UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

PARAMETRIZATION OF PARTON DISTRIBUTION FUNCTIONS WITH FANTÔMAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

FÍSICO

P R E S E N T A :

DANIEL MAXIMILIANO PONCE CHÁVEZ

TUTORA:

DRA. AURORE MARIE PASCALE
NICOLE COURTOY

CIUDAD DE MÉXICO, 2022

A mi abuelo Salvador,
cuyo afecto, ingenio y asombro
son entrañablemente recordados.
Que descanse en paz.

A mi sobrina Macarena,
que ha traído una nueva luz
a sus padres, abuelos y tíos.

AGRADECIMIENTOS

Me es complicado agradecer en plenitud y a detalle, pues la culminación de esta tesis no se habría logrado sin el apoyo de una gran cantidad de personas, entre ellas las y los profesores de la Facultad de Ciencias que han contribuido a mi formación. No obstante, algunas otras han sido esenciales durante de este proceso y me gustaría ofrecerles mi gratitud:

Principalmente, a la Dra. Aurore Courtoy por abrirme las puertas a este proyecto y por su dirección tan atenta y acertada, por su paciencia y por compartir sus conocimientos siempre que fue necesario. Al Dr. Pavel Nadolsky quien también coordina este proyecto y ha dirigido de una manera brillante y sabia al *Fantômas Development Team*. Asimismo, quiero agradecer a Lucas, Ryan y Varada quienes también son parte del equipo de desarrollo.

A mi madre y mi padre por seguir presentes, por asistirme para llegar hasta este punto y por ser mis guías persistentes. También a mi hermana y mi hermano, por ser tan semejantes y tan dispares al mismo tiempo, mostrándome una y otra lo que es el apoyo mutuo. Espero tenerlos a ustedes y su afecto por muchos años más.

A Alex, Braulio y Diego por estos primeros doce años de amistad y por espejearnos para poder seguir creciendo juntos. A mis amigos de la Facultad de Ciencias Políticas y Sociales por regalarme su compañía durante los últimos meses y a aquellos otros amigos con los que formé un vínculo más allá de los laboratorios y salones de clase.

Con especial énfasis, quiero agradecer a mi abuelo, quien durante el tiempo que compartimos juntos, mostró una fe ciega y un constante asombro por cada uno de mis acciones y logros, por minúsculos que fueran. Celebrando su vida, esta tesis es especialmente para él.

Por último, quiero agradecer a los proyectos No. IN111222 y IA101720 del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la Dirección General de Asuntos del Personal Académico (DGAPA) y al proyecto Ciencia de Frontera No. 51244 (F3) del Consejo Nacional de Ciencia y Tecnología (CONACyT).

RESUMEN

La presente tesis se sustenta en el marco de la cromodinámica cuántica (QCD) y la física de hadrones, situando al protón como una estructura hadrónica constituida por partículas fundamentales y mediadores de la fuerza fuerte; a saber, quarks antiquarks y gluones. Esta hipótesis es abarcada por el modelo de quarks-partones propuesto por Richard Feynman para dar una interpretación a las observaciones en experimentos de dispersión profundamente inelástica (DIS) desarrollados en aceleradores de partículas como SLAC, HERA, entre otros. Los experimentos DIS permiten sondear la estructura interna de un hadrón a través de la dispersión de un leptón, como resultado del intercambio de un fotón virtual de longitud de onda corta. En el régimen de altas energías, el leptón interactúa con los constituyentes internos. Estudiando la sección eficaz del proceso en cuestión, surgen funciones que parametrizan la distribución de los partones dentro de los hadrones, en términos del 4-momento fraccionario llevado por dichos constituyentes. Estas funciones reciben el nombre de funciones de distribución de partones (PDF) y son de suma importancia para describir la materia a un nivel fundamental, por lo que su determinación es de gran interés, sin embargo, las PDFs no pueden ser evaluadas explícitamente con el marco teórico mediante el cual fueron inferidas. La única vía disponible para su determinación es la parametrización a través de datos experimentales, siendo un procedimiento en la mira de esta tesis. La determinación de funciones de distribuciones de partones es una tarea compleja bajo constante actualización, como resultado de los avances continuos en el área experimental y computacional. Existen diversas colaboraciones científicas y proyectos a cargo de la determinación de funciones de distribución de partones.

El objetivo de esta tesis es introducir el proyecto *Fantômas4QCD* cuyo propósito es

parametrizar funciones de distribución de partones mediante un programa computacional llamado *Metamorph* creado en *C++* y *Mathematica*. El proyecto representa una colaboración liderada por la Dra. Aurore Courtoy y el Dr. Pavel Nadolsky, entre físicos de la Universidad Nacional Autónoma de México y Southern Methodist University. El proyecto ha progresado por etapas en las que el autor de la tesis ha participado: acercamiento a conceptos de QCD y DIS, escritura y revisión de código, comparación con resultados conocidos, redacción, etc. *Fantômas4QCD* busca parametrizar PDFs de manera eficiente e innovadora a través de *Metamorph*, implementando un algoritmo que se basa en una herramienta flexible perteneciente al sector de diseño gráfico: las curvas de Bézier. Con la finalidad de acercar al lector y contextualizar el uso en la parametrización de funciones de distribución de partones, se presenta una síntesis sobre el origen y la aplicación de las curvas de Bézier, así como las reglas y restricciones que deben cumplir las PDFs parametrizadas. Asimismo, por medio de ejemplos concretos, se presentan los resultados obtenidos con *Metamorph* al reproducir las parametrizaciones de PDFs más recientes de uno de los proyectos colaborativos pioneros en el área: CTEQ. Después, se discuten algunos criterios para guiar a los usuarios en la implementación de *Metamorph* de manera práctica y apropiada, partiendo de los resultados y observaciones realizadas por el *Fantômas Development Team*, durante la creación del proyecto, que tuvo sus inicios en el verano de 2021. Finalmente, se expone de manera breve el estado actual, aplicaciones y planes a futuro para *Fantômas4QCD*.

ABSTRACT

This thesis is supported by the Quantum Chromodynamics (QCD) and hadronic physics frameworks, which situate the proton as a hadronic structure constituted by fundamental particles and strong force mediators, namely, quarks, antiquarks and gluons. This hypothesis is encompassed within the quark-parton model proposed by Richard Feynman, aiming to give an interpretation to observations made in deep inelastic scattering experiments (DIS) developed in particle accelerators, such as SLAC, HERA, among others. DIS experiments allow probing the inner structure of hadrons by means of a scattered lepton, as a result of a short wavelength virtual photon exchange. In the regime of high energies, this lepton interacts with the inner hadron constituents. By carrying out the kinetic analysis and studying the cross section of the process under discussion, functions parametrizing the distribution of partons within hadrons arise, in terms of the fractional 4-momentum carried by such constituents. These functions are given the name of parton distribution functions (PDF) and they are of substantial significance to describe matter at a fundamental level. Consequently, PDF determination is a task of great interest. Nonetheless, these functions cannot be explicitly evaluated with the framework by which they were inferred. The only available way to determine these objects is by means of parametrization through experimental data, being a procedure in the aim of the present thesis. PDF determination is a complex task under constant updating, as a result of the continuous advances attained in experimental and computational fields. Multiple scientific collaborations and projects exist to determine parton distribution functions.

The objective of this thesis is to introduce the *Fantômas4QCD* project, whose purpose is to parametrize parton distribution functions by means of a program named *Metamorph* created with both *C++* and *Mathematica*. This project emerges as a collaboration lead by Dr. Aurore Courtoy and Dr. Pavel Nadolsky, between physicists of UNAM and Southern Methodist University. The project was developed by stages in which the author of the thesis took part: approaching to QCD and DIS, code writing and review, comparing with known results, writing, etc. *Fantômas4QCD* seeks to parametrize PDFs in an efficient and innovative manner with *Metamorph*, implementing an algorithm based on a flexible tool belonging to the graphic design sector: Bézier curves. With the purpose of approaching readers to the topic and give context to the usage in parton distribution parametrization, we present a synthesis of the origin and applications of Bézier curves as well as the rules and constraints that parametrized PDFs must follow. At the same time, through concrete examples, we present the results obtained with *Metamorph* by replicating the PDF parametrizations produced by one of the pioneering projects in the field: CTEQ. Then, we discuss some criteria intended to guide users in the implementation of *Metamorph* in a practical and appropriate way, arising from the results and observations made by the *Fantômas Development Team* while creating the project, which finds its beginnings in the summer of 2021. Finally, we discuss the current state, applications and future plans for *Fantômas4QCD*.

CONTENTS

Agradecimientos **3**

Resumen **5**

List of Figures **11**

List of Tables **13**

1 Introduction **15**

2 Theoretical Background **21**

2.1 The Standard Model 21

2.2 Strong Force and QCD 25

2.3 Scattering and Hadronic Structure 27

2.3.1 Deep Inelastic Scattering 29

2.3.2 Bjorken Scaling 32

2.3.3 The Quark-Parton Model 33

2.3.4 Parton Distribution Functions and Non-Perturbative QCD 34

3 Methodology **43**

3.1 State of the Art: PDF Parametrization 43

3.1.1 CTEQ and CT18 44

3.2 Fantômas4QCD 48

3.2.1 Why Fantômas? 48
3.2.2 Bézier Curves and Bernstein Polynomials 50
3.2.3 Fitting with Bézier Curves 52
3.2.3.1 Functional Form 52
3.2.3.2 The Bézier Curve Matrix Method 53
3.2.3.3 Error Function Minimization 54

4 Results **57**
4.1 *Metamorph Mathematica* Notebook 57
4.1.1 Examples 60
4.1.1.1 Up Valence 60
4.1.1.2 Down Valence 62
4.1.1.3 Gluon 64
4.1.1.4 Strange (Antistrange) 65
4.2 Scaling and Control Point Selection 66
4.2.1 Supplementary Discussion 69

5 Conclusions **71**

Bibliography **73**

LIST OF FIGURES

Figure 2.1: Confined colored quarks within a neutral color proton, interacting via gluon exchange 26

Figure 2.2: Single-photon e^-p elastic scattering 27

Figure 2.3: DIS as an inclusive process 29

Figure 2.4: Short wavelength virtual photon resolving the quark structure within a hadron 33

Figure 2.5: Experimental data of proton's $F_2(x)$ at different Q^2 scales 34

Figure 2.6: Valence quarks $F_2(x)$ structure function 35

Figure 2.7: Sets of experimental measurements of the running coupling constant 40

Figure 3.1: CT18's weighed PDFs for different partons at two different energy scales . 46

Figure 3.2: CT18's weighed parametrized PDFs at $Q = 1.3$ GeV 47

Figure 3.3: Film poster of 1965 Hunebelle's *Fantômas se déchaîne* 48

Figure 3.4: Flying *DS* replica 49

Figure 3.5: *Citroën DS* sketch by the acclaimed designer Flaminio Bertoni 49

Figure 3.6: Bézier curves depicted as successive linear interpolations 50

Figure 4.1: Input *.card* file example for *Metamorph* 57

Figure 4.2: Softsign and linear bounded scaling functions 59

Figure 4.3: Plots of CT18's and *Metamorph*'s fit for the up valence modulation function with evenly spaced control points 60

Figure 4.4: Ratio plot between CT18's and *Metamorph*'s up valence modulators, remaining close to 1 throughout the entire domain 61

Figure 4.5: Plots of CT18's and *Metamorph*'s fit for the down valence modulation function with unevenly spaced control points 62
Figure 4.6: Ratio plot between CT18's and *Metamorph*'s down valence modulators, remaining close to 1 throughout the entire domain 63
Figure 4.7: Plots of CT18's and *Metamorph*'s fit for the gluon modulation function with evenly spaced control points 64
Figure 4.8: Plots of CT18's and *Metamorph*'s fit for the strange modulation function with unevenly spaced control points 65
Figure 4.9: Ratio plot between CT18's and *Metamorph*'s gluon and strange modulators respectively, remaining close to 1 throughout the entire domain 66
Figure 4.10: CT18's and *Metamorph*'s Up-Valence modulator, with 11 unevenly distributed control points and $f = 1$ 67

LIST OF TABLES

Table 2.1: Chart of fermions arranged by generation 21

Table 2.2: Bosons arranged by interaction [1] 23

Table 2.3: Comparative strength of the fundamental forces at subatomic scales 23

Table 3.1: CT18's best-fit parameters for up valence quarks and gluons at $Q_0 = 1.3$ GeV [2]. 46

Table 4.1: *Metamorph*'s and CT18's fitting coefficients for the up valence modulator . . 61

Table 4.2: *Metamorph*'s fitting coefficients for the down valence modulator 63

Table 4.3: *Metamorph*'s fitting coefficients for the gluon modulator 64

CHAPTER 1

INTRODUCTION

Curiosity is a human attribute which acts as a stimulus for knowledge flourishing and a guiding light for scientific progress. Naturally, ever since early times, humankind has made high endeavors in pursuing an extensive and thorough conception of the physical structure of the Universe, our home, at each conceivable length scale. Being an essential questioning, through history and still to this day, numerous scientists have done the generous duty of dedicating their lifetime research to expand the boundaries of comprehension on this matter. Particularly, thanks to these efforts, a theoretical-experimental physics framework has been instituted with the purpose of exploring the structure of matter at the smallest and fundamental level. This branch is known as *particle physics*. Therefore, in order to acknowledge and assimilate the current status of particle physics and reveal the arguments motivating this thesis, it is worthwhile to chronologically revisit some of the milestones that shaped the subject as we know it today.

The first inquiries on the structure of matter at the lowest scale were conducted by ancient civilizations as a result of their philosophical doctrines and meditations regarding the place taken by living and non-living entities in the natural world. In the Indian civilization, a primal notion of "atom" already existed since the 8th century BCE, presented by the philosopher Uddālaka Āruṇi [3, 4]. Āruṇi stated that objects were composed of small particles, invisible to the eye. Throughout the following centuries, various atomic theories, each with their own traits, were developed by multiple schools of thought. Take, for example, the

atomism presented in the *Vaiśeṣika-sūtra*, an elemental compendium of principles written by the philosopher Kaṇāda (c. 6th $\sim$ 2nd century BCE) in which the characteristics and properties of atoms are detailed, asserting that such objects are simple and indivisible, but being too small to be observed, they have to be perceived by their effects [4]. Although somewhat naive, this inferential approach constitutes the basis for several modern high-energy physics experiments where subatomic lengths/high resolution distances are involved. Kaṇāda also asserts that these invisible atoms are "eternal" as single particles but not as aggregates, *i.e.* combined into compounds [4, 5], for they can be successively split into smaller components; a conjecture to which we can adhere in our search for the ultimate constituents of matter.

On the other hand, the Greek atomism has its roots, at least, in the 6th century BCE, proposed by the Milesian school [6], although its concrete founders were Leucippus and his pupil Democritus. It is a natural philosophy intended to account for the origins of things by means of natural causes and effects, without divine intervention [7]. By addressing the problem of the genesis of material things and the nature of coming into being and change, put forward by earlier schools of thought, Leucippus and Democritus proposed the existence of atoms and their characterization following strong logical principles [8]: shapes, sizes, movement, etc. These atoms cluster together to constitute larger substances, but the variation in their arrangement accounts for the perceivable differences between these aggregates [6, 8]. Together with these properties, atoms are proposed as indivisible, considering the debate on the outcome if material objects are divided *ad infinitum*.

Without deepening into ontological affairs, we must bear in mind that although these dogmas might come across as old-fashioned, setting apart the deeply intertwined ideas lying in the scientific domain and the ones remaining in the philosophical realm, leaves us with a *protoscientific* way of thinking [3] which certainly shaped the modern embodiment of science we rely on to a great extent. In fact, some antique conceptions are now held as fundamental physical postulates; conservation/rearrangement/origin of matter, transmutation, etc. They were subtly manifested throughout these philosophical discourses heavily motivated on a natural philosophy perspective based, to a certain degree, on metaphysical backgrounds.

By the end of the ancient era, scrutiny on the matter that concerns us suffered a significant and prolonged hindering, primarily due to the predominance of Christianity which firmly opposed the atomistic school of thought. During this period, religious communities and philosophical currents took an active stand on atomism. Excepting few notable cases (*e.g.* Islamic atomism), the ones who were in favor did not generally contribute to the development of atomism at the time, but preserve it for a subsequent resurgence which took place centuries later in the course of the Renaissance and Enlightenment periods. Some of the greatest intellectuals of the epoch revitalized the debate on the structure of matter and put atomism [1] *back on the map* with a modern view, *e.g* Pierre Gassendi, Giordano Bruno, Isaac Newton, Robert Boyle and even Galileo Galilei [2], among others [8]. Not only that, but these new manifestations of atomism adopted the mathematical perception of nature strongly put into effect by Newton, Galileo and Kepler themselves.

It was not until several years later that the debate on the structure of matter evolved from a merely speculative subject to a legitimate problem of experimental evidence, once the technology and the apparatus of science at the time were in a mature facet. With the joint efforts of physicists and chemists of the 19th century, the theory of atoms regained not only strength but also a wide acceptance by the scientific community by the end of the century. Some of the landmarks responsible for this are John Dalton's atomic model and the periodic table devised by Dmitri Mendeleev. Still, the work of other scientists (omitted for scope reasons) has to be praised as well, for it paved the way for what was to come: the end of the debate on atoms.

An accurate description of atoms remained unresolved until the early years of the 20th century. Based on arguments from different perspectives; thermodynamics, stoichiometry, electrochemistry, etc., the long-lasting debate on the existence of atoms and whether the underlying theory was the correct way to describe matter seemed to be coming to an end, tipping the balance in favor of those who supported the theory throughout all those years. Nonetheless, such description did not turn out as expected.

[1] And the corpuscular philosophy as well [6].

[2] Yet, some of these scientists had a Christian worldview.

A sudden wave of previously unseen and extraordinary phenomena, mostly in the physics domain, dethroned atoms from the revered status of fundamental constituents they were meant to possess. A few remarkable examples are the cathode-ray experiments conducted by Thomson leading to the discovery of the electron [3] and a subsequent model of a composite atom, or the series of discoveries eventually giving rise to quantum mechanics, whose fathomless impact is irrefutable. However, there is major experimental result disproving atoms as the smallest units of matter, consequence of the Geiger-Marsden experiments serving as inspiration for the scattering experiments detailed further in this thesis. Directed by Ernest Rutherford, the experiment consisted in scattering α-particles incoming from a beam aimed at a thin foil of gold. The observed scattering pattern showed that atoms consisted mostly of empty space and a concentration of positive charge in a nucleus at the center of the atom, around which electrons are located [4]. By studying the hydrogen atom, Rutherford reached further and designated the positive charges in the nucleus as *protons*, appearing in the same number of electrons for the atom to remain neutral. If the gold nucleus consisted only of protons, then roughly half of the atomic mass would be missing. It was determined eventually that the rest of the mass was carried by electrically neutral particles contained in the nucleus as well, which we now know as *neutrons* discovered by James Chadwick in 1932.

