

# Analysis of HERA data with a PDF parametrization inspired by quantum statistical mechanics

---

Marco Bonvini,<sup>a</sup> Franco Buccella,<sup>a</sup> Francesco Giuli,<sup>b</sup> Federico Silvetti<sup>c</sup>

<sup>a</sup>*INFN, Sezione di Roma 1, Piazzale Aldo Moro 5, 00185 Roma, Italy*

<sup>b</sup>*CERN, 1211 Geneva 23, Switzerland*

<sup>c</sup>*Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, UK*

*E-mail:* [marco.bonvini@roma1.infn.it](mailto:marco.bonvini@roma1.infn.it), [franco.buccella@roma1.infn.it](mailto:franco.buccella@roma1.infn.it),  
[francesco.giuli@cern.ch](mailto:francesco.giuli@cern.ch), [federico.silvetti@durham.ac.uk](mailto:federico.silvetti@durham.ac.uk)

ABSTRACT: We present a determination of the parton distribution functions (PDFs) of the proton from HERA data using a PDF parametrization inspired by a quantum statistical model of the proton dynamics. This parametrization is characterised by a very small number of parameters, yet it leads to a reasonably good description of the data, comparable with other parametrizations on the market. It may thus provide an alternative to standard parametrizations, useful for studying parametrization bias and to possibly simplify the fit procedure thanks to the small number of parameters. Interestingly, the model reproduces key physical features, such as a  $\bar{d}$  distribution larger than  $\bar{u}$ , that HERA data alone are not able to constrain when using more flexible parametrizations. Moreover, polarized distributions are described in the model by the same parameters of the unpolarized ones, giving us the possibility of extracting both types of distributions within the same fit.

ARXIV EPRINT: [2311.08785](https://arxiv.org/abs/2311.08785)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The PDF parametrization from the statistical model</b>	<b>2</b>
<b>3</b>	<b>Setup of the fit and benchmark</b>	<b>5</b>
<b>4</b>	<b>Fit with the new parametrization</b>	<b>7</b>
<b>5</b>	<b>More flexible QSPDF parametrization</b>	<b>12</b>
<b>6</b>	<b>Implications</b>	<b>17</b>
6.1	Comparison with previous model determinations	17
6.2	The anti-up anti-down asymmetry in the proton	19
6.3	Additional data	21
6.4	Polarized PDFs	22
6.5	Future directions	25
<b>7</b>	<b>Conclusions</b>	<b>27</b>
<b>A</b>	<b>Implementation of the sum rules</b>	<b>28</b>
	<b>References</b>	<b>29</b>

---

## 1 Introduction

Parton distributions functions (PDFs) describe the longitudinal momentum fraction  $x$  distributions of partons (quarks and gluons) inside the proton and they are a key ingredient for the theoretical description of collisions with protons in the initial state. For this reason, in the Large Hadron Collider (LHC) era, a huge effort from both the theory and experimental communities to improve their determination took place. PDFs parametrize a low-scale, non-perturbative dynamics of the proton, and cannot thus be determined using perturbation theory. Therefore, PDFs are usually fitted from data, mostly coming from the HERA collider deep inelastic scattering (DIS) experiments, but also with ever increasing LHC inputs.

PDF fits are performed by several groups [1–8] and differ by many aspects: from the theory description of the data to the technicalities of the fitting procedure, from the datasets included in the fit<sup>1</sup> to the evaluation of uncertainties. One distinctive aspect of the various PDF fits is the choice of the functional form used to parametrize the  $x$  dependence of PDFs at the initial scale (at any other scale PDFs are obtained by perturbative DGLAP evolution). Some PDF fits use very few parameters, e.g. HERAPDF depends on 14 free parameters [1], while other fits use very flexible parametrizations, e.g. NNPDF with hundreds of parameters [5]. The use of so many parameters in the NNPDF framework is possible thanks to the application of machine learning techniques to the

---

<sup>1</sup>In particular, some PDF fits are based on a global set of experimental data, including (collider and fixed-target) DIS, (collider and fixed-target) Drell-Yan, jet production,  $V$ +jet, etc., while other fits are based on more limited sets, e.g. HERAPDF [1] uses only HERA DIS data.

minimisation of the  $\chi^2$ . However, most PDF fitting groups use more traditional techniques, which are unable to deal with so many parameters. In these traditional fitting frameworks, having flexible parametrizations with a small number of parameters is a value.

In this work we consider a parametrization inspired by a quantum statistical model of the proton dynamics. This parametrization is characterised by a very small number of parameters. We use it to fit PDFs from the combined HERA data, using next-to-next-to-leading order (NNLO) theory without and with the inclusion of small- $x$  resummation, and it leads to a reasonably good fit despite its limited flexibility, somewhat comparable with other parametrizations on the market. We believe that this parametrization can be used as a complement to more standard ones, both to study parametrization bias and perhaps to facilitate the fit (having few parameters). For accurate PDF determination, we also consider adding parameters to increase the flexibility of the parametrization, especially at small  $x$ . With just two extra parameters, the quality of the fit becomes competitive with standard parametrizations.

Interestingly, physical features such as a  $\bar{d}$  distribution larger than  $\bar{u}$  come out automatically from the chosen functional form, even though the HERA data alone are not able to constrain them. We verify that this property is stable upon inclusion of additional data. Moreover, the physical model can also describe polarized parton distributions with the same parameters. If we trust the model, this feature gives the possibility to fit simultaneously polarized and unpolarized distributions without the need of introducing new degrees of freedom. We investigate this possibility finding very promising results.

The paper is structured as follows. In section 2 we introduce the parametrization from the statistical model. In section 3 we discuss the setup of our fit and introduce a benchmark fit with parametrization à la HERAPDF. In section 4 we test the parametrization against HERA data and compare with our benchmark and with other public PDFs. In section 5 we introduce a more flexible version of the parametrization which improves the fit quality and study its model uncertainty and its comparison with other parametrizations. In section 6 we discuss our result in view of the physical model behind it, and consider possible implications and future directions. We conclude in section 7.

## 2 The PDF parametrization from the statistical model

A field-theoretical computation of PDFs requires dealing with non-perturbative dynamics, and it is therefore very difficult to achieve. Even numerical techniques like lattice QCD are not (yet<sup>2</sup>) able to satisfactorily determine PDFs, essentially because PDFs are defined in terms of bilocal operators separated by a light cone distance which cannot be described on a Euclidean lattice [11].

Usually, PDFs are determined by fitting them to data using an arbitrary parametrization of the  $x$  dependence of the PDFs at an “initial” scale  $\mu_0$ , typically chosen at the border between perturbative and non-perturbative QCD ( $\mu_0 \sim 1$  GeV). PDFs are then evolved at different scales solving the perturbative DGLAP equation [12–14]. The goal in choosing the parametrization is usually to minimise the bias with sufficiently flexible functional forms while at the same time keeping the number of parameters small enough (for better numerical performances and to avoid overfitting). A notable exception is the parametrization used by the NNPDF collaboration, which uses neural networks to parametrize PDFs with hundreds of parameters: this makes any bias completely negligible, but requires the use of sophisticated machine learning techniques to perform the fit and to determine the uncertainties [5, 15].

In Ref. [16] (see also Refs. [17–36]) a different approach has been proposed, where the functional form of the PDFs are obtained through a statistical model of the parton dynamics in the proton. The

---

<sup>2</sup>Attempts to determine PDFs in lattice QCD make use of alternative definitions of PDFs, denoted quasi-PDFs [9] and pseudo-PDFs [10], which are well defined in Euclidean space-time and are related in some limit to ordinary light-cone PDFs.

model assumes that the partons could be treated as massless particles forming an ideal quantum gas at equilibrium at the initial scale  $\mu_0$  in a finite volume, characterised by an effective temperature. Working in the infinite momentum frame (à la Feynman), the transverse degrees of freedom can be neglected and the dynamics can be described in terms of the longitudinal momentum fraction  $x$ . As a result, the model predicts a (very biased) functional form for the PDFs in terms of very few parameters.

Because the model does not take into account the QCD interaction, information like the factorization scheme is absent in the parametrization. This makes the model clearly incomplete. For instance, the resulting PDFs do not contain information on the scale at which they are supposed to be computed, as the scale dependence is a consequence of the factorization of collinear singularities in QCD. We thus do not expect the model to give a reliable description per se, but we want to investigate if it can provide a suitable baseline for a parametrization.

To this end, the plain model must be supplemented with “phenomenological” modifications. Some of them were introduced already in the original publications [16–18], others have been considered later on [20–27, 29, 30, 34]. Crucially, the various incarnations of the model describe separately the individual polarizations of quarks, thus providing in principle a description of polarized and unpolarized PDFs in terms of the same parameters.

In this work, we will consider the simple parametrization proposed in Ref. [18]. More flexible functional forms depending on more parameters may provide a better description of the data, but in this analysis we want to keep the number of parameters as small as possible. Let us first introduce the function

$$h_{\pm}(x; b, X) = \frac{x^b}{\exp\left(\frac{x-X}{\bar{x}}\right) \pm 1}, \quad (2.1)$$

which descends from Fermi-Dirac ( $h_+$ ) and Bose-Einstein ( $h_-$ ) distributions, supplemented by a phenomenological power term  $x^b$  to describe the small- $x$  asymptotic behaviour. Within the model, all PDFs can be written as linear combination of this function with different values of the parameters  $b$  and  $X$ , the latter representing a “chemical potential”. The parameter  $\bar{x}$  plays the role of a “temperature”, and it is common to all PDFs. For this reason, we do not write it explicitly as an argument.

Let us start with the quark PDFs. Being spin- $\frac{1}{2}$  particles, quarks follow a Fermi-Dirac distribution, and are thus described in terms of the  $h_+$  function. The PDFs at the initial scale  $\mu_0$  for each polarization (indicated with an up or down arrow) are given by

$$xq^{\uparrow}(x, \mu_0^2) = AC_q^{\uparrow} h_+(x; b, X_q^{\uparrow}) + \tilde{A}h_+(x; \tilde{b}, 0), \quad (2.2a)$$

$$x\bar{q}^{\uparrow}(x, \mu_0^2) = \bar{A}\bar{C}_q^{\uparrow} h_+(x; \bar{b}, -X_q^{\uparrow}) + \tilde{A}h_+(x; \tilde{b}, 0), \quad (2.2b)$$

where  $C_q^{\uparrow}$ ,  $\bar{C}_q^{\uparrow}$  and  $X_q^{\uparrow}$  are parameters depending on the quark species and polarization, while  $b$ ,  $\bar{b}$ ,  $\tilde{b}$  as well as the normalizations  $A$ ,  $\bar{A}$  and  $\tilde{A}$  are flavour-independent parameters.<sup>3</sup> In each equation the first term is of “valence nature”, and dominates at high  $x$ . The second term, called “diffractive” term in the original literature, is identical for all quark flavours and helicities and represents a contribution of “sea nature”, which thus is expected to dominate at small  $x$ . The unpolarized quark distribution is just the sum of the distributions with opposite helicities. For example, for an antiquark we have

$$\begin{aligned} x\bar{q}(x, \mu_0^2) &\equiv xq^{\uparrow}(x, \mu_0^2) + x\bar{q}^{\downarrow}(x, \mu_0^2) \\ &= \tilde{A}[\bar{C}_q^{\uparrow} h_+(x; \bar{b}, -X_q^{\uparrow}) + \bar{C}_q^{\downarrow} h_+(x; \bar{b}, -X_q^{\downarrow})] + 2\tilde{A}h_+(x; \tilde{b}, 0), \end{aligned} \quad (2.3a)$$

---

<sup>3</sup>Note that as far as the parameters  $C_q^{\uparrow}$  and  $\bar{C}_q^{\uparrow}$  are unconstrained, the normalizations  $A$  and  $\bar{A}$  are redundant as they can be reabsorbed in a redefinition of those parameters. However, we will see that simple incarnations of the model constrain the values of  $C_q^{\uparrow}$  and  $\bar{C}_q^{\uparrow}$  so that the normalizations  $A$  and  $\bar{A}$  are no longer redundant.

while for a valence quark distribution  $q_v = q - \bar{q}$  we have

$$\begin{aligned} xq_v(x, \mu_0^2) &\equiv xq^\uparrow(x, \mu_0^2) + xq^\downarrow(x, \mu_0^2) - x\bar{q}^\uparrow(x, \mu_0^2) - x\bar{q}^\downarrow(x, \mu_0^2) \\ &= A[C_q^\uparrow h_+(x; b, X_q^\uparrow) + C_q^\downarrow h_+(x; b, X_q^\downarrow)] - \bar{A}[\bar{C}_q^\uparrow h_+(x; \bar{b}, -X_q^\downarrow) + \bar{C}_q^\downarrow h_+(x; \bar{b}, -X_q^\uparrow)]. \end{aligned} \quad (2.3b)$$

Note that the valence distribution does not depend anymore on the diffractive (sea) term. Finally, the gluon follows a Bose-Einstein distribution with vanishing potential (i.e. a Planck distribution)

$$xg(x, \mu_0^2) = A_g h_-(x; b_g, 0) \quad (2.3c)$$

depending on a new normalization  $A_g$  and an exponent  $b_g$ .<sup>4</sup>

We immediately observe that the gluon distribution is very limited, as it depends only on 3 parameters ( $\bar{x}, b_g, A_g$ ). Differently, quark distributions depend on many more parameters. However, there are some relations that constrain some of them, as we shall now see.

Let us start from the asymptotic behaviours. At small  $x$ , we expect the gluon and sea quark to behave (grow) in the same way. The gluon grows as  $xg(x) \sim x^{b_g-1}$ , while the quarks are dominated at small  $x$  by the sea term<sup>5</sup> and thus behave as  $xq(x) \sim x^{\bar{b}}$ . Therefore, we find the relation

$$b_g = \bar{b} + 1 \quad (2.4)$$

that allows us to remove one of the two parameters.

We now focus on the parameters  $C_q^{\uparrow\downarrow}, \bar{C}_q^{\uparrow\downarrow}$ . In the statistical model, they are not independent parameters, rather they are related to one another and possibly with other parameters of the model. In Refs. [21, 23, 26, 30] the original model has been extended to describe transverse degrees of freedom. After integrating over transverse momentum to get collinear PDFs, the coefficients  $C_q^{\uparrow\downarrow}, \bar{C}_q^{\uparrow\downarrow}$  are given by

$$C_q^{\uparrow\downarrow} = \log\left(1 + e^{Y_q^{\uparrow\downarrow}}\right) \quad \bar{C}_q^{\uparrow\downarrow} = \log\left(1 + e^{-Y_q^{\uparrow\downarrow}}\right) \quad (2.5)$$

in terms of the parameters  $Y_q^{\uparrow\downarrow}$ , denoted ‘‘transverse potentials’’. Note that there are two independent  $Y$ ’s for each quark flavour, fixing the four  $C$ ’s for each quark. We observe that in previous studies, e.g. Ref. [18], the  $C$  parameters were simply given in terms of the chemical potentials  $X_q^{\uparrow\downarrow}$  through the relations

$$C_q^{\uparrow\downarrow} = X_q^{\uparrow\downarrow}, \quad \bar{C}_q^{\uparrow\downarrow} = \frac{1}{X_q^{\uparrow\downarrow}} \quad (2.6)$$

(note that in the second equation the order of the arrows changes). In this way, these four degrees of freedom for each quark flavour are completely fixed, leaving a parametrization that is rather constrained. The choice Eq. (2.6) of Ref. [18] was justified by the agreement with data, and later observed [26, 30] to be in decent agreement with the parametrization Eq. (2.5), up to a rescaling of  $A$  and  $\bar{A}$ . In this work, we only consider the simplest model, i.e. we adopt Eq. (2.6) in order to reduce the number of parameters to a minimum, keeping in mind that more flexibility can be achieved by adopting Eq. (2.5) instead.

Finally, we have to take into account the sum rules. Specifically we have two quark number sum rules and the momentum sum rule, that allow us to fix three additional parameters. We

<sup>4</sup>Note that we do not consider separate polarizations for the gluon, as suggested in the original literature for the model [18, 19]. This is perhaps not ideal, as there is experimental evidence that gluons carry a non-zero polarization contributing to the proton spin [37]. Phenomenological extensions of the model accounting for the gluon polarization have been discussed in the literature [25, 27], however it is not clear whether these extension are in agreement with the model assumptions. We will discuss gluon polarization with greater detail in section 6.4. When we consider unpolarized PDFs only, our assumption does not represent a severe limitation.

<sup>5</sup>The ‘‘valence’’ terms depending on  $b$  and  $\bar{b}$  contribute to the valence PDFs which cannot grow at small  $x$ , so we expect  $b$  and  $\bar{b}$  to be positive.

choose them to be the three normalizations  $A, \bar{A}, A_g$ . We stress that, differently from standard parametrizations,  $A$  and  $\bar{A}$  are not directly the normalizations of  $u$  and  $d$  valence distributions. This makes the implementation of the sum rules in the fitting code not straightforward. We give technical details in Appendix A.

One peculiar feature of the PDF parametrization Eqs. (2.3) is that it does not vanish at  $x = 1$ , as all other PDF parametrizations on the market do (to our knowledge). In particular the function  $h_{\pm}$  in Eq. (2.1) in  $x = 1$  becomes

$$h_{\pm}(1; b, X) = \frac{1}{\exp\left(\frac{1-X}{\bar{x}}\right) \pm 1}. \quad (2.7)$$

As we shall see, in the fits  $X < 1$  always, thus this function is exponentially suppressed. The suppression is rather strong, thanks to the value of  $\bar{x} \sim 0.1$  which is common to all fits with this parametrization, and to the fact that the largest value of  $X$  from the fit is smaller than 0.5. So practically the resulting PDFs are indistinguishable from zero in  $x = 1$ . Moreover, we recall that  $x = 1$  corresponds to the elastic scattering limit, which is no longer described by the QCD factorization theorem.

