Neural network QCD analysis of charged hadron fragmentation functions in the presence of SIDIS data

Maryam Soleymaninia,[1,*] Hadi Hashamipour,[1,†] and Hamzeh Khanpour[2,1,3,‡]

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

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

[3]*Department of Theoretical Physics, Maynooth University, Maynooth, County Kildare W23 F2H6, Ireland*

(Received 24 February 2022; accepted 24 May 2022; published 15 June 2022)

In this paper, we present a QCD analysis to extract the fragmentation functions (FFs) of unidentified light charged hadron entitled as `SHK22.h` from high-energy lepton-lepton annihilation and lepton-hadron scattering datasets. This analysis includes the data from all available single inclusive electron-positron annihilation processes and semi-inclusive deep-inelastic scattering (SIDIS) measurements for the unidentified light charged hadron productions. The SIDIS data that has been measured by the COMPASS experiment could allow the flavor dependence of the FFs to be well constrained. We exploit the analytic derivative of the neural network for fitting of FFs at next-to-leading-order (NLO) accuracy in the perturbative QCD. The Monte Carlo method is implied for all sources of experimental uncertainties and the parton distribution functions as well. Very good agreements are achieved between the `SHK22.h` FFs set and the most recent QCD fits available in literature, namely, `JAM20` and `NNFF1.1h`. In addition, we discuss the impact arising from the inclusion of SIDIS data on the extracted light-charged hadron FFs. The global QCD resulting at NLO for charged hadron FFs provides valuable insights for applications in present and future high-energy measurement of charged hadron final state processes.

DOI: 10.1103/PhysRevD.105.114018

I. INTRODUCTION

In perturbative quantum chromodynamics (pQCD), the hard-scattering processes in which a hadron is observed in the final state, include an integral part in which to be called fragmentation functions (FFs) in the theoretical framework. They are process independent and universal quantities, and they show a nonperturbative transition of a parton into a hadron. FFs depend on the fraction of the longitudinal momentum of the parton taken by the hadron and the scale of energy [1]. FFs have a critical role in the current experimental programs at Jefferson Lab, Future Electron-Ion Collider (EIC) [2,3], Future Circular Collider (FCC) [4,5], and the LHC, and this aspect of their role leads to the main motivation for studying the collinear FFs in several phenomenological studies.

Since FFs are nonperturbative quantities, they need to be determined from a QCD analysis of the corresponding experimental datasets. The core experimental datasets are the single-inclusive electron-positron annihilation (SIA) from several collaborations and at different ranges of center of mass energy from 10.5 GeV up to the M_Z [6–13]. In order to disentangle all the different flavors of FFs for quarks and antiquarks, in addition to the SIA data sample, one needs to take into account some other observables. Hence, the determination of FFs in the global QCD analyses also include the data on semi-inclusive deep-inelastic scattering (SIDIS) processes [14–17] and the single-inclusive hadron production in proton-proton collisions [18–22].

Several theoretical analyses have been exploited on SIA, SIDIS, and pp collisions datasets in QCD analysis to constrain the FFs of identified light charged hadrons [23–28], unidentified light charged hadrons [25,29,30], and heavy hadrons [31–33].

There are two methods to calculate the FFs of unidentified light charged hadrons. In the first method, for every flavor, they can be calculated as a sum of the FF sets of all identified light charged hadrons produced in the fragmentation of the given parton. Alternatively, the FFs of unidentified charged hadrons are implemented

*Maryam_Soleymaninia@ipm.ir
†H_Hashamipour@ipm.ir
‡Hamzeh.Khanpour@cern.ch

independently from a QCD analysis included the unidentified charged hadron experimental data directly.

Before discussing our analysis, we will first review the FF sets of unidentified charged hadrons which have been recently calculated. In our recent analysis entitled SGKS20 [27], we implemented the FFs of unidentified charged hadrons up to next-to-next-to-leading order (NNLO) by taking advantage of the first method and determined in a simultaneous fit the FF sets for pion, kaon, proton, and the residual light charged hadrons. All the available SIA data for pion, kaon, and proton and unidentified charged hadrons production have been considered in this analysis.

Another recent analysis for light charged hadron has been done by the JAM Collaboration [25] up to the next-to-leading-order (NLO). Their analysis includes all available SIDIS and SIA data for pion, kaon, and unidentified charged hadrons to calculate the FFs of pion, kaon, and charged hadrons. In addition, they have used data from inclusive DIS and Drell-Yan lepton-pair production to calculate the parton distribution functions (PDFs) simultaneously with the FFs. Accounting for unidentified charged hadrons, they have used the first method and add a fitted residual correction to the sum.

The other analysis of unidentified charged-hadron FFs has been presented by the NNPDF Collaboration [30]. They have utilized the second method to calculate the FFs up to the NLO accuracy. The proton-proton data for unidentified charge hadron production has been added by means of Bayesian reweighting to the analysis based only on SIA datasets. They have tried to complement their analysis of Ref. [34] with the measurements of the charged hadron spectra in pp collisions. Their study demonstrated that the inclusion of pp data in a FF fit could provide a stringent constraint on the gluon distribution FF.

Another analysis of unidentified charged hadrons, based on the second method, has been done by SGK18 in which the FFs of charged hadrons have been done up to NNLO by including all the unidentified charged hadrons from SIA experimental datasets [29].

The main aim of this paper is to revisit our previous QCD analysis in Ref. [29] to implement a global QCD analysis for FFs of charged hadrons by adding the SIDIS datasets to the data sample, and applying the neural network (NN) technique. In this analysis, the hadron FFs are fitted directly from all the experimental data for unidentified light charged hadrons production from SIA and SIDIS processes. Our main goal in this study is the inclusion of COMPASS SIDIS experimental data [14] as the only dataset for the charged hadron production from the SIDIS process.

In recent years, machine learning (ML) has spread through all subjects of particle physics, specially collider physics. One of the encouraging areas of application of such methods is improving our knowledge of nonperturbative quantities of nucleons such as PDFs and FFs [23,28,35]. In light of this fact, we decided to use such a method based on the artificial neural networks to extract the light charge hadron FFs from QCD analysis of the corresponding datasets. Modern optimization techniques are utilized in this project to minimize the bias of FFs parametrization by taking advantage of the neural network and also the Monte Carlo sampling method as a proper statistical treatment of experimental data uncertainties to obtain the probability density distribution from the data. For this purpose, we use the publicly available code MontBlanc in this analysis which can be obtained from [36].

This code is devoted to the extraction of collinear distributions of fragmentation functions. The code is an open-source package that provides a framework for the determination of the FFs, for many different kinds of analyses in QCD. So far, it has been developed to determine the FFs of the pion from experimental data for SIA and SIDIS datasets [23], and in our most recent study to determine the fragmentation functions of $\Xi^-/\bar{\Xi}^+$ [37]. MontBlanc can analyze the SIA data up to NNLO and the SIDIS data up to NLO in perturbation theory. The framework in this code is a combination of the Monte Carlo method to map the uncertainty distributions of FFs and neural networks to parametrize the FFs.

The structure of the paper is as follows: In Sec. II, we review the theoretical formalism for the inclusive hadron production in electron-positron annihilation and semi-inclusive deep-inelastic scattering process, and the timelike evolution equation. The parametrization of FFs in terms of neural network are also discussed in detail in this section. Section III includes our fitting methodology. We also illustrate the Monte Carlo methodology adopted in our analysis to calculate the uncertainties of FFs and the optimal fit. This section also summarizes the SIA and SIDIS experimental datasets analyzed in this study, and the possible tensions between the datasets are also examined. The main results of SHK22.h are presented in Sec. IV. This section includes the SHK22.h fit quality and the numerical results for the differential cross sections, and detailed comparison of theory predictions with the analyzed experimental datasets. We present the SHK22.h light charged hadron FFs and detailed comparisons with other results available in the literature, namely, JAM20 and NNFF1.1h FFs. We also discuss in this section the impact arising from the inclusion of SIDIS data on the extracted light-charged hadron FFs. Finally, we summarize our conclusions in Sec. V and outline possible future developments.

II. THEORETICAL SETUP

In this section, we present the theoretical backgrounds of the standard collinear factorization and discuss the perturbative and nonperturbative parts of the cross-section measurements in the SIA and multiplicities in SIDIS processes. Then, we present the timelike evolution equations and the splitting functions used for the FFs. Finally, we discuss the parametrization of the FFs in terms of the

neural networks in the presence of the SIDIS data. The neural network architecture for all the fitted FFs and the Monte Carlo uncertainty propagation also will be discussed in this section.

A. SIA and SIDIS factorization

In the standard collinear factorization, we separate the QCD cross sections into the perturbative partonic hard factors which are convoluted with the nonperturbative partonic or hadronic distribution functions.

In the present analysis, we consider the SIA process which is given by

$$e^+ + e^- \to (\gamma, Z^0) \to h^\pm + X, \tag{1}$$

and the semi-inclusive charged hadron production in the lepton-nucleon deep inelastic scattering that can be written as

$$\ell + N \to \ell + h^+/h^- + X. \tag{2}$$

According to the collinear factorization theorem, the cross sections for the two processes mentioned above can be written as

$$\begin{aligned}\sigma^{\mathrm{SIA}} &= \hat{\sigma} \otimes \mathrm{FF},\\ \sigma^{\mathrm{SIDIS}} &= \hat{\sigma} \otimes \mathrm{PDF} \otimes \mathrm{FF},\end{aligned} \tag{3}$$

where the $\hat{\sigma}$ indicates the process dependent perturbative partonic cross section. The parton distribution function and FF are nonperturbation functions.

The details of the computation of the SIA cross sections are provided in some studies available in the literature, and we refer the reader to the Refs. [28,29] for a clear review.

The basic cross sections for the charged hadron production with four-momentum p_h in deep inelastic scattering of a lepton with momentum l from a nucleon with momentum p can be written as

$$\begin{aligned}\frac{d\sigma^h}{dxdydz_h} = \frac{2\pi\alpha^2}{Q^2}\Bigg[&\frac{(1+(1-y)^2)}{y} 2F_1^h(x,z_h,Q^2)\\ &+ \frac{2(1-y)}{y} F_L^h(x,z_h,Q^2)\Bigg],\end{aligned} \tag{4}$$

which are functions of the Bjorken scaling variable $x = \frac{Q^2}{2p.q}$, the charged hadrons fragmentation scaling variable $z_h = \frac{p_h.p}{q.p}$, energy transfer or inelasticity $y = \frac{Q^2}{xs}$, and the four momentum transfer squared of the virtual photon $Q^2 = -q^2$. In this equation, the α indicates the fine-structure constant.

