Master thesis

Master of High Energy, Astrophysics and Physic Computational

Prepared at : Laboratoire de l'Accélérateur Linéaire (LAL)

The impact of LHC measurements on the precision of the Parton Distribution Functions of the Proton

---

Submitted by :
TNOURJI Abdellah

Under the guidance of :
Mohamed EL Kacimi
Zhiqing Zhang
Louis Fayard

Defensed in:

Faculty of science Semlalia on **22/07/2019**

Jury :

Prof. Mohamed EL KACIMI (FSSM)
Prof. Mustapha OULEN (FSSM)
Prof. Abderrahim ADAHCHOUR (FSSM)

2018/2019

Acknowledgment

After almost five months of study at LAL I find that there are many people I would like to thank for making my time here so memorable. First and foremost I would like to thank my supervisor **Zhang Zhiqing** and **Louis Fayard** for their advice and expertise, along with their encouragement and interest in my progress as physicist. I wouldn't have made it this far without their guidance and direction.

During my internship I've had the pleasure of working within the PDF collaboration. From the collaboration I have received an incredible amount of support, learnt a great deal, and made many good friends. Special thanks go to **Sasha Glasov** and **Francesco Giuli** who shared their experience with me and without their help and guidance it will be impossible to complete this wonderful work.

I would also like to thank **Prof. Mohamed EL KACIMI** for kindly agreeing to be the second referee of this internship and all his help and support during my master degree and my last CERN summer student and during my preparation to my defense.

I wish to express my thanks to **Prof. Driss GOUJDAMI** for his helps during my internship. I would also to express my deepest thanks to **Prof. Abderrahim ADAHCHOUR** for accepting to be a member of the jury of my master defense.

Special thanks go to **Marc Escalier** and **Lydia Fayard** for their help and welcome during my internship and all ATLAS-LAL team.

I would to express my thanks and gratitude to the PHEA laboratory at Marrakech specially the master coordinator **Prof. Mustapha OULEN** and the

laboratory director **Prof. Mohamed CHABAB** and all laboratory teams for them support and hard work to present and prepare a very good master degree and conditions to get our diploma.

I express my deepest thanks to my friend **ATMANI Hicham** for helping me to integrate into the ATLAS-LAL group and to my friend **BELFKIR Mohamed** for his valuables information on programming. Many thanks to all my friends in my master class for the best and enjoyable moments.

I specifically would like to acknowledge and extend my sincere thanks to **my family**, particularly my parents who have supported me throughout my life and studies so far. This work is dedicated to them.

Abstract

Recent measurements of top quark pair production cross section and triple-differential cross section for the Drell-Yan process, which are performed at the LHC, are studied using Hessian profiling technique to obtain their impact on the parton distribution functions (PDFs). It is explained how the Hessian profiling method may be used to assess the impact of these new data on PDFs and consequently on their predictions. We discuss issues observed with xFitter and the solutions we come up with. We establish the first comparison between the new software package, ePump (error PDF Updating Method Package) and xFitter. Special care is given to discuss the potential of each programs and the assumptions made of this theoretical framework compared to a treatment by the full global-analysis program.

Contents

List of Figures

Chapter 1

Introduction

The determination of the quark and gluon structure of the proton is a central component of the precision phenomenology program at the Large Hadron Collider (LHC). This internal structure is quantified in the collinear QCD factorisation framework by the Parton Distribution Functions (PDFs), which encode the information related to the momentum distribution of quarks and gluons within the proton. Being driven by lowscale nonperturbative dynamics, PDFs cannot be computed from first principles, at least with current technology, and therefore they need to be determined using experimental data. Most of these data come from legacy experiments, such as Deep Inelastic Scattering (DIS) experiments, various fixed target hadron experiments, and the Fermilab Tevatron collider. LHC experimental results are beginning to be used in global PDF fits, and in the coming decades new knowledge of PDFs will come from measurements at ATLAS, CMS, and LHCb. This global QCD analysis program involves combining the most PDF-sensitive data and the highest precision QCD and electroweak calculations available within a statistically robust fitting methodology. See Ref. [1].

PDFs parametrise the unknown non-perturbative dynamics of the proton. As a universal property of protons, the PDFs may be determined from available experimental data and then applied in the calculation of predictions for other experiments, therefore making the application of QCD in hadron collisions into a predictive theory which may be tested via comparison to data (see chapter 3).

This report is motivated by the fact that the recent years have seen a number of important breakthroughs in our understanding of the quark and gluon structure of the proton. To begin with, the impressive recent progress in NNLO QCD calculations has now made it possible to include essentially all

relevant collider cross sections consistently into a NNLO global analysis, from top quark differential distributions [2] to inclusive jets [3] and dijets [4], isolated photons [5], and the p_{T} distribution of Z bosons [6, 7]. There has also been a recent explosion in the number of tools available for PDF studies, from the open-source fitting framework xFitter [8] to new fast (N)NLO interfaces and public codes for the PDF evolution [9, 10] and the efficient calculation of hadronic cross sections [11] and recently, ePump [14].

Constraining PDFs and their uncertainties is now an intense research program. The systematic uncertainty in the PDF models arises from the 1) experimental uncertainties of the input data used in a global fit, 2) any theoretical assumptions made by the fitting groups, and/or 3) the chosen parameterizations characterizing the functional forms of the PDFs themselves.

In order to obtain the most comprehensive PDF constraints, different theoretical groups perform the global QCD fits of the experimental data. To study the impact of new experimental measurements, one can perform a QCD global analysis by including the new data to the base data. To estimate the impact of new experimental measurements on the PDFs, we can use the approximate methods that can be used instead of a complete QCD fit, as an alternative approach. In this regard, one can use the Bayesian Monte Carlo reweighting and Hessian profiling techniques [13], as the approximate methods.

The goal of this report is to find the impact of the new measurements of the production cross section of the top quark pair ($t\bar{t}$) [12] on the modern CT14 PDF sets (see section 3.5.2) using the Hessian profiling technique[13], without performing a complete baseline global PDF fit. In this report, the QCD analysis is performed based on xFitter open source framework [8] and ePump [14]. The ePump package is not a substitute for full global fitting, but can be used as a tool to probe the effects of new data. The recent top quark production data which are not included in the main xFitter and ePump packages are added.

The focus of this report is to show the new top quark data can be used to constrain the PDFs (especially in the central value) and uncertainties or both for the s-quark and gluon PDFs at the large-x using xFitter and ePump [14], and a comparison of these two packages is shown for the first time. We will also shed light on the impact of Drell-Yan (ZD3) data on PDFs using xFitter package but no comparison with ePump packages is made for this case.

This report is arranged as follows. In chapter 2, a brief description of the LHC collider and the ATLAS detector. In chapter 3 we shall provide a concise discussion of the theoretical structure of parton distributions, where they arise in the calculation of DIS cross-sections and further theoretical background relevant to the reliable determination of PDFs from experimental data. Also, a brief review of the Hessian profiling method is described. Chapter 4 describes the two packages used in analysis. In chapter 5, the data samples and the theoretical calculation and tools of the present analysis are explained. Finally, the results obtained in the paper are summarized in chapter 6, where the comparison of xFitter and ePump is shown. A summary of the manuscript is presented in chapter 7.

Chapter 2

LHC and ATLAS detector

2.1 The Large Hadron Collider

The LHC is a super-conducting accelerator and collider installed in a 27 km long circular tunnel that is buried 100 m underground. The LHC is located at the European Organization for Nuclear Research (CERN). It sits across the border of France and Switzerland, near the city of Geneva. A diagram of the LHC is shown in Fig. 2.1. The tunnel was originally constructed between 1984 and 1989 for the CERN LEP machine.

The LHC collides protons at four locations along the ring of the machine, corresponding to the location of the four LHC experiments: ALICE, ATLAS, CMS, and LHCb. Inside the LHC, beams of protons travel in opposite directions in separate beam pipes. They are guided around the accelerator ring by a strong magnetic field, achieved with super-conducting magnets. The LHC is designed to produce collisions with a center of mass energy of up to $\sqrt{s} = 14$ TeV.

The LHC is only the final stage in a series of machines used to accelerate the protons to increasingly higher energies. Protons, obtained from hydrogen atoms, begin the chain in a linear accelerator called Linac 2. The Linac 2 accelerates the protons to 50 MeV. The protons are then injected in to the Proton Synchrotron (PS) Booster, which accelerates them to 1.4 GeV. After the PS Booster, the protons are sent to the PS where they are accelerated to 25 GeV. They are then sent to the Super PS (SPS) where they are accelerated to 450 GeV. They are finally injected into the LHC where they are accelerated to their final energy. Under normal operating conditions, the colliding beams will circulate for many hours at a time (Fig. 2.1).

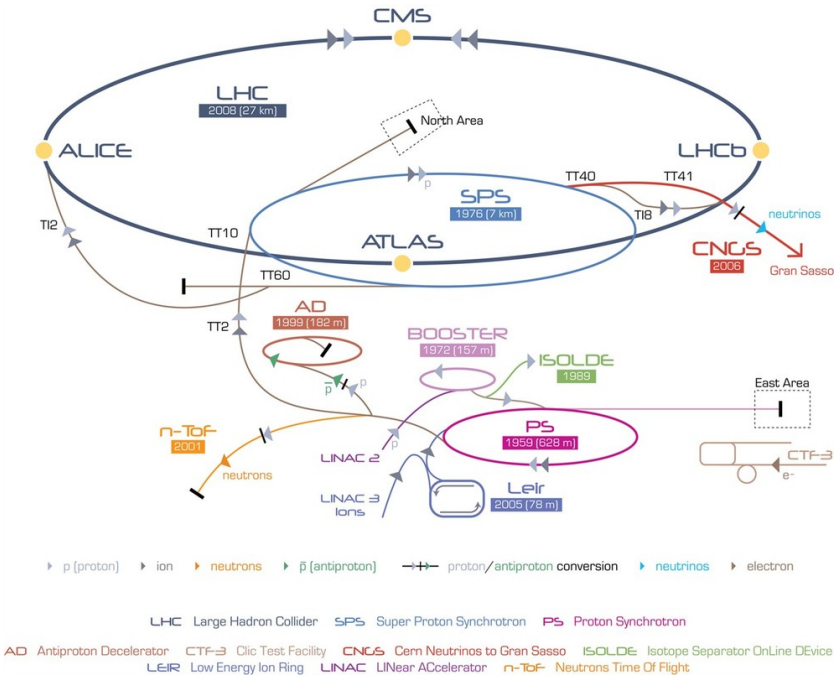


Figure 2.1: Overview of the CERN accelerator complex.

As a consequence of the acceleration scheme, the proton beams circulate the ring in bunches. Under nominal operating conditions, each proton beam has 2808 bunches, with each bunch containing about 10^{11} protons. These bunches are a few centimeters long and about 16 μm wide when they collide. As a result, each bunch crossing produces many pp interactions. The 2012 running had as many as 30 interactions per bunch crossing (Fig. 2.2).

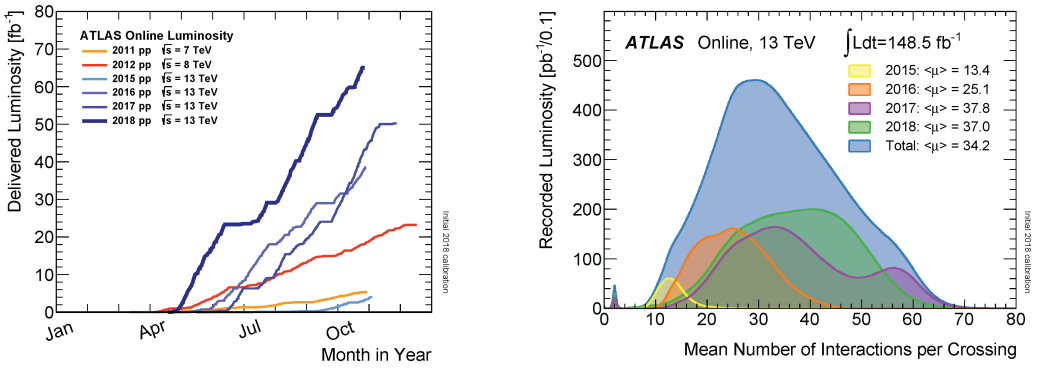
2.1.1 LHC performance

The other important characteristic of the LHC data is the luminosity. The luminosity is proportional to the number of collisions produced by the accelerator:

$$N_{\text{events}} = \mathcal{L} \times \sigma \,. \tag{2.1}$$

The rate of interactions (R_{inel}) is related to the inelastic cross section of the proton–proton collision(σ_{inel}), and the luminosity can be expressed as [16] :

$$\mathcal{L} = \frac{R_{\text{inel}}}{\sigma_{\text{inel}}} \,. \tag{2.2}$$



(a) (b)

Figure 2.2: (a) Cumulative luminosity versus day for 2011 – 2018 delivered to ATLAS during stable beams for high energy pp collisions [15]. (b) Mean number of interactions per bunch crossing, showing the 13 TeV data from 2015 – 2018 [15]

The rate can be related to the average number of inelastic interactions per bunch crossing (μ), the number of bunches per ring (k_b) and the revolution frequency (f_{rev}):

$$\mathcal{L} = \frac{\mu f_{\text{rev}} k_b}{\sigma_{\text{inel}}} \,. \tag{2.3}$$

The luminosity is the quality factor for colliders, measuring the intensity of the beam, and can be written in terms of the accelerator parameters:

$$\mathcal{L} = \frac{N_b^2 f_{\text{rev}} k_b \gamma}{4\pi \epsilon_n \beta^*} F \,, \tag{2.4}$$

where :

- N_b the number of protons per bunch ($\sim 10^{11}$),
- k_b is the number of bunch per beam (2808),
- γ relativistic factor,
- ϵ_n the normalized transverse beam emittance - characterizes its spread in coordinate and momentum phase space,
- the beta-function at the interaction point, determined by the magnets configuration,

- β^* the beta-function at the interaction point, determined by the magnets configuration,
- F is the geometric luminosity reduction factor due to the crossing angle at the interaction point.

2.1.2 The LHC upgrade plan

The LHC baseline program until 2025 is shown schematically in Fig. 2.3. After entering into the nominal energy regime of 1314 TeV centre-of-mass energy in 2015, it is expected that the LHC will reach the design luminosity of $1 \times 10^{34}\,\mathrm{cm}^{-2}s^{-1}$. This peak value should give a total integrated luminosity of about 40 fb^{-1} per year. In the period 2015-2022 the LHC will hopefully further increase the peak luminosity. Margins in the design of the nominal LHC are expected to allow, in principle, about two times the nominal design performance. The baseline program for the next ten years is depicted in Fig. 2.3, while Fig. 2.4 shows the possible evolution of peak and integrated luminosity.

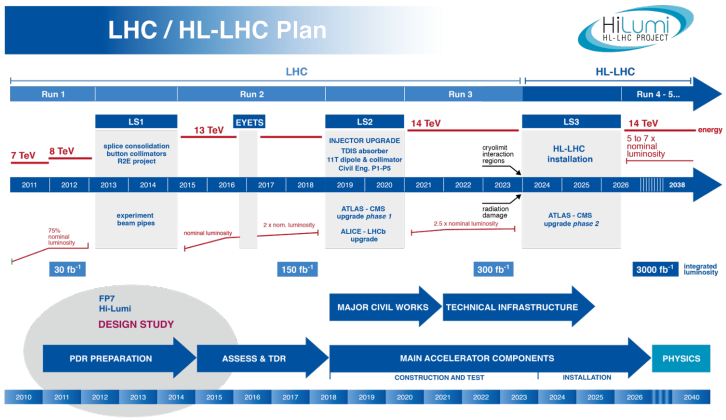


Figure 2.3: LHC baseline plan for the next decade and beyond showing the energy of the collisions (upper red line) and luminosity (lower green lines). The first long shutdown (LS1) in 2013-2014 allowed the design parameters of beam energy and luminosity to be reached. The second long shutdown (LS2) in 2019-2020, consolidates luminosity and reliability as well as the upgrading of the LHC injectors. After LS3, 2023-2025, the machine will be in the High Luminosity configuration (HL-LHC).

The main objective of the High Luminosity LHC (HL-LHC) design study

was to determine a set of beam parameters and the hardware configuration that will enable the LHC to reach the following targets:

- A peak luminosity of 5×10^{34} cm$^{-2}s^{-1}$, with levelling, allowing:
- An integrated luminosity of 250 fb^{-1} per year with the goal of 3000 fb^{-1} in about a dozen years after the upgrade. This integrated luminosity is about ten times the expected luminosity reach of the first twelve years of the LHC lifetime.

The overarching goals are the installation of the main hardware for the HL-LHC and the commissioning of the new machine configuration during LS3, scheduled for 2023-2025, while taking all actions to assure a high efficiency in operation until 2035.

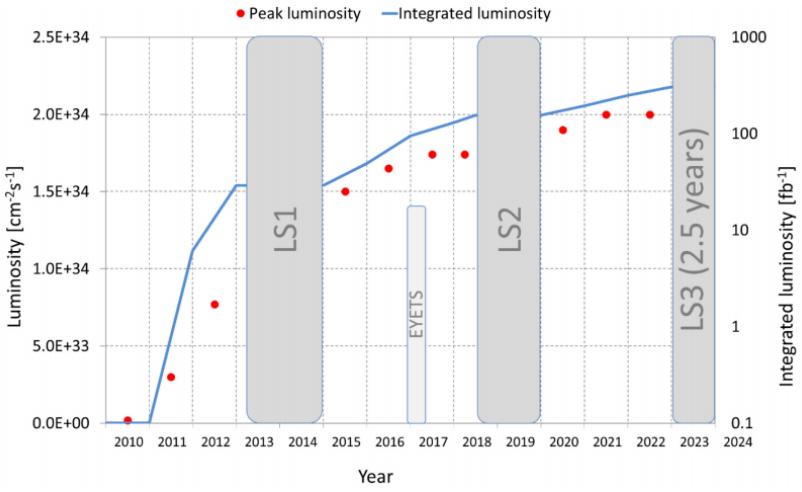


Figure 2.4: LHC luminosity plan for the next decade, both peak (red dots) and integrated (blue line). Main shutdown periods are indicated.

2.2 ATLAS detector

The ATLAS (A Toroidal LHC ApparatuS) experiment (Fig. 2.5) [17] is a multipurpose particle detector with a forward-backward symmetric cylindrical geometry. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a

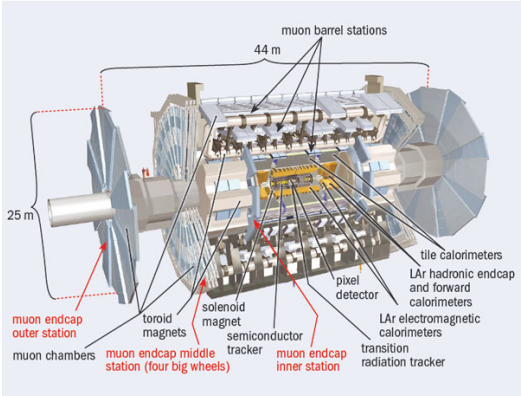


Figure 2.5: The ATLAS detector.

muon spectrometer incorporating three large super-conducting toroid magnets.