We conclude the section by counting the number of free parameters that we have in our fit. As we will only perform a fit to HERA data, we do not parametrize the strange distribution independently, as the data are not sufficiently powerful to distinguish it from the  $\bar{d}$  distribution. We thus take it to be a fixed fraction of  $\bar{d}$  distribution,

$$s(x, \mu_0^2) = \bar{s}(x, \mu_0^2) = \frac{f_s}{1 - f_s} \bar{d}(x, \mu_0^2), \quad f_s = 0.4, \quad (2.8)$$

which is a standard choice adopted by HERAPDF [1]. We are thus left with 5 PDFs to fit, i.e.  $u_v, d_v, \bar{u}, \bar{d}, g$ . According to the parametrizations given above, the free parameters to be fitted are

$$\bar{x}, b, \bar{b}, \tilde{b}, \bar{A}, X_u^{\uparrow}, X_u^{\downarrow}, X_d^{\uparrow}, X_d^{\downarrow}, \quad (2.9)$$

for a total of 9 parameters. For comparison, the default HERAPDF parametrization has 14 free parameters.

### 3 Setup of the fit and benchmark

Having established the form of the parametrization that we want to use, we now discuss the setup of our fit. We use the public `xFitter` toolkit, using a setup that is close to the one used for the determination of HERAPDF2.0 [1], with some notable differences:

- First of all, the paper [18] where we take our PDF parametrization advocates that it should be used at the initial scale  $\mu_0 = 2$  GeV. We therefore consider this scale as our default parametrization scale, which is higher than the HERAPDF2.0 scale which is 1.38 GeV.
- In order to avoid backward evolution, we then keep data only above 2 GeV, namely we have to cut out the  $Q^2 = 3.5$  GeV<sup>2</sup> bin of the HERA dataset.
- For the same reason, since we want to generate the charm PDF perturbatively, we have to raise the charm matching scale  $\mu_c$  above the parametrization scale [38]. In our work we set it to  $\mu_c = 1.38m_c = 2.01$  GeV, with  $m_c = 1.46$  GeV.
- To implement the displaced charm threshold, we have to use the `APFEL` evolution code [39] rather than the default `QCDNUM` code [40], as only the former implements this feature. This implies that we have to change the variable flavour number scheme from TR [41–43] to FONLL [44]. In practice, in our NNLO fits we use FONLL-C [44].

Contribution to $\chi^2$	HERAPDF2.0	Our HERAPDF-like fit
subset NC $e^+$ 920	444/377	415/363
subset NC $e^+$ 820	66/70	66/68
subset NC $e^+$ 575	219/254	217/249
subset NC $e^+$ 460	217/204	213/200
subset NC $e^-$	219/159	214/159
subset CC $e^+$	45/39	45/39
subset CC $e^-$	56/42	56/42
correlation term + log term	91 + 5	91 + 15
<b>Total <math>\chi^2</math>/d.o.f.</b>	<b>1363/1131</b>	<b>1333/1106</b>

**Table 1.** Total  $\chi^2$  per degrees of freedom (d.o.f.) and the partial  $\chi^2$  per number of data points (n.d.p.) of each subset of the inclusive HERA dataset, for HERAPDF2.0 and a HERAPDF-like fit obtained with the new setting introduced here.

The other settings of fit (masses, couplings,  $\chi^2$  definition, minimization strategy, etc) are kept as in HERAPDF2.0 [1].

We now present a HERAPDF-like NNLO fit with these settings, using the default HERAPDF2.0 parametrization, and compare it with the public HERAPDF2.0. The default HERAPDF2.0 parametrization is given by [1]

$$xg(x, \mu_0^2) = A_g x^{B_g} (1-x)^{C_g} - A'_g x^{B'_g} (1-x)^{25} \quad (3.1a)$$

$$xu_v(x, \mu_0^2) = A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} \left[ 1 + E_{u_v} x^2 \right] \quad (3.1b)$$

$$xd_v(x, \mu_0^2) = A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}} \quad (3.1c)$$

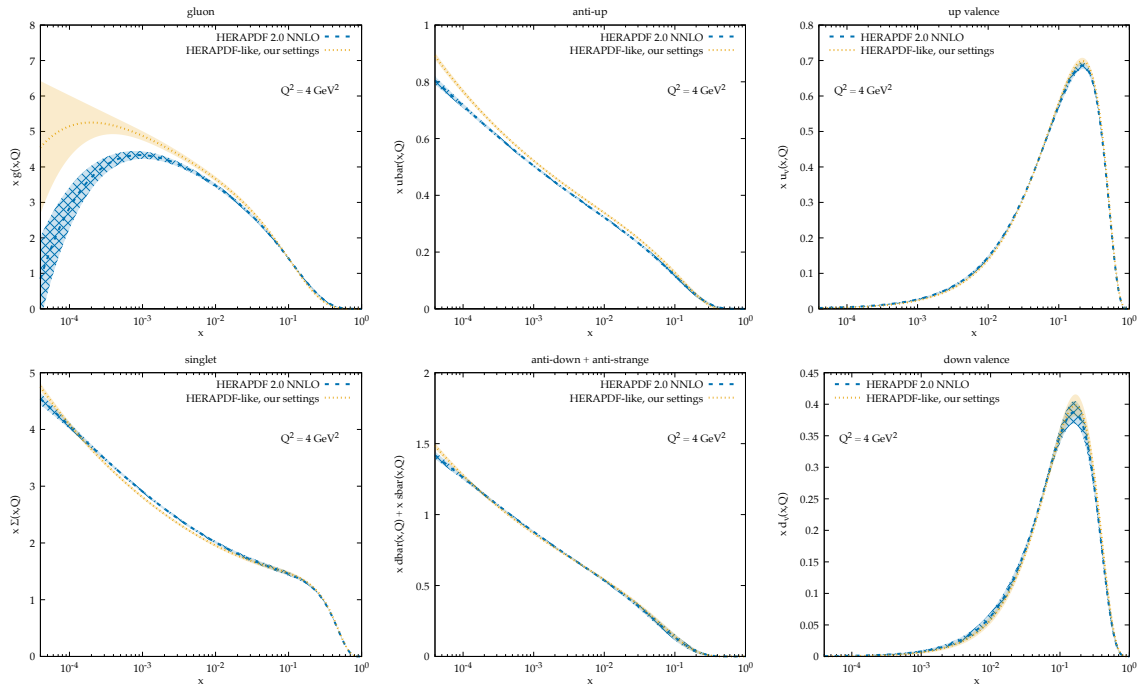
$$x\bar{u}(x, \mu_0^2) = A_{\bar{u}} x^{B_{\bar{u}}} (1-x)^{C_{\bar{u}}} \left[ 1 + D_{\bar{u}} x \right] \quad (3.1d)$$

$$x\bar{d}(x, \mu_0^2) = A_{\bar{d}} x^{B_{\bar{d}}} (1-x)^{C_{\bar{d}}}. \quad (3.1e)$$

Our HERAPDF-like fit will serve as a baseline for our next studies with the parametrization presented in Sect. 2. We choose the HERAPDF parametrization for this baseline because it is the simplest among the mainstream PDF parametrizations, using the smallest set of parameters (14 free parameters in total). Any other mainstream parametrization on the market has more parameters and it is therefore expected to be more flexible and possibly lead to higher fit quality.

We start by showing in Table 1 the  $\chi^2$  breakdown for the two fits. For each subset the contribution to the  $\chi^2$  over the number of data points is shown, as well as the contributions to the  $\chi^2$  from the correlations and the logarithmic term (see Ref. [1] for the definition and meaning of these pieces). Our fit has a total  $\chi^2$  which is smaller by 30 units, which is compatible (within statistical fluctuations) with the reduction of datapoints by 25 units. The small improvement is likely due to the better description of the  $E = 920$  GeV dataset which contains the small- $x$  data, which are not well described by fixed-order perturbation theory [1, 45–48]. The cut bin at  $Q^2 = 3.5$  GeV<sup>2</sup> contains the data at smallest  $x$ , thus its absence in the new fit leads to a 29 units smaller  $\chi^2$  of the  $E = 920$  GeV dataset with just 14 less datapoints.

We now show the effect of the new settings to the PDFs. In Figure 1 we show a comparison of HERAPDF2.0 and our new fit with HERAPDF parametrization at the scale  $Q = 2$  GeV for the gluon, total quark singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs. We observe that the valence distributions as well as the medium-large  $x$  behaviour of all PDFs remain almost unchanged in the two fits. Differences are instead present in the small- $x$  region, for the sea contribution to the quarks and more markedly to the gluon. These differences are certainly due, at least in part, by the smaller



**Figure 1.** Comparison of the original HERAPDF2.0 fit (dashed blue) with the one with our modified settings (dotted yellow) for the gluon, total singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs. The uncertainty shown is only the “experimental” one, namely the one coming from the uncertainty on the parameters determined from the fit.

dataset, and in particular by the absence of the small- $x$  data of the  $Q^2 = 3.5 \text{ GeV}^2$  bin, which also leads to an increased uncertainty in the gluon PDF at small  $x$ .<sup>6</sup>

## 4 Fit with the new parametrization

We now consider the parametrization described in section 2. We perform a fit with the same settings described above, simply changing the PDF parametrization, that we dub QSPDF. Comparing it with the HERAPDF-like fit just discussed in section 3 we are able to test the effect of the different parametrization alone, disentangled from any other effect.

The  $\chi^2$  breakdown for this new fit is shown in Table 2, to be compared with the last column of Table 1. We immediately observe an overall deterioration of the fit quality, with the  $\chi^2$  increasing by 51 units, from 1333 to 1384. The number of degrees of freedom also increases slightly (5 units) due to the smaller number of parameters in the QSPDF parametrization, but this is not enough to explain the increase in the  $\chi^2$ .

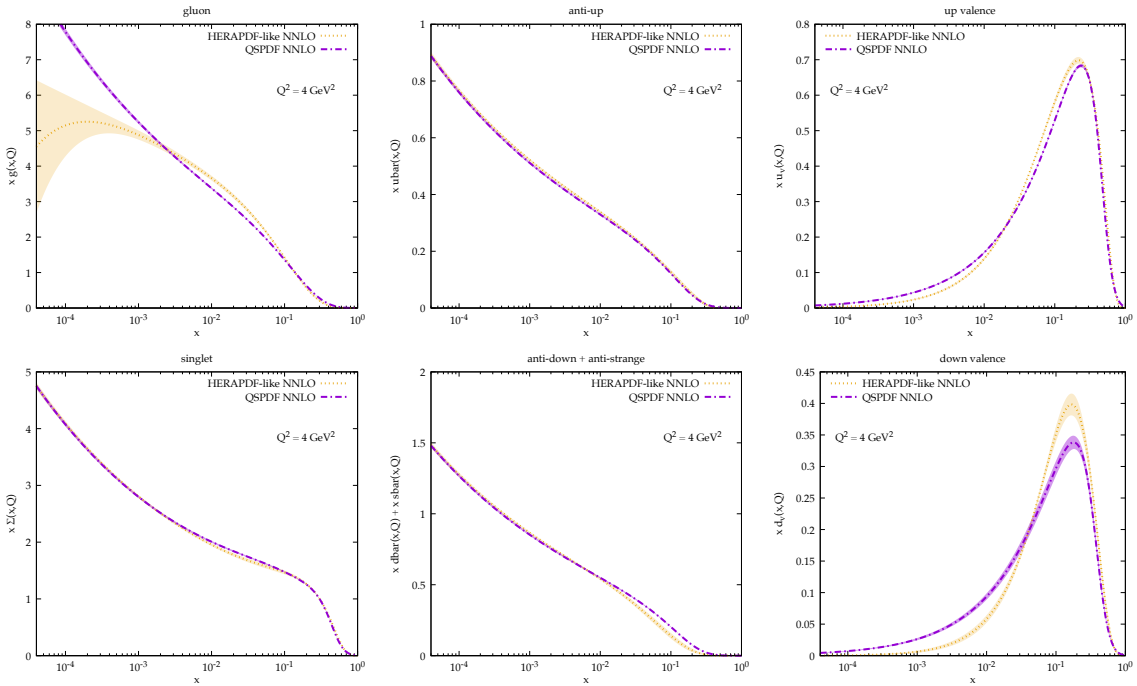
Looking carefully at the tables, we see that the dataset exhibiting the largest deterioration is again the  $E = 920 \text{ GeV}$  dataset, namely the one containing the majority of small- $x$  data. We suspect indeed that the origin of the large  $\chi^2$  comes from a bad description of the low- $x$  data, which is in turn due to the limited flexibility of the QSPDF parametrization at small  $x$ , in particular for the gluon. We will come back to this point later.

<sup>6</sup>On top of this, there is a technical difference in the way uncertainties are calculated. HERAPDF2.0 uses the Pumplin procedure [49], while we use the simpler HESSE approach of MINUIT [50] which may lead to larger uncertainties.



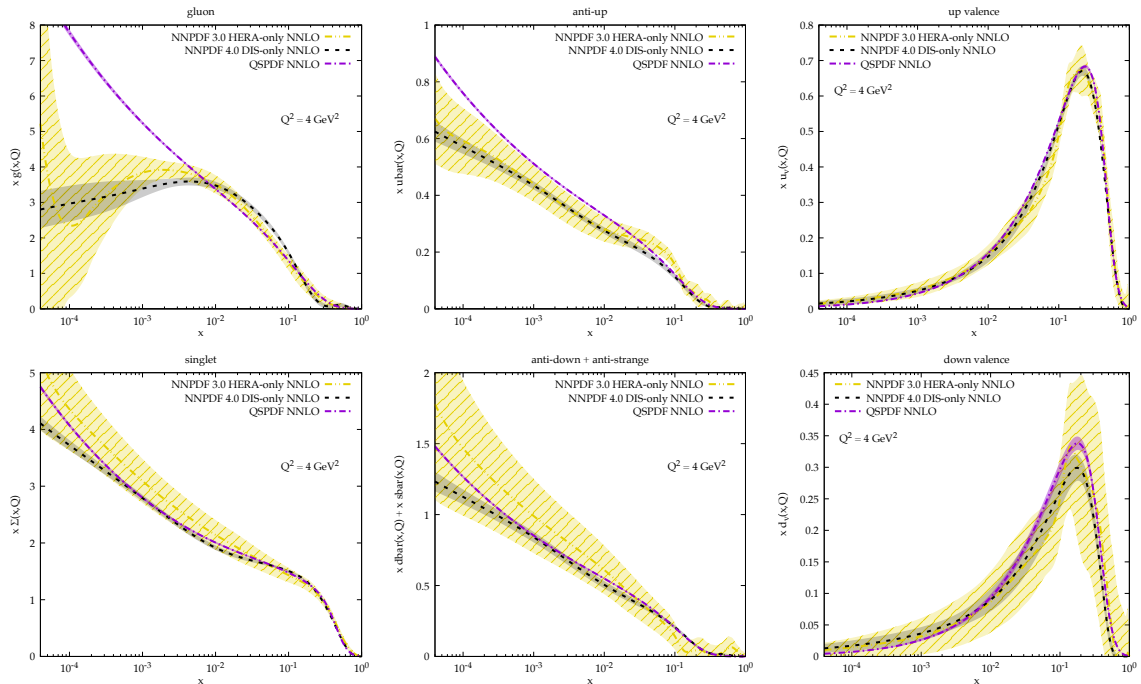
Contribution to $\chi^2$	QSPDF (NNLO)	QSPDF (NNLO+NLL $x$ )	HERAPDF-like (NNLO+NLL $x$ )
subset NC $e^+$ 920	452/363	447/363	408/363
subset NC $e^+$ 820	71/68	66/68	63/68
subset NC $e^+$ 575	224/249	229/249	216/249
subset NC $e^+$ 460	221/200	231/200	218/200
subset NC $e^-$	222/159	225/159	219/159
subset CC $e^+$	46/39	48/39	46/39
subset CC $e^-$	61/42	61/42	54/42
correlation term + log term	98 - 11	88 - 28	80 + 1
<b>Total <math>\chi^2</math>/d.o.f.</b>	<b>1384/1111</b>	<b>1369/1111</b>	<b>1304/1106</b>

**Table 2.** Same as table 1, showing the  $\chi^2$  breakdown for QSPDF at NNLO and NNLO+NLL $x$  as well as our HERAPDF-like fit at NNLO+NLL $x$ .



**Figure 2.** Comparison of QSPDF (dot-dashed purple) with our HERAPDF-like fit (dotted orange) for the gluon, total singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs, showing experimental uncertainty band.

In Figure 2 we show the comparison of our HERAPDF-like fit with the QSPDFs. There are some marked differences between the two PDF sets. Starting with the quarks, we observe a small distortion of the  $u_v$  distribution below the peak, with QSPDFs being smaller at medium  $x$  and larger at small  $x$ . A similar but bigger effect is present on the  $d_v$  distribution as well, where also the height of the peak is smaller in the QSPDF fit. The anti-quark PDFs behave the same at small  $x$ , while at medium-large  $x$  the QSPDFs for the total singlet and the  $\bar{d} + \bar{s}$  distribution are slightly larger compared to the PDF uncertainty. Finally, a big difference is present in the gluon PDF from medium to low  $x$ , with the QSPDF gluon rising at small  $x$  compared to the HERAPDF-like gluon which bends down at  $x \sim 10^{-4}$ . Moreover, the PDF uncertainty of the QSPDFs is very small everywhere, especially in the gluon where instead the HERAPDF parametrization gives a much



**Figure 3.** Comparison of QSPDF (dot-dashed purple) with NNPDF3.0 HERA-only (dot-dot-dashed yellow) and NNPDF4.0 DIS-only (dashed black) for the gluon, total singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs. The NNPDF uncertainty band covers various sources of uncertainty, including those coming from parametrization choice.

larger uncertainty in the small- $x$  region.

These differences, especially in the gluon, are due to the very constrained parametrization which limits its flexibility. However, before drawing conclusions, it is instructive to compare the QSPDFs with other PDFs on the market. We consider the NNPDF3.0 set that has been obtained fitting only HERA data [51], which is a dataset very similar to what we are using here, and the NNPDF4.0 set that has been obtained fitting only DIS data [5], which contains more data but it is still closer to our dataset than a global fit.

