

Improved description of the HERA data with a new simple PDF parametrization

Francesco Giuli*

*University of Rome Tor Vergata and INFN, Sezione di Roma 2,
Via della Ricerca Scientifica 1, 00133 Roma, Italy*
E-mail: francesco.giuli@roma2.infn.it

Marco Bonvini

*INFN, Sezione di Roma 1,
Piazzale Aldo Moro 5, 00185 Roma, Italy*
E-mail: marco.bonvini@roma1.infn.it

A new parametrization for the parton distribution functions with a higher flexibility in the small- x region is presented. It has been implemented in the `xFitter` open-source PDF fitting tool, and compared to the default `xFitter` parametrization, used for the determination of the HERAPDF set. It has been found that the combined inclusive HERA I+II data can be described using NNLO theory with a significantly higher quality than HERAPDF2.0: the χ^2 is reduced by more than 60 units, having used only four more parameters. Our result highlights a significant parametrization bias in the default `xFitter` parametrization at small x , which would lead to even more dramatic effects when used for higher energy colliders, where the small- x region is more relevant. We also find that the inclusion of small- x resummation leads to a further reduction by approximately 30 extra units in χ^2 . In this contribution, we review the results of the recent paper "A new simple PDF parametrization: improved description of the HERA data" (arXiv:1902.11125).

*XXVII International Workshop on Deep-Inelastic Scattering and Related Subjects - DIS2019
8-12 April, 2019
Torino, Italy*

*Speaker.

1. The new proposed parametrization

Among many others, parton distribution functions (PDFs) represent a fundamental aspect of perturbative QCD (pQCD) in presence of incoming protons. These objects describe the longitudinal momentum fraction x of partons within the proton. Currently the most accurate and reliable way to determine PDFs is through fits to data; thus the resulting fitted distributions depend on various aspects of this procedure e.g. the perturbative order of partonic cross sections or DGLAP splitting functions, the way heavy quarks are treated, χ^2 definition, minimization methods, the choice of the PDF parametrization at the initial scale Q_0^2 , etc. Here we review the results of Ref. [1], where a new flexible and simple parametrization has been proposed and successfully used to determine PDFs.

Starting from the default parametrization used in `xFitter` [2, 3] (which is the one used for the HERAPDF set), namely

$$xf(x, \mu_0^2) = Ax^B(1-x)^C \left[1 + Dx + Ex^2 \right] - A' x^{B'} (1-x)^{C'}, \quad (1.1)$$

we propose a new extension, designed to add more flexibility in the small- x region, while keeping the number of fitted parameters small. A polynomial in $\log x$, which gives flexibility in the low- x region, is added on top of the polynomial in x , which gives flexibility in the high- x region. These two polynomials can be combined considering a multiplicative option

$$xf(x, \mu_0^2) = Ax^B(1-x)^C \left[1 + Dx + Ex^2 \right] \left[1 + F \log x + G \log^2 x + H \log^3 x \right] \quad (1.2)$$

or an additive option

$$xf(x, \mu_0^2) = Ax^B(1-x)^C \left[1 + Dx + Ex^2 + F \log x + G \log^2 x + H \log^3 x \right]. \quad (1.3)$$

These two options have been tested and it has been found that the additive parametrization results in smoother shapes and smaller χ^2 in the fit. So, the actual parametrization used in our fits to the inclusive HERA data is

$$xg(x, \mu_0^2) = A_g x^{B_g} (1-x)^{C_g} \left[1 + F_g \log x + G_g \log^2 x \right] \quad (1.4a)$$

$$xu_v(x, \mu_0^2) = A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} \left[1 + E_{u_v} x^2 + F_{u_v} \log x + G_{u_v} \log^2 x \right] \quad (1.4b)$$

$$xd_v(x, \mu_0^2) = A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}} \quad (1.4c)$$

$$x\bar{u}(x, \mu_0^2) = A_{\bar{u}} x^{B_{\bar{u}}} (1-x)^{C_{\bar{u}}} \left[1 + D_{\bar{u}} x + F_{\bar{u}} \log x \right] \quad (1.4d)$$

$$x\bar{d}(x, \mu_0^2) = A_{\bar{d}} x^{B_{\bar{d}}} (1-x)^{C_{\bar{d}}} \left[1 + D_{\bar{d}} x + F_{\bar{d}} \log x \right]. \quad (1.4e)$$

This new parametrization depends on 18 free parameters to be fitted at the starting scale, which is to be compared with the HERAPDF2.0 parametrization, which depends on 14 free parameters.