The structure functions F_1^h and F_L^h in Eq. (4) are the relevant inclusive DIS structure functions in which at NLO accuracy is given by

$$\begin{aligned}&F_1^h(x,z_h,Q^2)\\ &= \frac{1}{2}\sum_{q,\bar{q}} e_q^2 \Bigg\{ q(x,Q^2) D_q^h(z_h,Q^2)\\ &+ \frac{\alpha_s(Q^2)}{2\pi}[q \otimes C_{qq}^1 \otimes D_q^h + q \otimes C_{gq}^1 \otimes D_g^h\\ &+ g \otimes C_{qg}^1 \otimes D_q^h](x,z_h,Q^2)\Bigg\},\end{aligned} \tag{5}$$

$$\begin{aligned}F_L^h(x,z_h,Q^2) = \frac{\alpha_s(Q^2)}{2\pi}\sum_{q,\bar{q}} e_q^2 [&q \otimes C_{qq}^L \otimes D_q^h\\ &+ q \otimes C_{gq}^L \otimes D_g^h\\ &+ g \otimes C_{qg}^L \otimes D_q^h](x,z_h,Q^2).\end{aligned} \tag{6}$$

The convolution symbol $\otimes$ in the equations above is defined as

$$\begin{aligned}&q(x) \otimes C(x,z_h) \otimes D^h(z_h)\\ &= \int_x^1 \frac{dx'}{x'} \int_{z_h}^1 \frac{dz_h'}{z_h'} q\left(\frac{x}{x'}\right) c(x',z_h') D\left(\frac{z_h}{z_h'}\right).\end{aligned} \tag{7}$$

In Eqs. (5) and (6), the PDFs inside the nucleon are denoted by q, $\bar{q}$, and g, and D_q^h, $D_{\bar{q}}$, and D_g denote the FFs. The hard scattering coefficient functions $C_{ij}^{1,L}$ related to the F_1^h and F_L^h structure functions admit the usual perturbative expansion. Currently, these coefficient functions are known up to $\mathcal{O}(\alpha_s)$, i.e., NLO and can be found, for example, in Refs. [38,39]. Although the coefficient functions for structure function in the SIA process are known up to the $\mathcal{O}(\alpha_s^2)$, i.e., NNLO, the full set of NNLO accuracy corrections are not known for the SIDIS process, and the structure functions are only known up to NLO. Hence, our QCD calculations in the perturbative part are limited to the $\mathcal{O}(\alpha_s)$, i.e., NLO, accuracy.

It should be noted here that we determine the charged hadron FFs in the zero-mass variable-flavor-number scheme (ZM-VFNS) in which all the active flavors are considered to be massless. However, the masses of heavy quarks require to be introduced during the subschemes to determine the number of active flavors based on the heavy-quark thresholds. In this analysis, the charm and bottom masses are considered to be fixed at $m_c = 1.51$ GeV and $m_b = 4.92$ GeV, respectively.

As a final point, we should highlight here that we have used the proton PDF set `NNPDF31` [35] at NLO accuracy to calculate the cross section in the SIDIS process. The SIDIS data included in our QCD fit, have been measured by the COMPASS Collaboration in which the muon beam collides with the lithium (^{6}LiD) target. In our present study, we focus on the analysis with the proton PDFs without considering the nuclear corrections. We plan to revisit this analysis in the near future to study the impact of such a

nuclear effect along with the target mass corrections (TMC) and hadron mass corrections.

B. pQCD and timelike evolution

The integrated FFs are universal and process independent quantities in the sense that $D^h(z,Q)$ is the same in processes like e^+e^- annihilation, SIDIS, and hadronic collisions. FFs depend on an additional parameter called the renormalization scale μ because of the QCD dynamics. Based on the pQCD approach, the structure of the evolution equations for the unpolarized integrated FFs generally is given by

$$\frac{dD_i^h(z,\mu^2)}{d\ln\mu^2}=\frac{\alpha_s(\mu^2)}{2\pi}\sum_j\int_z^1\frac{dm}{m}P_{ji}(m,\alpha_s(\mu^2))D_j^h\left(\frac{z}{m},\mu^2\right), \tag{8}$$

where P_{ji} is the matrix for the timelike splitting functions and have a perturbative expansion of the form,

$$P_{ji}(m,\alpha_s(\mu^2))=P_{ji}^{(0)}(m)+\frac{\alpha_s(\mu^2)}{2\pi}P_{ji}^{(1)}(m)+\left(\frac{\alpha_s(\mu^2)}{2\pi}\right)^2P_{ji}^{(2)}(m)+\ldots. \tag{9}$$

The NLO timelike splitting functions P_{ji} have been computed in Refs. [40,41]. Usually the evolution equation is decomposed into a singlet sector comprising the gluon and the sum of all quark and antiquark FFs, and the nonsinglet sector for quark-antiquark and flavor differences.

The range of applicability for the FFs is limited to the medium-to-large range of z value. There are two reasons for such limitation. First, the strong singular behavior in the timelike splitting functions when $z\to 0$. Second, the produced hadrons in the final state are considered to be massless. Supplementary to this, the timelike splitting functions have a logarithmic piece $\simeq\ln^2 z/z$ in the NLO part which leads to negative FFs for $z\ll 1$ in the influence of the Q^2 evolution and it leads to unphysical, negative cross sections. In addition, at small z, the finite mass corrections become more important. Therefore, in this global QCD analysis we limit ourselves to the kinematic regions in which mass corrections and the singularity of small-z behavior in the evolution kernels are negligible, as discussed in Sec. III B.

C. Neural network and flavor decomposition

As we discussed in Ref. [29], inclusive SIA data allow for the determination of only the summed quark and antiquark FFs by including the total inclusive, light-, charm-, and bottom-quark tagged cross sections

$$D_{u^+}^{h^\pm},\quad D_{d^++s^+}^{h^\pm},\quad D_{c^+}^{h^\pm},\quad D_{b^+}^{h^\pm},\quad D_g^{h^\pm}. \tag{10}$$

It has been shown that adding the SIDIS datasets to the data sample could provide a direct constraint on the individual q and $\bar{q}$ FFs for light quarks. We adopt the following parametrization basis for h^+,

$$D_u^{h^+},\quad D_{\bar{u}}^{h^+},\quad D_{d+s}^{h^+},\quad D_{\bar{d}+\bar{s}}^{h^+},\quad D_{c^+}^{h^+},\quad D_{b^+}^{h^+},\quad D_g^{h^+}. \tag{11}$$

Taking into account the SIDIS data in the QCD fit, the above combinations of quark FFs $D_{u^+}^{h^\pm}$ and $D_{d^++s^+}^{h^\pm}$ in Eq. (10) are considered to be decomposed. The heavy distributions are assumed to be symmetric. It reads,

$$D_q^{h^+}=D_{\bar{q}}^{h^+},\qquad q=c,b. \tag{12}$$

Hence, by adding SIDIS COMPASS data, the number of independent distributions increases to the seven. We observed that under these flavor combinations of quark FFs for hadron production, the best fit quality and accuracy can be achieved. In order to choose the best parametrization basis, we study different scenarios. However, in the most general case, we disentangle all light flavors. The datasets included cannot constrain well enough all 6 light, 2 heavy quarks, and gluon FFs. In another case, we assume symmetry just between $D_d^{h^+}=D_s^{h^+}$ and a separate parametrization are implied for the $D_{\bar{d}}^{h^+}$ and $D_{\bar{s}}^{h^+}$. In particular, we find this assumption leads to a deterioration of the quality of FFs and the fit, and then we omitted such assumptions.

We finally note that the COMPASS measurements have been reported for both positive (h^+) and negative (h^-) charged hadrons separately. We consider a relation based on charge conjugation between h^+ and h^- as follows:

$$D_{q(\bar{q})}^{h^-}(z,Q)=D_{\bar{q}(q)}^{h^+}(z,Q),\qquad D_g^{h^-}(z,Q)=D_g^{h^+}(z,Q). \tag{13}$$

Hence, one can obtain the h^- FFs in terms of the h^+ FFs.

The h^+ FFs for every parton flavor in terms of a neural network is defined at the initial scale of $Q_0=5$ GeV, which is given by

$$zD_i^{h^+}(z,Q_0)=(N_i(z,\theta)-N_i(1,\theta))^2. \tag{14}$$

Here $N_i(z,\theta)$ denotes the output of the neural network and θ stands for the (internal) parameters of the neural network. Note that the result of the neural network at $z=1$ is subtracted in order to satisfy the requirement that FFs should vanish at this point. Also, the result is squared to make sure that FFs always stay positive. A few remarks on the construction of neural networks are in order. First, in

this analysis we use a simple yet efficient[1] neural network structure which has only one hidden layer. Second, we choose to use 20 neurons (nodes) in the hidden layer; admittedly this number is somewhat arbitrary, smaller number of nodes may result in an equally accurate fit. In an effort to examine the effect of choosing an alternative NN architecture on the obtained results, we performed another analysis with different configurations, i.e., {1-9-9-7} architecture. Our examination shows that the results are basically unchanged, which is in agreement with other studies available in the literature [43]. We believe that this stability despite changing the NN shows that our FFs are not driven by hyperparameters such as the number of hidden layers or the number of nodes in each layer but by input experimental data. Third, the number of replicas in this analysis is 200, regardless, requirements of the replica method can be achieved by a smaller number of replicas, for example, 100 [23].

III. FITTING PROCEDURE

In this section, we first discuss the minimization strategy and the uncertainty propagation estimation using the Monte Carlo method. Then we illustrate a comprehensive set of measurements of the charged hadron production in electron-positron annihilation SIA, and the lepton-nucleon SIDIS processes as well. We also discuss the kinematic cuts on the experimental data in which a description in the framework of pQCD can be expected to work well. Finally, we comment on the tension between COMPASS data with some of other SIA datasets analyzed in the `SHK22.h` study.

A. Minimization and uncertainty propagation method

It goes without saying that any measurement in high-energy particle physics has an uncertainty associated with it, this introduces the problem of understanding the effect that this produces on other quantities referred to as uncertainty propagation. Two widely used methods to propagate the experimental uncertainties to the FFs or observables are (i) the Hessian method and (ii) the Monte Carlo (MC) or replica method. The Monte Carlo approach is nowadays widely used in various QCD analyses [35,44–47]. This method estimates the parameters posterior probability distribution by performing a number of fits. Every fit is independent and performed on a pseudo dataset (a replica) resulting in an optimal set. The results of all fits performed then *learn* the probability distribution which defines both the central value (mean of the probability distribution) and the uncertainties of FFs (standard deviation of the probability distribution). In order to properly account for the uncertainty of the PDFs used in SIDIS observables, we use a similar method as developed and adopted in Ref. [23] to ensure that the PDF uncertainty is propagated into FFs. In SIDIS calculations each time a different proton replica of `NNPDF3.1` is chosen at random from the `NNPDF31_nlo_pch_as_0118` set.