The plots are shown in Figure 3. We immediately notice that the NNPDF uncertainties are larger than QSPDFs, generally due to the very flexible parametrization, with the ones of the older NNPDF3.0 fit being larger than the newer NNPDF4.0 fit, partly due to the larger dataset and partly to the improved fitting methodology in the latter. Within uncertainties, there is a sufficiently good agreement between QSPDFs and the NNPDF sets for the total singlet, the  $\bar{d} + \bar{s}$  and the valence distributions. In particular, we notice that the  $u_v$  and  $d_v$  PDFs of the QSPDF set have a shape very similar to the NNPDF ones, despite the differences with the HERAPDF-like fit. Because of the unbiased nature of the NNPDF parametrization, we thus conclude that the  $u_v$  and  $d_v$  distributions are probably better described by the QSPDF parametrization than by the HERAPDF parametrization.<sup>7</sup>

<sup>7</sup>To confirm this, we have tried to fit directly the  $u_v$  and  $d_v$  distributions from the QSPDF set using the HERAPDF parametrization Eq. (3.1), noticing that indeed the HERAPDF parametrization is not able to reproduce the shape of the QSPDF valence distributions in the medium/small- $x$  region. Conversely, the QSPDF parametrization is able to reproduce the valence distributions of our HERAPDF-like fit, except for the high- $x$  tail which is however mostly unconstrained by data. Note that adding more polynomial contributions in the HERAPDF parametrization of the valence PDFs, as done in many PDF studies (see e.g. [7]), will likely give enough flexibility to reproduce a QSPDF valence shape.

In contrast, the  $\bar{u}$  distribution in QSPDF is somewhat higher at medium-small  $x$  than both NNPDF predictions. This behaviour is dictated by the sea term of the QSPDF parametrization, which is in turn linked to the gluon. The gluon is very different from the NNPDF one, as the latter tends to flatten below  $x \sim 10^{-2}$  in both fit versions, while the QSPDF gluon keeps growing at small  $x$ . This behaviour of the QSPDF gluon is a consequence of its very constrained functional form, which is not able to reproduce a shape similar to the NNPDF one (and not even of the HERAPDF-like fit).

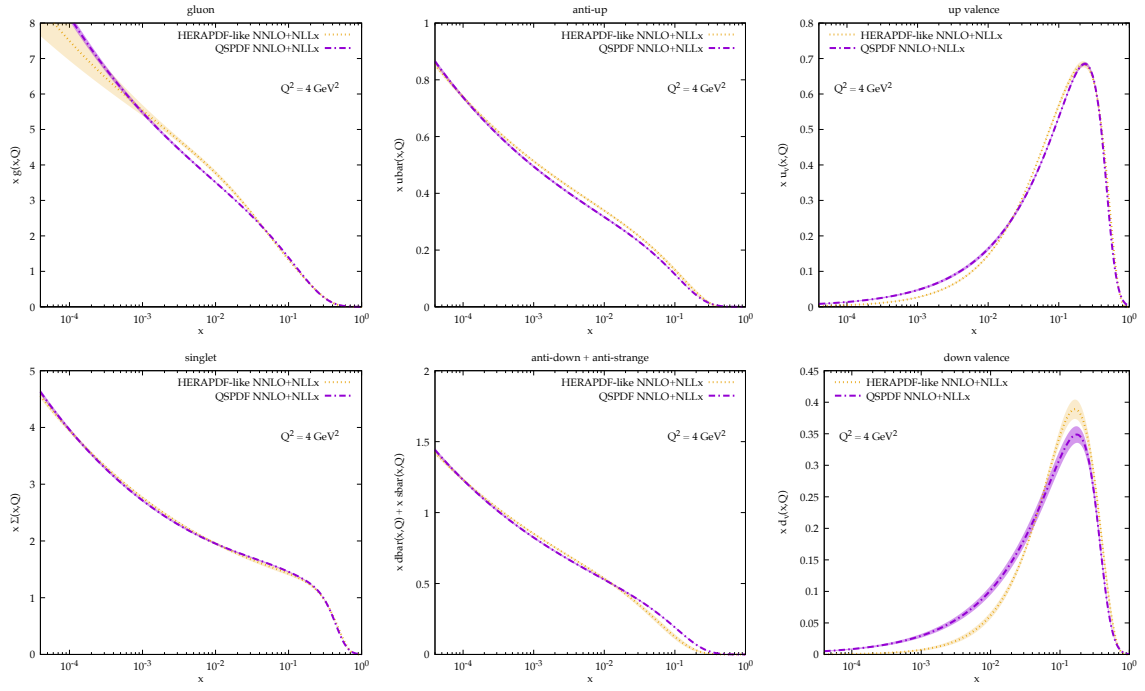
We have thus found various hints that the gluon parametrization Eq. (2.3c) is not sufficient to accurately describe the data at small  $x$ . For this reason, we will consider in section 5 a more flexible parametrization for the gluon PDF. However, we must recall that the shape of the gluon PDF at small  $x$  is strongly dependent on the perturbative order of the fit, due to the presence of enhanced logarithmic terms in the perturbative ingredients that make their perturbative expansion unstable at small  $x$ . This instability can be cured by resumming these logarithms to all orders [52–70], leading to predictions that are more reliable in that region. Interestingly, adding the resummation of small- $x$  logarithms in PDF fits leads to a gluon PDF that rises at small  $x$  at small scales [47, 48, 71]. Therefore, it may be possible that the QSPDF is able to give a better description of the gluon PDF when small- $x$  resummation is turned on.

To verify this, we have performed fits to the HERA data with small- $x$  resummation at next-to-leading logarithmic (NLL $x$ ) accuracy (included in `xFitter` through `APFEL` interfaced to the `HELL` resummation code [72–75], version 3.0) using both the QSPDF and the HERAPDF parametrizations. The fit quality is reported in the last two columns of table 2. For QSPDFs we observe a reduction of the  $\chi^2$  from NNLO to NNLO+NLL $x$  of 15 units, which is significant but not substantial. Conversely, in our HERAPDF-like fit the  $\chi^2$  reduces by 29 units, which is more notable given that it was already lower than the QSPDF one. These numbers confirm [47, 48, 71] that the inclusion of small- $x$  resummation is beneficial and improves the agreement with data, but they also show that the QSPDF parametrization is not flexible enough at small  $x$  to give a good description even when small- $x$  resummation is included.

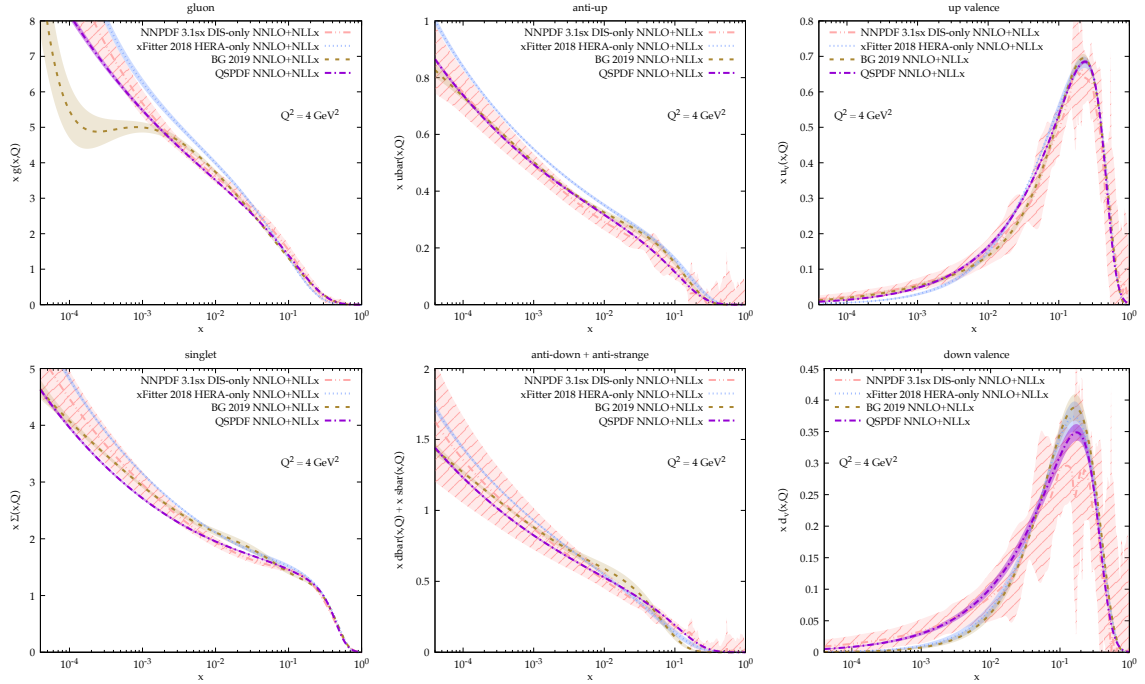
As far as PDFs are concerned, we plot in figure 4 both “resummed” PDF sets, using the same structure as before. We observe that most PDFs are essentially unchanged, with small differences visible only in the  $\bar{u}$  distribution, with the exception of the gluon PDF, that changes significantly in the HERAPDF-like fit, getting much closer to the QSPDF gluon, which is instead basically unchanged. This shows that indeed the QSPDF parametrization is more suitable for fitting PDFs with all-order resummation of small- $x$  logarithms than without.

To conclude, we also compare the QSPDF fit at NNLO+NLL $x$  with analogous resummed fits from Refs. [47, 48, 71] in figure 5. There are differences at medium-high  $x$  and in the valence distributions between the sets that are due to the parametrization and the dataset which are not very useful to compare now. Let’s focus instead on the low- $x$  region. We see in general good agreement in the quark sea contributions, with the uncertainty from the NNPDF3.1sx fit being large enough to cover all other curves (except the  $\bar{u}$  distribution from the 2018 `xFitter` study with resummation [48] which is slightly higher but still very close). In the gluon PDF we see a general tendency to grow at small  $x$ , with the BG 2019 fit from Ref. [71] having a peculiar shape that makes it different from the other sets. This shape is due to the particular parametrization as well as the use of a newer version of the resummation code `HELL` with respect to the previous two fits, differing from the previous version by subleading logarithmic contributions [74, 75], as documented in Ref. [71] itself. This newer version of `HELL` is the same used here, but both QSPDF and HERAPDF-like parametrizations are not flexible enough to produce a similar shape.

As a final observation, we note that the shape of the gluon obtained with small- $x$  resummation is similar to what is obtained with the recent MSHT (approximate) N<sup>3</sup>LO fit [76]. Indeed the small- $x$  logarithms appearing at this order behave in a way similar to their all-order resummation,



**Figure 4.** Same as figure 2 but including small- $x$  resummation at NNLO+NLL $x$  accuracy.



**Figure 5.** Same as figure 4, but comparing QSPDF with NNPf3.1sx [47] (dot-dot-dashed salmon), the 2018 xFitter low- $x$  study of Ref. [48] (dotted blue) and BG 2019 [71] (dashed brown).

at least in a region of intermediate  $x \sim 10^{-3}$ , thus providing a sort of approximation of the all-order behaviour in that region. Performing a QSPDF fit at (approximate) N<sup>3</sup>LO using xFitter is

Contribution to $\chi^2$	QSPDFflex (NNLO)	QSPDFflex (NNLO+NLL $x$ )
subset NC $e^+$ 920	412/363	401/363
subset NC $e^+$ 820	71/68	67/68
subset NC $e^+$ 575	224/249	220/249
subset NC $e^+$ 460	220/200	223/200
subset NC $e^-$	226/159	228/159
subset CC $e^+$	55/39	53/39
subset CC $e^-$	63/42	62/42
correlation term + log term	80 – 17	75 – 17
<b>Total <math>\chi^2</math>/d.o.f.</b>	<b>1334/1109</b>	<b>1311/1109</b>

**Table 3.** Same as table 1, showing the  $\chi^2$  breakdown for QSPDFflex at NNLO and NNLO+NLL $x$ .

however not possible at the moment due to the lack of the necessary theoretical ingredients in the code.

## 5 More flexible QSPDF parametrization

In section 4 we have seen that the QSPDF parametrization is able to give a reasonable description of the PDFs at medium-high  $x$ , but it is not sufficiently flexible at small  $x$  to describe the data well. In particular, the gluon PDF parametrization Eq. (2.3c) is very constrained and cannot produce the variety of shapes obtained in PDF fits using more flexible parametrizations.

In this section we thus consider a minimal modification of the QSPDF parametrization that increases the flexibility of the gluon PDF at small  $x$ . To do so, we follow the suggestion of Ref. [71] of using a polynomial in  $\log x$  to model the shape at small  $x$ , and modify the gluon parametrization Eq. (2.3c) as

$$xg(x, \mu_0^2) = A_g h_-(x; b_g, 0) [1 + F_g \log x + G_g \log^2 x]. \quad (5.1)$$

The polynomial contribution in  $\log x$  does not modify the behaviour of the PDF at large  $x$ , where the statistical model is meant to be physically motivated, and gives additional degrees of freedom to model the low- $x$  region, where instead the model behaviour  $x^{b_g}$  is just phenomenological.

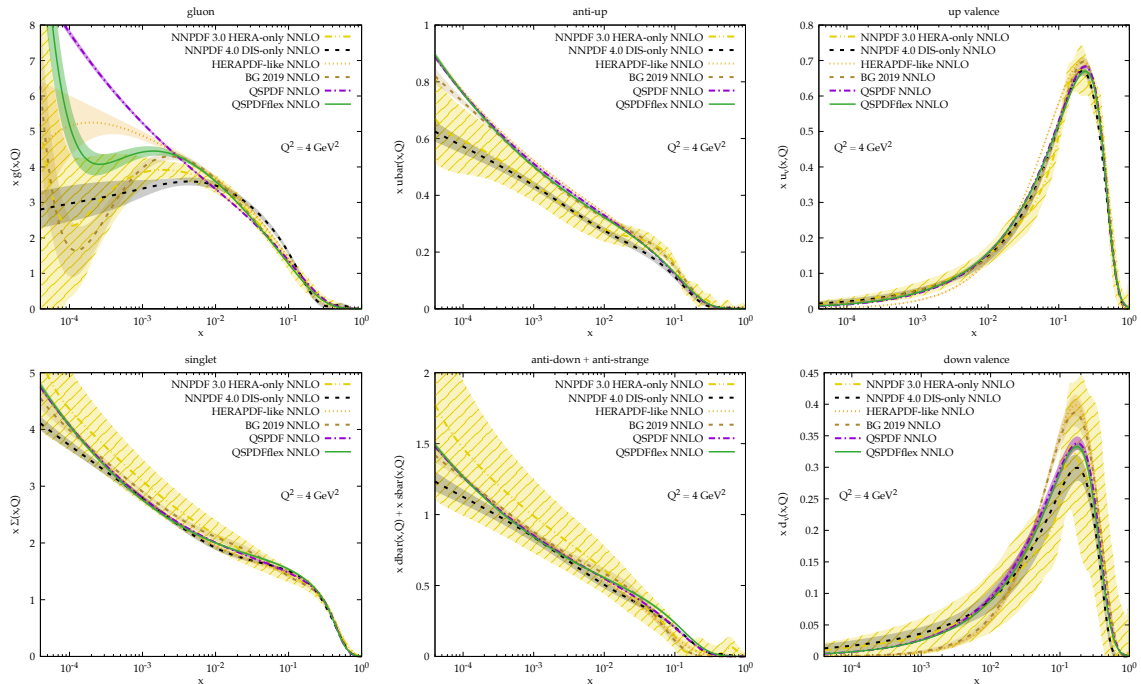
In Eq. (5.1),  $F_g$  and  $G_g$  are two new parameters to be fitted.<sup>8</sup> Moreover, we also decide to unlink  $b_g$  from  $\tilde{b}$ , considering it a free parameter. In this way we have three extra free parameters with respect to the QSPDF parametrization of section 2. However, we have verified that with this new choice of parametrization for the gluon it is possible to fix the value of the  $\tilde{b}$  parameter entering the antiquark parametrization without decreasing the fit quality. We choose as default value  $\tilde{b} = b$ , and we verified that other reasonable choices (e.g.  $\tilde{b} = b/2$ ) do not change the fit quality. According to this procedure, we have to a total of 11 free parameters to be fitted,

$$\bar{x}, b, \tilde{b}, b_g, F_g, G_g, \tilde{A}, X_u^\uparrow, X_u^\downarrow, X_d^\uparrow, X_d^\downarrow, \quad (5.2)$$

which is just two more parameters with respect to the default QSPDF parametrization, and three parameters less than the HERAPDF parametrization. We dub this alternative parametrization QSPDFflex.

We start considering the fit quality of the QSPDFflex parametrization, both at NNLO and with small- $x$  resummation at NNLO+NLL $x$ . We report the  $\chi^2$  breakdown in table 3. We observe immediately that now the quality is comparable to the analogous fits obtained with the HERAPDF

<sup>8</sup>We use the same notation of Ref. [71] for a direct comparison.



