Eur. Phys. J. C manuscript No.
(will be inserted by the editor)

HERAFitter

Open Source QCD Fit Project

Version 0.8 (svn 1406)

HERAFitter developers team,
P. Belov[1,12], S. Camarda[1], A.M. Cooper-Sarkar[2], C. Diaconu[3], A. Gizhko[1], A. Glazov[1], V. Kolesnikov[4], A. Luszczak[5], V. Myronenko[1], H. Pirumov[1], R. Plačakytė[1], K. Rabbertz[6], V. Radescu[1], A. Sapronov[4], A. Schöning[10], S. Shushkevich[1], W. Slominski[7], P. Starovoitov[1], M. Sutton[8], J. Tomaszewska[9], O. Turkot[1], G. Watt[11], K. Wichmann[1],

[1]DESY, Hamburg, Germany
[2] Department of Physics, University of Oxford, Oxford, United Kingdom
[3] CPPM, IN2P3-CNRS, Univ. Mediterranee, Marseille, France
[4] Joint Institute for Nuclear Research (JINR),Joliot-Curie 6, 141980, Dubna, Moscow Region, Russia
[5] T. Kosciuszko Cracow University of Technology
[6] Institut für Experimentelle Kernphysik, Karlsruhe, Germany
[7] Jagiellonian University, Institute of Physics, Ul. Reymonta 4, PL-30-059 Cracow, Poland
[8] University of Sussex, Department of Physics and Astronomy, Sussex House, Brighton BN1 9RH, United Kingdom
[9] Warsaw University of Technology, Faculty of Physics, Koszykowa 75, 00-662 Warsaw, Poland
[10] Physikalisches Institut, Universität Heidelberg, Heidelberg, Germany
[11] Universität Zürich, Physik-Institut Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
[12] Current address: Department of Physics, St. Petersburg State University, Ulyanovskaya 1, 198504 St. Petersburg, Russia

Received: date / Accepted: date

Abstract The paper presents the `HERAFitter` project which provides a framework for Quantum Chromodynamics (QCD) analyses related to the proton structure. The main processes sensitive to the Parton Distribution Functions (PDFs) of the proton are Deep-Inelastic-Scattering in ep collisions at HERA and Drell Yan, jet and top quark production in pp ($p\bar{p}$) collisions at the LHC (Tevatron). Data of such recent measurements are included into `HERAFitter` and can be used for PDF determination based on the concept of the factorisable nature of the cross sections of hard scattering measurements into process dependent partonic scattering and universal PDFs. `HERAFitter` provides a comprehensive choice of options in the treatment of the experimental data uncertainties, a large number of theoretical and methodological options via interfaces to external software packages which are described here.

Keywords PDFs · QCD · Fit

Contents

1 Introduction

In the era of the Higgs discovery [1, 2] and extensive searches for signals of new physics at the LHC it is crucial to have accurate Standard Model (SM) predictions for hard scattering processes in hadron-hadron collisions. The most common approach to calculate the SM cross sections for such reactions is to use collinear factorisation in perturbative QCD (pQCD) [3]:

$$\sigma(\alpha_s,\mu_R,\mu_F) = \sum_{a,b}\int_0^1 dx_1\, dx_2\, f_a(x_1,\alpha_s,\mu_F) f_b(x_2,\alpha_s,\mu_F) \times \hat{\sigma}^{ab}(x_1,x_2;\alpha_s,\mu_R,\mu_F). \quad (1)$$

Here the cross section σ for any hard-scattering inclusive process $ab \to X + all$ is expressed as a convolution of Parton Distribution Functions (PDFs) f_a and f_b with the partonic cross section $\hat{\sigma}^{ab}$. The PDFs describe the probability of finding a specific parton a (b) in the first (second) proton carrying a fraction x_1 (x_2) of its momentum. The sum over indices a and b in Eq. 1 indicates the various kinds of partons, i.e. gluons, quarks and antiquarks of different flavours, that are considered as the constituents of the proton. Both the PDFs and the partonic cross section depend on the strong coupling α_s, and the factorisation and renormalisation scales, μ_F and μ_R, respectively. The partonic cross sections are calculable in pQCD, but the PDFs cannot yet be predicted in QCD they must rather be determined from measurement. PDFs are assumed to be universal such that different scattering reactions can be used to constrain them [4, 5].

Measurements of the inclusive Neutral Current (NC) and Charged Current (CC) Deep-Inelastic-Scattering (DIS) at the ep collider HERA provide crucial information for determining the PDFs. For instance, the gluon density relevant for calculating the dominant gluon-gluon fusion contribution to Higgs production at the LHC can be accurately determined at low and medium x from the HERA data alone. Many processes in pp and $p\bar{p}$ collisions at LHC and Tevatron, respectively, probe PDFs in the kinematic ranges, inaccessible by DIS measurements. Therefore inclusion of the LHC and Tevatron data in the QCD analysis of the proton structure provide additional constraints on the PDFs, improving either their precision, or providing important information of the correlations of PDF with the fundamental QCD parameters like strong coupling or quark masses. In this context, the processes of interest at hadron colliders are Drell Yan (DY) production, W and Z asymmetries, associated production of W or Z bosons and heavy quarks, top quark, jet and prompt photon production.

Open-source QCD platform `HERAFitter` encloses the set of tools necessary for a comprehensive global QCD analysis of hadron-induced processes even on the early stage of the experimental measurement. It has been developed for determination of PDFs and extraction of fundamental QCD parameters like heavy quark masses or the strong coupling constant. This tool provides also the basis for comparisons of different theoretical approaches and can be used for direct tests of the impact of new experimental data in the QCD analyses. The processes that are currently included in `HERAFitter` framework are listed in Tab. 1. The functionality of `HERAFitter` is schematically illustrated in Fig. 1 and can be represented by the four main blocks:

Data	Process	Reaction	Theory calculations, schemes
HERA	DIS NC	$ep \to eX$	TR', ACOT ZM (`QCDNUM`) FFN (`OPENQCDRAD`, `QCDNUM`)
	DIS CC	$ep \to \nu_e X$	ACOT, ZM (`QCDNUM`) FFN (`OPENQCDRAD`)
	DIS jets	$ep \to e$ jets	`NLOJet++` (`fastNLO`)
	DIS heavy quarks	$ep \to ec\bar{c}X$, $ep \to eb\bar{b}X$	ZM (`QCDNUM`), TR', ACOT, FFN (`OPENQCDRAD`, `QCDNUM`)
Fixed Target	DIS NC	$ep \to eX$	ZM (`QCDNUM`), TR', ACOT
Tevatron, LHC	Drell Yan	$pp(\bar{p}) \to l\bar{l}X$, $pp(\bar{p}) \to l\nu X$	`MCFM` (`APPLGRID`)
	top pair	$pp(\bar{p}) \to t\bar{t}X$	`MCFM` (`APPLGRID`), `HATHOR`
	single top	$pp(\bar{p}) \to tl\nu X$, $pp(\bar{p}) \to tX$, $pp(\bar{p}) \to tWX$	`MCFM` (`APPLGRID`)
	jets	$pp(\bar{p}) \to$ jets$X$	`NLOJet++` (`APPLGRID`), `NLOJet++` (`fastNLO`)
LHC	DY+heavy quarks	$pp \to VhX$	`MCFM` (`APPLGRID`)

Table 1 The list of processes available in the `HERAFitter` package. The refernces for the individual calculations and their implementations are given in the text.

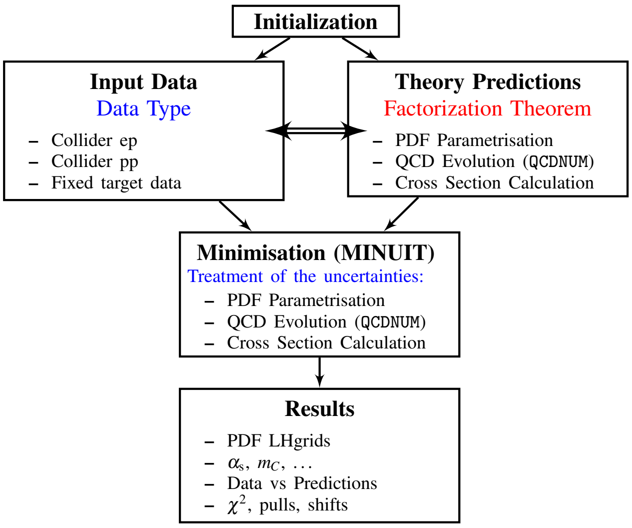

Fig. 1 Schematic structure of the `HERAFitter` program.

Input data: All relevant cross section measurements from the various reactions are stored internally in `HERAFitter` with the full information on their uncorrelated and correlated uncertainties. HERA I data sets are the basis of any proton PDF extraction, and they are used by all global PDF groups [6–10]. Additional measurements provide constraints to the sea flavour decomposition (such as the new results from the LHC), as well as constraints to PDFs in the kinematic phase-space regions not covered precisely by HERA I data, such as the high x region for the gluon and valence quark distributions.

Theory predictions: Predictions for cross section of different processes are obtained using the factorisation approach (Eq. 1). PDFs are parametrised at a starting scale Q_0^2 by a chosen functional form with a set of free parameters **p**. These PDFs are then evolved from Q_0^2 to the scale of the measurement using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [11–15] evolution equations as implemented in `QCDNUM` [16], and then convoluted with the hard parton cross sections calculated using a relevant theory program (as listed in Tab. 1).