In order to perform the QCD analysis, one mainly applies the maximum log-likelihood method which in turn reduces to minimum χ^2 under the usual assumptions. In this case, the problem of finding the optimal parameters of a parametric form or optimal neural network parameter is equivalent to minimizing the χ^2 function at hand. There are a few ways to minimize a χ^2 function in a QCD fit which is based on the neural network; one that readily comes to mind is explicit differentiation and calculation of the global minimum directly, this approach is inefficient and sometimes straight impossible. It is therefore natural to look for numerical methods such as the genetic algorithm used by `NNPDF` for PDFs [48], stochastic gradient descent methods used by `nNNPDF` for nuclear PDFs [49], and trust-region methods as provided by `Ceres Solver` [50] and utilized by `MAPFF` [23] and `SHKS22` [37]. In this analysis, we adopt the later method that is implemented in the `MontBlanc` package [36]. The χ^2 function that is subject to minimization is defined as follows:

$$\chi^{2(k)} \equiv (\mathbf{T}(\boldsymbol{\theta}^{(k)}) - \boldsymbol{x}^{(k)})^{\mathrm{T}} \cdot \mathbf{C}^{-1} \cdot (\mathbf{T}(\boldsymbol{\theta}^{(k)}) - \boldsymbol{x}^{(k)}). \quad (15)$$

In the equation above, the $\boldsymbol{x}^{(k)}$ is the kth replica, $\mathbf{T}(\boldsymbol{\theta}^{(k)})$ is the theoretical prediction for the kth replica based on the parameters of neural network ($\boldsymbol{\theta}^{(k)}$), and $\mathbf{C}$ is the covariance matrix of the data that contains all information on the uncertainties and correlations. In view of the fact that neural networks by construction are redundant, i.e., the number of parameters is typically much bigger than that of a functional form of parametrization. For this reason the χ^2 is by convention normalized to N_{dat}, the number of data points.

B. Dataset selection

In the present analysis, we make use of all available experimental data on the charged hadron production in SIA and SIDIS processes to determine the unidentified light-charge hadron FFs. In our previous analyses [27,29,51], we have included all available SIA experimental datasets to determine the FFs of charged hadron production. The SIA measurements are reported as a sum of the observables for the positive and negative charged hadron production. However, the observables in the SIDIS process are separated into positive and negative charged hadron production.

The kinematic coverage in the (z, Q) plane of the SIA and SIDIS datasets analyzed in `SHK22.h` analysis are shown in Fig. 1. The data points for SIA are shown as blue, the SIDIS data points are shown as green; and the gray points are excluded by kinematic cuts as discussed in the text. They datasets contain all analyzed flavor-untagged

[1]This happens because the universal approximation theorem [42] states that a simple feed forward neural network such as one used here can represent any function in a specific interval.

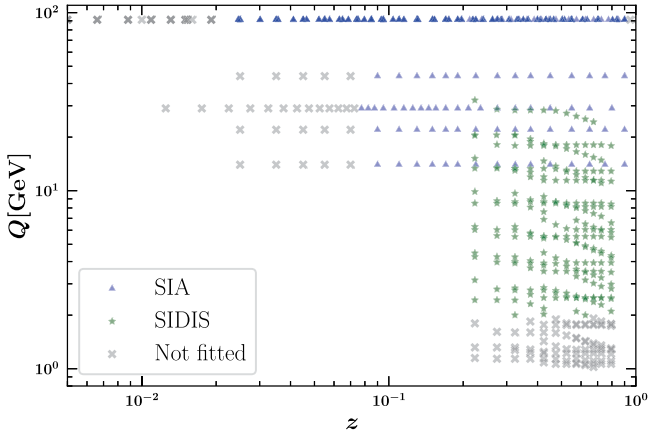

FIG. 1. Kinematic coverage in the (z, Q) plane of the SIA and SIDIS datasets analyzed in SHK22.h study. The data points for SIA are shown as blue, the SIDIS data points are shown as green; and the gray points are excluded by kinematic cuts as discussed in the text.

and tagged measurements which are reported by different experiments. These datasets include the TASSO [8] experiment at DESY; the TPC [9] experiment at SLAC, ALEPH [10], DELPHI [11], and OPAL [12] experiments at CERN and SLD [13] experiments at SLAC.

As one can see from Table I, there are several different measured observables available for these datasets. The experimental collaborations have reported total inclusive and light-, charm-, and bottom-tagged cross sections. Determination of the separate FFs for light and heavy quark flavors is provided by the light and heavy flavor tagged measurements. For a detailed discussion of the SIA datasets, we refer the reader to our previous study on the light-charged hadron FFs [29].

One of our main aims and motivations in this analysis is to revisit our previous QCD analysis [29] by including the available SIDIS data to the SIA data sample. The COMPASS Collaboration has measured the multiplicities of the charged hadrons produced in semi-inclusive scattering. They have used a 160 GeV muon beam and a target (^{6}LiD). COMPASS has measured the differential multiplicity for positive and negative charged hadrons separately, which is given by

$$\frac{dM^h(x,z,Q^2)}{dz} = \frac{d^3\sigma^h(x,z,Q^2)/dxdQ^2dz}{d^2\sigma^{DIS}(x,Q^2)/dxdQ^2},$$

where the numerator is given by the differential SIDIS cross section for charged hadron production and the denominator is given by the differential inclusive DIS cross section. The cross sections at leading order (LO) can be expressed in terms of PDFs $q(x,Q^2)$ and FFs $D_q^h(z,Q^2)$,

$$\frac{dM^h(x,z,Q^2)}{dz} = \frac{\sum_q e_q^2 q(x,Q^2) D_q^h(z,Q^2)}{\sum_q e_q^2 q(x,Q^2)}.$$

The kinematic cuts have been imposed on the photon virtuality $Q^2 > 1$ (GeV/c)^2, on the Bjorken scaling variable $0.004 < x < 0.4$, on the scale variable in the final state $0.2 < z < 0.85$ and on the inelasticity $0.1 < y < 0.7$.

Although the cleanest way to access the FFs for hadron production in the final state is an electron-positron annihilation process and also the FFs are the only nonperturbative objects in the observables, SIA has several limitations which can be addressed by the SIDIS process. On one hand, while the extraction of flavor-separated FFs is difficult in a QCD analysis based on the SIA only, the data from SIDIS experiments are crucial to getting a direct constrain on the separation of quark and anti-quark FFs. On the other hand, the range of the center-of-mass energy at which FFs are probed in SIA covers from $Q = 10$ GeV to $Q = M_Z$. However, SIDIS data cover lower scales of energy, from $Q \sim 1$ to $Q \sim 6$ GeV.

Considering the discussions presented in Sec. II B, we apply kinematic cuts on the experimental data for which a description in the framework of pQCD and timelike evolution can be expected to work well. Hence, we exclude the range of very small values of z from SIA datasets. For the SIA data points, we use the kinematic cuts on z as $0.02 \leq z \leq 0.9$ for data at a center-of-mass energy of M_Z, and $0.075 \leq z \leq 0.9$ for other data points. As a matter of fact, at low- energy scale Q, higher order perturbative corrections are necessary to have acceptable theory predictions. So the perturbative QCD corrections up to NLO accuracy are unreliable for low Q. Hence, we exclude the range of very small values of Q from SIDIS datasets. For the COMPASS SIDIS, we implemented cuts of $Q > 2$ GeV.

Finally, we include in total $N_{\rm dat} = 684$ data points in our analysis after kinematic cuts which include the $N_{\rm dat} = 314$ data points for SIDIS, and $N_{\rm dat} = 370$ for the SIA.

C. Compatibility of TASSO 35 GeV

In this section, we comment on the possible tensions between the SIA datasets. As one can see from datasets reported in Table I, one of the sources of SIA datasets is the TASSO 35 GeV which we do not include in the list of the datasets. Our detailed study on the individual χ^2 shows that there is a tension between this dataset and the COMPASS datasets. We first perform the calculation of the light-charged hadron FFs using the SIA experimental data only, and we achieve an acceptable description for the TASSO dataset at 35 GeV with the individual χ^2 per data point of 1.56. However, when the COMPASS SIDIS data are added to the QCD fit, one cannot obtain an optimal description of the TASSO 35 GeV dataset in the fit, and a large χ^2 per data point is achieved for it, $\chi^2/N_{\rm dat} = 8.37$. The origin of this behavior can be related to a tension between the TASSO 35 GeV and the COMPASS data.

We should note here that the drop of matching between the theoretical predictions and all other TASSO

TABLE I. The χ^2 values per data point for the individual datasets, total SIA, and total SIDIS included in the SHK22.h analysis. The number of data points $N_{\rm dat}$ after the kinematic cuts and the global χ^2 values are also displayed.

Experiment	$\chi^2/N_{\rm dat}$	$N_{\rm dat}$
TASSO 14 GeV $h^{\pm}$	1.791	14
TASSO 22 GeV $h^{\pm}$	1.254	14
TASSO 44 GeV $h^{\pm}$	2.912	14
TPC $h^{\pm}$	0.659	21
ALEPH $h^{\pm}$	0.825	32
DELPHI total $h^{\pm}$	0.610	21
DELPHI *uds* $h^{\pm}$	0.380	21
DELPHI bottom $h^{\pm}$	1.028	21
OPAL total $h^{\pm}$	1.821	19
OPAL *uds* $h^{\pm}$	0.794	19
OPAL charm $h^{\pm}$	0.599	19
OPAL bottom $h^{\pm}$	0.299	19
SLD total $h^{\pm}$	1.047	34
SLD *uds* $h^{\pm}$	0.946	34
SLD charm $h^{\pm}$	1.034	34
SLD bottom $h^{\pm}$	1.102	34
COMPASS h^{-}	0.907	157
COMPASS h^{+}	1.338	157
Global dataset	**1.079**	**684**

experimental data are also examined, and have been seen for all other scales of energy 14, 22, 35, and 44 GeV. As a matter of fact, the extracted values of χ^2 per data points for other data points are also worsened, which in order of 2.3 for the TASSO 14, 1.9 for the TASSO 22, 4.5 the for TASSO 44, and 5.3 for the TASSO 35 GeV after the inclusion of the COMPASS data. Since the matching drop for the TASSO measurements at 14, 22, and 44 GeV are milder than those of the 35 GeV one, and the extracted ranges of χ^2 per data point seem to be acceptable for them, we decided to exclude only the TASSO 35 GeV measurements in the fit. Notwithstanding, we have performed a separate analysis which included the TASSO35 dataset and noticed that the central value of the distributions does not affect, whereas the uncertainty estimates are now larger at small z values. We presume that this large uncertainty band is an overestimate and not truthful. This finding also points to the tension between the TASSO35 and COMPASS datasets.

IV. SHK22.H NUMERICAL RESULTS

In this section, we present the main results and findings for the determination of the FFs of the light charged hadron, called SHK22.h, in which most of the available and updated SIA and SIDIS measurements are added to the data sample and analyzed up to the NLO accuracy in perturbative QCD.

We first present the fit quality and discuss in detail the term of both the individual and the total datasets included in our analysis. Then we present the data and theory comparison, both for the SIA and SIDIS datasets analyzed in SHK22.h.

We also illustrate the resulting light charged hadron FFs and their uncertainties, for all parton species, focusing on the comparison of the extracted NLO FFs with the publicly available JAM20 and NNFF1.1h analyses. We discuss the interplay between the SIA and SIDIS experimental data, and the stability of the light charged hadron FFs upon inclusion of the SIDIS datasets.

