

The Structure of the Proton in the LHC Precision Era

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

^a*Institute of Nuclear and Particle Physics and Shanghai Key Laboratory for Particle Physics and Cosmology,
Shanghai Jiao Tong University, Shanghai, 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, emphasizing 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

1	Introduction	3
2	The global QCD analysis framework	4
2.1	A brief history of PDF fits	4
2.2	QCD factorization in deep-inelastic scattering	7
2.3	QCD factorization in hadronic collisions	9
2.4	The DGLAP evolution equations	9
2.5	Heavy quark structure functions	11
3	Experimental data and theoretical calculations	14
3.1	Overview	14
3.2	Deep-inelastic scattering	16
3.3	Inclusive jets	18
3.4	Inclusive gauge boson production	22
3.5	The p_T of Z bosons	27
3.6	Direct photon production	31
3.7	Top quark production	34

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

3.8	Charm production in pp collisions	38
3.9	Other PDF-sensitive process	38
3.10	Fast interfaces	39
4	Fitting methodology	43
4.1	PDF parametrization	43
4.1.1	Functional form	43
4.1.2	Sum rules	44
4.1.3	Flavour assumptions	44
4.2	Data/theory agreement and minimization	45
4.2.1	Definition of χ^2	45
4.2.2	Minimization of χ^2	46
4.3	PDF uncertainties	47
4.3.1	The Hessian method	47
4.3.2	The Monte Carlo method	49
4.3.3	The Lagrange multiplier method	50
4.4	Treatment of theory parametric uncertainties	51
4.5	Combined and reduced PDF sets	51
4.6	Approximate methods	55
4.6.1	Hessian profiling	55
4.7	Delivery	56
5	PDF analyses: state of the art	58
5.1	CT	58
5.2	MMHT	60
5.3	NNPDF	64
5.4	ABM	67
5.5	CJ	69
5.6	HERAFitter/xFitter	70
5.7	PDF efforts in ATLAS	73
5.8	PDF efforts in CMS	74
6	The proton structure	76
6.1	The gluon	76
6.2	Quark flavor separation	77
6.3	The strange content of the proton	79
6.4	The charm content of the proton	81
7	QED corrections and the photon PDF	83
7.1	Photon-induced processes	83
7.2	Electroweak corrections	89
8	Implications for LHC phenomenology	91
8.1	Higgs production cross-sections	91
8.2	PDF uncertainties and searches for new massive particles	92
8.3	Precision measurements of SM parameters	95
9	The future of PDF determinations	97
9.1	Theoretical uncertainties	97
9.1.1	MHOU on matrix element calculation	97
9.1.2	MHOU on determined PDFs	98
9.2	PDFs and lattice calculations	100
9.3	Parton distributions at future colliders	100

9.3.1	PDFs at the Large Hadron electron Collider	101
9.3.2	PDFs at a 100 TeV hadron collider	102

10 Conclusions **104**

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 for the calculation 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

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

59 This Report focuses only on one of the main aspects of the internal structure of nucleons, namely collinear un-
60 polarized PDFs, which are its most relevant feature for the exploration of the high-energy frontier at the LHC. There
61 are however many other fascinating aspects of the inner life of protons that due to space limitations cannot be cov-
62 ered here, since each of these important topics would deserve a separated Report. These include, among others, the
63 determination of its spin structure by means of the polarized PDFs [28, 29]; the nuclear modifications of the free-
64 proton PDFs [30, 31], relevant for the understanding of cold nuclear matter effects and the RHIC and LHC heavy-ion
65 program; or the to the three-dimensional imaging of nucleons in terms of transverse-momentum-dependent PDFs
66 (TMD-PDFs) [32]. We note only here that progress in some of these other aspects of the proton structure also affect
67 unpolarized PDF fits, for example nuclear corrections are frequently used to include neutrino DIS structure functions
68 taken on heavy nuclear targets.

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

85 **2. The global QCD analysis framework**

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

90 *2.1. A brief history of PDF fits*

91 The first direct measurement of proton structure was performed in the the pioneering experiments of Hofstadter on
92 elastic electron–nucleon scattering [33, 34]. By examining the deviations from the simple Mott scattering formulae

93 for point-like particles, the finite extent of the proton could be resolved, and the charge radius of the proton was
94 determined to be ~ 0.7 fm to within a few percent precision.

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

102 This situation changed in 1967 with the new experimental data on deep inelastic scattering (DIS) provided by the
103 SLAC 20 GeV linear accelerator. The SLAC-MIT collaboration were surprised to find that, in contrast to the case
104 of elastic lepton-proton scattering, the two form factors associated with the DIS cross section, the so-called structure
105 functions, were roughly independent of Q^2 [37, 38]. Moreover, these appeared to exhibit the scaling behaviour
106 predicted by Bjorken in 1969 [39], namely that the structure functions should depend only on the ratio of Q^2 to the
107 lepton energy loss ν in the proton rest frame¹.

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

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

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

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

139 Up to this point all DIS data was taken with fixed target experiments and hence limited to the higher $x \gtrsim 0.01$
140 region. This changed in 1992 when HERA high energy collider at DESY started taking data. This collided 920 GeV
141 protons with ≈ 27.5 GeV electrons for most of the run period, allowing the previously unexplored region down to

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

142 $x \sim 10^{-4}$ region to be probed at high Q^2 . By 1994 data from HERA were included for the first in the MRS(A) [67]
143 and CTEQ3 [68] global fits. These were also the first fits to include data from the Tevatron $p\bar{p}$ collider, with in
144 particular the W asymmetry data providing new information on the quark flavour decomposition. In the years that
145 followed further public releases within these approaches were produced, with in particular the increasingly precise
146 HERA measurements, and Tevatron data on jet production placing new and important constraints.

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

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

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

179 The approaches described above differ greatly in many respects, both in the choice of input data sets, and the
180 treatment of the corresponding theory predictions. However, while there are significant differences in the precise
181 choice of parameterisation, in all cases these rely on parameterising the PDFs in terms of reasonably contained,
182 $O(20 - 40)$, number of free variables. Moreover, while the precise prescription may vary, these are again all based
183 on the ‘Hessian’ linear error propagation procedure. A different approach, first discussed in [87], was taken by the
184 NNPDF collaboration. Here the PDF functional forms are based on neural networks, allowing many more ($O(200 -$
185 $300)$) free parameters. In addition, rather than constructing the PDF error from the χ^2 variation about the best fit
186 values, a ‘Monte Carlo’ (MC) approach is taken, with a large enough sample of PDF ‘replica’ sets each fit to randomly
187 distributed pseudo-data generated according to the measured data values and their uncertainties. The first NNPDF1.0
188 fit was reported in [88], at NLO and to a range of DIS and fixed target data. Subsequently, NNPDF2.1 [89] provided
189 the first NNLO PDF set within this approach, and included Tevatron data for the first time.

190 In recent times, data from the LHC has played an increasingly important role in PDF determination. The CT14 [18],
191 MMHT14 [19] and NNPDF2.3 [90] sets included LHC data on jets, W and Z boson production, and top pair produc-
192 tion for the first time. In addition, ABMP16 [20] was the first set from this group to include input from hadropro-
193 duction processes, with data on W and Z boson and single top production from the Tevatron and LHC and top pair

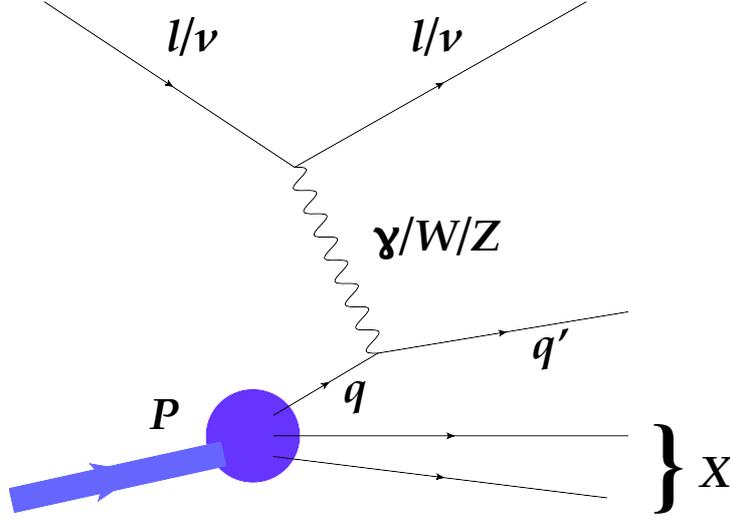


Figure 1: Schematic representation of the deep-inelastic scattering process. An energetic lepton (electron, muon or neutrino) scatters off one of the quarks in the proton by means of the interchange of a gauge boson (γ , W^\pm or Z). The large virtuality of the gauge boson, $Q \gg \Lambda_{\text{QCD}}$,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

194 production data at the LHC fit. As we will see in this review, these data, which are being produced with increasingly
 195 high precision, are now providing some of the most stringent constraints on the PDFs.

196 In parallel to these developments, there has been increasing focus on the use of PDFs as precise tools for LHC
 197 physics, emphasising the need for clear benchmarking exercises between sets and PDF combinations, to provide an
 198 overall PDF uncertainty. The PDF4LHC Working Group, formed in 2006, has played a significant role in this, with
 199 the benchmarking described in [5] leading to first so-called PDF4LHC recommendation [91] for the use of PDFs and
 200 their uncertainties at the LHC. This has subsequently been updated in [2] (see [1] for an alternative approach).

201 2.2. QCD factorization in deep-inelastic scattering

202 The importance of the deep-inelastic scattering (DIS) process for PDF fits cannot be under-emphasized. Instru-
 203 mental for the discovery of quarks, DIS represents since then the backbone of global PDF fits, and indeed historically
 204 the first PDF fits were based on DIS data alone. For three decades, DIS experiments were based on fixed-target kine-
 205 matics, where an energetic lepton (electron or a muon) was accelerated and then focused into an stationary target. A
 206 major breakthrough in our understanding of the proton structure started in 1992 with the commissioning of the HERA
 207 collider at DESY in Hamburg (Germany). The first (and so-far only one) lepton-proton collider, HERA operated
 208 between 1992 and 2007, collecting data on neutral and charged current events in electron and positron scattering of
 209 protons for a center-of-mass energy of $\sqrt{s} = 318$ GeV (after the Run II update).

210 The deep-inelastic scattering process is schematically represented in Fig. 1. In this process, an energetic lepton,
 211 which can be either charged (electron or muon) or neutral (a neutrino) scatters off a proton (or some other hadron) by
 212 means of the interchange of an electroweak gauge boson, a virtual photon γ^* or a W^\pm or Z boson. The large virtuality
 213 Q of the gauge boson, $Q \gg \Lambda_{\text{QCD}}$, ensures that the process can be described within QCD factorization in terms of
 214 coefficient functions and parton distributions, as we show below.

215 Denoting by k and k' the four-momenta of the incoming and outgoing leptons, by q the four-momentum of the
 216 exchanged gauge boson and P the incoming proton's momentum, the DIS process is defined in terms of a few invariant
 217 quantities, namely

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

218 Here x is known as the Bjorken variable, and although here it is defined purely in terms of the kinematics of the initial
 219 and final-state particles, it can be shown that in the parton model it corresponds to the momentum fraction carried

220 by the struck parton. Recall that by momentum conservation $q = k' - k$, and thus all the variables in Eq. (1) can be
 221 determined by the knowledge of the incoming momenta of the lepton k and of the proton P as well as the outgoing
 222 momentum of the lepton k' without any reference to the final hadronic state X . The center of mass energy W of the
 223 quark-photon collision is given by

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

224 where we have neglected proton mass effects. The value $x = 1$ corresponds to the *elastic limit*, where the proton and
 225 the lepton scatter off each other elastically so that the former remains intact after the collision.

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

$$\frac{d^2\sigma^{\text{NC},\ell^\pm}}{dx dQ^2}(x, y, Q^2) = \frac{2\pi\alpha^2}{xQ^4} \left[Y_+ F_2^{\text{NC}}(x, Q^2) \mp Y_- x F_3^{\text{NC}}(x, Q^2) - y^2 F_L^{\text{NC}}(x, Q^2) \right], \quad (3)$$

232 where we have defined

$$Y_\pm = 1 \pm (1-y)^2. \quad (4)$$

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

$$\tilde{\sigma}^{\text{NC},\ell^\pm}(x, y, Q^2) = \left[\frac{2\pi\alpha^2}{xQ^4} Y_+ \right]^{-1} \frac{d^2\sigma^{\text{NC},\ell^\pm}}{dx dQ^2}(x, y, Q^2), \quad (5)$$

234 which the main motivation that this reduced cross-section has a closer relation with the dominant structure function
 235 $F_2(x, Q^2)$. In the case of *charged current DIS*, either when neutrinos are used as projectiles or when the incoming
 236 charged leptons interact with the proton by means of the exchange of a charged weak gauge boson W^\pm , the differential
 237 cross-sections are given by:

$$\begin{aligned} \frac{d^2\sigma^{\text{CC},\ell^\pm}}{dx dQ^2}(x, y, Q^2) &= \frac{G_F^2}{4\pi x} \left(\frac{M_W^2}{M_W^2 + Q^2} \right)^2 \\ &\times \frac{1}{2} \left[Y_+ F_2^{\text{CC},\ell^\pm}(x, Q^2) \mp Y_- x F_3^{\text{CC},\ell^\pm}(x, Q^2) - y^2 F_L^{\text{CC},\ell^\pm}(x, Q^2) \right]. \end{aligned} \quad (6)$$

238 which similarly as in the NC case, is customarily rescaled to define a reduced cross section

$$\tilde{\sigma}^{\text{CC},\ell^\pm}(x, y, Q^2) = \left[\frac{G_F^2}{4\pi x} \left(\frac{M_W^2}{M_W^2 + Q^2} \right)^2 \right]^{-1} \frac{d^2\sigma^{\text{CC},\ell^\pm}}{dx dQ^2}(x, y, Q^2), \quad (7)$$

239 in a way that the experimental measurements have a closer relation to the DIS structure functions and thus with the
 240 underlying parton distributions.

241 By exploiting the *QCD factorization theorem*, it can be shown that the general expression of the DIS structure
 242 functions that appear in Eqns. (3) and (6) can be written schematically in the following way,

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

243 where $C_i(x/y, \alpha_s(Q^2))$ are known as the *coefficient functions* and $q_i(y, Q^2)$ are the parton distribution functions. The
 244 coefficient functions represent the cross-section for the partonic process $q_i + \gamma^* \rightarrow X$, and can be computed in per-
 245 turbation theory as a series expansion in the strong coupling α_s . They encode the short distance dynamics of the
 246 parton-boson collision. The PDFs $q_i(y, Q^2)$ on the other hand cannot be computed from first principles since they are

247 determined by long distance non-perturbative QCD dynamics. Therefore, they need to be parametrized and extracted
 248 from a global analysis of hard scattering measurements. The crucial importance of the factorized expression Eq. (8) is
 249 that while the coefficient functions (or in general the partonic cross-sections) are process dependent, the PDFs instead
 250 are universal, and therefore we can measure them in some processes and use them to provide pure predictions for
 251 other processes.

252 2.3. QCD factorization in hadronic collisions

253 Following the discussion of QCD factorization in the deep-inelastic scattering process, here we discuss QCD
 254 factorization in proton-proton collisions. Similar as the DIS structure functions, the production cross sections can
 255 be factorized as convolutions of two universal parton distribution functions and the process-dependent partonic cross
 256 sections. For example the Drell-Yan production cross section can be expressed as [? ?]

$$\frac{d^2\sigma^{DY,l^\pm}}{dydQ^2}(y, Q^2) = \sum_{a,b=\{q,g\}} \int_{\tau_1}^1 dx_1 q_{a/p}(x_1, \mu^2) \int_{\tau_2}^1 dx_2 q_{b/p}(x_2, \mu^2) \quad (9)$$

$$\times \frac{d^2\hat{\sigma}_{ab}^{DY,l^\pm}}{dydQ^2}(x_1, x_2, y, Q^2, \mu^2), \quad (10)$$

257 where y and Q^2 are rapidity and invariant mass square of the dilepton, the lower limits $\tau_{1,2} = \sqrt{Q^2/s} e^{\pm y}$ with s be
 258 the center of mass energy of the two incoming protons. The sum runs over all partonic channels. The partonic cross
 259 section describing hard scattering of QCD partons can be computed in perturbation theory as a series expansion of the
 260 strong coupling constant,

$$\frac{d^2\hat{\sigma}_{ab}^{DY,l^\pm}}{dydQ^2}(x_1, x_2, y, Q^2, \mu^2) = \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu^2)}{2\pi} \right)^n \frac{d^2\hat{\sigma}_{ab}^{(n)DY,l^\pm}}{dydQ^2}. \quad (11)$$

261 The factorization scale μ can be chosen so as to maintain a better convergence of the series. The physical cross
 262 sections as products of PDFs and partonic cross sections are independent of the choice of the factorization scale. In
 263 Drell-Yan production the conventional scale choice is $\mu^2 = Q^2$.

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

$$\sigma = \sum_{a,b=\{q,g\}} \int_{M^2}^s \frac{d\hat{s}}{\hat{s}} \mathcal{L}_{ab}(\hat{s}, \mu^2) \hat{\sigma}_{ab}(\hat{s}, M^2, \mu^2), \quad (12)$$

266 where \hat{s} is the center of mass energy of the two incoming partons, the parton-parton luminosity can be defined as [92]

$$\mathcal{L}_{ab}(\tau, \mu^2) = \frac{1}{s} \int_{\tau/s}^1 \frac{dx}{x} q_{a/p}(\tau/sx, \mu^2) q_{b/p}(x, \mu^2). \quad (13)$$

267 The partonic cross section depends only on the kinematic variable $z \equiv M^2/\hat{s}$ and the factorization scale

$$\hat{\sigma}_{ab}(\hat{s}, M^2, \mu^2) = \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu^2)}{2\pi} \right)^n C_{ab}^{(n)}(z, \mu^2). \quad (14)$$

268 The coefficient functions $C^{(n)}(z, \mu^2)$ are known to NNLO for Drell-Yan production [93] and to N³LO for Higgs boson
 269 production via gluon fusion in the limit of infinite top quark mass [94].

270 2.4. The DGLAP evolution equations

271 As indicated in Eq. (8), parton distributions $q_i(x, Q^2)$ depend on two variables: the Bjorken variable x , which at
 272 leading order can be identified with the momentum fraction carried by this specific parton; and Q^2 , the virtuality of
 273 the gauge boson. While the dependence of the PDFs with x is indeed determined by non-perturbative dynamics, and
 274 therefore cannot be computed from first principles, the dependence with Q^2 can instead be computed in perturbation
 275 theory. The reason is that the dependence of the PDFs in Q^2 arises when the initial-state collider singularities of

276 the partonic cross-section are regularised, and these singularities have a universal expression. For this reason, the
 277 dependence of the PDFs in Q^2 can in principle be computed in perturbation theory up to any given order.

278 The dependence of the PDFs in Q^2 can be determined by solving a series of integro-differential equations known
 279 as the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi evolution equations, or the DGLAP equations for short. These
 280 equations have the generic form

$$Q^2 \frac{\partial}{\partial Q^2} f_i(x, Q^2) = \sum_j P_{ij}(x, \alpha_s(Q^2)) \otimes f_j(x, Q^2), \quad (15)$$

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

$$P_{ij}(x, \alpha_s(Q^2)) = \sum_{n=0} \left(\frac{\alpha_s(Q^2)}{2\pi} \right)^{n+1} P_{ij}^{(n)}(x), \quad (16)$$

282 and where \otimes denotes the convolution

$$f(x) \otimes g(x) \equiv \int_x^1 \frac{dy}{y} f(y) g\left(\frac{x}{y}\right), \quad (17)$$

283 which appears ubiquitously in QCD calculation. The splitting functions Eq. (16) depend on the type of initial and
 284 final state parton that is involved in the splitting. At leading order, the DGLAP splitting functions are given by

$$P_{qq} = \frac{4}{3} \left[\frac{1+x^2}{(1-x)_+} \right], \quad (18)$$

$$P_{qg} = \frac{1}{2} [x^2 + (1-x^2)], \quad (19)$$

$$P_{gq} = \frac{4}{3} \left[\frac{1+(1-x)^2}{x} \right], \quad (20)$$

$$P_{gg} = 6 \left[\frac{1-x}{x} + x(x-1) + \frac{x}{(1-x)_+} \right]. \quad (21)$$

288 Note that both P_{gg} and P_{qg} have a singularity at $x=0$: this fact is responsible for the rapid growth with x of the gluons
 289 and consequently of the sea quarks as well at small- x .

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

$$\begin{aligned} \sigma(x, Q^2) &\equiv \sum_{i=1}^{n_f} (q_i + \bar{q}_i)(x, Q^2), \\ T_3(x, Q^2) &\equiv (u + \bar{u} - d - \bar{d})(x, Q^2), \\ T_8(x, Q^2) &\equiv (u + \bar{u} + d + \bar{d} - 2(s + \bar{s}))(x, Q^2), \end{aligned} \quad (22)$$

$$\begin{aligned} V(x, Q^2) &\equiv \sum_{i=1}^{n_f} (q_i - \bar{q}_i)(x, Q^2), \\ V_3(x, Q^2) &\equiv (u - \bar{u} - d + \bar{d})(x, Q^2), \\ V_8(x, Q^2) &\equiv (u - \bar{u} + d - \bar{d} - 2(s - \bar{s}))(x, Q^2), \end{aligned} \quad (23)$$

293 has the important property that all the PDF combinations except for Σ , known as the total quark singlet, evolve
 294 independently using their own specific splitting functions. These combinations, known as non-singlet flavour combi-
 295 nations, have thus a simple evolution structure. The singlet PDF instead evolves coupled to the gluon.

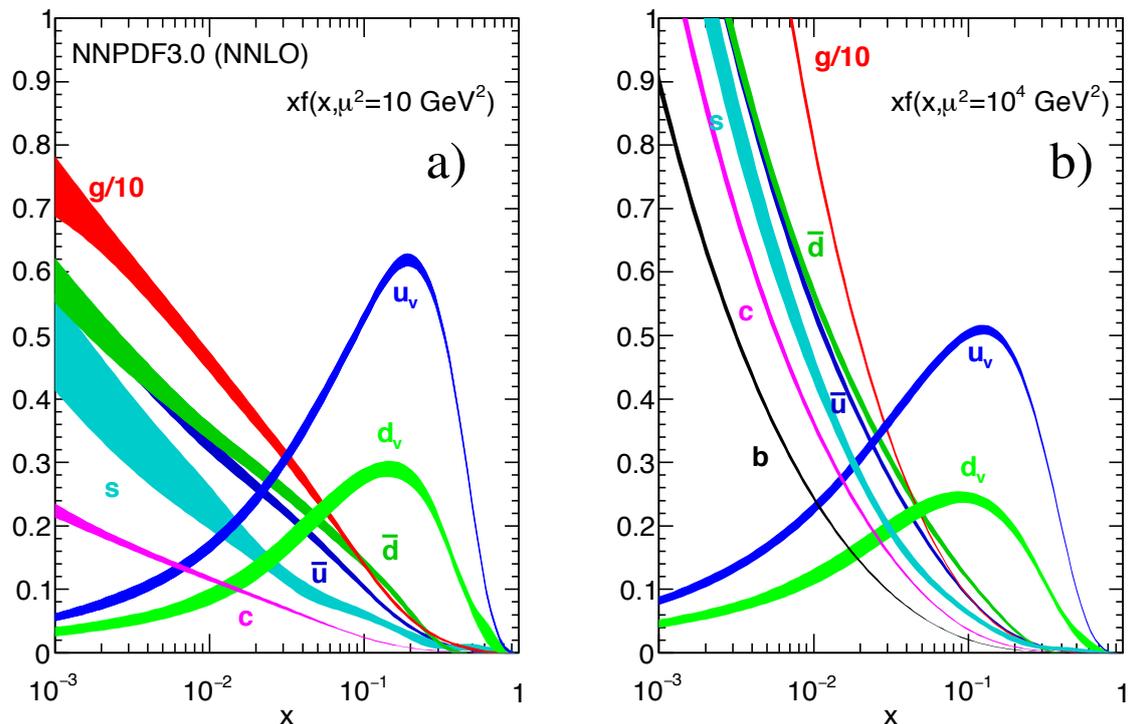


Figure 2: The effect of the DGLAP evolution in NNPDF3.0 NNLO, comparing the PDFs at a low scale of $Q^2 = 10 \text{ GeV}^2$ (left) with the same PDFs evolved up to a typical LHC scale of $Q^2 = 10^4 \text{ GeV}^2$ (right plot). In this plot the PDFs are shown together with the corresponding one-sigma PDF uncertainty band.

296 The splitting functions Eq. (16) are known up to $\mathcal{O}(\alpha_s^3)$ (NNLO), and thus PDF evolution can be performed up
 297 to this order. Several public codes implement the numerical solution of the DGLAP equations, either using x -space
 298 methods such as HOPPET, APFEL or QCDNUM, or Mellin (N -space) methods such as PEGASUS. These various codes
 299 have been benchmarked in various occasions, finding agreement at the level of $\mathcal{O}(10^{-5})$ or better.

300 In order to illustrate the impact of the DGLAP evolution on the PDFs, in Fig. 2 we show the NNPDF3.0 NNLO
 301 set comparing the PDFs at a low scale of $Q^2 = 10 \text{ GeV}^2$ (left) with the same PDFs evolved up to a typical LHC
 302 scale of $Q^2 = 10^4 \text{ GeV}^2$ (right plot). In this plot the PDFs are shown together with the corresponding one-sigma PDF
 303 uncertainty band. From this comparison we see that the effects of the evolution are relatively mild in the non-singlet
 304 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
 305 very steep growth at small- x . This steep growth is driven by the small- x structure of the splitting functions Eq. (21)
 306 Another interesting aspect that can be observed from Fig. 2 is that the valence PDFs xu_V and xd_V are integrable, and
 307 the fact that they have a similar shape but $u_V \approx 2d_V$ is a consequence of the valence sum rules which fix the value of
 308 their integrals.

309 2.5. Heavy quark structure functions

310 The production of heavy quarks in lepton-proton scattering deserves a separate discussion. For the HERA kine-
 311 matics, the contribution of the charm structure function F_2^c to the total inclusive structure function F_2^p can be as high
 312 as 25%, so it is crucial to compute it with high accuracy. The main difference with respect of lepto-production of light
 313 quarks is that here the effects of a finite mass of the heavy quark need to be taken into account. There are various
 314 theoretical schemes that have been proposed for the computation of heavy quark production in DIS:

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

- The fixed-flavor number scheme (FFN), 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-VFN) 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-VFN calculation of heavy quark DIS structure functions. Although for illustration purposes we will focus on the FONLL derivation from Ref. [95], the construction of related GM-VFN schemes such as ACOT, S-ACOT and TR 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

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

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 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

$$F^{(n_l)}(x, Q^2) = x \int_x^1 \frac{dy}{y} \sum_{i=q, \bar{q}, g} C_i^{(n_l)} \left(\frac{x}{y}, \frac{Q^2}{m^2}, \alpha_s^{(n_l)}(Q^2) \right) f_i^{(n_l)}(y, Q^2). \quad (26)$$

where now the massive coefficient functions $C_i^{(n_l)}$ include the full mass dependence and the heavy quark is not treated as a massless parton anymore. In this scheme, the PDFs and α_s satisfy evolution equations with n_l active quarks. The construction of the GM-VFN structure functions is based on two steps². First of all first of all to express PDFs and α_s in the massless scheme by means of the matching conditions

$$\alpha_s^{(n_l+1)}(Q^2) = \alpha_s^{(n_l)}(Q^2) + \sum_{i=2}^{\infty} c_i(L) \times \left(\alpha_s^{(n_l)}(m^2) \right)^i, \quad (27)$$

$$f_i^{(n_l+1)}(x, Q^2) = \int_x^1 \frac{dy}{y} \sum_{j=q, \bar{q}, g} K_{ij} \left(\frac{x}{y}, L, \alpha_s^{(n_l)}(Q^2) \right) f_j^{(n_l)}(y, Q^2), \quad (28)$$

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,

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

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

$$F^{(n_l, 0)}(x, Q^2) = x \int_x^1 \frac{dy}{y} \sum_{i=q, \bar{q}, g} B_i^{(0)} \left(\frac{x}{y}, \frac{Q^2}{m^2}, \alpha_s^{(n_l+1)}(Q^2) \right) f_i^{(n_l+1)}(y, Q^2), \quad (30)$$

²As mentioned, we follow the FONLL notation, but conceptually a very similar approach is followed in the construction of other GM-VFN schemes

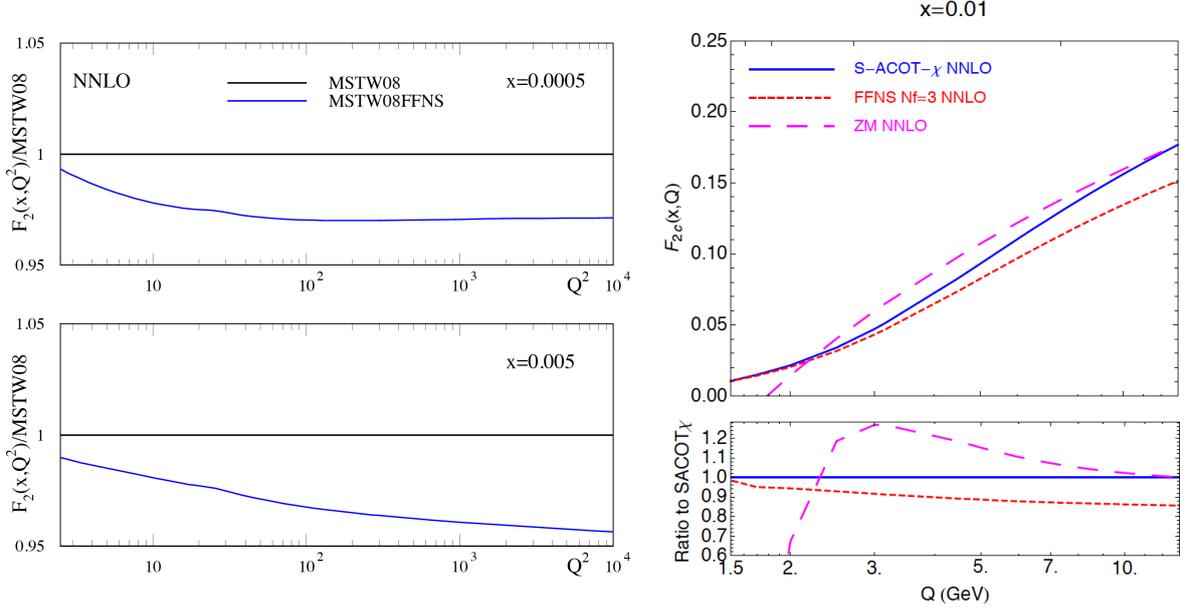


Figure 3: Left plot: the inclusive proton structure function $F_2^2(x, Q^2)$ at NNLO as a function of Q^2 for two different values of x in the RT' GM-VFNS as compared to the FFNS calculation. Right plot: the NNLO charm structure function $F_2^c(x, Q)$ as a function of Q for $x = 0.01$ comparing the S-ACOT-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

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

348 The FONLL approximation for F is then given by

$$F^{\text{FONLL}}(x, Q^2) = F^{(d)}(x, Q^2) + F^{(n)}(x, Q^2), \quad (31)$$

$$F^{(d)}(x, Q^2) \equiv [F^{(n+1)}(x, Q^2) - F^{(n,0)}(x, Q^2)] \quad (32)$$

349 where Eq. (32) is constructed out of the massless-scheme expression $F^{(n+1)}$, and the massless limit $F^{(n,0)}$ Eq. (30) of
 350 the massive-scheme expression. It is thus clear to see that in the limit where $Q \gg m$, the FONLL structure function
 351 reduces to the massless calculation, while for $Q \sim m$ the FONLL result coincides with the massive calculation up to
 352 subleading (higher order) terms.

353 To illustrate the numerical impact the heavy quark mass effects have in deep-inelastic structure functions, in Fig. 3
 354 we show the inclusive proton structure function $F_2^2(x, Q^2)$ at NNLO [96] as a function of Q^2 for two different values of
 355 x in the RT' GM-VFNS as compared to the FFNS calculation. We see that differences can be as large as few percent,
 356 comparable or larger with the precision of available DIS data. In the same figure we show the NNLO charm structure
 357 function $F_2^c(x, Q)$ as a function of Q for $x = 0.01$ comparing the S-ACOT- γ GM scheme with the corresponding ZM
 358 and FFN scheme calculations, where we can observe how the S-ACOT- γ calculation smoothly interpolates between
 359 the FFN scheme at low values of Q and the massless result at high Q .

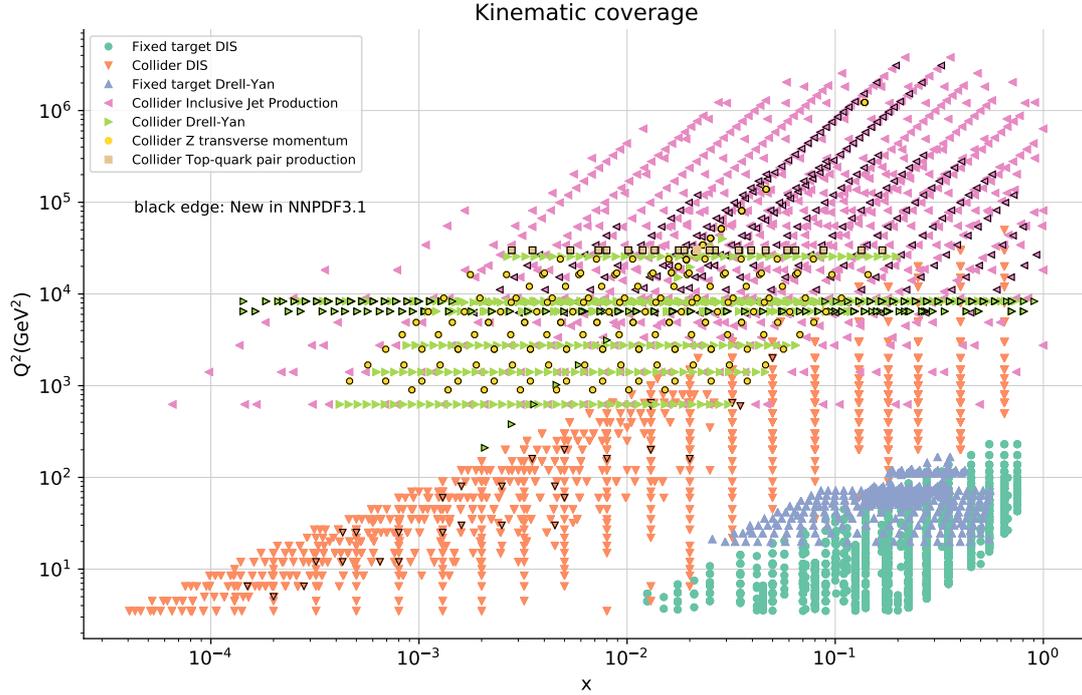


Figure 4: Typical kinematical coverage in the (x, Q^2) plane for the dataset included in a global analysis, in this case NNPDF3.1. For hadronic observables, leading order kinematics are assumed to map each data bin to a pair of (x, Q^2) values. The various datasets are clustered into families of related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

360 3. Experimental data and theoretical calculations

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

368 3.1. Overview

369 To begin this section, first of all we provide a general overview of the datasets that are available for PDF studies
 370 and then we move to discuss each process separately. In Fig. 4 we show a representative kinematical coverage in the
 371 (x, Q^2) plane for the dataset included in a global analysis, in this case the recent NNPDF3.1 fit [97]. For hadronic
 372 observables, leading order kinematics are assumed to map each data bin to a pair of (x, Q^2) values. In order to
 373 facilitate visualization, different measurements corresponding to the same family of processes have been clustered
 374 together. Only the data points that survive the kinematical cuts on Ref. [97] are plotted, which in the case of DIS
 375 corresponds to $Q^2 \geq 3.5 \text{ GeV}^2$ and $W^2 \geq 12.5 \text{ GeV}^2$.

376 From Fig. 4 we see the a global dataset provides a rather complete coverage of a wide region in the (x, Q^2) plane.
 377 The small- x and Q^2 region is dominated by the inclusive HERA structure function measurements, which provide
 378 information down to $x \simeq 3 \cdot 10^{-5}$. The large- x region is covered by various families of processes, from the fixed-target
 379 DIS structure functions at low Q to the collider jet, Drell-Yan and top-quark pair production at large Q . The very
 380 high Q region, up to a few TeV, is covered only by the inclusive jet production data from ATLAS and CMS. Recalling

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

381 that until relatively recent times most PDF fits were based only on DIS and DY fixed-target data with some Tevatron
382 measurements, it is remarkable the breath of the experimental information that is now available for PDF fits, including
383 processes like the $Z p_T$ and the $t\bar{t}$ differential distribution which have only recently been considered in this respect.

384 Another way to provide a general overview of the experimental data entering a modern global analysis is presented
385 in Table 1, where we summarize the various hard-scattering processes which are used to constrain PDFs in a global
386 analysis. In each case we indicate the hadronic-level process, the corresponding dominant partonic level process,
387 as well as the partons which are constrained by each specific process in a given range of x . Note that the latter are
388 necessarily approximate, and only indicate in a qualitative way the x region that dominates the PDF sensitivity of
389 each measurement. This table emphasizes the vital necessity that the global PDF fit has of including a set of input
390 measurements as broad as possible, in order to constrain all the various quark flavour combinations and the gluon in
391 the phenomenologically relevant region of x . We also note that the medium-to-low- x region, $x \lesssim 0.01$, is only covered
392 by the HERA collider structure functions and also by some of the LHC measurements. The deep low x region, below
393 the coverage of the HERA data which stops as $x \simeq 5 \cdot 10^{-5}$, can only be accessed via D meson production and perhaps
394 also by J/ψ production as well.

395 In the rest of this section, we discuss the various processes that can be used to constrain the parton distributions in

396 a global analysis one by one. We follow the same structure for each process: first of all we review the PDF sensitivity,
 397 then we discuss the available measurements, followed by a description of the state of the art of the corresponding
 398 theoretical calculations, to conclude with some representative illustrations of the impact of each family of processes
 399 in the PDF fit.

400 In the following, we restrict the discussion to theoretical calculations based on fixed-order perturbative QCD; see
 401 Refs. [98] and [99] for studies of the impact of the PDF fit of theory calculations based on all-order resummations of
 402 logarithmically enhanced terms at small- x and large- x respectively.

403 3.2. Deep-inelastic scattering

404 As for the rest of the processes in this section, first of all we review the PDF sensitivity, then the available data,
 405 the status of theory calculations and finally illustrate its impact on PDFs.

406 PDF sensitivity

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

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

$$415 \quad [F_2^\gamma, F_2^{\gamma Z}, F_2^Z] = x \sum_{i=1}^{n_f} [e_i^2, 2e_i g_V^i, g_V^{i2} + g_A^{i2}] (q_i + \bar{q}_i) , \quad (33)$$

$$[F_3^\gamma, F_3^{\gamma Z}, F_3^Z] = x \sum_{i=1}^{n_f} [0, 2e_i g_A^i, 2g_V^i g_A^i] (q_i - \bar{q}_i) , \quad (34)$$

416 while the longitudinal structure function vanishes in this model, $F_L = 0$, and the superscripts in the LHS indicate the
 417 gauge boson which is being interchanged, as well as the contribution from the γZ interference term. In Eqns. (33)
 418 and (34), e_i is the electric charge of the quark of flavour i and the weak couplings are given by $g_V^i = \pm \frac{1}{2} - 2e_i \sin^2 \theta_W^2$
 419 and $g_A^i = \pm \frac{1}{2}$, where the \pm depends on whether the quark i is of the u type or of the d type. The sum runs over all
 420 the n_f quarks that are active for the specific scale at which the scattering takes place. From Eqns. (33) and (34) we
 421 see that the main limitation of neutral current structure functions is that they provide limited access to quark flavour
 422 separation and in particular they cannot separate quarks from antiquarks, unless one goes to very high Q^2 values where
 423 the suppression induced by the Z boson propagator can be ignored.

424 In the case of charged-current DIS, the corresponding expressions for the structure functions in the parton model,
 425 assuming that we are above the charm threshold but below the top quark threshold, and the CKM suppressed transi-
 426 tions can be neglected, are given by

$$\begin{aligned} F_2^{W^-} &= 2x(u + \bar{d} + \bar{s} + c) , \\ F_3^{W^-} &= 2x(u - \bar{d} - \bar{s} + c) , \\ F_2^{W^+} &= 2x(d + \bar{u} + \bar{c} + s) , \\ F_3^{W^+} &= 2x(d - \bar{u} - \bar{c} + s) , \end{aligned} \quad (35)$$

427 and where again the longitudinal structure function $F_L^{W^\pm} = 0$ vanishes in this model. By comparing the NC and
 428 CC expressions, we find that the main difference them is that in the latter the F_3^W structure function, which provides
 429 information on the difference between quark flavours, is not suppressed with respect to F_2^W . For this reason, including
 430 DIS charged current structure functions in global fits is often motivated to improve the discrimination between quarks
 431 and antiquarks, which is not possible from neutral current structure functions.

432 These quark parton model expressions are also valid at leading order in perturbative QCD, once the effects of the
 433 DGLAP evolution are accounted for as described in Sect. 2.4. Starting at NLO, the contribution from the gluon PDFs
 434 must also be included. It is thus clear that inclusive DIS structure functions will be weakly sensitive to $g(x, Q^2)$, since
 435 only information via scaling violations or via the small $\mathcal{O}(\alpha_s)$ effects in the coefficient functions is available. The
 436 exception is the longitudinal structure function F_L , that while vanishing at LO, starting from NLO it is non-zero and
 437 has direct sensitivity to the gluon PDF. Indeed, it can be shown that this structure function can be written

$$F_L(x, Q^2) = \frac{\alpha_s(Q^2)}{\pi} \left[\frac{4}{3} \int_0^1 \frac{dy}{y} \left(\frac{x}{y}\right)^2 F_2(y, Q^2) + 2 \sum_i e_i^2 \int_x^1 \frac{dy}{y} \left(\frac{x}{y}\right)^2 (1 - x/y) g(x, Q^2) \right], \quad (36)$$

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

440 *Experimental data*

441 Since the pioneering DIS experiments at SLAC in the late 60s and early 70s that demonstrated the existence of
 442 quarks, there have been many measurements of deep-inelastic scattering structure functions. These measurements
 443 have been performed using either electrons, positrons or muons as projectiles scattering off protons, deuterons and
 444 neutrons, either for fixed-target or for collider kinematics. We now discuss this various measurements in turn.

445 To begin with, the fixed-target DIS measurements available for PDF fits can be divided into neutral current and
 446 charged current datasets. The former include the measurements of the proton and deuteron structure functions by the
 447 BCDMS collaboration [100, 101] using muons as projectiles; the measurements of the proton and deuteron structure
 448 functions by the NMC collaboration [102, 103], as well as of the ratio between deuteron to proton structure functions,
 449 F_2^d/F_2^p ; and the SLAC measurements of the proton and deuteron NC structure functions [104], among others. In this
 450 category we should also mention the measurements of the proton, neutron and deuteron structure functions at large- x
 451 and small- Q performed by JLAB experiments such as CLAS [105]. These JLAB measurements are excluded from
 452 most PDF fits since only a handful of points survive the typical DIS kinematical cuts in x and Q , but are included in the
 453 CJ fits. Moreover, there exist also rather old measurements of structure functions from the EMC collaboration [106].
 454 Despite their age, the EMC measurements of F_2^c have never been repeated and thus provide unique information on the
 455 charm content of the proton at large- x .

456 Concerning the fixed-target charged current DIS measurements, these can be divided into inclusive structure func-
 457 tion measurements and into charm production in neutrino-induced DIS, the so-called dimuon process (since the charm
 458 quark hadronizes into a D meson which then decays semi-leptonically), schematically represented in Fig. 5. Both
 459 types of measurements provide useful information on quark flavour separation, with the dimuon process specifically
 460 allows accessing the strange content of the nucleon. In the category of charged-current neutrino-induced inclusive
 461 DIS cross-sections we have the CDHSW, CCFR [107, 108] and CHORUS [109] and NuTeV [110] measurements
 462 on nuclear targets. On the other hand, charm production in neutrino scattering, the dimuon process, has been mea-
 463 sured by the CCFR and NuTeV [111, 112] and and CHORUS [113] collaborations on the same nuclear targets as the
 464 corresponding inclusive measurements.

465 Concerning DIS measurements from the HERA lepton-proton collider, the most updated inclusive measurements
 466 are the NC and CC differential cross-sections using electron and positron projectiles from the combination of the
 467 Run I and Run II data-taking periods [21]. This HERA legacy combination of DIS inclusive structure functions
 468 supersedes all previous inclusive measurements from H1 and ZEUS, including the Run I combined dataset [85] as
 469 well as the separate measurements by the two experiments from Run II [114, 115, 116, 117]. The impact of replacing
 470 these individual datasets by the final HERA combination of inclusive structure functions has been studied by different
 471 groups [118, 119, 120], finding that the impact of this replacement is quite moderate in general.

472 In addition, we also have HERA measurements of charm and bottom structure functions. As indicated in Fig. 5,
 473 heavy quark structure functions offer direct information on the gluon PDF, as well as on the treatment of heavy quark
 474 mass effects in the theoretical calculation. Charm structure functions in addition are an important ingredient for the
 475 determination of the charm mass m_c together with the PDFs. In particular, the latest heavy flavour measurements from
 476 HERA include the combined NC cross-sections of charm production in DIS, $\tilde{\sigma}_c$ [121] and the H1 and ZEUS data on
 477 the bottom structure function $F_2^b(x, Q^2)$ [122, 123]. While F_2^b is known to have a small impact in the global fit, it is
 478 relevant for specific applications, for instance the determination of the bottom quark mass m_b from the PDF fit.

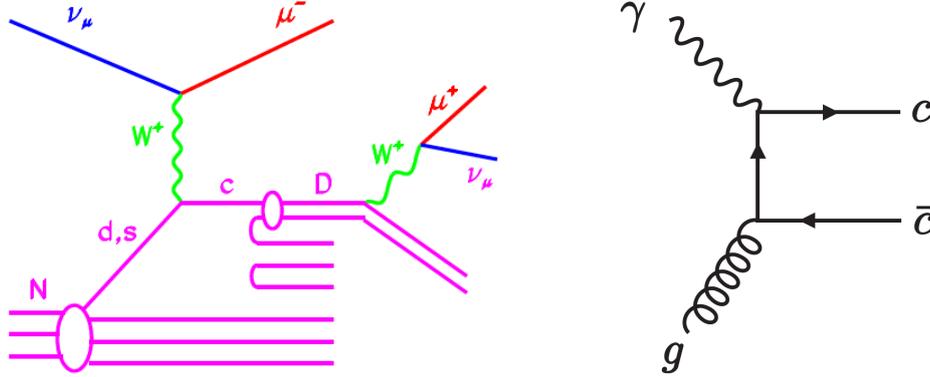


Figure 5: Left plot: D meson production in charged-current neutrino-induced DIS. This is known as the dimuon process since events are tagged where the D meson decays semi-leptonically, with the pair of oppositely charged muons providing a clean signature. Right plot: Charm production in neutral current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

479 We also note that previous measurements of the longitudinal structure function F_L by the H1 and ZEUS collabo-
 480 rations [115] are now superseded by the final inclusive HERA combination.

481 *Theoretical calculations and tools*

482 From the theoretical point of view, the coefficient functions of DIS structure functions in the neutral current case
 483 they are available up to $\mathcal{O}(\alpha_s^3)$ in the massless limit and up to $\mathcal{O}(\alpha_s^2)$ taking into account heavy quark mass effects,
 484 though there has been considerable recent progress towards the completion of the $\mathcal{O}(\alpha_s^3)$ calculation of massive DIS
 485 structure functions. For charged current structure functions, massless coefficients are available up to $\mathcal{O}(\alpha_s^3)$ and
 486 massive coefficient functions up to $\mathcal{O}(\alpha_s^2)$ [124].

487 *3.3. Inclusive jets*

488 **Jun**

489 Inclusive jet production at hadron colliders have been providing the dominant constraints on gluon PDF at large- x
 490 since the run of Fermilab Tevatron. The definition of jet cross sections starts from a well defined jet algorithm. It
 491 is usually chosen to be infrared and collinear safe thus can be calculated in perturbative QCD for hard scattering
 492 at high energies. The most used jet algorithm at the LHC is anti- k_T jet algorithm [125] provided with the 4-vector
 493 recombination scheme. Other common choices include k_T algorithm [126, 127], Cambridge-Aachen algorithm [128],
 494 and also the Midpoint algorithm [129] once used at Tevatron. There exist non-perturbative corrections when match-
 495 ing parton-level cross sections to those directly measured at hadron-level, e.g., due to underlying events and QCD
 496 hadronizations. Ususally the non-perturbative corrections are provided as multiplicative factors derived from leading-
 497 order event generators. The non-perturbative corrections can be significant at low- p_T , reaching as large as 20%, and
 498 small at high- p_T , at a level of one percent [130]. Variations of these non-perturbative corrections can be treated as a
 499 correlated systematic error. The experimental data can thus be compared with various parton-level predictions multi-
 500 plied by the correction factors. There are also predictions on matching the next-to-leading order calculations with the
 501 parton showering and hadronization [131] which can be directly compared with the data at hadron-level.

502 *PDF sensitivity*

At LO the jet production at hadron colliders includes the following subprocesses

$$\begin{aligned}
 &gg \rightarrow gg, \quad gg \rightarrow q\bar{q}, \quad gq \rightarrow gq, \quad q\bar{q} \rightarrow gg, \\
 &q\bar{q} \rightarrow q\bar{q}, \quad q\bar{q} \rightarrow q'\bar{q}', \quad q\bar{q}' \rightarrow q\bar{q}', \quad qq \rightarrow qq, \quad qq' \rightarrow qq',
 \end{aligned}
 \tag{37}$$

503 and the charge conjugate processes. Kinematics of the two leading jets in the final state can be characterized by
 504 their rapidities y_1, y_2 and transverse momenta $p_{T,1}, p_{T,2}$, which are balanced at LO and equal p_T . The momentum
 505 fractions carried by the two incoming partons can be reconstructed as

$$x_1 = \frac{p_T}{\sqrt{s}}(e^{y_1} + e^{y_2}), \quad x_2 = \frac{p_T}{\sqrt{s}}(e^{-y_1} + e^{-y_2}), \quad (38)$$

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

$$x_1 x_2 = \frac{4p_T^2 \cosh^2 y^*}{s}, \quad x_1/x_2 = e^{2y_b}. \quad (39)$$

508 Beyond LO there can be multiple jets in the final state from additional QCD radiations. Thus jet production can probe
 509 the gluon PDF directly at LO in QCD and quark and anti-quark PDFs universally.

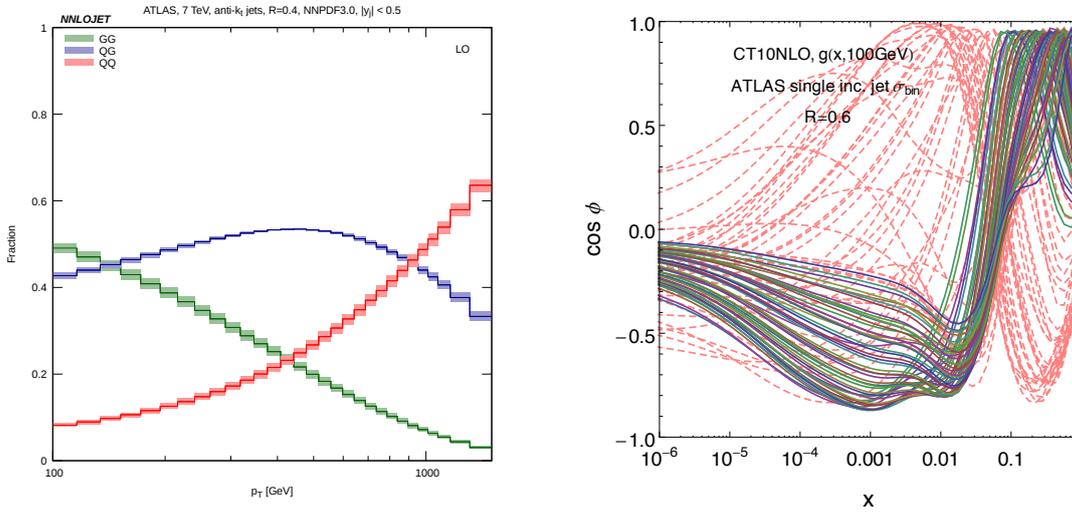


Figure 6: Left: Fractional contributions from different partonic channels to the single inclusive jet production at the LHC 7 TeV at LO in the central rapidity region [132]. Right: Correlations between binning cross sections from ATLAS on the single inclusive jet production at the LHC 7 TeV and the gluon PDF; dashed curves correspond to experiment bins at low p_T .

510 Jet production can be measured in various ways experimentally. The most commonly used jet cross sections for
 511 constraining PDFs are the double differential distribution in jet p_T and rapidity for single inclusive jet production where
 512 one count all jets in a single event and fill them into the same distribution. The double differential cross section can
 513 be sensitive to different flavor combinations. Left plot in Figure 6 shows the fractional contributions from different
 514 subprocesses to the single inclusive jet production in central rapidity region at the LHC. Subchannels involving gluons
 515 are dominant at low- p_T and still significant at high- p_T . The major PDF variations of the jet cross sections arise from
 516 the gluon PDF since the quark PDFs are better constrained with the DIS data. That is illustrated in Figure 6 on the PDF
 517 correlations between the single inclusive jet cross sections (using the ATLAS binning [133], each curve corresponds
 518 to one experimental bin) and the gluon PDF at various x values. Figure 6 shows that the single inclusive jet production
 519 can further constrain the gluon PDF in a wide range of x , $10^{-3} \sim 1$. There are also measurements on double differential
 520 cross sections of inclusive dijet production, e.g., with respect to y^* and invariant mass of the two leading jets, or even
 521 triple differential cross sections, e.g., with respect to y_b, y^* , and average p_T of the two leading jets. Through such
 522 refined binnings one can probe different initial states more efficiently. Those bins with large y_b usually receive more
 523 contributions from gluon initial states. On another hand bins with large y^* and large $p_{T,ave}$ are dominated by initial
 524 states with two valence quarks and can potentially further constrain the d -valence PDF at high- x .

525 *Experimental data*

526 In this section we summarize available measurements on jet production at hadron colliders that are relevant for
 527 constraining PDFs from Tevatron Run II to LHC Run II.

- 528 • The measurements on double differential cross sections in single inclusive jet production from CDF collabora-
 529 tion [134, 135] and D0 collaboration [136, 137] at Tevatron Run II (1.96 TeV).
- 530 • The measurements on double differential cross sections in single inclusive jet production from ATLAS collab-
 531 oration [138, 133, 139, 140] and CMS collaboration [141, 142, 130] at LHC Run I (7 and 8 TeV).
- 532 • The measurements on double differential cross sections in inclusive dijet production from ATLAS collabora-
 533 tion [138, 133, 143] and CMS collaboration [144, 141] at LHC Run I (7 and 8 TeV).
- 534 • The measurement on triple differential cross sections in inclusive dijet production from CMS collaboration [?]]
 535 at LHC Run I (8 TeV).
- 536 • The measurements on ratio of double differential cross sections in single inclusive jet production at different
 537 center of mass energies, 2.76, 7 and 8 TeV, from ATLAS collaboration [139] and CMS collaboration [130] at
 538 LHC Run I.
- 539 • The measurement on double differential cross sections in single inclusive jet production from CMS collabora-
 540 tion [145] at LHC Run I (13 TeV).

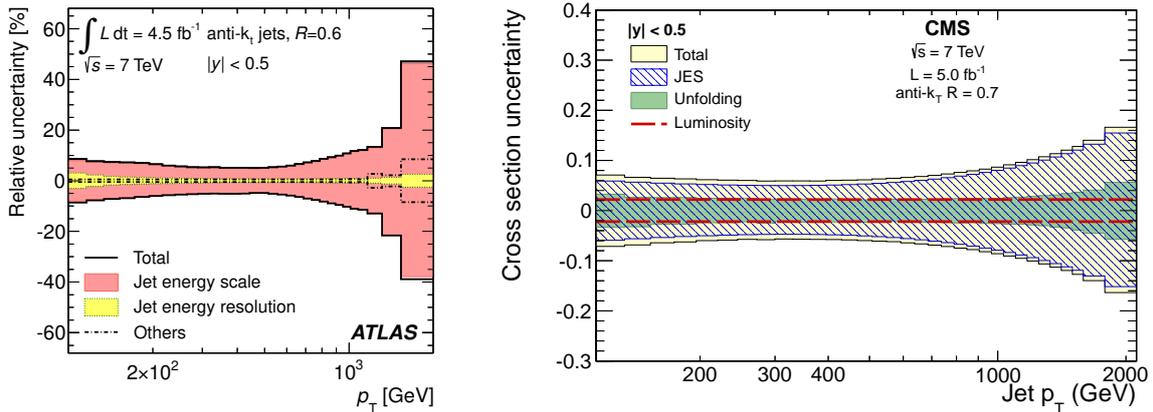


Figure 7: Representative systematic errors in single inclusive jet measurement at LHC 7 TeV in the central rapidity region, from ATLAS (left) [140] and CMS (right) [142]. The luminosity error is not included in the ATLAS plot.

