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

HERAFitter

Open Source QCD Fit Project

Version 0.91 (svn 1447)

S. Alekhin[16,17], O. Behnke[1], P. Belov[1,12], M. Botje[18], D. Britzger[1], S. Camarda[1], A.M. Cooper-Sarkar[2], K. Daum[30,31], C. Diaconu[3], J. Feltesse[19], A. Gizhko[1], A. Glazov[1], A. Guffanti[20], M. Guzzi[1], F. Hautmann[13,14,15], H. Jung[1], V. Kolesnikov[4], H. Kowalski[1], O. Kuprash[1], A. Kusina[21], S. Levonian[1], K. Lipka[1], B. Lobodzinski[29], K. Lohwasser[16], A. Luszczak[5], B. Malaescu[25], R. McNulty[28], V. Myronenko[1], S. Naumann-Emme[1], K. Nowak[1], F. Olness[21], E. Perez[23], H. Pirumov[1], R. Plačakytė[1], K. Rabbertz[6], V. Radescu[1], R. Sadykov[24], G. Salam[26,27], A. Sapronov[4], A. Schöning[10], T. Schörner-Sadenius[1], S. Shushkevich[1], W. Slominski[7], H. Spiesberger[22], P. Starovoitov[1], M. Sutton[8], J. Tomaszewska[9], O. Turkot[1], A. Vargas[1], G. Watt[11], K. Wichmann[1]

[1] Deutsches Elektronen-Synchrotron (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] Institute for Particle Physics Phenomenology, Durham University, Durham, DH1 3LE, United Kingdom
[12] Current address: Department of Physics, St. Petersburg State University, Ulyanovskaya 1, 198504 St. Petersburg, Russia
[13] Dept. of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, United Kingdom
[14] Rutherford Appleton Laboratory, Chilton OX11 0QX, United Kingdom
[15] Dept. of Theoretical Physics, University of Oxford, Oxford OX1 3NP, United Kingdom
[16] Deutsches Elektronen-Synchrotron (DESY), Platanenallee 6, D15738 Zeuthen, Germany
[17] Institute for High Energy Physics,142281 Protvino, Moscow region, Russia
[18] Nikhef, Science Park, Amsterdam, the Netherlands
[19] CEA, DSM/Irfu, CE-Saclay, Gif-sur-Yvette, France
[20] Niels Bohr Institute, University of Copenhagen, Denmark
[21] Southern Methodist University, Dallas, Texas
[22] WA ThEP, Johannes-Gutenberg-Universität Mainz, D-55099 Mainz, Germany
[23] CERN, European Organization for Nuclear Research, Geneva, Switzerland
[24] Joint Institute for Nuclear Research, Joliot-Curie str. 6, Dubna, 141980, Russia
[25] Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université, Paris-Diderot and CNRS/IN2P3, Paris, France
[26] CERN, PH-TH, CH-1211 Geneva 23, Switzerland
[27] LPTHE; CNRS UMR 7589; UPMC Univ. Paris 6; Paris 75252, France
[28] University College Dublin, Dublin 4, Ireland
[29] Max Planck Institut Für Physik, Werner Heisenberg Institut, Föhringer Ring 6, Muńchen
[30] Fachbereich C, Universität Wuppertal, Wuppertal, Germany
[31] Rechenzentrum, Universität Wuppertal, Wuppertal, Germany

Received: date / Accepted: date

Abstract `HERAFitter` [1] is an open-source package which provides a framework for the determination of the parton distribution functions (PDFs) of the proton and for multifold analyses in Quantum Chromodynamics (QCD).

Measurements of lepton-proton deep inelastic scattering (DIS) and of proton-proton (proton-antiproton) collisions at hadron colliders are included in the `HERAFitter` package, and are used to probe and constrain the partonic content of the proton.

The partonic distributions are determined by using the factorisation properties of the hadronic cross sections in which short-distance perturbatively calculable hard scatterings and long-distance contributions that are the non-perturbative universal PDFs, are factorised.

The `HERAFitter` platform provides a broad choice of options for the treatment of the experimental uncertainties and a common environment where a large number of theoretical calculations and methodological options are used to perform detailed QCD analyses. The general structure of `HERAFitter` together with available methods are described in this paper.

Keywords PDFs · QCD · Fit · proton structure

Contents

1 Introduction

The constant inflow of new experimental measurements with unprecedented accuracy from hadron colliders is a remarkable challenge for the high energy physics community to provide higher-order theory predictions and to develop efficient tools and methods for data analysis. The recent discovery of the Higgs boson [2, 3] and the extensive searches for signals of new physics in LHC proton-proton collisions demand high-precision computations to test the validity of the Standard Model (SM) and factorisation in Quantum Chromodynamics (QCD). According to collinear factorisation in perturbative QCD (pQCD) hadronic inclusive cross sections are written as

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

where 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 represent the probability of finding a specific parton a (b) in the first (second) proton carrying a fraction x_1 (x_2) of its momentum. Indices a and b in the Eq. 1 indicate 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 whereas PDFs cannot be computed analytically in QCD, they must rather be determined from measurements. 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. The gluon density in small and medium x can be accurately determined solely from the HERA data. Many processes in pp and $p\bar{p}$ collisions at LHC and Tevatron, respectively, probe PDFs in the kinematic ranges, complementary to the DIS measurements (see Fig 1). 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 asymmetries, associated production of W or Z bosons and heavy quarks, top quark, jet and prompt photon production.

The open-source QCD platform `HERAFitter` encloses the set of tools necessary for a comprehensive global QCD analysis of hadron-induced processes even at the early stage of the experimental measurement. It has been developed for determination of PDFs and extraction of fundamental QCD parameters such as the heavy quark masses or the strong coupling constant. This platform also provides 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.

This paper is organised as follows. The structure and overview of `HERAFitter` is presented in section 2. Section 3 discusses the various processes and corresponding theoretical calculations performed in the DGLAP [6–10] formalism, available in `HERAFitter`. Section 4 presents various

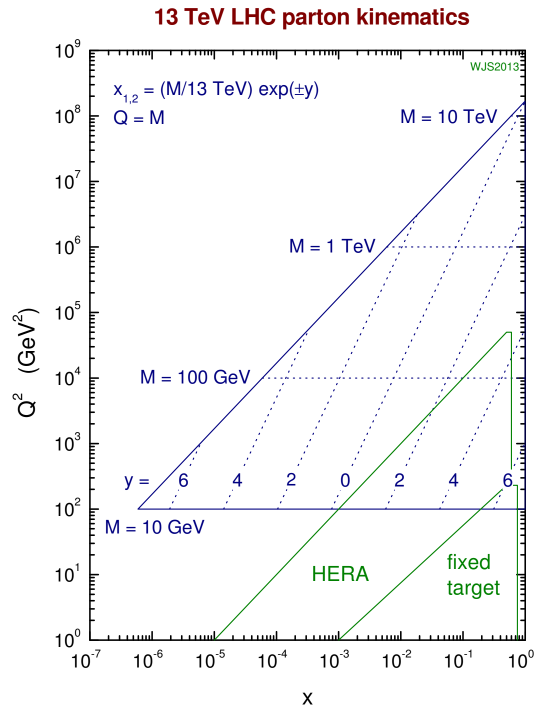

Fig. 1 The parton kinematic plane with the approximate region sensitivity to the PDFs of LHC and DIS experiments.

fast techniques employed by the theory calculations used in HERAFitter. Section 5 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 6. Specific applications of the package are given in section 7 and the summary is presented in section 8.

