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

# HERAFitter

## Open Source QCD Fit Project

**Version 0.99 (svn - 1638, post draft v3 circulation)**

**S. Alekhin[1,2], O. Behnke[3], P. Belov[3,4], S. Borroni[3], M. Botje[5], D. Britzger[3], S. Camarda[3], A.M. Cooper-Sarkar[6], K. Daum[7,8], C. Diaconu[9], J. Feltesse[10], A. Gizhko[3], A. Glazov[3], A. Guffanti[11], M. Guzzi[3], F. Hautmann[12,13,14], A. Jung[15], H. Jung[3,16], V. Kolesnikov[17], H. Kowalski[3], O. Kuprash[3], A. Kusina[18], S. Levonian[3], K. Lipka[3], B. Lobodzinski[19], K. Lohwasser[1,3], A. Luszczak[20], B. Malaescu[21], R. McNulty[22], V. Myronenko[3], S. Naumann-Emme[3], K. Nowak[3,6], F. Olness[18], E. Perez[23], H. Pirumov[3], R. Plačakytė[3], K. Rabbertz[24], V. Radescu[3], R. Sadykov[17], G.P. Salam[25,26], A. Sapronov[17], A. Schöning[27], T. Schörner-Sadenius[3], S. Shushkevich[3], W. Slominski[28], H. Spiesberger[29], P. Starovoitov[3], M. Sutton[30], J. Tomaszewska[31], O. Turkot[3], A. Vargas[3], G. Watt[32], K. Wichmann[3]**

[1] Deutsches Elektronen-Synchrotron (DESY), Platanenallee 6, D–15738 Zeuthen, Germany
[2] Institute for High Energy Physics,142281 Protvino, Moscow region, Russia
[3] Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany
[4] Current address: Department of Physics, St. Petersburg State University, Ulyanovskaya 1, 198504 St. Petersburg, Russia
[5] Nikhef, Science Park, Amsterdam, the Netherlands
[6] Department of Physics, University of Oxford, Oxford, United Kingdom
[7] Fachbereich C, Universität Wuppertal, Wuppertal, Germany
[8] Rechenzentrum, Universität Wuppertal, Wuppertal, Germany
[9] CPPM, IN2P3-CNRS, Univ. Mediterranee, Marseille, France
[10] CEA, DSM/Irfu, CE-Saclay, Gif-sur-Yvette, France
[11] Niels Bohr Institute, University of Copenhagen, Denmark
[12] School of Physics and Astronomy, University of Southampton, UK
[13] Rutherford Appleton Laboratory, Chilton OX11 0QX, United Kingdom
[14] Dept. of Theoretical Physics, University of Oxford, Oxford OX1 3NP, United Kingdom
[15] FERMILAB, Batavia, IL, 60510, USA
[16] Elementaire Deeltjes Fysica, Universiteit Antwerpen, B 2020 Antwerpen, Belgium
[17] Joint Institute for Nuclear Research (JINR), Joliot-Curie 6, 141980, Dubna, Moscow Region, Russia
[18] Southern Methodist University, Dallas, Texas
[19] Max Planck Institut Für Physik, Werner Heisenberg Institut, Föhringer Ring 6, Muńchen
[20] T. Kosciuszko Cracow University of Technology
[21] Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université, Paris-Diderot and CNRS/IN2P3, Paris, France
[22] University College Dublin, Dublin 4, Ireland
[23] CERN, European Organization for Nuclear Research, Geneva, Switzerland
[24] Institut für Experimentelle Kernphysik, Karlsruhe, Germany
[25] CERN, PH-TH, CH-1211 Geneva 23, Switzerland
[26] leave from LPTHE; CNRS UMR 7589; UPMC Univ. Paris 6; Paris 75252, France
[27] Physikalisches Institut, Universität Heidelberg, Heidelberg, Germany
[28] Jagiellonian University, Institute of Physics, Reymonta 4, PL-30-059 Cracow, Poland
[29] PRISMA Cluster of Excellence, Institut für Physik (WA THEP), Johannes-Gutenberg-Universität, D-55099 Mainz, Germany
[30] University of Sussex, Department of Physics and Astronomy, Sussex House, Brighton BN1 9RH, United Kingdom
[31] Warsaw University of Technology, Faculty of Physics, Koszykowa 75, 00-662 Warsaw, Poland
[32] Institute for Particle Physics Phenomenology, Durham University, Durham, DH1 3LE, United Kingdom

Received: date / Accepted: date

**Abstract** `HERAFitter` [1] is an open-source package that provides a framework for the determination of the parton distribution functions (PDFs) of the proton and for many different kinds of analyses in Quantum Chromodynamics (QCD). It encodes results from a wide range of experimental measurements in lepton-proton deep inelastic scattering and proton-proton (proton-antiproton) collisions at hadron colliders. Those are complemented with a variety of theoretical options for calculating PDF-dependent cross section predictions corresponding to the measurements. The data and theoretical predictions are brought together through numerous methodological options for carrying out PDF fits and plotting tools to help visualise the results. While primarily based on the approach of collinear factorisation, `HERAFitter` also provides facilities for fits of dipole models and transverse-momentum dependent PDFs. This paper describes the general structure of `HERAFitter` and its wide choice of options.

**Keywords** PDFs · QCD · Fit · proton structure

## Contents



## 1 Introduction

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 calculations and computations to test the validity of the Standard Model (SM) and factorisation in Quantum Chromodynamics (QCD). Using collinear factorisation, hadron inclusive cross sections may be written as

$$\sigma(\alpha_s(\mu_R^2), \mu_R^2, \mu_F^2) = \sum_{a,b} \int_0^1 dx_1\, dx_2 f_a(x_1, \mu_F^2) f_b(x_2, \mu_F^2) \\ \times\ \hat{\sigma}^{ab}(x_1, x_2; \alpha_s(\mu_R^2), \mu_R^2, \mu_F^2) \\ +\ \mathcal{O}\left(\frac{\Lambda_{QCD}^2}{Q^2}\right) \tag{1}$$

where the cross section $\sigma$ is expressed as a convolution of Parton Distribution Functions (PDFs) $f_a$ and $f_b$ with the parton cross section $\hat{\sigma}^{ab}$, involving a momentum transfer $q$ such that $Q^2 = |q^2| \gg \Lambda_{QCD}^2$. At Leading-Order (LO), 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. The indices $a$ and $b$ in 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. The PDFs depend on the factorisation scale, $\mu_F$, while the parton cross sections depend on the strong coupling, $\alpha_s$, and the factorisation and renormalisation scales, $\mu_F$ and $\mu_R$. The parton cross sections $\hat{\sigma}^{ab}$ are calculable in perturbative QCD (pQCD) whereas PDFs are non-perturbative and are usually constrained by global fits to a variety of experimental data. The assumption that PDFs are universal, within a particular factorisation scheme [4–8], is crucial to this procedure. Recent review articles on PDFs can be found in Refs. [9, 10].

A precise determination of PDFs as a function of $x$ requires large amounts of experimental data that cover a wide kinematic region and that are sensitive to different kinds of partons. Measurements of inclusive Neutral Current (NC) and Charge Current (CC) Deep Inelastic Scattering (DIS) at the lepton-proton ($ep$) collider HERA provide crucial information for determining the PDFs. Different processes in proton-proton ($pp$) and proton-antiproton ($p\bar{p}$) collisions at the LHC and the Tevatron, respectively, provide complementary information to the DIS measurements. The PDFs are determined from $\chi^2$ fits of the theoretical predictions to the data [11–15]. The rapid flow of new data from the LHC experiments and the corresponding theoretical developments, which are providing predictions for more complex processes at increasingly higher orders, has motivated the development of a tool to combine them together in a fast, efficient, open-source platform.

This paper describes the open-source QCD fit platform HERAFitter, which includes a set of tools to facilitate global QCD analyses of $pp$, $p\bar{p}$ and $ep$ scattering data. It has been developed for the determination of PDFs and the extraction of fundamental parameters of QCD such as the heavy quark masses and the strong coupling constant. It also provides a common platform for the comparison of different theoretical approaches. Furthermore, it can be used to test the impact of new experimental data on the PDFs and on the SM parameters.

This paper is organised as follows: The general structure of HERAFitter is presented in section 2. In section 3 the various processes available in HERAFitter and the corresponding theoretical calculations, performed within the framework of collinear factorisation and the DGLAP [16–20] formalism, are discussed. In section 4 tools for fast calculations of the theoretical predictions are presented. In section 5 the methodology to determine PDFs through fits based on various $\chi^2$ definitions is explained. In particular, different treatments of correlated experimental uncertainties are presented. Alternative approaches to the DGLAP formalism are presented in section 6. The organisation of the HERAFitter code is discussed in section 7, specific applications of the package are persented in section 8, which is followed by a summary in section 9.

## 2 The HERAFitter Structure

The diagram in Fig. 1 gives a schematic overview of the HERAFitter structure and functionality, which can be divided into four main blocks:

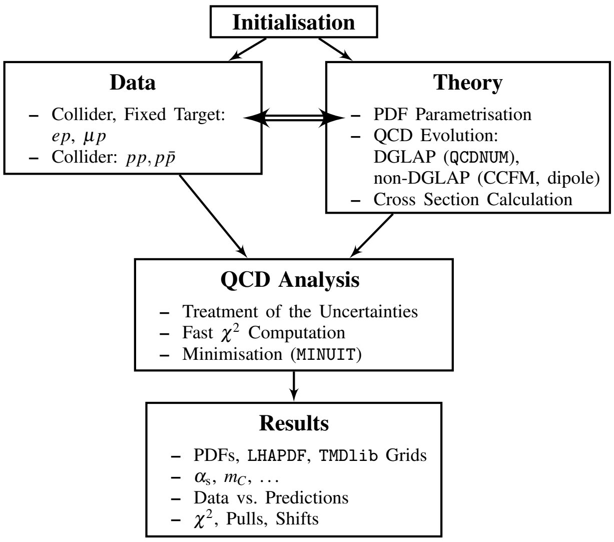


**Fig. 1** Schematic overview of the HERAFitter program.

| Experimental Data | Process | Reaction | Theory calculations, schemes |
|---|---|---|---|
| HERA, Fixed Target | DIS NC | $ep \to eX$<br>$\mu p \to \mu X$ | TR′, ACOT, ZM (QCDNUM), FFN (OPENQCDRAD, QCDNUM), TMD (uPDFevolv) |
| HERA | DIS CC | $ep \to \nu_e X$ | ACOT, ZM (QCDNUM), FFN (OPENQCDRAD) |
| | DIS jets | $ep \to e$ jets$X$ | NLOJet++ (fastNLO) |
| | DIS heavy quarks | $ep \to ec\bar{c}X$,<br>$ep \to eb\bar{b}X$ | TR′, ACOT, ZM (QCDNUM), FFN (OPENQCDRAD, QCDNUM) |
| Tevatron, LHC | Drell-Yan | $pp(\bar{p}) \to l\bar{l}X$,<br>$pp(\bar{p}) \to l\nu X$ | MCFM (APPLGRID) |
| | top pair | $pp(\bar{p}) \to t\bar{t}X$ | MCFM (APPLGRID), HATHOR, DiffTop |
| | single top | $pp(\bar{p}) \to tl\nu X$,<br>$pp(\bar{p}) \to tX$,<br>$pp(\bar{p}) \to tWX$ | 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 experimental data and theory calculations implemented in the HERAFitter package. The references for the individual calculations and schemes are given in the text.