The outcome of an extensive dilemma lasting centuries involving remarkable scientific and philosophical figures of all sorts was the proof of atoms not being the last level of description, but complex structures built by electrons, protons and neutrons and still, the study of atoms comes about as a thorough branch of physics in itself, the same circumstance occurring with nuclei. The idea of atoms as indivisible entities still resonates to these days, since the name itself implies it, nonetheless, we are currently aware of the fact that they are not the primary objects once claimed to be. As it happens, neither are the nucleons, *i.e.* protons and neutrons. Yet, we must maintain the convictions which motivated the hunt for the elementary and utmost indivisible constituents of matter in the first place, regardless of the time and space we find ourselves in, for it seems to be a desire deeply enrooted in the human intellect.

[3] As we will detail in the following section, the electron is, in fact, one of the elementary units of matter.

[4] Known as the Rutherford model.

After this brief digression, we may naturally ask ourselves: *where do we stand, then?* The branch of particle physics bloomed significantly throughout the previous century, due to a sophisticated evolution in both experimental and theoretical domains, aided by technological breakthroughs. Thanks to this and to the emergence of quantum physics, scientists have established a systematic approach to explore matter at the smallest scale by means of high-energy scattering experiments. Consequently, in the span of a hundred and a few more years, we have progressed from a description based on atoms to a wider and refined picture of multiple elementary particles being the building blocks of the Universe: the quark scheme, proposed by Murray Gell-Mann and George Zweig, following the ancient tradition of pursuing the elementary units of matter; presumably quarks and leptons. Our understanding on the structure of matter at the current accessible and smallest scale must not be limited to merely identify such particles, but also elucidate the complex mechanism by which these constituents interact. In response, the Standard Model offers a description of the fundamental forces of nature in terms of exchanges of field quanta between matter constituents. Particularly, our focus is on the strong force, disclosed by Quantum Chromodynamics, responsible for the arrangement of quarks forming composite particles. Accessing the content of such particles, that is, detecting individual quarks is not quite possible. In consequence, an experimental method and a subsequent framework can be conceived to treat quarks as locally "free": Deep Inelastic Scattering (DIS) and the *quark-parton model*, respectively.

As will be further detailed, DIS experiments are the most efficient tool within our reach to explore the quark-composite particles (collectively called *hadrons*), deriving in a series of key results and observations eventually encompassed by the *quark-parton model* devised by Richard Feynman, putting us on track towards our object of study: **Parton Distribution Functions** (PDFs). In general terms, these functions are mathematical objects intended to parametrize the distribution of quarks (and strong force mediators) within hadrons, although, they cannot be evaluated by the quark-parton and QCD schemes, making the explicit determination of such objects a complex task. The aim of the present work is to introduce an innovative tool to parametrize PDFs in terms of experimental data produced in DIS experiments, named *Metamorph*. The collaborative project in charge of this tool is called *Fantômas4QCD*.

In order to fulfill our purpose, the way of proceeding throughout this thesis will be as follows: first, we will survey some of the theoretical foundations supporting the *Fantômas4QCD* project, namely the Standard Model which provides a picture of matter and fundamental interactions in terms of elemental particles, the strong force responsible for holding quarks together, its subsequent QCD framework and the Deep Inelastic Scattering experiments enabling to explore the inner structure of hadrons. Then, we will examine some relevant formulas from the kinematic analysis of such experiments from which parton distribution functions are obtained and the reasons impeding their exact evaluation, thus leading to the task of extracting information from phenomenology and parametrizing these objects via experimental data. Next, a brief review on the current state of PDF parametrization will be given, emphasizing specifically in CTEQ, a pioneering collaboration in the field. This will be followed by the description of the innovative methodology proposed by the *Fantômas4QCD* project, employing a tool from Computer Aided Geometric Design (CAGD): Bézier curves. The name of the project was inspired by the French fictional character Fantômas and his flying *Citroën DS*; a car designed with the very same tool. Afterwards, we will detail how the *Metamorph* program was created by the *Fantômas Development Team*, followed by the results obtained intending to reproduce the parametrizations from CTEQ's latest set of PDFs. Then, a discussion on the usage of *Metamorph* will be presented, together with a series of guidelines and observations made by the team during the development process. Finally, we will describe the current stage of the *Fantômas4QCD* project and some applications expected to be implemented in the near future.

The results and reflections encompassed in this introduction only embody a small segment of what particle physics is all about. As of today, we find ourselves in an advanced phase of uncovering the underlying mechanisms giving physical structure to matter. In light of recent experimental breakthroughs, the domain of particle physics appears to be progressing at a fast pace towards a long-awaited description of the Universe at an elemental level, putting this branch constantly back on the map. Nonetheless, we have seen that the search for a final theory is far from new. In a way, we are taking part in a duty lasting centuries and yet, we are closer than ever, but there still is a long way to go.

CHAPTER 2

THEORETICAL BACKGROUND

2.1 The Standard Model

All physical events taking place in the Universe must be understood as interactions involving matter and energy. Then, it is indispensable, yet intuitive, to classify thoroughly these entities in terms of objects, or more precisely, particles; discerning between the ones being responsible for such interactions and the ones constituting matter. Accounting for this, the Standard Model establishes that matter is constituted by elementary ½-spin ($\hbar$ units) particles, following Fermi statistics, that is, obeying the Pauli exclusion principle, giving them the name of ***fermions***: six leptons and six quarks, all remaining structureless at the current energy scales [9].

Leptons spin = 1/2		
Flavor	Mass GeV/c^2	Electric charge
ν_e electron neutrino	$<1\times10^{-8}$	0
e electron	0.000511	−1
ν_μ muon neutrino	<0.0002	0
μ muon	0.106	−1
ν_τ tau neutrino	<0.02	0
τ tau	1.7771	−1

Quarks spin = 1/2		
Flavor	Approx. Mass GeV/c^2	Electric charge
u up	0.003	2/3
d down	0.006	−1/3
c charm	1.3	2/3
s strange	0.1	−1/3
t top	175	2/3
b bottom	4.3	−1/3

Table 2.1: Chart of fermions arranged by generation (increasing mass), each containing two types particles. [1]

Leptons, commonly seen as generalizations or replicas of electrons, are comprised of three different flavors of electrically charged particles, each paired by weak force processes with a distinct flavor of neutral lepton, known as neutrinos. At the same time, the group of quarks is made up of six different flavors. Ordinary matter is composed of electrons and stable particles, *i.e.* protons and neutrons, made of the two lightest flavors, that is, up and down quarks. Following quantum rules and conservation laws, the rest of quark flavors can combine constituting composite but unstable particles, decaying into lighter particles with up and down quarks in the same way as heavy leptons decay into electrons; both cases requiring the weak force to intervene.

Although leptons and quarks are both classified as fermions, several attributes set them apart. Leptons mutually interact via the weak and electromagnetic forces. Quarks, on the other hand, interact not only by the weak and electromagnetic forces, but also by the strong force. Furthermore, table 2.1 exhibits leptons having integer electric charges, whereas quarks carry fractional charges, as a consequence of a fundamental property called ***color confinement***; while leptons can be found as free, isolated particles, quarks are restricted to exist in combinations with other quarks and not individually [9, 10]. It is essential to note that these units of matter exist along with their antiparticles (having the exact same mass and lifetime, but opposite electric charge and magnetic moment [10]), in such a way that only two types of combined-quark bound states can be found in nature: *mesons* being unstable, composite particles made of one quark and one antiquark, and *baryons*, being constituted solely by quarks, with the proton as the most stable baryon, followed by the neutron. These two groups of particles are altogether called hadrons [10, 11]. Aside from protons and neutrons, hadrons were generally first detected in cosmic ray experiments, decades before the Standard Model was devised. These “strange” and novel particles did not fit within the present-time model, composed of electrons, protons, neutrons, neutrinos and photons, consequently suggesting a broader scheme [12]. Hadrons are currently a major subject of study in modern physics, and the technological progress achieved in recent decades has enabled us to possess an alternative source for finding such particles: high-energy particle accelerators, such as Stanford’s Linear Accelerator, the LHC, or the now defunct *Tevatron* and HERA.

Conversely, the Standard Model provides a compelling picture regarding the fundamental forces of nature. Preserving the approach of physics Nobel laureate Hideki Yukawa, forces are depicted as *virtual quanta* exchange interactions between massive particles [9, 13]. In accordance with the quantum scheme, discrete "mediators" transmit information about the specific type of force corresponding to the interaction taking place. The source of such field quanta are fermions in table 2.1, and they distinguished by being integer-spin particles (which can be massless), following Bose-Einstein statistics, earning them the name of ***bosons***:

Unified Electroweak spin = 1		
Name	Mass GeV/c^2	Electric charge
γ photon	0	0
W^-	80.4	−1
W^+	80.4	+1
Z^0	91.187	0

Strong (color) spin = 1		
Name	Mass GeV/c^2	Electric charge
g gluon	0	0

Table 2.2: Bosons arranged by interaction [1].

Table 2.2 displays data of field bosons involved in each interaction; photons corresponding to electromagnetism, $W^{\pm}$ and Z^0 present at weak interactions and the gluon, accounting for strong interactions. Electrically neutral bosons are their own antiparticles, *i.e.* photon, Z^0 and the gluon, while W^+ and W^- are mutual antiparticles.

Each force has an effective range, with a relative magnitude compared to the rest. The following scale, given by Griffiths [14] and Perkins [10], is rather ambiguous and should just be taken as an overall outline:

Force	*Strong*	*Electromagnetic*	*Weak*	*Gravity*
Strength	1	$10^{-2} \sim 10^{-3}$	$10^{-7} \sim 10^{-14}$	$10^{-39} \sim 10^{-43}$

Table 2.3: Comparative strength of the fundamental forces at subatomic scales, employing the strong force as reference.

Though gravity has been formulated since Newton's *Principia* in the seventeenth century, and generally theorized by Einstein more than two centuries later, this fundamental force is not considered within the Standard Model, due to two crucial reasons: in contrast to the other forces, it is substantially weaker at the quantum regime and it does not have a thorough quantum field theory being renormalizable, a process later described. Still, a hypothetical massless boson has been proposed: the graviton, with spin 2 and zero electric charge, nonetheless, its direct detection is currently far from possible [15]. Distinct theoretical frameworks have been devised during recent decades, but a conclusive viewpoint is still under development.

Table 2.2 is organized in a way that the strong force is isolated from the two other forces, the reason being that weak and electromagnetic interactions of both leptons and quarks can be described by a single gauge theory where both forces are unified: the electroweak theory, making use of the formalism of *Quantum Electrodynamics* (QED), motivated by observations of both interactions having the same strength at high-energy scales in experiments realized during the 1970s [10]. In contrast, the gauge theory describing the strong interaction between quarks is *Quantum Chromodynamics* (QCD), being the framework supporting this thesis.

Undoubtedly, the Standard Model has been the most successful theory regarding elementary particles giving structure to matter and describing fundamental interactions between them. Despite how elegant this framework is, certain limitations and unsolved difficulties impede the Standard Model from being a final theory; the first of them being the dismissal of gravity due to the arguments previously discussed. Next, this framework does not provide a direct reasoning on why there are exactly three generations of fermions. Finally, it does not account for further, but related, physical problems well beyond the scope of the present thesis and the Standard Model itself, such as the asymmetry of matter and antimatter or the prediction of the so called *dark matter* [10]. Extended frameworks commonly referred as *Physics Beyond the Standard Model* intend to address these issues with alternative approaches. Regardless of these adversities, the Standard Model and its inherent gauge theories have proven multiple times that scientists are on the right path to describe matter and the fundamental interactions at the elemental level.

2.2 Strong Force and QCD

Sufficient evidence has already been introduced supporting the statement that neither atoms nor nuclei are the minimal units of matter. The collection of experimental results acquired throughout the preceding century and recent years have proven the existence of an underlying rich structure within nucleons and other composite particles, ruled by a sophisticated interplay between quarks exchanging gluons due to the strong force; a process detailed by Quantum Chromodynamics. Said interaction is not only responsible for quarks clumping together forming hadrons, but for the residual force between hadrons as well, *e.g.* overpowering the electromagnetic repulsion between nucleons, allowing them to set up nuclei. As shown in table 2.3, the strong force possesses the greatest magnitude within a short range between interacting hadrons.

In order to give a brief QCD outline, we will examine one of the two types of hadrons: baryons, which are bound states of three ½-spin quarks [1]. Two archetypal examples of such are protons, constituted by two up quarks and one down quark (*uud*) and neutrons, having two down quarks and one up quark (*udd*). Alternative quark combinations are possible, namely Δ^{++} with three up quarks (*uuu*) and electric charge of 2 (elementary charge units), or Δ^{-} with three down quarks (*ddd*) and electric charge of -1. However, these two instances appear to be problematic. The exclusion principle forbids two identical fermions from having the same quantum numbers; a circumstance seeming unavoidable, since both Delta baryons are composed of three quarks with the same flavor [11, 12]. This matter is solved by endowing quarks with a new quantum number acting as an analog of electric charge for strong interactions: color; a suitable abstract property relating color mixing and particle physics. By doing this, each quark within a hadron possesses a color quantum number in such a way that the hadronic bound state is colorless (white). This condition must be compelled for a hadron to exist; non-neutral color states are prohibited. Quarks have the primary colors: red, green and blue, and antiquarks have the respective complementary colors: cyan, magenta and yellow.

[1] Exotic baryons with a number of quarks other than three have been observed/hypothesized, *e.g. pentaquarks*, though, these objects will not be considered in this thesis.

Therefore, quarks cannot be found as isolated particles since the color neutrality property would be infringed, forcing them to be confined in hadronic bound states. The neutral color hadron configurations are the following: three quarks (red, green, blue) producing baryons, three antiquarks (cyan, magenta, yellow) constituting antibaryons, and quark/antiquark combinations (primary and corresponding complementary colors) forming mesons/antimesons. This constraint prevents hadrons from having non-integer electric charges as well. The color scheme might appear as just being an artifice; this is far from the truth, since it links QCD with group theory, providing a $SU(3)$ color transformation symmetry, maintaining the Lagrangian invariant under local phase transformations and making QCD a non-Abelian [2] gauge theory. Experimental evidence validates $SU(3)$ as the underlying structure of QCD [16].

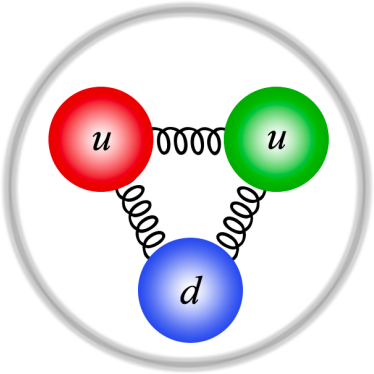


Figure 2.1: Diagram depicting confined colored quarks within a neutral color proton, interacting via gluon exchange.

Once the required framework has been established, it is appropriate to steer back towards the experimental domain, since much of the theoretical work was motivated to account for the observations made at particle accelerators during the 1960s and following years [17]. Similar to Rutherford's experiment where the angular distribution of deflected α-particles off a nucleus implied the presence of scattering centers of positive charge within nuclei, by a procedure detailed in section 2.3.1, Deep Inelastic Scattering (DIS) experiments at SLAC revealed not only that protons have internal scattering centers, but also that such objects are structureless [3]. These particles were encompassed in Murray Gell-Mann's and George Zweig's proposal: the quark scheme. It is persuasive to proceed embracing the quark framework readily, nonetheless, this hypothesis will be momentarily omitted, until the parton model and the quark picture are reacquainted in section 2.3.3. The chronological formalism presented by Aitchison and Hey [9], Halzen and Martin [11] and Close [17] will be followed.

[2] Gluons also carry color, enabling them to interact with other gluons and self-couple.

[3] At least at the probing energy scales currently accessible [9].

2.3 Scattering and Hadronic Structure

In view of the fact that color confinement forbids having access to individual quarks, experimental techniques must be devised to acquire an insight on the structure of hadrons. Scattering beams of leptons (electrons, commonly) off nuclear targets is the most effective procedure to achieve so. As stated in section 2.1, hadrons are not units of matter, implying that the process under discussion consists in scattering an electron off a charge distribution with internal structure. The scale of description reached with these experiments depends on the momentum carried by both the incident electron and the exchanged virtual photon probing the target. With the aim of arriving at parton distribution functions [4], the lepton-hadron scattering process will be described and relevant formulas will be outlined.

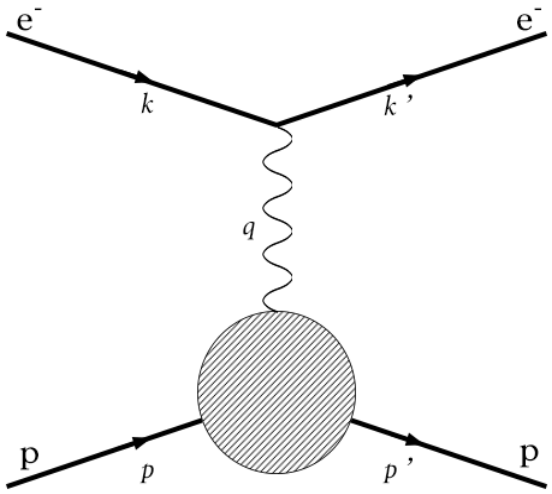


Figure 2.2: Single-photon $e^{-}p$ elastic scattering. [5]

Consider an incident electron with initial 4-momentum k interacting at the lowest order, that is, via single-photon exchange with a proton carrying initial 4-momentum p and rest mass M [9]. The virtual photon acts as a probe to explore the proton's structure, with resolving power wavelength [18]: $1/Q$ where Q is the 4-momentum transferred to the proton, as defined in equation (2.1).

[4] Still undefined to this point but they are the main concern of this thesis.

[5] Figure based on [9] and [11].