2 HERAFitter Structure

`HERAFitter` is a flexible open-source platform for the QCD analyses of different experimental measurements, providing a versatile environment for benchmarking studies. It is widely used within LHC experiments [11–17].

The functionality of `HERAFitter` is schematically illustrated in Fig. 2 and it can be divided in four main blocks:

Input data: Different available measurements from the various processes are implemented in the `HERAFitter` package including the full information on their uncorrelated and correlated uncertainties. HERA data are sensitive to light quark and gluon densities mostly through scaling violations, covering low and medium x ranges. These data are the basis of any proton PDF extraction, and are used by all global PDF groups [18–22]. However, improvements in precision of PDFs require additional constraints on the gluon and quark distributions at high x, better understanding of heavy quark distributions and decomposition of the light-quark sea. For these purposes, the measurements of the fixed-target experiments, Tevatron and LHC are of particular importance. The processes that are currently available in `HERAFitter` framework are listed in Tab. 1.

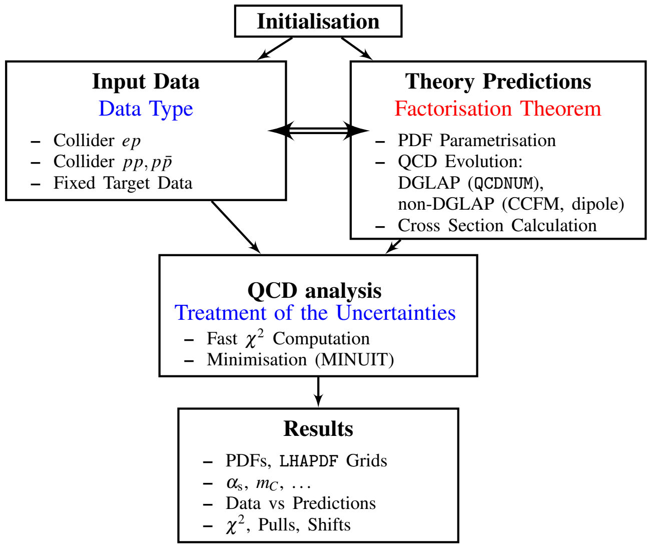

Fig. 2 Schematic structure of the `HERAFitter` program.

Theory predictions: Predictions for cross section of different processes are obtained using the factorisation approach (Eq. 1). The PDFs are parametrised at a starting input scale Q_0^2 by a chosen functional form with a set of free parameters $\mathbf{p}$. These PDFs are evolved to the scale of the measurement Q^2, $Q^2 > Q_0^2$. The evolution follows either DGLAP [6–10] (as implemented in `QCDNUM` [23]), CCFM [24–27] or dipole models [28–30]. The prediction of a particular process cross section is obtained by a convolution of the evolved PDFs and the partonic cross section, calculated at a certain order in QCD with a relevant theory program (as listed in Tab. 1).

QCD analysis: The PDFs are extracted from a least square fit by minimising the χ^2 function with respect to free parameters. The χ^2 function is formed from the input data and the theory prediction. The χ^2 is minimised iteratively with respect to $\mathbf{p}$ using the MINUIT [31] 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 as described in section 5.2). In addition, `HERAFitter` allows to study different statistics assumptions for the distribu-

Data	Process	Reaction	Theory calculations, schemes
HERA	DIS NC	$ep \to eX$	TR', ACOT ZM (QCDNUM) FFN (OPENQCDRAD, QCDNUM), TMD (uPDFevolv)
	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 references for the individual calculations and their implementations are given in the text.

tions of the systematic uncertainties i.e. Gauss [32] (see section 5.3).

Results: The resulting PDFs are provided in a format ready to be used by the LHAPDF library [33, 34] (or by TMDlib [35]). HERAFitter drawing tools can be used to display the PDFs with their uncertainty at a chosen scale. A first set of PDFs extracted by using HERAFitter is HERAPDF1.0 [36], shown in Fig. 3, which is based on HERA I data. Since then several other PDF sets were produced within the HERA [37] and LHC [38] collaborations. The comparison of data used in the fit to the theory predictions are also produced. In Fig. 4, a comparison of inclusive NC data from the HERA I running period with predictions based on HERAPDF1.0 is shown. Also shown are theory predictions, obtained using the nuisance parameter method, which accounts for correlated systematic shifts when using the nuisance parameter method that accounts for correlated systematic uncertainties. The consistency of the measurements and the theory is expressed by pulls, defined as a difference between data and theory divided by the uncorrelated error of the data. In each kinematic bin of the measurement, pulls are provided in units of sigma.

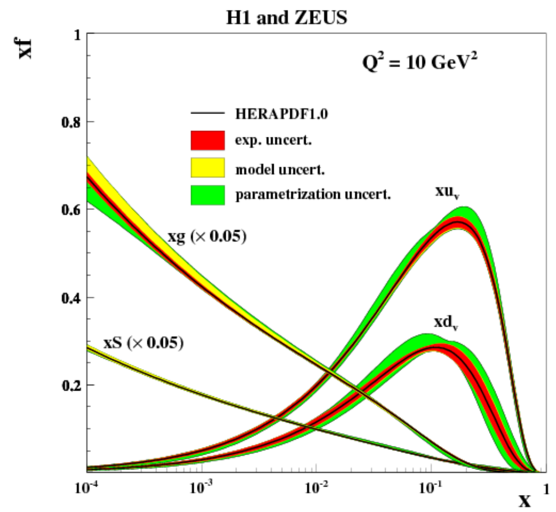

Fig. 3 Distributions of valence (xu_v, xd_v), sea (xS) and the gluon (g) densities in HERAPDF1.0 [36]. The gluon and the sea distributions are scaled down by a factor of 20. The experimental, model and parametrisation uncertainties are shown as colored bands.

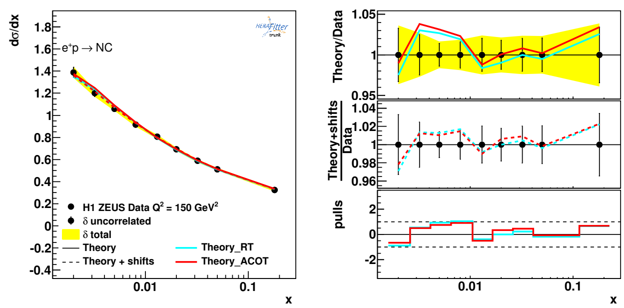

Fig. 4 An illustration of the consistency of HERA measurements [36] and the theory predictions, obtained in HERAFitter with the default drawing tool.

3 Theoretical Input

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

3.1 Deep Inelastic Scattering and Proton Sructure

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. [39]) and it will only be briefly summarised here. DIS is the process where 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 absolute squared four-momentum of the exchange boson, Q^2, the Bjorken x, and the inelasticity y, related by $y = Q^2/sx$, where s is the squared centre-of-mass

(c.o.m) energy.
The NC cross section can be expressed in terms of generalised structure functions:

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

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, $x\tilde{F}_3$ becomes important at high Q^2 and $\tilde{F}_L$ is sizable only at high y.
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:

$$\begin{aligned} \sigma_{CC}^{e^+p} &\approx x[\bar{u}+\bar{c}] + (1-y)^2 x[d+s], \\ \sigma_{CC}^{e^-p} &\approx x[u+c] + (1-y)^2 x[\bar{d}+\bar{s}]. \end{aligned} \quad (3)$$

The QCD predictions for the DIS structure functions are obtained by convoluting the PDFs with the respective coefficient functions. The DIS measurements span in the kinematic range from low to high Q^2, such that the treatment of heavy quarks (charm and beauty) and of their masses becomes important. Several schemes exist and the implemented variants in HERAFitter are briefly discussed as follows.