2. PDF determination at NNLO

In order to directly compare our fit results with HERAPDF2.0, we use the same definition of the χ^2 , namely [4]

$$\chi^2 = \sum_i \frac{[D_i - T_i (1 - \sum_j \gamma_j b_j)]^2}{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i} + \sum_j b_j^2 + \sum_i \log \frac{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i}{\delta_{i,\text{uncor}}^2 D_i^2 + \delta_{i,\text{stat}}^2 D_i^2}, \quad (2.1)$$

Differences in the fit setup	Old setup, same as [4]	New setup, same as [10]
heavy flavour scheme	TR	FONLL
initial scale μ_0	1.38 GeV	1.6 GeV
charm matching scale μ_c	m_c	$1.12m_c$
charm mass m_c	1.43 GeV	1.46 GeV

Table 1: Summary of the differences in the theoretical setup between the fit of HERAPDF2.0 [4] and the new fits presented in this and in the following sections (which is the same of Ref. [10]).

Contribution to χ^2	Old parametrization [10]	New parametrization
subset NC e^+ 920 $\tilde{\chi}^2$ /n.d.p.	451/377	406/377
subset NC e^+ 820 $\tilde{\chi}^2$ /n.d.p.	68/70	74/70
subset NC e^+ 575 $\tilde{\chi}^2$ /n.d.p.	220/254	222/254
subset NC e^+ 460 $\tilde{\chi}^2$ /n.d.p.	218/204	225/204
subset NC e^- $\tilde{\chi}^2$ /n.d.p.	215/159	217/159
subset CC e^+ $\tilde{\chi}^2$ /n.d.p.	44/39	37/39
subset CC e^- $\tilde{\chi}^2$ /n.d.p.	57/42	50/42
correlation term + log term	100 + 15	79 + 2
Total χ^2 /d.o.f.	1388/1131	1312/1127

Table 2: Total χ^2 per degrees of freedom (d.o.f.) and the partial $\tilde{\chi}^2$ per number of data points (n.d.p.) of each subset of the inclusive HERA dataset, for HERAPDF2.0 and our fit obtained with the parametrization Eq. (1.4). The second and third terms of Eq. (2.1), denoted correlation and log terms respectively, are also shown. The FONLL scheme is used, having raised $\mu_c/m_c = 1.12$, $\mu_0 = 1.6$ GeV and $m_c = 1.46$ GeV, namely the setting of Ref. [10].

where the measured data are represented by D_i , the corresponding theoretical prediction by T_i , $\delta_{i,\text{uncor}}$ and $\delta_{i,\text{stat}}$ represent the uncorrelated systematic and the statistical uncertainties on the measured data respectively, γ_j describe the correlated systematics which are accounted for using the nuisance parameters b_j . The sums over i extend over all data points, while the sum over j runs over the various sources of correlated systematics.

Instead of the ‘‘optimized’’ version [5] of the Thorne-Roberts scheme [6, 7] used in HERAPDF2.0 [4], the FONLL scheme [8] is used as heavy quark mass scheme. The differences between the two fit setups are summarized in Tab. 1. The results of the fit in terms of χ^2 , switching from the old parametrization to the new one in FONLL scheme, are presented in Tab. 2. A significant reduction of 76 units of the total χ^2 is observed, which is much larger than the increase of 4 units in the number of parameters.

Moving to the PDF comparison, the gluon, u_v and d_v distributions at the scale $Q^2 = 3$ GeV² are shown in Fig. 1, where our fit results are plotted along with the HERAPDF2.0. It can be noticed that the shape is generally smoother for HERAPDF2.0, while a richer structure in the medium- and small- x region is present in our PDFs. The comparison between our PDFs and a NNPDF3.0 set obtained fitting only HERA data [9] is also shown in Fig. 1. This choice has been made because NNPDF has the most flexible parametrization available on the market. It is noticeable that our PDFs lie inside the NNPDF uncertainty bands in most regions of x , while the HERAPDF2.0 PDFs lie outside in many more cases. Furthermore, we observe that the gluon shape predicted by NNPDF

