (2011).
- [11] O. Samoylov *et al.*, *Nucl. Phys.* **B876**, 339 (2013).
- [12] N. Kalantarians, C. E. Keppel, and M. E. Christy, *Phys. Rev. C* **96**, 032201 (2017).
- [13] A. Accardi, F. Arleo, W. K. Brooks, D. D’Enterria, and V. Muccifora, *Riv. Nuovo Cimento* **32**, 439 (2009).
- [14] A. Airapetian *et al.*, *Phys. Lett. B* **666**, 446 (2008).
- [15] A. Airapetian *et al.*, *Phys. Rev. D* **89**, 097101 (2014).
- [16] M. Stolarski, *Phys. Rev. D* **92**, 098101 (2015).
- [17] E. Leader, A. V. Sidorov, and D. B. Stamenov, *Phys. Rev. D* **90**, 054026 (2014).
- [18] E. Leader, A. V. Sidorov, and D. B. Stamenov, *Phys. Rev. D* **93**, 074026 (2016).
- [19] S. Chatrchyan *et al.*, *J. High Energy Phys.* 04 (2011) 050.
- [20] S. Chatrchyan *et al.*, *Phys. Rev. Lett.* **109**, 111806 (2012).
- [21] S. Chatrchyan *et al.*, *Phys. Rev. D* **90**, 032004 (2014).
- [22] M. Aaboud *et al.*, *Eur. Phys. J. C* **77**, 367 (2017).
- [23] V. Khachatryan *et al.*, *Eur. Phys. J. C* **76**, 469 (2016).
- [24] G. Aad *et al.*, *Phys. Rev. Lett.* **109**, 012001 (2012).
- [25] S. Chatrchyan *et al.*, *J. High Energy Phys.* 02 (2014) 013.
- [26] G. Aad *et al.*, *J. High Energy Phys.* 05 (2014) 068.
- [27] A. M. Sirunyan *et al.*, *Eur. Phys. J. C* **79**, 269 (2019).
- [28] M. Aaboud *et al.*, *Eur. Phys. J. C* **78**, 110 (2018); **78**, 898(E) (2018).
- [29] S. Alekhin, J. Blümlein, and S. Moch, *Phys. Lett. B* **777**, 134 (2018).
- [30] I. Borsa, R. Sassot, and M. Stratmann, *Phys. Rev. D* **96**, 094020 (2017).
- [31] D. De Florian, G. A. Lucero, R. Sassot, M. Stratmann, and W. Vogelsang, *Phys. Rev. D* **100**, 114027 (2019).
- [32] J. M. Campbell and R. K. Ellis, *Phys. Rev. D* **60**, 113006 (1999).
- [33] Y. L. Dokshitzer, *Sov. Phys. JETP* **46**, 641 (1977).
- [34] V. N. Gribov and L. N. Lipatov, *Sov. J. Nucl. Phys.* **15**, 438 (1972).
- [35] G. Altarelli and G. Parisi, *Nucl. Phys.* **B126**, 298 (1977).
- [36] S. Navas *et al.* (Particle Data Group), *Phys. Rev. D* **110**, 030001 (2024).
- [37] D. Adamiak, T. Anderson, C. Cocuzza, W. Melnitchouk, N. Sato, and Y. Zhou (to be published).
- [38] E. Moffat, W. Melnitchouk, T. C. Rogers, and N. Sato, *Phys. Rev. D* **104**, 016015 (2021).
- [39] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock, and A. Bodek, *Phys. Lett. B* **282**, 475 (1992).
- [40] A. C. Benvenuti *et al.*, *Phys. Lett. B* **237**, 592 (1990).
- [41] M. Arneodo *et al.*, *Nucl. Phys.* **B483**, 3 (1997).
- [42] M. Arneodo *et al.*, *Nucl. Phys.* **B487**, 3 (1997).
- [43] H. Abramowicz *et al.*, *Eur. Phys. J. C* **75**, 580 (2015).
- [44] C. Adolph *et al.*, *Phys. Lett. B* **764**, 1 (2017).
- [45] C. Adolph *et al.*, *Phys. Lett. B* **767**, 133 (2017).
- [46] R. Aaij *et al.*, *J. High Energy Phys.* 12 (2014) 079.
- [47] R. Aaij *et al.*, *J. High Energy Phys.* 01 (2016) 155.
- [48] V. M. Abazov *et al.*, *Phys. Rev. Lett.* **112**, 151803 (2014); **114**, 049901(E) (2015).
- [49] T. Aaltonen *et al.*, *Phys. Rev. Lett.* **102**, 181801 (2009).