Zero-Mass Variable Flavour Number (ZM-VFN): In this scheme [40], the heavy quark densities are included in the proton for Q^2 values above a threshold $\sim m_h^2$ (heavy quark mass) and they 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 with $Q^2 \gg m_h^2$. In HERAFitter this scheme is available for the DIS structure function calculation via interface to the QCDNUM [23] package and it benefits from the fast QCDNUM convolution engine.

Fixed Flavour Number (FFN): In this scheme [41–43] only the gluon and the light quarks are considered as partons within the proton and massive quarks 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. 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) [44]. Through QCDNUM, the calculation of the heavy quark contributions to DIS structure functions are available at Next-to-Leading-Order (NLO), at $O(\alpha_s)$, and only electromagnetic exchange contributions are taken into account. Through the ABM implementation the heavy quark contributions to CC structure functions are available and, for the NC case, the QCD corrections to the coefficient functions at Next-to-Next-to Leading Order (NNLO) are provided at the best currently known approximation [45]. The ABM implementation also includes the running-mass definition of the heavy quark mass [46]. 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.

General-Mass Variable-Flavour Number (GM-VFN): It this scheme [47], heavy quark production is treated for $Q^2 \le m_h^2$ in the FFN scheme and for $Q^2 \gg m_h^2$ in a fully massive scheme. The recent series of PDF groups that use this scheme are MSTW, CT(CTEQ), NNPDF, and HERAPDF. HERAFitter implements different variants of the GM-VNS scheme and they are presented below:

- **GM-VFN Thorne-Roberts scheme:** The Thorne-Roberts (TR) scheme [48] 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 [49]. There are two different variants of the TR' schemes: TR' standard (as used in MSTW PDF sets [18, 49]) and TR' optimal [50], with a smoother transition across the heavy quark threshold region. Both variants are accessible within the HERAFitter package at LO, NLO and NNLO.
- **GM-VFN ACOT scheme:** The Aivazis-Collins-Olness-Tung (ACOT) scheme belongs to the group of VFN factorisation schemes that use the renormalization method of Collins-Wilczek-Zee (CWZ) [51]. This scheme unifies the low scale $Q^2 < m_h^2$ and high scale $Q^2 > m_h^2$ regions with a smooth interpolation across the full energy regime. Within the ACOT package, different variants of the ACOT scheme are available: ACOT-Full [52], S-ACOT-χ [53, 54], ACOT-ZM [52], $\overline{\text{MS}}$ 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 4).

3.2 Electroweak Corrections to DIS

Calculations of higher-order electroweak corrections to DIS scattering at HERA are available in HERAFitter in the onshell scheme. In this scheme 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 [55]. The code provides the running of α using the most recent parametrisation of the hadronic contribution to Δ_α [56], as well as an older version from Burkhard [57].

3.3 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 a 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:

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

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

$$\overline{\sigma}^{D(4)} = F_2^{D(4)} - \frac{y^2}{1+(1-y)^2}F_L^{D(4)}. \qquad (5)$$

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 x_{IP}, Q^2, β, t.

The diffractive PDFs in `HERAFitter` are implemented following the prescription of ZEUS collaboration [58].

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

Drell Yan 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).

The LO DY triple differential cross section in invariant mass M, boson rapidity y and c.o.m lepton scattering angle $\cos\theta$, for NC, can be written as [59, 60]:

$$\frac{d^3\sigma}{dMdyd\cos\theta} = \frac{\pi\alpha^2}{3MS}\sum_q P_q\left[f_q(x_1,Q^2)f_{\bar{q}}(x_2,Q^2)+(q\leftrightarrow\bar{q})\right], \qquad (6)$$

where S is the squared c.o.m 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 form:

$$\frac{d^3\sigma}{dMdyd\cos\theta} = \frac{\pi\alpha^2}{48S\sin^4\theta_W}\frac{M^3(1-\cos\theta)^2}{(M^2-M_W^2)+\Gamma_W^2 M_W^2} \sum_{q_1,q_2} V_{q_1q_2}^2 f_{q_1}(x_1,Q^2)f_{q_2}(x_2,Q^2), \qquad (7)$$

where $V_{q_1q_2}$ is the Cabibbo-Kabayashi-Masakawa (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. This form provides easy means to apply kinematic cuts to theory predictions to emulate data.

Currently, the predictions for DY and W and Z production are available to NNLO and W, Z in association with heavy flavour quarks - to NLO. There are several possibilities for obtaining the theoretical predictions for DY production in `HERAFitter`.

The NLO and NNLO calculations are computing power and time consuming and k-factor or fast grid techniques must be employed (see section 4 for details), interfaced to programs such as `MCFM` [61–63], available for NLO calculations, or FEWZ [64] and DYNNLO [65] for NLO and NNLO.

3.5 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. [18]) 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 currently only known to NLO, although calculations for higher-order contributions to jet production in proton-proton collisions are now quite advanced [66–68]. Within `HERAFitter`, programs as `MCFM` or `NLOJet++` [69, 70] 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 fast grid techniques are used to facilitate the QCD analyses including jet cross section measurements. in ep, pp and $p\bar{p}$ collisions (for details see section 4).

3.6 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 [71]. Precise predictions for the total $t\bar{t}$ cross section have become available to full NNLO recently [72]. They can be used within `HERAFitter` via an interface to the program `HATHOR` [73]. Differential $t\bar{t}$ cross section predictions can be used with `MCFM` [63, 74–77] at NLO accuracy interfaced to `HERAFitter` with fast grid techniques.

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. Predictions for single-top production are available only at NLO accuracy using `MCFM`.

4 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 the 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 as follows.

4.1 k-factor Technique

The 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.

This procedure, however, 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, appropriate treatments of the heavy quarks require computationally slow calculations. 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. For ACOT case, due to long computation time, the k-factors are used in the default settings in `HERAFitter`. Fig. 5 illustrates the PDFs extracted from the QCD fits to the HERA data, for which the “FAST” method for ACOT was used as a cross check to the main results [36].

4.2 Fast Grid Techniques

Fast grid techniques exploit the factorisable nature of the cross sections and 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 the time consuming higher-order calculation (see Eq. 1) only

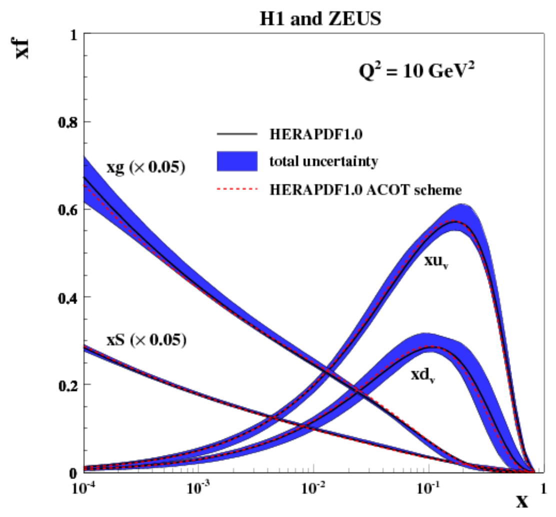

Fig. 5 Overview showing the $u-$ and d-valence, the total sea (scaled), and gluon (scaled) PDFs of the NLO HERAPDF1.0 set [36] with their total uncertainty at the scale of $Q^2 = 10$ GeV^2 obtained using the TR' scheme and compared to the PDFs obtained with the ACOT scheme using the k-factor technique (red).

