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

HERAFitter

Open Source QCD Fit Project

HERAFitter **team**

[1] Version 0.1 (svn 1349)

Received: date / Accepted: date

Abstract The HERAFitter project is presented which provides a framework for Quantum Chronodynamics (QCD) analyses related to the proton structure in the context of multi-processes and multi-experiments. Based on the concept of factorisable nature of the cross sections into universal parton distribution functions (PDFs) and process dependent partonic scattering cross sections, HERAFitter allows determination of PDFs from the various hard scattering measurements. The main processes and data sets that are currently included are Deep-Inelastic-Scattering (DIS) in ep collisions at HERA and Drell Yan (DY), jet and top quark production in pp ($p\bar{p}$) collisions at the LHC (Tevatron). A large number of theoretical and methodological options is available within HERAFitter via interfaces to external software packages.

Keywords PDFs · QCD · Fit

1 Introduction

In the era of the Higgs discovery and scrupulous searches for signals of new physics at the LHC it is crucial to have accurate Standard Model (SM) predictions for many hard scattering processes such as the Higgs production at the LHC. A most common approach to calculate the SM cross sections for such reactions is to use perturbative QCD collinear factorisation:

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

Here the cross section $\sigma^{pp\to H+X}$ for inclusive Higgs production is expressed as a convolution of Parton Distribution Functions (PDF) f_a and f_b with the partonic cross section $\hat{\sigma}^{ab\to H+X}$. The PDFs describe the probability of finding a specific parton a (b) in the first (second) proton carrying the fraction x_1 (x_2) of its momentum. The sum in Eq. 1 in indeces a and b is over all different kind of partons, i.e. gluons and the various quarks and antiquarks flavours, that are considered as the constituents of the proton. Both the PDFs and the partonic cross section depend on the strong coupling constant α_s, the factorisation and renormalisation scales, μ_F and μ_R, respectively. The partonic cross sections are calculable in pQCD while the PDFs cannot be determined solely with pQCD but are assumed to be universal. This permits the use of different scattering reactions to constrain the PDFs; in particular one can use specific reaction data for determining the PDFs and then take these PDFs for predicting other processes via Eq. 1.

Key information on the PDFs is provided by the Deep Inelastic Scattering (DIS) data from the ep collider HERA. For instance, the gluon density relevant for calculating the dominant gluon-gluon fusion contribution to the Higgs production at the LHC can be accurately determined from the HERA data alone. Specific data from the Tevatron $p\bar{p}$ and the LHC pp collider can help to further constrain the PDFs. The most sensitive processes at the $p\bar{p}$ colliders are Drell Yan production, W and Z asymmetries, top quark production, and jet production.

HERAFitter represents a QCD analysis framework that aims at determining precise PDFs by integrating all the PDF sensitive information from HERA, Tevatron and the LHC. The processes that are currently included in HERAFitter framework are listed in Tab. 1. The basic functionality of HERAFitter is shown in Fig. 1 and consists of four parts: **needs to update figure!**

Input data: All relevant cross section data from the various reactions are stored internally in HERAFitter with the full information on their uncorrelated and correlated uncertainties.

Data	Type	Reaction	Theory calculation
HERA	DIS NC	$ep \to ep$	QCDNUM, RT, ACOT
HERA	DIS CC	$ep \to \nu_e p$	QCDNUM, RT, ACOT
HERA	DIS jets	$ep \to eX$	FastNLO (NLOJet++)
HERA	DIS heavy quark	$ep \to ep$	ZM (QCDNUM), RT, ACOT, FFNS (ABM,QCDNUM)
Fixed Target	DIS NC	$ep \to ep$	ZM (QCDNUM), RT, ACOT
Tevatron, LHC	Drell Yan	$pp(\bar{p})$	APPLGRID (MCFM)
Tevatron, LHC	W charge asym	$pp(\bar{p})$	APPLGRID (MCFM)
Tevatron, LHC	top	$pp(\bar{p})$	APPLGRID (MCFM), HATHOR
Tevatron, LHC	jets	$pp(\bar{p})$	APPLGRID (NLOJet++) FastNLO (NLOJet++)
LHC	DY+heavy quark	$pp(\bar{p})$	APPLGRID (MCFM)

Table 1 The list of processes available in `HERAFitter` .

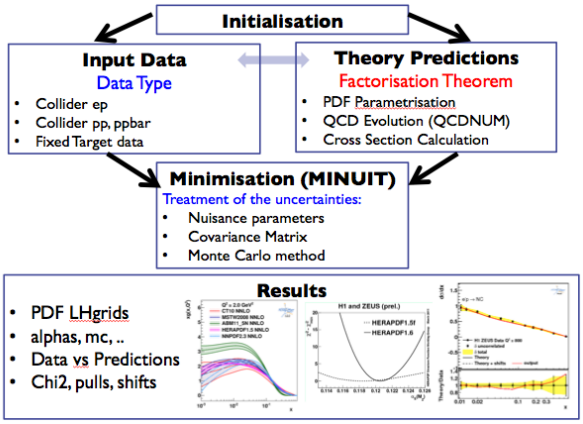

Fig. 1 Schematic structure of the `HERAFitter` program.

Theory predictions: Predictions are obtained relying on factorisation approach (Eq. 1). PDFs are parametrised at a starting scale Q_0 by a chosen functional form with a set of free parameters **p**. Then they evolved from Q_0 to the scale of the measurement using Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [1, 2, 3, 4, 5] evolution equations as implemented in QCDNUM [6], and then multiplied (Eq. 1) with the hard parton cross sections calculated by a specific theory program (also listed in Tab. 1).