A. Fit quality

In Table I we report the value of the χ^2 per data point, $\chi^2/N_{\rm dat}$, for the individual datasets for both SIA and SIDIS included in the SHK22.h analysis. This table also includes the number of data points that pass the kinematic cuts. The values of the χ^2 per data point for the total datasets is shown as well.

Considering the numbers presented in this table, a few remarks for the individual and total datasets are in order. As can be seen, the global χ^2 per data point in our fit, equal to 1.080, indicates, in general, a very good description of the entire datasets. Remarkably, a comparable fit quality is observed for both the SIA and SIDIS datasets separately.

Concerning the fit quality of the individual SIDIS experiments, we see that for the h^- production at COMPASS, we obtain a better χ^2 per data point with respect to the COMPASS h^+.

A closer look to the χ^2 per data point presented in this table reveals that acceptable descriptions are achieved almost for all of the individual datasets analyzed in the SHK22.h fit, with two main exceptions.

First, for some datasets reported in the table, the $\chi^2/N_{\rm dat}$ value is still large: this specifically happens for the $h^{\pm}$ light charged hadron production in TASSO 14 GeV, TASSO 44 GeV, and OPAL total inclusive.

From this table we also observe that the $\chi^2/N_{\rm dat}$ value for the DELPHI *uds* and OPAL bottom $h^{\pm}$ is anomalously small. This finding was already observed and reported in some previous FF analyses, which is likely due to the overestimate of the uncorrelated systematic uncertainty. We refer the reader to the Refs. [45,52–54] for more details.

B. Theory and data comparison

In order to assess, it would be instructive to look at the comparison between the data and the NLO theory predictions obtained with the SHK22.h light charged hadron FFs for all SIA and some selected SIDIS datasets.

We start with detailed comparisons with the SIA data analyzed in this work. For all results presented in SHK22.h, the upper panels represent the absolute distributions while the lower ones display the ratio to the experimental central values analyzed in SHK22.h. In Fig. 2, we compare the NLO theory predictions with the inclusive datasets from the

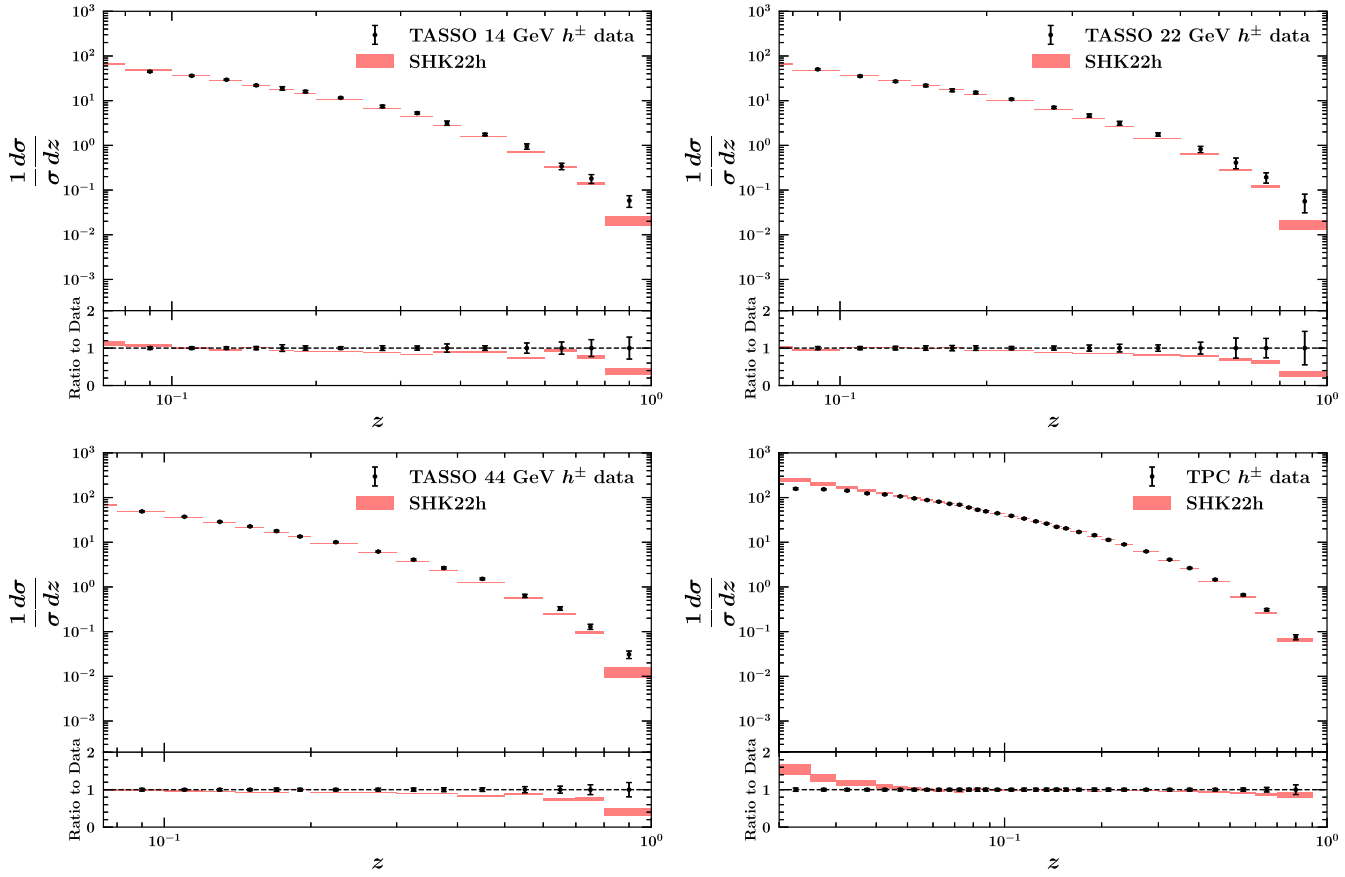

FIG. 2. The data and theory comparison for the TASSO 14, TASSO 22, and TASSO 44 GeV Collaborations at $\sqrt{s} < M_Z$ for inclusive datasets at NLO. The same comparison also are shown for the TPC data. The lower panels display the ratio to the experimental central values.

TASSO 14, TASSO 22, and TASSO 44 GeV Collaborations. The same comparison is also shown for the TPC data. As one can see, overall, satisfying agreements are achieved, however, with exception at high z. Our theory predictions do not satisfy the high-z TASSO data, which is the origin of the slightly high-χ^2 value reported in Table I for these datasets. This specific feature is particularly pronounced for TASSO 44 data, and more moderate, but still significant, for TASSO 14 and 22. For the TPC data, some deviation can be seen for small values of z, but with the χ^2 value reported in Table I, the description of TPC data seems to be still convincing.

In Fig. 3, we present detailed comparisons of the NLO theory predictions with the total inclusive ALEPH, DELPHI, OPAL, and SLD data. Comparisons with the uds-tagged data from DELPHI, OPAL, and SLD data are also presented as well. Consistently with the χ^2 values reported in Table I, the description of these datasets is desirable, with one exception for the OPAL inclusive data. Some small deviation for large-z data can be seen from the plots presented in Fig. 3. Another important finding that emerges for the comparison is that the experimental data points for the inclusive measurements by the OPAL Collaboration fluctuate at high-z around the theoretical predictions by an amount that seems to be typically larger than the calculated uncertainties. This should explain the poor χ^2 values reported in Table I for this specific dataset.

We now turn to the comparisons of our NLO theory predictions with the charm and bottom-tagged data from OPAL, SLD, DELPHI, and SLD. The corresponding plots are shown in Fig. 4. Once again, the goodness of the χ^2 values reported in Table I is reflected in a general good description of the charm and bottom-tagged data.

Figures 5 and 6 present the data and theory comparison for some of the COMPASS multiplicities datasets for the h^+ and h^-, respectively. Each panel shows a distribution as a function of z corresponding to a bin in x and y. As above, the lower panels display the ratio to the experimental central values. A remarkable feature of the distributions shown in these figures is the very nice agreements between the light charged hardon COMPASS data and the `SHK22.h` NLO theory prediction.

While agreement for the COMPASS h^- data is noticeable for all ranges of z and for all bins in x and y, some differences between the COMPASS h^+ data and theory predictions can be seen, more specifically for the small value of z in which the theoretical predictions overshoot the

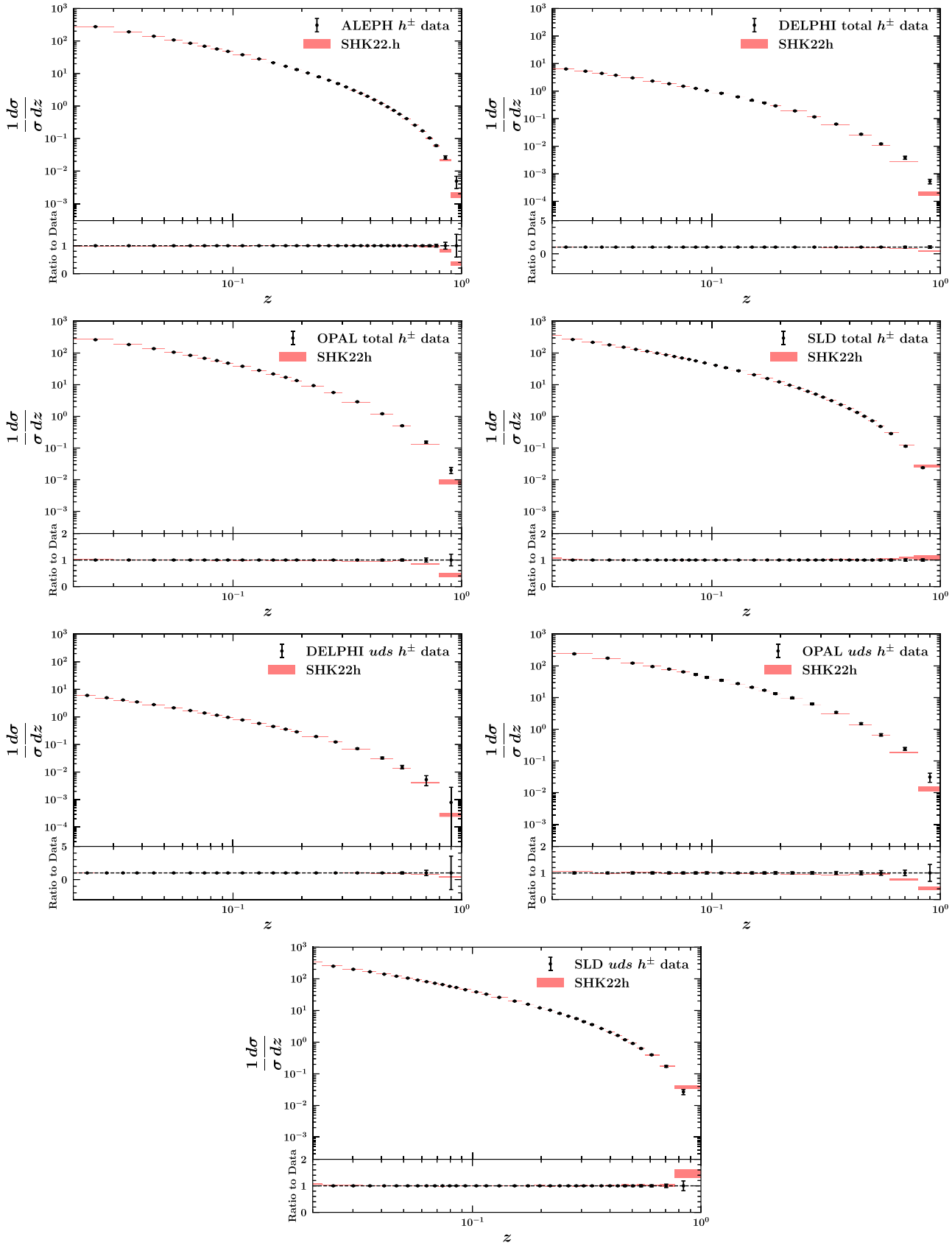

FIG. 3. The data and theory comparison for the ALEPH, DELPHI, OPAL, and SLD Collaborations at $\sqrt{s} = M_Z$ for inclusive and light flavor-tagged datasets. The lower panels display the ratio to the experimental central values.