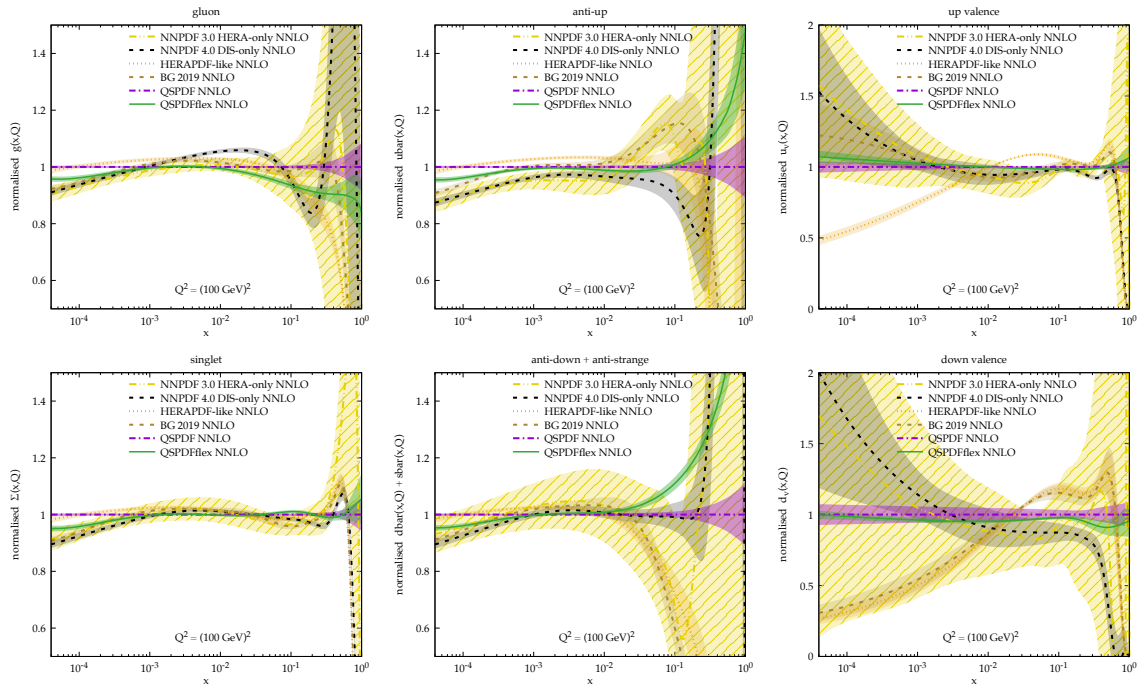
**Figure 6.** Comparison of QSPDF (dot-dashed purple) and QSPDFflex (solid green) at NNLO with our HERAPDF-like fit (dotted orange), NNPDF3.0 HERA-only (dot-dot-dashed yellow), NNPDF4.0 DIS-only (dashed black) and BG (dashed brown) for the gluon, total singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs.

parametrization. As expected, the improvement is driven by a better description of the  $E = 920$  GeV dataset containing the small- $x$  data. More precisely, taking into account the different number of parameters, the fit quality is basically identical for QSPDFflex and HERAPDF-like, making each parametrization as good as the other. As such, each of them represent a measure of the parametrization bias of the other, and could be used for constructing a parametrization uncertainty.

We now move to comparing the PDFs. In figure 6 we show the QSPDFflex set at NNLO together with all the NNLO PDFs considered so far. In particular, we see that gluon PDF in the new fit (solid green curve) is rather different from the previous QSPDF gluon, as now there is a non-trivial shape that tends to reduce the size of the gluon below  $x \sim 10^{-3}$  before rising again at  $x \sim 10^{-4}$ . This shape is similar to the one obtained in the BG fit of Ref. [71], even though in that case the drop and growth are stronger, and also close to the NNPDF3.0 HERA-only fit. More in general, all PDF parametrizations except the QSPDF one predict a gluon that either flattens or decreases before possibly growing again at small  $x$ . As far as the other PDFs are concerned, the difference between the QSPDFflex and the QSPDF sets is very small or totally negligible, as expected given that the biggest change is in the parametrization of the gluon.

We also consider in figure 7 the same PDFs at the electroweak scale  $Q = 100$  GeV, plotted as a ratio to QSPDFs. For gluon and anti-quark PDFs we see a generally good agreement between the various sets in the small- $x$  region, with differences at most at the 10% level, even though not always within the uncertainty. This shows that DGLAP evolution reduces the discrepancies between the sets, especially in the gluon PDF. At larger  $x \gtrsim 0.1$  the uncertainties get larger as well as the differences between sets, due to the lack of direct experimental constraints this region, which is thus strongly dependent on the functional forms adopted. The valence distributions are characterised by larger relative uncertainties and differences (note also the larger range shown in the plot), especially in the low- $x$  region. In particular, we observe that the HERAPDF-like fit predicts a very different





**Figure 7.** Same as figure 6 but for  $Q = 100$  GeV, showing the various curves as ratios to QSPDF.

$u_v$  distribution from all the other sets. Similarly, the  $d_v$  distribution of the HERAPDF-like set is very different from QSPDF(flex) and NNPDF, but almost identical to the BG set. The large uncertainties and differences at large  $x$  also shows that the use of non-vanishing functions in  $x = 1$ , Eq. (2.7), in place of a more standard power suppression in  $1 - x$ , does not represent an issue when describing the data.

To further verify the impact of using the functional form based on Fermi-Dirac and Bose-Einstein distributions, we plot in figure 8 a zoom of the large- $x$  region of figure 6. We can appreciate that gluon and anti-quark distributions are essentially indistinguishable from other sets where PDFs vanish in  $x = 1$ , and are compatible with them within uncertainty in the  $x \rightarrow 1$  region. Quark PDFs are instead more sensitive to the functional form. In particular, the valence distributions of QSPDF and QSPDFflex tend to be higher than the HERAPDF-like fit for  $x \gtrsim 0.7$ , staying visibly larger than zero at  $x = 1$ . This difference is inherited by the total singlet PDF. We note however that QSPDF and QSPDFflex are fully within the NNPDF3.0 uncertainty band even at very high  $x$ , due to the large uncertainty bands of the set in turn due to the lack of data in this region. Note that the NNPDF4.0 set has a much reduced uncertainty. This is partly due to the larger dataset, that includes several DIS data at rather high- $x$  (e.g. BCDMS [77] reaches  $x = 0.75$ , NMC [78] reaches  $x = 0.68$ , CHORUS [79] reaches  $x = 0.65$ ), and partly also to the improved fitting methodology [5] that favours smoother replicas allowing for smaller uncertainties in the extrapolation region  $x \gtrsim 0.75$ . Since NNPDF4.0 explicitly assumes a power behaviour  $(1 - x)^\alpha$  vanishing at large  $x$  ( $\alpha > 0$ ), the small uncertainty in the extrapolation region  $x \gtrsim 0.75$  is not necessarily reliable. Therefore, the inconsistency of our quark PDFs with NNPDF4.0 in this region is not indicative of a preference towards a stronger large- $x$  suppression.

When including the resummation of small- $x$  logarithms, the agreement between the various curves improves. This is shown in figure 9, again comparing the new QSPDFflex set with all the PDF sets with resummation considered before. Again, the gluon is the one that exhibits the biggest

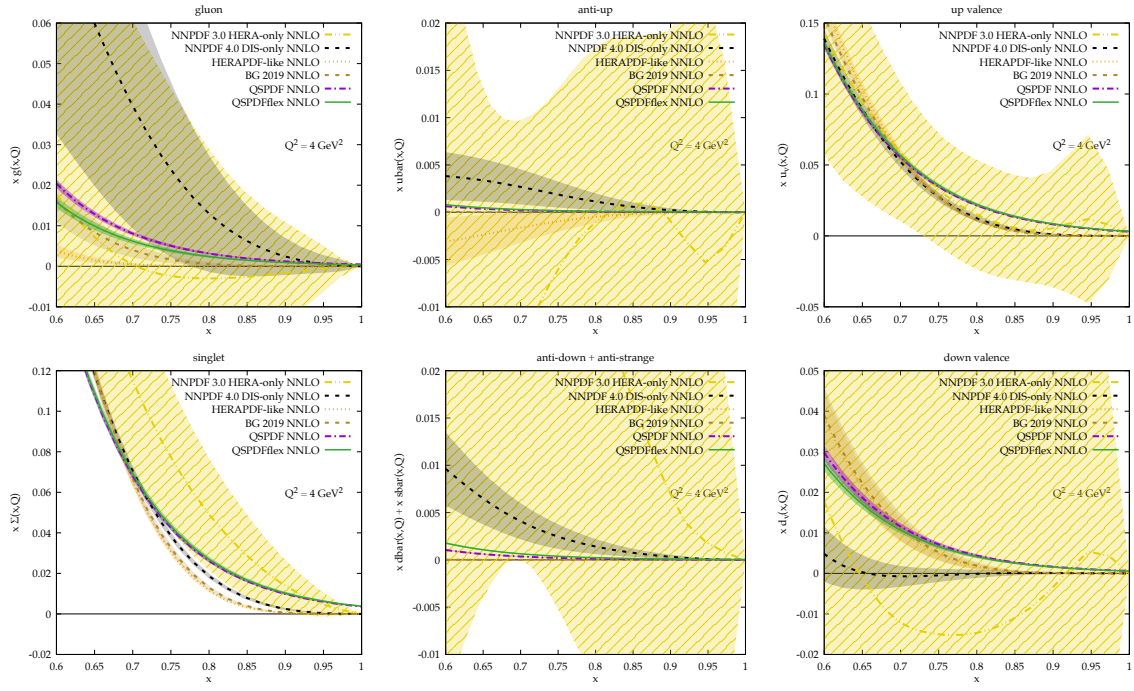


Figure 8. Same as figure 6 zoomed in at large  $x$ .

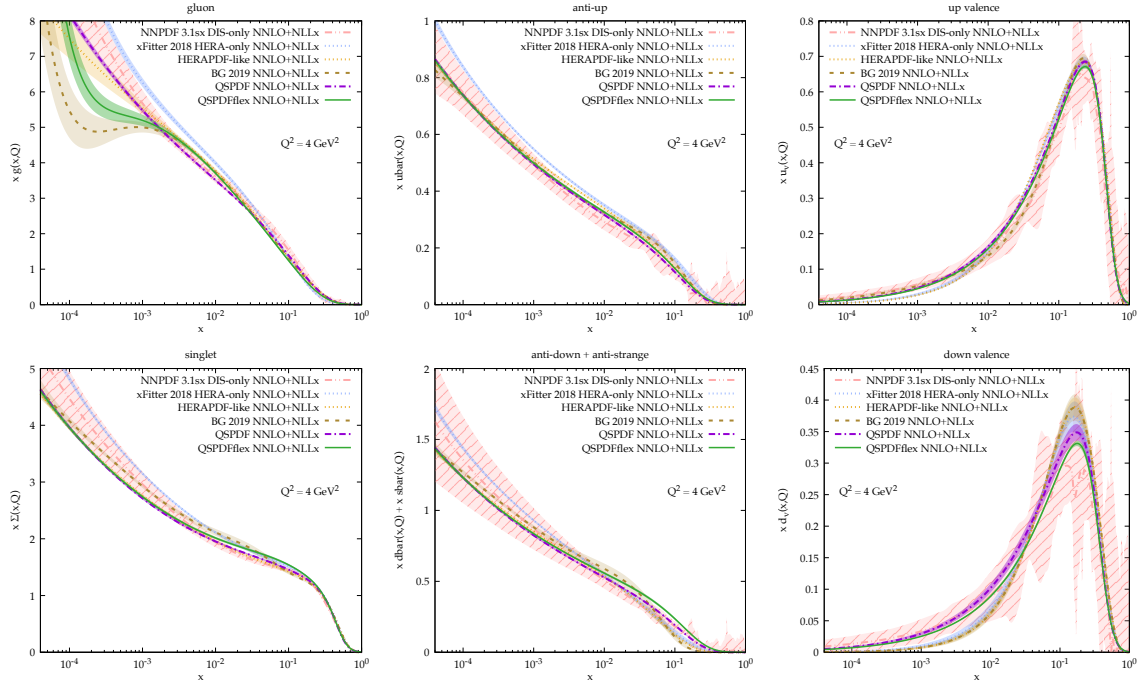
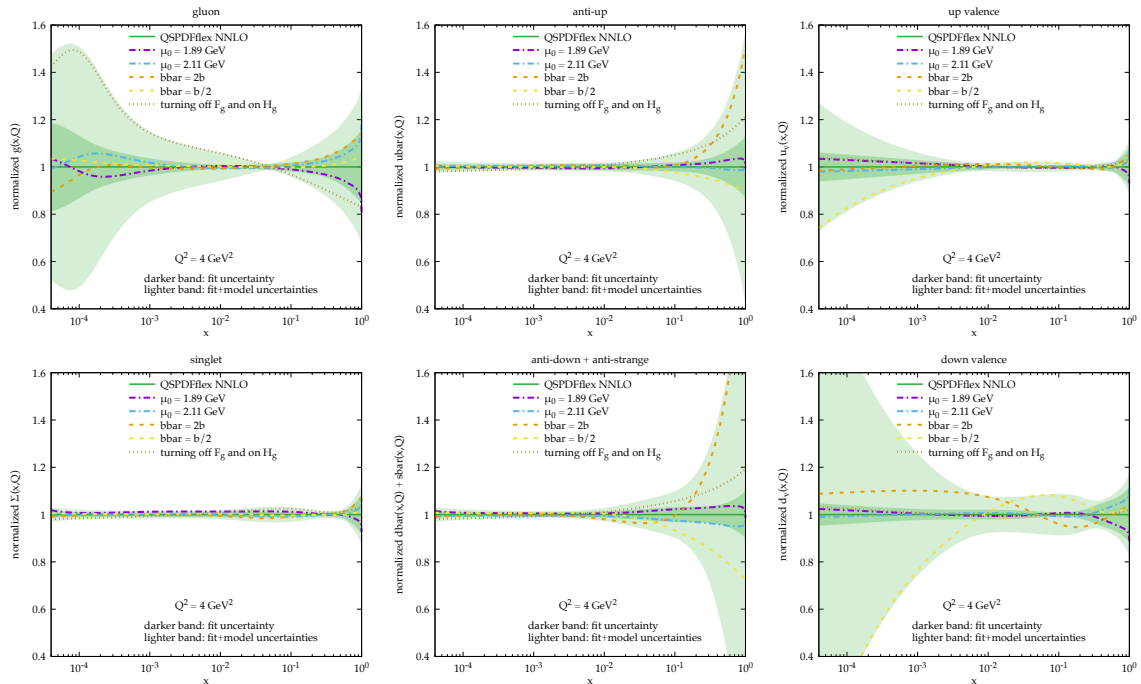


Figure 9. Comparison of QSPDF (dot-dashed purple) and QSPDFflex (solid green) at NNLO+NLL $x$  with our HERAPDF-like fit (dotted orange), NNPDF3.1sx DIS-only (dot-dot-dashed salmon), xFitter 2018 (dotted blue) and BG (dashed brown) for the gluon, total singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs.





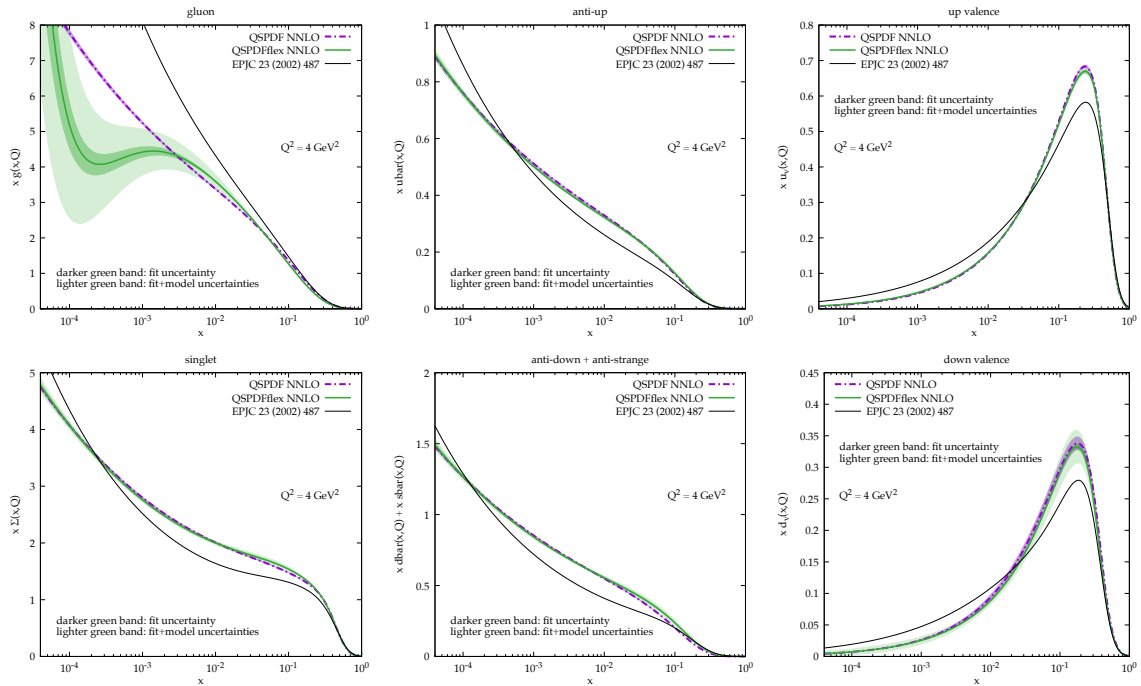
**Figure 10.** Model uncertainties at NNLO obtained by various variations as indicated in the text. The plots are presented as ratios to the central PDF, for the gluon, total singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs.

difference between QSPDF and QSPDFflex, but now the QSPDFflex gluon grows as the QSPDF one, but with some oscillations that resemble those of the BG set [71]. We keep seeing that the QSPDFflex is close to QSPDF for the other PDFs, but now there is a more marked difference in the sea quarks and in the  $d_v$  distribution at medium  $x$ .

We now discuss the uncertainties. We observe that the QSPDFflex has similar (small) uncertainties to QSPDF, except for the gluon which has a larger band at small  $x$ , due to the presence of extra parameters. For the more promising QSPDFflex parametrization we also consider model uncertainties, in particular parametrization uncertainty. Specifically, we investigate the effect of changing the initial scale of the parametrization, raising<sup>9</sup> or lowering it by 0.11 GeV, and of modifying parameters or functional form: we vary  $\bar{b}$  up and down to  $\bar{b} = 2b$  and  $\bar{b} = b/2$ , and replace the linear logarithmic gluon term  $F_g \log x$  with a cubic term  $H_g \log^3 x$  (following an analogous variation performed in Ref. [71]). The relative effect of each individual variation is shown at NNLO in figure 10 (the relative uncertainties at NNLO+NLL $x$  are similar). We notice that the gluon uncertainty is dominated at medium-small  $x$  by the effect of changing the parametrization with a cubic logarithmic terms in place of the linear one, as expected. Its effect on other PDFs is mild, and concentrated at high  $x$ , indirectly induced by the momentum sum rule. The  $\bar{b}$  variations have larger effects on the (anti)quarks, and in particular on the valence distributions at medium-low  $x$ , most importantly for the  $d_v$ . The variation of the fit scale  $\mu_0$  are mild and similar in all PDFs, giving bigger effects at small and high  $x$ .

To conclude the section, we show the actual PDFs of the QSPDFflex set including the model uncertainties in Fig. 11. We also show the QSPDF set for comparison. The lighter-green uncertainty band represents the total (fit+model) uncertainty, obtained by summing in quadrature the fit

<sup>9</sup>When we raise the initial scale, we also have to raise the charm threshold to keep the condition  $\mu_c > \mu_0$ . The contribution of raising  $\mu_c$  is less important than that of raising  $\mu_0$ , and indeed the effect is rather symmetric to the lower variation of  $\mu_0$  where  $\mu_c$  is kept fixed.