- [50] J. Adam *et al.*, *Phys. Rev. D* **103**, 012001 (2021).
- [51] V. M. Abazov *et al.*, *Phys. Rev. D* **76**, 012003 (2007).
- [52] T. A. Aaltonen *et al.*, *Phys. Lett. B* **692**, 232 (2010).
- [53] V. M. Abazov *et al.*, *Phys. Rev. Lett.* **101**, 062001 (2008).
- [54] A. Abulencia *et al.*, *Phys. Rev. D* **75**, 092006 (2007).
- [55] B. I. Abelev *et al.*, *Phys. Rev. Lett.* **97**, 252001 (2006).
- [56] H. Albrecht *et al.*, *Z. Phys. C* **44**, 547 (1989).
- [57] J. P. Lees *et al.*, *Phys. Rev. D* **88**, 032011 (2013).
- [58] M. Leitgab *et al.*, *Phys. Rev. Lett.* **111**, 062002 (2013).
- [59] R. Brandelik *et al.*, *Phys. Lett.* **100B**, 357363 (1981).
- [60] M. Althoff *et al.*, *Z. Phys. C* **22**, 307 (1984).
- [61] W. Braunschweig *et al.*, *Z. Phys. C* **42**, 189 (1989).
- [62] X.-Q. Lu, Heavy quark jets from e^+e^- annihilation at 29 GeV, thesis, Johns Hopkins University, 1986.
- [63] H. Aihara *et al.*, *Phys. Rev. Lett.* **61**, 1263 (1988).
- [64] R. Itoh *et al.*, *Phys. Lett. B* **345**, 335 (1995).
- [65] D. Buskulic *et al.*, *Z. Phys. C* **66**, 355 (1995).
- [66] P. Abreu *et al.*, *Eur. Phys. J. C* **5**, 585 (1998).
- [67] R. Akers *et al.*, *Z. Phys. C* **63**, 181 (1994).
- [68] K. Abe *et al.*, *Phys. Rev. D* **69**, 072003 (2004).
- [69] M. Althoff *et al.*, *Z. Phys. C* **17**, 5 (1983).
- [70] R. Brandelik *et al.*, *Phys. Lett. B* **100**, 357 (1981).
- [71] D. Buskulic *et al.*, *Phys. Lett. B* **357**, 487 (1995); **364**, 247 (E) (1995).
- [72] K. Ackerstaff *et al.*, *Eur. Phys. J. C* **7**, 369 (1999).
- [73] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk, *Phys. Rev. D* **101**, 074020 (2020).
- [74] C. Cocuzza, C. E. Keppel, H. Liu, W. Melnitchouk, A. Metz, N. Sato, and A. W. Thomas, *Phys. Rev. Lett.* **127**, 242001 (2021).
- [75] R. Hyndman and Y. Fan, *Am. Stat.* **50**, 361 (1996).
- [76] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Eur. Phys. J. C* **75**, 204 (2015).
- [77] M. Boggione, A. Dotson, L. Gamberg, S. Gordon, J. O. Gonzalez-Hernandez, A. Prokudin, T. C. Rogers, and N. Sato, *J. High Energy Phys.* 10 (2019) 122.
- [78] H. Bhatt *et al.*, *Phys. Lett. B* **865**, 139485 (2025).
- [79] R. Abdul Khalek, V. Bertone, A. Khoukli, and E. R. Nocera, *Phys. Lett. B* **834**, 137456 (2022).
- [80] R. D. Ball *et al.*, *Eur. Phys. J. C* **82**, 428 (2022).
- [81] T.-J. Hou *et al.*, *Phys. Rev. D* **103**, 014013 (2021).
- [82] A. Accardi, T. Hobbs, and W. Melnitchouk, *J. High Energy Phys.* 11 (2009) 084.
- [83] J. V. Guerrero, J. J. Ethier, A. Accardi, S. W. Casper, and W. Melnitchouk, *J. High Energy Phys.* 09 (2015) 169.
- [84] J. V. Guerrero and A. Accardi, *Phys. Rev. D* **97**, 114012 (2018).

- [85] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo, and M. Ubiali, *Eur. Phys. J. C* **80**, 1168 (2020).
- [86] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *Eur. Phys. J. C* **81**, 341 (2021).
- [87] M. Hirai, S. Kumano, T.-H. Nagai, and K. Sudoh, *Phys. Rev. D* **75**, 094009 (2007).
- [88] G. Aad *et al.*, *Eur. Phys. J. C* **82**, 438 (2022).
- [89] <https://www.lhapdf.org/>.