The Structure of the Proton in the LHC Precision Era

Jun Gao[a], Lucian Harland-Lang[b], Juan Rojo[c]

[a]*Instiutue of Nuclear and Particle Physics and Shanghai Key Laboratory for Particle Physics and Cosmology, Shangai Jiao Tong University, Shangai, China*
[b]*Department of Physics and Astronomy, University College London, WC1E 6BT, United Kingdom*
[c]*Department of Physics and Astronomy, VU University, De Boelelaan 1081, 1081HV Amsterdam, and Nikhef, Science Park 105, NL-1098 XG Amsterdam, The Netherlands*

Abstract

We review recent progress in the determination of the parton distribution functions (PDF) of the proton, with emphasis on application for precision phenomenology at the Large Hadron Collider (LHC). We start by discussing the general theoretical framework underlying the global QCD analysis of the internal proton structure in terms of quarks and gluons. We then present a detailed overview of the hard-scattering measurements, and the corresponding theory predictions, that are used in state-of-the-art PDF fits, emphasising the crucial role that NNLO calculations play. We introduce the methodology used to extract PDFs from the data in the global analysis, and then review and compare the most recent releases from the various PDF fitting collaborations. We discuss the role that QED corrections and the photon PDF play in modern PDF analysis. We provide representative examples of the implications of PDF fits for high-precision LHC phenomenological applications. We conclude this report by discussing some selected topics relevant for the future of PDF determinations, including the treatment of theoretical uncertainties, the connection with lattice QCD calculations, and the role of PDFs at future high-energy colliders beyond the LHC.

Keywords: Parton Distributions, Quantum Chromodynamics, Large Hadron Collider, Higgs boson, Standard Model, Electroweak theory

Contents

Email addresses: `jung49@sjtu.edu.cn` (Jun Gao), `l.harland-lang@ucl.ac.uk` (Lucian Harland-Lang), `j.rojo@vu.nl` (Juan Rojo)

1. Introduction

The determination of the quark and gluon structure of the proton is a central component of the precision phenomenology program at the Large Hadron Collider (LHC). This internal structure of nucleons is quantified in the collinear QCD factorization framework by the Parton Distribution Functions (PDFs), which encode the probability of finding quarks and gluons inside the proton carrying a given amount of its momentum. Being driven by low-scale non-perturbative dynamics, PDFs cannot currently be computed from first principles, at least with current technology, and therefore they need to be determined from experimental data from a variety of hard-scattering cross-sections in lepton-proton and proton-proton collisions. This program, known as the *global QCD analysis*, involves combining the most PDF-sensitive data and the highest precision QCD and electroweak calculations available within a statistically robust fitting methodology. See Refs. [1, 2, 3, 4, 5, 6, 7, 8, 9] for recent reviews on PDF determinations.

A strong motivation to improve our understanding of the internal structure of the proton is provided by the fact that parton distributions and their associated uncertainties play a decisive role in several LHC applications. To begin with, they represent one of the dominant theoretical uncertainties for the determination of the Higgs boson couplings [10], where any deviation from the tightly fixed SM predictions would indicate a smoking gun for new physics. PDF uncertainties also affect the production of new high-mass resonances, as those predicted by many Beyond the Standard Model scenarios [11], since they probe PDFs at large values of the momentum fraction x which are poorly constrained by available data. A third example is provided by the measurement of precision SM parameters at hadron colliders, such as the W mass [12] or the strong coupling constant $\alpha_s(Q)$. These can be sensitive to BSM effects (for instance via virtual effects of new particles too heavy to be produced directly) and in many cases PDF uncertainties are also one of the limiting factors of the measurements. Beyond the LHC, there are also several other instances where PDFs play an important role, for instance in astroparticle physics, such as for the accurate predictions for signal [13] and background [14] events at neutrino telescopes. And needless to say, parton distributions will keep playing an important role for any future higher-energy collider involving hadrons in the initial state [15, 16], and therefore improving PDFs helps in shaping fit the physics potential of such future colliders.

A number of collaborations provide regular updated of their PDF sets, see [17, 18, 19, 20, 21, 22, 23] and references there in. Differences between these various analysis arise at the level of the choice of the input fitted dataset, the theoretical calculations of cross-sections, and methodological choices for the parametrization of PDFs, the estimate and propagation of PDF uncertainties, and the treatment of external parameters. For instance, while some PDF fits are based on a global dataset, including the widest possible variety of experimental constraints, some others are based on reduced datasets (for example, without jet data) or even on a single dataset, as the HERAPDF2.0 set which is based only on the HERA inclusive structure functions. Despite these differences, it has been shown that, under some well-specified conditions, PDF sets can be statistically combined among them into a unified set. The most popular realization of this combination paradigm are the PDF4LHC15 sets [2], which combine the CT14, MMHT14, and NNPDF3.0 sets using the Monte Carlo method, and are subsequently reduced to small number of Hessian eigenvectors or MC replicas to facilitate phenomenological applications.

This Report is motivated by the fact that the recent years have seen a number of rather important breakthroughs in our understanding of the quark and gluon structure of the proton. To begin with, the impressive recent progress in NNLO QCD calculations has now made possible to include essentially all relevant collider cross-sections consistently into a NNLO global analysis, from top-quark differential distributions to inclusive jets and dijets, isolated photons, and the p_T distribution of Z bosons, among others. These theoretical developments have been matched by the availability of high-precision measurements from ATLAS, CMS, and LHCb at $\sqrt{s}$ = 7, 8 and 13 TeV, in several cases with statistical uncertainties at the per-mile level and systematic errors at the few-percent level. The combination of these state-of-the art calculations and high-precision data provides a great opportunity to constrain PDFs, but it also represents a challenge to verify if the global QCD framework can satisfactorily accommodate them.

Another important topic that has attracted a lot of attention recently is the role that QED and electroweak effects, and specifically the photon PDFs, play in global fits of parton distributions. Recent progress has demonstrated that the photon PDF can be computed with few-percent accuracy [24], improving on previous model and data-driven determinations, with direct implications for LHC cross-sections. Another important development is the realization that the charm PDF can be treated on an equal footing as the light quarks in the global fit [25], allowing to stabilize the m_c dependence, improve the agreement with high-precision data, and making possible direct comparison with non-perturbative models of the charm content of the proton [26]. From the methodological point of view, there have been several improvements in the way that PDFs are parametrized and the various associated sources of uncertainty estimated among the PDF fitting groups. In addition, there has also been a recent explosion in the number of tools available for PDF studies from the open-source fitting framework `xFitter` [27], to new fast (N)NLO interfaces and public codes for the PDF evolution and the efficient calculation of hadronic cross-sections. It is therefore the goal of this Report to present a detailed overview of these various recent developments, and how they have modified our present understanding of the quark and gluon structure of the proton, with emphasis on the resulting phenomenological applications.

This Report focuses only on one of the main aspects of the internal structure of nucleons, namely collinear unpolarized PDFs, which are its most relevant feature for the exploration of the high-energy frontier at the LHC. There are however many other fascinating aspects of the inner life of protons that due to space limitations cannot be covered here, since each of these important topics would deserve a separated Report. These include, among others, the determination of its spin structure by means of the polarized

PDFs [28, 29]; the nuclear modifications of the free-proton PDFs [30, 31], relevant for the understanding of cold nuclear matter effects at the RHIC and LHC heavy-ion program; or the three-dimensional imaging of nucleons in terms of transverse-momentum-dependent PDFs (TMD-PDFs) [32]. We note only here that progress in some of these other aspects of the proton structure also affect unpolarized PDF fits, for example nuclear corrections are frequently used to include neutrino DIS structure functions taken on heavy nuclear targets.

The structure of this Report is as follows. First of all in Sect. 2 we review the theoretical foundations of the global PDF analysis framework, specifically the QCD factorization theorems of lepton-hadron and hadron-hadron collisions and the scale dependence of the PDFs. Then in Sect. 3 we discuss the hard-scattering experimental data, as well as the corresponding state-of-the-art theoretical calculations, that are used to constrain the PDFs in modern global analyses. We continue in Sect. 4 presenting the methodological framework of PDF fits, including the various approaches to parametrized the PDFs and to estimate and propagate the uncertainties from theory and data to physical cross-sections. In Sect. 5 we summarize the main features of the different PDF collaborations that provide regular updates of their PDF fits, and then in Sect. 6 we compare them, assessing their differences and similarities for different aspects of the proton structure such as the gluon PDF, quark-flavour separation, and the strange and charm content of the proton. We then move in Sect. 7 to discuss a topic that has received a lot of attention recently, namely the role that QED and electroweak corrections play in PDF fits, with emphasis on the photon content of the proton. In Sect. 8 we highlight a number of representative examples of the role of PDFs and their uncertainties for the LHC precision physics program. In the last part of this Report, Sect 9 we discuss some of the topics that are likely to play an important role for the future of PDF determinations, such as the quantification of theoretical uncertainties, the interplay with lattice QCD calculations, and the application of PDFs for future higher energy lepton-proton and proton-proton colliders. Finally we conclude and summarize this Report in Sect. 10.

2. The global QCD analysis framework

In the first section of this Report, we first present a brief historical account of PDF determinations and then introduce the foundations of global PDF analysis, namely the QCD factorization of lepton-hadron and hadron-hadron collisions. We also discuss the scale dependence of parton distributions as encoded in the DGLAP evolution equations and briefly address the important topic of mass effects in deep-inelastic structure functions.

2.1. A brief history of PDF fits

The first direct measurement of proton structure was performed in the the pioneering experiments of Hofstadter on elastic electron–nucleon scattering [33, 34]. By examining the deviations from the simple Mott scattering formulae for point–like particles, the finite extent of the proton could be resolved, and the charge radius of the proton was determined to be ~ 0.7 fm to within a few percent precision.

Although this result hints at an underlying substructure, the serious possibility that the proton is composite originated with the idea proposed independently by Zweig [35] and Gell–Mann [36] in 1964. By postulating the existence of three 'aces' (Zweig's term) or 'quarks' (Gell–Mann's) with fractional electric charge and baryon number, and spin–$1/2$, the complex structure of the hadrons and meson multiplets could be simply explained. However, Zweig and Gell–Mann were understandably cautious about interpreting these objects as physical particles of finite mass, rather than simply convenient mathematical devices, as the mechanism for binding such quarks together was not understood and stable quarks had not been seen experimentally.

This situation changed in 1967 with the new experimental data on deep inelastic scattering (DIS) provided by the SLAC 20 GeV linear accelerator. The SLAC–MIT collaboration were surprised to find that, in contrast to the case of elastic lepton–proton scattering, the two form factors associated with the DIS cross section, the so–called structure functions, were roughly independent of Q^2 [37, 38]. Moreover, these appeared to exhibit the scaling behaviour predicted by Bjorken in 1969 [39], namely that the structure functions should depend only on the ratio of Q^2 to the lepton energy loss ν in the proton rest frame[1].

These observations led Feynman to introduce the parton model [41], in which the incident lepton scatters incoherently and instantaneously from the point–like 'parton' constituents of the proton. This concept, developed further in [42], naturally explains the observed Bjorken scaling behaviour, with the point–like partons in this simple picture providing no additional scale through which Bjorken scaling could be broken. At the same time Callan and Gross [43] showed that the DIS structure functions obey a simple relation for the case of spin–$\frac{1}{2}$ quark constituents, a finding that was also supported by the data [44]. These partons were therefore naturally associated with the quarks of Gell–Mann and Zweig. The demonstration of asymptotic freedom in 1973 in strongly interacting non–abelian gauge theories [45, 46] provided a simple explanation for the observed absence of free quarks, through the process of confinement, and the QCD parton model became the established approach.

A natural ingredient of this parton model is the probability distributions of the partons themselves, that is the PDFs. The first studies concentrated on developing simple models for these objects based on the limited experimental input available, for example: in [42] phase space considerations were used to conclude that the PDFs must also include a contribution from the now well known sea of quark–antiquark pairs in addition to the valence quarks; in [47] a gluon PDF was introduced to account for the observed quark momentum fractions in a physically reasonable way, and simple x dependencies of the PDFs were predicted according to general Regge theory and phase space expectations.

The idea of fitting a freely parameterised set of PDF followed soon after these first studies. In [48] the approach of [47] was extended to a more general phenomenological form, and a 4–parameter fit to the quark PDFs was performed to the available data on proton and neutron structure functions. As the amount and type of data increased the forms of the PDFs became increasingly general, see for example [49, 50]. Although the momentum fraction carried by the gluon could be determined by the missing contribution to DIS appearing in the momentum sum rule, it was only possible to fit its shape following the observation of scaling violations in the structure functions, first seen at FNAL [51] in 1974. Such Q^2 dependent deviations from simple Bjorken scaling occur due to higher–order QCD corrections to DIS and were directly connected through the DGLAP equation [52, 53, 54, 55] in 1977 to the Q^2 evolution of the PDFs. This allowed the first determinations of the shape of the gluon to be made in [56] (see also [57]).

The subsequent LO fits of [58] (based on [59]) to fixed target structure function and neutrino DIS data, and [60], which also included J/ψ meson and muon pair hadroproduction, were widely used for a range of phenomenological applications. By the late 1980s PDF fits at NLO in the strong coupling were standard, with the earlier analyses of [61, 62] fitting to fixed target DIS and the subsequent fits of [63, 64, 65] including prompt photon and Drell–Yan hadroproduction. The 'dynamical' PDF set of [66] were produced with the assumption that at low scale the quark sea vanished and the gluon becomes proportional to the valence quark distributions, themselves determined from DIS data.

Up to this point all DIS data was taken with fixed target experiments and hence limited to the higher $x \gtrsim 0.01$ region. This changed in 1992 when HERA high energy collider at DESY started taking data. This

[1] Indications of this scaling were also observed at the DESY electron synchrotron in the same year [40].

collided 920 GeV protons with $\approx$ 27.5 GeV electrons for most of the run period, allowing the previously unexplored region down to $x \sim 10^{-4}$ region to be probed at high Q^2. By 1994 data from HERA were included for the first in the MRS(A) [67] and CTEQ3 [68] global fits. These were also the first fits to include data from the Tevatron $p\overline{p}$ collider, with in particular the W asymmetry data providing new information on the quark flavour decomposition. In the years that followed further public releases within these approaches were produced, with in particular the increasingly precise HERA measurements, and Tevatron data on jet production placing new and important constraints.

These PDF sets corresponded to the best fit only, that is, no precise estimate of the uncertainty on the PDFs due to the errors on the data in the fit were included, beyond simple studies where a range of fits under different input assumptions might be performed to give some estimate of the spread. This was an acceptable situation when the uncertainties on the hadron collider data were sufficiently large, however as the data precision increased this rapidly became an issue. In the 1996 CDF measurement [69] of inclusive jet production, for example, there was an apparent excess of events at high jet $E_\perp$ that was interpreted at the time as a possible sign for new physics. In the subsequent study of [70] it was shown that the gluon PDF could be modified in a way that still fit all available data, including the CDF jets. Clearly a precise evaluation of the PDF uncertainties was needed to avoid such a situation.

The first attempts to produce such uncertainties, based on linear propagation of the experimental systematic and statistical errors through to the PDFs, considered a restricted set of DIS data [71, 72, 73, 74]. The extension of these methods to the wider data set included in a global PDF fit was a complicated problem, both from a purely technical point of view, but also more conceptually. In particular, more conventional statistical approaches to evaluating the uncertainty on the fitted PDF parameters, such as a standard '$\Delta\chi^2 = 1$' variation, are only appropriate when fitting perfectly consistent data sets with purely Gaussian errors against a well–defined theory. For PDF fits none of these criteria are fulfilled: different data sets are often found to be highly improbable, with a χ^2 per degree of freedom well above one, the experimental systematic uncertainties will not generally be Gaussian in nature, and the fixed order perturbative theory calculation will carry its own (usually omitted) uncertainties. These issues were addressed in the CTEQ [75, 76] and MRST [77] PDF releases in 2002, with the basic idea being to allow the χ^2 to vary from the minimum by a larger degree, or 'tolerance', to account for the departure from the textbook statistical situation.

The calculation of the NNLO splitting functions in 2004 [78, 79] provided the necessary tools to go to NNLO in PDF fits, and with the release of the MSTW08 [80] and CT10 [81] sets (the successors to the MRST and CTEQ sets, respectively) NNLO became the standard for global PDFs. At the same time the ABKM09 [82] NNLO PDFs were released. These were based on the earlier studies of [71, 83, 84], and fit to a reduced data set of DIS and fixed target Drell–Yan and dimuon production, with a classical '$\Delta\chi^2 = 1$' error treatment applied. A further set to consider a reduced data sample to appear at this time was the HERAPDF1.0 [85] PDFs. These included only the combined H1 and ZEUS measurements from the HERA Run I phase, with the aim of determining the PDFs from a completely consistent DIS data sample. This allowed the PDF uncertainty to again be described without the introduction of a larger tolerance factor, while the uncertainties due to model assumptions and choice of parameterisation were included in addition. This NLO set was extended to NNLO in the HERAPDF2.0 [21] PDFs, which used the final combined HERA I + II data sample. The NNLO JR09 [86] set included a range of DIS and fixed target data, applying both a 'standard' fitting approach and the 'dynamical' approach of [66]. The subsequent JR14 [22] set included a range of data updates, including jet production from the Tevatron.

The approaches described above differ greatly in many respects, both in the choice of input data sets, and the treatment of the corresponding theory predictions. However, while there are significant differences in the precise choice of parameterisation, in all cases these rely on parameterising the PDFs in terms of reasonably contained, $O(20-40)$, number of free variables. Moreover, while the precise prescription may vary, these are again all based on the 'Hessian' linear error propagation procedure. A different approach, first discussed in [87], was taken by the NNPDF collaboration. Here the PDF functional forms are based on neural networks, allowing many more ($O(200-300)$) free parameters. In addition, rather than constructing the PDF error from the χ^2 variation about the best fit values, a 'Monte Carlo' (MC) approach is taken, with a large enough sample of PDF 'replica' sets each fit to randomly distributed pseudo–data generated according to the measured data values and their uncertainties. The first NNPDF1.0 fit was reported in [88], at NLO and to a range of DIS and fixed target data. Subsequently, NNPDF2.1 [89] provided the first NNLO PDF set within this approach, and included Tevatron data for the first time.

In recent times, data from the LHC has played an increasingly important role in PDF determination. The CT14 [18], MMHT14 [19] and NNPDF2.3 [90] sets included LHC data on jets, W and Z boson production, and top pair production for the first time. In addition, ABM12 [91] was the first set from this group to include input from LHC, with data on W and Z boson production and top quark pair production. As we will see in this review, these data, which are being produced with increasingly high precision, are now providing some of the most stringent constraints on the PDFs.

In parallel to these developments, there has been increasing focus on the use of PDFs as precise tools for LHC physics, emphasising the need for clear benchmarking exercises between sets and PDF combinations, to provide an overall PDF uncertainty. The PDF4LHC Working Group, formed in 2006, has played a significant role in this, with the benchmarking described in [5] leading to first so–called PDF4LHC recommendation [92] for the use of PDFs and their uncertainties at the LHC. This has subsequently been updated in [2] (see [1] for an alternative approach).

2.2. QCD factorization in deep-inelastic scattering

The importance of DIS for PDF fits cannot be overemphasised. This process was instrumental in the discovery of quarks, and has since then represented the backbone of global PDF fits. The DIS mechanism is schematically represented in Fig. 1. Here, an energetic lepton, which can be either charged (electron or muon) or neutral (a neutrino) scatters off a proton (or some other hadron) by means of the interchange of a virtual photon γ^* or a $W^\pm$ or Z boson. The large virtuality Q of the gauge boson, $Q \gg \Lambda_{\rm QCD}$, ensures that the process can be described within QCD factorization in terms of coefficient functions and parton distributions, as we show below.

The DIS process is defined in terms of a few invariant quantities, namely

$$x \equiv \frac{Q^2}{2P\cdot q}\,, \quad Q^2 \equiv q^2\,, \quad y \equiv \frac{q\cdot P}{k\cdot P}\,, \tag{1}$$

where k and k' are the four–momenta of the incoming and outgoing leptons, q is the four–momentum of the exchanged gauge boson and P is incoming proton's momentum. Here x is known as the Bjorken variable, and although it is defined purely in terms of the kinematics of the initial and final-state particles, it can be shown that in the parton model it corresponds to the momentum fraction carried by the struck parton. Recall that by momentum conservation $q = k' - k$, and thus all the variables in Eq. (1) can be determined by the knowledge of the incoming momenta of the lepton k and of the proton P as well as the outgoing momentum of the lepton k' without any reference to the final hadronic state X. The centre of mass energy W of the quark-photon collision is given by

$$W^2 = (P+q)^2 = Q^2\frac{1-x}{x} + m_p^2\,. \tag{2}$$

The value $x = 1$ corresponds to the elastic limit, where the proton remains intact after the collision.

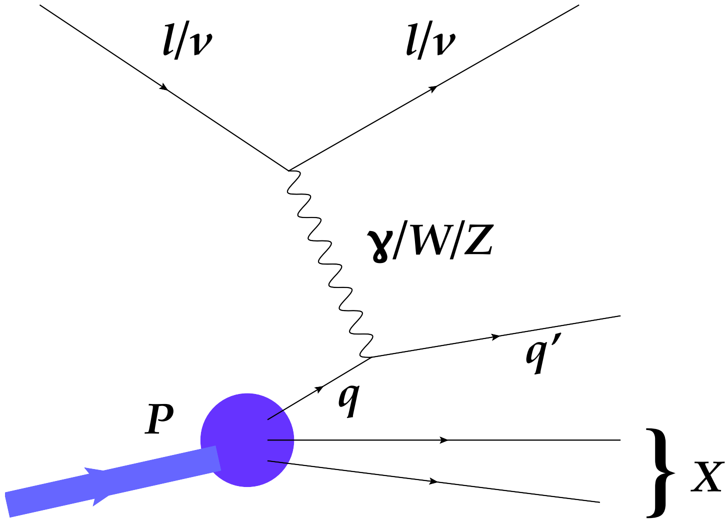


Figure 1: Schematic representation of the deep-inelastic scattering process. An energetic lepton (electron, muon or neutrino) scatters off one of the quarks in the proton by means of the interchange of a gauge boson (γ, $W^{\pm}$ or Z). The large virtuality of the gauge boson, $Q \gg \Lambda_{\rm QCD}$, ensures that the process can be described within QCD factorization in terms of coefficient functions and parton distributions.

Differential cross sections in DIS are therefore measured in terms of two of the three kinematic variables in Eq. (1), for instance as a function of (x, Q^2) or (x, y). Using Lorentz invariance and kinematical arguments, it can be shown that the DIS cross sections can be expressed in terms of a series of independent structure functions that describe the dynamics of the interaction between the gauge boson and the hadron. In the neutral current (NC) case, that is, where either a virtual photon γ^* or a Z boson is exchanged, the DIS differential cross section for a charged lepton $\ell^{\pm}$ scattering off a proton can be decomposed in terms of structure functions as follows:

$$\frac{d^2\sigma^{{\rm NC},\ell^{\pm}}}{dxdQ^2}(x,y,Q^2) = \frac{2\pi\alpha^2}{xQ^4}\left[Y_+F_2^{NC}(x,Q^2) \mp Y_-xF_3^{NC}(x,Q^2) - y^2F_L^{NC}(x,Q^2)\right] , \tag{3}$$

where we have defined

$$Y_{\pm} = 1 \pm (1-y)^2 . \tag{4}$$

In most cases, experimental measurements are given in terms of a reduced cross-section, defined as

$$\widetilde{\sigma}^{{\rm NC},e^{\pm}}(x,y,Q^2) = \left[\frac{2\pi\alpha^2}{xQ^4}Y_+\right]^{-1}\frac{d^2\sigma^{{\rm NC},e^{\pm}}}{dxdQ^2}(x,y,Q^2) , \tag{5}$$

which is more closely related to the dominant structure function $F_2(x, Q^2)$, and thus the underlying PDFs. In the case of charged current (CC) DIS, when neutrinos are used as projectiles or when the incoming charged leptons interact with the proton by means of the exchange of a charged weak gauge boson $W^{\pm}$, the differential cross-sections are given by:

$$\begin{aligned} \frac{d^2\sigma^{{\rm CC},e^{\pm}}}{dxdQ^2}(x,y,Q^2) &= \frac{G_F^2}{4\pi x}\left(\frac{M_W^2}{M_W^2+Q^2}\right)^2 \\ &\times \frac{1}{2}\left[Y_+F_2^{CC,e^{\pm}}(x,Q^2) \mp Y_-xF_3^{CC,e^{\pm}}(x,Q^2) - y^2F_L^{CC,e^{\pm}}(x,Q^2)\right] . \end{aligned} \tag{6}$$

which is generally rescaled to define a reduced cross section

$$\widetilde{\sigma}^{\mathrm{CC,e}^{\pm}}(x,y,Q^2) = \left[\frac{G_F^2}{4\pi x}\left(\frac{M_W^2}{M_W^2+Q^2}\right)^2\right]^{-1} \frac{d^2\sigma^{\mathrm{CC,e}^{\pm}}}{dxdQ^2}(x,y,Q^2)\,, \tag{7}$$

similarly to the NC case.

By exploiting the QCD factorization theorem, it can be shown that the general expression for the DIS structure functions can be written schematically as

$$F(x,Q^2) = x\int_x^1 \frac{dy}{y}\sum_i C_i\left(\frac{x}{y},\alpha_s(Q^2)\right) f_i(y,Q^2), \tag{8}$$

where $C_i\left(x/y,\alpha_s(Q^2)\right)$ are known as the coefficient functions and $f_i(y,Q^2)$ are the PDFs. The coefficient functions represent the cross section for the partonic process $q_i+\gamma^*\to X$, and can be computed in perturbation theory as a series expansion in the strong coupling α_s. While these encode the short distance dynamics of the parton–boson collision, the PDFs are determined by long distance non–perturbative QCD dynamics, and can therefore not be computed using perturbative methods. Therefore, they need to the parametrised and extracted from a global analysis of hard scattering measurements. This is possible due to the crucial factorization property of Eq. (8); while the coefficient functions (or in general the partonic cross-sections) are process dependent, the PDFs instead are universal. The PDFs extracted in such a global fit can therefore be used to make predictions for other PDF–dependent processes.

2.3. *QCD factorization in hadronic collisions*

In a similar way to the DIS structure functions for electron–proton collisions, the production cross sections in proton–proton collisions can be factorized as convolutions of two universal PDFs and a process-dependent partonic cross section. For example, the Drell-Yan production cross section can be expressed as [93, 94]

$$\frac{d^2\sigma^{\mathrm{DY}}}{dydQ^2}(y,Q^2,\mu_R^2,\mu_F^2) = \sum_{a,b=q,\bar{q},g}\int_{\tau_1}^1 dx_1 f_a(x_1,\mu_F^2)\int_{\tau_2}^1 dx_2 f_b(x_2,\mu_F^2)\,\frac{d^2\hat{\sigma}_{ab}^{\mathrm{DY}}}{dydQ^2}(x_1,x_2,y,Q^2,\mu_R^2,\mu_F^2), \tag{9}$$

where y and Q^2 are the rapidity and invariant mass square of the lepton pair, and s is the centre–of–mass energy of the two incoming protons, while μ_F (μ_R) are the factorization (renormalization) scales. The lower limits on integration are $\tau_{1,2} = \sqrt{Q^2/s}\,e^{\pm y}$. The partonic cross section can be computed as a perturbative expansion in α_S:

$$\frac{d^2\hat{\sigma}_{ab}^{\mathrm{DY}}}{dydQ^2}(x_1,x_2,y,Q^2,\mu_R^2,\mu_F^2) = \sum_{n=0}^{\infty}\left(\frac{\alpha_s(\mu_R^2)}{2\pi}\right)^n \frac{d^2\hat{\sigma}_{ab}^{(n)\,\mathrm{DY}}}{dydQ^2}. \tag{10}$$

The introduction of PDFs requires the introduction of a factorization scale μ_F, below which additional collinear emissions are absorbed into the PDFs. To all orders, the physical cross section, as a product of the PDFs and partonic cross section is independent of the choice of the factorization scale. However at any truncated order in the perturbative series there is a some higher order dependence, which can be minimised by choosing a suitable value of μ_F so as to maintain a better convergence of the series. In Drell-Yan production the conventional scale choice is $\mu_F^2 = Q^2$.

In case of total inclusive cross section for a narrow resonance production with mass M, the cross section can be factorized as

$$\sigma = \sum_{a,b=q,\bar{q},g} \int_{M^2}^{s} \frac{d\hat{s}}{\hat{s}} \mathcal{L}_{ab}(\hat{s}, \mu_F^2)\, \hat{s}\hat{\sigma}_{ab}(\hat{s}, M^2, \mu_R^2, \mu_F^2), \tag{11}$$

where $\hat{s}$ is the center of mass energy of the two incoming partons, the parton-parton luminosity can be defined as [95]

$$\mathcal{L}_{ab}(\tau, \mu_F^2) = \frac{1}{s} \int_{\tau/s}^{1} \frac{dx}{x} f_a(\tau/sx, \mu_F^2)\, f_b(x, \mu_F^2). \tag{12}$$

The partonic cross section depends only on the kinematic variable $z \equiv M^2/\hat{s}$ and $\mu_{F,R}$

$$\hat{s}\hat{\sigma}_{ab}(\hat{s}, M^2, \mu_R^2, \mu_F^2) = \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu_R^2)}{2\pi} \right)^n C_{ab}^{(n)}(z, \mu_R^2, \mu_F^2). \tag{13}$$

The coefficient functions $C^{(n)}(z, \mu_R^2, \mu_F^2)$ are known to NNLO for Drell-Yan production [96] and to N^{3}LO for Higgs boson production via gluon fusion in the limit of infinite top quark mass [97].

2.4. The DGLAP evolution equations

As discussed above, the PDFs depend on two variables: the Bjorken variable x, which at leading order can be identified with the momentum fraction carried by the considered parton, and the scale Q^2, which in DIS corresponds to the virtuality of the exchanged gauge boson. While the dependence of the PDFs on x is determined by non–perturbative dynamics, and therefore cannot be computed perturbatively, the situation is different for the Q^2 variable. Here, the Q^2 dependence of the PDFs is introduced when higher–order initial–state collider singularities of the partonic cross section are regularised. Such singularities correspond to generic long distance QCD dynamics, and therefore have a universal expression.

For this reason, the Q^2 dependence of the PDFs can in principle be computed in perturbation theory up to any given order. This is determined by a series of integro–differential equations known as the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations, which have the generic form

$$Q^2 \frac{\partial}{\partial Q^2} f_i(x, Q^2) = \sum_j P_{ij}(x, \alpha_s(Q^2)) \otimes f_j(x, Q^2), \tag{14}$$

where $P_{ij}(x, \alpha_s(Q^2))$ are the Altarelli-Parisi splitting functions, which can be computed in perturbation theory

$$P_{ij}(x, \alpha_s(Q^2)) = \sum_{n=0} \left(\frac{\alpha_s(Q^2)}{2\pi} \right)^{n+1} P_{ij}^{(n)}(x)\,, \tag{15}$$

and where $\otimes$ denotes the convolution

$$f(x) \otimes g(x) \equiv \int_x^1 \frac{dy}{y} f(y) g\left(\frac{x}{y}\right), \tag{16}$$

which appears ubiquitously in QCD calculation. The splitting functions Eq. (15) depend on the type of initial and final state parton that is involved in the splitting. At leading order, the DGLAP splitting functions are given by

$$P_{qq} = \frac{4}{3} \left[\frac{1 + x^2}{(1 - x)_+} \right], \tag{17}$$

$$P_{qg} = \frac{1}{2}\left[x^2 + (1-x^2)\right] , \tag{18}$$

$$P_{qg} = \frac{4}{3}\left[\frac{1+(1-x)^2}{x}\right] , \tag{19}$$

$$P_{gg} = 6\left[\frac{1-x}{x} + x(x-1) + \frac{x}{(1-x)_+}\right] . \tag{20}$$

Note that both P_{gg} and P_{qg} have a singularity at $x = 0$: this fact is responsible for the rapid growth at lower x of the gluons and consequently of the sea quarks in this region.

The structure of the DGLAP evolution equations is significantly simplified if we use specific linear combinations of PDFs. For instance, below the charm threshold, where there are only $n_f = 3$ active quarks, the following combination

$$\begin{aligned}
\Sigma(x,Q^2) &\equiv \sum_{i=1}^{n_f} (q_i + \bar{q}_i)(x,Q^2) , \\
T_3(x,Q^2) &\equiv \left(u + \bar{u} - d - \bar{d}\right)(x,Q^2) , \\
T_8(x,Q^2) &\equiv \left(u + \bar{u} + d + \bar{d} - 2(s+\bar{s})\right)(x,Q^2) , \qquad (21) \\
V(x,Q^2) &\equiv \sum_{i=1}^{n_f} (q_i - \bar{q}_i)(x,Q^2) , \\
V_3(x,Q^2) &\equiv \left(u - \bar{u} - d + \bar{d}\right)(x,Q^2) , \\
V_8(x,Q^2) &\equiv \left(u - \bar{u} + d - \bar{d} - 2(s-\bar{s})\right)(x,Q^2) , \qquad (22) \\
& \qquad (23)
\end{aligned}$$

has the important property that all the PDF combinations except for Σ, known as the total quark singlet, evolve independently using their own specific splitting functions. These combinations, known as non-singlet flavour combinations, therefore obey a particularly simple evolution equation. As the $g \to q\bar{q}$ splitting can only generate an overall $q+\bar{q}$ combination, only the singlet PDF evolution is explicitly coupled to the gluon.

The splitting functions Eq. (15) are known up to $\mathcal{O}\left(\alpha_s^3\right)$ (NNLO), and thus PDF evolution can be performed up to this order. Several public codes implement the numerical solution of the DGLAP equations, with the `HOPPET`, `APFEL` and `QCDNUM` codes using x-space methods, while the `PEGASUS` code performs the evolution in Mellin space. These codes have undergone detailed benchmarking studies, with agreement at the level of $\mathcal{O}\left(10^{-5}\right)$ or better being found.

In order to illustrate the impact of the DGLAP evolution on the PDFs, in Fig. 2 we show the PDF4LHC NNLO Hessian set (with 100 eigenvectors) comparing the PDFs at a low scale of $Q^2 = 10$ GeV^2 (left) with the same PDFs evolved up to a typical LHC scale of $Q^2 = 10^4$ GeV^2 (right plot). In this plot the PDFs are shown together with the corresponding one-sigma PDF uncertainty band. From this comparison we see that the effects of the evolution are relatively mild in the non-singlet combinations $u_V = u - \bar{u}$ and $d_V = d - \bar{d}$, but they are dramatic on the gluon and the sea quarks, where they induce a very steep growth at small-x. This steep growth is driven by the small-x structure of the splitting functions Eq. (20) Another interesting aspect that can be observed from Fig. 2 is that the valence PDFs xu_V and xd_V are integrable, and the fact that they have a similar shape but with $u_V \simeq 2d_V$ is a consequence of the valence sum rules which fix the value of their integrals.

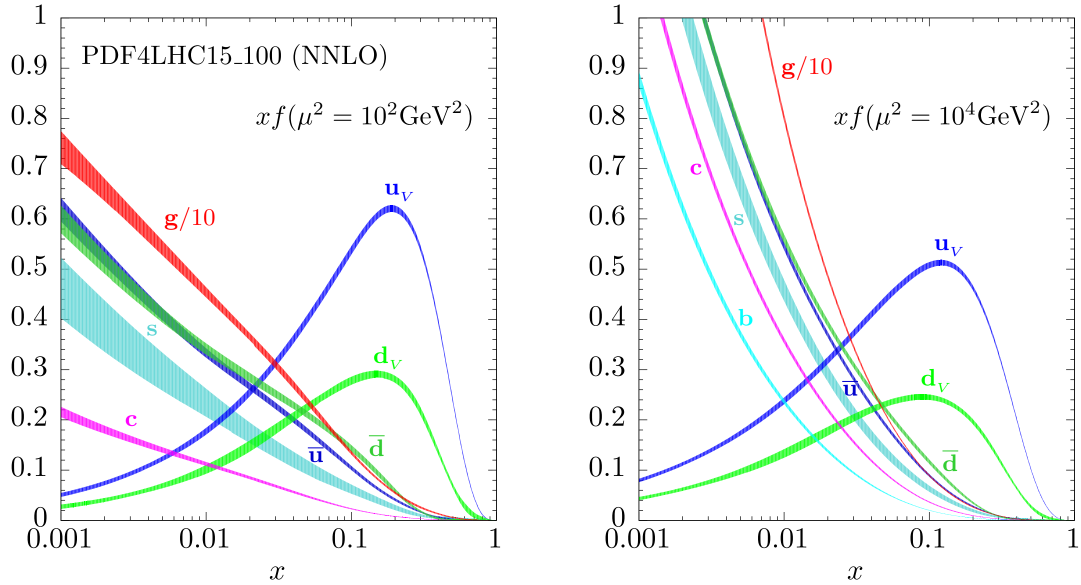


Figure 2: The effect of the DGLAP evolution in the PDF4LHC NNLO Hessian set (with 100 eigenvectors), comparing the PDFs at a low scale of $Q^2 = 10$ GeV^2 (left) with the same PDFs evolved up to a typical LHC scale of $Q^2 = 10^4$ GeV^2 (right plot). In this plot the PDFs are shown together with the corresponding one-sigma PDF uncertainty band.

2.5. Heavy quark structure functions

The contribution of the charm structure function F_2^c to the total inclusive structure function F_2^p at HERA can be as high as 25%, and so it is crucial to compute it with high accuracy. In such a case, the finite heavy quark mass must be taken into account. There are various theoretical schemes that have been proposed for the computation of heavy quark production in DIS:

- The Zero-Mass Variable Flavour Number scheme (ZM-VFNS), where all heavy quark mass effects are ignored but potentially large logarithms of $\ln Q/m$ are resummed into the heavy quark parton distribution. This is also known as the massless scheme.

- The fixed-flavor number scheme (FFNS), where the heavy quark is always treated as a massive particle and never as a massless parton irrespective of the value of the scale Q. In this scheme the heavy quark PDF does not exist and the number of active flavours is always kept fixed. This scheme takes into account heavy quark mass effects in the coefficient functions, but does not resum logarithmically enhanced terms of the form $\ln Q/m$ that become numerically relevant at high scales.

- The General-Mass Variable Flavor Number scheme (GM-VFNS) combines the advantage of the massive and massless calculations in an interpolated scheme which is valid for any value of the scale Q.

Here we review the basic steps that enter into the construction of the GM-VFNS calculation of heavy quark DIS structure functions. Although for illustration purposes we will focus on the FONLL derivation from Ref. [98], the construction of related GM-VFNS such as ACOT [**?**], S-ACOT [99] and TR [100] has most of these steps in common. We start by the expression of a generic DIS structure function $F(x, Q^2)$, in a kinematical regime where one has n_l light flavours and a single heavy flavour of mass $= m$. In the massless scheme, accurate when $W \gg 4m^2$, the expression of F in terms of PDFs and coefficient functions is the following

$$F^{(n_l+1)}(x,Q^2) = x\int_x^1 \frac{dy}{y} \sum_{i=q,\bar{q},h,\bar{h},g} C_i^{(n_l+1)}\left(\frac{x}{y}, \alpha_s^{(n_l+1)}(Q^2)\right) f_i^{(n_l+1)}(y,Q^2), \tag{24}$$

where q are the light quarks and h is the heavy quark. As indicated from the sum, in this scheme the heavy quark is treated as a massless parton, with all finite mass effects therefore neglected.

Now, in the massive (or decoupling) scheme, which is most suitable when $W \approx 4m^2$ and thus heavy quark mass effects must be accounted for, this structure function reads

$$F^{(n_l)}(x,Q^2) = x\int_x^1 \frac{dy}{y} \sum_{i=q,\bar{q},g} C_i^{(n_l)}\left(\frac{x}{y}, \frac{Q^2}{m^2}, \alpha_s^{(n_l)}(Q^2)\right) f_i^{(n_l)}(y,Q^2). \tag{25}$$

where now the massive coefficient functions $C_i^{(n_l)}$ includes the full mass dependence, and the heavy quark is on longer treated as a massless initial–state parton. In this scheme, the PDFs and α_s satisfy evolution equations with n_l active quarks. The construction of the GM-VFNS structure functions is based on two steps. First of all to express PDFs and α_s in the massless scheme by means of the matching conditions

$$\begin{aligned}
\alpha_s^{(n_l+1)}(Q^2) &= \alpha_s^{(n_l)}(Q^2) + \sum_{i=2}^{\infty} c_i(L) \times \left(\alpha_s^{(n_l)}(m^2)\right)^i , && (26)\\
f_i^{(n_l+1)}(x,Q^2) &= \int_x^1 \frac{dy}{y} \sum_{j=q,\bar{q},g} K_{ij}\left(\frac{x}{y}, L, \alpha_s^{(n_l)}(Q^2)\right) f_j^{(n_l)}(y,Q^2) , && (27)
\end{aligned}$$

where $L \equiv \log Q^2/m^2$, and then using these transformed expressions to write down $F^{(n_l)}$ in terms of PDFs and α_s in the massless scheme,

$$F^{(n_l)}(x,Q^2) = x\int_x^1 \frac{dy}{y} \sum_{i=q,\bar{q},g} B_i\left(\frac{x}{y}, \frac{Q^2}{m^2}, \alpha_s^{(n_l+1)}(Q^2)\right) f_i^{(n_l+1)}(y,Q^2), \tag{28}$$

Once we have expressed both $F^{(n_l)}$ and $F^{(n_l+1)}$ in terms of PDFs and α_s in the massless scheme, the second step is to match the two expressions while removing any double counting. This way we will maintain the main advantages of the two schemes (heavy quark mass effects in $F^{(n_l)}$, resummation of large $\ln Q^2/m^2$ logarithms in $F^{(n_l+1)}$) within a single scheme that is valid for any scale Q. To achieve this, one defines the massless limit of the massive scheme structure function as follows

$$F^{(n_l,0)}(x,Q^2) = x\int_x^1 \frac{dy}{y} \sum_{i=q,\bar{q},g} B_i^{(0)}\left(\frac{x}{y}, \frac{Q^2}{m^2}, \alpha_s^{(n_l+1)}(Q^2)\right) f_i^{(n_l+1)}(y,Q^2), \tag{29}$$

where in the coefficient functions $B_i^{(0)}$ all the terms which are power suppressed of the form m/Q are neglected, and the only dependence on the heavy quark mass m is on logarithms of the form $\ln Q/m$,

The FONLL approximation for F is then given by

$$\begin{aligned}
F^{\mathrm{FONLL}}(x,Q^2) &= F^{(d)}(x,Q^2) + F^{(n_l)}(x,Q^2), && (30)\\
F^{(d)}(x,Q^2) &\equiv \left[F^{(n_l+1)}(x,Q^2) - F^{(n_l,0)}(x,Q^2)\right] && (31)
\end{aligned}$$

where Eq. (31) is constructed out of the massless-scheme expression $F^{(n_l+1)}$, and the massless limit $F^{(n_l,0)}$ of the massive-scheme expression as in Eq. (29). It is thus clear to see that in the limit where $Q \gg m$, the

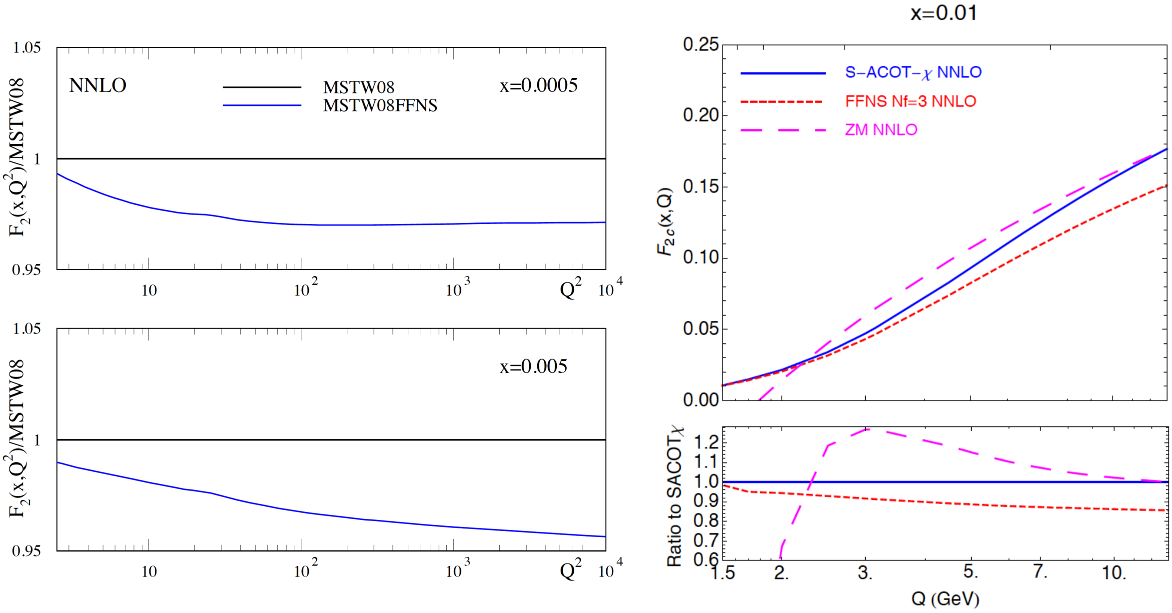


Figure 3: Left plot: the inclusive proton structure function $F_2(x, Q^2)$ at NNLO as a function of Q^2 for two different values of x in the TR' GM-VFNS as compared to the FFNS calculation. Right plot: the NNLO charm structure function $F_2^c(x, Q)$ as a function of Q for $x = 0.01$ comparing the S-ACOT-χ GM-VFNS with the corresponding ZM and FFN scheme calculations.

FONLL structure function reduces to the massless calculation, while for $Q \sim m$ the FONLL result coincides with the massive calculation up to subleading (higher order) terms.

To illustrate the numerical impact the heavy quark mass effects have in deep-inelastic structure functions, in Fig. 3 we show the inclusive proton structure function $F_2(x, Q^2)$ at NNLO [101] as a function of Q^2 for two different values of x in the TR' GM-VFNS [102] as compared to the FFNS calculation. We see that differences can be as large as few percent, comparable or larger with the precision of available DIS data. In the same figure we show the NNLO charm structure function $F_2^c(x, Q)$ as a function of Q for $x = 0.01$ comparing the S-ACOT-χ [103] GM-VFNS with the corresponding ZM and FFN scheme calculations, where we can observe how the S-ACOT-χ calculation smoothly interpolates between the FFN scheme at low values of Q and the massless result at high Q.

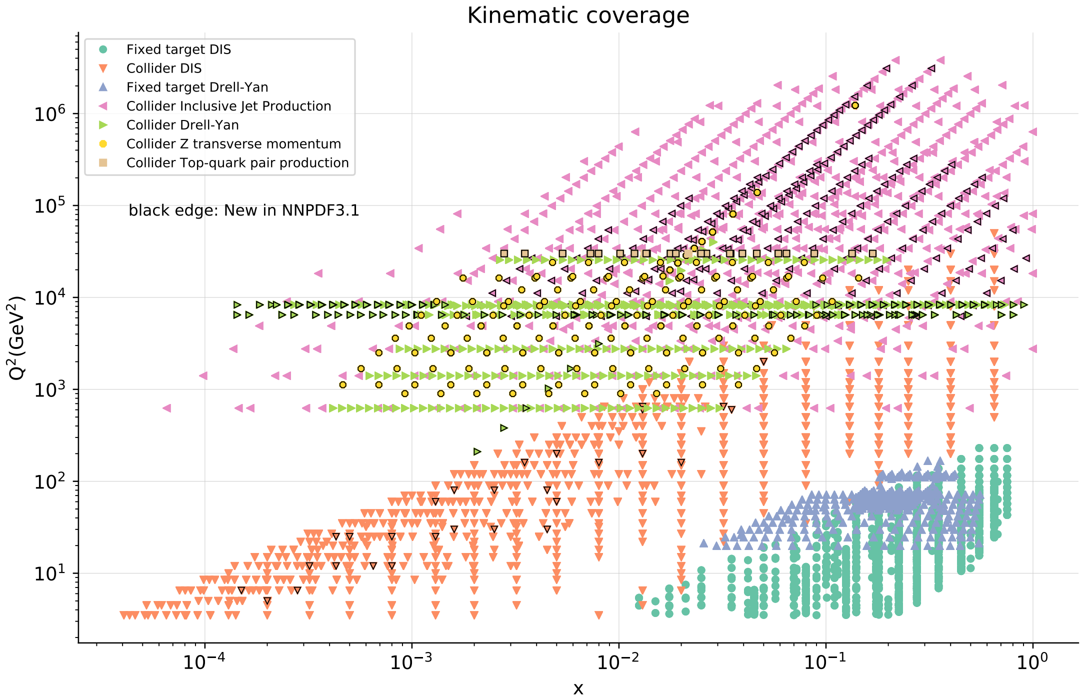


Figure 4: Typical kinematical coverage in the (x, Q^2) plane for the dataset included in a global analysis, in this case NNPDF3.1. For hadronic observables, leading order kinematics are assumed to map each data bin to a pair of (x, Q^2) values. The various datasets are clustered into families of related processes.

3. Experimental data and theoretical calculations

In this section we discuss the experimental data that is used in the global PDF analysis, as well as the status of the corresponding theoretical calculations and fast interfaces for their inclusion in the PDF fits. For each process we discuss first the PDF sensitivity, then the available data and state-of-the-art theory calculations, and finally illustrate its impact on PDFs. We start with a general overview of the datasets that are available for PDF studies and then we move to discuss each process separately, starting from DIS and then moving to inclusive jet and weak boson production, the p_T of Z bosons, direct photon and top-quark pair production, and charm production. In the last part of this section we discuss the important topic of fast (N)NLO interfaces.

3.1. Overview

We begin this section with a general overview of the datasets that are available for PDF studies, before moving on to discuss each process separately. In Fig. 4 we show a representative kinematical coverage in the (x, Q^2) plane for the dataset included in a global analysis, in this case the recent NNPDF3.1 fit [104]. We can see that a global dataset provides a rather wide coverage in the (x, Q^2) plane. The low–x and Q^2 region is dominated by the inclusive HERA structure function measurements, which provide information down to $x \sim 3 \cdot 10^{-5}$. The high–x region is covered by various processes, from fixed–target DIS structure functions at low Q^2 to collider jet, Drell–Yan and top–quark pair production at large Q^2. The very high Q^2 region, up to a few TeV^2, is only covered by inclusive jet production data from ATLAS and CMS. Until relatively recently most PDF fits were based only on DIS and fixed–target data, with some data from the Tevatron included. The breath of experimental information that is now included in the latest PDF fits is therefore quite impressive, with data from processes such as the Z p_T and the $t\bar{t}$ differential distribution only recently being considered for the first time.

In Table 1 we present another overview of the data entering a modern global PDF analysis. Here, we summarize the various hard scattering processes which are used to constrain PDFs in a global analysis. In each case we indicate the hadron–level process, the corresponding dominant parton–level process, as well as the partons which are constrained in each case and the corresponding range of x. Note that the latter are necessarily approximate, and only indicate in a qualitative way the x region that dominates the PDF sensitivity of each measurement. The necessity to include as broad a set of input datasets as possible, in order to constrain all quark flavour combinations and the gluon in the phenomenologically relevant region of x, is clear. We also note that the medium–to–low-x region, $x \lesssim 0.01$, is only covered by the HERA collider structure functions, and by some LHC data. The very low–x region, below the coverage of the HERA data, $x \lesssim 5 \cdot 10^{-5}$, can only be accessed via D meson production and exclusive J/ψ production.

In the rest of this section, we discuss the various processes that can be used to constrain the parton distributions in a global analysis one by one. We follow the same structure for each process: first of all we review the PDF sensitivity, then we discuss the available measurements, followed by a description of the state of the art of the corresponding theoretical calculations, to conclude with some representative illustrations of the impact of each family of processes in the PDF fit.

In the following, we restrict the discussion to theoretical calculations based on fixed–order perturbative QCD; see Refs. [105] and [106] for studies of the impact of the PDF fit of theory calculations based on all-order resummations of logarithmically enhanced terms at small-x and large-x respectively.

3.2. Deep-inelastic scattering

PDF sensitivity

Before the establishment of QCD as the renormalizable quantum field theory of the strong interaction, the results of DIS experiments were interpreted in the context of the so–called quark parton model. In this model, the proton was composed by non-interacting, co-moving quarks each of them carrying a given fraction x of its total momentum, and the DIS structure functions have particularly simple expressions in terms of the PDFs. Moreover, in this model the PDFs have a simple probabilistic interpretation, with $q_i(x)\Delta x$ giving the probability of finding a quark of flavour i inside the proton carrying out a momentum fraction in the range $[x, x + \Delta x]$. The expressions of the DIS structure functions in the quark parton model therefore provide a useful way to illustrate the PDF sensitivity of this process.

For the NC DIS structure functions F_2 and F_3, the quark parton model expressions are given by

$$\left[F_2^{\gamma}, F_2^{\gamma Z}, F_2^{Z}\right] = x \sum_{i=1}^{n_f} \left[e_i^2, 2e_i g_V^i, g_V^{i2} + g_A^{i2}\right] (q_i + \bar{q}_i) \ , \tag{32}$$

$$\left[F_3^{\gamma}, F_3^{\gamma Z}, F_3^{Z}\right] = x \sum_{i=1}^{n_f} \left[0, 2e_i g_A^i, 2g_V^i g_A^i\right] (q_i - \bar{q}_i) \ , \tag{33}$$

while the longitudinal structure function vanishes in this model, $F_L = 0$, and the superscripts on the LHS indicate the gauge boson which is being interchanged, as well as the contribution from the γZ interference term. In Eqns. (32) and (33), e_i is the electric charge of the quark of flavour i and the weak couplings are given by $g_V^i = \pm\frac{1}{2} - 2e_i \sin^2\theta_W^2$ and $g_A^i = \pm\frac{1}{2}$, where the $\pm$ corresponds to a u or d type quark. The sum runs over all the n_f quarks that are active for the specific scale at which the scattering takes place. From

	Process	Subprocess	Partons	x range
	$\ell^{\pm}\,\{p,n\} \to \ell^{\pm} + X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^{\pm}\, n/p \to \ell^{\pm} + X$	$\gamma^*\, d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
Fixed Target	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu})\, N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu\, N \to \mu^-\mu^+ + X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu}\, N \to \mu^+\mu^- + X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
	$e^{\pm}\, p \to e^{\pm} + X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
	$e^+\, p \to \bar{\nu} + X$	$W^+\,\{d, s\} \to \{u, c\}$	d, s	$x \gtrsim 0.01$
Collider DIS	$e^{\pm} p \to e^{\pm}\, c\bar{c} + X$	$\gamma^* c \to c,\ \gamma^* g \to c\bar{c}$	c, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^{\pm} p \to e^{\pm}\, b\bar{b} + X$	$\gamma^* b \to b,\ \gamma^* g \to b\bar{b}$	b, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^{\pm} p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
	$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
Tevatron	$p\bar{p} \to (W^{\pm} \to \ell^{\pm}\nu) + X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
	$p\bar{p} \to (Z \to \ell^+\ell^-) + X$	$uu, dd \to Z$	u, d	$x \gtrsim 0.05$
	$p\bar{p} \to t\bar{t} + X$	$qq \to t\bar{t}$	q	$x \gtrsim 0.1$
	$pp \to \text{jet} + X$	$gg, qg, q\bar{q} \to 2j$	g, q	$0.001 \lesssim x \lesssim 0.5$
	$pp \to (W^{\pm} \to \ell^{\pm}\nu) + X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u, d, \bar{u}, \bar{d}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X$	$q\bar{q} \to Z$	$q, \bar{q}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X,\ p_{\perp}$	$gq(\bar{q}) \to Zq(\bar{q})$	$g, q, \bar{q}$	$x \gtrsim 0.01$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, Low mass	$q\bar{q} \to \gamma^*$	$q, \bar{q}, g$	$x \gtrsim 10^{-4}$
LHC	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, High mass	$q\bar{q} \to \gamma^*$	$\bar{q}$	$x \gtrsim 0.1$
	$pp \to W^+\bar{c}, W^-c$	$sg \to W^+c,\ \bar{s}g \to W^-\bar{c}$	$s, \bar{s}$	$x \sim 0.01$
	$pp \to t\bar{t} + X$	$gg \to t\bar{t}$	g	$x \gtrsim 0.01$
	$pp \to D, B + X$	$gg \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to J/\psi, \Upsilon + pp$	$\gamma^*(gg) \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to \gamma + X$	$gq(\bar{q}) \to \gamma q(\bar{q})$	g	$x \gtrsim 0.005$

Table 1: Overview of the various hard-scattering processes which are used to constrain PDFs in a global analysis. In each case we indicate the hadronic-level process and the corresponding dominant partonic level process, as well as the partons which are constrained by each specific process in a given range of x. This table is an extended version of Table 1 of [80]. The x ranges are merely indicative and based on the approximate leading-order kinematics.

Eqns. (32) and (33) we see that the main limitation of the NC structure functions is that they provide limited access to quark flavour separation and in particular they cannot separate quarks from antiquarks, unless one goes to very high Q^2 values where the suppression induced by the Z boson propagator can be ignored.

In the case of CC DIS, the corresponding expressions for the structure functions in the parton model, assuming that we are above the charm threshold but below the top quark threshold, and the CKM suppressed transitions can be neglected, are given by

$$\begin{aligned}
F_2^{W^-} &= 2x\left(u + \bar{d} + \bar{s} + c\right) , \\
F_3^{W^-} &= 2x\left(u - \bar{d} - \bar{s} + c\right) , \\
F_2^{W^+} &= 2x\left(d + \bar{u} + \bar{c} + s\right) , \\
F_3^{W^+} &= 2x\left(d - \bar{u} - \bar{c} + s\right) ,
\end{aligned} \tag{34}$$

where again the longitudinal structure function $F_L^{W^\pm} = 0$ vanishes in this model. By comparing the NC and CC expressions, we can see that the main difference between them is that in the latter case the F_3^W structure function, which provides information on the difference between quark flavours, is not suppressed with respect to F_2^W. For this reason, CC structure functions are generally included in global fits in order to improve the discrimination between quarks and anti–quarks.

These quark parton model expressions are also valid at LO in perturbative QCD, once the effects of the DGLAP evolution are accounted for as described in Sect. 2.4. It is only at NLO that the contribution from the gluon PDF must also be included, and therefore the inclusive DIS structure functions will only be weakly sensitive to the gluon PDF, either through scaling violations (that is, the effect on the quark DGLAP evolution) or via the small $O(\alpha_s)$ contribution to the coefficient functions. The exception is the longitudinal structure function F_L, which vanishes at LO, and at NLO is non-zero and directly sensitive to the gluon PDF. Indeed, it can be shown that this structure function is given by

$$F_L(x, Q^2) = \frac{\alpha_s(Q^2)}{\pi}\left[\frac{4}{3}\int_0^1 \frac{dy}{y}\left(\frac{x}{y}\right)^2 F_2(y, Q^2) + 2\sum_i e_i^2 \int_x^1 \frac{dy}{y}\left(\frac{x}{y}\right)^2 (1 - x/y) g(x, Q^2)\right] , \tag{35}$$

which is known as the Altarelli-Martinelli relation. For this reason, F_L measurements can, in principle, provide direct constraints on the gluon in particular at low-x.

Finally, as well as the inclusive structure functions it is possible to determine the heavy quark structure functions experimentally, by selecting DIS events with charm or bottom mesons in the final state. The LO process proceeds via $\gamma g \to q\bar{q}$, see Fig. 5 (Right), and therefore heavy quark structure functions offer direct information on the gluon PDF, as well as on the treatment of heavy quark mass effects in the theoretical calculation. Charm structure functions in addition are an important ingredient for the determination of the charm mass m_c together with the PDFs. While data on F_2^b is known to have a small impact in the global fit, it is relevant for specific applications, for instance the determination of the bottom quark mass m_b from the PDF fit.

Experimental data

Since the pioneering DIS experiments at SLAC in the late 60s and early 70s, there have been many measurements of the DIS structure functions. These have been performed using either electrons, positrons or muons as projectile, and scattering off protons, deuterons and neutrons, either for fixed–target or for collider kinematics. We now discuss this various measurements in turn.

To begin with, the fixed-target DIS measurements available for PDF fits can be divided into neutral current and charged current datasets. In the NC case this includes:

- Proton and deuteron structure function data by the BCDMS collaboration [107, 108], using muons as projectiles.
- Proton and deuteron structure function data by the NMC collaboration [109, 110], as well as of the ratio between deuteron to proton structure functions, F_2^d/F_2^p.

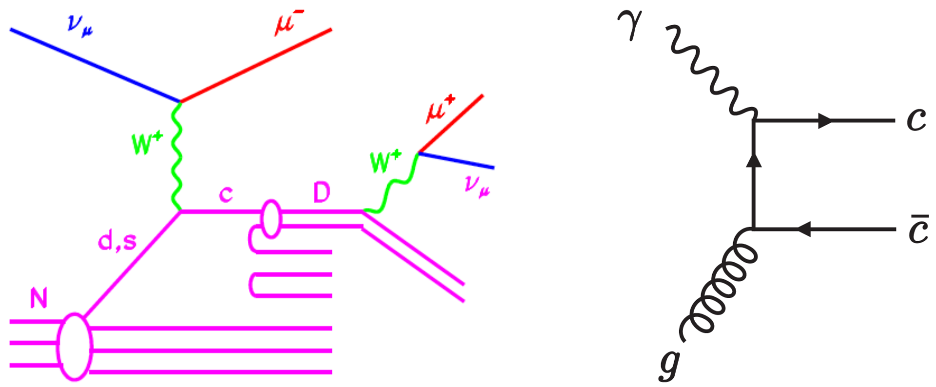


Figure 5: Left plot: D meson production in charged-current neutrino-induced DIS. This is known as the dimuon process since events are tagged where the D meson decays semi-leptonically, with the pair of oppositely charged muons providing a clean signature. Right plot: Charm production in neutral current DIS at leading order, highlighting the sensitivity of this process to the gluon PDF.

- SLAC measurements of the proton and deuteron NC structure functions [111].
- Proton, neutron and deuteron structure function data at high-x and low–Q^2 performed by JLAB experiments such as CLAS [112]. While these are excluded from most PDF fits by the typical DIS cuts in x and Q^2, these are included in the CJ fits.
- Older structure function data from the EMC collaboration [113]. Despite their age, the EMC measurements of F_2^c have never been repeated and thus provide unique information on the charm content of the proton at high–x.

In the CC case we have:

- Inclusive structure function measurements due to neutrino beams on nuclear targets, by the CDHSW, CCFR [114, 115] and CHORUS [116] and NuTeV [117] collaborations.
- Charm production in neutrino–induced DIS, often referred to as dimuon production, since the charm quark hadronizes into a D meson which then decays semi-leptonically, see Fig. 5. Data has been taken by the CCFR and NuTeV [118, 119] and and CHORUS [120] collaborations on the same nuclear targets as the corresponding inclusive measurements, and also by the NOMAD collaboration [121].

For the DIS measurements from the HERA lepton-proton collider we have:

- The final measurements of the NC and CC differential cross-sections using electron and positron projectiles from the the combination of the Run I and Run II data-taking periods [21]. These supersede all previous inclusive measurements from H1 and ZEUS, including the Run I inclusive measurements from H1 and ZEUS, including the Run I combined dataset [85] as well as the separate measurements by the two experiments from Run II [122, 123, 124, 125].
- The latest heavy flavour measurements from HERA include the combined NC cross-sections of charm production in DIS, $\widetilde{\sigma}_c$ [126] and the H1 and ZEUS data on the bottom structure function $F_2^b(x, Q^2)$ [127, 128].

This HERA legacy combination of DIS inclusive structure functions supersedes all previous inclusive measurements from H1 and ZEUS, including the Run I combined dataset [85] as well as the separate measurements by the two experiments from Run II [122, 123, 124, 125]. The impact of replacing these individual datasets by the final HERA combination of inclusive structure functions has been studied by different groups [129, 130, 131], finding that the impact of this replacement is quite moderate in general.

We also note that previous measurements of the longitudinal structure function F_L by the H1 and ZEUS collaborations [123] are now superseded by the final inclusive HERA combination.

Theoretical calculations and tools

The coefficient functions of DIS structure functions in the neutral current case are available up to $O\left(\alpha_s^3\right)$ in the massless limit and up to $O\left(\alpha_s^2\right)$ taking into account heavy quark mass effects, though there has been considerable recent progress towards the completion of the $O\left(\alpha_s^3\right)$ calculation of massive DIS structure functions. For charged current structure functions, massless coefficients are available up to $O\left(\alpha_s^3\right)$ and massive coefficient functions up to $O\left(\alpha_s^2\right)$ [132].

3.3. Inclusive jets

Since the first run of the Tevatron at Fermilab, inclusive jet production at hadron colliders has provided the dominant constraint on the gluon PDF at large–x. The definition of jet cross sections starts from a well defined jet algorithm, which is usually chosen to be infrared and collinear safe so that the corresponding parton–level cross section can be calculated in perturbative QCD for hard scattering at high energies. The most commonly used jet algorithm at the LHC is the anti–k_T algorithm [133], provided with the 4–vector recombination scheme. Other common choices include the k_T algorithm [134, 135], the Cambridge-Achen algorithm [136], as well as the Midpoint algorithm [137], which was sometimes used at the Tevatron.

When comparing to the calculated parton–level cross section to the experimentally measured jet cross section, it is essential to correct these to the hadron level. That is, additional non–perturbative corrections due, for example, to the underlying event and hadronization effects, must be accounted for. These are usually provided by the experimental collaborations as multiplicative factors derived from leading–order event generators. The size of such corrections can be significant at low-p_T, as high as $\sim 20\%$, while at high-p_T they are generally small, at the percent leve [138]. Variations of these non–perturbative corrections, by considering for example difference generator predictions, are then treated as an additional source of correlated systematic error. Although PDF fits typically use parton–level predictions, results also exist which include the matching of NLO calculations to parton shower and hadronization [139], which can be directly compared with the data at hadron–level.

PDF sensitivity

At LO jet production at hadron colliders includes the following subprocesses

$$\begin{aligned}&gg \to gg,\ gg \to q\bar{q},\ gq \to gq,\ q\bar{q} \to gg,\\&q\bar{q} \to q\bar{q},\ q\bar{q} \to q'\bar{q}',\ q\bar{q}' \to q\bar{q}',\ qq \to qq,\ qq' \to qq',\end{aligned} \tag{36}$$

along with the charge conjugate processes. Thus, jet production is sensitive to both the gluon and quark PDFs. The kinematics of the two leading jets in the final state can be characterized by their rapidities $y_{(1,2)}$ and their transverse momenta $p_{T,(1,2)}$. At LO we have $p_{T,1} = p_{T,2} = p_T$, and the momentum fractions carried by the two incoming partons are given by

$$x_1 = \frac{p_T}{\sqrt{s}}(e^{y_1} + e^{y_2}), \quad x_2 = \frac{p_T}{\sqrt{s}}(e^{-y_1} + e^{-y_2}), \tag{37}$$

where $\sqrt{s}$ is the center of mass energy of the two incoming hadrons. If we instead consider the rapidity of the jet in the centre–of–mass frame of the dijet system, $y^* \equiv (y_1 - y_2)/2$, and the boost of the dijet $y_b \equiv (y_1 + y_2)/2$, we have

$$x_1 x_2 = \frac{4p_T^2 \cosh^2 y^*}{s}, \quad x_1/x_2 = e^{2y_b}. \tag{38}$$

Beyond LO there can be multiple jets in the final state from additional QCD radiation.

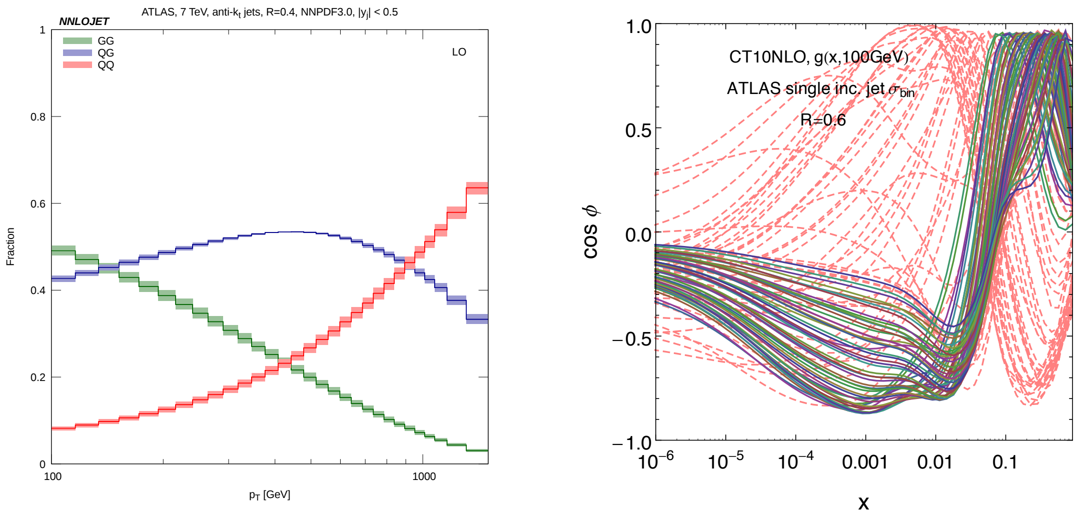


Figure 6: Left: Fractional contributions from different partonic channels to the single inclusive jet production at the LHC 7 TeV at LO in the central rapidity region [140]. Right: Correlations between binning cross sections from ATLAS on the single inclusive jet production at the LHC 7 TeV and the gluon PDF; dashed curves correspond to experiment bins at low p_T.

Experimentally, jet production can be measured in various ways. The most commonly used type for PDF fits is the single inclusive jet cross section, double differential in the jet p_T and rapidity. Here, one count all jets in a single event and includes them in the same distribution. Such a double differential cross section can be sensitive to different flavor combinations, depending on the kinematic region considered. In Fig. 6 (left) the fractional contributions from the different parton–level subprocesses to the inclusive jet cross section in central rapidity region at the LHC is shown, as a function of the jet p_T. We can see that at low p_T the channels involving initial–state gluons are dominant, while at higher p_T the $q\bar{q}$ contribution increases, but nonetheless with a sizeable gluon–induced fraction. As the quark PDFs are generally already well constrained by DIS data in these kinematic regions, jet data is therefore dominantly sensitive to the gluon PDF. This is illustrated in Fig. 6, which show the correlations between the inclusive jet cross section and the gluon PDF at various x values. This follows the ATLAS binning [141], with each curve corresponding to one bin. From this we can see that the inclusive jet production can further constrain the gluon PDF in a wide range of x, $10^{-3} \sim 1$.

In addition to the single inclusive case, there are also measurements of the double differential cross sections for inclusive dijet production, that is with repect to y^* and invariant mass of the two leading jets, or even triple differential cross sections, e.g., with respect to y_b, y^*, and average p_T of the two leading jets. Through such refined binning one can probe different initial states more efficiently. The large y_b region usually receives more contributions from gluon initial states, while at large y^* and p_T initial states with two valence quarks dominate, allowing the d-valence PDF at high-x to be further constrained.

Experimental data

The currently available measurements on jet production at hadron collider which are relevant for constraining the PDFs are as follows:

- The double differential single inclusive jet production cross section data from the CDF [142, 143] and D0 [144, 145] collaboration, at Tevatron Run II (1.96 TeV).
- The double differential single inclusive jet production cross section data from the ATLAS [146, 141, 147, 148] and CMS [149, 150, 138] collaborations at LHC Run I (7 and 8 TeV).
- The double differential inclusive dijet production cross section data from the ATLAS [146, 141, 151] and CMS [152, 149] collaborations at LHC Run I (7 and 8 TeV).
- The triple differential inclusive dijet production cross section data from the CMS collaboration [153] at LHC Run I (8 TeV).
- The measurements of the ratio of double differential cross sections in single inclusive jet production at different centre–of–mass energies, 2.76, 7 and 8 TeV, from the ATLAS [147] and CMS [138] collaborations, at LHC Run I.
- The double differential single inclusive jet production cross section data from the CMS collaboration [154] at LHC Run II (13 TeV).
- More recently, measurements of triple differential dijet cross sections are becoming available, see e.g. the recent CMS analysis [155] at 8 TeV.

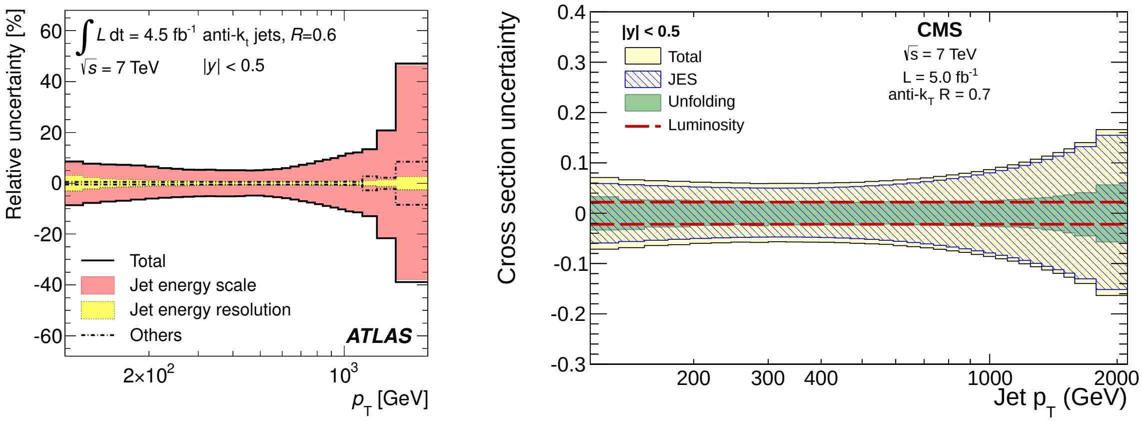


Figure 7: Representative systematic errors in single inclusive jet measurement at LHC 7 TeV in the central rapidity region, from ATLAS (left) [148] and CMS (right) [150]. The luminosity error is not included in the ATLAS plot.

Due to the complexity of jet reconstruction and calibration there are a large number of experimental systematic uncertainties, with $\sim 50-100$ correlated systematic errors for $100 \sim 200$ data points, in the case of both ATLAS and CMS. In the most recent ATLAS and CMS 7 TeV measurements [148, 150], the total correlated experimental uncertainties are at a level of about 5 $\sim$ 20% in most regions. On the other hand, the uncorrelated systematic errors and statistical errors are at one percent level or less in general, and therefore the uncertainty on such data is generally completely systematics dominated. The typical experimental systematics from both ATLAS and CMS are shown in Fig. 7, and are seen to be dominated by the jet energy scale [148, 150]. The increasing precision of the LHC jet data, and the generally small uncorrelated errors, makes it rather challenging to fit the jet data well across the entire kinematic region in e.g. the case of the ATLAS 7 TeV measurement. A full account of these issues will almost certainly require a better understanding of both the experimental systematics and sources of theoretical errors that have not generally been included in PDFs fits previously. In addition, for measurements of the ratios of the double differential inclusive jet production cross sections at different centre–of–mass the experimental systematic errors largely cancel out, although the statistical uncertainties are somewhat larger, see e.g. the CMS 2.76, 7 and 8 TeV [138] and ATLAS 2.76 and 7 TeV [147] measurements.

Theoretical calculations and tools

The NLO QCD corrections to single inclusive jet and inclusive dijet production was first calculated in the early 90's [156, 157], and has been implemented in two numerical programs, NLOjet++ [158, 159] and MEKS [160]. Recently, the NNLO QCD corrections to the same process have been completed for all partonic channels [161, 162, 163], with the exception of some sub–leading colour contributions. The calculation is based on the Antenna subtraction method [164, 165] for isolating the infrared singularities in QCD real radiations.