After interacting, the electron is deflected with final 4-momentum k' at an angle θ (between k and k' 4-momenta, measured in the laboratory frame) whereas the proton carries final 4-momentum p'. In agreement with Halzen and Martin [11], for this process, the only available scalar variable at the proton vertex is:

$$\textbf{Squared 4-momentum transfer: } q^2 = (p - p')^2 = -Q^2 < 0 \tag{2.1}$$

The invariant transition amplitude of the Feynman diagram in figure 2.2 is obtained by enhancing the scattering theory, implementing elements of quantum field theory together with a perturbative approach [19]. This quantity describes the transition probability between the initial and final states of the scattering process. Computing the transition amplitude leads to the calculation of the differential cross section which describes the number of scattered particles coming out in a solid angle Ω. Consequently, the electron-proton scattering differential cross section (σ) per unit solid angle (Ω) is given as the product of two terms:

$$\left(\frac{d\sigma}{d\Omega}\right) = \left(\frac{d\sigma}{d\Omega}\right)_{\text{P-L}} \times \ |F(q^2)|^2 \tag{2.2}$$

where the first term on the RHS corresponds to the scattering cross section of an electron with a point-like target (P-L). The second term corresponds to a form factor (F), accounting for the hadronic extended structure of the proton, dictating how this particle deviates from a point-like behavior. Perkins intuitively describes the form factor as follows: "$|F(q^2)|^2$ measures the probability that the nucleon 'holds together' and recoils under the impact q." [10]. The form factor is depicted as a 'blob' in figure 2.2 and it is a q^2-dependent function following two principles: Lorentz and gauge invariance. Form factors are determined experimentally by measuring the scattering cross section and modifying known theoretical cross sections to match the observed results. The final expression of equation (2.2) must portray the proton as a charge distribution and the magnetic moment of the proton taking part in the interaction. Additionally, the proton is not assumed to be infinitely massive; in the regime of low momentum transfers ($q^2 \to 0$), the interaction causes the hadron to recoil or resonate while retaining its unit structure. This implies that the final hadronic state is known, identifying this example as an exclusive process [10].

Changing the Ω-dependence to Q^2, the explicit expression of equation (2.2) in the laboratory frame is:

$$\frac{d\sigma}{d(Q^2)}\Big|_{\text{Lab}} = \frac{\alpha^2}{4k^2\sin^4(\theta/2)}\frac{\pi}{kk'}\left[\frac{G_E^2+\tau G_M^2}{1+\tau}\cos^2\left(\frac{\theta}{2}\right)+2\tau G_M^2\sin^2\left(\frac{\theta}{2}\right)\right] \tag{2.3}$$

being an alternative expression of the Rosenbluth formula, with α as the fine structure constant, $\tau = Q^2/4M^2$ and G_E, G_M being Q^2-dependent form factors related to the proton's charge and magnetic moment, respectively. The factor outside the brackets is related to Rutherford scattering, *i.e.* an electrically charged particle interacting with a fixed Coulomb potential. Considering the factor $\cos^2\left(\frac{\theta}{2}\right)$ alone inside the brackets of equation 2.3 accounts for the scattering of electrons from spinless point-like particles in the relativistic regime [9]. The second factor inside the brackets accounts for the hadron's ½-spin nature.

2.3.1 Deep Inelastic Scattering

Similarly, Deep Inelastic Scattering (DIS) consists of a beam of incoming leptons scattering off a target hadron with the condition of being a hard process, that is, large energy loss and momentum transfers (q^2) from the incident electron are required, in order to achieve an enhanced spatial resolution with the probing photon.

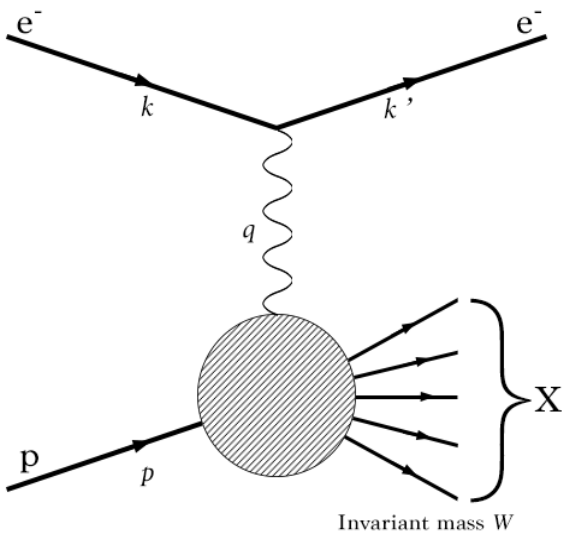


Figure 2.3: DIS as an inclusive process. [6]

[6] Figure based on [9] and [20]. In this regime, the probability of elastic scattering is substantially low.

DIS represents the most straightforward way of exploring the internal structure of hadrons, analogously to Rutherford scattering probing the nuclear structure within atoms. Upon reaching such levels of momentum transfer, elastic conditions are no longer met, since the hadron extracts some kinetic energy from the electron. Consequently, the hadron does not behave as a point-like structure anymore, but as a composite system instead: the probing virtual photon breaks up the hadron and interacts quasi-freely with the individual point-like constituents, revealing an internal structure. In contrast to elastic scattering, there are two independent variables in this process; q^2 4-momentum transfer as introduced in equation (2.1) and:

$$\textbf{Rest frame energy transfer: } \nu = (p \cdot q)/M = E - E' \tag{2.4}$$

Assuming single-photon exchange, this interaction is depicted in figure 2.3 where the outgoing electron is detected by measuring its scattering angle and kinetic energy, since both variables are now independent. On the other hand, the initial proton identity is not preserved; X is an unrestricted hadronic system consisting of the sum of all possible hadronic final states, making DIS an inclusive process. Therefore, the inclusive cross section is "a sum over the cross sections for all the possible hadronic states" [9]. The electron mass is neglected and the effective mass of the hadronic system is given by:

$$W^2 = p'^2 = (p+q)^2 = M^2 + 2M\nu + q^2$$

In both elastic and DIS cases, the differential cross section is proportional to the contraction between a leptonic tensor $L_{\mu\nu}$ and a hadronic tensor $W^{\mu\nu}$, respectively associated to the electron and proton vertices in the DIS Feynman diagram of figure 2.3. The latter serves as a parametrization of the unknown form of the current at the hadronic term of the propagator for the invariant amplitude [11]. It also accounts for scattering from the electric charge and magnetic moment of the target hadron. In DIS, $W^{\mu\nu}$ must enforce a momentum conservation restriction for the sum of final states X, using both independent variables: ν and Q^2. It is parametrized via Lorentz invariance and current conservation arguments.

The explicit forms of the leptonic and hadronic tensors are given by [9]:

$$\begin{cases} L_{\mu\nu} = 2\left(k'_\mu k_\nu + k'_\nu k_\mu + \dfrac{q^2}{2}\, g_{\mu\nu}\right) \\ W^{\mu\nu} = \left(-g^{\mu\nu} + \dfrac{q^\mu q^\nu}{q^2}\right) W_1(\nu, Q^2) + \dfrac{1}{M^2}\left(p^\mu - \dfrac{p\cdot q}{q^2}\, q^\mu\right)\left(p^\nu - \dfrac{p\cdot q}{q^2}\, q^\nu\right) W_2(\nu, Q^2) \end{cases}$$

With suitable changes of variable, the corresponding unpolarized (averaged over the lepton and hadron spins) differential cross section is given by:

$$\left.\frac{d^2\sigma}{dQ^2 d\nu}\right|_{\text{Lab}} = \frac{\alpha^2}{4k^2\sin^4(\theta/2)}\frac{\pi}{kk'}\left[W_2(\nu, Q^2)\cos^2\left(\frac{\theta}{2}\right) + 2W_1(\nu, Q^2)\sin^2\left(\frac{\theta}{2}\right)\right] \tag{2.5}$$

where W_1 and W_2 are structure functions replacing form factors from the elastic regime, depending on the two DIS variables (2.1) and (2.4). These functions are derived from a hadronic tensor accounting for strong interactions. For the same reason, it is reasonable to think that these structure functions are complex objects, yet, James D. Bjorken rather predicted a simple behavior for these functions at high Q^2 levels, as described in section 2.3.2.

It remains yet to be detailed, but it will be assumed in advance that the point-like constituents within the proton involved in DIS are Dirac ½-spin particles and the incident electron scatters elastically off them. Therefore, it is natural to construct a cross section similar to that of electron-point-like fermion scattering, where the i^{th} constituent has charge e_i and mass m_i:

$$\left.\frac{d^2\sigma^i}{dQ^2 d\nu}\right|_{\text{Lab}} = \frac{\alpha^2}{4k^2\sin^4(\theta/2)}\frac{\pi}{kk'}\left[e_i^2\cos^2\left(\frac{\theta}{2}\right) + 2e_i^2\frac{Q^2}{4m_i^2}\sin^2\left(\frac{\theta}{2}\right)\right]\delta\left(\nu - \frac{Q^2}{2m_i}\right) \tag{2.6}$$

Because hadrons contain multiple constituent particles, direct comparison with DIS cross section in (2.5) and (2.6), shows that structure functions corresponding to the i^{th} constituent are [9]:

$$\begin{cases} W_1^i = e_i^2\dfrac{Q^2}{4m_i^2}\delta\left(\nu - \dfrac{Q^2}{2m_i}\right) & (2.7) \\ W_2^i = e_i^2\delta\left(\nu - \dfrac{Q^2}{2m_i}\right) & (2.8) \end{cases}$$

Although structure functions represent scattering of an electron off the fundamental particles constituting a proton, these functions cannot be directly implemented. Structure functions must be weighed by functions accounting for different contributions the constituents may have to the hadronic bound state, bringing us closer to parton distribution functions.

2.3.2 Bjorken Scaling

As previously stated, DIS occurs in the regime of high-Q^2 and high-ν, nonetheless, when both variables are progressively increased towards Bjorken's limit ($Q^2, \nu \to \infty$), a particular property is observed: W_1 and W_2 structure functions within the DIS cross section (equation (2.5)) scale to single variable functions as follows:

$$\begin{cases} MW_1(\nu, Q^2) \to F_1(x) \\ \nu W_2(\nu, Q^2) \to F_2(x) \end{cases} \tag{2.9}$$

Both $F_1(x)$ and $F_2(x)$ are finite and Q^2 scale invariant. Bjorken remarkably predicted this precise behavior on the structure functions based on Gell-Mann's quark model and DIS data originated at SLAC [9, 21]. Consequently, the scaling phenomenon was named after him. It is now imperative to give the definition of the scaling variable before proceeding. While Q^2 and ν tend to infinity in Bjorken's limit, their ratio remains finite. By employing the two DIS independent variables, Bjorken's dimensionless scaling variable is defined as:

$$x \equiv \frac{Q^2}{2M\nu} \tag{2.10}$$

The x variable [7] corresponds to the fraction of the hadron's total momentum momentum carried away by an individual point-like hadron constituent. This interpretation relies strongly on the hypothesis that, with sufficient resolution power, the exchanged virtual photon interacts with point-like scattering centers behaving as structureless Dirac particles which constitute the proton. The interaction takes place at short distance and time scales, causing the impacted constituents to appear being free, *i.e.* not interacting via strong force. This, in fact, brings us back to an elastic scattering regime within the proton.

[7] This variable will be of great significance throughout the following chapters.

2.3.3 The Quark-Parton Model

Throughout the preceding DIS kinematic analysis, various presuppositions were made. Richard Feynman encompassed these assumptions in a single framework based on Bjorken scaling: the *quark-parton model*, which justifies the single exchange virtual photon interaction at a short time and distance scale, making point-like hadron constituents to behave as free. These constituents were given the name of ***partons*** by Feynman [22] and they move collectively in a direction parallel to that of the proton, each carrying an x-fraction of the proton's momentum. Naturally, an inevitable question arises along with this proposal: physically speaking, what exactly are partons?

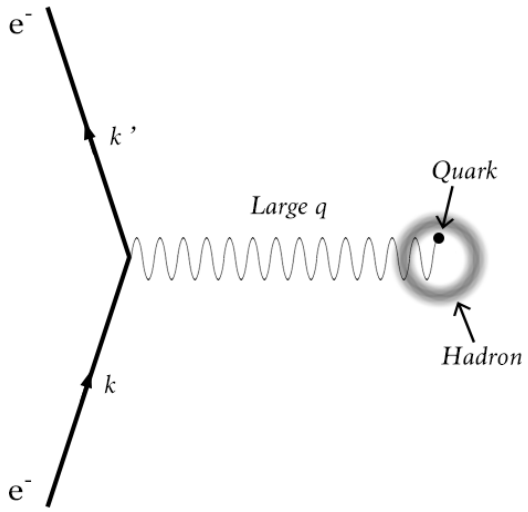


Figure 2.4: Short wavelength virtual photon resolving the quark structure within a hadron. [8]

Before presenting equation (2.6), we assumed that ½-spin particles constitute the proton (and hadrons in general). It would be futile to introduce the Standard Model in section 2.1 without establishing an explicit link between DIS phenomena and Gell-Mann's quark picture. With extensive experimental evidence [9] it is confirmed that, in fact, partons are fermions having the same quantum numbers as quarks. Since antiquarks are also fermions, they are included in the quark-parton model as well. Then, hadrons are made of quarks and antiquarks

[8] Figure based on [11].

[9] The following texts give brief treatments on data supporting this assertion: [9, 10, 12, 17, 23].

with individual flavors and quantum numbers. Recalling from section 2.1, distinct parton combinations produce distinct hadrons: baryons, having multiples of ½-spin, are built with three quarks while mesons, having integer-spin, are built with quarks and antiquarks. In the particular case under analysis, protons are baryons composed by two up quarks and one down quark (*uud*), whose electric charges add up to the elementary electric charge. The collection of quarks and antiquarks dictating hadrons' quantum numbers are also known as *valence quarks*. However, valence quarks given by the quark-parton model do not reconstruct faithfully the DIS picture by themselves.

2.3.4 Parton Distribution Functions and Non-Perturbative QCD

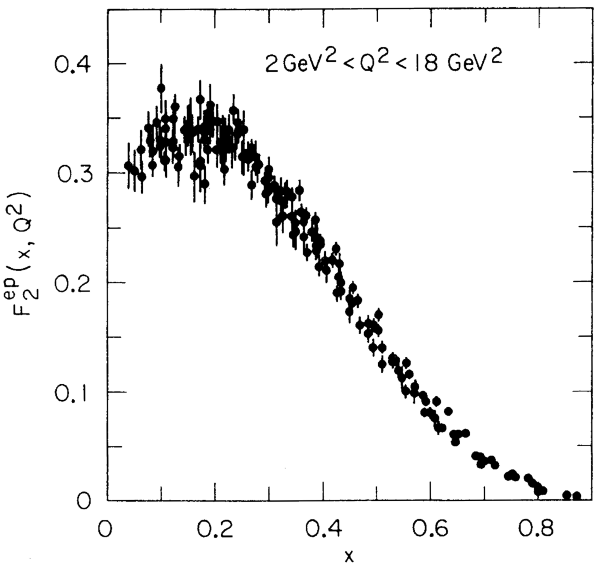


Figure 2.5: Experimental data of proton's $F_2(x)$ at different Q^2 scales. [10]

The quark-parton model serves as a naive, but intuitive approach detailing the internal structure of hadrons. Despite possessing a simple description of parton "arrangement" within hadrons, sensible care must be taken regarding the approximations considered in the previous sections. As stated in section 2.3.2, specific conditions are assumed for DIS to occur: at Bjorken's limit, according to the quark-parton model, quarks behave "freely", *i.e.* mutual strong interactions are ignored. Similarly, final-state interactions are assumed to occur

[10] Figure extracted from [24].

at much larger distances and time scales, so they can be ignored as well. Of course, these circumstances do not fully embody how quarks interact, since they are not free particles, but objects forced to be confined within hadrons. As a consequence, QCD must surface once again to describe how quarks interact by means of strong force.

An evidence of this is that, if the quark-parton model holds true, then, for the $F_2(x)$ proton structure function (equation (2.9)) a peak would be expected for $x \approx 1/3$, implying a roughly uniform distribution of the hadron's total momentum between the three valence quarks, but DIS experiments reveal a different behavior, depicted in figure 2.5, reflecting that quarks contain not only valence quarks but also a "sea" of virtual quark-antiquark pairs where valence quarks exist, produced by gluons. The desired behavior shows up if only valence quarks are accounted for, which is observed when plotting the subtraction of the proton and neutron F_2 structure functions, canceling sea quark contributions: $F_2^p(x) - F_2^n(x)$

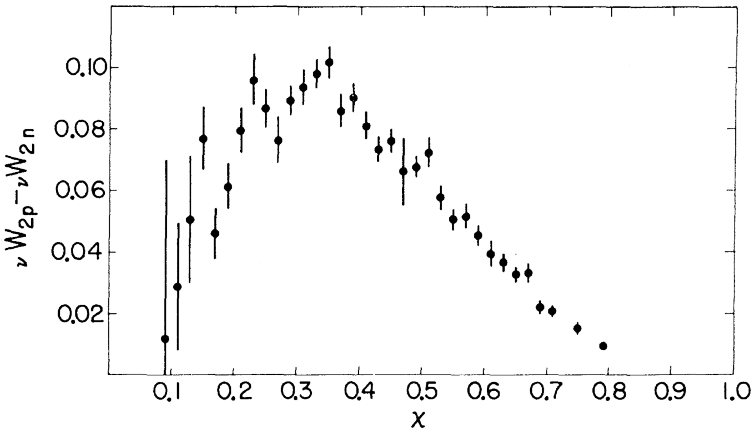


Figure 2.6: Valence quarks $F_2(x)$ structure function.
The peak is located near $x \approx 0.35$. [11]

The quark-parton model is then generalized to include valence quarks, gluons and sea quarks, all carrying an individual fraction of the hadron's momentum. Once the parton scheme has been fully situated, the main matter of this thesis can be ultimately defined. Understanding that partons contribute differently to the hadron, it is necessary to introduce a function reckoning the contribution of each parton in order to gain a complete description.

[11] Figure extracted from [25].