PDF fit: PDFs are extracted from a least square fit by constructing a χ^2 from the input data and the theory prediction. The χ^2 is minimized iteratively with respect to the PDF parameters using the MINUIT [17] program. Various choices of accounting for the experimental uncertainties are employed in `HERAFitter`, either using a nuisance parameter method for the correlated systematic uncertainties, or a covariance matrix method (see details in section 4.2). In addition, `HERAFitter` allows to study different statistics assumptions for the distributions of the systematic uncertainties (i.e. Gauss or log-normal) [18]. In the χ^2 minimization, the fitted parameters **p** and their estimated uncertainties are produced.

Results: The resulting PDFs are provided in a format ready to be used by the `LHAPDF` library [19, 20]. `HERAFitter` drawing tools can be used to display the PDFs with the uncertainty at a chosen scale. A first set of PDFs extracted by `HERAFitter` is HERAPDF1.0 [21] (Fig. 2) which is based on HERA I data. Since then several other PDF sets were produced within the HERA and LHC collaborations. In addition to PDF display, the figures comparing the data used in the fit and the relevant theory predictions are produced. In Fig. 3, a comparison of inclusive NC data from the HERA I running period with predictions based on HERAPDF1.0. It also illustrates the comparison to the theory predictions which are adjusted by the systematic uncertainty shifts when using the nuisance parameter method that accounts for correlated systematic uncertainties. As an additional consistency check between data and the theory predictions, pull information, defined as the difference between data and prediction divided by the uncorrelated uncertaintly of the data, is displayed in units of sigma shifts for each given data bin.

The `HERAFitter` project provides a versatile environment for benchmarking studies and a flexible platform for the QCD interpretation of analyses within the LHC experiments, as already demonstrated by several publicly available results using the `HERAFitter` framework [22–28].

The outline of this paper is as follows. Section 2 discusses the various processes and corresponding theoretical calculations performed in the DGLAP [11–15] formalism that are available in `HERAFitter`. Section 3 presents various techniques employed by the theory calculations used in `HERAFitter`. Section 4 elucidates the methodology of determining PDFs through fits based on various χ^2 definitions used in the minimisation procedure. Alternative approaches to the DGLAP formalism are presented in section 5. Specific applications of the package are given in section 6.

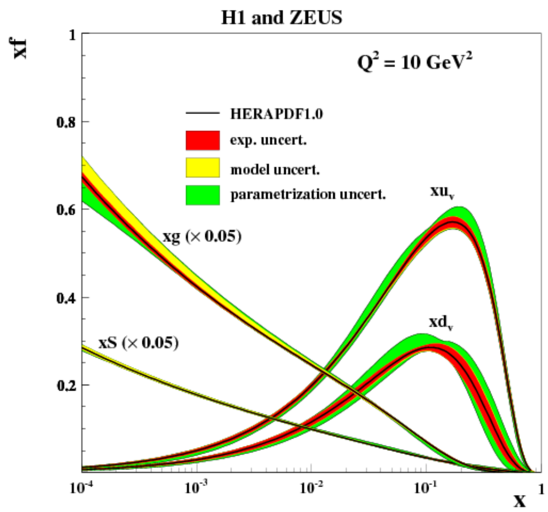

Fig. 2 Summary plots of valence, total sea (scaled) and gluon densities (scaled) with their experimental, model and parametrisation uncertainties at the scale of $Q^2 = 10$ GeV^2 for the HERAPDF1.0 PDF set at NLO [21].

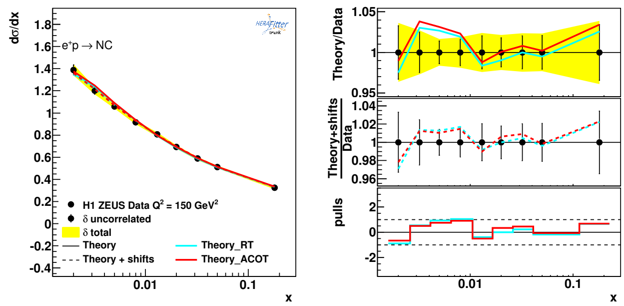

Fig. 3 An illustration of the `HERAFitter` drawing tools comparing the measurements (in this case HERA I) to the predictions of the fit. In addition, ratio plots are also provided together with the pull distribution (right panel).

2 Theoretical Input

In this section the theoretical formalism for various processes available in `HERAFitter` is described.

2.1 Deep Inelastic Scattering Formalism and Schemes

Deep Inelastic Scattering (DIS) data provide the backbone of any PDF fit. The formalism that relates the DIS measurements to pQCD and the PDFs has been described in detail in many extensive reviews (see e.g. [29]) and will only be briefly recapped here. DIS is a lepton scattering off the constituents of the proton by a virtual exchange of a NC or CC vector boson and, as a result, a scattered lepton and a multihadronic final state are produced. The DIS kinematic variables are the negative squared four-momentum of the exchange boson, Q^2, the Bjorken x, and the inelasticity y, where $y = Q^2/sx$ and s is the squared centre-of-mass energy.

The NC cross section can be expressed in terms of generalised structure functions:

$$\frac{d^2\sigma^{e^\pm p}_{NC}}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}\left[Y_+\tilde{F}_2^\pm \mp Y_- x\tilde{F}_3^\pm - y^2\tilde{F}_L^\pm\right], \tag{2}$$

where $Y_\pm = 1 \pm (1-y)^2$. The generalised structure functions $\tilde{F}_{2,3}$ can be written as linear combinations of the proton structure functions $F_2, F_{2,3}^{\gamma Z}$ and $F_{2,3}^{Z}$ associated to pure photon exchange terms, photon-Z interference terms and pure Z exchange terms respectively. Structure function $\tilde{F}_2$ is the dominant contribution to the cross section, $x\tilde{F}_3$ is important at high Q^2 and $\tilde{F}_L$ is sizable only at high y. In the framework of pQCD the structure functions are directly related to the PDFs, i.e. in leading order (LO) F_2 is the weighted momentum sum of quark and anti-quark distributions, $F_2 \approx x\sum e_q^2(q+\overline{q})$, xF_3 is related to their difference, $xF_3 \approx x\sum 2e_q a_q(q-\overline{q})$ (where a_q is the axial-vector quark coupling and e_q the quark electric charge) and F_L vanishes. At higher orders, terms related to the gluon density distribution ($\alpha_s g$) appear, in particular F_L is strongly related to the low-x gluon.

The inclusive CC ep cross section can be expressed in terms of another set of structure functions and in LO the e^+p and e^-p cross sections are sensitive to different quark flavour densities:

$$\begin{aligned} e^+: \ \tilde{\sigma}^{e^+p}_{CC} &= x\overline{U} + (1-y)^2 xD \\ e^-: \ \tilde{\sigma}^{e^-p}_{CC} &= xU + (1-y)^2 x\overline{D}. \end{aligned} \tag{3}$$

Here U and D denote the sum over up- and down-type quarks; the latter include also strange and beauty quarks and the former charm quarks. Beyond LO, the QCD predictions for the DIS structure functions are obtained by convoluting the PDFs with the respective coefficient functions (hard process matrix elements). The treatment of heavy charm and beauty quark production is a crucial point in these calculations and several schemes exist:

- In the Fixed Flavour Number (FFN) scheme [30–32] only the gluon and the light quarks are considered as partons within the proton and massive quarks (with mass m_h) are produced perturbatively in the final state. The lowest order process is the fusion of a gluon in the proton with a boson from the lepton to produce a heavy quark and an antiquark. The modern series of PDFs that use this scheme as default are ABM and JR PDF groups.
- In the Zero-Mass Variable Flavour Number (ZM-VFN) scheme [33] the heavy quark densities are included in the proton for Q^2 values above a threshold $\sim m_h^2$ and are treated as massless in both the initial and final states. The lowest order process is the scattering of a heavy quark in the proton with the lepton via (electroweak) boson exchange. This scheme is expected to be reliable only in the region $Q^2 \gg m_h^2$. This is the scheme that was used in the past by PDF groups.
- In the General-Mass Variable-Flavour Number (GM-VFN) scheme [34] heavy quark production is treated for $Q^2 \leq m_h^2$ in the FFN scheme and for $Q^2 \gg m_h^2$ in a fully massive scheme. The modern series of PDF groups that use this scheme are MSTW, CT(CTEQ), NNPDF, and HERAPDF.

All three schemes are available in `HERAFitter`. In the following the implemented variants are briefly discussed.

FFN scheme: In `HERAFitter` this scheme can be accessed via the `QCDNUM` implementation or through the interface to the open-source code `OPENQCDRAD` (as implemented by the ABM group) [35]. The latter implementation also includes the running mass definition of the heavy quark mass [36]. The running mass scheme has the advantage of reducing the sensitivity of the DIS cross sections to higher order corrections, and improving the theoretical precision of the mass definition. In `QCDNUM`, the calculation of the heavy quark contributions to DIS structure functions are available at Next-to-Leading-Order (NLO) and only electromagnetic exchange contributions are taken into account. In the ABM implementation the heavy quark contributions to CC structure functions are available and, for the NC case, the QCD corrections to the massive Wilson coefficients at Next-to-Next-to Leading Order (NNLO) are provided at the best currently known approximation [37].

ZM-VFN scheme: The scheme is available for the DIS structure function calculation via interface to the `QCDNUM` package.