Fig. 8 (Left) shows the NNLO QCD corrections to inclusive jet production at the 7 TeV LHC, with the anti–k_T algorithm and a central scale choice of the leading jet p_T. The NNLO QCD corrections are seen to be significant at low-p_T, leading to a 10% increase with respect to NLO, while at high-p_T the NNLO corrections are small. The NLO scale variations bands are asymmetric at low-p_T and, interestingly, largely underestimate the perturbative uncertainties. EW corrections can be significant at high-p_T for central rapidities due to the presence of large EW Sudakov logarithms, but are well below 1% for a rapidity greater than 1 [166].

There are ambiguities in choosing the appropriate QCD scale even in the simplest case of single inclusive jet production. In particular, one can take either the p_T of the individual jet or the leading jet in the event. While these variables are the same at LO, where the two jets are produced back–to–back, at higher orders there exist more than two jets which can have large differences in p_T. The NNLO predictions using these two choices for the central scales are studied in [140], and are found they lead to vary significantly. This is shown in Fig. 8 (Right), where at high-p_T the two predictions converge as expected, but at low and intermediate p_T, there are significant differences of the central values in comparison to the size of scale variations. Indeed, the two error bands do not even overlap. Although it seems that the NNLO predictions using the individual jet p_T as the central scale tend to follow the trends of ATLAS data better, clearly further investigations are needed to resolve the ambiguity of scale choice in the NNLO predictions.

As well as fixed–order predictions, there are various theoretical calculations including analytic QCD resummation [167, 168, 169, 170, 171]. It has been shown in [170] for the case of inclusive jet production at the LHC, that the approximate NNLO predictions from the expansion of threshold resummation agree well with the exact NNLO predictions for the all–gluon channel at large p_T. Over the full rapidity range the threshold expansion reproduces the fixed–order results down to a p_T of about 400 GeV with the same value shifted to lower p_T for large rapidity region. Such approximate NNLO predictions have been used in previous global analysis involving jet data [19, 17]. The jet cross sections are also sensitive to the jet algorithm used, in particular on the value of the cone size or the distance parameter. A larger cone size

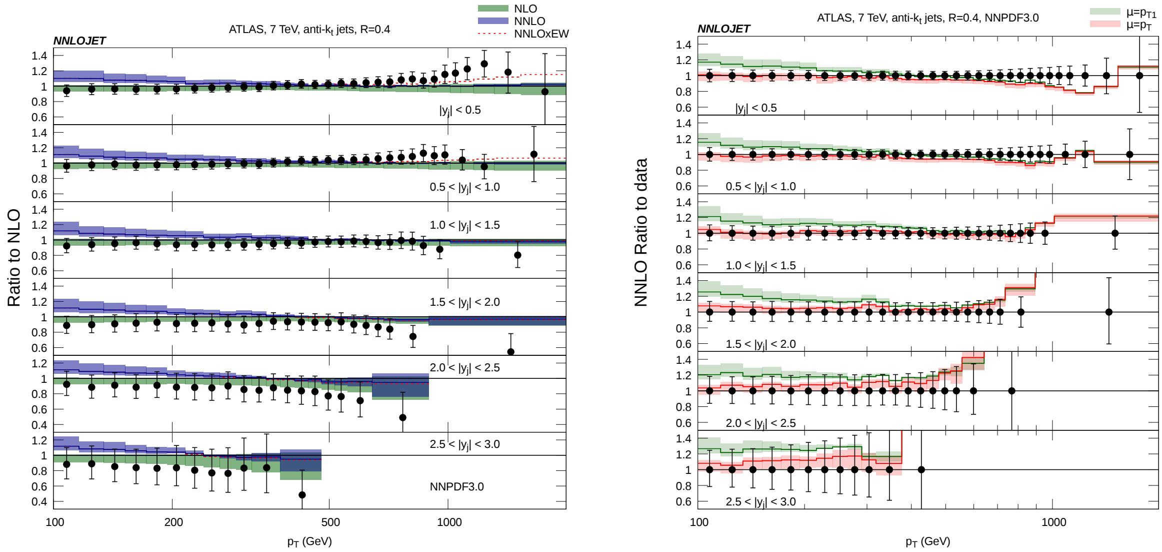


Figure 8: Left: Predictions on single inclusive jet production at the LHC 7 TeV using ATLAS binning and anti-k_T algorithm with $R = 0.4$, with a central scale choice of leading jet p_T and scale variations by varying renormalization and factorization scales simultaneously by a factor of 2 [140]. Right: For the same setup comparing the NNLO predictions using a central scale choice of the leading jet p_T (green) and the individual jet p_T (red) [140].

usually leads to a larger inclusive cross section and better convergence in the perturbative expansion. While this also reduces the non-perturbative corrections from QCD hadronization, it increases the correction from underlying events. At the LHC, ATLAS uses distance parameters of 0.4 and 0.6, while CMS uses 0.5 and 0.7.

Impact on PDFs

Jet data from the Tevatron and LHC Run I have already played an important role in global analyses [172, 17, 18, 19], although in the NNLO fits these only currently apply NLO or approximate NNLO theoretical predictions. Indeed, it was found that removing all jet data from the global analyses can lead to an increase of the gluon PDF uncertainties at large-x by at least a factor of two [17]. There are also independent studies from CMS [173, 138, 153] on the effects of jet data on the PDFs, based on NLO fits and using the xFitter program [27]. Fig. 9, taken from [153] shows the impact of the CMS 8 TeV jet data on the gluon PDF, by adding the data into a base fit with HERA DIS data only [21]. The inclusion of both the single inclusive jet data and the inclusive dijet data leads to a sizeable reduction in the gluon PDF uncertainty at large x. Meanwhile in the same fit a reduction of the PDF uncertainty in the valence quark at high x is also observed, providing a complementary constraint to Drell-Yan and fixed-target DIS data. With the full NNLO predictions on jet production now available we can expect significant advances in pinning down the gluon PDF at large x, in particular using the increasingly precise inclusive jet data from LHC Run I and Run II.

3.4. *Inclusive gauge boson production*

Now we turn to discuss the inclusive production of electroweak gauge boson. This process has been of enormous historical importance since it provided a first window on the quark flavour separation in the

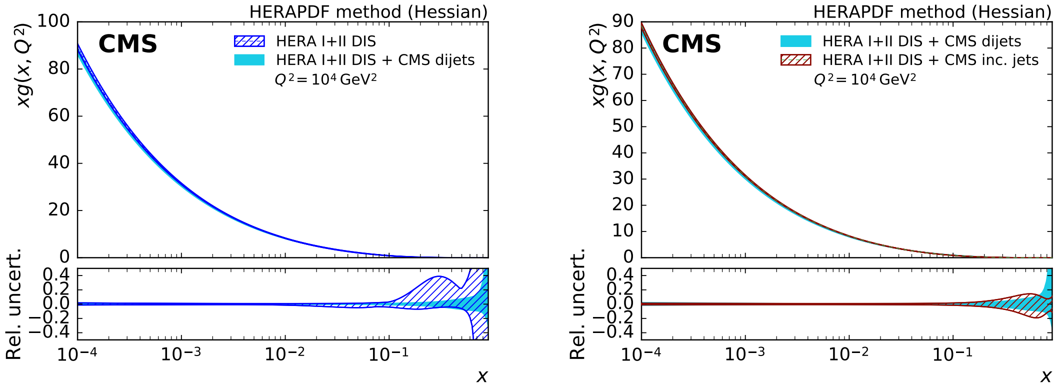


Figure 9: Effects of the CMS 8 TeV jet data on the gluon PDF when adding up to a PDF fit with HERA DIS data only. The left plot compares the fit with only DIS data and the fit with CMS dijet data in addition. The right plot compares the fit to HERA DIS plus CMS dijet data and fit to HERA DIS plus CMS single jet data [153].

proton beyond the information contained on DIS structure functions. Nowadays, Drell-Yan cross-sections provide the backbone of global PDF fits together with the fixed-target and HERA structure function data.

PDF sensitivity

The lowest order contributions to W and Z/γ^* production proceed via

$$u\bar{d},\ c\bar{s} \quad (u\bar{s},\ c\bar{d}) \to W^+\ , \tag{39}$$

$$d\bar{u},\ s\bar{c} \quad (s\bar{u},\ d\bar{c}) \to W^-\ , \tag{40}$$

$$q\bar{q} \to Z/\gamma^*\ , \tag{41}$$

where we show the Cabibbo suppressed contributions in brackets and q corresponds to all active flavours. These processes can therefore tell us about the flavour decomposition of the proton. To examine the dominant PDF sensitivity we can approximate the CKM matrix as diagonal, and thus ignore the bracketed contributions. In this case it is informative to consider the ratio of W^+ to W^- production

$$R_\pm = \frac{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W}{\mathrm{d}\sigma(W^-)/\mathrm{d}y_W} = \frac{u(x_1)\bar{d}(x_2) + c(x_1)\bar{s}(x_2) + 1 \leftrightarrow 2}{d(x_1)\bar{u}(x_2) + s(x_1)\bar{c}(x_2) + 1 \leftrightarrow 2}\ . \tag{42}$$

and the W asymmetry

$$A_W = \frac{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W - \mathrm{d}\sigma(W^-)/\mathrm{d}y_W}{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W + \mathrm{d}\sigma(W^-)/\mathrm{d}y_W} = \frac{u(x_1)\bar{d}(x_2) + c(x_1)\bar{s}(x_2) - d(x_1)\bar{u}(x_2) - s(x_1)\bar{c}(x_2) + 1 \leftrightarrow 2}{u(x_1)\bar{d}(x_2) + c(x_1)\bar{s}(x_2) + d(x_1)\bar{u}(x_2) + s(x_1)\bar{c}(x_2) + 1 \leftrightarrow 2}\ . \tag{43}$$

We will for simplicity consider the W rapidity, rather than the experimentally observable rapidity of the charged lepton from the W decay, in what follows. These variables are clearly correlated; we will comment further on this at the end.

Thus these ratios are in general sensitive to a fairly non–trivial combination of quark and anti–quark PDFs at $x_{1,2} = \frac{M_W}{\sqrt{s}} e^{\pm y_W}$. While these expression completely define the PDF sensitivity of these observables at LO, it is informative to consider various kinematic limits, where these expressions simplify and more straightforward approximate dependences become apparent. Including only the (dominant) u and d contributions, we can in particular consider the cases of central and forward W production

$$\text{Central}: \quad y_W, \sim 0 \quad x_1 \sim x_2 = x_0, \quad \bar{u}(x_{1,2}) \sim \bar{d}(x_{1,2}) \,, \tag{44}$$

$$\text{Forward}: \quad y_W \gtrsim 2, \quad x_1 \gg x_2, \quad q(x_1) \sim q_V(x_1),\, \bar{u}(x_2) \sim \bar{d}(x_2) \,, \tag{45}$$

where $x_0 = M_W/\sqrt{s}$ and $q = u, d$. At the LHC we have $x_0 = 0.005 - 0.01$, while in the forward region $x_2 \ll 1$, and therefore the $\bar{d} \sim \bar{u}$ approximation is a very good one. For the case of negative W rapidity we can of course simply interchange $x_1 \leftrightarrow x_2$.

In the central region, applying the simplification of (44) and dropping the c, s contributions we find

$$R_\pm \sim \frac{u(x_0)}{d(x_0)} \,, \tag{46}$$

$$A_W \sim \frac{u_V(x_0) - d_V(x_0)}{u(x_0) + d(x_0)} \,. \tag{47}$$

Thus A_W is sensitive to the valence difference, while $R_\pm$ is sensitive to the ratio of u to d at $x_1 \sim x_2 \sim x_0$. For these reasonably low x values, the valence u and d quarks are fairly small, and so we roughly expect $R_\pm \sim 1$ and $A_W \sim 0$, with the departures from these values being due to the precise flavour content of the proton, in particular the fact that the valence distributions are not completely negligible in this region.

In the forward region, applying the simplification of (45) and again dropping the c, s contributions we find

$$R_\pm \sim \frac{u_V(x_1)}{d_V(x_1)} \,, \tag{48}$$

$$A_W \sim \frac{u_V(x_1) - d_V(x_1)}{u_V(x_1) + d_V(x_1)} \,. \tag{49}$$

Thus these provide (equivalent) sensitive constraints on the u/d ratio at high x

Considering now the case of Z production, then for forward production we find

$$\frac{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W + \mathrm{d}\sigma(W^-)/\mathrm{d}y_W}{\mathrm{d}\sigma^Z/\mathrm{d}y_Z} \approx \frac{u_V(x_1) + d_V(x_1)}{0.29 u_V(x_1) + 0.37 d_V(x_1)} \,, \tag{50}$$

where the factors in the denominator come from the electroweak Z–quark couplings. For the central region a similar result evaluated at x_0, is found, up to an overall factor of 2. Thus, the $W^\pm$ and Z cross sections provide very similar information about the u and d quarks.

Up to this point we have omitted the contribution from the strange quarks to W and Z production. Generally speaking this is washed out when considering ratio observables, justifying their omission above, although the W asymmetry displays some sensitivity to the strange difference $s - \bar{s}$. On the other hand the contribution to the absolute cross sections is not negligible, in particular at lower x. Thus for example the Z cross section at central rapidity becomes, for five active flavours

$$\frac{\mathrm{d}\sigma^Z}{\mathrm{d}y_{ll}} \sim 0.29\Big(u(x_0)\bar{u}(x_0) + c(x_0)\bar{c}(x_0)\Big) + 0.37\Big(d(x_0)\bar{d}(x_0) + s(x_0)\bar{s}(x_0) + b(x_0)\bar{b}(x_0)\Big) \,. \tag{51}$$

and so, provided the absolute cross section data are sufficiently accurate and the other quark flavours are sufficiently well determined, this may for example be sensitive to the currently less well determined strange quark distribution. Moreover, this is not a case of a simple overall normalization; as the Z rapidity increases the valence u, d contributions will become increasingly dominant, and the contribution from the strange (and the heavy flavours) will decrease. Thus the shape of the Z rapidity distribution is sensitive to the proton strangeness, as well as the heavy flavour PDFs. Similar considerations also apply for the absolute $W^{\pm}$ cross sections.

Moving away from the Z peak region, the Drell–Yan process is dominated by an off–shell intermediate photon, with

$$\frac{\mathrm{d}\sigma^{\mathrm{DY}}}{\mathrm{d}y_{\mathrm{ll}}} \sim \sum_i e_i^2 \Big(q(x_1)\overline{q}(x_2) + q(x_2)\overline{q}(x_1) \Big) \,. \tag{52}$$

Thus in comparison to (51) a different combination of the quark and anti–quark PDFs is probed, due to the differing electromagnetic couplings. In particular, the relative $u\overline{u}$ to $d\overline{d}$ contribution is now a factor of ~ 5 higher in comparison to the Z cross section. At the LHC, low mass Drell–Yan production therefore provides complementary flavour information in the low to intermediate x region. In addition, as the cuts on the final–state lepton transverse momenta tend to increase the relative important of the higher order contributions, for which the Z $p_\perp$ can be non–zero, this can be sensitive to the gluon PDF at lower x, which contributes through the NLO $g \to q\overline{q}$ splitting. High mass Drell-Yan production is sensitive to the $q, \overline{q}$ PDFs at high x, in particular the anti–quarks, which are less well determined in this region.

A further constraint is provided by considering the Drell–Yan process on fixed proton and neutron (in practice, deuteron) targets. By using isospin symmetry the PDFs between the proton and the neutron can be related

$$u^p = d^n \qquad d^p = u^n \,, \tag{53}$$

allowing an extra handle on the proton flavour decomposition. In particular, such fixed target experiments generally have larger acceptance in the $x_1 \gg x_2$ region (where x_1 is defined with respect to the proton beam) for which the first term in (52) is dominant, with $q(x_1) \sim q_V(x_1)$. It is then straightforward to show that

$$\frac{\sigma^{pn}}{\sigma^{pp}} \sim \frac{\overline{d}(x_2)}{\overline{u}(x_2)} \,. \tag{54}$$

That is, they are sensitive to quark sea decomposition in the intermediate to high $x_2 \sim 0.01 - 0.3$ region probed by these fixed target experiments [174]. This however comes with the added complication that the nuclear corrections accounting for the fact that the neutron is bound in a deuteron nucleus, and therefore the 'free' neutron PDF is not directly probed. Fixed target pp scattering alone does not suffer from this issue, and is sensitive to the quark sea (dominantly, the $\overline{u}$) in the same x region, but is much less directly sensitive to the $\overline{d}/\overline{u}$ decomposition.

Turning now to the case of W, Z production at the Tevatron, the fact that we have $p\overline{p}$ collisions affects the flavours probed. In particular, we can use charge–conjugation symmetry to write

$$q^p = \overline{q}^{\overline{p}} \,. \tag{55}$$

In fact, it is straightforward to show that in the region of valence quark dominance, the cross section ratio $R^{\pm}$ and the asymmetry A_W are again sensitive to the u/d ratio and the valence difference $u_V - d_V$, while the Z cross section again provides similar information to $W^{\pm}$ cross section sum. Nonetheless, these conclusions are only approximately true, and the presence of a $\overline{p}$ beam provides complementary flavour information.

Finally, we have considered above the distributions with respect to the (unobservable) rapidity of the W boson to simplify the discussion. In general we should correctly account for the kinematics, as well as weight the corresponding $q\overline{q}$ contributions by the appropriate W decay distributions. This is in fact provides a further handle on the flavour sensitivity of this observable, as by changing the $p_\perp$ cut on the charged lepton, different weights of the different quark contributions are achieved, see e.g. [80, 175] for further details. Nonetheless, the forward and central W rapidity regions are certainly correlated with the equivalent lepton rapidity regions that are measured experimentally, and so the above discussion provides a qualitative guide for the PDF sensitivity of W boson production. However, as we will discuss below, the current simulation codes for W and Z production include the full kinematics of the leptonic decays, and therefore there is no need to explicitly correct back to the W rapidity.

Experimental data

A non–exhaustive list of the available data is as follows:

- The most precise fixed target Drell–Yan data come from the E866/NuSea [176] experiment at Fermilab, while the E906/SeaQuest experiment [177] will extend out to higher x, and is currently taking data.
- The Tevatron collider has produced a range of data on W and Z production, including measurements of the Z rapidity distribution [178, 179] and in W production both the lepton [180, 181] and the W [182, 183] asymmetries.
- Early LHC measurements of the the Z rapidity distribution presented by CMS [184] and ATLAS [185].
- CMS Drell–Yan data at 7 TeV [186], for $15 < M_{ll} < 1500$ GeV and at 8 TeV [187], which increased the upper mass limit to 2000 GeV. These are presented double differentially in the rapidity and invariant mass of the lepton pair.
- ATLAS 7 TeV Drell–Yan invariant mass distribution (integrated over rapidity) at high [188] ($116 < M_{ll} < 1500$ GeV) and low [189] ($26 < M_{ll} < 66$) invariant masses.
- CMS [190, 191] 7 TeV W asymmetry, and ATLAS [185] W^+ and W^- cross section data.
- ATLAS high precision W and Z, γ^* data [192], using the full $4.6\,\text{fb}^{-1}$ data set at 7 TeV. The Drell–Yan rapidity distribution is presented double differentially in three intervals of lepton pair mass, over the $45 < M_{ll} < 150$ GeV range. In the Z peak and higher mass regions the measurement was also extended out to $|y_{ll}| = 3.6$.
- LHCb Z rapidity distributions at 7 [193, 194], 8 [195] and 13 [196] TeV.
- LHCb Lepton rapidity distributions for W^+ and W^- production at 7 [197], and 8 [198, 199] TeV.
- In [198] cross section ratios between the 7 and 8 TeV W and Z measurements are presented, with the cancellation in various systematic uncertainties providing a more precise PDF sensitivity.

Thus at the LHC multiple measurements have been presented. The Z/γ^* data are available over a wide range of invariant masses, providing extensive coverage in x. The W data are increasingly presented as individual cross sections, including the correlated error information, to provide the maximum possible constraints. While in the majority of cases, the ATLAS and CMS measurements are limited to the central rapidity region, that is a lepton pseudorapidity of $|\eta_l| < 2.4$, this reach is extended by exploiting the forward acceptance of the LHCb detector, for which $2 < \eta_l < 4.5$ is accessed. This allows the high and low x region to be probed. The most recent ATLAS W and Z, γ^* data [192], which uses the full $4.6\,\text{fb}^{-1}$ data set at 7 TeV demonstrates the level of precision that is now being achieved. The Z rapidity distribution

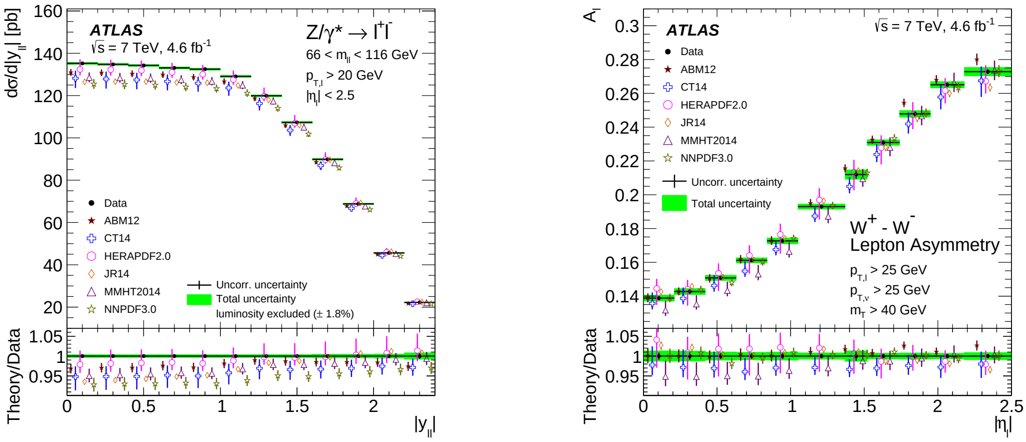


Figure 10: ATLAS 7 TeV measurement of Z rapidity distribution (left) and W asymmetry (right), taken from [192].

and W asymmetry are shown in Fig. 10 (in the latter case the individual W measurements are available). The high experimental precision is clear, in particular in the Z distribution where excluding the luminosity uncertainty it is as low as $\sim 0.3\%$ at central rapidity, while the error on the PDF predictions, as well as the spread between sets, is significantly larger. The impact of such data can therefore be sizeable.

Theoretical calculations and tools

W and Z boson production is arguably the simplest process one can consider at a hadron collider, and indeed it was the first hadroproduction process for which the NNLO calculation became available, with the total cross sections being calculated in the early 90s [96][2]. A decade later, in [201, 202] the NNLO corrections to the differential W and Z rapidity distributions was presented for the first time. A more direct comparison with experimental observables was provided in [203, 204] which presented the NNLO calculation fully differential in the final–state leptons, including in addition spin correlations, finite width effects and $\gamma - Z$ interference. This was accompanied by the public release of the `FEWZ` simulation code, with subsequent improvements reported in [205, 206] and [207], where NLO EW corrections (first calculated in [208, 209, 210, 211, 212, 213]) were included. The `DYNNLO` [214] parton level MC provides an alternative tool for generating W and Z production, again including spin correlations, finite width effects, and $\gamma - Z$ interference, but currently without EW corrections. This code allows for arbitrary user-defined cuts on the final–state partons and leptons to be imposed and histograms to be made, in contrast to `FEWZ`, where a selection of pre-determined cuts and histograms may be applied.

These two codes differ in their theoretical treatment of the processes, in particular in the method that is applied to achieve the (non–trivial) cancellation of IR singularities at intermediates steps in the calculation. While `FEWZ` uses the local 'sector decomposition' method [203, 204] that provides an automated method for extracting and cancelling the IR poles, `DYNNLO` applies an alternative non–local 'q_T-subtraction' approach [215] which uses the transverse momentum q_T of the produced W or Z as a cut variable, treating the calculation in a different way above and below some $q_T^{\rm cut}$. It is unfortunately now quite well established that these codes can give non–negligible differences in their predictions for identical input parameters. For example, in the recent ATLAS high precision W and Z/γ^* analysis [198] the difference in the fiducial cross

[2] An error in the one–loop real emission contribution was reported in [200].

section predictions can be as high as $\sim 1\%$, that is larger than the experimental uncertainties. This is due to the differing subtraction procedures, which affects the predicted boson $p_\perp$ distributions. The differences between the predicted cross sections are generally more significant when more restrictive cuts on the final–state leptons are imposed; for the total W, Z cross sections the codes agree to within 0.2% [198]. A closer investigation of this issue and its impact on PDF determination will clearly be essential.

More recently the `MCFM` event generator [216] has extended the NLO simulation of W and Z production to NNLO [217]. This takes a similar non–local approach to `DYNNLO`, but using the N–jettiness variable rather than the q_T. Here, it is shown that a careful and process–dependent choice of the cut on the 0–jettiness variable, $\tau_0^{\rm cut}$ (the equivalent of $q_T^{\rm cut}$ above) is required in order to balance the requirements of sufficient statistical precision and control over systematic power corrections that increase in importance as this cut is increased.

Finally, event generators including transverse momentum resummation are also available. The `DYRes` [218] code combines NNLO fixed–order with NNLL resummation, while `ResBos` [219] combines NLO fixed–order with NNLL resummation. However, typical observables that are used in PDF fits are chosen to be largely insensitive to such resummation effects, which are most important as the W, Z transverse momentum becomes small, and so these codes are in general not used in PDF analyses.

Impact on PDFs

As described in Section 3.4 inclusive gauge boson production has played a crucial role in determining the quark flavour decomposition of the proton. Indeed, these have been included in all major PDF analyses for some time, from earlier fixed target data through to measurements at the Tevatron and increasingly at the LHC. Two recent LHC results are show in Fig. 11. In the left panel we show the CMS fit [220] to the down valence quark distribution. The baseline fit is to the HERA I+II data only, which is compared to the result including the CMS 8 TeV W boson production data. The change in shape and sizeable reduction in the PDF uncertainty over a wide range of x is clear.

In the right panel we show the impact on the strange quark fraction relative to the light quark sea

$$R_s = \frac{s+\overline{s}}{\overline{u}+\overline{d}}\,, \tag{56}$$

of the ATLAS high precision W and Z/γ^* data [192]. As described in Section 3.4, provided the light quark flavours are sufficiently well determined, and the data are sufficiently precise, the size and shape of the W, Z rapidity distributions can provide constraints on the strange quark PDFs. This is clear from the figure, where the fit to the ATLAS data predict a significantly higher value of R_s in comparison to previous PDF fits, which do not include the ATLAS data. A hint of this effect is seen in the earlier 'ATLAS-epWZ12' result [221], but it is only with the more recent high precision data that a clear effect becomes apparent.

3.5. *The p_T of Z bosons*

The LHC has provided precision measurements of inclusive transverse momentum spectra of the Z boson produced in hadronic collisions, which may be exploited for the purposes of PDF fitting. There are three distinct regions of the p_T spectrum. At small $p_T \ll m_Z$, the fixed–order predictions diverge due to higher–order logarithms generated by soft gluon radiation. Here, QCD resummation is needed to maintain reliable predictions, see [222, 223, 224, 225, 226, 227, 228, 229]. Such predictions require additional non–perturbative input that cannot be calculated from first principles [230, 226], and therefore the Z p_T distribution cannot be reliably used for the extraction of the collinear PDFs in this region.

At large $p_T \gg m_Z$, the fixed–order predictions can also receive large logarithmic contributions due to soft gluon radiation at the partonic threshold of the Z boson and the recoiling jet [231, 232]. It has been

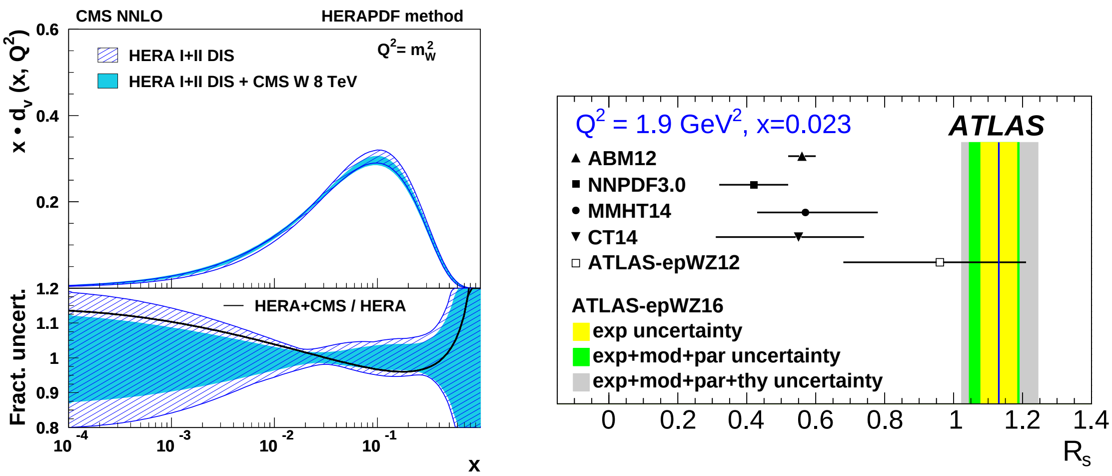


Figure 11: (Left) Down valence distribution, with the CMS fit to HERA I+II data only, and including the CMS 8 TeV W production data. Taken from [220]. (Right) The ratio of the strange quark to the light quark sea, R_s, with different PDF predictions and the result of the ATLAS fit to HERA I+II data and the high precision W and Z/γ^* data shown. Taken from [192].

shown that those contributions can increase the the accumulated cross sections with $p_T > 200\,\text{GeV}$ by $\sim 5\%$ compared to the NLO prediction at the LHC [231]. For intermediate $p_T \sim m_Z$, the fixed–order predictions can be trusted and therefore in this region the distribution can provide additional constraints on the PDFs, in particular the gluon.

PDF sensitivity

At LO Z boson production with finite transverse momentum includes the following subprocesses

$$q\bar{q} \to Zg,\;\; gq \to Zq,\;\; g\bar{q} \to Z\bar{q}. \tag{57}$$

In the leptonic channel, the kinematics of the Z boson, namely the transverse momentum p_T and rapidity y_Z, can be reconstructed from the momenta of the lepton pair produced in the Z decay. The momentum fractions of the initial–state partons are given by

$$x_1 = \frac{m_T}{\sqrt{s}}e^{y_Z} + \frac{p_T}{\sqrt{s}}e^{y_j}, \quad x_2 = \frac{m_T}{\sqrt{s}}e^{-y_Z} + \frac{p_T}{\sqrt{s}}e^{-y_j}, \tag{58}$$

where $\sqrt{s}$ is the center of mass energy of the two incoming hadrons, m_T is the transverse mass of the Z boson and y_j is the rapidity of the recoiling parton. For inclusive production with respect to the hadronic recoil, that is integrating over y_j, these momentum fractions are therefore not uniquely determined, although for LO kinematics lower limits can be derived from the above equation. Usually experiments measure the double differential cross sections in p_T and y_Z at the Z peak, although the off-shell region, where the contributions from virtual photon can be important, can also be considered.

The cross sections at moderate and large transverse momentum are dominated by contributions from the gluon and quark scattering and are strongly correlated with the gluon PDF in the region relevant for Higgs boson production at the LHC. That is illustrated in Fig. 12, which shows the PDF induced correlations between the cross sections in different p_T bin, in the rapidity interval $0 < |y_Z| < 0.4$, and the gluon, down- and up-quark PDFs at various x values [233]. We can see that indeed the correlations with the gluon at $x \sim 10^{-2}$ almost reach 0.9. Moderate correlations with quark PDFs at $x \sim 10^{-3}$ are also observed.

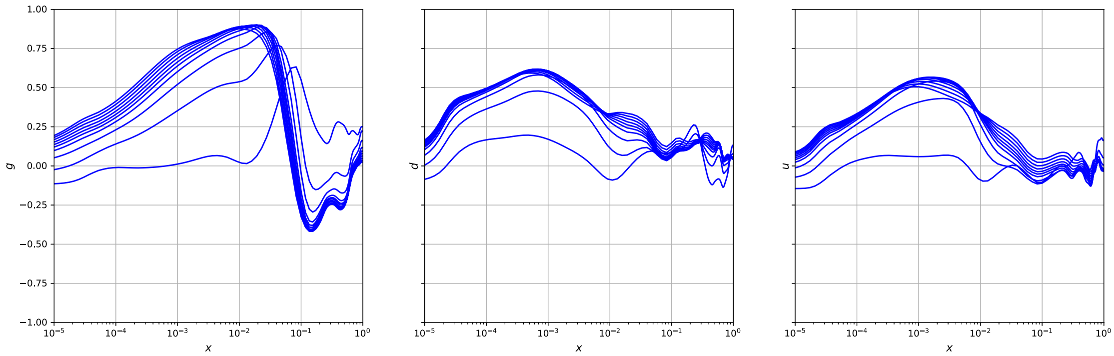


Figure 12: Correlations between the cross sections in various p_T bins and the gluon, down- and up-quark PDFs as a function of x [233]. The binning corresponds to the ATLAS measurement [234] with rapidity interval $0 < |y_Z| < 0.4$.

Experimental data

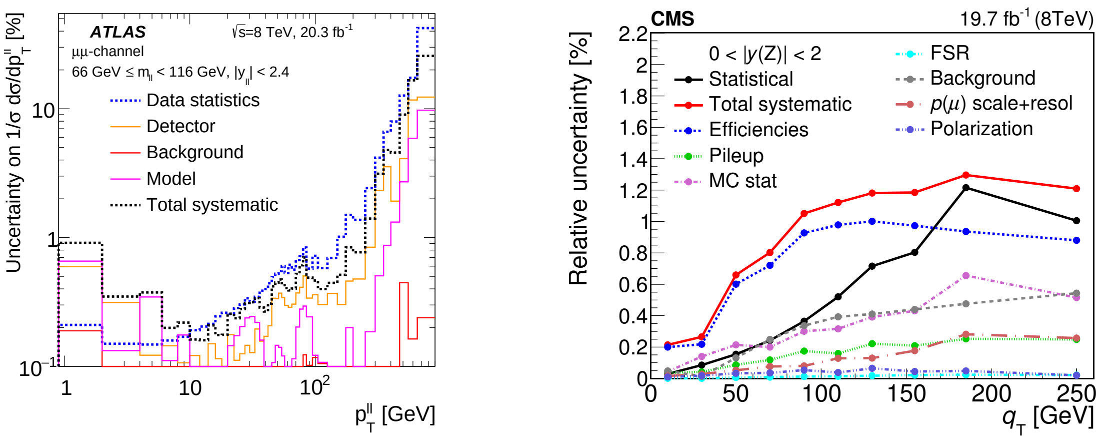


Figure 13: Left plot: relative experimental errors on the measured normalized p_T spectra of the Z boson in dimuon channel (dressed) from ATLAS 8 TeV [234] . Right plot: relative experimental errors on the measured normalized p_T spectra of the Z boson in dimuon channel from CMS 8 TeV [235] .

The experimental measurements of the p_T spectra of the Z boson have reached the percent level for both ATLAS [234] and CMS [235] in LHC Run I, due to the clean dilepton final state as well as the high statistics of the signal. The ATLAS measurement extracts the cross sections at three different ('Born', 'bare' and 'dressed') particle levels when considering the effect of final–state photon radiation. The Born and bare levels are defined from the lepton kinematics before and after final–state radiations, while the dressed level is defined by further combining the momentum of the lepton with photons radiated within a certain cone.

The distributions can be presented with the Z boson rapidity integrated over, or separated into different rapidity intervals, and can be on or off the Z–peak. In addition, measurements of the distributions with respect to the angular variable ϕ^*_η [234], which is proportional to $p_{T,Z}$ at small transverse momentum, are available. As ϕ^*_η only depends on the direction of the lepton momenta, which are better measured than the momenta themselves, this allows the experimental systematics to be reduced.

We summarise the available measurements on p_T spectra of the Z boson relevant to constraining the PDFs below:

- The normalized Z p_T distribution in different rapidity intervals by the ATLAS collaboration [236, 237] at LHC Run I (7 TeV).
- The normalized and unnormalized distributions of lepton pairs with respect to p_T or ϕ^*_η in different rapidity intervals by the ATLAS collaboration [234] at LHC Run I (8 TeV).
- The normalized Z p_T distribution integrated over rapidity by the CMS collaboration [184] at LHC Run I (7 TeV).
- The normalized and unnormalized double differential Z distribution in p_T and rapidity by the CMS collaboration [235] at LHC Run I (8 TeV).
- The normalized Z p_T distribution and the ratio to the W p_T distribution by the CMS collaboration [238] at LHC Run I (8 TeV).
- The unnormalized Z distribution in ϕ^*_η in the forward region by the LHCb collaboration [193, 195] at LHC Run I (7 and 8 TeV).
- The normalized Z distribution in p_T or ϕ^*_η in the forward region by the LHCb collaboration [194] at LHC Run I (7 TeV).
- The unnormalized distribution of the lepton pair with respect to ϕ^*_η in different rapidity intervals by the D0 collaboration [239] at Tevatron Run II (1.96 TeV).

A summary of the experimental uncertainties for the ATLAS and CMS 8 TeV measurements [234, 235] of the normalized Z p_T distribution is shown in Fig. 13. The luminosity uncertainty and some of the systematic errors largely cancel in the normalized distributions. Both ATLAS and CMS have measured the Z p_T up to about 1 TeV, while ATLAS has a finer binning at small p_T. For ATLAS, the statistical errors are well within 1% for p_T smaller than 200 GeV and are $O(10\%)$ at the higher p_T tail. The total systematic errors start at $\sim$ 2 per mil for $p_T \sim 10$ GeV, and are within 1% in most of the region for both ATLAS and CMS. The PDF uncertainties from individual PDF groups are about 2%, which is already larger than the experimental errors in general, even before considering the spread between different PDFs; such data can therefore provide valuable PDF constraints.

Theoretical calculations and tools

The NLO QCD corrections to the Z p_T distribution were calculated decades ago [240, 241, 242], while more recently the EW corrections have been studied extensively [243, 244, 245, 246]. The NLO QCD corrections are found to be sizeable at LHC energies, and large QCD scale variations are found in the predicted p_T spectra, rendering such predictions inappropriate for PDF determination. However, the NNLO QCD corrections have very recently been calculated by two independent groups, in one case using the antenna subtraction method [164, 247, 248] and in the other the N-jettiness subtraction method [249, 250]; these are found to be in good agreement. While the original calculations are for Z+jet production, these can readily be translated to the case of inclusive production of Z boson at finite p_T.

At NNLO, the theoretical uncertainty due to the QCD scale variation is fount to be greatly reduced, allowing the Z boson p_T spectra data to be included for the first time in precision PDF determination.

Moreover, these calculation include the leptonic decays of the Z boson and thus the parton–level selection cuts may be applied to the theoretical predictions, allowing a direct comparison with the measured fiducial cross sections without relying any experimental phase space extrapolation. Fig. 14 (taken from [248]) shows the NLO and NNLO predictions for the unnormalized and normalized Z boson p_T spectra at the 8 TeV LHC. The central values of the renormalization and factorization scale are set to the transverse mass of the Z boson, with scale variations calculated by varying these simultaneously by a factor of 2 up and down. The NNLO corrections are moderate for the unnormalized distribution, about 5% at low p_T and 9% at high p_T. The remaining scale variations range from 1% to 6% depending on value of p_T. The EW corrections are small at moderate transverse momentum but can be sizeable in the tail region, reaching $\sim -10\%$ for p_T greater than 600 GeV. However, as the statistical errors in the tail region are currently quite large, this prevents a direct probe of these EW effects. For the normalized distribution, the denominator used is the inclusive Z production cross section at NNLO in same fiducial region and with independent scale variations. The size of the QCD corrections are found to be similar to the unnormalized case.

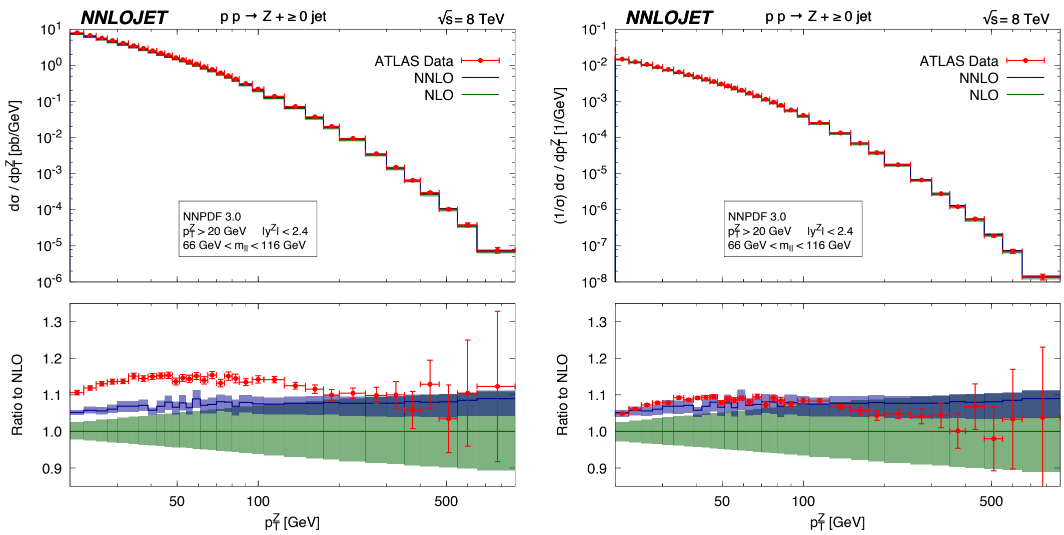


Figure 14: Left(right) plot shows the unnormalized(normalized) transverse momentum distribution of the inclusive Z boson production at LHC 8 TeV [248]. The green and blue bands denote the NLO and NNLO predictions with scale variations. The fiducial cuts on charged leptons are $p_{T,l} > 20\,\text{GeV}$ and $|\eta_l| < 2.4$.

A detailed phenomenlogical study and comparison of the NNLO calculation to the ATLAS and CMS 8 TeV measurements has been presented in [247, 248]. Very good agreement between the NNLO theory and data for the normalized distribution in ranges from 20 GeV $< p_T <$ 900 GeV, in all rapidity intervals, is observed. The conclusions are similar for the CMS data. However, there is some discrepancy in the comparison to the ATLAS unnormalized distributions, see Fig. 14, with the data tending to overshoot the theory over a wide p_T range. On the other hand, the NNLO prediction for the shape of the p_T distribution is in good agreement with the data down to a p_T value of 4 GeV, and is largely improved in comparison to the NLO predictions.

Impact on PDFs

The impact of the Z boson p_T data at LHC Run I has been studied very recently within a global analysis framework [233, 251]. In these studies an additional uncorrelated error of $\sim 1\%$ has been added to all p_T

bins to account for the theoretical uncertainty due for example to the MC integration error in the NNLO calculations. Without including these errors, it was found that NNLO predictions can not describe the data well, especially in the case of the normalized distributions. Some tension is also found between the ATLAS 7 TeV normalized p_T distribution [237] and the 8 TeV p_T distribution from both ATLAS and CMS [234, 235]. The ATLAS 7 TeV data also pulls the PDFs in a very different direction with respect to the HERA inclusive DIS data [233]. In [251] it is concluded that the inclusion of the ATLAS 7 TeV normalized data in the global analyses does not appear to be justified.

In Fig. 15 the impact of the ATLAS and CMS 8 TeV data on the NNPDF3.1 global analyses [251] is shown, by comparing the changes of the PDFs in the analysis with and without the Z p_T data sets included. The uncertainty in the gluon PDF is seen to be slightly reduced in the x region of $10^{-2} \sim 10^{-1}$. In the same region the gluon PDF receives constraints from the precision measurements on top-quark pair production and HERA inclusive DIS, both of which are present in the same analysis. It is also found that the 8 TeV data lead to a moderate reduction in the PDF uncertainty on the total strangeness.

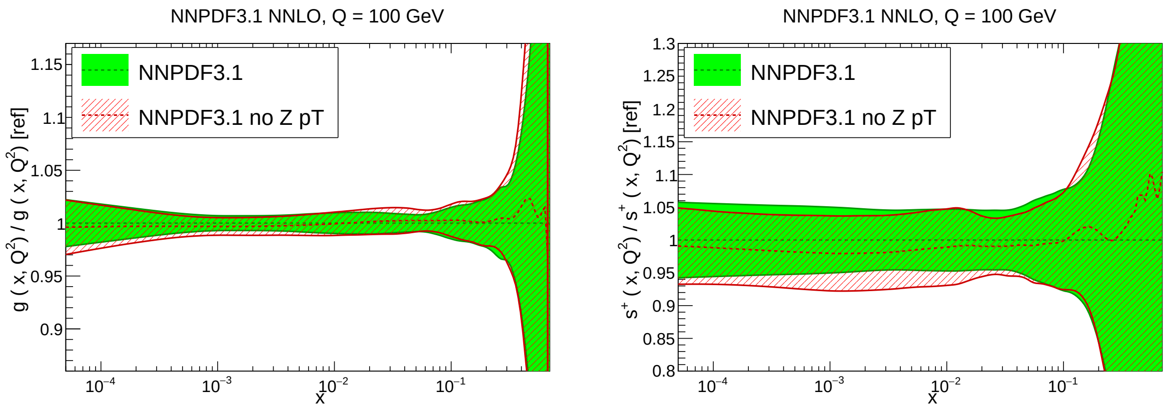


Figure 15: Impact of the Z boson transverse momentum measurements from ATLAS and CMS 8 TeV on the gluon PDF and the total strangeness in the NNPDF3.1 global analyses [251].

3.6. Direct photon production

In this section we discuss the PDF constraints that can be derived from the isolated photon production process.

PDF sensitivity

The LO parton–level processes for ‘direct’ photon production, where the photon is produced by point–like emission from a quark, are given by

$$\text{Compton :} \qquad qg \to q\gamma \, , \tag{59}$$

$$\text{Annihilation :} \qquad q\bar{q} \to g\gamma \, . \tag{60}$$

The Compton process gives the dominant contribution, in particular at the LHC. For LO kinematics the momentum fraction carried by the incoming gluon is directly proportional to the transverse energy $E_\perp^\gamma$ of the produced photon, and thus for higher $E_\perp^\gamma$ this process provides a direct probe of the gluon PDF at high x. Moreover, this represents the highest rate electroweak process at the LHC, while the produced photon

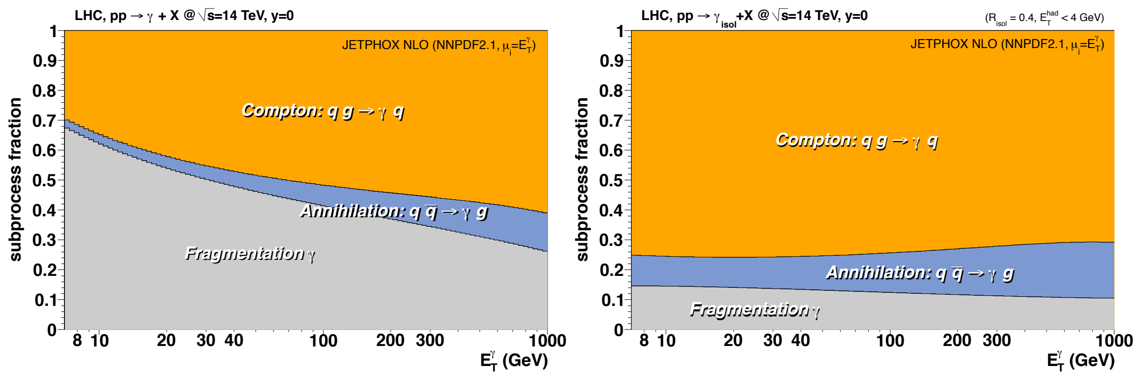


Figure 16: Relative contributions from Compton (qg), annihilation ($q\overline{q}$) and fragmentation to prompt photon production at central rapidities at the 14 TeV LHC, before (left) and after (right) the application of isolation cuts. Figures taken from [252].

directly reflects the production kinematics, without for example requiring any additional hadronization corrections, as in the case of jet production. This can therefore provide a valuable tool with which to constrain the gluon.

However, this process is not without its complications. In particular, the 'direct' process described above is not the only way in which high $E_\perp$ photons can be produced in hadronic collisions. We must also include the 'fragmentation' contribution whereby a standard $2 \to 2$ QCD scatter involving a final–state quark (or anti–quark) produces a photon through a collinear $q \to q\gamma$ emission. While the parton–level process carries an extra power of α_s compared to direct production, the collinearly enhanced photon emission is effectively of order α/α_s, and thus this enters at the same order. Technically speaking, this fragmentation emerges from the higher order corrections to the direct process. These correspond to multiple collinear splittings of a high $p_\perp$ parton which end up with a photon, and that can be absorbed into universal 'fragmentation functions'. These cannot be calculated perturbatively, but rather must be fit to data, for example in e^+e^- annihilation to hadrons. This introduces a potentially significant additional source of uncertainty.

In fact, the situation is greatly improved by noting that physically this fragmentation process corresponds to same multiple emission process that generates final–state jets, and indeed such fragmentation photons are typically accompanied by significant additional hadronic activity in the vicinity. This is to be contrasted with direct emission, where at LO the produced photon and outgoing quark are produced completely back–to–back. The direct mechanism may therefore be greatly enhanced by introducing 'isolation' criteria whereby the total sum of the transverse energy of the hadrons present in some cone R centred on the photon is less than a given value. These also reduce the additional 'non–prompt' background due to the electromagnetic decay of hadrons. The impact of such a cut is shown in Fig. 16, where it seen that the contribution from the less well known fragmentation contribution is small. This also demonstrates the dominance of the direct Compton production process.

In fact, isolated photon production represented one of the first PDF constraints considered, and was used in such early fits as [63, 64, 65]. However, the difficulties in describing the fixed target E706 [253, 254] data raised questions about the reliability of this process for PDF fits and potential sensitivities to non–perturbative effects. Combined with the increasing availability of high precision jet data from the Tevatron, which also constrain the high x gluon, this lead to the process falling out of favour in the PDF fitting community. The last PDF set to include any such data is the MRST99 [255] fit.

However, the subsequent studies of [256, 252] (see also [257] for a study of the related γ + jet process). have shown that by increasing the $\sqrt{s}$ from fixed target to collider energies and, as discussed above, imposing a suitable isolation condition on the produced photon, the process may be brought under reasonable theoretical control. Moreover, a comparison of the NLO pQCD predictions to the ATLAS measurement [258, 259] discussed below shows an adequate description of the data, albeit with fairly large $\sim 10-15\%$ scale variation uncertainties. We may expect this situation to improve further with the recent NNLO calculation discussed in Section 3.6. Thus isolated photon production may well provide a useful tool for LHC PDF constraints in the future.

Experimental data

The available collider data on isolated photon production is summarised below:

- The most recent data at 1.96 TeV from CDF [260] and D0 [261] extends out to $E_\perp^\gamma < 0.5$ TeV and 0.3 TeV, respectively, while the photon pseudorapidity is restricted to have $|\eta^\gamma| \lesssim 1$. In the CDF case this corresponds to the full Run II 9.5 fb^{-1} data set, and so represents the final legacy measurement.
- The ATLAS 7 TeV measurement [262], out to $E_\perp^\gamma < 1$ TeV and at 8 [258] and 13 [259] TeV, extending to $E_\perp^\gamma < 1.5$ TeV. These correspond to the full available integrated luminosities of 4.6 fb^{-1} and 20.2 fb^{-1} at 7 and 8 TeV, respectively, while the 13 TeV measurement uses a 3.2 fb^{-1} data set.
- The most precise CMS data at 7 TeV [263], corresponding to 36 pb^{-1} of integrated luminosity and extending to $E_\perp^\gamma < 0.4$ TeV.
- Data from a smaller sample at 2.76 TeV have also been taken by ATLAS [264] and CMS [265]. In all cases the photon pseudorapidity is restricted to have $|\eta^\gamma| \lesssim 2.4$.

Theoretical calculations and tools

For the past 15 years, the theoretical state–of–the–art was provided by the `JETPHOX` [266] MC generator, which implements both the direct and the fragmentation contributions consistently at NLO. The NLO EW corrections have also been calculated in [267]. However, recently the first NNLO calculation of direct photon production has been reported [268]. The NNLO prediction for the ATLAS 8 TeV data [258] is compared to the NLO and found to lie consistently within the NLO scale uncertainty band, with the central value being $\sim 5\%$ higher. Moreover, the NNLO scale uncertainty is found to be greatly reduced, giving a $\sim \pm 2-3\%$ uncertainty.

While the default description of the data is found to be quite poor, including the LL EW Sudakov corrections of [246], and evaluating the coupling α at the scale M_Z, as recommended in [246], the description is improved. In particular, the EW corrections are found to reduce the cross section by as much as 10%, that is significantly outside the QCD scale variation band, at the highest $E_\perp$, improving the shape description. The results of the `PeTeR` [269], which combines the NLO calculation with N^3LL threshold resummation in addition to these EW corrections, is found to lie close to the NNLO + EW prediction, but with a larger uncertainty band, indicating that the data may not be too sensitive to such additional resummation effects. Therefore, while the NNLO calculation is a very encouraging step towards including isolated photon data in high precision PDF fits, there are clearly some further theoretical issues to be investigated, relating to the impact of EW corrections and, as discussed in [268] the choice of photon isolation, which can also affect the NNLO comparison.

Impact on PDFs

Currently no up to date studies of the impact of isolated photon data on the PDFs have been performed, in particular taking into account the new NNLO calculation and the high precision LHC data. However,

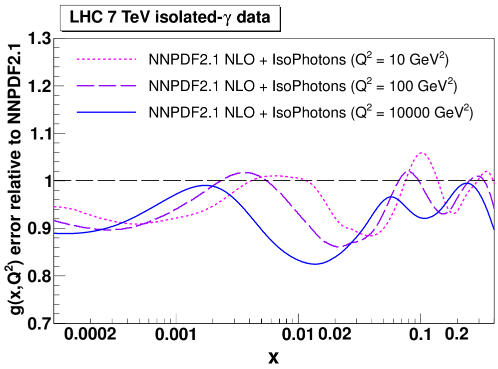


Figure 17: Relative reduction in the NNPDF2.1 NLO gluon PDF uncertainty when including a 36 pb^{-1} LHC 7 TeV isolated photon data set via reweighting. Taken from [252].

in [252] (see also [256]) the impact of a range of data, including the earlier 36 pb^{-1} ATLAS and CMS measurements at 7 TeV, on the PDFs has been studied in detail through a reweighting of the NNPDF2.1 set. In Fig. 17 the impact of the this LHC data on the gluon PDF is shown. A significant reduction in the uncertainty, of up to 20%, is found in the intermediate x region. Interestingly, this overlaps with the kinematically relevant region for Higgs boson production via gluon fusion at the LHC, and indeed a $\sim$ 20% reduction in the Higgs production cross section is found. Given these results correspond to a reasonably limited LHC data set, it will be interesting to see the impact of the latest data, as well as the NNLO corrections.

3.7. Top quark production

In this section we discuss the PDF information that can be obtained from top quark pair production measurements, and at the end we also review the constraints can could potentially be obtained from single top production.

PDF sensitivity

The production of top quark pairs at hadron colliders is driven by the gluon-gluon luminosity. Therefore, provided that other sources of theoretical uncertainties such as missing higher orders and the values of the top mass m_t can be kept under control, including top quark production data into the global PDF fit has the potential to constrain the gluon in the large-x region, which is affected by large uncertainties.

To illustrate the kinematical sensitivity of top quark pair production to the gluon, in Fig. 18 we show the correlation coefficient $\rho\,[g(x,Q), d\sigma]$ between the gluon PDF at $Q = 100$ GeV and the theory predictions for the differential distributions in $y_{t\bar{t}}$ and $m_{t\bar{t}}$ at $\sqrt{s} = 8$ TeV, as a function of x. Each curve corresponds to specific measurement bin. The higher the absolute value of the correlation coefficient, the bigger the sensitivity to the gluon for those specific values of x. We observe that this sensitivity is high for values of x up to $x \simeq 0.6 - 0.7$, beyond the reach of other processes sensitive to the gluon such as inclusive jet production. Moreover, the availability of differential distributions significantly extends the kinematical coverage beyond that provided by the total inclusive cross sections.

Experimental data

The available data on top quark pair production are summarised below:

- Earlier measurements, presented at the total cross section level, have been performed first at the Tevatron [270] and then by ATLAS and CMS [271, 272, 273, 274, 275, 276].

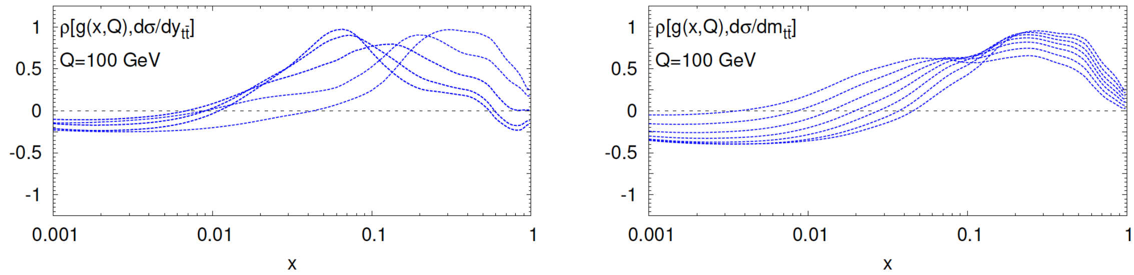


Figure 18: The correlation coefficient between the gluon PDF at $Q = 100$ GeV and the theory predictions for the differential distributions in $y_{t\bar{t}}$ (left) and $m_{t\bar{t}}$ (right plot) at $\sqrt{s} = 8$ TeV, as a function of x. Each curve corresponds to a specific measurement bin. The higher the absolute value of the correlation coefficient, the bigger the sensitivity to the gluon in those specific values of x.

- Single-inclusive differential distributions of top quark pair production have been presented by ATLAS [277] and CMS [278]. These include measurements both at the level of top-level observables extrapolated to the full phase space (p_T^t, $y_{t\bar{t}}$, $m_{t\bar{t}}$), as well as at the level of observables contributed in terms of directly observable quantities (charged lepton p_T and rapidity, b-tagged jet kinematics etc).

- Double differential distributions for top quark pair production may also be performed, as illustrated by the recent CMS measurement [279] of normalized double differential distributions, *i.e.*, as a function of $p_{t\bar{t}}$ and $m_{t\bar{t}}$.

The differential measurements are often presented normalized to the total cross-section, in order to benefit from a number of cancellations between experimental systematic uncertainties, but absolute measurements are also available.

Theoretical calculations and tools

The NNLO QCD calculation of the total $t\bar{t}$ production cross section has been available since 2013 [280, 281, 282], including the resummation of logarithmically enhanced threshold corrections up to NNLL. More recently, the full NNLO corrections to the single inclusive distributions in top quark pair production have been computed [283, 284]. Differential NNLO results are available for the rapidity of the top quark and the top-pair system, y_t and $y_{t\bar{t}}$, the transverse momentum of the top quark, p_T^t, and the invariant mass of the top-pair pair $m_{t\bar{t}}$, though not for other variables such as $p_T^{t\bar{t}}$ since these vanish at leading order. When differential distributions probe the TeV regions, electroweak corrections (including photon-initiated processes) also become relevant and need to be included in the theoretical calculations. In [285] (see also [286]), the NNLO QCD calculation where combined with the state–of–the–art NLO EW corrections, in the latter case including not only the $O\left(\alpha_s^2\alpha\right)$ but also the $O\left(\alpha_s\alpha^2\right)$ and $O\left(\alpha^3\right)$ contributions. This study showed that an accurate description of the tails of the kinematical distributions, such as the high-p_T^t and high-$m_{t\bar{t}}$ regions, must include NLO EW corrections.

An important limitation of the calculations discussed above is that they are restricted to stable top quarks. On the other hand, when experimental measurements are presented at the top quark level, they are extrapolated from the fiducial cross-sections using some theoretical model, thus possibly biasing the result by an amount which is difficult to quantify. Ideally, one would like a fully differential calculation with NNLO corrections included both for production and decay, in order to directly compare with experimentally observable quantities. An important milestone in this respect was the recent calculation of top-quark pair-production and decay [287] which allows providing predictions for observables constructed from top-quark leptonic and b-tagged jet final states, based on an approximation to the exact NNLO corrections to production and exact NNLO corrections to the decay.

Concerning the tools for the inclusion of top quark differential data into PDF fits, there exist two basic approaches. The first one is based on computing `APPLgrids` for the NLO calculation using either `MCFM` or `Sherpa` (see also Sect. 3.10), and then supplementing these with the NNLO/NLO bin-by-bin K-factors from [283, 284]. An improved strategy has been made feasible by the recent availability of `FastNLO` tables [288] that allow the efficient calculation of NNLO top quark pair distributions for arbitrary PDF sets and input $\alpha_s(m_Z)$ values. The latter option provides a more precise evaluation of the PDF–dependent NNLO corrections, although as shown explicitly in [289] the dependence of these K-factors on the PDF set is very small.

Impact on PDFs

The availability of the NNLO calculation of the total cross-sections for top quark pair production has made it possible to include top quark data from the Tevatron and the LHC consistently into a NNLO PDF fit for the first time. By applying Bayesian reweighting to NNPDF2.3, it was shown in Ref. [290] that top quark data could reduce the PDF uncertainties in the large-x gluon by up to 20% for $x \simeq 0.2$ (see also previous related work in [291]). Several other global fits, such as ABMP16 and MMHT14, also include total $t\bar{t}$ cross-sections in their default fits. While these results provided an encouraging indication of the PDF constraining potential of $t\bar{t}$ production, the full exploitation of this potential clearly required the use of differential distributions.

The impact of the $\sqrt{s} = 8$ TeV top quark pair differential data from ATLAS and CMS on the NNPDF3.0 fit was quantified in [289]. Here, it was shown that the constraints on the large-x gluon were at this point competitive with those provided by inclusive jet production, despite the much reduced number of experimental data points. See also [292] for related work based on approximate NNLO calculations. An important result of the investigations of [289] was that the constraints from the normalized distributions were in general superior to those from their absolute counterparts, most likely because of the cancellation of systematic uncertainties in this case. In addition, top quark differential distributions at 8 TeV from the LHC have been included in the recent NNPDF3.1 global analysis, and other groups have also studied the impact of this data into their PDF fits in a preliminary form. A challenge in the study of [289] was the observed tension between some of the ATLAS and CMS distributions, such as $m_{t\bar{t}}$, which prevented their simultaneous inclusion in the global fit. While the underlying cause of these discrepancies is still under investigation, this limitation was bypassed by identifying pairs of distributions which could be fitted with good quality at the same time and that exhibited comparable constraining power.

In order to illustrate the impact of the top quark data on the large-x gluon, in Fig. 19 we show how the PDF uncertainties of the NNPDF2.3 gluon are reduced once the Tevatron and the LHC 7 and 8 TeV inclusive top-quark pair cross-section data are included by means of Bayesian reweighting [290]. As can be seen, at the level of total cross section data the impact is still moderate, with PDF error reduction being at most $\simeq 20\%$. On the other hand, in Fig. 19 we also show the impact on the gg luminosity of the normalized 8 TeV ATLAS and CMS $t\bar{t}$ differential distributions, compared to a baseline fit based on the NNPDF3.0 settings but without the jet data (since the NNLO corrections to jet production were not available at the time). One clearly sees here how the impact is much more significant, highlighting the increase in constraining power of the differential distributions are compared to the total cross section data, specially in the large-x region, where PDF uncertainties can be reduced by more than a factor of 2.

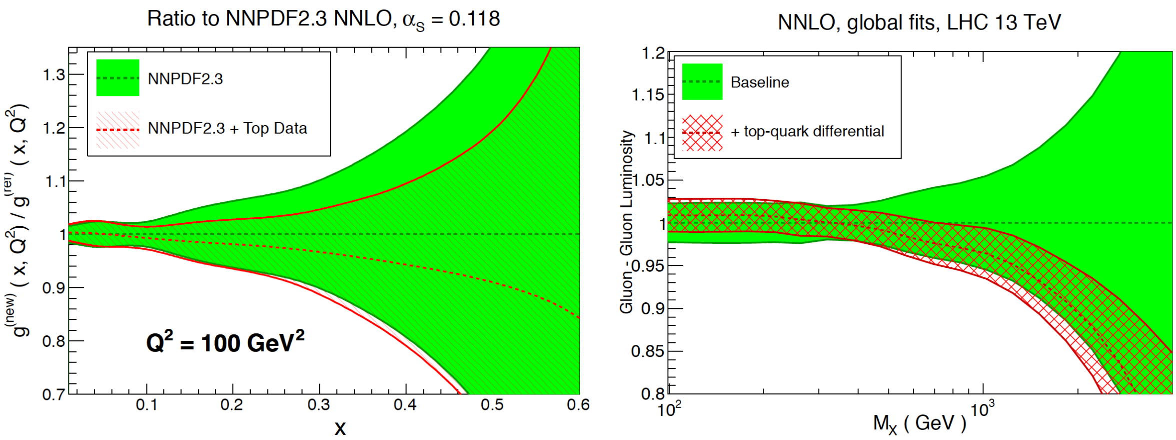


Figure 19: Left plot: the impact of the LHC 7 and 8 TeV inclusive top-quark pair cross-section data on the large-x gluon of NNPDF2.3 [290]. Right plot: the impact of the LHC 8 TeV differential distributions in top-quark pair production on the gg luminosity [289], compared with a baseline fit based on the NNPDF3.0 global analysis without the jet data.

Single top production

In addition to top quark pair production, single top production provides also in principle provides a useful PDF–sensitive information. Such a process can proceed via he scattering of a b with a light quark, see Fig. 20 (Left) for a typical diagram, and will therefore provide information about the b–quark PDF. In addition, due to the presence of the b in the initial state it provides unique testing ground for the different heavy quark flavour schemes used in the calculation described in Section 2.5. That is, one can use a $n_f = 4$ scheme, a $n_f = 5$ scheme, or a matched scheme interpolating between the two, see the discussion in Refs. [293, 294]. State or the art calculations are based on NNLO QCD theory both for the total cross sections and for differential distributions [295**?**], and LHC measurements at 8 TeV and 13 TeV of total cross sections (including ratios of top to anti–top production) as well as single inclusive distributions are already available [296, 297].

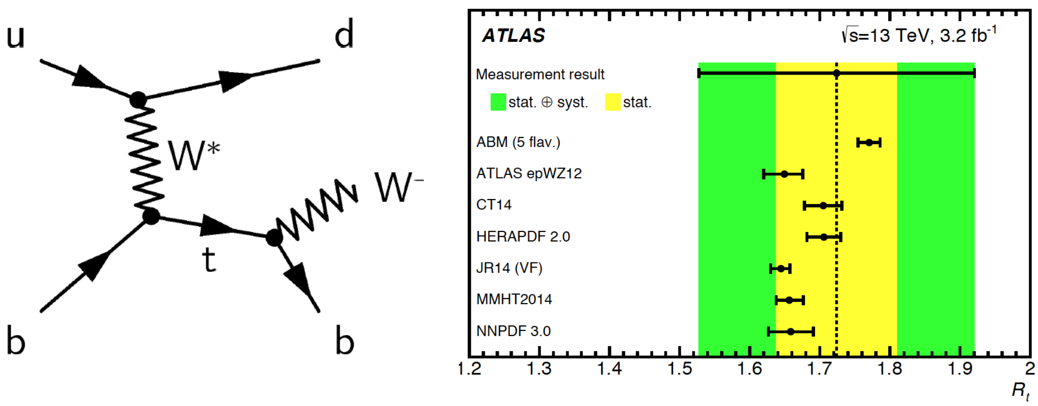


Figure 20: Left plot: one of the Feynman diagrams for single-top production at leading order, illustrating its sensitivity to the b-quark PDF. Right plot: comparison of the theoretical predictions for the ratio $R_t = \sigma_t/\sigma_{\bar{t}}$ from different PDF sets and the corresponding ATLAS measurements at $\sqrt{s} = 13$ TeV from [297].

Moreover, since the production of top and anti–top quarks is generated by different initial state partons, cross section ratios such as $R_t \equiv \sigma_t/\sigma_{\bar{t}}$ can provide information on the quark flavour separation. To illustrate this point, we show in Fig. 20 (Right) a comparison of the theoretical predictions for the R_t ratio from different PDF sets and the corresponding ATLAS measurements at $\sqrt{s} = 13$ TeV from [297]. While experimental uncertainties are still large, due to the limited dataset, we can see that the measurement may eventually become sensitive to differences between PDF sets. Moreover, similar comparisons could also be performed for differential distributions, either at the level of top kinematic variables or at the level of observable quantities constructed from leptons and b-jets.

3.8. Charm production in pp collisions

The production of heavy quark mesons at hadron colliders is driven by the gluon-gluon luminosity, and therefore it provides a sensitive probe to the gluon PDF at medium and small-x. In particular, charmed meson production in the forward region covered by LHCb gives information on the gluon at values of x as small as $x \simeq 10^{-6}$, well below the kinematic reach of the HERA structure function data, and thus in a region where PDF uncertainties are very large due to the very limited amount of experimental information available.

The precision determination of the small-x gluon PDF impacts LHC phenomenology indirectly, by means of its impact on the Monte Carlo modeling of soft and semi-hard dynamics, which depend on the gluon at very small-x. Understanding the gluon in this region is also crucial for ultra-high energy neutrino astronomy and cosmic ray production. In the former case, the small-x gluon is relevant both for the calculation of signal event rates, via the interaction cross-section between UHE neutrinos and target nucleons (ice or water), as well as for the calculation of the rates for the dominant background process, the production of charm quarks in cosmic ray collisions in the atmosphere which then decay into so–called 'prompt' neutrinos and which dominate the atmospheric neutrino flux at high energies.

The LHCb experiment has presented measurements of charm meson production at $\sqrt{s} = 5, 7$ and 13 TeV. These are double differential cross-sections as a function of the transverse momentum p_T^D and rapidity y^D of the produced D mesons, for the following species: $D^{\pm}$, D^0, D^* and D_s, together with the corresponding complex conjugates. In addition to these double-differential distributions for the three values of $\sqrt{s}$, also double ratios between center of mass energies, have been provided, in particular between 13 TeV and 7 TeV and between 13 TeV and 5 TeV.

In Fig. 21 we show a comparison between the small-x gluon at $Q^2 = 4$ GeV^2 for NNPDF3.0 with the corresponding result after different combinations of the charm production data at LHCb have been included in the fit. We show the central value and one-sigma PDF uncertainty bands for the $N^7 + N^{13/5}$ and the $N^5 + N^7 + N^{13}$ combinations, as well as the central value for the $N^5 + N^{13/7}$ case. We can see that the charm data reduces the uncertainty on the small-x PDF uncertainties by up to almost an order of magnitude.

3.9. Central Exclusive Production

Centeral Exclusive Production. The Central Exclusive Production (CEP) process occurs when an object X and nothing else is produced in a hadronic collision, while the hadrons themselves remain intact after the collision. The photoproduction of heavy vector mesons, see Fig. 22, is one example of such a process which has possible implications for PDF determination. Thus for pp collisions, while one proton elastically emits a photon, the other interacts via t–channel two gluon exchange. This may therefore access the gluon PDF at a comparatively low scale $Q^2 \sim M_V^2$ and $x \sim M_V/\sqrt{s}$, where it is so far quite poorly determined.

LHCb have measured the exclusive production of J/ψ and $\psi(2S)$ mesons at 7 TeV [299] and Υ production at 7 and 8 TeV [299]. Preliminary LHCb data on J/ψ and $\psi(2S)$ production at 13 TeV has also been reported in [300]. This uses the newly installed `HeRSCheL` shower counters, which greatly extend the

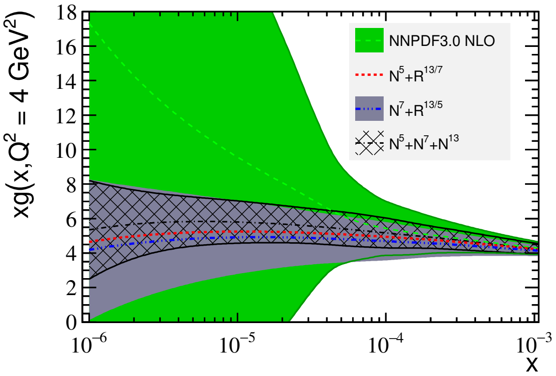


Figure 21: Left plot: comparison between the small-x gluon at $Q^2 = 4$ GeV^2 in NNPDF3.0 with the corresponding result after different combinations of the charm production data at LHCb have been included in the fit. We show the central value and one-sigma PDF uncertainty bands for the $N^7 + N^{13/5}$ and the $N^5 + N^7 + N^{13}$ combinations, as well as the central value for the $N^5 + N^{13/7}$ case

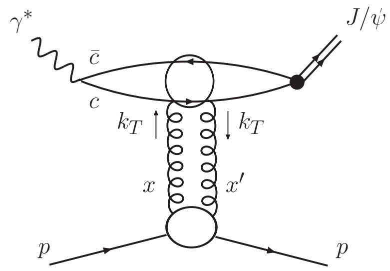


Figure 22: Photoproduction of heavy vector meson J/ψ [298].

rapidity coverage for vetoing additional particle production, reducing significantly the non–exclusive background. In addition, ALICE have measured exclusive J/ψ production in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV [301]. Due to the large Z^2 enhancement for photon emission from the Pb ion, this is dominated by the process where the lead ion emits a photon.

While clearly the photoproduction process shown in Fig. 22 proceeds through an initial–state gluon interaction, this does not correspond to a standard inclusive process where PDFs are generally introduced. However, as discussed in [298] under certain assumptions this process can be related to the standard gluon PDF and may therefore serve as a probe of it at low x and Q^2. This analysis has subsequently been performed at NLO [302] (see also [303]) for the cases of J/ψ and Υ production. However, here it is found that the NLO correction is significantly larger than, and of the opposite sign to, the LO contribution, indicating a lack of perturbative stability and casting some doubt on its viability as a PDF probe. Nonetheless, work in the direction of at least partially solving this issue has been reported most recently in [304], where the stability is shown to be improved through judicious choice of factorization and renormlization scale, and by imposing a cut on the NLO contribution to avoid double counting. It is also worth noting that the perturbative stability is naturally improved somewhat by considering the production of the higher mass Υ.

3.10. Fast interfaces

Given the highly CPU time–consuming nature of global PDF fits, the direct evaluation of the lengthy (N)NLO hadronic cross sections during the PDF fit is not feasible. For this reason, until around 2008 PDF fits included hadronic data using LO hadronic cross-sections supplemented by bin–by–bin K-factors, defined as

$$K^i_{\rm NLO} \equiv \frac{\sigma_i^{\rm NLO}}{\sigma_i^{\rm LO}} \,, \tag{61}$$

using the same PDF set in the numerator and in the denominator. To ensure consistency of the procedure, these K factors were computed iteratively until convergence was achieved. However, this approximation is known to have several deficiencies, the most important one being the reduced sensitivity to those partonic initial states that only enter the cross-section at NLO.

In order to improve over this unsatisfactory situation, the method of *fast interfaces* was proposed. In these methods, the most CPU time consuming part of an NLO calculation, namely the evaluation of the partonic matrix elements over a large number of events, is precomputed *a priori* using a complete interpolation basis for the input PDFs. This way, the hadronic cross-sections can be reconstructed *a posteriori* by means of a very efficient matrix multiplication of the PDFs evaluation in a grid of (x, Q) points and the precomputed partonic matrix elements at the same grid points. These tools have become very popular and are therefore used in the majority of former PDF fits.

