THE IMPACT OF pPb W/Z LHC DATA ON nCTEQ15

PARTON DISTRIBUTION FUNCTIONS

WITH AN EMPHASIS ON REFITTING THE nCTEQ15 nPDF SET

Approved by:

_____

Dr. Fredrick Olness
Professor

_____

Dr. Stephen Sekula
Professor - Committee Chair

_____

Dr. Jingbo Ye
Professor

_____

Dr. Joseph F. Owens III
Professor - Florida State University

THE IMPACT OF pPb W/Z LHC DATA ON nCTEQ15

PARTON DISTRIBUTION FUNCTIONS

WITH AN EMPHASIS ON REFITTING THE nCTEQ15 nPDF SET

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Physics

by

Eric Godat

B.A., Physics, Southwestern University
B.A., Mathematics, Southwestern University
M.S., Physics, Southern Methodist University

May 19, 2018

ACKNOWLEDGMENTS

As no physicist can work in a vacuum, I have had the privilege to work with some amazing physicists and even better people. First and foremost is my advisor, Fred Olness, for without his mentorship and vision none of this dissertation would exist. I'd like to thank Steve Sekula for his friendship and guidance as well as his service as the chair of my dissertation committee. Thank you to the other members of my dissertation committee (Jeff Owens, Jingbo Ye and Kent Hornbostel) for their advice and perspective as I produced and defended this work. I also wish to express my gratitude to my colleagues in the `nCTEQ` collaboration (D.B. Clark, T. Ježo, C. Keppel, K. Kovařík, A. Kusina, F. Lyonnet, J. Morfín, F.I. Olness, F.J. Owens, I. Schienbein, J.Y. Yu,) for the countless hours of consultation, discussion and camaraderie. Additionally, I'd like to thank the `xFitter` collaboration for fostering my growth as a global physicist.

Thank you to the Southern Methodist University Physics Department for giving me the opportunity to do this work. To the Office of Graduate Studies, I offer my appreciation for their generous support through the Dissertation Fellowship. A special thank you to Lacey Breaux for helping me navigate the bureaucratic ins and outs of SMU and to my fellow graduate students for fostering my growth and for commiserating when classes were neigh impossible and the code didn't compile.

Furthermore, I could not have done this work without the incredible support from my family. I'd like to thank my mom, Martha, for never questioning my ability to succeed even when I did; my dad, Peter, for kindling my love of physics at an early age and for teaching me that knowing is not as important as understanding; my stepdad, Dale, for reminding me that everyday can be the best day if you want it to be; my stepmom, Lene, for being the

iv

model of hard work without compromising a bit of humor; my sister, Rachel, for keeping me grounded and humble and always having my back; and my stepbrother, Tyler, for being my co-op wingman whenever I needed to come up for air.

Thank you to Freckles, for being happy to see me, whenever I came home, since Day 1 and to Dexter for being a constant source of furry comedy.

Finally, to Christina, thank you for being my teammate and perpetual well of motivation. Without you, this process would have been impossible.

Godat, Eric

B.A., Physics, Southwestern University
B.A., Mathematics, Southwestern University
M.S., Physics, Southern Methodist University

The Impact of pPb W/Z LHC data on `nCTEQ15`
Parton Distribution Functions
With An Emphasis on Refitting the `nCTEQ15` nPDF Set

Advisor: Dr. Fredrick Olness
Doctor of Philosophy degree conferred May 19, 2018
Dissertation completed March 29, 2018

As the LHC experiments increase their statistics and reduce their systematics, it is increasingly important to reduce the theoretical uncertainties so we can perform precision comparisons. The parton distribution functions (PDFs) which describe the hadronic structure of protons and nuclei are the critical link between the experimental data and the proposed theoretical models; hence, a better understanding of the PDFs is essential for improved predictions. Nuclear PDFs, in particular, have proven valuable in this regard, especially for flavor differentiation. I introduce "nCTEQ+LHC," the first PDF set fit using the nCTEQ formalism to include data from the LHC. This PDF set also represents the first results of the new (C++ based) "nCTEQ++" fitting code. To incorporate many new LHC processes into nCTEQ++, I developed and tested a novel way of uniting various theory tools in a modular framework to enable fast PDF fitting of complex higher-order theory calculations using modern grid techniques. This combination of tools provides a foundation for a wide variety of future analysis involving both proton and nuclear data.

TABLE OF CONTENTS

CHAPTER

viii

LIST OF FIGURES

LIST OF TABLES

To my parents

for their unwavering support

and encouragement.

Chapter 1

Introduction

The primary focus of my work has been to improve the way parton distribution functions (PDFs) are determined, utilized and configured for calculations. PDFs represent a vital phenomenological link between theoretical calculations and experimental data. These PDFs describe the probability of a parton inside a hadron[1] with a specific fraction of the hadron's total momentum and a given energy. PDFs are determined by fitting[2] a parameterization to experimental data and are presented as a series of values stored in a discrete three dimensional grid referenced by partonic flavor, hard scattering energy and momentum fraction. Additionally, each PDF is accompanied by a set of error PDFs that represent some inherent uncertainty in the determination of the PDF fit. Once a PDF has been determined, it is possible to make predictions and compare to experimental measurements with the ultimate goal of minimizing the uncertainty in the PDF and accurately predicting future experimental results. However, this iterative process of PDF fitting is neither straightforward nor trivial, and choices driven by limits in methodology, data and computation all complicate the process. In this dissertation, I will detail my efforts to increase the ease of which PDFs are fit as well as my work to improve understanding of PDFs in general. I also use the results of my work to look forward and predict potential future physics results.

The remainder of this chapter will go in to the theoretical motivation and background of PDFs. This starts with the Standard Model of particle physics and the methods employed by high-energy physicists to better understand the foundational building blocks of the universe as we know it. I will discuss elements of the Standard Model relevant to PDF fitting as well

---

[1]Typically a proton, but neutrons and heavier nuclei can also be represented.

[2]Fitting is a key concept in this dissertation and will be explained in detail later. However, it can be described simply as finding the best set of values for the coefficients in a polynomial such that the resulting curve describes the data you wish to fit.

as the motivations for studying nuclear PDFs and the corrections needed to accommodate their added complexity.

In Chapter 2, I will present a software package, `ManeParse`, I developed that improves the ability for users to manipulate and perform calculations with PDFs inside the `Mathematica` framework. `ManeParse` represents my effort to improve the field of PDF fitting as a whole as well as a tool that was employed throughout the remainder of this work. `ManeParse` is a user-friendly interface that allows the user to load PDF sets in a variety of formats directly into `Mathematica` and perform calculations directly without any necessary external links, a utility that had only existed in a limited capacity prior to its publication.

In an effort to understand the impact of new data on an existing PDF release, in Chapter 3, I detail a process known as PDF reweighting. Since the `nCTEQ15` PDF release does not include any data from the Large Hadron Collider (LHC), it became important to understand the possible effect of this data on the PDF fit. However, the complexity in including a new data set in the existing `nCTEQ` fitting code meant that it was useful to first estimate the impact of this new data. To do so we employed a technique known as Bayesian reweighting using an augmented version of `FEWZ` [1] [2] to allow for asymmetric LHC beams. The results of this study determined that this data was significant enough to warrant inclusion in a new `nCTEQ` PDF fit.

Chapter 4 delves deeper into the mechanics of PDF fitting as I present my work on heavy flavor variable number schemes with the `xFitter` collaboration. When PDFs are fit at higher orders, discontinuities appear as a result of quark mass thresholds. When fitting at low energies, it is acceptable to neglect the contributions of the heavy quarks, leaving you with a three-flavor fitting scheme. As you evolve to higher energies, this scheme begins to introduce errors as the charm, and eventually bottom, quark mass thresholds are passed. This problem is exacerbated when data at higher energy, often spanning a mass threshold or simply falling above it, is included in the fit. When this occurs, multiple fitting schemes can be used simultaneously within the same fit introducing discontinuities and enhancing the uncertainty in the fit. However, by making a different choice of energy to make the switch

between quark flavor numbering schemes, opposed to simply at the quark masses, presents a possibility of mitigating the introduced uncertainty. This chapter contains the results from the study into these variable number schemes.

Finally, in Chapter 5, I reach the main effort of this work: including LHC data into a `nCTEQ` PDF fit. This begins with an introduction to the `nCTEQ` framework for PDF fitting and a summary of the most current `nCTEQ15` PDF release. From there, I explain the reworked `nCTEQ` fitting code and a series of validations I performed to confirm that the significant improvements made to the code base did not hinder the functionality needed to produce PDF fits. I then explain the series of steps I developed to include the LHC data into the `nCTEQ` fit utilizing `MCFM` as well as `APPLgrid` utilities. This culminates with the presentation of a pair of new PDF fits, one representing simply the `nCTEQ15` parameters including the LHC data and one representing a relaxation of the strange quark parameters that had been previously fixed due to a lack of constraining data. These fits are then compared to the results from the PDF reweighting study presented in Chapter 3. Additional conclusions as well as directions for future study and predictions for future experimental results using these new fits can be found in Chapter 6.

1.1. Standard Model

Long ago the ancient Greek philosopher Leucippus and his pupil Democritus proposed that there existed a fundamental discrete unit of the universe that was indivisible. This unit, which they called an atom, came in different shapes and sizes and was based on their theological beliefs rather than any scientific insight; however there does seem to be some element of scientific truth to their theory. As many a reader will know, an atom in its modern usage refers to the smallest unit of a chemical element that retains the properties of that element, so calling an atom an "atom" is in fact a misnomer as an atom actually consists of some number of protons and neutrons making up a nucleus and a cloud of electrons. Yet

even the protons and neutrons in the nucleus are not indivisible[3]. As predicted by Richard Feynman in 1969, these "nucleons" are actually comprised of particles known as partons.
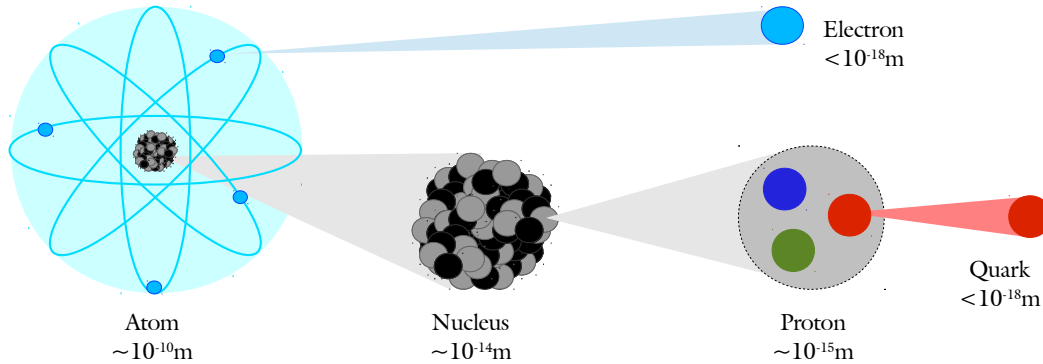
Figure 1.1: Graphical representation of relative sizes of subatomic structures.

In order to observe these more fundamental objects, we need to use higher and higher energy probes. As seen in Fig. 1.1, a proton is roughly on the scale of a femtometer (10^{-15}m). Thus using de Broglie's equation, Eq. 1.1, to relate momentum and wavelength, we can determine that we need a probe that carries approximately 1 gigaelectron-volt (GeV) worth of momentum in order to be sensitive enough to observe its properties. If we want to observe partons and electrons at smaller scales, we need even higher energy probes. This concept of using energetic probes to understand internal structure will come up again, but first let us discuss the theoretical formulation that governs this subatomic regime.

$$\lambda = \frac{h}{p} \tag{1.1}$$

The Standard Model (SM) is a quantum field theory (QFT) that is built on the ideas of gauge invariance, relativity and quantum mechanics. The theory can be divided into three main groups of particles by the quantum spin of the particles in each group.

---

[3]As of this writing, electrons are believed to be point particles classically constrained to a radius of $2.8179403227(19) \times 10^{-15}$ m [3].

The first of these groups consists of the spin-1 gauge bosons and can be written concisely as a group with three quantum numbers (Color, weak isospin and weak hypercharge) as seen in Eq. 1.2. These bosons are charge carriers and correspond to three of the four fundamental forces of nature[4].

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.2}$$

The first term in Eq. 1.2, $SU(3)_C$, refers to the strong force mediated by gluons. Gluons, carriers of the strong force, couple to color charged objects (quarks) and are described by quantum chromodynamics (QCD). Some features the strong force (QCD) relevant to the study of PDFs will be discussed in further detail later. The remaining terms, $SU(2)_L \otimes U(1)_Y$, describe the electroweak force, a high energy unification of the electromagnetic force and the weak force. The gauge bosons for the electroweak force, after spontaneous symmetry breaking, are the massive W and Z bosons that mediate the weak force and the massless photon (γ), understood using quantum electrodynamics (QED).

The second group of particles in the SM consists of spin-1/2 fermions, which can be broken down into two families, one of leptons and another of quarks. Each of these families have three generations that are functionally identical to each other save for differences in mass. Quarks can carry each of the quantum numbers from the SM and are the only fermions that carry color. There are six varieties, or "flavors", of quarks: up, down, charm, strange, top and bottom. Leptons can be divided into two groups, the first consisting of the charged electron, muon and tau and the second consisting of their associated neutral neutrinos. Leptons are colorless objects but can carry weak isospin and/or weak hypercharge.

The third, and relatively new group, represents the Higgs mechanism and the associated spin-0 Higgs boson. The Higgs mechanism generates the spontaneous symmetry breaking of the electroweak force and leads to the masses of the W/Z bosons. Additionally, via the

---

[4]The electromagnetic, strong and weak forces; gravity is not described by the Standard Model.

Yukawa interaction, the Higgs gives mass to the quarks and charged leptons. The Higgs, at least at leading order, does not couple to the massless gluons or photon [4].

1.1.1. Quantum Chromodynamics

The study of PDFs is inherently governed by the QCD properties of the SM. QCD itself is a renormalizable non-Abelian gauge theory with $SU(3)_C$ symmetry that is pertubative at high energies, but nonpertubative at low energies. As mentioned above, quarks are "color" charged objects and thus interact with gluons. This charge comes in three varieties: red, blue and green for quarks along with the anti-color of each, anti-red, anti-blue and anti-green for anti-quarks. Color neutral objects can be formed by having one of each color (rgb), one of each anti-color $(\bar{r}\bar{g}\bar{b})$ or having a color anti-color pair (e.g. $r\bar{r}$). Gluons are the charge carriers in QCD and allow quarks to change color by emitting a colored gluon[5]. For example, a red quark and a blue quark could exchange a red anti-blue gluon causing both quarks to change color (from red to blue and blue to red respectively): $q_r + g_{b\bar{r}} \rightarrow q_b$. Since gluons are inherently colored objects, they are able to self-interact[6] (unlike photons), a fact that will be relevant later when discussing asymptotic freedom. It is this feature that distinguishes the strong force (QCD) from the electromagnetic force (QED).

Armed with this understanding of color, it is easy to see how hadrons, particles consisting of bound states of quarks and/or anti-quarks, are formed. A proton, for example, is a $(u_r u_g d_b)$ state where each quark represents a different color. But what is to prevent two quarks from being the same color within the proton or even further, what is to prevent a bare quark from existing? Since no state with net color charge has ever been observed in nature, QCD includes a property requiring all observable states to be color neutral. This means that the only way a proton can be constructed is for each quark to represent a different color and

---

[5]There are 8 gluons possible in QCD, naively one might expect (3 colors × 3 anti-colors =) 9 gluons but the linear combination of $R\bar{R} + G\bar{G} + B\bar{B}$ represents a color singlet and is not allowed under SU(3) symmetry, thus the remaining 8 linearly independent combinations for a color octet.

[6]E.g. $g_{r\bar{g}} + g_{b\bar{r}} \rightarrow g_{b\bar{g}}$

Standard Model of Elementary Particles
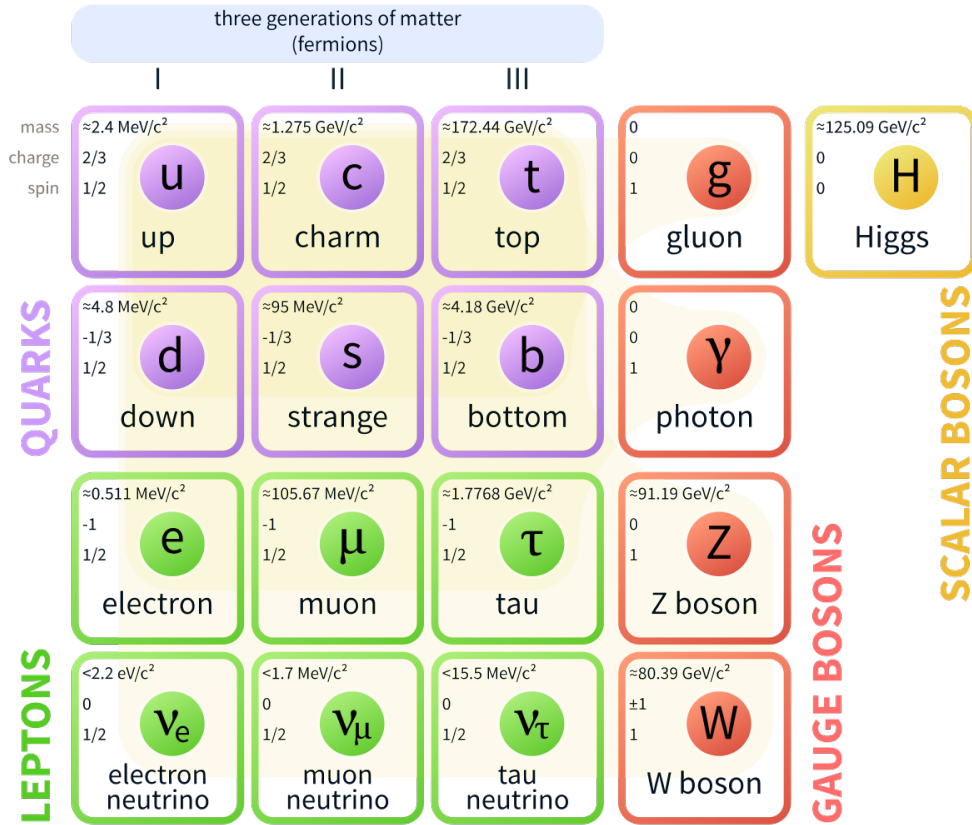
Figure 1.2: Graphical representation of the Standard Model of Particle Physics.

that free quarks cannot exist on their own. Additionally, this prevents gluons from existing on their own[7]. This property is known as confinement.

1.1.2. Asymptotic Freedom

At this point, a clever reader might ask why one couldn't simply extract a quark from its confined state, much like removing an electron from an atomic bound state to form an ion. To answer this, let's first go back to introductory electromagnetism.

---

[7]Free quarks and gluons can exist briefly in the aftermath of hadronic collisions due to the uncertainty principle but they rapidly form jets of color neutral hadrons.

$$F = \frac{k \ q_1 \ q_2}{r^2} \qquad (1.3)$$

When we examine Coulomb's law, Eq. 1.3, we find that the strength between two electrically charged objects, is directly related to the strength of the electric charge and inversely to the distance between them. If we were to decrease that distance enough, we must leave the realm of classical physics and instead look at the quantum effects between the electric charges (QED). Here the vacuum around the electron becomes polarized; virtual electron-positron pairs pop into existence around the electron and align in its electric field. The virtual electrons are repelled away, leaving a cloud of virtual positrons attracted to the electron. These positively charged positrons effectively shield the charge of the electron, reducing the "bare" charge of the electron to the effective, classical charge for the electron we are used to. This shielding is known as "renormalization" of the electric charge.

If we wanted to measure the "bare" charge of an electron, we could fire a second electron with high enough energy to penetrate the charge-shielding cloud of the first electron. When this is done, we find that the charge and correspondingly the strength of the electromagnetic force increases asymptotically. To put it more simply, in QED, the smaller the length scale, the stronger the force and conversely, the larger the length scale, the weaker the force.

Since QCD and QED share a similar structure, we could attempt to do a similar test for two quarks instead of electrons. Starting with a single colored quark, we see that the vacuum once again polarizes, this time with virtual quark-antiquark pairs and gluons. But here, QCD begins to behave differently. As mentioned above, gluons, unlike the photon, can self-interact which allows for gluon shielding in addition to the quark-antiquark shielding. This leads to a renormalization of the color charge that actually causes an enhancement of the effective color over the "bare" color of the quark. Simplifying again, in QCD, the smaller the length scale, the weaker the force and conversely, the larger the length scale, the stronger the force. This property, known as asymptotic freedom, allows quarks to move freely when closely bound together, while simultaneously needing a large amount of energy to be pried

from their bound state. This energy, it turns out, is enough to create a quark-antiquark pair[8] such that all quarks continue to satisfy confinement.

> For an analogy of this behavior, imagine a rubber band, cut so that the ends are free. If you were to stretch this rubber band far enough, eventually it would snap leaving two rubber bands. If we think of the ends of the rubber band as quarks (or anti-quarks) and the amount of stretching to be the energy needed to free the quarks, we see how the newly formed quark-antiquark pair allows our original quarks to separate yet prevents them from existing freely.

The strength of the force between two charged objects changing as a function of scale, is known as a running coupling. In pertubative QCD, this coupling, α_s, can be calculated at leading order using Eq. 1.4, where Q^2 represents the energy scale of the interaction, Λ_{QCD}^2 represents the energy scale of QCD interactions ($\sim 1\,\mathrm{GeV}^2$ or $1\,\mathrm{fm}$) and b is a constant related to the number of quark flavors. The role of α_s in PDF fits will be examined further in Chapter 5.

$$\alpha_s(Q^2) = \frac{1}{b \; \ln(Q^2/\Lambda_{QCD}^2)} \tag{1.4}$$

At this point, QCD has been described a theoretical framework for strong interactions, but physics is an experimental science and thus any good theory must make testable predictions. The following section details the phenomenological elements of QCD that form the foundations for PDF analyses.

1.2. Phenomenology

Phenomenology is the study of applying theoretical models, in our case QCD, to experimental observables in high energy physics and vice versa. As we have mentioned previously, in order to study objects at very small scales, one needs to use highly energetic probes. These types of experiments are referred to as scattering experiments and form the basis of

---

[8]Via the mass-energy relation, Einstein's famous $E = mc^2$

particle physics all the way back to Rutherford's discovery of the atomic nucleus. Below, I will cover two types of modern scattering experiments relevant to the motivation behind and determination of PDFs. Following that, I will return to discussing properties of QCD, namely Bjorkën Scaling and factorization.

1.2.1. Scattering

In order to study the structure of the proton, one must determine a choice of probe. Referring back to Fig. 1.2 we might consider an electron as a valid option; they can be isolated (unlike quarks), accelerated to high energies due to their electric charge, plus they are stable and readily available. We might also consider other probes. Neutrinos are the lightest fermion but rarely interact thus requiring heavy targets, and quarks complicate things due to their confinement. Nevertheless, all three of those particles are used in scattering experiments.

Scattering experiments that involve a lepton interacting with a proton are known as "deep inelastic scattering" experiments (DIS). In these experiments, a lepton with enough energy to penetrate the proton (deep), exchanges momentum with a quark within the proton, causing the quark to recoil (inelastic) and effectively kicking a quark out of the proton. These experiments then measure the scattering angle and energy of the lepton after the collision to determine the specifics of the interaction which can then be related to the fraction of the momentum the quark carried and the hard scattering energy. Deep inelastic scattering allows us to study several aspects of the proton, and because of the relative cleanliness of these observables, DIS data is important for constraining uncertainties in PDF fits; however because leptons cannot couple to gluons, this data gives no insight in that regard.

The Drell-Yan (DY) process utilizes quarks to probe the structure of the proton. In this type of experiment, a hadron, e.g. a proton, is accelerated such that one of its quarks can interact with an anti-quark from another proton, annihilating into an intermediate electroweak bosons, which subsequently decays into two leptons. These two leptons are easily observable in a detector and thus make DY a popular calibration tool in hadron colliders.

More on DY can be found in Chapters 3 and 5 as both W and Z production data studied in those chapters is the result of DY.

Going back to DIS, if we examine the formula for the DIS differential cross section[9] (Eq. 1.5), we can see that it consists of a sum of three structure functions, F_1, F_2, F_3.

$$\frac{d^2\sigma^i}{dxdy} = (\frac{4\pi\alpha^2}{xyQ^2}\eta^i)\Big\{y^2 xF_1^i + (1 - y - \frac{x^2y^2M^2}{Q^2}) \ F_2^i$$
$$\mp \ (y - \frac{y^2}{2}) \ xF_3^i\Big\}$$

(1.5)

These functions describe the internal structure of a hadron when probed by a lepton[10] as a function of momentum fraction, x and hard scattering energy, Q. While the differential cross section depends on y, the inelasticity of the collision, and M, the mass of the hadron, the structure functions do not. These functions, a selection of which are listed below as Eq. 1.6, can be observed experimentally and provide insight into the flavor structure of PDFs[11]. These functions provide a direct relationship between an experimental observable (left) and a theoretically defined abstract object (right)[12], a concept that will be important later the discussion on PDFs.

---

[9]Cross section is a common observable in scattering experiments and is effectively a measure of the probability of a specific interaction occurring. More details, can be found in the Particle Data Group review article on Structure Functions [4], including the relevant coefficients, definitions of the kinematic variables and the η^i factor.

[10]The DIS differential cross section can also be written as a product of the Leptonic (L) and Hadronic tensors (W), shown here:
$$\frac{d^2\sigma^i}{dxdy} = \frac{2\pi y\alpha^2}{Q^4}\sum_j \eta_j L_j^{\mu\nu} W_{\mu\nu}^j$$

[11]The longitudinal structure function can be shown at Leading-Order to be $F_L = 0$ and thus $F_2 = 2xF_1$. This result is known as the Callan-Gross relation.

[12]Here the notation $F_2^{\ell p}$ refers to the structure function resulting from lepton-proton scattering, where the lepton, ℓ, could be an electron, e, a muon, μ, or a neutrino, ν. The letters on the right side of the equations (u, d, c, s) represent different quark flavors involved in the interaction.

$$\begin{aligned}
F_2^{ep} &= \frac{4}{9}x[u + \bar{u} + c + \bar{c}] \\
&+ \frac{1}{9}x[d + \bar{d} + s + \bar{s}] \\
F_2^{\nu p} &= 2x[d + s + \bar{u} + \bar{c}] \\
F_2^{\bar{\nu} p} &= 2x[u + c + \bar{d} + \bar{s}] \\
F_3^{\nu p} &= 2[d + s - \bar{u} - \bar{c}] \\
F_3^{\bar{\nu} p} &= 2[u + c - \bar{d} - \bar{s}]
\end{aligned} \tag{1.6}$$

1.2.2. Bjorkën Scaling and The Partonic Model

However, these structure functions provide a interesting window into QCD as well, in the form of Bjorkën scaling. As we see in Fig. 1.3, the structure function remains essentially constant in energy Q^2 for all but the most extreme values of x. This independence of the structure on Q^2 was predicted in 1969 by James Bjorkën and put simply, states that if the momentum of the probe becomes large, the proton stops behaving as a single object and instead the probe scatters off the internal partons as though they were independent point particles. Scaling directly corresponded with Richard Feynman's parton model, which stated that nucleons are made of constituent particles called partons.

So now that we know the proton is constructed of point-like particles that we can scatter probes off of, let's revisit the DIS structure functions in the context of the parton model. Assume the proton is a simple collection of three free quarks, then we would assume that each of them carries $\frac{1}{3}$ the total momentum of the proton ($x = \frac{1}{3}$). However, we know this to be impossible because, at minimum, the quarks within the proton are bound to each other by the strong force. These interquark interactions lead to a smearing of the momentum, from a single value to a range of possible values. As we see in Fig. 1.3, at the extreme values of x, the Q dependence grows even further. This is a result of our leptonic probe "seeing" increasingly more of these interquark interactions.
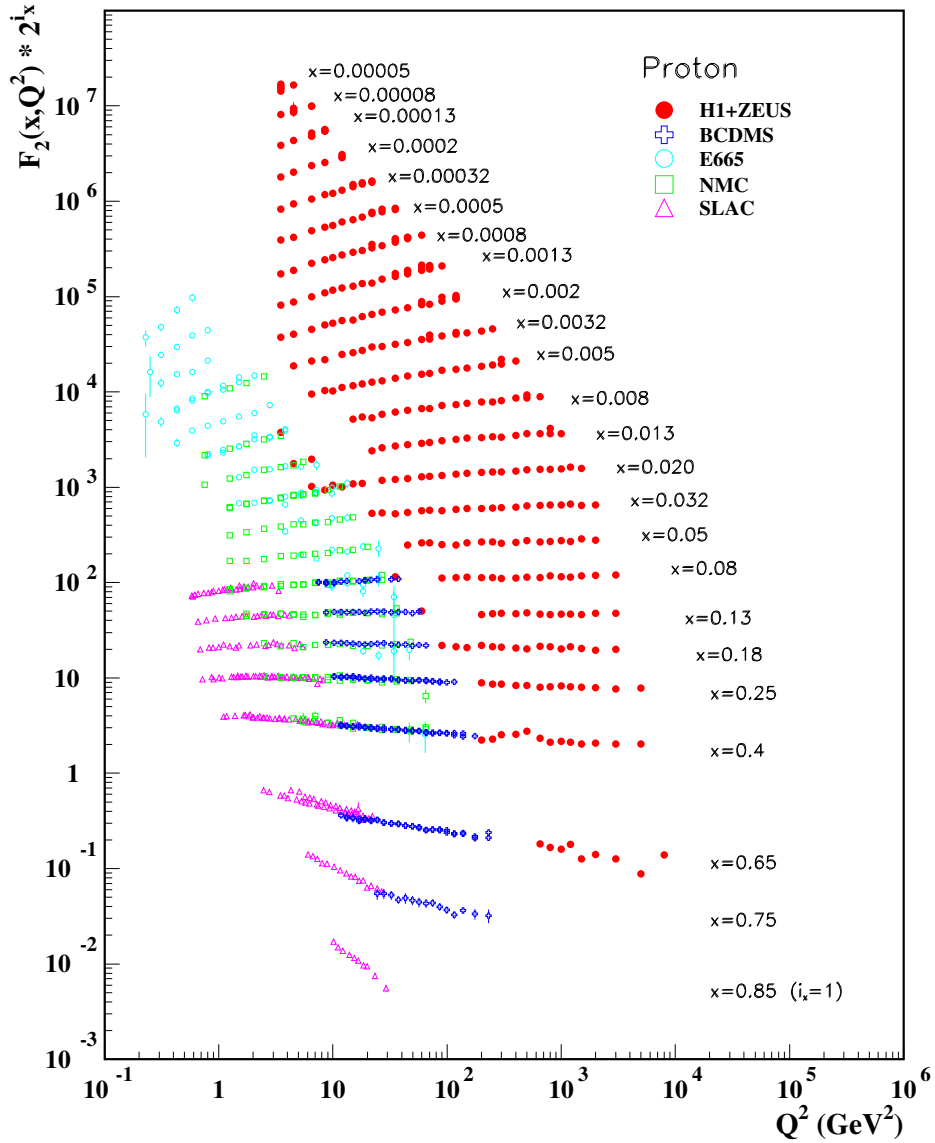
Figure 1.3: Plot of F_2^p structure function as a function of Q^2 from a collection of DIS experiments. Scaling (Q^2 independence) is evident for the central values of x ($x \approx 0.13$); while at the more extreme values of x, scaling is violated as described by the DGLAP evolution equation characterizing the QCD interactions within the proton.

13

1.2.3. Factorization

If we return to Eq. 1.5 equipped with our new understanding of partonic structure, we can schematically rewrite the equation in a simpler, albeit less technical, way.

$$d\sigma_{DIS} \; = \; \sum_i f_i(x, Q^2) \otimes \hat{d\sigma}^i \tag{1.7}$$

What we have done is split the equation into two pieces: the first represents the probability that the quark our lepton interacts with has a given fraction of the hadron's momentum and energy; the second represents the interaction between the lepton and the quark as point particles. If we then sum this over all the possible quarks, we get back out differential cross section. From a theoretical standpoint, this would be an excellent feat of mathematical gymnastics because it allows us to rewrite the cross section as a convolution of a low-energy, nonpertubative, soft piece and a high-energy, pertubative, hard piece. Additionally it allows us to replace several terms that can only be empirically determined (F_1, F_2, F_3) with only one, $f_i(x, Q^2)$. Fortunately, this technique works because of the differences in the energy scales of the interactions, and is known as factorization.

From before when we discussed the running coupling in QCD, α_s, we introduced two energy scales that occur in hadronic collisions. The first, Q^2, represents the energy exchanged between our lepton and our quark in their point-like (hard) interaction and must be relatively large, since the lepton must be able to resolve the quark within the proton (remember the inverse energy-length relationship discussed earlier). The second, Λ^2_{QCD}, is relatively small ($\sim 1\,\mathrm{GeV}^2$), and represents the characteristic nonpertubative hadronic scale of the proton (soft). So why does this separation of scales matter? If we remember a unique feature of QCD, is asymptotic freedom. That means that for high enough energies (or equivalently short enough distances), such as Q^2, the quarks in the proton behave as free particles and we can use perturbation theory to calculate the interaction between them and the incoming lepton. However, the same effect works against us for the low energy (long distance) interquark interactions within the proton, causing them to be nonpertubative and

thus only determinable through experiment. This ability to separate the pertubative and nonpertubative pieces in an experiment does not hold for all processes, but is viable for most processes present in collider environments, where $Q >> \Lambda_{QCD}$. Our factorized cross section can then be written as convolution of a nonpertubative parton distribution function (PDF), $f_i(x, Q^2)$ and a pertubative interaction piece, $(\hat{d\sigma})$.

Another advantage of factorization comes from the PDF term. As it turns out, once the PDF is determined in one experiment (e.g. DIS), that PDF remains viable for other types of experiments (e.g. DY or jet production). This allows for a simple replacement of the pertubative interaction term in the cross section formula and you are then prepared to calculate any experimental observables. (For example, we can calculate: $d\sigma_{DIS} = f \otimes \hat{d\sigma}_{DIS} \Rightarrow d\sigma_{DY} = f \otimes \hat{d\sigma}_{DY} \otimes f$ where f remains the same universal PDF in both calculations.) This universality of the PDF is important for PDF fitting, a subject discussed at length in Chapter 5. Lets take a moment to try to better understand factorization and PDFs with an analogy.

Imagine making blueberry pancakes: you start with a list of known ingredients, some flour, milk, baking powder to get a standard batter. Then you drop in your handful of blueberries and mix it all together. Then you heat up a griddle until it sizzles and pour on the batter. A little bubbling and a flip later, you stack the finished pancakes on a plate. But lets say you wanted to know exactly what each bite you took from your double stack would taste like. So lets look back at the process that got us here. We had our bowl of batter, our long range interaction, where we knew from the recipe, the Standard Model, what our pancakes should look like but there is some randomness and uncertainty there. Then we get the pour where we go from batter to pancake, this is our PDF or in this case our blueberry distribution function, where the theoretical pancake from the picture in the cookbook meets the experimental pancake that we see on our griddle. Once on the griddle the number of blueberries and their placement are set, cooked into place by our experiment. Finally we stack our two pancakes and cut into them, this is our short range interaction, where we get either a boring no blueberry bite

or a juicy double blueberry bite. What factorization gives us is the ability to separate the process of making pancakes into a batter phase and double stacked pancake phase. You could extend this idea further and imagine pouring thousands of batches of batter and keeping track of how the pancakes stack up and eventually you'd have a functional understanding of how pancake batter pours, the same idea is applied to how we derive the PDF. You could then imagine trying the same batter with different berries or even chocolate chips and take even more delicious data, each time improving upon your understanding of the PDF.

This concludes the broad study of QCD properties relevant to this work. In the next section, we were introduced to parton distribution functions formally, since PDFs represent the core component of the remainder of this dissertation, the following sections will present some of properties governing PDFs and their derivation as well as introduce nuclear PDFs.

1.3. Parton Distribution Functions

Parton distribution functions represent a vital link between experimental observables and theoretical calculations, but since they are nonpertubative, where do they come from? As we saw with the DIS structure functions before, PDFs are determined with experimental data; but, in addition to that there is an underlying theoretical framework to the PDF. The combination of these two results in PDFs being parameterized fits to experimental data. In order to do this there two elements that must be discussed: fitting and parameterization.

1.3.1. Fitting

If we imagine a series of random data points and were told to draw the "best" line to describe those data points, it could be very difficult to determine one "best" line from another and it would become even more difficult if more points were added. To combat this, a test can be designed to measure the "best-ness" of one line against some other line. This measurement, called χ^2 and shown in Eq. 1.8, forms the basis of fitting PDFs.

$$\chi^2 = \sum_i \frac{(D_i - T_i)^2}{\sigma_i^2} \tag{1.8}$$

In this function, D_i refers to a data point, σ_i refers to the error on that data point, and T_i represents the theoretical prediction for that point given our PDF. Since PDFs represent smooth functions[13] that seek to describe experimental observables, it is important that we minimize the χ^2 function to obtain the "best" fit. In order to do this we vary the PDF by modifying parameters that affect the shape of our function, but more on this in a bit. This idea of fitting a smooth function to a series of points is a bit naive though. In reality, our data points are scattered over several orders of magnitude in Q^2 as we seek to include as much data as possible. We also have data from different experiments and processes to consider. To solve these problems, we go back to QCD.

In order to compare our PDFs to data spread across a wide range of Q^2, we utilize a result related to Bjorkën scaling. First, we fit our PDFs at a low energy, $Q_0 \sim 1 \text{GeV}$, then we evolve our PDF up to the energy of the data. If scaling held perfectly for all values of x, then we would be done, e.g. $f(x, Q_0) = f(x, Q_1)$; but, as discussed earlier, that is not the case. Instead we must utilize the DGLAP evolution equation[14] which tells us how PDFs determined at Q_0 are related to PDFs evaluated at Q_1.

In order to accommodate different types of data in PDF fitting, we rely heavily on factorization. As mentioned earlier, factorization allows us to fit a universal PDF with one data set and one type of observable and then use that PDF in calculations for an entirely different purpose. For example, a PDF fit to DIS data with a very small Q^2 could be used in calculations involving DY data for very high Q^2. However, there are caveats to doing this, as the PDF might poorly describe this new data and thus not be the "best" fit it was

---

[13]This is not necessarily true for higher order PDFs, a thorough discussion of which can be found in Chapter 4.

[14]This equation can be derived from the Renormalization Group Equation. This introduces a renormalization scale, μ into our PDFs which enters the logarithmic divergences, $\ln(Q/\mu)$, in the evolution. For a single scale problem, we choose the renormalization scale to be equal to the hard scattering energy, $\mu = Q$, in order to eliminate the logarithms. More information on this can be found in [5].

thought to be. This leads to an iterative process - fit, compare to new data, and refit - that underlies PDF determination studies as discussed in Chapters 3 and 5.

1.3.2. Parameterization

One might ask though, what exactly is this function we are fitting to all these data points? In the `nCTEQ` collaboration [6] [7], we utilize a parameterization function of the form:

$$f_i(x, Q_0) = x^{-\alpha}(1 - x)^{\beta} P_i(x) \tag{1.9}$$

for each quark flavor and the gluon. Here, α and β represent positive real numbers and $P_i(x)$ represents a smooth function for the parton. For this parameterization, the values of α and β control the shape of the PDF and also ensure the momentum sum rule is not violated. The momentum sum rule simply states that the fraction of the momentum, x, of a single hadron cannot exceed the total momentum of the hadron[15], or equivalently, the sum of the momentum of the partons must be the total momentum of the hadron (see equation 1.10).

$$\int_0^1 dx \sum_i x f_i(x, Q_0) = 1 \tag{1.10}$$

Additionally, PDFs are bound by the structure of the hadrons they describe. In the case of a proton, this number sum rule guarantees that the proton PDF will have the expected valence quark[16] structure (uud), Eq. 1.11 and no net strangeness, 1.12.

$$\int_0^1 dx \ f_{u_v}(x, Q_0) = 2, \quad \int_0^1 dx \ f_{d_v}(x, Q_0) = 1 \tag{1.11}$$

$$\int_0^1 dx \ (f_s(x, Q_0) - f_{\bar{s}}(x, Q_0)) = 0 \tag{1.12}$$

---

[15]For nuclear PDFs, x can exceed 1, and is instead bound by the number of nucleons. However, most nuclear PDF analyses continue to enforce the single hadron momentum sum rule for simplicity and because corrections to the PDF for $x \geq 1$ are negligibly small.

[16]Here, u_v and d_v refer to the quarks that give the proton its quantum numbers and can be represented by $u - \bar{u}$ and $d - \bar{d}$, respectively.

While these sum rules are needed to ensure that the PDF accurately depicts nature, they also require some knowledge of the partonic structure of the proton. For example, the gluon momentum is generally fit using the remaining momentum from the momentum sum rule which then requires an accurate accounting of the momenta from all quark flavors. This differentiation needs to then come from different types of data and different observables. As we saw in Eq. 1.6, the DIS structure functions, when combined in certain ways, can provide some insight. However, some of these structure functions describe neutrino DIS that require heavy nuclei as targets in order to get reasonable cross sections.

1.4. Nuclear Corrections

From chemistry, we know that nuclei of heavy elements are constructed from combinations of protons and neutrons. However, as discovered by the European Muon Collaboration (EMC) and illustrated in Fig. 1.4, the structure of the nucleus as it pertains to partons is much more complicated than simple addition. EMC found that the F_2 structure function (and correspondingly the PDF) for iron ($^{56}_{26}$Fe), was suppressed for certain values[17] of x. This effect, later known as the EMC effect, describes a suppression of the heavy nuclear cross section as compared to the lighter deuterium ($^{1}_{2}$D).

This discovery and others, prompted the study of nuclear PDFs (nPDFs) to try to determine the structure of these heavy nuclei as a function of their atomic mass, A. Currently, there are two schools of thought for determining this A-dependence: the first is to rescale any proton PDF with an A-dependent multiplicative factor (Eq. 1.13) and the second is to construct an individual nucleus for each element by fitting a bound proton PDF for that element (Eq. 1.14).

$$f_i^{(A)}(x, Q^2) = R_i^A(x, Q^2) f_i^p(x, Q^2) \tag{1.13}$$

---

[17]The EMC saw this effect only for $0.3 < x < 0.8$. However, there are nuclear effects resulting from different sources present in other regions of x. Shadowing, $x < 0.1$, refers the suppression of the nuclear PDF that results from destructive interference between the nucleon and the produced virtual boson. Conversely, anti-shadowing represents an enhancement due to constructive interference, $0.1 < x < 0.3$. Fermi motion, $0.8 < x$, is due to the quantum motion of the nucleons.
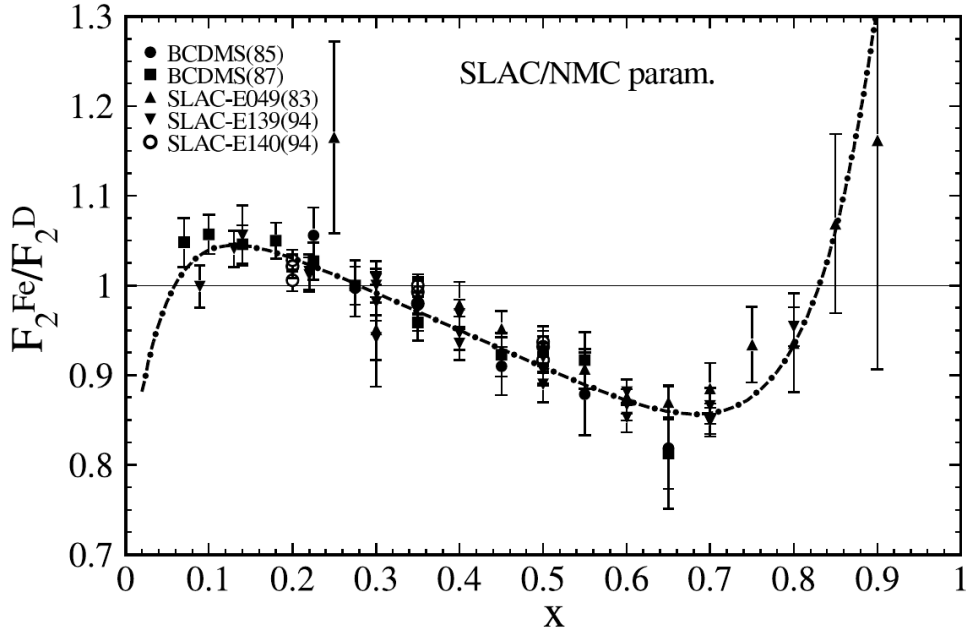
Figure 1.4: The nuclear correction ratio as a function of x compared to SLAC and BCDMS data [8] [6]. There are four regions of this plot representing:
(shadowing) $< 0.1 <$ (anti-shadowing) $< 0.3 <$ (EMC effect) $< 0.8 <$ (Fermi motion).

The first method, preferred by fitting groups such as EPS [9] and HKN [10], allows for any PDF to be used as the base proton PDF. This allows these groups to simply release their multiplicative ratios. However, this method can introduce theoretical errors into the nPDF that are not solely a result of the underlying PDF, instead being a combination of the PDF's errors and the error introduced by the ratio. It also requires that predictions be calibrated when using different proton PDFs, else it can be difficult to tease out whether results are an effect of the nuclear correction or due to a difference in base PDF.

$$f_i^{(A,Z)}(x, Q^2) = \frac{Z}{A} f_i^{p/A}(x, Q^2) + \frac{A-Z}{A} f_i^{n/A}(x, Q^2) \tag{1.14}$$

The second method is utilized by the `nCTEQ` collaboration and will be the applied method throughout this dissertation. This type of nuclear correction relies on fitting a bound proton nPDF for a given nuclei, then constructing a bound neutron nPDF by utilizing approximate

isospin symmetry[18]. The bound proton PDF for a specific nuclei is generated by generalizing Eq. 1.9 such that $\alpha \to \alpha(A)$ and $\beta \to \beta(A)$. From there a nucleus is constructed with Z protons and $A - Z$ neutrons. The collaboration can then release the set of bound proton nPDFs and the set of fully constructed nPDFs. This has advantages of consistency when making predictions and the ability for a user to custom build nuclei or isotopes of their choice but lacks the flexibility to change the base PDF offered by the first method.

The study of nPDFs and heavy nuclear interactions will continue in Chapter 3, where I discuss an attempt to understand how the inclusion of newly released data from lead ion runs at the Large Hadron Collider will affect the `nCTEQ15` nPDF set using Bayesian reweighting, and again in Chapter 5, where I include this data into a new `nCTEQ` fit.

In the following chapter, I detail a `Mathematica` package, `ManeParse`, that is written to aid in the study of PDFs and nPDFs by allowing users access to PDFs within the `Mathematica` environment without the use of outside software packages. The user is then able to perform calculations with the PDFs and visualize them using `Mathematica`'s comprehensive mathematical and plotting libraries.

Chapter 4 does not directly relate to nPDFs but instead relates to my work with the `xFitter` collaboration in the study of PDF fitting. This work examined the use of varying the number of quark flavors used in PDF fitting as a function of energy by manipulating the point at which a given flavor would become active rather than assuming the traditional value of the quark mass.

The final chapter provides a summary of my work as well as applications for my findings from Chapter 5. Additionally, I present possible avenues for future work in the field of PDF research.

---

[18]Isospin symmetry imposes $f_u^{p/A} \Leftrightarrow f_d^{n/A}$ and $f_{\bar{u}}^{p/A} \Leftrightarrow f_{\bar{d}}^{n/A}$

Chapter 2

MangParse

As discussed in Chapter 1, Parton Distribution Functions (PDFs) are essential non-perturbative inputs for calculation of any observable with hadronic initial states. These PDFs are released by individual groups in discrete grids as a function of the Bjorken-x and energy scale Q. The LHAPDF project maintains a repository of PDFs from various groups in a new standardized `LHAPDF6` format; additionally older formats such as the `CTEQ` PDS grid format are still in use.

In this chapter, I present `ManeParse`[1], a package that provides access to PDFs within `Mathematica` to facilitate calculation and plotting. The program is self-contained so there are no external links to any `FORTRAN`, `C` or `C++` programs. The package includes the option to use the built-in `Mathematica` interpolation or a custom cubic Lagrange interpolation routine which allows for flexibility in the extrapolation (particularly at small x-values). `ManeParse` is fast enough to enable simple calculations (involving even one or two integrations) in the `Mathematica` framework.

`ManeParse` greatly improves the ability for a user to visualize a PDF in a straightforward way. This simplicity, combined with a flexible and robust set of parsing routines, makes for a very powerful tool when prototyping PDF analyses and validating results. As such, `ManeParse` was used several times to that effect thought this dissertation. Additional information about obtaining and using `ManeParse` can be found in Appendix A.

In the development of `ManeParse`, I worked as part of a small team consisting of myself, Ben Clark and Fred Olness. When I use "we", it is this team to which I am referencing. As part of this team, I was responsible for designing and writing the `LHAPDF6` parsing routine (`pdfParseLHA` and the `pdfCalc` package as well as contributing on the `pdfError` package. I

---

[1]Available for download: https://ncteq.hepforge.org/mma/index.html

was also solely responsible for the creation and maintenance of `Mathematica` documentation, external users manual and downloadable demo. I am also the primary developer for fixing bugs and adding features for updates to the package.

2.1. What is `ManeParse`?

Parton Distribution Functions (PDFs) are essential elements for making predictions involving hadrons (protons and nuclei) in the initial state. For example, at the LHC, we can compute the Higgs production cross section (σ) using the formula $\sigma_{pp \to H} = \sum_{a,b} f_{a/P} \otimes f_{b/P} \otimes \omega_{ab \to H}$ where PDFs $f_{a/P}$ and $f_{b/P}$ give the probability density for finding partons "a" and "b" in the two proton beams, and the hard cross section, $\omega_{ab \to H}$, gives the probability density for partons a and b producing the Higgs, H. The PDFs cannot be computed from first principles at this time, so they must be extracted using fits to experimental data.[2] This analysis is performed by a number of collaborations, and the PDFs are generally distributed as a grid of values in x and Q which must be interpolated to generate the PDF $f_{a/P}(x, Q)$ for flavor "a" in hadron "P" at momentum fraction x and energy scale Q.

`ManeParse`[3] is a flexible, modular, lightweight, stand-alone package used to provide access to a wide variety of PDFs within `Mathematica`. To illustrate the flexibility, in Fig. 2.1 we show how `ManeParse` can work simultaneously with different PDF sets from a variety of groups.[4] This figure displays the selected PDF sets listed in Table 2.3. Some of the sets are in the `LHAPDF6` grid format [13], and others are in the older PDS grid format. [14] These sets also have different numbers of active flavors, N_F, different values for the initial evolution scale, Q_0, different values for the heavy quark masses, $\{m_c, m_b, m_t\}$, and they can represent either free protons or protons bound in nuclei. Nevertheless, `ManeParse` is able to easily compare and contrast sets from different groups in a common framework.

---

[2]Lattice QCD has made great strides in computing PDFs in recent years, however their efforts remain too preliminary to be relevent to this work. [11, 12]

[3]The `ManeParse` program was originally developed to run on the SMU computing cluster "ManeFrame" which is a play on words inspired by the school mascot, Peruna the pony.

[4]All plots presented here have been generated in `Mathematica`.
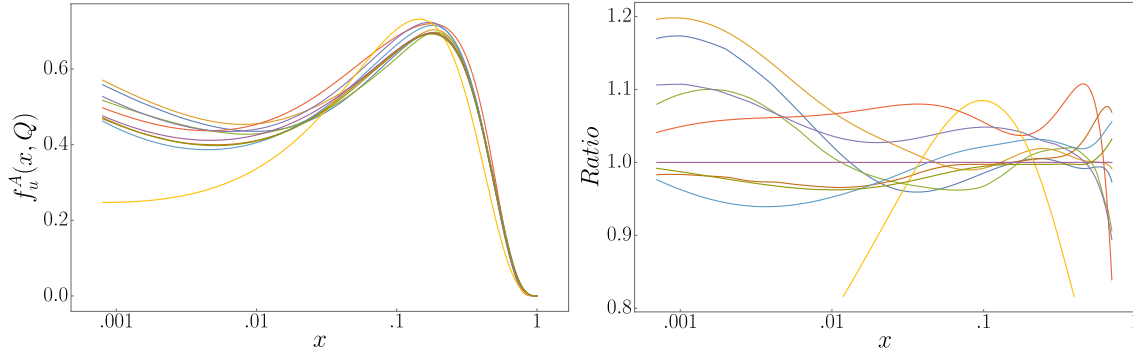
Figure 2.1: a) We display $x f_u^A(x, Q)$ for the up-quark at $Q = 2$ GeV as a function of x for the 10 PDFs listed in Table 2.3. b) We display the ratio of the PDFs in a) compared to CT10 proton PDF ($A = 1$) as a function of x. While we don't identify them individually, the one curve (yellow) that distinctly deviates from the others is the nuclear PDF for lead $A = 208$.

As `ManeParse` is a stand-alone code, this complements a number of other available programs such as the QCDNUM program, [15] the APFEL program including the web-plotter,[5] the Transverse Momentum Dependent (TMD) distributions plotter hosted at DESY,[6] and also the Durham HepData online PDF plotting and calculation tool.[7] The online tools provide the ability to quickly plot PDFs, ratios, and luminosities. Then with `ManeParse`, it is easy to take the next step and compute cross sections and other user-selected quantities in the `Mathematica` environment.

In this dissertation, we describe the key features of `ManeParse` available to the user. In Section 4.5, we sketch a minimal example of how the program is used. In Section 2.3, we provide some details of how the PDFs are parsed, stored and interpolated. In Section 2.4, we display some example plots that are easily constructed using `ManeParse`. In Section 2.5, we provide examples of the functions in the `pdfError` module. Finally, we discuss files provided by `ManeParse` and how to obtain the external PDF files.

---

[5]Details can be found in Ref. [16] and online at: http://apfel.mi.infn.it/

[6]Details can be found in Refs. [17, 18] and online at: http://tmdplotter.desy.de/

[7]Details can be found online at: http://hepdata.cedar.ac.uk/pdf/pdf3.html

2.2. A simple example

We begin by outlining a simple example of how `ManeParse` may be used. After loading the `ManeParse` packages into `Mathematica`, the user can enter the following commands:

```
Get[pdfParseLHA.m]
iSet1=pdfParseLHA[LHA_file.info,LHA_file.dat]
pdfFunction[iSet1,iParton,x,Q]


Get[pdfParseCTEQ.m]
iSet2=pdfParseCTEQ[PDS_file.pds]
pdfFunction[iSet2,iParton,x,Q]
```

The first and fourth line load the parsing subpackages included in `ManeParse`. Loading either of these, causes the `pdfCalc` package to be loaded as well. The second line reads an `LHAPDF6` formatted external data file (`LHA_File.dat`) and its associated information file (`LHA_File.info`), and generates an internal PDF set that is referenced by the integer `iSet1`. The fifth line reads a PDS formatted external data file[8] (`PDS_File.pds`) and generates an internal PDF set that is referenced by the integer `iSet2`.

After reading these data files, the user is provided with the core function for computing the PDFs: `pdfFunction[iSet,iParton,x,Q]`. Here, `iSet` selects the individual PDF set, `iParton` selects the parton flavor as shown in Table 2.1, and $\{x,Q\}$ specify the momentum fraction, x, and the energy scale, Q, in GeV.

`pdfFunction` performs the bulk of the work for the `ManeParse` program, so the package has been optimized for speed to make it practical to perform single or double integrals in a reasonable amount of time; specifically, the `pdfFunction` call generally takes less than 1 ms per core on a standard laptop or desktop.

---

[8]Note that the `LHAPDF6` files have both a data file and an info file whereas the older `CTEQ` PDS files have only a data file.

Additionally, `ManeParse` can handle an arbitrary number of PDF sets and can switch between sets without delay. When the external PDF file is parsed, the data is stored internally (about 1 Mb per PDF set) and the `iSet` variable essentially functions as a pointer to the set; thus, it is trivial to loop over many PDF sets as was done in Fig. 2.1. This feature contrasts to some of the older `FORTRAN` programs, which could only store a fixed number of sets in memory and often had to re-read the data files.

These are the key elements of the package, however, we also provide many auxiliary functions described below. Consistent with the `Mathematica` convention, all our public functions begin with the prefix "`pdf`". One can obtain a complete list with the command `?pdf*`. The usage message for individual functions is displayed in a similar manner to:

`?pdfFunction`

pdfFunction[setNumber, flavor, x, Q]

- This function returns the interpolated value of the PDF for the `.pds`/`.dat` file specified by *setNumber*, for the given *flavor* and value of Bjorken x and scale Q.

- *Warning*: The results of this function are only reliable between the maximum and minimum values of x and Q in the `.pds`/`.dat` file.[9]

2.3. Inside the `ManeParse` Package

2.3.1. Overview of package

`ManeParse` internally consists of four modules (or sub-packages) as illustrated in Fig. 2.2. The modular structure of `ManeParse` allows for separate parsers for the `LHAPDF6` (`pdfParseLHA`) and PDS (`pdfParseCTEQ`) grids which read the individual file types and pass the information on to a common calculation (`pdfCalc`) module.

---

[9]If interpolation outside the given grid is requested by the user, `ManeParse` is equipped to handle this. The `Mathematica` interpolator will throw a warning message and proceed to use built-in extrapolation techniques. The `ManeParse` interpolator will extrapolate using the behavior defined with `pdfSetXpower`.

flavor #	0 or 21	±1	±2	±3	±4	±5	±6
parton	gluon	down/dbar	up/ubar	strange/sbar	charm/cbar	bottom/bbar	top/tbar

Table 2.1: The standard Monte Carlo (MC) flavor numbering convention [19] used within `ManeParse`. This differs from the mass-ordered convention used in many older `CTEQ` releases. `ManeParse` converts these releases into the MC ordering.
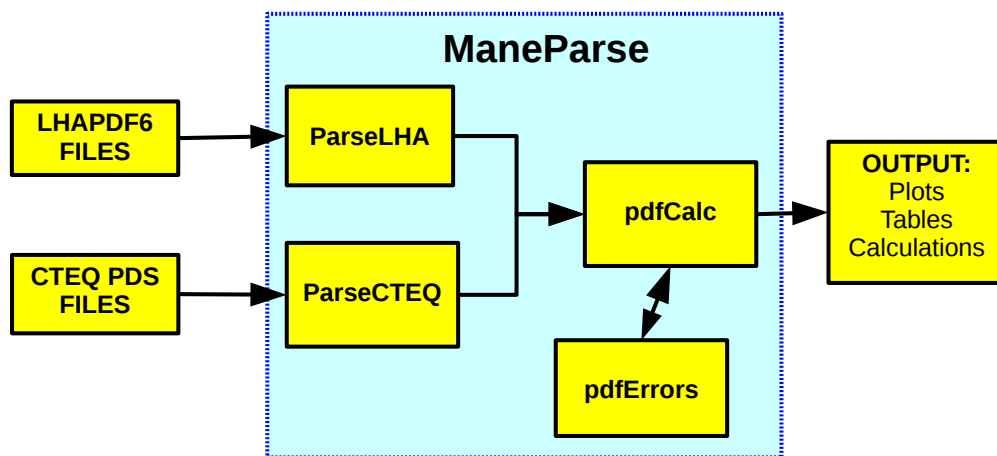
Figure 2.2: A schematic overview of the `ManeParse` package and the individual modules.

The new `LHAPDF6` format is intended as a standard that all groups can use to release their results. Additionally, many older PDF sets have been converted into this format.

The `ManeParse` modular structure provides flexibility, as the user can use both `LHAPDF6` and PDS format, or even write a custom parser to read a set that is not in one of these formats.

The error PDFs module (`pdfError`) uses `pdfCalc` to construct PDF uncertainties, luminosities, and correlations as illustrated in Sec. 2.5.

The key elements of each PDF set include the 3-dimensional $\{x, Q, N_F\}$ grid and the associated information, which is stored as a set of `Mathematica` rules. We now describe the features and some details of these structures.

2.3.2. The PDF $\{x, Q, N_F\}$ grid

The parsing routines `pdfParseLHA` and `pdfParseCTEQ` read the external files and assemble the PDF sets into a common data structure that is used by the `pdfCalc` module. The central structure is a 3-dimensional grid of PDF values in $\{x, Q, N_F\}$ space, which uses vectors $\{x_{vec}, Q_{vec}\}$ to specify the grid points. The spacing of $\{x_{vec}, Q_{vec}\}$ need not be uniform; typically, Q_{vec} uses logarithmic spacing, and x_{vec} is commonly logarithmic at small x and linear at large x. Different spacings in x_{vec} and Q_{vec} do not pose a problem for the `pdfCalc` package, as the grid points are simply interpolated to provide the PDF at a particular point in $\{x, Q, N_F\}$. The user is agnostic to the specific grid spacing chosen in a PDF release.

2.3.2.1. N_F Convention

The N_F flavor dimension is determined by the `iSet` value passed to `pdfFunction`. The association between the grid slice in N_F and `iSet` is specified in the `LHAPDF6` info file using the "`key:data`" format such as "`Flavors:` $[-5, -4, -3, -2, -1, 1, 2, 3, 4, 5, 21]$". This tells us which partons are in the grid, and their proper order.[10] Note: we use the standard Monte Carlo (MC) convention[11] throughout `ManeParse` where $d = 1$ and $u = 2$ rather than the mass-ordered convention (see Table 2.1).[12] The standard MC convention also labels the gluon as `iParton` $= 21$; for compatibility, the gluon in `ManeParse` can be identified with either `iParton` $= 21$ or `iParton` $= 0$.

`ManeParse` is able to work with PDF sets with different numbers of flavors. For example, in Fig. 2.1, the NNPDF set includes $N_F = 6$ where `iParton` $= \{\bar{t}, \ldots, t\}$, while most of the other sets have $N_F = 5$. If a flavor, `iParton`, is not defined, `pdfFunction` will return zero. This feature allows the user to write a sum over all quarks $\sum f_i(x, Q)$ for $i = \{-6, \ldots, 6\}$ without worrying whether some PDF sets might have less than 6 active flavors.

---

[10]For the PDS files, this information is contained in the header of the data file so there is not a separate info file; `pdfParseCTEQ` extracts the proper association.

[11]See Ref. [19] "Review of Particle Physics," Chapter 34 entitled "Monte Carlo particle numbering scheme."

[12]Caution is required here as many of the older `CTEQ` releases use the mass-ordered convention with $u = 1$ and $d = 2$. `ManeParse` converts these mass-ordered sets into the MC ordering.

Additionally, the `ManeParse` framework has the flexibility to handle new particles such as a 4$^{\text{th}}$ generation of quarks with `iParton` $= \{b', t'\} = \{7, 8\}$ or a light gluino with `iParton` $= \tilde{g} = 1000021$ PDF by identifying the flavor index, `iParton`, with the appropriate grid position in the `LHAPDF6` info file.

2.3.2.2. Q Sub-Grids

At NNLO and beyond, the PDFs can become discontinuous across the mass flavor thresholds. This is illustrated using the NNLO MSTW set in Fig. 2.3 where we observe a discontinuity of both the gluon and b-quark PDF across the b-quark threshold at $m_b = 4.75$ GeV. `ManeParse` accommodates this by using sub-grids in Q as illustrated in Fig. 2.4-a); for example, we use separate grids below and above the threshold at $Q = m_b = 4.75$ GeV. When we call the PDF at a specific Q value, `ManeParse` looks up the relevant heavy quark thresholds, $\{m_c, m_b, m_t\}$, to determine which sub-grid to use for the interpolation. For $Q < m_b$, sub-grid #2 ($N_F = 4$) is used, and for $Q \geq m_b$, sub-grid #3 ($N_F = 5$) is chosen.

Note that for the x value (10^{-4}) displayed in Fig. 2.3, the b-quark PDF is negative for Q just above m_b; this is the correct higher-order result and justifies (in part) why we do not force the PDFs to be positive definite. This behavior also makes sense in terms of the momentum sum rule, which we will discuss in Sec. 2.4.3.

2.3.2.3. An N_F-dependent PDF: $f(x, Q, N_F)$

Note, the use of sub-grids in Q also enables the use of overlapping N_F ranges as in a hybrid scheme as described in Ref. [20]; in this case, we generalize the PDF so that it also becomes a function of the number of flavors: $f(x, Q, N_F)$. This feature is useful if, for example, we are performing a fit to data in the region $Q \sim m_b$; we can perform a consistent $N_F = 4$ flavor fit even if some of the data are above the $N_f = 5$ threshold ($Q > m_b$) by selecting $f(x, Q, N_F = 4)$; thus, we avoid encountering any discontinuities in the region of the data.[13] We illustrate this generalized case for $f(x, Q, N_F)$ in Fig. 2.4-b). Here, the user

---

[13]Note that the APFEL PDF evolution library [21] is in the process of implementing these features.
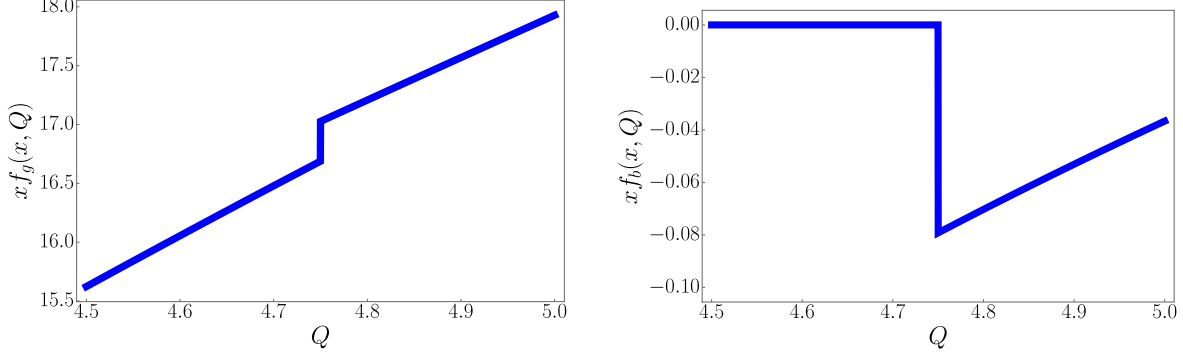
Figure 2.3: The discontinuity of the gluon (left) and b-quark (right) PDFs across the $m_b = 4.75$ GeV flavor threshold; the horizontal axis is Q (in GeV), and the vertical axis is $x\,f(x,Q)$. The curves are for the MSTW2008nnlo68cl PDF with $x = 10^{-4}$. Note that the gluon and b-quark shift in opposite directions to ensure the momentum sum rule is satisfied.
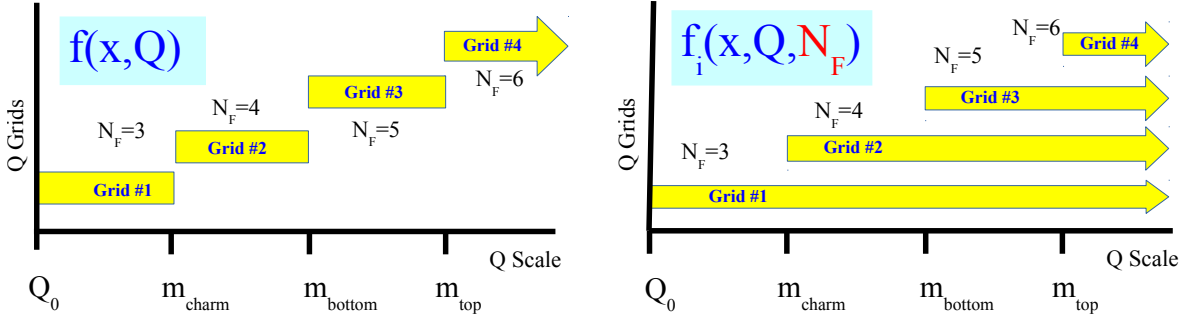
Figure 2.4: A schematic representation of the Q sub-grids used to handle discontinuities across the heavy quark thresholds at $\{m_c, m_b, m_t\}$. Fig. a) shows the conventional arrangement for $f(x,Q)$ with non-overlapping sub-grids; for a given Q, the N_F flavor dimension is uniquely determined. Fig. b) shows a flavor-number-dependent PDF $f(x,Q,N_F)$ where the user has the freedom to choose the N_F flavor dimension value (and hence the sub-grid).

has the freedom to choose the active number of flavors, N_F, rather than being forced to transition at the quark mass values as in Fig. 2.4-a).

2.3.3. The LHAPDF6 Info File

In addition to the 3-dimensional $\{x, Q, N_F\}$ grid, there is auxiliary material associated with each PDF set. In the LHA format, each PDF collection has an associated "info" file which contains the additional data in a YAML format,[14] whereas in the CTEQ PDS format

---

[14] "YAML Ain't Markup Language" http://yaml.org/

YAML	Mathematica
key: "data"	"key" → "data"
SetDesc: "nCTEQ15 ..."	"SetDesc" → "nCTEQ15 ..."
NumFlavors: 5	"NumFlavors" → 5
Flavors: [-5,-4,-3,-2,-1,1,2,3,4,5,21]	"Flavors"→{-5,-4,-3,-2,-1,1,2,3,4,5,21}
AlphaS_Qs: [1.299999e+00, ...]	"AlphaS_Qs"→ $\{1.299999 \times 10^{+00}, ...\}$
UnknownKey: data	"UnknownKey"→"data"

Table 2.2: Sample YAML entries contained in the `LHAPDF6` info file, and the corresponding rules passed to `Mathematica`. The rules for a specific PDF set are obtained using the `pdfGetInfo[iSet]` function.

files, the auxiliary information is contained at the top of each PDS data file. Each parser interprets this information and builds a list of `Mathematica` rules.

The basic syntax of YAML is [`key:` `''data'']`, and the LHA parser converts this into a `Mathematica` rule as {`''key''→''data''`}. This can be viewed within `ManeParse` using the function `pdfGetInfo[iSet]`, and Table 2.2 demonstrates the some sample mappings between the two.

If "`key`" is known to be a number, "`data`" is converted from a string into a number. This behavior applies to values such as {NumFlavors, QMin, MTop, ...}. If "`key`" is known to be a list such as {Flavors, AlphaS_Qs}, "`data`" is converted from a string into a `Mathematica` list. If "key" is unknown, "`data`" is left as a string. This means that `ManeParse` can handle any unknown "`key`", and the user can modify these rules after the fact, or introduce a custom modification by identifying "`key`" to the parser.

2.3.4. Interpolation

Once the 3-dimensional $\{x,Q,N_F\}$ grid and auxiliary rules are given to the `pdfCalc` module, we are ready to interact with the PDFs. When the user calls for $f_i(x,Q)$, the `pdfCalc` module will determine the appropriate N_F index and Q grid and do a 4-point

interpolation in the 2-dimensional $\{x, Q\}$ space. For the interpolation we use a 4-point Lagrange interpolation given by:[15]

$$g(x) = c_0(x)\, y_0 + c_1(x)\, y_1 + c_2(x)\, y_2 + c_3(x)\, y_3$$

where $y_k = g(x_k)$ are the PDF values at the grid points, and the coefficients are given by:

$$c_j(x) = \prod_{0 \leq m \leq 3}' \frac{(x - x_m)}{(x_j - x_m)}$$

where the prime $(')$ indicates the restriction $j \neq m$ in the product. This formula has the feature that the interpolated curve will always contain the grid points $\{x_i, y_i\}$. The grid points do not need to be equally spaced.

To perform the 2-dimensional interpolation, we extract a 4×4 sub-grid in $\{x, Q\}$ space; we first compute 4 interpolations in x-space, and then use these to perform a 4-point interpolation in Q-space. Generally, `pdfCalc` will interpolate $\{x, Q\}$ values with 2 grid points on each side, but at the edges of the grid, it will use a 3-1 split. It also will extrapolate beyond the limits of the grid and will return a number, even if it is unphysical. Except for setting $f_i(x, Q) = 0$ for $x > 1$, we do not check bounds, as this would slow the computation; in the sample files, we do provide examples of how the user can implement particular boundaries if desired.

Additionally, we allow the interpolated PDF to be negative. At very large x this can happen due to numerical uncertainty, but there are also instances where a negative PDF is the physical result, such as at NNLO (illustrated in Fig. 2.3). Within `Mathematica`, it is easy for the user to impose particular limits (*i.e.* positivity) if desired. The interpolation can be performed either with the `Mathematica Interpolate` function (default) or a custom 4-point Lagrange interpolator and is set with the `pdfSetInterpolator` function. We set the `Mathematica Interpolate` function as the default, as it is slightly faster, but the custom 4-point Lagrange interpolator often will provide better extrapolation of the PDFs beyond

---

[15]We present the interpolation formulas in the x-variable; an equivalent form is used for the Q interpolation.

the grid boundaries and has some adjustable parameters which are useful in the small x region.

The PDF typically increases as $1/x^a$ at small x where $a \sim 1.5$; thus, we can improve the interpolation by scaling the PDF by a factor of x^a which is implemented by replacing $y_k \rightarrow x^a \, g(x_k)$ and $g(x) \rightarrow g(x)/x^a$. This is why many of the PDF programs fundamentally compute with $xf(x)$ rather than $f(x)$. To return $f_i(x, Q)$ we divide by x, but to avoid dividing by zero we internally impose a default minimum x value of $x_{min} = 10^{-30}$. The default scaling factor for the the custom interpolator is $a = 1$, but this can be set with the `pdfSetXpower[a]` function.

2.3.4.1. Interpolation Quality

By construction, the interpolation curve will always intersect the grid values: $g(x_i, Q_j) \equiv f(x_i, Q_j)$ if x_i and Q_j are grid points. Therefore, the numerical uncertainty arises from how we connect these grid points. We have bench-marked many of the PDF sets to ensure our interpolations are accurate across the defined grid in $\{x, Q\}$ space. For the PDS files, our interpolation (with scaling $1/x^a$ for $a = 1.0$) uses the same algorithm as the benchmark `CTEQ FORTRAN`, so our results easily match to better than one part in 10^3. The `LHAPDF6` interpolation uses a logarithmic bi-cubic interpolation in the central region, and switches to linear near the grid boundaries.[16] To illustrate the range of numerical uncertainty, we will show how the interpolation changes as we vary the a power. We will also compare with the built-in `Mathematica` interpolator. If a different interpolation is required, the a-parameter can be tuned, or the user can supply a custom interpolation routine.

In Fig. 2.5, we show the ratio of the interpolated value for the gluon PDF compared to the default `Mathematica` interpolation. We select a Q value which is precisely a grid point, and then show the variation as a function of x between these grid values. Figs. 2.5a, 2.5b, 2.5c show the results for three ranges of x, {small, mid, large}, while Figs. 2.5d, 2.5e show the

---

[16]`LHAPDF6` has validated a number of PDF sets, and these generally match both our interpolator, with $a = 1$, and the `Mathematica` interpolator to 1 part in 10^{-3}.
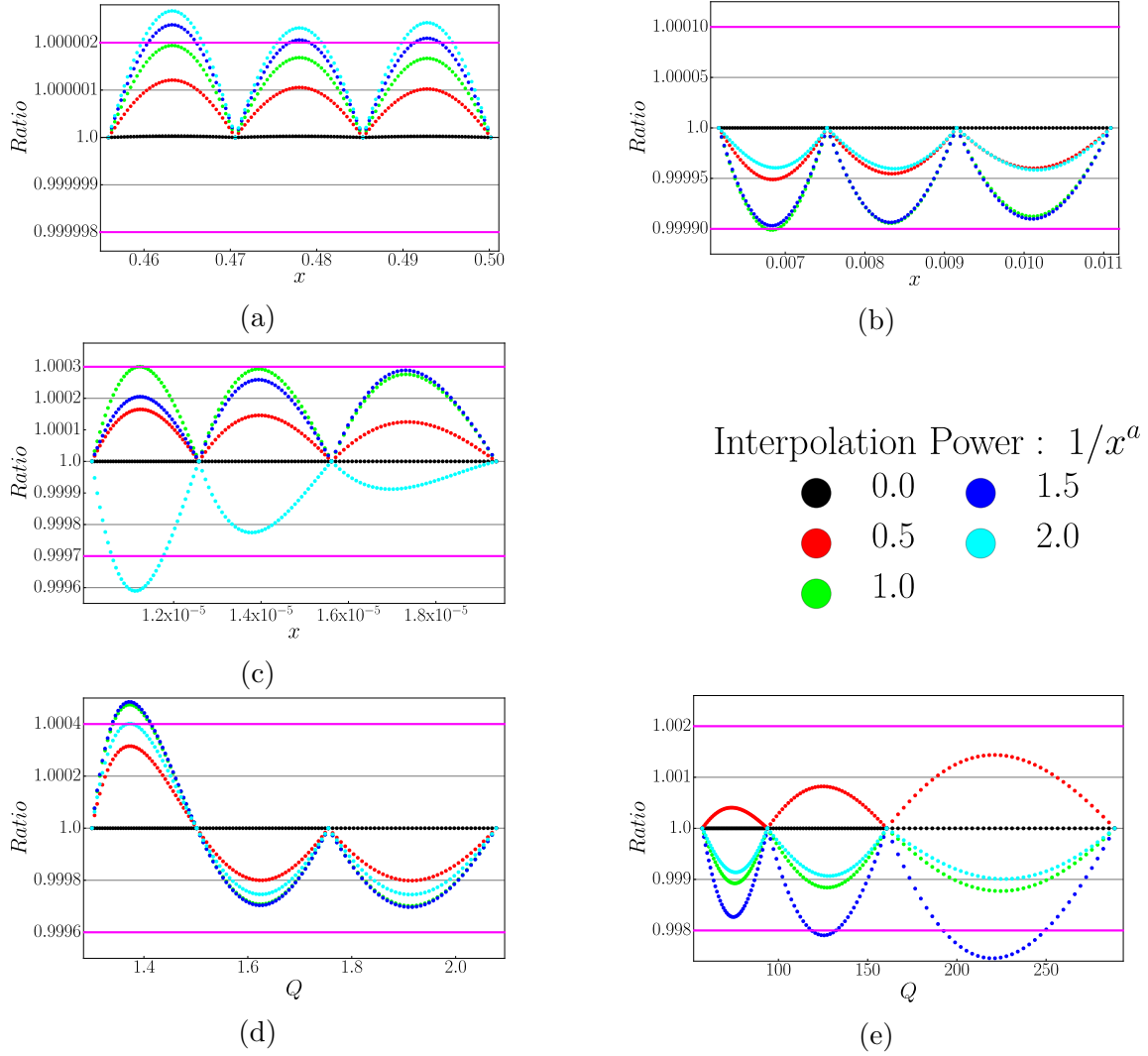
Figure 2.5: We show the numerical variation of the interpolation for the CT10 central set by presenting the ratio of the `ManeParse` interpolator with different a values to the `Mathematica` interpolator. The range of the CT10 grid is $x = [10^{-8}, 1]$ with 150 points, and $Q = [1.3, 34515]\, GeV$ with 24 points. In (a)-(c), we display the variation in x for fixed $Q = 1.3\, GeV$ (which is a grid value). In (d)-(e), we display the variation in Q for fixed $x = 0.0110878$ (which is a grid value). We have drawn horizontal guide-lines to indicate the approximate numerical variation. The ratios were plotted as points rather than lines to avoid any line-smoothing of the graphics output.

results for small Q and mid Q. In all five plots, we observe that the interpolated curves match exactly at the grid values (x_k), as they should. In between the grid values, we see there is a variation depending on the details of the interpolation and the particular value of the scaling power a. We have varied the scaling power over the range $a =\{0.0,0.5,1.0,1.5,2.0\}$. The scaling power $a = 0$ matches with the default `Mathematica` interpolation routine, while $a = 1$ compensates for the $1/x$ PDF behavior at small x.

In Fig. 2.5a, we observe that the variation is quite small in the large x range $(x \sim 0.5)$, of order $\sim 3 \times 10^{-6}$. For many calculations, such as Higgs and W/Z boson production, the mid x range of $(x \sim 0.01)$, seen in Fig. 2.5b, is the most relevant region and here we find the variation to be a bit larger, of order $\sim 1 \times 10^{-4}$. At the small x range $(x \sim 10^{-5})$, Fig. 2.5c, we find the largest variation which can be of order $\lesssim 10^{-3}$; this is partly because the PDFs are diverging in the limit $x \to 0$, so the relative error increases.[17]

We now investigate the quality of the interpolation in the Q variable. In 2.5d, we show the small Q range, $(Q \sim Q_0)$. Here, the steps in Q are about 20% apart and we see the variation is of order $\sim 5 \times 10^{-4}$. At the larger Q range in 2.5e, the steps in Q are up to 100% apart and we see the variation is of order $\sim 10^{-3}$; if increased accuracy is required here, the obvious solution would be to include more grid points in Q.

In general, we expect $a = 1$ yields the best representation of the PDFs, and the spread between $a = 0$ and $a = 1$ is a reasonable estimate of the uncertainty. Computing the momentum sum rule (*c.f.*, Table 2.3) can also provide a useful check.

We find that ratios of PDFs are more sensitive to the interpolation than the PDFs themselves. For illustrative purposes, in Fig. 2.6, we show an example of a poor interpolation generated with the `Mathematica` interpolator compared to a good interpolation by the custom 4-point Lagrange interpolation with the default $a = 1$ scaling; in general, we find the custom 4-point Lagrange interpolation computes smoother ratios and provides better extrapolation beyond the grid limits.

---

[17]The PDFs typically exhibit a rise at small x of the form $1/x^a$. At smaller Q values, the exponent is commonly slightly larger than 1, and increases with increasing Q toward an asymptotic limit in the range $a \sim [1.5, 1.7]$. Note, the momentum sum rule requires $a < 2$. [22]
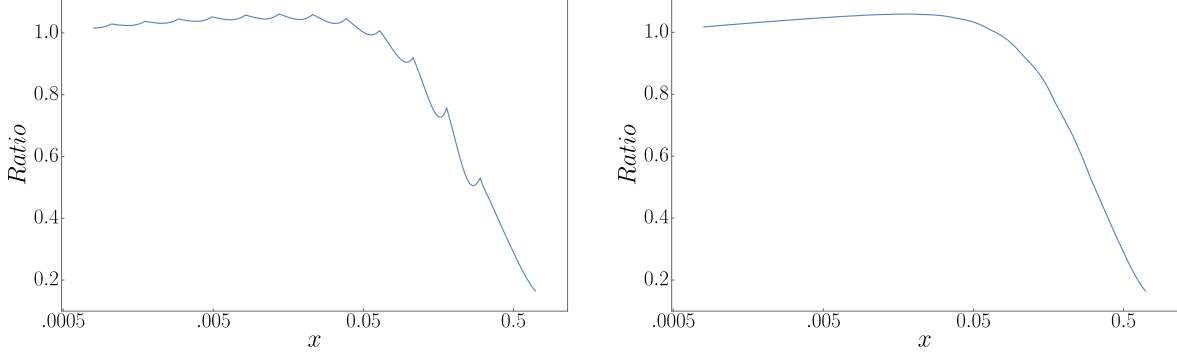
Figure 2.6: The ratio of PDFs can sometimes lead to interpolation problems; we display the ratio of two gluon PDFs at $Q = 100$ GeV. Fig.-a) on the left was generated with the default `Mathematica` interpolator, and Fig.-b) on the right was generated with the custom 4-point Lagrange interpolation with the default scaling of $a = 1$.

2.3.5. α_S Function

For some of the PDF sets, the value of $\alpha_S(Q)$ is provided as a list of points associated with Q_{vec}. For these sets, we interpolate $\alpha_S(Q)$ to provide a matched function called `pdfAlphaS[iSet,Q]`; this is displayed in Fig. 2.7 for a sample PDF set.[18] The `pdfGetInfo[iSet]` function will display the information associated with the corresponding PDF set (including any α_S values). If the PDF set does not have any α_S information, the `pdfAlpha` function will return `Null`. In Fig. 2.7-a) we display $\alpha_S(Q)$ for the NNPDF set, and in Fig. 2.7-b) we enlarge the region near $m_b = 4.18$ GeV to display the discontinuity. In general, $\alpha_S(Q)$ will be discontinuous at NNLO and higher and at all mass thresholds, $\{m_c, m_b, m_t\}$.

2.4. Sample Plots & Calculations

The advantage of importing the PDF sets into Mathematica is that we have the complete set of built-in tools that we can use for calculating and graphing. We illustrate some of these features here.

---

[18]Since at Leading Order (LO), $\alpha_S(Q) = 1/[\beta_0 \ln(Q^2/\Lambda^2)]$, we obtained improved results by interpolating in $1/\alpha_S(Q)$.
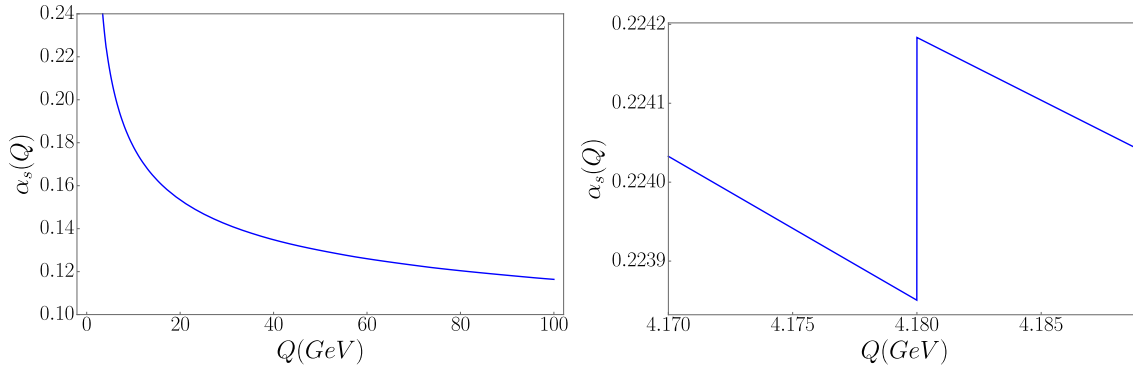
Figure 2.7: $\alpha_S(Q)$ vs. Q in GeV from NNPDF. Note the discontinuity across the $m_b = 4.18$ GeV threshold which is enlarged in Fig. b).

2.4.1. Graphical Examples

To highlight the graphical capabilities, in Figure 2.8 we display a selection of PDFs using both linear (left) and log (right) scale. Using the flexible graphics capabilities of `Mathematica` it is easy to automatically generate such plots for different PDF sets.

2.4.2. Small x Extrapolation

Sometimes it is useful to extrapolate to low x values beyond the limits of the PDF grid; for example, the study of high energy cosmic ray experiments that use very small x extrapolations. [23, 24] We provide the command `pdfLowFunction[iSet,iParton,x,Q,power]` which allows the user to choose the extrapolation power in the small x region.[19] An example is displayed in Fig. 2.9 for the nCTEQ15 proton PDF. The minimum x value for this set for the grid is $x_{min} = 5 \times 10^{-6}$; beyond this limit `pdfLowFunction` will extrapolate using the form $1/x^a$. In this example, we vary the power from 0.4 to 1.6; using the `Mathematica` integration routines it is easy to find that this range of variation in the small x behavior will only change the momentum fraction of the gluon by 0.5%.

---

[19]The "`a`" argument is optional; the default power is 1.0. We use a separate function `pdfLowFunction` so as not to slow the computation of `pdfFunction`.
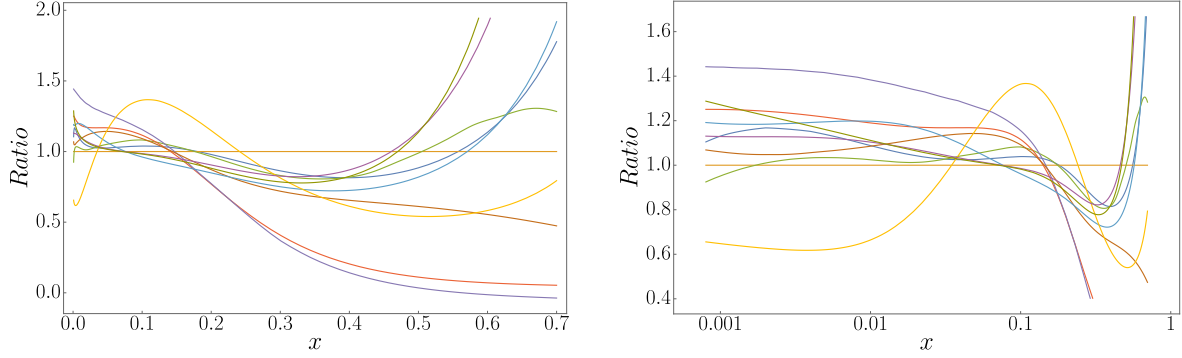
Figure 2.8: Sample linear and log ratio plots of the gluon PDFs from Table 2.3 compared to CT14 as a function of x at $Q = 2.0$ GeV.
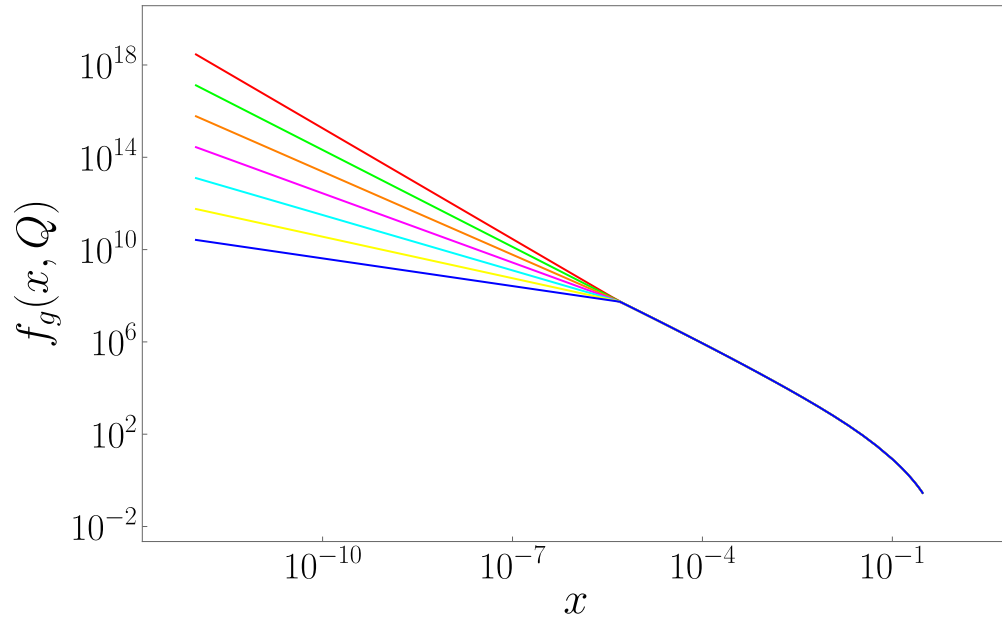
Figure 2.9: Small x extrapolation of the gluon PDF from the nCTEQ15 proton at $Q = 100$ GeV using `pdfLowFunction`. Here, $x_{min} = 5 \times 10^{-6}$, and the extrapolation exponent $1/x^a$ is set to $a = \{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$.

PDF Set		Total	$\overline{u}$	$\overline{d}$	g	d	u	s	c	b
MSTW2008nnlo68cl [25]		99.87	3.3	3.8	43.5	14.6	29.3	2.0	0.7	0
CT14nnlo [26]		100.01	3.1	3.7	43.4	14.6	29.7	2.0	0.8	0
NNPDF30_nnlo_as_0118_nf_6 [27]		99.98	3.2	3.7	43.6	14.6	29.4	2.2	0.8	0
HERAPDF20_NLO_VAR [28]		99.98	3.9	3.0	41.7	14.6	31.2	2.2	0.6	0
abm12lhc_5_nnlo [29]		100.14	2.9	3.5	43.4	14.8	30.4	2.0	0.7	0
CJ15nlo [30]		99.96	3.0	3.7	43.3	15.1	29.8	1.8	0.7	0
nCTEQ15_1_1 [7]		100.10	3.1	3.8	43.0	15.0	30.2	1.8	0.7	0
nCTEQ15_208_82 [7]		99.99	2.7	3.4	44.6	17.0	27.2	1.8	0.8	0
ct10.pds [31]		99.97	3.0	3.7	43.4	14.6	29.6	2.2	0.7	0
ctq66m.pds [14]		99.98	2.9	3.6	43.6	14.5	29.4	2.3	0.7	0

Table 2.3: We compute the momentum sum rule, Eq. 2.1, (in percent) for the individual partons at $Q = 3$ GeV. Partons $\{\bar{s}, \bar{c}, \bar{b}\}$ are not shown, but are equal to $\{s, c, b\}$. The totals sum to 100% within uncertainties of integration and interpolation. Here the colors matched with each set correspond to that set in Fig. 2.1 and Fig. 2.8

2.4.3. Momentum Sum Rules

The PDFs satisfy a several of momentum and number sum rules, and this provides a useful cross check on the results. The momentum sum rule:

$$\sum_i \int_0^1 dx\, x\, f_i(x, Q) = 1, \tag{2.1}$$

says that the total momentum fraction of the partons must sum to 100%. If any single parton flavor were not imported correctly, this cross-check would be violated; hence, this provides a powerful "sanity check" on our implementation. In Table 2.3 we display the partonic momentum fractions (in percent) and the total; for each PDF set the momentum sum rule checks within numerical accuracy.[20]

While Table 2.3 presented the momentum fraction for a single Q value (3 GeV), it is interesting to see how these values change with the energy scale. In Fig. 2.10 we show the momentum carried by each PDF flavor (in percent) as a function of Q in GeV. We can see the heavy quarks, $\{c, b, t\}$ enter as we cross the flavor mass thresholds. In the limit of large Q, the $\{\bar{u}, \bar{d}, \bar{s}\}$ PDFs approach each other asymptotically.

2.4.4. Nuclear Correction Factors

Given the PDFs, it is then trivial to build up simple calculations. In Fig. 2.11 we display the nuclear correction factors F_2^A / F_2^N for a variety of nuclei. Here, the F_2 structure functions are related to the PDFs via $F_2^A(x, Q) = x \sum_q e_q^2 f_{q/A}(x, Q)$ at leading order where F_2^N is an isoscalar, and F_2^A is the scaled structure function[21] for nuclei A. We have also superimposed the uncertainty bands; we will discuss this more in the following Section.

---

[20]Numerical uncertainties arise from the extrapolation down to $x \to 0$, the interpolation, and the integration precision.