The inner detector (ID) is the innermost part of the ATLAS detector used to reconstruct the primary and secondary vertices, and charged 3 particle tracks with high efficiencies over the pseudorapidity range of $|| < 2.5$. It consists of three sub-components all present in a magnetic field parallel to the beam axis: pixel detector (Pixel), semiconductor tracker (SCT), and transition radiation tracker (TRT), each of which is comprised of a barrel and two endcaps (see Fig.2.6)

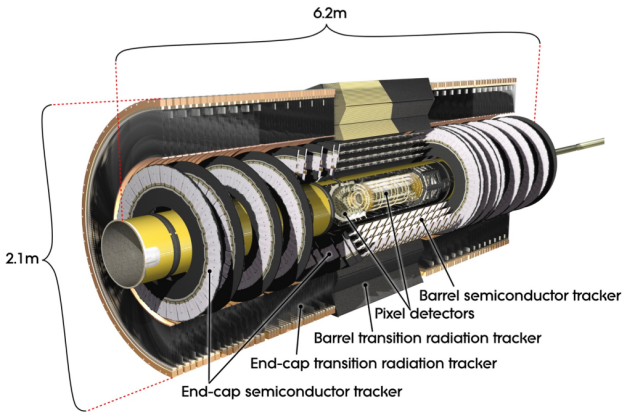


Figure 2.6: The inner-detector system (ID).

The calorimeter system (Fig. 2.7) covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic (EM) calorimetry is provided by high-granularity lead/liquid-argon (LAr) calorimeters, with an additional

thin LAr presampler covering $|\eta| < 1.8$ to correct for upstream energy-loss fluctuations. The EM calorimeter is divided into a barrel section covering $|\eta| < 1.475$ and two endcap sections covering $1.375 < |\eta| < 3.2$. For $|\eta| < 2.5$ it is divided into three layers in depth, which are finely segmented in η and ϕ.

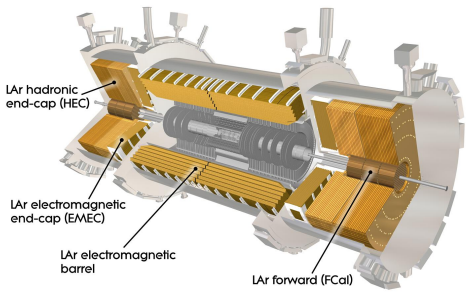


Figure 2.7: The ATLAS calorimeter.

Hadronic calorimetry (Fig. 2.7) is provided by a steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$ and two copper/LAr hadronic endcap calorimeters covering $1.5 < |\eta| < 3.2$. The solid-angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules in $3.1 < |\eta| < 4.9$, optimized for electromagnetic and hadronic measurements, respectively.

The muon spectrometer (MS) (Fig. 2.8) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by superconducting air-core toroids. The precision chamber system covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, complemented by cathode strip chambers in the forward region. The muon trigger system covers the range $|\eta| < 2.4$ with resistive plate chambers in the barrel, and thin gap chambers in the endcap regions. A three-level trigger system is used to select events for offline analysis [18]. The level-1 trigger is implemented in hardware and uses a subset of detector information to reduce the event rate to a design value of at most 75 kHz. This is followed by two software-based trigger levels which together reduce the event rate to about 300 Hz.

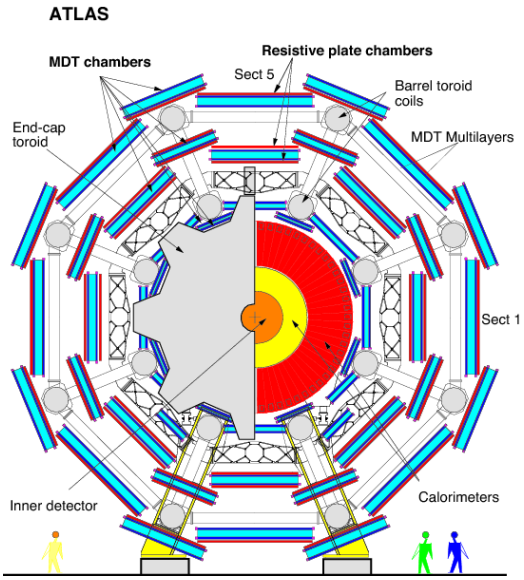


Figure 2.8: The muon spectrometer (MS).

Chapter 3

Parton distribution functions

The quarks inside the proton will interact with each other through the exchange of gluons. The dynamics of this interacting system will result in a distribution of quark momenta within the proton. These distributions are expressed in terms of PDFs.

Parton distributions are one of the central pillars of perturbative QCD, factorizing as they do the perturbatively incalculable long distance dynamics present in calculations involving hadronic initial states. Combined with the perturbative description of the short-distance cross-section what could seem at first a hopeless situation is alleviated, and QCD becomes a predictive and useful theory when applied to hadronic scattering.

In this chapter a brief overview of how PDFs arise in QCD calculations will be presented. We shall explore the prototypical example of the DIS of leptons off a hadronic target, first in the naive parton model arising before the advent of QCD and then with the QCD-improved parton model which allows for an excellent description of DIS measurements across a wide range of hard scales.

The treatment of heavy quarks in parton distributions is a particularly delicate issue and therefore will also be discussed here. Finally there will be some exploration of the general properties of parton distributions in order to provide a summary of the available theoretical constraints upon PDFs.

3.1 Partons in deep inelastic scattering

We shall begin by introducing PDFs as they arise in the early parton model. The model was originally introduced by Feynman and Bjorken [19, 20] in the late 1960s in an effort to understand the scattering behavior of hadronic states and successfully describes many properties observed in early deep inelastic scattering experiments.

In this process, a charged lepton l probes a proton P by the exchange of a gauge boson. For simplicity we shall describe here the neutral current process where a photon is exchanged. In the inelastic regime where the momentum transfer to the target proton is large, the proton does not survive the scattering process and fragments into an arbitrary hadronic final state X. The process $l(k) + P(p) \to l(k') + X$ is illustrated at tree level in Fig. 3.1.

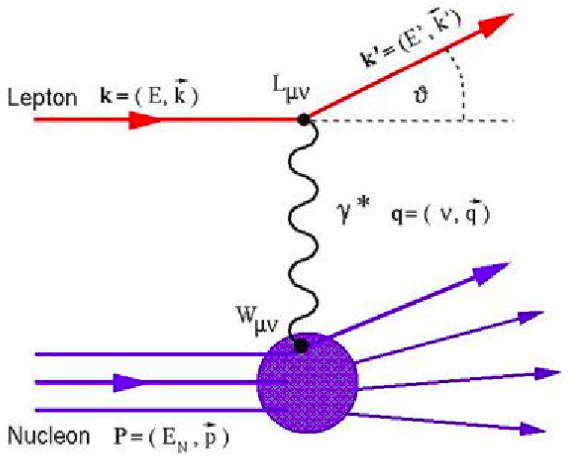


Figure 3.1: Deep inelastic scattering of a charged lepton off a proton target.

Kinematic variables for inelastic scattering:

In this system we can define the following standard DIS kinematic variables.

The hadronic final state resulting from the break-up of the proton usually consists of many particles. The invariant mass of this hadronic system,denoted W, depends on the four momentum of virtual photon,

$$W^2 = -(P + q)^2 \,, \tag{3.1}$$

where Q^2 denotes the momentum transfer from the electron to the target proton,

$$Q^2 = -q^2 \,, \tag{3.2}$$

when writing in term of four-momenta of the initial and final state electrons:

$$Q^2 = -(k - k')^2 \,. \tag{3.3}$$

In inelastic scattering, the energies are sufficiently high that the electron mass can be neglected and therefore, to a very good approximation:

$$Q^2 = 4E(k)E(k')\sin^2\left(\frac{\theta}{2}\right) \,. \tag{3.4}$$

Bjorken x

It is a Lorentz-invariant dimensionless quantity,

$$x = \frac{Q^2}{2P \cdot q} \,. \tag{3.5}$$

The range of possible values of x can be found by writing the four-momentum of the hadronic system in terms of that of the virtual photon,

$$x = \frac{Q^2}{Q^2 + W^2 - m_p^2} \,. \tag{3.6}$$

Because there are three valence quarks in the proton, and quarks and anti-quarks can be produced together only in pairs, the hadronic final state in an inelastic scattering process must include at least one baryon (qqq). Consequently, the invariant mass of the final-stat hadronic system is always greater than the mass of the proton (which is the lightest baryon), thus

$$W^2 \geq m_p^2 \,.$$

Also since $Q^2 \geq 0$, the relation of (3.6) implies that x is always in the range:

$$0 \leq x \leq 1 \,. \tag{3.7}$$

The value of x expresses the "elasticity" of the scattering process. The extreme case of $x = 1$ is equivalent to $W^2 = m_p^2$, and therefore corresponds to elastic scattering.

y and ν

A second dimensionless Lorentz-invariant quantity, the inelasticity y, is defined as:

$$y = \frac{k' \cdot q}{k' \cdot k} \,. \tag{3.8}$$

In the frame where the proton is at rest, $k' = (m_p, 0, 0, 0)$, the momenta of the initial-state lepton, the final-state lepton and the virtual photon can be written as

$$k = (E_k, 0, 0, 0), k' = (E_{k'}, E_{k'} \sin\theta, 0, E_{k'} \cos\theta) \text{ and } q = (E_k - E_{k'}, \vec{p}_k - \vec{p}_{k'})$$

and therefore

$$y = 1 - \frac{E_{k'}}{E_k} \,, \tag{3.9}$$

hence y can be identified as the fractional energy lost by the electron in the scattering process in the frame where the proton is initially at rest. In this frame, the energy of the final stat hadronic system is always greater than the energy of the initial-state proton, $E_W \geq m_p$, which implies the initial-state lepton must lose energy. Consequently, y is constrained to be in the range:

$$0 \leq y \leq 1 \,. \tag{3.10}$$

Sometimes it is more convenient to work in terms of energies, rather than the fractional energy loss described by y. In this case the related quantity

$$\nu = \frac{P \cdot q}{m_p} \tag{3.11}$$

is often used. In the frame where the initial-state proton is at rest,

$$\nu = E_k - E_{k'} \tag{3.12}$$

is simply the energy lost by the initial-state lepton.

Relationship between kinematic variables

For a given centre-of-mass energy s, the kinematics of inelastic scattering are fully defined by specifying two independent observable which are usually chosen to be two of Lorentz-invariant quantities, Q^2, x, y and ν. Provide the chosen quantities are independent, the other two quantities then can be determined through the relations that follow from the definitions,

$$Q^2 = -q^2, x = \frac{Q^2}{2P \cdot q}, y = \frac{k' \cdot q}{k' \cdot k}, \nu = \frac{P \cdot q}{m_p} \,. \tag{3.13}$$

For example, it immediately can be seen that x is related to Q^2 and ν by:

$$x = \frac{Q^2}{2m_p \cdot \nu}\,. \tag{3.14}$$

Furthermore, for a fixed centre-mass-energy,

$$s = (k+P)^2 = k^2 + P^2 + 2k^2P^2 = 2k^2P^2 + m_p^2 + m_k^2\,. \tag{3.15}$$

Since $m_k^2 << m_p^2$, to a good approximation,

$$2k \cdot P \simeq s - m_p^2\,, \tag{3.16}$$

and then from the definitions of (3.13), it can be seen that Q^2 is related to x and y by:

$$Q^2 = (s - m_p^2)xy\,. \tag{3.17}$$

Hence, for a fixed centre-of-mass energy, the kinematics of inelastic scattering can be described by any two of the Lorentz-invariant quantities x, Q^2, y and ν, with the exception of y and ν, which are not independent.

3.1.1 QCD factorization in DIS

The large virtuality Q of the gauge boson, $Q \gg \lambda_{QCD}$, ensures that the process can be described within the perturbative QCD factorization framework in terms of coefficient functions and parton distributions, as we show below.

As discussed in the previous section, only two of the variables in (3.13) are independent, and therefore differential cross sections in DIS are measured for instance as a function of (x, Q^2) or (x, y). Using Lorentz invariance and kinematic arguments, it can be shown that the DIS cross sections can be expressed in terms of a series of independent "structure functions" that describe the dynamics of the interaction between the gauge boson and the incoming hadron. In the neutral current (NC) case, that is, where either a virtual photon γ^* or a Z boson is exchanged, the DIS differential cross section for a charged lepton $l^\pm$ scattering off a proton can be decomposed in terms of structure functions as follows:

$$\begin{aligned}\frac{d^2\sigma^{\mathrm{NC},l^\pm}(x,y,Q^2)}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}[\, & Y_+ F_2^{\mathrm{NC}}(x,Q^2) \mp Y_- x F_3^{\mathrm{NC}}(x,Q^2) \\ & - y^2 F_L^{\mathrm{NC}}(x,Q^2)]\end{aligned} \tag{3.18}$$

where we defined

$$Y_\pm = 1 \pm (1-y)\,. \tag{3.19}$$

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

$$\tilde{\sigma}^{\mathrm{NC},l^{\pm}}(x,y,Q^2) = \left[\frac{2\pi\alpha^2}{xQ^4}Y_+\right]^{-1} \frac{d^2\sigma^{\mathrm{NC},l^{\pm}}(x,y,Q^2)}{dxdQ^2}\,, \tag{3.20}$$

which is more closely related to the dominant structure function $F_2(x,Q^2$, and thus to the underlying PDFs of the proton.

In the case of charged current (CC) DIS, when neutrinos are used as projectiles or when the incoming charged leptons interact with the proton by means of the exchange of a charged weak gauge boson $W^{\pm}$, the differential cross sections are given by:

$$\begin{aligned}\frac{d^2\sigma^{\mathrm{CC},l^{\pm}}(x,y,Q^2)}{dxdQ^2} &= \frac{G_F^2}{4\pi x}\left(\frac{M_W^2}{M_W^2+Q^2}\right)^2 \\ &\times \frac{1}{2}\left[Y_+F_2^{\mathrm{CC}}(x,Q^2) \mp Y_-xF_3^{\mathrm{CC}}(x,Q^2) - y^2F_L^{\mathrm{CC}}(x,Q^2)\right]\end{aligned} \tag{3.21}$$

which is generally rescaled to define a reduced cross section:

$$\tilde{\sigma}^{\mathrm{CC},l^{\pm}}(x,y,Q^2) = \left[\frac{G_F^2}{4\pi x}\left(\frac{M_2^2}{M_W^2+Q^2}\right)^2\right]^{-1} \frac{d^2\sigma^{\mathrm{CC},l^{\pm}}(x,y,Q^2)}{dxdQ^2}\,, \tag{3.22}$$

similarly to the NC case. In (3.21) and (3.22), $l^{\pm}$ labels either the incoming or outgoing charged lepton.

According to the QCD factorization theorem, the general expression for the DIS structure functions can be written schematically as

$$F(x,Q^2) = x\int_x^1 \frac{dy}{y}\sum_i C_i\left(\frac{x}{y},\alpha_s(\mu_R\right),\mu_F,Q)f_i(y,\mu_F)\,, \tag{3.23}$$

where the C_i are known as the coefficient functions, $f_i(y,\mu_F)$ are the PDFs, and $\mu_F(\mu_R)$ are the factorization (renormalization) scales, typically set to $\mu_F = \mu_R = Q$. The coefficient functions represent the cross section for the partonic scattering process $q_i + \gamma^* \to X$, and can be computed in perturbation theory as a series expansion in the strong coupling α_s, as well as in the electroweak coupling α_w if these corrections are included.

While the coefficient functions encode the short distance dynamics of the parton-boson collision, the PDFs are instead determined by long distance

non-perturbative QCD dynamics, and so cannot be computed using perturbative methods. On the other hand, the crucial factorization property of (3.23) is that while the coefficient functions (or in general the partonic cross sections) are process dependent, the PDFs themselves are universal. **This allows us to parameterize and extract the PDFs from a global analysis of hard scattering measurements. These can then be used to make predictions for other PDFdependent processes.**

3.1.2 QCD factorization in hadronic collisions

In a similar way to the DIS structure functions for electronproton collisions, the production cross sections in proton-proton collisions can be factorized in terms of the convolution between two universal PDFs and a processdependent partonic cross section. For example, the Drell-Yan production cross section, $\sigma^{\rm DY}(pp \to l^+l^- + X)$, can be expressed [21, 22] as:

$$\frac{d^2\sigma^{\rm DY}(y, Q^2, \mu_R^2, \mu_F^2)}{dydQ^2} = \sum_{a,b=q,\tilde{q},g} \int_{\tau_1}^1 dx_1 f_a(x_1, \mu_F^2) \int_{\tau_2}^1 dx_2 f_b(x_2, \mu_F^2) \frac{d^2\widehat{\sigma}_{ab}^{\rm DY}(x_1, x_2, y, Q^2, \mu_R^2, \mu_F^2)}{dydQ^2} \tag{3.24}$$

where y and Q^2 are the rapidity and invariant mass squared of the lepton pair, and s is the centre-of-mass energy of the two incoming protons, while $\mu_F(\mu_R)$ are the factorization (renormalization) scales. The lower integration limits are $\tau_{1,2} = \sqrt{Q^2/s}e^{\pm y}$. The partonic cross sections that appear in (3.24) can be computed as a perturbative expansion in α_s:

$$\frac{d^2\widehat{\sigma}_{ab}^{\rm DY}}{dydQ^2}(x_1, x_2, y, Q^2, \mu_R^2, \mu_F^2) = \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu_R^2)}{2\pi}\right)^n \frac{d^2\widehat{\sigma}^{(n)\,\rm DY}_{ab}}{dydQ^2} \,. \tag{3.25}$$

From (3.24) we observe that the definition of the PDFs, once perturbative QCD corrections are accounted for, requires the introduction of a factorization scale μ_F, below which additional collinear emissions are absorbed into a PDF redefinition. To all orders, the physical cross section, as a product of the PDFs and partonic cross section, is independent of the choice of the factorization scale. However, at any fixed order in the perturbative series, **there will be some sensitivity due to the missing higher orders, which can be minimized by choosing a suitable value of μ_F so as to maintain a better convergence of the series.** In Drell-Yan production, the conventional scale choice is $\mu_F = Q^2$, namely the invariant mass of the dilepton pair.

3.1.3 The DGLAP evolution equations

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

For this reason, the Q^2 dependence of the PDFs can be computed in QCD perturbation theory up to any given order. This dependence is determined by a series of integro-differential equations, known as the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [23, 24], which have the generic form:

$$Q^2 \frac{\partial}{\partial Q^2} f_i(x, Q^2) = \sum_j P_{ij}(x, \alpha_s(Q^2)) \otimes f_j(x, Q^2)\,, \tag{3.26}$$

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

$$P_{ij}(x, \alpha_s(Q^2)) = \sum_{n=0} \left(\frac{\alpha_s(Q^2)}{2\pi}\right)^{n+1} P_{ij}^{(n)}(x), \tag{3.27}$$

and where $\otimes$ denotes the convolution:

$$f(x) \otimes g(x) = \int_x^1 \frac{dy}{y} f(y) g(\frac{x}{y}) \tag{3.28}$$

which appears ubiquitously in QCD calculations. The splitting functions (3.27) depend on the type of initial and final state parton that is involved in

the splitting. At leading order, the DGLAP splitting functions are given by

$$P_{qq} = \frac{4}{3}\left[\frac{1+x^2}{(1-x)_+} + \frac{3}{2}\delta(1-x)\right] , \tag{3.29}$$

$$P_{qg} = \frac{1}{2}\left[x^2 + (1-x^2)\right] , \tag{3.30}$$

$$P_{gq} = \frac{4}{3}\left[\frac{1+(1-x)^2}{x}\right] , \tag{3.31}$$

$$P_{gg} = 6\left[\frac{1-x}{x} + x(x-1) + \frac{x}{(1-x)_+} + \frac{33-3n_f}{6}\delta(1-x)\right] . \tag{3.32}$$

Note that both P_{gg} and P_{qg} have a singularity at $x = 0$: this fact is responsible for the rapid growth at small x of the gluons and consequently of the sea quarks in this region. The overall coefficients of the splitting functions are related to the QCD color factors. Some splitting functions exhibit an apparent singularity at $x = 1$, which cancels against those due to virtual corrections and is regularized by means of the plus prescriptions, defined as:

$$\int_0^1 dx f(x)\left[\frac{1}{1-x}\right]_+ = \int_0^1 dx (f(x) - f(1))\left[\frac{1}{1-x}\right] . \tag{3.33}$$

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

$$\Sigma(x,Q^2) \equiv \sum_{n=i}^{n_f}(q_i + \bar{q}_i)(x,Q^2) , \tag{3.34}$$

$$T_3(x,Q^2) \equiv (u + \bar{u} - d - \bar{d})(x,Q^2) , \tag{3.35}$$

$$T_8(x,Q^2) \equiv (u + \bar{u} - d - \bar{d} - 2(s + \bar{s}))(x,Q^2) , \tag{3.36}$$

$$V(x,Q^2) \equiv \sum_{n=i}^{n_f}(q_i - \bar{q}_i)(x,Q^2) , \tag{3.37}$$

$$V_3(x,Q^2) \equiv (u - \bar{u} - d + \bar{d})(x,Q^2) , \tag{3.38}$$

$$V_8(x,Q^2) \equiv (u + \bar{u} + d - \bar{d} - 2(s - \bar{s}))(x,Q^2) . \tag{3.39}$$

Several public codes implement the numerical solution of the DGLAP equations, with the HOPPET [25], APFEL [26] and QCDNUM [27] codes using x-space methods, while the PEGASUS [28] code performs the evolution in Mellin (moment). These codes have undergone detailed benchmarking studies, with agreement at the level of $\mathcal{O}(10^5)$ or better being found [29, 30].