GM-VFN Thorne-Roberts scheme: The Thorne-Roberts (TR) scheme [38] was designed to provide a smooth transition from the massive FFN scheme at low scales $Q^2 < m_h^2$ to the massless ZM-VFNS scheme at high scales $Q^2 \gg m_h^2$. However, the original version was technically difficult to implement beyond NLO, and was updated to the TR' scheme [39] which is simpler (and closer to the ACOT-scheme, see below). There are two different variants of the TR' schemes: TR' standard (as used in MSTW PDF sets [6, 39]) and TR' optimal [40], with a smoother transition across the heavy quark threshold region. Both of these variants are accessible within the `HERAFitter` package at LO, NLO and NNLO.

GM-VFN ACOT scheme: The Aivazis-Collins-Olness-Tung scheme belongs to the group of VFN factorisation schemes that use the renormalization method of Collins-Wilczek-Zee (CWZ) [41]. This scheme unifies the low scale $Q^2 < m_h^2$ and high scale $Q^2 > m_h^2$ regions; thus, it provides a smooth interpolation across the full energy regime. It is built upon the massive factorization theorem by Collins [41] to incorporate the heavy quark masses for $Q^2 > m_h^2$; hence, it can be consistently applied order by order in the perturbation theory. Within the ACOT package, different variants of the ACOT scheme are available: ACOT-Full, S-ACOT-χ, ACOT-ZM, $\overline{\text{MS}}$ at LO and NLO. For the longitudinal structure function higher order calculations are also available. The ACOT-Full implementation takes into account the quark masses and it reduces to ZM $\overline{\text{MS}}$ scheme in the limit of masses going to zero, but it has the disadvantage that it is computationally intensive (addressed in section 3).

Calculations of higher-order electroweak corrections to DIS scattering at HERA are available in `HERAFitter`, performed in the on-shell scheme where the gauge bosons masses M_W and M_Z are treated symmetrically as basic parameters together with the top, Higgs and fermion masses. These electroweak corrections are based on the EPRC package [42]. The code provides the running of α using the most recent parametrisation of the hadronic contribution to Δ_α [43], as well as an older version from Burkhard [44].

2.2 Drell Yan processes in pp or $p\bar{p}$ collisions

The Drell Yan (DY) process provides further valuable information about PDFs. In pp and $p\bar{p}$ scattering, the Z/γ and W production probe bi-linear combinations of quarks. Complementary information on the different quark densities can be obtained from the W asymmetry (d, u and their ratio), the ratio of the W and Z cross sections (sensitive to the flavor composition of the quark sea, in particular to the s density), and associated W and Z production with heavy quarks (sensitive to s and c quark densities).

Presently, the predictions for Drell-Yan and W and Z production are known to NNLO and W, Z in association with heavy flavour quarks are known to NLO. There are several possibilities for obtaining the theoretical predictions for DY production in `HERAFitter`. At LO an analytic calculation is available within the package and described below:

The LO DY triple differential cross section in invariant mass M, boson rapidity y and Centre of Mass lepton Scattering (CMS) angle $\cos\theta$, for NC, can be written as [45, 46]:

$$\frac{d^3\sigma}{dMdy\cos\theta} = \frac{\pi\alpha^2}{3MS}\sum_q P_q\left[f_q(x_1,Q^2)f_{\bar{q}}(x_2,Q^2) + (q \leftrightarrow \bar{q})\right], \tag{4}$$

where S is the squared CMS beam energy, $x_{1,2} = \frac{M}{\sqrt{S}}\exp(\pm y)$, $f_q(x_1,Q^2)$ is the parton number density, and P_q is a partonic cross section.

The expression for CC scattering has a simpler form:

$$\frac{d^3\sigma}{dMdy\cos\theta} = \frac{\pi\alpha^2}{48S\sin^4\theta_W}\frac{M^3(1-\cos\theta)^2}{(M^2-M_W^2)+\Gamma_W^2 M_W^2} \sum_{q_1,q_2} V_{q_1q_2}^2 f_{q_1}(x_1,Q^2)f_{q_2}(x_2,Q^2), \tag{5}$$

where $V_{q_1q_2}$ is the CKM quark mixing matrix and M_W and Γ_W are the W boson mass and decay width.

The simple form of these expressions allows the calculation of integrated cross sections without the use of Monte-Carlo (MC) techniques which often introduce statistical fluctuations. In both NC and CC expressions PDFs factorise as functions dependent only on boson rapidity y and invariant mass M, while the integral in $\cos\theta$ can be computed analytically.

The NLO and NNLO calculations are highly demanding in terms of the computing power and time, and k-factor or fast grid techniques must be employed (see section 3 for details), interfaced to programs such as `MCFM` [47–49], available for NLO calculations, or FEWZ [50] and DYNNLO [51] for NLO and NNLO.

2.3 Jet production in ep and pp or $p\bar{p}$ collisions

Jet production at high transverse momentum is sensitive to the high-x gluon PDF (see e.g. [6]) and can thus increase the precision of the gluon PDF determination, which is particularly important for the Higgs production and searches for new physics. Jet production cross sections are only currently known to NLO, although NNLO calculations are now quite advanced [52–54]. Within `HERAFitter` the programs such `MCFM` and `NLOJet++` [55, 56] may be used for the calculation of jet production. Similarly to the DY case, the calculation is very demanding in terms of computing power. Therefore, to allow the possibility to include ep, pp or $p\bar{p}$ jet cross section measurements in QCD fits to extract PDFs and α_s fast grid techniques are used (see section 3).

2.4 Top-quark production in pp and $p\bar{p}$ collisions

Top-quark pairs ($t\bar{t}$) are produced at hadron colliders dominantly via gg fusion and $q\bar{q}$ annihilation. Measured $t\bar{t}$ cross sections provide additional constraints in particular on the gluon density at medium to high values of x, on α_s and on the top-quark mass, m_t. Single top quarks are produced via electroweak interactions and single-top cross sections can be used, for example, to probe the ratio of the u and d densities in the proton as well as the b-quark PDF. Precise predictions for the total $t\bar{t}$ cross section have become available to full NNLO recently [57]. They can be used within

HERAFitter via an interface to the program HATHOR [58]. Differential $t\bar{t}$ cross sections and predictions for single-top production can be used with HERAFitter at NLO accuracy from MCFM [49, 59–62] in combination with fast grid techniques.

3 Computational Techniques

More precise measurements require theoretical predictions with equally improved accuracy in order to maximize their impact in PDF fits. Perturbative calculations, however, get more and more involved with increasing number of Feynman diagrams at each higher order. Nowadays even the most advanced perturbative techniques in combination with recent computing hardware do not lead to sufficiently small turn-around times. The direct inclusion of computationally demanding higher-order calculations into iterative fits therefore is not possible. Relying on the fact that a full repetition of the perturbative calculation for arbitrary changes in input parameters is not necessary at each iteration step, two methods have been developed to resolve this problem: the techniques of k-factors and *fast grids*. Both are available in HERAFitter and described in the following.

k-factor technique:

k-factors are defined as the ratio of the prediction of a higher-order (slow) pQCD calculation to a lower-order (fast) calculation. Because the k-factors depend on the phase space probed by the measurement they have to be stored into a table in dependence of the relevant kinematic variables. Before the start of a fitting procedure the table of k-factors has to be computed once for a given PDF with the time consuming higher-order code. In subsequent iteration steps the theory prediction is derived from the fast lower-order calculation multiplied by the pre-tabulated k-factors.

However, this procedure neglects the fact that the k-factors are process dependent and, as a consequence, they have to be re-evaluated for the newly determined PDF at the end of the fit in order to check for any changes. Usually, the fit is repeated until input and output k-factors have converged. In summary, this technique avoids to iterate the higher-order calculation at each step, but still requires a couple of repetitions depending on the analysis.

– In DIS, the special case occurs of accurate but computationally slow calculations of the heavy flavour schemes. For this purpose, "FAST" heavy flavour schemes are implemented in HERAFitter with k-factors defined as the ratio of calculations at the same perturbative order but for massive vs. massless quarks, e.g. NLO (massive)/NLO (massless). In the HERAFitter implementation, these k-factors are calculated only for the starting PDF and hence, the "FAST" heavy flavour schemes should only be used for quick checks, i.e. full heavy flavour schemes are recommended (with an exception of ACOT case where, due to long computation time, the k-factors are used in the default settings).
This "FAST" method was employed in the QCD fits to the HERA data shown in Fig. 4. In this case, the ACOT scheme was used as a cross check of the central results [21].

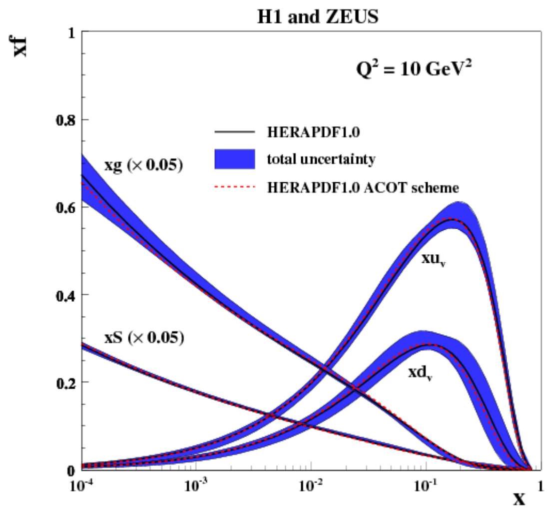

Fig. 4 Overview showing the $u-$ and d-valence, the total sea (scaled), and gluon (scaled) PDFs of the NLO HERAPDF1.0 set with their total uncertainty at the scale of $Q^2 = 10$ GeV^2 obtained using the TR' scheme and compared to the PDFs obtained with the ACOT scheme using the k-factor technique (red).