[21]More specifically, F_2^N is the average of the proton and neutron $(p + n)/2$ and F_2^A is composed of Z protons, (A-Z) neutrons, and scaled by A to a make it "per nucleon:" $[Z\, p + (A - Z)\, n]/A$.
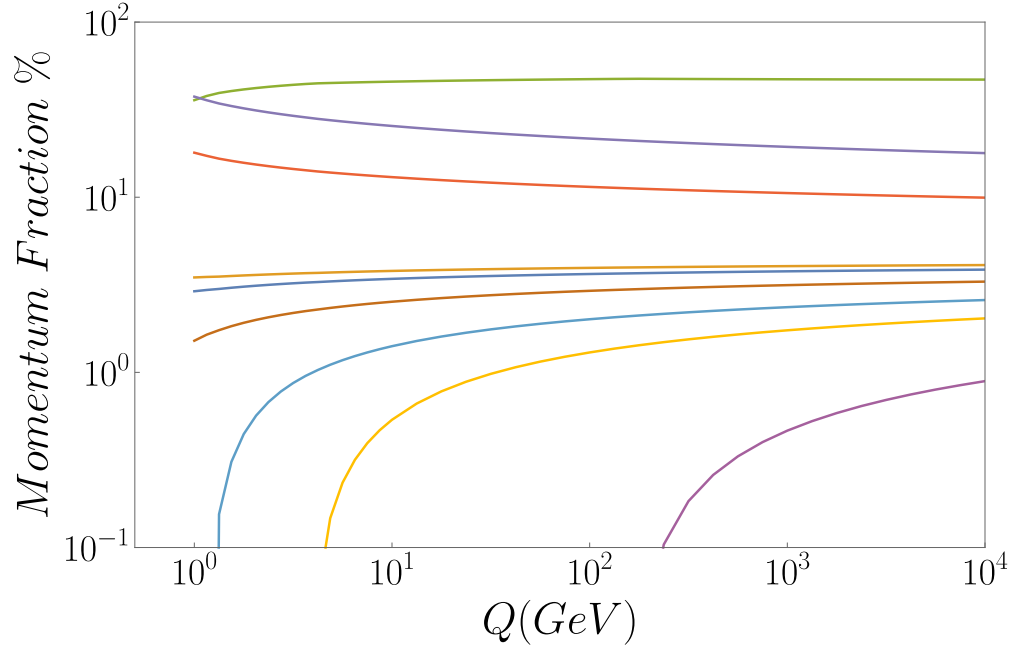
Figure 2.10: The integrated momentum fraction Eq. 2.1 of the PDF flavors vs. Q in GeV for the NNPDF set. At large Q the curves are (in descending order) $\{g, u, d, \bar{u}, \bar{d}, s, c, b, t\}$.
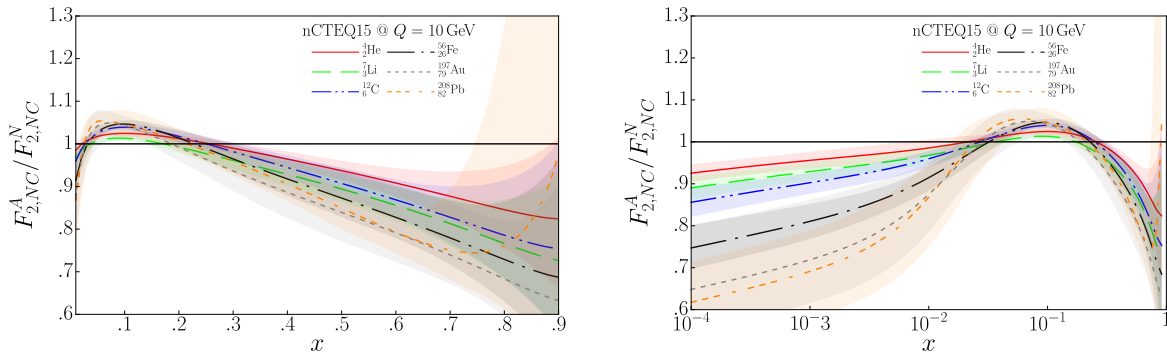
Figure 2.11: Nuclear correction ratios F_2^A/F_2^N vs. x for $Q = 10$ GeV for the nCTEQ15 PDFs over an iso-scalar target. The left plot is on a linear scale, and the right plot is a log scale. This figure is comparable to Fig. 1 of Ref. [6].

2.4.5. Luminosity

Using the integration capabilities of `Mathematica` it is easy to compute the differential parton-parton luminosity[22] for partons a and b: [32]

$$\frac{d\mathcal{L}_{ab}}{d\hat{s}} = \frac{1}{s}\frac{1}{1+\delta_{ab}}\int_{\tau}^{1}\frac{dx}{x}\, f_a(x,\sqrt{\hat{s}})\, f_b(\frac{\tau}{x},\sqrt{\hat{s}}) + (a \leftrightarrow b), \qquad (2.2)$$

where $\tau = \hat{s}/s$ ($\hat{s}$ being the energy scale of the interaction), and the cross section is

$$\sigma = \sum_{a,b}\int \left(\frac{d\hat{s}}{\hat{s}}\right)\left(\frac{d\mathcal{L}_{ab}}{d\hat{s}}\right)(\hat{s}\,\hat{\sigma}_{ab}). \qquad (2.3)$$

Note, the luminosity definition of Eq. 2.2 has dimensions of a cross section $(1/\hat{s})$, and in Eq. 2.3 we multiply by a scaled (dimensionless) cross section $(\hat{s}\,\hat{\sigma}_{ab})$.

We define the `pdfLuminosity` function to compute Eq. 2.2. The hadron-hadron production cross section for producing particle of mass $\sqrt{\hat{s}} = M_X$ is proportional to the luminosity times the scaled partonic cross section $\hat{s}\sigma$ as in Eq. 2.3. In Fig. 2.12 we display the differential luminosity $dL_{a\bar{a}}/dM_X^2$ for parton– anti-parton $(a\bar{a})$ combinations; this luminosity would be appropriate if we were interested in estimating the size of the cross section for the process of quark–anti-quark annihilation into a Higgs boson, $b\bar{b} \to H$, for example.[23]

2.4.6. W Boson Production

Next, we compute a simple leading-order (LO) cross section for W^+ boson production at the Tevatron proton–anti-proton collider (1.96 TeV) and the LHC proton-proton collider (8 TeV). Schematically, the cross section is $\sigma(W^+) = f_a \otimes f_b \otimes \omega_{ab \to W^+}$. There are two convolution integrals, but the constraint that the partonic energies sum to the boson mass W^+ eliminates one. [33, 34] Hence, this can easily be performed inside of `Mathematica`, and

---

[22]There are other definitions of the luminosity in the literature which are dimensionless such as $\mathcal{L} = f_a \otimes f_b$.

[23]`ManeParse` also has the capability to handle custom PDFs. This allows the user to explore a wide variety of phenomena, such as intrinsic heavy quarks, as long as the custom PDFs are written in either `LHAPDF6` or `CTEQ` format
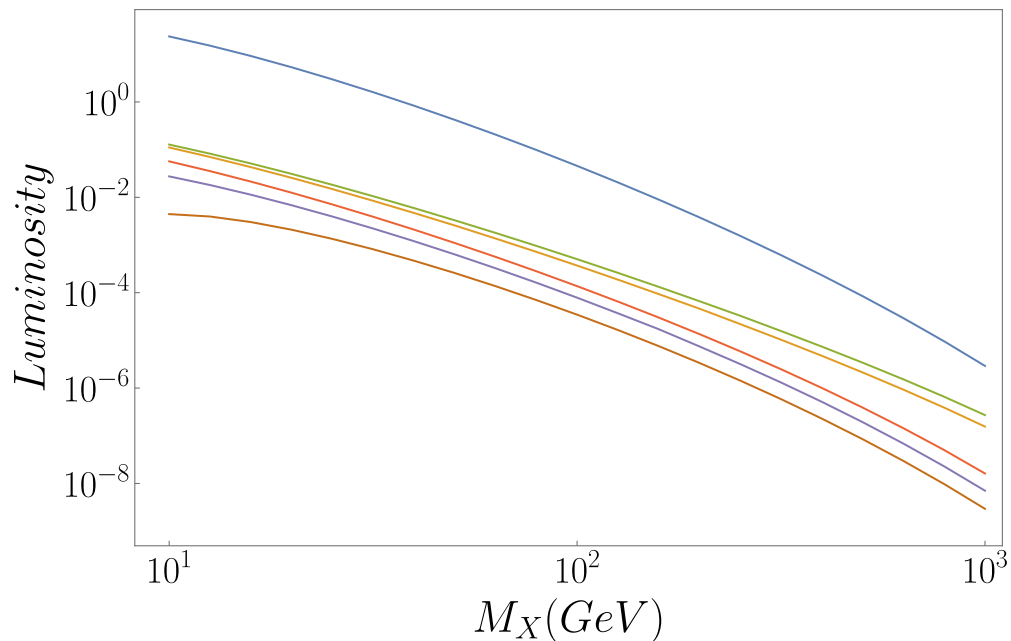
Figure 2.12: The differential parton-parton luminosity $dL_{a\bar{a}}/dM_X^2$ vs. M_x in GeV at $\sqrt{s} = 14\,\text{TeV}$ for (in descending order) $a = \{g, u, d, s, c, b\}$.

the results are displayed in Fig. 2.13. It is interesting to note the much larger width of the rapidity distribution at the LHC as well as the increased relative contribution of the heavier quark channels (such as $c\bar{s}$ and $u\bar{s}$).

2.5. Error PDFs & Correlations

2.5.1. PDF Uncertainties

We now examine some of the added features provided by the `pdfError` module. To accommodate the PDF errors, it is common for the PDF groups to release a set of grids to characterize the uncertainties; the number of PDFs in each error set is typically in the range 40 to 100, but can in principle be as many as 1000.

As `Mathematica` handles lists naturally, we can exploit this feature to manipulate the error PDFs. The `pdfFamilyParseLHA` and `pdfFamilyParseCTEQ` functions will read an entire directory of PDFs and return the associated set numbers as a list; this list can then be used to manipulate the entire group of error PDFs.
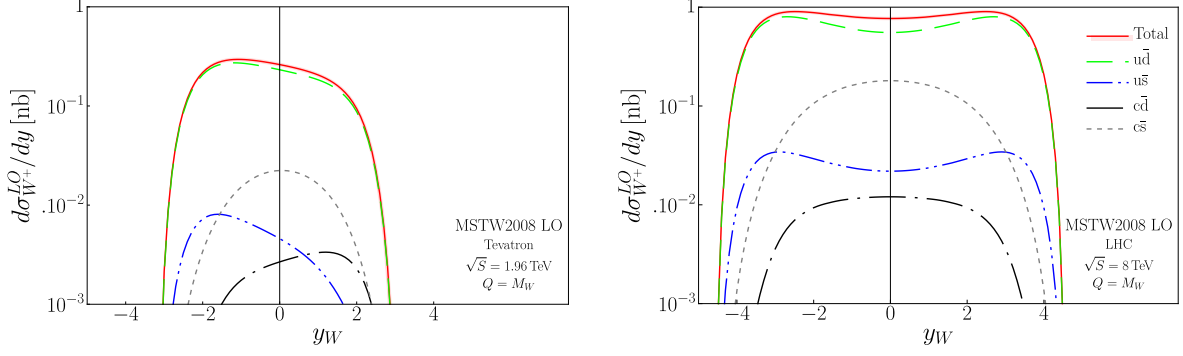
Figure 2.13: Leading-Order W^+ production cross section, $d\sigma/dy$ at the Tevatron ($p\bar{p}$, 1.96 TeV) and the LHC (pp, 8 TeV). We display the total cross section and the individual partonic contributions.
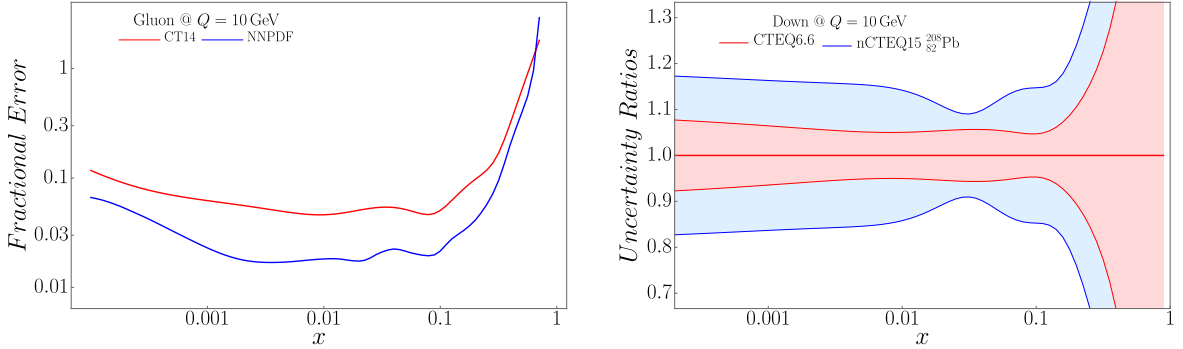
Figure 2.14: The fractional PDF uncertainty vs. x at $Q = 10$ GeV. a) The upper (red) curve is CT14 using the `pdfHessianError` function, and the lower (blue) curve is the NNPDF using the `pdfMCError` function for the gluon. (Note, these curves do not necessarily represent the same confidence level.) b) The down quark PDF uncertainty band for the CTEQ6.6 PDFs (inner, red) and the `nCTEQ15` lead 208 (outer, blue);

For example, we can use this feature to read the 100 PDFs of the NNPDF set displayed in Fig. 2.14, capture the returned list of `iSet` values, and pass this to the plotting function; we'll describe this more in the following.

When working with the error PDFs, the first step is to take the list of `iSet` values and obtain a list of the PDF values. Constructing the PDF error depends on whether the set is based on the Hessian or the Monte Carlo method.

The Hessian PDF error sets can be organized as follows $\{X_0,\ X_1^+,\ X_1^-,\ X_2^+,\ X_2^-, ..., X_N^+,\ X_N^-\}$ where X_0 represents the central set, $\{X_1^+,\ X_1^-\}$ represent the plus and minus directions

along eigenvector #1, and so on up to eigenvector N. For the Hessian PDF sets, there should be an odd number equal to $2N + 1$ where N is the number of eigenvector directions. The PDF errors can then be constructed using symmetric, plus, or minus definitions: [32, 35]

$$\Delta X_{sym}^{Hess} = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left[X_i^+ - X_i^-\right]^2} \tag{2.4}$$

$$\Delta X_{plus}^{Hess} = \sqrt{\sum_{i=1}^{N}\left[\max\left\{X_i^+ - X_0,\, X_i^- - X_0,\, 0\right\}\right]^2} \tag{2.5}$$

$$\Delta X_{minus}^{Hess} = \sqrt{\sum_{i=1}^{N}\left[\max\left\{X_0 - X_i^+,\, X_0 - X_i^-,\, 0\right\}\right]^2} \tag{2.6}$$

These can be computed using the function `pdfHessianError[iSet,(method)]`, and can take an optional "`method`" argument, {``sym'',``plus'',``minus'' }, to specify which formula is used to compute the error; the default being ``sym''.

We next turn to the Monte Carlo sets. For example, the NNPDF set (#3 in Table 2.3) has 101 elements; the "zeroth" set is the central set, and the remaining 100 replica sets span the PDF uncertainty space. The central set is the average of all the sets, and the PDF error is the standard deviation of the 100 replica sets. For these sets, `pdfMCCentral` will return the central PDF value. `pdfMCError[iSet,(method)]` will return the associated error. This function can also take an optional "`method`" argument, {``sym'', ``plus'',``minus'' }, defined by Eqs. 2.7, 2.8 and 2.9. [36, 37]

The modification from the Hessian case is due to the MC error PDFs using replica sets, not eigenvector pairs.[24] The formula for ΔX_{sym}^{MC} is a straightforward extension of the Hessian case:

---

[24]See the `LHAPDF6` reference [13] for a more complete description of the error definitions and calculation.

$$\Delta X_{sym}^{MC} \;=\; \sqrt{\frac{1}{N_{rep}} \sum_{i=1}^{N} [X_i - X_0]^2} \quad . \tag{2.7}$$

where N_{rep} counts the 100 replica sets not including the "zeroth" central set. This quantity is simply the standard deviation of the values. The $1/\sqrt{N_{rep}}$ factor compensates for the fact that Monte Carlo sets can have an arbitrary number of replicas, in contrast to the Hessian sets which have a fixed number of eigenvector sets.

It is possible to define extensions for Monte Carlo "plus" and "minus" uncertainties as: [38]

$$\Delta X_{plus}^{MC} \;=\; \sqrt{\frac{1}{N_{rep}^{+}} \sum_{i=1}^{N} [\max\{X_i - X_0,\, 0\}]^2} \tag{2.8}$$

$$\Delta X_{minus}^{MC} \;=\; \sqrt{\frac{1}{N_{rep}^{-}} \sum_{i=1}^{N} [\max\{X_0 - X_i,\, 0\}]^2} \quad , \tag{2.9}$$

where $N_{rep}^{\pm}$ are the number of replicas above/below the mean.

In Fig. 2.14-a), we compute the fractional PDF error for the CT14 PDF gluon using the `pdfHessianError` function with the ``sym'' formula of Eq. 2.4. The same is done for the NNPDF set `pdfMCError` function, using Eq. 2.7. As expected, we see the uncertainty increase both as $x \to 1$ and at very small x values.

In Fig. 2.14-b), we compute the error bands for the down quark in the CTEQ6.6 proton PDF and also the `nCTEQ15` lead-208 PDF; as expected, we see the uncertainties on the nuclear PDF are larger than the proton PDF uncertainties.

2.5.2. Correlation Angle

Finally, we can compute the correlation cosines via the relation: [14]

$$
\begin{aligned}
\cos \varphi &= \frac{\overrightarrow{\nabla} X \cdot \overrightarrow{\nabla} Y}{\triangle X \, \triangle Y} \\
&= \frac{1}{4 \triangle X \, \triangle Y} \sum_{i=1}^{N} \left(X_i^+ - X_i^- \right) \left(Y_i^+ - Y_i^- \right) .
\end{aligned} \tag{2.10}
$$

We have implemented separate functions `pdfHessianCorrelation` and `pdfMCCorrelation` as the computation of the uncertainty in the denominator $\triangle X \, \triangle Y$ could depend on Eqs. 2.4, 2.5 and 2.6 or Eqs. 2.7, 2.8 and 2.9.

In Fig. 2.15 we display an example where we show the correlation cosine between the W^+ cross section and the partonic flavors for both the Tevatron and LHC. We observe the behavior of the flavors is quite similar except for the u and d quarks which stand out at large x.

The cosine of the correlation angle indicates the degree to which the error on a particular parton's PDF contributes to the uncertainty on some function of the PDFs, usually a physical observable. A value close to one for some parton indicates that the PDF error on the observable is being driven by the error on that parton's PDF. Similarly, a value close to zero indicates that the error on the parton's PDF does not contribute significantly to the error on the observable. More details can be found in Ref. [14].

2.6. Conclusions

We have presented the `ManeParse` package which provides PDFs within the `Mathematica` framework. This is designed to work with any of the `LHAPDF6` format PDFs (through the module of my own design), and is extensible to other formats such as the CTEQ PDS format. `ManeParse` can also work with nuclear PDFs such as the nCTEQ15 sets.
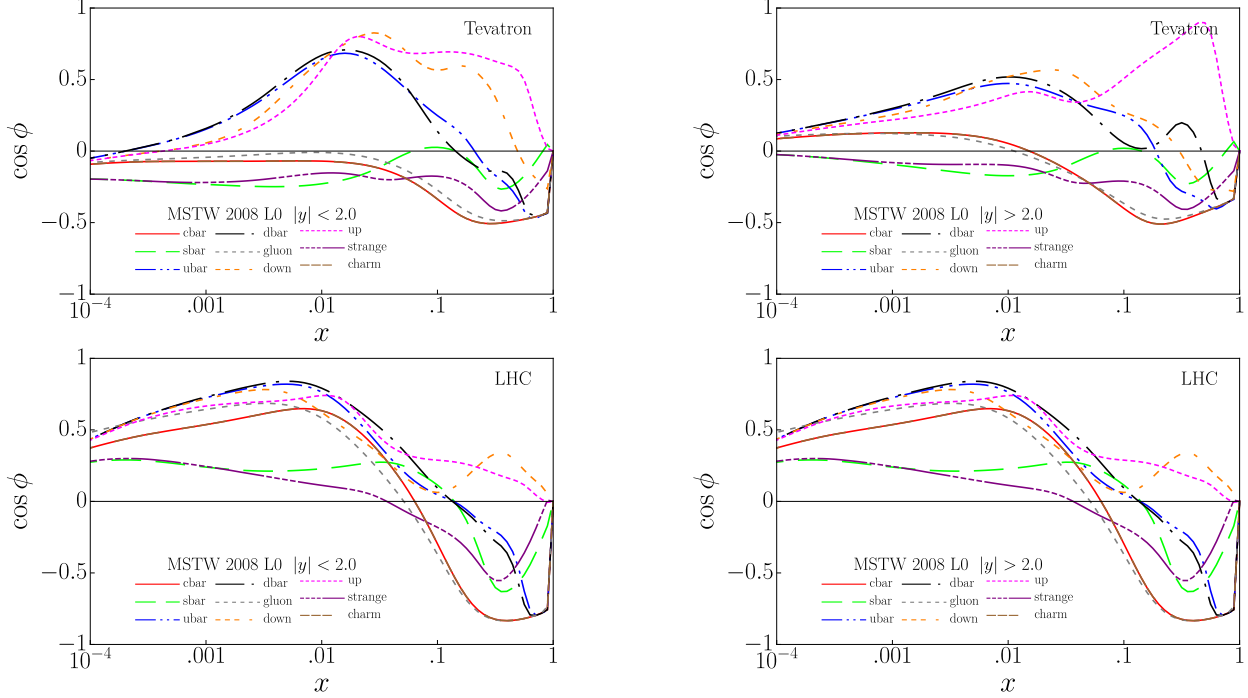
Figure 2.15: The cosine of the correlation angle, $\cos\phi$, as in Eq. 2.10, as a function of Bjorken-x for the leading-order W^+ cross section and each of the partonic flavors. Differences between Tevatron (top) $p\bar{p}$ collisions at $\sqrt{S} = 1.96$ TeV and LHC (bottom) pp collisions at $\sqrt{S} = 8$TeV are visible in the high-x region in both the central ($|y| < 2$) and high absolute rapidity ($|y| > 2$) regions.

The `ManeParse` package implements a number of novel features. It adapts YAML relations into `Mathematica` rules including unknown keys, and can handle discontinuities in both the PDFs and $\alpha_s(Q^2)$. We have implemented a flexible interpolation with a tunable parameter, and it can extrapolate to small x with a variable power. Additionally, we have implemented functions to facilitate the calculation of PDF uncertainties for both Hessian and Monte Carlo PDF sets.

`ManeParse` provides many tools to simplify calculations involving PDFs, and is fast enough such that even one or two convolutions can easily be computed within the `Mathematica` framework. We illustrated these features with examples of W production, luminosity calculations, nuclear correction factors, and N_F-dependent PDFs.

In summary, the `ManeParse` package is a versatile, flexible, user-extensible tool that can be used by beginning users to make simple PDF plots, as well as by advanced users investi-

gating subtle features of higher-order discontinuities and PDF uncertainty calculations. My work in particular expands the usability of the package with a fully-functional demo and detailedd user information.

Chapter 3

PDF Reweighting

In this chapter, I present a preliminary study to understand the effects of adding LHC
W/Z data into an `nCTEQ` nPDF fit. In order to do so without fully refitting the `nCTEQ15`
PDFs, a technique known as PDF reweighting was implemented. PDF reweighting allows
you to study the effects new data might have on the PDF but does not allow you to alter
the underlying PDF itself. This limitation constrains the effectiveness of using reweighting as an analysis tool and procludes the resulting reweighted PDF from being a true fit
to data. However, the results from this work provided a framework on which the refitting
analysis performed in Chapter 5 was built as well as providing vital insight into the nature
of the relationship between PDFs and data. Chapter 5 follows the results from this chapter
and contains a comparison between the results of reweighting study presented here and the
refitting analysis presented there. I performed this analysis as a member of the `nCTEQ` collaboration and this chapter is written in the collaboration's voice. Individually I contributed
several of the necessary cross checks and validations throughout the reweighting process as
well as some of the interpretation of the resulting PDFs.

3.1. Introduction

Vector boson production in hadron collisions is a well understood process and serves as
one of the "standard candle" measurements at the LHC. $W^{\pm}$ and Z bosons are numerously
produced in heavy ion pPb and PbPb[1] collisions at the LHC and can be used to gain insight
into the structure of nuclear parton distribution functions (nPDFs). As the $W^{\pm}$ and Z
bosons couple weakly, their interaction with the nuclear medium is negligible which makes

---

[1]Here pPb and PbPb refer to collisions betweeen a proton and a lead nucleus and two lead nuclei,
respectively.

		Observable	Cuts (GeV)	Figure
pPb	ATLAS	$d\sigma(Z \to \ell^+\ell^-)/dy_Z$ [39]	$\lvert y_Z^{CM}\rvert < 3.5;\ 60 < m_{\ell^+\ell^-} < 120$	Fig. 3.4
		$d\sigma(W^+ \to \ell^+\nu)/dy_{\ell^+}$ [40]	$p_T^{\ell^\pm} > 25;\ m_T^{\ell^\pm} > 40;\ \lvert\eta_{lab}^{\ell^\pm}\rvert < 2.4$	Fig. 3.8a
		$d\sigma(W^- \to \ell^-\bar\nu)/dy_{\ell^-}$ [40]	$p_T^{\ell^\pm} > 25;\ m_T^{\ell^\pm} > 40;\ \lvert\eta_{lab}^{\ell^\pm}\rvert < 2.4$	Fig. 3.8b
	CMS	$d\sigma(Z \to \ell^+\ell^-)/dy_Z$ [41]	$\lvert\eta_{lab}^{\ell^\pm}\rvert < 2.4;\ 60 < m_{\ell^+\ell^-} < 120;\ p_T^{\ell^+(\ell^-)} > 20$	Fig. 3.5
		$d\sigma(W^+ \to \ell^+\nu)/dy_{\ell^+}$ [42]	$p_T^{\ell^\pm} > 25;\ \lvert\eta_{lab}^\pm\rvert < 2.4$	Fig. 3.7a
		$d\sigma(W^- \to \ell^-\bar\nu)/dy_{\ell^-}$ [42]	$p_T^{\ell^\pm} > 25;\ \lvert\eta_{lab}^\pm\rvert < 2.4$	Fig. 3.7b
	ALICELHCb	$\sigma(Z \to \ell^+\ell^-)$ [43]	$60 < m_{\ell^+\ell^-} < 120;$ $p_T^{\ell^+(\ell^-)} > 20; 2.0 < \eta^{\ell^\pm} < 4.5; -4.5 < \eta_{\ell^\pm} < -2.0$	Fig. 3.6
		$\sigma(W^+ \to \ell^+\nu)$ [44]	$p_T^{\ell^\pm} > 10;\ 2.03 < \eta_{lab}^{\ell^\pm} < 3.53;\ -4.46 < \eta_{lab}^{\ell^\pm} < -2.96$	Fig. 3.9a
		$\sigma(W^- \to \ell^-\bar\nu)$ [44]	$p_T^{\ell^\pm} > 10;\ 2.03 < \eta_{lab}^{\ell^\pm} < 3.53;\ -4.46 < \eta_{lab}^{\ell^\pm} < -2.96$	Fig. 3.9b
PbPb	ATLAS	$1/\sigma_{tot}d\sigma/dy_Z$ [45]	$66 < m_{\ell^+\ell^-} < 116;\ \lvert y_Z\rvert < 2.5$	Fig. 3.10a
		A_ℓ [46]	$p_T^\ell < 25;\ \lvert\eta_{lab}^\ell\rvert < 2.5;\ m_T > 40;\ p_T^{miss} < 25$	Fig. 3.11a
	CMS	$1/\sigma_{tot}d\sigma/dy_Z$ [47]	$60 < m_{\ell^+\ell^-} < 120;\ \lvert y_Z\rvert < 2.0$	Fig. 3.10b
		A_ℓ [48]	$p_T^\ell < 25;\ \lvert\eta_{lab}^\ell\rvert < 2.1;\ m_T > 40$	Fig. 3.11b

Table 3.1: LHC data sets considered in this analysis. The observables studied include differential and total cross section ($d\sigma, \sigma$) and lepton asymmetry (A_ℓ). The cuts implemented vary by experiment but were applied to the following variables by the experiments: rapidity (y), invariant mass of the leptons (m), transverse momentum (p_T), pseudorapidity (η), and transverse mass (m_T)

these processes one of the cleanest probes of the nuclear structure available at the LHC. The possibility of using vector boson production data to constrain nPDFs was previously considered [49], and this demonstrated the strong potential for the proton-lead data (especially the asymmetries) to constrain the nuclear PDFs. The current LHC measurements for $W^\pm$ and Z production include rapidity and transverse momentum distributions for both proton-lead (pPb) and lead-lead (PbPb) collisions [39–48]. Some of these data were already used (along with jet and charged particle production data) in a recent analysis [50, 51] employing

Beam Energy [TeV]	3.5	4	6.5	7
$\sqrt{s_{pp}}$	7.00	8.00	13.00	14.00
$\sqrt{s_{PbPb}}$	2.76	3.15	5.12	5.52
$\sqrt{s_{pPb}}$	4.40	5.02	8.16	8.79

Table 3.2: The CM energy per nucleon for pp, pPb and PbPb collisions *vs.* the proton beam energy in TeV units.

a reweighting method to estimate the impact of these data on EPS09 [9] and DSSZ [52] nPDFs.[2]

The LHC heavy ion $W^{\pm}/Z$ data set is especially interesting as it can help to resolve the long-standing dilemma regarding the heavy flavor components of the proton PDFs. Historically, this has been an important issue as nuclear target data (especially ν-DIS) have been essential in identifying the individual parton flavors [26,27,54,55]; however, this means that the uncertainties of the heavy flavors are intimately tied to the (large) nuclear uncertainties. The LHC heavy ion $W^{\pm}/Z$ data has the potential to improve this situation due to the following two key features. First, this data is in a kinematic regime where the heavier quark flavors (such as strange and charm) contribute substantially. Second, by comparing the proton $W^{\pm}/Z$ data with the heavy ion results we have an ideal environment to precisely characterize the nuclear corrections. The combination of the above can not only improve the nuclear PDFs, but also the proton PDFs which are essential for any LHC study.

In this chapter I present predictions for vector boson production in pPb and PbPb collisions at the LHC obtained using `nCTEQ15` nuclear parton distributions, and perform a comprehensive comparison to the available LHC data. We also identify the measurements which have the largest potential to constrain the nPDFs, and perform a reweighting study which allows us to estimate the effects of including these data in an nPDF fit.

---

[2]A new global analysis including pPb LHC data has since been released by the EPPS collaboration [53]. However, it had not been published at the time of this analysis.

The rest of the chapter is organized as follows. Sec. 3.2 is devoted to predictions of vector boson production at the LHC in nuclear collisions. In particular, we provide an overview of the kinematic range probed by the $W^\pm/Z$ data and discuss the tools we will use for the calculation. Then we present our predictions for pPb and PbPb collisions at the LHC and compare them with the experimental data and other theoretical predictions. In Sec. 3.3 we perform a reweighting using `nCTEQ15` distributions to assess the impact of the nuclear data on the nPDFs. Finally, Sec. 3.4 summarizes our results and observations.
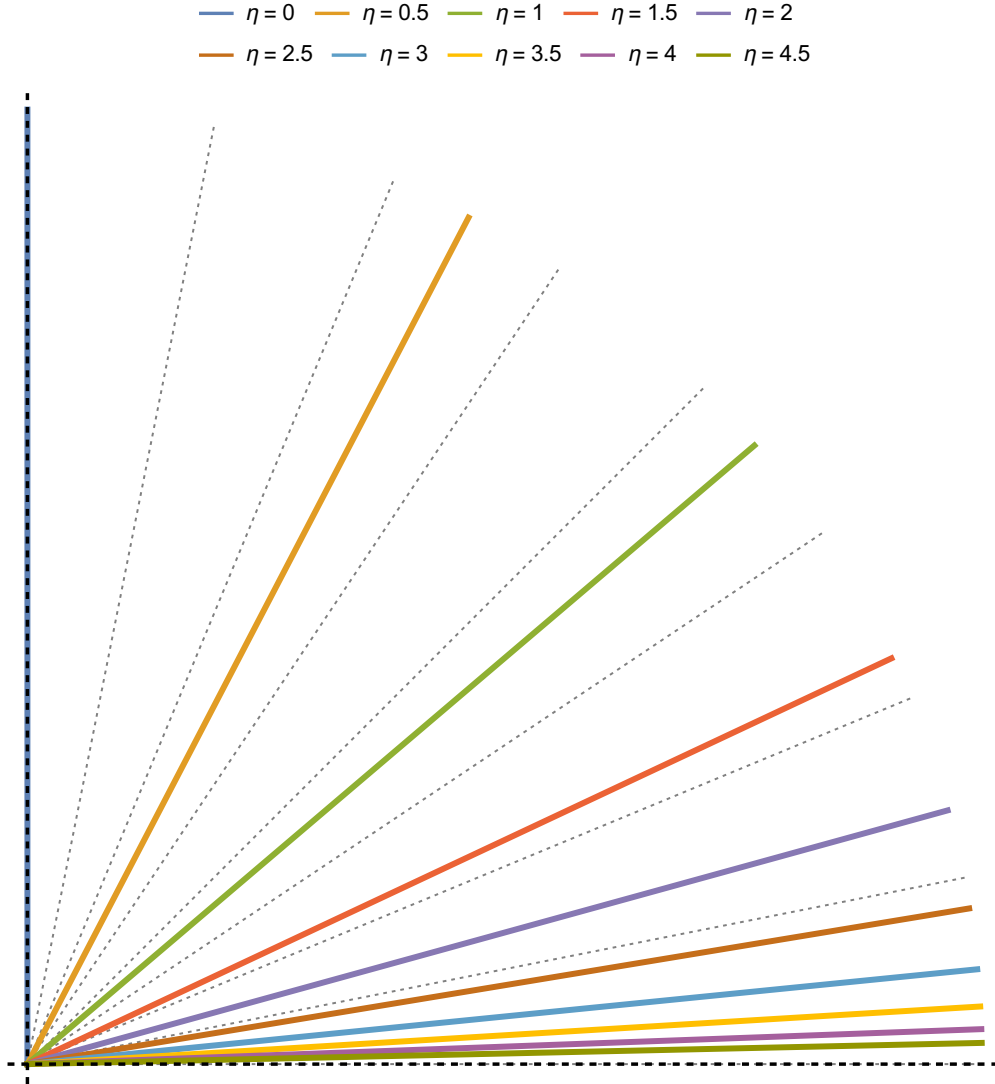
Figure 3.1: As a reminder to the reader, rapidity, y, is an measurement of the angle of a particle with respect to the beam axis in a collider experiment. It is defined by: $y = \frac{1}{2}\ln\left|\frac{E+p_z}{E-p_z}\right|$. In the limit where the rest mass of the particle is negligable in regard to its momentum, rapidity becomes equivalent to pseudorapidity. Pseudorapidity, η, can be defined by: $\eta = \frac{1}{2}\ln\left|\frac{|\vec{p}|+p_z}{|\vec{p}|-p_z}\right| = -\ln\left|\tan(\frac{\theta}{2})\right|$. Here pseudorapidities from 0 (perpendicular to the beam axis, $\theta = \frac{\pi}{2}$) to 4.5 have been plotted for reference.
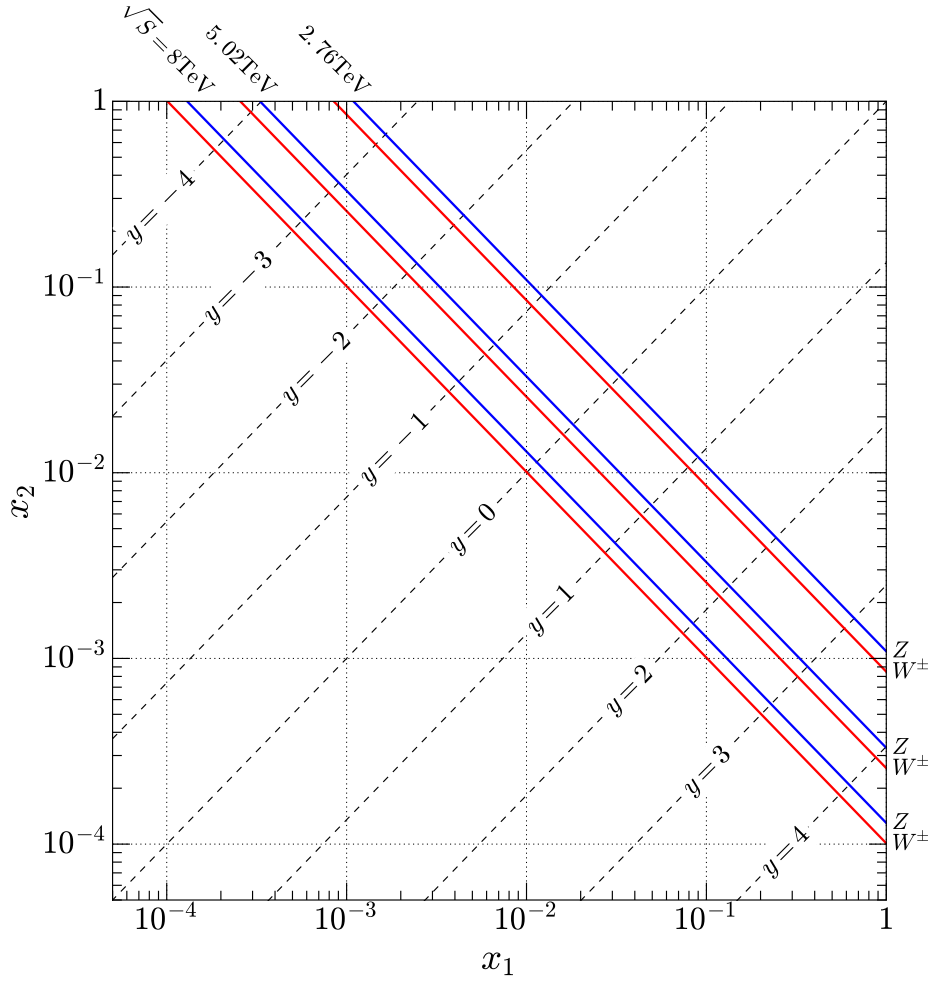
Figure 3.2: The kinematic (x_1, x_2) space explored by the measurements in this study. We display lines of constant $\tau = M_V/\sqrt{s}$ where M_V is the invariant mass of the produced $W^\pm/Z$ vector boson, as well as the center of mass (CM) rapidity y. In case of pPb collisions, we use the standard convention where x_1 corresponds to the proton and x_2 to the Pb momentum fraction.

3.2. $W^\pm/Z$ Production at the LHC

We begin by presenting our predictions for $W^\pm$ and Z boson production in nuclear collisions at the LHC using the recently published `nCTEQ15` PDFs [7].

3.2.1. Experimental data and theoretical setup

For the theoretical calculations in our study we use the FEWZ (Fully Exclusive W, Z production) [1, 2] program version 2.1. Even though FEWZ can compute W and Z production with decays up to next-to-next-to-leading order, we work at next-to-leading order (NLO) to be consistent with the order of evolution of the nPDFs.[3]

As FEWZ is designed to handle pp or $p\bar{p}$ collisions, we have extended it so that two different PDF sets can be used for the two incoming beams as required for the pPb collisions.

For the lead PDFs we use the `nCTEQ15` nPDFs [7], while we use the CT10 distributions [31] for the free protons; the only exception is the use of MSTW2008 PDFs [25] for the LHCb Z boson measurement [43] in order to match the original LHCb publication. Additionally, we compare these results with predictions calculated using nuclei made out of free proton PDFs, and in some cases free proton PDFs supplemented with EPS09 nuclear corrections [9].

We will consider LHC data on $W^\pm$ and Z boson production from the ALICE, ATLAS, CMS, and LHCb experiments. The exhaustive list of data sets that we use is provided in Table 3.1 along with the experimental kinematical cuts implemented in the analysis. While there are measurements for both the rapidity and transverse momentum distributions, for this study we will focus only on the rapidity measurements.

Using the transverse momentum (p_T) distributions to study the PDFs is more intricate as it requires resummations in the low p_T region where the cross section is maximal; we reserve this for a future study.

In Fig. 3.2 we display the kinematic space probed by the $W^\pm/Z$ production process [33]. We translate between the $\{x_1, x_2\}$ and the $\{y, \tau\}$ variables for three values of the collider center of mass (CM) energy, $\sqrt{s}$. Table 3.2 lists the CM energy per nucleon as a function of

---

[3]The CT10 proton PDFs used in the theoretical calculations are also at NLO.

the nominal proton beam energy which is determined from the relation:

$$\sqrt{s_{N_1 N_2}} = \sqrt{s_{pp}} \sqrt{\frac{Z_{N_1}}{A_{N_1}}} \sqrt{\frac{Z_{N_2}}{A_{N_2}}} \, , \tag{3.1}$$

where in case of lead we have $A = 208$ and $Z = 82$. Additionally for asymmetric collisions there is a rapidity shift, δy, between the CM and the laboratory (LAB) frame:

$$\delta y = \frac{1}{2} \ln \left[\frac{E_{N_1}}{E_{N_2}} \right] \, , \tag{3.2}$$

and in particular for the case of pPb collisions, $E_{Pb} = (Z_{Pb}/A_{Pb})E_p$ giving $\delta y_{pPb} = \frac{1}{2} \ln \left(\frac{82}{208} \right) \simeq -0.465$, i.e. $y_{CM} = y_{LAB} - 0.465$.

For the asymmetric case of pPb, we use the convention where x_1 is the proton momentum fraction, and x_2 is the lead momentum fraction. Thus, for pPb at large y_{CM} we have a large proton x_1 and a small lead x_2; conversely, at small y_{CM} we have a small proton x_1 and a large lead x_2.

In Fig. 3.2, the pair of lines with $\sqrt{s}$=2.76 TeV corresponds to PbPb collisions with a beam energy of 3.5 TeV per proton, and $\sqrt{s}$=5.02 TeV corresponds to pPb collisions with a beam energy of 4 TeV per proton.

3.2.2. Comparison to Proton-Lead (pPb) data

We first consider the LHC pPb collisions at $\sqrt{s} = 5.02$ TeV. The distributions are shown in the CM frame, and include the appropriate rapidity shift according to Eq. (3.2). In Fig. 3.3, we display the kinematic range of the pPb data bins (central values) in the plane (y, x_2) where y is the rapidity in the CM frame of the relevant vector boson or lepton, and x_2 the lead parton momentum fraction. As expected, there is little data below $x \sim 10^{-3}$ and most of the constraints from these LHC data are in the low- to mid-x region.

Figs. 3.4, 3.5 and 3.6 show our predictions for the ATLAS [39], CMS [41] and LHCb [43] Z boson production measurements, respectively. In all three cases, results obtained with the nCTEQ15 nPDFs are shown along with those obtained with a lead nucleus composed
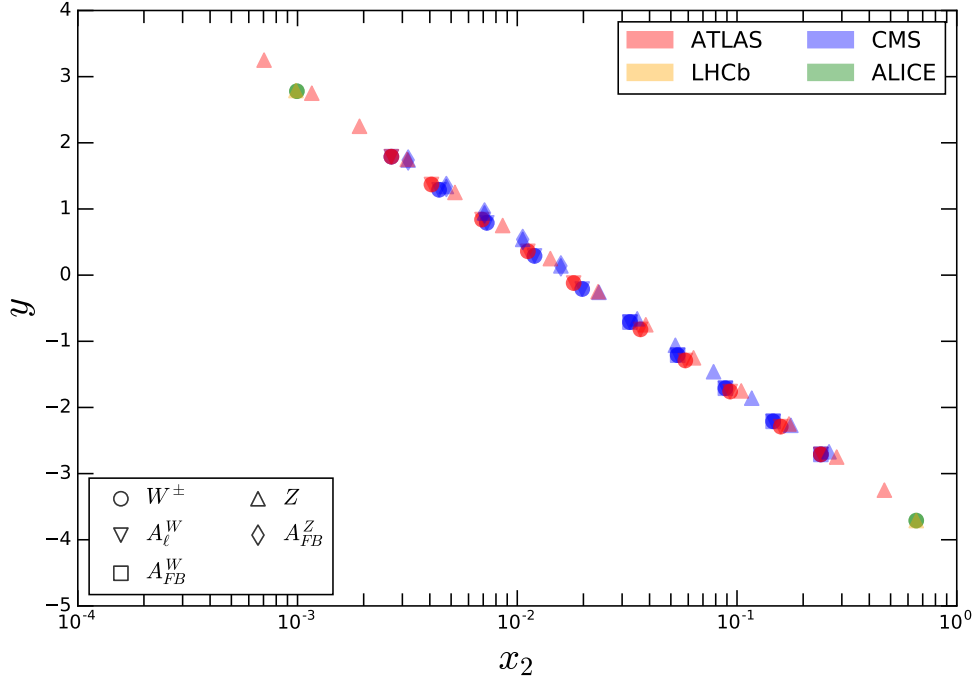
Figure 3.3: Range of the pPb data used for reweighting. y is rapidity in the CM frame and x_2 is momentum of the parton from the lead beam.

by Z protons and $A - Z$ neutrons, assuming isospin symmetry and using CT10 PDFs; the ratio of predictions over the data is shown in the lower panel. Note that the errors shown for the `nCTEQ15` predictions are for nuclear uncertainties only (and only for the beam with momentum fraction x_2) which means that the PDF error of the proton beam is not accounted for.[4] Furthermore, the errors shown for the pPb predictions using lead nuclei constructed from CT10 and MSTW2008 proton PDFs are only for the beam with momentum fraction x_2. By comparing the proton uncertainties (CT10 and MSTW2008) to the nuclear uncertainties, we see that the nuclear uncertainties are much larger.

Examining Figs. 3.4, 3.5 and 3.6, it is interesting to note the following.

1. The data and theory are generally compatible (without significant tension) both with and without nuclear corrections; this situation may change as the experimental errors and nuclear uncertainties are reduced.

---

[4]For the symmetric case of PbPb collisions the errors on both beams are taken into account.
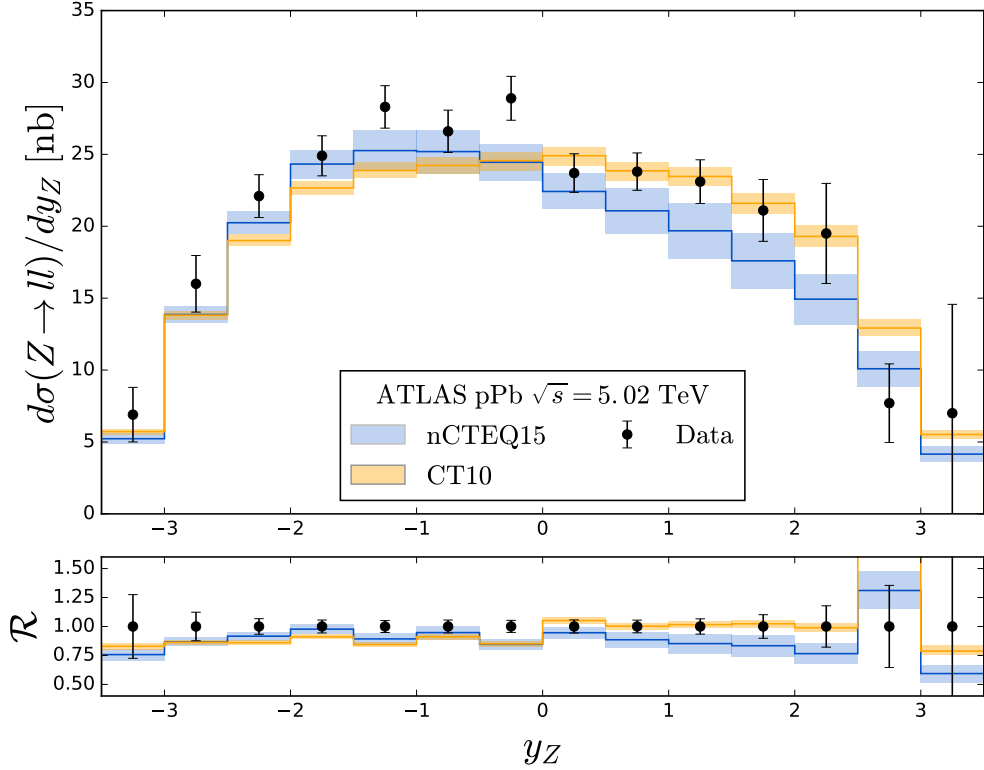
Figure 3.4: ATLAS Z production in pPb collisions.

2. Focusing on the ATLAS and CMS comparison of Figs. 3.4 and 3.5, we observe that the distributions peak at negative rapidities $y_Z \sim -1$. Referring to Fig. 3.2, this corresponds to an enhancement of the $q\bar{q}$ proton-lead luminosity over the pure proton one in the x_2 region ~ 0.05.

3. Focusing on the LHCb data of Fig. 3.6, we find good agreement for negative y, but poor agreement for the positive y data point. Despite these differences, the large uncertainties will yield a reduced impact in our subsequent reweighting procedure.

We now turn our attention to W^+ and W^- production at the LHC. In Figs. 3.7, 3.8 and 3.9 we compare the data obtained by CMS [42], ATLAS [40] and ALICE [44] for $W^{\pm}$ production with theoretical predictions obtained with `nCTEQ15` and CT10 PDFs.

We find the W^- CMS and ATLAS data are adequately described the shape in the negative rapidity range ($y_{\ell^-} < 0$), but the tensions grow as we move to larger rapidity. This effect is magnified for the case of W^+ where we see substantive deviations at large rapidity ($y_{\ell^+} > 1$).
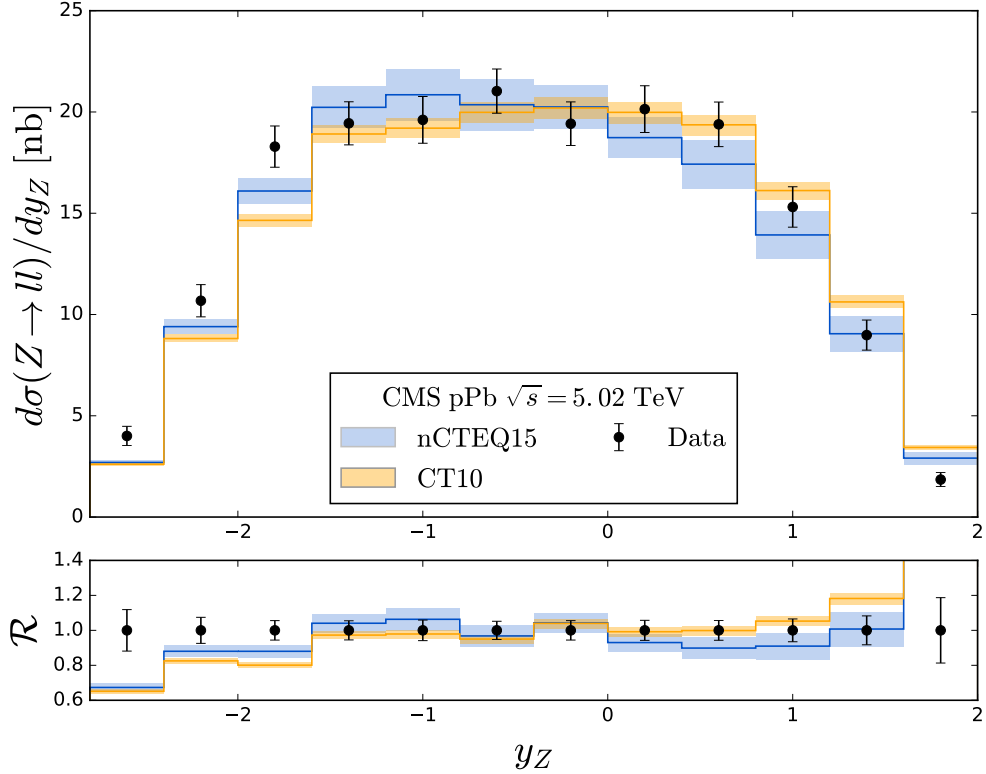
Figure 3.5: CMS Z production in pPb collisions.

Referring to Fig. 3.2, these deviations are in the smaller x_2 region ($\sim 3 \times 10^{-3}$) where we might expect nuclear shadowing of the $u\bar{d}$ and $d\bar{u}$ luminosities.[5] However, this low x_2 range is unconstrained by the data currently used in nPDF fits, so these results come from an extrapolation of the larger x_2 region. It is interesting to observe that a delayed shadowing (which shifts the shadowing down to smaller x_2 values) would improve the comparison of the data with the theory in the larger $y_{\ell^\pm}$ region; this type of behavior was observed in the nuclear corrections extracted from the neutrino-DIS charged current data [6, 56, 57]. Taking into account the errors from both the experimental data and the theoretical predictions, no definitive conclusions can be drawn at the present time. Notwithstanding, this data has the

---

[5]The nuclear correction factors are typically defined as the ratio of the nuclear quantity to the proton or isoscalar quantity. At large x ($\gtrsim 0.2$) in the EMC region the nuclear quantities are suppressed relative to the proton. In the intermediate region $x \sim 0.1$ we find "anti-shadowing" where the nuclear results are enhanced. Finally, at smaller x (a few $\times 10^{-2}$) we have the "shadowing" region where the nuclear results are suppressed.
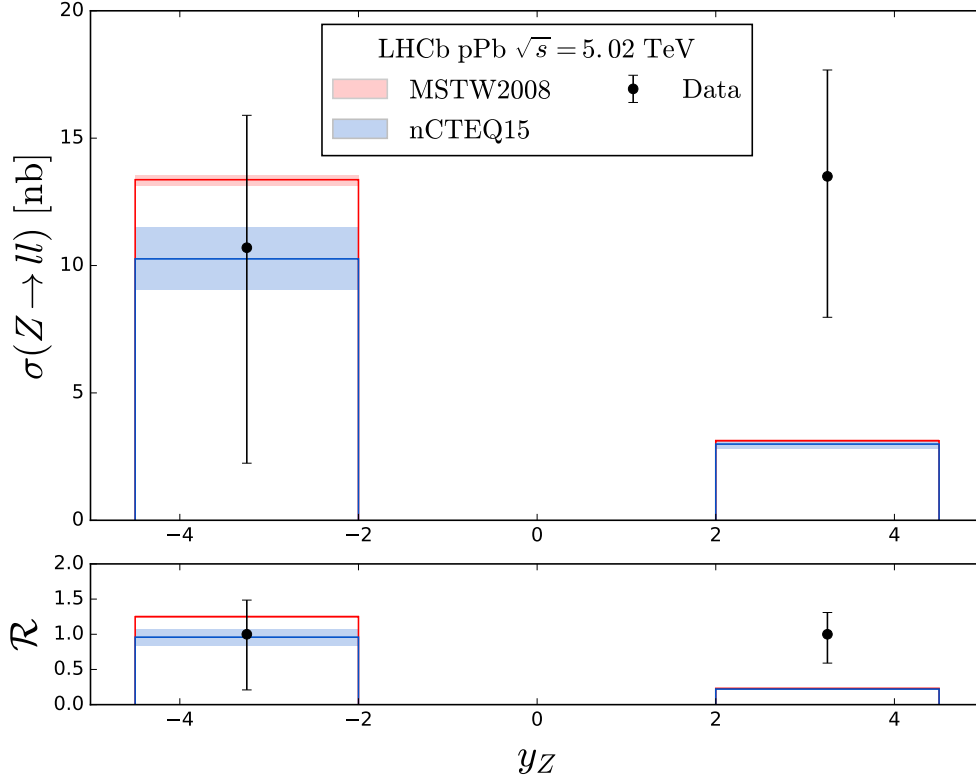
Figure 3.6: LHCb Z production in pPb collisions.

potential to influence the nPDF fits, especially in the small x_2 region, if the uncertainties could be reduced.

Finally, the ALICE data (Fig. 3.9) currently have large uncertainties, and we expect they will have a minimal impact on the reweighting.

3.2.3. Comparison to Lead-Lead data

We now consider the LHC PbPb collisions at $\sqrt{s} = 2.76$ TeV. As these beams are symmetric, we now have $y_{CM} = y_{lab}$. Again, we will use nCTEQ15 [7] and CT10 [31] PDFs for the theoretical predictions. Results from ATLAS and CMS collaborations are available in the form of either event yields (Z boson production) or charge asymmetries (A_ℓ).

We show in Fig. 3.10a and 3.10b we present the comparison of the ATLAS [45] and CMS [47] data with theoretical predictions with nCTEQ15 and CT10 PDFs. Note that the differential cross sections have been normalized to the total cross section. The PbPb data
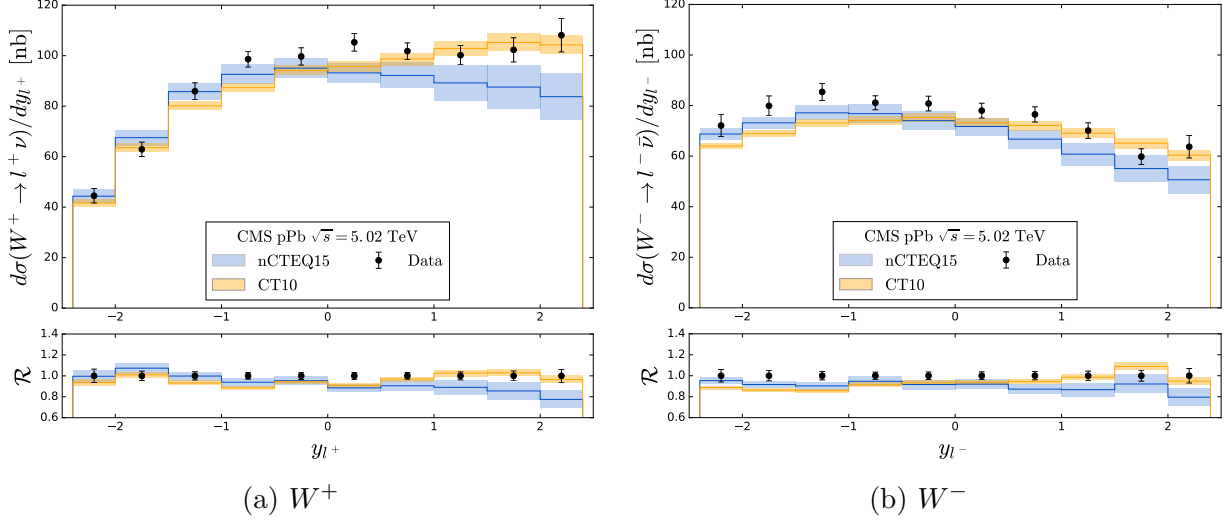
Figure 3.7: CMS $W^{\pm}$ production in pPb collisions at the LHC.

generally exhibits no tension as the distributions are well described across the kinematical range; however, this is in part due to the large uncertainties due to two nuclei in the initial state.

The measurement of charge asymmetries can provide strong constraints on the PDF fits as many of the systematic uncertainties cancel in such ratios. In Fig. 3.11 we compute the lepton ($\ell = [\mu, e]$) charge asymmetry $A_\ell(y_\ell)$:

$$A_\ell(y_\ell) = \frac{dN(W^+ \to \ell^+ \nu_\ell) - dN(W^- \to \ell^- \bar{\nu}_\ell)}{dN(W^+ \to \ell^+ \nu_\ell) + dN(W^- \to \ell^- \bar{\nu}_\ell)} \tag{3.3}$$

for W^+ and W^- bosons as measured by the ATLAS [46] and CMS [48] experiments. Unfortunately, it appears that the dependence on the nuclear corrections largely cancels out in the ratio as the nuclear nCTEQ15 result is indistinguishable from the CT10 proton result. Hence, these charge asymmetry ratios cannot constrain the nuclear corrections at the present time.
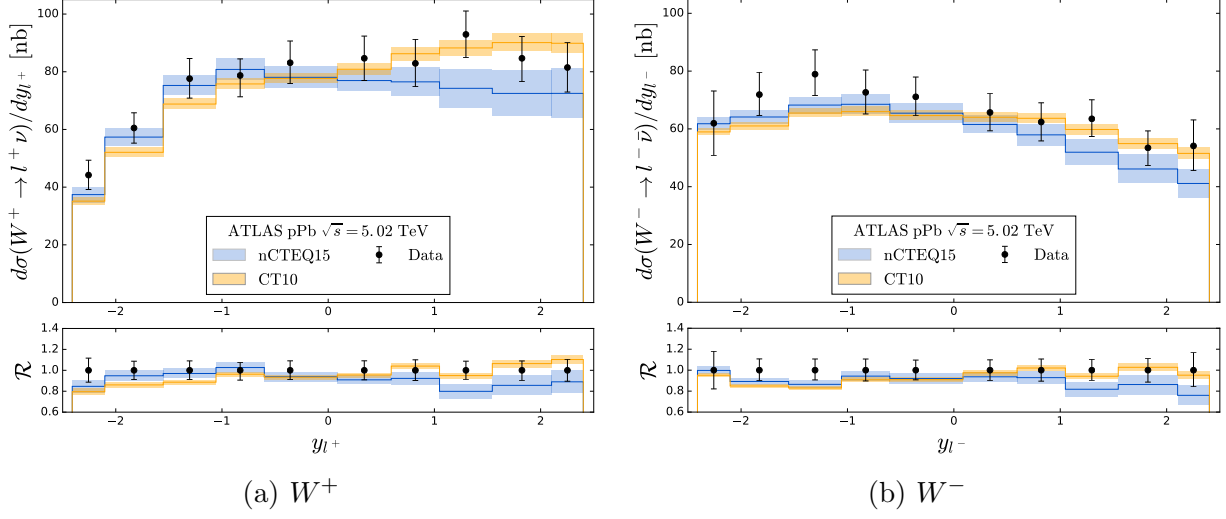
Figure 3.8: ATLAS $W^{\pm}$ production in pPb collisions at the LHC.

3.2.4. $W^{\pm}/Z$ Cross Section Correlations

In order to analyze our results more quantitatively, it is very useful to look at PDF correlations. In particular, we are interested in assessing the importance of the strange quark in our results. We first review some standard definitions before presenting our analysis.

The definition of the correlation cosine of two PDF-dependent observables X and Y is [14]

$$
\begin{aligned}
\cos\phi &= \frac{\vec{\nabla}X \cdot \vec{\nabla}Y}{\Delta X \Delta Y} \\
&= \frac{1}{4\Delta X \Delta Y} \sum_i \left(X_i^{(+)} - X_i^{(-)} \right) \left(Y_i^{(+)} - Y_i^{(-)} \right),
\end{aligned}
\tag{3.4}
$$

where ΔX is the PDF error of the corresponding observable. For the `nCTEQ15` PDFs this corresponds to the symmetric error given by

$$
\Delta X = \frac{1}{2} \sqrt{ \sum_i^N \left(X_i^{(+)} - X_i^{(-)} \right)^2 } \quad .
\tag{3.5}
$$

$X_i^{(\pm)}$ is the observable evaluated along the $\pm$ error PDF eigenvector i, and the summation runs over all eigenvector directions.
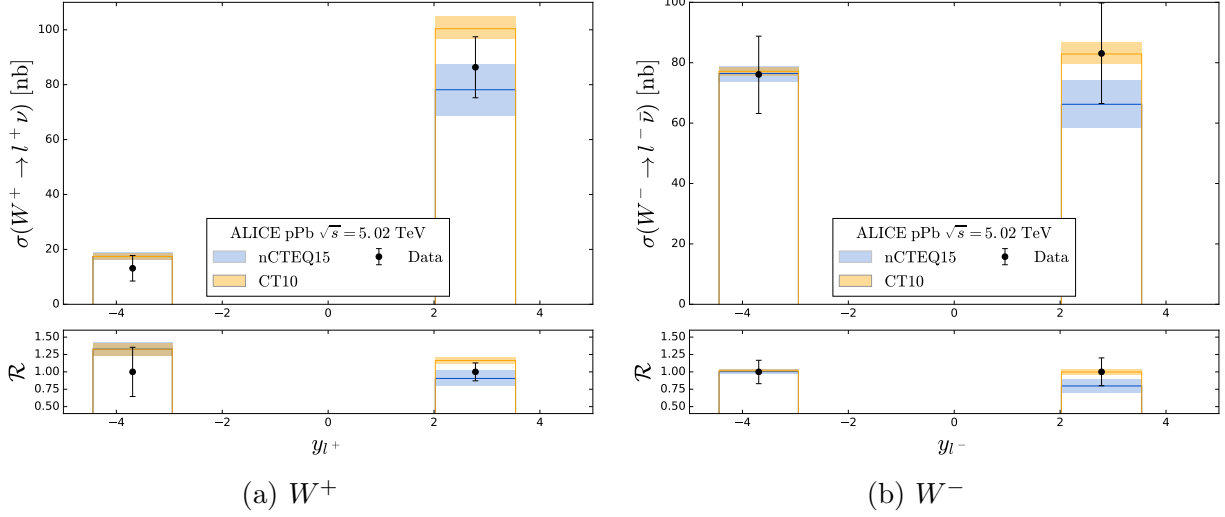
Figure 3.9: ALICE $W^{\pm}$ production in pPb collisions at the LHC.

In our case we are interested in observables $X, Y \in \{\sigma_Z, \sigma_{W^+}, \sigma_{W^-}\}$. Here, we focus on the planes formed by the (W^+, W^-) and the $(Z, W^{\pm})$ boson production cross sections to visualize the correlations.

Fig. 3.12 shows the correlations of the W^+ and W^- production cross sections for pPb collisions at the LHC in comparison with the CMS and ATLAS measurements. Similarly, in Fig. 3.13 we display the results for Z and $W^{\pm}$ bosons. The results are shown for three different rapidity regions, $y < -1$, $|y| < 1$, $y > 1$, and for several PDFs sets. For the proton side we always use the CT10 PDFs and for the lead side we examine three results: i) `nCTEQ15`, ii) CT10, and iii) CT10 PDFs supplemented by the nuclear corrections from EPS09 (CT10+EPS09). Finally, the central predictions are supplemented with uncertainty ellipses illustrating correlations between the cross sections. The ellipses are calculated in the following way [14],

$$
\begin{aligned}
X &= X_0 + \Delta X \cos\theta, \\
Y &= Y_0 + \Delta Y \cos(\theta + \phi),
\end{aligned}
\tag{3.6}
$$

64
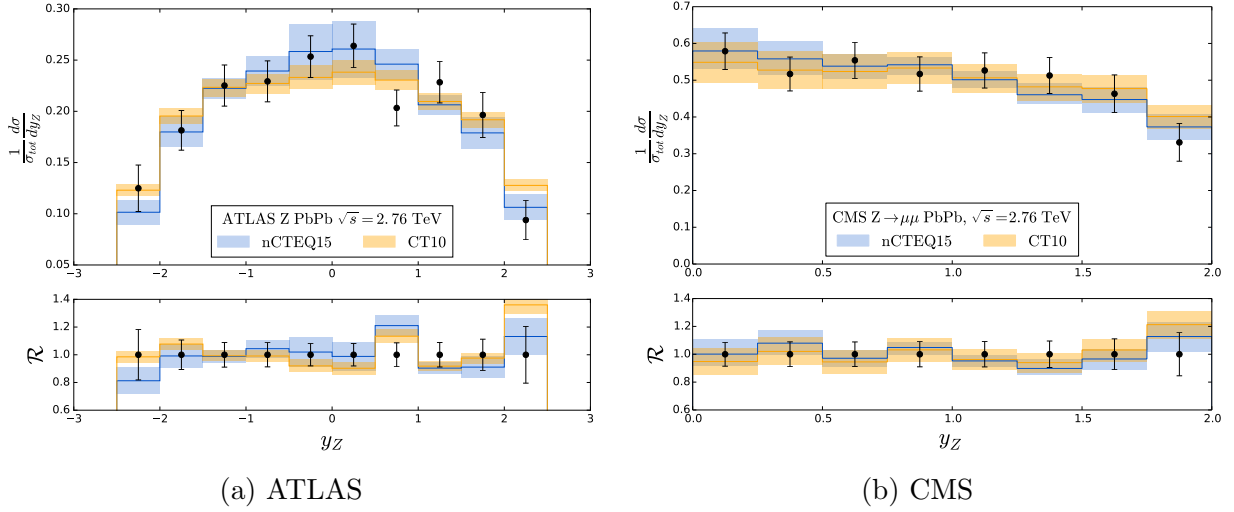
Figure 3.10: Z boson production cross section normalized to the total cross section for PbPb collisions at the LHC with $\sqrt{s} = 2.76$ TeV as measured by the ATLAS and CMS collaborations. Corresponding predictions obtained with `nCTEQ15` and CT10 PDFs are also shown.

where X, Y represent PDF-dependent observables, X_0 (Y_0) is the observable calculated with the central PDF, ΔX (ΔY) is defined in Eq. (3.5), ϕ is the correlation angle defined in Eq. (3.4), and θ is a parameter ranging between 0 and 2π.

From Figs. 3.12 and 3.13 one can generally observe that the ellipses for the different PDF sets overlap. Furthermore, the central predictions for all three PDF sets lie in the overlapping area of the three ellipses. However, a trend can be observed as a function of the rapidity:

1. For negative rapidities $(y < -1)$, the central predictions from the nuclear PDFs (nCTEQ15, EPS09) are closer to the experimental data as they yield larger cross sections than the uncorrected (proton) CT10 PDFs. This can be understood because the lead x_2 values probed in this rapidity bin lie in the region $x_2 \sim 10^{-1}$ where the nPDFs are enhanced due to anti-shadowing (cf., Fig. 9 in Ref. [7]). Due to the larger uncertainties associated with the `nCTEQ15` predictions, the ATLAS and CMS cross sections lie within the 1σ ellipse. Conversely, the measured data lie outside the uncorrected (proton) CT10 error ellipsis.

65
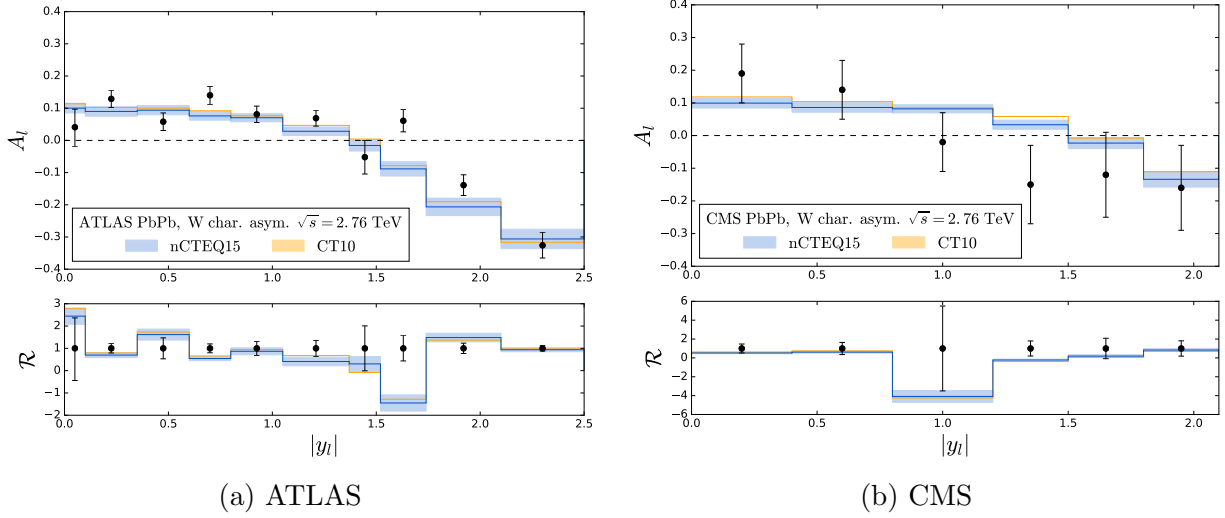
Figure 3.11: W charge asymmetry for PbPb collisions at the LHC with $\sqrt{s} = 2.76$ TeV as measured by the ATLAS and CMS collaborations. Corresponding predictions obtained with `nCTEQ15` and CT10 PDFs are also shown.

2. For the central rapidity bin ($|y| < 1$), the predictions from all three PDF sets lie generally very close together. In this case, the probed x_2 values lie in the range $0.007 \leq x_2 \leq 0.05$ which is in the transition zone from the anti-shadowing to the shadowing region. We find the LHC W^+ and W^- cross sections in Fig. 3.12 tend to lie above the theory predictions. Examining the Z cross section of Fig. 3.13, we find the CMS data agrees closely with the theory predictions, while the ATLAS data is larger by approximately 1σ.

3. For the positive rapidity bin ($y > 1$), we find the central predictions from CT10 match the $W^\pm$ data very closely, but slightly overshoot the Z data. The nuclear PDFs (nCTEQ15, EPS09) undershoot the $W^\pm$ data by a bit more than 1σ, but agree with the Z cross section within 1σ. Here, the probed x_2 values are $\lesssim 0.007$; in this region the lead PDFs are poorly constrained and the corresponding cross sections are dependent on extrapolations of the PDF parameterization in this region.

Interpreting the above set of results appears complicated, so we will try and break the problem down in to smaller components. We now compute the same results as above, but

66
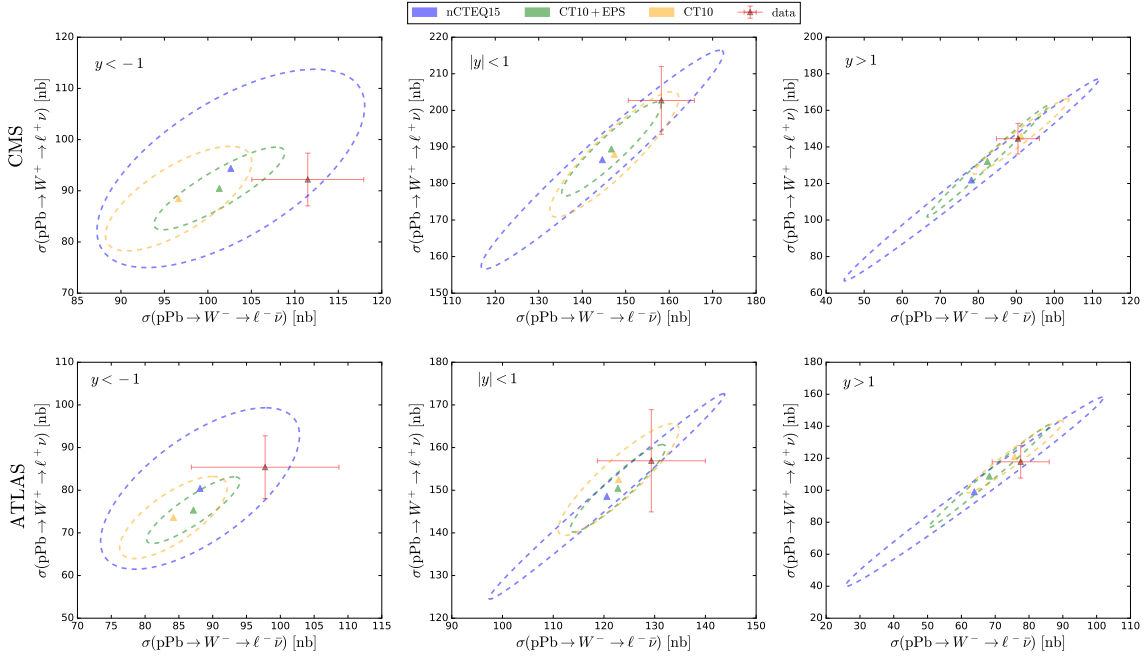
Figure 3.12: Correlations between W^+ and W^- cross sections calculated with different PDFs overlaid with the corresponding LHC data from CMS and ATLAS.

using only 2 flavors (one family) of quarks: $\{u, d\}$; specifically, these plots are produced by zeroing the heavy flavor components (s, c, b), but keeping (u, d) and the gluon. For the Z production this eliminates the $s\bar{s}$ and (the smaller) $c\bar{c}$ contributions, while for W^+/W^- production it is the $\bar{s}c/s\bar{c}$ contribution which drives the change. While the charm PDF does play a role in the above (the bottom contribution is minimal), $c(x)$ is generated radiatively by the process $g \to c\bar{c}$ (we assume no intrinsic component); thus, it is essentially determined by the charm mass value and the gluon PDF. In contrast, the "intrinsic" nature of the strange PDF leads to its comparably large uncertainties. For example, if we compare the free-proton PDF baselines (CTEQ6.1, CT10), the strange quark exhibits substantial differences while the charm (and bottom) distributions are quite similar; this pattern then feeds into the nPDFs. Therefore, the strange quark PDF will be the primary focus of the following discussion.
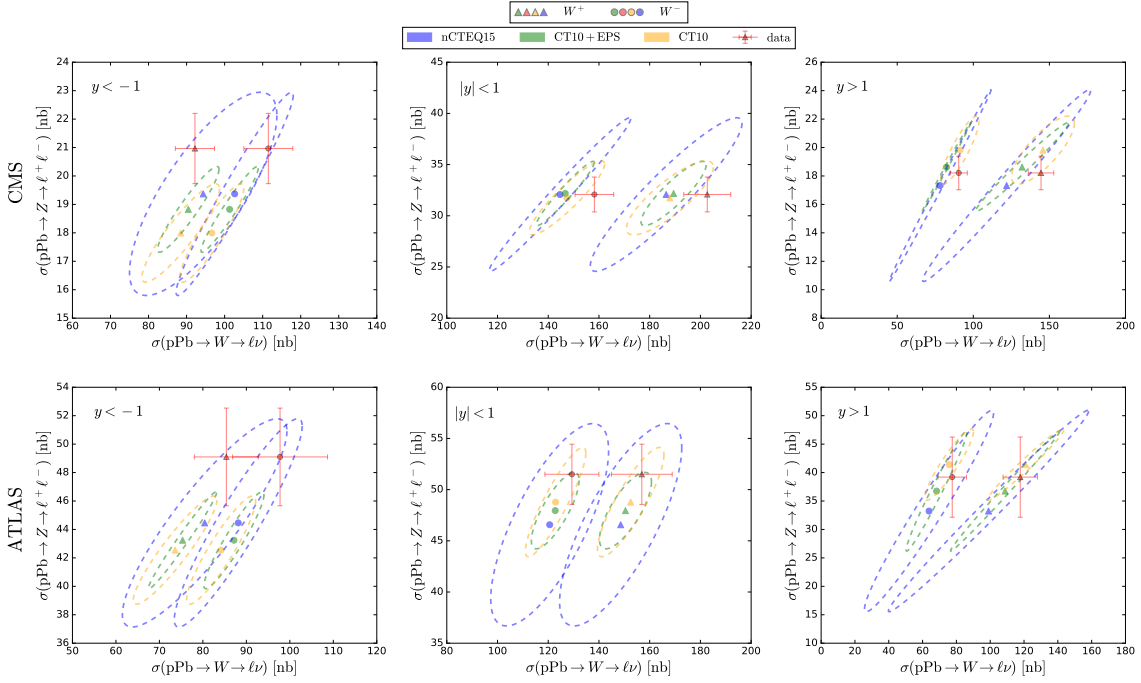
Figure 3.13: Correlations between Z and W^+/W^- cross sections calculated with different PDFs overlaid with the corresponding LHC data from CMS and ATLAS.

In Figs. 3.14 and 3.15 we compare the 5 flavor and 2 flavor results using nCTEQ15, CT10+EPS09, CTEQ6.1+ EPS09, and CT10. We have added CTEQ6.1+EPS09 as CTEQ6.1 was the baseline used for the EPS09 fit.

Examining Fig. 3.14, the shift of the 2 flavor results compared to the 5 flavor results can be as large as $\sim$30% and reflects the contributions of the strange and charm quarks.

For the 5 flavor case ($\blacktriangle$), the calculations are scattered to the low side of the data in both W^+ and W^-. The CT10 result is the closest to the data, but due to the larger uncertainties of nCTEQ15, the data point is within range of both of their ellipses. We also observe that the CT10+EPS09 and CTEQ6.1+EPS09 results bracket the nCTEQ15 value; again, this is due to the very different strange quark PDF associated with CT10 and CTEQ6.1.

For the 2 flavor case ($\bullet$), all the nuclear results (nCTEQ15, CT10+EPS09, CTEQ6.1+EPS09) coalesce, and they are distinct from the non-nuclear result (CT10). This pattern suggests that the nuclear corrections of nCTEQ15 and EPS09 for the $\{u, d\}$ flavors are quite similar,

and the spread observed in the 5 flavor case comes from differences of $s(x)$ in the underlying base PDF. Thus we infer that the difference between the nuclear results and the proton result accurately represents the nuclear corrections for the 2 flavor case (for $\{u, d\}$), but for the 5 flavor case it is a mix of nuclear corrections and variations of the underlying sea quarks.

Fig. 3.15 displays the same information as Fig. 3.14 except it is divided into rapidity bins. As we move from negative y to positive y we move from high x where the nPDFs are well constrained to small x where the nPDFs have large uncertainties (*cf.*, Fig. 3.3). Thus, it is encouraging that at $y < -1$ we uniformly find the nuclear predictions yield larger cross sections than the proton results (without nuclear corrections) and thus lie closer to the LHC data.

Conversely, for $y > 1$ we find the nuclear predictions yield smaller cross sections than the proton results. The comparison with the LHC data varies across the figures, but this situation suggests a number of possibilities.

First, the large nPDF uncertainties in this small x_2 region could be improved using the LHC data.

Second, the lower nPDF cross sections are partly due to the nuclear shadowing in the small x region; if, for example, this shadowing region were shifted to even lower x values, this would increase the nuclear results. Such a shift was observed in Refs. [6, 56, 57] using charged current neutrino-DIS data, and this would move the nuclear predictions of Fig. 3.12 at $y > 1$ toward the LHC data.

Finally, we note that measurements of the strange quark asymmetry [58] indicate that $s(x) \neq \bar{s}(s)$ which is unlike what is used in the current nPDFs; this would influence the $W^\pm/Z$ cross sections separately as (at leading-order) [33] $W^+ \sim \bar{s}c$, $W^- \sim s\bar{c}$, and $Z \sim \bar{s}s$. As the strange PDF has a large impact on the $W^\pm/Z$ measurements, this observation could provide incisive information on the individual s and $\bar{s}$ distributions.

These points are further exemplified in Fig. 3.16 which displays $W^\pm$ production for both 2 and 5 flavors as a function of lepton rapidity $y_{\ell\pm}$. For large $y_{\ell\pm}$, (small lead x_2) the CT10 proton result separates from the collective nuclear results; presumably, this is due to the

nuclear shadowing at small x_2. Again, we note that in this small x_2 region there are minimal experimental constraints and the nPDFs come largely from extrapolation at higher x_2 values. Additionally, by comparing the 2 and 5 flavor results, we clearly see the impact of the heavier flavors, predominantly the strange quark PDF.

Furthermore, different strange quark PDFs in the baseline PDFs compared in Figs. 3.12 and 3.13, make it challenging to distinguish nuclear effects from different strange quark distributions. Thus, we find that the extraction of the nuclear corrections is intimately intertwined with the extraction of the proton strange quark PDF, and we must be careful to separately distinguish each of these effects. Fortunately, the above observations can help us to accomplish this.
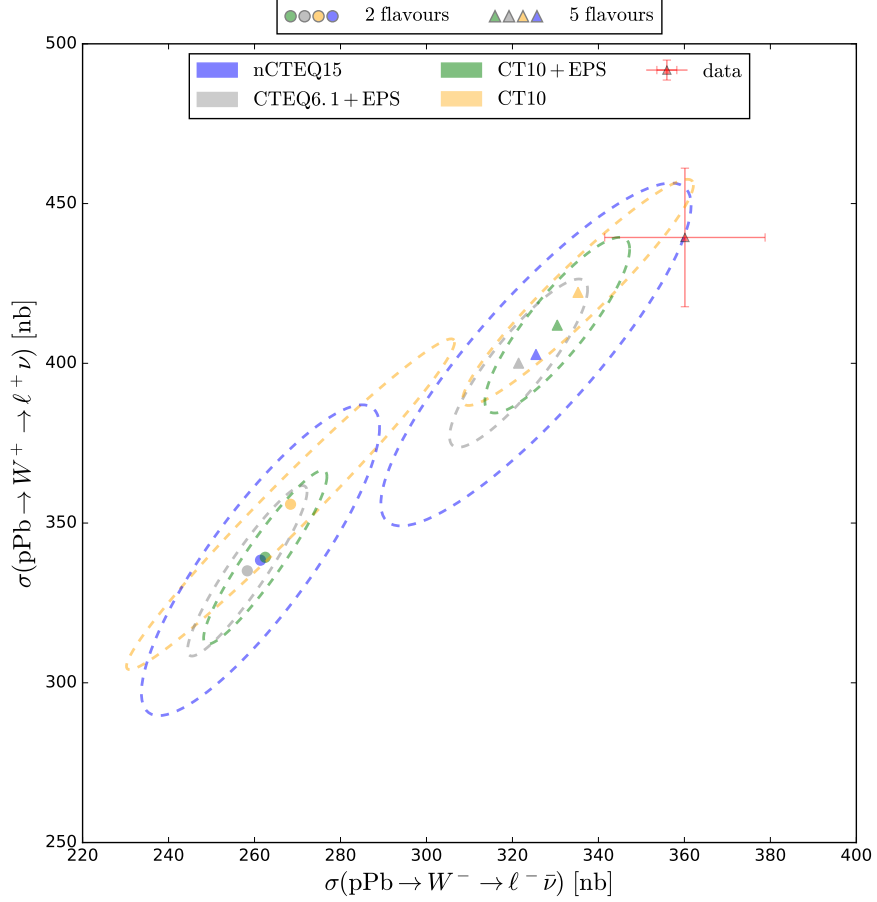
Figure 3.14: Comparison of correlations between W^+ and W^- cross sections for the case when only one family of quarks $\{u, d\}$ is included and when all families are accounted for. We show here results for `nCTEQ15`, EPS09+CT10, EPS09+CTEQ6.1 and CT10 PDFs overlaid with the CMS data.
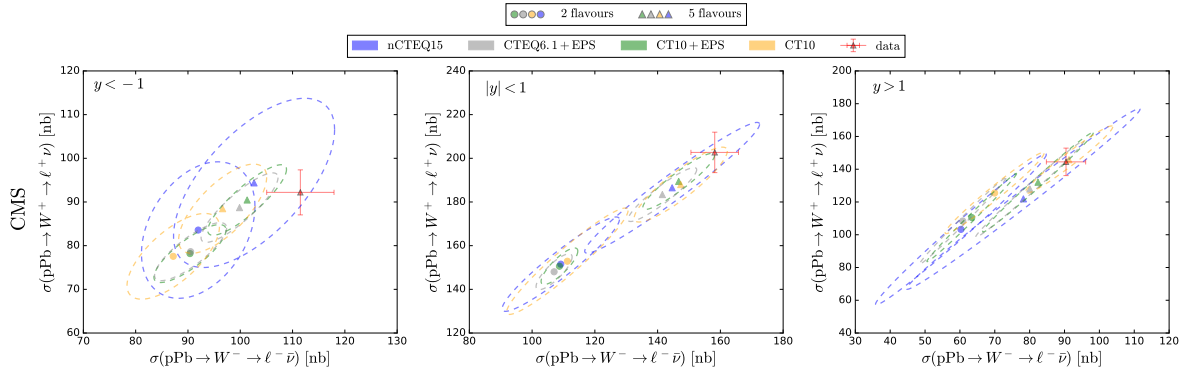
Figure 3.15: Same as Fig. 3.14 but divided into rapidity bins.

71

(a) W^+ 5 flavours

(b) W^+ 2 flavours

(c) W^- 5 flavours

(d) W^- 2 flavours
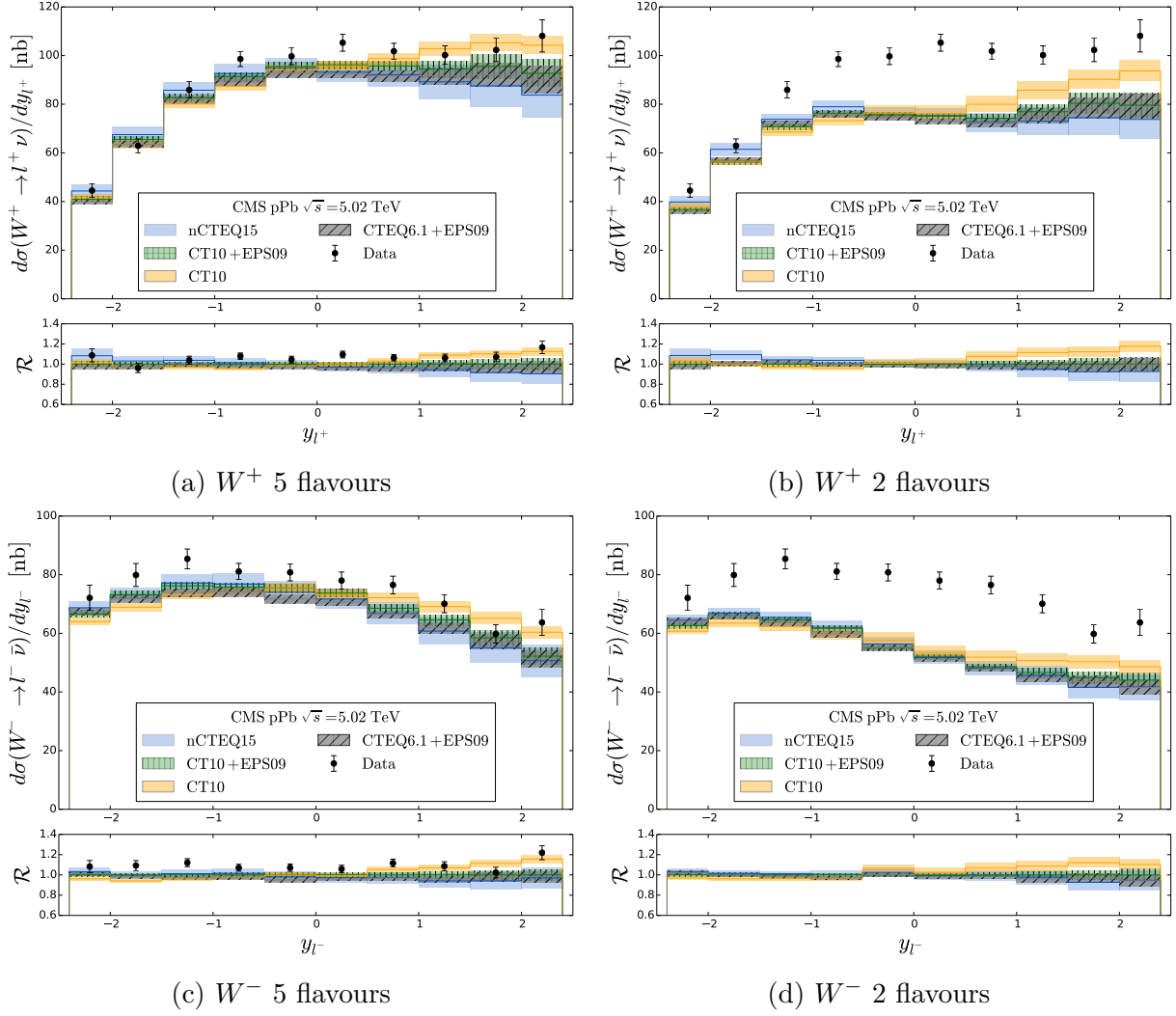
Figure 3.16: Rapidity distributions for $W^\pm$ cross sections measured by CMS compared with predictions from the `nCTEQ15`, EPS09+CT10, EPS09+CTEQ6.1 and CT10 PDFs. Figures (a) and (c) show the results for 5 flavors, while Figures (b) and (d) show the 2 flavors results.

3.3. Reweighting
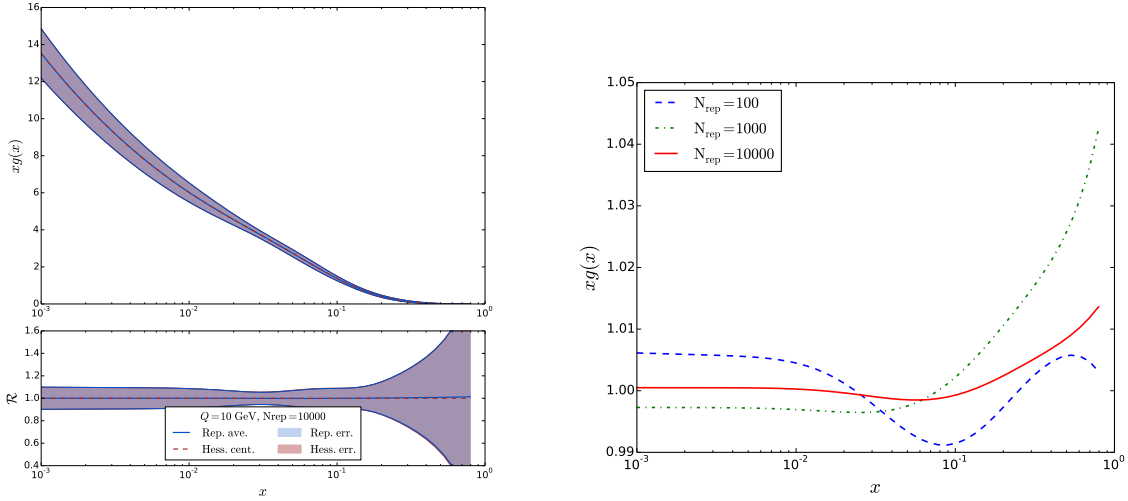
(a) Comparison of the `nCTEQ15` central gluon PDF and Hessian error bands with the average and variance calculated from $N_{\rm rep} = 10^4$ replicas.

(b) Ratio of the average gluon PDF calculated using $N_{\rm rep} = \{10^2, 10^3, 10^4\}$ replicas to the `nCTEQ15` central gluon PDF.

Figure 3.17: Comparison of the `nCTEQ15` Hessian gluon distribution and its reproduction in terms of replicas at a scale of $Q = 10$ GeV.

In this section we perform a reweighting study to estimate the possible impact of the $W^\pm/Z$ data on `nCTEQ15` lead PDFs. For this purpose we will use only the pPb data sets.

We refrain from using PbPb data as typically the agreement of these data with current nPDFs is much better (in part due to the large uncertainties), so the impact in the reweighting analysis will be minimal. Secondly the factorization in lead-lead collisions is not firmly established theoretically [59] such that the interpretation may be complicated.

3.3.1. Basics of PDF reweighting

In this section we summarize the PDF reweighting technique and provide formulas for our specific implementation of this method. Additional details can be found in the literature [60–64].

In preparation for the reweighting, we need to convert the `nCTEQ15` set of Hessian error PDFs into a set of PDF replicas [50, 65] which serve as a representation of the underlying probability distribution. The PDF replicas can be defined by a simple formula,[6]

$$
f_k = f_0 + \sum_{i=1}^{N} \frac{f_i^{(+)} - f_i^{(-)}}{2} R_{ki},
\tag{3.7}
$$

where f_0 represents the best fit (central) PDF, $f_i^{(+)}$ and $f_i^{(-)}$ are the plus and minus error PDFs corresponding to the eigenvector direction i, and N is the number of eigenvectors defining the Hessian error PDFs. Finally, R_{ki} is a random number from a Gaussian distribution centered at 0 with standard deviation of 1, which is different for each replica (k) and each eigen-direction (i).

After producing the replicas, we can calculate the average and variance of any PDF-dependent observable as moments of the probability distribution:

$$
\begin{aligned}
\langle \mathcal{O} \rangle &= \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \mathcal{O}(f_k), \\
\delta \langle \mathcal{O} \rangle &= \sqrt{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \left(\mathcal{O}(f_k) - \langle \mathcal{O} \rangle \right)^2} \quad .
\end{aligned}
\tag{3.8}
$$

In particular, it can be done for the PDFs themselves; we should be able to reproduce our central PDF f_0 by the average $\langle f \rangle$, and the (68% c.l.) Hessian error bands $\Delta f = \frac{1}{2}\sqrt{\sum_i^N (f_i^{(+)} - f_i^{(-)})^2}$ by the corresponding variance $\delta \langle f \rangle$. Of course, the precision at which we are able to reproduce Hessian central PDFs and corresponding uncertainties depends on how well we reproduce the underlying probability distribution, and this will depend on the number of replicas, N_{rep}, we use. In the following we use $N_{\text{rep}} = 10^4$ which allows for a very good reproduction of both central and error PDFs (within $\sim$ 0.1% or better).

---

[6]A detailed discussion on the construction of replicas from Hessian PDF sets in the case of asymmetric errors can be found in ref. [66].

We note here that since the `nCTEQ15` error PDFs correspond to the 90% confidence level (c.l.) we need to convert the obtained uncertainties such that they correspond to the 68% c.l.[7] The conversion is done using the following approximate relation between the 68% c.l. and 90% c.l. Hessian uncertainties: $\Delta_{90}^{H}\mathcal{O} \approx 1.645\,\Delta_{68}^{H}\mathcal{O}$.

In Fig. 3.17 we perform the above exercise and determine if our procedure is self consistent. Specifically, in Fig. 3.17a we display the central value and uncertainty bands for the original gluon PDF and those generated from the replicas; they are indistinguishable. Additionally, in Fig. 3.17b we demonstrate the convergence of the average of replicas to the central Hessian PDF for $N_{rep} = \{10^2, 10^3, 10^4\}$. For $N_{rep} = 10^4$ the central gluon is reproduced to better than 1% except at the highest x values. This is certainly a sufficient accuracy considering the size of the PDF errors. Even the $N_{rep} = 10^2$ and $N_{rep} = 10^3$ replicas yield good results except at larger x ($\gtrsim 0.1$) where the PDFs are vanishing and the uncertainties are large. Since our computational cost will be mostly dictated by the number of Hessian error PDFs, we will use $N_{rep} = 10^4$ to get a better representation of the underlying probability distribution.