Minimization: PDFs are extracted from a least square fit by constructing a χ^2 from the input data and the theory prediction. The χ^2 is minimized iteratively with respect to the PDF parameters using the MINUIT[7] program.

Results: The fitted parameters **p** and their estimated uncertainties are produced. The resulting PDFs are provided in a ready to be used LHAPDF library format and can be graphically displayed at arbitrary scales with their one sigma uncertainties bands. To demonstrate the fit consistency, plots are provided in which the input data are compared to the fitted theory predictions.

The `HERAFitter` program permits determination of the PDFs from the variours measurements of the cross section at ep, $p\bar{p}$ or pp colliders. It includes various options for theoretical models and choices to account for the experimental uncertainties. Therefore, this project represents not only an ideal environment for benchmarking studies, but also a unique support for the QCD interpretation of analyses within the LHC experiments, as already demonstrated by several publicly available results using the `HERAFitter` framework [8, 9, 10, 11, 12, 13, 14].

The outline of this paper is as follows. Section 2 discusses the various processes and corresponding theoretical calculations that are available in `HERAFitter` . Section 3 elucidates the methodology of determining PDFs through fits based on various χ^2 definitions used in the minimisation procedure. Specific applications of the package are given in section 4.

2 Theoretical Input

The PDFs are determined by measuring the agreement between experimental data and corresponding theory models. Models which are available in `HERAFitter` for various processes are described in the following text.

2.1 Deep Inelastic Scattering Formalism and Schemes

DIS data provide the tightest constraints on the PDFs so far. Deep inelastic scattering is the lepton scattering on the constituents of the proton by a virtual exchange of a neutral (NC) or charged (CC) boson and, as a result, a scattered lepton and a multihadronic final state are produced. The DIS kinematic variables are the negative squared four-momentum of the exchange boson, Q^2, the scaling variable x, which can be related in the parton model to the fraction of momentum carried by the struck quark, and the inelasticity parameter y, which is the fraction of the energy transferred to the hadronic vertex.
The NC (and similarly CC) cross section can be expressed in terms of 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 $Y_\pm = 1 \pm (1-y)^2$ with y being the inelasticity. The structure function $\tilde{F}_2$ is the dominant contribution to the cross section, $x\tilde{F}_3$ is important at high Q^2 and $\tilde{F}_L$ is sizable only at high y. In the framework of perturbative QCD the structure functions are directly related to the parton distribution functions, i.e. in leading order (LO) F_2 is the momentum sum of quark and anti-quark distributions, $F_2 \approx x\sum e_q^2(q+\bar{q})$, and xF_3 is related to their difference, $xF_3 \approx x\sum 2e_q a_q(q-\bar{q})$. At higher orders, terms related to the gluon density distribution ($\alpha_s g$) appear.
In analogy to neutral currents, the inclusive CC ep cross section can be expressed in terms of structure functions and in

LO the e^+p and e^-p cross sections are sensitive to different quark densities:

$$\begin{aligned} e^+ &: \ \tilde{\sigma}_{CC}^{e^+p} = x[\bar{u}+\bar{c}] + (1-y)^2 x[d+s] \\ e^- &: \ \tilde{\sigma}_{CC}^{e^-p} = x[u+c] + (1-y)^2 x[\bar{d}+\bar{s}]. \end{aligned} \quad (3)$$

The QCD predictions for the DIS structure functions are obtained by convoluting the PDFs with the coefficient functions calculated using various schemes, i.e. the general mass Variable-Flavour number (GM-VFN) [15] schemes or the Fixed-Flavour number (FFN) [16, 17, 18], The following VFN schemes with various treatments for the heavy quark thresholds are considered in `HERAFitter` : The Thorne Roberts (TR) scheme with its variants at NLO and NNLO [19, 20] as provided by the MSTW group, the ACOT scheme with its variants at LO and NLO as provided by the CTEQ group. In addition, the zero-mass variable flavour number scheme (ZM-VFNS) where heavy quark densities are included in the proton for $Q^2 >> m_h^2$ but are treated as massless in both the initial and final states can be used in `HERAFitter` . The FFN scheme is available via the QCDNUM implementation and via the `OPENQCDRAD` [21] interface. Each of these schemes is briefly discussed below.

GM-VFN Thorne-Roberts scheme: The Thorne-Roberts (TR) scheme smoothly connect the two regions: scales below ($Q^2 < m_h^2$) and scales much above the heavy quark scale threshold ($Q^2 >> m_h^2$). There are two different variants of the TR schemes: TR standard (as used in MSTW PDF sets [20, 22]) and TR optimal [23], with a smoother transition across the heavy quark mass scales and both of them are accessible within the `HERAFitter` package. The calculations are available to NLO and NNLO. In addition, a fast version of the scheme is available (i.e. RT FAST) by using the $k-factor$ technique where $k-factors$ are defined as the ratio between massless (accessed by QCDNUM) and massive scheme. The $k-factors$ are only calculated for the PDF parameters at the first fit iteration hence, the recommended is the full TR scheme.