once for the set of interpolating functions. 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 [78] to facilitate the inclusion of notoriously time consuming jet cross sections at NLO into PDF fits. The APPLGRID [79] 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 pre-produced 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 [78] has been interfaced to the NLOJet++ program [69] for the calculation of jet production in DIS [80] as well as 2- and 3-jet production in hadron-hadron collisions at NLO [70, 81]. 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 [82] following Ref. [83].
  The latest version of fastNLO [84] allows for a creation of tables where renormalisation and factorisation scales can be chosen freely as a function of two pre-defined observables, e.g. jet transverse momentum $p_\perp$ and Q for DIS. fastNLO can be obtained from [85], where numerous pre-calculated grid tables for jet cross sections can be downloaded as well.
  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 [15].
– In the APPLGRID package [79, 86], in addition to the jet cross sections from NLOJet++ in $pp(\bar{p})$ and DIS processes, the calculations of DY production are implemented. The look-up tables (grids) can be generated with modified versions of the MCFM parton level generator for DY [61–63]. Alternative values of the strong coupling constant as well as a posteriori variation of the renormalisation and factorisation scales can be freely chosen in the calculation of the theory predictions with the APPLGRID tables. For NNLO predictions in HERAFitter 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 [11]. An illustration of ATLAS PDFs extracted using the k-factors is shown in Fig. 6 together with the comparison to global PDF sets CT10 [19] and NNPDF2.1 [20].

5 Fit Methodology

There is a considerable number of choices available when performing a QCD analysis on e.g. the functional parametrisation form, the heavy quarks mass values, 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,

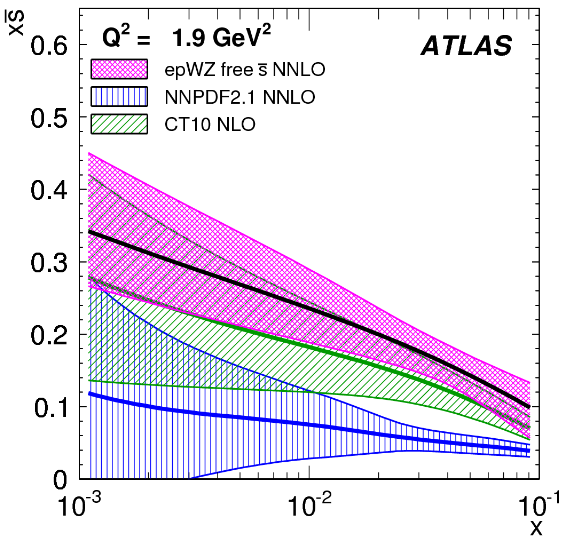

Fig. 6 The strange anti-quark density versus x for the ATLAS epWZ free sbar NNLO fit (magenta band) compared to predictions from NNPDF2.1 (blue hatched) and CT10 (green hatched) at $Q^2 = 1.9$ GeV^2. The ATLAS fit was performed using k-factor method for NNLO corrections. The figure is taken from [11].

either related to the inclusion of a new theoretical calculation, or of new approaches to treat uncertainties.

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

5.1 Functional Forms for PDF parametrisation

The PDFs are parametrised at a starting scale, chosen to be below charm mass. In `HERAFitter` various functional forms to parametrise PDFs can be used:

Standard Polynomials: A polynomial form is used to parametrise the x-dependence of the PDFs:

$$xf(x) = Ax^B(1-x)^C P_i(x), \tag{8}$$

The standard polynomial form is most commonly used by the PDF groups. In HERA PDFs, the parametrised PDFs 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. The $P_i(x)$ for the HERAPDF [36] style takes the 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.

Bi-Log-Normal Distributions: This parametrisation is motivated by multi-particle statistics and holds the following functional form:

$$xf(x) = ax^{p-b\log(x)}(1-x)^{q-d\log(1-x)}. \tag{9}$$

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 PDFs are multiplied by a $(1-x)$ term to ensure that they vanish as $x \to 1$. The resulting parametric form is

$$xg(x) = A_g\,(1-x)\sum_{i=0}^{N_g-1} A_{g_i} T_i\left(-\frac{2\log x - \log x_{\min}}{\log x_{\min}}\right), \tag{10}$$

$$xS(x) = (1-x)\sum_{i=0}^{N_S-1} A_{S_i} T_i\left(-\frac{2\log x - \log x_{\min}}{\log x_{\min}}\right). \tag{11}$$

Here the sum runs over i 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.
A study of the parametrisation uncertainty at low Bjorken $x \leq 0.1$ for PDFs was presented in [87]. Figure 7 shows the comparison of the gluon density detemined from the HERA data with the standard and the Chebyshev parametrisation.

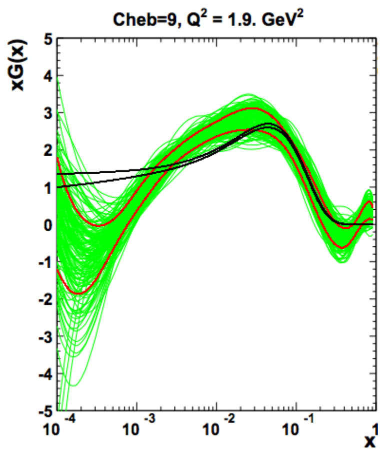

Fig. 7 The gluon density is shown at the starting scale. The black lines correspond to the error band of the gluon distribution using a standard parameterisation and it is to be compared to the case of the Chebyshev parameterisation [87].

External PDFs: HERAFitter provides the possibility to access external PDF sets, which can be used to construct theoretical predictions for the various processes of interest as implemented in HERAFitter. This is possible via an interface to LHAPDF [33, 34] which provides access to the global PDF sets available at different orders 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.

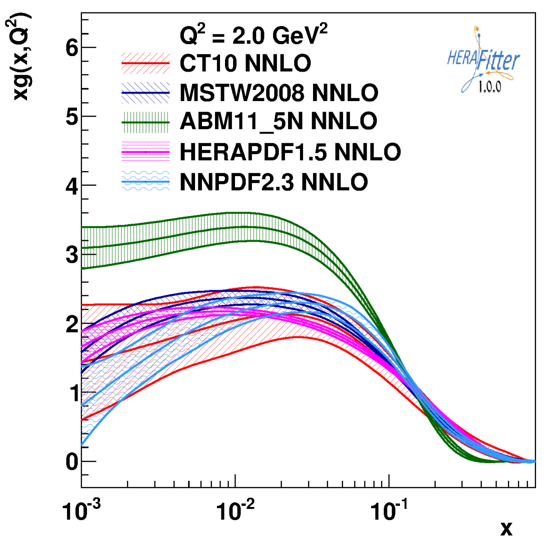

Fig. 8 Gluon density as extracted by various PDF groups at the scale of $Q^2 = 2$ GeV^2, plotted using the drawing tools from HERAFitter.