3.2 Treatment of heavy quarks

In the QCD parton model, the assumption that all the quarks contributing in the theory are massless is made, an approximation that becomes increasingly untenable when investigating scattering processes with a hard scale approaching a quarks physical mass. A careful treatment of terms depending on quark masses is therefore vital for making theoretical predictions to a data set that spans heavy quark mass thresholds.

Dealing with heavy quark mass effects is a delicate issue in that different treatments generally have different regions of applicability. The specific combination of approaches to quark masses used when confronting a data set with a broad reach in hard scale is known as a heavy quark scheme, although not necessarily in the spirit of factorization or renormalization schemes as the choice often lies in the particulars of the approximation rather than in some arbitrary shuffling of parameters. A heavy quark scheme choice can therefore potentially lead to differences with alternative calculations that do not in principle vanish in the limit of an all-orders calculation.

The space of heavy quark renormalization schemes is bounded by two regimes where the treatment is fairly simple, the fixed flavour number scheme (FFNS) and the zero-mass variable flavour number scheme (ZM-VFNS). The remaining schemes, known as general-mass variable flavour number schemes (GM-VFNS) aim to interpolate between the FFNS and ZM-VFNS, reducing to the simpler calculations in certain kinematic limits. Motivated by observations suggesting that a more careful treatment of quark mass effects is phenomenologically relevant at the LHC [33], a number of such schemes have arisen in an attempt to better describe experimental data. These typically differ by sub-leading terms in the method of interpolation between the two limiting regimes.

We shall now outline in more details The FFN, ZM-VFNS schemes and general mass schemes.

3.2.1 The FFN and ZM-VFNS schemes

In the FFN scheme (Fig. 3.2) only the gluon and the light quarks are considered as partons within the proton, massive quarks are produced perturbatively in the final state.

In addition, a recent variation of the fixed-flavour number scheme in which

the running mass definition of the heavy quark mass is used in the MS scheme [31]. This variant is realized via the interface to the open-source code OPENQCDRAD [32]. This scheme has the advantage of reducing the sensitivity of the DIS cross sections to higher order corrections, and improving the theoretical precision of the mass definition. In QCDNUM [27], the calculation of the heavy quark contributions to DIS structure functions are available at NLO and only electromagnetic exchange contributions are taken into account.

In the zero-mass variable flavour number scheme (ZM-VFNS) heavy quark densities are included in the proton for $Q^2 \gg m_H^2$ but they are treated as massless in both the initial and final states. This scheme is accurate in the region where Q^2 is much greater than m_H^2 but becomes unreliable for Q^2 $\sim$m$_H^2$.

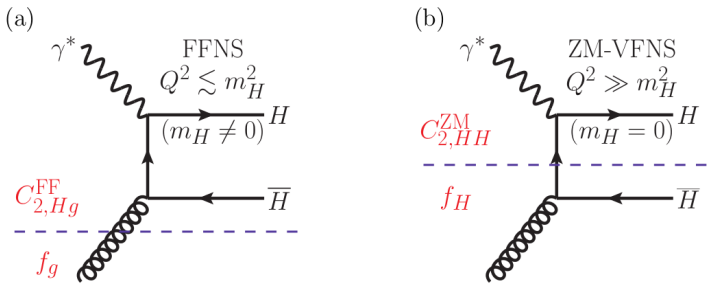


Figure 3.2: Diagrams contributing at leading-order to the heavy flavour structure function, in (a) the fixed flavour number scheme (FFNS) valid for $Q^2 \lesssim m_H^2$, and (b) the zero-mass variable flavour number scheme (ZM-VFNS) valid for $Q^2 \gg m_H^2$.

3.2.2 General mass schemes

Analyses of QCD measurements are often performed by making a choice between using a suitable FFN scheme at scales in the region of heavy quark mass thresholds or a ZM scheme at high scales where the associated powers of m_H^2/Q^2 can be safely neglected. In either case the treatment of heavy quarks is at least unambiguous, with the ZM approach yielding a simpler procedure as there is no requirement to calculate coefficient functions with the heavy quark masses intact.

For analyses of a large data set, potentially spanning several heavy quark

thresholds and extending to very high scales, the desire to improve the perturbative reliability of the calculations has led to the development of a number of hybrid or general mass schemes. In such schemes the treatments generally reduce to the FFN regime at low scales and the ZM treatment at high scales, with the intermediate regime handled via some interpolation between the two. In other word, the General-Mass Variable Flavour Number scheme (GM-VFNS) combines the advantage of the massive and massless calculations by means of an interpolated scheme which is valid for any value of the scale Q, and that matches the FFN and ZM-VFN schemes at small and large values of Q, respectively.

3.3 General features of parton distributions

The number of independent PDFs to be determined is dependent upon the choice of initial scale, as quark distributions that can be considered heavy with respect to Q_0^2 may be generated perturbatively through the DGLAP procedure outlined previously. The typical choice is to determine the parton distributions at some scale $m_s^2 < Q_0^2 \leq m_c^2$ such that the flavours c, b, t are produced by evolution. These scale choices minimize the number of distributions to be determined while remaining perturbatively reliable.

As the remaining seven distributions,the gluon, the u, d, s quarks and their antiquarks,are fundamentally a parametrization of the nonperturbative dynamics of the proton, they are by definition out of reach of a perturbative analysis. There are however some general statements that may be made of their x-dependence that are independent of the hard scale. The most important of which are the parton distribution sum rules which constrain the relative normalization of PDFs.

Firstly, the momentum sum rule (MSR) ensures that the parton distributions fractional momenta sum to the momentum of the parent proton

$$\int_0^1 dx \left[dx\Sigma(x, Q^2) + xg(x, Q^2 \right] = 1 \tag{3.40}$$

where Σ is the singlet distribution. Following this are the quark valence sum rules. These fix the quark distributions such that the resulting proton has

the appropriate quantum numbers,

$$\text{up-valence:} \int_0^1 dx(f_u(x,Q^2) - f_{\bar{u}}(x,Q^2)) = 2 \tag{3.41}$$

$$\text{down-valence:} \int_0^1 dx(f_d(x,Q^2) - f_{\bar{d}}(x,Q^2)) = 1 \tag{3.42}$$

$$\text{strange-valence:} \int_0^1 dx(f_s(x,Q^2) - f_{\bar{s}}(x,Q^2)) = 0\,. \tag{3.43}$$

From these rules we may infer additional constraints upon individual PDFs.

The MSR suggests a form for the large-x behaviour of the distributions, in that they should parametrically tend to zero as $x \to 1$. The number sum rules in (3.41), (3.42) and (3.43) require the valence-type distributions to be integrable over the whole x-range. While there is no requirement for the singlet and gluon distributions to be integrable, their first moments must be, as required by the MSR. Combining these three constraints we may parametrize the large and small-x behaviour of both valence-like and gluon or singlet-like distributions as:

$$f_V(x,Q_0^2) = N_V x^{\alpha_V}(1-x)^{\beta_V} r_V(x)\,, \tag{3.44}$$

$$f_\Sigma(x,Q_0^2) = N_\Sigma x^{\alpha_\Sigma}(1-x)^{\beta_\Sigma} r_\Sigma(x)\,. \tag{3.45}$$

In these expressions, the parameters α and β control the small and large-x PDF behaviour respectively. The β should be such that the PDFs tend to zero smoothly at large-x, and the α such that the valence distributions are integrable, and the first moment of the gluon and singlet are integrable. The overall PDF normalisations N being constrained via the appropriate sum rules.

Finally, what remains in the determination of the distributions are the remainder terms $r(x)$ which describe the PDFs between the two x-limits. Their determination is considerably more complex and is a ongoing source of research.

3.4 PDF determination

Understanding the functional structure of parton distributions is a complex task that has been subject to a number of approaches over the years. As nonperturbative quantities describing the behaviour of QCD bound states,

in principle they may be subject to analysis using Lattice QCD methods. While a great deal of effort and progress has been made in understanding PDFs through nonperturbative methods [34, 35], results remain short of providing distributions for practical application at hadron colliders.

The majority of PDF analyses are therefore performed analogously to the determination of many other QCD parameters; via a fit to appropriate experimental data. The fundamental difficulty in PDF fits being that they are determinations of functions rather than single parameters and therefore one must attempt to find some optimum solution in an (in principle) infinite-dimensional functional parameter space. This is of course complicated by having only a finite set of experimental data points upon which to perform a fit. Moreover as the applications involving PDFs have become more precise, a detailed understanding of the uncertainties in the determination of PDFs has become vital. The problem of PDF fitting is therefore one of finding a reliable estimator for a probability distribution in a space of functions.

The complexity of the task, along with the inherent ambiguities in the QCD treatment of data, led to the emergence of several competing methodologies and determinations. Today there are a diverse array of fitting groups producing sets of parton distribution functions, the most important of which being the ABM [36, 37], CTEQ/CT [38, 39], HERAPDF [40, 41], NNPDF [42, 43], MSTW [92] and MMHT [93]. All of these sets are publicly available though the standard LHAPDF interface [57].

Typically PDF sets are provided for a variety of theory input parameters such as perturbative order, and value of the strong coupling. All modern PDF sets now include a quantitative assessment of their associated uncertainties. In this chapter we shall review the method used in modern PDF determination.

3.4.1 Parametrization

A set of PDFs is a set of functions, one for each parton entering the factorized expressions, Because PDFs at different scales are related by the evolution equations, the goal is to determine a set of functions for $0 < x < 1$ at some reference scale Q_0^2.

There are in principle thirteen independent PDFs in a given hadron (six quarks and antiquarks and the gluon); however, in practice, charm and heavier quark PDFs in the nucleon are not independently determined in

all current PDF sets, and are instead assumed only to be generated by QCD radiation. The (moderate) impact of introducing an independent (non-perturbative) charm PDF, so that charm does not vanish below the threshold for its radiation. In practice, in many cases, it turns out to be convenient to express the six light quark PDFs as suitable linear combinations, like the singlet combination of Eq. (3.34).

Once a suitable set of basis PDFs has been chosen, all existing PDF determinations are based on choosing a parametrization of PDFs at the reference scale. A standard choice, adopted by most PDF fitting groups, is to assume that:

$$f_i(x, Q_0^2) = x^{\alpha_i}(1-x)^{\beta_i} g_i(x) \tag{3.46}$$

where $g_i(x)$ tends to a constant for both $x \to 0$ and $x \to 1$. This choice is motivated by the expectation that PDFs behave as a power of x as $x \to 0$ due to Regge theory [45], and as a power of $(1-x)$ as $x \to 1$ due to quark counting rules (see, e.g., Ref. [46] and references therein).

Specific choices for the function $g_i(x)$ differ between groups. Common choices are x a polynomial or the exponential of a polynomial in x or x, with more parameters used to describe PDFs for which more information is available (such as the gluon) in comparison to those (such as the strange PDF) that are poorly constrained by the data. Typical contemporary PDF sets based on this choice of functional form are parametrized by about $20-30$ parameters. For example, MSTW2008 [47] uses the following basis for their determination:

$$\begin{aligned} g, \\ q_v \equiv q - \bar{q}, \\ \Delta \equiv \bar{d} - \bar{u}, \\ S \equiv 2(\bar{u} + \bar{d}) + s + \bar{s}, \\ s^{\pm} \equiv s \pm \bar{s}, \end{aligned} \tag{3.47}$$

where g is the gluon PDF and the q_v correspond to the u, d quark valence PDFs. These fully parameterize the degrees of freedom to be determined. A functional form in x is then chosen for each of the distributions (the value of Q^2 is kept fixed at the input scale for fitting). While all groups include the limiting-x description of Eq. (3.45), the choice of parametrization for the remainder function r varies substantially between fitting groups. As an example, the valence quark PDF q_v parametrization in MSTW2008 is provided by the expression:

$$x q_v(x, Q_0^2) = a x^b (1-x)^c (1 + d\sqrt{x} + ex), \tag{3.48}$$

and the equivalent parametrization in CT10 [48] is

$$xq_v(x, Q_0^2) = ax^b(1-x)^c \exp(cx + dx^2 + e\sqrt{x}), \tag{3.49}$$

where the $(a, \cdots, e)$ are the parameters to be determined in the fit. In total the MSTW08 basis has 30 free parameters (taking into account sum rule constraints), the CT10 parametrization is a little less flexible, having 26 free parameters. The problem is now reduced to finding the optimum parameters for the 7 PDFs that minimize some measure of fit quality.

3.4.2 Error propagation from experimental data points

Uncertainties in global PDF analyses can be divided into two general categories, which may loosely be called "theoretical" and "experimental". The theoretical errors include those associated with the choice of the form of the input parametrization, the neglected higher-order QCD and electroweak corrections, parton recombination and other higher-twist corrections, the choice of data sets and kinematic cuts, the choice of nuclear corrections for the neutrino-initiated data and the treatment of heavy flavours. These uncertainties are often difficult to quantify a priori until a better calculation or prescription becomes available.

On the other hand, in principle there are well-defined procedures for propagating experimental uncertainties on the fitted data points through to the PDF uncertainties. Three main methods have been used to do this:

1. The Lagrange multiplier method [49, 50], which does not rely on linear error propagation, but requires the ability to perform a global fit.

2. The Hessian method [51], which is based on linear error propagation and involves the production of eigenvector PDF sets suitable for convenient use by the end user.

3. The use of Monte Carlo sampling [52, 53], which has recently been used in conjunction with neural networks to determine NLO PDFs from a DIS-only fit [54].

The first two methods were originally used by CTEQ [38, 39] and then by MRST [55]. In principle, the Lagrange multiplier method is superior to the Hessian approach, but it suffers from the enormous practical disadvantage

that a series of new global fits has to be done every time one considers a new quantity. Fortunately, it turns out that for those quantities that have been considered by both methods, the uncertainties have been found to be comparable.

3.5 PDF analyses: state of the art

3.5.1 Various fitting groups

Various fitting groups currently produce general-purpose sets of PDFs of the nucleon, with most of the groups having a long history which goes back at least a couple of decades. Six of these groups have been providing regular updates of their PDFs. All of these sets are publicly available though the standard LHAPDF interface [57], though CT10 NNLO and HERAPDF1.5 have not been presented in a journal publication.

	MSTW08	CT10	NNPDF2.3	HERAPDF1.5	ABM11	JR09
HERA DIS	✔	✔	✔	✔	✔	✔
Fixed-target DIS	✔	✔	✔	✗	✔	✔
Fixed-target DY	✔	✔	✔	✗	✔	✔
Tevatron W+Z+jets	✔	✔	✔	✗	✗	✗
LHC W+Z+jets	✗	✗	✔	✗	✗	✗

Figure 3.3: Data included in various NNLO PDF sets. [56]

	MSTW08	CT10	NNPDF2.3	HERAPDF1.5	ABM11	JR09
No. of PDFs	7	6	7	5	6	5
Statistics	Hess.+DT	Hess.+DT	MC	Hess.+Model+Parm.	Hess.	Hess.+T
PDF parms.	20+8	25	259	14	24	12
Heavy quarks	VFN TR	VFN ACOT	VFN FONLL	VFN TR	FFN	FFN

Figure 3.4: Main features of various NNLO PDF sets (see text for details) [56].

The main feature which distinguishes PDF sets is the data on which they are based they are summarized in 3.3. The main choices which underlie the PDF sets are summarized 3.4. Only three groups (MSTW08 [47], CT10 [48], and NNPDF2.3 [58]) make a fully global fit, defined here to be a fit including HERA and fixed-target DIS data, fixed-target Drell-Yan production, and Tevatron data on W, Z and jet production. The NLO version of the JR09 fit, GJR08 [59], does include some Tevatron jet data. The NNPDF2.3 set

is the only one to include LHC data (only in this comparison). Concerning HERA data, note that CT10 and NNPDF2.3 include the combined HERA I inclusive data [61], MSTW08 and JR09 instead include the older separate data from H1 and ZEUS, ABM11 includes combined HERA I data but only with the cut $Q^2 < 1000 GeV^2$, and HERAPDF1.5 additionally includes the preliminary combined HERA II inclusive data [60]. The kinematical coverage of the NNPDF2.3 data set is shown in Figure 1, with the x and Q^2 values shown determined using leading-order parton kinematics.

3.5.2 CT14nnlo

The CT10 parton distribution functions were published at next-to-leading order (NLO) in 2010 [62], followed by the CT10 next-to-next-to leading order (NNLO) parton distribution functions in 2013 [48]. These PDF ensembles were determined using diverse experimental data from fixed-target experiments, HERA and the Tevatron collider, but without data from the LHC. The new CT14 global analysis [63] includes data from the LHC for the first time, as well as updated data from the Tevatron and from HERA experiments. Various CT14 PDF sets have been produced at the leading order (LO), NLO and NNLO and are available from LHAPDF [57].

Features of the CT14 analysis. The CT14 PDFs are determined from data on inclusive high momentum transfer processes, for which perturbative QCD is expected to be reliable. For example, in the case of deep inelastic lepton scattering, only data with $Q > 2$ GeV and $W > 3.5$ GeV are used.

Data (see Fig. 3.5) in this region are expected to be relatively free of non-perturbative effects, such as higher twists or nuclear corrections.

The new LHC measurements [64, 65] of W/Z cross sections directly probe flavor separation of u and d (anti-)quarks in an x-range around 0.01 that was not directly assessed by the previously available experiments. The updated measurements of electron charge asymmetry from the DØ collaboration [66] included in the CT14 analysis probe the d quark PDF at $x > 0.1$. To better estimate variations in relevant PDF combinations, such as $d(x,Q)/u(x,Q)$ and $\bar{u}(x,Q)/\bar{d}(x,Q)$, the number of free PDF parameters is increased to 28, compared to 25 in CT10 NNLO.