GM-VFN ACOT scheme:
The Aivazis-Collins-Olness-Tung scheme belongs to the group of VFN factorisation schemes that use the renormalization method of Collins-Wilczek-Zee (CWZ) [24]. This scheme involves a mixture of the $\overline{\text{MS}}$ scheme for light and heavy (when the factorisation scale is larger than the heavy quark mass) partons and the zero-momentum subtraction renormalisation scheme for graphs with heavy quark lines (if the factorisation scale is smaller than the mass of the heavy quark threshold).
Within the ACOT package, different variants of the ACOT scheme are available: ACOT-Full, S-ACOT-χ, ACOT-ZM, $\overline{\text{MS}}$ at LO and NLO. For the longitudinal structure function higher order calculations are also available.
The ACOT-Full implementation takes into account the quark masses and it reduces to ZM $\overline{\text{MS}}$ scheme in the limit of masses going to zero, but it has the disadvantage of being quite slow. Therefore the $k-factor$ technique has been adopted within the `HERAFitter` framework which are defined in two different ways: as the ratio between same order calculations but massless vs massive (i.e. NLO (ZM-VFNS)/NLO (ACOT), or as the ratio between LO (massless)/NLO (massive), both giving similar results.

Fixed-Flavour Number Scheme:
In the FFN scheme only the gluon and the light quarks are considered as partons within the proton and massive quarks are produced perturbatively in the final state. In `HERAFitter` this scheme can be accessed via the QCDNUM implementation or interface to the open-source code OPENQCDRAD (ABM) [21]. The later implementation also includes the running mass definition of the heavy quark mass [25]. This scheme has the advantage of reducing the sensitivity of the DIS cross sections to higher order corrections, and improving the theoretical precision of the mass definition. In QCDNUM, the calculation of the heavy quark contributions to DIS structure functions are available at NLO and only electromagnetic exchange contributions are taken into account. In the ABM implementation, the QCD corrections to the massive Wilson coefficients up to the currently best known approximate NNLO for the NC heavy-quark production [26] and up to NLO for the CC case are available.

The calculations of higher-order electroweak corrections to DIS scattering at HERA are performed in the on-shell scheme where the gauge bosons masses M_W and M_Z are treated symmetrically as basic parameters together with the top, Higgs and fermion masses.
In the `HERAFitter` the electroweak corrections for the DIS process are based on the EPRC package [27]. The code provides the running of α using the most recent parametrisation of the hadronic contribution to Δ_α [28], as well as an older one from Burkhard [29].

2.2 Diffractive PDFs

Similar 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 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 process, the proton vertex factorisation approach is assumed where the diffractive DIS is mediated by the exchange of hard Pomeron and a secondary Reggeon. The factorisable pomeron picture has proved remarkably successful for the description of most of these data.

In addition to x, Q^2 and the squared four-momentum transfer t (the undetected momentum transfer to the proton system), the mass M_X of the diffractively produced final state provides a further degree of freedom. 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

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

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

The diffractive PDFs in `HERAFitter` are implemented following the prescription of ZEUS publication [30] and can be used to reproduce the main results.

2.3 Alternative to DGLAP DIS models

Different approaches that are alternative to DGLAP formalism can be used to analyse DIS data in `HERAFitter` . Those include several dipole models and transverse momentum dependent, or unintegrated PDFs, uPDFs. Both approaches are discussed below.

2.3.1 DIPOLE models

The dipole picture provides an alternative approach to the virtual photon-proton scattering at low x because it 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 [31]. 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 virtual photon fluctuates into a quark-antiquark pair and subsequently interacts with the target, and 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 [32], the colour glass condensate approach to the high parton density regime Iancu-Itakura-Munier (IIM) model [33] and a modified GBW model which takes into account the effects of DGLAP evolution Bartels-Golec-Kowalski (BGK) [34].

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

$$\sigma_{\rm dip}(x, r^2) = \sigma_0\left(1 - \exp\left[-\frac{r^2}{4R_0^2(x)}\right]\right), \qquad (6)$$

here r corresponds to the transverse separation between the quark and the antiquark, and R_0^2 is an x dependent scale parameter which has corresponds to a saturation radius, $R_0^2(x) = (x/x_0)^{\lambda}$. The free fitted parameters are the cross-section normalisation σ_0 as well as x_0 and λ.

IIM model: The IIM model assumes an improved expression for the dipole cross section which is based on the Balitsky-Kovchegov equation [35]. The explicit formula for $\sigma_{\rm dip}$ can be found in [33]. The free 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

$$\sigma_{\rm dip}(x, r^2) = \sigma_0\left(1 - \exp\left[-\frac{\pi^2 r^2 \alpha_s(\mu^2) xg(x,\mu^2)}{3\sigma_0}\right]\right). \qquad (7)$$

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

BGK model with valence quarks:
The dipole models are valid in the low-x region only, where the valence quark contribution is small, of the order of 5%. The new HERA F_2 data have a precision which is better than 2 %. Therefore, in the `HERAFitter` the contribution of the valence quarks is taken from the PDF fits and added to the original BGK model, this is uniquely possible within the `HERAFitter` framework.

2.3.2 Transverse Momentum Dependent (unintegrated PDF) with CCFM

Here another alternative approach to collinear DGLAP evolution is presented. In high energy factorization [36] the measured cross section is written as a convolution of the partonic

cross section $\hat{\sigma}(k_t)$, which depends on the transverse momentum k_t of the incoming parton, with the k_t-dependent parton distribution function $\tilde{\mathcal{A}}(x,k_t,p)$ (transverse momentum dependent (TMD) or unintegrated uPDF):

$$\sigma = \int \frac{dz}{z} d^2k_t \hat{\sigma}(\frac{x}{z},k_t)\tilde{\mathcal{A}}(x,k_t,p) \tag{8}$$

