Using analytic models to describe effective PDFs

S. A. Ochoa-Oregon,[1] D. F. Rentería-Estrada,[2] R. J. Hernández-Pinto,[1] G. F. R. Sborlini,[3] and P. Zurita[4,5]

[1]*Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Sinaloa, Ciudad Universitaria, CP 80000 Culiacán, Sinaloa, México*

[2]*Instituto de Física Corpuscular, Universitat de València—Consejo Superior de Investigaciones Científicas, Parc Científic, 46980 Paterna, Valencia, Spain*

[3]*Departamento de Física Fundamental e IUFFyM, Universidad de Salamanca, 37008 Salamanca, Spain*

[4]*Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany*

[5]*Departamento de Física Teórica & IPARCOS, Universidad Complutense de Madrid, E-28040 Madrid, Spain*

(Received 29 April 2024; accepted 5 August 2024; published 29 August 2024)

Parton distribution functions play a pivotal role in hadron collider phenomenology. They are nonperturbative quantities extracted from fits to available data, and their scale dependence is dictated by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi evolution equations. In this article, we discuss machine-assisted strategies to efficiently compute parton distribution functions (PDFs) explicitly incorporating the scale dependence. Analytical approximations to the PDFs as functions of x and Q^2, including up to next-to-leading-order effects in quantum chromodynamics, are obtained. The methodology is tested by reproducing the HERAPDF2.0 set and implementing the analytical expressions in benchmarking codes. It is found that the computational time cost of evaluating the distributions is reduced by ∼50%, while the precision of the simulations stays well under control.

DOI: 10.1103/PhysRevD.110.036019

I. INTRODUCTION

Breaking the precision frontier in particle physics is a challenging task. The tiny discrepancies among experiments and theoretical predictions might hide new phenomena, and this forces theoreticians to refine as much as possible their methodologies to produce accurate simulations of particle collisions. Even if a plethora of powerful methods are available (see, e.g., Ref. [1]), most of them require enormous computing resources to achieve the intended precision. In this direction, the purpose of this investigation is to reduce the resource consumption of higher-order calculations for perturbative quantum field theory while keeping the precision achieved by the most advanced theoretical results. This will allow us to provide fast and reliable results, and also to reduce the environmental impact associated with high-energy research [2].

Most of the current higher-order cross-section simulations rely on the factorization theorem [3–5]—i.e., splitting the whole calculation into perturbative and nonperturbative sectors. The process-dependent perturbative contribution is given in terms of well-defined analytical functions (or highly efficient numerical representations of them). The universal, nonperturbative part, encoded in parton distribution functions (PDFs), cannot be computed from first principles: only their evolution with the factorization scale chosen can be calculated via the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [6] evolution equations. The determination of the PDFs is, instead, achieved by performing global fits to existing data. For practical implementations, the outcomes of these fits are given as tables in the relevant kinematic variables with a function that interpolates over them. The interpolation requires us to perform several evaluations, and the routines are written to be as fast as possible to allow for the efficient computation of physical observables. However, some cross sections might require a very large number of trials (and running time) to achieve relevant precision, due to the nature of the process. Thus, it is worth exploring the possibility of reducing the CPU/GPU time required by providing an alternative form for the PDFs. Therefore, in this work, we aim to determine an analytical functional form for a set of known PDFs, by approximating their (x, Q^2) behavior. A different, global alternative would be to identify approximations to closed analytic solutions of DGLAP equations, a highly nontrivial problem in the context of coupled integrodifferential equations [7].

To achieve this ambitious objective, we profit from artificial intelligence, machine learning, and similar tools implemented within *Mathematica* built-in functions. By using machine-assisted techniques, we identify suitable functional dependencies to accurately describe the PDF sets. Our thesis is that, avoiding the interpolation over the grids, it will drastically reduce the time required to perform the simulations/computations. Furthermore, the PDFs would then be written in terms of a few parameters, reducing the storage required for them. Also, having access to analytic formulas for PDFs will allow us to calculate their derivatives, both in x and Q, which are relevant for performing the matching of F_2 in the threshold regions.

Throughout this work, we focus on the set of collinear proton PDFs extracted by the xFitter Collaboration, HERAPDF2.0 [8]. We would like to highlight that our methodology is applicable to any PDF set, independently of the order at which it is calculated: if the PDF set exists, it is possible to find an analytic approximation with the techniques explained in this article. For this reason, we stress that our study does not replace, in any way, the need of performing global fits. On the contrary, it is only by having the PDF sets already determined that we can search *a posteriori* for a functional form. Our analytic approximations to a PDF set constitute a first proof of concept, and we show that it has successfully passed several physically motivated quality checks.

This work is organized as follows: In Sec. II, we present our proposed analytical model for a set of collinear, unpolarized proton PDFs, using HERAPDF2.0 as a baseline. The methodology used to determine the functional form is explained in detail in Sec. III. Our results and their comparison with the baseline PDFs can be found in Sec. IV. In Sec. V, we carefully quantify the errors induced by our analytic approximations and check the validity of the sum rules. There, we also check the time required to run benchmarking codes using both the available PDF grids and our *a posteriori* analytical PDFs. We conclude by summarizing our findings in Sec. VI.

II. STARTING HYPOTHESIS

Traditionally, proton PDFs [9] are determined by proposing a functional form in Bjorken x for the parton densities (or a linear combination of them) at some chosen initial scale Q_0^2. Giving values to the parameters involved and evolving the PDFs to the experimental scales Q^2 using the DGLAP equations, one obtains the distributions that, convoluted with the partonic cross sections, are compared with data. By repeating the procedure until an adequate description is achieved (according to some chosen criteria), the best-fit parameters are determined. This is the case of many sets of proton and nuclear PDFs (e.g., [10–16]) and fragmentation functions (e.g., [17,18]). Another possibility that does not use a proposed parametrization but rather relies on neural networks is also employed by the NNPDF Collaboration (see, e.g., Ref. [19] and references within).

Regarding the parametric form of the distributions, they are usually chosen to be an Euler beta function, with some extra flexibility given by a multiplicative polynomial or exponential. Schematically,

$$f_i(x, Q_0^2) = N_i x^{\alpha_i}(1-x)^{\beta_i} P(x, c_{ij}), \tag{1}$$

where x is, in the Breit frame, the fraction of momentum of the proton carried by the parton; i indicates a parton flavor or a combination of them (selected to do the evolution); and $P(x, c_{ij})$ is some function of x with coefficients c_{ij}. This form is flexible enough that very different shapes can be achieved by varying the parameters. After the best fit is found, tables in x and Q^2 are made available (in modern times, through LHAPDF [20,21]), with fast interpolating routines.

To determine an analytical expression for the collinear proton PDFs, which we test with the HERAPDF2.0 set at next-to-leading-order (NLO) accuracy in QCD, we propose that the Q^2 dependence of the PDFs is given by an extension of Eq. (1), with each parameter acquiring a Q^2 dependence. In other words, we start from the assumption that

$$f_i(x, Q^2) = N_i(Q^2) x^{\alpha_i(Q^2)}(1-x)^{\beta_i(Q^2)} \times P(x, c_{ij}(Q^2)), \tag{2}$$

with P being, e.g., a polynomial in x. The aim of our work is to determine these functions using machine learning (ML) methods, and to explore to what extent the Q^2 dependence can be reproduced with these functional forms.

III. METHODOLOGY

In order to obtain the coefficients from Eq. (2), we performed a two-step fitting procedure. First, we fixed a PDF set within the LHAPDF framework: in our study, it was HERAPDF2.0_NLO_EIG, obtained by the xFitter group, which uses solely data from the HERA collider. Then, we generated a grid of $\mathcal{O}(5000)$ random points in $\{x, Q\}$ with $x \in [10^{-4}, 0.65]$ and $Q \in [Q_{\text{Min}}, 1000]$, evaluating the corresponding PDF in these points. The range in x was chosen in accordance with the recommendations in Ref. [8]. For the gluon and light quarks, we set $Q_{\text{Min}} = Q_0 \approx 1.37$ GeV, while we used $Q_{\text{Min}} = 1.5$ GeV and $Q_{\text{Min}} = 4.5$ GeV for the charm and bottom quarks, respectively.