Following a common philosophy, two main tools have been developed, `APPLgrid` [305] and `FastNLO` [306]. More recently, the `aMCfast` interface [307] to `MadGraph5_aMC@NLO` [308] has been developed. Given the automated character of this code, `aMCfast` allows producing fast interpolation of arbitrary NLO processes, defined by the user at run time. It is also possible to produce fast grids for NLO calculations matched to parton shower Monte Carlos, which opens the way to include in the PDF fit hadron-level cross-sections such as W boson in association with charmed mesons or forward D meson production at LHCb.

We now describe the basic strategy of fast interpolation methods. We use the notation of the `APPLgrid` paper for concreteness, but the general method is very similar in `FastNLO` and `aMCfast`. These are based on representing the PDFs in (x, Q^2) by means of a suitable interpolation basis, computing a physical cross-section for a basis PDF set, and then reconstructing the same observable *a posteriori* using an arbitrary PDF set. Therefore, to begin, with, one expands an arbitrary PDF $f(x, Q^2)$ in terms of a suitable basis of interpolating polynomials

$$f(x,Q^2) = \sum_{i=0}^{n}\sum_{j=0}^{m} f_{k+i,\kappa+j} I_i^{(n)}\left(\frac{y(x)}{\delta y} - k\right) I_j^{(m)}\left(\frac{\tau(Q^2)}{\delta\tau} - \kappa\right) \,, \tag{62}$$

where n and m are the interpolation orders in x and Q^2 respectively, $y(x) = \ln 1/x + a(1-x)$ and $\tau(Q^2) = \ln\left(\ln Q^2/\Lambda^2\right)$, and $I_i^{(n)}, I_j^{(m)}$ are interpolating functions, for instance Lagrange interpolating polynomials, though Eq. (62) holds generically for other choices. k and κ are defined as

$$k(x) = {\rm int}\left(\frac{y(x)}{\delta y} - \frac{n-1}{2}\right), \; \kappa(Q^2) = {\rm int}\left(\frac{\tau(Q^2)}{\delta\tau} - \frac{m-1}{2}\right), \tag{63}$$

with ${\rm int}(u)$ be the largest integer that is smaller than u.

After the representation Eq. (62) has been constructed, we need to evaluate cross-sections using the interpolation basis. Let us consider first for simplicity a hypothetical DIS structure function F that receives contributions from a single flavour. The NLO cross-section is typically computed by means of Monte Carlo program that generates a large number N of events, each one with weight ω_m and with associated values x_m and Q^2_m. If p_m is the order of α_s for this specific event, the total cross-section can be written as

$$F = \sum_{t=1}^{N} \omega_t \left(\frac{\alpha_s(Q_t^2)}{2\pi} \right)^{p_t} f(x_t, Q_t^2) \,. \tag{64}$$

The fast interpolation can be constructed by, instead of computing F as in Eq. (64), introducing a weight grid $W^{(p)}_{i_y,i_\tau}$, and for each event only a fraction of the grid nodes is updated according to the expression

$$W^{(p_t)}_{k+i,\kappa+j} \rightarrow W^{(p_t)}_{k+i,\kappa+j} + \omega_t I_i^{(n)} \left(\frac{y(x_t)}{\delta y} - k \right) I_j^{(m)} \left(\frac{\tau(Q_t^2)}{\delta \tau} - \kappa \right) . \tag{65}$$

Conceptually, the weight grid $W^{(p)}_{i_y,i_\tau}$ is the equivalent of computing the structure function F but for a given combination of interpolating polynomials as opposed to the original PDF.

The important factor here is that the most CPU time intensive computation, the calculation of the MC weights ω_m, needs only to be done once to fill the grid $W^{(p)}_{i_y,i_\tau}$, and the PDF can be decided a posteriori at virtually no extra computational cost. Indeed, it can be shown that the structure function can be reconstructed a posteriori using the weight grid using

$$F = \sum_p \sum_{i_y} \sum_{i_\tau} W^{(p)}_{i_y,i_\tau} \left(\frac{\alpha_s(Q^2_{i_\tau})}{2\pi} \right)^p f(x_{i_y}, Q^2_{i_\tau}) \,. \tag{66}$$

In other words, the only information which is needed is the value of the PDFs and the strong coupling at the grid nodes i_y, i_τ. The method can be straightforwardly generalized to hadron-hadron collisions and to a generic composition of the initial parton state, taking into account that now the formula includes two PDFs. In proton-proton collisions, the analog of Eq. (66) is given by

$$\sigma = \sum_p \sum_{l=0}^{n_{\rm sub}} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W^{(p)(l)}_{i_{y_1},i_{y_2},i_\tau} \left(\frac{\alpha_s(Q^2_{i_\tau})}{2\pi} \right)^p \mathcal{L}^{(l)} \left(x_{1,i_{y_1}}, x_{2,i_{y_2}}, Q^2_{i_\tau} \right) , \tag{67}$$

where we have indicated that there are n_l contributing partonic subprocesses, each with the corresponding luminosity $\mathcal{L}^{(l)}$, which depend on the cross-section upon consideration.

In order to illustrate the high precision that these fast interfaces can achieve, we show two representative examples in Fig. 23. First of all, we show the ratio between the `NLOjet++` calculation of inclusive jet production at 7 TeV in the rapidity interval $2 \leq y \leq 3$ and the corresponding *a posteriori* calculation based on `APPLgrid`, for different values of the factorization and renormalization scales. One sees that the differences between the original and the interpolated calculation are at the few permille level. Then we show the transverse momentum distribution of photons in the $pp \rightarrow \gamma + \mathrm{jet}$ process at 7 TeV, comparing the original `MadGraph5_aMC@NLO` calculation with the *a posteriori* result based on `aMCfast` and `APPLgrid`. The lower insets show the ratio between the two calculations for different choices of μ_R and μ_F. Here, we also find excellent agreement between the original and interpolated calculations, now at the sub-permille level. In all these methods, the interpolation accuracy can be arbitrarily increased by using denser grids in x and Q^2.

While these fast interface represent a very significant improvement in terms of CPU efficiency as compared to the original NLO calculations, one limitation of this approach can be seen from the master formula for proton–proton collisions Eq. (64): each time the PDF set is varied, one needs to recompute its values

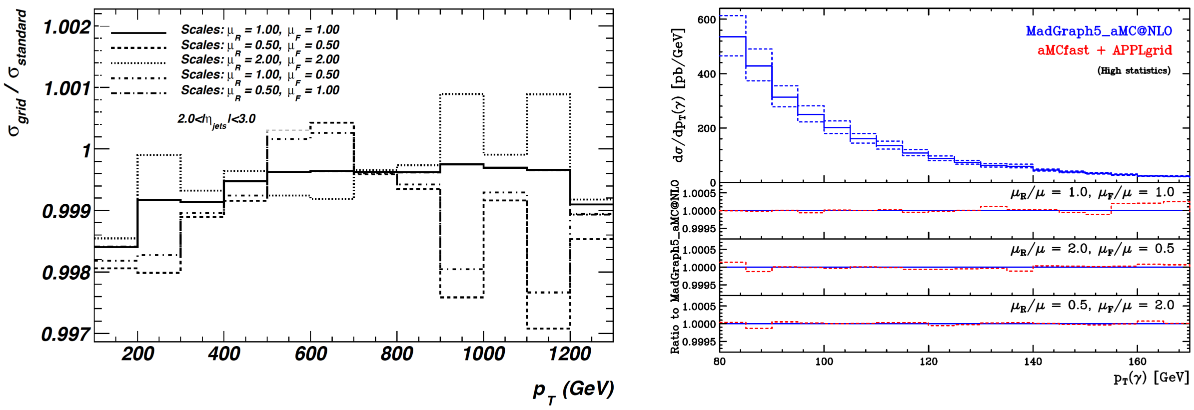


Figure 23: Left plot: the ratio between the NLOjet++ calculation of inclusive jet production at 7 TeV in the rapidity interval $2 \leq y \leq 3$ and the corresponding *a posteriori* calculation based on APPLgrid, for different values of the factorization and renormalization scales. Right plot: the transverse momentum distribution of photons in the $pp \to \gamma + \mathrm{jet}$ process at 7 TeV, comparing the original MadGraph5_aMC@NLO calculation with the *a posteriori* result based on aMCfast and APPLgrid. The lower insets show the ratio between the two calculations for different choices of μ_R and μ_F.

in the (x, Q^2) nodes. In a PDF fit, this means that each time the input parametrization is modified during the iterative minimization, the DGLAP evolution equations need to be solved again, before the PDFs can be convoluted with the interpolated coefficient functions to obtain the hadronic cross-section. To improve upon this shortcoming, recently the APFELgrid tool has been developed [309]. The goal of APFELgrid is to combine the interpolated partonic cross-sections provided by APPLgrid with the DGLAP evolution factors provided by APFEL, in a way that hadronic cross-sections can be reconstructed from a matrix multiplication requiring only as input the values of the PDFs at the x grid nodes at the input evolution scale Q_0. This combination then lead to a very significant improvement in computation speed as compared to Eq. (64) without any loss of numerical accuracy, and leads to much faster PDF fits. Mathematically, the APFELgrid method allows expression an arbitrary hadronic cross-section as follows

$$\sigma_{pp\to X} = \sum_{k,l} \sum_{\delta,\gamma} \widetilde{W}_{kl,\delta\gamma} f_k(x_\delta Q_0^2) f_l(x_\gamma, Q_0)\,, \tag{68}$$

in terms of the PDFs at the parametrization scale Q_0, where k, l run over all active parton flavours and δ, γ run over the nodes of the x interpolating grid.

To gauge the improvements in computational efficiency that can archived by this method, in Fig. 24 we show a comparison of the timings per data point between the original APPLgrid computation of hadronic cross-sections, Eq. (66), with the same calculation based on the APFELgrid combination, Eq. (68), for a variety of LHC datasets. As we see from this figure, the improvement in computational speed is between a factor 100 and a factor 1000 depending on the specific dataset. This means that PDF fits based on APFELgrid will be much faster, by up to two orders of magnitude, which is a very attractive property of this method.

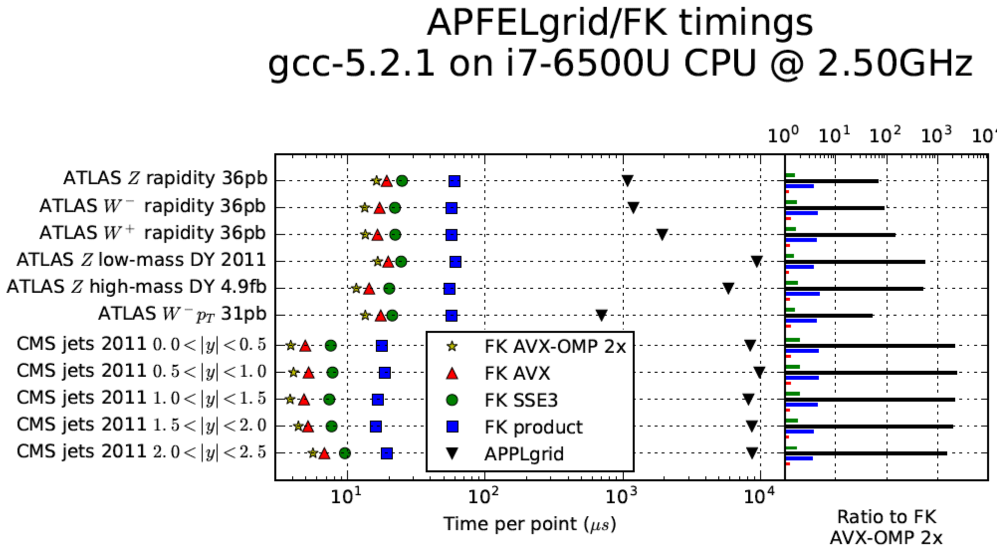


Figure 24: Comparison of the timings per data point between the original APPLgrid computation of hadronic cross-sections, Eq. (66), with the same calculation based on the APFELgrid combination, Eq. (68), for a variety of LHC datasets. We find that the improvement in computational speed is between a factor 100 and a factor 1000 depending on the specific dataset.

4. Fitting methodology

In this section we present the framework of the global QCD analysis of parton distribution functions. First of all we discuss how PDFs are parametrized at the reference scale, and review the theoretical constraints that should be imposed to this parametrization such as the momentum and valence sum rules and positivity. Then we discuss how to quantify the agreement of data and theory, and review various methods used in global analysis for minimization of the figure of merit χ^2. Next we review the various methods available to estimate and propagate PDF uncertainties, with emphasis on the three most important ones: the Hessian, the Monte Carlo, and the Lagrange multiplier methods. In this section we also discuss how to combine individual PDF sets within a single PDF set.

4.1. PDF parametrization

4.1.1. Functional form

To extract the PDFs some form of parameterisation in x must be assumed, which can then be fit to the available data. As described in Sect. 2.4, given the PDFs at some reference scale Q_0, DGLAP evolution determines the PDFs at any other scale μ. Thus the PDFs are typically parameterised at a low scale $Q_0^2 \sim 1-2\,\mathrm{GeV}^2$, which can then be evolved up to the scale relevant to e.g., LHC physics. These universally take the form

$$xf(x, Q_0^2) = A_f x^{a_f}(1-x)^{b_f} I_f(x)\ . \tag{69}$$

The $(1-x)^{b_f}$ term, with $b_f > 0$, ensures that the PDFs vanish in the elastic $x \to 1$ limit, as we would expect on basic physical grounds. Such a form is also expected from the quark counting rules [310]. Here, in this elastic limit all the momentum is carried by the struck parton and the remaining n_s quark become spectators. An analysis of the scaling behaviour for elastic scattering then predicts $b_f = 2n_s - 1$, that is $b_f = 3, 5$ and 7 for valence, sea and gluon distributions, respectively. The x^{a_f} form dominates at low x; in this region, the PDFs are related to the high energy parton–proton scattering amplitudes, which may be calculated using the tools of Regge theory. This predicts such a simple power–like form, with the precise value of the power a_f being related to the leading Regge trajectory that is exchanged; for non–singlet distributions (e.g. the valence quarks) this predicts $a_f \sim 0.5$ and for singlet distributions (e.g. the gluon and the sea) this predicts $a_f \sim 0$. The above discussion only correspond to quite general expectations, which do not for example account for the scale dependence of the PDFs. Thus while the high and low x form of (69) is taken, for modern fits the values of the powers themselves are more generally left free where there is sufficient data to constrain them.

The $I_f(x)$ is the interpolating function, which determines the behaviour of the PDFs away from the $x \to 0$ and 1 limits, where it tends to a constant value. This is assumed to be a smoothly varying function of x, for which a variety of choices have been made in the literature. The simplest ansatz, which has been very widely used, is to take a basic polynomial form in x (or $\sqrt{x}$), such as

$$I_f(x) = 1 + c_f\sqrt{x} + d_f x + \ldots\ . \tag{70}$$

Forms of this type are for example taken by CJ, HERAPDF, and in the previous MSTW08 set. A similar approach, but where the polynomial enters as the exponent of a power of x or a simple exponential function, are taken by ABMP and earlier CT sets, respectively.

Such a choice is appropriate for a relatively small number of parameters c, d. However, as the precision and amount of the data included in the fit increases it becomes essential to allow for an increasingly flexible parameterisation. As discussed in [311], simply adding more parameters to (70) can quickly run into the technical issues that large coefficients appear, with large cancellations between the terms. This leads to an unstable χ^2 minimisation and implausibly large variations in x in certain regions. This issue may be solved by instead expanding the interpolating function in terms of a basis of suitably chosen functions

$$I_f(x) = \sum_{i=1}^{n} \alpha_{f,i} P_i(y(x)) \; , \tag{71}$$

where $y(x)$ is some simple function of x. Two choices for the functions P_i are the Chebyshev and Bernstein polynomials, which are used in the MMHT14 and CT14 sets, respectively. These are chosen as each order of the polynomials is strongly peaked at different values of y, and hence x, significantly reducing the degree of correlation between the terms. In addition, as the order is increased these tend to probe smaller scale variations in x, so that the smoothness requirement for $I(x)$ naturally leads to smaller coefficients α at higher i. Thus, while formally equivalent to the simply polynomial expansion in Eq. (70), these are much more convenient for fitting as the number of free parameters n is increased.

An alternative approach is taken by the NNPDF group. Here, the interpolating function is modelled with a multi–layer feed forward neural network, see Sect. 5.3 for more details. In practice, this allows for a greatly increased number of free parameters, with the latest default fit having 37 per PDF, that is $\sim$ an order of magnitude higher than other sets. The form of (69) is still assumed, but these are pre–processing factors that speed up the minimisation procedure but which do not in principle have to be explicitly included. Nonetheless, the study of [312] has shown that the NNPDF fit does exhibit high and low x behaviour that is consistent with (69), providing further support for such an assumed form.

4.1.2. Sum rules

The valence uud structure of the proton, with zero strangeness, is expressed in the three number sum rules

$$\int_0^1 \mathrm{d}x \left[u(x,Q^2) - \overline{u}(x,Q^2) \right] = 2 \; , \tag{72}$$

$$\int_0^1 \mathrm{d}x \left[d(x,Q^2) - \overline{d}(x,Q^2) \right] = 1 \; , \tag{73}$$

$$\int_0^1 \mathrm{d}x \left[s(x,Q^2) - \overline{s}(x,Q^2) \right] = 0 \; , \tag{74}$$

thus for the valence distributions we must have $a_f > 0$ for the exponents in (69) or these integrals will diverge. In others words, we have the well known result that the xf valence distributions vanish as $x \to 0$. Although not shown explicitly, a similar constraint applies to the heavy quark PDFs as to the strange PDF. In the absence of any intrinsic heavy flavour, these are automatically satisfied.

The sum of PDFs must also obey the momentum sum rule

$$\int_0^1 \mathrm{d}x \, x \left(\sum_{n_f} (q(x,Q^2) + \overline{q}(x,Q^2)) + g(x,Q^2) + \cdots \right) = 1 \; , \tag{75}$$

which expresses the simple physical requirement that the total proton momentum must be equal to the sum of its constituents. We have suppressed contributions from any additional parton, for example the photon (or even electroweak bosons), which is included within a given a set. Thus for non–valence distributions the exponent a_f may be negative, but must be greater than -1 to avoid giving a divergent contribution to the momentum sum rule.

The above 4 sum rules provide additional constraints on the input PDFs, and are typically applied to fix certain parameters, for example the overall normalization A_f of a given set. Provided these sum rules are satisfied at the input scale, it follows straightforwardly from the form of the DGLAP evolution that they will be satisfied at any other scale μ; the $g \to q\overline{q}$ splitting can generate no net $q - \overline{q}$ component, and the DGLAP evolution reshuffles the momentum carried between the different partons, but of course generates no momentum violation.

4.1.3. Quark flavour assumptions

Assuming that there are n_f active quark flavors at the input parametrization scale Q_0, there will be in general $2n_f + 1$ PDFs to be parametrized and fitted from data. Assuming that the heavy quark PDFs are generated perturbatively, in addition to the gluon, in many cases the remaining 6 light quarks PDFs parametrized are not those in the *flavour basis*, namely

$$u, \quad \bar{u}, \quad d, \quad \bar{d}, \quad s, \quad \bar{s}\,, \tag{76}$$

but rather other convenient linear combinations, *e.g.* the valence $u_V = u - \bar{u}$ and $d_V = d - \bar{d}$ distributions are often used. To give one example, the MMHT14 analysis takes as fitting basis, in addition to the gluon,

$$u_V, \quad d_V, \quad \overline{d} - \overline{u}, \quad s + \overline{s}, \quad s - \overline{s}, \quad s + \overline{s} + 2(u + \overline{u} + d + \overline{d})\,. \tag{77}$$

As another example, the NNPDF3.0 fit parametrizes the PDFs at the input evolution scale in the so-called *evolution basis*, defined as the eigenvectors of the DGLAP evolution equations,

$$\begin{aligned} \Sigma &= u + \bar{d} + d + \bar{d} + s + \bar{s} \\ T_3 &= u + \bar{d} - d - \bar{d}\,, \\ T_8 &= u + \bar{d} + d + \bar{d} - 2s - 2\bar{s}\,, \\ V &= u - \bar{d} + d - \bar{d} + s - \bar{s} \\ V_3 &= u - \bar{d} - d + \bar{d}\,, \\ V_8 &= u - \bar{d} + d - \bar{d} - 2s + 2\bar{s}\,, \end{aligned} \tag{78}$$

in addition to the gluon PDF. However, as any particular basis can be trivially related to another by a linear transformation, the physics should not depend on this choice. On the other hand, different flavour assumptions do often lead to different results in regions with limited experimental constraints, such as the large-x region.

Historically, the strange quark has been less well determined than the u and d quark PDFs, and indeed in many earlier fits this was fixed according to

$$s = \overline{s} \propto \overline{u} + \overline{d}\,. \tag{79}$$

Such a choice is still taken in the CJ15 and HERAPDF fits, due to the more restricted data set. With the increase in available data, the total strangeness $s + \overline{s}$ is now freely parameterised in all global fits. While the sum rule Eq. (74) requires there to be no overall strangeness in the proton, at a given x value there is no requirement for the $s - \overline{s}$ distribution to vanish, and indeed non–perturbative approaches such as the "meson cloud model" [313] predict a non-zero strange asymmetry. However, the strange difference $s - \overline{s}$ is generally quite poorly determined and still broadly consistent with zero within current uncertainties. From the latest global fits, only MMHT14 and the NNPDF3 sets fit the strange difference, while for all other sets it is still assumed that $s = \overline{s}$. Note also that at NNLO, even if $s_V = s - \bar{s}$ is set to zero at the initial evolution asymmetry, a non-zero strangeness (as well as charm and bottom) asymmetry will be generated dynamically by the DGLAP evolution equations [314].

The above discussion assumes that the charm PDF is generated due to perturbative $g \to c\bar{c}$ splittings, in which case it is generated purely by DGLAP evolution and is determined in terms of the light quark and gluon PDFs. If the charm PDF is instead fitted, the input flavour assumptions need thus to be modified. In the case of NNPDF3.1, the evolution basis of Eq. (78) is supplemented with $c^+ = c + \bar{c}$, which is freely parametrized with a neural network, while it is assumed that $c^- = c - \bar{c} = 0$. This option is also adopted in other recent studies where the charm PDF is fitted, such as in the CT14 IC analysis. Note that in general different flavour assumptions concerning the parametrized charm PDF are conceivable, for instance Eq. (78) could be generalized by adding

$$T_{15} = u + \bar{d} + d + \bar{d} + s + \bar{s} - 3c - 3\bar{c}\,, \tag{80}$$

though this option would have the drawback that the connection with charm-sensitive observables is far less direct.

An important issue related to the PDF parametrization is that of the positivity. While, beyond LO, PDFs are scheme dependent quantities and thus in principle can become negative, physical observables such as cross-sections and structure functions should always be positive-definite. This constrain is incorporated in the (N)NLO global fits in different ways. For instance, in CT14 all PDFs are made by construction positive-definite, while MMHT14 allows the small-x gluon PDF to become negative. In the case of the NNPDF family of fits, no positivity constraints are imposed at the PDF level, but during the fit the strict positivity of a range of physical cross-sections is imposed by means of a Lagrange multiplier. Specifically, in the NNPDF3 sets the positivity of the following cross-sections is imposed at $Q^2 = 5$ GeV^2: F_2^u,F_2^d, F_2^s, F_L, $\sigma_{\rm DY}^{u\bar{u}}$, $\sigma_{\rm DY}^{d\bar{d}}$, and $\sigma_{\rm DY}^{s\bar{s}}$. Note that in general this positivity constraint applies to all conceivable cross-sections, including for instance those that involve hypothetical new particles, and is not restricted to the actual cross-sections that are accessible experimentally.

4.2. Data/theory agreement and minimization

4.2.1. Definition of χ^2

The quality of a global fit,is usually expressed in terms of the log–likelihood function, or χ^2. When the correlations between the experimental systematic errors are not available the χ^2 as a function of the PDF parameters is given by

$$\chi^2(\{a\}) = \sum_{k=1}^{N_{pt}} \frac{1}{\sigma_k^2} \left(D_k - T_k\right)^2\,, \tag{81}$$

where N_{pt} is number of data points, and σ_k are the total experimental errors, given by adding the statistical and systematic errors in quadrature. T_k are theoretical predictions, which depend on the PDF parameters $\{a\}$, and D_k are the central values of the experimental measurement.

Modern experiments provide correlated sources of systematic error, in addition to the statistical and uncorrelated systematics. The simplest example is the luminosity error in collider experiments, which is fully correlated among all measurements from the same data sample, but typically there are many other sources that are introduced in the process of any given analysis. In this case, the χ^2 has the form [76]

$$\chi^2(\{a\},\{\lambda\}) = \sum_{k=1}^{N_{pt}} \frac{1}{s_k^2} \left(D_k - T_k - \sum_{\alpha=1}^{N_\lambda} \beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha^2, \tag{82}$$

for N_λ sources of correlated error. Here, s_k represents the total uncorrelated error, which is constructed by adding the statistical error and uncorrelated systematic errors in quadrature. Each source of correlated systematic error is described by a nuisance parameter λ_α, with the error $\beta_{i,\alpha}$ correlated among all data points $\{i\}$. Thus the induced systematic shift to experimental measurement is $\sum_\alpha \beta_{k,\alpha}\lambda_\alpha$. The second sum on right side of Eq. (82) includes the penalty terms to the χ^2, assuming standard normal distributions for the nuisance parameters.

In global analyses we are more interested on the PDF parameters than these nuisance parameters, and so for any given set $\{a\}$ we can first minimise the χ^2 with respect to the nuisance parameters λ_α to give the profiled log-likelihood function $\chi^2(\{a\}) \equiv \chi^2(\{a\}, \{\hat{\lambda}\})$. While naïvely we might worry that this would be a computationally intensive exercise, the simple quadratic dependence of the χ^2 on the λ_α allows the profiled nuisance parameter $\hat{\lambda}_\alpha$ to be solved for analytically, assuming purely Gaussian errors. Explicitly, we have

$$\hat{\lambda}_\alpha = \sum_{i=1}^{N_{pt}} \frac{(D_i - T_i)}{s_i} \sum_{\delta=1}^{N_\lambda} A^{-1}_{\alpha\delta} \frac{\beta_{i,\delta}}{s_i}, \tag{83}$$

with

$$A_{\alpha\beta} = \delta_{\alpha\beta} + \sum_{k=1}^{N_{pt}} \frac{\beta_{k,\alpha}\beta_{k,\beta}}{s_k^2}. \tag{84}$$

By substituting $\hat{\lambda}_\alpha$ into the χ^2 we obtain the profiled χ^2 as a function of the PDF parameters,

$$\chi^2(\{a\}) = \sum_{i,j=1}^{N_{pt}} (T_i - D_i)(\mathrm{cov}^{-1})_{ij}(T_j - D_j), \tag{85}$$

with the covariance matrix and its inverse given by

$$(\mathrm{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}, \;\; (\mathrm{cov}^{-1})_{ij} = \frac{\delta_{ij}}{s_i^2} - \sum_{\alpha,\beta=1}^{N_\lambda} \frac{\beta_{i,\alpha}}{s_i^2} A^{-1}_{\alpha\beta} \frac{\beta_{j,\beta}}{s_j^2}. \tag{86}$$

Thus, the profiled χ^2 is fully determined by the covariance matrix, which itself constructed analytically in terms of the experimental statistical and systematic errors. In certain circumstances, for example the case of most LHCb measurements , the experiments publish the covariance matrix directly, instead of a full breakdown of the experimental systematics.

One final subtlety concerning the construction of the covariance matrix is due to the fact that experimental systematic errors are usually presented as relative errors $\sigma_{i,\alpha}$ with respect to the data, that is

$$(\mathrm{cov})_{ij} = s_i^2 \delta_{ij} + \left(\sum_{\alpha=1}^{N_c} \sigma^{(c)}_{i,\alpha}\sigma^{(c)}_{j,\alpha} + \sum_{\alpha=1}^{N_L} \sigma^{(c)}_{i,\alpha}\sigma^{(c)}_{j,\alpha} \right) D_i D_j. \tag{87}$$

Here, we have further separated these sources into N_c additive and N_L multiplicative errors; in the former case this counts those errors that are absolute in size, while in the latter those sources (such as the luminosity) which genuinely correspond to a relative uncertainty on the data. These have quite different statistical interpretations, and indeed it is known that the above experimental definition of the covariance matrix will result in a D'Agostini bias of the multiplicative errors [315] when used in a PDF fit. Instead, we should use the so-called 't_0' definition of the covariance matrix, given by

$$(\mathrm{cov})_{ij} = s_i^2 \delta_{ij} + \left(\sum_{\alpha=1}^{N_c} \sigma^{(c)}_{i,\alpha}\sigma^{(c)}_{j,\alpha} D_i D_j + \sum_{\alpha=1}^{N_L} \sigma^{(c)}_{i,\alpha}\sigma^{(c)}_{j,\alpha} T_i^0 T_j^0 \right). \tag{88}$$

That is, we should rescale the multiplicative errors not by the data but by the theory prediction T_i^0, from the last iteration of the χ^2 minimization. An alternative prescription is the t definition [4], where the multiplicative errors are rescaled by the same theoretical prediction as in the comparison to the data,

$$(\mathrm{cov})_{ij} = s_i^2\delta_{ij} + \left(\sum_{\alpha=1}^{N_c} \sigma_{i,\alpha}^{(c)}\sigma_{j,\alpha}^{(c)} D_i D_j + \sum_{\alpha=1}^{N_L} \sigma_{i,\alpha}^{(c)}\sigma_{j,\alpha}^{(c)} T_i T_j\right), \tag{89}$$

or the *extended-*t_0 and *extended-t* definitions, where both the additive and multiplicative errors are rescaled by the corresponding theory. A detailed discussion of the various possible prescriptions can be found in [4, 316].

4.2.2. Minimization of χ^2

The central PDF fits are determined by finding the global minimum of the log-likelihood function $\chi^2(\{a\})$. For PDF sets with a moderate, $\sim 10--40$, number of free parameters, numerical gradient–based algorithms are typically used. As a simple example, in Newton's method, the trial solution for the global minimum is given by

$$a_i^{\mathrm{trial}} = a_i^0 - \sum_{j=1}^{n_{\mathrm{par}}} H_{ij}^{-1} d_j \,, \tag{90}$$

for the ith PDF parameter. Here a^0 is an arbitrary starting point in the PDF parameter space, d is the gradient and H^{-1} is the inverse of the Hessian matrix (defined in Sect. 4.3.1) at the same point. This solution is exact assuming a purely quadratic shape for the χ^2, although in practice it can deviate significantly from this when it is far away from the global minimum. The above solution is therefore typically applied iteratively until the desired degree of convergence is achieved. However, the method will fail if the Hessian matrix H is not positive–definite, and can suffer from numerical instabilities. Various quasi–Newton methods have been proposed to overcome these complexities in real applications, such as the Levenberg-Marquardt method used in MSTW/MMHT analyses [80], which is applies a dynamically determined combination of Newton's method and the steepest decent method.

In gradient based methods, the gradient and Hessian matrix must be calculated numerically by means of finite differences. Another class of widely used gradient based algorithms are the variable metric methods (VMM), where it is not necessary to calculate the Hessian matrix numerically. Instead the matrix H is updated iteratively based only on information of the gradients. VMM is the default algorithm in the MINUIT package [317] and is used in CTEQ-TEA analyses [318].

As the number of free parameters is increased, the above methods will begin to suffer from numerical instabilities and issues with local minima. For the NNPDF analysis, where the typical number of parameters is an order of magnitude higher than in other sets, a genetic algorithm is the appropriate choice, as demonstrated in [88]. The basic idea is to start from an ensemble of arbitrarily chosen samples of the PDF parameters. Random mutations with possible crossing–overs are the applied to generate a larger group of new samples. Those candidates predicting a lower χ^2 are then selected to form a new ensemble with the same size. This procedure is then iterated until a suitable convergence criterion is met, while care is taken to prevent overfitting.

4.3. PDF uncertainties

A number of methods have been proposed to determine in a systematic way the uncertainties associated to a PDF fit, and allow these uncertainties to be propagated to cross-sections predictions. These fall into three main classes, known as Hessian, Monte Carlo and Lagrange multiplier methods. Each will be explained in turn in the following sections.

4.3.1. The Hessian method

The Hessian method to quantify PDF uncertainties was first developed in [75]. In this Section, we will describe the basic ingredients of this method and of their subsequent refinements, mostly following the discussion of [80].

Given the χ^2 estimator, the best-fit values correspond to those for which this estimator has a global minimum, $\chi^2_{\rm min}$. In the vicinity of this minimum, the χ^2 can be approximated in terms of a quadratic expansion of the form

$$\Delta\chi^2 \equiv \chi^2 - \chi^2_{\rm min} = \sum_{i,j=1}^{n_{\rm par}} H_{ij}\left(a_i - a_i^0\right)\left(a_j - a_j^0\right) , \tag{91}$$

where the $n_{\rm par}$ fit parameters are denoted by $\{a_1, \ldots, a_{n_{\rm par}}\}$, and the best-fit values that minimize the χ^2 are indicated by $\{a_1^0, \ldots, a_{n_{\rm par}}^0\}$. In the quadratic expansion Eq. (91), we have introduced the Hessian matrix, defined as the matrix of second derivatives of the χ^2 with respect to the fit parameters, namely

$$H_{ij} \equiv \frac{1}{2}\frac{\partial^2\chi^2}{\partial a_i \partial a_j}\bigg|_{\{\vec{a}\}=\{\vec{a}^0\}} . \tag{92}$$

This Hessian matrix contains all the information necessary to quantify the PDF uncertainties. Indeed, for a generic function $\langle\mathcal{F}[\{a_i\}]\rangle$ that depends on the PDFs and thus indirectly on the fit parameters, the associated uncertainty can be computed by means of linear error propagation

$$\sigma_{\mathcal{F}} = T\left(\sum_{i,j}^{n_{\rm par}} \frac{\partial\mathcal{F}}{\partial a_i}\left(H\right)^{-1}\frac{\partial\mathcal{F}}{\partial a_j}\right)^{1/2} , \tag{93}$$

where $T = \sqrt{\Delta\chi^2}$ is the tolerance factor that determines the matching between the allowed range of parameter variations around the best-fit values and the associated confidence interval of the PDF uncertainties. While textbook statistics suggest that $T = 1$ corresponds to a 68% confidence interval, in the context of a global fit there is ample evidence that somewhat larger values for the tolerance are required in the Hessian method, in particular to account for inconsistent experiments, theoretical uncertainties, and for methodological uncertainties such as the specific choice of functional form [311].

The main limitation of Eq. (93) is that in general the derivatives $\partial\mathcal{F}/\partial a_i$ are unknown. This problem can be bypassed by diagonalizing the Hessian matrix and then representing PDF uncertainties in terms of orthogonal eigenvalues. After this diagonalization procedure, Eq. (93) has the simpler form

$$\sigma_{\mathcal{F}} = \frac{1}{2}\left(\sum_{i,j}^{n_{\rm par}}\left[\mathcal{F}(S_i^+) - \mathcal{F}(S_i^-)\right]\right)^{1/2} , \tag{94}$$

where $S_i^\pm$ corresponds to the i-th eigenvector associated to positive and negative variations with respect to the best fit value. Using the eigenvectors $\{S_i^\pm\}$ it is also possible to compute asymmetric PDF uncertainties using the prescription of Ref. [319].

Concerning the determination of the tolerance factor $T = \sqrt{\Delta\chi^2}$, the original studies by the CTEQ and MRST group used values of $T = 10$ and $T = \sqrt{50}$ respectively. In more recent releases, the determination of this tolerance has been refined. In the case of the MSTW08 analyses for example (as well as the subsequent MMHT14 set), the tolerance is determined dynamically for each eigenvector by demanding that all data sets are included within the 68% confidence level variation. To illustrate this, in Fig. 25 we show the

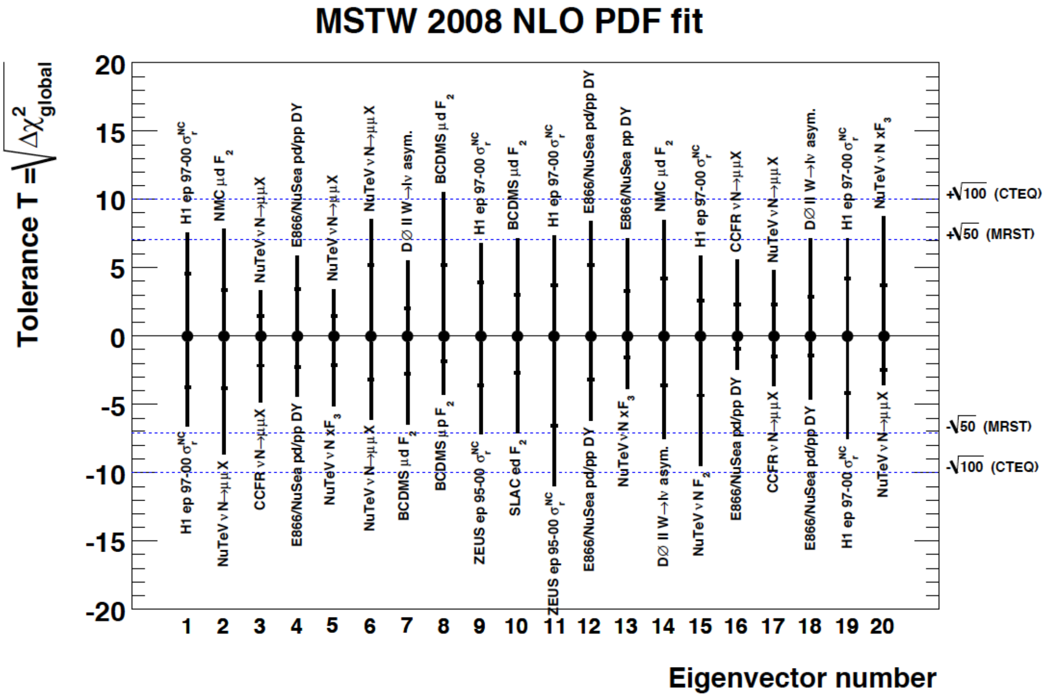


Figure 25: The individual tolerance for each eigenvector, determined by the criterion that each separate experiment should be described within 90% CL. In each case the figure indicates the name of the experiment that determines the tolerance for the various eigenvector directions.

individual tolerance for each eigenvector of the MSTW08 global analysis, determined by the criterion that each separate experiment should be described within 90% CL. For the various eigenvector directions, the figure indicates the name of the experiment that determines the tolerance. The fact that many different experiments are responsible for determining this tolerances emphasizing the crucial importance of using a very wide dataset in the global PDF analysis.

4.3.2. The Monte Carlo method

In the Monte Carlo method, the propagation of the experimental data uncertainties to the parton distributions is achieved by means constructing a Monte Carlo representation of the probability distribution associated to the data. This implies generating a large number $N_{\rm rep}$ of artificial replicas of the original data, so called pseudo–data, which encodes the same information on central values, variances and correlations as that provided by the experiment. In particular, given an experimental measurement of a hard-scattering cross-section denoted generically by $F_I^{(\rm exp)}$ with total uncorrelated uncertainty $\sigma_I^{(\rm stat)}$, $N_{\rm sys}$ fully correlated systematic uncertainties $\sigma_{I,c}^{(\rm corr)}$ and N_a (N_r) absolute (relative) normalization uncertainties $\sigma_{I,n}^{(\rm norm)}$, the artificial MC replicas are constructed using the following expression

$$F_I^{(\rm art)(k)} = S_{I,N}^{(k)} F_I^{(\rm exp)} \left(1 + \sum_{c=1}^{N_{\rm sys}} r_{I,c}^{(k)} \sigma_{I,c}^{(\rm corr)} + r_I^{(k)} \sigma_I^{(\rm stat)} \right) , \quad k = 1, \ldots, N_{\rm rep} , \tag{95}$$

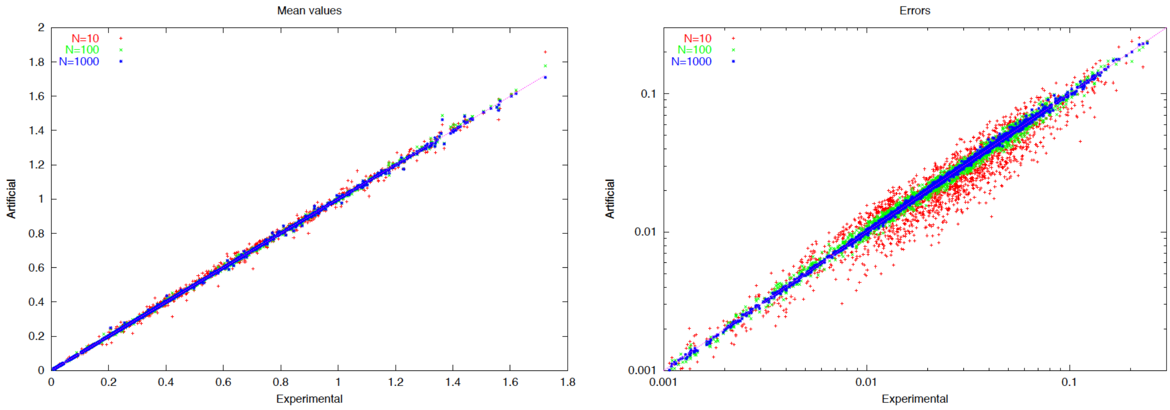


Figure 26: The scatter between the mean values (left) and variances (right plot) of all the data points included in the analysis of [321], comparing the original experimental values with the results obtained from the MC representation for different number $N_{\rm rep}$ of replicas.

where the normalization prefactor is given by

$$S^{(k)}_{I,N} = \prod_{n=1}^{N_a}\left(1 + r^{(k)}_{I,n}\sigma^{(\rm norm)}_{I,n}\right)\prod_{n'=1}^{N_r}\sqrt{1 + r^{(k)}_{I,n'}\sigma^{(\rm norm)}_{I,n'}}. \tag{96}$$

Here the variables $r^{(k)}_{I,c}, r^{(k)}_{I}, r^{(k)}_{p,n}$ are univariate gaussian random numbers. Eq. (95) represents the fluctuations of the pseudo-data replicas around the measured central values by the amount allowed by the experimental uncertainties. Note that for each replica the random fluctuations associated to a given fully correlated systematic uncertainty will be the same for all data points, $r^{(k)}_{I,c} = r^{(k)}_{I',c}$. The same condition holds for the normalization uncertainties.

An important question in the Monte Carlo method is how many replicas $N_{\rm rep}$ need to be generated in order to achieve a faithful representation of the underlying probability density in the space of data. To this purpose, a number of statistical estimators were constructed in Ref. [320]. It was found that $N_{\rm rep} = 10$ replicas are enough to reproduce central values, $N_{\rm rep} = 100$ for the variances and that $N_{\rm rep} = 1000$ to satisfactorily reproduce the data correlations. Subsequent analysis have shown that this statement holds for a generic input dataset. To illustrate this point, in Fig. 26 we show the scatter between the mean values and variances of all the data points included in the analysis of [321], comparing the original experimental values with the results obtained from the MC representation for different number $N_{\rm rep}$ of replicas. We find that indeed for central values, the averages computed using only $N_{\rm rep} = 10$ replicas agree with the original data, but that for variances this is not the case, where $N_{\rm rep} = 100$ replicas are required.

Once the Monte Carlo sampling of the experimental data has been achieved, a separate PDF fit is performed in each replica. This can be done using traditional polynomial functional forms or other interpolators such as artificial neural networks. The resulting sample of $N_{\rm rep}$ PDF replicas realizes the concept of the probability density in the space of parton distributions. The calculation of the resulting PDF uncertainties and their propagation to generic cross-sections can be performed using textbook methods. Note that in this approach the PDF uncertainty propagation is fully general, and in particular is not restricted to the Gaussian approximation. For instance, in the Monte Carlo method the expectation function of a generic cross-section

$\mathcal{F}[\{q\}]$ is evaluated as an average over the replica sample,

$$\langle \mathcal{F}[\{q\}] \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{F}[\{q^{(k)}\}] \,, \tag{97}$$

and the corresponding uncertainty is then determined as the variance of the Monte Carlo sample,

$$\sigma_{\mathcal{F}} = \left(\frac{1}{N_{\rm rep}-1} \sum_{k=1}^{N_{\rm rep}} \left(\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle \right)^2 \right)^{1/2} \,. \tag{98}$$

These formulae may also be used for the determination of central values and uncertainties of the parton distribution themselves, in which case the functional $\mathcal{F}$ is identified with the parton distribution $q : \mathcal{F}[\{q\}] \equiv q$.

In the case of a fully consistent dataset, the Monte Carlo method to estimate the PDF uncertainties is expected to coincide with the Hessian method described in Sect. 4.3.1 for a standard tolerance $\Delta\chi^2 = 1$. This equivalence was explicitly demonstrated in the HERA-LHC workshop proceedings [322]. In Fig. 27 we show the gluon PDF at $Q = 2$ GeV in this HERA-LHC benchmark fit, based on HERA inclusive structure function data, where the one-sigma PDF uncertainties computed with the Hessian method (black lines) are compared to those of the Monte Carlo method (red lines), finding good agreement. In this figure, each of the green curves corresponds to an individual MC replica. In the left fit, the normalization and systematic uncertainties in the MC replicas from Eq. (95) fluctuate according to a multi-Gaussian distribution, while in the right fit they fluctuate instead according to a log-normal distribution, showing that the PDFs depend only weakly on the specific assumptions about the specific probability distribution of the experimental systematic uncertainties.

Finally, we note that a Hessian representation of a Monte Carlo PDF sets can be accurately constructed using the `mc2h` algorithm developed in Ref. [323]. This technique is discussed in more detail in Sect 4.5.

4.3.3. The Lagrange multiplier method

The Lagrange multiplier method was originally developed in Ref. [324, 318], and is a generalization of the χ^2 minimization procedure. As in the Hessian case, the first step is to find the PDF parameters $\{a_i^0\}$ that minimize the global $\chi^2(\{a_i\})$. Then one has to select a specific physical quantity that depends on the PDFs, such a DIS structure function or a cross-section, which we denote generically by $\mathcal{F}(\{a_i\})$, which takes the value $\mathcal{F} = 0 = \mathcal{F}(\{a_i^0\})$ at the global fit minimum. The goal of the Lagrange multiplier method is to determine the PDF uncertainty associated to $\mathcal{F}_0$ without making any assumption on the specific behaviour of the χ^2 around the global minimum.

In order to achieve this, the global fit χ^2 is modified by introducing the physical quantity $\mathcal{F}$ as a Lagrange multiplier, so that the new function that needs to minimized is now given by

$$\Psi(\lambda, \{a_i\}) = \chi^2(\{a_i\}) + \lambda \mathcal{F}(\{a_i\}) \,. \tag{99}$$

Now for each specific value of λ, denoted by λ_α, the minimization of Eq. (99) will lead to a different set of best-fit PDF parameters, which we indicate by $\{a_i^{\rm (min)}(\lambda_\alpha)\}$. Mathematically, these parameters are the result of a constrained PDF fit where the value of the physical observable has been fixed to $\mathcal{F}_\alpha = \mathcal{F}(\{a_i^{\rm (min)}(\lambda_\alpha)\})$. The resulting PDF set of this constrained fit is now indicated by S_α.

The main result of this procedure is establishing a parametric relation between the value of the physical quantity $\mathcal{F}$ and the global fit χ^2 by means of the Lagrange multiplier λ. This means that we can determine the

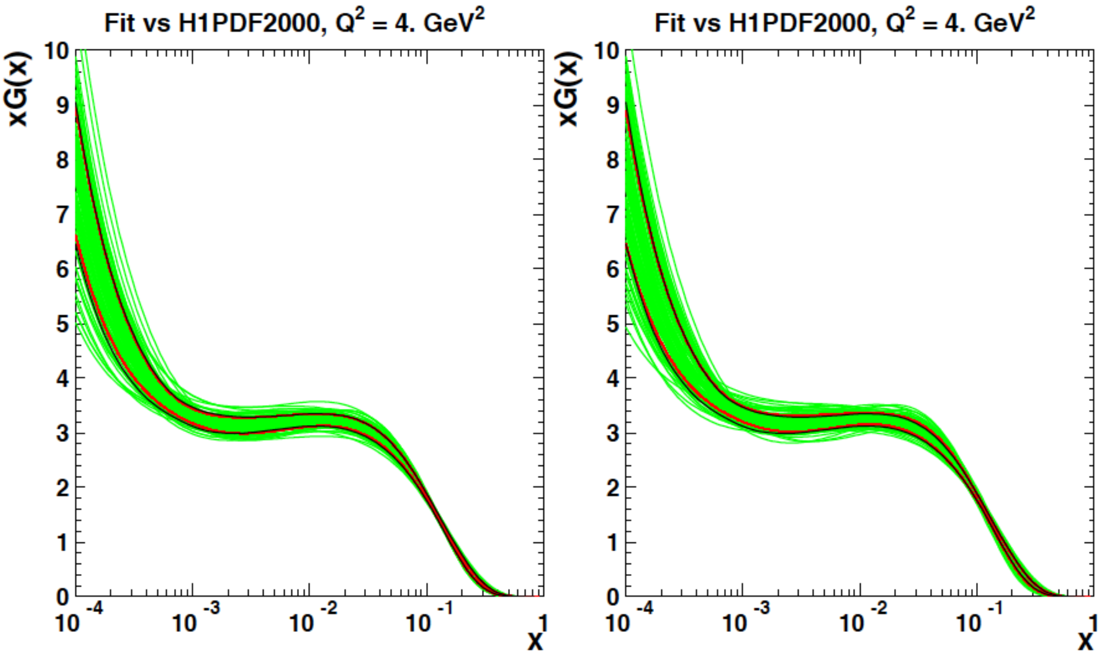


Figure 27: The gluon PDF at $Q = 2$ GeV in the HERA-LHC benchmark fit of Ref. [322], where the one-sigma PDF uncertainties computed with the Hessian method (black lines) are compared to those of the Monte Carlo method (red lines), finding good agreement. Each of the green curves corresponds to an individual MC replica. In the left fit, the normalization and systematic uncertainties in the MC replicas from Eq. (95) fluctuate according to a multi-Gaussian distribution, while in the right fit they fluctuate instead according to a log-normal distribution.

PDF uncertainty associated to $\mathcal{F}$ imposing that the χ^2 satisfies $\chi^2 = \chi^2_{\rm min} + \Delta\chi^2$ with $\Delta\chi^2 = T^2$ representing the tolerance, as introduced in the previous section. Is clear that the main advantage of the Lagrange multiplier method as compared to the Hessian method is that one does not need to restrict to the quadratic expansion or linear error propagation, since the PDF uncertainties in this method are determined only by the values of the χ^2 and not by its specific shape. On the other hand, an important restriction of the method is that the PDF error analysis for each specific physical quantity $\mathcal{F}$ requires redoing a large number of new PDF fits, and this is not only very CPU time intensive but it can also not be done outside the PDF fitting collaboration.

The Lagrange multiplier method is schematically illustrated in Fig. 28. In the left plot we show a two-dimensional projection of the PDF parameter space, indicating the contours in χ^2 for fixed values of the physical quantity $\mathcal{F}$. The parametric relation is provided by the value of multiplier λ. In the right plot we show how the PDF uncertainty associated to $\mathcal{F}$ for a given confidence interval is determined by the condition that the global χ^2 should not grow beyond the tolerance $\Delta\chi^2$. As in the case of the Hessian method, the specific value of the tolerance $T = \sqrt{\Delta\chi^2}$ is an input to the method and must be determined independently.

4.4. *Treatment of theory parametric uncertainties*

PDF determination also depends on theoretical QCD inputs, namely the strong coupling constant and the heavy quark masses. The choice of these input parameters therefore acts as an additional source of PDF uncertainty. As there can be strong correlations between the PDFs and these inputs, a complete evaluation

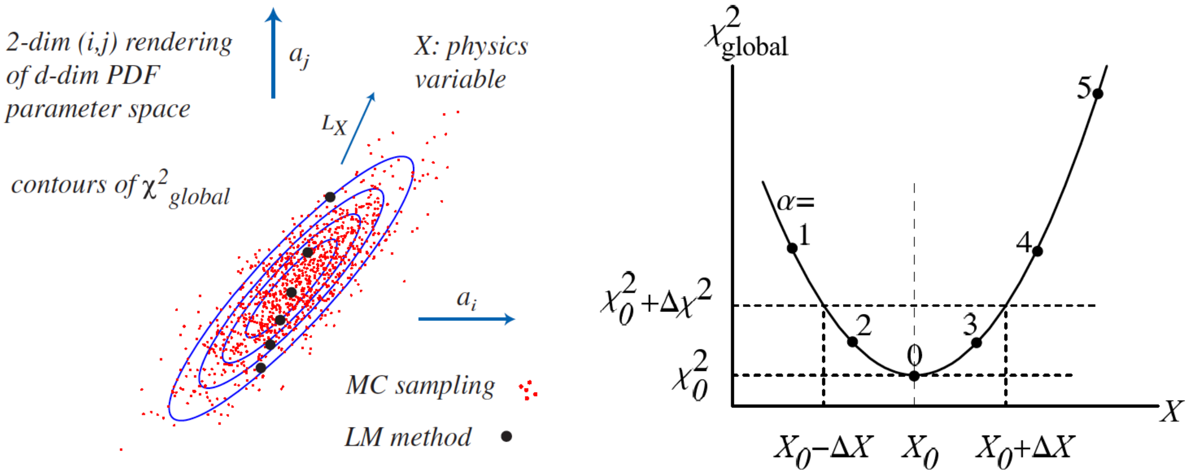


Figure 28: Schematic representation of the Lagrange Multiplier method. In the left plot we show a two-dimensional projection of the PDF parameter space, indicating the contours in χ^2 for fixed values of the physical quantity $\mathcal{F}$. In the right plot we show how the PDF uncertainty associated to $\mathcal{F}$ for a given confidence interval is determined by the condition that the global χ^2 should not grow beyond the tolerance $\Delta\chi^2$.

of the PDF errors requires a consistent combination of these with the PDF parametric uncertainties. On the another hand, the global analysis can also provide an independent determination of those QCD parameters, which can contribute to the world average values.

4.4.1. Strong coupling constant

The current world average value for the strong coupling constant is $\alpha_s(M_Z) = 0.1181 \pm 0.0011$. This is extracted from 6 subsets of measurements, namely τ decay, lattice results, e^+e^- jets and event shapes, structure functions, EW precision fits, and $t\bar{t}$ cross sections at LHC. These are combined with theoretical predictions at NNLO or higher orders using the χ^2 averaging method [325]. The 2015 PDF4LHC combined PDF sets are based on a slightly different value of $\alpha_s(M_Z) = 0.1180 \pm 0.0015$ [2], i.e. rounded to a value that is often used in global fits and with a somewhat more conservative uncertainty band. Individual PDF groups also extract values of $\alpha_s(M_Z)$ including the uncertainties that are solely from their global analyses.

The choice of strong coupling constant affects a global PDF analysis in two principle ways, through the DGLAP evolution of the PDFs themselves, and the perturbative QCD predictions for the processes that enter the fit. To study these effects a scan over different values of $\alpha_s(M_Z)$ value is typically performed. For each choice of α_s the best-fit of PDFs is found and the χ^2 profile is constructed. The best–fit value of $\alpha_s(M_Z)$ is then identified and the uncertainty on this can be determined in a similar way to the standard PDF uncertainties, using either a '$\Delta\chi^2 = 1$' or a tolerance criteria. Fig. 29 shows the χ^2 profile from MMHT and NNPDF NNLO global analyses. The extracted $\alpha_s(M_Z)$ values at NNLO are 0.1172 ± 0.0013 [326] and 0.1173 ± 0.0007 [327] respectively. The CT and ABMP groups have also extracted values of the strong coupling, finding at NNLO the values 0.115 ± 0.003 [18] and 0.1147 ± 0.0008 [20], that is with lower central values than MMHT and NNPDF. The error reported by the CT group is much larger than the other groups due to the stronger tolerance condition used. There is therefore a large spread in the best-fit values from the different PDF groups, and so the combined 0.1156 ± 0.0021 which enters the world average has a much larger error than those reported by individual groups. At NLO the global analyses prefers a $\alpha_s(M_Z)$ value that is about 0.002 ∼ 0.003 higher than at NNLO, compensating for the missing higher–order corrections.

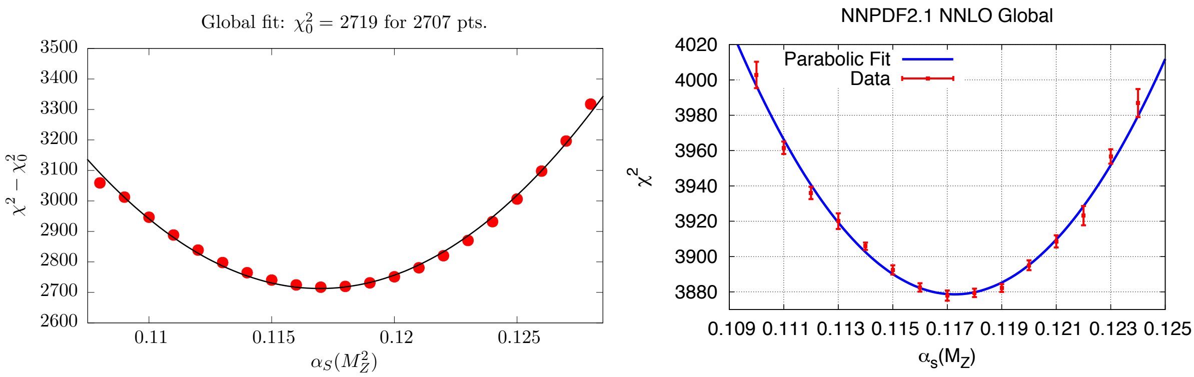


Figure 29: Left plot: the profile of global χ^2 as in a scan of $\alpha_s(M_Z)$ in MMHT2014 NNLO analysis [326]. Right plot: the profile of global χ^2 as in a scan of $\alpha_s(M_Z)$ in NNPDF2.1 NNLO analysis [327]. The error bars indicate fluctuations of the χ^2 due to finite number of MC replicas.

The choice of strong coupling constant obviously has a significant impact on the predictions for various important processes at hadron colliders, such Higgs boson production via gluon fusion and top quark pair production, both of which are proportional to α_s^2 at LO. As mentioned above it is crucial to account for the correlations between α_s and PDFs when evaluating the full uncertainties of observables at hadron colliders. For example, it is well known that the gluon PDF is anti-correlated with α_s in the small and intermediate x regions due to the constraints from scaling violations of inclusive structure functions, which can partly compensate the change of cross sections due to change of α_s in the matrix elements. In principle in global analysis one can treat $\alpha_s(M_Z)$ in exactly the same way as other PDF parameters, e.g., in the Hessian method, by calculating the full Hessian matrix, determining the eigenvector directions and the uncertainties along each direction. In this way the PDF+α_s uncertainty on any observable can be evaluated using the Hessian error PDFs in exactly the same way as for the standard case where only PDF uncertainties are included. The ABM and later ABMP group follows exactly this procedure.

The downside of this approach is that it is not possible to separate the PDF and α_s uncertainties, and each error PDF will be associated with a different value of α_s. A much convenient but completely equivalent method has been proposed in [328]. Here, it has been shown that, under the quadratic approximation for the χ^2 the full PDF+α_s uncertainty can be calculated by simply adding the usual PDF uncertainty and the α_s uncertainty in quadrature, with the eigenvectors for PDF uncertainties constructed with α_s fixed to its best–fit value. The α_s uncertainty is then calculated through one additional eigenvector (with two directions) constructed by fixing $\alpha_s(M_Z)$ to its upper and lower limits and then fitting the remaining PDF parameters in the usual way. The equivalence of the above two approaches is shown in Fig. 30 for the gluon and charm quark PDFs. This latter approach is now adopted by CT, MMHT, and NNPDF collaborations due to its simple form and ease of use. Note that the upper and lower limits on $\alpha_s(M_Z)$ can come either from the fit itself, as in the case of MMHT 2014, or can be chosen according to the world average, as in CT14 and NNPDF3.1. Changes of the α_s uncertainty for different input errors on α_s can be easily obtained by a linear rescaling [2].

4.4.2. Heavy quark masses

Global PDF analyses also rely on the input of heavy quark (charm, bottom and top) masses. In an analysis that uses the GM-VFNS (see Sect. 2.5) the charm and bottom quark masses enter through the

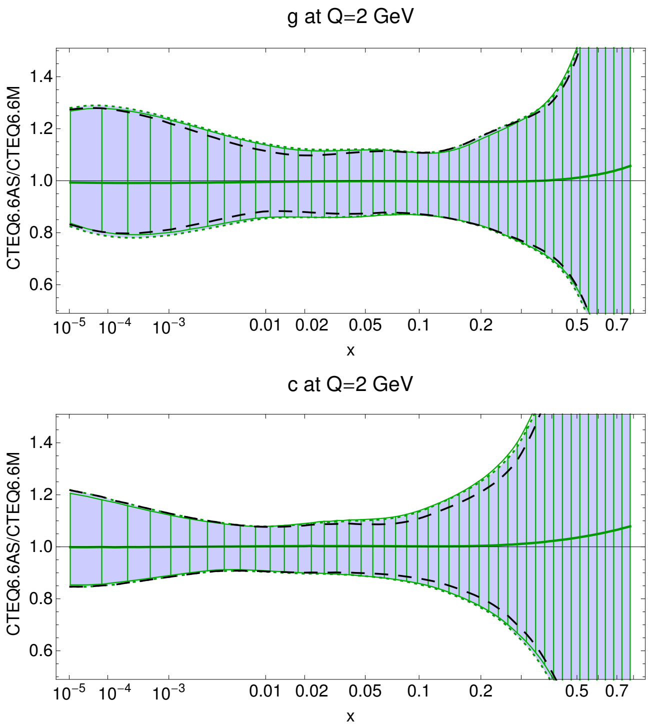


Figure 30: Comparison of the PDF+α_s uncertainties for the gluon and charm quark PDFs using the full eigenvectors including $\alpha_s(M_Z)$ in the Hessian matrix (filled error band with dotted borders) and with the separate PDF and α_s uncertainties added in quadrature (hatched band with solid borders) [328]. The dashed lines represent the PDF uncertainties only.

running of α_s, the boundary conditions for the switching of active flavors, as well as the predictions for the inclusive DIS structure functions and for open charm/bottom production in DIS or hadron–hadron collisions. The dependence on the top quark mass is less pronounced unless top quark production data are included in the analysis. In this case as the $gg \to t\bar{t}$ process is strongly dependent on s α_s, the top quark mass and the gluon PDF, this will induce a strong correlation between these three objects. The world averages for m_c and m_b in $\overline{\text{MS}}$ scheme are [325]

$$m_c(m_c) = 1.27 \pm 0.03\,\text{GeV}, \;\text{and}\; m_b(m_b) = 4.18 \pm 0.035\,\text{GeV}\,, \tag{100}$$

which can be translated into the pole masses as [329]

$$m_c(m_c) = 1.5 \pm 0.2\,\text{GeV}, \;\text{and}\; m_b(m_b) = 4.9 \pm 0.2\,\text{GeV}\,, \tag{101}$$

by using the 3–loop conversion for the bottom quark together with a known relation between the bottom and charm masses [330]. The large uncertainties here are due to the fact that the pole mass is in fact not well defined due to the diverging series, i.e. there is a renormalon ambiguity of order $\sim 0.1-0.2$ GeV. On the other hand, this effect largely cancels in the difference of the two masses, and therefore the above uncertainties are highly correlated. The majority of PDF groups use the pole mass as input, as the relevant coefficient functions and matrix elements are calculated in on–shell scheme. In particular, CT14 take a default value of $m_{c(b)} = 1.3(4.75)$ GeV, MMHT14 take 1.4(4.75) GeV and NNPDF3.1 take 1.51(4.92) GeV. Both the

CT14 and MMHT 2014 NNLO analyses prefer a lower charm quark mass of about 1.3 GeV [18, 329] if it is treated as a free parameter, and fitted to data which is consistent with the conversion from the world average value. In ABMP16 the $\overline{\text{MS}}$ masses are extracted directly from the fit, giving $m_c(m_c) = 1.252 \pm 0.018$ GeV and $m_b(m_b) = 3.84 \pm 0.12$ GeV [20].

As in the case of α_s the uncertainty due to the heavy quark masses can be calculated by constructing an additional eigenvector from fits with alternative mass values. The full uncertainty can then be obtained by adding it in quadrature to the PDF uncertainty obtained with the default choice of heavy quark masses. For example, the CT14, MMHT 2014 and NNPDF3.1 analyses all provide a series of best–fit PDFs with m_c or m_b fixed to alternative values around their default choices. However, there has so far not been an agreement on a common choice of the heavy quark masses and their errors in global PDF analyses, although this is foreseen for the next PDF4LHC recommendation. Fig. 31 shows dependence of the predicted total cross sections for weak boson and Higgs boson production at the 13 TeV LHC on the choice of charm quark pole mass used in CT14 NNLO analysis [26]. It is found that varying m_c by 0.2 GeV has a negligible effect on the Higgs boson cross section and induces at most a 2% change in the weak boson cross sections. This is well within the PDF uncertainties. Similar conclusion has been reached in the MMHT 2014 analysis [329]. In addition, in both the CT14 and NNPDF3.1 analyses, it is observed that the effect of varying m_c can be partly cancelled by changes of the non–perturbative component of the charm PDFs.

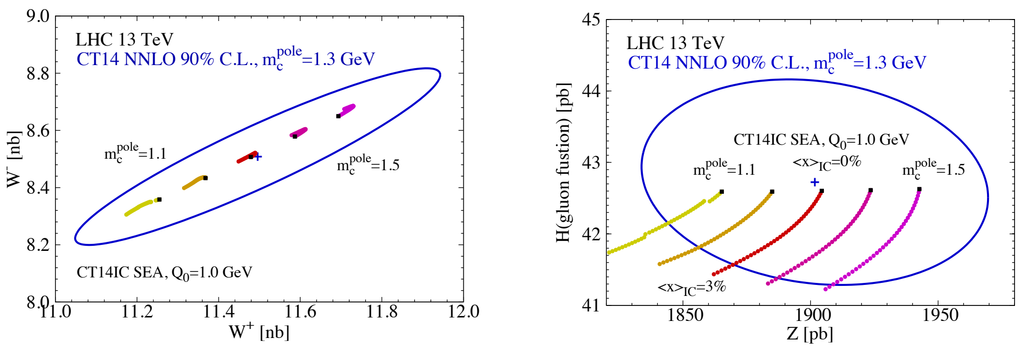


Figure 31: Dependence of the total cross sections for weak boson and Higgs boson production at the 13 TeV LHC 13 on the choice of charm quark mass in the CT14 NNLO analysis [26]. The ellipse indicates the PDF uncertainties at 90% C.L. Also shown is the dependence of the cross sections on the assumed momentum fraction carried by non–perturbative charm PDF at the initial scale.

4.5. Combined and reduced PDF sets

Individual PDF sets from different groups are widely used when comparin precision theoretical predictions with LHC measurements and in the assessment of the accuracy of PDF sets themselves. However, for many LHC applications an assessment of the *total* PDF uncertainty for certain observables, by taking into account predicitons from all applicable PDF sets, is required. This will for example be the case in the extraction of the couplings of the Higgs boson, or the calculation of signal and background rates in searches for BSM physics. For these purposes a statistical procedure is needed with which to combine results different PDF sets. Such a statistical combination is much more complicated than in the case of for example the world average of the strong coupling constant or heavy–quark masses, since it combines functions which have in principle an infinite number of degrees of freedom. The prescription must accommodate the fact that the individual PDF sets are not identical either in their central values or in their uncertainties, and it should account for possible correlations between PDF sets from different groups. In addition, it will be desirable to maintain a compact set of PDFs in the final combination.

The 2010 PDF4LHC recommendation proposed the use of a simple envelope prescription [5, 4, 92]. That is, the PDF determinations from different groups are treated as instances of a probability distribution affected by unknown sources of systematics rather than statistically distributed instances of an underlying probability distribution. This envelope prescription can also only be applied at the level of individual observables, without information on PDF induced correlations. Given the better understanding of current PDF determinations, the relatively good agreement between global set, and the high precision demands for LHC Run II studies, such a prescription is therefore certainly inadequate.

With the above considerations in mind, the updated 2015 PDF4LHC recommendation was proposed as a replacement [2]. There are certain criteria for the individual PDF sets to be considered for the combination. First, the individual PDF sets should be based on a global determination with a large number of datasets from a variety of experiments, that is DIS and hadron–hadron scattering in fixed–target and collider experiments. Second, the hard cross sections for DIS and hadron–hadron scattering processes used in the extraction should be evaluated up to two loops in QCD in a GMVFN scheme, with a maximum number of 5 active quark flavors. Third, all known experimental and procedural sources of uncertainties should be properly accounted for, including the experimental uncertainties propagated from data, uncertainties due to the incompatability of different data sets, and uncertainties due to the functional form of PDFs. It was decided that the combination should be carried out with a central value of $\alpha_s(M_Z) = 0.118$ at both NNLO and NLO and with the uncertainties of $\alpha_s(M_Z)$ taken to be 0.0015, consistent with the PDG world-average [331]. The heavy quark masses used in individual PDF sets are not currently required to be the same[3] but should be compatible with their world–average values. The existing PDF sets satisfying all of the above requirements at present have been identified as CT14 [18], MMHT2014 [19], and NNPDF3.0 [17]. The PDF4LHC 2015 PDF sets are therefore statistical combination of these three global analyses.

The combination can only be carried out efficiently using the Monte-Carlo method, as different PDF determinations adopt different forms for the PDF parametrizations. In the first step the CT14 and MMHT2014 PDFs, which are originally in Hessian form, are converted into their Monte-Carlo representations by applying the Watt–Thorne method with symmetric formula [332]. It has been validated that a MC ensemble with 300 replicas is sufficient to reproduce the central value and uncertainties of the original Hessian PDFs to high precision. The NNPDF3.0 PDF set is already in a Monte-Carlo form with 1000 repicas. Following the idea of individual PDF determinations as equally likely representations of an underlying probability distribution, a combined PDF set is then build by taking 300 MC replicas from each input PDF sets and merging them equally. The resulting combined PDF set, an ensemble of 900 MC replicas, is referred to as the MC900 or PDF4LHC15_prior, and represents the combined probability distribution of the PDFs. However, such a set of 900 PDFs would be unmanageably large for most applications, in particular given the time and storage cost required for complicated NNLO calculations and experimental simulations. Therefore, various methods have been applied to reduce the size of the combined sets, while minimizing the information loss according to various statistical measures.

The first method applies the META–PDFs framework [333]. Here, a flexible functional form with Bernstein polynomials is chosen to parametrize the PDFs at an initial scale. Each replica in the MC900 ensemble is then represented by a group of PDF parameters through a fit to the chosen parametrizations, by minimizing a metric function. The prior probability distribution of PDFs is thus transformed into probability

[3]It will be desirable in the future for all PDF groups provide error sets with common choices of heavy-quark masses and furthermore to include the uncertainties due to the mass inputs, similarly to the case of α_s.

distributions in the PDF parameter space. The covariance matrix of the PDF parameters is calculated,

$$cov(a_l, a_m) = \frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} (a_l^{(k)} - a_l^{(0)})(a_m^{(k)} - a_m^{(0)}), \tag{102}$$

where $a_l^{(0,k)}$ denotes the fitted PDF parameters from the central set and the k-th MC replicas and N_{rep} is the total number of MC replicas. The covariance matrix can be diagonalized by an orthogonal transformation. Eigenvectors are then calculated and ordered according to their impact on the PDF uncertainties with a designed error metric; the eigenvectors with smaller contributions can be dropped according to the accuracy required. Finally a central PDF set and a group of orthogonal error PDF sets are generated under the assumption of a multi–Gaussian distribution. They can then be used in a similar way as the conventional Hessian PDF set. For example, the 68% cl uncertainty or 1σ error is given by

$$\delta^{\text{PDF}} X = \sqrt{\sum_{i=1}^{N_{eig}} (X_i - X_0)^2}, \tag{103}$$

where X_0 is the prediction on observable X given by the central set and X_i is the prediction given by the i-th error set. Note there is only one error set along each eigenvector/orthogonal direction since symmetric Gaussian distributions are assumed in this case.

The second method is to use the MC2Hessian algorithm with Singular Value Decomposition, followed by the Principle Componenent Analysis [323]. The idea is to first discretize the PDFs with $N_x N_{\text{pdf}}$ observables which are the PDF values at the corresponding grid point, where N_x denotes the total number of grid points in momentum fraction x and N_{pdf} is the number of total independent flavors. A $N_x N_{\text{pdf}} \times N_x N_{\text{pdf}}$ covariance matrix on all those PDF values can be constructed from all the MC replicas,

$$cov_{ll'} = \frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} X_{lk} X_{kl'}^T = \frac{1}{N_{\text{rep}} - 1} X X^T \tag{104}$$

where X_{lk} is the PDF value on l-th grid point given by k-th MC replica subtracted with the corresponding value from central PDF set, N_{rep} is the total number of MC replicas. The above covariance matrix can be rewritten in its singular value decomposition form

$$cov_{ll'} = \frac{1}{N_{\text{rep}} - 1} (U S V^T)(U S V^T)^T, \tag{105}$$

where S is a diagonal matrix constructed out from singular values of X, V is an orthogonal $N_{\text{rep}} \times N_{\text{rep}}$ matrix of coefficients, and U is a $N_x N_{\text{pdf}} \times N_{\text{rep}}$ matrix containing orthogonal eigenvectors of the covariance matrix with nonzero eigenvalues. Indeed the matrix V gives a Hessian basis built upon linear combinations of original MC replicas, which reproduces fully the covariance matrix given by the original MC replicas. This basis can be further truncated using Principle Component Analysis with a certain error metric, resulting in a smaller Hessian PDF set similar to the case of META-PDFs.

A third option is provided by the compressed Monte-Carlo (CMC) method [334]. In this case an ensemble of pseudo-MC replica PDFs (CMC-PDFs) are generated. The CMC-PDFs have a different statistical interpretation compared to the native MC PDFs. However, certain statistical measures, such as the mean, covariance matrix, skewness, kurtosis and the Kilmoforov distance can be reconstructed in a similar way to the native MC PDFs. The CMC-PDFs aim to preserve some of the non-Gaussian features in the prior given by MC900 in addition to the Gaussian features, for which the Hessian form is more adequate. Note that in the current prescription some of the non-Gaussian behaviours from individual PDF sets have been smoothed out due to the symmetric formula used in converting Hessian PDFs to MC replicas. The compression starts with a figure of merit,

$$ERF = \sum_k \frac{1}{N_k} \sum_i \left(\frac{C_i^{(k)} - O_i^{(k)}}{O_i^{(k)}} \right), \tag{106}$$

where k runs over the number of chosen statistical estimators, N_k is a normalization factor, $O_i^{(k)}$ is the value of the k-th estimator calculated at the generic point (x_i, Q_i) from the prior and $C_i^{(k)}$ is the corrsponding value of the same estimator in the compressed set. The compressed set is simply a subset of the MC900 ensemble. For any given number of total MC replicas, the compressed set is chosen by minimization of the above error function using a generic algorithm.