Generally, the evolution of $\tilde{\mathcal{A}}(x,k_t,p)$ can proceed via the BFKL, DGLAP or via the CCFM evolution equations. In HERAFitter an extension of the CCFM [37, 38, 39, 40] evolution has been implemented. Since the evolution cannot be easily obtained in a closed form, first a kernel $\tilde{\mathcal{A}}(x'',k_t,p)$ is determined from the MC solution of the CCFM evolution equation, and is then folded with the non-perturbative starting distribution $\mathcal{A}_0(x)$ [41]:

$$\begin{aligned} x\mathcal{A}(x,k_t,p) &= x\int dx' \int dx'' \mathcal{A}_0(x)\tilde{\mathcal{A}}\left(x'',k_t,p\right)\delta(x'\cdot x''-x) \\ &= \int dx' \int dx'' \mathcal{A}_0(x)\tilde{\mathcal{A}}\left(x'',k_t,p\right)\frac{x}{x'}\delta(x''-\frac{x}{x'}) \\ &= \int dx' \mathcal{A}_0(x')\cdot\frac{x}{x'}\tilde{\mathcal{A}}\left(\frac{x}{x'},k_t,p\right). \end{aligned} \tag{9}$$

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

The calculation of the cross section according to Eq.(8) involves a multidimensional Monte Carlo integration which is time consuming and suffers from numerical fluctuations, and therefore cannot be used directly in a fit procedure. Instead the following procedure is applied:

$$\begin{aligned} \sigma_r(x,Q^2) &= \int_x^1 dx_g \mathcal{A}(x_g,k_t,p)\hat{\sigma}(x,x_g,Q^2) \\ &= \int_x^1 dx' \mathcal{A}_0(x')\cdot\tilde{\sigma}(x/x',Q^2). \end{aligned} \tag{10}$$

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

$$x\mathcal{A}_0(x,k_t) = Nx^{-B_g}\cdot(1-x)^{C_g}(1-D_g x) \tag{11}$$

with free parameters N, B_g, C_g, D_g.

The calculation of the ep cross section follows eq.(8), with the off-shell matrix element including quarks masses taken from [36] in its implementation in CASCADE [42]. In addition to the boson gluon fusion process, valence quark initiated $\gamma q \to q$ processes are also included, with the valence quarks taken from [43].

2.4 Drell Yan processes

The Drell Yan (DY) process constrain all different quark combinations providing valuable information about PDFs. Presently, the calculations of the DY processes are known for many observables up to NNLO order, for example, MCFM [44] package is available for NLO calculations, FEWZ [45] and DYNNLO [46] for NLO and NNLO. Due to the complicated nature of these calculation involving an increased number of diagrams with each additional order, they are too slow to be used iteratively in a fit. There are several methods available to speed-up such calculations two of which are implemented into HERAFitter : the $k-factor$ approximation from lower (LO) to higher order (NLO) and the so-called grid technique using an interface to the APPLGRID, both of them shortly described below.

$k-factor$ **technique:** The leading order DY triple differential cross section in invariant mass M, boson rapidity y and CMS lepton scattering angle $\cos\theta$, for the neutral current, can be written as [47, 48]:

$$\frac{d^3\sigma}{dMdyd\cos\theta} = \frac{\pi\alpha^2}{3MS}\sum_q P_q\left[F_q(x_1,Q^2)F_{\bar{q}}(x_2,Q^2)+(q\leftrightarrow\bar{q})\right], \tag{12}$$

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

$$\begin{aligned} P_q &= e_l^2 e_q^2(1+\cos^2\theta) \\ &+ e_l e_q \frac{2M^2(M^2-M_Z^2)}{\sin^2\theta_W\cos^2\theta_W\left[(M^2-M_Z^2)^2+\Gamma_Z^2M_Z^2\right]} \\ &\left[aA_q(1+\cos^2\theta)+2bB_q\cos\theta\right] \\ &+ \frac{M^4}{\sin^4\theta_W\cos^4\theta_W\left[(M^2-M_Z^2)^2+\Gamma_Z^2M_Z^2\right]} \\ &\left[(a^2+b^2)(A_q^2+B_q^2)(1+\cos^2\theta)+8abA_qB_q\cos\theta\right]. \end{aligned} \tag{13}$$

Here θ_W is the Weinberg angle, M_Z and Γ_Z are Z boson mass and width, a,b,A_q,B_q,e_l,e_q are electro-weak couplings.

The expression for charged current scattering has a simpler form.

$$\begin{aligned} \frac{d^3\sigma}{dMdyd\cos\theta} &= \frac{\pi\alpha^2}{48S\sin^4\theta_W}\frac{M^3(1-\cos\theta)^2}{(M^2-M_W^2)+\Gamma_W^2M_W^2} \\ &\sum_{q_1,q_2} V_{q_1q_2}^2 F_{q_1}(x_1,Q^2)F_{q_2}(x_2,Q^2), \end{aligned} \tag{14}$$

where $V_{q_1q_2}$ is the CKM quark mixing matrix and M_W and Γ_W are W boson mass and decay width.
The simple form of these expressions allows the calculation of integrated cross sections without utilization of Monte-Carlo techniques which often introduce statistical fluctuations. In both neutral and charged current expressions the parton distribution functions factorise as functions dependent only on boson rapidity y and invariant mass M. The integral in $\cos\theta$ can be computed analytically and integrations in y and M can be performed

with the Simpson method. The $\cos\theta$ parts are kept in the equation explicitly because their integration is asymmetric for data in lepton η bins and also because of the need to apply lepton $p_\perp$ cuts.
The fact that PDF functions factorise, allows high speed calculations when performing parameter fits over lepton rapidity data. In this case the factorised part of the expression which is independent of PDFs can be calculated only once for all minimisation iterations. The leading order code in `HERAFitter` package implements this optimisation and uses fast convolution routines provided by QCDNUM. Currently the full width LO calculations are optimised for lepton pseudorapidity and boson rapidity distributions with the possibility to apply lepton $p_\perp$ cuts. This flexibility allows the calculations to be performed within the phase space corresponding to the available measurement.
The calculated leading order cross sections are multiplied by $k-factors$ to obtain predictions at NLO.