Fast grid technique:

Fast grid techniques exploit the fact that iterative PDF fitting procedures do not impose completely arbitrary changes to the types and shapes of the parameterised functions that represent each PDF. Instead, it can be assumed that a generic PDF can be approximated by a set of interpolating functions with a sufficient number of strategically well-chosen support points. The quality, i.e. the accuracy of this approximation, can be tested and optimised by a number of means, the simplest one being an increase in the number of support points. Ensuring an approximation bias that is negligibly small for all practical purposes this method can be used to perform only once for the set of interpolating functions the time consuming higher-order calculation involving complicated convolutions, see Eq. 1. The repetition of a cross section evaluation for a particular PDF set then is very fast and implies only sums over the set of interpolators multiplied by factors depending on the respective PDF. The described approach applies equally to processes involving one or two hadrons in the initial state as well as to the renormalisation and factorisation scale dependence in the convolution of the PDFs with the partonic cross section.

This technique was pioneered in the `fastNLO` project [63] to facilitate the inclusion of notoriously time consuming jet cross sections at NLO into PDF fits. The `APPLGRID` [64] package extended first a similar methodology to DY production. While differing in their interpolation and optimisation strategies, both packages construct tables with grids for each bin of an observable in two steps: In the first step the accessible phase space in the parton momentum fractions x and the renormalisation and factorisation scales μ_R and μ_F is explored in order to optimize the table size. The second step consists of the actual grid construction and filling for the requested observables. Higher-order cross sections can then be restored very efficiently from the preproduced grids while varying externally provided PDF sets, μ_R and μ_F, or the strong coupling $\alpha_s(Q)$. The approach can in principal be extended to arbitrary processes, but requires to establish an interface between the higher-order theory programs and the fast interpolation frameworks. Work in that direction is ongoing for both packages. They are described in some more detail in the following:

- The `fastNLO` project [63] has been interfaced to the `NLOJet++` program [55] for the calculation of jet production in DIS [65] as well as 2- and 3-jet production in hadron-hadron collisions at NLO [56, 66]. To demonstrate the applicability to higher-orders, threshold corrections at 2-loop order, which approximate the NNLO for the inclusive jet cross section, have been included into the framework as well [67] following Ref. [68].
  Since it is completely sufficient in most cases, the original `fastNLO` tables of version 1.4 could be evaluated with arbitrary scale factors for the renormalisation scale μ_R, but only with a fixed set of predefined scale factors for μ_F. The significantly updated version 2 of `fastNLO` [69] allows in addition to the less space consuming tables with fixed μ_F factors the creation of tables, where μ_R and μ_F can be chosen freely as a function of two pre-defined observables, e.g. jet $p_\perp$ and Q for DIS.
  `fastNLO` can be obtained from [70], where numerous precalculated grid tables for jet cross sections can be downloaded as well. The latest release is compatible with `LHAPDF` versions 5 or 6, includes the α_s evolution package `CRunDec` [71], and can also be linked to use the `HOPPET` or `QCDNUM` evolution. The C++ library part can be accessed via an optional Python interface.
  Dedicated `fastNLO` libraries and tables required for comparison to particular datasets are included in the `HERAFitter` package. In this case, the evaluation of the strong coupling constant is taken consistently with the PDF evolution from the `QCDNUM` code. The interface to the `fastNLO` tables from within `HERAFitter` was used in a recent CMS analysis, where the impact on the extraction of the PDFs from the inclusive jet cross section is investigated [26]. The influence on the gluon density by the CMS inclusive jet data is illustrated in Fig. 5.

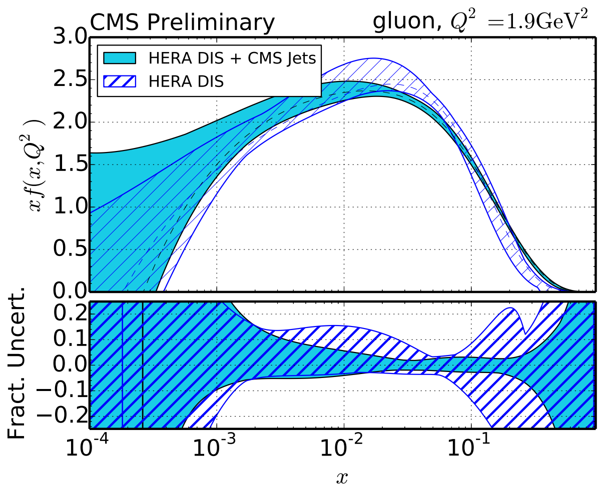

Fig. 5 The gluon density as a function of x as derived from HERA inclusive DIS data alone (cyan) and in combination with CMS inclusive jet data from 2011 (blue hatched), where bands represent the total uncertainty of the PDFs. The PDFs are shown at the starting scale $Q^2 = 1.9$ GeV^2.

- The `APPLGRID` package [64], which is also available from [72], implements in addition to the jet cross section from `NLOJet++` calculations of DY production. The look-up tables (grids) can be generated with modified versions of the MCFM parton level generator for DY [47–49]. For NNLO predictions k-factors can be applied.
  The `HERAFitter` interface to `APPLGRID` was used by the ATLAS collaboration to extract the strange quark density of the proton from W and Z cross sections [22]. An illustration of ATLAS PDFs extracted using k-factors is shown in Fig. 6 togehter with the comparison to global PDF sets CT10 [7] and NNPDF2.1 [8].

4 Fit Methodology

There are a considerable number of choices available when performing a QCD fit analysis which require careful investigation (i.e. functional parametrisation form, heavy quark

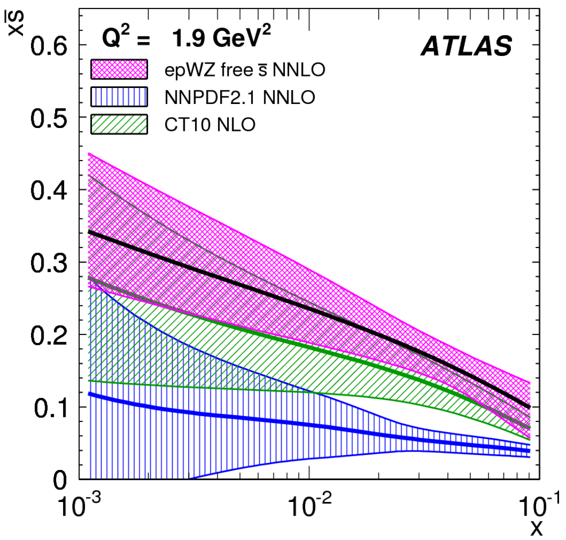

Fig. 6 The strange anti-quark density versus x for the ATLAS epWZ free sbar NNLO fit (magenta band) compared to predictions from NNPDF2.1 (blue hatched) and CT10 (green hatched) at $Q^2 = 1.9$ GeV^2.

masses, alternative theoretical calculations, method of minimisation, interpretation of uncertainties etc.). It is desirable to be able to discriminate or quantify the effect of the chosen ansatz, ideally within a common framework, and `HERAFitter` is optimally designed for such tests. The methodology employed by `HERAFitter` relies on a flexible and modular framework that allows for independent integration of the state-of-the-art techniques, either related to the inclusion of a new theoretical calculation, or to new approaches to treat uncertainties.

In this section we briefly describe the available options in `HERAFitter` ranging from the functional form used to parametrise PDFs and the choice of the form of the χ^2 function, to different methods to assess the experimental uncertainties on extracted PDFs.

In addition, as an alternative approach to a complete QCD fit, the Bayesian reweighting method, which is also available in the `HERAFitter`, is described in this section.

4.1 Functional Forms for PDF parametrisation

The PDFs are parametrised at a starting scale which is chosen by the user. Various functional forms can be tested using free parameters to be extracted from the fit:

Standard Polynomials: The term standard is understood to refer to a simple polynomial that interpolates between the low and high x regions:

$$xf(x) = Ax^B(1-x)^C P_i(x), \tag{6}$$

Standard forms are commonly used by PDF groups. The parametrised PDFs at HERA are the valence distributions xu_v and xd_v, the gluon distribution xg, and the u-type and d-type sea $x\bar{U}$, $x\bar{D}$, where $x\bar{U} = x\bar{u}$, $x\bar{D} = x\bar{d} + x\bar{s}$ at the starting scale chosen below the charm mass threshold. The $P_i(x)$ for the HERAPDF [21] style takes the simple Regge-inspired form $(1+\varepsilon\sqrt{x}+Dx+Ex^2)$ with additional constraints relating to the flavour decomposition of the light sea. For the CTEQ style, $P_i(x)$ takes the form $e^{a_3 x}(1+e^{a_4}x+e^{a_5}x^2)$. QCD number and momentum sum-rules are used to determine the normalisations A for the valence and gluon distributions. The sum-rules can be evaluated analytically.

Log-Normal Distributions: A bi-log-normal distribution to parametrise the x dependence of the PDFs is also available in `HERAFitter`. This parametrisation is motivated by multiparticle statistics [18]. The following functional form can be used:

$$xf(x) = x^{p-b\log(x)}(1-x)^{q-\log(1-x)}. \tag{7}$$

This function can be regarded as a generalisation of the standard functional form described above. In order to satisfy the QCD sum rules this parametric form requires numerical integration.