Regarding the parametrization, we draw inspiration from Ref. [8], where at the initial scale Q_0, the authors proposed

$$xu_v(x) = A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}(1 + E_{u_v} x^2), \tag{3}$$

$$xd_v(x) = A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}}, \tag{4}$$

$$x\bar{U}(x) = A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}(1 + D_{\bar{U}} x^2), \tag{5}$$

$$x\bar{D}(x) = A_{\bar{D}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{D}}}, \tag{6}$$

$$x\bar{s}(x) = f_s x\bar{D}(x) \tag{7}$$

for the valence and sea distributions of light quarks (u, d, and s), and

$$xg(x) = A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{C'_g} \tag{8}$$

for the gluon. In these formulas, we are using $f_s = 0.4$ at Q_0 and $s \equiv \bar{s}$, as well as

$$\bar{U} = \bar{u}, \qquad \bar{D} = \bar{d} + \bar{s}. \tag{9}$$

When the scales probed are above the production threshold of a heavy quark, said quark will be radiatively produced and thus have a nonzero distribution (unless intrinsic heavy flavors are considered, which is beyond the scope of the present work). In particular, in HERAPDF2.0, the optimized general mass variable-flavor-number scheme RTOPT was used to treat charm and bottom PDFs [22]. Due to its radiative nature, the gluon has a noticeable impact in the determination of sea- and heavy-quark distributions.

Motivated by this, we proposed functional forms like Eq. (2) to fit $u_v(x, Q^2)$ and $d_v(x, Q^2)$, while we relied on gluon-like sums of Euler beta functions—i.e., Eq. (8)—for the remaining distributions. To perform a quantitative study of the goodness of the approximation, we used the so-called *integral error*, defined according to

$$\Delta_i(Q^2) = \frac{I\left[f_i^{\mathrm{ML}}(x, Q^2), Q^2\right] - I\left[f_i^{\mathrm{HERA}}(x, Q^2), Q^2\right]}{I\left[f_i^{\mathrm{HERA}}(x, Q^2), Q^2\right]}, \tag{10}$$

with f_i^{ML} and f_i^{HERA} being the PDFs corresponding to the flavor i obtained with our analytic approximation (ML-PDF from here onwards) and with the HERAPDF2.0 set, respectively. The integration operator $I[f, Q^2]$ is given by

$$I[f, Q^2] = \int_{10^{-4}}^{1} dx\, f(x, Q^2). \tag{11}$$

We would like to highlight that this definition of the error is suitable to control the validity of the sum rules. Also, since the PDFs contribute to the hadronic cross section through a convolution with the partonic cross section, we observe that the integral error successfully provides an error estimation for the physical observables. Both tests are discussed extensively in Sec. V.

Besides this error definition, we explored additional estimators, such as the *error shape*,

$$\tilde{\Delta}_i(Q^2) = \frac{1}{N}\sum_{j=1}^{N}\left|1 - \frac{f_i^{\mathrm{ML}}(x_j, Q^2)}{f_i^{\mathrm{HERA}}(x_j, Q^2)}\right|, \tag{12}$$

where $\{x_j\}_{j=1,\ldots,N}$ is a partition of $x \in [10^{-4}, 1]$. This definition is particularly sensitive to fluctuations or oscillations around the original PDF, and it leads to a noticeable overestimation of the error. For example, we found typical error shapes of $\mathcal{O}(10\%\text{–}50\%)$ for a large range of Q values, in spite of percent- and even subpercent-level deviations of the sum rules and other physical observables (see Sec. V). For this reason, we consider the *integral error* as a more reliable estimator.

IV. RESULTS

In this section, we explicitly report the expressions for the analytic approximations to all the PDF flavors provided by HERAPDF2.0 at NLO. We fitted the u_v, $\bar{u}$, d_v, $\bar{d}$, g, s, c, and b distributions following the functional forms proposed in each of the following subsections. For each value of Q, we extracted the coefficients $\{A_i, B_i, \ldots\}$, performing a fit in x with the *Mathematica* built-in function NonlinearModelFit [23].

Then, we performed a second fit to determine the Q dependence of these coefficients. In this step, we relied on machine-assisted techniques that provide suitable functional forms to build the *Ansätze*. In particular, we made use of the *Mathematica* built-in routine FindFormula to generate an approximation to the coefficients $\{A_i, B_i, \ldots\}$ as functions of Q. Then, we used these approximations to propose an *Ansatz* for each coefficient, including similar expressions to those found by FindFormula. The functions obtained after this two-step fitting procedure, together with the full set of analytic ML-PDFs, are publicly available in an ancillary file uploaded to the Zenodo repository [24]. Given that the final number of parameters is quite large, we refrain from presenting in the manuscript a table with the fitted parameters and functional forms.

A. Gluon distribution

To obtain an approximation to the gluon distribution, we proposed a functional form inspired by Eq. (8). Concretely, we used

$$xg(x, Q^2) = f_1(x, Q^2) - \Theta(x_{C,g} - x) f_2(x, Q^2), \tag{13}$$

with f_i, for $i = 1, 2$, as given in Eq. (1). In this way, for values of x below the threshold $x_{C,g}$, Eq. (13) offers more flexibility to fit the PDF, in a fully analogous way to that used in Ref. [8].

Furthermore, we noticed that the Q dependence in the coefficient varies significantly in the whole Q range considered (i.e., from Q_0 to 1000 GeV). Hence, we split the analysis into four separate regions:

$$R_1 = \{Q_0 \le Q \le 2.5\ \mathrm{GeV}\}, \tag{14}$$

$$R_2 = \{2.5\ \mathrm{GeV} \le Q \le 5\ \mathrm{GeV}\}, \tag{15}$$

$$R_3 = \{5\ \text{GeV} \leq Q \leq 150\ \text{GeV}\}, \tag{16}$$

$$R_4 = \{150\ \text{GeV} \leq Q \leq 1000\ \text{GeV}\}, \tag{17}$$

which we determined by examining the Q behavior of the gluon density. Consequently, we defined

$$\begin{aligned} xg(x,Q^2) &= A_g(Q^2)x^{B_g(Q^2)}(1-x)^{C_g(Q^2)} \\ &\quad - \Theta(x_{C,g}-x)A'_g(Q^2)x^{B'_g(Q^2)} \\ &\quad \times (1-x)^{C'_g(Q^2)} \end{aligned} \tag{18}$$

for $Q \in R_1$, and

$$\begin{aligned} xg(x,Q^2) &= A_g(Q^2)x^{B_g(Q^2)}(1-x)^{C_g(Q^2)} \\ &\quad - \Theta(x_{C,g}-x)A'_g(Q^2)x^{B'_g(Q^2)} \\ &\quad \times (1-x)^{C'_g(Q^2)}[1+D'_g(Q^2)x^2] \end{aligned} \tag{19}$$

for $Q \in \{R_2, R_3, R_4\}$, together with $x_{C,g} = 0.1$. Again, this last value was fixed by an exploratory procedure.

We emphasize that, since the analysis was done independently for each region, the functions $\{A_g(Q^2), B_g(Q^2), C_g(Q^2)\}$ and $\{A'_g(Q^2), B'_g(Q^2), C'_g(Q^2), D'_g(Q^2)\}$ in Eq. (19) have different behaviors in R_2, R_3, and R_4.

In Fig. 1, we show our analytic ML-PDF approximation to $xg(x,Q^2)$ with respect to the corresponding gluon PDF from HERAPDF2.0_NLO_EIG at 10 GeV (blue), 100 GeV (green), and 1000 GeV (orange). Even if small fluctuations occur at low x, the overall agreement is very good. Furthermore, the agreement remains when choosing different values of Q, thanks to splitting the analysis into four regions.

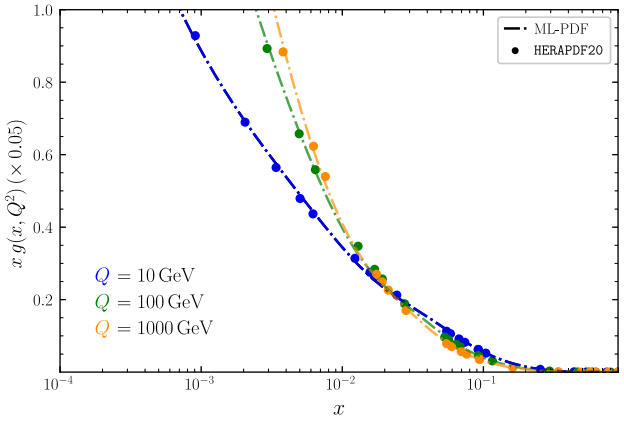

FIG. 1. Comparison between our analytic ML-PDF approximation (dashed lines) and HERAPDF2.0 (solid dots) for gluon distributions, at three different values of Q. Notice that the gluon PDFs are multiplied by a factor of 0.05 in order to better appreciate the differences among them at different scales.

