HERAFitter - an open source QCD fit framework

Pavel Starovitov
on behalf of the `HERAFitter` team

DESY (Deutsches Elektronen-Synchrotron)
Notkestrasse, 85, 22607 Hamburg, Germany

The patron distribution functions (PDFs) describe the parton content of the proton. They cannot be calculated from the first principle and are to be determined empirically by fitting experimental observables to quantum chromodynamics (QCD) predictions. The `HERAFitter` project aims to provide a framework for QCD analyses of proton structure at leading order (LO), next-to-leading order (NLO) and next-to-next-to-leading (NNLO) orders in perturbative QCD. The framework includes various modules and interfaces enabling a large number of theoretical and methodological options. It also allows to study the impact of the new experimental data on the PDFs from *ep*, *pp* and $p\bar{p}$ scattering processes. The fast development of the project involves active communication with theorists and experimentalist as well as a close collaboration with the main PDF fitting groups. Full information about the project, downloads and documentation can be found in `http://herafitter.org`.

1. `HERAFitter` project

The hard-scattering cross section is a convolution of the parton distribution functions, coupling constant and the perturbatively calculable matrix elements of a hard scattering process. PDFs, which cannot be derived from the theory and are to be determined by a fit of predictions to experimental observables using the DGLAP evolution equation [1]. `HERAFitter` project is an open source QCD fit framework designed for the extraction of parton distribution functions of the proton. The framework includes a large number of theoretical models and allows to study the impact of the new experimental data on the PDFs. A schematic structure of the `HERAFitter` framework is illustrated in Figure 1 and explained below. Further information about the project with available releases and documentation (manual and release notes) can be found in `http://herafitter.org`.

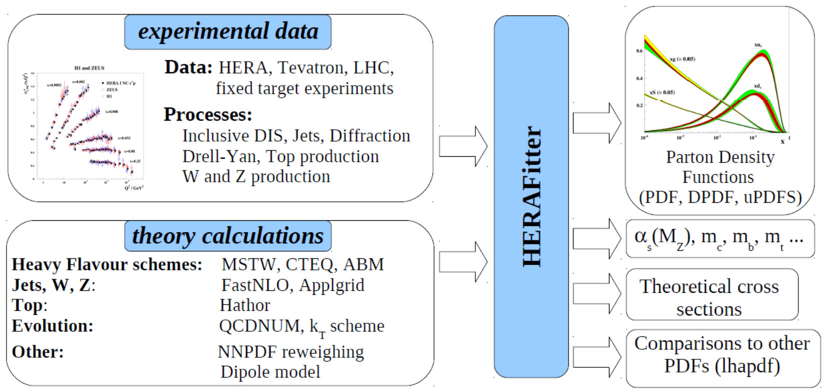

Fig. 1. Schematic structure of the `HERAFitter` program: The "experimental data" block illustrates variety of processes which can be included into the framework and are matched to the corresponding theoretical calculation illustrated in the block below; after the QCD fit is performed the resulting PDFs together with theoretical cross sections are provided as shown on the right side of the figure.

2. Functionality

The procedure to determine PDFs in the `HERAFitter` framework is the following. First, the PDFs are parametrized at the starting scale Q_0^2 chosen to be below the charm mass threshold. Then, the fits to the experimental data are performed using DGLAP evolution in the $\overline{MS}$ scheme as implemented in the QCDNUM [2] package. The evolution is available at leading (LO), next-to-leading (NLO) or next-to-next-to-leading (NNLO) orders.

During the fitting procedure, the measured and predicted cross sections together with their corresponding errors are used to build a χ^2 and determine the parameters of the PDF model. The main functionality of the code which includes the treatment of the experimental data, calculation of the theory predictions, χ^2 minimization, parametrization and optional tools are summarized below.