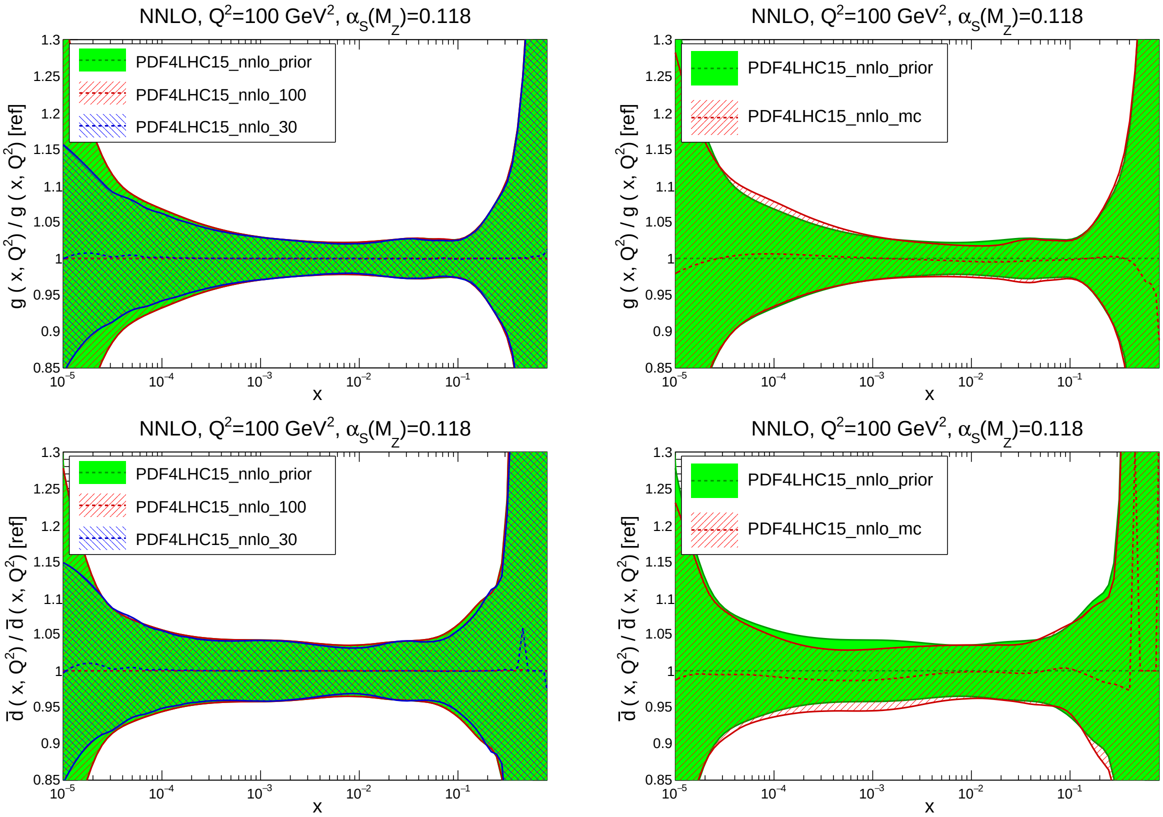


Figure 32: Comparison of the gluon and $\bar{d}$-quark PDFs at a scale of $Q^2 = 100\,\mathrm{GeV}^2$ between the prior and the two reduced Hessian sets, and between the prior and the compressed MC set, normalized to the central value of the prior [2].

Finally, in the 2015 PDF4LHC recommendation there are three reduced PDF sets provided to the public: a Hessian set with 30 error PDFs (PDF4LHC15_30), a Hessian set with 100 error PDFs (PDF4LHC15_100), and a compressed MC set with 100 replicas (PDF4LHC15_mc). All of these are constructed from the same prior (MC900), but have a slightly different focus in each case. The symmetric PDF uncertainties of any observables can be calculated using Eq. (103) for Hessian sets and the usual master formula for MC PDFs.

Fig. 32 shows the comparison of the central/mean value and the uncertainties of the gluon and $\bar{d}$ quark PDFs for the prior and the three reduced sets. The agreement between the Hessian set with 100 eigenvectors and the prior is good for all PDF combinations in the complete range of x. The Hessian set with 30 eigenvectors also shows good agreement with the prior in the x range related to precision physics measurements, but gives slightly smaller uncertainty in the extrapolation regions at small- and large-x as a tradeoff of fewer error PDFs. The compressed MC set also agrees well with the prior in most of the region for the mean and uncertainty except for small overall fluctuations.

The PDF4LHC recommendation for the usage of different these PDF4LHC15 sets depends on the particular case under consideration:

- Use individual PDF sets, and, in particular, as many of the modern PDF sets as possible for comparisons between data and theory for standard model measurements.
- Use the PDF4LHC15_mc sets for searches for BSM phenomena where non–Gaussian behaviour could be important.
- Use the PDF4LHC15_30 sets for calculation of PDF uncertainties in situations when computational speed is needed, or a more limited number of error PDFs may be desirable.
- Use the PDF4LHC15_100 sets for calculation of PDF uncertainties in precision observables.

The cases listed above are not exclusive, and one or the other will be more appropriate depending on the theoretical interpretation of a given experimental measurement. In addition, there are two further PDFs with $\alpha_s(M_Z) = 0.1165$ and 0.1195 in the PDF4LHC15 sets, provided for estimation of the uncertainty due to α_s input. The corresponding uncertainty at 68% cl for the observable X is given by

$$\delta^{\alpha_s} X = \frac{X(\alpha_s = 0.1195) - X(\alpha_s = 0.1165)}{2}, \tag{107}$$

where $X(\alpha_s)$ is the value calculated using the PDF together with the hard matrix elements evaluated at that α_s value. The combined PDF+α_s uncertainty is then computed as follows

$$\delta^{\mathrm{PDF}+\alpha_s} X = \sqrt{(\delta^{\mathrm{PDF}} X)^2 + (r \cdot \delta^{\alpha_s} X)^2}, \tag{108}$$

where the rescaling factor $r = 1$ is recommendated but can be varied according to user's choice of uncertainty on $\alpha_s(M_Z)$.

It is also noted that the PDF4LHC15 PDF sets can be further reduced to compact sets with around ten eigenvectors or less if the applications are restricted to a certain group of observables, e.g., the cross sections and distributions in Higgs boson production at the LHC. That can be achieved either through the data set diagonalization method [335, 336] or the singular value decomposition method [337].

4.6. Approximate methods

Now we turn to discuss two approximate methods that can, under certain circumstances, be used instead of a full fledged PDF fit, namely the Bayesian reweighting of Monte Carlo replicas [338, 339] and the profiling of Hessian sets [340]. The main advantage of these two techniques is that they can be used *e. g.* to quantify the impact of new experiments on a pre-existing fit based only on publicly available information, in particular the `LHAPDF` grids. On the other hand these methods have a number of limitations, and are not able to account for the effect of methodological changes, for example in the input PDF parametrization, or of modifications in the theoretical calculations.

4.6.1. Bayesian Monte Carlo reweighting

The Bayesian reweighting method [338, 339] can be applied to any Monte Carlo set to quantify the impact at the PDF level of a new experimental measurement. The basic idea is that, starting from a sample of $N_{\rm rep}$ MC replicas all carrying equal weight, the compatibility of each replica with the new experimental dataset can be quantified by computing a series of new weights for each replica given by

$$\omega_k = \frac{\left(\chi^2_k\right)^{(n-1)/2} e^{-\chi^2_k/2}/N_{\rm rep}}{\frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}\left(\chi^2_k\right)^{(n-1)/2} e^{-\chi^2_k/2}} \,, \quad k=1,\ldots,N_{\rm rep}\,, \tag{109}$$

where χ^2_k is the goodness–of–fit estimator between the replica k and the new experimental measurement. For instance, if for a given replica the agreement with the new experiment is very poor, its χ^2_k will be large and thus the weight of this specific replica will be exponentially suppressed. Note that by definition these new weights ω_k are appropriately normalized, and from the statistical point of view, they can be interpreted as the probability of the replicas f_k, given the χ^2_k for the new experimental measurement.

One of the limitations of the Bayesian reweighting method is that it entails a given loss of information as compared to the initial prior, because some of the original $N_{\rm rep}$ MC replicas will carry a very small weight, meaning that they have been effectively discarded. One suitable estimator to quantify this efficiency loss is the so-called Shannon entropy, which allows the effective number of replicas left out after the reweighting to be evaluated. This is defined as

$$N_{\rm eff} \equiv \exp\left[\frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}\omega_k \ln\left(N_{\rm rep}/\omega_k\right)\right]\,, \tag{110}$$

where by construction, $0 \le N_{\rm eff} \le N_{\rm rep}$. The interpretation of this effective number of replicas is that a reweighted PDF set carries the same amount of information as a direct refit based on $N_{\rm eff}$ replicas. Clearly, the smaller $N_{\rm eff}$ is, the more the new dataset constrains the PDFs, but on the other hand if $N_{\rm eff}$ becomes small enough, the reweighting method loses validity and a full refit becomes necessary.

An advantage of the Bayesian reweighting method is that it provides a way to estimate if the experimental uncertainties have been either under or overestimated, assuming that theoretical uncertainties are under control. To achieve this, it is possible to rescale the total experimental uncertainties of the data by a factor α, and then use inverse probability in order to evaluate the probability density associated to the rescaling parameter α, namely

$$\mathcal{P}(\alpha) \propto \frac{1}{\alpha}\sum_{k=1}^{N_{\rm rep}}\omega_k(\alpha)\,, \tag{111}$$

where the weights $\omega_k(\alpha)$ are computed using Eq. (109) but replacing χ^2_k by χ^2_k/α^2, and therefore represent the probability of f_k given the new data with rescaling error. If this probability density Eq. (111) peaks far above (below) one, then this suggest that the uncertainties in the data have been under (over) estimated, providing a useful handle to assess the compatibility of a new measurement with a prior PDF analysis.

The Bayesian Monte Carlo method has been studied in many PDF applications, both for proton (unpolarized and polarized) PDFs, see for instance Refs. [252, 341, 290, 172, 28], and for nuclear PDFs [342, 343, 344, 31]. Crucially, its explicit validity has been assessed by comparing the reweighted results with those of a direct refit, finding good agreement in all cases. We note that some authors have advocated a different functional form for the weights than that of Eq. (109), see for instance [345, 74, 346]. However, so far only the definition of Eq. (109) has been explicitly demonstrated to lead to equivalent results in comparison to

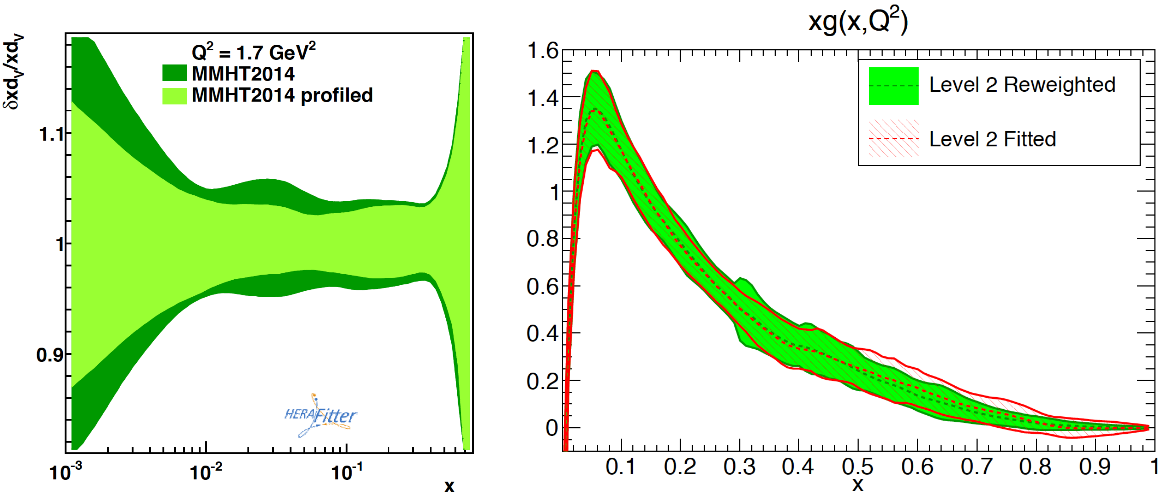


Figure 33: Left plot: the impact of the Tevatron W and Z data on the MMHT2014 NLO fit, estimated by the Hessian profiling method of Ref. [340]. Right plot: the gluon PDF in the NNPDF3.0 closure tests, quantifying the impact of the collider inclusive jet data, and comparing the results of the Bayesian reweighting with those of a direct refit.

a direct refit, within the reweighting accuracy. It is also worth mentioning that the derivation of Eq. (109) assumes a native Monte Carlo PDF set, and that it does not necessarily apply to MC sets that are obtained from native Hessian sets using the conversion method of [332], see for example the discussions in [344, 31]

In order to provide an illustrative example of the Bayesian reweighting method, in Fig. 33 we show the gluon PDF in the NNPDF3.0 closure tests [17], estimating the impact of the collider inclusive jet data and comparing the results of the Bayesian reweighting with those of a direct refit. In this study, the prior was a set of $N_{\rm rep} = 1000$ replicas obtained with NNPDF2.3-like dataset but without any collider inclusive jet production data included. The pseudo-data were generated using the MSTW08 NLO set, though similar results were obtained using other priors. We observe that there is good agreement between the approximate Bayesian reweighting method and the exact refit results.

4.6.2. Hessian profiling

For a Hessian PDF set, the so-called profiling technique provides a closely related method to approximately quantity the impact of a new experimental measurement. This method is based on the minimization of a χ^2 estimator that compares the theoretical predictions obtained with a given input Hessian PDF set with the new experimental measurements. This estimator takes into account both the experimental uncertainties and the effects from the PDF variations (as encoded by the Hessian eigenvectors) and is defined as follows:

$$\chi^2\left(\beta_{\rm exp},\beta_{\rm th}\right) = \frac{1}{\Delta_i^2}\sum_{i=1}^{N_{\rm dat}}\left(\sigma_i^{\rm exp} + \sum_j \Gamma_{ij}^{\rm exp}\beta_{j,\rm exp} - \sigma_i^{\rm th} + \sum_k \Gamma_{ik}^{\rm th}\beta_{k,\rm exp}\right)^2 + \sum_j \beta_{j,\rm exp}^2 + \sum_k \beta_{k,\rm th}^2\,, \qquad (112)$$

where $\beta_{j,\rm exp}$ are the nuisance parameters corresponding to the set of fully correlated experimental systematic uncertainties, and $\beta_{k,\rm th}$ are the nuisance parameters corresponding to the PDF Hessian eigenvectors. Δ_i is the total experimental uncorrelated uncertainty, and $N_{\rm dat}$ is the number of data points of the measurement which is being added into the PDF fit. Finally, the matrices $\Gamma_{ij}^{\rm exp}$ and $\Gamma_{ik}^{\rm th}$ encode the effects of the corresponding nuisance parameters on the experimental data and on the theory predictions, respectively.

Upon minimization of the χ^2 estimator Eq. (112), the corresponding values of the theoretical nuisance parameters, denoted by $\beta^{\rm min}_{k,\rm th}$, can be interpreted as leading to PDFs that have been optimized (hence the name "profiled") to describe this new specific measurement. Note also that in general the profiling will modify both the central value and the total PDF uncertainty. For example, the new measurement might reduce the allowed range of variation of a given eigenvector, if the original variation leads to large values of Eq. (112).

As in the case of the Bayesian reweighting method, there are a number of limitations of the the Hessian profiling method that limit the cases where it can be used to replace a complete refit. First of all, it assumes that the optimal PDF parametrization will not be modified by the addition of the new experiment. It is well known that this condition does not necessarily holds, for instance new experiments might require the use of more flexible input PDF parametrizations in order to achieve an optimal description, and this cannot be accounted for with the profiling method. Secondly, the standard version of the Hessian profiling method assumes that the PDF uncertainties are defined by the $\Delta\chi^2 = 1$ criterion, which is generally not the case for global Hessian PDF fits, see Sect. 4.3. For this reason, the impact of the data as estimated by Hessian profiling will in general differ in comparison to the result of a full refit. However, this limitation can be eliminated by using a tolerance criterion that mimics the one used in the prior Hessian PDF set, see for example Ref. [31].

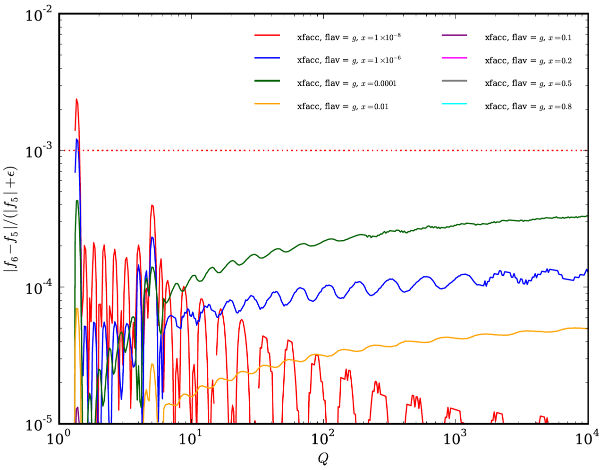
As an example of the applications of the Hessian profiling method, in Fig. 33 we show the the impact of the Tevatron W and Z data on the MMHT2014 NLO set from Ref. [340], estimated by Hessian profiling. An important point to emphasise is that this exercise was performed using completely public tools, in this case the MMHT2014 `LHAPDF` grids, and the experimental information of the Tevatron W and Z measurements, without any additional input from the authors of the original MMHT2014 analysis.

4.7. Delivery

The final part of a PDF fit is of course to make it available to any potential user. In the beginning this was achieved by means of (x, Q) interpolation tables and the corresponding driver codes which were specific to each PDF group. This was however far from optimal since standardization was very difficult, with programs requiring PDFs as input having to be adapted each time a new PDF set was released. A first step towards PDF access standardization was achieved with the release in 1993 of `PDFLIB` [347] as part of the CERN Program Library software. This allowed a unique interface for calling PDFs to be used without the need to add external files on a case by case basis. In addition to the PDFs, the value of $\alpha_s(m_Z)$ used in each specific fit could be accessed.

The next step in this standardisation process came with the release in 2005 of LHAPDF, the Les Houches Accord on PDFs [348, 349], which was developed as a functional replacement for `PDFLIB`. In order to ensure backwards compatibility, LHAPDF included LHAPDF glue, a PDFLIB-like interface. One of the main motivations to release LHAPDF was the realization that dealing with a large number of error PDF sets, that had then recently became available, was extremely cumbersome with `PDFLIB`. In particular, LHAPDF was organized around the concept of *PDF set*, which was constituted by the central (average) member as well as the corresponding error PDF sets. As in the case of `PDFLIB`, LHAPDF was written in Fortran 77, although later a C/C++ interface was also developed.

While the Fortran incarnation of LHAPDF was very popular and widely used, at some point its further development became very challenging in particular due to the intrinsic limitations of Fortran 77 as its native language. In particular, since Fortran 77 required to allocate space for all available PDFs at compilation time, the memory footprint eventually become impossible to handle and LHAPDF v5.9.1 was the last release. To overcome these limitations, a complete rewriting of LHAPDF from scratch in `C++` was completed



Process/PDF	t_5	t_6	t_5/t_6
Cross-section integrations, 1M phase space points			
CT10			
$pp \to jj$	23'10"	9'17"	2.5
$pp \to \ell\ell$	4'12"	2'02"	2.1
$pp \to H$ (ggF)	0'20"	0'15"	1.3
NNPDF23nlo			
$pp \to jj$	54'40"	9'28"	5.8
$pp \to \ell\ell$	8'06"	2'33"	3.2
$pp \to H$ (ggF)	0'25"	0'11"	2.3
CKKW event generation, 100k $pp \to \leq 4$ jet events			
CT10			
Weighted	43'02"	35'47"	1.2
Unweighted	5h04'39"	4h30'26"	1.1
NNPDF23nlo			
Weighted	47'47"	27'20"	1.7
Unweighted	6h44'47"	4h48'26"	1.4

Figure 34: Left plot: the relative difference between LHAPDF v5 and v6 for $g(x, Q)$ for different values of x as a function of Q, using CT10 as input PDF. Right plot: the timing improvement in v6 as compared to v5, t_6/t_5, for a cross-section integration of 1M phase space points with `Sherpa` and for CKKW event generation of 100k $pp \to 4$ jet events.

in 2014, dubbed LHAPDF6 [350]. In addition to reducing static memory requirements by orders of magnitude, this `C++` incarnation of LHAPDF offered improved CPU performance and improved interpolation and extrapolation functionalities. Moreover, its cascading meta-data system ensures that software releases are completely decoupled from the availability of novel PDF sets. To ensure backwards compatibility, Fortran 77 interfaces are also provided.

In terms of interpolation accuracy, LHAPDF6 reproduces the v5 results down to residual differences of at most 0.1%. This is illustrated in Fig. 34, where we show the relative difference between LHAPDF v5 and v6 for $g(x, Q)$ for different values of x as a function of Q, using CT10 as input PDF. In Ref. [350] it was also shown that LHAPDF6 improves also the CPU timings as compared to v5 by a factor between 2 and 6. This is seen in the right table in Fig. 34, which represents the timing improvements in v6 as compared to v5, t_6/t_5, for a cross-section integration of 1M phase space points with `Sherpa` [351] and for CKKW event generation of 100k $pp \to 4$ jet events.

Currently LHAPDF6 has established itself as the almost universal software to access PDFs. Its current version is 6.1 and more than 700 PDF sets can be accessed. In addition to unpolarized parton distributions, the flexibility of the LHAPDF6 framework makes it suitable to release other types of non-perturbative QCD objects, and indeed also polarized PDFs and nuclear PDFs are available.

5. PDF analyses: state of the art

Here we review the latest developments from the main PDF fitting groups. The comparison among them is left for the next section.

5.1. CT

The CTEQ-TEA global analysis was established by Wu-Ki Tung et. al. in the early 1990s with the CTEQ1 PDFs [65]. The most recent release of general purpose PDFs from the collaboration are the CT14 PDF sets [18], which include the nominal sets as well as alternative sets with different choices of α_s and the maximum number of active flavors. The PDFs are parameterised at the starting scale $Q_0 = 1.3\,\text{GeV}$ using the form (69) described in Sect. 4.1.1 In pre–CT14 analyses the interpolating function I_f was chosen as an exponential of a polynomial in x or $\sqrt{x}$, such that positivity conditions on the PDFs at the initial scale were enforced. In CT14 analysis an improved parametrization choice was introduced, with for example for the u-valence

$$P_{u_v} = d_0 p_0(y) + d_1 p_1(y) + d_2 p_2(y) + d_3 p_3(y) + d_4 p_4(y), \tag{113}$$

where $y = \sqrt{x}$ and p_n are the fourth order Bernstein polynomials, given by

$$p_0(y) = (1-y)^4,\ \ p_1(y) = 4y(1-y)^3,\ \ p_2(y) = 6y^2(1-y)^2,\ \ p_3(y) = 4y^3(1-y),\ \ p_4(y) = y^4. \tag{114}$$

Namely, the interpolating function is chosen as a fourth-order polynomial in y with an expansion in the basis of Bernstein polynomials. As discussed in Sect. 4.1.1, this greatly increases the stability of the fit within the Hessian approach. In the CT14 case the positivity of PDFs at $Q_0 = 1.3\,\text{GeV}$ in fact emerges automatically as a consequence of the fit to data. The CT14 PDFs have a total number of free parameters of 28 in the PDF parametrization; using a more flexible parametrizations, by adding higher-order polynomials, is found to have a small effect on both the best–fit and the estimated PDF uncertainties in the region that is well constrained by data.

The CT14 global analysis includes a variety of experimental data. The majority comes from the inclusive DIS and semi–inclusive DIS measurements of the structure functions and the reduced cross section measurements from fixed-target experiments (BCDMS [352, 353], NMC [110], CCFR [115, 114, 118], NuTeV [119], CDHSW [354]) or HERA experiments [355, 126, 85, 123]. A Q cut of 2 GeV and W cut of 3.5 GeV are adopted in the selection of DIS data to minimize non–perturbative effects from either nuclear corrections or higher–twists corrections. Thus no further nuclear or higher twists corrections are included in theory predictions in CT14 except for those already applied in the unfolding of experimental data. For the NC DIS process, the CT14 analysis ultilises a treatment of heavy-quark mass effects up to NNLO, through a type of GM-VFN scheme, known as S-ACOT-χ [103]. For CC DIS, the theory is only implement at NLO, which is judged to be sufficient given the relatively small number of data points and their large experimental errors. Drell-Yan production data from fixed-target experiments (E605 [356], E866 [176]) and W/Z boson production data from Tevatron [357, 358, 359, 360, 179] including the new D0 electron charge asymmetry data [181], are also fit.

The Tevatron W, Z data provide further discriminations on quark flavors in large-x region, with the W asymmetry data probing the average slope of d/u ratio at $x \gtrsim 0.1$, see Section 3.4 for more discussion. NNLO predictions from ResBos [219, 361, 226, 362] are used for the W/Z boson production data, with a p_T cut imposed on the charged leptons, and incorporating soft gluon resummation effects at small p_T of the vector boson. These resummed predictions provide a better description of the p_T spectrum of the charged leptons. The updated D0 electron charge asymmetry data [181] shows a large impact on the d/u PDF ratio at large-x comparing to CT10 and CT10W [363, 81]. In CT10 fits the D0 lepton charge asymmetry data resulted in larger asymptotic value of d/u though tensions were found between different subsets of the data or the D0 data and other DIS experiments. As shown in Fig. 35, for CT14 the updated D0 electron charge asymmetry data shows better agreements with other data sets in the global analysis and drives the d/u ratio to a lower value close to CTEQ6.6 [364] at large-x. The d/u ratio in CT14 also shows good agreement with the extraction from CJ12 [365], which is based on independent large–x and low–W DIS data, including power corrections and deuteron corrections. Similar data on W/Z boson production from LHC Run I are also included from the ATLAS [185], CMS [191, 190] and LHCb [366] experiments, which further extend the coverage to the intermediate and small x region. In addition, single inclusive jet production from the Tevatron [143, 144] and LHC [141, 149] are fit, providing the dominant constraint on the gluon PDF at large x, with the latter data also extending the coverage to the intermediate x region. For inclusive jet production at hadron colliders only the NLO predictions were available at the time of the CT14 fit, and therefore this is used in the NNLO fit. This will be updated with the recent NNLO calculations [161] in future CTEQ-TEA analysis.

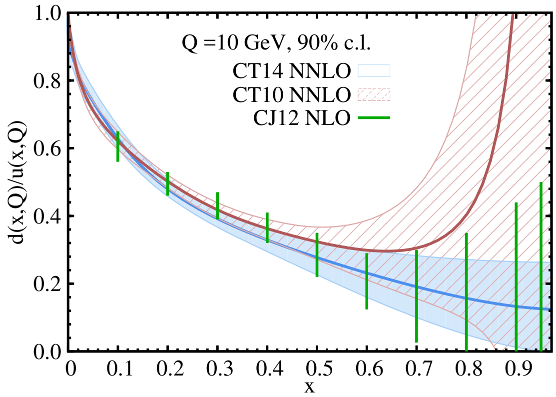


Figure 35: A comparison of 90% C.L. uncertainties on the d/u PDF ratio at $Q = 10$ GeV for CT14 NNLO (solid blue) and CT10 NNLO (dashed red), and CJ12 (green lines) error ensembles [18].

The CTEQ-TEA group uses the Hessian method with certain tolerance conditions for the nominal fits in the determination of PDF uncertainties at 90% c.l.. That is supplemented with a Lagrange multiplier (LM) scan for certain important observables or for PDFs in region poorly constrained by data. In pre-CT10 analyses it was found that within a global χ^2 tolerance of $\Delta\chi^2 = 100$ (for more than 2000 data points) the fits agree with all experiments at 90% c.l.. In latter CTEQ-TEA analyses a more efficient dynamic tolerance criteria is adopted to account for agreement with individual data set. It is constructed from an equivalent Gaussian variable, e.g.,

$$S_n = \sqrt{2\chi^2(N_n)} - \sqrt{2N_n - 1}, \tag{115}$$

where N_n is the total number of data points in data set n and $\chi^2(N_n)$ represents the χ^2 of the fit to that data set. S_n follows a normal distribution given the number of data points is large enough. Thus a value of S_n greater than 1.3 will be excluded at 90% c.l.. We add a second layer of penalty to the global χ^2 when determing the boundaries of confidence intervals, called a Tier-2 penalty,

$$P = \sum_{n=1}^{N_{exp}} (S_n/S_{n,best})^{16}, \tag{116}$$

where the sum runs over all data sets included and we normalize S_n to its value in the best fit to account for poor fit to certain experiments. The power of 16 is introduced so that the penalty will reach the tolerance of 100 as soon as any data set shows disagreement at 90% c.l.. The tolerance criteria then changes to $\Delta\chi^2$+P=100. Fig. 36 shows the distribution of $S_{n,best}$ for all 33 experiments included in CT14 analysis. The distribution is wider than a normal distribution, indicating the presence of disagreement, or tensions, between some of the included experiments.

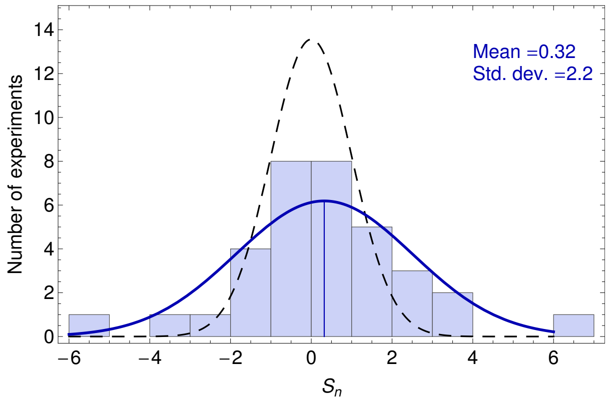


Figure 36: Best-fit values of the equivalent Gaussian variables for 33 experiments in the CT14 NNLO global analysis [18].

With the best-fit and $2D$ error sets, the asymmetric errors for any QCD observable X can be calculated through the master formula

$$(\delta X)_+ = \sqrt{\sum_{i=1}^{D}[max(X_{+i} - X_0, X_{-i} - X_0, 0)]^2}, \qquad (\delta X)_- = -\sqrt{\sum_{i=1}^{D}[max(X_0 - X_{+i}, X_0 - X_{-i}, 0)]^2}, \quad (117)$$

where X_0 is the prediction from central set, X_{+i} and X_{-i} are from two error sets in the direction of i-th eigenvector. The errors can be scaled down to 68% c.l. with a factor of 1.64 assuming Gaussian distributions. As mentioned earlier the CTEQ-TEA analysis also uses the Lagrange multiplier method [324] to crosscheck the error estimation from nominal Hessian sets. In the CT14 analysis, Lagrange multiplier scans have been performed for the cross sections of Higgs boson production via gluon fusion and of the top quark pair production at the LHC. In such scans the best-fits and the associated χ^2 are found for each fixed value of the observable studied. Then the PDF uncertainties on the observable are determined from the χ^2 profile obtained using the same tolerance criteria as in the Hessian method. Fig. 37 shows the good agreement of the 90% C.L. uncertainties for the Higgs cross sections from the CT14 Hessian PDFs and the CT14 LM scans. The latter can be read off from the intersection of the horizontal line $\Delta\chi^2 = 100$ and the various curves. The LM method does not rely on the linear approximation, and therefore serves as a robust check of the Hessian results.

There are a few other specialities of the CTEQ-TEA global analysis. The CTEQ-TEA analyses uses the world average of strong coupling constant $\alpha_s(M_Z)$ as an input. Usually the nominal fit is performed with $\alpha_s(M_Z) = 0.118$ at both NLO and NNLO, but additional fits with alternative α_s choices are also provided. The fit itself provides a much weaker constraint on α_s than the world average. Similarly the pole mass of charm quark and bottom quark are chosen to be close to the world average values, with $m_c = 1.3\,\text{GeV}$ and

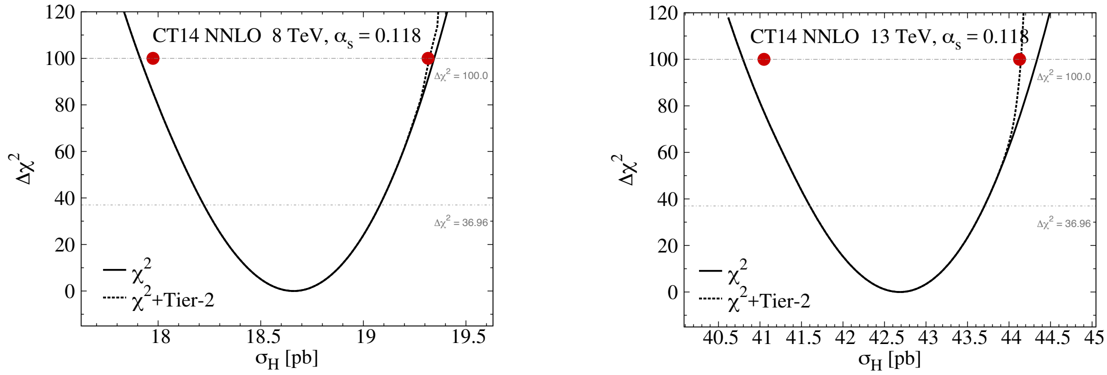


Figure 37: Dependence of the increase in χ^2 in the constrained CT14 fit on the expected cross section σ_H at the LHC 8 and 13 TeV [18], for $\alpha_s(M_Z)$=0.118. The solid and dashed curves are for the constrained fits without and with the Tier-2 penalties, respectively. The red dots correspond to the upper and lower 90% C.L. limits calculated by the Hessian method.

m_b = 4.75 GeV. The CTEQ-TEA group also provides specialized fits with non-perturbative charm quark PDFs. A sea-like or valence-like charm distribution is added to the nominal parametrization and then fitted to data. Limits on the momentum fraction carried by the fitted charm at the initial scale Q_0 = 1.3 GeV are then derived. In the most recent CT14 analysis, the limits are 1.6% for the Sea-like model and 2.1% for the BHPS model both at 90% c.l. [367]. There are also CT14 QED PDFs [368] based on a radiative ansatz for the inelastic component of the photon PDFs. The 90% C.L. limit on the momentum fraction of the proton carried by the photon is 0.11% at Q_0 = 1.3 GeV, as derived from fit to the ZEUS measurement on isolated photon production [369].

5.2. MMHT

The MMHT14 PDFs [19] are the successor to the MSTW08 [80] set, which derives from the earlier MRST and MRS studies. The first NLO fit [61] to DIS data was performed in the late 80s, while in the mid 90s the MRS(A) [67] fit was released, including data from HERA and the Tevatron for the first time. This corresponded to a truly global analysis, fitting to fixed target, DIS and hadroproduction data to constrain the PDFs as precisely as possible. Subsequent releases have all built on this approach, but with significant advances achieved over the years due to improvements in both theory and experiment. The MRST98 release [370] was the first set to include a full treatment of heavy flavours within the GM–VFNS developed in [371], and discussed further in Sect. 2.5. This was motivated by the new HERA measurements of the charm structure function, which demonstrated the importance of a consistent treatment of charm production and low and high scales; indeed, the introduction of this flavour scheme resulted in an improved description of such data. The MRST02 release [77] included a full treatment of PDF errors for the first time, described further below, while the MRST04 [372] set went to NNLO for the first time.

These elements were all incorporated in the major MSTW08 [80] release. This presented a global fit to a range of DIS data from HERA and fixed proton and nuclear targets, fixed target Drell–Yan and dimuon production and W, Z and jet production at the Tevatron, with $O(2500)$ data points in total. Fits were performed up to NNLO in the strong coupling, with an improved dynamical error treatment, and with an up to date heavy flavour scheme applied. This aimed to provide a PDF set for use at the LHC, which began operation soon after the release, and was subsequently very widely used in LHC phenomenological studies and experimental analyses. This fit was updated in the latest MMHT14 [19] set, which includes a number of theoretical and experimental updates. In particular for the first time LHC data on W, Z, $t\bar{t}$ and jet production are included, as well as updated HERA data on the charged, neutral, charm and longitudinal structure functions, and updated Tevatron W and Z measurements. As in earlier fits, for DIS data a Q^2 cut of 2 GeV^2 and W^2 cut of 15 GeV^2 is imposed to avoid sensitivity to higher twist corrections.

In the case of Tevatron jet production, in the absence of a full NNLO calculation at the time, an approximation to the NNLO corrections based on the threshold corrections of [167] was applied in the NNLO MSTW08 fit, with the judgement being made that the difference between this and the full NNLO result would be expected to be smaller than the systematic uncertainties on the data, which itself provided the only direct constraint on the gluon at high x. At the LHC much of the jet data are quite far from threshold, while those that do not probe a kinematically similar region to the Tevatron data, and so at NNLO these are not included in the MMHT14 fit. For the $t\bar{t}$ data the top mass is allowed to be determined from the fit, with the pole mass value of $m_t = 172.5 \pm 1$ GeV taken as an input. This gives a value at NNLO that is consistent with the world average, while at NLO it is somewhat lower.

The MSTW PDFs were parameterised in terms of simple polynomials in x, with 29 free parameters. However, in [373] it was shown that this parameterisation was not sufficiently adaptive to describe the Tevatron and LHC W asymmetry data. In particular, it was necessary to introduce a more flexible basis for the interpolating function described in Sect. 4.1.1, with

$$I_f(x) = \sum_i^n \alpha_{f,i} T_i(y(x)) \; , \tag{118}$$

where T_i is a Chebyshev polynomial of order i and $y(x) = 1 - 2\sqrt{x}$ is chosen so as to sample a wide range of x, and has the additional advantage that this provides a half–integer separation in powers of x, as expected on Regge theory grounds. In order to determine how many parameters n are needed, in [373] pseudo–data points with a constant percentage error were generated for the required distributions, in terms of a very large order polynomial with additional smoothness constraints applied. The fractional deviation from the true PDF, as well as the decrease in χ^2, were then determined as the number of parameters were increased, until no further significant improvement was observed and the level of agreement was well below the PDF uncertainty for the set. In this way $n = 4$ was arrived at as a good choice with which to parameterise the u_V, d_V, $s + \overline{s}$ and light quark sea S distributions. Fitting to the same MSTW08 data set, these resulted in some improvement in the fit quality, but with the only significant change in the PDF being in the u_V at lower x. This was found to lie outside the previous PDF uncertainty band, and the additional flexibility provided a greatly improved description of W asymmetry data.

In the MMHT14 set, this Chebyshev parameterisation is used at $Q_0^2 = 1\,\mathrm{GeV}^2$ for the u_V, d_V, $s + \overline{s}$ and light quark sea S distributions, while for the gluon a term with $n = 2$ Chebyshevs is included, but with a second term still present, as in MSTW08, which has a different low x power and provides the additional flexibility at low x that is required by the HERA data; this has the effect that the gluon at NLO and higher becomes negative at low x and Q^2. Standard polynomial parameterisations are taken for the less constrained $s - \overline{s}$ and $\overline{d} - \overline{u}$ distributions, although as the data becomes more precise we can expect this to change.

A further improvement described in [373] that is included in the MMHT14 set is in the treatment of the non–perturbative corrections that should in general be applied when considering DIS data on deuteron targets, to account for the binding of the proton and neutron within the deuteron. While in MSTW08 and earlier fits, a fixed shadowing correction at small x was applied, a more flexible approach is now taken. In particular the deuteron corrections are freely parameterised in terms of a function $c(x)$, which is determined along with its corresponding uncertainties from the PDF fit. This resulted in a significantly

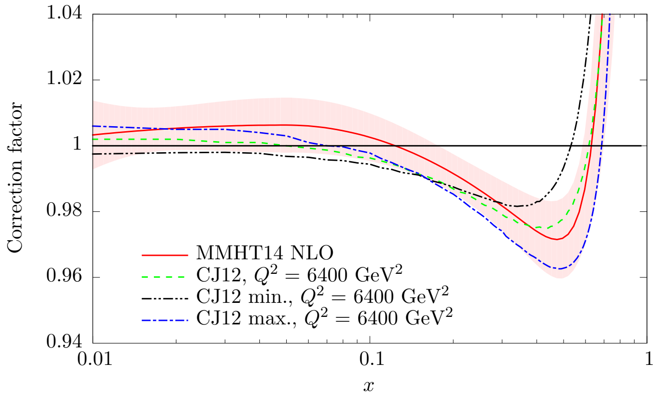


Figure 38: Fitted MMHT14 deuteron correction factors with uncertainty, compared to the CJ12 [365] predictions. Taken from [19].

improved description of the BCDMS deuteron structure function data, the E866 Drell–Yan asymmetry and the Tevatron lepton asymmetry data, with some significant changes in d_V. The result of the MMHT14 fit is shown in Fig. 38 and compared against different model predictions used in the CJ12 [365] analysis. Interestingly, very good agreement is found with the CJ12mid prediction, demonstrating the power of global PDF fits to extract additional physical information beyond the PDFs themselves.

In MMHT14, the Hessian approach is applied to calculate the PDF errors, with the ‘dynamical’ tolerance criteria described in Sect. 4.3.1 taken. For MMHT14 the 68% uncertainties are calculated using this procedure. In the fit there are 37 free PDF parameters in total, however in the error determination certain parameter directions are found to be largely degenerate, leading to departures from quadratic χ^2 behaviour. This is corrected by fixing some parameters when calculating the error eigenvectors, reducing the number of 25, that is 50 directions.

Other theoretical updates in the MMHT14 include the treatment of the $D \to \mu$ branching ratio, which is required in the fit to dimuon production in DIS. This is now determined from the fit but with the measurement of [374], which is not determined from dimuon production data, included as a data point. The result is somewhat lower than that taken in MSTW08, corresponding to a larger strangeness, but the most dramatic effect is that the $\sim$ 10% uncertainty on the branching ratio allows for a much larger strangeness uncertainty when fitting to the same data. Other smaller improvements include an updated treatment of nuclear corrections and a multiplicative, rather than additive, treatment of systematic uncertainties where appropriate.

In contrast to other global fits, where it is taken as an input, in MMHT the value of the strong coupling is allowed to be determined by the fit, it being argued that valuable information can be provided from global PDF fits about this object. This in addition serves as a consistency test on the overall fit; if the extracted value is in strong tension with the world average then this might indicate that further work is needed. In the fit the preferred values at NLO and NNLO are indeed found to be consistent with the world average, and including this as an additional data point is not found to affect the fit significantly. In particular, the detailed study of [326] found best fit values of $\alpha_S(M_Z^2) = 0.1201 \pm 0.0015$ at NLO and $\alpha_S(M_Z^2) = 0.1172 \pm 0.0013$, to be compared with the world average value of $\alpha_S(M_Z^2) = 0.1181 \pm 0.0013$. The NNLO χ^2 profile for $\alpha_S(M_Z^2)$ and the corresponding individual constraints from the most constraining data sets are shown in Fig. 39. The corresponding PDF sets for a range of α_S values, from 0.108 to 0.128 in steps of 0.001, are publicly available.

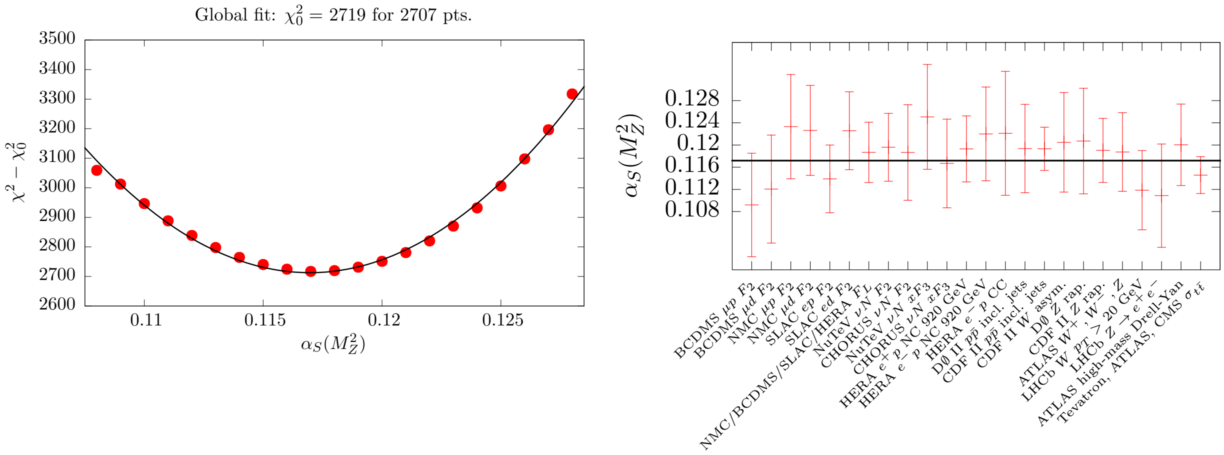


Figure 39: (Left) Global χ^2 values as a function of $\alpha_S(M_Z^2)$. (Right) Best fit $\alpha_S(M_Z^2)$ together with the upper and lower 1σ constraints from the most constraining data sets. Both figures correspond to NNLO fits and are taken from [326].

For the heavy quarks, the GM-VFNS scheme is taken, applying the 'optimal' scheme of [100] which improves the smoothness of the transition region where the number of active flavours is increased by one. The charm and bottom pole masses are fixed to $m_c(m_b) = 1.4(4.75)$ GeV, but a detailed study of the heavy quark mass dependence was performed in [329]. In particular, when m_c and m_b were taken as free parameters in the fit, the preferred values were found to be somewhat lower than these defaults, possibly suggesting a $\overline{MS}$ scheme may be more appropriate, although the impact of such a choice is expected to be fairly small. Sets with a range of charm and bottom masses, as well as with 3 or 4 active flavours, are made available.

A subsequent study within the MMHT framework on the PDF impact of the final HERA I + II combined data [21], which was released following MMHT14, was performed in [375]. This was found to lead to some reduction in the PDF uncertainties, principally in the gluon, with little change in the central values. It was therefore decided not to release an updated set but rather to wait until more precise and varied LHC data became available, as well as theoretical calculations such as NNLO jet production. Subsequent work towards a new PDF set has been presented in for example [376], where a first fit at NNLO to jet data is presented, while the impact of new LHC data in the fit is seen to be significant. In addition, the first steps towards the inclusion of the photon PDF within the MMHT framework are presented, see Sect. 7 for more discussion. A further public release is therefore anticipated in the near future.

5.3. NNPDF

As discussed in Sect. 4, the NNPDF fitting methodology is based on the combination of three main components: i) the use of artificial neural networks as universal unbiased interpolants, ii) the Monte Carlo method to estimate and propagate PDF uncertainties, and iii) Genetic Algorithms minimization for the training of the neural networks. Here we review the main developments in the NNPDF family of PDF fits.

The NNPDF methodology was first presented in [87], where it was used to produce a neural network based determination of the proton, deuteron and non-singlet DIS structure functions from the fixed target data from NMC and BCDMS. As a first phenomenological application, this determination was used to extract the strong coupling constant $\alpha_S(m_Z)$ from scaling violations of truncated moments of structure functions [377]. This analysis was subsequently extended [321] to include as well the F_2^p measurements from the H1 and ZEUS experiments at the HERA collider. Note that a determination of structure functions is purely data-driven, with no theoretical input required at this point.

When moving from fitting structure functions to PDFs, there are a number of simplifications, for instance one needs to fit only a 1D function $q_i(x, Q_0)$ as opposed to a 2D function $F_2^p(x, Q^2)$, but also technical complications, the most important one being able to compute DIS structure functions starting from the neural-network based parametrization of $q_i(x, Q_0)$. First of all, the usual ANN training algorithm of back-propagation cannot be used in this case, due to the convolution of the PDFs with the DGLAP evolution kernels and the DIS coefficients functions. To overcome this limitation, it was demonstrated how Genetic Algorithms can be efficiently used for ANN training under a non-trivial mapping between the latter and the experimental data, and used to extract the QCD vacuum condensates from hadronic tau decay data [378]. An efficient method to solve the DGLAP evolution equations in N-space was also developed, called the `Fast Kernel` method. With these ingredients at hand, it became possible for the first time to apply the NNPDF methodology to a determination of parton distributions, starting from a fit of the non-singlet combination $q_{NS}(x, Q_0)$ [320] and them moving to a first full-fledged NLO PDF fit based on neutral-current DIS structure function data [379] in the NNPDF1.0 analysis.

Subsequently, the global NNPDF fits were improved both by adding new experimental data, updating the theoretical calculations and/or refining the fitting methodology. The NNPDF1.2 analysis [380] relaxed the previous assumption that the strange sea was proportional to the light quark sea, $s = \bar{s} = \kappa\left(\bar{u} + \bar{d}\right)$, and parametrized both s^+ and s^- using neural networks, exploiting the constraints from the NuTeV dimuon charged-current neutrino scattering data. Two important phenomenological consequences of this analysis were, first of all, the demonstration that the PDF uncertainties associated with s^- were enough to complete wipe out the NuTeV anomaly [381] in the determination of the weak mixing angle $\sin^2\theta_W$; and second, a direct extraction of the CKM matrix element V_{cs} with a precision compatible with that of the PDG average.

In 2010, the NNPDF2.0 set was released [88], which constituted the first truly global PDF fit from the NNPDF collaboration. In addition to the NC and CC DIS structure function data included in previous releases, NNPDF2.0 included in addition fixed-target Drell-Yan cross-sections from the Fermilab E605 and E866 experiments, inclusive jet production measurements from CDF and D0 and the Tevatron as well as the differential rapidity distributions of the Z boson also from the Tevatron. From the theoretical point of view, NNPDF2.0 was still based on the zero-mass VFN scheme, and thus charm and bottom structure function data from HERA were not included. A good overall description of all experiments in the global fit was found. NNPDF2.0 was also the first PDF set to include the recently release HERA combination of H1 and ZEUS structure function data for the Run I data period [85]. The NNPDF2.0 was one of the sets included in the PDF4LHC 2011 recommendation [92] for the usage of PDFs at the LHC.

While NNPDF2.0 demonstrated that the NNPDF methodology could be successfully applied to a global determination of parton distributions, there were still a number of important limitations from the theoretical point of view. First, the use of a ZM-VFN scheme neglected heavy quark mass effects in the DIS structure functions, which were known to be important for the description of the low-x, low-Q^2 HERA data. Second, all NNPDF fits so far were based on NLO theory, and NNLO accuracy was essential to match the corresponding precision of important partonic hard-scattering cross-sections such as Higgs production in gluon fusion. The first of these theory limitations was removed with the release of NNPDF2.1 [382], which was based on the FONLL general-mass VFN for the calculation of DIS structure functions, which allowed the HERA charm and bottom structure functions data to be fit. This analysis also showed that the impact of heavy quark mass effects was less drastic then previously reported, with the cross-section predictions between NNPDF2.0 and 2.1 typically agreeing at the one-sigma level. The NNPDF2.1 fit was also used to produce a determination of the strong coupling $\alpha_s(m_Z)$ from the global dataset [383].

The second of these theoretical limitations was removed shortly after, with the release of a NNLO version of NNPDF2.1 [89]. This PDF set was based on the same dataset as its NLO counterpart, but with

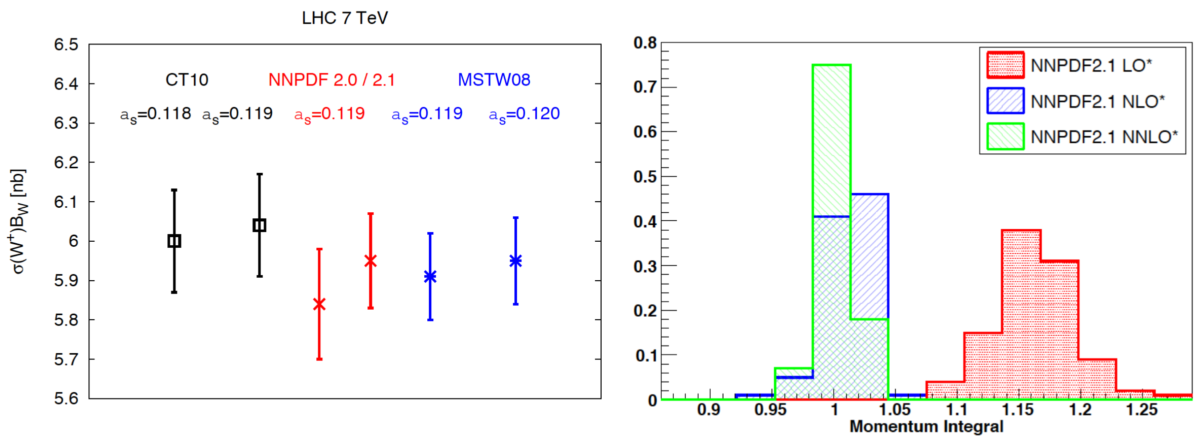


Figure 40: Left plot: comparison between the NNPDF2.0 and 2.1 predictions for the inclusive $W+$ production cross-section at the LHC 7 TeV, which illustrates the phenomenological impact of heavy quark mass effects. Right plot: the distribution of the momentum integral Eq. (119) among the MC replicas for the variants of the NNPDF2.1 LO, NLO and NNLO fits that do not impose explicitly the momentum sum rule.

the DIS and hadronic cross-sections computed at NNLO, in the former case using the FONLL-C GM-VFN scheme. In the same publication, the first NNPDF LO sets were also presented. The availability of NNPDF2.1 fits at LO, NLO and NNLO allowed a systematic study of the perturbative convergence of the global fit, finding in particular reasonable agreement at the one-sigma level between the NLO and NNLO versions. The consistency of the global QCD analysis framework was also tested by performing fits without imposing the momentum sum rule and then verifying a posteriori that the global fit result was consistent with the QCD expectation, finding indeed that at NNLO

$$[M] \equiv \int_0^1 dx \left(g(x, Q^2) + \Sigma(x, Q^2) \right) = 1.002 \pm 0.014 \,. \tag{119}$$

The NNPDF2.1 NNLO analysis was also used to perform a determination of the strong coupling constant [327], finding a value $\alpha_s(m_Z) = 0.1173 \pm 0.0007^{\rm stat} \pm 0.0009^{\rm pert}$, a result which is still included in the PDG global average of α_s [325].

The main advantage of the reweighting method is allowing to gauge the impact of new data without having to do a PDF fit and based only on public information, so that it was not restricted to PDF fitters anymore.

In the early 90s the availability of the HERA structure function measurements became a game–changer for global fits, and from 2010 the LHC experiments started producing a wealth of PDF-sensitive information, which promised to impact global fits in a similar significant way. With this motivation, in 2012 the NNPDF2.3 set was released [90], and was the first PDF fit to include LHC data, in particular electroweak gauge boson production from ATLAS, CMS and LHCb as well as jet production from ATLAS. As with all subsequent releases, NNPDF2.3 was available at LO, NLO and NNLO and was based on the FONLL general mass scheme. The NNPDF2.3 set became the baseline PDF set in several popular Monte Carlo event generators, such as `Pythia8` and `aMC@NLO`.

Following the release of NNPDF2.3, it was realized that the increase in complexity required to include the many new experiments that were either available or about to be released could not be satisfactory tackled

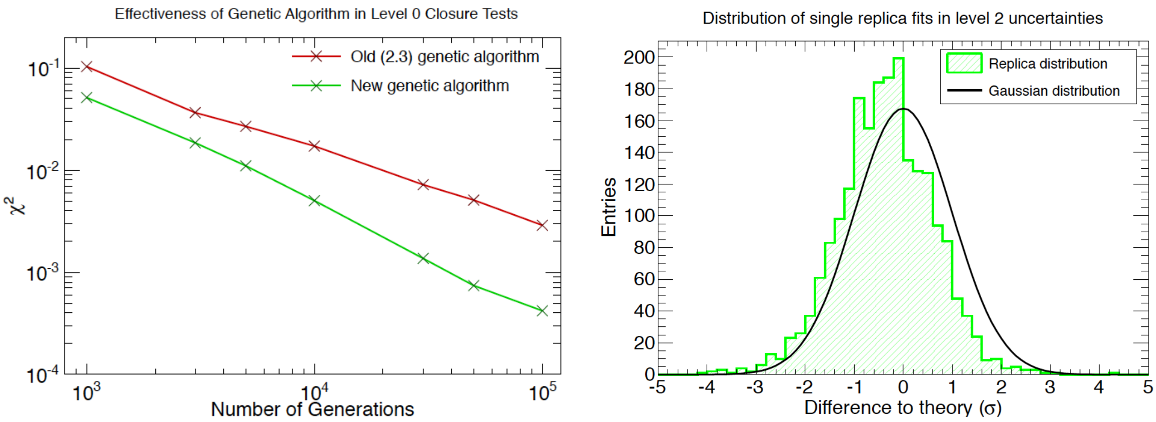


Figure 41: Representative results of the closure tests presented in the NNPDF3.0 analysis. Left plot: in a level 0 closure tests, where the pseudo-data is generated without any statistical fluctuations, the χ^2 should decrease monotonically as a function of the number of GA iterations, down to arbitrarily small values. Right plot: the distribution of the difference between theory and data in units of the error of the latter among each of the Monte Carlo replicas. This distribution is consistent with the Gaussian predicted by statistics.

with the current, `FORTRAN77`-based code. With this motivation, a complete rewriting of the NNPDF global analysis framework into `C++` and `Python` was undertaken, a two-year long effort that culminated with the release of the NNPDF3.0 set [17]. In addition to including many new LHC experiments on jets, vector boson production, W+charm and top production, the main result of NNPDF3.0 was the systematic validation of the complete fitting methodology based on statistically robust closure tests. In these closure tests, pseudo-data was generated based on some "truth" PDFs, and then a PDF fit was performed: if the resulting PDF central values and uncertainties were consistent with the (known) input PDFs, the the closure test can be considered successful. In Fig. 41 we show some representative results of the closure tests presented in the NNPDF3.0 analysis. In the left plot we show the results of a level 0 closure test, where the pseudo-data is generated without any statistical fluctuations, the χ^2 should decrease monotonically as a function of the number of GA iterations, down to arbitrarily small values. And in the right plot we show the distribution of the difference between theory and data, in units of the error of the latter among each of the Monte Carlo replicas. This is consistent with the expected Gaussian distribution.

A recent development in the NNPDF family of global analyses concerns the treatment of the charm PDF. In all previous PDF sets NNPDF assumed that the charm PDF was generated dynamically from the gluons and light quarks, as dictated by the DGLAP evolution starting from the charm mass threshold $\mu_c \simeq m_c$. However, a possible non-perturbative component of the charm PDF would invalidate this assumption, a hypothesis which can ultimately be tested against experimental data. In addition, treating the charm PDF on an equal footing with the gluon and light quark PDFs offers other potential advantages, such as a reduced dependence on the value of m_c and an improved data/theory agreement from the more flexible input PDF parametrization. With this motivation, a variant of the NNPDF3.0 fit with a fitted charm PDF was studied in [25]. By parametrising the charm PDF with an artificial neural network with 37 free parameters, we found that fitting charm leads to an improved χ^2 for several experiments, stabilized the dependence of the fit with respect to the value of m_c and moreover allowing for the first time a satisfactory description, $\chi^2/N_{\rm dat} \simeq 1$, of the EMC charm structure function data. The resulting charm PDF can be compared with non-perturbative models [384], and some tantalizing evidence for a large-x non-perturbative charm component in the proton

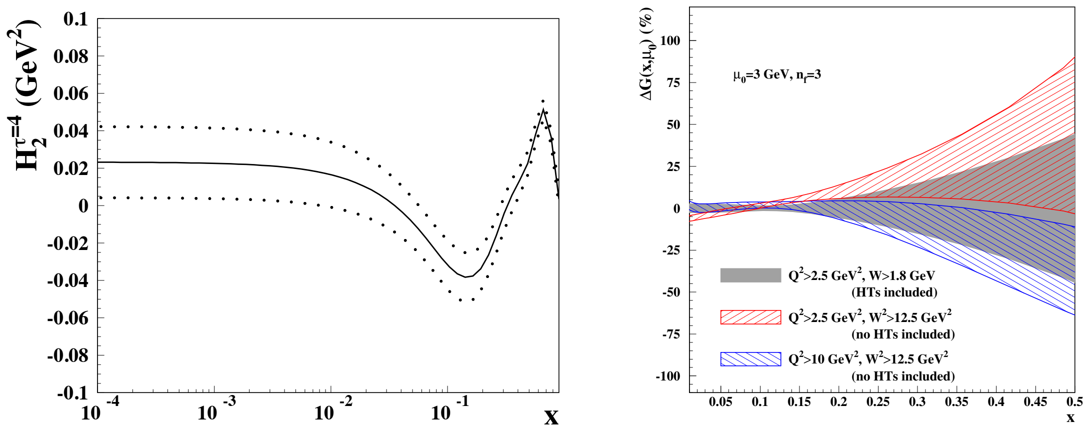


Figure 42: (Left) The higher twist coefficient for the F_2 structure function from the ABMP16 fit, including 1–σ uncertainty. (Right) Percentage difference in ABMP $n_f = 3$ gluon distribution between the default result and fits performed with higher W^2 cuts, without higher twist corrections. The 1–σ uncertainty bands are shown. Plots taken from [20].

was found. Predictions for a number of LHC process such as Z+charm and large-p_T D meson production were performed, showing the potential of Run II data to disentangle the charm content of the proton.

The most recent incarnation of the NNPDF global analysis is the NNPDF3.1 set. The main motivation for this release was the availability of a large number of high-precision collider measurements providing PDF sensitivity information, including some that for the first time could be used in a PDF fit, such as differential distributions in top quark pair production and the p_T of Z bosons. In addition, due to the impressive progress in NNLO QCD and NLO electroweak calculations, the theoretical predictions for these processes have now become available, allowing a consistent inclusion of all these observables in the NNLO fit. The second main motivation was to provide a state-of-the-art PDF set without assuming that charm is generated perturbatively, that is, providing baseline global PDF fits with fitted charm. Some of the new experiments included in NNPDF3.1 were the $t\bar{t}$ distributions from ATLAS and CMS, the legacy LHCb inclusive W and Z measurements from Run II, the D0 W asymmetries in the muon and electron channel, the p_T of Z bosons from ATLAS and CMS at $\sqrt{s} = 8$ TeV, as well as several other inclusive gauge boson and jet production measurements from ATLAS and CMS.

5.4. ABM

The ABMP16 [20] set is the latest PDF fit following on from the ABM11 [385], ABM12 [91] and ABKM09 [82] sets. These are based on the earlier studies of [71, 83, 84] to HERA and fixed proton and deuteron target DIS data, with the ABKM09 fit [82] and those that follow it including in addition fixed target Drell–Yan and dimuon production data from neutrino DIS on fixed nuclear targets. The PDFs are parameterised in terms of polynomials in x, with the latest fits including 25 free parameters. In the context of this fit to such a reduced data set, the use of the classical '$\Delta\chi^2 = 1$' criteria for determination of the PDF errors is applied. All sets from ABKM09 onwards go to NNLO in the strong coupling.

Two notable features of these fits are the use of a purely FFNS for the charm and bottom quark contributions in the fit and the treatment of higher–twist effects. In the latter case no attempt is made to impose a cut to remove the region of sensitivity to such effects. Rather, a lower cut of $W > 1.8$ GeV is imposed for the DIS data then is typically applied in other PDF fits. The structure functions are then given by

$$F_i(x, Q^2) = F_i^{\rm TMC}(x, Q^2) + \frac{H_i^{\tau=4}}{Q^2} \;, \tag{120}$$

where $i = 2, T$. Thus x dependent and Q^2 independent twist–4 corrections H_i^4 are included. While the effect of these dies off with increasing Q^2, at lower scales they can have a significant effect. These are then parameterised in terms of cubic splines defined at x_k ($k = 1...7$) points roughly linearly spaced between $x =$ 0 and 1, which are then determined from the fit. The result for the F_2 correction is shown in Fig. 42 (left), and is found to be inconsistent with zero, in particular at higher x. The effect of these corrections, and of conversely omitting them and including a more stringent W^2 cut on the DIS data is shown in Fig. 42 (right) for the extracted gluon PDF. The fit with the cut of $W^2 > 12.5$ and no higher twist corrections is found to prefer a somewhat larger gluon at higher x, and in some regions lies outside the 1–σ uncertainty band of the default fit.

In addition to the higher twist corrections included in (120), the structure function functions also include target mass corrections, that is the impact of terms $\sim M_N^2/Q^2$, where M_N is the nucleon mass. These are taken into account according the Georgi–Politzer prescription [386] (see also [385]), with

$$F_2^{\rm TMC}(x, Q^2) = \frac{x^2}{\xi^2\gamma^3} F_2(x, Q^2) + 6\frac{x^3 M_N^2}{Q^2\gamma^4} \int_\xi^1 \frac{{\rm d}\xi'}{\xi'} F_2(\xi', Q^2) \;, \tag{121}$$

where $\xi = 2x/(1+\gamma)$ and $\gamma = (1 + 4x^2 M_N^2/Q^2)^{1/2}$, and a similar result holds for F_T. Thus, as $Q^2 \to \infty$ the corrected $F_i^{\rm TMC}$ reduce to the regular F_i.

As mentioned above, the ABMP fit in addition uses a purely fixed flavour scheme to describe the DIS data. That is, this is fit with $n_f = 3$ light quark PDFs with the heavy c, b treated as massive final–state partons which can be produced at order NLO and higher. It is argued that the bulk of the DIS data can be described within this scheme. The Tevatron and LHC collider as well as fixed–target DY data, on the other hand, for which $\mu_F^2 \gg m_{c,b}^2$, is treated using a 5 flavour set evolved from the same input by means of the NNLO matching conditions [82]. PDF sets for $n_f = 3, 4$ and 5 active flavours are made publicly available.

A further feature of note is that the strong coupling α_S is determined from the fit. In ABM11 this was found to be $\alpha_S(M_Z^2) = 0.1134 \pm 0.0011$ at NNLO, that is in some tension with the PDG world average value of $\alpha_S(M_Z^2) = 0.1181 \pm 0.0013$ (the dominant uncertainty in which is determined by lattice QCD) used by the CT and NNPDF collaborations, and the value extracted by MMHT. While in [385, 20] the omission of higher twist corrections is found to lead to a sizeable increase in α_s, in contrast in [101] the use of the FFNS within the MSTW framework is found to lead to a smaller extracted value consistent with that seen by ABM(P), while higher twist effects are found to have less of an impact.

In the ABM11 fit [385], the heavy quark masses $m_{c,b}$ were included in the $\overline{MS}$ scheme for the first time, in contrast to other PDF fits. This allows the values to be constrained directly from the PDG results without relying on the perturbative transformation between the $\overline{MS}$ and pole masses, which is known to be poorly convergent. Thus, the quark masses are left free in the fit but with the PDG values added in as pseudo–data points. For the charm mass, the DIS data included in this fit are then found to give a comparable error to the PDG value.

The ABM12 fit [91] included hadron collider data for the first time, with a range of LHC W and Z boson and top pair production data at the LHC and Tevatron fit. The latest ABMP16 fit [20] includes an increased LHC set, including single top for the first time, as well as Tevatron lepton asymmetry data. In addition, the HERA I+II combined data set and updated NOMAD and CHORUS data on dimuon production are fit. For the $t\bar{t}$ data the mass m_t is treated in the $\overline{MS}$ scheme and is determined from the fit, giving

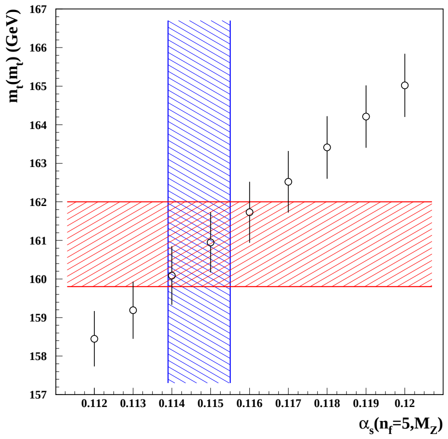


Figure 43: The $\overline{\text{MS}}$ value of the top quark mass $m_t(m_t)$ obtained in the ABMP16 fit for variants of $\alpha_S(M_Z^2)$ (data points) and for the best fit values (hatched bands). Plot taken from [20].

$m_t(m_t) = 160.9 \pm 1.1$ GeV. This is consistent with the PDG value of $160.0^{+4.8}_{-4.3}$ GeV, although this clearly has quite a large uncertainty, as it based on a single Tevatron measurement. The result is shown in Fig. 43, with the masses extracted at different $\alpha_S(M_Z^2)$ are also given. The correlation between m_t and α_s is clear; as discussed in [20] further information can be provided here by considering single top production data.

Interestingly, the extracted value of the strong coupling, $\alpha_S(M_Z^2) = 0.1147 \pm 0.0008$, is somewhat larger in this fit than in the previous sets, due dominantly to the HERA I+II combined data, although this is still lower than the world average value. The ABMP16 PDFs are available as error sets at NNLO for 3,4 and 5 fixed flavours, and for a range of $\alpha_S(M_Z^2)$ values, from 0.112–0.120 in steps of 0.001, in the 5 flavour scheme.

5.5. CJ

The CTEQ-Jefferson Lab (CJ) Collaboration has performed a series of global PDF analyses [387, 388] with the latest PDF set being CJ15 [23], following the previous CJ12 set [365]. The analyses are carried out at NLO in QCD only, and focus on utilising DIS data at the highest-x values applicable to a perturbative QCD treatment. The kinematic selection cuts are chosen to be $Q^2 > 1.69\,\text{GeV}^2$ and $W^2 > 3\,\text{GeV}^2$ so as to keep data points at low-Q and high-x. This results in about 1300 more data points from proton and deuteron targets, roughly a 50% increase as comparing to standard cuts. These additional data points provide valuable information on the PDFs at large-x, into the $x \gtrsim 0.7$ region where the constraints for most global analyses are indirect or purely from extrapolation. In particular, the deuterium data can improve on the determination of d quark at large-x, for which the proton DIS data are less sensitive.

For the treatment of heavy-quark mass effects in DIS structure functions, CJ12 uses a ZM-VFN scheme with heavy-quark masses implemented as the flavor thresholds. CJ15 uses a more adequate GM-VFN scheme S-ACOT [99] to better describe data over a wide kinematic range, including the threshold regions. It is found that the implementation of the GM-VFN scheme leads to large changes in the gluon PDF at large-x. Going to low-Q and large-x involves further complications to the theoretical predictions for the DIS structure functions, as finite Q^2 corrections to the leading-twist calculation, i.e. power corrections of $\mathcal{O}(1/Q^2)$. must be taken into account. CJ analyses adopt the standard OPE expression for the target mass corrections (TMCs) which allows structure functions at finite Q^2 be expressed in terms of their massless

($M^2/Q^2 \sim 0$) values through the scaling variable $\rho^2 = 1+4x^2M^2/Q^2$ [389, 390]. For other subleading $1/Q^2$ corrections including higher twists, they are parametrized by a phenomenological function form [23].

Another important aspect concerns the nuclear corrections for processes with deuteron targets, which become significant in the intermediate and large-x region and are equally important for low and high Q values. The nuclear corrections account for Fermi motion, binding, and nucleon off-shell effects, and can be implemented as convolutions with nuclear smearing functions. In the CJ12 analysis three PDF fits are provided with different models of deuterium corrections, CJ12min, CJ12mid and CJ12max, corresponding to mild to strong corrections. The corrections are only applied at the level of structure functions. The CJ15 analysis employ a phenomenological parametrization for part of the deuterium corrections with free parameters fitted to data, reducing the model dependence and increasing the flexibility of the fit. The deuterium corrections are formulated at the parton level and can therefore also be applied to non DIS processes. For example, the total quark PDF in deuteron can be written as $q^d = q^{d(\text{on})} + q^{d(\text{off})}$, with the on-shell and off-shell components given by [391, 392],

$$\begin{aligned} q^{d(\text{on})}(x, Q^2) &= \int \frac{dz}{z} f^{(\text{on})}(z) q^N(x/z, Q^2) \,, \\ q^{d(\text{off})}(x, Q^2) &= \int \frac{dz}{z} f^{(\text{off})}(z) \delta f^N(x/z, Q^2) q^N(x/z, Q^2), \end{aligned} \tag{122}$$

where $q^N(x, Q^2)$ is the quark PDF in free nucleons. The on-shell and off-shell smearing functions $f^{(\text{on})}$ and $f^{(\text{off})}$ can be calculated systematically within the weak binding approximation, using the deuteron wave functions [393, 394]. The nominal CJ15 PDF fit is based on AV18 wave functions [395], but alternative fits with CD-Bonn [396], WJC-1 and WJC-2 [397] wave functions are also provided. The off-shell correction $\delta f^N(x)$ in Eq. (122) is parametrized as [393]

$$\delta f^N(x) = C(x - x_0)(x - x_1)(1 + x_0 - x). \tag{123}$$

The two zeros x_0, x_1 and the normalization C are free parameters fitted to data with the constraint of maintaining the total number of valence quarks in the nucleon. It was found that different wave function models give similar quality of fits to the global data set and result in changes in the PDFs that are well within the uncertainties, since their differences can be largely compensated by the parametrization of the off-shell corrections.

In the CJ15 analysis two new data sets are found to have significant effects on constraining the d-quark PDF at large-x beside of the deuteron data. These are the measurement of the F_2 structure function of a nearly free neutron inside a deuterium nucleus from the BONuS experiment [398, 112] at Jefferson Lab using a spectator tagging technique, and the the lepton and reconstructed W boson charge asymmetry measurements from D0 Run 2, with full luminosities [180, 181]. In Fig. 44 the plot on the left shows impact of different data sets on PDF uncertainty (90% C.L.) of d/u ratio in CJ15 fits. It was found that at $x \lesssim 0.3$ the DIS data from deuteron target can reduce the PDF uncertainty on d/u by almost 50%. For $x \gtrsim 0.3$ the W asymmetry data provides the dominant constraint. Besides, the constraint from deuteron DIS data dies out for $x \gtrsim 0.6$ which turns into fit to the deuterium off-shell corrections. In Fig. 44 (Right) the comparison of the d/u ratios from CJ15, MMHT14, CT14 and JR14 PDFs is shown. They are in good agreements within PDF uncertainties as x goes to 1. The CJ15 PDF set has smaller PDF error on d/u in general, with an extrapolated value

$$d/u \longrightarrow_{x\to 1} 0.09 \pm 0.03, \tag{124}$$

at the 90% C.L., due to the new data sets on constraining d-quark at large-x. On the other hand with the additional data sets that are less sensitive to the nuclear corrections, i.e., the D0 W asymmetry data and

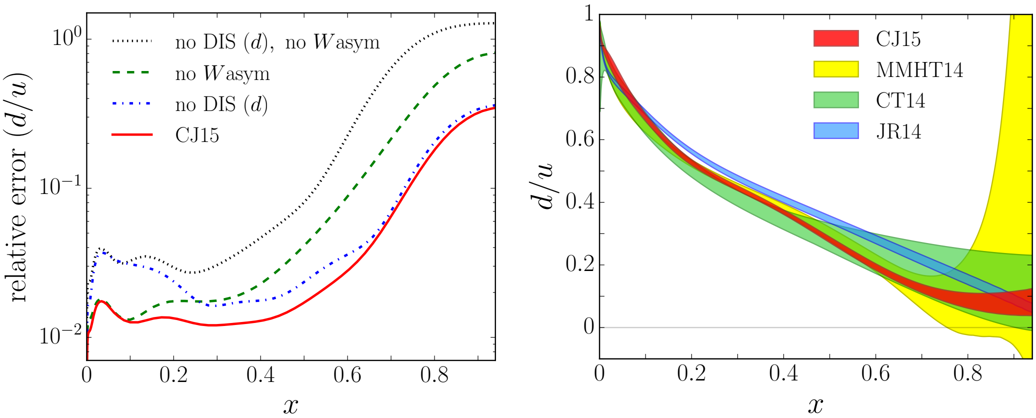


Figure 44: Left plot: relative error (90% C.L.) on the d/u PDF ratio as a function of x at $Q^2 = 10\,\mathrm{GeV}^2$ from the CJ15 fit compared with errors obtained in fits excluding certain data sets [23]. Right plot: comparison of the d/u ratio at $Q^2 = 10\,\mathrm{GeV}^2$ for different PDF sets, including CJ15, MMHT14, CT14, and JR14, with PDF uncertainties shown for 90% C.L. except for MMHT14 at 68% C.L. [23]

BONuS data, the CJ15 analysis is able to pin down the deuterium corrections through the interplay with deuteron DIS data. Fig. 45 gives the deuteron to isoscalar nucleon ratio F_2^d/F_2^N from the CJ15 fits with different input wave functions together with the 90% C.L. uncertainty of the CJ15 nominal fit shown in the coloured bands. Significant corrections are found for $x \gtrsim 0.7$. The ratio is insensitive to the choice of wave functions since it is only the combination of the wave function and the off-shell corrections that are constrained by current data.