*Data:* Measurements from various processes are implemented in the HERAFitter package including the full information on their uncorrelated and correlated uncertainties. HERA inclusive scattering data are sensitive to quark and to gluon PDFs through scaling violations and the longitudinal structure function $F_L$. These data are the backbone of any proton PDF extraction, and are used by all global PDF groups [11–15]. They can be supplemented by HERA measurements sensitive to heavy quarks and by HERA jet measurements, which have direct sensitivity to the gluon PDF. However, the kinematic range of HERA data mostly covers low and medium ranges in $x$. 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, measurements from fixed-target experiments, the Tevatron and the LHC can be used. The processes that are currently available within the HERAFitter framework are listed in Tab. 1.

*Theory:* The PDFs are parametrised at a starting 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 $Q^2$, $Q^2 > Q_0^2$. By default, the evolution uses the DGLAP formalism [16–20] as implemented in QCDNUM [21]. Alternatively, the CCFM evolution [22–25] as implemented in uPDFevolv [26] can be chosen. The prediction of the cross section for a particular process is obtained, assuming factorisation, by the convolution of the evolved PDFs with the corresponding hard-process parton scattering cross section. Available theory calculations are listed in Tab. 1. Predictions using dipole models [27–29] can also be obtained.

*QCD Analysis:* The PDFs are determined in a least squares fit, minimising a $\chi^2$ function that is constructed from the input data and theory predictions, with the `MINUIT` [30] program. In `HERAFitter` various choices are available for the treatment of experimental uncertainties. Correlated experimental uncertainties can be accounted for using a nuisance parameter method or a covariance matrix method as described in section 5.2. Different statistical assumptions for the distributions of the systematic uncertainties, e.g. Gaussian or LogNormal [31], can also be studied (see section 5.3).

*Results:* The resulting PDFs are provided in a format ready to be used by the `LHAPDF` library [32, 33] or by `TMDlib` [34]. `HERAFitter` drawing tools can be used to display the PDFs with their uncertainties at a chosen scale. As an example, the first set of PDFs extracted using `HERAFitter` from HERA I data, HERAPDF1.0 [35], is shown in Fig. 2 (taken from Ref. [35]). Note that following conventions, the PDFs are displayed as parton momentum distributions $xf(x,\mu_F^2)$.

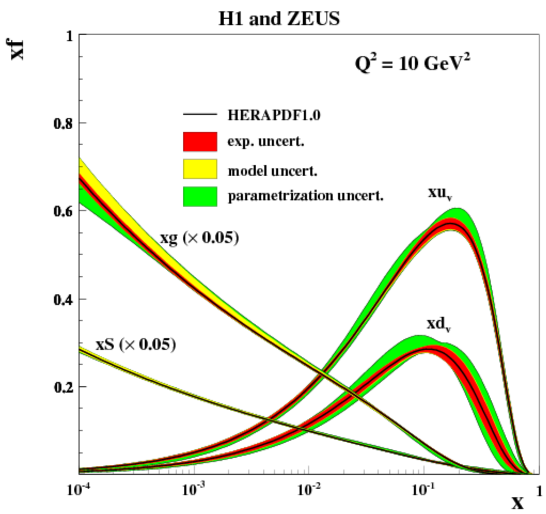


**Fig. 2** Distributions of valence ($xu_v$, $xd_v$), sea ($xS$) and the gluon ($g$) PDFs in HERAPDF1.0 [35]. The gluon and the sea distributions are scaled down by a factor of 20. The experimental, model and parametrisation uncertainties are shown as coloured bands.

## 3 Theoretical formalism using DGLAP evolution

In this section the theoretical formalism based on DGLAP [16–20] evolution is described.

A direct consequence of factorisation (Eq. 1) is that the scale dependence or "evolution" of the PDFs can be predicted by the renormalisation group equations. By requiring physical observables to be independent of $\mu_\mathrm{F}$, a representation of the parton evolution in terms of the DGLAP equations is obtained:

$$\frac{d\, f_a(x,\mu_\mathrm{F}^2)}{d\log\mu_\mathrm{F}^2} = \sum_{b=q,\bar{q},g} \int_x^1 \frac{dz}{z} P_{ab}\left(\frac{x}{z};\mu_\mathrm{F}^2\right) f_b(z,\mu_\mathrm{F}^2)\,, \tag{2}$$

where the functions $P_{ab}$ are the evolution kernels or splitting functions, which represent the probability of finding parton $a$ in parton $b$. They can be calculated as a perturbative expansion in $\alpha_s$. Once PDFs are determined at the initial scale $\mu_F^2 = Q_0^2$, their evolution to any other scale $Q^2 > Q_0^2$ is entirely determined by the DGLAP equations. The PDFs are then used to calculate cross sections for various different processes. Alternative approaches to DGLAP evolution, valid in different kinematic regimes, are also implemented in `HERAFitter` and will be discussed in section 6.

### 3.1 Deep Inelastic Scattering and Proton Structure

The formalism that relates the DIS measurements to pQCD and the PDFs has been described in detail in many extensive reviews (see e.g. Ref. [36]) and it is only briefly summarised here. DIS is the process where a lepton scatters off the partons in the proton by a virtual exchanged of a neutral ($\gamma/Z$) or charged ($W^\pm$) vector boson and, as a result, a scattered lepton and a multi-hadronic final state are produced. The common DIS kinematic variables are the scale of the process $Q^2$, which is the absolute squared four-momentum of the exchange boson, Bjorken $x$, which can be related in the parton model to the momentum fraction that is carried by the struck quark, and the inelasticity $y$. These are 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:

$$\frac{d^2\sigma^{e^\pm p}_{NC}}{dxdQ^2} = \frac{2\pi\alpha^2 Y_+}{xQ^4}\sigma^{e^\pm p}_{r,NC}, \tag{3}$$

$$\sigma^{e^\pm p}_{r,NC} = \tilde{F}_2^\pm \mp \frac{Y_-}{Y_+} x\tilde{F}_3^\pm - \frac{y^2}{Y_+}\tilde{F}_L^\pm, \tag{4}$$

where $Y_\pm = 1 \pm (1-y)^2$ and $\alpha$ is the electromagnetic coupling constant. The generalised structure functions $\tilde{F}_{2,3}$ can be written as linear combinations of the proton structure functions $F_2^\gamma, F_{2,3}^{\gamma Z}$ and $F_{2,3}^Z$, which are associated with pure photon exchange terms, photon-$Z$ interference terms and pure $Z$ exchange terms, respectively. The 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

$y$. In the framework of pQCD the structure functions are directly related to the PDFs, i.e. at leading order (LO) $F_2$ is the weighted momentum sum of quark and anti-quark distributions, $xF_3$ is related to their difference, and $F_L$ vanishes. At higher orders, terms related to the gluon distribution appear, in particular $F_L$ is strongly related to the low-$x$ gluon.
The inclusive CC $ep$ cross section, analogous to the NC $ep$ case, can be expressed in terms of another set of structure functions, $\tilde{W}$:

$$\frac{d^2\sigma^{e^\pm p}_{CC}}{dxdQ^2} = \frac{1\pm P}{2}\frac{G_F^2}{2\pi x}\left[\frac{m_W^2}{m_W^2+Q^2}\right]\sigma^{e^\pm p}_{r,CC} \tag{5}$$

$$\sigma^{e^\pm p}_{r,CC} = Y_+\tilde{W}_2^\pm \mp Y_- x\tilde{W}_3^\pm - y^2\tilde{W}_L^\pm, \tag{6}$$

where $P$ represents the lepton beam polarisation. At LO in $\alpha_s$, the CC $e^+p$ and $e^-p$ cross sections are sensitive to different combinations of the quark flavour densities.

Beyond LO, the QCD predictions for the DIS structure functions are obtained by convoluting the PDFs with appropriate hard-process scattering matrix elements, which are referred to as coefficient functions.

The DIS measurements span a large range of $Q^2$ from a few GeV$^2$ to about $10^5$ GeV$^2$, crossing heavy quark mass thresholds, thus the treatment of heavy quark (charm and beauty) production and the chosen values of their masses become important. There are different schemes for the treatment of heavy quark production. Several variants of these schemes are implemented in `HERAFitter` and they are briefly discussed below.

*Zero-Mass Variable Flavour Number (ZM-VFN)*:
In this scheme [37], the heavy quarks appear as partons in the proton at $Q^2$ values above $\sim m_h^2$ (heavy quark mass) and they are then treated as massless in both the initial and final states of the hard scattering process. The lowest order process is the scattering of the lepton off the heavy quark via electroweak boson exchange. This scheme is expected to be reliable in the region where $Q^2 \gg m_h^2$. In `HERAFitter` this scheme is available for the DIS structure function calculation via the interface to the `QCDNUM` [21] package, thus it benefits from the fast `QCDNUM` convolution engine.

*Fixed Flavour Number (FFN)*:
In this rigorous quantum field theory scheme [38–40], 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 heavy quark-antiquark pair production via boson-gluon fusion. In `HERAFitter` this scheme can be accessed via the `QCDNUM` implementation or through the interface to the open-source code `OPENQCDRAD` [41] as implemented by the ABM group. This scheme is reliable for $Q^2 \sim m_h^2$. 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 `OPENQCDRAD` implementation the heavy quark contributions to CC structure functions are also available and, for the NC case, the QCD corrections to the coefficient functions in Next-to-Next-to Leading Order (NNLO) are provided in the best currently known approximation [42]. The `OPENQCDRAD` implementation uses the running heavy quark mass in the $\overline{\text{MS}}$ scheme [43].