Having defined the replicas we can apply the reweighting technique to estimate the importance of a given data set on our current PDFs. The idea is based on Bayes theorem which states that the posterior distribution representing the probability of a hypothesis (new probability distribution representing the PDFs if we would perform a fit including the new data set we are using in the reweighting) is a product of the prior probability (PDFs without the new data set) and an appropriate likelihood function. This allows us to assign a weight to each of the replicas generated earlier according to eq. (3.7).

In the context of Hessian PDFs using a global tolerance criterion the appropriate weight definition is given by a modified Giele-Keller expression [50, 60, 63, 64][8]

$$w_k = \frac{e^{-\frac{1}{2}\chi_k^2/T}}{\frac{1}{N_{\text{rep}}}\sum_i^{N_{\text{rep}}} e^{-\frac{1}{2}\chi_i^2/T}}, \tag{3.9}$$

---

[7]The 68% c.l. is necessary to correspond with the variance of the PDF set defined below.

[8]In the context of Monte Carlo PDF sets a NNPDF weight definition should be used [62].

where T is the tolerance criterion used when defining Hessian error PDFs[9] and χ^2_k represents the χ^2 of the data sets considered in the reweighting procedure for a given replica k. The pPb $W^\pm$ and Z data do not provide correlated errors (the published errors are a sum of statistical and systematic errors in quadrature)[10] so it is sufficient for our analysis to use a basic definition of the χ^2 function given by

$$\chi^2_k = \sum_j^{N_{\text{data}}} \frac{(D_j - T^k_j)^2}{\sigma^2_j}, \tag{3.10}$$

where index j runs over all data points in the data set(s), N_{data} is the total number of data points, D_j is the experimental measurement at point j, σ_j is the corresponding experimental uncertainty, and T^k_j is the corresponding theoretical prediction calculated with PDFs given by replica k.

With the above prescription we can now calculate the weights needed for the reweighting procedure. The expectation value and variance of any PDF-dependent observable can now be computed in terms of weighted sums:

$$
\begin{aligned}
\langle \mathcal{O} \rangle_{\text{new}} &= \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} w_k \mathcal{O}(f_k), \\
\delta \langle \mathcal{O} \rangle_{\text{new}} &= \sqrt{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} w_k \left(\mathcal{O}(f_k) - \langle \mathcal{O} \rangle_{\text{new}} \right)^2}.
\end{aligned}
\tag{3.11}
$$

For our reweighting analysis we will only use the pPb data sets. Because the uncertainty of the nuclear PDFs dominates the proton PDFs, it is sufficient to only vary the lead PDFs. Consequently, the pPb cross sections are linear in the lead uncertainties, and we can compute the reweighting by evaluating cross sections only on the Hessian error PDFs (32+1 in case

[9]In the case of the nCTEQ15 PDFs, the tolerance criterion is $T = 35$ which corresponds to a 90% c.l., the detailed explanation on how it was defined can be found in appendix A of [7]. The tolerance factor used in this analysis corresponds to the 68% c.l. which we obtain by rescaling the above: $T \approx 35/1.645^2 \sim 13$.

[10]In our analysis we also add the normalization errors in quadrature to the statistical and systematic ones.

of `nCTEQ15`) instead of the individual replicas ($N_{\mathrm{rep}} = 10^4$)

$$\sigma_k = f^{\mathrm{P}} \otimes \hat{\sigma} \otimes \left[f_0^{\mathrm{Pb}} + \sum_i^N \frac{f_i^{\mathrm{Pb}(+)} - f_i^{\mathrm{Pb}(-)}}{2} R_{ki} \right]. \tag{3.12}$$

A similar decomposition can be used for pp or PbPb data to reduce the number of necessary evaluations. However, because of the quadratic dependence on the PDFs, the reduction is smaller and does not necessarily lead to lower computational costs.

We will compare the χ^2 for each experiment calculated with the initial PDFs (before reweighting) and with the PDFs after the reweighting procedure; this will allow us to estimate the impact of each individual data set. We do this using the following formula

$$\chi^2 = \sum_j^{N_{\mathrm{data}}} \frac{(D_j - \langle T_j \rangle)^2}{\sigma_j^2}, \tag{3.13}$$

where $\langle T_j \rangle$ is a theory prediction calculated as an average over the (reweighted or not-reweighted) replicas according to eq. (3.11) (with or without weights).

Finally, the effectiveness of the reweighting procedure can be (qualitatively) estimated by computing the *effective number of replicas* defined as [62]:

$$N_{\mathrm{eff}} = \exp \left[\frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} w_k \ln(N_{\mathrm{rep}/w_k}) \right]. \tag{3.14}$$

N_{eff} provides a measure of how many of the replica sets are effectively contributing to the reweighting procedure. By definition, N_{eff} is restricted to be smaller than N_{rep}. However, when $N_{\mathrm{eff}} \ll N_{\mathrm{rep}}$ it indicates that there are many replicas whose new weight (after the reweighting procedure) is sufficiently small that they provide a negligible contribution to the updated probability density. This typically happens when the new data is not compatible with the data used in the original fit, or if the new data introduces substantial new information; in both cases, the procedure becomes ineffective and a new global fit is recommended.

3.3.2. Reweighting using CMS $W^{\pm}$ rapidity distributions
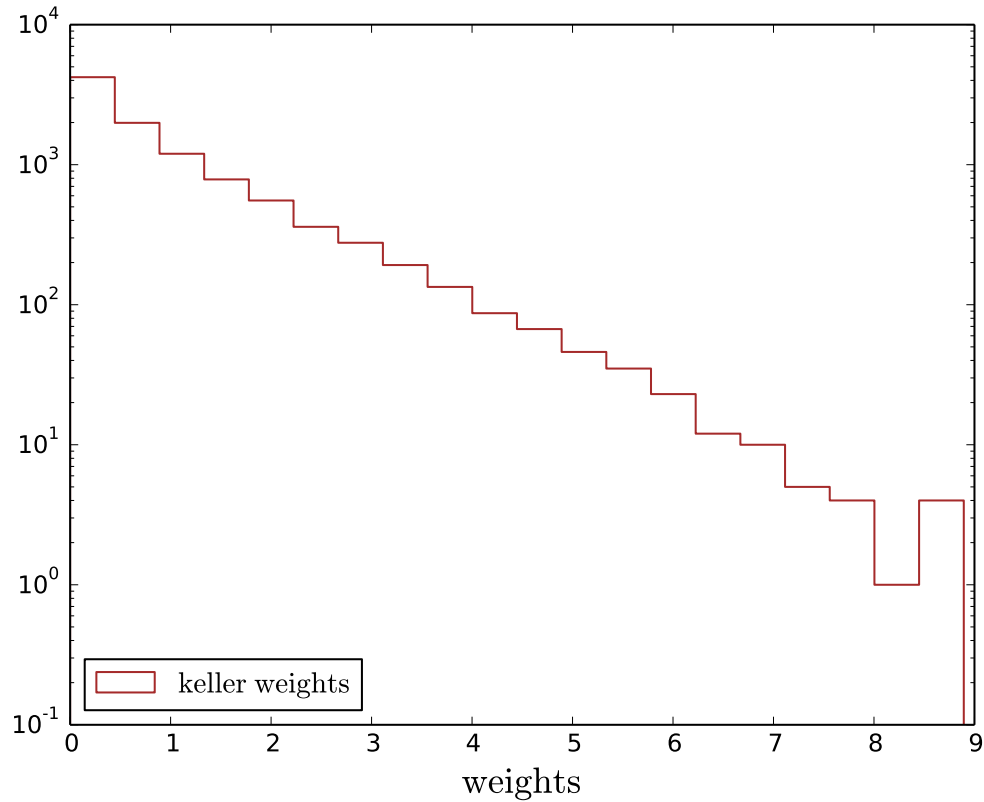
Figure 3.18: Weight distribution after reweighting using rapidity distributions of charged leptons from CMS $W^{\pm}$ production data.

As an example, we consider the reweighting using the CMS $W^{\pm}$ production data from pPb collisions [42]. In this example we use rapidity distributions of charged leptons originating from the decay of both W^+ and W^- bosons with $N_{\mathrm{rep}} = 10^4$ replicas leading to $N_{\mathrm{eff}} = 5913$.

In Fig. 3.18 we display the distribution of the weights obtained from the reweighting procedure. We see that the magnitudes of the weights are reasonable; they extend up to ~ 9 with a peak at the lowest bin. It will be useful to compare this distribution with later results using different observables and data sets.
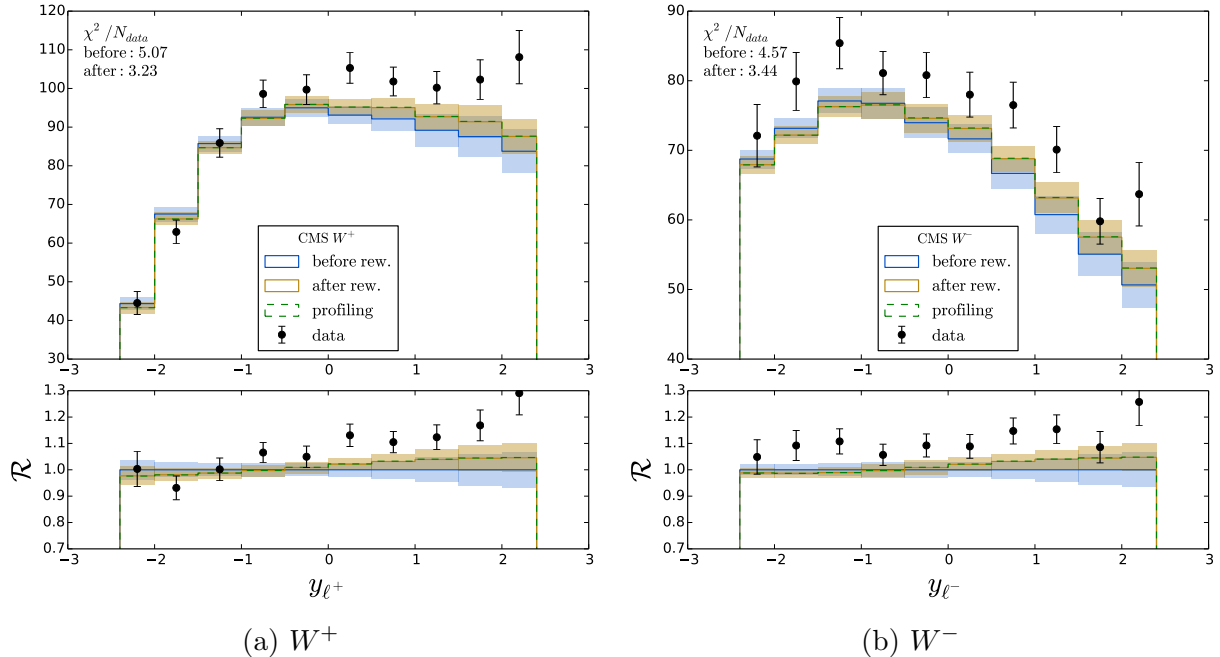
Figure 3.19: Comparison of data and theory before and after the reweighting procedure for the rapidity distributions of charged leptons in $W^{\pm}$ production measured by CMS [42].

In Fig. 3.19 we show the comparison of the data to theory before and after the reweighting procedure.[11] As expected, we see that after the reweighting procedure the description of the data is improved. This is true for both the W^{+} (left figure) and W^{-} (right figure) cases. We can quantify the improvement of the fit by examining the χ^2/N_{data} for the individual distributions. For the W^{+} case, the χ^2/N_{data} is improved from 5.07 before reweighting to 3.23 after reweighting. Similarly, for W^{-} the χ^2/N_{data} is improved from 4.57 to 3.44. The amount of change due to the reweighting procedure should be proportional to the experimental uncertainties of the incorporated data; this is the same as we would expect from a global fit. For $W^{\pm}$ production investigated here, the uncertainties are quite substantial, and the effects are compounded by the lack of correlated errors.

Finally, we show the effect of the reweighting on the PDFs themselves. In Fig. 3.20, we display PDFs for the up quark and gluon at a scale of $Q = 10$ GeV. We can see that

---

[11]We note here the difference of PDF uncertainties compared to the plots presented in Sec. 3.2; this is caused by the fact that now we use the 68% c.l. errors whereas in Sec. 3.2 we have used the 90% c.l. errors that are provided with the nCTEQ15 PDFs. This holds for all plots in Sec. 3.3.
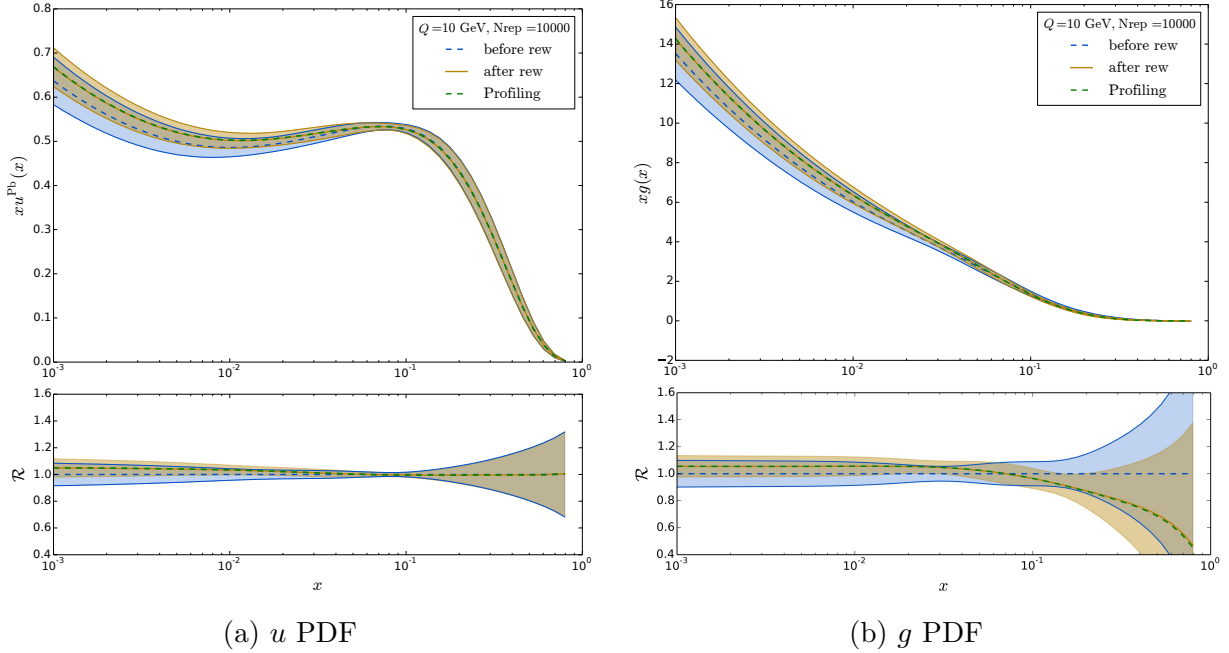
Figure 3.20: Comparison of PDFs at $Q = 10$ GeV before and after the reweighting procedure using only the CMS rapidity distribution data set. The lower plots show the ratio compared to the central (average) distribution before the reweighting.

the reweighting has the largest effects in the low x region, and this holds also for the other flavors as well. Generally the effects at intermediate and large x values are limited, with the exception of the gluon which is poorly constrained and exhibits a substantial change for large x.

In Figs. 3.19 and 3.20, in addition to the reweighting results, we also show results calculated using the Hessian profiling method [64]. The Hessian profiling should agree precisely with our reweighting calculations, and this can serve as an independent cross-check of our results. Indeed, in the figures we observe that the profiling exactly matches the reweighted results. In the following figures we will display only the reweighting results, but in all presented cases we have checked that these two methods agree.

3.3.3. Using Asymmetries instead of differential cross sections
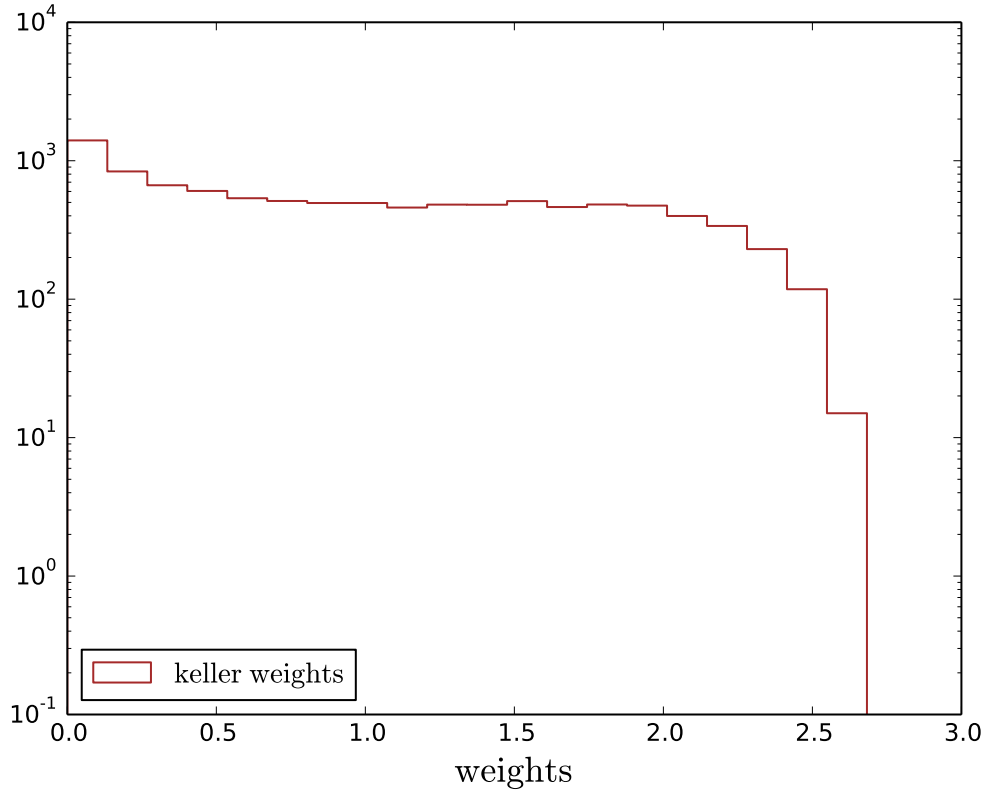
Figure 3.21: Weight distribution after reweighting using the lepton charge asymmetry and forward-backward asymmetry from CMS $W^{\pm}$ production data.

In this section we will re-investigate the reweighting analysis from the previous section employing the CMS $W^{\pm}$ production data. Instead of using rapidity distributions (as in the previous section), we will use two types of asymmetries which are constructed with the charged leptons. The lepton charge asymmetry is

$$A_\ell(y_\ell) = \frac{dN_{l^+} - dN_{l^-}}{dN_{l^+} + dN_{l^-}} \ , \tag{3.15}$$

and is defined per bin in the rapidity of the charged lepton where $N_{l^{\pm}}$ represents the corresponding number of observed events in a given bin. For the purpose of the theory calculation, $N_{l^{\pm}}$ will be replaced by the corresponding cross-section in a given bin.

(a) Charge Asymmetry

(b) Forward-backward Asymmetry
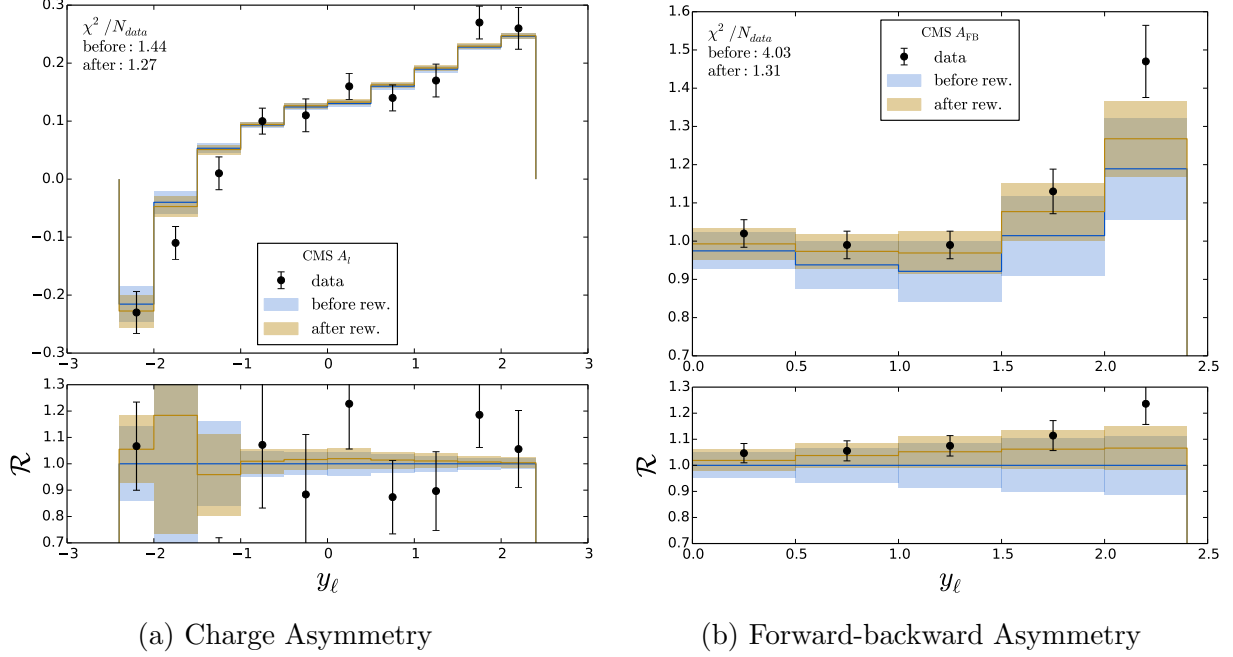
Figure 3.22: Comparison of data and theory before and after the reweighting procedure using the charge asymmetry (left) and the forward-backward asymmetry (right) from $W^{\pm}$ production data by CMS [42].

It is useful to consider the expression for the charge asymmetry at leading order in the parton model assuming a diagonal CKM matrix:

$$A_\ell = \frac{\substack{u(x_1)\bar{d}(x_2)+\bar{d}(x_1)u(x_2)+c(x_1)\bar{s}(x_2)+\bar{s}(x_1)c(x_2) \\ -\bar{u}(x_1)d(x_2)-d(x_1)\bar{u}(x_2)-\bar{c}(x_1)s(x_2)-s(x_1)\bar{c}(x_2)}}{\substack{u(x_1)\bar{d}(x_2)+\bar{d}(x_1)u(x_2)+c(x_1)\bar{s}(x_2)+\bar{s}(x_1)c(x_2) \\ +\bar{u}(x_1)d(x_2)+d(x_1)\bar{u}(x_2)+\bar{c}(x_1)s(x_2)+s(x_1)\bar{c}(x_2)}} . \tag{3.16}$$

Here, the partons with momentum fraction x_1 are in the proton, and those with momentum fraction x_2 are inside the lead. At large negative rapidities (small x_1, large x_2), we have $f(x_1) = \bar{f}(x_1)$ for all parton flavors ($f = u, d, s, c$) and the expression for the asymmetry simplifies to the following form

$$A_\ell \rightarrow \frac{d(x_1)u_v(x_2) - u(x_1)d_v(x_2) - c(x_1)s_v(x_2) + s(x_1)c_v(x_2)}{d(x_1)u_v(x_2) + u(x_1)d_v(x_2) + c(x_1)s_v(x_2) + s(x_1)c_v(x_2)} . \tag{3.17}$$

(a) u PDF   (b) g PDF
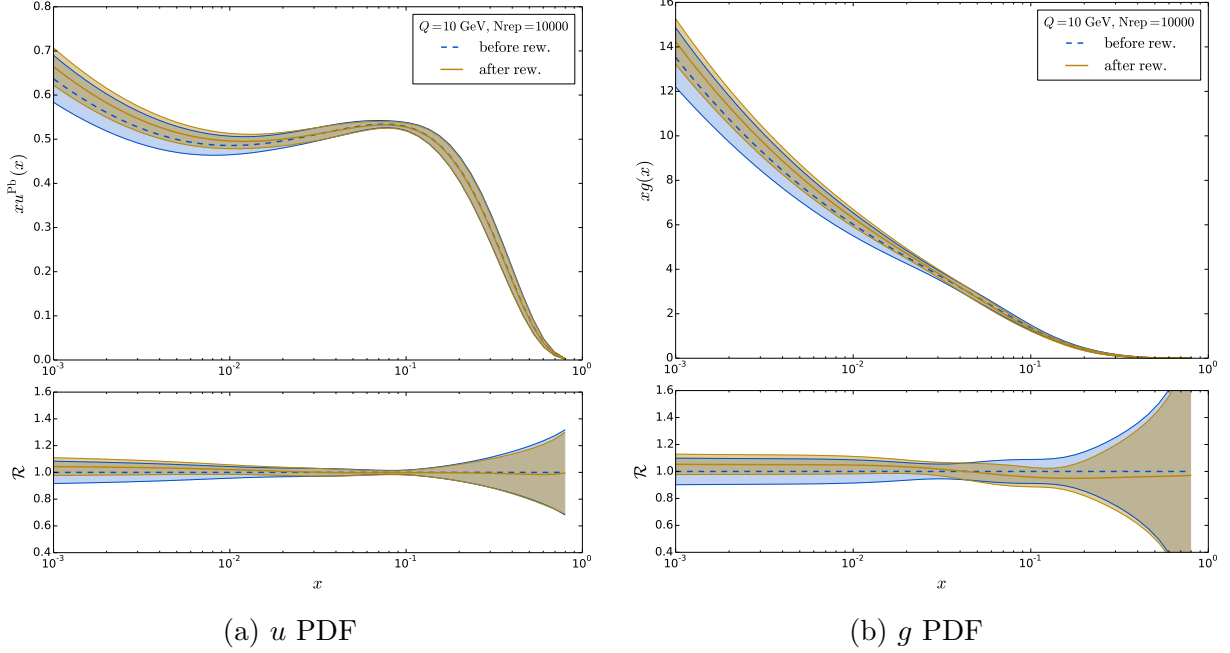
Figure 3.23: Comparison of PDFs at a scale of $Q = 10$ GeV before and after the reweighting procedure using asymmetries measured by CMS [42]. The lower plots show the ratio compared to the central (average) distribution before the reweighting.

Assuming $c_v(x_2) = c(x_2) - \bar{c}(x_2) = 0$ and $s_v(x_2) = s(x_2) - \bar{s}(x_2) = 0$, as it is the case in all the existing nPDF sets, the expression further simplifies

$$A_\ell \to \frac{d(x_1)u_v(x_2) - u(x_1)d_v(x_2)}{d(x_1)u_v(x_2) + u(x_1)d_v(x_2)} \simeq \frac{u_v(x_2) - d_v(x_2)}{u_v(x_2) + d_v(x_2)} \,. \tag{3.18}$$

In the last equation, we have used the fact that the small x_1 up and down PDFs are very similar [67]. Since the $d_v^{\mathrm{Pb}}(x_2) > u_v^{\mathrm{Pb}}(x_2)$, we expect the asymmetry to be negative at large negative rapidities. One can also observe that the asymmetry calculated with either $n_f = 2$ or $n_f = 5$ will be the same. A non-zero strange asymmetry $(s(x_2) > \bar{s}(x_2))$ would lead to a decrease of the A_ℓ asymmetry, thereby improving the description of the CMS data.

Conversely, at large positive rapidities (small x_2, large x_1), we have $f(x_2) = \bar{f}(x_2)$ for all parton flavors $(f = u, d, s, c)$ and the expression for the asymmetry becomes

$$A_\ell \to \frac{u_v(x_1)d(x_2) - d_v(x_1)u(x_2) + c_v(x_1)s(x_2) - s_v(x_1)c(x_2)}{u_v(x_1)d(x_2) + d_v(x_1)u(x_2) + c_v(x_1)s(x_2) + s_v(x_1)c(x_2)} \,. \tag{3.19}$$

Again, assuming $c(x_1) = \bar{c}(x_1)$ and $s(x_1) = \bar{s}(x_1)$, this expression further simplifies to

$$A_\ell \to \frac{u_v(x_1)d(x_2) - d_v(x_1)u(x_2)}{u_v(x_1)d(x_2) + d_v(x_1)u(x_2)} \simeq \frac{u_v(x_1) - d_v(x_1)}{u_v(x_1) + d_v(x_1)} \,, \tag{3.20}$$

where we have again used $u(x_2) \simeq d(x_2)$ at small x_2. Since $u_v(x_1) > d_v(x_1)$ in the proton, we expect a positive asymmetry in the kinematic region of large positive rapidities. Furthermore, the reweighting of the nuclear PDFs will have very little impact on the charge asymmetry in this limit even if the precision of the data will increase in the future.

Another asymmetry used by CMS is the forward-backward asymmetry. This is defined as a ratio of the number of events in the forward and backward region in a given rapidity bin:

$$A_{\rm FB}^\pm(y_\ell) = \frac{dN_{l^\pm}(+y_{\rm lab})}{dN_{l^\pm}(-y_{\rm lab})} \,. \tag{3.21}$$

This asymmetry is defined separately for the W^+ and W^- cases. It can also be combined into a single quantity, the forward-backward asymmetry of charge-summed W bosons:

$$A_{\rm FB}(y_\ell) = \frac{dN_{l^+}(+y_{\rm lab}) + dN_{l^-}(+y_{\rm lab})}{dN_{l^+}(-y_{\rm lab}) + dN_{l^-}(-y_{\rm lab})} \,. \tag{3.22}$$

This is the quantity we will use for our analysis in this section.

We now use the asymmetries of Eqs. (3.15) and (3.22) to perform a reweighting of the `nCTEQ15` lead PDFs. These asymmetries are just combinations of the rapidity distributions used in Sec. 3.3.2, and if both are employed at the same time they should encode similar information to the rapidity distributions themselves. In the literature it is sometimes argued that the asymmetries are more sensitive to the PDFs and in turn are better suited to performing PDF fits [42, 49, 50]. We will empirically check this statement by comparing reweighting predictions using rapidity distributions and the above mentioned asymmetries.

In the following, we present the results of the reweighting using the lepton charge asymmetry and forward-backward asymmetry of charge-summed W bosons. In this case, the effective number of replicas is $N_{\rm eff} = 7382$.

The distribution of weights is displayed in Fig. 3.21, and we can see that compared to the reweighting using directly the rapidity distributions (Fig. 3.18), the weights are smaller extending only to around ~ 2.7 and more evenly distributed.

In Fig. 3.22 we show a comparison of data and theory before and after the reweighting procedure. In the case of the charge asymmetry we do not see a large improvement, but this is not surprising as there is already good agreement between the data and theory before the reweighting. We note that the χ^2/N_{data} before the reweighting is 1.44 and 1.27 after the reweighting.

In the case of the forward-backward asymmetry the initial agreement between data and theory is not as good and the corresponding improvement is much larger; χ^2/N_{data} changes from 4.03 to 1.31.

We now show the effect of the reweighting procedure on the PDFs. In Fig. 3.23 we display the PDFs for the up quark and gluon at a scale of $Q = 10\,\text{GeV}$. We can see that in both cases the effect is limited to the low x region and does not exceed few percent. The results for other flavors are similar, and overall the asymmetries with the current experimental uncertainties seem to have rather small effect on the nPDFs.

In particular it seems that using asymmetry ratios yields a reduced impact, at least compared to the rapidity distributions of Sec. 3.3.2. This is possibly due to the fact that much of the information on the nuclear corrections is lost when constructing the ratios. However, asymmetries can be still useful to explore the very forward and backward regions of the rapidity distributions (corresponding to higher/lower x values) where experimental uncertainties are typically large but can cancel in the ratios.

3.3.4. Including all the data sets
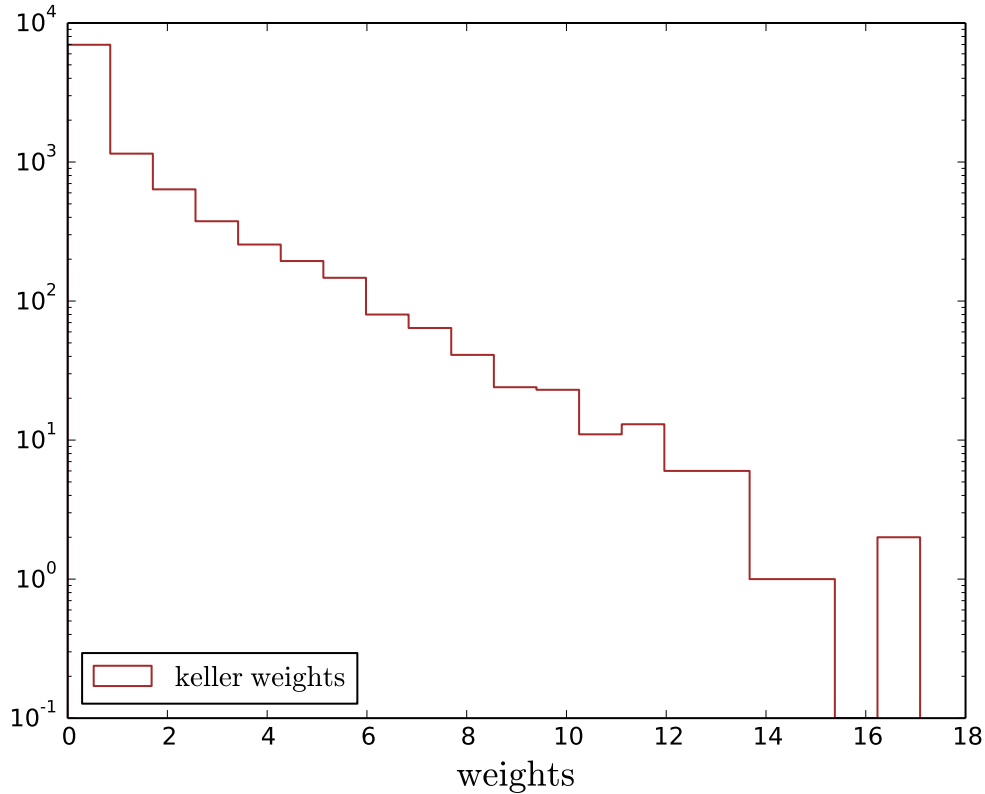
Figure 3.24: Weight distribution after reweighting using all LHC pPb data on W/Z production.

Due to large experimental uncertainties, the effect of individual data sets presented in Sec. 3.2 on the lead PDFs is rather limited. The largest constraint is obtained from the CMS $W^{\pm}$ data [42] (Secs. 3.3.2), and from (preliminary) ATLAS $W^{\pm}$ data [40]. In order to maximize the effects on the PDFs, we now employ all proton-lead data sets from Tab. 3.1 to perform the reweighting of the nCTEQ15 lead PDFs. Note that we use both the rapidity distributions and the asymmetries; although this can be regarded as "double counting", it is a common practice in proton PDF analyses, e.g. [26].

As the impact of the reweighting on the theory predictions for ALICE $W^{\pm}$ production data [44], LHCb Z data [43] and both ATLAS [39] and CMS [41] Z production data is very small, we will not show the corresponding comparisons of theory predictions before and after the reweighting. We do note that in the majority of these cases the χ^2/N_{data} has improved
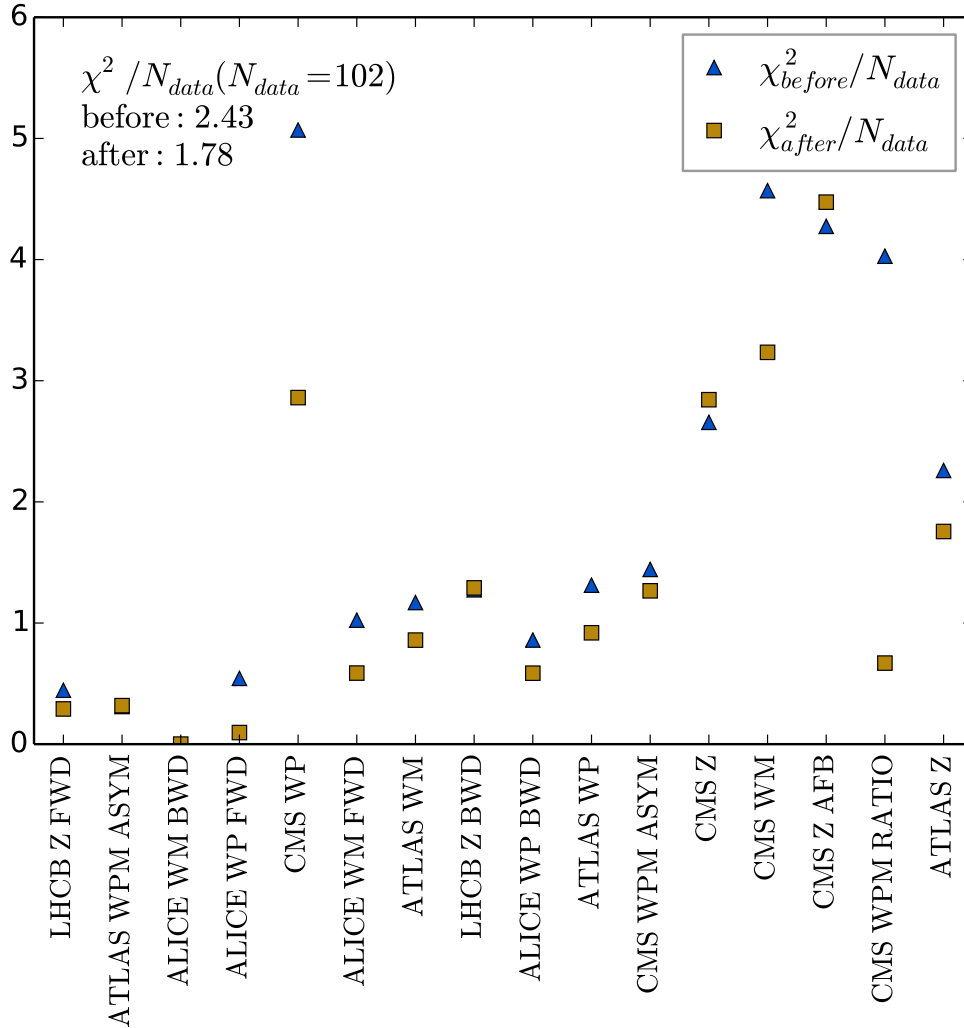
Figure 3.25: χ^2 per experiment before and after the reweighting procedure using all LHC pPb data.

indicating that the data sets are compatible, cf. Fig. 3.25. However, the initial χ^2 for these data sets was already very small which reflects the large experimental uncertainties of these data sets and their limited constraining power on the nPDFs.

We start by examining the distribution of weights of the new replicas which is displayed in Fig. 3.24. We see that the distribution is steeply falling in a similar manner to the one from Fig. 3.18 obtained using only CMS $W^\pm$ rapidity distributions, but it extends to higher values of ~ 17. These results are not very surprising as the CMS $W^\pm$ data set is the one introducing the most constraints. We also note that the reweighting procedure results in
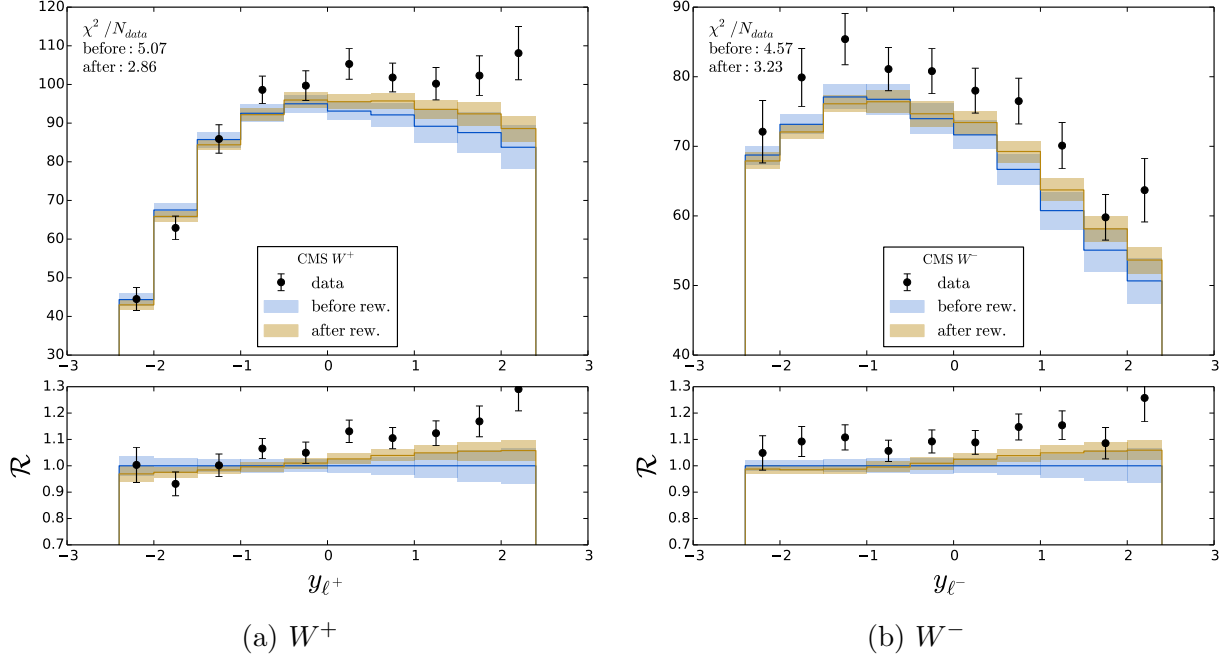
Figure 3.26: Comparison of data and theory before and after the reweighting procedure using all LHC pPb data. The results for the CMS W^+ (left) and W^- (right) distributions are shown.

the effective number of replicas $N_{\text{eff}} = 3603$ which is around 40% of the number of initial replicas. This suggests that the reweighting procedure should still yield reliable results.

Now we turn to the comparison of data with the theory predictions before and after the reweighting procedure. In Fig. 3.26 we show the predictions for the CMS $W^\pm$ data [42], and in Fig. 3.27 we show the corresponding predictions for the ATLAS $W^\pm$ data [40]. We can see that in both cases we observe an improvement in the data description that is confirmed by the corresponding χ^2/N_{data} values (see figures). The χ^2 values tell us also that the largest effect comes from the CMS data which has smaller errors and for which the initial description (before the reweighting) was worse than in the ATLAS case.

Furthermore, comparing the values of χ^2/N_{data} for the CMS $W^\pm$ data after the reweighting using all data sets and using only CMS data (Sec. 3.3.2) we see further improvement of χ^2/N_{data} when more data is included. This shows that the different data sets are compatible with each other and that they pull the results in the same direction.
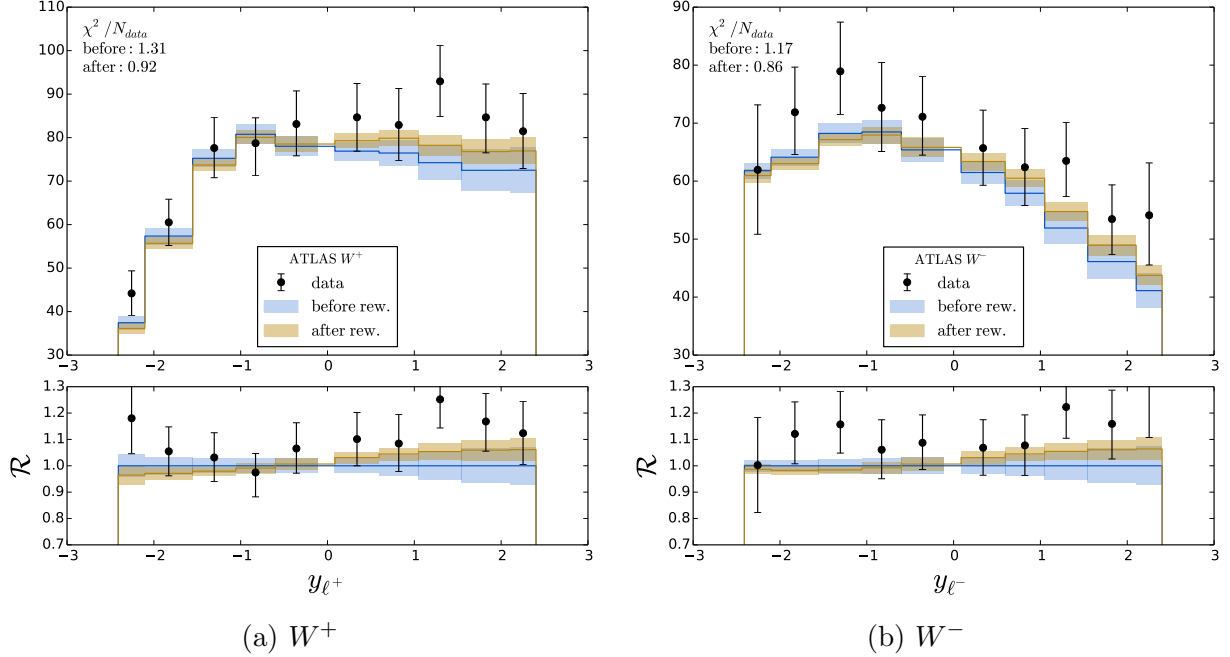
Figure 3.27: Comparison of data and theory before and after the reweighting procedure using all LHC pPb data. The results for the ATLAS W^+ (left) and W^- (right) distributions are shown.

In addition, we show in Fig. 3.25 the χ^2/N_{data} before and after the reweighting for each of the experiments, as well as the χ^2/N_{data} combining all 102 data points from the different experiments. This highlights the fact that the CMS $W^\pm$ measurement yields the largest impact on the PDFs out of all the considered data sets.

Finally, in Figs. 3.28-3.30 we present the effects of the reweighting on the $\{u, d, \bar{u}, \bar{d}, g, s\}$ distributions in lead for a scale $Q = 10$ GeV. The effects are similar when looking at different scales. From the figures we can see that changes in the PDFs are generally affecting the low-x distributions, and to a lesser extent the moderate to high-x distributions.

When considering the ratios of PDFs, the effects of the reweighting appear to be quite substantial at large x, especially for the gluon; however, as is evident from looking at the plots of the PDFs directly, they are approaching zero at large x so the impact for physical observables is minimal.

Furthermore, when interpreting the results of the reweighting analysis it is important to remember that this method can only estimate the effects a given data set might have on the

(a) u PDF (b) d PDF
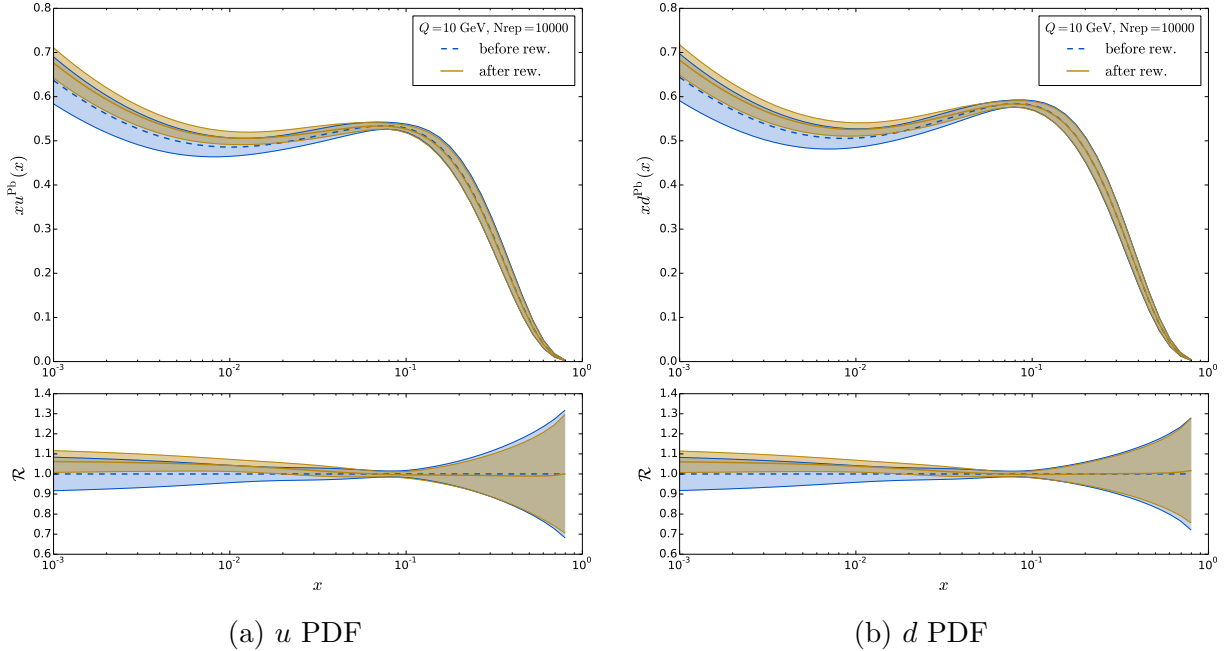
Figure 3.28: u and d PDFs before and after the reweighting using all LHC pPb data sets.

PDFs; it is not equivalent to a full fit. For example, a reweighting analysis cannot be used to explore new parameters or other dimensions that are not already spanned by the original PDF uncertainty basis. In particular, this study has shown us that the strange quark PDF can play an important role in the LHC pPb production of W/Z. As our current $s(x)$ is parameterized proportional to $\bar{u}(x) + \bar{d}(x)$, this restricts our ability to vary the strange PDF independently;[12] hence, an independent fit (in progress) is needed to better the impact of this data on the nPDFs.

3.3.5. Comparison with EPPS16

During the course of our analysis, a new global fit including LHC data (EPPS16 [53]) has been released. This gives us an opportunity to compare the results of our reweighting study with these new PDFs. We note here that this is a qualitative comparison as the data sets used in these two studies are different. Another important difference is that the EPPS16 fit has more parameters to describe the sea-quark PDFs as compared to the nCTEQ15 analysis;

---

[12]This point was explored in more detail in ref. [68].
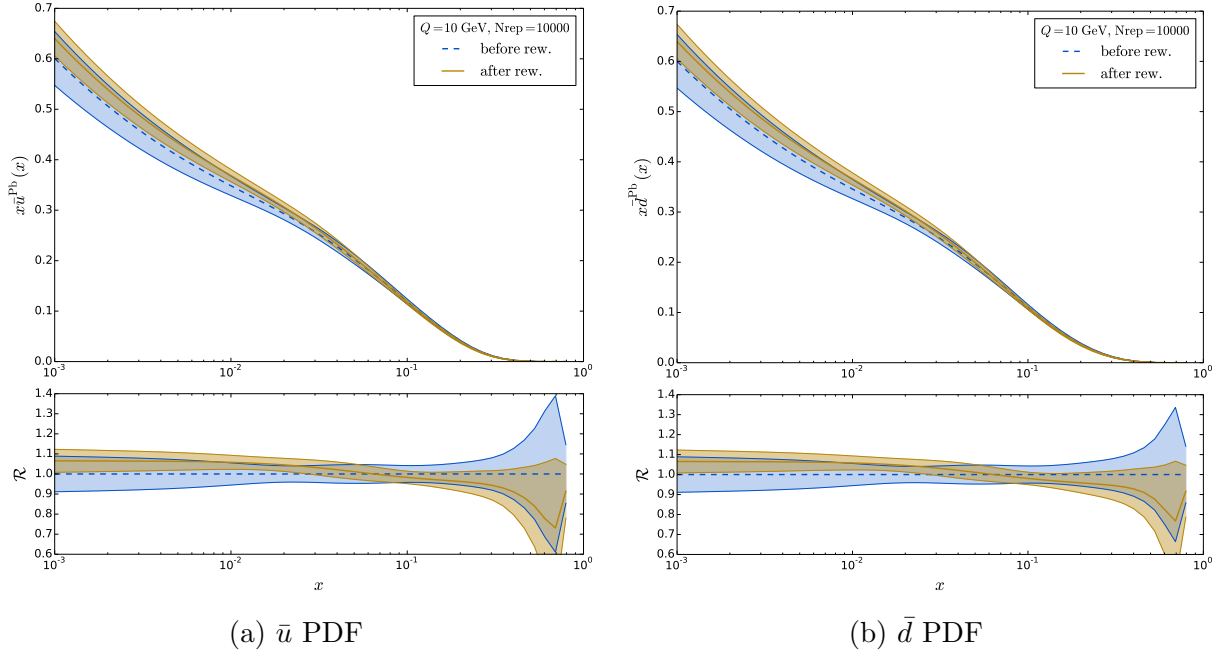
Figure 3.29: $\bar{u}$ and $\bar{d}$ PDFs before and after the reweighting using all LHC pPb data sets.

this provides EPPS16 additional flexibility to accommodate all the considered data. As mentioned earlier, our reweighting of nCTEQ15 cannot compensate for our more restrictive parametrization, so this must be considered when evaluating these comparisons.

In Figs. 3.31 and 3.32 we present a comparison of u, d, $\bar{u}$, $\bar{d}$, g and s for the nCTEQ15 PDFs before and after the reweighting, with the EPPS16 distributions at the scale of 80 GeV. There are a number of trends which emerge:

1. In the low x region, the reweighted nCTEQ15 PDFs approach the EPPS16 distributions; for the g and s PDFs, the central values are very close. The effect of the reweighting appears mostly in this region where (prior to the LHC data) there were minimal constraints on the PDFs. Therefore, adding the LHC data is able to significantly adjust the PDFs in this region.

2. In the intermediate x range ($\sim 3 \times 10^{-2}$), the central values of the EPPS16 and both reweighted and initial nCTEQ15 PDFs coincide, and their uncertainty bands are also similar (except for the strange quark). This region was previously constrained by pre-LHC data, and we observe minimal changes in this region.

91
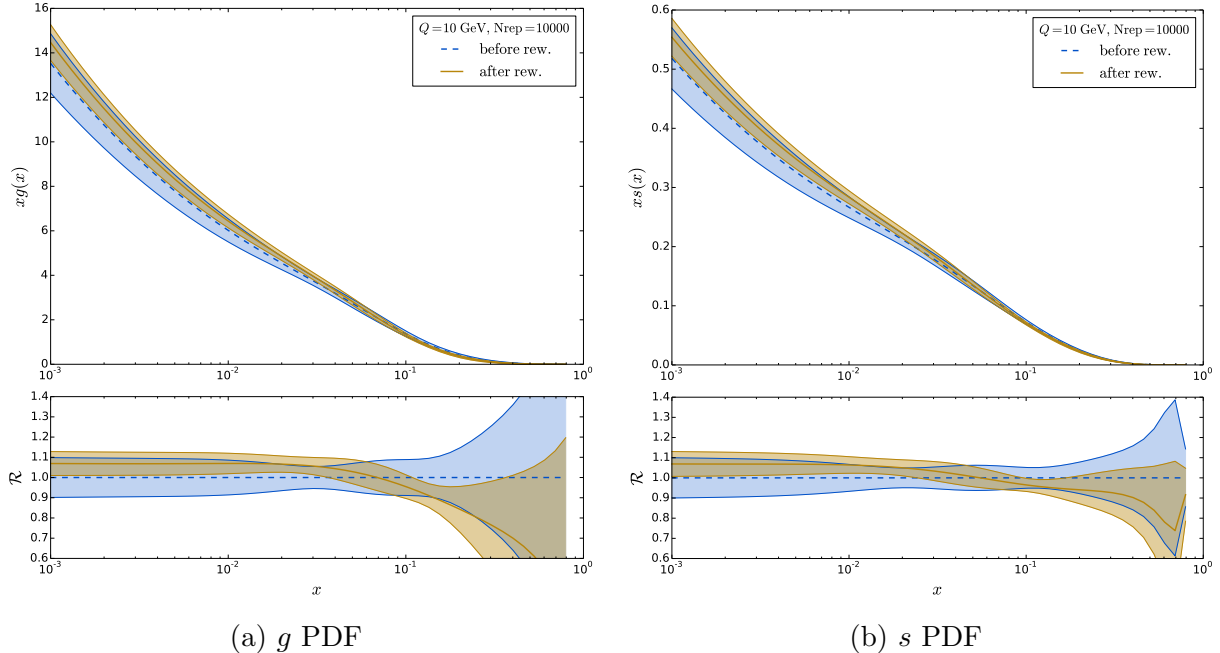
Figure 3.30: g and s PDFs before and after the reweighting using all LHC pPb data sets.

3. On the contrary, where x is large, the differences are more important with no consistent pattern. This is a challenging region as the absolute value of the PDFs is small, and the nCTEQ15 parameterization may not be sufficiently flexible to accommodate the new data. Additionally, the inclusion of certain data sets in the EPPS16 analysis (such as the CHORUS ν-Pb data [69]) can have a significant impact.

Finally, we also see that the EPPS16 PDFs have consistently larger uncertainty bands (especially at low x). As the nCTEQ15 uncertainty bands in this region are essentially extrapolated from larger x results, the EPPS16 uncertainties are probably a more realistic assessment. The issue of PDF parameterization is a perennial challenge for the nuclear PDFs as there is less data and more degrees of freedom as compared to the proton PDFs. The common solution is to impose assumptions on the nPDF parameters, or to limit the flexibility of the parameterization, and thereby underestimate the uncertainty. These issues highlight the importance of including this new LHC data in the nPDF analyses as they not only will help determine the central fits, but also provide for more reliable error estimation.

(a) u PDF

(b) d PDF

(c) $\bar{u}$ PDF

(d) $\bar{d}$ PDF
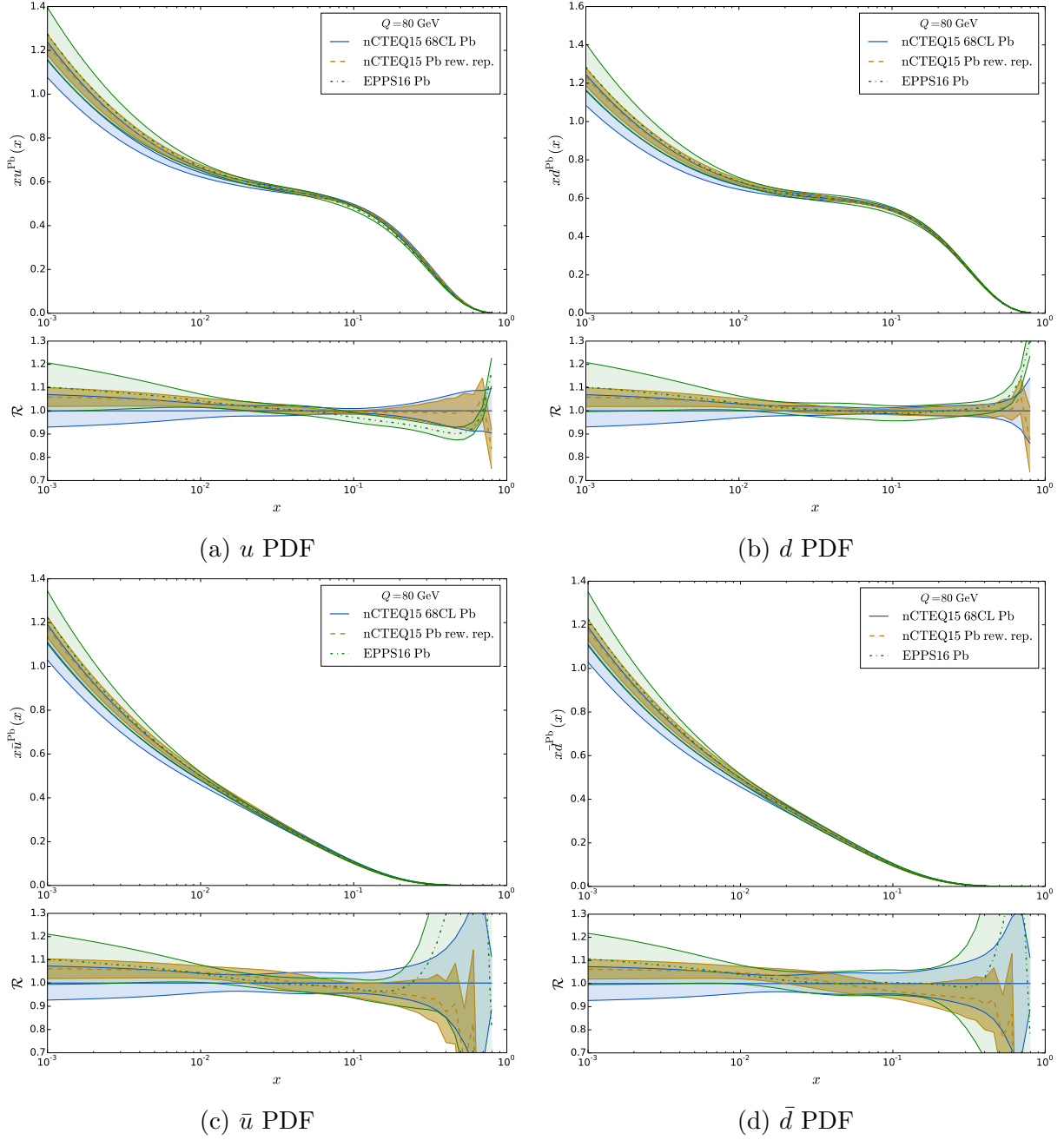
Figure 3.31: Comparison of the nCTEQ15 PDFs before and after the reweighting using all pPb data sets with the EPPS16 PDFs including LHC data. The EPPS16 error bands include only the nuclear errors (unlike what is provided in LHAPDF where also the proton baseline errors are included) and they are calculated using symmetric formula.

(a) g PDF

(b) s PDF
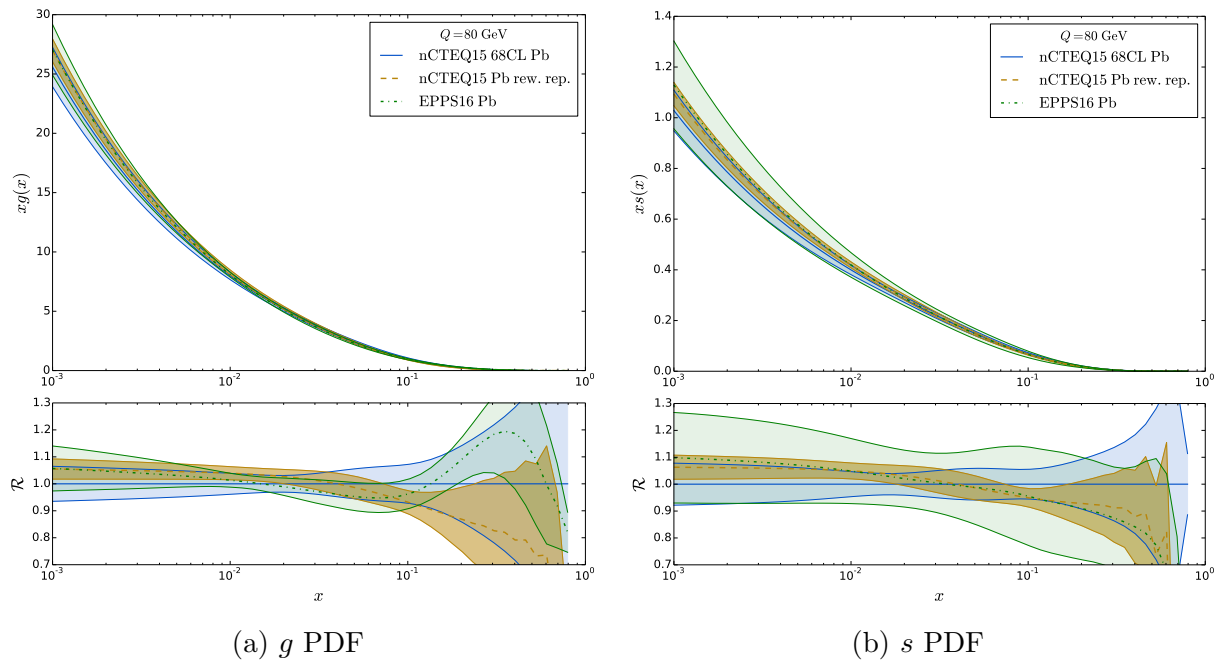
Figure 3.32: Continuation of Fig. 3.31.

3.4. Conclusions

We have presented a comprehensive study of vector boson production $(W^\pm, Z)$ from lead collisions at the LHC. This LHC lead data is of particular interest for a number of reasons:

1. Comparisons with LHC proton data can determine nuclear corrections for large A values; this is a kinematic $\{x, Q^2\}$ range very different from nuclear corrections provided by fixed-target measurements.

2. The $W^\pm, Z$ lead data are sensitive to the heavier quark flavors (especially the strange PDF), so this provides important information on the nuclear flavor decomposition.

3. Improved information on the nuclear corrections from the LHC lead data can also help reduce proton PDF uncertainties as fixed-target nuclear data is essential for distinguishing the individual flavors.

Predictions from the recent `nCTEQ15` nPDFs are generally compatible with the LHC experimental data; however, this is partially due to the large uncertainties from both the nuclear corrections and the data. We do see suggestive trends (for example $W^\pm$ production in pPb at large $y_{\ell+}$) which may impose influential constraints on the nPDF fits as the experimental uncertainties are reduced. Intriguingly, the large rapidity W/Z data seem to prefer nuclear PDFs with no or delayed shadowing at small x, similar to what has been observed in ν-Fe DIS. This observation was validated by our reweighting study that demonstrated the impact of the W/Z pPb data on nPDFs.

The uncertainties of the currently available data are relatively large, and correlated errors are not yet available. Fortunately, we can look forward to more data (with improved statistics) in the near future as additional heavy ion runs are scheduled.

While the above reweighting technique provides a powerful method to quickly assess the impact of new data, there are limitations. For example, the reweighting method cannot introduce or explore new degrees of freedom. Thus, if the original fit imposes artificial constraints (such as linking the strange quark PDF to the up and down sea distributions), this limitation persists for the reweighted PDF [68].

Most importantly, our correlation study (Sec.II.D) demonstrated the importance of the strange quark distribution for the vector boson (W/Z) production at the LHC, possibly even pointing to a nuclear strangeness asymmetry ($s(x) > \bar{s}(x)$). The comparison of the 2 flavor and 5 flavor results illustrates how flavor decomposition and nuclear corrections can become entangled. Therefore, it is imperative to separately control the strange PDF and the nuclear correction factor if we are to obtain unambiguous results. The investigations performed in this paper provide a foundation for improving our determination of the PDFs in lead, especially the strange quark component. Combining this information in a new nCTEQ fit across the full A range can produce improved nPDFs, and thus yield improved nuclear correction factors. These improved nuclear correction factors, together with the LHC W/Z production data for pp, can refine our knowledge of the strange PDF in the proton.

Chapter 4

Heavy Flavor Variable Number Schemes

In a brief detour from the study of the effects of LHC data on `nCTEQ` nPDFs, this chapter provides a deeper look at the theoretical choices that underlie PDF fitting, specifically as those choices pertain to heavy flavor number schemes. While this analysis has little direct effect on fitting `nCTEQ15` with LHC data, the lessons learned here proved invaluable when validating `nCTEQ++` (See Section 5.2).

4.1. Introduction

The global analyses of PDFs has progressed significantly in recent years. On the experimental front, there is data ranging from the fixed-target regime at low energy, on to HERA and the LHC at very high energies. On the theoretical front, the analysis can be performed not only at NLO, but now at NNLO. To capitalize on these advances, it is essential to include a proper treatment of the heavy quarks to enable high precision phenomenological analysis of measurements.

The Variable Flavor Number Scheme (VFNS) allows us to deal with the heavy quark mass scale across the full kinematic range by varying the number of active flavors (N_F) in the DGLAP QCD evolution [29, 70–79]. At low energy scales, the DGLAP evolution only involves N_F light flavors, and there is no PDF for the heavy quark. At high energy, the heavy quark PDF is included in the DGLAP evolution so that there are now $N_F + 1$ active flavors. To combine the above N_F and $N_F + 1$ sub-schemes into a single VFNS, we must define an energy scale μ_m where we match these together; this will be the scale where we introduce the heavy quark PDF.
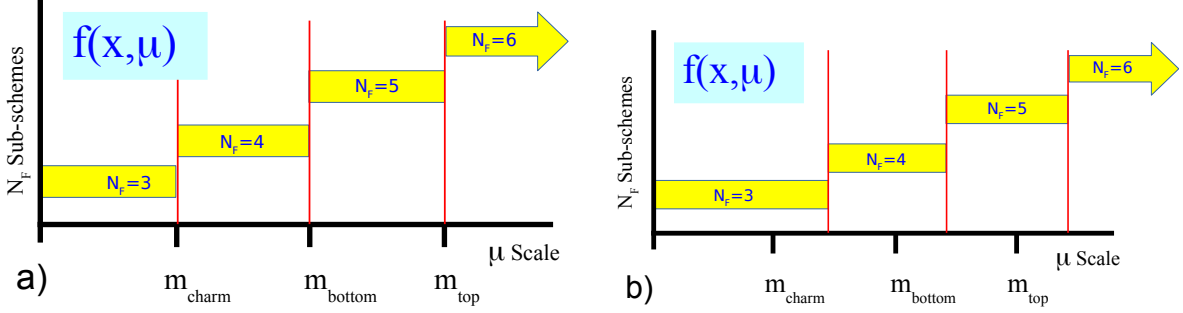
Figure 4.1: An illustration of the separate N_F renormalization sub-schemes which define a VFNS. Historically, the matching scales μ_m were chosen to be exactly the mass values $m_{c,b,t}$ as in Figure-a. Figure-b is a generalized case where the matching scales μ_m are chosen to be different from the mass values.

Historically, the matching scale μ_m was taken to be the heavy quark mass m_H. At the matching scale, the PDFs and $\alpha_S(\mu)$ for $N_F + 1$ are defined in terms of the N_F quantities by the following boundary conditions:

$$
f_i^{(N_F+1)}(x,\mu_m) \;=\; \sum_j \mathcal{M}_i^j \;\otimes\; f_j^{(N_F)}(x,\mu_m) \tag{4.1}
$$

$$
\alpha_S^{(N_F+1)}(\mu_m) \;=\; \alpha_S^{(N_F)}(\mu_m) \;\times
$$

$$
\times \left(1 + \sum_{n=1}^{\infty} \sum_{k=0}^{n} c_{n\,k} \left[\alpha_S^{(N_F)}(\mu_m) \right]^n \; \ln^k \frac{\mu_m^2}{m_H^2} \right) \;. \tag{4.2}
$$

The matching matrix $\mathcal{M}_i^j$ and coefficients $c_{n\,k}$ can be perturbatively computed.[1]

The new `xFitter` 2.0.0 program[2] links to the `APFEL` code [21] which has implemented generalized matching conditions that enable the switch from N_F to $N_F + 1$ at an arbitrary matching scale μ_m. This allows us to introduce the heavy quark PDF at any scale—not just at $\mu_m = m_H$; this flexibility provides a number of advantages. For example, as the matching

---

[1]The perturbative coefficients of $\mathcal{M}_i^j$ at NLO are available in Refs. [80, 81], and at NNLO in Ref. [82]. m_H is the mass of the $N_F + 1$ flavor quark. For $\alpha_S(\mu)$, the $c_{n\,k}$ coefficients are available in the Particle Data Group review of Quantum Chromodynamics [83].

[2]Information on the `xFitter` program can be found at www.xFitter.org, and in Refs. [84, 85].

scale moves to higher scales, the theory at the lower scales effectively becomes a Fixed Flavor Number Scheme (FFNS); yet we still retain a VFNS at the higher scales.

The choice of the matching scale μ_m, like the choice of VFNS or FFNS, amounts to a theoretical scheme choice. As such, the variation of μ_m represents a source of theoretical uncertainty. The variable matching scale implemented in `xFitter` provides a new incisive tool to study the impact of these choices across a broad kinematic region. Additionally, as we move from NLO to NNLO calculations, new features are encountered, and these compel us to reexamine some of the foundational elements used to construct this theoretical framework.

Reconsidering the historical choice $\mu_m = m_H$ is of particular relevance for heavy-quark initiated processes at the LHC. In this context, the benefits of the FFNS close to the threshold region and of the VFNS at higher scales are often simultaneously needed to describe the data. Therefore, a careful choice of the matching scales could help formulate a matching prescription between FFNS and VFNS able to achieve this goal in a very simple fashion [86].

This study will examine the combined HERA data set and evaluate the impact of the matching scale on the features of the fit of PDFs. In Sect. 4.2, we review the key elements of the VFNS used in this study. Sect. 4.3, shows the impact of the matching scale μ_m on the PDFs. In Sect. 4.4, we perform a fit of the combined HERA data sets at both NLO and NNLO, and investigate the effect of the matching scale μ_m. Sect. 4.5 presents an example of how the μ_m flexibility can be used as a tool to evaluate a recent suggestion for a N_F dependent PDF. Sect. 4.6 summarizes the general characteristics and conclusions of this study.

4.2. Variable Flavor Number Scheme (VFNS)

Here we will outline the key concepts of the heavy quark VFNS which are relevant for this investigation.

4.2.1. The matching scale μ_m

A generalized formulation of the VFNS factorization is based on the Collins-Wilczek-Zee (CWZ) renormalization scheme which involves a sequence of sub-schemes parameterized by the number of active quark flavors (N_F) [87,88]. For each sub-scheme, the N_F (active) flavors are renormalized using the $\overline{\text{MS}}$ scheme while the heavy (inactive) flavors are renormalized using zero-momentum subtraction. This ensures that to all orders in perturbation theory (i) the results are gauge invariant, (ii) the results for the active N_F flavors match the standard $\overline{\text{MS}}$ results, and (iii) the heavy (inactive) flavors manifestly decouple.[3] Specifically, both the DGLAP evolution kernels for the N_F active PDFs and the renormalization group equation for $\alpha_S^{(N_F)}(\mu)$ are pure $\overline{\text{MS}}$.

To connect the separate N_F sub-schemes into a single scheme that spans the full kinematic range, we must choose a matching scale μ_m which will relate the sub-schemes. This is where we define the PDFs and α_S of the $N_F + 1$ scheme in terms of the N_F scheme, *cf.* Eqs. (4.1) and (4.2). A schematic representation of this is displayed in Fig. 4.1.

For example, at scales $\mu_c < \mu < \mu_b$ the scheme has $N_F = 4$ active flavors $\{u, d, s, c\}$ with 4-flavor PDFs and $\alpha_S^{(4)}(\mu)$; the bottom quark is **not** treated as a parton and $f_b^{(4)}(x, \mu) = 0$.

At the scale $\mu = \mu_b$, we can compute the 5-flavor PDFs and $\alpha_S^{(5)}(\mu)$ in terms of the 4-flavor quantities; the boundary conditions are non-trivial and the PDFs and $\alpha_S(\mu)$ are not necessarily continuous. This scheme has $N_F = 5$ active flavors $\{u, d, s, c, b\}$, and the bottom quark is included in the DGLAP evolution.

4.2.2. Historical choice of $\mu_m = m_{c,b,t}$

Historically, the matching scale μ_m was commonly taken to be exactly equal to the mass of the heavy quark $\mu_m = m_{c,b,t}$; this was a convenient choice for a number of reasons.

---

[3]For the CWZ scheme with N_F (active) flavors and an arbitrary number of heavy (inactive) flavors, the evolution of the PDFs and $\alpha_S^{(N_F)}(\mu)$ will involve only the active N_F flavors; the inactive heavy flavors can be ignored.

For example, the generic NLO matching condition for the PDFs at the $N_F = 4$ to $N_F = 5$ transition is [20]:

$$f_i^{(5)}(x, \mu_b) = \left\{ \delta_{ij} + \frac{\alpha_S^{(4)}(\mu_b)}{2\pi} \left[c_0^{ij} + c_1^{ij} \ln\left(\frac{\mu_b^2}{m_b^2}\right) \right] \right\} \otimes f_j^{(4)}(x, \mu_b) \qquad (4.3)$$

where c_0^{ij} and c_1^{ij} are perturbatively calculable coefficient functions. Note that the right-hand side uses 4-flavor PDFs and α_S, while the left-hand side uses 5-flavors.

The choice $\mu_b = m_b$ will cause the logarithms to vanish, and this greatly simplifies the matching relations. Additionally, at NLO in the $\overline{\text{MS}}$ scheme the constant term c_0^{ij} in the matching equation coincidentally vanishes [82]. The net result is that for $\mu_b = m_b$, the PDFs will be continuous (but not differentiable) at NLO. This is historically why μ_m was set to $m_{c,b,t}$.

However, at NNLO and beyond the situation is more complex; in particular, the higher-order terms corresponding to c_0^{ij} will be non-zero, and the matching of both the PDFs and $\alpha_S(\mu)$ will be discontinuous. Consequently, the freedom to arbitrarily choose the matching scale μ_m (and decide where to place the discontinuities) will have a number of advantages, as the next subsection will demonstrate.

4.2.3. Smooth matching across flavor thresholds
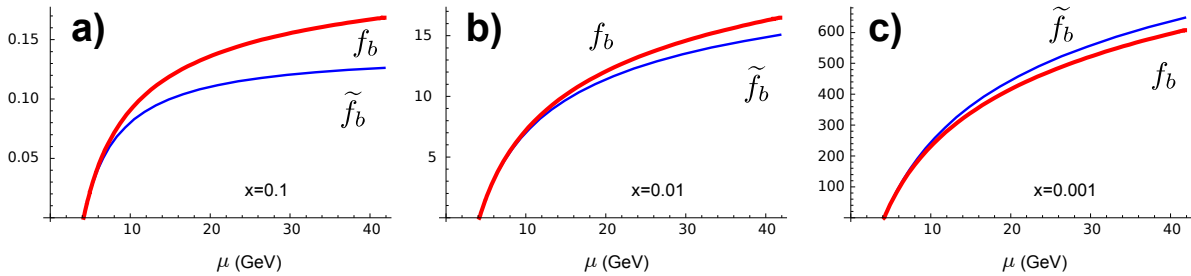
Figure 4.2: The comparison of the DGLAP-evolved PDF $f_b(x, \mu)$ and the perturbatively calculated $\tilde{f}_b(x, \mu)$ as a function of μ for selected x values. For $\mu \to m_b$ we find the functions match precisely: $\tilde{f}_b(x, \mu) \to f_b(x, \mu)$. We have used NNPDF30_lo_as_118_nf_6 as the base PDF set.

To gauge the impact of the contributions of the heavy quark PDFs in a process independent manner, we can compare the DGLAP-evolved heavy quark PDF $f_b(x,\mu)$ with a perturbatively computed quantity: $f\widetilde{f}_b(x,\mu)$. At NLO, $f\widetilde{f}_b(x,\mu)$ takes a gluon PDF and convolutes it with a perturbative (DGLAP) splitting $g \to b\bar{b}$ [89,90]; this can be thought of as a "perturbatively" computed bottom PDF. The result at NLO is:

$$f\widetilde{f}_b(x,\mu) = \frac{\alpha_S}{2\pi} \; P_{g\to b\bar{b}} \otimes f_g \;\; \ln\left[\frac{\mu^2}{m_b^2}\right] \;\;\; . \tag{4.4}$$

The difference between $f_b(x,\mu)$ and $f\widetilde{f}_b(x,\mu)$ is due to the higher order terms which are resummed by the heavy quark DGLAP evolution.[4]

To better understand these quantities, we compute DIS bottom production at NLO in a 5-flavor VFNS, and find the cross section to be [72]:

$$\sigma_{VFNS} = \sigma_{b\to b} \otimes \left[f_b(x,\mu) - f\widetilde{f}_b(x,\mu)\right] + \underbrace{\sigma_{g\to b} \otimes f_g(x,\mu)}_{\sim\, \sigma_{FFNS}} \;\;\; . \tag{4.5}$$

Here, $\sigma_{b\to b} \otimes f_b$ is the LO term, and $\sigma_{b\to b} \otimes f\widetilde{f}_b$ is the subtraction (SUB) term. The unsubtracted NLO term $\sigma_{g\to b} \otimes f_g$ corresponds (approximately) to a FFNS calculation. Here, $\sigma_{b\to b}$ is proportional to a delta function which makes the convolution trivial.

Thus, the combination $(f_b - f\widetilde{f}_b)$ represents (approximately) the difference between a VFNS and FFNS result.[5] These quantities are displayed in Fig. 4.2. In the region $\mu \sim m_b$, $f_b(x,\mu)$ and $f\widetilde{f}_b(x,\mu)$ match precisely; it is this cancellation which (at NLO) ensures physical quantities will have a smooth transition across the flavor threshold.

At larger μ scales, $f_b(x,\mu)$ and $f\widetilde{f}_b(x,\mu)$ begin to diverge; this indicates that the resummed heavy quark logarithms are becoming sizable. The details clearly depend on the specific x values. For large x ($x \sim 0.1$) we find $f_b(x,\mu) > f\widetilde{f}_b(x,\mu)$, while for small x

---

[4]In Eq. (4.4), $f\widetilde{f}_b(x,\mu)$ includes the single splitting ($g \to b\bar{b}$); in contrast, the DGLAP evolution of $f_b(x,\mu)$ sums an infinite tower of splittings. Note, we have used the NNPDF30_lo_as_118_nf_6 PDFs to precisely match the order of the splitting kernels in the NLO calculation.

[5]The above correspondences are only approximate as the VFNS and FFNS also differ in α_S and the PDFs.

$(x \sim 0.001)$ the result is $f_b(x, \mu) < f\widetilde{f_b}(x, \mu)$; finally, for intermediate x $(x \sim 0.01)$ the two terms nearly balance even for sizable μ scales.

While the QCD theory ensures proper matching, this is not so easy to implement in a general numeric calculation for all observables, especially for complex observables involving multiple numeric integrations. In particular, the cancellation of Fig. 4.2 requires that the quark masses $m_{c,b,t}$, the strong coupling α_S, and the order of the PDF evolution are exactly matched in (i) the DGLAP evolution that generates the PDFs, (ii) the partonic cross sections that are convoluted with the PDFs, and (iii) the fragmentation function (if used).

In practice, there are almost always slight differences. A typical analysis might use a variety of PDFs from different PDF groups, together with a selection of fragmentation functions; each of these will be generated with a specific set of quark masses and α_S values which are most likely different. Thus, it is essentially inevitable that the cancellations exhibited in Fig. 4.2 will be spoiled leading to spurious contributions which can be substantive.

Instead of setting the matching scale at the heavy quark mass $\mu_m = m_{c,b,t}$, xFitter provides the flexibility to delay the matching scale μ_m to a few multiples of the heavy quark mass; this will avoid the need for the delicate cancellation in the $\mu_m \sim m_{c,b,t}$ region, and the results will be numerically more stable.

As an extreme example, one could imagine delaying the matching scale to infinity ($\mu_m \to \infty$) which would amount to a FFNS; here, the disadvantage is that the FFNS does not include the resummation of the higher-order heavy quark logs which have been demonstrated to improve the fit to the data [91]. Using the new flexibility of the xFitter program, it is possible to investigate the trade-offs between a large and small value for the matching scale μ_m.

A separate example is present in the transverse momentum (p_T) distributions for heavy quark production ($pp \to b\bar{b}$) using the (general mass) GM-VFNS [92, 93]. If we compute this in an $N_F = 5$ flavor scheme, the contribution from the $b\bar{b} \to b\bar{b}$ sub-process with an exchanged t-channel gluon will be singular at $p_T = 0$. For a scale choice of the transverse

mass $\mu = \sqrt{p_T^2 + m_b^2}$ (a common choice), the singularity can be cured by either a different scale choice, or by delaying the switch to the 5-flavor scheme to a higher scale, *e.g.*, $\mu_b \sim 2m_b$.

4.2.4. Discontinuities

At NNLO both the PDFs and the $\alpha_S(\mu)$ will necessarily have discontinuities when matching between the N_F to $N_F + 1$ flavor schemes as specified by Eqs. (4.1) and (4.2). If we are analyzing a high precision experiment and arbitrarily impose a matching at the quark masses $\mu_m = m_{c,b,t}$, this may well introduce discontinuities within the kinematic range of some precision data. While it is true that these discontinuities simply reflect the theoretical uncertainties, it is disconcerting to insert them in the middle of a precision data set.

The ability to vary the matching scale μ_m provides us with the option to shift the location of these discontinuities for a particular analysis. For example, to analyze the high-precision charm production HERA data, we necessarily are working in the region of the bottom mass scale (~ 4.5 GeV). Both the PDFs and $\alpha_S(\mu)$ will be discontinuous at the matching scale which transitions between the $N_F = 4$ and $N_F = 5$ schemes. If the matching scale is chosen in the region $\mu_m \sim m_b$, these discontinuity will appear in the region of the data. Instead, we can shift the matching μ_m to a higher scale (for example, set μ_m to $2m_b$ or $3m_b$) and thus analyze the charm production data in a consistent $N_F = 4$ flavor framework. Yet, we still retain the transition to $N_F = 5$ flavors so that processes such as LHC data at high scales are computed including the bottom PDF.

4.3. The matching scale μ_m

Having sketched the characteristics of a flexible matching scale μ_m, we will examine the specific boundary condition and the impact on the global fit of the PDFs.

4.3.1. Impact of matching on the PDFs
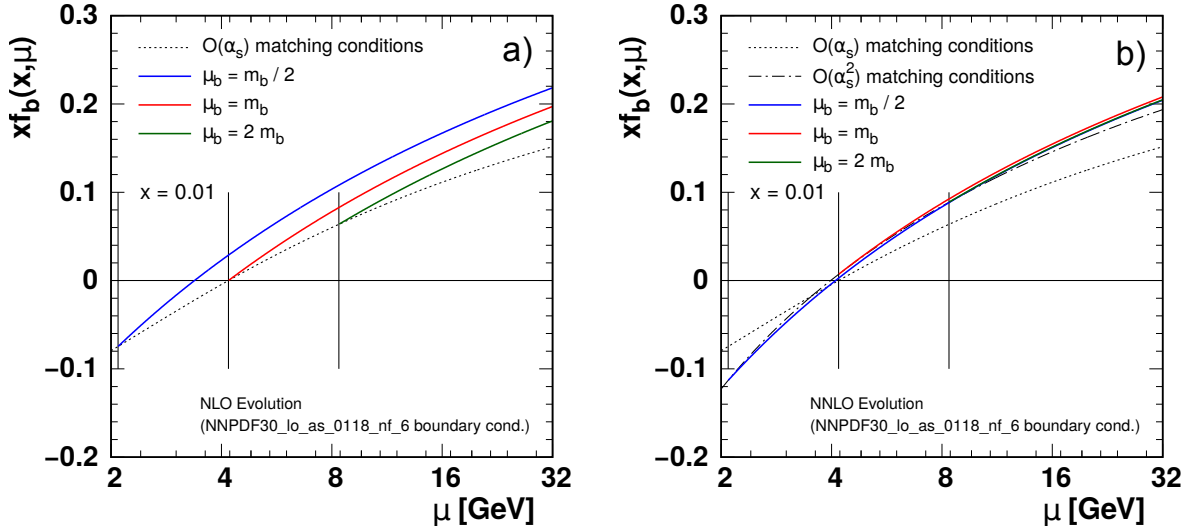
Figure 4.3: We display the b-quark PDF $x\, f_b^{(5)}(x, \mu)$ for different choices of the matching scales $\mu_m = \{m_b/2, m_b, 2m_b\}$ (indicated by the vertical lines) computed at NLO (Fig.-a) and NNLO (Fig.-b).

Fig. 4.3 displays the effect of different values of the bottom matching scale μ_b on the bottom-quark PDF for both the NLO and NNLO cases.[6] At NLO, the matching conditions are schematically:[7]

$$f_b^{(5)}(x, \mu_b) = \frac{\alpha_S^{(4)}(\mu_b)}{2\pi}\ \left[c_0^{bg} + c_1^{bg}\ L\right] \otimes f_g^{(4)}(x, \mu_b) \tag{4.6}$$

where $L = \ln(\mu_b^2/m_b^2)$. The superscripts $\{4, 5\}$ identify the number of active flavors N_F. The gluon and the light quarks also have matching conditions analogous to Eq. (4.6).

As already mentioned, if we choose to match at $\mu_b = m_b$ then $L = 0$ and $f_b^{(5)}(x, \mu_m)$ will start from zero at $\mu_b = m_b$. This coincidental zero ($c_0^{ij} = 0$) is the historic reason why

---

[6] A first study of the impact of moving the bottom matching scale with respect to the bottom mass was already done in Ref. [94] in the context of $\bar{b}bH$ production at the LHC using a matched scheme. The approach developed in this study was more recently applied to the 13 TeV LHC in Ref. [95].

[7] At NNLO, the bottom-quark matching condition also receives contributions from the light quarks as well as gluons; this has been included in the calculation.
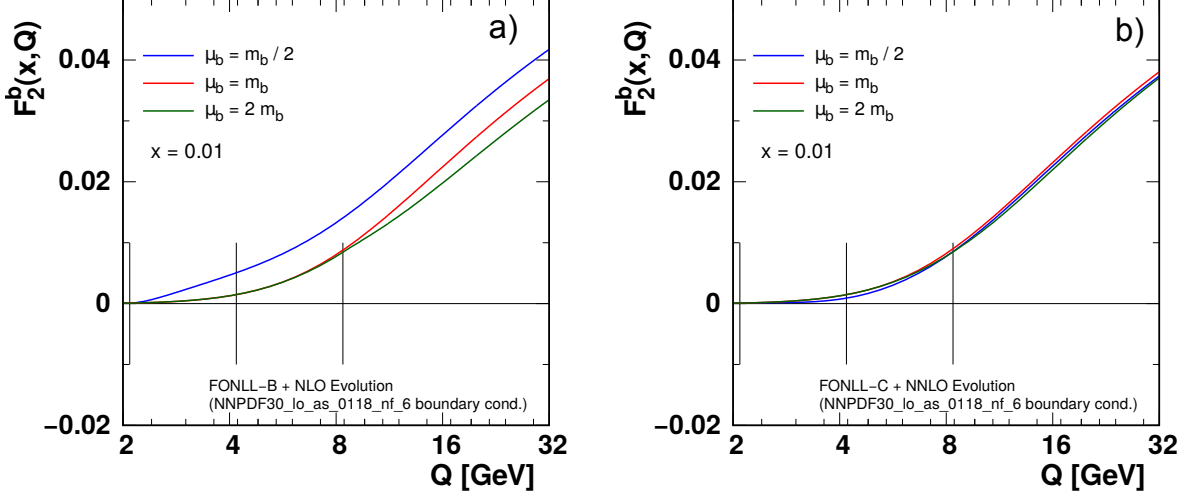
Figure 4.4: We display $F_2^b(x, Q)$ for different choices of the matching scales $\mu_m = \{m_b/2, m_b, 2m_b\}$ (indicated by the vertical lines) computed at NLO (Fig.-a) and NNLO (Fig.-b). Here, we have chosen $\mu = Q$. For details on the FONNL calculation see Ref. [75].

most NLO analyses perform the matching at $\mu_b = m_b$; if both the c_0^{ij} and $c_1^{ij} L$ terms can be ignored, then the PDFs are continuous (but not differentiable) across the matching scale.[8]

At NNLO this is no longer the case; the NNLO constant term at $\mathcal{O}(\alpha_S^2)$ does not vanish and the PDFs will have a discontinuity regardless of the choice of matching scale. Although the difference is subtle, the (red) curve for $\mu_b = m_b$ does start exactly from zero for the NLO calculation (Fig. 4.3-a), while for the NNLO calculation (Fig. 4.3-b) it starts from a small non-zero value.

As we vary the matching μ_b in the vicinity of m_b, the sign of $f_b^{(5)}(x, \mu_b)$ is controlled by the log term $(c_1^{ij} L)$. For $\mu_b < m_b$ this combination will drive $f_b^{(5)}(x, \mu_b)$ negative, and this will be compensated (in the sum rule for example) by a positive shift in the 5-flavor gluon. Thus, QCD ensures that both momentum and number sum rules are satisfied to the appropriate order.

Comparing different $f_b^{(5)}(x, \mu)$ curves computed with the NLO matching conditions (Fig. 4.3-a) at large μ scales, there are obvious differences in the curves. This reflects the difference between the single log contribution $(c_1^{ij} L)$ computed by the matching condition of Eq. (4.6)

---

[8] While the VFNS framework is compatible with an intrinsic charm or bottom PDF, we do not introduce these into the current study. For additional details, see Refs. [96–99].

and the resummed contributions computed by the DGLAP evolution equation. Specifically, the NLO matching includes the $\alpha_S L$ contribution, but is missing $\alpha_S^2 L^2$ and higher terms; this is what gives rise to the differences of Fig. 4.3-a. Obviously, the $\alpha_S^2 L^2$ contributions can be important.

Comparing the different $f_b^{(5)}(x, \mu)$ curves computed with the NNLO matching conditions (Fig. 4.3-b) at large μ scales, the differences in the curves are greatly reduced compared to the NLO case. The NNLO result includes both the $\alpha_S L$ and $\alpha_S^2 L^2$ contributions, but is missing $\alpha_S^3 L^3$ and higher orders. Clearly the inclusion of the $\alpha_S^2 L^2$ contributions dramatically reduces the effect of the different choices of the μ_m matching scale.

Finally, we wish to emphasize that ultimately the choice of μ_m amounts to a choice of scheme. In the limit that perturbation theory is computed to all orders, the infinite tower of logarithms resummed by the DGLAP evolution equations (in the $N_F + 1$-flavor scheme) will be explicitly summed in the matching conditions (in the N_F-flavor scheme). In a practical sense, while the differences at NLO are substantive, at NNLO the residual differences at large μ scale are much smaller. This reduced sensitivity on the choice of μ_m provides increased flexibility and precision in our fits, as will be illustrated in the following sections.

4.3.2. Impact of matching on $F_2^b(x, Q)$

Having examined the PDFs in the previous section we now turn to a physical observable, $F_2^b(x, Q)$.

Fig. 4.4-a) shows the NLO result for $F_2^b(x, Q)$ which will receive contributions from the LO process ($\gamma b \to b$) as well as the NLO ($\gamma g \to b\bar{b}$) process. For $\mu < \mu_b$, $f_b^{(5)}(x, \mu) = 0$ and only the gluon initiated process contributes. For $\mu \gtrsim \mu_b$, the bottom PDF turns on (*cf.* Fig. 4.3), and the heavy quark initiated process now contributes. Because the PDFs, $\alpha_S(\mu)$, and m_b are all carefully matched in this calculation, the cancellation outlined in Sect. 4.2.3 ensures that the prediction for the physical observable is relatively smooth in this region.

(a)

(b)
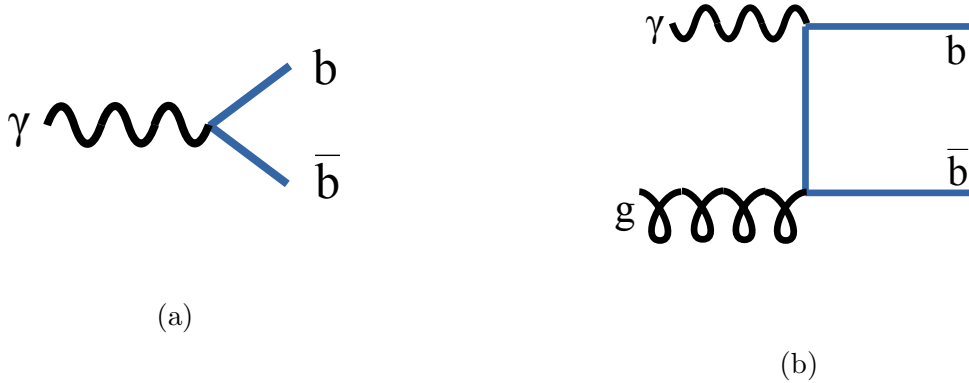
Figure 4.5: The LO process $\gamma b \to b$ (4.5a) and the NLO process $\gamma g \to b\bar{b}$(4.5b).

Fig. 4.4-b) shows the NNLO result for $F_2^b(x, Q)$. As with the PDF matching of Fig. 4.3-b), the additional NNLO contributions significantly reduce the impact of the different matching scales so that the prediction for $F_2^b(x, Q)$ is now very insensitive to μ_b.