As another important modification, CT14 employs a novel flexible parametrization for the PDFs, based on the use of Bernstein polynomials (reviewed in the Appendix of Ref. [63]). The shape of the Bernstein poly-

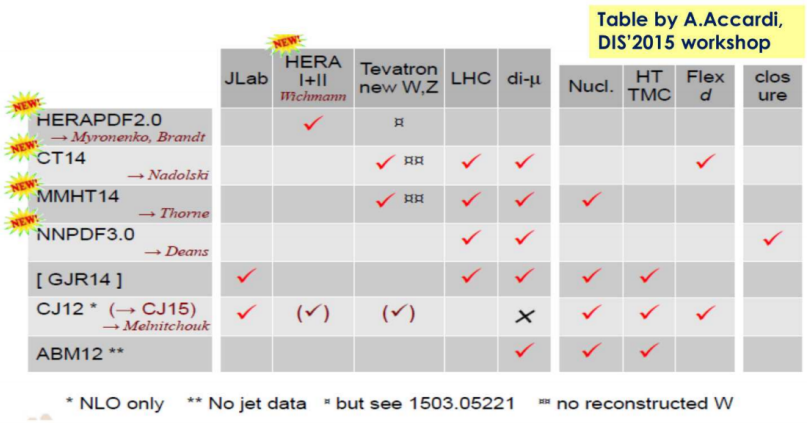


Figure 3.5: CT14 NNLO data sets. [67]

nomials is such that a single polynomial is dominant in each given x range, reducing undesirable correlations among the PDF parameters that sometimes occurred in CT10. In the asymptotic limits of $x \to 0$ or x$\to$1, the new parametrization forms allow for the possibility of arbitrary constant ratios of d/u or $\bar{d}/\bar{u}$, in contrast to the more constrained behavior assumed in CT10. The CT14 PDF error sets is obtained using two techniques, the Hessian method [51] and Monte Carlo sampling [52, 53]. Lagrange multiplier studies [49, 50] have also been used to verify the Hessian uncertainties, especially in regions not well constrained by data. This applies at NNLO and NLO; no error sets are provided at LO due to the difficulty of defining meaningful uncertainties at that order. A central value of $\alpha_s(M_Z^2)$ of 0.118 has been assumed in the global fits at NLO and NNLO, but PDF sets at alternative values of $\alpha_s(MZ^2)$ are also provided. CT14 prefers $\alpha_s(M_Z^2) = 0.115^{+0.006}_{-0.004}$ at NNLO (0.117 ± 0.005 at NLO) at 90% confidence level (C.L.). These uncertainties from the global QCD fits are larger than those of the data from LEP and other experiments included into the world average. Thus, the central PDF sets are obtained using the value of 0.118, which is consistent with the world average value and was recommended by the PDF4LHC group.

In Fig. 3.6 one observes that the CT14 NNLO PDFs have a softer strange-quark distribution at low x and a somewhat softer gluon at high x d-quark has increased by 5% at $x \simeq 0.05$, after ATLAS and CMS W/Z production data sets at 7 TeV were included. At $x \gtrsim 0.1$, the update of the D$\emptyset$ charge asymmetry data set in the electron channel has reduced the magnitude of the d quark PDFs by a large amount, and has moderately increased the $u(x, Q)$ distribution.

3.6 The future of PDFs determination

Parton distributions have become increasingly relevant with the advent of the LHC. After the landmark discovery of a Higgs-like boson in July 2012, focus now shifts to characterizing the properties of this new particle, as well as increasingly difficult searches for indications of other new physics. As a consequence, demand will grow to reduce the unavoidable uncertainties associated with the PDFs in calculations of both signal and background. On the one hand, it will be necessary to bring under complete control the uncertainties in the region of electroweak symmetry breaking. In addition, searches for new physics will involve heavy final states, and thus, they will involve knowledge of PDFs in the large $x \lesssim 0.5$ region where they are currently very poorly known. To achieve these goals, it will be necessary to construct PDFs which satisfy a number of criteria, that not so long ago characterized an ideal PDF determination, but in the LHC era have become necessary requirements, namely, in decreasing order of importance:

1. The range and precision of **data sets** must be as wide as possible, cover currently unexplored kinematic regions, and include new LHC processes, which will gradually remove current discrepancies between PDF determinations.

2. The **parametrization** (see Section 3.4.1) should be sufficiently general and demonstrably unbiased, either by using a sufficiently large number of parameters, or by careful a posteriori checks of parametrization independence.

3. Computations should be performed at the highest available perturbative order, and in particular, at the order which is subsequently to be used in the computation of partonic cross sections. This is currently NNLO, but the need for the inclusion of various kinds of all-orders resummation is becoming increasingly important.

4. The treatment of heavy quarks (see Section 3.2) will have to include mass-suppressed terms in the coefficient functions, while also resumming logarithmically enhanced terms via the evolution equations. Such a treatment, like the schemes discussed in Section 3.2, is a minimum requirement: this is currently standard for DIS, but applications to hadronic observables are so far limited. Also, the dependence of results on the choice of value for the heavy-quark masses will have to be studied more systematically, with PDF sets made available for several values of the heavy-quark masses.

3.7 Hessian profiling technique

The impact of a new data set on a given PDF set can be performed by using a QCD global fit analysis using the experimental data. As an alternative approach, an approximate method can be used instead of a complete QCD fit. The profiling technique is the approximate method that can be applied for PDFs extracted by Hessian method [68].

There are generally two techniques that they can be applied to find the impact of new experimental data on a preexisting PDF. Bayesian Monte Carlo reweighting [69, 70] and Hessian profiling techniques [71]. It should be noted that these approximate methods have a number of limitations. For example, if the impact of new measurements is very large, these methods can not be useful and in particular are not able to explain the effect on the input PDF parametrization, or in the theoretical calculations. Therefore, not only when using these approximate methods some care should be taken but also we should care when interpreting their results.

The Hessian profiling technique is based on the χ^2 minimization method using a comparison between the theoretical predictions extracted with a given input Hessian PDF set and the new experimental data. According to this method, the χ^2 definition with taking into account the uncertainties of experimental data and the effects from the variations of PDF which is encoded by the Hessian eigenvectors, is as following [72, 70]:

$$\chi^2(\beta_{\rm exp},\beta_{\rm th}) = \sum_{i=1}^{N_{\rm data}} \frac{\left(\left[\sigma_i^{\rm exp} + \sum_j \Gamma_{ij}^{exp}\beta_{j,\rm exp}\right] - \left[\sigma_i^{\rm th} + \sum_k \Gamma_{ik}^{\rm th}\beta_{k,\rm th}\right]\right)^2}{\delta_i^2} + \sum_i \beta_{j,\rm exp}^2 + \sum_k \beta_{k,\rm th}^2\,, \tag{3.50}$$

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

After minimizing the χ^2 in Eq. (3.50), the corresponding values of the theoretical $\beta_{k,\rm th}^{\rm min}$ parameters can be interpreted as leading to optimized PDFs

("profiled") to explain the new specific measurement. In the next sections it will be seen how profiling method modifies both central values and total PDF uncertainties.

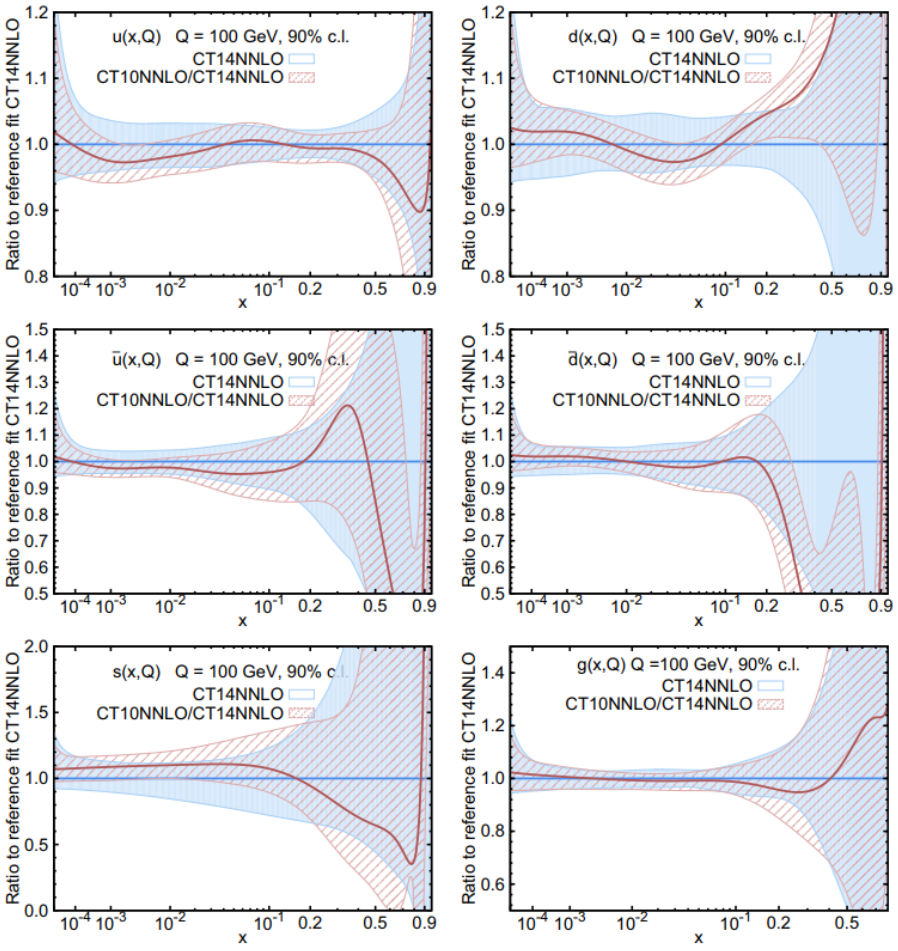


Figure 3.6: Comparison of 90% C.L. PDF uncertainties from CT14 NNLO (solid blue) and CT10 NNLO (red dashed) error sets. Both error bands are normalized to the respective central CT14 NNLO PDFs. [67]

Chapter 4

xFitter and ePump

In this chapter we shall explain in more detail the two packages, that are used throughout this analysis. The first section will be reserved to xFitter followed by ePump. We will not show any comparison with these two packages in this chapter. This will be a subject for another section.

4.1 xFitter

The xFitter package [68] aims at providing a framework for QCD analyses related to proton structure in the context of multi-processes and multi-experiments. The framework includes modules or interfaces enabling a large number of theoretical and methodological options, as well as a large number of relevant data sets from HERA, Tevatron and LHC.

A schematic structure of the xFitter is illustrated in Fig. 4.1 which encapsulates all the current functionality of the platform.

PDFs are the essential components that allow us to make theoretical predictions for experimental measurements of protons and hadrons. The precision of the PDF analysis has advanced tremendously in recent years, and these studies are now performed with very high precision at NLO and NNLO in perturbation theory. In the following we will explore the most important features of the xFitter package:

1. The xFitter project grew out of PDF efforts of H1 and ZEUS which became the HERAFitter project in 2012, and renamed the xFitter project in 2015. xFitter is continually being updated, and version **2.0.1 (Old-Fashioned)** is released in 2019 with many improvements and new features.

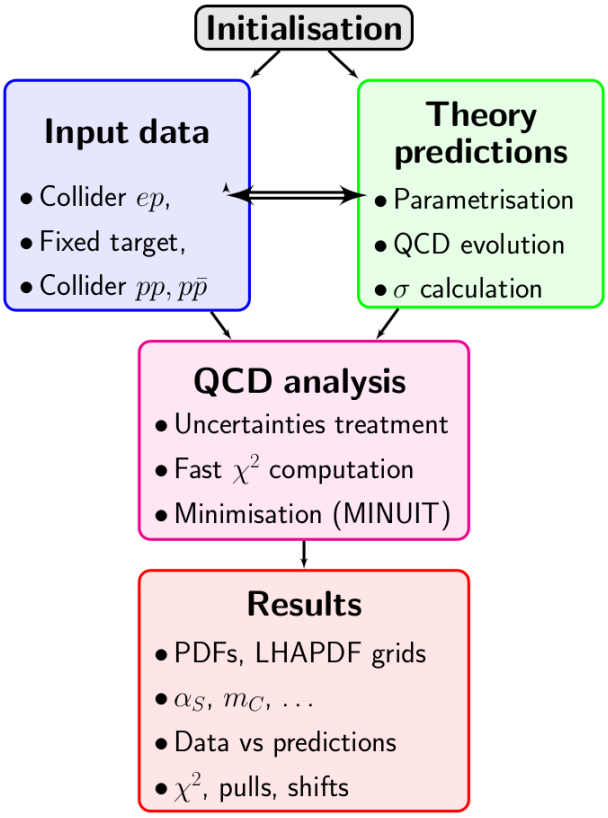


Figure 4.1: Schematic structure of the xFitter program.

2. The xFitter can perform PDF fits, assess the impact of new data (profiling), compare existing PDF sets, and perform a variety of other tasks.

3. The xFitter framework has already been used for more than 40 analyses including many LHC studies [68].

4. The framework of xFitter is modular to allow for various theoretical and methodological options, and contains interfaces to QCDNUM [27], APFEL [26], LHAPDF [57], APPLGRID [82], APFELGRID [85], FastNLO [86], HATHOR [87], among other packages.

5. xFitter is able to read and write PDFs in the LHAPDF6 format, Fig. 4.2.

6. xFitter can also generate comparison plots of data vs. theory, and an example is shown in Fig. 4.3.

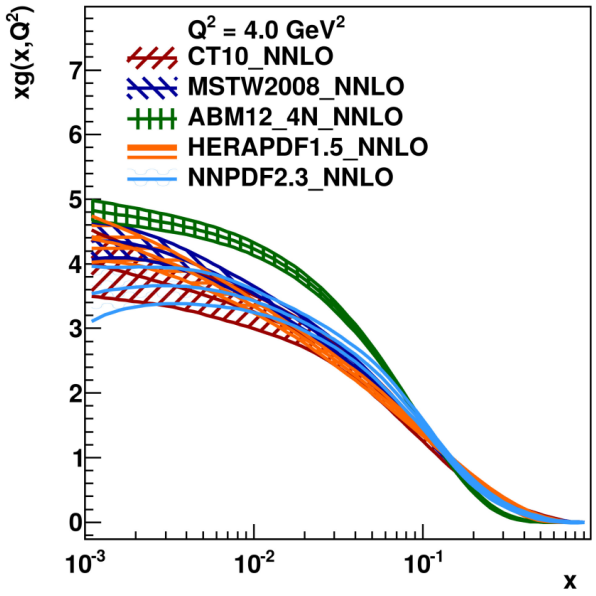


Figure 4.2: Gluon PDFs with uncertainties at $Q^2 = 4.0$ GeV^2. [89]

7. There are a variety of options for the definition of the χ^2 function and the treatment of experimental uncertainties.

8. An important application of xFitter is to understand how a particular data set or experiment will impact the PDFs, which is the main objective of this paper. More examples and details are given in the next sections.

   Furthermore, xFitter is able to perform PDF profiling and reweighting studies. The reweighting method allows xFitter to update the probability distribution of a PDF uncertainty set (such as a set of NNPDF replicas) to reflect the influence of new data inputs. For the PDF profiling (see Section 3.7 for more details), xFitter compares data and MC predictions based on the χ^2-minimization, and then constrains the individual PDF eigenvector sets taking into account the data uncertainties.

9. xFitter can also be used for certain heavy ion analyses.

10. xFitter has the ability to handle both pole masses and $\bar{M}S$ running masses in the FONNL scheme.

The xFitter program is a versatile, flexible, modular, and comprehensive

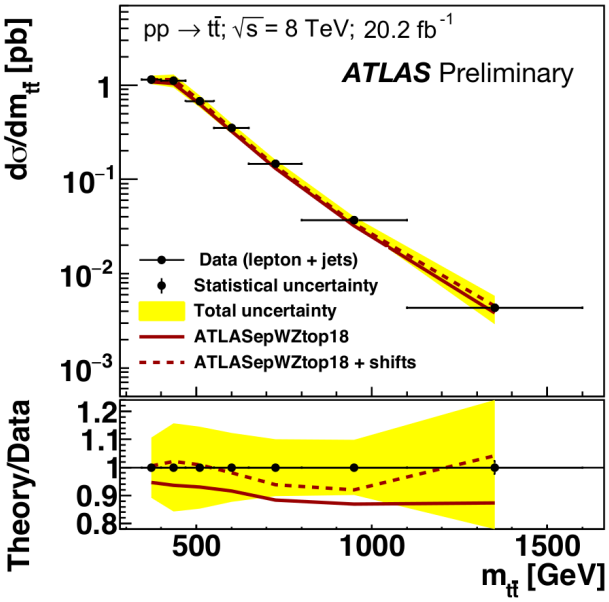


Figure 4.3: Normalized $t\bar{t}$ differential cross-sections as a function of the invariant mass ($m_{t\bar{t}}$). [88]

tool that can facilitate analyses of the experimental data and theoretical calculations.

4.2 ePump

The ePump package [14] can be used to update or optimize a set of PDFs, including the best-fit PDF set and Hessian eigenvector pairs of PDF sets (i.e., error PDFs), and to update any other set of observables. The ePump code is written in C++ and consists of two main executables, UpdatePDFs and OptimizePDFs.

A schematic structure of the ePump is illustrated in Fig. 4.4 which encapsulates all the current functionality of the platform.

In this report, we will not optimize any set of PDFs but instead we will update a set of PDFs, in other word, we shall demonstrate how to use ePump to update CT14 PDFs when new experimental data are included in the global fit. The main feature of ePump are listed below:

1. ePump may be used to update any PDF set containing Hessian error

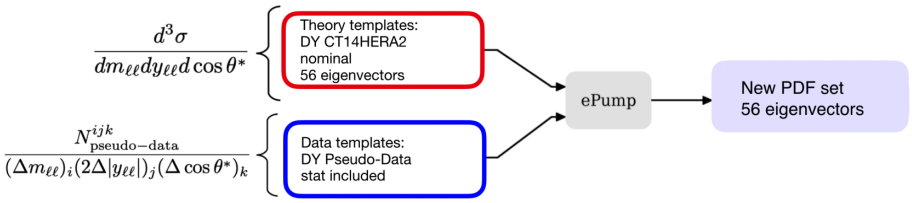


Figure 4.4: The ePump package requires two inputs to generate an updated PDF set: an existing Theory template of a PDF set (parameters + uncertainties) and binned Data template of (pseudo-) data, including statistical uncertainties from integrated luminosity assumptions. [90]

PDFs in LHAPDF format (".dat" files) or CTEQ format (".pds" files), and the updated central and error PDFs are output in the same format as the input files (i.e., either ".dat" or ".pds" files). Fig. 4.5.

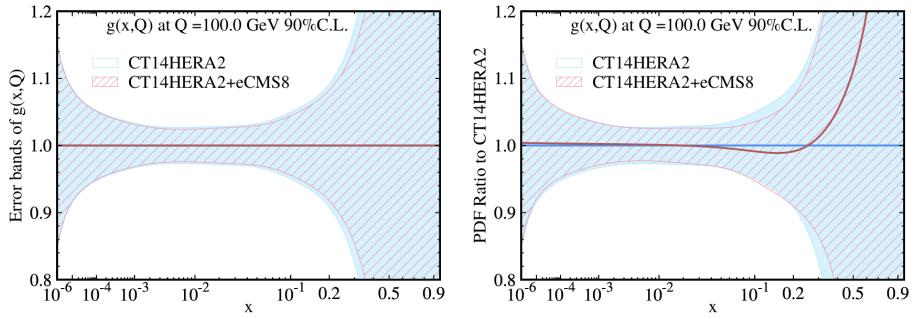


Figure 4.5: Comparison between the gluon-PDF at $Q = 100$ GeV for CT14HERA2 NNLO and for CT14HERA2 updated by ePump using the CMS 8 TeV double-differential inclusive jet cross section measurements as a function of jet rapidity (y) and transverse momentum (q_T). The left curves and error bands are normalized to their respective central fits, while the right curves and error bands are each normalized to the CT14HERA2NNLO central PDF. [14]

2. It is flexible enough to accommodate different non-global tolerance criteria.

3. Either a global tolerance value may be used, or dynamical tolerances may be used, in which case a file of tolerance-squared values, defined

as $(T_i^{\pm})^2$ for each $\pm$ error PDF, in a ".tol" file must be included with the PDF set.