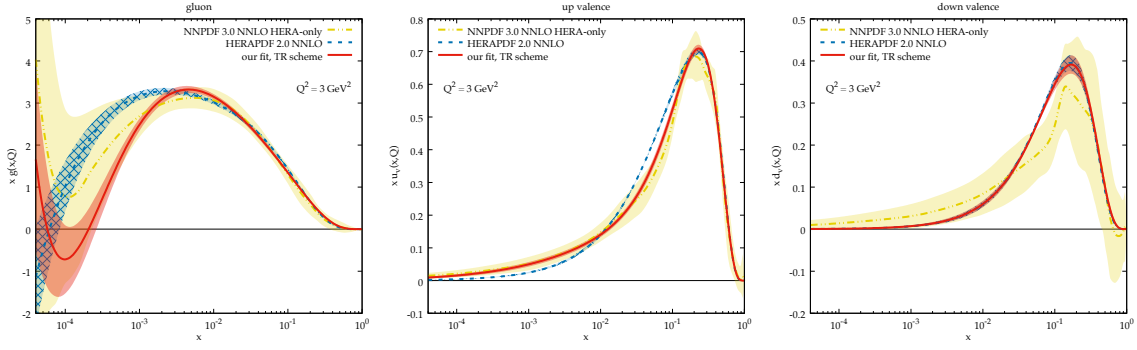


Figure 1: Comparison of our fit (solid red) with HERAPDF2.0 (dashed blue) and NNPDF3.0 HERA-only (dot-dot-dashed yellow) for the gluon, u_v and d_v PDFs. The uncertainty shown is only the “experimental” one, namely the one coming from the uncertainty on the parameters determined from the fit. For NNPDF, this uncertainty actually covers other kinds of uncertainties, such as those coming from parametrization bias.

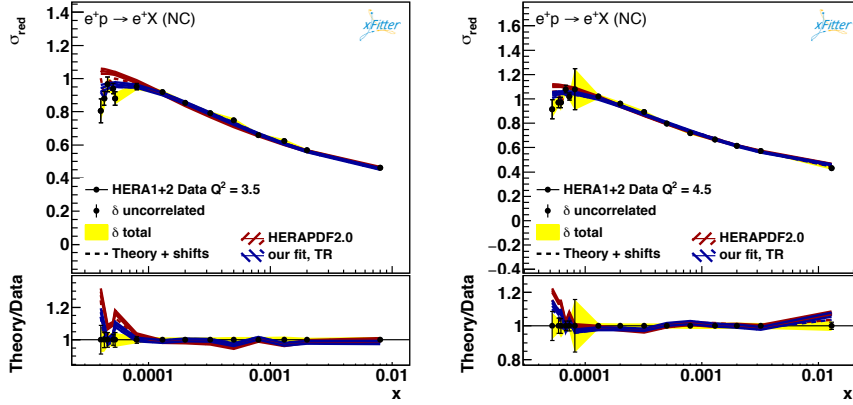


Figure 2: An example of comparison of the theoretical predictions from our and HERAPDF2.0 with data: low- Q^2 neutral-current reduced cross section from the neutral-current $E_p = 920$ GeV dataset.

is very similar to ours and instead quite different from HERAPDF2.0.

The comparison of HERA data with theoretical predictions using both our fit and HERAPDF2.0 has been inspected in detail. It can be seen in Tab. 2 that the agreement is at the same level, apart from the low- x and low- Q^2 data (contained in the first dataset of the list), where a clear improvement is manifest. Fig. 2 shows the two lowest Q^2 bins included in the fit, namely at $Q^2 = 3.5$ GeV² and $Q^2 = 4.5$ GeV², which contain the data at lowest x ; here, the flexibility of our parametrization in the small- x regime allows to better describe this region.

3. Reducing correlations for a stabler fit

A strong correlation between the parameters of the fit may lead to instabilities. In our parametrization, the parameters governing the small- x region (B, F, G, H) turned out to have significant correlations. In order to reduce such correlation, it is useful to redefine the parameters such that they each multiply a function whose contribution is dominant in a restricted region. Bernstein polynomials provide an easy way to achieve this goal. For instance, a generic polynomial of degree n in x

in the range $0 < x < 1$ can be conveniently expressed as a linear combination of the Bernstein basis polynomials

$$\binom{n}{k} (1-x)^k x^{n-k}, \quad k = 0, \dots, n, \quad (3.1)$$

each of which is peaked in a different region of x . The variable x can also be replaced by a function of x which still ranges from 0 to 1. For instance in Ref. [11] this basis was used in CT fits, but replacing x with \sqrt{x} . In our case, the most obvious choice is to use $\log \frac{1}{x}$, which however ranges from 0 to infinity. To circumvent this difficulty, we simply consider a limited range in x in which we reparametrize our polynomial in $\log \frac{1}{x}$ in terms of Bernstein polynomials. Since the data only extend to a small but finite value of x , we consider the range $x_0 < x < 1$, with $x_0 \sim 10^{-4}$. Therefore, we can use as a basis for our parametrization the polynomials

$$\binom{n}{k} (1-y(x))^k y(x)^{n-k}, \quad k = 0, \dots, n, \quad \text{with } y(x) = \frac{\log \frac{1}{x}}{\log \frac{1}{x_0}}. \quad (3.2)$$