The above smooth transition of $F_2^b(x, Q)$ from the $N_F = 4$ to the $N_F = 5$ scheme holds even though the PDFs and $\alpha_S(\mu)$ have discontinuities. Because we have used consistent choices for $\{m_b, f_i^{(N_F)}, \alpha_S\}$, the cancellation of Sect. 4.2.3 applies, and the effect of any discontinuities in the physical observable will be of higher order. Conversely, a mismatch in $\{m_b, f_i^{(N_F)}, \alpha_S\}$ would spoil this cancellation and result in unphysical large contributions when $f_b^{(5)}(x, \mu)$ is introduced. This is precisely the case where shifting the matching scale μ_b to a higher scale such as $2m_b$ or $3m_b$ would help avoid these problems.

It is interesting to note that as we compute even higher orders, the discontinuities in the PDFs and $\alpha_S(\mu)$ will persist at lower order; but, any discontinuities in the physical observables will systematically decrease order by order.

4.4. The PDF fits

4.4.1. xFitter, APFEL, and data sets

To study the effects of varying the matching scales for the charm and bottom quark we will perform a series of fits to various data sets. Since we are varying the matching scales in the vicinity of m_c and m_b, we want data that constrain the PDFs in this region. For this purpose, we include the very precise combined HERA data sets as these provide strong constraints in the region $\mu \sim m_{c,b}$, and also extend up to higher scales [28, 100–102]. In particular, the HERA measurement of the charm and bottom cross sections are included as they are sensitive to the choice of μ_c and μ_b.

These fits are performed with the xFitter program using the APFEL evolution code [16, 21, 103]. The DIS calculations use the FONLL-B scheme for the NLO calculations, and the FONLL-C scheme for the NNLO calculations; these are both $\mathcal{O}(\alpha_S^2)$ prescriptions, and the details are specified in Ref. [75]. We use $m_c = 1.45$ GeV, $m_b = 4.5$ GeV, $\alpha_S(M_Z) = 0.118$ for both the NLO and NNLO calculations. The fit is performed using pole masses, but the formalism can be used equally well with the $\overline{\text{MS}}$ definition of the heavy quark masses [104]. For the PDFs, we use a HERAPDF 14-parameter functional form with initial QCD evolution scale $Q_0^2 = 1.0$ GeV^2 and strangeness fraction $f_s = 0.4$; the other QCD fit settings and constraints are similar to the analysis of Ref. [104].

The minimization of the χ^2 is performed using MINUIT [105]. The correlations between data points caused by systematic uncertainties are taken into account in the "Correlated χ^{2}" contribution. A "Log penalty χ^{2}" arises from the likelihood transition to χ^2 when the scaling of the errors is applied [84, 106].

The full sets of data are listed in Tabs. 4.1-4.4, and the reference for each data set is cited in Tab. 4.1. The combined inclusive HERA data (HERA1+2) from Ref. [28] includes both neutral current (NC) and charged current (CC) results for electrons (em) and positrons (ep) at a variety of energies. The charm cross sections from Ref. [101] include the combined

H1-ZEUS results. The bottom cross sections from ZEUS are presented in Ref. [102] and those from H1 in Ref. [100].

4.4.2. Impact of matching on the fits: charm
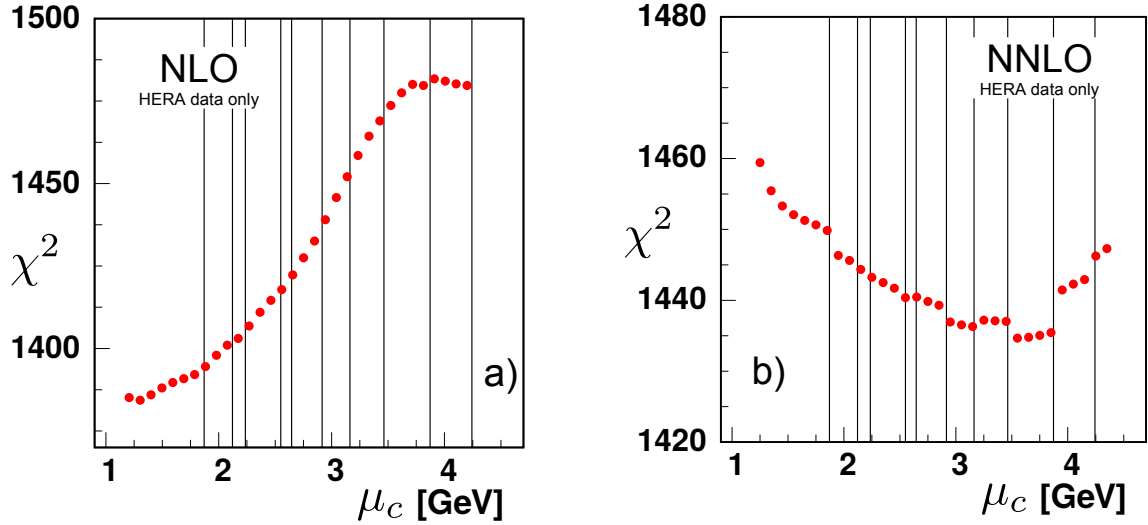
Figure 4.6: χ^2 vs. the charm matching scale μ_c at a) NLO and b) NNLO for all data sets. The bin boundaries for the HERA data set "HERA1+2 NCep 920" are indicated by the vertical lines.

The charm cross section data are expected to be sensitive to the treatment of the charm PDF in the threshold region, and this is reflected in the results of Figures 4.6, 4.7 and Tables 4.1, 4.2.

Fig. 4.6 displays the results for varying the charm quark matching scale μ_c both for the NLO and NNLO calculations.[9] Comparing the NLO and NNLO cases, the NLO result ranges over ~ 100 units in χ^2, while the NNLO varies over ~ 25 units of χ^2. This difference in the χ^2 variation reflects the effects of the higher order terms; it is reassuring to see that the μ_c dependence decreases at higher orders.

---

[9] For these scans we hold the bottom matching fixed at $\mu_b = m_b$ and keep $\mu_c < m_b$ so the ordering of the mass thresholds is not inverted.
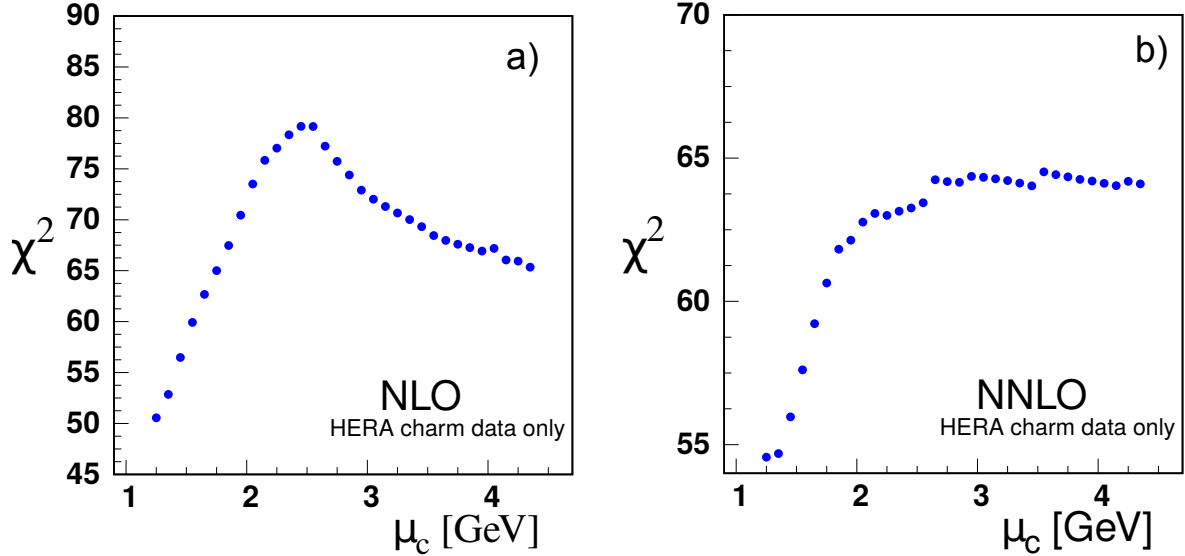
Figure 4.7: χ^2 vs. the charm matching scale μ_c at a) NLO and b) NNLO for only the H1-ZEUS combined charm production data; note, this includes the correlated χ^2 contribution from Tables 4.1 and 4.2.

At NLO, the matching conditions pick up the contribution of only the single log term L (Eq. (4.6)), while at NNLO we pick up both the L and L^2 terms. In contrast, the DGLAP evolved charm PDF resums the above, as well as an infinite tower of logs: $\sum_{n=1}^{\infty} \sum_{k=0}^{n} \alpha_S^n L^k$.

Examining the NLO analysis of Fig. 4.6-a, we find that at low scales, the χ^2 increases with increasing μ_c scale. While our plot extends slightly below the charm mass, it is not obvious if there is actually a minimum in μ_c. It is problematic to compute with μ_c values much lower than m_c as α_S becomes large and the charm PDF negative. Thus, the optimal computational range for μ_c appears to be in the region of m_c.

Focusing on the charm data alone as shown in Fig. 4.7-a, the situation is not so clear; the χ^2 increases with increasing μ_c, but again there does not appear to be a minimum at low μ_c values. Moving to large μ_c, the χ^2 values initially increase, but then decrease as μ_c approaches m_b. As we want to maintain the ordering $\mu_c < \mu_b$, we cannot go to larger scales unless we increase μ_b. While this is allowed, it is more complex to explore the two-dimensional $\{\mu_c, \mu_b\}$ parameter space; hence, we limit the present study to variation of a single scale.

111

Charm NLO	$\mu_c =$	$1\,m_c$	$2\,m_c$	$3\,m_c$
Charm cross section H1-ZEUS combined [101]		46 / 47	61 / 47	54 / 47
H1 F2 Beauty Vertex [100]		3.1 / 12	2.8 / 12	2.7 / 12
Beauty cross section ZEUS Vertex [102]		12 / 17	12 / 17	12 / 17
HERA1+2 CCep [28]		44 / 39	44 / 39	45 / 39
HERA1+2 CCem [28]		52 / 42	47 / 42	48 / 42
HERA1+2 NCem [28]		220 / 159	228 / 159	227 / 159
HERA1+2 NCep 820 [28]		65 / 70	70 / 70	68 / 70
HERA1+2 NCep 920 [28]		414 / 377	433 / 377	471 / 377
HERA1+2 NCep 460 [28]		221 / 204	217 / 204	225 / 204
HERA1+2 NCep 575 [28]		216 / 254	224 / 254	222 / 254
Correlated χ^2 total (charm)		86 (10.5)	91 (12.5)	105 (11.3)
Log penalty χ^2 total (charm)		+6.7 (+0.1)	-0.7 (-0.4)	-1.2 (-0.2)
Total χ^2 / dof		1386 / 1207	1430 / 1207	1479 / 1207

Table 4.1: The χ^2 values at NLO for individual data sets for a selection of the charm matching scales μ_c. The contribution of the charm data contained in the "Correlated χ^2" and in the "Log penalty χ^2" terms is indicated separately in the parentheses.

The χ^2 results for each individual data set is summarized in Tab. 4.1. The data sets with the largest effects are i) the H1-ZEUS combined charm cross section data, and ii) the very precise "HERA1+2NCep 920" set. The sensitivity of the "HERA1+2NCep 920" set is due to a large number of data points with small uncertainties.

Turning to the NNLO analysis of Fig. 4.6-b and the results of Tab. 4.2, a number of points are evident. Again, the two data sets with the largest impact are the H1-ZEUS combined charm cross section data, and the "HERA1+2NCep 920" set. In Fig. 4.6 the vertical lines indicate the bin boundaries for the "HERA1+2NCep 920" data set.

Scanning in χ^2, discrete jumps are evident. As we vary the matching scale, certain data bins move between the $N_F = 3$ and $N_F = 4$ schemes, shifting the χ^2 by one or two units

Charm NNLO	$\mu_c =$	$1\,m_c$	$2\,m_c$	$3\,m_c$
Charm cross section H1-ZEUS combined		45 / 47	50 / 47	50 / 47
H1 F2 Beauty Vertex		3.5 / 12	3.5 / 12	3.3 / 12
Beauty cross section ZEUS Vertex		13 / 17	13 / 17	13 / 17
HERA1+2 CCep		43 / 39	43 / 39	43 / 39
HERA1+2 CCem		55 / 42	55 / 42	54 / 42
HERA1+2 NCem		217 / 159	217 / 159	217 / 159
HERA1+2 NCep 820		66 / 70	64 / 70	66 / 70
HERA1+2 NCep 920		444 / 377	433 / 377	442 / 377
HERA1+2 NCep 460		218 / 204	219 / 204	216 / 204
HERA1+2 NCep 575		220 / 254	218 / 254	219 / 254
Correlated χ^2 total (charm)		111 (10.8)	109 (11.3)	110 (14.5)
Log penalty χ^2 total (charm)		+18 (-1.1)	+18 (-1.8)	+15 (-1.8)
Total χ^2 / dof		1453 / 1207	1439 / 1207	1447 / 1207

Table 4.2: The χ^2 values at NNLO for individual data sets for a selection of the charm matching scales μ_c. The contribution of the charm data contained in the "Correlated χ^2" and in the "Log penalty χ^2" terms is indicated separately in the parentheses.

which is visible in Fig. 4.6-b). These jumps reflect the underlying theoretical uncertainty arising from the choice of N_F.

In Fig. 4.6-b the total NNLO variation of χ^2 is reduced compared to the NLO case, and the minimum global χ^2 is now in the region $\mu_c \sim 2m_c$. Focusing on the charm data alone in Fig. 4.7-b, again it is not obvious if there is actually a minimum in μ_c. Given the limitations of computing with $\mu_c \ll m_c$, the optimal computational range again appears to be in the general region of m_c.

While it may be tempting to try and optimize the matching scale for each data set, recall that μ_m represents a choice of scheme, and thus reflects an inherent theoretical uncertainty; a specific choice of μ_m will not reduce this uncertainty.

This situation can also be found in complex global fits where the final result may be a compromise of data sets which are in tension; this is why a tolerance factor is often introduced. This complexity is evident when examining the details of Tables 4.1 and 4.2 which demonstrate the minimum χ^2 for individual data sets is not simply correlated; this will be discussed further in Section 4.4.4. An additional challenge of analyzing the charm case is that μ_c can only vary over the limited dynamic range between $\sim m_c$ and μ_b. This will not be an issue for the bottom quark (because $m_t \gg m_b$), which is considered in the following section.

4.4.3. Impact of matching on the fits: bottom
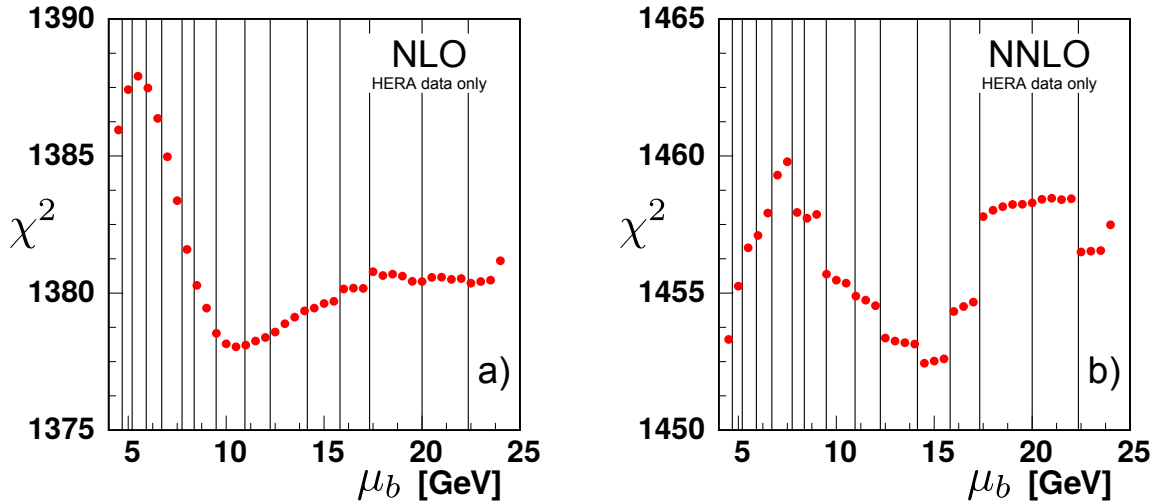
Figure 4.8: χ^2 vs. the bottom matching scale μ_b at a) NLO and b) NNLO for all data sets. The bin boundaries for the HERA data set "HERA1+2 NCep 920" are indicated by the vertical lines.

The results for varying the bottom quark matching scale μ_b both for the NLO and NNLO calculations, is presented in Fig. 4.8. This figure highlights the ranges of χ^2; the NLO result ranges over approximately ~ 10 units in χ^2, and the NNLO varies by about the same amount.

The reduced χ^2 variation as compared to the charm case reflects, in part, the decrease in the strong coupling $\alpha_S(m_b) < \alpha_S(m_c)$ which also diminishes the higher order contributions.
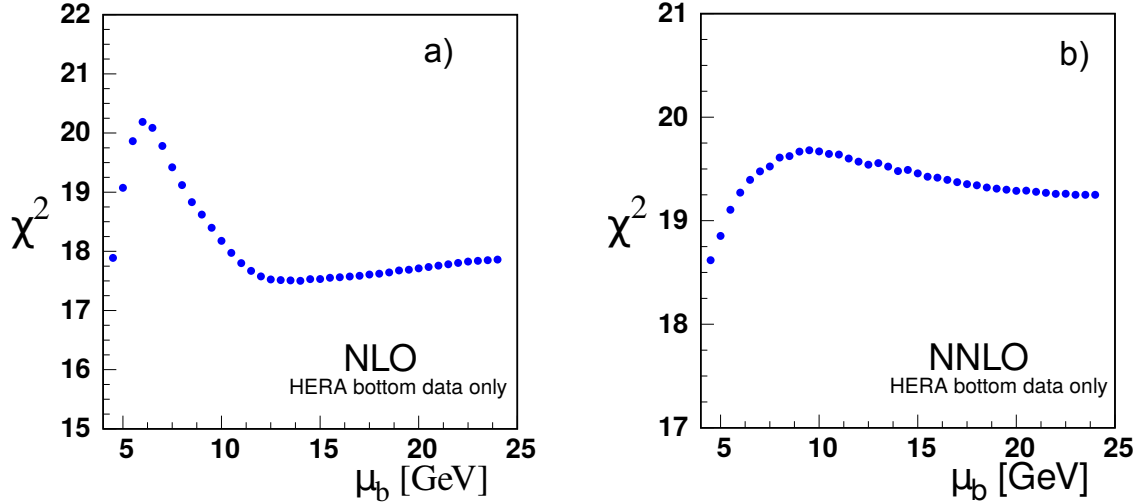
Figure 4.9: χ^2 vs. the bottom matching scale μ_b at a) NLO and b) NNLO for only the bottom data; note, this includes the H1 and ZEUS beauty data as well as the correlated χ^2 contribution from Tables 4.3 and 4.4.

Fig. 4.6 with Fig. 4.8 there is a χ^2 range of ~ 100 vs. ~ 10 for NLO, and ~ 15 vs. ~ 10 for NNLO.

Examining the NLO analysis of Fig. 4.8-a, there is a slight minimum for χ^2 in the region $\mu_b \sim 2m_b$ with relatively flat behavior at larger μ_b scales. Correspondingly, there is a similar behavior when we focus on only the bottom data of Fig. 4.9-a. The χ^2 results for each individual data set are summarized in Tab. 4.3.

The data sets with the largest effects are i) the very precise "HERA1+2NCep 920" set, and ii) the separate H1 and ZEUS bottom cross section data. The H1 and ZEUS bottom cross sections display some minimal χ^2 variation in the region $\mu_b \sim m_b$, but then is relatively flat out to very high scales ($\mu_b \sim 14m_b$). It is primarily the "HERA1+2NCep 920" set which drives the shape of the χ^2 curve in the $\mu_b \sim m_b$ region. Compared to the charm results, the interpretation of the bottom cross section data requires some care as the number of data points is smaller, and the relative uncertainty larger.

Turning to the NNLO analysis of Fig. 4.8-b, the variation of the χ^2 curve is within ~ 8 units across the range of the plot. The resolution of the vertical χ^2 scale accentuates the

Bottom NLO	$\mu_b =$ $1\,m_b$	$3\,m_b$	$5\,m_b$	$10\,m_b$	$14\,m_b$
Charm cross section H1-ZEUS combined	46 / 47	46 / 47	46 / 47	46 / 47	46 / 47
H1 F2 Beauty Vertex	3.1 / 12	3.2 / 12	3.1 / 12	3.2 / 12	3.2 / 12
Beauty cross section ZEUS Vertex	12 / 17	12 / 17	12 / 17	12 / 17	14 / 17
HERA1+2 CCep	44 / 39	44 / 39	44 / 39	44 / 39	44 / 39
HERA1+2 CCem	52 / 42	52 / 42	52 / 42	53 / 42	53 / 42
HERA1+2 NCem	220 / 159	219 / 159	220 / 159	219 / 159	219 / 159
HERA1+2 NCep 820	65 / 70	65 / 70	65 / 70	65 / 70	65 / 70
HERA1+2 NCep 920	414 / 377	410 / 377	410 / 377	412 / 377	412 / 377
HERA1+2 NCep 460	221 / 204	221 / 204	221 / 204	219 / 204	220 / 204
HERA1+2 NCep 575	216 / 254	216 / 254	216 / 254	216 / 254	216 / 254
Correlated χ^2 total (bottom)	86 (0.8)	86 (0.8)	86 (0.8)	87 (0.8)	89 (0.8)
Log penalty χ^2 total (bottom)	+6.7 (-0.1)	+4.2 (-0.1)	+4.5 (-0.1)	+6.6 (-0.1)	+7.3 (-0.1)
Total χ^2 / dof	1386 / 1207	1379 / 1207	1380 / 1207	1383 / 1207	1388 / 1207

Table 4.3: The χ^2 values at NLO for individual data sets for a selection of the bottom matching scales μ_b. The contribution of the bottom data contained in the "Correlated χ^2" and in the "Log penalty χ^2" terms is indicated separately in the parentheses.

discrete jumps as the data bins move between the $N_F = 4$ and $N_F = 5$ schemes. The bin boundaries for the "HERA1+2NCep 920" data set are indicated with vertical lines.

Focusing on the bottom data alone as shown in Fig. 4.9-b, the χ^2 profile is flat within one unit across the plot range.

For both Fig. 4.8-b and Fig. 4.9-b, the χ^2 variation is within a reasonable "tolerance" factor for the global fit; thus, the matching scale μ_b can vary within this range with minimal impact on the resulting fit.

The scale μ_b can extend up to larger scales, and Tabs. 4.3 and 4.4 display the results for $10m_b$ and $14m_b$. The pattern across the various data sets is consistent, and the overall χ^2 values rise slowly.

Bottom NNLO	$\mu_b =$	$1\,m_b$	$3\,m_b$	$5\,m_b$	$10\,m_b$	$14\,m_b$
Charm cross section H1-ZEUS combined		45 / 47	45 / 47	45 / 47	45 / 47	45 / 47
H1 F2 Beauty Vertex		3.5 / 12	3.7 / 12	3.7 / 12	3.6 / 12	3.6 / 12
Beauty cross section ZEUS Vertex		13 / 17	13 / 17	13 / 17	13 / 17	14 / 17
HERA1+2 CCep		43 / 39	43 / 39	43 / 39	42 / 39	42 / 39
HERA1+2 CCem		55 / 42	55 / 42	55 / 42	55 / 42	56 / 42
HERA1+2 NCem		217 / 159	216 / 159	220 / 159	218 / 159	218 / 159
HERA1+2 NCep 820		66 / 70	66 / 70	66 / 70	66 / 70	66 / 70
HERA1+2 NCep 920		444 / 377	445 / 377	445 / 377	451 / 377	453 / 377
HERA1+2 NCep 460		218 / 204	219 / 204	219 / 204	217 / 204	218 / 204
HERA1+2 NCep 575		220 / 254	219 / 254	219 / 254	219 / 254	219 / 254
Correlated χ^2 total (bottom)		111 (0.9)	112 (0.9)	112 (0.9)	114 (0.9)	116 (0.9)
Log penalty χ^2		+18	+17	+15	+18	+18
Total χ^2 / dof		1453 / 1207	1453 / 1207	1457 / 1207	1463 / 1207	1470 / 1207

Table 4.4: The χ^2 values at NNLO for individual data sets for a selection of the bottom matching scales μ_b. The contribution of the bottom data contained in the "Correlated χ^2" and in the "Log penalty χ^2" terms is indicated separately in the parentheses.

4.4.4. Comparisons

To facilitate comparisons of the NLO and NNLO results, Fig. 4.10 displays the ratio χ^2/χ_0^2 for charm (on the left) and bottom (on the right) where χ_0^2 is the value of the χ^2 at $\mu_m = m_H$. Similarly, Fig. 4.11 displays the same ratio for only the heavy quark data sets. By plotting χ^2/χ_0^2, we can better compare the fractional variation of χ^2 across the matching scale values.

The motivation for the scaled plot of Figs. 4.10 and 4.11 is that the overall χ^2 values are different; specifically, those of the NNLO are greater than the NLO. This counter intuitive result has been observed in other analyses [27,28], and it has been suggested that this may be improved by resumming the singular $\ln[1/x]$ terms in the higher order splitting kernels [107].

Here, we first make some observations specific to Figures 4.10 and 4.11:

- At NLO for the case of charm, the optimal computational scale for μ_c is in the general range $\mu_c \sim m_c$ for both the inclusive data set (Fig. 4.10-a) and the charm data set

117
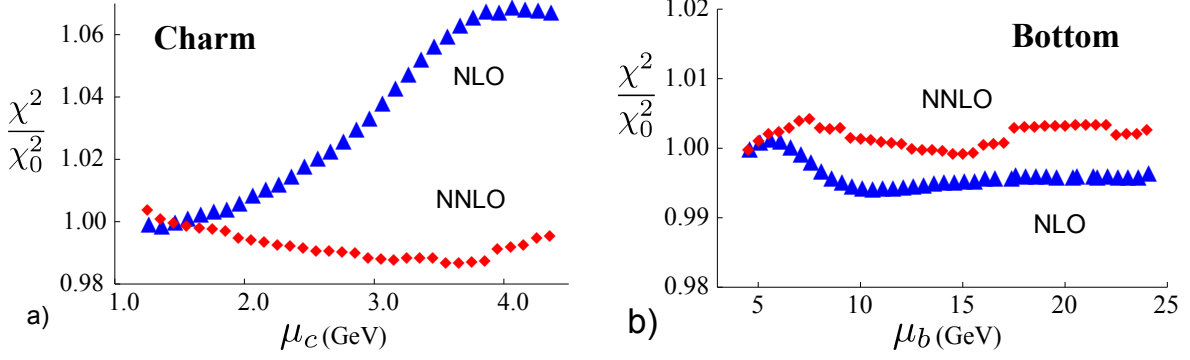
Figure 4.10: The ratio (χ^2/χ_0^2) of total χ^2 values (all data sets combined) from Figs. 4.6 and 4.8, as a function of the a) charm and b) bottom matching scale $\mu_{c,b}$ in GeV. χ_0^2 is the χ^2 value for μ_m equal to the quark mass. The triangles (blue ▲) are NLO and the diamonds (red ♦) are NNLO.
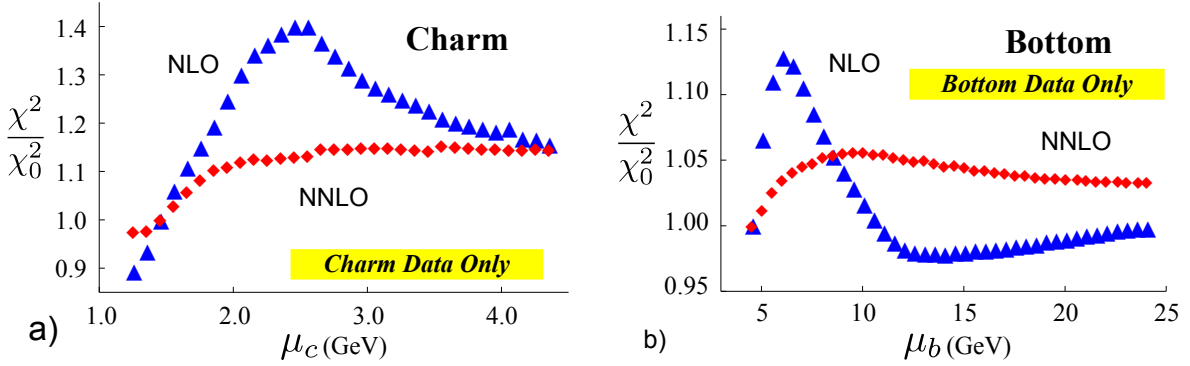
Figure 4.11: The ratio (χ^2/χ_0^2) of partial χ^2 values (charm/bottom data only) from Figs. 4.7 and 4.9 as a function of the a) charm and b) bottom matching scale $\mu_{c,b}$ in GeV. χ_0^2 is the χ^2 value for μ_m equal to the quark mass. The triangles (blue ▲) are NLO and the diamonds (red ♦) are NNLO.

(Fig. 4.11-a). For lower scales $(\mu_c \ll m_c)$, $\alpha_S(\mu)$ is large and the charm PDFs are negative. For higher scales $(\mu_c \gg m_c)$, χ^2/χ_0^2 increases.

- At NLO for the case of bottom, the optimal scale for μ_b is in the general range $\mu_b \sim 2m_b$. For the inclusive data set (Fig. 4.10-b) the χ^2/χ_0^2 variation is very mild ($\sim 1\%$), while for the bottom data set (Fig. 4.11-b) the χ^2/χ_0^2 variation is larger ($\sim 10\%$).

- At NNLO for the case of charm, the χ^2/χ_0^2 variation is reduced. For the inclusive data set (Fig. 4.10-a) the χ^2/χ_0^2 variation is very mild ($\sim 2\%$), while for the charm data set

(Fig. 4.11-a) the χ^2/χ_0^2 variation is larger ($\sim 10\%$). There is no obvious optimal choice for the μ_c scale.

- At NNLO for the case of bottom, the χ^2/χ_0^2 variation is reduced and a matching scale choice in the region $\mu_b \sim m_b$ appears to be optimal. For the inclusive data set (Fig. 4.10-b) the χ^2/χ_0^2 variation is very mild ($\sim 1\%$), while for the bottom data set (Fig. 4.11-b) the χ^2/χ_0^2 variation is slightly larger ($\sim 5\%$).

While the detailed characteristics of the above fits will depend on specifics of the analysis, there are two general patterns which emerge: i) the χ^2 variation of the NNLO results are generally reduced compared to the NLO results, and ii) the relative χ^2 variation across the bottom transition is reduced compared to the charm transition. For example, although the global χ^2 can be modified by different choices of data sets and weight factors, these general properties persist for each individual data set of Tables 4.1–4.4; in fact, we see that the bulk of the data sets are quite insensitive to the details of the heavy quark matching scale. Additionally, there are a variety of prescriptions for computing the heavy flavor contributions; these primarily differ in how the higher order contributions are organized. As a cross check, we performed a NLO fit using the FONNL-A scheme; while the absolute value of χ^2 differed, the above general properties persisted.

The net result is that we can now quantify the theoretical uncertainty associated with the transition between different N_F sub-schemes. In practical applications, if we choose $\mu_c \sim m_c$, the impact of the $N_F = 3$ to $N_F = 4$ transition is reduced as this is often below the minimum kinematic cuts of the analysis (e.g. Q_{min}^2 and W_{min}^2). Conversely, the $N_F = 4$ to $N_F = 5$ transition is more likely to fall in the region of fitted data; hence, it is useful to quantify the uncertainty associated with the μ_b choice.

4.5. An example: N_F-dependent PDFs

The variable matching scale μ_m can be used as an incisive tool to explore various aspects of the PDFs and global fits. As an example, Ref. [20] introduced an N_F-dependent PDF

$f_i(x, \mu, N_F)$ where N_F is the active number of flavors in the VFNS. This extension provides additional flexibility in the region of the heavy quark thresholds; however, the implementation of Ref. [20] only used a fixed matching scale of $\mu_m = m_H$. Using xFitter we can improve on this concept by generating PDFs with a variable μ_m scale. We illustrate this below and provide example grids at xFitter.org.
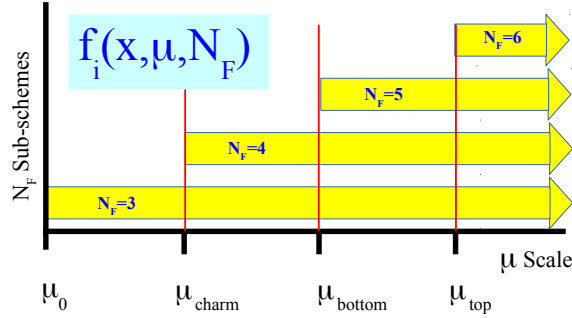
Figure 4.12: An illustration of the separate N_F renormalization sub-schemes which define the VFNS. In contrast to Fig. 4.1-a), each of the N_F sub-schemes are available for all scales above μ_m. The particular scheme can be specified by choosing N_F when calling the PDF, i.e. $f_i(x, \mu, N_F)$. This illustration shows a matching scale of $\mu_m = m_H$.
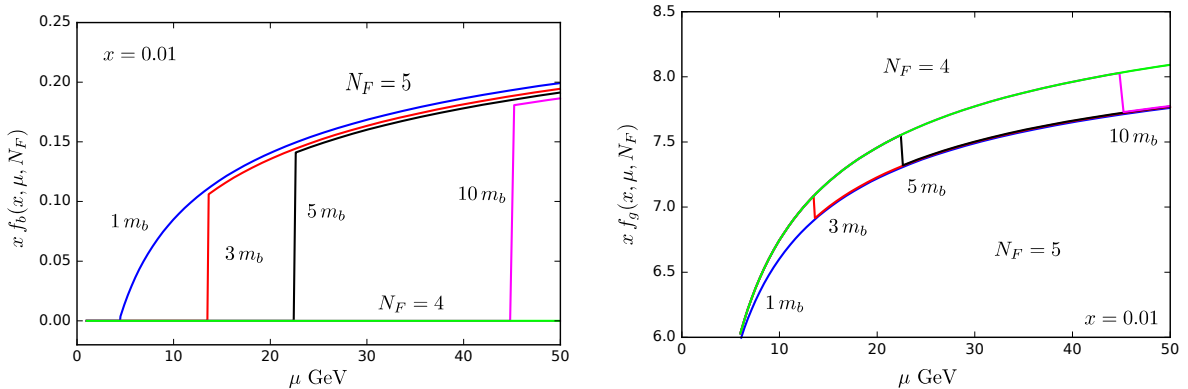
Figure 4.13: N_F-dependent PDFs $x f_i(x, \mu, N_F)$ for the bottom quark (left) and gluon (right) with variable matching scales for $\mu_b = \{1, 3, 5, 10, \infty\} \times m_b$ {blue, red, black, magenta, green} with $x = 0.01$ as a function of μ in GeV. The vertical lines in the plots show the transition from the $N_F = 4$ to $N_F = 5$ flavor scheme.

The usual PDF can be generalized to include an N_F-dependence [20]: $f_i(x, \mu) \to f_i(x, \mu, N_F)$. In this approach, the many $N_F = \{3, 4, 5, ...\}$ flavor schemes coexist, and they can be selected by specifying the number of active flavors N_F along with the other arguments of the PDF.

This concept is represented pictorially in Fig. 4.12. All the N_F sets of PDFs are available above the matching scale μ_m. For example, with an N_F-dependent PDF, one could simultaneously analyze selected data sets with $N_F = 4$ and others with $N_F = 5$ even if they overlap kinematically; the user has the flexibility (and responsibility) to select N_F.

Note in Fig. 4.12 that the various N_F grids are not individual fits but are related analytically via the flavor threshold matching conditions. Operationally, they are generated from an initial PDF $f_i(x, \mu_0, N_F = 3)$ and $\alpha_S(\mu_0)$ at the starting scale μ_0. The $N_F = 3$ grid is generated by evolving from μ_0 to μ_{max}. The $N_F = 4$ grid is then generated by matching at μ_c (which may or may not equal m_c), and evolving up to scale μ_{max}. The $N_F = 5$ and $N_F = 6$ grids are generated in a similar manner.[10] This process ensures that all the PDFs $f_i(x, \mu, N_F)$ are analytically related to the PDF and α_S boundary conditions at μ_0.

To provide an explicit illustration of the above, we have generated a set of PDF grids with a variety of matching scales (μ_b) for the matching between the schemes with $N_F = 4$ and $N_F = 5$ active flavours: $\mu_b = \{1, 3, 5, 10, \infty\} \times m_b$. We focus on μ_b as this is the flavor transition most likely to fall within a particular data set. For the initial PDF we use the NNLO bottom fit with $\mu_b = 1\, m_b$ of Table 4.4, and we evolve at NNLO. The PDFs are fixed such that they all match at the initial evolution scale $\mu_0 = 1.0$ GeV with the same value of $\alpha_S(\mu_0) = 0.467464$.

This is illustrated in Fig. 4.13 where we display the bottom quark and gluon PDFs as a function of μ in GeV. As we evolve up in μ, we explicitly see the transition from $N_F = 4$ to $N_F = 5$ flavors at each respective μ_b threshold. For these particular kinematic values, the discontinuity of the bottom PDF is positive while that of the gluon is negative; this ensures the momentum sum rule is satisfied. Furthermore, we observe the spread in the bottom PDF at large μ is broader than that of Fig. 4.3. In Fig. 4.13, while the values of α_S all coincide at μ_0, the evolution across the different μ_b thresholds result in different α_S values at large μ scales. This is in contrast to Fig. 4.3 where the values of α_S all coincide at the large scale

---

[10]Note the $N_F = \{3, 4, 5, 6\}$ grids are stored in separate LHAPDF data files; they can be combined into an effective N_F dependent PDF as illustrated in Refs. [20, 108].

$\mu = M_Z$. Additionally, note that the illustration in Fig. 4.3 is based on the NNPDF3.0 PDF set while Fig. 4.13 is based on our fit from Table 4.4.

Because the $N_F = 4$ and $N_F = 5$ grids are available concurrently, we can choose to analyze the HERA data in an $N_F = 4$ flavor scheme for arbitrarily large scales, but simultaneously allow LHC data to be analyzed in a $N_F = 5$ flavor scheme throughout the full kinematic region even down to low scales.

In this illustration, the PDFs revert to $N_F = 4$ below μ_b; however, this is not required. For example the $N_F = 5$ PDFs could be evolved backwards from μ_b to provide values at scales $\mu < \mu_b$. Both APFEL [21] and QCDNUM [15,109] have this capability.[11]

For bottom at NNLO using the results from Tab. 4.4 for the inclusive data set, we observe the μ_b variation is minimal. Thus, a choice in the range $\mu_b \sim [m_b, 5m_b]$ yields a $\Delta\chi^2 \leq (1457 - 1453) \sim 4$ units out of ~ 1450. This minimal χ^2 dependence means we can shift the μ_b matching scale if, for example, we want to avoid a N_F flavor transition in a specific kinematic region. While these results should be checked with additional data sets, the insensitivity to μ_b, especially at NNLO, is an important result as the ability to displace the $N_F = 4$ and $N_F = 5$ transition can be beneficial when this threshold comes in the middle of a data set.

Combined with the variable heavy quark threshold, the N_F dependent PDFs provide additional flexibility to analyze multiple data sets in the optimal theoretical context.

4.6. Conclusions

In this study we have examined the impact of the heavy flavor matching scales μ_m on a PDF fit to the combined HERA data set.

The choice of μ_m allows us to avoid delicate cancellations in the region $\mu_m \sim m_H$ as illustrated in Fig. 4.2. Additionally, the discontinuities associated with the $N_F = 4$ to

---

[11]However, it is generally advisable not to backwards evolve too far in μ as this can become unstable [22, 110].

$N_F = 5$ transition can be shifted so that these discontinuities do not lie in the middle of a specific data set.

Using `xFitter` and `APFEL` to study the μ_m dependence of a global PDF fit to the HERA data, we can extract the following general features. For the charm matching scale, μ_c, there is a large variation of χ^2 at NLO, but this is significantly reduced at NNLO. In contrast, for the bottom matching scale, μ_b, there is a relatively small variation of χ^2 at both NLO and NNLO.

These observations can be useful when performing fits. While charm has a larger χ^2 variation (especially at NLO), the charm quark mass $m_c \sim 1.45$ GeV lies in a region which is generally excluded by cuts in Q^2 and/or W^2.

On the contrary, the χ^2 variation for the bottom quark is relatively small at both NLO and NNLO. Since the bottom quark mass $m_b \sim 4.5$ GeV is in a region where there is abundance of precision HERA data, this flexibility allows us to shift the heavy flavor threshold (and the requisite discontinuities) away from any particular data set. Functionally, this means that we can analyze the HERA data using an $N_F = 4$ flavor scheme up to relatively large μ scales, and then perform the appropriate NNLO matching (with the associated constants and log terms) so that we can analyze the high-scale LHC data in the $N_F = 5$ or even $N_F = 6$ scheme.

These variable heavy flavor matching scales μ_m allow us to generalize the transition between a FFNS and a VFNS, and provides a theoretical "laboratory" which can quantitatively test proposed implementations. We demonstrated this with the example of the N_F-dependent PDFs. Having the quantitative results for the χ^2 variation of the $\mu_{c,b}$ scales, one could systematically evaluate the impact of using different matching scale choices for the $f_i(x, \mu, N_F)$.

In conclusion, we find that the ability to vary the heavy flavor matching scales μ_m, not only provides new insights into the intricacies of QCD, but also has practical advantages for PDF fits.

Chapter 5

nCTEQ15+LHC

Armed with an understanding of the underlying mechanics of PDF fitting from Chapter 4 and the results from the reweighting study from Chapter 3, we now begin the process of refitting the `nCTEQ15` PDF set to include W/Z boson pPb data from Run 1 of the Large Hadron Collider (LHC). This chapter will first cover the `nCTEQ` fitting framework and the existing `nCTEQ15` fit [7]. From there, I introduce `nCTEQ++`, a new version of the `nCTEQ` fitting code updated from `FORTRAN` to C++. This new code contains several improvements over the existing code but represents a fundamental shift in the fitting process. As a result, I present several validations performed to compare `nCTEQ++` to the now defunct `FORTRAN` code.

One feature of `nCTEQ++` is the ability to add new data sets into a fit with considerably less effort than before. In prior iterations, one would need to write standalone theory prediction modules for any new process you wished to include in the fit[1]. This was difficult, time consuming and inevitably discouraged the inclusion of new data. Building on the success of using `FEWZ` in the reweighting procedure, I developed a tool chain to connect data from LHC experiments to theory predictions usable in `nCTEQ++`. Furthermore, this chain is usable for processes beyond W/Z production at the LHC, making it far easier and faster to include data in future fits. More on this, including validations for each of the links, can be found in Section 5.3.

Finally, in Section 5.4, I present the results of a new fit including LHC data. Using `nCTEQ++` and the tool chain mentioned above, it was possible to perform a set of fits with

---

[1]This was most recently done to include pion production data from RHIC in `nCTEQ15`. However, for simplicity, fits performed in this chapter will not include this data opting instead to use `nCTEQ15-np`, a PDF set without inclusive pion data, as the baseline PDF. More information on this distinction can be found in [7].

different combinations of open parameters and compare them to both `nCTEQ15` and the results from Chapter 3.

5.1. Existing `nCTEQ15`

Before diving into a new `nCTEQ` fit, it is first important to discuss thoroughly the framework behind `nCTEQ15` as it will provide the foundation on which the new fit is constructed. As such, this section will rely heavily of the work done in [7] to produce the `nCTEQ15` fit. Many of the definitions, formulas and details of that fit will be reproduced here as well as a detailed accounting of the data they used. This is necessary because the new fit utilizes much of the same machinery and includes the data used in `nCTEQ15`.

In the first section, I will explain the structure underlying the `nCTEQ15` fit. This framework remains largely unchanged despite the implementation of `nCTEQ++` and the addition of LHC data in the new fit. I then provide the data present in `nCTEQ15` in Section 5.1.2. And, finally, I provide the results from the `nCTEQ15` analysis.

5.1.1. `nCTEQ15` Framework

As discussed in Section 1.4, the `nCTEQ` collaboration implements their nuclear corrections in the form of a bound proton PDF, $f_i^{p/A}(x, Q)$. This bound proton PDF is a parameterized fit to experimental data and represents the effective bound proton within a nucleus containing A nucleons[2]. Using isospin, a bound neutron PDF can be similarly determined and a full nucleus consisting of A protons and $A - Z$ neutrons can be constructed following Eq. 5.1.

$$f_i^{(A,Z)}(x, Q) = \frac{Z}{A} f_i^{p/A}(x, Q) + \frac{A - Z}{A} f_i^{n/A}(x, Q),$$ (5.1)

This full nuclear PDF is then useful when making theoretical calculations such as DIS structure functions (Eq. 5.2) or Drell-Yan cross sections (Eq. 5.3).

---

[2]This PDF is simply a constructed entity designed to describe an average nucleon while the constructed full nuclear PDF, $f_i^{A,Z}(x, Q)$, is what is actually used in calculations of experimental observables.

$$F_2^A(x, Q^2) \;=\; \sum_i f_i^{(A,Z)}(x, Q^2) \otimes C_{2,i}(x, Q^2) \tag{5.2}$$

$$d\sigma_{AB \to l\bar{l}X} \;=\; \sum_{ij} f_i^{(A_1,Z_1)} \otimes f_j^{(A_2,Z_2)} \otimes d\hat{\sigma}^{ij \to l\bar{l}X} \tag{5.3}$$

When calculating the DIS structure functions, the ACOT variable flavor number scheme [71, 72, 88, 111] at next-to-leading order (NLO) in QCD is used. This scheme only includes dominant target mass effects in the structure functions; however, a fuller treatment of these corrections is unnecessary due to the kinematic cuts enforced on the data present in the fit[3].

For theory calculations throughout the analysis, the renormalization and factorization scales have been set to be equivalent: $\mu = \mu_R = \mu_F$. For deep inelastic scattering (DIS), this equates to $\mu^2 = Q^2$ and represents the virtual mass of the exchanged vector boson between the leptonic probe and the quark. In Drell-Yan processes (DY), the scale is set to the invariant mass of the leptonic daughters of the produced vector boson, i.e. $\mu^2 = M^2$. It is also important to note that K-factor techniques [32] were implemented in order to accelerate the fitting process, specifically the evaluation of the NLO cross section, but for nCTEQ15 final fits were done with full NLO calculations.

5.1.1.1. Parameterization

$$f_i(x, Q_0) = x^{-\alpha}(1-x)^\beta P_i(x)$$

In Section 1.3.2, Eq. 1.9, reproduced above, described a simple parameterization that underlies the determination of any nCTEQ PDF fit [6, 56, 112]. This parameterization, expanded

---

[3]Target mass corrections are prevalent mostly at large x and low Q^2. A full description of the kinematic cuts implemented in nCTEQ15 can be found in Section 5.1.2

below as Eq. 5.4, follows from the parameterization used in prior CTEQ fits [35,113,114] for the free proton with the exception that this parameterization introduces a nuclear dependence[4].

$$xf_i^{p/A}(x, Q_0) = c_0\, x^{c_1}(1-x)^{c_2} e^{c_3 x}(1 + e^{c_4}x)^{c_5}$$

$$\text{for} \quad i = u_v, d_v, g, \bar{u} + \bar{d}, s + \bar{s}, s - \bar{s}$$

$$\frac{\bar{d}(x, Q_0)}{\bar{u}(x, Q_0)} = e^{c_1} x^{c_2 - 1}(1 - x)^{c_3} + (1 + c_4 x)(1 - x)^{c_5} \tag{5.4}$$

The nuclear dependence is introduced by modifying the c_k parameters as seen in Eq. 5.5. This modification guarantees that for $A = 1$, the parameterization will return the free proton PDF, as expected. The parameterization is fit at the input scale Q_0, with $Q_0 = 1.3$ GeV to match the style of the CTEQ free proton fits [113, 114]. Here the free proton would be determined by the p_k coefficients.

$$c_k \to c_k(A) \equiv p_k + a_k \left(1 - A^{-b_k}\right),$$

$$k = \{1, \ldots, 5\}. \tag{5.5}$$

In the original nCTEQ15 fit, there was insufficient data to constrain the strange quark in the parameterization. To correct for this, the strange distribution was fixed to be a fraction, κ, of the sum of the $\bar{u}$ and $\bar{d}$ distributions, as shown in Eq. 5.6. The A-dependent term, $\kappa(A)$, represents a normalization factor parameterized much the same way as the $c_k(A)$ parameters in Eq. 5.5. However, with the inclusion of W/Z production data from the LHC, it is possible to fit the strange quark distributions along with the other flavors[5].

---

[4]This parameterization represents the seven flavors fit in nCTEQ analyses: $u_v, d_v, g, \bar{u} + \bar{d}, \frac{\bar{d}}{\bar{u}}, s + \bar{s}, s - \bar{s}$. These represent the u and d valence quark distributions, the gluon distribution, the sum of the u and d sea quark distributions, the total strange distribution and the distribution of strange asymmetry, respectively. These flavors can be translated into the familiar quark flavors:
$s = \frac{(s+\bar{s})+(s-\bar{s})}{2}, \quad \bar{s} = \frac{(s+\bar{s})-(s-\bar{s})}{2}, \quad \bar{u} = \frac{\bar{u}+\bar{d}}{1+\frac{\bar{d}}{\bar{u}}}, \quad \bar{d} = (\frac{\bar{d}}{\bar{u}})(\frac{\bar{u}+\bar{d}}{1+\frac{\bar{d}}{\bar{u}}}), \quad u = u_v + \bar{u}, \quad d = d_v + \bar{d}.$

[5]The $s - \bar{s}$ flavor, representing the strange asymmetry, remains fixed to 0 due to lack of data in both nCTEQ15 and all new fits.

$$s^{p/A}(x, Q_0) = \bar{s}^{p/A}(x, Q_0) = \frac{\kappa(A)}{2}\left(\bar{u}^{p/A} + \bar{d}^{p/A}\right) \tag{5.6}$$

The c_0 normalization coefficients[6] are constrained by the sum rules described in Section 1.3.2. Schematically, this is done as follows:

1. u_v and d_v normalized using number sum rule

2. g normalized with fit parameter

3. Remaining momentum split between s and $\bar{s}$, $\bar{u}$, and $\bar{d}$

4. Fix s and $\bar{s}$

5. Momentum sum rule fixes $\bar{u}$, and $\bar{d}$

The number sum rules, Eq. 5.7, constrain the u_v and d_v normalization parameters for all values of A.

$$\int_0^1 dx \; f_{u_v}(x, Q_0) = 2, \quad \int_0^1 dx \; f_{d_v}(x, Q_0) = 1 \tag{5.7}$$

In nCTEQ15, the A-dependent momentum fraction of the gluon is parameterized by:

$$\int_0^1 dx \, x g^{p/A}(x, Q_0) = M_g e^{\left[p_0^g + a_0^g\left(1 - A^{-b_0^g}\right)\right]} , \tag{5.8}$$

where the combination of M_g and p_0^g represent the gluon momentum fraction in the free proton, however in nCTEQ++, these parameters have been combined into a single parameter:

$$M_g e^{\left[p_0^g + a_0^g\left(1 - A^{-b_0^g}\right)\right]} \;\Rightarrow\; \tilde{p}_0^g e^{\left[a_0^g\left(1 - A^{-b_0^g}\right)\right]} , \tag{5.9}$$

in an effort to reduce complexity in the parameterization and normalization definitions. Beyond that, one additional normalization parameter can be determined by the momentum sum rule:

---

[6]$c_0 = p_0$ for $A = 1$; $p_{k \neq 0}$ are fixed to those of a free proton fit; CTEQ 66 [14] for nCTEQ15.

$$\int_0^1 dx \sum_i x f_i(x, Q_0) = 1 \ . \tag{5.10}$$

The remaining normalization parameters are left as free parameters to be fit or are fixed by additional assumptions. One such assumption is that the momentum fraction of the combination of $s + \bar{s}$ is given by:

$$\int_0^1 dx \, x \left(s^{p/A}(x, Q_0) + \bar{s}^{p/A}(x, Q_0) \right) = \tag{5.11}$$

$$\frac{\kappa}{(2+\kappa)} \left(1 - \int_0^1 dx \sum_i x f_i^{p/A} \right) \left[p_{0,0}^{s+\bar{s}} + a_{0,1}^{s+\bar{s}} \left(1 - A^{-b_{0,2}^{s+\bar{s}}} \right) \right] \text{ with } i = u_v, d_v, g \ ,$$

which binds the strange quark distribution to that of the valence quark and the gluon distributions. That leaves the remaining normalization parameter[7] $(\bar{u} + \bar{d})$ to be determined by the momentum sum rule and does not introduce any more free parameters.

As nCTEQ fits follow the same framework of a free proton PDF fit, the momentum fraction of all protons fit are restricted to the range $0 < x < 1$. This constraint is artificially introduced, as the momentum fraction of a parton within a nucleus could obtain a momentum fraction greater than 1. This could occur when a nucleon within the nucleus obtains a momentum greater than that of the average nucleon, allowing the momentum fraction of the parton to climb all the way up to A. However, this would require restructuring the parameterization, the sum rules (Eq. 5.7 and Eq. 5.10), and the implementation of the DGLAP evolution equation. The current implementation of the DGLAP equation assumes $x < 1$, and would not necessarily be valid in this region. At present, all indications are the PDF above $x = 1$ is negligibly small and can be ignored, thus this constraint is applied almost unilaterally[8] throughout the nuclear fitting community and will continue in nCTEQ++.

---

[7]For the ratio of $\bar{d}/\bar{u}$, there is no normalization. Thus the remaining normalization fixed by the sum rule is $\bar{u} + \bar{d}$.

[8]The nuclear PDF analysis presented in [115] is the first fit to use a framework, which at least hypothetically, allows for $x > 1$, however the parameterization used in this analysis remains limited to $x < 1$.

As introduced in Section 1.3.1, in order to best describe the data in a parameterized PDF fit, one seeks to minimize a χ^2 function. A simple expression of this function, shown in Eq. 5.12, relates experimental data values, D_i, to the theoretical predictions, T_i. The set of parameters, $\{a_j\}$, are fit at the input scale, $Q_0 = 1.3$ GeV for `nCTEQ15`, and are adjusted in order to find a stable minimum for the χ^2 function. σ_i represents the errors on the experimental measurements, which are determined by adding the uncorrelated systematic and statistical error in quadrature: $\sigma_i^2 = \sigma_{sys_i}^2 + \sigma_{stat_i}^2$.

$$\chi^2(\{a_j\}) = \sum_i \frac{[D_i - T_i(\{a_j\})]^2}{\sigma_i^2} \tag{5.12}$$

In `nCTEQ15`, as has been done in prior `nCTEQ` analyses [6,8,56,112], this simple description of the χ^2 is rewritten, using [35] as a guide, to allow for correlated uncertainties in the data. To do this, the χ^2 is first decomposed as a sum of χ_n^2 for the n experimental data sets.

$$\chi^2(\{a_j\}) = \sum_n \chi_n^2(\{a_j\}) , \tag{5.13}$$

These new χ_n^2 represent the partial χ^2 for each experiment and can be described as the familiar form for a χ^2 but with an additional correction term.

$$\chi_n^2(\{a_j\}) = \sum_i \frac{[D_i - T_i(\{a_j\})]^2}{\alpha_i^2} - \sum_{k,k'} B_k \, A_{kk'}^{-1} \, B_{k'} , \tag{5.14}$$

The correction term provides the ability to represent correlated uncertainties in the χ^2. Here the index, i, runs over the data points, while k, k' run over the correlated errors. Here, α_i^2 replaces σ_i^2 as the quadratic sum of uncertainties, but allowing the systematic errors to be uncorrelated. The correlated uncertainties are represented by Eq. 5.15, with β_{ik} being the sources of the correlated systematics.

$$B_k(\{a_j\}) = \sum_i \frac{\beta_{ik}\left[D_i - T_i(\{a_j\})\right]}{\alpha_i^2} \ ,$$

$$A_{kk'} = \delta_{kk'} + \sum_i \frac{\beta_{ik}\beta_{ik'}}{\alpha_i^2} \ , \tag{5.15}$$

Decomposing the χ^2 equation this way also allows for the introduction of a weighting term. Here, w_n is the weight given to each experiment[9], but for `nCTEQ15` and all fits presented in this dissertation, $w_n = 1$.

$$\chi^2(\{a_j\}) = \sum_n w_n \ \chi_n^2(\{a_j\}) \ , \tag{5.16}$$

It is important to note that this prescription for calculating χ^2 only includes uncertainty on the experimental data and is agnostic to any uncertainty due to theoretical choices and higher order (NNLO, N3LO, ...) corrections.

5.1.2. Experimental Data

In this section, I will describe the experimental data sets used in the `nCTEQ15` fit [7]. This data is not only relevant to understanding `nCTEQ15`, but the new fits including LHC data, described later in Section 5.4, include this data as well.

The `nCTEQ15` fit includes data from deep inelastic scattering experiments (DIS) and Drell-Yan lepton pair production data[10]. A full catalog of the data used in the fit can be found in Tables 5.1-5.3.

Each type of data set is necessary to include in the fit because the different types of data provide differing constraints when fitting. DIS data represents a majority of the data in the fit and is essential in constraining the u and d valence distributions, as well as the

---

[9]A non-zero weight could be used to preferentially study the impact of a particular data set on a fit. This would be useful should a data set provide a strong constraint on a parameter that otherwise would lack the sensitivity to adaquately fit.

[10]Again, I will neglect the pion production data included in `nCTEQ15`, opting instead to use only data included in `nCTEQ15-np`.

sea distributions. However this type of data is relatively agnostic to the gluon and strange quark distributions. Drell-Yan data is useful in when trying to isolate contributions from the u and d quarks and provides some, if minimal, information about the gluon distribution[11].It is important to note that with the inclusion of more processes, further and more diverse constraints can be added to the fit, reducing uncertainties and improving the fit PDF.

In order to avoid troublesome kinematic ranges when fitting, cuts were introduced on both Q and W. The cut on Q is designed to curb possible effects of non-factorisable higher twist contributions present at low Q, where $\frac{Q^2}{\Lambda_{QCD}^2}$ is small, as well as to ensure the range for DIS. The cut on W, a composite variable defined in Eq. 5.17, eliminates a portion of the large x region to limit the effects of target mass corrections. Both sets of cuts are compatible with cuts used in the free proton analysis on which the parameterization is based.

$$W^2 = \frac{Q^2(1-x)}{x} + m_p^2 \tag{5.17}$$

The cuts for `nCTEQ15` are as follows:

- DIS: $Q > 2$ GeV and $W > 3.5$ GeV

- DY: $2 < M < 300$ GeV
  (with M being the invariant mass of the produced lepton pair)

For `nCTEQ15-np`, 708 data points survive the cuts: 616 DIS and 92 DY. This number is relatively small amongst nPDF fitting groups (EPS09 [9], for example has 929 data points and EPPS16 [53], including LHC data, fits 1811 data points.). However, this is a result of other fitting groups utilizing more relaxed cuts and completely neglecting to cut on W. These looser cuts may provide more data points to fit with but it comes at the cost of potentially introducing troublesome higher twist effects and target mass corrections. For

---

[11]Due to the lack of pion production data in `nCTEQ15-np`, the gluon lacks some constraints available in the `nCTEQ15` fit. This was considered acceptable given that the pion data adds complexity to the fit by introducing an additional dependence on the fragmentation function, which would have made determining the effects of the W/Z LHC data more difficult. Future analyses will include both pion data and W/Z production data from the LHC.

large nuclei, target mass corrections are especially tricky as the high density of partons present in the nucleus that are not relevant to the interaction can cause these corrections to be underestimated [59, 135]. It is possible to mitigate these effects as it has been shown in [136, 137], that these effects can be tempered by using ratios of observables.

Figure 5.1 shows the full set of DIS and DY data used in the `nCTEQ15` fit plotted with respect to the kinematic variables x and Q^2. The dashed lines represent the kinematic cuts imposed on the data, with only data above and to the left of the cut is included in the fit. When the data is visualized in this way, it becomes clear that the fit PDF exceeds the kinematic range covered by the data and thus relies heavily on extrapolation in those regions[12]. The restriction of the data to this region of $x - Q^2$ space provides one of the arguments for the inclusion of LHC W/Z data. We will revisit this concept in Section 5.4.

5.1.3. `nCTEQ15-np` Results

As `nCTEQ15-np` will provide the baseline for the new fits discussed in this chapter, this section will provide a brief summary of this PDF set. As mentioned earlier, `nCTEQ15-np` is equivalent to the `nCTEQ15` PDF fit but with the exclusion of the inclusive pion production data. As the fits including LHC data will also not include pion data, `nCTEQ15-np` is preferred over `nCTEQ15` as a baseline.

5.1.3.1. *Parameterization*

The `nCTEQ15-np` PDFs are fit at an initial scale of $Q_0 = 1.3$ GeV using the parameterization detailed in Section 5.1.1.2. As seen in Table 5.4, for each flavor there are ~ 10 parameters that describe the x and A dependence that can be opened and fit. Due to a lack of data to constrain this many free parameters, `nCTEQ15-np` was fit with only 16 open parameters (highlighted in bold in Table 5.4).

---

[12]Roughly, the kinematic range of the data in Fig. 5.1 covers 2 orders of magnitude in both x and Q^2. PDFs are often fit over 4 orders of magnitude in Q^2 and 5 orders of magnitude in x.
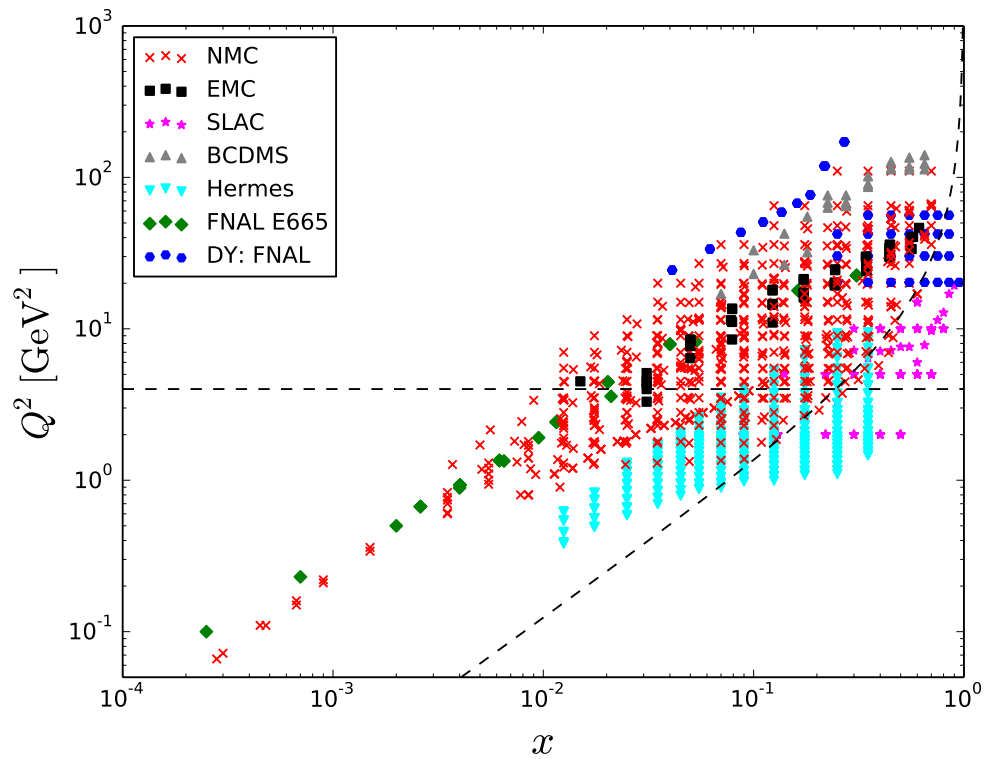
Figure 5.1: The kinematic regime spanned by the DIS and DY data used in `nCTEQ15`. The dashed lines represent the kinematic cuts enforced on the data ($Q > 2$ GeV, $W > 3.5$ GeV), with only the points above and to left of both cuts being included in the fit.

$\mathbf{F_2^A/F_2^D}$: Observable	Experiment	ID	Ref.	# data	# data after cuts	χ^2
D	NMC-97	5160	[116]	292	201	247.81
He/D	Hermes	5156	[117]	182	17	14.33
	NMC-95,re	5124	[118]	18	12	15.64
	SLAC-E139	5141	[119]	18	3	1.08
Li/D	NMC-95	5115	[120]	24	11	14.56
Be/D	SLAC-E139	5138	[119]	17	3	1.00
C/D	FNAL-E665-95	5125	[121]	11	3	0.75
	SLAC-E139	5139	[119]	7	2	1.75
	EMC-88	5107	[122]	9	9	7.45
	EMC-90	5110	[123]	9	0	0.00
	NMC-95	5113	[120]	24	12	9.38
	NMC-95,re	5114	[118]	18	12	18.63
N/D	Hermes	5157	[117]	175	19	9.88
	BCDMS-85	5103	[124]	9	9	4.59
Al/D	SLAC-E049	5134	[125]	18	0	0.00
	SLAC-E139	5136	[119]	17	3	1.86
Ca/D	NMC-95,re	5121	[118]	18	12	27.44
	FNAL-E665-95	5126	[121]	11	3	0.86
	SLAC-E139	5140	[119]	7	2	2.09
	EMC-90	5109	[123]	9	0	0.00
Fe/D	SLAC-E049	5131	[126]	14	2	1.10
	SLAC-E139	5132	[119]	23	6	6.07
	SLAC-E140	5133	[127]	10	0	0.00
	BCDMS-87	5101	[128]	10	10	4.41
	BCDMS-85	5102	[124]	6	6	2.18
Cu/D	EMC-93	5104	[129]	10	9	7.64
	EMC-93(chariot)	5105	[129]	9	9	8.08
	EMC-88	5106	[122]	9	9	5.90
Kr/D	Hermes	5158	[117]	167	12	9.09
Ag/D	SLAC-E139	5135	[119]	7	2	1.87
Sn/D	EMC-88	5108	[122]	8	8	20.61
Xe/D	FNAL-E665-92	5127	[130]	10	2	0.62
Au/D	SLAC-E139	5137	[119]	18	3	1.74
Pb/D	FNAL-E665-95	5129	[121]	11	3	1.48
Total:				1205	414	449.90

Table 5.1: The DIS F_2^A/F_2^D data sets used in the `nCTEQ15` fit. The table details values of χ^2 for each experiment, the specific nuclear targets, references, and the number of data points with and without kinematic cuts. This and the following tables originally come from Ref. [7].

$\mathbf{F_2^A/F_2^{A'}}$:					# data	
Observable	Experiment	ID	Ref.	# data	after cuts	χ^2
C/Li	NMC-95,re	5123	[118]	25	7	5.90
Ca/Li	NMC-95,re	5122	[118]	25	7	1.35
Be/C	NMC-96	5112	[131]	15	14	4.04
Al/C	NMC-96	5111	[131]	15	14	5.23
Ca/C	NMC-95,re	5120	[118]	25	7	4.16
	NMC-96	5119	[131]	15	14	5.25
Fe/C	NMC-96	5143	[131]	15	14	9.36
Sn/C	NMC-96	5159	[132]	146	111	64.88
Pb/C	NMC-96	5116	[131]	15	14	6.85
Total:				296	202	107.02

Table 5.2: The DIS $F_2^A/F_2^{A'}$ data sets used in the `nCTEQ15` fit. The same details for each data set as in Tab. 5.1 are listed here.

$\sigma_{DY}^{pA}/\sigma_{DY}^{pA'}$:					# data	
Observable	Experiment	ID	Ref.	# data	after cuts	χ^2
C/H2	FNAL-E772-90	5203	[133]	9	9	7.80
Ca/H2	FNAL-E772-90	5204	[133]	9	9	2.70
Fe/H2	FNAL-E772-90	5205	[133]	9	9	3.15
W/H2	FNAL-E772-90	5206	[133]	9	9	7.25
Fe/Be	FNAL-E886-99	5201	[134]	28	28	23.48
W/Be	FNAL-E886-99	5202	[134]	28	28	24.31
Total:				92	92	68.70

Table 5.3: The Drell-Yan process data sets used in the `nCTEQ15` fit. The same details for each data set as in Tab. 5.1 are listed here.

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
p_0^g	(0.38232)	–	–	–	–	–	–	$p_0^{s+\bar{s}}$	(0.500)	–	–
p_1^g	(0.38232)	$p_1^{u_v}$	(0.630)	$p_1^{d_v}$	(0.513)	$p_1^{\bar{d}+\bar{u}}$	(-0.324)	$p_1^{s+\bar{s}}$	(-0.324)	$p_1^{\bar{d}/\bar{u}}$	(10.075)
p_2^g	(3.03369)	$p_2^{u_v}$	(2.934)	$p_2^{d_v}$	(4.211)	$p_2^{\bar{d}+\bar{u}}$	(8.116)	$p_2^{s+\bar{s}}$	(8.116)	$p_2^{\bar{d}/\bar{u}}$	(4.957)
p_3^g	(4.39448)	$p_3^{u_v}$	(-2.369)	$p_3^{d_v}$	(-2.375)	$p_3^{\bar{d}+\bar{u}}$	(0.413)	$p_3^{s+\bar{s}}$	(0.413)	$p_3^{\bar{d}/\bar{u}}$	(15.167)
p_4^g	(2.35917)	$p_4^{u_v}$	(1.266)	$p_4^{d_v}$	(0.965)	$p_4^{\bar{d}+\bar{u}}$	(4.754)	$p_4^{s+\bar{s}}$	(4.754)	$p_4^{\bar{d}/\bar{u}}$	(17.000)
p_5^g	(-3.000)	$p_5^{u_v}$	(1.718)	$p_5^{d_v}$	(3.000)	$p_5^{\bar{d}+\bar{u}}$	(0.614)	$p_5^{s+\bar{s}}$	(0.614)	$p_5^{\bar{d}/\bar{u}}$	(9.948)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
a_0^g	(-0.256)	–	–	–	–	–	–	$a_0^{s+\bar{s}}$	(0.167)	–	–
a_1^g	**-0.001**	$a_1^{u_v}$	**-2.71781**	$a_1^{d_v}$	**0.272**	$a_1^{\bar{d}+\bar{u}}$	**0.406**	$a_1^{s+\bar{s}}$	(0.406)	$a_1^{\bar{d}/\bar{u}}$	(0.000)
a_2^g	(0.000)	$a_2^{u_v}$	**-0.162**	$a_2^{d_v}$	**-0.196**	$a_2^{\bar{d}+\bar{u}}$	(0.415)	$a_2^{s+\bar{s}}$	(0.415)	$a_2^{\bar{d}/\bar{u}}$	(0.000)
a_3^g	(0.383)	$a_3^{u_v}$	(0.018)	$a_3^{d_v}$	(0.085)	$a_3^{\bar{d}+\bar{u}}$	(-0.759)	$a_3^{s+\bar{s}}$	(0.000)	$a_3^{\bar{d}/\bar{u}}$	(0.000)
a_4^g	**0.055**	$a_4^{u_v}$	**12.083**	$a_4^{d_v}$	(3.874)	$a_4^{\bar{d}+\bar{u}}$	(-0.203)	$a_4^{s+\bar{s}}$	(0.000)	$a_4^{\bar{d}/\bar{u}}$	(0.000)
a_5^g	**0.002**	$a_5^{u_v}$	**-1.140**	$a_5^{d_v}$	**-0.080**	$a_5^{\bar{d}+\bar{u}}$	**-0.084**	$a_5^{s+\bar{s}}$	(0.000)	$a_5^{\bar{d}/\bar{u}}$	(0.000)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
b_0^g	**-0.038**	–	–	–	–	–	–	$b_0^{s+\bar{s}}$	(0.104)	–	–
b_1^g	**-1.341**	$b_1^{u_v}$	(0.006)	$b_1^{d_v}$	(0.466)	$b_1^{\bar{d}+\bar{u}}$	(0.172)	$b_1^{s+\bar{s}}$	(0.172)	$b_1^{\bar{d}/\bar{u}}$	(0.000)
b_2^g	(0.000)	$b_2^{u_v}$	(0.524)	$b_2^{d_v}$	(0.440)	$b_2^{\bar{d}+\bar{u}}$	(0.290)	$b_2^{s+\bar{s}}$	(0.290)	$b_2^{\bar{d}/\bar{u}}$	(0.000)
b_3^g	(0.520)	$b_3^{u_v}$	(0.073)	$b_3^{d_v}$	(0.107)	$b_3^{\bar{d}+\bar{u}}$	(0.298)	$b_3^{s+\bar{s}}$	(0.000)	$b_3^{\bar{d}/\bar{u}}$	(0.000)
b_4^g	**-0.516**	$b_4^{u_v}$	(0.038)	$b_4^{d_v}$	(-0.018)	$b_4^{\bar{d}+\bar{u}}$	(0.888)	$b_4^{s+\bar{s}}$	(0.000)	$b_4^{\bar{d}/\bar{u}}$	(0.000)
b_5^g	**-1.418**	$b_5^{u_v}$	(0.615)	$b_5^{d_v}$	(-0.236)	$b_5^{\bar{d}+\bar{u}}$	(1.353)	$b_5^{s+\bar{s}}$	(0.000)	$b_5^{\bar{d}/\bar{u}}$	(0.000)

Table 5.4: Values of the parameters of the `nCTEQ15-np` fit at the initial scale $Q_0 = 1.3\text{GeV}$. Values in bold represent the free parameters and values in parentheses are fixed in the fit. The first block of parameters, the p_k's refer to the free proton parameters. The normalization parameters, determined by the sum rules are not listed.

For `nCTEQ15-np`, the total χ^2 was found to be 625.6 for 708 data points. This yields a χ^2 per degree of freedom of 0.884 (625.6 / 708) which indicates a good fit. The χ^2 for each data set is presented in Fig. 5.30. Of note is the DIS set (ID:5108) which consists of Sn/D data from EMC-88; this set has proven difficult to accommodate in both this fit as well as for other analyses [10] [52].
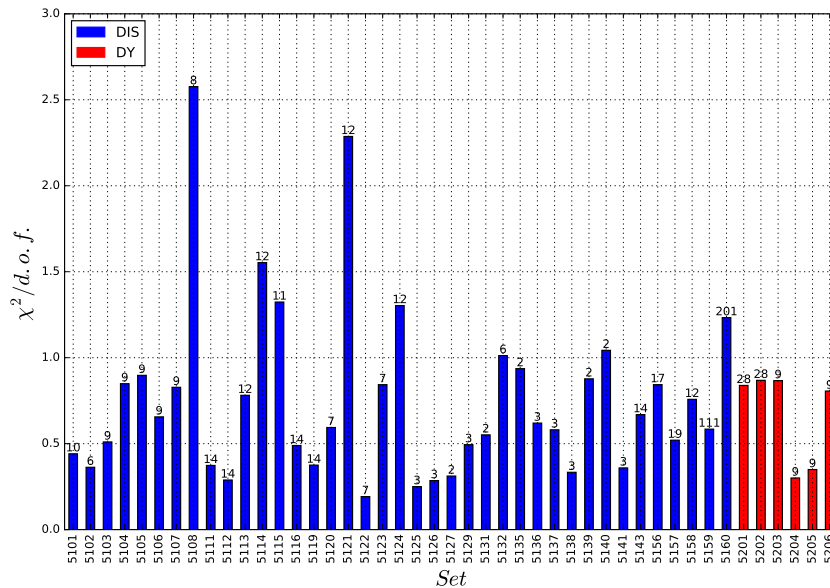
Figure 5.2: Listing of the χ^2/d.o.f. for each data set included in `nCTEQ15-np`. The individual data sets are identified by the ID number corresponding to those in Tables 5.1-5.3. The DIS data is displayed in **blue** and have ID numbers corresponding to 51XX, while DY data is displayed in **red** and have ID's corresponding to 52XX.

5.1.3.3. `nCTEQ15-np` *PDFs*

Given that a nPDF set contains a full family of bound proton PDFs, it is prudent to examine the A dependence of the resulting PDFs. Figure 5.3 shows the central value for a number of different nuclei from $A = 1$ (proton/hydrogen) to $A = 208$ (lead). By examining these figures, one can see a softening of the gluon and the sea quark distributions at high x for larger values of A by the increase in the PDF at low x. The valence quark distributions

see a similar softening but the effect is reduced with respect to the other flavors. A more direct comparison between the A values of $A = 1$ and $A = 208$ along with the ratio of the PDFs can be found in Fig. 5.4 and 5.5. Here the ratios represent the effective nuclear correction applied to the proton PDF in order to produce the lead PDF. This correction is a result of the A dependent parameterization.

To this point, the material covered concerning `nCTEQ15` and `nCTEQ15-np` has been necessary for not only understanding the `nCTEQ` fitting framework but also establishing a benchmark against which any new fit including LHC data will be compared. Before diving into a description of a new fit, it is important to describe the updated `nCTEQ++` fitting code and the other computational machinery adapted to make improve the fitting process.
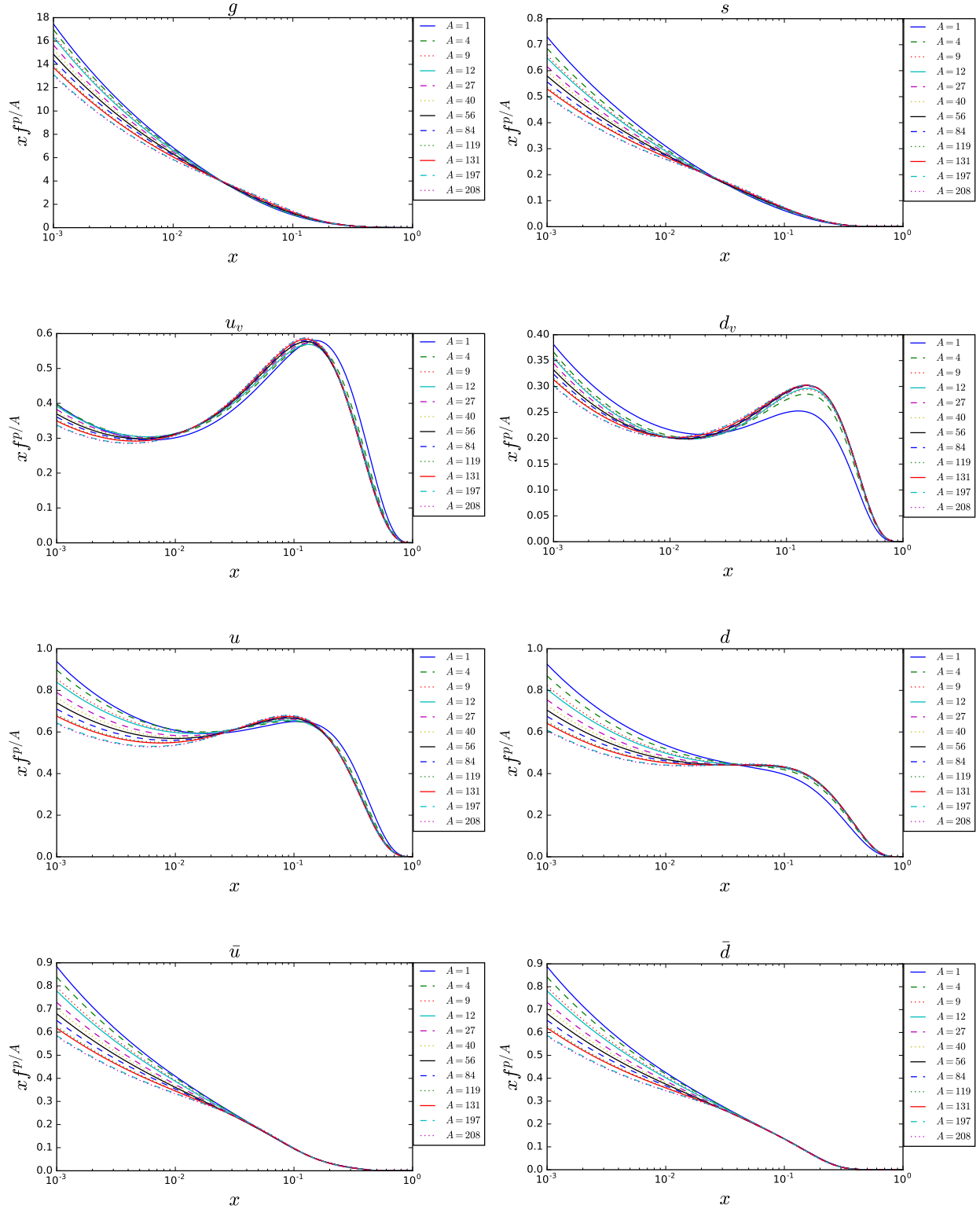
Figure 5.3: `nCTEQ15-np` bound proton PDFs for $Q = 10$ GeV for the free proton ($A = 1$) to lead ($A = 208$).
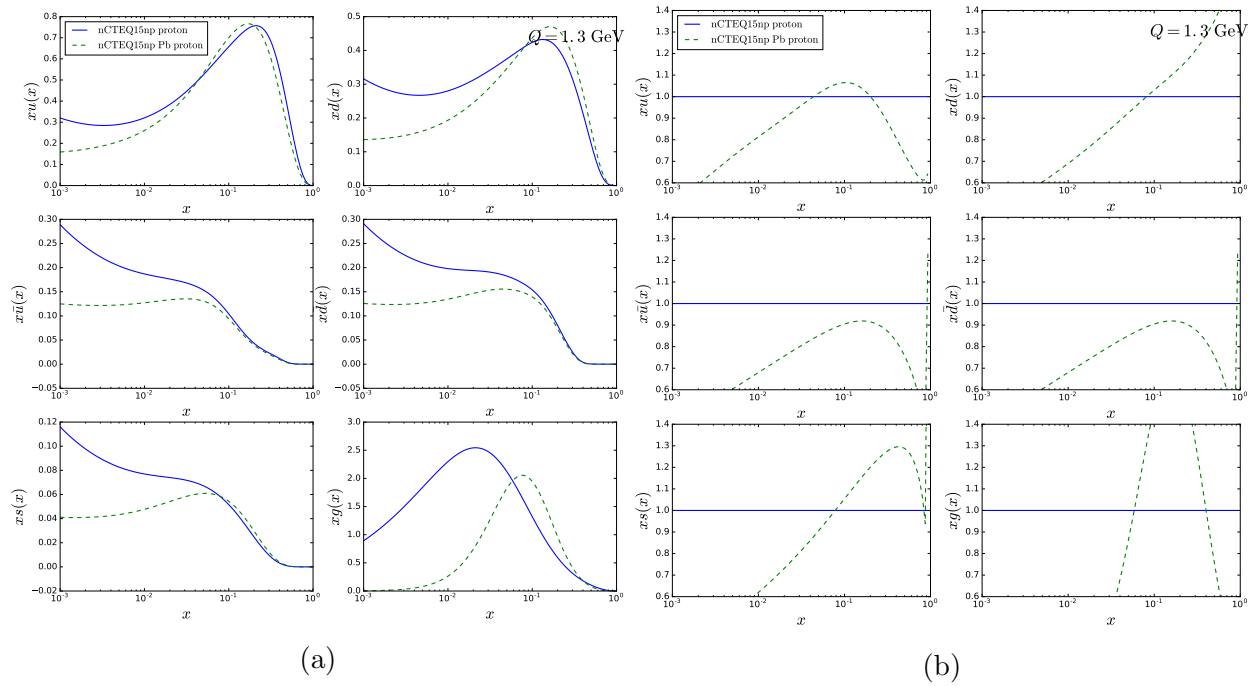
Figure 5.4: `nCTEQ15-np` bound proton PDF compared to the bound proton in lead 5.4a: for $Q = 1.3$ GeV as well as the ratio between them, 5.4b. The ratio plots represent the effective nuclear correction to the proton.
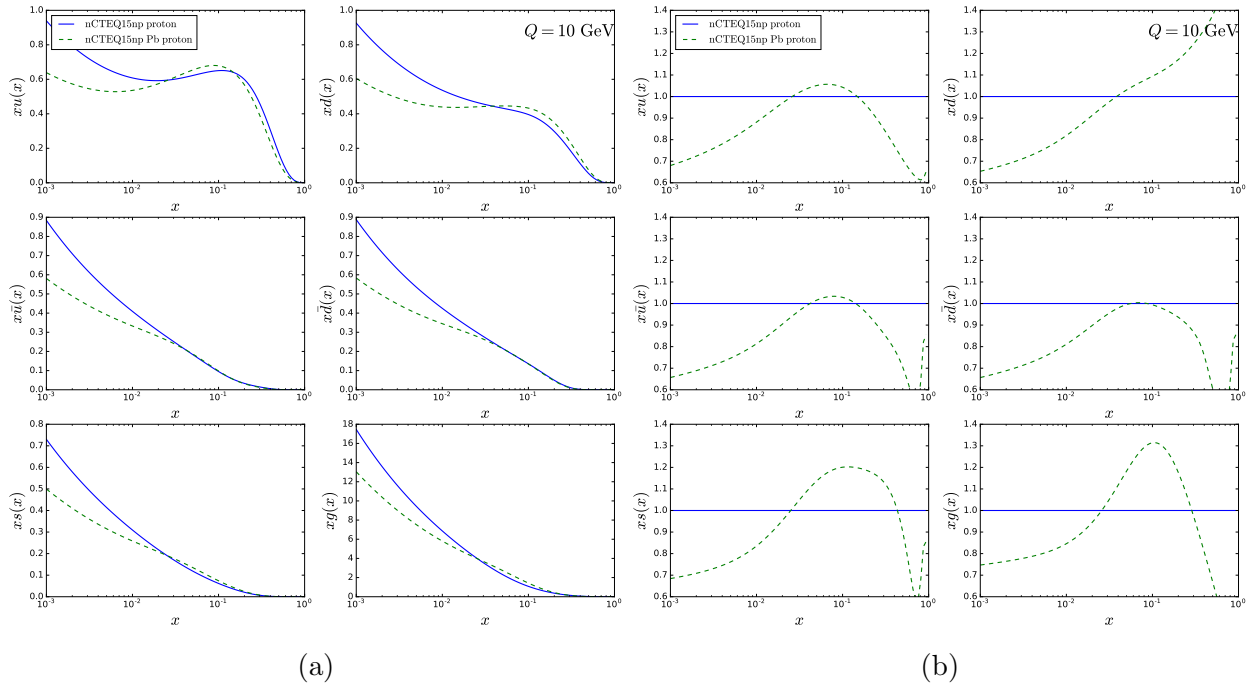
Figure 5.5: `nCTEQ15-np` bound proton PDF compared to the bound proton in lead 5.6a: for $Q = 10$ GeV as well as the ratio between them, 5.6b. The ratio plots represent the effective nuclear correction to the proton.

5.2. nCTEQ++

The original nCTEQ code was a FORTRAN-based code derived and modified from the CTEQ proton fitting code [8]. This code had many advantages but ultimately lacked flexibility when adding new processes. Each new process required a dedicated module written specifically to perform the theory calculations for that process. That module would then have to be implemented into the FORTRAN coding base without disrupting any of the existing processes. This was incredibly time intensive and required a fair amount of coding expertise to accomplish. As an example, the inclusive pion production module took several months to build and implement in this code base in order to produce nCTEQ15.

In an effort to modernize the nCTEQ code and port it from FORTRAN into C++, nCTEQ++ was created. This code was designed to take advantage of some of the improvements provided by the PDF fitting community. New tools such as LHAPDF6[13] [13], MCFM[14] [138], HOPPET[15] [139] and APPLgrid[16] [140] have all been integrated in this new code in various ways. These tools also solve the flexibility issue in the old FORTRAN code by providing opportunities for external theory modules to be implemented directly into the fitting process without sacrificing speed (a previous issue with the complexity in the FORTRAN code).

However, with the implementation of all of the new features, it became necessary to confirm that nCTEQ++ could replicate the results of its predecessor. With a top to bottom rewrite like this, it was unlikely that the codes would ever produce identical results (if simply due to the repair of some small bugs that had plagued the FORTRAN). Thus I performed a series of validations to confirm that I was able to identify and quantify any and all differences between the new and old codes. Once these variations were cataloged, I then examined the parameter space from nCTEQ15-np in nCTEQ++ to determine if there was any change to the PDF fits as a result of the transition. This was necessary in order to prevent the results

---

[13]Les Houches Accord Parton Distribution Functions: https://lhapdf.hepforge.org/

[14]Monte Carlo for FeMtobarn processes: https://mcfm.fnal.gov/

[15]Higher Order Perturbative Parton Evolution Toolkit: https://hoppet.hepforge.org/

[16]The APPLgrid project: https://applgrid.hepforge.org/

of any future fit from being attributed or misattributed to differences in the fitting code as opposed to physics results.

5.2.1. Validation

The most significant change in the process of fitting came with the implementation of the `HOPPET` PDF evolution code [139]. `HOPPET` allows the evolution of the parameterized PDF from the input scale, Q_0 to the energy scale for a given data set to be done quickly without needing to write in an internal evolution routine. Additionally this has the advantage of being maintained and improved by an external group to `nCTEQ`, and is widely recognized tool within the PDF fitting community. Despite the inclusion of `HOPPET` in `nCTEQ++` being an overall improvement, it lead to some inevitable differences with the `FORTRAN` code. These differences are detailed in the following sections.

5.2.1.1. Differences in Solving for α_S

In Section 1.1.2, I presented the following equation for calculating the running coupling for QCD, α_S, at leading order.

$$\alpha_s(Q^2) = \frac{1}{b \, \ln(Q^2/\Lambda_{QCD}^2)} \tag{5.18}$$

However, when performing PDF fits to higher orders, this equation is no longer sufficient and additional correction terms are needed to properly determine $\alpha_S(Q^2)$. In order to accurately determine the value of α_S to the order of one's fit, you must solve Eq. 5.19 to the order desired.

$$\frac{d\alpha_S}{d\ln(Q^2)} = \beta(\alpha_S(Q^2)) = -(b_0\alpha_S^2 + b_1\alpha_S^3 + b_2\alpha_S^4 + ...) \tag{5.19}$$

This equation defines the β function which determines the running of α_S. The β function can be calculated pertubatively allowing it to be matched to the order of the fitting. The

coefficients, b_i are functions of the number of flavors, n_f, in the fit and, for b_2 and above, are renormalization scheme dependent. The first two b_i coefficients are presented here:

$$b_0 = \frac{33 - 2n_f}{12\pi} \quad , \quad b_1 = \frac{153 - 19n_f}{24\pi^2} \quad . \tag{5.20}$$

Due to this reliance on solving the β function pertubatively, it is possible to use different differential equation solution strategies to reach a value for α_S. In the nCTEQ FORTRAN code the β function was solved as a truncated analytic series, effectively calculating the solution to one term past the order of the fit (e.g. for an NLO fit, the series would be truncated after the b_2 term). This technique is simple and effective as nCTEQ fits never go past NLO. However, with eyes to future fits, nCTEQ++ transitioned from that truncated fit to the Runge-Kutta solution implemented in HOPPET [139].

Additionally, when solving for α_S, it is necessary to "pin down" α_S at a single experimentally determined value from which to start the evolution (either up, down or both). In the FORTRAN code, this was done by setting $\alpha_S(M_Z = 91.118 \text{ GeV}) = 0.118$. In nCTEQ++, it is possible to choose the value for this matching but for all fits shown in this dissertation α_S was matched at M_Z as it was in the FORTRAN code.

As seen from the b_i coefficients in the β function, there is also a dependence on the number of active flavors when fitting α_S. While dependencies of this nature in fitting are described at length in Chapter 4, for α_S it is sufficient to use a simple switch in the number of active flavors at the quark masses. In the FORTRAN code this is simplified further, using $n_f = 4$ below $Q = m_b = 4.5 \text{ GeV}$ and $n_f = 5$ above m_b (neglecting the top quark entirely[17]). In nCTEQ++, an additional switch from $n_f = 5$ to $n_f = 6$ may occur at $Q = m_t = 175 \text{ GeV}$.

Figure 5.6 shows the the total effect of these differences in α_S between the FORTRAN code and nCTEQ++. Since both codes match α_S at M_Z, the differences in solution method become more dramatic as $Q \Rightarrow Q_0 = 1.3 \text{ GeV}$ where the value of α_S increases asymptotically. The inclusion of a $n_f = 6$ threshold causes the comparison to diverge above the m_t.

---

[17]This was sufficient as very few of the data points fit in nCTEQ15 were near the mass of the top quark
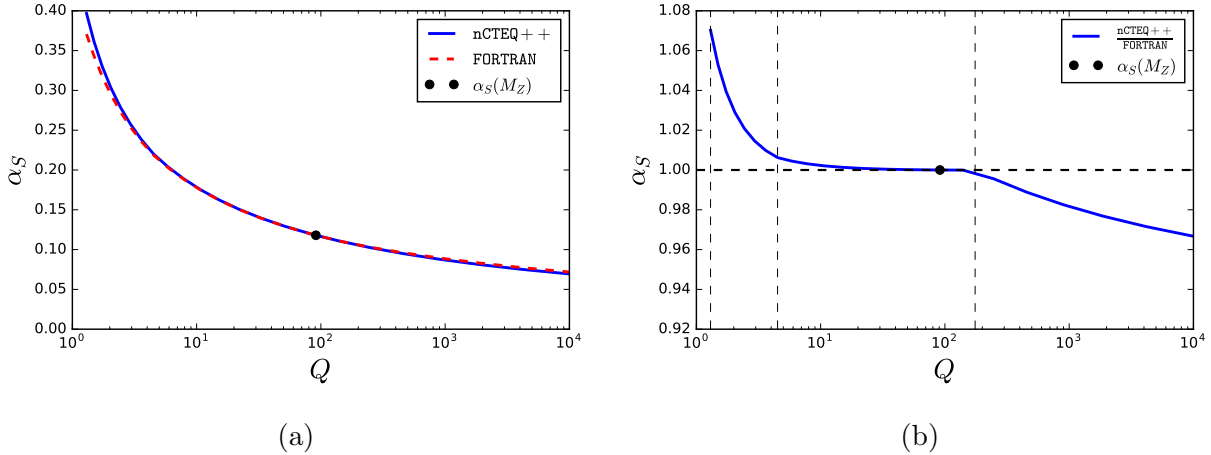
Figure 5.6: Comparison between the α_S implementation (a) in the nCTEQ FORTRAN and nCTEQ++ as well as (b) the ratio of α_S in nCTEQ++ to nCTEQ FORTRAN. Both implementations match $\alpha_S(M_Z = 91.118 \text{ GeV}) = 0.118$. Additionally the vertical lines represent $m_c = 1.3$ GeV, $m_b = 4.5$ GeV, $m_t = 175$ GeV from left to right.

5.2.1.2. Hoppet Evolution

This difference in the running of $\alpha_S(Q)$ has an effect on the evolution as well. In nCTEQ++, the evolution of the PDF is performed by HOPPET where in the nCTEQ FORTRAN code, the evolution is done by an internal routine. As noted in Fig. 5.7, 5.8 and 5.9, there is a slight difference in the results of these evolution routines. The PDFs in these figures are identical with the exception of the evolution code and the corresponding difference in α_S. This is demonstrated by taking the ratio of the two at the input scale, $Q_0 = 1.3$ GeV, where the evolution is not performed. For a more comprehensive comparison of these PDFs, see Appendix B.1.