4. For each data set of observables to be used to update The PDFs, one must supply a ".data" file and a ".theory" file.

   - The ".data" file contains the experimental values and errors for each of the data points in a particular data set.
   - The errors can be included in several different formats, including a table of uncorrelated statistical and systematic errors and correlated systematic errors.
   - The ".theory" file contains a list of the theoretical predictions for the observables in the data set for each of the PDFs in the original best-fit and Hessian error PDF set.

5. The code takes only a few seconds to run on an early-2013 MacBook Pro, so it is quick and easy to try different combinations of data sets to compare their impact on the PDFs.

The complete ePump package can be found here: ePump link.

Chapter 5

Data sets and theoretical framework

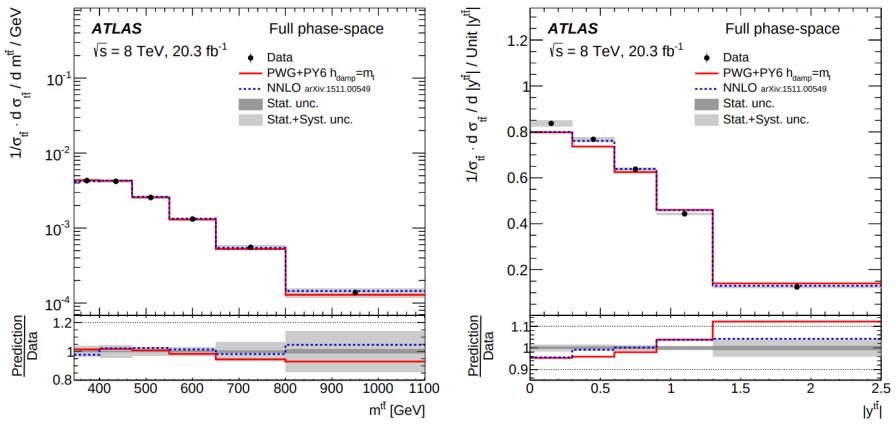
5.1 Data sets

The profiling is preformed using two different classes of data samples measured by ATLAS: $t\bar{t}$ production in the lepton+jets channel; $t\bar{t}$ production in the dilepton channel and the triple-differential cross section of the Drell-Yan process. The details of these samples are given below.

The $t\bar{t}$ data have been measured at 8 TeV using 20.2 fb^{-1} of data in the lepton+jets [73] and dilepton [74] decay modes.

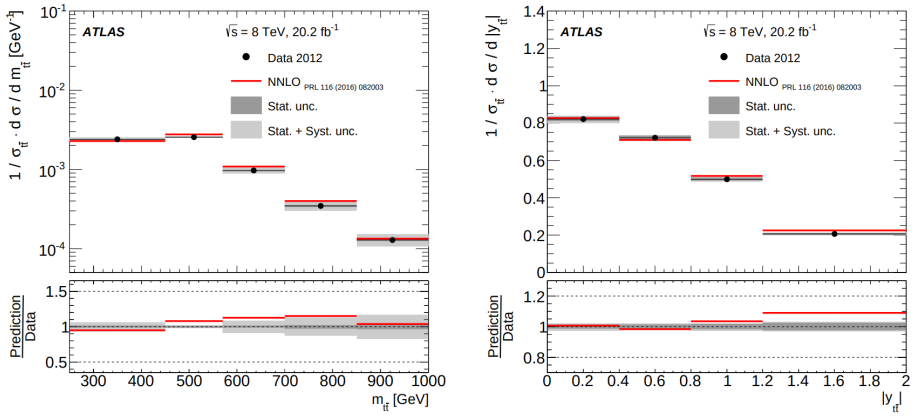
In the lepton+jets channel the differential spectra that are sensitive to the gluon PDF are: the mass of the $t\bar{t}$ pair, $m_{t\bar{t}}$ (Fig. 5.2), the rapidity of the $t\bar{t}$ pair, $y_{t\bar{t}}$, the average top-quark rapidity, y_t and the average top-quark transverse momentum, p_{T}^{t} (Fig. 5.2). Each of these spectra has full information on systematic bin-to-bin correlations. **There are 55 such sources of systematic uncertainty and these are also correlated between the different spectra**.

In the dilepton channel, the spectra for the mass of the $t\bar{t}$ pair, $m_{t\bar{t}}$, and the rapidity of the $t\bar{t}$ pair, $y_{t\bar{t}}$, are used. Note that although the transverse momentum distributions of the $t\bar{t}$ pair, $p_{\mathrm{T}}^{t\bar{t}}$, are available for both channels these are not used since the predictions are effectively NLO, rather than NNLO. The $t\bar{t}$ data are sensitive to the gluon PDF for $x > 0.01$. For the dilepton data the correlations are provided as a total covariance matrix for each spectrum separately.



(a) (b)

Figure 5.1: Full phase-space normalized differential cross-section as a function of the (a) invariant mass ($m_{t\bar{t}}$) and (b) absolute value of the rapidity ($y_{t\bar{t}}$) of the $t\bar{t}$ system compared to NNLO theoretical calculations. [73]



(a) $m_{t\bar{t}}$ (b) $|y_{t\bar{t}}|$

Figure 5.2: Normalized $t\bar{t}$ differential cross-sections as a function of the (a) invariant mass ($m_{t\bar{t}}$) and (b) absolute value of the rapidity ($|y_{t\bar{t}}|$) of the ($t\bar{t}$ system at $\sqrt{s} = 8$ TeV measured in the dilepton $e\mu$ channel compared with theoretical QCD calculations at full NNLO accuracy. [76]

The ATLAS precise measurement of the triple-differential cross section for the Drell-Yan process $Z/\gamma^* \rightarrow l^+l^-$ [75], where l is an electron or a muon,

comes from a sample of 20.2 fb^{-1} of pp collisions data at a centre-of-mass energy of $\sqrt{s} = 8$ TeV collected in 2012. The measurement is performed for invariant masses of the lepton pairs, m_{ll}, between 46 and 200 GeV.

The data are presented in bins of invariant mass, absolute dilepton rapidity, $|y_{ll}|$, and the angular variable $\cos\theta^*$ between the outgoing lepton and the incoming quark in the Collins-Soper frame. The measurements are performed in the range $|y_{ll}| < 2.4$ in the muon channel, and extended to $|y_{ll}| < 3.6$ in the electron channel. The cross sections are used to determine the Z boson forward-backward asymmetry as a function of $|y_{ll}|$ and m_{ll}. The measurements achieve high-precision, below the percent level in the pole region, excluding the uncertainty in the integrated luminosity, and are in agreement with predictions. These precision data are sensitive to the parton distribution functions and the effective weak mixing angle.

5.2 Theoretical framework

The NNLO predictions for top-quark pair production [77] are supplied in the form of FastNLO (interfaced to xFitter package) grids [78, 79] for the data in the lepton+jets channel. The predictions for $m_{t\bar{t}}$, $y_{t\bar{t}}$, and y_t are given for the renormalization and factorization scale equal to $H_{\mathrm{T}}/4$, where $H_{\mathrm{T}} = \sqrt{m_t^2 + (p_{\mathrm{T}}^t)^2} + \sqrt{m_t^2 + (p_{\mathrm{T}}^{\bar{t}})^2}$, and the predictions for p_{T}^t are given for scale equal to $m_{\mathrm{T}}/2$, where $m_{\mathrm{T}} = \sqrt{m_t^2 + (p_{\mathrm{T}}^t)^2}$ and $m_t = 173.3$ GeV.

For the dilepton channel APPLGRID is interfaced to MCFM to produce NLO grids and a K-factor defined as

$$\sigma_{\mathrm{NNLO}} = K \times \sigma_{\mathrm{NLO}} \tag{5.1}$$

is used to correct from NLO to NNLO predictions, using K-factors from Ref. [80]. The NNLO/NLO K-factors are $\sim 7\%$ above unity and flat for $y_{t\bar{t}}$, and vary between 6% and 12% above unity for $m_{t\bar{t}}$, as $m_{t\bar{t}}$ rises from 345 GeV to 1600 GeV. Electroweak corrections for the $t\bar{t}$ spectra are also considered using the additional K-factors given in reference [81]. For the 8 TeV data these corrections can be up to 1% for $y_{t\bar{t}}$ and y_t, up to 2% for $m_{t\bar{t}}$, and up to 4% for p_{T}^t.

The xFitter package uses the APPLGRID code [82] interfaced to the MCFM program [83, 84] for fast calculation of the differential Z/γ^* boson cross sections at NLO in QCD and LO in EW and a K-factor technique (5.1) is

used to correct from NLO to NNLO predictions in QCD and from LO to NLO predictions in EW. These K-factors are close to unity within $1 - 2\%$.

Chapter 6

Results and discussion

The strategy followed in this report is as follow:

In the first two sections we show the impact on triple-differential cross section for the Drell-Yan process and $t\bar{t}$ production in the lepton+jets channel and $t\bar{t}$ production in the dilepton channel from ATLAS on CT14nnlo PDFs sets (see Section 3.5.2). In addition, only xFitter framework is used and no comparison with ePump in shown.

The second part will be restricted to discussed some issues observed with xFitter and the fix we come up with. Finally, a comparison of xFitter and ePump is shown for the first time.

6.1 The Impact of ATLAS data on CT14nnlo

6.1.1 Set up

The CT14 parton distribution functions, in 90% confidence level are available in LHAPDF library [57] which is interfaced to xFitter.

To study the impact of triple-differential cross section for the Drell-Yan process and top quark cross section measurements on a given PDF set, the Hessian profiling method is used (see Section 3.7). This approximate method incorporates the information contained in new measurements into an existing specific PDF sets without the need for refitting.

Each data (ZD3 and top quark) is used to update the proton PDFs using

the profiling method, utilized by the Thorne-Roberts (TR) [91] GM-VFN scheme. The values of top quark mass, m_t, and strong coupling constant at Z boson mass, $\alpha_s(M_Z^2)$, are set to 173.3 GeV and 0.118, respectively.

6.1.2 The Impact of Z3D data

Using the profiling technique introduced in Section 3.7, the agreement between data and predictions can be quantitatively assessed.

Profiling PDFs, by introducing the data presented here, provides a shifted set of parton distributions with generally reduced uncertainties. The effect of the triple-differential cross section for the Drell-Yan process on the sea-quark distribution is examined (Figs. 6.1 and 6.2).

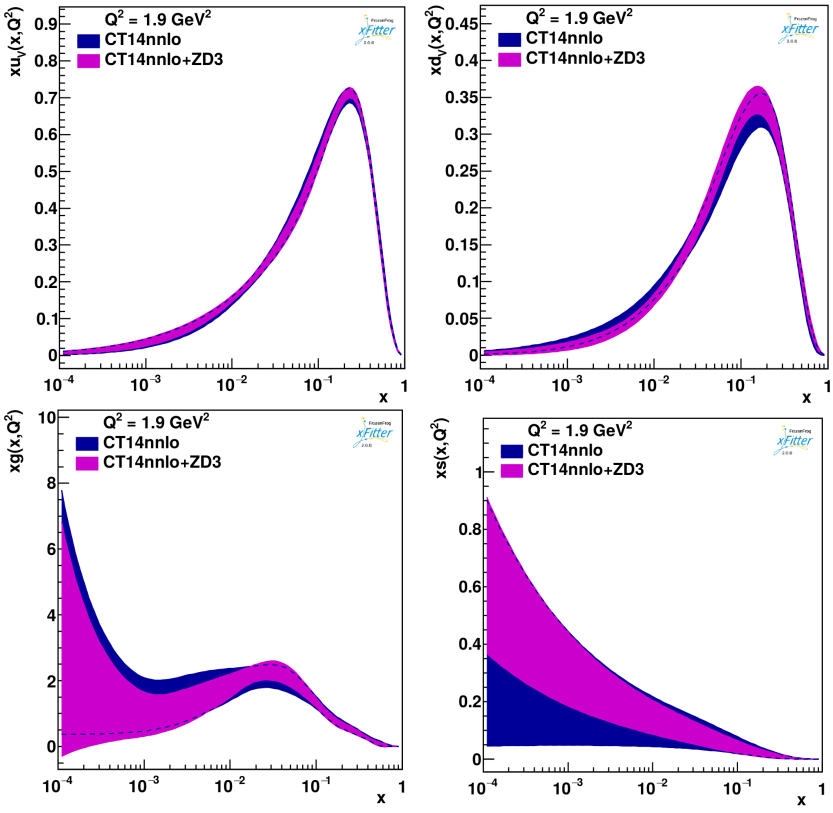


Figure 6.1: Distributions of xu, xd, xg and xs PDFs as a function of Bjorken-x at a scale of $Q^2 = 1.9$ GeV^2 for the CT14 PDF set before and after profiling with the new Z3D data.

The strange-quark distribution is significantly increased and the uncertainties are reduced. Also, Some reduction of the uncertainty is also observed for the valence-quark distributions, xu_v/xu_v and xd_v/xd_v, as is illustrated in Figs. 6.1 and 6.2.

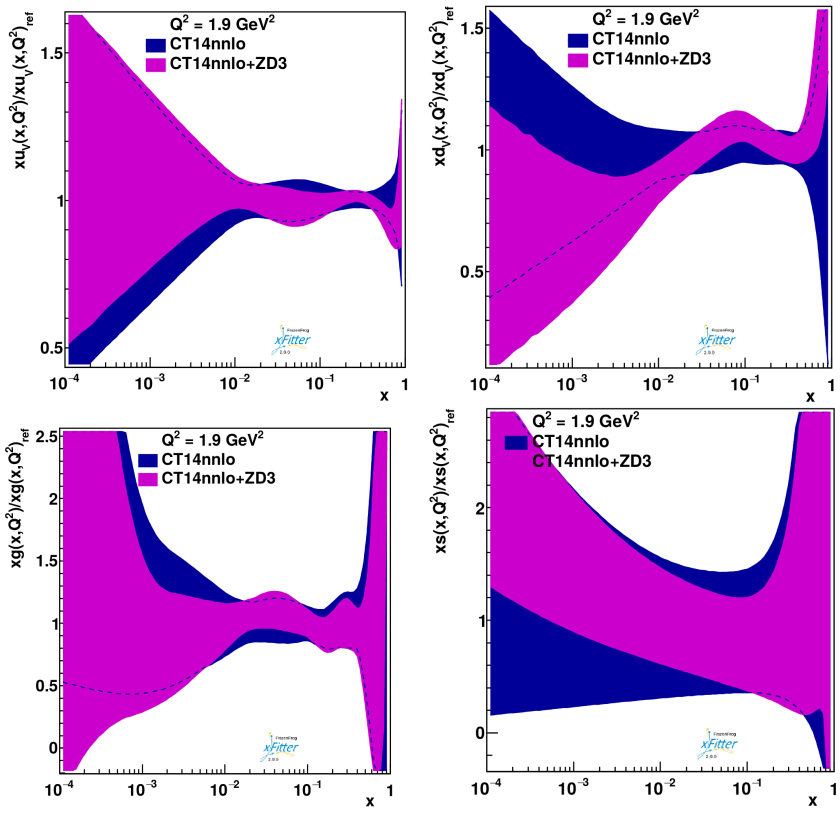


Figure 6.2: Distributions of xu_v/xu_v, xd_v/xd_v, xg/xg and xs/xs relative uncertainties as a function of Bjorken-x at a scale of $Q^2 = 1.9$ GeV^2 for the CT14 PDF set before and after profiling with the new Z3D data.

6.1.3 The impact of top data

The new top quark cross section data provide significant constraints on the central values and their uncertainties of xs and xg PDFs. The most significant impact of new measurements is observed only on the gluon PDF ratio.

Profile the $t\bar{t}$ lepton+jets spectra to CT14nnlo

The impact of lepton+jets spectra, $m_{t\bar{t}}$ and p_T^t, on CT14 is shown in this section in Figs 6.3 and A.1, where the CT14nnlo+Toplp label stands for the profiled CT14nnlo. The $m_{t\bar{t}}$ and p_{T}^t both lead to a harder gluon distribution and a reduced high-x uncertainty on the gluon PDF.

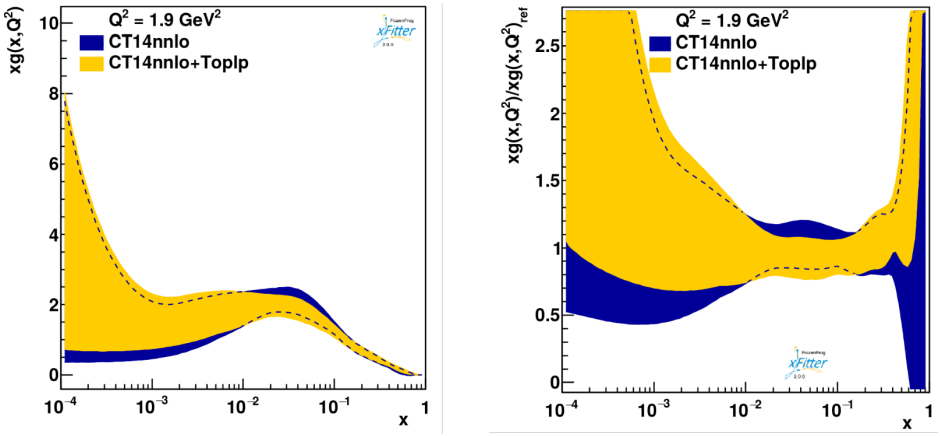


Figure 6.3: The gluon PDFs and relative uncertainties extracted from profiled CT14, PDF sets at 1.9 GeV^2 as a function of x. The results obtained after the profiling procedure are compared with corresponding same features before profiling.

The comparison between the original and profiled parton distribution of xu_v, xd_v, $x\bar{u}$, $x\bar{u}$, and $x\Sigma$ extracted from CT14 PDFs are presented in Fig. A.1.

Profile the $t\bar{t}$ dilepton spectra to CT14nnlo

The impact of dilepton $t\bar{t}$ spectra, $m_{t\bar{t}}$ and $y_{t\bar{t}}$, on CT14 is shown in this section in Figs. 6.4 and A.2, where the CT14nnlo+Topdilp label represents the profiled CT14nnlo.

For the lepton+jets spectra the $m_{t\bar{t}}$ and p_{T}^t data support a somewhat harder gluon, but in the dilepton case the effects are milder. Also, the impact of this data on xs is more relevant than lepton+jets data (Fig. A.2).