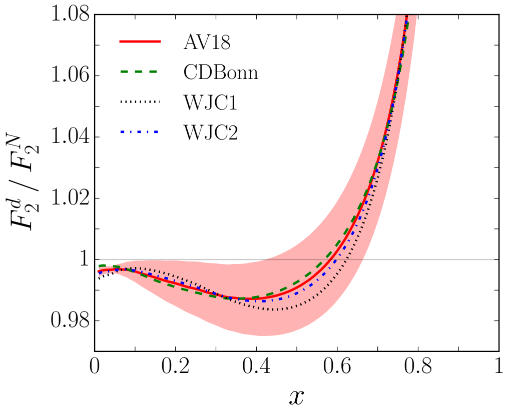


Figure 45: Ratio of deuteron to isoscalar nucleon structure functions F_2^d/F_2^N at $Q^2 = 10\,\mathrm{GeV}$ for CJ15 fits with different models of wave functions [23]; colored band is the 90% C.L. error for CJ15 main fit (with AV18 wave functions).

5.6. *HERAfitter/xFitter*

For many years, the H1 and ZEUS collaborations performed QCD analyses of their structure function data, first separately and then together based on the H1+ZEUS combined datasets. The backbone of these analysis was the neutral- and charged-current inclusive structure function measurements, in some cases supplemented by the charm production structure functions and DIS jet cross-sections. The main results from these studies were the HERAPDF family of PDF fits, which include HERAPDF1.0 [85], based on the Run II data, and HERAPDF2.0 [21], based on the final combination of inclusive measurements from Runs I+II. In Fig. 46 we show the results of the HERAPDF1.0 analysis for the u_V, d_V and S quarks and the gluon. In the HERAPDF methodology, the total PDF uncertainty is divided into three types of errors: experimental uncertainties, propagated from the statistical and systematic uncertainties in the fitted data, model uncertainties, for instance due to α_s and Q_0 variations,and parametrization uncertainties, reflecting the spread from different comparable choices of input functional form for the PDFs.

The expertise developed by these QCD analyses of HERA structure function data lead to the development and release of `HERAfitter` [27], a publicly available open–source PDF fitting toolbox. This was developed as an extension of the H1 and ZEUS internal PDF fitting codes, that were extensively tested and applied in various QCD analyses of HERA inclusive and charm data, including the HERAPDF sets. Despite its name, `HERAfitter` was not restricted to the analysis of HERA data, and could also be used for the PDF interpretation of measurements from fixed-target DIS and proton-proton collisions. The flexibility of this open–source software tool also allows QCD analyses beyond unpolarized fixed-order PDF fits to be performed, such as fits of transverse-momentum dependent (TMD) parton distributions and fragmentation functions. Recently, `HERAfitter` was renamed `xFitter`, to emphasize that this code is a general fitting toolbox not necessarily related to or involving the analysis of HERA inclusive structure function data.

The `xFitter` framework includes modules that allow for various theoretical and methodological options, and is capable of fitting to a large number of data sets from HERA, Tevatron and LHC. For example, polarized and unpolarized PDF evolution can be performed using either `APFEL` or `QCDNUM`, and a number of fixed and variable flavor number schemes are implemented, such as the FFN scheme from `OpenQCDrad` and the S-ACOT, TR and FONLL general-mass schemes. In addition to PDF fitting, a large number of their functionalities are available, such as the approximate inclusion of new datasets in existing PDF sets by means of either Bayesian reweighting or Hessian profiling, and a wide variety of PDF plotting options. In Fig. 46 we show a schematic representation of the `xFitter` code structure. The first part is the initialization, where the fit settings are specified in the steering file. This involves a number of choices, in particular selecting the fitted data sets and the theory and methodology settings such as the specific PDF parametrisation or the scheme for heavy quark structure functions. Then the PDF fit is performed, where the fit parameters are determined by means of `MINUIT`-based minimization including the propagation of experimental results. The final result is the `LHAPDF6` grid file, together with various PDF and data/theory comparison plots.

The `HERAfitter/xFitter` framework has been used in many ATLAS and CMS PDF interpretation studies, discussed in Sects. **??** and **??** respectively. In addition, the `HERAfitter/xFitter` developer's team has released a number of dedicated PDF studies, including:

- A QCD analysis of the legacy W and Z boson production measurements at the Tevatron [340], including the precise W asymmetries in the electron and muon channels by the D0 collaboration, together with the HERA structure function data. This analysis demonstrated that these measurements, which are now included in most global PDF fits, provide useful information on quark flavour separation at medium and large-x.

- A determination of the running charm quark mass $m_c(m_c)$ from HERA structure function data within the framework of the FONLL general-mass variable-flavour number scheme [399]. This study demonstrated that the best fit value of $m_c(m_c)$ was consistent when using FONLL as compared to a fit performed in the fixed-flavor number scheme, as expected from general theoretical considerations. This

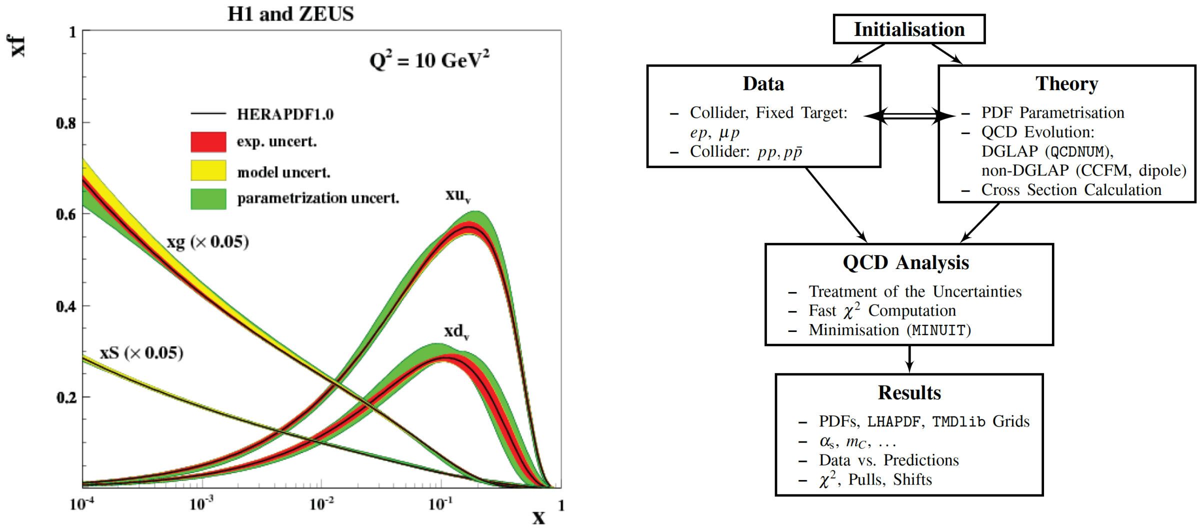


Figure 46: Left plot: the HERAPDF1.0 determination of parton distributions, based on the analysis of the combined HERA structure functions from Run I. Right plot: schematic representation of the `xFitter` code structure, see text for more details.

value was also consistent with previous determinations of the running mass from HERA data and with the global PDG average.

- A determination of the photon PDF $x\gamma(x, Q^2)$ [400] from the measurement of Drell-Yan high mass cross-sections at 8 TeV from the ATLAS collaboration. This was the first analysis where LHC data was included in a QED fit of the photon PDF directly, rather than using reweighting methods, by means of an extension of the `aMCfast` interface to account for photon-initiated contributions in `MadGraph5_aMC@NLO`. The results of this analysis showed that the high-mass DY data indeed allowed important constrains on the photon PDF at intermediate x, although the resulting PDF uncertainties were still not competitive with those from the more recent determinations discussed in Sect. 7.1.

Additional work based on `xFitter` include studies of PDFs with correlated uncertainties between different perturbative orders [401], non-DGLAP evolution equations [402], and the determination of transverse-momentum dependent PDFs [32], In Fig. 47 we show two representative PDF-related analyses performed by the `xFitter` Developer's Team. First, we show the impact on the d_V PDF of the Tevatron W and Z data when added to an HERA-only fit, comparing the impact of the lepton-level measurements with that of the boson-level measurements, from the `xFitter` analysis of Ref. [340]. We also show the χ^2 profile of a `xFitter` fit based on the inclusive HERA and charm data, as a function of the running mass $m_c(m_c)$ from Ref. [399]. In this analysis charm structure functions were computed with `APFEL` in the FONLL-C general mass scheme. As discussed above, this analysis finds a value of the running charm mass $m_c(m_c) = 1.335 \pm 0.043$ which is consistent with the PDG average as well as with previous determinations based on HERA data.

Concerning future developments, the `xFitter` code is now being rewritten from Fortran to C++, to ensure modularity and to facilitate its maintenance and the addition of novel theoretical ingredients. Several new external codes and additional features are being implemented, such as the possibility of new parametrization options like Chebyshev polynomials, the fast convolution option for hadronic cross-sections

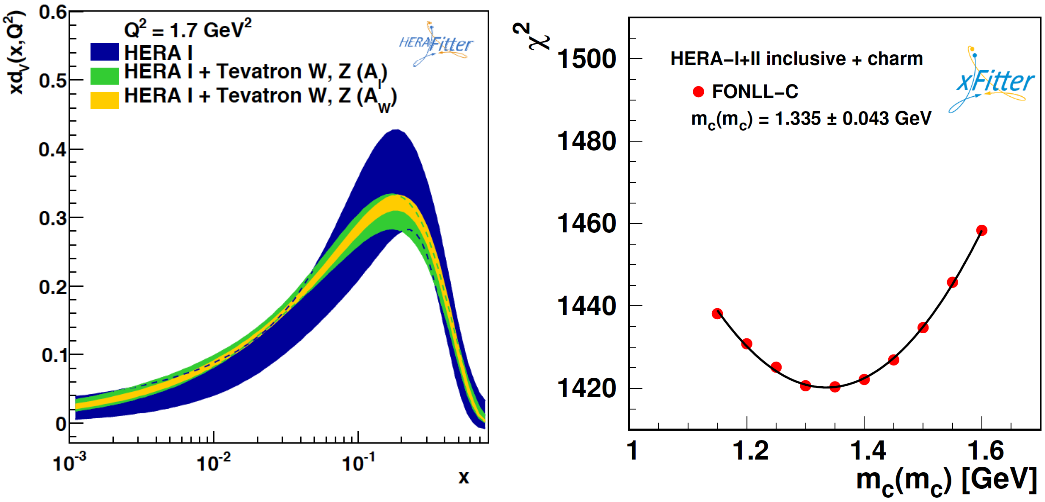


Figure 47: Two representative analyses of PDF-related studies performed by the `xFitter` Developer's Team. Left plot: the impact on d_V of the Tevatron W and Z data on a HERA-only fit, comparing the impact of the lepton-level measurements with that of the boson-level measurements, from Ref. [340]. Right plot: the χ^2 profile of a fit based on the inclusive HERA and charm data, as a function of the running mass $m_c(m_c)$ from Ref. [399]. In this analysis charm structure functions were computed with `APFEL` in the FONLL-C general mass scheme.

as realized in `APFELgrid` [309], more flexible PDF parametrizations including the charm and the photon PDF, and improvements in the QED evolution interface.

5.7. PDF efforts by the LHC collaborations

As discussed in Sect. 3, the LHC experiments have provided a large number of experimental measurements with important PDF sensitivity, most of which are now part of the toolbox of global PDF fits. In addition to presenting the results of such measurements, the ATLAS and CMS collaborations have developed an active program of PDF interpretation studies, aimed to quantify the constrains of their data on the proton structure.

In all cases these are performed using the `xFitter` framework described in Sect 5.6. Thus the PDFs are parameterised at $Q_0^2 = 1.9\,\mathrm{GeV}^2$ in terms of simple polynomials in x. Fits are then performed with an increasing number of free parameters introduced up to the point when no further improvement in χ^2 is observed. This generally leads to ~ 15 free parameters in the fit, with the precise number depending on the particular analysis. Experimental uncertainties are calculated using the standard '$\Delta\chi^2 = 1$' criteria, and as in the HERAPDF fit, additional model and parameterisation uncertainties are determined. In all cases either the HERA I DIS [85], or in later studies the I + II combination [21] are included in a baseline fit, before assessing the impact of the corresponding CMS data, which are then fit in addition.

Such studies represent an important contribution to the PDF fitting community, not only because they demonstrate the PDF impact of specific measurements, but also because they provide an internal cross-check that the information required for PDF fits, in particular the full experimental covariance matrix, is ready to be used. We describe the individual efforts from ATLAS and CMS below.

5.7.1. ATLAS

A representative selection of ATLAS PDF interpretation studies is given below:

- The ATLAS measurements of W^+, W^- and Z rapidity distributions at 7 TeV from the 2010 dataset were used in Ref. [221], in combination with HERA DIS data, to determine the strange content of the proton. The full cross-correlations between the three rapidity distributions were accounted for, and while W^+ and W^- measurements constrained the up and down quarks and antiquarks, the Z measurement constrained the strangeness. This analysis found that the strange sea was not suppressed as compared to the up and down quark sea.

- The recent study [192], based on the updated W^+, W^- and Z rapidity distributions at 7 TeV from the 2011 dataset, corresponding to a greatly improved precision in comparison to 2010. An analysis of this data, combined with HERA DIS data, was found to prefer a strange quark sea that is symmetric with respect to the light quark sea, consistent with [221]. The PDF uncertainties in the strangeness determination were significantly reduced in comparison to the PDFs determined from the analysis of the 2010 data. The issue of the strange content of the proton will be discussed in more detail in Sect. 6.3.

- PDF fits based on jet production measurements have also been performed by ATLAS. For instance, in [147] an analysis of the HERA DIS data supplemented with inclusive jet cross sections at $\sqrt{s}$ = 2.76 TeV and $\sqrt{s}$ = 7 TeV (from the 2010 run) was performed. It was shown that an improved constraint could be achieved by considering the ratio $R_{7/2.76}$ of jet cross-sections is used, given that many experimental and theoretical uncertainties cancel when taking such ratios between different centre-of-mass energies [403].

In Fig. 48 we show some representative results of PDF interpretation studies performed within the ATLAS collaboration. In the left plot, we show the results of a PDF fit quantifying the effect on the gluon of the ATLAS inclusive jet measurements at $\sqrt{s}$ = 2.76 TeV and 7 TeV, in comparison to a HERA–only fit. In the right plot we show the determination of the strangeness ratio $R_s(x = 0.023, Q^2 = 1.9\,\mathrm{GeV}^2)$ for a HERA–only fit and including the 2011 ATLAS measurements of the $W^\pm$ and Z rapidity distributions at 7 TeV. The results of the `xFitter` analysis, denoted by ATLAS-epWZ16, are compared with the predictions from various PDF fits.

5.7.2. CMS

A representative selection of CMS PDF studies is given below:

- In [191] the 7 TeV measurement of the W charge asymmetry, as well as $W + c$ production [404], is fit at NLO, and improvements in the determination of the up and down valence quark PDFs due to the W asymmetry, and the strange quark PDFs due to the $W + c$ data, are demonstrated.

- In [220] a fit to the the 8 TeV differential W boson production data [220] is performed at NNLO, again showing improvements in the determination of the up and down valence quark PDFs.

- In [173] the 7 TeV inclusive jet measurement is fit at NLO, and the significant impact of these data on the gluon PDF in particular is demonstrated. A study is also performed here using the MC method for PDF determination, allowing for a more flexible PDF parameterisation, and consistent results are found but with larger PDF uncertainties.

- A NLO fit to the 8 TeV jet data is performed in [138], and a direct comparison to the 7 TeV case is shown, with the impact found to be very similar.

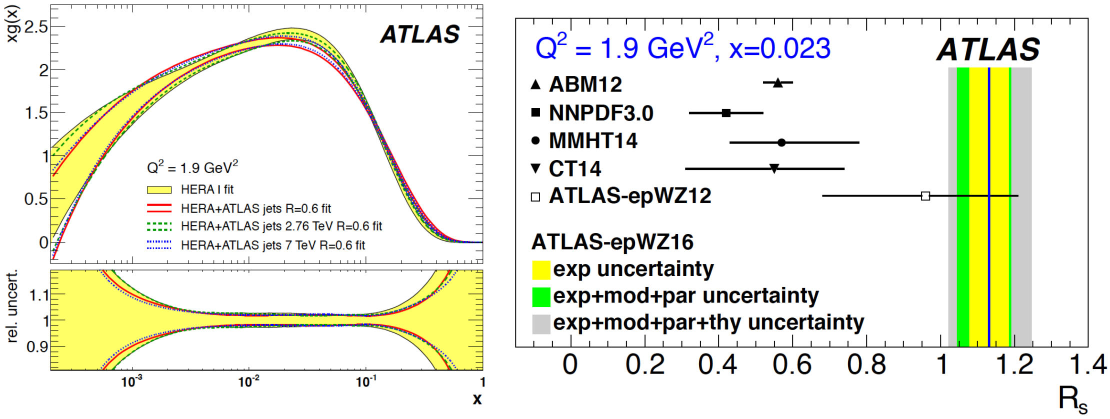


Figure 48: Two representative results of the PDF fitting efforts performed within the ATLAS collaboration. Left plot: a PDF fit quantifying the effect on the gluon from the HERA-only fit of the ATLAS inclusive jet measurements at $\sqrt{s}$ = 2.76 TeV and 7 TeV, from Ref. [147]. Right plot: the determination of the strangeness ratio $R_s(x = 0.023, Q^2 = 1.9\,\text{GeV}^2)$ for a fit to HERA data and the 2011 ATLAS measurements of the $W^\pm$ and Z rapidity distributions at 7 TeV, where the results of the `xFitter` analysis, denoted by ATLAS-epWZ16, are compared with the predictions from various PDF fits. Taken from [192].

- In [279] a NLO fit to the 8 TeV double differential top pair production data is compared to a baseline fit that includes the 8 TeV W boson production data [220]. The impact of including the data differential in different kinematic variables is assessed, and a sizeable reduction in the uncertainty on the gluon PDF in particular is found for $x > 0.01$, with the largest constraint coming from the rapidity, $y_{t\bar{t}}$, and invariant mass, $M_{t\bar{t}}$ of the top pair.

- More recent preliminary results including PDF fits to triple differential dijet production at 8 TeV [155] and the top pair production cross section at 5.02 TeV [405] have been presented.

In Fig. 49 (Left) we show the impact on the gluon PDF of the CMS $W^\pm$ data [220], the double differential top pair production data [279] and the inclusive jet production data [138], in all cases at 8 TeV. This is seen to lead to a sizeable reduction in the uncertainty at higher x, in a way that is consistent between the data sets in the probed x region. The $t\bar{t}$ differential data are competitive with the the measurement. In Fig. 49 (Right) we show the impact of the CMS $W^\pm$ boson charge asymmetry measurement at 8 TeV on the down valence distribution, in comparison to a HERA–only fit. The impact on the shape, and reduction in uncertainties, achieved by the asymmetry data is clear.

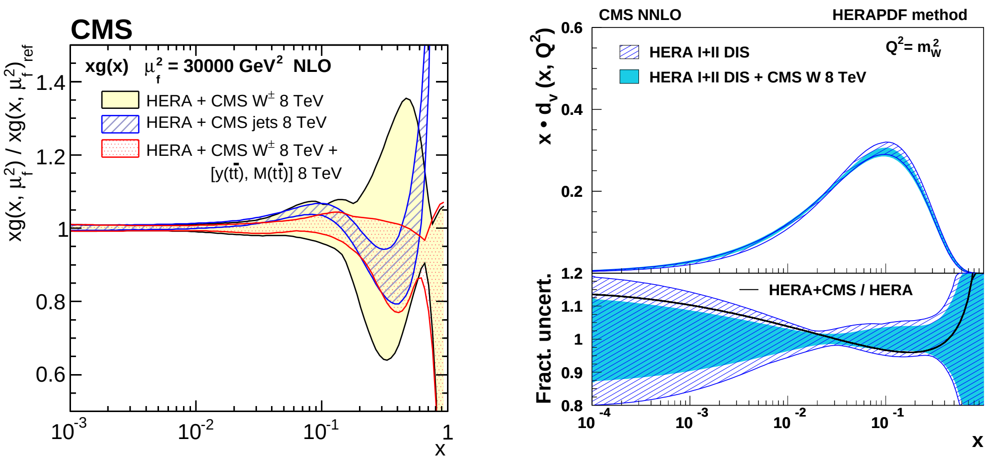


Figure 49: (Left) The gluon distribution at $\mu^2 = 30000\,\mathrm{GeV}^2$, as obtained from a PDF fit to HERA DIS data and CMS $W^\pm$ boson charge asymmetry measurements [220], the CMS inclusive jet production cross sections [138], and the $W^\pm$ boson charge asymmetry plus the double-differential $t\bar{t}$ cross section [279], in all cases at 8 TeV. All presented PDFs are normalized to the results from the fit using the DIS and $W^\pm$ boson charge asymmetry measurements, and the total uncertainty band in each fit is shown. Taken from [279]. (Right) The down valence distribution at $\mu^2 = M_W^2$ as obtained from a PDF fit to the HERA DIS data and CMS $W^\pm$ boson charge asymmetry measurement at 8 TeV, with the total PDF uncertainties shown. In the lower panel the distributions are normalized to 1. Taken from [220].

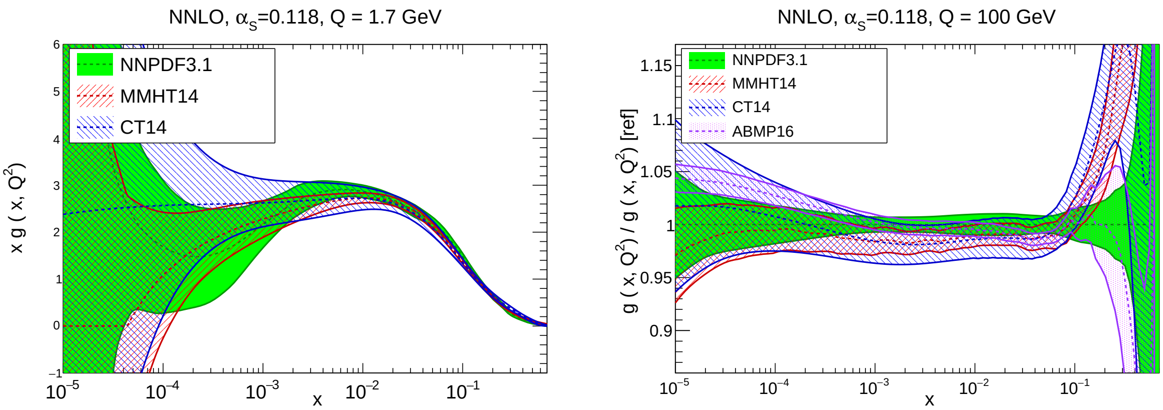


Figure 50: The gluon PDF $xg(x, Q^2)$ at $Q = 1.7$ GeV (left) and $Q = 100$ GeV (right) comparing the ABMP16, CT14, MMHT14, and NNPDF3.1 NNLO sets with $\alpha_S(m_Z) = 0.118$. In the right plots, results are normalized to the central value of NNPDF3.1.

6. The proton structure

Following the discussion of the various approaches to PDF fitting in the previous section, here we compare the results of the most representative state-of-the-art PDF fits. This comparison is organized in terms of specific PDF flavour combinations relevant for phenomenology. We begin by discussing the gluon, before turning to the quark flavor separation followed by the large-x behaviour of the PDFs, and subsequently studying the strange and charm content of the proton.

In the following we compare ABMP16, CT14, MMHT14, and NNPDF3.1 NNLO sets, all with $\alpha_S(m_Z) = 0.118$. Note that for the low-scale comparisons the ABMP16 curve cannot be included since the set with $\alpha_S(m_Z) = 0.118$ is only available in the $n_f = 5$ scheme, and therefore can only be used above the bottom quark threshold. We will only show a representative selection of PDF comparisons: other restults, including with PDF sets not shown here, can be simply produced using the `APFEL-WEB` online PDF plotting interface [406].

6.1. The gluon

In Fig. 50 we show gluon PDF $xg(x, Q^2)$ at a low hadronic scale $Q = 1.7$ GeV (left plot) and at a typical LHC scale $Q = 100$ GeV (right plot). We find that in general there is reasonable agreement between the four sets of PDF considered within uncertainties. This remains true at small-x, where the PDF uncertainties become rather large, due to the lack of experimental constraints. While the central value of the MMHT14 gluon becomes negative for $x \lesssim 5\times10^{-5}$, the CT14 result exhibits a flat behaviour, and the NNPDF3.1 gluon increases rapidly. The agreement with ABMP16 becomes significantly worse if the PDF set corresponding to their best-fit $\alpha_s(m_Z) = 0.1149$ value is used.

Perhaps the most important discrepancy between the gluon PDFs from the four groups arises in the large-x region, where the NNPDF3.1 result (and even more markedly ABMP16) is rather softer in comparison to CT14 and MMHT14. For example, at $x \simeq 0.2$ the differences between the NNPDF3.1 and CT14 central values are at the 2-σ level. One of the reasons for these differences could be related to the use of different datasets to constrain the large-x gluon, and specifically by the use of top-quark differential distributions in NNPDF3.1, which have been shown to lead to a softer large-x gluon as compared to the same fit without any $t\bar{t}$ data included [289]. Indeed, the CT14 and previous NNPDF3.0 set (which do not fit such

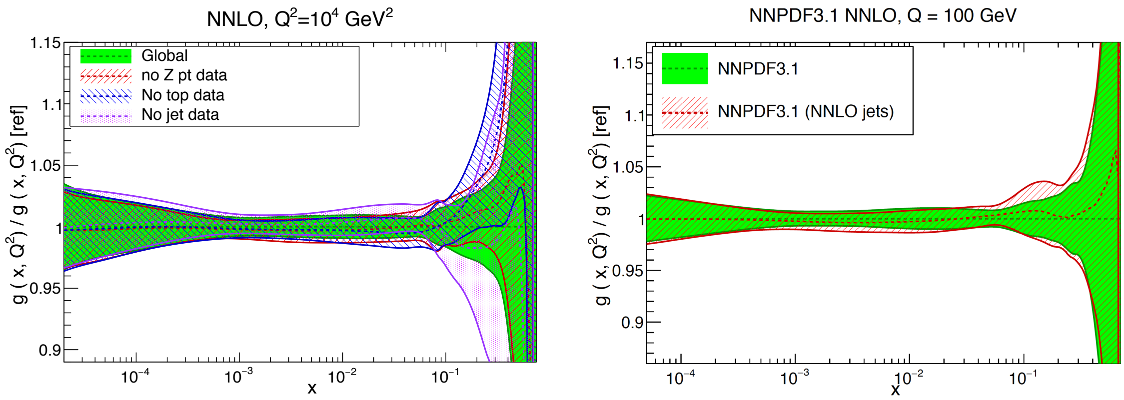


Figure 51: Left: comparison of the NNPDF3.1 NNLO global fit at $Q = 100$ GeV with the corresponding fits where the Z p_T, top quark, or inclusive jet data have been removed. Right plot: same as before, now comparing with the NNPDF3.1 NNLO fit where the ATLAS and CMS 7 TeV inclusive jet data have been treated using exact NNLO theory.

data) are in better agreement within uncertainties . We also note that in the large-x region PDF uncertainties are quite large, leading to siginificant theoretical uncertainties for the production of new BSM heavy particles, as will be discussed in Sect. 8.2. In this respect, it will be interesting to compare the large-x gluon PDF from the three global sets once they include a similar dataset.

It is worth emphasising that one of the most important differences of the current version of PDF fits as compared to previous versions is the fact that several datasets provide independent constraints on the large-x gluon. Until recently, the gluon at large-x was only constrained in the PDF fit by inclusive jet cross-sections, and to a lesser extent by DIS data via scaling violations. However, we now have at least three datasets that can constrain the large-x gluon, namely inclusive jets, the p_T distribution of Z bosons, and top quark differential distributions. In all cases, NNLO calculations are now available. To illustrate the robustness of the resulting gluon, in Fig. 51 we show a comparison of the NNPDF3.1 NNLO global fit at $Q = 100$ GeV with the corresponding fits where the Z p_T, top quark, or inclusive jet data have been removed. We can observe that the four fits agree within PDF uncertainties, highlighting that these three families of physical processes have statistically consistent pulls on the large x gluon.

Another point that is relevant to the large-x gluon in global fit are the settings for the theoretical calculation of inclusive jet cross-sections. Until 2016, only the NLO calculation was available, and different groups treated jet data in the global fit in different ways, from adding the NLO scale errors as additional systematic uncertainties as in CT14 and NNPDF3.1, to using the threshold approximation [167] to the full NNLO result as in MMHT14, to excluding jet data altogether as advocated by ABMP16. The availability of the complete NNLO calculation makes these different approaches obsolete, and future iterations of the various PDF fits will be able to fit to inclusive jet data using the exact NNLO theory. This said, there is evidence that, at least for specific settings of the NLO calculation, the inclusion of jet data with NNLO theory has a moderate phenomenological relevance. This is because, if the jet p_T scale is adopted as central renormalization and factorization scale, and a not too small value of R is used, the NNLO/NLO K-factor is a few percent at most. To illustrate this point, in Fig. 51 we compare the baseline NNPDF3.1 NNLO fit (where jet data are treated at NLO, with scale uncertainties as additional systematic error) to the same fit where exact NNLO theory is used for the ATLAS and CMS 7 TeV data. We see that the resulting differences are small at the PDF level, and at the χ^2 level one finds [251] a small but non-negligible improvement once

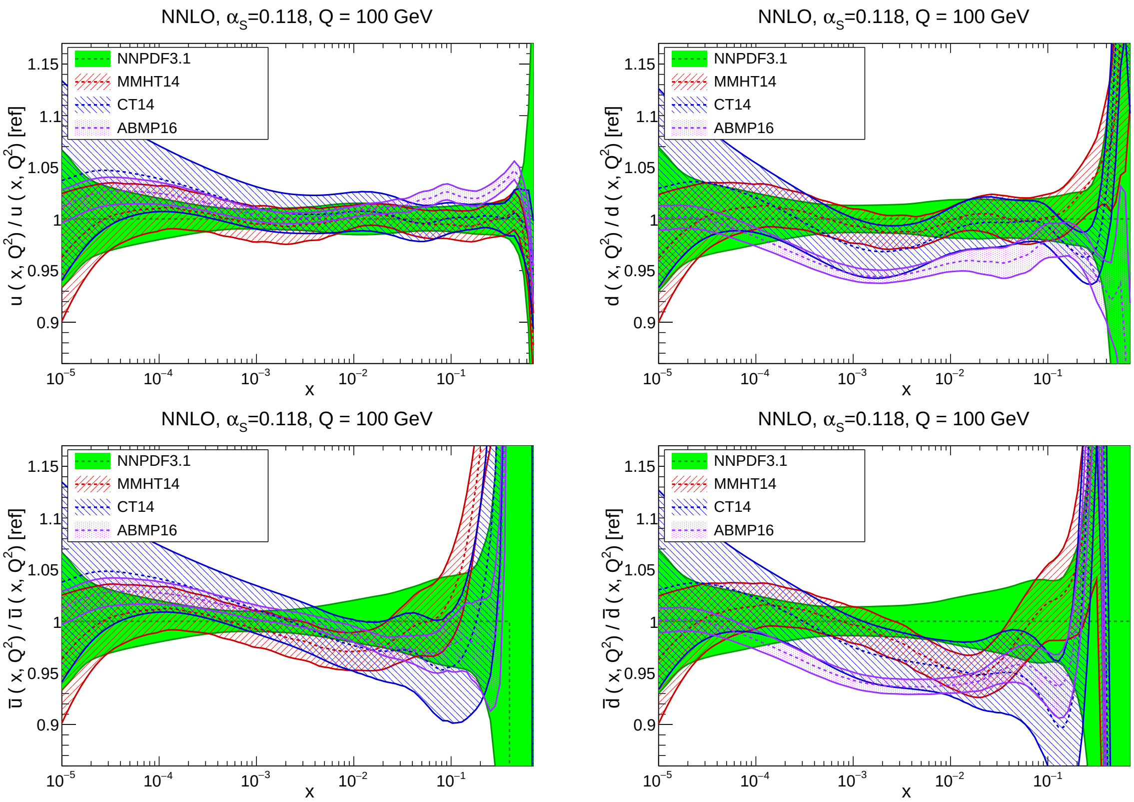


Figure 52: Same as Fig. 50 (right), now comparing the up, down, anti-up, and anti-down quark PDFs.

NNLO theory is used.

6.2. *Quark flavor separation*

In Fig. 52 we show the up, down, anti-up, and anti-down quark PDFs at $Q = 100$ GeV. The up quark, $u(x, Q_0)$ is one of the better constrained PDFs, in particular at large-x, due to fixed–target DIS data. For this PDF, we find good agreement within uncertainties in the entire range of x, with the only exception being ABMP16, which overshoots the other three sets in the large-x region. For the down quark, $d(x, Q_0)$, the spread between the central values is larger, and the PDF uncertainties are also comparatively increased. Here we find good agreement between CT14, MMHT14, and NNPDF3.1 within uncertainties for the entire range of x, while ABMP16 is around 5% lower than the central NNPDF3.1 value at intermediate values of x. The PDF uncertainties are the largest at high-x, with CT14, MMHT14 and ABMP16 on the other hand pointing in different directions, with the NNPDF3.1 central value lying somewhere in the middle. One of the possible sources of difference between the groups is the treatment of deuteron nuclear corrections in the fitting of the F_2^d structure functions [19], though this effect is known to be localized in the region around $x \simeq 0.1$ [407].

For the light antiquark PDFs, $\bar{u}$ and $\bar{d}$, there is reasonable agreement between the various sets within uncertainties for $\bar{u}$, while this agreement is marginal for $\bar{d}$. In the latter case, the ABMP16 result is again around $\simeq$ 5% smaller than the NNPDF3.1 central value. As in the case of the quark PDFs, we see significant differences at large-x; in this region there are limited experimental constraints, and thus the methodological differences in each PDF fit can have a rather more marked impact. Similarly to the gluon, these large PDF

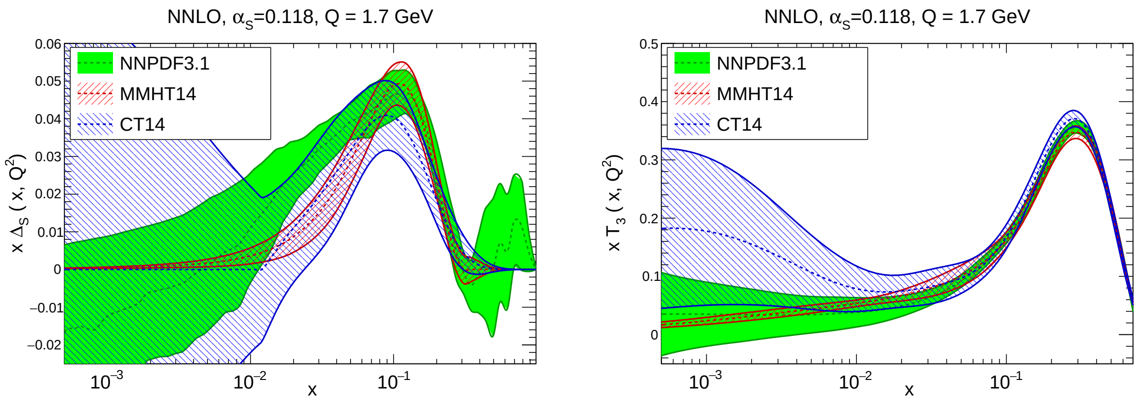


Figure 53: Same as Fig. 50 (left), now comparing the sea quark asymmetry $\Delta_S = \bar{d} - \bar{u}$ (left) and the quark isotriplet $T_3 = u + \bar{u} - d - \bar{d}$ (right plot).

uncertainties at high-x have phenomenological consequences, for instance for the production of a heavy W' or Z' boson, or the pair production of a squark-antisquark pair $\tilde{q}\tilde{q}^*$, both processes being driven by the quark-antiquark luminosity.

Another useful way to compare the quark flavour separation between the various PDF group is to plot flavour combinations that can be directly related to physical cross-sections. In Fig. 53 we compare the sea quark asymmetry $\Delta_S = \bar{d} - \bar{u}$ and the quark isotriplet $T_3 = u + \bar{u} - d - \bar{d}$ at $Q = 1.7$ GeV for CT14, MMHT14, and NNPDF3.1. The former flavour combination is closely related to the W asymmetries in collider Drell-Yan production, while the latter is directly sensitive to the difference between the proton and deuteron DIS structure functions, $F_2^p - F_2^d$. From this comparison, we see that for Δ_S the general shape is similar between the three groups, although there are large differences in the estimate of the PDF uncertainties, both at small and large x, which in some cases can be traced back to the PDF parametrization assumptions. The agreement is reasonably good both in terms of central values and of uncertainties for T_3, although here again the small-x behaviour does differ among the three groups.

From the previous comparisons we can see that the differences in the quark flavour separation between the various groups are mostly localised in the large-x region. With this in mind, in Fig. 54 we again show the up and down quark PDFs, but focusing on the large-x region, using a linear scale in the x axis. From this comparison we can see that PDF uncertainties are the largest in NNPDF3.1. In terms of central values, there is reasonable agreement for u, less so for d. Note that in the NNPDF fits the PDFs are not forced to be positive (although the physical cross-sections are indeed positive-definite) and therefore the down PDF can become negative at large-x, although its central value is always positive. An alternative approach to compare the behaviour of PDF sets at large-x, and in doing so comparing with non-perturbative models such as the quark counting rules, is the effective exponent method discussed in Ref. [312].

6.3. The strange content of the proton

The size and shape of the strange PDF has recently attracted a lot of debate. On the one hand, most PDF fits find a suppressed strangeness as compared to the non-strange light quark sea, a pull driven mostly by the deep-inelastic neutrino inclusive F_2 and charm production ("dimuon") data. On the other hand, high-precision collider data from the LHC has instead exhibited the opposite trend, with a recent QCD analysis from ATLAS based on the W, Z 7 TeV rapidity distributions from the 2011 dataset finding a strange sea that is in fact larger than the non-strange sea. Given the importance of strangeness for many phenomenological

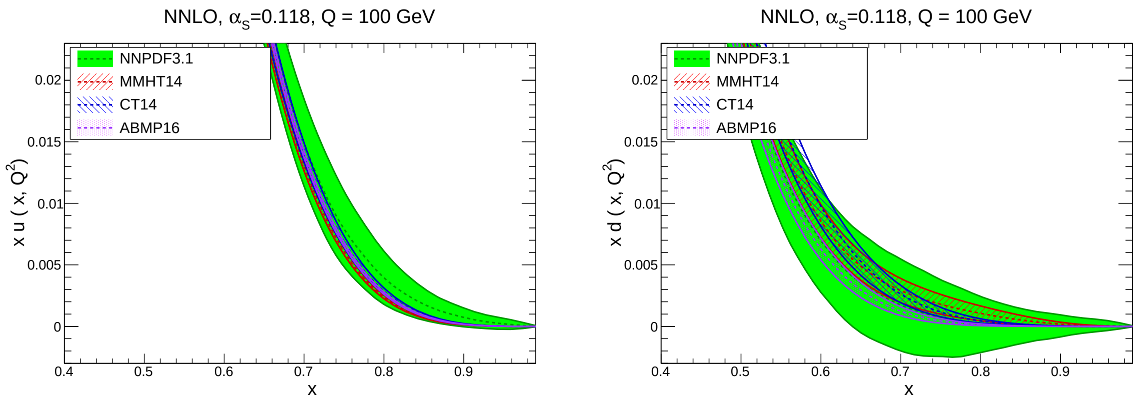


Figure 54: Same as Fig. 52, now focusing on the large-x region of the up quark (left) and down quark (right plot) PDFs.

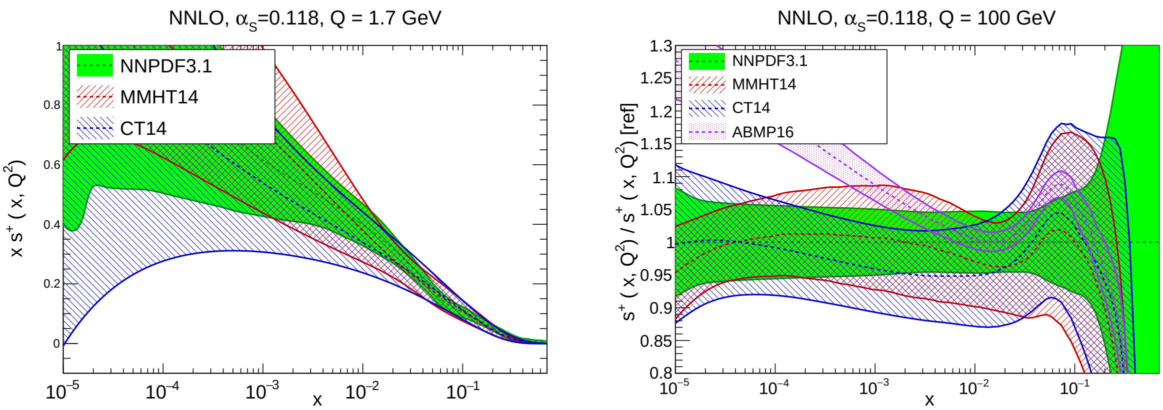


Figure 55: Same as Fig. 50 for the total strangeness $xs^+(x, Q^2)$.

applications, for instance the measurement of the W mass, it will be important to resolve this issue in the future.

In Fig. 55 we show the total strange PDF $xs^+(x, Q^2)$ at $Q = 100$ GeV, in the same format as that of Fig. 50 The strangeness-sensitive datasets included in the four analysis are rather different, both in terms of the neutrino fixed-target data and the LHC collider data. For example, only the ABMP16 fit includes the NOMAD dimuon data [121], while only NNPDF3.1 includes the ATLAS W, Z 2011 rapidity distributions. We can see that there is reasonable agreement within uncertainties between the four groups except for ABMP16 for $x \lesssim 10^{-3}$, which has a much harder strangeness than the other groups. We also note that the differences in the size of the strange PDF uncertainty can vary by up to a factor ~ 5, with ABMP16 having the smallest uncertainties.

A more physically transparent method to assess the strange content of the proton is given by the ratio of the strange to the non-strange sea quark PDFs, defined as

$$R_s(x, Q^2) = \frac{s(x, Q^2) + \bar{s}(x, Q^2)}{\bar{u}(x, Q^2) + \bar{d}(x, Q^2)} \,. \quad (125)$$

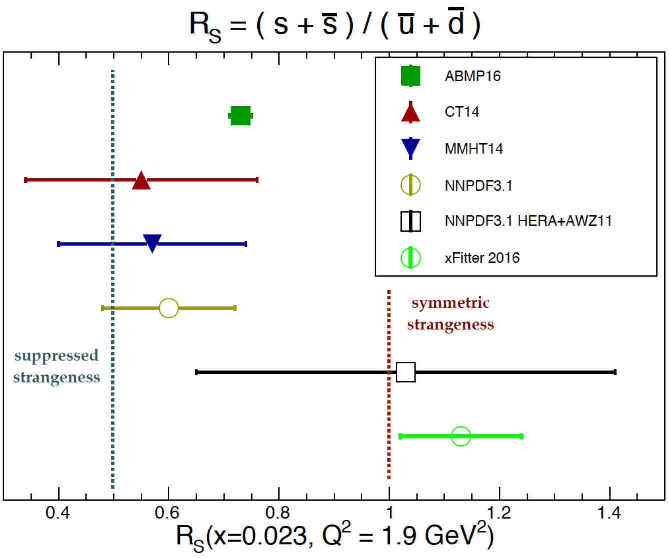


Figure 56: The ratio of strange to non-strange sea quarks $R_s(x, Q^2)$, Eq. (125) for $x = 0.023$ and $Q^2 = 1.9$ GeV^2. We compare the results of various global PDF fits with those of the ATLAS/`xFitter` interpretation study as well as with those of a NNPDF3.1 fit based on the same dataset as the ATLAS study. The vertical lines indicate two possible scenarios for the strange PDFs, namely a suppression of size $R_s \simeq 0.5$ and then a strange sea which is symmetric with the non-strange one, $R_S \simeq 1$.

In this ratio, a symmetric strange sea would correspond to $R_S \simeq 1$, while a suppressed strangeness instead leads to $R_S \ll 1$. Traditionally, the constraints from neutrino dimuon production have suggested a value $R_S \sim 1/2$ in most global fits. In Fig. 56 we show the ratio of strange to non-strange sea quarks $R_s(x, Q^2)$, Eq. (125), for $Q^2 = 1.9$ GeV^2 and $x = 0.023$. We compare the results of the various global PDF fits with those of the ATLAS/`xFitter` study [192], which includes the recent ATLAS W, Z high precision data, as well as with those of a NNPDF3.1 fit based on the same dataset as the ATLAS study. The vertical lines indicate the two possible scenarios for the strange PDFs, namely a suppression of size $R_s \simeq 0.5$ and a strange sea which is symmetric with the non-strange one, $R_S \simeq 1$. We can see that the ABMP16, CT14, MMHT14 and NNPDF3.1 have a preference for a suppressed strangeness. On the other hand, this comparison also shows that if only the HERA and ATLASWZ11 data are considered, the NNPDF3.1 analysis yields an unsuppressed strangeness, although with PDF uncertainties rather larger than for the `xFitter analysis`. This comparison demonstrates that the opposite pull between the low-energy neutrino data and the high-energy collider data is genuine, although the tension is not dramatic, as indicated by the fact that the NNPDF3.1 global and HERA+ATLASWZ11 results agree within PDF uncertainties.

One limitation of the comparison summarised in Fig. 56 is that it is restricted to a specific point $x \simeq 0.023$. Tn Fig. 57 we therefore show the $R_s(x, Q^2)$ ratio as a function of x both a low and at high scales. There are a number of interesting features that can be observed from this comparison. First, we observe that DGLAP evolution automatically increases the value of R_S, since as we go to higher values of Q the sea component dominates over the valence components. Second, we find a consistent strangeness content for the four groups in most of the range of x, although the corresponding uncertainties in each case can vary quite a lot.

Another important point from this comparison is that clearly any statement about whether or not strangeness is suppressed depends on the region of x that is being considered. For instance, in the MMHT14 analysis

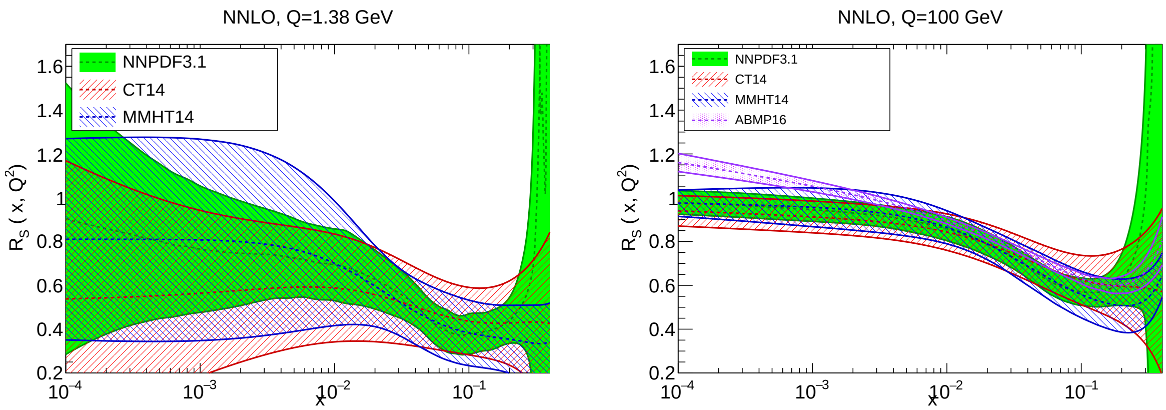


Figure 57: The ratio of strange to non-strange sea quarks $R_s(x, Q^2)$, Eq. (125), as a function of x for $Q = 1.38$ GeV (left plot) and for $Q = 100$ GeV (right plot).

for $Q = 1.38$ GeV the value of R_S changes from around 0.4 at $x \simeq 0.1$ to around 0.8 for $x \simeq 0.007$. So different x regions exhibit different amounts of suppression with respect to the light sea quarks, and therefore the question of the suppression (or lack thereof) of the strange PDF is a more nuanced issue than what is sometimes stated. In any case is clear from the comparison of Fig. 57 that a symmetric strange sea in the entire range of x is not favoured by any of the four fits shown here, in particular in the region around $x \simeq 0.1$ and above. In this respect, future data from the LHC will help to to shed some light on this important issue.

6.4. *The charm content of the proton*

The charm content of the proton is a topic that has recently received quite a lot of attention (see [408] for a review). As discussed in Sect. 4, there are two different approaches to treat the charm PDF within the global QCD analysis. On the one hand, one can assume that the charm PDF is generated entirely from perturbative evolution, and thus compute the charm PDF from the gluon and light quark PDFs starting from the charm threshold $\mu_c \simeq m_c$ by means of the DGLAP evolution equations. On the other hand, it is also possible to release this assumption and treat the charm PDF on an equal footing to the light quark PDFs, namely introducing a functional form for $c(x, Q_0)$ with parameters to be determined by experimental data.

Until recently, in most global fits the charm PDF was generated using the perturbative evolution *ansatz*, and then separately, in dedicated intrinsic charm studies, variants of these global fits were performed with specific models for the charm PDF. In these studies, the parameters of the model charm PDF, typically its overall normalization, where constrained by experimental data, see for instance Refs. [409, 367, 26, 410]. An alternative approach is taken by the NNPDF3.1 global analysis, which fits the charm PDF using the same parametrization as for the light quarks. In all cases, the dominant constraints on the fitted charm PDF come from processes sensitive to initial-state charm, such as the charm structure functions of the EMC experiment [113], other fixed-target DIS datasets, and collider electroweak gauge boson production.

For the first approach, the CT14IC analysis provides a recent and representative example. Here, the charm PDF is parametrized according to two theoretical scenarios. First, using either the exact of the approximate BHPS model [384], which predicts a valence-like charm PDF at the input scale. In the case of the approximate solution of the model, we have

$$c(x, Q_0) = \frac{1}{2}Ax^2\left[\frac{1}{3}(1-x)(1+10x+x^2) - 2x(1+x)\ln(1/x)\right] , \tag{126}$$

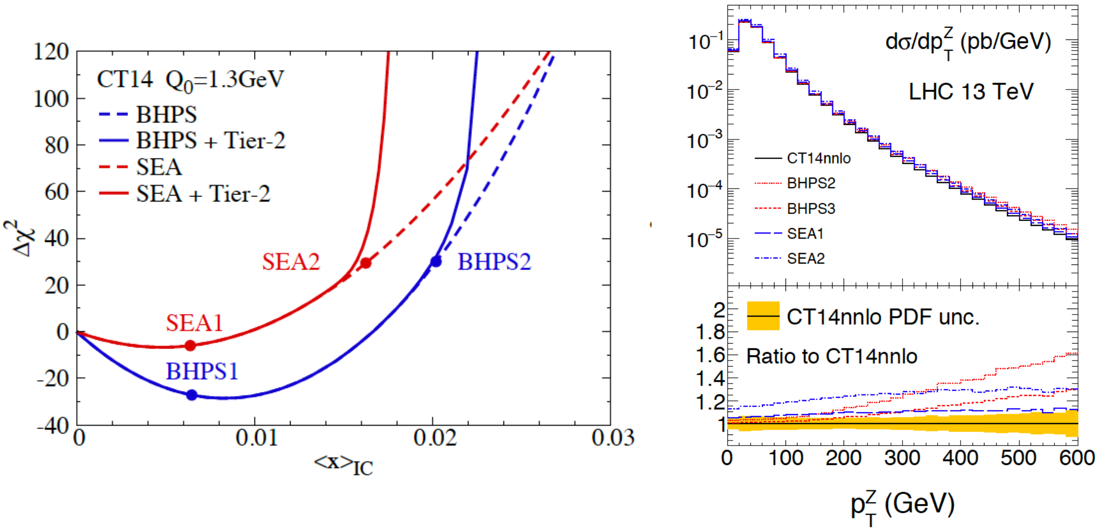


Figure 58: Left: the deviation of the χ^2 in the CT14IC fits, with respect to the best-fit value of the CT14 fit with perturbative charm, as a function of $\langle x\rangle_{\rm IC}$ Results are shown for the BHPS and SEA models, with the "1" points labeling the preferred value of $\langle x\rangle_{\rm IC}$, with those labelled with "2" indicate the largest values of the charm momentum fraction allowed by the fit tolerance criteria. Right: the transverse momentum distribution of Z bosons in the $pp \to Z + c$ process at 13 TeV, comparing the CT14 NNLO result with various of the CT14 IC models, as a function of p_T^Z.

while a more complicated, non-analytic expression is used for the exact BHPS solution. The other scenario explored in the CT14IC study is the SEA model, where the charm PDF is parametrized by a "sea-like" function taken to be proportional to the light quark sea PDFs, namely

$$c(x, Q_0) = A\left(\bar{d}(x, Q_0) + \bar{u}(x, Q_0)\right) . \tag{127}$$

In these models, the overall normalization A of the fitted charm is a free parameter to be determined from the experimental data. In Fig. 58 we plot the deviation of the χ^2 in the CT14IC fits, with respect to the best-fit value of the CT14 fit with perturbative charm, as a function of the charm momentum fraction $\langle x\rangle_{\rm IC} = C(Q_0)$, where we have defined

$$C(Q^2) \equiv \int_0^1 dx\, x\left(c(c, Q^2) + \bar{c}(x, Q^2)\right) . \tag{128}$$

Results are shown for the BHPS and SEA models, with the '1' points labeling the preferred value of $\langle x\rangle_{\rm IC}$, with those labelled with '2' indicating the largest values of the charm momentum fraction allowed by the fit tolerance criteria. We observe that the BHPS model is preferred, leading to a best-fit value of $\langle x\rangle_{\rm IC} \simeq 0.6\%$.

As mentioned above, a different approach to fitted charm is adopted by the NNPDF collaboration [25, 251]. In this case, the charm PDF is treated on an equal footing to the light quark PDFs, and therefore it is parametrized with a 37-parameter artificial neural network with 2-5-3-1 architecture,

$$c^+(x, Q_0) = c(x, Q_0) + \bar{c}(x, Q_0) = x^{a_{c^+}}(1-x)^{b_{c^+}}\,{\rm NN}_{c^+}(x)\,, \tag{129}$$

where a_{c^+}, b_{c^+} are the corresponding preprocessing exponents, whose range is determined by an iterative procedure. The charm asymmetry, on the other hand, is assumed to vanish $c^-(x, Q_0) = 0$. The charm PDF is then determined at the input evolution scale $Q_0 = 1.64$ GeV from the experimental data, finding

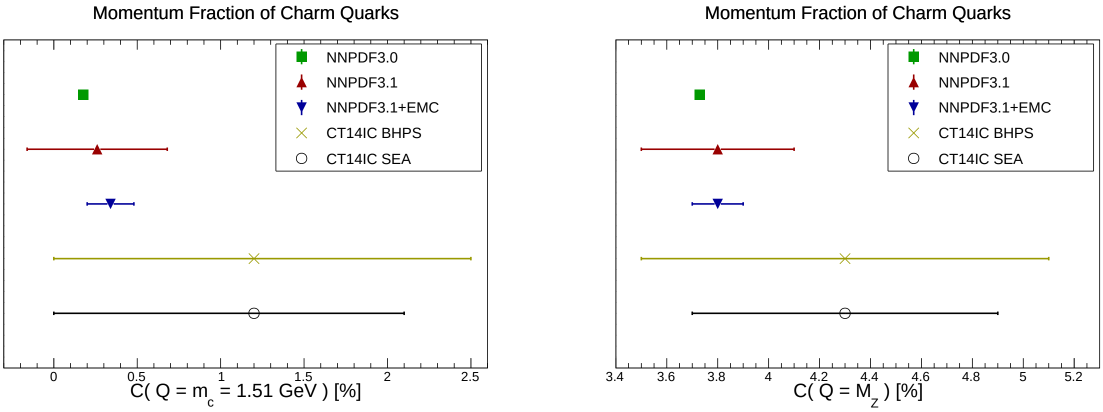


Figure 59: The momentum fraction carried by charm quarks, $C(Q)$ Eq. (128), both at a low scale $Q = 1.51$ GeV (left) and at a high scale $Q = M_Z$ (right plot). We compare NNPDF3.0 (perturbative charm) with NNPDF3.1 (based on fitted charm) with and without the inclusion of the EMC charm data, as well as with the BHPS and SEA scenarios of the CT14IC fits. See text for more details.

that the recent LHC high-precision electroweak gauge boson production measurements provide the best constraints [251]. DIS structure functions are treated with the FONLL general-mass VFN scheme, modified to account for initial-state massive contributions [411, 412].

One of the potential benefits of this model-independent approach is that it improves the stability of the fitted PDFs with respect to the value of the charm mass m_c, since its variations can be re-absorbed into the fitted charm boundary condition. To illustrate this point, in Fig. 60 we show the dependence of the quark-antiquark PDF luminosity at the LHC 13 TeV in the NNPDF3IC fits with the value of the charm mass m_c used in the fit. We find that even for a relatively large variation of $\delta m_c = \pm 0.14$ GeV, the $q\bar{q}$ luminosity is very stable in most of the M_X range.

The amount of charm present inside the proton is most usefully quantified by its momentum fraction, defined in Eq. (128). In the case of perturbative charm, by construction we have $C(Q^2 < \mu_c^2) = 0$, while if there is a non-perturbative charm component in the proton in general we have $C(Q^2) \neq 0$ at all values of the scale Q^2. In Fig. 59 we show the momentum fraction carried by charm quarks both at a low scale $Q = 1.51$ GeV and at a high scale $Q = M_Z$, comparing the results from NNPDF3.0, based on perturbative charm, with those from NNPDF3.1, based on fitted charm, with and without the inclusion of the EMC charm data, as well as with the BHPS and SEA scenarios of the CT14IC fits. In the latter case, the uncertainty bands indicates the range between no intrinsic charm and the maximum amount of IC allowed within the CT14 tolerance, with the central value corresponding to the best-fit.

The comparisons of Fig. 59 highlight first of all that when the charm PDF is generated perturbatively its uncertainties are very small, this is not necessarily the case once it is fitted. Reassuringly, once charm is fitted (NNPDF3.1), the results with perturbative charm (NNPDF3.0) are consistent within uncertainties. The NNPDF3.1 analysis also finds that while adding the EMC charm data helps in reducing the PDF uncertainties on $\langle x \rangle_{\rm IC}$ by around a factor 3, even without it one achieves a quite competitive charm determination, due to the precision collider electroweak data. The CT14IC results are consistent within PDF errors with the NNPDF3.1IC, although the highest values within the CT14IC are disfavored by the latest LHC data. The rapid growth of $C(Q)$ from $Q = 1.51$ GeV to $Q = M_Z$, driven by the perturbative component, is also clear.

In Fig. 60 we show a comparison of the fitted charm PDF at $Q = 1.65$ GeV between the NNPDF3IC set and the different model scenarios considered in the CT14IC analysis. We see that NNPDF3IC prefers a valence-like shape, along the lines of the BHPS model, though uncertainties are still large. The CT14IC BHPS results tend to have the maximum at slightly lower values of x; note also that they develop a perturbative tail since the plot is performed at a value $Q > Q_0$. The CT14IC SEA models predict that the enhancement of the charm PDF is localized at medium and small-x, while in the valence region the fitted charm agrees with the perturbative *ansatz*.

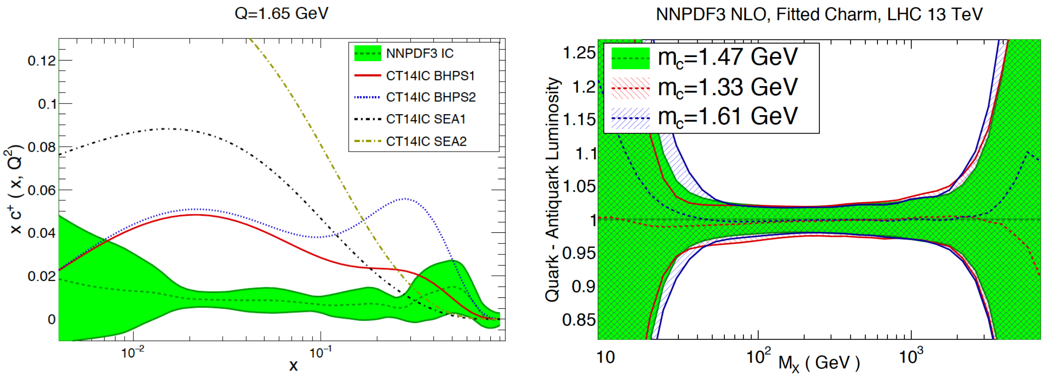


Figure 60: Left: comparison of the fitted charm PDF at $Q = 1.65$ GeV between the NNPDF3IC set and the different models of the CT14IC analysis. Right: the dependence of the quark-antiquark PDF luminosity at the LHC 13 TeV in the NNPDF3IC fits with the value of the charm mass m_c used in the fit.

If the charm content of the nucleon is indeed different from the one predicted by the perturbative ansatz, there are a number of phenomenological consequences that could be studied at the LHC. To begin with, it would modify the kinematic distributions of the Z bosons in the $pp \to Z + c$ process [413, 414, 415], leading to an enhancement of the cross-section which grows with the value of p_T^Z. To illustrate this point, in Fig. 58 we show the transverse momentum distribution of Z bosons in the $pp \to Z + c$ process at $\sqrt{s} = 13$ TeV, comparing the CT14 NNLO result with the CT14 IC models, as a function of p_T^Z. Depending on the specific model considered, enhancements of up to 50% are predicted. A closely related process is photon production in association with charm mesons [416, 417], which is however theoretically less clean as it is affected by the poorly-understood parton-to-photon fragmentation component. Another important process where intrinsic charm would make a difference is open D meson production [418, 419], in particular at large p_T and at forward rapidities, which enhance the sensitivity to the large-x region. By exploiting the information from these various processes, we can hope in the future to shed more light on this topic.

7. QED corrections and the photon PDF

In this section we explore a topic that have received a lot of attention in PDF fits in the recent years, namely the role of QED and weak corrections, in particular concerning photon-initiated processes. Here first of all we discuss the role of QED corrections and the photon PDFs, and then we review pure weak corrections to hard scattering matrix elements arising from virtual massive weak boson exchange.

7.1. Photon-induced processes

It has been over a decade since the calculation of the splitting functions at NNLO in α_s [78, 79] provided the necessary tools to be able to carry out NNLO PDF fits. Moreover, we have seen in Sect. 3 that for the

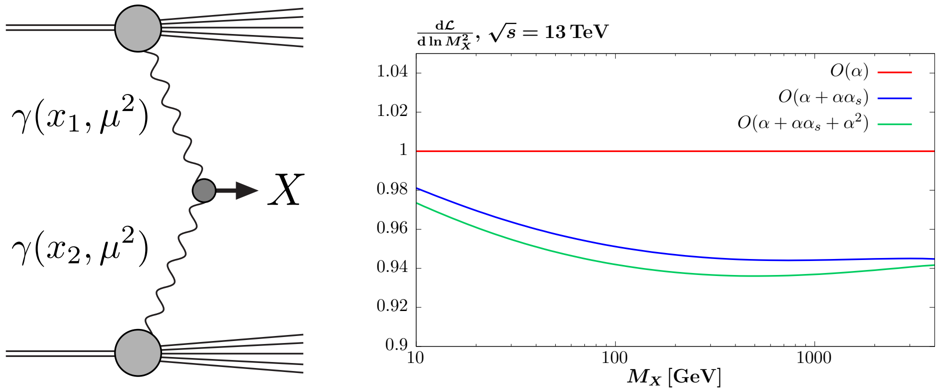


Figure 61: Left: schematic diagram for the photon–initiated production of a system X, and the corresponding photon PDFs. Right: the $\gamma\gamma$ luminosity as a function of the invariant mass, M_X, of the produced final-state. The ratio to results with $O(\alpha\alpha_S)$ and $O(\alpha^2)$ to the leading $O(\alpha)$ DGLAP evolution shown. Calculated using the approach described in [376].

majority of PDF sensitive observables, the perturbative calculation calculation is available at this NNLO order. Given that the data from the LHC are available at increasing precision, to below the percent level, NNLO PDF fits are essential to match this unprecedented precision and have naturally become the standard. However, a simple counting of powers of α_S indicates that

$$\alpha_s^2(M_Z) \sim \frac{1}{70}\,, \qquad \alpha_{\rm EM}(M_Z^2) \sim \frac{1}{130}\,. \tag{130}$$

That is, we may roughly expect the NNLO QCD and NLO EW corrections to be of the same order of magnitude. While such an argument clearly neglects the non–trivial differences in the structure of the QCD and EW corrections, this nonetheless serves as a warning that we must at least consider the impact of going to NLO EW if we are to claim percent-precision to LHC cross-sections.

A specific type of EW correction of particular relevance to PDF studies is the contribution from photon–initiated processes, such as those shown schematically in Fig. 61. As this involves a photon in the initial state, this requires the introduction of a PDF for the photon in the proton[4]. This is included in complete analogy to the QCD partons, and moreover as it involves a massless boson in the initial–state, higher order QED $q \to q\gamma$ and $\gamma \to q\bar{q}$ splitting will generate collinear singularities that must be absorbed into the corresponding PDFs. In other words, this will produce QED corrections to the DGLAP evolution of the PDFs. Another important type of EW corrections relevant for PDF fits, namely those associated to virtual massive weak boson exchange, are discussed in Sect. 7.2.

[4]For brevity, we will refer to this throughout as the photon PDF, but this should not be confused with the partonic content of the photon itself, which often receives a similar label, see e.g. [420].

QED corrections to DGLAP evolution

The introduction of the photon PDF requires the following straightforward extension of the DGLAP evolution equations,

$$
\begin{aligned}
Q^2 \frac{\partial}{\partial Q^2} g(x, Q^2) &= \sum_{q,\bar{q},g} P_{ga}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{g\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2), \\
Q^2 \frac{\partial}{\partial Q^2} q(x, Q^2) &= \sum_{q,\bar{q},g} P_{qa}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{q\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2), \\
Q^2 \frac{\partial}{\partial Q^2} \gamma(x, Q^2) &= P_{\gamma\gamma} \otimes \gamma(x, Q^2) + \sum_{q,\bar{q},g} P_{\gamma a}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2).
\end{aligned}
\tag{131}
$$

The splitting functions can then be expanded in powers of both the QCD and QED coupling

$$
P_{ij} = \sum_{m,n} \left(\frac{\alpha_S}{2\pi}\right)^m \left(\frac{\alpha}{2\pi}\right)^n P_{ij}^{(m,n)} \,. \tag{132}
$$

The lowest order QED splitting function $P_{\gamma q}^{(0,1)}$ is due to the same type of Feynman diagram as in the LO QCD case for $P_{gq}^{(1,0)}$, with the gluon simply replaced by a photon, and similarly for P_{qq} and $P_{q\gamma}$. Thus these are trivially related by suitable adjustments of the colour factors and inclusion of the electric charges e_q of the quark, with

$$
P_{qq}^{(0,1)} = \frac{e_q^2}{C_F} P_{qq}^{(1,0)} \,, \qquad P_{q\gamma}^{(0,1)} = \frac{e_q^2}{T_R} P_{qg}^{(1,0)} \,, \qquad P_{\gamma q}^{(0,1)} = \frac{e_q^2}{C_F} P_{gq}^{(1,0)} \,, \qquad P_{\gamma\gamma}^{(0,1)} = -\frac{2}{3} \sum_f e_f^2 \, \delta(1-x) \,, \tag{133}
$$

where for the $P_{\gamma\gamma}$ case only the Abelian contribution is present and the sum is over all fermions in the loop, that is quarks and leptons[5]. The calculation of the $O(\alpha_s\alpha)$ $(m, n = 1)$ terms, where the $P_{g\gamma}$ and $P_{\gamma g}$ splittings enter for the first time, is given in [422], while the $O(\alpha^2)$ $(m = 0, n = 2)$ terms are given in [423]. Publicly available implementations of the DGLAP evolution including these QED corrections are provided by the `APFEL` [424] and `QEDEVOL` [425] packages.

The impact of the $O(\alpha\alpha_S)$ and $O(\alpha^2)$ corrections on the $\gamma\gamma$ luminosity, defined as

$$
\mathcal{L}_{\gamma\gamma} = \frac{1}{s} \int_{M_X^2/s}^1 \frac{\mathrm{d}x}{x} \, \gamma(x, M_X^2) \, \gamma\left(\frac{M_X^2}{xs}, M_X^2\right) \,, \tag{134}
$$

is shown in Fig. 61. The $O(\alpha\alpha_S)$ corrections have a fairly small but clearly non–negligible impact on the luminosity, giving up to a $\sim 5\%$ negative correction. As we would expect, the $O(\alpha^2)$ corrections are significantly smaller, but can reach the percent level. Note that in general these corrections will depend on the PDF set used. The results of Fig. 61 have been computed using the MMHT framework [376], which is closely based on the LUXqed formalism described below. Using the NNPDF3.0QED instead, the $O(\alpha)$ result for $M_X = 200$ GeV is lower, with $R \simeq 0.9$.

As discussed in detail in [426], the $P_{\gamma\gamma}$ self–energy contribution to the DGLAP evolution of the photon PDF is intimately connected to the choice of renormalization scale for the initial–state photon coupling

[5]To be consistent, and in particular to preserve momentum fully, this requires the introduction of lepton PDFs in the proton. However as discussed in [421] the contribution from these is generally of limited phenomenological relevance, and can be safely neglected. Note in any case that including leptons in the running of the QED coupling $\alpha(Q)$ is still required.

to the hard process. It is well known in QED that for on–shell external photons the coupling receives no renormalization, and is completely determined to be $\alpha(0)$. However, the $P_{\gamma\gamma}$ term breaks this simple picture, and we should instead use $\alpha(\mu_F)$ in the caslculation. Physically, the photon substructure is being resolved by the introduction of a photon PDF and the contribution from $\gamma \to q\overline{q}$ splittings in the evolution, such that a purely on–shell renormalization scheme is no longer appropriate. This has been confirmed at NLO EW order in [427], where it is shown that using the on–shell scheme leads to uncanceled fermion–mass singularities in the hard cross section.

The photon PDF

The first attempts at describing the photon PDF can be divided into two distinct categories, either being phenomenological approaches that model the photon PDF, as in the MRST2004QED [428] and more recent CT14QED [368] sets, or the data-driven approach of the NNPDF2.3/3.0QED [429, 430] sets. The first attempt to include the photon in a PDF set was provided by MRST2004QED. Here, a simple model for the photon PDF at input scale Q_0 due to one–photon emission off the valence quarks was assumed. In other words, the quark valence distributions were frozen at Q_0 and the LO QED DGLAP evolution for the photon is integrated between the light quark mass m_q and Q_0, so that

$$\gamma(x, Q_0^2) = \frac{\alpha}{2\pi}\left[\frac{4}{9}\log\left(\frac{Q_0^2}{m_u^2}\right)u(x, Q_0) + \frac{1}{9}\log\left(\frac{Q_0^2}{m_d^2}\right)d(x, Q_0)\right] \otimes \frac{1 + (1 - x)^2}{x}\ . \tag{135}$$

The CT14QED set generalised this approach, allowing additional freedom in the normalization of the photon, which was fit to ZEUS data [369] on isolated photon production[6]. Thus, within such approaches the photon PDF is completely *predicted* within the specific model, up to any freedom in the model parameters, such as the choice of quark masses for MRST2004QED or the overall normalization for CT14QED.

On the other hand, the NNPDF2.3QED [429] set (subsequently updated to NNPDF3.0QED [430]), freely parameterises the photon PDF at input. In other words, the photon is treated on exactly the same footing as the QCD partons. This is then extracted from a fit (or more precisely, a Bayesian reweighting) to DIS and LHC W, Z data; in the former case the constraint comes purely from the effect on the PDF evolution, with no explicit photon–initiated contribution included. However, in general the contribution of photon–initiated process are small, leading to significant uncertainties on the extracted photon PDF.

More recently, there has been a great deal of progress in our understanding of the photon PDF. One crucial point that was missed in the above approaches is the long range nature of QED. That is, the proton is itself an electrically charged object which can coherently emit a photon, with the proton remaining intact afterwards. The possibility for such elastic photon emission is of course very well established. Elastic *ep* scattering is an extremely well measured process, providing for example the first measurement of the proton charge radius [33, 34] in the 1950s, with further precise measurements of this process [431] continuing to this day. Theoretically, the well known equivalent photon approximation (EPA) [432] provides a precise foundation for describing the elastic scattering process in terms of a flux of coherently emitted photons from the proton.

The connection of this fact to the photon PDF was discussed in [433] and more recently in [434, 435]. Following the equivalent photon approximation, it is straightforward to show that elastic photon emission leads to a contribution to the photon PDF at a scale Q_0 ~1 GeV given by

$$\gamma_{\rm el}(x, Q_0^2) = \frac{1}{x}\frac{\alpha}{\pi}\int_{\frac{x^2 m_p^2}{1-x}}^{Q_0^2}\frac{{\rm d}Q^2}{Q^2}\left[\left(1 - x - \frac{x^2 m_p^2}{Q^2}\right)F_E(Q^2) + \frac{x^2}{2}F_M(Q^2)\right]\ , \tag{136}$$

[6] In fact, as we will discuss below, this has been supplemented with the elastic component to give the inclusive set CT14QEDinc.

where F_E and F_M are the elastic and magnetic form factors of proton, which are related to the electric and magnetic charge distributions in the proton. These are steeply falling functions of Q^2 that are probed very precisely in a range of elastic ep scattering experiments, see e.g. [431].

To demonstrate the connection of this elastic component to the inclusive photon PDF, if we for simplicity omit the small backreaction that the photon has on the quark and gluon PDFs via the evolution equations, then we can solve Eq. (131) to get [436]

$$\gamma(x,\mu^2) = \frac{1}{\alpha(\mu^2)}\left(\alpha(Q_0^2)\,\gamma(x,Q_0^2) + \int_{Q_0^2}^{\mu^2}\frac{\mathrm{d}Q^2}{Q^2}\alpha(Q^2)\sum_{q,\bar{q},g} P_{\gamma a}(x,\alpha_s(Q^2))\otimes f_a(x,Q^2)\right)\,, \tag{137}$$

$$\equiv \gamma_{\rm input}(x,\mu^2) + \gamma_{\rm evol}(x,\mu^2)\,. \tag{138}$$

Thus the photon is given separately in terms of an input at low scale Q_0 and an evolution component due to the usual DGLAP $q \to q\gamma$ emission for $Q^2 > Q_0$. The latter is completely determined in terms of the quark and gluon PDFs, leaving the input photon at Q_0, which is dominantly due to elastic emission. Thus this already provides quite a strong constraint on the photon PDF; as we will see below, the impact in comparison to the model-independent NNPDF approach can be dramatic.

However, even for relatively low photon virtualities, $Q^2 < Q_0$, the emission may also be inelastic, such that the proton no longer remains intact afterwards. In other words we have

$$\gamma(x,Q_0^2) = \gamma_{\rm el}(x,Q_0^2) + \gamma_{\rm inel}(x,Q_0^2)\,, \tag{139}$$

In [434, 435] fairly simple phenomenological models for this inelastic component, given by suitable generalizations of (135), were taken, while the CT14QED set allowed an additionally free normalization to be fitted to ZEUS data on isolated photon production, as described above.