Chebyshev Polynomials:

A flexible Chebyshev polynomial based parametrisation can be used for the gluon and sea densities. The polynomials use $\log x$ as an argument to emphasize the low x behavior. The parametrisation is valid for $x > x_{\min} = 1.7 \times 10^{-5}$. The PDFs are multiplied by $(1-x)$ term to ensure that they vanish as $x \to 1$. The resulting parametric form is

$$xg(x) = A_g(1-x)\sum_{i=0}^{N_g-1} A_{g_i} T_i\left(-\frac{2\log x - \log x_{\min}}{\log x_{\min}}\right), \tag{8}$$

$$xS(x) = (1-x)\sum_{i=0}^{N_S-1} A_{S_i} T_i\left(-\frac{2\log x - \log x_{\min}}{\log x_{\min}}\right). \tag{9}$$

Here the sum over i runs up to $N_{g,S} = 15$ order Chebyshev polynomials of the first type T_i for the gluon, g, and sea-quark, S, density, respectively. The normalisation A_g is given by the momentum sum rule. The advantages of this parametrisation are that the momentum sum rule can be evaluated analytically and that for $N \geq 5$ the fit quality is already similar to the standard Regge-inspired parametrisation with a similar number of parameters.
Such study of the parametrisation uncertainty at low Bjorken $x \leq 0.1$ for PDFs was presented in [73]. Figure 7 shows that the accuracy of the HERA data allows the gluon density to be determined in the kinematic range of $0.0005 \leq x \leq 0.05$ with a reduced parametrisation uncertainty. An additional regularisation prior leads to a significantly reduced uncertainty for $x \leq 0.0005$.

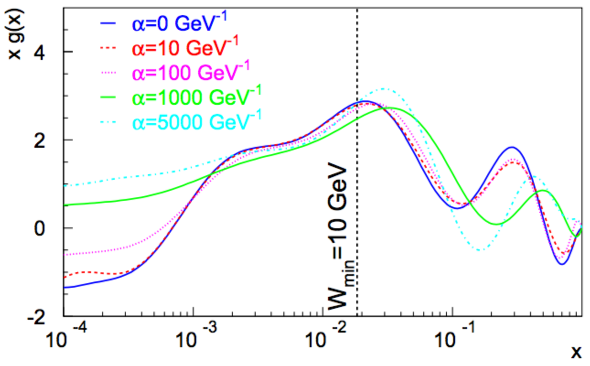

Fig. 7 Gluon PDF at the scale of $Q^2 = 1.9$ GeV^2 for various values of the length-prior weight using the Chebyshev parametrisation expanded to the 15th order.

External PDFs: `HERAFitter` also provides the possibility to access external PDF sets, which can be used to construct theoretical predictions for the various processes implemented in `HERAFitter`. This is possible via an interface to `LHAPDF` [19, 20] which provides access to the global PDF sets available at LO, NLO or NNLO evolved either locally through the `HERAFitter` or taken as provided by the `LHAPDF` grids. Figure 8 is produced with the drawing tools available in `HERAFitter` and illustrates the PDFs accessed from `LHAPDF`.

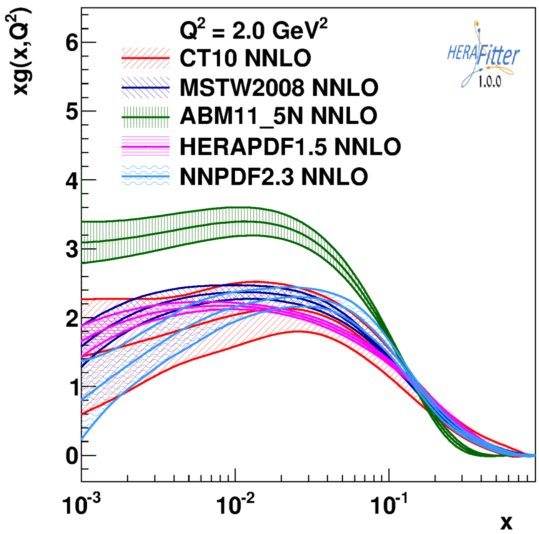

Fig. 8 Gluon density as extracted by various PDF groups at the scale of $Q^2 = 2$ GeV^2, plotted using the drawing tools from `HERAFitter`.

4.2 χ^2 representation

The PDF parameters are extracted from a χ^2 minimization process. For experimental uncertainties there are various forms to represent the χ^2 function, e.g. using a covariance matrix or representing them by nuisance parameters. In addition, there are various methods to deal with correlated systematic (or statistical) uncertainties (e.g. different scaling options, etc.). Here we summarise the options available in `HERAFitter`.

Covariance Matrix Representation: For a data point μ_i with a corresponding theory prediction m_i, the χ^2 function for the case when experimental uncertainties are given as a covariance matrix $C_{i,j}$ over data bins i and j, can be expressed in the following form:

$$\chi^2(m) = \sum_{i,j}(m_i - \mu_i)C^{-1}_{ij}(m_j - \mu_j). \qquad (10)$$

The covariance matrix can be decomposed into statistical, uncorrelated and correlated systematic contributions:

$$C_{ij} = C^{stat}_{ij} + C^{uncor}_{ij} + C^{sys}_{ij}. \qquad (11)$$

This representation can not single out the effect of a particular source of systematic uncertainty.

Nuisance Parameters Representation:

$$\chi^2(m,b) = \sum_i \frac{\left[\mu_i - m_i\left(1 - \sum_j \gamma^i_j b_j\right)\right]^2}{\delta^2_{i,\mathrm{unc}} m_i^2 + \delta^2_{i,\mathrm{stat}}\, \mu_i m_i \left(1 - \sum_j \gamma^i_j b_j\right)} + \sum_j b_j^2. \qquad (12)$$

Here μ_i is the measured central value at a point i with relative statistical $\delta_{i,\mathrm{stat}}$ and relative uncorrelated systematic uncertainty $\delta_{i,\mathrm{unc}}$. Further, γ^i_j quantifies the sensitivity of the measurement μ_i at the point i to the correlated systematic source j. The function χ^2 depends in addition on the set of systematic nuisance parameters b_j. This definition of the χ^2 function assumes that systematic uncertainties are proportional to the central prediction values (multiplicative errors), whereas the statistical errors scale with the square root of the expected number of events. The systematic shift nuisance parameters b_j as well as the PDF parameters are free parameters of the fit. Thus the fit determines the best fit to the data taking into account correlated systematic shifts of the data.

Mixed Form: It can happen that various parts of the systematic and statistical uncertainties are stored in different forms. A situation can be envisaged when the correlated systematic experimental uncertainties are provided as nuisance parameters, but the statistical bin-to-bin correlations are given in the form of a covariance matrix. `HERAFitter` offers the possibility to include such information, when provided, as well as any other mixed form of treating statistical, uncorrelated and correlated systematic uncertainties.

4.3 Treatment of the Experimental Uncertainties

Three distinct methods for propagating experimental uncertainties to PDFs are implemented in `HERAFitter` and reviewed here: the Hessian, Offset, and Monte Carlo method.

Hessian method: The technique developed by [74] presents an estimate of PDF uncertainties reflecting the experimental precision of data used in the QCD fit by examining the behaviour of χ^2 in the neighborhood of the minimum. This is known as the Hessian or error matrix method. The Hessian matrix is built by the second derivatives of χ^2 at the minimum. The PDF eigenvectors are obtained through an iterative procedure used to diagonalise the Hessian matrix and rescale the eigenvectors to adapt the step sizes to their natural scale.

Offset method:
Another method to propagate the correlated systematic experimental uncertainties from the measurements to PDFs [75] is Offset method. It uses also the χ^2 function for the central fit for which only uncorrelated uncertainties are taken into account to get the best PDF parameters. The goodness of fit can no longer be judged from the χ^2 since correlated uncertainties are ignored. The correlated systematic uncertainties of the data are then used to estimate the errors on the PDF parameters as follows. The cross section is varied by one sigma shift from the central value for each systematic source and the fit is performed. This is done for both positive and negative one sigma shifts. After this has been done for all sources the resulting deviations of each of these fits from the central PDF parameters are added in quadrature.
In most cases, the uncertainties estimated through the offset method are larger than those from the Hessian method, as the offset method does not use the information on correlated systematic uncertainties optimally.

Monte Carlo method: The PDF uncertainties can be estimated using a Monte Carlo technique [76, 77]. The method consists in preparing replicas of data sets by allowing the central values of the cross sections to fluctuate within their systematic and statistical uncertainties taking into account all point-to-point correlations. The preparation of the data is repeated for a large N (> 100 times) and for each of these replicas a QCD fit is performed to extract the PDF set. The PDF central values and uncertainties are estimated using the mean values and standard deviations over the replicas.
The MC method was checked against the standard error estimation of the PDF uncertainties as used by the Hessian method. A good agreement was found between the methods when employing for the MC approach the assumption that uncertainties (statistical and systematic) follow Gaussian distribution [18]. This comparison is illustrated in Fig. 9. Similar findings were observed also in the MSTW global analysis [78].

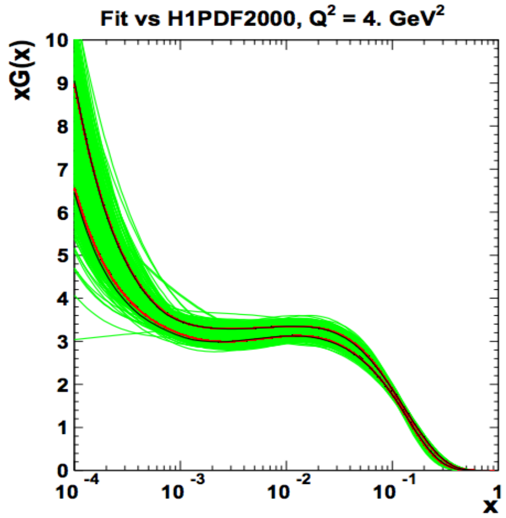

Fig. 9 Comparison between the standard error calculations as employed by the Hessian approach (black lines) and the MC approach assuming Gaussian distribution for uncertainty distributions, shown here for each replica (green lines) together with the evaluated standard deviation (red lines).