This effect of α_S on the evolution of the PDFs behaves as expected when comparing to Fig. 5.6. We match α_S at the scale M_Z, so the two results compare well in this region; however, as we move away (to either low or high scales) the numerical difference accumulate. We observe the PDFs are identical at the initial evolution scale Q_0, but then the differences in α_S cause them to differ as we evolve. Above the scale of m_b, the α_S values are comparable and the ratios of the PDFs remain relatively stable to larger scales. We reach the top mass scale m_t, the α_S values differ as HOPPET switches to $n_f = 6$ while the nCTEQ FORTRAN code

146
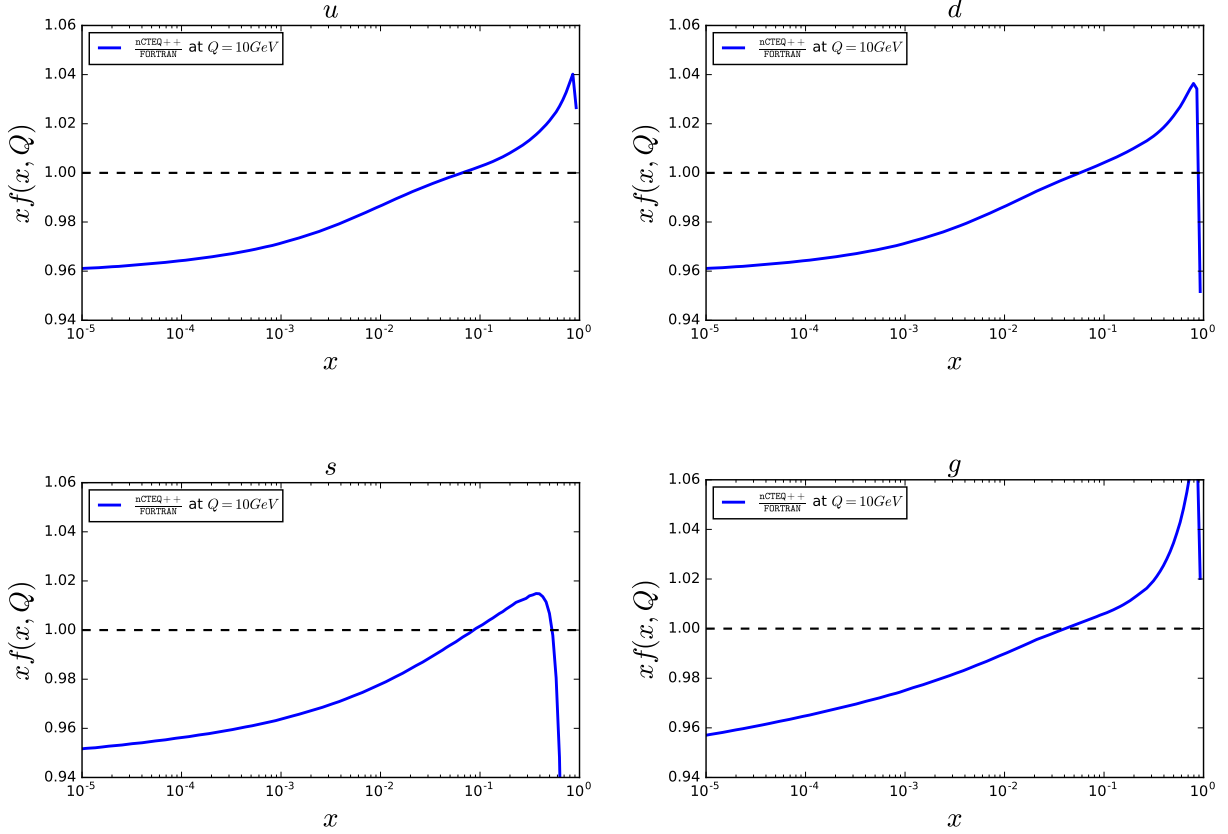
Figure 5.7: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here Q is fixed to be 10 GeVand multiple flavors are displayed. The difference between the evolution codes is maintains a similar shape for each flavor across all values of x. This difference is primarily due to differences in the α_S evolutions explained in Fig. 5.6. The effect of α_S is less directly identifiable when viewed from this slice of the x, Q and flavor space in which the PDFs reside.

keeps $n_f = 5$; this difference is reflected in the PDF ratios at large scale primarily in the gluon PDFs which communicates most directly (at order α_S^1) with the top PDF.

Both the `nCTEQ` internal evolution code and `HOPPET` have been extensively benchmarked [139,141,142]. Because `HOPPET` is a public code and is maintained and updated by an external group, it was chosen to provide the evolution in `nCTEQ++`[18].

---

[18]Additionally, `HOPPET` is an accepted evolution code by PDF4LHC [143]

5.2.1.3. Determining Effects of `nCTEQ++` *on* `nCTEQ15-np` χ^2

Since `nCTEQ++` is the fitting code of choice in the remainder of this dissertation, it is important to recalculate the χ^2 for `nCTEQ15-np` in within the new framework. In the `FORTRAN` code, the χ^2 was found to be 625.6 for 708 data points[19] or a $\chi^2/$d.o.f of 0.883. In `nCTEQ++`, this χ^2 was found to be 640.8 for the same 708 data points, giving a $\chi^2/$d.o.f. of 0.905. This gives an absolute difference of 15.2 between the two codes (note the `nCTEQ15-np` tolerance was $T = \Delta\chi^2 = 35$), and a percent difference of 2.39%.

5.2.2. Parameter Scans

Given that the parameters fit in `nCTEQ15-np` are no longer at a minimum in `nCTEQ++`, one dimensional scans of the parameter space were performed to determine how far the parameters at the old minimum had drifted from the actual minimum. This study was done by opening individual parameters originally fit in `nCTEQ15-np` and allowing them to float in a range around their "minimum" value to map out their parameter space.

Figure 5.10 displays the scans for each of the 16 parameters originally fit in `nCTEQ15-np`. Here the blue point indicates the parameter's original minimum value in the `nCTEQ15-np` where the χ^2 was found to be 640.768. The black horizontal lines represent ± 35 units of χ^2 from the minimum in `nCTEQ15-np`.[20]

Though scanning in this way gives a simplified interpretation of the parameter space, as it neglects any and all interplay between the parameters themselves, it does provide a useful estimate as to the difference on a parameter by parameter basis between the fitting procedures. When evaluated in the context of the tolerance criteria from `nCTEQ15`, the scans also allow for a "goodness" of fit comparison across fitting procedures.

---

[19]As a reminder, these 708 data points are for `nCTEQ15-np` and do not include the pion production data included in `nCTEQ15`, nor does this χ^2 value. For completeness, the χ^2 for `nCTEQ15` including the pion data, as stated in [7], was 587.4 for 740 data points, or 0.81 $\chi^2/$d.o.f. (18 free parameters)

[20]This choice of tolerance criteria is described at length in Appendix A of [7], the paper associated with `nCTEQ15`.

The `nCTEQ15-np` tolerance was $T = \Delta\chi^2 = 35$; thus, any result in the range between the middle and upper lines of Fig. 5.10 would be considered acceptable. Correspondingly, if the minimum for the new `nCTEQ++` fit is in the range between the middle and lower lines of Fig. 5.10, then the original `nCTEQ15-np` fit would be considered acceptable.

Since the new minimum for every parameter fit in `nCTEQ15` is within the $T = 35$ tolerance criteria, it was determined that the `nCTEQ15-np` fit is a sufficient baseline for the present comparative studies. This direct comparison between the published `nCTEQ15-np` results and the new fits performed in Section 5.4 will allow the study of the impact of LHC W/Z data on actual the actual fits without being mired in comparisons to some intermediate result.

Figure 5.8: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here the parton flavor is fixed to be $\bar{u}$ and multiple Q values of note are displayed. The difference between the evolution codes is maintains a similar shape for each evolved value of Q across all values of x. However, it is possible to see signs of the impact of α_S specifically where Q is small but above Q_0 (2GeV, m_b). For those values, the PDFs start identically at Q_0 and then begin to diverge up to m_b where the magnitude of the difference stabilizes becausee the α_S values match, see Fig. 5.6. It is worth noting that at Q_0 the PDFs are identical; for $x \to 1$ the PDFs are ~ 0 leading the ratio of the PDFs to become unstable. Here $Q_0 = 1.3$GeV, $m_b = 4.5$GeV and $M_Z = 91.188$GeV.
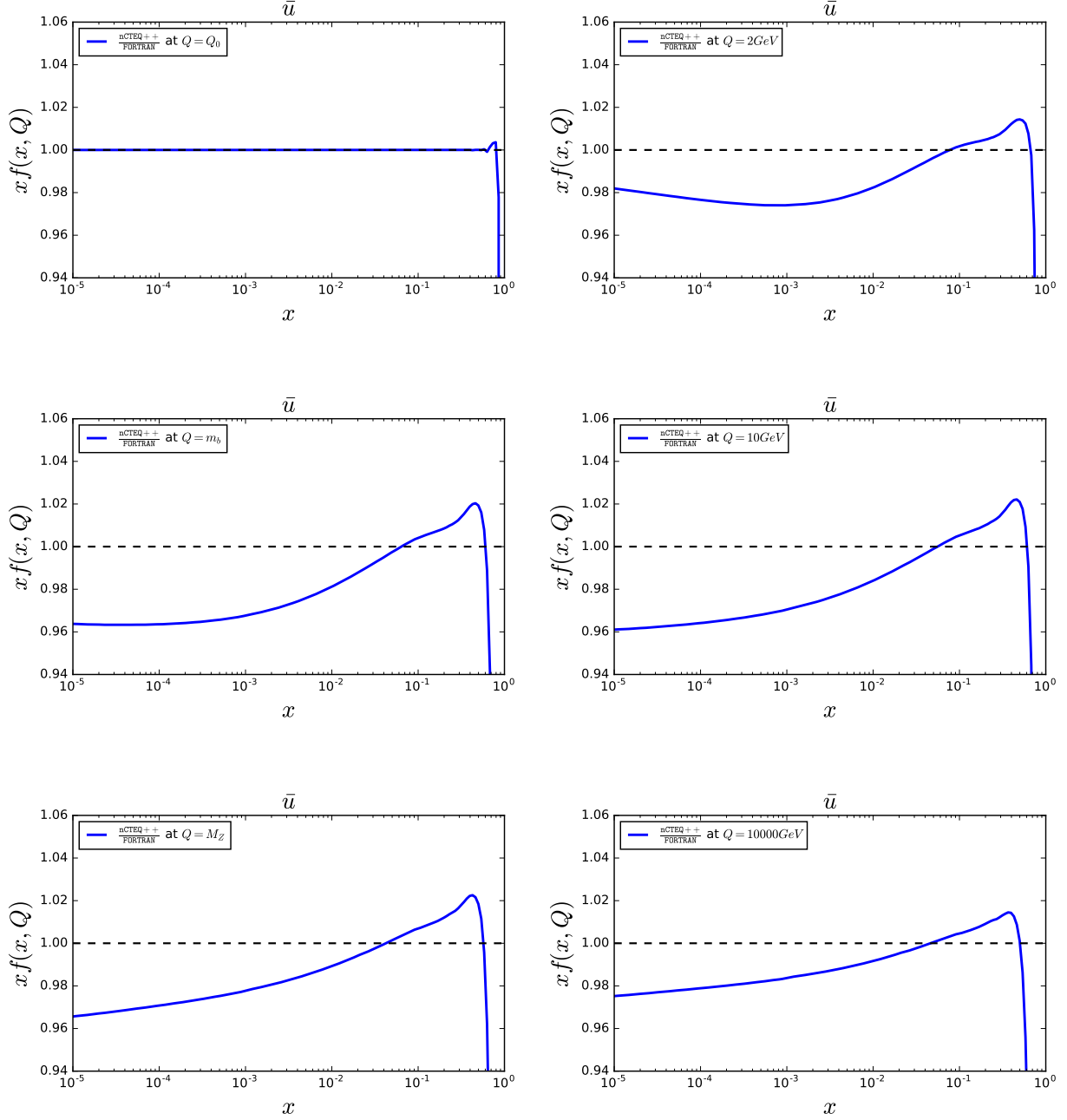
150
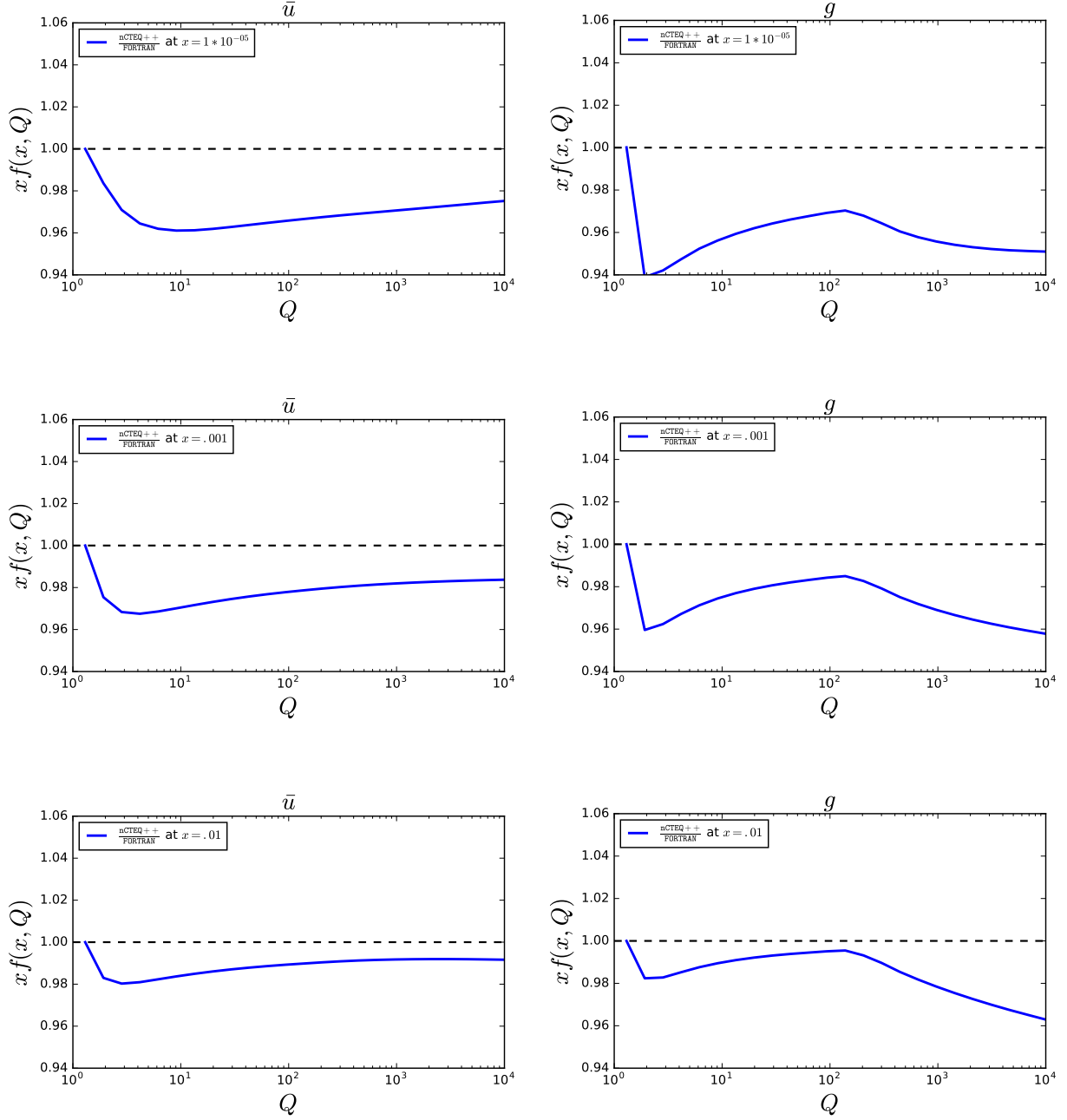
Figure 5.9: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here the gluon and $\bar{u}$ distributions are shown for multiple x values. When viewing the PDF's Q dependence in this way, it is apparent that the differences in α_S and the handling of the top quark are the cause of the characteristic shape seen in Figures 5.7 and 5.8.
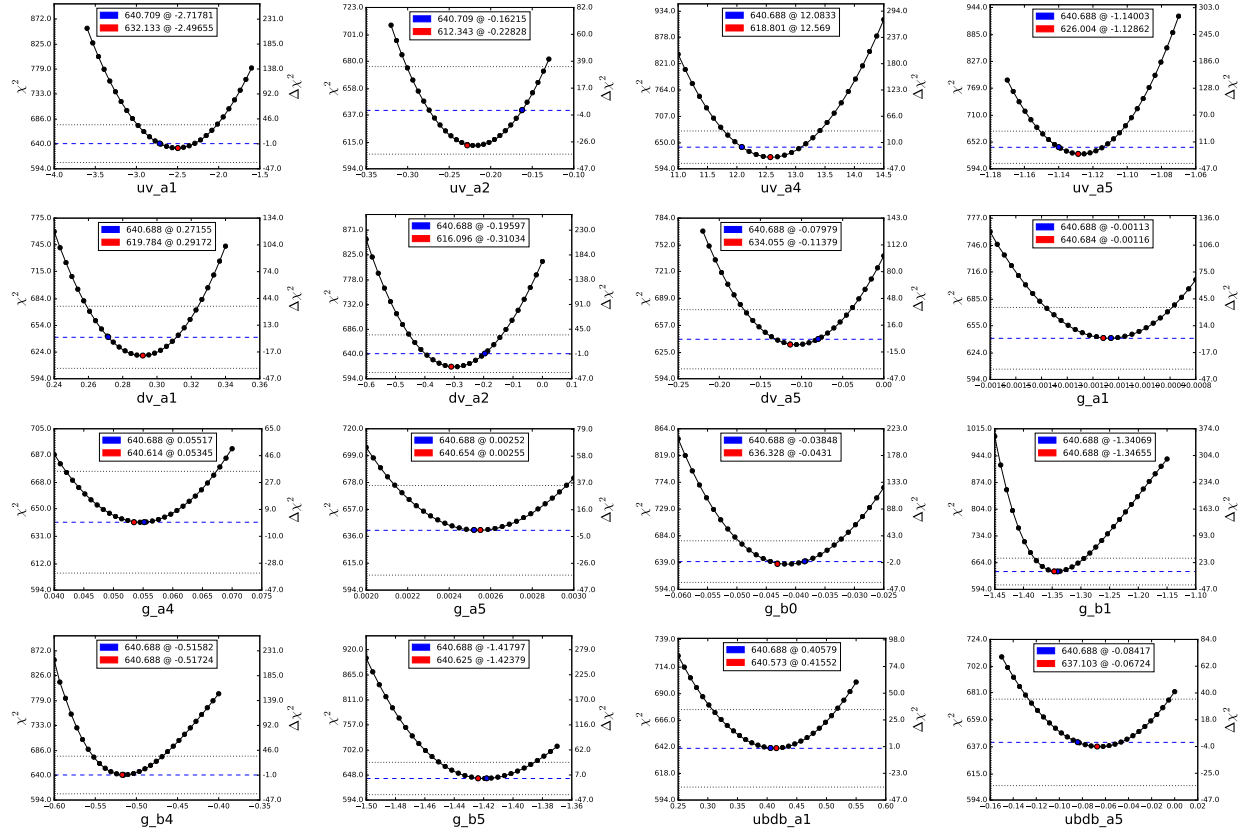
Figure 5.10: One dimensional parameter scans for the 16 free parameters fit in `nCTEQ15-np`. Each scan was performed by opening a single parameter while the others remained fixed, eliminating any correlation between parameters. The **blue** marker indicates the value of the parameter and the associated χ^2 from the minimum previously found in the `nCTEQ FORTRAN` fitting code. The **red** marker indicates where the parameter is at a minimum in `nCTEQ++` as well as the χ^2 at that point. The dotted horizontal black lines indicate ± 35 units of χ^2. Recall that the tolerance criteria in `nCTEQ15` is $T = 35$ [7]. That all of the parameters are minimized in `nCTEQ++` within this tolerance criteria, indicates that `nCTEQ15-np` can be used as a baseline against which future fits can be compared without needing to refit `nCTEQ15-np` in `nCTEQ++`.

5.3. Theory Tool Chain

With `nCTEQ++` validated as a viable fitting code and `nCTEQ15-np` established as the benchmark against which the future fits are compared, let me begin to describe the process by which LHC W/Z data will be incorporated into these new fits. While the primary goal of this work was simply to study the impact that these data sets might have on an `nCTEQ` fit, a secondary, and possibly more significant achievement, was the construction of a chain of software tools that allow for the addition of new data sets and processes into `nCTEQ++` to be fast, flexible, and robust. This procedure represents a significant improvement over the `FORTRAN nCTEQ` code, taking tasks that once took months of dedicated work down to a matter of days. This section will explain the links in this chain as well as the numerous validations performed for each of them.

5.3.1. Overview

In Chapter 3, I introduced the `FEWZ` [1][21] software tool for making theory calculations for W/Z production at the LHC as part of the study in PDF reweighting. In that study, the code was modified in such a way that it was possible to perform not just symmetric collisions between nuclei (e.g. pp and PbPb), but also asymmetric ones (e.g. pPb). Additionally, the individual cuts, binning and observables from several W/Z data sets were compiled and written into the code such that it was possible to make comparisons between theoretical calculations and the experimental data.

After the completion of the reweighting study, it became evident that a refit of `nCTEQ15` would eventually be necessary in order to fully incorporate the LHC data. As fitting can be a time consuming process in its own right, being able to make quick theory calculations is incredibly important. In order to satisfy this desired functionality, an interface to `APPLgrid` was incorporated in `nCTEQ++`. `APPLgrid` provides a structure (called grids) for theoretical calculations to be computed independently from (and in advance of) a fitting calculation. This effectively converts the slow convolution integral from the factorization theorem into

---

[21]Fully Exclusive W, Z Production through NNLO in pQCD: http://gate.hep.anl.gov/fpetriello/FEWZ.html

a series of very fast matrix multiplications[22]. However, this drastic improvement in speed comes at the cost of reduced flexibility as each grid becomes dependent of the inputs used to create it. This can be used to some benefit as the cuts and binning from an experiment can be hardwired directly into the precalculated theory calculation, but unfortunately so is the underlying PDF used in the initial calculation. When fitting a PDF, the PDF is constantly changing and the "shadow" of this past PDF would provide an insurmountable complication to using the grids.

As I mentioned, `APPLgrid` provides a structure for the theory calculations to be stored, but it does not actually make the predictions themselves. Ideally, `FEWZ` would be able to provide the theory calculations that would be stored in the grids, but there currently does not exist a simple way to interface these two codes. Fortunately, `APPLgrid` does have an existing interface to `MCFM`[23]. Additionally, with `MCFM` and an `APPLgrid` utility known as `applgrid-combine` it is possible to eliminate the residual dependence of the grids on the underlying PDF. However, `MCFM` can only provide theory predictions for symmetric collisions and is unable to match the pPb W/Z data; with such a reliance on `MCFM` to calculate the theory for `APPLgrid`, it crucial to be able to validate that `MCFM` is providing reasonable predictions. However, by combining `FEWZ` and `MCFM` there is now a feasible series of steps through which one can take a data set and produce PDF independent gridded theory predictions. This series of steps is illustrated in Fig. 5.11 with the numbers in the figure corresponding to the following steps:

1. pPb W/Z production data from the LHC was compared to a modified version of `FEWZ` as part of the PDF reweighting analysis (Section 3.2).

---

[22]This matrix multiplication process is known as "convoluting" a grid with a PDF and is provided by an internal `APPLgrid` utility (`applgrid-convolute`). Clarification is made here to avoid confusion later.

[23]This interface is known as `mcfm-bridge` and can be found at: https://applgrid.hepforge.org/news.htm. Unfortunately, support of this interface is incompatible with `MCFM` releases past Version 6.8 as `MCFM` introduced support for `OpenMP` in Version 7.0 [144]. `APPLgrid` was not written to be "threadsafe" and therefore is unable to be multi-threaded.

2. Once `FEWZ`, run in asymmetric mode, is matched to a data set, it is re-run symmetrically with the same settings. This allows the cuts and binning that match pPb data to be ported into pp (Section 3.2).

3. The symmetric run from `FEWZ` can then be compared to the results of `MCFM` (Section 5.3.2).

4. `MCFM` is then run to produce `APPLgrid` grids by way of the `mcfm-bridge`. Using Mane-Frame[24] to run several instances of `MCFM` with different Monte Carlo seeds in parallel, multiple replicas of each grid are generated. Each individual replica can be convoluted with the PDF used in `MCFM` and matched to the corresponding instance of `MCFM` (Section 5.3.3).

5. The `applgrid-combine` utility, provided in `APPLgrid`, then compresses the multiple replica grids into a single combined grid. This grid no longer compares to any one individual instance of `MCFM`. If enough replicas were combined, the combined grid will be independent of the underlying PDF (Section 5.3.4).

6. The combined grid can then be convoluted with asymmetric PDFs to compare to `FEWZ` and completing the validation (Section 5.3.5).

7. In `nCTEQ++`, theory predictions corresponding to each W/Z production data set are provided by the combined grid associated with that set (Section 5.4).

The following sections will contain more detail about the individual steps as well as provide information on the validations and cross checks performed at each step.

5.3.2. `FEWZ` and `MCFM`

The first link in the chain, matching data to `FEWZ` and the development of asymmetric `FEWZ`, was described in Chapter 3. Thus I'll begin with comparing the results of a symmetric

---

[24]ManeFrame refers to the Southern Methodist University high-performance computing cluster. More information can be found here: https://www.smu.edu/Academics/CSC

run of `FEWZ` to an equivalent run in `MCFM`. Since both codes have been benchmarked [145,146], this simply provided verification that the results for specific processes could be replicated.

First `FEWZ` was run with identical input as that of the asymmetric pPb run except with symmetric PDFs instead (pp or PbPb, it does not matter which). From there `MCFM` was run with the same cuts, binning and initial settings as that of the symmetric `FEWZ` run and the results were compared.

Figures 5.12-5.15, contain the comparisons run for each of the LHC data sets that would eventually be included in the new fits[25]. With differences on the order of $\sim 1\%$ (near the numeric precision of the Monte Carlo used in the runs), it proved safe to conclude that indeed `MCFM` could replicate the results from symmetric `FEWZ`. Additional comparisons can be found in Appendix C.1.

5.3.3. `MCFM` with ApplGrid

Having confirmed that `MCFM` can replicate the results from `FEWZ`, it is now possible to introduce `APPLgrid` into the procedure. First, the Monte Carlo events generated for the process of interest are generated in `MCFM`. These events are then passed to the `mcfm-bridge`, where the cuts and binning that match those in the original data set are applied. The resulting events, now encoded with information specific to both the process and the individual data set, are stored in grids based on the observables specified for the given process. These observables are set by the user but for this case are primarily rapidity and transverse momentum distributions of the leptonic daughters of the W/Z bosons.

In an isolated run, the grids can then be convoluted with the PDF used to generate them to reproduce the output from the original `MCFM` run[26], as shown in Fig. 5.16. However, the utility of the grids is not limited to simply recreating `MCFM` output. By creating `APPLgrid` grids from several different runs of `MCFM`, with different Monte Carlo seeds, and combining

---

[25]As a reminder, these comparisons do not match the data directly; instead the results represent a pseudo-experiment for symmetric beams at the energies seen in pPb collisions.

[26]Once the grids are combined, the resulting grid no longer compares to a specific instance of `MCFM`. Instead the combined grid contains an functional average result of all the replicas but with enhanced numerical statistics.

them, it is possible to eliminate the dependence of the PDF used to initiate the `MCFM` events [147]. Once this PDF independence is eliminated, it is then possible to use the grids as theory predictions while fitting.

5.3.4. Determining PDF Independence of Produced `APPLgrid` Grids

In order to quantify the threshold for which it is possible to claim PDF independence for the `APPLgrid` grids, I performed several test cases in which different numbers of grids were combined, convoluted and compared to each other. Figure 5.17 details some of the main results from this study; a thorough compendium of these results can be found in Appendix D.

In this study, two sets of ten grids were produced with the same `MCFM` inputs save for the underlying PDF. I then produced combined grids from the results for each PDF; one replica grid was left uncombined, one grid was produced from three replicas and one was a combination of all ten replicas. The resulting grids were then convoluted with a third, distinct, PDF to prevent any enhancement of the underlying PDF.

As shown in Fig. 5.17d, once 10 replica grids are combined into a single grid, the dependence on the underlying PDF drops to within $\pm 0.02\%$, which is on the order of the numerical noise from the integration in `MCFM` ($\pm .01\%$ for the conditions used in these tests)[27]. Although this exercise shows that a set of 10 replica grids would be sufficient to ensure PDF independence to the precision needed, for the grids used as theory predictions for fitting, 15 replica grids were combined to further ensure PDF independence.

5.3.5. `FEWZ` and `APPLgrid`

Once established that the grids produced are independent of the underlying PDF, they can be convoluted with PDFs to make theory predictions. In Fig. 5.18, the prediction from `FEWZ`, seen earlier in 5.12, was shown to match these convoluted grids with similar precision to that of `MCFM`.

---

[27]Note: 10 grids were also found to be sufficient to obtain high numerical precision in [147]

Finally, we reach the end of the chain of tools as we are able to make theory predictions for data sets from asymmetric collisions using the `APPLgrid` grids produced from symmetric `MCFM`. Figure 5.19 shows this comparison, including the prediction from `FEWZ` as reference. Now that it is possible to grid the theory predictions for the pPb LHC data we would like to include in the new fits with `nCTEQ++`, I will return to where I left off at the end Section 5.1.
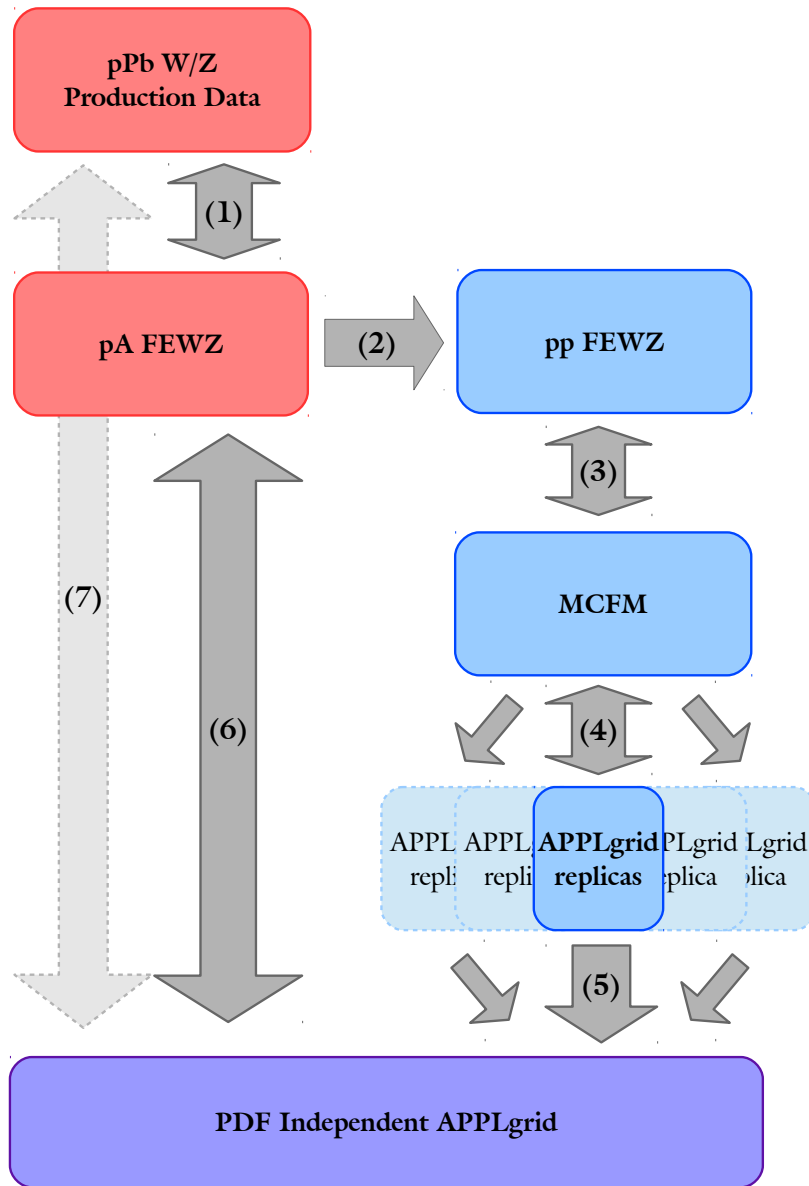
Figure 5.11: A schematic representation of how `APPLgrid` PDF independent grids are produced and validated against `FEWZ`, `MCFM` and experimental data. These grids go on to provide the input theory in `nCTEQ++` for the W/Z LHC data included in the new fits.

In this graphic there are several features: double arrows ($\updownarrow$) represent comparisons and validations that have been made, single arrows ($\Rightarrow$) represent a process for which the inputs cannot be directly compared to the outputs, **red** boxes refer to processes that utilize asymmetric inputs (specifically pPb), **blue** boxes refer to processes that utilize only symmetric inputs (e.g. pp), **purple** box refers to a process that is independent of the choice of PDF and finally each numbered arrow refers to a description in the text in a subsection of Section 5.3.1.
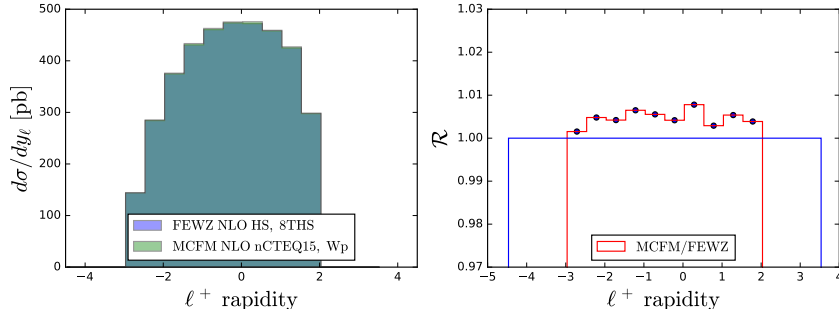
Figure 5.12: Comparison of `MCFM` and `FEWZ` for differential cross section as a function of lepton rapidity for W^+ production at NLO in pPb collisions from CMS [42].
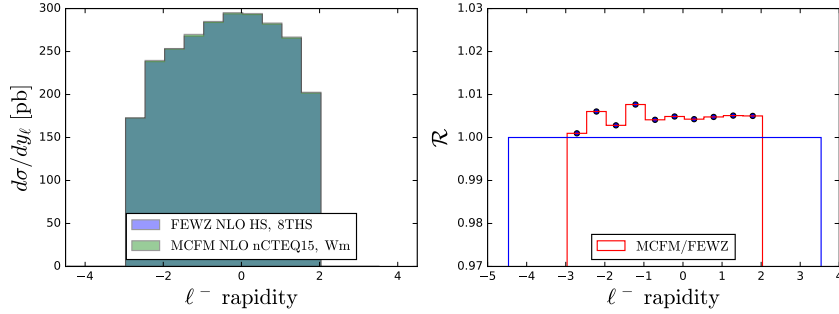
Figure 5.13: Comparison of `MCFM` and `FEWZ` for differential cross section as a function of lepton rapidity for W^- production at NLO in pPb collisions from CMS [42].
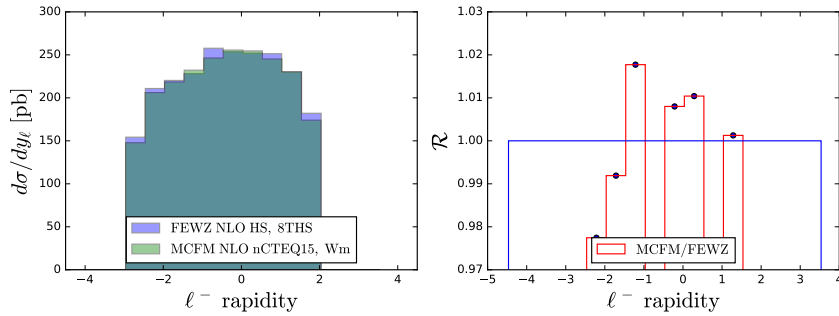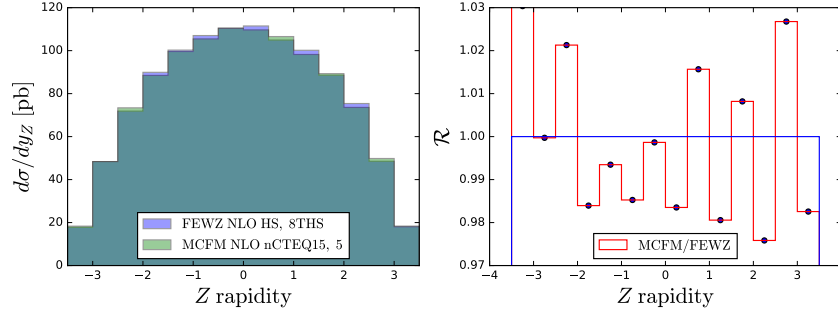

Figure 5.14: Comparison of `MCFM` and `FEWZ` for differential cross section as a function of lepton rapidity for W^- production at NLO in pPb collisions from ATLAS [40]. The larger differences seen here are due to the difficulty of implementing varied the bin widths used in this data set in `MCFM`. This is solved when generating the grids as `mcfm-bridge` provides fully customizable binning.

160

Figure 5.15: Comparison of `MCFM` and `FEWZ` for differential cross section as a function of lepton rapidity for for Z production at NLO in pPb collisions from ATLAS [39].
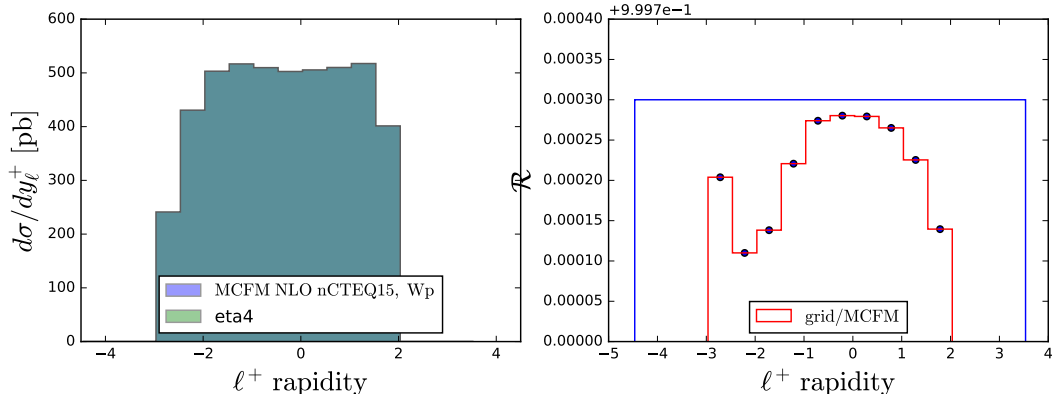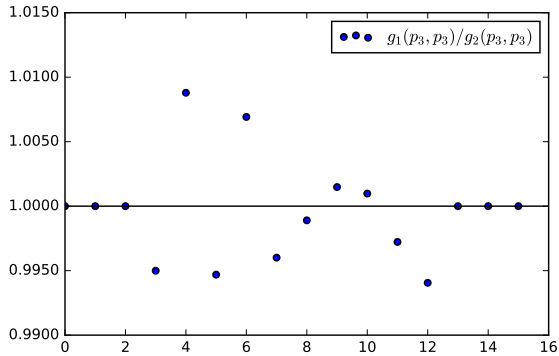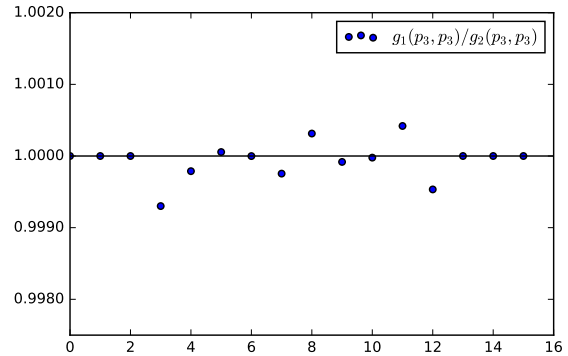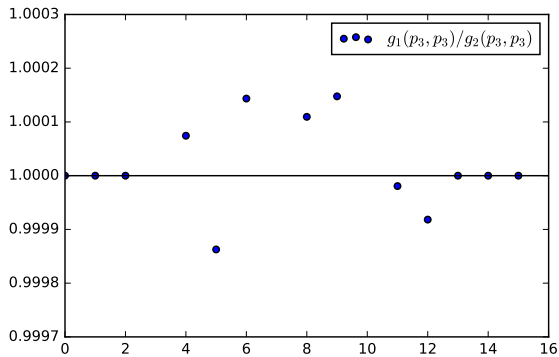

Figure 5.16: `MCFM` output histogram compared to the `APPLgrid` grid produced with the same settings and convoluted with the same PDFs used in the original `MCFM` process. The label "eta4" refers to the grid associated with the rapidity of the charged lepton produced in the decay of the W^+ boson.
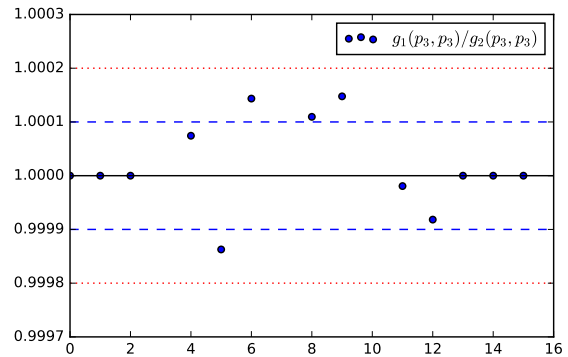
(a) Uncombined Single Grid

(b) Combined - 3 Grids

(c) Combined - 10 Grids

(d)

Figure 5.17: Presented here are three results of the dependence of the `APPLgrid` grids on the underlying PDFs used in the generation of the `MCFM` events. The dependence is shown as a ratio of convoluted grid values for each bin. If the grids were infinitely precise and independent, one would expect all the points to fall on the horizontal line at 1.0, however the numerical precision of the Monte Carlo limits the measure of independence to $\pm0.01\%$. The label "g_k" refers to the different set of grids, while the PDF convoluted with the grid is labled "p_3". 5.17a contains a single uncombined grid and displays a dependence on the order of $\pm1\%$. 5.17b contains a a grid resulting from the combination of 3 grids and displays a dependence on the order of $\pm0.1\%$. 5.17c contains a a grid resulting from the combination of 10 grids and displays a dependence on the order of $\pm0.01\%$. This is shown explicity in 5.17d where the dashed line $= 0.01\%$ and the dotted line $= 0.02\%$. Additional comparisons including convolutions with different PDFs can be found in Appendix D.
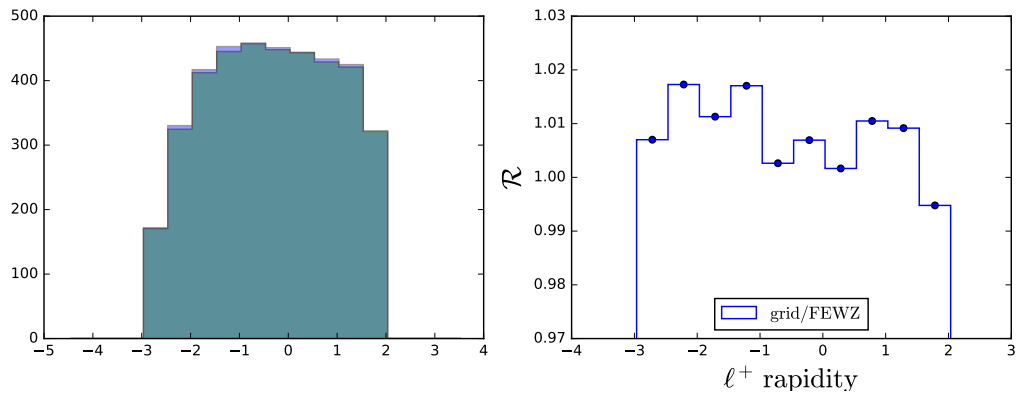
Figure 5.18: Theory predictions from `FEWZ` compared to the `APPLgrid` grid convoluted with the same pPb PDFs used in the original `FEWZ` run. The convoluted grid prediciton in purple is only slightly visible above the `FEWZ` prediction in green.
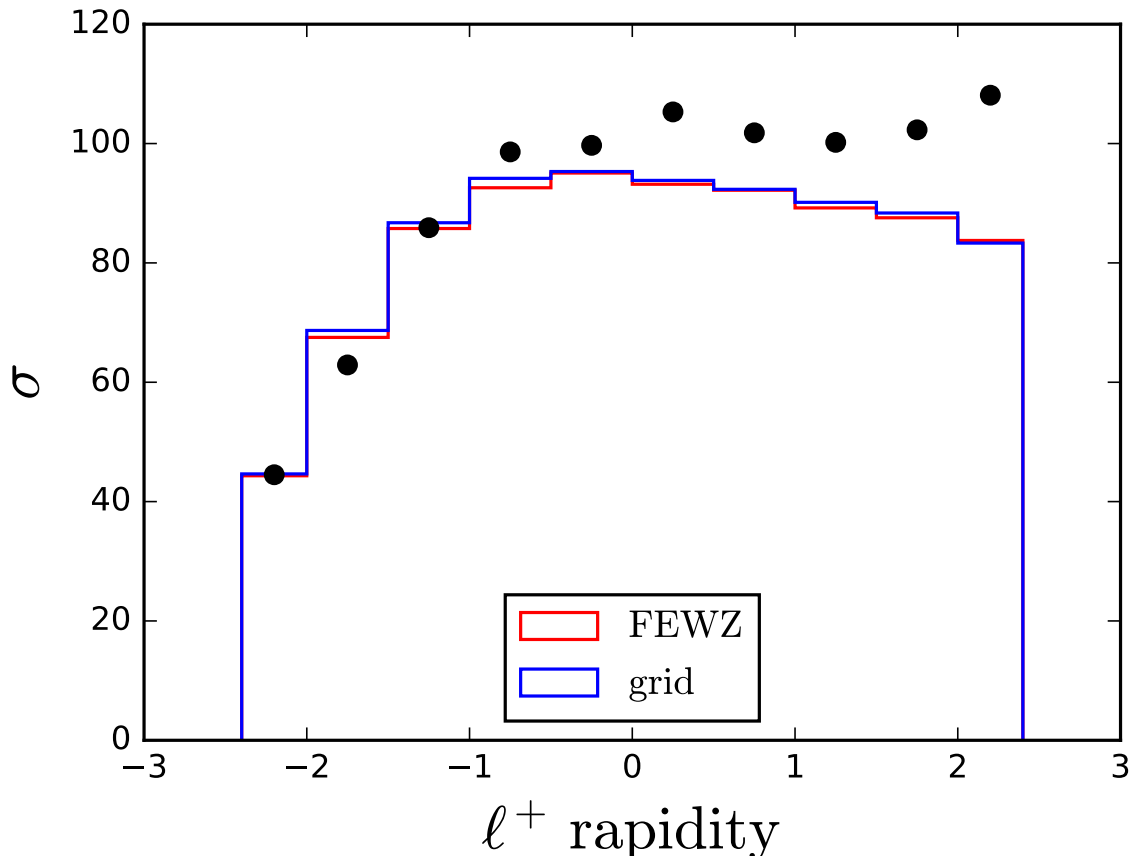

Figure 5.19: Theory predictions from `FEWZ`, a convoluted `APPLgrid` grid produced with by combining replica grids representing `MCFM` events and the associated data set (**Black** points), W^+ production from CMS [42], all shown in one comparison.

5.4. nCTEQ15 Fit with LHC data

The goal of including data from LHC experiments into a `nCTEQ` fit is the driving force behind the developments discussed in Section 5.3. As was found in Chapter 3, the disagreement between the experimental data and `nCTEQ15` was large enough that even a reweighted PDF was not able to bridge the gap. This meant that refitting the parameters from `nCTEQ15` would be necessary in order to get better agreement with the data.

In this section, I will detail two fits performed with `nCTEQ++`. Both fits use `nCTEQ15-np` as a baseline, meaning the parameters the were fit started at the values noted in Table 5.4, and include the data sets discussed in Section 5.4.1. The first fit, called `nCTEQ+LHC`, uses a similar set of parameters to `nCTEQ15-np` and opens no new parameters. The second fit, `nCTEQ+LHCs`, includes all the parameters opened in the first fit but also opens three strange parameters.

5.4.1. Data Sets Included

Data from the LHC affords the new fits the opportunity to probe a kinematic regime that lacked constraint in past `nCTEQ` fits. In Figure 5.20, it becomes clear that the LHC data is vastly separated in (x, Q^2) space from any of the previously included DIS or DY data. Relating this to rapidity in Figure 5.21, this new data allows us to constrain our cross section predictions in rapidity regions where no prior constraint existed, e.g. for pPb W^+ production, any prediction for positive rapidities was merely extrapolation in prior fits. This reliance on extrapolation gives some indication as to the disagreement seen between `nCTEQ15` and the data in this region.

The 4 data sets included in `nCTEQ+LHC` come from pPb collisions and represent both the ATLAS and CMS experiments. They are Z production in ATLAS, W^+ production in CMS and W^- production in both ATLAS and CMS. A detailed list of the data, with citations, can be found in Table 5.5. In total, the 708 DIS and DY data points included in `nCTEQ15-np`
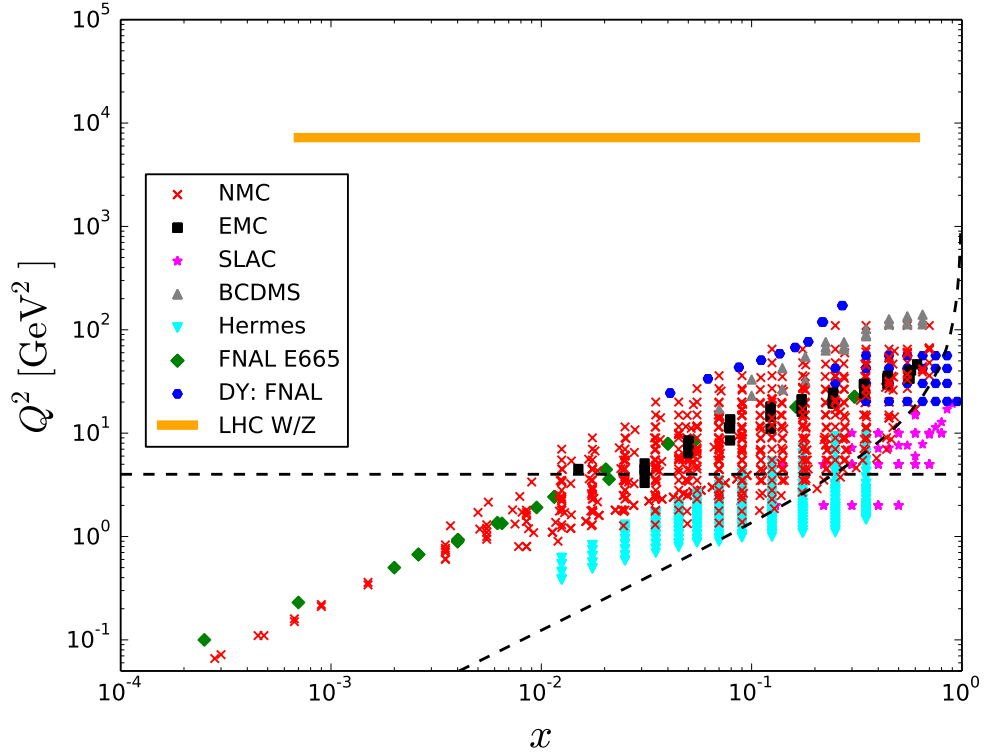
164

Figure 5.20: Here the W/Z production data from the LHC is added kinematic range of the data shown in Fig. 5.1. This data falls well above the kinematic cuts from `nCTEQ15` and includes the only data fit above Q^2 of ~ 100 GeV^2 or below $x = 10^{-2}$. This illustrates the complete kinematic range covered in the new fits described in section 5.4.2 and 5.4.3

are supplemented by 44 additional data points[28], yielding a total of 752 data points in the new fits.

---

[28]Note: 14 data points for Z production and 30 data points for $W^{\pm}$ production

Observable	Experiment	ID	Ref.	# data	after cuts
$d\sigma(W^- \to \ell^- \bar{\nu})/dy_{\ell^-}$	ATLAS	6211	[40]	10	10
$d\sigma(W^- \to \ell^- \bar{\nu})/dy_{\ell^-}$	CMS	6231	[42]	10	10
$d\sigma(W^+ \to \ell^+ \nu)/dy_{\ell^+}$	CMS	6233	[42]	10	10
$d\sigma(Z \to \ell^+ \ell^-)/dy_Z$	ATLAS	6215	[39]	14	14
Total:				44	44

Table 5.5: The W/Z production data used in the `nCTEQ15+LHC` fits. The same details for each data set as in Tab. 5.1 are listed here with the exception of the χ^2, which is presented in the sections pertaining to each fit.
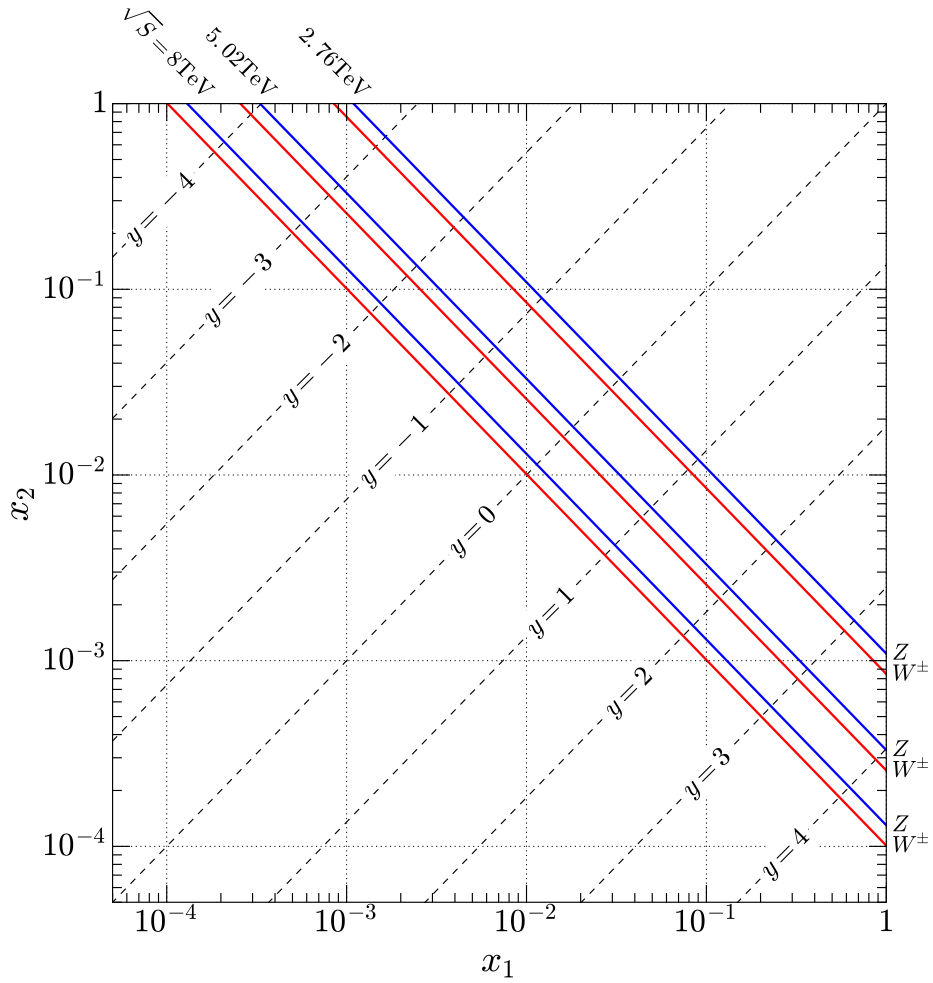
Figure 5.21: The kinematic (x_1, x_2) space explored by the experimental data used included in the new fits. The red and blue lines represent constant $\tau = M_V/\sqrt{s}$ where M_V is the invariant mass of the produced $W^\pm/Z$ vector boson (respectively), as well as the center of mass rapidity y. Here x_1 corresponds to the proton and x_2 to the Pb momentum fraction.

5.4.2. `nCTEQ+LHC`

The first of the new fits, `nCTEQ+LHC`, aimed to simply update `nCTEQ15-np` to include LHC data. This fit opened parameters for the valence quarks, the $\bar{u} + \bar{d}$, and the gluon. All of these parameters had been opened in the original `nCTEQ15-np` fit so they were natural candidates to reopen for this fit. The values of the parameters all originated at the values from `nCTEQ15-np` and the final values after the fit are shown in Table 5.6. Since all the LHC data that was added was from pPb experiments, it was only prudent to open either a_k or b_k parameters in the fit. As seen in the relationship shown here:

$$c_k(A) = p_k + a_k \left(1 - A^{-b_k}\right) \quad , \tag{5.21}$$

the coefficients used in the parameterization are dependent on A but when only adding additional data for a single nucleus, it is very difficult to discriminate whether changes to the a_k and b_k parameters are a result of the inclusion of the data or feedback in the fitting procedure. As such, only a_k parameters were opened in `nCTEQ+LHC`.

Figure 5.22 shows how the parameters in `nCTEQ+LHC` change as a function of A. For most of the parameters, the value of the parameter changes rapidly for small A before eventually plateauing for large A. This is very similar to the result from `nCTEQ15` [7] and seemingly indicates that the parameterization for the non-gluon parameters might be suboptimal at describing the dependence on A. It is also clear that the parameters corresponding to the behavior at small x, c_1, are more strongly correlated with A than those that describe behavior at large x. The gluon parameters especially exhibit this behavior as the small x parameters linearly rise in A throughout the entire range.

As the parameterization and subsequently the parameters are dependent on A, it is important to examine the PDFs resulting from the fit at different values of A. In Fig. 5.23, you can see that as A increases, the value of the PDF is suppressed at low x and enhanced at large x. This result is similar to that of `nCTEQ15-np` (Fig. 5.3) except that the spread of

the different PDFs is much smaller for this fit. This is directly a result of the shift of the shadowing-anti-shadowing point to lower x.

Since the goal was to study the impact of the LHC data on `nCTEQ15-np`, Figure 5.29 compares the bound lead PDF for several flavors fit in `nCTEQ+LHC` to the equivalent PDF from `nCTEQ15-np` at 10 GeV. In this comparison, the fit for Pb shifts the shadowing-anti-shadowing point to a lower x than was seen in the `nCTEQ15-np` fit and lends itself to a larger PDF value in the low x region. This shift at low x is like due to the constraint in this kinematic region offered by the LHC data.

For `nCTEQ+LHC`, the total χ^2 was found to be 770.081 for 752 data points. This gives a χ^2 per degree of freedom of 1.03645 ($770.081/(752-9$ free parameters$)$). While this indicates a good fit, Fig. 5.25 shows that the χ^2 per degree of freedom for the LHC sets is rather large[29]. This indicated that additional parameters might need to be opened in order to accommodate this data.

5.4.2.1. Parameter Scans for *nCTEQ+LHC*

From Figure 5.26, it is very clear that, for the 9 parameters fit in `nCTEQ+LHC`, each one is in a minimum. This is important for the stability of the fit; it is possible that a parameter could fall into a small local minimum, tricking the minimizer into not looking for a larger global minimum. The shape of the parameter space is also revealed by these scans and, for the most part, it is very smooth. There are some slight numerical noise in the gluon parameters but as it increases the χ^2 as opposed to lowering it, these are tolerable.

---

[29]A thorough examination of the χ^2 for each data set including the LHC sets can be found in Appendix E

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
p_0^g	(0.38232)	−	−	−	−	−	−	$p_0^{s+\bar{s}}$	(0.500)	−	−
p_1^g	(0.38232)	$p_1^{u_v}$	(0.630)	$p_1^{d_v}$	(0.513)	$p_1^{\bar{d}+\bar{u}}$	(-0.324)	$p_1^{s+\bar{s}}$	(-0.324)	$p_1^{\bar{d}/\bar{u}}$	(10.075)
p_2^g	(3.03369)	$p_2^{u_v}$	(2.934)	$p_2^{d_v}$	(4.211)	$p_2^{\bar{d}+\bar{u}}$	(8.116)	$p_2^{s+\bar{s}}$	(8.116)	$p_2^{\bar{d}/\bar{u}}$	(4.957)
p_3^g	(4.39448)	$p_3^{u_v}$	(-2.369)	$p_3^{d_v}$	(-2.375)	$p_3^{\bar{d}+\bar{u}}$	(0.413)	$p_3^{s+\bar{s}}$	(0.413)	$p_3^{\bar{d}/\bar{u}}$	(15.167)
p_4^g	(2.35917)	$p_4^{u_v}$	(1.266)	$p_4^{d_v}$	(0.965)	$p_4^{\bar{d}+\bar{u}}$	(4.754)	$p_4^{s+\bar{s}}$	(4.754)	$p_4^{\bar{d}/\bar{u}}$	(17.000)
p_5^g	(-3.000)	$p_5^{u_v}$	(1.718)	$p_5^{d_v}$	(3.000)	$p_5^{\bar{d}+\bar{u}}$	(0.614)	$p_5^{s+\bar{s}}$	(0.614)	$p_5^{\bar{d}/\bar{u}}$	(9.948)

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
a_0^g	(-0.256)	−	−	−	−	−	−	$a_0^{s+\bar{s}}$	(0.167)	−	−
a_1^g	**-0.00624**	$a_1^{u_v}$	**-3.72300**	$a_1^{d_v}$	**0.29370**	$a_1^{\bar{d}+\bar{u}}$	**0.62475**	$a_1^{s+\bar{s}}$	(0.406)	$a_1^{\bar{d}/\bar{u}}$	(0.000)
a_2^g	(0.000)	$a_2^{u_v}$	**-0.19173**	$a_2^{d_v}$	**-0.30716**	$a_2^{\bar{d}+\bar{u}}$	(0.415)	$a_2^{s+\bar{s}}$	(0.415)	$a_2^{\bar{d}/\bar{u}}$	(0.000)
a_3^g	(0.383)	$a_3^{u_v}$	(0.018)	$a_3^{d_v}$	(0.085)	$a_3^{\bar{d}+\bar{u}}$	(-0.759)	$a_3^{s+\bar{s}}$	(0.000)	$a_3^{\bar{d}/\bar{u}}$	(0.000)
a_4^g	**-0.09693**	$a_4^{u_v}$	(12.083)	$a_4^{d_v}$	(3.874)	$a_4^{\bar{d}+\bar{u}}$	(-0.203)	$a_4^{s+\bar{s}}$	(0.000)	$a_4^{\bar{d}/\bar{u}}$	(0.000)
a_5^g	**0.00494**	$a_5^{u_v}$	(-1.140)	$a_5^{d_v}$	(-0.080)	$a_5^{\bar{d}+\bar{u}}$	**-0.16552**	$a_5^{s+\bar{s}}$	(0.000)	$a_5^{\bar{d}/\bar{u}}$	(0.000)

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
b_0^g	(-0.038)	−	−	−	−	−	−	$b_0^{s+\bar{s}}$	(0.104)	−	−
b_1^g	(-1.341)	$b_1^{u_v}$	(0.006)	$b_1^{d_v}$	(0.466)	$b_1^{\bar{d}+\bar{u}}$	(0.172)	$b_1^{s+\bar{s}}$	(0.172)	$b_1^{\bar{d}/\bar{u}}$	(0.000)
b_2^g	(0.000)	$b_2^{u_v}$	(0.524)	$b_2^{d_v}$	(0.440)	$b_2^{\bar{d}+\bar{u}}$	(0.290)	$b_2^{s+\bar{s}}$	(0.290)	$b_2^{\bar{d}/\bar{u}}$	(0.000)
b_3^g	(0.520)	$b_3^{u_v}$	(0.073)	$b_3^{d_v}$	(0.107)	$b_3^{\bar{d}+\bar{u}}$	(0.298)	$b_3^{s+\bar{s}}$	(0.000)	$b_3^{\bar{d}/\bar{u}}$	(0.000)
b_4^g	(-0.516)	$b_4^{u_v}$	(0.038)	$b_4^{d_v}$	(-0.018)	$b_4^{\bar{d}+\bar{u}}$	(0.888)	$b_4^{s+\bar{s}}$	(0.000)	$b_4^{\bar{d}/\bar{u}}$	(0.000)
b_5^g	(-1.418)	$b_5^{u_v}$	(0.615)	$b_5^{d_v}$	(-0.236)	$b_5^{\bar{d}+\bar{u}}$	(1.353)	$b_5^{s+\bar{s}}$	(0.000)	$b_5^{\bar{d}/\bar{u}}$	(0.000)

Table 5.6: Values of the parameters of the nCTEQ+LHC fit at the initial scale $Q_0 = 1.3\,\text{GeV}$. Values in bold represent the free parameters and values in parentheses are fixed in the fit. The first block of parameters, the p_k's refer to the free proton parameters. The normalization parameters, determined by the sum rules are not listed.
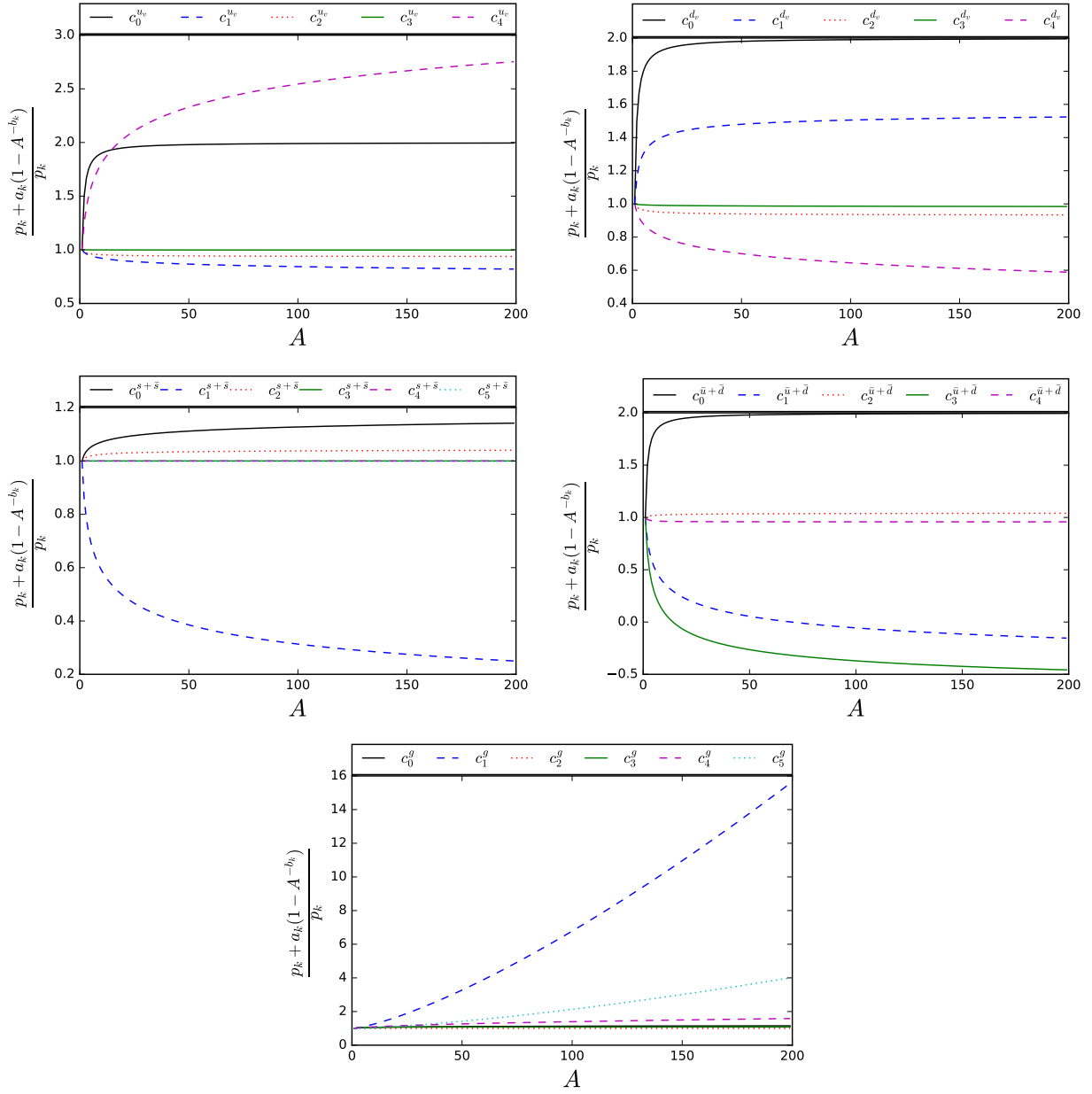
Figure 5.22: A detailed look at the A dependence of the parameters fit in `nCTEQ+LHC` as described by Eq. 5.5. As `nCTEQ+LHCs` has very similar parameters, the strange was included as well. Each of these plots have been noramlized to the free proton parameters such that only the nuclear parameters contribute. The parameters for `nCTEQ+LHC` and `nCTEQ+LHCs` can be found in 5.6 and 5.7, respectively.
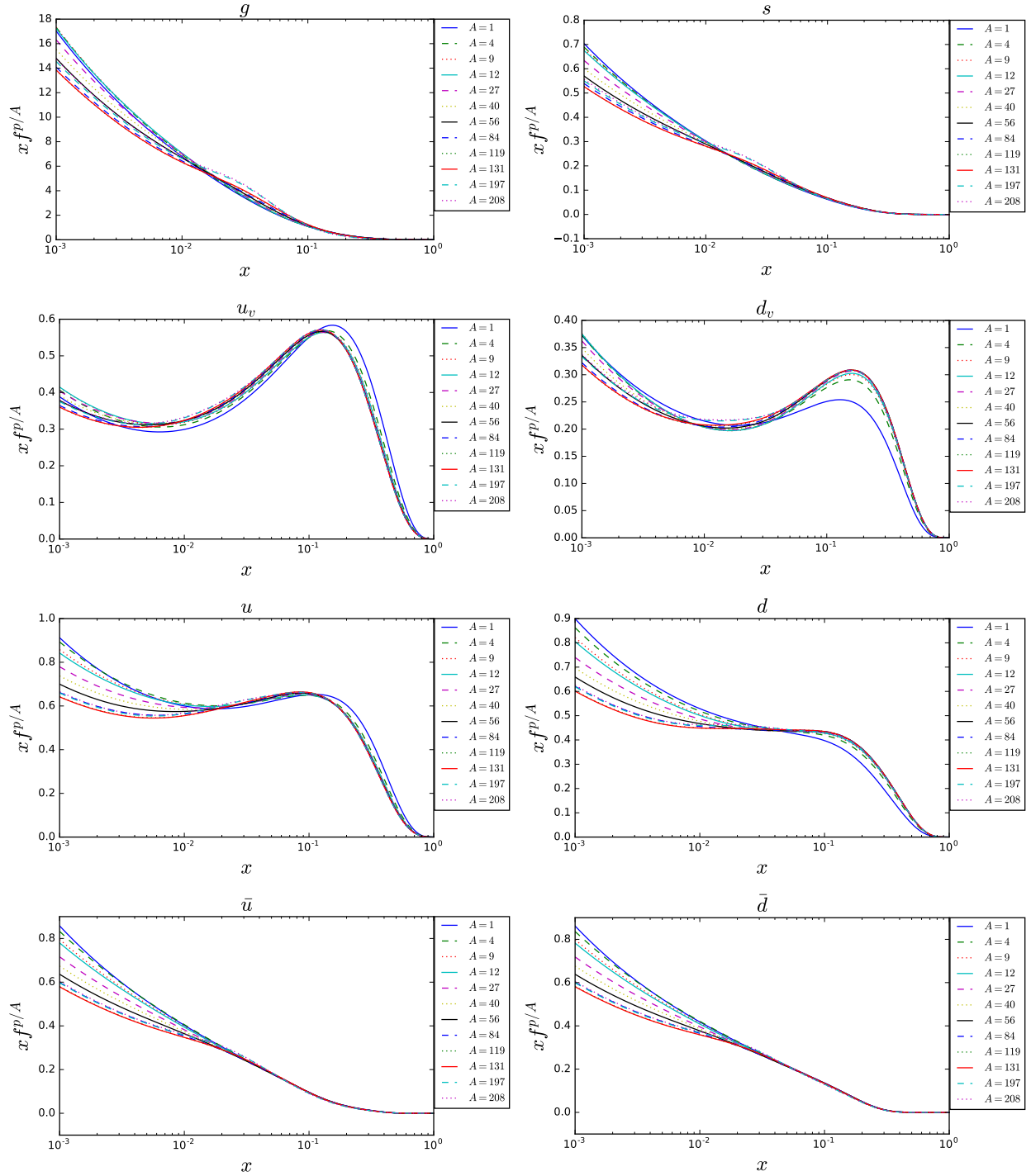
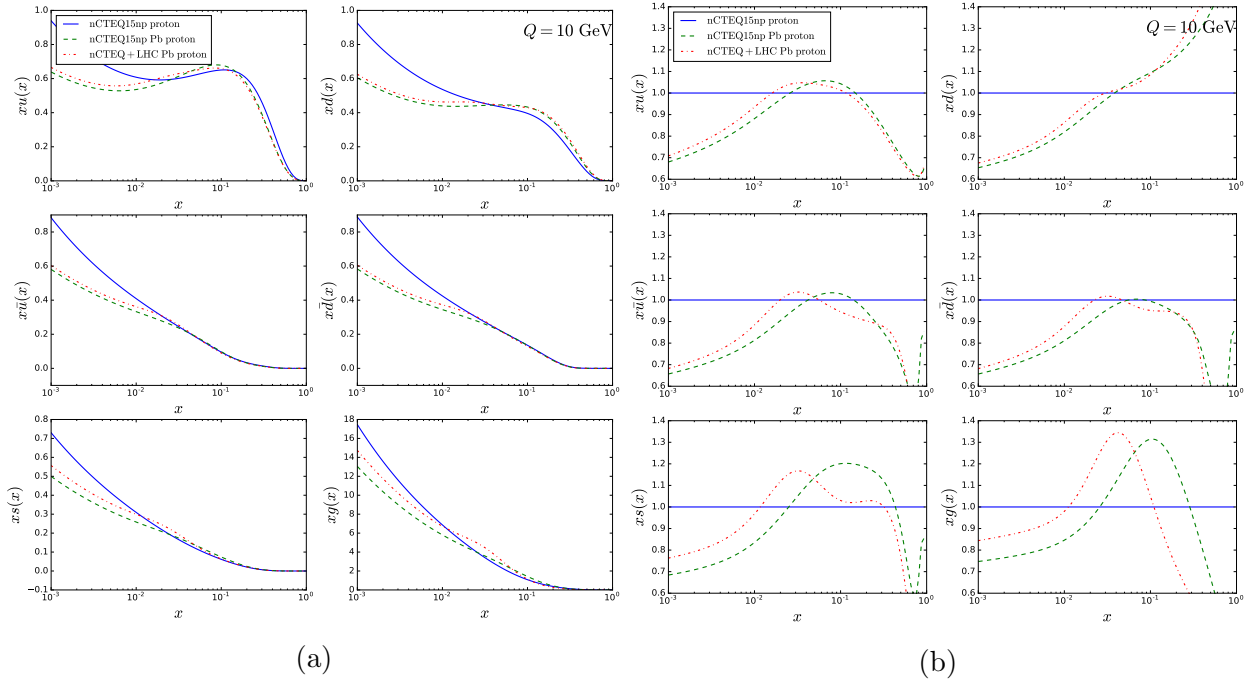Figure 5.23: nCTEQ+LHC bound proton PDFs for $Q = 10$ GeV for the free proton ($A = 1$) to lead ($A = 208$).

Figure 5.24: `nCTEQ15-np` bound proton PDF compared to the bound proton in lead for `nCTEQ+LHC` as well as the pound proton in lead for `nCTEQ15-np` 5.24a: for $Q = 10$ GeV as well as the ratio between them, 5.24b.. The proton parameters remained fixed for `nCTEQ+LHC`, so showing the proton PDF for `nCTEQ+LHC` would be redundant. The ratio plots represent the effective nuclear correction to the proton.
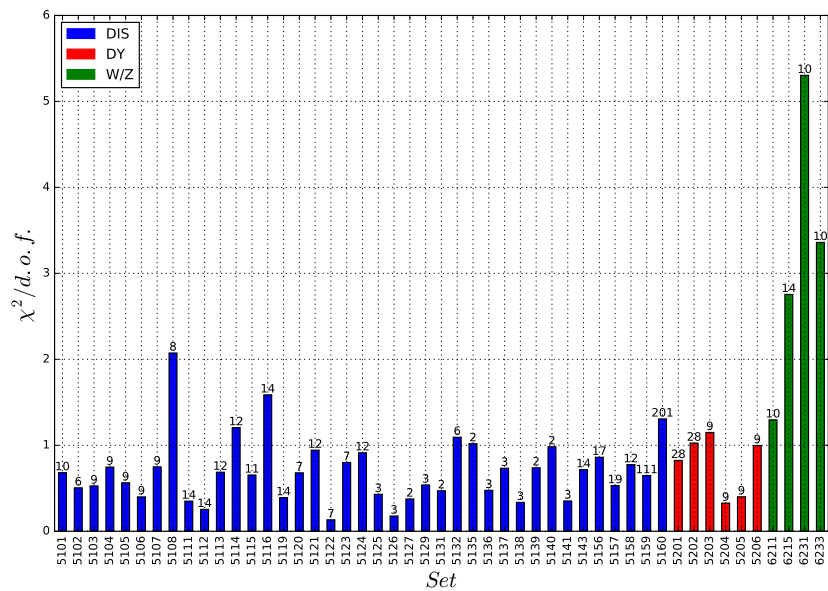
Figure 5.25: Listing of the χ^2/d.o.f. for each data set included in `nCTEQ+LHC`. The individual data sets are identified by the ID number corresponding to those in Tables 5.1-5.3 as well as the LHC data included in 5.5. The DIS data is displayed in **blue** and have ID numbers corresponding to 51XX, while DY data is displayed in **red** and have ID's corresponding to 52XX. The LHC data is displayed in **green** and have ID numbers corresponding to 62XX.
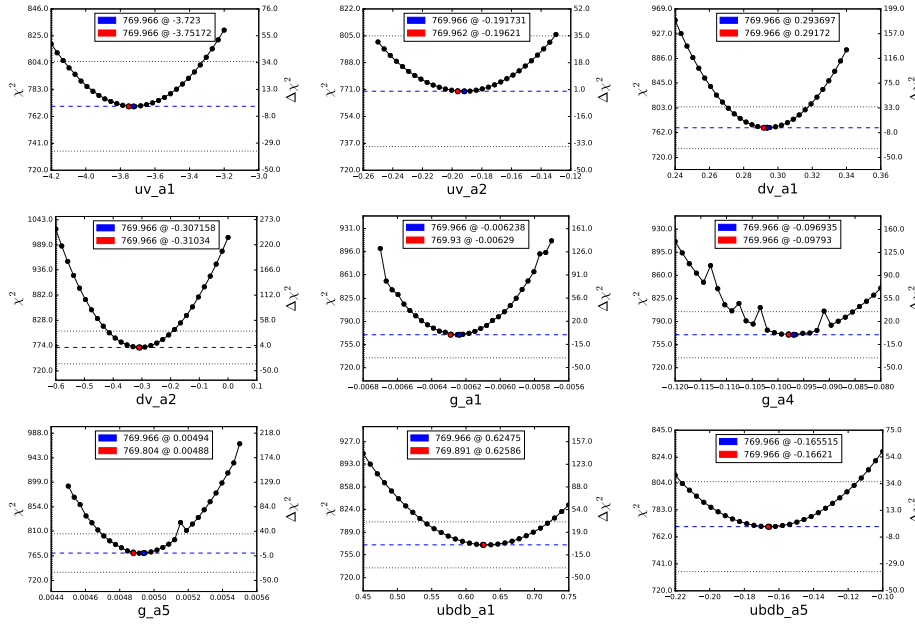
Figure 5.26: One dimensional parameter scans for the 9 free parameters fit in `nCTEQ+LHC`. Each scan was performed by opening a single parameter while the others remained fixed, eliminating any correlation between parameters. The **blue** marker indicates the value of the parameter and the associated χ^2 from the minimum previously found in the fit. The **red** marker indicates where the parameter is at a minimum in the scan as well as the χ^2 at that point, These two markers should be in essentially the same location. The dotted horizontal black lines indicate ± 35 units of χ^2.

5.4.3. nCTEQ+LHCs

(a)   (b)
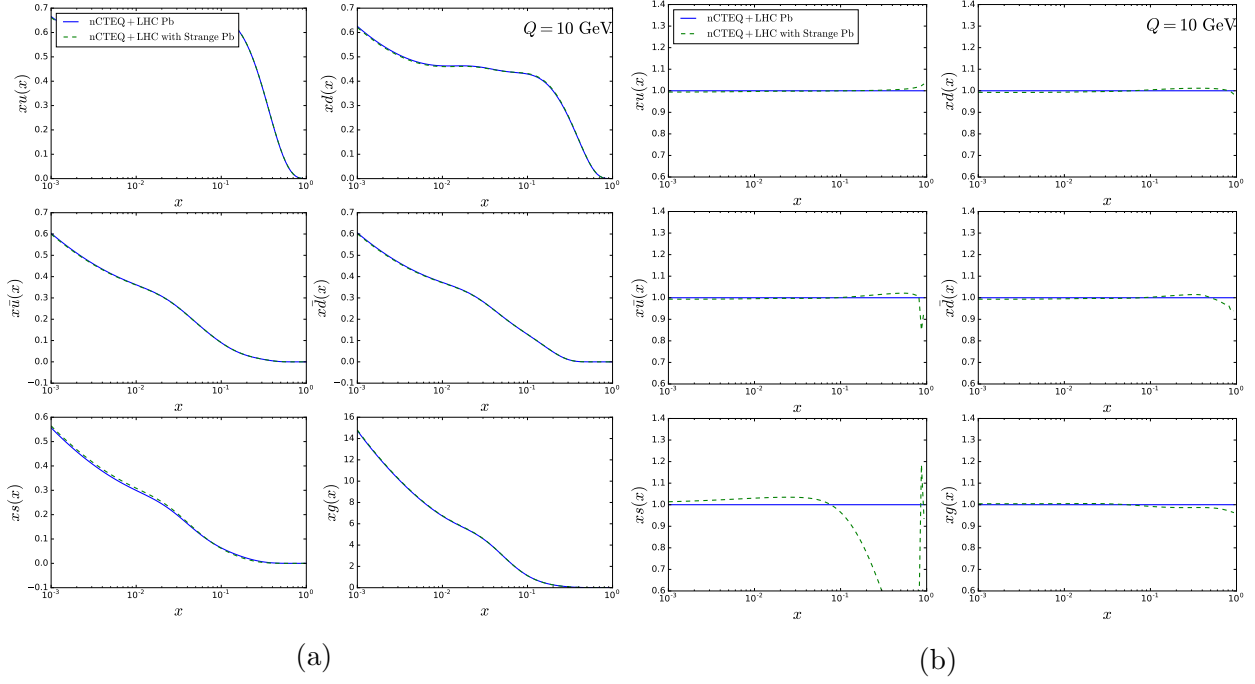
Figure 5.27: `nCTEQ+LHC` bound lead PDF compared to the bound lead PDF for `nCTEQ+LHCs`. As is clear from 5.27a, the PDFs from these two fits are very similar, with the only noticeable difference coming in the strange quark PDF. This would be somewhat expected as the only difference betweeen the two fits is the freedom of the strange quark parameters. In the ratios, 5.27b the difference in the strange distribution is more clear. Additionally, there are in fact slight changes to the other flavors as a secondary result of opening the strange parameters.

One set of parameters worth examining were those of the strange quark distribution. Since the strange asymmetry $(s - \bar{s} = 0)$ is fixed in the `nCTEQ` parameterization, this meant the only flexibility in the strange could come from $s+\bar{s}$. As such, `nCTEQ+LHCs` was determined by taking the resulting parameters from `nCTEQ+LHC`, opening three strange parameters and refitting. The reasoning for starting with the `nCTEQ+LHC` parameters for this fit was that, since those parameters represented a stable minimum for the χ^2 function, opening any additional parameters could only improve the fit rather than relying on the minimization routine to relocate the stable minimum with the additional strange parameters complicating the parameter space[30]. The parameters found to minimize this fit can be found in Table 5.7.

---

[30]A fit of this kind, identical in procedure to `nCTEQ+LHC` (starting from the `nCTEQ15-np` parameters) but with the three opened strange parameters was actually performed and the same minimum was found, albeit

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
p_0^{g}	(0.38232)	–	–	–	–	–	–	$p_0^{s+\bar{s}}$	(0.500)	–	–
p_1^{g}	(0.38232)	$p_1^{u_v}$	(0.630)	$p_1^{d_v}$	(0.513)	$p_1^{\bar{d}+\bar{u}}$	(-0.324)	$p_1^{s+\bar{s}}$	(-0.324)	$p_1^{\bar{d}/\bar{u}}$	(10.075)
p_2^{g}	(3.03369)	$p_2^{u_v}$	(2.934)	$p_2^{d_v}$	(4.211)	$p_2^{\bar{d}+\bar{u}}$	(8.116)	$p_2^{s+\bar{s}}$	(8.116)	$p_2^{\bar{d}/\bar{u}}$	(4.957)
p_3^{g}	(4.39448)	$p_3^{u_v}$	(-2.369)	$p_3^{d_v}$	(-2.375)	$p_3^{\bar{d}+\bar{u}}$	(0.413)	$p_3^{s+\bar{s}}$	(0.413)	$p_3^{\bar{d}/\bar{u}}$	(15.167)
p_4^{g}	(2.35917)	$p_4^{u_v}$	(1.266)	$p_4^{d_v}$	(0.965)	$p_4^{\bar{d}+\bar{u}}$	(4.754)	$p_4^{s+\bar{s}}$	(4.754)	$p_4^{\bar{d}/\bar{u}}$	(17.000)
p_5^{g}	(-3.000)	$p_5^{u_v}$	(1.718)	$p_5^{d_v}$	(3.000)	$p_5^{\bar{d}+\bar{u}}$	(0.614)	$p_5^{s+\bar{s}}$	(0.614)	$p_5^{\bar{d}/\bar{u}}$	(9.948)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
a_0^{g}	(-0.256)	–	–	–	–	–	–	$a_0^{s+\bar{s}}$	**0.08526**	–	–
a_1^{g}	**-0.00625**	$a_1^{u_v}$	**-3.74399**	$a_1^{d_v}$	**0.30672**	$a_1^{\bar{d}+\bar{u}}$	**0.67121**	$a_1^{s+\bar{s}}$	**0.45734**	$a_1^{\bar{d}/\bar{u}}$	(0.000)
a_2^{g}	(0.000)	$a_2^{u_v}$	**-0.20501**	$a_2^{d_v}$	**-0.28874**	$a_2^{\bar{d}+\bar{u}}$	(0.415)	$a_2^{s+\bar{s}}$	**3.30345**	$a_2^{\bar{d}/\bar{u}}$	(0.000)
a_3^{g}	(0.383)	$a_3^{u_v}$	(0.018)	$a_3^{d_v}$	(0.085)	$a_3^{\bar{d}+\bar{u}}$	(-0.759)	$a_3^{s+\bar{s}}$	(0.000)	$a_3^{\bar{d}/\bar{u}}$	(0.000)
a_4^{g}	**-0.09695**	$a_4^{u_v}$	(12.083)	$a_4^{d_v}$	(3.874)	$a_4^{\bar{d}+\bar{u}}$	(-0.203)	$a_4^{s+\bar{s}}$	(0.000)	$a_4^{\bar{d}/\bar{u}}$	(0.000)
a_5^{g}	**0.00497**	$a_5^{u_v}$	(-1.140)	$a_5^{d_v}$	(-0.080)	$a_5^{\bar{d}+\bar{u}}$	**-0.18145**	$a_5^{s+\bar{s}}$	(0.000)	$a_5^{\bar{d}/\bar{u}}$	(0.000)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
b_0^{g}	(-0.038)	–	–	–	–	–	–	$b_0^{s+\bar{s}}$	(0.104)	–	–
b_1^{g}	(-1.341)	$b_1^{u_v}$	(0.006)	$b_1^{d_v}$	(0.466)	$b_1^{\bar{d}+\bar{u}}$	(0.172)	$b_1^{s+\bar{s}}$	(0.172)	$b_1^{\bar{d}/\bar{u}}$	(0.000)
b_2^{g}	(0.000)	$b_2^{u_v}$	(0.524)	$b_2^{d_v}$	(0.440)	$b_2^{\bar{d}+\bar{u}}$	(0.290)	$b_2^{s+\bar{s}}$	(0.290)	$b_2^{\bar{d}/\bar{u}}$	(0.000)
b_3^{g}	(0.520)	$b_3^{u_v}$	(0.073)	$b_3^{d_v}$	(0.107)	$b_3^{\bar{d}+\bar{u}}$	(0.298)	$b_3^{s+\bar{s}}$	(0.000)	$b_3^{\bar{d}/\bar{u}}$	(0.000)
b_4^{g}	(-0.516)	$b_4^{u_v}$	(0.038)	$b_4^{d_v}$	(-0.018)	$b_4^{\bar{d}+\bar{u}}$	(0.888)	$b_4^{s+\bar{s}}$	(0.000)	$b_4^{\bar{d}/\bar{u}}$	(0.000)
b_5^{g}	(-1.418)	$b_5^{u_v}$	(0.615)	$b_5^{d_v}$	(-0.236)	$b_5^{\bar{d}+\bar{u}}$	(1.353)	$b_5^{s+\bar{s}}$	(0.000)	$b_5^{\bar{d}/\bar{u}}$	(0.000)

Table 5.7: Values of the parameters of the `nCTEQ+LHCs` fit at the initial scale $Q_0 = 1.3\text{GeV}$. Values in bold represent the free parameters and values in parentheses are fixed in the fit. The first block of parameters, the p_k's refer to the free proton parameters. The normalization parameters, determined by the sum rules are not listed.

As seen in Fig. 5.27, opening the strange parameters actually had little to no effect on `nCTEQ+LHC`, with only the strange quark showing any noticeable variation (which would be expected when opening those parameters). This similarity is mirrored in the χ^2 for the fit, which was found to be 766.990 for 752 data points or a χ^2 per degree of freedom of 1.03647 ($766.990/(752-12$ free parameters$)$). This represents a percent difference in total χ^2 between the two fits of 0.4% and a 0.002% difference in χ^2 per degree of freedom.

with slight discrepancies on the order of the numerical precision in χ^2. As the results from that fit are equivalent to the results from `nCTEQ+LHCs`, only results from `nCTEQ+LHCs` will be displayed here.

This minute difference between the two fits indicates that opening the strange parameters had a negligible effect when describing the data, despite the LHC data being more sensitive to the strange than any data previously included in `nCTEQ15`. It also indicates that some additional parameters could be introduced, possibly the $s - \bar{s}$ or a fit normalization (Section 5.4.5), that might have a stronger impact on the χ^2. However, for completeness, the results of `nCTEQ+LHCs` can be found in Fig. 5.28-5.30.

5.4.3.1. *Parameter Scans for `nCTEQ+LHCs`*

Once again scanning the parameter space gives some insight into the behavior of the fit, shown in Fig. 5.31. For the 9 non-strange parameters, they remain well behaved and sit in a global minimum, just as they did in `nCTEQ+LHC`. However the 3 strange parameters are another story. The a_0 parameter is flat for a large span of the parameter space. This is especially difficult for `Minuit` to minimize and can lead to large uncertainties in the fit parameters and instability in the fit itself. The a_1 and a_2 parameters are also fairly flat, with the scan needing to cover a large range in order to reveal the parabolic shape of the parameter space. This further justifies the result that `nCTEQ+LHCs` does not have the necessary constraint to provide any real insight into the strange distribution.
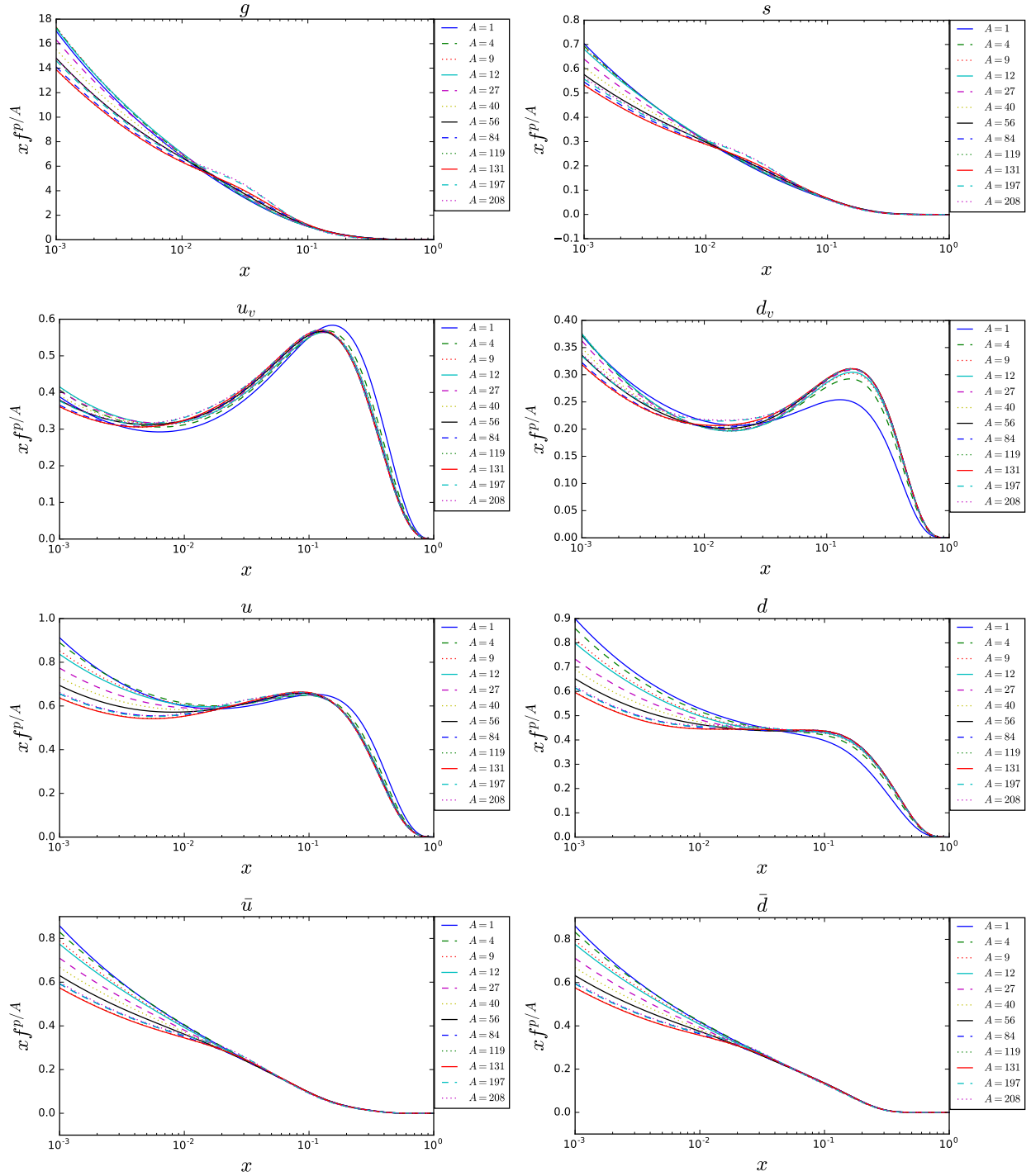
Figure 5.28: `nCTEQ+LHCs` bound proton PDFs for $Q = 10$ GeV for the free proton ($A = 1$) to lead ($A = 208$).
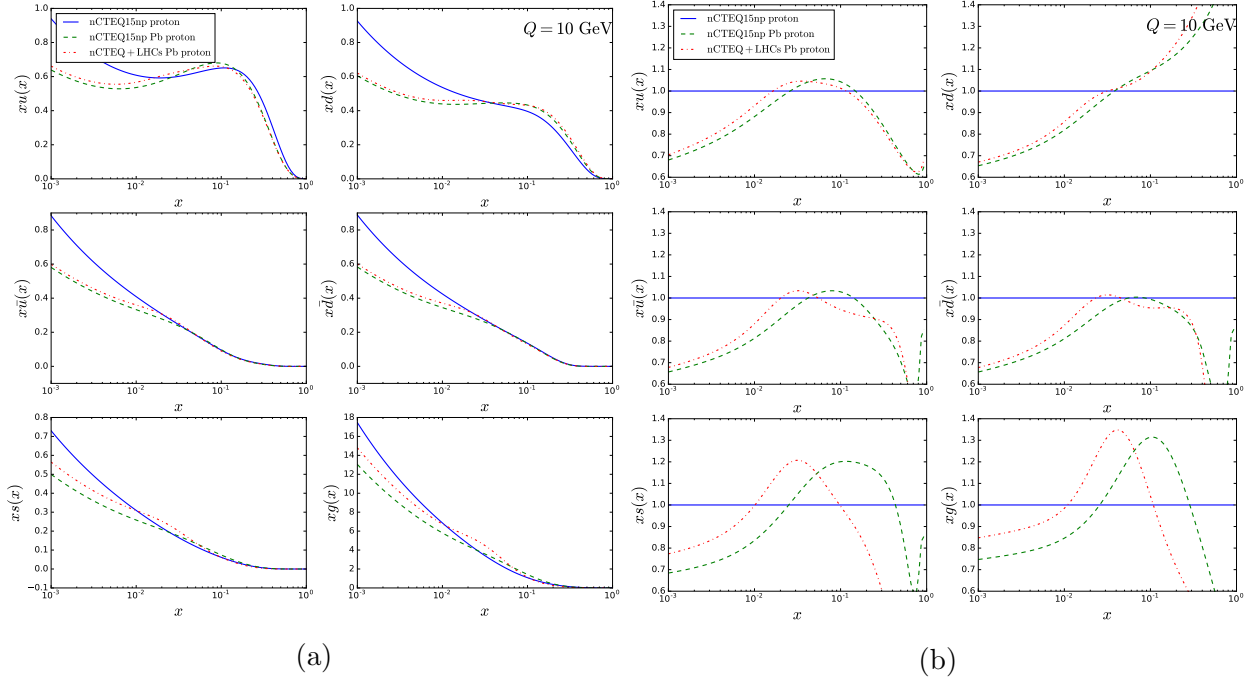
Figure 5.29: `nCTEQ15-np` bound proton PDF compared to the bound proton in lead for `nCTEQ+LHCs` as well as the pound proton in lead for `nCTEQ15-np` 5.29a: for $Q = 10$ GeV as well as the ratio between them, 5.29b. The proton parameters remained fixed for `nCTEQ+LHCs`, so showing the proton PDF for `nCTEQ+LHCs` would be redundant. The ratio plots represent the effective nuclear correction to the proton.
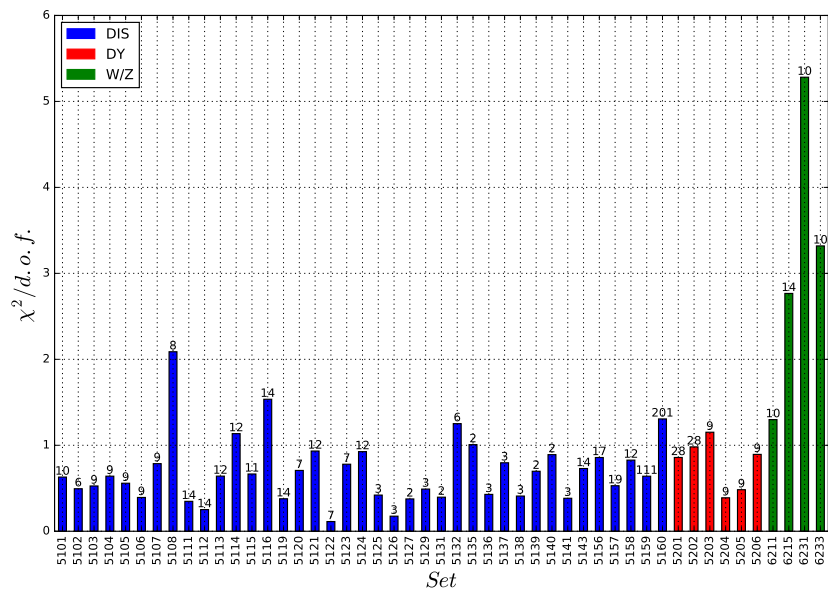
Figure 5.30: Listing of the χ^2/d.o.f. for each data set included in `nCTEQ+LHCs`. The individual data sets are identified by the ID number corresponding to those in Tables 5.1-5.3 as well as the LHC data included in 5.5. The DIS data is displayed in **blue** and have ID numbers corresponding to 51XX, while DY data is displayed in **red** and have ID's corresponding to 52XX. The LHC data is displayed in **green** and have ID numbers corresponding to 62XX.

