UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata
Dipartimento di Fisica

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata
Ciclo XXVII

Parton distribution functions with QED corrections

Settore Scientifico Disciplinare FIS/02

Tesi di Dottorato di:
Stefano CARRAZZA

Supervisore: Prof. Stefano FORTE

Coordinatore: Prof. Marco BERSANELLI

Anno Accademico 2014-2015

Committee of the final examination:

External Referee:
Prof. José I. LATORRE

External Members:
Dr. Michelangelo L. MANGANO
Prof. Paolo GAMBINO

Internal Members:
Prof. Alessandro VICINI

Final examination:

July 6, 2015
Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

MIUR subjects:

FIS/02 - Fisica Teorica, Modelli e Metodi Matematici

PACS:

02.70.Uu, 07.05.Mh, 12.38.-t, 13.60.Hb, 89.20.Ft, 12.15.Lk

Keywords:

Parton Distribution Functions (PDFs), Electroweak corrections to PDFs, Photon PDF

Official Collaboration logos:

Cover illustration:

S. Carrazza, *PDF covariance*, generated with 1000 replicas of NNPDF3.0 NLO.

Internal illustrations:

All plots have been produced by Stefano Carrazza
Typeset by Stefano Carrazza using LaTeX

To Vera Lucia, Dante e Roxana

Abstract

We present the first unbiased determination of parton distribution functions (PDFs) with electroweak corrections. The aim of this thesis is to provide an exhaustive description of the theoretical framework and the technical implementation which leads to the determination of a set of PDFs which includes the photon PDF and quantum electrodynamics (QED) contributions to parton evolution. First, we introduce and motivate the need of including electroweak corrections to PDFs, providing phenomenological examples and presenting an overview of the current state of the art in PDF fits. The theoretical implications of such corrections are then described through the implementation of the combined QCD⊗QED evolution in `APFEL`, a public code for the solution of the PDF evolution developed particularly for this thesis. We proceed by presenting the new structure of the Neural-Network PDF (NNPDF) methodology used for the extraction of this set of PDFs with QED corrections. We then provide a first determination of the full set of PDFs based on deep-inelastic scattering data and LHC data for W and Z/γ^* Drell-Yan production, using leading-order QED and NLO or NNLO QCD: the so-called NNPDF2.3QED set of PDFs. We perform a preliminary investigation of the phenomenological implications of NNPDF2.3QED set, in particular, focusing on the photon-induced corrections to direct photon production at HERA, high-mass dilepton and W pair production at the LHC and finally, providing a first determination of lepton PDFs through the `APFEL` evolution. We conclude with a summary of the technological upgrades required for the improvement of future PDF determinations with electroweak corrections.

List of Publications

Refereed publications

- V. Bertone, S. Carrazza and E. R. Nocera, *Reference results for time-like evolution up to $\mathcal{O}(\alpha_s^3)$*, JHEP **1503** (2015) 46
  arXiv:1501.00494, DOI:10.1007/JHEP03(2015)046

- R. D. Ball *et al.*, *Parton distributions for the LHC Run II*, JHEP (2015)
  arXiv:1410:8849, DOI:10.1007/JHEP04(2015)040

- S. Carrazza *et al.*, *APFEL Web: a web-based application for the graphical visualization of parton distribution functions*, J. Phys. G: Nucl. Part. Phys. **45** 057001
  arXiv:1410.5456, DOI:10.1088/0954-3899/42/5/057001

- S. Carrazza and J. Pires, *Perturbative QCD description of jet data from LHC Run-I and Tevatron Run-II*, JHEP **1410** (2014) 145
  arXiv:1407.7031, DOI: 10.1007/JHEP10(2014)145

- P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, Eur. Phys. J. C **74** (2014) 8, 3024
  arXiv:1404.5630, DOI: 10.1140/epjc/s10052-014-3024-y

- V. Bertone, S. Carrazza and 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

- R. D. Ball *et al.*, *Parton distributions with QED corrections*, Nucl. Phys. B **877** (2013) 290
  arXiv:1308.0598, DOI: 10.1016/j.nuclphysb.2013.10.010

- R. D. Ball *et al.*, *Parton Distribution Benchmarking with LHC Data*, JHEP **1304** (2013) 125
  arXiv:1211.5142, DOI: 10.1007/JHEP04(2013)125

- 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

Publications in reports

- J. M. Campbell *et. al.*, *Working Group Report: Quantum Chromodynamics*, Snowmass community summer study 2013
  arXiv:1310.5189

Publications in conference proceedings

- S. Carrazza, *Disentangling electroweak effects in Z-boson production*, Nuovo Cimento, La Thuile 2014
  arXiv:1405.1728

- S. Carrazza, S. Forte and J. Rojo, *Parton Distributions and Event Generators*, ISMD 2013
  arXiv:1311.5887

- S. Carrazza, *Towards the determination of the photon parton distribution function constrained by LHC data*, PoS DIS **2013** (2013) 279
  arXiv:1307.1131

- S. Carrazza, *Towards an unbiased determination of parton distributions with QED corrections*, Les Rencontres de Moriond, p.357-360 2013
  arXiv:1305.4179

Acknowledgements

First of all, I am infinitely grateful to my parents who have always taken part of my life and supported my activities. They have provided me an excellent cultural environment during my existence and they have always helped me to fulfill my dreams. I dedicate this thesis to them and to my sister.

I also would like to thank my supervisor, Prof. Stefano Forte, who guided me during my doctorate with passion, determination and a lot patience. I am immensely thankful for his encouragement in becoming part of the scientific community, in particular by motivating and supporting my participation at several doctoral schools and international conferences.

I thank all my collaborators for their support and motivation too. In particular, a special thanks to Juan Rojo, Valerio Bertone, Nathan Hartland and Emanuele Nocera. Thanks to these interactions I have learned and developed a solid scientific approach.

I would like to thank Prof. José I. Latorre for accepting to be the referee of this work. I thank as well the external and internal members of the final examination committee: Dr. Michelangelo L. Mangano, which provided me an exciting experience as visitor at CERN, Prof. Paolo Gambino and Prof. Alessandro Vicini.

Finally, I would like to thank all people from the Milan Department of Physics with whom I have interacted during these last wonderful years, in particular the members of the Milan Theoretical Physics Group: Giancarlo Ferrera, João Pires. I take this opportunity also to thank Alfio Ferrara, Silvia Salini and Massimo Florio in the context of the EIBURS project.

Stefano Carrazza

Contents

Introduction

During the past years we have witnessed several discoveries predicted by the Standard Model (SM) of particle physics, for example, the discovery of the tau neutrino [1], the top quark [2], and recently in 2012, the Higgs boson [3, 4] and its subsequent phenomenological characterization, thanks to the measurements performed by the Large Hadron Collider (LHC) at CERN.

The great success of the SM is practically due to the two theories which it is based: the Quantum Chromodynamics (QCD) and the Electroweak theory. The QCD is the theory of strong interactions between quarks, antiquarks and gluons (*partons*), meanwhile Quantum Electrodynamics (QED) and Weak interactions are described by the unified Electroweak theory. Both theories are in a continuous development of calculation techniques which improves the accuracy of theoretical predictions since the latter half of the 20th century. It is interesting to remark that even if the gauge groups of such theories are factored ($SU(3) \times SU(2) \times U(1)$) the theory is unique. First of all, we notice that these interactions are connected to each other through the mediation of common fundamental particles, moreover, from a technical point of view, the CKM matrix mixes the strong and weak interactions and the commutation of the generators of the strong and weak interactions constraints their form. This remark is essential in the context of this thesis where we present explicitly the combination of the QCD and QED theories.

Parton distribution functions (PDFs) are one of the most important ingredients for a realistic computation of any particle physics observable thanks to the collinear factorization property of QCD states. This formalism expresses any cross-section, σ, as the convolution product

$$\sigma = \hat{\sigma} \otimes f, \tag{1}$$

where the elementary hard cross-section $\hat{\sigma}$ is convoluted with f the PDF. The hard cross-section is computed in QCD and it depends just on the physical process, meanwhile PDFs cannot be computed using perturbative QCD because of the confinement property of QCD. PDFs carry the probability that a nucleon contains a parton with a certain momentum fraction, this information is process independent and thus are extracted from experimental data.

Motivated by the need for greater precision phenomenology at the LHC, the inclusion of electroweak corrections, in particular QED, to hadron collider processes is essential. From the technological point of view, this goal requires the development of computational tools which include such corrections in the hard cross-section calculations [5–7] and on the other hand, it also requires a precise determination of sets of PDFs with QED corrections.

In these last three years I have been working on topics which cover an extensive and detailed set of arguments from collider phenomenology with emphasis on PDFs. In this thesis I focus the discussion on the inclusion of QED corrections to PDFs, taking the opportunity to present several studies performed during my PhD.

We start from the theoretical aspects of this implementation, such as the upgrade of parton evolution equations, the inclusion of photon-related contributions in the computation of deep inelastic scattering (DIS) and Drell-Yan processes. This discussion is then followed by a technical description of the framework required for the determination of a set of PDFs with QED corrections.

Here we include QED corrections up to leading order (LO) in $\mathcal{O}(\alpha)$, to next-to-leading (NLO, *i.e.* $\mathcal{O}(\alpha_s^2)$) and next-to-next-to-leading (NNLO) order QCD computations. This choice can be motivated by the naïve comparison of the similar magnitude of the coupling constants $\alpha_s^2(M_Z^2)$ and $\alpha(M_Z^2)$, which suggests that LO QED corrections and NLO QCD corrections are of a similar size, *e.g.*

$$\frac{\mathcal{O}(\alpha_s^2)}{\mathcal{O}(\alpha)} \to \frac{\alpha_s^2(M_Z^2)}{\alpha(M_Z^2)} = \frac{0.118^2}{1/127} \sim 1.78. \tag{2}$$

and thus non-negligible QED effects are expected when computing predictions beyond the NLO QCD. This observation also suggests that measurements from the LHC contain useful information for an accurate determination of sets of PDFs with electroweak corrections.

The inclusion of QED corrections to PDFs assumes the presence of the photon particle as an additional parton of the nucleon which interacts with other partons. This assumption is translated by the definition of a photon parton distribution function which could be obtained from a fit to experimental data.

In this work we determine a set of PDFs with QED corrections which includes a photon PDF and its uncertainties extracted from DIS and LHC hadronic data using the Monte Carlo approach adopted by the Neural Network PDF (NNPDF) methodology. In fact, a precise determination of the photon PDF is needed for reliable computations of high mass searches, W mass determination, WW production and for several new physics signals, such as the cross-section for Z' and W' production. It is important to highlight that the methodology for PDF determination is a complex topic subjected to studies and discussions, so in this thesis we present some of these aspects, such as the most remarkable methodological choices adopted by the most active groups of PDFs. The reader is invited to check the results presented in the works listed in (*vii*) for a complete overview of the technical aspects of such issues.

Finally, the phenomenological implications of this set of QED corrected PDFs and the impact of the photon PDF are presented at the end of the discussion. We conclude with a short summary of potential improvements which are required from the point of view of new data measurements and theory developments which are expected to improve the accuracy of this set of PDFs.

We organize the discussion of this thesis following the scheme presented in Figure A:

Chapter 1: Parton distribution functions. We review the theoretical formalism of the parton model, providing a brief description of DIS and Drell-Yan processes, together with the DGLAP formulation. These concepts are then used in the determination of a set of PDFs with QED corrections. The formalism is then

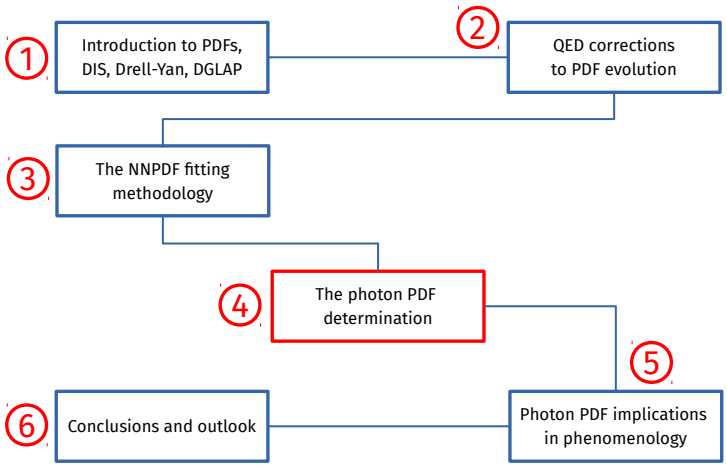

Figure A: Schematic summary of the topics discussed in this thesis.

followed by an overview of the general features of modern PDF determination, following the layout of the benchmarking exercise with LHC data performed in Ref. [8].

Chapter 2: QED corrections to PDF evolution. The combined QCD⊗QED DGLAP evolution equations are presented together with the numerical implementation in `APFEL`, a PDF evolution library [9]. We describe the features and the upgrades that `APFEL` received since its initial publication, such as the *combined* and *unified* evolution solutions. We validate the results by comparing the `APFEL` evolution with other public codes. Finally, we present `APFEL Web` [10], a web-based application for the graphical visualization of parton distribution functions that regroups in a centralized system tools for the manipulation of PDFs.

Chapter 3: The NNPDF methodology. We review the methodology used for the determination of PDFs from a global fit to experimental data. The discussion starts from the presentation of the NNPDF methodology [11, 12] which is complemented by a description of the new code structure in `C++`. This code was developed in order to improve the performance and simplify the determination of modern sets of PDFs with LHC data. We conclude this chapter with the description of the NNPDF2.3 set of PDFs, which was employed as the baseline technology for the determination of the set of PDFs with QED corrections.

Chapter 4: The photon PDF determination. We present the details of the first determination of a set of PDFs with QED corrections and the respective photon PDF based on the NNPDF methodology: the so-called NNPDF2.3QED

set. The results presented in this chapter are partially based on the published Refs. [13–15]. The photon PDF is first extracted from a fit to DIS data and then, consequently, reweighted by LHC γ^*/Z high mass and low mass measurements and W, Z rapidity distributions. We show PDF comparison plots for the photon PDF and we measure the impact of QED corrections to sets of PDFs without QED corrections.

Chapter 5: Phenomenological implications of the photon PDF. We investigate the impact of the NNPDF2.3QED set of PDFs, with emphasis on the photon PDF, looking at several observables, such as direct photon production at HERA, searches for new massive electroweak gauge boson, W pair production at the LHC presented and high and low mass Drell-Yan in Ref. [13, 16]. We conclude the discussion with a preliminary guess for the lepton PDFs obtained through the `APFEL` evolution.

Chapter 6. Conclusions and outlook. We conclude with a summary of the most relevant results presented in this thesis. Furthermore, we provide an outlook about future technical developments, in terms of experimental data and theory developments required to constraint and reduce the uncertainties of the photon PDF and improve the accuracy of sets of PDFs with QED corrections.

Chapter 1

Parton distribution functions

In the first part of this chapter we review the basic concepts of deep-inelastic scattering (DIS) process and the definition of parton distribution functions. Then we present the Drell-Yan process in hadron collisions [17, 18] and the DGLAP evolution equations which are essential in PDF determination and particularly important when including QED corrections to PDFs.

On the second half of this chapter we discuss about the general features of modern parton distributions, presenting the current state of the art in PDF determination through the results of the benchmarking exercise of Ref. [8] performed between the most active PDF groups.

1.1 Deep-inelastic scattering

Deep-inelastic scattering is a fundamental process which have been used for testing the validity of perturbative QCD. This process played an important role in the historical development of the theory but it still has a relevant role in PDF determination from experimental data, for example measurements at HERA (H1 and ZEUS [19, 20]), SLAC [21] and BCDMS [22].

1.1.1 DIS kinematics and the parton model

We consider the scattering of a charged lepton $l(k^\mu)$, with four-momenta k^μ, off a hadron target $h(p^\mu)$, such as

$$l(k^\mu) + h(p^\mu) \to l'(k'^\mu) + X, \tag{1.1}$$

where $l'(k')$ is the scattered lepton and X is the hadronic final state, see Figure 1.1 for a graphical representation of this process. We define the space-like lepton momentum transfer $q = k - k'$ in terms of differences between the incoming and outgoing leptons four-momenta. Then, the standard variables used in DIS are

$$\begin{aligned} Q^2 &= -q^2, \\ p^2 &= M^2, \\ W^2 &= (p+q)^2, \\ s &= (p+k)^2, \end{aligned} \tag{1.2}$$

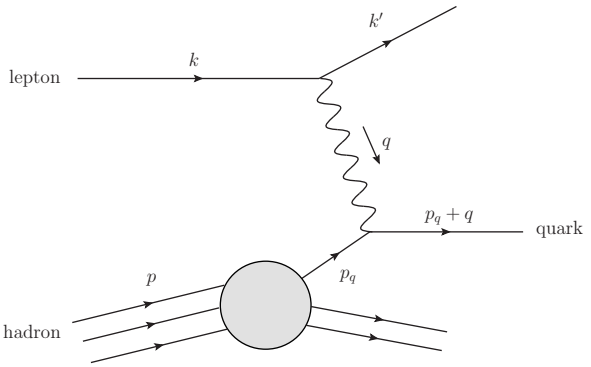

Figure 1.1: Example of deep-inelastic scattering in QCD.

where Q^2 is the virtuality of electroweak vector boson exchange, M the hadron mass, W the invariant mass of the hadronic final state and s the square of the lepton-hadron center of mass energy.

If we consider the picture of a hadron composed by pointlike massless partons, then it is natural to introduce the *longitudinal momentum fraction* ξ, where $0 \leq \xi \leq 1$, of the hadron's total momentum p. This suggests that for a given hadron composed by n_f partons, there exists a probability distribution function $f_i(\xi)$ which translates the probability that the hadron contains a parton i carrying a longitudinal fraction ξ. This concept is the basis of the parton model which was proposed by Feynman in 1969 [23] even before the formulation of QCD.

In this framework, we note $p_q = \xi p$ the momentum of the scattered parton. Satisfying the mass-shell constraint for the outgoing parton, we define the so-called Bjorken variable x as the momentum fraction ξ of a parton inside the hadron

$$m_q^2 = (p_q + q)^2 \simeq 2\xi p \cdot q - Q^2 = \frac{Q^2}{x}\xi - Q^2, \quad \xi = \left(1 + \frac{m_q^2}{Q^2}\right) x \simeq x. \tag{1.3}$$

This picture, also noted as the Bjorken limit, defined when $Q^2, p \cdot q \to \infty$ with x fixed, probes the structure of the incoming hadron at short distances. The Bjorken variable x and the energy fraction transferred by the scattered lepton are defined as

$$x = \frac{Q^2}{2p \cdot q}, \quad y = \frac{p \cdot q}{p \cdot k} = \frac{Q^2}{x \cdot s}, \tag{1.4}$$

where $0 \leq x \leq 1$ and if $x = 1$ the scattering is totally elastic.

As we have already mentioned, the determination of parton distribution functions from DIS data is possible through a fitting procedure. Since we have introduced the parton model concept for PDFs, if we consider the proton as the target hadron, there are the so-called sum rules which implies some constraints to PDF fits. The proton

consists of three valence quarks uud, this yields to the following rules

$$\int_0^1 dx\,[f_u(x) - f_{\bar{u}}(x)] = 2, \quad \int_0^1 dx\,[f_d(x) - f_{\bar{d}}(x)] = 1, \tag{1.5}$$

$$\text{MSR} = \int_0^1 dx\,x\left(\sum_{i=q}[f_i(x) + f_{\bar{i}}(x)] + f_g(x)\right) = 1. \tag{1.6}$$

Equations 1.5 and 1.6 are respectively known as the valence and momentum sum rules. It is important to highlight that thanks to the isospin symmetry in QCD, the proton PDFs are also expressed in terms of neutron PDFs: $f_u^n = f_d$, $f_d^n = f_u$ and $f_{\bar{u}}^n = f_{\bar{d}}$. The validity of these equalities is limited to the framework of pure QCD processes. Indeed, when considering QED corrections to QCD the isospin symmetry breaking introduces the electric charge of quarks and hence such simplification is no more possible.

1.1.2 DIS in perturbative QCD

Given the basic idea behind the naïve parton model, it is possible to formulate the DIS process through the quantum field theory formalism. For simplicity let us consider the neutral current electron-proton scattering process with a virtual photon γ^* exchange. At the first order in perturbation theory the matrix element of this process is

$$T = \frac{ie^2}{q^2}[\bar{u}(k')\gamma^\mu u(k)]\langle X|j_\mu(0)|p\rangle. \tag{1.7}$$

where j_μ is the electromagnetic current. From the last expression we observe that the amplitude squared is factored into the leptonic and hadronic tensors

$$|T|^2 \propto L_{\mu\nu}W^{\mu\nu}. \tag{1.8}$$

The leptonic tensor expression is trivially extracted from a simple QED computation, meanwhile the hadronic tensor cannot be completely determined, *i.e.*

$$L_{\mu\nu} = e^2\text{Tr}[\not{k}'\gamma_\mu\not{k}\gamma_\nu] = 4e^2(k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu}k\cdot k'), \tag{1.9}$$

$$W_{\mu\nu} = \frac{1}{4\pi}\int d^4x\, e^{iq\cdot x}\langle p|\,[j_\mu(x), j_\nu(0)]\,|p\rangle. \tag{1.10}$$

However, requiring the current conservation, $q\cdot W = 0$, one may parametrize the hadronic tensor in terms of two real scalar structure functions F_1 and F_2

$$W^{\mu\nu} = \left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right)F_1(x,\,Q^2) + \left(p^\mu + \frac{q^\mu}{2x}\right)\left(p^\nu + \frac{q^\nu}{2x}\right)\frac{1}{p\cdot q}F_2(x,\,Q^2). \tag{1.11}$$

Both functions, $F_1(x,Q^2)$ and $F_2(x,Q^2)$, parametrizes the structure of the target hadron in terms of x, Q^2 which are correlated to p, q. If we compute explicitly the leptonic tensor at leading order (LO), negleting the proton mass, one obtains

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[[1+(1-y)^2]F_1(x,Q^2) + \frac{(1-y)}{x}F_L(x,Q^2)\right] \tag{1.12}$$

where $F_L = F_2 - 2xF_1$ is the *longitudinal structure function*. Furthermore, $F_L = 0 \Leftrightarrow F_2 = 2xF_1$ is the Callan-Gross relation, a consequence of quarks having spin 1/2. In this equation we clearly see that the dynamics of strong interactions are represented by the structure functions of the incoming hadron.

When considering the parton model with no QCD corrections the structure functions are simply given by

$$F_2(x,Q^2) = 2xF_1(x,Q^2) = \sum_{q,\bar{q}} \int_0^1 \frac{dy}{y} f_q(y) x e_q^2 \delta\left(1 - \frac{x}{y}\right) = \sum_{q,\bar{q}} e_q^2 x f_q(x). \tag{1.13}$$

This result anticipates the factorization theorem [17] which generalizes Eq. (1.13) to all order in QCD. The factorization theorem states that any structure function F is factorized by weighting the parton structure functions with PDFs

$$\begin{aligned} F(x,Q^2) &= \sum_{i=q,\bar{q},g} \int_0^1 dy \int_0^1 dz\, C_i(z,Q^2)\, f_i(y)\, \delta(x-yz) \\ &= \sum_{i=q,\bar{q},g} \int_x^1 \frac{dy}{y} C_i\left(\frac{x}{y}, Q^2\right) f_i(y) \\ &= \sum_{i=q,\bar{q},g} C_i(x,Q^2) \otimes f_i(x), \end{aligned} \tag{1.14}$$

where the $C_i(x,Q^2)$ are the so-called coefficient functions or Wilson coefficients. We have also introduced the Mellin convolution product $\otimes$ which is defined as

$$f(x) \otimes g(x) \equiv \int_0^1 dy \int_0^1 dz f(y) g(z) \delta(x-yz) = \int_x^1 \frac{dy}{y} f\left(\frac{x}{y}\right) g(y). \tag{1.15}$$

In Eq. (1.14) the Wilson coefficients carries the information from high-energy contributions and so their exact formulation is process depend and it is calculable in perturbation theory. On the other hand the functions f_i, the PDFs, enclose the low-energy contributions and thus are non-perturbative and universal quantities which characterizes the intrinsic components of the hadron.

The calculation of the coefficient functions beyond the LO shows an *ultraviolet* (UV) and *infrared* (IR) divergences. The complete calculation of such divergences which are fully documented in Refs. [17, 18] is beyond the scope of this short review, however we summarize in the next paragraphs the most important results.

The UV divergences arising from the loop contribution are typically treated using the dimensional regularization and renormalization techniques. Concerning the IR divergences, we observe the cancellation of soft and final state collinear singularities thanks to the completely inclusive final state, which is IR safe.

In order to provide an example of the removal of the uncancelled initial state collinear singularities, lets consider the $\gamma^* q \to gq$ processes. In Figure 1.2 we present the diagrams which contribute to the lowest-order corrections to the partonic cross-section $\mathcal{O}(\alpha_s)$. The usual procedure consists in introducing the infrared cutoff μ^2 which can be chosen arbitrarily small and a bare distribution $f_q^{(0)}$ of a quark in a proton, thus we define $f_q(x,\mu_F^2)$ as the renormalized and measurable distribution

$$f_q(x,\mu_F^2) = f_q^{(0)}(x) + \frac{\alpha_s}{2\pi} \int_x^1 \frac{d\xi}{\xi} f_q^{(0)}(\xi) \left[P_{qq}\left(\frac{x}{\xi}\right) \ln \frac{Q^2}{\mu^2} + \kappa\left(\frac{x}{\xi}\right) \right] + \cdots, \tag{1.16}$$

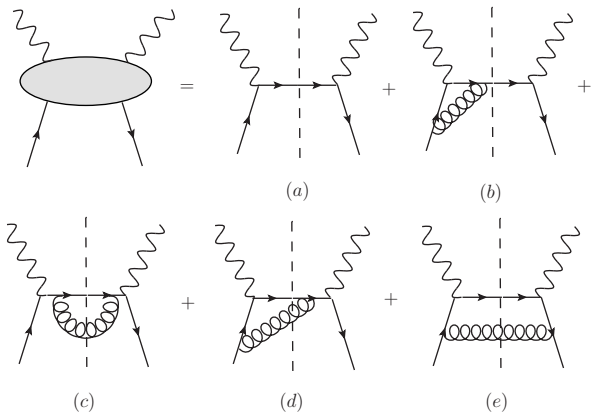

Figure 1.2: Diagrams contributing to the $\mathcal{O}(\alpha_s)$ QCD corrections with initial state quarks and anti-quarks.

where $\mu_F^2 \geq \mu^2$ is the mass factorization scale at which the quark distribution is measured, $\kappa(x)$ is a calculable function and $P_{qq}(x)$ is known as the $q \to q$ splitting function.

Using the expression of Eq. (1.16) into the QCD corrected structure function we obtain

$$F_2(x,Q^2) = x\sum_{q,\bar{q}} e_q^2 \int_x^1 \frac{d\xi}{\xi} f_q(\xi,\mu_F^2)\left[\delta\left(1-\frac{x}{\xi}\right) + \frac{\alpha_s}{2\pi}P_{qq}\left(\frac{x}{\xi}\right)\ln\frac{Q^2}{\mu_F^2} + \cdots\right], \quad (1.17)$$

which is independent of the infrared cutoff μ^2 and when setting $\mu_F^2 = Q^2$ as usual in DIS computations, the $f_q(x,Q^2)$ can be determined from structure function data at any scale.

This discussion is easily generalized also to the initial state gluons, and to other renormalization schemes, for example the $\overline{\text{MS}}$ scheme.

1.2 Hard processes in hadron collisions

Another important result of the factorization theorem is the study of processes and observables at hadron colliders such as the LHC. The high energy collision of hadrons induce soft interactions of the constituent partons, and therefore such interactions cannot be treated with perturbative QCD, but as in DIS the lowest-order QCD prediction should accurately describe the process.

The parton model cross-section for hadron-hadron colliders is defined as

$$\sigma_{AB} = \sum_{i,j\in\{q,\bar{q},g\}} \int dx_1 dx_2 f_i(x_1) f_j(x_2) \hat{\sigma}_{ij\to X}, \quad (1.18)$$

where two partons enter into a hard collision from which a final state X emerges. In this equation, the subprocess cross-section $\hat{\sigma}$ is weighted by the PDFs extracted respectively from the beam A and target B. The formal domain of validity of this

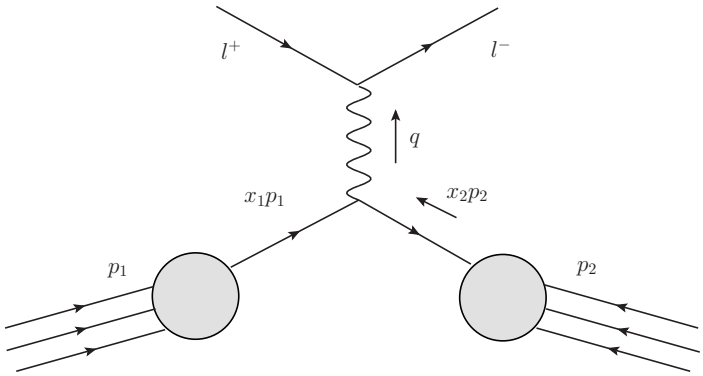

Figure 1.3: A pictorial representation of the Drell-Yan process.

definition is the asymptotic scaling limit $M^2, \hat{s} \to \infty$, with $\tau = M^2/\hat{s}$ fixed, which is the analogous of the Bjorken limit in DIS.

One of the most relevant process in hadron-hadron collision is the production of lepton pair l^+l^- with large invariant mass-squared, $M^2 = (p_{l^+} + p_{l^-})^2 \gg 1\,\text{GeV}^2$, through quark-antiquark annihilation, the so-called Drell-Yan (DY) process represented in Figure 1.3. Such process is extremely important to describe Z/γ^* and W production in high-energy collisions. It is possible to proof that the inclusion of QCD corrections to this process generates the same IR behavior observed in DIS, where PDFs have been defined as renormalized scale dependent objects as in Eq. (1.16). Thus, this is also the case for hard scattering process in hadron collisions. In this particular setup then Eq. (1.18) becomes

$$\sigma_{\text{DY}} = \sum_q \int dx_1 dx_2 f_q(x_1, M^2) f_{\bar{q}}(x_2, M^2) \hat{\sigma}_{q\bar{q}\to l^+l^-}. \tag{1.19}$$

where the PDFs are called at the M^2 scale. In this framework we identify the square of the invariance mass as

$$M^2 = x_1 x_2 \hat{s}, \tag{1.20}$$

where the variables x_1 and x_2 are defined as

$$x_1 = \frac{M}{\sqrt{\hat{s}}} e^y, \quad x_2 = \frac{M}{\sqrt{\hat{s}}} e^{-y}, \tag{1.21}$$

where y is the rapidity of the virtual photon.

Important measurements from the LHC have been performed during the last years which are relevant in PDF determination, *e.g.* the ATLAS measurements of the Z/γ^* high-mass [24] and W, Z rapidity distributions [25] and LHCb low-mass measurements [26]. In Chapter 4 data from these experiments are used in order to provide a reliable constraint on the photon PDF uncertainties.

1.3 DGLAP evolution equations

The definition of the renormalized PDFs presented in the previous sections shows the need of evolution equations which describes the variation of $f_q(x,\mu_F^2)$ with μ_F^2. By differentiating Eq. (1.16) with respect to $\ln\mu_F^2$ we obtain the renormalization group equation for the quark distribution:

$$\mu_F^2\frac{\partial}{\partial\mu_F^2}f_q(x,\mu_F^2)=\frac{\alpha_s(\mu_F^2)}{2\pi}\int_x^1\frac{d\xi}{\xi}P_{qq}\left(\frac{x}{\xi},\alpha_s(\mu_F^2)\right)f_q(\xi,\mu_F^2). \tag{1.22}$$

This is the so-called Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation. With DGLAP evolution equations we compute PDFs distributions at any given value of μ_F^2, by solving the system of integro-differential equation which requires just the initial condition of the PDFs.

The most important ingredients of DGLAP equations are the splitting functions. The splitting functions depend on the type of the parton splitting, and they have a perturbative expansion in the running coupling $\alpha_s(\mu_F^2)$. Currently, in the QCD framework splitting functions have been computed up to $\mathcal{O}(\alpha_s^3)$ [27, 28]. In Chapter 2 we discuss in detail the solution of this system of equation in the framework of combined QCD⊗QED evolution, meanwhile in the next lines we present some basic concepts about the solution of the DGLAP equations.

At leading-order the splitting functions contributions are

$$P_{qq}^{(0)}(x) = C_F\left[\frac{1+x^2}{(1-x)_+}+\frac{3}{2}\delta(1-x)\right], \tag{1.23}$$

$$P_{qg}^{(0)}(x) = T_R\left[x^2+(1-x)^2\right],\ T_R=\frac{1}{2}, \tag{1.24}$$

$$P_{gq}^{(0)}(x) = C_F\left[\frac{1+(1-x)^2}{x}\right], \tag{1.25}$$

$$P_{gg}^{(0)}(x) = 2C_A\left[\frac{x}{(1-x)_+}+\frac{1-x}{x}+x(1-x)\right] \tag{1.26}$$

$$+\left(\frac{11}{6}C_A-\frac{4}{6}n_fT_R\right)\delta(1-x), \tag{1.27}$$

where $C_F=4/3$ and $C_A=3$ are the QCD color factors, and the *plus* refers to the prescription

$$\int_0^1 dx\frac{f(x)}{(1-x)_+}=\int_0^1 dx\frac{f(x)-f(1)}{1-x}. \tag{1.28}$$

In order to solve the DGLAP evolution equations in an efficient way, we split the system of equations into two subsystems: the singlet and non-singlet sectors. Given a system with $n_f=6$ flavors, where $f_i=u,d,s,c,b,t$, we introduce a new PDF basis, known as evolution basis, by first defining

$$f_i^{\pm}=f_i\pm\bar{f}_i. \tag{1.29}$$

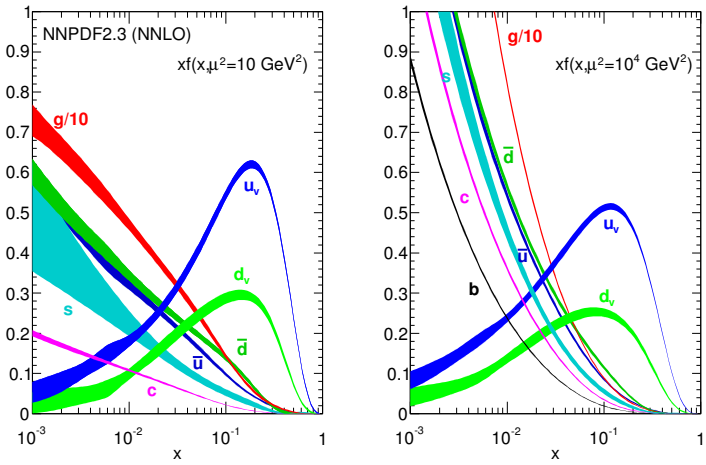

Figure 1.4: Example of PDFs evolution obtained in NNLO NNPDF2.3 global analysis [11] at scales $\mu_F^2 = 10\,\text{GeV}^2$ (left plot) and $\mu_F^2 = 10^4\,\text{GeV}^2$ (right plot), with $\alpha_s(M_Z^2) = 0.118$. This plot was produced for the PDG 2013 edition.

The non-singlet sector evolves accordingly to Eq. (1.22), and it is composed by valences and triplets

$$\text{Valences:} \quad V_i \equiv f_i^-, \qquad \text{Triplets:} \quad \begin{cases} T_3 \equiv u^+ - d^+ \\ T_8 \equiv u^+ + d^+ - 2s^+ \\ T_{15} \equiv u^+ + d^+ + s^+ - 3c^+ \\ T_{24} \equiv u^+ + d^+ + s^+ + c^+ - 4b^+ \\ T_{35} \equiv u^+ + d^+ + s^+ + c^+ + b^+ - 5t^+ \end{cases} . \tag{1.30}$$

On the other hand, the non singlet sector couples all quarks to the gluon PDF, so we define the singlet PDF as

$$\Sigma(x,\mu_F^2) = \sum_i f_i^+ \equiv \sum_i f_i(x,\mu_F^2) + f_{\bar{i}}(x,\mu_F^2). \tag{1.31}$$

Then, the coupled singlet system reads

$$\mu_F^2 \frac{\partial}{\partial \mu_F^2} \begin{pmatrix} \Sigma(x,\mu_F^2) \\ g(x,\mu_F^2) \end{pmatrix} = \frac{\alpha_s(\mu_F^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} \begin{pmatrix} P_{qq}\left(\frac{x}{\xi},\alpha_s\right) & 2n_f P_{qg}\left(\frac{x}{\xi},\alpha_s\right) \\ P_{qg}\left(\frac{x}{\xi},\alpha_s\right) & P_{gg}\left(\frac{x}{\xi},\alpha_s\right) \end{pmatrix} \begin{pmatrix} \Sigma(\xi,\mu_F^2) \\ g(\xi,\mu_F^2) \end{pmatrix}. \tag{1.32}$$

In the last paragraphs we have presented the DGLAP equations in x-space, however by looking at Eq. (1.22) we identify the Mellin convolution and so we are able to translate the same set of equations in the Mellin N-space where the DGLAP has an analytic solution. For example, for the non-singlet we obtain

$$\mu_F^2 \frac{\partial}{\partial \mu_F^2} f_{\text{NS}}(N,\mu_F^2) \quad = \quad \frac{\alpha_s(\mu_F^2)}{2\pi} \gamma_{qq}(N,\alpha_s(\mu_F^2)) f_{\text{NS}}(N,\mu_F^2), \tag{1.33}$$

where we applied the Mellin transform to the PDFs and splitting functions, defining the so-called anomalous dimension

$$f_i(N,\mu_F^2) = \int_0^1 dx\, x^{N-1} f_i(x,\mu_F^2), \quad \gamma_{ij}(N,\mu_F^2) = \int_0^1 dx\, x^{N-1} P_{ij}(x,\alpha_s(\mu_F^2)). \tag{1.34}$$

The anomalous dimension at LO are given by

$$\gamma_{qq}^{(0)}(N) = C_F\left[-\frac{1}{2} + \frac{1}{N(N+1)} - 2\sum_{k=2}^{N}\frac{1}{k}\right], \tag{1.35}$$

$$\gamma_{qg}^{(0)}(N) = T_R\left[\frac{2+N+N^2}{N(N+1)(N+2)}\right], \tag{1.36}$$

$$\gamma_{gq}^{(0)}(N) = C_F\left[\frac{2+N+N^2}{N(N^2-1)}\right], \tag{1.37}$$

$$\gamma_{gg}^{(0)}(N) = 2C_A\left[-\frac{1}{12} + \frac{1}{N^2-N} + \frac{1}{(N+1)(N+2)} - \sum_{k=2}^{N}\frac{1}{k}\right] - \frac{2n_f T_R}{3} \tag{1.38}$$

In both spaces, the solution of the DGLAP evolution is possible to derive by solving the respective integro-differential systems of equations. The N-space solution is trivial to obtain through the simple analytic solution for both sectors when using the basis presented in Eqs. 1.30 and 1.31. On the other hand, the solution in the x-space representation is highly non-trivial, so a numerical approach solution, based *e.g.* on the Runge-Kutta method is preferable. Technical details of this solution will be presented in Chapter 2.

The approach used here is easily generalized by the Wilson Operator Product Expansion (OPE) which provides a powerful computational tool for the determination of the anomalous dimensions, and it provides a more abstract determination of the DGLAP equation from the the renormalization group equations [29, 30].

Finally, in Figure 1.4 we show an example of PDF evolution in function of x, using the physical basis where $f_v = f^-$. In this case, the evolution is performed from the initial scale $\mu_0^2 = 2$ GeV^2 to $\mu_F^2 = 10$ GeV^2 (left plot) and $\mu_F^2 = 10^4$ GeV^2 (right plot). Thanks to the DGLAP evolution, PDF determination from data with different energy scales is much simpler because we are able to select a initial scale $\mu_F^2 = Q_0^2$ where the PDF is parametrized when comparing predictions to the data evolve the PDF to the experiment energy value.

A final remark concerns the notation, in the next sections and chapters the factorization scale will be noted in terms of the energy of the processes: $\mu_F^2 = Q^2$.

1.4 Characterization of modern PDFs

After introducing the origin and definition of PDFs, we conclude this chapter by showing some general features of modern PDF determination from experimental data. Nowadays this topic is studied by several groups and each group provides its own sets of PDFs. The main differences between these sets are due to the technical choices of each group, *i.e.* the experimental data included in the fit, the theoretical choices for the computation of predictions, the PDF functional form parametrization and finally,

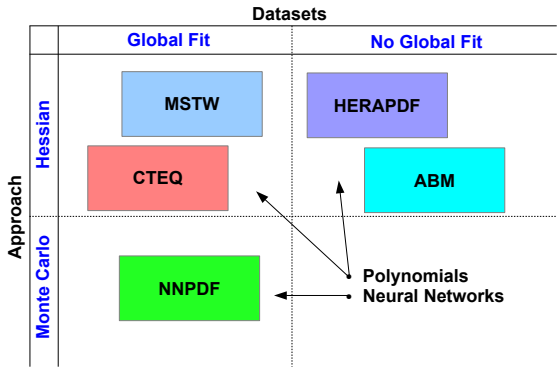

Figure 1.5: Pictorial representation of the PDF groups discussed in this section.

the fitting algorithm. In the next paragraph we present an overview of the most active groups of PDFs.

- The ABM collaboration provides sets of PDFs based on DIS and Drell-Yan data at NLO and NNLO. The ABM PDFs are parametrized by 6 independent PDFs using polynomials (25 free parameters). The minimization algorithm is based in the Hessian method, where the PDF uncertainties are given by symmetric eigenvectors. This collaboration has released ABM11 [31] which uses the combined HERA-I data, $\overline{\text{MS}}$ running heavy quark masses for DIS structure functions [32], and provides PDF sets for a range of values of α_s in a fixed flavor number scheme (FFNS) with $n_f = 5$.

- The CT collaboration extracts PDFs from a global dataset that includes DIS, Drell-Yan, W, Z production and jet data using the Hessian approach at LO, NLO and NNLO. The PDFs are also parametrized by 6 polynomials (26 free parameters) and the uncertainties are delivered through eigenvectors. The CT collaboration has released the CT10 set of PDFs [33,34] using the NNLO implementation of the S-ACOT-χ variable flavor number scheme (VFNS) for heavy quark structure functions [35].

- The HERAPDF collaboration provides PDF sets based on HERA-only DIS data at NLO and NNLO. The approach is the Hessian one, in combination with 5 polynomial independent PDFs (14 free parameters). The recent HERAPDF1.5 set [36, 37] contains the combined HERA-I dataset and the inclusive HERA-II data from H1 [38] and ZEUS [20]. This is the only set of PDFs where uncertainties are provided in terms of variations of fit parameters and experimental uncertainties.

- The MSTW collaboration releases PDF sets using a global dataset at LO, NLO and NNLO. The fit is performed by 7 independent PDFs, parametrized by polynomials (20 free parameters). The MSTW PDFs are based on the Hessian ap-

PDF set	Ref.	$\alpha_s^{(0)}$(NLO)	α_s range (NLO)	$\alpha_s^{(0)}$(NNLO)	α_s range (NNLO)
ABM11	[31]	0.1181	[0.110, 0.130]	0.1134	[0.104, 0.120]
CT10	[33]	0.118	[0.112, 0.127]	0.118	[0.112, 0.127]
HERAPDF1.5	[36, 37]	0.1176	[0.114, 0.122]	0.1176	[0.114, 0.122]
MSTW08	[39]	0.1202	[0.110, 0.130]	0.1171	[0.107, 0.127]
NNPDF2.3	[11]	all	[0.114, 0.124]	all	[0.114, 0.124]

Table 1.1: PDF sets described in this section. The table contains information about the available α_s range at NLO and NNLO for the PDF central value together with $\alpha_s^{(0)}$ for which PDF uncertainties are provided. For ABM11 the α_s varying PDF sets are only available for the $n_f = 5$ set. NNPDF always provides uncertanties for every α_s in the range.

proach. Here we use the MSTW08 PDFs [39] which was available together with the other sets considered in this section, although the MMHT2014 [40] set of PDFs has been released recently.

- The NNPDF collaboration determines PDFs at LO, NLO and NNLO from a global dataset like CT and MSTW collaborations. The NNPDF approach uses the Monte Carlo sampling method for the determination of PDF uncertainties. The parametrization consists in 7 PDFs based on artificial neural networks (ANN), for a total of 259 free parameters trained by a genetic algorithm (GA). A complete description of the NNPDF methodology is presented in Chapter 3. In this thesis we focus the discussion on the NNPDF2.3 [11] set even if this set has been recently superseded by the NNPDF3.0 [12]. The NNPDF2.3 set implements the FONLL VFNS at NNLO [41], and it also includes relevant LHC data for which the experimental correlation matrix is available.

In Figure 1.5 we show a pictorial representation of the PDF groups listed above. In Table 1.1 we summarize the PDF sets that will be compared with the common value of $\alpha_s(M_Z^2) = 0.118$. We will show results for PDFs, parton luminosities and physical cross-sections. We do not include in this comparison the JR09 PDF set [42] because it is available only for a single value of $\alpha_s(M_Z^2)$.

All the above groups provide versions of the respective PDF sets both at NLO and at NNLO, however here we will show only the NNLO PDFs. Results at NLO and for a wider range of α_s values is available from an online catalog of plots at `HepForge`:

`http://nnpdf.hepforge.org/html/pdfbench/catalog`.

1.4.1 Parton distributions and parton luminosities

In this section we compare the PDFs of the groups presented in Section 1.4 and then parton luminosities at NNLO for $\alpha_s = 0.118$. Some of the sets provide PDF errors exclusively for some default value of α_s. For those sets we take the central replica for the PDFs values at $\alpha_s = 0.118$ but we use the uncertainties of the PDF set at the default value of α_s.

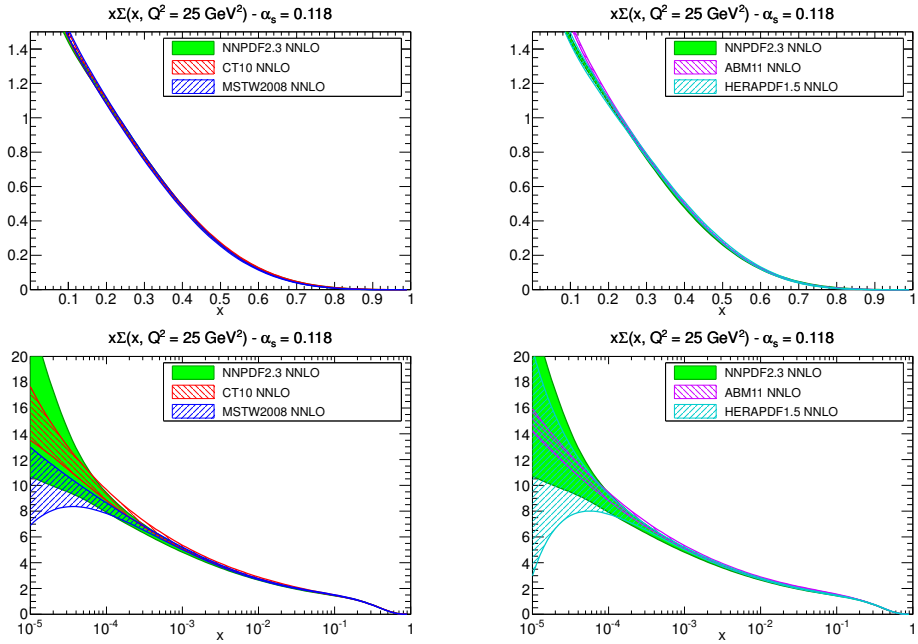

Figure 1.6: The singlet PDFs comparison at $Q^2 = 25$ GeV^2 between the NNLO PDF sets with $\alpha_s = 0.118$, on a linear scale (upper plots) and on a logarithmic scale (lower plots). The plots on the left show the comparison between NNPDF2.3, CT10 and MSTW08, while the plots on the right compare NNPDF2.3, HERAPDF1.5 and ABM11.

PDF comparison

We compare PDFs at $Q^2 = 25$ GeV^2, which is above the b-quark threshold knowing that the ABM11 set provides multiple values of α_s only when $n_f = 5$. The comparisons are organized in the following:

- For each PDFs flavor and combination we compare two sub-groups of sets: NNPDF2.3, CT10 and MSTW08 and then NNPDF2.3, ABM11 and HERAPDF1.5. The first sub-group considers sets determined from fits to a global dataset, meanwhile in the second group we still use NNPDF2.3 as reference which is compared to PDFs obtained from reduced datasets.

- In all plots, PDF uncertainties do not contain the α_s uncertainty, except for the ABM11 PDFs, where the α_s uncertainty is treated on a equal footing to the PDF parameters in the covariance matrix. The ABM11 and HERAPDF results also include an uncertainty on quark masses, while other groups provide sets with a variety of masses.

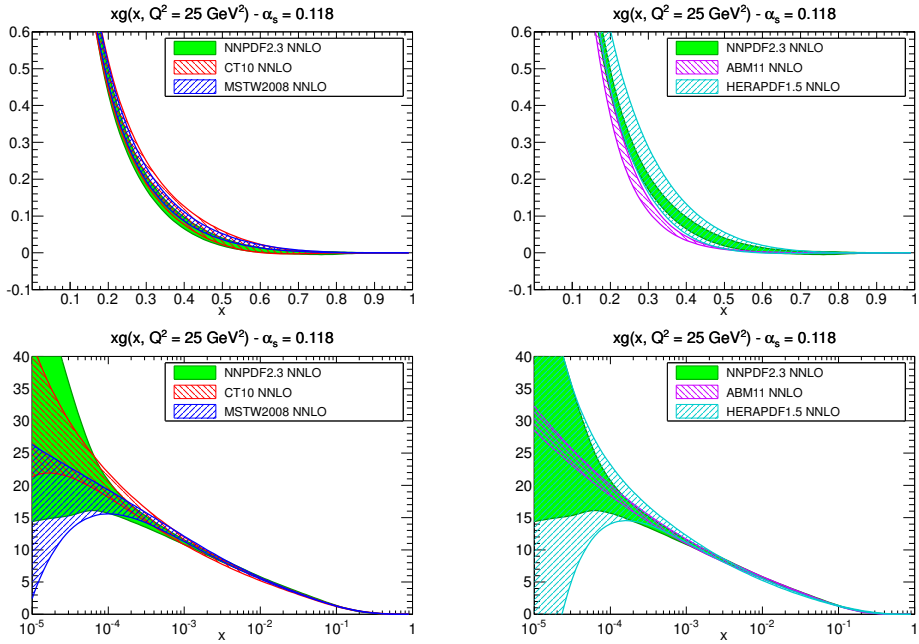

Figure 1.7: Same as Figure 1.6, but for the gluon PDF.

In Figure 1.6 we show the singlet PDF, as defined in Eq. (1.31), both on linear (upper plots) and on logarithmic (lower plots) scales, while in Figure 1.7 we show the equivalent comparison for the gluon PDFs.

The agreement is good between all the sets for the singlet, though the uncertainty band at small x is rather wider for NNPDF and HERAPDF. There is also reasonable agreement for the gluon between CT10, MSTW and NNPDF sets, where the PDF 1-σ uncertainty bands overlap for all the range of x. Differences are larger for ABM11, in particular, at small x the ABM11 gluon has smaller uncertainties than other groups, even for x values where there is little constraint from the data, due perhaps to the more restrictive underlying PDF parametrization. The ABM11 gluon at high x is smaller than that of CT, MSTW and NNPDF, meanwhile the uncertainty band overlaps that of HERAPDF in most places. The HERAPDF1.5 gluon at large x has larger uncertainties due to the lack of collider data, while at small x it is close to the other PDF sets as expected, since in this region it is only the precise HERA-I data that provides constraints to the gluon.