5.2 Representation of χ^2

The PDF parameters are extracted in a χ^2 minimisation process. The construction of the χ^2 accounts for the experimental uncertainties. There are various forms that can be used to represent the experimental uncertainties, e.g. using covariance matrices or providing nuisance parameters for dependence of each systematic source for each measurement data point. In addition, there are various methods to deal with correlated systematic (or statistical) uncertainties. 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 can be expressed in the following form:

$$\chi^2(m) = \sum_{i,j}(m_i - \mu_i)C^{-1}_{ij}(m_j - \mu_j). \qquad (12)$$

Here, the experimental uncertainties are given in a form of covariance matrix $C_{i,j}$ for measurements in bins i an j. The covariance matrix can be decomposed into statistical, uncorrelated and correlated systematic contributions:

$$C_{ij} = C^{stat}_{ij} + C^{uncor}_{ij} + C^{sys}_{ij}. \qquad (13)$$

With this representation the particular effect of a certain systematic source of the uncertainty can no longer be distinguished from others.

Nuisance Parameters Representation: The χ^2 form is expressed as

$$\chi^2(m,b) = \sum_i \frac{\left[\mu_i - m_i\left(1 - \sum_j \gamma^i_j b_j\right)\right]^2}{\delta^2_{i,\mathrm{unc}} m_i^2 + \delta^2_{i,\mathrm{stat}}\,\mu_i m_i \left(1 - \sum_j \gamma^i_j b_j\right)} + \sum_j b_j^2. \qquad (14)$$

Here ,μ_i is the central value of the measurement i with its 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 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 uncertainties scale with the square root of the expected number of events.

During the χ^2 minimisation, the nuisance parameters b_j and the PDFs are determined.

Mixed Form Representation: In some cases, the statistical and systematic uncertainties are provided in different forms. For example, the correlated experimental systematic uncertainties are available as nuisance parameters but the bin-to-bin statistical correlations are given in a form of covariance matrix. HERAFitter offers possibilities to include also the mixed form of treating statistical, uncorrelated and correlated systematic uncertainties.

5.3 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 in [88] presents an estimate of PDF uncertainties reflecting the experimental precision of data used in the QCD fit by examining the behavior 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 Hessian matrix is diagonalised through an iterative procedure and its PDF eigenvectors are obtained, which correspond to the orthogonal sources of uncertainties on the obtained PDF.

Offset method: The Offset method [89] uses also the χ^2 function for the central fit for which only uncorrelated uncertainties are taken into account. The goodness of the fit can no longer be judged from the χ^2 since correlated uncertainties are ignored. Instead, the correlated systematic uncertainties of the measurement are used to estimate the errors on the PDF parameters as follows. The cross section is varied by $\pm 1\sigma$ from the central value for each systematic source and the fit is performed. 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.

Monte Carlo method: The Monte-Carlo technique [90, 91] can be used to determine PDF uncertainties. 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 large N (>100 times) and for each of these replicas a QCD fit is performed. The PDF central values and experimental 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 once the Gaussian distribution of statistic and systematic uncertainties is assumed in the MC approach [32]. This comparison is illustrated in Fig. 9. Similar findings were reported by the MSTW global analysis [92].

Generally, the experimental uncertainties using nuisance parameters 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:

$$f_i(b_j) = \omega^i_j b_j^2 + \gamma^i_j b_j, \tag{15}$$

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

$$\omega^i_j = \frac{1}{2}\left(S^{+}_{ij} + S^{-}_{ij}\right), \qquad \gamma^i_j = \frac{1}{2}\left(S^{+}_{ij} - S^{-}_{ij}\right) \tag{16}$$

For this case the definition of the χ^2 from Eq. 14 is extended with the parabolic approximation for asymmetric uncertainties, such that the expected cross section is adjusted to be

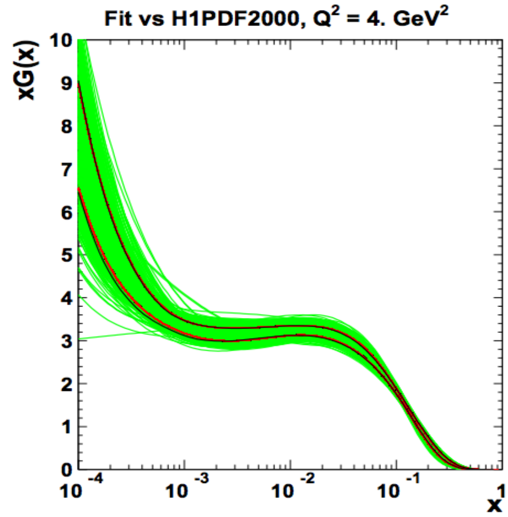

Fig. 9 Comparison between the standard error calculations as employed by the Hessian approach (black lines) and the MC approach (with more than 100 replicas) assuming Gaussian distribution for uncertainty distributions, shown here for each replica (green lines) together with the evaluated standard deviation (red lines) [32]. The black lines in the figure are mostly covered by the red lines.

$$m_i(1 - \sum_j \gamma^i_j b_j) \to m_i \left(1 - \sum_j b_j(\omega^i_j b_j + \gamma^i_j)\right). \tag{17}$$

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

5.4 Treatment of the Theoretical Input Parameters

The results of a QCD fit depend not only on the input data but also on the input parameters used by the theoretical calculations. Nowadays, recent 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 the value of the starting scale itself. `HERAFitter` provides possibility of different user choices of various input parameters of the theory.

5.5 Bayesian Reweighting Techniques

As an alternative to a complete QCD fit, the reweighting method (Bayesian Reweighting) is available in `HERAFitter`. The method provides a fast estimate of the impact of new data on PDFs. The original suggestion [90] was developed by the NNPDF collaboration [93, 94] and later extended [92] 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_{rep} parton distribution functions and observables $\mathscr{O}(\text{PDF})$ are conventionally calculated from the average of the predictions obtained from the ensemble:

$$\langle \mathscr{O}(\text{PDF}) \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \mathscr{O}(\text{PDF}_k). \quad (18)$$

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:

$$w_k = \frac{(\chi_k^2)^{\frac{1}{2}(N_{\text{data}}-1)} e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} (\chi_k^2)^{\frac{1}{2}(N_{\text{data}}-1)} e^{-\frac{1}{2}\chi_k^2}}, \quad (19)$$

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

$$\chi^2(y, \text{PDF}_k) = \sum_{i,j=1}^{N_{\text{data}}} (y_i - y_i(\text{PDF}_k)) \sigma_{ij}^{-1} (y_j - y_j(\text{PDF}_k)) \quad (20)$$

From all the resulting PDF replicas, those providing predictions incompatible with the measurements are discarded. Therefore, reweighted PDFs encompass less replicas than used in the input.

6 Alternatives to DGLAP formalism

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.

6.1 DIPOLE models

The dipole picture provides an alternative approach to the proton-virtual photon 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 [95]. 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 [28], the colour glass condensate approach to the high parton density regime called the Iancu-Itakura-Munier (IIM) dipole model [29] and a modified GBW model which takes into account the effects of DGLAP evolution called the Bartels-Golec-Kowalski (BGK) dipole model [30].

GBW model: In the GBW model the dipole-proton cross section σ_{dip} is given by

$$\sigma_{\text{dip}}(x, r^2) = \sigma_0 \left(1 - \exp\left[-\frac{r^2}{4R_0^2(x)}\right]\right), \quad (21)$$

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 [96]. The explicit formula for σ_{dip} can be found in [29]. 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

$$\sigma_{\text{dip}}(x, r^2) = \sigma_0 \left(1 - \exp\left[-\frac{\pi^2 r^2 \alpha_s(\mu^2) x g(x, \mu^2)}{3\sigma_0}\right]\right). \quad (22)$$

The factorisation 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 Eq. 8 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 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 5% to 15% for x from 0.0001 to 0.01 [97]. The new HERA F_2 measurements have a precision which is better than 2%. Therefore, in `HERAFitter` the contribution of the valence quarks can be taken from the PDF fits and added to the original BGK model [98, 99].