Profile the $t\bar{t}$ dilepton and lepton+jets spectra to CT14nnlo

In the present section the $m_{t\bar{t}}$ and p_{T}^t spectra from lepton+jets and the $y_{t\bar{t}}$ and $m_{t\bar{t}}$ spectrum from the dilepton data are profiled together, see Figs. 6.5

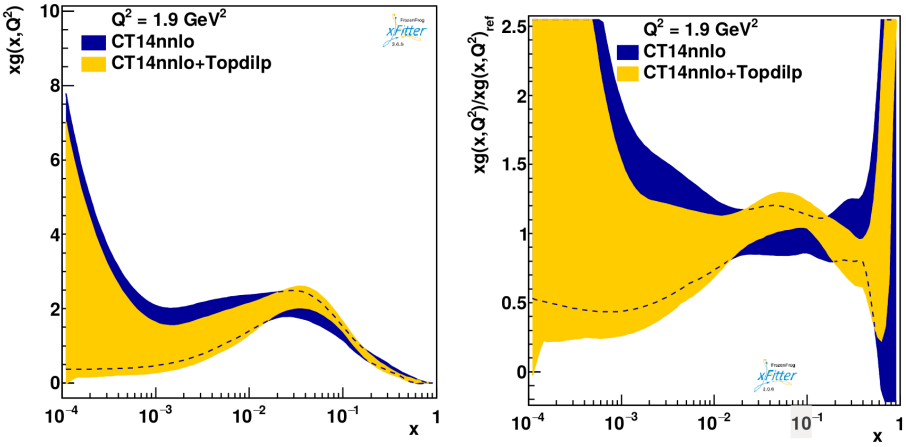


Figure 6.4: The gluon PDFs and relative uncertainties extracted from profiled CT14, PDF sets at 1.9 GeV^2 as a function of x. The results obtained after the profiling procedure compared with corresponding same features before profiling.

and A.3 where, the CT14nnlo+Toplp+dilp label corresponds to the profiled CT14nnlo.

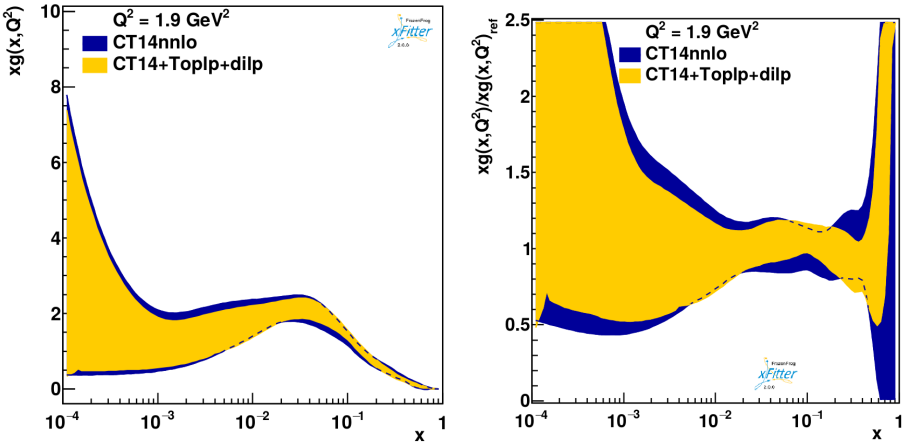


Figure 6.5: The gluon PDFs and relative uncertainties extracted from profiled CT14, PDF sets at 1.9 GeV^2 as a function of x. The results obtained after the profiling procedure e.g., adding the $m_{t\bar{t}}$ and $y_{t\bar{t}}$ spectra from di-lepton and the $m_{t\bar{t}}$ and p_{T}^{t} spectra from lepton+jets, compared with corresponding same features before profiling.

The dilepton $y_{t\bar{t}}$ data soften the gluon found from the lepton+jets spectra

and contributes to the reduction in uncertainty of the high-x gluon.

The profiling procedure using new set of top quark pair production data improves the strange xs and gluon distributions of CT14. The profiling affects the shape of the PDF, since we observed a small impact on $x\bar{u}$.

These findings are interesting and show the significance of the top quark production cross section data to constrain gluon and strange PDFs.

6.2 Issues with xFitter

When first running xFitter on $m_{t\bar{t}}$ lepton+jets cross section data, we encountered some technical issues that we present here.

6.2.1 Large total error and size of error bars

This issue with xFitter is observed when we try to compare the $m_{t\bar{t}}$ lepton+jets cross section we obtain with the result published in the ATLAS PUB Note [88] in Fig. 6.6.

Although, the ATLAS PUB Note performed a fit and we are using profiling, it will not affect the value of total error and the size of error bars, since we are using the same data files.[1]

Before going into the details, the current issues observed with xFitter will not affect the profiling (and the fits) but only the plotting.[2]

Two features are observed, on one hand, we found a larger total error band (yellow band in Fig. 6.6) compared to that in the PUB Note, and on the other hand, a large size of error bars.

Large total error

By default, xFitter is calculating the total error without symmetrization of sys_up and sys_down. Therefore, there is an over estimation of the total uncertainties. Also, the uncertainty in Lumi, is not added into the calculation of the total error. Hence, we have got a large total error shown

[1] This is confirmed after many discussions with Francesco Giuli.
[2] This is tested and it was confirmed by Sasha Glazov.

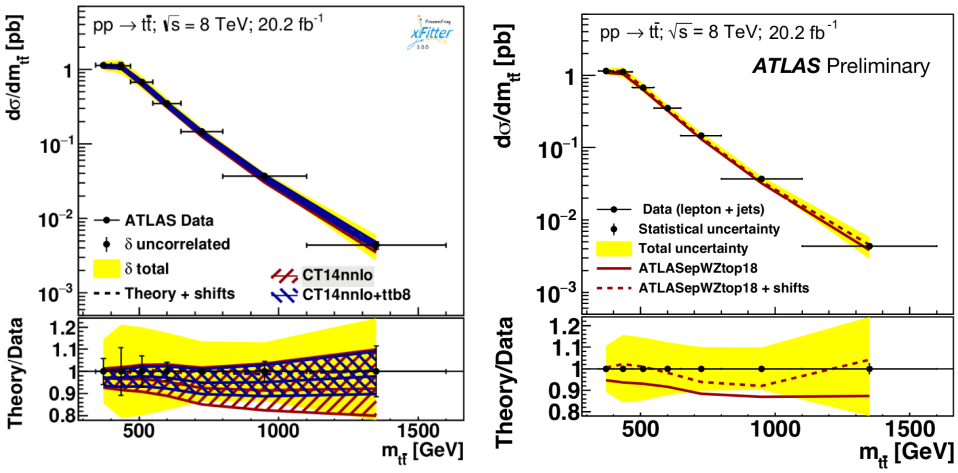


Figure 6.6: Normalized $t\bar{t}$ differential cross-sections as a function of the invariant mass ($m_{t\bar{t}}$). Left: our results and right: the plot shown in the ATLAS PUB Note.

in Fig. 6.6.

xFitter is doing something like:

$$\text{TotalError}^2 = \sum(\text{stat}^2 + \text{StatConst}^2 + \text{UncorError}^2 + \text{UncorConst}^2 + \text{sys\_up}^2 + \text{sys\_down}^2) \tag{6.1}$$

(6.3)

where :

- StatConst = statistical uncertainties, in case of error re-scaling[3].
- UncorError = uncorrelated uncertainties.
- UncorConst = uncorrelated uncertainties, in case of error re-scaling[4].
- stat = statistical uncertainties.
- sys_up(dwon) = asymmetric systematic errors.

[3]in our case StatConst = 0, since no re-scaling in requested.
[4]in our case UncorConst = 0, since no re-scaling in requested.

So the correct form shall be:

$$\sqrt{\text{Stat}^2 + \text{StatConst}^2 + \text{UncorError}^2 + \text{UncorConst}^2 + \sum \text{BETA}^2 + \text{Lumi}^2} \tag{6.4}$$

where :

- $\sum \text{BETA}^2$ = the quadratic sum of all systematic uncertainties after the symmetrization. without including uncorr.uncertainties.

Larger size of error bars

Generally, xFitter computes the error by adding quadratically, the statisrical error and the uncorrelated uncertainties as:

$$\text{uncor} = \sqrt{\text{stat}^2 + \text{unc}^2}\,. \tag{6.5}$$

The difference between the two errors bars shown in Fig. 6.6 is due to the fact that the plot in the ATLAS PUB Note shows only the statistical error in Eq. (6.5). In Fig. 6.6, the size of error bars in our result is lager because we are including the uncorrelated uncertainties into the calculation of Eq. (6.5).

One more thing to add, we also found a mistake in the input data files for the top data. The mistake is that one uncertainty source (ps_model) is treated as correlated[5] rather than the uncorrelated error (ps_model). Such mistake will result a double counting in Eq. (6.5), hence, larger size of error bars.

In addition, our solution is flexible enough to work for both, symmetric and asymmetric error sources.

Result after correction

Taking into account all these corrections, we can now see a better agreement between our result and that in the ATLAS PUB Note (Fig. 6.7).

Given that we both are using the same xFitter but we have got different result (Fig 6.6), this may be explained by the fact that the systematic uncertainties shown in the ATLAS Note were already symmetrized with a presumably different input file. Figure 6.7 shows that the same results are obtained when using the symmetrized systematic uncertainties.

[5] In this case we have: "ps_model+ uncor" and "ps_model- uncor". Hence, from the Eq. (6.5) we will have a double counting.

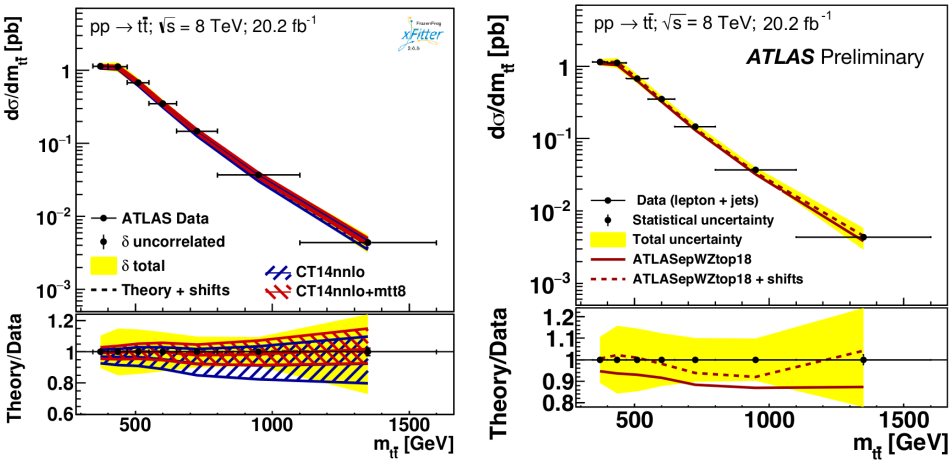


Figure 6.7: Normalized $t\bar{t}$ differential cross-sections as a function of the invariant mass ($m_{t\bar{t}}$). Left: our result after corrections and right: the plot shown in the ATLAS PUB Note.

6.3 xFitter vs. ePump

Previously, we showed how top data affect the shape of PDFs using xFitter. Now in this section will explore the difference between xFitter and ePump packages that are already introduced in chapter 4.

- **xFitter**
  - xFitter is not obviously to install with all the necessary interfaces.
  - xFitter is a framework for QCD analyses (profiling, fit, $\cdots$).
  - xFitter interfaces to a large number of theoretical and methodological options (fastNLO, APPLgrig, $\cdots$).
  - The breakdown of systematics may be asymmetric.
  - xFitter can only read LHAPDF formats.
  - currently, it is not supporting asymmetric errors when a correlation matrix is provided [6].
  - xFitter can generate comparison plots of data vs. theory, and an example is shown in Fig. 4.4.
  - Only global tolerance value may be used.

[6] As it was confirmed by Sasha Glazov.

- The code takes few time to run.
- xFitter uses directly SM parameters to calculate PDFs.[7]

- **ePump**
  - Easy to install.
  - ePump is a package used mainly to update a set of PDFs by using a new data set.
  - The breakdown of systimatics must be symmetrized, otherwise, ePump will not work.
  - ePump can read either CTEQ or LHAPDF formats.
  - Theoretical predictions must be computed before using fastNLO, APPLgrig, $\cdots$.
  - Either a global or dynamical tolerance value may be used, in which case a file of tolerance-squared values, defined as $(T_i^{\pm})^2$ for each $\pm$ error PDF, in a ".tol" file must be included with the PDF set.
  - The code takes only a few seconds to run. So it is quick and easy to try different combinations of data sets to compare their impact on the PDFs.
  - ePump doesn't use directly any of SM parameters for the input PDFs, since the input PDFs do have dependence on SM parameters such as W and Z boson masses and quarks masses and the theory file must be consist with these SM parameters.

6.3.1 LHAPDF vs. CTEQ

The first step for the comparison of xFitter and ePump is by comparing CT14 in LHADPF (note here as CT14_LHAPDF) format and CTEQ (note here CT14_CTEQ) format used by ePump.

Obviously, CT14_LHAPDF and CT14_CTEQ should be the same PDFs and no difference must be observed, but a little discrepancy was noticed at low x and Q.

The comparison is shown in Fig. 6.8. The blue error bands correspond to CT14_LHAPDF and the red error bands correspond to

[7] xFitter is controlled by three files: steering.txt, minuit.in.txt and ewparam.txt. Here ewparam.txt controls electroweak parameters such as W and Z boson masses and CKM matrix parameters. For more information see [68]

CT14_CTEQ/CT14_LHAPDF.

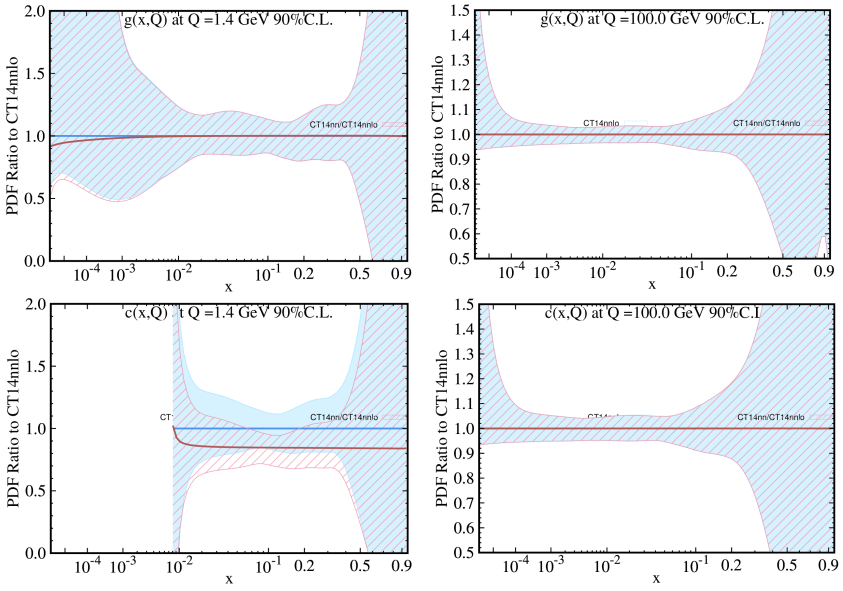


Figure 6.8: Comparison between the gluon-PDF at $Q = 1.4$ GeV and $Q = 100$ GeV for CT14_LHAPDF and for CT14_CTEQ. The error bands are each normalized to the CT14_LHAPDF central PDF. The blue error bands correspond to CT14_LHAPDF and The red error bands correspond to CT14_CTEQ/CT14_LHAPDF.

The difference between the LHAPDF and CTEQ format is only apparent at low x and at low Q, mainly, $Q = 1.4$ GeV. This difference seems to disappeared at higher Q ($Q > 1.7, 100$ GeV).

This small difference in interpolation probably arises when we convert CT14_CTEQ to CT14_LHAPDF. The large discrepancy observed with the charm PDF in Fig. 6.8 which showed an off-set through the whole range of x is treated differently in both format, since the charm quark generated mainly due to radiation from gluon splitting, this difference should be small at scale (Q) above the charm quark mass ($m_c = 1.275$ GeV^2).

6.3.2 xFitter vs. ePump

Based on the previous results, it will be wise to not use different format of CT14 for this comparison. Also, it is important to use the same experimental

data sets and the theoretical prediction for both code.

Hence, in our case, we made sure that xFitter and ePump satisfied the following:

1. Both codes are using the same LHAPDF PDFs[8].
2. The differential spectra in the lepton+jets channel, $m_{t\bar{t}}$ (Fig. 5.2) and $y_{t\bar{t}}$ are used for this comparison. Also, one should provide this data in the specific formats compatible with xFitter and ePump.
3. The theoretical prediction obtained by xFitter with fastNLO, are used for ePump.

The impact of $m_{t\bar{t}}$

The impact of the differential spectra in the lepton+jets channel is already shown in section 6.1.3. Our purpose, in this section, is to locate if there is any differences in profiling techniques used by both packages.

The first attempt toward this goal, is updating CT14 by adding $t\bar{t}$ differential cross-section $m_{t\bar{t}}$ using both xFitter and ePump. Figure 6.9 shows the impact of $t\bar{t}$ differential cross-section $m_{t\bar{t}}$ on CT14.

In Fig. 6.9 CT14nnlo is the reference (blue curve). CT14nnlo-mttbxFitter is the standard CT14 but updated by xFitter to include $m_{t\bar{t}}$ data (green curve). CT14nnlo+mttb8+ePump is the standard CT14 but updated by ePump to include $m_{t\bar{t}}$ data (red curve).

For completeness, we also compare in Fig. 6.10 the changes in the other flavor (u, d, and s) PDFs after updating CT14.

These plots show that including these data makes the gluon-PDF become softer when x is larger than around 0.2, and it reduces the uncertainty in the gluon PDF in the same region of x. More importantly, xFitter and ePump showed a wired behavior. Although there is a basic definition of the χ^2, Eq. (3.50), important differences may arise depending on the treatment of correlated uncertainties.

The disagreement between xFitter and ePump is only apparent for gluon-PDFs and s quark. In addition, at high Q, we observe a good agreement

[8] we can't use CTEQ format since xFitter is not compatible with it.

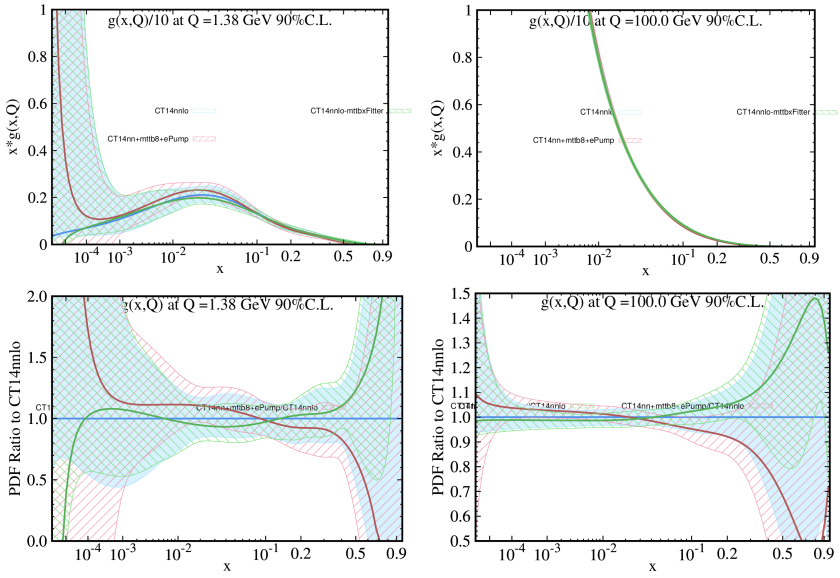


Figure 6.9: Comparison of the gluon-PDF at the scale $Q = 1.38$ and $Q = 100$ GeV for three analyses: CT14, CT14nnl-mttbxFitter and CT14nnlo+mttb8+ePump. In the bottom, the left and right curves and error bands are each normalized to the CT14nnlo central PDF.

between the two packages. For u and d quarks there is a small discrepancy that is mostly related to the fact that this data is not sensitive or has no great impact on quarks. Also, ePump uses symmetric systematic uncertainties and xFitter uses asymmetric ones.

The impact of $m_{t\bar{t}} + y_{t\bar{t}}$

In Fig. 6.9 CT14nnlo is the reference (blue curve). CT14-ytt-mttbxFitter is the standard CT14 but updated by xFitter to include $m_{t\bar{t}}$ data and $y_{t\bar{t}}$ data (green curve). CT14nn+mttb8+ytt+ePump is the standard CT14 but updated by ePump to include $m_{t\bar{t}}$ and $y_{t\bar{t}}$ data (red curve).