B. Down-quark distributions

In this case, the optimal fit was achieved by considering d_v and $\bar{d}$ distributions. Therefore, we wrote

$$\begin{aligned} xd_v(x,Q^2) &= A_{d_v}(Q^2)x^{B_{d_v}(Q^2)}(1-x)^{C_{d_v}(Q^2)} \\ &\quad \times [1 + D_{d_v}(Q^2)x^2 + E_{d_v}(Q^2)x^4 \\ &\quad + F_{d_v}(Q^2)x^6], \end{aligned} \tag{20}$$

$$\begin{aligned} x\bar{d}(x,Q^2) &= A_{\bar{d}}(Q^2)x^{B_{\bar{d}}(Q^2)}(1-x)^{C_{\bar{d}}(Q^2)} \\ &\quad - \Theta(x_{C,\bar{d}}-x)A'_{\bar{d}}(Q^2)x^{B'_{\bar{d}}(Q^2)}(1-x)^{C'_{\bar{d}}(Q^2)} \\ &\quad \times [1 + D'_{\bar{d}}(Q^2)x^2 + E'_{\bar{d}}(Q^2)x^4], \end{aligned} \tag{21}$$

with $x_{C,\bar{d}} = 0.04$. We deemed this value to be the best choice to describe the PDFs by repeating the procedure for a range of $x_{C,\bar{d}}$ and selecting the optimal. Then, the d-quark distribution can be recovered by computing

$$d(x,Q^2) = d_v(x,Q^2) + \bar{d}(x,Q^2). \tag{22}$$

In Fig. 2, we present a comparison between our analytic ML-PDF approximations to $xd_v(x,Q^2)$ (upper plot) and

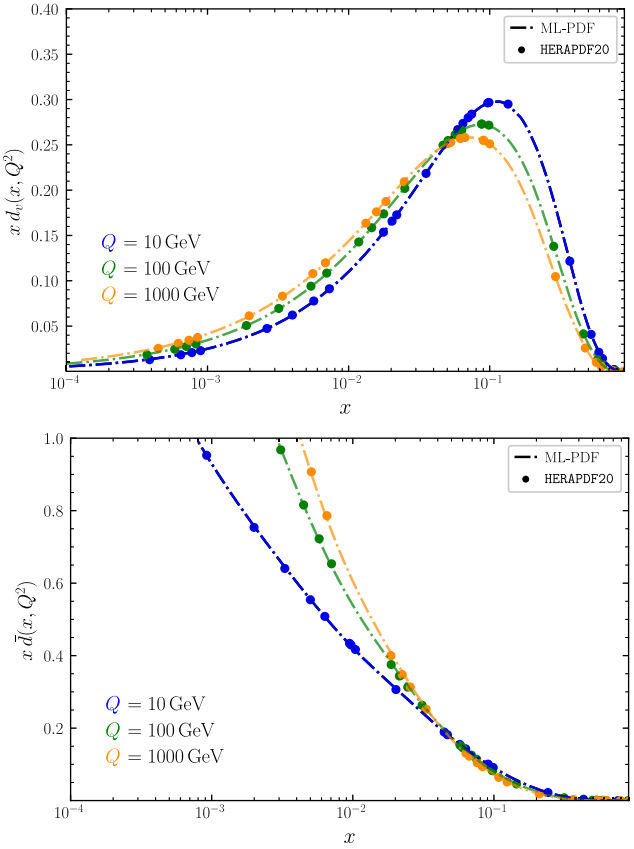

FIG. 2. Comparison between our analytic ML-PDF approximations (dashed lines) and HERAPDF2.0 (solid dots), for xd_v (upper plot) and $x\bar{d}$ (lower plot), at three different values of Q.

$x\bar{d}(x, Q^2)$ (lower plot) and the HERAPDF2.0_NLO_EIG PDF set. We appreciate a very good agreement for different values of Q.

C. Up-quark distributions

The optimal fit was found by considering u_v and $\bar{u}$ distributions. To this end, we proposed

$$\begin{aligned} xu_v(x, Q^2) = & A_{u_v}(Q^2)x^{B_{u_v}(Q^2)}(1-x)^{C_{u_v}(Q^2)} \\ & \times [1 + D_{u_v}(Q^2)x + E_{u_v}(Q^2)x^2 \\ & + F_{u_v}(Q^2)x^3 + G_{u_v}(Q^2)x^4 + H_{u_v}(Q^2)x^5] \end{aligned} \tag{23}$$

to approximate u_v, while a slightly more complicated *Ansatz* was used for $\bar{u}$. In that case, we split the analysis into two regions:

$$R_1 = \{Q_0 \leq Q \leq 10 \text{ GeV}\}, \tag{24}$$

$$R_2 = \{10 \text{ GeV} \leq Q \leq 1000 \text{ GeV}\}. \tag{25}$$

Then, we defined

$$\begin{aligned} x\bar{u}(x, Q^2) = & A_{\bar{u}}(Q^2)x^{B_{\bar{u}}(Q^2)}(1-x)^{C_{\bar{u}}(Q^2)} \\ & \times [1 + D_{\bar{u}}(Q^2)x + E_{\bar{u}}(Q^2)x^2 \\ & + F_{\bar{u}}(Q^2)x^3 + G_{\bar{u}}(Q^2)x^4 + H_{\bar{u}}(Q^2)x^5] \end{aligned}$$

for $Q \in R_1$, and

$$\begin{aligned} x\bar{u}(x, Q^2) = & A_{\bar{u}}(Q^2)x^{B_{\bar{u}}(Q^2)}(1-x)^{C_{\bar{u}}(Q^2)} \\ & \times [1 + D_{\bar{u}}(Q^2)x + E_{\bar{u}}(Q^2)x^3] \\ & - \Theta(x_{C,\bar{u}} - x)A'_{\bar{u}}(Q^2)x^{B'_{\bar{u}}(Q^2)} \\ & \times (1-x)^{C'_{\bar{u}}(Q^2)}[1 + D'_{\bar{u}}(Q^2)x^2] \end{aligned} \tag{26}$$

for $Q \in R_2$. Here, we used the cut $x_{C,\bar{d}} = 0.01$, inspired by the gluon PDF parametrization defined by HERAPDF. After fitting these two distributions, we can define

$$u(x, Q^2) = \bar{u}(x, Q^2) + u_v(x, Q^2) \tag{27}$$

and obtain the u-quark distribution.

We show the comparison between our analytic ML-PDF approximations to $xu_v(x, Q^2)$ (upper plot) and $x\bar{u}(x, Q^2)$ (lower plot) with respect to the HERAPDF2.0_NLO_EIG PDF set in Fig. 3. Again, the agreement is excellent for the complete range of Q considered in this analysis.