In Figure 1.8 we show the total strangeness $s^+(x, Q^2)$, see Eq. (1.29), on a logarithmic scale. NNPDF2.3, MSTW08, ABM11 agree at the 1-σ level, however CT10 is slightly higher, which is justified by the different treatments of heavy-quark mass effects near threshold in charged current structure functions and implementation of NuTeV data. The CT10, MSTW, and NNPDF groups use a general-mass variable flavor number (GM-VFNS) scheme, which in the case of MSTW and NNPDF turns out to be close to the fixed-flavor number scheme (FFNS) in neutrino charm production

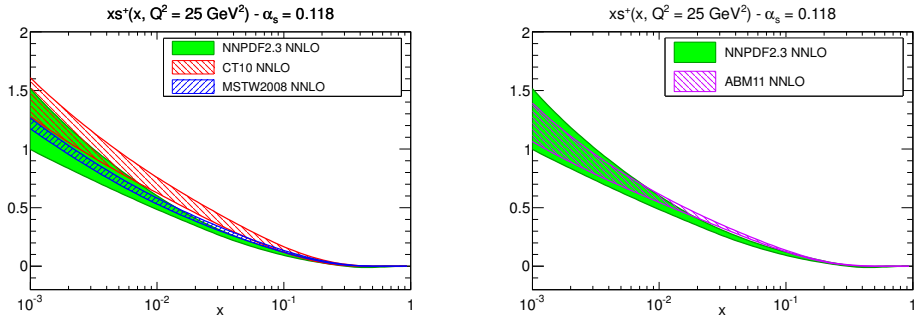

Figure 1.8: Comparison of the total strange PDFs at $Q^2 = 25$ GeV^2 between different NNLO PDF sets on a logarithmic scale. On the left plot we show the comparison between NNPDF2.3, CT10 and MSTW08, while in the plot on the right we compare only NNPDF2.3 and ABM11. HERAPDF1.5 is not show because it does not have an independent parametrization of strangeness.

in the region relevant to data [43, 44]. The ABM11 uses FFNS for neutrino charm production, while HERAPDF1.5 does not use the dimuon data and fixes strangeness to be a fraction of the total quark sea.

Studies from the ATLAS collaboration have shown that the inclusive W, Z production with the 36 pb^{-1} data prefers a larger strange PDF [45] with large uncertainties than the one typically extracted from the neutrino dimuon data. In the NNPDF2.3 analysis [11] this behavior is confirmed, ATLAS data prefers a larger strangeness, but the uncertainties are still sizable so the global fit still prefers the softer strange PDF favored by the NuTeV dimuon data. This issue should be clarified in future when including more data from the LHC, from more inclusive electroweak vector boson production data and the exclusive $W + c$ data.

We conclude this comparison analysis section with other flavor combinations:

- The non-singlet distributions T_3 and V PDFs, defined in Eq. (1.30) in Figure 1.9.

- The quark sea asymmetry $\Delta_S = \bar{d} - \bar{u}$ and the strangeness asymmetry $s^- = s - \bar{s}$ in Figure 1.10.

We observe a reasonable agreement for T_3 and V, except for ABM11, where T_3 is higher at large x due to a larger u distribution. The HERAPDF1.5 PDF uncertainties in T_3 are rather larger, reflecting the fact that HERA data does not provide much information on quark flavor separation. Concerning the quark sea asymmetry all sets are in a agreement apart from the HERAPDF1.5, which does not include the Drell-Yan and electroweak boson production data and cannot separate $\bar{u}$ and $\bar{d}$ flavors. Finally, the only sets that provide an independent parametrization of the strangeness asymmetry PDF are MSTW08 and NNPDF2.3, showing a reasonable agreement within uncertainties.

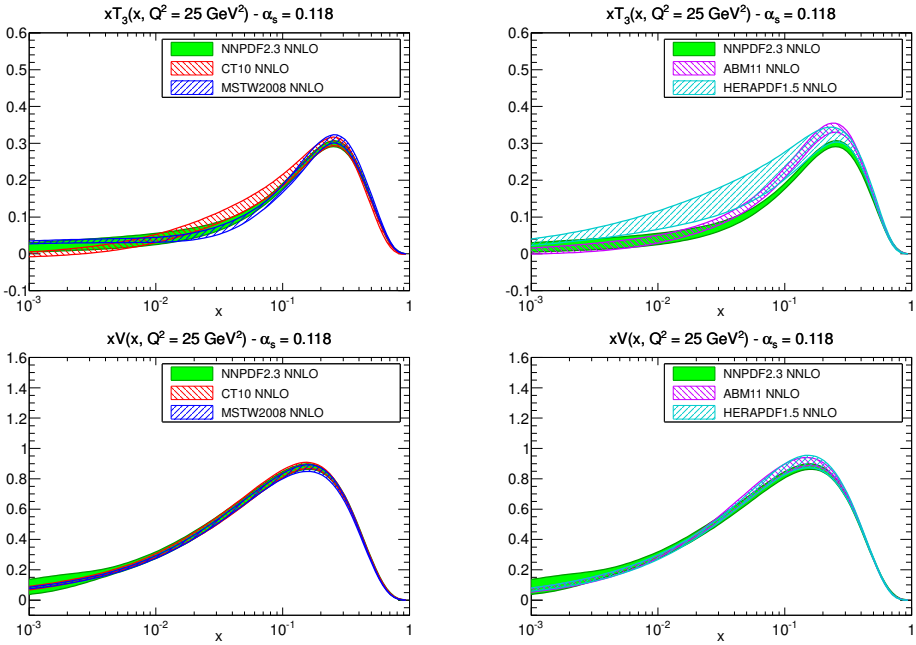

Figure 1.9: Same as Figure 1.6 for the non-singlet triplet $T_3(x)$ and the total valence $V(x)$ PDFs.

PDF luminosities

At a hadron collider the factorized observables for production of a heavy final state with mass M_X depend on parton distributions through a parton luminosity, which, following our introduction in Sect. 1.2 and Ref. [46], is defined as

$$\Phi_{ij}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} f_i\left(x_1, M_X^2\right) f_j\left(\tau/x_1, M_X^2\right) \ , \tag{1.39}$$

where $f_i(x, M^2)$ is a PDF at a scale M^2, and $\tau \equiv M_X^2/s$. Following the criteria applied to the PDF comparison, also here all parton luminosities are compared at $\alpha_s = 0.118$. The NNPDF2.3 set is used as reference for the parton luminosities ratios, and we assume a center-of-mass energy of $\sqrt{s} = 8$ TeV which is close to the energy achieved by the LHC Run-I collisions.

The gluon-gluon and quark-gluon luminosities are shown in Figure 1.11, and the quark-quark and quark-antiquark luminosities are shown in Figure 1.12. A reasonably good agreement is observed between the NNPDF2.3, MSTW08 and CT10 PDF sets for the full range of invariant masses. The PDF uncertainties increase dramatically at $M_X > 1$ TeV. Future data from the LHC such as the high-mass Drell-Yan process should be able to provide constraints in this important region. For HERAPDF1.5, there is generally an agreement in central values, but the uncertainty is rather larger in some x ranges, particularly for the gluon luminosity, but also to some extent for

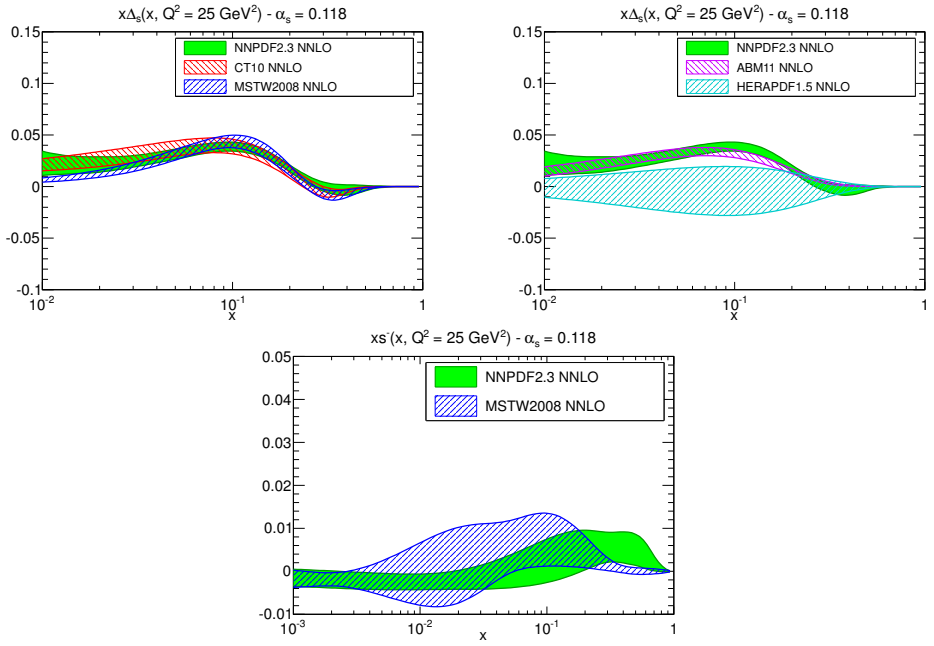

Figure 1.10: Same as Figure 1.9 for the the sea asymmetry Δ_S and the strange asymmetry s^- PDFs. In the latter case we show only the results for MSTW08 and NNPDF2.3, the only PDF sets that introduce an independent parametrization of the strangeness asymmetry.

the quark-antiquark one. For ABM11 instead, the quark-quark and quark-antiquark luminosity are systematically higher by over 5% below 1 TeV, and above this the quark-antiquark luminosity becomes much softer than either NNPDF2.3 or MSTW08. The gluon-gluon luminosity becomes smaller than all the other PDFs at high invariant masses, overlapping only with the very large HERAPDF1.5 uncertainty.

It is also useful to compare the relative PDF uncertainties in the parton luminosities. In Figure 1.13 we show this relative PDF uncertainty for the quark-antiquark and gluon-gluon luminosities. We see clearly the much larger HERAPDF1.5 uncertainty, and that at high invariant mass, the uncertainty in the ABM11 gluon-gluon luminosity becomes smaller.

The larger quark-antiquark luminosity from ABM11 as compared to the other PDF sets could be inferred from the PDF comparison plots at lower Q^2, the ABM gluon is a little larger than the central value of the other groups below about $x = 0.05$, and this drives more quark and antiquark evolution at small x values. It has been recently suggested [47] from the results of a NLO fit to DIS data only, that some of these features could receive a contribution from the different ABM treatment of heavy-quark masses (see also [48]). While CT, MSTW and NNPDF use different versions of the variable flavour number scheme [35, 41, 49], which are broadly equivalent to one another up to small subleading terms, ABM11 uses a fixed flavour number scheme

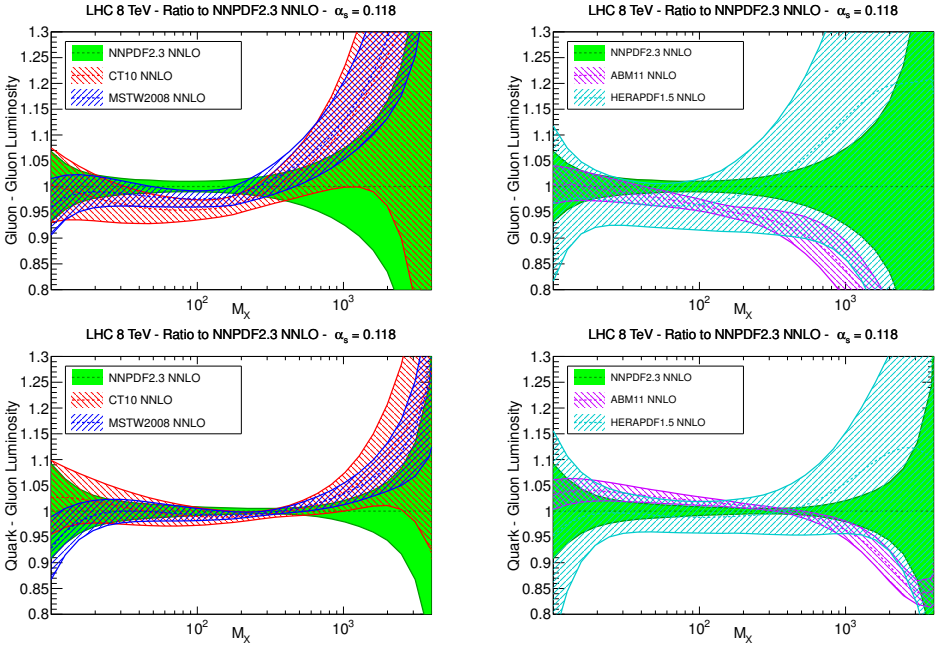

Figure 1.11: The gluon-gluon (upper plots) and quark-gluon (lower plots) luminosities, Eq. (1.39), with $\alpha_s = 0.118$, at LHC $\sqrt{s} = 8$ TeV. The NNPDF2.3 set is used as reference for both comparison groups.

	Q_0^2 (GeV^2)	$Q^2_{\rm min}$ (GeV^2)	$W^2_{\rm min}$ (GeV^2)
ABM11	9	2.5	3.24
CT10	1.69	4.0	12.25
HERAPDF1.5	1.9	3.5	-
MSTW08	1	2.0	15.0
NNPDF2.3	2.0	3.0	12.5

Table 1.2: The values of the initial evolution scale where the PDFs are parametrized, Q_0^2, and the kinematical cuts in Q^2 and $W^2 = Q^2\ (1/x - 1)$ applied to the fitted DIS dataset, $Q^2_{\rm min}$ and $W^2_{\rm min}$, in the present work and in other recent PDF determinations.

for heavy-quark PDFs. This may explain the increase in the medium-x and small-x light quarks and gluons, and the corresponding softer large-x gluon required by the momentum sum rule, found in the ABM fits [47], though more studies are required to confirm this point.

An alternative interpretation proposed to explain these differences between ABM11 and the other groups resides on the treatment of the kinematical cuts of the DIS data. These cuts control the impact of higher twists contributions. All groups undertake measures to minimize the impact of higher twists, in particular the CT10, MSTW08 and NNPDF2.3 fits suppress this contribution with a minimal cut on $W^2 =$

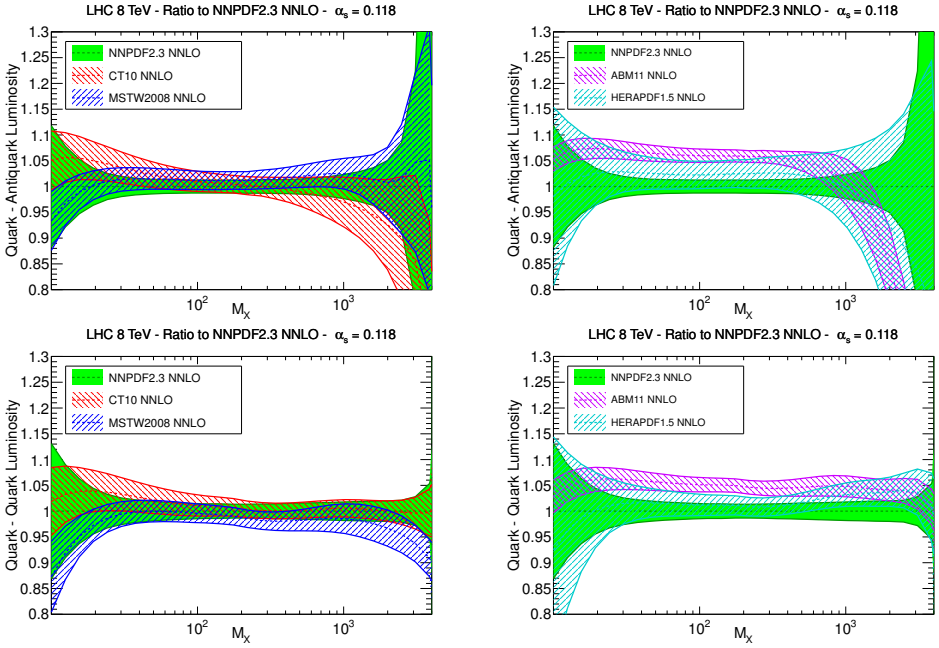

Figure 1.12: Same as Figure 1.11 for the quark-antiquark (upper plots) and quark-quark (lower plots) luminosities.

$Q^2\,(1/x-1)$.

In Table 1.2 we show a summary of the values of the initial evolution scale Q_0^2 where the PDFs are parametrized, together with the lower kinematical cuts $Q^2_{\rm min}$ and $W^2_{\rm min}$ applied to the fitted DIS data sets for each PDF group. The ABM11 fit also imposes an upper cut $Q^2_{\rm max} = 10^3$ GeV^2 on the HERA data. It is well known that, the larger the dataset, the more robust are the PDFs with respect to variations in these cuts. For instance, stability under variation of the default MSTW08 kinematical cuts was studied in Ref. [50]. The inclusion of higher twists in MRST fits has previously been shown to lead to only a small effect on high-Q^2 PDFs [51], and an ongoing extension of the study in [47] suggests this is qualitatively the same with more up-to-date PDFs. This conclusion has been confirmed in similar studies by NNPDF [52].

1.4.2 LHC inclusive cross-sections

We conclude this chapter by describing the behavior of the cross-sections predictions at 8 TeV for various benchmark processes and compare the results for all NNLO PDF sets used in the previous section. Also here we consider only PDF uncertainties, negleting a careful assessment of all relevant theoretical uncertainties into consideration for each of the studied processes.

In Figure 1.14 we show the inclusive cross-sections for electroweak gauge boson production, W^+, W^- and Z, at 8 TeV with $\alpha_s(M_Z^2) = 0.118$, meanwhile in Figure 1.15

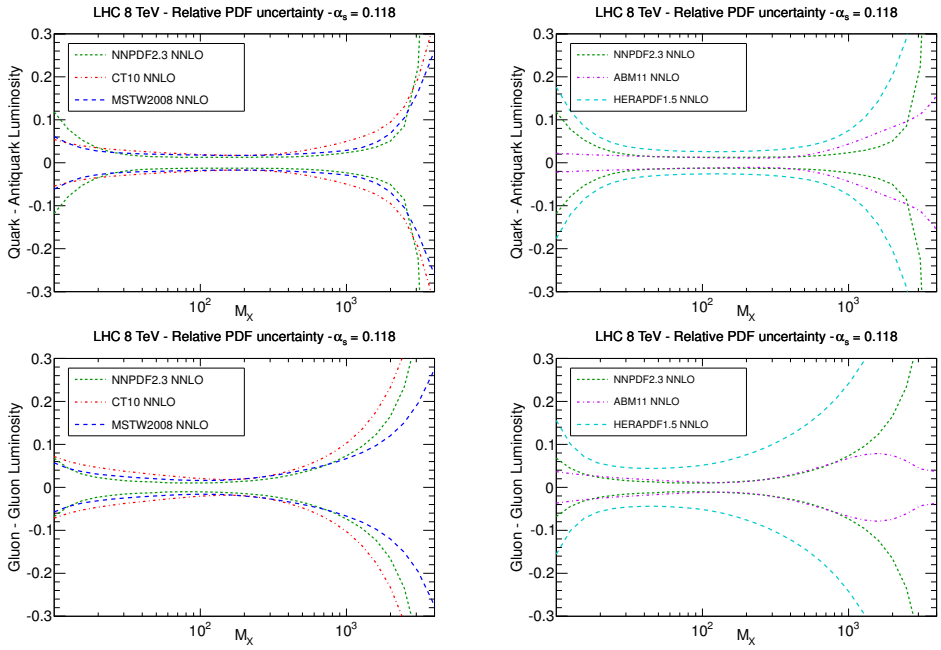

Figure 1.13: The relative PDF uncertainties in the quark-antiquark luminosity (upper plots) and in the gluon-gluon luminosity (lower plots), with $\alpha_s = 0.118$ at the LHC $\sqrt{s} = 8$ TeV.

we present the W^+/W^- and W/Z cross-section ratios. In both cases, the predictions have been computed at NNLO using the `Vrap` code [53] with the central scale choice $\mu_R = \mu_F = M_V$. The CMS measurements [54] are plotted together with the theoretical predictions showing a good agreement between NNPDF2.3, CT10, MSTW08 and HERAPDF1.5, as already observed in the quark-antiquark luminosity in Figure 1.12. The comparison with ABM11 leads to systematically higher cross-sections which is also consistent with the larger luminosities.

The Higgs boson production cross-section is another important process for LHC phenomenology. In Figure 1.16 we compare several predictions for the LHC Standard Model Higgs boson cross-section at 8 TeV between the NNLO PDF sets. The left hand plots show results for $\alpha_s(M_Z^2) = 0.117$, while on the right $\alpha_s(M_Z^2) = 0.119$. This choice is made in order to quantify the impact of the α_s variation. In all cases the same value of α_S is used consistently in both the PDFs and in the matrix element calculation and we take $m_H = 125$ GeV. The codes and setups used for the formulation of these plots are listed below:

- The Higgs boson production cross-sections in the gluon fusion channel have been computed with the `iHixs` code [55] where the central scale has been taken to be $\mu_F = \mu_R = m_H$, consistent with the Higgs cross-section working group recommendations [56].

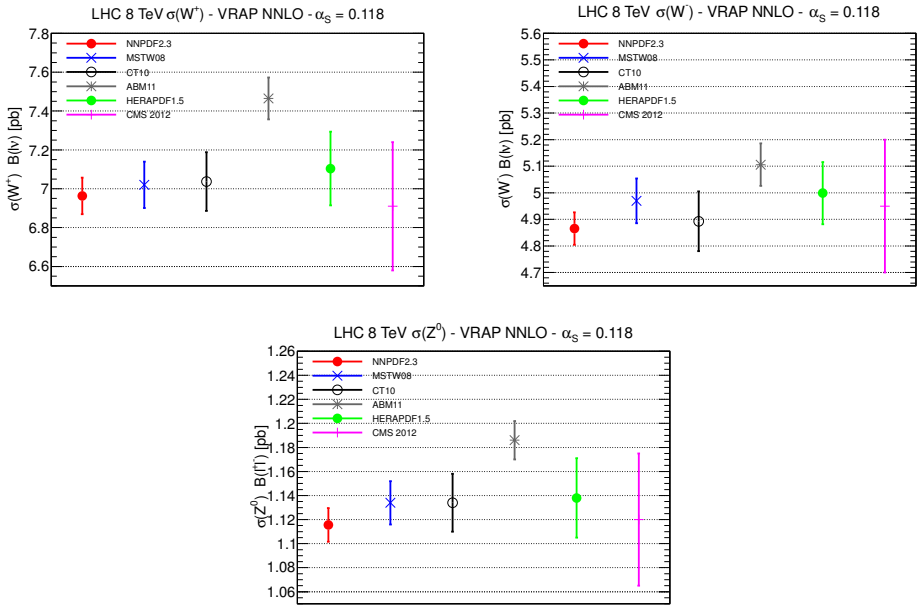

Figure 1.14: Comparison of the predictions for inclusive cross-sections for electroweak gauge boson production between different PDF sets at LHC 8 TeV. In all cases the branching ratios to leptons have already been taken into account. From top to bottom and from left to right we show the W^+, W^-, and Z inclusive cross-sections. All cross-sections are compared at a common value of $\alpha_s(M_Z) = 0.118$. We also show the recent CMS 8 TeV measurements.

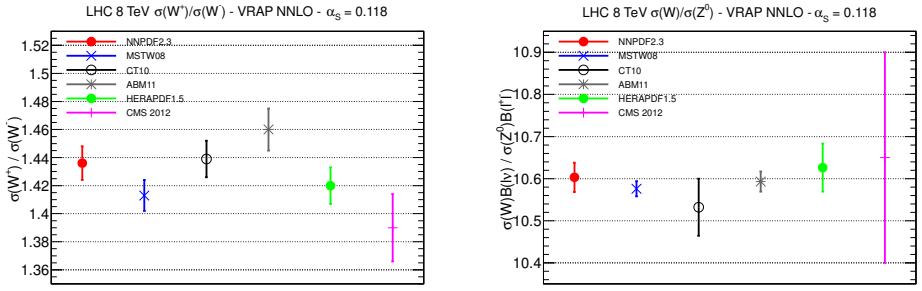

Figure 1.15: Same as Figure 1.14 for the W^+/W^- and W/Z ratios.

- The Higgs production in the Vector Boson Fusion (VBF) channel has been computed at NNLO with the `VBF@NNLO` code [57], with $\mu_F = \mu_R = m_H$.
- The Higgs production in association with W and Z bosons has been computed at NNLO with the `VH@NNLO` program [58, 59]. Also here the scale choice is $\mu_F = \mu_R = m_H$.
- The Higgs production in association with a top quark pair, $t\bar{t}H$, has been computed at LO with the `MCFM` program [60], with the scale choice $\mu_F = \mu_R = 2m_t + m_H$.

In summary we observe that the comparison between PDF sets is unaffected by changes in α_s. The variation of α_s produces shifts of the absolute value of the cross-section. ABM11 and HERAPDF1.5 for the gluon fusion fall within the envelope composed by the NNPDF2.3, CT10 and MSTW08 PDFs. However, the HERAPDF1.5 uncertainty is bigger than this envelope. For VBF, WH and $t\bar{t}H$ production, there is a reasonable agreement between CT10, MSTW08 and NNPDF2.3 both in central values and in the size of PDF uncertainties. ABM11, on the other hand, leads to rather different results, despite the fact that a common value of α_s is being used. For quark-initiated processes, like VBF and WH, the ABM11 cross-section is higher than that of the other sets, specially for WH production. For $t\bar{t}H$, which has its largest contribution from gluon-initiated diagrams, the ABM11 cross-section is smaller. The HERAPDF1.5 PDF uncertainties are distinctly larger compared to three global fits, especially for ggH and $t\bar{t}H$. This can be attributed to the poorly constrained large-x gluon in the HERA-only fits and, in the case of $t\bar{t}H$, less constrained sea quarks.

Finally, we conclude the comparisons with the inclusive top quark pair production cross-section, which has been computed at NNLO$_{\rm approx}$+NNLL with the `top++` code [61–66] as implemented in `v1.3`, which includes the complete NNLO corrections to the $q\bar{q} \to t\bar{t}$, with the central scale $\mu_F = \mu_R = m_t$. The settings of the theoretical calculations are the default ones in Ref. [67]. In all calculations we use $m_t = 173.2$ GeV.

In Figure 1.17 we show the approximate NNLO top quark pair production cross-section at 8 TeV for different NNLO PDF sets with $\alpha_s(M_Z^2) = 0.117$ and $\alpha_s(M_Z^2) = 0.119$. Also in this case theoretical predictions are compared to the recent CMS measurements [68] in terms of the average of the cross-section in the dilepton and lepton+jets final states. The $t\bar{t}$ total cross-section has some sensitivity to the value of α_s. This sensitivity has been recently used by CMS to provide the first ever determination of α_s from top cross-sections [69]. For the $t\bar{t}$ cross-section, we see a reasonable agreement between NNPDF2.3, CT10 and MSTW08, while ABM11 is somewhat lower. The HERAPDF1.5 central value is in good agreement with the global fits but, as usual, the PDF uncertainties are larger.

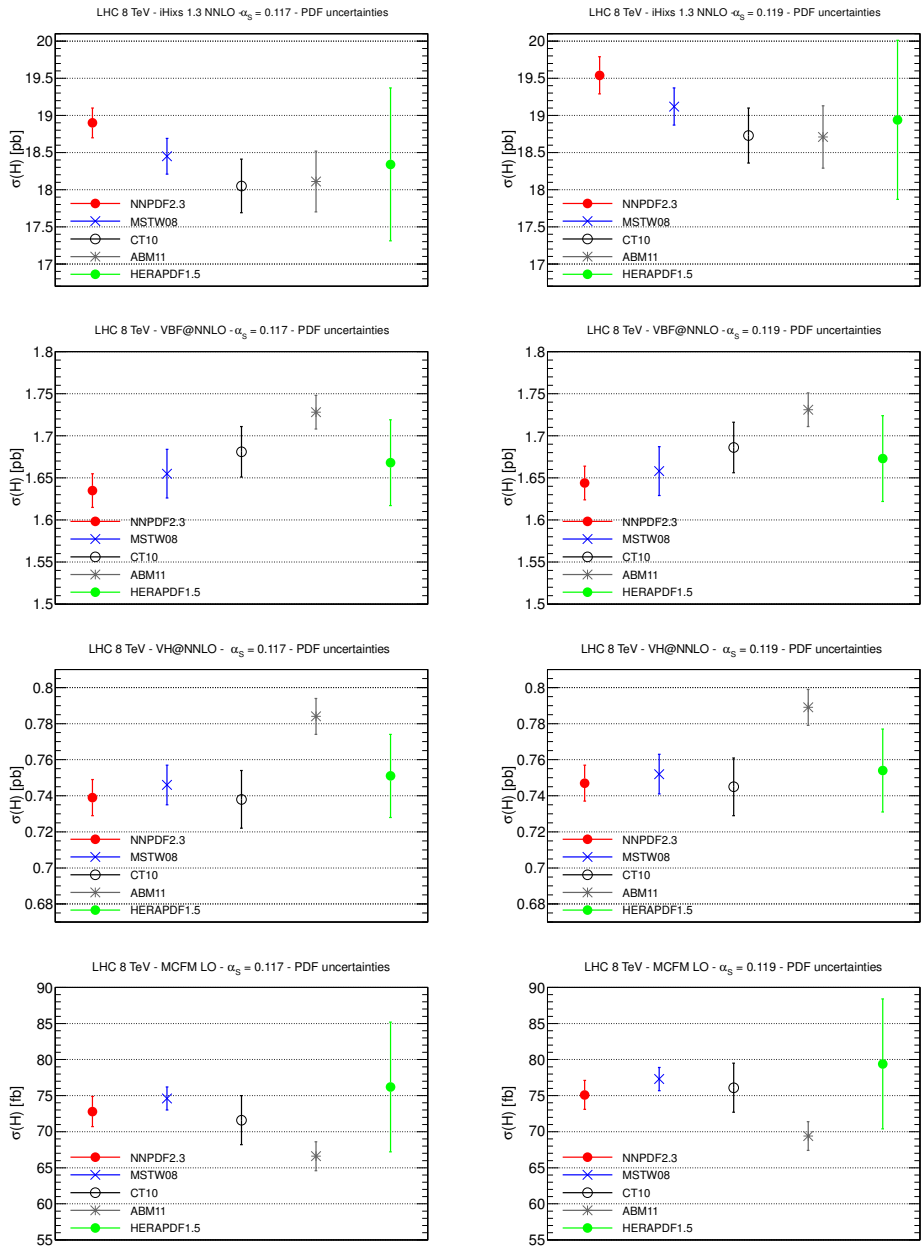

Figure 1.16: Comparison of the predictions for the LHC Standard Model Higgs boson cross-sections at 8 TeV between various NNLO PDF sets. From top to bottom we show gluon fusion, vector boson fusion (VBF), associated production (with W), and associated production with a $t\bar{t}$ pair. The left hand plots show results for $\alpha_S(M_Z) = 0.117$, while on the right we have $\alpha_S(M_Z) = 0.119$.

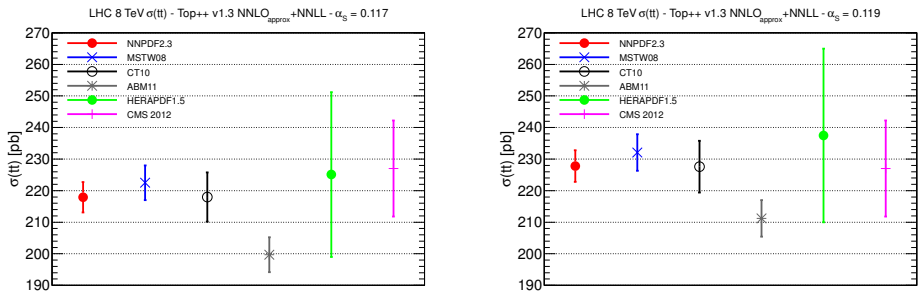

Figure 1.17: Comparison of the predictions for the top quark pair production at LHC 8 TeV between various NNLO PDF sets. Left plot: results for $\alpha_S(M_Z) = 0.117$. Right plot: results for $\alpha_S(M_Z) = 0.119$. In both cases we also show the CMS 8 TeV measurement.

Chapter 2

QED corrections to PDF evolution

In this chapter we introduce the theoretical framework and the numerical implementation of QED corrections to parton evolution equations. We organize this chapter as follows: in Sect. 2.1 we motivate the inclusion of QED corrections to DGLAP and we introduce the `APFEL` library which was developed specifically for this project. Then, in Sect. 2.2, the combined DGLAP equations are presented explicitly and the solution strategy is discussed in detail. The numerical techniques used in `APFEL` are summarized in Sect. 2.3, validation and benchmarking results against other public codes are presented in Sect. 2.4. Finally, in Sect. 2.5 we conclude the discussion with the description of `APFEL Web`, a spin-off of the `APFEL` library. The DGLAP solution developed in this chapter have been applied to the photon PDF determination in Chapter 4.

2.1 Introduction to APFEL

Following the discussion started in the introduction of this thesis, we recall that during the last years a great effort has been made for the achievement of PDFs determined using NLO and NNLO QCD theory [8, 70, 71]. However, at present, the level of accuracy in theoretical predictions and experimental uncertainties is such that QED and electroweak (EW) corrections are required for the precision physics at the LHC.

There are several examples of predictions for hadron collider processes with QED and EW corrections, which have been computed in the last years. A full review of such processes is presented in Ref. [72], from which we can mention:

- the inclusive W and Z production [5, 73–82]
- the W and Z boson production in association with jets [83–85], diboson production [86–88], dijet production [89, 90] and top quark pair production [91–95]

The combination of QCD and QED calculations at hadron colliders requires PDFs with QCD⊗QED DGLAP evolution equations [96–98]. Many studies have been performed during the last 20 years about the numerical solution and optimization of the QCD DGLAP evolution equations, many of which have become public tools [99–106]. On the other hand, much less effort has been invested to the solutions of QCD⊗QED DGLAP [107–109], in particular, to the best of our knowledge, before the release of

this work, the only public codes which offered the possibility to obtain an estimation of such corrections were

- the `partonevolution` [108, 110] library, which is limited to NLO QCD corrections and it does not contain a modern interface the `LHAPDF` [111] library (used for accessing all published PDF sets), and in addition it does not allow to explore different possibilities for the combination of the QCD and QED evolution equations.
- the MRST2004QED set of PDFs [109], which until the set of PDFs presented in this thesis, was the only set which included QED corrections, where the photon PDF is based on model assumptions. However, this set of PDFs delivers precomputed evolution encoded in the grid, which denies the possibility to perform systematic studies of the DGLAP equation for different initial conditions.

Therefore, in this work we present `APFEL` [9], which stands for *A Parton distribution Function Evolution Library*. `APFEL`'s goal is to fill the need of a public tool, accurate and flexible that can be used to perform PDF evolution up to NNLO in QCD and LO in QED, both in the fixed-flavor-number (FFN) and in the variable-flavor-number (VFN) schemes, and using either pole or $\overline{\text{MS}}$ heavy quark masses. `APFEL` is designed to meet the needs of PDF fits, providing large control of evolution parameters like the heavy quark thresholds, the coupling running solution, and many others.

`APFEL` is implemented in `Fortran77` with wrappers in `C++` and `Python`. It is publicly available from the `HepForge` website[1]. `APFEL` is part of the family of codes which solves the DGLAP equations using x-space methods, which typically use a representation of the PDFs on a grid in x and μ_F^2 together with higher-order interpolation techniques for the solution of the intergro-differential equations [99, 101–104].

This methodology is widely used by other pure QCD evolution libraries such as: `HOPPET` [99] and `QCDNUM` [101]. Other tools, like the best-know `PEGASUS` [105], solve the DGLAP equations in N-space, by transforming the evolution equations into Mellin space (see Sect. 1.3) which are then analytically solved and inverted back to x-space using complex-variable methods [100, 105, 106, 108, 110]. The main drawback of the N-space methods, however, is the fact that they require the analytical Mellin transform of the initial PDFs which is possible only for some very specific functional forms, which is unlikely the case for the PDF sets in `LHAPDF` which are delivered in function of the x variable. A third approach is provided by the hybrid method adopted in the `FastKernel` methodology, the internal code used in the NNPDF fits [112, 113], where DGLAP equations are solved in Mellin space and then used to determine the x-space evolution operators, which are convoluted with the x-space PDFs to perform the evolution.

2.2 DGLAP evolution with QED corrections

In this section we present the strategy that `APFEL` adopts in order to perform the DGLAP evolution of PDFs when QCD and QED effects are taken into account.

First, we present the QED evolution equations, and then we show how to define an evolution basis which solves the system. In this work we suggest two different

[1] http://apfel.hepforge.org/

approaches for the solution of the combined system: the *unified* solution and the so-called *coupled* solution where QCD and QED equations are solved separately and then combined. We show that the coupled approach provides a good approximation of the unified solution.

In next sections we limit the discussion to the QED sector, however more details about QCD corrections to DGLAP evolution equations up to NNLO are available in Refs. [114–124], and the structure of their solutions has also been discussed in great detail in the literature, see for example Refs. [70, 99, 105].

2.2.1 Solving the QED evolution equations

The implementation of the QED corrections to the DGLAP evolution equations leads to the inclusion of additional terms which contain QED splitting functions [96–98], proportional to the QED coupling α, convoluted with the PDFs. There are several possibilities to solve the combined QCD⊗QED DGLAP evolution equations, and, as opposed to previous works, `APFEL` adopts:

- the *coupled* solution: a fully factorized approach where the QCD and the QED factorization procedures can be regarded as two independent steps that lead to two independent factorization scales on which all PDFs depend.

- the *unified* solution: the procedure where the QCD and QED sectors are solved by an unique system of equations.

In the next paragraphs and in Sect. 2.2.2 we describe the coupled approach, meanwhile we devote Sect. 2.2.3 for the unified method. For simplicity, in both discussions we assume that no heavy quark threshold is crossed during the DGLAP evolution *i.e.* it is valid only when PDF evolution is performed in the FFN scheme, however the generalization to the VFN scheme is in `APFEL`, the documented in Sect. 2.3 of Ref. [9].

In the case where QED corrections are included up to $\mathcal{O}(\alpha)$ and the mixed subleading terms $\mathcal{O}(\alpha\alpha_s)$ are neglected, the QCD evolution with respect to μ and the QED evolution with respect to ν will be given by two fully decoupled equations:

$$\begin{aligned} \mu^2 \frac{\partial}{\partial \mu^2} \mathbf{q}(x,\mu,\nu) &= \mathbf{P}^{\rm QCD}(x,\alpha_s(\mu)) \otimes \mathbf{q}(x,\mu,\nu)\,, \\ \nu^2 \frac{\partial}{\partial \nu^2} \mathbf{q}(x,\mu,\nu) &= \mathbf{P}^{\rm QED}(x,\alpha(\nu)) \otimes \mathbf{q}(x,\mu,\nu)\,, \end{aligned} \tag{2.1}$$

where $\mathbf{P}^{\rm QCD}$ and $\mathbf{P}^{\rm QED}$ are respectively the QCD and QED matrices of splitting functions and $\mathbf{q}(x,\mu,\nu)$ is a vector containing all the parton distribution functions. Let us recall that in the presence of QED corrections, the photon PDF $\gamma(x,\mu,\nu)$ should also be included in $\mathbf{q}(x,\mu,\nu)$. The independent solutions of the differential equations in Eq. (2.1), irrespective of the numerical technique used, will give as a result two different evolution operators: $\mathbf{\Gamma}^{\rm QCD}$, that evolves the array $\mathbf{q}$ in μ while keeping ν constant, and $\mathbf{\Gamma}^{\rm QED}$, that evolves $\mathbf{q}$ in ν while keeping μ constant. If the QCD evolution takes place between μ_0 and μ_1 and the QED evolution between ν_0 and

ν_1, we will have that:

$$\begin{aligned}\mathbf{q}(x,\mu_1,\nu) &= \mathbf{\Gamma}^{\rm QCD}(x|\mu_1,\mu_0)\otimes\mathbf{q}(x,\mu_0,\nu)\,,\\ \mathbf{q}(x,\mu,\nu_1) &= \mathbf{\Gamma}^{\rm QED}(x|\nu_1,\nu_0)\otimes\mathbf{q}(x,\mu,\nu_0)\,.\end{aligned} \tag{2.2}$$

Once the QCD and QED evolution operators in Eq. (2.2) have been calculated, one can combine them to obtain a coupled evolution operator $\mathbf{\Gamma}^{\rm QCD\otimes QED}$ that evolves PDFs both in the QCD and in the QED scales, that is:

$$\mathbf{q}(x,\mu_1,\nu_1) = \mathbf{\Gamma}^{\rm QCD\otimes QED}(x|\mu_1,\mu_0;\nu_1,\nu_0)\otimes\mathbf{q}(x,\mu_0,\nu_0)\,. \tag{2.3}$$

Before discussing the derivation of the combined evolution operator $\mathbf{\Gamma}^{\rm QCD\otimes QED}$, we present the strategy used in `APFEL` to solve the QED DGLAP equations in Eq. (2.1). At leading order, the QED equations for the evolution of the quark and photon PDFs, dropping for simplicity the dependence on the QCD factorization scale μ, read:

$$\begin{aligned}\nu^2\frac{\partial}{\partial\nu^2}\gamma(x,\nu) &= \frac{\alpha(\nu)}{4\pi}\left[\left(\sum_i N_c e_i^2\right)P_{\gamma\gamma}^{(0)}(x)\otimes\gamma(x,\nu)+\sum_i e_i^2 P_{\gamma q}^{(0)}(x)\otimes(q_i+\bar{q}_i)(x,\nu)\right]\,,\\ \nu^2\frac{\partial}{\partial\nu^2}q_i(x,\nu) &= \frac{\alpha(\nu)}{4\pi}\left[N_c e_i^2 P_{q\gamma}^{(0)}(x)\otimes\gamma(x,\nu)+e_i^2 P_{qq}^{(0)}(x)\otimes q_i(x,\nu)\right]\,,\\ \nu^2\frac{\partial}{\partial\nu^2}\bar{q}_i(x,\nu) &= \frac{\alpha(\nu)}{4\pi}\left[N_c e_i^2 P_{q\gamma}^{(0)}(x)\otimes\gamma(x,\nu)+e_i^2 P_{qq}^{(0)}(x)\otimes\bar{q}_i(x,\nu)\right]\,,\end{aligned} \tag{2.4}$$

where $\gamma(x,\nu)$, $q_i(x,\nu)$ and $\bar{q}_i(x,\nu)$ are respectively the PDFs of the photon, the i-th quark and the i-th antiquark, e_i the quark electric charge, $N_c = 3$ the number of colors and $\alpha(\nu)$ the running fine structure constant. In this work we neglect the impact of lepton PDFs. Note that at this order the gluon PDF does not enter the QED evolution equations. The leading-order QED splitting functions $P_{ij}^{(0)}(x)$ are given by:

$$\begin{aligned}P_{q\gamma}^{(0)}(x) &= 2\left[x^2+(1-x)^2\right]\,,\\ P_{\gamma q}^{(0)}(x) &= 2\left[\frac{1+(1-x)^2}{x}\right]\,,\\ P_{\gamma\gamma}^{(0)}(x) &= -\frac{4}{3}\delta(1-x),\\ P_{qq}^{(0)}(x) &= 2\frac{1+x^2}{(1-x)_+}+3\delta(1-x)\,.\end{aligned} \tag{2.5}$$

The index i in Eq. (5.8) runs over the active quark flavors at a given scale ν.

It should be noted that, in the presence of QED effects, the usual momentum sum rule is modified to take into account the contribution coming from the photon PDF. Therefore, provided that the input PDFs respect the momentum sum rule, the QED evolution should satisfy the equality:

$$\int_0^1 dx\; x\left\{\sum_i(q_i+\bar{q}_i)(x,\mu,\nu)+g(x,\mu,\nu)+\gamma(x,\mu,\nu)\right\} = 1\,, \tag{2.6}$$

for any value of the scales μ and ν. An important test of the numerical implementation of DGLAP evolution in the presence of QED effects is to check that Eq. (2.6) indeed holds at all scales.

As in the case of QCD, an important practical issue that needs to be addressed when solving the QED DGLAP evolution equations is the choice of the PDF basis. The use of the flavor basis $\mathbf{q} = \{\gamma, u, \bar{u}, d, \bar{d}, ...\}$ requires the solution of a system of thirteen coupled equations which in turns leads to a cumbersome numerical implementation. This problem can be overcome by choosing a suitable PDF basis, the evolution basis, that maximally diagonalizes the QED splitting function matrix. Note that this optimized basis will be different from that used in QCD, due to the presence of the electric charges e_i in Eq. (5.8) that are different between up- and down-type quarks. This difference between up- and down-type quarks, in the presence of QED effects, is also responsible for the dynamical generation of isospin symmetry breaking between proton and neutron PDFs.

2.2.2 Basis for the coupled QCD⊗QED solution

For the coupled approach we adopt a PDF basis for the QED evolution which was originally suggested in Ref. [108], defined by the following singlet and non-singlet PDF combinations:

$$\text{Singlet}: \quad \mathbf{q}^{\rm SG} = \begin{pmatrix} \gamma \\ \Sigma \equiv u^+ + c^+ + t^+ + d^+ + s^+ + b^+ \\ \Delta_\Sigma \equiv u^+ + c^+ + t^+ - d^+ - s^+ - b^+ \end{pmatrix},$$

$$\text{Non-Singlet}: \quad q_i^{\rm NS} = \left\{ \begin{array}{c} \Delta_{uc} \equiv u^+ - c^+, \\ \Delta_{ds} \equiv d^+ - s^+, \\ \Delta_{sb} \equiv s^+ - b^+, \\ \Delta_{ct} \equiv c^+ - t^+, \\ u^-, \\ d^-, \\ s^-, \\ c^-, \\ b^-, \\ t^- \end{array} \right\}, \quad i = 1, \ldots, 10\,, \tag{2.7}$$

where $q^\pm \equiv q \pm \bar{q}$. Similarly to the QCD notation introduced in Sect. 1.3, the singlet distributions are those that couple to the photon PDF $\gamma(x, \nu)$, while the non-singlet distributions evolve multiplicatively and do not couple to the photon.

With the choice of basis of Eq. (2.7), the original thirteen-by-thirteen system of coupled equations in the flavor basis reduce to a three-by-three system of coupled equations and ten additional decoupled differential equations. Expressing the QED DGLAP equations given in Eq. (5.8) in terms of this evolution basis, we find that the singlet PDFs evolve as follows:

$$\nu^2 \frac{\partial}{\partial \nu^2} \begin{pmatrix} \gamma \\ \Sigma \\ \Delta_\Sigma \end{pmatrix} = \frac{\alpha(\nu)}{4\pi} \begin{pmatrix} e_\Sigma^2 P_{\gamma\gamma}^{(0)} & \eta^+ P_{\gamma q}^{(0)} & \eta^- P_{\gamma q}^{(0)} \\ \theta^- P_{q\gamma}^{(0)} & \eta^+ P_{qq}^{(0)} & \eta^- P_{qq}^{(0)} \\ \theta^+ P_{q\gamma}^{(0)} & \eta^- P_{qq}^{(0)} & \eta^+ P_{qq}^{(0)} \end{pmatrix} \otimes \begin{pmatrix} \gamma \\ \Sigma \\ \Delta_\Sigma \end{pmatrix}, \tag{2.8}$$

where, using the fact that $e_u^2 = e_c^2 = e_t^2$ and $e_d^2 = e_s^2 = e_b^2$, we have defined:

$$\begin{aligned} e_\Sigma^2 &\equiv N_c(n_u e_u^2 + n_d e_d^2)\,, \\ \eta^\pm &\equiv \frac{1}{2}\left(e_u^2 \pm e_d^2\right)\,, \\ \theta^\pm &\equiv 2N_c n_f \left[\left(\frac{n_u - n_d}{n_f}\right)\eta^\pm + \eta^\mp\right]\,, \end{aligned} \tag{2.9}$$

where n_u and n_d are the number of up- and down-type active quark flavors, respectively, and $n_f = n_u + n_d$. The non-singlet PDFs, instead, obey the multiplicative evolution equation:

$$\nu^2 \frac{\partial}{\partial \nu^2} q_i^{\rm NS}(x,\nu) = e_i^2 P_{qq}^{(0)}(x) \otimes q_i^{\rm NS}(x,\nu)\,, \tag{2.10}$$

where the electric charge $e_i^2 = e_u^2$ for the up-type distributions $q_i^{\rm NS} = \Delta_{uc}, \Delta_{ct}, u^-, c^-, t^-$ while $e_i^2 = e_d^2$ for the down-type distributions $q_i^{\rm NS} = \Delta_{ds}, \Delta_{sb}, d^-, s^-, b^-$. Let us mention that strictly speaking Eq. (2.10) is valid only if all the quark flavors are present in the evolution, that is for $n_f = 6$. For $3 \le n_f \le 5$, some non-singlet PDF (Δ_{uc}, Δ_{sb} and Δ_{ct}) will not evolve independently, since they can be written as a linear combination of singlet PDFs. For instance, below the charm threshold, $\Delta_{uc} = u^+ = (\Sigma + \Delta_\Sigma)/2$.

The solution of Eqs. (2.8) and (2.10) determines the QED evolution operators that evolve the singlet and non-singlet PDFs from the initial scale ν_0 to some final scale ν according to the equations:

$$\begin{aligned} \mathbf{q}^{\rm SG}(x,\nu) &= \mathbf{\Gamma}^{\rm SG}_{\rm QED}(x|\nu,\nu_0) \otimes \mathbf{q}^{\rm SG}(x,\nu_0)\,, \\ q_i^{\rm NS}(x,\nu) &= \Gamma^{\rm NS}_{{\rm QED},i}(x|\nu,\nu_0) \otimes q_i^{\rm NS}(x,\nu_0)\,, \end{aligned} \tag{2.11}$$

where the singlet evolution operator $\mathbf{\Gamma}^{\rm SG}_{\rm QED}$ is a three-by-three matrix while the non-singlet evolution operators $\Gamma^{\rm NS}_{{\rm QED},i}$ form an scalar array. In Sect. 2.3 we will show how to compute numerically these evolution operators solving the corresponding integro-differential equations by means of higher-order interpolation techniques.