**Figure 11.** Comparison of QSPDFflex (solid green), showing also the total (fit+model) uncertainty, with QSPDF (dot-dashed purple) and the older determination of Ref. [18] (solid black, without uncertainties), for the gluon, total singlet,  $\bar{u}$ ,  $\bar{d} + \bar{s}$ ,  $u_v$  and  $d_v$  PDFs.

uncertainty (corresponding to the darker-green band) and each individual model variation with respect to the central PDF. We note that the gluon and  $d_v$  distributions have visibly larger total bands, while for the others the difference is less marked, but still visible in some regions (especially low  $x$  and medium-high  $x$ ).

## 6 Implications

Having established the possibility of fitting HERA data with reasonable/good quality with the QSPDF and QSPDFflex parametrizations, we now want to comment on some possible physical implications of this study.

### 6.1 Comparison with previous model determinations

First of all, we may consider the success in fitting the data as a success of the statistical model behind the QSPDF parametrization. It is true that the original model of Ref. [18] leading to the QSPDF parametrization does not allow to fit the data with high quality, but still the quality is reasonable and moreover we have seen that most of the problems come from the limited flexibility of the gluon at small  $x$ . Admittedly, the statistical model of Ref. [18] is designed to describe the high- $x$  region of the PDFs, and the ingredients needed to describe the small  $x$  region, i.e. the  $x^b$  factors and the “diffractive” (sea) quark terms, are only phenomenological. Therefore, we believe that the QSPDFflex parametrization of section 5 is still in agreement with the original ideas of the statistical model, as in particular it does not change the description of the large- $x$  PDFs. In this sense, the better description achieved with the QSPDFflex parametrization can be considered as a success of the statistical model.

Fitted param.	QSPDF NNLO	QSPDF NNLO+NLL $x$	QSPDFflex NNLO	QSPDFflex NNLO+NLL $x$	Ref. [18]
$\bar{x}$	$0.0950 \pm 0.0011$	$0.0932 \pm 0.0012$	$0.0964 \pm 0.0017$	$0.0971 \pm 0.0018$	0.09907
$b$	$0.557 \pm 0.009$	$0.54 \pm 0.03$	$0.538 \pm 0.012$	$0.545 \pm 0.012$	0.40962
$\bar{b}$	$0.00016 \pm 0.00003$	$0.000019 \pm 0.000003$	$b$	$b$	$2b$
$\tilde{b}$	$-0.1700 \pm 0.0018$	$-0.172 \pm 0.003$	$-0.179 \pm 0.005$	$-0.169 \pm 0.005$	$-0.25347$
$b_g$	$\tilde{b} + 1$	$\tilde{b} + 1$	$0.440 \pm 0.013$	$0.434 \pm 0.016$	$\tilde{b} + 1$
$F_g$	-	-	$0.212 \pm 0.004$	$0.207 \pm 0.004$	-
$G_g$	-	-	$0.0116 \pm 0.0004$	$0.0111 \pm 0.0004$	-
$\bar{A}$	$0.159 \pm 0.003$	$0.151 \pm 0.004$	$0.146 \pm 0.006$	$0.156 \pm 0.006$	0.08318
$X_u^\uparrow$	$0.410 \pm 0.006$	$0.415 \pm 0.007$	$0.407 \pm 0.008$	$0.404 \pm 0.008$	0.46128
$X_u^\downarrow$	$0.21 \pm 0.02$	$0.216 \pm 0.019$	$0.18 \pm 0.03$	$0.18 \pm 0.03$	0.29766
$X_d^\uparrow$	$0.11 \pm 0.03$	$0.13 \pm 0.03$	$0.13 \pm 0.02$	$0.12 \pm 0.03$	0.22775
$X_d^\downarrow$	$0.292 \pm 0.008$	$0.292 \pm 0.009$	$0.276 \pm 0.008$	$0.278 \pm 0.008$	0.30174
$A$	3.4	3.2	3.6	3.6	1.74938
$\bar{A}$	0.00002	0.00006	0.13	0.08	1.90801
$A_g$	17	18	11	11	14.27535

**Table 4.** Values of the fitted parameters (including uncertainties) for the QSPDF and QSPDFflex sets, both at NNLO and NNLO+NLL $x$ . In the last three lines we also provide the values of the additional normalization parameters that are determined from the sum rules (indicating the central value only). In the last column we also report the values of the parameters from the determination of Ref. [18], which were given without uncertainty and with the same number of digits shown here.

As we mentioned before, the model is agnostic about QCD interactions and therefore it does not know about the perturbative order and the factorization scheme and scale dependence. In this sense, the good description of the data may seem surprising. However, we have to recall that we are using the model as a provider for a parametrization, and it is the parametrization that works well. We can also guess why it is so. Essentially, the model assumes the original Feynman’s parton model, i.e. the LO approximation of the QCD factorization theorem. It is well known that LO PDFs have shapes that are similar to  $\overline{\text{MS}}$ -scheme NLO and NNLO PDFs at large  $x$ , and so it is likely that the same parametrization, with different parameters, is able to describe also NNLO PDFs. At small  $x$  differences are more marked between various orders, and indeed we had to modify the parametrization of the gluon PDF in the QSPDFflex set to obtain a reasonable description.

It is interesting to compare the values of the model parameters with previous determinations. In table 4 we list the values of the parameters for the four fits considered, namely QSPDF and QSPDFflex, both at NNLO and NNLO+NLL $x$ , as well as those from Ref. [18]. We immediately notice that the values of the parameters is very similar across all our fits. The only exception is  $\bar{b}$ , which is fixed to be equal to  $b$  in the QSPDFflex fits, whose value is very different from the value found in the QSPDF fits. In fact, the very small value of the  $\bar{b}$  parameter in the QSPDF fits, which describes a strong small- $x$  growth of the antiquark valence component, is compensated by a very small value of  $\bar{A}$  computed by the sum rules (see appendix A), making the effect of that term very small.

Comparing our numbers with the values obtained in Ref. [18] we see that there is generally a good agreement between the parameters. In particular, the “effective temperature”  $\bar{x}$ , which governs the large- $x$  drop of all PDFs, is very stable across the various determinations. The quark “potentials” are in good agreement, but they are all smaller in our fits, in particular  $X_d^\uparrow$  and to a lesser extent  $X_u^\downarrow$ . This difference is likely due to the absence in our determination of information from polarized distributions. The  $b$  parameter governing the small- $x$  drop of the valence distributions

is always slightly larger in our fits, as well as the  $\bar{b}$  parameter governing the small- $x$  growth of sea quarks, which in turn determines a larger  $\bar{A}$  coefficient to compensate. In the last three lines of the table the values of the parameters  $A, \bar{A}, A_g$ , determined from the sum rules, are also shown. As the value of  $\bar{A}$  strongly depends on the value of  $\bar{b}$ , it is very different in the various families of fits.

The effect of these different parameters is shown in figure 11, where the PDFs from Ref. [18] are compared with our QSPDF(flex) NNLO sets. We see clear distortions in all distributions between the old and new sets. At high  $x$  all distributions behave in the same way, as this region is predominantly governed by the  $\bar{x}$  parameter which is almost the same for all PDFs and secondarily by the quark potentials  $X_q^{\uparrow\downarrow}$  which are similar. The PDFs of Ref. [18] are all harder at small  $x$ , including the valence distributions, and compensate this with smaller quark sea distributions at medium  $x$  and with smaller valence peaks. These differences are certainly due to the fact that Ref. [18] fits a bunch of DIS data from different experiments, including polarized data, all at a  $Q^2$  scale close to  $\mu_0^2 = 4 \text{ GeV}^2$ , thus containing different information with respect to our dataset. As a comparison, we have computed the  $\chi^2$  of the PDF set of Ref. [18] with our fit setting, finding more than 6000 units, showing that this PDF set is very far from giving an acceptable description of HERA data.

## 6.2 The anti-up anti-down asymmetry in the proton

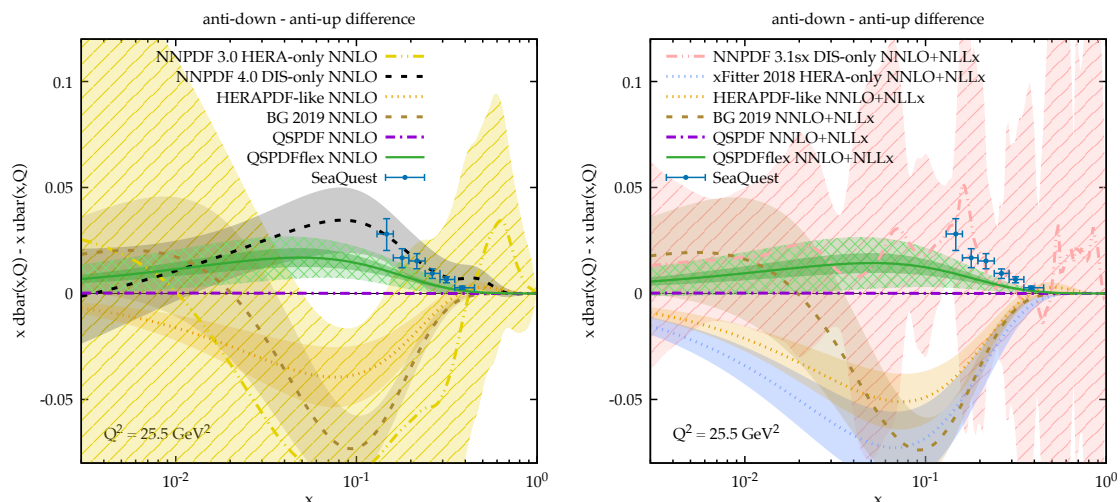
An important implication of the statistical model is that the difference  $\bar{d} - \bar{u}$  is greater than zero. A positive value for the first Mellin moment of this difference was first determined by the NMC experiment [80], which found a defect in the Gottfried sum rule [81]. This confirmed the conjecture by Niegawa and Sasaki [82] and by Feynman and Field [83] that, as a consequence of Pauli principle, in the proton there are more  $\bar{d}$  than  $\bar{u}$ .

The positivity of  $\bar{d} - \bar{u}$  is a consequence of the values of the potentials  $X_{u,d}^{\uparrow\downarrow}$  that we can understand analytically. Looking at Eq. (2.3b) we note that, for each polarization, there is one term that dominates over the other. Using explicitly Eq. (2.6) and grouping terms with the same potential,

$$\begin{aligned} xq_v(x, \mu_0^2) &= AX_q^{\uparrow} h_+(x; b, X_q^{\uparrow}) - \frac{\bar{A}}{X_q^{\uparrow}} h_+(x; \bar{b}, -X_q^{\uparrow}) \\ &\quad + AX_q^{\downarrow} h_+(x; b, X_q^{\downarrow}) - \frac{\bar{A}}{X_q^{\downarrow}} h_+(x; \bar{b}, -X_q^{\downarrow}), \end{aligned} \quad (6.1)$$

we see that, for positive potentials  $X_q^{\uparrow\downarrow} > 0$  as given by the fit, the second term proportional to  $\bar{A}$  in each line is exponentially suppressed with respect to each first term by a factor  $\exp(-2X_q^{\uparrow\downarrow}/\bar{x})$  at medium/large  $x$ . Therefore, the size of the valence distribution is dominated by the  $A$  terms, and in particular by the polarization component characterised by the largest potential, which is  $\max(X_u^{\uparrow}, X_u^{\downarrow}) = X_u^{\uparrow}$  for the up quark and  $\max(X_d^{\uparrow}, X_d^{\downarrow}) = X_d^{\downarrow}$  for the down quark<sup>10</sup> (table 4). Moreover, since the valence distribution of the up quark is larger (roughly by a factor of two) than the valence distribution for the down, it follows that  $\max(X_u^{\uparrow}, X_u^{\downarrow}) > \max(X_d^{\uparrow}, X_d^{\downarrow})$ , namely  $X_u^{\uparrow} > X_d^{\downarrow}$ , which is indeed verified in all fits. From the fit results we also see that the potential for the subdominant polarizations is always larger for the up quark than for the down quark,  $X_u^{\downarrow} > X_d^{\uparrow}$ . While this is not directly related to striking features of the parametrization, it is a property that ensures that the shapes of the valence distributions are in agreement with data and with the sum

<sup>10</sup>Note that since the parametrization of quarks is simply the sum of the two independent polarization components, differing only by the potentials, there is a symmetry in the parametrizations for the exchange  $X_q^{\uparrow} \leftrightarrow X_q^{\downarrow}$ . Therefore, which potential is larger is just a matter choice, i.e. the fit can equivalently find a minimum or its symmetric where the potentials are swapped. The hierarchy shown here reproduces the hierarchy found in the original publications where additional information from polarized data were included.



**Figure 12.** The difference  $\bar{d} - \bar{u}$  in all the PDF sets considered so far, at NNLO (left) and NNLO+NLL $x$  (right). For the QSPDFflex result the total uncertainty band is also shown in lighter green with a crossed pattern. The scale of the plot is  $Q^2 = 25.5 \text{ GeV}^2$  to match that of the SeaQuest data points [84, 85] also shown in the plot.

rules.<sup>11</sup> This inequality immediately implies that  $\bar{d} > \bar{u}$ . Indeed their difference is given by

$$x\bar{d}(x, \mu_0^2) - x\bar{u}(x, \mu_0^2) = \bar{A} \left[ \frac{1}{X_d^\uparrow} h_+(x; \bar{b}, -X_d^\uparrow) + \frac{1}{X_d^\downarrow} h_+(x; \bar{b}, -X_d^\downarrow) - \frac{1}{X_u^\uparrow} h_+(x; \bar{b}, -X_u^\uparrow) - \frac{1}{X_u^\downarrow} h_+(x; \bar{b}, -X_u^\downarrow) \right], \quad (6.2)$$

which is dominated by the smallest potentials, and it is positive if  $\min(X_u^\uparrow, X_u^\downarrow) > \min(X_d^\uparrow, X_d^\downarrow)$ , namely  $X_u^\downarrow > X_d^\uparrow$ , which is exactly the condition mentioned before.<sup>12</sup>

The numerical results are reported in figure 12 for several PDF sets at fixed order (left plot) and with small- $x$  resummation (right plot). We notice that all PDF sets obtained using only HERA data tend to predict a negative  $\bar{d} - \bar{u}$  difference in the valence region, with the exception of the QSPDF and QSPDFflex sets. The inclusion of data sensitive to quark flavours, e.g. charged current data from DIS experiments as in the NNPDF4.0 DIS-only set, twists the situation by predicting a positive  $\bar{d} - \bar{u}$  difference in the valence region. This is further confirmed in global PDF fits (see e.g. Ref. [7]). We thus conclude that HERA data alone not only are not able to predict the flavour separation, but they also tend to be in better agreement with a negative  $\bar{d} - \bar{u}$  difference irrespective of the parametrization used, again with the exception of QSPDFs.

Now let us focus on our QSPDF and QSPDFflex fits. We observe that the latter gives a positive  $\bar{d} - \bar{u}$  distribution, in agreement with the discussion above about the values of the potentials. However, the QSPDF predicts essentially the same values for  $\bar{d}$  and  $\bar{u}$ , giving a vanishing difference. This effect is the result of the very small value of  $\bar{b}$  coming from the fit that forces  $\bar{A}$  to be almost zero. As the difference  $\bar{d} - \bar{u}$  is proportional to  $\bar{A}$ , Eq. (6.2), it becomes consequently very small.

<sup>11</sup>We were not able to prove that this inequality has to be satisfied based on strict conditions. Therefore, we can only say that it is very unlikely that a good fit can violate this hierarchy, and indeed all fits so far led to values satisfying this hierarchy.

<sup>12</sup>This conclusion assumes that  $\bar{A} > 0$ . This is always the case in the fits considered here, but it may not be the case in general. However,  $\bar{A} > 0$  ensures that the anti-quark distributions are positive in the high- $x$  region, so we find it reasonable to assume that this will always be the case in good fits.

We suspect that the small value of  $\bar{b}$  found by the fit is driven by this effect: HERA data favours a negative  $\bar{d} - \bar{u}$ , and the fit finds the value of the parameters that makes it as close as possible to negative, i.e. zero. Conversely, in the QSPDFflex parametrization where  $\bar{b}$  is fixed to  $b$ , this flexibility is no longer present and the difference  $\bar{d} - \bar{u}$  remains significantly positive.

One may wonder why in the QSPDFflex parametrization we are allowed to fix  $\bar{b} = b$  while in the QSPDF parametrization this leads to a sizeable increase in  $\chi^2$ . We suspect that this is due to the greater flexibility of the gluon parametrization, which is less linked to the diffractive (sea) term, which is then more flexible and can thus better describe the antiquark distributions without the need for an extra degree of freedom. We may also guess that when including data sensitive to quark flavours leaving  $\bar{b}$  as a free parameter can still predict a positive  $\bar{d} - \bar{u}$  difference, this time induced by the data and not by a parametrization bias. To see this, we also report in the figure the recent data from the SeaQuest collaboration [84, 85]. We see that they scale in reasonable agreement with QSPDFflex, but they are higher, in closer agreement with the NNPDF4.0 DIS-only set (despite the fact that it does not include them). We will investigate in section 6.3 the stability of this result upon inclusion of additional data.