D. Strange-quark distribution

For the strange quark, we took the usual assumption that $\bar{s} \equiv s$, a common choice for PDFs at NLO that is also

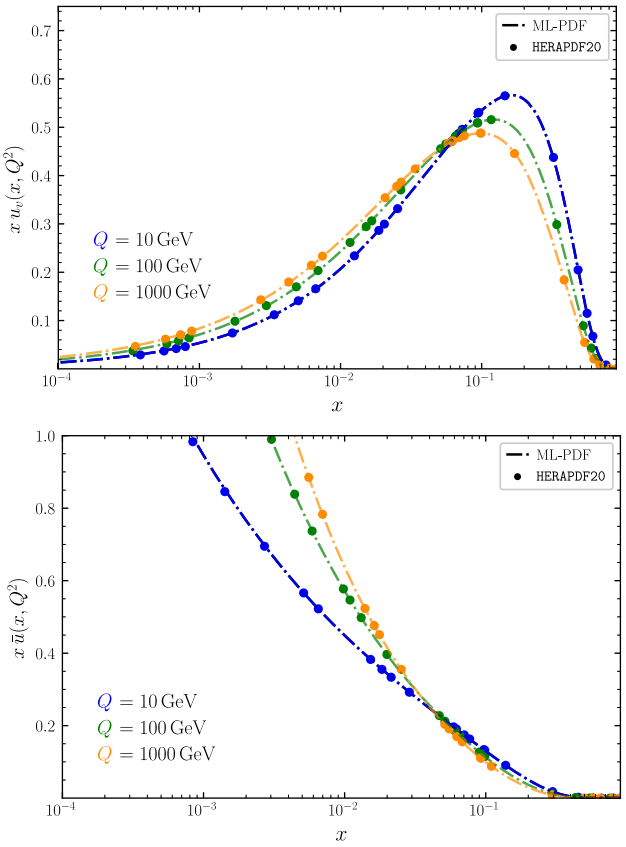

FIG. 3. Comparison between our analytic ML-PDF approximations (dashed lines) and HERAPDF2.0 (solid dots), for xu_v (upper plot) and $x\bar{u}$ (lower plot), at three different values of Q.

imposed for charm and bottom PDFs. After unsuccessfully trying several functional forms based on Euler beta functions, we noticed that the behaviors below and above $Q = 4.5$ GeV were slightly different. Thus, we used this threshold and split the Q analysis into two regions:

$$R_1 = \{Q_0 \leq Q \leq 4.5 \text{ GeV}\}, \tag{28}$$

$$R_2 = \{4.5 \text{ GeV} \leq Q \leq 1000 \text{ GeV}\}. \tag{29}$$

We defined

$$\begin{aligned} xs(x, Q^2) = & A_s(Q^2)x^{B_s(Q^2)}(1-x)^{C_s(Q^2)} \\ & - \Theta(x_{C,s}(R_1) - x)A'_s(Q^2)x^{B'_s(Q^2)} \\ & \times (1-x)^{C'_s(Q^2)} \end{aligned} \tag{30}$$

and

$$\begin{aligned} xs(x, Q^2) = & A_s(Q^2)x^{B_s(Q^2)}(1-x)^{C_s(Q^2)} \\ & - \Theta(x_{C,s}(R_2) - x)A'_s(Q^2)x^{B'_s(Q^2)} \\ & \times (1-x)^{C'_s(Q^2)}[1 + D'_s(Q^2)x^2], \end{aligned} \tag{31}$$

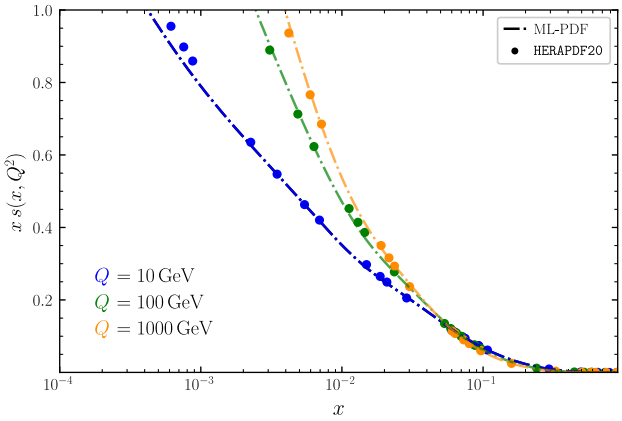

FIG. 4. Comparison between our analytic ML-PDF strange distribution approximation (dashed lines) and HERAPDF2.0 (solid dots). We considered three different values of Q.

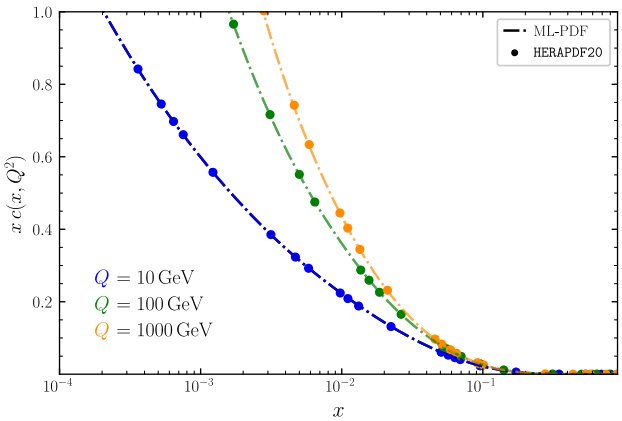

FIG. 5. Comparison between our analytic ML-PDF charm distribution approximation (dashed lines) and HERAPDF2.0 (solid dots). We considered three different values of Q.

where $x_{C,s}(R_1) = 0.25$ and $x_{C,s}(R_2) = 0.1$ in R_1 and R_2, respectively. Once again, these values were determined by exploring a range of values and choosing the best ones (i.e., the ones that gave a smaller value for the integral error).

In Fig. 4, we show our analytic ML-PDF approximation to $xs(x,Q^2)$ (dashed lines) with respect to the corresponding strange PDF from HERAPDF2.0_NLO_EIG (solid dots). At low x (i.e., below 10^{-3}), our fit slightly undershoots the prediction from HERAPDF. Still, the agreement is very good for $x > 10^{-3}$, and in particular, for different values of Q.

E. Charm-quark distribution

As mentioned above, we considered the heavy quarks as generated radiatively by the gluons, and thus fixed their distributions to zero when the scale is below their respective production threshold. Above it, in the case of the charm quark, we split the analysis into two different regions in Q. Explicitly, we considered

$$R_1 = \{1.47\ \mathrm{GeV} < Q \leq 3\ \mathrm{GeV}\}, \tag{32}$$

$$R_2 = \{3\ \mathrm{GeV} \leq Q \leq 1000\ \mathrm{GeV}\}, \tag{33}$$

since $m_c^{\mathrm{pole}} = 1.47$ GeV according to the HERAPDF2.0 fit including NLO QCD corrections [25]. Thus, we defined

$$\begin{aligned} xc(x,Q^2) &= A_c(Q^2)x^{B_c(Q^2)}(1-x)^{C_c(Q^2)}[1+D_c(Q^2)x^2] \\ &\quad - \Theta(x_{C,c}(R_1)-x)A_c'(Q^2)x^{B_c'(Q^2)} \\ &\quad \times (1-x)^{C_c'(Q^2)}[1+D_c'(Q^2)x^2] \end{aligned} \tag{34}$$

and

$$\begin{aligned} xc(x,Q^2) &= A_c(Q^2)x^{B_c(Q^2)}(1-x)^{C_c(Q^2)} \\ &\quad \times (1+D_c(Q^2)x+E_c(Q^2)x^2) \\ &\quad - \Theta(x_{C,c}(R_2)-x)A_c'(Q^2)x^{B_c'(Q^2)} \\ &\quad \times (1-x)^{C_c'(Q^2)}(1+D_c'(Q^2)x^2) \end{aligned} \tag{35}$$

where $x_{C,c}(R_1) = 0.1$ and $x_{C,c}(R_2) = 0.05$ in R_1 and R_2, respectively.

We present a comparison between our analytic ML-PDF approximation to $xc(x,Q^2)$ (dashed lines) versus the values provided by HERAPDF2.0_NLO_EIG (solid dots) in Fig. 5. The agreement is impressive, both in x and in Q.