Once the QED evolution operators in Eq. (2.11) have been computed by means of some suitable numerical method, one needs to combine them with the corresponding QCD evolution operators. In order to perform the combination, we can write Eq. (2.11) in a matrix form introducing in the PDF basis also the gluon PDF $g(x,\nu,\mu)$. Taking into account the fact that at leading order in QED the gluon PDF does not evolve, reintroducing the dependence on the QCD factorization scales μ and dropping for simplicity the dependence on x, we can write Eq. (2.11) as follows:

$$\underbrace{\begin{pmatrix} g(\mu,\nu) \\ \mathbf{q}^{\rm SG}(\mu,\nu) \\ q_1^{\rm NS}(\mu,\nu) \\ \vdots \\ q_{10}^{\rm NS}(\mu,\nu) \end{pmatrix}}_{\mathbf{q}(\mu,\nu)} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{\Gamma}^{\rm SG}_{\rm QED} & 0 & \dots & 0 \\ 0 & 0 & \Gamma^{\rm NS}_{{\rm QED},1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Gamma^{\rm NS}_{{\rm QED},10} \end{pmatrix}}_{\mathbf{\Gamma}^{\rm QED}(\nu,\nu_0)} \otimes \underbrace{\begin{pmatrix} g(\mu,\nu_0) \\ \mathbf{q}^{\rm SG}(\mu,\nu_0) \\ q_1^{\rm NS}(\mu,\nu_0) \\ \vdots \\ q_{10}^{\rm NS}(\mu,\nu_0) \end{pmatrix}}_{\mathbf{q}(\mu,\nu_0)}\,. \tag{2.12}$$

In the above expression, we have denoted by $\mathbf{q}(\mu,\nu)$ the fourteen-dimensional vector that contains all PDF combinations in the QED evolution basis of Eq. (2.7) plus the gluon PDF. Of course, a similar expression as that of Eq. (2.11) will hold for the solution of the QCD DGLAP evolution equations:

$$\widetilde{\mathbf{q}}(\mu,\nu) = \widetilde{\mathbf{\Gamma}}^{\rm QCD}(\mu,\mu_0)\otimes\widetilde{\mathbf{q}}(\mu_0,\nu)\,, \tag{2.13}$$

where in this case the vector $\widetilde{\mathbf{q}}$ is given in the QCD evolution basis, which is a different linear combination of the quark, anti-quark, gluon and photon PDFs as compared to the corresponding QED evolution basis. The two basis are related by an invertible fourteen-by-fourteen rotation matrix $\mathbf{T}$ that transforms the vector $\widetilde{\mathbf{q}}$ into the vector $\mathbf{q}$:

$$\mathbf{q} = \mathbf{T}\cdot\widetilde{\mathbf{q}} \quad\Longrightarrow\quad \widetilde{\mathbf{q}} = \mathbf{T}^{-1}\cdot\mathbf{q}\,. \tag{2.14}$$

Using Eq. (2.14) and the condition $\mathbf{T}\cdot\mathbf{T}^{-1} = \mathbf{1}$, the solution of the QED evolution equations Eq. (2.12) can be rotated as follows:

$$\widetilde{\mathbf{q}}(\mu,\nu) = \underbrace{\left[\mathbf{T}^{-1}\cdot\mathbf{\Gamma}^{\rm QED}(\nu,\nu_0)\cdot\mathbf{T}\right]}_{\widetilde{\mathbf{\Gamma}}^{\rm QED}(\nu,\nu_0)}\otimes\widetilde{\mathbf{q}}(\mu,\nu_0)\,. \tag{2.15}$$

where $\widetilde{\mathbf{\Gamma}}^{\rm QED}(\nu,\nu_0)$ is now the QED evolution operator expressed in the QCD evolution basis. Eqs. (2.13) and (2.15) determine the QCD and the QED evolution, respectively, of PDFs in the QCD evolution basis and can therefore be consistently used to construct a combined QCD⊗QED evolution operator. In the following, we drop all the tildes since it is understood that PDFs and evolution operators are always expressed in the QCD evolution basis.

Now, when combining QCD and QED evolution operators we are faced with an inherent ambiguity. Given that QCD and QED evolutions take place by means of the matrix evolution operators $\mathbf{\Gamma}^{\rm QCD}$ and $\mathbf{\Gamma}^{\rm QED}$ that do not commute,

$$[\mathbf{\Gamma}^{\rm QCD},\mathbf{\Gamma}^{\rm QED}] \neq 0\,, \tag{2.16}$$

this implies that performing first the QCD evolution followed by the QED evolution leads to a different result if the opposite order is assumed. We can then define the two possible cases:

$$\mathbf{\Gamma}^{\rm QCED}(\mu,\mu_0;\nu,\nu_0) \equiv \mathbf{\Gamma}^{\rm QED}(\nu,\nu_0)\otimes\mathbf{\Gamma}^{\rm QCD}(\mu,\mu_0)\,, \tag{2.17}$$

$$\mathbf{\Gamma}^{\rm QECD}(\mu,\mu_0;\nu,\nu_0) \equiv \mathbf{\Gamma}^{\rm QCD}(\mu,\mu_0)\otimes\mathbf{\Gamma}^{\rm QED}(\nu,\nu_0)\,, \tag{2.18}$$

and the condition in Eq. (2.16) implies that:

$$\mathbf{\Gamma}^{\rm QCED}(\mu,\mu_0;\nu,\nu_0)\otimes\mathbf{q}(\mu_0,\nu_0) \neq \mathbf{\Gamma}^{\rm QECD}(\mu,\mu_0;\nu,\nu_0)\otimes\mathbf{q}(\mu_0,\nu_0)\,. \tag{2.19}$$

However, using the analytical solution of the QCD and QED DGLAP equations in Mellin space and the Baker-Campbell-Hausdorff formula, it is possible to show that:

$$[\mathbf{\Gamma}^{\rm QCD},\mathbf{\Gamma}^{\rm QED}] = \mathcal{O}(\alpha\alpha_s)\,, \tag{2.20}$$

A careful analysis of the expansions of the two combined evolution operators in Eqs. (2.17) and (2.18) shows that they have a similar perturbative structure:

$$\mathbf{\Gamma}^{\rm QCED} = \sum_{n=0}^{\infty} (\alpha \boldsymbol{A} + \alpha_s \boldsymbol{B})^n + \alpha\alpha_s \boldsymbol{C} + \mathcal{O}(\alpha^2)\,, \tag{2.21}$$

$$\mathbf{\Gamma}^{\rm QECD} = \sum_{n=0}^{\infty} (\alpha \boldsymbol{A} + \alpha_s \boldsymbol{B})^n - \alpha\alpha_s \boldsymbol{C} + \mathcal{O}(\alpha^2)\,. \tag{2.22}$$

These expansions suggest a third possibility for the combined evolution operator given by the average of the $\mathbf{\Gamma}^{\rm QCED}$ and $\mathbf{\Gamma}^{\rm QECD}$ operators:

$$\mathbf{\Gamma}^{\rm QavD} \equiv \frac{\mathbf{\Gamma}^{\rm QCED} + \mathbf{\Gamma}^{\rm QECD}}{2}\,, \tag{2.23}$$

so that the subleading terms $\mathcal{O}(\alpha\alpha_s)$ cancel and the perturbative remainder is $\mathcal{O}(\alpha^2)$.

A possible objection to this approach is that, in the case in which μ and ν are very different from each other, this procedure might lead to the presence of numerically large, unresummed logarithms. So, in order to suppress the impact of these potentially large (subleading) logarithms, we have implemented in `APFEL` the combination of QCD and QED evolutions not over the whole (possibly large) $[Q_0, Q]$ range, but rather dividing it in small intervals $[Q_0, Q_1]$, $[Q_1, Q_2]$, ..., $[Q_N, Q]$, and performing the combination on each interval. This procedure ensures that no artificially large logarithm of two widely different scales appears in the solution.

In Sect. 2.4 we will show that the `QavD`, `QCED` and `QECD` solutions implemented with this strategy turn out to be good approximations to the unified approach, which is equivalent to the MRST2004QED [109] and `partonevolution` [108,110] implementations, all of them different by $\mathcal{O}(\alpha^2)$ terms only.

2.2.3 Basis for the unified QCD⊗QED solution

Another common choice for the solution of the combined DGLAP consists in solving the unified QCD⊗QED system with a specific basis which satisfies both evolution equations. In order to diagonlize as much as possible the evolution matrix in the presence of QED corrections avoiding unnecessary couplings between parton distributions, we propose the following evolution basis

$$\begin{array}{ll}
\texttt{1}: g & \\
\texttt{2}: \gamma & \\
\texttt{3}: \Sigma = \Sigma_u + \Sigma_d & \texttt{9}: V = V_u + V_d \\
\texttt{4}: \Delta_\Sigma = \Sigma_u - \Sigma_d & \texttt{10}: \Delta_V = V_u - V_d \\
\texttt{5}: T_1^u = u^+ - c^+ & \texttt{11}: V_1^u = u^- - c^- \\
\texttt{6}: T_2^u = u^+ + c^+ - 2t^+ & \texttt{12}: V_2^u = u^- + c^- - 2t^- \\
\texttt{7}: T_1^d = d^+ - s^+ & \texttt{13}: V_1^d = d^- - s^- \\
\texttt{8}: T_2^d = d^+ + s^+ - 2b^+ & \texttt{14}: V_2^d = d^- + s^- - 2b^-\,,
\end{array} \tag{2.24}$$

where we have introduced the singlet, triplets and valences combinations

$$\Sigma_u = \sum_{k=1}^{n_u} u_k^+ \,, \quad \Sigma_d = \sum_{k=1}^{n_d} d_k^+ \,, \tag{2.25}$$

$$V_u = \sum_{k=1}^{n_u} u_k^- \,, \quad V_d = \sum_{k=1}^{n_d} d_k^- \,. \tag{2.26}$$

When considering leading-order QED corrections to DGLAP, the equations for this basis are divided into three sub-systems: the singlet, the valence-sector and the non-singlet. For the singlet sector we have

$$\begin{aligned} \mu^2 \frac{\partial}{\partial \mu^2} \begin{pmatrix} g \\ \gamma \\ \Sigma \\ \Delta_\Sigma \end{pmatrix} &= \left[\begin{pmatrix} \widetilde{P}_{gg} & 0 & \widetilde{P}_{gq} & 0 \\ 0 & 0 & 0 & 0 \\ 2n_f \widetilde{P}_{qg} & 0 & \widetilde{P}_{qq} & 0 \\ \frac{n_u - n_d}{n_f} 2n_f \widetilde{P}_{qg} & 0 & \frac{n_u - n_d}{n_f}(\widetilde{P}_{qq} - \widetilde{P}^+) & \widetilde{P}^+ \end{pmatrix} \right. \\ &+ \left. \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & e_\Sigma^2 P_{\gamma\gamma}^{(0)} & \eta^+ P_{\gamma q}^{(0)} & \eta^- P_{\gamma q}^{(0)} \\ 0 & \theta^- P_{q\gamma}^{(0)} & \eta^+ P_{qq}^{(0)} & \eta^- P_{qq}^{(0)} \\ 0 & \theta^+ P_{q\gamma}^{(0)} & \eta^- P_{qq}^{(0)} & \eta^+ P_{qq}^{(0)} \end{pmatrix} \right] \begin{pmatrix} g \\ \gamma \\ \Sigma \\ \Delta_\Sigma \end{pmatrix} , \end{aligned} \tag{2.27}$$

where we separate the splitting matrix into two elements: the first matrix contains the QCD splitting functions $\widetilde{P}_{ij}$ meanwhile the second contains LO QED splittings $P_{ij}^{(0)}$. Note that the QED splitting matrix is identical to Eq. (2.8). For the QCD sector we have introduce the usual notation [114–124] in terms of flavor singlet (S) and non-singlet (V) quantities:

$$\begin{aligned} \widetilde{P}_{q_i q_j} &= \widetilde{P}_{\bar{q}_i \bar{q}_j} = \delta_{ij} \widetilde{P}_{qq}^V + \widetilde{P}_{qq}^S \\ \widetilde{P}_{\bar{q}_i q_j} &= \widetilde{P}_{q_i \bar{q}_j} = \delta_{ij} \widetilde{P}_{\bar{q}q}^V + \widetilde{P}_{\bar{q}q}^S \\ \widetilde{P}_{q_i g} &= \widetilde{P}_{\bar{q}_i g} = \widetilde{P}_{qg} \\ \widetilde{P}_{g q_i} &= \widetilde{P}_{g \bar{q}_i} = \widetilde{P}_{gq} \,. \end{aligned} \tag{2.28}$$

It follows the definition of $\widetilde{P}^\pm$, $\widetilde{P}_{qq}$ and $\widetilde{P}^V$ as

$$\begin{aligned} \widetilde{P}^\pm &\equiv \widetilde{P}_{qq}^V \pm \widetilde{P}_{q\bar{q}}^V \\ \widetilde{P}_{qq} &\equiv \widetilde{P}^+ + n_f(\widetilde{P}_{qq}^S + \widetilde{P}_{q\bar{q}}^S) \quad , \\ \widetilde{P}^V &\equiv \widetilde{P}^- + n_f(\widetilde{P}_{qq}^S - \widetilde{P}_{q\bar{q}}^S) \end{aligned} \tag{2.29}$$

The second system to solve is the valence-sector defined as

$$\mu^2 \frac{\partial}{\partial \mu^2} \begin{pmatrix} V \\ \Delta_V \end{pmatrix} = \left[\begin{pmatrix} \widetilde{P}^V & 0 \\ \frac{n_u - n_d}{n_f}(\widetilde{P}^V - \widetilde{P}^-) & \widetilde{P}^- \end{pmatrix} + \begin{pmatrix} \eta^+ P_{qq}^{(0)} & \eta^- P_{qq}^{(0)} \\ \eta^- P_{qq}^{(0)} & \eta^+ P_{qq}^{(0)} \end{pmatrix} \right] \begin{pmatrix} V \\ \Delta_V \end{pmatrix} . \tag{2.30}$$

Finally, we have the non-singlet equations for the remaining evolution flavors:

$$\begin{aligned}
\mu^2\frac{\partial T^u_{1,2}}{\partial\mu^2} &= (\widetilde{P}^+ + e_u^2 P^{(0)}_{qq})T^u_{1,2}\,,\\
\mu^2\frac{\partial T^d_{1,2}}{\partial\mu^2} &= (\widetilde{P}^+ + e_d^2 P^{(0)}_{qq})T^d_{1,2}\,,\\
\mu^2\frac{\partial V^u_{1,2}}{\partial\mu^2} &= (\widetilde{P}^- + e_u^2 P^{(0)}_{qq})V^u_{1,2}\,,\\
\mu^2\frac{\partial V^d_{1,2}}{\partial\mu^2} &= (\widetilde{P}^- + e_d^2 P^{(0)}_{qq})V^d_{1,2}\,.
\end{aligned} \tag{2.31}$$

The basis presented here is just an example of possible choice for the unified solution, which is implemented in `APFEL` as `QUniD` solution, however many other choices are possible.

2.3 Numerical techniques

In this section we will present the numerical techniques that `APFEL` uses to solve the DGLAP evolution equations. The same numerical techniques presented here are applied to both QCD and QED DGLAP evolution equations thanks to the same formal structure. In order to show the general strategy, we will see how `APFEL` solves the QCD evolution equations but keeping in mind that the same procedure applies to the QED ones as well.

The DGLAP evolution equations can be written as:

$$\mu^2\frac{\partial q_i(x,\mu)}{\partial\mu^2} = \int_x^1 \frac{dy}{y} P_{ij}\left(\frac{x}{y},\alpha_s(\mu)\right) q_j(y,\mu)\,, \tag{2.32}$$

where $P_{ij}\left(x,\alpha_s(\mu)\right)$ are the usual splitting functions up to some perturbative order in α_s. If we make the following definitions:

$$\begin{aligned}
t &\equiv \ln(\mu^2)\,,\\
\tilde{q}(x,t) &\equiv xq(x,\mu)\,,\\
\tilde{P}_{ij}(x,t) &\equiv xP_{ij}(x,\alpha_s(\mu))\,,
\end{aligned} \tag{2.33}$$

Eq. (2.32) becomes:

$$\frac{\partial\tilde{q}_i(x,t)}{\partial t} = \int_x^1 \frac{dy}{y}\tilde{P}_{ij}\left(\frac{x}{y},t\right)\tilde{q}_j(y,t)\,. \tag{2.34}$$

In order to numerically solve the above equation, we choose to express PDFs in terms of an interpolation basis over an x grid with N_x+1 points. This way we can write:

$$\tilde{q}(y,t) = \sum_{\alpha=0}^{N_x} w_\alpha^{(k)}(y)\tilde{q}(x_\alpha,t)\,, \tag{2.35}$$

where $\{w_\alpha^{(k)}(y)\}$ is a set of interpolation functions of degree k. In `APFEL` we have chosen to use the Lagrange interpolation method and therefore the interpolation functions read:

$$w_\alpha^{(k)}(x) = \sum_{j=0, j\leq\alpha}^{k} \theta(x - x_{\alpha-j})\theta(x_{\alpha-j+1} - x) \prod_{\delta=0,\delta\neq j}^{k} \left[\frac{x - x_{\alpha-j+\delta}}{x_\alpha - x_{\alpha-j+\delta}}\right] . \tag{2.36}$$

Notice that Eq. (2.36) implies that:

$$w_\alpha^{(k)}(x) \neq 0 \quad \text{for} \quad x_{\alpha-k} < x < x_{\alpha+1} . \tag{2.37}$$

Now we can rewrite Eq. (2.34) as follows:

$$\frac{\partial \tilde{q}_i(x,t)}{\partial t} = \sum_\alpha \left[\int_x^1 \frac{dy}{y} \tilde{P}_{ij}\left(\frac{x}{y}, t\right) w_\alpha^{(k)}(y)\right] \tilde{q}_j(x_\alpha, t) . \tag{2.38}$$

In the particular case in which the x variable in Eq. (2.38) coincides with one of the x-grid nodes, say x_β, the evolution equations take the following discretized form:

$$\frac{\partial \tilde{q}_i(x_\beta,t)}{\partial t} = \sum_\alpha \underbrace{\left[\int_{x_\beta}^1 \frac{dy}{y} \tilde{P}_{ij}\left(\frac{x_\beta}{y}, t\right) w_\alpha^{(k)}(y)\right]}_{\Pi_{ij,\beta\alpha}(t)} \tilde{q}_j(x_\alpha, t) . \tag{2.39}$$

From Eq. (2.37) follows the condition:

$$\Pi_{ij,\beta\alpha}(t) \neq 0 \quad \text{for} \quad \beta \leq \alpha . \tag{2.40}$$

In addition, the computation $\Pi_{ij,\beta\alpha}$ in Eq. (2.39) can be simplified to:

$$\Pi_{ij,\beta\alpha}(t) = \int_a^b \frac{dy}{y} \tilde{P}_{ij}\left(\frac{x_\beta}{y}, t\right) w_\alpha^{(k)}(y) , \tag{2.41}$$

where the integration bounds are given by:

$$a \equiv \max(x_\beta, x_{\alpha-k}) \quad \text{and} \quad b \equiv \min(1, x_{\alpha+1}) . \tag{2.42}$$

Alternatively, by means of a change of variable, the integral in Eq. (2.41) can be rearranged as follows:

$$\Pi_{ij,\beta\alpha}(t) = \int_c^d \frac{dy}{y} \tilde{P}_{ij}(y,t) w_\alpha\left(\frac{x_\beta}{y}\right) , \tag{2.43}$$

where the new integration bounds are defined as:

$$c \equiv \max(x_\beta, x_\beta/x_{\alpha+1}) \quad \text{and} \quad d \equiv \min(1, x_\beta/x_{\alpha-k}) . \tag{2.44}$$

One central aspect of the numerical methods used in `APFEL` is the use of an interpolation over a logarithmically distributed x grid. In this case, the interpolation coefficients in Eq. (2.36) can be expressed as

$$w_\alpha^{(k)}(x) = \sum_{j=0, j\leq\alpha}^{k} \theta(x - x_{\alpha-j})\theta(x_{\alpha-j+1} - x) \prod_{\delta=0,\delta\neq j}^{k} \left[\frac{\ln(x) - \ln(x_{\alpha-j+\delta})}{\ln(x_\alpha) - \ln(x_{\alpha-j+\delta})}\right] . \tag{2.45}$$

If in addition the x grid is logarithmically distributed, *i.e.* such that $\ln(x_\beta) - \ln(x_\alpha) = (\beta - \alpha)\Delta$, where the step Δ is a constant, one has that the interpolating functions read:

$$w_\alpha^{(k)}(x) = \sum_{j=0,\, j\leq\alpha}^{k} \theta(x - x_{\alpha-j})\theta(x_{\alpha-j+1} - x) \prod_{\delta=0,\delta\neq j}^{k} \left[\frac{1}{\Delta}\ln\left(\frac{x}{x_\alpha}\right)\frac{1}{j-\delta} + 1\right], \quad (2.46)$$

so that the dependence on x of the interpolating function $w_\alpha^{(k)}(x)$ is through the function $\ln(x/x_\alpha)$ only. Therefore, it can be shown that in Eq. (2.43) $w_\alpha^{(k)}(x_\beta/y)$ depends only on the combination $[(\beta-\alpha)\Delta - \ln y]$ and thus $\Pi_{ij,\beta\alpha}$ depends only on the difference $(\beta - \alpha)$.

One can use this information, together with the condition in Eq. (2.40), to represent $\Pi_{ij,\beta\alpha}(t)$ as a matrix, where β is the row index and α the column index. Such a representation of $\Pi_{ij,\beta\alpha}(t)$ reads:

$$\Pi_{ij,\beta\alpha}(t) = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{N_x} \\ 0 & a_0 & a_1 & \cdots & a_{N_x-1} \\ 0 & 0 & a_0 & \cdots & a_{N_x-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_0 \end{pmatrix}. \quad (2.47)$$

The knowledge of the first row of the matrix $\Pi_{ij,\beta\alpha}(t)$ is enough to determine all the other entries. This feature, which is based on the particular choice of the interpolation procedure, leads to a more efficient computation of the evolution operators since it reduces by a factor N_x the number of integrals to be computed.

After the presentation of the interpolation method, we turn to discuss the actual computation of the evolution operators. Any splitting function, be it QED or QCD at any given perturbative order, has the following general structure:

$$\tilde{P}_{ij}(x,t) = xP_{ij}^R(x,t) + \frac{xP_{ij}^S(x,t)}{(1-x)_+} + P_{ij}^L(t)x\delta(1-x)\,, \quad (2.48)$$

where $P_{ij}^R(x,t)$ is the regular term, $P_{ij}^S(x,t)$ is the coefficient of the plus-distribution term, and $P_{ij}^L(t)$ is the coefficient of the local term proportional to the delta functions. It is useful to recall here that the general definition of plus-distribution in the presence of arbitrary integration bounds is given by:

$$\int_c^d dy \frac{f(y)}{(1-y)_+} = \int_c^d dy \frac{f(y) - f(1)\theta(d-1)}{1-y} + f(1)\ln(1-c)\theta(d-1)\,. \quad (2.49)$$

Moreover, each of the functions P_{ij} appearing in Eq. (2.48) has the usual perturbative expansion that at N^{k}LO reads:

$$P_{ij}^J(x,t) = \sum_{n=0}^{k} a_s^{n+1}(t) P_{ij}^{J,(n)}(x), \quad \text{with} \quad J = R, S, L\,, \quad (2.50)$$

where we have defined $a_s \equiv \alpha_s/4\pi$.

Taking the above considerations into account and using the fact that $w_\alpha^{(k)}(x_\beta) = \delta_{\beta\alpha}$, we can write the evolution operators in terms of the various parts of the splitting functions as follows:

$$\Pi_{ij,\beta\alpha}(t) =$$

$$\sum_{n=0}^{k} a_s^{n+1}(t)\Bigg\{\int_c^d dy\left[P_{ij}^{R,(n)}(y)w_\alpha\left(\frac{x_\beta}{y}\right)+\frac{P_{ij}^{S,(n)}(y)}{1-y}\left(w_\alpha\left(\frac{x_\beta}{y}\right)-\delta_{\beta\alpha}\theta(d-1)\right)\right]$$

$$+\left[P_{ij}^{S,(n)}(1)\ln(1-c)\theta(d-1)+P_{ij}^{L,(n)}\right]\delta_{\beta\alpha}\Bigg\} \equiv \sum_{n=0}^{k} a_s^{n+1}(t)\Pi_{ij,\beta\alpha}^{(n)}\,, \tag{2.51}$$

where the coefficients $\Pi_{ij,\beta\alpha}^{(n)}$ are independent of the energy scale t, and need to be evaluated a single time once the x interpolation grid and the evolution parameters have been defined.

Now we will show that Eq. (2.51) respects the symmetry conditions of Eq. (2.47). We can distinguish two cases: 1) $d < 1$ and 2) $d = 1$. In the case 1), due to the presence of the Heaviside functions $\theta(d-1)$, Eq. (2.51) reduces to:

$$\Pi_{ij,\beta\alpha}^{(n)} = \int_c^d dy\left[P_{ij}^{R,(n)}(y)+\frac{P_{ij}^{S,(n)}(y)}{1-y}\right]w_\alpha\left(\frac{x_\beta}{y}\right)+P_{ij}^{L,(n)}\delta_{\beta\alpha}\,, \tag{2.52}$$

which clearly follows Eq. (2.47). In the case 2), instead, we have:

$$\Pi_{ij,\beta\alpha}^{(n)} = \int_c^1 dy\left[P_{ij}^{R,(n)}(y)w_\alpha\left(\frac{x_\beta}{y}\right)+\frac{P_{ij}^{S,(n)}(y)}{1-y}\left(w_\alpha\left(\frac{x_\beta}{y}\right)-\delta_{\beta\alpha}\right)\right]$$
$$+\left[P_{ij}^{S,(n)}(1)\ln(1-c)+P_{ij}^{L,(n)}\right]\delta_{\beta\alpha}\,, \tag{2.53}$$

and apparently, if $\alpha = \beta$, the term proportional to $\ln(1-c)$ could break the symmetry. However, from Eq. (2.44), we know that in this case:

$$c = \max(x_\beta, x_\beta/x_{\beta+1}) = \frac{x_\beta}{x_{\beta+1}}\,, \tag{2.54}$$

because $x_{\beta+1} \leq 1$. In addition, on a logarithmically distributed grid we have that $x_{\beta+1} = x_\beta \exp(\Delta)$. Therefore, it turns out that:

$$\ln(1-c) = \ln\left(1-\frac{x_\beta}{x_{\beta+1}}\right) = \ln[1-\exp(-\Delta)]\,, \tag{2.55}$$

which is a constant which does not depend on the indices α and β and therefore satisfies Eq. (2.47).

At this point, the DGLAP equations imply that the discretized PDFs evolve between two scales t and t_0 according to the following matrix equation:

$$\tilde{q}_i(x_\beta, t) = \sum_{\gamma,k} \Gamma_{ik,\beta\gamma}(t, t_0)\tilde{q}_k(x_\gamma, t_0)\,, \tag{2.56}$$

where it follows from Eq. (2.39) that the evolution operators are given by the solution of the system:

$$\begin{cases} \dfrac{\partial \Gamma_{ij,\alpha\beta}(t,t_0)}{\partial t} = \displaystyle\sum_{\gamma,k} \Pi_{ik,\alpha\gamma}(t)\Gamma_{kj,\gamma\beta}(t,t_0) \\ \\ \Gamma_{ij,\alpha\beta}(t_0,t_0) = \delta_{ij}\delta_{\alpha\beta} \end{cases} \tag{2.57}$$

Eq. (2.57) is a set of coupled first order ordinary linear differential equations for the evolution operators $\Gamma_{ij,\alpha\beta}(t,t_0)$. In `APFEL` Eq. (2.57) is solved using a fourth-order adaptive step-size control Runge-Kutta (RK) algorithm. Note that no interpolation in t is involved, the solution of the differential equations in t is only limited by the precision of the RK method. Once the evolved PDFs at the grid values $\tilde{q}_i(x_\beta,t)$ have been determined by means of the evolution operators in Eq. (2.56), the value of these same PDFs for arbitrary values of x will be computed using again higher-order interpolation.

A final consideration concerning the choice of interpolating grid in x is needed. As is well known, an accurate solution of the DGLAP equations requires a denser grid at large x, where PDFs have more structure than at small-x. In `APFEL` it is not possible to use an x-grid with variable spacing that allows to have a denser grid at large x and at the same time to maintain the symmetry that allows to substantially reduce the number of integrals to be evaluated, see Eq. (2.47). In fact, a logarithmically distributed x grid necessarily leads to a looser grid in the large-x region, thus potentially degrading the evolution accuracy there. To overcome this problem, `APFEL` implements the possibility of using different interpolation grids according to the value of x in which PDFs need to be evaluated.

The basic idea is the following. The evolution of a given set of PDFs from the initial condition at the scale μ_0 up to some other scale μ is determined by the convolution between the evolution operators and the boundary conditions, which implies performing and integral between x and one. This convolution, when discretized on an interpolation x grid, corresponds to Eq. (2.56). It is clear that such operation will use only those x_β nodes of the interpolation grid that fall in the range between x and one.

Therefore, the computation of the PDF evolution in the large-x region using a logarithmically spaced interpolation grid with a small value of $x_{\rm min}$ will be certainly inefficient, since the convolution would use only a small number of points in the large-x region such that $x_\beta \leq x \leq 1$, discarding those with $x < x_\beta$. In order to avoid this problem and simultaneously achieve a good accuracy and performance over the whole range in x, `APFEL` gives the possibility to use different interpolating grids, each with a different value of $x_{\rm min}$, interpolation degree and number of points. Then, to compute the evolution of the PDFs for the point x, the program will automatically select the grid with the largest value of $x_{\rm min}$ compatible with the condition $x_{\rm min} \leq x$.

The use of $n \geq 2$ subgrids increases slightly the time taken by initialization phase, since more evolution operators need to be precomputed, and also the actual evolution is somewhat slower than in the case with a single grid ($n = 1$), with the important trade-off of a much more accurate result in the large-x region. As default settings, `APFEL` uses $n = 3$ interpolation grids, with interpolation order $3, 5$ and 5, number of points $N_x = 80, 50$ and 40 and $x_{\rm min} = 10^{-5}, 0.1$ and 0.8 respectively.

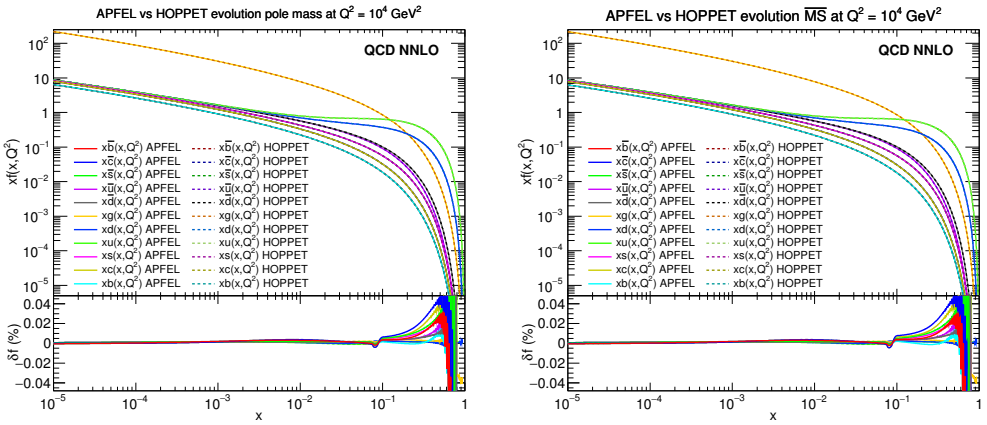

Figure 2.1: Comparison between PDFs evolved at NNLO in QCD using `APFEL` and `HOPPET`, from Q_0^2 =2 GeV^2 up to Q^2 =10^4 GeV^2, using the Les Houches PDF benchmark settings. The comparison is performed in the pole mass scheme (left) and in the $\overline{\rm MS}$ scheme (right). The lower plots show the percent differences between the two codes.

2.4 Validation and benchmarking

In this section we first perform a detailed benchmarking of `APFEL` against `HOPPET` finding good agreement for the QCD evolution up to NNLO, both with pole and $\overline{\rm MS}$ heavy quark masses. Then we turn to the validation of the combined QCD⊗QED evolution. We verify the consistency of the different methods for the solution of the combined QCD⊗QED evolution equations, showing that the coupled solution is numerically equivalent to the unified solution when constructed iteratively in small steps in Q. Finally, we compare the predictions of `APFEL` with: the `partonevolution` code, the internal MRST2004QED evolution and the `QCDNUM` library [101, 125].

2.4.1 QCD evolution

We validate the QCD evolution in `APFEL` by comparing it with the results from the `HOPPET` program, version 1.1.5, up to NNLO, and using both pole and $\overline{\rm MS}$ heavy quark masses. The settings are the same as in the original Les Houches PDF evolution benchmark [126]. In the case of $\overline{\rm MS}$ masses, we take the $\overline{\rm MS}$ Renormalization-Group-Invariant charm mass $m_c(m_c)$ to have the same numerical values as the pole masses. In all the comparisons in this section, the interpolation settings in `APFEL` are the default ones discussed in Sect. 2.3.

Results for the evolved PDFs at $Q^2 = 10^4$ GeV^2 for both `HOPPET` and `APFEL` are shown in Fig. 2.1. The left plot shows the results using pole masses, while the right plot corresponds to the case of $\overline{\rm MS}$ masses. Fig. 2.1 also shows the percent difference between both predictions, to show the excellent agreement obtained for the whole range in x, being at most $\sim 0.04\%$ at large-x, where PDFs have more structure.

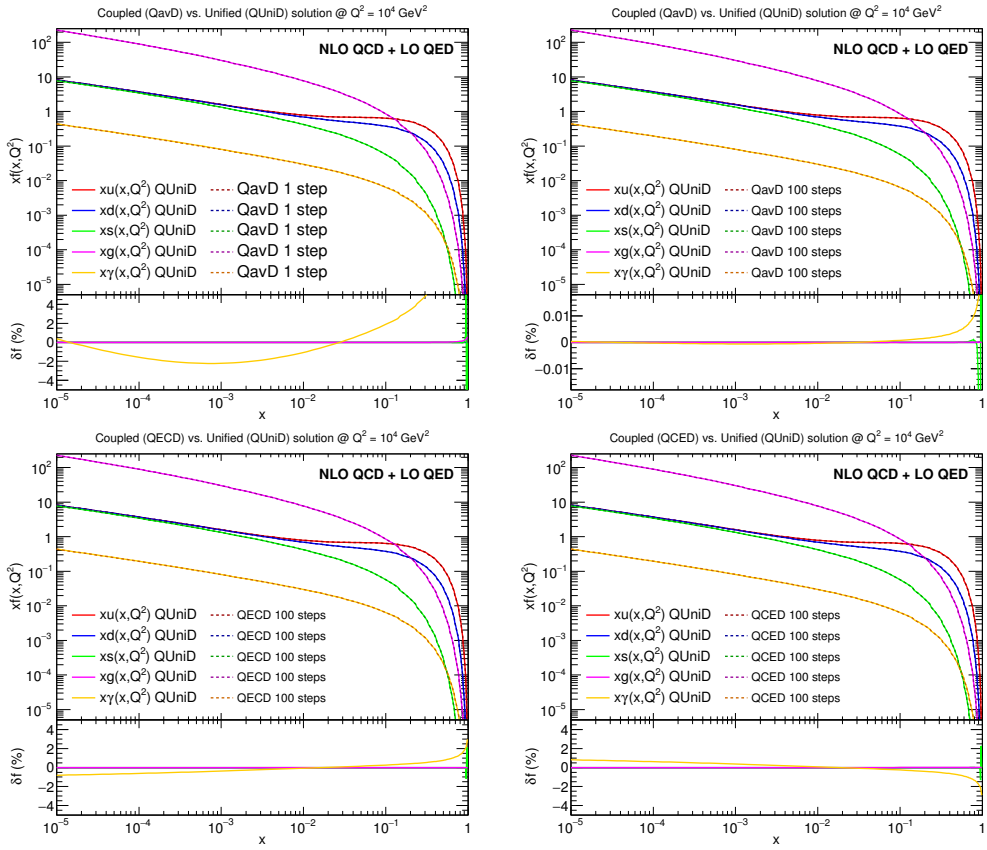

Figure 2.2: Comparison between PDFs evolved with APFEL with the combined QCD⊗QED DGLAP. We show in the plots of the top the comparison between the `QUniD` and the `QavD` solutions using 1 step (left plot) and 100 steps (right plot). The bottom plots show the comparison of `QUniD` and `QECD` (plot on the left) and `QCED` (plot on the right) both using 100 steps. The evolution is performed between $Q_0^2 = 2$ GeV^2 and $Q^2 = 10^4$ GeV^2 in the VFN scheme at NLO in QCD and LO in QED using the Les Houches PDF setup [126], supplemented by the ansatz $\gamma(x, Q_0) = 0$.

2.4.2 QCD⊗QED evolution

Consistency of the coupled solution

Before comparing `APFEL` to other libraries we first analyze the numerical impact of the coupled approach in function of the operators introduced in Sect. 2.2.2.

Using the same settings of the Les Houches PDF evolution benchmark [126], supplemented by the ansatz $\gamma(x, Q_0) = 0$ we evolve the PDFs at NLO in QCD and LO in QED between $Q_0^2 = 2$ GeV^2 and $Q^2 = 10^4$ GeV^2 using the VFN scheme. In the top left plot of Figure 2.2 we show the comparison between the unified solution `QUniD` and the average solution `QavD`, performed with a single step between Q_0^2 and Q^2. In the bottom panel of each plot we show the percentage difference between the two results: a

good agreement is found for all flavors except for the photon PDF, where at $x \sim 10^{-3}$ a peak of -2% difference is observed, followed by more important discrepancies at large x. The right plot of the same figure shows the comparison where the full range $[Q_0^2, Q^2]$ has been divided into 100 logarithmically spaced intervals. In this condition we obtain a good agreement between both solutions for all flavors. This result confirms that the average solution `QavD` is free of numerically large scale logarithms when introducing a moderate number of Q^2 intervals.

At this point we turn to consider the `QECD` and the `QCED` solutions. The result for 100 steps evolution is shown in the bottom plots of Fig. 2.2, where the `QECD` (left plot) and the `QCED` (right plot) solutions are compared to the unified solution `QUniD`. We observe evident discrepancies for the photon PDF: the `QECD` solution underestimates the photon evolution at small x meanwhile the `QCED` solution overestimates in the same region. It is important to highlight that both solution are not able to reproduce the same level of accuracy of the average solution, even if we require the same number of steps. This suggests that these solutions introduce artificially large logarithms and that an effective way to cancel them is to perform the evolution in smaller steps combining sequentially the results. In this regime the `QECD` and `QCED` solutions coincide to a good approximation with that of the `QavD` solution, so that all three strategies lead to the same numerical accuracy.

In the next paragraphs, in order to simplify the analysis, we will use exclusively the `QUniD` solution when comparing the combined QCD⊗QED evolution to other codes.

Comparison with `partonevolution`

We start the benchmarking exercise by comparing the results of the combined QCD⊗QED DGLAP evolution in `APFEL` with those of the public `partonevolution` code [108,110], version 1.1.3.

To perform the benchmark, we use `APFEL` with the same settings used in the original publication [108] to present the numerical results of `partonevolution`, *i.e.* we take the input PDFs from the toy model used in the benchmarking exercise of Ref. [127], given by:

$$\begin{aligned} xu_v(x) &= A_u x^{0.5}(1-x)^3\,, & xd_v(x) &= A_d x^{0.5}(1-x)^4\,,\\ xS(x) &= A_S x^{-0.2}(1-x)^7\,, & xg(x) &= A_g x^{-0.2}(1-x)^5\,,\\ xc(x) &= 0\,, & x\bar{c}(x) &= 0\,, \end{aligned} \tag{2.58}$$

at the initial scale $Q_0^2 = 4$ GeV^2, with a SU(3) symmetric sea that carries 15% of the proton's momentum at Q_0^2, and only four active quarks are considered even above the bottom threshold. This toy model should not be confused with that used in the Les Houches PDF benchmark study, used elsewhere in this paper. In addition, the photon PDF is set to zero at the initial scale, that is $\gamma(x, Q_0^2) = 0$.

In order to set up the baseline, we ran the two codes at NLO QCD only, switching off the QED corrections, in the FFN scheme with $n_f = 4$. As can be seen from the left plot of Fig. 2.3, a good agreement is achieved. The results about the combined QCD⊗QED evolution are summarized on the right plot of Fig. 2.3, where we compare the evolution of quark, gluon and photon PDFs given by the two codes, using the `QUniD` solution implemented in `APFEL`. With these settings the evolution of quarks and gluon is essentially identical, with differences at most being $\mathcal{O}(0.01\%)$, while differences in

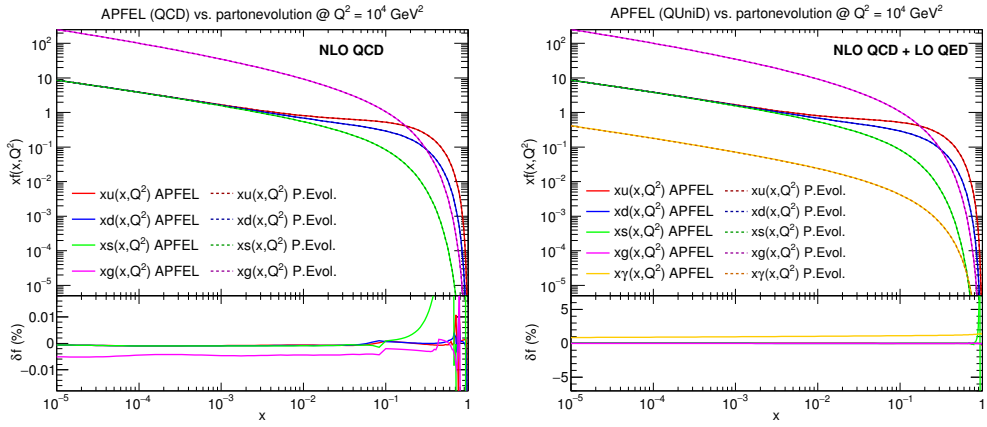

Figure 2.3: Comparison between PDFs evolved using `APFEL` and `partonevolution`, from Q_0^2 =4 GeV^2 up Q_0^2 =10^4 GeV^2. The same settings of the PDF benchmark study of Ref. [127] have been used. On the left plot we show evolution at NLO in QCD (without QED corrections), meanwhile on the right plot we consider the QCD⊗QED evolution. For each comparison, we also show the percent differences with respect to the `partonevolution` results.

the evolution of $\gamma(x, Q^2)$ are below the few percent level except at the largest values of x. More substantial differences appear for the photon PDF, in this case the solutions differ by up to 1%, both at small and large-x, however this level of agreement is still acceptable in view of the technical differences between both codes. As we will see in the next paragraphs the quality of the comparison is much better when using codes with a modern implementations of the QCD⊗QED combined evolution.

Comparison with MRST2004QED

At this point we compare `APFEL` to the QED evolution used in the determination of the MRST2004QED parton distributions [109]. Though the original evolution code is not publicly available, the evolution which was used can be indirectly accessed via the public `LHAPDF` grids. In this case, it is not possible to use the Les Houches benchmark settings, and we are instead forced to use the same boundary conditions for the PDFs at Q_0 as those used in the MRST2004QED fit as well as the same values of the heavy quark masses and reference coupling constants. The available MRST2004QED fit was obtained at NLO in QCD in the VFN scheme, therefore it is possible to perform a meaningful comparison with the results of their evolution by using `APFEL` at NLO with the same settings.

The comparison between the `APFEL` predictions and the MRST2004QED evolution is shown in Fig. 2.4, where PDFs have been evolved using `APFEL` and the internal MRST evolution from Q_0^2 =1.25 GeV^2 up to Q^2 =10^4 GeV^2. An excellent agreement is found for all flavors. We observe differences of -1% at most for the $\gamma(x, Q^2)$ PDF at large-x, meanwhile for quark and gluon PDFs the discrepancies are smaller.

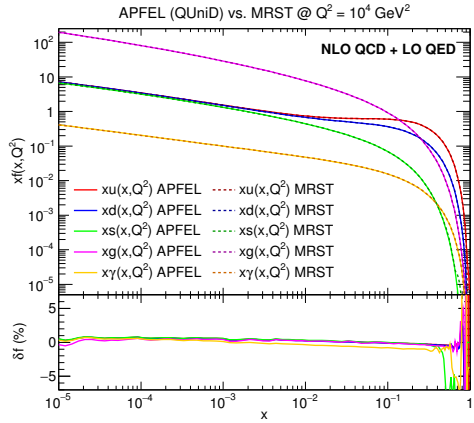

Figure 2.4: Comparison between PDFs evolved using `APFEL` and the internal MRST2004QED parton evolution, from Q_0^2 =1.25 GeV^2 up to Q^2 =10^4 GeV^2 at NLO in QCD and LO in QED using the VFN scheme. The boundary conditions for the PDFs are the same as those of the MRST2004QED fit.

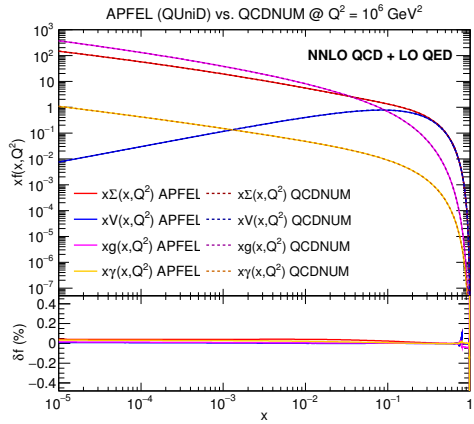

Figure 2.5: Comparison between PDFs evolved using `APFEL` and `QCDNUM` evolution, from Q_0^2 =2 GeV^2 up to Q^2 =10^6 GeV^2 at NNLO in QCD and LO in QED using the VFN scheme. The same settings of the PDF benchmark study of Ref. [127] have been used. We show PDFs in the evolution basis presented in Sect. 2.2.3.

Comparison with QCDNUM

We conclude this section by performing the comparison with the recent implementation of the combined QCD⊗QED evolution in the `QCDNUM` library [101, 125].

In Figure 2.5 we show the comparison of both codes from $Q_0^2 = 2$ GeV^2 up to $Q^2 = 10^6$ GeV^2 at NNLO in QCD and LO in QED, using the VFN scheme. The boundary condition for the input PDFs are the same of the PDF benchmark study of Ref. [127]. In this case, instead of plotting the single quark PDFs we have plotted

the singlet and valence PDFs defined in Sect. 2.2.3. The level of agreement between APFEL and QCDNUM is extremely good for all flavors, we observe differences of 0.02% at most in all cases.

In conclusion, the found a good level of agreement for all comparison performed in this section. This guarantees that APFEL implements correctly the QCD and QCD⊗QED evolutions, therefore it can be used in PDF fits.

2.5 APFEL Web

We conclude this chapter by presenting APFEL Web, a spin-off of the APFEL library, which has been ported to an online centralized server system. This service is designed with the objective to provide a fast and complete set of tools for PDF comparison, luminosities, DIS observables and theoretical prediction computed through the APPLgrid interface [128] with an user-friendly Web-application interface. The advantage of this system resides on the possibility to setup PDF evolution in real time, and perform quick comparison of the effects due to different configurations. In this respect, APFEL Web provides also a timely replacement to the HepData online PDF plotter[2].

APFEL Web is a Web-based application attached to a computer cluster, available online at:

http://apfel.mi.infn.it/

It contains PDF grids from LHAPDF5 and LHAPDF6 libraries and it allows users to evolve PDFs using custom configurations provided by the APFEL library. Computational results are presented in the format of plots which are produced by the ROOT framework.

This article is organized as follows. In Sect. 2.5.1 we document the application design and we explain the model scheme developed for this project. In Sect. 2.5.2 we discuss how to use the Web-application and obtain results. Finally, in Sect. 2.5.3 we present our conclusion and directions for future work.

2.5.1 Application design

The APFEL Web project is divided into two parts: the server-side and the cluster-side. The separation is a real requirement because the service needs to interact with multiple users and computational jobs at the same time. In the following we start from the description of the Web framework developed for the server-side and then we show how the combination is performed.

The Web framework and interface

For the development of the Web interface we have used the Django Web framework[3]. Django is a high-level Python Web framework which provides a high-performing solution for custom and flexible Web-applications. Moreover the choice of Python as programming language instead of PHP or Java, is motivated by the need of a simple interface to interact with the server system, by simplifying the implementation of the communication between server and cluster sides.

[2] http://hepdata.cedar.ac.uk/pdf/pdf3.html
[3] https://www.djangoproject.com/

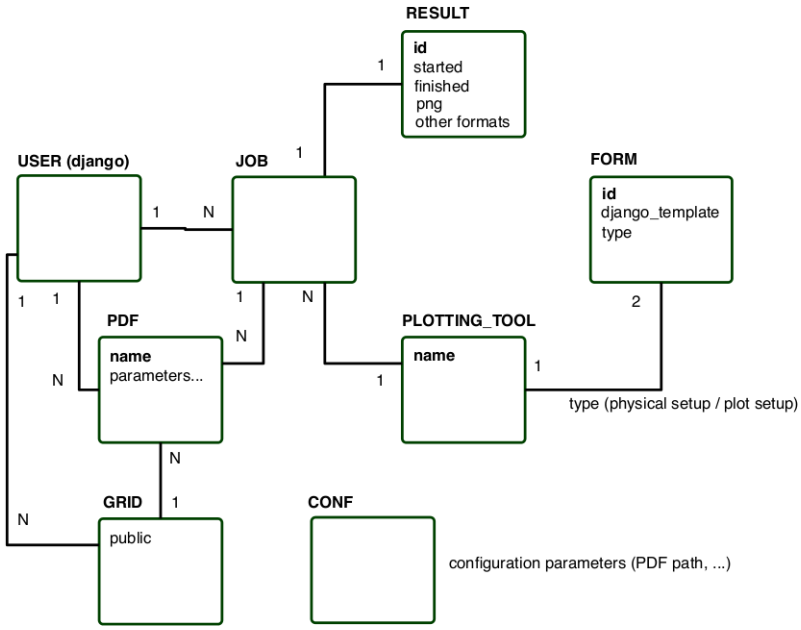

Figure 2.6: A static design scheme of the `APFEL Web` application model. The boxes represent a simplified view of the main components of this Web-application. Solid lines with 1/N labels represent the one/many relationships for each component of the application. Labels inside the boxes are examples of the database entry keys associated to the model.

Following the `Django` data model we have chosen to stored data in a `PostgreSQL`[4] database which should provide a good performance for our query requirements. We use the authentication system provided by the `Django` framework in order to create a personal user Web-space, so users can save privately personal configurations and start long jobs without need to be connected over the whole calculation time.

In Figure 2.6 we show a schematic view of the Web-application model used in `APFEL Web`. Starting from the top-left element, users have access to `PDF` objects which store in the database the information about the PDF: e.g. the set name, the PDF uncertainty treatment and the library for the treatment of PDF evolution. Users have the option to choose PDF sets from the `LHAPDF` library or, if preferred, upload their own private grid using the `LHAPDF5 LHgrid` and `LHAPDF6` formats. Users are able to run jobs after setting up the PDF grid objects: there are seven job types which are classified in the image as plotting tools and will be described in detail in Sect. 2.5.2. For each plotting tool there are customized input Web-forms, implemented with the `Django models` framework, which collect information and store it in the database before the

[4] http://www.postgresql.org

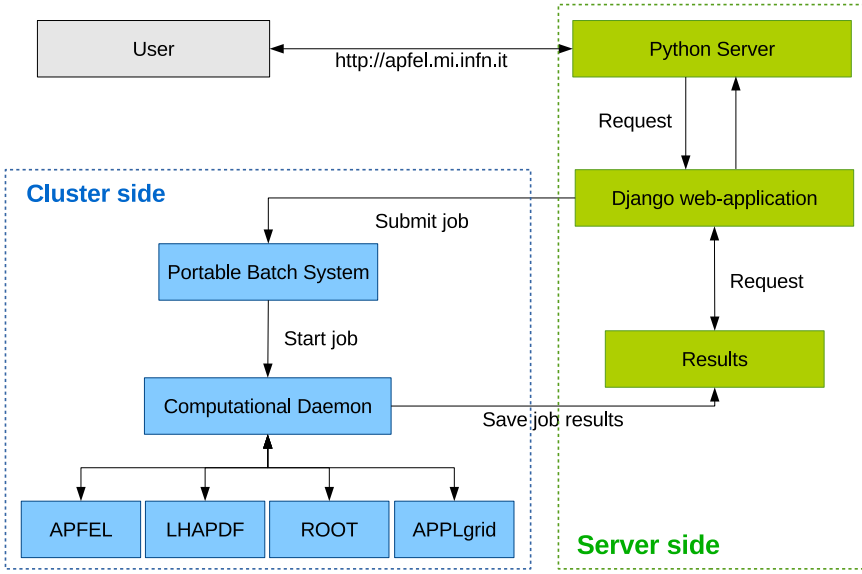

Figure 2.7: Deployment layout of `APFEL Web`.

job submission. When a job finalizes, it stores images and `ROOT` files to the server disk, which are then downloaded by the user. General configuration information such as the path of the PDF grids and libraries are stored directly into the `Django` settings.