APPLGRID technique: The APPLGRID [49] package allows the fast computation of NLO cross sections for particular processes for arbitrary sets of proton parton distribution functions. The package implements calculations of DY production as well as jet production in $pp(\bar{p})$ collisions and DIS processes.
The approach is based on storing the perturbative coefficients of NLO QCD calculations of final-state observables measured in hadron colliders in look-up tables. The PDFs and the strong couplings are included during the final calculations, e.g. during PDF fitting. The method allows variation of factorization and renormalization scales in calculations.
The look-up tables (grids) can be generated with modified versions of MCFM parton level generator [50, 51] or NLOjet++ [52] code for the jet production. The model input parameters are in fact pre-set following the MCFM input steering card, while binning and definitions of the cross section observables are set in the APPLGRID code. The grid parameters, Q^2 binning and interpolation orders are also defined in the code.
APPLGRID constructs the grid tables in two steps: *(i)* exploration of the phase space in order to optimize the memory storage and *(ii)* actual grid construction in the phase space corresponding to the requested observables. Afterwards the NLO cross sections are restored from the grids using externally provided PDFs, α_S, factorization and renormalization scales. QCD NNLO $k-factors$ can be applied if requested.

2.5 Cross Sections for $t\bar{t}$ production in pp or $p\bar{p}$ collisions

Top-quark pairs ($t\bar{t}$) are mainly produced at hadron colliders via gg fusion and $q\bar{q}$ annihilation thus providing possibility to constrain the gluon density in the proton. In `HERAFitter` the program HATHOR [53] is interfaced which allows the calculation of the expected total $t\bar{t}$ cross section at $p\bar{p}$ and pp colliders up to approximate NNLO accuracy. Version 1.3 of HATHOR includes the exact NNLO for $q\bar{q} \to t\bar{t}$ [54] as well as a new high-energy constraint on the approximate NNLO calculation obtained from soft-gluon resummation [55]. The default choice for renormalization and factorization scale in $t\bar{t}$ production is the top-quark mass, m_t. The pole mass scheme is typically employed for m_t but HATHOR also supports calculations in the $\overline{\text{MS}}$ scheme.

2.6 Jet production

Similarly to DY case, the calculation of higher order jet cross sections is very demanding in terms of computing power. Therefore, in order to enable the inclusion of jet-cross section measurements in PDF and α_s fits, the perturbative coefficients have to be pre-computed in a PDF and α_s independent way. For this purpose, two grid tools are interfaced to the `HERAFitter` , APPLGRID (see 2.4) and FastNLO.

FastNLO technique: The fastNLO project [56, 57, 58] uses multi-dimensional interpolation techniques to convert the convolutions of perturbative coefficients with parton distribution functions and the strong coupling into simple products. The perturbative coefficients are calculated by the `NLOJET++` program [59] where calculations for jet-production in DIS [60] as well as in hadron-hadron collisions [52, 61] with threshold-corrections of $\mathcal{O}$(NNLO) for inclusive jet cross sections [62] are available.
The fastNLO libraries are included in the `HERAFitter` package and in order to include a new measurement into the PDF fit, only the fastNLO tables have to be specified. These tables include all necessary information about the perturbative coefficients and the calculated process for all bins of a certain dataset. The fastNLO tables are conventionally calculated for multiple factors of the factorization scale, and the renormalization scale factor can be chosen freely. Some of the fastNLO tables already allow for the free choice [58] of the renormalization and the factorization scale as a function of two pre-defined observables. The evaluation of the strong coupling constant, which enters the cross section calculation, is taken consistently from the QCDNUM evolution code.

3 Methodology

Nowadays, there are considerable number of choices available when performing a QCD fit analysis which require a careful investigation (i.e. input parametrisation form, threshold values for heavy quarks, various theory prescriptions, method of minimisation, interpretation of uncertaintes and so on). It is desirable to be able to discriminate or quantify the effect of chosen ansatz, ideally within a framework that provides such cross checks and `HERAFitter` is optimaly designed for such tests. The methodology employed by `HERAFitter` relies on a flexible and modular framework that allows for independent integration of the state-of-the-art techniques, either related to the inclusion of a new theoretical calculation, or to a new approaches to treat uncertainties.

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

In addition, the reweighting method - an alternative approach to a complete QCD fit, available in the `HERAFitter` is also described in this section. It can provide for an estimate of an impact of new data, as advocated alreay by the NNPDF collaboration [63, 64]. The method has been extended to work not only on the replica method, but also on the eigenvectors (as introduced by MSTW group [65]).

An important factor for a fesible QCD fit which is performed iteratively through the χ^2 minimisation process, represents the performance in terms of how long a calculation takes for each given data point. In `HERAFitter` this is achieved by optimising the time of calculations relyng on innovative techniques such as cache option, fast evolution kernels, grid techniques making the platform a practical engine for iterative usage.