It is sometimes argued that this scheme reduces the sensitivity of the DIS cross sections to higher order corrections [42]. It is also known to have smaller non-perturbative corrections than the pole mass scheme [44].

*General-Mass Variable Flavour Number (GM-VFN)*:
In this scheme [45], heavy quark production is treated for $Q^2 \sim m_h^2$ in the FFN scheme and for $Q^2 \gg m_h^2$ in the massless scheme with a suitable interpolation in between. The details of this interpolation differ between implementations. The PDF groups that use GM-VFN schemes are MSTW, CT (CTEQ), NNPDF, and HERAPDF. `HERAFitter` implements different variants of the GM-VFN scheme.

- *GM-VFN Thorne-Roberts scheme:* The Thorne-Roberts (TR) scheme [46] was designed to provide a smooth transition from the massive FFN scheme at low scales $Q^2 \sim m_h^2$ to the massless ZM-VFNS scheme at high scales $Q^2 \gg m_h^2$. Because the original version was technically difficult to implement beyond NLO, it was updated to the TR$'$ scheme [47]. There are two variants of the TR$'$ schemes: TR$'$ standard (as used in MSTW PDF sets [11, 47]) and TR$'$ optimal [48], with a smoother transition across the heavy quark threshold region. Both TR$'$ 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 renormalisation method of Collins-Wilczek-Zee (CWZ) [49]. This scheme unifies the low scale $Q^2 \sim m_h^2$ and high scale $Q^2 > m_h^2$ regions in a coherent framework across the full energy range. Within the ACOT package, the following variants of the ACOT scheme are available: ACOT-Full [50], S-ACOT-$\chi$ [51, 52], ACOT-ZM [50], $\overline{\text{MS}}$ at LO and NLO. For the longitudinal structure function higher order calculations are also available. A comparison of PDFs extracted from QCD fits to the HERA data with the TR$'$ and ACOT-Full schemes is illustrated in Fig. 3 (taken from [35]).

### 3.2 Electroweak Corrections to DIS

Calculations of higher-order electroweak corrections to DIS at HERA are available in `HERAFitter` in the on-shell scheme.

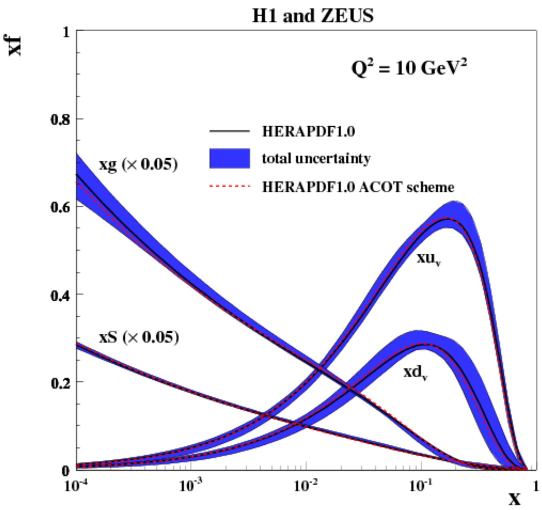


**Fig. 3** Distributions of valence ($xu_v$, $xd_v$), sea ($xS$) and the gluon ($g$) PDFs in HERAPDF1.0 [35] with their total uncertainties at the scale of $Q^2 = 10$ GeV$^2$ obtained using the TR$'$ scheme and compared to the PDFs obtained with the ACOT scheme using the $k$-factor technique (red). The gluon and the sea distributions are scaled down by a factor of 20.

In this scheme, the masses of the gauge bosons $m_W$ and $m_Z$ are treated as basic parameters together with the top, Higgs and fermion masses. These electroweak corrections are based on the EPRC package [53]. The code calculates the running of the electromagnetic coupling $\alpha$ using the most recent parametrisation of the hadronic contribution [54] as well as an older version from Burkhard [55].

### 3.3 Diffractive PDFs

About 10% of deep inelastic interactions at HERA are diffractive, such that the interacting proton stays intact ($ep \to eXp$). The proton is well separated from the rest of the hadronic final state by a large rapidity gap. This is interpreted as the dissociation of the virtual photon into a hadronic system $X$ with an invariant mass much smaller than the photon-proton c.o.m. energy $W = ys - Q^2 + m_p^2(1-y)$, where $m_p$ is the proton mass. Such a process is often assumed to be mediated by the exchange of a hard Pomeron or a secondary Reggeon with vacuum quantum numbers. This factorisable Pomeron picture has proved remarkably successful in the description of most of the diffractive data. Diffractive parton distributions (DPDFs) can be determined from QCD fits to diffractive cross sections in a similar way to the determination of the standard PDFs [56].

In addition to the usual DIS variables $x$, $Q^2$, extra kinematic variables are needed to describe the diffractive process. These are the squared four-momentum transfer of the exchanged Pomeron or Reggeon, $t$, and the mass $m_X$ of the diffractively produced final state. In practice, the variable $m_X$ is often replaced by the dimensionless quantity $\beta = \frac{Q^2}{m_X^2+Q^2-t}$. In models based on a factorisable Pomeron, $\beta$ may be viewed at LO as the fraction of the Pomeron longitudinal momentum, $x_{IP}$, which is carried by the struck parton, $x = \beta x_{IP}$, where $P$ denotes the momentum of the proton.
For the inclusive case, the diffractive cross-section reads as:

$$\frac{d^4\sigma}{d\beta\, dQ^2 dx_{IP}\, dt} = \frac{2\pi\alpha^2}{\beta Q^4}\left(1+(1-y)^2\right)\overline{\sigma}^{D(4)}(\beta, Q^2, x_{IP}, t) \qquad (7)$$

with the "reduced cross-section":

$$\overline{\sigma}^{D(4)} = F_2^{D(4)} - \frac{y^2}{1+(1-y)^2} F_L^{D(4)}. \qquad (8)$$

The diffractive structure functions can be expressed as convolutions of calculable coefficient functions with the diffractive quark and gluon distribution functions, which in general depend on $x_{IP}$, $Q^2$, $\beta$ and $t$.

The diffractive PDFs [57, 58] in HERAFitter are implemented as a sum of two factorised contributions:

$$\Phi_{IP}(x_{IP},t)\, f_a^{IP}(\beta, Q^2) + \Phi_{IR}(x_{IP},t)\, f_a^{IR}(\beta, Q^2), \qquad (9)$$

where $\Phi(x_{IP},t)$ are the Reggeon and Pomeron fluxes. The Reggeon PDFs, $f_a^{IR}$ are fixed as those of the pion, while the Pomeron PDFs, $f_a^{IP}$, can be obtained from a fit to the data.

### 3.4 Drell-Yan Processes in $pp$ or $p\bar{p}$ Collisions

The Drell-Yan (DY) process provides valuable information about PDFs. In $pp$ and $p\bar{p}$ scattering, the $Z/\gamma^*$ and $W$ production probe bi-linear combinations of quarks. Complementary information on the different quark densities can be obtained from the $W^\pm$ asymmetry ($d$, $u$ and their ratio), the ratio of the $W$ and $Z$ cross sections (sensitive to the flavour composition of the quark sea, in particular to the $s$-quark distribution), and associated $W$ and $Z$ production with heavy quarks (sensitive to $c$- and $b$-quark densities). Measurements at large boson transverse momentum $p_T \gtrsim m_{W,Z}$ are potentially sensitive to the gluon distribution [59].

At LO the DY NC cross section triple differential in invariant mass $m$, boson rapidity $y$ and lepton scattering angle $\cos\theta$ in the parton c.o.m. frame can be written as [60, 61]:

$$\frac{d^3\sigma}{dm\,dy\,d\cos\theta} = \frac{\pi\alpha^2}{3ms}\sum_q \hat{\sigma}^q(\cos\theta, m) \times \left[f_q(x_1,m^2) f_{\bar{q}}(x_2,m^2) + (q \leftrightarrow \bar{q})\right], \qquad (10)$$

where $s$ is the squared c.o.m. beam energy, the parton momentum fractions are given by $x_{1,2} = \frac{m}{\sqrt{s}}\exp(\pm y)$, $f_q(x_1, m^2)$ are the PDFs at the scale of the invariant mass, and $\hat{\sigma}^q$ is the parton-parton hard scattering cross section.

The corresponding triple differential CC cross section has the form:

$$\frac{d^3\sigma}{dmdyd\cos\theta} = \frac{\pi\alpha^2}{48s\sin^4\theta_W}\frac{m^3(1-\cos\theta)^2}{(m^2-m_W^2)+\Gamma_W^2 m_W^2} \times \sum_{q_1,q_2} V_{q_1q_2}^2 f_{q_1}(x_1,m^2) f_{q_2}(x_2,m^2), \qquad (11)$$

where $V_{q_1q_2}$ is the Cabibbo-Kobayashi-Maskawa (CKM) quark mixing matrix and $m_W$ and $\Gamma_W$ are the $W$ boson mass and decay width, respectively.

The simple LO form of these expressions allows for the analytic calculations of integrated cross sections. In both NC and CC expressions the PDFs depend only on the boson rapidity $y$ and invariant mass $m$, while the integral in $\cos\theta$ can be evaluated analytically even for the case of realistic kinematic cuts.

Beyond LO, the calculations are often time-consuming and Monte Carlo generators are often employed. Currently, the predictions for $W$ and $Z/\gamma^*$ production are available up to NNLO and the predictions for $W$, $Z$ in association with heavy flavour quarks is available to NLO.

There are several possibilities to obtain the theoretical predictions for DY production in `HERAFitter`. The NLO and NNLO calculations are time consuming and $k$-factor or *fast grid* techniques must be employed (see section 4 for details), which are interfaced to programs such as `MCFM` [62–64], available for NLO calculations, or `FEWZ` [65] and `DYNNLO` [66] for NLO and NNLO, with electro-weak corrections estimated using `MCSANC` [67, 68].

### 3.5 Jet Production in $ep$ and $pp$ or $p\bar{p}$ Collisions

The cross section for production of high $p_T$ hadronic jets is sensitive to the high-$x$ gluon PDF (see e.g. Ref. [11]). Therefore this process can be used to improve the determination of the gluon PDF, which is particularly important for Higgs production and searches for new physics. Jet production cross sections are currently known only to NLO. Calculations for higher-order contributions to jet production in $pp$ collisions are in progress [69–71]. Within `HERAFitter`, the `NLOJet++` program [72, 73] may be used for calculations 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 $pp$ or $p\bar{p}$ Collisions

At the LHC, top-quark pairs ($t\bar{t}$) are produced dominantly via $gg$ fusion. Thus, LHC measurements of the $t\bar{t}$ cross section provide additional constraints on the gluon distribution at medium to high values of $x$, on $\alpha_s$ and on the top-quark mass, $m_t$ [74]. Precise predictions for the total $t\bar{t}$ cross section are available to NNLO [75]. They can be computed within `HERAFitter` via an interface to the program `HATHOR` [76]. Differential $t\bar{t}$ cross section predictions at NLO can be obtained using `MCFM` [64, 77–80] interfaced to `HERAFitter` with *fast grid* techniques.