The impact of the differential spectra in the lepton+jets channel $y_{t\bar{t}}$ on CT14 is relevant. Adding $y_{t\bar{t}}$ to CT14 shows clearly, that the profiling methodology used by xFitter and ePump are far to be the same, despite using all of the same inputs, very different profiling are obtained by each code. In Figs. 6.11 and 6.12 the profiling obtained with xFitter has the largest difference.

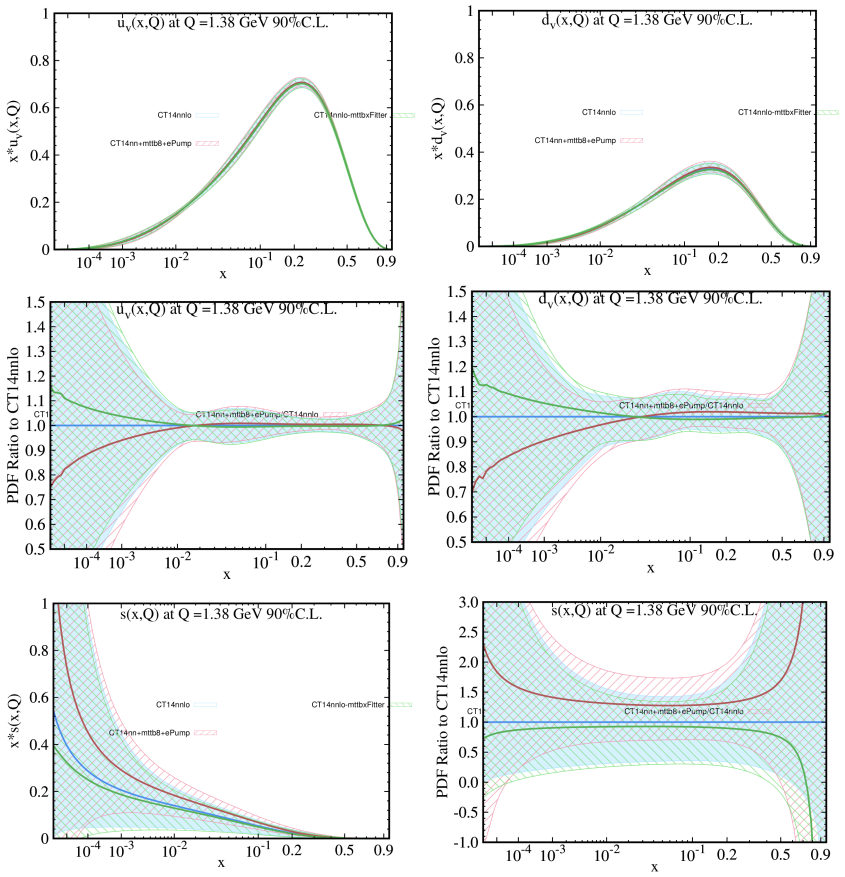


Figure 6.10: Same as Fig. 6.9 but at the scale $Q = 1.38$ GeV and for the updated xu, xd and xs PDFs.

xFitter uses the Hessian updating method that it directly works with the (small set of) Hessian PDFs and it is a simpler and much faster way to estimate the effects of the new data. The advantage is that this method directly calculates the minimum of the χ^2 function within the Hessian approximation, ePump extends the Hessian updating method to be used to update any Hessian PDF sets obtained from an earlier global analysis.

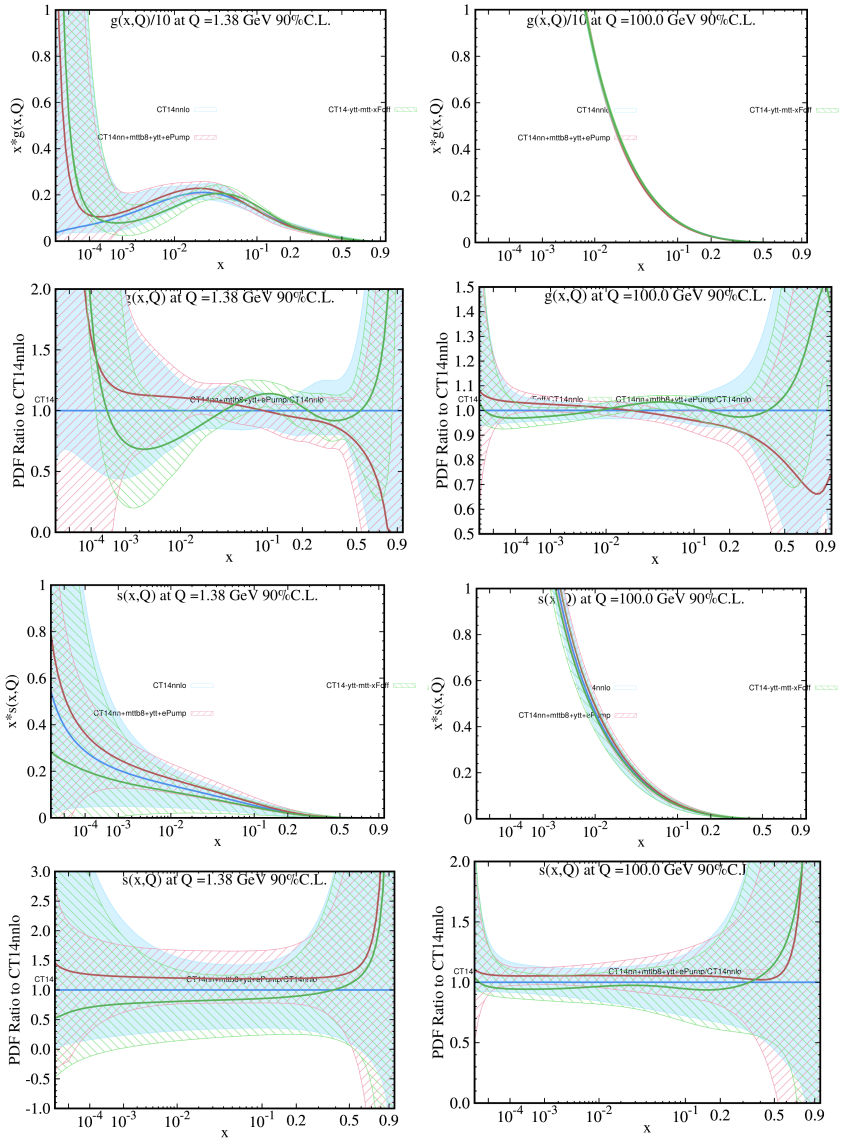


Figure 6.11: Comparison of the gluon-PDF ans strange-PDF at the scale $Q = 1.38$ and $Q = 100$ GeV for three analyses: CT14, CT14-ytt-mttbxFitter and CT14nn+mttb8+ytt+ePump. There related curves and error bands are each normalized to the CT14nnlo central PDF.

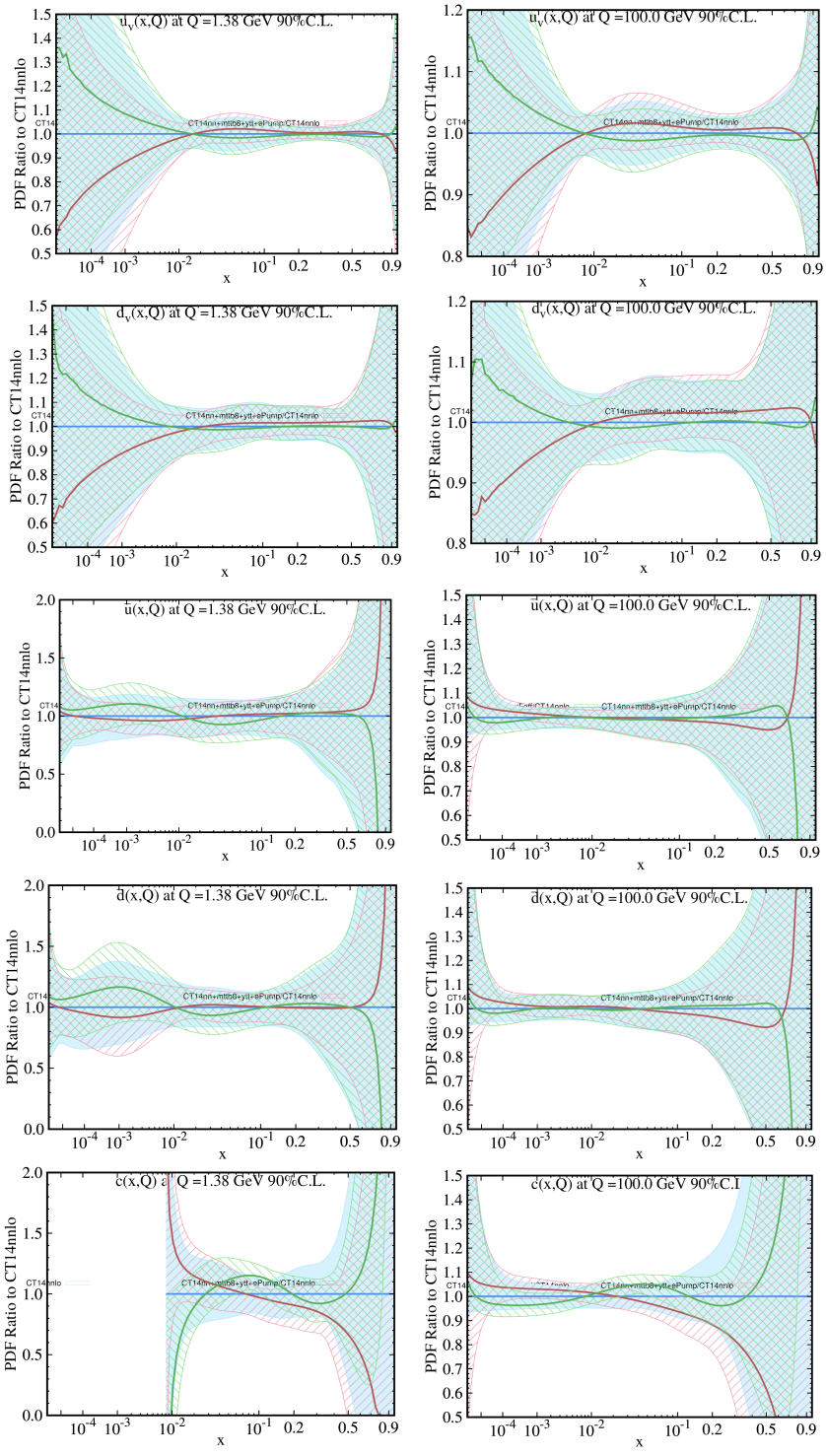


Figure 6.12: Same as Fig. 6.11 but the related curves and error bands are each normalized to the CT14nnlo central PDF for the update PDFs : $xu, xd, x\bar{u}, x\bar{d}$ and xc PDFs.

Chapter 7

Conclusion

In this report, we have discussed the impact of LHC measurements upon the extraction of parton distribution functions.

The PDFs determination in global analyses is a complex procedure, which needs the parametrization using the fits of experimental data. Although different PDF parameterizations are available for a general user, finding the impact of new measurements of the data on PDFs without doing a global QCD analysis would be useful. For example, we can find which kind of PDFs can be constrained in the presence of a specific new data. In this regard, the Hessian profiling technique is a good choice.

We have discussed how to investigate the effects that a new set of Drell-Yean and top quark pair production measurements have within an existing PDF set. Using the profiling formalism we have determined the impact of the recent Drell-Yean and top quark pair production data on the CT14 PDFs. The results of the profiling on quarks and gluon PDFs, their relative uncertainties, and on the PDF ratios with respect to before profiling procedure are shown.

A significant reduction of the relative gluon and xs uncertainties is observed in the medium and large x for CT14 PDF set. The profiling procedure using new set of Drell-Yean and top quark affected the strange xs and gluon distributions at the CT14 PDF sets. These findings are interesting and show the significance of the Drell-Yean and top quark production cross section data to constrain gluon and strange PDFs.

The next step in this analysis will be to fix issues observed with xFitter and the solutions we come up with. Again, these issues have no effect on

profiling and fits. Such discovery is saluted by **Sasha Glasov**. At the end, a comparison of xFitter and ePump is shown for the first time.

Appendix A

Impact of Top data on sea quarks

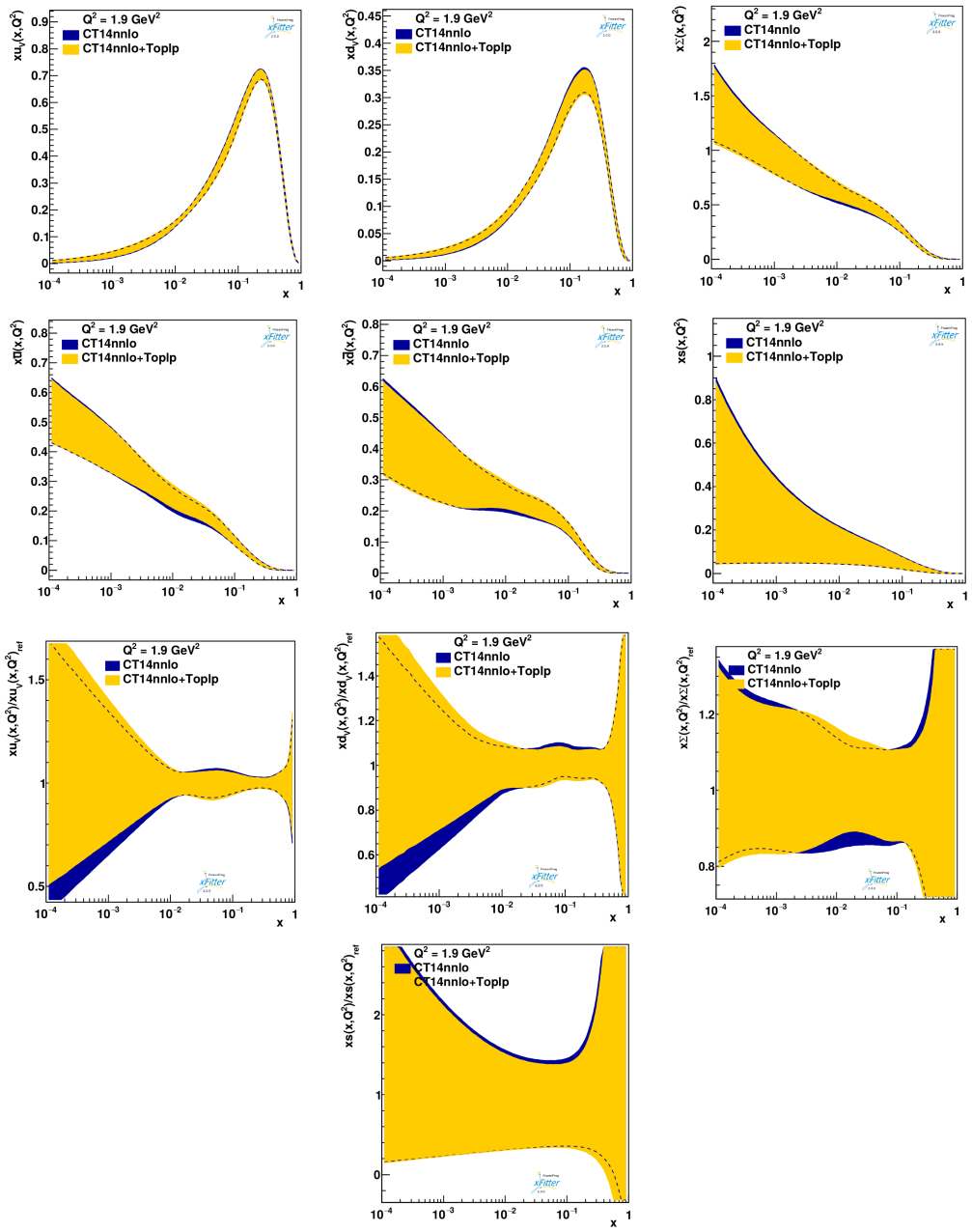


Figure A.1: The parton distribution of xu_v, xd_v, $x\bar{u}$, $x\bar{u}$, $x\Sigma$, xs and the related uncertainties of xu_v, xd_v, $x\Sigma$ and xs extracted from CT14 PDFs. The results obtained after the profiling procedure e.g., adding the $m_{t\bar{t}}$ and p_{T}^{t} spectra from lepton+jets, compared with corresponding same features before profiling. Newly added top quark data obviously constrained distributions of $\delta x\Sigma/x\Sigma$.

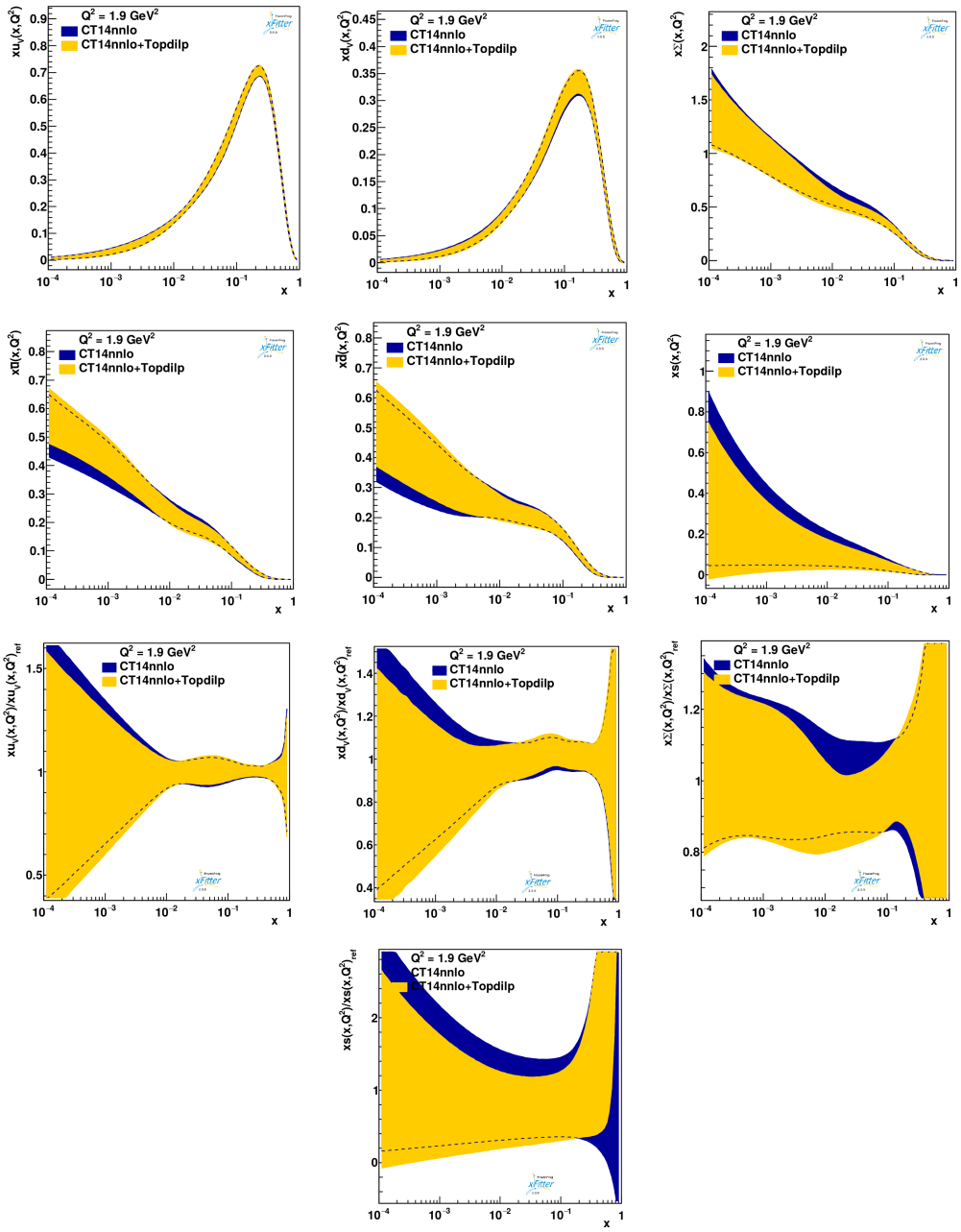


Figure A.2: The parton distribution of xu_v, xd_v, $x\bar{u}$, $x\bar{u}$, $x\Sigma$, xs and the related uncertainties of xu_v, xd_v, $x\Sigma$ and xs extracted from CT14 PDFs. The results obtained after the profiling procedure e.g., adding the $m_{t\bar{t}}$ and $y_{t\bar{t}}$ spectra from di-lepton, compared with corresponding same features before profiling. Newly added top quark data constrained distributions of $\delta x\Sigma/x\Sigma$ and xs/xs.