3.1 Functional Forms for PDF parametrisation

The PDFs are parametrised at the starting scale bellow the charm mass threshold, chosen by the user. Various functional forms can be tested using desired number of free parameters to be extracted through the fit:

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

$$xf(x) = Ax^B(1-x)^C P_i(x), \tag{15}$$

Standard forms are commonly used by PDF groups. The parametrised PDFs at HERA are the valence distributions xu_v and xd_v, the gluon distribution xg, and the u-type and d-type sea $x\bar{U}$, $x\bar{D}$, where $x\bar{U} = x\bar{u}$, $x\bar{D} = x\bar{d} + x\bar{s}$. The $P_i(x)$ for the HERAPDF style takes the simple form of $(1+Dx+Ex^2)$ with additional constraints due to flavour decomposition insensitivity for $q\bar{s}$ fromother light sea quark contributions. For the CTEQ style, $P_i(x)$ takes te form of $e^{a_3 x}(1+e^{a_4}x+e^{a_5}x^2)$.

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

$$xf(x) = x^{p-b\log(x)}(1-x)^{q-\log(1-x)}. \tag{16}$$

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

Chebyshev Polynomials:

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

$$xg(x) = A_g\,(1-x) \sum_{i=0}^{N_g-1} A_{g_i} T_i \left(-\frac{2\log x - \log x_{min}}{\log x_{min}} \right) \tag{17}$$

$$xS(x) = (1-x) \sum_{i=0}^{N_S-1} A_{S_i} T_i \left(-\frac{2\log x - \log x_{min}}{\log x_{min}} \right). \tag{18}$$

Here the sum over i runs up to $N_{g,S} = 15$ order Chebyshev polynomials of the first type T_i for the gluon, g, and sea-quark, S, density, respectively. The normalisation A_g is given by the momentum sum rule.

The advantages of this parametrisation are that the momentum sum rule can be evaluated analytically and that for $N \geq 5$ the fit qulaity is already similar to a standard Regge-inspired parametrisation with a similar number of parameters.

External PDFs: `HERAFitter` provides also possibility to access the external PDF sets, which can then be used in constructing the theoretical predictions rather than the ones restricted only to the `HERAFitter` framework. This is possible via interface to LHAPDF which commonly provides access to the global PDF sets available at LO, NLO or NNLO evolved either locally through the `HERAFitter` or taken as provided by the LHAPDF grid.

3.2 Chisquare representation

The PDF parameters are extracted from the χ^2 minimization process. There are various forms to represent the χ^2 function, i.e. covariance matrix or decomposed into nuisance parameters. In addition, there are various methods in dealing

with the 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 for the case when experimental uncertainties are given in a covariance matrix over data bins $C_{i,j}$ can be expressed in the following form:

$$\chi^2(m) = \sum_{i,j}(m_i - \mu_i)C_{ij}^{-1}(m_j - \mu_j). \quad (19)$$

The χ^2 function depends on the theory parameters m^i (denoted as the vector m). The covariance matrix can be decomposed in statistical, uncorrelated and correlated systematic contributions:

$$C_{ij} = C_{ij}^{stat} + C_{ij}^{uncor} + C_{ij}^{sys}. \quad (20)$$

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

Nuisance Parameters Representation:

$$\chi^2(m,b) = \sum_i \frac{\left[m^i - \sum_j \gamma_j^i m^i b_j - \mu^i\right]^2}{\delta_{i,\mathrm{stat}}^2 \mu^i \left(m^i - \sum_j \gamma_j^i m^i b_j\right) + \left(\delta_{i,\mathrm{uncor}} m^i\right)^2} + \sum_j b_j^2. \quad (21)$$

Here μ^i is the measured central value at a point i with relative statistical $\delta_{i,stat}$ and relative uncorrelated systematic uncertainty $\delta_{i,unc}$. Further, γ_j^i quantifies the sensitivity of the measurement μ^i at the point i to the systematic source j. The function χ^2 depends in addition on the set of systematic uncertainties b_j (b). This definition of the χ^2 function takes into account that systematic uncertainties are proportional to the central values (multiplicative errors), whereas the statistical errors scale with the square roots of the expected number of events.

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

3.3 Treatment of the Experimental Uncertainties

HERAFitter provides three methods in assessing the experimental uncertainties on PDFs: Hessian, Offset, and Monte Carlo method, which are described below.

Hessian method: The technique developed by [66] presents an estimate of PDF uncertainties reflecting the experimental precision of used data in the QCD fit by examining the behaviour of χ^2 in the neighborhood of the minimum. This is known as Hessian or error matrix method. The Hessian matrix is build by the second derivatives of χ^2 at the minimum. The PDF eigenvectors are obtained through an iterative procedure used to diagonalise the Hessian matrix and rescale the eigenvectors to adapt the step sizes to their natural scale.

Offset method:

There is another method to propagate the systematic experimental uncertainties from the measurements to PDFs [67], which has the practical advantage that does not require the inversion of a large measurement covariance matrix. It uses also the χ^2 function for the central fit for which only uncorrelated uncertainties are taken into account to get the best PDF parameters. However, the goodness of fit can no longer be judged since correlated uncertainties are ignored. In the offset method, the systematic uncertainties on PDFs are estimated from fits where each systematic source is offset by its given one sigma shift to the central cross section values, after which the resulting deviation from the central PDF parameters are added in quadrature.

In most cases, the uncertainties estimated through offset method are larger than these from the Hessian method, as offset method is not so efficient.

Monte Carlo method: The PDF uncertainties can be estimated using a Monte Carlo technique [68, 69]. The method consists in preparing replicas of data sets by allowing the central values of the cross sections to fluctuate within their systematic and statistical uncertainties taking into account all point-to-point correlations. The preparation of the data is repeated for a large N (> 100 times) and for each of these replicas a NLO QCD fit is performed to extract the PDF set. The PDF central values and uncertainties are estimated using the mean values and RMS over the replicas.