In our specific case, we actually mix a polynomial in x and in $\log \frac{1}{x}$, Eq. (1.3). These two variables, or better x and $y(x)$, tend to zero in opposite limits, and therefore describe opposite regions. The best option to separate off the two regions described by these two polynomials is to consider two different Bernstein polynomials, one in x and one in $y(x)$, suppressing each with the $k = n$ contribution of the other. However, this option does no longer correspond to the polynomial we used in our fits, due to the presence of contributions $x^a \log^b \frac{1}{x}$ with both $a > 0$ and $b > 0$, which are absent in Eq. (1.3). Therefore, we propose a simpler choice, in which the x polynomial is treated as a ‘‘correction’’ to the $\log \frac{1}{x}$ polynomial. Our most generic parametrization Eq. (1.3) then becomes

$$\begin{aligned} xf(x, \mu_0^2) &= Ax^B (1-x)^C \left[1 + F \log x + G \log^2 x + H \log^3 x + Dx + Ex^2 \right] \\ &= Ax^B (1-x)^C \left[(1-y)^3 + 3F'y(1-y)^2 + 3G'y^2(1-y) + H'y^3 + 2D'x(1-x) + E'x^2 \right], \end{aligned} \quad (3.3)$$

where the new ‘‘primed’’ parameters should be much less correlated among each other, thereby leading to a stabler minimization procedure. Simpler versions with less coefficients can be constructed in similar ways. For instance, when the $\log^3 x$ term is not used, as in our default parametrization Eq. (1.4), we simply have

$$\begin{aligned} xf(x, \mu_0^2) &= Ax^B (1-x)^C \left[1 + F \log x + G \log^2 x + Dx + Ex^2 \right] \\ &= Ax^B (1-x)^C \left[(1-y)^2 + 2F''y(1-y) + G''y^2 + 2D'x(1-x) + E'x^2 \right]. \end{aligned} \quad (3.4)$$

Similarly, one can switch off either the D or the E term, which leaves the other unmodified. We plan to test this new form of the parametrization once the new version of `xFitter` will be released.

References

- [1] Bonvini, M. and Giuli, F., *A new simple PDF parametrization: improved description of the HERA data*, [1902.11125]
- [2] Alekhin, S. et al., *HERAFitter*, *Eur. Phys. J. C* **75** (2015) 7 [1410.4412]

- [3] Bertone, V. et al., *xFitter 2.0.0: An Open Source QCD Fit Framework*, *PoS DIS2017* (2018) 203 [1709.01151]
- [4] ZEUS, H1 collaboration, *Combination of measurements of inclusive deep inelastic $e\text{-}p$ scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J. C* **75** (2015) 580 [1506.06042]
- [5] Thorne, R.S., *The Effect of Changes of Variable Flavour Number Scheme on PDFs and Predicted Cross Sections*, *Phys. Rev.* **D86** (2012) 074017 [1201.6180]
- [6] Thorne, R. S. and Roberts, R. G., *An Ordered analysis of heavy flavor production in deep inelastic scattering*, *Phys. Rev.* **D57** (1998) 6871-6898 [hep-ph/9709442]
- [7] Thorne, R.S., *A Variable-flavor number scheme for NNLO*, *Phys. Rev.* **D73** (2006) 054019 [hep-ph/0601245]
- [8] Forte, S. et al., *Heavy quarks in deep-inelastic scattering*, *Nucl. Phys.* **B834** (2010) 116-162 [1001.2312]
- [9] Ball, R. D. et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [1410.8849]
- [10] Abdolmaleki, H. et al., *Impact of low- x resummation on QCD analysis of HERA data*, *Eur. Phys. J. C* **78** (2018) 621 [1802.00064]
- [11] Dulat, S. et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.* **D93** (2016) 033006 [1506.07443]