F. Bottom-quark distribution

Finally, we studied the bottom-quark distribution. Similarly to the case of the charm quark, we started the fitting for values of Q above the mass of the bottom—i.e., $Q > 4.5$ GeV. We noticed that the quality of our fit significantly increased if we divided the Q range into two regions:

$$R_1 = \{4.5\ \mathrm{GeV} \leq Q \leq 15\ \mathrm{GeV}\}, \tag{36}$$

$$R_2 = \{15\ \mathrm{GeV} \leq Q \leq 1000\ \mathrm{GeV}\}. \tag{37}$$

As we did for the other flavors, we then proposed

$$\begin{aligned} xb(x,Q^2) &= A_b(Q^2)x^{B_b(Q^2)}(1-x)^{C_b(Q^2)}[1+D_b(Q^2)x^2] \\ &\quad - \Theta(x_{C,b}(R_1)-x)A_b'(Q^2)x^{B_b'(Q^2)} \\ &\quad \times (1-x)^{C_b'(Q^2)}[1+D_b'(Q^2)x^2] \end{aligned} \tag{38}$$

and

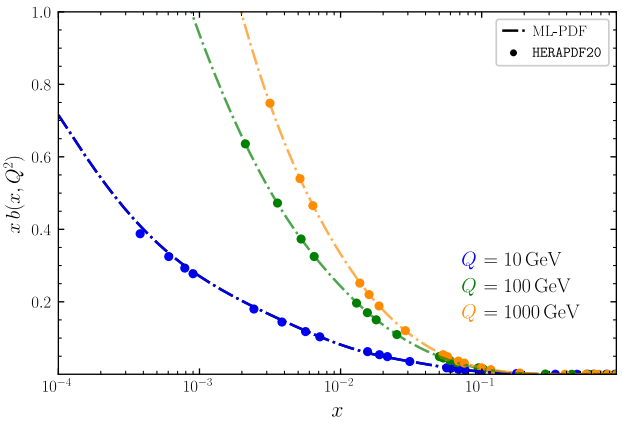

FIG. 6. Comparison between our analytic bottom ML-PDF approximation (dashed lines) and HERAPDF2.0 (solid dots). We considered three different values of Q.

$$\begin{aligned} xb(x,Q^2) = {} & A_b(Q^2)x^{B_b(Q^2)}(1-x)^{C_b(Q^2)} \\ & \times [1 + D_b(Q^2)x + E_b(Q^2)x^2] \\ & - \Theta(x_{C,b}(R_2) - x)A'_b(Q^2)x^{B'_b(Q^2)} \\ & \times (1-x)^{C'_b(Q^2)}[1 + D'_b(Q^2)x^2], \end{aligned} \tag{39}$$

where $x_{C,b}(R_1) = 0.1$ and $x_{C,b}(R_2) = 0.05$ in R_1 and R_2, respectively.

As for the other flavors, in Fig. 6, we show our analytic ML-PDF approximation to $xb(x, Q^2)$ (dashed lines) with respect to the corresponding bottom PDF from HERAPDF2.0_NLO_EIG (solid dots). The agreement is very good for different values of Q, from 10 to 500 GeV. Even if some discrepancies arise for $x < 5 \times 10^{-4}$, we can state that our formulas globally provide a reliable approximation to the bottom PDF.

V. EFFICIENCY AND QUALITY BENCHMARKS

In this section, we discuss quantitatively the quality of the ML-PDF approximations found in Sec. IV and compare the time required to compute some chosen observables.

In order to estimate the discrepancies between our analytic ML-PDFs and the original HERAPDF2.0 distributions in a phenomenologically relevant way, we rely on our definition of integral error given in Eq. (10). In Fig. 7, we show that the integral error for almost all the distributions (i.e., u_v, d_v, $\bar{u}$, $\bar{d}$, c, b, and g) is below 0.5% in the $Q \in [Q_0, 1000\ \text{GeV}]$ range. For the strange-quark PDF, we appreciate a larger deviation at higher values of Q, reaching up to 1.5% error for $Q \approx 1000$ GeV. Although it is not present in the plots, we want to highlight that the integral error for u and d is also well under control, being below the percent level for the whole range of Q values explored. This is expected from Eqs. (22) and (27), propagating the errors shown in Fig. 7.

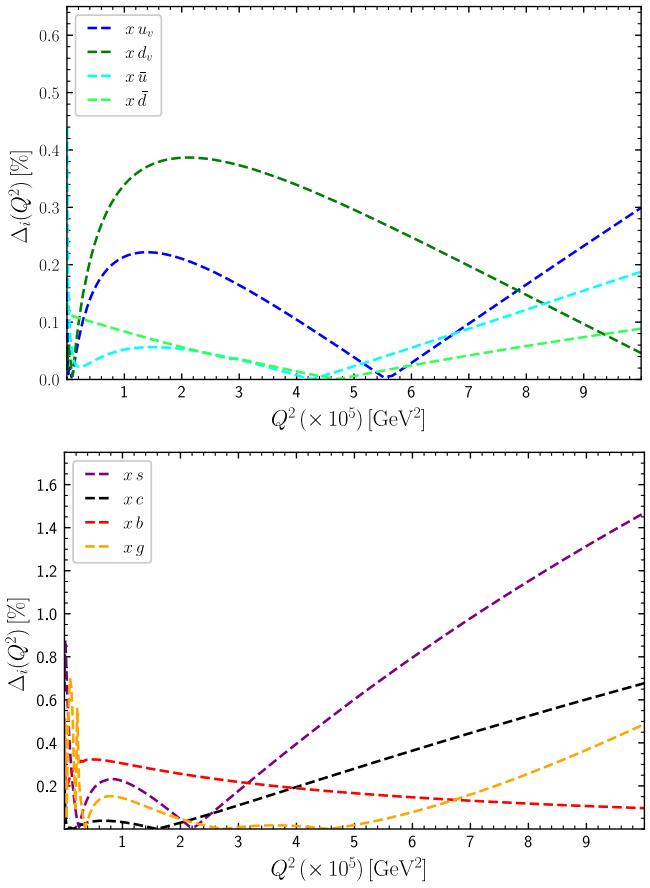

FIG. 7. Integral error for our analytic ML-PDF approximations with respect to HERAPDF distributions. In the upper plot, we show up- and down-quark distributions, while s, c, b, and gluon PDFs are presented in the lower plot.

As we explained in Sec. III, the integral error is expected to provide a more reliable estimation on the impact of using our ML-PDFs instead of the original PDF set in a physical calculation. Still, for the sake of completeness, we present here an analysis of the approximation error as a function of x. For this purpose, we sample with an exponential distribution the range $Q \in [10\ \text{GeV}, 1000\ \text{GeV}]$ using $N = 50{,}000$ random points. Then, we define

$$\bar{\Delta}_i(x) = \frac{1}{N}\sum_{j=1}^{N}\left|1 - \frac{f_i^{\text{ML}}(x, Q_j^2)}{f_i^{\text{HERA}}(x, Q_j^2)}\right| \tag{40}$$

as an estimator of the error in x. This definition is similar to the *error shape* mentioned in Sec. III, and it corresponds to an average of the relative errors as a function of x. In Fig. 8, we show the results for up and down quarks (upper plot), for strange, charm, and bottom quarks, and for the gluon (lower plot). For u_v, d_v, $\bar{u}$ and $\bar{d}$ quarks, $\bar{\Delta}$ is below 10% for $x \in [10^{-3}, 0.3]$. Similarly, for the other quark flavors, the error is below 20% in $x \in [10^{-3}, 0.3]$. For the gluon, the error is slightly larger, $\mathcal{O}(10\%\text{–}25\%)$ in the central

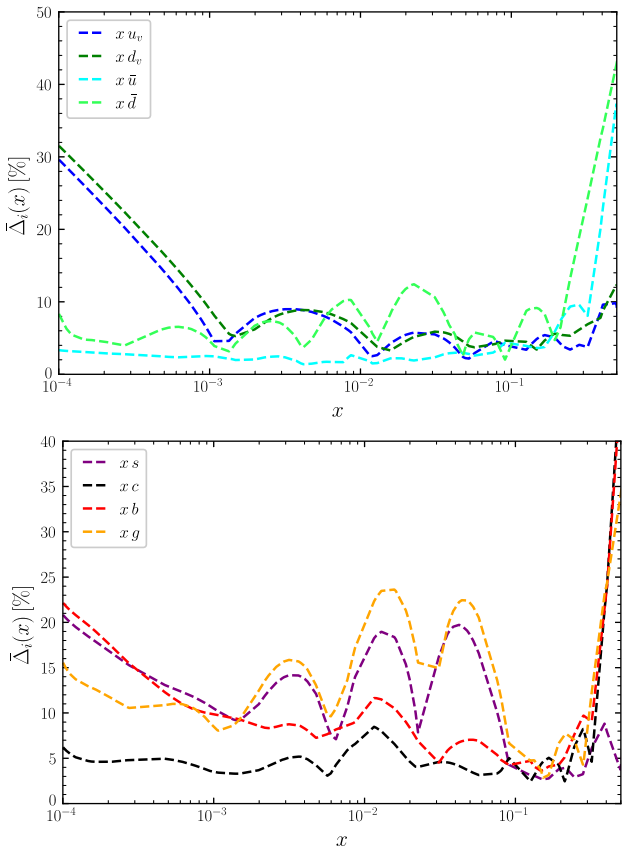

FIG. 8. Estimation of the error as a function of x, averaged over $Q \in [10\ \text{GeV}, 1000\ \text{GeV}]$ for our analytic ML-PDF approximations with respect to HERAPDF2.0 distributions. In the upper plot, we show up- and down-quark distributions, while s, c, b, and gluon PDFs are presented in the lower plot.

x region. All the distributions tend to increase their average relative error for $x > 0.3$–0.4, since the PDFs decrease a couple of orders of magnitude for larger values of x. Also, the valence distributions show this behavior in the low-x region. In both cases, small absolute discrepancies translate into large relative fluctuations. This effect is shadowed in the definition of integral error, because the contributions of the PDFs in those regions are rather small.