All the main experimental data sets relevant for PDFs can be fitted within the `HERAFitter` framework: inclusive deep inelastic scattering (DIS) cross sections from HERA and fixed target experiments, Drell-Yan cross sections, inclusive jet production data (ep, pp and $p\bar{p}$) and heavy quark structure functions. The DIS structure functions may be computed in a variety of heavy quark schemes including the fixed-flavor (FFN) and variable flavor number (VFN) schemes. VFN schemes with various treatments for the heavy quark thresholds include the Thorne Roberts (TR) scheme at LO, NLO and NNLO [3, 4] as provided by the MSTW group, the ACOT scheme at LO and NLO as provided by the CTEQ group. The QCDNUM also provides the calculations of the DIS structure functions in the zero-mass FFN schemes. The FFN scheme is alternatively available via the OPENQCDRAD [7] interface in which the running mass definition [8] is implemented.

The calculations of production cross sections at hadron colliders with NLO accuracy require a huge amount CPU time in order to reach a high statistical precision of the result. To facilitate the use of NLO calculations in the fitting framework the grid methods were adopted by the `HERAFitter` team. These techniques rely on the factorization theorem. It decouples the hard scattering coefficients from PDFs and stores them on the grid allowing a fast recalculation of the cross section in every fit iteration. APPLGRID [9] (or FastNLO [10]) technique is used as a fast interface to jet [6] and electroweak [5] cross sections calculators. An independent treatment for the electro-weak corrections is applied as k-factors, using external packages such as SANC or FEWZ to calculate them.

The predictions of Drell Yan production in $pp(\bar{p})$-collisions can be provided using two independent implementations. The first one uses calculations at the LO which and is extendable to NLO by applying a k-factors. Alternatively, one can obtain the full NLO predictions directly from the NLO calculations as implemented in MCFM [5] by interfacing it to APPLGRID library. Any of two methods can be chosen by a framework user.

The program HATHOR [11] interfaced to `HERAFitter` provides the calculation of the total $t\bar{t}$ cross section at $p\bar{p}$ and pp colliders up to approximate NNLO accuracy. Various definitions of χ^2 can be chosen in the `HERAFitter` for the minimization procedure which are based on the use of nuisance parameters or on the full covariance matrix. For a single data set, the χ^2 function is defined in a simple form

$$\chi^2 = \sum_i \frac{\left[m^i - \sum_j \gamma^i_j m^i b_j - \mu^i\right]^2}{\left(\delta_{i,\mathrm{stat}} m^i\right)^2 + \left(\delta_{i,\mathrm{uncor}}\, m^i\right)^2} + \sum_j b_j^2, \tag{1}$$

or can be evolved to the so called "scaled" form:

$$\chi^2 = \sum_i \frac{\left[m^i - \sum_j \gamma^i_j m^i b_j - \mu^i\right]^2}{\delta^2_{i,\mathrm{stat}} \mu^i \left(m^i - \sum_j \gamma^i_j m^i b_j\right) + \left(\delta_{i,\mathrm{uncor}}\, m^i\right)^2} + \sum_j b_j^2. \tag{2}$$

Here μ^i is the measured central value at a point i with relative statistical $\delta_{i,stat}$ and relative uncorrelated systematic $\delta_{i,unc}$ uncertainties, the quantity m^i is the theoretical prediction, γ^i_j represents relative correlated systematic uncertainties and b_j – their shifts. In the case of the covariance matrix, the

χ^2 function takes the form

$$\chi^2 = \sum_{ij}(m^i - \mu^i) C^{-1}_{\text{tot } ij}(m^j - \mu^j), \tag{3}$$

where the $C^{-1}_{\text{tot } ij}$ is the total covariance matrix given by the sum of the statistical and systematic covariance matrices. In addition, the different approaches can be combined together, e.g. some systematic uncertainties can be treated using the matrix method, others can be treated using nuisance parameters.

Several different parametrization forms for the PDFs at the staring scale can be chosen in the `HERAFitter`. First, the standard functional form can be used to parameterize PDFs, second, the bi-log normal functional form is available and third, forms based on generalized polynomials such as the Chebyshev can be used.