181
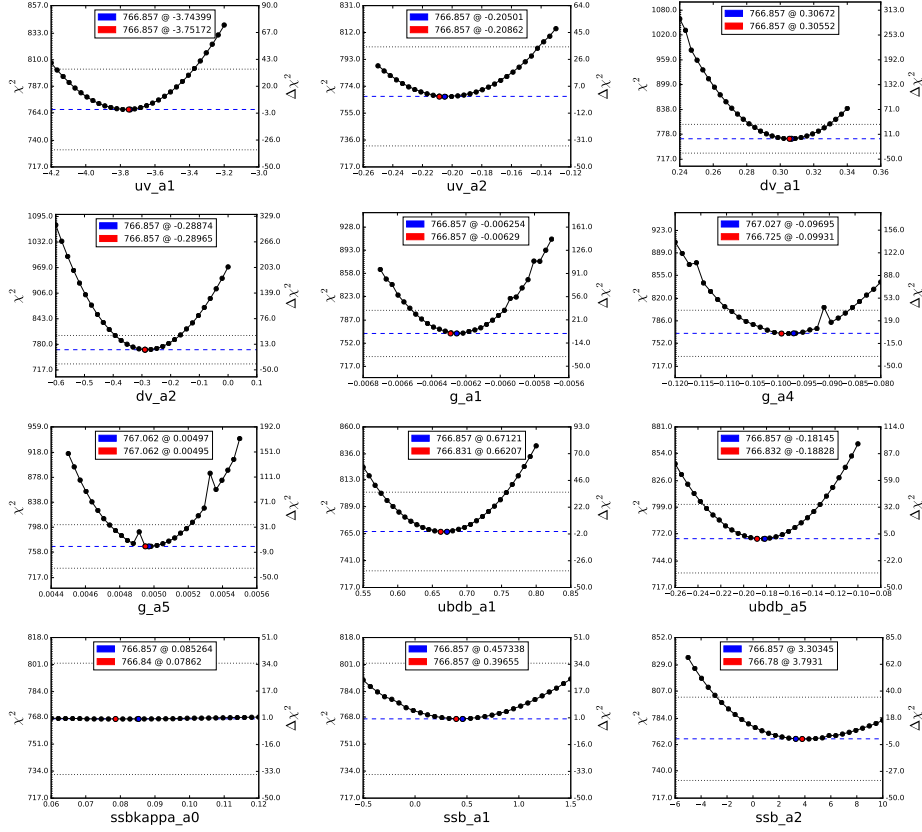
Figure 5.31: One dimensional parameter scans for the 12 free parameters fit in `nCTEQ+LHCs`. Each scan was performed by opening a single parameter while the others remained fixed, eliminating any correlation between parameters. The **blue** marker indicates the value of the parameter and the associated χ^2 from the minimum previously found in the fit. The **red** marker indicates where the parameter is at a minimum in the scan as well as the χ^2 at that point, These two markers should be in essentially the same location. The dotted horizontal black lines indicate ± 35 units of χ^2.

5.4.4. Comparison to Results from PDF Reweighting
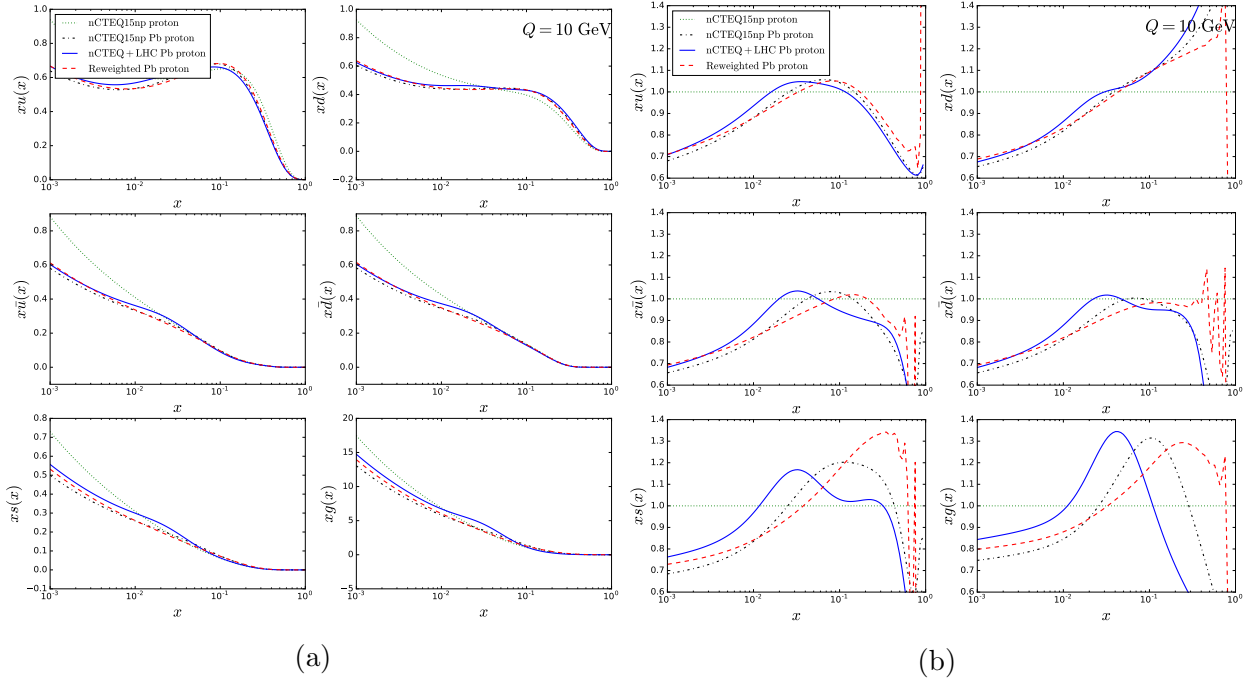
Figure 5.32: Here lead PDFs from `nCTEQ15-np`, the reweighting study (3) and `nCTEQ+LHC` are all compared to the `nCTEQ15-np` proton. The results from the reweighting show movement in the direction of the new fit from `nCTEQ15-np`, particularly in the low x region ($x < 10^{-2}$); the region where the LHC data provides the only constraint on the PDF.

The aim of the PDF reweighting study in Chapter 3 was to estimate and predict the behavior that a PDF refit with LHC data might exhibit. Now that such a fit is available, an evaluation of how well those predictions did is possible. In Figure 5.32, lead PDFs from `nCTEQ15-np`, the reweighting study and `nCTEQ+LHC` are all compared to the baseline `nCTEQ15-np` proton. Encouragingly, the reweighted PDF appears to shift in the direction of `nCTEQ+LHC` from `nCTEQ15-np`, particularly in the low x region ($x < 10^{-2}$). This is the kinematic region previously unconstrained in `nCTEQ15` and that is covered with the inclusion of the LHC data. There is a slight caveat when looking at the reweighting results, especially for the gluon PDF as the reweighted PDF used was `nCTEQ15` and not `nCTEQ15-np`. This means the underlying PDF in the reweighting has some additional constrains that are lacking in `nCTEQ+LHC`.

5.4.5. Normalization of LHC Data

An additional consideration when considering the results from `nCTEQ+LHC`, it that any overall normalization uncertainty was neglected. Currently `nCTEQ++` does not contain the functionality to fit a normalization parameter and enforce a penalty on the χ^2 but it can be done manually after the fact. This was done specifically for the four LHC data sets included in `nCTEQ+LHC`, as shown in Figure 5.34.

$$\chi^2_N(\{a_j\}) = \left(\frac{1 - f_N}{\sigma_N^{norm}}\right)^2 + \sum_i \frac{[D_i - f_N T_i(\{a_j\})]^2}{\sigma_i^2} \tag{5.22}$$

This normalization has a dramatic effect on the χ^2 for these sets, shown in Table 5.8, reducing the χ^2/d.o.f by up to half for one of the sets. This new χ^2 can be related to the normalization uncertainty for each data set by adding a penalty term to the χ^2 calculation, shown in Eq. 5.22 [9]. Here f_N represents the normalization factor, $1 \pm \%$ uncertainty and the uncertainty in the luminosity normalization for that data set is represented by σ_N^{norm}. This penalty is constructed such that for no normalization, there is no penalty to the χ^2 and for any non zero normalization, the penalty[31] adds (% error/% norm)^2 to the χ^2.

	Data ID:	6211	6231	6233	6215
`nCTEQ15-np` [7]	χ^2 per d.o.f:	1.55	6.91	7.73	3.16
Reweighting (3)	χ^2 per d.o.f:	0.87	3.27	2.95	1.76
`nCTEQ+LHC` (5.4.2)	χ^2 per d.o.f:	1.30	5.30	3.36	2.75
`nCTEQ+LHC` ($1 \times \sigma_N$)	χ^2 per d.o.f:	0.92(+0.10)	2.77(+0.10)	1.66(+0.10)	1.96(+0.07)
`nCTEQ+LHC` ($4\sigma_N^{ATLAS}$, $2\sigma_N^{CMS}$)	χ^2 per d.o.f:	0.42(+1.60)	1.33(+0.40)	1.39(+0.40)	0.94(+1.14)

Table 5.8: Comparison of the χ^2 per d.o.f for several PDF sets. The normalized `nCTEQ+LHC` fit represents the χ^2 (with the penalty term in parentheses) recalculated with Eq. 5.22 with $\sigma_N = (2.7, 3.5)$ for (ATLAS, CMS) respectively. As a reminder set IDs of 621X refer to ATLAS and 623X refer to CMS.

---

[31]E.g. (7% error/3.5% norm)$^2 \rightarrow ((2 \times 3.5)/3.5)^2 \rightarrow 2^2 =$ penalty of 4 added to the χ^2
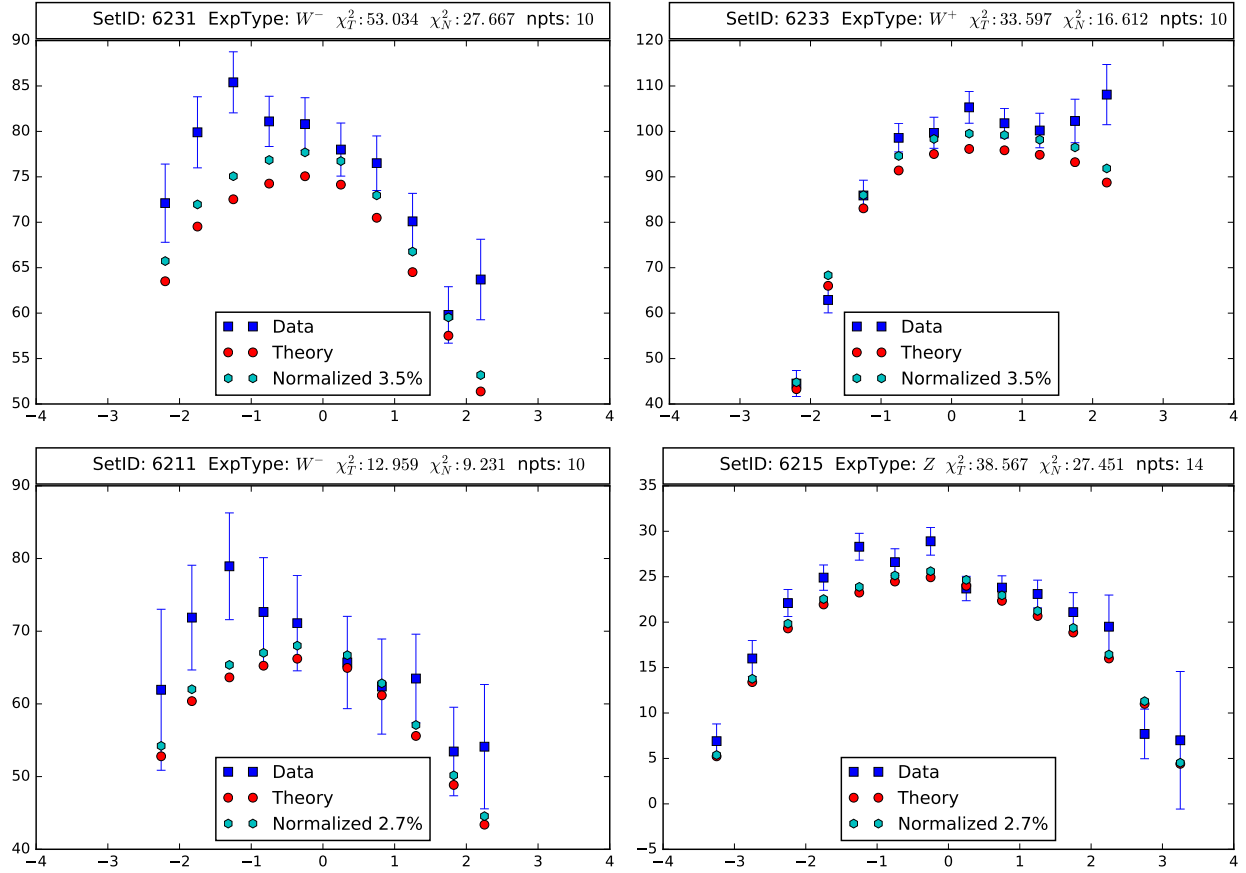
Figure 5.33: Comparison of theory predictions to the data for each of the LHC data sets included in `nCTEQ+LHC` as well as a normalized theory prediction. The normalization for each set represents $1 \times \sigma_N$ for that experiment. The χ_N^2 is calculated according to Eq. 5.22.

It is also possible in `nCTEQ++` to apply a fixed normalization prior to fitting. In Figure 5.36, an example of a fit using an optimal normalization demonstrates the potential of fitting this way. Here the normalization, described in the last row of Tab. 5.8, was applied to the LHC data prior to fitting. The fit was then performed in the same way as `nCTEQ+LHC` and `nCTEQ+LHCs`. The results, Fig. 5.37, show a significant reduction to the χ^2/d.o.f. for the LHC data, putting these data sets on par with the χ^2/d.o.f. of the other data sets included in the fit.

These optimally normalized fits also allow for a comparison to the results from the reweighting analysis from Chapter 3. Figure 5.36 includes these results as well as the results
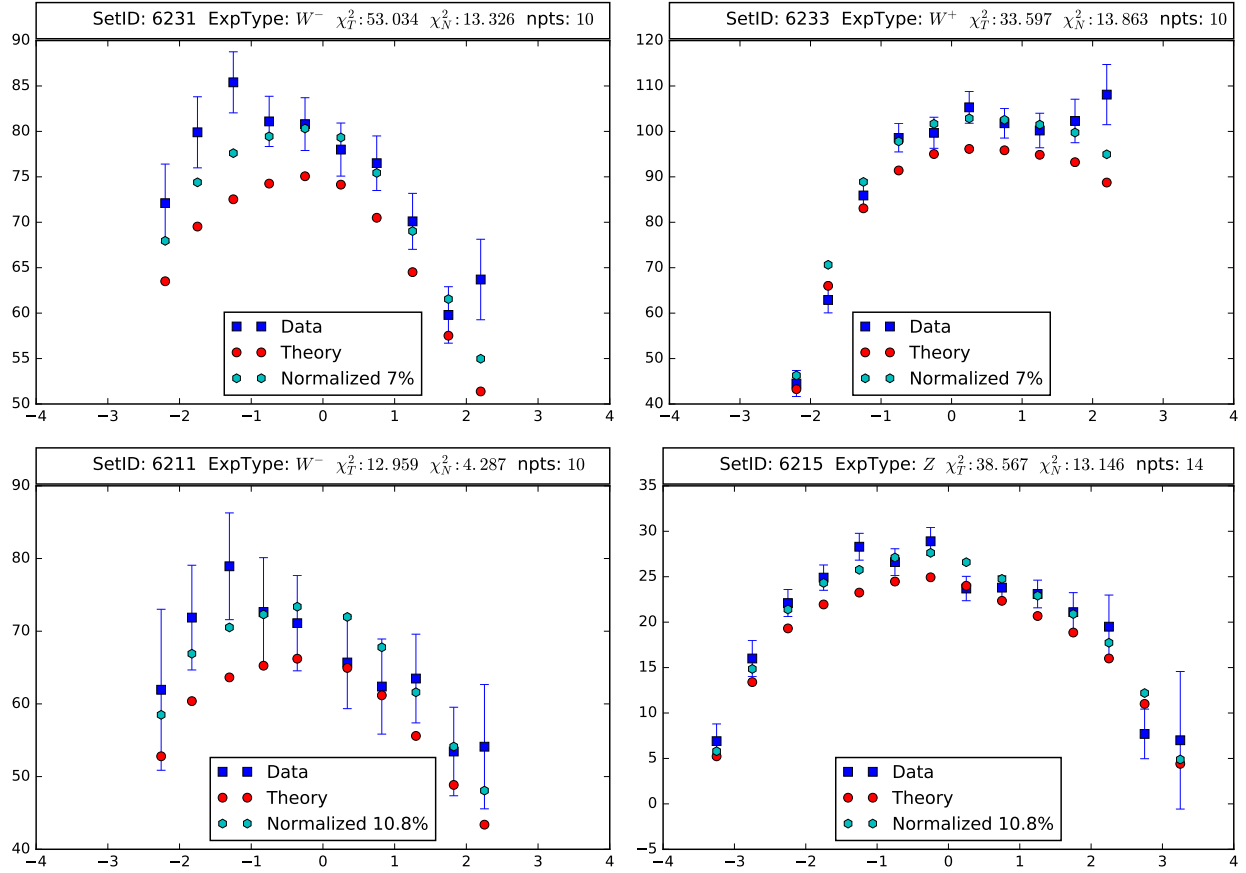
Figure 5.34: Comparison of theory predictions to the data for each of the LHC data sets included in `nCTEQ+LHC` as well as a normalized theory prediction. The normalization for each set represents $4 \times \sigma_N$ for the ATLAS sets and $2 \times \sigma_N$ for the CMS sets. This shows an example of what a fit normalization in `nCTEQ++` might allow for. The χ^2_N is calculated according to Eq. 5.22, but neglects the χ^2 penalty term ($+16 \chi^2$ for ATLAS, $+4 \chi^2$ for CMS).

from the normalized fits. At low x particularly, it becomes clear that the reweighting analysis hinted at, but understated, the direction the data would pull a refit `nCTEQ15` PDF.

Additionally, once the parameterization was freed from having to overcome the normalization, these fits show how opening parameters respond to the inclusion of the W/Z data. For the strange quark specifically, there is a large enhancement at low x that was not seen in the fits without the normalization. This indicates that dynamic fitting of the normalization is likely necessary in order to fully understand the effect these data sets have on the strange quark distribution.
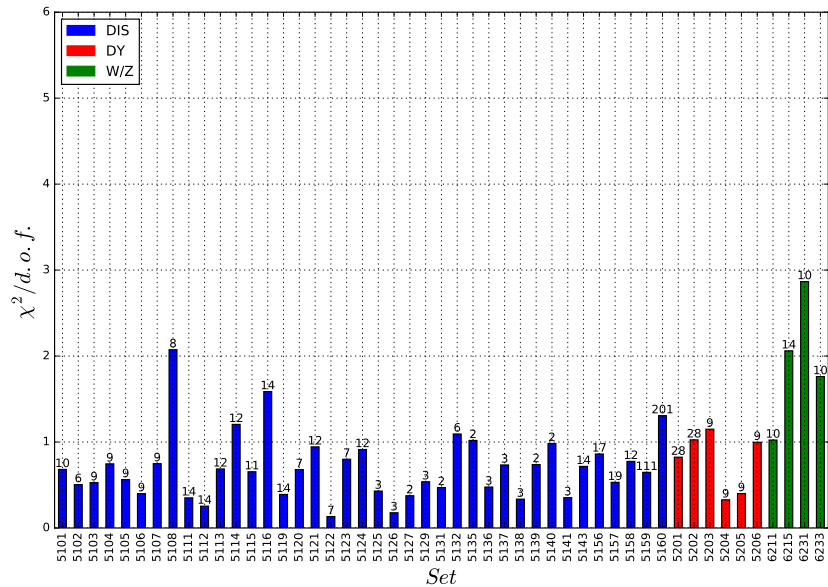
Figure 5.35: Listing of the $\chi^2/\text{d.o.f.}$ for each data set included in `nCTEQ+LHC` with the LHC data normalized by $1 \times \sigma_N$ (including the χ^2 penalty for the normalization). The individual data sets are identified by the ID number corresponding to those in Tables 5.1-5.3 as well as the LHC data included in 5.5. The DIS data is displayed in **blue** and have ID numbers corresponding to 51XX, while DY data is displayed in **red** and have ID's corresponding to 52XX. The LHC data is displayed in **green** and have ID numbers corresponding to 62XX.

(a) Gluon

(b) Gluon - Ratio to `nCTEQ15` Pb

(c) Strange Quark

(d) Strange Quark - Ratio to `nCTEQ15` Pb
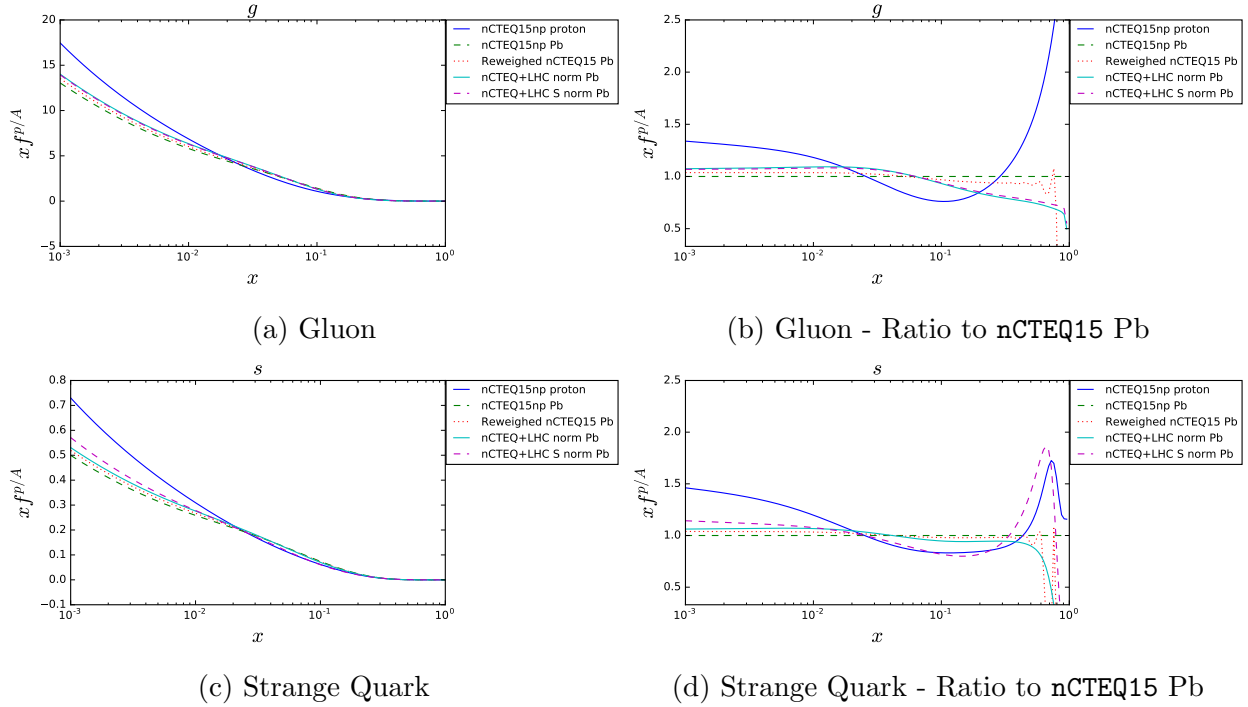
Figure 5.36: The gluon and strange quark PDFs described by the fits (one mirroring `nCTEQ+LHC` and one mirroring `nCTEQ+LHCs`) with a fixed optimal normalization applied prior to fitting. This represents a possible fit `nCTEQ++` might allow for once dynamic fitting of normalizations is implemented. The normalization for each set represents $4 \times \sigma_N$ for the ATLAS sets and $2 \times \sigma_N$ for the CMS sets. There was no χ^2 normalization penalty applied in these fits. The PDF resulting from the reweighting analysis in Ch. ?? is also included in this plot.
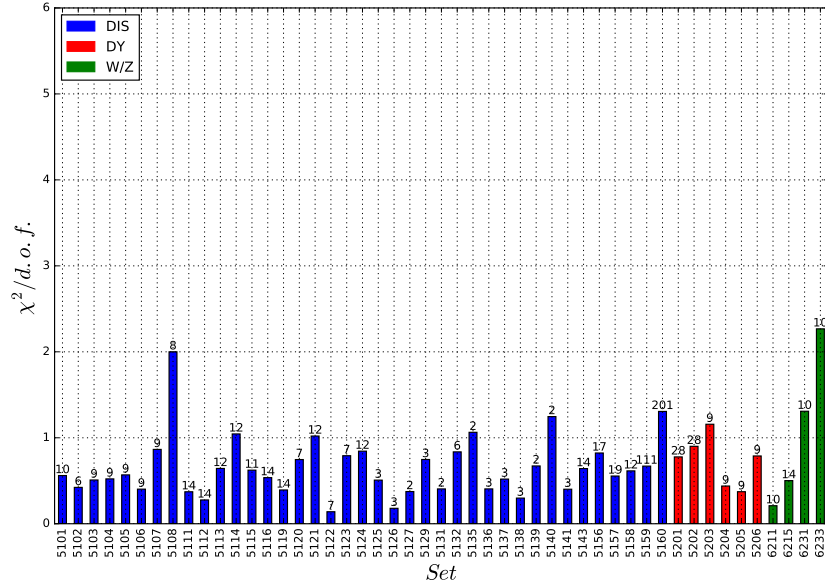
Figure 5.37: Listing of the χ^2/d.o.f. for each data set included in a fit with the LHC data normalized prior to fitting by $4 \times \sigma_N$ for sets 621X and $2 \times \sigma_N$ for sets 623X. This represents an optimal normalization without a χ^2 penalty applied. The individual data sets are identified by the ID number corresponding to those in Tables 5.1-5.3 as well as the LHC data included in 5.5. The DIS data is displayed in **blue** and have ID numbers corresponding to 51XX, while DY data is displayed in **red** and have ID's corresponding to 52XX. The LHC data is displayed in **green** and have ID numbers corresponding to 62XX.

5.4.6. Conclusions

`nCTEQ+LHC` and `nCTEQ+LHCs` both represent successful fits of LHC data. `nCTEQ+LHC` is the first `nCTEQ` fit to achieve a number of milestones. It is the first fit to be performed by `nCTEQ++`, a nontrivial indication that the code was successfully migrated from the old `FORTRAN nCTEQ` code. `nCTEQ+LHC` also represents theory predictions that are a result of `APPLgrid` grids produced from `MCFM` as part of the long chain of theory prediction tools detailed in Section 5.3. It also represents the first inclusion of LHC data into an `nCTEQ` fit and opens the door for future LHC processes to be included in later fits. `nCTEQ+LHCs` is also significant because it marks the first `nCTEQ` fit to include data that allows for any reasonable constraint on the strange quark distribution.

As exciting as these fits are, they are not without their lumps. `nCTEQ+LHC` made it clear that the fit needs to include a normalization parameter in order to better describe the W/Z production data. Unfortunately, `nCTEQ+LHCs` showed that there is still not enough constraint on the strange quark to get any significant information on it, possibly due to the limited number of data points or an inflexible parameterization. There are also lingering concerns about the gluon PDF that was left unconstrained in `nCTEQ15-np`, and as a result `nCTEQ+LHC`, due to the lack of implementation of the pion data present in `nCTEQ15`.

In the next chapter, I will summarize the comprehensive work done in this dissertation as well as make some comments on the direction of this work going forward.

Chapter 6

Conclusion

Having completed a novel pair of parton distribution function (PDF) fits in Chapter 5 representing the first of their kind within the nCTEQ framework, let us step back and examine the entirety of this work. I started by presenting the Standard Model of Particle Physics and the theoretical framework of quantum chromodynamics. This introduced quarks and gluons as the constituent particles of protons and, more broadly, nuclei. These ideas, together with some phenomenological principles, led to the description of PDFs. These PDFs turned out to be intricate objects that required complex and finely tuned machinery in order to determine them.

ManeParse, introduced in Chapter 2, was the first example of such machinery discussed. The goal of ManeParse was to provide the user with a Mathematica package for studying these PDFs entirely independent of external software. I, as part of a small team of developers, achieved this by writing a series of robust parsers for external data files, an internal interpolation routine comparable to the internal Mathematica routine, and an error determination module used for the determination of PDF error bands. With ManeParse, PDFs can be loaded quickly, manipulated, used in calculations and plotted in a clean and efficient manner within Mathematica. This provided a needed utility within the PDF fitting community.

In the next chapter, Chapter 3, I began to discuss PDF fitting and the inclusion of LHC data into these fits. By introducing PDF reweighting as a process for estimating the potential impact a data set might have, I, as part of nCTEQ, determined that certain LHC W/Z production data sets would have a meaningful impact on the nCTEQ15 PDF set. The work presented in this chapter would lay the ground work for the PDF fitting presented in a later chapter. One of the conclusions from this reweighting study is that the LHC data

might provide sufficient constraints to allow for the strange quark parameter to be freed up in a `nCTEQ` PDF fit for the first time. This conclusion was tested with `nCTEQ+LHCs`, described in Chapter 5.

Before approaching a full PDF fit, I described the work I did with the `xFitter` collaboration on heavy flavor variable number schemes in Chapter 4. These schemes provide someone fitting a PDF the flexibility to move the threshold for switching the number of active flavors. This can help to limit the theoretical uncertainties introduced by the discontinuities that arise when fitting at higher orders. Importantly, my work here led me to consider how discontinuities and flavor thresholds contribute to the overall structure of a fit, an issue that arose when validating α_S evolutions in `nCTEQ++`.

Finally, in Chapter 5, I reach the culmination of all the previous chapters with the development of `nCTEQ+LHC` and `nCTEQ+LHCs`. These fits both have elements of significance to them as `nCTEQ+LHC` represents the first `nCTEQ` PDF fit including LHC W/Z production data, while `nCTEQ+LHCs` also includes LHC data but with several strange quark parameters opened for the first time. An additional achievement associated with these fits is the development of `nCTEQ++` fitting code to replace the original `nCTEQ FORTRAN` code. This new fitting code provides numerous advantages but required thorough testing and validation before any results could be produced. One of these innovations was the ability to outsource the theory predictions with powerful precalculated grid techniques by linking to `APPLgrid`. As was true for the code itself, these predictions had to be rigorously tested and verified to produce credible predictions. Armed with these powerful new tools and the new fits I produced, I look to the future of nuclear PDF analyses and some potential improvements for `nCTEQ++` and `nCTEQ+LHC`.

6.1. Physics at the FCC

Using the `nCTEQ+LHC` fit, it is possible to consider physics beyond the LHC. The Future Circular Collider (FCC), is a proposed collider to move beyond the LHC and once completed may be able to produce energies close to 100TeV [148]. While there is a wealth of possible

data to include in PDF fits before this experiment is even close to turning on, through the use of `ManeParse`, we are able to make a simple W^+ production cross section prediction for a beam of this energy.

In Figure 6.1, the dominant contribution to the total cross section is that from $u\bar{d}$. However, even at the jump in energy from the Tevatron to the LHC, contributions from heavy quark interactions (namely $c\bar{s}$) start to have an effect. This is even further visible in Figure 6.2, where the contributions from $c\bar{s}$ at central rapidity are within a factor of 3 to those of $u\bar{d}$. This would indicate that as higher and higher energies are achieved at hadron colliders, further understanding of the PDF for the strange quark and heavy quarks will become increasingly necessary.

It is worth noting here as well that the top quark contribution ($t\bar{b}$) is not shown here despite it being a non-negligible contribution. Currently the top quark PDF is not included in `nCTEQ` fits and thus cannot be used in this prediction. This represents one of the many necessary hurdles that would be necessary to overcome for the FCC.

(a) Tevatron (b) LHC
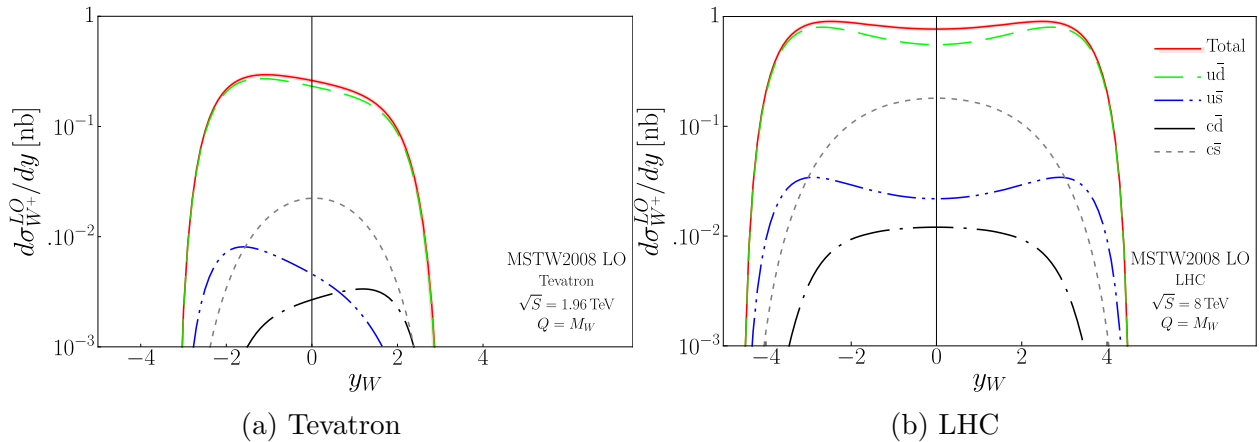
Figure 6.1: Reproduced from Figure 2.13. Leading-Order W^+ production cross section, $d\sigma/dy$ at the Tevatron ($p\bar{p}$, 1.96 TeV) and the LHC (pp, 8 TeV). The red line indicates the total cross section while the individual parton contributions are represented in different colors.
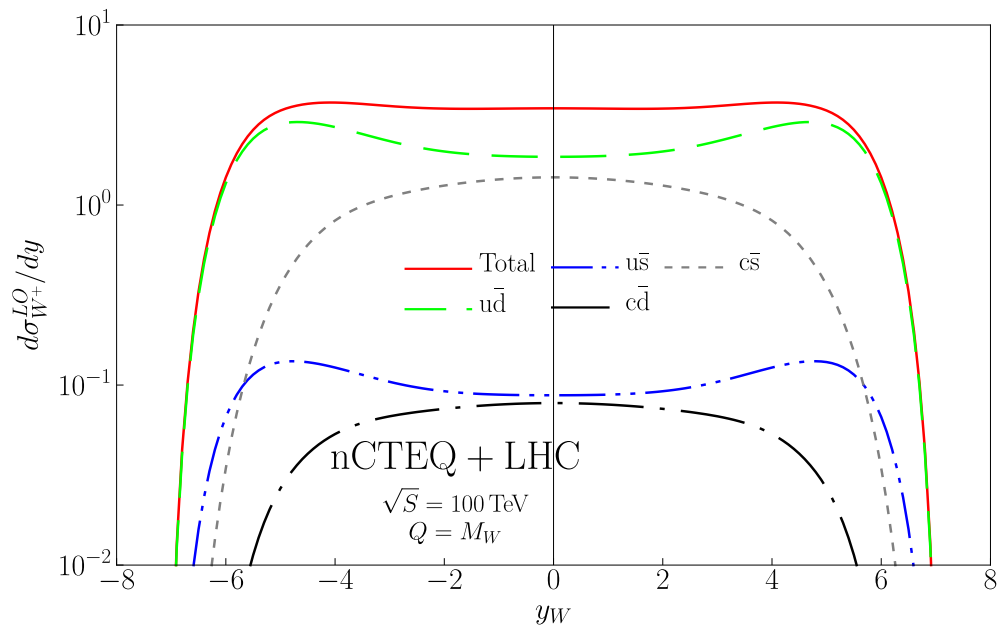
Figure 6.2: Prediction for Leading-Order W^+ production cross section, $d\sigma/dy$ at the FCC (pp, 100 TeV). Of note is the large contribution from $c\bar{s}$, especially at central rapidity. This would warrant an improved understanding of the strange quark PDF as well as the PDFs for the heavy quarks.

6.2. To the Future with `nCTEQ++`

Within `nCTEQ`, a lot of effort has gone into the development, validation and implementation of `nCTEQ++` for PDF fitting and this code was designed with the future in mind. Explained below are a handful of the many possibilities for improving upon existing and creating new PDF fits with this new code. I will primarily focus on ways to improve `nCTEQ+LHC` and `nCTEQ+LHCs`, but the principles can be abstracted to apply to just about any `nCTEQ` fit.

6.2.1. Adding Error Bands

The ability to quantify the uncertatinty of a PDF set is important when preforming a global analysis. A Hessian Error analysis [7] like one implemented in `nCTEQ15` has yet to be implemented in `nCTEQ++` but this will be a crucial component for future fits.

6.2.2. Additional Data Sets

Back in Chapter 3 when choosing W/Z data to use to reweight the PDFs, data from all four major LHC experiments was selected and studied. However, for `nCTEQ+LHC` only includes four of these sets and was limited to a pair of sets from each of ATLAS and CMS. This reduction of the data sets was a result of the time it took to innovate and validate the chain of theory tools for producing `APPLgrid` grids to use as theory production. Now that this process is established and has been proven to be an effective technique for including new data into a fit, more of the pPb and PbPb sets used in the reweighting study could conceivably be included in `nCTEQ+LHC`.

Another data set missing in `nCTEQ+LHC` that has been included in the past is the inclusive pion production data included in `nCTEQ15`. This data provided `nCTEQ15` with more constraint on the gluon parameters than `nCTEQ15-np` has and its inclusion in `nCTEQ++` would allow future fits to use `nCTEQ15` as a baseline, instead of `nCTEQ15-np`.

There are several other nuclear data sets that are also available. For example, EPPS16 included a huge number of neutrino DIS data points from the CHORUS [69] experiment (824 points), as well as, dijet data from CMS [149]. Other neutrino experiments such as

NuTeV [150] [151] and MINERνA [152] also have data that could be considered for inclusion for a future fit. There are even some interesting results from the LHCb experiment where they utilize their capability of running in fixed target mode with its internal gas target SMOG [153] [154]. There are even new tools being developed that use machine learning algorithms [155] to provide insight into which of these data sets might be most profitable for inclusion a fit, similar to, but arguably more flexible than, the statistical approach of PDF reweighting.

6.2.3. Expanding with MCFM

With this wealth of available data, being able to quickly and accurately generate theory predictions becomes the main constraint for fitting new processes. Fortunately, `MCFM` can provide theory predictions for hundreds of processes[1] and, by extension, `APPLgrid` gridded theory predictions. These processes can then be include in `nCTEQ++` quite easily by following the steps in Section 5.3.

6.2.4. Open Additional Parameters

When examining new data, it is natural to consider opening more and more parameters. `nCTEQ+LHC` and `nCTEQ+LHCs` were limited by a number of factors to only having 9 and 12 parameters open, where as `nCTEQ15-np` only had 16 open parameters. These numbers represent only a fraction of the total parameters available for fitting and with the speed afforded by using `APPLgrid` to precalculate the theory predictions, run time is no longer as constraining as it once was for fitting a large number of parameters. Now the limitations are more a function of data availability and precision, as well as the use of a fixed parameterization.

When examining the ~ 60 nuclear parameters available with the current `nCTEQ` parameterization, there are definitely areas that can be probed further. For instance, adding data sets with only a single nuclei, Pb, precluded opening both the a_k and b_k parameters for each

---

[1]The list of processes included in `MCFM` can be found in the `MCFM` v6.8 Users Guide found here: https://mcfm.fnal.gov/

196

flavor, but with data from different nuclei, this would no longer be the case. The addition of inclusive pion data could allow more gluon parameters to be constrained, or even work with the LHC W/Z data in this. As I showed in `nCTEQ+LHCs`, opening the strange quark parameters was not as fruitful as it was hypothesized to be, but that's not to say that some future data might not help to constrain this.

While most of these parameters rely on new data to constrain them, a vital parameter missing from `nCTEQ+LHC` and `nCTEQ+LHCs` is a free normalization parameter tied to the luminosity uncertainty in the normalization of the data (as well as other normalization uncertainties). I touched on this effect briefly at the end of Chapter 5 by applying a fixed normalization and penalty to the χ^2 for the LHC data in the fit. This however is not an ideal solution and should be implemented into `nCTEQ++` directly when fitting. An example of what this might look like for the ATLAS and CMS data from `nCTEQ+LHC` can be found in Fig. 6.3.

This does not represent a true minimization of the χ^2 by fitting a normalization parameter, but it provides preliminary indication that fitting the normalization would be able to improve the overall χ^2 of a fit.

6.2.5. Improved Parameterization

In addition to all of these potential improvements, it is possible to rework the underlying parameterization in `nCTEQ++` itself. The current parameterization is structured like the `FORTRAN nCTEQ` code, which itself was structured like the `CTEQ` fitting code. In `nCTEQ++`, there is the possibility to change the parameterization without reworking the entire code structure. This was possible due to the class structure in `C++` and would allow future fits to test out difference parameterizations without any significant hindrance. To some extent, this has already been done to allow `nCTEQ++` to read `LHAPDF6` files and use them in calculations using an "LHA Parameterization." While this particular parameterization is extremely limited (e.g. It can not be used when attempting to fit), future parameterizations would not face such restrictions.
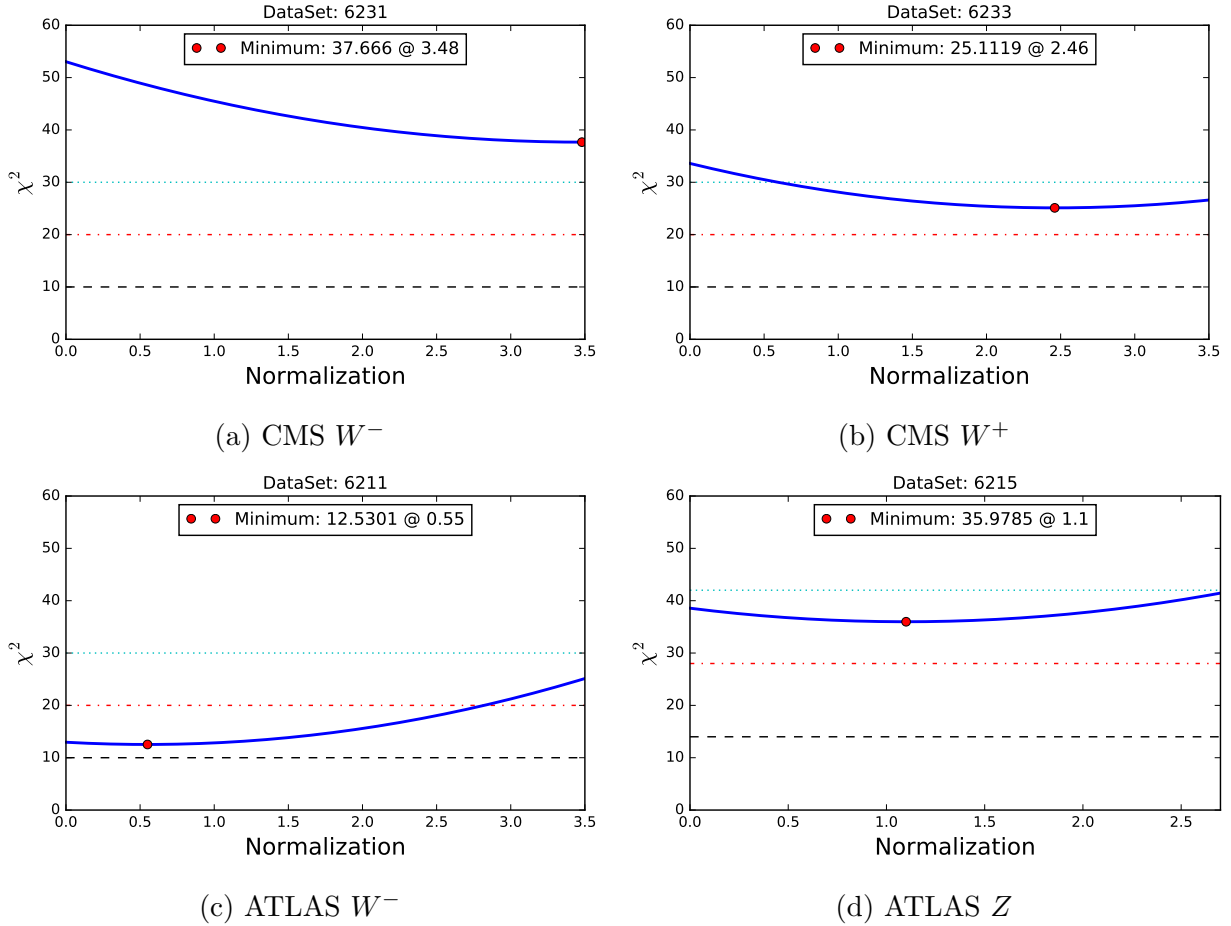
Figure 6.3: An example of what a fit normalization in `nCTEQ++` might allow for. Here the χ_N^2 is calculated according to Eq. 5.22 including the χ^2 penalty. The x-axis ranges from no normalization (0) out to $1 \times \sigma$ for each set (2.7 for ATLAS and 3.5 for CMS). The horizontal lines represent 1, 2, and 3 χ^2 per degree of freedom; dashed, dash-dotted, and dotted respectively.

One especially interesting option for a future parameterization would be to unfix the strange quark asymmetry, $s - \bar{s} = 0$. This idea was touched on briefly in Chapter 3 as the LHC data in that study exhibited some tension in this regard. In order to implement this, additional parameters would have to be introduced as well as some functional form, polynomial or otherwise.

Another interesting option would be to rework the A dependence for the coefficients in the parameterization. When looking at how the parameters themselves changed with increasing A, Fig 6.4, the parameters change rapidly at small A but quickly level off for $A < 50$

in most cases[2] (the gluon being the exception). While this behavior is clearly favored in `nCTEQ+LHC` and `nCTEQ15-np` (As seen in Section 5.4), it might be that a different nuclear parameterization could do a better job of describing the data.

6.3. Closing Remarks

The vast majority of this work was done in order to improve the understanding of parton distribution functions. I did this both by building tools to better study PDFs, as well as producing my own PDF fits. `ManeParse` has proven itself to be a useful tool already by providing a simple structure to load and cross check PDFs. PDF reweighting was crucial in the development of `nCTEQ+LHC` while being a useful technique in its own right. My work with `xFitter` on heavy flavor variable number schemes was also incredibly valuable for validating `nCTEQ++` and producing `nCTEQ+LHC`. Finally, the fits presented in this work, `nCTEQ+LHC` and `nCTEQ+LHCs`, are the first of their kind both by including LHC data in a `nCTEQ` fit and by being the first time the strange quark has been able to be fit within `nCTEQ`.

---

[2]For perspective: calcium has an $A = 40$ and iron has an $A = 56$. `nCTEQ` fits cover A values up to lead, $A = 208$.
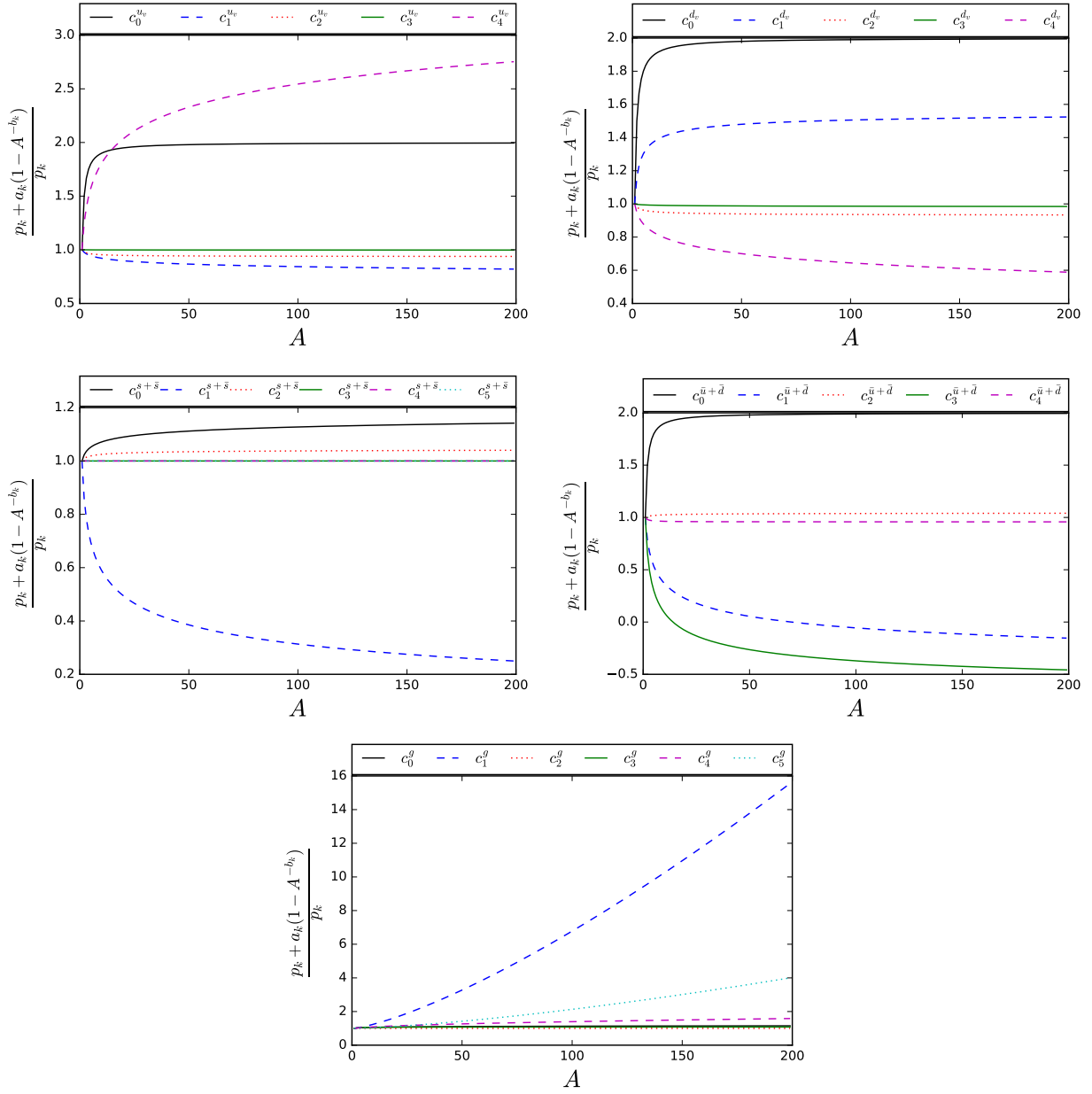
Figure 6.4: Reproduced from Fig. 5.22: A detailed look at the A dependence of the parameters fit in `nCTEQ+LHC` as described by Eq. 5.5. As `nCTEQ+LHCs` has very similar parameters, the strange was included as well. Each of these plots have been normalized to the free proton parameters such that only the nuclear parameters contribute. The parameters for `nCTEQ+LHC` and `nCTEQ+LHCs` can be found in 5.6 and 5.7, respectively.

Appendix A

ManeParse Distribution Files

A.1. `ManeParse` Distribution Files

The `ManeParse` package is distributed as a gzipped tar file (about 2.6Mb), and this is available at `cteq.org` or `ncteq.HepForge.org`.

When this is unpacked, the `ManeParse` modules {`pdfCalc`, `pdfErrors`, `pdfParseCTEQ`, `pdfParseLHA` } will be in the `./MP_Packages/` directory.

There is a `Demo.nb` Mathematica notebook which will illustrate the basic functionality of the program; we also include a `Demo.pdf` file so the user can see examples of the correct output.

We do not distribute any PDF files, so these must be obtained from the `LHAPDF6` website[1] or the CTEQ website.[2] The `README` file will explain how to run the `MakeDemo.py` python script to download and set up the necessary directories for the PDF files.[3]

The `MakeDemo.py` script will also run the Perl script `noe2.perl` on the CT10 data files. Older versions of these files use a two digit exponent (e.g. 1.23456E-12), but occasionally three digits are required in which case the value is written as 1.23456-123 instead of 1.23456E-123. While the GNU compiler writes and reads this properly, other programs (including `Mathematica`) do not, so the `noe2.perl` script fixes this. This script can also be run interactively, in which case it will print out any lines that are modified.

---

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

[2] http://cteq.org/

[3] Python is not essential to `ManeParse` as the files can be setup manually.

There is a manual in both `Mathematica` format (manual_v1.nb) and PDF format (manual_v1.pdf); this allows the user to execute the notebook directly, but also see how the output should look. The manual provides examples of all the functions of `ManeParse`.

There is also a glossary file `User.pdf` which provides a list and usage of all the commands.

A.2.   A Simple Example

First we define some directory paths. You should adjust for your particular machine. Note, for LHAPDF6, the individual "dat" and "info" files are stored in subdirectories.

pacDir="../ManeParse/Demo/packs"

pdfDir="../LHAPDF"

subDir1=pdfDirlg **"/MSTW2008nnlo68cl"**

subDir2=pdfDirlg **"NNPDF30_nnlo_as_0118_nf_6"**

ctqDir="../ManeParse/Demo/PDF_Sets/PDS"

Next, we load the `ManeParse` packages. The `pdfCalc` package is automatically loaded by both `pdfParseLHA` and `pdfParseCTEQ`, so we do not need to do this separately.

Get[pacDirlg **"/pdfParseLHA.m"];**

Get[pacDirlg **"/pdfParseCTEQ.m"];**

Get[pacDirlg **"/pdfErrors.m"];**

`pdfParseLHA` will read the PDF set and assign an "iSet" number, which in this case is 1.

iSetMSTW=

pdfParseLHA[

   subDir1lg **"/MSTW2008nnlo68cl.info",**

   subDir1lg **"/MSTW2008nnlo68cl_0000.dat"]**

Out[...]:=1

The "iSet" numbers are assigned sequentially, and are returned by `pdfParseLHA` which we use to define the variable `iSetMSTW` (=1 in this example). We can then evaluate the PDF values.

iParton=0; (* Gluon *)

x=0.1;

q=10.;

pdfFunction[iSetMSTW,iParton,x,q]

Out[...]:=11.714

Next, we can read in an NNPDF PDF set.

iSetNNPDF=

pdfParseLHA[

   subDir2lg "/NNPDF30_nnlo_as_0118_nf_6.info",

   subDir2 lg "/NNPDF30_nnlo_as_0118_nf_6_0000.dat"]

Out[...]:=2

We can then evaluate this PDF. We find it is similar (but not identical) to the value above.

pdfFunction[iSetNNPDF,iParton,x,q]

Out[...]:=11.8288

Finally, we load a `ctq66` PDF file in the older "pds" format using the `pdfParseCTEQ` function; note this only takes a single file as the "info" details are contained in the "pds" file header.

iSetC66=pdfParseCTEQ[

      ctqDirlg "/ctq66.00.pds"];

Out[...]:= 3

pdfFunction[iSetC66,iParton,x,q]

Out[...]:= 11.0883

Now that we have these functions defined inside of Mathematica, we can make use of all the numerical and graphical functions. Detailed working examples are provided in the auxiliary files.

A.3. N_F-Dependent PDF Example

We provide an example of implementing the N_F-dependent PDFs within the `ManeParse` framework using the matched set of PDFs[4] with $N_F = \{3, 4, 5, 6\}$ from Ref. [20]. We load the `ManeParse` packages as above, and then read in the grid files which are in "pds" format.

pdfDir="../vfnsnf";

iSetNF3=pdfParseCTEQ[pdfDirlg **"/nf3_q1.2.pds"]**

iSetNF4=pdfParseCTEQ[pdfDirlg **"/nf4_q1.2.pds"]**

iSetNF5=pdfParseCTEQ[pdfDirlg **"/nf5_q1.2.pds"]**

iSetNF6=pdfParseCTEQ[pdfDirlg **"/nf6_q1.2.pds"]**

`pdfParseCTEQ` returns the "iSet" number and we store these in {`iSetNF3, ...` }. The below function `pdfNF` allows the user to choose N_F, and then returns the appropriate PDF.

Clear[pdfNF,nf,iParton,x,q];

pdfNF[nf_,iParton_,x_,q_]:=Module[{iSet=0},

    If[nf==3,iSet=iSetNF3];

    If[nf==4,iSet=iSetNF4];

    If[nf==5,iSet=iSetNF5];

    If[nf==6,iSet=iSetNF6];

---

[4]These PDF sets are available at http://ncteq.hepforge.org/.
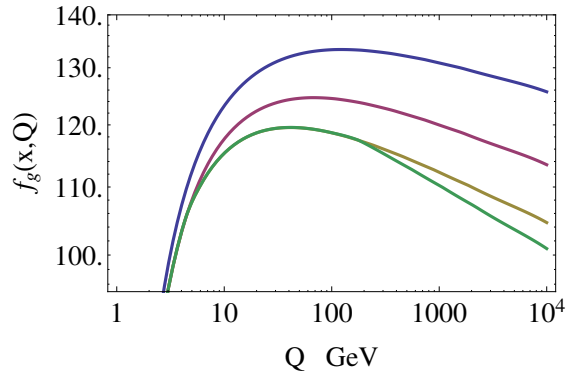
Figure A.1: We display the gluon PDF $f_g(x, Q)$ at $x = 0.03$ vs. Q for $N_F = \{3, 4, 5, 6\}$; $N_F = 3$ is the largest, and $N_F = 6$ is the smallest curve.

```
If[iSet==0,Return[Null]];

Return[pdfFunction[iSet,iParton,x,q]]

]
```

Note in the `pdfNF` function, the "iSet" variable is local to the `Module`. We now compute some sample values.

```
iParton=0; (* Gluon *)
x=0.03;
q=10.;
{pdfNF[3,iParton,x,q], pdfNF[4,iParton,x,q],
 pdfNF[5,iParton,x,q], pdfNF[6,iParton,x,q]}
```
Out[...]:={123.288, 117.694, 115.331, 115.341}

As we have taken $Q = 10$ GeV, we are above the charm and bottom transition, but below the top transition; hence the $N_F = \{5, 6\}$ results are the same, but the $N_F = \{3, 4\}$ values differ.

In Fig. A.1 we display the gluon PDF vs. Q for $N_F = \{3, 4, 5, 6\}$. We observe as we activate more flavors in the PDF evolution the gluon is reduced as a function of N_F. This decrease in the gluon PDF will be (partially) compensated by the new N_F channels.

Appendix B

Further Validations of `nCTEQ++`

B.1. Evolution Code Comparison

This section follows from Section 5.2.1.2 and provides a more thorough comparison of two PDFs, one evolved with the `nCTEQ FORTRAN` code and the other with the `HOPPET` evolution routine implemented in `nCTEQ++`. The comparisons are broken down by parton flavor and are displayed at Q values of interest when fitting PDFs. There is also a section for validation performed at initial scale for Q. The PDF these comparisons are based on is `nCTEQ15_1_1`, which is the `nCTEQ15` bound proton PDF [7] and can be found at https://ncteq.hepforge.org/ncteq15/ncteq15.html.

B.1.1. Validation at Q_0
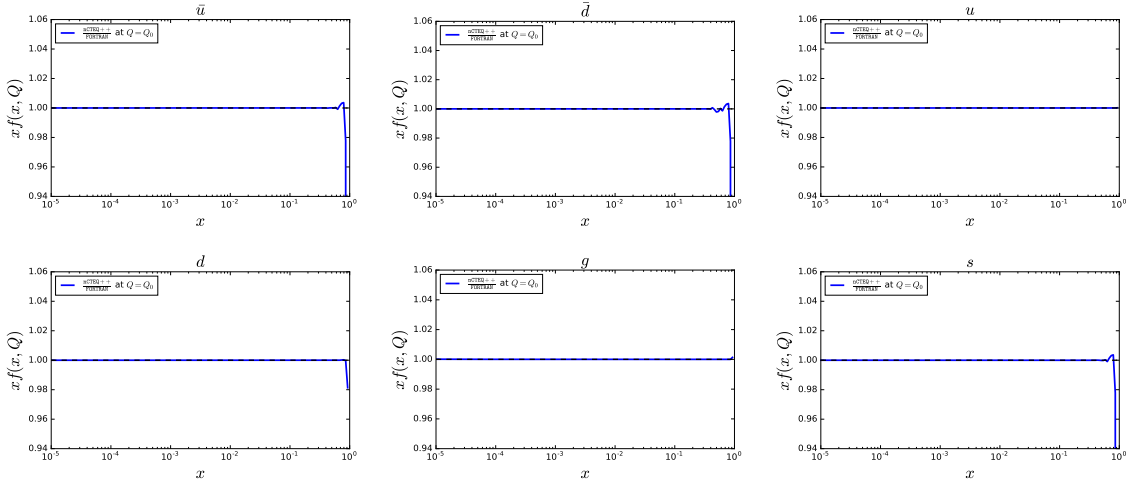
Figure B.1: At Q_0 there is no sign of a variation between the two evolution codes. This is expected as the PDF is fit at Q_0 and evolved up to the desired Q. Here $Q_0 = 1.3$GeV. At very large x the PDFs are ~ 0 and the ratio becomes unstable, causing the behavior seen on the far right of each plot.

The matching of the PDFs at Q_0 provides the necessary validation that the PDFs are indeed identical and any difference must arise from evolution procedures. The following subsections are not validations, in so much as they provide a catalog of the differences in the behavior of the evolution codes for a comprehensive list of x, Q and parton flavor combinations. These differences are attributed to the algorithmic differences in solving for α_S and differences in the number of flavors present in the evolution, especially the handling of the top quark in the α_S evolution.

B.1.2. Sea Quarks: $\bar{u}$ and $\bar{d}$
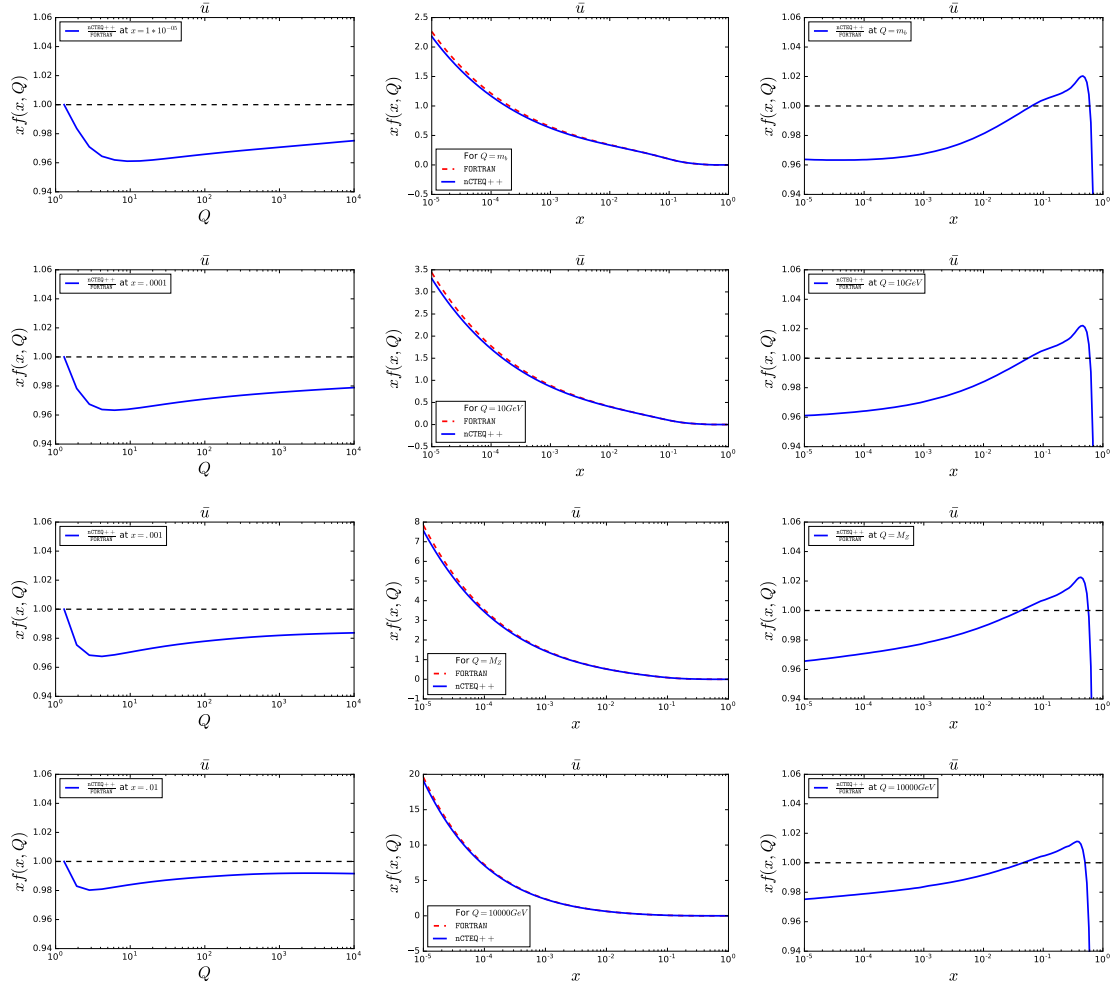
Figure B.2: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here the parton flavor is fixed to be $\bar{u}$ and multiple Q and x values are displayed. Here $m_b = 4.5\text{GeV}$ and $M_Z = 91.188\text{GeV}$.
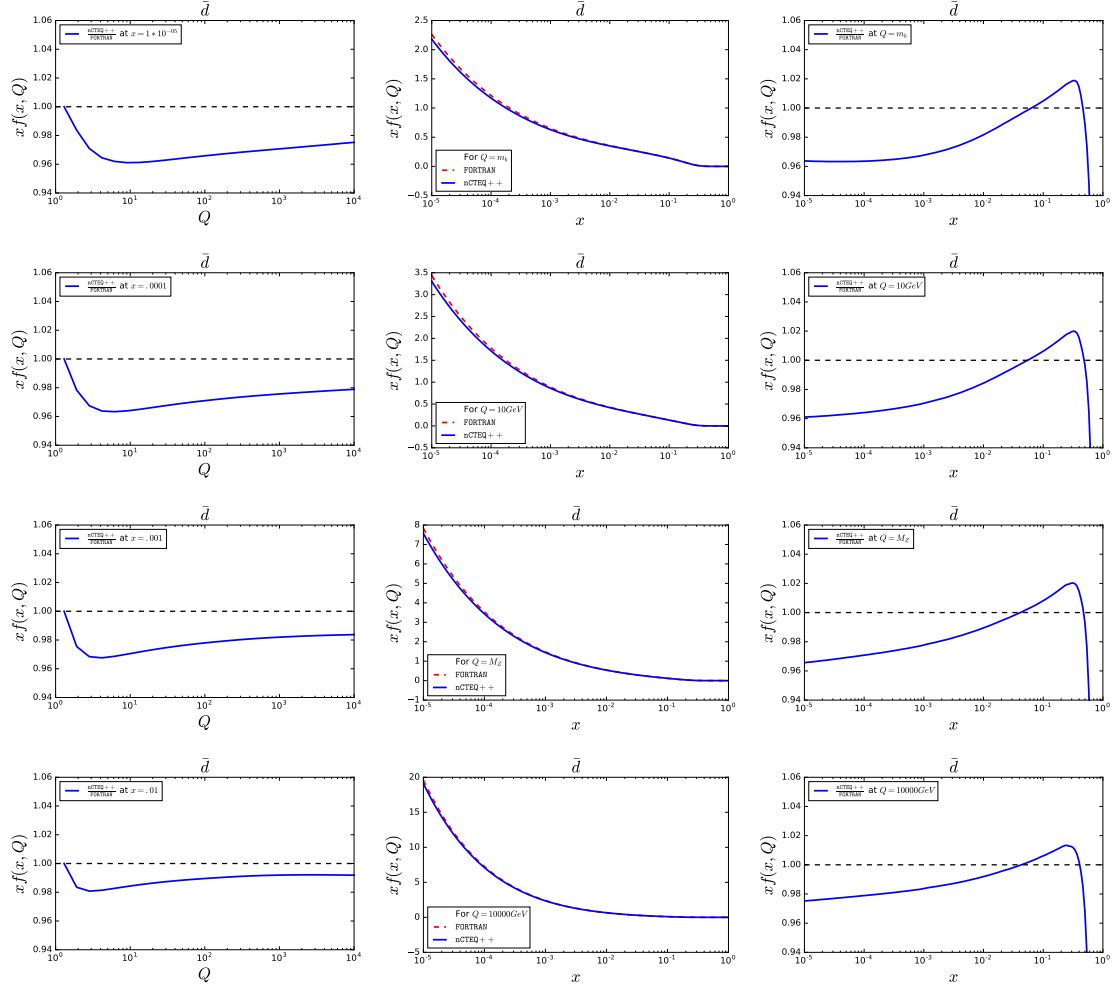
Figure B.3: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here the parton flavor is fixed to be $\bar{d}$ and multiple Q and x values are displayed. Here $m_b = 4.5\text{GeV}$ and $M_Z = 91.188\text{GeV}$.

B.1.3. Valence + Sea Quarks: u and d
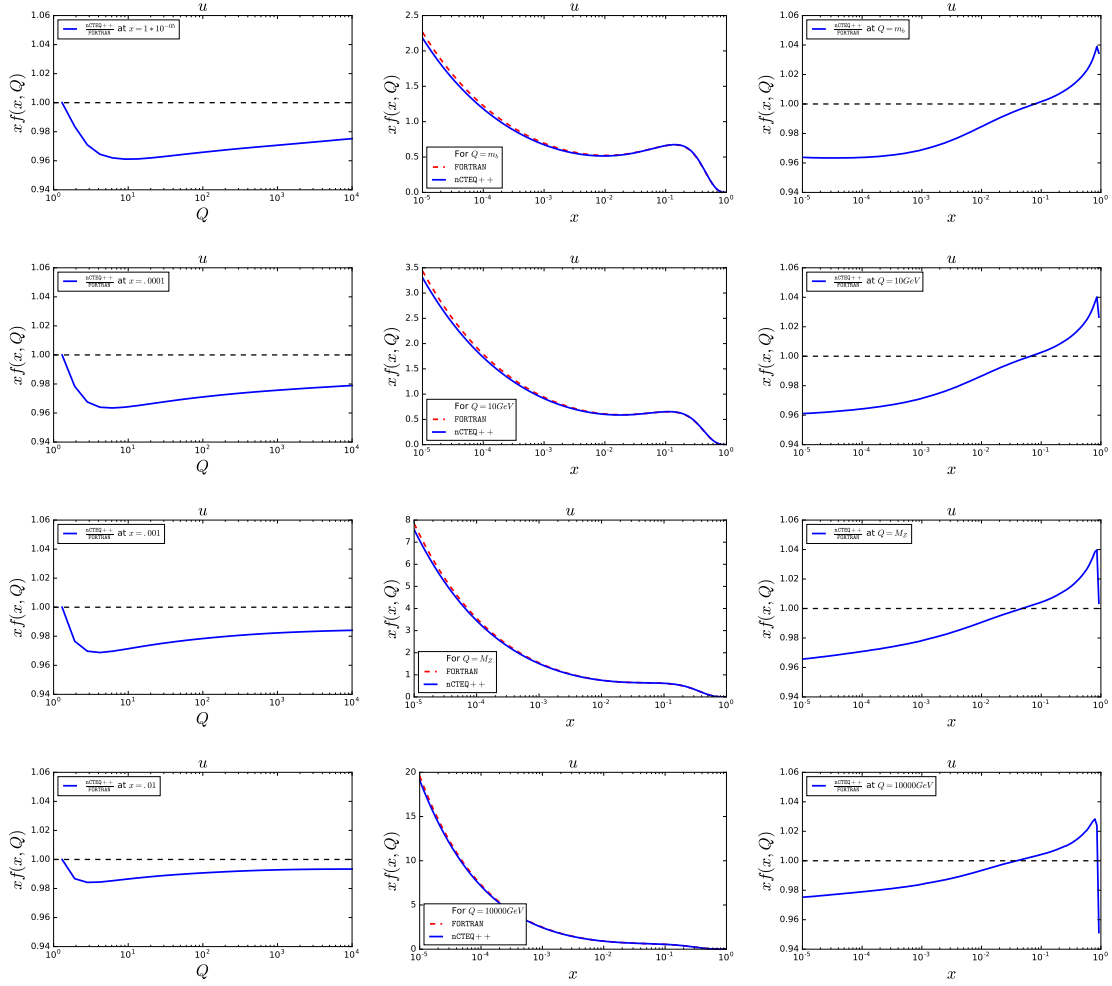
Figure B.4: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here the parton flavor is fixed to be u and multiple Q and x values are displayed. Here $m_b = 4.5$GeV and $M_Z = 91.188$GeV.
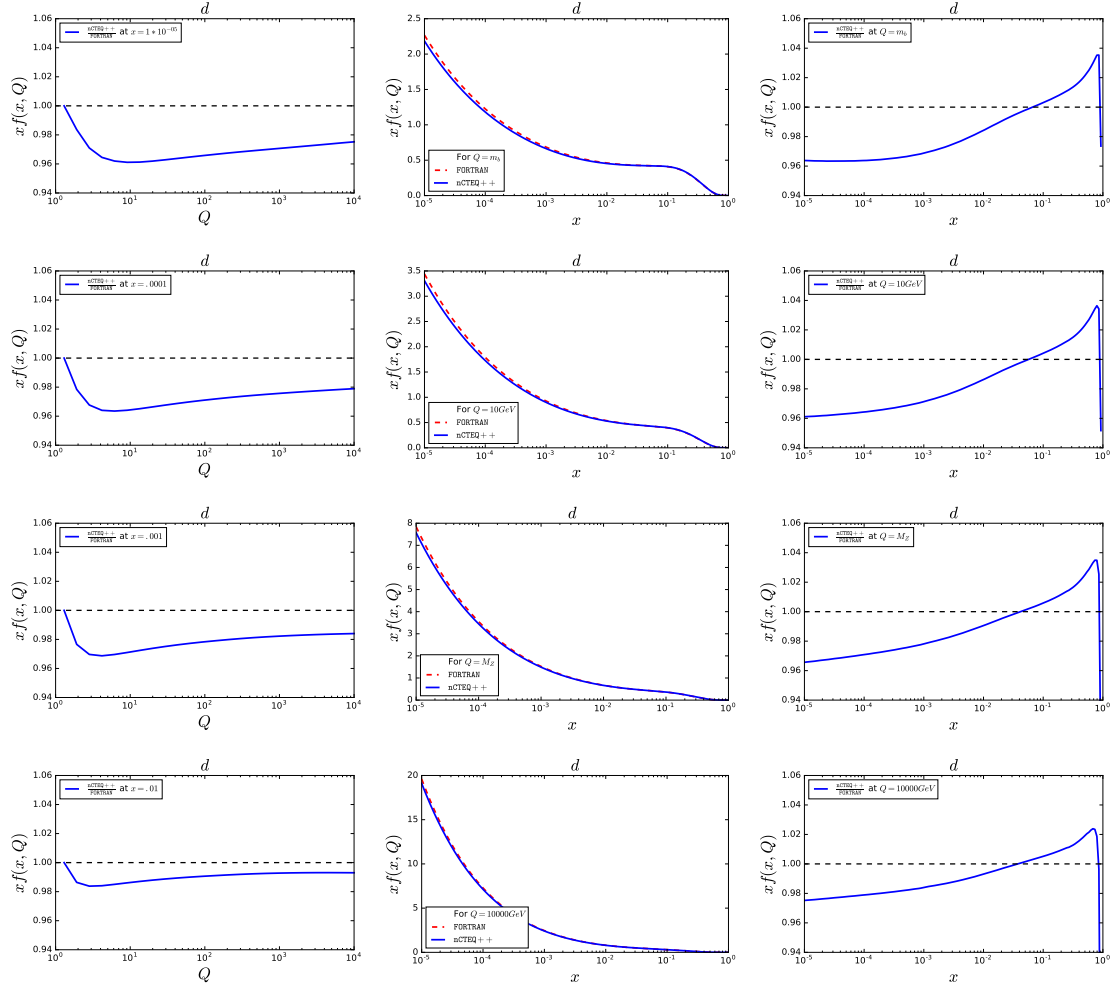
Figure B.5: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here the parton flavor is fixed to be d and multiple Q and x values are displayed. Here $m_b = 4.5\mathrm{GeV}$ and $M_Z = 91.188\mathrm{GeV}$.

B.1.4. Gluon and Strange
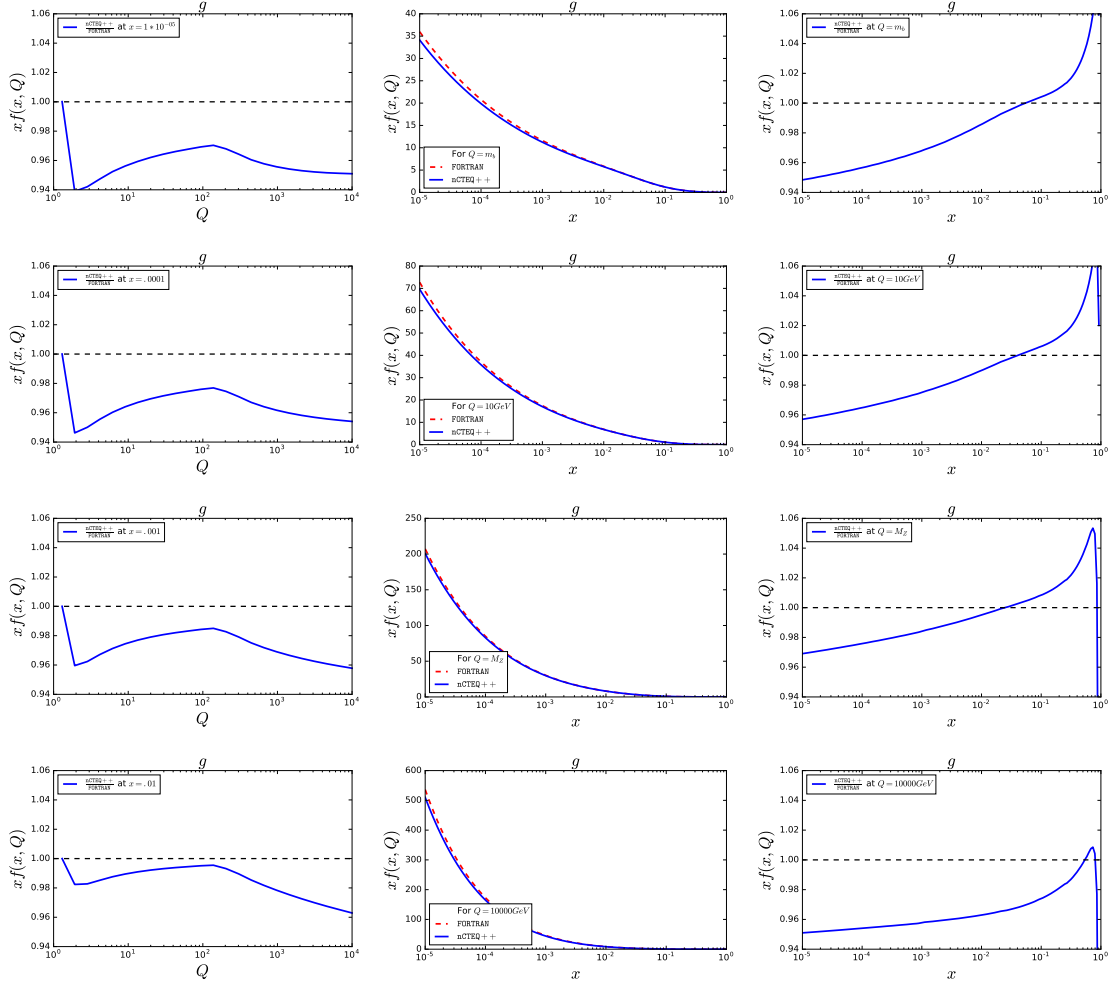
Figure B.6: The ratio of a PDF evolved using the `HOPPET` implementation in `nCTEQ++` to the same PDF evolved using the existing `FORTRAN nCTEQ` code. Here the parton flavor is fixed to be g and multiple Q and x values are displayed. Here $m_b = 4.5\text{GeV}$ and $M_Z = 91.188\text{GeV}$.
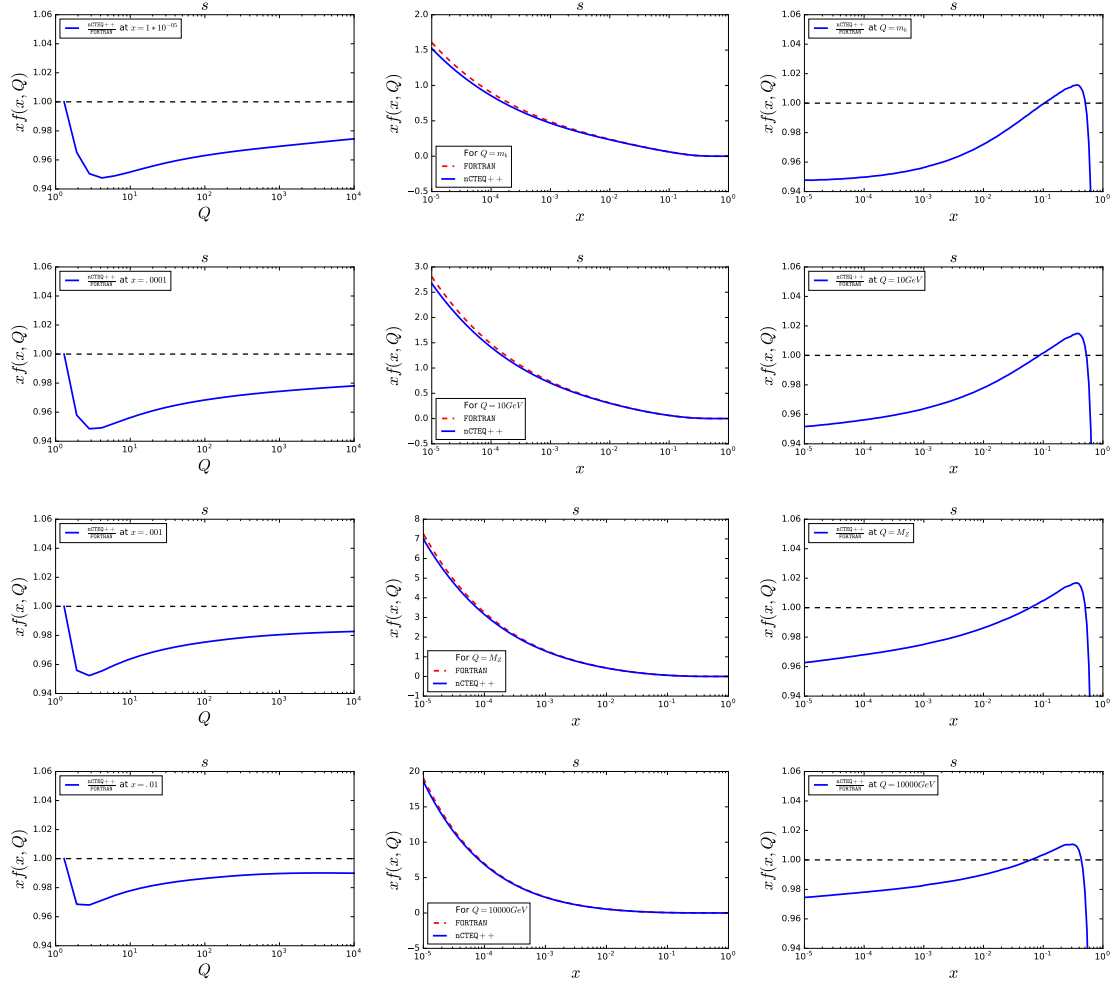
Figure B.7: The ratio of a PDF evolved using the HOPPET implementation in nCTEQ++ to the same PDF evolved using the existing FORTRAN nCTEQ code. Here the parton flavor is fixed to be s and multiple Q and x values are displayed. Here $m_b = 4.5\text{GeV}$ and $M_Z = 91.188\text{GeV}$.

Further Validations of Theory Prediction Tools

This Appendix contains several validations done for the chain of theory tools detailed in Section 5.3.

C.1. Additional `FEWZ` to `MCFM` Comparisons

This section catalogs the different theory calculations in `FEWZ` that were replicated in `MCFM`. Generally speaking, the comparisons were done for Leading Order (LO) and Next to Leading Order (NLO), rapidity of the boson and the leptons, and transverse momentum of the boson and leptons. Additionally data sets included come from the ATLAS and CMS collaborations for both PbPb and pPb data sets.

Of note, out-of-the-box `MCFM` version 6.8 does not have the ability to accomodate cuts on boson rapidity nor does it permit asymmetric rapidity cuts for the leptons. As both of these were needed for the following validations, the transverse mass and lepton veto cuts, respectively, already present in `MCFM` were modified internally to obtain additional functionality and perform the needed cuts.

C.1.1. Z^0

$PbPb \to Z \to \ell^+ + \ell^-$

C.1.1.1. CMS PbPb [47]

No PbPb data has been included in `nCTEQ++` as of yet. The analysis of this set was designed to be a proof of concept test for the validation procedure. It is included here should someone wish to include this data in `nCTEQ++` in the future.