6.2 Transverse Momentum Dependent (Unintegrated) PDFs with CCFM

QCD calculations of multiple-scale processes and complex final-states require in general transverse-momentum dependent (TMD) [100], or unintegrated, parton density and parton decay functions [101–109]. TMD factorisation has been proven recently [100] for inclusive DIS. For special processes in hadron-hadron scattering, like heavy flavor or vector boson (including Higgs) production, TMD factorisation has also been proven in the high-energy limit (small x) [110–112]

In the framework of high-energy factorisation [110, 113, 114] the DIS cross section can be written as a convolution in both longitudinal and transverse momenta of the TMD parton density function $\mathcal{A}(x,k_t,\mu)$ with off-shell partonic matrix elements, as follows

$$\sigma_j(x,Q^2) = \int_x^1 dz \int d^2k_t\ \hat{\sigma}_j(x,Q^2,z,k_t)\ \mathcal{A}(z,k_t,\mu) \qquad (23)$$

with the DIS cross sections σ_j, $(j=2,L)$ related to the structure functions F_2 and F_L. The hard-scattering kernels $\hat{\sigma}_j$ of Eq. (23), are k_t-dependent and the evolution of the transverse momentum dependent gluon density $\mathcal{A}$ is obtained by combining the resummation of small-x logarithmic contributions [115–117] with medium-x and large-x contributions to parton splitting [6, 9, 10] according to the CCFM evolution equation [26, 118, 119].

The factorisation formula (23) allows resummation of logarithmically enhanced $x \to 0$ contributions to all orders in perturbation theory, both in the hard scattering coefficients and in the parton evolution, taking fully into account the dependence on the factorisation scale μ and on the factorisation scheme [120, 121].

The cross section σ_j, $(j=2,L)$ is calculated in a FFN scheme, where only the boson-gluon fusion process ($\gamma^* g^* \to q\bar{q}$) is included. The masses of the quarks are explicitly included with the light and heavy quark masses being free parameters. In addition to $\gamma^* g^* \to q\bar{q}$, the contribution from valence quarks is included via $\gamma^* q \to q$ as described later by using a CCFM evolution of valence quarks [122, 123].

CCFM Grid Techniques:
The CCFM evolution cannot be written easily in an analytic closed form. For this reason a Monte Carlo method is employed, which is however time-consuming, and cannot be used in a straightforward manner in a fit program. Following the convolution method introduced in [123, 124], the kernel $\tilde{\mathcal{A}}(x'',k_t,p)$ is determined from the Monte Carlo solution of the CCFM evolution equation, and then folded with the non-perturbative starting distribution $\mathcal{A}_0(x)$:

$$\begin{aligned} x\mathcal{A}(x,k_t,p) &= x\int dx' \int dx'' \mathcal{A}_0(x')\tilde{\mathcal{A}}(x'',k_t,p)\,\delta(x'x''-x) \\ &= \int dx' \mathcal{A}_0(x') \cdot \frac{x}{x'}\,\tilde{\mathcal{A}}\left(\frac{x}{x'},k_t,p\right) \end{aligned} \qquad (24)$$

with k_t being the transverse momentum of the propagator gluon and p being the evolution variable.
The kernel $\tilde{\mathcal{A}}$ incorporates all of the dynamics of the evolution. It is determined on a grid of $50 \otimes 50 \otimes 50$ bins in x, k_t, p. The binning in the grid is logarithmic, except for the longitudinal variable x where 40 bins in logarithmic spacing below 0.1, and 10 bins in linear spacing above 0.1 are used.
The calculation of the cross section according to Eq. (23) involves a multidimensional Monte Carlo integration which is time consuming and suffers from numerical fluctuations. This cannot be employed directly in a fit procedure involving the calculation of numerical derivatives in the search for the minimum. Instead the following equation is applied:

$$\begin{aligned} \sigma(x,Q^2) &= \int_x^1 dx_g \mathcal{A}(x_g,k_t,p)\hat{\sigma}(x,x_g,Q^2) \\ &= \int_x^1 dx' \mathcal{A}_0(x') \cdot \tilde{\sigma}(x/x',Q^2) \end{aligned} \qquad (25)$$

Here, first $\tilde{\sigma}(x',Q^2)$ is calculated numerically with a Monte Carlo integration on a grid in x for the values of Q^2 used in the fit. Then the last step in Eq.(25) is performed with a fast numerical gauss integration, which can be used in standard fit procedures.

Functional Forms for TMD parameterisation:
For the starting distribution $\mathcal{A}_0$, at the starting scale Q_0, the following form is used:

$$x\mathcal{A}_0(x,k_t) = Nx^{-B} \cdot (1-x)^C \left(1-Dx+E\sqrt{x}\right) \exp[-k_t^2/\sigma^2] \ , \qquad (26)$$

with $\sigma^2 = Q_0^2/2$ and the free parameters N, B, C, D, E.
Valence quarks are treated using the method of [122] as described in [123] with a starting distribution taken from any collinear PDF. At every scale p the flavor sum rule is fulfilled.
The TMD parton densities can be plotted either with `HERAFitter` provided tools or with TMDplotter [35].

7 Applications of `HERAFitter`

`HERAFitter` is an open source code and it can be downloaded from [1] together with its supporting documentation. A README file is provided within the package together with fast grid theory files (described in 4) which are associated with the properly formatted data files availabe in `HERAFitter`. The source code contains all the relevant information to perform QCD fits with HERA DIS data as a default set. The performance time depends on the fitting options and varies from 10 minutes (using 'FAST' techniques as described in 4) to several hours when full uncertainties are estimated. The HERAFitter code is a combination of C++ and Fortran 77 libraries with minimal dependencies, i.e. for the default fitting options no external dependences are required except QCDNUM evolution program [23] and CERN libs. The ROOT libaries are only required for the drawing tools and when invoking APPLGRID . There are also cache options, fast evolution kernels, and usage of the OpenMP (Open Multi-Processing) interface which allows parallel applications of the GM-VFNS theory predictions in DIS. In addition, the HERAFitter references and GNU public licence are provided together with the main source code.

For the following LHC analyses of SM processes the HERAFitter package was used: inclusive Drell-Yan and W and Z production [11, 13, 14], inclusive jets [12, 15] production. At HERA, the results of QCD analyses using HERAFitter are published for the inclusive H1 measurements [16] and the recent combination of charm production measurements in DIS [17]. A determination of the transverse momentum dependent gluon density using precision HERA data obtained with HERAFitter has been reported in [125].

The HERAFitter platform has been already used to produce PDF grids from the QCD analyses performed at HERA [36] and at the LHC, using measurements from ATLAS [11, 12] (ATLAS PDF sets [38]) which can be used to study predictions for SM or beyond SM processes. Moreover, HERAFitter provides a possibility to perform impact studies for possible future colliders as demonstrated by the QCD studies at the LHeC [126].

Recently a study based on a set of parton distribution functions determined with the HERAFitter program using HERA data was performed [127]. 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.