The best way to see the impact of the measurement on the PDFs is to perform a QCD fit. As an alternative to the PDF fit, a Bayesian reweighting technique can be used to study the impact of experimental data on PDFs. Bayesian reweighting can be used with Monte Carlo method as first employed by the NNPDF Collaboration [12] or using the Hessian Eigenvector Method as proposed in [13].

The `HERAFitter` provides PDFs in a standard LHAPDF format which can be used by theoretical calculation and Monte Carlo simulation programs. Basic plotting tools and several examples of the fits are included in the `HERAFitter` package.

3. Latest developments (HERAFitter-0.3.1)

The latest developments which are included in the newest `HERAFitter` release (HERAFitter-0.3.1 beta) are summarized below:

- improved treatment of the experimental data uncertainties which can be treated as asymmetric (using Toy MC method), as Offset method (the minimization is performed without taking into account correlated uncertainty sources but nuisance parameters are shifted by ± 1 in the uncertainty determination) and using the covariance matrix representation for data sets which have uncertainties reported in this form;

- the possibility to study the bias introduced by the parametrization form, a flexible PDF parametrization and MC method now can be employed. There are two regularization techniques, data driven and external regularization based on the χ^2 penalty term, which can be used to estimate parametrization induced biases;

- PDF Bayesian reweighting based on eigenvectors (Hessian method) as explained in the section above,
- unintegrated PDFs based on the k_T factorization (CCFM) evolution as an alternative approach to colinear DGLAP evolution;
- a Bartels-Golec-Kowalski(BGK) [14] dipole model which takes into account the effects of DGLAP evolution;
- inclusion of additional data sets from LHC and Tevatron and other updates like upgraded interface format to FastNLO and newer electroweak DIS program.

4. Results obtained using `HERAFitter`

The `HERAFitter` framework is actively used by HERA and LHC experiments. At HERA, the results of QCD analyses using `HERAFitter` are published in the recent combination of charm production measurements in DIS [15] and inclusive H1 measurements at high Q^2 with longitudinally polarized lepton beams [16]. The `HERAFitter` framework has been used in the QCD studies with Z and W cross sections measured by the ATLAS Collaboration to determine the strange quark density of the proton [17] as depicted in Figure 2. It was shown that the Z and W measurements introduce a novel sensitivity to the strange quark density at $x \sim 0.01$. The ratio of the strange to the down sea quark density is found to be $r_s = 1.00^{+0.25}_{-0.28}$ at Bjorken $x = 0.023$ and the initial scale of the QCD fit $Q_0^2 = 1.9$ GeV^2.

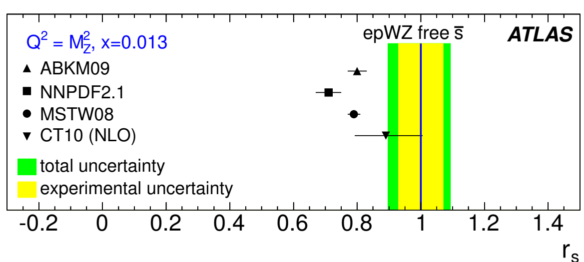

Fig. 2. Predictions obtained from NNLO pQCD analyses of the ratio $r_s = 0.5(s + \bar{s})/\bar{d}$, at $Q^2 = M_Z^2$ and $x = 0.013$. Points: Global fit analyzes using the PDF uncertainties as quoted in [17]; Bands: ATLAS measurement [17]; inner yellow band — experimental uncertainty; outer green band — total uncertainty.

Another QCD analysis using the `HERAFitter` framework was performed on the inclusive ATLAS jet data at $\sqrt{s} = 2.76$ TeV and $\sqrt{s} = 7$ TeV [18].

It was demonstrated that by including the ATLAS jet data, a harder gluon distribution and a softer sea-quark distribution in the high Bjorken-x region are obtained with respect to the fit of HERA data only. The results are presented in Figure 3.