### 6.3 Additional data

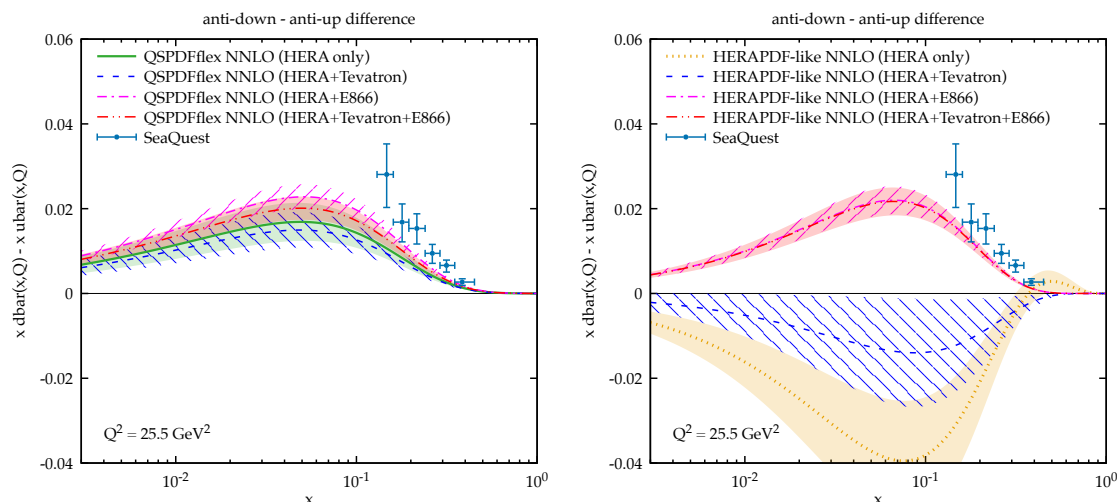
In this section we investigate the possibility of including additional data in the fit, with emphasis on the large- $x$  behaviour of PDFs and in particular on the  $\bar{d} - \bar{u}$  difference. Unfortunately, we are limited by the datasets available in `xFitter` (implementing new dataset ourselves would require a significant amount of work which is far beyond the scope of this article). As we are interested in the large- $x$  region, we identified two datasets that are relevant for us: old fixed-target Drell-Yan from E866 [86] (39 datapoints) and Tevatron CDF and D0  $Z$  rapidity distributions [87, 88] (56 datapoints). In particular, the former data are given as the ratio of proton-deuteron cross section over proton-proton cross section, giving direct access to the up/down antiquark asymmetry.

We have performed fits to HERA+E866, HERA+Tevatron and HERA+E866+Tevatron, using both QSPDFflex and HERAPDF-like parametrization. Unfortunately, the theoretical description of these data in `xFitter` is limited to NLO, so the fits use inconsistently NNLO theory for DIS and DGLAP evolution and NLO for Drell-Yan.<sup>13</sup> We must thus expect some increase in the reduced  $\chi^2$  that we indeed see with both parametrizations. For QSPDFflex, the  $\chi^2/\text{d.o.f}$  increases from the HERA-only value of 1.20 to 1.21 for HERA+Tevatron and 1.25 for HERA+E866 and HERA+E866+Tevatron. For HERAPDF-like parametrization, the  $\chi^2/\text{d.o.f}$  increases from the HERA-only value of 1.21 to 1.26 for HERA+E866 and 1.25 for HERA+E866+Tevatron, while it decreases slightly to 1.20 for HERA+Tevatron. The fact that these variations are very similar among the two parametrizations shows that the QSPDFflex parametrization does a good job in fitting these data as well as HERA data.

As far as PDFs are concerned, the effect of the new data is essentially negligible for QSPDFflex, while there are some differences for the HERAPDF-like fits, especially at high  $x$ . In particular, we focus on the  $\bar{d} - \bar{u}$  difference, shown in figure 13. With the QSPDFflex parametrization, the inclusion of Tevatron data tends to lower slightly the prediction, while in contrast E866 leads to a larger asymmetry, pushing the PDFs closer to the SeaQuest data also shown in the plot. When both datasets are included, the prediction is larger than the HERA-only fit, but still very close to it. Overall, the four predictions are in good agreement and we can conclude that QSPDFflex is stable upon inclusion of different datasets, also confirming that the positive value of  $\bar{d} - \bar{u}$  is a feature of the parametrization.

With the HERAPDF-like parametrization, the inclusion of Tevatron data brings the prediction for  $\bar{d} - \bar{u}$  closer to zero, but still negative in similarity with the HERA-only fit and with similarly

<sup>13</sup>We note however that the lack of NNLO corrections should be less significant for E866, as the data are given as a ratio of cross sections which is less sensitive to perturbative corrections.



**Figure 13.** The difference  $\bar{d} - \bar{u}$  in QSPDFflex (left) and HERAPDF-like (right) fits to different datasets. The scale of the plot is  $Q^2 = 25.5 \text{ GeV}^2$  to match that of the SeaQuest data points [84, 85] also shown in the plot.

large uncertainty. Once E866 data are included, the prediction becomes positive and with smaller uncertainty, and it is close to the SeaQuest datapoints. Moreover, it is stable upon inclusion of Tevatron data. This shows that the HERAPDF parametrization is more flexible at high  $x$  and it leads to inaccurate results when data are not sufficiently constraining, while it provides a stable result once constraining data are included. This behaviour is what is usually expected from a fit with unbiased parametrization.

We observe that the HERAPDF-like result with E866 data is very similar to the analogous QSPDFflex result. This means that the QSPDFflex is also accurate, with the difference that it was so also before the inclusion of E866 data. This is the effect of the (physically motivated) bias present in this parametrization.

#### 6.4 Polarized PDFs

As already stressed, the parametrization of the quark PDFs is made in terms of contributions from the individual quark polarizations. This means that the same parameters would in principle allow to determine also polarized PDFs within the statistical model. Using Eq. (2.2), the parametrizations for polarized quark and antiquark PDFs are given by

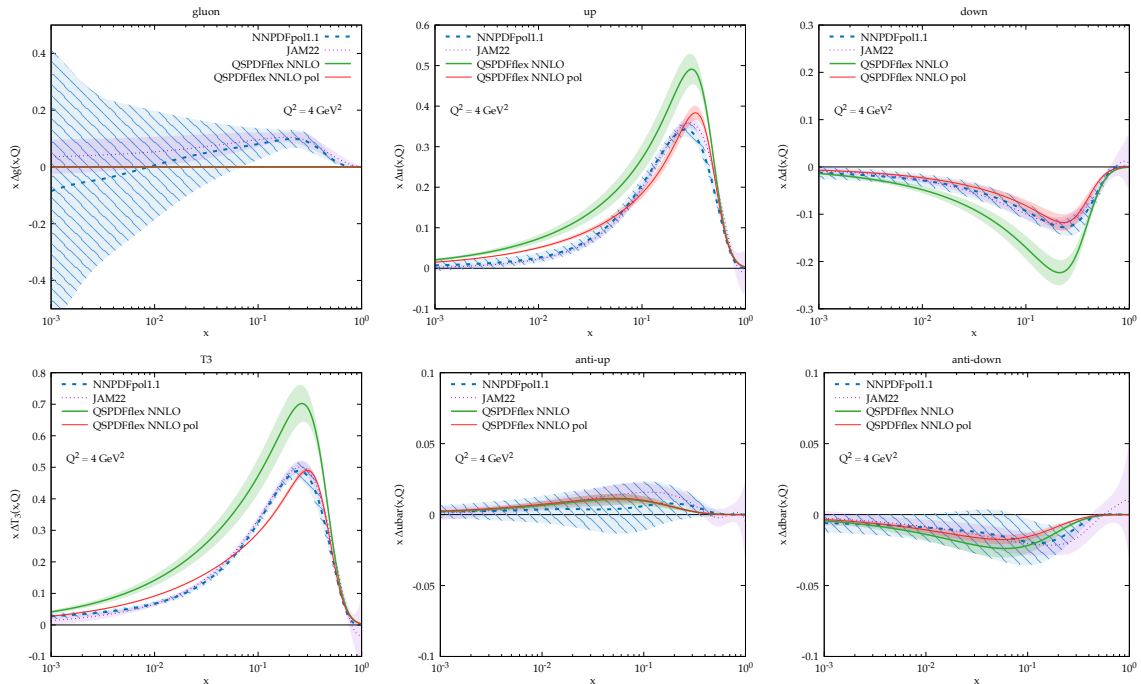
$$\begin{aligned} x\Delta q(x, \mu_0^2) &\equiv xq^\uparrow(x, \mu_0^2) - xq^\downarrow(x, \mu_0^2) \\ &= A[X_q^\uparrow h_+(x; b, X_q^\uparrow) - X_q^\downarrow h_+(x; b, X_q^\downarrow)], \end{aligned} \quad (6.3a)$$

$$\begin{aligned} x\Delta \bar{q}(x, \mu_0^2) &\equiv x\bar{q}^\uparrow(x, \mu_0^2) - x\bar{q}^\downarrow(x, \mu_0^2) \\ &= \bar{A} \left[ \frac{1}{X_q^\downarrow} h_+(x; \bar{b}, -X_q^\downarrow) - \frac{1}{X_q^\uparrow} h_+(x; \bar{b}, -X_q^\uparrow) \right], \end{aligned} \quad (6.3b)$$

while the model assumes no polarization for gluons, i.e.  $\Delta g = 0$  (we will comment on this assumption later in this section).

Of course, a fit of unpolarized PDFs is not able to distinguish the individual polarizations. However, if polarized data are included in the fit, it becomes possible to simultaneously determine polarized and unpolarized PDFs without changing the parametrization (i.e., without adding extra parameters). The agreement (see figure 11) at  $x \gtrsim 0.4$  with the valence distributions found in





**Figure 14.** Comparison of polarized PDFs obtained from our QSPDFflex NNLO fit (with and without imposing the triplet sum rule) and the NNPdfpol1.0 [89] and NNPdfpol1.1 [90] sets.

[18], where the information from polarized scattering was well described, lets us reasonably think that the statistical model may be able to describe both unpolarized and polarized distributions in a satisfactory way. Testing this in practice requires quite some work, as the current `xFitter` infrastructure should be modified to introduce this possibility, and we therefore leave this task to future work. Here we perform a simpler test.

First of all, we consider the QSPDFflex set fitted from unpolarized data and plot polarized PDFs Eq. (6.3) obtained with those parameters out of the box. As stressed, we do not expect to find agreement with direct determinations from polarized data [37, 89–97], but we want to understand how far we are. Therefore, in figure 14 we show the polarized PDFs constructed using Eqs. (6.3) from our QSPDFflex NNLO unpolarized fit, compared with NNPdfpol1.1 [90] and JAM22 [95].<sup>14</sup> Specifically, we plot  $\Delta u$ ,  $\Delta \bar{u}$ ,  $\Delta d$ ,  $\Delta \bar{d}$  and also the triplet combination

$$\Delta T_3 = \Delta u + \Delta \bar{u} - \Delta d - \Delta \bar{d} \quad (6.4)$$

at the fit scale  $Q^2 = 4\text{GeV}^2$ . We also show the gluon  $\Delta g$  for completeness.

We observe that the QSPDFflex anti-quark polarized PDFs are perfectly compatible with NNPdfpol1.1 within the (large) uncertainty of the latter. They are also close to the JAM22 determination, except in the region  $0.1 \lesssim x \lesssim 0.4$  where JAM22 is slightly larger (in absolute value). The quark polarized PDFs instead are not in agreement, as  $\Delta u$ ,  $\Delta d$  and  $\Delta T_3$  are all larger (in absolute value) than their NNPdfpol1.1 and JAM22 counterparts. Nevertheless, the shapes are very similar. Finally, we notice that the NNPdfpol1.1 and JAM22 find a polarized gluon which

<sup>14</sup>We have considered the latest version of these two families of polarized PDF fits. To our knowledge, the DSSV polarized PDFs [37, 91, 93] are not publicly available in LHAPDF format. In any case, the PDFs shown are good representative of polarized PDFs, as for instance the most recent DSSV determination is in good agreement [93] with NNPdfpol1.1. Note also that for JAM22 we have considered only the positive gluon solution, as the negative one strongly violates the positivity bound  $|\Delta g| \leq g$ , see the discussion in Ref. [98].



is significantly larger than zero at medium/large  $x$ , as confirmed also by the DSSV analysis [37], while in our parametrization  $\Delta g$  is assumed to be zero at the fit scale  $Q^2 = 4\text{GeV}^2$ .

All in all, the agreement found is rather remarkable, given that our set is constructed without any information from polarized data. While as already mentioned performing a simultaneous fit of unpolarized and polarized data requires a significant amount of work, we can try to include some information on polarized PDFs in our fit with a little effort. Specifically, isospin symmetry (which is expected to hold to high accuracy) implies the so-called triplet sum rule

$$\int_0^1 dx \Delta T_3(x, \mu^2) = \left| \frac{g_A}{g_V} \right| = 1.2754 \pm 0.0013 \quad (6.5)$$

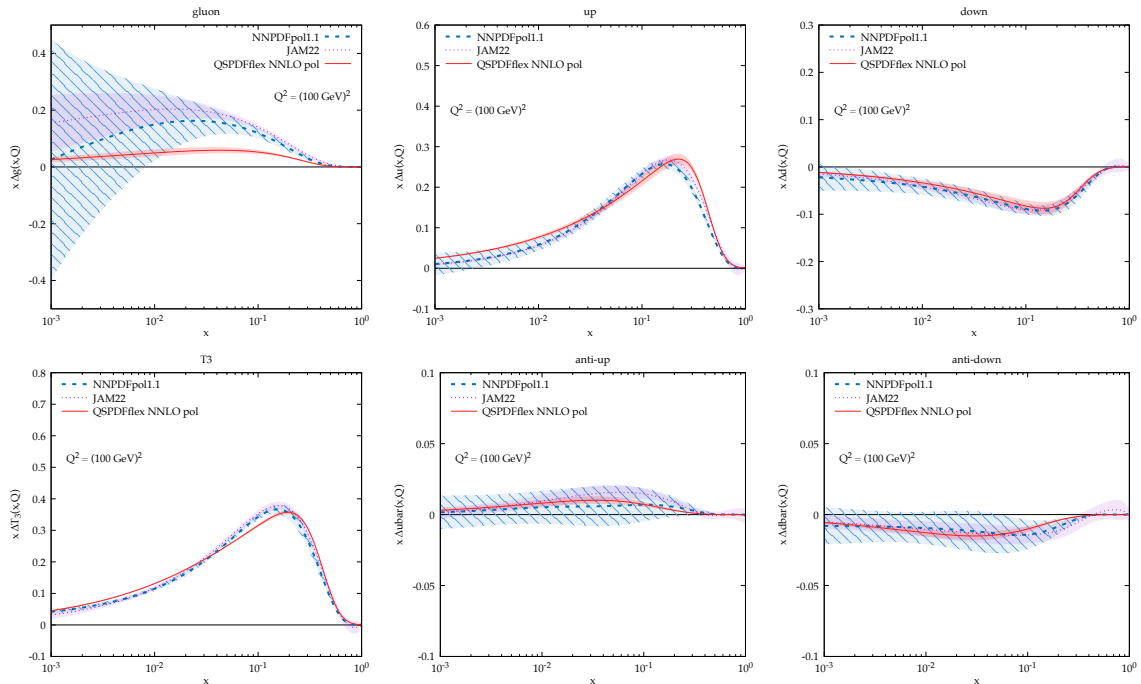
where  $g_A$  and  $g_V$  are the axial and vector electroweak couplings which can be derived from neutron decay. The reported experimental value is taken from the latest PDG average [99]. Notably, the first moment on the left-hand side is scale independent, and so the sum rule is valid at any scale.

We have thus performed an additional NNLO fit to HERA data with the QSPDFflex parametrization in which we impose the triplet sum rule. Practically, we do so as if it were a datapoint, to account for the experimental uncertainty. The result of the fit has a slightly worse  $\chi^2/\text{d.o.f.}$ : from 1334/1109 without the sum rule to 1342/1110 with the sum rule. This deterioration is driven by the charged-current positron subset, whose partial  $\chi^2$  increases from 55 to 62, over 39 datapoints, while all other subset are essentially unchanged. Interestingly, the partial  $\chi^2$  from the triplet sum rule is extremely small (close to zero), despite the very small (permille) uncertainty on the experimental value, showing that the QSPDFflex parametrization is able to easily accommodate such a physical constraint.

The resulting polarized PDFs are shown in the same figure 14, in red. The simple imposition of the sum rule improves the agreement with NNPDFpol1.1 and JAM22 significantly. Among the quarks, only the shape of  $\Delta u$ , which in turn affects  $\Delta T_3$ , is not compatible with NNPDFpol1.1, although it is very close. We are thus tempted to hope that if a single constraint was able to achieve such a good agreement, the inclusion of polarized data in the fit could further improve the agreement without a significant deterioration of the fit quality. We must also note that the polarized PDF determinations of NNPDFpol1.1, JAM22 and DSSV have been obtained using NLO theory, while our fit is NNLO accurate: this difference may also contribute to the disagreement. We also stress that the unpolarized PDFs of our new fit are essentially unchanged with respect to the QSPDFflex set without the triplet sum rule.

The difference on the polarized gluon PDF deserves a separate discussion. The assumption of a zero gluon polarization at the fit scale  $\mu_0$  descends from the equilibrium condition which is at the core of the statistical model, which in turn implies a vanishing potential for the gluon PDF leading to the Plank distribution Eq. (2.3c). However, DGLAP evolution brings partons away from equilibrium, as it only describes splittings and not recombination. Indeed, recombination is supposed to be a necessary ingredient only in a strongly interacting regime, which may happen either at very high density (i.e. at very small  $x$ , leading to saturation) or in the non-perturbative region at low  $Q^2$ . Therefore, the QSPDF(flex) parametrization is best suited to describe PDFs at the border between non-perturbative and perturbative regimes, namely at the border between strong dynamics and DGLAP realm. The choice  $\mu_0 = 2 \text{ GeV}$  for this border may not be optimal — indeed, admittedly 2 GeV is a scale which is certainly in the perturbative regime. This choice was used in the original literature [18, 30] and justified a posteriori by the ability to obtain a good description of the data. However, meanwhile data have improved, leading in particular to the striking evidence of a polarized gluon at 2 GeV from RICH data, which may be due to DGLAP evolution from a lower equilibrium scale.

To understand if this may be the case, we have tried to evolve the PDFs to a higher scale, specifically to  $Q = 100 \text{ GeV}$ . For our set (we now consider only the one obtained with the triplet sum



**Figure 15.** Same as figure 14, but at  $Q = 100$  GeV and showing only the QSPDFflex NNLO result with the triplet sum rule compared with the other public sets.

rule) we used APFEL++ [39, 100] to perform the polarized evolution (at NLO). The resulting PDFs are shown in figure 15. We notice that the QSPDFflex polarized gluon, thanks to the evolution, is no longer vanishing, and in particular it is positive as the NNPDFpol1.1 and JAM22 ones. It is still smaller than those and not compatible with them at medium  $x$ , as a consequence of the differences at the starting scale. We also observe that the agreement of quark PDFs improves after evolution.