- PDF set: nCTEQ15_208_82

- Cut: $60 \leq m_{34} \leq 120$

- Cut: $-2 \leq y_Z \leq 2$

Leading Order
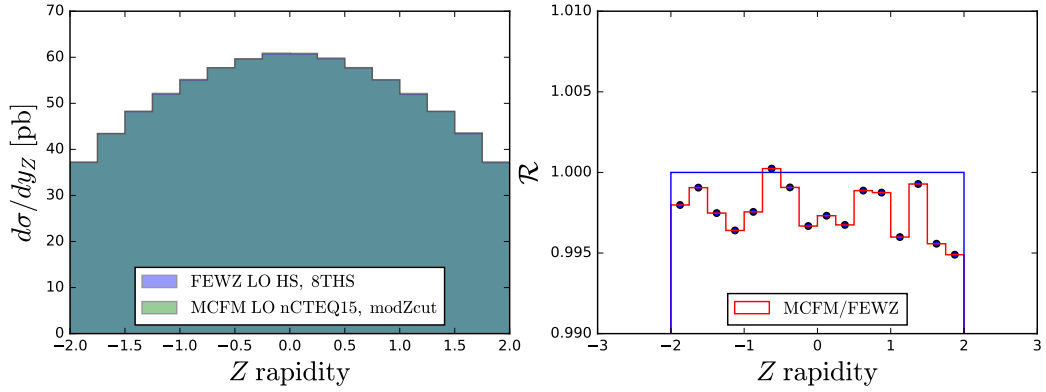
Figure C.1: Differential Cross Section vs Rapidity of Z at LO
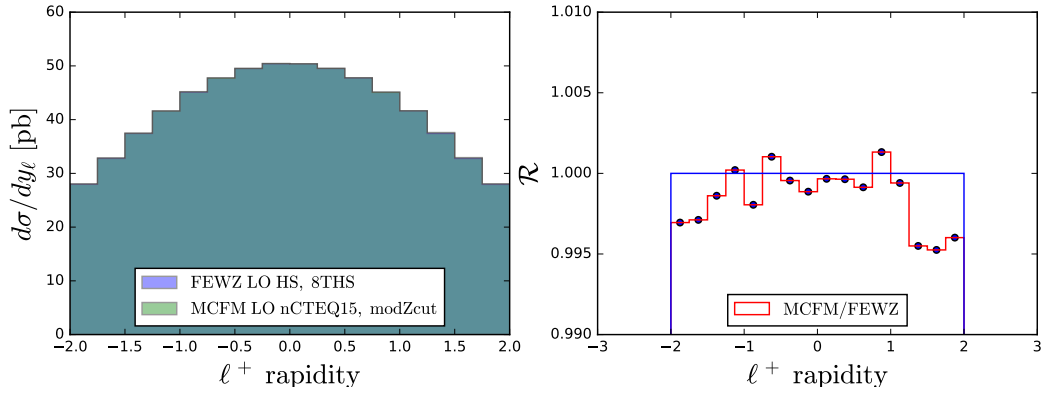
Figure C.2: Differential Cross Section vs Rapidity of ℓ^+ at LO
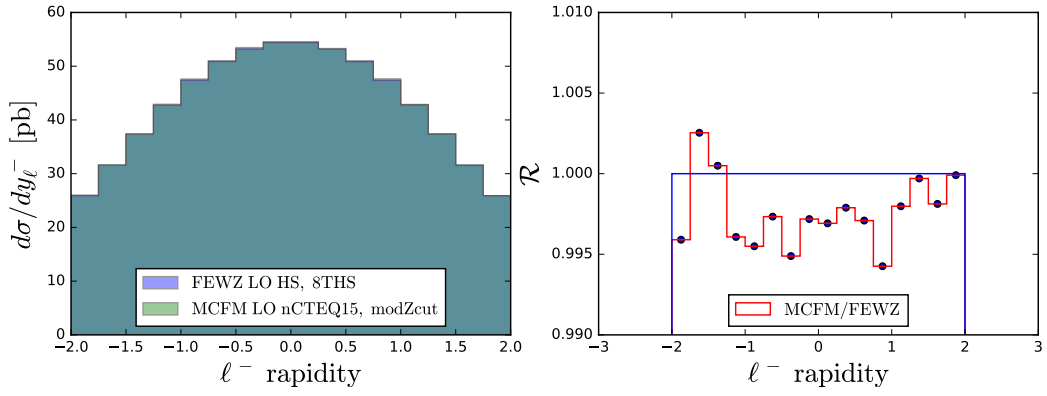
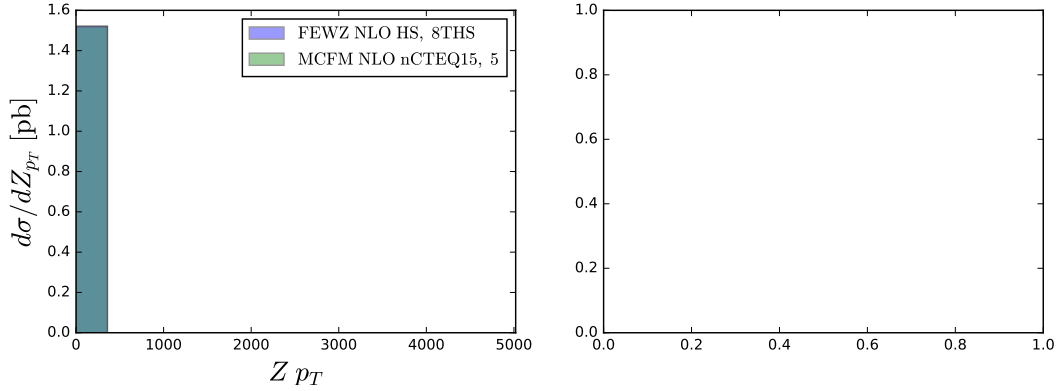Figure C.3: Differential Cross Section vs Rapidity of ℓ^- at LO
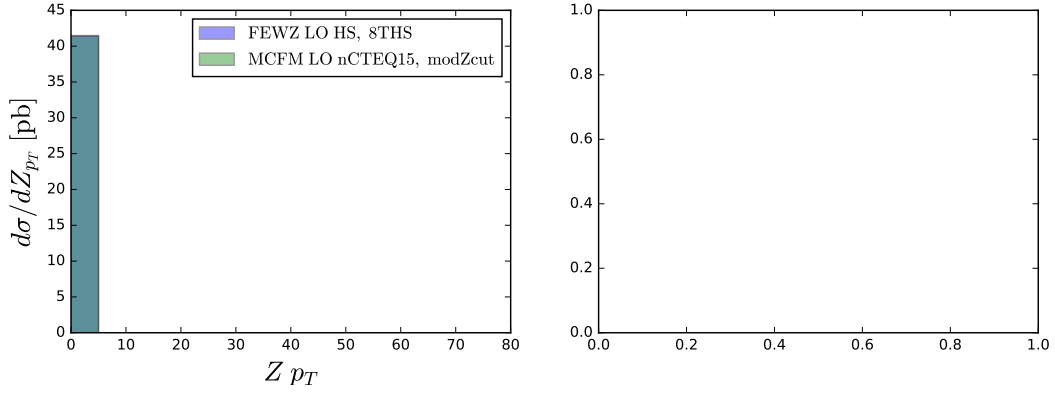
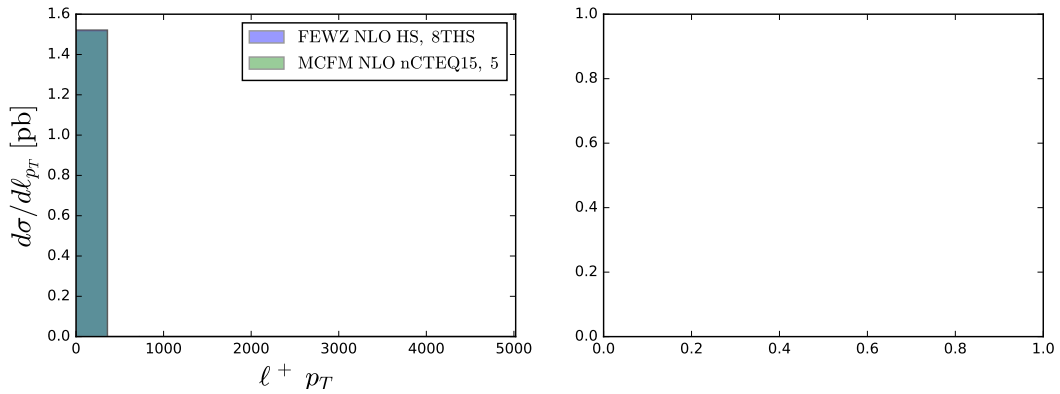
Figure C.4: Differential Cross Section vs p_T of Z at LO
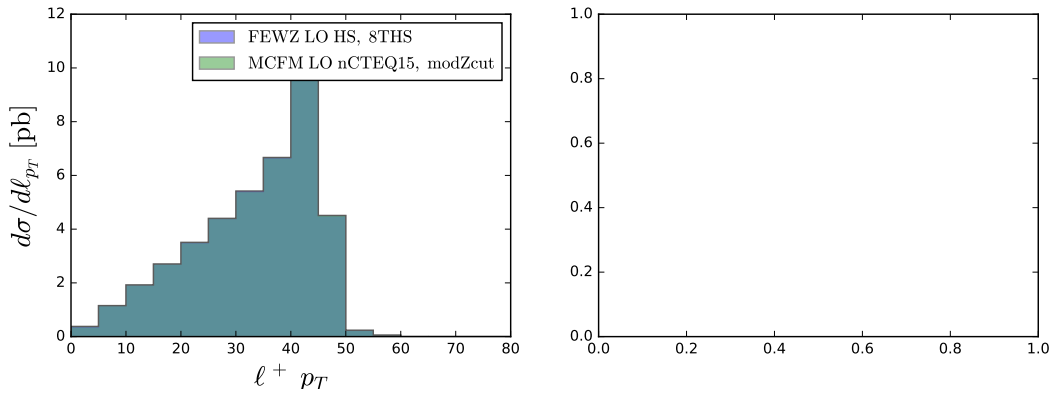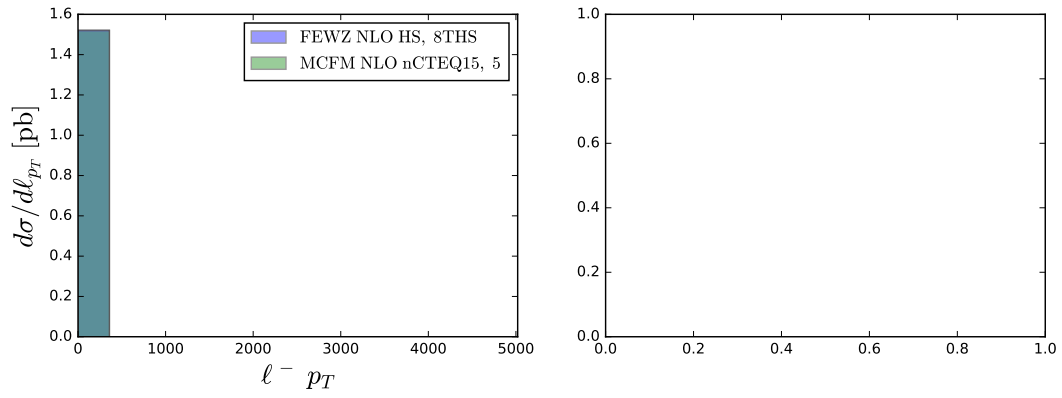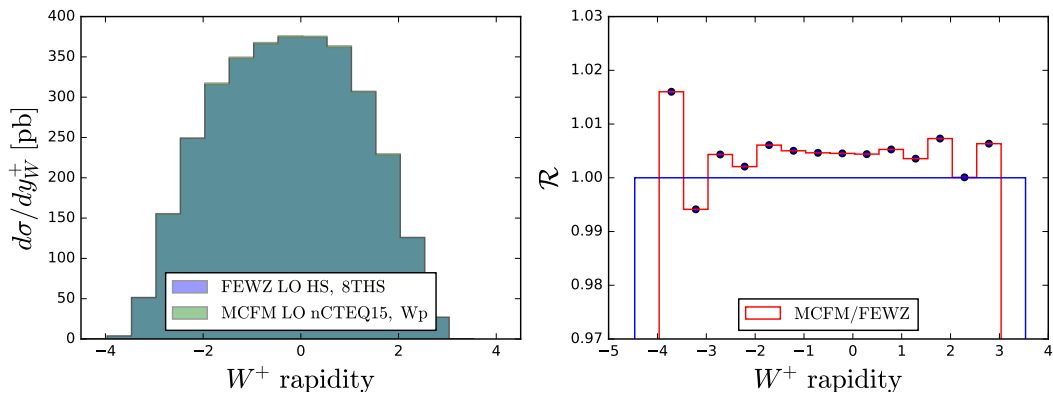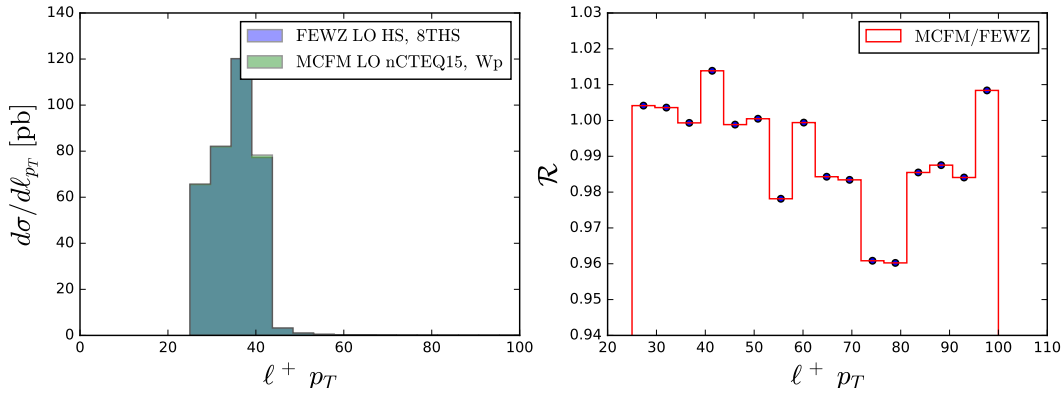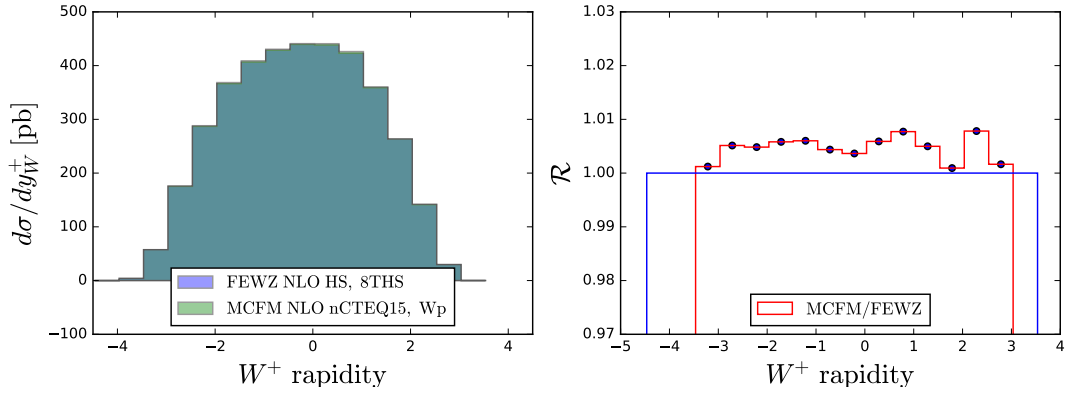

Figure C.5: Differential Cross Section vs p_T of ℓ^+ at LO
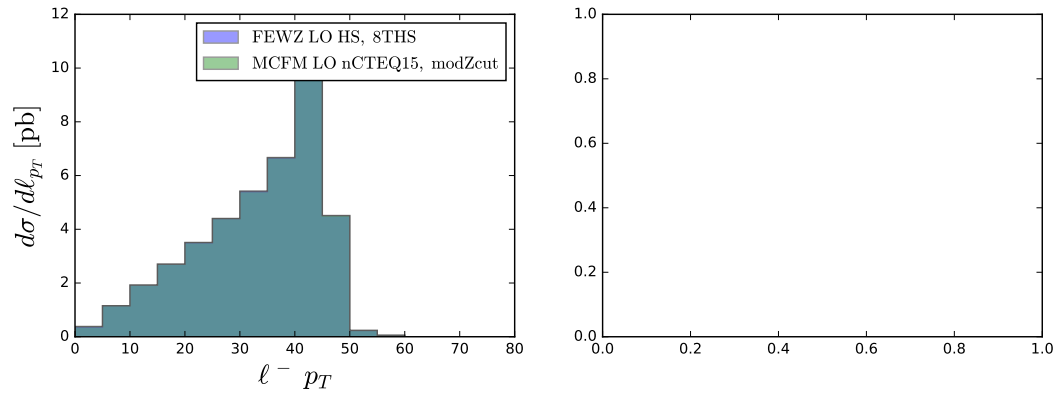
Figure C.6: Differential Cross Section vs p_T of ℓ^- at LO

Next to Leading Order
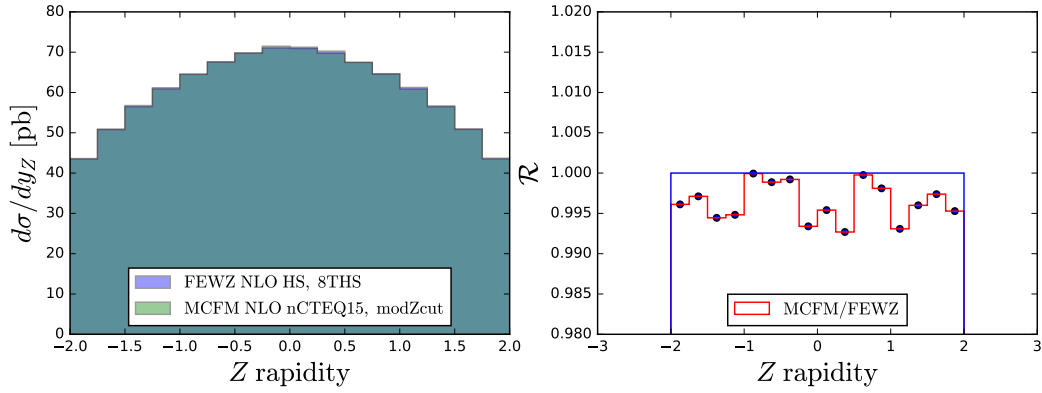
Figure C.7: Differential Cross Section vs Rapidity of Z at NLO
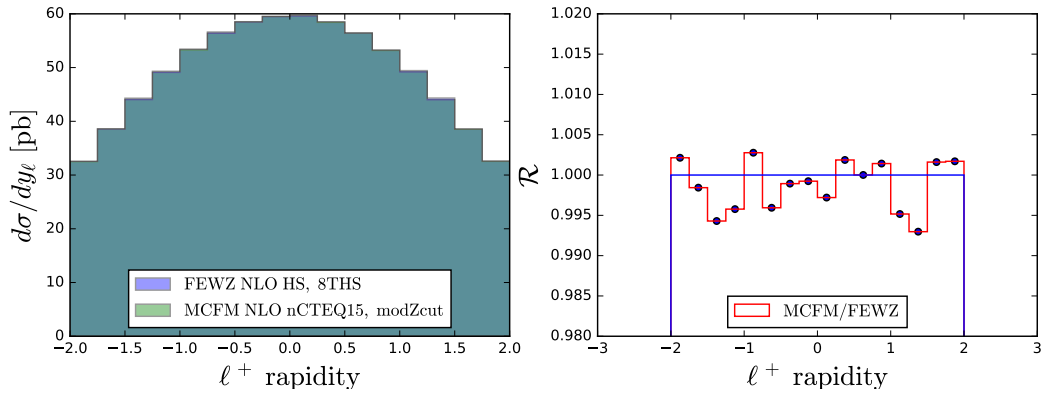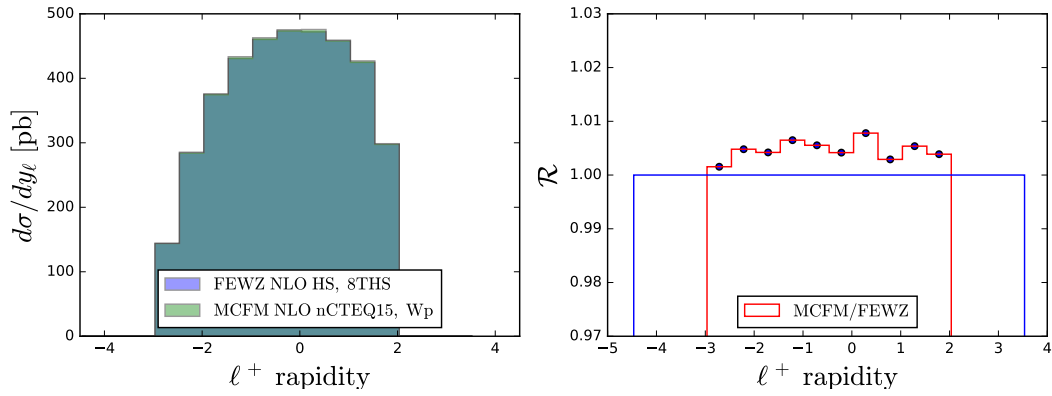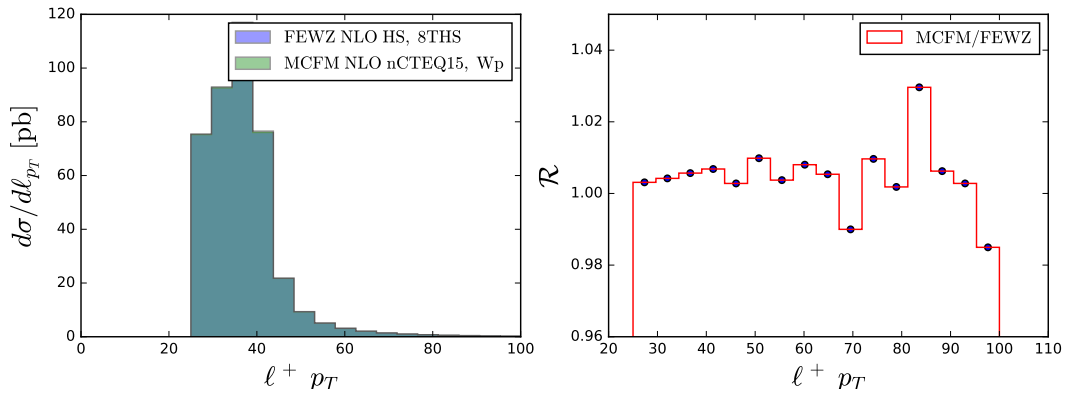

Figure C.8: Differential Cross Section vs Rapidity of ℓ^+ at NLO
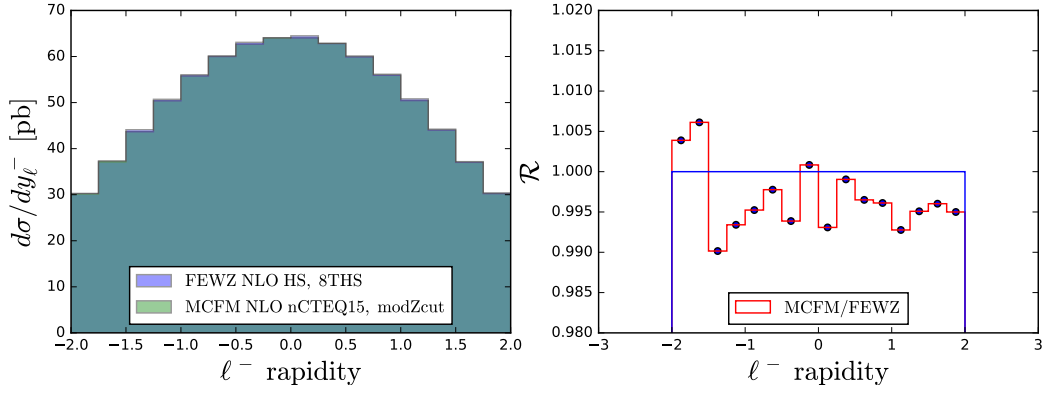
Figure C.9: Differential Cross Section vs Rapidity of ℓ^- at NLO
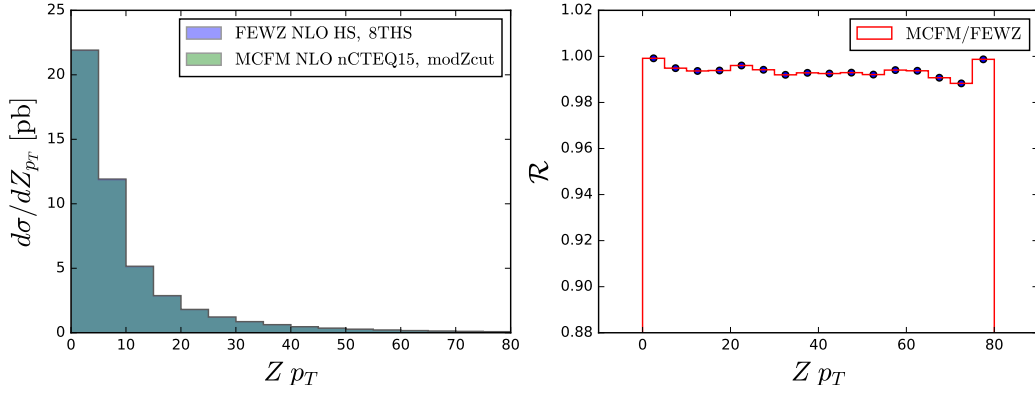
Figure C.10: Differential Cross Section vs p_T of Z at NLO
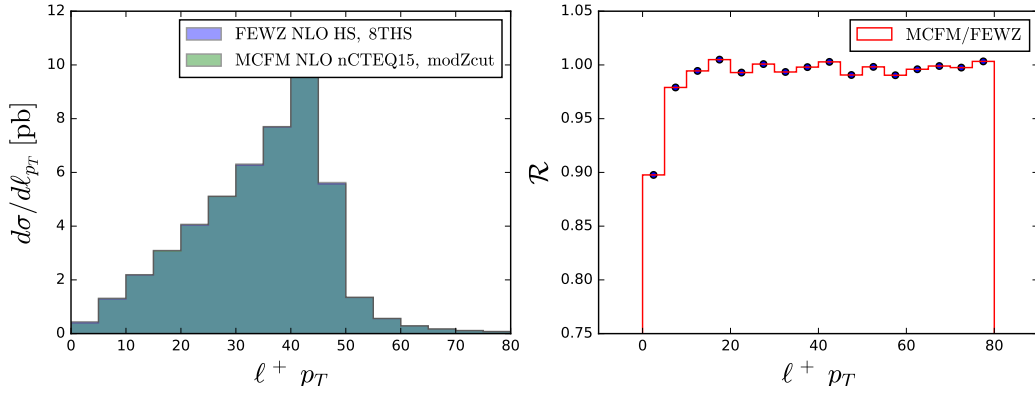
Figure C.11: Differential Cross Section vs p_T of ℓ^+ at NLO
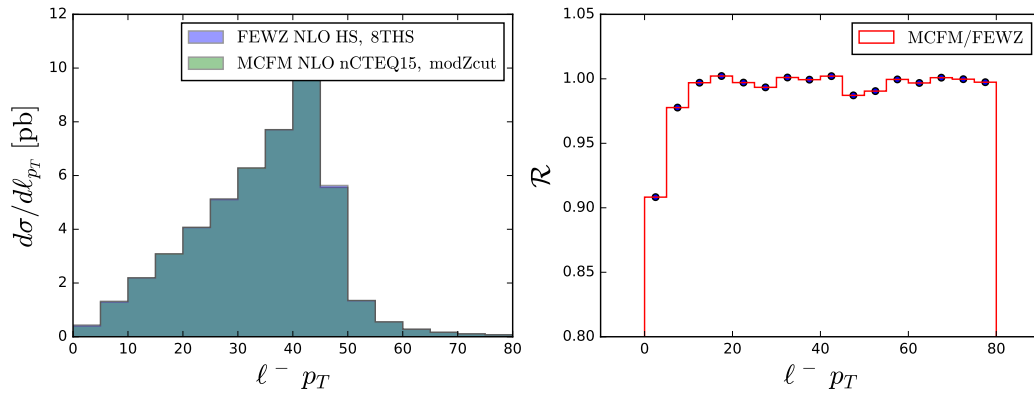
Figure C.12: Differential Cross Section vs p_T of ℓ^- at NLO

- PDF set: nCTEQ15_208_82

- Cut: $60 \leq m_{34} \leq 120$

- Cut: $20 \leq pT_{\ell\pm}$

- Cut: $-2.4 \leq y_Z \leq 2.4$
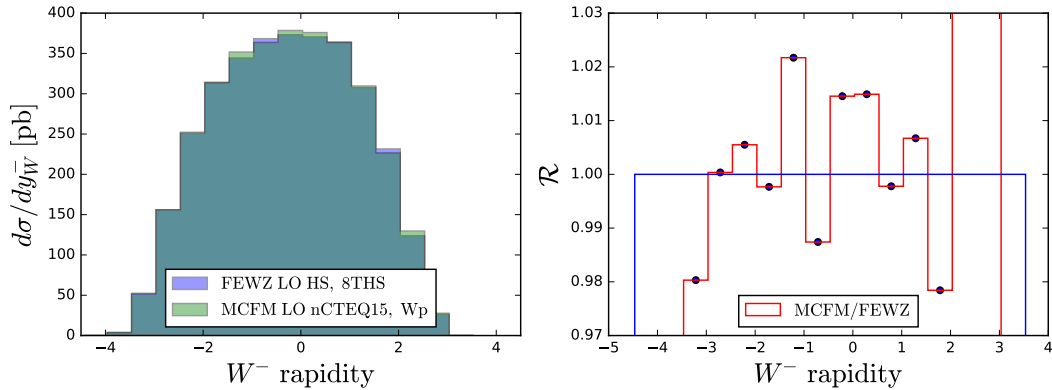
- Rapidity Shift: 0.465

Leading Order
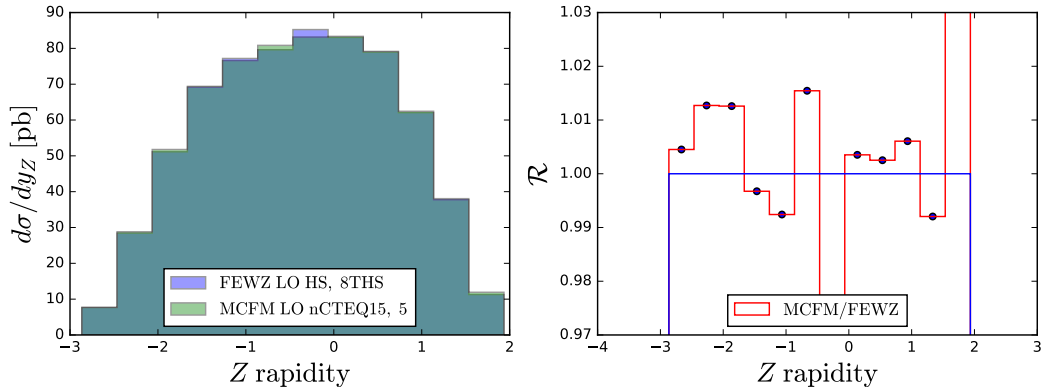
Figure C.13: Differential Cross Section vs Rapidity of Z at LO
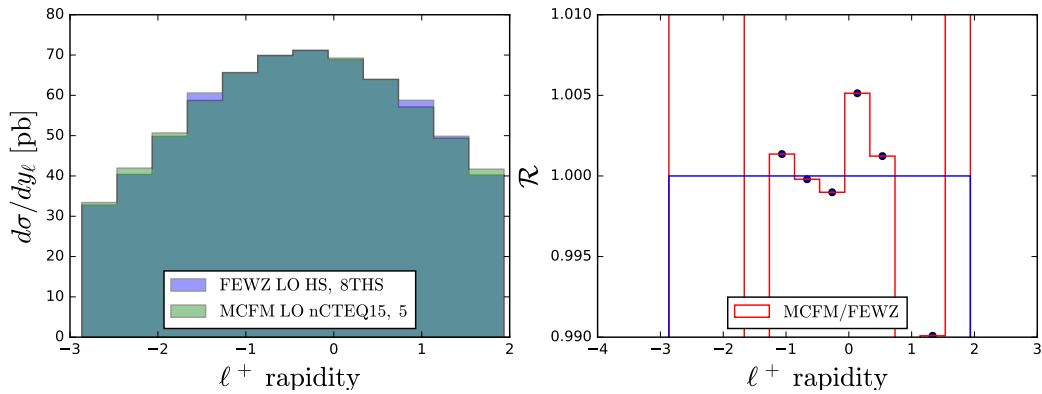
Figure C.14: Differential Cross Section vs Rapidity of ℓ^+ at LO
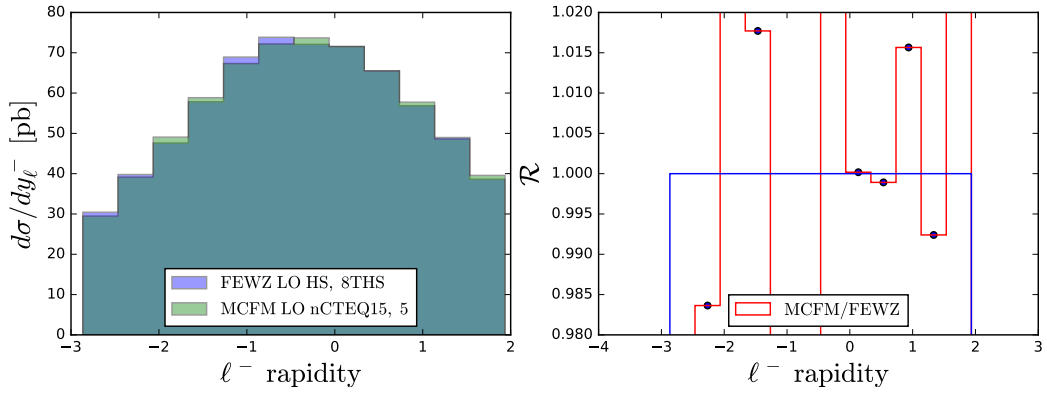
Figure C.15: Differential Cross Section vs Rapidity of ℓ^- at LO
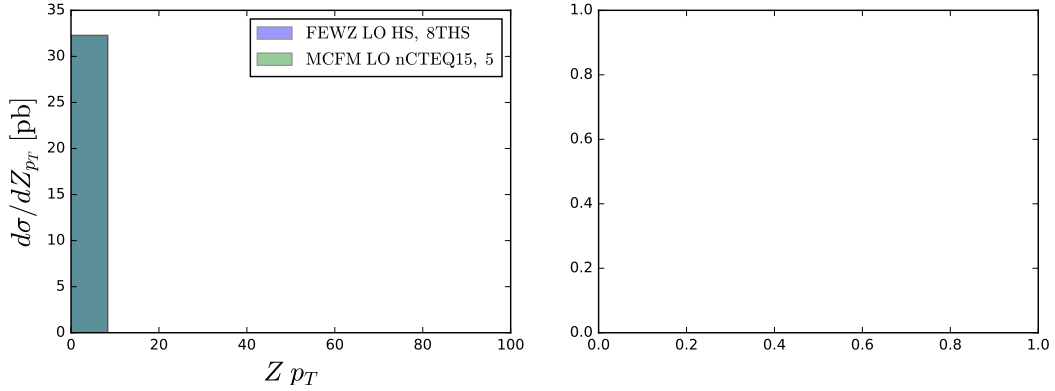
Figure C.16: Differential Cross Section vs p_T of Z at LO
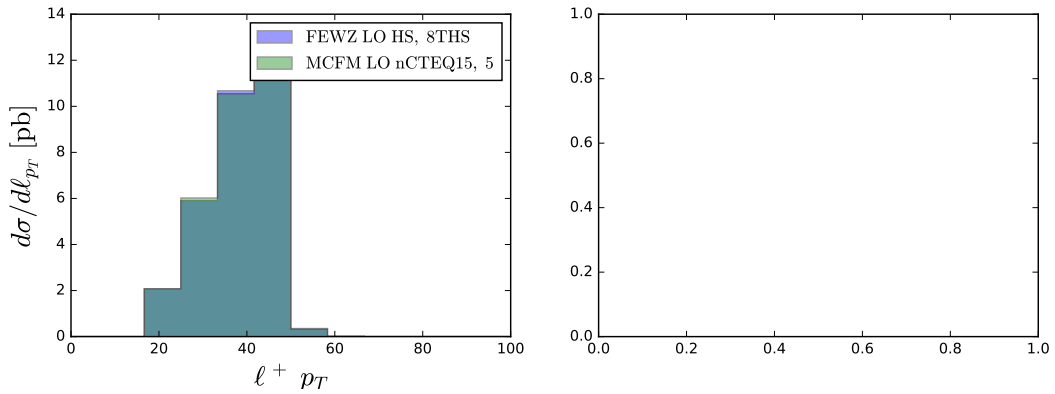
Figure C.17: Differential Cross Section vs p_T of ℓ^+ at LO
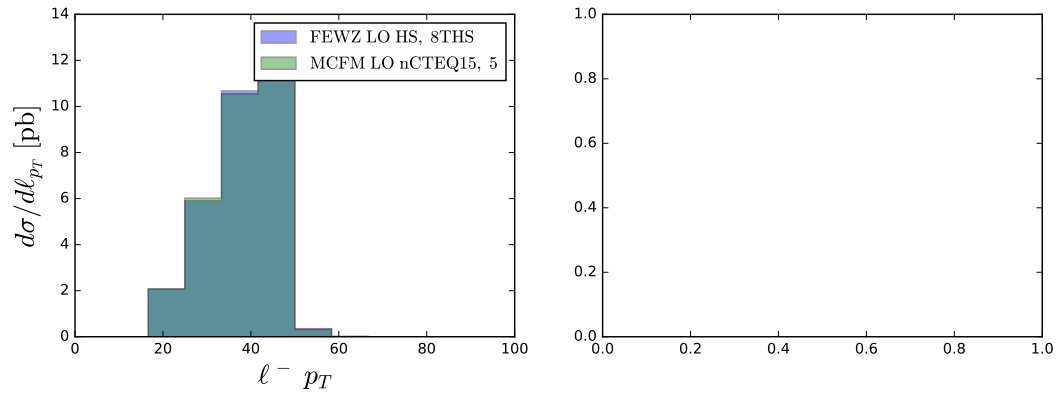
Figure C.18: Differential Cross Section vs p_T of ℓ^- at LO

Next to Leading Order
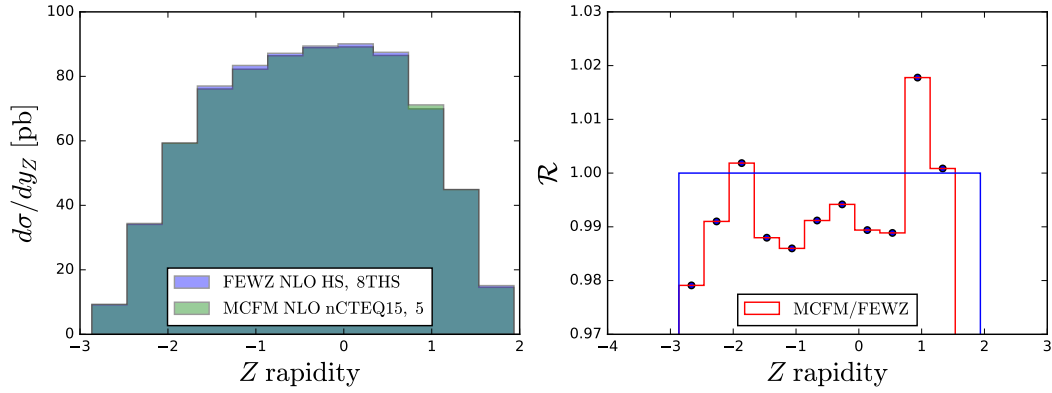
Figure C.19: Differential Cross Section vs Rapidity of Z at NLO
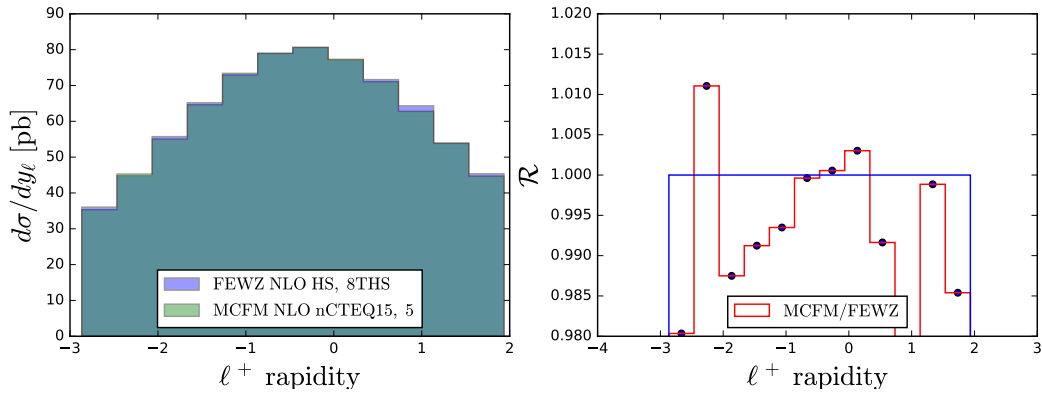
Figure C.20: Differential Cross Section vs Rapidity of ℓ^+ at NLO

225
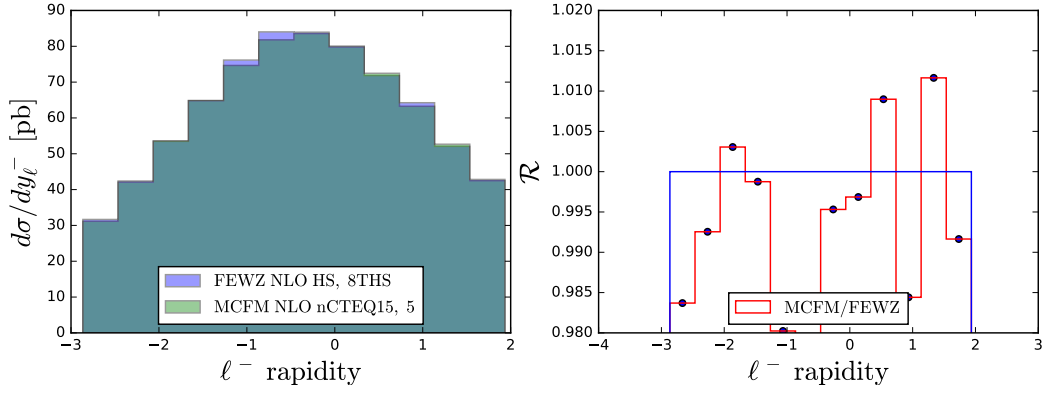
Figure C.21: Differential Cross Section vs Rapidity of ℓ^- at NLO
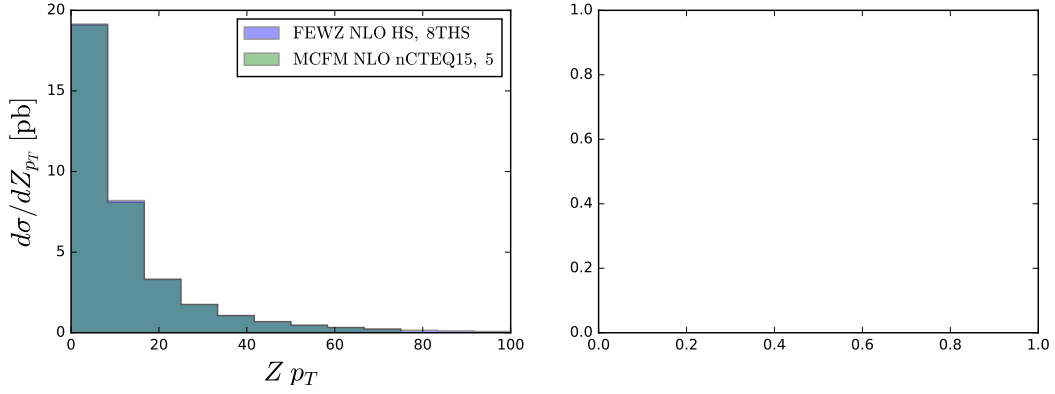
Figure C.22: Differential Cross Section vs p_T of Z at NLO
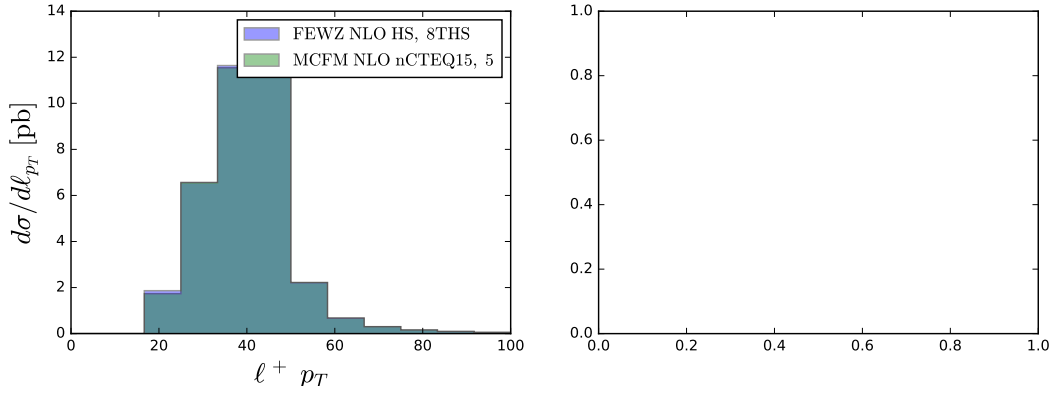
Figure C.23: Differential Cross Section vs p_T of ℓ^+ at NLO

226
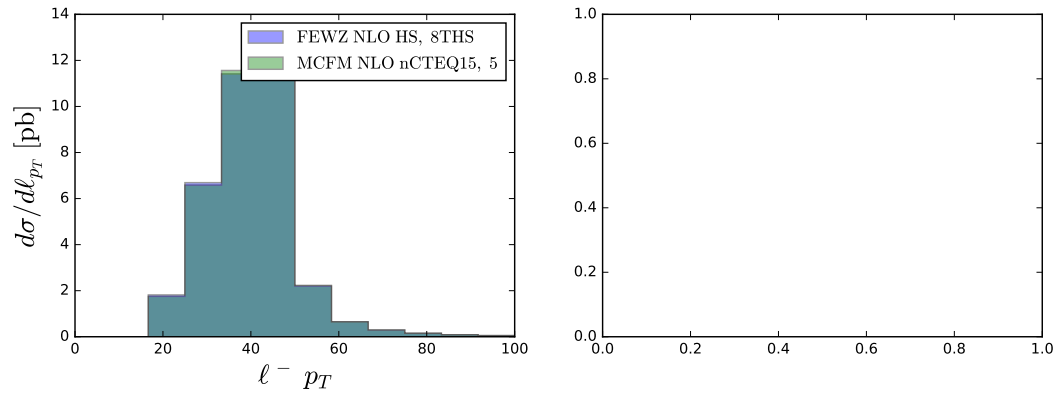
Figure C.24: Differential Cross Section vs p_T of ℓ^- at NLO

- PDF set: nCTEQ15_208_82

- Cut: $66 \leq m_{34} \leq 116$

- Cut: $-3.5 \leq y_Z \leq 3.5$

Leading Order
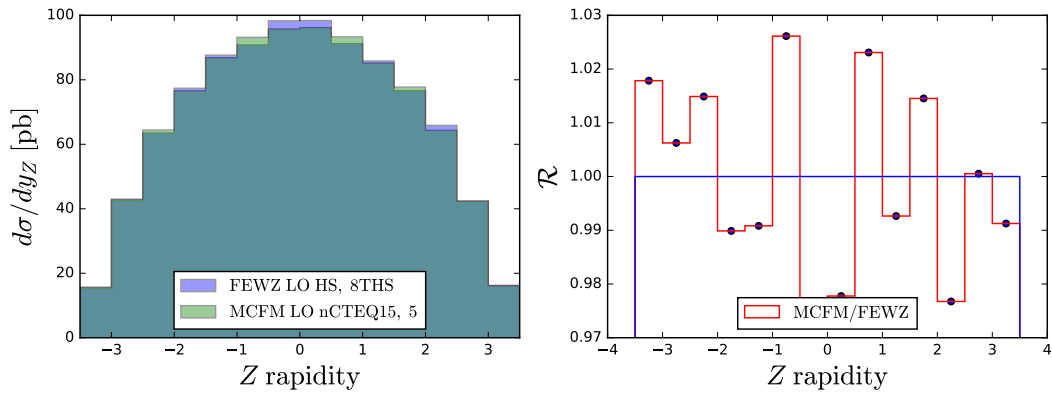
Figure C.25: Differential Cross Section vs Rapidity of Z at LO
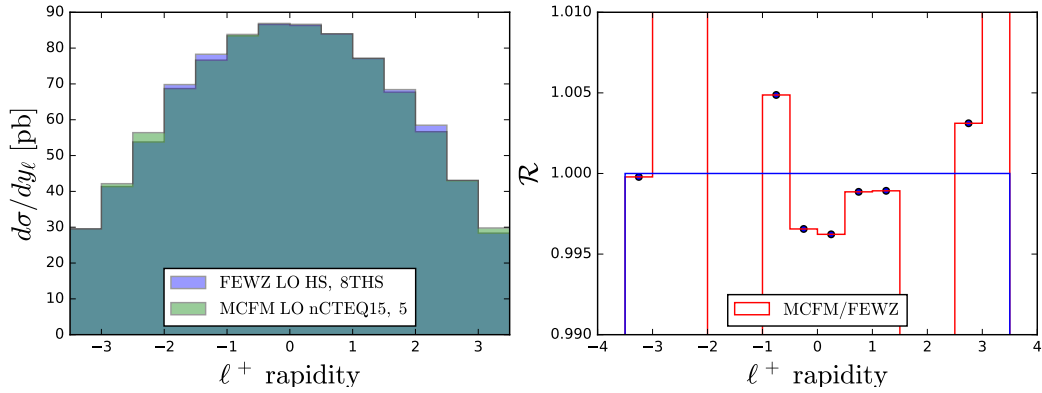
Figure C.26: Differential Cross Section vs Rapidity of ℓ^+ at LO
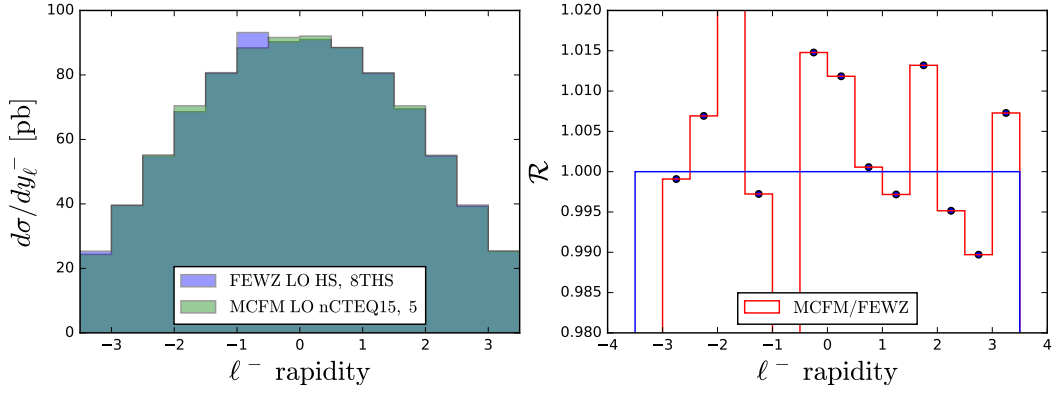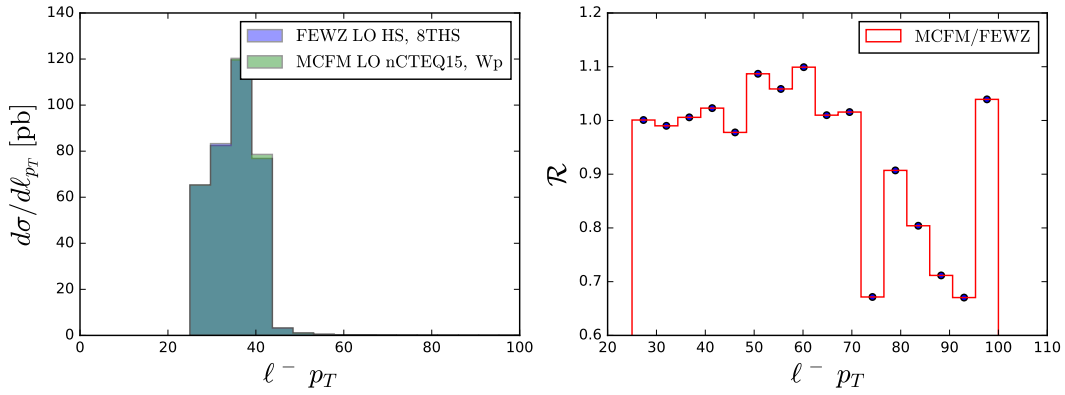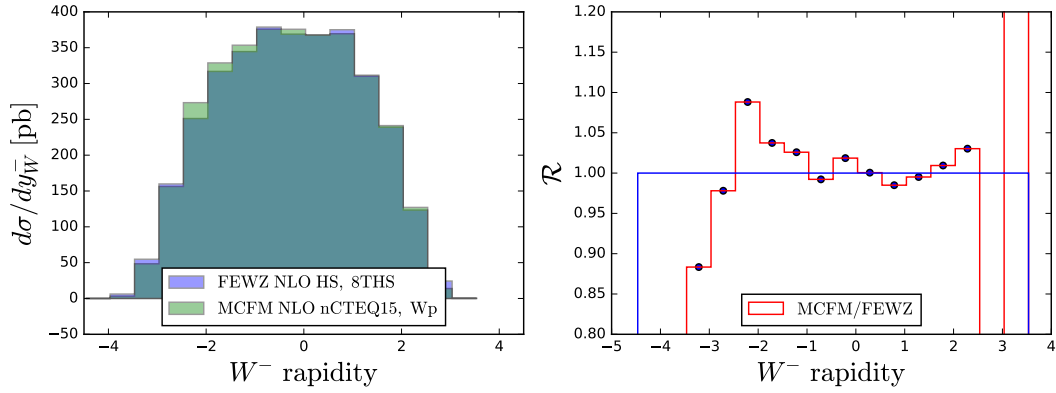
Figure C.27: Differential Cross Section vs Rapidity of ℓ^- at LO
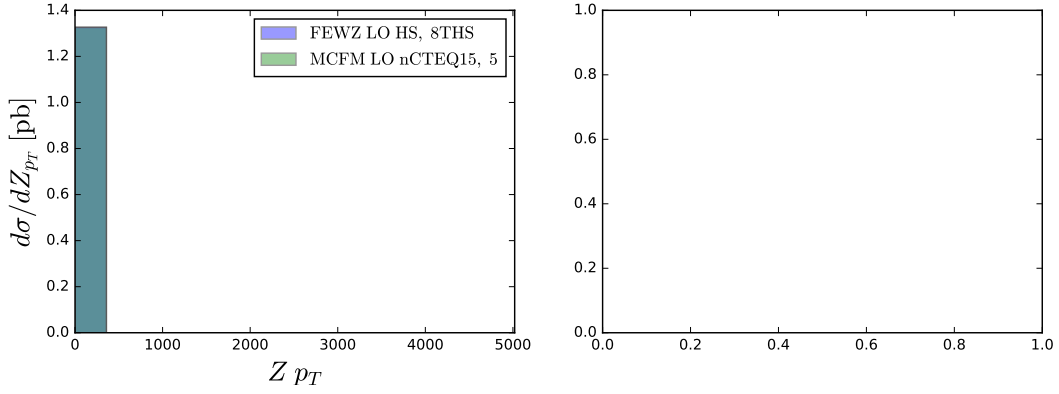
Figure C.28: Differential Cross Section vs p_T of Z at LO
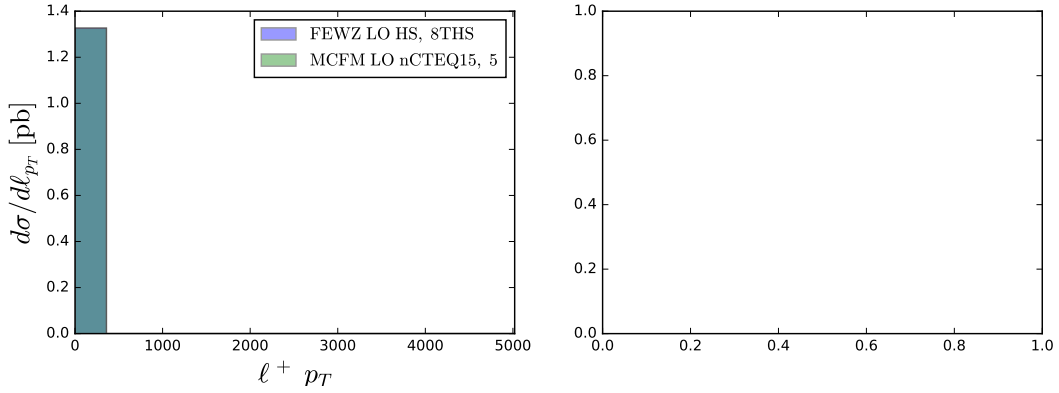
Figure C.29: Differential Cross Section vs p_T of ℓ^+ at LO
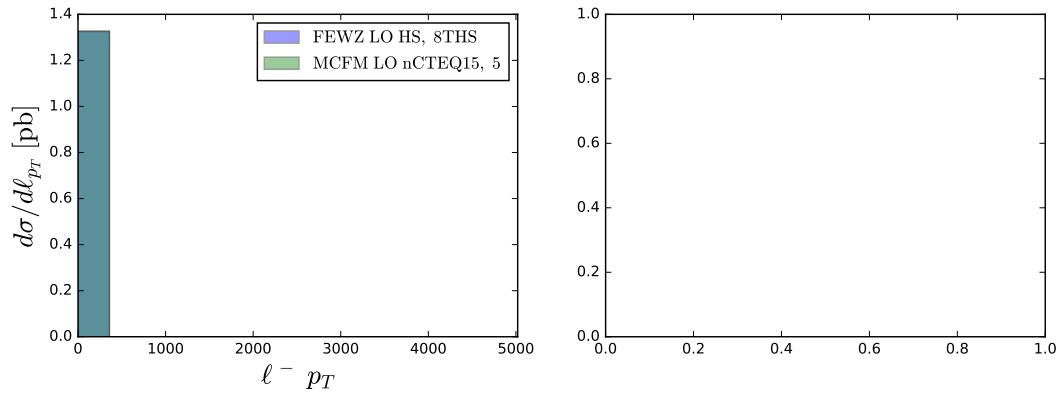
Figure C.30: Differential Cross Section vs p_T of ℓ^- at LO

Next to Leading Order
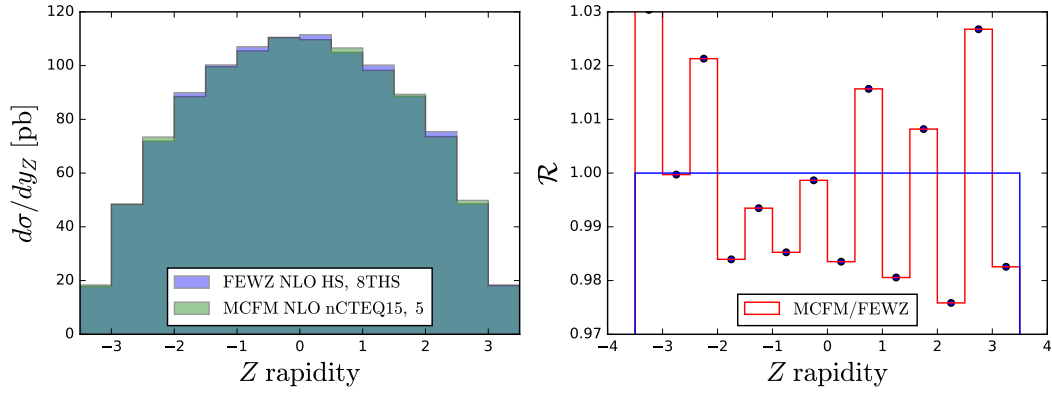
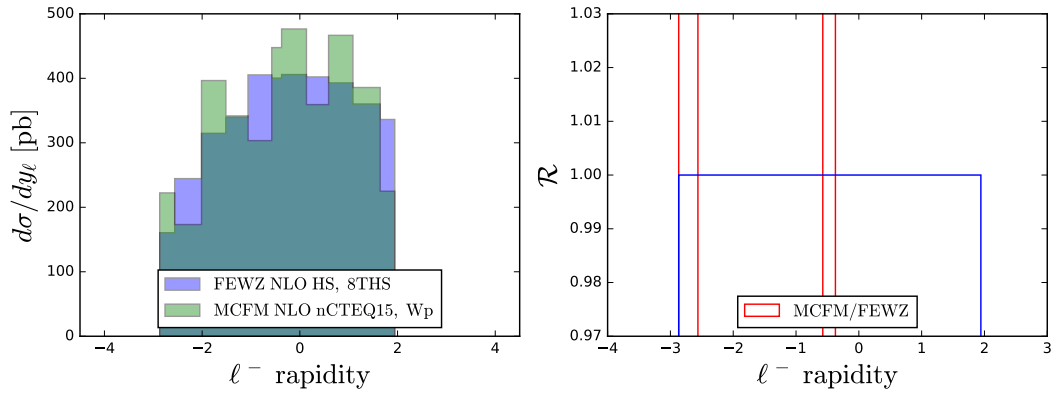Figure C.31: Differential Cross Section vs Rapidity of Z at NLO
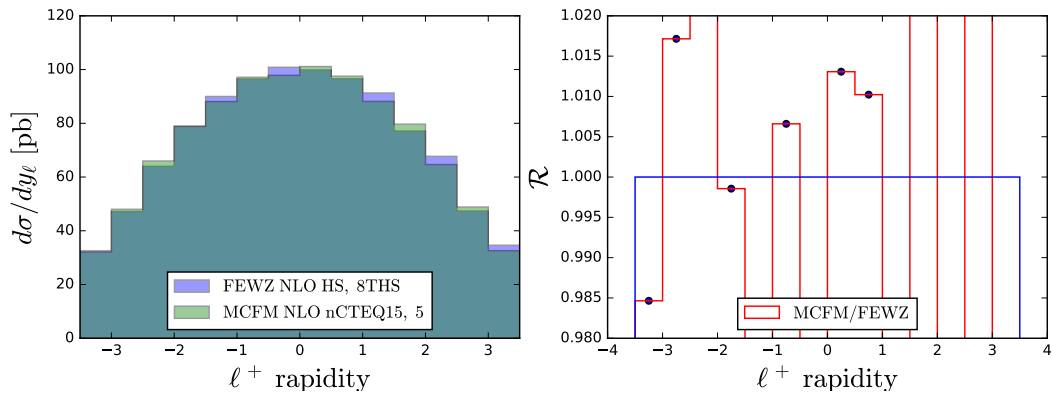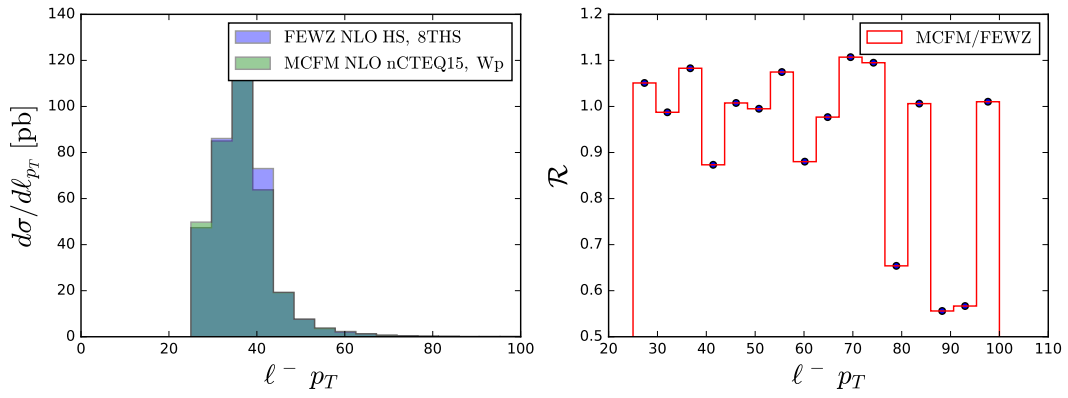

Figure C.32: Differential Cross Section vs Rapidity of ℓ^+ at NLO
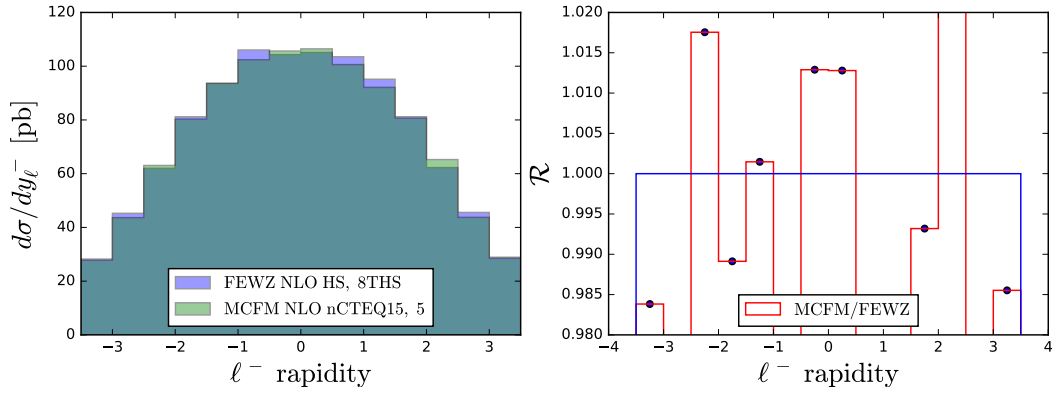
Figure C.33: Differential Cross Section vs Rapidity of ℓ^- at NLO

Figure C.34: Differential Cross Section vs p_T of Z at NLO

Figure C.35: Differential Cross Section vs p_T of ℓ^+ at NLO

Figure C.36: Differential Cross Section vs p_T of ℓ^- at NLO

C.1.2. W^+

$$pPb \rightarrow W^+ \rightarrow \nu + \ell^+$$

C.1.2.1.   CMS pPb [42]

- PDF set: nCTEQ15_208_82

- Cut: $25 \leq pT^{\ell+}$

- Cut: $-2.4 \leq \eta_\pm \leq 2.4$

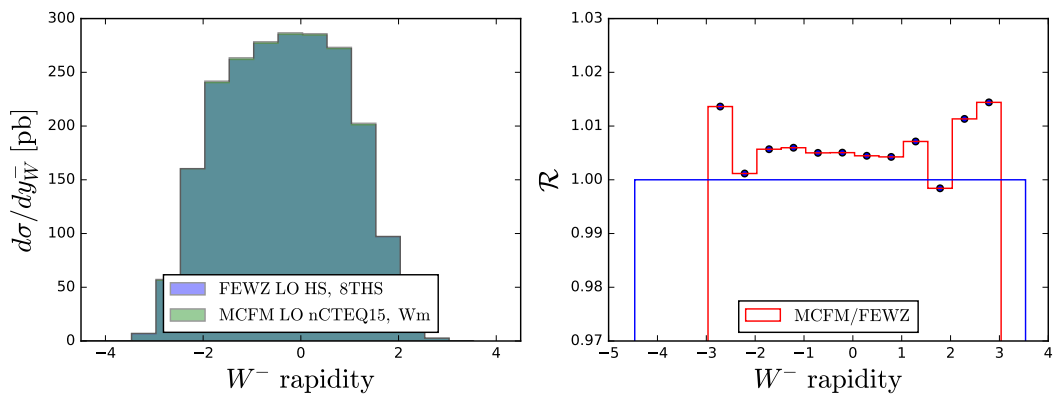- Rapidity Shift: 0.465

Leading Order

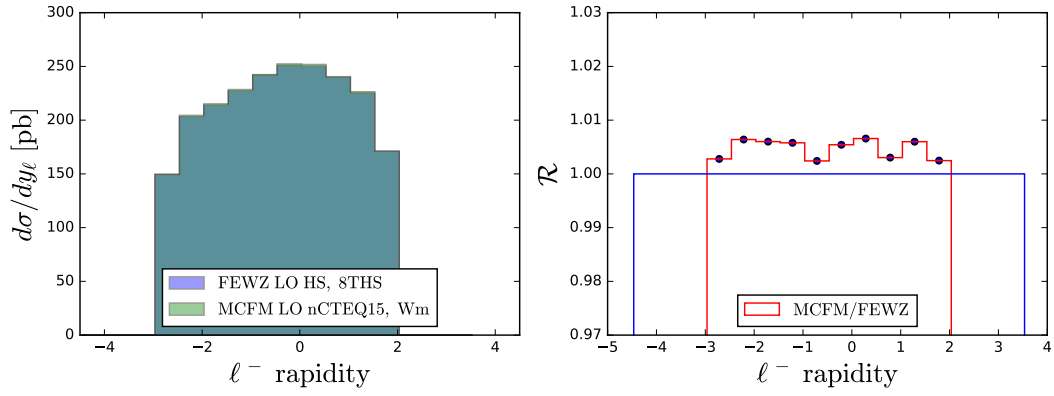Figure C.37: Differential Cross Section vs Rapidity of W^+ at LO
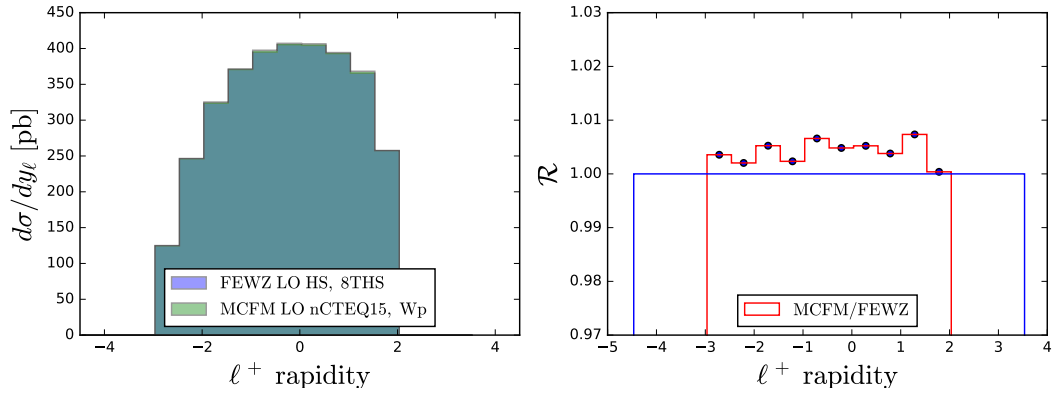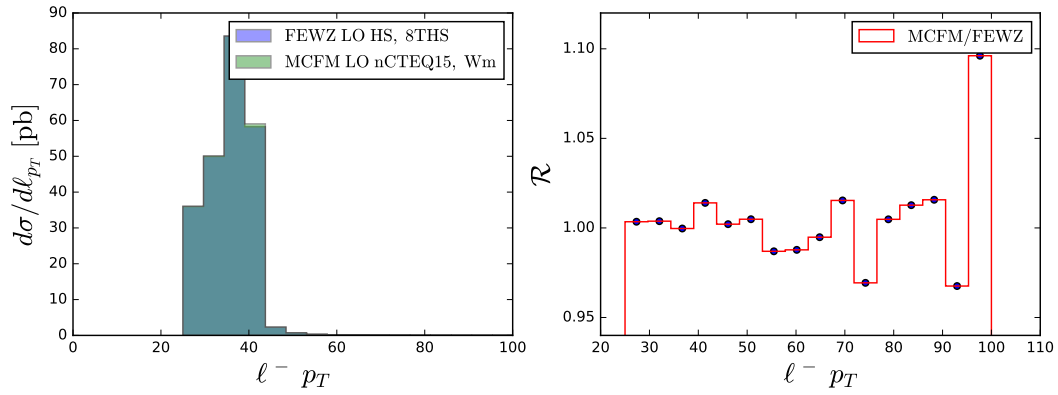
Figure C.38: Differential Cross Section vs Rapidity of ℓ^+ at LO

Figure C.39: Differential Cross Section vs p_T of ℓ^+ at LO
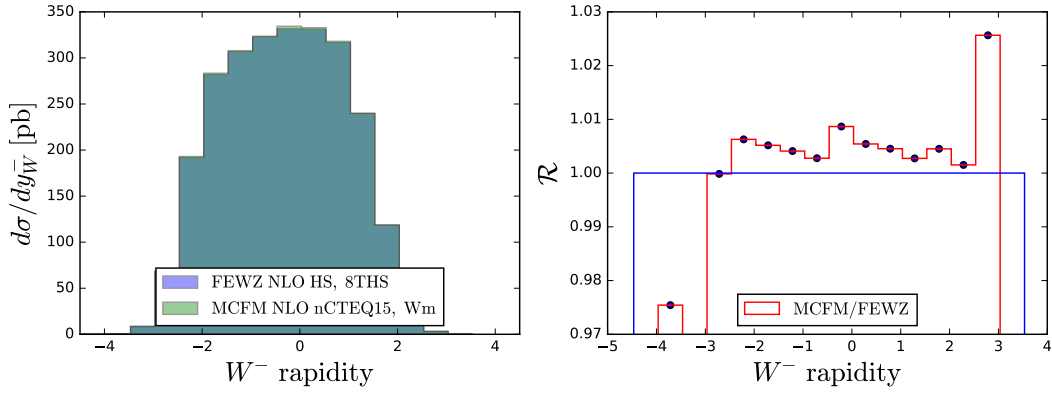
Next to Leading Order

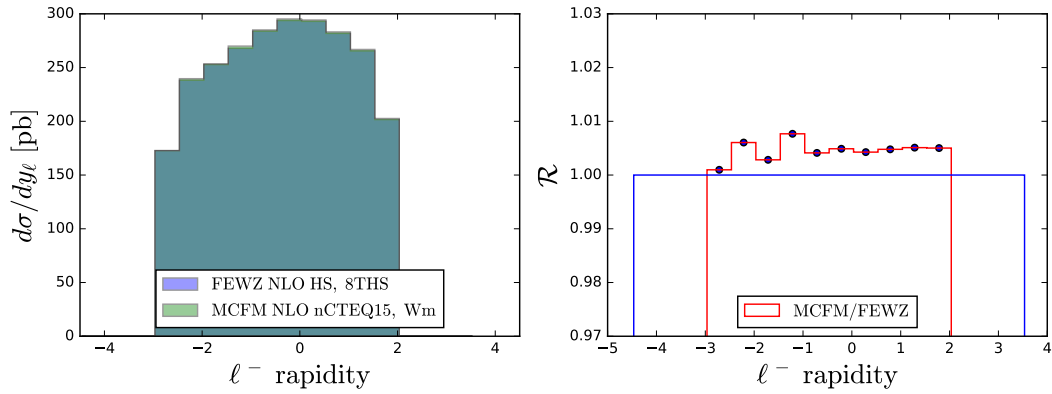Figure C.40: Differential Cross Section vs Rapidity of W^+ at NLO
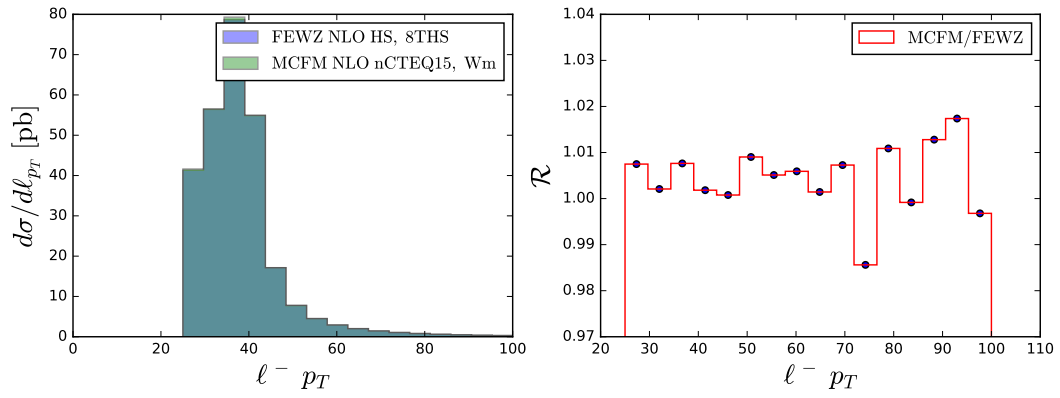

Figure C.41: Differential Cross Section vs Rapidity of ℓ^+ at NLO

Figure C.42: Differential Cross Section vs p_T of ℓ^+ at NLO

This data set proved incredibly difficult for `MCFM` to handle as the original data is missing a bin at central rapidity. There exists a possible modification to `MCFM` that can be made to handle this irregular binning but it was unable to be implemented into the code at the time. In principle, as the `mcmf-bridge` interface between `APPLgrid` and `MCFM` provides its own binning for the `MCFM` events, there should not be an issue implementing this data set into `nCTEQ++`. As of this writing, that validation has not been performed and the data set has not been included.

- PDF set: nCTEQ15_208_82

- Cut: $40 \leq mT^{\ell+}$

- Cut: $25 \leq pT^{\ell+}$

- Cut: $-2.4 \leq \eta_{\pm} \leq 2.4$

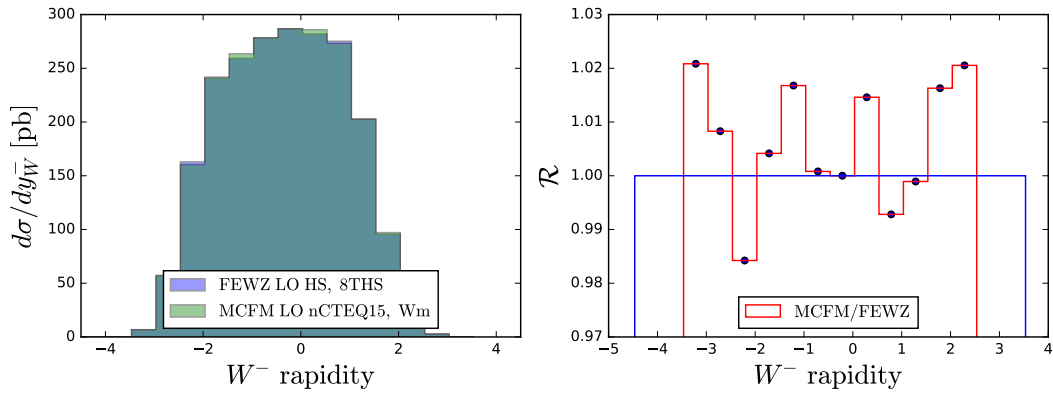- Rapidity Shift: 0.465

Leading Order

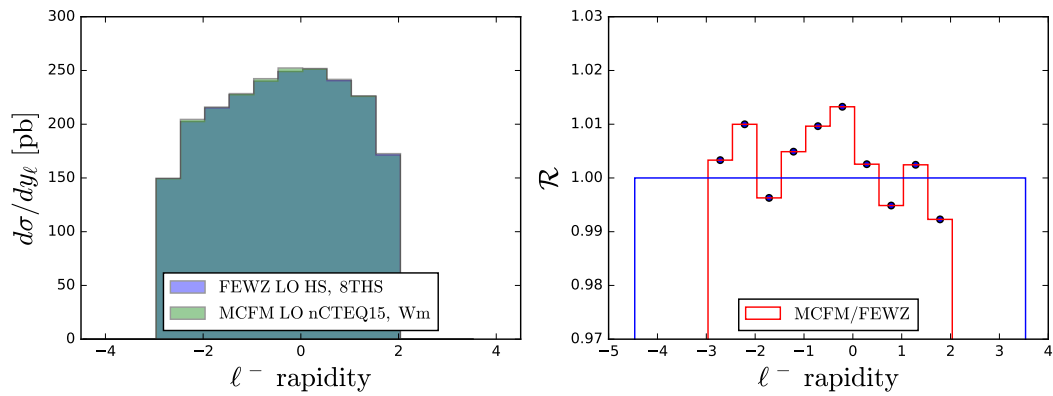Figure C.43: Differential Cross Section vs Rapidity of W^+ at LO
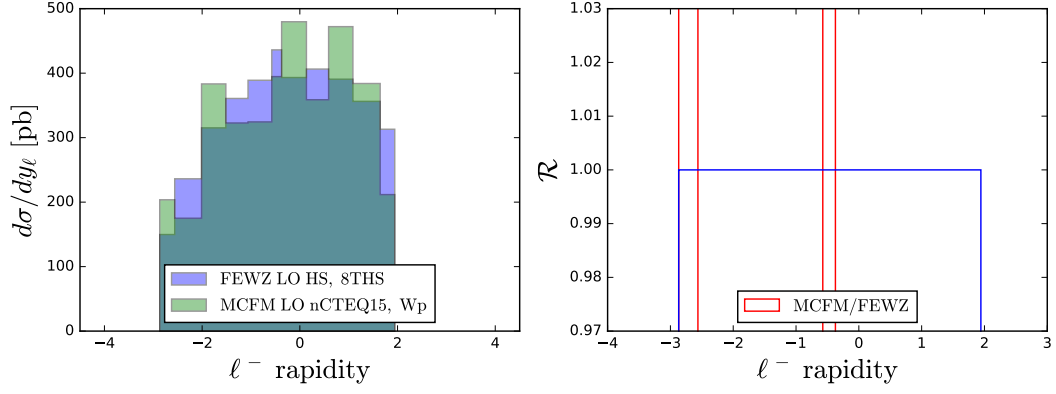
Figure C.44: Differential Cross Section vs Rapidity of ℓ^- at LO

Figure C.45: Differential Cross Section vs p_T of ℓ^- at LO

Next to Leading Order

Figure C.46: Differential Cross Section vs Rapidity of W^+ at NLO
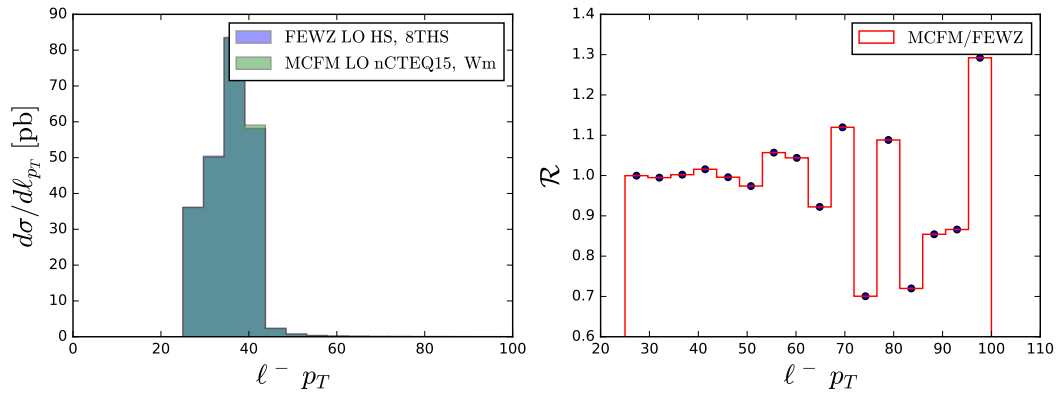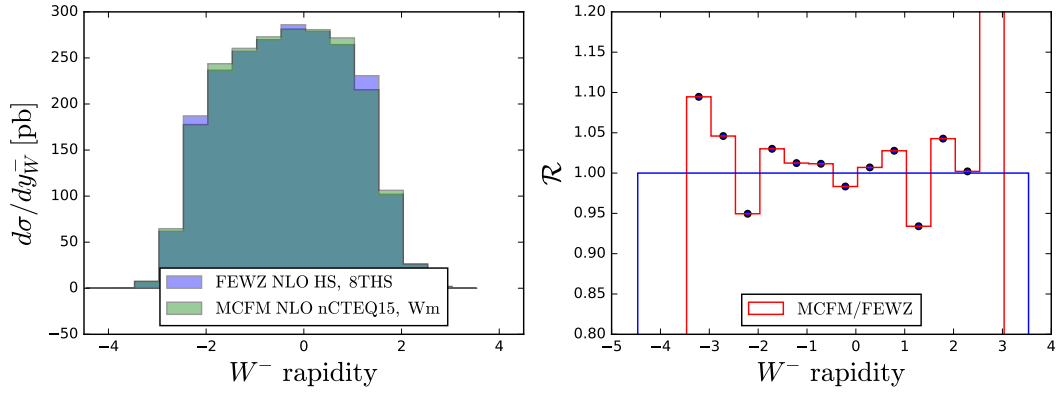

Figure C.47: Differential Cross Section vs Rapidity of ℓ^- at NLO

Figure C.48: Differential Cross Section vs p_T of ℓ^- at NLO

C.1.3. W^-

$$pPb \to W^- \to \nu + \ell^-$$

C.1.3.1. CMS pPb [42]

- PDF set: nCTEQ15_208_82

- Cut: $25 \leq pT^{\ell^-}$

- Cut: $-2.4 \leq \eta_\pm \leq 2.4$

- Rapidity Shift: 0.465

Leading Order

Figure C.49: Differential Cross Section vs Rapidity of W^- at LO

Figure C.50: Differential Cross Section vs Rapidity of ℓ^- at LO

Figure C.51: Differential Cross Section vs p_T of ℓ^- at LO

Next to Leading Order

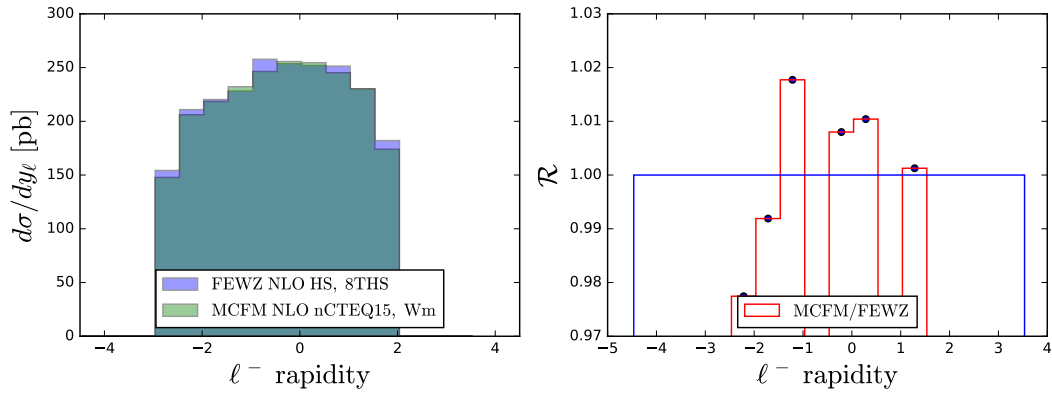Figure C.52: Differential Cross Section vs Rapidity of W^- at NLO
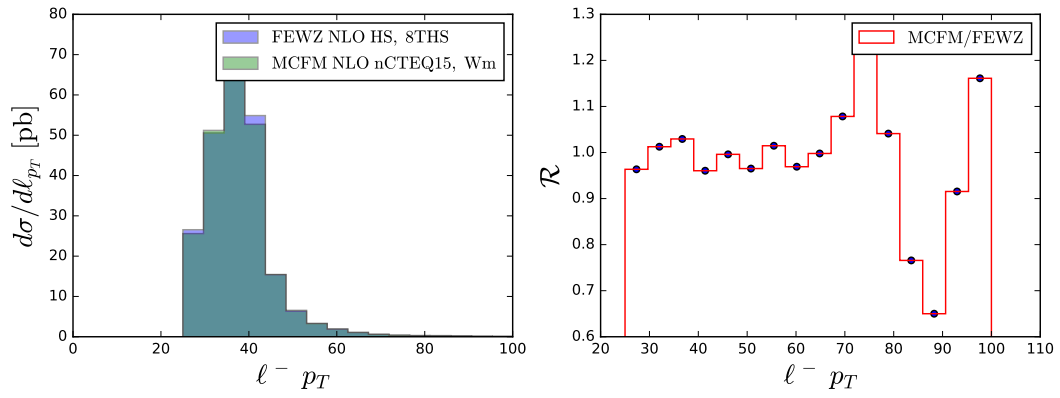

Figure C.53: Differential Cross Section vs Rapidity of ℓ^- at NLO

Figure C.54: Differential Cross Section vs p_T of ℓ^- at NLO

- PDF set: nCTEQ15_208_82

- Cut: $40 \leq mT^{\ell^-}$

- Cut: $25 \leq pT^{\ell^-}$

- Cut: $-2.4 \leq \eta_\pm \leq 2.4$

- Rapidity Shift: 0.465

Leading Order

Figure C.55: Differential Cross Section vs Rapidity of W^- at LO

Figure C.56: Differential Cross Section vs Rapidity of ℓ^- at LO

Figure C.57: Differential Cross Section vs p_T of ℓ^- at LO

Next to Leading Order

Figure C.58: Differential Cross Section vs Rapidity of W^- at NLO

Figure C.59: Differential Cross Section vs Rapidity of ℓ^- at NLO

Figure C.60: Differential Cross Section vs p_T of ℓ^- at NLO

Appendix D

Analysis of PDF Dependence for MCFM produced APPLgrid Grids

This Appendix contains a comprehensive analysis of the dependence of APPLgrid grids on the underlying PDF used in the MCFM events from which they are generated. This appendix follows from Section 5.3.4.

D.1. Overview

For this analysis of the PDF dependence of grids generated using APPLgrid linked through the mcfm-bridge to MCFM, MCFM was set for W^+ to a lepton and a neutrino with kinematic cuts:

$$-2.865 < y_{lep} < 1.935 \qquad\qquad p_{Tlep} > 25$$

The grids used in this study are for the lepton rapidity (eta4). There are three runs compared below. First are solo grids; these are produced by running MCFM with APPLgrid twice, once to initialize the grid and a second time to fill them. Second, three grids were combined using an APPLgrid utility to produce a grid that is less dependent on the underlying PDF. The three grids were generated using the same inputs from MCFM but utilize a different random seed. Finally, the process that was done for three grids was replicated for ten grids, further decreasing the dependence on the underlying PDF. If there was no PDF dependence in the grid then one would expect the grids to have differences on the level of the numerical noise in the calculation (0.01%).

Since the grids are to be utilized in calculating observables, they were convoluted with different PDFs when determining their PDF dependence. For this exercise, the following PDF sets were used. The numerical label refers to how the PDF is identified in the labels in the plots.