At CMS, there are several analyses which use `HERAFitter` for PDF constraints, i.e. Drell-Yan, W+charm, W asymmetry data and inclusive jets. The combined analysis of HERA I data and CMS W asymmetry measurement [20] shows that CMS data change the central values of the light valence quark PDFs and decreases its uncertainties (see Figure 4).

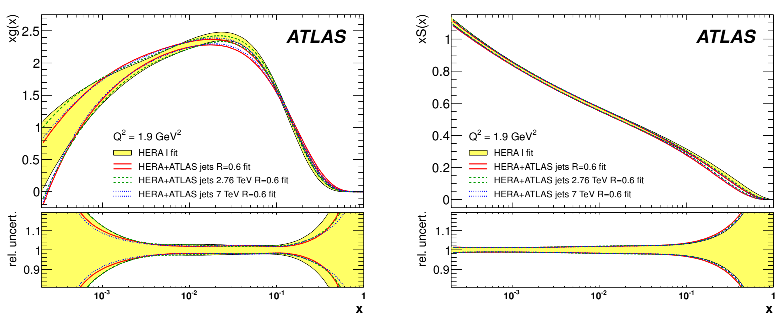

Fig. 3. Momentum distributions of the gluon, $xg(x)$, (left) and total sea, $xS(x)$,(right) together with their relative experimental uncertainty as a function of x for $Q^2 = 1.9$ GeV^2. The filled area indicates a fit to HERA data only. The bands show fits to HERA data in combination with both ATLAS jet data sets, and with the individual ATLAS jet data sets separately, each for jets of $R = 0.6$ size. For each fit the uncertainty in the PDF is centered on unity.

Adding the CMS jet data [19] to the HERA I data set (see Figure 5) results in a harder gluon distribution in high Bjorken-x region in agreement with ATLAS analysis [18].

From the theory side, the work is ongoing on updating the ACOT scheme module (in collaboration with CTEQ group members), inclusion of photon PDF in QCDNUM (on former a publication is foreseen). The QCD studies of LHeC data have been also performed with `HERAFitter` and published in [21].

5. Summary

`HERAFitter` contains all necessary ingredients to study the proton PDFs, it incorporates variety of different data processes and theory calculations, contains many useful tools and is an optimal platform for various benchmarking studies. `HERAFitter` is the first open source package to perform

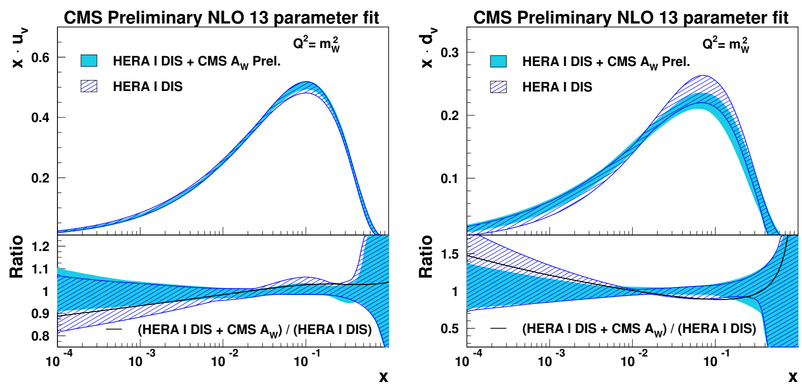

Fig. 4. Distributions of u-valence (left) and d-valence (right) quarks as a function of x at the scale. The results of the 13-parameter fixed strangeness fit to the HERA I DIS data and CMS muon asymmetry measurements (shaded band), and to HERA only (hashed band) are compared.