We thus conclude that it is very likely that the optimal scale for the QSPDF(flex) parametrization is a smaller one. This has been explored in the literature on the statistical parametrization. For example, Refs. [25, 27] use  $\mu_0 = 1$  GeV. Nevertheless, they find that even starting from such a low scale data are better described if a non-zero gluon polarization is assumed at the fit scale. Moreover, DGLAP evolution from 1 GeV to 2 GeV at fixed order, be it NLO or NNLO or even higher, is likely inaccurate due to the large value of  $\alpha_s$  in this region that makes missing higher order contributions very sizable. Choosing a starting scale below 1 GeV would make this issue even worse and must thus be avoided. We conclude that further studies are needed to understand what the best strategy could be. Perhaps allowing a non-zero gluon polarization is the simplest solution, but a functional form compatible with the model assumptions must be worked out.

## 6.5 Future directions

The results presented so far show that the QSPDF(flex) parametrization has a number of virtues, due to the possibility of producing unpolarized and polarized PDFs with sensible physical behaviour with a very small number of parameters. For this reason, on top of being a useful complement to standard parametrizations, it is worth considering it for further studies. In this respect, it would be interesting to perform a global fit, including unpolarized data from other DIS experiment as well as collider data from Tevatron and LHC, and eventually also polarized data.

As we already mentioned, for the moment this task cannot be performed straight away in `xFitter` due to the limited amount of available datasets and the lack of theoretical predictions for

polarized observables. In any case, it is easy to foresee that the simple parametrizations that we have considered so far will not be suitable to describe a large variety of data. For such a study, the statistical model parametrization must be made more flexible. We can increase the flexibility of the parametrization in three ways:

- adopting Eq. (2.5) rather than Eq. (2.6) for defining the coefficients  $C^{\uparrow\downarrow}, \bar{C}^{\uparrow\downarrow}$ , which introduces two extra parameters for each quark flavour to be fitted and thus gives more flexibility to model the medium/high- $x$  region of quark PDFs;<sup>15</sup>
- modifying the low- $x$  behaviour of quark PDFs in the same way we did for the gluon, Eq. (5.1), namely multiplying each term in the parametrization by a polynomial in  $\log x$ , which gives much more flexibility at small  $x$  without altering the large- $x$  region where the model is physically motivated;
- providing an independent parametrization for the strange quark PDF, needed for fitting data beyond HERA, using e.g. the parametrization proposed in Ref. [101] in the context of the statistical model;
- changing (lowering) the fit scale  $\mu_0$  at which PDFs are parametrized to find an optimal value where the assumptions of the model are reasonably satisfied, and possibly introduce a parametrization for polarized gluons (see discussion at the end of section 6.4).

All these modifications are perfectly consistent with the model hypotheses, and can thus be viewed as a natural extension of the study in this work. As a result of this extension, the parametrization would depend on many more parameters, reducing some of the advantages of the statistical parametrization, but keeping the important physical properties, e.g. the possibility of determining unpolarized and polarized PDFs with the same parameters. Note also that some of the new parameters (in particular those modelling the small- $x$  behaviour) may be redundant and could possibly be eliminated, but this can only be decided after performing the fit.

The extended parametrization proposed above would be much more flexible, making it comparable with other fixed-form flexible parametrizations on the market. Despite the flexibility, such an improved QSPDF parametrization would still differ from other parametrizations in one key aspect: the  $x \rightarrow 1$  behaviour. Indeed, the statistical model is characterized by a non-vanishing limit in  $x = 1$ , Eq. (2.7), while all other parametrizations, including the very flexible NNPDF one, assume a vanishing power-like behaviour of the form  $(1-x)^\alpha$  with  $\alpha > 0$ . We have already seen in section 5 that differences are significant for  $x \gtrsim 0.7$ : the inclusion of data sensitive to such high values of  $x$ , such as LHC jet data or future EIC data, would thus allow us to test which functional form is more suitable for their description. On top of looking at the fit quality, it could be possible to perform a simple test to verify whether the data prefer vanishing or non-vanishing PDFs in  $x = 1$ : after having fitted high- $x$  data with the extended QSPDF parametrization, one could multiply each PDF parametrization by a  $(1-x)^\alpha$  factor, with the same value of  $\alpha$  for all PDFs,<sup>16</sup> and fit data again. If the fit selects a (positive) value of  $\alpha$  significantly different from zero *and* the  $\chi^2$  reduces by more than one unity (corresponding to the extra fitted parameter), then one must conclude that data prefer vanishing PDFs in  $x = 1$ . Conversely, if at least one of the conditions above are not satisfied, the non-vanishing behaviour predicted by the statistical model has to be considered as compatible with the data.

<sup>15</sup>In fact, Eq. (2.6) is an approximation of Eq. (2.5), so the use of the latter is the most consistent way of using the model to parametrize quark PDFs.

<sup>16</sup>Using different values of  $\alpha$  for each PDFs may lead to overfitting, as the extended QSPDF parametrization is already rather flexible at high  $x$ . Moreover here the goal is to understand whether data prefer vanishing or a non-vanishing PDFs in  $x = 1$ , and this is better tested by the simplest modification of the parametrization that adds the least flexibility.

On top of this high- $x$  test, another powerful validation of the model consists in the ability of fitting in a satisfactory way both unpolarized and polarized data. The results of section 6.4 are very encouraging, but the actual test can only come once polarized and more unpolarized data are included in the fit. As discussed in section 6.4, a key issue to be faced is the polarized gluon PDF: while at equilibrium it makes sense to assume zero polarization, it is not obvious that it is possible to really start the fit at the equilibrium scale, thus requiring the introduction of a parametrization for the polarized gluon PDF. Finding such a parametrization in a way that is compatible with the statistical model requires a theory study that is left to a future work.

## 7 Conclusions

In this work we have considered a PDF parametrization inspired by a statistical model of the proton dynamics and tested it in fits to HERA data through the public `xFitter` code. The idea behind the use of such parametrization is the opposite of the standard practice: while usually one tries to consider very flexible functional forms to reduce the parametrization bias, here we consider very biased functional forms to reduce the number of fitting parameters. The hope is that the bias introduced in the parametrization be justified by the physical model which the functional form is derived from, thus leading to a reasonably good fit with all the advantages of a small set of parameters. To our knowledge, this is the first PDF fit based on the statistical model performed at NNLO and NNLO+NLL $x$  accuracy (previous ones were only NLO accurate).

We have considered two versions of the parametrization. One is the simplest original version coming from the model [18], denoted QSPDF, which has 9 free parameters. The other one is a variant in which we add more flexibility to the gluon at small  $x$ , denoted QSPDFflex, which depends on 11 free parameters. The QSPDF parametrization allows to reasonably describe HERA data, but the quality of the fit is not particularly high, due to an extremely limited functional form of the gluon PDF which does not allow to describe low- $x$  data well. The QSPDFflex parametrization, instead, leads to a good description of the data comparable with other PDF parametrizations which depend on more parameters. For instance, the  $\chi^2$  obtained with QSPDFflex is essentially the same as that obtained with a HERAPDF-like parametrization. We can thus conclude that the QSPDFflex parametrization is working well and it can provide a useful complement to standard parametrization, e.g. to study parametrization bias (or equivalently to estimate a parametrization uncertainty) or as a new starting point for constructing more flexible parametrizations.

In terms of PDF comparison, we have noticed that the shapes predicted by QSPDF and QSPDFflex differ in several respects from the ones of a HERAPDF-like parametrization. The valence distributions differ noticeably, and the QSPDF(flex) results are in better agreement with other PDF sets (NNPDF, BG) based on more flexible parametrizations. Conversely, at small  $x$  the QSPDF shapes, in particular for the gluon, are very constrained and thus differ from all other PDF sets on the market. The QSPDFflex set cures this problem, allowing the gluon to behave in better agreement with other public sets. It has to be noticed however that the shape of the gluon at small  $x$  is very sensitive also to the dataset and the theory ingredients, and so different sets may differ visibly.

Related to this last observation, we have also considered fits in which theoretical predictions are supplemented by small- $x$  resummation. In these fits we expected the gluon to be more in agreement among different sets, because the inclusion of the resummation stabilises the fit in the small- $x$  region and predicts a gluon that tends to rise more steeply. We have indeed found a good agreement between various resummed fits on the market and our QSPDF(flex) fit, as well as a HERAPDF-like fit. The inclusion of small- $x$  resummation also produces a reduction of the  $\chi^2$ , in agreement with previous studies [47, 48, 71].

Finally, we have made some considerations on how our results impact the statistical model itself. The success of the QSPDF(flex) parametrization in describing HERA data can be seen as a sort of validation of the model. Also the fact that QSPDFflex predicts a positive  $\bar{d} - \bar{u}$  distribution even if the HERA data do not contain enough information to separate these flavours can be seen as an indication that the model is reliable. We have verified that this prediction is stable upon inclusion of additional data at high  $x$ . Moreover, if we were to trust the model, then it would allow to describe with the very same parameters polarized PDFs as well. We have verified that our fit to unpolarized data produces polarized PDFs that are remarkably similar to those fitted from data, and the simple inclusion of the triplet sum rule in the fit further improves the agreement. This shows that it is probably possible to simultaneously fit with the same parameters both unpolarized and polarized data, and we leave this investigation to future work.

The QSPDF and QSPDFflex sets at NNLO and NNLO plus small- $x$  resummation can be downloaded at the address [l.infn.it/qspdf](http://l.infn.it/qspdf). These sets should not be regarded as general-purpose PDFs, because they have been obtained using a reduced dataset and the parametrization adopted is very minimal. When trying to fit more data, including other DIS experiments and collider (LHC) measurements, it is very likely that the very simple parametrizations proposed, even the QSPDFflex one, will not be able to achieve a satisfactory fit quality. To get a competitive fit, it is possible to improve the parametrization in three respects. First, it is possible to parametrize the strange PDF independently, using e.g. the formulation of Ref. [101]. Second, we can give more flexibility to the high- $x$  region by using less constrained values for the  $C_q^{\uparrow\downarrow}$  parameters, e.g. introducing the so-called “transverse potentials”, Eq. (2.5). Third, we can further model the small- $x$  region by introducing polynomials in  $\log x$  [71] as we did for the gluon, Eq. (5.1), also for the quark PDFs. Including also polarized data in the fit, such as those from RHIC, further requires an extension of the model to introduce a polarized gluon PDF. We leave the study of these extensions to future work.

## Acknowledgments

We are grateful to Francesco Tramontano for collaboration during various phases of this work. We thank Mandy Cooper-Sarkar and Sasha Glazov for useful comments on the manuscript. The work of MB is partly supported by the Italian Ministry of University and Research (MUR) grant PRIN 2022SNA23K funded by the European Union – Next Generation EU.

## A Implementation of the sum rules

The parameters of the fit are constrained by the quark number and momentum sum rules:

$$2 = \int_0^1 dx u_v(x, \mu_0^2), \quad (\text{A.1})$$

$$1 = \int_0^1 dx d_v(x, \mu_0^2), \quad (\text{A.2})$$

$$1 = \int_0^1 dx x [g(x, \mu_0^2) + u(x, \mu_0^2) + \bar{u}(x, \mu_0^2) + d(x, \mu_0^2) + \bar{d}(x, \mu_0^2) + s(x, \mu_0^2) + \bar{s}(x, \mu_0^2)]. \quad (\text{A.3})$$

Usually, they are used to fix the normalization of  $u_v$ ,  $d_v$  and the gluon respectively. However, the parametrization Eq. (2.3) does not have an overall normalization factor for valence distributions, making the implementation of the quark number sum rules non trivial.

The parametrization Eq. (2.3) for valence quarks has the form of the sum of two contributions, one proportional to the parameter  $A$  and the other to  $\bar{A}$ . Crucially, these two parameters are the same for the up and the down quarks. Therefore, it is still possible to use the quark number sum

rules Eqs. (A.1) and (A.2) to determine  $A$  and  $\bar{A}$ , but in order to do so we have to solve the algebraic system

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} K_u & \bar{K}_u \\ K_d & \bar{K}_d \end{pmatrix} \begin{pmatrix} A \\ \bar{A} \end{pmatrix} \equiv \mathcal{K} \begin{pmatrix} A \\ \bar{A} \end{pmatrix} \quad (\text{A.4})$$

with

$$K_q = \int_0^1 \frac{dx}{x} [C_q^\uparrow h_+(x; b, X_q^\uparrow) + C_q^\downarrow h_+(x; b, X_q^\downarrow)], \quad (\text{A.5})$$

$$\bar{K}_q = - \int_0^1 \frac{dx}{x} [\bar{C}_q^\uparrow h_+(x; \bar{b}, -X_q^\downarrow) + \bar{C}_q^\downarrow h_+(x; \bar{b}, -X_q^\uparrow)]. \quad (\text{A.6})$$

The solution is given by

$$A = \frac{1}{\det(\mathcal{K})} \det \begin{pmatrix} 2 & \bar{K}_u \\ 1 & \bar{K}_d \end{pmatrix} \quad \bar{A} = \frac{1}{\det(\mathcal{K})} \det \begin{pmatrix} K_u & 2 \\ K_d & 1 \end{pmatrix}. \quad (\text{A.7})$$

In order to implement this procedure in `xFitter`, we have introduced a new PDF decomposition in which the valence parts of quarks and antiquarks (those proportional to  $A$  and  $\bar{A}$  respectively) are considered as independent PDFs, as well as the diffractive (sea) term. After the determination of  $A$  and  $\bar{A}$  according to Eq. (A.7), they are combined together to form the parametrization for valence quarks and antiquarks as in Eq. (2.3).

The sum rules integrals are computed numerically in `xFitter`. In particular, the integral is divided into two regions, one in  $0.1 < x < 1$  which is sampled linearly, and one in  $10^{-6} < x < 0.1$  which is sampled logarithmically. Therefore, the generic sum rule integrals are approximated as

$$\int_0^1 dx x^N f(x, \mu_0^2) \simeq \int_{x_0}^1 dx x^N f(x, \mu_0^2) \quad (N = 0, 1) \quad (\text{A.8})$$

with  $x_0 = 10^{-6}$ . The neglected region below this value is usually harmless. However, if the integrand is divergent and not integrable in  $x = 0$ , the approximation still gives a finite value thanks to the cutoff  $x_0$ . This may be problematic, as the non-integrability is a manifestation of bad values of the parameters of the PDFs, which must be avoided in the fit. The divergence of the integral is thus a way to put barriers to some parameters.

To this end, it is important to improve the approximation Eq. (A.8) by adding the contribution below  $x_0$ . We assume that all PDFs at small  $x$  have a power like behaviour

$$f(x, \mu_0^2) \xrightarrow{x \rightarrow 0} \alpha x^\beta, \quad (\text{A.9})$$

which is the case for most parametrizations, including the ones we use in this work. The integral from 0 to  $x_0 = 10^{-6}$  can then be approximated by

$$\int_0^{x_0} dx x^N f(x, \mu_0^2) \simeq \alpha \frac{x_0^{\beta+N+1}}{\beta + N + 1}.$$

We have implemented this additional contribution in `xFitter`, with a numerical extrapolation of  $\beta$  and  $\alpha$ . In this way, values of  $\beta \leq -1 - N$  become also practically forbidden, thus providing the aforementioned barriers to the proper parameters. Effectively, for the parametrization of Sect. 2, this constraint forces the following conditions on the  $b$  parameters:  $b, \bar{b}, b_g > 0$  and  $\tilde{b} > -1$ .

## References

- [1] H1, ZEUS collaboration, H. Abramowicz et al., *Combination of measurements of inclusive deep inelastic  $e^\pm p$  scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J. C* **75** (2015) 580 [[1506.06042](#)].



- [2] S. Alekhin, J. Blümlein, S. Moch and R. Placakyte, *Parton distribution functions,  $\alpha_s$ , and heavy-quark masses for LHC Run II*, *Phys. Rev. D* **96** (2017) 014011 [[1701.05838](#)].
- [3] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, *Phys. Rev. D* **103** (2021) 014013 [[1912.10053](#)].
- [4] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin and R. S. Thorne, *Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs*, *Eur. Phys. J. C* **81** (2021) 341 [[2012.04684](#)].
- [5] NNPDF collaboration, R. D. Ball et al., *The path to proton structure at 1% accuracy*, *Eur. Phys. J. C* **82** (2022) 428 [[2109.02653](#)].
- [6] A. Accardi, T. J. Hobbs, X. Jing and P. M. Nadolsky, *Deuterium scattering experiments in CTEQ global QCD analyses: a comparative investigation*, *Eur. Phys. J. C* **81** (2021) 603 [[2102.01107](#)].
- [7] ATLAS collaboration, G. Aad et al., *Determination of the parton distribution functions of the proton using diverse ATLAS data from pp collisions at  $\sqrt{s} = 7, 8$  and  $13$  TeV*, *Eur. Phys. J. C* **82** (2022) 438 [[2112.11266](#)].
- [8] CMS collaboration, A. Tumasyan et al., *Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **02** (2022) 142 [[2111.10431](#)].
- [9] X. Ji, *Parton Physics on a Euclidean Lattice*, *Phys. Rev. Lett.* **110** (2013) 262002 [[1305.1539](#)].
- [10] A. V. Radyushkin, *Quasi-parton distribution functions, momentum distributions, and pseudo-parton distribution functions*, *Phys. Rev. D* **96** (2017) 034025 [[1705.01488](#)].
- [11] H.-W. Lin et al., *Parton distributions and lattice QCD calculations: a community white paper*, *Prog. Part. Nucl. Phys.* **100** (2018) 107 [[1711.07916](#)].
- [12] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys. B* **126** (1977) 298.
- [13] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438.
- [14] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, *Sov. Phys. JETP* **46** (1977) 641.
- [15] NNPDF collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663 [[1706.00428](#)].
- [16] C. Bourrely, F. Buccella, G. Miele, G. G. Migliore, J. Soffer et al., *Fermi-Dirac distributions for quark partons*, *Z. Phys. C* **62** (1994) 431 [[hep-ph/9410375](#)].
- [17] F. Buccella, G. Miele and N. Tancredi, *Quantum statistical parton distributions and the spin crisis*, *Prog. Theor. Phys.* **96** (1996) 749 [[hep-ph/9604230](#)].
- [18] C. Bourrely, J. Soffer and F. Buccella, *A Statistical approach for polarized parton distributions*, *Eur. Phys. J. C* **23** (2002) 487 [[hep-ph/0109160](#)].
- [19] C. Bourrely, F. Buccella and J. Soffer, *Recent tests for the statistical parton distributions*, *Mod. Phys. Lett. A* **18** (2003) 771 [[hep-ph/0211389](#)].
- [20] C. R. V. Bourrely, J. Soffer and F. Buccella, *The Statistical parton distributions: Status and prospects*, *Eur. Phys. J. C* **41** (2005) 327 [[hep-ph/0502180](#)].
- [21] C. Bourrely, J. Soffer and F. Buccella, *The Extension to the transverse momentum of the statistical parton distributions*, *Mod. Phys. Lett. A* **21** (2006) 143 [[hep-ph/0507328](#)].
- [22] C. Bourrely, F. Buccella and J. Soffer, *Semiinclusive DIS cross sections and spin asymmetries in the quantum statistical parton distributions approach*, *Phys. Rev. D* **83** (2011) 074008 [[1008.5322](#)].