8 Summary

The HERAFitter project is a unique platform for QCD analyses to study the structure of the proton. The project successfully encapsulates a wide variety of QCD tools to facilitate investigations of the experimental data and theoretical calculations. HERAFitter is the first open source platform which is optimal for benchmarking studies. It allows for direct comparisons of various theoretical approaches under the same settings, a variety of different methodologies in treating of the experimental and model uncertainties. The growth of HERAFitter benefits from its flexible modular structure driven by QCD advances.

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, RFBR grant 12-02-91526-CERN a and a dedicated funding of the Initiative and Networking Fond of Helmholtz Association SO-072. We also 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. *HERAFitter*, https://www.herafitter.org.
2. G. Aad *et al.* [ATLAS Collaboration], Phys.Lett. **B716**, 1 (2012), [1207.7214].
3. S. Chatrchyan *et al.* [CMS Collaboration], Phys.Lett. **B716**, 30 (2012), [1207.7235].
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. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).
7. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 675 (1972).
8. L. N. Lipatov, Sov. J. Nucl. Phys. **20**, 94 (1975).
9. Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
10. G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
11. G. Aad *et al.* [ATLAS Collaboration], Phys. Rev. Lett. **109**, 012001 (2012), [arXiv:1203.4051].
12. G. Aad *et al.* [ATLAS Collaboration], Eur.Phys.J. **73**, 2509 (2013), [arXiv:1304:4739].
13. G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. **B725**, 223 (2013), [arXiv::1305.4192].
14. S. Chatrchyan *et al.* [CMS Collaboration], submitted to Phys. Rev. **D** (2014), [arXiv:1312.6283].
15. S. Chatrchyan *et al.* [CMS Collaboration], CMS PAS **SMP-12-028** (2014).
16. F. Aaron *et al.* [H1 Collaboration], JHEP **1209**, 061 (2012), [arXiv:1206.7007].
17. H. Abramowicz *et al.* [H1 and ZEUS Collaborations], Eur. Phys. J. **C73**, 2311 (2013), [arXiv:1211.1182].
18. A. Martin, W. Stirling, R. Thorne, and G. Watt, Eur. Phys. J. C **63**, 189 (2009), [arXiv:0901.0002], URL `http://mstwpdf.hepforge.org/`.
19. 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/`.
20. 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/`.

21. S. Alekhin, J. Bluemlein, and S. Moch (2013), [1310.3059].
22. P. Jimenez-Delgado and E. Reya, Phys.Rev. **D80**, 114011 (2009), [0909.1711], URL `http://www.het.physik.tu-dortmund.de/pdfserver/index.html`.
23. M. Botje (2010), http://www.nikef.nl/h24/qcdnum/index.html, [arXiv:1005.1481].
24. M. Ciafaloni, Nucl. Phys. B **296**, 49 (1988).
25. S. Catani, F. Fiorani, and G. Marchesini, Phys. Lett. B **234**, 339 (1990).
26. S. Catani, F. Fiorani, and G. Marchesini, Nucl. Phys. B **336**, 18 (1990).
27. G. Marchesini, Nucl. Phys. B **445**, 49 (1995).
28. K. Golec-Biernat and M. Wüsthoff, Phys. Rev. D **59**, 014017 (1999), [hep-ph/9807513].
29. E. Iancu, K. Itakura, and S. Munier, Phys. Lett. **B590**, 199 (2004), [hep-ph/0310338].
30. J. Bartels, K. Golec-Biernat, and H. Kowalski, Phys. Rev. D **66**, 014001 (2002), [hep-ph/0203258].
31. F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).
32. M. Dittmar, S. Forte, A. Glazov, and S. Moch (2009), Altarelli, G. and others (contributing authors), [arXiv:0901.2504].
33. M. R. Whalley, D. Bourilkov, and R. Group (2005), [hep-ph/0508110].
34. *LHAPDF*, URL `http://lhapdf.hepforge.org`.
35. [TMD Collaboration], to be published.
36. F. Aaron *et al.* [H1 and ZEUS Collaborations], JHEP **1001**, 109 (2010), [arXiv:0911.0884].
37. *HERAPDF1.5LO, NLO and NNLO* (H1prelim-13-141 and ZEUS-prel-13-003, H1prelim-10-142 and ZEUS-prel-10-018, H1prelim-11-042 and ZEUS-prel-11-002), available via: http://lhapdf.hepforge.org/pdfsets.
38. *ATLAS NNLO epWZ12*, available via: http://lhapdf.hepforge.org/pdfsets.
39. R. Devenish and A. Cooper-Sarkar (2011), *Deep Inelastic Scattering*, ISBN: 0199602255,9780199602254.
40. J. C. Collins and W.-K. Tung, Nucl. Phys. B **278**, 934 (1986).
41. E. Laenen *et al.*, Phys. Lett. **B291**, 325 (1992).
42. E. Laenen *et al.*, Nucl. Phys. **B392**, 162, 229 (1993).
43. S. Riemersma, J. Smith, and van Neerven. W.L., Phys. Lett. **B347**, 143 (1995), [hep-ph/9411431].
44. S. Alekhin, *OPENQCDRAD*, a program description and the code are available via: http://www-zeuthen.desy.de/~alekhin/OPENQCDRAD.
45. H. Kawamura, N. Lo Presti, S. Moch, and A. Vogt, Nucl.Phys. **B864**, 399 (2012).
46. S. Alekhin and S. Moch, Phys. Lett. **B699**, 345 (2011), [arXiv:1011.5790].
47. R. Demina, S. Keller, M. Kramer, S. Kretzer, R. Martin, *et al.* (1999), [hep-ph/0005112].
48. R. S. Thorne and R. G. Roberts, Phys. Rev. D **57**, 6871 (1998), [hep-ph/9709442].
49. R. S. Thorne, Phys. Rev. **D73**, 054019 (2006), [hep-ph/0601245].
50. R. S. Thorne, Phys. Rev. D **86**, 074017 (2012), [arXiv:1201.6180].
51. J. C. Collins, Phys.Rev. **D58**, 094002 (1998), [hep-ph/9806259].
52. M. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, Phys.Rev. **D50**, 3102 (1994), [hep-ph/9312319].
53. M. Kramer, F. I. Olness, and D. E. Soper, Phys. Rev. **D62**, 096007 (2000), [hep-ph/0003035].
54. S. Kretzer, H. Lai, F. Olness, and W. Tung, Phys.Rev. **D69**, 114005 (2004), [hep-ph/0307022].
55. H. Spiesberger, Private communication.
56. F. Jegerlehner, Proceedings, LC10 Workshop **DESY 11-117** (2011).
57. H. Burkhard, F. Jegerlehner, G. Penso, and C. Verzegnassi, in CERN Yellow Report on ”Polarization at LEP” 1988.
58. S. Chekanov *et al.* [ZEUS Collaboration], Nucl. Phys. **B831**, 1 (2010), [hep-ex/09114119].
59. S. D. Drell and T.-M. Yan, Phys. Rev. Lett. **25**, 316 (1970).
60. M. Yamada and M. Hayashi, Nuovo Cim. **A70**, 273 (1982).
61. J. M. Campbell and R. K. Ellis, Phys. Rev. **D60**, 113006 (1999), [arXiv:9905386].
62. J. M. Campbell and R. K. Ellis, Phys. Rev. **D62**, 114012 (2000), [arXiv:0006304].
63. J. M. Campbell and R. K. Ellis, Nucl. Phys. Proc. Suppl. **205-206**, 10 (2010), [arXiv:1007.3492].
64. Y. Li and F. Petriello, Phys.Rev. **D86**, 094034 (2012), [arXiv:1208.5967].
65. G. Bozzi, J. Rojo, and A. Vicini, Phys.Rev. **D83**, 113008 (2011), [arXiv:1104.2056].
66. A. Gehrmann-De Ridder, T. Gehrmann, E. Glover, and J. Pires, Phys. Rev. Lett. **110**, 162003 (2013), [arXiv:1301.7310].
67. E. Glover and J. Pires, JHEP **1006**, 096 (2010), [arXiv:1003.2824].
68. J. Currie, A. Gehrmann-De Ridder, E. Glover, and J. Pires, JHEP **1401**, 110 (2014), [1310.3993].
69. Z. Nagy and Z. Trocsanyi, Phys.Rev. **D59**, 014020 (1999), [hep-ph/9806317].
70. Z. Nagy, Phys.Rev.Lett. **88**, 122003 (2002), [hep-ph/0110315].
71. S. Chatrchyan *et al.* [CMS Collaboration], Phys.Lett. **B728**, 496 (2014), [1307.1907].
72. M. Czakon, P. Fiedler, and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013), [1303.6254].