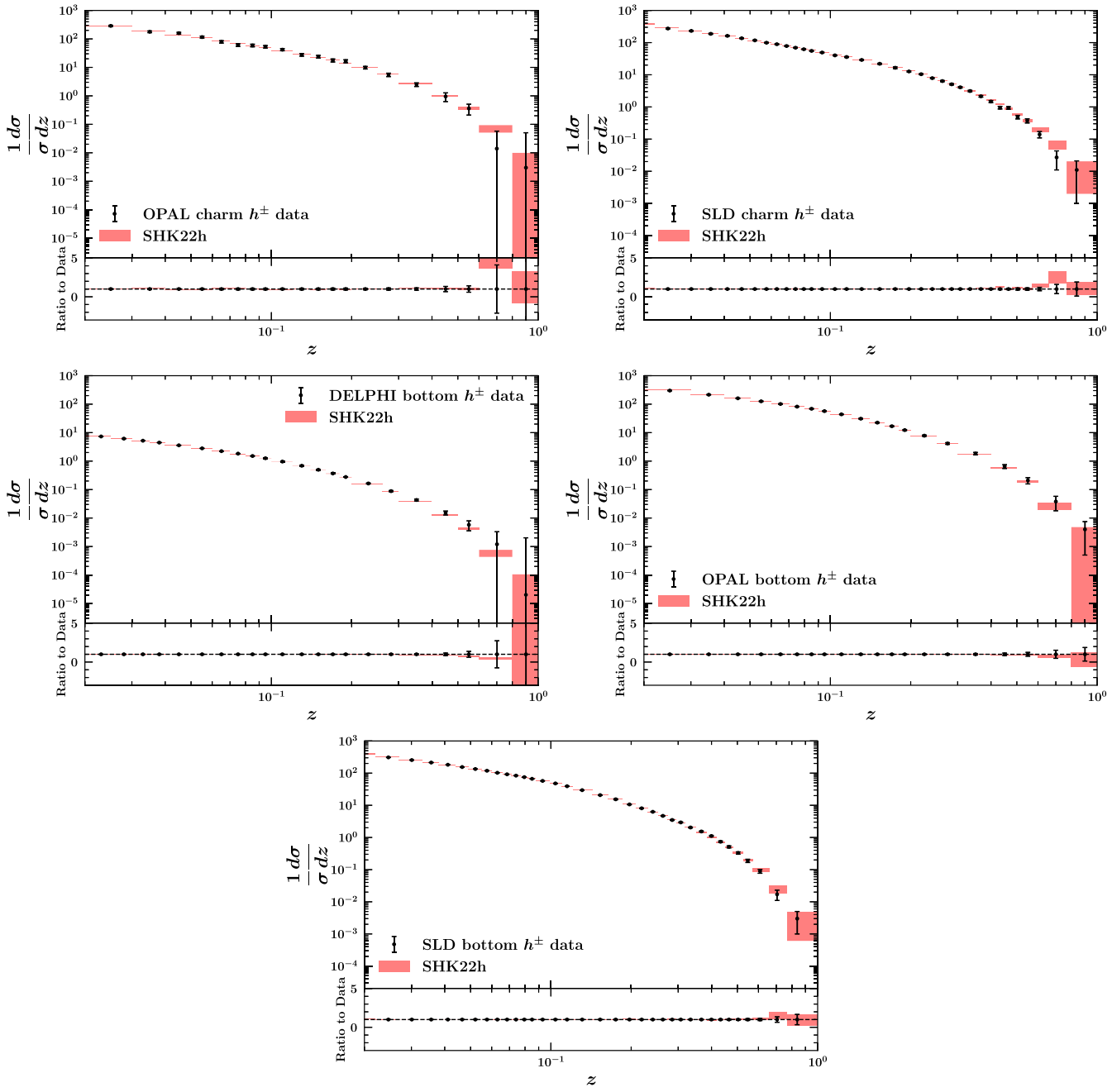

FIG. 4. The data and theory comparison for the charm and bottom-tagged data from OPAL, SLD, DELPHI, and SLD. The lower panels display the ratio to the experimental central values.

data. This is also consistent with the poor χ^2 for COMPASS h^+ data reported in Table I.

As a short summary, in general, an overall good agreement between the analyzed datasets and the NLO theoretical predictions is achieved for all experiments, consistent with the total values of χ^2 reported in this section. Remarkably, the SHK22.h NLO theoretical predictions and both SIA and SIDIS data are in reasonable agreement also in the small and large-z values with exception of few datasets that we discussed above.

C. The SHK22.h light-charged hadron FFs set

We are now in a position to discuss the SHK22.h light-charged hadron FFs sets. In order to study in detail the extracted FFs sets, we compare our best-fit results to other recent counterparts available in the literature, the JAM20 [25] and NNFF1.1h [30] analyses.

We display the light-charged hadron FFs parametrized in SHK22.h fits, and their uncertainties in Fig. 7. We present the 7 hadronic species at $Q = 5$ GeV which are $zD_g^{h^+}(z,Q)$, $zD_d^{h^+}(z,Q)$, $zD_{\bar{d}}^{h^+}(z,Q)$, $zD_u^{h^+}(z,Q)$,

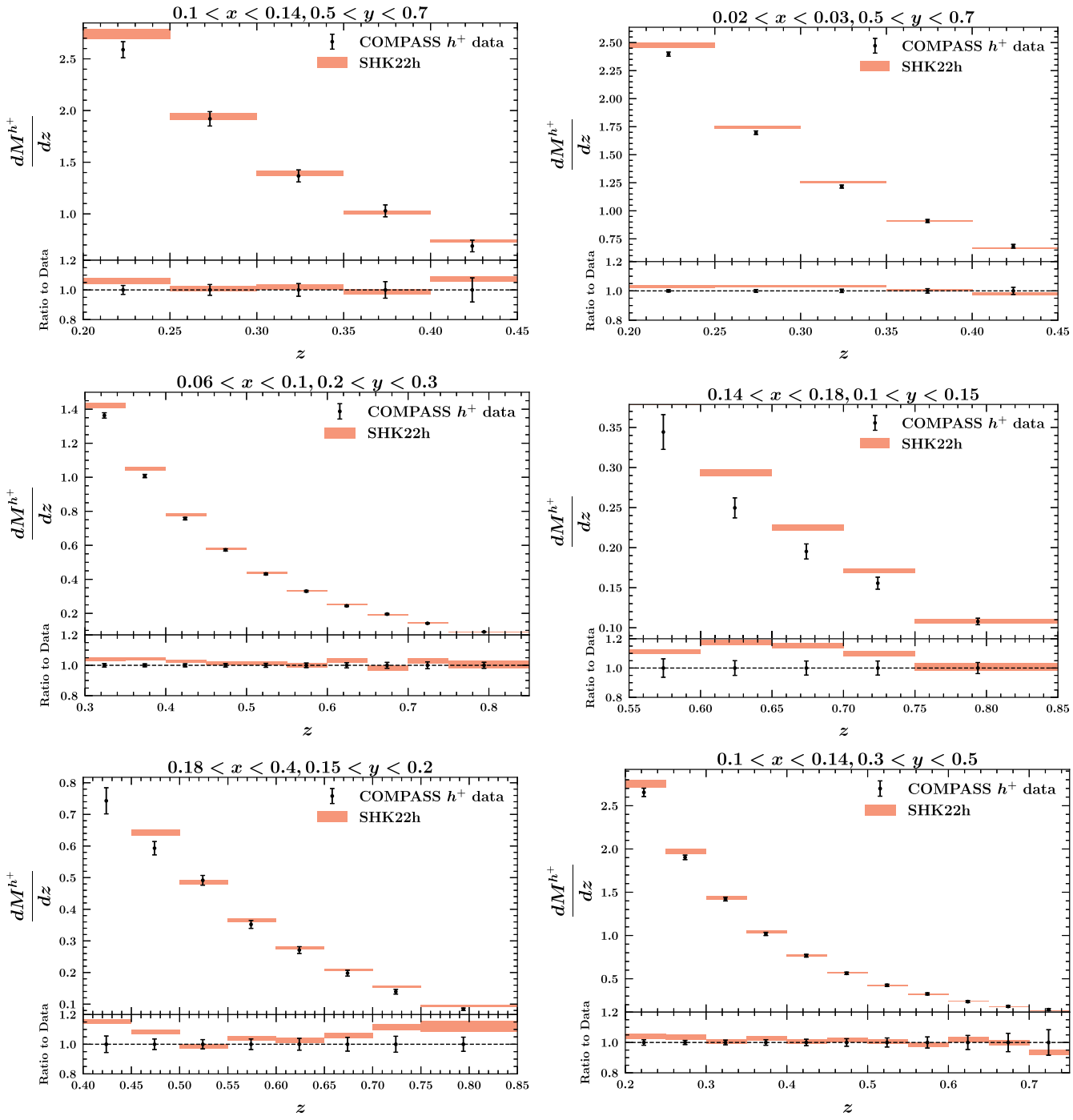

FIG. 5. The data and theory comparison for COMPASS multiplicities datasets for h^+. Each panel shows a distribution as a function of z corresponding to a bin in x and y. The lower panels display the ratio to the experimental central values.

$zD_u^{h^+}(z,Q)$, $zD_{c^+}^{h^+}(z,Q)$, and $zD_{b^+}^{h^+}(z,Q)$. The upper panel of each plot presents the absolute distributions, while the lower panels display the ratio to the `SHK22.h`. It is to be noted that `NNFF1.1h` only extracted the gluon, c, b quark, and flavor singlet combination in their analysis. However, the authors have given instructions to disentangle the up and down contributions in Appendix A of [28] and also produced the related `LHAPDF` format grids, and the plots that are presented here use such prescriptions.

Concerning the shapes of the light-charged hadron FFs, there are a number of interesting similarities and differences between these three different sets as can be seen from the comparisons presented in Fig. 7.