541 Due to the complexity of jet reconstruction and calibration there exist a large amount of experimental system-
 542 atic uncertainties for the cross sections measurement, e.g., about 50 ~ 100 correlated systematic errors for a total
 543 number of data points of 100 ~ 200, for both ATLAS and CMS. In the most recent ATLAS and CMS 7 TeV mea-
 544 surements [140, 142] the total correlated experimental uncertainties are at a level of about 5 ~ 20% in most region.
 545 For the same measurements the uncorrelated systematic errors and statistical errors are at one percent level or less
 546 in general. Figure 7 shows the typical experimental systematics from both ATLAS and CMS which are dominated by
 547 the jet energy scale [140, 142]. The precision of the LHC data, especially with the small uncorrelated errors, makes
 548 it rather challenging to fit the jet data well, e.g., for the ATLAS 7 TeV measurement. It requires better understanding
 549 on the experimental systematics and also possible theoretical errors. Besides, there are measurements on ratios of
 550 the double differential cross sections between single inclusive jet production at different center of mass energies, e.g.,
 551 CMS 2.76, 7 and 8 TeV [130], ATLAS 2.76 and 7 TeV [139], for which the experimental systematic errors are largely
 552 cancelled though the statistical errors are slightly larger. There are also measurements on kinematic distributions in

553 inclusive dijet production, including double differential cross sections or even triple differential cross sections in one
 554 of recent CMS analysis [146]. In the last case the cross sections are measured in three dimensional bins of the total
 555 boost, the rapidity difference and average p_T of the two leading jets, which can provide more refined separations on
 556 sensitivities to different flavors and different Bjorken- x .

557 *Theoretical calculations and tools*

558 The NLO QCD corrections to single inclusive jet and inclusive dijet production was first calculated in Refs. [147,
 559 148] in early 90's. The NLO calculation has been implemented in two numerical programs NLOjet++ [149, 150]
 560 and MEKS [151]. Recently the NNLO QCD corrections to the same process have been completed for all partonic
 561 channels [152, 153, 154] except missing for some sub-leading contributions in QCD colors. The calculation is based
 562 on the Antenna subtraction method [155, 156] for isolating the infrared singularities in QCD real radiations. Left
 563 plot in Figure 8 summarizes the NNLO QCD corrections to the single inclusive jet production at LHC 7 TeV with
 564 anti- k_T algorithm and a central scale choice of the leading jet p_T . The NNLO QCD corrections are significant at
 565 low- p_T leading to a 10% increase of the cross sections while the NLO corrections there are negative. At high- p_T
 566 region the NNLO corrections are small and stabilize the theoretical predictions. The NLO scale variations bands
 567 are asymmetric at low- p_T and largely underestimate the perturbative uncertainties. The EW corrections can be significant
 568 at high- p_T for central rapidities due to the large EW Sudakov logarithms, and are well below 1% for rapidity greater
 569 than 1 [157]. There are ambiguities on choosing the *correct* QCD scales even for single inclusive jet production, i.e.,
 570 either using p_T of the individual jet or the leading jet. They are the same at LO but can be very different when there
 571 exist more jets and they have large differences in p_T . Ref. [132] further compares the NNLO predictions using these
 572 two choices for the central scales. Right plot in Figure 8 shows the two NNLO predictions converge at high- p_T as
 573 expected. However, at low and intermediate p_T , there are significant differences of the central values comparing to
 574 the size of scale variations. The two error bands even do not overlap. It is observed that the NNLO predictions using
 575 individual p_T as the central scale tend to follow the trends of ATLAS data better. Further investigations are needed to
 resolve the ambiguity of scale choice in the NNLO predictions.

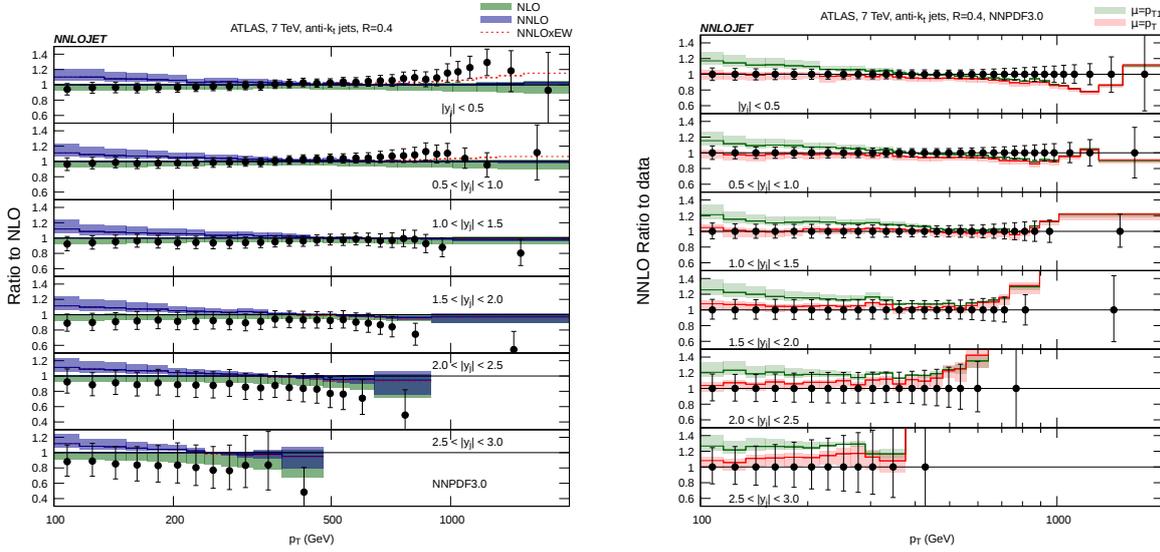


Figure 8: Left: Predictions on single inclusive jet production at the LHC 7 TeV using ATLAS binning and anti- k_T algorithm with $R = 0.4$, with a central scale choice of leading jet p_T and scale variations by varying renormalization and factorization scales simultaneously by a factor of 2 [132]. Right: 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) [132].

576 Besides, there are theoretical calculations including analytic QCD resummation [158, 159, 160, 161, 162]. It has
 577 been shown in Ref. [161] that approximate NNLO predictions from expansion of threshold resummation agree well
 578

579 with the exact NNLO predictions for all-gluon channel at large p_T of single inclusive jet production at the LHC. For
580 full rapidity range the threshold expansions reproduce well fixed-order results down to p_T of about 400 GeV with
581 the same value shifted to lower p_T for large rapidity region. The approximate NNLO predictions have been used in
582 previous global analysis involving jet data [19, 17]. The jet cross sections are also sensitive to the jet algorithm used
583 especially on the value of cone size or the distance parameter. A larger cone size usually leads to more inclusive cross
584 sections and better convergence in perturbative expansion. That also reduces the non-perturbative corrections from
585 QCD hadronization but increases corrections from underlying events. At the LHC in the nominal jet measurement,
586 ATLAS uses distance parameters of 0.4 and 0.6, and CMS uses 0.5 and 0.7 instead.

587 *Impact on PDFs*

588 The jet data from Tevatron and LHC Run I have already set important role in global analyses [163, 17, 18, 19]
589 though the NNLO analyses only apply NLO or approximate NNLO theoretical predictions currently. It was found
590 that removing all jet data from the global analyses can lead to an increase of the gluon PDF uncertainties at large- x
591 by at least a factor of two [17]. There are also independent studies from CMS [164, 130?] on effects of the jet data
592 on constraining PDFs based on NLO fits using the xFitter program [27]. Figure 9 from Ref. [?] shows impact of the
593 CMS 8 TeV jet data on the gluon PDF by adding the data into a base fit with HERA DIS data only [21]. Inclusion of
594 both the single inclusive jet data and the inclusive dijet data lead to a great reduction of the gluon PDF uncertainties
595 at large x . Meanwhile in the same fit one also observe reduction of the PDF uncertainties of the valence quark at x close
596 to one which can be complementary to the constraint from Drell-Yan and fixed-target DIS data. With the full NNLO
597 predictions on jet production available we are now more confident on pinning down the gluon PDF at large x in the
598 global analyses with the more precise inclusive jet data from LHC Run I and Run II.

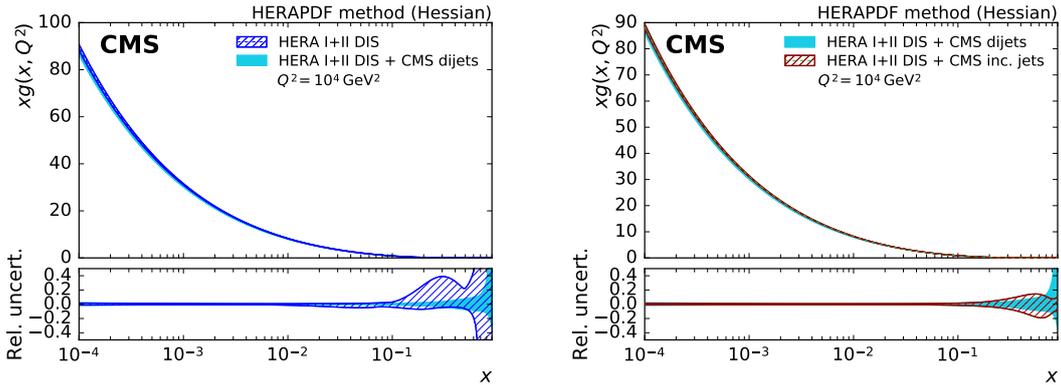


Figure 9: Effects of the CMS 8 TeV jet data on the gluon PDF when adding up to a PDF fit with HERA DIS data only. The left plot compares the fit with only DIS data and the fit with CMS dijet data in addition. The right plot compares the fit to HERA DIS plus CMS dijet data and fit to HERA DIS plus CMS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

599 *3.4. Inclusive gauge boson production*

600 Now we turn to discuss the inclusive production of electroweak gauge boson. This process has been of enormous
601 historical importance since it provided a first window on the quark flavour separation in the proton beyond the in-
602 formation contained on DIS structure functions. Nowadays, Drell-Yan cross-sections provide the backbone of global
603 PDF fits together with the fixed-target and HERA structure function data.

604 *PDF sensitivity*

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

$$605 \quad \bar{u}\bar{d}, \bar{c}\bar{s} \quad (u\bar{s}, c\bar{d}) \rightarrow W^+, \quad (40)$$

$$606 \quad \bar{d}\bar{u}, \bar{s}\bar{c} \quad (s\bar{u}, d\bar{c}) \rightarrow W^-, \quad (41)$$

$$607 \quad q\bar{q} \rightarrow Z/\gamma^*, \quad (42)$$

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

$$R_{\pm} = \frac{d\sigma(W^+)/dy_W}{d\sigma(W^-)/dy_W} = \frac{u(x_1)\bar{d}(x_2) + c(x_1)\bar{s}(x_2) + 1 \leftrightarrow 2}{d(x_1)\bar{u}(x_2) + s(x_1)\bar{c}(x_2) + 1 \leftrightarrow 2}. \quad (43)$$

609 and the W asymmetry

$$A_W = \frac{d\sigma(W^+)/dy_W - d\sigma(W^-)/dy_W}{d\sigma(W^+)/dy_W + d\sigma(W^-)/dy_W} = \frac{u(x_1)\bar{d}(x_2) + c(x_1)\bar{s}(x_2) - d(x_1)\bar{u}(x_2) - s(x_1)\bar{c}(x_2) + 1 \leftrightarrow 2}{u(x_1)\bar{d}(x_2) + c(x_1)\bar{s}(x_2) + d(x_1)\bar{u}(x_2) + s(x_1)\bar{c}(x_2) + 1 \leftrightarrow 2}. \quad (44)$$

610 We will for simplicity consider the W rapidity, rather than the experimentally observable rapidity of the charged lepton
 611 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 expressions 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

$$\text{Central :} \quad y_W \sim 0 \quad x_1 \sim x_2 = x_0, \quad \bar{u}(x_{1,2}) \sim \bar{d}(x_{1,2}), \quad (45)$$

$$\text{Forward :} \quad y_W \gtrsim 2, \quad x_1 \gg x_2, \quad q(x_1) \sim q_V(x_1), \quad \bar{u}(x_2) \sim \bar{d}(x_2), \quad (46)$$

612 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
 613 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
 614 interchange $x_1 \leftrightarrow x_2$.

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

$$R_{\pm} \sim \frac{u(x_0)}{d(x_0)}, \quad (47)$$

$$A_W \sim \frac{u_V(x_0) - d_V(x_0)}{u(x_0) + d(x_0)}. \quad (48)$$

615 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
 616 reasonably low x values, the valence u and d quarks are fairly small, and so we roughly expect $R_{\pm} \sim 1$ and $A_W \sim 0$,
 617 with the departures from these values being due to the precise flavour content of the proton, in particular the fact that
 618 the valence distributions are not completely negligible in this region.

In the forward region, applying the simplification of (46) and again dropping the c, s contributions we find

$$R_{\pm} \sim \frac{u_V(x_1)}{d_V(x_1)}, \quad (49)$$

$$A_W \sim \frac{u_V(x_1) - d_V(x_1)}{u_V(x_1) + d_V(x_1)}. \quad (50)$$

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

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

$$\frac{d\sigma(W^+)/dy_W + d\sigma(W^-)/dy_W}{d\sigma^Z/dy_Z} \approx \frac{u_V(x_1) + d_V(x_1)}{0.29u_V(x_1) + 0.37d_V(x_1)}, \quad (51)$$

621 where the factors in the denominator come from the electroweak Z -quark couplings. For the central region a similar
 622 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
 623 information about the u and d quarks.

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

$$\frac{d\sigma^Z}{dy_{\parallel}} \sim 0.29(u(x_0)\bar{u}(x_0) + c(x_0)\bar{c}(x_0)) + 0.37(d(x_0)\bar{d}(x_0) + s(x_0)\bar{s}(x_0) + b(x_0)\bar{b}(x_0)). \quad (52)$$

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

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

$$\frac{d\sigma^{\text{DY}}}{dy_{\parallel}} \sim \sum_i e_i^2 (q(x_1)\bar{q}(x_2) + q(x_2)\bar{q}(x_1)). \quad (53)$$

637 Thus in comparison to (52) a different combination of the quark and anti–quark PDFs is probed, due to the differing
 638 electromagnetic couplings. In particular, the relative $u\bar{u}$ to $d\bar{d}$ contribution is now a factor of ~ 5 higher in compar-
 639 ison to the Z cross section. At the LHC, low mass Drell–Yan production therefore provides complementary flavour
 640 information in the low to intermediate x region. In addition, as the cuts on the final–state lepton transverse momenta
 641 tend to increase the relative important of the higher order contributions, for which the Z p_{\perp} can be non–zero, this can
 642 be sensitive to the gluon PDF at lower x , which contributes through the NLO $g \rightarrow q\bar{q}$ splitting. High mass Drell–Yan
 643 production is sensitive to the q, \bar{q} PDFs at high x , in particular the anti–quarks, which are less well determined in this
 644 region.

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

$$u^p = d^n \quad d^p = u^n, \quad (54)$$

647 allowing an extra handle on the proton flavour decomposition. In particular, such fixed target experiments generally
 648 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
 649 term in (53) is dominant, with $q(x_1) \sim q_V(x_1)$. It is then straightforward to show that

$$\frac{\sigma^{pn}}{\sigma^{pp}} \sim \frac{\bar{d}(x_2)}{\bar{u}(x_2)}. \quad (55)$$

650 That is, they are sensitive to quark sea decomposition in the intermediate to high $x_2 \sim 0.01 - 0.3$ region probed by
 651 these fixed target experiments [165]. This however comes with the added complication that the nuclear corrections
 652 accounting for the fact that the neutron is bound in a deuteron nucleus, and therefore the ‘free’ neutron PDF is not
 653 directly probed. Fixed target pp scattering alone does not suffer from this issue, and is sensitive to the quark sea
 654 (dominantly, the \bar{u}) in the same x region, but is much less directly sensitive to the \bar{d}/\bar{u} decomposition.

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

$$q^p = \bar{q}^{\bar{p}}. \quad (56)$$

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

661 Finally, we have considered above the distributions with respect to the (unobservable) rapidity of the W boson to
 662 simplify the discussion. In general we should correctly account for the kinematics, as well as weight the corresponding
 663 $q\bar{q}$ contributions by the appropriate W decay distributions. This in fact provides a further handle on the flavour
 664 sensitivity of this observable, as by changing the p_{\perp} cut on the charged lepton, different weights of the different quark
 665 contributions are achieved, see e.g. [80, 166] for further details. Nonetheless, the forward and central W rapidity
 666 regions are certainly correlated with the equivalent lepton rapidity regions that are measured experimentally, and so
 667 the above discussion provides a qualitative guide for the PDF sensitivity of W boson production. However, as we will
 668 discuss below, the current simulation codes for W and Z production include the full kinematics of the leptonic decays,
 669 and therefore there is no need to explicitly correct back to the W rapidity.

670 *Experimental data*

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

- 672 • The most precise fixed target Drell–Yan data come from the E866/NuSea [167] experiment at Fermilab, while
 673 the E906/SeaQuest experiment [168] will extend out to higher x , and is currently taking data.
- 674 • The Tevatron collider has produced a range of data on W and Z production, including measurements of the Z
 675 rapidity distribution [169, 170] and in W production both the lepton [171, 172] and the W [173, 174] asymme-
 676 tries.
- 677 • Early LHC measurements of the the Z rapidity distribution presented by CMS [175] and ATLAS [176].
- 678 • CMS Drell–Yan data at 7 TeV [177], for $15 < M_{ll} < 1500$ GeV and at 8 TeV [178], which increased the upper
 679 mass limit to 2000 GeV. These are presented double differentially in the rapidity and invariant mass of the lepton
 680 pair.
- 681 • ATLAS 7 TeV Drell–Yan invariant mass distribution (integrated over rapidity) at high [179] ($116 < M_{ll} < 1500$
 682 GeV) and low [180] ($26 < M_{ll} < 66$) invariant masses.
- 683 • CMS [181, 182] 7 TeV W asymmetry, and ATLAS [176] W^+ and W^- cross section data.
- 684 • ATLAS high precision W and Z, γ^* data [183], using the full 4.6 fb^{-1} data set at 7 TeV. The Drell–Yan rapidity
 685 distribution is presented double differentially in three intervals of lepton pair mass, over the $45 < M_{ll} < 150$
 686 GeV range. In the Z peak and higher mass regions the measurement was also extended out to $|y_{ll}| = 3.6$.
- 687 • LHCb Z rapidity distributions at 7 [184, 185], 8 [186] and 13 [187] TeV.
- 688 • LHCb Lepton rapidity distributions for W^+ and W^- production at 7 [188], and 8 [189, 190] TeV.
- 689 • In [189] cross section ratios between the 7 and 8 TeV W and Z measurements are presented, with the cancellation
 690 in various systematic uncertainties providing a more precise PDF sensitivity.

691 Thus at the LHC multiple measurements have been presented. The Z/γ^* data are available over a wide range
 692 of invariant masses, providing extensive coverage in x . The W data are increasingly presented as individual cross
 693 sections, including the correlated error information, to provide the maximum possible constraints. While in the
 694 majority of cases, the ATLAS and CMS measurements are limited to the central rapidity region, that is a lepton
 695 pseudorapidity of $|\eta_l| < 2.4$, this reach is extended by exploiting the forward acceptance of the LHCb detector, for
 696 which $2 < \eta_l < 4.5$ is accessed. This allows the high and low x region to be probed. The most recent ATLAS W
 697 and Z, γ^* data [183], which uses the full 4.6 fb^{-1} data set at 7 TeV demonstrates the level of precision that is now
 698 being achieved. The Z rapidity distribution and W asymmetry are shown in Fig. 10 (in the latter case the individual
 699 W measurements are available). The high experimental precision is clear, in particular in the Z distribution where
 700 excluding the luminosity uncertainty it is as low as $\sim 0.3\%$ at central rapidity, while the error on the PDF predictions,
 701 as well as the spread between sets, is significantly larger. The impact of such data can therefore be sizeable.

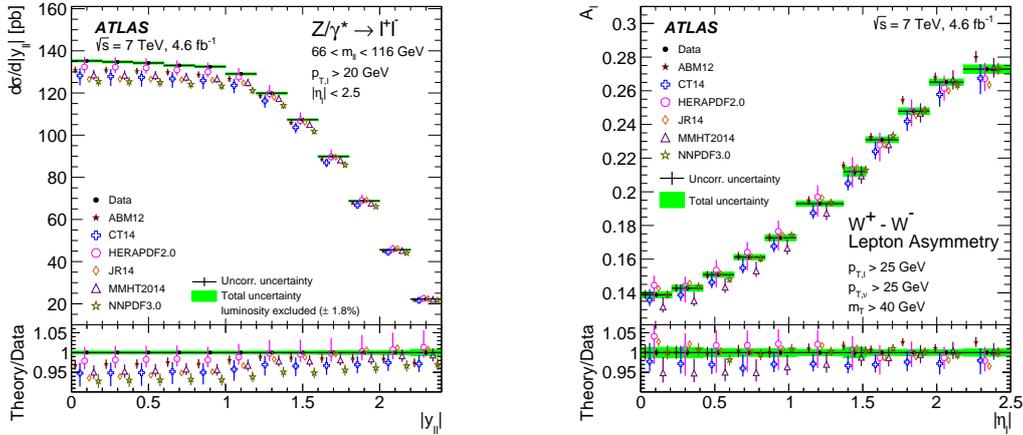


Figure 10: ATLAS 7 TeV measurement of Z rapidity distribution (left) and W asymmetry (right), taken from [183].

702 *Theoretical calculations and tools*

703 W and Z boson production is arguably the simplest process one can consider at a hadron collider, and indeed it
 704 was the first hadroproduction process for which the NNLO calculation became available, with the total cross sections
 705 being calculated in the early 90s [93]³. A decade later, in [192, 193] the NNLO corrections to the differential W and
 706 Z rapidity distributions was presented for the first time. A more direct comparison with experimental observables was
 707 provided in [194, 195] which presented the NNLO calculation fully differential in the final-state leptons, including in
 708 addition spin correlations, finite width effects and $\gamma - Z$ interference. This was accompanied by the public release of the
 709 FEWZ simulation code, with subsequent improvements reported in [196, 197] and [198], where NLO EW corrections
 710 (first calculated in [199, 200, 201, 202, 203, 204]) were included. The DYNLLO [205] parton level MC provides an
 711 alternative tool for generating W and Z production, again including spin correlations, finite width effects, and $\gamma - Z$
 712 interference, but currently without EW corrections. This code allows for arbitrary user-defined cuts on the final-state
 713 partons and leptons to be imposed and histograms to be made, in contrast to FEWZ, where a selection of pre-determined
 714 cuts and histograms may be applied.

715 These two codes differ in their theoretical treatment of the processes, in particular in the method that is applied
 716 to achieve the (non-trivial) cancellation of IR singularities at intermediate steps in the calculation. While FEWZ uses
 717 the local ‘sector decomposition’ method [194, 195] that provides an automated method for extracting and cancelling
 718 the IR poles, DYNLLO applies an alternative non-local ‘ q_T -subtraction’ approach [206] which uses the transverse
 719 momentum q_T of the produced W or Z as a cut variable, treating the calculation in a different way above and below
 720 some q_T^{cut} . It is unfortunately now quite well established that these codes can give non-negligible differences in their
 721 predictions for identical input parameters. For example, in the recent ATLAS high precision W and Z/γ^* analysis [189]
 722 the difference in the fiducial cross section predictions can be as high as $\sim 1\%$, that is larger than the experimental
 723 uncertainties. This is due to the differing subtraction procedures, which affects the predicted boson p_{\perp} distributions.
 724 The differences between the predicted cross sections are generally more significant when more restrictive cuts on the
 725 final-state leptons are imposed; for the total W , Z cross sections the codes agree to within 0.2% [189]. A closer
 726 investigation of this issue and its impact on PDF determination will clearly be essential.

727 More recently the MCFM event generator [207] has extended the NLO simulation of W and Z production to
 728 NNLO [208]. This takes a similar non-local approach to DYNLLO, but using the N -jettiness variable rather than
 729 the q_T . Here, it is shown that a careful and process-dependent choice of the cut on the 0-jettiness variable, τ_0^{cut} (the
 730 equivalent of q_T^{cut} above) is required in order to balance the requirements of sufficient statistical precision and control
 731 over systematic power corrections that increase in importance as this cut is increased.

732 Finally, event generators including transverse momentum resummation are also available. The DYRes [209] code

³An error in the one-loop real emission contribution was reported in [191].

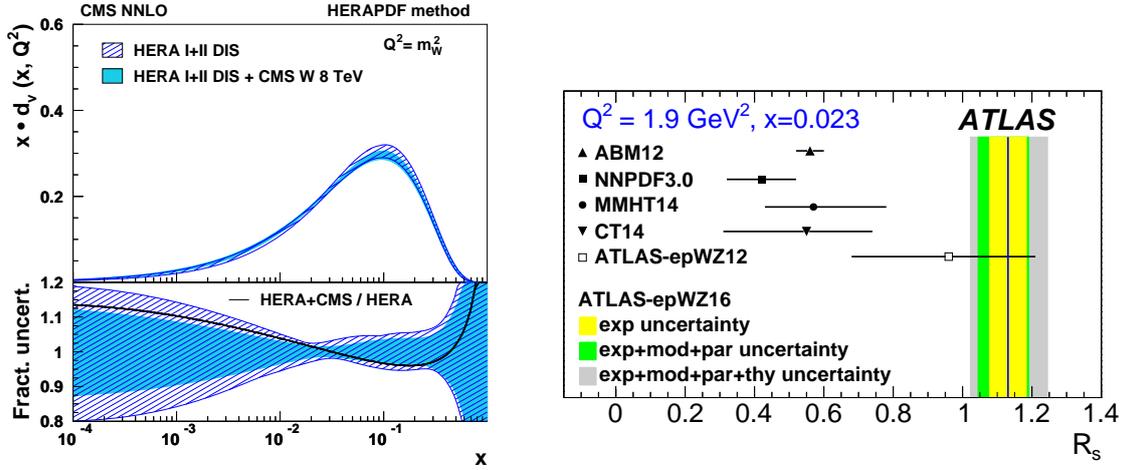


Figure 11: (Left) Down valence distribution, with the CMS fit to HERA I+II data only, and including the CMS 8 TeV W production data. Taken from [211]. (Right) The ratio of the strange quark to the light quark sea, R_s , with different PDF predictions and the result of the ATLAS fit to HERA I+II data and the high precision W and Z/ γ^* data shown. Taken from [183].

733 combines NNLO fixed-order with NNLL resummation, while ResBos [210] combines NLO fixed-order with NNLL
 734 resummation. However, typical observables that are used in PDF fits are chosen to be largely insensitive to such
 735 resummation effects, which are most important as the W, Z transverse momentum becomes small, and so these codes
 736 are not currently used in PDF analyses. **possibly link to resummation section**

737 Impact on PDFs

738 As described in Section 3.4 inclusive gauge boson production has played a crucial role in determining the quark
 739 flavour decomposition of the proton. Indeed, these have been included in all major PDF analyses for some time, from
 740 earlier fixed target data through to measurements at the Tevatron and increasingly at the LHC. Two recent LHC results
 741 are show in Fig. 11. In the left panel we show the CMS fit [211] to the down valence quark distribution. The baseline
 742 fit is to the HERA I+II data only, which is compared to the result including the CMS 8 TeV W boson production data.
 743 The change in shape and sizeable reduction in the PDF uncertainty over a wide range of x is clear.

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

$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}}, \quad (57)$$

745 of the ATLAS high precision W and Z/ γ^* data [183]. As described in Section 3.4, provided the light quark flavours are
 746 sufficiently well determined, and the data are sufficiently precise, the size and shape of the W, Z rapidity distributions
 747 can provide constraints on the strange quark PDFs. This is clear from the figure, where the fit to the ATLAS data
 748 predict a significantly higher value of R_s in comparison to previous PDF fits, which do not include the ATLAS data.
 749 A hint of this effect is seen in the earlier ‘ATLAS-epWZ12’ result [212], but it is only with the more recent high
 750 precision data that a clear effect becomes apparent.

751 3.5. The p_T of Z bosons

752 **Jun**

753 The running of LHC allows precision measurements of inclusive transverse momentum spectra of the Z boson
 754 produced in hadronic collisions. There are three distinct regions for the full p_T spectra. At small $p_T \ll m_Z$ (mass
 755 of the Z boson), the fixed-order predictions diverge due to the soft gluon radiations. QCD resummations are needed
 756 to maintain reliable predictions which have received a lot of attentions [213, 214, 215, 216, 217, 218, 219, 220].
 757 The distributions at small p_T thus are sensitive to additional non-perturbative input [221, 217] and are inadequate

758 for extraction of the collinear PDFs. At large $p_T \gg m_Z$, the fixed-order predictions can also receive large logarithm
759 contributions due to soft gluon radiations at partonic threshold of the Z boson and the recoiling jet [222, 223]. It
760 was found that those contributions can bring an additional 5% increase to the accumulated cross sections with $p_T >$
761 200 GeV comparing to the NLO predictions at the LHC [222]. For intermediate $p_T \sim m_Z$, the fixed-order predictions
762 should be reliable and the distribution can provide additional constraint notably for the gluon PDF.

763 PDF sensitivity

764 At LO the Z boson production with finite transverse momentum at hadron colliders includes the following sub-
765 processes

$$q\bar{q} \rightarrow Zg, \quad gq \rightarrow Zq, \quad g\bar{q} \rightarrow Z\bar{q}. \quad (58)$$

766 Kinematics of the Z boson, namely the transverse momentum p_T and rapidity y_Z , can be reconstructed from momen-
767 tums of the decayed charged leptons. In the case of inclusive with respect to the hadronic recoils the momentum
768 fractions of the initial state partons are not uniquely determined

$$x_1 = \frac{m_T}{\sqrt{s}} e^{y_Z} + \frac{p_T}{\sqrt{s}} e^{y_j}, \quad x_2 = \frac{m_T}{\sqrt{s}} e^{-y_Z} + \frac{p_T}{\sqrt{s}} e^{-y_j}, \quad (59)$$

769 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
770 the rapidity of the recoiling parton to be integrated over. At LO lower limits on x_1 and x_2 probed can be derived from
771 above equation. Usually experiments measure the double differential cross sections in p_T and y_Z around the Z boson
772 mass. The invariant mass window of the charged leptons can be also shifted to off-shell region where the contributions
773 from virtual photon can be important.

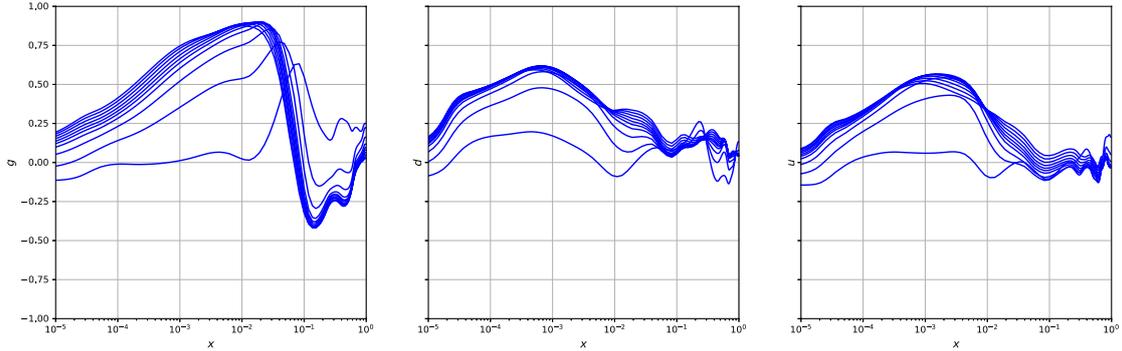


Figure 12: Correlations between the cross sections in various p_T bins and the gluon, down- and up-quark PDFs as a function of x [?]. The binning is chosen as in the ATLAS measurement [224] with rapidity interval $0 < |y_Z| < 0.4$.

774 The cross sections at moderate and large transverse momentum are dominated by contributions from the gluon and
775 quark scattering and are strongly correlated with the gluon PDF in the region relevant for Higgs boson production at
776 the LHC. That is further illustrated in Figure 15 which shows the PDF induced correlations between the binning cross
777 sections in p_T with rapidity interval $0 < |y_Z| < 0.4$ and the gluon, down- and up-quark PDFs at various x values [?].
778 The correlations to gluon PDFs at $x \sim 10^{-2}$ almost reach 0.9 as expected. Moderate correlations with quark PDFs at
779 $x \sim 10^{-3}$ are also observed.

780 Experimental data

781 The experimental measurements on p_T spectra of the Z boson have reached a precision of percent or subpercent
782 as from both ATLAS [224] and CMS [225] at LHC Run I due to the cleanest of the dilepton final state as well as
783 the high statistics of the signal. In addition the ATLAS measurement extracts the cross sections at 3 different particle
784 levels concerning the final-state photon radiations. The Born and bare levels are defined from the lepton kinematics

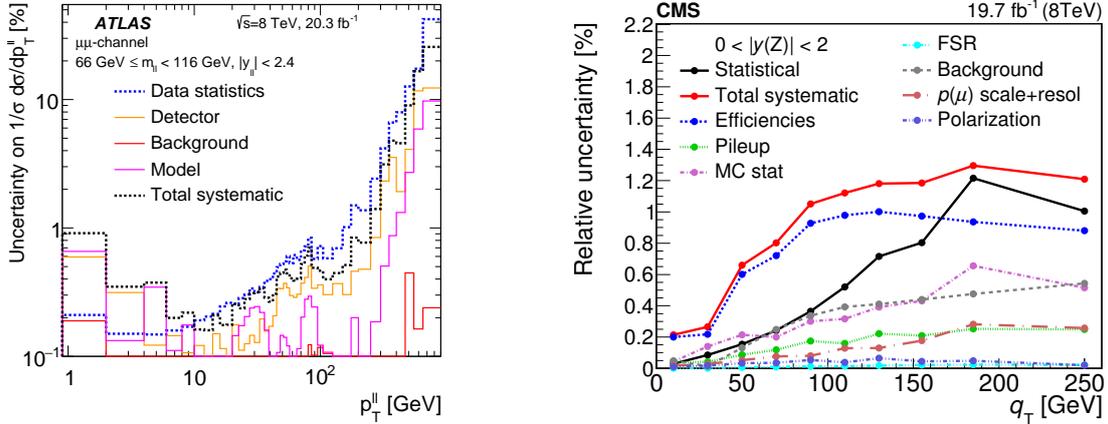


Figure 13: Left plot: relative experimental errors on the measured normalized p_T spectra of the Z boson in dimuon channel (dressed) from ATLAS 8 TeV [224]. Right plot: relative experimental errors on the measured normalized p_T spectra of the Z boson in dimuon channel from CMS 8 TeV [225].

785 before and after final-state radiations. The dressed level is defined by further combining the momentum of the lepton
 786 with photons radiated within a certain cone. We summarize the available measurements on p_T spectra of the Z boson
 787 from recent that are relevant for constraining PDFs.

- 788 • The measurements on normalized differential distribution of Z boson in p_T with different rapidity intervals from
 789 ATLAS collaboration [226, 227] at LHC Run I (7 TeV).
- 790 • The measurement on both normalized and unnormalized differential distribution of Drell-Yan lepton pair in p_T
 791 or ϕ_η^* with different rapidity intervals from ATLAS collaboration [224] at LHC Run I (8 TeV).
- 792 • The measurement on normalized differential distribution of Z boson in p_T integrated over rapidity from CMS
 793 collaboration [175] at LHC Run I (7 TeV).
- 794 • The measurement on both normalized and unnormalized double differential distribution of Z boson in p_T and
 795 rapidity from CMS collaboration [225] at LHC Run I (8 TeV).
- 796 • The measurement on normalized differential distribution of Z boson in p_T and the ratio to distribution of W
 797 boson from CMS collaboration [?] at LHC Run I (8 TeV).
- 798 • The measurements on unnormalized differential distribution of Z boson in ϕ_η^* in forward region from LHCb
 799 collaboration [184, 186] at LHC Run I (7 and 8 TeV).
- 800 • The measurement on normalized differential distribution of Z boson in p_T or ϕ_η^* in forward region from LHCb
 801 collaboration [185] at LHC Run I (7 TeV).
- 802 • The measurement on unnormalized differential distribution of Drell-Yan lepton pair in ϕ_η^* with different rapidity
 803 intervals from D0 collaboration [?] at Tevatron Run II (1.96 TeV).

804 Figure 13 shows a summary of various experimental uncertainties for the ATLAS and CMS 8 TeV measure-
 805 ments [224, 225] on the normalized p_T distribution around the Z pole. The luminosity error and some of the sys-
 806 tematics cancels largely in the normalized distributions. Both ATLAS and CMS measured p_T upto about 1 TeV with
 807 ATLAS has more fined binning at small p_T . The statistical errors are well within 1% for p_T smaller than 200 GeV
 808 and reach tens percents at the high p_T tail for ATLAS. The total systematic errors start at about two permil for $p_T \sim$
 809 10 GeV, and are within 1% in most of the region for both ATLAS and CMS. The PDF uncertainties as from individual
 810 PDF groups are about 2% which is already larger than the experimental errors in general not counting the spread

811 of different PDFs. The distributions can be measured by integrating over rapidity of the Z boson or separating into
 812 different rapidity intervals, or for off Z -peak invariant mass region. Besides, there are also measurements on distribu-
 813 tions of the angular variable ϕ_η^* [224] which is proportional to $p_{T,Z}$ at small transverse momentum. The experimental
 814 systematics can be further reduced for ϕ_η^* since it only depends on directions of the lepton momentum which is better
 815 measured than the momentum itself.

816 *Theoretical calculations and tools*

817 The NLO QCD corrections were calculated decades ago [228, 229, 230], and found to be sizable for the energy of
 818 the LHC. The EW corrections were studied extensively in recent years [231, 232, 233, 234]. The NLO predictions for
 819 the inclusive p_T spectra are inadequate for precision determination of PDFs due to its large theoretical uncertainties
 820 as can be either estimated from the size of the NLO corrections or the QCD scale variations. Fortunately the NNLO
 821 QCD corrections have just been calculated by two groups independently utilizing different methods, the antenna
 822 subtraction method [155, 235, 236] and the N -jettiness subtraction method [237, 238]⁴. The theoretical uncertainties
 823 are now greatly reduced which allows the inclusion of the Z boson p_T spectra data in the precision determination of
 824 PDFs. The results from two groups are in good agreements in general which further validates the NNLO calculations.
 825 Besides, the calculations can include leptonic decays of the Z boson and thus selection cuts at parton level. That allows
 826 a direct comparison with the measured fiducial cross sections without relying on experimental extrapolation on the
 827 phase space. Figure 14 shows the NLO and NNLO predictions on the unnormalized and normalized p_T spectra of the
 828 Z boson at LHC 8 TeV as from Ref. [236]. The selection cuts on the reconstructed Z boson are demonstrated in the
 829 plot while the fiducial cuts on charged leptons are explained in the figure caption. Central values of renormalization
 830 and factorization scale are set to transverse mass of the Z boson with scale variations calculated by varying them
 831 simultaneously by a factor of 2. The NNLO corrections are moderate for the unnormalized distribution, about 5%
 832 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
 833 EW corrections are small at moderate transverse momentum and can be sizable in the tail region, e.g., reach $\sim -10\%$
 834 for p_T greater than 600 GeV. However, the current large statistical errors in the tail region prevent a direct probe
 835 of effects from the EW corrections. The size of QCD corrections are similar for the normalized distribution. The
 836 denominator used is the inclusive Z production cross section at NNLO in same fiducial region and with independent
 837 scale variations.

838 Refs. [235, 236] have presented a detailed phenomenological study of the NNLO calculations comparing to the
 839 ATLAS and CMS 8 TeV measurements. There are very good agreements between the NNLO predictions and data
 840 on the normalized distribution for p_T ranges from 20 GeV to 900 GeV in all rapidity intervals. The conclusions
 841 are similar for comparison with the CMS 8 TeV data. Note that certain discrepancies are found for comparison to
 842 the ATLAS unnormalized distributions as shown in Figure 14. Moreover, it has been found that with the NNLO
 843 predictions now follow the same trend of the central values of data down to a p_T value of 4 GeV, largely improved as
 844 comparing to the NLO predictions.

845 *Impact on PDFs*

846 The impact of data on transverse momentum distribution of Z boson at LHC Run I have been studied very re-
 847 cently within the framework of global analyses in Refs. [239]. In above two analyses an additional uncorrelated
 848 errors of about 1% have been added for all p_T bins to account for additional theoretical errors, e.g., due to statistical
 849 uncertainties in MC integration of the NNLO calculations. It has been found that without the theoretical uncertainties
 850 the NNLO predictions can not describe the data well especially for the case of normalized distributions. Meanwhile
 851 certain tensions have been found as between the ATLAS 7 TeV normalized p_T distribution [227] and the 8 TeV p_T
 852 distribution from both ATLAS and CMS [224, 225]. The ATLAS 7 TeV data also pulls the PDFs to very different
 853 direction with respect to the HERA inclusive DIS data [?]. Ref. [239] concludes that the inclusion of the ATLAS
 854 7 TeV normalized data in the global analyses does not appear to be justified. Figure 15 shows impact of the ATLAS
 855 and CMS 8 TeV data in the NNPDF3.1 global analyses [239] by comparing the changes of PDFs in the analysis
 856 with and without the Z p_T data sets. The uncertainty of the gluon PDF have been slightly reduced in the x region of
 857 $10^{-2} \sim 10^{-1}$. In the same region the gluon PDF receives constraints from the precision measurements on top-quark

⁴The original calculations are for Z +jet production which can be easily translated to inclusive production of Z boson at finite p_T .

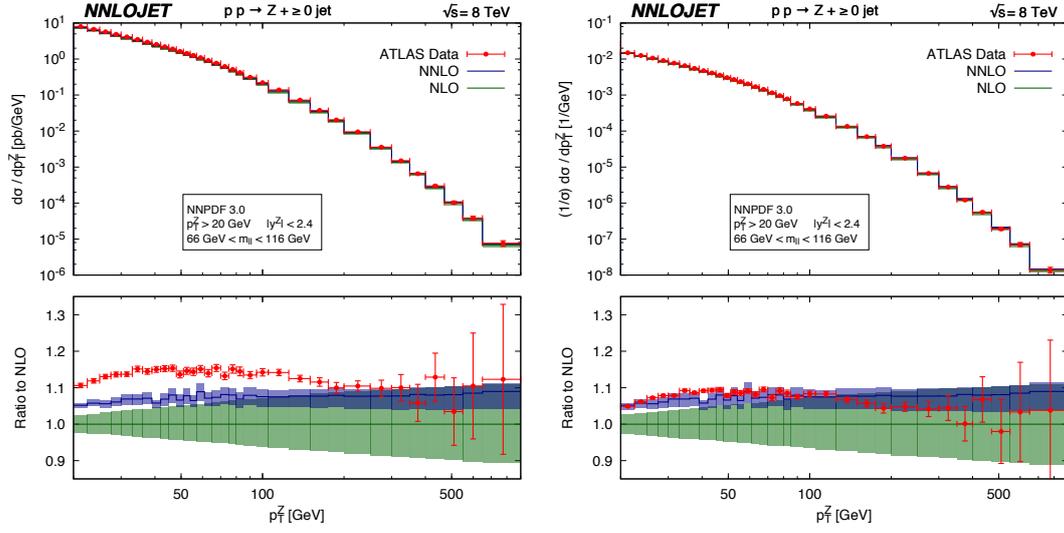


Figure 14: Left(right) plot shows the unnormalized(normalized) transverse momentum distribution of the inclusive Z boson production at LHC 8 TeV [236]. The green and blue bands denote the NLO and NNLO predictions with scale variations. The fiducial cuts on charged leptons are $p_{T,l} > 20$ GeV and $|\eta_l| < 2.4$.

858 pair production and HERA inclusive DIS that are also present in the same analyses. It is also found that the 8 TeV
 859 data lead to a moderate reduction of the PDF uncertainty of the total strangeness.

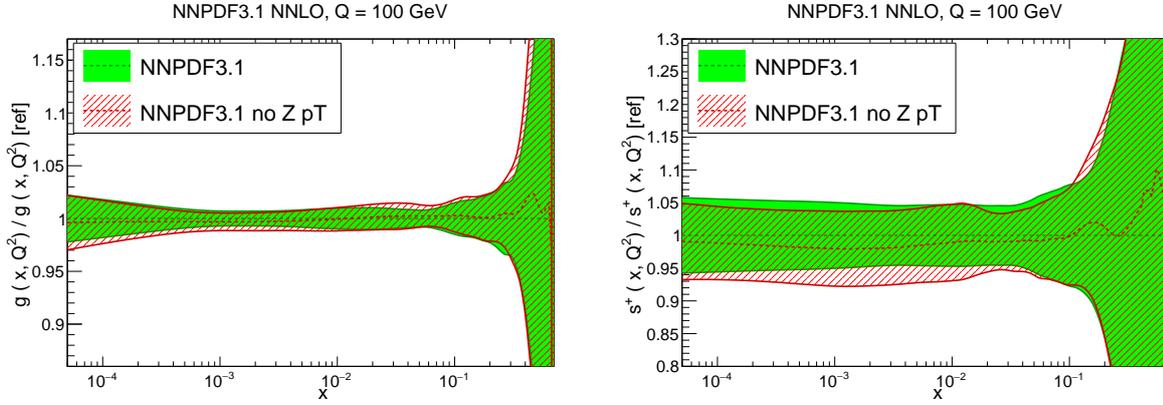


Figure 15: Impact of the Z boson transverse momentum measurements from ATLAS and CMS 8 TeV on the gluon PDF and the total strangeness in the NNPDF3.1 global analyses [239].

860 3.6. Direct photon production

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

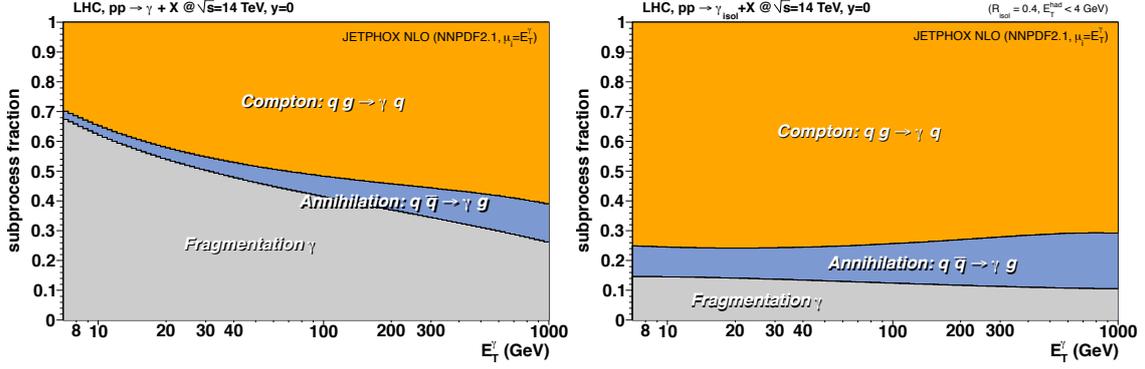


Figure 16: Relative contributions from Compton (qg), annihilation ($q\bar{q}$) and fragmentation to prompt photon production at central rapidities at the 14 TeV LHC, before (left) and after (right) the application of isolation cuts. Figures taken from [240].

862 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

$$\text{Compton : } \quad qg \rightarrow q\gamma, \quad (60)$$

$$\text{Annihilation : } \quad q\bar{q} \rightarrow g\gamma. \quad (61)$$

863 The Compton process gives the dominant contribution, in particular at the LHC. For LO kinematics the momentum
 864 fraction carried by the incoming gluon is directly proportional to the transverse energy E_T^γ of the produced photon,
 865 and thus for higher E_T^γ this process provides a direct probe of the gluon PDF at high x . Moreover, this represents the
 866 highest rate electroweak process at the LHC, while the produced photon directly reflects the production kinematics,
 867 without for example requiring any additional hadronization corrections, as in the case of jet production. This can
 868 therefore provide a valuable tool with which to constrain the gluon.

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

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

887 In fact, isolated photon production represented one of the first PDF constraints considered, and was used in such
 888 early fits as [63, 64, 65]. However, the difficulties in describing the fixed target E706 [241, 242] data raised questions
 889 about the reliability of this process for PDF fits and potential sensitivities to non-perturbative effects. Combined with
 890 the increasing availability of high precision jet data from the Tevatron, which also constrain the high x gluon, this lead

891 to the process falling out of favour in the PDF fitting community. The last PDF set to include any such data is the
892 MRST99 [243] fit.

893 However, the subsequent studies of [244, 240] have shown that by increasing the \sqrt{s} from fixed target to collider
894 energies and, as discussed above, imposing a suitable isolation condition on the produced photon, the process may be
895 brought under reasonable theoretical control. Moreover, a comparison of the NLO pQCD predictions to the ATLAS
896 measurement [245, 246] discussed below shows an adequate description of the data, albeit with fairly large $\sim 10\text{--}15\%$
897 scale variation uncertainties. We may expect this situation to improve further with the recent NNLO calculation
898 discussed in Section 3.6. Thus isolated photon production may well provide a useful tool for LHC PDF constraints in
899 the future.

900 *Experimental data*

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

- 902 • The most recent data at 1.96 TeV from CDF [247] and D0 [248] extends out to $E_{\perp}^{\gamma} < 0.5$ TeV and 0.3 TeV,
903 respectively, while the photon pseudorapidity is restricted to have $|\eta^{\gamma}| \lesssim 1$. In the CDF case this corresponds to
904 the full Run II 9.5 fb^{-1} data set, and so represents the final legacy measurement.
- 905 • The ATLAS 7 TeV measurement [249], out to $E_{\perp}^{\gamma} < 1$ TeV and at 8 [245] and 13 [246] TeV, extending to
906 $E_{\perp}^{\gamma} < 1.5$ TeV. These correspond to the full available integrated luminosities of 4.6 fb^{-1} and 20.2 fb^{-1} at 7 and 8
907 TeV, respectively, while the 13 TeV measurement uses a 3.2 fb^{-1} data set.
- 908 • The most precise CMS data at 7 TeV [250], corresponding to 36 pb^{-1} of integrated luminosity and extending to
909 $E_{\perp}^{\gamma} < 0.4$ TeV.
- 910 • Data from a smaller sample at 2.76 TeV have also been taken by ATLAS [251] and CMS [252]. In all cases the
911 photon pseudorapidity is restricted to have $|\eta^{\gamma}| \lesssim 2.4$.

912 *Theoretical calculations and tools*

913 For the past 15 years, the theoretical state-of-the-art was provided by the JETPHOX [253] MC generator, which
914 implements both the direct and the fragmentation contributions consistently at NLO. The NLO EW corrections have
915 also been calculated in [254]. However, recently the first NNLO calculation of direct photon production has been
916 reported [255]. The NNLO prediction for the ATLAS 8 TeV data [245] is compared to the NLO and found to lie
917 consistently within the NLO scale uncertainty band, with the central value being $\sim 5\%$ higher. Moreover, the NNLO
918 scale uncertainty is found to be greatly reduced, giving a $\sim \pm 2 - 3\%$ uncertainty.

919 While the default description of the data is found to be quite poor, including the LL EW Sudakov corrections
920 of [234], and evaluating the coupling α at the scale M_Z , as recommended in [234], the description is improved. In
921 particular, the EW corrections are found to reduce the cross section by as much as 10%, that is significantly outside
922 the QCD scale variation band, at the highest E_{\perp} , improving the shape description. The results of the PeTeR [256],
923 which combines the NLO calculation with N^3LL threshold resummation in addition to these EW corrections, is found
924 to lie close to the NNLO + EW prediction, but with a larger uncertainty band, indicating that the data may not be too
925 sensitive to such additional resummation effects. Therefore, while the NNLO calculation is a very encouraging step
926 towards including isolated photon data in high precision PDF fits, there are clearly some further theoretical issues to
927 be investigated, relating to the impact of EW corrections and, as discussed in [255] the choice of photon isolation,
928 which can also affect the NNLO comparison.

929 *Impact on PDFs*

930 Currently no up to date studies of the impact of isolated photon data on the PDFs have been performed, in particular
931 taking into account the new NNLO calculation and the high precision LHC data. However, in [240] (see also [244])
932 the impact of a range of data, including the earlier 36 pb^{-1} ATLAS and CMS measurements at 7 TeV, on the PDFs has
933 been studied in detail through a reweighting of the NNPDF2.1 set. In Fig. 17 the impact of this LHC data on the
934 gluon PDF is shown. A significant reduction in the uncertainty, of up to 20%, is found in the intermediate x region.
935 Interestingly, this overlaps with the kinematically relevant region for Higgs boson production via gluon fusion at the
936 LHC, and indeed a $\sim 20\%$ reduction in the Higgs production cross section is found. Given these results correspond

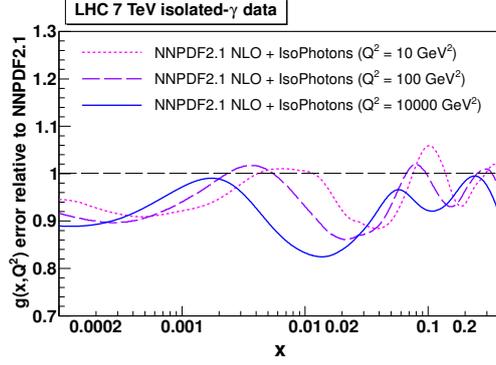


Figure 17: Relative reduction in the NNPDF2.1 NLO gluon PDF uncertainty when including a 36 pb^{-1} LHC 7 TeV isolated photon data set via reweighting. Taken from [240].

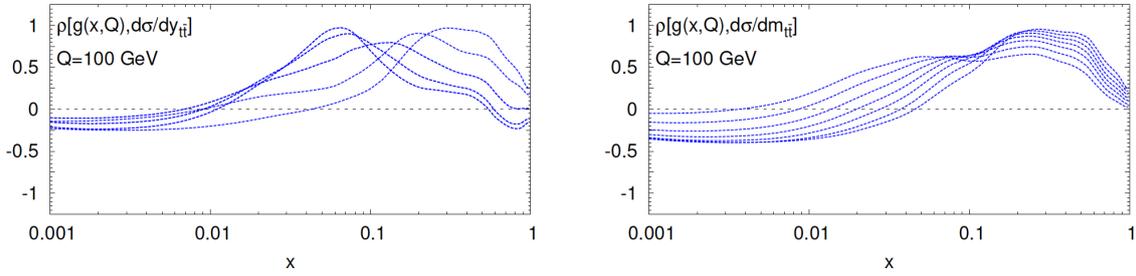


Figure 18: The correlation coefficient between the gluon PDF at $Q = 100 \text{ GeV}$ and the theory predictions for the differential distributions in y_{i1} (left) and m_{i1} (right plot) at $\sqrt{s} = 8 \text{ TeV}$, as a function of x . Each curve corresponds to specific measurement bin. The higher the absolute value of the correlation coefficient, the bigger the sensitivity to the gluon in those specific values of x .

937 to a reasonably limited LHC data set, it will be interesting to see the impact of the latest data, as well as the NNLO
 938 corrections.

939 3.7. Top quark production

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

942 PDF sensitivity

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

947 To illustrate the kinematical sensitivity on the gluon of top quark pair production, in Fig. 18 we show the correla-
 948 tion coefficient $\rho[g(x, Q), d\sigma]$ between the gluon PDF at $Q = 100 \text{ GeV}$ and the theory predictions for the differential
 949 distributions in y_{i1} and m_{i1} at $\sqrt{s} = 8 \text{ TeV}$, as a function of x . Each curve corresponds to specific measurement bin.
 950 The higher the absolute value of the correlation coefficient, the bigger the sensitivity to the gluon in those specific
 951 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
 952 processes sensitive to the gluon such as inclusive jet production. Moreover, the availability of differential distributions
 953 significantly extends the kinematical coverage of total inclusive cross-sections.

954 *Experimental data*

955 The first measurements of top quark pair production were presented at the level of total cross-sections, first at
956 the Tevatron [257] and then by ATLAS and CMS [258, 259, 260, 261, 262, 263]. Subsequently, single-inclusive
957 differential distributions of top quark pair production have also been presented by ATLAS [264, 265] and CMS, which
958 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
959 well as at the level of observables contributed in terms of directly observable quantities (charged lepton p_T and rapidity,
960 b -tagged jet kinematics etc). These measurements are often presented normalized to the total cross-section, in order to
961 benefit from a number of cancellations between experimental systematic uncertainties, but absolute measurements are
962 also available. It is also possible to measure double differential distributions in top quark pair production, as illustrated
963 by the recent CMS measurement [266] of normalized double differential distributions, *i.e.*, as a function of $p_{t\bar{t}}$ and $m_{t\bar{t}}$.

964 *Theoretical calculations and tools*

965 The total $t\bar{t}$ production cross-section is known up to NNLO-QCD since 2013 [267, 268, 269], including the re-
966 summation of logarithmically enhanced threshold corrections up to NNLL. More recently, the full NNLO corrections
967 to single inclusive distributions in top quark pair production have also been computed [270, 271]. Differential NNLO
968 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
969 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 $p_T^{\bar{t}}$ since these
970 vanish at leading order. When differential distributions probe the TeV regions, electroweak corrections (including
971 photon-initiated processes) also become relevant and need to be included in the theoretical calculations. In Ref. [272]
972 (see also [273]), the NNLO QCD calculation where combined with the state-of-the-art NLO EW corrections in the
973 latter case including not only contributions of $\mathcal{O}(\alpha_s^2\alpha)$ but also those of order $\mathcal{O}(\alpha_s\alpha^2)$ and $\mathcal{O}(\alpha^3)$. This study showed
974 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,
975 necessarily must include NLO EW corrections.

976 An important limitation of the calculations discussed above is that they are restricted to stable top quarks. On
977 the other hand, when experimental measurements are presented at the top quark level, they are extrapolated from the
978 fiducial cross-sections using some theoretical model, thus possibly biasing the result by an amount which is difficult to
979 quantify. Ideally, one would like a fully differential calculation with NNLO corrections included both for production
980 and decay, in order to directly compare with experimentally observable quantities. An important milestone in this
981 respect was the recent calculation of top-quark pair-production and decay [274] which allows providing predictions
982 for observables constructed from top-quark leptonic and b -tagged jet final states, based on an approximation to the
983 exact NNLO corrections to production and exact NNLO corrections to the decay.

984 Concerning the tools for the inclusion of top quark differential data into PDF fits, there exist two basic approaches.
985 The first one is based on computing APPLgrids for the NLO calculation using either MCFM or Sherpa (see also
986 Sect. 3.10), and then supplementing these with the NNLO/NLO bin-by-bin K -factors from [270, 271]. An improved
987 strategy has been made feasible by the recent availability of FastNLO tables [275] that allow the efficient calculation
988 of NNLO top quark pair distributions for arbitrary PDF sets and input $\alpha_s(m_Z)$ values. The latter option bypasses the
989 need of recomputing the K -factors in each iteration of the fit, although on the other hand as shown explicitly in [276]
990 the PDF dependence of these K -factors is very small.