Single top quarks are produced via electroweak interactions and the measurement of their production cross section 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 to NLO accuracy using `MCFM`.

DiffTop

## 4 Computational Techniques

Precise measurements require accurate theoretical predictions in order to maximise their impact in PDF fits. Perturbative calculations become more complex and time-consuming at higher orders due to the increasing number of relevant Feynman diagrams. The direct inclusion of computationally demanding higher-order calculations into iterative fits is thus not possible currently. However, a full repetition of the perturbative calculation for small changes in input parameters is not necessary at each step of the iteration. Two methods have been developed which take advantage of this to solve the problem: the $k$-factor technique and the *fast grid* technique. Both are available in `HERAFitter`.

### 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 using the same PDF. Because the $k$-factors depend on the phase space probed by the measurement, they have to be stored including their dependence on the relevant kinematic variables. Before the start of a fitting procedure, a table of $k$-factors is computed once for a fixed PDF with the time consuming higher-order code. In subsequent iteration steps the theory prediction is derived from the fast lower-order calculation by multiplying the pre-tabulated $k$-factors.

This procedure, however, neglects the fact that the $k$-factors are PDF dependent, and as a consequence, they have to be re-evaluated for the newly determined PDF at the end of the fit for a consistency check. The fit must be repeated until input and output $k$-factors have converged. In summary, this technique avoids iteration of the higher-order calculation at each step, but still requires typically a few re-evaluations.

In HERAFitter, the $k$-factor technique is also used for the fast computation of the time-consuming GM-VFN schemes for heavy quarks in DIS. "FAST" heavy-flavour schemes are implemented 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). These $k$-factors are calculated only for the starting PDF and hence, the "FAST" heavy flavour schemes should only be used for quick checks. Full heavy flavour schemes should be used by default. However, for the ACOT scheme, due to exceptionally long computation times, the $k$-factors are used in the default setup of HERAFitter.

### 4.2 *Fast Grid* Techniques

*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 judiciously chosen support points. The accuracy of this approximation is checked and optimised such that the approximation bias is negligibly small compared to the experimental and theoretical accuracy. This method can be used to perform the time consuming higher-order calculations (Eq. 1) only once for the set of interpolating functions. Further iterations of the calculation for a particular PDF set are fast, involving only sums over the set of interpolators multiplied by factors depending on the PDF. This approach can be used to calculate the cross sections of processes involving one or two hadrons in the initial state and to assess their renormalisation and factorisation scale variation.

This technique was pioneered by the fastNLO project [81] to facilitate the inclusion of time consuming NLO jet cross section predictions into PDF fits. The APPLGRID [82] project developed an alternative method and, in addition to jets, extended its applicability to other scattering processes, such as DY and heavy quark pair production in association with boson production. The packages differ in their interpolation and optimisation strategies, but both of them 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 $\mu_\mathrm{R}$ and $\mu_\mathrm{F}$ is explored in order to optimise the table size. In the second step the grid is filled for the requested observables. Higher-order cross sections can then be obtained very efficiently from the pre-produced grids while varying externally provided PDF sets, $\mu_\mathrm{R}$ and $\mu_\mathrm{F}$, or the strong coupling $\alpha_s(\mu_R)$. This approach can in principle be extended to arbitrary processes. This requires an interface between the higher-order theory programs and the fast interpolation frameworks. Currently available processes for each package are as follows:

- The fastNLO project [81] has been interfaced to the NLOJet++ program [72] for the calculation of jet production in DIS [83] as well as 2- and 3-jet production in hadron-hadron collisions at NLO [73, 84]. Threshold corrections at 2-loop order, which approximate the NNLO for the inclusive jet cross section, have also been included into the framework [85] following Ref. [86].
  The latest version of the fastNLO convolution program [87] allows for the creation of tables in which renormalisation and factorisation scales can be varied as a function of two pre-defined observables, e.g. jet transverse momentum $p_\perp$ and $Q$ for DIS. Recently, the differential calculation of top-pair production in hadron collisions at approximate NNLO [88] has been interfaced to fastNLO. The fastNLO code is available online [89]. Jet cross-section grids computed for the kinematics of various experiments can be downloaded from this site.
  Dedicated fastNLO libraries and tables with theory predictions for comparison to particular cross section measurements are included into the HERAFitter package. For the HERAFitter implementation, the evaluation of the strong coupling constant is done consistently with the PDF evolution from the QCDNUM code.
- In the APPLGRID package [82, 90], in addition to jet cross sections for $pp(p\bar{p})$ and DIS processes, calculations of DY production are also implemented. The grids are generated with the customised versions of the MCFM parton level DY generator [62–64]. Variation of the renormalisation and factorisation scales is possible a posteriori, when calculating theory predictions with the APPLGRID tables, and independent variation of $\alpha_S$ is also allowed. For higher-order predictions, the $k$-factors technique can also be applied within the APPLGRID framework.
  As an example, the HERAFitter interface to APPLGRID was used by the ATLAS [91] and CMS [92] collaborations to extract the strange quark distribution of the proton. The ATLAS strange PDF extracted employing these techniques is displayed in Fig. 4 together with a comparison to the global PDF sets CT10 [12] and NNPDF2.1 [13] (taken from [91]).

## 5 Fit Methodology

When performing a QCD analysis to determine PDFs there are various assumptions and choices to be made concerning, for example, the functional form of the input parametrisation, the treatment of heavy quarks and their mass values, alternative theoretical calculations, alternative representations of the fit $\chi^2$ and for different ways of treating correlated sys-

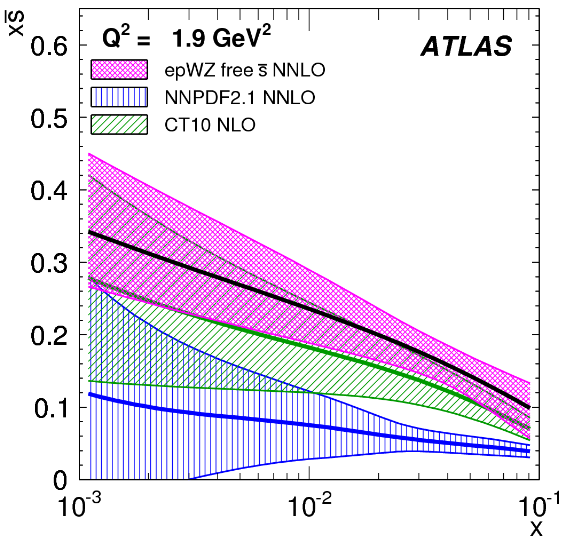


**Fig. 4** The strange antiquark distribution versus $x$ for the ATLAS epWZ free $\bar{s}$ NNLO fit [91] (magenta band) compared to predictions from NNPDF2.1 (blue hatched) and CT10 (green hatched) at $Q^2 = 1.9$ GeV$^2$. The ATLAS fit was performed using a $k$-factor approach for NNLO corrections.

tematic uncertainties. It is useful to discriminate or quantify the effect of a chosen ansatz 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 state-of-the-art techniques, either related to the inclusion of a new theoretical calculation, or of new approaches to treat data and their uncertainties.

In this section we describe the available options for the fit methodology 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 can be parametrised using several predefined functional forms and flavour decompositions:

*Standard Polynomials:* The standard polynomial form is the most commonly used. A polynomial functional form is used to parametrise the $x$-dependence of the PDFs, where index $j$ denotes each parametrised PDF flavour:

$$xf_j(x) = A_j x^{B_j} (1-x)^{C_j} P_j(x). \tag{12}$$

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, which is chosen below the charm mass threshold. The form of polynomials $P_j(x)$ can be varied. The form $(1+\epsilon_j\sqrt{x}+D_j x+E_j x^2)$ is used for the HERAPDF [35] with additional constraints relating to the flavour decomposition of the light sea. This parametrisation is termed HERAPDF-style. The polynomial can also be parametrised in the CTEQ-style, where $P_j(x)$ takes the form $e^{a_3 x}(1+e^{a_4}x+e^{a_5}x^2)$ and, in contrast to the HERAPDF-style, this is positive by construction. QCD number and momentum sum rules are used to determine the normalisations $A$ for the valence and gluon distributions, and the sum-rule integrals are solved analytically.

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

$$xf_j(x) = a_j x^{p_j - b_j \log(x)} (1-x)^{q_j - d_j \log(1-x)}. \tag{13}$$

This function can be regarded as a generalisation of the standard polynomial form described above, however, numerical integration of Eq. 13 is required in order to impose the QCD sum rules.

*Chebyshev Polynomials:* A flexible parametrisation based on the Chebyshev polynomials can be employed for the gluon and sea distributions. Polynomials with argument $\log(x)$ are considered for better modelling the low-$x$ asymptotic behaviour of those PDFs. The polynomials are multiplied by a factor of $(1-x)$ to ensure that they vanish as $x \to 1$. The resulting parametric form reads

$$xg(x) = A_g\,(1-x) \sum_{i=0}^{N_g-1} A_{g_i} T_i \left( -\frac{2\log x - \log x_{\min}}{\log x_{\min}} \right), \tag{14}$$

$$xS(x) = (1-x) \sum_{i=0}^{N_S-1} A_{S_i} T_i \left( -\frac{2\log x - \log x_{\min}}{\log x_{\min}} \right), \tag{15}$$

where $T_i$ are first-type Chebyshev polynomials of order $i$. The normalisation factor $A_g$ is derived from the momentum sum rule analytically. Values of $N_{g,S}$ to 15 are allowed, however the fit quality is already similar to that of the standard-polynomial parametrisation from $N_{g,S} \geq 5$ and has a similar number of free parameters. Fig. 5 (taken from [93]) shows a comparison of the gluon distribution obtained with the parametrisation Eqs. 14, 15 to the standard-polynomial one, for $N_{g,S} = 9$.

*External PDFs:* `HERAFitter` also provides the possibility to access external PDF sets, which can be used to compute theoretical predictions for the cross sections for all the processes available in `HERAFitter`. This is possible via an interface to `LHAPDF` [32, 33] providing access to the global PDF sets. `HERAFitter` also allows one to evolve PDFs from `LHAPDF` using `QCDNUM`. Fig. 6 illustrates a comparison of various gluon PDFs accessed from `LHAPDF` as produced with the drawing tools available in `HERAFitter`.