Concerning Web-security in `APFEL Web`, the user's account and its information are protected by the `Django Middleware` framework. Undesirable users, such as spam-bots, are filtered by a security question during the registration form. Finally, all users have a limited disk quota which disable job submissions when exceeded.

Computation engine and server deployment

In parallel to the Web development, the most important component of `APFEL Web` is the computational engine that we called `apfeldaemon`. The program is a generalization of the open source `APFEL GUI` code in `C++` with the inclusion of the database I/O procedures. The job configuration and the PDF grids are read from the database, and the computation is performed upon request by the user. In order to solve the problem correlated with the usage of two different interfaces to PDF grids, i.e. `LHAPDF5` and `LHAPDF6`, the `apfeldaemon` is composed by two binaries which are linked to the respective libraries: the Web-application checks the PDF grid version and it starts the computation procedure with the correct daemon.

In Figure 2.7 we show the scheme of the Web-application structure. Users from Web browsers send requests to a `Python` server which in our case is implemented by `gunicorn` and `nginx`[5]. The `Python` server performs the request using the `Django`

[5] `http://gunicorn.org` and `http://nginx.org/`

framework, at this level PDF objects and jobs are prepared and saved in the database, additionally eventual job results are collected in a dedicated view. From the computational point of view the layout is very simple and clearly illustrated by the left side of Fig. 2.7. We have set up a Portable Batch System (PBS)[6] for the multi-core server which receives job submissions and is able to automatically handle the job queue, avoiding the unpleasant situation of server overloading. Jobs are submitted by the `Django` application which passes the job identification number, this value is read by the `apfeldaemon` and it performs a query at the corresponding database entry, then it collects the relevant information to start the correct job. When the job finalizes the `apfeldaemon` modifies the job status in the database, so the Web interface notifies the user of the job status.

The `apfeldaemon` program was designed and compiled with performance as priority, in fact there are relevant computational speed improvements when comparing to the previous `APFEL GUI` program almost due to the clear separation between the `GUI` and the calculation engine. In order to provide to the reader an idea of the typical processing time per job, we estimate that one job requires two seconds to process a single PDF set when producing a PDF comparison plots, meanwhile for luminosity and observables jobs, the system takes up to one minute per PDF set when including the uncertainty treatment.

2.5.2 Plotting tools

While the use of the Web-interface should be self-explanatory, here we describe and show examples of job results that a user is able to obtain from `APFEL Web`.

The first step consists in the creation of custom "PDF objects" in the user's workspace. The following points explain how to create such objects:

1. select a PDF grid from the `LHAPDF5` and/or `LHAPDF6` libraries and determine the treatment of the PDF uncertainty among: no error, Monte Carlo approach, Hessian eigenvectors (68 and 90% c.l.) and symmetric eigenvectors. When selecting a PDF set the system proposes automatically an uncertainty type based on the PDF collaboration name.

2. import a new `LHAPDF` grid file, with the only requirement that it is provided either in the `LHAPDF5 LHgrid` or in the `LHAPDF6` format. The main target for this feature are the members of the PDF collaborations which can perform comparisons with preliminary sets of PDFs before the publication in `LHAPDF`.

3. set the evolution library by choosing between the `LHAPDF` interpolation routines or the `APFEL` custom evolution.

We provide the following computational functions, which are illustrated in Figures 2.8 and 2.9:

- "`Plot PDF Members`": it plots for projections in x all the members of a PDF set for a single parton flavor at a given energy scale Q. See the top-left image in Fig. 2.8 where we show the replicas of `NNPDF2.3 NLO` [11] together with

[6] An example of PBS open source implementation: http://www.adaptivecomputing.com/

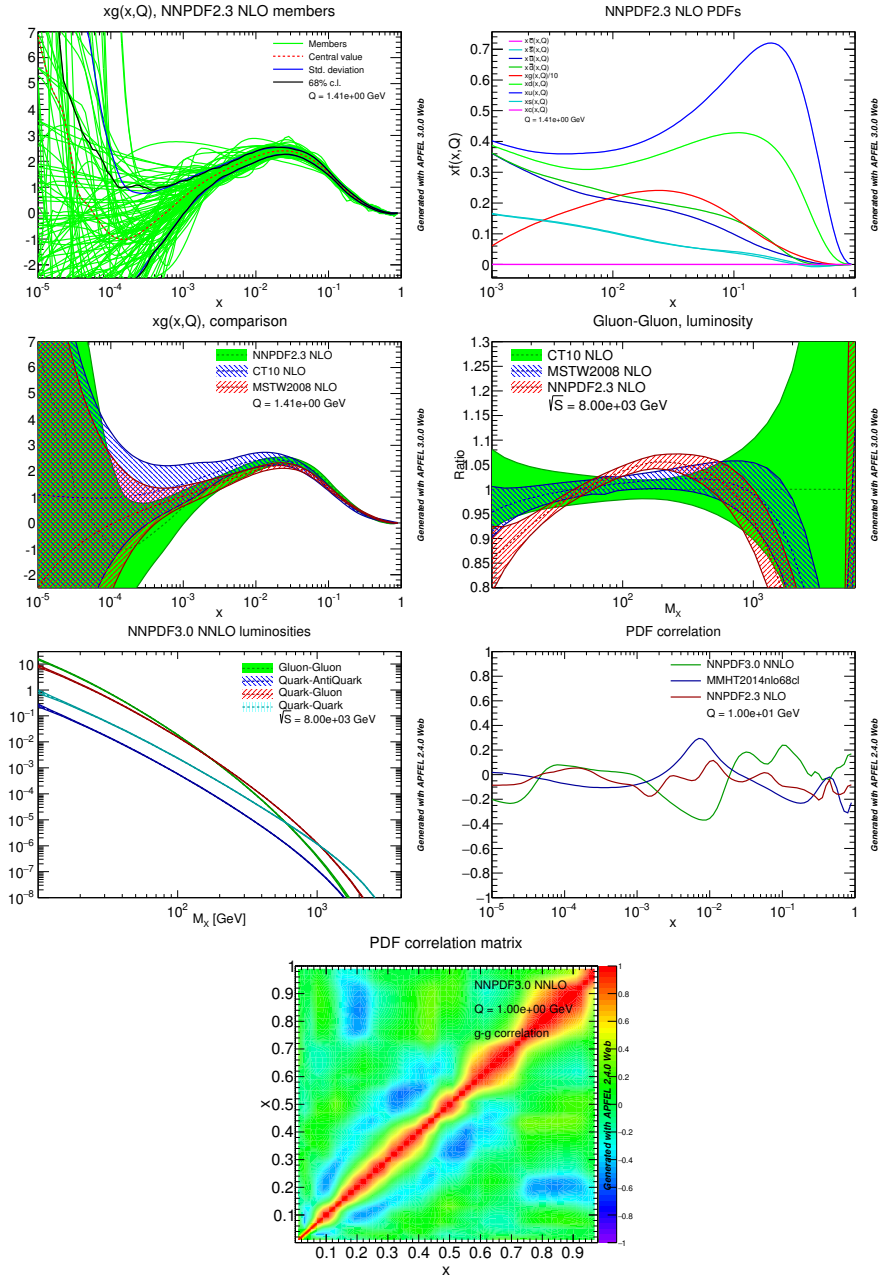

Figure 2.8: Examples of output generated with `APFEL Web`. Plots are presented in the following order, clockwise from top-left: PDF members, multiple PDF flavors, PDF comparison in x, gg-channel luminosity, all luminosities, PDF correlations and correlation matrix.

its central value and Monte Carlo uncertainty band, these last features are options which can be disabled by the user. This plotting tool accepts only a single PDF set at each time in order to avoid too many information in a single plot. We provide the possibility to choose between the usual parton flavors, i.e. $b, t, c, s, d, u, g, \gamma, q_i^{\pm} = q_i \pm \bar{q}_i$ with $q_i = u, d, s, c, b, t$, and the combination of them: $(\Sigma, V, V_3, V_{15}, V_{24}, V_{35}, T_3, T_{15}, T_{24}, T_{35}, \Delta_s)$ [44], the so called evolution basis.

- "`Plot multiple PDF flavors`": each PDF flavor is plotted together in the same canvas at a fixed energy scale. We also provide the possibility to scale PDF flavors by a predetermined numeric factor in order to produce plots similar to the PDG [129]. An example of PDF flavor plot is presented in the top-right of Fig. 2.8 where the gluon PDF is scaled by a factor 10.

- "`Compare PDFs in` x": this tool compares the same flavor of multiple PDF sets and the respective uncertainties at a given energy scale for projections in x. We provide the possibility to compute the absolute value or the just the ratio respect to a reference PDF set. The second row left image of Fig. 2.8 shows the comparison between `NNPDF2.3 NLO` [11], `CT10 NLO` [34] and `MSTW2008 NLO` [39] sets at $Q = 1$ GeV.

- "`Compare PDFs in` Q": this tool compares the same flavor of multiple PDF sets and the respective uncertainties at a fixed x-value as a function of the energy scale Q.

- "`Compare PDF Luminosity`": it performs the computation of parton luminosities [130] normalized to a reference PDF set at a given center of mass energy. There are several channels available: gg, $q\bar{q}$, qg, cg, bg, qq, $c\bar{c}$, $b\bar{b}$, $\gamma\gamma$, γg, etc. In the second row right plot of Fig. 2.8 we show an example of gg-luminosity at $\sqrt{s} = 8$ TeV using the PDF sets presented above with `CT10 NLO` as reference PDF set.

- "`All PDF Luminosities`": for a given set of PDFs this tool compares the gg, $q\bar{q}$, qg and qq luminosities in a single plot. The third row left image of Fig. 2.8 shows an example of the output for `NNPDF3.0 NLO` at $\sqrt{s} = 8$ TeV.

- "`Compare PDF Correlations`": it performs the comparison of PDF correlations between pairs of PDFs flavors for multiple sets of PDFs. The correlation coefficients are obtained through the `LHAPDF6` interface. The third row right image of Fig. 2.8 shows an example of the output for this plotting tool.

- "`PDF Correlations Matrix`": for a given set of PDFs this tool computes the correlation matrix for pairs of PDF flavors in a grid of x-points. The correlation coefficients are computed automatically through the `LHAPDF6` interface. An example of such tool is shown in the bottom image of Fig. 2.8.

- "`DIS in` x`/DIS in` Q": it computes DIS observables as functions of x or Q for different heavy quark schemes and perturbative orders, including the Fixed Flavor Number scheme (FFNS), the Zero Mass Variable Number scheme (ZMVN),

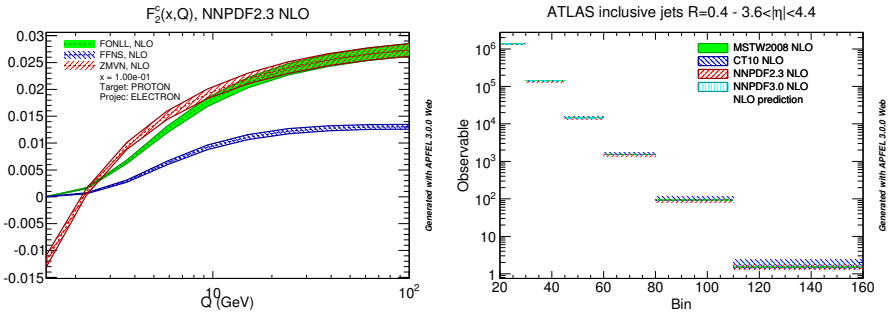

Figure 2.9: On the left, an example of DIS observable computed by `APFEL Web`: $F_2^c(x,Q)$. On the right, another example about the `APPLgrid` observables tool used for the computation of predictions for ATLAS 2010 inclusive jets $R = 0.4$ at $\sqrt{s} = 7$ TeV [132].

and the FONLL scheme [41] where the choice of a NLO prediction implies using the FONLL-A scheme, while choosing NNLO leads to using the FONLL-C scheme. A detailed explanation of all possible configurations is presented in Sect. 4.3 of Ref. [9]. An example of such tool is presented in the left plot of Fig. 2.9.

- "`APPLgrid observables`": this tool provides a simple a fast interface to theoretical predictions through the `APPLgrid` library [128]. The system already provides several grids that are available from the `APPLgrid` website[7] but also from the NNPDF collaboration [12] and `aMCfast` [131]. This function allows users to compute the central value and the respective uncertainties for multiple PDF sets. On the right plot of Fig. 2.9 we show the output of this tool for the predictions of ATLAS 2010 inclusive jets $R = 0.4$ at $\sqrt{s} = 7$ TeV [132].

For all the tools presented above, the Web interface provides options for customizing the graphics, like setting the plot title, axis ranges, axis titles and curve colors. `APFEL Web` also provides the possibility to save plots and the associated data in multiple formats, including: PNG, EPS, PDF, .C (`ROOT`) and .root (`ROOT`).

Finally, it is important to highlight that the results produced by `APFEL Web` for PDF comparison and parton luminosities from different PDF sets have been verified against the PDF benchmarking exercise of Ref. [8].

2.5.3 Usage statistics

The `APFEL Web` application was released on October 7, 2014. Five months after the release we already have 131 registered users from 20 countries, and an average of 258 visits each month. In Figure 2.10 we show a pictorial representation of the total unique visits by country during the period between the release date to March 2015.

At the current date, the server has successfully completed more than 3500 jobs. In the left plot of Figure 2.11 the distribution of jobs selected by the users is shown in

[7] http://applgrid.hepforge.org/

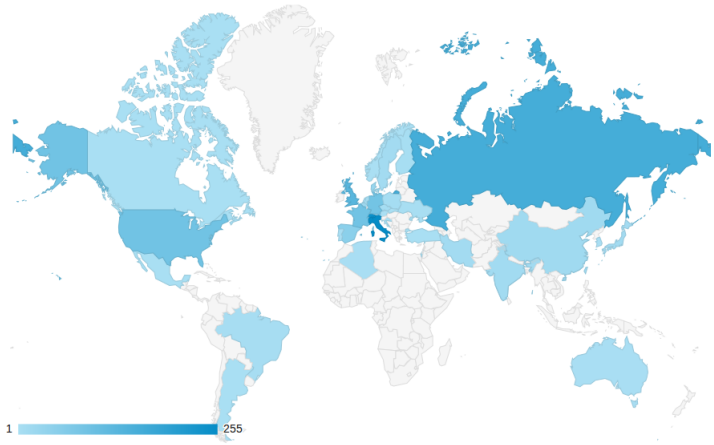

Figure 2.10: Unique sessions by country from October 2014 to March 2015 (1293 visits).

percentages. The PDF comparison and luminosity are the most popular tools, followed by PDF members and all PDF flavors plots. The right plot of Figure 2.11 presents a pie chart with the country affiliation of users registered in the `APFEL` website. Top users are from Switzerland (mainly from CERN), UK, USA followed by users spread across all continents. These results, obtained in a relative short time period, are rewarding showing that there is an international community of physicist interested in the features provided by `APFEL Web`.

Finally, in Figure 2.12 we show the local time of the day preferred by users for the submission of jobs. The polar axis shows the time of day, meanwhile the radius the total number of job submissions. There are two peaks of activity, the first at 12am and the second at 6pm. Furthermore we observe a continuous operation cycle from 9am to 10pm. Possibly, these results can be interpreted as another advantage of having an online server interface, accessible from any device connected to internet at any time.

Thanks to its flexibility and user-friendliness, we believe that in the coming months and years `APFEL Web` has the potential to become a widely used tool worldwide.

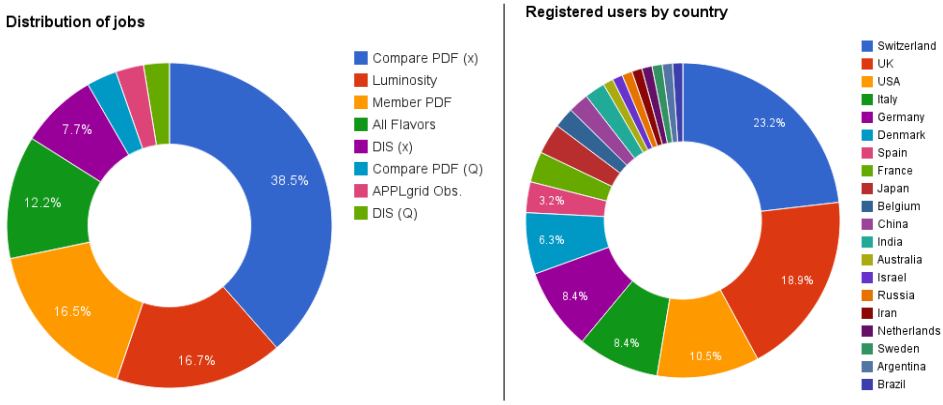

Figure 2.11: On the left plot, the distribution of the plotting tools selected by the users when sending jobs. On the right plot, the fraction of the registered users organized by country. In both cases, the legend elements are organized in descending order. The results refer for the period from October 2014 to March 2015.

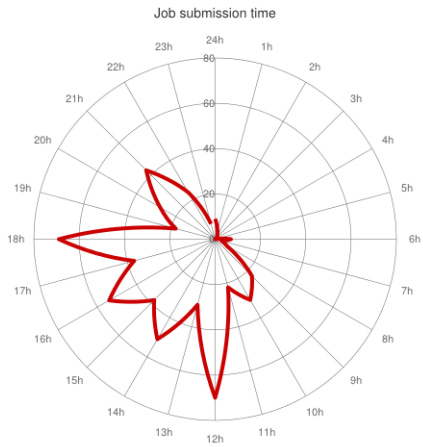

Figure 2.12: Number of jobs submitted in function of the local time of day.

Chapter 3

The NNPDF methodology

In this chapter we present the Neural Network Parton Distribution Function (NNPDF) methodology. We provide an overview of the NNPDF approach to PDFs, which is then used in Chapter 4 for the determination of a set of PDFs with QED corrections. In Sect. 3.1 we begin with the description of the NNPDF methodology which is then followed in Sect. 3.2 by a technical presentation of the new modern framework developed specifically for this project and for the next generation of NNPDF fits. In Sect. 3.3, we conclude the chapter with the characterization of the NNPDF2.3 set of PDFs, which is the baseline configuration used in the fit with QED corrections.

3.1 Introduction to NNPDF

The NNPDF Collaboration is the only group which implements the Monte Carlo approach to a global fit of PDFs instead of the usual Hessian method. The goal of this strategy is to provide an unbiased determination of PDFs with reliable uncertainty. The approach implemented in NNPDF is based on advanced computational techniques, such as:

- The Monte Carlo treatment of experimental data.
- The parametrization of PDFs with artificial neural networks.
- The minimization strategy based on Genetic Algorithm.

In an initial step, the original experimental data is transformed into a Monte Carlo ensemble of replicas. In this procedure, the ensemble of artificial data replicas follows a multi-Gaussian distribution centered around the central value of each data point and with the variance based on the statistical, systematic and normalization uncertainties, encoded in the experimental covariance matrix. The total number of replicas is selected in such a way that it is large enough to produce the statistical properties of the original data to the desired accuracy.

Each of the Monte Carlo data replica is then fitted by PDFs parametrized with artificial neural networks (ANN). The use of ANNs instead of selecting a specific functional forms, *e.g.* based on polynomial, guarantees no bias due to the parametrization. In fact, neural networks with large architectures are able to imitate the behavior of any functional form.

The last stage of the NNPDF methodology is the fitting strategy. As in any other fitting procedure, we define a figure of merit which compares the theoretical predictions of physical observables, obtained through the convolution of PDFs, to the respective data replica. In this case, as we are dealing with a large number of parameters, the ANNs are trained by a Genetic Algorithm (GA) which shows a good performance in comparison to algorithms based on Newton's methods.

The Monte Carlo representation of the underlying probability density associated to a given set of PDFs has several advantages as compared with the traditional Hessian approach. The most important advantage of the MC method is that it does not require the selection of a fixed functional form. This feature lets discard any bias associated to the PDF parametrization. Moreover, it also does not assume that the underlying PDF uncertainties are Gaussian, as the Hessian method does, and so, it does not rely on the linear approximation to propagate uncertainties from the original data to the PDFs. Technical details about each of the previous points will be addressed in the next section.

3.2 A modern implemention of the NNPDF framework

We show the details of the NNPDF methodology from the point of view of the implementation of a new code framework. The main motivation for updating the NNPDF code resides on the need for flexibility and performance. There are several advantages in reformulating the methodology in a modern object-oriented approach. First of all, the possibility to have more expressiveness, which allows the inheritance of data structures, introducing layers of abstraction between several components of the code. From the NNPDF practical point of view, this strategy is translated by a huge simplification of the framework, where data, theory and fitting are completely independent elements, which can be easily extended and optimized. These technical advantages reflect an easy a fast development of specific projects, for example the QED corrections to PDFs, as presented here, the determination of Nuclear PDFs and Fragmentation Functions [133].

On the other hand, with the current inclusion of a substantial number of LHC datasets in a global PDF determination, we face performance issues due to the complexity of adding new hadronic observables into the fitting framework. These issues reflect an increasingly computational cost of running fits. This trend is supposed to grow in the next years, due to future new LHC measurements. The main cause of these performance issues resides on the NNPDF computationally intensive Genetic Algorithm minimization. So, in order to deal with such problems, we have developed a modern fitting code based on two object-oriented languages: `C++` and `Python`. This choice, as already mentioned before, allows the inclusion of new datasets achieving a highly efficient implementation of the minimization algorithms which is not possible to achieve in the previous `Fortran77` implementation.

In what follows we describe the technical choices and code structure of the new code framework through the description of the NNPDF methodology. In this thesis, we focus on the NNPDF2.3 setup because the QED corrections have been applied to this fitting configuration using a preliminary version of the updated framework. However, note that the NNPDF Collaboration have recently presented a new set of PDFs, the NNPDF3.0 [12], where the new framework is used by default.

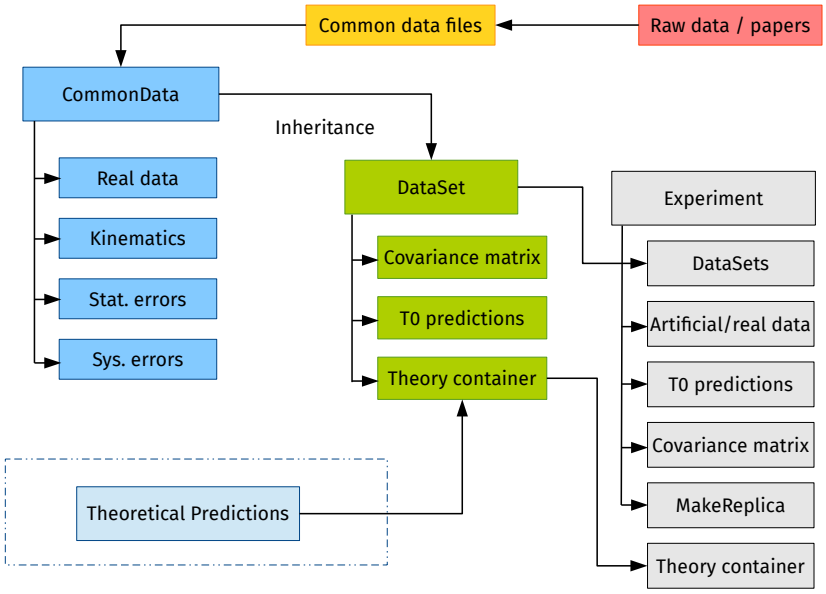

Figure 3.1: Pictorial representation of the NNPDF data management layout. The dashed blue box indicates a simplification in the diagram.

3.2.1 Data treatment

The implementation of the Monte Carlo artificial data generation starts from the construction of the experimental covariance matrix. For each experiment, the current framework first groups together the respective datasets, in order to take into account eventual cross-correlations, and then creates the final covariance matrix. For a given experiment let us consider the measurement of two observables O_I and O_J, so, the experimental covariance matrix reads

$$\mathrm{cov}_{ij} = O_{I,i} O_{J,j} \left(\sum_{l=1}^{N_c} \sigma_{i,l}\sigma_{j,l} + \sum_{n=1}^{N_a} \sigma_{i,n}\sigma_{j,n} + \sum_{n=1}^{N_r} \sigma_{i,n}\sigma_{j,n} + \delta_{ij}\sigma_{i,s}^2 \right) , \tag{3.1}$$

where i and j run over the experimental points, and the various uncertainties given as relative values, are:

- $\sigma_{i,l}$, the N_c correlated systematic uncertainties,
- $\sigma_{i,n}$ the N_a absolute and N_r relative normalization uncertainties,
- $\sigma_{i,s}$, the statistical uncertainty.

Before defining the artificial replica generation we introduce the total uncertainty for the $i-$th point, in terms of

$$\sigma_{i,\mathrm{tot}} = \sqrt{\sigma_{i,s}^2 + \sigma_{i,c}^2 + \sigma_{i,N}^2} , \tag{3.2}$$

where $\sigma^2_{i,c}$ and $\sigma^2_{i,N}$ are respectively the total correlated and the total normalization uncertainties defined as

$$\sigma^2_{i,c} = \sum_{l=1}^{N_c} \sigma^2_{i,l} \,, \tag{3.3}$$

and

$$\sigma^2_{i,N} = \sum_{n=1}^{N_a} \sigma^2_{i,n} + \sum_{n=1}^{N_r} \left(\frac{1}{2}\sigma_{i,n}\right)^2 . \tag{3.4}$$

Note that in Eq. (3.1) we have introduced the definition of the experimental covariance matrix, however in a real fit, such matrix is replaced by the so called t_0 covariance matrix where the observables O_I are extracted from predictions obtained with a prior set of PDFs, rater than the original data, avoiding the known bias presented in Ref. [134].

At this stage, we generate $k = 1, \ldots, N_{\rm rep}$ artificial replicas of the original data points by shifting with a multi-Gaussian distribution defined as

$$O^{({\rm art})(k)}_{I,i} = O_{I,i} \left(1 + \sum_{l=1}^{N_c} r^{(k)}_{i,l}\sigma_{i,l} + r^{(k)}_i \sigma_{i,s}\right) \prod_{n=1}^{N_a} \left(1 + r^{(k)}_{i,n}\sigma_{i,n}\right) \prod_{n=1}^{N_r} \sqrt{1 + r^{(k)}_{i,n}\sigma_{i,n}} \,, \tag{3.5}$$

where the univariate Gaussian random numbers, $r^{(k)}_{i,l}, r^{(k)}_i, r^{(k)}_{i,n}$, generate fluctuations of the artificial data around the central value given by the experiments. For each replica k, if two experimental data points have correlated systematics or normalization uncertainties, then the fluctuations associated to such uncertainty are taken the same for both points.

In Figure 3.1 we show a simplified picture of the code structure used for the manipulation of data and the generation of MC artificial replicas. Experimental data is stored in files with a common layout, which contains the process type information, the experimental kinematics for each data point, the experimental central values, the full breakdown of experimental systematic uncertainties and the choice of additive/multiplicative treatment of systematic uncertainties. These files are obtained from the conversion of raw data information extracted directly from publications of experimental collaborations. From a programatically point of view, this information is read from the common data files when the `CommonData` container is initialized and allocated in memory. From `CommonData` we have created the inherited `DataSet` class which implements the covariance matrix using both definitions: experimental and t_0. This class also loads in memory the associated theoretical prediction model which will be discussed in details in Sect. 3.2.3. Note that the information contained in `DataSet` is not used directly in the fit. The final element of the data layout is the `Experiment` class, which groups together datasets from the same experiment, constructs the covariance matrix taking into account eventual cross-correlations, and provide the algorithm for generating the MC artificial replicas given by Eq. (3.5). This class is used directly in the fit of PDFs, and it is easily generalized for any kind of experimental data.

3.2.2 PDF parametrization

Concerning the PDF parametrization, the artificial neural networks used in NNPDF fits consist of connected nodes organized in layers. In order to evaluate the network,

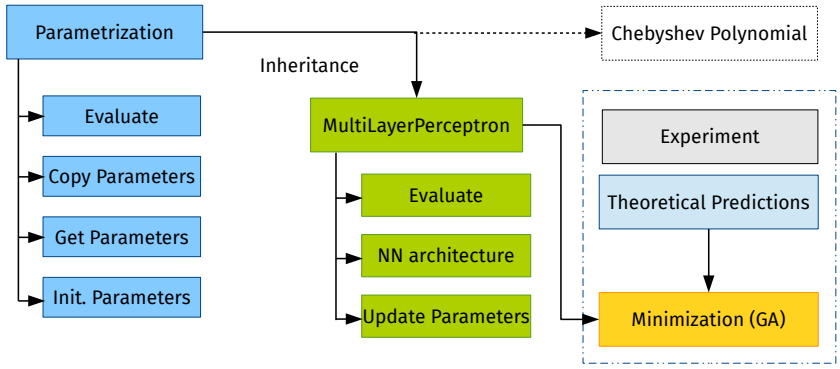

Figure 3.2: Pictorial representation of the NNPDF parametrization layout.

the nodes in the input layer are set with the required x and $\log x$ values and then the activation of nodes in successive layers are calculated according to

$$\xi_i^{(l)} = g\left(\sum_j w_{ij}^{(l)} \xi_j^{(l-1)} + \theta_i^l\right) \tag{3.6}$$

$$g(a) = \frac{1}{1+e^{-a}} \tag{3.7}$$

where $\xi_i^{(l)}$ is the activation of the i-th node in the l-th layer of the network, $w_{ij}^{(l)}$ are the weights from that node to the nodes in the previous layer and θ_i^l is the threshold for that node. The weights and the thresholds are the parameters in the fit which are changed during the Genetic Algorithm minimization. This implementation is known as a multi-layer feed-forward neural network model (MLP). There is an exception to Eq. (3.7) in the last layer, where in order to allow for an unbounded output a linear activation function $g(a) = a$ is used instead. The flexibility of the fitting code allows us to easily explore other choices, for instance a quadratic output of the last layer, $g(a) = a^2$, has been used in studies of the PDF positivity in leading order fits, including special configurations where only a single PDF flavors is positive defined, *e.g.* the photon PDF (cfr. Chap. 4).

In Figure 3.2 we present the parametrization layout implemented in the new framework. An abstract container, called `Parametrization` implements *virtual* methods for the evaluation and manipulation of parameters for a generic input function. From this class we can inherit different functions, in particular, for the NNPDF methodology we have implemented the neural networks of Eq. (3.7) in the `MultiLayerPerceptron` container. In the diagram we show a dashed line with another example of parametrization, the Chebyshev polynomial. This container provides methods for the evaluation, and modification of weights and thresholds of a given ANN architecture by the minimization algorithm. This new framework also provides several features such as the

possibility to choose the input scale, the parametrization basis, preprocessing, and the implementation of PDF positivity.

Parametrization basis

In the NNPDF fits, PDFs are parametrized at a reference scale Q_0^2. The choice of Q_0^2 has no effect whatsoever on the results of the fit because the DGLAP evolution evolves the input parametrization from the initial scale to the energy of the experimental data point. PDFs are expressed in terms of a set of basis functions for quark, antiquark and gluon PDFs already introduced in Chap. 1. For the NNPDF2.3 we define the following basis:

$$\begin{aligned}
\Sigma(x,Q_0^2) &= \left(u+\bar{u}+d+\bar{d}+s+\bar{s}\right)(x,Q_0^2)\\
T_3(x,Q_0^2) &= \left(u+\bar{u}-d-\bar{d}\right)(x,Q_0^2)\\
V(x,Q_0^2) &= \left(u-\bar{u}+d-\bar{d}+s-\bar{s}\right)(x,Q_0^2)\\
\Delta_S(x,Q_0^2) &= \left(\bar{d}-\bar{u}\right)(x,Q_0^2)\\
s^+(x,Q_0^2) &= \left(s+\bar{s}\right)(x,Q_0^2)\\
s^-(x,Q_0^2) &= \left(s-\bar{s}\right)(x,Q_0^2)\\
g(x,Q_0^2)\,. &
\end{aligned} \tag{3.8}$$

In the PDF basis above we do not introduce an independent parametrization for the charm and anticharm PDFs (intrinsic charm), however the new framework provides the possibility to easily activate any combination or flavor parametrization.

This basis was chosen in NNPDF2.3 because it directly relates physical observables to PDFs, by making the leading order expression of some physical observables in terms of the basis functions particularly simple: for example, T_3 is directly related to the difference in proton and deuteron deep-inelastic structure functions $F_2^p - F_2^d$, and Δ_S is simply expressed in terms of Drell-Yan production in proton-proton and proton-deuteron collisions, for which there is data for example from the E866 experiment. On the other hand, with the current code we can show that several other basis choices does not affect the results: our results are independent of the basis change, as recently presented in details in the NNPDF3.0 paper [12].

Each PDF is then parametrized by the ANN of Eq. (3.7) with architecture 2-5-3-1 at the reference scale Q_0^2 times a preprocessing factor:

$$f_i(x,Q_0) = A_i \hat{f}_i(x,Q_0^2); \quad \hat{f}_i(x,Q_0^2) = x^{-\alpha_i}(1-x)^{\beta_i}\,\mathrm{NN}_i(x) \tag{3.9}$$

where A_i is an overall normalization constant, and f_i and $\hat{f}_i$ denote the normalized and un-normalized PDF respectively. The preprocessing term $x^{-\alpha_i}(1-x)^{\beta_i}$ is simply there to speed up the minimization, without biasing the fit. In the case of the s^- parametrization we introduce an auxiliary term such as

$$s^-(x,Q_0^2) = A_{s^-}\hat{s}^-(x,Q_0^2) - s_{\mathrm{aux}}(x,Q_0^2)\,, \tag{3.10}$$

where $s_{\mathrm{aux}}(x,Q_0^2) = A_{s^-}x^{-\gamma_{s^-}}(1-x)^{\delta_{s^-}}$, with exponents chosen in such a way that $s_{\mathrm{aux}}(x,Q_0^2)$ peaks in the valence region, not interfering with the small-x and large-x behavior of $s^-(x,Q_0^2)$.

Out of the seven normalization constants, A_i in Eq. (3.9), three can be constrained by the valence sum rules, sea asymmetry and the momentum sum rule. Which particular combinations depends of course of the choice of basis. With the basis, Eq. (3.9), these constraints lead to

$$A_g = \frac{1-\int_0^1 dx x\Sigma(x,Q_0^2)}{\int_0^1 dx\, x\, \hat{g}(x,Q_0^2)}; \quad A_V = \frac{3}{\int_0^1 dx\, \hat{V}(x,Q_0^2)}; \quad A_{\Delta_S} = \frac{1-\int_0^1 dx\, \hat{T}_3(x,Q_0^2)}{\int_0^1 dx\, 2\hat{\Delta}_S(x,Q_0^2)}\,. \tag{3.11}$$

The other normalization constants can be set arbitrarily to unity, that is $A_\Sigma = A_{T_3} = A_{s^-} = A_{s^+} = 1$: the overall size of these PDFs is then determined by the size of the fitted network. The finiteness of sum rule integrals Eq. (3.11) is enforced by discarding during the Genetic Algorithm minimization (see Sect. 3.2.4 below) any mutation for which the integrals would diverge. This condition, in particular, takes care of those NN configurations that lead to a too singular behavior at small-x.

Effective preprocessing exponents

We have introduced in Eq. (3.9) the preprocessing concept which absorbs in a prefactor the bulk of the fitted behavior so that ANN only has to fit deviations from it. This choice is motivated by a performance improvement during the fit. However, it is important to implement an automatic mechanism that performs the choice of these coefficients without biasing the result. As in previous NNPDF fits, this is done by randomizing the preprocessing exponents, choosing a different value for each replica within a suitable range. We first define the effective asymptotic exponents as follows:

$$\alpha_{\text{eff},i}(x) = \frac{\ln f_i(x)}{\ln 1/x}\,, \quad \beta_{\text{eff},i}(x) = \frac{\ln f_i(x)}{\ln(1-x)}\,. \tag{3.12}$$

Then, we perform a fit where the algorithm chooses a random set of coefficients between a wide starting range for the preprocessing exponents for each PDF. The effective exponents Eq. (3.12) are then computed for all replicas at $x = 10^{-6}$ and 10^{-3} for the low-x exponent α_i and at $x = 0.95$ and 0.65 for the large-x exponent β_i, for all PDFs (except for the gluon and singlet small-x exponent, α_i, which is computed at $x = 10^{-6}$). The fit is then repeated by taking as new range for each exponent the envelope of twice the 68% confidence interval for each x value. The process is then iterated until convergence, with a tolerance of few percent. From a practical point of view, the convergence i typically fast, even in the cases where the fitted dataset is varied significantly or for example when the minimization algorithm is modified.

This procedure ensures that the final effective exponents are well within the range of variation both in the region of the smallest and largest x data points, and in the asymptotic region (these two regions coincide for the gluon and singlet at small x), thereby ensuring that the allowed range of effective exponents is not artificially reduced by the preprocessing, either asymptotically or sub-asymptotically.

3.2.3 Theoretical predictions

As we have anticipated at the beginning of this section, the most computationally intensive task for the PDF fitting technology is the computation of theoretical predictions. Indeed, any PDF determination involves an iterative procedure where all the

data points included in the fit need to be recomputed a very large number of times for different functional forms of the input PDFs. The computation of physical observables in the NNPDF framework is based upon the `FastKernel` method introduced in Refs. [11,130]. Here we recall the basic concepts necessary to explain the structure of the new code.

The FastKernel methodology

Let us consider a grid of points in x, where each PDF flavor at a given scale Q^2 is represented in terms of $f_i(x_\alpha, Q^2)$ with $\alpha = 1, \ldots, N_x$ where the index i identifies the parton flavor, and the index α enumerates the points on the grid. DIS observables, which are linear in the PDFs, can be computed using a precomputed kernel $\hat{\sigma}^{I,J}_{\alpha j}$:

$$O_I(x_J, Q_J^2) = \sum_{j=1}^{N_{\rm pdf}} \sum_{\alpha=1}^{N_x} \hat{\sigma}^{I,J}_{\alpha j} f_j(x_\alpha, Q_0^2)\,, \tag{3.13}$$

where the index I labels the physical observable, x_J and Q_J are the corresponding kinematical variables for each specific experimental data point J, j runs over the parton flavors and α runs over the x-grid points. The kernel $\hat{\sigma}$ just introduced is referred to as an `FKTable`. A similar expression is available for the hadronic observables, which are written as a convolution of two PDFs, and computed in terms of an (hadronic) `FKTable` $\hat{W}^{I,J}_{kl\gamma\delta}$:

$$O_I(x_J, Q_J^2) = \sum_{k,l=1}^{N_{\rm pdf}} \sum_{\gamma,\delta=1}^{N_x} \hat{W}^{I,J}_{kl\gamma\delta} f_k(x_\gamma, Q_0^2) f_l(x_\delta, Q_0^2)\,, \tag{3.14}$$

where the indices k, l run over the parton flavors, and the indices γ, δ count the points on the interpolating grids.

In the fitting code, for each experimental dataset I we have a separate `FKTable` that encodes all the theory information. In Figure 3.3 we show the components encoded in a `FKTable` file. These tables encode all the information about the theoretical description of the observables such as: the perturbative order, the value of the strong coupling, the choice of scales, the QCD and electroweak perturbative corrections (C-factors), or the prescription for the evolution. The modification of any of the of theoretical description of a given observable is reflected in a new `FKtable`. The convolutions of the `FastKernel` tables with the PDFs at the initial scale are generic, and do not require any knowledge about the theoretical framework. On the other hand, the tables also contain information about the grid of points in x used for the interpolation and the so-called flavor map which optimizes the grids size by indicating all the available non-zero flavor channels. Notice that this layout implements a clean separation of the theoretical assumption from the fitting procedure. In particular, during the fitting procedure the tables are always kept fixed and treated as an external input. The only shared information between these tables and fit is the initial scale Q^2.

One important remark about the differences between the `FastKernel` approach in comparison to fast NLO calculators such as `FastNLO` [135], `APPLgrid` [128] and `aMCfast` [131], is that it includes PDF evolution into the precomputed tables, while the other approaches require as input the PDFs evolved at the scales where experimental

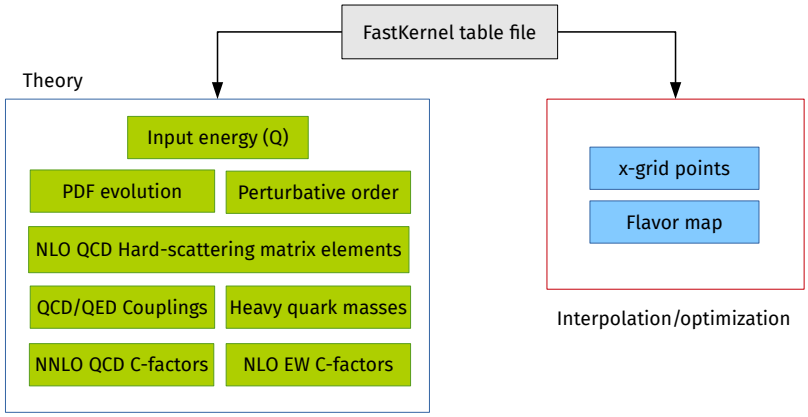

Figure 3.3: Graphical summary of the `FKTable` layout.

Observable	`APPLgrid`	`FKTable`	optimized `FKTable`
W^+ production	1.03 ms	0.41 ms (2.5x)	0.32 ms (3.2x)
Inclusive jet production	2.45 ms	20.1 μs (120x)	6.57 μs (370x)

Table 3.1: Comparison of `APPLgrid` and `FKTable` convolution timings. Results are provided for two different observables: the total cross-section for W^+ production and for inclusive jet production for typical cuts of p_T and rapidity. In parenthesis we show the relative speed-up compared to the the reference convolution based on `APPLgrid`. In the last column we use SSE acceleration in the convolution product.

data is provided. The inclusion of PDF evolution is essential to reduce drastically the computational cost of running PDF fits. Note also that the generic structure of the `FastKernel` methodology holds for any fast NLO calculator as well as for any PDF evolution code. For example, in NNPDF2.3 and later we use our own internal Mellin-space `FKgenerator` code for PDF evolution and DIS observables. A future version of this combination, planned for the next NNPDF release, combines the x-space evolution from `APFEL` with the usual `FastKernel` combination algorithm (the so-called `APFELcomb` project). This shows how flexible the code is: the `FastKernel` tables are independent elements from the NNPDF framework, which can be computed with external tools, specialized in the computation of theoretical predictions.

The main advantage of the `FastKernel` methodology in comparison to *e.g.* `APPLgrid` or `FastNLO` is that PDF evolution is precomputed and stored in the `FKTable` itself. This point is particularly relevant when performing a fit to data distributed along a large range of Q^2 values, *e.g.* the inclusive jet production, where an equivalent large number of PDF evolutions are needed. In these cases, the inclusion of PDF evolution improves drastically the performance of the fit. The improvement due to this acceleration is quantified in Table 3.1.

The code layout for the `FastKernel` procedure is presented in Figure 3.4. The

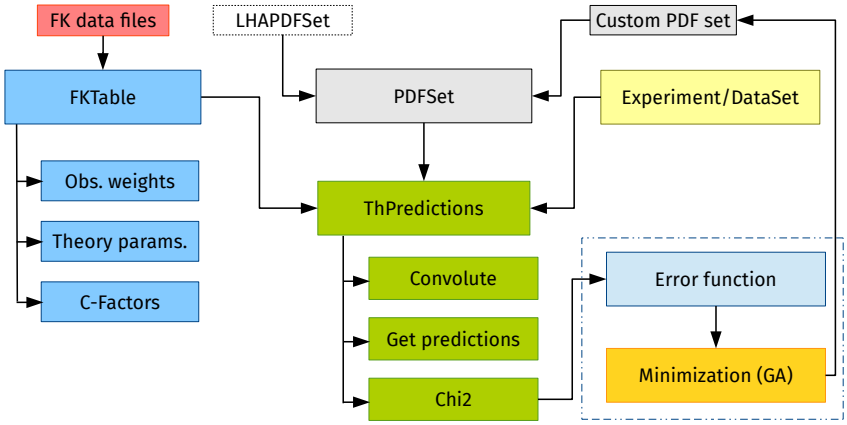

Figure 3.4: Pictorial representation of the NNPDF theoretical predictions framework.

`FKTable` class reads the `FastKernel` objects from files stored on disk. For each dataset, this class makes available the convolution kernel, the theoretical setup and the eventual C-factors to other modules of the code. As we have explained previously, the new fitting code has been designed with an explicit separation between experiment and theory. Therefore, the kinematic cuts upon an experimental dataset can now be performed algorithmically by selecting the points in the `CommonData` format which pass the required cuts according to their bundled kinematic information, and matching with the equivalent points in the `FKTable`. This is a considerable improvement over the earlier regeneration of the precomputed theory tables due to the monolithic treatment of the experimental data in the `Fortran77` code. Note that this layout allows the introduction of PDF positivity constrains through the convolution of PDFs with artificial observables encoded in `FastKernel` tables, which are tested during the minimization algorithm and in the case of violation it penalizes the error function.

The PDF convolution is performed in the `ThPredictions` class, which takes as input: a PDF set through the abstract `PDFSet` class and a `FKTable` object, which can be passed automatically from the `DataSet` and `Experiment` classes. PDFs are accessible through the `LHAPDFSet` interface, which calls PDFs from the `LHAPDF` library, or by any other custom set obtained by extending the `PDFSet` class, this is exactly what the minimization algorithm does. The `ThPredictions` object provides methods for the `FastKernel` convolution product. This class computes theoretical predictions but also determines the χ^2 to data when used in combination with `Experiment/DataSet`.

Concerning optimizations, in order to ensure a fast and efficient minimization procedure, the `FKTable` class has been designed such that the `FastKernel` table is stored with the optimal alignment in machine memory for use with `SIMD` (Single Instruction Multiple Data) instructions, which allow for an acceleration of the observable calculation by performing multiple numerical operations simultaneously. The large size of a typical `FastKernel` product makes the careful memory alignment of the `FastKernel` table and PDFs extremely beneficial. A number of `SIMD` instruction sets are available

depending on the individual processor. By default we use a 16-byte memory alignment for suitability with Streaming SIMD Extensions (SSE) instructions, although this can be modified by a parameter to 32-bytes for use with processors enabled with Advanced Vector Extensions (AVX). The product itself is performed both with SIMD instructions and, where available, OpenMP is used to provide acceleration using multiple CPU cores, parallelizing the computation of predictions for each experimental data point. We have also investigated about a further level of improvement of the FastKernel product by using GPUs, while presenting no technical objections, has so far not been developed due to scalability concerns on available computing clusters. Moreover, several technologies such as NVIDIA CUDA[1] or OpenCL[2] show optimal performance only on dedicated devices, disfavoring portability.

The performance improvements are clearly visible when comparing with the calculation of the hadronic convolution Eq. (3.14) using the optimized settings with that using non-optimized settings. To illustrate this point, we compare in Table 3.1 the timings for a couple of representative LHC observables, for the convolution performed using APPLgrid, the standard double-precision version of the FKTable implementation, and the optimized FKTable implementation using the SSE-accelerated calculation, for two representative observables. The results shows a massive improvement in speed by precomputing the PDF evolution in the FKTable, with further improvements obtained by the careful optimization of the FastKernel product, and even further gains possible when combined with OpenMP on a multiprocessor platform, dividing the computational cost by the total number of available cores.

3.2.4 Minimization algorithm

The minimization is performed using Genetic Algorithms, which are especially suitable for dealing with very large parameter space. Note that the current ANN architecture (2-5-3-1) corresponds to 37 free parameters for each PDF, *i.e.* a total of 259 free parameters, to be compared to less than a total of 30 free parameters for PDF fits based on conventional polynomial functional forms. Because of the extreme flexibility of the fitting functions and the large number of parameters, the optimal fit is not necessarily the absolute minimum of the χ^2 which might correspond to an 'overfit' in which not only the desired best fit is reproduced, but also statistical fluctuation about it. As a consequence, a stopping criterion is needed on top of the minimization method. In the next paragraphs we discuss in turn the GA and the stopping strategies implemented in NNPDF.

Genetic Algorithms

In the new framework, we have performed a careful analysis of the Genetic Algorithm minimization procedure utilized in previous NNPDF determinations. Instead of reproducing the previous methodology, we have introduced new features only if they resulted in faster fitting.

The GA algorithm implemented here consists of three main steps: mutation, evaluation and selection. The minimization procedure of each PDF replica, is completely

[1] www.nvidia.com
[2] https://www.khronos.org/opencl/

independent from each other, so the procedure can be parallelized on multiple machines. Starting from a large number of mutants, PDF sets are generated based on a parent set from the previous generation. The goodness of fit to the data for each mutant is then calculated, with the error function

$$\chi^{2(k)} = \frac{1}{N_{\rm dat}} \sum_{i,j=1}^{N_{\rm dat}} \left(O_{I,i}^{({\rm art})(k)} - O_{I,i}^{({\rm NN})(k)} \right) (\mathrm{cov}_{t_0})_{i,j}^{-1} \left(O_{J,j}^{({\rm art})(k)} - O_{J,j}^{({\rm NN})(k)} \right) , \quad (3.15)$$

where $O_{I,i}^{({\rm NN})(k)}$ is the prediction for replica k of an observable I at a data point i computed with the ANN parametrization, and cov_{t_0} the covariance matrix based on the t_0 prescription explained in Sect. 3.2.1.

The best fit mutant is identified and passed on to the next generation, while the rest are discarded. The algorithm is then iterated until a set of stopping criteria are satisfied. The number of mutants tested each generation is now set to 80 for all generations, removing the two GA 'epochs' used in previous determinations. The choice of this number is arbitrary and depends on the total number of generations. All mutants are generated from the single best mutant from the previous generation.

To generate each mutant, the weights of the neural networks from the parent PDF set are altered by mutations. In fits before NNPDF3.0 the mutations have consisted of point changes, where individual weights or thresholds in the networks were mutated at random. However, investigations of strategies for training neural networks [136] have found that employing coherent mutations across the whole network architecture instead leads to improved fitting performance. The general principle that explains this is that of changing multiple weights which are related by the structure of the network, leading to improvements in both the speed and quality of the training.

In the NNPDF3.0 fits we use a nodal mutation algorithm, which gives for each node in each network an independent probability of being mutated. If a node is selected, its threshold and all of the weights are mutated according to

$$w \to w + \frac{\eta r_\delta}{N_{\rm ite}^{r_{\rm ite}}} , \quad (3.16)$$

where η is the baseline mutation size, r_δ is a uniform random number between -1 and 1, different for each weight, $N_{\rm ite}$ is the number of generations elapsed and $r_{\rm ite}$ is a second uniform random number between 0 and 1 shared by all of the weights. An investigation performed on closure test fits in Sect. 4 of Ref. [12] found that the best value for η is 15, while for the mutation probability the optimal value turns out to be around 5%, which corresponds to an average of 3.15 nodal mutations per mutant PDF set.