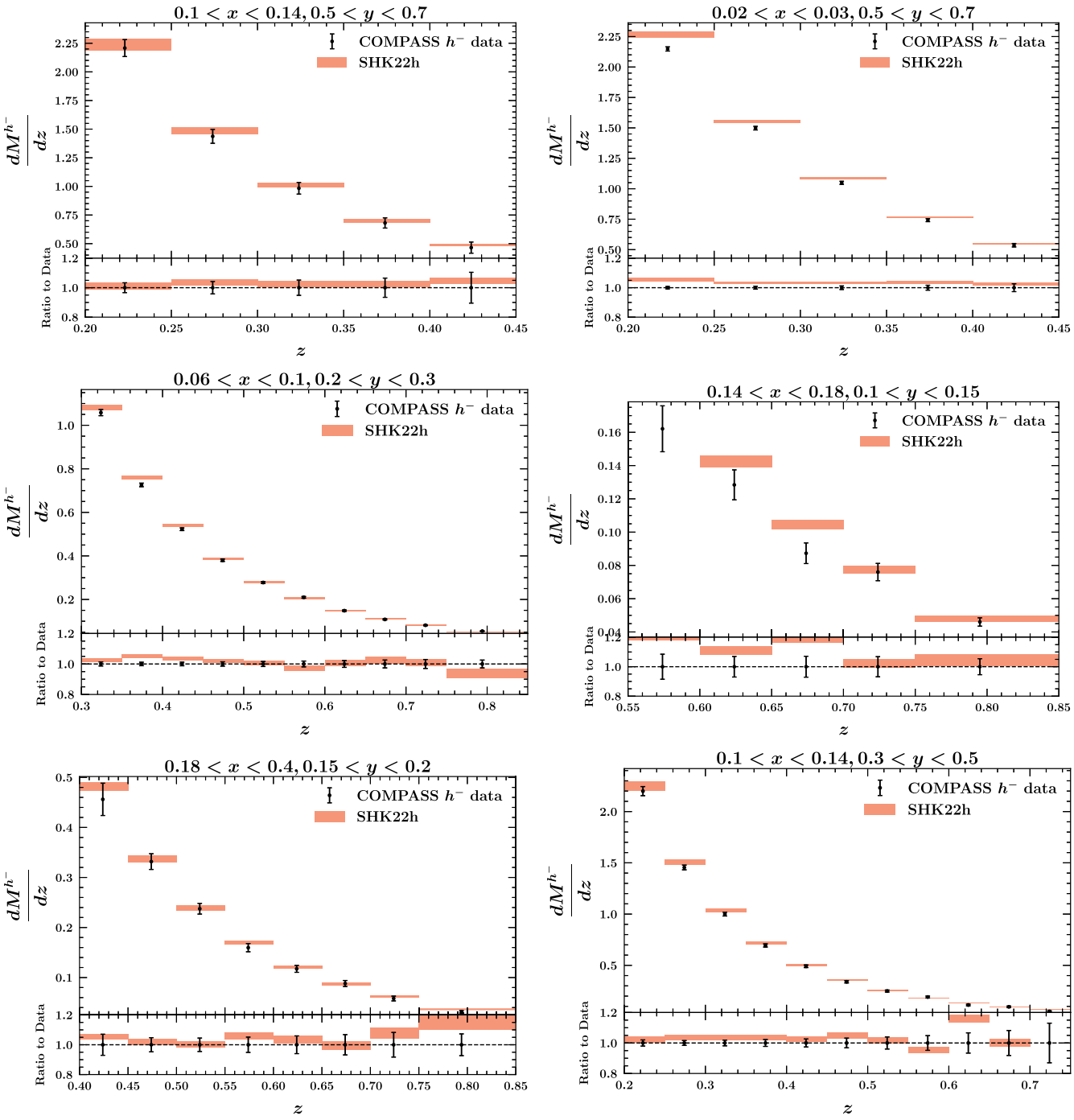

FIG. 6. Same as Fig. 5 but for h^-.

We start with the bottom quark FF $zD_{b^+}^{h^+}(z,Q)$. As one can see from Fig. 7, in terms of the central distribution, these three sets are in very good agreement, except for the high-z region. The uncertainty of the $zD_{b^+}^{h^+}(z,Q)$ FFs deserves a separate comment. The uncertainty for the `NNFF1.1h` is relatively large over all regions of z, while for the `JAM20` it is very narrow. For `SHK22.h`, the calculated uncertainty is slightly large for the large value of z due to the lack of data in this region. Although, it still remained smaller than `NNFF1.1h` over all the range of z. The same findings hold for the charm-quark FF $zD_{c^+}^{h^+}(z,Q)$ with the exception that the central value of `SHK22.h` and `NNFF1.1h` is smaller than those of the `JAM20` for low value of z; $z < 2 \times 10^{-1}$. An interesting difference can be seen for the $zD_{\bar{u}}^{h^+}(z,Q)$ FF between `SHK22.h` and `NNFF1.1h` for medium to large values of z in which

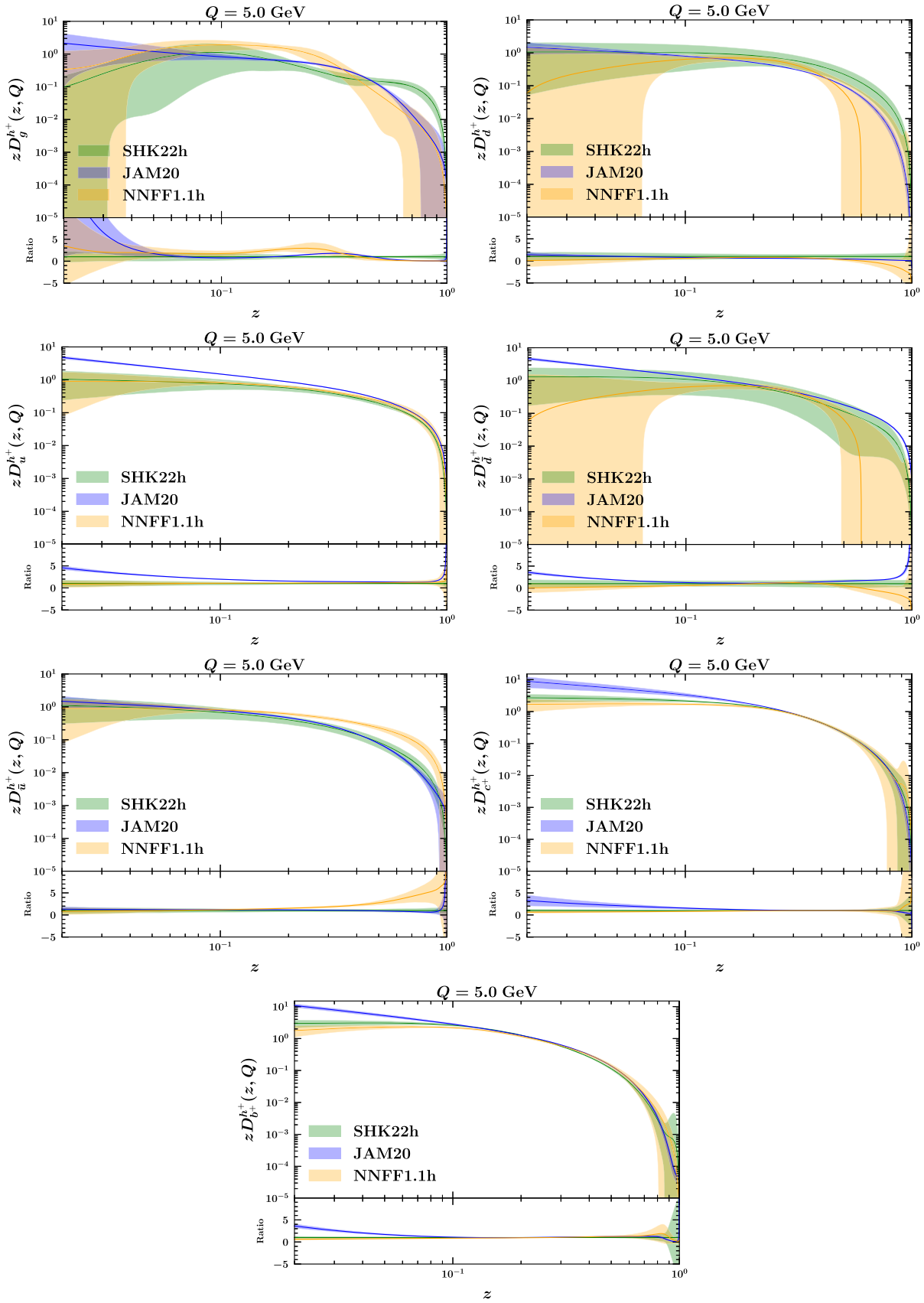

FIG. 7. Upper panel indicates the comparison of our FF sets for charged hadrons with `JAM20` [25] and `NNFF1.1h` [30] FF sets at $Q = 5$ GeV at NLO accuracy. Lower panel represents the ratios of the all sets to the central value of `SHK22.h`.

the NNFF1.1h result is larger than SHK22.h, while for the case of $zD_u^{h^+}(z,Q)$ they are in good agreement. Moderate differences on up-quark FF are observed for the SHK22.h and JAM20 for almost all regions of z while SHK22.h and NNFF1.1h FFs remain consistent.

The most pronounced differences both in shape and uncertainty bands are observed for the gluon $zD_g^{h^+}(z,Q)$, down-quark $zD_d^{h^+}(z,Q)$, and $zD_{\bar{d}}^{h^+}(z,Q)$ FFs. For the $zD_d^{h^+}(z,Q)$ and $zD_{\bar{d}}^{h^+}(z,Q)$ FFs, the differences in shape and uncertainty bands among the three FF sets are more marked for large z rather than the small region of z. As one can see, a fair agreement for the central value is observed only in the region of $z < 2 \times 10^{-1}$. For the medium to large z region, the SHK22.h $zD_{\bar{d}}^{h^+}(z,Q)$ FF is larger than NNFF1.1h and smaller than the JAM20. For the case of $zD_d^{h^+}(z,Q)$, our results are larger than those of the two others in most of the z range. In terms of uncertainty bands, we obtained larger error bands in respect to the JAM20 analysis, which is expected, considering their functional parametrization. For the NNFF1.1h, the central value of $zD_d^{h^+}(z,Q)$ and $zD_{\bar{d}}^{h^+}(z,Q)$ tend to zero for large values of z, $z > 0.6$, with much wider error bands for all region of z.

The central value and uncertainty bands of the gluon FFs $zD_g^{h^+}(z,Q)$, deserve separate comments. As one can see from Fig. 7, there are noticeable differences both in terms of central values and uncertainty bands between these three different sets.