The usage of the nuisance parameters for the experimental uncertainty treatment in QCD fits is quite common and has an advantage of the flexible assessment of such uncertainties on PDFs. Generally, the experimental uncertainties are symmetrised when QCD fits are performed, however often the provided uncertainties are rather asymmetric. `HERAFitter` provides the possibility to use asymmetric systematic uncertainties. The technical implementation relies on the assumption that asymmetric uncertainties can be described by a parabolic function, as given below:

$$f_i(b_j) = \omega^i_j b_j^2 + \gamma^i_j b_j, \tag{13}$$

where the coefficients ω^i_j, γ^i_j are defined as up and down shifts of the cross sections to a nuisance parameter, $S^{\pm}_{ij}$,

$$\omega^i_j = \frac{1}{2}\left(S^-_{ij} + S^+_{ij}\right), \qquad \gamma^i_j = \frac{1}{2}\left(S^-_{ij} + S^+_{ij}\right) \tag{14}$$

For this case the definition of the χ^2 from Eq. 12 is extended with the parabolic approximation for asymmetric uncertainties, such that the expected cross section is adjusted to be

$$m_i(1 - \sum_j \gamma^i_j b_j) \rightarrow m_i \left(1 - \sum_j b_j(\omega^i_j b_j + \gamma^i_j)\right). \tag{15}$$

The minimisation is performed using fixed number of iterations (typically ten), with rapid convergence.

4.4 Treatment of the Theoretical Input Parameters

The results of a QCD fit depend not only on the input data but also on the input theoretical ansatz, which is also uncertain. Nowadays, modern PDF sets try to address the impact of the choices of theoretical parameters by providing alternative PDFs with different choices of the mass of the charm quarks m_c, mass of the bottom quarks m_b and the value of $\alpha_s(M_Z)$, etc. Another important input is the choice of the functional form for the PDFs at the starting scale and indeed the value of the starting scale itself. `HERAFitter` provides a platform in which such choices can readily be varied within a common framework.

4.5 Bayesian Reweighting Techniques

As an alternative to a complete QCD fit, the reweighting method (Bayesian Reweighting) is available in `HERAFitter`. Because no fit is performed, the method provides a fast estimate of the impact of new data. The original suggestion [76] was developed by the NNPDF collaboration [79, 80] and later extended [78] to work not only on the NNPDF replicas, but also on the eigenvectors provided by most PDF groups.

The Bayesian Reweighting technique uses the PDF probability distributions which are modified with weights to account for the difference between theory predictions and new data. In the NNPDF method the PDFs are constructed as ensembles of $N_{\rm rep}$ parton distribution functions and observables $\mathscr{O}(\mathrm{PDF})$ are conventionally calculated from the average of the predictions obtained from the ensemble $\langle\mathscr{O}(\mathrm{PDF})\rangle = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}\mathscr{O}(\mathrm{PDF}_k)$. In the case of PDF uncertainties provided by standard Hessian eigenvector error sets, this can be achieved by creating the k-th random replica by introducing random fluctuations around the central PDF set.

As a next step, the initial PDF probability distributions are updated by applying weights w_k, calculated as:

$$w_k = \frac{(\chi^2_k)^{\frac{1}{2}(N_{\rm data}-1)}e^{-\frac{1}{2}\chi^2_k}}{\frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}(\chi^2_k)^{\frac{1}{2}(N_{\rm data}-1)}e^{-\frac{1}{2}\chi^2_k}}, \tag{16}$$

where $N_{\rm data}$ is the number of new data points, k denotes the specific replica for which the weight is calculated and χ^2_k is a difference between a given data point y_i and its theoretical prediction obtained with the k-th PDF replica:

$$\chi^2(y,\mathrm{PDF}_k) = \sum_{i,j=1}^{N_{\rm data}}(y_i - y_i(\mathrm{PDF}_k))\sigma^{-1}_{ij}(y_j - y_j(\mathrm{PDF}_k)) \tag{17}$$

The new, reweighted PDFs commonly are chosen to be based upon a smaller number of PDF sets compared to the input because replicas that are incompatible with the data are discarded in order to create a more stream-lined PDF set.

4.6 Performance Optimisation

The above mentioned features make `HERAFitter` a powerful project that encapsulates state of the art developments to debates on reaching the ultimate experimental precision.

An important factor for a feasible QCD fit which is performed by iterative χ^2 minimisation, is performance in terms of how long a calculation takes for each given data point. The performance of the `HERAFitter` code is greatly improved with several special built-in options including the $k-factor$ techniques (see section 3) and the grid techniques for the fast calculation of cross sections of particular processes for arbitrary sets of PDFs. There are also cache options, fast evolution kernels, and usage of the OpenMP (Open Multi-Processing) interface which allows parallel applications of some of the heavy flavour scheme theory predictions in DIS.

5 Alternative to DGLAP formalisms

Different approaches that are alternatives to the DGLAP formalism can be used to analyse DIS data in `HERAFitter`. These include several different dipole models and the use of transverse momentum dependent, or unintegrated PDFs, uPDFs. These approaches are discussed below.

5.1 DIPOLE models

The dipole picture provides an alternative approach to virtual photon-proton scattering at low x which allows the description of both inclusive and diffractive processes. In this approach, the virtual photon fluctuates into a $q\bar{q}$ (or $q\bar{q}g$) dipole which interacts with the proton [81]. The dipoles can be viewed as quasi-stable quantum mechanical states, which have very long life time $\propto 1/m_p x$ and a size which is not changed by scattering. The dynamics of the interaction are embedded in the dipole scattering amplitude.

Several dipole models which assume different behavior of the dipole-proton cross sections are implemented in `HERAFitter`: the Golec-Biernat-Wüsthoff (GBW) dipole saturation model [82], the colour glass condensate approach to the high parton density regime called the Iancu-Itakura-Munier (IIM) dipole model [83] and a modified GBW model which takes into account the effects of DGLAP evolution called the Bartels-Golec-Kowalski (BGK) dipole model [84].

GBW model: In the GBW model the dipole-proton cross section $\sigma_{\rm dip}$ is given by

$$\sigma_{\rm dip}(x,r^2) = \sigma_0\left(1 - \exp\left[-\frac{r^2}{4R_0^2(x)}\right]\right), \tag{18}$$

where r corresponds to the transverse separation between the quark and the antiquark, and R_0^2 is an x-dependent scale parameter which represents the spacing of the gluons in the proton. $R_0^2(x) = (x/x_0)^\lambda$ is called the saturation radius. The fitted parameters are the cross-section normalisation σ_0 and x_0 and λ. This model gives exact Bjorken scaling when the dipole size r is small.

IIM model: The IIM model assumes an improved expression for the dipole cross section which is based on the Balitsky-Kovchegov equation [85]. The explicit formula for σ_{dip} can be found in [83]. The fitted parameters are an alternative scale parameter $\tilde{R}$, x_0 and λ.

BGK model: The BGK model modifies the GBW model by taking into account the DGLAP evolution of the gluon density. The dipole cross section is given by

$$\sigma_{\mathrm{dip}}(x,r^2) = \sigma_0 \left(1 - \exp\left[-\frac{\pi^2 r^2 \alpha_s(\mu^2) x g(x,\mu^2)}{3\sigma_0}\right]\right). \tag{19}$$

The factorization scale μ^2 has the form $\mu^2 = C_{bgk}/r^2 + \mu_0^2$. This model relates to the GBW model using the idea that the spacing R_0 is inverse to the gluon density. The gluon density parametrized at some starting scale Q_0^2 by $xg(x) = A_g x^{-\lambda_g}(1-x)^{C_g}$ is evolved to larger scales using DGLAP evolution. The fitted parameters for this model are σ_0, μ_0^2 and three parameters for the gluon density: A_g, λ_g, C_g. The parameter C_{bgk} is kept fixed: $C_{bgk} = 4.0$.

BGK model with valence quarks:
The dipole models are valid in the low-x region only, where the valence quark contribution is small. The new HERA F_2 data have a precision which is better than 2%. Therefore, in `HERAFitter` the contribution of the valence quarks is taken from the PDF fits and added to the original BGK model, this is uniquely possible within the `HERAFitter` framework.