Following Bjorken scaling, **Parton Distribution Functions** (PDFs) are defined as *functions describing the probability of a hadron containing an i-parton carrying an x-fraction of the hadron's momentum p, at the energy scale Q^2* [11]:

$$f_i(x) = \frac{dP_i}{dx} \tag{2.11}$$

where the i-index denotes the distinct parton types within the hadron. Being probability functions, PDFs serve as weights tempering individual parton contributions within the variables shown in section 2.3.1 at each x-value, similar to statistical probability density functions. Accounting all parton contributions for an individual hadron, parton distribution functions must satisfy a normalization condition [12]:

$$\sum_i \int_0^1 x\, f_i(x)\, dx = 1$$

This condition is known as the ***momentum sum rule***, indicating that momenta of all partons must add up to the hadron's total momentum. Once PDFs are defined, the overall ν and Q^2 dependent W_j structure functions consist of the weighted sum of the corresponding i^{th}-parton W_j^i structure functions in equations (2.7) and (2.8). Expressing the i^{th}-parton mass as an overly simplified approximation ($m_i = xM$) gives:

$$\begin{cases} W_1(\nu, Q^2) = \sum_i \int_0^1 f_i(x')\, W_1^i\, dx' = \sum_i \int_0^1 \frac{Q^2}{4M^2x'^2}\, f_i(x')\, e_i^2\, \delta\left(\nu - \frac{Q^2}{2Mx'}\right) dx' & (2.12) \\ W_2(\nu, Q^2) = \sum_i \int_0^1 f_i(x')\, W_2^i\, dx' = \sum_i \int_0^1 f_i(x')\, e_i^2\, \delta\left(\nu - \frac{Q^2}{2Mx'}\right) dx' & (2.13) \end{cases}$$

If we employ the Dirac delta function composition identity,

$$\delta\left(g(x)\right) = \frac{\delta(x - x_0)}{|g'(x_0)|}$$

where x_0 satisfies $g(x_0) = 0$, factors in equations (2.12) and (2.13) transform into:

$$\delta\left(\nu - \frac{Q^2}{2Mx'}\right) = \frac{Q^2}{2M\nu}\delta\left(x' - \frac{Q^2}{2M\nu}\right) = \frac{x}{\nu}\,\delta\left(x' - \frac{Q^2}{2M\nu}\right) = \frac{x}{\nu}\,\delta\left(x' - x\right)$$

[12] It is the first moment of the Mellin integral transform.

The previous result allows to simplify $W_1(\nu, Q^2)$ and $W_2(\nu, Q^2)$ into single variable functions by computing the integrals:

$$\begin{cases} W_1(x) = \sum_i \frac{x}{\nu} \int_0^1 \frac{Q^2}{4M^2x'^2}\, f_i(x')\, e_i^2\, \delta\,(x'-x)\, dx' = \frac{1}{2M} \sum_i f_i(x)\, e_i^2 & (2.14) \\ W_2(x) = \sum_i \frac{x}{\nu} \int_0^1 f_i(x')\, e_i^2\, \delta\,(x'-x)\, dx' = \sum_i \frac{x}{\nu}\, f_i(x)\, e_i^2 & (2.15) \end{cases}$$

We then proceed to define scaling functions $F_j(x)$:

$$F_1(x) \equiv MW_1(\nu, Q^2) = \frac{1}{2} \sum_i f_i(x)\, e_i^2 \tag{2.16}$$

$$F_2(x) \equiv \nu W_2(\nu, Q^2) = \sum_i x f_i(x)\, e_i^2 \tag{2.17}$$

being the exact same functions from section 2.3.2. From equations (2.16) and (2.17), it is shown that scaling functions satisfy the property:

$$F_1(x) = \frac{1}{2x} F_2(x)$$

known as the Callan-Gross relation, which is a consequence of the fact that quarks are ½-spin particles [16, 17]. Revisiting the proton example, the corresponding structure functions are:

$$\begin{cases} F_1^p(x) = \frac{1}{2} \left\{ \frac{4}{9} \left[u(x) + \bar{u}(x)\right] + \frac{1}{9} \left[d(x) + \bar{d}(x) + s(x) + \bar{s}(x)\right] + \ldots \right\} & (2.18) \\ F_2^p(x) = x \left\{ \frac{4}{9} \left[u(x) + \bar{u}(x)\right] + \frac{1}{9} \left[d(x) + \bar{d}(x) + s(x) + \bar{s}(x)\right] + \ldots \right\} & (2.19) \end{cases}$$

where $u(x)$, $\bar{u}(x)$, $d(x)$, $\bar{d}(x)$, $s(x)$, $\bar{s}(x)$ denote up, down and strange quarks and antiquarks parton distribution functions; a notation that will be subsequently followed when referring to a PDF of a specific parton. The ellipsis in equations (2.18) and (2.19) denote possible contributions from additional partons.

Meanwhile, parton distribution functions have to comply with physical constraints. Let $f_i(x)$ denote a parton distribution function:

- **Positivity:** PDFs can be restricted with bounds, guaranteeing a positive cross section.

$$f_i(x) \geq 0$$

- **Domain:** provided that x represents the parton's fractional momentum (equation (2.10)), then $f_i(x)$ must be defined for $x \in [0, 1]$.

- **Valence sum rule:** integrals of contributing valence quarks PDFs must give the quantum numbers. Particularly, for protons:

$$\int_0^1 u_v(x)\, dx = \int_0^1 [u(x) - \bar{u}(x)]\, dx = 2$$
$$\int_0^1 d_v(x)\, dx = \int_0^1 [d(x) - \bar{d}(x)]\, dx = 1$$
$$\int_0^1 s_v(x)\, dx = \int_0^1 [s(x) - \bar{s}(x)]\, dx = 0$$

  Therefore:

$$\int_0^1 [u(x) - \bar{u}(x)] - [d(x) - \bar{d}(x)]\, dx = 1$$

- **Momentum sum rule:** momenta of all contributing quarks, antiquarks, sea quarks and gluons must add up to the total momentum:

$$\sum_i \int_0^1 x\, f_i(x)\, dx = 1$$

To satisfy both valence and momentum sum rules, PDFs must be integrable. This, together with additional theoretical arguments, leads to restrictions on the behavior of PDFs:

- **Small-x behavior:** in the region of $x \to 0$, sea-quark contributions domain. $f_i(x)$ must fall as a power law:

$$\lim_{x \to 0} f(x) \longrightarrow Ax^a$$

- **Large-x behavior:** known as the elastic limit, in the region of $x \to 1$ where valence quarks predominate by carrying most of the proton's the momentum, valence PDFs must fall as a power law:

$$\lim_{x \to 1} f(x) \longrightarrow B(1 - x)^b$$

Allowed values of a and b are such that the integrability condition is satisfied.

Before moving forward, it is timely to summarize briefly the situation where we currently stand: DIS enables exploring the underlying structure of hadrons in the most direct way. Additionally, the quark-parton model characterizes this structure in terms of the fundamental particles described by the Standard Model. On the other hand, the kinematic analysis provides formulas that have been well verified with DIS experiments and the frameworks mentioned before are proven to be successful. Simultaneously, parton distribution functions are objects of great relevance, since they are intimately related to the inner structure of hadrons, revealing how partons are "distributed" within. It appears that we are on the right track to describe matter at the fundamental level, nonetheless, there still is one major concern which has not been discussed yet: ***how are parton distribution functions determined?*** It occurs that parton distribution functions cannot be evaluated by the theory derived from the frameworks that we have presented so far.

To completely address this problem and search for viable solutions, we have to briefly detour towards QCD's domain once more. As stated in section 2.2, Quantum Chromodynamics is a non-Abelian gauge field theory, whose associated Lagrangian remains invariant under local gauge transformations of the color $SU(3)$ group. The QCD Lagrangian is given by [16]:

$$\mathcal{L}_{\text{Classical}} = \sum_{\text{Flavors}} \bar{q}_j(x)(i\not{D} - m)_{jk}\, q_k(x) - \frac{1}{4} F^A_{\alpha\beta} F^{\alpha\beta}_A \tag{2.20}$$

where $\bar{q}(x)$ and $q(x)$ are quark and antiquark spacetime dependent fields, m is the quark mass, $\not{D} = \gamma_\mu D^\mu$ is the gauge covariant derivative and $F^A_{\alpha\beta}$ is the field strength tensor assembled with the gluon field $\mathcal{A}^A_\alpha$:

$$F^A_{\alpha\beta} = \partial_\alpha \mathcal{A}^A_\beta - \partial_\beta \mathcal{A}^A_\alpha - g f^{ABC} \mathcal{A}^B_\alpha \mathcal{A}^C_\beta$$

Focusing on the current issue, we will only analyze the term g in the last factor on the RHS; the gauge coupling constant determining the strength of interaction between quarks, antiquarks and gluons with color, being the charge of QCD. In quantum field theories, divergent integrals of 4-momenta appear frequently when observable quantities are calculated. The procedure dealing with such integrals is called *renormalization*. Renormalization allows expressing formulas with physical parameters, rather than with the original terms [9].

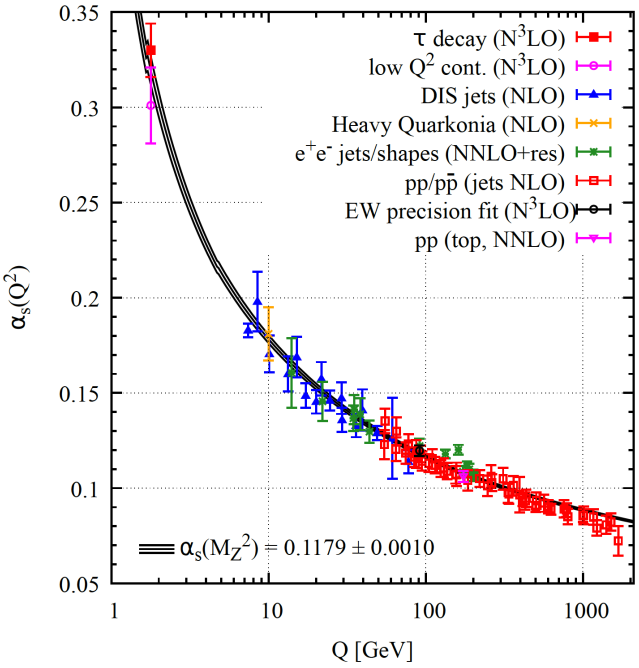


Figure 2.7: Sets of experimental measurements of the running coupling constant. [13]

After renormalization, the coupling constant in QCD becomes free of divergences for the price of becoming scale dependent instead, that is, Q-dependent [16, 18]. As an alternative to the coupling constant g, the running coupling constant is rather used, defined as:

$$\alpha_s = \frac{g^2}{4\pi}$$

being also a function of Q, acting as an effective parameter analogous to effective charges in nuclear physics. The plot in figure 2.7 depicts how the running coupling constant behaves in terms of Q. As Q-energy levels progressively increase (implying shorter length and time scales), the strong coupling between partons becomes weaker, in contrast to electromagnetic interactions becoming weaker at larger distances instead; a behavior called **asymptotic freedom**, present in large momentum transfer processes, corresponding to the smooth and monotonic decreasing region in the plot of figure 2.7. Asymptotic freedom was discovered in 1973 by physicists David Gross and Frank Wilczek while working at Princeton and David Politzer who was a PhD student as well, at Harvard. Thanks to their contributions in QCD, they were awarded with the Nobel prize in physics in 2004 [26].

[13] Figure extracted from [1].

Taking advantage of this behavior and employing the only tool within our current reach, perturbation theory, power series in α_s are deployed to perform necessary calculations. This scheme constitutes what is known as *Perturbative QCD*, whereas the regime of strong coupling constant (low-Q) is described by *Non-Perturbative QCD*. There is a major issue: the hadronic structure belongs to regions of low-Q where power series in the coupling constant α_s do not converge, whereas perturbative QCD domains in regions of high-Q momentum transfer, implying that structure functions and parton distribution functions cannot be predicted with a perturbative approach. It is not fully understood yet how the QCD Lagrangian in equation (2.20) incorporates information regarding the hadronic bound state describing the binding of partons within the hadron.

In conclusion, the explicit x-dependence of parton distribution functions is not known and the formalism of perturbative QCD cannot be employed for this matter. Nonetheless, this framework can still be suitable, since once PDFs are determined independently, perturbative QCD gives the evolution of these functions at distinct Q^2 squared energy orders, by means of the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations in the perturbative regime: $\alpha_s(Q^2) << 1$ [27]. The arguments just exhibited should convince the reader that a fundamental task must be carried out: determination of parton distribution functions through parametrization.

CHAPTER 3

METHODOLOGY

The preceding chapter served for two main purposes: set the grounds for parton distribution functions and also present the theoretical barriers that impede a direct evaluation of PDFs from first principles coming from the models examined until now. Still, given the relevance of parton distribution functions, a method of accessing these objects must be devised, giving rise to PDF parametrization, which is the general objective of the present thesis.

PDF parametrization must be understood as the elemental procedure of acquiring analytical expressions of parton distribution functions from sets of finite data using global QCD analysis [28, 29]. Fits are produced with cross section data produced at high-energy scattering experiments with different x-values and Q energy levels, taking place in particle accelerators such as the *Hadron-Electron Ring Accelerator*, *SLAC National Accelerator Laboratory* or the *Large Hadron Collider*, among others.

3.1 State of the Art: PDF Parametrization

Dating from the early 1980s [29], PDF parametrization is not a novel duty, however, the sustained improvements attained in the experimental and computational fields keep this subject constantly renewed. In addition to this, PDF parametrization is further diversified by evolving these functions at different approximation orders of the DGLAP equations in terms of powers of the running coupling constant $\alpha_s(Q^2)$ [27]; NLO (Next-to-Leading-Order) and

NNLO (Next-to-Next-to-Leading-Order) are the most widely used orders. At the same time, PDF parametrization is conducted by different groups around the world. Some collaborations endeavoring on this task are: ABM, HERAPDF, CT, MSHT and NNPDF [29], each of them following a different procedure and uncertainty assessment. Going into details, there are two main fitting methods currently, contemplating QCD constraints and predictions:

- An analytical functional form is proposed. The objective is to find a the best fit by determining the set of parameters which globally minimize some version of the statistical goodness-of-fit χ^2 function, comparing the produced model with experimental data and addressing additional possible sources of uncertainty [29].

- PDFs are proposed as the product of a fixed and a free function. The latter is determined with artificial neural networks as interpolating tools and Monte Carlo methods. Only continuity is assumed in order to avoid biases from full functional forms, yielding an arbitrarily large number of parameters. An error is minimized as well [30].

Since PDFs cannot be explicitly derived from theory, along with parametrizations, the associated uncertainties have to be determined and assessed as well. The CT collaboration has classified PDF's uncertainties with the following contributions: experimental, theoretical, parametrization and methodological [29]. Each method has its strengths and weaknesses, *e.g.* the analytical method possesses the disadvantage of adding a source of bias by proposing a specific functional form, which might be restrictive, while the second one does not. At the same time, the neural network approach has a cost drawback for being computationally expensive. The analytical method is rather uncomplicated in this aspect, offering simpler algorithms. We will focus on this approach.

3.1.1 CTEQ and CT18

PDF parametrization is accomplished thanks to collaborative efforts made by research teams from all over the world; the *Coordinate Theoretical-Experimental Project on QCD* (CTEQ) is one of them. Founded in 1990, CTEQ is currently comprised of theoretical and experimental scientists from twenty three universities and six national laboratories.

CTEQ has a wide program regarding Quantum Chromodynamics, its issues and implications. PDF fitting has been in this project's agenda for the last three decades, with improving results as more experimental data becomes available. Naturally, CTEQ has become one of the pioneering projects in this field. Furthermore, CTEQ produces sets of PDFs periodically by analyzing experimental DIS data; the latest set of PDFs being CT18. CT (CTEQ-Tung et. Al) is a branch of CTEQ, established by Professor and founding member Wu-Ki Tung [31].

In CT18, a few hundred different parametrization forms were tested [29], keeping two principal goals in mind: functional forms must fit experimental data without overfitting and the uncertainty associated to the parametrization of choice must be understood [2]. CT18's general PDF functional form is as follows:

$$f_i(x, Q_0) = a_0 x^{a_1 - 1}(1 - x)^{a_2} \times P_i(y; a_3, a_4, \ldots) \tag{3.1}$$

where a_i are free parameters to be determined, P_i is a function identified as *modulator*, constituted by a sum of Bernstein polynomials which depend on $y = g(x)$, an argument known as scaling function. CT18 quantifies uncertainties with the goodness-of-fit functions, given by:

$$\chi^2(a, \lambda) = \sum_{k=1}^{N_p} \frac{1}{s_k^2} \left(D_k - T_k(a) - \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha \beta_{k\alpha} \right)^2 + \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha^2 \tag{3.2}$$

with N_p the number of experimental data values, s_k the total uncorrelated error comprised of uncorrelated statistical and systematic errors, on the central measurement D_k. T_k is the theoretical value produced by the model, corresponding to the value D_k, depending on the set of free parameters $\{a_1, a_2, \ldots\}$ as shown in equation (3.2). Lastly, each D_k central value can depend on a set of N_λ systematic uncertainties, which may be correlated over the considered data points. Accounting for this, CTEQ evaluates these uncertainties with a nuisance parameter λ_α sampled from a standard normal distribution actually estimated with its change $\lambda_\alpha \beta_{k\alpha}$. Additional treatment on this function can be found in CT18's article [2].

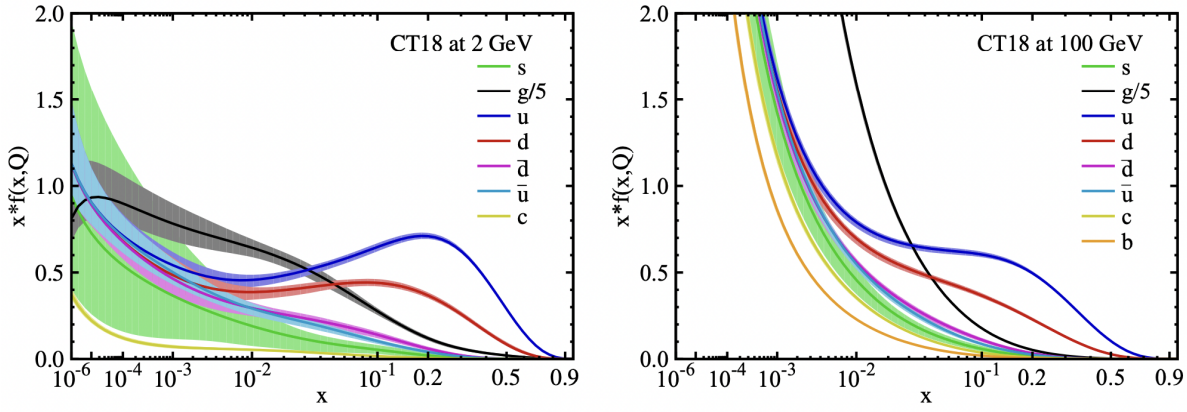


Figure 3.1: Set of CT18's weighed PDFs for different partons at two different energy scales $Q = 2$ GeV and $Q = 100$ GeV, respectively. [1]

Figure (3.1) presents gluon's and different quark flavors' central PDFs, along with respective error bands accounting for PDF uncertainties. These fits correspond to the lowest χ^2 value, considering CTEQ's parametrization in equation (3.1) and they were produced employing NLO and NNLO analysis from HERA and LHC data sets, which are selected by meeting statistical criteria to give promising results.