3.4 Treatment of the Theoretical Input Parameters

The results of a QCD fit depends not only on the input data but also on the input theoretical ansatz, which is also uncertain. Nowadays, the modern PDFs try to address the impact of this ansatz on the resulting PDFs by assessing an uncertainty on the choice of the initial parameter, such as mass of charm m_c, mass of the bottom quarks m_b. Another important input is the choice of the functional form for the PDFs at the starting scale. For this, HERAFitter provides a series of choices ranging from simple functional forms to more complex forms such as Chebyshev Polynomials with larger

flexibility. Larger flexibility usually requires some regularisation methods in order for the results to be physical.

3.5 Performance Optimisation

The above mentioned features make HERAFitter a powerful project that encapsulates the state of the art developments from struggles on reaching atmost experimental precision to the most up-to-date theory developments.

The performance of the HERAFitter code is greatly improved with several special build-in options including the $k-factor$ techniques (described in section 2.1), the grid techniques for the fast calculational of cross sections of particular processes for arbitrary sets of PDFs (sections 2.4 and 2.6) and usage of the openMP (Open Multi-Processing) interface which allows parallel applications of some of the heavy flavour scheme theory prediction calculations in DIS.

4 Applicability of HERAFitter

HERAFitter has been successfully integrated in the high energy community as a much needed mean in providing understanding and interpretation of new measurements in context of QCD theory, a field limitted by the PDF precision. The applicability of the HERAFitter platform is reflected in usage and extraction of the theory parameters from data such as PDFs, strong coupling, heavy quark masses, quantitative asssessment of the fit quality with full detail information in provided experimental and theoretical uncertainties. Importantly, it also produces PDF grids that could be used further for an improved and more discriminating powers of predictions for beyond SM processes, as well as for impact studies of possible future colliders using pseudo-data.

The series of PDF grids extracted so far extend from the QCD analyses performed at HERA (HERAPDF series [70]), now extented to the LHC with published measurements from ATLAS [8] (the first ever ATLAS PDF sets [71]).

A list of new results that have been based on HERAFitter platform includes the SM processes studied at the LHC in inclusive W and Z production [8, 10, 11], in inclusive jets [9, 12], and also extended to the top measurements (which benefit from recent theoretical advancements). At HERA, the results of QCD analyses using HERAFitter are published in inclusive H1 measurements [13] and the recent combination of charm production measurements in DIS [14]. The HERAFitter framework also provides an unique possibility to make impact studies for future colliders which can be illustrated by the preformed QCD studies exploring the potential of the LHeC data [72].

5 Summary

The HERAFitter project is a unique framework for QCD analyses to study the proton structure which incorporates the most of the experimental data processes from hadron colliders sensitive to PDFs and variety of the up-to date theory calculations. It has flexible modular structure and contains many different useful tools for PDF interpretation. HERAFitter is the first open source platform optimal for various benchmarking studies and is excessively used by various experimental and theoretical high energy physics communities.

References

1. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).
2. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 675 (1972).
3. L. N. Lipatov, Sov. J. Nucl. Phys. **20**, 94 (1975).
4. Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
5. G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
6. M. Botje (2010), http://www.nikef.nl/h24/qcdnum/index.html, [arXiv:1005.1481].
7. F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).
8. G. Aad *et al.* [ATLAS Collaboration], Phys. Rev. Lett. **109**, 012001 (2012), [arXiv:1203.4051].
9. G. Aad *et al.* [ATLAS Collaboration], Eur.Phys.J. **73**, 2509 (2013), [arXiv:1304:4739].
10. G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. **B725**, 223 (2013), [arXiv::1305.4192].
11. S. Chatrchyan *et al.* [CMS Collaboration], submitted to Phys. Rev. **D** (2014), [arXiv:1312.6283].
12. S. Chatrchyan *et al.* [CMS Collaboration], CMS PAS **SMP-12-028** (2014).
13. F. Aaron *et al.* [H1 Collaboration], JHEP **1209**, 061 (2012), [arXiv:1206.7007].
14. H. Abramowicz *et al.* [H1 and ZEUS Collaborations], Eur. Phys. J. **C73**, 2311 (2013), [arXiv:1211.1182].
15. R. Demina, S. Keller, M. Kramer, S. Kretzer, R. Martin, *et al.* (1999), [hep-ph/0005112].
16. E. L. *et al.*, Phys. Lett. **B291**, 325 (1992).
17. E. L. *et al.*, Nucl. Phys. **B392**, 162, 229 (1993).
18. S. Riemersma, J. Smith, and van Neerven. W.L., Phys. Lett. **B347**, 143 (1995), [hep-ph/9411431].
19. R. S. Thorne and R. G. Roberts, Phys. Rev. D **57**, 6871 (1998), [hep-ph/9709442].
20. R. S. Thorne, Phys. Rev. **D73**, 054019 (2006), [hep-ph/0601245].
21. S. Alekhin, *OPENQCDRAD*, a program description and the code are available via: http://www-zeuthen.desy.de/∼alekhin/OPENQCDRAD.