As with the removal of the fast- and slow-epochs and their replacement with a single set of GA parameters, the Targeted Weighted Training (TWT) procedure adopted in previous fits has also been dropped. This was originally introduced in order to avoid imbalanced training between datasets. With the considerably larger dataset of NNPDF3.0 along with numerous methodological improvements, such an imbalance is no longer observed even in fits without weighted training. Whereas previously the minimization was initiated with a TWT epoch in which the fit quality to individual datasets was minimized neglecting their cross-correlations, in NNPDF3.0 the minimization always includes all available cross-correlations between experimental datasets.

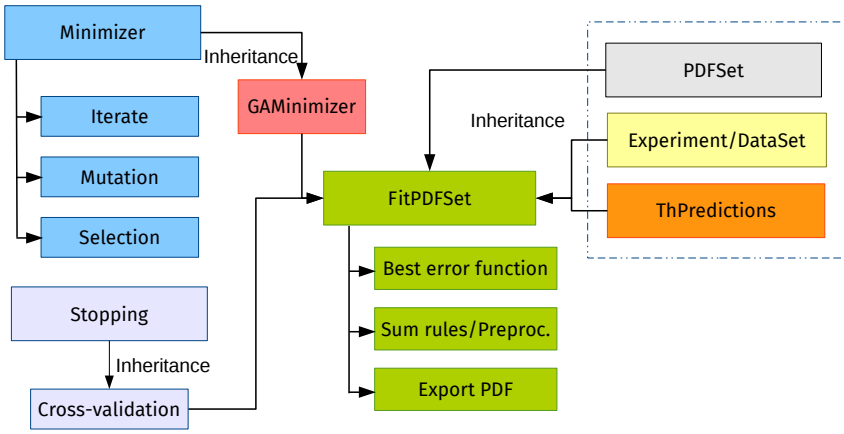

Figure 3.5: Layout of the NNPDF minimization framework.

Stopping criterion

The stopping criterion for the GA is the cross-validation method. This is based on the idea of separating the data in two sets, a training set, which is fitted, and a validation set, which is not fitted. The GA minimizes the χ^2 of the training set, while the χ^2 of the validation set is monitored along the minimization, and the optimal fit is achieved when the validation χ^2 stops improving.

In PDF fits before NNPDF3.0 this stopping criterion was implemented by monitoring a moving average of the training and validation χ^2, and stopping when the validation moving average increased while the training moving average decreased by an amount which exceeded suitably chosen threshold values. The moving average prevented the fit from stopping due to statistical fluctuations, but introduced a certain arbitrariness since the value of these three parameters (the length of the moving average and the two thresholds) had to be tuned.

In NNPDF3.0 the previous stopping criterion is replaced by the so-called *look-back* method which stores the PDF parametrization for the iteration where the fit reaches the absolute minimum of the validation χ^2 within a given maximum number of generations. This method reduces the level of arbitrariness introduced in the previous strategy, however it keeps the total number of iterations for all replicas.

In Figure 3.5 we finalize the description of the new framework with the minimization layout. The output set of PDFs is allocated in the `FitPDFSet` class, inherited from `PDFSet`, which drives the minimization and stores the best mutant and its error function for each iteration of the GA. This class is also responsible for the computation of the normalization coefficients and preprocessing of the neural networks. At the end of the minimization the best PDF parametrization is exported to file. From the minimization point of view, we have coded an abstract `Minimizer` class with *virtual* methods for the GA iteration, mutation and selection. This class is extended in `GAMinimizer` with the technical choices explained in the previous section. It contains all elements for a fast computation of training and validation χ^2 from `ThPredictions`.

The cross-validation data split is performed at level of `Experiment` class, at the beginning of the program, note that this procedure is parallelized for each artificial replica. The last point of the code structure is the `GAMinimizer` connection to the `Stopping` class. This class is easily extended with *i.e. look-back* method.

3.3 NNPDF2.3

Now that we have presented the NNPDF methodology through the new code framework, we conclude this chapter by describing the NNPDF2.3 fitting configuration in terms of PDF parametrization, minimization setup and the description of the data included in this fit. Here, we present this set of PDFs instead of the most recent NNPDF3.0 because QED correction has been obtained from the baseline NNPDF2.3 set. For a complete discussion about the phenomenological impact of this set of PDFs we address the reader to Sect. 1.4 in Chap. 1.

3.3.1 Fit configuration

The PDF parametrization used in the NNPDF2.3 was already shown in Sect. 3.2.2. In Table 3.2 the range of the small- and large-x preprocessing exponents used in this fit are presented for each element of the fitting basis. In the NNPDF2.3 the preprocessing exponents are the same for both NLO and NNLO determinations. These ranges have been redetermined self-consistently for different fits: for example, for fits to reduced datasets, wider ranges are obtained due to the experimental information being less constraining.

The mutation parameters of the Genetic Algorithm used in NNPDF2.3 are presented in the left Table 3.3: for each PDF basis element we show the number of mutations $N_{\rm mut}$ and the respective mutation sizes η. It interesting to note that this configuration has changed in NNPDF3.0 by applying a mutation probability of 5% per network node, and the mutation size to $\eta = 15$.

In NNPDF2.3 we used the cross-validation method with Targeted Weighted Training (TWT) for the first $N_{\rm gen}^{\rm mut} = 2500$. In this first phase of the minimization, we use a large number of mutants $N_{\rm mut}^{a} = 80$, which is then reduced to $N_{\rm mut}^{b} = 30$. The dynamic stopping condition, based on the variation of the moving average of the validation and training χ^2 (see Sect. 3.2.4), is activated after $N_{\rm gen}^{\rm wt} = 10000$. The moving average criterion is complemented by a minimum training $E_{\rm rm}^{\rm min} = 6$. The maximum number of allowed iterations is $N_{\rm gen}^{\rm max} = 50000$. All these parameters are summarized on the right Table 3.3.

3.3.2 Experimental data

After presenting the main characteristics of the NNPDF2.3 methodology, we now discuss about the data set used by this fit. Concerning non-LHC data, the NNPDF2.3 data set includes at NLO and NNLO:

- NMC [137,138], BCDMS [22,139] and SLAC [21] deep-inelastic scattering (DIS) fixed target data;

PDF	NNPDF2.3 NLO and NNLO	
	$[\alpha_{\min}, \alpha_{\max}]$	$[\beta_{\min}, \beta_{\max}]$
Σ	[1.05, 1.35]	[2.55, 3.45]
g	[1.05, 1.35]	[3.55, 4.45]
T_3	[0.00, 0.50]	[2.55, 3.45]
V	[0.00, 0.50]	[2.55, 3.45]
Δ_S	[-0.95, -0.65]	[12.0, 14.0]
s^+	[1.05, 1.35]	[2.55, 3.45]
s^-	[0.00, 0.50]	[2.55, 3.45]

Table 3.2: The small- and large-x preprocessing exponents in Eq. 3.9 randomly chosen in NNPDF2.3.

NNPDF2.3 Single Parameter Mutation		
PDF	N_{mut}	η
Σ	2	10, 1
g	3	10, 3, 0.4
T_3	2	1, 0.1
V	3	8, 1, 0.1
Δ_S	3	5, 1, 0.1
s^+	2	5, 0.5
s^-	2	1, 0.1

NNPDF2.3 Minimization Setup	
Parameter	Value
$N_{\text{gen}}^{\text{wt}}$	10000
$N_{\text{gen}}^{\text{mut}}$	2500
$N_{\text{gen}}^{\text{max}}$	50000
$E_{\text{tr}}^{\text{min}}$	6
N_{mut}^{a}	80
N_{mut}^{b}	30

Table 3.3: The mutation parameters are shown for the NNPDF2.3 determination. In the right table, parameters controlling the maximum fit length, number of mutants, target weighted training settings are shown.

- the combined HERA-I DIS data set [140], HERA F_L [19] and F_2^c structure function data [141–147], ZEUS HERA-II DIS cross-sections [148, 149], CHORUS inclusive neutrino DIS [150], and NuTeV dimuon production data [151, 152];
- fixed-target E605 [153] and E866 [154–156] Drell-Yan production data;
- CDF W asymmetry [157] and CDF [158] and D0 [159] Z rapidity distributions;
- CDF [160] and D0 [161] Run-II one-jet inclusive cross-sections.

The kinematical cuts of DIS data are the usual $Q^2_{\min} = 3$ GeV^2 and $W^2_{\min} = 12.5$ GeV^2. We included also all currently available LHC data for which the experimental covariance matrix has been provided:

- the ATLAS W and Z lepton rapidity distributions from the 2010 data set [25];
- the CMS W electron asymmetry from the 2011 data set [162];
- the LHCb W lepton rapidity distributions from the 2010 data set [163];
- the ATLAS inclusive jet cross-sections from the 2010 run with $R = 0.4$ [132].

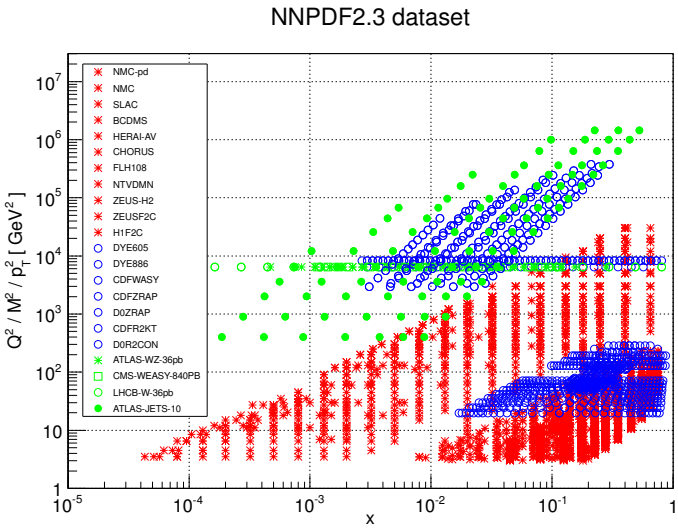

Figure 3.6: The kinematical coverage of the experimental data used in the NNPDF2.3 PDF determination.

More recent measurements from the 2011 and 2012 runs, which are very relevant for PDF fits, like the CMS and LHCb low mass Drell-Yan differential distributions [26,164] and the inclusive jets and dijets from ATLAS and CMS [165,166] have been included in the NNPDF3.0 release.

The kinematical coverage of the LHC data sets included in the NNPDF2.3 analysis with the corresponding average experimental uncertainties for each data set are summarized in Tab. 3.4.[3]. A scatter plot of the kinematical plane for all experimental data from NNPDF2.3 is shown in Fig. 3.6. The LHC electroweak data span a larger range in Bjorken-x than the Tevatron data thanks to the extended rapidity coverage (up to $\eta = 4.5$), while the inclusive jets span a much wider kinematical range both in x and Q^2 than the one accessible at the Tevatron. In Tab. 3.5 we also give the total number of data points used for PDF fitting, both for the NLO and the NNLO global sets.

The theoretical predictions for LHC electroweak boson production have been computed at NLO with `MCFM` [167,168] interfaced with the `APPLgrid` library for fast NLO calculations [128]. The NNLO predictions are obtained by means of local C-factors. These have been computed using the `DYNNLO` code [169]. The kinematical cuts applied to the calculation of the NLO cross sections are now discussed in turn. For the ATLAS data, these are the following:

- cuts for the W lepton rapidity distributions

$$p_T^l \geq 20 \text{ GeV}, \qquad p_T^\nu \geq 25 \text{ GeV}, \qquad m_T < 40 \text{ GeV}, \qquad |\eta_l| \leq 2.5;$$

[3]For jets, we plot only the x value of the parton with smallest x, given by $x = 2\frac{p_T}{\sqrt{s}}e^{-|\eta|}$

Data Set	Ref.	$N_{\rm dat}$	$[\eta_{\rm min}, \eta_{\rm max}]$	$\sigma_{\rm stat}$ (%)	$\sigma_{\rm sys}$ (%)	$\sigma_{\rm norm}$ (%)
CMS We^- asy. 840 pb^{-1}	[162]	11	[0, 2.4]	2.1	4.7	0
ATLAS W$^+$ 36 pb^{-1}	[25]	11	[0, 2.4]	1.4	1.3	3.4
ATLAS W$^-$ 36 pb^{-1}	[25]	11	[0, 2.4]	1.6	1.4	3.4
ATLAS Z 36 pb^{-1}	[25]	8	[0, 3.2]	2.8	2.4	3.4
LHCb W$^+$ 36 pb^{-1}	[163]	5	[2, 4.5]	4.7	11.1	3.4
LHCb W$^-$ 36 pb^{-1}	[163]	5	[2, 4.5]	3.4	7.8	3.4
LHCb Z 36 pb^{-1}	[163]	5	[2, 4.5]	24	4.7	3.4
ATLAS Incl. Jets 36 pb^{-1}	[132]	90	[0, 4.5]	10.2	23.4	3.4

Table 3.4: The number of data points, kinematical coverage and average statistical, systematic and normalization percentage uncertainties for each of the experimental LHC data sets considered for the NNPDF2.3 analysis.

Fit	NLO	NNLO
NNPDF2.3 noLHC	3341	3360
NNPDF2.3 Collider only	1217	1236
NNPDF2.3	3487	3506

Table 3.5: Total number of data points for the various global sets used for PDF fitting.

- cuts for the Z rapidity distribution

$$p_T^l \geq 20 \text{ GeV}, \qquad 66 \text{ GeV} \leq m_{l^+l^-} \leq 116 \text{ GeV}, \qquad \eta_{l^+,l^-} \leq 4.9.$$

In fact, ATLAS measures separately the rapidity distributions in both the electron and muon channels, and then combines them into a common data set. The above kinematical cuts correspond to the combination of electrons and muons, but differ from the cuts applied in individual leptonic channels. For Z rapidity distributions we have explicitly verified that results are unchanged if the cut on the rapidity of the leptons from the Z decay is removed.

For the CMS W electron asymmetry, the only cut is $p_T^e \geq 35$ GeV, with the same binning in electron rapidity as in Ref. [162]. Finally, for the LHCb we have included in our determination only the W data because at that time the Z data was being reanalyzed. The kinematical cuts for LHCb are:

- cuts for the W muon rapidity distributions

$$p_T^\mu \geq 20 \text{ GeV}, \qquad 2.0 \leq \eta^\mu \leq 4.5;$$

For all three data sets, we performed extensive cross-checks at NLO using two different codes, `DYNNLO` and `MCFM`: we checked that, once common settings are adopted, the results of the `MCFM` and `DYNNLO` runs agree to better than 1% for all the data bins. In the particular case of the ATLAS W and Z distributions, we also found good agreement with the `APPLgrid` tables used in the recent `HERAfitter` analysis of ATLAS data [45].

Concerning jet data, we have included the measurements from the Tevatron experiments, which are important for constraining the gluon PDF, together with the extended kinematics coverage provided by the LHC jet data. From the 2010 36pb^{-1} data set inclusive jet and dijet production have been measured by CMS [170, 171] and

ATLAS [132], however only ATLAS give the full experimental covariance matrix. The covariance matrix is particularly important for these data because they are highly correlated.

The theoretical calculation of NLO jet production cross sections in hadron collisions can be carried out by exclusive parton level Monte Carlo codes such as `NLOjet++` [172] and `EKS-MEKS` [173,174]. These MC codes provide NLO predictions which are consistently included in a global PDF analysis using the fast NLO interfaces implemented in `FastNLO` [135,175] or `APPLgrid` [128].

The full NNLO corrections to the inclusive jet production were unknown at that time. Only recently, results about the exact gluons-only channel have been published in Refs. [176,177], but the full channel prediction is still missing. At that time only the threshold corrections to the inclusive jet p_T distribution were available [178], thus the inclusion of jet data into an NNLO analysis is necessarily approximate. On the other hand, in NNPDF3.0 these threshold corrections have been replaced by the improved predictions based on threshold resummation published in Ref. [179], after applying a rejection criterion [180] of kinematical regions based on the difference to the exact gluons-only channel prediction.

We compute inclusive jet cross-sections using `NLOjet++` interfaced to `APPLgrid`. The jet reconstruction parameters are identical to those used in the experimental analysis [181]. The NLO calculation uses the anti-k_T algorithm [182], and the factorization and renormalization scales are set to be $p_T^{\rm max}$, the transverse momentum of the hardest jet in each event. We choose to include in the analysis the data with $R = 0.4$. These data are less sensitive to nonperturbative corrections from the underlying event and pileup as compared to the $R = 0.6$ data [183, 184], and though they are a bit more sensitive to hadronization effects, all in all the nonperturbative parton to hadron correction factors are smaller for $R = 0.4$ than for $R = 0.6$. We have checked that the results are essentially unchanged, both in terms of impact on PDFs and at the level of the χ^2 description if the $R = 0.6$ data is used instead of the $R = 0.4$ data.

On top of the 86 sources of fully correlated systematic errors, the ATLAS jet spectra have an additional source of uncertainty due to the theoretical uncertainty in the computation of the hadron to parton nonperturbative correction factors. We take these nonperturbative corrections and their associated uncertainties from the ATLAS analysis, where they are obtained from the variations of different leading order Monte Carlo programs. It is clear from Ref. [181] that for a given Monte Carlo model the nonperturbative correction is strongly correlated between data bins, and thus conservatively we treat it as an additional source of fully correlated systematic uncertainty, to be added to the covariance matrix.

Because NNLO corrections to jet cross-sections are not available, hadron collider jet data can only be included in a NNLO fit within some approximations. Here, the NNLO theoretical predictions for CDF and D0 inclusive jet data are obtained using the approximate NNLO matrix element obtained from threshold resummation [178] as implemented in the `FastNLO` framework [135,175]. For ATLAS data the threshold approximation is expected to be worse because of the higher centre-of-mass energy, and thus we simply used the NLO matrix element with NNLO PDFs and α_s. It was checked in Ref. [44] that the difference between fits with approximate NNLO jet matrix elements, and fits with purely NLO matrix elements is significantly smaller than the difference between fits with and without jet data. These choices have been updated

in the NNPDF3.0 determination by including consistently threshold resummation C-factors [179] only for data bins where the exact gluons-only predictions are close to the approximation [180] and excluding all the other bins.

Chapter 4

The photon PDF determination

In this chapter we present the determination of the NNPDF2.3QED set of PDFs. This is the first NNPDF set with QED corrections. As we have already mentioned at the beginning of this thesis, thanks to the need of precise phenomenology at the LHC [70, 185], PDFs are determined using the NNLO order in QCD. However, at this level of accuracy, also LO QED corrections ($\mathcal{O}(\alpha)$) become relevant. Some examples about the impact of QED and EW corrections to various hadron collider processes have been studied in detail, *i.e.* the inclusive W and Z production [5, 73–82], W and Z boson production in association with jets [83–85], dijet production [89, 90] and top quark pair production [91–95].

As we have seen in Chapter 2, the first step to obtain a set of PDF with QED corrections consist in the implementation of such corrections to PDF evolution, together with the addition of a new parton: the photon PDF. Before the determination of NNPDF2.3QED, we find in literature only one PDF set with QED corrections: the MRST2004QED set [109]. In this pioneering work, the photon PDF was determined based on a model inspired by photon radiation off constituent quarks (though consistency with some HERA data was checked a posteriori), and therefore not provided with a PDF uncertainty.

The aim of this chapter is to show how we construct a PDF set including QED corrections, with a photon PDF parametrized in the same way as all the other PDFs, and determined from a fit to hard-scattering experimental data using the NNPDF methodology. The goal is to construct a PDF set where

- QCD corrections are included up to NLO or NNLO;
- QED corrections are included to LO;
- the photon PDF is obtained from a fit to deep-inelastic scattering (DIS) and Drell-Yan (both low mass, on-shell W and Z production, and high mass) data;
- all other PDFs are constrained by the same data included in the NNPDF2.3 PDF determination [11], see Chapter 3.

We will consider negligible the impact of the lepton PDF, as well as weak contributions to evolution equations [186, 187].

In principle, this goal could be achieved by simply performing a global fit including QED and QCD corrections both to perturbative evolution and to hard matrix elements, and with data which constrain the photon PDF. In practice, this would require the availability of a fast interface, like `APPLgrid` [128] or `FastNLO` [175], to codes which include QED corrections to processes which are sensitive to the photon PDF, such as single or double gauge boson production. Because such interfaces are not available, we adopt instead a reweighting procedure, which turns out to be sufficiently accurate to accommodate all relevant existing data.

In Figure 4.1 we summarize the steps for the construction of this special set of PDFs:

1. In the first step, we construct a set of PDFs (NNPDF2.3QED DIS-only), including a photon PDF, by performing a fit to DIS data only, based on the same DIS data used for NNPDF2.3 (see Sect. 3.3.2 in Chap. 3), and using either NLO or NNLO QCD and LO QED theory. To leading order in QED, the photon PDF only contributes to DIS through perturbative evolution (just like the gluon PDF to leading order in QCD). Therefore, the photon PDF is only weakly constrained by DIS data, and thus the photon PDF in the NNPDF2.3QED DIS-only set is affected by large uncertainties. The result is a pair of PDF sets: NNPDF2.3QED DIS-only, NLO or NNLO, according to how QCD evolution has been treated.

2. Then, each replica of the photon PDF from the NNPDF2.3QED DIS-only set is combined with a random PDF replica of a set of the default NNPDF2.3 PDFs, fitted to the global dataset. This works because of the small correlation between the photon PDF and other PDFs, as we shall explicitly check. Also, the violation of the momentum sum rule that this procedure entails is not larger than the uncertainty on the momentum sum rule in the global QCD fit. The procedure is performed using NLO or NNLO NNPDF2.3 PDFs, for three values of $\alpha_s(M_z) = 0.117, 0.118, 0.119$. The photon PDF determined in the NNPDF2.3QED DIS-only fit is in fact almost independent of the value of α_s within this range. This leads to several sets of PDF replicas, which we call NNPDF2.3QED prior, at the scale Q_0^2.

3. At this stage, we evolve the NNPDF2.3QED prior set to all Q^2 using combined QCD$\otimes$QED evolution equations, to LO in QED and either to NLO or NNLO in QCD and with the appropriate value of α_s, using the strategy explained in details in Chap. 2 with the `APFEL` implementation.

4. The LHC W and Z/γ^* production data are now included in the fit by Bayesian reweighting [188] of the NNPDF2.3QED prior PDF set.

5. Finally, the set of reweighted replicas is then unweighted [189] in order to obtain a standard set of 100 replicas of our final NNPDF2.3QED set.

As we will see, the photon PDF in NNPDF2.3QED turns out to be in good agreement with that from the MRST2004QED set at medium large $x \gtrsim 0.03$, while for smaller x values it is substantially smaller (by about a factor three for $x \sim 10^{-3}$), though everywhere affected by sizable uncertainties, typically of order 50%.

This chapter is organized as follows. In Sect. 4.1 we also discuss the first step of our procedure, namely, the determination of NNPDF2.3QED DIS-only PDF set. The

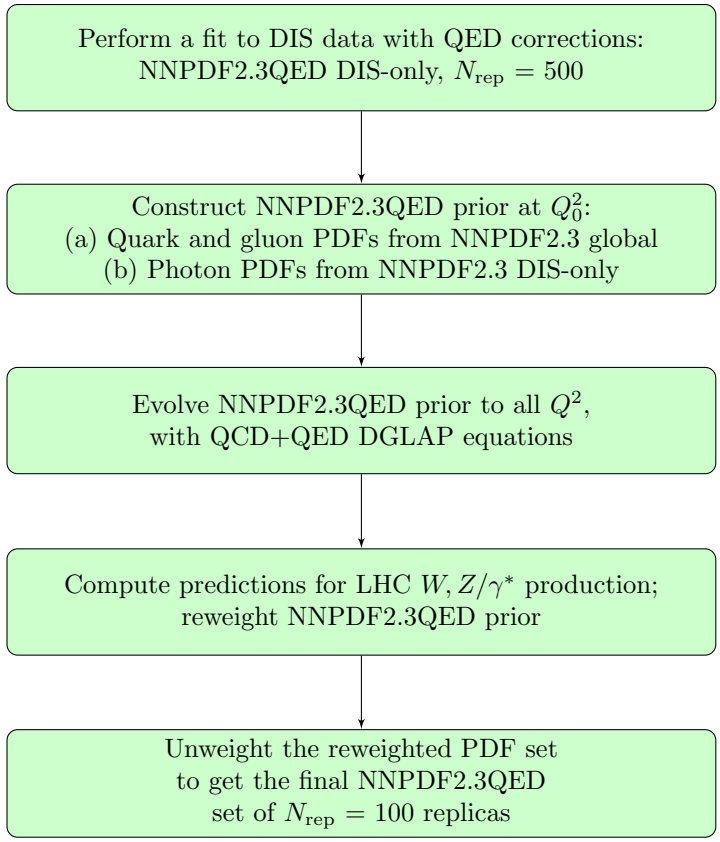

Figure 4.1: Flow-chart for the construction of the NNPDF2.3QED set.

subsequent steps, namely the construction of the NNPDF2.3QED prior set, and its reweighting and unweighting leading to the final NNPDF2.3QED set are presented in Sect. 4.2. Finally, phenomenological investigations of this set of PDFs are presented in Chap. 5.

4.1 Deep-inelastic scattering with QED corrections

4.1.1 Fitting PDFs with QED corrections

Let us now proceed with a first determination of the photon PDF from a fit to deep-inelastic data. We want to include QED corrections to DIS at LO, i.e., more accurately, the leading log level. This means that the splitting functions are computed to $\mathcal{O}(\alpha)$, while all partonic cross-sections (coefficient functions) are determined to lowest order in α. Because the photon is electrically neutral, the photon deep-inelastic coefficient function only starts at $\mathcal{O}(\alpha^2)$, while quark coefficient functions start at $\mathcal{O}(\alpha)$. This means that at LO the photon coefficient function vanishes, and the photon only con-

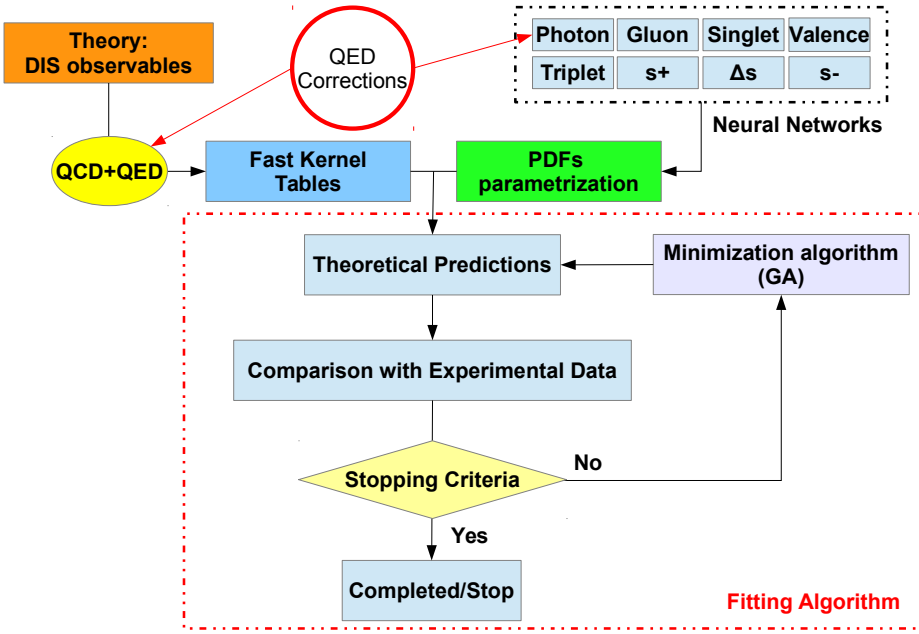

Figure 4.2: Graphical representation of the fitting strategy.

tributes to DIS through its mixing with quarks due to perturbative evolution. This is fully analogous to the role of the gluon in the standard LO QCD description of DIS: the gluon coefficient function only starts at $\mathcal{O}(\alpha_s)$ while the quark coefficient function starts at $\mathcal{O}(1)$, so at LO the gluon only contributes to deep-inelastic scattering through its mixing with quarks upon perturbative evolution.

An important issue when including QED corrections is the choice of the factorization scheme in the subtraction of QED collinear singularities [82, 190]. Different factorization schemes differ by next-to-leading log terms. Because our treatment of QED evolution is at the leading log level, our results do not depend on the choice of factorization scheme. This means that if our photon PDF is used in conjunction with a next-to-leading log computation of QED cross-sections, the latter can be taken in any (reasonable) factorization scheme. The difference in results found when changing the QED factorization scheme should be considered to be part of the theoretical uncertainty. However, in practice, in some schemes the perturbative expansion may show faster convergence (so, for example, next-to-leading log results are closer to leading-log ones in some schemes than others). We will indeed see in the next section that when DIS data are combined with Drell-Yan data it is advantageous to use the DIS factorization scheme, which is defined by requiring that the deep-inelastic structure function F_2 is given to all orders by its leading-order expression [82, 190].

The starting point of our fit to DIS data including QED corrections is the NNPDF2.3 PDF determination, in terms of experimental data, theory settings and methodol-

ogy. We will perform fits at NLO and NNLO in QCD, for three different values of $\alpha_s(M_Z) = 0.117$, 0.118 and 0.119, all with LO QED evolution. Unless otherwise stated, in the following all results, tables, and plots will use the $\alpha_s = 0.119$ PDF sets.

We add to the NNPDF default set of seven independent PDF combinations a new, independently parametrized PDF for the photon, in a completely analogous way to all other PDFs (see Sect. 3.2.2), with a small modification related to positivity to be discussed below:

$$\gamma(x, Q_0^2) = (1-x)^{m_\gamma} x^{-n_\gamma} \mathrm{NN}_\gamma(x), \tag{4.1}$$

where $\mathrm{NN}_\gamma(x)$ is a multi-layer feed-forward neural network with 2-5-3-1 architecture, with a total of 37 parameters to be determined by experimental data, and the prefactor is a preprocessing function used to speed up minimization, and on which the final result should not depend. The preprocessing function is parametrized by the exponents m_γ and n_γ, whose values are chosen at random for each replica, with uniform distribution in the range

$$1 \leq m_\gamma \leq 20, \qquad -1.5 \leq n_\gamma \leq 1.5. \tag{4.2}$$

We have explicitly checked that the results are independent on the preprocessing range, by computing for each replica the effective small- and large-x exponents [13], defined as

$$n_\gamma[\gamma(x,Q^2)] = \frac{\ln \gamma(x,Q^2)}{\ln 1/x}\,, \qquad m_\gamma[\gamma(x,Q^2)] = \frac{\ln \gamma(x,Q^2)}{\ln(1-x)}\,, \tag{4.3}$$

and verifying that the range of the effective exponents at small- and large-x respectively is well within the range of variation of the preprocessing exponents, thus showing that the small- and large-x behaviour of the best-fit PDFs is not constrained by the choice of preprocessing but rather determined by experimental data.

A graphical representation of the strategy described above is shown in Figure 4.2. The DIS predictions and the combined QCD⊗QED evolution are encoded in `FastKernel` tables. The photon PDF parametrization is added to the other flavors of the NNPDF2.3 basis. The convolution between both elements produce theoretical prediction which are compared to experimental data using the standard NNPDF minimization strategy, presented in details in Chap. 3.

Parton distributions must satisfy positivity conditions which follow from the requirement that, even though PDFs are not directly physically observable, they must lead to positive-definite physical cross-sections [191]. Leading-order PDFs are directly observable, and thus they must be positive-definite: indeed, they admit a probabilistic interpretation. Because we treat QED effects at LO, the photon PDF must be positive definite. This is achieved, as in the construction of the NNPDF2.1 LO PDF sets [44], by squaring the output of the neuron in the last (linear) layer of the neural network $\mathrm{NN}_\gamma(x)$, so that $\mathrm{NN}_\gamma(x)$ is a positive semi-definite function.

Once QED evolution is switched on, isospin is no longer a good symmetry, and thus it can no longer be used to relate the PDFs of the proton and neutron. Because deuteron deep-inelastic scattering data are used in the fit, in principle this requires an independent parametrization for proton and neutron PDFs. Experimental data for the neutron PDFs would then no longer provide a useful constraint, and in particular they

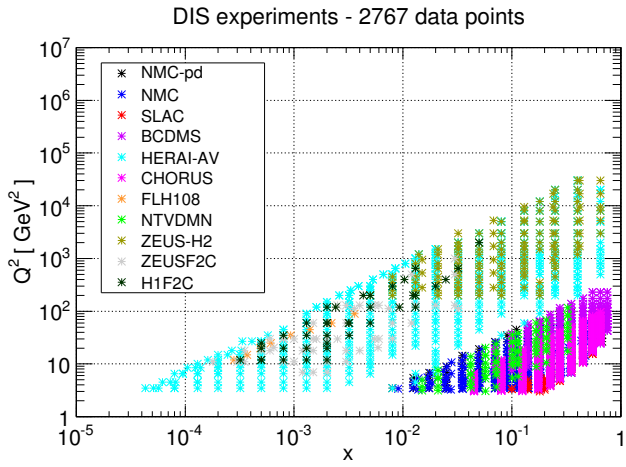

Figure 4.3: Kinematic coverage of the experimental DIS data used in the determination of the photon PDF.

would no longer constrain the isospin triplet PDF. Whereas future PDF fits including substantially more LHC data might allow for an accurate PDF determination without using deuteron data, this does not seem to be possible at present.

There are two separate issues here: one, is the amount of isospin violation in the quark and gluon PDFs, and the second is the amount of isospin violation in the photon PDF. At the scale at which PDFs are parametrized, which is of the order of the nucleon mass, we expect isospin violating effects in the quark and gluon PDFs to be of the same order as that displayed in baryon spectroscopy, which is at the per mille level, much below the current PDF uncertainties (isospin violations of this order have been predicted, among others, on the basis of bag model estimates [192]). The second is the amount of isospin violation in the photon distribution itself: this could be somewhat larger (perhaps at the percent level), however any reasonable amount of isospin violation in the photon is way below the uncertainty on the photon PDF. Therefore, we will assume that no isospin violation is present at the initial scale.

Of course, even with isospin conserving PDFs at the starting scale, isospin violation is then generated by QED evolution: this is consistently accounted for when solving the evolution equations, by determining separate solutions for the proton and neutron so that at any scale $Q \neq Q_0$, $u^p(x,Q^2) \neq d^n(x,Q^2)$ and $d^p(x,Q^2) \neq u^n(x,Q^2)$. Because of the larger electric charge of the up quark, the dynamically generated photon PDF ends up being larger for the proton than it is for the neutron.

In Ref. [109] isospin violation was parametrized on the basis of model assumptions. We will compare our results for isospin violation to those of this reference in Sect. 4.2.2 below: we will see that while indeed the amount of isospin violation in the photon PDF from that reference is somewhat larger than our own, it is much smaller than the relevant uncertainty.

	NLO		NNLO	
Experiment	QCD	QCD+QED	QCD	QCD+QED
Total	1.10	1.10	1.10	1.10
NMC-pd	0.88	0.87	0.88	0.88
NMC	1.68	1.70	1.67	1.69
SLAC	1.36	1.40	1.08	1.10
BCDMS	1.17	1.16	1.24	1.23
CHORUS	1.01	1.01	0.98	0.99
NTVDMN	0.54	0.54	0.56	0.54
HERAI-AV	1.01	1.01	1.04	1.03
FLH108	1.34	1.34	1.25	1.24
ZEUS-H2	1.26	1.25	1.24	1.25
ZEUS F_2^c	0.75	0.75	0.76	0.78
H1 F_2^c	1.55	1.50	1.41	1.39

Table 4.1: The χ^2 values per data point for individual experiments computed in the NNPDF2.3 DIS-only NLO and NNLO PDF sets, in the QCD-only fits compared to the results with combined QCD⊗QED evolution. All χ^2 values have been obtained using $N_{\rm rep}$=100 replicas with $\alpha_s(M_Z) = 0.119$. Normalization uncertainties have been included using the experimental definition of the covariance matrix, see App. A of Ref. [8], while in the actual fitting the t_0 definition was used [193].

4.1.2 The photon PDF from DIS data

We have performed two fits at NLO and NNLO to DIS data only, with the same settings used for NNPDF2.3, but with QED corrections in the PDF evolution now included, as discussed in Chap. 2. The kinematic coverage of experimental DIS data used in this fit is presented in Figure 4.3.

The χ^2 for the fit to the total dataset and the individual DIS experiments are shown in Table 4.1, with and without QED corrections, and with QCD corrections included either at NLO or at NNLO. The χ^2 listed in the table use the so-called experimental definition of the χ^2, in which normalization uncertainties are included in the covariance matrix: this definition is most suitable for benchmarking purposes, as it is independent of the fit results, but it is unsuitable for minimization as it would lead to biased fit results. It is clear that there is essentially no difference in fit quality between the QCD and QED⊗QCD fits. Indeed, a direct comparison of the PDFs obtained in the pairs of fits with and without QED corrections show that they differ very little.

In order to assess this difference quantitatively, in Figure 4.4 we plot the distance between central values and uncertainties of individual combinations of PDFs in the NLO QCD fit before and after the inclusion of QED corrections. We refer to Appendix A for the definition of distance. Recall that for a set of $N_{\rm rep}$ PDF replicas, $d \sim 1$ corresponds to PDFs extracted from the same underlying distribution, *i.e.* to statistically equivalent PDF sets, while $d \sim \sqrt{N_{\rm rep}}$ (so $d \sim 10$ in our case) corresponds to PDFs extracted from distributions whose means or central values differ by one σ. The distances are shown in Figure 4.4 for the NLO fit: it is clear that all PDFs but the gluon from the sets with and without QED corrections are statistically equivalent, while the gluon shows a change in the valence region of less than half σ. These results

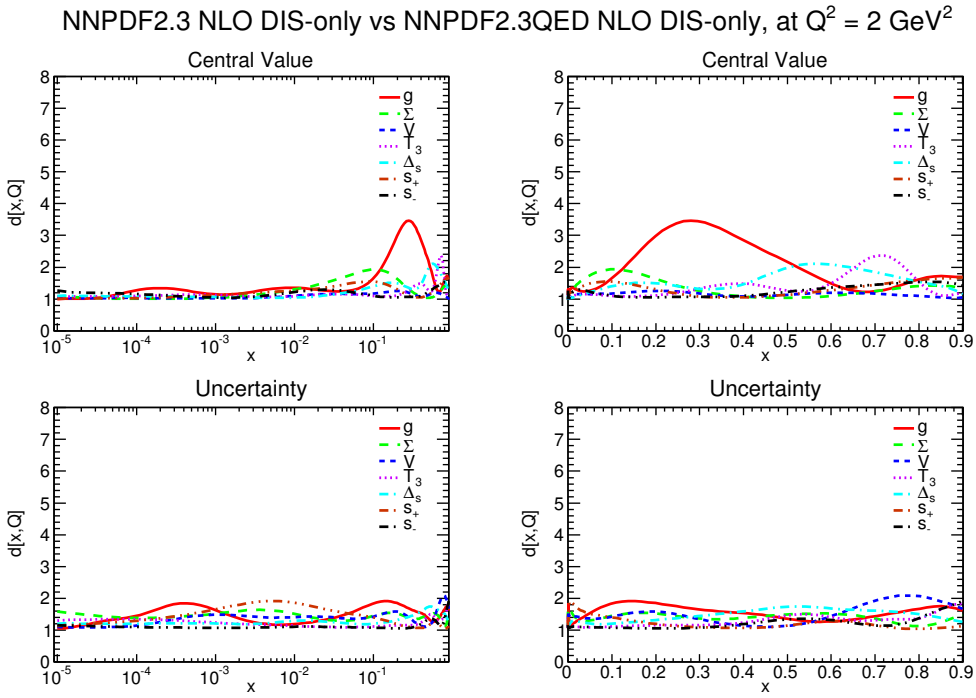

Figure 4.4: Distances between PDFs in the NNPDF2.3 NLO DIS-only fit and the fit including QED corrections, at the input scale of Q_0^2=2 GeV^2. Distances between central values (top) and uncertainties (bottom) are shown, on a logarithmic (left) or linear (right) scale in x.

are unchanged when QCD is treated at NNLO order.

The fact that the inclusion of a photon PDF has a negligible impact on other PDFs can be also seen by determining the correlation between the photon and other PDFs. Results are shown in Figure 4.5. The correlation is negligible at the input scale, meaning that the particular shape of the photon in each replica has essentially no effect on the other PDFs of that replica. In particular, this correlation is much smaller than that which arises at a higher scale (also shown in Figure 4.5), due to the mixing of PDFs with the photon induced by PDF evolution.

Hence, at the initial scale $Q_0^2 = 2$ GeV^2 the sets with and without QED corrections differ mainly because of the presence of a photon PDF in the latter. The photon PDF determined in the NLO fit is shown in Figure 4.6 at $Q_0^2 = 2$ GeV^2: the individual replicas, the mean value, the one-σ range and the 68% confidence interval are all shown. The MRST2004QED photon PDF is also shown. It is clear that positivity imposes a strong constraint on the photon PDF, which is only very loosely constrained by DIS data. As a consequence, the probability distribution of replicas is very asymmetric: some replicas may have large positive values of $\gamma(x, Q^2)$, but positivity always ensures that no replica goes below zero. It follows that the usual gaussian assumptions cannot be made, and in particular there is a certain latitude in how to define the uncertainty.

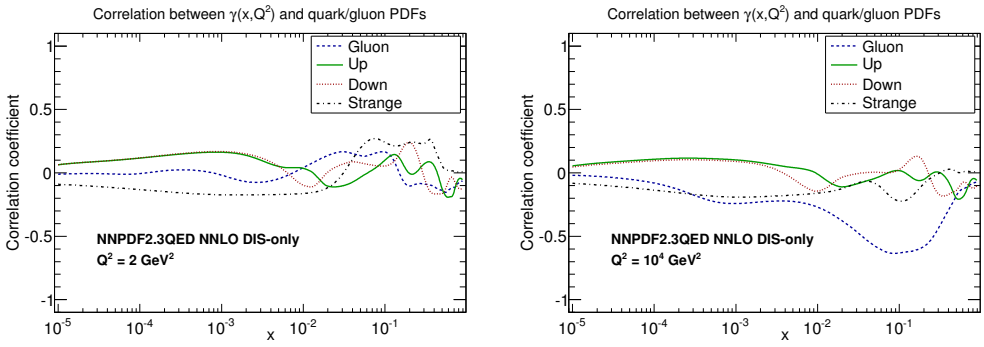

Figure 4.5: Correlation between the photon and other PDFs in the NNPDF2.3QED NLO DIS-only fit, shown as a function of x at the input scale Q_0^2=2 GeV^2 (left) and at $Q^2 = 10^4$ GeV^2.

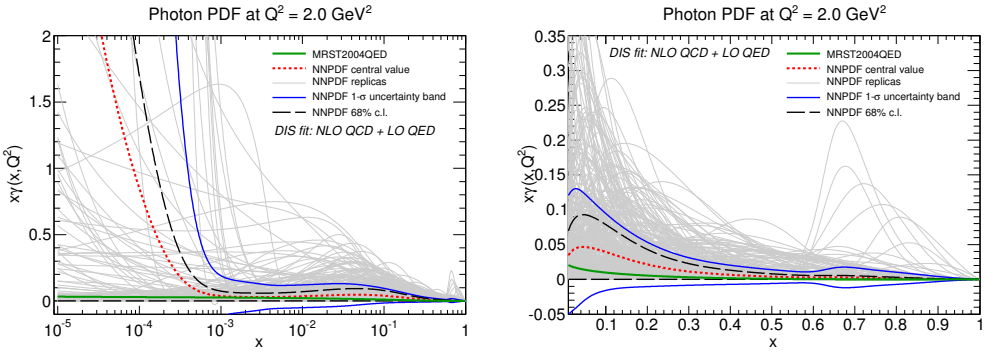

Figure 4.6: The photon PDF determined from the NNPDF2.3QED NLO DIS-only fit, in a linear (left plot) and logarithmic (right plot) scales, $N_{\rm rep} = 500$. We show the central value (mean), the individual replicas and the PDF uncertainty band defined as a one σ sigma interval and as a symmetric 68% confidence level centered at the mean. The MRST2004QED photon PDF is also shown.

Here and in the remainder of this paper we will always define central values as the mean of the distribution, and uncertainties as symmetric 68% confidence levels centered at the mean, namely, as the symmetric interval centered at the mean such that 68% of the replicas falls within it. All uncertainty bands will be determined in this way, unless otherwise stated. Because of the accumulation of replicas just above zero, the lower edge of the uncertainty band on the photon PDF at the initial scale turns out to be very close to zero. Again, results are essentially unchanged when the fit is done using NNLO QCD theory.

As discussed in Sect. 4.1.1, we have determined the effective exponents Eq. (4.3) for the photon PDF, and compared them to the range of variation of the preprocessing exponents Eq. (4.2). Given the very loose constraints that the data impose on the photon PDF, it is especially important to make sure that preprocessing imposes no

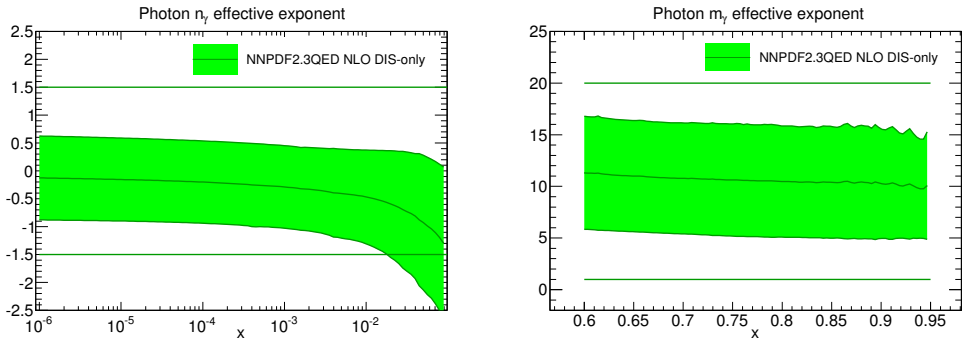

Figure 4.7: One-σ range for the effective exponents Eq. (4.3) for the photon PDF, compared to the range of variation of the preprocessing exponents Eq. (4.2) (shown as horizontal lines).

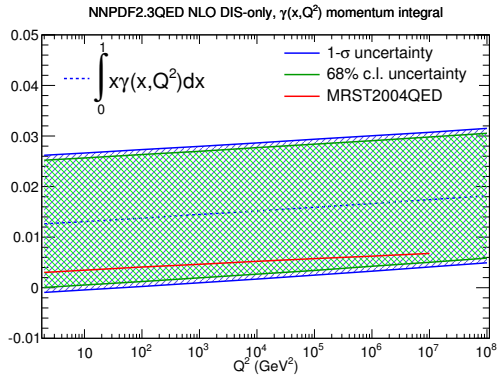

Figure 4.8: The momentum fraction carried by the photon PDF in the NLO fit as a function of scale. The MRST2004QED result is also shown.

bias. The comparison is shown in Figure 4.7: it is clear that the effective exponents are well within the range chosen for the preprocessing exponents, so that no bias is being introduced.

The photon PDF at the initial scale shown in Figure 4.6 is essentially compatible with zero, and it remains small even at the top of its uncertainty band; it is consistent with the MRST2004QED photon PDF within its large uncertainty band.

The momentum fraction carried by the photon is accordingly small: it is shown as a function of scale in Figure 4.8 for the NLO fit; results at NNLO are very similar. At the input scale $Q_0^2 = 2$ GeV^2 we find

$$\int_0^1 x\gamma\left(x, Q_0^2\right) = (1.26 \pm 1.26)\ \% \ , \tag{4.4}$$

The symmetric 68% confidence level uncertainty of Eq. (4.4) turns out to be quite close to the standard deviation $\sigma = 1.36\%$. Hence, even at the top of its uncertainty

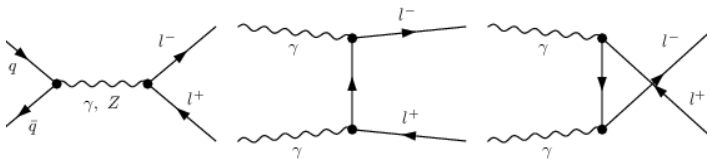

Figure 4.9: Feynman diagrams for the Born-level partonic subprocesses which contribute to the production of dilepton pairs in hadronic collisions.

range the photon momentum fraction hardly exceeds 2%, and it is compatible with zero to one σ. The momentum fraction carried by the the MRST2004QED photon (also shown in Figure 4.8) is well below 1%, and thus compatible with our own within uncertainties

4.2 The photon PDF from W and Z production at the LHC

As we have seen in the last section the photon PDF $\gamma(x, Q^2)$ determined from a fit to DIS data is affected by large uncertainties. This suggests that its impact on predictions for hadron collider processes to which the photon PDF contributes already at leading order could be substantial, and thus, conversely, that data on such processes might provide further constraints. In this section we use the simplest of such processes, namely, electroweak gauge boson production, to constrain the photon PDF.

At hadron colliders, the dilepton production process receives contributions at Born level both from quark-initiated neutral current Z/γ^* exchange and from photon-initiated diagrams, see Figure 4.9, and thus the contributions from $\gamma(x, Q^2)$ must be included even in a pure leading-order treatment of QED effects. Photon-initiated contributions to dilepton production at hadron colliders were recently emphasized in Ref. [82], where $\mathcal{O}(\alpha)$ radiative corrections to this process [5,73,75–82] were reassessed, and also kinematic cuts to enhance the sensitivity to $\gamma(x, Q^2)$ were suggested.

Beyond the Born approximation, radiative corrections to the neutral-current process, as well as the charged-current process, which starts at $\mathcal{O}(\alpha)$ (see Figure 4.10 for some representative Feynman diagrams) may be comparable in size to the Born level contribution, because the suppression due to the extra power of α might be compensated by the enhancement arising from the larger size of the quark-photon parton luminosity in comparison to the photon-photon luminosity. However, a full inclusion of $\mathcal{O}(\alpha)$ corrections would require solving evolution equations to NLO in the QED and mixed QED$\otimes$QCD terms, so it is beyond the scope of this work; we will nevertheless discuss an approximate inclusion of such corrections which, while not allowing us to claim more than LO accuracy in QED, should ensure that NLO QED corrections are not unnaturally large.

We use neutral and charged-current Drell-Yan production data from the LHC to further constrain the photon PDF, thereby arriving at our final NNPDF2.3QED PDF sets. This is obtained by combining the photon PDF from NNPDF2.3 DIS-only set discussed in the previous section with the standard NNPDF2.3 PDF set, and then

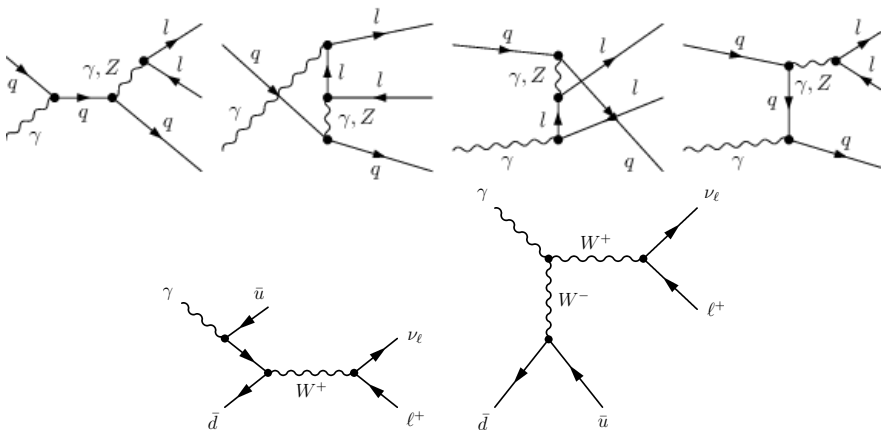

Figure 4.10: Some Feynman diagrams for $\mathcal{O}(\alpha)$ photon-initiated partonic subprocesses which contribute to neutral current (top row) and charged current (bottom row) dilepton pair production in hadronic collisions.

using gauge boson production data to reweight the result. We discuss first this two-step fitting procedure, and then the ensuing NNPDF2.3QED PDF set and its features.