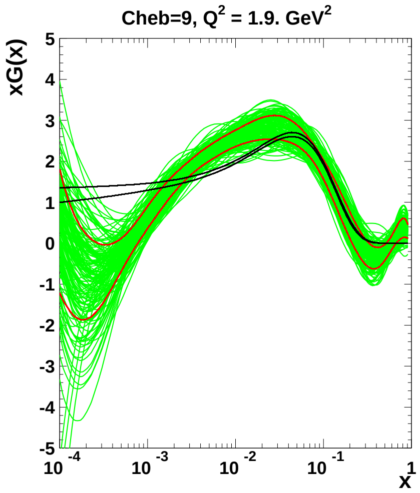


**Fig. 5** The gluon density is shown at the starting scale $Q^2 = 1.9$ GeV$^2$. The black lines correspond to the uncertainty band of the gluon distribution using a standard parametrisation and it is compared to the case of the Chebyshev parametrisation [93]. The uncertainty band for the latter case is estimated using the Monte Carlo technique (see section 5.3) with the green lines denoting fits to data replica. Red lines indicate the standard deviation about the mean value of these replicas.

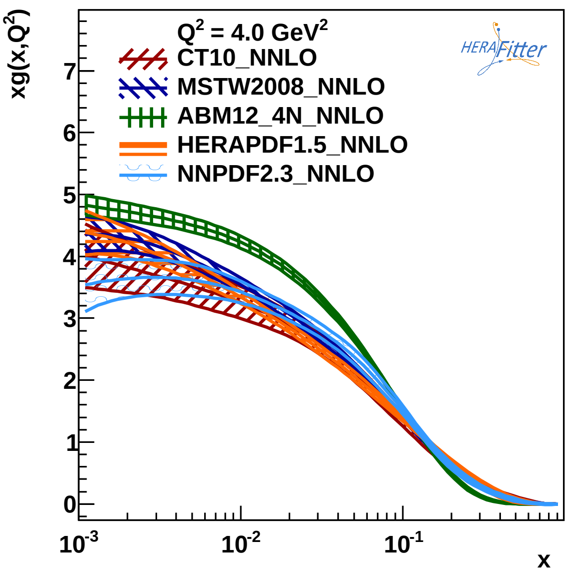


**Fig. 6** The gluon PDF as extracted by various PDF groups at the scale of $Q^2 = 4$ GeV$^2$, plotted using the drawing tools from `HERAFitter`.

## 5.2 Representation of $\chi^2$

The PDF parameters are determined in `HERAFitter` by minimisation of a $\chi^2$ function taking into account correlated and uncorrelated measurement uncertainties. There are various forms of $\chi^2$, e.g. using a covariance matrix or providing nuisance parameters to encode the dependence of each correlated systematic uncertainty for each measured data point. The options available in `HERAFitter` are the following:

*Covariance Matrix Representation:* For a data point $\mu_i$ with a corresponding theory prediction $m_i$, the $\chi^2$ function can be expressed in the following form:

$$\chi^2(m) = \sum_{i,k}(m_i - \mu_i)C_{ik}^{-1}(m_k - \mu_k), \tag{16}$$

where the experimental uncertainties are given as a covariance matrix $C_{ik}$ for measurements in bins $i$ and $k$. The covariance matrix $C_{ik}$ is given by a sum of statistical, uncorrelated and correlated systematic contributions:

$$C_{ik} = C_{ik}^{stat} + C_{ik}^{uncor} + C_{ik}^{sys}. \tag{17}$$

Using this representation one cannot distinguish the effect of each source of systematic uncertainty.

*Nuisance Parameter Representation:* In this case, the $\chi^2$ is expressed as

$$\chi^2(m,b) = \sum_i \frac{\left[\mu_i - m_i\left(1 - \sum_j \gamma_j^i b_j\right)\right]^2}{\delta_{i,\mathrm{unc}}^2 m_i^2 + \delta_{i,\mathrm{stat}}^2 \mu_i m_i \left(1 - \sum_j \gamma_j^i b_j\right)} + \sum_j b_j^2, \tag{18}$$

where, $\delta_{i,\mathrm{stat}}$ and $\delta_{i,\mathrm{unc}}$ are relative statistical and uncorrelated systematic uncertainties of the measurement $i$. Further, $\gamma_j^i$ quantifies the sensitivity of the measurement to the correlated systematic source $j$. The function $\chi^2$ depends on the set of systematic nuisance parameters $b_j$. This definition of the $\chi^2$ 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 $\chi^2$ minimisation, the nuisance parameters $b_j$ and the PDFs are determined, such that the effect of different sources of systematic uncertainties can be distinguished.

*Mixed Form Representation:* In some cases, the statistical and systematic uncertainties of experimental data 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 the form of a covariance matrix. `HERAFitter` offers the possibility to include such mixed forms of information.

Any source of measured systematic uncertainty can be treated as additive (i.e. as absolute uncertainty) or multiplicative (i.e. as a relative uncertainty). The statistical uncertainties can be included as additive or following the Poisson statistics. Minimisation with respect to nuisance parameters is performed analytically, however, for more detailed studies of correlations individual nuisance parameters can be included into the `MINUIT` minimisation.

## 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 (Eigenvector) method:* The PDF uncertainties reflecting the data experimental uncertainties are estimated by examining the shape of the $\chi^2$ function in the neighbourhood of the minimum [94]. Following the approach of Ref. [94], the Hessian matrix is defined by the second derivatives of $\chi^2$ on the fitted PDF parameters. The matrix is diagonalised and the Hessian eigenvectors are computed. Due to orthogonality these vectors correspond to independent sources of uncertainty in the obtained PDFs.

*Offset method:* The Offset method [95] uses the $\chi^2$ function for the central fit, but only uncorrelated uncertainties are taken into account. The goodness of the fit can no longer be judged from the $\chi^2$ since correlated uncertainties are ignored. The correlated uncertainties are propagated into the PDF uncertainties by performing variants of the fit with the experimental data varied by $\pm 1\sigma$ from the central value for each systematic source. The resulting deviations of the PDF parameters from the ones obtained in the central fit are statistically independent, and they can be combined in quadrature to derive a total PDF systematic uncertainty.
The uncertainties estimated by the offset method are generally larger than those from the Hessian method.

*Monte Carlo method:* The Monte Carlo (MC) technique [96, 97] can also be used to determine PDF uncertainties. The uncertainties are estimated using pseudo-data replicas (typically $> 100$) randomly generated from the measurement central values and their systematic and statistical uncertainties taking into account all point-to-point correlations. The QCD fit is performed for each replica and the PDF central values and their experimental uncertainties are estimated from the distribution of the PDF parameters obtained in these fits, by taking the mean values and standard deviations over the replicas.
The MC method has been checked against the standard error estimation of the PDF uncertainties obtained by the Hessian method. A good agreement was found between the methods provided that Gaussian distributions of statistical and systematic uncertainties are assumed in the MC approach [31]. A comparison is illustrated in Fig. 7. Similar findings were reported by the MSTW global analysis [98].
Since the MC method requires large number of replicas, the eigenvector representation is a more convenient way to store the PDF uncertainties. It is possible to transform MC to eigenvector representation as shown by [99]. Tools to perform this transformation are provided with HERAFitter and were recently employed for the representation of correlated sets of PDFs at different perturbative orders [100].

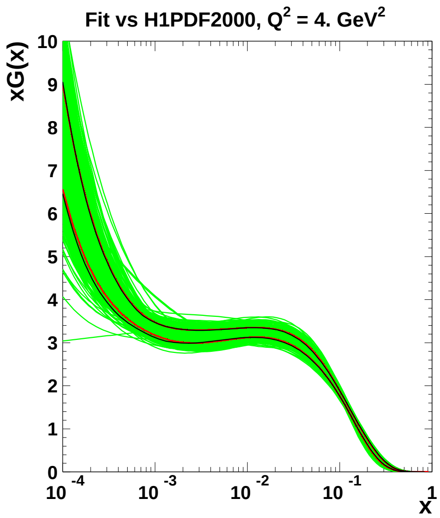


**Fig. 7** Comparison between the standard error calculations as employed by the Hessian approach (black lines) and the MC approach (with more than 100 replicas) assuming Gaussian distribution for uncertainty distributions, shown here for each replica (green lines) together with the evaluated standard deviation (red lines) [31]. The black and red lines in the figure are superimposed because agreement of the methods is so good that it is hard to distinguish them.

The nuisance parameter representation of $\chi^2$ in Eq. 18 is derived assuming symmetric experimental errors, however, the published systematic uncertainties are often asymmetric. HERAFitter provides the possibility to use asymmetric systematic uncertainties. The implementation relies on the assumption that asymmetric uncertainties can be described by a parabolic function. The nuisance parameter in Eq. 18 is modified as follows

$$\gamma_j^i \to \omega_j^i b_j + \gamma_j^i, \tag{19}$$

where the coefficients $\omega_j^i$, $\gamma_j^i$ are defined from the maximum and minimum shifts of the cross sections due to a variation of the systematic uncertainty $j$, $S_{ij}^{\pm}$,

$$\omega_j^i = \frac{1}{2}\left(S_{ij}^{+} + S_{ij}^{-}\right), \qquad \gamma_j^i = \frac{1}{2}\left(S_{ij}^{+} - S_{ij}^{-}\right). \tag{20}$$

## 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 in the theoretical calculations. Nowadays, PDF groups 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 $\alpha_s(m_Z)$. Other important aspects are the choice of the functional form for the PDFs at the starting scale and the value of the starting scale itself. `HERAFitter` provides the possibility of different user choices of all these inputs.

### 5.5 Bayesian Reweighting Techniques

As an alternative to performing a full QCD fit, `HERAFitter` allows the user to assess the impact of including new data in an existing fit using the Bayesian Reweighting technique. The method provides a fast estimate of the impact of new data on PDFs. Bayesian Reweighting was first proposed for PDF sets delivered in the form of MC replicas by [96] and further developed by the NNPDF Collaboration [101, 102]. More recently, a method to perform Bayesian Reweighting studies starting from PDF fits for which uncertainties are provided in the eigenvector representation has been also developed [98]. The latter is based on generating replica sets by introducing Gaussian fluctuations on the central PDF set with a variance determined by the PDF uncertainty given by the eigenvectors. Both reweighting methods are implemented in `HERAFitter`.