991 *Impact on PDFs*

992 The availability of the NNLO calculation of the total cross-sections for top quark pair production made possible
993 to include for the first time top quark data from the Tevatron and the LHC consistently into a NNLO PDF fit. By
994 applying Bayesian reweighting to NNPDF2.3, it was shown in Ref. [277] that indeed top quark data could reduce the
995 PDF uncertainties on the large- x gluon by up to 20% around $x \simeq 0.2$ (see also previous related work in [278]). Several
996 other global fits, such as ABMP16 and MMHT14, also include now total $t\bar{t}$ cross-sections in their default fits. While
997 these results provided a tantalizing hint of the PDF constraining potential of $t\bar{t}$ production, the full exploitation of this
998 potential clearly required the use of differential distributions.

999 The impact of the $\sqrt{s} = 8$ TeV top quark pair differential data from ATLAS and CMS on the NNPDF3.0 fit
1000 was quantified in [276]. There it was shown that the constraints on the large- x gluon were at this point competitive
1001 with those provided by inclusive jet production, despite the much reduced number of experimental data points. See
1002 also [279] for related work based on approximate NNLO calculations. An important result of the investigations

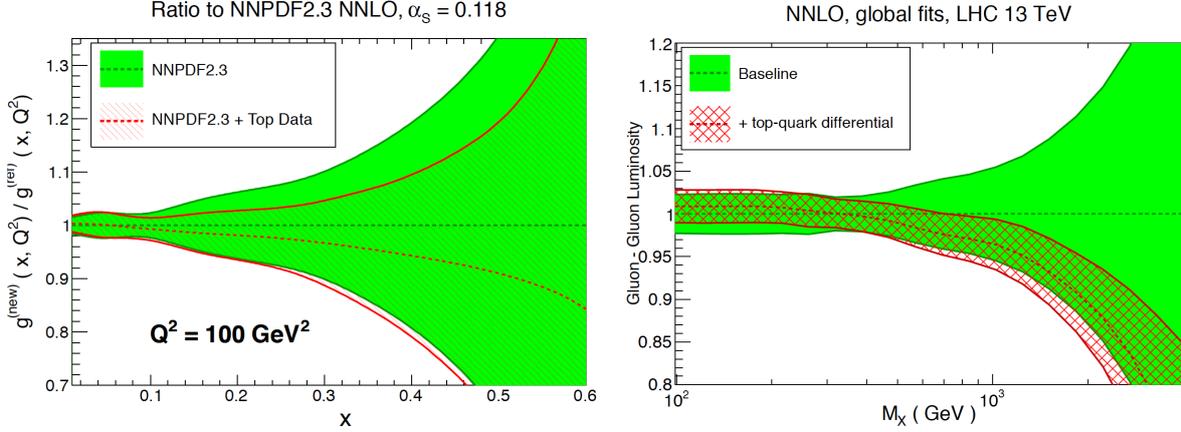


Figure 19: Left plot: the impact of the LHC 7 and 8 TeV inclusive top-quark pair cross-section data on the large- x gluon of NNPDF2.3 [277]. Right plot: the impact of the LHC 8 TeV differential distributions in top-quark pair production on the gg luminosity [276], compared with a baseline fit based on the NNPDF3.0 global analysis without the jet data.

1003 of [276] was that the constraints from the normalized distributions were in general superior to those from their absolute
 1004 counterparts, most likely because of the cancellation of systematic uncertainties that arises when the distributions are
 1005 normalized. In addition, top quark differential distributions at 8 TeV from the LHC have also been included in the
 1006 recent NNPDF3.1 global analysis, and other groups have also studied the impact of this data into their PDF fits in a
 1007 preliminary form. A challenge in the study of [276] was the observed tension between some of the ATLAS and CMS
 1008 distributions, such as the $m_{t\bar{t}}$, which prevented their simultaneous inclusion in the global fit. While the underlying issue
 1009 of these discrepancies are still being investigated, this limitation was bypassed by identifying pairs of distributions
 1010 which could be fitted with good quality at the same time and that exhibited comparable constraining power.

1011 In order to illustrate the impact of the top quark data on the large- x gluon, in Fig. 19 we show how the PDF
 1012 uncertainties of the NNPDF2.3 gluon are reduced once the Tevatron and the LHC 7 and 8 TeV inclusive top-quark
 1013 pair cross-section data are included by means of Bayesian reweighting [277]. As can be seen, at the level of total
 1014 cross-section data the impact is still moderate, with PDF error reduction being at most $\approx 20\%$. On the other hand,
 1015 in Fig. 19 we also show the impact on the gg luminosity of the normalized 8 TeV ATLAS and CMS $t\bar{t}$ differential
 1016 distributions, compared to a baseline fit based on the NNPDF3.0 settings but without the jet data (since the NNLO
 1017 corrections to jet production were not available at the time). One clearly sees here how the constraints are much more
 1018 significant, highlighting the increase in constraining power of the differential distributions are compared to the total
 1019 cross-sections data, specially in the large- x region, where PDF uncertainties can be reduced by more than a factor of
 1020 2.

1021 *Single top production*

1022 In addition to top quark pair production, also single top production provides in principle interesting PDF-sensitive
 1023 information. Specifically, single top could provide constraints on the b -quark PDF, given that as shown in Fig. 20,
 1024 some of the leading order contributions arise from the scattering of a b with a light quark. In addition, single top
 1025 production is a unique testing group for different heavy quark flavour schemes used in the calculation, *i.e.*, one can
 1026 use a $n_f = 4$ scheme, a $n_f = 5$ scheme, or a matched scheme interpolating between the two, see the discussion in
 1027 Refs. [280, 281]. State of the art calculations are based on NNLO QCD theory both for total cross-sections and for
 1028 differential distributions [282], and LHC measurements at 8 TeV and 13 TeV of total cross-sections (including ratios
 1029 of top over anti-top production) as well as single inclusive distributions with the corresponding covariance matrix are
 1030 already available [283, 284].

1031 From Fig. 20 is clear that single top production could be interesting in PDF fits, since it provides sensitivity to
 1032 the bottom content of the proton. Moreover, since the production of top and anti-top quarks depends on different
 1033 initial state partonic luminosities, it is possible to construct ratios such as $R_t \equiv \sigma_t / \sigma_{\bar{t}}$ that provide information on

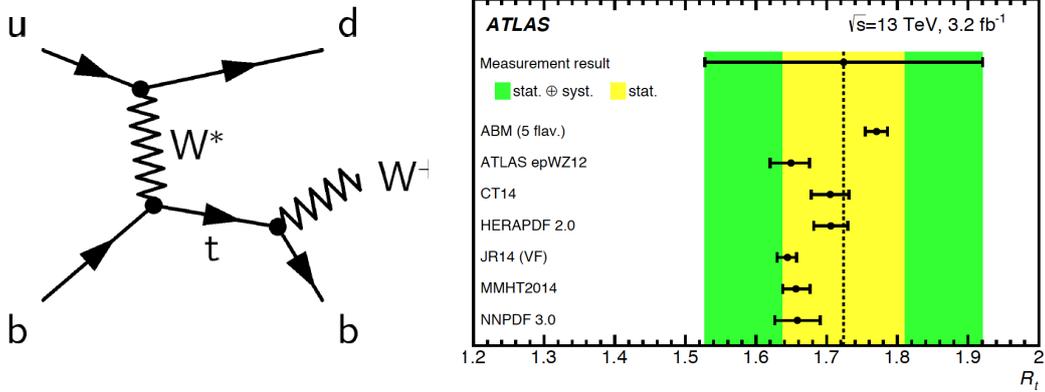


Figure 20: Left plot: one of the Feynman diagrams for single-top production at leading order, illustrating its sensitivity to the b -quark PDF. Right plot: comparison of the theoretical predictions for the ratio $R_t = \sigma_t / \sigma_{\bar{t}}$ from different PDF sets and the corresponding ATLAS measurements at $\sqrt{s} = 13$ TeV from [284].

1034 quark flavour separation. To illustrate this point, we show in Fig. 20 a comparison of the theoretical predictions for
 1035 the R_t ratio from different PDF sets and the corresponding ATLAS measurements at $\sqrt{s} = 13$ TeV from [284]. While
 1036 experimental uncertainties are still large due to the limited dataset, we also see that the measurement might eventually
 1037 become sensitive to differences between PDF sets, for instance the ABM prediction is higher as compared to the other
 1038 predictions. Moreover, similar comparisons could also be performed at the level of differential distributions, either at
 1039 the level of top kinematic variables or at the level of observable quantities constructed from leptons and b -jets.

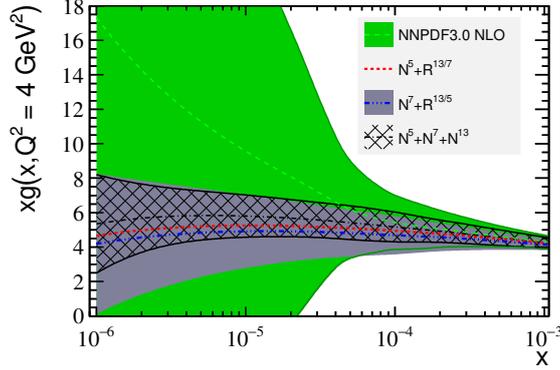


Figure 21: Left plot: comparison between the small- x gluon at $Q^2 = 4 \text{ GeV}^2$ in NNPDF3.0 with the corresponding result after different combinations of the charm production data at LHCb have been included in the fit. We show the central value and one-sigma PDF uncertainty bands for the $N^7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

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 $x \approx 10^{-6}$, well below the kinematic reach of the HERA structure function data, and thus 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 to 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 at very small- x 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 the so called “prompt” neutrinos and which dominate the atmospheric neutrino flux at large 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 $Q^2 = 4 \text{ GeV}^2$ 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. We see that the reduction of the small- x PDF uncertainties can be as large as almost an order of magnitude.

3.9. Other PDF-sensitive process

In this final part of the section, we briefly review other processes that have been suggested as possible ingredients of the global PDF analysis, but that have been studied rather less than the rest of processes in this section.

Central 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

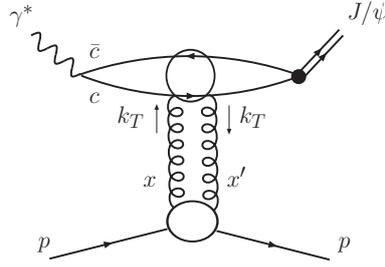


Figure 22: Figure from 1307.7099. perhaps don't need a figure- if we decide to keep, will find better one...

1071 via t -channel two gluon exchange. This may therefore access the gluon PDF at a comparatively low scale $Q^2 \sim M_V^2$
 1072 and $x \sim M_V/\sqrt{s}$, where it is so far quite poorly determined.

1073 LHCb have measured the exclusive production of J/ψ and $\psi(2S)$ mesons at 7 TeV [285] and Υ production at 7
 1074 and 8 TeV [285]. Preliminary LHCb data on J/ψ and $\psi(2S)$ production at 13 TeV has also been reported in [286].
 1075 This uses the newly installed HeRSChE L shower counters, which greatly extend the rapidity coverage for vetoing ad-
 1076 ditional particle production, reducing significantly the non-exclusive background. In addition, ALICE have measured
 1077 exclusive J/ψ production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV [287]. Due to the large Z^2 enhancement for photon
 1078 emission from the Pb ion, this is dominated by the process where the lead ion emits a photon.

1079 While clearly the photoproduction process shown in Fig. 22 proceeds through an initial-state gluon interaction,
 1080 this does not correspond to a standard inclusive process where PDFs are generally introduced. However, as discussed
 1081 in [288] under certain assumptions this process can be related to the standard gluon PDF and may therefore serve as
 1082 a probe of it at low x and Q^2 . This analysis has subsequently been performed at NLO [289] (see also [290]) for the
 1083 cases of J/ψ and Υ production. However, here it is found that the NLO correction is significantly larger than, and
 1084 of the opposite sign to, the LO contribution, indicating a lack of perturbative stability and casting some doubt on its
 1085 viability as a PDF probe. Nonetheless, work in the direction of at least partially solving this issue has been reported
 1086 most recently in [291], where the stability is shown to be improved through judicious choice of factorization and
 1087 renormalization scale, and by imposing a cut on the NLO contribution to avoid double counting. Other possibilities,
 1088 such as for example considering cross section ratios as discussed in **section**, may also improve the situation. It is
 1089 also worth noting that the perturbative stability is naturally improved somewhat by considering the production of the
 1090 higher mass Υ .

1091 *W, Z+jets production.* ...

1092 *Photon+jet production.* In addition to the inclusive photon production process discussed in Sect. 3.6, it has also been
 1093 suggested that the photon+jet process could also be used to PDF studies [292].

1094 3.10. Fast interfaces

1095 Given the highly CPU time-consuming character of global PDF fits, the direct evaluation of the lengthy (N)NLO
 1096 hadronic cross-sections during the PDF fit is not feasible. For this reason, until around 2008 PDF fits included hadronic
 1097 data using LO hadronic cross-sections supplemented by bin-by-bin K -factors, defined as

$$K_{\text{NLO}}^i \equiv \frac{\sigma_i^{\text{NLO}}}{\sigma_i^{\text{LO}}}, \quad (62)$$

1098 using the same PDF set in the numerator and in the denominator. To ensure consistency of the procedure, these K
 1099 factors were computed iteratively until convergence was achieved. However, this approximation is known to have
 1100 several deficiencies, the most important one being the reduced sensitivity to those partonic initial states that enter the
 1101 cross-section only starting from NLO.

1102 In order to improve over this unsatisfactory situation, the method of *fast interfaces* was proposed. In these methods,
 1103 the most CPU time consuming part of an NLO calculation, namely the evaluation of the partonic matrix elements over
 1104 a large number of events, is precomputed *a priori* using a complete interpolation basis for the input PDFs. This way,
 1105 the hadronic cross-sections can be reconstructed *a posteriori* by means of a very efficient matrix multiplication of the
 1106 PDFs evaluation. in a grid of (x, Q) points and the precomputed partonic matrix elements at the same grid points.
 1107 These tools have become very popular and are therefore used in the majority of former PDF fits.

1108 Following a common philosophy, two main tools have been developed

- 1109 • APPLgrid [293]
- 1110 • FastNLO [294]

1111 More recently, the aMCfast interface [295] to MadGraph5_aMC@NLO [296] has been developed. Given the auto-
 1112 mated character of this code, aMCfast allows producing fast interpolation of arbitrary NLO processes, defined by
 1113 the user at run time. It is also possible to produce fast grids for NLO calculations matched to parton shower Monte
 1114 Carlos, which opens the way to include in the PDF fit hadron-level cross-sections such as W boson in association with
 1115 charmed mesons or forward D meson production at LHCb.

1116 We now describe the basic strategy of fast interpolation methods - we use the notation of the APPLgrid paper
 1117 for concreteness, but the general method is very similar in FastNLO and aMCfast. These methods are based on
 1118 representing the PDFs in the (x, Q^2) by means of a suitable interpolation basis, computing a physical cross-section for
 1119 this basis PDFs, and then reconstructing the same observable *a posteriori* using an arbitrary PDF set. Therefore, to
 1120 begin, with, one expands an arbitrary PDF $f(x, Q^2)$ in terms of a suitable basis of interpolating polynomials

$$f(x, Q^2) = \sum_{i=0}^n \sum_{j=0}^m f_{k+i, \kappa+j} I_i^{(n)} \left(\frac{y(x)}{\delta y} - k \right) I_j^{(m)} \left(\frac{\tau(Q^2)}{\delta \tau} - \kappa \right), \quad (63)$$

1121 where n and m are the interpolation orders in x and Q^2 respectively; $y(x) = \ln 1/x + a(1-x)$ and $\tau(Q^2) = \ln \ln Q^2/\Lambda^2$;
 1122 and $I_i^{(n)}, I_j^{(m)}$ are interpolating functions, for instance Lagrange interpolating polynomials, though Eq. (63) holds
 1123 generically for other choices.

1124 After the representation Eq. (63) has been constructed, we need to evaluate cross-sections using the interpolation
 1125 basis. Let us consider first for simplicity a hypothetical DIS structure function F that receives contributions from a
 1126 single flavour. The NLO cross-section is typically computed by means of Monte Carlo program that generates a large
 1127 number N of events, each one with weight ω_m and with associated values x_m and Q_m^2 . If p_m is the order of α_s for this
 1128 specific event, the total cross-section can be written as

$$F = \sum_{m=1}^N \omega_m \left(\frac{\alpha_s(Q_m^2)}{2\pi} \right)^{p_m} f(x_m, Q_m^2). \quad (64)$$

1129 The fast interpolation can be constructed by, instead of computing F as in Eq. (64), introducing a weight grid $W_{i_y, i_\tau}^{(p)}$,
 1130 and for each event only a fraction of the grid nodes is updated according to the expression

$$W_{k+i, \kappa+j}^{(p)} \rightarrow W_{k+i, \kappa+j}^{(p)} + \omega_m I_i^{(n)} \left(\frac{y(x_m)}{\delta y} - k \right) I_j^{(m)} \left(\frac{\tau(Q_m^2)}{\delta \tau} - \kappa \right). \quad (65)$$

1131 Conceptually, the weight grid $W_{i_y, i_\tau}^{(p)}$ is the equivalent of computing the structure function F but for a given combination
 1132 of interpolating polynomials as opposed to the original PDF.

1133 The important factor here is that the most CPU time intensive computation, the calculation of the MC weights
 1134 ω_m , needs only to be done once to fill the grid $W_{i_y, i_\tau}^{(p)}$, and the the PDF can be decided a posteriori at virtually no extra
 1135 computational cost. Indeed, it can be shown that the structure function can be reconstructed a posteriori using the
 1136 weight grid using

$$F = \sum_p \sum_{i_y} \sum_{i_\tau} W_{i_y, i_\tau}^{(p)} \left(\frac{\alpha_s(Q_{i_\tau}^2)}{2\pi} \right) f(x_{i_y}, Q_{i_\tau}^2). \quad (66)$$

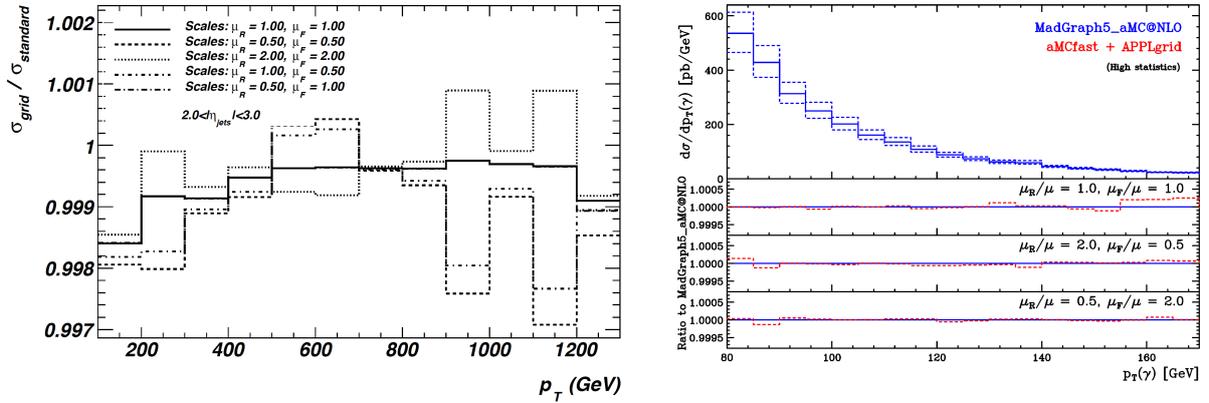


Figure 23: Right plot: the ratio between the NLOjet++ calculation of inclusive jet production at 7 TeV in the rapidity interval $2 \leq y \leq 3$ and the corresponding *a posteriori* calculation based on APPLgrid, for different values of the factorization and renormalization scales. Left plot: the transverse momentum distribution of photons in the $pp \rightarrow \gamma + \text{jet}$ 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 .

1137 In other words, the only information which is needed is the value of the PDFs and the strong coupling at the grid nodes
 1138 i_y, i_τ . The method can be straightforwardly generalized to hadron-hadron collisions and to a generic composition of
 1139 the initial parton state, taking into account that now the formula includes two PDFs. In proton-proton collisions, the
 1140 analog of Eq. (66) is given by

$$\sigma = \sum_p \sum_{l=0}^{n_{\text{sub}}} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W_{i_{y_1}, i_{y_2}, i_\tau}^{(p)(l)} \left(\frac{\alpha_s(Q_{i_\tau}^2)}{2\pi} \right) \mathcal{L}^{(l)}(x_1^{i_{y_1}}, x_2^{i_{y_2}}, Q_{i_\tau}^2), \quad (67)$$

1141 where we have indicated that there are n_l contributing partonic subprocesses, each with the corresponding luminosity
 1142 $\mathcal{L}^{(l)}$, which depend on the cross-section upon consideration.

1143 In order to illustrate the high precision that these fast interfaces can achieve, we show two representative examples
 1144 in Fig. 23. First of all, we show the ratio between the NLOjet++ calculation of inclusive jet production at 7 TeV in the
 1145 rapidity interval $2 \leq y \leq 3$ and the corresponding *a posteriori* calculation based on APPLgrid, for different values of
 1146 the factorization and renormalization scales. One sees that the differences between the original and the interpolated
 1147 calculation are at the few permille level. Then we show the transverse momentum distribution of photons in the
 1148 $pp \rightarrow \gamma + \text{jet}$ process at 7 TeV, comparing the original MadGraph5_aMC@NLO calculation with the *a posteriori* result
 1149 based on aMCfast and APPLgrid. The lower insets show the ratio between the two calculations for different choices
 1150 of μ_R and μ_F . Also here we find excellent agreement between the original and interpolated calculations, now at the
 1151 sub-permille level. In all these methods, the interpolation accuracy can be arbitrarily increase *i.e.* by using denser
 1152 grids in x and Q^2 .

1153 While these fast interface represent a very significant improvement in terms of CPU efficiency as compared to
 1154 the original NLO calculations, one limitation of the approach can be seen from the master formula for proton-proton
 1155 collisions Eq. (64): each time the PDF set is varied, one needs to recompute its values in the (x, Q^2) nodes. In a PDF
 1156 fit, this means that each time the input parametrization is modified during the iterative minimization, the DGLAP
 1157 evolution equations need to be solved again, before the PDFs can be convoluted with the interpolated coefficient
 1158 functions to obtain the hadronic cross-section. To improve this shortcoming, recently the APFELgrid tool has been
 1159 developed [297]. The goal of APFELgrid is to combine the interpolated partonic cross-sections provided by APPLgrid
 1160 with the DGLAP evolution factors provided by APFEL, in a way that hadronic cross-sections can be reconstructed from
 1161 a matrix multiplication requiring only as input the values of the PDFs at the x grid nodes at the input evolution scale
 1162 Q_0 . This combination then lead to a very significant improvement in computation speed as compared to Eq. (64)

APFELgrid/FK timings gcc-5.2.1 on i7-6500U CPU @ 2.50GHz

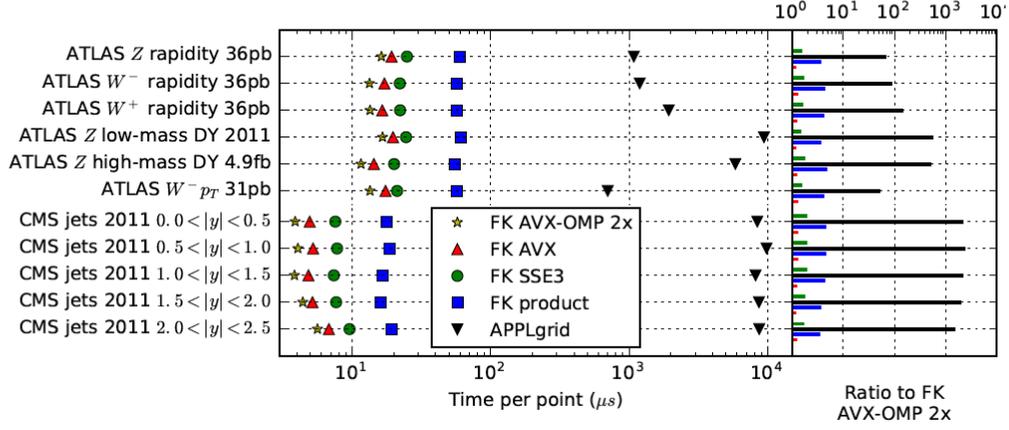


Figure 24: Comparison of the timings per data point between the original APPLgrid computation of hadronic cross-sections, Eq. (66), with the same calculation based on the APFELgrid combination, Eq. (68), for a variety of LHC datasets. We find that the improvement in computational speed is between a factor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

1163 without any loss of numerical accuracy, and leads to much faster PDF fits. Mathematically, the APFELgrid method
1164 allows expression an arbitrary hadronic cross-section as follows

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

1165 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
1166 the nodes of the x interpolating grid.

1167 To gauge the improvements in computational efficiency that can archived by this method, in Fig. 24 we show
1168 a comparison of the timings per data point between the original APPLgrid computation of hadronic cross-sections,
1169 Eq. (66), with the same calculation based on the APFELgrid combination, Eq. (68), for a variety of LHC datasets. As
1170 we see from this figure, the improvement in computational speed is between a factor 100 and a factor 1000 depending
1171 on the specific dataset. This means that PDF fits based on APFELgrid will be much faster, by up to two orders of
1172 magnitude, which is a very attractive property of this method.

1173 **4. Fitting methodology**

1174 In this section we present the framework of the global QCD analysis of parton distribution functions. First of all
 1175 we discuss how PDFs are parametrized at the reference scale, and review the theoretical constraints that should be
 1176 imposed to this parametrization such as the momentum and valence sum rules and positivity. Then we review the
 1177 various methods available to estimate and propagate PDF uncertainties, with emphasis on the three most important
 1178 ones: the Hessian, the Monte Carlo, and the Lagrange multiplier methods. The next building block of the QCD fitting
 1179 program is the minimization. In this section we also discuss how to combine individual PDF sets within a single PDF
 1180 set.

1181 *4.1. PDF parametrization*

1182 *4.1.1. Functional form*

1183 To extract the PDFs some form of parameterisation in x must be assumed, which can then be fit to the available
 1184 data. As described in **section**, given the PDFs at some reference scale Q_0 , DGLAP evolution determines the PDFs
 1185 at any other scale μ . Thus the PDFs are typically parameterised at a low scale $Q_0^2 \sim 1 - 2 \text{ GeV}^2$, which can then be
 1186 evolved up to the scale relevant to e.g., LHC physics. These universally take the form

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

1187 The $(1-x)^{b_f}$ term, with $b_f > 0$, ensures that the PDFs vanish in the elastic $x \rightarrow 1$ limit, as we would expect on
 1188 basic physical grounds. Such a form is also expected from the quark counting rules [298]. Here, in this elastic limit
 1189 all the momentum is carried by the struck parton and the remaining n_s quark become spectators. An analysis of the
 1190 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
 1191 distributions, respectively. The x^{a_f} form dominates at low x ; in this region, the PDFs are related to the high energy
 1192 parton–proton scattering amplitudes, which may be calculated using the tools of Regge theory. This predicts such a
 1193 simple power–like form, with the precise value of the power a_f being related to the leading Regge trajectory that is
 1194 exchanged; for non–singlet distributions (e.g. the valence quarks) this predicts $a_f \sim 0.5$ and for singlet distributions
 1195 (e.g. the gluon and the sea) this predicts $a_f \sim 0$. The above discussion only correspond to quite general expectations,
 1196 which do not for example account for the scale dependence of the PDFs. Thus while the high and low x form of (69)
 1197 is taken, for modern fits the values of the powers themselves are more generally left free where there is sufficient data
 1198 to constrain them.

1199 The $I_f(x)$ is the interpolating function, which determines the behaviour of the PDFs away from the $x \rightarrow 0$ and 1
 1200 limits, where it tends to a constant value. This is assumed to be a smoothly varying function of x , for which a variety
 1201 of choices have been made in the literature. The simplest ansatz, which has been very widely used, is to take a basic
 1202 polynomial form in x (or \sqrt{x}), such as

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

1203 Forms of this type are for example taken by CJ, HERAPDF, and in the previous MSTW08 set. A similar approach, but
 1204 where the polynomial enters as the exponent of a power of x or a simple exponential function, are taken by ABMP
 1205 and earlier CT sets, respectively.

1206 Such a choice is appropriate for a relatively small number of parameters c, d . However, as the precision and
 1207 amount of the data included in the fit increases it becomes essential to allow for an increasingly flexible parameter-
 1208 isation. As discussed in [299], simply adding more parameters to (70) can quickly run into the technical issues that
 1209 large coefficients appear, with large cancellations between the terms. This leads to an unstable χ^2 minimisation and
 1210 implausibly large variations in x in certain regions. This issue may be solved by instead expanding the interpolating
 1211 function in terms of a basis of suitably chosen functions

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

1212 where $y(x)$ is some simple function of x . Two choices for the functions P_i are the Chebyshev and Bernstein polynomi-
 1213 als, which are used in the MMHT14 and CT15 sets, respectively. These are chosen as each order of the polynomials
 1214 is strongly peaked at different values of y , and hence x , significantly reducing the degree of correlation between the

1215 terms. In addition, as the order is increased these tend to probe smaller scale variations in x , so that the smoothness
 1216 requirement for $I(x)$ naturally leads to smaller coefficients α at higher i . Thus, while formally equivalent to the simply
 1217 polynomial expansion (70), these are much more convenient for fitting as the number of free parameters n is increased.

1218 An alternative approach is taken by the NNPDF group. Here, the interpolating function is modelled with a multi-
 1219 layer feed forward neural network, see **section** for more details. In practice, this allows for a greatly increased number
 1220 of free parameters, with the latest default fit having 37 per PDF, that is \sim an order of magnitude higher than other sets.
 1221 The form of (69) is still assumed, but these are pre-processing factors that speed up the minimisation procedure but
 1222 which do not in principle have to be explicitly included. Nonetheless, the study of [300] has shown that the NNPDF
 1223 fit does exhibit high and low x behaviour that is consistent with (69), providing further support for such an assumed
 1224 form.

1225 4.1.2. Sum rules

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

$$\int_0^1 dx [u(x, Q^2) - \bar{u}(x, Q^2)] = 2, \quad (72)$$

$$\int_0^1 dx [d(x, Q^2) - \bar{d}(x, Q^2)] = 1, \quad (73)$$

$$\int_0^1 dx [s(x, Q^2) - \bar{s}(x, Q^2)] = 0, \quad (74)$$

1226 thus for the valence distributions we must have $a_f > 0$ for the exponents in (69) or these integrals will diverge. In
 1227 other words, we have the well known result that the xf valence distributions vanish as $x \rightarrow 0$. Although not shown
 1228 explicitly, a similar constraint applies to the heavy quark PDFs as to the strange PDF. In the absence of any intrinsic
 1229 heavy flavour, these are automatically satisfied.

1230 The sum of PDFs must also obey the momentum sum rule

$$\int_0^1 dx x \left(\sum_{n_f} (q(x, Q^2) + \bar{q}(x, Q^2)) + g(x, Q^2) + \dots \right) = 1, \quad (75)$$

1231 which expresses the simple physical requirement that the total proton momentum must be equal to the sum of its
 1232 constituents. The ‘...’ corresponds to any additional parton, for example the photon (or even electroweak bosons),
 1233 which is included within a given a set. Thus for non-valence distributions the exponent a_f may be negative, but must
 1234 be greater than -1 to avoid giving a divergent contribution to the momentum sum rule.

1235 The above 4 sum rules provide additional constraints on the input PDFs, and are typically applied to fix certain
 1236 parameters, for example the overall normalization A_f of a given set. Provided these sum rules are satisfied at the input
 1237 scale, it follows straightforwardly from the form of the DGLAP evolution that they will be satisfied at any other scale
 1238 μ ; the $g \rightarrow q\bar{q}$ splitting can generate no net $q - \bar{q}$ component, and the DGLAP evolution reshuffles the momentum
 1239 carried between the different partons, but of course generates no momentum violation.

1240 4.1.3. Flavour assumptions

1241 For n_f active flavours there are $2n_f + 1$ PDFs, that is 7 if we omit any non-perturbative charm. In addition to
 1242 the gluon, in many cases the remaining 6 light quarks PDFs are not individually parameterised, but rather convenient
 1243 linear combinations, e.g. the valence u_V, d_V distributions, are used. To give one example, MMHT14 take

$$u_V, \quad d_V, \quad \bar{d} - \bar{u}, \quad s + \bar{s}, \quad s - \bar{s}, \quad s + \bar{s} + 2(u + \bar{u} + d + \bar{d}), \quad g. \quad (76)$$

1244 However, as any particular basis can be trivially related to another by a simple linear transformation, the physics
 1245 should not depend on this choice.

1246 Historically the strange quark has been less well determined than the u, d quark PDFs, and indeed in many earlier
 1247 fits this was fixed according to

$$s = \bar{s} \propto \bar{u} + \bar{d}. \quad (77)$$

1248 Such a choice is still taken in the CJ15 and HERAPDF fits, due to the more restricted data set. With the increase in
 1249 available data, the total strangeness $s + \bar{s}$ is now freely parameterised in all global fits. While the sum rule (74) requires
 1250 there to be no overall strangeness in the proton, at a given x value there is no requirement for the $s - \bar{s}$ distribution to
 1251 vanish, and indeed non-perturbative approaches such as the ‘meson cloud model’ [301] predict a strange asymmetry.
 1252 However, the strange difference $s - \bar{s}$ is generally quite poorly determined and still broadly consistent with 0 within
 1253 current uncertainties. Thus will MMHT14 and the latest NNPDF sets fit the strange difference, in all other sets it is
 1254 still assumed that $s = \bar{s}$.

1255 **something about intrinsic charm?? Juan- you mentioned positivity constraints. Could have subsection on**
 1256 **this, maybe more NNPDF related...**

1257 4.2. Data/theory agreement and minimization

1258 4.2.1. Definition of χ^2

1259 The agreement between theory and experimental data or quality of a global fit are usually expressed in term of the
 1260 log-likelihood function or χ^2 . In case that correlations of experimental systematic errors are not available the figure
 1261 of merit χ^2 as a function of the PDF parameters can be expressed as

$$\chi^2(\{a\}) = \sum_{k=1}^{N_{pt}} \frac{1}{\sigma_k^2} (D_k - T_k)^2, \quad (78)$$

1262 where the sum runs over all N_{pt} experimental data points, and σ_k is the total experimental errors by adding the statis-
 1263 tical error and all systematic errors in quadrature. T_k are theoretical predictions that depends on the PDF parameters
 1264 $\{a\}$. D_k are central values of the experimental measurement. In modern experiments there exist statistical errors, un-
 1265 correlated systematic errors and correlated systematic errors, for example, the luminosity error in collider experiments
 1266 are fully correlated among all measurements with the same data sample. The figure of merit χ^2 thus can be written
 1267 as [76]

$$\chi^2(\{a\}, \{\lambda\}) = \sum_{k=1}^{N_{pt}} \frac{1}{s_k^2} \left(D_k - T_k - \sum_{\alpha=1}^{N_l} \beta_{k,\alpha} \lambda_\alpha \right)^2 + \sum_{\alpha=1}^{N_l} \lambda_\alpha^2, \quad (79)$$

1268 where the first sum runs over all N_{pt} experimental data points, and the second sum runs over all the correlated
 1269 systematic errors of N_l sources. s_k represent the total uncorrelated errors constructed by adding the statistical error
 1270 and uncorrelated systematic errors in quadrature. Each source of correlated systematic error is described by a nuisance
 1271 parameter λ_α with the error $\beta_{i,\alpha}$ correlated among all experimental data points $\{i\}$. Thus the induced systematic shift
 1272 to experimental measurement is $\sum_{\alpha} \beta_{k,\alpha} \lambda_\alpha$. The second sum on right side of Eq. (79) are the penalty terms to χ^2 by
 1273 simply assuming standard normal distributions of the nuisance parameter.

1274 In the global analyses we are more interested on the PDF parameters than those nuisance parameters. For any
 1275 PDF parameter $\{a\}$ one can minimize the χ^2 with respect to the nuisance parameter λ_α first that gives the profiled
 1276 log-likelihood function $\chi^2(\{a\}) \equiv \chi^2(\{a\}, \{\hat{\lambda}\})$. The profiled nuisance parameter $\hat{\lambda}_\alpha$ can be solved analytically in above
 1277 assumed Gaussian case since the dependence of χ^2 on λ_α has a simple quadratic form. Explicitly, one have

$$\hat{\lambda}_\alpha = \sum_{i=1}^{N_{pt}} \frac{D_i - T_i}{s_i} \sum_{\delta=1}^{N_l} A_{\alpha\delta}^{-1} \frac{\beta_{i,\delta}}{s_i}, \quad (80)$$

1278 with

$$A_{\alpha\beta} = \delta_{\alpha\beta} + \sum_{k=1}^{N_{pt}} \frac{\beta_{k,\alpha} \beta_{k,\beta}}{s_i}. \quad (81)$$

1279 By substituting $\hat{\lambda}_\alpha$ into the χ^2 one further obtain the profiled χ^2 as a function of the PDF parameters,

$$\chi^2(\{a\}) = \sum_{i,j=1}^{N_{pt}} (T_i - D_i) (\text{cov}^{-1})_{ij} (T_j - D_j), \quad (82)$$

1280 with the covariance matrix and its inverse given by

$$(\text{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_i} \beta_{i,\alpha} \beta_{j,\alpha}, \quad (\text{cov}^{-1})_{ij} = \frac{\delta_{ij}}{s_i^2} - \sum_{\alpha,\beta=1}^{N_i} \frac{\beta_{i,\alpha}}{s_i^2} A_{\alpha\beta}^{-1} \frac{\beta_{j,\beta}}{s_j^2}. \quad (83)$$

1281 Namely the profiled χ^2 is fully determined by the covariance matrix that constructed from the experimental statistical
 1282 and systematic errors. In certain circumstance, e.g., most of the measurements at LHCb, the experiments only publish
 1283 the covariance matrix instead of a full breakdown of the experimental systematics.

1284 There is one subtlety concerning the construction of the covariance matrix due to the fact that experimental sys-
 1285 tematic errors are usually presented in relative errors $\sigma_{i,\alpha}$,

$$(\text{cov})_{ij} = s_i^2 \delta_{ij} + \left(\sum_{\alpha=1}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} + \sum_{\alpha=1}^{N_L} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} \right) D_i D_j. \quad (84)$$

1286 The correlated systematic errors are further separated into additive errors (N_c in total) and multiplicative errors (N_L
 1287 in total), like the luminosity uncertainties. They have different statistical interpretations though all normalized to the
 1288 experimental central values. It is noted that above *experimental* definition of the covariance matrix will result in a
 1289 D'Agostini bias of the multiplicative errors [302] when using in the actual PDF fit. The correct way to handle the
 1290 multiplicative errors in actual fit is to use the following so-called “ t_0 ” definition of the covariance matrix,

$$(\text{cov})_{ij} = s_i^2 \delta_{ij} + \left(\sum_{\alpha=1}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} D_i D_j + \sum_{\alpha=1}^{N_L} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} T_i^0 T_j^0 \right), \quad (85)$$

1291 namely rescaling the multiplicative errors by a specific theory prediction T_i^0 , the prediction from last iteration of the
 1292 χ^2 minimization. Alternative prescriptions could be the t definition [4] of rescaling the multiplicative errors to the
 1293 same theoretical prediction as in the comparison to data,

$$(\text{cov})_{ij} = s_i^2 \delta_{ij} + \left(\sum_{\alpha=1}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} D_i D_j + \sum_{\alpha=1}^{N_L} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} T_i T_j \right), \quad (86)$$

1294 or the *extended- t_0* and *extended- t* definitions of rescaling both additive and multiplicative errors to the corresponding
 1295 theory. A detailed discussion of various possible prescriptions can be found in [4?].

1296 4.2.2. Minimization of χ^2

1297 The best-fit or central value of the fit of the PDF parameters are determined by minimization of the log-likelihood
 1298 function $\chi^2(\{a\})$. Modern PDF analysis usually involves a large amount of free PDF parameters, e.g., ~ 10 – 40 for PDF
 1299 groups with traditional analytical parametrization, and a few hundreds in case of parametrization with neural networks.
 1300 In the first case of minimization with moderate dimensions, the gradient based algorithms can be very efficient. For
 1301 example, in Newton's method, the trial solution of the global minimum is given by $a_0 - Vd$, where a_0 is an arbitrary
 1302 starting point in the PDF parameter space, d and V are gradient and inverse of the Hessian matrix at the same point.
 1303 The solution is exact assuming a quadratic shape of the χ^2 . In practice the χ^2 can deviate significantly from a parabola
 1304 when it is far away from the global minimum. Above procedure must be applied iteratively till it converges. Besides,
 1305 the method fails if matrix V is not positive-defined and suffers from numerical instabilities. Various quasi-Newton
 1306 methods have been proposed to overcome those complexities in real applications, e.g., Levenberg-Marquardt method
 1307 as used in MSTW/MMHT analyses [80], which is essentially a dynamic mixture of Newton method and steepest
 1308 decent method. For PDF analysis with gradient based methods, usually the gradient and Hessian matrix can only be
 1309 calculated numerically by means of finite differences. There is another class of widely used gradient based algorithms,
 1310 variable metric methods (VMM). In VMM it is not necessary to calculate the Hessian matrix numerically which is
 1311 time consuming. Instead the matrix V is updated iteratively based only on information of the gradients. The VMM
 1312 is the default algorithm in the MINUIT package [?] and is used for CTEQ-TEA analyses [303]. On another hand,
 1313 in case of a few hundreds of free parameters, obviously the gradient based method is inappropriate which will suffer

1314 greatly from numerical instability and local minimums. The genetic algorithm turns out to be a proper choice as
 1315 demonstrated in the NNPDF analysis [88]. The basic idea is to start from an ensemble of arbitrary chosen samples of
 1316 the PDF parameters. Random mutations with possible crossing-overs are applied to generate a larger group of new
 1317 samples. Those candidates predicted lower χ^2 are then selected to form a new ensemble with the same size. Similar
 1318 as before the procedure is iterated until a suitable convergence criterion is met and meanwhile preventing overfitting.

1319 4.3. PDF uncertainties

1320 For a long time, PDF fits were performed without a systematic assessment of the associated uncertainties. At most,
 1321 variants of the central fit with different settings were provided, and the comparison between the results of different
 1322 PDF groups were used as an estimate of the intrinsic PDF uncertainties. This situation was however clearly far from
 1323 satisfactory, as exemplified by two famous anomalies, the high E_T jets at the Tevatron and the leptoquarks anomaly
 1324 at HERA, both of which attracted quite some interest as possible indications of new physics Beyond the Standard
 1325 Model, while on the other hand it was shown that these anomalies reflected limitations of our modeling of the internal
 1326 structure of the proton.

1327 With this motivation, around 2000 a number of methods were devoted to determine in a systematic way the
 1328 uncertainties associated to a PDF fit, and moreover to allow these uncertainties to be propagated to phenomenological
 1329 cross-sections. These methods fall into three main classes, although within each method there are variations. The
 1330 first one is the *Hessian method*, which is based on the parabolic expansion of the χ^2 in the vicinity of the minimum,
 1331 parametrized by a number of orthogonal eigenvectors within some fixed tolerance. The second technique is the *Monte*
 1332 *Carlo method*, where a representation of the probability distribution of the experimental data is constructed by beams
 1333 of a large number of *replicas*. PDF fits are then performed separately on each of these Monte Carlo replicas, and the
 1334 resulting ensemble of PDFs represents a probability density in the space of parton distributions. The third technique
 1335 is known as the *Lagrange multiplier method*.

1336 4.3.1. The Hessian method

1337 The Hessian method to quantify PDF uncertainties was first developed in [75]. We describe now the basic in-
 1338 gredients of this method and of their subsequent refinements, following mostly the discussion of Ref. [80]. Given
 1339 the χ^2 estimator that quantifies that data/theory agreement, the best-fit values corresponds to the those for which this
 1340 estimator has an absolute minimum, χ^2_{\min} . In the vicinity of this minimum, the χ^2 can be approximated in terms of a
 1341 quadratic expansion of the form

$$\Delta\chi^2 \equiv \chi^2 - \chi^2_{\min} = \sum_{i,j=1}^{n_{\text{par}}} H_{ij} (a_i - a_i^0) (a_j - a_j^0), \quad (87)$$

1342 where the n_{par} fit parameters are denoted by $\{a_1, \dots, a_{n_{\text{par}}}\}$, and the best-fit values that minimize the χ^2 are indicated
 1343 by $\{a_1^0, \dots, a_{n_{\text{par}}}^0\}$. In the quadratic expansion Eq. (87), we have introduced the Hessian matrix, defined as the matrix
 1344 of second derivatives of the χ^2 with respect to the fit parameters, namely

$$H_{ij} \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} \Big|_{\{\vec{a}\}=\{\vec{a}^0\}}. \quad (88)$$

1345 This Hessian matrix contains all the information necessary to quantify the PDF. Indeed, for a generic function $\langle \mathcal{F}[\{a_i\}] \rangle$
 1346 that depends on the PDFs and thus indirectly on the fit parameters, the associated uncertainty can be computed by
 1347 means of linear error propagation

$$\sigma_{\mathcal{F}} = T \left(\sum_{i,j}^{n_{\text{par}}} \frac{\partial \mathcal{F}}{\partial a_i} (H)^{-1} \frac{\partial \mathcal{F}}{\partial a_j} \right)^{1/2}, \quad (89)$$

1348 where $T = \sqrt{\Delta\chi^2}$ the *tolerance factor* that determine the matching between the allowed range of parameter variations
 1349 around the best-fit values and the associated confidence interval of the PDF uncertainties. While textbook statistics
 1350 suggest that $T = 1$ corresponds to a 68% confidence interval, in the context of a global fit there is ample evidence that

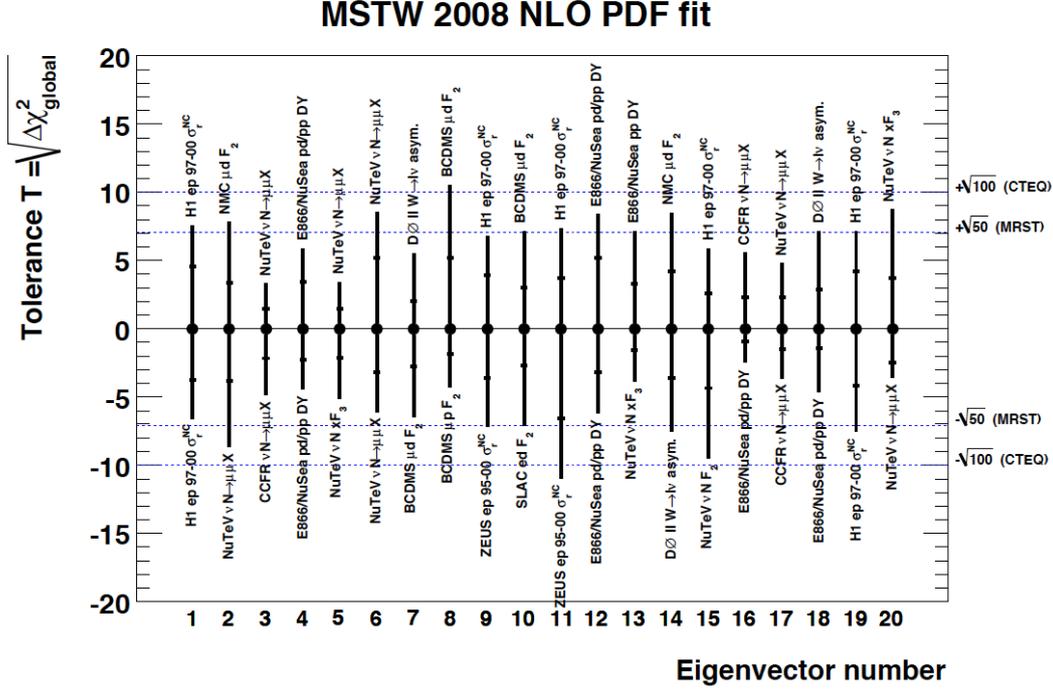


Figure 25: The individual tolerance for each eigenvector, determined by the criterion that each separate experiment should be described within 90% CL. In each case the figure indicates the name of the experiment that determines the tolerance for the various eigenvector directions.

1351 somewhat larger values for the tolerance are required in the Hessian method, in particular to account for inconsistent
 1352 experiments and for methodological uncertainties such as the specific choice of functional form [299].

1353 The main limitation of Eq. (89) is that in general the derivatives $\partial\mathcal{F}/\partial a_i$ are unknown. This problem can be
 1354 bypassed by diagonalizing the Hessian matrix and then representing PDF uncertainties in terms of orthogonal eigen-
 1355 values. After this diagonalization procedure, Eq. (89) has the simpler form

$$\sigma_{\mathcal{F}} = \frac{1}{2} \left(\sum_{i,j}^{n_{\text{par}}} [\mathcal{F}(S_i^+) - \mathcal{F}(S_i^-)] \right)^{1/2}, \quad (90)$$

1356 where S_i^{\pm} correspond to the i -th eigenvector associated to positive and negative variations with respect to the best fit
 1357 value. Using the eigenvectors $\{S_i^{\pm}\}$ it's also possible to compute asymmetric PDF uncertainties using the prescription
 1358 of Ref. [304].

1359 Concerning the determination of the tolerance factor $T = \sqrt{\Delta\chi^2}$, the original studies by the CTEQ and MRST
 1360 group used values of $T = 10$ and $T = \sqrt{50}$ respectively. In more recent releases, the determination of this tolerance
 1361 has been refined. In the case of the MSTW08 analyses for example, the tolerance is determined dynamically for each
 1362 eigenvector demanding that all data sets are included within the 68% confidence level variation. To illustrate this,
 1363 in Fig. 25 we show the individual tolerance for each eigenvector of the MSTW08 global analysis, determined by the
 1364 criterion that each separate experiment should be described within 90% CL. For the various eigenvector directions,
 1365 the figure indicates the name of the experiment that determines the tolerance. The fact that many different experiments
 1366 are responsible for determining this tolerances emphasizing the crucial importance of using a very wide dataset in the
 1367 global PDF analysis.

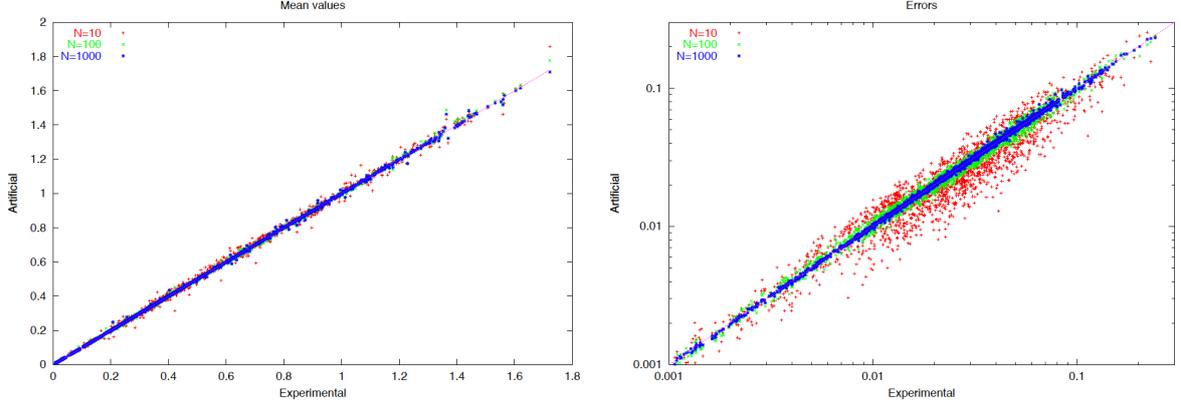


Figure 26: The scatter between the mean values (left) and variances (right plot) of all the data points included in the analysis of [306], comparing the original experimental values with the results obtained from the MC representation for different number N_{rep} of replicas.

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_{rep} of artificial replicas of the original data (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 $F_I^{(\text{exp})}$ with total uncorrelated uncertainty $\sigma_I^{(\text{stat})}$, N_{sys} fully correlated systematic uncertainties $\sigma_{I,c}^{(\text{corr})}$ and N_a (N_r) absolute (relative) normalization uncertainties $\sigma_{I,n}^{(\text{norm})}$, the artificial MC replicas are constructed using the following expression

$$F_I^{(\text{art})(k)} = S_{I,N}^{(k)} F_I^{(\text{exp})} \left(1 + \sum_{c=1}^{N_{\text{sys}}} r_{I,c}^{(k)} \sigma_{I,c}^{(\text{corr})} + r_I^{(k)} \sigma_I^{(\text{stat})} \right), \quad k = 1, \dots, N_{\text{rep}}, \quad (91)$$

where the normalization prefactor is given by

$$S_{I,N}^{(k)} = \prod_{n=1}^{N_a} (1 + r_{I,n}^{(k)} \sigma_{I,n}^{(\text{norm})}) \prod_{n'=1}^{N_r} \sqrt{1 + r_{I,n'}^{(k)} \sigma_{I,n'}^{(\text{norm})}}. \quad (92)$$

Here the variables $r_{I,c}^{(k)}$, $r_I^{(k)}$, $r_{p,n}^{(k)}$ are univariate gaussian random numbers. Eq. (91) 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, $r_{I,c}^{(k)} = r_{I',c}^{(k)}$. The same condition holds for the normalization uncertainties.

An important question in the Monte Carlo method is how many replicas N_{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. [305]. It was found that $N_{\text{rep}} = 10$ replicas are enough to reproduce central values, $N_{\text{rep}} = 100$ for the variances and that $N_{\text{rep}} = 1000$ replicas are required to satisfactory 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 [306], comparing the original experimental values with the results obtained from the MC representation for different number N_{rep} of replicas. We find that indeed for central values, the averages computed using only $N_{\text{rep}} = 10$ replicas agree with the original data, but that for variances this is not the case, and $N_{\text{rep}} = 100$ replicas are required in this case.

Once the Monte Carlo sampling of the experimental data has been performed, a separate PDF fit is performed in each replica. Note that this can be done using traditional polynomial functional forms or other interpolators such as

1393 artificial neural networks. The resulting sample of N_{rep} PDF replicas realizes the concept of the probability density in
 1394 the space of parton distributions. The calculation of the resulting PDF uncertainties and their propagation to generic
 1395 cross-sections can be performed using textbook methods. Note that in this approach PDF uncertainty propagation
 1396 is fully general, and in particular is not restricted to the Gaussian approximation. For instance, in the Monte Carlo
 1397 method the expectation function of a generic cross-section $\mathcal{F}[\{q\}]$ is evaluated as an average over the replica sample,

$$\langle \mathcal{F}[\{q\}] \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{set}}} \mathcal{F}[\{q^{(k)}\}], \quad (93)$$

1398 and the corresponding uncertainty is then determined as the variance of the Monte Carlo sample,

$$\sigma_{\mathcal{F}} = \left(\frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} (\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle)^2 \right)^{1/2}. \quad (94)$$

1399 These formulae may also be used for the determination of central values and uncertainties of the parton distribution
 1400 themselves, in which case the functional \mathcal{F} is identified with the parton distribution q : $\mathcal{F}[\{q\}] \equiv q$.

1401 In the case of a fully consistent dataset, the Monte Carlo method to estimate the PDF uncertainties is expected
 1402 to coincide with the Hessian method described in Sect. 4.3.1 for a standard tolerance $\Delta\chi^2 = 1$. This equivalence
 1403 was explicitly demonstrated in the HERA-LHC workshop proceedings [307]. In Fig. 27 we show the gluon PDF at
 1404 $Q = 2$ GeV in this HERA-LHC benchmark fit, based on HERA inclusive structure function data, where the one-sigma
 1405 PDF uncertainties computed with the Hessian method (black lines) are compared to those of the Monte Carlo method
 1406 (right lines), finding good agreement. In this figure, each of the green curves corresponds to an individual MC replica.
 1407 In the left fit, the normalization and systematic uncertainties in the MC replicas from Eq. (91) fluctuate according
 1408 to a multi-Gaussian distribution, while in the right fit they fluctuate instead according to a log-normal distribution,
 1409 showing that the PDFs depend only weakly on the specific assumptions about the specific probability distribution of
 1410 the experimental systematic uncertainties.

1411 Finally, we want to point out that a Hessian representation of a Monte Carlo PDF sets can be accurately constructed
 1412 using the `mc2h` algorithm developed in Ref. [308]. This technique is discussed in more detail in Sect 4.5.

1413 4.3.3. The Lagrange multiplier method

1414 The Lagrange multiplier method was originally developed in Ref. [309, 303]. This method is a generalization of
 1415 the χ^2 minimization procedure. As in the Hessian case, the first step is to find the PDF parameters $\{a_i^0\}$ that minimize
 1416 the global $\chi^2(\{a_i\})$. Then one has to select a specific physical quantity that depends on the PDFs, such a DIS structure
 1417 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
 1418 global fit minimum. The goal of the Lagrange multiplier method is to determine the PDF uncertainty associated to \mathcal{F}_0
 1419 without making any assumption on the specific behaviour of the χ^2 around the global minimum.

1420 In order to achieve this, the global fit χ^2 is modified by introducing the physical quantity \mathcal{F} as a Lagrange multi-
 1421 plier, so that the new function that needs to be minimized is now given by

$$\Psi(\lambda, \{a_i\}) = \chi^2(\{a_i\}) + \lambda \mathcal{F}(\{a_i\}). \quad (95)$$

1422 Now for each specific value of λ , denoted by λ_α , the minimization of Eq. (95) will lead to a different set of best-fit
 1423 PDF parameters, which we indicate by $\{a_i^{(\min)}(\lambda_\alpha)\}$. Mathematically, these parameters are the result of a constrained
 1424 PDF fit where the value of the physical observable has been fixed to $\mathcal{F}_\alpha = \mathcal{F}(\{a_i^{(\min)}(\lambda_\alpha)\})$. The resulting PDF set of
 1425 this constrained fit is now indicated by S_α .