22. A. D. Martin, Eur. Phys. J. C **63**, 189 (2009).
23. R. S. Thorne (2012), [arXiv:1201.6180].
24. J. C. Collins, Phys.Rev. **D58**, 094002 (1998), [hep-ph/9806259].
25. S. Alekhin and S. Moch, Phys. Lett. **B699**, 345 (2011), [arXiv:1011.5790].
26. K. H., N. Lo Presti, S. Moch, and A. Vogt, Nucl.Phys. **B864**, 399 (2012).
27. H. Spiesberger, Private communication.
28. Jegerlehner, Proceedings, LC10 Workshop **DESY 11-117** (2011).
29. H. Burkhard, F. Jegerlehner, G. Penso, and C. Verzegnassi, in CERN Yellow Report on "Polarization at LEP" 1988.
30. S. Chekanov *et al.* [ZEUS Collaboration], Nucl. Phys. **B831**, 1 (2010), [hep-ex/09114119].
31. N. N. Nikolaev and B. Zakharov, Z.Phys. **C49**, 607 (1991).
32. K. Golec-Biernat and M. Wüsthoff, Phys. Rev. D **59**, 014017 (1999), [hep-ph/9807513].
33. E. Iancu, K. Itakura, and S. Munier, Phys. Lett. **B590**, 199 (2004), [hep-ph/0310338].
34. J. Bartels, K. Golec-Biernat, and H. Kowalski, Phys. Rev. D **66**, 014001 (2002), [hep-ph/0203258].
35. I. Balitsky, Nucl. Phys. B **463**, 99 (1996), [hep-ph/9509348].
36. S. Catani, M. Ciafaloni, and F. Hautmann, Nucl. Phys. B **366**, 135 (1991).
37. M. Ciafaloni, Nucl. Phys. B **296**, 49 (1988).
38. S. Catani, F. Fiorani, and G. Marchesini, Phys. Lett. B **234**, 339 (1990).
39. S. Catani, F. Fiorani, and G. Marchesini, Nucl. Phys. B **336**, 18 (1990).
40. G. Marchesini, Nucl. Phys. B **445**, 49 (1995).
41. H. Jung and F. Hautmann (2012), [arXiv:1206.1796].
42. H. Jung, S. Baranov, M. Deak, A. Grebenyuk, F. Hautmann, *et al.*, Eur.Phys.J. **C70**, 1237 (2010), [arXiv:1008.0152].
43. M. Deak, F. Hautmann, H. Jung, and K. Kutak, *Forward-Central Jet Correlations at the Large Hadron Collider* (2010), [arXiv:1012.6037].
44. A. Falkowski, M. L. Mangano, A. Martin, G. Perez, and J. Winter (2012), [arXiv:1212.4003].
45. Y. Li and F. Petriello, Phys.Rev. **D86**, 094034 (2012), [arXiv:1208.5967].
46. G. Bozzi, J. Rojo, and A. Vicini, Phys.Rev. **D83**, 113008 (2011), [arXiv:1104.2056].
47. S. D. Drell and T.-M. Yan, Phys. Rev. Lett. **25**, 316 (1970).
48. M. Yamada and M. Hayashi, Nuovo Cim. **A70**, 273 (1982).
49. T. Carli *et al.*, Eur. Phys. J. **C66**, 503 (2010), [arXiv:0911.2985].
50. J. M. Campbell and R. K. Ellis, Phys. Rev. **D60**, 113006 (1999), [arXiv:9905386].
51. J. M. Campbell and R. K. Ellis, Nucl. Phys. Proc. Suppl. **205-206**, 10 (2010), [arXiv:1007.3492].
52. Z. Nagy, Phys.Rev.Lett. **88**, 122003 (2002), [hep-ph/0110315].
53. M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, *et al.*, Comput.Phys.Commun. **182**, 1034 (2011), [arXiv:1007.1327].
54. P. Bärnreuther, M. Czakon, and A. Mitov (2012), [arXiv:1204.5201].
55. S. Moch, P. Uwer, and A. Vogt, Phys.Lett. **B714**, 48 (2012), [hep-ph/1203.6282].
56. T. Kluge, K. Rabbertz, and M. Wobisch, pp. 483–486 (2006), [hep-ph/0609285].
57. M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober [fastNLO Collaboration] (2011), [arXiv:1109.1310].
58. D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch [fastNLO Collaboration] (2012), [arXiv:1208.3641].
59. Z. Nagy and Z. Trocsanyi, Phys.Rev. **D59**, 014020 (1999), [hep-ph/9806317].
60. Z. Nagy and Z. Trocsanyi, Phys.Rev.Lett. **87**, 082001 (2001), [hep-ph/0104315].
61. Z. Nagy, Phys.Rev. **D68**, 094002 (2003), [hep-ph/0307268].
62. N. Kidonakis and J. Owens, Phys.Rev. **D63**, 054019 (2001), [hep-ph/0007268].
63. R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, *et al.*, Nucl.Phys. **B855**, 608 (2012), [arXiv:1108.1758].
64. R. D. Ball *et al.* [NNPDF Collaboration], Nucl.Phys. **B849**, 112 (2011), [arXiv:1012.0836].
65. G. Watt and R. Thorne, JHEP **1208**, 052 (2012), [arXiv:1205.4024].
66. J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, *et al.*, Phys.Rev. **D65**, 014013 (2001), [hep-ph/0101032].
67. M. Botje, J.Phys. **G28**, 779 (2002), [hep-ph/0110123].
68. W. T. Giele and S. Keller, Phys.Rev. **D58**, 094023 (1998), [hep-ph/9803393].
69. W. T. Giele, S. Keller, and D. Kosower (2001), [hep-ph/0104052].
70. F. Aaron *et al.* [H1 and ZEUS Collaborations], JHEP **1001**, 109 (2010), [arXiv:0911.0884].
71. *ATLAS NNLO epWZ12*, availble via: https://lhapdf.hepforge.org/pdfsets.
72. J. L. Abelleira Fernandez *et al.* [LHeC Study Group], Journal of Phys. **G**, 075001 (2012), [arXiv:1206.2913].