5.2 Transverse Momentum Dependent PDFs with CCFM

Here another alternative approach to collinear DGLAP evolution is presented. In high energy factorization [86] the measured cross section is written as a convolution of the partonic cross section $\hat{\sigma}(k_t)$, which depends on the transverse momentum k_t of the incoming parton, with the k_t-dependent parton distribution function $\tilde{\mathscr{A}}(x,k_t,p)$ (transverse momentum dependent (TMD) or unintegrated uPDF):

$$\sigma = \int \frac{dz}{z} d^2k_t \hat{\sigma}(\frac{x}{z},k_t) \tilde{\mathscr{A}}(x,k_t,p) \tag{20}$$

where p is the factorization scale. Generally, the evolution of $\tilde{\mathscr{A}}(x,k_t,p)$ can proceed via the BFKL[**?**] DGLAP [11–15] or via the CCFM [87–90] evolution equations. In `HERAFitter`, an extension of the CCFM evolution has been implemented. Since the evolution cannot be easily obtained in a closed form, first a kernel $\tilde{\mathscr{A}}(x'',k_t,p)$ is determined from the MC solution of the CCFM evolution equation, and is then folded with a non-perturbative starting distribution $\mathscr{A}_0(x)$ [91]:

$$\begin{aligned} x\mathscr{A}(x,k_t,p) &= x\int dx' \int dx'' \mathscr{A}_0(x) \tilde{\mathscr{A}}(x'',k_t,p)\, \delta(x'\cdot x'' - x) \\ &= \int dx' \int dx'' \mathscr{A}_0(x) \tilde{\mathscr{A}}(x'',k_t,p)\, \frac{x}{x'} \delta(x'' - \frac{x}{x'}) \\ &= \int dx' \mathscr{A}_0(x') \cdot \frac{x}{x'} \tilde{\mathscr{A}}\left(\frac{x}{x'},k_t,p\right). \end{aligned} \tag{21}$$

The kernel $\tilde{\mathscr{A}}$ includes all the dynamics of the evolution, Sudakov form factors and splitting functions and is determined in a grid of $50 \otimes 50 \otimes 50$ bins in x, k_t, p.

The calculation of the cross section according to Eq.(20) involves a multidimensional Monte Carlo integration which is time consuming and suffers from numerical fluctuations, and therefore cannot be used directly in a fit procedure. Instead the following procedure is applied:

$$\begin{aligned} \sigma_r(x,Q^2) &= \int_x^1 dx_g \mathscr{A}(x_g,k_t,p) \hat{\sigma}(x,x_g,Q^2) \\ &= \int_x^1 dx' \mathscr{A}_0(x') \cdot \tilde{\sigma}(x/x',Q^2). \end{aligned} \tag{22}$$

The kernel $\tilde{\mathscr{A}}$ has to be provided separately and is not calculable within the program. A starting distribution $\mathscr{A}_0$, at the starting scale Q_0, of the following form is used:

$$x\mathscr{A}_0(x,k_t) = Nx^{-B_g} \cdot (1-x)^{C_g}(1 - D_g x) \tag{23}$$

with free parameters N, B_g, C_g, D_g.

The calculation of the ep cross section follows eq.(20), with the off-shell matrix element including quark masses taken from [86] in its implementation in `CASCADE` [92]. In addition to the boson gluon fusion process, valence quark initiated $\gamma q \to q$ processes are included, with the valence quarks taken from [93].

5.3 Diffractive PDFs

Similarly to standard DIS, diffractive parton distributions (DPDFs) can be derived from QCD fits to diffractive cross sections. At HERA about 10% of deep inelastic interactions are diffractive leading to events in which the interacting proton stays intact ($ep \to eXp$). In the diffractive process the proton appears well separated from the rest of the hadronic final state by a large rapidity gap and this is interpreted as the diffractive dissociation of the exchanged virtual photon to produce a hadronic system X with mass much smaller than W and the same net quantum numbers as the exchanged photon. For such processes, the proton vertex factorisation approach is assumed where diffractive DIS is mediated by the exchange of hard Pomeron or a secondary Reggeon. The factorisable pomeron picture has proved remarkably successful in the description of most of these data.

In addition to the usual variables x, Q^2, one must consider the squared four-momentum transfer t (the undetected momentum transfer to the proton system) and the mass M_X of the diffractively produced final state. In practice, the variable M_X is often replaced by $\beta = \frac{Q^2}{M_X^2+Q^2-t}$. In models based on a factorisable Pomeron, β may be viewed as the fraction of the pomeron longitudinal momentum which is carried by the struck parton, $x = \beta x_{IP}$.

For the inclusive case, the diffractive cross-section can be expressed as:

$$\frac{d\sigma}{d\beta\, dQ^2 dx_{IP}\, dt} = \frac{2\pi\alpha^2}{\beta Q^4} \left(1+(1-y)^2\right) \overline{\sigma}^{D(4)}(\beta, Q^2, x_{IP}, t) \quad (24)$$

where the "reduced cross-section" , $\overline{\sigma}$, is defined as

$$\overline{\sigma}^{D(4)} = F_2^{D(4)} - \frac{y^2}{1+(1-y)^2} F_L^{D(4)} = F_T^{D(4)} + \frac{2(1-y)}{1+(1-y)^2} F_L^{D(4)}. \quad (25)$$

With $x = x_{IP}\beta$ we can relate this to the standard DIS formula. The diffractive structure functions can be expressed as convolutions of the calculable coefficient functions with diffractive quark and gluon distribution functions, which in general depend on all of x_{IP}, Q^2, β, t.

The diffractive PDFs in HERAFitter are implemented following the prescription of ZEUS publication [94] and can be used to reproduce the main results.

6 Application of HERAFitter

The HERAFitter project has successfully incorporated a wide variety of tools to facilitate investigations of the HEP experimental data and theoretical calculations. It provides a versatile interface for understanding and interpreting new data and the derived PDFs in the context of precision QCD theory. The HERAFitter platform not only allows the extraction of PDFs but also of theory parameters such as the strong coupling and heavy quark masses. The parameters and distributions are ouput with a quantitative asssessment of the fit quality with fully detailed information on experimental and theoretical uncertainties. The results are also output to PDF grids that can be used to study predictions for SM or beyond SM processes, as well as for the study of the impact of future collider measurements (using pseudo-data).

So far the HERAFitter platform has been used to produce grids from the QCD analyses performed at HERA (HERAPDF series [21]), and their extension to the LHC using measurements from ATLAS [22, 23] (the first ever ATLAS PDF sets [95]).

New results that have been based on the HERAFitter platform include the following SM processes studied at the LHC: inclusive Drell-Yan and $W$and Z production [22, 24, 25]; inclusive jets [23, 26] production. At HERA, the results of QCD analyses using HERAFitter are published for inclusive H1 measurements [27] and the recent combination of charm production measurements in DIS [28]. The HERAFitter framework also provides an unique possibility to make impact studies for future colliders as illustrated by the QCD studies that have been performed to explore the potential of the LHeC data [96].

In addition, a recent study based on a set of parton distribution functions determined with the HERAFitter program using HERA data was performed [97]. It addresses the issue of correlations between uncertainties for the LO, NLO and NNLO sets. These sets are then propagated to study uncertainties for ratios of cross sections calculated at different orders in QCD and a reduction of overall theoretical uncertainty is observed.

7 Summary

The HERAFitter project is a unique platform for QCD analyses to study the structure of the proton. It incorporates not only the crucial data on Deep Inelastic Scattering from HERA but also data from the hadron colliders which are sensitive to Parton Distribution Functions. A variety of up-to-date theory calculations are available for each process at LO, NLO and NNLO when possible. HERAFitter has flexible modular structure and contains many different useful tools for PDF interpretation. HERAFitter is the first open source platform which is optimal for benchmarking studies.

Acknowledgements HERAFitter developers team acknowledges the kind hospitality of DESY and funding by the Helmholtz Alliance "Physics at the Terascale" of the Helmholtz Association. We are grateful to the DESY IT department for their support of the HERAFitter developers. Additional support was received from BMBF-JINR cooperation program, Heisenberg-Landau program and RFBR grant 12-02-91526-CERN a. We aslo acknowledge Nathan Hartland with Luigi Del Debbio for contributing to the implementation of the Bayesian Reweighting technique and would like to thank R. Thorne for fruitful discussions.

References

1. G. Aad *et al.* [ATLAS Collaboration], Phys.Lett. **B716**, 1 (2012), [1207.7214].
2. S. Chatrchyan *et al.* [CMS Collaboration], Phys.Lett. **B716**, 30 (2012), [1207.7235].
3. J. C. Collins *et al.* (1989), *Factorization of Hard Processes (in QCD in Perturbative Quantum Chromodinamics)*, ISBN: 9971-50-564-9, 9971-50-565-7.
4. E. Perez and E. Rizvi, Rep.Prog.Phys. **76**, 046201 (2013), [1208.1178].
5. S. Forte and G. Watt, Ann.Rev.Nucl.Part.Sci. **63**, 291 (2013), [1301.6754].
6. A. Martin, W. Stirling, R. Thorne, and G. Watt, Eur. Phys. J. C **63**, 189 (2009), [arXiv:0901.0002], URL http://mstwpdf.hepforge.org/.