1426 The main result of this procedure is establishing a parametric relation between the value of the physical quantity
 1427 \mathcal{F} and the global fit χ^2 by means of the Lagrange multiplier λ . This means that we can determine the PDF uncertainty
 1428 associated to \mathcal{F} imposing that the χ^2 satisfies $\chi^2 = \chi_{\text{min}}^2 + \Delta\chi^2$ with $\Delta\chi^2 = T^2$ representing the tolerance, as introduced
 1429 in the previous section. It is clear that the main advantage of the Lagrange multiplier method as compared to the Hessian
 1430 method is that one does not need to restrict to the quadratic expansion nor in linear error propagation, since the PDF
 1431 uncertainties in the LM are determined only by the values of the χ^2 but never by its specific shape. On the other hand,
 1432 an important restriction of the method is that the PDF error analysis for each specific physical quantity \mathcal{F} requires

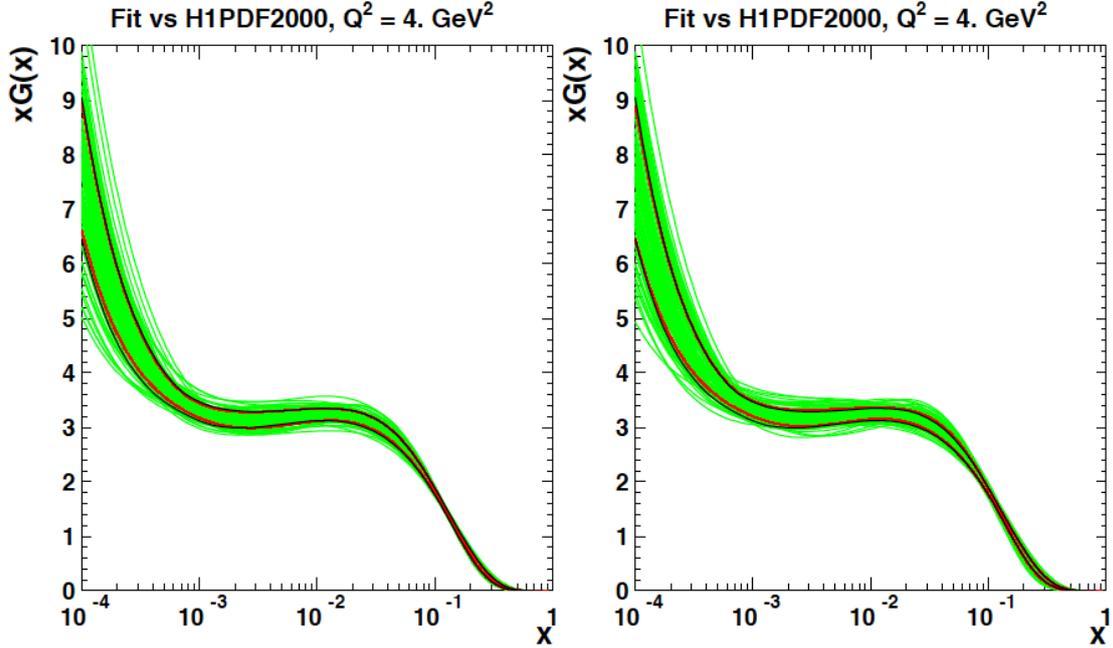


Figure 27: The gluon PDF at $Q = 2$ GeV in the HERA-LHC benchmark fit of Ref. [307], where the one-sigma PDF uncertainties computed with the Hessian method (black lines) are compared to those of the Monte Carlo method (right lines), finding good agreement. Each of the green curves corresponds to an individual MC replica. In the left fit, the normalization and systematic uncertainties in the MC replicas from Eq. (91) fluctuate according to a multi-Gaussian distribution, while in the right fit they fluctuate instead according to a log-normal distribution.

1433 redoing a large number of new PDF fits, and this is not only very CPU time intensive but also it cannot be done outside
 1434 the PDF fitting collaboration.

1435 The Lagrange multiplier method is schematically illustrated in Fig. 28. In the left plot we show a two-dimensional
 1436 projection of the PDF parameter space, indicating the contours in χ^2 for fixed values of the physical quantity \mathcal{F} . The
 1437 parametric relation is provided by the value of multiplier λ . In the right plot we show how the PDF uncertainty
 1438 associated to \mathcal{F} for a given confidence interval is determined by the condition that the global χ^2 should not grow
 1439 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
 1440 input to the method and must be determined independently.

1441 4.4. Treatment of theory parametric uncertainties

1442 Treatment of PDF theory parametric uncertainties due to the strong coupling, the heavy quark masses etc

1443 4.5. Combined and reduced PDF sets

1444 Jun

1445 Individual PDF sets from different groups are widely used when testing precision theoretical predictions with the
 1446 LHC measurements and in assessment of the accuracy of PDF sets themselves. However, in many LHC applications
 1447 it requires assessment of the *total* PDF uncertainties on certain observables by taking into account predictions from all
 1448 applicable PDF sets. That can be the case for example in extraction of the couplings of the Higgs boson, or calculations
 1449 of signal and background rates in searches of new physics beyond the standard model and so on. For the purpose a
 1450 statistical procedure is needed to combine results from all those PDF sets. Such a statistical combination is much more
 1451 complicated than for example the world average of strong coupling constant or heavy-quark masses, since it combines
 1452 functions which have infinite number of degrees of freedom in principle. The prescription must accommodate the

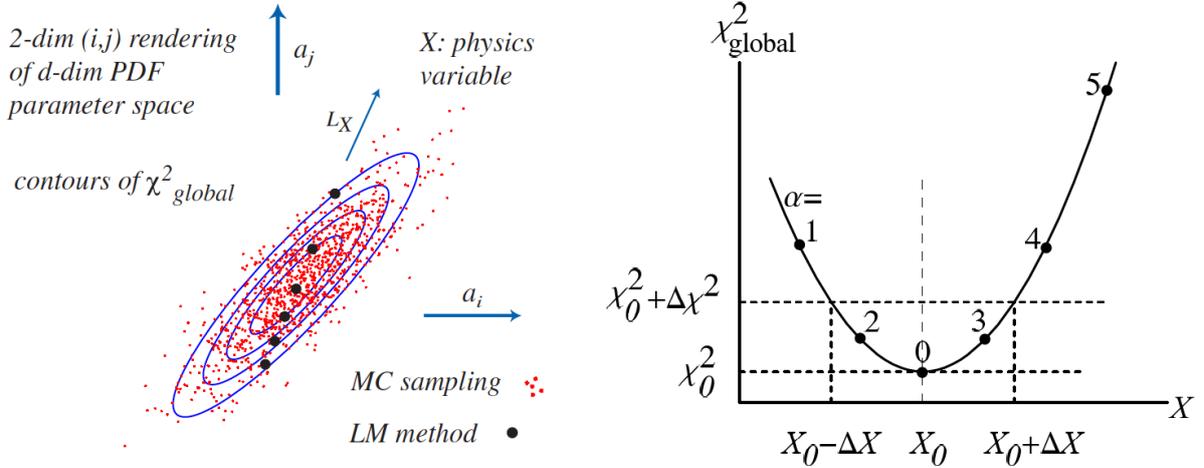


Figure 28: Schematic representation of the Lagrange Multiplier method. In the left plot we show a two-dimensional projection of the PDF parameter space, indicating the contours in χ^2 for fixed values of the physical quantity \mathcal{F} . In the right plot we show how the PDF uncertainty associated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

1453 fact that the individual PDF sets are not identical either in their central values or in their uncertainties. It should also
 1454 account for possible correlations between PDF sets from different groups. Besides it will be desirable to maintain a
 1455 compact set of PDFs in the final combination.

1456 The 2010 PDF4LHC recommendation proposed to use an envelope prescription [5, 4, 91]. Namely PDF deter-
 1457 minations from different groups are treated as instances of a probability distribution affected by unknown sources of
 1458 systematics rather than statistically distributed instances of an underlying probability distribution. Also the envelope
 1459 prescription can only be applied at the level of individual observables without information of PDF induced correla-
 1460 tions. Such a prescription is certainly inadequate now given better understanding on current PDF determinations,
 1461 their good agreements, and the high precision demanded for LHC Run II study. The 2015 PDF4LHC recommenda-
 1462 tion came out as a replacement [2]. There are certain criteria for the individual PDF sets to be considered for the
 1463 combination. First the individual PDF sets should be based on a global determination with large number of datasets
 1464 from variety of experiments, e.g., DIS and hadron-hadron scattering in fixed-target and collider experiments. The
 1465 hard cross sections for DIS and hadron-hadron scattering processes used in the extraction should be evaluated up to
 1466 two loops in QCD in a GMVFN scheme with a maximum number of active quark flavors of 5. Besides, all known
 1467 experimental and procedural sources of uncertainties should be properly accounted for, which includes experimental
 1468 uncertainties propagated from data, uncertainties due to incompatibility of different data sets, and uncertainties due
 1469 to the functional form of PDFs, etc. It is decided that the combination should be carried out with a central value of
 1470 $\alpha_s(M_Z) = 0.118$ at both NNLO and NLO and with the uncertainties of $\alpha_s(M_Z)$ taken to be 0.0015, consistent with the
 1471 PDG world-average [310]. The heavy-quark masses used in individual PDFs sets are not necessary to be the same at
 1472 current stage⁵ but should be compatible with their world-average values. The existing PDF sets satisfying all above
 1473 requirements at present have been identified as CT14 [18], MMHT2014 [19], and NNPDF3.0 [17]. The PDF4LHC
 1474 2015 PDF sets are thus statistical combination of those three global analyses.

1475 The combination can be only carried out efficiently using the Monte-Carlo method since different PDF determina-
 1476 tions adopt different from of PDF parametrizations. In the first step the CT14 and MMHT2014 PDFs as originally in
 1477 Hessian form are converted into their Monte-Carlo representations based on the Watt-Thorne method with symmetric
 1478 formula [311]. It has been validated that a MC ensemble with 300 replicas is sufficient to reproduce the central and

⁵It will be desirable in the future all PDF groups provide error sets with common choices of heavy-quark masses and further include the uncertainties due to the mass inputs similar as for the α_s .

1479 unceratinties of the original Hessian PDF to very good extend. The NNP3.0 is already in a Monte-Carlo form with
 1480 1000 repicas. Following the idea of individual PDF determinations are equally likely representations of an underly-
 1481 ing probability distribution, a combined PDF set is build by taking 300 MC replicas from each input PDF sets and
 1482 merging them equally. The resulting combined PDF set, an ensemble of 900 MC replicas, is referred to as MC900
 1483 or PDF4LHC15_prior, and represents the combined probability distribution of the PDFs. Meanwhile, as mentioned
 1484 earlier, such a large set with 900 PDFs would be unmanageable in most of the applications especially considering the
 1485 time and storage cost for complicated NNLO calculations and experimental simulations. Various methods have been
 1486 applied to reduce the size of the combined set while minimizing information loss on statistical measures in the large
 1487 ensemble, which lead to the reduced PDF sets.

1488 The first method is to use the META-PDFs framework [312]. A flexible functional form with Bernstein polyno-
 1489 mials is chosen to parametrize PDFs at an initial scale. Each replica in the MC900 ensemble is then represented by a
 1490 group of PDF parameters through fit to the chosen parametrizations by minimizing a metric function. The prior prob-
 1491 ability distribution of PDFs thus transforms to probability distributions in the PDF parameter space. The covariance
 1492 matrix of the PDF parameters is calculated,

$$cov(a_l, a_m) = \frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} (a_l^{(k)} - a_l^{(0)})(a_m^{(k)} - a_m^{(0)}), \quad (96)$$

1493 where $a_l^{(0,k)}$ denotes the fitted PDF parameters from the central set and the k -th MC replicas, N_{rep} is the total number
 1494 of MC replicas. The covariance matrix can be diagonalized by an orthogonal transformation. Eigenvectors are found
 1495 and ordered according to their impact on the PDF unceratinties with a designed error metric. The eigenvectors with
 1496 smaller contributions can be dropped according to the accuracy required. Finally a central PDF set and a group of
 1497 orthogonal error PDF sets are generated with assumption of a multi-Gaussian distribution. They can used in a similar
 1498 way as the conventional Hessian PDF set. E.g., the 68% cl uncertainty or 1σ error is given by

$$\delta^{\text{PDF}} X = \sqrt{\sum_{i=1}^{N_{\text{eig}}} (X_i - X_0)^2}, \quad (97)$$

1499 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.
 1500 Note there is only one error set along each eigenvector/orthogonal direction since symmetric Gaussian distributions
 1501 are assumed in this case. The second method is to use the MC2Hessian algorithm with Singular Value Decomposition
 1502 followed by Principle Component Analysis [308]. The idea is to first discretize the PDFs with $N_x N_{\text{pdf}}$ observables
 1503 which are the PDF values at the corresponding grid point. N_x denotes the total number of grid points in momentum
 1504 fraction x . 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
 1505 values can be constructed from all the MC replicas,

$$cov_{ll'} = \frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} X_{lk} X_{k'l'}^T = \frac{1}{N_{\text{rep}} - 1} X X^T \quad (98)$$

1506 where X_{lk} is the PDF value on l -th grid point given by k -th MC replica subtracted with the corresponding value from
 1507 central PDF set, N_{rep} is the total number of MC replicas. The above covariance matrix can be rewritten in its singular
 1508 value decomposition form

$$cov_{ll'} = \frac{1}{N_{\text{rep}} - 1} (U S V^T)(U S V^T)^T, \quad (99)$$

1509 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
 1510 coefficients, and U is a $N_x N_{\text{pdf}} \times N_{\text{rep}}$ matrix containing orthogonal eigenvecgtors of the covariance matrix with nonzero
 1511 eigenvalues. Indeed the matrix V gives a Hessian basis bulit upon linear combinations of original MC replicas, which
 1512 reproduces fully the covariance matrix given by original MC replicas. This basis can be further truncated using
 1513 Principle Component Analysis with certain error matric which results in a smaller Hessian PDF set similar to the
 1514 case of META-PDFs. A third option is provided by the compressed Monte-Carlo (CMC) method [313]. In this
 1515 case an ensemble of pseudo-MC replica PDFs (CMC-PDFs) are generated. The CMC-PDFs have different statistical

1516 interpretation as the native MC PDFs. However, certain statistical measures, e.g., mean, covariance matrix, skewness,
 1517 kurtosis and the Kilmoforov distance can be reconstructed in a similar way as in the native MC PDFs. The CMC-PDFs
 1518 are aiming at preserving some of the non-Gaussian features in the prior give by MC900 in addition to the Gaussian
 1519 approximation for which the Hessian form is more adequate. Note in current prescription some of the non-Gaussian
 1520 behaviors from individual PDF sets have been smoothed out due to the symmetric formula used in converting Hessian
 1521 PDFs to MC replicas. The compression starts with a figure of merit,

$$ERF = \sum_k \frac{1}{N_k} \sum_i \left(\frac{C_i^{(k)} - O_i^{(k)}}{O_i^{(k)}} \right), \quad (100)$$

1522 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
 1523 estimator calculated at the generic point (x_i, Q_i) from the prior and $C_i^{(k)}$ is the corresponding value of the same estimator
 1524 in the compressed set. The compressed set is simply a subset of the MC900 ensemble. For any given number of total
 1525 MC replicas, the compressed set is chosen by minimization of above error function using a generic algorithm.

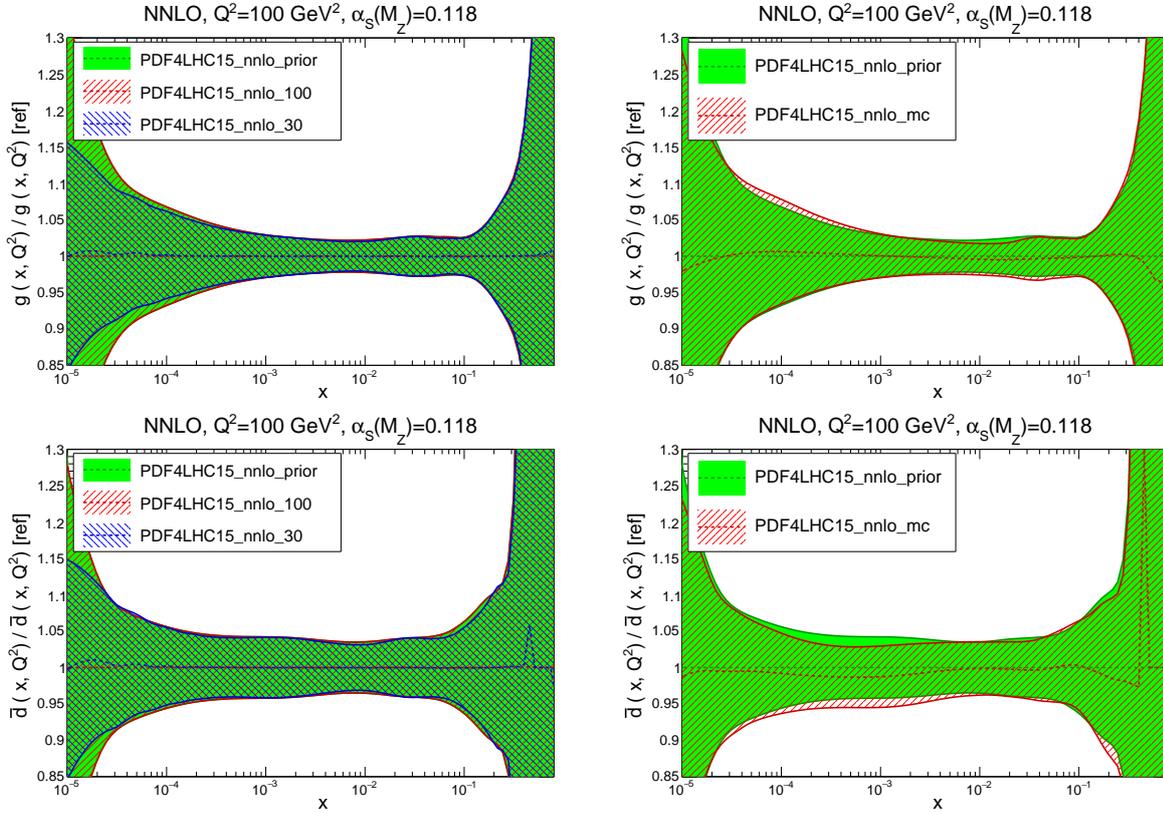


Figure 29: Comparison of the gluon and \bar{d} -quark PDFs at a scale of $Q^2 = 100 \text{ GeV}^2$ between the prior and the two reduced Hessian sets, and between the prior and the compressed MC set, normalized to the central value of the prior [2].

1526 Finally in the 2015 PDF4LHC recommendation there are three reduced PDF sets provided to public, a Hessian
 1527 set with 30 error PDFs (PDF4LHC15_30), a Hessian set with 100 error PDFs (PDF4LHC15_100), and a compressed
 1528 MC set with 100 replicas (PDF4LHC15_mc). All of them are constructed from the same prior (MC900) with slightly
 1529 different focus. The symmetric PDF uncertainties of any observables can be calculated using Eq. (97) for Hessian sets
 1530 and the usual master formula for MC PDFs. Figure 29 shows the comparison of the central/mean and the uncertainties
 1531 of the gluon and \bar{d} quark PDFs from the prior and the three reduced sets. The agreement between the Hessian set with

1532 100 eigenvectors and the prior is good for all PDF combinations in the complete range of x . The Hessian set with 30
 1533 eigenvectors also shows good agreement with the prior in the x range related to precision physics measurements, but
 1534 gives slightly smaller uncertainty in the extrapolation regions at small- and large- x as a tradeoff of fewer error PDFs.
 1535 The compressed MC set also agrees well with the prior in most of the region for mean and uncertainty except for
 1536 small overall fluctuations. The PDF4LHC recommendation also suggests usage of different PDF4LHC15 sets,

- 1537 • Use individual PDF sets, and, in particular, as many of the modern PDF sets as possible from comparison
 1538 between data and theory for standard model measurements.
- 1539 • Use the PDF4LHC15_mc sets for searches for phenomena beyond the standard model where the non-Gaussian
 1540 behaviors could be important.
- 1541 • Use the PDF4LHC15_30 sets for calculation of PDF uncertainties in situations when computational speed is
 1542 needed, or a more limited number of error PDFs may be desirable.
- 1543 • Use the PDF4LHC15_100 sets for calculation of PDF uncertainties in precision observables.

1544 The cases listed above are not exclusive: one or the other should be more adequate depending on the theoretical
 1545 interpretation of a given experimental measurement. Besides, there are two additional PDFs with $\alpha_s(M_Z) = 0.1165$
 1546 and 0.1195 in the PDF4LHC15 sets for estimation of the uncertainty due to α_s input. The corresponding uncertainty
 1547 at 68% cl for the observable X is given by

$$\delta^{\alpha_s} X = \frac{X(\alpha_s = 0.1195) - X(\alpha_s = 0.1165)}{2}, \quad (101)$$

1548 where $X(\alpha_s)$ is the calculated value by using the PDF together with the hard matrix elements evaluated at that α_s
 1549 value. The combined PDF+ α_s uncertainty is then computed as follows

$$\delta^{\text{PDF}+\alpha_s} X = \sqrt{(\delta^{\text{PDF}} X)^2 + (r \cdot \delta^{\alpha_s} X)^2}, \quad (102)$$

1550 where the rescaling factor $r = 1$ as recommended but can be varied according to user's choice on uncertainty of
 1551 $\alpha_s(M_Z)$.

1552 It is also noted that the PDF4LHC15 PDF sets can be further reduced to compact sets with around ten eigenvectors
 1553 or less if the applications are restricted to a certain group of observables, e.g., the cross sections and distributions in
 1554 Higgs boson production at the LHC. That can be achieved either through the data set diagonalization method [314,
 1555 315] or the singular value decomposition method [316].

1556 4.6. Approximate methods

1557 In this subsection we discuss approximate methods that can in certain circumstances be used instead of a full
 1558 fledged PDF fit: the Bayesian reweighting of Monte Carlo replicas [317, 318] and the profiling of Hessian sets [319].

1559 4.6.1. Hessian profiling

1560 The impact that a new experimental measurement has when added to a Hessian PDF set can be estimated by means
 1561 of the so-called profiling technique. This can be achieved by minimizing a χ^2 estimator that compares the theoretical
 1562 predictions obtained with a given input Hessian PDF set with the new experimental measurements. This estimator
 1563 takes into account both the experimental uncertainties and the effects from the PDF variations (Hessian eigenvectors)
 1564 and is defined by

$$\chi^2(\beta_{\text{exp}}, \beta_{\text{th}}) = \sum_{i=1}^{N_{\text{data}}} \left(\sigma_i^{\text{exp}} + \sum_j \Gamma_{ij}^{\text{exp}} \beta_{j,\text{exp}} - \sigma_i^{\text{th}} + \sum_k \Gamma_{ik}^{\text{th}} \beta_{k,\text{exp}} \right)^2 / \Delta_i^2 + \sum_j \beta_{j,\text{exp}}^2 + \sum_k \beta_{k,\text{th}}^2, \quad (103)$$

1565 where $\beta_{j,\text{exp}}$ are the nuisance parameters corresponding to the set of fully correlated experimental systematic uncer-
 1566 tainties, and $\beta_{k,\text{th}}$ are the nuisance parameters associated to the PDF eigenvectors. In this equation, Δ_i stands for the
 1567 total experimental uncorrelated uncertainty, and N_{dat} is the number of data points of the measurement which is being

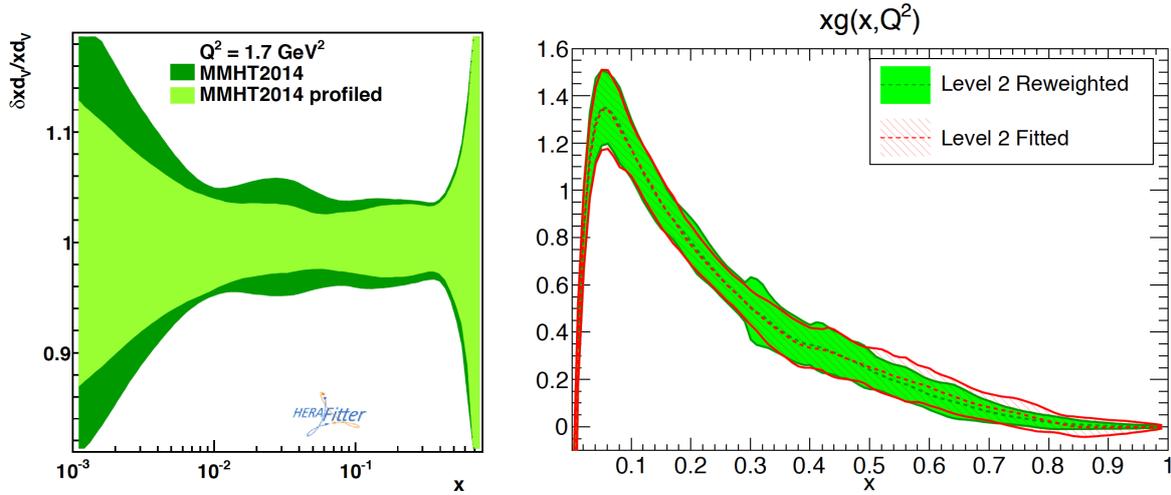


Figure 30: Left plot: the impact of the Tevatron W and Z data on the MMHT2014 NLO fit, estimated by the Hessian profiling method in [319]. Right plot: the gluon PDF in the NNPDF3.0 closure tests, estimating the impact of the collider inclusive jet data and comparing the results of the Bayesian reweighting with those of a direct refit.

1568 added into the PDF fit. The matrices Γ_{ij}^{exp} and Γ_{ik}^{th} encode the influence of the corresponding nuisance parameters on
 1569 the experimental data and on the theory predictions respectively.

1570 Upon minimization of this χ^2 estimator, Eq..., the corresponding values of the theoretical nuisance parameters
 1571 $\beta_{k,\text{th}}^{\text{min}}$ can be interpreted as leading to PDFs optimized (“profiled”) to describe this new specific measurement. Note
 1572 that in general the profiling will modify both the central value and the total PDF uncertainty.

1573 There are two major limitations of this Hessian profiling method. First of all, it assumes that the optimal PDF
 1574 parametrization will not be modified by the addition of the new experiment, and in many cases this is not true, since
 1575 new experiments might require more flexible input parametrizations. Secondly, the vanilla version of the Hessian
 1576 profiling method assumes that PDF uncertainties are defined by the $\Delta\chi^2 = 1$ criterion, which is not the case of the
 1577 global Hessian PDF fits as reviewed in Sect. 4.3. For this reason, the impact of the data as estimated by the Hessian
 1578 profiling will in general differ as compared to what would be obtained if the data would be included in a global Hessian
 1579 fit.

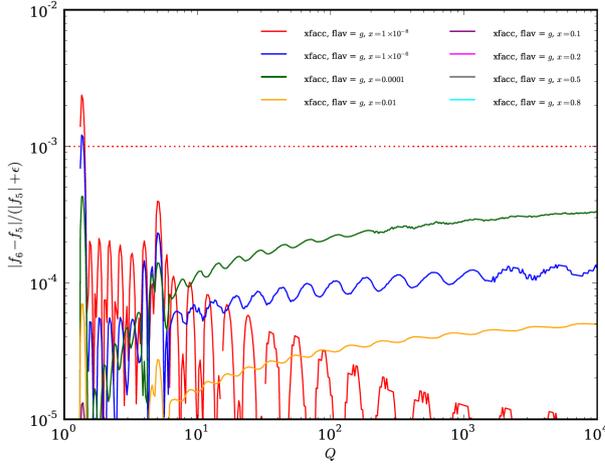
1580 *Bayesian Monte Carlo reweighting*

1581 We now discuss the Bayesian Monte Carlo reweighting.

1582 In Fig. 30 we show the gluon PDF in the NNPDF3.0 closure tests, estimating the impact of the collider inclusive
 1583 jet data and comparing the results of the Bayesian reweighting with those of a direct refit. In this study, the prior was
 1584 a set of $N_{\text{rep}} = 1000$ replicas obtained with NNPDF2.3-like dataset but without any collider inclusive jet production
 1585 data included. The pseudo-data were generated using MSTW08. We observe the very good agreement between the
 1586 approximate Bayesian reweighting method and the exact refit results.

1587 *4.7. Delivery*

1588 The final part of a PDF fit is of course to make it available to any potential user. In the beginning this was achieved
 1589 by means of (x, Q) interpolation tables and the corresponding driver codes which were specific from each PDF group.
 1590 This was however far from optimal since standardization was very difficult, with programs requiring PDFs as input
 1591 having to be adapted each time a new PDF set was released. A first step towards PDF access standardization was
 1592 achieved with the release in 1993 of PDFLIB [320] as part of the CERN Program Library software. This made
 1593 possible to have a unique interface for calling PDFs without the need of adding external files on a case by case basis.
 1594 In addition to the PDFs, also the value of $\alpha_s(m_Z)$ used in each specific fit could be accessed.



Process/PDF	t_5	t_6	t_5/t_6
Cross-section integrations, 1M phase space points			
CT10			
$pp \rightarrow jj$	23'10"	9'17"	2.5
$pp \rightarrow \ell\ell$	4'12"	2'02"	2.1
$pp \rightarrow H$ (ggF)	0'20"	0'15"	1.3
NNPDF23nlo			
$pp \rightarrow jj$	54'40"	9'28"	5.8
$pp \rightarrow \ell\ell$	8'06"	2'33"	3.2
$pp \rightarrow H$ (ggF)	0'25"	0'11"	2.3
CKKW event generation, 100k $pp \rightarrow \leq 4$ 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 31: 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 $pp \rightarrow 4$ jet events.

1595 The next step of this standardization process came with the release of in 2005 of LHAPDF, the Les Houches Ac-
1596 cord on PDFs [321, 322], which was developed as a functional replacement for PDFLIB. In order to ensure backwards
1597 compatibility, LHAPDF included LHAPDF glue, a PDFLIB-like interface. One of the main motivations to release
1598 LHAPDF was the realization that dealing with a large number of error PDF sets, that had then recently become avail-
1599 able, was extremely cumbersome with PDFLIB. In particular, LHAPDF was organized around the concept of *PDF*
1600 *set*, which was constituted by the central (average) member as well as the corresponding error PDF sets. As PDFLIB,
1601 LHAPDF was written in Fortran 77, although later a C/C++ interface was also developed.

1602 While the Fortran incarnation of LHAPDF was very popular and widely used, at some point its further develop-
1603 ment became very challenging in particular due to the intrinsic limitations of Fortran 77 as its native language. In
1604 particular, since Fortran 77 required to allocate space for all available PDFs at compilation time, the memory footprint
1605 eventually become impossible to handle and LHAPDF v5.9.1 was the last release. To overcome these limitations, a
1606 complete rewriting of LHAPDF from scratch in C++ was completed in 2014, dubbed LHAPDF6 [323]. In addition
1607 to reducing static memory requirements by orders of magnitude, this C++ incarnation of LHAPDF offered improved
1608 CPU performance and improved interpolation and extrapolation functionalities. Moreover, its cascading meta-data
1609 system ensures that software releases are completely decoupled from the availability of novel PDF sets. Also to
1610 ensure backwards compatibility, Fortran 77 interfaces are also provided.

1611 In terms of interpolation accuracy, LHAPDF6 reproduces the v5 results down to residual differences of at most
1612 0.1%. This is illustrated in Fig. 31, where we show the relative difference between LHAPDF v5 and v6 for $g(x, Q)$
1613 for different values of x as a function of Q , using CT10 as input PDF. In Ref. [323] it was also shown that LHAPDF6
1614 improves also the CPU timings as compared to v5 by a factor between 2 and 6, as shown by the right table in Fig. 31,
1615 which represents the timing improvements in v6 as compared to v5, t_6/t_5 , for a cross-section integration of 1M phase
1616 space points with Sherpa [324] and for CKKW event generation of 100k $pp \rightarrow 4$ jet events.

1617 Currently LHAPDF6 has established itself as the almost universal software to access PDFs. Its current version of
1618 v6.1 and more than 700 PDF sets can be accessed. In addition to unpolarized parton distributions, the flexibility of
1619 the LHAPDF6 framework makes it suitable to release other types of non-perturbative QCD objects, and indeed also
1620 polarized PDFs and nuclear PDFs are available.

1621 **5. PDF analyses: state of the art**

1622 Here we review the latest developments from the main PDF fitting groups. The comparison among them is left for
1623 the next section.

1624 *5.1. CT*

1625 **Jun** The CTEQ-TEA global analysis of QCD was established by Wu-Ki Tung et. al. since early 1990's with the
1626 CTEQ1 PDFs [65]. The most recent release on general purpose PDFs from the CTEQ-TEA collaboration are CT14
1627 PDF sets [18], including the nominal sets as well as alternative sets with different choices of α_s and maximum number
1628 of active flavors. The parton distributions are parametrized analytically with a form of products of polynomials and
1629 the asymptotic shapes at small and large x . The parametrizations are defined at an initial scale of $Q_0 = 1.3$ GeV and
1630 then evolved to higher scales with DGLAP evolution. Be specific the general form of parametrization adopted can be
1631 written as

$$xf_a(x, Q_0) = x^{a_1}(1-x)^{a_2}P_a(x), \quad (104)$$

1632 where $P_a(x)$ is a multiplicative factor for flavor a . In pre-CT14 analyses P_a was chosen as an exponential of a
1633 polynomial in x or \sqrt{x} so the positivity conditions of PDFs at the initial scale are enforced. In CT14 analysis a better
1634 style of parametrization was introduced, e.g., for the u -valence

$$P_{u_v} = d_0p_0(y) + d_1p_1(y) + d_2p_2(y) + d_3p_3(y) + d_4p_4(y), \quad (105)$$

1635 with $y = \sqrt{x}$ and p_n be the 4-th order Bernstein polynomials, e.g.,

$$p_0(y) = (1-y)^4, \quad p_1(y) = 4y(1-y)^3, \quad p_2(y) = 6y^2(1-y)^2, \quad p_3(y) = 4y^3(1-y), \quad p_4(y) = y^4. \quad (106)$$

1636 Namely the multiplicative factor is chosen as a fourth-order polynomial of y with expansion in the basis of Bernstein
1637 polynomials. Since each Bernstein polynomial is strongly peaked at a different value of y it reduces correlations
1638 between those coefficients or PDF parameters d_n and renders a stable Hessian approach. In the CT14 case the positivity
1639 of PDFs at $Q_0 = 1.3$ GeV emerges automatically as a consequence of the fit to data. CT14 PDFs have a total number of
1640 free parameters of 28 in the PDF parametrization. More flexible parametrizations with even higher-order polynomials
1641 are found to have small effects on either the best-fit or the estimated PDF uncertainties comparing to CT14 in region
1642 well constrained by the data.

1643 The CT14 global analysis includes a variety of experimental data. The majority of the data are from the in-
1644 clusive DIS and semiinclusive DIS measurements on structure functions or reduced cross sections from fixed-target
1645 experiments (BCDMS [325, 326], NMC [103], CCFR [108, 107, 111], NuTeV [112], CDHSW [327]) or HERA ex-
1646 periments [328, 121, 85, 115]. A Q cut of 2 GeV and W cut of 3.5 GeV are adopted in selection of those data points
1647 to minimize non-perturbative effects from either nuclear corrections or higher twists corrections. Thus no further
1648 nuclear or higher twists corrections are included in theory predictions in CT14 except for those already applied in
1649 the unfolding of experimental data. The CT14 analysis utilizes a treatment of heavy-quark mass effects upto NNLO
1650 for the NC DIS processes realized in one of the GM-VFN scheme, the S-ACOT- χ scheme [329]. The calculations
1651 implemented for charged-current DIS processes are at NLO only given their small number of data points as well
1652 as large experimental errors. There are also Drell-Yan production data from fixed-target experiments (E605 [330],
1653 E866 [167]) and W/Z boson production data from Tevatron [331, 332, 333, 334, 170] including the new D0 electron
1654 charge asymmetry data [172]. They provide further discriminations on quark flavors at large- x region. Especially,
1655 NNLO predictions from ResBos [210, 335, 217, 336] are used for W/Z boson production data with p_T cut on the
1656 charged leptons incorporating soft gluon resummation effects at small p_T of the vector boson. The resummed predic-
1657 tions render a better description of the p_T spectrum of the decayed charged leptons. The updated D0 electron charge
1658 asymmetry data [172] shows a large impact on the d/u PDF ratio at large- x comparing to CT10 and CT10W [337, 81].
1659 The Tevatron W asymmetry data can probe the average slope of d/u ratio at large- x , about 0.1 or larger. In CT10 fits
1660 the D0 lepton charge asymmetry data resulted in larger asymptotic value of d/u though tensions were found between
1661 different subsets of the data or the D0 data and other DIS experiments. In CT14 the updated D0 electron charge
1662 asymmetry data shows better agreements with other data sets in the global analysis and drives the d/u ratio to a lower
1663 value close to CTEQ6.6 [338] at large- x as shown in Figure 32. The d/u ratio in CT14 also shows good agreements

1664 with extractions from CJ12 [339] that are based on independent large- x and low- W DIS data with power corrections
 1665 and deuteron corrections. Similar data on W/Z boson production from LHC Run I are also included from the AT-
 1666 LAS [176], CMS [182, 181] and LHCb [340] experiments. They further extend the coverage to intermediate and
 1667 small x region. Besides, single inclusive jet production from Tevatron [135, 136] and LHC [133, 141] provide domi-
 1668 nant constraint on gluon PDF at large x region with the latter also show coverage extended to intermediate x region.
 1669 For inclusive jet production at hadron colliders only the NLO predictions are available at the time of the CT14 fit
 1670 carried out. They will be updated with the recent NNLO calculations [152] in future CTEQ-TEA analysis.

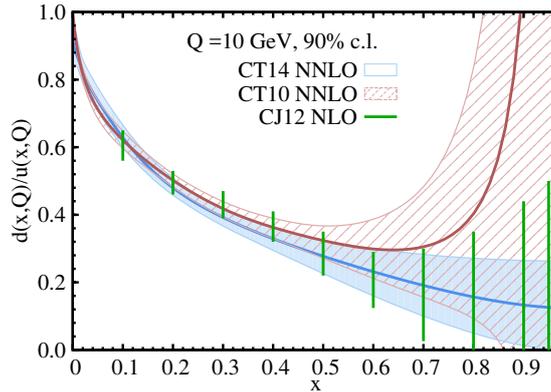


Figure 32: A comparison of 90% C.L. uncertainties on the d/u PDF ratio at $Q = 10$ GeV for CT14 NNLO (solid blue) and CT10 NNLO (dashed red), and CJ12 (green lines) error ensembles [18].

1671 The CTEQ-TEA group use the Hessian method with certain tolerance conditions for the nominal fits in the de-
 1672 termination of PDF uncertainties. That is supplemented with a Lagrange multiplier (LM) scan for certain important
 1673 observables or in less constrained region by data. In the Hessian method first the log-likelihood χ^2 is constructed from
 1674 agreement of theory and experimental data as a function of all the PDF parameters. The central/best-fit PDF is found
 1675 through minimization of the global χ^2 . The Hessian matrix with respect to all free PDF parameters (a total number
 1676 of D) is calculated through finite differences of the χ^2 around the global minimum. A total number of D eigenvector
 1677 directions are determined through diagonalization of the Hessian matrix. The boundaries of 90% c.l. intervals along
 1678 each eigenvector direction are then determined according a tolerance criteria similar to the case of hypothesis test,
 1679 instead of using a parameter fit criteria of $\Delta\chi^2 = 1$. That is due to the possible tensions between different data sets as
 1680 well as the remaining dependence on the PDF parametrizations. In pre-CT10 analyses it was found with a global χ^2
 1681 tolerance of $\Delta\chi^2 = 100$ (for more than 2000 data points) the fit can agree with all the experiments at 90% c.l.. In latter
 1682 CTEQ-TEA analyses a more efficient dynamic tolerance criteria is adopted to account for agreement with individual
 1683 data set. It is constructed from an equivalent Gaussian variable, e.g.,

$$S_n = \sqrt{2\chi^2(N_n)} - \sqrt{2N_n - 1}, \quad (107)$$

1684 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
 1685 follows a normal distribution given the number of data points is large enough. Thus a value of S_n greater than 1.3
 1686 will be excluded at 90% c.l.. We add a second layer of penalty to the global χ^2 when determining the boundaries of
 1687 confidence intervals, called Tier-2 penalty,

$$P = \sum_{n=1}^{N_{exp}} (S_n/S_{n,best})^{16}, \quad (108)$$

1688 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
 1689 to certain experiments. The power of 16 is introduced so the penalty will reach the tolerance of 100 as soon as any data
 1690 set shows disagreement at 90% c.l.. The tolerance criteria then changes to $\Delta\chi^2 + P = 100$. Figure 34 shows distribution

1691 of $S_{n,best}$ for all 33 experiments included. The distribution is wider than a normal distribution indicating the presence
 1692 of disagreement, or tensions, between some of the included experiments.

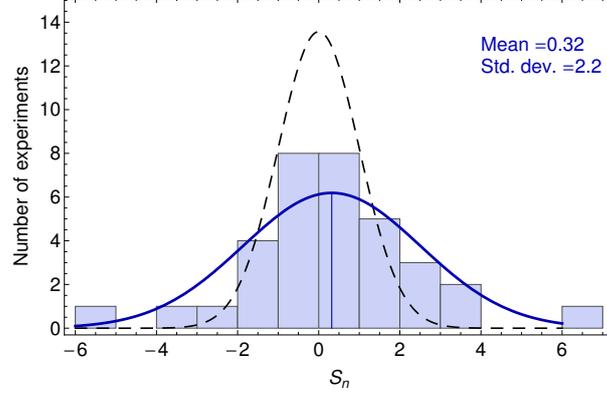


Figure 33: Best-fit values of the equivalent Gaussian variables for 33 experiments in the CT14 NNLO global analysis [18].

1693 With the best-fit and 2D error sets, the asymmetric errors for any QCD observable X can be calculated through the
 1694 master formula

$$(\delta X)_+ = \sqrt{\sum_{i=1}^D [\max(X_{+i} - X_0, X_{-i} - X_0, 0)]^2}, \quad (\delta X)_- = -\sqrt{\sum_{i=1}^D [\max(X_0 - X_{+i}, X_0 - X_{-i}, 0)]^2}, \quad (109)$$

1695 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
 1696 errors can be scaled down to 68% c.l. by dividing with a factor of 1.64 assuming Gaussian distributions. As mentioned
 1697 earlier the CTEQ-TEA analysis also uses Lagrange multiplier method [309] to crosscheck the error estimation from
 1698 nominal Hessian sets. In the CT14 analysis, a Lagrange multiplier scan has been performed for the cross sections of
 1699 Higgs boson production via gluon fusion and of the top quark pair production at the LHC. In such scans the best-fits
 1700 and the associated χ^2 are found for each fixed value of the observable studied. Then the PDF uncertainties on the
 1701 observable are determined from the χ^2 profile obtained based on the same tolerance criteria as in the Hessian method.
 1702 Figure 34 shows the good agreement of the 90% C.L. uncertainties for the Higgs cross sections from the CT14 Hessian
 1703 PDFs and the CT14 LM scan. The latter one can be read as intersection of $\Delta\chi^2 = 100$ and various curves. The LM
 1704 method does not rely on the linear approximation. Thus it serves as a robust check of the Hessian results.

1705 There are a few other specialities of the CTEQ-TEA global analysis. The CTEQ-TEA analyses use world average
 1706 on strong coupling constant $\alpha_s(M_Z)$ as an input. Usually the nominal fit is done with $\alpha_s(M_Z) = 0.118$ at both NLO and
 1707 NNLO. Additional fits with alternative α_s choice are also provided. It has been shown that the resulting uncertainties
 1708 can be added in quadrature to the PDF uncertainties [341]. The fit itself alone shows much weaker constraint on
 1709 α_s than the world average. Similarly the pole mass of charm quark and bottom quark are chosen to be close to the
 1710 world average, $m_c = 1.3$ GeV and $m_b = 4.75$ GeV. The CTEQ-TEA group also provides specialized fits with non-
 1711 perturbative charm PDFs. A sea-like or valence-like charm distribution is added to the nominal parametrization and
 1712 then fitted to data. Limits on momentum fraction carried by the fitted charm at initial scale $Q_0 = 1.3$ GeV are then
 1713 derived. In most recent CT14 analysis, the limits are 1.6% for Sea-like model and 2.1% for BPHS model both at
 1714 90% c.l. [342]. There are also CT14 QED PDFs [343] based on radiative ansatz for the inelastic component of the
 1715 photon PDFs. The 90% C.L. limit on momentum fraction of the proton carried by photon is 0.11% at $Q_0 = 1.3$ GeV
 1716 as derived from fit to the ZEUS measurement on isolated photon production [344].

1717 5.2. MMHT

1718 The MMHT14 PDFs [19] are the successor to the MSTW08 [80] set, which derives from the earlier MRST and
 1719 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

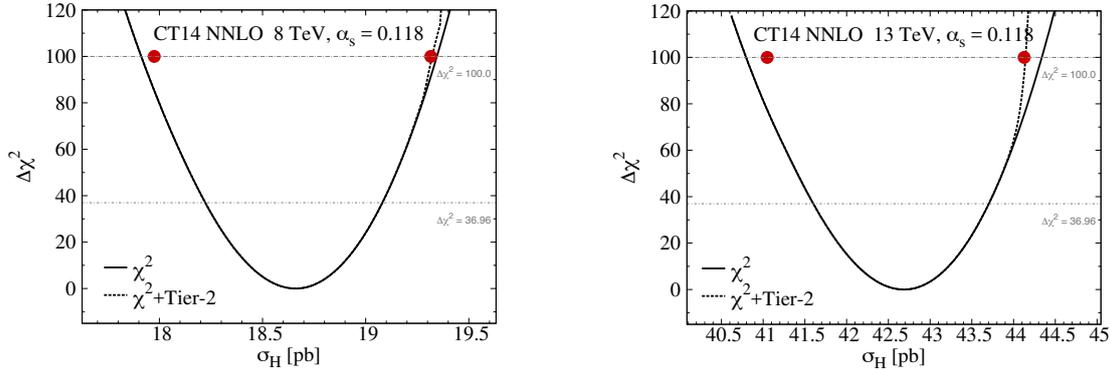


Figure 34: Dependence of the increase in χ^2 in the constrained CT14 fit on the expected cross section σ_H at the LHC 8 and 13 TeV [18], for $\alpha_s(M_Z)=0.118$. The solid and dashed curves are for the constrained fits without and with the Tier-2 penalties, respectively. The red dots correspond to the upper and lower 90% C.L. limits calculated by the Hessian method.

1720 was released, including data from HERA and the Tevatron for the first time. This corresponded to a truly global anal-
 1721 ysis, fitting to fixed target, DIS and hadroproduction data to constrain the PDFs as precisely as possible. Subsequent
 1722 releases have all built on this approach, but with significant advances achieved over the years due to improvements in
 1723 both theory and experiment. The MRST98 release [345] was the first set to include a full treatment of heavy flavours
 1724 within the GM-VFNS developed in [346], and discussed further in Section 2.5. This was motivated by the new HERA
 1725 measurements of the charm structure function, which demonstrated the importance of a consistent treatment of charm
 1726 production and low and high scales; indeed, the introduction of this flavour scheme resulted in an improved descrip-
 1727 tion of such data. The MRST02 release [77] included a full treatment of PDF errors for the first time, described further
 1728 below, while the MRST04 [347] set went to NNLO for the first time.

1729 These elements were all incorporated in the major MSTW08 [80] release. This presented a global fit to a range of
 1730 DIS data from HERA and fixed proton and nuclear targets, fixed target Drell-Yan and dimuon production and W , Z
 1731 and jet production at the Tevatron, with $O(2500)$ data points in total. Fits were performed up to NNLO in the strong
 1732 coupling, with an improved dynamical error treatment, and with an up to date heavy flavour scheme applied. This
 1733 aimed to provide a PDF set for use at the LHC, which began operation soon after the release, and was subsequently
 1734 very widely used in LHC phenomenological studies and experimental analyses. This fit was updated in the latest
 1735 MMHT14 [19] set, which includes a number of theoretical and experimental updates. In particular for the first time
 1736 LHC data on W , Z , $t\bar{t}$ and jet production are included, as well as updated HERA data on the charged, neutral, charm
 1737 and longitudinal structure functions, and updated Tevatron W and Z measurements. As in earlier fits, for DIS data a
 1738 Q^2 cut of 2 GeV^2 and W^2 cut of 15 GeV^2 is imposed to avoid sensitivity to higher twist corrections.

1739 In the case of Tevatron jet production, in the absence of a full NNLO calculation at the time, an approximation to
 1740 the NNLO corrections based on the threshold corrections of [158] was applied in the NNLO MSTW08 fit, with the
 1741 judgement being made that the difference between this and the full NNLO result would be expected to be smaller than
 1742 the systematic uncertainties on the data, which itself provided the only direct constraint on the gluon at high x . At the
 1743 LHC much of the jet data are quite far from threshold, while those that do not probe a kinematically similar region to
 1744 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
 1745 to be determined from the fit, with the pole mass value of $m_t = 172.5 \pm 1 \text{ GeV}$ taken as an input. This gives a value at
 1746 NNLO that is consistent with the world average, while at NLO it is somewhat lower.

1747 The MSTW PDFs were parameterised in terms of simple polynomials in x , with 29 free parameters. However,
 1748 in [348] it was shown that this parameterisation was not sufficiently adaptive to describe the Tevatron and LHC
 1749 W asymmetry data. In particular, it was necessary to introduce a more flexible basis for the interpolating function
 1750 described in Section 4.1.1, with

$$I_f(x) = \sum_i^n \alpha_{f,i} T_i(y(x)), \quad (110)$$

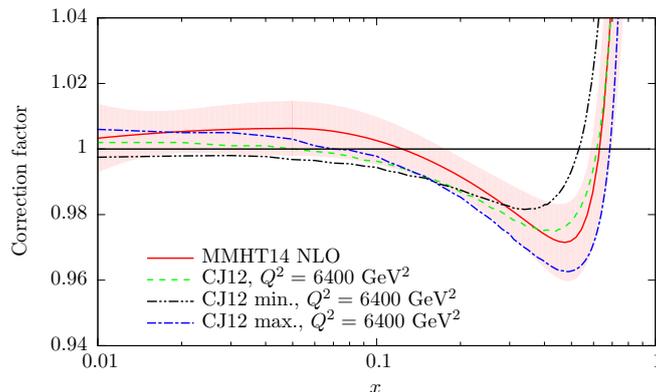


Figure 35: Fitted MMHT14 deuteron correction factors with uncertainty, compared to the CJ12 [339] predictions. Taken from [19].

1751 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
 1752 x , and has the additional advantage that this provides a half-integer separation in powers of x , as expected on Regge
 1753 theory grounds. In order to determine how many parameters n are needed, in [348] pseudo-data points with a constant
 1754 percentage error were generated for the required distributions, in terms of a very large order polynomial with additional
 1755 smoothness constraints applied. The fractional deviation from the true PDF, as well as the decrease in χ^2 , were then
 1756 determined as the number of parameters were increased, until no further significant improvement was observed and
 1757 the level of agreement was well below the PDF uncertainty for the set. In this way $n = 4$ was arrived at as a good
 1758 choice with which to parameterise the u_V , d_V , $s + \bar{s}$ and light quark sea S distributions. Fitting to the same MSTW08
 1759 data set, these resulted in some improvement in the fit quality, but with the only significant change in the PDF being
 1760 in the u_V at lower x . This was found to lie outside the previous PDF uncertainty band, and the additional flexibility
 1761 provided a greatly improved description of W asymmetry data.

1762 In the MMHT14 set, this Chebyshev parameterisation is used at $Q_0^2 = 1 \text{ GeV}^2$ for the u_V , d_V , $s + \bar{s}$ and light
 1763 quark sea S distributions, while for the gluon a term with $n = 2$ Chebyshevs is included, but with a second term still
 1764 present, as in MSTW08, which has a different low x and provides the additional flexibility at low x that is required
 1765 by the HERA data; this has the effect that the gluon at NLO and higher becomes negative at low x and Q^2 . Standard
 1766 polynomial parameterisations are taken for the less constrained $s - \bar{s}$ and $\bar{d} - \bar{u}$ distributions, although as the data
 1767 becomes more precise we can expect this to change.

1768 A further improvement described in [348] that is included in the MMHT14 set is in the treatment of the non-
 1769 perturbative corrections that should in general be applied when considering DIS data on deuteron targets, to account
 1770 for the binding of the proton and neutron within the deuteron. While in MSTW08 and earlier fits, a fixed shadowing
 1771 correction at small x was applied, a more flexible approach is now taken. In particular the deuteron corrections are
 1772 freely parameterised in terms of a function $c(x)$, which is determined along with its corresponding uncertainties from
 1773 the PDF fit. This resulted in a significantly improved description of the BCDMS deuteron structure function data,
 1774 the E866 Drell-Yan asymmetry and the Tevatron lepton asymmetry data, with some significant changes in d_V . The
 1775 result of the MMHT14 fit is shown in Fig. 35 and compared against different model predictions used in the CJ12 [339]
 1776 analysis. Interestingly, very good agreement is found with the CJ12mid prediction, demonstrating the power of global
 1777 PDF fits to extract additional physical information beyond the PDFs themselves.

1778 In MMHT14, the Hessian approach is applied to calculate the PDF errors, with the ‘dynamical’ tolerance criteria
 1779 described in Section 4.3.1 taken. This accounts for the fact that we are not generally fitting consistent data sets
 1780 with purely Gaussian errors and a well defined theory, by appealing to a weaker ‘hypothesis-testing’ procedure that
 1781 simply requires that all fitted data sets are described with the required confidence limit. For MMHT14 the 68%
 1782 uncertainties are calculated using this procedure. In the fit there are 37 free PDF parameters in total, however in
 1783 the error determination certain parameter directions are found to be largely degenerate, leading to departures from
 1784 quadratic χ^2 behaviour. This is corrected by fixing some parameters when calculating the error eigenvectors, reducing

1817 are presented, see Section 7 for more discussion. A further public release is therefore anticipated in the near future.

1818 5.3. NNPDF

1819 As discussed in Sect. 4, the NNPDF fitting methodology is based on the combination of three main components:
1820 i) the use of artificial neural networks as universal unbiased interpolants, ii) the Monte Carlo method to estimate and
1821 propagate PDF uncertainties, and iii) Genetic Algorithms minimization for the training of the neural networks. Here
1822 we review the main developments in the NNPDF family of PDF fits.

1823 The NNPDF methodology was first presented in [87], where it was used to produce a neural network based
1824 determination of the proton, deuteron and non-singlet DIS structure functions from the fixed target data from NMC and
1825 BCDMS. As a first phenomenological application, this determination was used to extract the strong coupling constant
1826 $\alpha_S(m_Z)$ from scaling violations of truncated moments of structure functions [355]. This analysis was subsequently
1827 extended [306] to include as well the F_2^p measurements from the H1 and ZEUS experiments at the HERA collider.
1828 Note that a determination of structure functions is purely data-driven, with no theoretical input required at this point.

1829 When moving from fitting structure functions to PDFs, there are a number of simplifications, for instance one
1830 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 of technical complica-
1831 tions, the most important one being able to compute DIS structure functions starting from the neural-network based
1832 parametrization of $q_i(x, Q_0)$. First of all, the usual ANN training algorithm of back-propagation cannot be used in
1833 this case, due to the convolution of the PDFs with the DGLAP evolution kernels and the DIS coefficients functions.
1834 To overcome this limitation, it was demonstrated how Genetic Algorithms can be efficiently used for ANN training
1835 under a non-trivial mapping between the latter and the exp data, and used to extract the QCD vacuum condensates
1836 from hadronic tau decay data [356]. An efficient method to solve the DGLAP evolution equations in N -space was also
1837 developed, called the `Fast Kernel` method. With these ingredients at hand, it became possible for the first time to
1838 apply the NNPDF methodology to a determination of parton distributions, starting from a fit of the non-singlet com-
1839 bination $q_{NS}(x, Q_0)$ [305] and them moving to a first full-fledged NLO PDF fit based on neutral-current DIS structure
1840 function data [357], the NNPDF1.0 analysis.

1841 Subsequently, the global NNPDF fits were improved both by adding new experimental data, updated theoretical
1842 calculations and/or refining the fitting methodology. To begin with, the NNPDF1.2 analysis [358] released the previ-
1843 ously used assumption that the strange sea was proportional to the light quark sea, $s = \bar{s} = \kappa(\bar{u} + \bar{d})$, and parametrized
1844 both s^+ and s^- using neural networks, exploiting the constraints from the NuTeV dimuon charged-current neutrino
1845 scattering data. Two important phenomenological consequences of this analysis were, first of all, the demonstration
1846 that the PDF uncertainties associated with s^- were enough to completely wipe out the NuTeV anomaly [359] in the
1847 determination of the weak mixing angle $\sin^2 \theta_W$; and second, a direct extraction of the CKM matrix element V_{cs} with
1848 a precision compatible with that of the PDG average.

1849 In 2010, the NNPDF2.0 set was released [88], which constituted the first truly global PDF fit from the NNPDF
1850 collaboration. In addition to the NC and CC DIS structure data included in previous releases, NNPDF2.0 included in
1851 addition fixed-target Drell-Yan cross-sections from the Fermilab E605 and E866 experiments, inclusive jet production
1852 measurements from CDF and D0 and the Tevatron as well as the differential rapidity distributions of the Z boson also
1853 from the Tevatron. From the theoretical point of view, NNPDF2.0 was still based on the zero-mass VFN scheme,
1854 and thus charm and bottom structure function data from HERA were not included. A good overall description of all
1855 experiments in the global fit was found. NNPDF2.0 was also the first PDF set to include the recently release HERA
1856 combination of H1 and ZEUS structure function data for the Run I data period [85]. The NNPDF2.0 was one of the
1857 sets included in the PDF4LHC 2011 recommendation [91] for the usage of PDFs at the LHC.