- [23] C. Bourrely, F. Buccella and J. Soffer, *The transverse momentum dependent statistical parton distributions revisited*, *Int. J. Mod. Phys. A* **28** (2013) 1350026 [[1302.4281](#)].
- [24] C. Bourrely, F. Buccella and J. Soffer,  *$W^\pm$  bosons production in the quantum statistical parton distributions approach*, *Phys. Lett. B* **726** (2013) 296 [[1308.3567](#)].
- [25] C. Bourrely and J. Soffer, *Statistical description of the proton spin with a large gluon helicity distribution*, *Phys. Lett. B* **740** (2015) 168 [[1408.7057](#)].
- [26] F. Buccella and S. Sohaily, *A Check-up for the Statistical Parton Model*, *Mod. Phys. Lett. A* **30** (2015) 1550203 [[1412.7683](#)].
- [27] C. Bourrely and J. Soffer, *New developments in the statistical approach of parton distributions: tests and predictions up to LHC energies*, *Nucl. Phys. A* **941** (2015) 307 [[1502.02517](#)].
- [28] C. Bourrely and J. Soffer, *Statistical approach of pion parton distributions from Drell–Yan process*, *Nucl. Phys. A* **981** (2019) 118 [[1802.03153](#)].
- [29] J. Soffer and C. Bourrely, *On the flavor structure of the light-quark sea distributions*, *Nucl. Phys. A* **991** (2019) 121607.
- [30] F. Buccella, S. Sohaily and F. Tramontano, *Low  $Q^2$  boundary conditions for DGLAP equations dictated by quantum statistical mechanics*, *J. Stat. Mech.* **1907** (2019) 073302.
- [31] C. Bourrely, F. Buccella and J.-C. Peng, *A new extraction of pion parton distributions in the statistical model*, *Phys. Lett. B* **813** (2021) 136021 [[2008.05703](#)].
- [32] C. Bourrely, W.-C. Chang and J.-C. Peng, *Pion Partonic Distributions in the Statistical Model from Pion-induced Drell-Yan and  $J/\Psi$  Production Data*, *Phys. Rev. D* **105** (2022) 076018 [[2202.12547](#)].
- [33] F. Buccella, *Status of the Quantum Statistical Approach to the Parton Distributions*, *PoS CORFU2021* (2022) 003.
- [34] L. Bellantuono, R. Bellotti and F. Buccella, *Planck formula for the gluon parton distribution in the proton*, *Mod. Phys. Lett. A* **38** (2023) 2350039 [[2201.07640](#)].
- [35] C. Bourrely, F. Buccella, W.-C. Chang and J.-C. Peng, *Extraction of Kaon Partonic Distribution Functions from Drell-Yan and  $J/\psi$  Production Data*, [[2305.18117](#)].
- [36] F. Silveti, *Resummation phenomenology and PDF determination for precision QCD at the LHC*, phd thesis, Sapienza, University of Rome, 9, 2023. [2403.20315](#).
- [37] D. de Florian, R. Sassot, M. Stratmann and W. Vogelsang, *Evidence for polarization of gluons in the proton*, *Phys. Rev. Lett.* **113** (2014) 012001 [[1404.4293](#)].
- [38] xFITTER DEVELOPERS TEAM collaboration, V. Bertone et al., *Impact of the heavy quark matching scales in PDF fits*, *Eur. Phys. J. C* **77** (2017) 837 [[1707.05343](#)].
- [39] APFEL collaboration, V. Bertone, S. Carrazza and J. Rojo, *APFEL: A PDF Evolution Library with QED corrections*, *Comput. Phys. Commun.* **185** (2014) 1647 [[1310.1394](#)].
- [40] M. Botje, *QCDNUM: Fast QCD Evolution and Convolution*, *Comput. Phys. Commun.* **182** (2011) 490 [[1005.1481](#)].
- [41] R. S. Thorne and R. G. Roberts, *An Ordered analysis of heavy flavor production in deep inelastic scattering*, *Phys. Rev. D* **57** (1998) 6871 [[hep-ph/9709442](#)].
- [42] R. S. Thorne, *A Variable-flavor number scheme for NNLO*, *Phys. Rev. D* **73** (2006) 054019 [[hep-ph/0601245](#)].
- [43] R. S. Thorne, *Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections*, *Phys. Rev. D* **86** (2012) 074017 [[1201.6180](#)].
- [44] S. Forte, E. Laenen, P. Nason and J. Rojo, *Heavy quarks in deep-inelastic scattering*, *Nucl. Phys. B* **834** (2010) 116 [[1001.2312](#)].



- [45] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *The impact of the final HERA combined data on PDFs obtained from a global fit*, *Eur. Phys. J. C* **76** (2016) 186 [[1601.03413](#)].
- [46] I. Abt, A. M. Cooper-Sarkar, B. Foster, V. Myronenko, K. Wichmann and M. Wing, *Study of HERA ep data at low  $Q^2$  and low  $x_{Bj}$  and the need for higher-twist corrections to standard perturbative QCD fits*, *Phys. Rev. D* **94** (2016) 034032 [[1604.02299](#)].
- [47] R. D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo and L. Rottoli, *Parton distributions with small- $x$  resummation: evidence for BFKL dynamics in HERA data*, *Eur. Phys. J. C* **78** (2018) 321 [[1710.05935](#)].
- [48] xFITTER DEVELOPERS' TEAM collaboration, H. Abdolmaleki et al., *Impact of low- $x$  resummation on QCD analysis of HERA data*, *Eur. Phys. J. C* **78** (2018) 621 [[1802.00064](#)].
- [49] J. Pumplin, D. R. Stump and W. K. Tung, *Multivariate fitting and the error matrix in global analysis of data*, *Phys. Rev. D* **65** (2001) 014011 [[hep-ph/0008191](#)].
- [50] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput. Phys. Commun.* **10** (1975) 343.
- [51] NNPDF collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [[1410.8849](#)].
- [52] G. P. Salam, *A Resummation of large subleading corrections at small  $x$* , *JHEP* **07** (1998) 019 [[hep-ph/9806482](#)].
- [53] M. Ciafaloni, D. Colferai and G. P. Salam, *Renormalization group improved small  $x$  equation*, *Phys. Rev. D* **60** (1999) 114036 [[hep-ph/9905566](#)].
- [54] M. Ciafaloni, D. Colferai, G. P. Salam and A. M. Stasto, *The Gluon splitting function at moderately small  $x$* , *Phys. Lett. B* **587** (2004) 87 [[hep-ph/0311325](#)].
- [55] M. Ciafaloni, D. Colferai, G. P. Salam and A. M. Stasto, *Renormalization group improved small  $x$  Green's function*, *Phys. Rev. D* **68** (2003) 114003 [[hep-ph/0307188](#)].
- [56] M. Ciafaloni, D. Colferai, G. P. Salam and A. M. Stasto, *A Matrix formulation for small- $x$  singlet evolution*, *JHEP* **08** (2007) 046 [[0707.1453](#)].
- [57] R. D. Ball and S. Forte, *Summation of leading logarithms at small  $x$* , *Phys. Lett. B* **351** (1995) 313 [[hep-ph/9501231](#)].
- [58] R. D. Ball and S. Forte, *Asymptotically free partons at high-energy*, *Phys. Lett. B* **405** (1997) 317 [[hep-ph/9703417](#)].
- [59] G. Altarelli, R. D. Ball and S. Forte, *Factorization and resummation of small  $x$  scaling violations with running coupling*, *Nucl. Phys. B* **621** (2002) 359 [[hep-ph/0109178](#)].
- [60] G. Altarelli, R. D. Ball and S. Forte, *An Anomalous dimension for small  $x$  evolution*, *Nucl. Phys. B* **674** (2003) 459 [[hep-ph/0306156](#)].
- [61] G. Altarelli, R. D. Ball and S. Forte, *Perturbatively stable resummed small  $x$  evolution kernels*, *Nucl. Phys. B* **742** (2006) 1 [[hep-ph/0512237](#)].
- [62] G. Altarelli, R. D. Ball and S. Forte, *Small  $x$  Resummation with Quarks: Deep-Inelastic Scattering*, *Nucl. Phys. B* **799** (2008) 199 [[0802.0032](#)].
- [63] R. S. Thorne, *Explicit calculation of the running coupling BFKL anomalous dimension*, *Phys. Lett. B* **474** (2000) 372 [[hep-ph/9912284](#)].
- [64] R. S. Thorne, *NLO BFKL equation, running coupling and renormalization scales*, *Phys. Rev. D* **60** (1999) 054031 [[hep-ph/9901331](#)].
- [65] R. S. Thorne, *The Running coupling BFKL anomalous dimensions and splitting functions*, *Phys. Rev. D* **64** (2001) 074005 [[hep-ph/0103210](#)].

- [66] C. D. White and R. S. Thorne, *A Global Fit to Scattering Data with NLL BFKL Resummations*, *Phys. Rev. D* **75** (2007) 034005 [[hep-ph/0611204](#)].
- [67] I. Z. Rothstein and I. W. Stewart, *An Effective Field Theory for Forward Scattering and Factorization Violation*, *JHEP* **08** (2016) 025 [[1601.04695](#)].
- [68] S. Catani, M. Ciafaloni and F. Hautmann, *GLUON CONTRIBUTIONS TO SMALL  $x$  HEAVY FLAVOR PRODUCTION*, *Phys. Lett. B* **242** (1990) 97.
- [69] S. Catani, M. Ciafaloni and F. Hautmann, *High-energy factorization and small  $x$  heavy flavor production*, *Nucl. Phys. B* **366** (1991) 135.
- [70] S. Catani and F. Hautmann, *High-energy factorization and small  $x$  deep inelastic scattering beyond leading order*, *Nucl. Phys. B* **427** (1994) 475 [[hep-ph/9405388](#)].
- [71] M. Bonvini and F. Giuli, *A new simple PDF parametrization: improved description of the HERA data*, *Eur. Phys. J. Plus* **134** (2019) 531 [[1902.11125](#)].
- [72] M. Bonvini, S. Marzani and T. Peraro, *Small- $x$  resummation from HELL*, *Eur. Phys. J. C* **76** (2016) 597 [[1607.02153](#)].
- [73] M. Bonvini, S. Marzani and C. Muselli, *Towards parton distribution functions with small- $x$  resummation: HELL 2.0*, *JHEP* **12** (2017) 117 [[1708.07510](#)].
- [74] M. Bonvini and S. Marzani, *Four-loop splitting functions at small  $x$* , *JHEP* **06** (2018) 145 [[1805.06460](#)].
- [75] M. Bonvini, *Small- $x$  phenomenology at the LHC and beyond: HELL 3.0 and the case of the Higgs cross section*, *Eur. Phys. J. C* **78** (2018) 834 [[1805.08785](#)].
- [76] J. McGowan, T. Cridge, L. A. Harland-Lang and R. S. Thorne, *Approximate  $N^3$ LO parton distribution functions with theoretical uncertainties: MSHT20a $N^3$ LO PDFs*, *Eur. Phys. J. C* **83** (2023) 185 [[2207.04739](#)].
- [77] BCDMS collaboration, A. C. Benvenuti et al., *A High Statistics Measurement of the Proton Structure Functions  $F_2(x, Q^{*2})$  and  $R$  from Deep Inelastic Muon Scattering at High  $Q^{*2}$* , *Phys. Lett. B* **223** (1989) 485.
- [78] NEW MUON collaboration, M. Arneodo et al., *Accurate measurement of  $F_2(d) / F_2(p)$  and  $R^{*d} - R^{*p}$* , *Nucl. Phys. B* **487** (1997) 3 [[hep-ex/9611022](#)].
- [79] CHORUS collaboration, G. Onengut et al., *Measurement of nucleon structure functions in neutrino scattering*, *Phys. Lett. B* **632** (2006) 65.
- [80] NEW MUON collaboration, P. Amaudruz et al., *The Gottfried sum from the ratio  $F_2(n) / F_2(p)$* , *Phys. Rev. Lett.* **66** (1991) 2712.
- [81] K. Gottfried, *Sum rule for high-energy electron - proton scattering*, *Phys. Rev. Lett.* **18** (1967) 1174.
- [82] A. Niegawa and K. Sasaki, *Adler Sum Rule and Quark Parton Distribution Functions in Nucleon*, *Prog. Theor. Phys.* **54** (1975) 192.
- [83] R. D. Field and R. P. Feynman, *Quark Elastic Scattering as a Source of High Transverse Momentum Mesons*, *Phys. Rev. D* **15** (1977) 2590.
- [84] SEAQUEST collaboration, J. Dove et al., *The asymmetry of antimatter in the proton*, *Nature* **590** (2021) 561 [[2103.04024](#)].
- [85] FNAL E906, SEAQUEST collaboration, J. Dove et al., *Measurement of flavor asymmetry of the light-quark sea in the proton with Drell-Yan dimuon production in  $p+p$  and  $p+d$  collisions at 120 GeV*, *Phys. Rev. C* **108** (2023) 035202 [[2212.12160](#)].
- [86] NUSEA collaboration, R. S. Towell et al., *Improved measurement of the anti- $d$  / anti- $u$  asymmetry in the nucleon sea*, *Phys. Rev. D* **64** (2001) 052002 [[hep-ex/0103030](#)].

- [87] CDF collaboration, T. A. Aaltonen et al., *Measurement of  $d\sigma/dy$  of Drell-Yan  $e^+e^-$  pairs in the Z Mass Region from  $p\bar{p}$  Collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Lett. B* **692** (2010) 232 [[0908.3914](#)].
- [88] D0 collaboration, V. M. Abazov et al., *Measurement of the Shape of the Boson Rapidity Distribution for  $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$  Events Produced at  $\sqrt{s}$  of 1.96-TeV*, *Phys. Rev. D* **76** (2007) 012003 [[hep-ex/0702025](#)].
- [89] NNPDF collaboration, R. D. Ball, S. Forte, A. Guffanti, E. R. Nocera, G. Ridolfi and J. Rojo, *Unbiased determination of polarized parton distributions and their uncertainties*, *Nucl. Phys. B* **874** (2013) 36 [[1303.7236](#)].
- [90] NNPDF collaboration, E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi and J. Rojo, *A first unbiased global determination of polarized PDFs and their uncertainties*, *Nucl. Phys. B* **887** (2014) 276 [[1406.5539](#)].
- [91] D. de Florian, R. Sassot, M. Stratmann and W. Vogelsang, *Extraction of Spin-Dependent Parton Densities and Their Uncertainties*, *Phys. Rev. D* **80** (2009) 034030 [[0904.3821](#)].
- [92] J. J. Ethier, N. Sato and W. Melnitchouk, *First simultaneous extraction of spin-dependent parton distributions and fragmentation functions from a global QCD analysis*, *Phys. Rev. Lett.* **119** (2017) 132001 [[1705.05889](#)].
- [93] D. De Florian, G. A. Lucero, R. Sassot, M. Stratmann and W. Vogelsang, *Monte Carlo sampling variant of the DSSV14 set of helicity parton densities*, *Phys. Rev. D* **100** (2019) 114027 [[1902.10548](#)].
- [94] JEFFERSON LAB ANGULAR MOMENTUM (JAM) collaboration, Y. Zhou, N. Sato and W. Melnitchouk, *How well do we know the gluon polarization in the proton?*, *Phys. Rev. D* **105** (2022) 074022 [[2201.02075](#)].
- [95] JEFFERSON LAB ANGULAR MOMENTUM (JAM) collaboration, C. Cocuzza, W. Melnitchouk, A. Metz and N. Sato, *Polarized antimatter in the proton from a global QCD analysis*, *Phys. Rev. D* **106** (2022) L031502 [[2202.03372](#)].
- [96] J. Karpie, R. M. Whitehill, W. Melnitchouk, C. Monahan, K. Orginos, J. W. Qiu et al., *Gluon helicity from global analysis of experimental data and lattice QCD Ioffe time distributions*, [[2310.18179](#)].
- [97] F. Hekhorn, G. Magni, E. R. Nocera, T. R. Rabemananjara, J. Rojo, A. Schaus et al., *Heavy Quarks in Polarised Deep-Inelastic Scattering at the Electron-Ion Collider*, [[2401.10127](#)].
- [98] D. de Florian, S. Forte and W. Vogelsang, *Higgs production at RHIC and the positivity of the gluon helicity distribution*, [[2401.10814](#)].
- [99] PARTICLE DATA GROUP collaboration, R. L. Workman et al., *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.
- [100] V. Bertone, *APFEL++: A new PDF evolution library in C++*, *PoS DIS2017* (2018) 201 [[1708.00911](#)].
- [101] C. Bourrely, J. Soffer and F. Buccella, *Strangeness asymmetry of the nucleon in the statistical parton model*, *Phys. Lett. B* **648** (2007) 39 [[hep-ph/0702221](#)].