The Bayesian Reweighting technique relies on the fact that MC replicas of a PDF set give a representation of the probability distribution in the space of PDFs. In particular, the PDFs are represented as ensembles of $N_{\text{rep}}$ equiprobable (*i.e.* having weights equal to unity) replicas, $\{f\}$. The central value for a given observable, $\mathscr{O}(\{f\})$, is computed as the average of the predictions obtained from the ensemble as

$$\langle \mathscr{O}(\{f\}) \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \mathscr{O}(f^k), \tag{21}$$

and the uncertainty as the standard deviation of the sample.

Upon inclusion of new data the prior probability distribution, given by theoriginal PDF set, is modified according to Bayes Theorem such that the weight of each replica, $w_k$, is updated according to

$$w_k = \frac{(\chi_k^2)^{\frac{1}{2}(N_{\text{data}}-1)} e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} (\chi_k^2)^{\frac{1}{2}(N_{\text{data}}-1)} e^{-\frac{1}{2}\chi_k^2}}, \tag{22}$$

where $N_{\text{data}}$ is the number of new data points, $k$ denotes the specific replica for which the weight is calculated and $\chi_k^2$ is the $\chi^2$ of the new data obtained using the $k$-th PDF replica. Given a PDF set and a corresponding set of weights, which describes the impact of the inclusion of new data, the prediction for a given observable after inclusion of the new data can be computed as the *weighted* average,

$$\langle \mathscr{O}(\{f\}) \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} w_k \mathscr{O}(f^k). \tag{23}$$

To simplify the use of a reweighted set, an unweighted set (*i.e.* a set of equiprobable replicas which incorporates the information contained in the weights) is generated according to the unweighting procedure described in [101]. The number of effective replicas of a reweighted set is measured by its Shannon Entropy [102]

$$N_{\text{eff}} \equiv \exp\left\{ \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} w_k \ln(N_{\text{rep}}/w_k) \right\}, \tag{24}$$

which corresponds to the size of a refitted equiprobable replica set containing the same amount of information. This number of effective replicas, $N_{\text{eff}}$, gives an indicative measure of the optimal size of an unweighted replica set produced with the reweighting/unweighting procedure. No extra information is gained by producing a final unweighted set that has a number of replicas (significantly) larger than $N_{\text{eff}}$. If $N_{\text{eff}}$ is much smaller than the original number of replicas the new data have great impact, however, it is unreliable to use the new reweighted set. In this case, instead, a full refit should be performed.

## 6 Alternatives to DGLAP Formalism

QCD calculations based on the DGLAP [16–20] evolution equations are very successful in describing all relevant hard scattering data in the perturbative region $Q^2 \gtrsim$ few GeV$^2$. At small-$x$ and small-$Q^2$ DGLAP dynamics may be modified by saturation and other (non-perturbative) higher-twist effects. Different approaches alternative to the DGLAP formalism can be used to analyse DIS data in `HERAFitter`. These include several dipole models and the use of transverse momentum dependent, or unintegrated PDFs (uPDFs).

### 6.1 Dipole Models

The dipole picture provides an alternative approach to proton-virtual photon scattering at low $x$ which can be applied to 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 [103, 104]. The dipoles can be considered 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 with the proton. The dynamics of the interaction are embedded in a dipole scattering amplitude.

Several dipole models, which assume different behaviours of the dipole-proton cross section, are implemented in `HERAFitter`: the Golec-Biernat-Wüsthoff (GBW) dipole saturation model [27], a modified GBW model which takes into account the effects of DGLAP evolution, termed the Bartels-Golec-Kowalski (BGK) dipole model [29] and the colour glass condensate approach to the high parton density regime, named the Iancu-Itakura-Munier (IIM) dipole model [28].

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

$$\sigma_{\text{dip}}(x,r^2) = \sigma_0 \left(1 - \exp\left[-\frac{r^2}{4R_0^2(x)}\right]\right), \tag{25}$$

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. $R_0^2$ takes the form, $R_0^2(x) = (x/x_0)^{\lambda} 1/\text{GeV}^2$, and is called the saturation radius. The cross-section normalisation $\sigma_0$, $x_0$, and $\lambda$ are parameters of the model fitted to the DIS data. This model gives exact Bjorken scaling when the dipole size $r$ is small.

*BGK model:* The BGK model is a modification of the GBW model assuming that the spacing $R_0$ is inverse to the gluon distribution and taking into account the DGLAP evolution of the latter. The gluon distribution, parametrised at some starting scale by Eq. 12, is evolved to larger scales using DGLAP evolution.

*BGK model with valence quarks:* The dipole models are valid in the low-$x$ region only, where the valence quark contribution to the total proton momentum is 5% to 15% for $x$ from 0.0001 to 0.01 [105]. The inclusive HERA measurements have a precision which is better than 2%. Therefore, `HERAFitter` provides the option of taking into account the contribution of the valence quarks

*IIM model:* The IIM model assumes an expression for the dipole cross section which is based on the Balitsky-Kovchegov equation [106]. The explicit formula for $\sigma_{\text{dip}}$ can be found in [28]. The alternative scale parameter $\tilde{R}$, $x_0$ and $\lambda$ are fitted parameters of the model.

### 6.2 Transverse Momentum Dependent PDFs

QCD calculations of multiple-scale processes and complex final-states can necessitate the use of transverse-momentum dependent (TMD) [8], or unintegrated parton distribution and parton decay functions [107–115]. TMD factorisation has been proven recently [8] for inclusive DIS. TMD factorisation has also been proven in the high-energy (small-$x$) limit [116–118] for particular hadron-hadron scattering processes, like heavy flavour, vector boson and Higgs production.

In the framework of high-energy factorisation [116, 119, 120] the DIS cross section can be written as a convolution in both longitudinal and transverse momenta of the TMD parton distribution function $\mathscr{A}\left(x,k_t,\mu_F^2\right)$ with the off-shell parton scattering matrix elements as follows

$$\sigma_j(x,Q^2) = \int_x^1 dz \int d^2k_t\, \hat{\sigma}_j(x,Q^2,z,k_t)\, \mathscr{A}\left(z,k_t,\mu_F^2\right), \tag{26}$$

where the DIS cross sections $\sigma_j$ $(j=2,L)$ are related to the structure functions $F_2$ and $F_L$. The hard-scattering kernels $\hat{\sigma}_j$ of Eq. 26, are $k_t$-dependent and the evolution of the transverse-momentum dependent gluon distribution $\mathscr{A}$ is obtained by combining the resummation of small-$x$ logarithmic contributions [121–123] with medium-$x$ and large-$x$ contributions to parton splitting [16, 19, 20] according to the CCFM evolution equation [24, 124, 125].

The factorisation formula (26) allows resummation of logarithmically enhanced small-$x$ contributions to all orders in perturbation theory, both in the hard-scattering coefficients and in the parton evolution, fully taking into account the dependence on the factorisation scale $\mu_F$ and on the factorisation scheme [126, 127].

The cross section $\sigma_j$ $(j=2,L)$ is calculated in a FFN scheme, using the boson-gluon fusion process $(\gamma^* g^* \to q\bar{q})$. The masses of the quarks are explicitly included as parameters of the model. In addition to $\gamma^* g^* \to q\bar{q}$, the contribution from valence quarks is included via $\gamma^* q \to q$ by using a CCFM evolution of valence quarks [128, 129].

*CCFM Grid Techniques:* The CCFM evolution cannot be written easily in an analytic closed form. For this reason, a MC method is employed, which is, however, time-consuming, and thus cannot be used directly in a fit program.

Following the convolution method introduced in [129, 130], the kernel $\tilde{\mathscr{A}}(x'',k_t,p)$ is determined from the MC solution of the CCFM evolution equation, and then folded with a non-perturbative starting distribution $\mathscr{A}_0(x)$

$$\begin{aligned} x\mathscr{A}(x,k_t,p) &= x\int dx' \int dx'' \mathscr{A}_0(x') \tilde{\mathscr{A}}\left(x'',k_t,p\right) \delta(x'x''-x) \\ &= \int dx' \mathscr{A}_0(x') \frac{x}{x'}\, \tilde{\mathscr{A}}\left(\frac{x}{x'},k_t,p\right), \end{aligned} \tag{27}$$

where $k_t$ denotes the transverse momentum of the propagator gluon and $p$ is the evolution variable.

The kernel $\tilde{\mathscr{A}}$ incorporates all of the dynamics of the evolution. It is defined 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 $x$ for which 40 bins in logarithmic spacing below 0.1, and 10 bins in linear spacing above 0.1 are used.

Calculation of the cross section according to Eq. 26 involves a time-consuming multidimensional MC integration, which suffers from numerical fluctuations. This cannot be employed directly in a fit procedure. Instead the following equation is applied:

$$\begin{aligned} \sigma(x,Q^2) &= \int_x^1 dx_g \mathscr{A}(x_g,k_t,p)\hat{\sigma}(x,x_g,Q^2) \\ &= \int_x^1 dx' \mathscr{A}_0(x')\tilde{\sigma}(x/x',Q^2), \end{aligned} \tag{28}$$

where first $\tilde{\sigma}(x',Q^2)$ is calculated numerically with a MC integration on a grid in $x$ for the values of $Q^2$ used in the fit. Then the last step in Eq. 28 is performed with a fast numerical Gauss integration, which can be used directly in the fit.

*Functional Forms for TMD parametrisation:* For the starting distribution $\mathscr{A}_0$, at the starting scale $Q_0^2$, the following form is used:

$$x\mathscr{A}_0(x,k_t) = Nx^{-B}(1-x)^C\left(1-Dx+E\sqrt{x}\right) \times \exp[-k_t^2/\sigma^2], \tag{29}$$

where $\sigma^2 = Q_0^2/2$ and $N,B,C,D,E$ are free parameters. Valence quarks are treated using the method of Ref. [128] as described in Ref. [129] with a starting distribution taken from any collinear PDF and imposition of the flavour sum rule at every scale $p$.
The TMD parton densities can be plotted either with `HERAFitter` tools or with `TMDplotter` [34].

## 7 `HERAFitter` Code Organisation

`HERAFitter` is an open source code licensed under the GNU general public licence. It can be downloaded from a dedicated webpage [1] together with its supporting documentation and *fast grid* theory files (described in section 4) associated with data files. The source code contains all the relevant information to perform QCD fits with HERA DIS data as a default set. [1] The execution time depends on the fitting options and varies from 10 minutes (using "FAST" techniques as described in section 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 dependencies are required except the `QCDNUM` evolution program [21]. The `ROOT` libraries are only required for the drawing tools and when invoking `APPLGRID`. Drawing tools built into `HERAFitter` provide a qualitative and quantitative assessment of the results. Fig. 8 shows an illustration of a comparison between the inclusive NC data from HERA I with the predictions based on HERAPDF1.0 PDFs. The consistency of the measurements and the theory can be expressed by pulls, defined as the 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 standard deviations. The pulls are also illustrated in Fig. 8.