1858 While NNPDF2.0 demonstrated that the NNPDF methodology could be successfully applied to a global determi-
1859 nation of parton distributions, there were still a number of important limitations from the theoretical point of view.
1860 First of all, the use of a ZM-VFN scheme neglected heavy quark mass effects in the DIS structure functions, which
1861 were known to be specially important for the description of the low- x , low- Q^2 HERA data. Secondly, all NNPDF fits
1862 so far were based in NLO theory, and NNLO accuracy was essential to match the corresponding precision of important
1863 partonic hard-scattering cross-sections such as Higgs production in gluon fusion. The first of these theory limitations
1864 was removed with the release of NNPDF2.1 [360], which was based on the FONLL general-mass VFN for the cal-
1865 culation of DIS structure functions, which in addition allowed to fit as well the HERA charm and bottom structure
1866 functions data. This analysis also showed that the impact of heavy quark mass effects was less drastic as previously

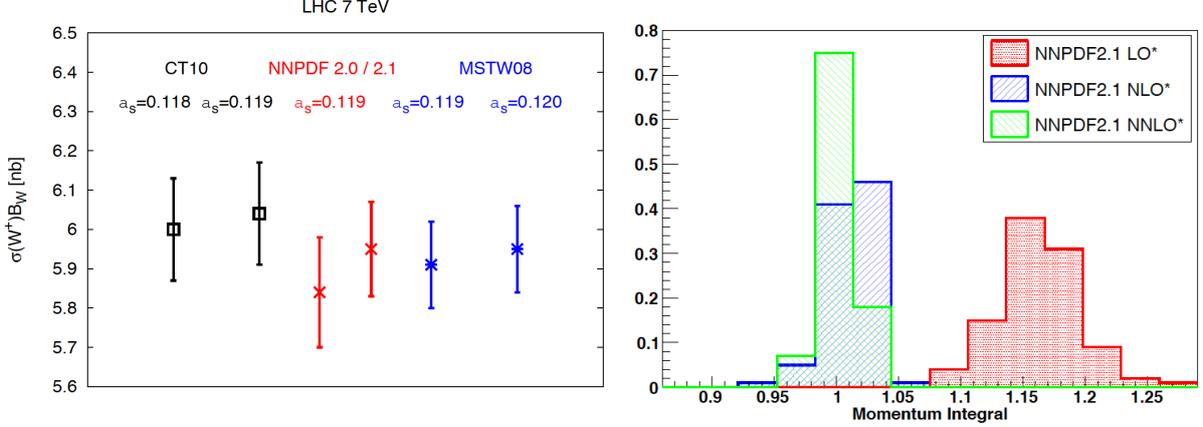


Figure 37: Left plot: comparison between the NNPDF2.0 and 2.1 predictions for the inclusive W^+ production cross-section at the LHC 7 TeV, which illustrates the phenomenological impact of heavy quark mass effects. Right plot: the distribution of the momentum integral Eq. (111) among the MC replicas for the variants of the NNPDF2.1 LO, NLO and NNLO fits that do not impose explicitly the momentum sum rule.

1867 reported, with the cross-section predictions between NNPDF2.0 and 2.1 typically agreeing at the one-sigma level.
 1868 The NNPDF2.1 fit was used to produce a determination of the strong coupling $\alpha_s(m_Z)$ from the global dataset [361].

1869 The second of these theoretical limitations was also removed shortly after, with the release of a NNLO version
 1870 of NNPDF2.1 [89]. This PDF set was based on the same dataset as its NLO counterpart, the differences is that now
 1871 DIS and hadronic cross-sections were computed at NNLO, in the former case using the FONLL-C GM-VFN scheme.
 1872 In the same publication, the first NNPDF LO sets were also presented. The availability of NNPDF2.1 fits at LO,
 1873 NLO and NNLO allowed a systematic study of the perturbative convergence of the global fit, finding in particular
 1874 reasonable agreement at the one-sigma level between the NLO and NNLO versions. The consistency of the global
 1875 QCD analysis framework was also tested by performing fits without imposing the momentum sum rule and then
 1876 verifying a posteriori that the global fit result was consistent with the QCD expectation, finding indeed that at NNLO

$$[M] \equiv \int_0^1 dx \left(g(x, Q^2) + \Sigma(x, Q^2) \right) = 1.002 \pm 0.014, \quad (111)$$

1877 in agreement with QCD. The NNPDF2.1 NNLO analysis was also used to perform a determination of the strong
 1878 coupling constant [362], finding a value $\alpha_s(m_Z) = 0.1173 \pm 0.0007^{\text{stat}} \pm 0.0009^{\text{pert}}$, a result which is still included in
 1879 the PDG global average of α_s [363].

1880 In parallel to these developments, it was realized that an alternative method to quantify the impact of new measure-
 1881 ments in the global fit was to exploit Bayesian reweighting [317, 318]. Building on earlier work by Giele, Keller and
 1882 Kosower [364, 74], it was demonstrated how Bayesian inference determines the weight of each Monte Carlo replica
 1883 of a given prior PDF set upon including a new dataset in the global fit, given by

$$\omega_k \propto \left(\chi_k^2 \right)^{(n-1)/2} e^{-\chi_k^2/2}, \quad k = 1, \dots, N_{\text{rep}}, \quad (112)$$

1884 where χ_k^2 has been computed using the theory predictions based on the k -th replica of the prior set and the new
 1885 experimental measurements. It was shown that, provided the impact of the new data was moderate, the reweighting
 1886 was fully equivalent (both in principle and in practice) as refitting. The main advantage of the reweighting method is
 1887 allowing to gauge the impact of new data without having to do a PDF fit and based only on public information, so that
 1888 it was not restricted to PDF fitters anymore.

1889 Just as in the early 90s the availability of the HERA structure function measurements became a game-changer for
 1890 global fits, also starting from 2010 the LHC experiments started producing a wealth of PDF-sensitive information,

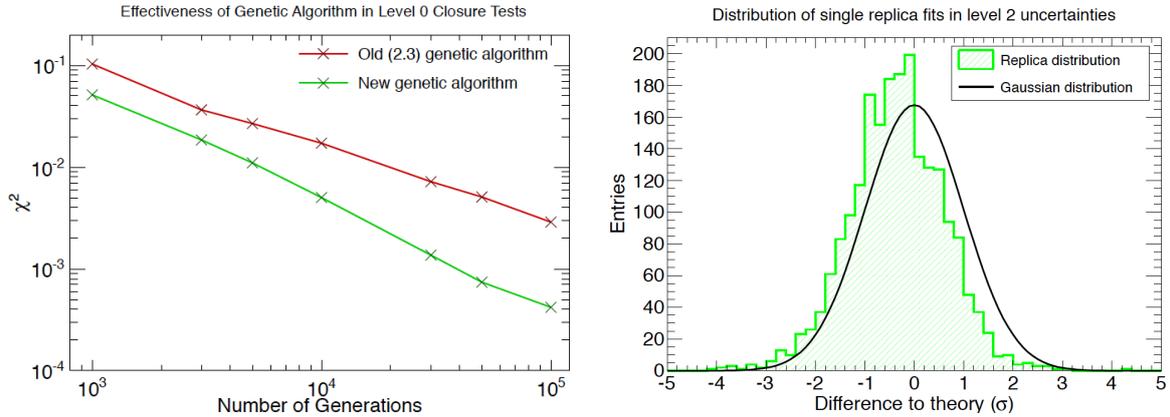


Figure 38: Representative results of the closure tests presented in the NNPDF3.0 analysis. Left plot: in a level 0 closure tests, where the pseudo-data is generated without any statistical fluctuations, the χ^2 should decrease monotonically as a function of the number of GA iterations, down to arbitrarily 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.

1891 which promised to also impact global fits in a significant way. With this motivation, in 2012 the NNPDF2.3 set
 1892 was released [90], which was back then the first PDF fit to include LHC data, in particular electroweak gauge boson
 1893 production from ATLAS, CMS and LHCb as well as jet production from ATLAS. As will all subsequent releases,
 1894 NNPDF2.3 was available at LO, NLO and NNLO and was based on the FONLL general mass scheme. The NNPDF2.3
 1895 became the baseline PDF set in several popular Monte Carlo event generators, such as Pythia8 and aMC@NLO.

1896 Following the release of NNPDF2.3, it was realized that the increase in complexity required to include the many
 1897 new experiments that were either available or about to be released could not be satisfactory tackled with the current,
 1898 FORTRAN77-based code. With this motivation, a complete rewriting of the NNPDF global analysis framework into
 1899 C++ and Python was undertaken, a two-year long effort that culminated with the release of the NNPDF3.0 set [17].
 1900 In addition to including many new LHC experiments on jets, vector boson production, W +charm and top production,
 1901 the main result of NNPDF3.0 was the systematic validation of the complete fitting methodology based on statistically
 1902 robust closure tests. In these closure tests, pseudo-data was generated based on some “truth” PDFs, and then a PDF
 1903 fit was performed: only if the resulting PDF central values and uncertainties were consistent with the (known) input
 1904 PDFs, the closure test can be considered succesful. In Fig. 38 we show some representative results of the closure
 1905 tests presented in the NNPDF3.0 analysis. In the left plot we show the results of a level 0 closure tests, where the
 1906 pseudo-data is generated without any statistical fluctuations, the χ^2 should decrease monotonically as a function of
 1907 the number of GA iterations, down to arbitrarily small values. And in the right plot we show the distribution of the
 1908 difference between theory and data, in units of the error of the latter among each of the Monte Carlo replicas. This
 1909 distribution is consistent with the Gaussian predicted by statistics.

1910 A recent development in the NNPDF family of global analysis concerns the treatment of the charm PDF. In all
 1911 previous PDF sets, similarly to the rest of the global fits, NNPDF assumed that the charm PDF was generated dynam-
 1912 ically from the gluons and light quarks, as dictated by the DGLAP evolution starting from the charm mass threshold
 1913 $\mu_c \simeq m_c$. However, a possible non-perturbative component of the charm PDF would invalidate this assumption, which
 1914 can thus ultimately be validated from experimental data. In addition, treating the charm PDF on an equal footing that
 1915 the gluon and light quark PDFs offers other potential advantages, such as a reduce dependence on the value of m_c and
 1916 an improved data/theory agreement from the more flexible input PDF parametrization. With this motivation, a variant
 1917 of the NNPDF3.0 fit with a fitted charm PDF was studied in [25]. By parametrising the charm PDF with an artificial
 1918 neural network with 37 free parameters, we found that fitting charm lead to an improved χ^2 for several experiments,
 1919 stabilized the dependence of the fit with respect to the value of m_c and moreover allowed for the first time a satisfactory
 1920 description, $\chi^2/N_{\text{dat}} \simeq 1$, of the EMC charm structure function data. The resulting charm PDF can be compared with

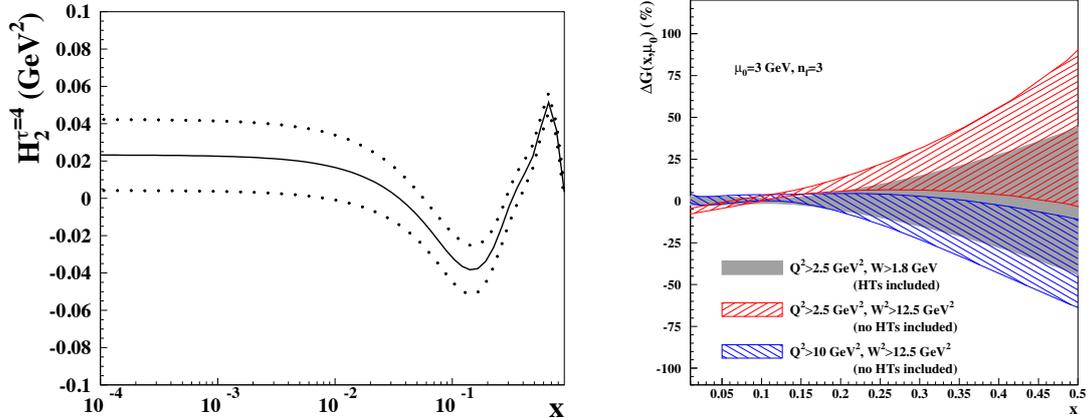


Figure 39: (Left) The higher twist coefficient for the F_2 structure function from the ABMP16 fit, including $1-\sigma$ uncertainty. (Right) Percentage difference in ABMP $n_f = 3$ gluon distribution between the default result and fits performed with higher W^2 cuts, without higher twist corrections. The $1-\sigma$ uncertainty bands are shown. Plots taken from [20].

1921 non-perturbative models [365], and some tantalizing evidence for a large- x non-perturbative charm component in the
 1922 proton was found. Predictions for a number of LHC process such as Z +charm and large- p_T D meson production were
 1923 performed, showing the potential of Run II data to disentangle the charm content of the proton.

1924 The most recent incarnation of the NNPDF global analysis is the recent NNPDF3.1 set. The main motivation for
 1925 this release was the availability of a large number of high-precision collider measurements providing PDF sensitivity
 1926 information, including some of them that for the first time could be used in a PDF fit such as differential distributions in
 1927 top quark pair production and the p_T of Z bosons. Equally important, thanks to the impressive progress in NNLO QCD
 1928 and NLO electroweak calculations, theoretical predictions were also available that allowed the consistent inclusion of
 1929 all these observables in the global NNLO fit. The second main motivation was to provide a state-of-the-art PDF set
 1930 without assuming that charm is generated perturbatively, that is, providing baseline global PDF fits with fitted charm.
 1931 Some of the new experiments included in NNPDF3.1 were the $t\bar{t}$ distributions from ATLAS and CMS, the legacy
 1932 LHCb inclusive W and Z measurements from Run II, the D0 W asymmetries in the muon and electron channel, the p_T
 1933 of Z bosons from ATLAS and CMS at $\sqrt{s} = 8$ TeV, as well as several other inclusive gauge boson and jet production
 1934 measurements from ATLAS and CMS.

1935 5.4. ABM

1936 The ABMP16 [20] set is the latest PDF fit following on from the ABM11 [366] and ABKM09 [82] sets. These are
 1937 based on the earlier studies of [71, 83, 84] to HERA and fixed proton and deuteron target DIS data, with the ABKM09
 1938 fit [82] and those that follow it including in addition fixed target Drell–Yan and dimuon production data from neutrino
 1939 DIS on fixed nuclear targets. The PDFs are parameterised in terms of polynomials in x , with the latest fits including
 1940 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
 1941 determination of the PDF errors is applied. All sets from ABKM09 onwards go to NNLO in the strong coupling.

1942 Two notable features of these fits are the use of a purely FFNS for the charm and bottom quark contributions in
 1943 the fit and the treatment of higher-twist effects. In the latter case no attempt is made to impose a cut to remove the
 1944 region of sensitivity to such effects. Rather, a lower cut of $W > 1.8$ GeV is imposed for the DIS data then is typically
 1945 applied in other PDF fits. The structure functions are then given by

$$F_i(x, Q^2) = F_i^{\text{TMC}}(x, Q^2) + \frac{H_i^{\tau=4}}{Q^2}, \quad (113)$$

1946 where $i = 2, T$. Thus x dependent and Q^2 independent twist-4 corrections H_i^4 are included. While the effect of these
 1947 dies off with increasing Q^2 at lower scales they can have a significant effect. These are then parameterised in terms of

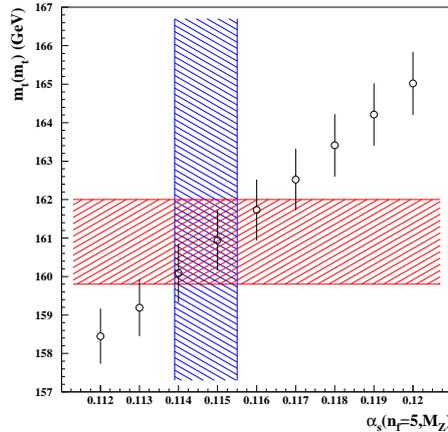


Figure 40: The \overline{MS} value of the top quark mass $m_t(m_t)$ obtained in the ABMP16 fit for variants of $\alpha_s(M_Z^2)$ (data points) and for the best fit values (hatched bands). Plot taken from [20].

1948 cubic splines defined at x_k ($k = 1 \dots 7$) points roughly linearly spaced between $x = 0$ and 1, which are then determined
 1949 from the fit. The result for the F_2 correction is shown in Fig. 39 (left), and is found to be inconsistent with zero, in
 1950 particular at higher x . The effect of these corrections, and of conversely omitting them and including a more stringent
 1951 W^2 cut on the DIS data is shown in Fig. 39 (right) for the extracted gluon PDF. The fit with the cut of $W^2 > 12.5$ and
 1952 no higher twist corrections is found to prefer a somewhat larger gluon at higher x , and in some regions lies outside the
 1953 $1-\sigma$ uncertainty band of the default fit.

1954 In addition to the higher twist corrections included in (113), the structure function functions also include target
 1955 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
 1956 according the Georgi–Politzer prescription [367] (see also [366]), with

$$F_2^{\text{TMC}}(x, Q^2) = \frac{x^2 \xi^2}{\gamma^3} F_2(x, Q^2) + 6 \frac{x^3 M_N^2}{Q^2 \gamma^4} \int_{\xi}^1 \frac{d\xi'}{\xi'} F_2(\xi', Q^2), \quad (114)$$

1957 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 $Q^2 \rightarrow \infty$ the corrected
 1958 F_i^{TMC} reduce to the regular F_i .

1959 As mentioned above, the ABMP fit in addition uses a purely fixed flavour scheme to describe the DIS data. That
 1960 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
 1961 produced at order NLO and higher. It is argued that the bulk of the DIS data can be described within this scheme. The
 1962 Tevatron and LHC collider data, on the other hand, for which $\mu_F^2 \gg m_{c,b}^2$, is treated using a 5 flavour set evolved from
 1963 the same input [82] **think this is true??**. PDF set for $n_f = 3, 4$ and 5 active flavours are made publicly available.

1964 A further feature of note is that the strong coupling α_s is determined from the fit. In ABM11 this was found to
 1965 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 $\alpha_s(M_Z^2) =$
 1966 0.1181 ± 0.0013 (the dominant uncertainty in which is determined by lattice QCD) used by the CT and NNPDF
 1967 collaborations, and the value extracted by MMHT. While in [366, 20] the omission of higher twist corrections is
 1968 found to lead to a sizeable increase in α_s , in contrast in [96] the use of the FFNS within the MSTW framework is
 1969 found to lead to a smaller extracted value consistent with that seen by ABM(P), while higher twist effects are found
 1970 to have less of an impact.

1971 In the ABM11 fit [366], the heavy quark masses $m_{c,b}$ were included in the \overline{MS} scheme for the first time, in
 1972 contrast to other PDF fits. This allows the values to be constrained directly from the PDG results without relying on
 1973 the perturbative transformation between the \overline{MS} and pole masses, which is known to be poorly convergent. Thus, the
 1974 quark masses are left free but in the fit but with the PDG values added in as pseudo–data points. For the charm mass,
 1975 the DIS data included in this fit are then found to give a comparable error to the PDG value.

1976 The latest ABMP16 fit [20] includes significant data updates, most notably for the first time hadron collider data,
 1977 from the LHC and Tevatron. In particular, a range of W and Z boson and single top production data from the Tevatron
 1978 and LHC and top pair production data at the LHC are included. In addition, the HERA I+II combined data set and
 1979 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}
 1980 scheme and is determined from the fit, giving $m_t(m_t) = 160.9 \pm 1.1$ GeV. This is consistent with the PDG value of
 1981 $160.0^{+4.8}_{-4.3}$ GeV, although this clearly has quite a large uncertainty, as it based on a single Tevatron measurement. The
 1982 result is shown in Fig. 40, with the masses extracted at different $\alpha_S(M_Z^2)$ are also given. The correlation between m_t
 1983 and α_S is clear; as discussed in [20] further information can be provided here by considering single top production
 1984 data.

1985 Interestingly, the extracted value of the strong coupling, $\alpha_S(M_Z^2) = 0.1147 \pm 0.0008$, is somewhat larger in this fit
 1986 than in the previous sets, due dominantly to the HERA I+II combined data, although this is still lower than the world
 1987 average value. The ABMP16 PDFs are available as error sets at NNLO for 3,4 and 5 fixed flavours, and for a range of
 1988 $\alpha_S(M_Z^2)$ values, from 0.112–0.120 in steps of 0.001, in the 5 flavour scheme.

1989 5.5. CJ

1990 **Jun** The CTEQ-Jefferson Lab (CJ) Collaboration has performed a series of global PDF analyses [368, 369] with
 1991 the latest PDF set CJ15 [23] and the previous generation of CJ12 [339]. The analyses are carried out at NLO in QCD
 1992 only with focus on utilizing DIS data at highest- x values applicable to perturbative QCD treatment. The kinematic
 1993 selection cuts are chosen to be $Q^2 > 1.69$ GeV² and $W^2 > 3$ GeV² so as to keep data points at low- Q and high- x region.
 1994 It results in about 1300 more data points from proton and deuteron targets, roughly 50% increase as comparing to the
 1995 standard cuts. Those additional data points provide valuable information on PDFs at large- x , e.g., $x > 0.7$, where
 1996 constraint for most of the global analyses are indirect or purely from extrapolation. Especially the deuterium data can
 1997 improve on determination of d quark at large- x which the DIS data on proton are less sensitive to.

1998 On treatment of heavy-quark mass effects in DIS structure functions, CJ12 uses a ZM-VFN scheme with heavy-
 1999 quark masses implemented as the flavor thresholds. CJ15 uses a more adequate GM-VFN scheme S-ACOT [370] to
 2000 better describe data over a wide kinematic range including the threshold regions. It is found that the implementation
 2001 of GM-VFN scheme leads to large changes on gluon PDF at large- x . Low- Q and large- x region involve further
 2002 complication on theoretical predictions of DIS structure functions. The finite Q^2 corrections to the leading-twist
 2003 calculation, i.e., power corrections of $\mathcal{O}(1/Q^2)$ must be taken into account. CJ analyses adopt the standard OPE
 2004 expression for the target mass corrections (TMCs) which allows structure functions at finite Q^2 be expressed in terms
 2005 of their massless ($M^2/Q^2 \sim 0$) values through the scaling variable $\rho^2 = 1 + 4x^2 M^2/Q^2$ [371, 372]. For other subleading
 2006 $1/Q^2$ corrections including the higher twists, they are parametrized by a phenomenological function form [23].

2007 Another important aspect is concerning the nuclear corrections for processes with deuteron targets which become
 2008 significant for intermediate and large- x region and are equally important for low and high Q values. The nuclear
 2009 corrections account for Fermi motion, binding, and nucleon off-shell effects can be implemented as convolutions
 2010 with nuclear smearing functions. In CJ12 analysis three PDF fits are provided with different models of deuteron
 2011 corrections, CJ12min, CJ12mid and CJ12max, with mild to strong corrections. The corrections are only applied at the
 2012 level of structure functions. The CJ15 analysis employ a phenomenological parametrization for part of the deuteron
 2013 corrections with free parameters fitted to data. It reduces the model dependence and increases the flexibility of the fit.
 2014 The deuteron corrections are formulated at the parton level thus can be applied to non DIS processes as well. For
 2015 example, the total quark PDF in the deuteron can be written as $q^d = q^{d(\text{on})} + q^{d(\text{off})}$, with the on-shell and off-shell
 2016 components [373, 374],

$$q^{d(\text{on})}(x, Q^2) = \int \frac{dz}{z} f^{(\text{on})}(z) q^N(x/z, Q^2), \quad q^{d(\text{off})}(x, Q^2) = \int \frac{dz}{z} f^{(\text{off})}(z) \delta f^N(x/z, Q^2) q^N(x/z, Q^2), \quad (115)$$

2017 with $q^N(x, Q^2)$ be the quark PDF in free nucleons. The on-shell and off-shell smearing functions $f^{(\text{on})}$ and $f^{(\text{off})}$
 2018 can be calculated systematically within the weak binding approximation using the deuteron wave function [375, 376].
 2019 The main CJ15 PDF fit is based on AV18 wave functions [377]. Alternative fits with CD-Bonn [378], WJC-1 and
 2020 WJC-2 [379] wave functions are also provided. Besides, the off-shell corrections $\delta f^N(x)$ Eq. (115) is parametrized
 2021 as [375]

$$\delta f^N(x) = C(x - x_0)(x - x_1)(1 + x_0 - x). \quad (116)$$

2022 The two zeros x_0 , x_1 and the normalization C are free parameters fitted to data with the constraint of maintaining the
 2023 total number of valence quarks in the nucleon. It was found that different wave function models give similar quality
 2024 of fits to the global data sets and result in PDFs well within the uncertainties since their differences can be largely
 2025 compensated by the parametrization of the off-shell corrections.

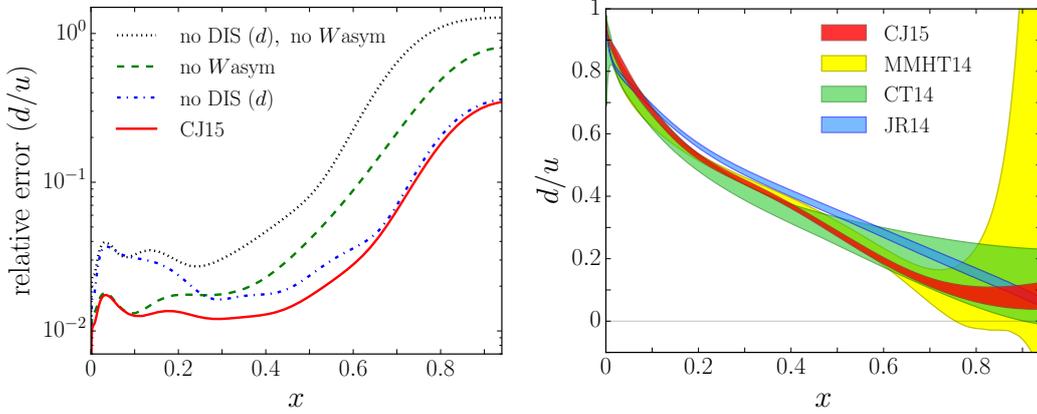


Figure 41: Left plot: relative error (90% C.L.) on the d/u PDF ratio as a function of x at $Q^2 = 10 \text{ GeV}^2$ from the CJ15 fit compared with errors obtained in fits excluding certain data sets [23]. Right plot: comparison of the d/u ratio at $Q^2 = 10 \text{ GeV}^2$ for different PDF sets, including CJ15, MMHT14, CT14, and JR14, with PDF uncertainties shown for 90% C.L. except for MMHT14 at 68% C.L. [23]

2026 In CJ15 analysis two new experimental data sets are found to have significant effects on constraining d -quark
 2027 PDF at large- x beside of the deuteron data. That includes the measurement on F_2 structure function of nearly free
 2028 neutron inside a deuterium nucleus from BONuS experiment [380, 105] at Jefferson Lab using a spectator tagging
 2029 technique. Another one is the lepton and reconstructed W boson charge asymmetry measurements from D0 Run 2 with
 2030 full luminosities [171, 172]. In Figure 41 the plot on the left shows impact of different data sets on PDF uncertainty
 2031 (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
 2032 uncertainty on d/u by almost 50%. For $x \gtrsim 0.3$ the W asymmetry data provides the dominant constraint. Besides,
 2033 the constraint from deuteron DIS data dies out for $x \gtrsim 0.6$ which turns into fit to the deuterium off-shell corrections.
 2034 In Figure 41 the plot on the right shows comparison of the d/u ratios from CJ15, MMHT14, CT14 and JR14 PDFs.
 2035 They are in good agreements when x goes to 1 considering the PDF uncertainties. The CJ15 PDF set has smaller PDF
 2036 error on d/u in general, with an extrapolated value

$$d/u \xrightarrow{x \rightarrow 1} 0.09 \pm 0.03, \quad (117)$$

2037 at the 90% C.L., due to the new data sets on constraining d -quark at large- x . On another hand with the additional
 2038 data sets that are less sensitive to the nuclear corrections, i.e., the D0 W asymmetry data and BONuS data, the CJ15
 2039 analysis is able to pin down the deuterium corrections through their interplay with the deuteron DIS data. Figure 42
 2040 gives the deuteron to isoscalar nucleon ratio F_2^d/F_2^N as from the CJ15 fits with different input of wave functions and
 2041 the 90% C.L. uncertainty for CJ15 main fit shown in the colored bands. Significant corrections are found for $x \gtrsim 0.7$.
 2042 The ratio is insensitive to the choice of wave functions since it is only the combination of the wave function and the
 2043 off-shell corrections that are subjected to constraint from current data.

2044 5.6. HERAFitter/xFitter

2045 During many years, the H1 and ZEUS collaborations performed QCD interpretation analyses of their structure
 2046 function data, first separately and then together based on the H1+ZEUS combined datasets. The backbone of these
 2047 analysis was always the neutral- and charged-current inclusive structure function measurements, in some cases sup-
 2048 plemented by the charm production structure functions and with the DIS jet cross-sections. The main results from

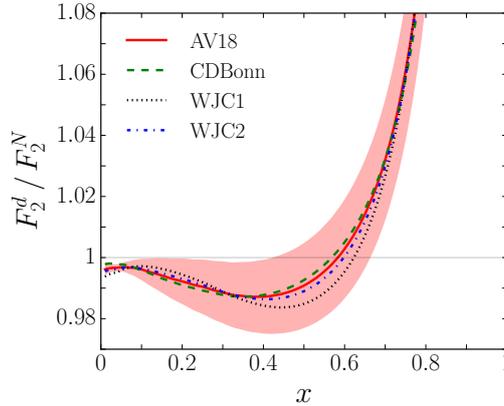


Figure 42: Ratio of deuteron to isoscalar nucleon structure functions F_2^d/F_2^N at $Q^2 = 10 \text{ GeV}$ for CJ15 fits with different models of wave functions [23]; colored band is the 90% C.L. error for CJ15 main fit (with AV18 wave functions).

2049 these studies were the HERAPDF family of PDF fits, which include HERAPDF1.0 [85], based on the Run II data,
 2050 and HERAPDF2.0 [21], based on the final combination of inclusive measurements from Runs I+II. In Fig. 43 we
 2051 show the results of the HERAPDF1.0 analysis for the u_V , d_V and S quarks and for the gluon. In the HERAPDF
 2052 methodology, the total PDF uncertainty is divided into three types of errors: experimental uncertainties, propagated
 2053 from the statistical and systematic uncertainties in the fitted data, model uncertainties, for instance due to α_s and Q_0
 2054 variations, as well as parametrization uncertainties, reflecting the spread from different comparable choices of input
 2055 functional form for the PDFs.

2056 The expertise developed by these QCD analysis of HERA structure function data eventually to the development
 2057 and release of HERAFitter [27], a publicly available open PDF fitting toolbox. HERAFitter was born as an extension
 2058 of the H1 and ZEUS internal PDF fitting codes that were extensively tested and applied in various QCD analysis of
 2059 HERA inclusive and charm data, including the various HERAPDF sets. Despite its name, HERAFitter was not
 2060 restricted to the analysis of the HERA data, and could be used for the PDF interpretation of measurements from
 2061 fixed-target DIS and proton-proton collisions as well. The flexibility of this open software tool makes also possible
 2062 to perform QCD analysis beyond unpolarized fixed-order PDF fits, such as fits of transverse-momentum dependent
 2063 (TMD) parton distributions and fragmentation functions. Recently, HERAFitter was renamed as xFitter, with the
 2064 motivation to to emphasize that this code is a general fitting toolbox not necessarily related or involving the analysis
 2065 of HERA inclusive structure function data.

2066 The xFitter framework includes modules allowing for a various theoretical and methodological options, capable
 2067 to fit a large number of relevant data sets from HERA, Tevatron and LHC. For instance, polarized and unpolarized
 2068 PDF evolution can be performed using either APFEL or QCDNUM, and a number of fixed and variable flavor number
 2069 schemes are implemented, such as the FFN scheme from OpenQCDrad and the S-ACOT, TR and FONLL general-
 2070 mass schemes. In addition to PDF fittings, a large number of their functionalities are available such as the approximate
 2071 inclusion of new datasets in existing PDF sets by means of either Bayesian reweighting or Hessian profiling, and a
 2072 wide variety of PDF plotting options. In Fig. 43 we show a schematic representation of the xFitter code structure.
 2073 The first part is the initialization, where the fit settings are specified in the steering file. This involves a number
 2074 of choices, in particular selecting the fitted datasets and the theory and methodology settings such as the specific
 2075 PDF parametrisation or the scheme for heavy quark structure functions. Then the PDF fit is performed, where the
 2076 fit parameters are determined by means of MINUIT-based minimization including the propagation of experimental
 2077 results. The final result of typical is the LHAPDF6 grid file, together with various PDF and data/theory comparison
 2078 plots.

2079 The HERAFitter/xFitter framework has been used in many ATLAS and CMS PDF interpretation studies, these
 2080 are discussed in Sects. 5.7 and 5.8 respectively. In addition, the HERAFitter/xFitter developer's team has released
 2081 a number of dedicated PDF studies, including:

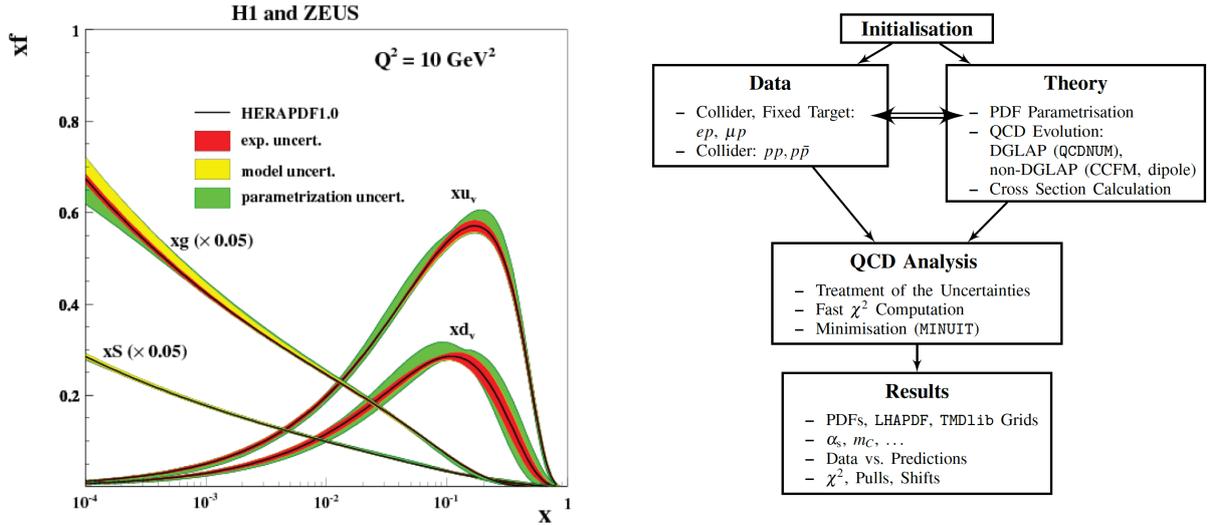


Figure 43: Left plot: the HERAPDF1.0 determination of parton distributions, based on the analysis of the combined HERA structure functions from Run I. Right plot: schematic representation of the xFitter code structure, see text for more details.

- A QCD analysis of the legacy W and Z boson production measurements at the Tevatron [319], 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 [381]. This study demonstrated that the best fit value of $m_c(m_c)$ was consistent in FONLL as compared to a fit performed in the fixed-flavor number scheme, as could be expected by theoretical considerations, as well as with previous determinations of the running mass from HERA data and with the global PDG average.
- A determination of the photon PDF $x\gamma(x, Q^2)$ [382] 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 is 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 constraints on the photon PDF at intermediate x , though the resulting PDF uncertainties were still not competitive with those from the LUXqed determination.

Additional studies based on xFitter are studies of PDFs with correlated uncertainties between different perturbative orders [383], non-DGLAP evolution equations [384], and the determination of transverse-momentum dependent PDFs [32]. In Fig. 43 we show two representative analyses of PDF-related studies performed by the xFitter Developer's Team. First of all, 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. [319]. 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. [381]. 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

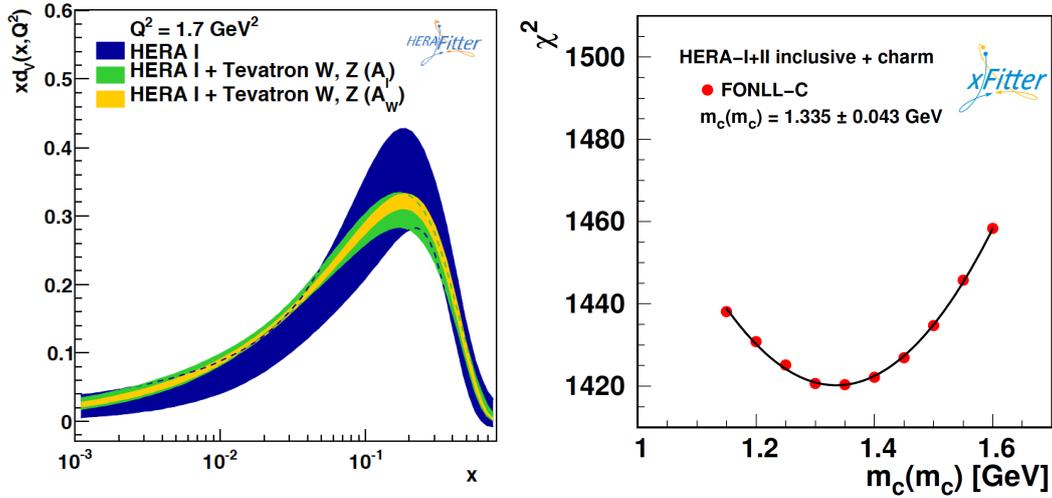


Figure 44: Two representative analyses of PDF-related studies performed by the xFitter Developer’s Team. Left plot: the impact on d_V of the Tevatron W and Z data on a HERA-only fit, comparing the impact of the lepton-level measurements with that of the boson-level measurements, from Ref. [319]. Right 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. [381]. In this analysis charm structure functions were computed with APFEL in the FONLL-C general mass scheme.

2110 and additional features are being implemented, such as the possibility of new parametrization options like Chebyshev
 2111 polynomials, the fast convolution option for hadronic cross-sections as realized in APFELgrid [297], more flexible
 2112 PDF parametrizations including the charm and the photon PDF, and improvements in the QED evolution interface.

2113 5.7. PDF efforts in ATLAS

2114 As discussed in Sect. 3, the ATLAS experiment has provided a large number of experimental measurements
 2115 with important PDF sensitivity, most of which are now part of the toolbox of global PDF fits. In addition to these
 2116 measurements, ATLAS has also developed a very active program of PDF interpretation studies, aimed to quantify
 2117 the constrains of their own measurements on the proton structure. These interpretation studies have been performed
 2118 using public PDF fitting tools, in particular the HERAFitter/xFitter framework described in Sect 5.6. These studies
 2119 represent an important contribution to the PDF fitting community, not only because they demonstrate the PDF impact
 2120 of specific measurements, but also because it provides an internal cross-check that all the information required for
 2121 PDF fits, in particular the full experimental covariance matrix, is ready to be used.

2122 A representative selection of ATLAS PDF interpretation studies is the following:

- 2123 • The ATLAS measurements of W^+ , W^- and Z rapidity distributions at 7 TeV from the 2010 dataset we used
 2124 in Ref. [212] to determine the strange content of the proton. This was allowed for the fact that the full cross-
 2125 correlations between the three rapidity distributions was accounted for, and while W^+ and W^- constrained the
 2126 up and down quarks and antiquarks, the Z measurements then fixed strangeness. This analysis found that the
 2127 strange sea was not suppressed as compared to the up and down quark sea.
- 2128 • This strangeness analysis was recently revisited in Ref. [183], which was based on the updated W^+ , W^- and Z
 2129 rapidity distributions at 7 TeV now from the 2011 dataset. Consistent results concerning the strange content of
 2130 the proton were found, again pointing out that the strange sea seems to be symmetric with respect to the light
 2131 quark sea. Given the high precision of this data, the PDF uncertainties in the strangeness determination were
 2132 significantly reduced as compared to the PDF interpretation analysis of the 2010 data. The important issue of
 2133 the strangeness content of the proton will be discussed in more detail in Sect. 6.3.

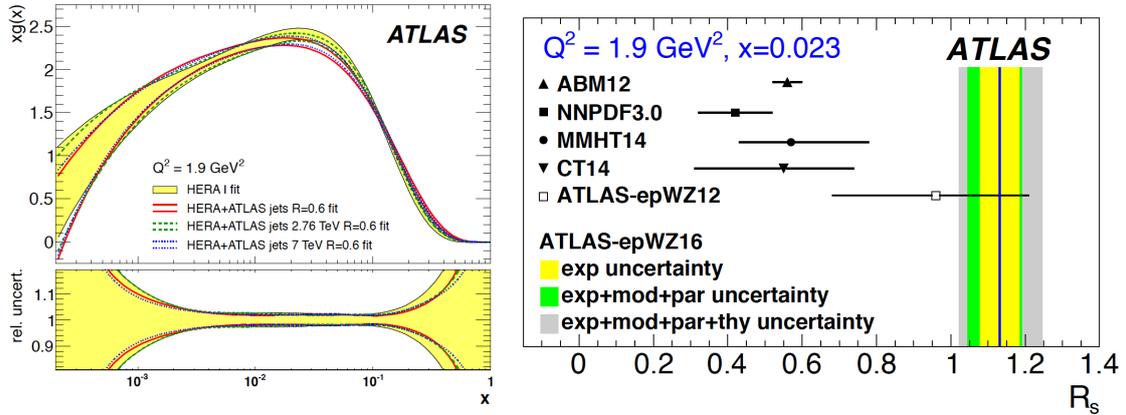


Figure 45: Two representative results of the PDF fitting efforts performed within the ATLAS collaboration. Left plot: a PDF fit quantifying the effect on the gluon from the HERA-only fit of the ATLAS inclusive jet measurements at $\sqrt{s} = 2.76$ TeV and 7 TeV, from Ref. [139]. Right plot: the determination 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.

- PDF fits based on jet production measurements have also been performed by ATLAS. For instance, in Ref. [139] a QCD analysis of HERA data supplemented with inclusive jet cross sections at $\sqrt{s} = 2.76$ TeV and $\sqrt{s} = 7$ TeV (from the 2010) run was performed. This analysis studies also the improved constraints that can be derived of 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 center-of-mass energies [385].

In Fig. 45 we show some representative results of PDF interpretation studies performed within the ATLAS collaboration. First of all we show the results of 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. [139]. Then we also show 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.

5.8. PDF efforts in CMS

The CMS collaboration have performed dedicated PDF fits in a variety of cases to assess the impact of their data. In all cases this uses the xFITTER (formerly known as HERAFITTER) platform, with the procedure for determining the PDFs following the approach of the HERAPDF fit. Thus the PDFs are parameterised at $Q_0^2 = 1.9 \text{ GeV}^2$ 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 leads to ~ 15 free parameters in the fit, with the precise number depending on the particular analysis. The TR' [386, 80] prescription for the GM-VFNS is used. 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.

In [182] the 7 TeV measurement of the W charge asymmetry, as well as $W + c$ production [387], 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. A similar impact on the valence quarks is also found in a fit to the 8 TeV differential W boson production data [211], which was performed at NNLO. In [164] 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

2162 flexible PDF parameterisation, and consistent results are found but with larger PDF uncertainties. A NLO fit to the 8
2163 TeV jet data is performed in [130], and a direct comparison to the 7 TeV case is shown, with the impact found to be
2164 very similar. In [266] a NLO fit to the 8 TeV double differential top pair production data is compared to a baseline
2165 fit that includes the 8 TeV W boson production data [211]. The impact of including the data differential in different
2166 kinematic variables is assessed, and a sizeable reduction in the uncertainty on the gluon PDF in particular is found
2167 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 **aside**
2168 **- is there double counting in this CMS study???** Finally, recent preliminary results including PDF fits to triple
2169 differential dijet production at 8 TeV [146] and the top pair production cross section at 5.02 TeV [388] have been
2170 presented.

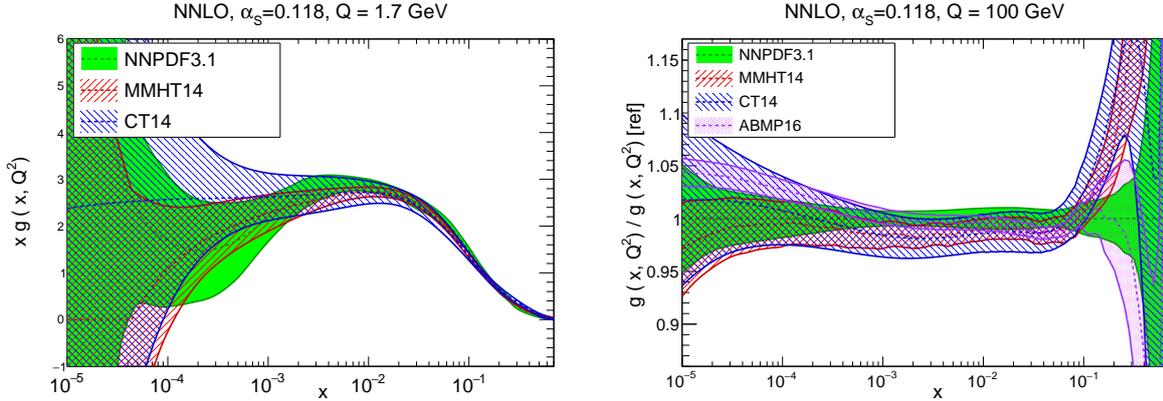


Figure 46: The gluon PDF $xg(x, Q^2)$ at $Q = 1.7$ GeV (left) and $Q = 100$ GeV (right) comparing the ABMP16, CT14, MMHT14, and NNPDF3.1 NNLO sets with $\alpha_s(m_Z) = 0.118$. In the right plots, results are normalized to the central value of NNPDF3.1.

2171 6. The proton structure

2172 Following the discussion of the various approaches to PDF fitting in the previous section, here we turn to compare
 2173 the results of the most representative state-of-the-art PDF fits among themselves. This comparison is organized in
 2174 terms of specific PDF flavour combinations relevant for phenomenology. We start discussing the gluon, then we turn
 2175 to quark flavor separation followed by the large- x behaviour of the PDFs, and subsequently we study the strange and
 2176 charm content of the proton.

2177 In the following we compare ABMP16, CT14, MMHT14, and NNPDF3.1 NNLO sets, all with $\alpha_s(m_Z) = 0.118$.
 2178 Note that for the low-scale comparisons the ABMP16 curve cannot be included since the set with $\alpha_s(m_Z) = 0.118$
 2179 is only available in the $n_f = 5$ scheme, and therefore can only be used above the bottom quark threshold. We wanted to
 2180 emphasize that here we will only show a representative number of PDF comparisons: other comparisons, including
 2181 with other PDF sets not shown here, can be easily produced using the APFEL-WEB online PDF plotting interface [389].

2182 6.1. The gluon

2183 Let us start by comparing the gluon PDF among these four PDF sets. In Fig. 46 show gluon PDF $xg(x, Q^2)$ at a low
 2184 hadronic scale $Q = 1.7$ GeV (left plot) and at a typical LHC scale $Q = 100$ GeV (right plot). From this comparison
 2185 we find that in general there is reasonable agreement between the four sets of PDF considered within uncertainties.
 2186 This is also true at small- x , where PDF uncertainties become rather large due to the lack of experimental constraints.
 2187 We note here that the MMHT14 gluon becomes negative for $x \lesssim 5 \times 10^{-5}$, while CT14 has exhibits a flat behaviour,
 2188 the NNPDF3.1 gluon shoots up. We have also verified that the agreement with ABMP16 becomes much worse if the
 2189 PDF set corresponding to their best-fit $\alpha_s(m_Z) = 0.1149$ value is used.

2190 Perhaps the most important discrepancy between the gluon PDF from the four groups is the one that arises in the
 2191 large- x region, where NNPDF3.1 (and even more markedly ABMP16) is rather softer than CT14 and MMHT14. For
 2192 instance, at $x \simeq 0.2$ the differences between the NNPDF3.1 and CT14 central values are at the $2\text{-}\sigma$ level in units of
 2193 the PDF uncertainties. One of the reasons for these differences could be related to the use of different datasets to
 2194 constrain the large- x gluon, and specifically by the use of top-quark differential distributions in the NNPDF3.1, which
 2195 have been shown to lead to a softer large- x gluon as compared to the same fit without any $t\bar{t}$ data included [276]. We
 2196 also note that in the large- x region PDF uncertainties are quite large, leading to significant theoretical uncertainties
 2197 for the production of new BSM heavy particles, as will be discussed in Sect. 8.2. In this respect, it will be interesting
 2198 to compare the large- x gluon PDF from the three global sets once they include a similar dataset.

2199 Here it is worth emphasizing that one of the most important differences of the current version of PDF fits as
 2200 compared to previous versions is that fact that several datasets provide independent constraints on the large- x gluon.
 2201 Until recently, the gluon at large- x was only constrained in the PDF fit by inclusive jet cross-sections, and to a lesser
 2202 extent by DIS data via scaling violations. On the other hand, now we have at least three datasets that allow to

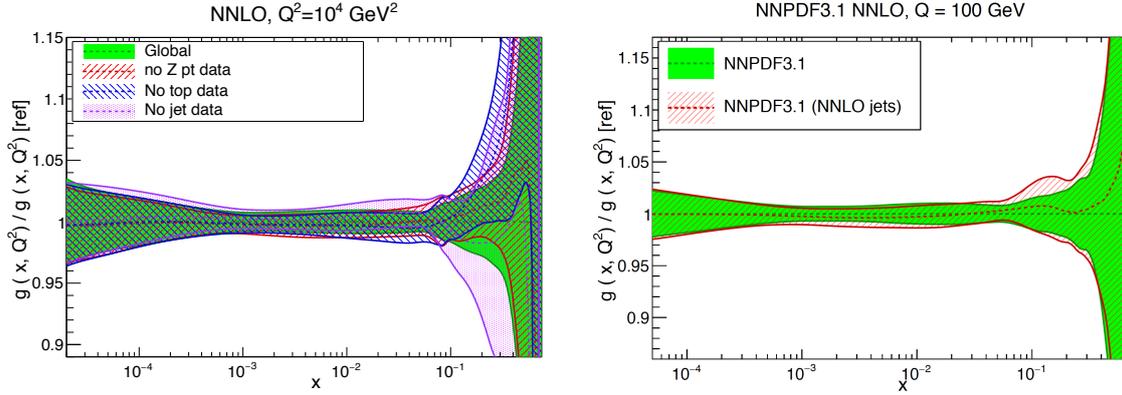


Figure 47: Left: comparison of the NNPDF3.1 NNLO global fit at $Q = 100$ GeV with the corresponding fits where the $Z p_T$, top quark, or inclusive jet data have been removed. Right plot: same as before, now comparing with the NNPDF3.1 NNLO fit where the ATLAS and CMS 7 TeV inclusive jet data have been treated using exact NNLO theory.

2203 robustly pin down the large- x gluon, and for the three of them NNLO calculations are available: inclusive jets, the p_T
 2204 distribution of Z bosons, and top quark differential distributions. To illustrate the robustness of the resulting gluon,
 2205 in Fig. 47 we show a comparison of the NNPDF3.1 NNLO global fit at $Q = 100$ GeV with the corresponding fits
 2206 where the $Z p_T$, top quark, or inclusive jet data have been removed. We can observe that the four fits agree within PDF
 2207 uncertainties, highlighting that these three families of physical processes have statistically consistent pulls on the large
 2208 x gluon. This is certainly an important result, given the relevance of the large- x gluon for many LHC applications.

2209 Another point directly to the large- x gluon in the global fit are the settings for the theoretical calculation of
 2210 inclusive jet cross-sections. Until 2016, only the NLO calculation has available, and different groups treated jet data
 2211 in the global fit in different ways, from adding the NLO scale errors as additional systematic uncertainties as CT14
 2212 and NNPDF3.1, to using the threshold approximation [158] to the full NNLO result as MMHT14, to excluding jet
 2213 data altogether as advocated by ABMP16. The availability of the complete NNLO calculation makes these different
 2214 approaches obsolete, and future iterations of the various PDF fits will inclusive jet data using always exact NNLO
 2215 theory. This said, there is evidence that, at least for specific settings of the NLO calculation, the inclusion of jet data
 2216 with NNLO theory has a moderate phenomenological relevance. This is because, if the jet p_T scale is adopted as
 2217 central renormalization and factorization scale, and a not too small value of R is used, the NNLO/NLO K -factor is
 2218 a few percent at most. To illustrate this point, in Fig. 47 we compare the baseline NNPDF3.1 NNLO fit (where jet data
 2219 are treated at NLO, with scale uncertainties as additional systematic error) to the same fit where exact NNLO theory
 2220 is used for the ATLAS and CMS 7 TeV data. We see that the resulting differences are small at the PDF level, and at
 2221 the χ^2 level one finds [239] a small but non-negligible improvement once NNLO theory is used.

2222 6.2. Quark flavor separation

2223 Next we turn to discuss the proton's quark flavour separation, namely the individual quarks and antiquarks. In
 2224 this respect, the use of fixed-target and collider Drell-Yan data provides a unique handle to disentangle quarks and
 2225 antiquarks, since inclusive DIS data is only sensitive to the total quark singlet combination $\Sigma(x, Q^2)$ (and also to the
 2226 triplet $T_3 = u + \bar{u} - d - \bar{d}$ if deuteron structure functions are included as well). To begin with, we show in Fig. 48 the up,
 2227 down, anti-up, and anti-down quark PDFs for the four PDF sets that we are comparing, at $Q = 100$ GeV. Let us start
 2228 by discussing $u(x, Q)$, which is one of the better constrained PDFs, specially at large- x , due to the fixed-target DIS
 2229 data. We find good agreement between uncertainties in the entire range of x , with the only exception being ABMP16,
 2230 which overshoots the other three sets in the large- x region.

2231 Concerning the down quark PDF $d(x, Q)$, here the spread between the central values of the four sets is larger
 2232 than for up quarks, and PDF uncertainties are also comparatively increased. Here we find good agreement between
 2233 CT14, MMHT14, and NNPDF3.1 within uncertainties for the entire range of x , while ABMP16 is around 5% lower
 2234 than the central NNPDF3.1 value at intermediate values of x . PDF uncertainties are the largest at high- x , with CT14

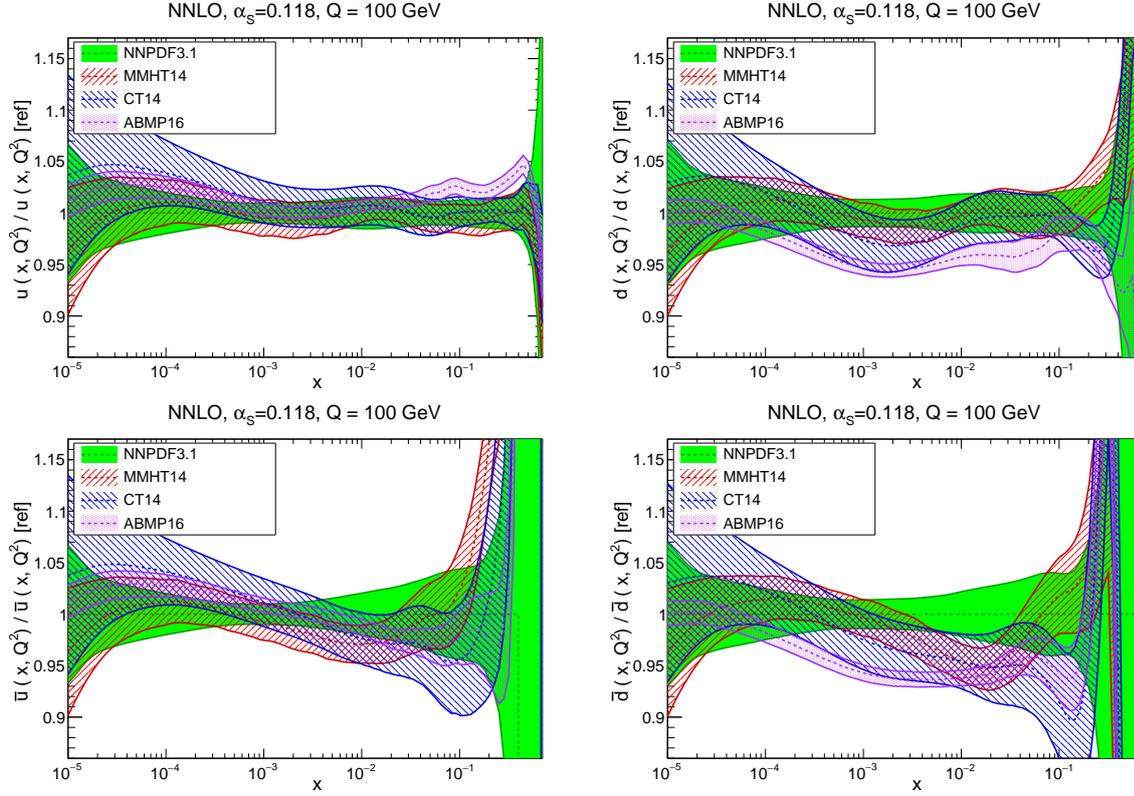


Figure 48: Same as Fig. 46 (right), now comparing the up, down, anti-up, and anti-down quark PDFs.

2235 and MMHT14 on the one hand and ABMP16 on the other hand pointing to different directions, with the NNPDF3.1
 2236 central value somewhere in the middle. Note that here one of the possible sources of differences between groups is
 2237 the treatment of deuteron nuclear corrections in the fitting of F_2^d structure functions [19], though this effect is known
 2238 to be localized in the region around $x \simeq 0.1$ [390].