With the purpose of fitting PDFs, we developed a new program: ***Fantômas4QCD***. In our results, some of CT18 PDFs are reproduced, following a specific procedure detailed in the next section. Consider two PDF examples: up valence quarks ($\boldsymbol{u_v}$) and the gluon ($\boldsymbol{g}$). In CT18's article, a_i free parameters are given for a particular non-perturbative initial energy scale: $Q_0 = 1.3$ GeV.

Best-fit Parameters	a_0	a_1	a_2	a_3	a_4	a_5	a_6
$\boldsymbol{u_v}$	3.385	0.763	3.036	$\sinh(1.502)$	$\sinh(-0.147)$	$\sinh(1.671)$	$1+\frac{a_1}{2}$
$\boldsymbol{g}$	2.690	0.531	3.148	$\sinh(3.032)$	$\sinh(-1.705)$	$1+\frac{2a_1}{3}$	-

Table 3.1: CT18's best-fit parameters for up valence quarks and gluons at $Q_0 = 1.3$ GeV. a_0 for $\boldsymbol{u_v}$ is obtained by normalizing the PDF with the valence sum rule, given in 2.3.4, while the gluon is normalized using benchmark values of the first Mellin moment: $<x>_g$, from previous PDF sets [2].

[1] Figure extracted from [2].

Following the functional form in equation (3.1), coefficients a_1 and a_2 are assumed to be the same for up and down valence quarks, since they dictate the asymptotic behavior in the PDF, as discussed in section 2.3.4. A similar condition is imposed for up, down and strange antiquarks [2]. These assumptions will be preserved in our procedure. At the same time, the scaling function is chosen to be $y = \sqrt{x}$ for both $\boldsymbol{u_v}$ and $\boldsymbol{g}$, due to its flexibility across the whole interval $0 < x < 1$. The parametrized PDF expressions are:

$$\begin{cases} f_{u_v}(x) = \boldsymbol{u_v}(x) = a_0 x^{a_1 - 1}(1-x)^{a_2} \times P^v_{\{a\}}(y) \\ f_g(x) = \boldsymbol{g}(x) = a_0 x^{a_1 - 1}(1-x)^{a_2} \times P^g_{\{a\}}(y) \end{cases} \tag{3.3}$$

with explicit modulation functions:

$$\begin{cases} P^v_{\{a\}}(y) = a_3(1-y)^4 + 4a_4(1-y)^3 y + 6a_5(1-y)^2 y^2 + 4a_6(1-y)y^3 + y^4 \\ P^g_{\{a\}}(y) = a_3(1-y)^3 + 3a_4(1-y)^2 y + 3a_5(1-y)y^2 + y^3 \end{cases} \tag{3.4}$$

Substituting the best-fit parameters in table 3.1 into expressions in (3.3) and (3.4) gives the following plots:

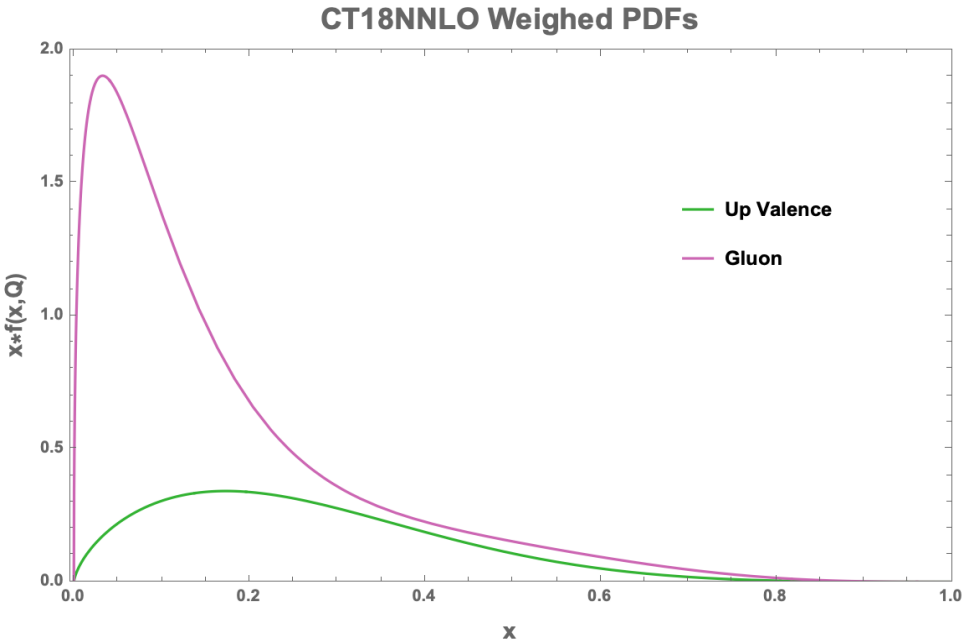


Figure 3.2: CT18's weighed parametrized PDFs at $Q = 1.3$ GeV. [3]

[2] Coefficients of the remaining partons are given in Appendix C of [2].

[3] This plot was reproduced with *Mathematica*, which was fundamental for the development of our program.

3.2 Fantômas4QCD

Fantômas4QCD is a novel project seeking to parametrize parton distribution functions in a different way, following CT's scheme. The *Fantômas4QCD* project consist of a program, named *Metamorph*, made up of a universal *C++* module and a *Mathematica* notebook built by the *Fantômas Development Team*, comprised of physicists from UNAM and Southern Methodist University, lead by Dr. Aurore Courtoy and Dr. Pavel Nadolsky. *Fantômas4QCD* represents an innovative and efficient approach compared to current parametrization methods, such as neural networks, etc. Via Bézier curve fitting, this alternative procedure employs a simple but powerful algorithm offering certain advantages, such as low number of parameters and required computing power.

3.2.1 Why Fantômas?

Fantômas is a famous and renowned French fictional character, or more precisely, *"anti-hero"*, created in 1911, prior to World War I; fruit of the intellects of writers Marcel Allain and Pierre Souvestre. Surrounded by elusiveness, anonymity and enigma, "Fantômas comes to the public as a sinister and shocking incarnation of wickedness" [32], being the material and intellectual author of terrible, but elegant unsolved crimes, earning him various nicknames such as *the Lord of Terror*, *the Genius of Evil*, etc [32]. Fantômas' legacy was born with thirty two novels written by both authors, published between 1911 and 1913. Souvestre died in 1914, but Allain carried on with eleven more novels. Naturally, the amorphous criminal and his exploits became an international source of creation, inspiring several pieces of media: films, comic books, television shows, etc. In particular, created in 1960's México, Fantômas had its own comic book series: "*Fantomas, La Amenaza Elegante*": the ruthless criminal was re-adapted as an avenging, but still elegant hero, helping to spread Fantômas' fame throughout Latin America.

Figure 3.3: Film poster of 1965 Hunebelle's *Fantômas se déchaîne*.

Figure 3.4: Flying *DS* replica.

Back in Europe, in the 1965 film *"Fantômas se déchaîne"* directed by André Hunebelle, the villain is seen making an escape in his flying vehicle; a winged *Citroën DS*. This was a french car produced from 1955 to 1975, distinguished by being innovative and futuristic. Among all the things that make *Citroën DS* a relevant invention, its design and the technological progress which stands for are the most important of all. Its influence is of such extent that it earned the third position in the Car of the Century ranking [33]. The *DS* model was designed by means of Bézier curves; a design tool having its origin in Computer Aided Geometric Design (CAGD). Gerald Farin, an expert on the topic, defined CAGD as "a discipline dealing with computational aspects of geometric objects" encompassing free-form curves, surfaces and volumes. This subject has numerous applications such as design and manufacturing, digital animation and scientific visualization [34, 35]. Due to its versatility, CAGD becomes suitable for parton distribution function parametrization, which is attained by means of one the major breakthroughs in this field: the theory of Bézier curves and surfaces. Thus, the *Fantômas4QCD* project finds its name and methodology inspired by both french icons: Fantômas and his flying *Citroën DS*.

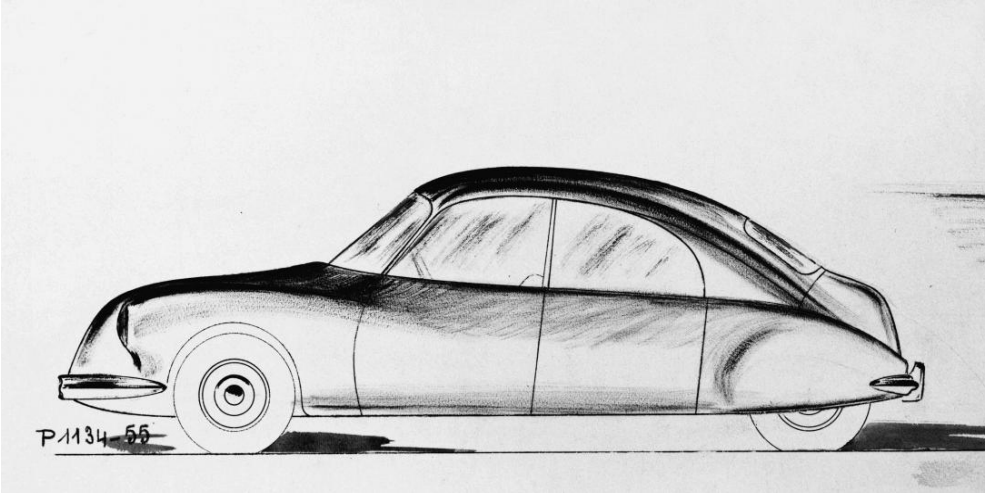
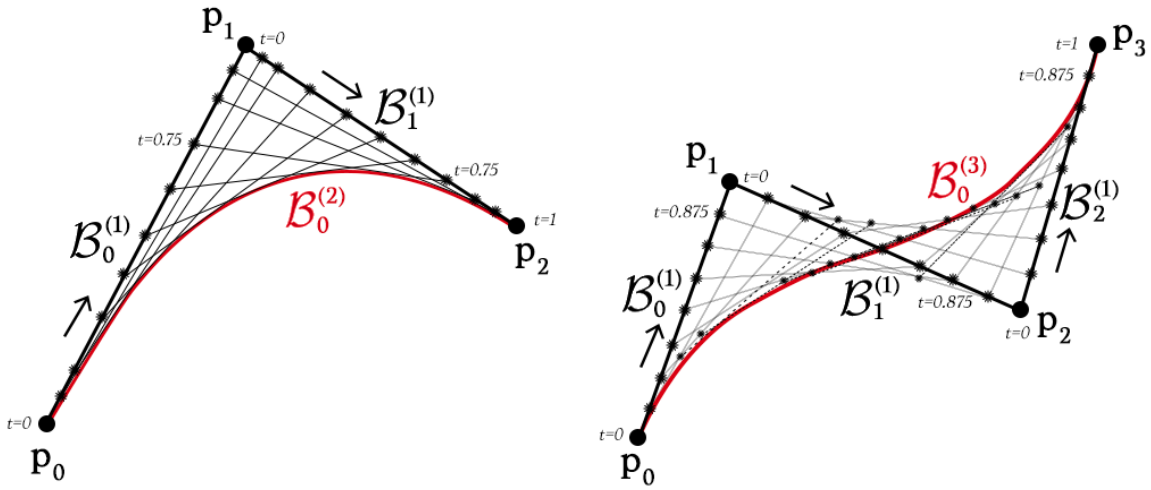


Figure 3.5: *Citroën DS* sketch by the acclaimed designer Flaminio Bertoni. [4]

[4] Extracted from *Citroën Origins* website: `https://www.citroenorigins.co.uk/en/cars/ds`

3.2.2 Bézier Curves and Bernstein Polynomials

Bézier curves are parametric curves acting as a flexible designing method and a commonly used tool in vector graphics programs due to being "well-suited for design work" [36]. This method was invented in 1959 by the physicist and mathematician Paul de Casteljau while working, in fact, at *Citroën*. De Casteljau devised a technique, presently known as *De Casteljau's algorithm*, to draw smooth curves by doing recursive linear interpolations. The algorithm is rather simple; consider the following example:



(a) Second order Bézier function. **(b)** Third order Bézier function.

Figure 3.6: Bézier curves depicted as successive linear interpolations.

- Take three $\mathbb{R}^3$-points: $\mathbf{p_0}$, $\mathbf{p_1}$, $\mathbf{p_2}$. Link successive points with straight segments, resulting in an open polygon called Bézier polygon.

- Generate parametric linear interpolations between points $\mathbf{p_0} \to \mathbf{p_1}$ and $\mathbf{p_1} \to \mathbf{p_2}$ with parameter $t \in [0, 1]$:

$$\mathcal{B}_{\mathbf{0}}^{(\mathbf{1})} = (1-t)\,\mathbf{p_0} + t\,\mathbf{p_1}$$
$$\mathcal{B}_{\mathbf{1}}^{(\mathbf{1})} = (1-t)\,\mathbf{p_1} + t\,\mathbf{p_2}$$

- Create a second order interpolation between points $\mathbf{p_0} \to \mathbf{p_2}$, iterating with $\mathcal{B}_{\mathbf{0}}^{(\mathbf{1})}$ and $\mathcal{B}_{\mathbf{1}}^{(\mathbf{1})}$ interpolations:

$$\mathcal{B}_{\mathbf{0}}^{(\mathbf{2})} = (1-t)\,\mathcal{B}_{\mathbf{0}}^{(\mathbf{1})} + t\,\mathcal{B}_{\mathbf{1}}^{(\mathbf{1})} = (1-t)^2\,\mathbf{p_0} + 2t(1-t)\,\mathbf{p_1} + t^2\,\mathbf{p_2} \tag{3.5}$$

- While t varies from 0 to 1, a curve starting at $\mathbf{p_0}$, ending in $\mathbf{p_2}$ is drawn (figure 3.6a).

Splitting equation 3.5 coordinate-wise:

$$\begin{cases} (\mathcal{B}_0^{(2)})_x = (1-t)^2\, x_0 + 2t(1-t)\, x_1 + t^2\, x_2 \\ (\mathcal{B}_0^{(2)})_y = (1-t)^2\, y_0 + 2t(1-t)\, y_1 + t^2\, y_2 \\ (\mathcal{B}_0^{(2)})_z = (1-t)^2\, z_0 + 2t(1-t)\, z_1 + t^2\, z_2 \end{cases} \tag{3.6}$$

where $\mathbf{p_l}$ has coordinates (x_l, y_l, z_l). Stripping coordinate values from equations in (3.6) reveals a perfect square trinomial structure, where coordinate values are seen as weights modulating the curvature along the line. De Casteljau's algorithm is then generalized to produce interpolations of higher orders, following the same powered-binomial structure. Figure 3.6b depicts an example of a third order Bézier interpolation. Thus, it is convenient to make use of Bernstein's polynomial basis, defined as:

$$B_{n,l}(t) = \binom{n}{l} t^l (1-t)^{n-l} \tag{3.7}$$

such that a n^{th} order interpolation is given by:

$$\mathcal{B}^{(\mathbf{n})}(t) = \sum_{l=0}^{n} \mathbf{p_l}\, B_{n,l}(t) = \sum_{l=0}^{n} \mathbf{p_l} \binom{n}{l} t^l (1-t)^{n-l} \tag{3.8}$$

Therefore, the curvature along the line is determined by the position of $\mathbf{p_l}$ control points and the resulting curve is called ***Bernstein-Bézier approximation***, since it is expressed as a linear combination of Bernstein basis polynomials. Although de Casteljau's method was invented first, later and independently, engineer Pierre Bézier conceived the same idea when working at *Renault*. Since de Casteljau's work was not published until years after, the theory of polynomial curves and surfaces with Bernstein basis was consequently named after Pierre Bézier [34].

Bézier curves were a groundbreaking advance in industrial design in the 1960s and remained as significant during the following decades. Therefore, the fact that Bézier curves have evolved into essential elements in each branch of modern-day computational design should be of no surprise. They are the central principle of operation of *Metamorph*.

3.2.3 Fitting with Bézier Curves

Though control points in ordinary Bézier curve design modulate the curvature along the line, in *Fantômas4QCD*'s approach, control points are used for curve fitting instead, which is applicable to determine parton distribution functions. By polynomial interpolation, a curve passing along k-control points of the form (x_i, f_i) is created. Control points are originated from given data, such as DIS experiments. Constructed by means of the Bernstein basis, *Fantômas4QCD*'s fits are flexible polynomials which can mimic a variety of behaviors of parton distribution functions and their uncertainties. These parametrizations must be adjustable enough to reach agreement with experimental data, while avoiding random fluctuations. At the same time, parametrized PDFs must satisfy the physical QCD-based constraints in 2.3.4.