In `HERAFitter` there are also available cache options for fast retrieval, fast evolution kernels, and the OpenMP (Open Multi-Processing) interface which allows parallel applications of the GM-VFNS theory predictions in DIS.

[1]Default settings in `HERAFitter` are tuned to reproduce the central HERAPDF1.0 set.

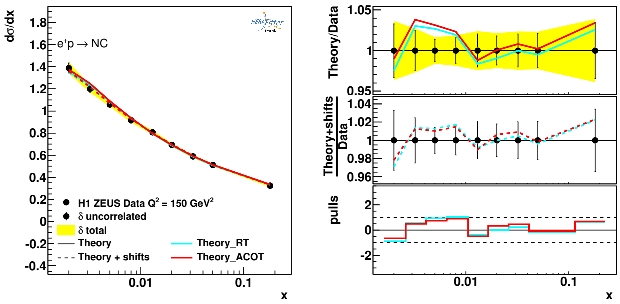


**Fig. 8** An illustration of the consistency of HERA measurements [35] and the theory predictions, obtained in `HERAFitter` with the default drawing tool.

## 8 Applications of `HERAFitter`

The `HERAFitter` program has been used in a number of experimental and theoretical analyses. This list includes several LHC analyses of SM processes, namely inclusive Drell-Yan and $W$ and $Z$ production [91, 92, 131–133], inclusive jet production [134], and inclusive photon production [135]. The results of QCD analyses using `HERAFitter` were also published by HERA experiments for inclusive [35, 136] and heavy flavour production measurements [137, 138]. The following phenomenological studies have been performed with `HERAFitter`: a determination of the transverse momentum dependent gluon distribution using precision HERA data [129], an analysis of HERA data within a dipole model [139], the study of the low-x uncertainties in PDFs determined from the HERA data using different parametrisations [93] and the impact of QED radiative corrections on PDFs [140]. A recent study based on a set of PDFs determined with `HERAFitter` and addressing the correlated uncertainties between different orders has been published in [100]. An application of the TMDs obtained with `HERAFitter` $W$ production at the LHC can be found in [141].

The `HERAFitter` framework has been used to produce PDF grids from QCD analyses performed at HERA [35, 142] and at the LHC [143], using measurements from ATLAS [91, 134]. These PDFs can be used to study predictions for SM or beyond SM processes. Furthermore, `HERAFitter` provides the possibility to perform various benchmarking exercises [144] and impact studies for possible future colliders as demonstrated by QCD studies at the LHeC [145].

## 9 Summary

`HERAFitter` is an open-source platform designed for studies of the structure of the proton. It provides a unique and flexible framework with a wide variety of QCD tools to facilitate analyses of the experimental data and theoretical calculations. `HERAFitter` allows for direct comparisons of various theoretical approaches under the same settings, different methodologies in treating the experimental and model uncertainties can be used for benchmarking studies. The progress of HERAFitter is driven by the latest QCD advances in theoretical calculations and in the precision of experimental data.

The HERAFitter code, in version 1.1.0, has sufficient options to reproduce the different theoretical choices made in MSTW, CTEQ and ABM fits. This will potentially make it a valuable tool for benchmarking and understanding differences between PDF fits. Such a study would however need to consider a range of further questions, such as the choices of data sets, treatments of uncertainties, input parameter values, $\chi^2$ definitions and so forth. We look forward to studying these questions in future work.

**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 the BMBF-JINR cooperation program, the Heisenberg-Landau program, the RFBR grant 12-02-91526-CERN a, the Polish NSC project DEC-2011/03/B/ST2/00220 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), [arXiv:1207.7214].
3. S. Chatrchyan *et al.* [CMS Collaboration], Phys.Lett. **B716**, 30 (2012), [arXiv:1207.7235].
4. J. C. Collins and D. E. Soper, Nucl.Phys. **B194**, 445 (1982).
5. J. C. Collins, D. E. Soper, and G. F. Sterman, Phys.Lett. **B134**, 263 (1984).
6. J. C. Collins, D. E. Soper, and G. F. Sterman, Nucl.Phys. **B261**, 104 (1985).
7. J. C. Collins, D. E. Soper, and G. F. Sterman, Adv.Ser.Direct.High Energy Phys. **5**, 1 (1988), [hep-ph/0409313].
8. J. Collins, *Foundations of perturbative QCD*, vol. 32 (Cambridge monographs on particle physics, nuclear physics and cosmology, 2011).
9. E. Perez and E. Rizvi, Rep.Prog.Phys. **76**, 046201 (2013), [arXiv:1208.1178].
10. S. Forte and G. Watt, Ann.Rev.Nucl.Part.Sci. **63**, 291 (2013), [arXiv:1301.6754].
11. A. Martin, W. Stirling, R. Thorne, and G. Watt, Eur. Phys. J. C **63**, 189 (2009), [arXiv:0901.0002], URL `http://mstwpdf.hepforge.org/`.
12. J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, *et al.*, Phys.Rev. **D89**, 033009 (2014), [arXiv:1302.6246], URL `http://hep.pa.msu.edu/cteq/public/`.
13. R. D. Ball *et al.*, Nucl.Phys. **B867**, 244 (2013), [arXiv:1207.1303], URL `https://nnpdf.hepforge.org/`.
14. S. Alekhin, J. Bluemlein, and S. Moch, Phys.Rev. **D89**, 054028 (2014), [arXiv:1310.3059].
15. P. Jimenez-Delgado and E. Reya, Phys.Rev. **D89**, 074049 (2014), [arXiv:1403.1852].
16. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).
17. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 675 (1972).
18. L. N. Lipatov, Sov. J. Nucl. Phys. **20**, 94 (1975).
19. Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
20. G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
21. M. Botje, Comput.Phys.Commun. **182**, 490 (2011), `http://www.nikef.nl/h24/qcdnum/index.html`, [1005.1481].
22. M. Ciafaloni, Nucl. Phys. B **296**, 49 (1988).
23. S. Catani, F. Fiorani, and G. Marchesini, Phys. Lett. B **234**, 339 (1990).
24. S. Catani, F. Fiorani, and G. Marchesini, Nucl. Phys. B **336**, 18 (1990).
25. G. Marchesini, Nucl. Phys. B **445**, 49 (1995).
26. F. Hautmann, H. Jung, and S. T. Monfared, Eur. Phys. J. C **74**, 3082 (2014), [arXiv:1407.5935].
27. K. Golec-Biernat and M. Wüsthoff, Phys. Rev. D **59**, 014017 (1999), [hep-ph/9807513].
28. E. Iancu, K. Itakura, and S. Munier, Phys. Lett. **B590**, 199 (2004), [hep-ph/0310338].
29. J. Bartels, K. Golec-Biernat, and H. Kowalski, Phys. Rev. D **66**, 014001 (2002), [hep-ph/0203258].
30. F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).
31. M. Dittmar, S. Forte, A. Glazov, S. Moch, G. Altarelli, *et al.* (2009), [arXiv:0901.2504].
32. M. Whalley, D. Bourilkov, and R. Group (2005), [hep-ph/0508110].
33. *LHAPDF*, URL `http://lhapdf.hepforge.org`.
34. F. Hautmann, H. Jung, M. Kramer, P. Mulders, E. Nocera, *et al.* (2014), [arXiv:1408.3015].
35. F. Aaron *et al.* [H1 and ZEUS Collaborations], JHEP **1001**, 109 (2010), [arXiv:0911.0884].
36. R. Devenish and A. Cooper-Sarkar (2011), *Deep Inelastic Scattering*, ISBN: 0199602255,9780199602254.
37. J. C. Collins and W.-K. Tung, Nucl. Phys. B **278**, 934 (1986).
38. E. Laenen *et al.*, Phys. Lett. **B291**, 325 (1992).
39. E. Laenen *et al.*, Nucl. Phys. **B392**, 162, 229 (1993).