2239 As for the light antiquark PDFs, $\bar{u}(x, Q)$ and $\bar{d}(x, Q)$, the comparisons of Fig. 48 indicate that there is a reasonable
 2240 agreement between the various sets within uncertainties for \bar{u} , while this agreement is marginal for \bar{d} . In the latter
 2241 case, the ABMP16 result is again around $\simeq 5\%$ smaller than the NNPDF3.1 central value. As in the case of the
 2242 quark PDFs, also for the antiquarks we see large differences at large- x : this is a region with limited experimental
 2243 constraints, and thus the methodological differences in each PDF fit can have a rather more marked impact here. As in
 2244 the case of the gluon, these large PDF uncertainties of the antiquarks at high- x have phenomenological consequences,
 2245 for instance for the production of a heavy W' or Z' boson, or for the pair production of a squark-antisquark pair $\tilde{q}\tilde{q}^*$,
 2246 both processes driven by the quark-antiquark luminosity.

2247 Another interesting way to compare the quark flavour separation between the various PDF group is to plot flavour
 2248 combinations that can be directly related to physical cross-sections. In Fig. 49 we compare the sea quark asymmetry
 2249 $\Delta_S = \bar{d} - \bar{u}$ and the quark isovector $T_3 = u + \bar{u} - d - \bar{d}$ at $Q = 1.7$ GeV between CT14, MMHT14, and NNPDF3.1.
 2250 The former flavour combination is closely related to the W asymmetries in Drell-Yan production at colliders, while
 2251 the latter is directly sensitive to the difference between the proton and deuteron DIS structure functions, $F_2^p - F_2^d$.
 2252 From this comparison, we see that for Δ_S the general shape is shared among the three groups, though there are large
 2253 differences in the estimate of the PDF uncertainties, both at small- x and at large x , which in some cases can be traced
 2254 back to the PDF parametrization assumptions. The agreement is reasonably good both in terms of central values and
 2255 of uncertainties for xT_3 , although here also the small- x behaviour does differ among the three groups.

2256 From the previous comparisons we have found that the differences in the quark flavour separation between the

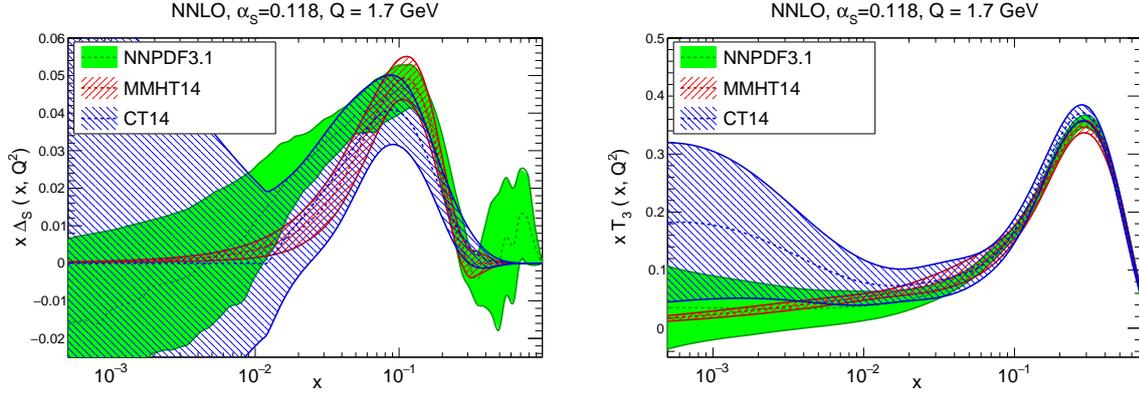


Figure 49: Same as Fig. 46 (left), now comparing the sea quark asymmetry $\Delta_S = \bar{d} - \bar{u}$ (left) and the quark isotriplet $T_3 = u + \bar{u} - d - \bar{d}$ (right plot).

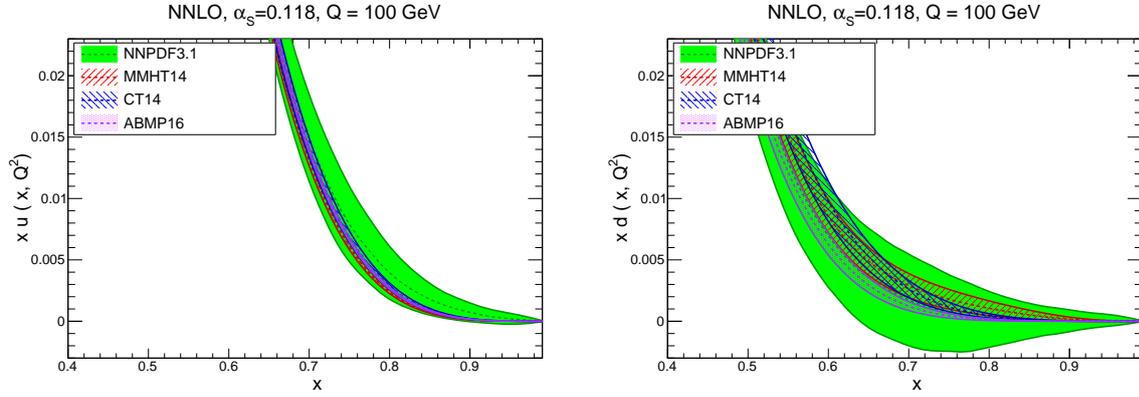


Figure 50: Same as Fig. 48, now focusing on the large- x region of the up quark (left) and down quark (right plot) PDFs.

2257 various groups are mostly localized in the large- x region. With this motivation, in Fig. 50 we show again the up and
 2258 down quark PDFs, now focusing on the large- x region, and therefore using a linear scale in the x axis. From this
 2259 comparison we see that PDF uncertainties are the largest in NNPDF3.1. In terms of central values, there is reasonable
 2260 agreement for u , less so for d . Note that in the NNPDF fits the PDFs are not forced to be positive (though the physical
 2261 cross-sections are indeed positive-definite) and therefore the down PDF can become negative at large- x , although its
 2262 central value becomes always positive. Let us also mention that an alternative approach to compare the behaviour of
 2263 PDF sets at large- x , and in doing so comparing with non-perturbative models such as the quark counting rules, is the
 2264 effective exponent method discussed in Ref. [300].

2265 6.3. The strange content of the proton

2266 Next we turn to discuss the issue of the strange content of the proton, in the light of the latest experimental
 2267 constrains both from fixed-target neutrino-nucleus scattering and from precision collider electroweak gauge boson
 2268 production data. The size and shape of the strange PDF has attracted a lot of debate recently. On the one hand, most
 2269 PDF fits find a suppressed strangeness as compared to the non-strange light quark sea, a pull driven mostly by the
 2270 deep-inelastic neutrino inclusive F_2 and charm production (“dimuon”) data. On the other hand, recent high-precision
 2271 collider data from the LHC has instead exhibited the opposite trend, with a recent QCD analysis from ATLAS based
 2272 on the W, Z 7 TeV rapidity distributions from the 2011 dataset finds even a strange sea larger than the non-strange sea.

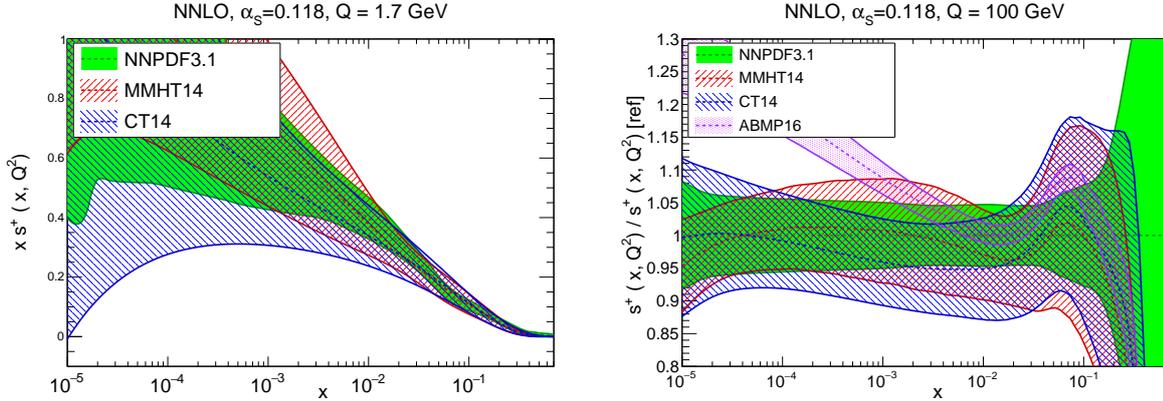


Figure 51: Same as Fig. 46 for the total strangeness $x s^+(x, Q^2)$.

2273 Given the importance of strangeness for many phenomenological applications, for instance the measurement of the W
 2274 mass, it would be important to be able to shed some light on this issue.

2275 To begin with, in Fig. 51 we show the total strange PDF $x s^+(x, Q^2)$ at $Q = 100$ GeV, comparing the results
 2276 from the different PDF groups, in the same format as that of Fig. 46. We mention here that the strangeness-sensitive
 2277 datasets included in by the four analysis are rather different, both in terms of the neutrino fixed-target data and in
 2278 terms of the LHC collider data. For instance, only the ABMP16 fit includes the NOMAD dimuon data [391], while
 2279 only NNPDF3.1 includes the ATLAS W, Z 2011 rapidity distributions. From this comparison we see that there is
 2280 reasonable agreement within uncertainties between the four groups except for ABMP16 for $x \lesssim 10^{-3}$, which has
 2281 a much harder strangeness than the other three groups. We also note that the differences in the size of the strange PDF
 2282 uncertainty can vary by up to a factor ~ 5 , with ABMP16 having the smallest uncertainties.

2283 A more physically adequate method to assess how sizable is the strange content of the proton is provided by the
 2284 ratio of the strange to the non-strange sea quark PDFs, defined as

$$R_s(x, Q^2) = \frac{s(x, Q^2) + \bar{s}(x, Q^2)}{\bar{u}(x, Q^2) + \bar{d}(x, Q^2)}. \quad (118)$$

2285 In this ratio, a symmetric strange sea would correspond to $R_s \simeq 1$, while a suppressed strangeness instead leads to
 2286 $R_s \ll 1$. Traditionally, the constraints from neutrino dimuon production have suggested a value $R_s \simeq 1/2$ is most
 2287 global fits. In Fig. 52 we show the ratio of strange to non-strange sea quarks $R_s(x, Q^2)$, Eq. (118), as a function of x for
 2288 $Q^2 = 1.9$ GeV². We compare the results of various global PDF fits with those of the ATLAS/xFitter interpretation
 2289 study as well as with those of a NNPDF3.1 fit based on the same dataset as the ATLAS study. The vertical lines
 2290 indicate the two possible scenarios for the strange PDFs, namely a suppression of size $R_s \simeq 0.5$ and then a strange
 2291 sea which is symmetric with the non-strange one, $R_s \simeq 1$. We can see that the ABMP16, CT14, MMHT14 and
 2292 NNPDF3.1 have a preference for a suppressed strangeness. On the other hand, this comparison also shows that if
 2293 only the HERA and ATLASWZ11 data are considered, the NNPDF3.1 analysis yields an unsuppressed strangeness,
 2294 although with PDF uncertainties rather larger than for the xFitter analysis. This comparison demonstrates that
 2295 the opposite pull between the low-energy neutrino data and the high-energy collider data is genuine, although the
 2296 tension is not dramatic, as indicated by the fact that the NNPDF3.1 global and HERA+ATLASWZ11 results agree
 2297 within PDF uncertainties.

2298 One limitation of the comparison summarized in Fig. 52 is that it is restricted to a specific point in x , namely
 2299 $x \simeq 0.023$. To compare with the situation for the rest of x range, in Fig. 53 we show the $R_s(x, Q^2)$ ratio as a function
 2300 of x both a low and at high scales. There are a number of interesting features that can be drawn from this comparison.
 2301 First of all, we observe that DGLAP evolution automatically increases the value of R_s , since as we go to higher
 2302 values of Q the sea component dominates over the valence components. Second, we find that the four groups that are
 2303 being compared here point to a consistent strangeness content in most of the range of x , although the corresponding
 2304 uncertainties in each case can change quite a lot.

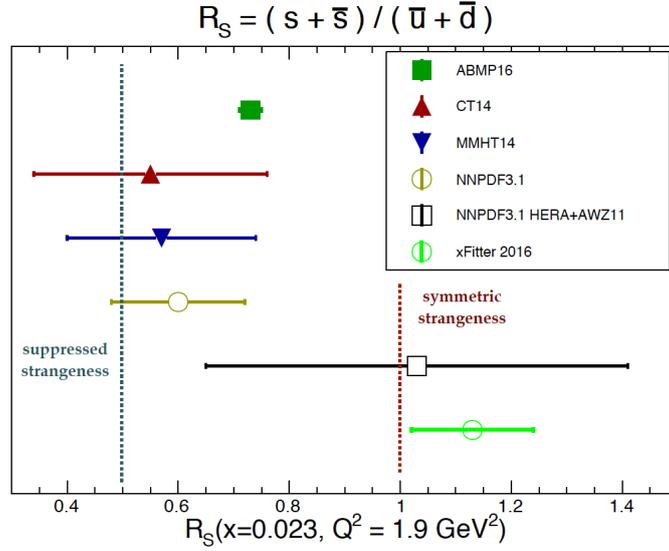


Figure 52: The ratio of strange to non-strange sea quarks $R_S(x, Q^2)$, Eq. (118) for $x = 0.023$ and $Q^2 = 1.9 \text{ GeV}^2$. We compare the results of various global PDF fits with those of the ATLAS/xFitter interpretation study as well as with those of a NNPDF3.1 fit based on the same dataset as the ATLAS study. The vertical lines indicate two possible scenarios for the strange PDFs, namely a suppression of size $R_S \approx 0.5$ and then a strange sea which is symmetric with the non-strange one, $R_S \approx 1$.

2305 Another important point of this comparison is that one can observe that any statement about whether or not
 2306 strangeness is suppressed depends on the region of x that is being considered. For instance, in the MMHT14
 2307 analysis for $Q = 1.38 \text{ GeV}$ the value of R_S changes from around 0.4 at $x \approx 0.1$ to around 0.8 for $x \approx 0.007$. So different x
 2308 regions exhibit different amounts of suppression with respect to the light sea quarks, and therefore the question of the
 2309 suppression (or lack thereof) of the strange PDF is a more nuanced issue than what is sometimes stated. In any case is
 2310 clear from the comparison of Fig 53 that a symmetric strange sea in the entire range of x is not favoured by any of the
 2311 four fits shown here, in particular in the region around $x \approx 0.1$ and above. In this respect, future data from the LHC
 2312 should be useful to shed some light on this important issue.

2313 6.4. The charm content of the proton

2314 We complete this anatomy of the internal structure of the protons by discussing its charm content, a topic that
 2315 has received quite a lot of attention recently. As discussed in Sect. 4, there are two different approaches to treat the
 2316 charm PDF in the global PDF analysis. On the one hand, one can assume that the charm PDF is generated entirely
 2317 from perturbative evolution, and thus compute the charm PDF from the gluon and light quark PDFs starting from the
 2318 charm threshold $\mu_c \approx m_c$. On the other hand, it is also possible to release this assumption and treat the charm PDF
 2319 on an equal footing to the light quark PDFs, namely introducing a parametrization to be fixed by experimental data.
 2320 Until recently, in most global PDF fits the charm was generated by perturbative evolution, and then separately, in
 2321 dedicated intrinsic charm studies, variants of the global fit with some model for the charm PDF were performed, with
 2322 the parameters of the model being constrained by experimental data, see for instance Refs [392, 342, 26, 393]. The
 2323 NNPDF3.1 global analysis instead fits the charm PDF in their baseline fits, using exactly the same parametrization as
 2324 for the rest of the light quark PDFs.

2325 The amount of charm present inside the proton can be quantified by its momentum fraction, defined as

$$C(Q^2) \equiv \int_0^1 dx x (c(c, Q^2) + \bar{c}(x, Q^2)). \quad (119)$$

2326 In the case of perturbative charm, by construction one has $C(Q^2 < \mu_c^2) = 0$, while if there is a non-perturbative
 2327 charm component in the proton in general $C(Q^2) \neq 0$ at all scales. In Sect. 54 we show the momentum fraction

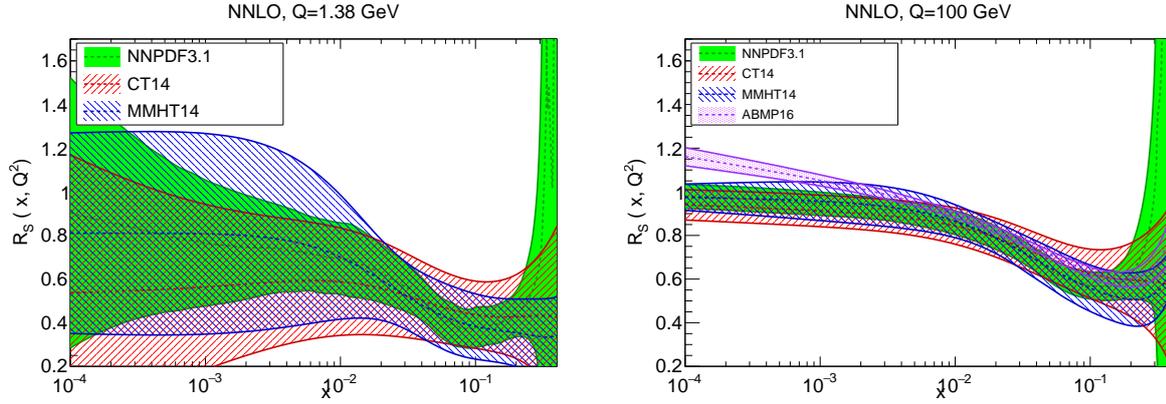


Figure 53: The ratio of strange to non-strange sea quarks $R_s(x, Q^2)$, Eq. (118), as a function of x for $Q = 1.38$ GeV (left plot) and for $Q = 100$ GeV (right plot).

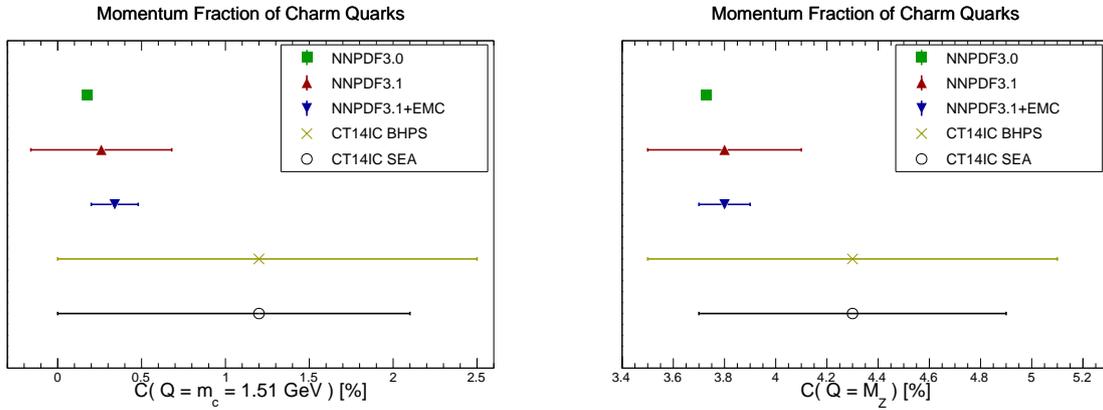


Figure 54: The momentum fraction carried by charm quarks, $C(Q)$, both at a low scale $Q = 1.51$ GeV (left) and at a high scale $Q = M_Z$ (right plot). We compare NNPDF3.0 (perturbative charm) with NNPDF3.1 (based on fitted charm) with and without the inclusion of the EMC charm data, as well as with the BHPS and SEA scenarios of the CT14IC fits. See text for more details.

2328 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
 2329 NNPDF3.0, based on perturbative charm, with those from NNPDF3.1, based on fitted charm, with and without the
 2330 inclusion of the EMC charm data, as well as with the BHPS and SEA scenarios of the CT14IC fits. In the latter case,
 2331 the uncertainty bands indicates the range between no intrinsic charm and the maximum amount of IC allowed within
 2332 the CT14 tolerance, with the central value corresponding to the best fit value.

2333 The comparison of Fig. 54 highlights first all that while when the charm PDF is generated perturbatively its
 2334 uncertainties are very small, this is not necessarily the case once it is fitted. As required by consistency, the results
 2335 of the NNPDF3.1 analysis include the NNPDF3.0 perturbative charm results. We also see that while adding the EMC
 2336 charm structure functions to the global analysis helps r

2337 **7. QED corrections and the photon PDF**

2338 In this section we explore a topic that have received a lot of attention in PDF fits in the recent years, namely the
 2339 role of QED and weak corrections, in particular concerning photon-initiated processes. Here first of all we discuss
 2340 the role of QED corrections and the photon PDFs, and then we review pure weak corrections to hard scattering matrix
 2341 elements arising from virtual massive weak boson exchange.

2342 *7.1. Photon-induced processes*

2343 *Introduction*

2344 It has been over a decade since the calculation of the splitting functions at NNLO in α_s [78, 79] provided the
 2345 necessary tools to be able to carry out NNLO PDF fits. Moreover, we have seen in Sect. 3 that for the majority of
 2346 PDF sensitive observables, the perturbative calculation calculation is available at this NNLO order. Given that the
 2347 data from the LHC are available at increasing precision, to below the percent level, NNLO PDF fits are essential to
 2348 match this unprecedented precision and have naturally become the standard. However, a simple counting of powers
 2349 of α_s indicates that

$$\alpha_s^2(M_Z) \sim \frac{1}{70}, \quad \alpha_{\text{EM}}(M_Z^2) \sim \frac{1}{130}. \quad (120)$$

2350 That is, we may roughly expect the NNLO QCD and NLO EW corrections to be of the same order of magnitude.
 2351 While such an argument clearly neglects the non-trivial differences in the structure of the QCD and EW corrections,
 2352 this nonetheless serves as a warning that we must at least consider the impact of going to NLO EW if we are to claim
 2353 percent-precision to LHC cross-sections.

2354 A specific type of EW correction of particular relevance to PDF studies is the contribution from photon-initiated
 2355 processes, such as those shown schematically in Fig. 55. As this involves a photon in the initial state, this requires
 2356 the introduction of a PDF for the photon in the proton⁶. This is included in complete analogy to the QCD partons,
 2357 and moreover as it involves a massless boson in the initial-state, higher order QED $q \rightarrow q\gamma$ and $\gamma \rightarrow q\bar{q}$ splitting will
 2358 generate collinear singularities that must be absorbed into the corresponding PDFs. In other words, this will produce
 2359 QED corrections to the DGLAP evolution of the PDFs. Another important type of EW corrections relevant for PDF
 2360 fits, namely those associated to virtual massive weak boson exchange, are discussed in Sect. 7.2.

2361 *QED corrections to DGLAP evolution*

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

$$\begin{aligned} Q^2 \frac{\partial}{\partial Q^2} g(x, Q^2) &= \sum_{q, \bar{q}, g} P_{ga}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{g\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2), \\ Q^2 \frac{\partial}{\partial Q^2} q(x, Q^2) &= \sum_{q, \bar{q}, g} P_{qa}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{q\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2), \\ Q^2 \frac{\partial}{\partial Q^2} \gamma(x, Q^2) &= P_{\gamma\gamma} \otimes \gamma(x, Q^2) + \sum_{q, \bar{q}, g} P_{\gamma a}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2). \end{aligned} \quad (121)$$

2362 The splitting functions can then be expanded in powers of both the QCD and QED coupling

$$P_{ij} = \sum_{m,n} \left(\frac{\alpha_S}{2\pi} \right)^m \left(\frac{\alpha}{2\pi} \right)^n P_{ij}^{(m,n)}. \quad (122)$$

2363 The lowest order QED splitting function $P_{\gamma q}^{(0,1)}$ is due to the same Feynman diagram as LO QCD case $P_{gq}^{(1,0)}$, with
 2364 the gluon simply replaced by a photon, and similarly for P_{qq} and $P_{q\gamma}$. Thus these are trivially related by suitable

⁶For 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. [394].

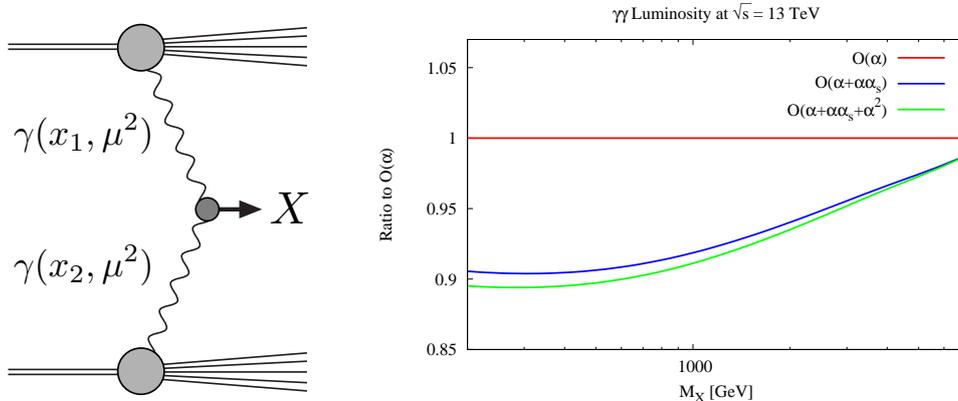


Figure 55: Left: schematic diagram for the photon–initiated production of a system X , and the corresponding photon PDFs. Right: the $\gamma\gamma$ luminosity as a function of the invariant mass, M_X , of the produced final-state, taken from [382]. The ratio to results with $O(\alpha\alpha_S)$ and $O(\alpha^2)$ to the leading $O(\alpha)$ DGLAP evolution shown.

adjustments of the colour factors and inclusion of the electric charges e_q of the quark, with

$$P_{qq}^{(0,1)} = \frac{e_q^2}{C_F} P_{qq}^{(1,0)}, \quad P_{q\gamma}^{(0,1)} = \frac{e_q^2}{T_R} P_{qg}^{(1,0)}, \quad P_{\gamma q}^{(0,1)} = \frac{e_q^2}{C_F} P_{gq}^{(1,0)}, \quad P_{\gamma\gamma}^{(0,1)} = -\frac{2}{3} \sum_f e_f^2 \delta(1-x), \quad (123)$$

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 leptons⁷. The calculation of the $O(\alpha_S)$ ($m, n = 1$) terms, where the $P_{g\gamma}$ and $P_{\gamma g}$ splittings enter for the first time, is given in [396], while the $O(\alpha^2)$ ($m = 0, n = 2$) terms are given in [397]. Publicly available implementations of the DGLAP evolution including these QED corrections are provided by the APFEL [398] and QEDEVOL [399] packages.

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

$$\mathcal{L}_{\gamma\gamma} = \frac{1}{s} \int_{M_X^2/s}^1 \frac{dx}{x} \gamma(x, M_X^2) \gamma\left(\frac{M_X^2}{xs}, M_X^2\right), \quad (124)$$

is shown in Fig. 55. The $O(\alpha_S)$ corrections have a fairly small but clearly non-negligible impact on the luminosity, giving up to a $\sim 10\%$ negative correction. As we would expect, the $O(\alpha^2)$ corrections are an order of magnitude smaller, but can reach the percent level. In both cases, the impact of these corrections decreases with increasing M_X , due to the smaller relative contribution from the PDF evolution in comparison to the input photon, see (127) below. Note that in general these corrections depend quite a bit on PDF set used, for instance the results of Fig. 55 have been computed using NNPDF3.0QED, while the ratio to the $O(\alpha)$ result for $M_X = 200$ GeV is $R \simeq 0.96$ is the HKR16 photon PDF is used instead.

As discussed in detail in [400], 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 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 calculation. Physically, the photon substructure is being resolved by the introduction of a photon PDF and the contribution from $\gamma \rightarrow q\bar{q}$ 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 [401], where it is shown that using the on-shell scheme leads to uncanceled fermion-mass singularities in the hard cross section.

⁷To be consistent, and in particular to preserve momentum fully, this requires the introduction of lepton PDFs in the proton. However as discussed in [395] 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.

2386 *The photon PDF*

2387 The first attempts at describing the photon PDF can be divided into two distinct categories, either being phe-
 2388 nomenological approaches that model the photon PDF, as in the MRST2004QED [402] and more recent CT14QED [343]
 2389 sets, or the data-driven approach of the NNPDF2.3/3.0QED [403, 404] sets. The first attempt to include the photon
 2390 in a PDF set was provided by MRST2004QED. Here, a simple model for the photon PDF at input scale Q_0 due to
 2391 one-photon emission off the valence quarks was assumed. In other words, the quark valence distributions were frozen
 2392 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

$$\gamma(x, Q_0^2) = \frac{\alpha}{2\pi} \left[\frac{4}{9} \log\left(\frac{Q_0^2}{m_u^2}\right) u(x, Q_0) + \frac{1}{9} \log\left(\frac{Q_0^2}{m_d^2}\right) d(x, Q_0) \right] \otimes \frac{1 + (1-x)^2}{x}. \quad (125)$$

2393 The CT14QED set generalised this approach, allowing additional freedom in the normalization of the photon, which
 2394 was fit to ZEUS data [344] on isolated photon production⁸. Thus, within such approaches the photon PDF is com-
 2395 pletely *predicted* within the specific model, up to any freedom in the model parameters, such as the choice of quark
 2396 masses for MRST2004QED or the overall normalization for CT14QED.

2397 On the other hand, the NNPDF2.3QED [403] set (subsequently updated to NNPDF3.0QED [404]), freely pa-
 2398 rameterises the photon PDF at input. In other words, the photon is treated on exactly the same footing as the QCD
 2399 partons. This is then extracted from a fit (or more precisely, a Bayesian reweighting) to DIS and LHC W, Z data; in the
 2400 former case the constraint comes purely from the effect on the PDF evolution, with no explicit photon-initiated con-
 2401 tribution included. However, in general the contribution of photon-initiated process are small, leading to significant
 2402 uncertainties on the extracted photon PDF.

2403 More recently, there has been a great deal of progress in our understanding of the photon PDF. The crucial point
 2404 that was missed in the above approaches is the long range nature of QED. That is, the proton is itself an electrically
 2405 charged object which can coherently emit a photon, with the proton remaining intact afterwards. The possibility for
 2406 such elastic photon emission is of course very well established. Elastic ep scattering is an extremely well measured
 2407 process, providing for example the first measurement of the proton charge radius [33, 34] in the 1950s, with further
 2408 precise measurements of this process [405] continuing to this day. Theoretically, the well known equivalent photon
 2409 approximation (EPA) [406] provides a precise foundation for describing the elastic scattering process in terms of a
 2410 flux of coherently emitted photons from the proton.

2411 The connection of this fact to the photon PDF was discussed in [407] and more recently in [408, 409]. Following
 2412 the equivalent photon approximation, it is straightforward to show that elastic photon emission leads to a contribution
 2413 to the photon PDF at a scale $Q_0 \sim 1$ GeV given by

$$\gamma_{\text{el}}(x, Q_0^2) = \frac{1}{x} \frac{\alpha}{\pi} \int_{\frac{x^2 m_p^2}{1-x}}^{Q_0^2} \frac{dQ^2}{Q^2} \left[\left(1 - x - \frac{x^2 m_p^2}{Q^2} \right) F_E(Q^2) + \frac{x^2}{2} F_M(Q^2) \right], \quad (126)$$

2414 where F_E and F_M are the elastic and magnetic form factors of proton, which are related to the electric and magnetic
 2415 charge distributions in the proton. These are steeply falling functions of Q^2 that are probed very precisely in a range
 2416 of elastic ep scattering experiments, see e.g. [405].

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. (121) to get [410]

$$\gamma(x, \mu^2) = \frac{1}{\alpha(\mu^2)} \left(\alpha(Q_0^2) \gamma(x, Q_0^2) + \int_{Q_0^2}^{\mu^2} \frac{dQ^2}{Q^2} \alpha(Q^2) \sum_{q, \bar{q}, g} P_{\gamma a}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) \right), \quad (127)$$

$$\equiv \gamma_{\text{input}}(x, \mu^2) + \gamma_{\text{evol}}(x, \mu^2). \quad (128)$$

2417 Thus the photon is given separately in terms of an input at low scale Q_0 and an evolution component due to the usual
 2418 DGLAP $q \rightarrow q\gamma$ emission for $Q^2 > Q_0$. The latter is completely determined in terms of the quark and gluon PDFs,

⁸In fact, as we will discuss below, this has been supplemented with the elastic component to give the inclusive set CT14QEDinc.

2419 leaving the input photon at Q_0 , which is dominantly due to elastic emission. Thus this already provides quite a strong
 2420 constraint on the photon PDF; as we will see below, the impact in comparison to the model-independent NNPDF
 2421 approach can be dramatic.

2422 However, even for relatively low photon virtualities, $Q^2 < Q_0$, the emission may also be inelastic, such that the
 2423 proton no longer remains intact afterwards. In other words we have

$$\gamma(x, Q_0^2) = \gamma_{\text{el}}(x, Q_0^2) + \gamma_{\text{inel}}(x, Q_0^2), \quad (129)$$

2424 In [408, 409] fairly simple phenomenological models for this inelastic component, given by suitable generalizations
 2425 of (125), were taken, while the CT14QED set allowed an additionally free normalization to be fitted to ZEUS data on
 2426 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 γ_{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 as⁹

$$\begin{aligned} x\gamma(x, \mu^2) = & \frac{1}{2\pi\alpha(\mu^2)} \int_x^1 \frac{dz}{z} \left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\mu^2} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \left[\left(z p_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2} \right) F_2\left(\frac{x}{z}, Q^2\right) - z^2 F_L\left(\frac{x}{z}, Q^2\right) \right] \right. \\ & \left. - \alpha^2(\mu^2) (z^2 + z p_{\gamma q} \ln(1-z)) F_2\left(\frac{x}{z}, \mu^2\right) \right\}. \end{aligned} \quad (130)$$

2427 Thus in this formalism the photon PDF is a derived quantity, which can be written purely in terms of the inclusive
 2428 DIS structure functions, which are known quite precisely from the experimental point of view.

2429 While the connection of Eq. (130) to the considerations above is not immediately clear, some similarity in the
 2430 overall form with Eq. (127) is apparent. Indeed, by substituting for $F_{2,L}$ in terms of the quark and gluon PDFs, at high
 2431 Q^2 this readily reduces to γ_{evol} in (127); this is precisely how the LUXqed photon PDF is calculated in this region. In
 2432 addition, using the known expression for the elastic contributions to $F_{2,L}$ reproduces Eq (126) when combined with
 2433 Eq. (127); this elastic contribution is also included. By using fits to the experimentally determined inelastic structure
 2434 functions at low Q^2 , including in the resonance region, it is shown in [24] that the remaining inelastic contribution,
 2435 and therefore the photon PDF in its entirety, is very precisely determined. In Fig. 56 we show an overview of the
 2436 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
 2437 that at small- x it is dominated by the PDF contribution, while at large- x the elastic contribution accounts for up to half
 2438 of the size of $\gamma(x, Q)$.

2439 It is worth emphasizing that the expression Eq. (130) does not rely on any explicit separation between an input and
 2440 evolution component to the photon as in (127), and corresponds to the exact result for the photon within the quoted
 2441 accuracy of [24]. Indeed, applying standard DGLAP above the starting scale Q_0 terms the power-like $\sim m_p^2/Q^2$
 2442 correction would be missed, while for $Q^2 > Q_0^2$ the contribution from the elastic component would be omitted and the
 2443 inelastic resonance component, which contributes at higher x in this region, would not be correctly modelled.

2444 However, from the point of view of a global PDF set it may be preferable to use Eq. (130) in a form that can
 2445 be more directly implemented within the standard fitting framework. That is, applying this approach after suitable
 2446 modification to calculate the input photon, which can then be included as part of the default input parton set for any
 2447 future fits and studies, see [354] for initial discussion.

2448 To illustrate the differences and similarities between these various determinations of $\gamma(x, Q)$, in Fig. we show
 2449 the comparison of the photon PDFs from CT14qed.inc, MRST2004, NNPDF2.3/3.0 and LUXqed, normalized to the
 2450 central value of the latter. It is clear from this comparison that the theoretical uncertainties associated to the LUXqed
 2451 determination are much smaller than in any other of the previous approaches. Interestingly, the LUXqed is in good
 2452 agreement within uncertainties with the model-independent NNPDF3.0QED fit in the entire range of x , though the
 2453 PDF uncertainties are much larger in the latter case specially at large x .

⁹Following the publication of [24] it was discovered that this expression had been already derived in the earlier papers of [411, 412], but without the correct limits on the Q^2 integral.

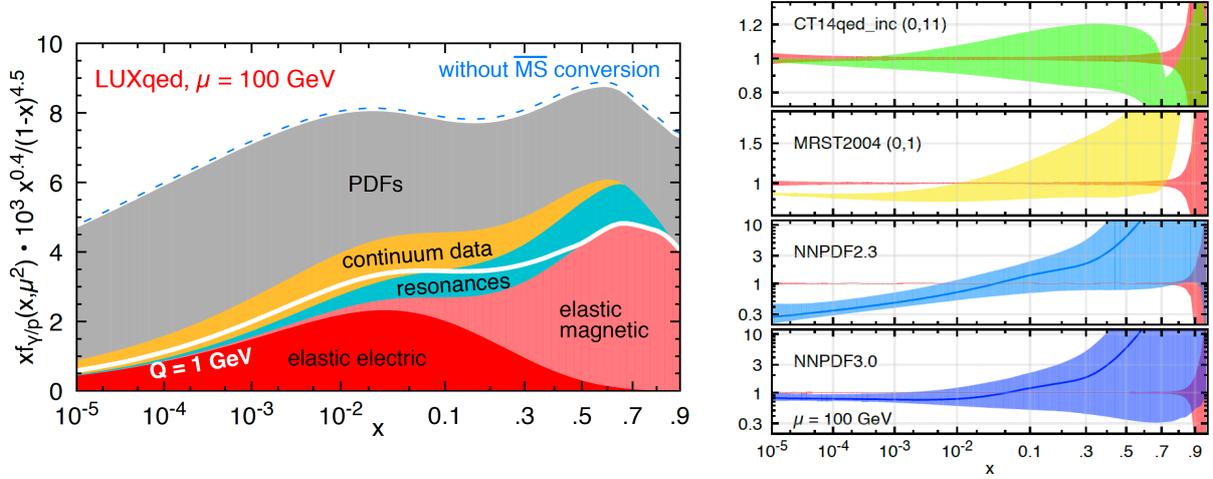


Figure 56: Left: overview of the various contributions to the photon PDF $\gamma(x, Q^2)$ in the LUXqed approach as a function of x at $Q = 100$ GeV. Right: comparison of the photon PDFs from CT14qed_inc, MRST2004, NNPDF2.3/3.0 and LUXqed, normalized to the central value of the latter.

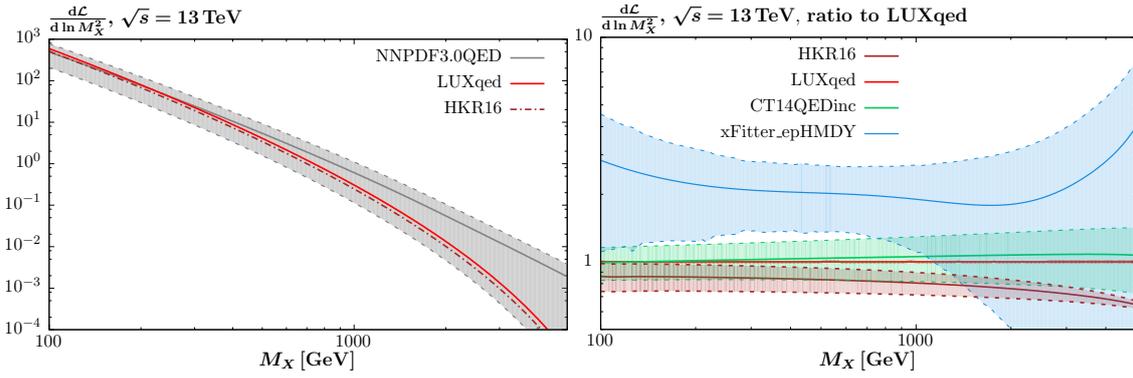


Figure 57: $\gamma\gamma$ luminosities at $\sqrt{s} = 13$ TeV. (Left) Absolute values for the HKR16, NNPDF3.0QED and LUXqed sets. (Right) Ratios of the CT14QED, HKR16 and xFitter_HMDYep sets to the LUXqed prediction. 68% uncertainty bands are shown, with the exception of the HKR16 set, where the error is due to model variation in the inelastic input (lower edge corresponds to elastic only).

2454 Phenomenology

2455 In Fig. 57 (right) we show the NNPDF3.0QED $\gamma\gamma$ luminosity at $\sqrt{s} = 13$ TeV, including the 68% C.L. error bands.
 2456 A large PDF uncertainty is evident, in particular at higher system mass M_X . As discussed above, the input component
 2457 in Eq. (128) is poorly determined within this approach, due to limited constraints placed by the available experimental
 2458 data. It is therefore unsurprising that the PDF errors should be most significant at higher mass, as here the relative
 2459 contribution from this input component is larger, due the reduced phase space for PDF evolution. In addition, the
 2460 central value of the luminosity is seen to lie towards the upper end of the uncertainty band. As discussed in [15, 409],
 2461 this exhibits a much gentler decrease with M_X in comparison to the QCD parton luminosities. However, also plotted
 2462 is the LUXqed result, and the difference is dramatic. The central value lies towards the lower end of the NNPDF band
 2463 at higher mass, with a PDF uncertainty that is smaller than the line width of the plot. We also show the result of [409],
 2464 labelled HKR16, which demonstrates a similar trend. Thus, simply applying basic physical conditions on the form of
 2465 the photon PDF, and including the dominant coherent input Eq. (126) gives a qualitatively similar result.

2466 Taking a closer look, in Fig. 57 (right) we show the ratio of the HKR16, CT14QED and xFitter_HMDYep results to
 2467 the LUXqed luminosity. The xFitter_HMDYep set is extracted in [382] by applying a similar agnostic methodology

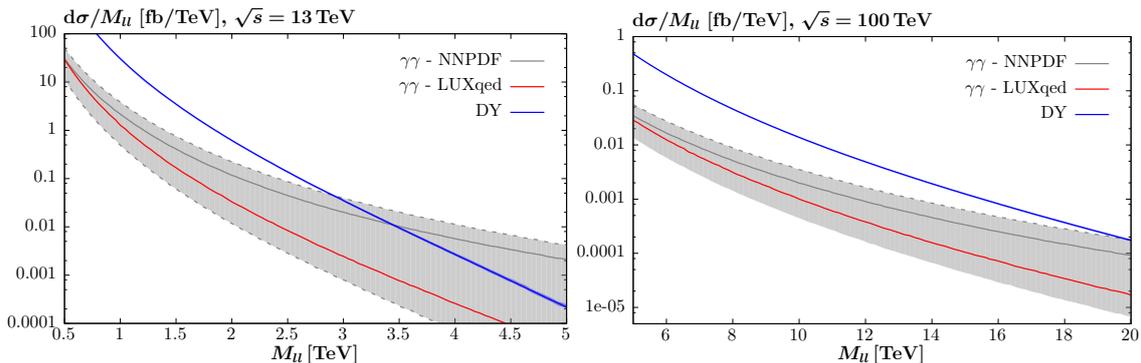


Figure 58: The differential lepton pair production cross sections at $\sqrt{s} = 13$ TeV and 100 TeV with respect to the invariant mass of the pair M_{ll} , for lepton $|\eta| < 2.5$ and $p_{\perp} > 20$ GeV. The photon-initiated contributions predicted using the LUXqed and NNPDF3.0QED sets, including the 68% C.L. uncertainty bands. The NLO Drell–Yan cross section, calculated with MCFM [207], is also shown.

2468 to NNPDF, but including the more constraining ATLAS high mass Drell–Yan data [413] in the fit; this therefore
2469 represents the most up to date set within such an approach. Again, the LUXqed uncertainty band is barely visible on
2470 the curve, varying from 1 – 2% over the considered mass interval. The CT14QEDinc prediction, which includes an
2471 elastic component, is consistent, but with larger $\sim 20 - 40\%$ uncertainties, due to the more limited constraints placed
2472 by the ZEUS isolated photon production data on the inelastic input¹⁰. The HKR prediction lies somewhat below the
2473 LUXqed result, outside of the quoted model variation band, in particular at larger M_X . This is due in large part to the
2474 lack of any explicit resonant contribution in the inelastic input, which becomes more important at higher x and hence
2475 M_X . Finally, the xFitter_HMDYep set is seen to have a sizeable uncertainty band (albeit smaller than the NNPDF3.0
2476 set [382]), and interestingly appears to lie somewhat above the LUXqed result. From this it is clear that any attempt
2477 to extract the photon PDF within such an approach will almost certainly not be competitive. More generally, we can
2478 see that the LUXqed set exhibits by far the smallest PDF uncertainties.

2479 Prior to these most recent developments, a range of phenomenological studies pointed out similar trends in the
2480 NNPDFQED predictions for the photon-initiated contributions to lepton and W pair [15, 414, 415] and $t\bar{t}$ [273]
2481 production. At high system invariant mass these could be significant, and even dominant over the standard channels,
2482 with a large PDF uncertainty. From Fig. 57 the reason for this is clear, being driven by the large PDF uncertainty in
2483 the $\gamma\gamma$ luminosity at high mass, and the relatively gentle decrease with mass in the central value. However, from the
2484 discussion above we know that using the NNPDF set will dramatically overestimate the uncertainty on the photon-
2485 initiated contribution. In Fig. 58 we show the lepton pair production cross section at high mass, at the $\sqrt{s} = 13$
2486 TeV LHC and a $\sqrt{s} = 100$ TeV FCC. We can see that indeed at the LHC, the NNPDF prediction for the photon-
2487 initiated contribution can even be larger than the standard Drell–Yan contribution. However, the up-to-date LUXqed
2488 prediction exhibits no such behaviour. The prediction is under good theoretical control, and gives a small, though not
2489 negligible, contribution.

2490 Thus, by considering the physics that generates the photon PDF, and recognising the dominance of the elastic emis-
2491 sion process, we already achieve a significant reduction in PDF uncertainty in comparison to the model-independent
2492 approach, even when accounting for the most sensitive data available in the latter case. Moreover, the additional
2493 information provided by Eq. (130) in combination with the high precision data on the inelastic (and elastic) proton
2494 structure functions provides extremely tight constraints on the photon PDF, resulting in a $\sim 1\%$ level PDF uncertainty.
2495 It is worth emphasising that while consistency tests are of course to be encouraged, this is not the result of a particular
2496 theoretical model, to be treated on the same footing as older PDF sets. The LUXqed set is a fundamentally experi-
2497 mental determination of photon; it is simply that by doing this directly in terms of the measured structure functions
2498 the tightest constraints can be achieved. Such information must be included in any future photon PDF set, and we

¹⁰In 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.

2499 have therefore moved beyond the era of large photon PDF uncertainties. Indeed, the photon PDF, which used to be
 2500 the poorest known of all the proton PDFs, is now the one characterized by the smallest uncertainties.

2501 7.2. Electroweak corrections

2502 In addition to the pure QED photon-initiated corrections discussed above, it can also be important to include in
 2503 the global fit other types of EW contributions, in particular those arising from virtual EW bosons. These corrections
 2504 are most important at larger invariant masses of the produced system, $Q \gg M_W$, where virtual EW effects receive
 2505 logarithmic enhancements, see Ref. [416] for a review. Specifically, the virtual exchange of soft or collinear weak
 2506 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
 2507 (negative) corrections for large values of Q . Given that many of the LHC datasets that enter into the global PDF have
 2508 sensitivity to the TeV region, from high-mass Drell-Yan production to the large p_T tail of Z production and inclusive
 2509 jets and dijets, the inclusion of pure EW corrections is in general required to achieve the best possible description of
 2510 experimental data in this region.

2511 The state-of-the-art of EW corrections is NLO, that is, $\mathcal{O}(\alpha_W^2)$, including as well in some cases mixed terms of the
 2512 form $\mathcal{O}(\alpha_W \alpha_s)$ and related terms. These corrections are available for most of the hadron collider processes that enter
 2513 a typical global fit. In particular, they are available for inclusive jet and dijet production [157], inclusive electroweak
 2514 gauge boson production at high p_T [234, 255] and high invariant mass $M_{ll(\nu)}$ [198, 197] and differential top quark pair
 2515 production [273, 272], among others. Most of these calculations are implemented in publicly available programs. For
 2516 instance, EW corrections to inclusive gauge boson production are available in programs such as FEWZ [198, 197] and
 2517 HORACE [203]. The latest version of MCFM [417] also allows the calculation weak corrections to Drell-Yan, top-quark
 2518 pair, and dijet production at hadron colliders. Recently, there has also been progress in the automation of the calcula-
 2519 tion of these corrections, both in the framework of MadGraph5_aMC@NLO [418] and of Sherpa/OpenLoops [419].

2520 In Fig. 59 we show two representative examples of NLO EW corrections for processes relevant for PDF determi-
 2521 nations, computed with MCFM at $\sqrt{s} = 13$ TeV [417]. In the left plot, we show the percentage NLO EW correction for
 2522 high-mass dilepton production as a function of M_{ll} , comparing also with the corresponding ZGRAD calculation. We
 2523 see that these corrections are negligible for $M_{ll} \lesssim 500$ GeV, but that they can become significant as we increase M_{ll} ,
 2524 for instance they become $\delta_{\text{wk}} \sim -20\%$ at 5 TeV. In the right plot, we show the same quantity, now for dijet production
 2525 as a function of the invariant mass of the dijet M_{jj} . The two curves correspond to two possible ways to combine NLO
 2526 QCD and EW corrections, known as additive (δ_{add}) and multiplicative (δ_{prod}). Here the corrections are more moderate
 2527 (since the Born is a pure QCD process) but they can still become up to few percent in the region accessible at the
 2528 LHC. The results of Fig. 59 illustrate how a careful inclusion of NLO EW corrections is important for the description
 2529 of the LHC data in the TeV region used for PDF determinations.

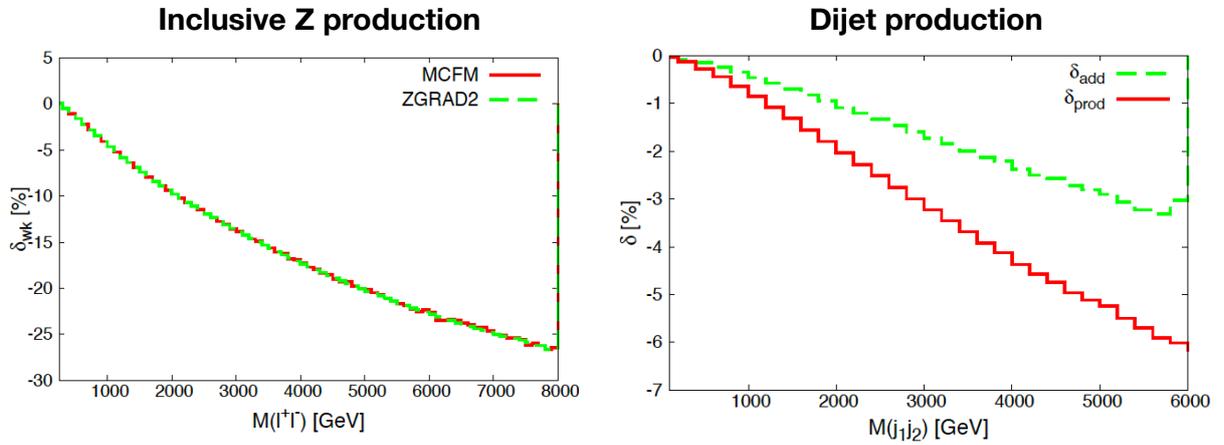


Figure 59: Two representative examples of NLO EW corrections for processes relevant for PDF determinations, computed with MCFM at $\sqrt{s} = 13$ TeV [417]. In the left plot, we show the percentage NLO EW correction for high-mass dilepton production as a function of M_{ll} , comparing also with the corresponding 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 (δ_{add}) and multiplicative (δ_{prod}).

2530 8. Implications for LHC phenomenology

2531 In this section we present an overview of some of the most representative implications of PDFs and their uncer-
2532 tainties for LHC phenomenology. First of all we discuss the role of PDFs for the predictions of the Higgs boson
2533 production cross-sections at the LHC. Then we assess what is the role of PDF uncertainties for the searches of new
2534 heavy resonances predicted by various Beyond the Standard Model scenarios. And we complete this section we high-
2535 light the importance of PDFs for the precision measurements of SM parameters such as the W mass or the strong
2536 coupling constant.

2537 8.1. Higgs production cross-sections

2538 In the Standard Model, once the Higgs mass is measured, all other parameters of the Higgs sectors, such as the
2539 strength it is coupling to fermions and vector bosons, are uniquely determined. On the other hand, deviations of
2540 these Higgs couplings with respect to the SM predictions are expected in generic BSM scenarios. Therefore, the
2541 precision measurements of the couplings of the Higgs boson represents a unique opportunity for BSM searches,
2542 since any deviation with respect to the the tightly fixed SM predictions would represent a *smoking gun* for New
2543 Physics. Crucially, realizing this program requires not only high precision experimental measurements of Higgs
2544 boson production and decay in various channels, but also the calculation of the SM cross-sections and decay rates
2545 with matching accuracy. And in this respect, PDFs are one of the largest sources of theoretical uncertainty affecting
2546 Higgs boson production cross-sections [10].

2547 Here we present a comparison of inclusive Higgs production cross-sections at 13 TeV with the latest releases of
2548 all PDF groups. The settings of this comparison, and the codes used for each process, are described in Ref. [97].
2549 Specifically, we show the dominant Higgs boson production modes at hadron colliders: gluon fusion, associated
2550 production with a $t\bar{t}$ pair, VZ associate production, and Higgs production in vector-boson fusion. In addition, we
2551 also show the results for double Higgs production in the dominant gluon-fusion channel. Results are provided for
2552 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
2553 we also indicate the result corresponding to their best fit value of $\alpha_s(m_Z) = 0.1149$. The theoretical settings for each
2554 cross-section calculation are based on state-of-the-art matrix element calculations, for instance the gluon-fusion and
2555 VBF results are computed at N3LO using the `ggHiggs` [420] and `vbf@n3lo` [421] codes respectively. We only show
2556 here the PDF uncertainties, and other sources of theoretical errors affecting these cross-sections are listed in *e.g.* the
2557 latest Higgs Cross-Section Working Group report [10]. Interestingly, the uncertainty associated to the input value of
2558 $\alpha_s(m_Z)$ can be comparable to the PDF uncertainties in some channels.

2559 There are a number of interesting features in the comparison of Fig. 60. First of all, it shows that in general there
2560 is good agreement between the three global fits, NNPDF3.1, CT14 and MMHT14 for all the Higgs boson production
2561 modes. The comparison between NNPDF3.1 and its predecessor NNPDF3.0 highlights good agreement for the gluon
2562 initiated channels, with a reduction of the PDF uncertainties in the former case, while for quark-initiated cross-sections
2563 such as VH and VBF there is an upper shift by around one sigma. Another remarkable feature of this comparison is
2564 that the recent ABMP16 set is also in reasonable with the rest of the groups, provided that the same common value of
2565 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
2566 in the calculation, there can significant differences in cross-sections, specially for the gluon initiated contributions.
2567 Specifically, ABMP16 is in this case around 7% (10%) lower than NNPDF3.1 for the gluon-fusion ($t\bar{t}$ associated
2568 production) cross-section.

2569 It is also worth mentioning here that PDF uncertainties are relevant not only for the extraction of Higgs couplings
2570 from inclusive cross-sections, but also for the differential measurements that will become available thanks to the large
2571 statistics that will be accumulated by the end of Run II as well as for the HL-LHC. To illustrate this point, in Fig. 61
2572 we show the PDF uncertainties in the p_T^h distribution of Higgs bosons produced in the gluon-fusion mode at the LHC
2573 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
2574 at around the $\sim 2\%$ level. However, these uncertainties will be increases as the LHC becomes sensitive to higher p_T
2575 values: as shown in the right plot of Fig. 61, at high invariant masses (high p_T values) the PDF uncertainties in the
2576 gluon-gluon luminosity become more significant.

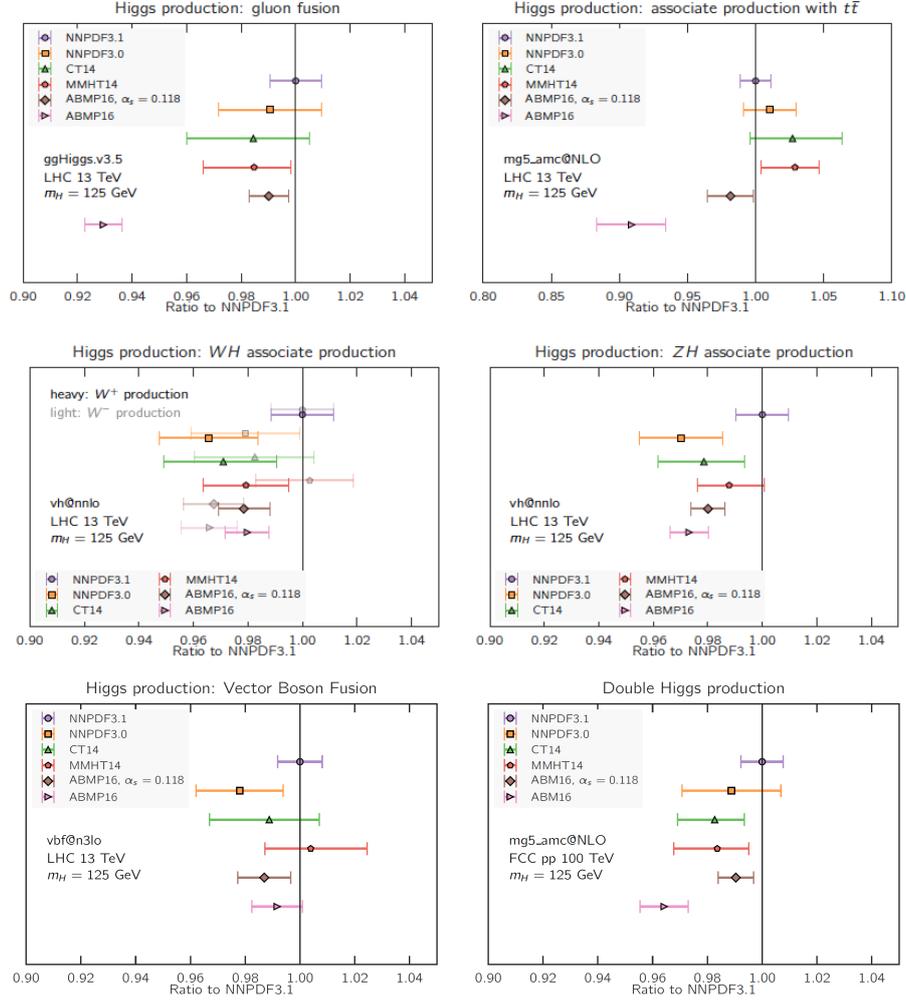


Figure 60: The PDF dependence of the most important Higgs production inclusive cross-sections at the LHC 13 TeV. The results are normalized to the central value of NNPDF3.1, and only PDF uncertainties are shown. See text for more details of the theoretical calculation.