4.2.1 The prior NNPDF2.3QED and its reweighting

As a first step towards the determination of a PDF set with inclusion of QED corrections, we use the photon PDF determined in the previous section from a fit to DIS data in conjunction with PDFs which retain all the information provided by the full NNPDF2.3 data set, which, on top of DIS, includes Drell-Yan and jet production data from the Tevatron and the LHC, as we have explained in details in Sect. 3.3.2.

We have seen in the previous section that all PDFs determined including QED corrections are statistically equivalent to their standard counterparts determined when QED corrections are not included, with the only exception of the gluon, which undergoes a change by less than half σ in a limited kinematic region. Furthermore, the photon in each PDF replica is essentially uncorrelated to the shape of other PDFs which are input to perturbative evolution, the only significant correlation being due to the mixing induced by the evolution itself. We can therefore simply combine the photon PDF obtained from the DIS fit of the previous section with the standard NNPDF2.3 PDFs at the starting scale $Q_0^2 = 2$ GeV^2. This procedure implies a certain loss of accuracy, which in particular appears as a violation of the momentum sum rule of the order of the momentum fraction carried by the photon at the initial scale Eq. (4.4), namely of order 1%. This is the accuracy to which the momentum sum rule would be verified if it were not imposed as a constraint in the fit [44].

The information contained in LHC Drell-Yan production data is included in the fit through the Bayesian reweighting method presented in Ref. [188, 189] and summarized in Appendix B. This method allows for the inclusion of new data without having to perform a full refit, by using Bayes' theorem to modify the prior probability distribution of PDF replicas in order to account for the information contained in the

Dataset	Observable	Ref.	$N_{\rm dat}$	$[\eta_{\rm min}, \eta_{\rm max}]$	$[M_{\rm ll}^{\rm min}, M_{\rm ll}^{\rm max}]$
LHCb γ^*/Z Low Mass	$d\sigma(Z)/dM_{ll}$	[26]	9	[2,4.5]	[5,120] GeV
ATLAS W, Z	$d\sigma(W^{\pm}, Z)/d\eta$	[25]	30	[-2.5,2.5]	[60,120] GeV
ATLAS γ^*/Z High Mass	$d\sigma(Z)/dM_{ll}$	[24]	13	[-2.5,2.5]	[116,1500] GeV

Table 4.2: Kinematical coverage of the three LHC datasets used to determinethe photon PDF.

new data. The ensuing replica set contains an amount of information, and thus allows for the computation of observables with an accuracy, that corresponds to an effective number of replicas $N_{\rm eff}$, which may be determined from the Shannon entropy of the reweighted set.

This new data only constrains significantly the photon PDF, hence we need to guarantee that good accuracy is obtained by starting with a large number of photon replicas. The initial prior set is thus obtained combining 500 photon PDF replicas with a standard set of 100 NNPDF2.3 replicas. In practice, this is done by simply producing five copies of the NNPDF2.3 100 replica set, and combining each of them at random with one of the 500 photon PDF replicas obtained from the QED fit to DIS data discussed in the previous section. The procedure is performed at NLO and NNLO, in each case combining the photon PDF from the combined QED⊗QCD fit to DIS data with the other PDFs from the corresponding standard NNPDF2.3 set. Furthermore, the procedure is repeated for three different values of $\alpha_s = 0.117, 0.118, 0.119$. We find no dependence of the photon PDF on the value of α_s, though there are minor differences between the photon determined using NLO or NNLO QCD theory in the DIS fit.

In each case, the set of $N_{\rm rep} = 500$ replicas is then evolved to all scales using combined QED⊗QCD evolution. Note that this in particular implies that no further violation of the momentum sum rule is introduced on top of that which was present at the initial scale, up to approximations introduced when solving the evolution equations.

In this work, the reweighting is performed using the following LHC datasets:

- LHCb low-mass Z/γ^* Drell-Yan production from the 2010 run [26]
- ATLAS inclusive W and Z production data from the 2010 run [25]
- ATLAS high-mass Z/γ^* Drell-Yan production from the 2011 run [24],

whose kinematic coverage is summarized in Table 4.2. Using data with three different mass ranges for the dilepton pairs, below, at, and above the W and Z mass, guarantees that both the low x (from low mass) and high x (from high mass) regions are covered.

For all the ATLAS data the experimental covariance matrix is available, hence the χ^2 may be computed fully accounting for correlated systematics. However, this is not the case for LHCb at that time: hence, the low-mass data are treated adding statistical and systematic errors in quadrature, and only including normalization errors in the covariance matrix. We have checked that if reweighting is performed using the diagonal covariance matrix, statistically indistinguishable results are obtained. This means that within the large uncertainty of the photon PDF, and due to the small impact of QED corrections on the quark and gluon PDFs, the lack of information on

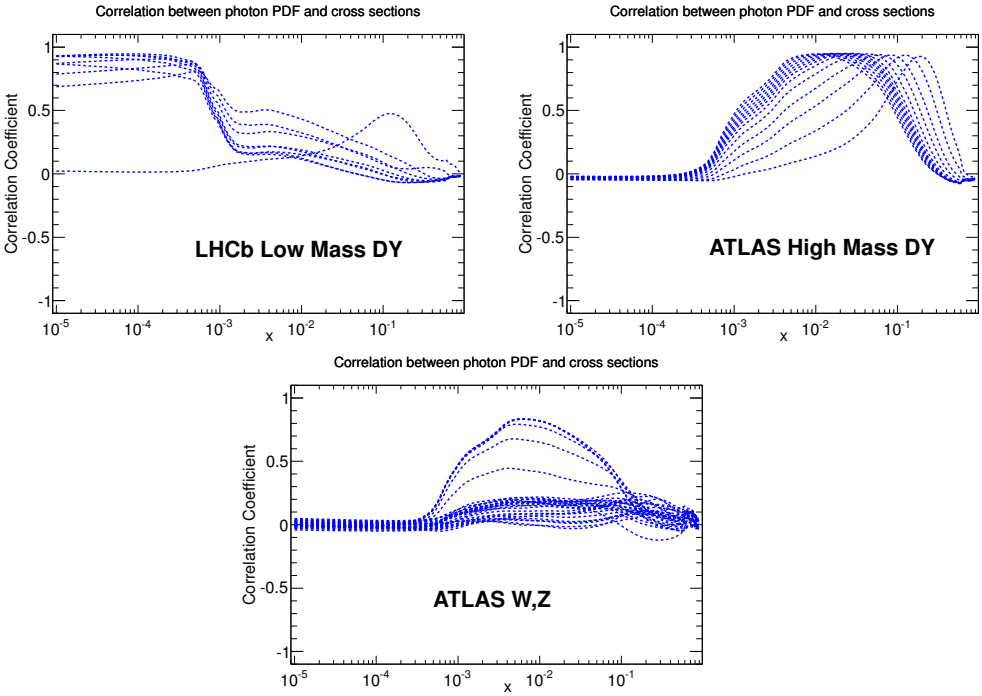

Figure 4.11: Correlation between the photon PDF and the LHC data of Tab. 4.2, shown as function of x for $Q^2 = 10^4$ GeV^2. Each curve corresponds to an individual data bin.

correlations for the LHCb experiment is immaterial. However, this implies that χ^2 values quoted for LHCb should only be taken as indicative. Unfortunately, at that time the CMS off-peak Drell-Yan data [164] was not publicly available, and thus could not be used in the present analysis.

The range of x for the photon PDF which is affected by each of the datasets of Table 4.2 can be determined quantitatively by computing the correlation coefficient (see [194] and Sect. 4.2 of Ref. [195]) between a given observable and the PDFs. The correlation coefficients computed using the NNPDF2.3QED NLO prior set are shown in Figure 4.11 for each bin in the experiments in Table 4.2. It is clear that the LHC data guarantee a good kinematic coverage for all $10^{-5} \lesssim x \lesssim 0.1$. The correlation is weaker for real W and Z production data, where the s-channel quark contribution dominates as the propagator goes on shell. The high-mass (low-mass) Drell-Yan data is thus essential to pin down $\gamma(x, Q^2)$ at large (small) Bjorken-x, where uncertainties are the largest. A preliminary determination of the photon distribution [15], which did not use the LHCb data, had significantly larger uncertainties at small x, consistently with the expectations based on the correlation plot of Figure 4.11.

Theoretical predictions for the datasets in Table 4.2 have been computed at NLO and NNLO in QCD using `DYNNLO` [169], supplemented with Born-level and $\mathcal{O}(\alpha)$ QED corrections using `HORACE` [5, 81]. Results from `DYNNLO` and `HORACE` have been

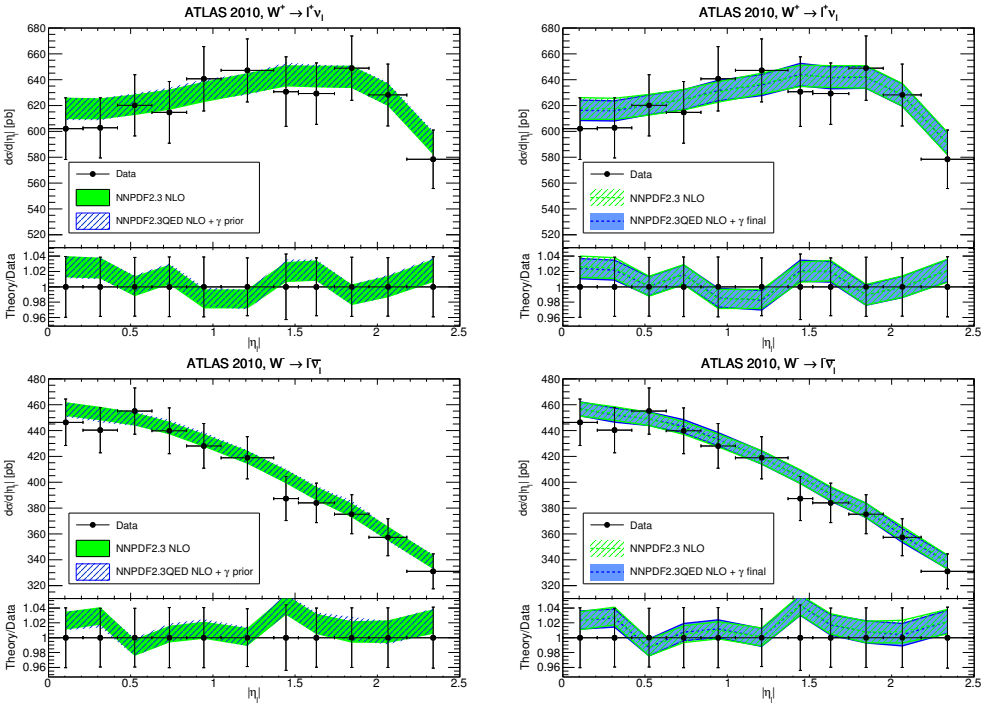

Figure 4.12: Comparison of the ATLAS W production data with NLO theoretical predictions obtained using PDFs before (left) and after (right) reweighting with the data of Tab. 4.2. In all plots we also show for comparison results obtained using the default NNPDF2.3 PDF set, with all QED corrections switched off. From top to bottom: W^+ and W^-. Error bands on the theoretical prediction correspond to one σ uncertainties. Experimental error bars give the total combined statistical and systematic uncertainty.

combined additively, avoiding double counting, in order to obtain a consistent combined QCD⊗QED theory prediction. The additive combination of QED and QCD corrections avoids introducing $\mathcal{O}(\alpha\alpha_s)$ terms, which are beyond the accuracy of our calculation. In the `DYNNLO` calculation, the renormalization and factorization scale have been set to the invariant mass of the dilepton pair in each bin. The `HORACE` default settings, with the renormalization and factorization set to the mass of the gauge boson, have been used for the ATLAS high-mass data, but we have also checked that for this data the choice is immaterial, in that the LO results obtained using `DYNNLO` and `HORACE` with the respective scale settings agree with each other.

For the LHCb low-mass data we have used a modified version of `HORACE` in which the scale choice is the same as in `DYNNLO`, since for these low scale data the choice of renormalization and factorization scale does make a significant difference. Note that the smallest mass values reached by these data correspond to momentum fractions $x \sim 10^{-3}$ in the central rapidity regions, for which, at the scale of the data, fixed order (unresummed) results are expected to be adequate (see Ref. [196], in particular

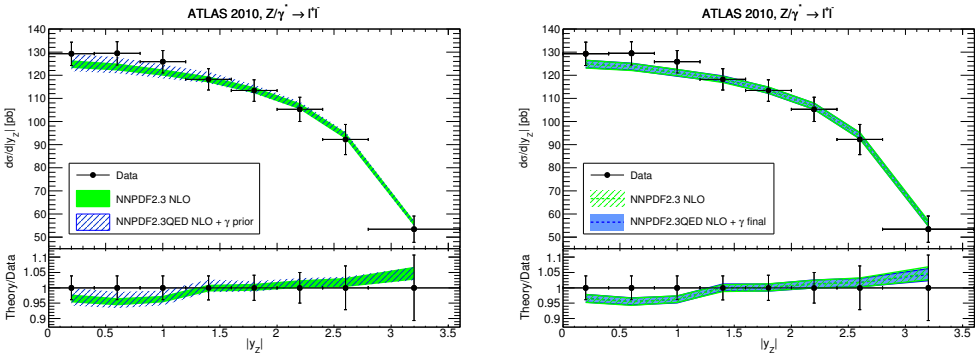

Figure 4.13: Same as Fig. 4.12, but for the neutral current data.

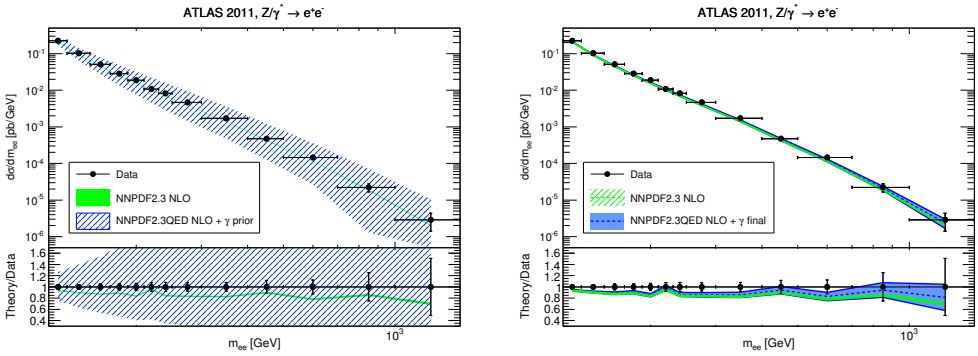

Figure 4.14: Same as Fig. 4.12, but for the ATLAS high-mass neutral-current data.

Figure 1). Indeed we shall see that our results are perturbatively stable in that the photon PDF at NLO and NNLO is very similar for all x (see Figs.4.16-4.17 below).

The same selection and kinematical cuts as in the corresponding experimental analysis has been adopted: in particular, the same requirements concerning lepton-photon final state recombination and the treatment of final state QED radiation have been implemented in the `HORACE` computations.

It should be noticed that, whereas the LHCb and ATLAS high-mass data are only being included now in the fit, the W and Z production data were already included in the original NNPDF2.3 PDF determination (where they turned out to have a moderate impact). Therefore, in principle a modified version of NNPDF2.3 in which these data are removed from the fit should have been used as a prior. In practice, however, this would make very little difference. We have verified that the inclusion of QED evolution affects minimally the prediction for this data, where differences are at the same level of the Monte Carlo integration uncertainty, recalling (see Figure 4.11) that the main impact of this data is in the $x \sim 0.01$ region. This means that the contributions to this process in the reweighting and in the original NNPDF2.3 fit in practice only differ because of the inclusion of the photon contribution. Furthermore, we have explicitly verified that if the ATLAS W and Z production data are excluded from the fit, the

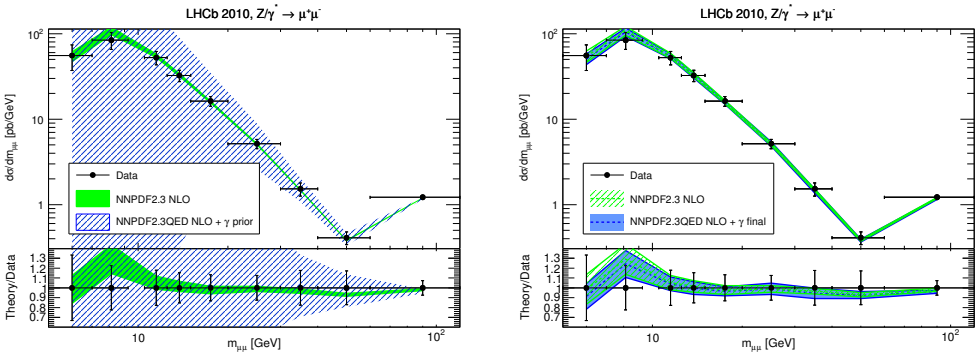

Figure 4.15: Same as Fig. 4.12, but for the LHCb low-mass neutral current data.

photon is systematically modified by a small but non-negligible amount (less then half σ at most) in the region $x \sim 10^{-3}$ where these data are expected to carry information (see Figure 4.11), while all other PDFs are essentially unaffected.

Whereas our computation is only accurate to leading order in QED, we did include $\mathcal{O}(\alpha)$ corrections to the electroweak gauge boson production process through `HORACE`, with the aim of avoiding unnaturally large NLO QED corrections. This raises several issues which we now discuss in turn.

As pointed out in Refs. [82, 190], usage of the leading-order expressions in QED for the DIS coefficient functions can be viewed as the choice of the DIS factorization scheme, in which deep-inelastic coefficient functions are taken to coincide to all orders with their leading-order expression, with higher order corrections factorized into the PDFs. Therefore, use of the DIS scheme for the QED corrections to the Drell-Yan process ensures that predictions for Drell-Yan obtained with PDFs determined using DIS data and LO QED are actually accurate up to NLO, modulo any NLO corrections from QED evolution. Therefore, we have used the DIS-scheme expressions for NLO corrections to Drell-Yan as implemented in `HORACE`. Of course, in practice, there will be NLO QED evolution effects, even though there is a certain overlap between the kinematic region of the HERA DIS data and that of the LHC Drell-Yan data, so we cannot claim NLO QED accuracy. However we expect this procedure to lead to greater stability of our results upon the inclusion of NLO QED corrections.

Radiative corrections related to final-state QED radiation have already been subtracted from the ATLAS data, but not from the LHCb data. Therefore, for ATLAS we have only included photon-induced processes in the `HORACE` runs, while for LHCb we have also included explicit $\mathcal{O}(\alpha)$ contributions from final-state QED radiation. Electroweak corrections, which are not subtracted from any of the data and which are not included in our calculation, could be potentially relevant in the high-mass region [82]. However, in practice they are always much smaller than the statistical uncertainty on the ATLAS data.

Finally, to NLO in QED the scheme used in defining electroweak couplings should be specified. The `DYNNLO` code uses the so-called G_μ scheme for the electroweak couplings, while `HORACE` also uses the G_μ scheme for charged-current production, but the improved Born approximation (IBA) for neutral-current production. We have verified

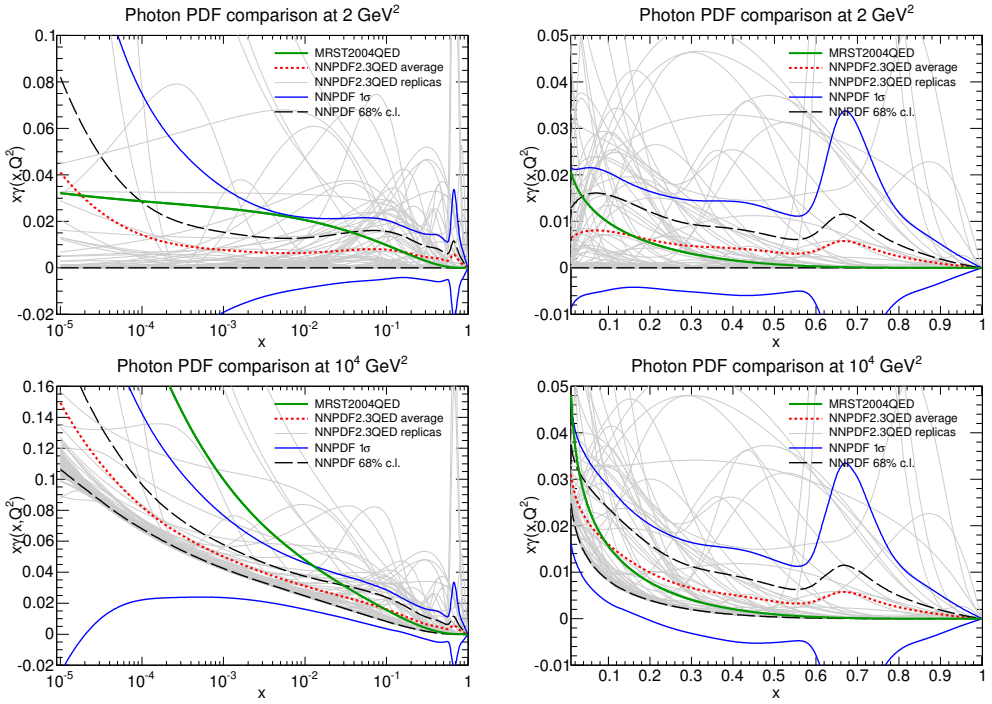

Figure 4.16: The NNPDF2.3QED NLO photon PDF at $Q^2 = 2$ GeV^2 and $Q^2 = 10^4$ GeV^2 plotted vs. x on a log (left) or linear (right) scale. The 100 replicas are shown, along with the mean, the one-σ, and the 68% confidence level ranges. The MRST2004QED photon PDF is also shown for comparison.

the differences in predictions between the two scheme are negligible in comparison to the statistical uncertainties of the Monte Carlo integrations, more details about the IBA scheme will be presented in Chap. 5.

4.2.2 The NNPDF2.3QED set

The NNPDF2.3QED PDF set is obtained by performing a reweighting of the prior $N_{\rm rep} = 500$ replica set with the data of Table 4.2. The procedure is performed at NLO and NNLO in QCD, with three different values of α_s in each case. The theoretical prediction used for reweighting is computed as discussed in the previous section, and the χ^2 used for reweighting is then determined from its comparison to the data, using the fully correlated systematics for the two ATLAS experiments, for which the covariance matrix is available, but adding statistical and systematic errors in quadrature for LHCb, for which information on correlations is not available. The ensuing weighted set of replicas is then unweighted [189] to obtain a standard set of $N_{\rm rep} = 100$ replicas.

The parameters of the reweighting are collected in Table 4.3: we show the χ^2 (divided by the number of data points) for the data of Table 4.2 before and after reweighting, the effective number of replicas after reweighting, and the mean value of α, the parameter which measures the consistency of the data which are used for

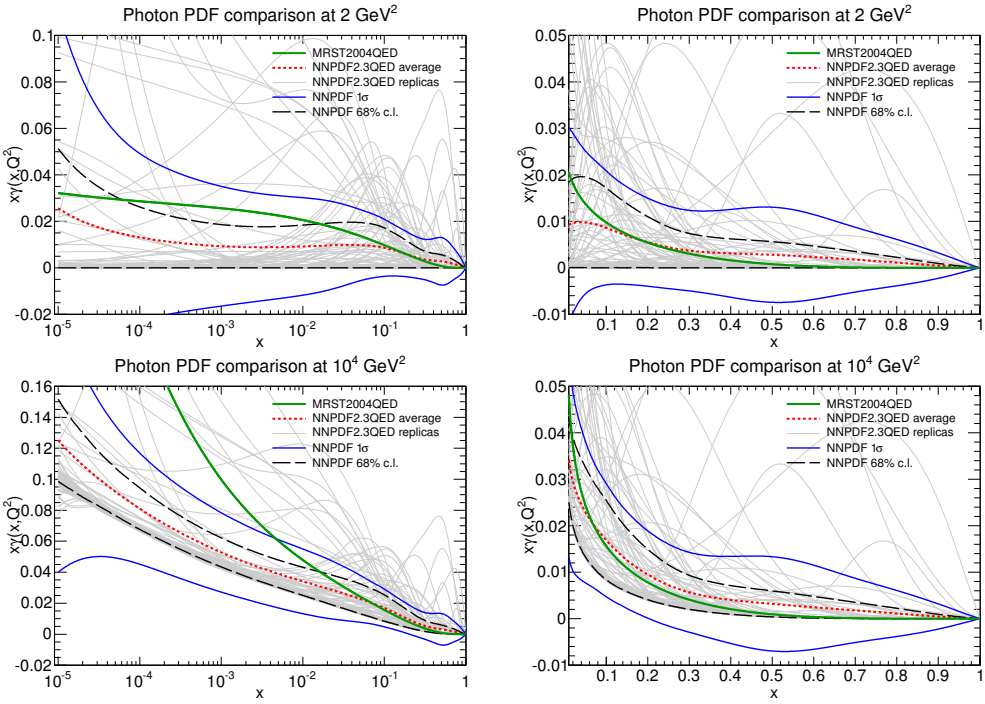

Figure 4.17: Same as 4.16 for the NNPDF2.3QED NNLO PDF set.

reweighting with those included in the prior set, by providing the factor by which the uncertainty on the new data must be rescaled in order of the two sets to be consistent (so $\alpha \sim 1$ means consistent data). Values are given for reweighting performed using each individual dataset, and the three datasets combined. All χ^2 values are computed using the experimental definition of the covariance matrix as in Table 4.1; the same form of the covariance matrix has also been used for reweighting for simplicity, as this choice is immaterial as discussed above.

In all cases the final effective number replicas turns out to be $N_{\rm eff} > 100$, thereby guaranteeing the accuracy of the final unweighted set. All sets show good compatibility with the prior datasets. The final χ^2 values show that the reweighted set provides an essentially perfect fit to the data; the low values for LHCb are a consequence of the fact that for this experiment the correlated systematics is not available so statistical and systematic errors are added in quadrature. Before reweighting the χ^2 of individual replicas shows wide fluctuations: indeed, its average and variance over the starting replica sample are given by $\langle\chi^2\rangle = 25.6 \pm 164.4$. After reweighting the value becomes $\langle\chi^2\rangle = 1.117 \pm 0.098$, thus showing that the χ^2 of indvidual replicas has become on average almost as good as that of the central reweighted prediction.

A first assessment of the impact of the photon-induced corrections and their effect on the photon PDF can be obtained by comparing the data to the theoretical prediction obtained using pure QCD theory and the default NNPDF2.3 set, QCD⊗QED with the prior photon PDF, and QED⊗QCD with the final NNPDF2.3QED set. The

NLO				
	LHCtot	ATLAS W,Z	ATLAS high mass DY	LHCb low-mass DY
$\chi^2_{\rm in}$	2.02	1.20	3.78	2.20
$\chi^2_{\rm rw}$	1.00	1.15	1.01	0.29
$N_{\rm eff}$	287	364	326	267
$\langle\alpha\rangle$	1.41	1.24	1.53	0.89

NNLO				
	LHCtot	ATLAS W,Z	ATLAS high mass DY	LHCb low-mass DY
$\chi^2_{\rm in}$	2.01	1.37	3.44	2.06
$\chi^2_{\rm rw}$	1.08	1.21	1.00	0.66
$N_{\rm eff}$	197	297	330	363
$\langle\alpha\rangle$	1.48	1.33	1.52	1.20

Table 4.3: Reweighting parameters in the construction of the final NNPDF2.3 sets. All χ^2 values are defined as in Tab. 4.1.

	NNPDF2.3QED NLO	NNPDF2.3QED NNLO	MRST2004QED
γ; $Q^2 = 2$ GeV^2	$(0.42 \pm 0.42)\%$	$(0.34 \pm 0.34)\%$	0.30%
γ; $Q^2 = 10^4$ GeV^2	$(0.68 \pm 0.42)\%$	$(0.61 \pm 0.34)\%$	0.52%
total; $Q^2 = 2$ GeV^2	$(100.43 \pm 0.44)\%$	$(100.32 \pm 0.34)\%$	99.95%
total; $Q^2 = 10^4$ GeV^2	$(100.38 \pm 0.43)\%$	$(100.29 \pm 0.36)\%$	99.92%

Table 4.4: Momentum fractions (in percentage) carried by the photon PDF (upper two rows) and by the sum of all partons in the proton (lower two rows) in the NNPDF2.3QED NLO, NNLO and MRST2004QED PDF sets at two different scales

comparison is shown in Figs. 4.12-4.15 for the NLO sets (the NNLO results are very similar): in the left plots we show the QED+QCD prediction obtained using the prior PDF set, and in the right plots the prediction obtained using the final reweighted sets, compared in both cases to the pure QCD prediction obtained using `DYNNLO` and the NNPDF2.3 set. At the W,Z peak, the impact of QED corrections is quite small, though, in the case of neutral current production, to which the photon-photon process contributes at Born level, when the prior photon PDF is used one can see the widening of the uncertainty band due to the large uncertainty of the photon PDF of Figure 4.6. At low or high mass, as one moves away from the peak, the large uncertainty on the prior photon PDF induces an increasingly large uncertainty on the theoretical prediction, substantially larger than the data uncertainty. This means that these data do constrain the photon PDF and indeed after reweighting the uncertainty is substantially reduced.

The final NNPDF2.3QED photon PDF obtained in the NLO and NNLO fits is respectively shown at $Q_0^2 = 2$ GeV^2 in Figure 4.16 and Figure 4.17. We display individual replicas, the central (mean) photon, and the one-σ and 68% confidence level ranges, as well as the MRST2004QED result. The improvement in accuracy in comparison to

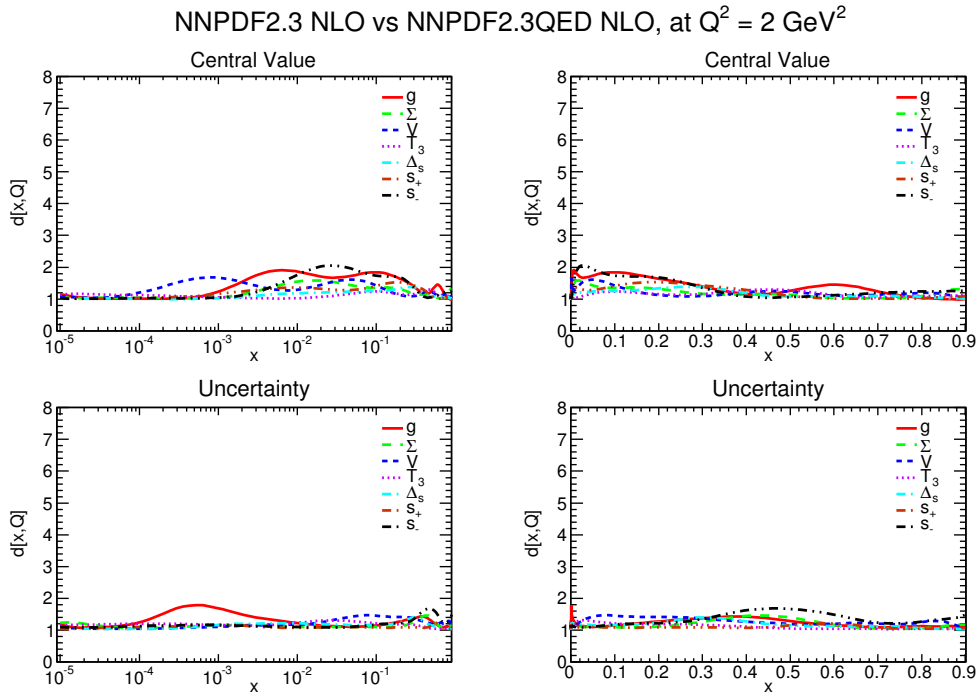

Figure 4.18: Distances between PDFs in the NNPDF2.3 and the NNPDF2.3QED NLO sets, at the input scale of Q_0^2=2 GeV^2. Distances between central values (top) and uncertainties (bottom) are shown, on a logarithmic (left) and linear (right) scale in x.

the prior PDF of Figure 4.6 is apparent, especially at small and at large x. Note also that, especially at large x, where the experimental information remains scarce (recall Figure 4.11), the positivity bound still plays an important role in constraining the photon PDF. Indeed, at the starting scale Q_0 the lower edge of the uncertainty band (determined as discussed in Sect. 4.1) is again very close to the positivity constraint, and consequently, even after having used the LHC data, the probability distribution of the photon PDF is significantly asymmetric, departing substantially from Gaussian. This should be kept in mind in phenomenological applications, in particular when computing uncertainties.

In Table 4.4 we show the momentum fraction carried by the photon PDF in NNPDF2.3QED at NLO and NNLO, both at a low and high scale: it is about half of a percent, compatible with zero within uncertainties, and mildly dependent on scale. The MRST2004QED values, also shown, are consistent within uncertainties. Note that the standard deviation would be almost twice the 68% confidence level interval given in the table. We also give the total momentum, which deviates from unity because of the slightly inconsistent procedure that we have followed in constructing the prior set, by combining the photon from a fit to DIS data with the other PDFs from the global NNPDF2.3 fit as discussed in Sect. 4.2.1 above. We also see that the total momentum fraction is not quite scale independent, because of the approximation introduced when

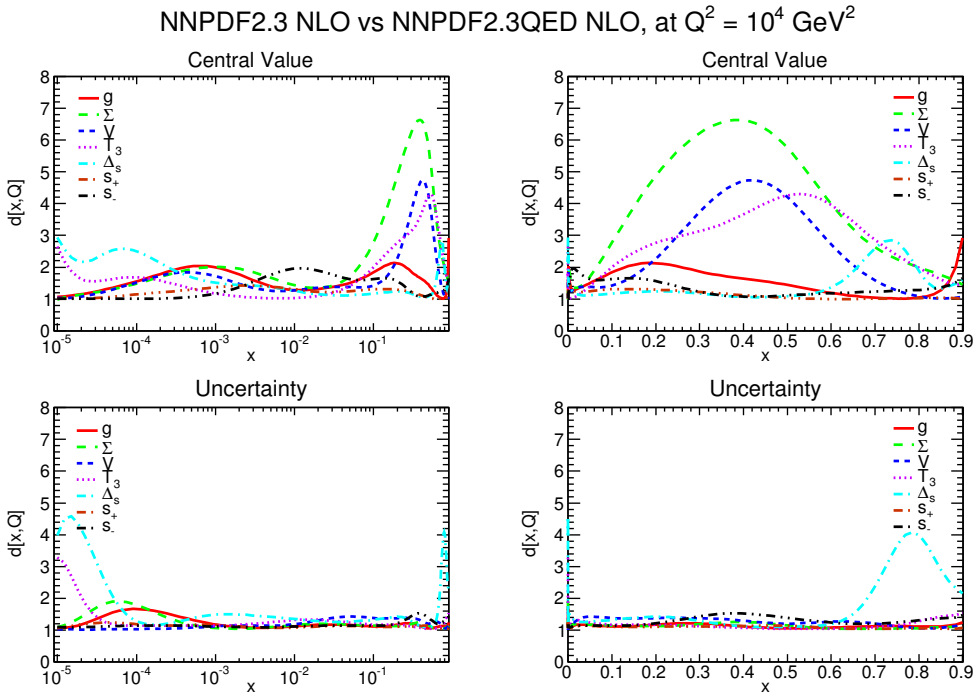

Figure 4.19: Same as Fig. 4.18 but now computed at $Q^2 = 10^4$ GeV^2.

neglecting terms of $\mathcal{O}(\alpha\alpha_s)$ in the solution of the combined QED⊗QCD evolution equations. Both effects are well below the 1% level.

All other PDFs at the initial scale Q_0 are left unaffected by the reweighting. This can be seen by computing the distances between PDFs in the starting NNPDF2.3 set and in the final NNPDF2.3QED set; they are displayed in Figure 4.18, at the scale $Q_0^2 = 2$ GeV^2 at which PDFs are parametrized: it is apparent that the distances are compatible with statistically equivalent PDFs. It is interesting to repeat the same comparison at $Q^2 = 10^4$ GeV^2 (Figure 4.19): in this case, statistically significant differences start appearing, as a consequence of the fact that the statistically equivalent starting PDFs in the two sets are then evolved respectively with and without QED corrections. However, the differences are below the one-σ level (and concentrated at large x), consistent with the conclusion that the new data are compatible with those used for the determination of the NNPDF2.3 PDF set.

In Figs. 4.16-4.17 the photon PDF from the MRST2004QED set is also shown for comparison. The MRST2004QED photon PDF is based on a model; an alternative (not publicly available) version of it, in which consituent rather than current quark masses are used as model parameters, has been used [25] to estimate the model uncertainty, though consituent masses are considered to be less appropriate by the authors of Ref. [109]. The MRST2004QED photon turns out to be in good agreement with the central NNPDF2.3QED prediction at medium and large x, but at small $x \lesssim 0.03$ it grows more quickly, and for $x \leq 10^{-2}$ it is larger and well outside the NNPDF2.3QED

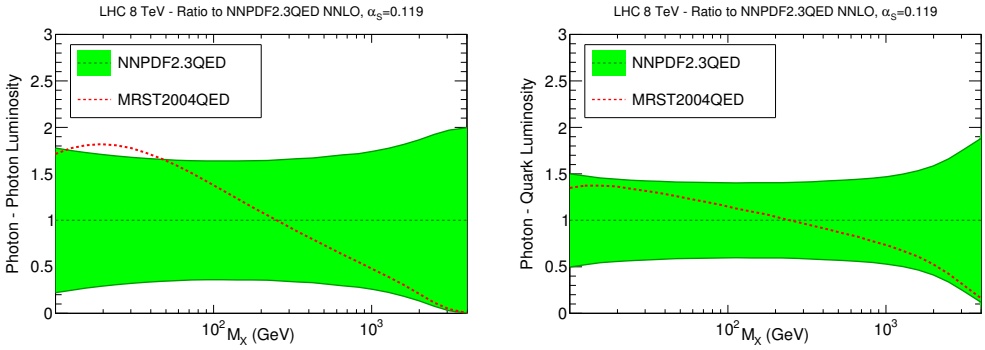

Figure 4.20: The photon-photon $\gamma\gamma$ (left) and photon-quark γq (right) parton luminosities at the LHC 8 TeV computed using MRST2004QED PDFs, shown as a ratio to the NNPDF2.3QED result. The 68% confidence level on the latter is also shown.

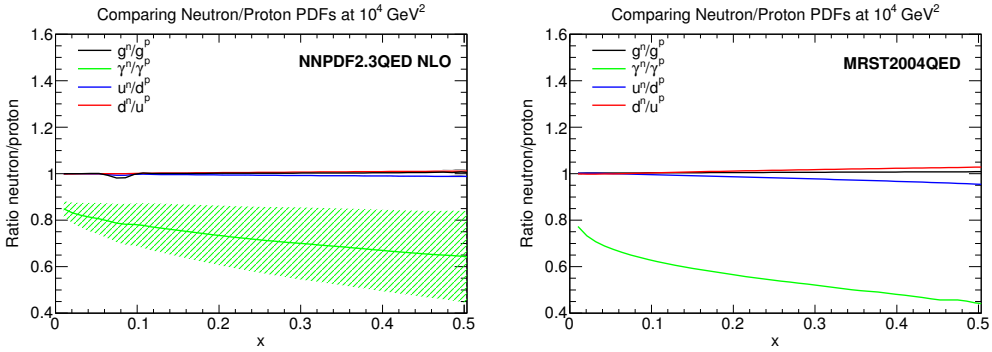

Figure 4.21: The ratio of the neutron to the proton PDFs in the NNPDF2.3QED NLO set at $Q^2 = 10^4$ GeV^2 (left) and MRST2004QED set (right). Results for the photon, gluon, up and down quark are shown. Error bands correspond to one-σ uncertainties.

uncertainty band.

It is also interesting to compare the NNPDF2.3QED and MRST2004QED sets at the level of the parton luminosities which enter the computation of hadronic processes. This comparison is shown in Figure 4.20. The two luminosities are in good agreement for invariant masses of the final state $M_X \sim 100$ GeV, but the agreement is less good for higher or lower final-state masses, with the MRST2004QED rather smaller at high mass and larger at low mass, where, for $M_X \sim 20$ GeV it is outside the NNPDF2.3QED uncertainty band. As we will see in the next section, these differences translate into differences in the predictions for electroweak processes at the LHC.

So far, we have shown results for the PDFs of the proton. Note, however that, as discussed in Sect. 4.1, even though we assume that isospin holds at the scale at which PDFs are parametrized, QED corrections to perturbative evolution introduce a violation of the isospin symmetry at all other scales. Therefore, we provide independent NNPDF2.3QED PDF sets for proton and neutron. The size of isospin violation is

expected to be comparable to the QED corrections themselves, so very small for quark and gluon distributions but more significant for the photon PDF. The expectation is borne out by Figure 4.21 where the ratio of the neutron to the proton PDF at $Q^2 = 10^4$ GeV^2 in NNPDF2.3QED NLO is compared to that in MRST2004QED set. The comparison shows that while the amount of isospin violation in the MRST2004QED photon PDF, which had a built-in model of non-perturbative isospin violation, is somewhat larger than our own, especially at large x, the difference is within the PDF uncertainty, as anticipated in Sect. 4.1.1. The amount of isospin violation on quark and gluon PDFs is extremely small, on the scale of PDF uncertainties, both for MRST2004QED and NNPDF2.3QED. The same conclusions hold if the NNLO set is used.

Chapter 5

Phenomenological implications of the photon PDF

In this chapter we investigate some examples of the use of the NNPDF2.3QED PDF set. We analyze several processes which are sensitive to photon-initiated contributions. In particular, we will start with the discuss of direct photon production at HERA, and then we show results about searches for new massive electroweak gauge bosons and W pair production at small p_T and large invariant mass, at LHC energies. After presenting the phenomenological impact for these processes, we then show details about availability of these sets of PDFs in Monte Carlo event generators. Finally, we conclude this chapter with a first determination of lepton PDFs using the `APFEL` evolution and sets with photon PDFs.

5.1 Photon-induced processes

5.1.1 Direct photon production at HERA

Deep-inelastic isolated photon production provides a direct handle on the photon parton distribution of the proton, through Compton scattering of the incoming electron off the photon component of the proton [198]. At the leading log level, this $\mathcal{O}(\alpha^2)$ partonic subprocess is the only contribution. In practice, however, the $\mathcal{O}(\alpha^3)$ quark-induced contributions [199] may be comparable (as for the Drell-Yan process discussed in Sect. 4.2) because of the larger size of the quark distribution. In Ref. [109], the total cross-section for this process computed at the leading log level using MRST2004QED PDFs was shown to be in reasonable agreement with HERA integrated cross-sections for prompt photon production data [200].

However, more recent HERA data [197] for the rapidity and transverse energy distribution of the photon do not agree well with either the fixed order [199] or the leading log [109,198] results for all values of the kinematics, suggesting that a calculation matching the leading-log resummation to the fixed order result would be necessary in order to obtain good agreement. In the absence of such a calculation, we did not use these data for the determination of the photon PDF.

Theoretical predictions obtained using the leading log calculation [109] and the NNPDF2.3QED or MRST2004QED PDF sets are compared in Fig. 5.1 to the ZEUS

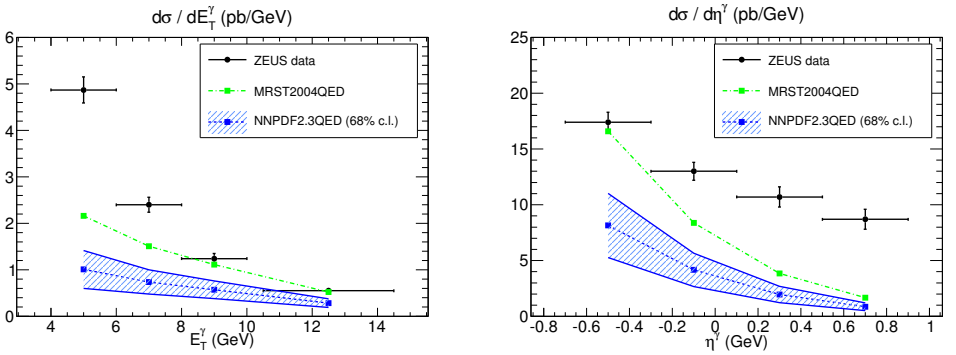

Figure 5.1: Comparison between the ZEUS data [197] for the photon transverse energy (left) and rapidity (right) distributions in deep-inelastic isolated photon production and the leading log theoretical prediction obtained using NNPDF2.3QED and MRST2004QED PDFs.

data of Ref. [197]. These predictions have been obtained using the code of Ref. [109]. The selection cuts are the same as in [197], namely

$$10 \leq Q^2 \leq 300 \text{ GeV}^2\,, \quad 4 \leq E_T^\gamma \leq 15 \text{ GeV}\,, \quad -0.7 \leq \eta^\gamma \leq 0.9\,. \tag{5.1}$$

The fact that the prediction is in better agreement with the data at large E_T is consistent with the expectation that the leading log approximation which is being used is more reliable in this region. However, as already mentioned, a fully matched calculation would be needed in order to consistently combine the leading log and fixed order results.

5.1.2 Searches for new massive electroweak gauge bosons

Heavy electroweak gauge bosons, denoted generically by W' and Z', have been actively searched at the LHC (see e.g. [201–204]), with current limits for $M_{V'}$ between 1 and 2 TeV depending on the model assumptions. The main background for such searches is the off-resonance production of W and Z bosons respectively. At such large invariant masses of the dilepton pair, photon-induced contributions, of the type shown in Figs. 4.9–4.10, are potentially large.

We have thus computed the theoretical predictions for high mass off-shell W and Z production using NNPDF2.3QED. We have calculated separately the $q\bar{q}$ initiated Born contributions, the Born term supplemented by photon-initiated processes, and the full set of $\mathcal{O}(\alpha)$ QED corrections, all determined with `HORACE` (hence using LO QCD theory) and the various electroweak scheme choices discussed in Sect. 4.2.2. We have used the following kinematical cuts, roughly corresponding to those used in the ATLAS and CMS searches

$$p_t^l \geq 25 \text{ GeV}\,, \quad |\eta^\gamma| \leq 2.4\,, \tag{5.2}$$

and we have generated enough statistics to properly populate the highest mass bins and reduce the impact of statistical fluctuations. Results are displayed in Fig. 5.2, for the

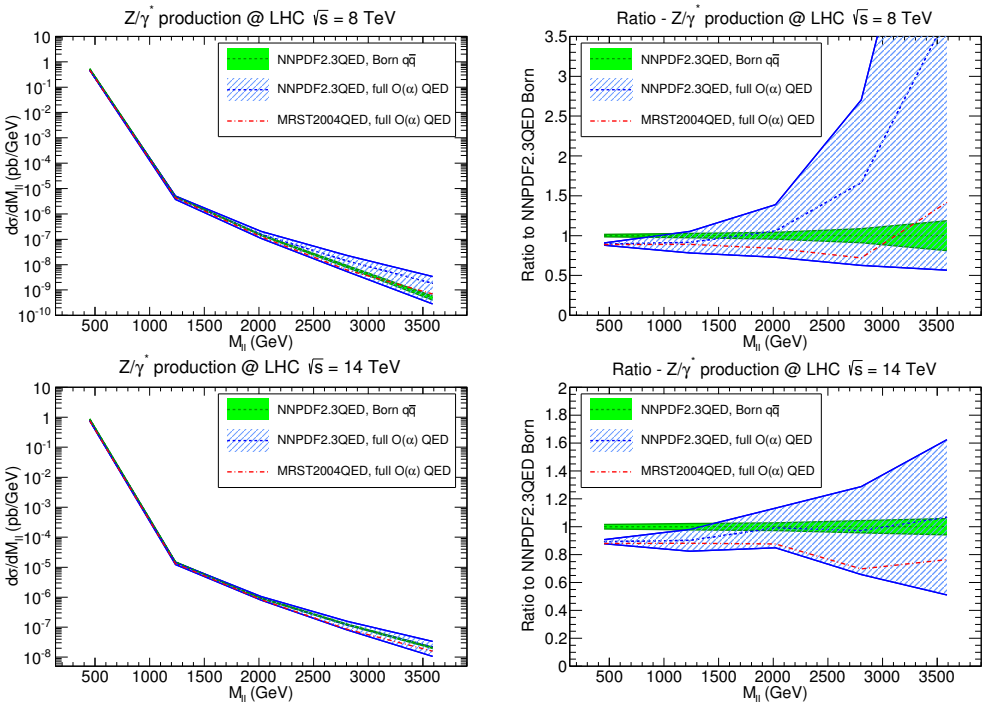

Figure 5.2: Neutral current Drell-Yan production at the LHC as a function of the invariant mass of the dilepton pair using NNPDF2.3QED and MRST2004QED PDFs. Theoretical predictions for the Born $q\bar{q}$ and the full $\mathcal{O}(\alpha)$ process (including photon-induced contributions) at the LHC 8 TeV (top) and LHC 14 TeV (bottom), are shown both on an absolute scale (left) or as a ratio to the central value of the Born $q\bar{q}$ cross-section from NNPDF2.3QED.

neutral-current and in Fig. 5.3 for charged-current dilepton production respectively. They are provided for LHC 8 TeV and LHC 14 TeV, shown both in an absolute scale and as a ratio to the central value of the Born $q\bar{q}$ cross-section from NNPDF2.3QED, using the NLO set.

The contribution from the photon-induced diagrams is generally not negligible. Especially in the neutral current case, in which the photon-induced contribution starts at Born level, the uncertainty induced by the QED corrections in the large invariant mass region is substantial, because the LHC data we used to constrain the photon PDF (recall in particular Tab. 4.2 and Fig. 4.11) have little effect there: the uncertainty is of order 20% for $M_{ll} \sim 1$ TeV at LHC 8 TeV, and it reaches the 50% level for $M_{ll} \sim 2$ TeV. Of course, for a given value of M_{ll}, the photon-induced uncertainties decrease when going to 14 TeV, since smaller values of x are probed, closer to the region of the data used for the current PDF determination.