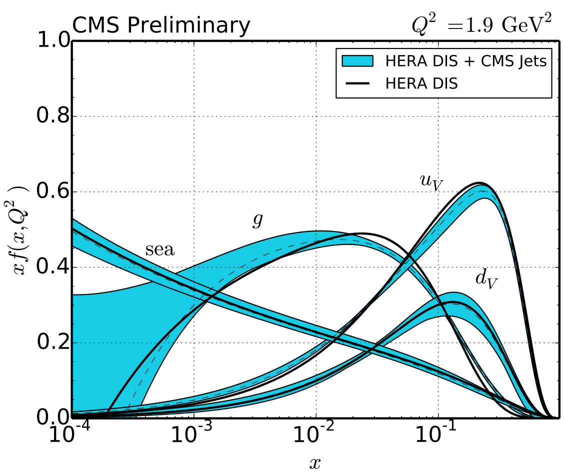

Fig. 5. Overview of the gluon, sea $\Sigma = 2(\bar{u} + \bar{d} + \bar{s})$, u valence, and d valence PDFs before (full line) and after (dashed line) including the CMS jet data into the fit. The PDFs are shown at the starting scale $Q^2 = 1.9$ GeV^2. In addition the total uncertainty including the CMS jet data is shown as a band around the central fit.

PDF fits and is actively used by experimental and theoretical high energy physics communities.

References

[1] V.N. Gribov and L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438, 675 (1972); G. Altarelli and G. Parisi, Nucl. Phys. B 126, 298 (1977); G. Curci, W. Furmanski, and R. Petronzio, Nucl. Phys. B **175**, 27 (1980); W. Furmanski and R. Petronzio, Nucl. Phys. B **97**, 473 (1980); S. Moch, J. Vermaseren, and A. Vogt, Nucl. Phys. B **688**, 101 (2004); A. Vogt, S. Moch, and J. Vermaseren, Nucl. Phys. B **691**, 129 (2004).

[2] M. Botje, http://www.nikef.nl/h24/qcdnum/index.html.

[3] R.S. Thorne and R.G. Roberts, Phys. Rev. D **57**, 6871 (1998).

[4] R.S. Thorne, Phys. Rev. D **73**, 054019 (2006).

[5] J. Campbell, K. Ellis, C. Williams, MCFM - Monte Carlo for FeMtobarn processes, a program description and the code are available via: http://mcfm.fnal.gov.

[6] Z. Nagy, NLOJet++, a program description and the code are available via: http://www.desy.de/∼znagy/Site/NLOJet++.html.

[7] S. Alekhin, OPENQCDRAD, a program description and the code are available via: http://www-zeuthen.desy.de/∼alekhin/OPENQCDRAD.

[8] S. Alekhin and S. Moch, Phys. Lett. B **699**, 345 (2011).

[9] T. Carli *et al.*, Eur. Phys. J C **66**, 503 (2010).

[10] D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch, "New features in version 2 of the fastNLO project," in the proceedings of the XX International Workshop on Deep Inelastic Scattering (DIS12), 26–30th March 2012, hep-ph/1208.3641; http://fastnlo.hepforge.org/.

[11] M. Aliev *et al.*, Comput. Phys. Commun. **182**, 1034 (2011).

[12] R.D. Ball *et al.*, Nucl. Phys. B **855**, 608 (2012);R.D. Ball *et al.* (NNPDF Collaboration), Nucl. Phys. B **849**, 112 (2011).

[13] G. Watt and R. Thorne, J. High Energy Phys. **08**, 052 (2012).

[14] J. Bartels, K. Golec-Biernat, and H. Kowalski, Phys. Rev. D **66**, 014001 (2002).

[15] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Eur. Phys. J C **73**, 2311 (2013).

[16] F. Aaron *et al.* (H1 Collaboration), J. High Energy Phys. **09**, 061 (2012).

[17] Georges Aad *et al.* (ATLAS Collaboration), Phys. Rev. Lett. **109**, 012001 (2012).

[18] Georges Aad *et al.* (ATLAS Collaboration), Eur. Phys. J C **73**, 2509 (2013).

[19] CMS Collaboration, Technical Report No. CMS-PAS-SMP-12-028, CERN, Geneva, 2013.

[20] CMS Collaboration, Technical Report No. CMS-PAS-SMP-12-021, CERN, Geneva, 2013.

[21] J. L. Abelleira Fernandez *et al.* (LHeC Study Group), J. Phys. G. **39**, 075001 (2012).