Given that the elastic component is directly determined from the form factors F_E and F_M, that are themselves measured from elastic ep scattering, it is natural to ask whether the inelastic component can be similarly determined. In other words, rather than relying on a phenomenological model, can $\gamma_{\rm inel}$ instead be calculated directly from the form factors for inelastic ep scattering, that is, from the proton structure functions? In the analysis of [24] it was shown that this is indeed the case, with the corresponding `LUXqed` PDF set made publicly available. In particular, they find that the photon PDF can be expressed as[7]

$$x\gamma(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)}\int_x^1\frac{dz}{z}\left\{\int_{\frac{x^2m_p^2}{1-z}}^{\frac{\mu^2}{1-z}}\frac{dQ^2}{Q^2}\alpha^2(Q^2)\left[\left(zp_{\gamma q}(z)+\frac{2x^2m_p^2}{Q^2}\right)F_2\left(\frac{x}{z},Q^2\right)-z^2F_L\left(\frac{x}{z},Q^2\right)\right]\right.$$
$$\left.-\,\alpha^2(\mu^2)z^2F_2\left(\frac{x}{z},\mu^2\right)\right\}. \tag{140}$$

Thus in this formalism the photon PDF is a derived quantity, which can be written purely in terms of the inclusive DIS structure functions, which are known quite precisely from the experimental point of view. More recently, the detailed study of [439] has demonstrated how this expression may be derived in a process independent way by using the operator definition of the photon PDF, as well as generalising this expression to the case of the polarized and transverse momentum dependent PDFs. This approach is also shown to provide quite simple derivation of the known $O(\alpha\alpha_s)$ and $O(\alpha^2)$ splitting functions, $P_{\gamma i}$.

[7]Following the publication of [24] it was discovered that this expression had been derived in the earlier papers of [437, 438], but without the correct limits on the Q^2 integral.

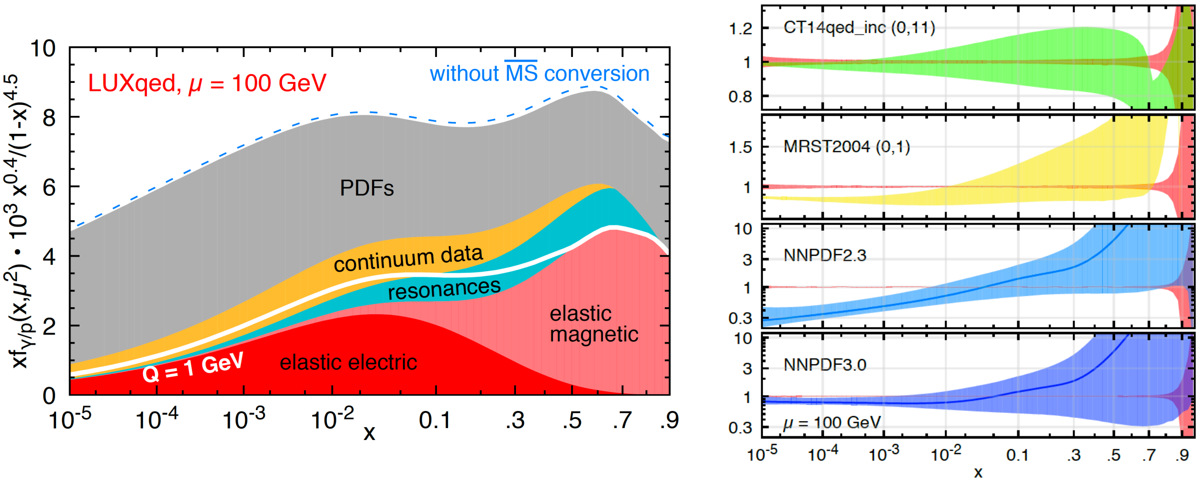


Figure 62: Left: overview of the various contributions to the photon PDF $\gamma(x, Q^2)$ in the LUXqed approach as a function of x at $Q = 100$ GeV. Right: comparison of the photon PDFs from CT14qed_inc, MRST2004, NNPDF2.3/3.0 and LUXqed, normalized to the central value of the latter.

While the connection of Eq. (140) to the considerations above is not immediately clear, some similarity in the overall form with Eq. (137) is apparent. Indeed, by substituting for $F_{2,L}$ in terms of the quark and gluon PDFs, at high Q^2 this readily reduces to $\gamma_{\rm evol}$ in (137); indeed this is how the `LUXqed` photon PDF is calculated in this region. In addition, using the known expression for the elastic contributions to $F_{2,L}$ reproduces Eq (136) when combined with Eq. (137); this elastic contribution is also included. By using fits to the experimentally determined inelastic structure functions at low Q^2, including in the resonance region, it is shown in [24] that the remaining inelastic contribution, and therefore the photon PDF in its entirely, is very precisely determined. In Fig. 62 we show an overview of the various contributions to the photon PDF $\gamma(x, Q^2)$ in the LUXqed approach as a function of x at $Q = 100$ GeV. We see that at small-x it is dominated by the PDF contribution, while at large-x the elastic contribution accounts for up to half of the size of $\gamma(x, Q)$.

It is worth emphasizing that the expression Eq. (140) does not rely on any explicit separation between an input and evolution component to the photon as in (137), and corresponds to the exact result for the photon within the quoted accuracy of [24, 439], valid to all orders in QED and QCD, and including non–perturbative corrections. Indeed, applying standard DGLAP above the starting scale Q_0 terms the power–like $\sim m_p^2/Q^2$ correction would be missed, while for $Q^2 > Q_0^2$ the contribution from the elastic component would be omitted and the inelastic resonance component, which contributes at higher x in this region, would not be correctly modelled.

However, from the point of view of a global PDF set it may be preferable to use Eq. (140) in a form that can be more directly implemented within the standard fitting framework. That is, applying this approach after suitable modification to calculate the input photon, which can then be included as part of the default input parton set for any future fits and studies, see [376] for initial discussion. An alternative iterative approach is proposed in [439].

To illustrate the differences and similarities between these various determinations of $\gamma(x, Q)$, in Fig. 62 we show the comparison of the photon PDFs from CT14qed_inc, MRST2004, NNPDF2.3/3.0 and LUXqed, normalized to the central value of the latter. It is clear from this comparison that the theoretical uncertainties associated with the LUXqed determination are much smaller than in any other of the previous

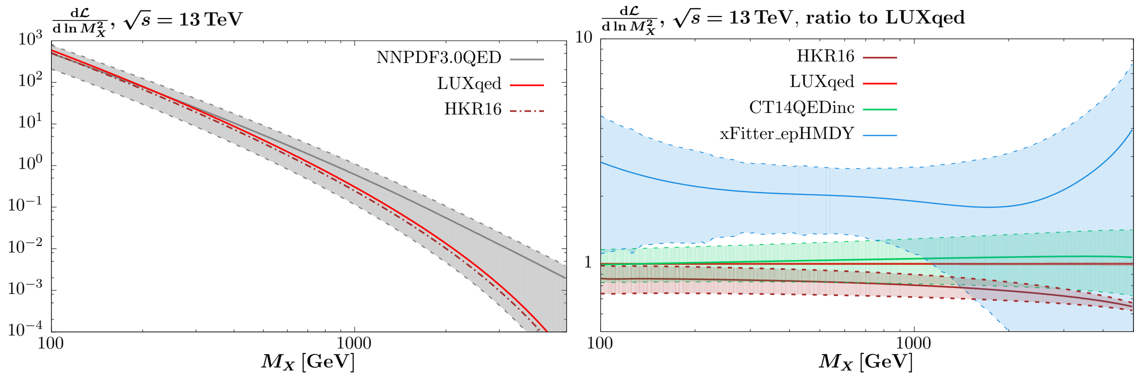


Figure 63: $\gamma\gamma$ luminosities at $\sqrt{s} = 13$ TeV. (Left) Absolute values for the HKR16, NNPDF3.0QED and LUXqed sets. (Right) Ratios of the CT14QED, HKR16 and `xFitter_HMDYep` sets to the LUXqed prediction. 68% uncertainty bands are shown, with the exception of the HKR16 set, where the error is due to model variation in the inelastic input (lower edge corresponds to elastic only).

approaches. Interestingly, the LUXqed pdf is in good agreement within uncertainties with the model-independent NNPDF3.0QED fit in the entire range of x, though the PDF uncertainties are much larger in the latter case specially at large x.

Phenomenology

In Fig. 63 (right) we show the NNPDF3.0QED $\gamma\gamma$ luminosity at $\sqrt{s} = 13$ TeV, including the 68% C.L. error bands. A large PDF uncertainty is evident, in particular at higher system mass M_X. As discussed above, the input component in Eq. (138) is poorly determined within this approach, due to limited constraints placed by the available experimental data. It is therefore unsurprising that the PDF errors should be most significant at higher mass, as here the relative contribution from this input component is larger, due the reduced phase space for PDF evolution. In addition, the central value of the luminosity is seen to lie towards the upper end of the uncertainty band. As discussed in [15, 435], this exhibits a much gentler decrease with M_X in comparison to the QCD parton luminosities. However, also plotted is the LUXqed result, and the difference is dramatic. The central value lies towards the lower end of the NNPDF band at higher mass, with a PDF uncertainty that is smaller than the line width of the plot. We also show the result of [435], labelled HKR16, which demonstrates a similar trend. Thus, simply applying basic physical conditions on the form of the photon PDF, and including the dominant coherent input Eq. (136) gives a qualitatively similar result.

Taking a closer look, in Fig. 63 (right) we show the ratio of the HKR16, CT14QED and xFitter_HMDYep results to the LUXqed luminosity. The `xFitter_HMDYep` set is extracted in [400] by applying a similar agnostic methodology to NNPDF, but including the more constraining ATLAS high mass Drell–Yan data [440] in the fit; this therefore represents the most up to date set within such an approach. Again, the LUXqed uncertainty band is barely visible on the curve, varying from $1-2\%$ over the considered mass interval. The CT14QEDinc prediction, which includes an elastic component, is consistent, but with larger $\sim 20-40\%$ uncertainties, due to the more limited constraints placed by the ZEUS isolated photon production data on the inelastic input[8]. The HKR prediction lies somewhat below the LUXqed result, outside of the quoted

[8] In addition, the ZEUS data are selected by requiring that at least on track associated with the proton side is reconstructed. This will remove the elastic component entirely, however while CT14QED extract the inelastic component only from this data, at least part of the inelastic will also be removed by this cut.

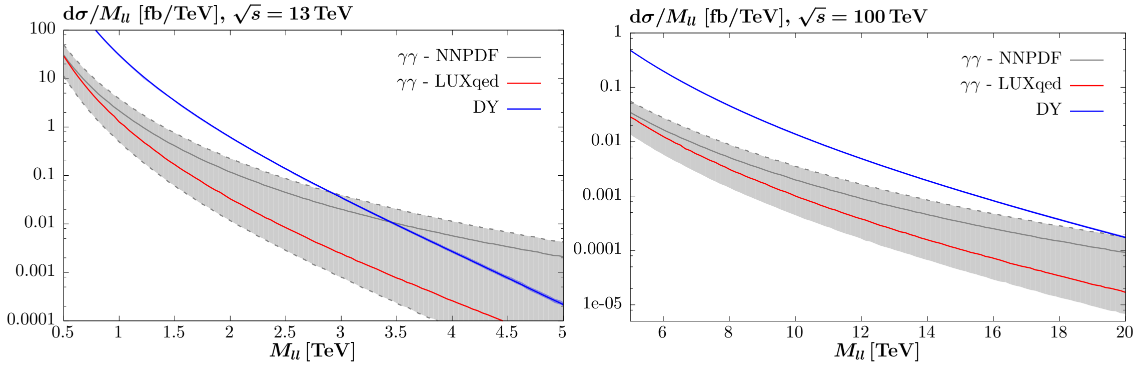


Figure 64: The differential lepton pair production cross sections at $\sqrt{s}$ = 13 TeV and 100 TeV with respect to the invariant mass of the pair M_{ll}, for lepton $|\eta| < 2.5$ and $p_\perp > 20$ GeV. The photon–initiated contributions predicted using the LUXqed and NNPDF3.0QED sets, including the 68% C.L. uncertainty bands. The NLO Drell–Yan cross section, calculated with `MCFM` [216], is also shown.

model variation band, in particular at larger M_X. This is due in large part to the lack of any explicit resonant contribution in the inelastic input, which becomes more important at higher x and hence M_X. Finally, the xFitter_HMDYep set is seen to have a sizeable uncertainty band (albeit smaller than the NNPDF3.0 set [400]), and interestingly appears to lie somewhat above the LUXqed result. From this it is clear that any attempt to extract the photon PDF within such an approach will almost certainly not be competitive. More generally, we can see that the LUXqed set exhibits by far the smallest PDF uncertainties.

Prior to these most recent developments, a range of phenomenological studies pointed out similar trends in the NNPDFQED predictions for the photon–initiated contributions to lepton and W pair [15, 441, 442] and $t\bar{t}$ [286] production. At high system invariant mass these could be significant, and even dominant over the standard channels, with a large PDF uncertainty. From Fig. 63 the reason for this is clear, being driven by the large PDF uncertainty in the $\gamma\gamma$ luminosity at high mass, and the relatively gentle decrease with mass in the central value. However, from the discussion above we know that using the NNPDF set will dramatically overestimate the uncertainty on the photon–initiated contribution. In Fig. 64 we show the lepton pair production cross section at high mass, at the $\sqrt{s}$ = 13 TeV LHC and a $\sqrt{s}$ = 100 TeV FCC. We can see that indeed at the LHC, the NNPDF prediction for the photon–initiated contribution can even be larger than the standard Drell–Yan contribution. However, the up–to–date LUXqed prediction exhibits no such behaviour. The prediction is under good theoretical control, and gives a small, though not negligible, contribution.

Thus, by considering the physics that generates the photon PDF, and recognising the dominance of the elastic emission process, we already achieve a significant reduction in PDF uncertainty in comparison to the model-independent approach, even when accounting for the most sensitive data available in the latter case. Moreover, the additional information provided by Eq. (140) in combination with the high precision data on the inelastic (and elastic) proton structure functions provides extremely tight constraints on the photon PDF, resulting in a $\sim 1\%$ level PDF uncertainty. It is worth emphasising that while consistency tests are of course to be encouraged, this is not the result of a particular theoretical model, to be treated on the same footing as older PDF sets. The LUXqed set is a fundamentally experimental determination of photon PDF; it is simply that by doing this directly in terms of the measured structure functions the tightest constraints can be achieved. Such information must be included in any future photon PDF set, and we have therefore moved beyond the era of large photon PDF uncertainties. Indeed, the photon PDF, which used to be the poorest known of all the proton PDFs, now has the smallest uncertainty.

7.2. Electroweak corrections

In addition to the QED photon–initiated corrections discussed above, it can also be important to include other EW contributions, in particular those arising from virtual EW bosons, in a PDF fit. These corrections are most important at larger invariant masses of the produced system, $Q \gg M_W$, where virtual EW contributions receive logarithmic enhancements, see Ref. [443] for a review. In particular, the virtual exchange of soft or collinear weak bosons leads to Sudakov logarithms of the form $\alpha_W \ln^2 Q^2/M_W^2$, where $\alpha_W = \alpha/\sin^2\theta_W$, which can lead to large (negative) corrections for large values of Q. Given that many of the LHC datasets that enter into the global PDF are sensitive to the TeV region, from high-mass Drell-Yan production to the large p_T tail of Z production and inclusive jets and dijets, the inclusion of such EW corrections is in general required to achieve the best possible description of experimental data in this region.

The state-of-the-art for EW corrections is NLO, that is, $O\left(\alpha_W^2\right)$, including in addition in some cases mixed terms of the form $O(\alpha_W\alpha_s)$ and related terms. These corrections are available for most of the hadron collider processes that enter a typical global fit, including inclusive jet and dijet production [166], inclusive electroweak gauge boson production at high p_T [246, 268] and high invariant mass $M_{ll(\nu)}$ [207, 206] and differential top quark pair production [286, 285]. Most of these calculations are implemented in publicly available programs. For instance, EW corrections to inclusive gauge boson production are available in programs such as `FEWZ` [207, 206] and `HORACE` [212]. The latest version of `MCFM` [444] also includes the calculation weak corrections to Drell-Yan, top-quark pair, and dijet production at hadron colliders. Recently, there has also been progress in the automation of the calculation of these corrections, both in the framework of `MadGraph5_aMC@NLO` [445] and of `Sherpa/OpenLoops` [446].

In Fig. 65 we show two representative examples of NLO EW corrections for processes relevant for PDF determinations, computed with `MCFM` at $\sqrt{s} = 13$ TeV [444]. In the left plot, we show the percentage NLO EW correction for high-mass dilepton production as a function of M_{ll}. The `ZGRAD` calculation is also shown. We see that these corrections are negligible for $M_{ll} \lesssim 500$ GeV, but that they can become significant as we increase M_{ll}, reaching $\delta_{\rm wk} \sim -20\%$ at 5 TeV. In the right plot, we show the same quantity, now for dijet production as a function of the invariant mass of the dijet M_{jj}. The two curves correspond to two possible ways to combine NLO QCD and EW corrections, known as additive ($\delta_{\rm add}$) and multiplicative ($\delta_{\rm prod}$). Here the corrections are more moderate (since the Born is a pure QCD process) but they can still become up to few percent in the region accessible at the LHC. The results of Fig. 65 illustrate how a careful inclusion of NLO EW corrections is important for the description of the LHC data in the TeV region used for PDF determinations.

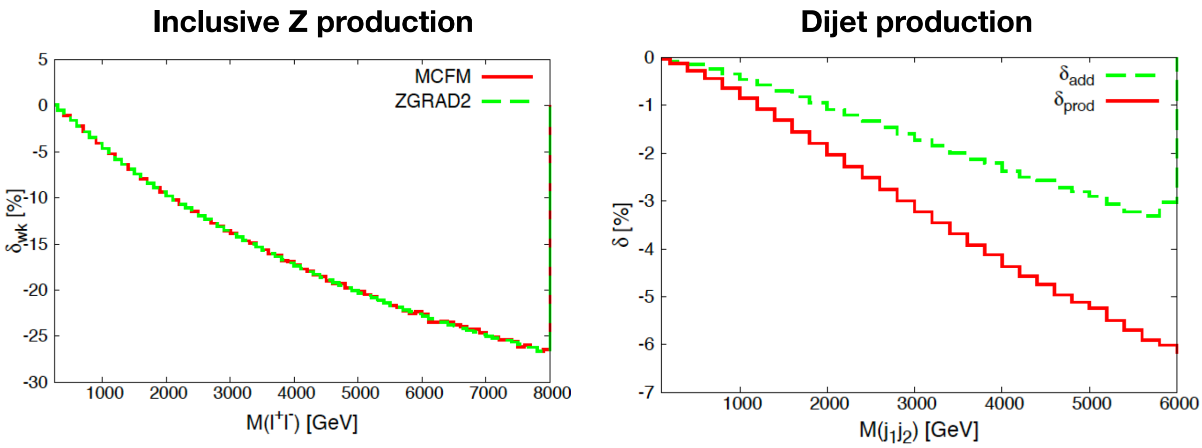


Figure 65: Two representative examples of NLO EW corrections for processes relevant for PDF determinations, computed with `MCFM` at $\sqrt{s}$ = 13 TeV [444]. In the left plot, we show the percentage NLO EW correction for high-mass dilepton production as a function of M_{ll}, comparing also with the corresponding `ZGRAD` calculation. In the right plot, we show the same quantity, now for dijet production as a function of the invariant mass of the dijet M_{jj}. The two curves correspond to two possible ways to combine NLO QCD and EW corrections, known as additive ($\delta_{\rm add}$) and multiplicative ($\delta_{\rm prod}$).

8. Implications for LHC phenomenology

In this section we present an overview of some of the most representative implications of PDFs and their uncertainties for LHC phenomenology. First of all we discuss the role of PDFs for the predictions of the Higgs boson production cross-sections at the LHC. Then we assess what is the role of PDF uncertainties for the searches of new heavy resonances predicted by various Beyond the Standard Model scenarios. And we complete this section we highlight the importance of PDFs for the precision measurements of SM parameters such as the W mass or the strong coupling constant.

8.1. Higgs production cross-sections

In the Standard Model, once the Higgs mass is measured, all other parameters of the Higgs sectors, such as the strength it is coupling to fermions and vector bosons, are uniquely determined. On the other hand, deviations of these Higgs couplings with respect to the SM predictions are expected in generic BSM scenarios. Therefore, the precision measurements of the couplings of the Higgs boson represents a unique opportunity for BSM searches, since any deviation with respect to the the tightly fixed SM predictions would represent a *smoking gun* for New Physics. Crucially, realizing this program requires not only high precision experimental measurements of Higgs boson production and decay in various channels, but also the calculation of the SM cross-sections and decay rates with matching accuracy. And in this respect, PDFs are one of the largest sources of theoretical uncertainty affecting Higgs boson production cross-sections [10].

Here we present a comparison of inclusive Higgs production cross-sections at 13 TeV with the latest releases of all PDF groups. The settings of this comparison, and the codes used for each process, are described in Ref. [104]. Specifically, we show the dominant Higgs boson production modes at hadron colliders: gluon fusion, associated production with a $t\bar{t}$ pair, VZ associate production, and Higgs production in vector-boson fusion. In addition, we also show the results for double Higgs production in the dominant gluon-fusion channel. Results are provided for NNPDF3.0 and 3.1, CT14, MMHT14 and for the ABMP16 NNLO sets for $\alpha_s(m_Z) = 0.118$, and for the latter case we also indicate the result corresponding to their best fit value of $\alpha_s(m_Z) = 0.1149$. The theoretical settings for each cross-section calculation are based on state-of-the-art matrix element calculations, for instance the gluon-fusion and VBF results are computed at N3LO using the `ggHiggs` [447] and `vbf@n3lo` [448] codes respectively. We only show here the PDF uncertainties, and other sources of theoretical errors affecting these cross-sections are listed in *e.g.* the latest Higgs Cross-Section Working Group report [10]. Interestingly, the uncertainty associated to the input value of $\alpha_s(m_Z)$ can be comparable to the PDF uncertainties in some channels.

There are a number of interesting features in the comparison of Fig. 66. First of all, it shows that in general there is good agreement between the three global fits, NNPDF3.1, CT14 and MMHT14 for all the Higgs boson production modes. The comparison between NNPDF3.1 and its predecessor NNPDF3.0 highlights good agreement for the gluon initiated channels, with a reduction of the PDF uncertainties in the former case, while for quark-initiated cross-sections such as VH and VBF there is an upper shift by around one sigma. Another remarkable feature of this comparison is that the recent ABMP16 set is also in reasonable with the rest of the groups, provided that the same common value of the strong coupling constant $\alpha_s(m_Z) = 0.118$ is used. On the other hand, if their best-fit value $\alpha_s(m_Z) = 0.1149$ is used in the calculation, there can significant differences in cross-sections, specially for the gluon initiated contributions. Specifically, ABMP16 is in this case around 7% (10%) lower than NNPDF3.1 for the gluon-fusion ($t\bar{t}$ associated production) cross-section.

It is also worth mentioning here that PDF uncertainties are relevant not only for the extraction of Higgs couplings from inclusive cross-sections, but also for the differential measurements that will become available thanks to the large statistics that will be accumulated by the end of Run II as well as for the HL-LHC.

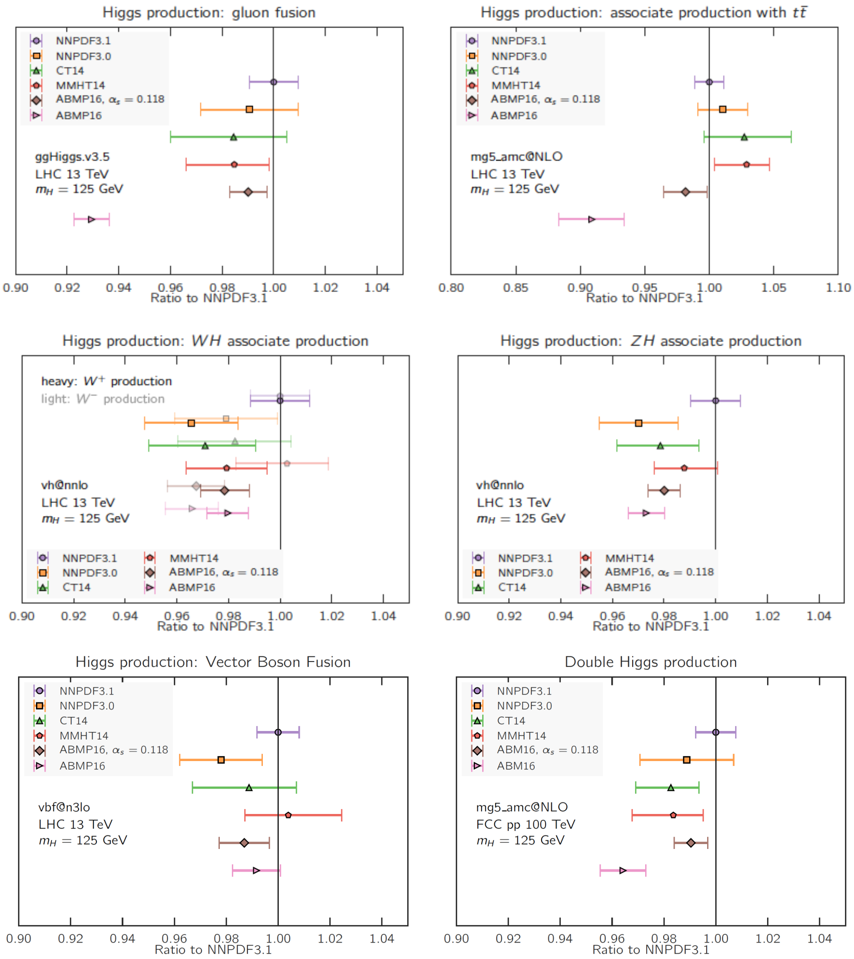


Figure 66: The PDF dependence of the most important Higgs production inclusive cross-sections at the LHC 13 TeV. The results are normalized to the central value of NNPDF3.1, and only PDF uncertainties are shown. See text for more details of the theoretical calculation.

To illustrate this point, in Fig. 67 we show the PDF uncertainties in the p_T^h distribution of Higgs bosons produced in the gluon-fusion mode at the LHC 13 TeV for $0 \leq p_T^h \leq 200$ GeV, computed using the PDF4LHC15 sets. In this case we find that PDF uncertainties are at around the $\sim 2\%$ level. However, these uncertainties will be increases as the LHC becomes sensitive to higher p_T values: as shown in the right plot of Fig. 67, at high invariant masses (high p_T values) the PDF uncertainties in the gluon-gluon luminosity become more significant.

8.2. PDF uncertainties and searches for new massive particles

Many scenarios of new physics beyond the Standard Model predict the existence of new heavy particles with masses around the TeV scale. Among many others, supersymmetry, composite Higgs, and extra dimensions, are some of the scenarios that motivate the search for new heavy resonances at the LHC in the high-mass tail of various kinematic distributions. In this respect, PDF uncertainties play an important role in setting robust exclusion limits based on available null results, and would become even more important in

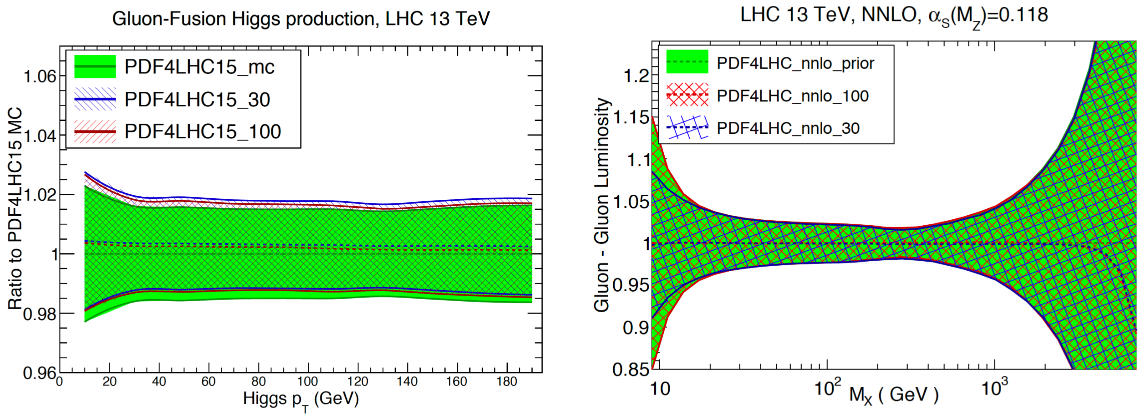


Figure 67: Left: the PDF uncertainties in the p_T^h distribution of Higgs bosons produced in the gluon-fusion mode at the LHC 13 TeV for $0 \leq p_T^h \leq 200$ GeV, computed using the PDF4LHC15 sets. Right: the gluon-gluon PDF luminosity with the same set now extending up to higher values of the invariant mass of the final state M_X.

the case of eventual discovery. The reason for this is that PDFs represent the dominant theoretical uncertainty for the production of new heavy particles in the TeV region, since these processes are sensitive to the large-x behaviour of quarks and gluons. And as discussed in Sect. **??**, PDF uncertainties are large in this region due to the limited amount of experimental constraints.

In order to quantify the size of the PDF uncertainties in the large invariant mass region relevant to these specific BSM searches, as well as the relative agreement between the PDF groups, it is useful to compare the PDF luminosities for $M_X \geq 1$ TeV. In this comparison we will restrict ourselves to ABMP16, CT14, MMHT14 and NNPDF3.1, in all cases using $\alpha_s(m_Z) = 0.118$. Results are shown in Fig. 68 for $\sqrt{s} = 13$, TeV normalized to the central value of the MMHT14 calculation. From the comparison in Fig. 68, we find that PDF uncertainties are small, at the few-percent level, up to $M_X \simeq 5$ TeV only for the quark-quark luminosities. This is explained by the fact that $\mathcal{L}_{qq}$ is dominated by the rather accurately known up and down quark valence PDFs, which are constrained *e.g.* by fixed-target DIS structure functions/ For all the other flavour combinations, PDF uncertainties are much larger, in particular for the quark-antiquark and gluon-gluon PDF luminosities.

In the case of the gluon-gluon luminosity, we find a rather large spread of the predictions between the different groups, with MMHT14 (ABMP16) leading to the largest (smallest) values of $\mathcal{L}_{gg}$. For instance, at $M_X \sim 5$ TeV, which is close to the upper limit of the kinematic coverage of the LHC, PDF uncertainties are almost $O(100\%)$. Even more more moderate invariant masses such as $M_X \sim 2.5$ TeV, the values of $\mathcal{L}_{gg}$ can vary between approximately $+10\%$ and -30% as compared to the central MMHT14 result. It is thus clear that these uncertainties would represent one of the limiting factors BSM characterization in the case of an eventual discovery. In the case of the quark-gluon luminosity, we observe a similar trend as in the gluon-gluon one but with reduced PDF uncertainties, due to the contribution of the well-constrained large-x quark PDFs.

PDF uncertainties, as well as differences between groups, are also large for the quark-antiquark PDF luminosity $\mathcal{L}_{q\bar{q}}$, also shown in Fig. 68. The reason for these behaviour is two-fold. On the one hand, large-x anti-quarks are notoriously difficult to pin down, although recent high-precision measurements from the Tevatron and the LHC are improving the situation. On the other hand, various groups parameterize the quark sea content of the proton with rather different assumptions [312], with the corresponding implications

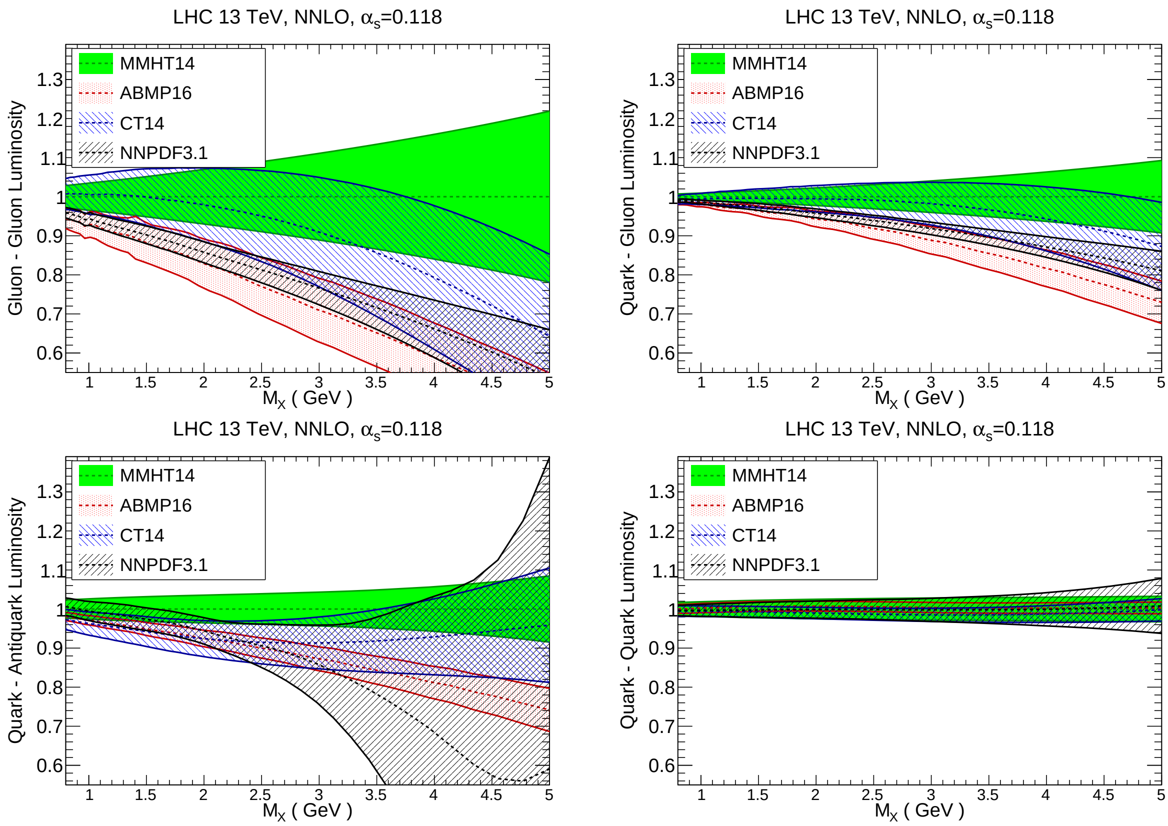


Figure 68: Comparison of PDF luminosities in the large invariant mass M_X region between MMHT14, ABMP16, CT14 and NNPDF3.1. From left to right and from top to bottom we show the results of the gluon-gluon, gluon-quark, quark-antiquark and quark-quark luminosities, normalized to the central value of MMHT14. In this comparison, NNLO PDFs with $\alpha_s(m_Z) = 0.118$ sets are used.

for the quark-antiquark luminosities. We find that the spread of the differences results ranges between +5% and −30% for $M_X = 3$ TeV, with PDF uncertainties becoming $O(100\%)$ for higher invariant masses. Note here that the PDF uncertainties are the largest for NNPDF3.1, despite this being the global analysis which includes a largest amount of LHC electroweak data sensitive to anti-quarks. This highlights the fact that methodological differences in the flavour assumptions and parametrization of anti-quarks are one of the dominant factors to explain the differences between the various groups in $\mathcal{L}_{q\bar{q}}$ at large values of M_X.

In order to illustrate the phenomenological consequences of these large PDF uncertainties at high M_X, in Fig. 69 we show the PDF uncertainties for high-mass graviton production in the Randall-Sundrum scenario [449, 450] induced by gluon-fusion at the LHC 8 TeV, computed with `MadGraph5` [451]. We compare the results of the NNPDF2.3 fit with those of the same fit including the constraints from top-quark production cross-sections. We observe that PDF uncertainties become $O(100\%)$ at large values of the graviton mass, consistent with the estimates from the gluon-gluon luminosity shown Fig. 68. We also see how these PDF uncertainties can be reduced by the inclusion of top quark pair production total cross-sections, highlighting the cross-talk between precision SM measurements and improving BSM searches.

In Fig. 69 we also show the K-factor for the NLO+NLL cross-sections including PDF uncertainties, normalized to the NLO result, for the production of a squark-anti-squark pair $\tilde{q}\tilde{q}^*$ at the LHC 13 TeV with

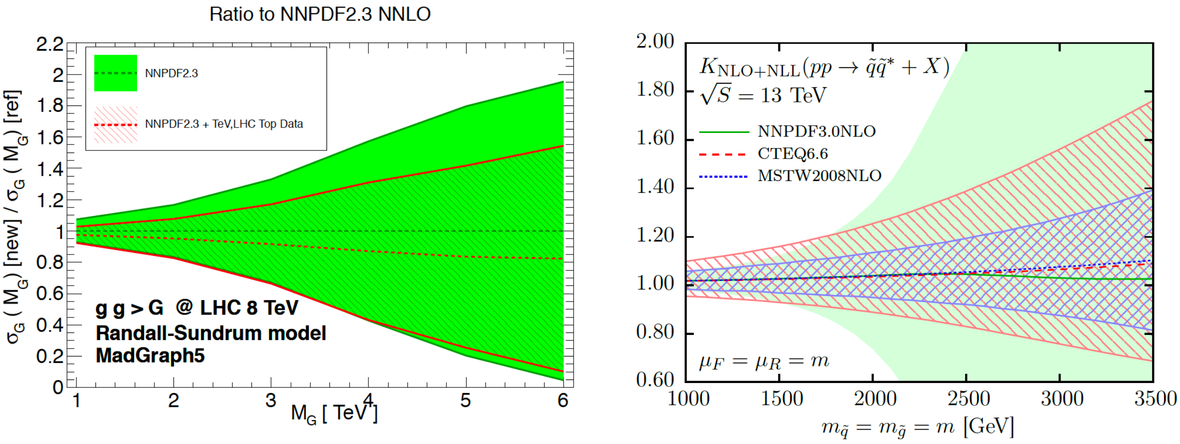


Figure 69: Left: the PDF uncertainties for high-mass graviton production in the Randall-Sundrum scenario induced by gluon-fusion at the LHC 8 TeV, computed with `MadGraph5`. We compare the results of the NNPDF2.3 fit with that of the same fit including the constraints from top-quark production cross-sections. Right: the K-factor for the NLO+NLL cross-section, including PDF uncertainties, for the production of a squark-anti-squark pair $\tilde{q}\tilde{q}^*$ at the LHC 13 TeV with various PDF sets.

various PDF sets, from Ref. [11]. This production channel is dominated by the quark-antiquark luminosity. Specifically, we compare the predictions of NNPDF3.0, CTEQ6.6, and MSTW2008, all at NLO. Note that by construction the central values of the three predictions are close since different trends cancel in this K-factor ratio, so the usefulness of this comparison is estimating the PDF uncertainties in each case. Here we also see that PDF uncertainties become very large at high-masses, in particular in the case of NNPDF3.0, reflecting the underlying behaviour of the quark-antiquark luminosities. Therefore, Fig. 69 highlights that, in the case of an eventual discovery of novel high-mass particles at the LHC, it will be crucial to improve our knowledge of large-x PDFs in order to be able to characterize the underlying BSM scenario,

In this respect, they way forward, as hinted already in Fig. 69, is the exploitation of high-precision collider data, mostly from the LHC, in order to pin down the large-x gluons and anti-quarks and thus reduce the PDF uncertainties associated to high-mass BSM particle production. For instance, Ref. [289] showed how by including the y_t and $y_{t\bar{t}}$ differential distributions from top quark pair production in a global PDF fit, it is possible to reduce the PDF uncertainties that affect the high-mass tail of the $m_{t\bar{t}}$ distribution by up to a factor two. This distribution is widely use for searches, for instance of new resonances that couple strongly to the top quark. More towards the future, it might conceivable to be able to provide indirect constraints on BSM, for instance on the coefficients of the SM-EFT [452] higher-order operators, by including these in the global PDF fit, along the lines of early studies aiming to constrain colored sparticles from Tevatron jet production [453].

8.3. *Precision measurements of SM parameters*

The precision measurements of SM parameters such as the mass of the W boson M_W or the running of the strong coupling constant $\alpha_S(Q)$ represent powerful ways of constraining BSM dynamics at the LHC. For instance, following the discovery of the Higgs boson, in the absence of new physics the standard model is an over-constrained theory: one can use a set of input parameters, such as the Higgs mass m_h and the top quark mass m_t, in the contact of the global electroweak precision fit [454], to predict other parameters, such as M_W. By comparing this indirect predictions of the W mass with direct experimental measurements,

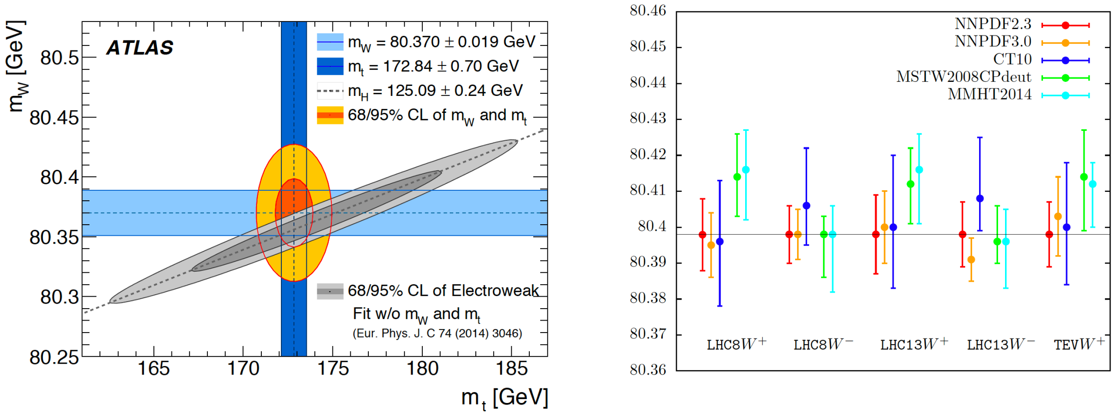


Figure 70: Left: comparison of the direct measurements of m_W, m_t, and m_h from ATLAS with the predictions from the global electroweak fit, from [458]. Right: estimate of the PDF uncertainties in the m_W determination using different PDF sets and collider scenarios, from [456] This estimate has been obtained from template fits to the p_T^l distribution, imposing the constraint that $p_T^W \leq 15$ GeV.

one can provide a stress-test of the SM, where any tension might indicate hints for BSM dynamics at scales higher than those that are currently accessible. The situation is fully analogous for the famous $g_\mu - 2$ anomaly, where a persistent 3 to 4-sigma discrepancy is found between the theoretical predictions of the muon anomalous magnetic moment [455] and the corresponding experimental measurement.

In order to make these comparison between indirect predictions and direct measurements as stringent as possible, it is important to improve the precision of the latter. And for many SM parameters, PDF uncertainties are one of the dominant uncertainties in their determination, providing another motivation for the need of improved PDFs. Focusing on the case of the W mass measurements, the role of PDF uncertainties has been quantified in detail in a number of studies, both from the phenomenological [456, 457, 12] and from the experimental point of view. In the latter case, the first direct measurement of M_W at the LHC has been recently presented by the ATLAS collaboration [458], yielding a total uncertainty of only 19 MeV, or which around half of it (9 MeV) is estimated to come from PDF uncertainties. In Fig. 70 we show a comparison of the direct measurements of m_W, m_t, and m_h from ATLAS with the predictions from the global electroweak fit, from [458]. We see that there is good agreement between the direct measurements and the indirect predictions, providing a highly-nontrivial validation test of the SM. Future measurements of m_W and m_t, as well as their combination with other experiments, should be able to reduce the uncertainties in this comparison.

In Fig. 70 we also show a phenomenological estimate of the PDF uncertainties associated to the m_W measurements using different PDF sets and collider scenarios, from [456] This estimate has been obtained from template fits to the p_T^l distribution, imposing the additional constraint that $p_T^W \leq 15$ GeV. A number of NNLO PDF sets are used in this comparison, in order to achieve a robust estimate of the PDF uncertainties. In general one finds that there is good agreement within PDF uncertainties, although in some cases this agreement is only marginal, as in the case of NNPDF3.0 and CT10 at the LHC 13 TeV for the W^- fits. From this study, one estimates that at the LHC 7 TeV PDF uncertainties using state-of-the-art sets are around 6 MeV, a similar number as the one in the ATLAS measurement.

Another SM parameter that could potentially provide indirect information on BSM dynamics is the QCD coupling $\alpha_s(Q)$, and specifically of its running at the TeV scale. It is well known that the presence of

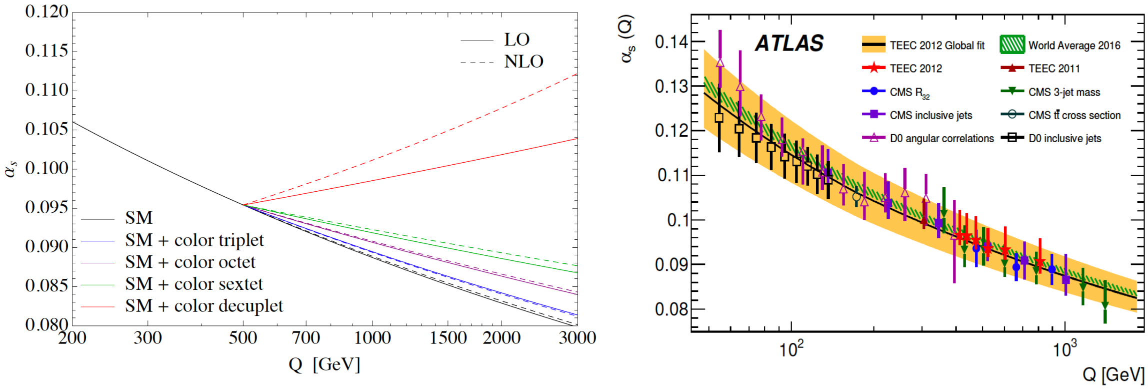


Figure 71: Left: the modification in the running of $\alpha_s(Q)$ induced by a new heavy colored fermion of mass $m = 0.5$ TeV as compared to the SM prediction, for various representations of its color gauge group, from [460]. Right: comparison of recent direct determinations of $\alpha_s(Q)$ at the Tevatron and the LHC as a function of Q, together with the PDG 2016 world average and with the results of the global ATLAS TEEC 2012 fit.

new colored heavy degrees of freedom will modify the QCD beta function and therefore lead to a different running with Q as compared to the corresponding SM prediction. This fact is for example at the basis of the improved agreement at high scales between the strong, weak and electromagnetic couplings in the case of low-scale supersymmetry, which suggest the unification of the three forces at a GUT scale of around $\Lambda_{\rm GUT} \sim 10^{16}$ GeV [459]. If these new heavy particles are at the TeV scale, the difference induced in the QCD coupling running could be accessible at the LHC, see *e.g.* [460] and Fig. 71, where we show the change in the running of $\alpha_s(Q)$ induced by a new heavy colored fermion of mass $m = 0.5$ TeV for various representations of its color gauge group, With this motivation, as well as the one to compare with other measurements of α_s at lower energies, the ATLAS and CMS collaborations have presented a number of measurements of both $\alpha_s(m_Z)$ and of $\alpha_s(Q)$ for individual Q bins, mostly from jet production [173, 461, 462] but also from top-quark pair production [463] (see also [464] for a review, and Fig. 71 for a graphical overview).

In these collider-based determinations of the strong coupling, PDF uncertainties, which are significant at the TeV scale (see Sect. 8.2), represent an important source of theoretical uncertainties. For instance, in the recent ATLAS determination of $\alpha_s(m_Z)$ from transverse energy-energy correlations (TEEC) at 8 TeV [461], the PDF uncertainty is $\delta_{\rm pdf} = 0.0018$, almost a factor 2 larger than the experimental uncertainty of $\delta_{\rm pdf} = 0.0011$. While in this analysis PDF uncertainties are sub-dominant with respect to the scale uncertainties, $\delta_{\rm scale} \simeq 0.006$, the latter were computed using NLO theory and can be reduced significantly by exploiting the NNLO calculation. Likewise, in the CMS analysis at 7 TeV, based on a QCD analysis of the inclusive jet cross-sections [173], one finds that the PDF uncertainties $\delta_{\rm pdf} = 0.0028$ are larger than the experimental uncertainties $\delta_{\rm exp} = 0.0019$, though still sub-dominant with respect to the large scale variations of the NLO calculation $\delta_{\rm scale} =^{+0053}_{-0.0024}$. In both cases, it is manifest that if one is able to reduce PDF uncertainties, and exploit the reduction of scale errors of the NNLO calculation, one can achieve a very competitive determination of $\alpha_s(m_Z)$ and to also perform stringent tests of its running in the TeV region.

As a related topic, we would like to mention that there are also proposals to measure the running of the electroweak running couplings at the LHC [465] and use these to impose model-independent constraints on new particles with electroweak quantum numbers without any assumptions about their decay properties. Also for this measurements PDFs are one of the dominant theoretical uncertainties, for instance in the high-mass tail of W and Z production at the LHC due to the quark-antiquark luminosity. Improving our knowledge of large-x anti-quarks will thus be helpful in these respect provide indirect constraints of new heavy electroweak sectors.

9. The future of PDF determinations

In the final section of this Report we discuss three topics that could play an important role in shaping global analyses of PDFs in the coming years. First of all we discuss the problem of theoretical uncertainties in fits of parton distributions, whose estimate is becoming more and more urgent given the size of PDF uncertainties in recent global analysis. Secondly, we summarize recent progress in lattice QCD computations of PDFs, including the first efforts towards a determination of their x-dependence, and suggest that in the near future lattice inputs could contribute to global PDF fits. Thirdly, we briefly review the status and plans for future high-energy colliders, such as the Large Hadron electron Collider (LHeC) or the Future Circular Collider (FCC), and the role that PDFs would play in these.

9.1. PDFs with theoretical uncertainties

The development of sophisticated methodologies for PDF fits, as well as the availability of a wealth of high-precision data, have reduced the PDF uncertainties in global analysis, arising mostly from experimental data and from procedural choices, to the few percent-level in the most constrained regions. At this level of accuracy, various theoretical uncertainties become more and more important, representing a major limitation for present and (even more) for future studies. Therefore, robustly accounting for these theoretical uncertainties is of the main goals of PDF fitters for the near future.

In this section, we focus specifically on the role of the theoretical uncertainties due to missing higher orders (MHOU) in the QCD coupling constant, namely those arising from the truncation of the asymptotic perturbative expansion. In this respect, there have been a number studies recently on how to estimate MHOU, although we are still far from a conclusive answer. In the following we first review progresses on MHOU of calculations of non-hadronic and hadronic processes and then several recent studies related to PDF determination.

We emphasize that these theoretical uncertainties from MHO should not be confused with the parametric theoretical uncertainties, that is, those arising from the choice of the values of input parameters such as α_s and m_c. These have been reviewed in Sect. 4.4, and it is well establish how to estimate its impact on the PDF fit and to propagate these parametric uncertainties into collider cross-sections.

9.1.1. MHOU on matrix element calculation

The most frequently used and probably also the simplest method of estimating the missing higher order corrections is the variation of the renormalization and factorization scales in a given fixed-order calculation. In the case of the total inclusive cross sections or decay rates with a single hard scale Q, usually one varies the QCD renormalization scale μ_R within the interval $[Q/r, rQ]$. The induced changes on the physical observable, either from three-point evaluations or from a scan over the entire range, are taken as the *uncertainty* of the MHO, assuming then typically either a Gaussian (or two half-Gaussians) or a flat distribution. The conventional choice is $r = 2$, which is found to work well in most cases, but that underestimates the true higher order corrections in certain cases, especially if the fixed-order calculation is carried out at leading order.

At hadron colliders, there exist in addition to the renormalization scale μ_R also the factorization scale μ_F, arising from the factorization of collider QCD divergences due to initial state hadrons. The two scales μ_R

and μ_F can be varied either simultaneously or independently within above range, with the later case usually further restricted to $1/r \leq \mu_F/\mu_R \leq r$. However, even for a single scale problem, there can still be different choices of the central or nominal scale, e.g., $Q/2$ or $2Q$, motivated by either QCD resummation or speed of convergence of the series [283], which leads to further ambiguities in the estimation of theoretical uncertainties from scale variations. There are also recent studies on utilizing the principal of maximum conformality on the choice QCD renormalization scale at different order which claims much smaller MHOU [466] than traditional scale variations.

Determining a suitable prescription for scale variations becomes more complicated when moving to differential observables, since there more hard scales including those related to the kinematics are involved. That usually requires a dynamic choice for the central scale, which often also depends on the specific distribution considered. For example, in a recent study on hadronic production of top-quark pairs [283] it shows that the preferable scale is half of the transverse mass of the top quark when studying the transverse momentum distribution of the top quark, and one fourth of the sum of transverse mass of top quark and anti-quark when studying rapidity distribution of the top quark and so on. Starting from a given choice of the nominal scale, then scale variations can be evaluated in a similar way as for the inclusive case and serve as estimations of MHOU. One further complexity arises concerning the correlations of the MHOU or scale variations in different regions of the distribution. Typically, they are assumed to be fully (anti-)correlated in the entire region which leads to very small theoretical uncertainties in case of a normalized distribution. In this respect, there have been attempts on decorrelating these scale variations based on consideration of kinematic dependence of the QCD corrections [467].

There exist other alternative proposals on estimating MHOU based on results at known orders, like the so-called Cacciari-Houdeau (CH) approach [468]. The basic idea is to express the full perturbation series in terms of the expansion parameter $\alpha_s(Q)$ and assume that all the expansion coefficients follow a same uniform bounded probability density distribution in the Bayesian sense. Bayesian inference can be used to calculate the probability density of the unknown higher-order coefficients given those known coefficients of lower orders. Thus MHOU including its probability density distribution (non-Gaussian in general) can be constructed. The original CH method was developed for the study of non-hadronic processes. Subsequently, in the modified CH ($\overline{\text{CH}}$) [469] approach, it was generalized to hadronic processes as well. There the expansion parameter has been adjusted to $\alpha_s(Q)/\lambda$, with the parameter λ determined from a global survey of selected processes with known higher-order corrections. To be specific, the best value of λ computed from the predicted probability density of higher orders is required to match the distribution from frequency count in the survey. For hadronic processes, the optimal value of λ is found to be about 0.6, meaning that the *true* perturbative expansion parameter is actually around $1.7\alpha_s(Q)$ rather than $\alpha_s(Q)$.

As another example, the series acceleration method [470] can also be applied to approximate the full result for physical observables based on the available information from a finite number terms of the asymptotic series, *e.g.*, using Levin-Weniger sequence transforms. In Ref. [470] it is assumed the theory prediction has a uniform distribution (in the Bayesian sense) between the last known partial sum of the perturbation series and its approximated value from Weniger δ-transform.

To illustrate how these various methods compare to each other, Fig. 72 shows the predictions for the production cross sections of the Higgs boson via gluon fusion at the LHC 8 TeV, calculated at LO, NLO, NNLO and approximate N^{3}LO with a nominal scale of $\mu_R = \mu_F = m_H$ [471]. It compares the MHOU as estimated from different approaches, including scale variations, CH, $\overline{\text{CH}}$ and the series acceleration method of [470] at various perturbative orders. Note that different approaches may have different interpretations on the uncertainties. In the case of the $\overline{\text{CH}}$ method, the λ value has been adjusted to give almost equal expansion coefficients for the known orders [471]. The $\overline{\text{CH}}$ method predicts larger uncertainty in general.

The different approaches turn our to give similar sizes for the MHOU associated to the N^{3}LO calculation, with the exception of the original CH. In addition, note that the series acceleration method also induces a shift on the central value of the prediction.

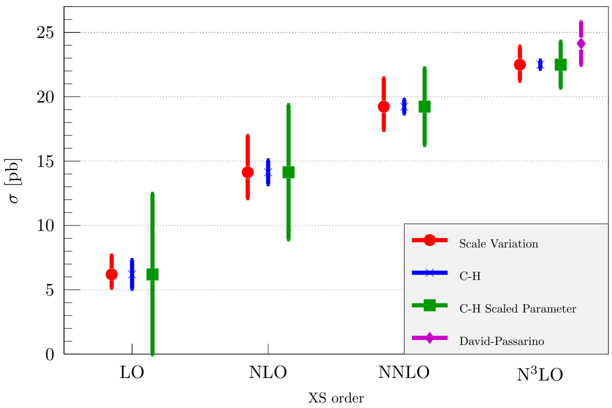


Figure 72: The cross section for Higgs production in gluon fusion calculated at increasing perturbative orders [471]. At each order the theoretical uncertainty is shown for using scale variation (red circles), the CH method (blue crosses), and the $\overline{\text{CH}}$ method (green squares); at N^{3}LO the Passarino-David uncertainty based on series acceleration method is also shown (purple diamonds).

9.1.2. *MHOU on parton distributions*

Global determinations of PDFs are based on perturbative calculations of matrix elements and DGLAP splitting kernels, suitably combined to predict a variety of physical cross-sections. In these perturbative calculations, in principle one should account for their associated MHOU, which then propagates into the resulting PDFs via the fitting of the theoretical predictions to the experimental data. Therefore, sophisticated treatments on the MHOU from different sources are required in order to study the impact on the PDFs, not unlike as the treatment of the experimental systematic uncertainties. Crucially, the correlations between theoretical predictions of different experimental bins of one process and further of different processes must be accounted for. Furthermore, when making any theoretical prediction, one should also take care of the correlations between the MHOU of the PDFs and of those coming from the MHOU of the process studied, since they may rely on the same perturbative expansion for the relevant matrix elements.

Due to the significant complexity of this problem, there is still no satisfactory solution, and the MHOU have not been included in any of the public PDFs from global determinations. However, it is possible to restrict ourselves to a region where we know only a single process is most likely dominant on the MHOU, *e.g.*, the inclusive jet cross sections as for gluon PDF at large-x, there can still exist simple prescriptions like using scale variations. To illustrate this, the left plot of Fig. 73 shows the impact of the choice of the QCD scales in calculations of the inclusive jet cross sections on the gluon PDFs at $Q = 2$ GeV for alternative CT10 NNLO fits [81] with two different χ^2 definitions (see Sect. 4.2.1). Note that the theoretical predictions on jet cross sections there are only at NLO though the PDFs are determined at NNLO.

From Fig. 73 we see that in the region with $x > 0.2$, the spread of the gluon PDFs by using scales of 0.5, 1 and 2 times the central scale (in this case the jet p_T) illustrates how here MHOU can be quite significant, most likely of the same order than the nominal PDF uncertainties. In a related study, also based on alternative CT10 NNLO fits, scale variations of the NLO inclusive jet cross sections are further decomposed into several correlated systematics described by five nuisance parameters [3]. By treating those

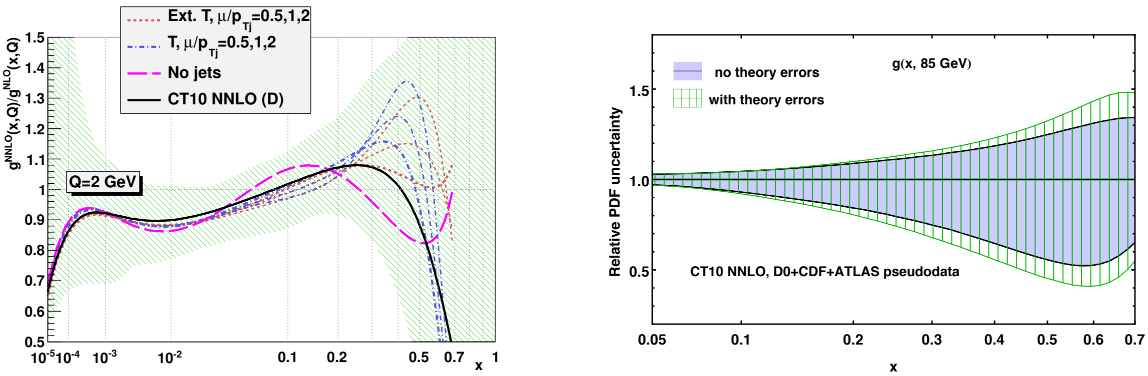


Figure 73: Left plot: dependence of the gluon PDF on the choice of QCD scales used in the calculation of inclusive jet cross sections in CT10 NNLO fits [81], normalized to the central value of the NLO gluon. Right plot: impact of the theoretical uncertainties from the inclusive jet cross sections included in the CT10 NNLO fits [3] in the resulting gluon PDFs.

systematics in a similar way as the experimental correlated systematic errors, it is possible to include the MHOU in the standard PDF uncertainty on the same footing as the experimental systematic uncertainties. As shown in the right plot of Fig. 73, the inclusion of theory errors from the jet cross-sections in the CT10 NNLO fit results in an increase of the gluon PDF uncertainty at large-x, consistent with the left panel of Fig. 73.

Another possibility to provide a rough estimate of the MHOU consists in checking the convergence of the fitted PDFs with increasing orders. Fig. 74 the shows the comparison of the nominal PDF uncertainties with the difference of the central PDFs fitted at NLO and NNLO for gluon and total singlet PDFs at $Q = 100$ GeV in the NNPDF3.0 fits [17]. This difference between PDF central values at NLO and NNLO provides a conservative upper bound of the MHOU associated to the NNLO PDFs. From this comparison, we can observe that there are regions where the shifts of NLO to NNLO PDFs are comparable or even larger than the conventional PDF uncertainties. In this respect, one could also apply the CH or $\overline{\text{CH}}$ approach based on the fitted PDFs at LO, NLO and NNLO. For example Ref. [471] found the resulting MHOU of the NNLO PDFs have negligible impact on the Higgs production cross section through gluon fusion, but on the other hand it could be relevant for the top-quark pair production.

9.2. Lattice QCD calculations of the proton structure

As discussed in Sect. 2.2, parton distributions arise from non-perturbative QCD dynamics. Therefore, since currently we are not able to solve analytically strongly-coupled non-Abelian gauge theories, it is very challenging to compute PDFs from first principles. Perhaps the only possibility in this respect, beyond model calculations, consists in exploiting recent progress in lattice QCD [472]. This method is based on discretizing the QCD Lagrangian in a finite-volume Euclidean lattice, which introduces naturally an ultraviolet cutoff, and then computing directly non-perturbative QCD quantities on this lattice and take the continuum limit. Perhaps the main advantage of lattice QCD calculations is that they require minimal external input, in particular only the hadronic mass scale $\Lambda_{\rm QCD}$ and the values of the quark masses, or alternatively, the physical pion and kaon masses.

Given that parton distributions have a formal definition in terms of the nucleon matrix elements of certain non-local operators (see the discussion in Sect. 2.2), in principle it should be possible to attempt to

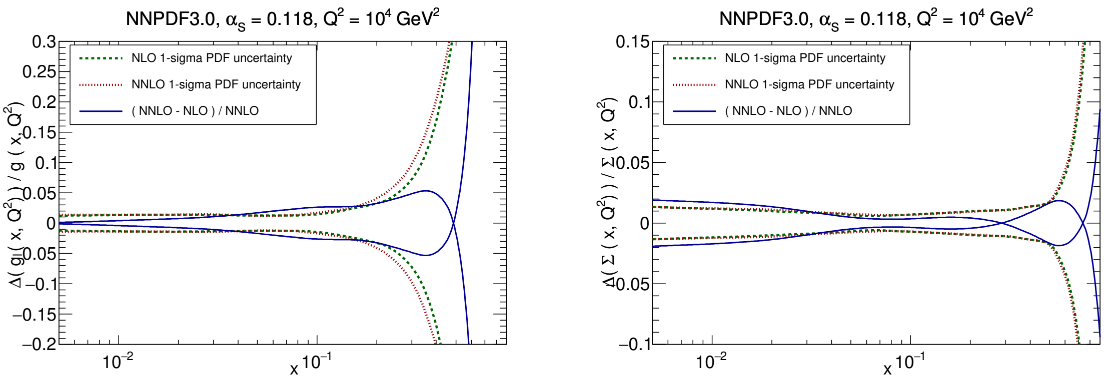


Figure 74: Left plot: comparisons between the nominal PDF uncertainties with the difference of the central PDFs determined at NLO and NNLO for the gluon PDF at $Q = 100$ GeV in the NNPDF3.0 fit. Right plot: same for the total quark singlet, $\Sigma(x, Q^2)$.

compute PDFs within lattice QCD. From the practical point of view, however, given the extremely CPU intensive nature of these calculations, most lattice QCD results on PDFs have been limited for a long time to the first two moments of non-singlet flavour combinations for large (unphysical) quark masses. These restrictions have been overcome in the recent years, with several groups providing now results of PDF moments for physical pion and kaon masses. Moreover, it is now also possible to go beyond flavor non-singlet operators, and compute in addition gluonic and quark singlet matrix elements. Even more recently, both conceptual and numerical breakthroughs in lattice QCD computations have allowed to move further, allowing the calculation of higher PDF moments both for flavour non-singlet and singlet operators, as well as the first attempts to evaluate parton distributions and related quantities directly in Bjorken-x space.

Here we briefly review some of these recent developments in lattice QCD calculations of PDFs. For a more detailed overview of this progress, together with the study of their interplay with state-of-the-art global analysis, see Ref. [473], a white paper that was produced as the outcome of the dedicated workshop "Parton Distributions and Lattice QCD calculation in the LHC era",[9] which took place in Oxford in March 2017, and that brought together experts from the two fields to explore the synergies and complementary aspects between the two approaches. The discussion and results shown in this section represent an brief excerpt of the material contained in this white paper.

Recent progress in lattice QCD calculations of PDFs and related quantities has been partly driven by a greatly improved control on the systematic uncertainties that enter the calculation of relatively simple quantities such as nucleon matrix elements, which correspond to low moments of the PDFs. These systematic uncertainties, among others, include using physical pion masses, reducing the excited-state contamination, and using large lattices to remove finite-size effects. Moreover, to make contact with the physical world and experimental data, the numerical results are extrapolated to the continuum and infinite-volume limits. In addition, the past decade has seen significant progress in the development of efficient algorithms for the generation of ensembles of gauge field configurations, which represent the QCD vacuum, and tools for extracting the relevant information from lattice QCD correlation functions.

Another important component of this progress has been the development of novel strategies to overcome the limitations in computing the first few moments [474, 475, 476] and make possible to determine

[9] http://www.physics.ox.ac.uk/confs/PDFlattice2017

more challenging quantities, such as gluon and flavor-singlet matrix elements, as well as directly calculate the Bjorken-x dependence of PDFs [477, 478, 479, 480], These developments have pushed lattice QCD calculations to the point where, for the first time, it is possible to provide information on the PDF shape of specific flavour combinations, both for quarks and for antiquarks, and meaningful comparison with global fits can start to be made.

As mentioned above, in order to be reliably used in phenomenological applications, lattice QCD calculations must demonstrate control over all relevant sources of systematic uncertainty introduced by the discretisation of QCD on the lattice. These include discretisation effects that vanish in the continuum limit; extrapolation from unphysically heavy pion masses; finite volume effects; and renormalisation of composite operators. Moreover, taking the continuum limit requires accurate determinations of the lattice spacing. These various sources of systematic uncertainty all need to be under control when confronting experimental data with lattice results, or vice versa. For a coherent assessment of the present state of lattice QCD calculations of various quantities, the degree to which each systematic has been controlled in a given calculation is an important consideration. The quality of individual lattice calculations can be quantitative assessed based on criteria such as those from the FLAG analysis of flavour physics on the lattice [481].

The traditional approach for lattice QCD calculations of parton distributions has been to determine the matrix elements of local twist-two operators, where twist is the dimension minus the spin, that can be related to the Mellin moments of PDFs. In principle, given a sufficient number of Mellin moments, PDFs can be reconstructed from the inverse Mellin transform. In practice, however, the calculation is limited to the lowest three moments, because power-divergent mixing occurs between twist-two operators on the lattice. Three moments are insufficient to reconstruct the momentum dependence of the PDFs without significant model dependence [482]. The lowest three moments do provide, however, useful information, both as benchmarks of lattice QCD calculations and as constraints in global extractions of PDFs. For instance, provided systematic uncertainties are kept under control, one can envisaging adding these lattice QCD calculations of PDF moments as an additional theoretical constrain to the global fit, on the same footing as the momentum and valence sum rules.

Here we briefly summarize the state-of-the-art of lattice QCD calculations of the first moment of unpolarized PDFs, which are those for which systematic uncertainties are under better control. See Ref. [473] for a more exhaustive set of comparisons, including those of lattice QCD calculations with global fits. The observables that are discussed here are defined as follows:

1. The first moment of the flavour triplet combination, $T_3 = u^+ - d^+$,

$$\langle x \rangle_{u^+-d^+}(\mu^2)\Big|_{\mu^2=Q^2} = \int_0^1 dx\, x \left\{ u(x,Q^2) + \bar{u}(x,Q^2) - d(x,Q^2) - \bar{d}(x,Q^2) \right\} . \qquad (141)$$

2. The first moment of the individual quark $q^+ = q + \bar{q}$ PDFs,

$$\langle x \rangle_{q^+=u^+,d^+,s^+,c^+}(\mu^2)\Big|_{\mu^2=Q^2} = \int_0^1 dx\, x \left\{ q(x,Q^2) + \bar{q}(x,Q^2) \right\} . \qquad (142)$$

3. The first moment of the gluon PDF,

$$\langle x \rangle_g(\mu^2)\Big|_{\mu^2=Q^2} = \int_0^1 dx\, x\, g(x,Q^2) . \qquad (143)$$

In Table 2 we show a selection of recent results for the moments defined in Eqns. (141-143). As can be seen, for $\langle x \rangle_{u^+-d^+}$ the lattice QCD uncertainties vary between 5% and 15%, with the quoted results not

Mom.	Collab.	Ref.	N_f	Status	discretisation	quark mass	finite volume	renormalisation	excited states		Value
$\langle x \rangle_{u^+-d^+}$	LHPC 14	[483]	2+1	P	■	★	★	★	★		0.140(21)
	ETMC 17	[484]	2	PreP	■	★	■	★	★	*	0.194(9)(11)
	RQCD 14	[485]	2	P	■	■	○	★	★	**	0.217(9)
$\langle x \rangle_{u^+}$	ETMC 17	[484]	2	PreP	■	★	■	★	★	*▸	0.453(57)(48)
$\langle x \rangle_{d^+}$	ETMC 17	[484]	2	PreP	■	★	■	★	★	*▸	0.259(57)(47)
$\langle x \rangle_{s^+}$	ETMC 17	[484]	2	PreP	■	★	■	★	★	*▸	0.092(41)(0)
$\langle x \rangle_g$	ETMC 17	[484]	2	PreP	■	★	■	○	★	*	0.267(22)(27)

* Study employing a single physical pion mass ensemble.
** Study employing a single ensemble with $m_\pi = 150$ MeV.
▸ The mixing with $\langle x \rangle_g$ is computed.

Table 2: Summary of recent lattice QCD calculations of the first moments of unpolarized PDFs, defined in Eqns. (141-143), evaluated at $\mu^2 = Q^2 = 4$ GeV^2. See Ref. [473] for more details about the computation of each entry in the table, as well as the description of the various sources of systematic uncertainties that affect them.

agreeing among themselves within errors. For the first moment of the gluon, $\langle x \rangle_g$, the uncertainties are around 10%, and for the individual total quark combinations they vary between 10% and 20%. So while current determinations of the first moments are unlikely to provide constraints on global PDF fits (where uncertainties are the few-percent level), future calculations with improved systematic errors might be able to make a difference. On the other hand, existing calculations can be already used to provide meaningful constraints on polarized PDFs, where uncertainties are rather larger than in the unpolarized case due to the scarcer dataset.

While the lowest moments of the PDFs provide crucial benchmarks to assess the reliability of lattice QCD calculations of the nucleon structure, as well as potentially useful information for global PDF fits, they do not allow reconstructing the complete x-dependence of the PDFs. In particular, the calculation of PDF moments is mostly insensitive to the small-x region. To bypass these limitations, recently a number of approaches have been developed, aiming to determine the x-dependence of PDFs directly from lattice QCD. One of the most important approaches goes under the same of *quasi-PDFs*, first formulated by Ji in Refs. [486, 487]. For simplicity, we focus in the following on the flavor nonsinglet case so that we can neglect the mixing with the gluons. The unpolarized quark quasi-PDF $\widetilde{q}(x, \Lambda, p_z)$ is defined as a momentum-dependent nonlocal static matrix element:

$$\widetilde{q}(x, \Lambda, p_z) \equiv \int \frac{dz}{4\pi} e^{-ixzp_z} \frac{1}{2} \sum_{s=1}^{2} \langle p, s | \bar{\psi}(z) \gamma_\alpha e^{ig \int_0^z A_z(z')dz'} \psi(0) | p, s \rangle \, , \tag{144}$$

where Λ is an UV cut-off scale, typically chosen to be the inverse lattice spacing $1/a$. Note that, because p is finite, the momentum fraction x can be larger than unity. As we see from Eq. (144), these quasi PDFs are defined for nucleon states at finite momentum. Therefore, in order to make contact with the standard collinear PDFs and thus with phenomenology, they must be related to the corresponding light-front PDF, for which the nucleon momentum is taken to infinity. In the large-momentum effective field theory (LaMET) approach, the quasi PDF $\widetilde{q}(x, \Lambda, p_z)$ can be related to the p_z-independent light-front PDF $q(x, Q^2)$ through

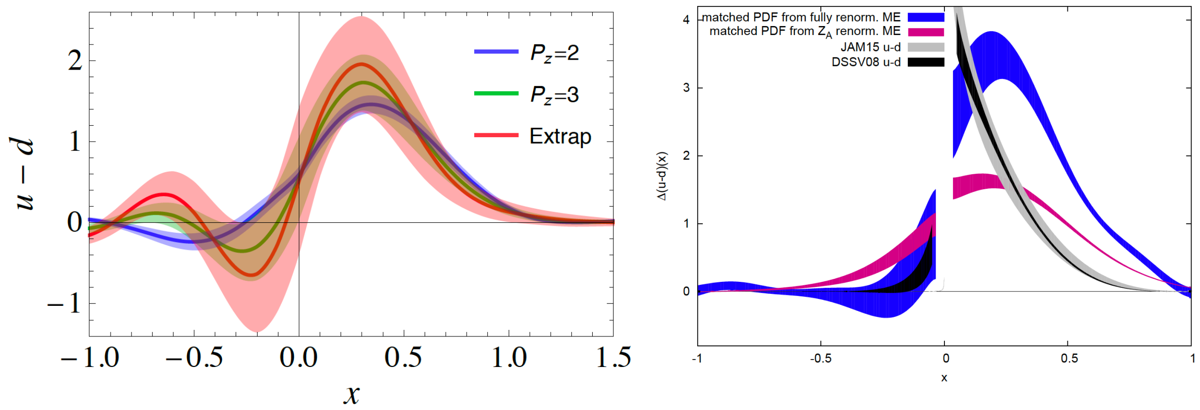


Figure 75: Left plot: the renormalized unpolarized isovector quark distribution, $u-d$, after one-loop matching and mass correction at the renormalization scale $\mu = 2.4$ GeV. The red band shows the extrapolation to infinite momentum. The negative-x part of this figure is related to the antiquark distribution by means of the following relation: $\bar{u}(x) - \bar{d}(x) = -u(-x) + d(-x)$ for $x > 0$. Right plot: comparison of polarized quark triplet, $\Delta u - \Delta d$, obtained from quasi PDF computed with either fully renormalized matrix elements (blue) or with bare matrix elements multiplied by the local axial current Z-factor, Z_A (magenta). For illustration purposes, we also show the results of two recent polarized PDF fits, namely DSSV08 and JAM15.

the following relation [486, 487]

$$\widetilde{q}(x, \Lambda, p_z) = \int_{-1}^{1} \frac{dy}{|y|} Z\left(\frac{x}{y}, \frac{\mu}{p_z}, \frac{\Lambda}{p_z}\right)_{\mu^2=Q^2} q(y, Q^2) + O\left(\frac{\Lambda^2_{\rm QCD}}{p_z^2}, \frac{m^2}{p_z^2}\right), \tag{145}$$

where μ is the renormalisation scale; Z is a matching kernel; and m is the nucleon mass. Here the $O\left(m^2/p_z^2\right)$ terms are target-mass corrections and the $O\left(\Lambda^2_{\rm QCD}/p_z^2\right)$ terms are higher twist effects, both of which are suppressed at large nucleon momentum. A complementary approach to the LaMET methods views instead the quasi PDF as a "lattice cross-section" from which the light-front PDF can be factorized [488, 489].