Currently, the uncertainty on QED corrections is typically estimated by varying the photon PDF between its MRST2004QED value and zero. Our results suggest that this might underestimate the size of the photon-induced contribution; it certainly

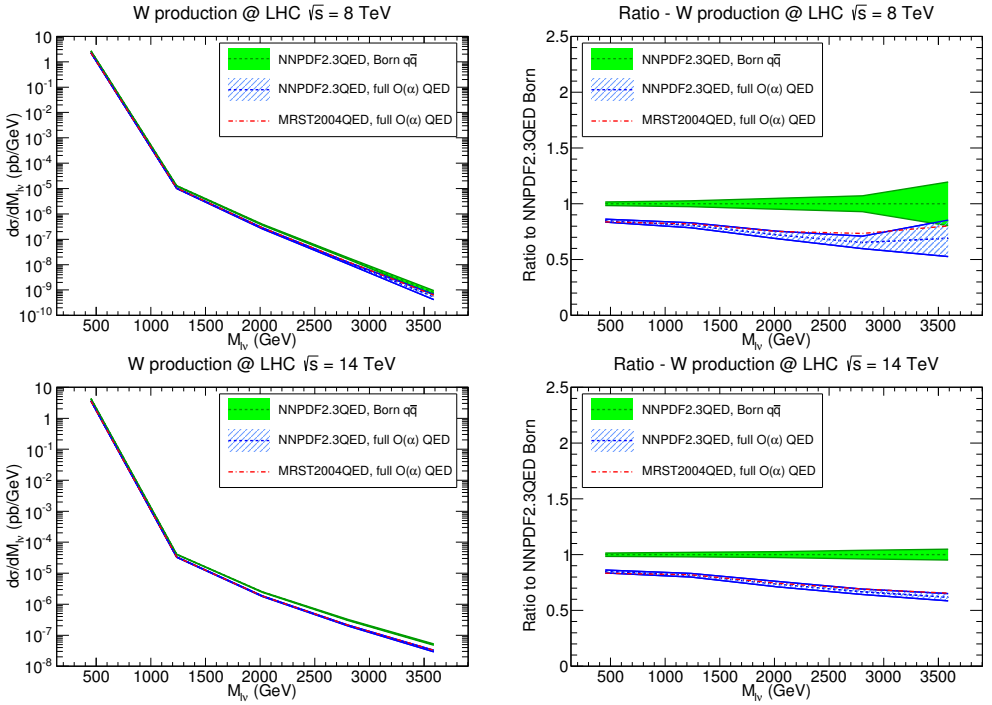

Figure 5.3: Same as Fig. 5.2 but for high-mass charged-current production.

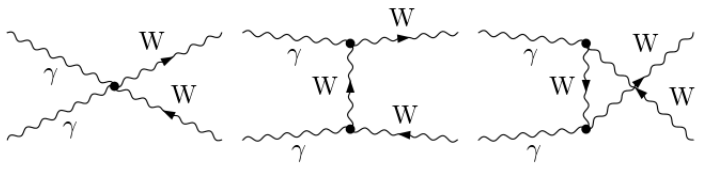

Figure 5.4: Tree-level diagrams for the LO processes $\gamma\gamma \to \mathrm{W}^-\mathrm{W}^+$, from Ref. [87].

does underestimate the uncertainty related to our current knowledge of it. This follows directly from the behavior of the luminosities of Fig. 4.20. In order to obtain more reliable exclusion limits for Z' and W' at the LHC, a more accurate determination of the photon PDF at large x might be necessary. This could come from the inclusion in the global PDF fit of new observables that are particularly sensitive to the photon PDF, such as W pair production, as we now discuss.

5.1.3 W pair production at the LHC

The production of pairs of electroweak gauge bosons is important, specifically for the determination of triple and quartic gauge boson couplings [205–207], and it is a significant background to searches [208–212] since several extensions to the Standard Model including warped extra dimensions [213] and dynamical electroweak symmetry-

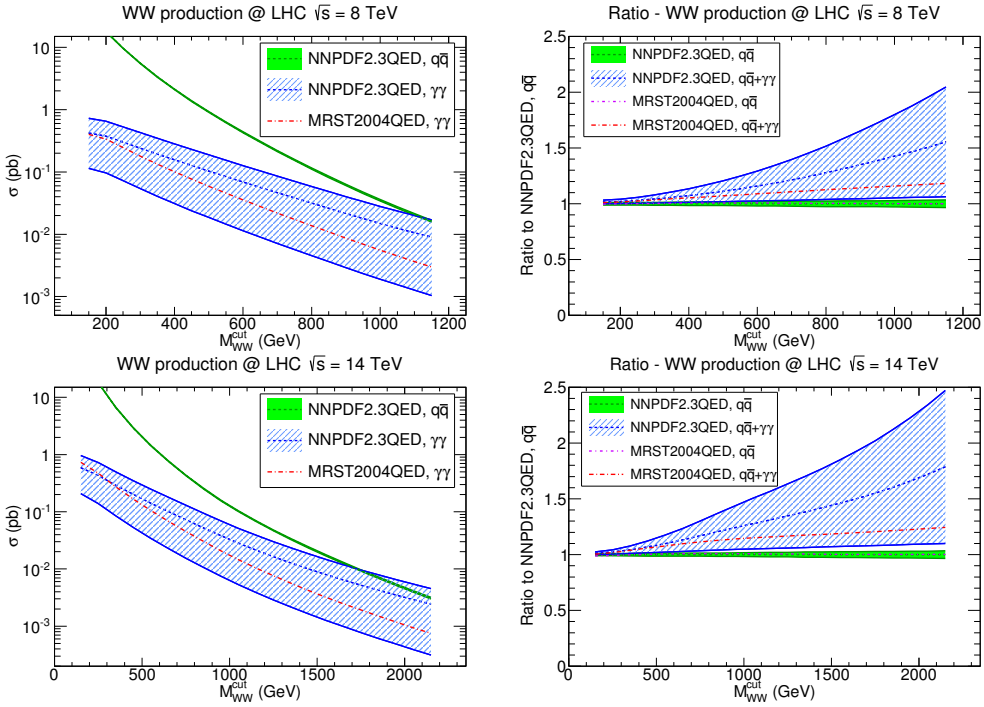

Figure 5.5: Photon-induced and quark-induced Born-level contributions to the production of a W pair with mass $M_{WW} > M_{WW}^{\rm cut}$ plotted as a function of $M_{WW}^{\rm cut}$ at the LHC 8 TeV (top) and LHC 14 TeV (bottom), computed with the code of Ref. [87] and NNPDF2.3QED NLO and MRST2004QED PDFs.

breaking models [214, 215] predict the existence of heavy resonances decaying to pairs of electroweak gauge bosons.

We consider now specifically the production of W boson pairs for large values of the invariant mass M_{WW} and moderate values of the transverse momentum $p_{T,W}$. Photon-induced contributions to this process start at Born level (see Fig. 5.4), and their contribution can be substantial, in particular at large values of M_{WW}. NLO QCD corrections, as well as the formally NNLO but numerically significant gluon-gluon initiated contributions, are known, and available in public codes such as `MCFM` [216]. Fixed-order electroweak corrections to W pair production are also known [87], as well as the resummation of large Sudakov electroweak logarithms at NNLL accuracy [217]; a recent review of theoretical calculations is in Ref. [86].

To estimate the impact of photon-induced contributions to WW production, predictions have been computed with either MRST2004QED or NNPDF2.3QED NLO PDFs. They have been provided by the authors of Ref. [87] using the code and settings of Ref. [87]. In particular, the kinematical cuts in the transverse momentum and rapidity of the W bosons are

$$p_{T,W} \geq 15 \text{ GeV}\,, \quad |y_W| \leq 2.5\,. \tag{5.3}$$

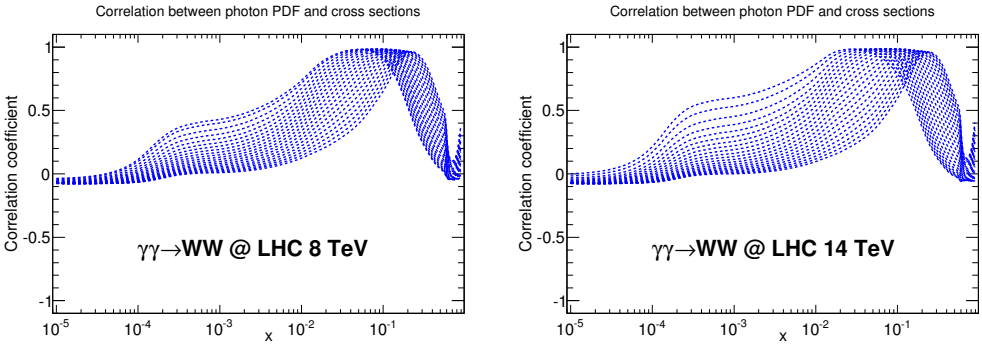

Figure 5.6: Correlations between the W pair production cross-section of Fig. 5.5 and the photon PDF from the NNPDF2.3QED NLO set for $Q = 10^4$ GeV^2. Each curve corresponds to one of 40 equally spaced bins in which the $M_{WW}^{\rm cut}$ range of Fig. 5.5 has been subdivided.

In Fig. 5.5 the cross-section for production of a W pair of mass $M_{WW} > M_{WW}^{\rm cut}$ is displayed as a function of $M_{WW}^{\rm cut}$, at the LHC 8 and 14 TeV. The Born $q\bar{q}$ and $\gamma\gamma$ initiated contributions are shown (computed using LO QCD), while we refer to Ref. [87] for the full $\mathcal{O}(\alpha)$ electroweak corrections, which depend only weakly on the photon PDF. It is clear that for large enough values of the mass of the pair the photon-induced contribution becomes increasingly important. Again, the relative size of the results obtained using NNPDF2.3QED or MRST2004QED PDFs can be inferred from the behavior of the luminosities shown in Fig. 4.20.

As in the case of Fig. 5.2, the large uncertainties found for large values of $M_{WW}^{\rm cut}$ reflect the lack of knowledge on the photon PDF at large $x \gtrsim 0.1$. Indeed, in Fig. 5.6 we display the correlation between the cross-section of Fig. 5.5 and the photon PDF at $Q^2 = 10^4$ GeV^2 as a function of x, obtained subdividing the range of $M_{WW}^{\rm cut}$ of Fig. 5.5 into 40 bins of equal width, and then computing the correlation for each bin. It is clear that this process is sensitive to the photon PDF at large x, where the data of Tab. 4.2 provide little or no constraint (recall Fig. 4.11). Hence, predictions for W pair production obtained using MRST2004QED or NNPDF2.3QED should be taken with care: NNPDF2.3QED provides a more conservative estimate of the uncertainties involved, but perhaps overestimates the range of reasonable photon PDF shapes. However, future measurements of this process could be used to pin down the photon PDF at large x, and thus in turn improve the accuracy of the prediction for very high mass Drell-Yan production discussed in Sect. 5.1.2 and Fig. 5.2, and conversely. Of course, in using either, or both of these channels for new physics searches, care should be taken that the sought-for new physics effects are not being hidden in the PDFs themselves, which could be done by introducing suitable kinematic cuts.

5.1.4 Disentangling electroweak effects in Z-boson production

In this section we estimate and compare to the PDF uncertainties the contributions to the invariant mass of the Drell-Yan Z-boson production due to electroweak corrections

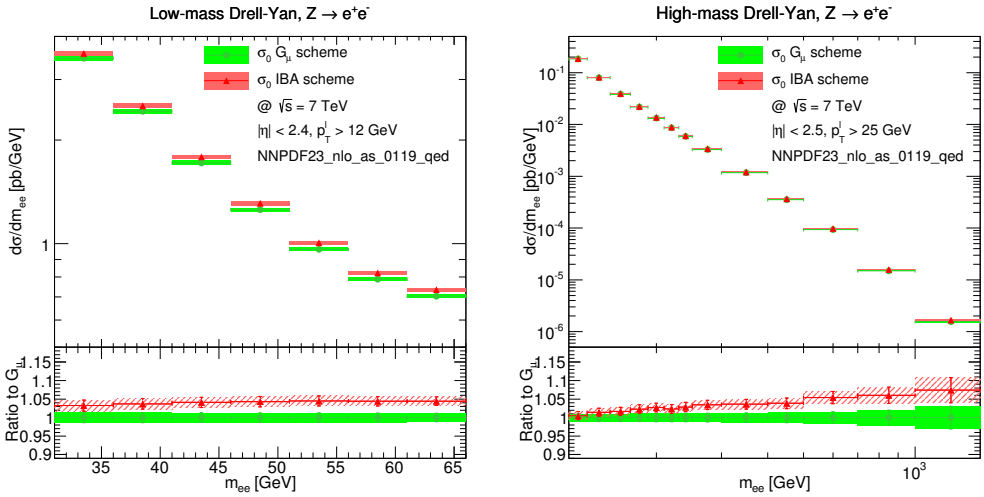

Figure 5.7: Born level predictions and respective ratios for low- (left) and high-mass (right) Drell-Yan, $Z \to e^+e^-$ production, using the IBA and the G_μ scheme. At low-mass there is a constant gap of 3-4% for all bins, while at high-mass, predictions increase progressively with the invariant mass, producing discrepancies of 7-8% in the last bin.

and the photon-induced channel, by considering the low-mass region, which is below the Z peak resonance and the high-mass tail.

In contrast to what was shown in Ref. [218] where predictions were computed with `FEWZ`, here we propose to combine two distinct parton level public codes: `DYNNLO` [219] for the NLO QCD prediction and `HORACE` [5] which provides the exact $\mathcal{O}(\alpha)$ electroweak radiative correction together with the photon-induced channel for the Z production. The motivation for this combination is the interest to measure the difference between predictions with electroweak effects at NLO/NNLO QCD accuracy computed in the improved Born approximation (IBA) instead of using electroweak correction computed by `FEWZ` in the G_μ scheme. The main difference between these choices is that effective couplings in the IBA reabsorb higher-order electroweak corrections and therefore it provides predictions in better agreement with experimental data.

Computations are performed exclusively with the NNPDF2.3QED set of PDFs with $\alpha_s = 0.119$, instead of using the respective LO and NNLO sets because here we will focus only on the NLO QCD accuracy and that is why we use a NLO set.

In the next sections, we first show the differences at Born level between the improved Born approximation (IBA), available in `HORACE` by default, and the G_μ scheme in `DYNNLO`, then, we proceed with the construction of the full prediction.

Comparing the improved Born approximation (IBA) with the G_μ scheme

In order to obtain realistic results, which are ready for comparisons with real data, we have selected the kinematic range and cuts inspired by recent measurements performed by the ATLAS experiment for low- and high-mass Drell-Yan differential cross-section

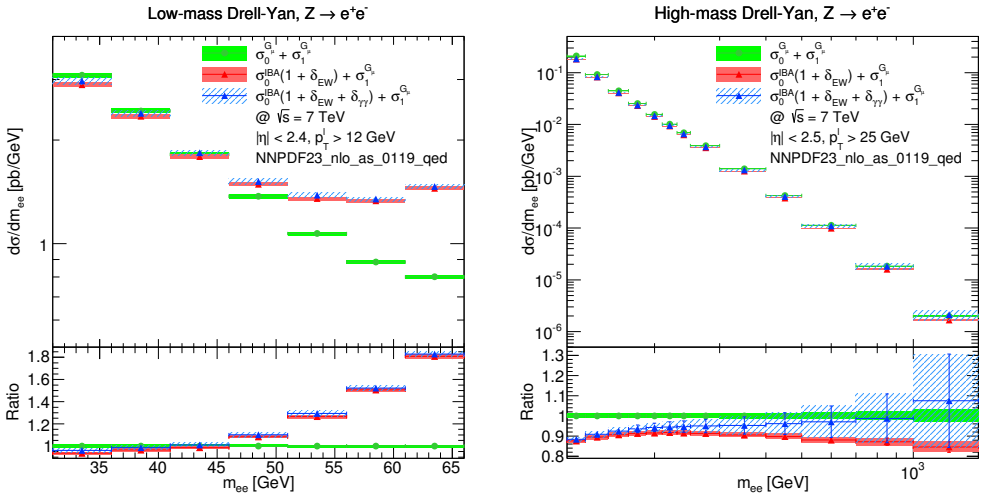

Figure 5.8: Comparison of predictions and respective ratios for low- (left) and high-mass (right) Drell-Yan, $Z \to e^+e^-$ production. We compare the NLO QCD prediction provided by DYNNLO (green distribution) with: the combined prediction with $\delta_{\rm EW}$ (red distribution) and with the $\delta_{\rm EW} + \delta_{\gamma\gamma}$ (blue distribution).

at $\sqrt{s} = 7$ TeV [24, 220].

Figure 5.7 shows the predicted distribution at Born level using the IBA (HORACE) and the G_μ scheme (DYNNLO) at low (left plot) and high (right plot) invariant mass regions, for the Drell-Yan process: $Z \to e^+e^-$. Here, the goal is to measure the numerical differences due to the choice of these methodologies.

For all distributions, the Monte Carlo uncertainty is below the percent level. The uncertainties shown in the figure have been calculated as the one-σ interval obtained after averaging over the 100 replicas provided by this set.

In the low-mass region, we have applied kinematic cuts to the lepton pair imposing: $p_T^l > 12$ GeV and $|\eta^l| < 2.4$ as in ATLAS [220]. In this region we observe an almost flat gap of 3-4% between the IBA and G_μ predictions, however in the bin $m_{ee} = 51-56$ GeV the difference is slightly higher.

On the other hand, in the high-mass region we have applied the following kinematic cuts: $p_T^l > 25$ GeV and $|\eta^l| < 2.5$ as in Ref. [24]. We observe a progressive increase of the central value prediction as a function of the invariant mass, reaching a maximum of 7-8% at the highest bin in m_{ee}. This suggests that the running of $\alpha(Q^2)$ in the IBA can play a crucial role when determining with accuracy the predictions in such region.

It is important to highlight that in both cases, PDF uncertainties are smaller than the observed differences induced by the choice of the scheme. These results are fully consistent with the IBA implementation discussed in Ref. [5]. In the sequel we are interested in combining electroweak effects with higher order QCD corrections in the IBA and then compare these results to pure QCD G_μ predictions.

Disentangling electroweak effects

At this point, we are interested in building a prediction based on IBA which includes NLO QCD with $\mathcal{O}(\alpha)$ correction and the photon-induced channel. We propose to extract the NLO correction from DYNNLO by removing its Born level, which contains the direct and strong dependence on the G_μ scheme, and combine the result with the HORACE prediction. Schematically this can be achieved by defining the quantities:

$$\sigma_{\texttt{DYNNLO}} = \sigma_0^{G_\mu} + \sigma_1^{G_\mu}, \tag{5.4}$$

$$\sigma_{\texttt{HORACE}} = \sigma_0^{\text{IBA}}(1 + \delta_{\text{EW}} + \delta_{\gamma\gamma}), \tag{5.5}$$

where σ_0^{IBA} and $\sigma_0^{G_\mu}$ are the Born levels presented in Figure 5.7, $\sigma_1^{G_\mu}$ the NLO QCD, δ_{EW} the $\mathcal{O}(\alpha)$ electroweak correction and $\delta_{\gamma\gamma}$ the photon-induced contribution.

The combination is then constructed in the following way:

$$\sigma_{\text{Total}} = \sigma_{\texttt{DYNNLO}} + \sigma_{\texttt{HORACE}} - \sigma_0^{G_\mu} \tag{5.6}$$

$$= \sigma_0^{\text{IBA}} + \sigma_0^{\text{IBA}}\delta_{\text{EW}} + \sigma_0^{\text{IBA}}\delta_{\gamma\gamma} + \sigma_1^{G_\mu}. \tag{5.7}$$

where we remove the DYNNLO Born level while we include the NLO QCD correction in the final prediction.

We are aware that using this methodology we improve the combination but we do not remove entirely the pure G_μ dependence at higher orders, however this is the best combination we can propose without applying technical modifications to both codes.

In Figure 5.8 we compare $\sigma_{\texttt{DYNNLO}}$ with σ_{Total}, the combination presented in Eq. 5.7, with and without the $\delta_{\gamma\gamma}$ term. For all distributions we compute the one-σ uncertainty except when including the photon-induced channel where we have used the 68% c.l. as in Ref. [221].

In the low-mass region the inclusion of $\mathcal{O}(\alpha)$ electroweak corrections has a strong impact on the last four bins, where differences can reach $\sim 80\%$ in comparison to the pure NLO QCD G_μ prediction, while the same correction for the high-mass distribution shows a moderate impact which is below $\sim 20\%$ for the highest invariant mass bin. This behavior is expected and derives from the shape of the Z-boson invariant mass: bins located in a region lower than the Z peak resonance undergoes large positive corrections while at high invariant mass we observe a change of sign of such corrections. It is important to highlight that modern data provided by the LHC experiments are already corrected by final-state photon radiation which carries a dominant fraction of the electroweak effects shown in Figure 5.8.

The photon-induced contribution has a moderate impact in the low-mass region while for high-mass it is dominant: this behavior is expected and due to the presence of the Z peak resonance where the photon-induced channel is negligible.

Also from these plots of Figure 5.8, it is important to emphasize again that modern PDF sets, as the NNPDF2.3QED, have uncertainties which are accurate enough to appreciate the differences due to scheme choices and electroweak effects, including the new photon PDF, which shows a stable behavior of uncertainties in all invariant mass regions except at very high-mass bins where uncertainties grow, reaching more than $\sim 20\%$. This situation will be improved in future by including more relevant and precise data to constrain the photon PDF.

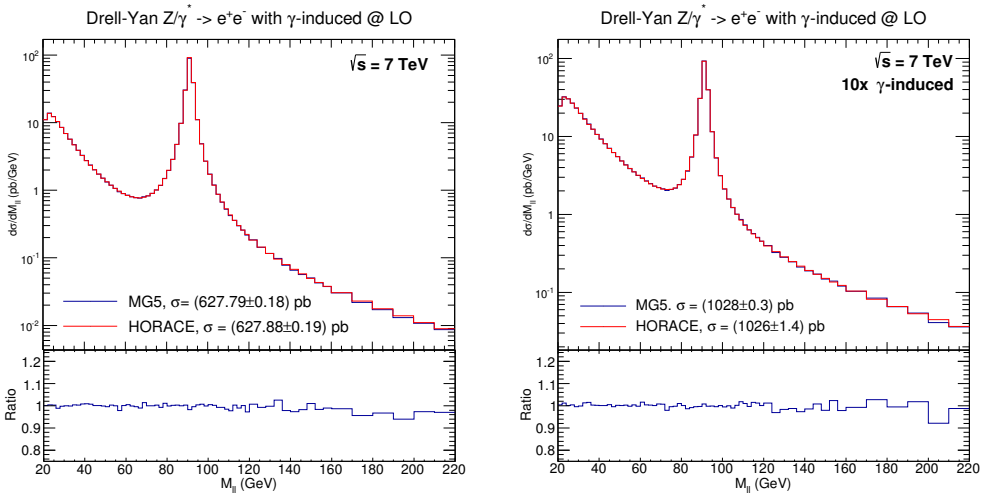

Figure 5.9: Comparison of HORACE and the MadGraph5_aMC@NLO implementation of the invariant mass for Drell-Yan Z/γ^* production at leading order with photon-induced contributions. The left plots shows predictions with the NNPDF2.3QED set of PDFs at $\sqrt{s} = 7$ TeV. On the right plot the photon PDF contribution is multiplied by a factor 10, in order to enlarge the photon-induced contributions and emphasize the good level of agreement.

5.2 Photon PDF in Monte Carlo event generators

After the release of the NNPDF2.3QED set of PDFs, several Monte Carlo event generators have implemented the possibility to activate photon-induced channels when computing predictions. We have released a fast standalone public code for the manipulation of these sets of PDFs independently of the LHAPDF library[1]. This code written in C++ and Fortran77 has been adapted and implemented in the core of the following Monte Carlo event generators: PYTHIA8 [222], MadGraph5_aMC@NLO [6] and SHERPA [223].

As an example, the NNPDF2.3QED set of PDFs can be used since PYTHIA8.1, where, in this release, presented in Ref. [222], we determine the updated fragmentation parameters with this new set of PDFs. We use minimum-bias, Drell-Yan, and underlying-event data from the LHC to constrain the initial-state-radiation and multi-parton-interaction parameters, combined with data from SPS and the Tevatron to constrain the energy scaling. Several distributions show significant improvements with respect to the current defaults, for both *ee* and *pp* collisions, though we emphasize that interesting discrepancies remain in particular for strange particles and baryons.

Another example of implementation of this set of PDFs in a MC event generator is displayed in Figure 5.9, where we show an example of a benchmarking comparison between HORACE and MadGraph5_aMC@NLO. In this figure, we compute the LO invariant mass of the Z/γ^* Drell-Yan production at $\sqrt{s} = 7$ TeV. On the left plot, we estimate

[1]The public code is available at https://github.com/scarrazza/nnpdfdriver

the distribution using the NNPDF2.3QED set of PDFs including photon-induced channels. The level of agreement is good in the peak and in the tails regions. However as the photon-induced contributions are small, at least two orders of magnitude smaller than the total inclusive cross-section, on the right plot we show the same process but now computed with the photon PDF from NNPDF2.3QED multiplied by a factor 10. This plot shows that the agreement is still good, confirming that the implementation in `MadGraph5_aMC@NLO` is correct.

This is an extremely important result, because it opens the possibilty of implementing a fast NLO interface for computations including electroweak corrections, and thus the photon-induced contributions through the `aMCfast` [131] code. With this code, we will be able to generate sets of `APPLgrid` [128] tables with weights associated to the photon contribution, and consequently, enabling the possibility to perform new fits of PDFs with QED corrections, improving the determination and uncertainties of the photon PDF by including more data in the fit and avoiding the reweighting strategy explained in Chap. 4.

5.3 Lepton PDFs

In the previous chapters, we have always neglected the PDFs of charged leptons, supposing that with the current methodology their determination is practically impossible from a fit to the available experimental data. In fact, the PDFs associated to $e^{\pm}$, $\mu^{\pm}$ and $\tau^{\pm}$ are expected to be much smaller than the photon PDF. However, when computing electroweak corrections to some hadron-collider processes, such as Drell-Yan, the presence of lepton PDFs requires the inclusion of new lepton-initiated channels which might have a non-negligible impact.

Currently, from literature we observe that only the photon content of the proton has been determined based either on model assumptions [109], the MRST2004QED set, or on a fit to data [221], the NNPDF2.3QED, but no attempt to estimated the lepton PDFs has ever been tried.

Therefore, for the conclusion of this chapter, we propose to give an estimate on the leptonic content of the proton. This will be achieved in the following steps:

- the implementation of the lepton PDF DGLAP evolution at LO in QED in the so called VFN scheme in `APFEL` [9];
- the determination of a guess for the lepton PDFs at the initial scale Q_0, based on the assumption that leptons are generated by photon splitting.

In the next sections, we discuss the details of both steps.

5.3.1 DGLAP Equation in the Presence of Photon and Leptons

Following the methodology presented in Chap. 2, we extend the DGLAP equations to include the evolution of photon and lepton PDFs at LO in QED. At LO in QED, leptons couple directly only to the photon. However, since the photon couples to quarks and so, indirectly to gluon, the lepton PDFs evolution depend on the evolution of all the other partons. Following the notation of Sect. 2.2, where QCD and QED evolutions are treated independently, the inclusion of leptons does not imply any change to the

QCD evolution, while QED evolution equations are modified with the inclusion of the leptonic terms in the photon evolution, together with the addition of lepton equations, namely

$$\begin{aligned}
\nu^2\frac{\partial\gamma}{\partial\nu^2} &= \frac{\alpha(\nu)}{4\pi}\left[\left(\sum_i N_c e_i^2\right)P_{\gamma\gamma}^{(0)}\otimes\gamma + \sum_i e_i^2 P_{\gamma q}^{(0)}\otimes(q_i+\bar{q}_i) + \sum_j P_{\gamma\ell}^{(0)}\otimes(\ell_j+\bar{\ell}_j)\right], \\
\nu^2\frac{\partial\ell_j}{\partial\nu^2} &= \frac{\alpha(\nu)}{4\pi}\left[P_{\ell\gamma}^{(0)}\otimes\gamma + P_{\ell\ell}^{(0)}\otimes\ell_j\right], \\
\nu^2\frac{\partial\bar{\ell}_j}{\partial\nu^2} &= \frac{\alpha(\nu)}{4\pi}\left[P_{\ell\gamma}^{(0)}\otimes\gamma + P_{\ell\ell}^{(0)}\otimes\bar{\ell}_j\right],
\end{aligned} \tag{5.8}$$

where γ, q_i, $\bar{q}_i$, ℓ_j and $\bar{\ell}_j$ are respectively the PDFs of the photon, the i-th quark, the i-th antiquark, j-th lepton and j-th anti-lepton, e_i the i-th quark electric charge, $N_c = 3$ the number of colors and α the running fine structure constant. Note also that the indices i and j in the first line of eq. (5.8) run over the n_f and n_ℓ number of active quarks and leptons at the scale ν, respectively. Note that the leading-order QED splitting functions satisfy the following identities: $P_{q\gamma}^{(0)} = P_{\ell\gamma}^{(0)}$, $P_{\gamma q}^{(0)} = P_{\gamma\ell}^{(0)}$ and $P_{qq}^{(0)} = P_{\ell\ell}^{(0)}$.

Combining the system of differential equations in eq. (5.8) with the pure-QCD DGLAP equations that govern the evolution of gluon and quarks, we obtain the full QCD⊗QED evolution in the presence of photon and leptons. We have implemented the solution of this system of differential equations in `APFEL` version 2.4.0.

Here, the fine-structure constant α runs with the renormalization scale that we take to be equal to the factorization scale μ_F. Consistently with evolution of PDFs, we consider the leading order running by solving the RG equation

$$\nu^2\frac{d\alpha}{d\nu^2} = \beta_{\rm QED}^{(0)}\alpha^2(\nu) \tag{5.9}$$

where

$$\beta_{\rm QED}^{(0)} = \frac{8}{12\pi}\left(N_c\sum_{i=1}^{n_f}e_i^2 + n_\ell\right), \tag{5.10}$$

with $N_c = 3$ the number of colors, e_i^2 the electric charge of the i-th quark, n_f the number of light quarks and n_ℓ the number of light leptons. Finally, as a boundary condition for the evolution we take $\alpha^{-1}(m_\tau) = 133.4$ as in Chap. 2.

5.3.2 Modeling the Lepton PDFs

The following step consists in the determination of the boundary condition for the initial scale PDFs. In this case, we can use the NNPDF2.3QED and MRST2004QED sets for the boundary conditions of quarks, gluons and photon. However, lepton PDFs cannot be extracted from data by means of a fit. The main reason for that is the fact that lepton PDFs are expected to be very small as compared to the quark and gluon PDFs, and even much smaller than the photon PDF. In particular, assuming a small intrinsic leptonic component in the proton, the lepton PDFs are expected to be of the

ID	PDF Set	Ref.	QCD	QED	Photon PDF	Lepton PDFs
A1	`apfel_nn23nlo0118_lept0`	[11]	NLO	LO	$\gamma(x,Q_0)=0$	Eq. (5.12)
A2	`apfel_nn23nnlo0118_lept0`	[11]	NNLO	LO	$\gamma(x,Q_0)=0$	Eq. (5.12)
B1	`apfel_nn23qedlo0118_lept0`	[224]	LO	LO	Internal	Eq. (5.12)
B2	`apfel_nn23qednlo0118_lept0`	[221]	NLO	LO	Internal	Eq. (5.12)
B3	`apfel_nn23qednnlo0118_lept0`	[221]	NNLO	LO	Internal	Eq. (5.12)
B4	`apfel_mrst04qed_lept0`	[109]	NLO	LO	Internal	Eq. (5.12)
C1	`apfel_nn23qedlo0118_lept`	[224]	LO	LO	Internal	Eq. (5.11)
C2	`apfel_nn23qednlo0118_lept`	[221]	NLO	LO	Internal	Eq. (5.11)
C3	`apfel_nn23qednnlo0118_lept`	[221]	NNLO	LO	Internal	Eq. (5.11)
C4	`apfel_mrst04qed_lept`	[109]	NLO	LO	Internal	Eq. (5.11)

Table 5.1: Summary of the sets of PDFs generated with `APFEL` with photons and leptons PDFs.

order of α times the photon PDF, where $\alpha \sim 10^{-2}$ is the fine structure constant. As a consequence, being the photon already very small as compared to quark and gluon PDFs, the contribution of leptons is expected to be extremely small and this clearly makes a reliable determination of the lepton PDFs from experimental data extremely hard.

As an alternative to the fit, we can try to guess the functional form of the PDFs of leptons just by assuming that light leptons, *i.e.* electrons and muons, are generated by photon splitting. At leading-logarithmic accuracy we can then guess their distributions at the initial scale Q_0 as:

$$\ell_\beta(x,Q_0) = \overline{\ell}_\beta(x,Q_0) = \frac{\alpha(Q_0)}{4\pi}\ln\left(\frac{Q_0^2}{m_\beta^2}\right)\int_x^1 \frac{dy}{y} P_{\ell\gamma}^{(0)}\left(\frac{x}{y}\right)\gamma(y,Q_0)\,, \tag{5.11}$$

with $\beta = e^\pm, \mu^\pm$. For the light lepton masses, we take $m_{e^\pm} = 0.510998928$ MeV and $m_{\mu^\pm} = 105.6583715$ MeV, as quoted in the PDG [129].

As far as the $\tau^\pm$ PDFs are concerned, since $m_{\tau^\pm} = 1.777$ GeV $\gtrsim Q_0$, we assume that they are dynamically generated at the threshold according to the usual scheme matching of the VFN scheme.

5.3.3 Preliminary results

In this section we discuss the results of the implementation of the lepton PDFs evolution in `APFEL`. The main goal of this work is to provide an estimate of the lepton PDFs. As discussed in the previous section, the determination of lepton PDFs from a direct fit to data is hard to achieve and thus the alternative is that of modeling initial scale lepton PDFs based on some theoretical assumption.

The model presented in the previous section is based on the assumption that lepton pairs are generated from photon splitting at the respective mass scale. At leading logarithmic accuracy, this results in the ansatz in eq. (5.11) for the light lepton PDFs. However, in order to test how sensitive the results are to the initial scale distributions, we also consider the zero-lepton ansatz where the lepton PDFs at the initial scale Q_0 are equal to zero, that is

$$\ell_\beta(x,Q_0) = \overline{\ell}_\beta(x,Q_0) = 0\,. \tag{5.12}$$

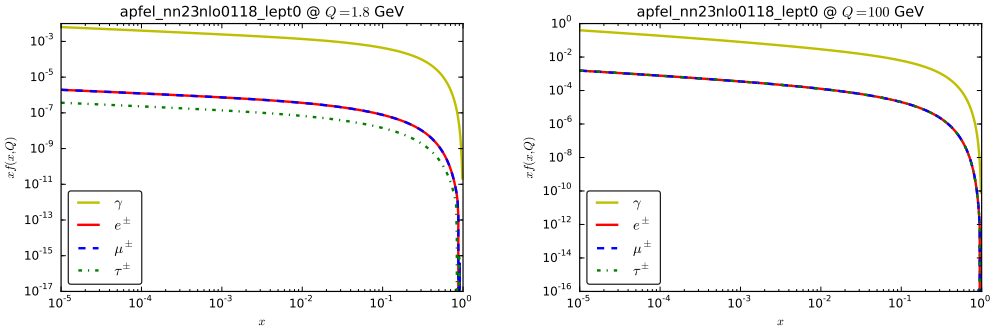

Figure 5.10: Leptons and photon PDF generated dynamically at NLO in QCD.

In this context the construction of PDF sets with leptons requires a pre-existing PDF set to which we add our model for the lepton distributions. Of course, in order to apply the ansatz in eq. (5.11) we need PDF sets that already contain a photon PDF. Presently, there are only two sets that contain a photon PDF: the MRST2004QED set [109] and the NNPDF2.3QED family [221], and we will use both of them to generate lepton PDFs. On the contrary, the ansatz in eq. (5.12) can be applied to any set so that lepton and photon distributions can be generated from any PDF set just by evolution.

In order to assess the effect of considering lepton PDFs in the DGLAP evolution, in this work we consider three different initial scale configurations that are also summarized in Table 5.1:

- Sets where both photon and lepton PDFs are set to zero at the initial scale Q_0 and dynamically generated by DGLAP evolution. For this configuration we have constructed the sets A1 and A2 in Table 5.1 based on NNPDF2.3 NLO and NNLO respectively.

- Sets where the photon distribution is present in the starting set but lepton PDFs are set to zero at the initial scale Q_0 (*i.e.* eq. (5.12)) and then evolved as discussed in Sect. 5.3.1. These configurations are based on the NNPDF2.3QED and MRST2004QED sets of PDFs and identified by the indices B1, B2, B3 and B4 in Table 5.1.

- Sets of PDFs extracted from NNPDF2.3QED and MRST2004QED but using the ansatz in eq. (5.11) for the lepton PDFs (sets C1, C2, C3 and C4 in Table 5.1).

The evolution of the PDF sets listed above is performed using `APFEL` as discussed in Sect. 5.3.1 and tabulated in the `LHAPDF6` format which allows for the inclusion of lepton PDFs in a straightforward manner. In the following we will quantify the differences of the different configurations by looking at PDFs, momentum fractions and luminosities.

In Fig. 5.10 we show the lepton and photon PDF central values for the A1 configuration. In this configuration photons and leptons are set to zero at $Q_0 = 1$ GeV and then dynamically generated by DGLAP evolution. The left plot shows PDFs at $Q = 1.8$ GeV, in this case electron and muon PDFs are identical (by definition), and

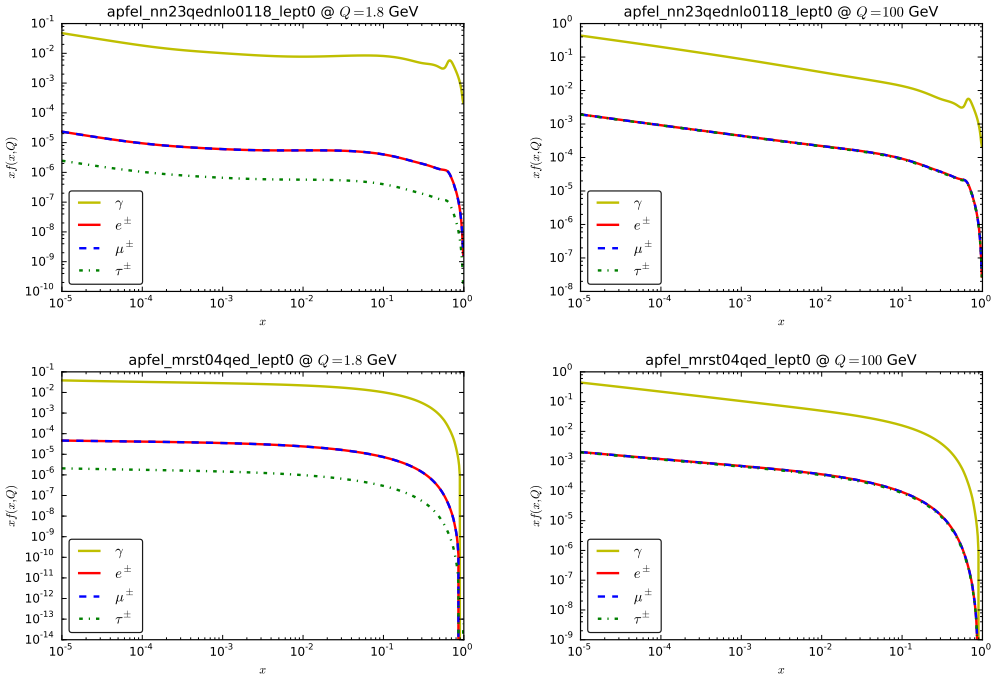

Figure 5.11: Lepton PDFs generated dynamically from the NNPDF2.3QED NLO (top) and MRST2004QED (bottom) photon PDFs.

the τ PDF has just been dynamically generated ($m_\tau = 1.777$ GeV). On the right plot, we display the same comparison but at $Q = 100$ GeV, showing that all lepton PDFs are close to each other. Similar results are obtained also with the NNPDF2.3 NNLO (A2).

Configurations B2 and B4 are shown in Fig. 5.11. For these configurations, the prior set of PDFs contains the photon PDF while the lepton PDFs are null at the initial scale Q_0 and generated dynamically by DGLAP evolution. A similar behaviors as for the configuration A is observed with the additional remark that lepton PDFs present an evident dependence on the shape of the photon PDF. Again, similar results are obtained for the NNPDF2.3QED NNLO (A2) set.

Now, let us consider the configuration of type C where, starting from a prior containing a photon PDF, the initial distributions for the leptons is determined using the ansatz in Eq. (5.11). In Fig. 5.12 we show the resulting lepton PDFs for the configurations C2 (top) and C4 (bottom), at $Q = 1.8$ GeV (left) and $Q = 100$ GeV (right). Again, the qualitative behavior is the same as for the configurations A and B.

In order to quantify the difference generated by the various initial conditions on the evolved lepton PDFs, in Fig. 5.13 we show the ratio plots to the configuration C for the light lepton PDFs produced starting from the NNPDF2.3 sets at NLO at $Q = 100$ GeV. For the electron PDFs (left plot), the ansatz in Eqs. (5.11) and (5.12) applied to a set with a photon PDF lead to similar results in the small-x region while

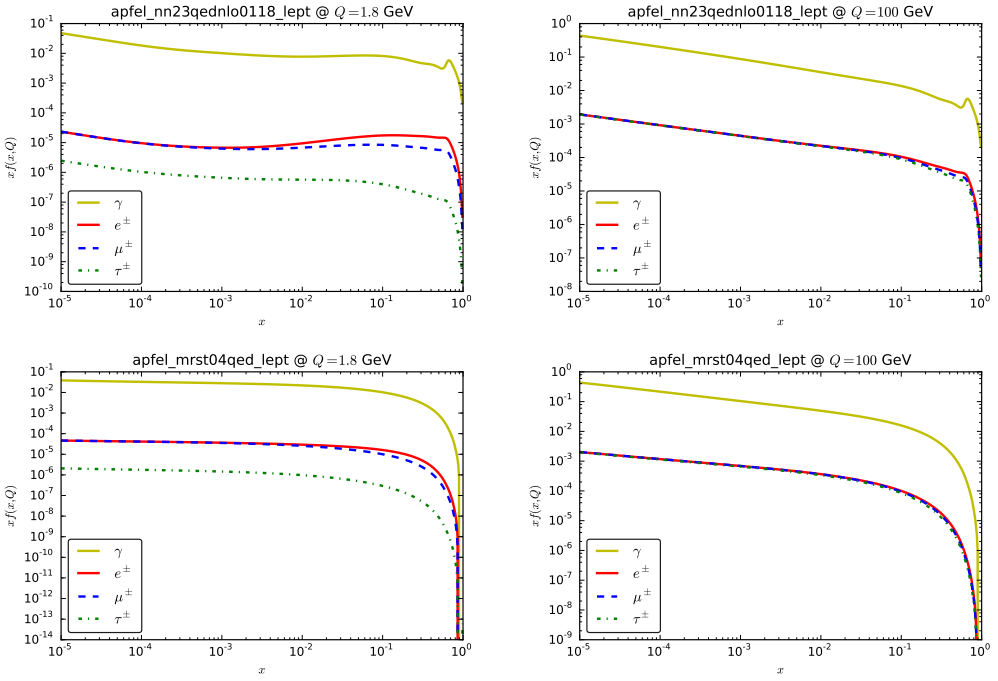

Figure 5.12: Lepton PDFs based on the ansatz of Eq. (5.11) and evolved with the NNPDF2.3QED NLO (top) and MRST2004QED (bottom) photon PDFs.

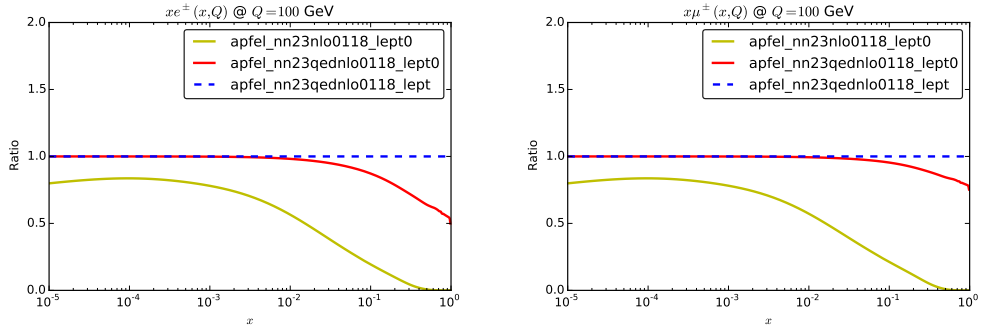

Figure 5.13: Ratio of electron and muon PDFs for each configuration.

difference up to 50% are observed in the larger-x region. The electron PDFs resulting from a set without a photon PDF are instead way below all over the x range. The same behavior is observed also for the muon PDFs (right plot in Fig. 5.13), with slightly less enhanced discrepancies as compared to the electrons.

Interesting information about the photon and lepton content of the proton is pro-

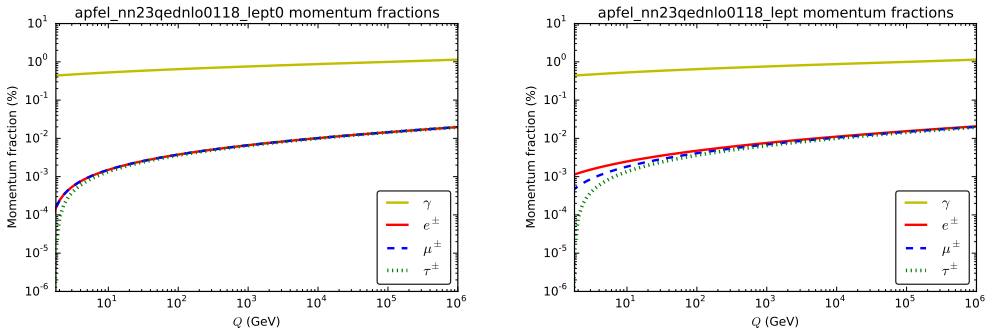

Figure 5.14: Momentum fractions for the photon and lepton PDFs.

vided by the respective momentum fraction defined as:

$$\mathrm{MF}_\gamma(Q) = \int_0^1 dx\, x\gamma(x,Q)\,, \quad \mathrm{MF}_{\ell^\pm}(Q) = \int_0^1 dx\, x\ell^\pm(x,Q)\,. \tag{5.13}$$

In Fig. 5.14 we plot the percent momentum fractions as a function of the energy Q for the configurations B2 (left) and C2 (right). While the photon PDF carries up to around 1% of the proton moment fraction of the proton, the lepton PDFs, independently from the parametrization conditions, carry a much smaller fraction around tow order of magnitude smaller than that carried by the photon. This is consistent with the fact that, for both parametrizations, lepton PDFs are proportional to α times the photon PDF ($\ell \propto \alpha \times \gamma$). In conclusion, lepton PDFs carry such a small fraction of the proton momentum that they do not cause a significant violation of the total momentum sum rule.

In the computation of hadron collider processes, PDFs factorize in the form of *parton luminosities* as defined in Eq. (1.39). Defining

$$\Phi_{\gamma\ell}\left(M_X\right) = \sum_{j=e^\pm,\mu^\pm,\tau^\pm} \Phi_{\gamma j}\left(M_X\right)\,, \tag{5.14}$$

in Fig. 5.15 we plot the $\Phi_{\gamma\gamma}$, $\Phi_{\gamma\ell}$, $\Phi_{e^+e^-}$, $\Phi_{\mu^+\mu^-}$ and $\Phi_{\tau^+\tau^-}$ parton luminosities as functions of M_X at $\sqrt{s} = 13$ TeV for the sets B2 and C2. The relative size of the plotted luminosities follows the expected pattern according to which the $\Phi_{\gamma\ell}$ luminosity is roughly suppressed by one power of α as compared to $\Phi_{\gamma\gamma}$ while $\Phi_{\ell^+\ell^-}$, with $\ell = e, \mu, \tau$, is suppressed by two powers.

We now turn to consider the uncertainties of the lepton PDFs. A realistic estimated of the lepton PDFs requires the estimation of uncertainty associated to each of them. To this end, using exactly the same procedure discussed in the previous sections, we generated lepton PDFs for all replicas of NNPDF2.3QED family sets. This eventually allowed us to estimate the uncertainty on each lepton PDF. In Fig. 5.16 we plot the lepton PDFs with the respective uncertainty for the configurations C2 and C3 at $Q = 100$ GeV. Uncertainties are calculated as the one-sigma interval (standard deviation) from the central value of each PDF flavor. As expected, the lepton PDF uncertainties follow the pattern of photon PDF uncertainty.

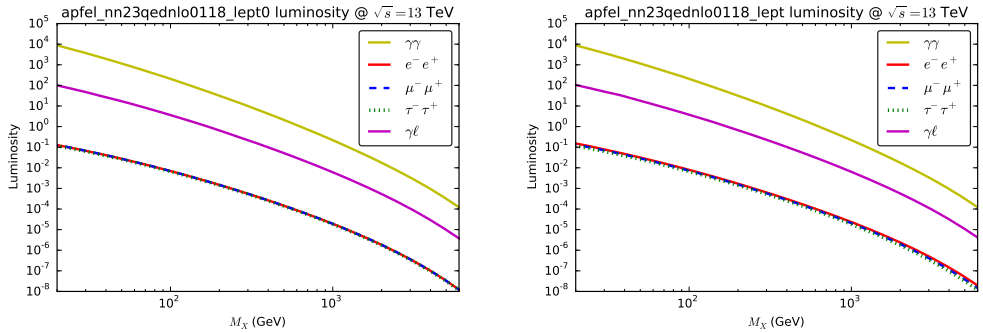

Figure 5.15: Parton luminosities for photon and lepton PDFs.

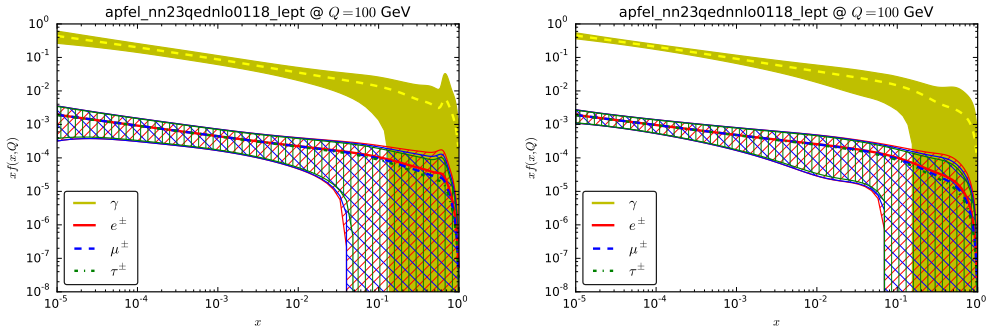

Figure 5.16: Uncertainties for the lepton PDFs at NLO (left) and NNLO (right) in QCD using the NNPDF2.3QED set (C2 and C3). Leptons are generated from Eq. (5.11).

In Fig. 5.17 we compute the correlation of PDFs for set C3 at $Q = 1.8$ GeV in a grid of $N_x = 50$ points in $x_1, x_2 = [10^{-5}, 1]$, for the flavors $(\tau, \mu, e, \gamma, g, d, u, s)$, defined as

$$\rho_{\alpha\beta}(x_1, x_2, Q) = \frac{N_{\rm rep}}{N_{\rm rep}-1}\left(\frac{\left\langle f_\alpha^{(k)}(x_1,Q) f_\beta^{(k)}(x_2,Q)\right\rangle_{\rm rep} - \left\langle f_\alpha^{(k)}(x_1,Q)\right\rangle_{\rm rep}\left\langle f_\beta^{(k)}(x_2,Q)\right\rangle_{\rm rep}}{\sigma_\alpha(x_1,Q)\cdot\sigma_\beta(x_2,Q)}\right), \tag{5.15}$$

where averages are taken over the $k = 1, \ldots, N_{\rm rep}$ replicas and where $\sigma_i(x, Q)$ are the corresponding standard deviations. Each row of this matrix is expressed in terms of $f_i \cdot N_x + x_j$, for $j = 1, \ldots, N_x$ and $i = \tau, \mu, e, \gamma, g, d, u, s$.

In the first place, we note a clear distinction between the QED (upper left square region) and the QCD sector (bottom right region). As expected, there are strong correlations between (τ, μ, e, γ) due to the fact that leptons are generated by photon splitting. A similar behavior is also observed for (g, d, u, s). The off-diagonal elements show that quark and gluon distributions are instead mildly correlated to lepton and photon PDFs. Similar results are obtained for the other configurations.