3.2.3.1 Functional Form

The functional form of a parton distribution function with support in $x \in [0, 1]$ consists in the product of two components: a fixed carrier function $\mathbf{F_c}$ reflecting QCD constraints, and a modulator $\mathbf{P}$, as specified in 3.1.1:

$$f(x) = \mathbf{F_c}\left(x; \{a\}\right) \times \mathbf{P}\left(g(x); \{c\}\right) \tag{3.9}$$

Following CT18's functional form, the modulator is chosen to be a Bézier curve of degree n:

$$\mathbf{P}(g(x); \{c\}) \equiv \mathcal{B}^{(n)}\left(g(x)\right) = \sum_{l=0}^{n} c_l \; B_{n,l}\left(g(x)\right)$$

Rewriting $y \equiv g(x) = x^f$ yields:

$$\mathbf{P}(y; \{c\}) = \sum_{l=0}^{n} c_l \; B_{n,l}\left(y\right) \tag{3.10}$$

with $\{c\} = \{c_l\}_{l=0}^{n}$ the set of coefficients for the Bernstein basis $B_{n,l}(y)$. The argument-scaling function is chosen to be $g(x) = x^f$, f being a real and positive power. The carrier function reflects the small and large-x behavior near the end points through the shape:

$$\mathbf{F_c}\left(x; \{a\}\right) = a_0 \; x^{a_1 - 1}(1 - x)^{a_2} \tag{3.11}$$

3.2.3.2 The Bézier Curve Matrix Method

Bézier curves, together with the arguments motivating their application in PDF parametrization were presented already. Now, we will introduce the algorithm that fits modulation functions in equation (3.9) with *Metamorph*: a Bézier curve of degree n

$$\mathcal{B}^{(n)}(y) = \sum_{l=0}^{n} c_l \; B_{n,l}(y), \tag{3.12}$$

is constructed from Bernstein basis polynomials:

$$B_{n,l}(y) \equiv \binom{n}{l} y^l (1-y)^{n-l}$$

The polynomial $\mathcal{B}^{(n)}(y)$ (now denoted as $\boldsymbol{\mathcal{B}}$) can be expressed as a vector in matrix form, in various symmetric ways [34, 36]:

$$\boldsymbol{\mathcal{B}} = \boldsymbol{T} \cdot \boldsymbol{M} \cdot \boldsymbol{C}, \tag{3.13}$$

$\boldsymbol{T}$ being the $1 \times (n+1)$ row matrix with entries $\{x^p\}_{p=0}^{n}$. If $y = x$, the argument in Bernstein basis is kept non-scaled. If, on the other hand, $y = x^f$ ($f \neq 1$), the scaling of the argument of the Bézier curve translates into the $\boldsymbol{T}$-matrix being expressed in terms of y:

$$\boldsymbol{T} = \begin{pmatrix} 1 & y & y^2 & \dots & y^p & \dots & y^{n-1} & y^n \end{pmatrix}$$

To obtain equation (3.13), use has been made of the binomial expansion of $(1-y)^{n-l}$. $\boldsymbol{M}$ is the square matrix of dimensions $(n+1) \times (n+1)$, whose elements run over indices p and l for the binomial coefficients:

$$\boldsymbol{M} = \binom{n}{l}\binom{n-l}{n-p}(-1)^{p-l} \tag{3.14}$$

or, equivalently, using the symmetries of the binomial coefficient $p \leftrightarrow n-p$. It is also a triangular matrix, reflecting the limits on the sum from the binomial expansion:

$$\boldsymbol{M} = \begin{pmatrix} \binom{n}{l}\binom{n-l}{n-p}(-1)^{p-l}|_{p=0,l=0} & 0 & 0 & \cdots & 0 \\ \cdots & \ddots & \vdots & \vdots & \vdots \\ \binom{n}{l}\binom{n-l}{n-p}(-1)^{p-l}|_{p,l=0} & \cdots & \binom{n}{l}\binom{n-l}{n-p}(-1)^{p-l}|_{p,l} & \vdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \vdots \\ \binom{n}{l}\binom{n-l}{n-p}(-1)^{p-l}|_{p=n,l=0} & \cdots & \binom{n}{l}\binom{n-l}{n-p}(-1)^{p-l}|_{p=n,l} & \cdots & \binom{n}{l}\binom{n-l}{n-p}(-1)^{p-l}|_{p=n,l=n} \end{pmatrix}$$

$\boldsymbol{C}$ is the column matrix of coefficients $\{c_l\}_{l=0}^{n}$ in equation (3.12), of dimensions $(n+1)\times 1$, which must be computed from given data to completely determine the Bézier fit.

$$\boldsymbol{C} = (c_0 \quad c_1 \quad \dots \quad c_l \quad \dots \quad c_n)^T$$

To proceed, a known column vector $\boldsymbol{P} = \{f_i\}_{i=1}^{k}$ of coordinates f_i at control points x_i must be given. $\boldsymbol{C}$ is determined by minimizing the distance between vectors $\boldsymbol{P}$ and $\boldsymbol{B}$, with $\boldsymbol{B} = \{\mathcal{B}(x_i)\}_{i=1}^{k}$ the Bézier polynomial vector evaluated at control points. Once $\boldsymbol{P}$ is provided, the matrix $\boldsymbol{T}$ hence becomes a $k \times (n+1)$-matrix:

$$\boldsymbol{T} = \begin{pmatrix} 1 & y_1 & y_1^2 & \dots & y_1^p & \dots & y_1^{n-1} & y_1^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_i & y_i^2 & \dots & y_i^p & \dots & y_i^{n-1} & y_i^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_k & y_k^2 & \dots & y_k^p & \dots & y_k^{n-1} & y_k^n \end{pmatrix}$$

3.2.3.3 Error Function Minimization

The fitting method just introduced is derived from Mike "Pomax" Kamerman's *Primer on Bézier Curves* [36]; a helpful and extensive guide to learn about the topic. However, this manual is based on a graphic design approach and not directly linked QCD. Still, similar to the assessment of the χ^2 function, by minimizing an error function, a useful expression is obtained to determine the best fitting parameters:

$$E(\boldsymbol{C}) = (\boldsymbol{P} - \boldsymbol{T}\cdot\boldsymbol{M}\cdot\boldsymbol{C})^2 = (\boldsymbol{P} - \boldsymbol{T}\cdot\boldsymbol{M}\cdot\boldsymbol{C})^T \cdot (\boldsymbol{P} - \boldsymbol{T}\cdot\boldsymbol{M}\cdot\boldsymbol{C}) \tag{3.15}$$

Known data is compared with output data from the proposed model, *i.e.* points computed by means of a Bézier curve employing the set of c_l coefficients (equation (3.12)), thus, the aimed task is to find the set of coefficients minimizing the error function. Taking the derivative of the error function and equating to zero gives:

$$\frac{\partial E(\boldsymbol{C})}{\partial \boldsymbol{C}} = -2\,\boldsymbol{T}^T(\boldsymbol{P} - \boldsymbol{T}\cdot\boldsymbol{M}\cdot\boldsymbol{C}_{\text{min}}) = 0$$

Expanding and solving for $\boldsymbol{C}_{\text{min}}$ leads to the general expression:

$$\boldsymbol{C}_{\text{min}} = \boldsymbol{M}^{-1}\cdot(\boldsymbol{T}^T\cdot\boldsymbol{T})^{-1}\cdot\boldsymbol{T}^T\cdot\boldsymbol{P}$$

The previous expression applies for non-square matrices, that is, if $n+1 < k$. In that case, a curve passing through all considered points might not exist, but rather minimizes the distance to such points. However, if $k = n+1$, *i.e.* all points are considered in the interpolation, $\boldsymbol{T}$ is a square matrix, therefore the minimization of the error function $E(\boldsymbol{C})$ leads to:

$$\boldsymbol{C}_{\text{min}} = \boldsymbol{M}^{-1} \cdot \boldsymbol{T}^{-1} \cdot (\boldsymbol{T}^{T})^{-1} \cdot \boldsymbol{T}^{T} \cdot \boldsymbol{P} = \boldsymbol{M}^{-1} \cdot \boldsymbol{T}^{-1} \cdot \boldsymbol{P} \tag{3.16}$$

Substituting $\boldsymbol{C}_{\text{min}}$ in equation (3.12) gives a squared difference equal to zero at known control points. This method was first introduced for parton distribution function fitting by Dr. Courtoy and Dr. Nadolsky [37].

CHAPTER 4

RESULTS

4.1 *Metamorph Mathematica* Notebook

The *Metamorph* program was created in both *Mathematica* and *C++*, but we will focus on the former. *Metamorph* was built remotely from Mexico City and Dallas, Texas, throughout the second half of 2021 by the *Fantômas Development Team* from UNAM and SMU, working in parallel on several tasks. *Mathematica*'s and *C++*'s code is kept under constant improvement. The algorithm previously shown is implemented within the code of both versions. *Metamorph* takes (x_i, f_i) control points and the Bézier fit degree as inputs from a steering *.card* file. Next, required matrices and the $\boldsymbol{C}$ matrix coefficients are computed, then returned together with the Bézier fit and given control points plots.

```
# Fantômas4QCD, from Mathematica, 2021-07-23
# Up-Valence,  CT18NNLO
3.385    0.7632   3.0361   : a0     a1       a2
0        0.5               :Mapping Mode     xPower
# x              Sm                C                          fp       fm
1.e-6            0.000189516       2.1514813246586657         1.e6     0
0.0616280625     0.406786          -0.1619841877067721        1.e6     0
0.24552025       0.720323          2.57417813937053760        1.e6     0
0.5516775625     0.273005          1.3759169204349406         1.e6     0
0.9801           0.0000231644      0.9914791708439334         1.e6     0
```

Figure 4.1: Input *.card* file example for *Metamorph*. These files can be written manually, or taken from *C++*'s and *Mathematica*'s *Metamorph* output.

The *Mathematica* notebook is structured by modules carrying out different tasks:

- ***ReadOptions***: imports *.card* text files with all required parameters to create a Bézier PDF fit, provided that files follow certain format (figure 4.1), which can be interpreted by both implementations to make back-to-back comparisons. The first two lines contain labels and information about the parton under treatment and the data set where values are taken from. The third line contains a_0, a_1 and a_2 carrier function parameters in equation (3.11), since *Metamorph* was employed to parametrize modulation functions in (3.9), fixing the carrier function with values given in CT18 [2]. The next line contains two parameters: ***Mapping Mode*** and ***xPower***; the first one corresponding to the three possible methods of reading s_i values in the second column (Sm) starting at the fifth line:

  - ***Mapping Mode = 0***: s_i values are passed directly as f_i PDF values, belonging to control points of the form (x_i, f_i)

$$f_i = s_i$$

Now, let $f(x)$ denote a parametrized PDF and $x_{min} = 0$, $x_{max} = 1$ initial and end points, respectively. $f(x)$ must satisfy the following condition: it is bounded between an upper (f_+) and a lower (f_-) known functions for $x \in [x_{min}, x_{max}]$ enforcing the positivity condition in section 2.3.4: $f_{\text{-}}(x) \leq f(x) \leq f_{\text{+}}(x)$. Respective $f_{\text{+i}}$ and $f_{\text{-i}}$ boundary values are given in the fourth and fifth column of the *.card* file, linked to each x_i-control point on the first column. The two remaining mapping modes compute f_i between the boundary values contemplating the positivity constraint:

  - ***Mapping Mode = 1***: f_i values are computed by means of a linear bounded scaling activation function

$$f_i = \frac{f_{\text{+i}} + f_{\text{-i}}}{2} + \frac{f_{\text{+i}} - f_{\text{-i}}}{2} \times s_i \text{ , with} - 0 \leq s_i \leq 1$$

  - ***Mapping Mode = 2***: f_i values are computed by means of the softsign activation function

$$f_i = \frac{f_{\text{+i}} + f_{\text{-i}}}{2} + \frac{f_{\text{+i}} - f_{\text{-i}}}{2} \times \frac{s_i}{1 + |s_i|} \text{ , with} - \infty < s_i < \infty$$

The behavior of both functions is depicted in figure 4.2. The ***xPower*** parameter corresponds to the value of the scaling function's f power, as described in subsection 3.2.3.1. Lastly, the third column (C) on the fifth line corresponds to the computed output c_l coefficients for the Bernstein basis in equation (3.12). ***ReadOptions*** module creates the required arrays to create a Bézier fit of PDFs or modulation functions.

- ***Bezier***: takes arrays acquired in ***ReadOptions*** and computes the coefficients matrix $\boldsymbol{C}$ by means of matrices $\boldsymbol{T}$, $\boldsymbol{P}$ and $\boldsymbol{M}$, following the method explained in 3.2.3.3. This module takes as parameters the number of control points k and the degree of the fitting polynomial n, with $n + 1 \leq k$. To reach *C++*'s *Metamorph* results and hold numerical precision, an additional command was implemented within matrix operations where precision should not be disregarded: *Rationalize*, converting numbers to nearby rational forms within a specified tolerance.

- ***BezierOutput***: returns the plot of the Bézier fit together with control points and the list of computed c_l Bernstein basis coefficients. ***BezierOutput*** exports a *.card* file with the same format as the input file. *C++*'s *Metamorph* can take as input files *Mathematica*'s *Metamorph* output and vice versa. Multiple bugs from both implementations were discovered by running the same examples in a parallel way.

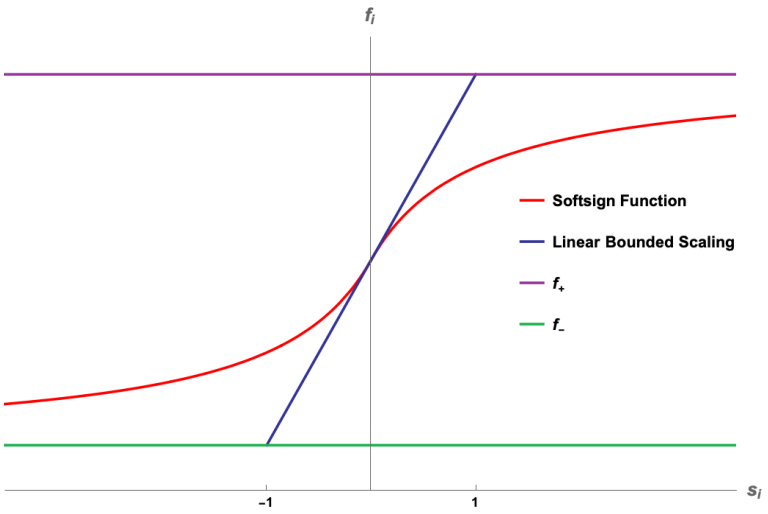


Figure 4.2: Softsign and linear bounded scaling functions, approaching the boundary values towards the domain endpoints.

4.1.1 Examples

As a practical application, *Metamorph* was employed to parametrize modulation functions in equation (3.9), based on CT18's results by taking their PDFs parametrizations and fixing the carrier function with parameters a_0, a_1 and a_2, given in [2]. Because values inside *.card* files are directly related to PDFs instead of modulators, s_i values must be divided by the known carrier functions evaluated at the respective x-control points to obtain modulator f_i values.

4.1.1.1 Up Valence

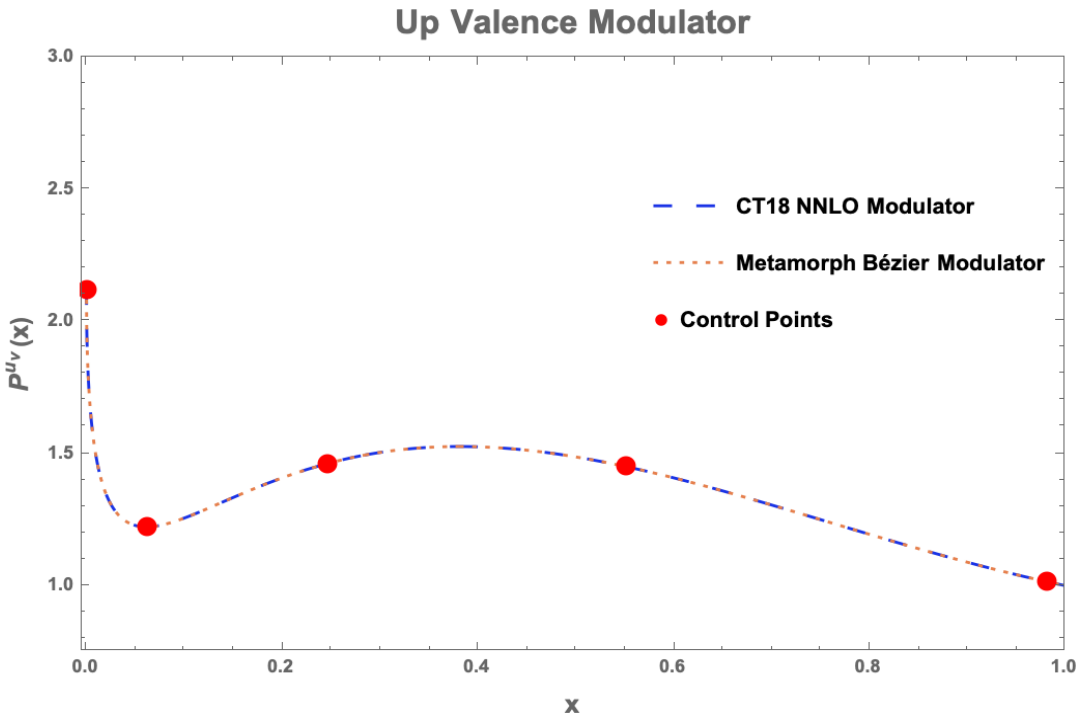


Figure 4.3: Plots of CT18's and *Metamorph*'s fit for the up valence modulation function with evenly spaced control points.

A Bézier fit of CT18's up valence modulator was produced with *Metamorph*, considering five evenly distributed control points. A $y = \sqrt{x}$ scaling was used as well, yielding a Bernstein polynomial of fourth degree:

$$\begin{aligned} P^{u_v}_{4,\text{Met}}\left(\sqrt{x}\right) \;=\; & c_0\left(1-\sqrt{x}\right)^4 + 4c_1\left(1-\sqrt{x}\right)^3\left(\sqrt{x}\right) + \\ & 6c_2\left(1-\sqrt{x}\right)^2\left(\sqrt{x}\right)^2 + 4c_3\left(1-\sqrt{x}\right)\left(\sqrt{x}\right)^3 + c_4\left(\sqrt{x}\right)^4 \end{aligned} \tag{4.1}$$

Because x represents Bjorken's fractional momentum, physical end points are $x_{\text{min}} = 0$ and $x_{\text{max}} = 1$, although, in order to avoid divisions by zero, the initial point was set as $x_{\text{min}} = 10^{-6}$ and the final point as $x_{\text{max}} = 0.99$, while ***Mapping Mode*** was set to 0, passing s_i values in figure 4.1 as f_i values, discarding $f_{\pm i}$ boundary values in this example. The $\{c_l\}_{l=0}^{4}$ output coefficients computed by *Metamorph* are shown in the following table, together with CT18's best fit parameters:

Up Valence Coefficients					
Metamorph	c_0	c_1	c_2	c_3	c_4
	2.134	-0.1472	2.565	1.382	1
CT18	2.134	-0.1472	2.565	1.382	1
Relative Error (%)	$< 1.6 \times 10^{-4}$	$< 4.9 \times 10^{-3}$	$< 2.9 \times 10^{-4}$	$< 5.7 \times 10^{-4}$	$< 3 \times 10^{-5}$

Table 4.1: *Metamorph*'s and CT18's fitting coefficients for the up valence modulator.

The up valence modulator fit was the first example worked out by the *Fantômas Development Team* with promising results, showing that the project was moving on the right track. Comparing both sets of coefficients, the induced error in this example is of, at most, 0.0049%, implying that results obtained with *Metamorph* are, indeed, reliable. Figure 4.4 shows the ratio between CT18's and *Metamorph*'s up valence modulators, depicting a precise performance since the quotient lies on the constant unit line from x_{min} to x_{max}.