The JAM20 analysis includes all available SIDIS and SIA, and the inclusive DIS and Drell-Yan lepton-pair production as well to calculate the PDFs simultaneously with the FFs. The analysis by the NNPDF Collaboration included the proton-proton data for unidentified charge hadron production by means of Bayesian reweighting to the analysis based only on the SIA datasets. Typically, the uncertainties of the NNFF1.1h FFs are much larger than our results and JAM20 at small and large values of z. The uncertainty band for our result is wider than JAM20 over the small value of z, and smaller for the high-z region. The smallness of the uncertainties for the JAM20 analysis are discussed in detail in Ref. [24]. We should stress here that the smaller uncertainty for all FFs presented in Fig. 7 reflect the more restrictive functional form used in the JAM20 analysis to parametrize their FFs at the input scale.

As we previously mentioned, the SIA cross sections are less sensitive to the gluon FF. As a consequence, one would expect that, in the presence of SIA data only, the gluon FF will be determined with larger uncertainties than other quark FFs. Hence, in the SHK22.h study, the SIDIS data added to the data sample to provide stronger constraint for the gluon density. In the next section, we present our study on the effect of SIDIS data on the extracted FFs.

D. Impact of SIDIS data on the SHK22.h FFs

In this section, we discuss the impact of SIDIS datasets on the extracted FFs. To this end, we compare the main results of the SHK22.h global QCD analysis which include the SIA and SIDIS experimental data with the analysis based on the SIA measurements only.

According to Eqs. (10) and(11), the flavor decompositions in the parametrization of FFs are not the same for the analysis with SIA data only, and the global analysis with both SIA and SIDIS datasets. Then, in order to investigate the impact of including SIDIS data in the SHK22.h FFs analysis, the comparison of FFs for different flavors have been shown in Fig. 8 at $Q = 5$ GeV.

From the comparisons presented in Fig. 8, one observes that the inclusion of SIDIS COMPASS datasets affects both the central values and the uncertainty bands of the extracted FFs.

Such differences are more pronounced for the gluon FF $zD_g^{h^+}$ in terms of both the central value and for the error bands. As one can see from the comparison between the SIA and SIDIS for the gluon FF, the inclusion of the SIDIS data leads to an enhancement of the $zD_g^{h^+}$ distribution for small values of z; $z < 0.2$ in comparison with a fit to the SIA data. We also see that at the large value of z, the SIA + SIDIS and SIA fits are in good agreement. For the medium value of z, the gluon distribution of the global SIA + SIDIS fit get suppressed with respect to the SIA fit. One can also see from Fig. 8 that the gluon FF uncertainty for all ranges of z is reduced. However, gluon FF enter the description of SIA and SIDIS at the same order of perturbation theory, so they are not different regarding the sensitivity to the FF of gluon. In order to constrain the gluon FF one needs to include proton-proton data in the analysis, which is sensitive to gluon FF already at LO [30]. Therefore, we believe that the reduction in the uncertainty of gluon FF is because of a significant increase in the statistics of the included experimental information from SIDIS observables.

For other FFs, we find, in general, a reasonable agreement between SIA and SIDIS fits, but also with important differences. As one can see, SIA and SIDIS fits are in good agreement for the central value of $zD_{c^+}^{h^+}$ and $zD_{b^+}^{h^+}$ FFs for all ranges of z. Reductions on the uncertainty of SIA + SIDIS also can be seen in respect to the SIA for both FFs. For the case of $zD_{u^+}^{h^+}$ FF, the SIA fits are larger than SIA + SIDIS for the range of z down to $z \sim 0.1$ and smaller elsewhere, however, a similar size of error bands are obtained for both SIA and SIA + SIDIS fits. For the case of $zD_{d^+ + s^+}^{h^+}$, a smaller uncertainty band is obtained and the central value for the SIA + SIDIS fit is larger than SIA over the whole range of z. Generally speaking, we find that the inclusion of the SIDIS data could affect the central value of the extracted FFs and leads to significant reductions of the uncertainty, and more specifically for the gluon FF.

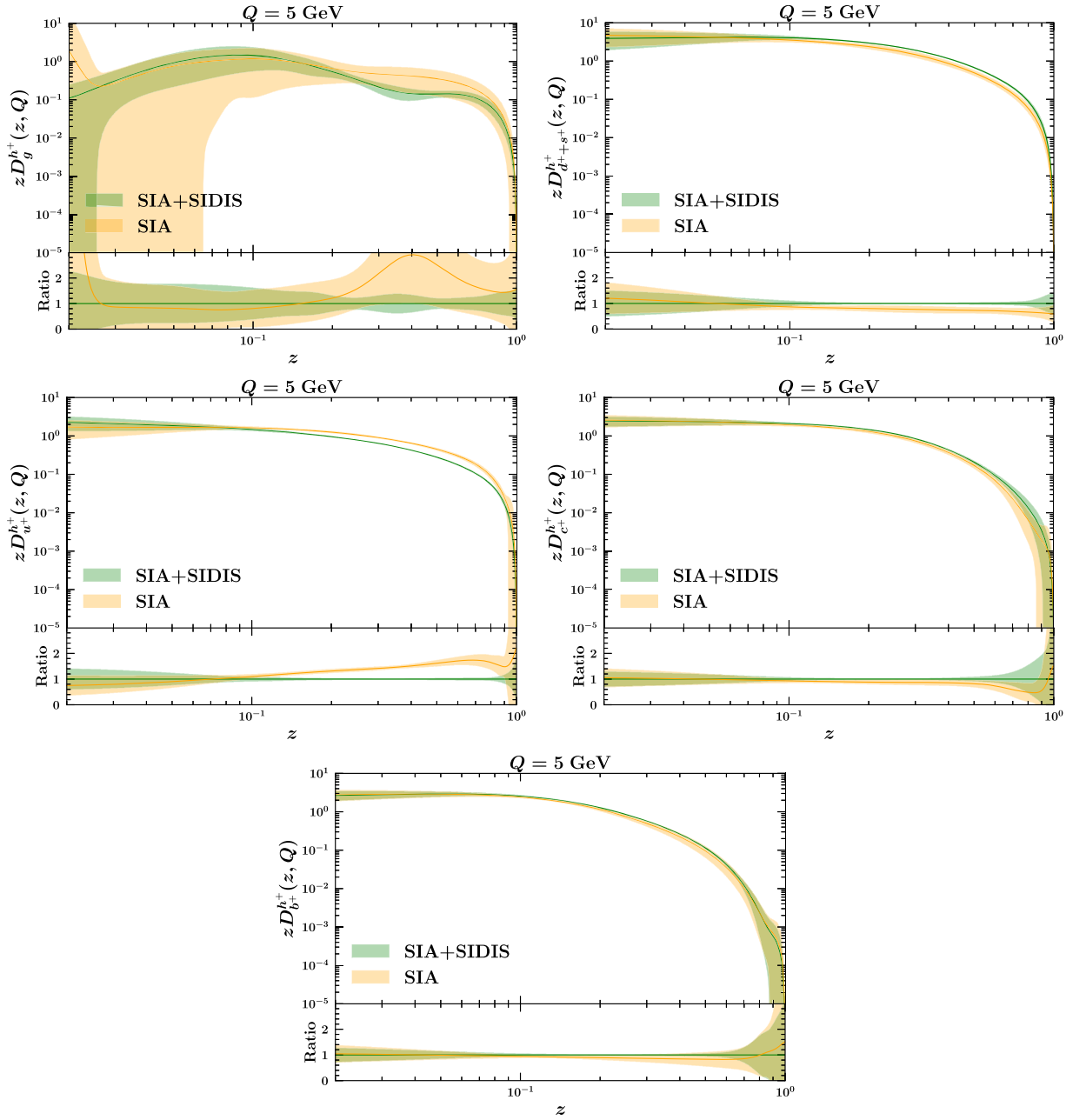

FIG. 8. Comparison of light-charge hadron FF sets obtained from only SIA data with the SHK22.h FFs which were obtained by a combination of SIA and SIDIS experimental data at NLO accuracy.

We finally note that the value of total χ^2 per data point increases from 0.8 in SIA data only fit to the 1.079 in the global analysis of SIA + SIDIS data. As we mentioned before in Sec. III C, the increasing of the value of the total χ^2 is related to the tension between TASSO and the COMPASS experimental datasets. As we reported, the values of χ^2 per data points for the TASSO data, more specifically TASSO 35 GeV, set significant increases after the inclusion the COMPASS data to the data sample.

V. SUMMARY AND CONCLUSIONS

In summary, we have presented a new global QCD analysis of light-charged hadron FFs, SHK22.h, by introducing several new features and some methodological improvements. On the methodological front, we have used the machine learning framework to extract the SHK22.h FFs sets, along with the Monte Carlo uncertainty analysis. This well-established fitting methodology is specifically designed to provide a faithful representation of the experimental uncertainties. This methodology is also useful to

minimize any bias related to the parametrization of the light-charged hadron FFs and to the minimization procedure as well.

In terms of the input datasets, in addition to the comprehensive set of high-energy lepton-lepton annihilation (SIA), we have added the lepton-hadron scattering (SIDIS) datasets to our data sample. We have shown that SIDIS datasets have significant effect on the FFs, and more specifically on the gluon FFs and the reduction of its uncertainty. The tension among some of the datasets included in our analysis is also studied and discussed in detail.

The detailed comparisons to the existing light-charged hadron FFs sets (`NNFF1.1h` and `JAM20`) fully demonstrate a reasonable agreement within the FFs error bands. Although, some discrepancies in flavor dependence were observed, more specifically for the gluon and down-quark FFs. The resulting NLO theory predictions for the SIA and SIDIS cross sections show very good agreement with the corresponding analyzed experimental datasets, as confirmed by the reported total χ^2 per data point.

Based on our findings in this study, one can conclude that adding the SIDIS data in the light-charged hadron study could lead to a much better level of precision of the extracted FFs.

In terms of future work, it would be interesting to revisit this analysis and study in detail the light-charged hadron FFs analysis described here considering the nuclear corrections in which we expect that it could affect the resulting FFs and their uncertainty, and could improve the description of the SIDIS data as well. Exploring the implications of such correction is left for future work.

The parametrizations of the `SHK22.h` light-charged hadron FFs presented in this paper are available in the standard `LHAPDF` format [55], and can be obtained from the authors upon request.

ACKNOWLEDGMENTS

The authors gratefully acknowledge many helpful discussions and comments by Hubert Spiesberger and Valerio Bertone that elevated and enriched the paper significantly. We thank the School of Particles and Accelerators, Institute for Research in Fundamental Sciences (IPM) for financial support provided for this project. M. S. is thankful to the Iran Science Elites Federation for the financial support. H. K. is also thankful to the Theoretical Physics Department of Maynooth University and, University of Science and Technology of Mazandaran for financial support provided for this research.