73. M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, *et al.*, Comput.Phys.Commun. **182**, 1034 (2011), [arXiv:1007.1327].
74. J. M. Campbell, R. Frederix, F. Maltoni, and F. Tramontano, Phys.Rev.Lett. **102**, 182003 (2009), [0903.0005].
75. J. M. Campbell and F. Tramontano, Nucl.Phys. **B726**, 109 (2005), [hep-ph/0506289].
76. J. M. Campbell, R. K. Ellis, and F. Tramontano, Phys.Rev. **D70**, 094012 (2004), [hep-ph/0408158].
77. J. M. Campbell and R. K. Ellis (2012), report FERMILAB-PUB-12-078-T, [1204.1513].
78. T. Kluge, K. Rabbertz, and M. Wobisch, pp. 483–486 (2006), [hep-ph/0609285].
79. T. Carli *et al.*, Eur. Phys. J. **C66**, 503 (2010), [arXiv:0911.2985].
80. Z. Nagy and Z. Trocsanyi, Phys.Rev.Lett. **87**, 082001 (2001), [hep-ph/0104315].
81. Z. Nagy, Phys.Rev. **D68**, 094002 (2003), [hep-ph/0307268].
82. M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober [fastNLO Collaboration] (2011), [arXiv:1109.1310].
83. N. Kidonakis and J. Owens, Phys.Rev. **D63**, 054019 (2001), [hep-ph/0007268].
84. D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch [fastNLO Collaboration] (2012), [arXiv:1208.3641].
85. http://fastnlo.hepforge.org, URL `http://fastnlo.hepforge.org`.
86. http://applgrid.hepforge.org, URL `http://applgrid.hepforge.org`.
87. A. Glazov, S. Moch, and V. Radescu, Phys. Lett. B **695**, 238 (2011), [arXiv:1009.6170].
88. J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, *et al.*, Phys.Rev. **D65**, 014013 (2001), [hep-ph/0101032].
89. M. Botje, J.Phys. **G28**, 779 (2002), [hep-ph/0110123].
90. W. T. Giele and S. Keller, Phys.Rev. **D58**, 094023 (1998), [hep-ph/9803393].
91. W. T. Giele, S. Keller, and D. Kosower (2001), [hep-ph/0104052].
92. G. Watt and R. Thorne, JHEP **1208**, 052 (2012), [arXiv:1205.4024].
93. R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, *et al.*, Nucl.Phys. **B855**, 608 (2012), [arXiv:1108.1758].
94. R. D. Ball *et al.* [NNPDF Collaboration], Nucl.Phys. **B849**, 112 (2011), [arXiv:1012.0836].
95. N. N. Nikolaev and B. Zakharov, Z.Phys. **C49**, 607 (1991).
96. I. Balitsky, Nucl. Phys. B **463**, 99 (1996), [hep-ph/9509348].
97. F. Aaron *et al.* [H1 Collaboration], Eur.Phys.J. **C71**, 1579 (2011), [1012.4355].
98. P. Belov, Doctoral thesis, Universität Hamburg (2013), [DESY-THESIS-2013-017].
99. A. Luszczak and H. Kowalski (2013), [1312.4060].
100. J. Collins, *Foundations of perturbative QCD*, vol. 32 (Cambridge monographs on particle physics, nuclear physics and cosmology., 2011).
101. S. M. Aybat and T. C. Rogers, Phys.Rev. **D83**, 114042 (2011), [1101.5057].
102. M. Buffing, P. Mulders, and A. Mukherjee, Int.J.Mod.Phys.Conf.Ser. **25**, 1460003 (2014), [1309.2472].
103. M. Buffing, A. Mukherjee, and P. Mulders, Phys.Rev. **D88**, 054027 (2013), [1306.5897].
104. M. Buffing, A. Mukherjee, and P. Mulders, Phys.Rev. **D86**, 074030 (2012), [1207.3221].
105. P. Mulders, Pramana **72**, 83 (2009), [0806.1134].
106. S. Jadach and M. Skrzypek, Acta Phys.Polon. **B40**, 2071 (2009), [0905.1399].
107. F. Hautmann, Acta Phys.Polon. **B40**, 2139 (2009).
108. F. Hautmann, M. Hentschinski, and H. Jung (2012), [1205.6358].
109. F. Hautmann and H. Jung, Nucl.Phys.Proc.Suppl. **184**, 64 (2008), [0712.0568].
110. S. Catani, M. Ciafaloni, and F. Hautmann, Phys. Lett. B **242**, 97 (1990).
111. J. C. Collins and R. K. Ellis, Nucl. Phys. B **360**, 3 (1991).
112. F. Hautmann, H. Jung, and V. Pandis, AIP Conf.Proc. **1350**, 263 (2011), [1011.6157].
113. S. Catani, M. Ciafaloni, and F. Hautmann, Nucl. Phys. B **366**, 135 (1991).
114. S. Catani, M. Ciafaloni, and F. Hautmann, Phys. Lett. B **307**, 147 (1993).
115. L. Lipatov, Phys.Rept. **286**, 131 (1997), [hep-ph/9610276].
116. V. S. Fadin, E. Kuraev, and L. Lipatov, Phys.Lett. **B60**, 50 (1975).
117. I. I. Balitsky and L. N. Lipatov, Sov. J. Nucl. Phys. **28**, 822 (1978).
118. M. Ciafaloni, Nucl. Phys. **B296**, 49 (1988).
119. G. Marchesini, Nucl. Phys. B **445**, 49 (1995), [hep-ph/9412327].
120. S. Catani and F. Hautmann, Nucl. Phys. B **427**, 475 (1994), [hep-ph/9405388].
121. S. Catani and F. Hautmann, Phys.Lett. **B315**, 157 (1993).
122. M. Deak, F. Hautmann, H. Jung, and K. Kutak, *Forward-Central Jet Correlations at the Large Hadron Collider* (2010), [arXiv:1012.6037].
123. F. Hautmann and H. Jung, Nuclear Physics B **883**, 1 (2014), [1312.7875].

124. H. Jung and F. Hautmann (2012), [arXiv:1206.1796].
125. F. Hautmann and H. Jung (2013), [1312.7875].
126. J. L. Abelleira Fernandez *et al.* [LHeC Study Group], Journal of Phys. **G**, 075001 (2012), [arXiv:1206.2913].
127. HERAFitter Developers Team and M. Lisovyi (2014), [arXiv:1404.4234].