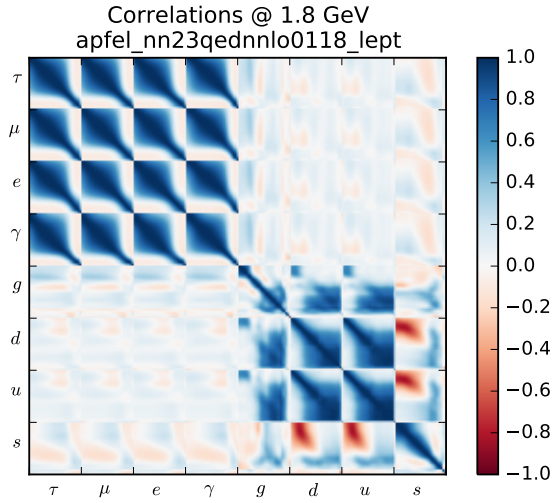

Figure 5.17: PDF correlation at $Q = 1.8$ GeV.

Finally, we remark that we are currently implementing the lepton-induced channels in the `MadGraph5_aMC@NLO` framework for Drell-Yan production. Preliminary results show that the contribution of lepton-induced channels is always similar or below to the photon-induced contribution for lepton pair, dijet and vector boson pair production at LHC ($\sqrt{s}$ =13 TeV) and FCC-hh ($\sqrt{s}$ =100 TeV) when applying reasonable kinematic cuts in lepton p_T and rapidity. The preliminary conclusion is that lepton-induced processes are rare for the SM predictions tested in this work, and so, fitting leptons in future sets of PDFs will not lead to significant results.

We would like to stress that this is the first study which provides a guess for the lepton PDFs. The sets of PDFs generated with `APFEL` are publicly available, so further analysis are encouraged, in particular on different configurations such as BSM. Moreover, the current evolution framework opens the possibility to eventually include the lepton PDFs contributions in PDF fits with QED correction.

Chapter 6

Conclusion and outlook

In this thesis we have presented a first determination of an unbiased set of PDFs with QED corrections using the NNPDF methodology: the NNPDF2.3QED set. In this set the photon PDF and its uncertainties are determined by deep-inelastic scattering and neutral- and charge-current Drell-Yan production data from the LHC. We have discussed about the phenomenological impact of the photon PDF, highlighting the lack of experimental information for large-x region, which induces large uncertainties related to electroweak corrections in processes which are relevant for new physics searches at the LHC, such as high mass gauge boson production and double gauge boson production.

This work has presented a series of important deliverables which have been developed particularly for the determination of this set of PDFs with QED corrections. Let us summarize these results:

- We have implemented `APFEL`, a new PDF evolution library that combines NNLO QCD corrections with LO QED effects in the solution of the DGLAP equations. This is the first public evolution code that performs the combined QCD$\otimes$QED evolution up to NNLO in QCD and LO in QED, both in the FFN and VFN schemes, and using either pole or $\overline{\rm MS}$ heavy quark masses. We provide two strategies for solving this combined system of evolution equations: the *coupled* and the *unified* solutions. We have presented a detailed benchmarking exercise between `APFEL` and other public available codes such as: `HOPPET`, `partonevolution`, `MRST2004QED` and `QCDNUM`.

- We have released `APFEL Web` a new Web-based application, born as a spin-off of the `APFEL` library. `APFEL Web` provides a user-friendly graphical user interface for the visualization of PDFs with a wide range of formats: absolute plots, ratio plots, compare PDFs from different groups, compare error PDF from a single set, plot all PDF flavor combinations at the same time, compute parton luminosities and finally compute DIS structure functions and `APPLgrid` observables. Moreover it provides a simple interface for the customization of PDF evolution.

- We have developed a new modern framework for the NNPDF methodology. This new framework provides a flexible and fast code structure to perform PDF fits, in multiple configurations. This code has been used for the determination of

this first set of PDFs with QED corrections, and it is the production code since NNPDF3.0.

- We have delivered the NNPDF2.3QED set of PDFs at NLO and NNLO in QCD and LO in QED, for $\alpha_s = 0.117, 0.118, 0.119$ values. In this set, the photon PDF is parametrized by an artificial neural-network and trained with the NNPDF methodology. The photon PDF is extracted for the first time from DIS data, and then by reweighting with neutral- and charge-current Drell-Yan production data from the LHC. The final results provides a first determination of the photon PDF and its uncertainties.

- Several phenomenological implications of this new set of PDFs have been studied for: the direct photon production at HERA, searches for new massive electroweak gauge bosons and W pair production at small p_T and large invariant mass, at LHC energies. We have also interfaced this set of PDFs for multiple Monte Carlo event generators: `PYTHIA`, `MadGraph5_aMC@NLO` and `SHERPA`.

- We present preliminary results about lepton PDFs. In fact, the inclusion of QED corrections requires extending the DGLAP evolution equations to include, in the first place, the photon PDF and, for consistency, also PDFs for the charged leptons $e^{\pm}$, $\mu^{\pm}$ and $\tau^{\pm}$. Here, we have shown how to construct those sets with `APFEL` considering multiple initial conditions for photon and lepton PDFs. We discuss about the size of these PDFs, its momentum fraction and luminosities. This is the first guess for the lepton PDFs, which is useful when considering electroweak corrections to some hadron-collider processes.

We plan to release in a near future new sets of PDFs with QED corrections after introducing some technical improvements in the procedure. First of all, we propose to extend the `APFEL` combined QCD$\otimes$QED evolution up to NLO in QED together with the inclusion of the subleading terms $\mathcal{O}(\alpha\alpha_s)$. Secondly, we need fast interfaces, such as `APPLgrid`, with electroweak corrections including the photon-initiated channels. Such interface is important because avoids the reweighting procedure that we have used for this first determination, moreover it opens the possibility to compute several photon-induced processes easily. Finally, the last important ingredient for an improved set of PDFs with QED corrections is the inclusion of new LHC data in regions where the photon PDF uncertainties are unconstrained. In particular, based on the results presented in this work and in the studies performed in Ref. [218], the most relevant data are: high- and low-mass Drell-Yan Z/γ^* production, diboson pair production at small-p_T and large invariant mass, dilepton rapidity distributions, small-p_T distribution for leptons among others.

In conclusion, future sets of PDFs with QED corrections will be releases after introducing the technological improvements listed in the last paragraph. Such sets of PDFs will enhance the quality and reliability of predictions.

Appendix A

Distance estimators

The distance estimator assesses the compatibility between two PDFs sets, and it tests whether two PDF sets are statistically equivalent.

Given a Monte Carlo sample of $N_{\rm rep}$ replicas representing the probability distribution of a given PDF set, $\left\{f^{(k)}\right\}$, the expectation value of the distribution as a function of x and Q^2 is given by

$$\bar{f}(x,Q^2) \equiv \left\langle f(x,Q^2)\right\rangle_{\rm rep} = \frac{1}{N_{\rm rep}} \sum_i^{N_{\rm rep}} f^{(k)}(x,Q^2)\,, \tag{A.1}$$

where the index (k) runs over all the replicas in the sample. The variance of the sample is estimated as

$$\sigma^2\left[f(x,Q^2)\right] = \frac{1}{N_{\rm rep}-1} \sum_i^{N_{\rm rep}} \left(f^{(k)}(x,Q^2) - \left\langle f(x,Q^2)\right\rangle_{\rm rep}\right)^2 . \tag{A.2}$$

The variance of the mean is, in turn, defined in terms of the variance of the sample by

$$\sigma^2\left[\left\langle f(x,Q^2)\right\rangle_{\rm rep}\right] = \frac{1}{N_{\rm rep}}\sigma^2\left[f(x,Q^2)\right] . \tag{A.3}$$

The variance of the variance itself can be computed using

$$\sigma^2\left[\sigma^2\left[f(x,Q^2)\right]\right] = \frac{1}{N_{\rm rep}}\left[m_4\left[f(x,Q^2)\right] - \frac{N_{\rm rep}-3}{N_{\rm rep}-1}\left(\sigma^2\left[f(x,Q^2)\right]\right)^2\right], \tag{A.4}$$

where $m_4\left[f(x,Q^2)\right]$ denotes the fourth moment of the probability distribution for $f(x,Q^2)$, namely

$$m_4\left[f(x,Q^2)\right] = \frac{1}{N_{\rm rep}}\left[\sum_{k=1}^{N_{\rm rep}}\left(f^{(k)}(x,Q^2) - \left\langle f(x,Q^2)\right\rangle_{\rm rep}\right)^4\right] . \tag{A.5}$$

The distance between two sets of PDFs, each characterized by a given distribution of the Monte Carlo replicas, denoted by $\left\{f^{(k)}\right\}$ and $\left\{g^{(k)}\right\}$, is defined as the square

root of the square difference of the PDF central values in units of the uncertainty of the mean:

$$d_{\bar{f},\bar{g}}(x,Q^2) = \sqrt{\frac{\left(\bar{f}-\bar{g}\right)^2}{\sigma^2\left[\bar{f}\right]+\sigma^2\left[\bar{g}\right]}}\,. \tag{A.6}$$

In Eq. (A.6), the denominator uses the variance of the mean of the distribution, defined as in Eq. (A.3). An analogous distance can be defined for the variances of the two samples:

$$d_{\sigma[f],\sigma[g]}(x,Q^2) = \sqrt{\frac{\left(\sigma^2\left[f\right]-\sigma^2\left[g\right]\right)^2}{\sigma^2\left[\sigma^2\left[f\right]\right]+\sigma^2\left[\sigma^2\left[g\right]\right]}}\,. \tag{A.7}$$

where now in the denominator we have the variance of the variance, Eq. (A.4).

The distances for the central values and for the variances defined in Eqs. (A.6) and (A.7) test whether the underlying distributions from which the two Monte Carlo samples $\left\{f^{(k)}\right\}$ and $\left\{g^{(k)}\right\}$ are drawn have respectively the same mean and the same standard deviation. In particular, it is possible to show that one expects these distances to fluctuate around $d \sim 1$ if the two samples do indeed come from the same distribution. Values of the distances around $d \sim \sqrt{N_{\rm rep}}$ indicates that the central values (the variances) of the two PDF sets differ by one standard deviation in units of the variance of the distribution Eq. (A.2) (in units of the variance of the variance Eq. (A.4)).

Appendix B

Bayesian reweighting

Let us consider that a set of experimental data is used to construct a probability distribution for PDFs, $\mathcal{P}_{\rm old}(f)$. With this probability distribution any observable can be obtained by performing averages over this ensemble, equally weighting each PDF.

Suppose that we would like to include new experimental data without applying the fitting procedure presented in Chap. 3. The only option that we have is to extract a new probability distribution $\mathcal{P}_{\rm new}$ by updating the weights, w_k associated to each individual PDF f_k of the prior ensemble.

From a practical point of view, the new data is assumed to have Gaussian errors, so we can write the relative probabilities of the new data for different choices of PDF in terms of the probability density of the χ to the new data

$$\mathcal{P}(\chi|f) \propto (\chi^2(y,f))^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2(y,f)}\,, \tag{B.1}$$

where $y = \{y_1, y_2, \ldots, y_n\}$ are the new n experimental data points and

$$\chi^2(y,f) = \sum_{i,j=1}^{n} \frac{(y_i - y_i[f])(y_j - y_j[f])}{\sigma_{ij}}\,, \tag{B.2}$$

where $y_i[f]$ is the value predicted for the data y_i using the PDF f, and σ_{ij} is the data uncertainties covariance matrix.

From statistical independece of the old and new data we have

$$\mathcal{P}_{\rm new}(f) = \mathcal{N}_\chi \mathcal{P}(\chi|f)\mathcal{P}_{\rm old}(f)\,, \tag{B.3}$$

where $\mathcal{N}_\chi$ is a normalization factor, independent of f. We can show that some observable $\mathcal{O}[f]$ is given in terms of N reweighted replicas f_k.

$$\langle \mathcal{O}_{\rm new} \rangle = \frac{1}{N} \sum_{k=1}^{N} w_k \mathcal{O}[f_k]\,, \tag{B.4}$$

where

$$w_k = \mathcal{N}_\chi \mathcal{P}(\chi|f_k) = \frac{(\chi_k^2)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N}\sum_{k=1}^{N} (\chi^2)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi_k^2}}\,, \tag{B.5}$$

with $\chi^2 \equiv \chi^2(y, f_k)$.

The definition of weights in Eq. (B.5) is used when new experimental data is included in the fit. We can also quantify the loss of efficiency by using the Shannon entropy to compute the effective number of replicas left after reweighting

$$N_{\text{eff}} = \exp\left(\frac{1}{N}\sum_{k=1}^{N} w_k \ln\frac{N}{w_k}\right) . \tag{B.6}$$

Finally, after reweighting a set of PDFs it is always possible to construct an unweighted set where each PDF has equally distributed weights. More details of the unweighting procedure can be found in Ref. [189].

Bibliography

[1] **ALEPH** Collaboration, D. Decamp et al., *Determination of the Number of Light Neutrino Species*, *Phys.Lett.* **B231** (1989) 519.

[2] **CDF** Collaboration, F. Abe et al., *Observation of top quark production in $\bar{p}p$ collisions*, *Phys.Rev.Lett.* **74** (1995) 2626–2631, [hep-ex/9503002].

[3] **ATLAS** Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys.Lett.* **B716** (2012) 1–29, [arXiv:1207.7214].

[4] **CMS** Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys.Lett.* **B716** (2012) 30–61, [arXiv:1207.7235].

[5] C. Carloni Calame, G. Montagna, O. Nicrosini, and 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].

[6] 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].

[7] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in FEWZ*, *Phys.Rev.* **D86** (2012) 094034, [arXiv:1208.5967].

[8] R. D. Ball, S. Carrazza, L. Del Debbio, S. Forte, J. Gao, et al., *Parton Distribution Benchmarking with LHC Data*, *JHEP* **1304** (2013) 125, [arXiv:1211.5142].

[9] V. Bertone, S. Carrazza, and J. Rojo, *APFEL: A PDF Evolution Library with QED corrections*, *Comput.Phys.Commun.* **185** (2014) 1647–1668, [arXiv:1310.1394].

[10] S. Carrazza, A. Ferrara, D. Palazzo, and J. Rojo, *APFEL Web: a web-based application for the graphical visualization of parton distribution functions*, arXiv:1410.5456.

[11] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, et al., *Parton distributions with LHC data*, *Nucl.Phys.* **B867** (2013) 244–289, [arXiv:1207.1303].

[12] **The NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, arXiv:1410.8849.

[13] **The NNPDF** Collaboration, R. D. Ball et al., *Unbiased determination of polarized parton distributions and their uncertainties*, *Nucl.Phys.* **B874** (2013) 36–84, [arXiv:1303.7236].

[14] S. Carrazza, *Towards an unbiased determination of parton distributions with QED corrections*, arXiv:1305.4179.

[15] S. Carrazza, *Towards the determination of the photon parton distribution function constrained by LHC data*, arXiv:1307.1131.

[16] S. Carrazza, *Disentangling electroweak effects in Z-boson production*, arXiv:1405.1728.

[17] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics*. Cambridge University Press, 1996.

[18] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*. Advanced book program. Westview Press Reading (Mass.), Boulder (Colo.), 1995.

[19] **H1** Collaboration, F. D. Aaron et al., *Measurement of the Proton Structure Function F_L at Low x*, *Phys. Lett.* **B665** (2008) 139–146, [arXiv:0805.2809].

[20] **ZEUS** Collaboration, A. Cooper Sarkar, *Measurement of high-Q2 neutral current deep inelastic e+p scattering cross sections with a longitudinally polarised positron beam at HERA*, arXiv:1208.6138.

[21] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock, and A. Bodek, *Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections*, *Phys. Lett.* **B282** (1992) 475–482.

[22] **BCDMS** Collaboration, A. Benvenuti et al., *A High Statistics Measurement of the Deuteron Structure Functions F2 (X, Q^2) and R From Deep Inelastic Muon Scattering at High Q^2*, *Phys.Lett.* **B237** (1990) 592.

[23] R. Feynman, *The behavior of hadrond collisions at extreme energies*, .

[24] **ATLAS** Collaboration, 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*, arXiv:1305.4192.

[25] **ATLAS** Collaboration, G. Aad et al., *Measurement of the inclusive $W^{\pm}$ and Z/γ cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys.Rev.* **D85** (2012) 072004, [arXiv:1109.5141].

[26] *Inclusive low mass Drell-Yan production in the forward region at $\sqrt{s} = 7$ TeV*, . LHCb-CONF-2012-013.

[27] A. Vogt, S. Moch, and J. Vermaseren, *The Three-loop splitting functions in QCD: The Singlet case*, *Nucl.Phys.* **B691** (2004) 129–181, [hep-ph/0404111].

[28] S. Moch, J. Vermaseren, and A. Vogt, *The Three loop splitting functions in QCD: The Nonsinglet case*, *Nucl.Phys.* **B688** (2004) 101–134, [hep-ph/0403192].

[29] H. Georgi and H. D. Politzer, *Electroproduction scaling in an asymptotically free theory of strong interactions*, *Phys.Rev.* **D9** (1974) 416–420.

[30] D. Gross and F. Wilczek, *Asymptotically free gauge theories*, *Phys.Rev.* **D9** (1974) 980–993.

[31] S. Alekhin, J. Blumlein, and S. Moch, *Parton Distribution Functions and Benchmark Cross Sections at NNLO*, *Phys.Rev.* **D86** (2012) 054009, [arXiv:1202.2281].

[32] S. Alekhin and S. Moch, *Heavy-quark deep-inelastic scattering with a running mass*, *Phys. Lett.* **B699** (2011) 345–353, [arXiv:1011.5790].

[33] P. Nadolsky, J. Gao, M. Guzzi, J. Huston, H.-L. Lai, et al., *Progress in CTEQ-TEA PDF analysis*, arXiv:1206.3321.

[34] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev.* **D82** (2010) 074024, [arXiv:1007.2241].

[35] M. Guzzi, P. M. Nadolsky, H.-L. Lai, and C.-P. Yuan, *General-Mass Treatment for Deep Inelastic Scattering at Two-Loop Accuracy*, *Phys.Rev.* **D86** (2012) 053005, [arXiv:1108.5112].

[36] **H1 and ZEUS** Collaboration, V. Radescu, *Combination and QCD analysis of the HERA inclusive cross sections*, *PoS* **ICHEP2010** (2010) 168.

[37] **ZEUS , H1** Collaboration, A. Cooper-Sarkar, *PDF Fits at HERA*, *PoS* **EPS-HEP2011** (2011) 320, [arXiv:1112.2107].

[38] **H1** Collaboration, F. Aaron et al., *Inclusive Deep Inelastic Scattering at High Q^2 with Longitudinally Polarised Lepton Beams at HERA*, *JHEP* **1209** (2012) 061, [arXiv:1206.7007].

[39] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C63** (2009) 189–285, [arXiv:0901.0002].

[40] L. Harland-Lang, A. Martin, P. Motylinski, and R. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, arXiv:1412.3989.

[41] S. Forte, E. Laenen, P. Nason, and J. Rojo, *Heavy quarks in deep-inelastic scattering*, *Nucl. Phys.* **B834** (2010) 116–162, [arXiv:1001.2312].

[42] P. Jimenez-Delgado and E. Reya, *Variable Flavor Number Parton Distributions and Weak Gauge and Higgs Boson Production at Hadron Colliders at NNLO of QCD*, *Phys. Rev.* **D80** (2009) 114011, [arXiv:0909.1711].

[43] R. D. Ball et al., *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, *Nucl. Phys.* **B849** (2011) 296–363, [arXiv:1101.1300].

[44] **NNPDF** Collaboration, R. D. Ball et al., *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, *Nucl.Phys.* **B855** (2012) 153–221, [arXiv:1107.2652].

[45] **ATLAS** Collaboration, G. Aad et al., *Determination of the strange quark density of the proton from ATLAS measurements of the $W \to \ell\nu$ and $Z \to \ell\ell$ cross sections*, *Phys.Rev.Lett.* **109** (2012) 012001, [arXiv:1203.4051].

[46] J. M. Campbell, J. W. Huston, and W. J. Stirling, *Hard interactions of quarks and gluons: A primer for LHC physics*, *Rept. Prog. Phys.* **70** (2007) 89, [hep-ph/0611148].

[47] R. Thorne, *The Effect of Changes of Variable Flavour Number Scheme on PDFs and Predicted Cross Sections*, *Phys. Rev.* **D86** (2012) 074017, [arXiv:1201.6180].

[48] A. M. Cooper-Sarkar, *Including heavy flavour production in PDF fits*, arXiv:0709.0191.

[49] R. Thorne, *A Variable-flavor number scheme for NNLO*, *Phys.Rev.* **D73** (2006) 054019, [hep-ph/0601245].

[50] R. Thorne and G. Watt, *PDF dependence of Higgs cross sections at the Tevatron and LHC: Response to recent criticism*, *JHEP* **1108** (2011) 100, [arXiv:1106.5789].

[51] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *Uncertainties of predictions from parton distributions. II: Theoretical errors*, *Eur. Phys. J.* **C35** (2004) 325–348, [hep-ph/0308087].

[52] **The NNPDF** Collaboration, R. D. Ball et al., *Theoretical issues in PDF determination and associated uncertainties*, *Phys.Lett.* **B723** (2013) 330–339, [arXiv:1303.1189].

[53] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, *Phys. Rev.* **D69** (2004) 094008, [hep-ph/0312266].

[54] **CMS** Collaboration, *Inclusive W/Z cross section at 8 TeV*, CMS-PAS-SMP-12-011.

[55] C. Anastasiou, S. Buehler, F. Herzog, and A. Lazopoulos, *Total cross-section for Higgs boson hadroproduction with anomalous Standard Model interactions*, *JHEP* **1112** (2011) 058, [arXiv:1107.0683].

[56] **LHC Higgs Cross Section Working Group** Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593.

[57] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, *Higgs production via vector-boson fusion at NNLO in QCD*, *Phys.Rev.Lett.* **105** (2010) 011801, [arXiv:1003.4451].

[58] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes a hadron collider*, *Phys.Lett.* **B579** (2004) 149–156, [hep-ph/0307206].

[59] O. Brein, R. V. Harlander, and T. J. Zirke, *vh@nnlo - Higgs Strahlung at hadron colliders*, arXiv:1210.5347.

[60] J. Campbell and R. K. Ellis, *Next-to-leading order corrections to W + 2jet and Z + 2jet production at hadron colliders*, *Phys. Rev.* **D65** (2002) 113007, [hep-ph/0202176].

[61] M. Czakon and A. Mitov, *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, arXiv:1112.5675.

[62] P. Baernreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, *Phys.Rev.Lett.* **109** (2012) 132001, [arXiv:1204.5201].

[63] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, arXiv:1210.6832.

[64] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, *JHEP* **1212** (2012) 054, [arXiv:1207.0236].

[65] M. Aliev et al., – *HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR*, *Comput. Phys. Commun.* **182** (2011) 1034–1046, [arXiv:1007.1327].

[66] S. Moch, P. Uwer, and A. Vogt, *On top-pair hadro-production at next-to-next-to-leading order*, *Phys.Lett.* **B714** (2012) 48–54, [arXiv:1203.6282].

[67] M. Cacciari, M. Czakon, M. L. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, *Phys.Lett.* **B710** (2012) 612–622, [arXiv:1111.5869].

[68] **CMS** Collaboration, *Cross section measurement in the di-lepton channel at 8 tev*, CMS-PAS-TOP-12-007.

[69] **CMS** Collaboration, *First Determination of the Strong Coupling Constant from the $t\bar{t}$ Cross Section*, CMS-PAS-TOP-12-022.

[70] S. Forte and G. Watt, *Progress in the Determination of the Partonic Structure of the Proton*, arXiv:1301.6754.

[71] A. De Roeck and R. Thorne, *Structure Functions*, *Prog.Part.Nucl.Phys.* **66** (2011) 727–781, [arXiv:1103.0555].

[72] K. Mishra, T. Becher, L. Barze, M. Chiesa, S. Dittmaier, et al., *Electroweak Corrections at High Energies*, arXiv:1308.1430.

[73] U. Baur, S. Keller, and D. Wackeroth, *Electroweak radiative corrections to W boson production in hadronic collisions*, *Phys.Rev.* **D59** (1999) 013002, [hep-ph/9807417].

[74] V. Zykunov, *Electroweak corrections to the observables of W boson production at RHIC*, *Eur.Phys.J.direct* **C3** (2001) 9, [hep-ph/0107059].

[75] S. Dittmaier and . Kramer, Michael, *Electroweak radiative corrections to W boson production at hadron colliders*, *Phys.Rev.* **D65** (2002) 073007, [hep-ph/0109062].

[76] U. Baur, O. Brein, W. Hollik, C. Schappacher, and D. Wackeroth, *Electroweak radiative corrections to neutral current Drell-Yan processes at hadron colliders*, *Phys.Rev.* **D65** (2002) 033007, [hep-ph/0108274].

[77] U. Baur and D. Wackeroth, *Electroweak radiative corrections to $p\bar{p} \to W^{\pm} \to \ell^{\pm}\nu$ beyond the pole approximation*, *Phys.Rev.* **D70** (2004) 073015, [hep-ph/0405191].

[78] A. Arbuzov, D. Bardin, S. Bondarenko, P. Christova, L. Kalinovskaya, et al., *One-loop corrections to the Drell–Yan process in SANC. (II). The Neutral current case*, *Eur.Phys.J.* **C54** (2008) 451–460, [arXiv:0711.0625].

[79] A. Arbuzov, D. Bardin, S. Bondarenko, P. Christova, L. Kalinovskaya, et al., *One-loop corrections to the Drell-Yan process in SANC. I. The Charged current case*, *Eur.Phys.J.* **C46** (2006) 407–412, [hep-ph/0506110].

[80] S. Brensing, S. Dittmaier, . Kramer, Michael, and A. Muck, *Radiative corrections to W^- boson hadroproduction: Higher-order electroweak and supersymmetric effects*, *Phys.Rev.* **D77** (2008) 073006, [arXiv:0710.3309].

[81] G. Balossini, G. Montagna, C. M. Carloni Calame, M. Moretti, O. Nicrosini, et al., *Combination of electroweak and QCD corrections to single W production at the Fermilab Tevatron and the CERN LHC*, *JHEP* **1001** (2010) 013, [arXiv:0907.0276].

[82] S. Dittmaier and 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].

[83] A. Denner, S. Dittmaier, T. Kasprzik, and A. Muck, *Electroweak corrections to W + jet hadroproduction including leptonic W-boson decays*, *JHEP* **0908** (2009) 075, [arXiv:0906.1656].

[84] A. Denner, S. Dittmaier, T. Kasprzik, and A. Muck, *Electroweak corrections to dilepton + jet production at hadron colliders*, *JHEP* **1106** (2011) 069, [arXiv:1103.0914].

[85] A. Denner, S. Dittmaier, T. Kasprzik, and A. Muck, *Electroweak corrections to monojet production at the LHC*, *Eur.Phys.J.* **C73** (2013) 2297, [arXiv:1211.5078].

[86] J. Baglio, L. D. Ninh, and M. M. Weber, *Massive gauge boson pair production at the LHC: a next-to-leading order story*, arXiv:1307.4331.

[87] A. Bierweiler, T. Kasprzik, H. Kuhn, and S. Uccirati, *Electroweak corrections to W-boson pair production at the LHC*, *JHEP* **1211** (2012) 093, [arXiv:1208.3147].

[88] M. Luszczak and A. Szczurek, *Subleading processes in production of W^+W^- pairs in proton-proton collisions*, arXiv:1309.7201.

[89] S. Moretti, M. Nolten, and D. Ross, *Weak corrections and high E(T) jets at Tevatron*, *Phys.Rev.* **D74** (2006) 097301, [hep-ph/0503152].

[90] S. Dittmaier, A. Huss, and C. Speckner, *Weak radiative corrections to dijet production at hadron colliders*, *JHEP* **1211** (2012) 095, [arXiv:1210.0438].

[91] W. Bernreuther, M. Fuecker, and Z. Si, *Mixed QCD and weak corrections to top quark pair production at hadron colliders*, *Phys.Lett.* **B633** (2006) 54–60, [hep-ph/0508091].

[92] J. H. Kuhn, A. Scharf, and P. Uwer, *Electroweak corrections to top-quark pair production in quark-antiquark annihilation*, *Eur.Phys.J.* **C45** (2006) 139–150, [hep-ph/0508092].

[93] W. Hollik and M. Kollar, *NLO QED contributions to top-pair production at hadron collider*, *Phys.Rev.* **D77** (2008) 014008, [arXiv:0708.1697].

[94] W. Hollik and D. Pagani, *The electroweak contribution to the top quark forward-backward asymmetry at the Tevatron*, *Phys.Rev.* **D84** (2011) 093003, [arXiv:1107.2606].

[95] J. Khn, A. Scharf, and P. Uwer, *Weak Interactions in Top-Quark Pair Production at Hadron Colliders: An Update*, arXiv:1305.5773.

[96] A. De Rujula, R. Petronzio, and A. Savoy-Navarro, *Radiative Corrections to High-Energy Neutrino Scattering*, *Nucl.Phys.* **B154** (1979) 394.

[97] J. Kripfganz and H. Perlt, *Electroweak radiative corrections and quark mass singularities*, *Z.Phys.* **C41** (1988) 319–321.

[98] J. Blumlein, *Leading log radiative corrections to deep inelastic neutra and charged current scattering at HERA*, *Z.Phys.* **C47** (1990) 89–94.

[99] G. P. Salam and J. Rojo, *A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)*, *Comput. Phys. Commun.* **180** (2009) 120–156, [arXiv:0804.3755].

[100] A. Cafarella, C. Coriano, and M. Guzzi, *Precision Studies of the NNLO DGLAP Evolution at the LHC with CANDIA*, *Comput.Phys.Commun.* **179** (2008) 665–684, [arXiv:0803.0462].

[101] M. Botje, *QCDNUM: Fast QCD Evolution and Convolution*, *Comput.Phys.Commun.* **182** (2011) 490–532, [arXiv:1005.1481].

[102] P. G. Ratcliffe, *A matrix approach to numerical solution of the DGLAP evolution equations*, *Phys.Rev.* **D63** (2001) 116004, [hep-ph/0012376].

[103] L. Schoeffel, *An Elegant and fast method to solve QCD evolution equations, application to the determination of the gluon content of the pomeron*, *Nucl.Instrum.Meth.* **A423** (1999) 439–445.

[104] C. Pascaud and F. Zomer, *A Fast and precise method to solve the Altarelli-Parisi equations in x space*, hep-ph/0104013.

[105] A. Vogt, *Efficient evolution of unpolarized and polarized parton distributions with qcd-pegasus*, *Comput. Phys. Commun.* **170** (2005) 65–92, [hep-ph/0408244].

[106] D. A. Kosower, *Evolution of parton distributions*, *Nucl.Phys.* **B506** (1997) 439–467, [hep-ph/9706213].

[107] H. Spiesberger, *QED radiative corrections for parton distributions*, *Phys.Rev.* **D52** (1995) 4936–4940, [hep-ph/9412286].

[108] M. Roth and S. Weinzierl, *QED corrections to the evolution of parton distributions*, *Phys.Lett.* **B590** (2004) 190–198, [hep-ph/0403200].

[109] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *Parton distributions incorporating QED contributions*, *Eur. Phys. J.* **C39** (2005) 155–161, [hep-ph/0411040].

[110] S. Weinzierl, *Fast evolution of parton distributions*, *Comput.Phys.Commun.* **148** (2002) 314–326, [hep-ph/0203112].

[111] A. Buckley, J. Ferrando, S. Lloyd, K. Nordstrom, B. Page, et al., *LHAPDF6: parton density access in the LHC precision era*, arXiv:1412.7420.

[112] **The NNPDF collaboration** Collaboration, L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo, *Neural network determination of parton distributions: The nonsinglet case*, *JHEP* **03** (2007) 039, [hep-ph/0701127].

[113] **The NNPDF** Collaboration, R. D. Ball et al., *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* **B809** (2009) 1–63, [arXiv:0808.1231].

[114] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nucl. Phys.* **B126** (1977) 298.

[115] V. N. Gribov and L. N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[116] 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.

[117] E. Floratos, D. Ross, and C. Sachrajda *Nucl.Phys.* **B129** (1977) 66.

[118] E. Floratos, D. Ross, and C. Sachrajda *Nucl.Phys.* **B152** (1979) 493.

[119] A. Gonzalez-Arroyo, C. Lopez, and F. Yndurain *Nucl.Phys.* **B153** (1979) 161.

[120] E. Floratos, C. Kounnas, and R. Lacaze *Nucl.Phys.* **B192** (1981) 417.

[121] G. Curci, W. Furmanski, and R. Petronzio *Nucl.Phys.* **B175** (1980) 27.

[122] S. Moch, J. Vermaseren, and A. Vogt *Nucl.Phys.* **B688** (2004) 101.

[123] S. Moch, J. Vermaseren, and A. Vogt *Phys. Lett.* **B691** (2004) 129.

[124] M. Buza, Y. Matiounine, J. Smith, R. Migneron, and W. L. van Neerven, *Heavy quark coefficient functions at asymptotic values Q $2 \gg m$ 2*, *Nucl. Phys.* **B472** (1996) 611–658, [hep-ph/9601302].

[125] R. Sadykov, *Impact of QED radiative corrections on Parton Distribution Functions*, arXiv:1401.1133.

[126] M. Dittmar et al., *Working Group I: Parton distributions: Summary report for the HERA LHC Workshop Proceedings*, hep-ph/0511119.

[127] J. Blumlein, S. Riemersma, M. Botje, C. Pascaud, F. Zomer, et al., *A Detailed comparison of NLO QCD evolution codes*, hep-ph/9609400.

[128] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, *Eur.Phys.J.* **C66** (2010) 503–524, [arXiv:0911.2985].

[129] **Particle Data Group** Collaboration, K. Olive et al., *Review of Particle Physics*, *Chin.Phys.* **C38** (2014) 090001.

[130] **The NNPDF collaboration** Collaboration, R. D. Ball et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, *Nucl. Phys.* **B838** (2010) 136–206, [arXiv:1002.4407].

[131] V. Bertone, R. Frederix, S. Frixione, J. Rojo, and M. Sutton, *aMCfast: automation of fast NLO computations for PDF fits*, *JHEP* **1408** (2014) 166, [arXiv:1406.7693].

[132] **ATLAS** Collaboration, 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].

[133] V. Bertone, S. Carrazza, and E. R. Nocera, *Reference results for time-like evolution up to $\mathcal{O}\left(\alpha_s^3\right)$*, *JHEP* **1503** (2015) 046, [arXiv:1501.0049].

[134] G. D'Agostini, *Bayesian reasoning in data analysis: A critical introduction*. World Scientific, 2003.

[135] T. Kluge, K. Rabbertz, and M. Wobisch, *Fast pQCD calculations for PDF fits*, hep-ph/0609285.

[136] D. J. Montana and L. Davis, *Training Feedforward Neural Networks Using Genetic Algorithms*, in *Proceedings of the 11th International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'89, (San Francisco, CA, USA), pp. 762–767, Morgan Kaufmann Publishers Inc., 1989.

[137] **New Muon** Collaboration, M. Arneodo et al., *Accurate measurement of $F_2(d)/F_2(p)$ and $R(d) - R(p)$*, *Nucl. Phys.* **B487** (1997) 3–26, [hep-ex/9611022].

[138] **New Muon** Collaboration, M. Arneodo et al., *Measurement of the proton and deuteron structure functions, $F_2(p)$ and $F_2(d)$, and of the ratio $\sigma(L)/\sigma(T)$*, *Nucl. Phys.* **B483** (1997) 3–43, [hep-ph/9610231].

[139] **BCDMS** Collaboration, A. C. Benvenuti et al., *A high statistics measurement of the proton structure functions $F_2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2*, *Phys. Lett.* **B223** (1989) 485.

[140] **H1 and ZEUS** Collaboration, A. F. et al., *Combined Measurement and QCD Analysis of the Inclusive ep Scattering Cross Sections at HERA*, arXiv:0911.0884.

[141] **ZEUS** Collaboration, J. Breitweg et al., *Measurement of $D^{*\pm}$ production and the charm contribution to F_2 in deep inelastic scattering at HERA*, *Eur. Phys. J.* **C12** (2000) 35–52, [hep-ex/9908012].

[142] **ZEUS** Collaboration, S. Chekanov et al., *Measurement of $D^{*\pm}$ production in deep inelastic $e^{\pm}p$ scattering at HERA*, *Phys. Rev.* **D69** (2004) 012004, [hep-ex/0308068].

[143] **ZEUS** Collaboration, S. Chekanov et al., *Measurement of $D^{\pm}$ and D0 production in deep inelastic scattering using a lifetime tag at HERA*, *Eur. Phys. J.* **C63** (2009) 171–188, [arXiv:0812.3775].

[144] **ZEUS** Collaboration, S. Chekanov et al., *Measurement of charm and beauty production in deep inelastic ep scattering from decays into muons at HERA*, *Eur. Phys. J.* **C65** (2010) 65–79, [arXiv:0904.3487].

[145] **H1** Collaboration, C. Adloff et al., *Measurement of $D^{*\pm}$ meson production and $F_2(c)$ in deep inelastic scattering at HERA*, *Phys. Lett.* **B528** (2002) 199–214, [hep-ex/0108039].

[146] **H1** Collaboration, F. D. Aaron et al., *Measurement of the D* Meson Production Cross Section and F_2^{ccbar}, at High Q^2, in ep Scattering at HERA*, *Phys. Lett.* **B686** (2010) 91–100, [arXiv:0911.3989].

[147] **H1** Collaboration, F. D. Aaron et al., *Measurement of the Charm and Beauty Structure Functions using the H1 Vertex Detector at HERA*, *Eur. Phys. J.* **C65** (2010) 89–109, [arXiv:0907.2643].

[148] **ZEUS** Collaboration, S. Chekanov et al., *Measurement of high-Q^2 neutral current deep inelastic e^-p scattering cross sections with a longitudinally polarised electron beam at HERA*, *Eur. Phys. J.* **C62** (2009) 625–658, [arXiv:0901.2385].

[149] **ZEUS** Collaboration, S. Chekanov et al., *Measurement of charged current deep inelastic scattering cross sections with a longitudinally polarised electron beam at hera*, *Eur. Phys. J.* **C61** (2009) 223–235, [arXiv:0812.4620].

[150] **CHORUS** Collaboration, G. Onengut et al., *Measurement of nucleon structure functions in neutrino scattering*, *Phys. Lett.* **B632** (2006) 65–75.

[151] **NuTeV** Collaboration, M. Goncharov et al., *Precise measurement of dimuon production cross-sections in nu/mu Fe and anti-nu/mu Fe deep inelastic scattering at the Tevatron*, *Phys. Rev.* **D64** (2001) 112006, [hep-ex/0102049].

[152] D. A. Mason, *Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data*, . FERMILAB-THESIS-2006-01.

[153] G. Moreno et al., *Dimuon production in proton - copper collisions at $\sqrt{s}$ = 38.8-GeV*, *Phys. Rev.* **D43** (1991) 2815–2836.

[154] **NuSea** Collaboration, J. C. Webb et al., *Absolute Drell-Yan dimuon cross sections in 800-GeV/c p p and p d collisions*, hep-ex/0302019.

[155] J. C. Webb, *Measurement of continuum dimuon production in 800-GeV/c proton nucleon collisions*, hep-ex/0301031.

[156] **FNAL E866/NuSea** Collaboration, R. S. Towell et al., *Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea*, *Phys. Rev.* **D64** (2001) 052002, [hep-ex/0103030].

[157] **CDF** Collaboration, T. Aaltonen et al., *Direct Measurement of the W Production Charge Asymmetry in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. Lett.* **102** (2009) 181801, [arXiv:0901.2169].

[158] **CDF** Collaboration, T. A. Aaltonen et al., *Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Lett.* **B692** (2010) 232–239, [arXiv:0908.3914].

[159] **D0** Collaboration, V. M. Abazov et al., *Measurement of the shape of the boson rapidity distribution for $p\bar{p} \to Z/gamma^* \to e^+e^- + X$ events produced at $\sqrt{s}$ of 1.96-TeV*, *Phys. Rev.* **D76** (2007) 012003, [hep-ex/0702025].

[160] **CDF** Collaboration, 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].

[161] **D0** Collaboration, 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. Lett.* **101** (2008) 062001, [arXiv:0802.2400].

[162] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys.Rev.Lett.* **109** (2012) 111806, [arXiv:1206.2598].

[163] **LHCb** Collaboration, R. Aaij et al., *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, arXiv:1204.1620.

[164] **CMS** Collaboration, *Measurement of the differential and double-differential Drell-Yan cross section in proton-proton collisions at 7 TeV*, CMS-PAS-SMP-13-003.

[165] **ATLAS** Collaboration, 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].

[166] **CMS** Collaboration, 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), no. 11 112002, [arXiv:1212.6660].

[167] J. M. Campbell, H. B. Hartanto, and C. Williams, *Next-to-leading order predictions for $Z\gamma$+jet and $Z\,\gamma\gamma$ final states at the LHC*, *JHEP* **1211** (2012) 162, [arXiv:1208.0566].

[168] J. Campbell, R. K. Ellis, and F. Tramontano, *Single top production and decay at next-to-leading order*, *Phys. Rev.* **D70** (2004) 094012, [hep-ph/0408158].

[169] S. Catani, G. Ferrera, and M. Grazzini, *W Boson Production at Hadron Colliders: The Lepton Charge Asymmetry in NNLO QCD*, *JHEP* **1005** (2010) 006, [arXiv:1002.3115].

[170] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the Inclusive Jet Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV*, *Phys.Rev.Lett.* **107** (2011) 132001, [arXiv:1106.0208].

[171] **CMS** Collaboration, 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].

[172] Z. Nagy, *Next-to-leading order calculation of three-jet observables in hadron hadron collision*, *Phys. Rev.* **D68** (2003) 094002, [hep-ph/0307268].

[173] S. D. Ellis, Z. Kunszt, and D. E. Soper, *The one jet inclusive cross-section at order α_s^3 quarks and gluons*, *Phys.Rev.Lett.* **64** (1990) 2121.

[174] J. Gao, Z. Liang, D. E. Soper, H.-L. Lai, P. M. Nadolsky, et al., *MEKS: a program for computation of inclusive jet cross sections at hadron colliders*, arXiv:1207.0513.

[175] **fastNLO** Collaboration, M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober, *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*, arXiv:1109.1310.

[176] J. Currie, A. Gehrmann-De Ridder, E. Glover, and J. Pires, *NNLO QCD corrections to jet production at hadron colliders from gluon scattering*, *JHEP* **1401** (2014) 110, [arXiv:1310.3993].

[177] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover, and J. Pires, *Second order QCD corrections to jet production at hadron colliders: the all-gluon contribution*, *Phys.Rev.Lett.* **110** (2013), no. 16 162003, [arXiv:1301.7310].

[178] N. Kidonakis and J. F. Owens, *Effects of higher-order threshold corrections in high-E(T) jet production*, *Phys. Rev.* **D63** (2001) 054019, [hep-ph/0007268].

[179] D. de Florian, P. Hinderer, A. Mukherjee, F. Ringer, and W. Vogelsang, *Approximate next-to-next-to-leading order corrections to hadronic jet production*, *Phys.Rev.Lett.* **112** (2014) 082001, [arXiv:1310.7192].

[180] S. Carrazza and J. Pires, *Perturbative QCD description of jet data from LHC Run-I and Tevatron Run-II*, *JHEP* **1410** (2014) 145, [arXiv:1407.7031].

[181] **ATLAS** Collaboration, 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].

[182] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **0804** (2008) 063, [arXiv:0802.1189].

[183] M. Dasgupta, L. Magnea, and G. P. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, *JHEP* **02** (2008) 055, [arXiv:0712.3014].

[184] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez, *Quantifying the performance of jet definitions for kinematic reconstruction at the LHC*, *JHEP* **12** (2008) 032, [arXiv:0810.1304].

[185] S. Forte, *Parton distributions at the dawn of the LHC*, *Acta Phys.Polon.* **B41** (2010) 2859–2920, [arXiv:1011.5247].

[186] M. Ciafaloni, P. Ciafaloni, and D. Comelli, *Towards collinear evolution equations in electroweak theory*, *Phys.Rev.Lett.* **88** (2002) 102001, [hep-ph/0111109].

[187] P. Ciafaloni and D. Comelli, *Electroweak evolution equations*, *JHEP* **0511** (2005) 022, [hep-ph/0505047].

[188] **The NNPDF** Collaboration, R. D. Ball et al., *Reweighting NNPDFs: the W lepton asymmetry*, *Nucl. Phys.* **B849** (2011) 112–143, [arXiv:1012.0836].

[189] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., *Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data*, *Nucl.Phys.* **B855** (2012) 608–638, [arXiv:1108.1758].

[190] K.-P. Diener, S. Dittmaier, and W. Hollik, *Electroweak higher-order effects and theoretical uncertainties in deep-inelastic neutrino scattering*, *Phys.Rev.* **D72** (2005) 093002, [hep-ph/0509084].

[191] G. Altarelli, S. Forte, and G. Ridolfi, *On positivity of parton distributions*, *Nucl. Phys.* **B534** (1998) 277–296, [hep-ph/9806345].

[192] J. Londergan and A. W. Thomas, *Charge symmetry violating contributions to neutrino reactions*, *Phys.Lett.* **B558** (2003) 132–140, [hep-ph/0301147].

[193] **The NNPDF collaboration** Collaboration, R. D. Ball et al., *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, *JHEP* **05** (2010) 075, [arXiv:0912.2276].

[194] F. Demartin, S. Forte, E. Mariani, J. Rojo, and A. Vicini, *The impact of PDF and alphas uncertainties on Higgs Production in gluon fusion at hadron colliders*, *Phys. Rev.* **D82** (2010) 014002, [arXiv:1004.0962].

[195] S. Alekhin, S. Alioli, R. D. Ball, V. Bertone, J. Blumlein, et al., *The PDF4LHC Working Group Interim Report*, arXiv:1101.0536.

[196] R. D. Ball, *Resummation of Hadroproduction Cross-sections at High Energy*, *Nucl.Phys.* **B796** (2008) 137–183, [arXiv:0708.1277].

[197] **ZEUS** Collaboration, S. Chekanov et al., *Measurement of isolated photon production in deep inelastic ep scattering*, *Phys.Lett.* **B687** (2010) 16–25, [arXiv:0909.4223].

[198] A. De Rujula and W. Vogelsang, *On the photon constituency of protons*, *Phys.Lett.* **B451** (1999) 437–444, [hep-ph/9812231].

[199] A. Gehrmann-De Ridder, T. Gehrmann, and E. Poulsen, *Isolated photons in deep inelastic scattering*, *Phys.Rev.Lett.* **96** (2006) 132002, [hep-ph/0601073].

[200] **ZEUS** Collaboration, S. Chekanov et al., *Observation of isolated high E(T) photons in deep inelastic scattering*, *Phys.Lett.* **B595** (2004) 86–100, [hep-ex/0402019].

[201] **CMS** Collaboration, S. Chatrchyan et al., *Search for leptonic decays of W ' bosons in pp collisions at* $\sqrt{s} = 7$ *TeV*, *JHEP* **1208** (2012) 023, [arXiv:1204.4764].

[202] **CMS** Collaboration, S. Chatrchyan et al., *Search for a* W' *boson decaying to a muon and a neutrino in pp collisions at* $\sqrt{s} = 7$ *TeV*, *Phys.Lett.* **B701** (2011) 160–179, [arXiv:1103.0030].

[203] **CMS** Collaboration, S. Chatrchyan et al., *Search for Resonances in the Dilepton Mass Distribution in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *JHEP* **1105** (2011) 093, [arXiv:1103.0981].

[204] **ATLAS** Collaboration, G. Aad et al., *Search for dilepton resonances in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, *Phys.Rev.Lett.* **107** (2011) 272002, [arXiv:1108.1582].

[205] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of* W^+W^- *and* ZZ *production cross sections in pp collisions at* $\sqrt{s} = 8$ *TeV*, *Phys.Lett.* **B721** (2013) 190–211, [arXiv:1301.4698].

[206] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of* W^+W^- *Production and Search for the Higgs Boson in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *Phys.Lett.* **B699** (2011) 25–47, [arXiv:1102.5429].

[207] **ATLAS** Collaboration, G. Aad et al., *Measurement of* W^+/W^- *production in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector and limits on anomalous* WWZ *and* $WW\gamma$ *couplings*, arXiv:1210.2979.

[208] **CMS** Collaboration, S. Chatrchyan et al., *Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV*, arXiv:1212.1910.

[209] **CMS** Collaboration, S. Chatrchyan et al., *Search for exotic resonances decaying into* WZ/ZZ *in pp collisions at* $\sqrt{s} = 7$ *TeV*, *JHEP* **1302** (2013) 036, [arXiv:1211.5779].

[210] **CMS** Collaboration, S. Chatrchyan et al., *Search for a* W' *or Techni-*ρ *Decaying into* WZ *in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *Phys.Rev.Lett.* **109** (2012) 141801, [arXiv:1206.0433].

[211] **ATLAS** Collaboration, G. Aad et al., *Search for resonant diboson production in the $WW/WZ \to l\nu jj$ decay channels with the ATLAS detector at $\sqrt{s} = 7$ TeV*, arXiv:1305.0125.

[212] **ATLAS** Collaboration, G. Aad et al., *Search for new phenomena in the WW to $\ell\nu\ell'\nu'$ final state in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys.Lett.* **B718** (2013) 860–878, [arXiv:1208.2880].

[213] L. Randall and R. Sundrum, *An Alternative to compactification*, *Phys.Rev.Lett.* **83** (1999) 4690–4693, [hep-th/9906064].

[214] J. Andersen, O. Antipin, G. Azuelos, L. Del Debbio, E. Del Nobile, et al., *Discovering Technicolor*, *Eur.Phys.J.Plus* **126** (2011) 81, [arXiv:1104.1255].

[215] E. Eichten and K. Lane, *Low-scale technicolor at the Tevatron and LHC*, *Phys.Lett.* **B669** (2008) 235–238, [arXiv:0706.2339].

[216] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **1107** (2011) 018, [arXiv:1105.0020].

[217] J. Kuhn, F. Metzler, A. Penin, and S. Uccirati, *Next-to-Next-to-Leading Electroweak Logarithms for W-Pair Production at LHC*, *JHEP* **1106** (2011) 143, [arXiv:1101.2563].

[218] R. Boughezal, Y. Li, and F. Petriello, *Disentangling radiative corrections using the high-mass Drell-Yan process at the LHC*, *Phys.Rev.* **D89** (2014), no. 3 034030, [arXiv:1312.3972].

[219] S. Catani and 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, [hep-ph/0703012].

[220] **ATLAS** Collaboration, G. Aad et al., *Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **1406** (2014) 112, [arXiv:1404.1212].

[221] **The NNPDF** Collaboration, R. D. Ball et al., *Parton distributions with QED corrections*, arXiv:1308.0598.

[222] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur.Phys.J.* **C74** (2014), no. 8 3024, [arXiv:1404.5630].

[223] S. Kallweit, J. M. Lindert, P. Maierhfer, S. Pozzorini, and M. Schnherr, *NLO electroweak automation and precise predictions for W+multijet production at the LHC*, arXiv:1412.5157.

[224] S. Carrazza, S. Forte, and J. Rojo, *Parton Distributions and Event Generators*, arXiv:1311.5887.