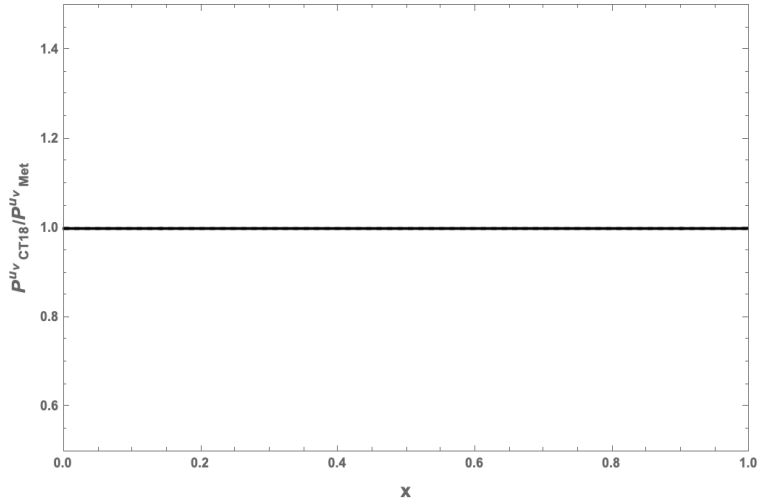


Figure 4.4: Ratio plot between CT18's and *Metamorph*'s up valence modulators, remaining close to 1 throughout the entire domain.

As mentioned in section 2.3.4, valence parton distribution functions must comply the valence sum rule, stating that their integrals give the quantum numbers. Multiplying the modulator in equation (4.1) by the modulator with corresponding parameters and integrating numerically with *Mathematica* gives the result:

$$\int_0^1 3.385\, x^{0.763-1}(1-x)^{3.036} \times P^{u_v}_{4,\text{Met}}(\sqrt{x})\, dx = 1.999$$

as expected, since protons contain two valence quarks.

4.1.1.2 Down Valence

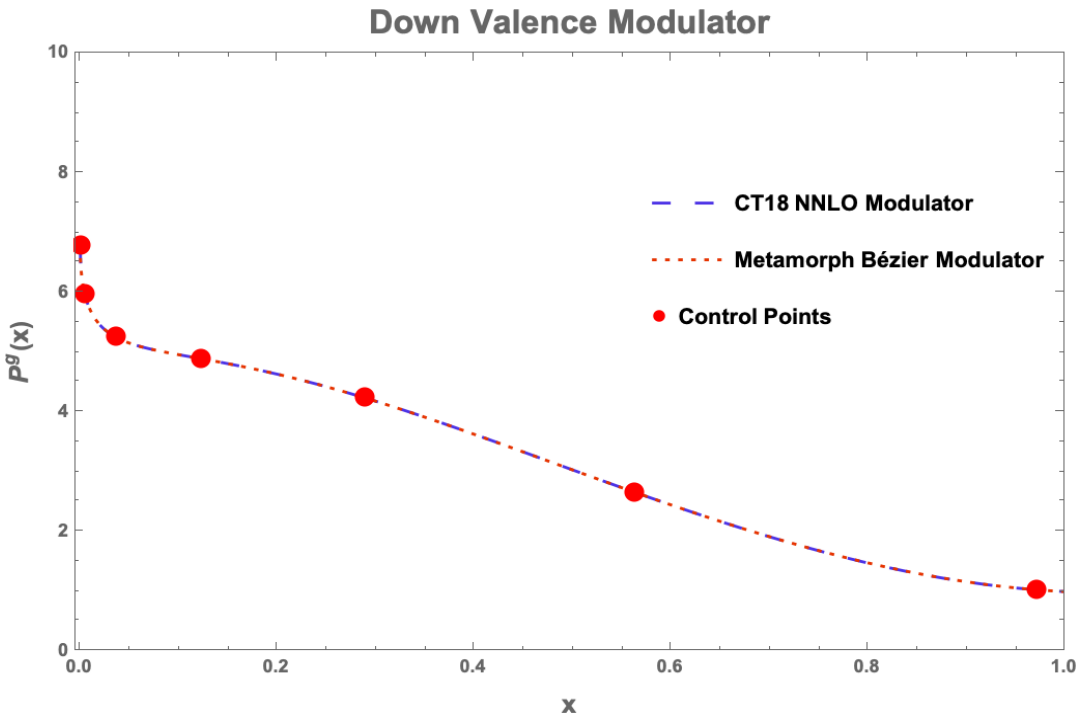


Figure 4.5: Plots of CT18's and *Metamorph*'s fit for the down valence modulation function with unevenly spaced control points.

Created with seven unevenly distributed control points (four of them clustered towards the low-x zone as a consequence of a power law distribution, detailed in section 4.2), this Bézier fit yields a Bernstein polynomial of sixth degree:

$$\begin{aligned} P^{d_v}_{6,\text{Met}}\left(\sqrt[3]{x}\right) \;=\; & c_0\left(1-\sqrt[3]{x}\right)^6 + 6c_1\left(1-\sqrt[3]{x}\right)^5\left(\sqrt[3]{x}\right) + 15c_2\left(1-\sqrt[3]{x}\right)^4\left(\sqrt[3]{x}\right)^2 + \\ & 20c_3\left(1-\sqrt[3]{x}\right)^3\left(\sqrt[3]{x}\right)^3 + 15c_4\left(1-\sqrt[3]{x}\right)^2\left(\sqrt[3]{x}\right)^4 + 6c_5\left(1-\sqrt[3]{x}\right)\left(\sqrt[3]{x}\right)^5 + c_6\left(\sqrt[3]{x}\right)^6 \end{aligned} \tag{4.2}$$

The $\{c_l\}_{l=0}^{6}$ output coefficients computed by *Metamorph* are shown in the following table:

Down-Valence Coefficients							
Metamorph	c_0	c_1	c_2	c_3	c_4	c_5	c_6
	6.799	6.401	4.097	4.669	6.899	1.397	0.9991

Table 4.2: *Metamorph*'s fitting coefficients for the down valence modulator.

This fit was produced with different settings than those given by CT18, aiming to show that *Metamorph* is a versatile and powerful tool. The main changes for this example are the degree of the Bézier fitting polynomial and the scaling function, chosen to be $y = \sqrt[3]{x}$. This implies that not only CT18's $\sqrt{x}$ scaling function is appropriate since adequate results were achieved, which is depicted by the modulators ratio plot in figure 4.6:

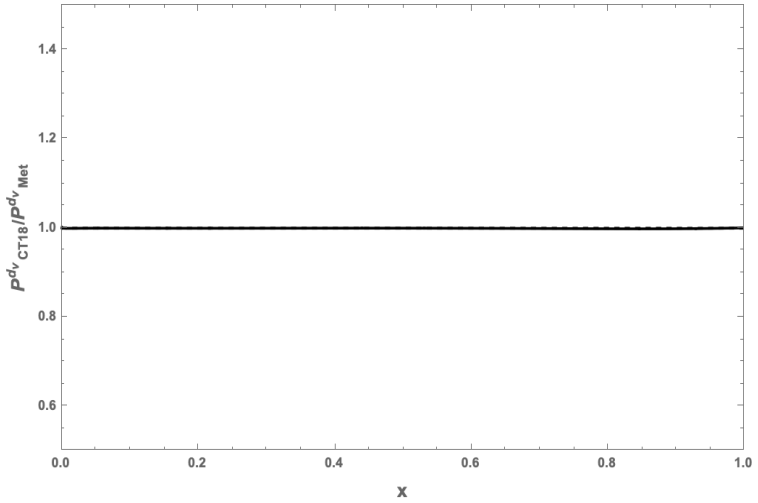


Figure 4.6: Ratio plot between CT18's and *Metamorph*'s down valence modulators, remaining close to 1 throughout the entire domain.

If we multiply the down valence modulation function in equation (4.2) by the respective carrier function, we can test again the PDF valence sum rule by integrating numerically with *Mathematica*, giving the result:

$$\int_0^1 0.490\, x^{0.763-1}(1-x)^{3.036} \times P^{d_v}_{6,\text{Met}}\left(\sqrt[3]{x}\right)\, dx = 1.001$$

which approaches to one; the number of down valence quarks confined within a single proton.

4.1.1.3 Gluon

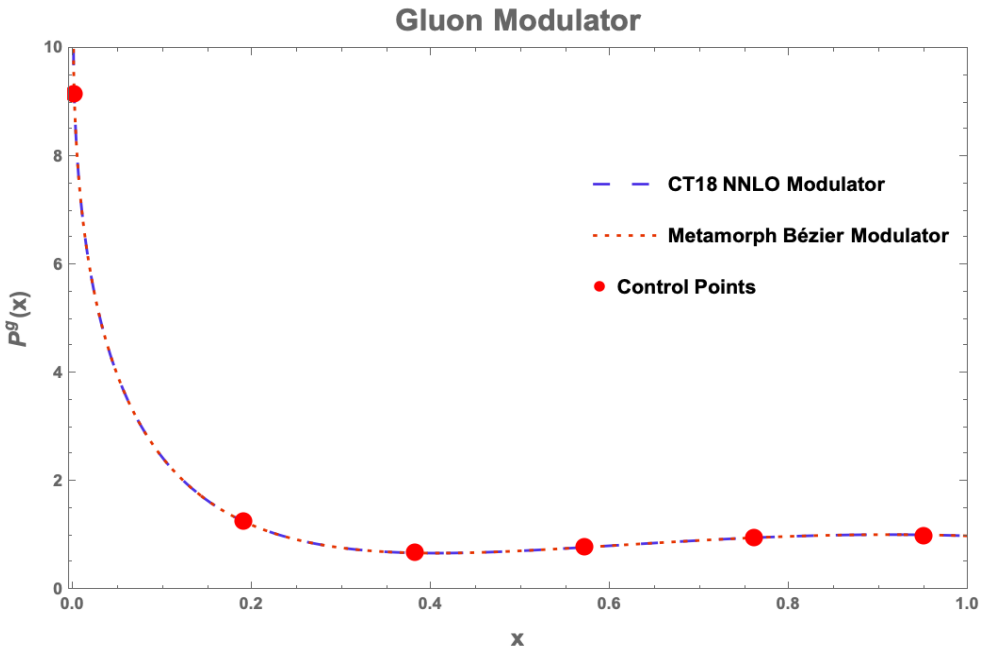


Figure 4.7: Plots of CT18's and *Metamorph*'s fit for the gluon modulation function with evenly spaced control points.

This Bézier fit was produced with six evenly spaced control points, yielding a fifth degree Bernstein polynomial with scaling function $\sqrt{x}$:

$$P^{g}_{5,\text{Met}}\left(\sqrt{x}\right) = c_0\left(1-\sqrt{x}\right)^5 + 5c_1\left(1-\sqrt{x}\right)^4\left(\sqrt{x}\right) + 10c_2\left(1-\sqrt{x}\right)^3\left(\sqrt{x}\right)^2 + 10c_3\left(1-\sqrt{x}\right)^2\left(\sqrt{x}\right)^3 + 5c_4\left(1-\sqrt{x}\right)\left(\sqrt{x}\right)^4 + c_5\left(\sqrt{x}\right)^5 \tag{4.3}$$

and $\{c_l\}_{l=0}^{5}$ output coefficients computed by *Metamorph*:

Gluon Coefficients						
Metamorph	c_0	c_1	c_2	c_3	c_4	c_5
	10.34	2.542	−0.1551	0.1144	1.212	1

Table 4.3: *Metamorph*'s fitting coefficients for the gluon modulator.

Similar to the previous example, a back-to-back coefficient comparison with CT18's coefficients cannot be performed, since settings were chosen to be different for this example. However, the ratio plot between CT18's and *Metamorph*'s gluon modulators exhibiting an adequate behavior over $x \in [0, 1]$ as depicted in figure 4.9a.

4.1.1.4 Strange (Antistrange)

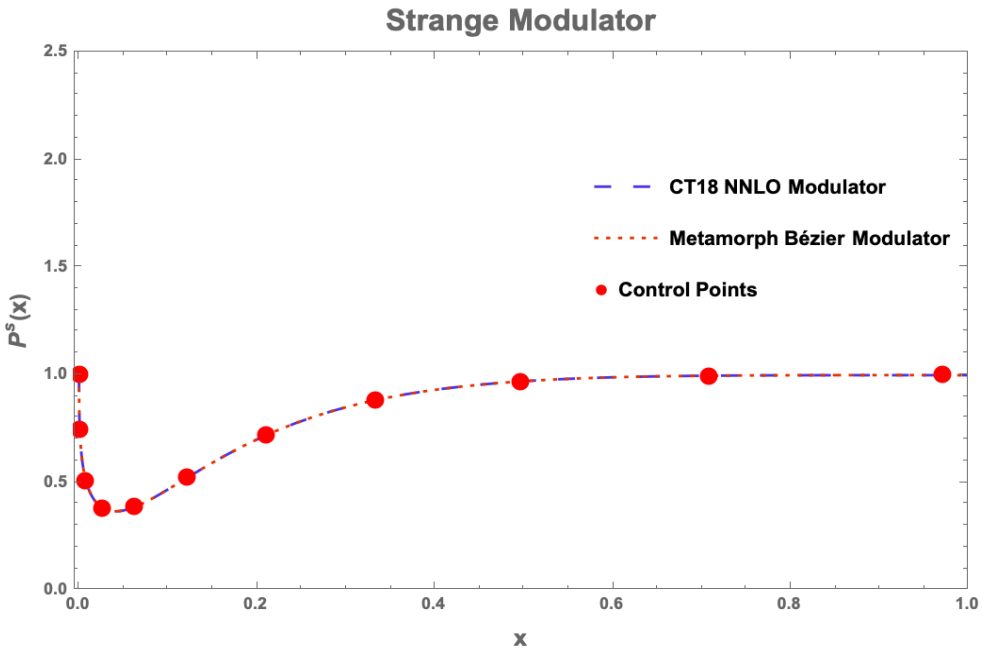


Figure 4.8: Plots of CT18's and *Metamorph*'s fit for the strange modulation function with unevenly spaced control points.

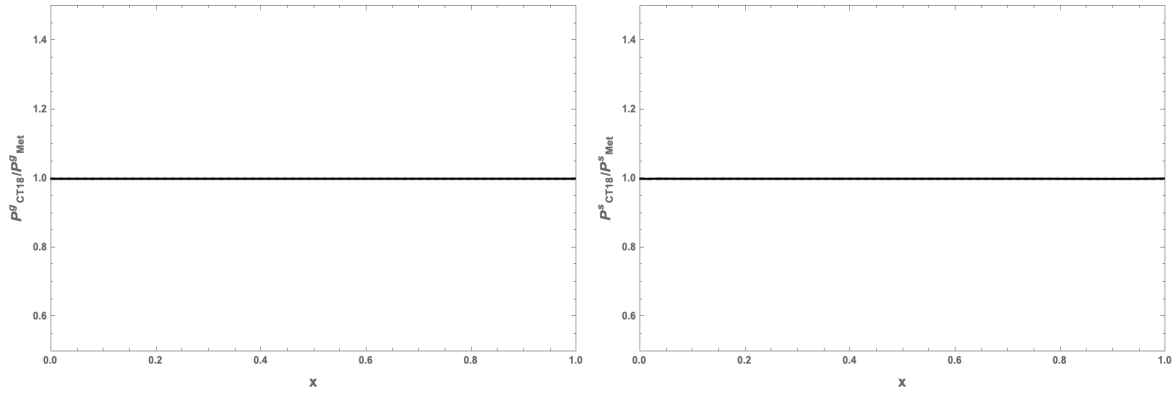
CT18 provides parameters for sea quarks modulator fits as well, employing a different argument for a modulation fifth degree Bernstein polynomial:

$$y \;=\; 1-(1-\sqrt{x})^{a_3}$$

fixing $a_3 = 4$. Now, we reproduce the fit for strange quarks, whose parton distribution function is assumed to be the same as antistrange quarks, with *Metamorph* using the same scaling function as the down valence fit: $y = \sqrt[3]{x}$.

Figure 4.8 shows *Metamorph*'s Bézier fit with eleven unevenly distributed control points, with the set of lowest suitable input parameters as follows: power law distribution $\phi = 3$ (detailed in the following section), scaling power $f = 1/3$, number of control points $k = 11$ and degree $n = 10$. This figure should capture the readers' attention due to the amount of required points to produce the fit, reasonably higher than previous examples. In the same way as the down valence modulator, most points are clustered towards low-x regions where

slope variations are significantly noticeable. By focusing on low-x points, care must be taken not forgetting to place sufficient control points towards high-x zones to faithfully reconstruct CT18's modulator over the entire domain as shown in the ratio plot in figure 4.9b, that being the reason of such high amount of control points. Simultaneously, the additional constraint on the scaling function had an evident impact when producing *Metamorph*'s fit, since fixing input parameters with similar values to those of the previous examples did not return satisfactory results, at least, near the low-x zone. Thus, the two distinct scaling argument functions are not precisely compatible, but they can reach an agreement by increasing the degree of the polynomial with power scaling and identifying the correct way of distributing control points. The ratio plot is shown in figure 4.9b. The explicit fitting polynomial of the strange modulation function is omitted due to being of high degree.



(a) Gluon. **(b)** Strange.

Figure 4.9: Ratio plot between CT18's and *Metamorph*'s gluon and strange modulators respectively, remaining close to 1 throughout the entire domain.

4.2 Scaling and Control Point Selection

As seen from chapters 3 and 4, *Metamorph* is quite flexible when it comes to parameters. All of them have a fundamental role within the algorithm producing Bézier fits, nonetheless, two parameters stand out due to their influence and sensitivity on the output and results: the scaling function $y = g(x)$ in 3.2.3.1 and x-control points; specifically, the amount considered, the distribution along the domain and the degree of the Bernstein polynomial generated.

Consequently, it is convenient to assess the selection of analytical parameters to use *Metamorph*'s flexibility to our advantage, in order to get the best possible output.

The results just presented required dozens of test runs. It may appear that the sets of parameters giving the best fits are merely determined by trial and error. This is not completely true, since proceeding that way is only helpful when it comes to fine tuning of the results. Consequently, by conceiving general usage criteria, fitting with *Metamorph* becomes more practical. Once the code was correctly developed, the subsequent months were assigned to the task of finding a strategy to determine the set of parameters giving the best output by analyzing the behavior of results for distinct parameter combinations, using the up valence and gluon illustrative examples as prototypes. The outcome of such analysis is a set of guidelines on how to implement *Metamorph* appropriately. These guidelines should be solely taken as suggestions, allowing the final user to play around with values to find the most suitable set of parameters, according to their particular needs.