40. S. Riemersma, J. Smith, and van Neerven. W.L., Phys. Lett. **B347**, 143 (1995), [hep-ph/9411431].
41. S. Alekhin, J. Blümlein, and S. Moch, *OPENQCDRAD*, http://www-zeuthen.desy.de/∼alekhin/OPENQCDRAD.
42. H. Kawamura, N. Lo Presti, S. Moch, and A. Vogt, Nucl.Phys. **B864**, 399 (2012).
43. S. Alekhin and S. Moch, Phys. Lett. **B699**, 345 (2011), [arXiv:1011.5790].
44. M. Beneke, Phys.Rept. **317**, 1 (1999), [hep-ph/9807443].
45. R. Demina, S. Keller, M. Kramer, S. Kretzer, R. Martin, *et al.* (1999), [hep-ph/0005112].
46. R. S. Thorne and R. G. Roberts, Phys. Rev. D **57**, 6871 (1998), [hep-ph/9709442].
47. R. S. Thorne, Phys. Rev. **D73**, 054019 (2006), [hep-ph/0601245].
48. R. S. Thorne, Phys. Rev. D **86**, 074017 (2012), [arXiv:1201.6180].
49. J. C. Collins, Phys.Rev. **D58**, 094002 (1998), [hep-ph/9806259].
50. M. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, Phys.Rev. **D50**, 3102 (1994), [hep-ph/9312319].
51. M. Kramer, F. I. Olness, and D. E. Soper, Phys. Rev. **D62**, 096007 (2000), [hep-ph/0003035].
52. S. Kretzer, H. Lai, F. Olness, and W. Tung, Phys.Rev. **D69**, 114005 (2004), [hep-ph/0307022].
53. H. Spiesberger, Private communication.
54. F. Jegerlehner, Proceedings, LC10 Workshop **DESY 11-117** (2011).
55. H. Burkhard, F. Jegerlehner, G. Penso, and C. Verzegnassi, in CERN Yellow Report on "Polarization at LEP" 1988.
56. J. C. Collins, Phys.Rev. **D57**, 3051 (1998), [hep-ph/9709499].
57. A. Aktas *et al.* [H1 Collaboration], Eur.Phys.J. **C48**, 715 (2006), [hep-ex/0606004].
58. S. Chekanov *et al.* [ZEUS Collaboration], Nucl. Phys. **B831**, 1 (2010), [hep-ex/09114119].
59. S. A. Malik and G. Watt, JHEP **1402**, 025 (2014), [arXiv:1304.2424].
60. S. D. Drell and T.-M. Yan, Phys. Rev. Lett. **25**, 316 (1970).
61. M. Yamada and M. Hayashi, Nuovo Cim. **A70**, 273 (1982).
62. J. M. Campbell and R. K. Ellis, Phys. Rev. **D60**, 113006 (1999), [arXiv:9905386].
63. J. M. Campbell and R. K. Ellis, Phys. Rev. **D62**, 114012 (2000), [arXiv:0006304].
64. J. M. Campbell and R. K. Ellis, Nucl. Phys. Proc. Suppl. **205-206**, 10 (2010), [arXiv:1007.3492].
65. Y. Li and F. Petriello, Phys.Rev. **D86**, 094034 (2012), [arXiv:1208.5967].
66. G. Bozzi, J. Rojo, and A. Vicini, Phys.Rev. **D83**, 113008 (2011), [arXiv:1104.2056].
67. D. Bardin, S. Bondarenko, P. Christova, L. Kalinovskaya, L. Rumyantsev, *et al.*, JETP Lett. **96**, 285 (2012), [1207.4400].
68. S. G. Bondarenko and A. A. Sapronov, Comput.Phys.Commun. **184**, 2343 (2013), [1301.3687].
69. A. Gehrmann-De Ridder, T. Gehrmann, E. Glover, and J. Pires, Phys. Rev. Lett. **110**, 162003 (2013), [arXiv:1301.7310].
70. E. Glover and J. Pires, JHEP **1006**, 096 (2010), [arXiv:1003.2824].
71. J. Currie, A. Gehrmann-De Ridder, E. Glover, and J. Pires, JHEP **1401**, 110 (2014), [arXiv:1310.3993].
72. Z. Nagy and Z. Trocsanyi, Phys.Rev. **D59**, 014020 (1999), [hep-ph/9806317].
73. Z. Nagy, Phys.Rev.Lett. **88**, 122003 (2002), [hep-ph/0110315].
74. S. Chatrchyan *et al.* [CMS Collaboration], Phys.Lett. **B728**, 496 (2014), [arXiv:1307.1907].
75. M. Czakon, P. Fiedler, and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013), [arXiv:1303.6254].
76. M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, *et al.*, Comput.Phys.Commun. **182**, 1034 (2011), [arXiv:1007.1327].
77. J. M. Campbell, R. Frederix, F. Maltoni, and F. Tramontano, Phys.Rev.Lett. **102**, 182003 (2009), [arXiv:0903.0005].
78. J. M. Campbell and F. Tramontano, Nucl.Phys. **B726**, 109 (2005), [hep-ph/0506289].
79. J. M. Campbell, R. K. Ellis, and F. Tramontano, Phys.Rev. **D70**, 094012 (2004), [hep-ph/0408158].
80. J. M. Campbell and R. K. Ellis (2012), report FERMILAB-PUB-12-078-T, [arXiv:1204.1513].
81. T. Kluge, K. Rabbertz, and M. Wobisch, pp. 483–486 (2006), [hep-ph/0609285].
82. T. Carli *et al.*, Eur. Phys. J. **C66**, 503 (2010), [arXiv:0911.2985].
83. Z. Nagy and Z. Trocsanyi, Phys.Rev.Lett. **87**, 082001 (2001), [hep-ph/0104315].
84. Z. Nagy, Phys.Rev. **D68**, 094002 (2003), [hep-ph/0307268].
85. M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober [fastNLO Collaboration] (2011), [arXiv:1109.1310].
86. N. Kidonakis and J. Owens, Phys.Rev. **D63**, 054019 (2001), [hep-ph/0007268].
87. D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch [fastNLO Collaboration] (2012), [arXiv:1208.3641].
88. M. Guzzi, K. Lipka, and S.-O. Moch (2014), [arXiv:1406.0386].
89. http://fastnlo.hepforge.org, URL `http://fastnlo.hepforge.org`.

90. http://applgrid.hepforge.org, URL `http://applgrid.hepforge.org`.
91. G. Aad *et al.* [ATLAS Collaboration], Phys. Rev. Lett. **109**, 012001 (2012), [arXiv:1203.4051].
92. S. Chatrchyan *et al.* [CMS Collaboration], Phys.Rev. **D90**, 032004 (2014), [arXiv:1312.6283].
93. A. Glazov, S. Moch, and V. Radescu, Phys. Lett. B **695**, 238 (2011), [arXiv:1009.6170].
94. J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, *et al.*, Phys.Rev. **D65**, 014013 (2001), [hep-ph/0101032].
95. M. Botje, J.Phys. **G28**, 779 (2002), [hep-ph/0110123].
96. W. T. Giele and S. Keller, Phys.Rev. **D58**, 094023 (1998), [hep-ph/9803393].
97. W. T. Giele, S. Keller, and D. Kosower (2001), [hep-ph/0104052].
98. G. Watt and R. Thorne, JHEP **1208**, 052 (2012), [arXiv:1205.4024].
99. J. Gao and P. Nadolsky, JHEP **1407**, 035 (2014), [arXiv:1401.0013].
100. HERAFitter Developers Team and M. Lisovyi (2014), [arXiv:1404.4234].
101. R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, *et al.*, Nucl.Phys. **B855**, 608 (2012), [arXiv:1108.1758].
102. R. D. Ball *et al.* [NNPDF Collaboration], Nucl.Phys. **B849**, 112 (2011), [arXiv:1012.0836].
103. N. N. Nikolaev and B. Zakharov, Z.Phys. **C49**, 607 (1991).
104. A. H. Mueller, Nucl. Phys. B **415**, 373 (1994).
105. F. Aaron *et al.* [H1 Collaboration], Eur.Phys.J. **C71**, 1579 (2011), [arXiv:1012.4355].
106. I. Balitsky, Nucl. Phys. B **463**, 99 (1996), [hep-ph/9509348].
107. S. M. Aybat and T. C. Rogers, Phys.Rev. **D83**, 114042 (2011), [arXiv:1101.5057].
108. M. Buffing, P. Mulders, and A. Mukherjee, Int.J.Mod.Phys.Conf.Ser. **25**, 1460003 (2014), [arXiv:1309.2472].
109. M. Buffing, A. Mukherjee, and P. Mulders, Phys.Rev. **D88**, 054027 (2013), [arXiv:1306.5897].
110. M. Buffing, A. Mukherjee, and P. Mulders, Phys.Rev. **D86**, 074030 (2012), [arXiv:1207.3221].
111. P. Mulders, Pramana **72**, 83 (2009), [arXiv:0806.1134].
112. S. Jadach and M. Skrzypek, Acta Phys.Polon. **B40**, 2071 (2009), [arXiv:0905.1399].
113. F. Hautmann, Acta Phys.Polon. **B40**, 2139 (2009).
114. F. Hautmann, M. Hentschinski, and H. Jung (2012), [arXiv:1205.6358].
115. F. Hautmann and H. Jung, Nucl.Phys.Proc.Suppl. **184**, 64 (2008), [arXiv:0712.0568].
116. S. Catani, M. Ciafaloni, and F. Hautmann, Phys. Lett. B **242**, 97 (1990).
117. J. C. Collins and R. K. Ellis, Nucl. Phys. B **360**, 3 (1991).
118. F. Hautmann, Phys.Lett. **B535**, 159 (2002), [hep-ph/0203140].
119. S. Catani, M. Ciafaloni, and F. Hautmann, Nucl. Phys. B **366**, 135 (1991).
120. S. Catani, M. Ciafaloni, and F. Hautmann, Phys. Lett. B **307**, 147 (1993).
121. L. Lipatov, Phys.Rept. **286**, 131 (1997), [hep-ph/9610276].
122. V. S. Fadin, E. Kuraev, and L. Lipatov, Phys.Lett. **B60**, 50 (1975).
123. I. I. Balitsky and L. N. Lipatov, Sov. J. Nucl. Phys. **28**, 822 (1978).
124. M. Ciafaloni, Nucl. Phys. **B296**, 49 (1988).
125. G. Marchesini, Nucl. Phys. B **445**, 49 (1995), [hep-ph/9412327].
126. S. Catani and F. Hautmann, Nucl. Phys. B **427**, 475 (1994), [hep-ph/9405388].
127. S. Catani and F. Hautmann, Phys.Lett. **B315**, 157 (1993).
128. M. Deak, F. Hautmann, H. Jung, and K. Kutak, *Forward-Central Jet Correlations at the Large Hadron Collider* (2010), [arXiv:1012.6037].
129. F. Hautmann and H. Jung, Nuclear Physics B **883**, 1 (2014), [arXiv:1312.7875].
130. H. Jung and F. Hautmann (2012), [arXiv:1206.1796].
131. G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. **B725**, 223 (2013), [arXiv:1305.4192].
132. G. Aad *et al.* [ATLAS Collaboration], JHEP **1406**, 112 (2014), [arXiv:1404.1212].
133. G. Aad *et al.* [ATLAS Collaboration], JHEP **1405**, 068 (2014), [arXiv:1402.6263].
134. G. Aad *et al.* [ATLAS Collaboration], Eur.Phys.J. **73**, 2509 (2013), [arXiv:1304:4739].
135. G. Aad *et al.* [ATLAS Collaboration], Tech. Rep. ATL-PHYS-PUB-2013-018, CERN, Geneva (2013).
136. F. Aaron *et al.* [H1 Collaboration], JHEP **1209**, 061 (2012), [arXiv:1206.7007].
137. H. Abramowicz *et al.* [H1 and ZEUS Collaborations], Eur. Phys. J. **C73**, 2311 (2013), [arXiv:1211.1182].
138. H. Abramowicz *et al.* [ZEUS Collaboration] (2014), [arXiv:1405.6915].
139. A. Luszczak and H. Kowalski, Phys.Rev. **D89**, 074051 (2013), [arXiv:1312.4060].
140. R. Sadykov (2014), [arXiv:1401.1133].
141. S. Dooling, F. Hautmann, and H. Jung, Phys.Lett. **B736**, 293 (2014), [arXiv:1406.2994].
142. *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.

143. *ATLAS NNLO epWZ12*, available via: http://lhapdf.hepforge.org/pdfsets.

144. J. Butterworth, G. Dissertori, S. Dittmaier, D. de Florian, N. Glover, *et al.* (2014), [arXiv:1405.1067].

145. J. L. Abelleira Fernandez *et al.* [LHeC Study Group], Journal of Phys. **G**, 075001 (2012), [arXiv:1206.2913].