Another important observation is the presence of oscillations in the relative error in the region $x \approx 0.05$–0.1, particularly for distributions that made use of a *gluon-like* functional form. This behavior is due to the matching of two Euler beta functions around $x_C = \mathcal{O}(0.1)$, which originates a fluctuation, since they have different signs. In fact, this justifies the reduced integral error, because these fluctuations around the original PDF cancel at integrand level.

Regarding the time needed for the use of the ML-PDFs and LHAPDF, we found a significant difference between simply calling the distributions and actually implementing them in a cross-section calculation. The results presented in the rest of this section were obtained using a desktop PC with a 16-core Intel i7-13700 processor. It is important to remark that nothing has been parallelized.

TABLE I. Comparison of the time (in seconds) required to compute N_{points} evaluations of HERAPDF2.0 within the LHAPDF framework, and our ML-PDF analytic approximation.

N_{points}	LHAPDF (s)	ML-PDFs (s)	Gain(%)
10^3	3.76×10^{-2}	2.92×10^{-4}	99.22
10^4	4.20×10^{-2}	2.50×10^{-3}	94.05
10^5	8.94×10^{-2}	2.50×10^{-2}	72.10
10^6	0.56	0.25	55.46
10^7	5.25	2.50	52.49
10^8	52.04	24.92	52.11

A. Run-time

We start by presenting, in Table I, the run-time difference when performing a call to our PDFs and LHAPDF, for sets of points in the (x, Q^2) space. As we can appreciate, for $N_{\text{points}} \leq 10^{3\text{–}4}$, the gain is quite substantial. This is due to the fact that calling LHAPDF has an overhead time for loading and reading the grid, which albeit small, dominates the total run-time when one evaluates a small number of points. Therefore, if we were interested in evaluating a handful of points we would be wise in choosing our ML-PDFs over LHAPDF. As we move to larger number of points $N_{\text{points}} \approx 10^5$, the weight of this overhead starts to dilute, and once we pass the mark $N_{\text{points}} \approx 10^6$, we reach a region where we are essentially comparing only the execution time of interpolation versus evaluation. In the last column, we display the time gain percentage, defined as

$$\text{gain}(\%) = 100 \times \frac{\text{time}_{\text{LHAPDF}} - \text{time}_{\text{ML-PDF}}}{\text{time}_{\text{LHAPDF}}}. \quad (41)$$

A negative gain would mean that the interpolation is faster than the direct evaluation. As can be seen from the table, the gain seems to approach a plateau around 50% (i.e., our ML-PDF are 2 times faster than LHAPDF), which is quite sizable.

From these numbers it appears that, for a code requiring Monte Carlo integration, it would be beneficial to use something akin to our ML-PDFs. We further investigate this in the next subsection.

B. Validity of the sum rules

One important physically motivated cross-check of our analytic ML-PDFs is the sum rules. For any energy scale Q, the sum rules are given by

$$S_1(Q^2) = \int_0^1 dx\, u_v(x, Q^2) = 2, \quad (42)$$

$$S_2(Q^2) = \int_0^1 dx\, d_v(x, Q^2) = 1 \quad (43)$$

for the up and down valence quarks, and

$$\int_0^1 dx[f_i(x,Q^2) - \bar{f}_i(x,Q^2)] = 0 \tag{44}$$

for all other quarks. Also, the PDFs must fulfill the condition that the total momenta carried by the constituent partons equal the momentum of the hadron. This is encoded in the momentum sum rule

$$S_3(Q^2) = \int_0^1 dx\, x\left[g(x,Q^2) + \sum_{i\in\{q,\bar{q}\}} f_i(x,Q^2)\right] = 1, \tag{45}$$

where the sum is carried out over all the active flavors of quarks. These constraints are often imposed when performing the PDF extraction from fits to experimental data. In the case of HERAPDF2.0, Eq. (44) is automatically fulfilled by construction (no $q-\bar{q}$ distinction for sea quarks), and this also holds for our analytic ML-PDFs.

In Fig. 9, we compare the results of Eqs. (42) and (43) for HERAPDF2.0 (blue) and our analytic ML-PDFs (orange) with

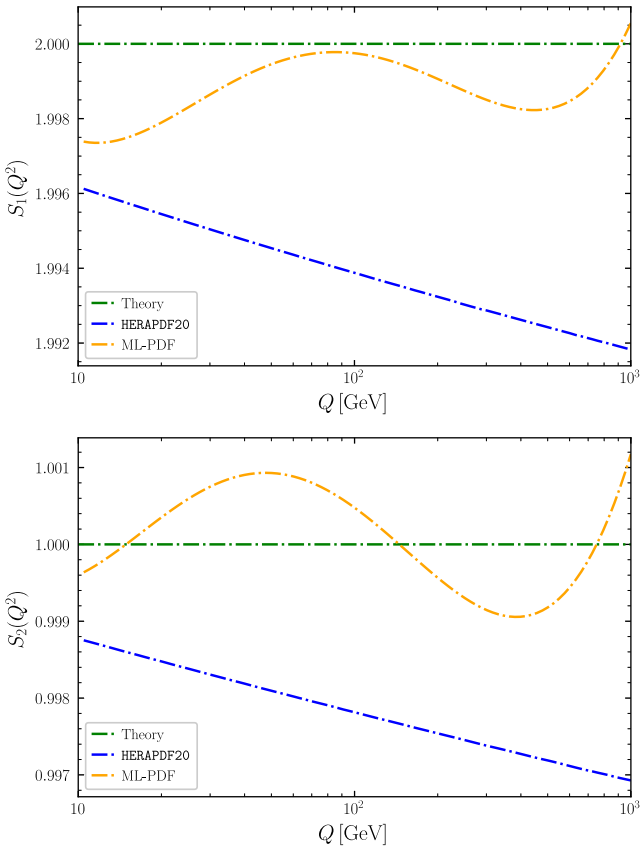

FIG. 9. Sum rules for up (upper) and down (lower) valence distributions, as a function of Q. We show the values calculated with HERAPDF2.0 (blue) and our analytic ML-PDF (orange), and also the theoretical exact value (green).

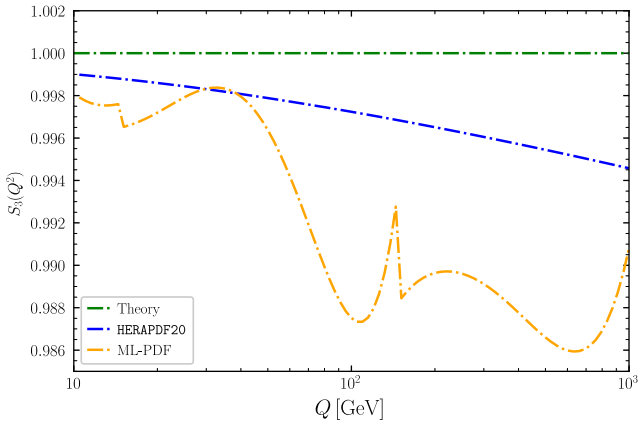

FIG. 10. Total momentum conservation rule, as a function of Q. We show the values calculated with HERAPDF2.0 (blue) and our analytic ML-PDF (orange), and also the theoretical exact value (green).

respect to the theoretical value (green line). In both cases, we consider $Q \in [10\ \text{GeV}, 1000\ \text{GeV}]$, and we find that our analytic ML-PDF leads to results closer to the theoretical value. Concretely, for up valence quarks, we get deviations of $\mathcal{O}(0.1\%)$ for ML-PDFs and $\mathcal{O}(0.2\%)$ for HERAPDF2.0, showing a trend for the latter to depart from the theoretical value in the high-Q region. Similarly, for down valence distributions, the predictions obtained with ML-PDFs oscillate around 1 within a band of $\mathcal{O}(0.2\%)$, while HERAPDF2.0 deviates more than $\mathcal{O}(0.3\%)$ for higher Q values.