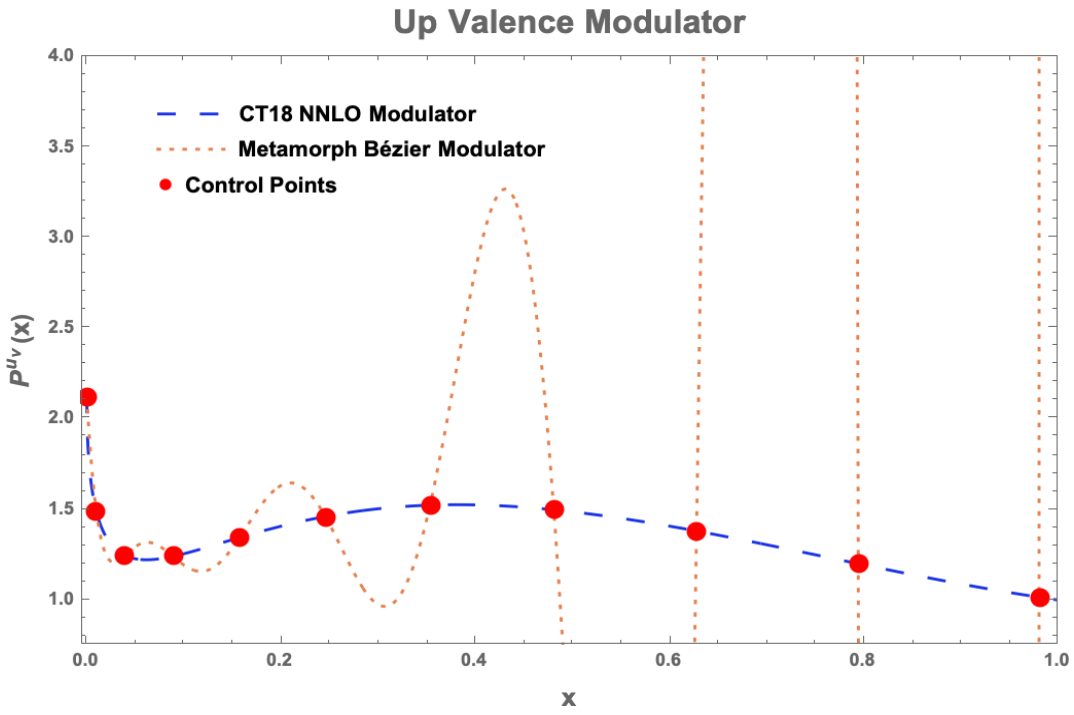


Figure 4.10: CT18's and *Metamorph*'s Up-Valence modulator, with 11 evenly distributed control points and $f = 1$.

First, through the dozens of test runs, it was determined that the best way to produce Bézier fits is by employing all available control points by means of equation (3.16). Therefore, the Bernstein polynomial's degree n must be equal to $k-1$, with k the amount of control points. At the same time, the user might be enticed to generate high degree fits with several control points, aiming to recover as many attributes as possible, so that the output polynomial resembles the intended behavior "accurately". This, in fact, is generally not true, since Runge's phenomenon can take place, producing unwanted oscillations between control points as depicted in figure 4.10. Thus, more control points does not necessarily imply more accuracy. It is then suggested to keep n low, as long as the desired behavior is still displayed.

At the same time, the distribution of control points is more than relevant. It is reasonable to think that a partition of evenly distributed control points along the domain is enough to cover all regions within the objective curve. While this is partly true, it does not depict the whole context thoroughly. Control point distribution can obey different rules, though it is essential to still include some as near as possible to the physical end points $x_{min} = 0$ and $x_{max} = 1$. For even spacing, points are given by an ordinary grid:

$$x_i = x_{\text{min}} + \frac{x_{\text{max}} - x_{\text{min}}}{n} \times i \text{ , for } \quad i = 0, \ldots, n$$

being useful when the objective plot is not monotonic over all regions in the domain, like the gluon example. Conversely, control points can follow a power law as well:

$$x_i = \left(x_{\text{min}} + \frac{x_{\text{max}} - x_{\text{min}}}{n} \times i \right)^{\phi} \text{ , for } \quad i = 0, \ldots, n$$

where ϕ is called *distribution power* ($\phi > 0$), applying when the objective curve is actively changing towards regions of low-x (down valence and strange examples in section 4.1.1). Early versions of *Metamorph* had the feature of control point computing with power law distributions within the ***Bézier*** module, however, this feature was removed in the final versions. Still, notes on how to create control points with both rules are given in recent versions. Then, the distribution criteria is actually broader: besides from end points, control points should be clustered near non-uniform zones where the function has pronounced slope variations, *i.e.* around local maxima and minima, etc.; a notion the user should constantly bear in mind.

Secondly, the behavior of the Bézier fit is determined by the scaling function $y = f(x)$. For a power scaling, according to CT18 [2] and the examples shown previously, it is found that scaling the argument as x^f in the modulator with $f = 1/2$ provides a smooth behavior due to being "very flexible across the whole interval $0 < x < 1$" [2], though, different powers might be as suitable and give great results, like the down valence and strange modulator fits, shown in figures 4.5 and 4.8, where a $\phi = 3$ power was used to distribute control points along the domain. Nonetheless, power scaling is not the only available option to adjust the Bernstein polynomial's argument. Moderately complex custom expressions can be applied, like the strange modulator's scaling in 4.1.1.4. The final user should examine in advance whether similar expressions are suitable, or if, on the other hand, it is more convenient to maintain power scaling thanks to its smoothness and versatility, as with the previous examples. Both choices can be easily implemented in *Metamorph*'s code, inside the ***Bezier*** module.

Power scaling was, in general, useful when producing fits by setting the right f power value and keeping low n degrees, in contrast to not scaling the argument at all, *i.e.* $f = 1$. It is seen, from figure 4.10, that the up valence modulator Bézier fit of tenth degree with no scaled argument returns poor results. While developing this example, with scaling $f = 1/2$, it was observed that the polynomial degree could be considerably increased employing more control points and simultaneously, keeping the desired behavior. This is, of course, excessive, since high precision was reached with polynomials of lower degrees and no noticeable changes could be detected with complex polynomials of higher degrees, but this, in fact, proves how scaling with rational powers has clear advantages over no scaling. Therefore, scaling could be set as a standard criteria when producing fits with *Metamorph*.

4.2.1 Supplementary Discussion

We have just argued that analytical parameters usage has a central role in performance and accuracy, determining whether *Metamorph* succeeds or fails when producing a fit displaying the intended behavior, however, there is still an independent setting which has not been discussed in depth: the mapping mode.

All of the examples previously shown have ***Mapping Mode*** set as 0, designating s_i values read from *.card* files directly as f_i-control point coordinates, giving appropriate outputs. Still, there was no allusion to mapping modes 1 and 2 in any of the results. Examples with these settings were omitted as a consequence unsatisfactory outputs with the implementation of *Metamorph* considered in section 4.1.1; parametrization of modulators in equation (3.9). A question may naturally raise: *what is the purpose of using mapping modes 1 and 2, then?*

Recalling from the structure of *Metamorph* in section 4.1, by setting the mapping mode to 1, f_i-control point coordinates are computed by means of a linear bound scaling activation function, whereas mapping mode 2 makes use of the softsign function. Both instances have the purpose of imposing penalties and constraints (such as positivity) on f_i values, linked to uncertainties as described in sections 2.3.4 and 4. Our examples have the defect of lacking a treatment of uncertainties, because the input control points produced from CT18's parametrizations were assumed to be error-free; a condition which does not reflect true experimental/theoretical circumstances, since "modern PDF parametrizations are provided with families of 'error PDF sets"' [29]. These parametrizations allow to locate parton distribution functions between error bands accounting for all uncertainties considered within the model; an example of such is shown in figure 3.1. The analysis of uncertainties in *Metamorph* was out of the scope of this thesis, but it is certainly under implementation now, as the *Fantômas4QCD* project is still progressing.

CHAPTER 5

CONCLUSIONS

The methodology and results encompassed in this thesis constitute the first phase of the *Fantômas4QCD* project, which concluded in the late 2021. During this stage, the development team had to become familiar with DIS phenomena, essential notions of QCD and PDFs before getting into coding. This goal was achieved with the aid of Dr. Aurore Courtoy and Dr. Pavel Nadolsky, both being experts on these topics, through various lectures given during June, 2021. Once these concepts were grasped, the development team then moved onto the tasks of writing, testing and debugging the code and reaching optimal numerical precision with both *Mathematica* and *C++*. The project is currently in its second phase, consisting on the implementation of the program to relevant physical problems such as the pion PDF and, in addition to CT18's results, setting up illustrative examples in order to distribute the program and final users to assimilate *Metamorph*'s structure and functioning.

Both versions of *Fantômas4QCD*'s *Metamorph* are expected to be publicly released in the following months, together with an extensive report. At the same time, major advances have been made recently on *C++*'s *Metamorph* end, but there is still a way to go. The *Fantômas Development Team* is currently working on the task of implementing *Metamorph* together with the latest version of *xFitter* (formerly *HERAFitter*): *2.2.0 FutureFreeze*. xFitter is a PDF fitting program commonly used by CTEQ, described as "*an open-source package that provides a framework for the determination of the parton distribution functions of the proton and for many different kinds of analyses in Quantum Chromodynamics.*"[38].

In retrospect, we can describe the course of action of this work as follows: with a perspective based on theoretical and experimental grounds, namely, the quark-parton model, QCD and DIS, the foundations of parton distribution functions were given, accompanied by their relevance for describing how matter is structured at a fundamental level and the underlying difficulties in their determination, setting PDF parametrization as a complex major task currently under performance. Consequently, the *Fantômas4QCD* project introduced in this thesis, emerged as an alternative to current PDF parametrization methods by offering a practical and innovative approach, supported by one of the cornerstones of graphic design: Bézier curves. Through illustrative examples, we showed that the *Metamorph* program written by the *Fantômas Development Team* is able to produce accurate fits with experimental data in an efficient and powerful way, based on CT18's PDF parametrizations. Not only a fitting program is provided, but the team also offers guidelines on how to implement their program properly in order to meet the needs of final users, making *Metamorph* a promising tool for PDF fitting at larger experimental projects.

Building the *Fantômas4QCD* project throughout the preceding months has been fruitful not only in the sense of providing an efficient, helpful and practical program for PDF fitting, but also for serving as a first approach for the students in the development team to Quantum Chromodynamics and high-energy physics, opening the door for them to professionally move forward towards programming-oriented fields, as well as theoretical projects, or expand their career paths and work on different branches of physics.

BIBLIOGRAPHY

[1] **Particle Data Group** Collaboration, R. L. Workman and Others, "Review of Particle Physics," PTEP (2022) 083C01.

[2] T.-J. Hou, J. Gao, T. J. Hobbs, K. Xie, S. Dulat, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, I. Sitiwaldi, D. Stump, and C.-P. Yuan, "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC," Phys. Rev. D **103** (Jan, 2021) 014013. https://link.aps.org/doi/10.1103/PhysRevD.103.014013.

[3] T. McEvilley, The Shape of Ancient Thought: Comparative Studies in Greek and Indian Philosophies. Aesthetics Today Series. Allworth Press, 2002. https://books.google.com.mx/books?id=Vpqr1vNWQhUC.

[4] M. Gangopadhyaya, Indian Atomism: History and Sources. Bagchi Indological Series 1. K. P. Bagchi & Company, 1980.

[5] H. Margenau, Physics and Philosophy: Selected Essays. D. Reidel Publishing Company, 3 ed., 1978.

[6] C. C. W. Taylor, The Atomists: Leucippus and Democritus: A Text and Translation with Commentary. University of Toronto Press, 1999.

[7] S. Berryman, "Ancient Atomism," in The Stanford Encyclopedia of Philosophy, E. N. Zalta, ed. Metaphysics Research Lab, Stanford University, Winter 2016 ed., 2016.

[8] B. Pullman and A. Reisinger, The Atom in the History of Human Thought. Oxford University Press, 2001. https://books.google.com.mx/books?id=IQs5hur-BpgC.

[9] I. J. R. Aitchison and A. J. G. Hey, Gauge Theories in Particle Physics: A Practical Introduction, vol. 1: From Relativistic Quantum Mechanics to QED. Institute of Physics Publishing, 3 ed., 2003.

[10] D. H. Perkins, Introduction to High Energy Physics. Cambridge University Press, 4 ed., 2000.

[11] F. Halzen and A. D. Martin, Quarks & Leptons: An Introductory Course in Modern Particle Physics. John Wiley & Sons, Inc., 1984.

[12] K. Gottfried and V. F. Weisskopf, Concepts of Particle Physics, vol. I. Oxford University Press, 1984.

[13] H. Yukawa, "On the Interaction of Elementary Particles. I," Proceedings of the Physico-Mathematical Society of Japan. 3rd Series **17** (1935) 48–57.

[14] D. Griffiths, Introduction to Elementary Particles. Wiley-VCH Verlag GmBH & Co. KGaA, 2 ed., 2008.

[15] T. Rothman and S. Boughn, "Can Gravitons be Detected?," Foundations of Physics **36** no. 12, (Nov, 2006) 1801–1825. https://doi.org/10.1007%2Fs10701-006-9081-9.

[16] R. K. Ellis, W. J. Stirling, and B. R. Webber, QCD and Collider Physics. Cambridge University Press, 2003.

[17] F. E. Close, An Introduction to Quarks and Partons. Academic Press Inc., 1979.

[18] R. G. Roberts, The Structure of the Proton: Deep Inelastic Scattering. Cambridge University Press, 1993.

[19] M. E. Peskin and D. V. Schroeder, An Introduction to Quantum Field Theory. Addison-Wesley Publishing Company, 1995.

[20] I. S. Hughes, Elementary Particles. Cambridge University Press, 3 ed., 1991.

[21] J. D. Bjorken and E. A. Paschos, "Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon," Phys. Rev. **185** (Sep, 1969) 1975–1982. https://link.aps.org/doi/10.1103/PhysRev.185.1975.

[22] R. P. Feynman, "The behavior of hadron collisions at extreme energies," in Special Relativity and Quantum Theory: A Collection of Papers on the Poincaré Group, M. E. Noz and Y. S. Kim, eds., pp. 289–304. Springer Netherlands, Dordrecht, 1988. https://doi.org/10.1007/978-94-009-3051-3_25.

[23] K. Gottfried and V. F. Weisskopf, Concepts of Particle Physics, vol. II. Oxford University Press, 1986.

[24] W. B. Atwood, "LEPTON NUCLEON SCATTERING," Prog. Math. Phys. **4** (11, 1979) 1–114.

[25] A. Bodek, M. Breidenbach, D. L. Dubin, J. E. Elias, J. I. Friedman, H. W. Kendall, J. S. Poucher, E. M. Riordan, M. R. Sogard, and D. H. Coward, "Comparisons of deep-inelastic $e-p$ and $e-n$ cross sections," Phys. Rev. Lett. **30** (May, 1973) 1087–1091. https://link.aps.org/doi/10.1103/PhysRevLett.30.1087.

[26] D. J. Gross, "The discovery of asymptotic freedom and the emergence of QCD," Proceedings of the National Academy of Sciences **102** no. 26, (2005) 9099–9108, https://www.pnas.org/doi/pdf/10.1073/pnas.0503831102. https://www.pnas.org/doi/abs/10.1073/pnas.0503831102.

[27] J. Feltesse, "Introduction to Parton Distribution Functions," Scholarpedia **5** no. 11, (2010) 10160.

[28] V. Barone and E. Pedrazzi, High Energy Particle Diffraction. Springer, 2002.

[29] K. Kovařík, P. M. Nadolsky, and D. E. Soper, "Hadronic structure in high-energy collisions," Rev. Mod. Phys. **92** (Nov, 2020) 045003. https://link.aps.org/doi/10.1103/RevModPhys.92.045003.

[30] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, "Neural Network Parametrization of Deep-Inelastic Structure Functions," Journal of High Energy Physics **2002** no. 05, (May, 2002) 062–062.

[31] CTEQ, "The Coordinated Theoretical-Experimental Project on QCD." http://cteq.org/. [Accessed: 08/02/2022].

[32] M. Priestman, The Cambridge Companion to Crime Fiction. Cambridge University Press, 2003.

[33] A. Law, "Car of the Century." Auto123.com, Dec., 1999. https://web.archive.org/web/20080102121653/http://www.auto123.com:80/en/info/news/news,view.spy?artid=1082. [Archived from the original on: 03/08/2006. Accessed: 08/10/2022].

[34] G. Farin, Curves and Surfaces for CAGD: A Practical Guide. Morgan Kauffman Publishers, 5 ed., 2002.

[35] G. Farin, J. Hoschek, and M.-S. Kim, Handbook of Computer Aided Geometric Design. North-Holland, 2002.

[36] M. Kamermans, "A Primer on Bézier Curves," 2011. https://pomax.github.io/bezierinfo/. [Accessed: 07/20/2022].

[37] A. Courtoy and P. M. Nadolsky, "Testing momentum dependence of the nonperturbative hadron structure in a global QCD analysis," Phys. Rev. D **103** no. 5, (2021) 054029, arXiv:2011.10078 [hep-ph].

[38] S. Alekhin, O. Behnke, P. Belov, S. Borroni, M. Botje, D. Britzger, S. Camarda, A. M. Cooper-Sarkar, K. Daum, C. Diaconu, J. Feltesse, A. Gizhko, A. Glazov, A. Guffanti, M. Guzzi, F. Hautmann, A. Jung, H. Jung, V. Kolesnikov, H. Kowalski, O. Kuprash, A. Kusina, S. Levonian, K. Lipka, B. Lobodzinski, K. Lohwasser, A. Luszczak, B. Malaescu, R. McNulty, V. Myronenko, S. Naumann-Emme, K. Nowak, F. Olness, E. Perez, H. Pirumov, R. Placakyte, K. Rabbertz, V. Radescu, R. Sadykov, G. P. Salam,

A. Sapronov, A. Schoening, T. Schoerner-Sadenius, S. Shushkevich, W. Slominski, H. Spiesberger, P. Starovoitov, M. Sutton, J. Tomaszewska, O. Turkot, A. Vargas, G. Watt, and K. Wichmann, "HERAFitter, Open Source QCD Fit Project," 2014. https://arxiv.org/abs/1410.4412.