2577 8.2. PDF uncertainties and searches for new massive particles

2578 Many scenarios of new physics beyond the Standard Model predict the existence of new heavy particles with
 2579 masses around the TeV scale. Among many others, supersymmetry, composite Higgs, and extra dimensions, are
 2580 some of the scenarios that motivate the search for new heavy resonances at the LHC in the high-mass tail of various
 2581 kinematic distributions. In this respect, PDF uncertainties play an important role in setting robust exclusion limits
 2582 based on available null results, and would become even more important in the case of eventual discovery. The reason
 2583 for this is that PDFs represent the dominant theoretical uncertainty for the production of new heavy particles in the
 2584 TeV region, since these processes are sensitive to the large- x behaviour of quarks and gluons. And as discussed in
 2585 Sect. ??, PDF uncertainties are large in this region due to the limited amount of experimental constraints.

2586 In order to quantify the size of the PDF uncertainties in the large invariant mass region relevant to these specific
 2587 BSM searches, as well as the relative agreement between the PDF groups, it is useful to compare the PDF luminosities
 2588 for $M_X \geq 1$ TeV. In this comparison we will restrict ourselves to ABMP16, CT14, MMHT14 and NNPDF3.1, in all
 2589 cases using $\alpha_s(m_Z) = 0.118$. Results are shown in Fig. 62 for $\sqrt{s} = 13$, TeV normalized to the central value of the
 2590 MMHT14 calculation. From the comparison in Fig. 62, we find that PDF uncertainties are small, at the few-percent

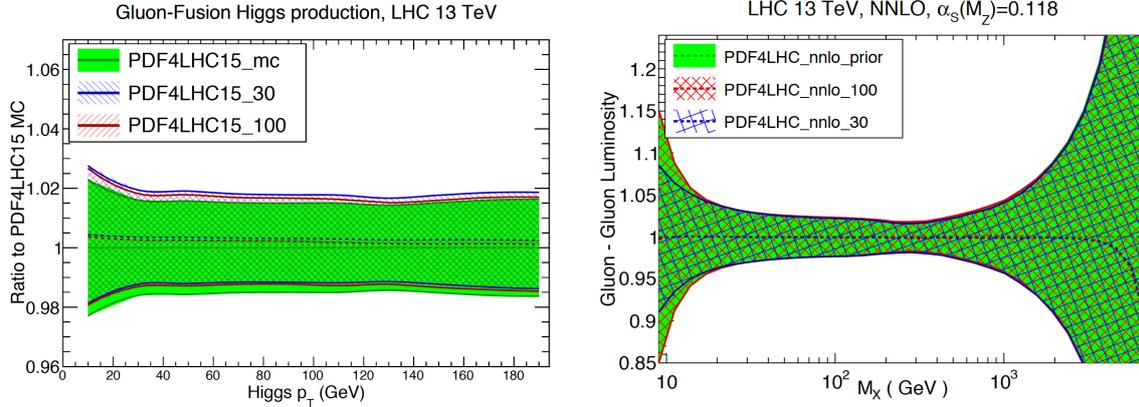


Figure 61: Left: the PDF uncertainties in the p_T^h distribution of Higgs bosons produced in the gluon-fusion mode at the LHC 13 TeV for $0 \leq p_T^h \leq 200$ GeV, computed using the PDF4LHC15 sets. Right: the gluon-gluon PDF luminosity with the same set now extending up to higher values of the invariant mass of the final state M_X .

2591 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
 2592 by the rather accurately known up and down quark valence PDFs, which are constrained *e.g.* by fixed-target DIS
 2593 structure functions/ For all the other flavour combinations, PDF uncertainties are much larger, in particular for the
 2594 quark-antiquark and gluon-gluon PDF luminosities.

2595 In the case of the gluon-gluon luminosity, we find a rather large spread of the predictions between the different
 2596 groups, with MMHT14 (ABMP16) leading to the largest (smallest) values of \mathcal{L}_{gg} . For instance, at $M_X \sim 5$ TeV, which
 2597 is close to the upper limit of the kinematic coverage of the LHC, PDF uncertainties are almost $\mathcal{O}(100\%)$. Even more
 2598 moderate invariant masses such as $M_X \sim 2.5$ TeV, the values of \mathcal{L}_{gg} can vary between approximately +10% and
 2599 -30% as compared to the central MMHT14 result. It is thus clear that these uncertainties would represent one of the
 2600 limiting factors BSM characterization in the case of an eventual discovery. In the case of the quark-gluon luminosity,
 2601 we observe a similar trend as in the gluon-gluon one but with reduced PDF uncertainties, due to the contribution of
 2602 the well-constrained large- x quark PDFs.

2603 PDF uncertainties, as well as differences between groups, are also large for the quark-antiquark PDF luminosity
 2604 $\mathcal{L}_{q\bar{q}}$, also shown in Fig. 62. The reason for these behaviour is two-fold. On the one hand, large- x anti-quarks are
 2605 notoriously difficult to pin down, although recent high-precision measurements from the Tevatron and the LHC are
 2606 improving the situation. On the other hand, various groups parameterize the quark sea content of the proton with rather
 2607 different assumptions [300], with the corresponding implications for the quark-antiquark luminosities. We find that the
 2608 spread of the differences results ranges between +5% and -30% for $M_X = 3$ TeV, with PDF uncertainties becoming
 2609 $\mathcal{O}(100\%)$ for higher invariant masses. Note here that the PDF uncertainties are the largest for NNPDF3.1, despite
 2610 this being the global analysis which includes a largest amount of LHC electroweak data sensitive to anti-quarks. This
 2611 highlights the fact that methodological differences in the flavour assumptions and parametrization of anti-quarks are
 2612 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 .

2613 In order to illustrate the phenomenological consequences of these large PDF uncertainties at high M_X , in Fig. 63
 2614 we show the PDF uncertainties for high-mass graviton production in the Randall-Sundrum scenario [422, 423] induced
 2615 by gluon-fusion at the LHC 8 TeV, computed with MadGraph5 [424]. We compare the results of the NNPDF2.3 fit
 2616 with those of the same fit including the constraints from top-quark production cross-sections. We observe that PDF
 2617 uncertainties become $\mathcal{O}(100\%)$ at large values of the graviton mass, consistent with the estimates from the gluon-gluon
 2618 luminosity shown Fig. 62. We also see how these PDF uncertainties can be reduced by the inclusion of top quark pair
 2619 production total cross-sections, highlighting the cross-talk between precision SM measurements and improving BSM
 2620 searches.

2621 In Fig. 63 we also show the K -factor for the NLO+NLL cross-sections including PDF uncertainties, normalized
 2622 to the NLO result, for the production of a squark-anti-squark pair $\tilde{q}\tilde{q}^*$ at the LHC 13 TeV with various PDF sets, from

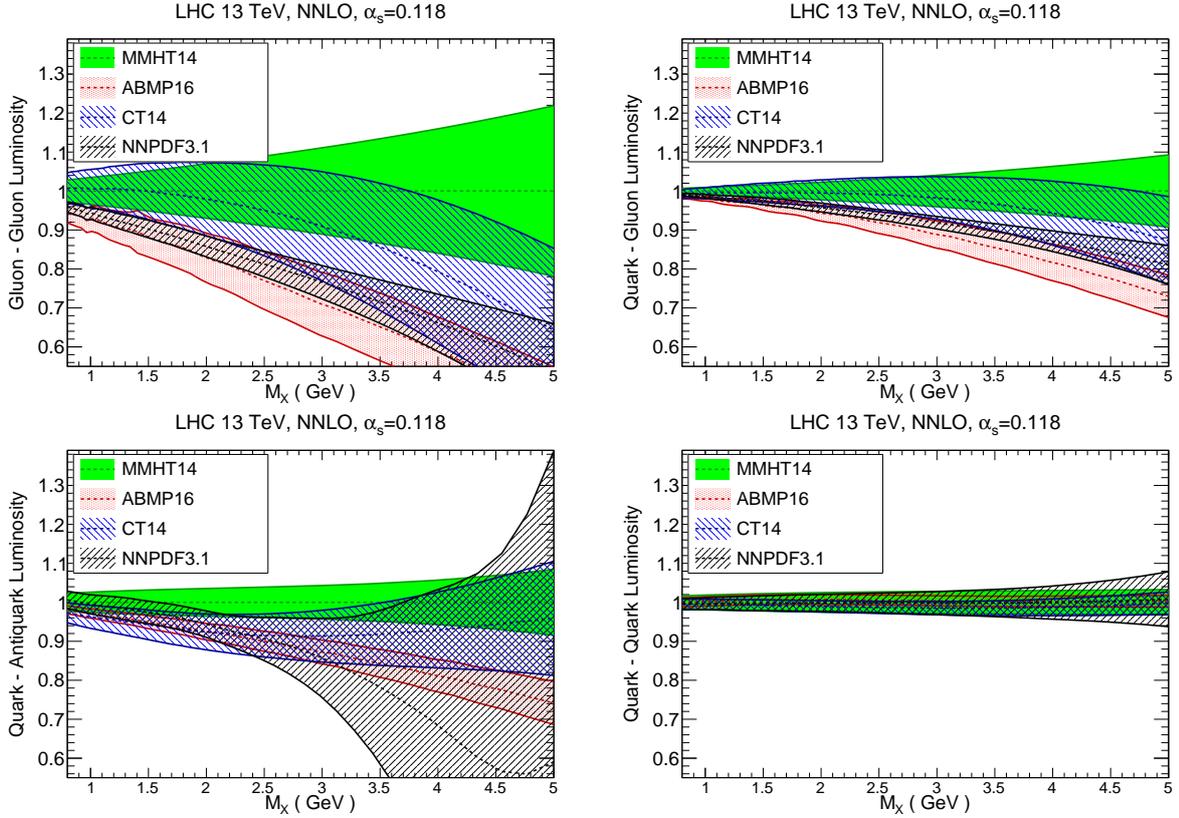


Figure 62: Comparison of PDF luminosities in the large invariant mass M_X region between MMHT14, ABMP16, CT14 and NNPDF3.1. From left to right and from top to bottom we show the results of the gluon-gluon, gluon-quark, quark-antiquark and quark-quark luminosities, normalized to the central value of MMHT14. In this comparison, NNLO PDFs with $\alpha_s(m_Z) = 0.118$ sets are used.

2623 Ref. [11]. This production channel is dominated by the quark-antiquark luminosity. Specifically, we compare the
 2624 predictions of NNPDF3.0, CTEQ6.6, and MSTW2008, all at NLO. Note that by construction the central values of the
 2625 three predictions are close since different trends cancel in this K -factor ratio, so the usefulness of this comparison is
 2626 estimating the PDF uncertainties in each case. Here we also see that PDF uncertainties become very large at high-
 2627 masses, in particular in the case of NNPDF3.0, reflecting the underlying behaviour of the quark-antiquark luminosities.
 2628 Therefore, Fig. 63 highlights that, in the case of an eventual discovery of novel high-mass particles at the LHC, it will
 2629 be crucial to improve our knowledge of large- x PDFs in order to be able to characterize the underlying BSM scenario,
 2630 In this respect, they way forward, as hinted already in Fig. 63, is the exploitation of high-precision collider data,
 2631 mostly from the LHC, in order to pin down the large- x gluons and anti-quarks and thus reduce the PDF uncertainties
 2632 associated to high-mass BSM particle production. For instance, Ref. [276] showed how by including the y_t and
 2633 $y_{t\bar{t}}$ differential distributions from top quark pair production in a global PDF fit, it is possible to reduce the PDF
 2634 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
 2635 for searches, for instance of new resonances that couple strongly to the top quark. More towards the future, it might
 2636 conceivable to be able to provide indirect constraints on BSM, for instance on the coefficients of the SM-EFT [425]
 2637 higher-order operators, by including these in the global PDF fit, along the lines of early studies aiming to constrain
 2638 colored sparticles from Tevatron jet production [426].

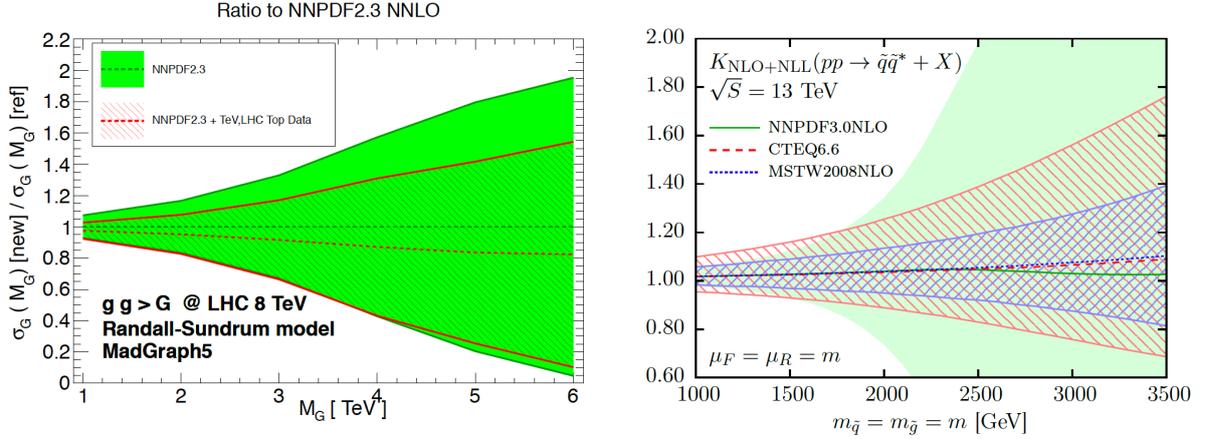


Figure 63: Left: the PDF uncertainties for high-mass graviton production in the Randall-Sundrum scenario induced by gluon-fusion at the LHC 8 TeV, computed with MadGraph5. We compare the results of the NNPDF2.3 fit with that of the same fit including the constraints from top-quark production cross-sections. 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.

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 context of the global electroweak precision fit [427], to predict other parameters, such as M_W . By comparing this indirect predictions of the W mass with direct experimental measurements, 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 [428] 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 [429, 430, 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 [431], 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. 64 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 [431]. 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. 64 we also show a phenomenological estimate of the PDF uncertainties associated to the m_W measurements using different PDF sets and collider scenarios, from [429]. 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 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

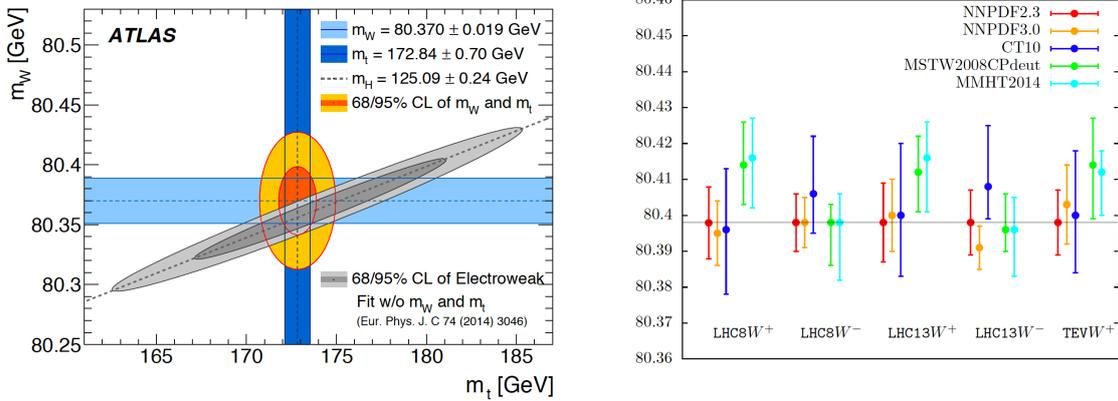


Figure 64: Left: comparison of the direct measurements of m_W , m_t , and m_h from ATLAS with the predictions from the global electroweak fit, from [431]. Right: estimate of the PDF uncertainties in the m_W determination using different PDF sets and collider scenarios, from [429]. This estimate has been obtained from template fits to the p_T^l distribution, imposing the constraint that $p_T^W \leq 15$ GeV.

2670 corresponding SM prediction. This fact is for example at the basis of the improved agreement at high scales between
 2671 the strong, weak and electromagnetic couplings in the case of low-scale supersymmetry, which suggest the unification
 2672 of the three forces at a GUT scale of around $\Lambda_{\text{GUT}} \sim 10^{16}$ GeV [432]. If these new heavy particles are at the TeV
 2673 scale, the difference induced in the QCD coupling running could be accessible at the LHC, see *e.g.* [433] and Fig. 65,
 2674 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
 2675 for various representations of its color gauge group. With this motivation, as well as the one to compare with other
 2676 measurements of α_s at lower energies, the ATLAS and CMS collaborations have presented a number of measurements
 2677 of both $\alpha_s(m_Z)$ and of $\alpha_s(Q)$ for individual Q bins, mostly from jet production [164, 434, 435] but also from top-quark
 2678 pair production [436] (see also [437] for a review, and Fig. 65 for a graphical overview).

2679 In these collider-based determinations of the strong coupling, PDF uncertainties, which are significant at the TeV
 2680 scale (see Sect. 8.2), represent an important source of theoretical uncertainties. For instance, in the recent ATLAS
 2681 determination of $\alpha_s(m_Z)$ from transverse energy-energy correlations (TEEC) at 8 TeV [434], the PDF uncertainty is
 2682 $\delta_{\text{pdf}} = 0.0018$, almost a factor 2 larger than the experimental uncertainty of $\delta_{\text{pdf}} = 0.0011$. While in this analysis PDF
 2683 uncertainties are sub-dominant with respect to the scale uncertainties, $\delta_{\text{scale}} \simeq 0.006$, the latter were computed using
 2684 NLO theory and can be reduced significantly by exploiting the NNLO calculation. Likewise, in the CMS analysis
 2685 at 7 TeV, based on a QCD analysis of the inclusive jet cross-sections [164], one finds that the PDF uncertainties
 2686 $\delta_{\text{pdf}} = 0.0028$ are larger than the experimental uncertainties $\delta_{\text{exp}} = 0.0019$, though still sub-dominant with respect
 2687 to the large scale variations of the NLO calculation $\delta_{\text{scale}} = {}^{+0.0053}_{-0.0024}$. In both cases, it is manifest that if one is able to
 2688 reduce PDF uncertainties, and exploit the reduction of scale errors of the NNLO calculation, one can achieve a very
 2689 competitive determination of $\alpha_s(m_Z)$ and to also perform stringent tests of its running in the TeV region.

2690 As a related topic, we would like to mention that there are also proposals to measure the running of the electroweak
 2691 running couplings at the LHC [438] and use these to impose model-independent constraints on new particles with
 2692 electroweak quantum numbers without any assumptions about their decay properties. Also for this measurements
 2693 PDFs are one of the dominant theoretical uncertainties, for instance in the high-mass tail of W and Z production at the
 2694 LHC due to the quark-antiquark luminosity. Improving our knowledge of large- x anti-quarks will thus be helpful in
 2695 these respect provide indirect constraints of new heavy electroweak sectors.

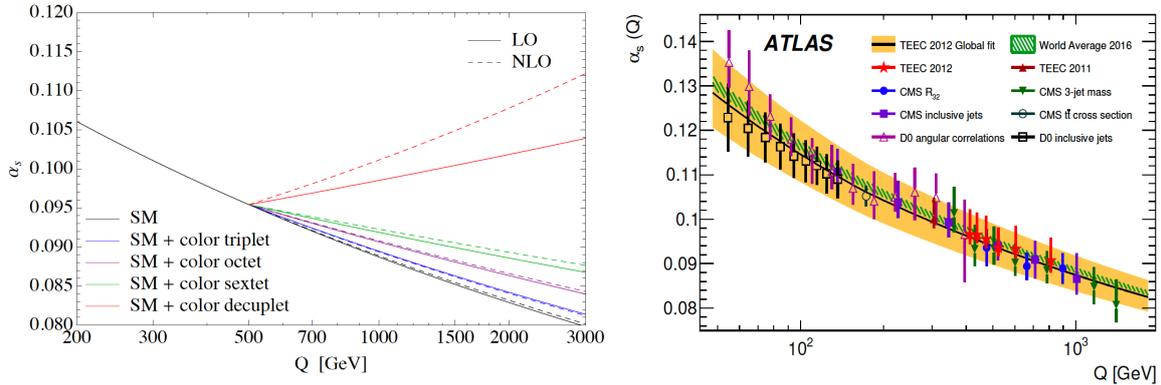


Figure 65: Left: the modification in the running of $\alpha_s(Q)$ induced by a new heavy colored fermion of mass $m = 0.5$ TeV as compared to the SM prediction, for various representations of its color gauge group, from [433]. Right: comparison of recent direct determinations of $\alpha_s(Q)$ at the 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.

2696 9. The future of PDF determinations

2697 In the final section of this Report we discuss three topics that could play an important role in shaping global
 2698 analyses of PDFs in the coming years. First of all we discuss the problem of theoretical uncertainties in fits of parton
 2699 distributions, whose estimate is becoming more and more urgent given the size of PDF uncertainties in recent global
 2700 analysis. Secondly, we summarize recent progress in lattice QCD computations of PDFs, including the first effort
 2701 towards a determination of their x -dependence, and suggest that in the near future lattice inputs could contribute to global
 2702 PDF fits. Thirdly, we briefly review the status and plans for future high-energy colliders and the role that PDFs would
 2703 play in these.

2704 9.1. Theoretical uncertainties

2705 As the development of methodologies of the PDF fits and more precise data available, the PDF uncertainties
 2706 in global determination due to experimental measurements have been reduced to percent-level in most constrained
 2707 region. Meanwhile, various theoretical uncertainties become more and more important and a major limitation on
 2708 future studies. In this section we focus on the theoretical uncertainties due to missing higher order (MHOU) QCD
 2709 corrections, namely by truncation of the asymptotic perturbative expansion, in contrast to those discussed before from
 2710 various input parameters of QCD. There have been extensive studies recently on how to estimate MHOU. In the
 2711 following we first review progresses on MHOU of calculations of non-hadronic and hadronic processes and then several
 2712 recent studies related to PDF determination.

2713 9.1.1. MHOU on matrix element calculation

2714 The mostly used and probably also the simplest way of estimating the missing higher order corrections is by
 2715 variation of the arbitrary QCD scales for example in a fixed-order calculation. In the case of the total inclusive cross
 2716 sections or decay rates with one hard scale Q , usually we vary the QCD renormalization scale μ_R within the interval
 2717 $[Q/r, rQ]$. The induced changes on the studied QCD observables, either from three point evaluations or a scan over
 2718 the entire range, are taken as the *uncertainty* of the MHO assuming a Gaussian or two half-Gaussian distributions
 2719 usually. The conventional choice is $r = 2$ which is found to work well in most cases but underestimate the true higher
 2720 order corrections in certain cases especially if the fixed-order calculation is carried out at leading order. At hadron
 2721 colliders there exist also QCD factorization scale μ_F from collinear factorization due to initial state hadrons. The two
 2722 scales μ_R and μ_F can be varied either simultaneously or independently within above range, with the later case further
 2723 restricted to $1/r \leq \mu_F/\mu_R \leq r$ usually. However, even for a single scale problem, there can still be different choices of
 2724 the central or nominal scale, e.g., $Q/2$ or $2Q$, motivated by either QCD resummation or speed of convergence of the

series [270], which leads to further ambiguities in 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 [?] than traditional scale variations.

Prescription of scale variations becomes complicated when moving to differential observables since more hard scales including those related to the kinematics are involved. That usually requires a dynamic choice for the nominal scale which can also depend on the distribution studied. For example, in a recent study on hadronic production of top-quark pairs [270] 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. From the *proper* 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 is concerning correlations of the MHOU or scale variations in different regions of the distribution. Normally 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. There has also been attempt on decorrelation of the scale variations based on consideration of kinematic dependence of the QCD corrections [?].

There are other interesting proposals on estimating MHOU based on results at known orders, like the Cacciari-Houdeau (CH) approach [439]. The basic idea is to express the full perturbation series in terms of the expansion parameter $\alpha_s(Q)$ and assume 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 is for study of non-hadronic processes. Later in the modified CH ($\overline{\text{CH}}$) [440] approach it is 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. Be specific for the best value of λ the predicted probability density of higher orders is required to match the distribution from frequency count in the survey. For hadronic processed the appropriate value of λ is found to be about 0.6. Besides, the series acceleration method [?] can also be applied to approximate full result on the observables based on informations of only finite terms of the asymptotic series, e.g., using Levin-Weniger sequence transforms. In Ref. [?] 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. Fig. 66 shows predictions on production cross sections of the Higgs boson via gluon fusion at the LHC 8 TeV, calculated at LO, NLO, NNLO and approximate N^3LO with a nominal scale of $\mu_R = \mu_F = m_H$ [441]. It compares the MHOU as estimated from different approaches, including scale variations, *CH*, $\overline{\text{CH}}$ and the series acceleration method in [?], at various orders. Note different approaches may have different interpretations on the uncertainties. In the case of $\overline{\text{CH}}$ the λ value has been adjusted to give almost equal expansion coefficients for the known orders [441]. The $\overline{\text{CH}}$ method predicts larger uncertainty in general. All approaches give similar size of uncertainties at N^3LO except for the *CH* method. The series acceleration method also has a shift on the central prediction by construction.

9.1.2. MHOU on determined PDFs

Global determination of PDFs utilize perturbative predictions on matrix elements on a variety of processes, including DGLAP splitting kernels and various production cross sections, of which the MHOU propagates into the resulting PDFs via fits of the theoretical predictions to the experimental data. Thus sophisticated treatments on the MHOU from different sources are required to study the impact on the PDFs, similar to the experimental systematics. Especially correlations between theoretical predictions of different experimental bins of one process and further of different processes must be accounted for. Further when making any theoretical predictions one should also be care of the correlations between MHOU of the PDFs and the MHOU of the process studied since they may rely on the same theory of the matrix elements. There is no satisfactory solution yet and the MHOU have not been included in any of the public PDFs from global determinations.

However, if focusing on a region where we know only a single process is 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. Left plot of Fig. 67 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 section ??). Note that the theoretical predictions on jet cross sections there are only at NLO though

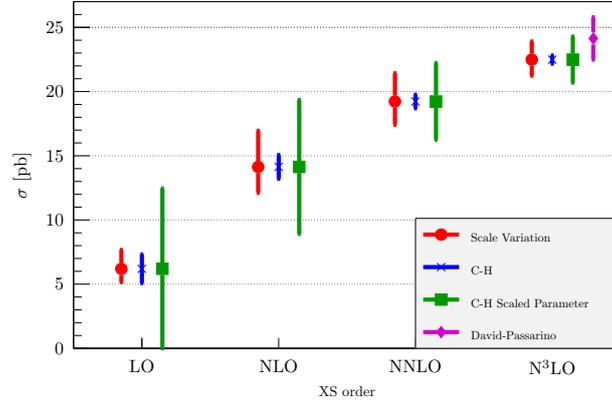


Figure 66: The cross section for Higgs production in gluon fusion calculated at increasing perturbative orders [441]. At each order the theoretical uncertainty is shown for using scale variation (red circles), the CH method (blue crosses), and the $\overline{\text{CH}}$ method (green squares); at N^3LO the Passarino-David uncertainty based on series acceleration method is also shown (purple diamonds).

2776 the PDFs are determined at NNLO. At $x > 0.2$ spread of the gluon PDFs by using scales of 0.5, 1 and 2 times the
 2777 individual jet p_T can be significant indicating possible large MHOU of the gluon PDFs there. In another study also
 2778 based on alternative CT10 NNLO fits, scale variations of the NLO inclusive jet cross sections are further decomposed
 2779 into several correlated systematics describing by five nuisance parameters [3]. By treating those systematics in a
 2780 similar way as the experimental correlated systematic errors the MHOU are included in the usual PDF uncertainty
 2781 automatically as shown in the right plot of Fig. 67 which results in an increase of the gluon PDF uncertainty at large- x .

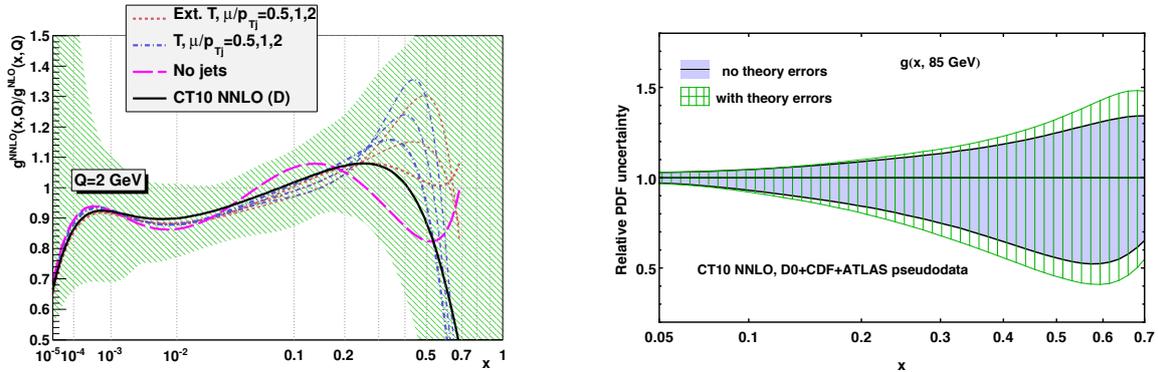


Figure 67: Left plot: dependence of the gluon PDFs on the choice of QCD scales in calculation of inclusive jet cross sections in CT10 NNLO fits [81]. Right plot: impact of the theoretical uncertainties of inclusive jet cross sections on the gluon PDFs in CT10 NNLO fits [3].

2782 Another possibility is by checking convergence of the fitted PDFs with increasing orders similar as in the case of
 2783 a total cross section. Fig. 68 shows comparisons of conventional PDF uncertainties with the difference of the central
 2784 PDFs fitted at NLO and NNLO for gluon and total singlet PDFs at $Q = 100$ GeV as in the NNPDF3.0 fits [17]. The
 2785 latter one can serve as a conservative estimation on upper bound of the MHOU of the NNLO PDFs. There are regions
 2786 where the shifts of NLO to NNLO PDFs are comparable or even larger than the conventional PDF uncertainties. Of
 2787 course one can also apply the CH or $\overline{\text{CH}}$ approach based on the fitted PDFs at LO, NLO and NNLO. For example
 2788 Ref. [441] found the resulting MHOU of the NNLO PDFs have negligible impact on the Higgs production cross
 2789 section through gluon fusion but can be important for the top-quark pair production at the LHC.

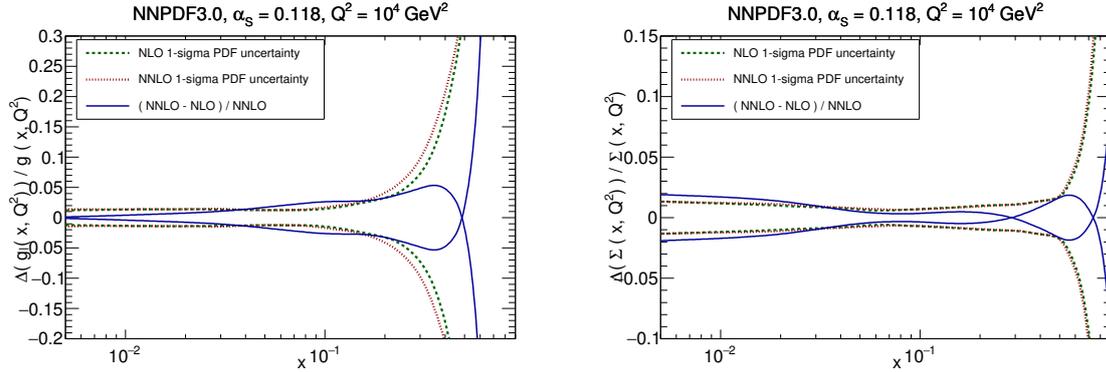


Figure 68: Left(right) plot shows comparisons between the conventional PDF uncertainties and the differences of the central PDFs determined at NLO and NNLO for the gluon(total singlet) PDF as from NNP3.0 [17].

2790 9.2. PDFs and lattice calculations

2791 As discussed in Sect. 2, the PDFs arise from non-perturbative dynamics and thus since currently we are not able
 2792 to solve QCD, they cannot be determined from first principles. However, there is a method that allows to evaluate
 2793 non-perturbative QCD quantities, namely lattice QCD. This method is based on discretizing the QCD Lagrangian in
 2794 a finite-volume finite-volume Euclidean lattice, which introduces naturally an ultraviolet cutoff, and then computing
 2795 directly non-perturbative QCD quantities. Perhaps the main advantage of lattice QCD calculations is that they require
 2796 minimal external input, in particular only the hadronic mass scale Λ_{QCD} and the values of the quark masses, or
 2797 alternatively the physical pion and kaon masses as their proxy.

2798 Given that parton distributions have a formal definition in terms of the nucleon matrix elements of certain oper-
 2799 ators, in principle it is possible to attempt to compute PDFs within lattice QCD. From the practical point of view
 2800 however, given the very CPU intensive nature of these calculations, most lattice QCD results on PDFs were limited to
 2801 the first two moments of non-singlet flavour combinations for large (unphysical) quark masses. These limitation have
 2802 been overcome in the recent years, with multiple groups providing now results of PDF moments for physical pion and
 2803 kaon masses and thus reducing one of the biggest systematic uncertainties that affected these computations, namely
 2804 the extrapolation to physical quark masses. Moreover, it is now possible to go beyond flavor non-singlet operators
 2805 and be able to compute also gluonic and quark singlet matrix elements. Even more recently, both conceptual and
 2806 numerical breakthroughs in lattice QCD computations have allowed to move further, both allowing the calculation
 2807 of higher PDF moments both for flavour non-singlet and singlet operators, as well as the first attempt to compute
 2808 parton distributions directly in Bjorken- x space. The latter was made possible by the development of the so-called
 2809 Large-Momentum Effective Theory (LaMET) [?], which introduces quasi-PDFs within nucleons corresponding to a
 2810 finite momentum boost in the co-moving reference frame [? ? ? ?]. Although we are still in the early days of direct
 2811 x -space PDF computations in lattice QCD, it is conceivable that recent progress might be able to provide competitive
 2812 constrains that can be incorporated in the global PDF analysis.

2813 The recent developments in lattice QCD calculations of PDFs are summarized in Fig. 69. In the left plot, we show
 2814 the $u(x) - d(x)$ difference as a function of Bjorken x , where the results of a recent lattice QCD calculation based on
 2815 quasi-PDFs is compared with the MSTW08 and CJ12 results. The in the right plot we show the momentum fraction
 2816 carried by sum of up and down quarks, $\langle x \rangle_{u+d}$ for different values of the parameter t_s (points with errors), compared
 2817 with the results from recent global PDF fits (horizontal band). These two representative results are telling of the
 2818 important progress in lattice QCD calculations of the proton structure, first of all being able to compute moments
 2819 of singlet combinations and of the gluon and second of starting explore the full x -dependence of the PDFs. These
 2820 developments suggest that in the future lattice QCD calculations could provide useful input for global PDF fits.

2821 9.3. Parton distributions at future colliders

2822 There exist are a number of different proposals for future high-energy colliders. Electron-positron colliders, such
 2823 as the ILC [442], CLIC [443], or TLEP/FCC-ee [444], offer the potential for ultra-high precision measurements of

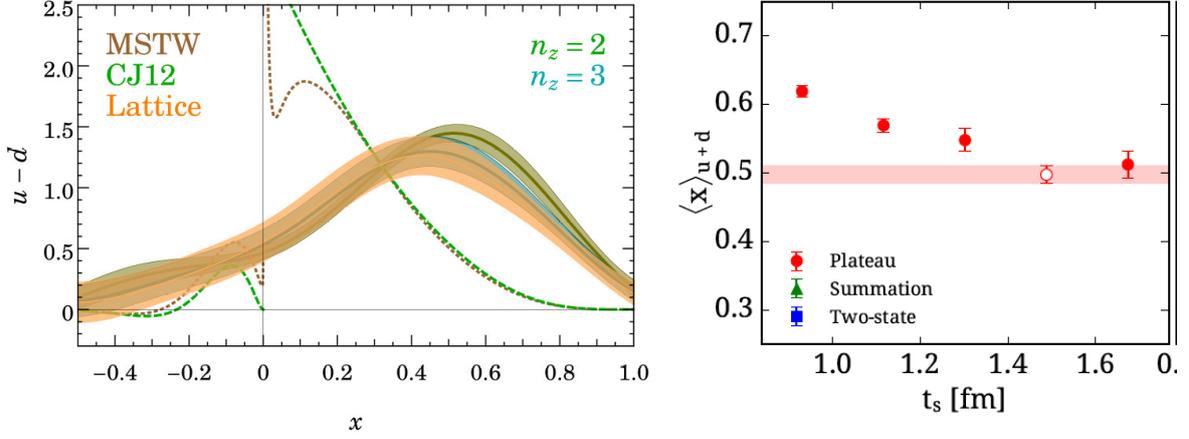


Figure 69: Left plot: the $u(x) - d(x)$ difference as a function of Bjorken x , where the results of a recent lattice QCD calculation based on quasi-PDFs is compared with the MSTW08 and CJ12 results. Right plot: the momentum fraction carried by sum of up and down quarks, $\langle x \rangle_{u+d}$ for different values of the parameter t_s (points with errors), compared with the results from recent global PDF fits (horizontal band).

2824 the Higgs, electroweak and top-sections. Hadron colliders with energy much greater than the LHC would allow to
 2825 continue the exploration of the high-energy frontier and the searches for dark matter particles, while allowing at the
 2826 same time unprecedented opportunities for the study of the Higgs sector such as the Higgs self-interactions. For
 2827 instance, there is ongoing work towards a circular collider hosted at the CERN site which would accelerate protons up
 2828 to the extreme energies of $\sqrt{s} = 100$ TeV [445, 15], dubbed FCC-hh, and there is also a similar machine under study
 2829 by the Chinese HEP community. Another avenue for future high energy collisions are based around electron-proton
 2830 colliders, essentially scaled-up versions of HERA, such as the LHeC [16] or FCC-ep. In the same category falls
 2831 the Electron Ion Collider (EIC) [446] which might start construction very soon either in the BNL or the JLAB sites.
 2832 Though its center of mass energy would not reach that of HERA, the possibility to polarize both leptons and protons
 2833 and to accelerate as well heavy nuclei offer unique possibilities to explore the internal structure of nucleon from many
 2834 different points of view. In this section we briefly review the role that PDFs would play first at the LHeC and then at
 2835 a future hadron collider at 100 TeV.

2836 9.3.1. PDFs at the Large Hadron electron Collider

2837 As discussed above, one of the possibilities for a future high energy collider would be to exploit the LHC protons
 2838 and collide them with a high energy lepton beam. This project, known as the Large Hadron electron Collider, would
 2839 then represent a scaled-up version of HERA, and as such would offer immense opportunities for improved determi-
 2840 nations of the proton structure down to very low x and high- Q^2 , as well as providing a wealth of information on
 2841 nuclear PDFs in a kinematic region where they are currently essentially unconstrained. Several options are now being
 2842 considered, with some preference now for synchronous operation during the final years of the HL-LHC upgrade.

2843 In Fig. 70 we show the kinematical coverage in the (x, Q^2) of several existing and proposed deep-inelastic scatter-
 2844 ing experiments. We observe how by starting from the fixed-target experiments and then moving to HERA, the LHeC
 2845 and finally the FCC-eh, as the center of mass energy increases, the kinematic reach extends both towards higher Q^2
 2846 and smaller x values. At the FCC-eh in particular, it should be able to cover the region down to $x \approx 10^{-7}$ without
 2847 leaving the perturbative region $Q \gtrsim 1$ GeV. In the same figure we also find the results of an xFitter PDF feasibility
 2848 study that compares the impact on the gluon PDF of adding either LHeC or FCC-eh (or both) pseudo-data in addition
 2849 to the HERA inclusive structure function dataset. The reduction of the PDF uncertainties down to very small- x values
 2850 reflects the extended kinematic reach of these possible future high energy lepton proton colliders.

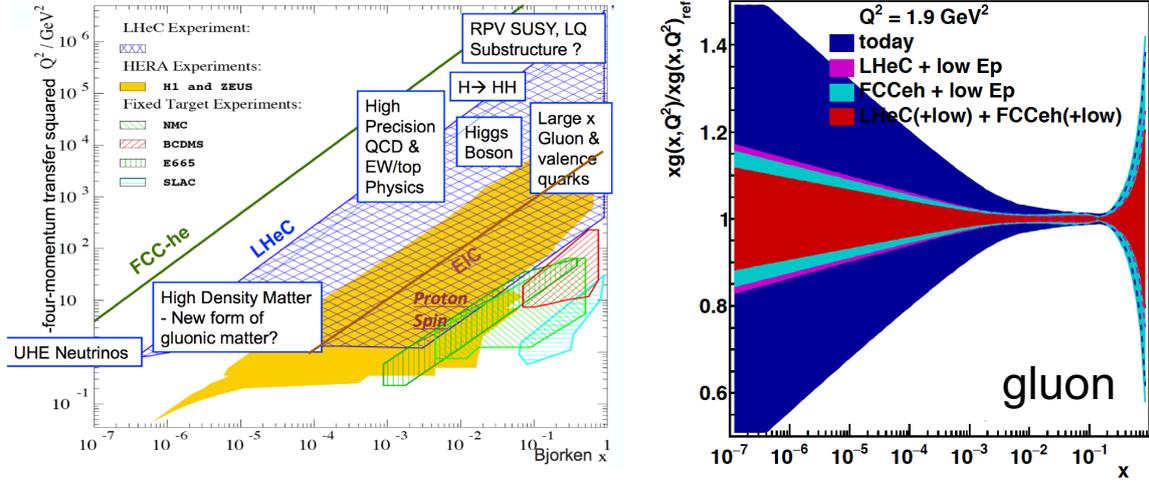


Figure 70: Left plot: kinematical coverage in the (x, Q^2) of several existing and proposed deep-inelastic scattering experiments. Starting from the fixed-target experiments and then moving to HERA, the LHeC and finally the FCC-eh, as the center of mass energy increases, the kinematic reach extends both towards higher Q^2 and smaller x values. Right plot: results of an `xFitter` PDF feasibility study that compares the impact of adding either LHeC or FCC-eh (or both) pseudo-data in addition to the HERA inclusive structure function dataset.

9.3.2. PDFs at a 100 TeV hadron collider

In order to illustrate the extended kinematical coverage that would be achieved at a 100 TeV proton-proton collider as compared to the one at the LHC, in Fig. 71 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 very significant increase in the kinematic coverage, and in particular that at the FCC even high-scale processes such as W, Z or h production become sensitive to the small- x region.

In Fig. 71 we represent the kinematical coverage in the (x, M_X) plane, where M_X is the invariant mass of the produced final state, for a $\sqrt{s} = 100$ TeV hadron collider, compared with the corresponding coverage of the LHC at $\sqrt{s} = 14$ TeV. We also indicate the coverage in M_X for phenomenologically important processes at the FCC, from low masses (such as Drell-Yan or low p_T jets), electroweak scale processes (such as Higgs, W, Z , or top production), and hypothetical new high-mass particles (such as a 2 TeV squark or a 20 TeV Z').

In Fig. 72 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, 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. 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.

Kinematics of a 100 TeV FCC

Plot by J. Rojo, Dec 2013

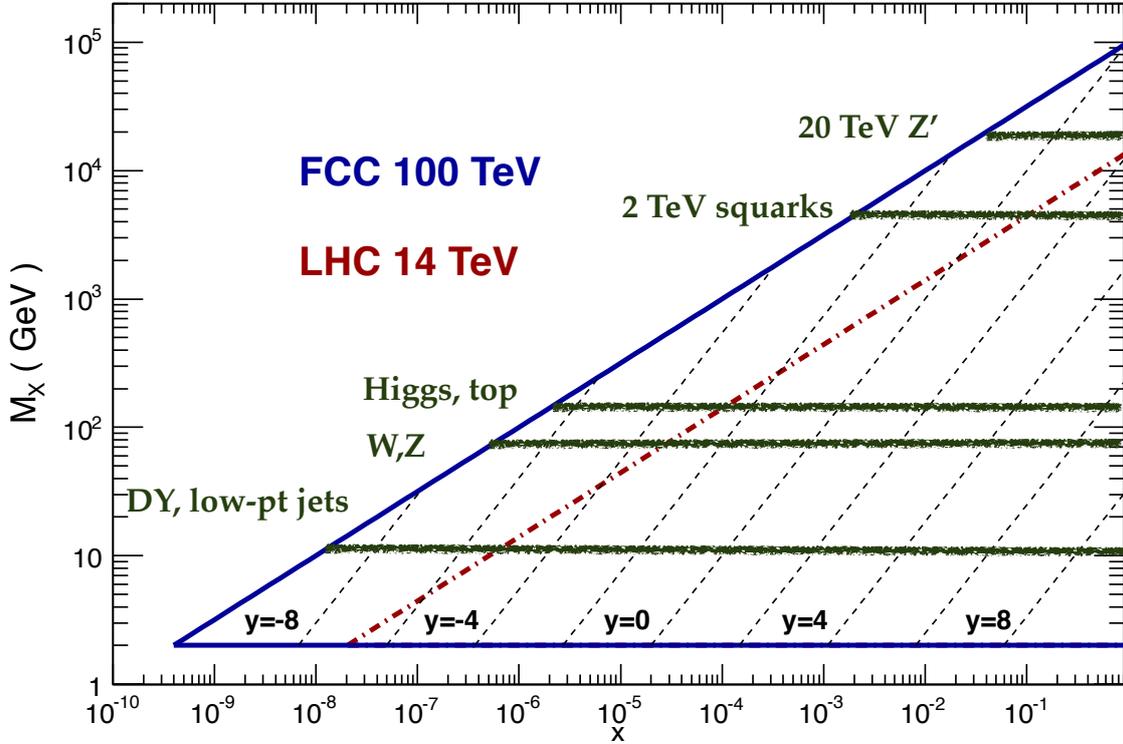


Figure 71: Kinematical coverage in the (x, M_X) plane of a $\sqrt{s} = 100$ TeV hadron collider (solid blue line), compared with the corresponding coverage of the LHC at $\sqrt{s} = 14$ TeV (dot-dashed red line).

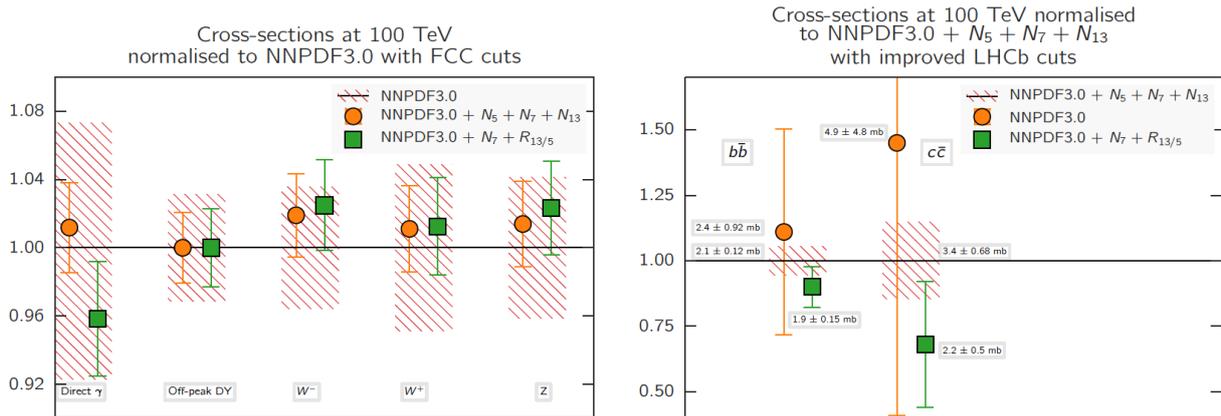


Figure 72: Comparison of cross-sections for different representative processes at the FCC with $\sqrt{s} = 100$ TeV, between the NNPDF3.0 predictions and those of the NNPDF3.0+LHCb sets, as discussed in the text. The acceptance cuts are different in each process. In the left plot we show the results 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.

2870 **10. Conclusions**

2871 The wealth of data that has been accumulated by the LHC so far, together with the additional data that will
2872 be collected in the coming two decades, has allowed the study and stress-test of the SM in an unprecedented way.
2873 Together with recent progress in theoretical calculations, this means that we are now entering the precision era of the
2874 LHC, aiming to compare data and theory at the few percent level or even less. Given the null results of BSM searches
2875 so far, a systematic high-precision analysis of the SM predictions and the LHC data might be one of most promising
2876 approaches to look for BSM dynamics at the LHC, for instance in the case that they manifest as subtle differences
2877 with the SM theory. And in this respect, the detailed mapping of the quark and gluon structure of the proton represents
2878 and important component of this LHC precision physics program.

2879 In this Report we have presented an overview of the most important recent developments in PDF determinations,
2880 with emphasis for their implications for LHC phenomenology. After a succinct review of the theoretical foundations
2881 of the global QCD analysis framework, we have reviewed recent progress both from the theoretical and the exper-
2882 imental point of view for those hard-scattering cross-sections used in PDF fits; we have compared the similarities
2883 and differences between the methodologies uses for the various PDF fitting collaborations; and then presented the
2884 state-of-the-art fits from each group and assessed what we can learn about the internal structure of the nucleons from
2885 various points of view. We have then discussed the role of QED corrections in PDF fits, in particular concerning
2886 photon-initiated processes, and presented some of the most representative examples of the applications of PDFs for
2887 LHC phenomenology, from the measurement of the Higgs couplings to the determination of the mass of the W boson.

2888 In the last section of these Report we have attempted to speculatively discuss some topics that very likely will
2889 play a crucial role in the near-term future of PDF determinations. One of these topics is that one of theoretical
2890 uncertainties, for instance arising from missing higher-order terms in the perturbative expansion. Given the size of
2891 PDF uncertainties in the latest sets, it is conceivable that these theory error are comparable (if not larger) than the
2892 nominal PDF uncertainties, and thus finding a statistically sound method to account for these is of utmost importance.
2893 Another topic that might affect the PDF fitting paradigm is that of the interplay with lattice QCD calculations, where
2894 recent progress both in computing Mellin moments of various flavor combinations as well in direct x -space calculation
2895 of PDFs suggest that in the future the constraints from lattice calculations should allow to improve the global PDF
2896 fits, in a similar way as *e.g.* the momentum sum rule. Finally we have summarized the importance of PDFs for future
2897 higher-energy colliders whose physics case is being discussed just now, such as a Large Hadron electron Collider or a
2898 new proton-proton collider with a center-of-mass energy of up to 100 TeV.

2899 We hope that this Report has managed to convey to the reader that the topic of PDF determinations is fascinating
2900 and lively one, with implications from the understanding of the non-perturbative dynamics of the strong interactions
2901 to searches for new BSM physics and ultra-high energy astrophysics. In some respect, PDF fits represent a unique
2902 stress-test of the SM and of the collinear QCD factorization framework, which is clearly in very good shape, given
2903 that we are now able to simultaneously describe a few tens of individual experiments, some of them with extremely
2904 small uncertainties at the per-mile level. PDF fits thrive at the the cross-roads of advanced data analysis, state-of-the-
2905 art perturbative calculations, and modern robust statistical methodology, and thus provide guidance for other similar
2906 global analyses efforts such as fits of the SMEFT coefficients. Moreover, progress in unpolarized PDF fits is also one
2907 of the main drivers of recent improvements of other related aspects of the proton structure, from polarized PDFs to
2908 nuclear PDFs.

2909 As we enter in the LHC precision era, ever-improving PDF determinations will keep providing a unique contribu-
2910 tion to this exciting exploration of the high-energy frontier. Taken into account that only around 15 years the first PDF
2911 sets with uncertainties were introduced, is clear that the requirements of this LHC precision program will further
2912 drive improvements in global PDF determinations, leading always to an ever more detailed picture of the inner life of
2913 the protons.

2914 **Acknowledgments**

2915 We are grateful to Giulia Zanderighi for giving us the opportunity to write this Report. We are grateful to many
2916 colleagues for extensive discussions and fruitful collaborations of the topic of PDF determinations. A necessarily in-
2917 complete list of people that we could like to thank includes Sergei Alekhin, Mandy Cooper-Sarkar, Stefano Forte, Jan

2918 Krenschmar, Sasha Glazov, Joey Huston, Katerina Lipka Michelangelo Mangano, Pavel Nadolsky, Voica Radescu,
2919 Robert Thorne, Gavin Salam, add all names

2920

2921 J. R. is supported by the European Research Council Starting Grant “PDF4BSM”. add all funding ID.

- 2922 [1] A. Accardi, et al., A Critical Appraisal and Evaluation of Modern PDFs, *Eur. Phys. J. C* 76 (8) (2016) 471. arXiv:1603.08906,
2923 doi:10.1140/epjc/s10052-016-4285-4.
- 2924 [2] J. Butterworth, et al., PDF4LHC recommendations for LHC Run II, *J. Phys. G* 43 (2016) 023001. arXiv:1510.03865, doi:10.1088/0954-
2925 3899/43/2/023001.
- 2926 [3] J. Rojo, et al., The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II, *J. Phys. G* 42 (2015) 103103.
2927 arXiv:1507.00556, doi:10.1088/0954-3899/42/10/103103.
- 2928 [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.
2929 arXiv:1211.5142, doi:10.1007/JHEP04(2013)125.
- 2930 [5] S. Alekhin, et al., The PDF4LHC Working Group Interim Report arXiv:1101.0536.
- 2931 [6] S. Forte, G. Watt, Progress in the Determination of the Partonic Structure of the Proton, *Ann.Rev.Nucl.Part.Sci.* 63 (2013) 291.
2932 arXiv:1301.6754, doi:10.1146/annurev-nucl-102212-170607.
- 2933 [7] S. Forte, Parton distributions at the dawn of the LHC, *Acta Phys.Polon.* B41 (2010) 2859. arXiv:1011.5247.
- 2934 [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-
2935 4885/76/4/046201.
- 2936 [9] A. De Roeck, R. S. Thorne, Structure Functions, *Prog.Part.Nucl.Phys.* 66 (2011) 727. arXiv:1103.0555, doi:10.1016/j.pnnp.2011.06.001.
- 2937 [10] D. de Florian, et al., Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector arXiv:1610.07922.
- 2938 [11] W. Beenakker, C. Borschensky, M. Krmer, A. Kulesza, E. Laenen, S. Marzani, J. Rojo, NLO+NLL squark and gluino production cross-
2939 sections with threshold-improved parton distributions, *Eur. Phys. J. C* 76 (2) (2016) 53. arXiv:1510.00375, doi:10.1140/epjc/s10052-016-
2940 3892-4.
- 2941 [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,
2942 *Phys.Rev. D* 83 (2011) 113008. arXiv:1104.2056, doi:10.1103/PhysRevD.83.113008.
- 2943 [13] A. Cooper-Sarkar, P. Mertsch, S. Sarkar, The high energy neutrino cross-section in the Standard Model and its uncertainty, *JHEP* 08 (2011)
2944 042. arXiv:1106.3723, doi:10.1007/JHEP08(2011)042.
- 2945 [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.
2946 arXiv:1511.06346, doi:10.1007/JHEP02(2016)130.
- 2947 [15] M. L. Mangano, et al., Physics at a 100 TeV pp collider: Standard Model processes arXiv:1607.01831.
- 2948 [16] J. Abelleira Fernandez, et al., A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and
2949 Detector, *J.Phys. G* 39 (2012) 075001. arXiv:1206.2913, doi:10.1088/0954-3899/39/7/075001.
- 2950 [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.
- 2951 [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 dis-
2952 tribution functions from a global analysis of quantum chromodynamics, *Phys. Rev. D* 93 (3) (2016) 033006. arXiv:1506.07443,
2953 doi:10.1103/PhysRevD.93.033006.
- 2954 [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. C*
2955 75 (5) (2015) 204. arXiv:1412.3989, doi:10.1140/epjc/s10052-015-3397-6.
- 2956 [20] S. Alekhin, J. Blumlein, S. Moch, R. Placakyte, Parton Distribution Functions, α_s and Heavy-Quark Masses for LHC Run
2957 II arXiv:1701.05838.
- 2958 [21] H. Abramowicz, et al., Combination of measurements of inclusive deep inelastic e^+p scattering cross sections and QCD analysis of HERA
2959 data, *Eur. Phys. J. C* 75 (12) (2015) 580. arXiv:1506.06042, doi:10.1140/epjc/s10052-015-3710-4.
- 2960 [22] P. Jimenez-Delgado, E. Reya, Delineating parton distributions and the strong coupling, *Phys.Rev. D* 89 (7) (2014) 074049. arXiv:1403.1852,
2961 doi:10.1103/PhysRevD.89.074049.
- 2962 [23] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, N. Sato, Constraints on large- x parton distributions from new weak boson production
2963 and deep-inelastic scattering data, *Phys. Rev. D* 93 (11) (2016) 114017. arXiv:1602.03154, doi:10.1103/PhysRevD.93.114017.
- 2964 [24] A. Manohar, P. Nason, G. P. Salam, G. Zanderighi, How bright is the proton? A precise determination of the photon parton distribution
2965 function, *Phys. Rev. Lett.* 117 (24) (2016) 242002. arXiv:1607.04266, doi:10.1103/PhysRevLett.117.242002.
- 2966 [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
2967 Content of the Proton, *Eur. Phys. J. C* 76 (11) (2016) 647. arXiv:1605.06515, doi:10.1140/epjc/s10052-016-4469-y.
- 2968 [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
2969 Distribution Functions from CTEQ-TEA Global Analysis arXiv:1707.00657.
- 2970 [27] S. Alekhin, et al., HERAFitter, *Eur. Phys. J. C* 75 (7) (2015) 304. arXiv:1410.4412, doi:10.1140/epjc/s10052-015-3480-z.
- 2971 [28] E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi, J. Rojo, A first unbiased global determination of polarized PDFs and their uncertainties,
2972 *Nucl.Phys. B* 887 (2014) 276. arXiv:1406.5539, doi:10.1016/j.nuclphysb.2014.08.008.
- 2973 [29] D. de Florian, R. Sassot, M. Stratmann, W. Vogelsang, Extraction of Spin-Dependent Parton Densities and Their Uncertainties, *Phys. Rev. D*
2974 80 (2009) 034030. arXiv:0904.3821, doi:10.1103/PhysRevD.80.034030.
- 2975 [30] K. J. Eskola, P. Paakkinen, H. Paukkunen, C. A. Salgado, EPPS16: Nuclear parton distributions with LHC data, *Eur. Phys. J. C* 77 (3) (2017)
2976 163. arXiv:1612.05741, doi:10.1140/epjc/s10052-017-4725-9.
- 2977 [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. C* 77 (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. B* 46 (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. Ganssaugue, Inelastic electron-proton scattering at fixed four-momentum transfer of 0.773 and 1.935 (gev/c)-squared, *Nucl. Phys. B* 13 (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. D* 5 (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. D* 4 (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. D* 8 (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. B* 73 (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. B* 128 (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. B* 126 (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. C* 13 (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. B* 155 (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. C* 17 (1983) 283. doi:10.1007/BF01571895.
- [60] D. W. Duke, J. F. Owens, q^2 Dependent Parametrizations of Parton Distribution Functions, *Phys. Rev. D* 30 (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. D* 37 (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. C* 39 (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. D* 39 (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. D* 42 (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. C* 52 (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. D*50 (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. D*51 (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. C*10 (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. C*14 (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. C*12 (2000) 243–262. arXiv:hep-ph/9907512, doi:10.1007/s100529900198.
- [74] W. T. Giele, S. A. Keller, D. A. Kosower, Parton distribution function uncertainties arXiv:hep-ph/0104052.
- [75] J. Pumplin, et al., Uncertainties of predictions from parton distribution functions. 2. The Hessian method, *Phys. Rev. D*65 (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: Experimental errors. ((T)), *Eur. Phys. J. C*28 (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. B*688 (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. B*691 (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. C*63 (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. D*89 (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. D*81 (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. D*63 (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^+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. D*79 (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. B*838 (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. B*855 (2012) 153. arXiv:1107.2652.
- [90] R. D. Ball, et al., Parton distributions with LHC data, *Nucl.Phys. B*867 (2013) 244. arXiv:1207.1303, doi:10.1016/j.nuclphysb.2012.10.003.
- [91] M. Botje, et al., The PDF4LHC Working Group Interim Recommendations arXiv:1101.0538.
- [92] 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.
- [93] 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. B*359 (1991) 343–405, [Erratum: *Nucl. Phys. B*644,403(2002)]. doi:10.1016/S0550-3213(02)00814-3, 10.1016/0550-3213(91)90064-5.
- [94] 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.
- [95] S. Forte, E. Laenen, P. Nason, J. Rojo, Heavy quarks in deep-inelastic scattering, *Nucl. Phys. B*834 (2010) 116–162. arXiv:1001.2312, doi:10.1016/j.nuclphysb.2010.03.014.
- [96] 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. C*74 (7) (2014) 2958. arXiv:1402.3536, doi:10.1140/epjc/s10052-014-2958-4.
- [97] R. D. Ball, et al., Parton distributions from high-precision collider data arXiv:1706.00428.
- [98] C. D. White, R. S. Thorne, A Global Fit to Scattering Data with NLL BFKL Resummations, *Phys. Rev. D*75 (2007) 034005. arXiv:hep-ph/0611204, doi:10.1103/PhysRevD.75.034005.
- [99] M. Bonvini, S. Marzani, J. Rojo, L. Rottoli, M. Ubiali, R. D. Ball, V. Bertone, S. Carrazza, N. P. Hartland, Parton distributions with threshold resummation, *JHEP* 09 (2015) 191. arXiv:1507.01006, doi:10.1007/JHEP09(2015)191.
- [100] 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. B*223 (1989) 485.