Preliminary results from lattice calculations of quasi PDFs have been rather encouraging [477, 478, 479, 480], although a number of important limitations still need to be overcome. To illustrate this progress in lattice calculations of x-space PDFs, in Fig. 75 we show the renormalized unpolarized isovector quark distribution, $u - d$, after one-loop matching and mass correction at the renormalization scale $\mu = 2.4$ GeV. The red band shows the extrapolation to infinite momentum, together with the associated uncertainties. The negative-x part of this figure is related to the antiquark distribution by means of the following relation:

$$\bar{u}(x) - \bar{d}(x) = -u(-x) + d(-x) \quad \text{for} \;\; x > 0. \tag{146}$$

Although these calculations are still in its infancy, they represent a promising approach to be able to complement global PDF fits with non-trivial information about the x dependence.

The same methodology can be of course also applied to other nucleon matrix elements, including the polarized PDFs. To show this, in Fig. 75 we also display the comparison of matched helicity PDFs for the same quark flavor combination, namely $\Delta u - \Delta q$. These results have been obtained from quasi PDF computed with either fully renormalized matrix elements or with bare matrix elements multiplied by the local (z=0) axial current Z-factor, Z_A. For illustration purposes, we also show the results of two recent polarized PDF fits, namely DSSV08 [29] and JAM15 [490]). The small-x region, where the current lattice QCD calculations are not reliable, is not shown. From this comparison we see that the lattice QCD results are still far from the global fits (and thus also from the experimental data), but as in the unpolarized case there is still ample room for further progress here.

Despite these remarkable developments, there still remain a number of important challenges that must be overcome before one can achieve a complete determination of the x-dependence of PDFs directly from lattice QCD that is competitive with the global PDF fits. In particular, excellent control over the various sources of systematic uncertainties that affect the calculation must be reached. Some of these are common to the calculations of PDF moments, as discussed above, but there are also a number of additional systematic errors specific to quasi-PDFs, such as those associated with the finite nucleon momentum of the lattice calculations and with the renormalisation of quasi PDFs. Once this is achieved, and given the recent fast progress, it is thus conceivable that in the future lattice QCD calculations of x-space PDFs can be used to constrain the global analysis.

9.3. Parton distributions at future colliders

Now we turn to discuss the role that PDFs would have in some of the recent proposals for future colliders involving hadrons in the initial state. There are three main families of possible future colliders now under active discussion. To begin with, electron-positron colliders, such as the ILC [491], CLIC [492] or TLEP/FCC-ee [493], offer the potential for ultra-high precision measurements of the Higgs, electroweak and top-quark sectors. On the other hand, hadron colliders with energy much greater than the LHC would allow to continue the exploration of the high-energy frontier and significantly extend the coverage of searches for new BSM particles, including Dark Matter candidates, while make possible at the same time unprecedented opportunities for the study of the Higgs sector such as the Higgs self-interactions. In this respect, there is ongoing work towards a circular collider hosted at the CERN site which would accelerate protons up to the extreme energies of $\sqrt{s} = 100$ TeV [494, 15], dubbed FCC-hh, and there is also a similar machine under study by the Chinese HEP community.

Another avenue for future high-energy collisions would be new machines based on electron-proton collisions, exploiting the successful strategy adopted with HERA. One of open proposals is the Large Hadron electron Collider (LHeC) [16], where the $E_p = 7$ TeV proton beam from the LHC would collide with an electron/positron beam with $E_e = 60$ GeV coming from a new LinAc, and that would be able to reach the region down to $x_{\rm min} \simeq 2 \cdot 10^{-6}$ at $Q^2 = 2$ GeV^2. A more extreme incarnation of the same idea corresponds to colliding these $E_e = 60$ GeV electrons with the $E_p = 50$ TeV beam of the FCC-hh. The resulting collider, dubbed FCC-eh, would be able to reach down to $x_{\rm min} \simeq 2 \cdot 10^{-7}$ at $Q^2 = 2$ GeV^2. These two machines would thus probe PDFs in the small-x region much deeper than HERA. In the same category falls the Electron Ion Collider (EIC) [495] which might start construction soon either in the BNL or the JLAB sites. The EIC would offer the possibility to polarize both leptons and protons and to accelerate as well heavy nuclei, although its $\sqrt{s}$ would be smaller than that of HERA.

In this section, we review the role that PDFs would play first at the LHeC/FCC-eh, and then at a future hadron collider with a center-of-mass energy of $\sqrt{s} = 100$ TeV.

9.3.1. PDFs at high-energy lepton-hadron colliders

As mentioned above, one of the possibilities for a future high-energy collider now under active discussion would be to exploit the LHC/FCC proton beam and collide it with a high energy lepton beam, which would be delivered by a new LinAc to be built at the CERN site. In the case of using the LHC beams, this project, known as the Large Hadron electron Collider (LHeC), would then represent a scaled-up version of HERA, and as such would offer immense opportunities for improved determinations of the proton structure

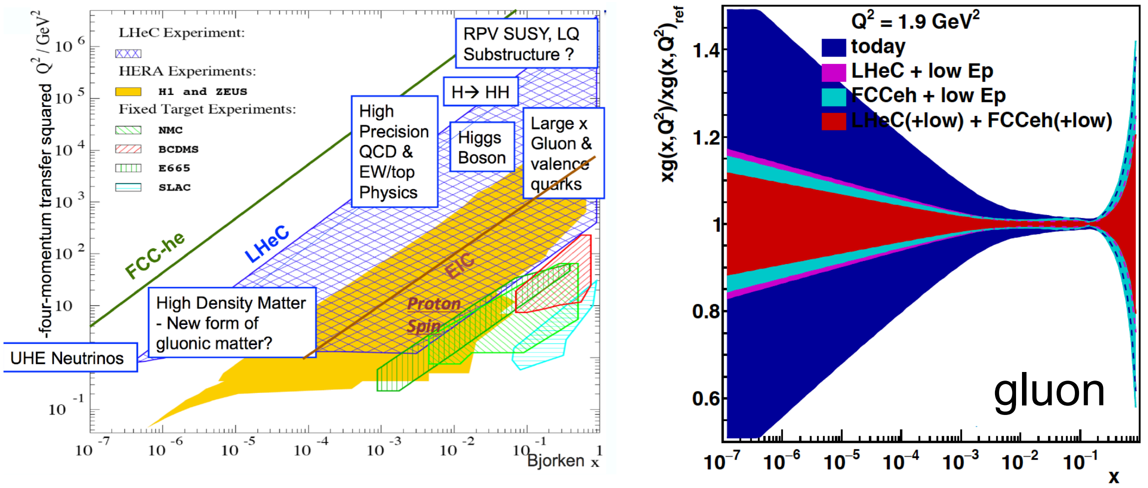


Figure 76: Left plot: kinematic coverage in the (x, Q^2) of several existing and proposed deep-inelastic scattering experiments. Starting from the fixed-target experiments and then moving to HERA, the LHeC and finally the FCC-eh, as the center of mass energy increases, the kinematic reach extends both towards higher Q^2 and smaller x values. Right plot: results of an `xFitter` PDF feasibility study that compares the impact on the gluon PDF of adding either LHeC or FCC-eh (or both) pseudo-data in addition to the HERA inclusive structure function dataset.

down to very low x and high-Q^2, as well as providing a wealth of information on nuclear PDFs in a kinematic region where they are currently essentially unconstrained. Several options are now being considered, with some preference now for synchronous operation during the final years of the HL-LHC upgrade, since then the LHeC program can be extended to include measurements of the Higgs sector.

In Fig. 76 we show the kinematic coverage in the (x, Q^2) of several existing and future deep-inelastic scattering experiments, including the EIC, the LHeC, and the FCC-eh. We observe how by starting from the fixed-target experiments and then moving to HERA, the LHeC and finally the FCC-eh, as the center of mass energy increases, the kinematic reach extends both towards higher Q^2 and smaller x values. At the FCC-eh in particular, it should be able to cover the region down to $x \simeq 10^{-7}$ without leaving the perturbative region $Q \gtrsim 1$ GeV. It is important to emphasize that the same coverage would be achieved for nuclear PDFs, extending by four or five orders of magnitude in x as compared to existing measurements.

One of the most important aspects of the LHeC/FCC-eh scientific case is being able to probe the proton/nuclear PDFs with an unprecedented precision, not only by means of inclusive structure functions but also with measurements of the strange, charm, and bottom structure functions, that provide a direct handle on the heavy flavour PDFs. As a related topic, the LHeC/FCC-eh would allow to measure the strong coupling constant $\alpha_s(M_Z)$ with per-mile uncertainties, for instance using jet production [496], and high-precision measurements of the electroweak sector parameters. To illustrate these potentialities, in Fig. 76 we also show the results of an `xFitter` PDF feasibility study that compares the impact on the gluon PDF of adding either LHeC or FCC-eh (or both) pseudo-data in addition to the HERA inclusive structure function dataset. The reduction of the PDF uncertainties down to very small-x values reflects the extended kinematic reach of these future high energy lepton proton colliders. A similar reduction of the PDF uncertainty is expected for the quark PDFs.

Another important aspect of the interplay between PDFs and the LHeC/FCC-eh is related the small-x resummation framework [497, 105, 498, 499]. This framework is based on extending the collinear DGLAP formalism to account for the all-order resummation of terms of the form $\alpha_s^k \ln^m(1/x)$, as implemented in the

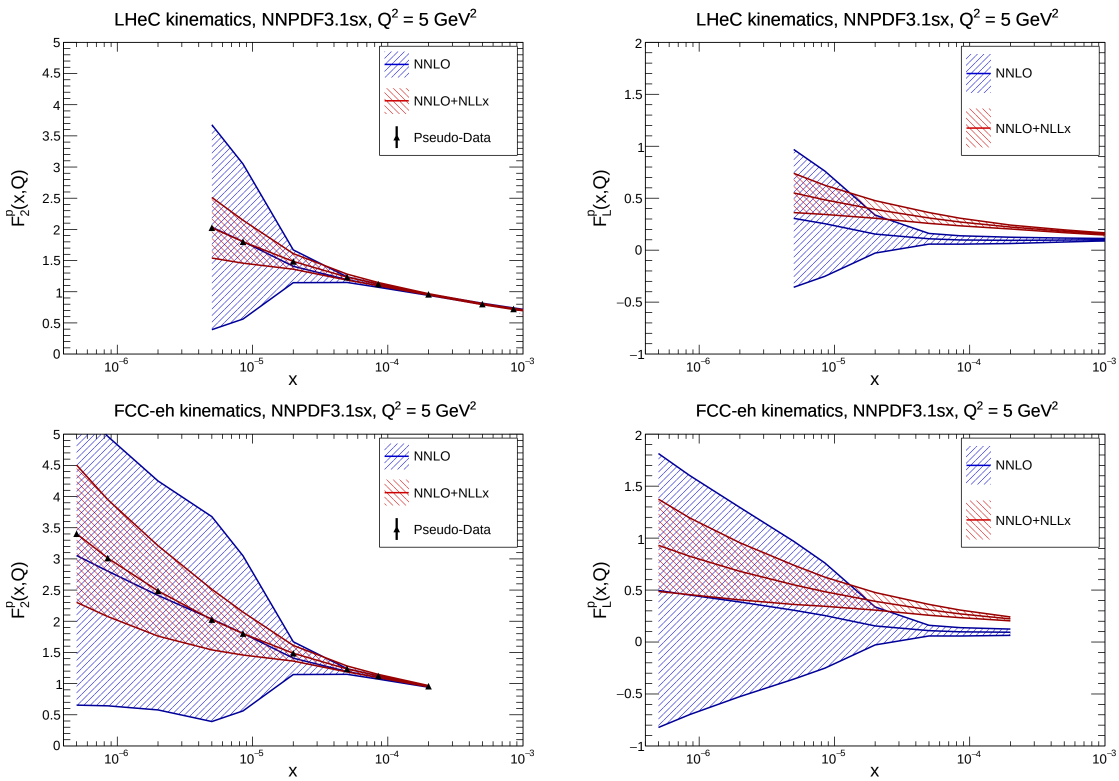


Figure 77: The theoretical predictions for the F_2 and F_L structure functions at the LHeC (upper) and FCC-eh (lower plots) using the NNPDF3.1sx NNLO and NNLO+NLLx fits at $Q^2 = 5$ GeV^2. In the case of the F_2^p structure function, we also show the expected total experimental uncertainties based on the simulated pseudo-data, assuming the NNLO+NLLx values as central prediction.

BFKL equation. Thanks to small-x resummation, the reliability of theoretical predictions for DIS structure functions and collider cross-sections can be extended down to much smaller values of x, as compared to the calculations based on the collinear DGLAP framework. Recently, a version of the NNPDF3.1 global analysis, called NNPDF3.1sx, based on NLO+NLLx and NNLO+NLLx theory has been presented [500, 501], providing unambiguous evidence for the onset of BFKL dynamics in the inclusive HERA structure function data. Therefore, given that the effects of small-x resummation are already important for the description of the HERA data, one expects them to become even more relevant for higher-energy lepton-proton colliders (see Fig. 76).

With the motivation of providing a first estimate of the relevance of small-x resummation for the LHeC/FCC-eh, in Fig. 77 we provide predictions for the F_2 and F_L structure functions using the NNPDF3.1sx NNLO and NNLO+NLLx fits at $Q^2 = 5$ GeV^2 for the kinematics of the LHeC and the FCC-eh. For these calculations, we have used `APFEL` to produced NNLO(+NLLx) predictions, each using as input the corresponding NNPDF3.1sx fits, for the most updated version of the simulated LHeC/FCC-eh pseudo-data kinematics. In the case of F_2, we also show the expected total experimental uncertainties based on the simulated pseudo-data, assuming the NNLO+NLLx curve as central prediction. The total uncertainties of the simulated pseudo-data are the few percent level, and therefore they are rather smaller than the PDF uncertainties in the complete kinematic range.

From the comparisons in Fig. 77, we see how the FCC-eh would allow probing the small-x region by about an order of magnitude deeper than the LHeC (which in turn extends HERA by about the same amount). The differences between NNLO and NNLO+NLLx are quite small for F_2, specially taking into account the large PDF uncertainties, implying that refitting the pseudo-data is required to first reduce PDF errors and then discriminate between the two theoretical scenarios. Given the small experimental errors, these inclusive F_2^p measurements would represent a sensitive probe of small-x dynamics. From this comparison we also see that differences are more marked for F_L, with central values differing by several sigma (in units of the PDF uncertainty) in most of the accessible kinematic range. This illustrates the sensitivity of F_L measurements to probe small-x QCD. We also note that small-x predictions based on non-linear effects ("saturation") have typically the opposite trend of small-x resummation (suppressing the structure functions as compared to the NNLO fixed-order calculation). Therefore, the measurements in Fig. 77 and related ones would open a unique window to the novel dynamical regime of QCD in at very small x.

9.3.2. PDFs at a 100 TeV hadron collider

Next we move to discuss parton distributions at the FCC-hh [502], a proposal for a future hadron collider with a center of mass energy of $\sqrt{s} = 100$ TeV. First of all, in order to illustrate the extended kinematic coverage that would be achieved at a 100 TeV proton-proton collider as compared to the one at the LHC, in Fig. 78 we compare the (x, M_X) coverage at 100 TeV and 14 TeV, where the dotted lines indicate the regions of constant rapidity at the FCC-hh. In addition, also indicate the relevant M_X regions for some representative processes, from low masses (Drell-Yan, low p_T jets), electroweak scale processes (Higgs, W, Z, top), new high-mass particles (squarks, Z'). It is clear that there is a significant increase in the kinematic coverage. A particularly interesting aspect is that at the FCC-hh even high-scale processes such as W, Z or h production become sensitive to the small-x region.

As discussed in the FCC Yellow Report [15], there are two main aspects of PDF phenomenology related to a 100 TeV collider. On the one hand, just as at the LHC, at 100 TeV parton distributions are one of the dominant systematic theoretical uncertainties for several cross-sections. In particular, electroweak scale cross-sections, such as W or h production, become sensitive to the small-x region where PDF uncertainties are currently large. To illustrate this point, in Fig. 79 we show the comparison of cross-sections for different representative processes at the FCC with $\sqrt{s} = 100$ TeV, between the NNPDF3.0 predictions and those of the NNPDF3.0+LHCb sets, see Ref. [503] for more details. The acceptance cuts are different in each process. In the left plot we show the results for direct photon production, off-peak Drell-Yan cross-sections, and inclusive weak boson production. In the right plot we show the fiducial cross-sections for $c\bar{c}$ and $b\bar{b}$ production. In all cases, and specially for heavy quark pair production, we can observe the reduction of PDF uncertainties that is derived once the NNPDF3.0+LHCb sets are used.

The other side of PDFs at the FCC-hh is the onset of new phenomenon that are absent at the lower energies of the LHC. These include the possibility of treating the top quark as a massless parton [504, 505], the need for resummation of "collinear" weak gauge boson radiation and the consequent introduction of electroweak PDFs [506, 507, 508], as well as the increased role for photon-induced processes [421]. Moreover, just as in the case of the LHeC/FCC-hh, the role of small-x resummation is expected to become more important at the FCC-hh than at the LHC, given the sensitivity of even standard candles such as W, Z and Higgs production to the small-x region.

Let us here provide two representative illustrations of these new PDF-related phenomena at the FCC-hh. As mentioned above, at 100 TeV the electroweak gauge bosons becomes effectively massless, and thus it is possible to construct electroweak PDFs with the corresponding evolution equations. In Fig. 80 we show the PDF of the W^+ boson normalized to that of the gluon, as a function of x for different scales: $q = 10^4$ GeV, 10^6 GeV, 10^8 GeV, computed using the framework of [506]. We observe that the dependence of the

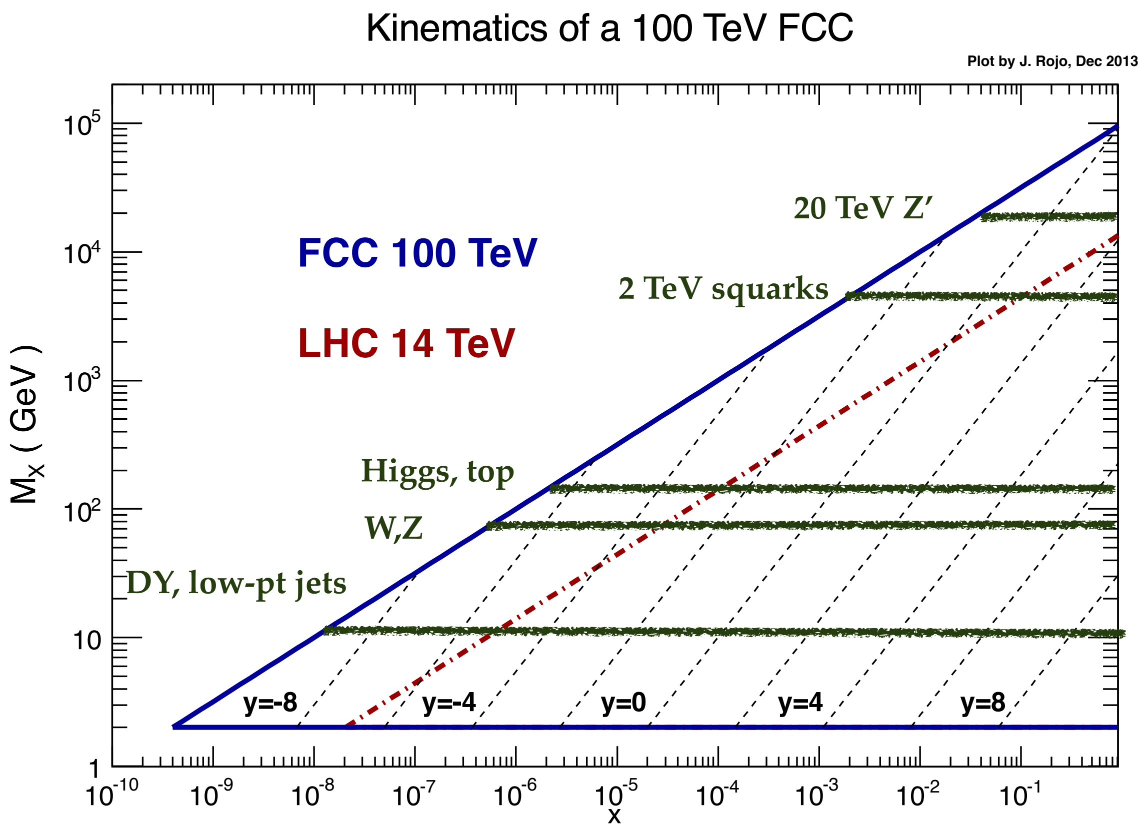


Figure 78: Kinematic coverage in the (x, M_X) plane of a $\sqrt{s}$ = 100 TeV hadron collider (solid blue line), compared with the corresponding coverage of the LHC at $\sqrt{s}$ = 14 TeV (dot-dashed red line).

W PDF with the energy q is rather mild. For most of the range of x, the W PDF is at most a few percent of the gluon PDF, while for $x \geq 0.1$ is becomes larger, up to 40% of the gluon PDF. This not not necessarily mean that the effects of the W PDF will be phenomenologically relevant: this can be assessed only at the cross-section level, comparing calculations with massive gauge bosons and those where these are treated as massless (and thus resummed into the electroweak PDFs).

Following a similar line of thought, at 100 TeV it is conceivable to treat the top quark as massless partons, much in the same way as at the LHC the bottom quark is treated as massless in most calculations. In Fig. 80 we show the cross-section inclusive Higgs production by $t\bar{t}$ associated production, comparing the results of the $n_f = 5$ scheme ($gg \to ht\bar{t}$), the $n_f = 6$ scheme ($t\bar{t} \to h$), and of their interpolation by means of the ACOT general-mass scheme. The comparison is performed as a function of the Higgs boson mass m_{H^0}. We find that the $n_f = 6$ calculation, where the top quark is treated as massless and resummed into a top PDF, is rather far from the matched calculation up to at least $m_{H^0} = 10$ TeV. This suggest that the massless top approximation is not suitable even for the extreme FCC energies. On the other hand, the concept of top PDF is still useful in order to improve fixed order calculations, using general-mass schemes such as ACOT or FONLL, but it should never be used in isolation.

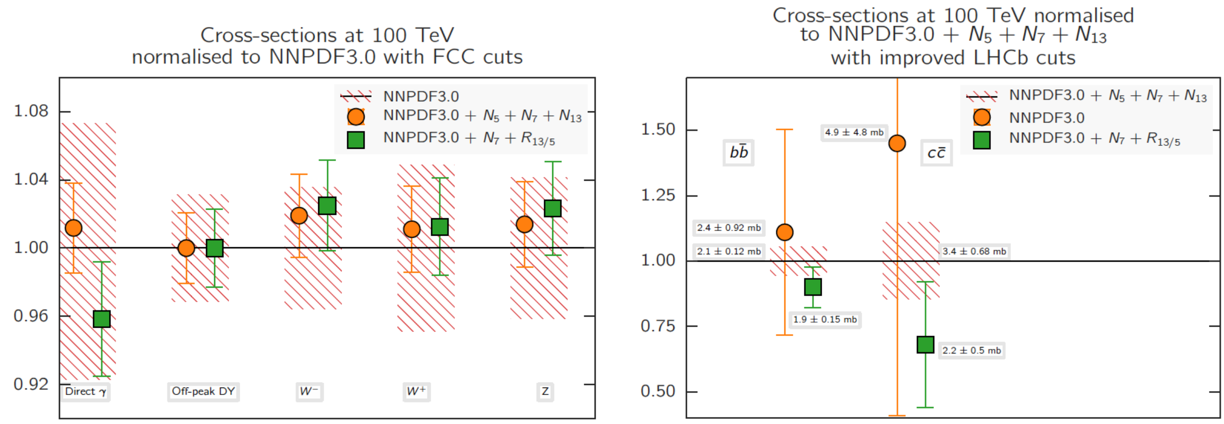


Figure 79: Comparison of cross-sections for different representative processes at the FCC with $\sqrt{s}$ = 100 TeV, between the NNPDF3.0 predictions and those of the NNPDF3.0+LHCb sets, as discussed in the text. The acceptance cuts are different in each process. In the left plot we show the results for direct photon production, off-peak Drell-Yan cross-sections, and inclusive weak boson production. In the right plot we show the fiducial cross-sections for $c\bar{c}$ and $b\bar{b}$ production.

10. Conclusions

The wealth of data that has been accumulated by the LHC so far, together with the additional data that will be collected in the coming two decades, has allowed the study and stress-test of the SM in an unprecedented way. Together with recent progress in theoretical calculations, this means that we are now entering the precision era of the LHC, aiming to compare data and theory at the few percent level or even less. Given the null results of BSM searches so far, a systematic high-precision analysis of the SM predictions and the LHC data might be one of most promising approaches to look for BSM dynamics at the LHC, for instance in the case that they manifest as subtle differences with the SM theory. And in this respect, the detailed mapping of the quark and gluon structure of the proton represents and important component of this LHC precision physics program.

In this Report we have presented an overview of the most important recent developments in PDF determinations, with emphasis for their implications for LHC phenomenology. After a succinct review of the theoretical foundations of the global QCD analysis framework, we have reviewed recent progress both from the theoretical and the experimental point of view for those hard-scattering cross-sections used in PDF fits; we have compared the similarities and differences between the methodologies uses for the various PDF fitting collaborations; and then presented the state-of-the-art fits from each group and assessed what we can learn about the internal structure of the nucleons from various points of view. We have then discussed the role of QED corrections in PDF fits, in particular concerning photon-initiated processes, and presented some of the most representative examples of the applications of PDFs for LHC phenomenology, from the measurement of the Higgs couplings to the determination of the mass of the W boson.

In the last section of these Report we have attempted to speculatively discuss some topics that very likely will play a crucial role in the near-term future of PDF determinations. One of these topics is that one of theoretical uncertainties, for instance arising from missing higher-order terms in the perturbative expansion. Given the size of PDF uncertainties in the latest sets, it is conceivable that these theory error are comparable (if not larger) than the nominal PDF uncertainties, and thus finding a statistically sound method to account for these is of utmost importance. Another topic that might affect the PDF fitting paradigm is that

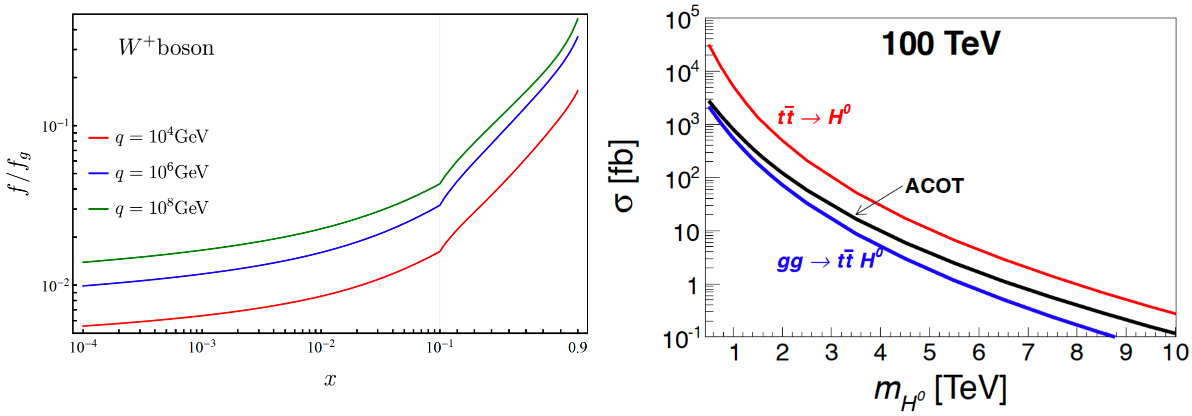


Figure 80: Left: the PDF of the W^+ boson normalized to that of the gluon, as a function of x for different scales: $q = 10^4$ GeV, 10^6 GeV, 10^8 GeV, from Ref. [506]. Right: the cross-section inclusive Higgs production by $t\bar{t}$ associated production, comparing the results of the $n_f = 5$ scheme ($gg \to ht\bar{t}$), the $n_f = 6$ scheme ($t\bar{t} \to h$), and of their interpolation by means of the ACOT general-mass scheme.

of the interplay with lattice QCD calculations, where recent progress both in computing Mellin moments of various flavor combinations as well in direct x-space calculation of PDFs suggest that in the future the constraints from lattice calculations should allow to improve the global PDF fits, in a similar way as *e.g.* the momentum sum rule. Finally we have summarized the importance of PDFs for future higher-energy colliders whose physics case is being discussed just now, such as a Large Hadron electron Collider or a new proton-proton collider with a center-of-mass energy of up to 100 TeV.

We hope that this Report has managed to convey to the reader that the topic of PDF determinations is fascinating and lively one, with implications from the understanding of the non-perturbative dynamics of the strong interactions to searches for new BSM physics and ultra-high energy astrophysics. In some respect, PDF fits represent a unique stress-test of the SM and of the collinear QCD factorization framework, which is clearly in very good shape, given that we are now able to simultaneously describe a few tens of individual experiments, some of them with extremely small uncertainties at the per-mile level. PDF fits thrive at the the cross-roads of advanced data analysis, state-of-the-art perturbative calculations, and modern robust statistical methodology, and thus provide guidance for other similar global analyses efforts such as fits of the SMEFT coefficients. Moreover, progress in unpolarized PDF fits is also one of the main drivers of recent improvements of other related aspects of the proton structure, from polarized PDFs to nuclear PDFs.

As we enter in the LHC precision era, ever-improving PDF determinations will keep providing a unique contribution to this exciting exploration of the high-energy frontier. Taken into account that only around 15 years the first PDF sets with uncertainties where introduced, is clear that the requirements of this LHC precision program will further drive improvements in global PDF determinations, leading always to an ever more detailed picture of the inner life of the protons.

Acknowledgments

We are grateful to Giulia Zanderighi for giving us the opportunity to write this Report. We are grateful to many colleagues for extensive discussions and fruitful collaborations of the topic of PDF determinations. A necessarily incomplete list of people that we could like to thank includes Sergei Alekhin, Mandy Cooper-Sarkar, Stefano Forte, Jan Krenschmar, Sasha Glazov, Joey Huston, Katerina Lipka Michelangelo Mangano, Pavel Nadolsky, Voica Radescu, Robert Thorne, Gavin Salam, ... `add all names`

J. R. is supported by the European Research Council Starting Grant "PDF4BSM". The work of J. G. is sponsored by Shanghai Pujiang Program. L. H. L thanks the Science and Technology Facilities Council (STFC) for support via grant award ST/L000377/1 `add all funding ID`.

[1] A. Accardi, et al., A Critical Appraisal and Evaluation of Modern PDFs, Eur. Phys. J. C76 (8) (2016) 471. arXiv:1603.08906, doi:10.1140/epjc/s10052-016-4285-4.

[2] J. Butterworth, et al., PDF4LHC recommendations for LHC Run II, J. Phys. G43 (2016) 023001. arXiv:1510.03865, doi:10.1088/0954-3899/43/2/023001.

[3] J. Rojo, et al., The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II, J. Phys. G42 (2015) 103103. arXiv:1507.00556, doi:10.1088/0954-3899/42/10/103103.

[4] R. D. Ball, S. Carrazza, L. Del Debbio, S. Forte, J. Gao, et al., Parton Distribution Benchmarking with LHC Data, JHEP 1304 (2013) 125. arXiv:1211.5142, doi:10.1007/JHEP04(2013)125.

[5] S. Alekhin, et al., The PDF4LHC Working Group Interim ReportarXiv:1101.0536.

[6] S. Forte, G. Watt, Progress in the Determination of the Partonic Structure of the Proton, Ann.Rev.Nucl.Part.Sci. 63 (2013) 291. arXiv:1301.6754, doi:10.1146/annurev-nucl-102212-170607.

[7] S. Forte, Parton distributions at the dawn of the LHC, Acta Phys.Polon. B41 (2010) 2859. arXiv:1011.5247.

[8] E. Perez, E. Rizvi, The Quark and Gluon Structure of the Proton, Rep.Prog.Phys. 76 (2013) 046201. arXiv:1208.1178, doi:10.1088/0034-4885/76/4/046201.

[9] A. De Roeck, R. S. Thorne, Structure Functions, Prog.Part.Nucl.Phys. 66 (2011) 727. arXiv:1103.0555, doi:10.1016/j.ppnp.2011.06.001.

[10] D. de Florian, et al., Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs SectorarXiv:1610.07922.

[11] W. Beenakker, C. Borschensky, M. Krmer, A. Kulesza, E. Laenen, S. Marzani, J. Rojo, NLO+NLL squark and gluino production cross-sections with threshold-improved parton distributions, Eur. Phys. J. C76 (2) (2016) 53. arXiv:1510.00375, doi:10.1140/epjc/s10052-016-3892-4.

[12] G. Bozzi, J. Rojo, A. Vicini, The Impact of PDF uncertainties on the measurement of the W boson mass at the Tevatron and the LHC, Phys.Rev. D83 (2011) 113008. arXiv:1104.2056, doi:10.1103/PhysRevD.83.113008.

[13] A. Cooper-Sarkar, P. Mertsch, S. Sarkar, The high energy neutrino cross-section in the Standard Model and its uncertainty, JHEP 08 (2011) 042. arXiv:1106.3723, doi:10.1007/JHEP08(2011)042.

[14] R. Gauld, J. Rojo, L. Rottoli, S. Sarkar, J. Talbert, The prompt atmospheric neutrino flux in the light of LHCb, JHEP 02 (2016) 130. arXiv:1511.06346, doi:10.1007/JHEP02(2016)130.

[15] M. L. Mangano, et al., Physics at a 100 TeV pp collider: Standard Model processesarXiv:1607.01831.

[16] J. Abelleira Fernandez, et al., A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector, J.Phys. G39 (2012) 075001. arXiv:1206.2913, doi:10.1088/0954-3899/39/7/075001.

[17] R. D. Ball, et al., Parton distributions for the LHC Run II, JHEP 04 (2015) 040. arXiv:1410.8849, doi:10.1007/JHEP04(2015)040.

[18] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, C. P. Yuan, New parton distribution functions from a global analysis of quantum chromodynamics, Phys. Rev. D93 (3) (2016) 033006. arXiv:1506.07443, doi:10.1103/PhysRevD.93.033006.

[19] L. A. Harland-Lang, A. D. Martin, P. Motylinski, R. S. Thorne, Parton distributions in the LHC era: MMHT 2014 PDFs, Eur. Phys. J. C75 (5) (2015) 204. arXiv:1412.3989, doi:10.1140/epjc/s10052-015-3397-6.

[20] S. Alekhin, J. Blumlein, S. Moch, R. Placakyte, Parton Distribution Functions, α_s and Heavy-Quark Masses for LHC Run IIarXiv:1701.05838.

[21] H. Abramowicz, et al., Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data, Eur. Phys. J. C75 (12) (2015) 580. arXiv:1506.06042, doi:10.1140/epjc/s10052-015-3710-4.

[22] P. Jimenez-Delgado, E. Reya, Delineating parton distributions and the strong coupling, Phys.Rev. D89 (7) (2014) 074049. arXiv:1403.1852, doi:10.1103/PhysRevD.89.074049.

[23] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, N. Sato, Constraints on large-x parton distributions from new

weak boson production and deep-inelastic scattering data, Phys. Rev. D93 (11) (2016) 114017. arXiv:1602.03154, doi:10.1103/PhysRevD.93.114017.

[24] A. Manohar, P. Nason, G. P. Salam, G. Zanderighi, How bright is the proton? A precise determination of the photon parton distribution function, Phys. Rev. Lett. 117 (24) (2016) 242002. arXiv:1607.04266, doi:10.1103/PhysRevLett.117.242002.

[25] R. D. Ball, V. Bertone, M. Bonvini, S. Carrazza, S. Forte, A. Guffanti, N. P. Hartland, J. Rojo, L. Rottoli, A Determination of the Charm Content of the Proton, Eur. Phys. J. C76 (11) (2016) 647. arXiv:1605.06515, doi:10.1140/epjc/s10052-016-4469-y.

[26] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, C. Schmidt, J. Winter, K. Xie, C. P. Yuan, CT14 Intrinsic Charm Parton Distribution Functions from CTEQ-TEA Global AnalysisarXiv:1707.00657.

[27] S. Alekhin, et al., HERAFitter, Eur. Phys. J. C75 (7) (2015) 304. arXiv:1410.4412, doi:10.1140/epjc/s10052-015-3480-z.

[28] E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi, J. Rojo, A first unbiased global determination of polarized PDFs and their uncertainties, Nucl.Phys. B887 (2014) 276. arXiv:1406.5539, doi:10.1016/j.nuclphysb.2014.08.008.

[29] D. de Florian, R. Sassot, M. Stratmann, W. Vogelsang, Extraction of Spin-Dependent Parton Densities and Their Uncertainties, Phys. Rev. D80 (2009) 034030. arXiv:0904.3821, doi:10.1103/PhysRevD.80.034030.

[30] K. J. Eskola, P. Paakkinen, H. Paukkunen, C. A. Salgado, EPPS16: Nuclear parton distributions with LHC data, Eur. Phys. J. C77 (3) (2017) 163. arXiv:1612.05741, doi:10.1140/epjc/s10052-017-4725-9.

[31] A. Kusina, F. Lyonnet, D. B. Clark, E. Godat, T. Jezo, K. Kovarik, F. I. Olness, I. Schienbein, J. Y. Yu, Vector boson production in pPb and PbPb collisions at the LHC and its impact on nCTEQ15 PDFs, Eur. Phys. J. C77 (7) (2017) 488. arXiv:1610.02925, doi:10.1140/epjc/s10052-017-5036-x.

[32] R. Angeles-Martinez, et al., Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects, Acta Phys. Polon. B46 (12) (2015) 2501–2534. arXiv:1507.05267, doi:10.5506/APhysPolB.46.2501.

[33] R. Hofstadter, R. W. McAllister, Electron Scattering From the Proton, Phys. Rev. 98 (1955) 217–218. doi:10.1103/PhysRev.98.217.

[34] R. W. Mcallister, R. Hofstadter, Elastic Scattering of 188-MeV Electrons From the Proton and the α Particle, Phys. Rev. 102 (1956) 851–856. doi:10.1103/PhysRev.102.851.

[35] G. Zweig, An SU(3) model for strong interaction symmetry and its breaking. Version 2, in: D. Lichtenberg, S. P. Rosen (Eds.), DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978, 1964, pp. 22–101.
URL `https://inspirehep.net/record/4674/files/cern-th-412.pdf`

[36] M. Gell-Mann, A Schematic Model of Baryons and Mesons, Phys. Lett. 8 (1964) 214–215. doi:10.1016/S0031-9163(64)92001-3.

[37] E. D. Bloom, et al., High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees, Phys. Rev. Lett. 23 (1969) 930–934. doi:10.1103/PhysRevLett.23.930.

[38] M. Breidenbach, J. I. Friedman, H. W. Kendall, E. D. Bloom, D. H. Coward, H. C. DeStaebler, J. Drees, L. W. Mo, R. E. Taylor, Observed Behavior of Highly Inelastic electron-Proton Scattering, Phys. Rev. Lett. 23 (1969) 935–939. doi:10.1103/PhysRevLett.23.935.

[39] J. D. Bjorken, Asymptotic Sum Rules at Infinite Momentum, Phys. Rev. 179 (1969) 1547–1553. doi:10.1103/PhysRev.179.1547.

[40] W. Albrecht, F. W. Brasse, H. Dorner, W. Flauger, K. H. Frank, J. Gayler, H. Hultschig, J. May, E. Ganssauge, Inelastic electron-proton scattering at fixed four-momentum transfer of 0.773 and 1.935 (gev/c)-squared, Nucl. Phys. B13 (1969) 1–8. doi:10.1016/0550-3213(69)90359-9.

[41] R. P. Feynman, Very high-energy collisions of hadrons, Phys. Rev. Lett. 23 (1969) 1415–1417. doi:10.1103/PhysRevLett.23.1415.

[42] J. D. Bjorken, E. A. Paschos, Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon, Phys. Rev. 185 (1969) 1975–1982. doi:10.1103/PhysRev.185.1975.

[43] C. G. Callan, Jr., D. J. Gross, High-energy electroproduction and the constitution of the electric current, Phys. Rev. Lett. 22 (1969) 156–159. doi:10.1103/PhysRevLett.22.156.

[44] G. Miller, et al., Inelastic electron-Proton Scattering at Large Momentum Transfers, Phys. Rev. D5 (1972) 528. doi:10.1103/PhysRevD.5.528.

[45] D. J. Gross, F. Wilczek, Ultraviolet Behavior of Nonabelian Gauge Theories, Phys. Rev. Lett. 30 (1973) 1343–1346. doi:10.1103/PhysRevLett.30.1343.

[46] H. D. Politzer, Reliable Perturbative Results for Strong Interactions?, Phys. Rev. Lett. 30 (1973) 1346–1349. doi:10.1103/PhysRevLett.30.1346.

[47] J. Kuti, V. F. Weisskopf, Inelastic lepton - nucleon scattering and lepton pair production in the relativistic quark parton model, Phys. Rev. D4 (1971) 3418–3439. doi:10.1103/PhysRevD.4.3418.

[48] R. McElhaney, S. F. Tuan, Some consequences of a modified Kuti Weisskopf quark parton model, Phys. Rev. D8 (1973) 2267–2272. doi:10.1103/PhysRevD.8.2267.

[49] V. D. Barger, R. J. N. Phillips, Quark - parton model relations in deep inelastic lepton scattering, Nucl. Phys. B73 (1974) 269–294. doi:10.1016/0550-3213(74)90020-0.

[50] I. Hinchliffe, C. H. Llewellyn Smith, Detailed Treatment of Scaling Violations in Asymptotically Free Gauge Theories, Nucl. Phys. B128 (1977) 93. doi:10.1016/0550-3213(77)90302-9.

[51] D. J. Fox, et al., Test of Scale Invariance in High-Energy Muon Scattering, Phys. Rev. Lett. 33 (1974) 1504. doi:10.1103/PhysRevLett.33.1504.

[52] G. Altarelli, G. Parisi, Asymptotic Freedom in Parton Language, Nucl. Phys. B126 (1977) 298–318. doi:10.1016/0550-3213(77)90384-4.

[53] V. N. Gribov, L. N. Lipatov, Deep inelastic ep scattering in perturbation theory, Sov. J. Nucl. Phys. 15 (1972) 438–450.

[54] Y. L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics. (in russian), Sov. Phys. JETP 46 (1977) 641–653.

[55] L. N. Lipatov, The parton model and perturbation theory, Sov. J. Nucl. Phys. 20 (1975) 94–102, [Yad. Fiz.20,181(1974)].

[56] M. Gluck, E. Hoffmann, E. Reya, Scaling Violations and the Gluon Distribution of the Nucleon, Zeit. Phys. C13 (1982) 119. doi:10.1007/BF01547675.

[57] L. Baulieu, C. Kounnas, A Direct Method for Computing QCD Predictions for Deep Inelastic Structure Functions, Nucl. Phys. B155 (1979) 429–446. doi:10.1016/0550-3213(79)90279-7.

[58] E. Eichten, I. Hinchliffe, K. D. Lane, C. Quigg, Super Collider Physics, Rev. Mod. Phys. 56 (1984) 579–707. doi:10.1103/RevModPhys.56.579.

[59] H. Abramowicz, et al., Neutrino and anti-neutrinos Charged Current Inclusive Scattering in Iron in the Energy Range 20-GeV ¡ Neutrino Energy ¡ 300-GeV, Z. Phys. C17 (1983) 283. doi:10.1007/BF01571895.

[60] D. W. Duke, J. F. Owens, q^2 Dependent Parametrizations of Parton Distribution Functions, Phys. Rev. D30 (1984) 49–54. doi:10.1103/PhysRevD.30.49.

[61] A. D. Martin, R. G. Roberts, W. J. Stirling, Structure Function Analysis and psi, Jet, W, Z Production: Pinning Down the Gluon, Phys. Rev. D37 (1988) 1161. doi:10.1103/PhysRevD.37.1161.

[62] M. Diemoz, F. Ferroni, E. Longo, G. Martinelli, Parton Densities from Deep Inelastic Scattering to Hadronic Processes at Super Collider Energies, Z. Phys. C39 (1988) 21. doi:10.1007/BF01560387.

[63] P. Aurenche, R. Baier, M. Fontannaz, J. F. Owens, M. Werlen, The Gluon Contents of the Nucleon Probed with Real and Virtual Photons, Phys. Rev. D39 (1989) 3275. doi:10.1103/PhysRevD.39.3275.

[64] P. N. Harriman, A. D. Martin, W. J. Stirling, R. G. Roberts, Parton Distributions Extracted From Data on Deep Inelastic Lepton Scattering, Prompt Photon Production and the Drell-Yan Process, Phys. Rev. D42 (1990) 798–810. doi:10.1103/PhysRevD.42.798.

[65] J. G. Morfin, W.-K. Tung, Parton distributions from a global QCD analysis of deep inelastic scattering and lepton pair production, Z. Phys. C52 (1991) 13–30. doi:10.1007/BF01412323.

[66] M. Gluck, E. Reya, A. Vogt, Radiatively generated parton distributions for high-energy collisions, Z. Phys. C48 (1990) 471–482. doi:10.1007/BF01572029.

[67] A. D. Martin, W. J. Stirling, R. G. Roberts, Parton distributions of the proton, Phys. Rev. D50 (1994) 6734–6752. arXiv:hep-ph/9406315, doi:10.1103/PhysRevD.50.6734.

[68] H. L. Lai, J. Botts, J. Huston, J. G. Morfin, J. F. Owens, J.-w. Qiu, W. K. Tung, H. Weerts, Global QCD analysis and the CTEQ parton distributions, Phys. Rev. D51 (1995) 4763–4782. arXiv:hep-ph/9410404, doi:10.1103/PhysRevD.51.4763.

[69] F. Abe, et al., Inclusive jet cross section in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV, Phys. Rev. Lett. 77 (1996) 438–443. arXiv:hep-ex/9601008, doi:10.1103/PhysRevLett.77.438.

[70] J. Huston, E. Kovacs, S. Kuhlmann, H. L. Lai, J. F. Owens, D. E. Soper, W. K. Tung, Large transverse momentum jet production and the gluon distribution inside the proton, Phys. Rev. Lett. 77 (1996) 444–447. arXiv:hep-ph/9511386, doi:10.1103/PhysRevLett.77.444.

[71] S. Alekhin, Extraction of parton distributions and alpha(s) from DIS data within the Bayesian treatment of systematic errors, Eur. Phys. J. C10 (1999) 395–403. arXiv:hep-ph/9611213, doi:10.1007/s100520050763.

[72] M. Botje, A QCD analysis of HERA and fixed target structure function data, Eur. Phys. J. C14 (2000) 285–297. arXiv:hep-ph/9912439, doi:10.1007/s100520000358.

[73] V. Barone, C. Pascaud, F. Zomer, A new global analysis of deep inelastic scattering data, Eur. Phys. J. C12 (2000) 243–262. arXiv:hep-ph/9907512, doi:10.1007/s100529900198.

[74] W. T. Giele, S. A. Keller, D. A. Kosower, Parton distribution function uncertaintiesarXiv:hep-ph/0104052.

[75] J. Pumplin, et al., Uncertainties of predictions from parton distribution functions. 2. The Hessian method, Phys. Rev. D65 (2001) 014013. arXiv:hep-ph/0101032, doi:10.1103/PhysRevD.65.014013.

[76] J. Pumplin, et al., New generation of parton distributions with uncertainties from global QCD analysis, JHEP 07 (2002) 012. arXiv:hep-ph/0201195.

[77] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Uncertainties of predictions from parton distributions. I: Experi-

mental errors. ((T)), Eur. Phys. J. C28 (2003) 455–473. arXiv:hep-ph/0211080, doi:10.1140/epjc/s2003-01196-2.

[78] S. Moch, J. A. M. Vermaseren, A. Vogt, The three-loop splitting functions in qcd: The non-singlet case, Nucl. Phys. B688 (2004) 101–134. arXiv:hep-ph/0403192.

[79] A. Vogt, S. Moch, J. A. M. Vermaseren, The Three-loop splitting functions in QCD: The Singlet case, Nucl. Phys. B691 (2004) 129–181. arXiv:hep-ph/0404111, doi:10.1016/j.nuclphysb.2004.04.024.

[80] A. D. Martin, W. J. Stirling, R. S. Thorne, G. Watt, Parton distributions for the LHC, Eur. Phys. J. C63 (2009) 189. arXiv:0901.0002, doi:10.1140/epjc/s10052-009-1072-5.

[81] J. Gao, et al., CT10 next-to-next-to-leading order global analysis of QCD, Phys.Rev. D89 (2014) 033009. arXiv:1302.6246, doi:10.1103/PhysRevD.89.033009.

[82] S. Alekhin, J. Blümlein, S. Klein, S. Moch, The 3-, 4-, and 5-flavor NNLO Parton from Deep-Inelastic- Scattering Data and at Hadron Colliders, Phys. Rev. D81 (2010) 014032. arXiv:0908.2766, doi:10.1103/PhysRevD.81.014032.

[83] S. I. Alekhin, Global fit to the charged leptons DIS data: alpha(s), parton distributions, and high twists, Phys. Rev. D63 (2001) 094022. arXiv:hep-ph/0011002, doi:10.1103/PhysRevD.63.094022.

[84] S. Alekhin, Parton distribution functions from the precise NNLO QCD fit, JETP Lett. 82 (2005) 628–631. arXiv:hep-ph/0508248.

[85] F. Aaron, et al., Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA, JHEP 1001 (2010) 109. arXiv:0911.0884, doi:10.1007/JHEP01(2010)109.

[86] P. Jimenez-Delgado, E. Reya, Dynamical NNLO parton distributions, Phys.Rev. D79 (2009) 074023. arXiv:0810.4274, doi:10.1103/PhysRevD.79.074023.

[87] S. Forte, L. Garrido, J. I. Latorre, A. Piccione, Neural network parametrization of deep-inelastic structure functions, JHEP 05 (2002) 062. arXiv:hep-ph/0204232.

[88] R. D. Ball, et al., A first unbiased global NLO determination of parton distributions and their uncertainties, Nucl. Phys. B838 (2010) 136. arXiv:1002.4407, doi:10.1016/j.nuclphysb.2010.05.008.

[89] R. D. Ball, et al., Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO, Nucl.Phys. B855 (2012) 153. arXiv:1107.2652.

[90] R. D. Ball, et al., Parton distributions with LHC data, Nucl.Phys. B867 (2013) 244. arXiv:1207.1303, doi:10.1016/j.nuclphysb.2012.10.003.

[91] S. Alekhin, J. Bluemlein, S. Moch, The ABM parton distributions tuned to LHC data, Phys.Rev. D89 (2014) 054028. arXiv:1310.3059, doi:10.1103/PhysRevD.89.054028.

[92] M. Botje, et al., The PDF4LHC Working Group Interim RecommendationsarXiv:1101.0538.

[93] J. C. Collins, D. E. Soper, The Theorems of Perturbative QCD, Ann. Rev. Nucl. Part. Sci. 37 (1987) 383–409.

[94] J. C. Collins, D. E. Soper, G. F. Sterman, Factorization of Hard Processes in QCD, Adv. Ser. Direct. High Energy Phys. 5 (1989) 1–91. arXiv:hep-ph/0409313.

[95] J. M. Campbell, J. W. Huston, W. J. Stirling, Hard interactions of quarks and gluons: A primer for LHC physics, Rept. Prog. Phys. 70 (2007) 89. arXiv:hep-ph/0611148, doi:10.1088/0034-4885/70/1/R02.

[96] R. Hamberg, W. L. van Neerven, T. Matsuura, A complete calculation of the order $\alpha - s^2$ correction to the Drell-Yan K factor, Nucl. Phys. B359 (1991) 343–405, [Erratum: Nucl. Phys.B644,403(2002)]. doi:10.1016/S0550-3213(02)00814-3, 10.1016/0550-3213(91)90064-5.

[97] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, B. Mistlberger, Higgs Boson Gluon-Fusion Production in QCD at Three Loops, Phys. Rev. Lett. 114 (21) (2015) 212001. arXiv:1503.06056, doi:10.1103/PhysRevLett.114.212001.

[98] S. Forte, E. Laenen, P. Nason, J. Rojo, Heavy quarks in deep-inelastic scattering, Nucl. Phys. B834 (2010) 116–162. arXiv:1001.2312, doi:10.1016/j.nuclphysb.2010.03.014.

[99] M. Kramer, 1, F. I. Olness, D. E. Soper, Treatment of heavy quarks in deeply inelastic scattering, Phys. Rev. D62 (2000) 096007. arXiv:hep-ph/0003035, doi:10.1103/PhysRevD.62.096007.

[100] R. Thorne, Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections, Phys.Rev. D86 (2012) 074017. arXiv:1201.6180, doi:10.1103/PhysRevD.86.074017.

[101] R. Thorne, The effect on PDFs and $\alpha_S(M_Z^2)$ due to changes in flavour scheme and higher twist contributions, Eur.Phys.J. C74 (7) (2014) 2958. arXiv:1402.3536, doi:10.1140/epjc/s10052-014-2958-4.

[102] R. S. Thorne, A Variable-flavor number scheme for NNLO, Phys.Rev. D73 (2006) 054019. arXiv:hep-ph/0601245, doi:10.1103/PhysRevD.73.054019.

[103] M. Guzzi, P. M. Nadolsky, H.-L. Lai, C.-P. Yuan, General-Mass Treatment for Deep Inelastic Scattering at Two-Loop Accuracy, Phys.Rev. D86 (2012) 053005. arXiv:1108.5112, doi:10.1103/PhysRevD.86.053005.

[104] R. D. Ball, et al., Parton distributions from high-precision collider dataarXiv:1706.00428.

[105] C. D. White, R. S. Thorne, A Global Fit to Scattering Data with NLL BFKL Resummations, Phys. Rev. D75 (2007) 034005. arXiv:hep-ph/0611204, doi:10.1103/PhysRevD.75.034005.

[106] M. Bonvini, S. Marzani, J. Rojo, L. Rottoli, M. Ubiali, R. D. Ball, V. Bertone, S. Carrazza, N. P. Hartland, Parton distribu-

tions with threshold resummation, JHEP 09 (2015) 191. arXiv:1507.01006, doi:10.1007/JHEP09(2015)191.

[107] A. C. Benvenuti, et al., A high statistics measurement of the proton structure functions $f_2(x, q^2)$ and r from deep inelastic muon scattering at high q^2, Phys. Lett. B223 (1989) 485.

[108] A. C. Benvenuti, et al., A high statistics measurement of the deuteron structure functions $f_2(x, q^2)$ and r from deep inelastic muon scattering at high q^2, Phys. Lett. B237 (1990) 592.

[109] M. Arneodo, et al., Accurate measurement of F_2^d/F_2^p and $R_d - R_p$, Nucl. Phys. B487 (1997) 3–26. arXiv:hep-ex/9611022, doi:10.1016/S0550-3213(96)00673-6.

[110] M. Arneodo, et al., Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T, Nucl. Phys. B483 (1997) 3–43. arXiv:hep-ph/9610231, doi:10.1016/S0550-3213(96)00538-X.

[111] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock, A. Bodek, Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections, Phys. Lett. B282 (1992) 475–482. doi:10.1016/0370-2693(92)90672-Q.

[112] S. Tkachenko, et al., Measurement of the structure function of the nearly free neutron using spectator tagging in inelastic ^{2}H(e, e'p)X scattering with CLAS, Phys. Rev. C89 (2014) 045206, [Addendum: Phys. Rev.C90,059901(2014)]. arXiv:1402.2477, doi:10.1103/PhysRevC.90.059901, 10.1103/PhysRevC.89.045206.

[113] J. J. Aubert, et al., Production of charmed particles in 250-GeV μ^+ - iron interactions, Nucl. Phys. B213 (1983) 31–64. doi:10.1016/0550-3213(83)90174-8.

[114] W. G. Seligman, et al., Improved determination of alpha(s) from neutrino nucleon scattering, Phys. Rev. Lett. 79 (1997) 1213–1216. arXiv:hep-ex/9701017, doi:10.1103/PhysRevLett.79.1213.

[115] U.-K. Yang, et al., Measurements of F_2 and $xF_3^\nu - xF_3^{\bar{\nu}}$ from CCFR ν_μ–Fe and $\bar{\nu}_\mu$–Fe data in a physics model independent way, Phys. Rev. Lett. 86 (2001) 2742–2745. arXiv:hep-ex/0009041, doi:10.1103/PhysRevLett.86.2742.

[116] G. Onengut, et al., Measurement of nucleon structure functions in neutrino scattering, Phys. Lett. B632 (2006) 65–75. doi:10.1016/j.physletb.2005.10.062.

[117] G. P. Zeller, et al., A Precise determination of electroweak parameters in neutrino nucleon scattering, Phys. Rev. Lett. 88 (2002) 091802, [Erratum: Phys. Rev. Lett.90,239902(2003)]. arXiv:hep-ex/0110059, doi:10.1103/PhysRevLett.88.091802.

[118] M. Goncharov, et al., Precise measurement of dimuon production cross-sections in ν_μFe and $\bar{\nu}_\mu$Fe deep inelastic scattering at the Tevatron, Phys. Rev. D64 (2001) 112006. arXiv:hep-ex/0102049, doi:10.1103/PhysRevD.64.112006.

[119] D. A. Mason, Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon dataFERMILAB-THESIS-2006-01. doi:10.1103/PhysRevLett.99.192001.

[120] A. Kayis-Topaksu, et al., Leading order analysis of neutrino induced dimuon events in the CHORUS experiment, Nucl. Phys. B798 (2008) 1–16. arXiv:0804.1869, doi:10.1016/j.nuclphysb.2008.02.013.

[121] O. Samoylov, et al., A Precision Measurement of Charm Dimuon Production in Neutrino Interactions from the NOMAD Experiment, Nucl.Phys. B876 (2013) 339. arXiv:1308.4750, doi:10.1016/j.nuclphysb.2013.08.021.

[122] F. Aaron, et al., Inclusive Deep Inelastic Scattering at High Q^2 with Longitudinally Polarised Lepton Beams at HERA, JHEP 1209 (2012) 061. arXiv:1206.7007, doi:10.1007/JHEP09(2012)061.

[123] F. Aaron, et al., Measurement of the Inclusive $e^\pm p$ Scattering Cross Section at High Inelasticity y and of the Structure Function F_L, Eur.Phys.J. C71 (2011) 1579. arXiv:1012.4355, doi:10.1140/epjc/s10052-011-1579-4.

[124] A. Cooper Sarkar, Measurement of high-Q^2 neutral current deep inelastic e+p scattering cross sections with a longitudinally polarised positron beam at HERAarXiv:1208.6138.

[125] H. Abramowicz, et al., Measurement of high-Q^2 charged current deep inelastic scattering cross sections with a longitudinally polarised positron beam at HERA, Eur.Phys.J. C70 (2010) 945–963. arXiv:1008.3493, doi:10.1140/epjc/s10052-010-1498-9.

[126] H. Abramowicz, et al., Combination and QCD Analysis of Charm Production Cross Section Measurements in Deep-Inelastic ep Scattering at HERA, Eur.Phys.J. C73 (2013) 2311. arXiv:1211.1182, doi:10.1140/epjc/s10052-013-2311-3.

[127] F. D. Aaron, et al., Measurement of the Charm and Beauty Structure Functions using the H1 Vertex Detector at HERA, Eur. Phys. J. C65 (2010) 89–109. arXiv:0907.2643, doi:10.1140/epjc/s10052-009-1190-0.

[128] H. Abramowicz, et al., Measurement of beauty and charm production in deep inelastic scattering at HERA and measurement of the beauty-quark mass, JHEP 09 (2014) 127. arXiv:1405.6915, doi:10.1007/JHEP09(2014)127.

[129] J. Rojo, Progress in the NNPDF global analysis and the impact of the legacy HERA combination, in: Proceedings, 2015 European Physical Society Conference on High Energy Physics (EPS-HEP 2015), 2015. arXiv:1508.07731.
URL `http://inspirehep.net/record/1391143/files/arXiv:1508.07731.pdf`

[130] L. A. Harland-Lang, A. D. Martin, P. Motylinski, R. S. Thorne, The impact of the final HERA combined data on PDFs obtained from a global fit, Eur. Phys. J. C76 (4) (2016) 186. arXiv:1601.03413, doi:10.1140/epjc/s10052-016-4020-1.

[131] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, C. P. Yuan, CTEQ-TEA parton distribution functions and HERA Run I and II combined data, Phys. Rev. D95 (3) (2017) 034003. arXiv:1609.07968, doi:10.1103/PhysRevD.95.034003.

[132] E. L. Berger, J. Gao, C. S. Li, Z. L. Liu, H. X. Zhu, Charm-Quark Production in Deep-Inelastic Neutrino Scattering at Next-to-Next-to-Leading Order in QCD, Phys. Rev. Lett. 116 (21) (2016) 212002. arXiv:1601.05430, doi:10.1103/PhysRevLett.116.212002.

[133] M. Cacciari, G. P. Salam, G. Soyez, The Anti-k(t) jet clustering algorithm, JHEP 0804 (2008) 063. arXiv:0802.1189, doi:10.1088/1126-6708/2008/04/063.

[134] S. Catani, Y. L. Dokshitzer, M. H. Seymour, B. R. Webber, Longitudinally invariant K_t clustering algorithms for hadron hadron collisions, Nucl. Phys. B406 (1993) 187–224. doi:10.1016/0550-3213(93)90166-M.

[135] S. D. Ellis, D. E. Soper, Successive combination jet algorithm for hadron collisions, Phys. Rev. D48 (1993) 3160–3166. arXiv:hep-ph/9305266, doi:10.1103/PhysRevD.48.3160.

[136] Y. L. Dokshitzer, G. D. Leder, S. Moretti, B. R. Webber, Better jet clustering algorithms, JHEP 08 (1997) 001. arXiv:hep-ph/9707323, doi:10.1088/1126-6708/1997/08/001.

[137] G. C. Blazey, et al., Run II jet physics, in: QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999, 2000, pp. 47–77. arXiv:hep-ex/0005012.
URL `http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-00-092`

[138] V. Khachatryan, et al., Measurement and QCD analysis of double-differential inclusive jet cross-sections in pp collisions at sqrt(s) = 8 TeV and ratios to 2.76 and 7 TeV, JHEP 03 (2017) 156. arXiv:1609.05331, doi:10.1007/JHEP03(2017)156.

[139] S. Alioli, K. Hamilton, P. Nason, C. Oleari, E. Re, Jet pair production in POWHEG, JHEP 1104 (2011) 081. arXiv:1012.3380, doi:10.1007/JHEP04(2011)081.

[140] J. Currie, E. W. N. Glover, A. Gehrmann-De Ridder, T. Gehrmann, A. Huss, J. Pires, Single jet inclusive production for the individual jet p_T scale choice at the LHC, in: 23rd Cracow Epiphany Conference on Particle Theory Meets the First Data from LHC Run 2 Cracow, Poland, January 9-12, 2017, 2017. arXiv:1704.00923.
URL `https://inspirehep.net/record/1589454/files/arXiv:1704.00923.pdf`

[141] G. Aad, et al., Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector, Phys. Rev. D86 (2012) 014022. arXiv:1112.6297.

[142] A. Abulencia, et al., Measurement of the Inclusive Jet Cross Section using the k_{T} algorithm in $p\overline{p}$ Collisions at $\sqrt{s}$=1.96 TeV with the CDF II Detector, Phys. Rev. D75 (2007) 092006. arXiv:hep-ex/0701051, doi:10.1103/PhysRevD.75.092006.

[143] T. Aaltonen, et al., Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p-pbar Collider Using a Cone-Based Jet Algorithm, Phys. Rev. D78 (2008) 052006. arXiv:0807.2204, doi:10.1103/PhysRevD.78.052006.

[144] V. M. Abazov, et al., Measurement of the inclusive jet cross-section in $p\bar{p}$ collisions at $s^{(1/2)}$ =1.96-TeV, Phys. Rev. Lett. 101 (2008) 062001. arXiv:0802.2400, doi:10.1103/PhysRevLett.101.062001.

[145] V. M. Abazov, et al., Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV, Phys. Rev. D85 (2012) 052006. arXiv:1110.3771, doi:10.1103/PhysRevD.85.052006.

[146] G. Aad, et al., Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector, Eur. Phys. J. C71 (2011) 1512. arXiv:1009.5908, doi:10.1140/epjc/s10052-010-1512-2.

[147] G. Aad, et al., Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s}$=2.76 TeV and comparison to the inclusive jet cross section at $\sqrt{s}$=7 TeV using the ATLAS detector, Eur.Phys.J. C73 (2013) 2509. arXiv:1304.4739, doi:10.1140/epjc/s10052-013-2509-4.

[148] G. Aad, et al., Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s}$ = 7 TeV using 4.5 fb^1 of data with the ATLAS detector, JHEP 02 (2015) 153, [Erratum: JHEP09,141(2015)]. arXiv:1410.8857, doi:10.1007/JHEP02(2015)153, 10.1007/JHEP09(2015)141.

[149] S. Chatrchyan, et al., Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the CMS detector, Phys.Rev. D87 (2013) 112002. arXiv:1212.6660, doi:10.1103/PhysRevD.87.112002.

[150] S. Chatrchyan, et al., Measurement of the ratio of inclusive jet cross sections using the anti-k_T algorithm with radius parameters R=0.5 and 0.7 in pp collisions at $\sqrt{s}$ = 7TeV, Phys. Rev. D90 (7) (2014) 072006. arXiv:1406.0324, doi:10.1103/PhysRevD.90.072006.

[151] G. Aad, et al., Measurement of dijet cross sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector, JHEP 1405 (2014) 059. arXiv:1312.3524, doi:10.1007/JHEP05(2014)059.

[152] S. Chatrchyan, et al., Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s}$ = 7 TeV, Phys. Lett. B700 (2011) 187–206. arXiv:1104.1693, doi:10.1016/j.physletb.2011.05.027.

[153] A. M. Sirunyan, et al., Measurement of the triple-differential dijet cross section in proton-proton collisions at sqrt(s) = 8 TeV and constraints on parton distribution functionsarXiv:1705.02628.

[154] V. Khachatryan, et al., Measurement of the double-differential inclusive jet cross section in protonproton collisions at $\sqrt{s}$ = 13 TeV, Eur. Phys. J. C76 (8) (2016) 451. arXiv:1605.04436, doi:10.1140/epjc/s10052-016-4286-3.

[155] C. Collaboration, Measurement of Triple-Differential Dijet Cross Sections at $\sqrt{s}$ = 8 TeV with the CMS Detector and Constraints on Parton Distribution Functions.

[156] S. D. Ellis, Z. Kunszt, D. E. Soper, Two jet production in hadron collisions at order alpha-s**3 in QCD, Phys.Rev.Lett. 69

(1992) 1496–1499. doi:10.1103/PhysRevLett.69.1496.
[157] Z. Kunszt, D. E. Soper, Calculation of jet cross-sections in hadron collisions at order alpha-s**3, Phys. Rev. D46 (1992) 192–221. doi:10.1103/PhysRevD.46.192.
[158] Z. Nagy, Three jet cross-sections in hadron hadron collisions at next-to-leading order, Phys.Rev.Lett. 88 (2002) 122003. arXiv:hep-ph/0110315, doi:10.1103/PhysRevLett.88.122003.
[159] Z. Nagy, Next-to-leading order calculation of three-jet observables in hadron hadron collision, Phys. Rev. D68 (2003) 094002. arXiv:hep-ph/0307268, doi:10.1103/PhysRevD.68.094002.
[160] J. Gao, Z. Liang, D. E. Soper, H.-L. Lai, P. M. Nadolsky, C. P. Yuan, MEKS: a program for computation of inclusive jet cross sections at hadron colliders, Comput. Phys. Commun. 184 (2013) 1626–1642. arXiv:1207.0513, doi:10.1016/j.cpc.2013.01.022.
[161] J. Currie, E. W. N. Glover, J. Pires, NNLO QCD predictions for single jet inclusive production at the LHCarXiv:1611.01460.
[162] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover, J. Pires, Second order QCD corrections to jet production at hadron colliders: the all-gluon contribution, Phys.Rev.Lett. 110 (2013) 162003. arXiv:1301.7310, doi:10.1103/PhysRevLett.110.162003.
[163] J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, N. Glover, J. Pires, S. Wells, Second order QCD corrections to gluonic jet production at hadron colliders, PoS LL2014 (2014) 001. arXiv:1407.5558.
[164] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, T. A. Morgan, Precise QCD predictions for the production of a Z boson in association with a hadronic jetarXiv:1507.02850.
[165] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, T. A. Morgan, NNLO QCD corrections for Z boson plus jet production, in: Proceedings, 12th International Symposium on Radiative Corrections (Radcor 2015) and LoopFest XIV (Radiative Corrections for the LHC and Future Colliders): Los Angeles, CA, USA, June 15-19, 2015, 2016. arXiv:1601.04569.
URL `https://inspirehep.net/record/1415788/files/arXiv:1601.04569.pdf`
[166] S. Dittmaier, A. Huss, C. Speckner, Weak radiative corrections to dijet production at hadron colliders, JHEP 1211 (2012) 095. arXiv:1210.0438, doi:10.1007/JHEP11(2012)095.
[167] N. Kidonakis, J. F. Owens, Effects of higher-order threshold corrections in high-E(T) jet production, Phys. Rev. D63 (2001) 054019. arXiv:hep-ph/0007268, doi:10.1103/PhysRevD.63.054019.
[168] M. C. Kumar, S.-O. Moch, Phenomenology of threshold corrections for inclusive jet production at hadron colliders, Phys. Lett. B730 (2014) 122–129. arXiv:1309.5311, doi:10.1016/j.physletb.2014.01.034.
[169] M. Klasen, G. Kramer, M. Michael, Next-to-next-to-leading order contributions to jet photoproduction and determination of α_s, Phys. Rev. D89 (7) (2014) 074032. arXiv:1310.1724, doi:10.1103/PhysRevD.89.074032.
[170] D. de Florian, P. Hinderer, A. Mukherjee, F. Ringer, W. Vogelsang, Approximate next-to-next-to-leading order corrections to hadronic jet production, Phys.Rev.Lett. 112 (2014) 082001. arXiv:1310.7192, doi:10.1103/PhysRevLett.112.082001.
[171] S. Carrazza, J. Pires, Perturbative QCD description of jet data from LHC Run-I and Tevatron Run-II, JHEP 10 (2014) 145. arXiv:1407.7031, doi:10.1007/JHEP10(2014)145.
[172] B. J. A. Watt, P. Motylinski, R. S. Thorne, The Effect of LHC Jet Data on MSTW PDFs, Eur.Phys.J. C74 (2014) 2934. arXiv:1311.5703, doi:10.1140/epjc/s10052-014-2934-z.
[173] V. Khachatryan, et al., Constraints on parton distribution functions and extraction of the strong coupling constant from the inclusive jet cross section in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C75 (6) (2015) 288. arXiv:1410.6765, doi:10.1140/epjc/s10052-015-3499-1.
[174] S. D. Ellis, W. J. Stirling, Constraints on isospin breaking in the light quark sea from the Drell-Yan process, Phys. Lett. B256 (1991) 258–264. doi:10.1016/0370-2693(91)90684-I.
[175] A. D. Martin, A. J. Mathijsseen, W. J. Stirling, R. S. Thorne, B. J. A. Watt, G. Watt, Extended Parameterisations for MSTW PDFs and their effect on Lepton Charge Asymmetry from W DecaysarXiv:1211.1215.
[176] R. S. Towell, et al., Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea, Phys. Rev. D64 (2001) 052002. arXiv:hep-ex/0103030, doi:10.1103/PhysRevD.64.052002.
[177] Fermilab E906 experiment, Drell-Yan Measurements of Nucleon and Nuclear Structure with the Fermilab Main Injector, D. F. Geesaman and P. E. Reimer, spokespersons; `http://www.phy.anl.gov/mep/SeaQuest/index.html`.
[178] V. M. Abazov, et al., Measurement of the shape of the boson rapidity distribution for $p\bar{p} \to Z/\gamma^* \to e^+e^- + X$ events produced at $\sqrt{s}$=1.96-TeV, Phys. Rev. D76 (2007) 012003. arXiv:hep-ex/0702025, doi:10.1103/PhysRevD.76.012003.
[179] T. A. Aaltonen, et al., Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV, Phys. Lett. B692 (2010) 232–239. arXiv:0908.3914, doi:10.1016/j.physletb.2010.06.043.
[180] V. M. Abazov, et al., Measurement of the muon charge asymmetry in $p\bar{p} \to$ W+X $\to \mu\nu$ + X events at $\sqrt{s}$=1.96 TeV, Phys.Rev. D88 (2013) 091102. arXiv:1309.2591, doi:10.1103/PhysRevD.88.091102.
[181] V. M. Abazov, et al., Measurement of the electron charge asymmetry in $\boldsymbol{p\bar{p} \to W + X \to e\nu + X}$ decays in $\boldsymbol{p\bar{p}}$ collisions at $\boldsymbol{\sqrt{s} = 1.96}$ TeV, Phys. Rev. D91 (3) (2015) 032007, [Erratum: Phys. Rev.D91,no.7,079901(2015)]. arXiv:1412.2862, doi:10.1103/PhysRevD.91.032007, 10.1103/PhysRevD.91.079901.

[182] T. Aaltonen, et al., Direct Measurement of the W Production Charge Asymmetry in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV, Phys. Rev. Lett. 102 (2009) 181801. arXiv:0901.2169, doi:10.1103/PhysRevLett.102.181801.

[183] V. M. Abazov, et al., Measurement of the W Boson Production Charge Asymmetry in $p\bar{p} \to W+X \to e\nu+X$ Events at $\sqrt{s}$ = 1.96 TeV, Phys. Rev. Lett. 112 (15) (2014) 151803, [Erratum: Phys. Rev. Lett.114,no.4,049901(2015)]. arXiv:1312.2895, doi:10.1103/PhysRevLett.114.049901, 10.1103/PhysRevLett.112.151803.

[184] S. Chatrchyan, et al., Measurement of the Rapidity and Transverse Momentum Distributions of Z Bosons in pp Collisions at $\sqrt{s}$ = 7 TeV, Phys. Rev. D85 (2012) 032002. arXiv:1110.4973, doi:10.1103/PhysRevD.85.032002.

[185] G. Aad, et al., Measurement of the inclusive $W^{\pm}$ and Z/γ^{*} cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s}$= 7 TeV with the ATLAS detector, Phys.Rev. D85 (2012) 072004. arXiv:1109.5141.

[186] S. Chatrchyan, et al., Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s}$ = 7 TeV, JHEP 1312 (2013) 030. arXiv:1310.7291, doi:10.1007/JHEP12(2013)030.

[187] V. Khachatryan, et al., Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at 8 TeV, Eur. Phys. J. C75 (4) (2015) 147. arXiv:1412.1115, doi:10.1140/epjc/s10052-015-3364-2.