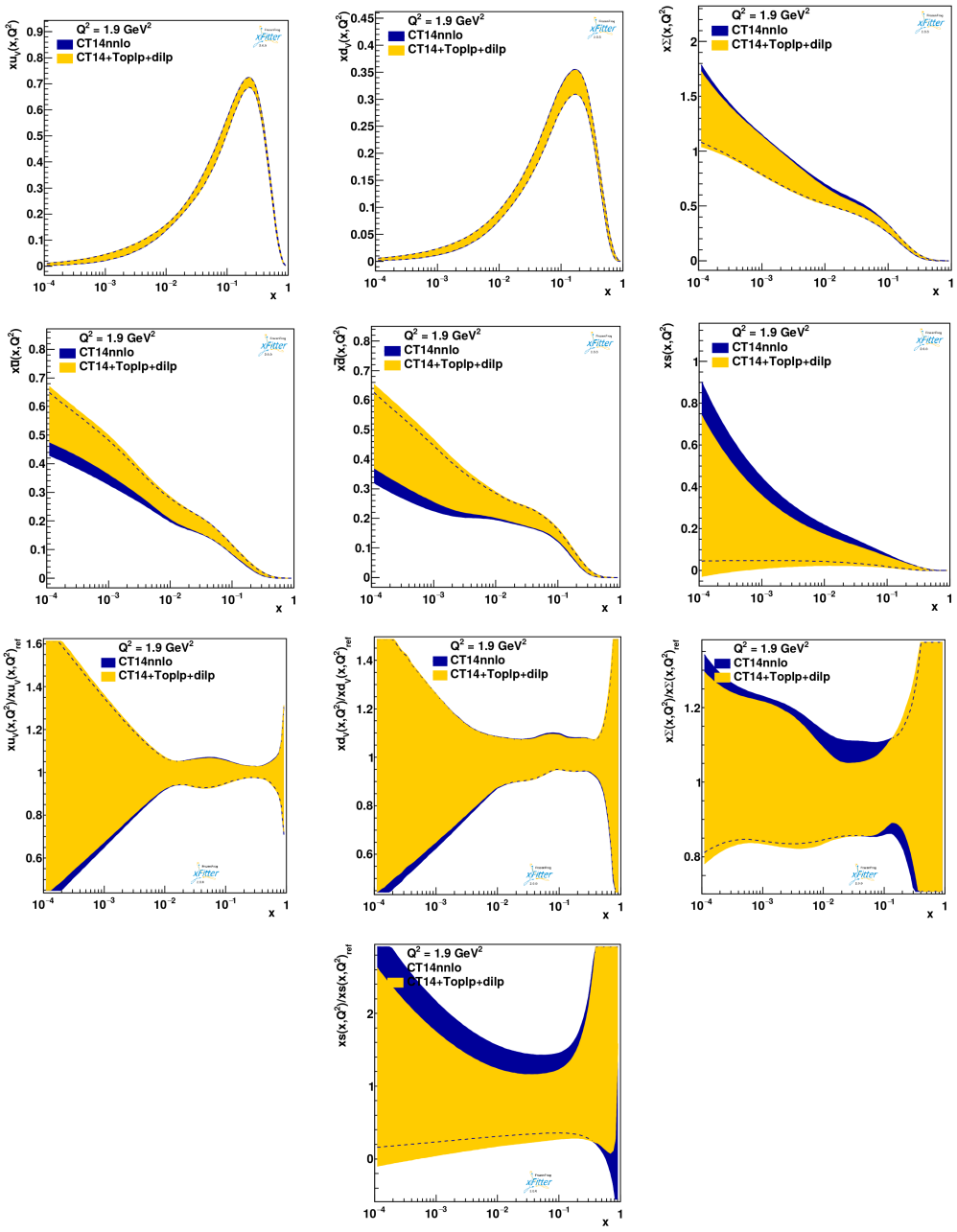


Figure A.3: The parton distribution of xu_v, xd_v, $x\bar{u}$, $x\bar{u}$, $x\Sigma$, xs and the related uncertainties of xu_v, xd_v, $x\Sigma$ and xs extracted from CT14 PDFs. The results obtained after the profiling procedure e.g., adding the $m_{t\bar{t}}$ and $y_{t\bar{t}}$ spectra from di-lepton and the $m_{t\bar{t}}$ and p_{T}^{t} spectra from lepton+jets, compared with corresponding same features before profiling. Newly added top quark data constrained distributions of $\delta x\Sigma/x\Sigma$ and xs/xs.

References

[1] Juan Rojon, et al. The PDF4LHC report on PDFs and LHC data: results from Run I and preparation for Run II.arXiv:1507.00556, doi:10.1088/0954-3899/42/10/103103

[2] M. Czakon, D. Heymes, A. Mitov, Dynamical scales for multi-TeV top-pair production at the LHC, JHEP 04 (2017) 071. arXiv:1606.03350, doi:10.1007/JHEP04(2017)071.

[3] J. Currie, E. W. N. Glover, J. Pires, Next-to-Next-to Leading Order QCD Predictions for Single Jet Inclusive Production at the LHC, Phys. Rev. Lett. 118 (7) (2017) 072002. arxiv:1611.01460

[4] J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, J. Pires, Precise predictions for dijet production at the LHC, Phys. Rev. Lett. 119 (15) (2017) 152001.arXiv:1705.10271

[5] J. M. Campbell, R. K. Ellis, C. Williams, Direct Photon Production at Next-toNext-to-Leading Order, Phys. Rev. Lett. 118 (22) (2017) 222001 arXiv:1612.04333

[6] R. Boughezal, C. Focke, X. Liu, F. Petriello, W-boson production in association with a jet at next-to-next-to-leading order in perturbative QCD, Phys. Rev. Lett. 115 (6) (2015) 062002. arXiv:1504.02131

[7] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, T. A. Morgan, The NNLO QCD corrections to Z boson production at large transverse momentum, JHEP 07 (2016) 133. arXiv:1605.04295

[8] S. Alekhin, et al., HERAFitter, Eur. Phys. J. C75 (7) (2015) 304. arXiv:1410.4412

[9] V. Bertone, S. Carrazza, J. Rojo, APFEL: A PDF Evolution Library with QED corrections, Comput.Phys.Commun. 185 (2014) 1647. arXiv:1310.1394

[10] M. Botje, QCDNUM: Fast QCD Evolution and Convolution, Comput.Phys.Commun. 182 (2011) 490532. arXiv:1005.1481

[11] T. Carli, et al., A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project, Eur.Phys.J. C66 (2010) 503. arXiv:0911.2985

[12] Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. arXiv:1511.04716

[13] H. Paukkunen and P. Zurita, JHEP 1412, 100 (2014). arXiv:1402.6623

[14] C. P. Y. C. Schmidt, J. Pumplin and P. Yuan, J. High Energ. Phys. (2018). arXiv:1806.07950

[15] ATLAS EXPERIMENT−Luminosity Public Results Run 2. Available online:ATLAS Luminosity

[16] The ATLAS Collaboration. Luminosity determination in pp collisions at s =7TeV using the ATLAS detector at the LHC. Eur.Phys. J. C, 71(1630), 2011. 33

[17] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003.

[18] ATLAS Collaboration, Performance of the ATLAS Trigger System in 2010,Eur. Phys. J. C 72 (2012) 1849,arXiv:1110.1530 [hep-ex].

[19] R.P. Feynman. Photon hadron interactions. W.A. Benjamin, New York, 1972.

[20] J.D. Bjorken. Asymptotic Sum Rules at Infinite Momentum. Phys.Rev., 179:15471553, 1969.

[21] J. C. Collins, D. E. Soper, The Theorems of Perturbative QCD, Ann. Rev. Nucl. Part. Sci. 37 (1987) 383409.

[22] J. C. Collins, D. E. Soper, G. F. Sterman, Factorization of Hard Processes in QCD, Adv. Ser. Direct. High Energy Phys. 5 (1989) 191. arXiv:hep-ph/0409313.

[23] G. Altarelli, G. Parisi, Asymptotic Freedom in Parton Language, Nucl. Phys. B126 (1977) 298318.doi:10.1016/0550-3213(77)90384-4

[24] L. N. Lipatov, The parton model and perturbation theory, Sov. J. Nucl. Phys. 20 (1975) 94102, [Yad. Fiz.20,181(1974)].

[25] G. P. Salam, J. Rojo, A Higher Order Perturbative Parton Evolution Toolkit (HOPPET), Comput. Phys. Commun. 180 (2009) 120156.arXiv:0804.3755, doi:10.1016/j.cpc.2008.08.010.

[26] V. Bertone, S. Carrazza, J. Rojo, APFEL: A PDF Evolution Library with QED corrections, Comput.Phys.Commun. 185 (2014) 1647.arXiv:1310.1394, doi:10.1016/j.cpc.2014.03.007

[27] M. Botje, QCDNUM: Fast QCD Evolution and Convolution, Comput.Phys.Commun. 182 (2011) 490532.arXiv:1005. 1481, doi:10.1016/j.cpc.2010.10.020.

[28] A. Vogt, Efficient evolution of unpolarized and polarized parton distributions with qcd-pegasus, Comput. Phys. Commun. 170 (2005) 6592.arXiv:hep-ph/0408244.

[29] M. R. Whalley, D. Bourilkov, R. C. Group, The Les Houches accord PDFs (LHAPDF) and LHAGLUE, in: HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B, 2005, pp. 575581.arXiv:hep-ph/0508110.

[30] M. Dittmar, et al., Parton Distributions.arXiv:0901.2504.

[31] S. Alekhin and S. Moch, Phys. Lett. B699, 345 (2011).arXiv:1011.5790

[32] S. Alekhin, OPENQCDRAD, a program description and the code are available via: http://www-zeuthen.desy.de/alekhin/OPENQCDRAD.

[33] W.K. Tung, H.L. Lai, A. Belyaev, J. Pumplin, D. Stump, et al. Heavy Quark Mass Effects in Deep Inelastic Scattering and Global QCD Analysis. JHEP, 0702:053, 2007, hep-ph/0611254.

[34] D. Dolgov, R. Brower, S. Capitani, John W. Negele, A. Pochinsky, et al. Moments of structure functions in full QCD. Nucl.Phys.Proc.Suppl., 94:303306, 2001, hep-lat/0011010.

[35] Wolfram Schroers. Parton distributions from the lattice.Nucl.Phys., A755:333336, 2005, hep-ph/0501156.

[36] S. Alekhin, J. Bluemlein, and S. Moch. The ABM parton distributions tuned to LHC data. Phys.Rev., D89:054028, 2014, 1310.3059.

[37] S. Alekhin, J. Blumlein, and S. Moch. Parton Distribution Functions and Benchmark Cross Sections at NNLO. Phys.Rev., D86:054009, 2012, 1202.2281.

[38] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, et al. The CT10 NNLO Global Analysis of QCD. Phys.Rev., D89:033009, 2014, 1302.6246.

[39] J.F. Owens, A. Accardi, and W. Melnitchouk. Global parton distributions with nuclear and finite-Q^2corrections. Phys.Rev., D87(9):094012, 2013, 1212.1702.

[40] F.D. Aaron et al. Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA. JHEP, 1001:109, 2010, 0911.0884.

[41] P. Belov: et al. Parton distribution functions at LO, NLO and NNLO with correlated uncertainties between orders. 2014, 1404.4234.

[42] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, et al. Parton distributions with LHC data. Nucl.Phys., B867:244289, 2013, 1207.1303.

[43] Richard D. Ball et al. A Determination of parton distributions with faithful uncertainty estimation. Nucl.Phys., B809:163, 2009, 0808.1231.

[44] Michiel Botje et al. The PDF4LHC Working Group Interim Recommenda- tions. 2011, 1101.0538.

[45] Tullio Regge, "Introduction to complex orbital momenta," Il Nuovo Cimento Series 10, Vol. 14, 1959, p. 951.

[46] R. G. Roberts, The structure of the proton. Cambridge Univ. Press, Cambridge, UK, 1990.

[47] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. Eur. Phys. J., C63:189285, 2009, 0901.0002.

[48] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. Phys.Rev., D82:074024, 2010, 1007.2241.

[49] J. Pumplin, D. R. Stump and W. K. Tung, Phys. Rev. D 65 (2002) 014011 [arXiv:hep- ph/0008191].

[50] C. Amsler et al. [Particle Data Group], Phys. Lett. B 667 (2008) 1.

[51] J. Pumplin et al., Phys. Rev. D 65 (2002) 014013 [arXiv:hep-ph/0101032].

[52] W. T. Giele and S. Keller, Phys. Rev. D 58 (1998) 094023 [arXiv:hep-ph/9803393].

[53] W. T. Giele, S. A. Keller and D. A. Kosower, arXiv:hep-ph/0104052.

[54] R. D. Ball et al. [NNPDF Collaboration], Nucl. Phys. B 809 (2009) 1 [arXiv:0808.1231 [hep-ph]].

[55] A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, Eur. Phys. J. C 28 (2003) 455 [arXiv:hep-ph/0211080].

[56] Progress in the Determination of the Partonic Structure of the Proton.arXiv:1301.6754

[57] https://lhapdf.hepforge.org/

[58] NNPDF Collaboration, R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, et al., Parton distributions with LHC data, Nucl.Phys. B867 (2013) 244289. arXiv:1207.1303

[59] M. Glck, P. Jimenez-Delgado, and E. Reya, Dynamical parton distributions of the nucleon and very small-x physics, Eur.Phys.J. C53 (2008) 355366.arXiv:0709.0614

[60] H1 and ZEUS Collaboration, F. D. Aaron et al., PDF fits including HERA-II high Q 2 data (HERAPDF1.5). H1prelim-10-141, ZEUS-prel-10-017.

[61] H1 and ZEUS Collaboration, F. D. Aaron et al., Combined Measurement and QCD Analysis of the Inclusive e p Scattering Cross Sections at HERA, JHEP 1001 (2010) 109,arXiv:0911.0884

[62] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P.M. Nadolsky, J. Pumplin, C.-P. Yuan, Phys. Rev. D82, 074024 (2010), 1007.2241

[63]] S. Dulat, T.J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, C.-P. Yuan (2015), 1506.07443

[64] G. Aad et al. (ATLAS), Phys. Rev. D85, 072004 (2012), 1109.5141

[65] R. Aaij et al. (LHCb), JHEP 06, 058 (2012), 1204.1620

[66] V.M. Abazov et al. (D0), Phys. Rev. D91, 032007 (2015), [Erratum: Phys. Rev.D91,no.7,079901(2015)], 1412.2862

[67] https://indico.cern.ch/event/361123/contributions/856249/attachments/1136197/1625877/

[68] xFitter, An open source QCD fit framework. http://xFitter.org [xFitter.org]

[69] R. D. Ball et al. [NNPDF Collaboration], Nucl. Phys. B 849, 112 (2011) Erratum: [Nucl. Phys. B 854, 926 (2012)] Erratum: [Nucl. Phys. B 855, 927 (2012)].arXiv:1012.0836

[70] S. Camarda et al. [HERAFitter developers Team], Eur. Phys. J. C 75, no. 9, 458 (2015).arXiv:1503.05221

[71] G. Watt and R. S. Thorne, JHEP 1208, 052 (2012).arXiv:1205.4024

[72] J. Gao, L. Harland-Lang and J. Rojo, Phys. Rept. 742, 1 (2018).arXiv:1709.04922

[73] ATLAS Collaboration, Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector,arXiv: 1511.04716 [hep-ex].

[74] ATLAS Collaboration, Measurement of top quark pair differential cross sections in the dilepton channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV with ATLAS, Phys. Rev. D 94 (2016) 092003.arXiv: 1607.07281 [hep-ex].

[75] ATLAS Collaboration, Measurement of the DrellYan triple-differential cross section in pp collisions at $\sqrt{s} = 8$ TeV.arXiv:1710.05167

[76] ATALS Collaboration, Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV with ATLAS.arXiv:1607.07281

[77] M. Czakon, D. Heymes and A. Mitov, fastNLO tables for NNLO top-quark pair differential distributions, (2017).arXiv:1704.08551

[78] T. Kluge, K. Rabbertz and M. Wobisch, FastNLO: Fast pQCD calculations for PDF fits, Deep inelastic scattering. Proceedings, 14th International Workshop, DIS 2006, Tsukuba, Japan, April 20-24, 2006, 2006 483. arXiv: hep-ph/0609285.

[79] D. Britzger, K. Rabbertz, F. Stober and M. Wobisch, New features in version 2 of the fastNLO project, Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012): Bonn, Germany, March 26-30, 2012, 2012 217.arXiv: 1208.3641 [hep-ph].

[80] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera and J. Rojo, Pinning down the large-x gluon with NNLO top-quark pair differential distributions.arXiv: 1611.08609 [hep-ph].

[81] M. Czakon et al., Top-pair production at the LHC through NNLO QCD and NLO EW.arXiv: 1705.04105 [hep-ph].

[82] T. Carli et al., A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project, Eur. Phys. J. C 66 (2010) 503.arXiv: 0911.2985 [hep-ph].

[83] J. M. Campbell and R. K. Ellis, An Update on vector boson pair production at hadron colliders.arXiv: hep-ph/9905386.

[84] J. M. Campbell and R. K. Ellis, MCFM for the Tevatron and the LHC.arXiv: 1007.3492 [hep-ph].

[85] Tancredi Carli, et al., A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project. Eur. Phys. J., C66:503524, 2010.

[86] D. Britzger et al. [fastNLO Collaboration], doi:10.3204/DESY-PROC-2012-02/165. D. Britzger, G. S. Klaus Rabbertz, F. Stober and M. Wobisch, PoS DIS 2015, 055 (2015).

[87] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer and M. Wiedermann, HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR, Comput. Phys. Commun. 182, 1034 (2011).

[88] ATLAS Collaboration. Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and $Z\backslash\gamma^*$ and $t\bar{t}$ cross sections. ATL-PHYS-PUB-2018-017

[89] https://www.researchgate.net/figure/A-selection-of-gluon-PDFs-with-uncertainties-at-Q-2-40-GeV-2-plotted-using-the$_fig1_3$19501574

[90] A New Method for Reducing PDF Uncertainties in the High-Mass Drell-Yan Spectrum. arXiv:1809.09481

[91] R. S. Thorne, Phys. Rev. D 86, 074017 (2012),arXiv:1201.6180

[92] https://mstwpdf.hepforge.org/

[93] http://www.hep.ucl.ac.uk/mmht/