- 3108 [101] 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
3109 at high q^2 , Phys. Lett. B237 (1990) 592.
- 3110 [102] 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,
3111 doi:10.1016/S0550-3213(96)00673-6.
- 3112 [103] 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
3113 (1997) 3–43. arXiv:hep-ph/9610231, doi:10.1016/S0550-3213(96)00538-X.
- 3114 [104] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock, A. Bodek, Precise measurements of the proton and deuteron structure functions from
3115 a global analysis of the SLAC deep inelastic electron scattering cross-sections, Phys. Lett. B282 (1992) 475–482. doi:10.1016/0370-
3116 2693(92)90672-Q.
- 3117 [105] S. Tkachenko, et al., Measurement of the structure function of the nearly free neutron using spectator tagging in inelastic $^2\text{H}(e,$
3118 $e'p)X$ scattering with CLAS, Phys. Rev. C89 (2014) 045206, [Addendum: Phys. Rev.C90,059901(2014)]. arXiv:1402.2477,
3119 doi:10.1103/PhysRevC.90.059901, 10.1103/PhysRevC.89.045206.
- 3120 [106] 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-
3121 3213(83)90174-8.
- 3122 [107] W. G. Seligman, et al., Improved determination of alpha(s) from neutrino nucleon scattering, Phys. Rev. Lett. 79 (1997) 1213–1216.
3123 arXiv:hep-ex/9701017, doi:10.1103/PhysRevLett.79.1213.
- 3124 [108] U.-K. Yang, et al., Measurements of F_2 and $xF_3^v - xF_3^{\bar{v}}$ from CCFR ν_μ -Fe and $\bar{\nu}_\mu$ -Fe data in a physics model independent way, Phys. Rev.
3125 Lett. 86 (2001) 2742–2745. arXiv:hep-ex/0009041, doi:10.1103/PhysRevLett.86.2742.
- 3126 [109] G. Onengut, et al., Measurement of nucleon structure functions in neutrino scattering, Phys. Lett. B632 (2006) 65–75.
3127 doi:10.1016/j.physletb.2005.10.062.
- 3128 [110] G. P. Zeller, et al., A Precise determination of electroweak parameters in neutrino nucleon scattering, Phys. Rev. Lett. 88 (2002) 091802,
3129 [Erratum: Phys. Rev. Lett.90,239902(2003)]. arXiv:hep-ex/0110059, doi:10.1103/PhysRevLett.88.091802.
- 3130 [111] M. Goncharov, et al., Precise measurement of dimuon production cross-sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe deep inelastic scattering at the Tevatron,
3131 Phys. Rev. D64 (2001) 112006. arXiv:hep-ex/0102049, doi:10.1103/PhysRevD.64.112006.
- 3132 [112] D. A. Mason, Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon dataFERMILAB-THESIS-2006-
3133 01. doi:10.1103/PhysRevLett.99.192001.
- 3134 [113] A. Kayis-Topaksu, et al., Leading order analysis of neutrino induced dimuon events in the CHORUS experiment, Nucl. Phys. B798 (2008)
3135 1–16. arXiv:0804.1869, doi:10.1016/j.nuclphysb.2008.02.013.
- 3136 [114] F. Aaron, et al., Inclusive Deep Inelastic Scattering at High Q^2 with Longitudinally Polarised Lepton Beams at HERA, JHEP 1209 (2012)
3137 061. arXiv:1206.7007, doi:10.1007/JHEP09(2012)061.
- 3138 [115] 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 ,
3139 Eur.Phys.J. C71 (2011) 1579. arXiv:1012.4355, doi:10.1140/epjc/s10052-011-1579-4.
- 3140 [116] A. Cooper Sarker, Measurement of high- Q^2 neutral current deep inelastic e+p scattering cross sections with a longitudinally polarised
3141 positron beam at HERAAarXiv:1208.6138.
- 3142 [117] H. Abramowicz, et al., Measurement of high- Q^2 charged current deep inelastic scattering cross sections with a longitudinally polarised
3143 positron beam at HERA, Eur.Phys.J. C70 (2010) 945–963. arXiv:1008.3493, doi:10.1140/epjc/s10052-010-1498-9.
- 3144 [118] J. Rojo, Progress in the NNPDF global analysis and the impact of the legacy HERA combination, in: Proceedings, 2015 European Physical
3145 Society Conference on High Energy Physics (EPS-HEP 2015), 2015. arXiv:1508.07731.
3146 URL <http://inspirehep.net/record/1391143/files/arXiv:1508.07731.pdf>
- 3147 [119] 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
3148 global fit, Eur. Phys. J. C76 (4) (2016) 186. arXiv:1601.03413, doi:10.1140/epjc/s10052-016-4020-1.
- 3149 [120] 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
3150 functions and HERA Run I and II combined data, Phys. Rev. D95 (3) (2017) 034003. arXiv:1609.07968, doi:10.1103/PhysRevD.95.034003.
- 3151 [121] H. Abramowicz, et al., Combination and QCD Analysis of Charm Production Cross Section Measurements in Deep-Inelastic ep Scattering
3152 at HERA, Eur.Phys.J. C73 (2013) 2311. arXiv:1211.1182, doi:10.1140/epjc/s10052-013-2311-3.
- 3153 [122] F. D. Aaron, et al., Measurement of the Charm and Beauty Structure Functions using the H1 Vertex Detector at HERA, Eur. Phys. J. C65
3154 (2010) 89–109. arXiv:0907.2643, doi:10.1140/epjc/s10052-009-1190-0.
- 3155 [123] H. Abramowicz, et al., Measurement of beauty and charm production in deep inelastic scattering at HERA and measurement of the beauty-
3156 quark mass, JHEP 09 (2014) 127. arXiv:1405.6915, doi:10.1007/JHEP09(2014)127.
- 3157 [124] 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-
3158 Leading Order in QCD, Phys. Rev. Lett. 116 (21) (2016) 212002. arXiv:1601.05430, doi:10.1103/PhysRevLett.116.212002.
- 3159 [125] 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-
3160 6708/2008/04/063.
- 3161 [126] S. Catani, Y. L. Dokshitzer, M. H. Seymour, B. R. Webber, Longitudinally invariant K_t clustering algorithms for hadron hadron collisions,
3162 Nucl. Phys. B406 (1993) 187–224. doi:10.1016/0550-3213(93)90166-M.
- 3163 [127] S. D. Ellis, D. E. Soper, Successive combination jet algorithm for hadron collisions, Phys. Rev. D48 (1993) 3160–3166. arXiv:hep-
3164 ph/9305266, doi:10.1103/PhysRevD.48.3160.
- 3165 [128] Y. L. Dokshitzer, G. D. Leder, S. Moretti, B. R. Webber, Better jet clustering algorithms, JHEP 08 (1997) 001. arXiv:hep-ph/9707323,
3166 doi:10.1088/1126-6708/1997/08/001.
- 3167 [129] 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,
3168 November 4-6, 1999, 2000, pp. 47–77. arXiv:hep-ex/0005012.
3169 URL http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-00-092
- 3170 [130] V. Khachatryan, et al., Measurement and QCD analysis of double-differential inclusive jet cross-sections in pp collisions at $\sqrt{s} = 8$ TeV
3171 and ratios to 2.76 and 7 TeV, JHEP 03 (2017) 156. arXiv:1609.05331, doi:10.1007/JHEP03(2017)156.
- 3172 [131] 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.
- [132] 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>
- [133] 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.
- [134] A. Abulencia, et al., Measurement of the Inclusive Jet Cross Section using the k_T algorithm in $p\bar{p}$ 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.
- [135] 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.
- [136] 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.
- [137] 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.
- [138] 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.
- [139] 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.
- [140] G. Aad, et al., Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using 4.5 fb¹ 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.
- [141] 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.
- [142] 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 $\sqrt{s} = 7$ TeV, Phys. Rev. D90 (7) (2014) 072006. arXiv:1406.0324, doi:10.1103/PhysRevD.90.072006.
- [143] 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.
- [144] 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.
- [145] 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.
- [146] C. Collaboration, Measurement of Triple-Differential Dijet Cross Sections at $\sqrt{s} = 8$ TeV with the CMS Detector and Constraints on Parton Distribution Functions.
- [147] S. D. Ellis, Z. Kunszt, D. E. Soper, Two jet production in hadron collisions at order α_s^3 in QCD, Phys.Rev.Lett. 69 (1992) 1496–1499. doi:10.1103/PhysRevLett.69.1496.
- [148] Z. Kunszt, D. E. Soper, Calculation of jet cross-sections in hadron collisions at order α_s^3 , Phys. Rev. D46 (1992) 192–221. doi:10.1103/PhysRevD.46.192.
- [149] 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.
- [150] 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.
- [151] 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.
- [152] J. Currie, E. W. N. Glover, J. Pires, NNLO QCD predictions for single jet inclusive production at the LHC arXiv:1611.01460.
- [153] 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.
- [154] 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.
- [155] 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 jet arXiv:1507.02850.
- [156] 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>
- [157] 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.
- [158] 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.
- [159] 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.
- [160] 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.
- [161] 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.

- 3238 [162] 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,
3239 doi:10.1007/JHEP10(2014)145.
- 3240 [163] 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,
3241 doi:10.1140/epjc/s10052-014-2934-z.
- 3242 [164] V. Khachatryan, et al., Constraints on parton distribution functions and extraction of the strong coupling constant from the inclusive jet cross
3243 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.
- 3244 [165] S. D. Ellis, W. J. Stirling, Constraints on isospin breaking in the light quark sea from the Drell-Yan process, Phys. Lett. B256 (1991)
3245 258–264. doi:10.1016/0370-2693(91)90684-I.
- 3246 [166] A. D. Martin, A. J. Mathijssen, W. J. Stirling, R. S. Thorne, B. J. A. Watt, G. Watt, Extended Parameterisations for MSTW PDFs and their
3247 effect on Lepton Charge Asymmetry from W Decays arXiv:1211.1215.
- 3248 [167] 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-
3249 ex/0103030, doi:10.1103/PhysRevD.64.052002.
- 3250 [168] Fermilab E906 experiment, Drell-Yan Measurements of Nucleon and Nuclear Structure with the Fermilab Main Injector, D. F. Geesaman
3251 and P. E. Reimer, spokespersons; <http://www.phy.anl.gov/mep/SeaQuest/index.html>.
- 3252 [169] V. M. Abazov, et al., Measurement of the shape of the boson rapidity distribution for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at $\sqrt{s}=1.96$ -
3253 TeV, Phys. Rev. D76 (2007) 012003. arXiv:hep-ex/0702025, doi:10.1103/PhysRevD.76.012003.
- 3254 [170] 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.
3255 Lett. B692 (2010) 232–239. arXiv:0908.3914, doi:10.1016/j.physletb.2010.06.043.
- 3256 [171] V. M. Abazov, et al., Measurement of the muon charge asymmetry in $p\bar{p} \rightarrow W+X \rightarrow \mu\nu + X$ events at $\sqrt{s}=1.96$ TeV, Phys.Rev. D88 (2013)
3257 091102. arXiv:1309.2591, doi:10.1103/PhysRevD.88.091102.
- 3258 [172] V. M. Abazov, et al., Measurement of the electron charge asymmetry in $p\bar{p} \rightarrow W + X \rightarrow e\nu + X$ decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV,
3259 Phys. Rev. D91 (3) (2015) 032007, [Erratum: Phys. Rev.D91,no.7,079901(2015)]. arXiv:1412.2862, doi:10.1103/PhysRevD.91.032007,
3260 10.1103/PhysRevD.91.079901.
- 3261 [173] 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
3262 (2009) 181801. arXiv:0901.2169, doi:10.1103/PhysRevLett.102.181801.
- 3263 [174] V. M. Abazov, et al., Measurement of the W Boson Production Charge Asymmetry in $p\bar{p} \rightarrow W + X \rightarrow e\nu + X$ Events at
3264 $\sqrt{s} = 1.96$ TeV, Phys. Rev. Lett. 112 (15) (2014) 151803, [Erratum: Phys. Rev. Lett.114,no.4,049901(2015)]. arXiv:1312.2895,
3265 doi:10.1103/PhysRevLett.114.049901, 10.1103/PhysRevLett.112.151803.
- 3266 [175] S. Chatrchyan, et al., Measurement of the Rapidity and Transverse Momentum Distributions of Z Bosons in pp Collisions at $\sqrt{s} = 7$ TeV,
3267 Phys. Rev. D85 (2012) 032002. arXiv:1110.4973, doi:10.1103/PhysRevD.85.032002.
- 3268 [176] 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 $\sqrt{s} = 7$
3269 TeV with the ATLAS detector, Phys.Rev. D85 (2012) 072004. arXiv:1109.5141.
- 3270 [177] S. Chatrchyan, et al., Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$
3271 TeV, JHEP 1312 (2013) 030. arXiv:1310.7291, doi:10.1007/JHEP12(2013)030.
- 3272 [178] V. Khachatryan, et al., Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at 8 TeV,
3273 Eur. Phys. J. C75 (4) (2015) 147. arXiv:1412.1115, doi:10.1140/epjc/s10052-015-3364-2.
- 3274 [179] 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,
3275 Phys.Lett. B725 (2013) 223. arXiv:1305.4192, doi:10.1016/j.physletb.2013.07.049.
- 3276 [180] 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)
3277 112. arXiv:1404.1212, doi:10.1007/JHEP06(2014)112.
- 3278 [181] S. Chatrchyan, et al., Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV,
3279 Phys.Rev.Lett. 109 (2012) 111806. arXiv:1206.2598, doi:10.1103/PhysRevLett.109.111806.
- 3280 [182] S. Chatrchyan, et al., Measurement of the muon charge asymmetry in inclusive pp to WX production at $\sqrt{s} = 7$ TeV and an improved
3281 determination of light parton distribution functions, Phys.Rev. D90 (2014) 032004. arXiv:1312.6283, doi:10.1103/PhysRevD.90.032004.
- 3282 [183] M. Aaboud, et al., Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS
3283 detector arXiv:1612.03016.
- 3284 [184] R. Aaij, et al., Measurement of the cross-section for $Z \rightarrow e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV, JHEP 1302 (2013) 106.
3285 arXiv:1212.4620, doi:10.1007/JHEP02(2013)106.
- 3286 [185] 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.
3287 arXiv:1505.07024, doi:10.1007/JHEP08(2015)039.
- 3288 [186] R. Aaij, et al., Measurement of forward $Z \rightarrow e^+e^-$ production at $\sqrt{s} = 8$ TeV, JHEP 05 (2015) 109. arXiv:1503.00963,
3289 doi:10.1007/JHEP05(2015)109.
- 3290 [187] 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.
3291 arXiv:1607.06495, doi:10.1007/JHEP09(2016)136.
- 3292 [188] 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,
3293 doi:10.1007/JHEP12(2014)079.
- 3294 [189] 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,
3295 doi:10.1007/JHEP01(2016)155.
- 3296 [190] R. Aaij, et al., Measurement of forward $W \rightarrow e\nu$ production in pp collisions at $\sqrt{s} = 8$ TeV, JHEP 10 (2016) 030. arXiv:1608.01484,
3297 doi:10.1007/JHEP10(2016)030.
- 3298 [191] R. V. Harlander, W. B. Kilgore, Next-to-next-to-leading order Higgs production at hadron colliders, Phys. Rev. Lett. 88 (2002) 201801.
3299 arXiv:hep-ph/0201206, doi:10.1103/PhysRevLett.88.201801.
- 3300 [192] C. Anastasiou, L. J. Dixon, K. Melnikov, F. Petriello, High precision QCD at hadron colliders: Electroweak gauge boson rapidity distribu-
3301 tions at NNLO, Phys. Rev. D69 (2004) 094008. arXiv:hep-ph/0312266, doi:10.1103/PhysRevD.69.094008.
- 3302 [193] C. Anastasiou, L. J. Dixon, K. Melnikov, F. Petriello, Dilepton rapidity distribution in the Drell-Yan process at NNLO in QCD, Phys. Rev.

- 3303 Lett. 91 (2003) 182002. arXiv:hep-ph/0306192, doi:10.1103/PhysRevLett.91.182002.
- 3304 [194] 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.
- 3305
- 3306 [195] 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.
- 3307
- 3308 [196] 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.
- 3309
- 3310 [197] 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.
- 3311
- 3312 [198] 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.
- 3313
- 3314 [199] F. A. Berends, R. Kleiss, Hard Photon Effects in W^\pm and Z^0 Decay, Z.Phys. C27 (1985) 365. doi:10.1007/BF01548639.
- 3315 [200] 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.
- 3316
- 3317 [201] 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.
- 3318
- 3319 [202] U. Baur, O. Brein, W. Hollik, C. Schappacher, D. Wackerth, 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.
- 3320
- 3321 [203] 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.
- 3322
- 3323 [204] 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.
- 3324
- 3325 [205] 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.
- 3326
- 3327 [206] 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.
- 3328
- 3329 [207] 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.
- 3330
- 3331 [208] 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.
- 3332
- 3333 [209] 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.
- 3334
- 3335 [210] 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.
- 3336
- 3337 [211] 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.
- 3338
- 3339 [212] 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.
- 3340
- 3341 [213] 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.
- 3342
- 3343 [214] 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.
- 3344
- 3345 [215] 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.
- 3346
- 3347 [216] 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.
- 3348
- 3349 [217] 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.
- 3350
- 3351 [218] 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.
- 3352
- 3353 [219] 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.
- 3354
- 3355 [220] 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.
- 3356
- 3357 [221] 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.
- 3358
- 3359 [222] 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.
- 3360
- 3361 [223] 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.
- 3362
- 3363 [224] G. Aad, et al., Measurement of the transverse momentum and ϕ_η^* distributions of Drell-Yan 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.
- 3364
- 3365 [225] V. Khachatryan, et al., Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton-proton collisions at 8 TeV, Phys. Lett. B749 (2015) 187–209. arXiv:1504.03511, doi:10.1016/j.physletb.2015.07.065.
- 3366
- 3367

- 3368 [226] G. Aad, et al., Measurement of the transverse momentum distribution of Z/γ^* bosons in proton-proton collisions at $\sqrt{s} = 7$ TeV with the
3369 ATLAS detector, *Phys.Lett. B*705 (2011) 415–434. arXiv:1107.2381, doi:10.1016/j.physletb.2011.10.018.
- 3370 [227] 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,
3371 *JHEP* 09 (2014) 145. arXiv:1406.3660, doi:10.1007/JHEP09(2014)145.
- 3372 [228] R. J. Gonsalves, J. Pawlowski, C.-F. Wai, QCD Radiative Corrections to Electroweak Boson Production at Large Transverse Momentum in
3373 Hadron Collisions, *Phys. Rev. D*40 (1989) 2245. doi:10.1103/PhysRevD.40.2245.
- 3374 [229] H. Baer, M. H. Reno, A Complete O (α_s) event generator for $p\bar{p} \rightarrow W^+X \rightarrow e^+$ neutrino X with parton showering, *Phys. Rev. D*44
3375 (1991) 3375–3378. doi:10.1103/PhysRevD.44.3375.
- 3376 [230] P. B. Arnold, R. P. Kauffman, W and Z production at next-to-leading order: From large $q(t)$ to small, *Nucl. Phys. B*349 (1991) 381–413.
3377 doi:10.1016/0550-3213(91)90330-Z.
- 3378 [231] E. Maina, S. Moretti, D. A. Ross, One loop weak corrections to γ / Z hadroproduction at finite transverse momentum,
3379 *Phys. Lett. B*593 (2004) 143–150, [Erratum: *Phys. Lett. B*614,216(2005)]. arXiv:hep-ph/0403050, doi:10.1016/j.physletb.2004.04.043,
3380 10.1016/j.physletb.2005.03.064.
- 3381 [232] J. H. Kuhn, A. Kulesza, S. Pozzorini, M. Schulze, Logarithmic electroweak corrections to hadronic Z+1 jet production at large transverse
3382 momentum, *Phys. Lett. B*609 (2005) 277–285. arXiv:hep-ph/0408308, doi:10.1016/j.physletb.2005.01.059.
- 3383 [233] J. H. Kuhn, A. Kulesza, S. Pozzorini, M. Schulze, One-loop weak corrections to hadronic production of Z bosons at large transverse
3384 momenta, *Nucl. Phys. B*727 (2005) 368–394. arXiv:hep-ph/0507178, doi:10.1016/j.nuclphysb.2005.08.019.
- 3385 [234] T. Becher, X. Garcia i Tormo, Electroweak Sudakov effects in W, Z and γ production at large transverse momentum, *Phys. Rev. D*88 (1)
3386 (2013) 013009. arXiv:1305.4202, doi:10.1103/PhysRevD.88.013009.
- 3387 [235] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, T. A. Morgan, The NNLO QCD corrections to Z boson production at
3388 large transverse momentum, *JHEP* 07 (2016) 133. arXiv:1605.04295, doi:10.1007/JHEP07(2016)133.
- 3389 [236] 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^*
3390 observables at the LHC, *JHEP* 11 (2016) 094. arXiv:1610.01843, doi:10.1007/JHEP11(2016)094.
- 3391 [237] R. Boughezal, J. M. Campbell, R. K. Ellis, C. Focke, W. T. Giele, X. Liu, F. Petriello, Z-boson production in association
3392 with a jet at next-to-next-to-leading order in perturbative QCD, *Phys. Rev. Lett.* 116 (15) (2016) 152001. arXiv:1512.01291,
3393 doi:10.1103/PhysRevLett.116.152001.
- 3394 [238] R. Boughezal, X. Liu, F. Petriello, Phenomenology of the Z-boson plus jet process at NNLO, *Phys. Rev. D*94 (7) (2016) 074015.
3395 arXiv:1602.08140, doi:10.1103/PhysRevD.94.074015.
- 3396 [239] R. D. Ball, et al., Parton distributions from high-precision collider data arXiv:1706.00428.
- 3397 [240] D. d’Enterria, J. Rojo, Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data, *Nucl.Phys.*
3398 *B*860 (2012) 311–338. arXiv:1202.1762.
- 3399 [241] 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-
3400 ex/9711017, doi:10.1103/PhysRevLett.81.2642.
- 3401 [242] L. Apanasevich, et al., Measurement of direct photon production at Tevatron fixed target energies, *Phys. Rev. D*70 (2004) 092009. arXiv:hep-
3402 ex/0407011, doi:10.1103/PhysRevD.70.092009.
- 3403 [243] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Parton distributions and the LHC: W and Z production, *Eur. Phys. J. C*14 (2000)
3404 133–145. arXiv:hep-ph/9907231, doi:10.1007/s100520050740, 10.1007/s100520000324.
- 3405 [244] R. Ichou, D. d’Enterria, Sensitivity of isolated photon production at TeV hadron colliders to the gluon distribution in the proton, *Phys. Rev.*
3406 *D*82 (2010) 014015. arXiv:1005.4529, doi:10.1103/PhysRevD.82.014015.
- 3407 [245] 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,
3408 *JHEP* 08 (2016) 005. arXiv:1605.03495, doi:10.1007/JHEP08(2016)005.
- 3409 [246] M. Aaboud, et al., Measurement of the cross section for inclusive isolated-photon production in pp collisions at $\sqrt{s} = 13$ TeV using the
3410 ATLAS detector arXiv:1701.06882.
- 3411 [247] 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,
3412 Submitted to: *Phys. Rev. D* arXiv:1703.00599.
- 3413 [248] 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. B*639 (2006) 151–158,
3414 [Erratum: *Phys. Lett. B*658,285(2008)]. arXiv:hep-ex/0511054, doi:10.1016/j.physletb.2007.06.047, 10.1016/j.physletb.2006.04.048.
- 3415 [249] 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
3416 using 4.6fb^{-1} , *Phys. Rev. D*89 (5) (2014) 052004. arXiv:1311.1440, doi:10.1103/PhysRevD.89.052004.
- 3417 [250] S. Chatrchyan, et al., Measurement of the Differential Cross Section for Isolated Prompt Photon Production in pp Collisions at 7 TeV,
3418 *Phys.Rev. D*84 (2011) 052011. arXiv:1108.2044, doi:10.1103/PhysRevD.84.052011.
- 3419 [251] G. Aad, et al., Centrality, rapidity and transverse momentum dependence of isolated prompt photon production in lead-lead col-
3420 lisions at $\sqrt{s_{NN}} = 2.76$ TeV measured with the ATLAS detector, *Phys. Rev. C*93 (3) (2016) 034914. arXiv:1506.08552,
3421 doi:10.1103/PhysRevC.93.034914.
- 3422 [252] S. Chatrchyan, et al., Measurement of isolated photon production in pp and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Phys. Lett. B*710 (2012)
3423 256–277. arXiv:1201.3093, doi:10.1016/j.physletb.2012.02.077.
- 3424 [253] S. Catani, M. Fontannaz, J. P. Guillet, E. Pilon, Cross-section of isolated prompt photons in hadron hadron collisions, *JHEP* 05 (2002) 028.
3425 arXiv:hep-ph/0204023, doi:10.1088/1126-6708/2002/05/028.
- 3426 [254] J. H. Kuhn, A. Kulesza, S. Pozzorini, M. Schulze, Electroweak corrections to hadronic photon production at large transverse momenta,
3427 *JHEP* 03 (2006) 059. arXiv:hep-ph/0508253, doi:10.1088/1126-6708/2006/03/059.
- 3428 [255] J. M. Campbell, R. K. Ellis, C. Williams, Direct photon production at next-to-next-to-leading order arXiv:1612.04333.
- 3429 [256] M. D. Schwartz, Precision direct photon spectra at high energy and comparison to the 8 TeV ATLAS data, *JHEP* 09 (2016) 005.
3430 arXiv:1606.02313, doi:10.1007/JHEP09(2016)005.
- 3431 [257] T. A. Aaltonen, et al., Combination of measurements of the top-quark pair production cross section from the Tevatron Collider, *Phys.Rev.*
3432 *D*89 (2014) 072001. arXiv:1309.7570, doi:10.1103/PhysRevD.89.072001.

- 3433 [258] 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
3434 using final states with two high-pt leptons, JHEP 1205 (2012) 059. arXiv:1202.4892, doi:10.1007/JHEP05(2012)059.
- 3435 [259] G. Aad, et al., Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 7$ TeV using kinematic information of lepton+jets
3436 events arXiv:ATLAS-CONF-2011-121, ATLAS-COM-CONF-2011-132.
- 3437 [260] 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
3438 jets arXiv:ATLAS-CONF-2013-097, ATLAS-COM-CONF-2013-112.
- 3439 [261] 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
3440 (2014) 024. arXiv:1312.7582, doi:10.1007/JHEP02(2014)024.
- 3441 [262] 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
3442 (2012) 067. arXiv:1208.2671, doi:10.1007/JHEP11(2012)067.
- 3443 [263] 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.
3444 B720 (2013) 83–104. arXiv:1212.6682, doi:10.1016/j.physletb.2013.02.021.
- 3445 [264] 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
3446 the ATLAS detector, Eur. Phys. J. C76 (10) (2016) 538. arXiv:1511.04716, doi:10.1140/epjc/s10052-016-4366-4.
- 3447 [265] 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.
3448 J. C75 (11) (2015) 542. arXiv:1505.04480, doi:10.1140/epjc/s10052-015-3709-x.
- 3449 [266] 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
3450 impact on parton distribution functions arXiv:1703.01630.
- 3451 [267] M. Czakon, P. Fiedler, A. Mitov, The total top quark pair production cross-section at hadron colliders through $\mathcal{O}(\alpha_s^4)$, Phys.Rev.Lett. 110
3452 (2013) 252004. arXiv:1303.6254, doi:10.1103/PhysRevLett.110.252004.
- 3453 [268] M. Czakon, A. Mitov, NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction, JHEP 1301 (2013) 080.
3454 arXiv:1210.6832, doi:10.1007/JHEP01(2013)080.
- 3455 [269] P. Baerzreuther, M. Czakon, A. Mitov, Percent level precision physics at the Tevatron: first genuine NNLO QCD corrections to $q\bar{q} \rightarrow t\bar{t}$
3456 + X arXiv:1204.5201.
- 3457 [270] M. Czakon, D. Heymes, A. Mitov, Dynamical scales for multi-TeV top-pair production at the LHC, JHEP 04 (2017) 071. arXiv:1606.03350,
3458 doi:10.1007/JHEP04(2017)071.
- 3459 [271] M. Czakon, D. Heymes, A. Mitov, High-precision differential predictions for top-quark pairs at the LHC, Phys. Rev. Lett. 116 (8) (2016)
3460 082003. arXiv:1511.00549, doi:10.1103/PhysRevLett.116.082003.
- 3461 [272] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, M. Zaro, Top-pair production at the LHC through NNLO QCD and NLO
3462 EW arXiv:1705.04105.
- 3463 [273] 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)
3464 479. arXiv:1606.01915, doi:10.1140/epjc/s10052-016-4318-z.
- 3465 [274] J. Gao, A. S. Papanastasiou, Top-quark pair-production and decay at high precision arXiv:1705.08903.
- 3466 [275] M. Czakon, D. Heymes, A. Mitov, fastNLO tables for NNLO top-quark pair differential distributions arXiv:1704.08551.
- 3467 [276] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera, J. Rojo, Pinning down the large- x gluon with NNLO top-quark pair differential
3468 distributions, JHEP 04 (2017) 044. arXiv:1611.08609, doi:10.1007/JHEP04(2017)044.
- 3469 [277] M. Czakon, M. L. Mangano, A. Mitov, J. Rojo, Constraints on the gluon PDF from top quark pair production at hadron colliders, JHEP
3470 1307 (2013) 167. arXiv:1303.7215, doi:10.1007/JHEP07(2013)167.
- 3471 [278] M. Beneke, P. Falgari, S. Klein, J. Piclum, C. Schwinn, et al., Inclusive Top-Pair Production Phenomenology with TOPIX, JHEP 1207
3472 (2012) 194. arXiv:1206.2454, doi:10.1007/JHEP07(2012)194.
- 3473 [279] M. Guzzi, K. Lipka, S.-O. Moch, Top-quark pair production at hadron colliders: differential cross section and phenomenological applications
3474 with DiffTop, JHEP 01 (2015) 082. arXiv:1406.0386, doi:10.1007/JHEP01(2015)082.
- 3475 [280] F. Maltoni, G. Ridolfi, M. Ubiali, b-initiated processes at the LHC: a reappraisal, JHEP 07 (2012) 022, [Erratum: JHEP04,095(2013)].
3476 arXiv:1203.6393, doi:10.1007/JHEP04(2013)095, 10.1007/JHEP07(2012)022.
- 3477 [281] S. Forte, D. Napoletano, M. Ubiali, Higgs production in bottom-quark fusion in a matched scheme, Phys. Lett. B751 (2015) 331–337.
3478 arXiv:1508.01529, doi:10.1016/j.physletb.2015.10.051.
- 3479 [282] M. Brucherseifer, F. Caola, K. Melnikov, On the NNLO QCD corrections to single-top production at the LHC, Phys. Lett. B736 (2014)
3480 58–63. arXiv:1404.7116, doi:10.1016/j.physletb.2014.06.075.
- 3481 [283] M. Aaboud, et al., Fiducial, total and differential cross-section measurements of t -channel single top-quark production in pp collisions at 8
3482 TeV using data collected by the ATLAS detector arXiv:1702.02859.
- 3483 [284] M. Aaboud, et al., Measurement of the inclusive cross-sections of single top-quark and top-antiquark t -channel production in pp collisions
3484 at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP 04 (2017) 086. arXiv:1609.03920, doi:10.1007/JHEP04(2017)086.
- 3485 [285] R. Aaij, et al., Updated measurements of exclusive J/ψ and $\psi(2S)$ production cross-sections in pp collisions at $\sqrt{s} = 7$ TeV, J. Phys. G41
3486 (2014) 055002. arXiv:1401.3288, doi:10.1088/0954-3899/41/5/055002.
- 3487 [286] T. L. Collaboration, Central exclusive production of J/ψ and $\psi(2S)$ mesons in pp collisions at $\sqrt{s} = 13$ TeV.
- 3488 [287] B. B. Abelev, et al., Exclusive J/ψ photoproduction off protons in ultra-peripheral p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, Phys. Rev. Lett.
3489 113 (23) (2014) 232504. arXiv:1406.7819, doi:10.1103/PhysRevLett.113.232504.
- 3490 [288] 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,
3491 JHEP 11 (2013) 085. arXiv:1307.7099, doi:10.1007/JHEP11(2013)085.
- 3492 [289] 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)
3493 035002. arXiv:1507.06942, doi:10.1088/0954-3899/43/3/035002.
- 3494 [290] D. Yu. Ivanov, A. Schafer, L. Szymanowski, G. Krasnikov, Exclusive photoproduction of a heavy vector meson in QCD, Eur. Phys.
3495 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,
3496 10.1140/epjc/s10052-015-3298-8.
- 3497 [291] 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.
- [292] 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.
- [293] 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. C*66 (2010) 503. arXiv:0911.2985, doi:10.1140/epjc/s10052-010-1255-0.
- [294] M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, F. Stober, Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes arXiv:1109.1310.
- [295] 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.
- [296] 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.
- [297] 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.
- [298] 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.
- [299] J. Pumplin, Parametrization dependence and $\Delta\chi^2$ in parton distribution fitting, *Phys.Rev. D*82 (2010) 114020. arXiv:0909.5176, doi:10.1103/PhysRevD.82.114020.
- [300] R. D. Ball, E. R. Nocera, J. Rojo, The asymptotic behaviour of parton distributions at small and large x , *Eur. Phys. J. C*76 (7) (2016) 383. arXiv:1604.00024, doi:10.1140/epjc/s10052-016-4240-4.
- [301] A. I. Signal, A. W. Thomas, Possible Strength of the Nonperturbative Strange Sea of the Nucleon, *Phys. Lett. B*191 (1987) 205. doi:10.1016/0370-2693(87)91348-7.
- [302] 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.
- [303] J. Pumplin, D. R. Stump, W. K. Tung, Multivariate fitting and the error matrix in global analysis of data, *Phys. Rev. D*65 (2001) 014011. arXiv:hep-ph/0008191, doi:10.1103/PhysRevD.65.014011.
- [304] P. M. Nadolsky, Z. Sullivan, PDF uncertainties in W H production at Tevatron arXiv:hep-ph/0110378, doi:OSTI/788251.
- [305] 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.
- [306] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, J. Rojo, Unbiased determination of the proton structure function $f_2(p)$ with estimation, *JHEP* 03 (2005) 080. arXiv:hep-ph/0501067.
- [307] M. Dittmar, et al., Parton Distributions arXiv:0901.2504.
- [308] S. Carrazza, S. Forte, Z. Kassabov, J. I. Latorre, J. Rojo, An Unbiased Hessian Representation for Monte Carlo PDFs, *Eur. Phys. J. C*75 (8) (2015) 369. arXiv:1505.06736, doi:10.1140/epjc/s10052-015-3590-7.
- [309] 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. D*65 (2001) 014012. arXiv:hep-ph/0101051, doi:10.1103/PhysRevD.65.014012.
- [310] J. **Particle Data Group**, Beringer, et al., Review of Particle Physics (RPP), *Phys.Rev. D*86 (2012) 010001. doi:10.1103/PhysRevD.86.010001.
- [311] 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.
- [312] J. Gao, P. Nadolsky, A meta-analysis of parton distribution functions, *JHEP* 1407 (2014) 035. arXiv:1401.0013, doi:10.1007/JHEP07(2014)035.
- [313] S. Carrazza, J. I. Latorre, J. Rojo, G. Watt, A compression algorithm for the combination of PDF sets, *Eur. Phys. J. C*75 (2015) 474. arXiv:1504.06469, doi:10.1140/epjc/s10052-015-3703-3.
- [314] J. Pumplin, J. Huston, H. L. Lai, W.-K. Tung, C. P. Yuan, Collider Inclusive Jet Data and the Gluon Distribution, *Phys. Rev. D*80 (2009) 014019. arXiv:0904.2424, doi:10.1103/PhysRevD.80.014019.
- [315] 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
- [316] 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>
- [317] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., Reweighting and Unweighting of Parton Distributions and the LHC μ lepton asymmetry data, *Nucl.Phys. B*855 (2012) 608–638. arXiv:1108.1758, doi:10.1016/j.nuclphysb.2011.10.018.
- [318] R. D. Ball, et al., Reweighting NNPDFs: the W lepton asymmetry, *Nucl. Phys. B*849 (2011) 112–143. arXiv:1012.0836, doi:10.1016/j.nuclphysb.2011.03.017.
- [319] S. Camarda, et al., QCD analysis of W- and Z-boson production at Tevatron, *Eur. Phys. J. C*75 (9) (2015) 458. arXiv:1503.05221, doi:10.1140/epjc/s10052-015-3655-7.
- [320] H. Plathow-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.
- [321] 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.
- [322] 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.

- 3563 [323] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, et al., LHAPDF6: parton density access in the LHC precision era, *Eur.Phys.J. C* 75
3564 (2015) 132. arXiv:1412.7420, doi:10.1140/epjc/s10052-015-3318-8.
- 3565 [324] 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.
- 3566 [325] 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
3567 Scattering at High Q^2 , *Phys. Lett. B* 223 (1989) 485. doi:10.1016/0370-2693(89)91637-7.
- 3568 [326] 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
3569 Scattering at High Q^2 , *Phys. Lett. B* 237 (1990) 592. doi:10.1016/0370-2693(90)91231-Y.
- 3570 [327] J. P. Berge, et al., A Measurement of Differential Cross-Sections and Nucleon Structure Functions in Charged Current Neutrino Interactions
3571 on Iron, *Z. Phys. C* 49 (1991) 187–224. doi:10.1007/BF01555493.
- 3572 [328] A. Aktas, et al., Measurement of $F_2(c\bar{c})$ and $F_2(b\bar{b})$ at high Q^2 using the H1 vertex detector at HERA, *Eur. Phys. J. C* 40 (2005) 349–359.
3573 arXiv:hep-ex/0411046, doi:10.1140/epjc/s2005-02154-8.
- 3574 [329] M. Guzzi, P. M. Nadolsky, H.-L. Lai, C.-P. Yuan, General-Mass Treatment for Deep Inelastic Scattering at Two-Loop Accuracy, *Phys.Rev.*
3575 *D* 86 (2012) 053005. arXiv:1108.5112, doi:10.1103/PhysRevD.86.053005.
- 3576 [330] G. Moreno, et al., Dimuon production in proton - copper collisions at $\sqrt{s} = 38.8$ -GeV, *Phys. Rev. D* 43 (1991) 2815–2836.
3577 doi:10.1103/PhysRevD.43.2815.
- 3578 [331] F. Abe, et al., Forward-backward charge asymmetry of electron pairs above the Z^0 pole, *Phys. Rev. Lett.* 77 (1996) 2616–2621.
3579 doi:10.1103/PhysRevLett.77.2616.
- 3580 [332] D. E. Acosta, et al., Measurement of the forward-backward charge asymmetry from $W \rightarrow e\nu$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV,
3581 *Phys. Rev. D* 71 (2005) 051104. arXiv:hep-ex/0501023, doi:10.1103/PhysRevD.71.051104.
- 3582 [333] V. M. Abazov, et al., Measurement of the muon charge asymmetry from W boson decays, *Phys. Rev. D* 77 (2008) 011106. arXiv:0709.4254,
3583 doi:10.1103/PhysRevD.77.011106.
- 3584 [334] V. M. Abazov, et al., Measurement of the ratios of the $Z/\gamma^* + \nu\bar{\nu}$ jet production cross sections to the total inclusive
3585 Z/γ^* cross section in p anti- p collisions at $s^{*1/2} = 1.96$ -TeV, *Phys. Lett. B* 658 (2008) 112–119. arXiv:hep-ex/0608052,
3586 doi:10.1016/j.physletb.2007.10.046.
- 3587 [335] C. Balazs, C. P. Yuan, Soft gluon effects on lepton pairs at hadron colliders, *Phys. Rev. D* 56 (1997) 5558–5583. arXiv:hep-ph/9704258,
3588 doi:10.1103/PhysRevD.56.5558.
- 3589 [336] M. Guzzi, P. M. Nadolsky, B. Wang, Nonperturbative contributions to a resummed leptonic angular distribution in inclusive neutral vector
3590 boson production, *Phys. Rev. D* 90 (1) (2014) 014030. arXiv:1309.1393, doi:10.1103/PhysRevD.90.014030.
- 3591 [337] H.-L. Lai, et al., New parton distributions for collider physics, *Phys. Rev. D* 82 (2010) 074024. arXiv:1007.2241,
3592 doi:10.1103/PhysRevD.82.074024.
- 3593 [338] P. M. Nadolsky, et al., Implications of CTEQ global analysis for collider observables, *Phys. Rev. D* 78 (2008) 013004. arXiv:0802.0007,
3594 doi:10.1103/PhysRevD.78.013004.
- 3595 [339] J. F. Owens, A. Accardi, W. Melnitchouk, Global parton distributions with nuclear and finite- Q^2 corrections, *Phys. Rev. D* 87 (9) (2013)
3596 094012. arXiv:1212.1702, doi:10.1103/PhysRevD.87.094012.
- 3597 [340] 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,
3598 doi:10.1007/JHEP06(2012)058.
- 3599 [341] H.-L. Lai, et al., Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions, *Phys. Rev. D* 82 (2010) 054021.
3600 arXiv:1004.4624, doi:10.1103/PhysRevD.82.054021.
- 3601 [342] 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,
3602 *PoS DIS2015* (2015) 166.
- 3603 [343] C. Schmidt, J. Pumplin, D. Stump, C. P. Yuan, CT14QED PDFs from Isolated Photon Production in Deep Inelastic Scat-
3604 tering arXiv:1509.02905.
- 3605 [344] S. Chekanov, et al., Measurement of isolated photon production in deep inelastic ep scattering, *Phys.Lett. B* 687 (2010) 16–25.
3606 arXiv:0909.4223, doi:10.1016/j.physletb.2010.02.045.
- 3607 [345] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Parton distributions: A New global analysis, *Eur. Phys. J. C* 4 (1998) 463–496.
3608 arXiv:hep-ph/9803445, doi:10.1007/s100529800904, 10.1007/s100520050220.
- 3609 [346] R. S. Thorne, R. G. Roberts, An Ordered analysis of heavy flavor production in deep inelastic scattering, *Phys. Rev. D* 57 (1998) 6871–6898.
3610 arXiv:hep-ph/9709442, doi:10.1103/PhysRevD.57.6871.
- 3611 [347] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Physical gluons and high-E(T) jets, *Phys. Lett. B* 604 (2004) 61–68. arXiv:hep-
3612 ph/0410230, doi:10.1016/j.physletb.2004.10.040.
- 3613 [348] A. Martin, A. T. Mathijssen, W. Stirling, R. Thorne, B. Watt, et al., Extended Parameterisations for MSTW PDFs and their effect on Lepton
3614 Charge Asymmetry from W Decays, *Eur.Phys.J. C* 73 (2) (2013) 2318. arXiv:1211.1215, doi:10.1140/epjc/s10052-013-2318-9.
- 3615 [349] T. Bolton, Determining the CKM parameter V_{cd} from νN charm production arXiv:hep-ex/9708014.
- 3616 [350] L. A. Harland-Lang, A. D. Martin, P. Motylinski, R. S. Thorne, Uncertainties on α_S in the MMHT2014 global PDF analysis and implications
3617 for SM predictions arXiv:1506.05682.
- 3618 [351] R. Thorne, Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections, *Phys.Rev. D* 86
3619 (2012) 074017. arXiv:1201.6180, doi:10.1103/PhysRevD.86.074017.
- 3620 [352] L. A. Harland-Lang, A. D. Martin, P. Motylinski, R. S. Thorne, Charm and beauty quark masses in the MMHT2014 global PDF analysis,
3621 *Eur. Phys. J. C* 76 (1) (2016) 10. arXiv:1510.02332, doi:10.1140/epjc/s10052-015-3843-5.
- 3622 [353] R. S. Thorne, L. A. Harland-Lang, A. D. Martin, P. Motylinski, The Effect of Final HERA inclusive Cross Section Data MMHT2014 PDFs,
3623 in: *Proceedings, 2015 European Physical Society Conference on High Energy Physics (EPS-HEP 2015)*, 2015. arXiv:1508.06621.
3624 URL <http://inspirehep.net/record/1389858/files/arXiv:1508.06621.pdf>
- 3625 [354] L. A. Harland-Lang, A. D. Martin, R. Nathvani, R. S. Thorne, MMHT PDFs: updates and outlook, in: *23rd Cracow Epiphany Conference*
3626 *on Particle Theory Meets the First Data from LHC Run 2 Cracow, Poland, January 9-12, 2017*, 2017. arXiv:1704.00162.
3627 URL <https://inspirehep.net/record/1589299/files/arXiv:1704.00162.pdf>

- [355] S. Forte, J. I. Latorre, L. Magnea, A. Piccione, Determination of α_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.
- [356] 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.
- [357] 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.
- [358] 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.
- [359] 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.
- [360] R. D. Ball, et al., Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology, Nucl. Phys. B849 (2011) 296. arXiv:1101.1300.
- [361] 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.
- [362] 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.
- [363] C. Patrignani, et al., Review of Particle Physics, Chin. Phys. C40 (10) (2016) 100001. doi:10.1088/1674-1137/40/10/100001.
- [364] 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.
- [365] 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.
- [366] 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.
- [367] H. Georgi, H. D. Politzer, Freedom at moderate energies: Masses in color dynamics, Phys. Rev. D14 (1976) 1829.
- [368] 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.
- [369] 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.
- [370] M. Kramer, I. 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.
- [371] H. Georgi, H. D. Politzer, Freedom at Moderate Energies: Masses in Color Dynamics, Phys. Rev. D14 (1976) 1829. doi:10.1103/PhysRevD.14.1829.
- [372] 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.
- [373] 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.
- [374] 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.
- [375] 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.
- [376] 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.
- [377] 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.
- [378] 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.
- [379] 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.
- [380] 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.
- [381] 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.
- [382] F. Giuli, et al., The photon PDF from high-mass Drell Yan data at the LHC arXiv:1701.08553.
- [383] 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.
- [384] F. Hautmann, H. Jung, A. Lelek, V. Radescu, R. Zlebick, Soft-gluon resolution scale in QCD evolution equations arXiv:1704.01757.
- [385] 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.
- [386] 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.
- [387] 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.
- [388] C. Collaboration, Measurement of the inclusive $\bar{t}\bar{t}$ cross section at $\sqrt{s} = 5.02$ TeV.
- [389] 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.
- [390] 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.
- [391] 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.
- [392] 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.
- [393] 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.
- [394] 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.
- [395] 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.
- [396] 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.
- [397] 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.
- [398] 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.
- [399] R. Sadykov, Impact of QED radiative corrections on Parton Distribution Functions arXiv:1401.1133.
- [400] 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.
- [401] S. Kallweit, J. M. Lindert, S. Pozzorini, M. Schonherr, NLO QCD+EW predictions for $2\ell\nu$ diboson signatures at the LHC arXiv:1705.00598.
- [402] 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.
- [403] 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.
- [404] V. Bertone, S. Carrazza, Combining NNPDF3.0 and NNPDF2.3QED through the APFEL evolution code, *PoS DIS2016* (2016) 031. arXiv:1606.07130.
- [405] 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.
- [406] 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.
- [407] 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.
- [408] 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.
- [409] 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.
- [410] 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.
- [411] 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.
- [412] A. Mukherjee, C. Pisano, Manifestly covariant analysis of the QED Compton process in $e p \rightarrow \ell e \gamma p$ and $e p \rightarrow \ell e \gamma X$, *Eur. Phys. J.* C30 (2003) 477–486. arXiv:hep-ph/0306275, doi:10.1140/epjc/s2003-01308-0.
- [413] 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.
- [414] D. Bourilkov, Photon-induced Background for Dilepton Searches and Measurements in pp Collisions at 13 TeV arXiv:1606.00523.
- [415] 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.
- [416] 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>
- [417] J. M. Campbell, D. Wackerroth, 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.
- [418] 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.
- [419] S. Kallweit, J. M. Lindert, P. Maierhofer, 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.
- [420] 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.
- [421] 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.

- 3758 [422] L. Randall, R. Sundrum, An Alternative to compactification, *Phys.Rev.Lett.* 83 (1999) 4690–4693. arXiv:hep-th/9906064,
3759 doi:10.1103/PhysRevLett.83.4690.
- 3760 [423] L. Randall, R. Sundrum, A Large mass hierarchy from a small extra dimension, *Phys.Rev.Lett.* 83 (1999) 3370–3373. arXiv:hep-ph/9905221,
3761 doi:10.1103/PhysRevLett.83.3370.
- 3762 [424] J. Alwall, et al., MadGraph/MadEvent v4: The New Web Generation, *JHEP* 09 (2007) 028. arXiv:0706.2334.
- 3763 [425] I. Brivio, M. Trott, The Standard Model as an Effective Field Theory arXiv:1706.08945.
- 3764 [426] 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,
3765 *Phys. Rev. D* 82 (2010) 114023. arXiv:1010.4315, doi:10.1103/PhysRevD.82.114023.
- 3766 [427] 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
3767 LHC and ILC, *Eur. Phys. J. C* 74 (2014) 3046. arXiv:1407.3792, doi:10.1140/epjc/s10052-014-3046-5.
- 3768 [428] M. Benayoun, P. David, L. DelBuono, F. Jegerlehner, Muon $g - 2$ estimates: can one trust effective Lagrangians and global fits?, *Eur. Phys.*
3769 *J. C* 75 (12) (2015) 613. arXiv:1507.02943, doi:10.1140/epjc/s10052-015-3830-x.
- 3770 [429] G. Bozzi, L. Citelli, A. Vicini, Parton density function uncertainties on the W boson mass measurement from the lepton transverse momen-
3771 tum distribution, *Phys. Rev. D* 91 (11) (2015) 113005. arXiv:1501.05587, doi:10.1103/PhysRevD.91.113005.
- 3772 [430] G. Bozzi, L. Citelli, M. Vesterinen, A. Vicini, Prospects for improving the LHC W boson mass measurement with forward muons, *Eur.*
3773 *Phys. J. C* 75 (12) (2015) 601. arXiv:1508.06954, doi:10.1140/epjc/s10052-015-3810-1.
- 3774 [431] M. Aaboud, et al., Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector arXiv:1701.07240.
- 3775 [432] S. Dimopoulos, S. Raby, F. Wilczek, Supersymmetry and the Scale of Unification, *Phys. Rev. D* 24 (1981) 1681–1683.
3776 doi:10.1103/PhysRevD.24.1681.
- 3777 [433] D. Becciolini, M. Gillioz, M. Nardecchia, F. Sannino, M. Spannowsky, Constraining new colored matter from the ratio of 3 to 2 jets
3778 cross sections at the LHC, *Phys. Rev. D* 91 (1) (2015) 015010, [Addendum: *Phys. Rev. D* 92, no. 7, 079905 (2015)]. arXiv:1403.7411,
3779 doi:10.1103/PhysRevD.91.015010, 10.1103/PhysRevD.92.079905.
- 3780 [434] M. Aaboud, et al., Determination of the strong coupling constant α_s from transverse energy-energy correlations in multijet events at $\sqrt{s} = 8$
3781 TeV using the ATLAS detector arXiv:1707.02562.
- 3782 [435] 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}
3783 $= 7$ TeV and first determination of the strong coupling constant in the TeV range, *Eur. Phys. J. C* 73 (10) (2013) 2604. arXiv:1304.7498,
3784 doi:10.1140/epjc/s10052-013-2604-6.
- 3785 [436] S. Chatrchyan, et al., Determination of the top-quark pole mass and strong coupling constant from the t t-bar production cross
3786 section in pp collisions at $\sqrt{s} = 7$ TeV, *Phys.Lett. B* 728 (2014) 496. arXiv:1307.1907, doi:10.1016/j.physletb.2014.08.040,
3787 10.1016/j.physletb.2013.12.009.
- 3788 [437] J. Rojo, Constraints on parton distributions and the strong coupling from LHC jet data, *Int. J. Mod. Phys. A* 30 (2015) 1546005.
3789 arXiv:1410.7728, doi:10.1142/S0217751X15460057.
- 3790 [438] 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.
3791 arXiv:1410.6810, doi:10.1007/JHEP02(2015)007.
- 3792 [439] M. Cacciari, N. Houdeau, Meaningful characterisation of perturbative theoretical uncertainties, *JHEP* 1109 (2011) 039. arXiv:1105.5152,
3793 doi:10.1007/JHEP09(2011)039.
- 3794 [440] E. Bagnaschi, M. Cacciari, A. Guffanti, L. Jenniches, An extensive survey of the estimation of uncertainties from missing higher orders in
3795 perturbative calculations arXiv:1409.5036.
- 3796 [441] S. Forte, A. Isgr, G. Vita, Do we need N³LO Parton Distributions?, *Phys.Lett. B* 731 (2014) 136–140. arXiv:1312.6688,
3797 doi:10.1016/j.physletb.2014.02.027.
- 3798 [442] 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
3799 Linear Collider Technical Design Report - Volume 2: Physics arXiv:1306.6352.
- 3800 [443] L. Linssen, A. Miyamoto, M. Stanitzki, H. Weerts, Physics and Detectors at CLIC: CLIC Conceptual Design Report arXiv:1202.5940,
3801 doi:10.5170/CERN-2012-003.
- 3802 [444] 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.
- 3803 [445] R. Contino, et al., Physics at a 100 TeV pp collider: Higgs and EW symmetry breaking studies arXiv:1606.09408.
- 3804 [446] D. Boer, M. Diehl, R. Milner, R. Venugopalan, W. Vogelsang, et al., Gluons and the quark sea at high energies: Distributions, polarization,
3805 tomography arXiv:1108.1713.