[188] G. Aad, et al., Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at $\sqrt{s}$=7 TeV with the ATLAS detector, Phys.Lett. B725 (2013) 223. arXiv:1305.4192, doi:10.1016/j.physletb.2013.07.049.

[189] G. Aad, et al., Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s}$ = 7 TeV using the ATLAS detector, JHEP 06 (2014) 112. arXiv:1404.1212, doi:10.1007/JHEP06(2014)112.

[190] S. Chatrchyan, et al., Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s}$ = 7 TeV, Phys.Rev.Lett. 109 (2012) 111806. arXiv:1206.2598, doi:10.1103/PhysRevLett.109.111806.

[191] S. Chatrchyan, et al., Measurement of the muon charge asymmetry in inclusive pp to WX production at $\sqrt{s}$ = 7 TeV and an improved determination of light parton distribution functions, Phys.Rev. D90 (2014) 032004. arXiv:1312.6283, doi:10.1103/PhysRevD.90.032004.

[192] M. Aaboud, et al., Precision measurement and interpretation of inclusive W^{+}, W^{-} and Z/γ^{*} production cross sections with the ATLAS detectorarXiv:1612.03016.

[193] R. Aaij, et al., Measurement of the cross-section for $Z \to e^{+}e^{-}$ production in pp collisions at $\sqrt{s}$ = 7 TeV, JHEP 1302 (2013) 106. arXiv:1212.4620, doi:10.1007/JHEP02(2013)106.

[194] R. Aaij, et al., Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s}$ = 7 TeV, JHEP 08 (2015) 039. arXiv:1505.07024, doi:10.1007/JHEP08(2015)039.

[195] R. Aaij, et al., Measurement of forward $\mathrm{Z} \to \mathrm{e^{+}e^{-}}$ production at $\sqrt{s}$ = 8 TeV, JHEP 05 (2015) 109. arXiv:1503.00963, doi:10.1007/JHEP05(2015)109.

[196] R. Aaij, et al., Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s}$ = 13 TeV, JHEP 09 (2016) 136. arXiv:1607.06495, doi:10.1007/JHEP09(2016)136.

[197] R. Aaij, et al., Measurement of the forward W boson cross-section in pp collisions at $\sqrt{s}$ = 7 TeV, JHEP 12 (2014) 079. arXiv:1408.4354, doi:10.1007/JHEP12(2014)079.

[198] R. Aaij, et al., Measurement of forward W and Z boson production in pp collisions at $\sqrt{s}$ = 8 TeV, JHEP 01 (2016) 155. arXiv:1511.08039, doi:10.1007/JHEP01(2016)155.

[199] R. Aaij, et al., Measurement of forward $W \to e\nu$ production in pp collisions at $\sqrt{s}$ = 8 TeV, JHEP 10 (2016) 030. arXiv:1608.01484, doi:10.1007/JHEP10(2016)030.

[200] R. V. Harlander, W. B. Kilgore, Next-to-next-to-leading order Higgs production at hadron colliders, Phys. Rev. Lett. 88 (2002) 201801. arXiv:hep-ph/0201206, doi:10.1103/PhysRevLett.88.201801.

[201] C. Anastasiou, L. J. Dixon, K. Melnikov, F. Petriello, High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO, Phys. Rev. D69 (2004) 094008. arXiv:hep-ph/0312266, doi:10.1103/PhysRevD.69.094008.

[202] C. Anastasiou, L. J. Dixon, K. Melnikov, F. Petriello, Dilepton rapidity distribution in the Drell-Yan process at NNLO in QCD, Phys. Rev. Lett. 91 (2003) 182002. arXiv:hep-ph/0306192, doi:10.1103/PhysRevLett.91.182002.

[203] K. Melnikov, F. Petriello, The W boson production cross section at the LHC through $O(\alpha_s^2)$, Phys. Rev. Lett. 96 (2006) 231803. arXiv:hep-ph/0603182, doi:10.1103/PhysRevLett.96.231803.

[204] K. Melnikov, F. Petriello, Electroweak gauge boson production at hadron colliders through O(alpha(s)**2), Phys. Rev. D74 (2006) 114017. arXiv:hep-ph/0609070, doi:10.1103/PhysRevD.74.114017.

[205] R. Gavin, Y. Li, F. Petriello, S. Quackenbush, FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order, Comput. Phys. Commun. 182 (2011) 2388–2403. arXiv:1011.3540, doi:10.1016/j.cpc.2011.06.008.

[206] R. Gavin, Y. Li, F. Petriello, S. Quackenbush, W Physics at the LHC with FEWZ 2.1, Comput.Phys.Commun. 184 (2013) 208–214. arXiv:1201.5896, doi:10.1016/j.cpc.2012.09.005.

[207] Y. Li, F. Petriello, Combining QCD and electroweak corrections to dilepton production in FEWZ, Phys.Rev. D86 (2012) 094034. arXiv:1208.5967, doi:10.1103/PhysRevD.86.094034.

[208] F. A. Berends, R. Kleiss, Hard Photon Effects in $W^{\pm}$ and Z^{0} Decay, Z.Phys. C27 (1985) 365. doi:10.1007/BF01548639.

[209] F. A. Berends, R. Kleiss, J. Revol, J. Vialle, QED Radiative Corrections and Radiative Decays of the Intermediate Weak

Bosons Produced in Proton - Anti-proton Collisions, Z.Phys. C27 (1985) 155. doi:10.1007/BF01642494.
[210] U. Baur, S. Keller, W. Sakumoto, QED radiative corrections to Z boson production and the forward backward asymmetry at hadron colliders, Phys.Rev. D57 (1998) 199–215. arXiv:hep-ph/9707301, doi:10.1103/PhysRevD.57.199.
[211] U. Baur, O. Brein, W. Hollik, C. Schappacher, D. Wackeroth, Electroweak radiative corrections to neutral current Drell-Yan processes at hadron colliders, Phys.Rev. D65 (2002) 033007. arXiv:hep-ph/0108274, doi:10.1103/PhysRevD.65.033007.
[212] C. Carloni Calame, G. Montagna, O. Nicrosini, A. Vicini, Precision electroweak calculation of the production of a high transverse-momentum lepton pair at hadron colliders, JHEP 0710 (2007) 109. arXiv:0710.1722, doi:10.1088/1126-6708/2007/10/109.
[213] S. Dittmaier, M. Huber, Radiative corrections to the neutral-current Drell-Yan process in the Standard Model and its minimal supersymmetric extension, JHEP 1001 (2010) 060. arXiv:0911.2329, doi:10.1007/JHEP01(2010)060.
[214] S. Catani, L. Cieri, G. Ferrera, D. de Florian, M. Grazzini, Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO, Phys.Rev.Lett. 103 (2009) 082001. arXiv:0903.2120, doi:10.1103/PhysRevLett.103.082001.
[215] S. Catani, M. Grazzini, An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC, Phys.Rev.Lett. 98 (2007) 222002. arXiv:hep-ph/0703012, doi:10.1103/PhysRevLett.98.222002.
[216] J. M. Campbell, R. K. Ellis, C. Williams, Vector boson pair production at the LHC, JHEP 1107 (2011) 018. arXiv:1105.0020, doi:10.1007/JHEP07(2011)018.
[217] R. Boughezal, J. M. Campbell, R. K. Ellis, C. Focke, W. Giele, X. Liu, F. Petriello, C. Williams, Color singlet production at NNLO in MCFM, Eur. Phys. J. C77 (1) (2017) 7. arXiv:1605.08011, doi:10.1140/epjc/s10052-016-4558-y.
[218] G. Bozzi, S. Catani, G. Ferrera, D. de Florian, M. Grazzini, Production of Drell-Yan lepton pairs in hadron collisions: Transverse-momentum resummation at next-to-next-to-leading logarithmic accuracy, Phys. Lett. B696 (2011) 207–213. arXiv:1007.2351, doi:10.1016/j.physletb.2010.12.024.
[219] C. Balazs, J.-w. Qiu, C. P. Yuan, Effects of QCD resummation on distributions of leptons from the decay of electroweak vector bosons, Phys. Lett. B355 (1995) 548–554. arXiv:hep-ph/9505203, doi:10.1016/0370-2693(95)00726-2.
[220] V. Khachatryan, et al., Measurement of the differential cross section and charge asymmetry for inclusive pp $\rightarrow$ W$^{\pm}$ + X production at $\sqrt{s}$ = 8 TeV, Eur. Phys. J. C76 (8) (2016) 469. arXiv:1603.01803, doi:10.1140/epjc/s10052-016-4293-4.
[221] G. Aad, et al., Determination of the strange quark density of the proton from ATLAS measurements of the W, Z cross sections, Phys.Rev.Lett.arXiv:1203.4051.
[222] J. C. Collins, D. E. Soper, G. F. Sterman, Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson Production, Nucl. Phys. B250 (1985) 199–224. doi:10.1016/0550-3213(85)90479-1.
[223] C. T. H. Davies, W. J. Stirling, Nonleading Corrections to the Drell-Yan Cross-Section at Small Transverse Momentum, Nucl. Phys. B244 (1984) 337–348. doi:10.1016/0550-3213(84)90316-X.
[224] R. K. Ellis, S. Veseli, W and Z transverse momentum distributions: Resummation in q_T space, Nucl. Phys. B511 (1998) 649–669. arXiv:hep-ph/9706526, doi:10.1016/S0550-3213(97)00655-X.
[225] J.-w. Qiu, X.-f. Zhang, QCD prediction for heavy boson transverse momentum distributions, Phys. Rev. Lett. 86 (2001) 2724–2727. arXiv:hep-ph/0012058, doi:10.1103/PhysRevLett.86.2724.
[226] F. Landry, R. Brock, P. M. Nadolsky, C. P. Yuan, Tevatron Run-1 Z boson data and Collins-Soper-Sterman resummation formalism, Phys. Rev. D67 (2003) 073016. arXiv:hep-ph/0212159, doi:10.1103/PhysRevD.67.073016.
[227] S. Mantry, F. Petriello, Transverse Momentum Distributions from Effective Field Theory with Numerical Results, Phys. Rev. D83 (2011) 053007. arXiv:1007.3773, doi:10.1103/PhysRevD.83.053007.
[228] T. Becher, M. Neubert, Drell-Yan Production at Small q_T, Transverse Parton Distributions and the Collinear Anomaly, Eur. Phys. J. C71 (2011) 1665. arXiv:1007.4005, doi:10.1140/epjc/s10052-011-1665-7.
[229] S. Catani, L. Cieri, D. de Florian, G. Ferrera, M. Grazzini, Universality of transverse-momentum resummation and hard factors at the NNLO, Nucl. Phys. B881 (2014) 414–443. arXiv:1311.1654, doi:10.1016/j.nuclphysb.2014.02.011.
[230] J.-w. Qiu, X.-f. Zhang, Role of the nonperturbative input in QCD resummed Drell-Yan Q_T distributions, Phys. Rev. D63 (2001) 114011. arXiv:hep-ph/0012348, doi:10.1103/PhysRevD.63.114011.
[231] T. Becher, G. Bell, C. Lorentzen, S. Marti, Transverse-momentum spectra of electroweak bosons near threshold at NNLO, JHEP 02 (2014) 004. arXiv:1309.3245, doi:10.1007/JHEP02(2014)004.
[232] N. Kidonakis, R. J. Gonsalves, NNLO soft-gluon corrections for the Z-boson and W-boson transverse momentum distributions, Phys. Rev. D89 (9) (2014) 094022. arXiv:1404.4302, doi:10.1103/PhysRevD.89.094022.
[233] R. Boughezal, A. Guffanti, F. Petriello, M. Ubiali, The impact of the LHC Z-boson transverse momentum data on PDF determinations, JHEP 07 (2017) 130. arXiv:1705.00343, doi:10.1007/JHEP07(2017)130.
[234] G. Aad, et al., Measurement of the transverse momentum and ϕ^*_η distributions of DrellYan lepton pairs in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector, Eur. Phys. J. C76 (5) (2016) 291. arXiv:1512.02192, doi:10.1140/epjc/s10052-016-4070-4.
[235] V. Khachatryan, et al., Measurement of the Z boson differential cross section in transverse momentum and rapidity in protonproton collisions at 8 TeV, Phys. Lett. B749 (2015) 187–209. arXiv:1504.03511, doi:10.1016/j.physletb.2015.07.065.

[236] G. Aad, et al., Measurement of the transverse momentum distribution of $Z/\gamma*$ bosons in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Phys.Lett. B705 (2011) 415–434. arXiv:1107.2381, doi:10.1016/j.physletb.2011.10.018.

[237] G. Aad, et al., Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, JHEP 09 (2014) 145. arXiv:1406.3660, doi:10.1007/JHEP09(2014)145.

[238] V. Khachatryan, et al., Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV, JHEP 02 (2017) 096. arXiv:1606.05864, doi:10.1007/JHEP02(2017)096.

[239] V. M. Abazov, et al., Measurement of the ϕ^*_η distribution of muon pairs with masses between 30 and 500 GeV in 10.4 fb^{-1} of $p\bar{p}$ collisions, Phys. Rev. D91 (7) (2015) 072002. arXiv:1410.8052, doi:10.1103/PhysRevD.91.072002.

[240] R. J. Gonsalves, J. Pawlowski, C.-F. Wai, QCD Radiative Corrections to Electroweak Boson Production at Large Transverse Momentum in Hadron Collisions, Phys. Rev. D40 (1989) 2245. doi:10.1103/PhysRevD.40.2245.

[241] H. Baer, M. H. Reno, A Complete O (alpha-s) event generator for $p\bar{p} \to W^+X \to e^+$ neutrino X with parton showering, Phys. Rev. D44 (1991) 3375–3378. doi:10.1103/PhysRevD.44.3375.

[242] P. B. Arnold, R. P. Kauffman, W and Z production at next-to-leading order: From large q(t) to small, Nucl. Phys. B349 (1991) 381–413. doi:10.1016/0550-3213(91)90330-Z.

[243] E. Maina, S. Moretti, D. A. Ross, One loop weak corrections to gamma / Z hadroproduction at finite transverse momentum, Phys. Lett. B593 (2004) 143–150, [Erratum: Phys. Lett.B614,216(2005)]. arXiv:hep-ph/0403050, doi:10.1016/j.physletb.2004.04.043, 10.1016/j.physletb.2005.03.064.

[244] J. H. Kuhn, A. Kulesza, S. Pozzorini, M. Schulze, Logarithmic electroweak corrections to hadronic Z+1 jet production at large transverse momentum, Phys. Lett. B609 (2005) 277–285. arXiv:hep-ph/0408308, doi:10.1016/j.physletb.2005.01.059.

[245] J. H. Kuhn, A. Kulesza, S. Pozzorini, M. Schulze, One-loop weak corrections to hadronic production of Z bosons at large transverse momenta, Nucl. Phys. B727 (2005) 368–394. arXiv:hep-ph/0507178, doi:10.1016/j.nuclphysb.2005.08.019.

[246] T. Becher, X. Garcia i Tormo, Electroweak Sudakov effects in W, Z and γ production at large transverse momentum, Phys. Rev. D88 (1) (2013) 013009. arXiv:1305.4202, doi:10.1103/PhysRevD.88.013009.

[247] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, T. A. Morgan, The NNLO QCD corrections to Z boson production at large transverse momentum, JHEP 07 (2016) 133. arXiv:1605.04295, doi:10.1007/JHEP07(2016)133.

[248] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, T. A. Morgan, NNLO QCD corrections for Drell-Yan p_T^Z and ϕ^* observables at the LHC, JHEP 11 (2016) 094. arXiv:1610.01843, doi:10.1007/JHEP11(2016)094.

[249] R. Boughezal, J. M. Campbell, R. K. Ellis, C. Focke, W. T. Giele, X. Liu, F. Petriello, Z-boson production in association with a jet at next-to-next-to-leading order in perturbative QCD, Phys. Rev. Lett. 116 (15) (2016) 152001. arXiv:1512.01291, doi:10.1103/PhysRevLett.116.152001.

[250] R. Boughezal, X. Liu, F. Petriello, Phenomenology of the Z-boson plus jet process at NNLO, Phys. Rev. D94 (7) (2016) 074015. arXiv:1602.08140, doi:10.1103/PhysRevD.94.074015.

[251] R. D. Ball, et al., Parton distributions from high-precision collider dataarXiv:1706.00428.

[252] D. d'Enterria, J. Rojo, Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data, Nucl.Phys. B860 (2012) 311–338. arXiv:1202.1762.

[253] L. Apanasevich, et al., Evidence for parton k_T effects in high p_T particle production, Phys. Rev. Lett. 81 (1998) 2642–2645. arXiv:hep-ex/9711017, doi:10.1103/PhysRevLett.81.2642.

[254] L. Apanasevich, et al., Measurement of direct photon production at Tevatron fixed target energies, Phys. Rev. D70 (2004) 092009. arXiv:hep-ex/0407011, doi:10.1103/PhysRevD.70.092009.

[255] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Parton distributions and the LHC: W and Z production, Eur. Phys. J. C14 (2000) 133–145. arXiv:hep-ph/9907231, doi:10.1007/s100520050740, 10.1007/s100520000324.

[256] R. Ichou, D. d'Enterria, Sensitivity of isolated photon production at TeV hadron colliders to the gluon distribution in the proton, Phys. Rev. D82 (2010) 014015. arXiv:1005.4529, doi:10.1103/PhysRevD.82.014015.

[257] L. Carminati, G. Costa, D. D'Enterria, I. Koletsou, G. Marchiori, J. Rojo, M. Stockton, F. Tartarelli, Sensitivity of the LHC isolated-gamma+jet data to the parton distribution functions of the proton, Europhys. Lett. 101 (2013) 61002. arXiv:1212.5511, doi:10.1209/0295-5075/101/61002.

[258] G. Aad, et al., Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, JHEP 08 (2016) 005. arXiv:1605.03495, doi:10.1007/JHEP08(2016)005.

[259] M. Aaboud, et al., Measurement of the cross section for inclusive isolated-photon production in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detectorarXiv:1701.06882.

[260] T. A. Aaltonen, et al., Measurement of the Inclusive-Isolated Prompt-Photon Cross Section in $p\bar{p}$ Collisions using the full CDF Data Set, Submitted to: Phys. Rev. DarXiv:1703.00599.

[261] V. M. Abazov, et al., Measurement of the isolated photon cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$-TeV, Phys. Lett. B639 (2006) 151–158, [Erratum: Phys. Lett.B658,285(2008)]. arXiv:hep-ex/0511054, doi:10.1016/j.physletb.2007.06.047, 10.1016/j.physletb.2006.04.048.

[262] G. Aad, et al., Measurement of the inclusive isolated prompt photons cross section in pp collisions at $\sqrt{s} = 7$TeV with the

ATLAS detector using $4.6 fb^{1}$, Phys. Rev. D89 (5) (2014) 052004. arXiv:1311.1440, doi:10.1103/PhysRevD.89.052004.
[263] S. Chatrchyan, et al., Measurement of the Differential Cross Section for Isolated Prompt Photon Production in pp Collisions at 7 TeV, Phys.Rev. D84 (2011) 052011. arXiv:1108.2044, doi:10.1103/PhysRevD.84.052011.
[264] G. Aad, et al., Centrality, rapidity and transverse momentum dependence of isolated prompt photon production in lead-lead collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV measured with the ATLAS detector, Phys. Rev. C93 (3) (2016) 034914. arXiv:1506.08552, doi:10.1103/PhysRevC.93.034914.
[265] S. Chatrchyan, et al., Measurement of isolated photon production in pp and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, Phys. Lett. B710 (2012) 256–277. arXiv:1201.3093, doi:10.1016/j.physletb.2012.02.077.
[266] S. Catani, M. Fontannaz, J. P. Guillet, E. Pilon, Cross-section of isolated prompt photons in hadron hadron collisions, JHEP 05 (2002) 028. arXiv:hep-ph/0204023, doi:10.1088/1126-6708/2002/05/028.
[267] J. H. Kuhn, A. Kulesza, S. Pozzorini, M. Schulze, Electroweak corrections to hadronic photon production at large transverse momenta, JHEP 03 (2006) 059. arXiv:hep-ph/0508253, doi:10.1088/1126-6708/2006/03/059.
[268] J. M. Campbell, R. K. Ellis, C. Williams, Direct photon production at next-to-next-to-leading orderarXiv:1612.04333.
[269] M. D. Schwartz, Precision direct photon spectra at high energy and comparison to the 8 TeV ATLAS data, JHEP 09 (2016) 005. arXiv:1606.02313, doi:10.1007/JHEP09(2016)005.
[270] T. A. Aaltonen, et al., Combination of measurements of the top-quark pair production cross section from the Tevatron Collider, Phys.Rev. D89 (2014) 072001. arXiv:1309.7570, doi:10.1103/PhysRevD.89.072001.
[271] G. Aad, et al., Measurement of the cross section for top-quark pair production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector using final states with two high-pt leptons, JHEP 1205 (2012) 059. arXiv:1202.4892, doi:10.1007/JHEP05(2012)059.
[272] G. Aad, et al., Measurement of the $t\bar{t}$ production cross-section in pp collisions at sqrts = 7 TeV using kinematic information of lepton+jets eventsarXiv:ATLAS-CONF-2011-121, ATLAS-COM-CONF-2011-132.
[273] G. Aad, et al., Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 8$ TeV using $e\mu$ events with b-tagged jetsarXiv:ATLAS-CONF-2013-097, ATLAS-COM-CONF-2013-112.
[274] S. Chatrchyan, et al., Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV, JHEP 1402 (2014) 024. arXiv:1312.7582, doi:10.1007/JHEP02(2014)024.
[275] S. Chatrchyan, et al., Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV, JHEP 1211 (2012) 067. arXiv:1208.2671, doi:10.1007/JHEP11(2012)067.
[276] S. Chatrchyan, et al., Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV with lepton + jets final states, Phys.Lett. B720 (2013) 83–104. arXiv:1212.6682, doi:10.1016/j.physletb.2013.02.021.
[277] G. Aad, et al., Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector, Eur. Phys. J. C76 (10) (2016) 538. arXiv:1511.04716, doi:10.1140/epjc/s10052-016-4366-4.
[278] V. Khachatryan, et al., Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV, Eur. Phys. J. C75 (11) (2015) 542. arXiv:1505.04480, doi:10.1140/epjc/s10052-015-3709-x.
[279] A. M. Sirunyan, et al., Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functionsarXiv:1703.01630.
[280] M. Czakon, P. Fiedler, A. Mitov, The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$, Phys.Rev.Lett. 110 (2013) 252004. arXiv:1303.6254, doi:10.1103/PhysRevLett.110.252004.
[281] M. Czakon, A. Mitov, NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction, JHEP 1301 (2013) 080. arXiv:1210.6832, doi:10.1007/JHEP01(2013)080.
[282] P. Baernreuther, M. Czakon, A. Mitov, Percent level precision physics at the Tevatron: first genuine NNLO QCD corrections to q qbar $\rightarrow$ t tbar + XarXiv:1204.5201.
[283] M. Czakon, D. Heymes, A. Mitov, Dynamical scales for multi-TeV top-pair production at the LHC, JHEP 04 (2017) 071. arXiv:1606.03350, doi:10.1007/JHEP04(2017)071.
[284] M. Czakon, D. Heymes, A. Mitov, High-precision differential predictions for top-quark pairs at the LHC, Phys. Rev. Lett. 116 (8) (2016) 082003. arXiv:1511.00549, doi:10.1103/PhysRevLett.116.082003.
[285] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, M. Zaro, Top-pair production at the LHC through NNLO QCD and NLO EWarXiv:1705.04105.
[286] D. Pagani, I. Tsinikos, M. Zaro, The impact of the photon PDF and electroweak corrections on $t\bar{t}$ distributions, Eur. Phys. J. C76 (9) (2016) 479. arXiv:1606.01915, doi:10.1140/epjc/s10052-016-4318-z.
[287] J. Gao, A. S. Papanastasiou, Top-quark pair-production and decay at high precisionarXiv:1705.08903.
[288] M. Czakon, D. Heymes, A. Mitov, fastNLO tables for NNLO top-quark pair differential distributionsarXiv:1704.08551.
[289] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera, J. Rojo, Pinning down the large-x gluon with NNLO top-quark pair differential distributions, JHEP 04 (2017) 044. arXiv:1611.08609, doi:10.1007/JHEP04(2017)044.
[290] M. Czakon, M. L. Mangano, A. Mitov, J. Rojo, Constraints on the gluon PDF from top quark pair production at hadron

colliders, JHEP 1307 (2013) 167. arXiv:1303.7215, doi:10.1007/JHEP07(2013)167.

[291] M. Beneke, P. Falgari, S. Klein, J. Piclum, C. Schwinn, et al., Inclusive Top-Pair Production Phenomenology with TOPIXS, JHEP 1207 (2012) 194. arXiv:1206.2454, doi:10.1007/JHEP07(2012)194.

[292] M. Guzzi, K. Lipka, S.-O. Moch, Top-quark pair production at hadron colliders: differential cross section and phenomenological applications with DiffTop, JHEP 01 (2015) 082. arXiv:1406.0386, doi:10.1007/JHEP01(2015)082.

[293] F. Maltoni, G. Ridolfi, M. Ubiali, b-initiated processes at the LHC: a reappraisal, JHEP 07 (2012) 022, [Erratum: JHEP04,095(2013)]. arXiv:1203.6393, doi:10.1007/JHEP04(2013)095, 10.1007/JHEP07(2012)022.

[294] S. Forte, D. Napoletano, M. Ubiali, Higgs production in bottom-quark fusion in a matched scheme, Phys. Lett. B751 (2015) 331–337. arXiv:1508.01529, doi:10.1016/j.physletb.2015.10.051.

[295] M. Brucherseifer, F. Caola, K. Melnikov, On the NNLO QCD corrections to single-top production at the LHC, Phys. Lett. B736 (2014) 58–63. arXiv:1404.7116, doi:10.1016/j.physletb.2014.06.075.

[296] M. Aaboud, et al., Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detectorarXiv:1702.02859.

[297] M. Aaboud, et al., Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector, JHEP 04 (2017) 086. arXiv:1609.03920, doi:10.1007/JHEP04(2017)086.

[298] S. P. Jones, A. D. Martin, M. G. Ryskin, T. Teubner, Probes of the small x gluon via exclusive J/ψ and Υ production at HERA and the LHC, JHEP 11 (2013) 085. arXiv:1307.7099, doi:10.1007/JHEP11(2013)085.

[299] R. Aaij, et al., Updated measurements of exclusive J/ψ and ψ(2S) production cross-sections in pp collisions at $\sqrt{s} = 7$ TeV, J. Phys. G41 (2014) 055002. arXiv:1401.3288, doi:10.1088/0954-3899/41/5/055002.

[300] T. L. Collaboration, Central exclusive production of J/ψ and $\psi(2S)$ mesons in pp collisions at $\sqrt{s} = 13$ TeV.

[301] B. B. Abelev, et al., Exclusive J/ψ photoproduction off protons in ultra-peripheral p-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, Phys. Rev. Lett. 113 (23) (2014) 232504. arXiv:1406.7819, doi:10.1103/PhysRevLett.113.232504.

[302] S. P. Jones, A. D. Martin, M. G. Ryskin, T. Teubner, Exclusive J/ψ and Υ photoproduction and the low x gluon, J. Phys. G43 (3) (2016) 035002. arXiv:1507.06942, doi:10.1088/0954-3899/43/3/035002.

[303] D. Yu. Ivanov, A. Schafer, L. Szymanowski, G. Krasnikov, Exclusive photoproduction of a heavy vector meson in QCD, Eur. Phys. J. C34 (3) (2004) 297–316, [Erratum: Eur. Phys. J.C75,no.2,75(2015)]. arXiv:hep-ph/0401131, doi:10.1140/epjc/s2004-01712-x, 10.1140/epjc/s10052-015-3298-8.

[304] S. P. Jones, A. D. Martin, M. G. Ryskin, T. Teubner, The exclusive J/ψ process at the LHC tamed to probe the low x gluon, Eur. Phys. J. C76 (11) (2016) 633. arXiv:1610.02272, doi:10.1140/epjc/s10052-016-4493-y.

[305] T. Carli, et al., A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project, Eur.Phys.J. C66 (2010) 503. arXiv:0911.2985, doi:10.1140/epjc/s10052-010-1255-0.

[306] M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, F. Stober, Theory-Data Comparisons for Jet Measurements in Hadron-Induced ProcessesarXiv:1109.1310.

[307] V. Bertone, R. Frederix, S. Frixione, J. Rojo, M. Sutton, aMCfast: automation of fast NLO computations for PDF fits, JHEP 1408 (2014) 166. arXiv:1406.7693, doi:10.1007/JHEP08(2014)166.

[308] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP 1407 (2014) 079. arXiv:1405.0301, doi:10.1007/JHEP07(2014)079.

[309] V. Bertone, S. Carrazza, N. P. Hartland, APFELgrid: a high performance tool for parton density determinations, Comput. Phys. Commun. 212 (2017) 205–209. arXiv:1605.02070, doi:10.1016/j.cpc.2016.10.006.

[310] S. J. Brodsky, G. R. Farrar, Scaling Laws at Large Transverse Momentum, Phys. Rev. Lett. 31 (1973) 1153–1156. doi:10.1103/PhysRevLett.31.1153.

[311] J. Pumplin, Parametrization dependence and $\Delta\chi^2$ in parton distribution fitting, Phys.Rev. D82 (2010) 114020. arXiv:0909.5176, doi:10.1103/PhysRevD.82.114020.

[312] R. D. Ball, E. R. Nocera, J. Rojo, The asymptotic behaviour of parton distributions at small and large x, Eur. Phys. J. C76 (7) (2016) 383. arXiv:1604.00024, doi:10.1140/epjc/s10052-016-4240-4.

[313] A. I. Signal, A. W. Thomas, Possible Strength of the Nonperturbative Strange Sea of the Nucleon, Phys. Lett. B191 (1987) 205. doi:10.1016/0370-2693(87)91348-7.

[314] S. Catani, D. de Florian, G. Rodrigo, W. Vogelsang, Perturbative generation of a strange-quark asymmetry in the nucleon, Phys. Rev. Lett. 93 (2004) 152003. arXiv:hep-ph/0404240, doi:10.1103/PhysRevLett.93.152003.

[315] R. D. Ball, et al., Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties, JHEP 05 (2010) 075. arXiv:0912.2276, doi:10.1007/JHEP05(2010)075.

[316] J. R. Andersen, et al., Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group ReportarXiv:1405.1067.

[317] F. James, M. Roos, Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations,

Comput. Phys. Commun. 10 (1975) 343–367. doi:10.1016/0010-4655(75)90039-9.

[318] J. Pumplin, D. R. Stump, W. K. Tung, Multivariate fitting and the error matrix in global analysis of data, Phys. Rev. D65 (2001) 014011. arXiv:hep-ph/0008191, doi:10.1103/PhysRevD.65.014011.

[319] P. M. Nadolsky, Z. Sullivan, PDF uncertainties in W H production at TevatronarXiv:hep-ph/0110378, doi:OSTI/788251.

[320] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, J. Rojo, Neural network determination of parton distributions: The nonsinglet case, JHEP 03 (2007) 039. arXiv:hep-ph/0701127.

[321] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, J. Rojo, Unbiased determination of the proton structure function f2(p) with estimation, JHEP 03 (2005) 080. arXiv:hep-ph/0501067.

[322] M. Dittmar, et al., Parton DistributionsarXiv:0901.2504.

[323] S. Carrazza, S. Forte, Z. Kassabov, J. I. Latorre, J. Rojo, An Unbiased Hessian Representation for Monte Carlo PDFs, Eur. Phys. J. C75 (8) (2015) 369. arXiv:1505.06736, doi:10.1140/epjc/s10052-015-3590-7.

[324] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, W. K. Tung, Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method, Phys. Rev. D65 (2001) 014012. arXiv:hep-ph/0101051, doi:10.1103/PhysRevD.65.014012.

[325] C. Patrignani, et al., Review of Particle Physics, Chin. Phys. C40 (10) (2016) 100001. doi:10.1088/1674-1137/40/10/100001.

[326] L. A. Harland-Lang, A. D. Martin, P. Motylinski, R. S. Thorne, Uncertainties on α_S in the MMHT2014 global PDF analysis and implications for SM predictionsarXiv:1506.05682.

[327] R. D. Ball, V. Bertone, L. Del Debbio, S. Forte, A. Guffanti, et al., Precision NNLO determination of $\alpha_s(M_Z)$ using an unbiased global parton set, Phys.Lett. B707 (2012) 66–71. arXiv:1110.2483, doi:10.1016/j.physletb.2011.11.053.

[328] H.-L. Lai, et al., Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions, Phys. Rev. D82 (2010) 054021. arXiv:1004.4624, doi:10.1103/PhysRevD.82.054021.

[329] L. A. Harland-Lang, A. D. Martin, P. Motylinski, R. S. Thorne, Charm and beauty quark masses in the MMHT2014 global PDF analysis, Eur. Phys. J. C76 (1) (2016) 10. arXiv:1510.02332, doi:10.1140/epjc/s10052-015-3843-5.

[330] A. H. Hoang, A. V. Manohar, Charm quark mass from inclusive semileptonic B decays, Phys. Lett. B633 (2006) 526–532. arXiv:hep-ph/0509195, doi:10.1016/j.physletb.2005.12.020.

[331] J. **Particle Data Group**, Beringer, et al., Review of Particle Physics (RPP), Phys.Rev. D86 (2012) 010001. doi:10.1103/PhysRevD.86.010001.

[332] G. Watt, R. S. Thorne, Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs, JHEP 1208 (2012) 052. arXiv:1205.4024, doi:10.1007/JHEP08(2012)052.

[333] J. Gao, P. Nadolsky, A meta-analysis of parton distribution functions, JHEP 1407 (2014) 035. arXiv:1401.0013, doi:10.1007/JHEP07(2014)035.

[334] S. Carrazza, J. I. Latorre, J. Rojo, G. Watt, A compression algorithm for the combination of PDF sets, Eur. Phys. J. C75 (2015) 474. arXiv:1504.06469, doi:10.1140/epjc/s10052-015-3703-3.

[335] J. Pumplin, J. Huston, H. L. Lai, W.-K. Tung, C. P. Yuan, Collider Inclusive Jet Data and the Gluon Distribution, Phys. Rev. D80 (2009) 014019. arXiv:0904.2424, doi:10.1103/PhysRevD.80.014019.

[336] J. Gao, Applications of the META parton distribution functions, in: Proceedings, 49th Rencontres de Moriond on QCD and High Energy Interactions: La Thuile, Italy, March 22-29, 2014, 2014, pp. 193–196.
URL `http://moriond.in2p3.fr/Proceedings/2014/Moriond_QCD_2014.pdf`

[337] S. Carrazza, Z. Kassabov, SMPDF Web: a web-based application for specialized minimal parton distribution functions, 2016. arXiv:1606.09248.
URL `https://inspirehep.net/record/1473186/files/arXiv:1606.09248.pdf`

[338] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data, Nucl.Phys. B855 (2012) 608–638. arXiv:1108.1758, doi:10.1016/j.nuclphysb.2011.10.018.

[339] R. D. Ball, et al., Reweighting NNPDFs: the W lepton asymmetry, Nucl. Phys. B849 (2011) 112–143. arXiv:1012.0836, doi:10.1016/j.nuclphysb.2011.03.017.

[340] S. Camarda, et al., QCD analysis of W- and Z-boson production at Tevatron, Eur. Phys. J. C75 (9) (2015) 458. arXiv:1503.05221, doi:10.1140/epjc/s10052-015-3655-7.

[341] R. Gauld, J. Rojo, Precision determination of the small-x gluon from charm production at LHCb, Phys. Rev. Lett. 118 (7) (2017) 072001. arXiv:1610.09373, doi:10.1103/PhysRevLett.118.072001.

[342] N. Armesto, H. Paukkunen, J. M. Penn, C. A. Salgado, P. Zurita, An analysis of the impact of LHC Run I protonlead data on nuclear parton densities, Eur. Phys. J. C76 (4) (2016) 218. arXiv:1512.01528, doi:10.1140/epjc/s10052-016-4078-9.

[343] N. Armesto, J. Rojo, C. A. Salgado, P. Zurita, Bayesian reweighting of nuclear PDFs and constraints from proton-lead collisions at the LHC, JHEP 11 (2013) 015. arXiv:1309.5371, doi:10.1007/JHEP11(2013)015.

[344] H. Paukkunen, P. Zurita, PDF reweighting in the Hessian matrix approach, JHEP 12 (2014) 100. arXiv:1402.6623,

doi:10.1007/JHEP12(2014)100.
[345] W. T. Giele, S. Keller, Implications of hadron collider observables on parton distribution function uncertainties, Phys. Rev. D58 (1998) 094023. arXiv:hep-ph/9803393, doi:10.1103/PhysRevD.58.094023.
[346] N. Sato, J. F. Owens, H. Prosper, Bayesian Reweighting for Global Fits, Phys. Rev. D89 (11) (2014) 114020. arXiv:1310.1089, doi:10.1103/PhysRevD.89.114020.
[347] H. Plothow-Besch, PDFLIB: A Library of all available parton density functions of the nucleon, the pion and the photon and the corresponding alpha-s calculations, Comput. Phys. Commun. 75 (1993) 396–416. doi:10.1016/0010-4655(93)90051-D.
[348] M. R. Whalley, D. Bourilkov, R. C. Group, The Les Houches accord PDFs (LHAPDF) and LHAGLUE, in: HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B, 2005, pp. 575–581. arXiv:hep-ph/0508110.
[349] D. Bourilkov, R. C. Group, M. R. Whalley, LHAPDF: PDF use from the Tevatron to the LHC, in: TeV4LHC Workshop - 4th meeting Batavia, Illinois, October 20-22, 2005, 2006. arXiv:hep-ph/0605240.
[350] A. Buckley, J. Ferrando, S. Lloyd, K. Nordstrm, B. Page, et al., LHAPDF6: parton density access in the LHC precision era, Eur.Phys.J. C75 (2015) 132. arXiv:1412.7420, doi:10.1140/epjc/s10052-015-3318-8.
[351] T. Gleisberg, et al., Event generation with SHERPA 1.1, JHEP 02 (2009) 007. arXiv:0811.4622, doi:10.1088/1126-6708/2009/02/007.
[352] A. C. Benvenuti, et al., A High Statistics Measurement of the Proton Structure Functions $F_2(x, Q^2)$ and R from Deep Inelastic Muon Scattering at High Q^2, Phys. Lett. B223 (1989) 485. doi:10.1016/0370-2693(89)91637-7.
[353] A. C. Benvenuti, et al., A High Statistics Measurement of the Deuteron Structure Functions $F_2(x, Q^2)$ and R from Deep Inelastic Muon Scattering at High Q^2, Phys. Lett. B237 (1990) 592. doi:10.1016/0370-2693(90)91231-Y.
[354] J. P. Berge, et al., A Measurement of Differential Cross-Sections and Nucleon Structure Functions in Charged Current Neutrino Interactions on Iron, Z. Phys. C49 (1991) 187–224. doi:10.1007/BF01555493.
[355] A. Aktas, et al., Measurement of F2($c\bar{c}$) and F2($b\bar{b}$) at high Q^2 using the H1 vertex detector at HERA, Eur. Phys. J. C40 (2005) 349–359. arXiv:hep-ex/0411046, doi:10.1140/epjc/s2005-02154-8.
[356] G. Moreno, et al., Dimuon production in proton - copper collisions at $\sqrt{s}$ = 38.8-GeV, Phys. Rev. D43 (1991) 2815–2836. doi:10.1103/PhysRevD.43.2815.
[357] F. Abe, et al., Forward-backward charge asymmetry of electron pairs above the Z^0 pole, Phys. Rev. Lett. 77 (1996) 2616–2621. doi:10.1103/PhysRevLett.77.2616.
[358] D. E. Acosta, et al., Measurement of the forward-backward charge asymmetry from $W \to e\nu$ production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV, Phys. Rev. D71 (2005) 051104. arXiv:hep-ex/0501023, doi:10.1103/PhysRevD.71.051104.
[359] V. M. Abazov, et al., Measurement of the muon charge asymmetry from W boson decays, Phys. Rev. D77 (2008) 011106. arXiv:0709.4254, doi:10.1103/PhysRevD.77.011106.
[360] V. M. Abazov, et al., Measurement of the ratios of the Z/gamma* + ¿= n jet production cross sections to the total inclusive Z/gamma* cross section in p anti-p collisions at s**(1/2) = 1.96-TeV, Phys. Lett. B658 (2008) 112–119. arXiv:hep-ex/0608052, doi:10.1016/j.physletb.2007.10.046.
[361] C. Balazs, C. P. Yuan, Soft gluon effects on lepton pairs at hadron colliders, Phys. Rev. D56 (1997) 5558–5583. arXiv:hep-ph/9704258, doi:10.1103/PhysRevD.56.5558.
[362] M. Guzzi, P. M. Nadolsky, B. Wang, Nonperturbative contributions to a resummed leptonic angular distribution in inclusive neutral vector boson production, Phys. Rev. D90 (1) (2014) 014030. arXiv:1309.1393, doi:10.1103/PhysRevD.90.014030.
[363] H.-L. Lai, et al., New parton distributions for collider physics, Phys. Rev. D82 (2010) 074024. arXiv:1007.2241, doi:10.1103/PhysRevD.82.074024.
[364] P. M. Nadolsky, et al., Implications of CTEQ global analysis for collider observables, Phys. Rev. D78 (2008) 013004. arXiv:0802.0007, doi:10.1103/PhysRevD.78.013004.
[365] J. F. Owens, A. Accardi, W. Melnitchouk, Global parton distributions with nuclear and finite-Q^2 corrections, Phys. Rev. D87 (9) (2013) 094012. arXiv:1212.1702, doi:10.1103/PhysRevD.87.094012.
[366] R. Aaij, et al., Inclusive W and Z production in the forward region at $\sqrt{s}$ = 7 TeV, JHEP 1206 (2012) 058. arXiv:1204.1620, doi:10.1007/JHEP06(2012)058.
[367] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. W. Huston, P. Nadolsky, J. Pumplin, C. R. Schmidt, D. Stump, C. P. Yuan, Heavy Flavors on CT14, PoS DIS2015 (2015) 166.
[368] C. Schmidt, J. Pumplin, D. Stump, C. P. Yuan, CT14QED PDFs from Isolated Photon Production in Deep Inelastic ScatteringarXiv:1509.02905.
[369] S. Chekanov, et al., Measurement of isolated photon production in deep inelastic ep scattering, Phys.Lett. B687 (2010) 16–25. arXiv:0909.4223, doi:10.1016/j.physletb.2010.02.045.
[370] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Parton distributions: A New global analysis, Eur. Phys. J. C4 (1998) 463–496. arXiv:hep-ph/9803445, doi:10.1007/s100529800904, 10.1007/s100520050220.
[371] R. S. Thorne, R. G. Roberts, An Ordered analysis of heavy flavor production in deep inelastic scattering, Phys. Rev. D57

(1998) 6871–6898. arXiv:hep-ph/9709442, doi:10.1103/PhysRevD.57.6871.
[372] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Physical gluons and high-E(T) jets, Phys. Lett. B604 (2004) 61–68. arXiv:hep-ph/0410230, doi:10.1016/j.physletb.2004.10.040.
[373] A. Martin, A. T. Mathijssen, W. Stirling, R. Thorne, B. Watt, et al., Extended Parameterisations for MSTW PDFs and their effect on Lepton Charge Asymmetry from W Decays, Eur.Phys.J. C73 (2) (2013) 2318. arXiv:1211.1215, doi:10.1140/epjc/s10052-013-2318-9.
[374] T. Bolton, Determining the CKM parameter V(cd) from nu N charm productionarXiv:hep-ex/9708014.
[375] R. S. Thorne, L. A. Harland-Lang, A. D. Martin, P. Motylinski, The Effect of Final HERA inclusive Cross Section Data MMHT2014 PDFs, in: Proceedings, 2015 European Physical Society Conference on High Energy Physics (EPS-HEP 2015), 2015. arXiv:1508.06621.
URL `http://inspirehep.net/record/1389858/files/arXiv:1508.06621.pdf`
[376] L. A. Harland-Lang, R. Nathvani, R. S. Thorne, A. D. Martin, MMHT PDFs: updates and outlook, Acta Phys. Polon. B48 (2017) 1011–1024. arXiv:1704.00162, doi:10.5506/APhysPolB.48.1011.
[377] S. Forte, J. I. Latorre, L. Magnea, A. Piccione, Determination of alpha(s) from scaling violations of truncated moments of structure functions, Nucl. Phys. B643 (2002) 477–500. arXiv:hep-ph/0205286, doi:10.1016/S0550-3213(02)00688-0.
[378] J. Rojo, J. I. Latorre, Neural network parametrization of spectral functions from hadronic tau decays and determination of qcd vacuum condensates, JHEP 01 (2004) 055. arXiv:hep-ph/0401047.
[379] R. D. Ball, et al., A determination of parton distributions with faithful uncertainty estimation, Nucl. Phys. B809 (2009) 1–63. arXiv:0808.1231, doi:10.1016/j.nuclphysb.2008.09.037.
[380] R. D. Ball, et al., Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering, Nucl. Phys. B823 (2009) 195–233. arXiv:0906.1958, doi:10.1016/j.nuclphysb.2009.08.003.
[381] S. Davidson, S. Forte, P. Gambino, N. Rius, A. Strumia, Old and new physics interpretations of the NuTeV anomaly, JHEP 02 (2002) 037. arXiv:hep-ph/0112302.
[382] R. D. Ball, et al., Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology, Nucl. Phys. B849 (2011) 296. arXiv:1101.1300.
[383] S. Lionetti, et al., Precision determination of α_s using an unbiased global NLO parton set, Phys. Lett. B701 (2011) 346–352. arXiv:1103.2369, doi:10.1016/j.physletb.2011.05.071.
[384] S. J. Brodsky, P. Hoyer, C. Peterson, N. Sakai, The Intrinsic Charm of the Proton, Phys. Lett. B93 (1980) 451–455. doi:10.1016/0370-2693(80)90364-0.
[385] S. Alekhin, J. Blümlein, S. Moch, Parton Distribution Functions and Benchmark Cross Sections at NNLO, Phys.Rev. D86 (2012) 054009. arXiv:1202.2281, doi:10.1103/PhysRevD.86.054009.
[386] H. Georgi, H. D. Politzer, Freedom at moderate energies: Masses in color dynamics, Phys. Rev. D14 (1976) 1829.
[387] A. Accardi, M. Christy, C. Keppel, P. Monaghan, W. Melnitchouk, et al., New parton distributions from large-x and low-Q^2 data, Phys.Rev. D81 (2010) 034016. arXiv:0911.2254, doi:10.1103/PhysRevD.81.034016.
[388] A. Accardi, W. Melnitchouk, J. Owens, M. Christy, C. Keppel, et al., Uncertainties in determining parton distributions at large x, Phys.Rev. D84 (2011) 014008. arXiv:1102.3686, doi:10.1103/PhysRevD.84.014008.
[389] H. Georgi, H. D. Politzer, Freedom at Moderate Energies: Masses in Color Dynamics, Phys. Rev. D14 (1976) 1829. doi:10.1103/PhysRevD.14.1829.
[390] L. T. Brady, A. Accardi, T. J. Hobbs, W. Melnitchouk, Next-to leading order analysis of target mass corrections to structure functions and asymmetries, Phys. Rev. D84 (2011) 074008, [Erratum: Phys. Rev.D85,039902(2012)]. arXiv:1108.4734, doi:10.1103/PhysRevD.84.074008, 10.1103/PhysRevD.85.039902.
[391] W. Melnitchouk, A. W. Schreiber, A. W. Thomas, Deep inelastic scattering from off-shell nucleons, Phys. Rev. D49 (1994) 1183–1198. arXiv:nucl-th/9311008, doi:10.1103/PhysRevD.49.1183.
[392] S. A. Kulagin, G. Piller, W. Weise, Shadowing, binding and off-shell effects in nuclear deep inelastic scattering, Phys. Rev. C50 (1994) 1154–1169. arXiv:nucl-th/9402015, doi:10.1103/PhysRevC.50.1154.
[393] S. A. Kulagin, R. Petti, Global study of nuclear structure functions, Nucl. Phys. A765 (2006) 126–187. arXiv:hep-ph/0412425, doi:10.1016/j.nuclphysa.2005.10.011.
[394] Y. Kahn, W. Melnitchouk, S. A. Kulagin, New method for extracting neutron structure functions from nuclear data, Phys. Rev. C79 (2009) 035205. arXiv:0809.4308, doi:10.1103/PhysRevC.79.035205.
[395] R. B. Wiringa, V. G. J. Stoks, R. Schiavilla, An Accurate nucleon-nucleon potential with charge independence breaking, Phys. Rev. C51 (1995) 38–51. arXiv:nucl-th/9408016, doi:10.1103/PhysRevC.51.38.
[396] R. Machleidt, The High precision, charge dependent Bonn nucleon-nucleon potential (CD-Bonn), Phys. Rev. C63 (2001) 024001. arXiv:nucl-th/0006014, doi:10.1103/PhysRevC.63.024001.
[397] F. Gross, A. Stadler, Covariant spectator theory of np scattering: Phase shifts obtained from precision fits to data below 350-MeV, Phys. Rev. C78 (2008) 014005. arXiv:0802.1552, doi:10.1103/PhysRevC.78.014005.
[398] N. Baillie, et al., Measurement of the neutron F2 structure function via spectator tagging with CLAS, Phys. Rev. Lett. 108

(2012) 142001, [Erratum: Phys. Rev. Lett.108,199902(2012)]. arXiv:1110.2770, doi:10.1103/PhysRevLett.108.199902, 10.1103/PhysRevLett.108.142001.

[399] V. Bertone, et al., A determination of $m_c(m_c)$ from HERA data using a matched heavy-flavor scheme, JHEP 08 (2016) 050. arXiv:1605.01946, doi:10.1007/JHEP08(2016)050.

[400] F. Giuli, et al., The photon PDF from high-mass Drell Yan data at the LHCarXiv:1701.08553.

[401] P. Belov, et al., Parton distribution functions at LO, NLO and NNLO with correlated uncertainties between orders, Eur. Phys. J. C74 (10) (2014) 3039. arXiv:1404.4234, doi:10.1140/epjc/s10052-014-3039-4.

[402] F. Hautmann, H. Jung, A. Lelek, V. Radescu, R. Zlebcik, Soft-gluon resolution scale in QCD evolution equationsarXiv:1704.01757.

[403] M. L. Mangano, J. Rojo, Cross Section Ratios between different CM energies at the LHC: opportunities for precision measurements and BSM sensitivity, JHEP 1208 (2012) 010. arXiv:1206.3557, doi:10.1007/JHEP08(2012)010.

[404] S. Chatrchyan, et al., Measurement of associated W + charm production in pp collisions at $\sqrt{s}$ = 7 TeV, JHEP 02 (2014) 013. arXiv:1310.1138, doi:10.1007/JHEP02(2014)013.

[405] C. Collaboration, Measurement of the inclusive $t\bar{t}$ cross section at $\sqrt{s}$ = 5.02 TeV.

[406] S. Carrazza, A. Ferrara, D. Palazzo, J. Rojo, APFEL Web: a web-based application for the graphical visualization of parton distribution functions, J.Phys. G42 (2015) 057001. arXiv:1410.5456, doi:10.1088/0954-3899/42/5/057001.

[407] R. D. Ball, et al., Theoretical issues in PDF determination and associated uncertainties, Phys.Lett. B723 (2013) 330. arXiv:1303.1189, doi:10.1016/j.physletb.2013.05.019.

[408] S. J. Brodsky, A. Kusina, F. Lyonnet, I. Schienbein, H. Spiesberger, R. Vogt, A review of the intrinsic heavy quark content of the nucleon, Adv. High Energy Phys. 2015 (2015) 231547. arXiv:1504.06287, doi:10.1155/2015/231547.

[409] J. Pumplin, H. L. Lai, W. K. Tung, The charm parton content of the nucleon, Phys. Rev. D75 (2007) 054029. arXiv:hep-ph/0701220, doi:10.1103/PhysRevD.75.054029.

[410] P. Jimenez-Delgado, T. Hobbs, J. Londergan, W. Melnitchouk, New limits on intrinsic charm in the nucleon from global analysis of parton distributions, Phys.Rev.Lett. 114 (8) (2015) 082002. arXiv:1408.1708, doi:10.1103/PhysRevLett.114.082002.

[411] R. D. Ball, V. Bertone, M. Bonvini, S. Forte, P. Groth Merrild, J. Rojo, L. Rottoli, Intrinsic charm in a matched general-mass scheme, Phys. Lett. B754 (2016) 49–58. arXiv:1510.00009, doi:10.1016/j.physletb.2015.12.077.

[412] R. D. Ball, M. Bonvini, L. Rottoli, Charm in Deep-Inelastic Scattering, JHEP 11 (2015) 122. arXiv:1510.02491, doi:10.1007/JHEP11(2015)122.

[413] T. Boettcher, P. Ilten, M. Williams, Direct probe of the intrinsic charm content of the proton, Phys. Rev. D93 (7) (2016) 074008. arXiv:1512.06666, doi:10.1103/PhysRevD.93.074008.

[414] G. Bailas, V. P. Goncalves, Phenomenological implications of the intrinsic charm in the Z boson production at the LHC, Eur. Phys. J. C76 (3) (2016) 105. arXiv:1512.06007, doi:10.1140/epjc/s10052-016-3941-z.

[415] P.-H. Beauchemin, V. A. Bednyakov, G. I. Lykasov, Yu. Yu. Stepanenko, Search for intrinsic charm in vector boson production accompanied by heavy flavor jets, Phys. Rev. D92 (3) (2015) 034014. arXiv:1410.2616, doi:10.1103/PhysRevD.92.034014.

[416] T. P. Stavreva, J. F. Owens, Direct Photon Production in Association With A Heavy Quark At Hadron Colliders, Phys. Rev. D79 (2009) 054017. arXiv:0901.3791, doi:10.1103/PhysRevD.79.054017.

[417] V. A. Bednyakov, M. A. Demichev, G. I. Lykasov, T. Stavreva, M. Stockton, Searching for intrinsic charm in the proton at the LHC, Phys. Lett. B728 (2014) 602–606. arXiv:1305.3548, doi:10.1016/j.physletb.2013.12.031.

[418] R. Vogt, S. J. Brodsky, QCD and intrinsic heavy quark predictions for leading charm and beauty hadroproduction, Nucl. Phys. B438 (1995) 261–277. arXiv:hep-ph/9405236, doi:10.1016/0550-3213(94)00543-N.

[419] B. A. Kniehl, G. Kramer, I. Schienbein, H. Spiesberger, Inclusive Charmed-Meson Production at the CERN LHC, Eur. Phys. J. C72 (2012) 2082. arXiv:1202.0439, doi:10.1140/epjc/s10052-012-2082-2.

[420] M. Gluck, E. Reya, I. Schienbein, Radiatively generated parton distributions of real and virtual photons, Phys. Rev. D60 (1999) 054019, [Erratum: Phys. Rev.D62,019902(2000)]. arXiv:hep-ph/9903337, doi:10.1103/PhysRevD.60.054019, 10.1103/PhysRevD.62.019902.

[421] V. Bertone, S. Carrazza, D. Pagani, M. Zaro, On the Impact of Lepton PDFs, JHEP 11 (2015) 194. arXiv:1508.07002, doi:10.1007/JHEP11(2015)194.

[422] D. de Florian, G. F. R. Sborlini, G. Rodrigo, QED corrections to the AltarelliParisi splitting functions, Eur. Phys. J. C76 (5) (2016) 282. arXiv:1512.00612, doi:10.1140/epjc/s10052-016-4131-8.

[423] D. de Florian, G. F. R. Sborlini, G. Rodrigo, Two-loop QED corrections to the Altarelli-Parisi splitting functions, JHEP 10 (2016) 056. arXiv:1606.02887, doi:10.1007/JHEP10(2016)056.

[424] V. Bertone, S. Carrazza, J. Rojo, APFEL: A PDF Evolution Library with QED corrections, Comput.Phys.Commun. 185 (2014) 1647. arXiv:1310.1394, doi:10.1016/j.cpc.2014.03.007.

[425] R. Sadykov, Impact of QED radiative corrections on Parton Distribution FunctionsarXiv:1401.1133.

[426] L. A. Harland-Lang, V. A. Khoze, M. G. Ryskin, Sudakov effects in photon-initiated processes, Phys. Lett. B761 (2016) 20–24. arXiv:1605.04935, doi:10.1016/j.physletb.2016.08.004.

[427] S. Kallweit, J. M. Lindert, S. Pozzorini, M. Schonherr, NLO QCD+EW predictions for $2\ell 2\nu$ diboson signatures at the LHCarXiv:1705.00598.

[428] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Parton distributions incorporating QED contributions, Eur. Phys. J. C39 (2005) 155. arXiv:hep-ph/0411040, doi:10.1140/epjc/s2004-02088-7.

[429] R. D. Ball, et al., Parton distributions with QED corrections, Nucl.Phys. B877 (2013) 290–320. arXiv:1308.0598, doi:10.1016/j.nuclphysb.2013.10.010.

[430] V. Bertone, S. Carrazza, Combining NNPDF3.0 and NNPDF2.3QED through the APFEL evolution code, PoS DIS2016 (2016) 031. arXiv:1606.07130.

[431] J. C. Bernauer, et al., Electric and magnetic form factors of the proton, Phys. Rev. C90 (1) (2014) 015206. arXiv:1307.6227, doi:10.1103/PhysRevC.90.015206.

[432] V. M. Budnev, I. F. Ginzburg, G. V. Meledin, V. G. Serbo, The Two photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation, Phys. Rept. 15 (1975) 181–281. doi:10.1016/0370-1573(75)90009-5.

[433] M. Gluck, C. Pisano, E. Reya, The Polarized and unpolarized photon content of the nucleon, Phys. Lett. B540 (2002) 75–80. arXiv:hep-ph/0206126, doi:10.1016/S0370-2693(02)02125-1.

[434] A. Martin, M. Ryskin, The photon PDF of the proton, Eur.Phys.J. C74 (2014) 3040. arXiv:1406.2118, doi:10.1140/epjc/s10052-014-3040-y.

[435] L. A. Harland-Lang, V. A. Khoze, M. G. Ryskin, The production of a diphoton resonance via photon-photon fusion, JHEP 03 (2016) 182. arXiv:1601.07187, doi:10.1007/JHEP03(2016)182.

[436] L. A. Harland-Lang, V. A. Khoze, M. G. Ryskin, The photon PDF in events with rapidity gaps, Eur. Phys. J. C76 (5) (2016) 255. arXiv:1601.03772, doi:10.1140/epjc/s10052-016-4100-2.

[437] H. Anlauf, H. D. Dahmen, P. Manakos, T. Mannel, T. Ohl, KRONOS: A Monte Carlo event generator for higher order electromagnetic radiative corrections to deep inelastic scattering at HERA, Comput. Phys. Commun. 70 (1992) 97–119. doi:10.1016/0010-4655(92)90095-G.

[438] A. Mukherjee, C. Pisano, Manifestly covariant analysis of the QED Compton process in e p —¿ e gamma p and e p —¿ e gamma X, Eur. Phys. J. C30 (2003) 477–486. arXiv:hep-ph/0306275, doi:10.1140/epjc/s2003-01308-0.

[439] A. V. Manohar, P. Nason, G. P. Salam, G. Zanderighi, The Photon Content of the ProtonarXiv:1708.01256.

[440] G. Aad, et al., Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, JHEP 08 (2016) 009. arXiv:1606.01736, doi:10.1007/JHEP08(2016)009.

[441] D. Bourilkov, Photon-induced Background for Dilepton Searches and Measurements in pp Collisions at 13 TeVarXiv:1606.00523.

[442] E. Accomando, J. Fiaschi, F. Hautmann, S. Moretti, C. H. Shepherd-Themistocleous, Photon-initiated production of a dilepton final state at the LHC: Cross section versus forward-backward asymmetry studies, Phys. Rev. D95 (3) (2017) 035014. arXiv:1606.06646, doi:10.1103/PhysRevD.95.035014.

[443] K. Mishra, et al., Electroweak Corrections at High Energies, in: Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013, 2013. arXiv:1308.1430.
URL `http://inspirehep.net/record/1246902/files/arXiv:1308.1430.pdf`

[444] J. M. Campbell, D. Wackeroth, J. Zhou, Study of weak corrections to Drell-Yan, top-quark pair, and dijet production at high energies with MCFM, Phys. Rev. D94 (9) (2016) 093009. arXiv:1608.03356, doi:10.1103/PhysRevD.94.093009.

[445] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, M. Zaro, Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons, JHEP 06 (2015) 184. arXiv:1504.03446, doi:10.1007/JHEP06(2015)184.

[446] S. Kallweit, J. M. Lindert, P. Maierhfer, S. Pozzorini, M. Schnherr, NLO electroweak automation and precise predictions for W+multijet production at the LHC, JHEP 04 (2015) 012. arXiv:1412.5157, doi:10.1007/JHEP04(2015)012.

[447] R. D. Ball, M. Bonvini, S. Forte, S. Marzani, G. Ridolfi, Higgs production in gluon fusion beyond NNLO, Nucl.Phys. B874 (2013) 746. arXiv:1303.3590, doi:10.1016/j.nuclphysb.2013.06.012.

[448] F. A. Dreyer, A. Karlberg, Vector-Boson Fusion Higgs Production at Three Loops in QCD, Phys. Rev. Lett. 117 (7) (2016) 072001. arXiv:1606.00840, doi:10.1103/PhysRevLett.117.072001.

[449] L. Randall, R. Sundrum, An Alternative to compactification, Phys.Rev.Lett. 83 (1999) 4690–4693. arXiv:hep-th/9906064, doi:10.1103/PhysRevLett.83.4690.

[450] L. Randall, R. Sundrum, A Large mass hierarchy from a small extra dimension, Phys.Rev.Lett. 83 (1999) 3370–3373. arXiv:hep-ph/9905221, doi:10.1103/PhysRevLett.83.3370.

[451] J. Alwall, et al., MadGraph/MadEvent v4: The New Web Generation, JHEP 09 (2007) 028. arXiv:0706.2334.

[452] I. Brivio, M. Trott, The Standard Model as an Effective Field TheoryarXiv:1706.08945.

[453] E. L. Berger, M. Guzzi, H.-L. Lai, P. M. Nadolsky, F. I. Olness, Constraints on color-octet fermions from a global parton

distribution analysis, Phys. Rev. D82 (2010) 114023. arXiv:1010.4315, doi:10.1103/PhysRevD.82.114023.

[454] M. Baak, J. Cth, J. Haller, A. Hoecker, R. Kogler, K. Mnig, M. Schott, J. Stelzer, The global electroweak fit at NNLO and prospects for the LHC and ILC, Eur. Phys. J. C74 (2014) 3046. arXiv:1407.3792, doi:10.1140/epjc/s10052-014-3046-5.

[455] M. Benayoun, P. David, L. DelBuono, F. Jegerlehner, Muon $g-2$ estimates: can one trust effective Lagrangians and global fits?, Eur. Phys. J. C75 (12) (2015) 613. arXiv:1507.02943, doi:10.1140/epjc/s10052-015-3830-x.

[456] G. Bozzi, L. Citelli, A. Vicini, Parton density function uncertainties on the W boson mass measurement from the lepton transverse momentum distribution, Phys. Rev. D91 (11) (2015) 113005. arXiv:1501.05587, doi:10.1103/PhysRevD.91.113005.

[457] G. Bozzi, L. Citelli, M. Vesterinen, A. Vicini, Prospects for improving the LHC W boson mass measurement with forward muons, Eur. Phys. J. C75 (12) (2015) 601. arXiv:1508.06954, doi:10.1140/epjc/s10052-015-3810-1.

[458] M. Aaboud, et al., Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detectorarXiv:1701.07240.

[459] S. Dimopoulos, S. Raby, F. Wilczek, Supersymmetry and the Scale of Unification, Phys. Rev. D24 (1981) 1681–1683. doi:10.1103/PhysRevD.24.1681.

[460] D. Becciolini, M. Gillioz, M. Nardecchia, F. Sannino, M. Spannowsky, Constraining new colored matter from the ratio of 3 to 2 jets cross sections at the LHC, Phys. Rev. D91 (1) (2015) 015010, [Addendum: Phys. Rev.D92,no.7,079905(2015)]. arXiv:1403.7411, doi:10.1103/PhysRevD.91.015010, 10.1103/PhysRevD.92.079905.

[461] M. Aaboud, et al., Determination of the strong coupling constant α_s from transverse energy-energy correlations in multijet events at $\sqrt{s} = 8$ TeV using the ATLAS detectorarXiv:1707.02562.

[462] S. Chatrchyan, et al., Measurement of the ratio of the inclusive 3-jet cross section to the inclusive 2-jet cross section in pp collisions at $\sqrt{s} = 7$ TeV and first determination of the strong coupling constant in the TeV range, Eur. Phys. J. C73 (10) (2013) 2604. arXiv:1304.7498, doi:10.1140/epjc/s10052-013-2604-6.

[463] S. Chatrchyan, et al., Determination of the top-quark pole mass and strong coupling constant from the t t-bar production cross section in pp collisions at $\sqrt{s} = 7$ TeV, Phys.Lett. B728 (2014) 496. arXiv:1307.1907, doi:10.1016/j.physletb.2014.08.040, 10.1016/j.physletb.2013.12.009.

[464] J. Rojo, Constraints on parton distributions and the strong coupling from LHC jet data, Int. J. Mod. Phys. A30 (2015) 1546005. arXiv:1410.7728, doi:10.1142/S0217751X15460057.

[465] D. S. M. Alves, J. Galloway, J. T. Ruderman, J. R. Walsh, Running Electroweak Couplings as a Probe of New Physics, JHEP 02 (2015) 007. arXiv:1410.6810, doi:10.1007/JHEP02(2015)007.

[466] M. Mojaza, S. J. Brodsky, X.-G. Wu, Systematic All-Orders Method to Eliminate Renormalization-Scale and Scheme Ambiguities in Perturbative QCD, Phys. Rev. Lett. 110 (2013) 192001. arXiv:1212.0049, doi:10.1103/PhysRevLett.110.192001.

[467] F. I. Olness, D. E. Soper, Correlated theoretical uncertainties for the one-jet inclusive cross section, Phys. Rev. D81 (2010) 035018. arXiv:0907.5052, doi:10.1103/PhysRevD.81.035018.

[468] M. Cacciari, N. Houdeau, Meaningful characterisation of perturbative theoretical uncertainties, JHEP 1109 (2011) 039. arXiv:1105.5152, doi:10.1007/JHEP09(2011)039.

[469] E. Bagnaschi, M. Cacciari, A. Guffanti, L. Jenniches, An extensive survey of the estimation of uncertainties from missing higher orders in perturbative calculationsarXiv:1409.5036.

[470] A. David, G. Passarino, How well can we guess theoretical uncertainties?, Phys. Lett. B726 (2013) 266–272. arXiv:1307.1843, doi:10.1016/j.physletb.2013.08.025.

[471] S. Forte, A. Isgr, G. Vita, Do we need N^{3}LO Parton Distributions?, Phys.Lett. B731 (2014) 136–140. arXiv:1312.6688, doi:10.1016/j.physletb.2014.02.027.

[472] R. Gupta, Introduction to lattice QCD: Course, in: Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2, 1997, pp. 83–219. arXiv:hep-lat/9807028.

[473] H.-W. Lin, et al., Parton Distributions and Lattice QCD calculationsarXiv:1710.aaaaa, doi:10.1103/PhysRevD.93.074005.

[474] M. Constantinou, Hadron Structure, PoS LATTICE2014 (2015) 001. arXiv:1411.0078.

[475] S. Syritsyn, Review of Hadron Structure Calculations on a Lattice, PoS LATTICE2013 (2014) 009. arXiv:1403.4686.

[476] H.-W. Lin, Lattice Hadron Structure: Applications within and beyond QCD, PoS LATTICE2012 (2012) 013. arXiv:1212.6849.

[477] H.-W. Lin, J.-W. Chen, S. D. Cohen, X. Ji, Flavor Structure of the Nucleon Sea from Lattice QCD, Phys. Rev. D91 (2015) 054510. arXiv:1402.1462, doi:10.1103/PhysRevD.91.054510.

[478] C. Alexandrou, K. Cichy, V. Drach, E. Garcia-Ramos, K. Hadjiyiannakou, K. Jansen, F. Steffens, C. Wiese, Lattice calculation of parton distributions, Phys. Rev. D92 (2015) 014502. arXiv:1504.07455, doi:10.1103/PhysRevD.92.014502.

[479] J.-W. Chen, S. D. Cohen, X. Ji, H.-W. Lin, J.-H. Zhang, Nucleon Helicity and Transversity Parton Distributions from Lattice QCD, Nucl. Phys. B911 (2016) 246–273. arXiv:1603.06664, doi:10.1016/j.nuclphysb.2016.07.033.

[480] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, F. Steffens, C. Wiese, Updated Lattice Results

for Parton Distributions, Phys. Rev. D96 (1) (2017) 014513. arXiv:1610.03689, doi:10.1103/PhysRevD.96.014513.

[481] S. Aoki, et al., Review of lattice results concerning low-energy particle physics, Eur. Phys. J. C77 (2) (2017) 112. arXiv:1607.00299, doi:10.1140/epjc/s10052-016-4509-7.

[482] W. Detmold, W. Melnitchouk, A. W. Thomas, Extraction of parton distributions from lattice QCD, Mod.Phys.Lett. A18 (2003) 2681–2698. arXiv:hep-lat/0310003, doi:10.1142/S0217732303012209.

[483] J. R. Green, M. Engelhardt, S. Krieg, J. W. Negele, A. V. Pochinsky, S. N. Syritsyn, Nucleon Structure from Lattice QCD Using a Nearly Physical Pion Mass, Phys. Lett. B734 (2014) 290–295. arXiv:1209.1687, doi:10.1016/j.physletb.2014.05.075.

[484] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, A. V. Avils-Casco, C. Wiese, The nucleon spin explained using lattice QCD simulationsarXiv:1706.02973.

[485] G. S. Bali, S. Collins, B. Glle, M. Gckeler, J. Najjar, R. H. Rdl, A. Schfer, R. W. Schiel, A. Sternbeck, W. Sldner, The moment $\langle x \rangle_{u-d}$ of the nucleon from $N_f = 2$ lattice QCD down to nearly physical quark masses, Phys. Rev. D90 (7) (2014) 074510. arXiv:1408.6850, doi:10.1103/PhysRevD.90.074510.

[486] X. Ji, Parton Physics on a Euclidean Lattice, Phys.Rev.Lett. 110 (26) (2013) 262002. arXiv:1305.1539, doi:10.1103/PhysRevLett.110.262002.

[487] X. Ji, Parton Physics from Large-Momentum Effective Field Theory, Sci. China Phys. Mech. Astron. 57 (2014) 1407–1412. arXiv:1404.6680, doi:10.1007/s11433-014-5492-3.

[488] Y.-Q. Ma, J.-W. Qiu, Extracting Parton Distribution Functions from Lattice QCD CalculationsarXiv:1404.6860.

[489] Y.-Q. Ma, J.-W. Qiu, QCD Factorization and PDFs from Lattice QCD Calculation, Int. J. Mod. Phys. Conf. Ser. 37 (2015) 1560041. arXiv:1412.2688, doi:10.1142/S2010194515600411.

[490] N. Sato, W. Melnitchouk, S. E. Kuhn, J. J. Ethier, A. Accardi, Iterative Monte Carlo analysis of spin-dependent parton distributions, Phys. Rev. D93 (7) (2016) 074005. arXiv:1601.07782, doi:10.1103/PhysRevD.93.074005.

[491] H. Baer, T. Barklow, K. Fujii, Y. Gao, A. Hoang, S. Kanemura, J. List, H. E. Logan, A. Nomerotski, M. Perelstein, et al., The International Linear Collider Technical Design Report - Volume 2: PhysicsarXiv:1306.6352.

[492] L. Linssen, A. Miyamoto, M. Stanitzki, H. Weerts, Physics and Detectors at CLIC: CLIC Conceptual Design ReportarXiv:1202.5940, doi:10.5170/CERN-2012-003.

[493] M. Bicer, et al., First Look at the Physics Case of TLEP, JHEP 01 (2014) 164. arXiv:1308.6176, doi:10.1007/JHEP01(2014)164.

[494] R. Contino, et al., Physics at a 100 TeV pp collider: Higgs and EW symmetry breaking studies, CERN Yellow Report (3) (2017) 255–440. arXiv:1606.09408, doi:10.23731/CYRM-2017-003.255.

[495] D. Boer, M. Diehl, R. Milner, R. Venugopalan, W. Vogelsang, et al., Gluons and the quark sea at high energies: Distributions, polarization, tomographyarXiv:1108.1713.

[496] J. Currie, T. Gehrmann, J. Niehues, Precise QCD predictions for the production of dijet final states in deep inelastic scattering, Phys. Rev. Lett. 117 (4) (2016) 042001. arXiv:1606.03991, doi:10.1103/PhysRevLett.117.042001.

[497] M. Ciafaloni, D. Colferai, G. Salam, A. Stasto, A Matrix formulation for small-x singlet evolution, JHEP 0708 (2007) 046. arXiv:0707.1453, doi:10.1088/1126-6708/2007/08/046.

[498] G. Altarelli, R. D. Ball, S. Forte, Small x Resummation with Quarks: Deep-Inelastic Scattering, Nucl. Phys. B799 (2008) 199–240. arXiv:0802.0032, doi:10.1016/j.nuclphysb.2008.03.003.

[499] M. Bonvini, S. Marzani, T. Peraro, Small-x resummation from HELL, Eur. Phys. J. C76 (11) (2016) 597. arXiv:1607.02153, doi:10.1140/epjc/s10052-016-4445-6.

[500] L. Rottoli, M. Bonvini, Towards small-x resummed parton distribution functions, 2017. arXiv:1707.01535.
URL `http://inspirehep.net/record/1608878/files/arXiv:1707.01535.pdf`

[501] M. Bonvini, et al., Parton distributions with small-x resummationarXiv:1709.aaaaa.

[502] J. Rojo, Parton Distributions at a 100 TeV Hadron Collider, PoS DIS2016 (2016) 275. arXiv:1605.08302.

[503] R. Gauld, J. Rojo, E. Slade, The small-x gluon from forward charm production: implications for a 100 TeV proton collider, in: 25th International Workshop on Deep Inelastic Scattering and Related Topics (DIS 2017) Birmingham, UK, April 3-7, 2017, 2017. arXiv:1705.04217.
URL `http://inspirehep.net/record/1599078/files/arXiv:1705.04217.pdf`

[504] T. Han, J. Sayre, S. Westhoff, Top-Quark Initiated Processes at High-Energy Hadron Colliders, JHEP 04 (2015) 145. arXiv:1411.2588, doi:10.1007/JHEP04(2015)145.

[505] S. Dawson, A. Ismail, I. Low, A Redux on ”When is the Top Quark a Parton?”, Phys.Rev. D90 (2014) 014005. arXiv:1405.6211, doi:10.1103/PhysRevD.90.014005.

[506] C. W. Bauer, N. Ferland, B. R. Webber, Standard Model Parton Distributions at Very High Energies, JHEP 08 (2017) 036. arXiv:1703.08562, doi:10.1007/JHEP08(2017)036.

[507] M. Ciafaloni, P. Ciafaloni, D. Comelli, Towards collinear evolution equations in electroweak theory, Phys.Rev.Lett. 88 (2002) 102001. arXiv:hep-ph/0111109, doi:10.1103/PhysRevLett.88.102001.

[508] J. Chen, T. Han, B. Tweedie, Electroweak Splitting Functions and High Energy ShoweringarXiv:1611.00788.