Finally, in Fig. 10, we show the results of computing Eq. (45) with our analytic ML-PDFs (orange) and HERAPDF2.0 (blue). In this case, HERAPDF2.0 better approaches the theoretical value, with discrepancies smaller than $\mathcal{O}(0.5\%)$. The predictions obtained with our analytic ML-PDFs show deviations of $\mathcal{O}(0.6\%)$ below 60 GeV, reaching up to $\mathcal{O}(1.5\%)$ for $Q \approx 500$ GeV. This behavior is driven by gluon PDF, since this parton carries more than 45% of proton momentum and our fit shows an increasing error for large Q (see Figs. 7 and 8). In any case, as we emphasized in the Introduction, this constitutes a first proof of concept, and we can claim that our ML-PDFs successfully passed all the cross-checks with percent-level (and even subpercent-level) precision.

C. Impact in physical observables

Here, we discuss the impact of using our analytic approximations to the PDFs in two realistic cross-section calculations.

We start by considering pion production in unpolarized proton-proton collisions at $\sqrt{S_{\text{cm}}} = 7$ TeV in the central rapidity region ($|\eta^\pi| < 0.5$). These simulations make use of a sequential code based on Ref. [26], modified to use the needed PDFs and fragmentation functions (FFs). The Monte-Carlo integrator is VEGAS, with 10^5 points, while the FF set used is DEHSS2014 [27], and all factorization and

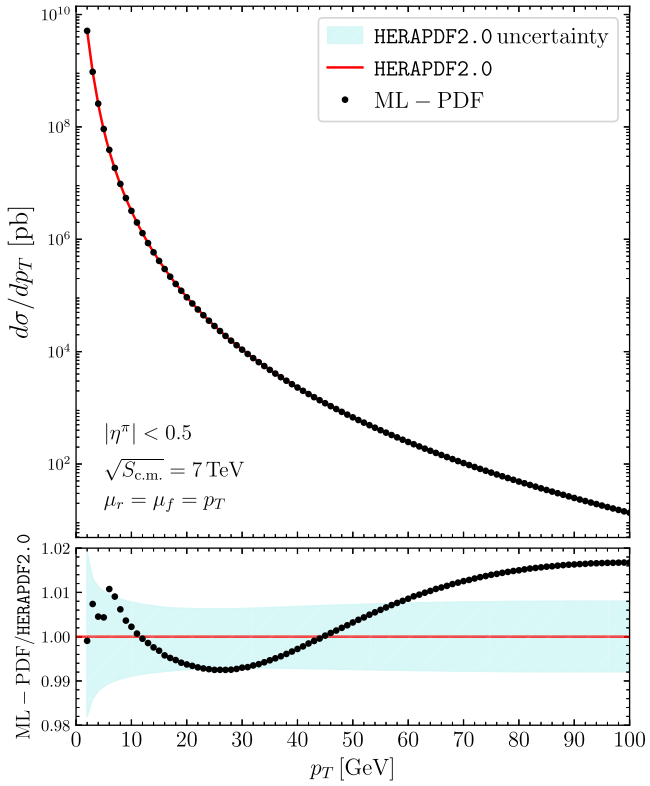

FIG. 11. Upper plot: differential cross section for the process $p + p \to \pi$ as a function of p_T, including up to NLO QCD corrections. We run the simulation using our analytic ML-PDFs (dots) and LHAPDF with HERAPDF2.0 (solid). Lower plot: ratio between the two cross sections. The discrepancies in the results are well below 2%, which is even less than the error introduced by scale variations (not included in this plots).

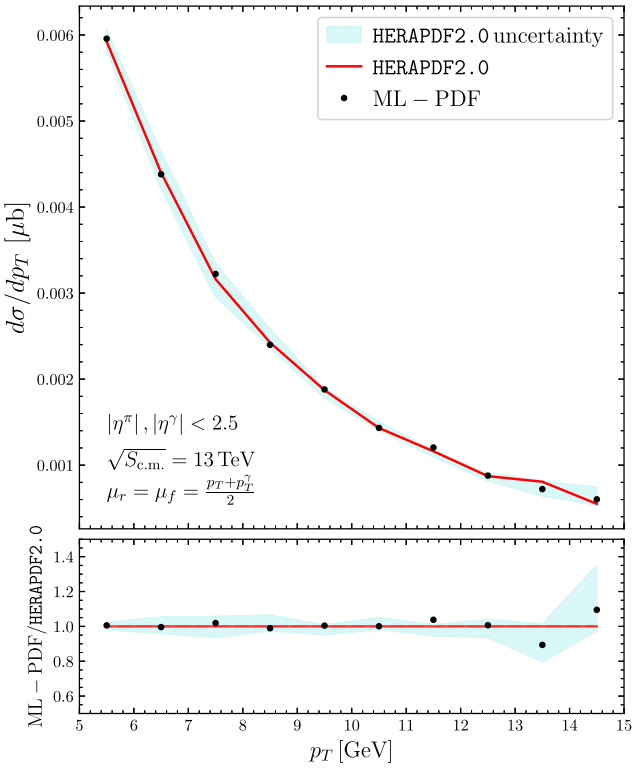

FIG. 12. Upper plot: differential cross-section for the process $p + p \to \gamma + \pi$ as a function of the transverse momentum of the pion (p_T), including up to NLO QCD corrections. We run the simulation using our analytic ML-PDFs (dots) and LHAPDF with HERAPDF2.0 (solid). Lower plot: ratio between the two cross sections.

renormalization scales have been taken to be equal to the transverse momentum of the pion (denoted p_T). In the upper plot of Fig. 11, we present the p_T spectrum using our ML-PDFs (black dots) and HERAPDF2.0 (red line), while the lower plot depicts the ratio between those two quantities. The (light blue) uncertainty band in the lower plot comes from the Hessian set of theoretical uncertainties of HERAPDF2.0; it is also present (but hardly noticeable) in the upper plot. From the lower panel, we can appreciate that the difference between the two simulations is well below 2%, and it mostly stays within the PDF uncertainty bands. This is much smaller than the typical perturbative error of NLO QCD calculations [generally larger than $\mathcal{O}(20\%)$]. Regarding the CPU time, it took 10 minutes, 28 seconds using LHAPDF and 9 minutes, 19 seconds with our ML-PDFs, running on the same system described before. From Table I, and considering the number of iterations used, one would expect a much larger gain. The outcome of this run serves to emphasize that the matter is far more subtle. It is worth highlighting that this reduction in the computational time is only due to our optimized ML-PDF, since all the other ingredients of the calculation remain the same in both simulations.

In the second example, we turned to a simulation that required a significantly larger amount of trials to reach a stable output. We tested our ML-PDFs with the calculation of $p + p \to \gamma + \pi$ including up to NLO QCD effects [28,29] at $\sqrt{S_{\text{cm}}} = 13$ TeV. We consider the central rapidity region, $|\eta^\pi|, |\eta^\gamma| \leq 2.5$, imposing $p_T^\gamma \in [30\ \text{GeV}, 1500\ \text{GeV}]$. We used the same FFs as before, and we took the renormalization and factorization scales to be equal to

$$\mu = \frac{p_T + p_T^\gamma}{2}. \tag{46}$$

In this case, we chose to run 10^8 points, both in order to be in what we expect (from Table I) to be a very low time-gain region and because the observable requires a large number of points to produce physical results. We show the outcome in Fig. 12. As before, the upper plot presents the cross section, while the lower plot demonstrates the ratio between the two computations. Given the p_T range explored, there are some oscillations of the output associated with statistical fluctuations. In fact, using 10^7 points gave unphysical results for some p_T values. We can then say that this observable would greatly benefit from using a higher

TABLE II. Comparison of the time (in seconds) required to compute two observables using LHAPDF and ML-PDFs. See text for details.