[1] A. Metz and A. Vossen, Parton fragmentation functions, Prog. Part. Nucl. Phys. **91**, 136 (2016).
[2] A. Accardi, J. L. Albacete, M. Anselmino, N. Armesto, E. C. Aschenauer, A. Bacchetta, D. Boer, W. K. Brooks, T. Burton, N. B. Chang *et al.*, Electron Ion Collider: The next QCD frontier: Understanding the glue that binds us all, Eur. Phys. J. A **52**, 268 (2016).
[3] A. C. Aguilar, Z. Ahmed, C. Aidala, S. Ali, V. Andrieux, J. Arrington, A. Bashir, V. Berdnikov, D. Binosi, L. Chang *et al.*, Pion and kaon structure at the electron-ion collider, Eur. Phys. J. A **55**, 190 (2019).
[4] A. Abada *et al.* (FCC Collaboration), FCC physics opportunities: Future circular collider conceptual design report volume 1, Eur. Phys. J. C **79**, 474 (2019).
[5] A. Abada *et al.* (FCC Collaboration), FCC-ee: The lepton collider: Future circular collider conceptual design report volume 2, Eur. Phys. J. Special Topics **228**, 261 (2019).
[6] R. Seidl *et al.* (Belle Collaboration), Update of inclusive cross sections of single and pairs of identified light charged hadrons, Phys. Rev. D **101**, 092004 (2020).
[7] J. P. Lees *et al.* (*BABAR* Collaboration), Production of charged pions, kaons, and protons in e^+e^- annihilations into hadrons at $\sqrt{s} = 10.54$ GeV, Phys. Rev. D **88**, 032011 (2013).
[8] W. Braunschweig *et al.* (TASSO Collaboration), Global jet properties at 14-GeV to 44-GeV center-of-mass energy in e^+e^- annihilation, Z. Phys. C **47**, 187 (1990).
[9] H. Aihara *et al.* (TPC/Two Gamma Collaboration), Charged-Hadron Inclusive Cross Sections and Fractions in e^+e^- Annihilation $\sqrt{s} = 29$ GeV, Phys. Rev. Lett. **61**, 1263 (1988).
[10] D. Buskulic *et al.* (ALEPH Collaboration), Measurement of alpha-s from scaling violations in fragmentation functions in $e + e-$ annihilation, Phys. Lett. B **357**, 487 (1995); **364**, 247(E) (1995).
[11] P. Abreu *et al.* (DELPHI Collaboration), $\pi^{\pm}$, $K^{\pm}$, p and $\bar{p}$ production in $Z^0 \to q\bar{q}$, $Z^0 \to b\bar{b}$, $Z^0 \to u\bar{u}, d\bar{d}, s\bar{s}$, Eur. Phys. J. C **5**, 585 (1998).
[12] K. Ackerstaff *et al.* (OPAL Collaboration), Measurements of flavor dependent fragmentation functions in $Z^0 \to q$ anti-q events, Eur. Phys. J. C **7**, 369 (1999).
[13] K. Abe *et al.* (SLD Collaboration), Production of $\pi+, \pi-, K+, K-$, p and $\bar{p}$ in light (uds), c, and b jets from Z^0 decays, Phys. Rev. D **69**, 072003 (2004).
[14] C. Adolph *et al.* (COMPASS Collaboration), Multiplicities of charged pions and charged hadrons from deep-inelastic scattering of muons off an isoscalar target, Phys. Lett. B **764**, 1 (2017).
[15] A. Airapetian *et al.* (HERMES Collaboration), Multiplicities of charged pions and kaons from semi-inclusive

deep-inelastic scattering by the proton and the deuteron, Phys. Rev. D **87**, 074029 (2013).
[16] R. Akhunzyanov *et al.* (COMPASS Collaboration), K^- over K^+ multiplicity ratio for kaons produced in DIS with a large fraction of the virtual-photon energy, Phys. Lett. B **786**, 390 (2018).
[17] G. D. Alexeev *et al.* (COMPASS Collaboration), Antiproton over proton and K^- over K^+ multiplicity ratios at high z in DIS, Phys. Lett. B **807**, 135600 (2020).
[18] S. Acharya *et al.* (ALICE Collaboration), Production of light-flavor hadrons in pp collisions at $\sqrt{s} = 7$ and $\sqrt{s} = 13$ TeV, Eur. Phys. J. C **81**, 256 (2021).
[19] L. Adamczyk *et al.* (STAR Collaboration), Neutral pion cross section and spin asymmetries at intermediate pseudorapidity in polarized proton collisions at $\sqrt{s} = 200$ GeV, Phys. Rev. D **89**, 012001 (2014).
[20] A. Adare *et al.* (PHENIX Collaboration), Inclusive cross section and double-helicity asymmetry for π^0 production at midrapidity in $p + p$ collisions at $\sqrt{s} = 510$ GeV, Phys. Rev. D **93**, 011501 (2016).
[21] S. Chatrchyan *et al.* (CMS Collaboration), Study of high-pT charged particle suppression in PbPb compared to pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV, Eur. Phys. J. C **72**, 1945 (2012).
[22] S. Chatrchyan *et al.* (CMS Collaboration), Charged particle transverse momentum spectra in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV, J. High Energy Phys. 08 (2011) 086.
[23] R. A. Khalek, V. Bertone, and E. R. Nocera, Determination of unpolarized pion fragmentation functions using semi-inclusive deep-inelastic-scattering data, Phys. Rev. D **104**, 034007 (2021).
[24] H. Abdolmaleki, M. Soleymaninia, H. Khanpour, S. Amoroso, F. Giuli, A. Glazov, A. Luszczak, F. Olness, and O. Zenaiev (xfitter Developers' Team), QCD analysis of pion fragmentation functions in the xFitter framework, Phys. Rev. D **104**, 056019 (2021).
[25] E. Moffat, W. Melnitchouk, T. C. Rogers, and N. Sato (Jefferson Lab Angular Momentum (JAM)), Simultaneous Monte Carlo analysis of parton densities and fragmentation functions, Phys. Rev. D **104**, 016015 (2021).
[26] M. Soleymaninia, H. Abdolmaleki, and H. Khanpour, First NNLO fragmentation functions of K^0_S and $\Lambda/\bar{\Lambda}$ and their uncertainties in the presence of hadron mass corrections, Phys. Rev. D **102**, 114029 (2020).
[27] M. Soleymaninia, M. Goharipour, H. Khanpour, and H. Spiesberger, Simultaneous extraction of fragmentation functions of light charged hadrons with mass corrections, Phys. Rev. D **103**, 054045 (2021).
[28] V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo (NNPDF Collaboration), A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties, Eur. Phys. J. C **77**, 516 (2017).
[29] M. Soleymaninia, M. Goharipour, and H. Khanpour, First QCD analysis of charged hadron fragmentation functions and their uncertainties at next-to-next-to-leading order, Phys. Rev. D **98**, 074002 (2018).
[30] V. Bertone, N. P. Hartland, E. R. Nocera, J. Rojo, and L. Rottoli (NNPDF Collaboration), Charged hadron fragmentation functions from collider data, Eur. Phys. J. C **78**, 651 (2018).
[31] M. Delpasand, S. M. Moosavi Nejad, and M. Soleymaninia, Λ_c^+ fragmentation functions from pQCD approach and the Suzuki model, Phys. Rev. D **101**, 114022 (2020).
[32] M. Benzke, B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, B-meson production in the general-mass variable-flavour-number scheme and LHC data, Eur. Phys. J. C **79**, 814 (2019).
[33] M. Salajegheh, S. M. Moosavi Nejad, H. Khanpour, B. A. Kniehl, and M. Soleymaninia, B-hadron fragmentation functions at next-to-next-to-leading order from a global analysis of e^+e^- annihilation data, Phys. Rev. D **99**, 114001 (2019).
[34] E. R. Nocera, Fragmentation functions of charged hadrons, Proc. Sci., DIS2017 (**2018**) 231 [arXiv:1709.03400].
[35] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions from high-precision collider data, Eur. Phys. J. C **77**, 663 (2017).
[36] https://github.com/MapCollaboration/MontBlanc.
[37] M. Soleymaninia, H. Hashamipour, H. Khanpour, and H. Spiesberger, Fragmentation functions for $\Xi^-/\bar{\Xi}^+$ using neural networks, arXiv:2202.05586.
[38] D. Graudenz, One particle inclusive processes in deeply inelastic lepton—nucleon scattering, Nucl. Phys. **B432**, 351 (1994).
[39] G. Altarelli, R. K. Ellis, G. Martinelli, and S. Y. Pi, Processes involving fragmentation functions beyond the leading order in QCD, Nucl. Phys. **B160**, 301 (1979).
[40] P. J. Rijken and W. L. van Neerven, $O(\alpha_s^2)$ contributions to the longitudinal fragmentation function in $e + e-$ annihilation, Phys. Lett. B **386**, 422 (1996).
[41] A. Mitov and S. O. Moch, QCD corrections to semi-inclusive hadron production in electron-positron annihilation at two loops, Nucl. Phys. **B751**, 18 (2006).
[42] B. C. Csáji *et al.*, Approximation with artificial neural networks, M.Sc. thesis, Faculty of Sciences, Etvs Lornd University, Hungary, 2001.
[43] R. A. Khalek, Exploring the substructure of nucleons and nuclei with machine learning, arXiv:2110.01924.
[44] H. Moutarde, P. Sznajder, and J. Wagner, Unbiased determination of DVCS compton form factors, Eur. Phys. J. C **79**, 614 (2019).
[45] N. Sato, J. J. Ethier, W. Melnitchouk, M. Hirai, S. Kumano, and A. Accardi, First Monte Carlo analysis of fragmentation functions from single-inclusive e^+e^- annihilation, Phys. Rev. D **94**, 114004 (2016).
[46] R. Abdul Khalek, J. J. Ethier, and J. Rojo (NNPDF Collaboration), Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider, Eur. Phys. J. C **79**, 471 (2019).
[47] N. Sato, W. Melnitchouk, S. E. Kuhn, J. J. Ethier, and A. Accardi (Jefferson Lab Angular Momentum), Iterative Monte Carlo analysis of spin-dependent parton distributions, Phys. Rev. D **93**, 074005 (2016).
[48] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions for the LHC Run II, J. High Energy Phys. 04 (2015) 040.
[49] R. Abdul Khalek, J. J. Ethier, J. Rojo, and G. van Weelden, nNNPDF2.0: Quark flavor separation in nuclei from LHC data, J. High Energy Phys. 09 (2020) 183.
[50] S. Agarwal, K. Mierle *et al.*, Ceres solver, http://ceres-solver.org.

[51] M. Soleymaninia, M. Goharipour, H. Khanpour, and H. Spiesberger, Simultaneous QCD analysis for identified and unidentified light charged hadrons, Proc. Sci., EPS-HEP2021 (**2022**) 407.

[52] D. de Florian, R. Sassot, M. Epele, R. J. Hernández-Pinto, and M. Stratmann, Parton-to-pion fragmentation reloaded, Phys. Rev. D **91**, 014035 (2015).

[53] M. Hirai, H. Kawamura, S. Kumano, and K. Saito, Impacts of B-factory measurements on determination of fragmentation functions from electron-positron annihilation data, Prog. Theor. Exp. Phys. **2016**, 113B04 (2016).

[54] D. de Florian, M. Epele, R. J. Hernandez-Pinto, R. Sassot, and M. Stratmann, Parton-to-kaon fragmentation revisited, Phys. Rev. D **95**, 094019 (2017).

[55] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, LHAPDF6: Parton density access in the LHC precision era, Eur. Phys. J. C **75**, 132 (2015).