7. J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, *et al.*, Phys.Rev. **D89**, 033009 (2014), [1302.6246], URL `http://hep.pa.msu.edu/cteq/public/`.
8. R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, *et al.*, Nucl.Phys. **B867**, 244 (2013), [1207.1303], URL `https://nnpdf.hepforge.org/`.
9. S. Alekhin, J. Bluemlein, and S. Moch (2013), [1310.3059].
10. P. Jimenez-Delgado and E. Reya, Phys.Rev. **D80**, 114011 (2009), [0909.1711], URL `http://www.het.physik.tu-dortmund.de/pdfserver/index.html`.
11. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).
12. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 675 (1972).
13. L. N. Lipatov, Sov. J. Nucl. Phys. **20**, 94 (1975).
14. Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
15. G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
16. M. Botje (2010), http://www.nikef.nl/h24/qcdnum/index.html, [arXiv:1005.1481].
17. F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).
18. M. Dittmar, S. Forte, A. Glazov, G. Moch, S. (conveners) Altarelli, and others (contributing authors) (2009), [arXiv:0901.2504].
19. M. R. Whalley, D. Bourilkov, and R. Group (2005), [hep-ph/0508110].
20. *LHAPDF*, URL `http://lhapdf.hepforge.org`.
21. F. Aaron *et al.* [H1 and ZEUS Collaborations], JHEP **1001**, 109 (2010), [arXiv:0911.0884].
22. G. Aad *et al.* [ATLAS Collaboration], Phys. Rev. Lett. **109**, 012001 (2012), [arXiv:1203.4051].
23. G. Aad *et al.* [ATLAS Collaboration], Eur.Phys.J. **73**, 2509 (2013), [arXiv:1304:4739].
24. G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. **B725**, 223 (2013), [arXiv::1305.4192].
25. S. Chatrchyan *et al.* [CMS Collaboration], submitted to Phys. Rev. **D** (2014), [arXiv:1312.6283].
26. S. Chatrchyan *et al.* [CMS Collaboration], CMS PAS **SMP-12-028** (2014).
27. F. Aaron *et al.* [H1 Collaboration], JHEP **1209**, 061 (2012), [arXiv:1206.7007].
28. H. Abramowicz *et al.* [H1 and ZEUS Collaborations], Eur. Phys. J. **C73**, 2311 (2013), [arXiv:1211.1182].
29. R. Devenish and A. Cooper-Sarkar (2011), *Deep Inelastic Scattering*, ISBN: 0199602255,9780199602254.
30. E. Laenen *et al.*, Phys. Lett. **B291**, 325 (1992).
31. E. Laenen *et al.*, Nucl. Phys. **B392**, 162, 229 (1993).
32. S. Riemersma, J. Smith, and van Neerven. W.L., Phys. Lett. **B347**, 143 (1995), [hep-ph/9411431].
33. J. C. Collins and W.-K. Tung, Nucl. Phys. B **278**, 934 (1986).
34. R. Demina, S. Keller, M. Kramer, S. Kretzer, R. Martin, *et al.* (1999), [hep-ph/0005112].
35. S. Alekhin, *OPENQCDRAD*, a program description and the code are available via: http://www-zeuthen.desy.de/∼alekhin/OPENQCDRAD.
36. S. Alekhin and S. Moch, Phys. Lett. **B699**, 345 (2011), [arXiv:1011.5790].
37. K. H., N. Lo Presti, S. Moch, and A. Vogt, Nucl.Phys. **B864**, 399 (2012).
38. R. S. Thorne and R. G. Roberts, Phys. Rev. D **57**, 6871 (1998), [hep-ph/9709442].
39. R. S. Thorne, Phys. Rev. **D73**, 054019 (2006), [hep-ph/0601245].
40. R. S. Thorne, Phys. Rev. D **86**, 074017 (2012), [arXiv:1201.6180].
41. J. C. Collins, Phys.Rev. **D58**, 094002 (1998), [hep-ph/9806259].
42. H. Spiesberger, Private communication.
43. F. Jegerlehner, Proceedings, LC10 Workshop **DESY 11-117** (2011).
44. H. Burkhard, F. Jegerlehner, G. Penso, and C. Verzegnassi, in CERN Yellow Report on "Polarization at LEP" 1988.
45. S. D. Drell and T.-M. Yan, Phys. Rev. Lett. **25**, 316 (1970).
46. M. Yamada and M. Hayashi, Nuovo Cim. **A70**, 273 (1982).
47. J. M. Campbell and R. K. Ellis, Phys. Rev. **D60**, 113006 (1999), [arXiv:9905386].
48. J. M. Campbell and R. K. Ellis, Phys. Rev. **D62**, 114012 (2000), [arXiv:0006304].
49. J. M. Campbell and R. K. Ellis, Nucl. Phys. Proc. Suppl. **205-206**, 10 (2010), [arXiv:1007.3492].
50. Y. Li and F. Petriello, Phys.Rev. **D86**, 094034 (2012), [arXiv:1208.5967].
51. G. Bozzi, J. Rojo, and A. Vicini, Phys.Rev. **D83**, 113008 (2011), [arXiv:1104.2056].
52. A. Gehrmann-De Ridder, T. Gehrmann, E. Glover, and J. Pires, Phys. Rev. Lett. **110**, 162003 (2013), [arXiv:1301.7310].
53. E. Glover and J. Pires, JHEP **1006**, 096 (2010), [arXiv:1003.2824].
54. J. Currie, A. Gehrmann-De Ridder, E. Glover, and J. Pires, JHEP **1401**, 110 (2014), [1310.3993].
55. Z. Nagy and Z. Trocsanyi, Phys.Rev. **D59**, 014020 (1999), [hep-ph/9806317].
56. Z. Nagy, Phys.Rev.Lett. **88**, 122003 (2002), [hep-ph/0110315].
57. M. Czakon, P. Fiedler, and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013), [1303.6254].

58. M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, *et al.*, Comput.Phys.Commun. **182**, 1034 (2011), [arXiv:1007.1327].
59. J. M. Campbell, R. Frederix, F. Maltoni, and F. Tramontano, Phys.Rev.Lett. **102**, 182003 (2009), [0903.0005].
60. J. M. Campbell and F. Tramontano, Nucl.Phys. **B726**, 109 (2005), [hep-ph/0506289].
61. J. M. Campbell, R. K. Ellis, and F. Tramontano, Phys.Rev. **D70**, 094012 (2004), [hep-ph/0408158].
62. J. M. Campbell and R. K. Ellis (2012), report FERMILAB-PUB-12-078-T, [1204.1513].
63. T. Kluge, K. Rabbertz, and M. Wobisch, pp. 483–486 (2006), [hep-ph/0609285].
64. T. Carli *et al.*, Eur. Phys. J. **C66**, 503 (2010), [arXiv:0911.2985].
65. Z. Nagy and Z. Trocsanyi, Phys.Rev.Lett. **87**, 082001 (2001), [hep-ph/0104315].
66. Z. Nagy, Phys.Rev. **D68**, 094002 (2003), [hep-ph/0307268].
67. M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober [fastNLO Collaboration] (2011), [arXiv:1109.1310].
68. N. Kidonakis and J. Owens, Phys.Rev. **D63**, 054019 (2001), [hep-ph/0007268].
69. D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch [fastNLO Collaboration] (2012), [arXiv:1208.3641].
70. http://fastnlo.hepforge.org, URL `http://fastnlo.hepforge.org`.
71. B. Schmidt and M. Steinhauser, Comput. Phys. Commun. **183**, 1845 (2012), [1201.6149].
72. http://applgrid.hepforge.org, URL `http://applgrid.hepforge.org`.
73. A. Glazov, S. Moch, and V. Radescu, Phys. Lett. B **695**, 238 (2011), [arXiv:1009.6170].
74. J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, *et al.*, Phys.Rev. **D65**, 014013 (2001), [hep-ph/0101032].
75. M. Botje, J.Phys. **G28**, 779 (2002), [hep-ph/0110123].
76. W. T. Giele and S. Keller, Phys.Rev. **D58**, 094023 (1998), [hep-ph/9803393].
77. W. T. Giele, S. Keller, and D. Kosower (2001), [hep-ph/0104052].
78. G. Watt and R. Thorne, JHEP **1208**, 052 (2012), [arXiv:1205.4024].
79. R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, *et al.*, Nucl.Phys. **B855**, 608 (2012), [arXiv:1108.1758].
80. R. D. Ball *et al.* [NNPDF Collaboration], Nucl.Phys. **B849**, 112 (2011), [arXiv:1012.0836].
81. N. N. Nikolaev and B. Zakharov, Z.Phys. **C49**, 607 (1991).
82. K. Golec-Biernat and M. Wüsthoff, Phys. Rev. D **59**, 014017 (1999), [hep-ph/9807513].
83. E. Iancu, K. Itakura, and S. Munier, Phys. Lett. **B590**, 199 (2004), [hep-ph/0310338].
84. J. Bartels, K. Golec-Biernat, and H. Kowalski, Phys. Rev. D **66**, 014001 (2002), [hep-ph/0203258].
85. I. Balitsky, Nucl. Phys. B **463**, 99 (1996), [hep-ph/9509348].
86. S. Catani, M. Ciafaloni, and F. Hautmann, Nucl. Phys. B **366**, 135 (1991).
87. M. Ciafaloni, Nucl. Phys. B **296**, 49 (1988).
88. S. Catani, F. Fiorani, and G. Marchesini, Phys. Lett. B **234**, 339 (1990).
89. S. Catani, F. Fiorani, and G. Marchesini, Nucl. Phys. B **336**, 18 (1990).
90. G. Marchesini, Nucl. Phys. B **445**, 49 (1995).
91. H. Jung and F. Hautmann (2012), [arXiv:1206.1796].
92. H. Jung, S. Baranov, M. Deak, A. Grebenyuk, F. Hautmann, *et al.*, Eur.Phys.J. **C70**, 1237 (2010), [arXiv:1008.0152].
93. M. Deak, F. Hautmann, H. Jung, and K. Kutak, *Forward-Central Jet Correlations at the Large Hadron Collider* (2010), [arXiv:1012.6037].
94. S. Chekanov *et al.* [ZEUS Collaboration], Nucl. Phys. **B831**, 1 (2010), [hep-ex/09114119].
95. *ATLAS NNLO epWZ12*, available via: http://lhapdf.hepforge.org/pdfsets.
96. J. L. Abelleira Fernandez *et al.* [LHeC Study Group], Journal of Phys. **G**, 075001 (2012), [arXiv:1206.2913].
97. H. Pirumov, M. Lisovyi, A. Glazov, and HERAFitter (2014), [arXiv:1404.XXXX].