Obs	LHAPDF (s)	ML-PDFs (s)	Gain(%)
$p + p \to \pi$	628.320	558.854	11.06
$p + p \to \gamma + \pi$	12452.273	8671.827	30.36

number of points. For our purpose, 10^8 points is enough. In the $p_T \leq 10$ GeV region, which dominates the cross section, the differences are within 5% and are feasible for further reduction. The runtime for LHAPDF was 207.5 minutes, or almost 3.45 hours, while the ML-PDF run took 144.5 minutes, or more than a full hour less. We summarize the runtime of the cross sections in Table II. This gain is much larger than in the previous test, and far more than expected from the table. Also, we notice that the predictions obtained with our ML-PDFs are compatible with the propagation of errors of HERAPDF2.0 into the observable (light blue band).

To conclude, the results of these two realistic calculations clearly support the potential of our ML-PDFs to reduce the computational cost of the simulations, keeping under control the uncertainties of the approximation.

VI. CONCLUSIONS

In this work, we provide an analytic approximation to PDFs using machine-assisted techniques to adjust both their x and Q dependence. Our starting hypothesis was the assumption that the x dependence could be reproduced by Eulerian-like functions, while all the Q dependence is embodied within the coefficients.

By doing so, we obtain a reliable approximation to HERAPDF2.0, taking into consideration up to NLO QCD corrections. We show that the integral error is under control for an ample range of x and Q values. In fact, our ML-PDFs were tested for $Q \in [Q_0, 1000\ \text{GeV}]$, presenting (for most of the distributions) deviations with respect to HERAPDF2.0 below the percent level. It is important to highlight that this is comparable with the error of the PDF sets themselves, and far smaller than the uncertainties introduced by truncating the perturbative expansion of observables in QCD. In fact, we find that the error induced by the ML-PDF in $p + p \to \pi$ is $\mathcal{O}(1\%)$, while for $p + p \to \gamma + \pi$, it is $\mathcal{O}(5\%)$ [30]. These errors are much smaller than the $\mathcal{O}(20\%–50\%)$—or even larger—uncertainties arising from scale variations. Furthermore, we found a non-negligible reduction of the runtime: $\mathcal{O}(11\%)$ for $p + p \to \pi$, and $\mathcal{O}(30\%)$ for $p + p \to \gamma + \pi$.

To conclude, we want to emphasize that the strategy explained in this article is fully applicable to any PDF set, and even to FFs, at any perturbative order (NLO, NNLO, and beyond), due to the fact that we made no assumptions about the order at which DGLAP evolution is truncated. The only required ingredient is a previously existent PDF, extracted from data, to be used as input in our formalism. Our only hypothesis is that the Q dependence is fully embodied within the coefficients $\{A, B, \ldots\}$ described in Sec. IV. Following this approach, we can avoid using interpolation methods to evaluate the distributions and speed up the calculation of PDFs (and, eventually, FFs), because we obtain a fit that already embodied the evolution in an analytically approximated way. This not only brings a reduction of the CPU time, but also leads to a reduction of the CO_2 footprint and points toward more efficient and sustainable practices in high-energy physics (HEP) [2]. For this reason, further developments to obtain analytic approximations to PDFs/FFs are expected, and this work can be considered as a first proof of concept, paving the way for future improvements. As a final remark, we want to stress that our approach does not replace the PDF extractions from data; in fact, they constitute an essential input and the basis for ML-PDF determinations such as the one presented here.

ACKNOWLEDGMENTS

We would like to thank L. Cieri, V. Mateu, and G. Rodrigo for fruitful comments about the manuscript. This work is supported by Spanish Government (Agencia Estatal de Investigación MCIN /AEI/10.13039/501100011033) Grants No. PID2020–114473GB-I00, No. PID2022-141910NB-I00 and Generalitat Valenciana Grants No. PROMETEO/2021/071 and No. ASFAE/2022/009. The work of D. F. R.-E. is supported by Generalitat Valenciana (CIGRIS/2022/145). S. A. O.-O and R. J. H.-P. are funded by CONAHCyT through Project No. 320856 (Paradigmas y Controversias de la Ciencia 2022) and Ciencia de Frontera 2021-2042. R. J. H.-P. is also funded by Sistema Nacional de Investigadores from CONAHCyT. P. Z. acknowledges support, at the early stages of this work, from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Research Unit FOR 2926, Grant No. 430915485. P. Z. is funded by the "Atracción de Talento" Investigador program of the Comunidad de Madrid (Spain) No. 2022-T1/TIC-24024. The work of G. S. was partially supported by EU Horizon 2020 research and the innovation program STRONG-2020 project under Grant Agreement No. 824093 and the H2020-MSCA-COFUND USAL4EXCELLENCE-PROOPI-391 project under Grant Agreement No. 101034371.

[1] G. Heinrich, Phys. Rep. **922**, 1 (2021).
[2] S. Banerjee *et al.*, arXiv:2306.02837.
[3] J. C. Collins, D. E. Soper, and G. F. Sterman, Adv. Ser. Dir. High Energy Phys. **5**, 1 (1989).
[4] S. Catani, D. de Florian, and G. Rodrigo, J. High Energy Phys. 07 (2012) 026.
[5] L. Cieri, P. K. Dhani, and G. Rodrigo, arXiv:2402.14749.
[6] G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).
[7] A. Simonelli, arXiv:2401.13663.
[8] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Eur. Phys. J. C **75**, 580 (2015).
[9] This also applies, with some differences, to other non-perturbative quantities feasible of phenomenological studies in pQCD, such as fragmentation functions, fracture functions, transverse-momentum-dependent distributions, etc.
[10] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 341 (2021).
[11] T.-J. Hou *et al.*, Phys. Rev. D **103**, 014013 (2021).
[12] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, Phys. Rev. D **96**, 014011 (2017).
[13] I. Helenius, M. Walt, and W. Vogelsang, Phys. Rev. D **105**, 094031 (2022).
[14] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, Phys. Rev. D **85**, 074028 (2012).
[15] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, Eur. Phys. J. C **82**, 413 (2022).
[16] P. Duwentäster, L. A. Husová, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K. F. Muzakka, F. I. Olness, I. Schienbein, and J. Y. Yu, Phys. Rev. D **104**, 094005 (2021).
[17] I. Borsa, R. Sassot, D. de Florian, M. Stratmann, and W. Vogelsang, Phys. Rev. Lett. **129**, 012002 (2022).
[18] I. Borsa, M. Stratmann, D. de Florian, and R. Sassot, Phys. Rev. D **109**, 052004 (2024).
[19] R. D. Ball *et al.* (NNPDF Collaboration), Eur. Phys. J. C **82**, 428 (2022).
[20] J. Butterworth *et al.*, J. Phys. G **43**, 023001 (2016).
[21] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, Eur. Phys. J. C **75**, 132 (2015).
[22] R. S. Thorne, Phys. Rev. D **86**, 074017 (2012).
[23] W. Research, NonlinearModelFit, https://reference.wolfram.com/language/ref/NonlinearModelFit.html (2008).
[24] S. Ochoa-Oregon, D. Rentería-Estrada, R. Hernández-Pinto, G. Sborlini, and P. Zurita, Ancillary files for "Using analytic models to describe effective PDFs," Zenodo (2024), https://doi.org/10.5281/zenodo.12745978.
[25] Z. Zhang (H1 and ZEUS Collaborations), Acta Phys. Pol. B Proc. Suppl. **8**, 957 (2015).
[26] B. Jager, A. Schafer, M. Stratmann, and W. Vogelsang, Phys. Rev. D **67**, 054005 (2003).
[27] R. J. Hernández-Pinto, M. Epele, D. de Florian, R. Sassot, and M. Stratmann, J. Phys. Conf. Ser. **761**, 012037 (2016).
[28] D. de Florian and G. F. R. Sborlini, Phys. Rev. D **83**, 074022 (2011).
[29] D. F. Rentería-Estrada, R. J. Hernández-Pinto, and G. F. R. Sborlini, Symmetry **13**, 942 (2021).
[30] In this work, we neglect to study the uncertainty in the parameters of the ML-PDFs. Such a study is beyond the scope of the present work, and we leave it to future studies.