Set	Label	`MCFM` Run	Cite
NNPDF23_nlo_as_0118	1	Yes	[156]
CT10nlo_as_0118	2	Yes	[31]
nCTEQ15_208_82	3	No	[7]

D.2. `MCFM` Output

PDF: NNPDF23_nlo_as_0118

- PreConditioning Itterations: 15

- Calls: 100,000

- Final Run Itterations: 30

- Calls: 500,000

Run	Cross Section	Percent Error
0	2283432 ± 283	.01%
1	2283838 ± 284	.01%
2	2283835 ± 285	.01%

PDF: CT10nlo_as_0118

- PreConditioning Itterations: 15

- Calls: 100,000

- Final Run Itterations: 30

- Calls: 500,000

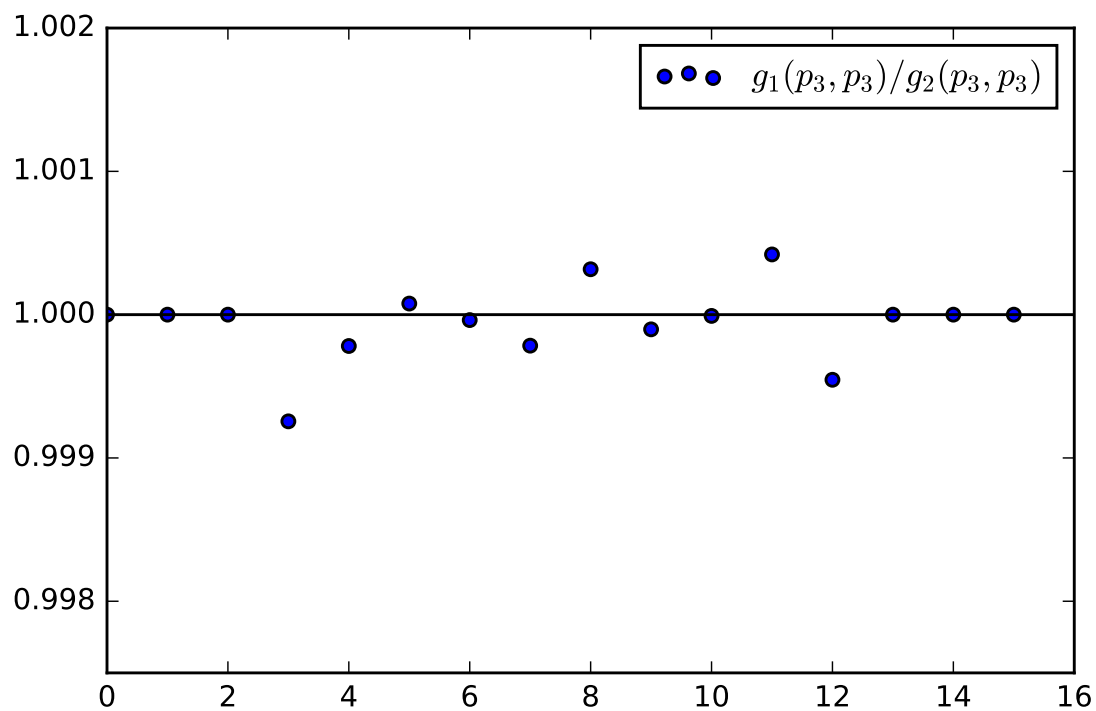| Run	Cross Section	Percent Error
0	2319380 ± 289	.01%
1	2319594 ± 289	.01%
2	2319676 ± 293	.01%

D.3. Ratio Plots

D.3.1. Uncombined Grids
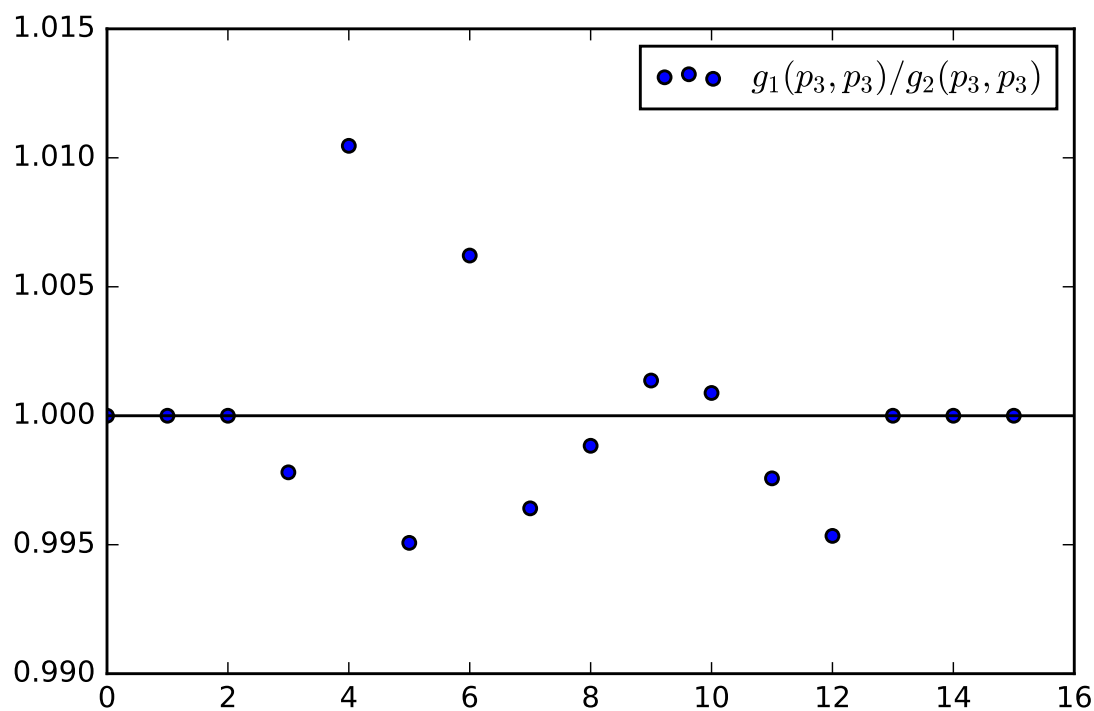
Figure D.1: Ratio of NNPDF grid to CT10 grid both convoluted with nCTEQ_208_82 pdf
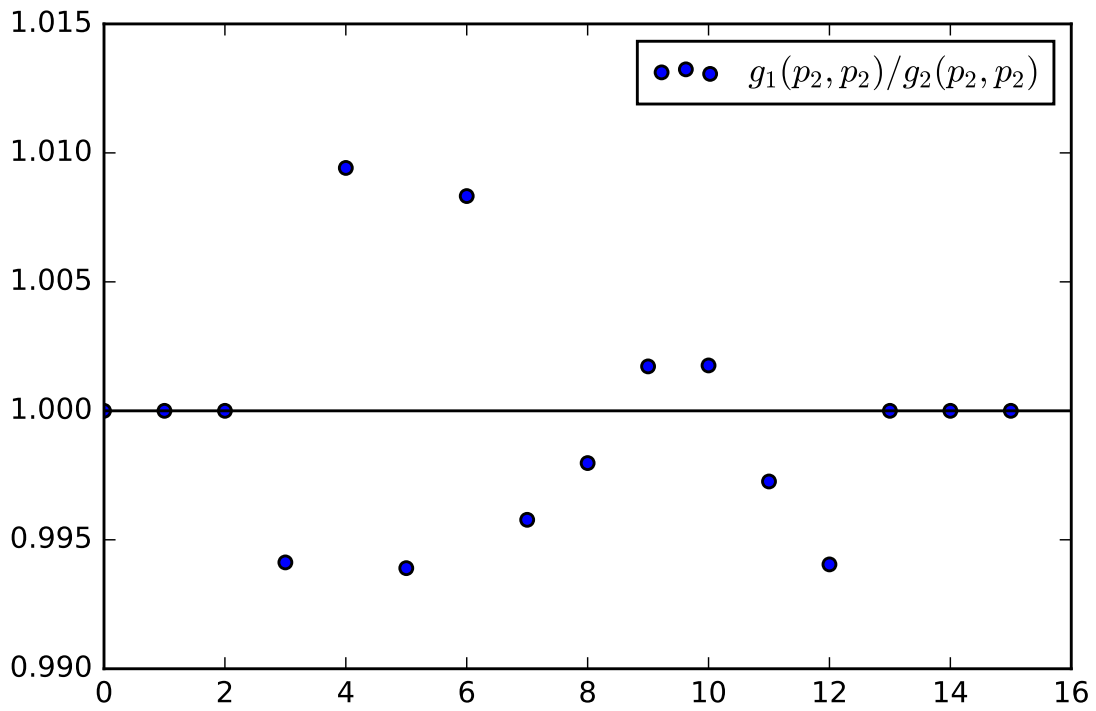
Figure D.2: Ratio of NNPDF grid to CT10 grid both convoluted with CT10 pdf
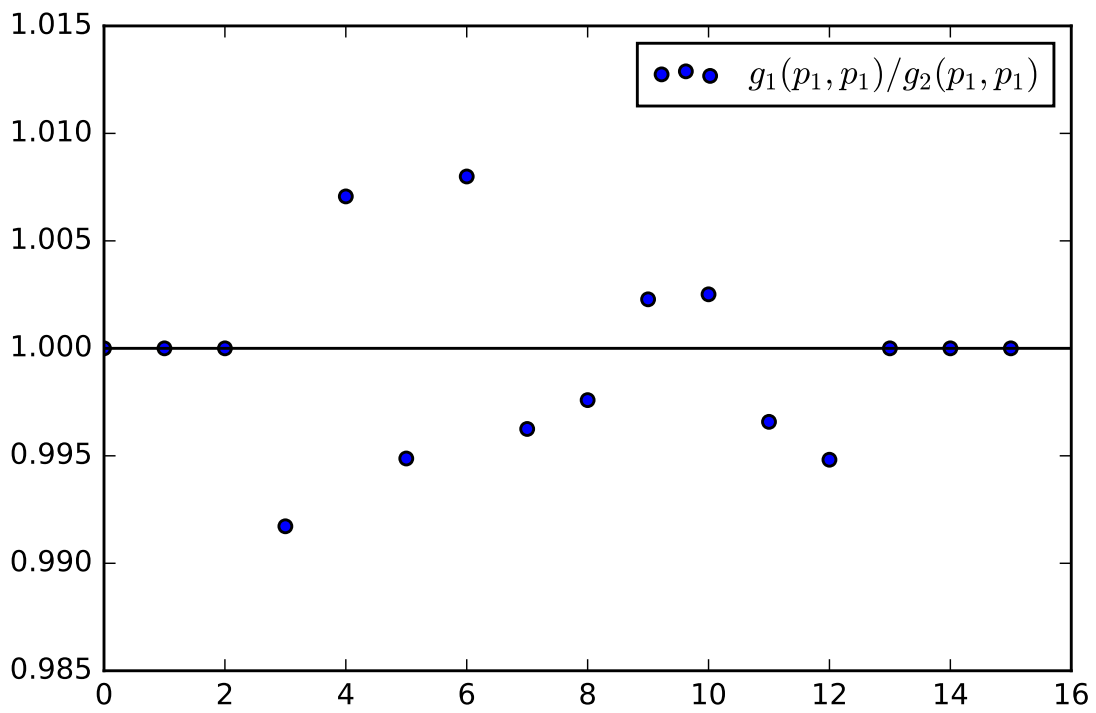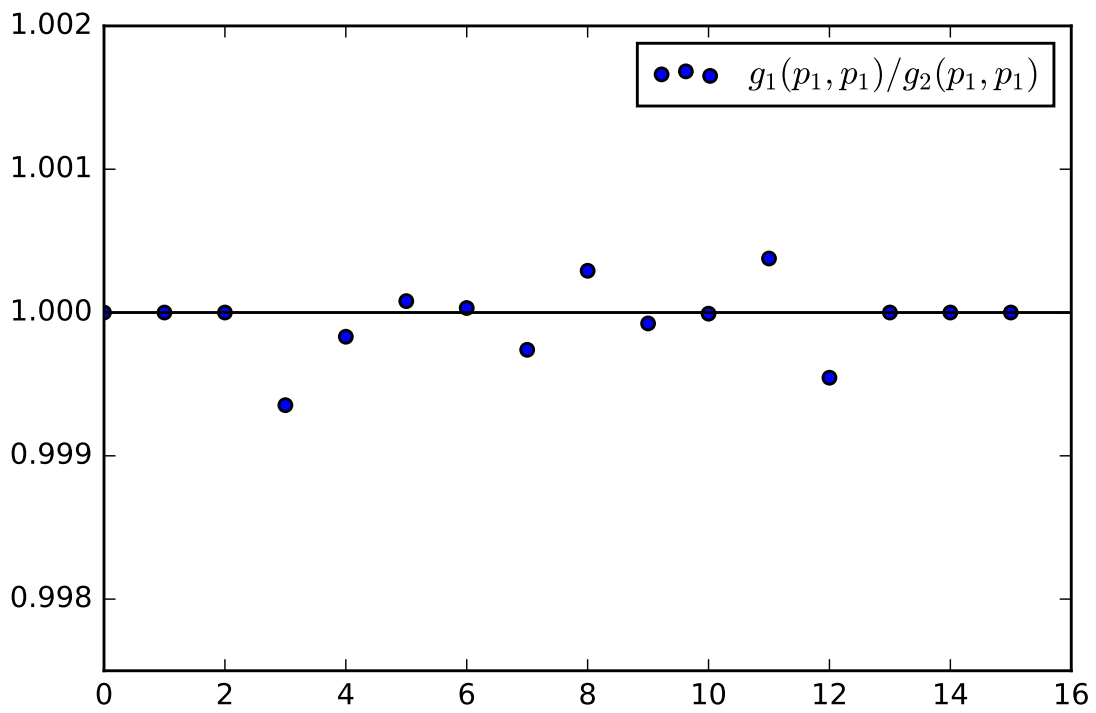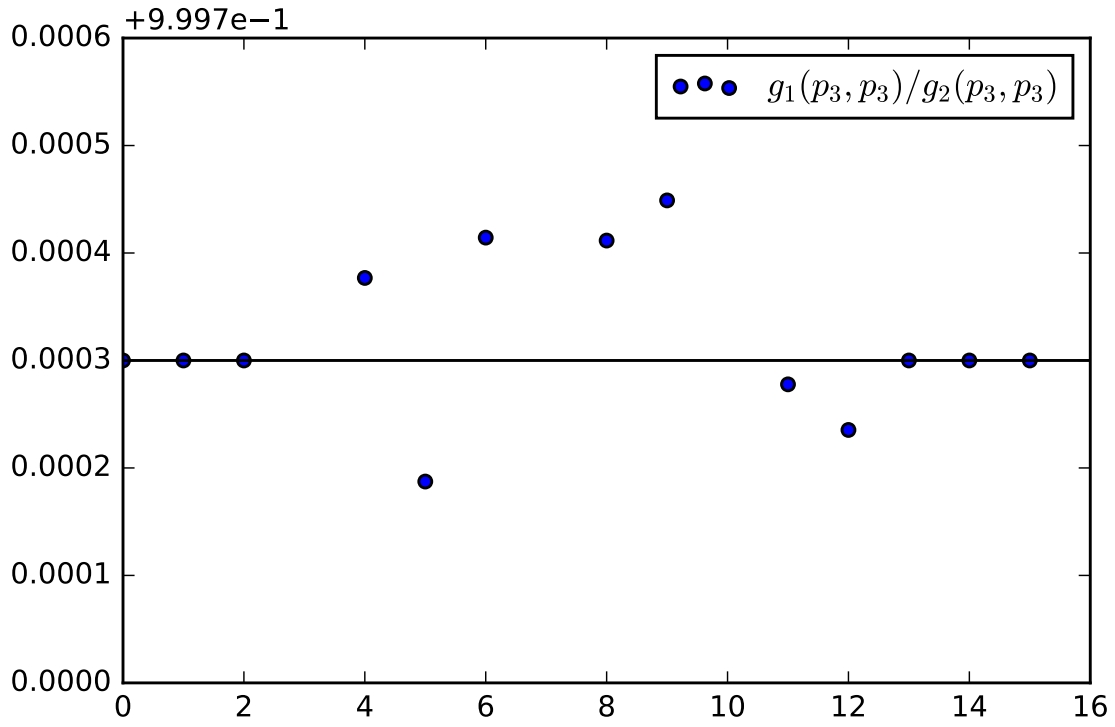
Figure D.3: Ratio of NNPDF grid to CT10 grid both convoluted with NNPDF pdf

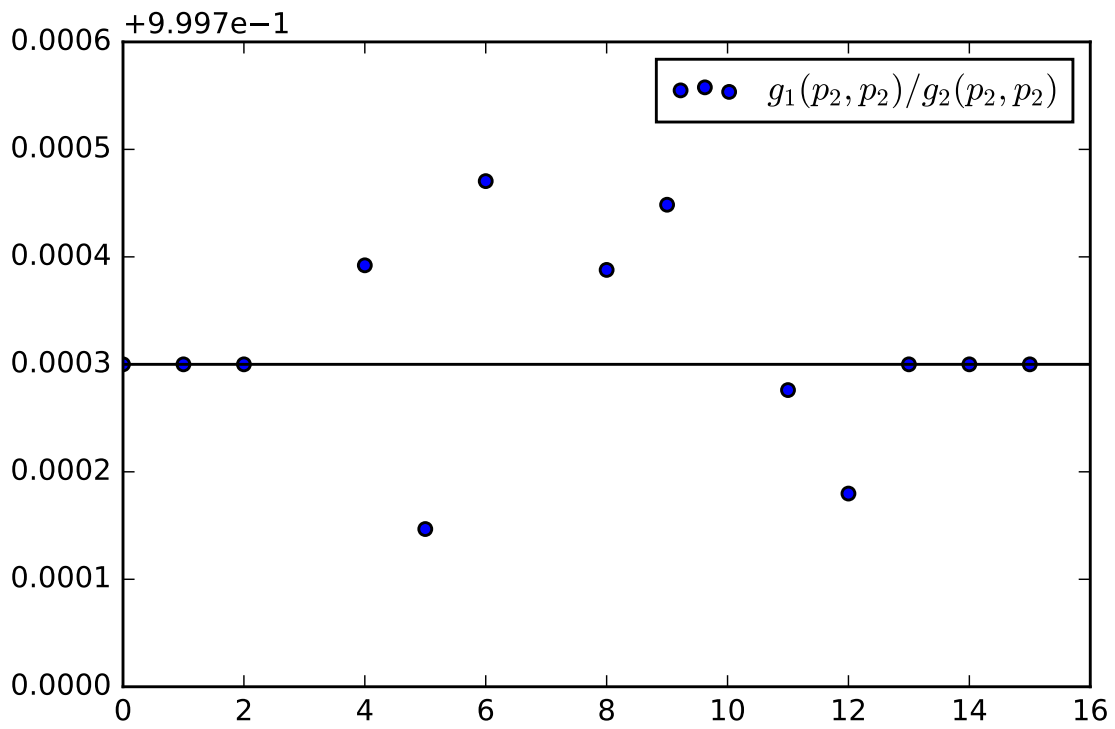Figure D.4: Ratio of NNPDF grid to CT10 grid both convoluted with nCTEQ_208_82 pdf
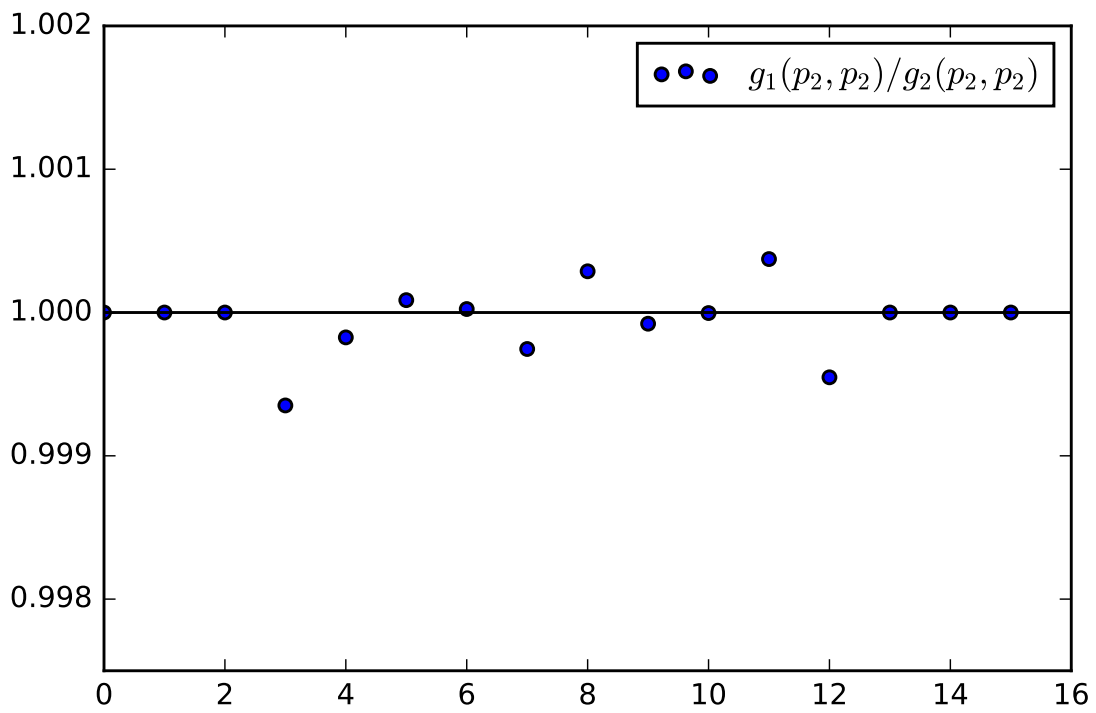
Figure D.5: Ratio of NNPDF grid to CT10 grid both convoluted with CT10 pdf

Figure D.6: Ratio of NNPDF grid to CT10 grid both convoluted with NNPDF pdf

D.3.3. Combined Grids - 10

Figure D.7: Ratio of NNPDF grid to CT10 grid both convoluted with nCTEQ_208_82 pdf

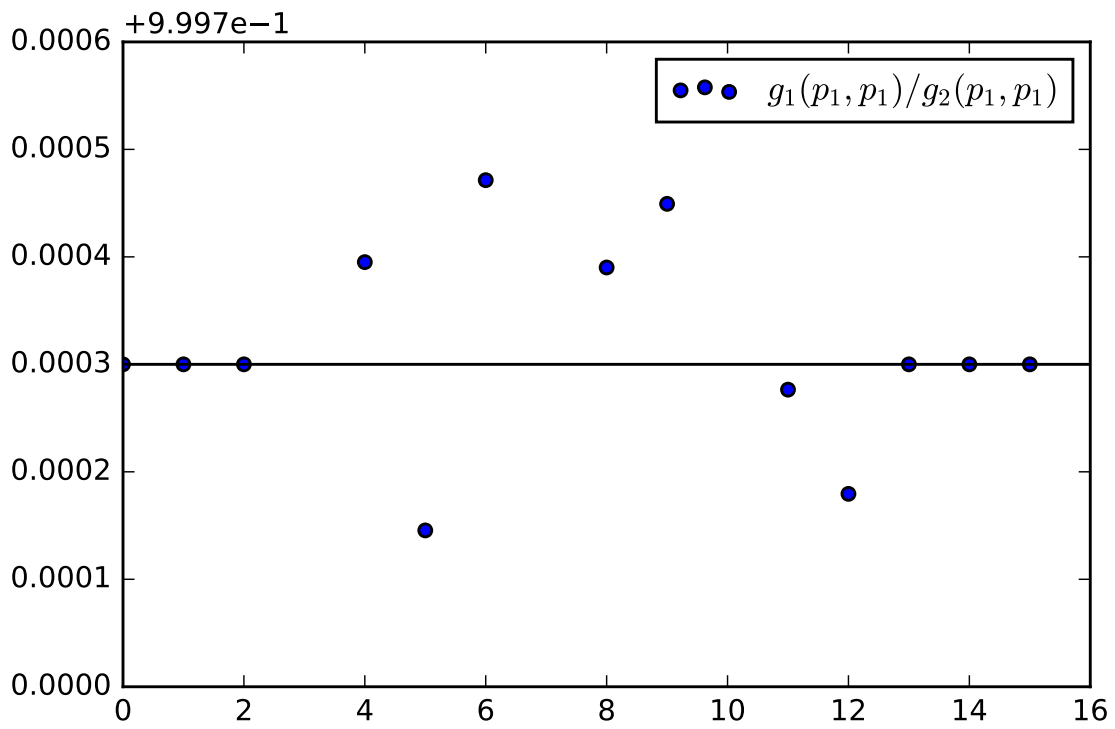Figure D.8: Ratio of NNPDF grid to CT10 grid both convoluted with CT10 pdf

Figure D.9: Ratio of NNPDF grid to CT10 grid both convoluted with NNPDF pdf

D.4. Additional Comparisons

Additionally, some comparisons between the different CT10 grids were done in an attempt to understand some of the numerical noise in the grids.
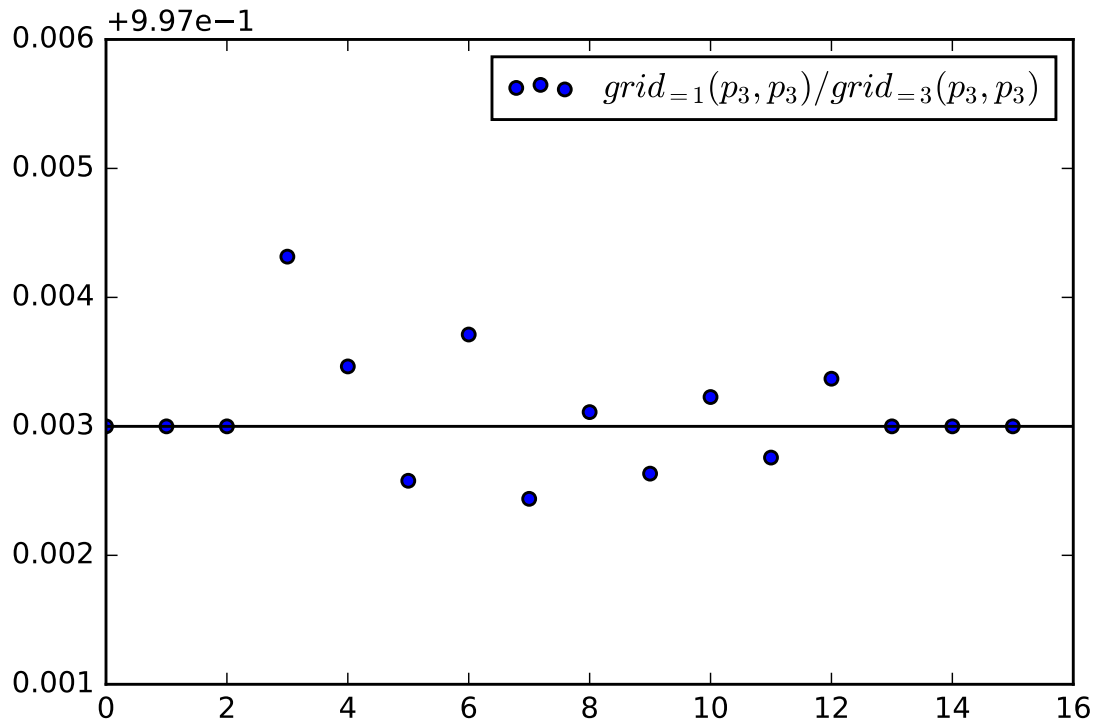
D.4.1. Ratio Plots: CT10 grids

Figure D.10: Ratio of CT10 uncombined grid to combined grid of 3 CT10 runs both convoluted with nCTEQ_208_82 pdf
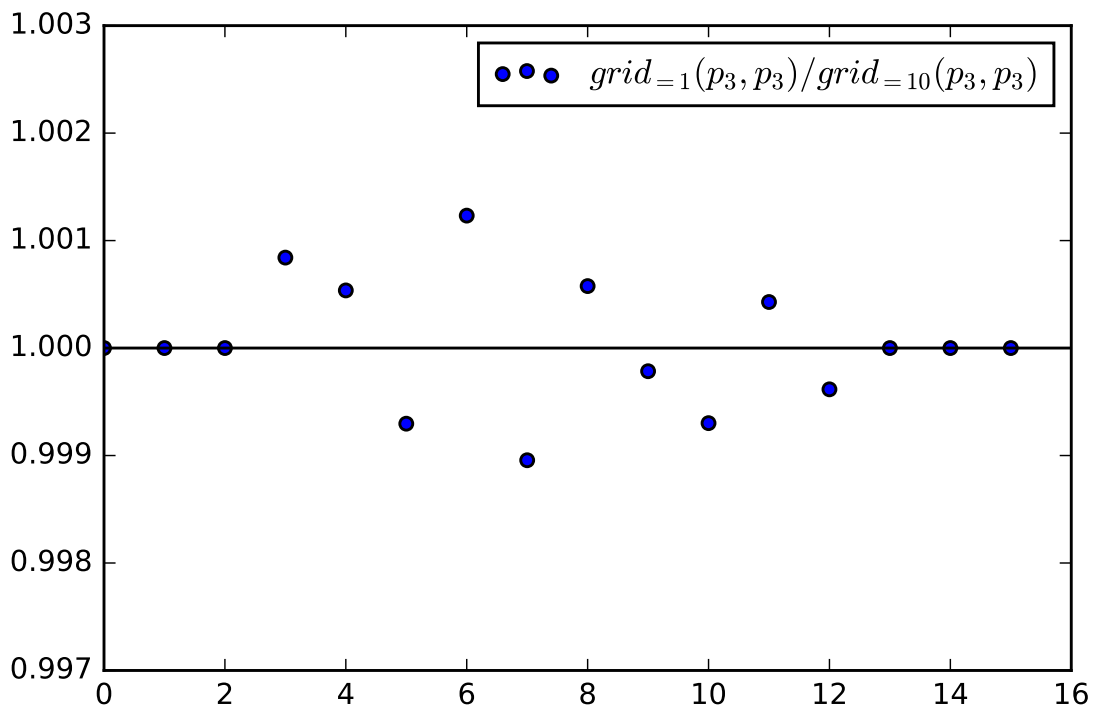
Figure D.11: Ratio of CT10 uncombined grid to combined grid of 10 CT10 runs both convoluted with nCTEQ_208_82 pdf
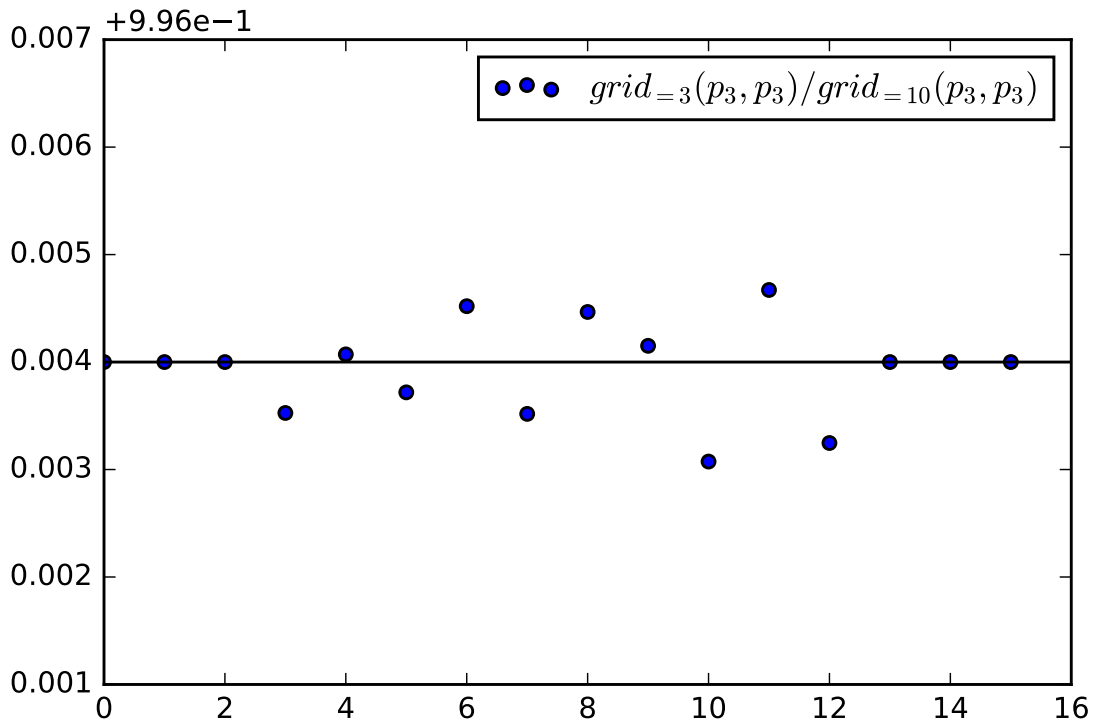
Figure D.12: Ratio of combined grid of 3 CT10 runs to combined grid of 10 CT10 runs both convoluted with nCTEQ_208_82 pdf
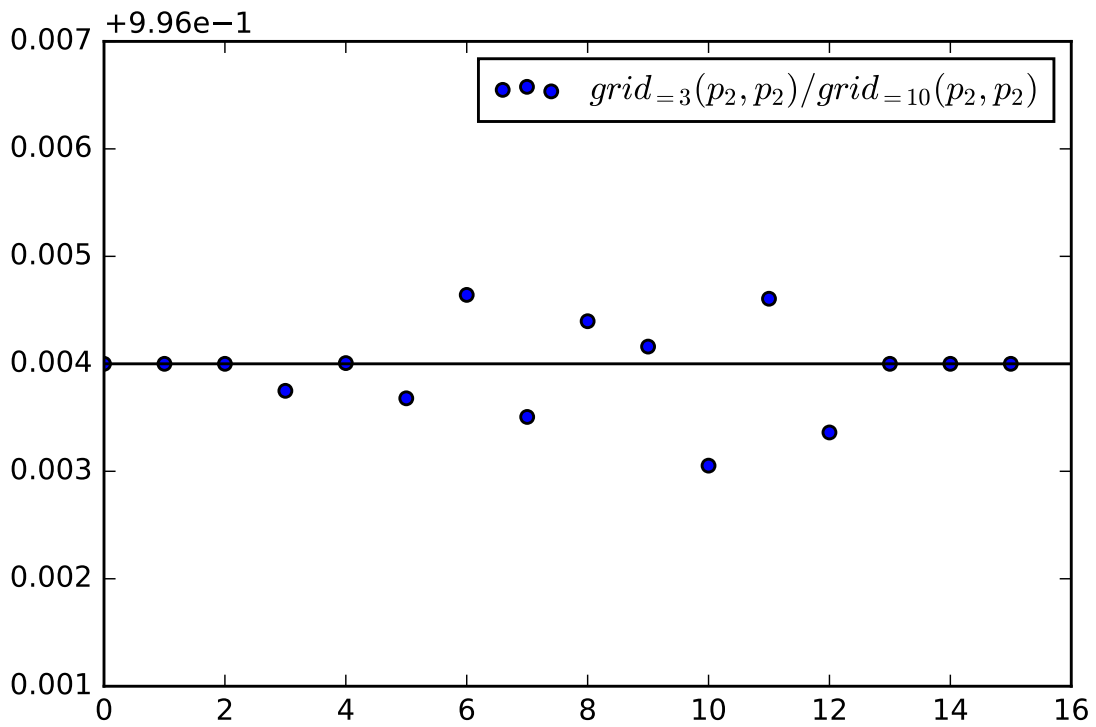
Figure D.13: Ratio of combined grid of 3 CT10 runs to combined grid of 10 CT10 runs both convoluted with CT10 pdf to eliminate any chance of the convoluted PDF contributing to the noise

Appendix E

nCTEQ+LHC: Data vs Theory

This Appendix includes the Data vs Theory plots for nCTEQ+LHC. The data described in these plots can be found in Tables 5.1, 5.2, 5.3 and 5.5 in Chapter 5. These tables include a lot of the relevant information, including the experiment and reference for each data set.

E.1. DIS Data
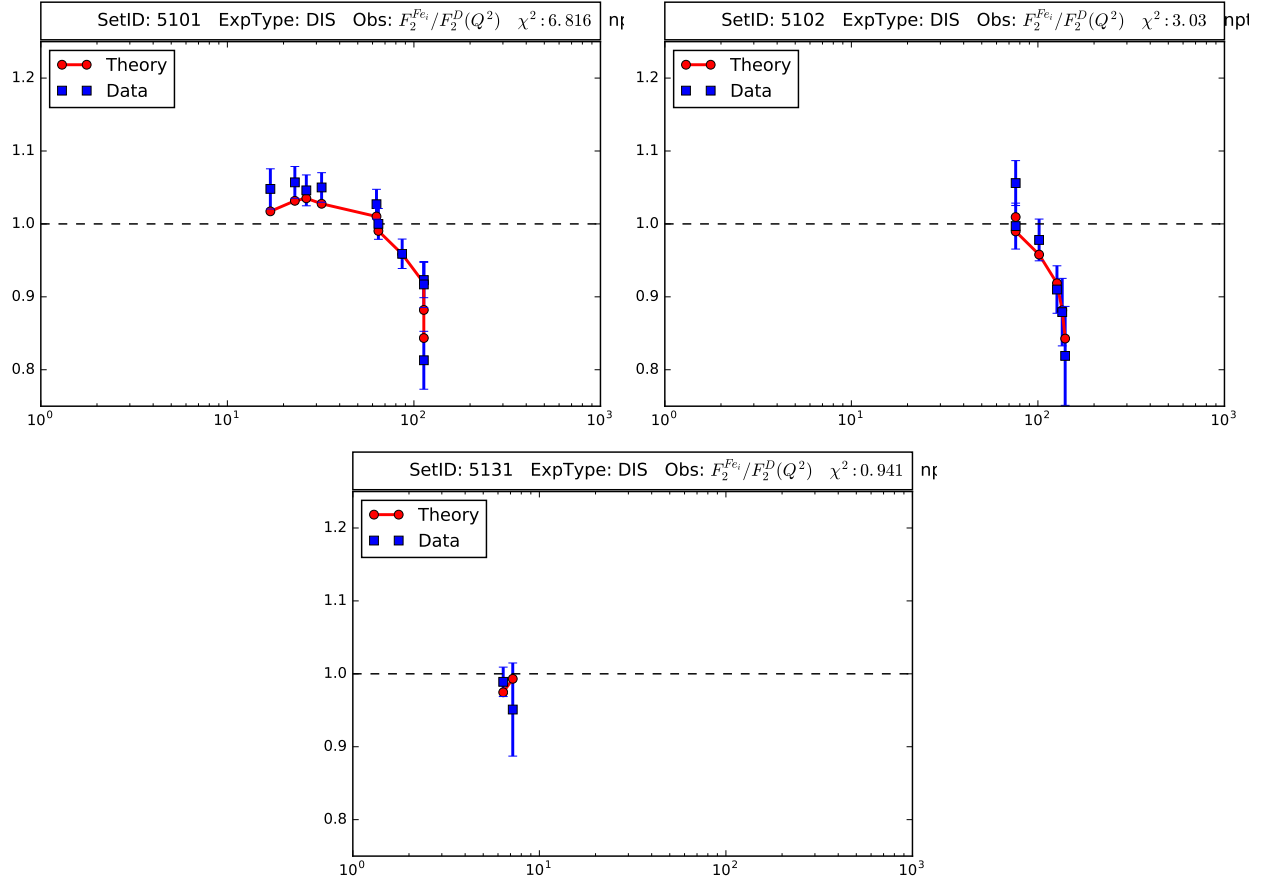
E.1.1. F_2^A/F_2^D

Data found in 5.1.

Figure E.1: `nCTEQ+LHC`: DIS data for Fe/D

Figure E.2: `nCTEQ+LHC`: DIS data for Fe/D

Figure E.3: `nCTEQ+LHC`: DIS data for N/D

267

Figure E.4: `nCTEQ+LHC`: DIS data for Cu/D

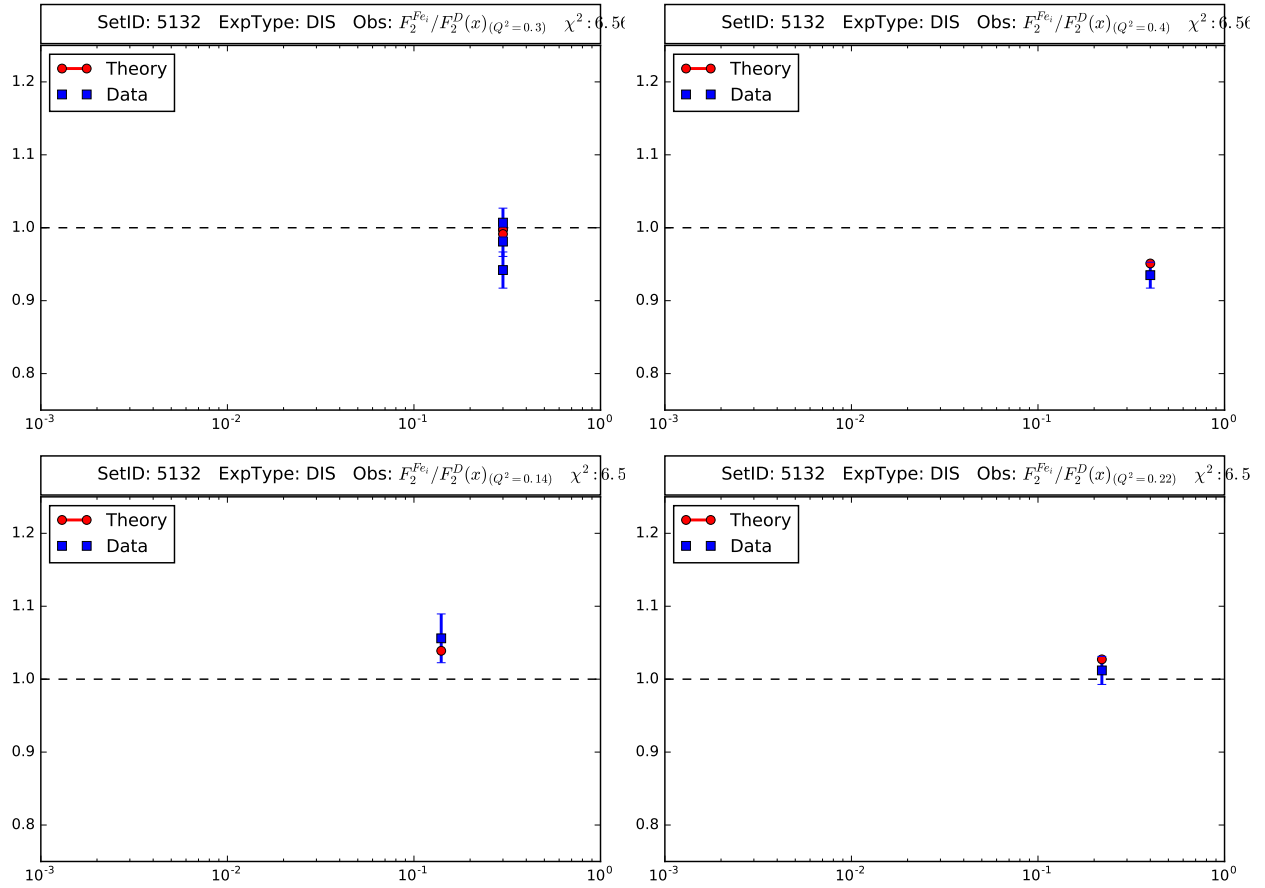Figure E.5: `nCTEQ+LHC`: DIS data for C/D

Figure E.6: `nCTEQ+LHC`: DIS data for Sn/D
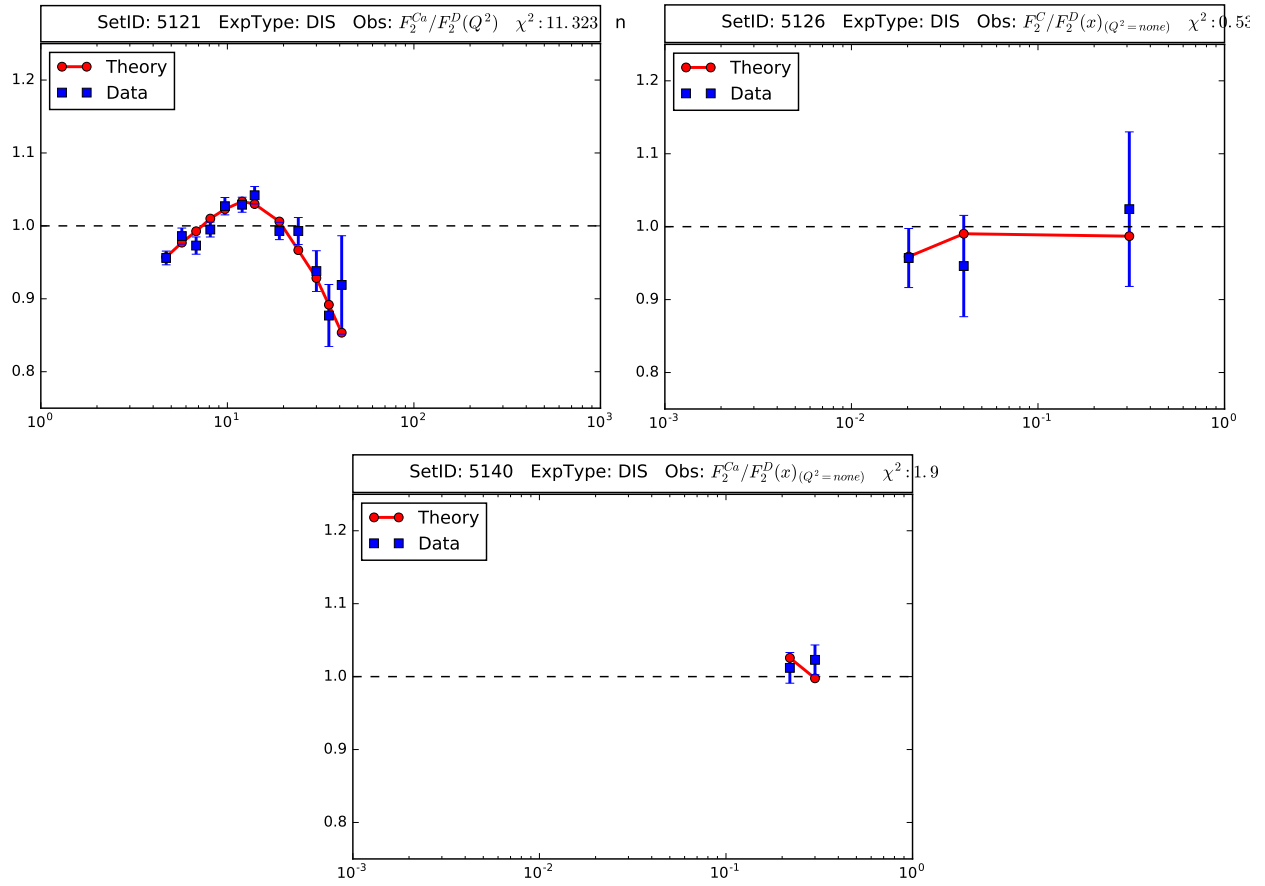
Figure E.7: `nCTEQ+LHC`: DIS data for Ca/D
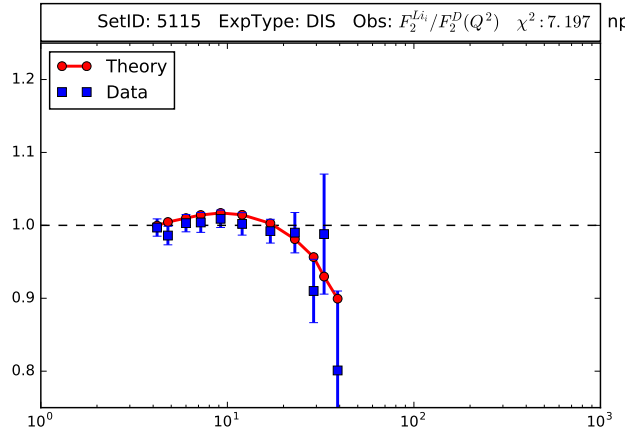
Figure E.8: `nCTEQ+LHC`: DIS data for Li/D
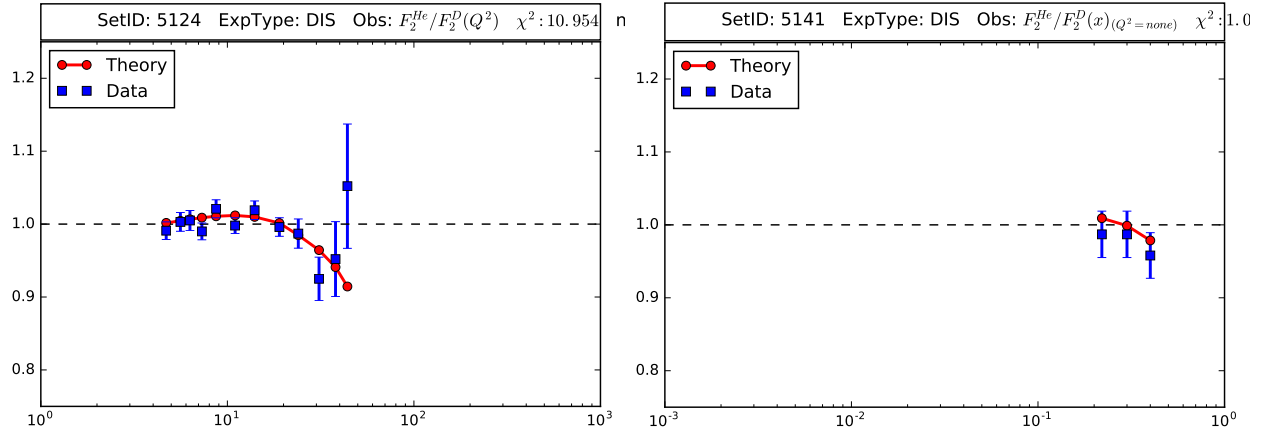
Figure E.9: `nCTEQ+LHC`: DIS data for He/D
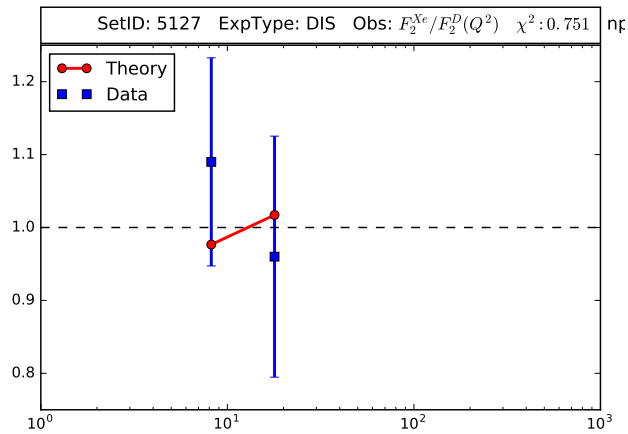
Figure E.10: `nCTEQ+LHC`: DIS data for Xe/D

271
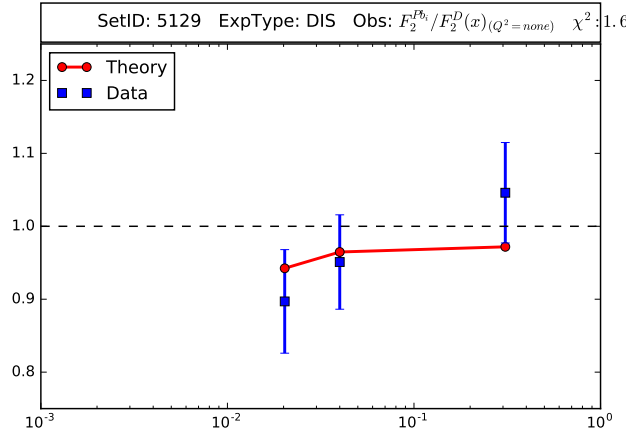
Figure E.11: **nCTEQ+LHC**: DIS data for Pb/D
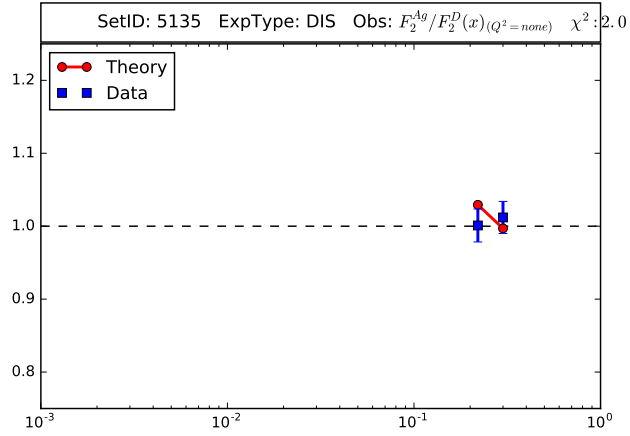
Figure E.12: **nCTEQ+LHC**: DIS data for Ag/D
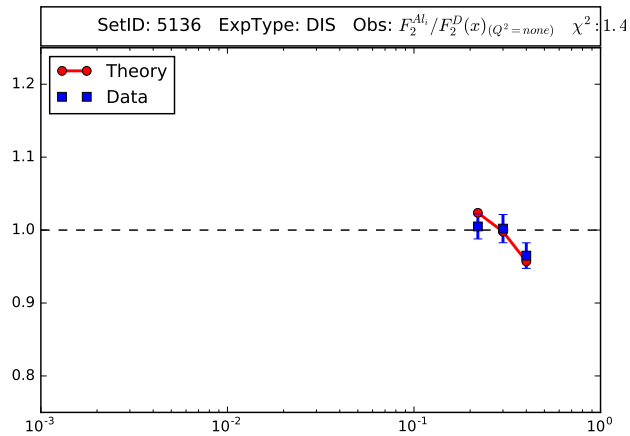
Figure E.13: **nCTEQ+LHC**: DIS data for Al/D
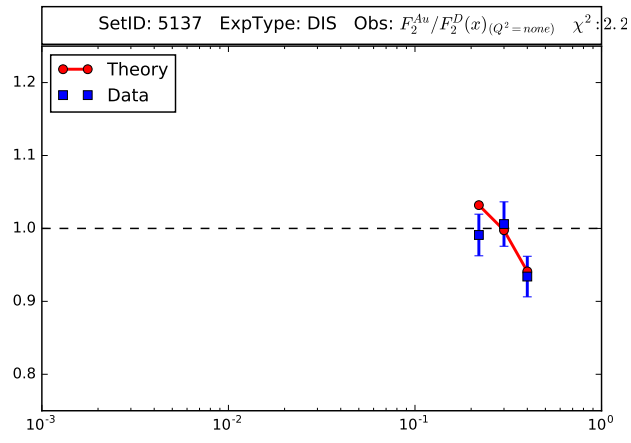
Figure E.14: `nCTEQ+LHC`: DIS data for Au/D
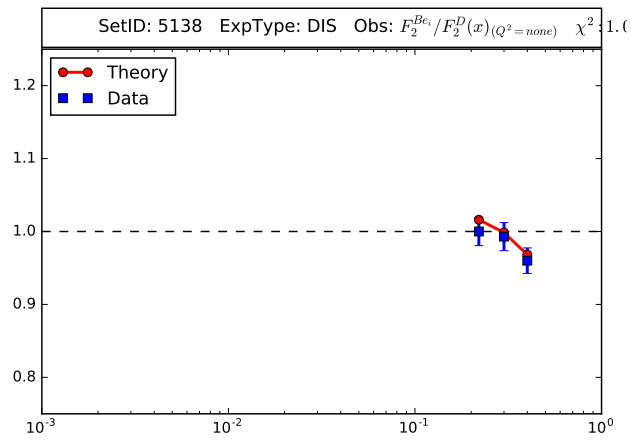
Figure E.15: `nCTEQ+LHC`: DIS data for Be/D
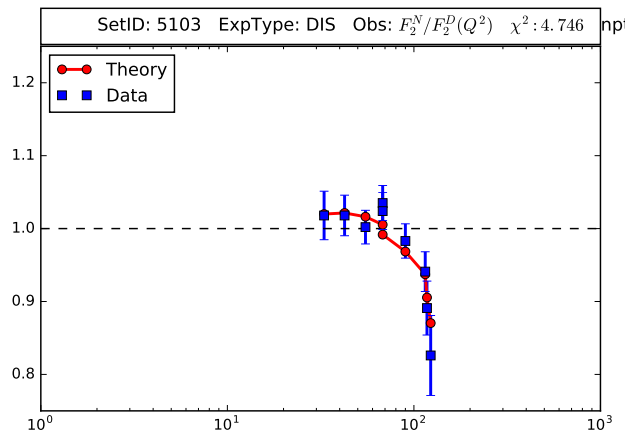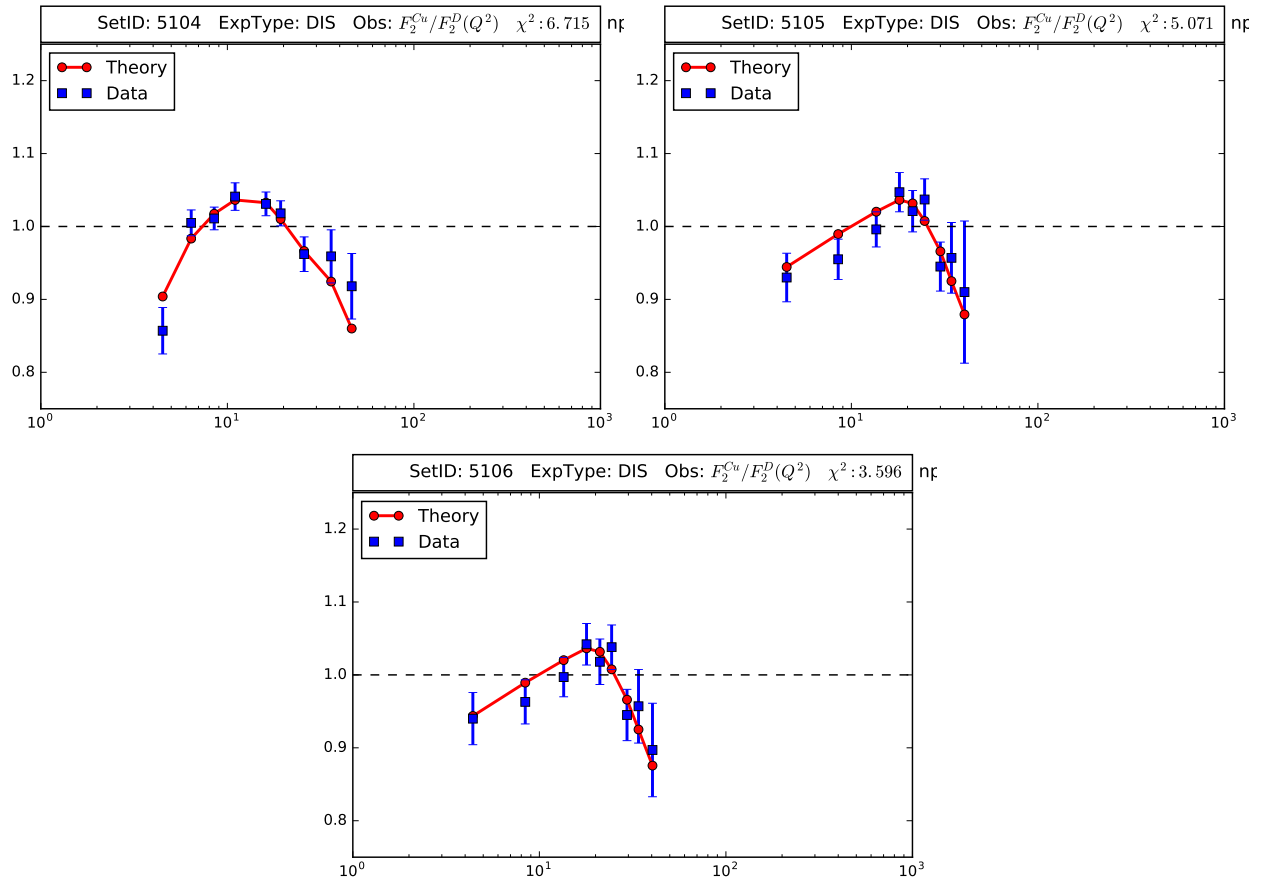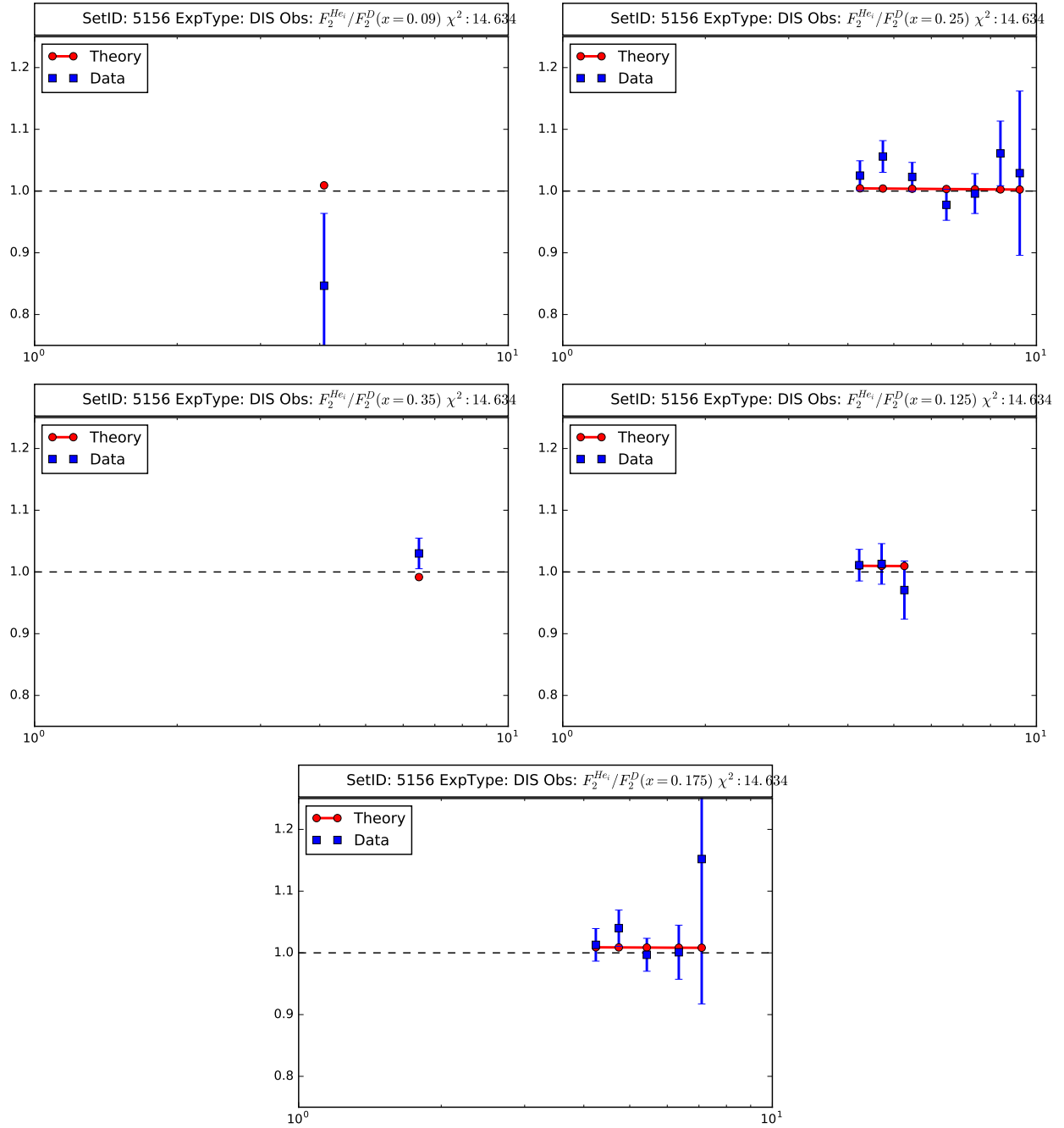
Figure E.16: `nCTEQ+LHC`: DIS data for He/D
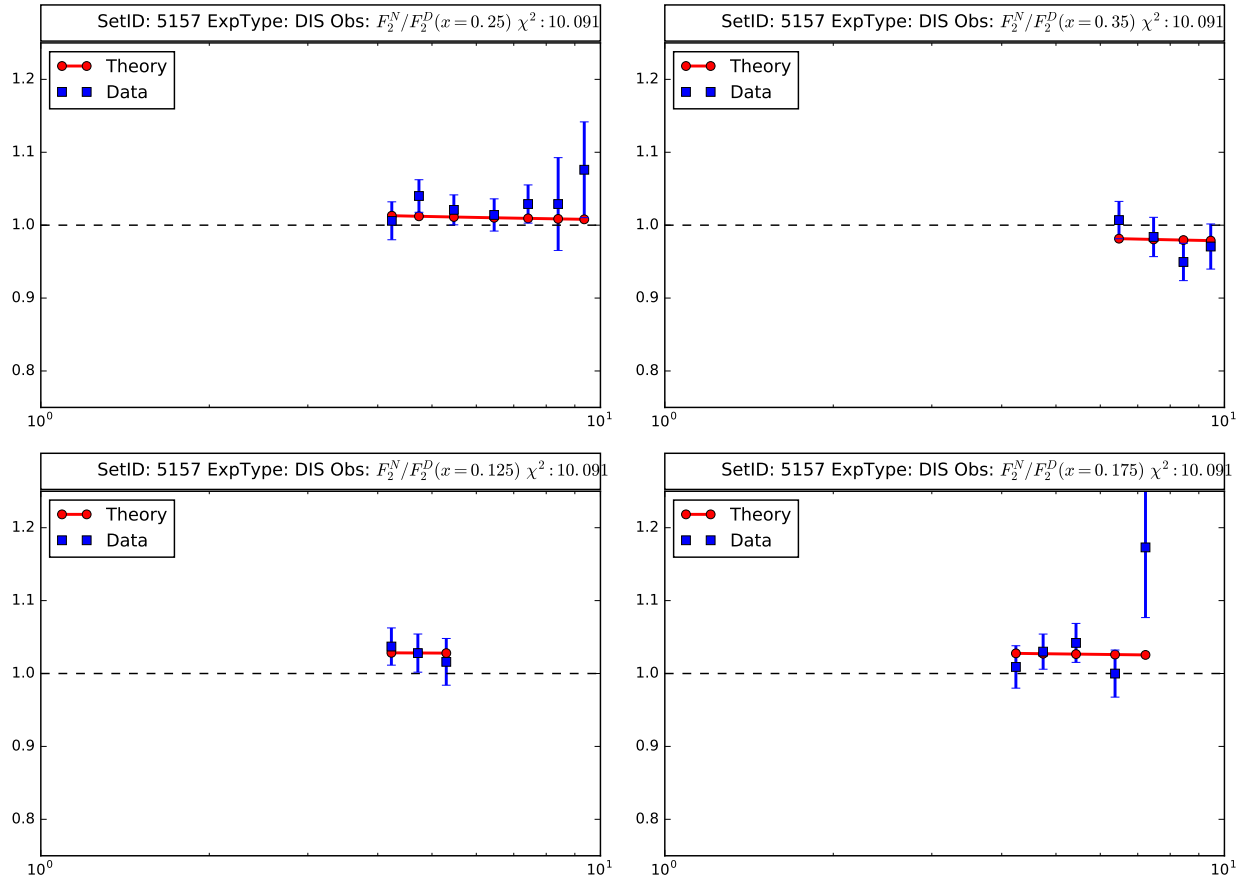
Figure E.17: `nCTEQ+LHC`: DIS data for N/D
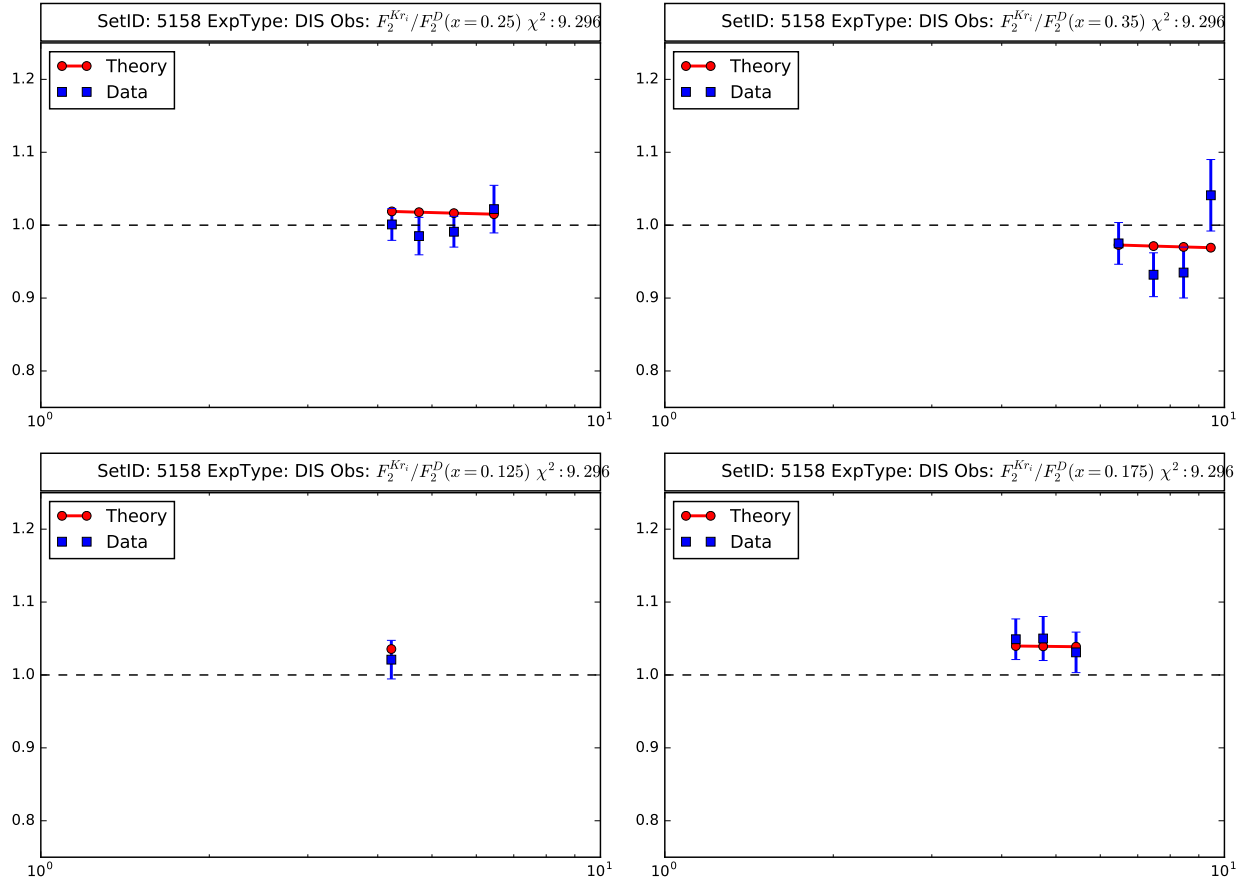
Figure E.18: `nCTEQ+LHC`: DIS data for K/D
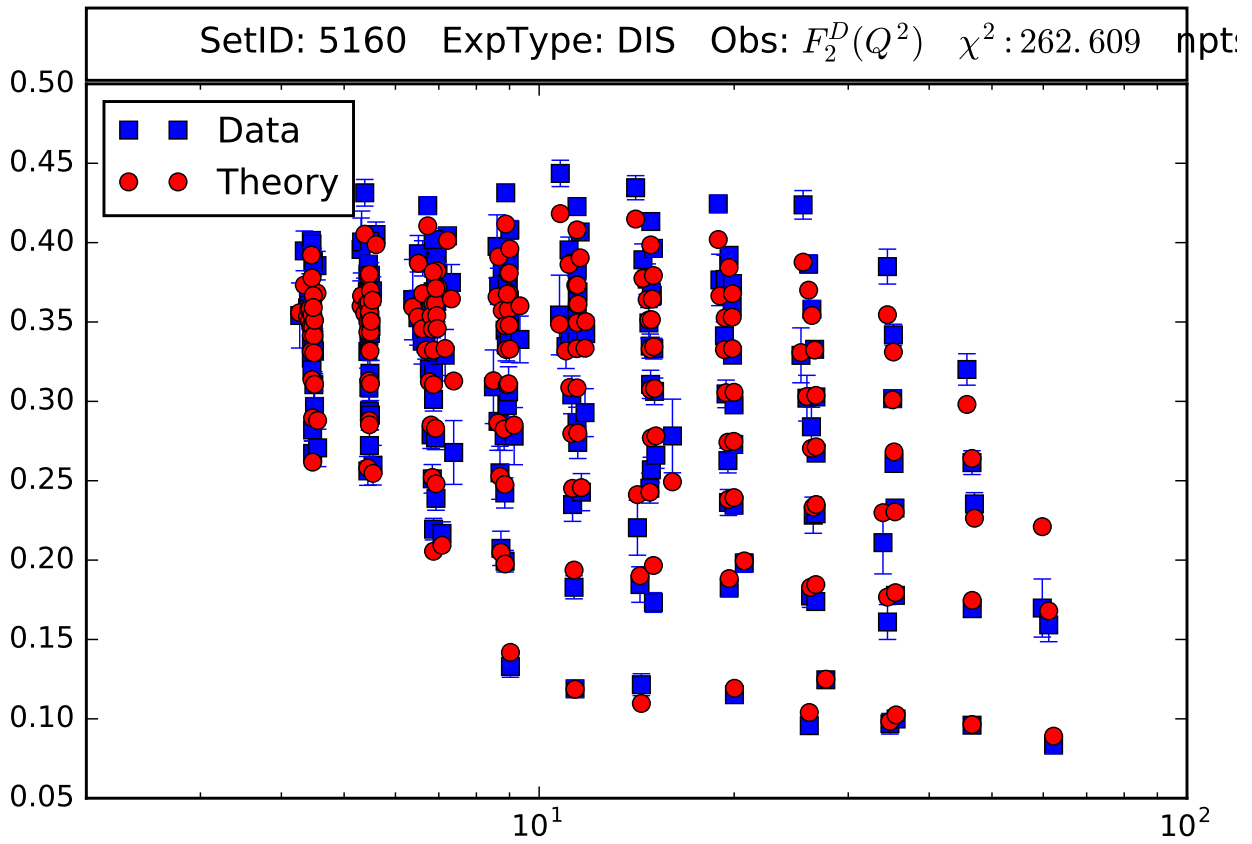
Figure E.19: `nCTEQ+LHC`: DIS data for Deuteron structure function

E.1.2. $F_2^A/F_2^{A'}$

Data found in 5.2.

Figure E.20: `nCTEQ+LHC`: DIS data for Al/C
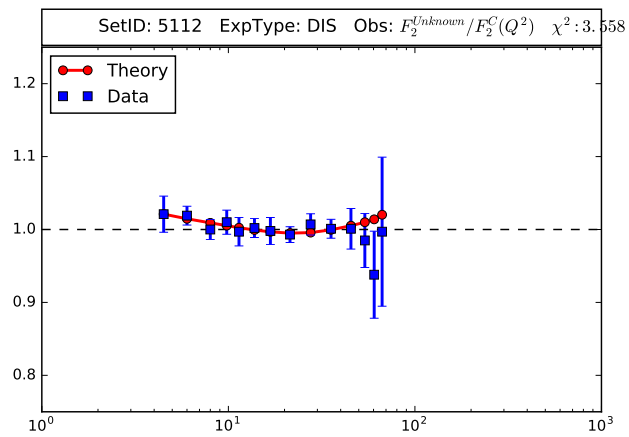
Figure E.21: `nCTEQ+LHC`: DIS data for Be/C
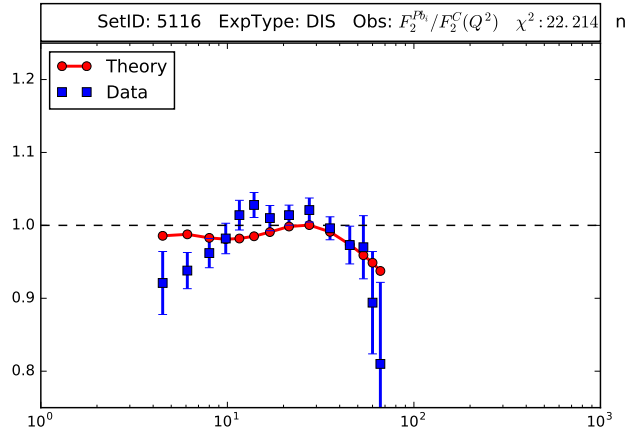
Figure E.22: `nCTEQ+LHC`: DIS data for Pb/C
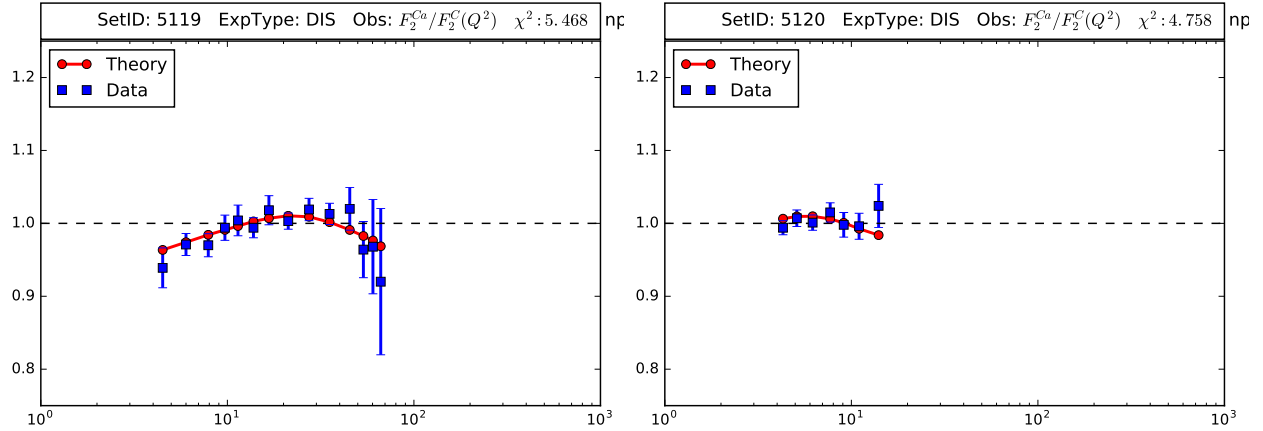
Figure E.23: `nCTEQ+LHC`: DIS data for Ca/C
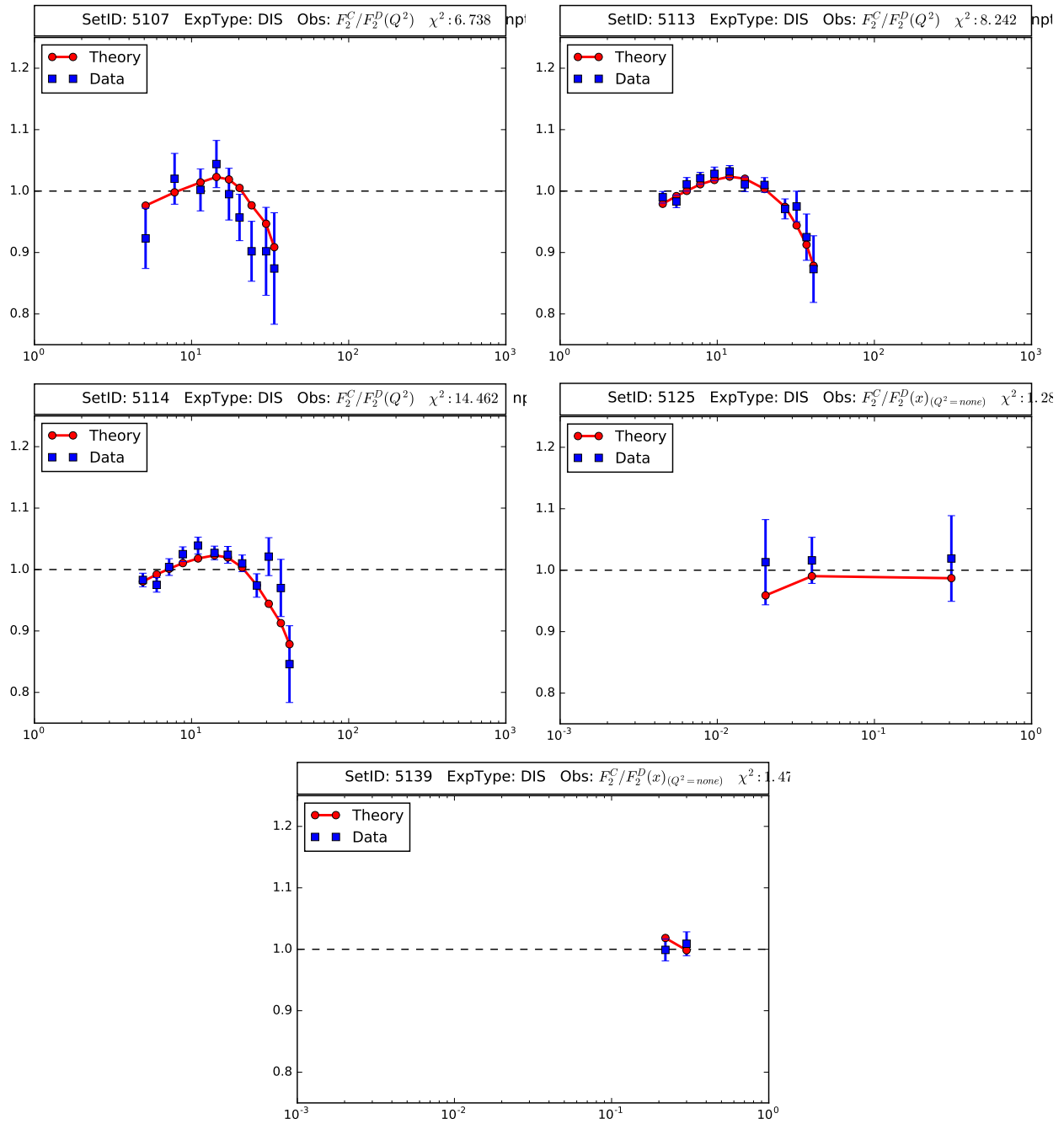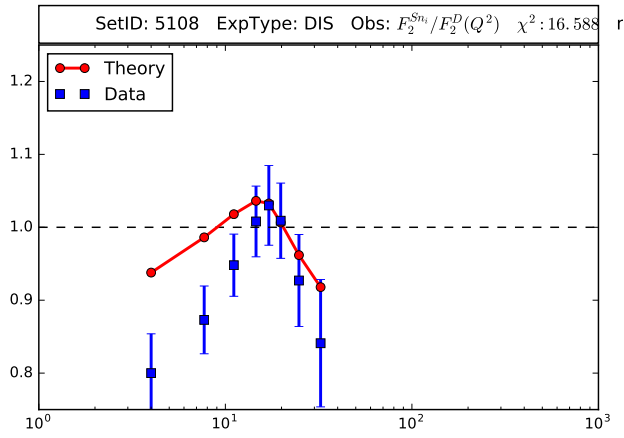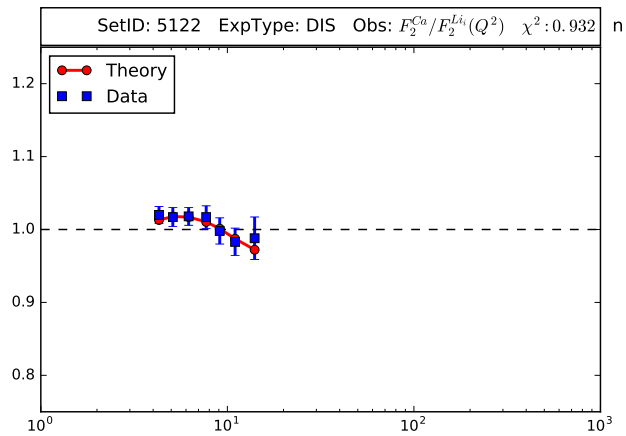

Figure E.24: `nCTEQ+LHC`: DIS data for Ca/Li
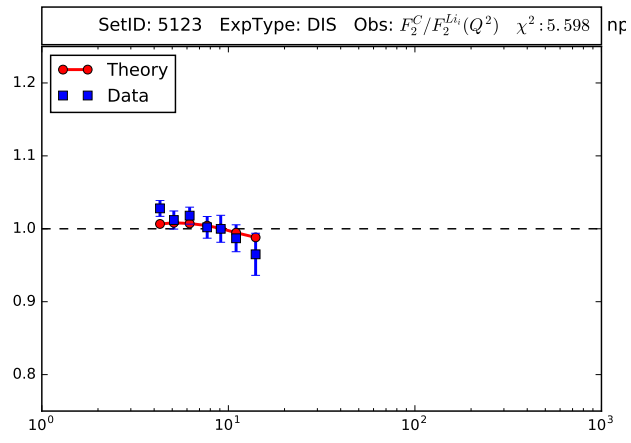
Figure E.25: `nCTEQ+LHC`: DIS data for C/Li
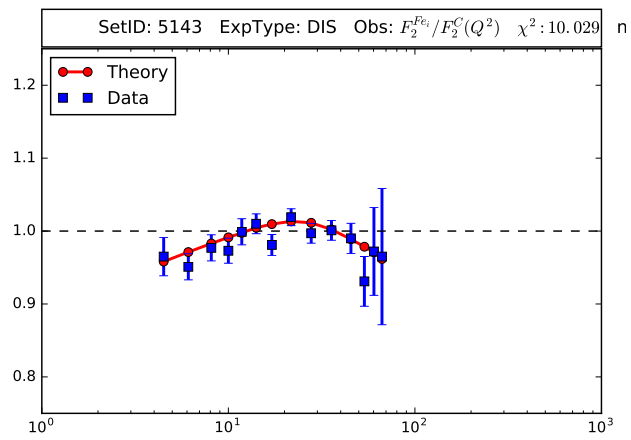
Figure E.26: `nCTEQ+LHC`: DIS data for Fe/C
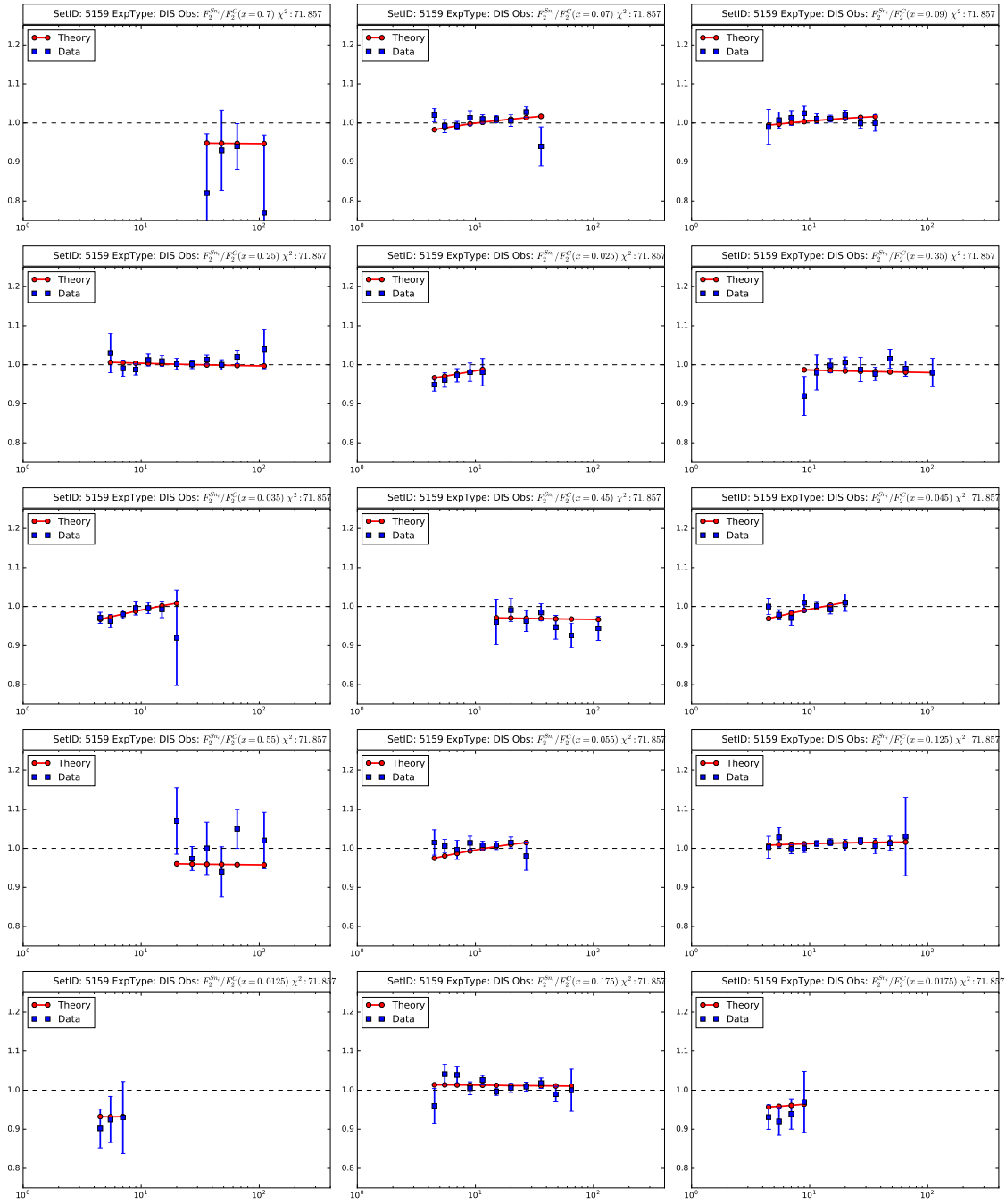
Figure E.27: `nCTEQ+LHC`: DIS data for Sn/C

E.2. DY Data

E.2.1. $\sigma^{pA}_{DY}/\sigma^{pA'}_{DY}$
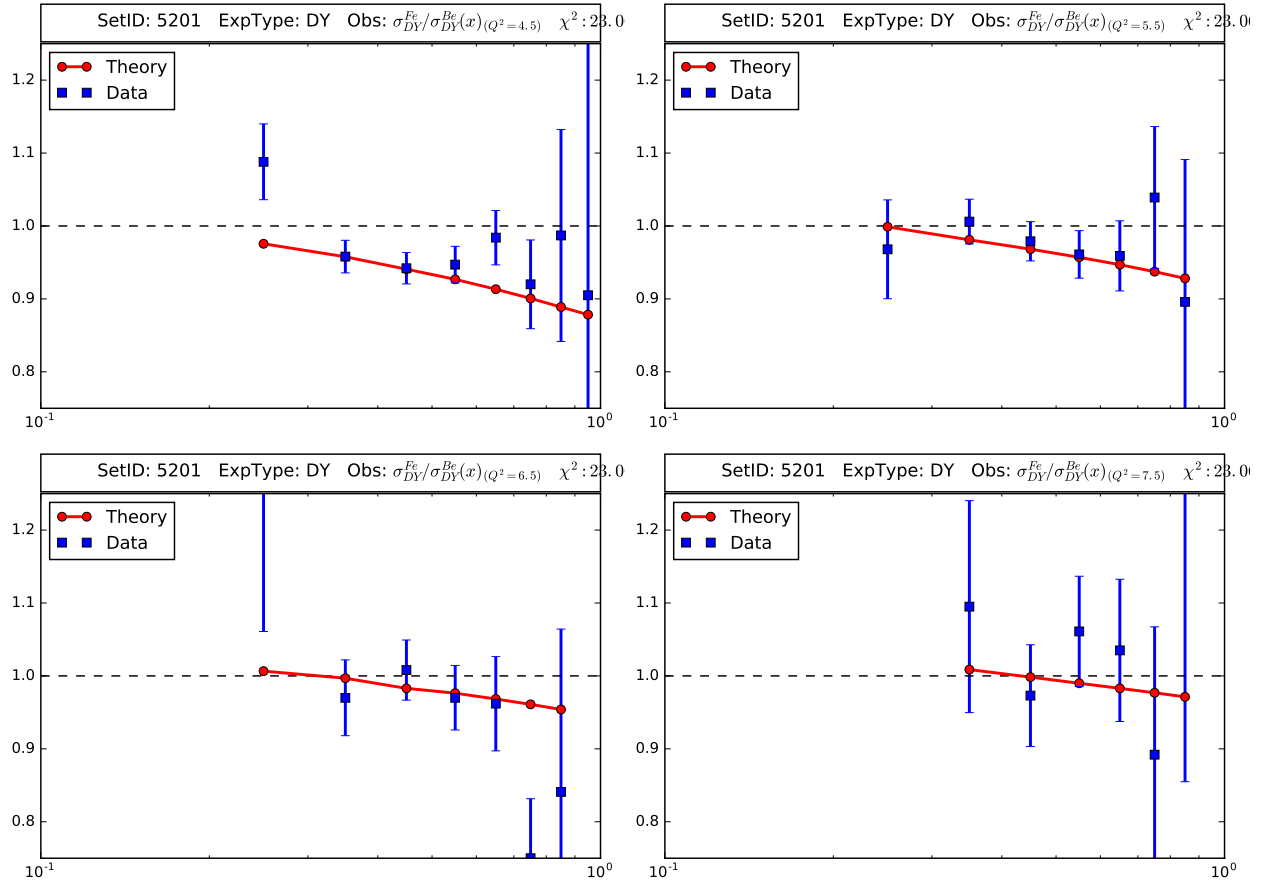
Data found in 5.3.

Figure E.28: `nCTEQ+LHC`: DY data for Fe/Be
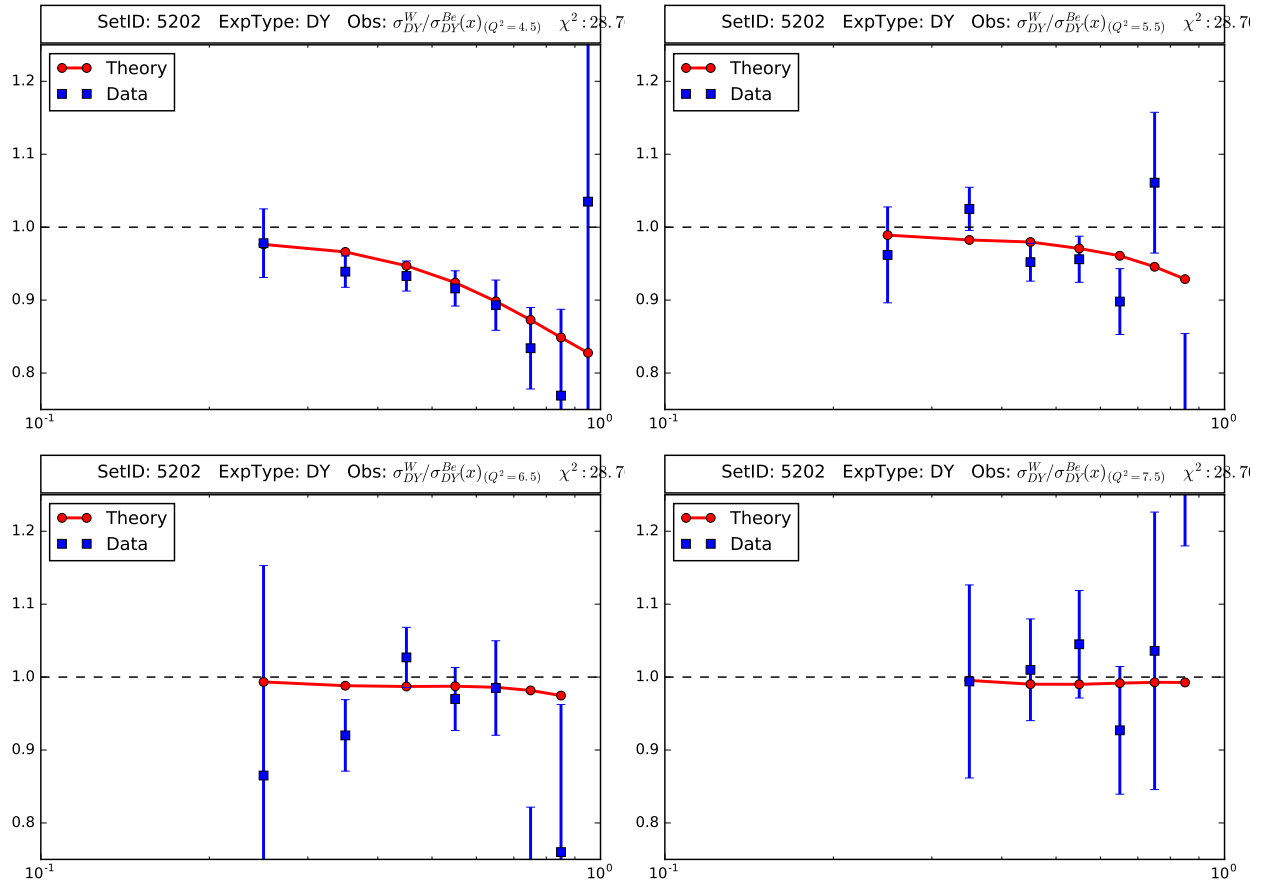
Figure E.29: `nCTEQ+LHC`: DY data for W/Be
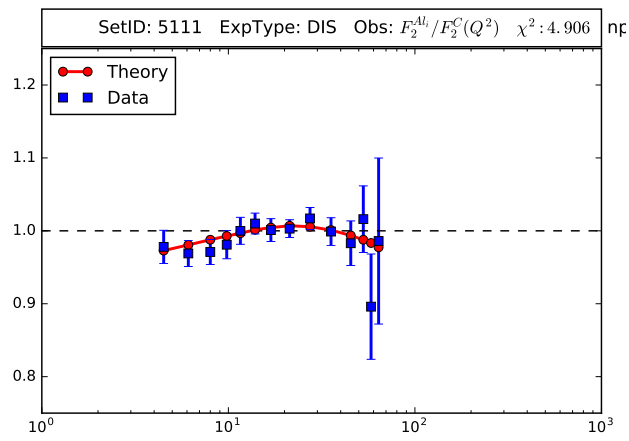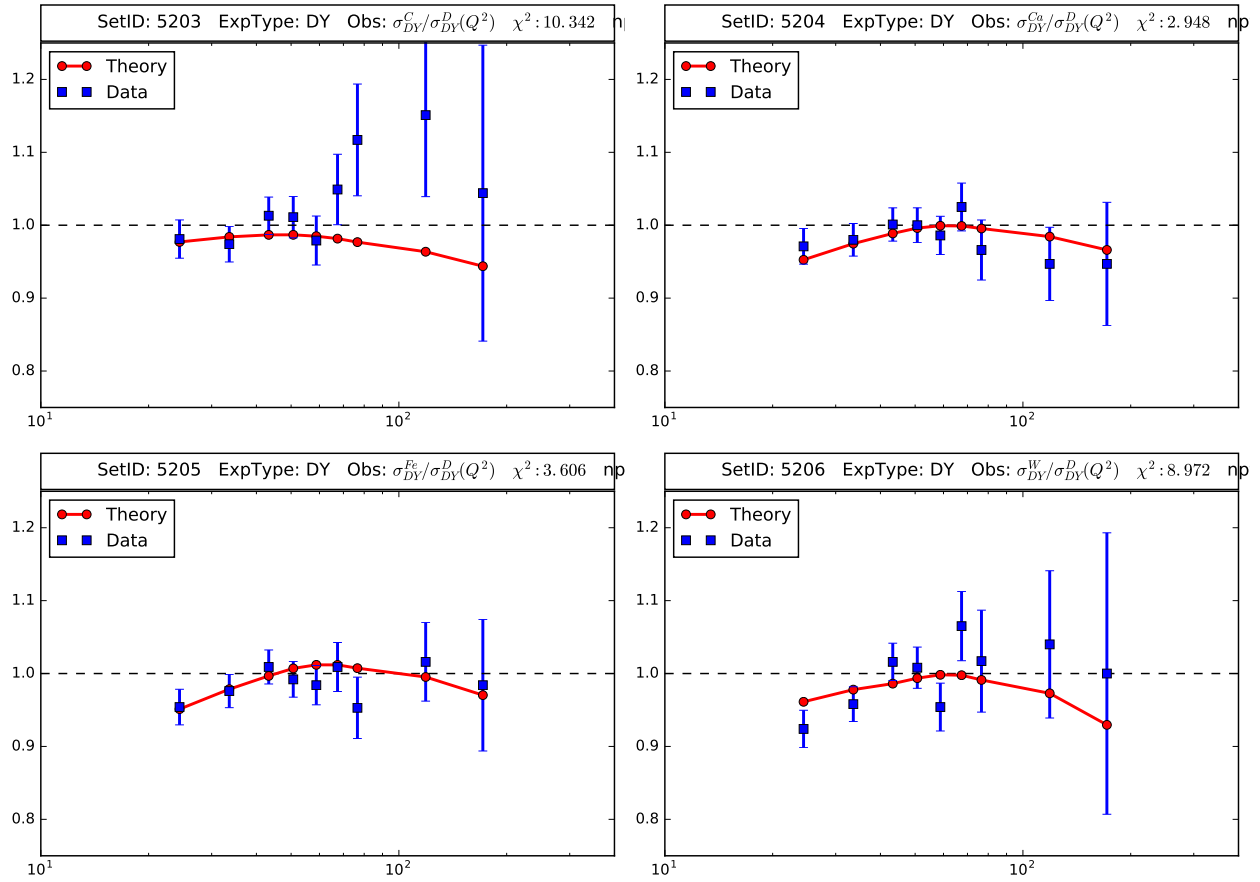
Figure E.30: nCTEQ+LHC: DY data for A/D

E.3. LHC Data

E.3.1. *W/Z* Production

Data found in 5.5.
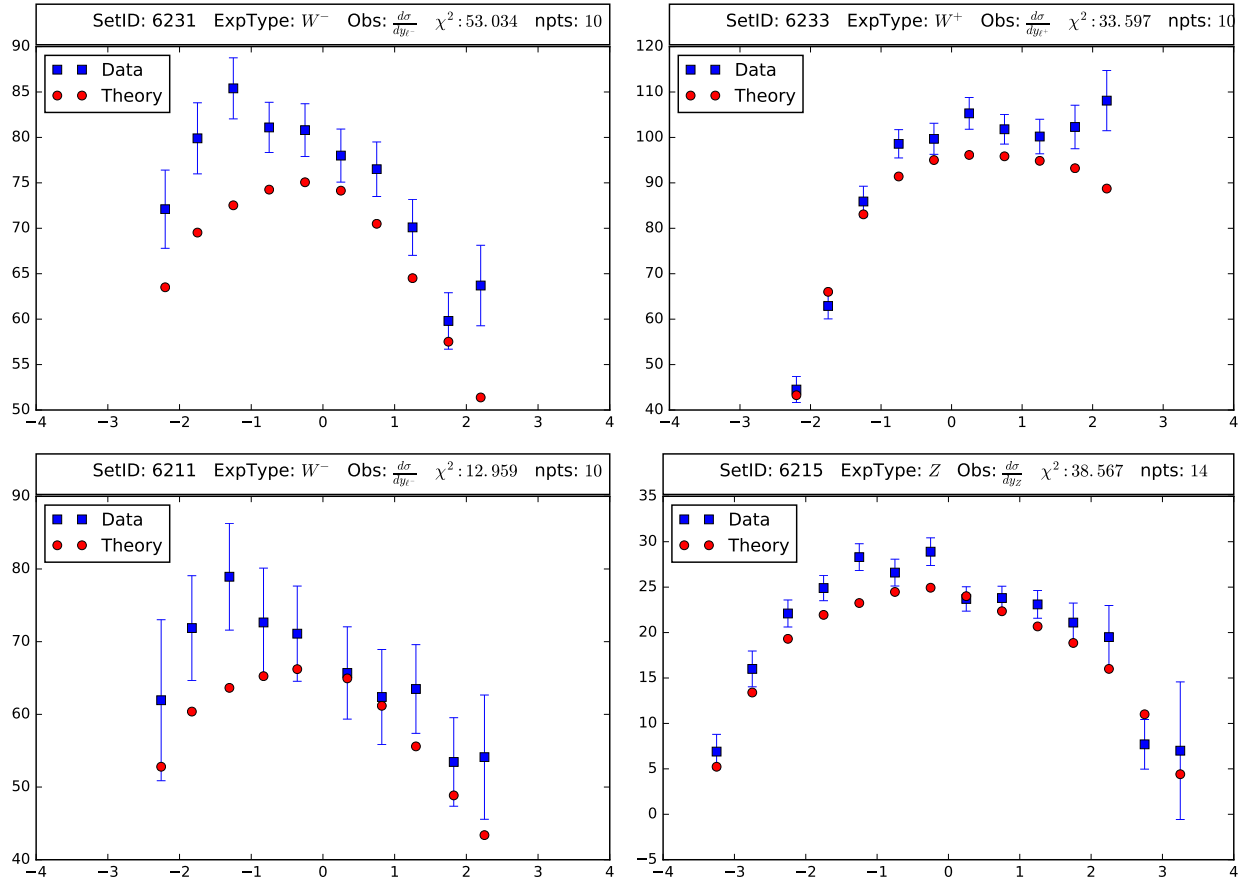
Figure E.31: `nCTEQ+LHC`: LHC pPb *W/Z* Production data

BIBLIOGRAPHY

[1] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order," *Comput. Phys. Commun.* **182** (2011) 2388–2403, `arXiv:1011.3540 [hep-ph]`.

[2] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "W Physics at the LHC with FEWZ 2.1," *Comput. Phys. Commun.* **184** (2013) 208–214, `arXiv:1201.5896 [hep-ph]`.

[3] P. J. Mohr, D. B. Newell, and B. N. Taylor, "CODATA Recommended Values of the Fundamental Physical Constants: 2014," *Rev. Mod. Phys.* **88** (2016) no. 3, 035009, `arXiv:1507.07956 [physics.atom-ph]`.

[4] **Particle Data Group** Collaboration, C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys.* **C40** (2016) no. 10, 100001.

[5] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.* **B126** (1977) 298–318.

[6] I. Schienbein, J. Y. Yu, K. Kovarik, C. Keppel, J. G. Morfin, F. Olness, and J. F. Owens, "PDF Nuclear Corrections for Charged and Neutral Current Processes," *Phys. Rev.* **D80** (2009) 094004, `arXiv:0907.2357 [hep-ph]`.

[7] K. Kovarik *et al.*, "nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework," *Phys. Rev.* **D93** (2016) no. 8, 085037, `arXiv:1509.00792 [hep-ph]`.

[8] I. Schienbein, J. Y. Yu, C. Keppel, J. G. Morfin, F. Olness, and J. F. Owens, "Nuclear parton distribution functions from neutrino deep inelastic scattering," *Phys. Rev.* **D77** (2008) 054013, `arXiv:0710.4897 [hep-ph]`.

[9] K. J. Eskola, H. Paukkunen, and C. A. Salgado, "EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions," *JHEP* **04** (2009) 065, `arXiv:0902.4154 [hep-ph]`.

[10] M. Hirai, S. Kumano, and T. H. Nagai, "Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order," *Phys. Rev.* **C76** (2007) 065207, `arXiv:0709.3038 [hep-ph]`.

[11] Y.-Q. Ma and J.-W. Qiu, "QCD Factorization and PDFs from Lattice QCD Calculation," *Int. J. Mod. Phys. Conf. Ser.* **37** (2015) 1560041, `arXiv:1412.2688 [hep-ph]`.

[12] C. Alexandrou, "Parton distribution functions from Lattice QCD," *Few Body Syst.* **57** (2016) no. 8, 621–626, arXiv:1602.08726 [hep-lat].

[13] A. Buckley, J. Ferrando, S. Lloyd, K. Nordstrm, B. Page, M. Rfenacht, M. Schnherr, and G. Watt, "LHAPDF6: parton density access in the LHC precision era," *Eur. Phys. J.* **C75** (2015) 132, arXiv:1412.7420 [hep-ph].

[14] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan, "Implications of CTEQ global analysis for collider observables," *Phys. Rev.* **D78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[15] M. Botje, "QCDNUM: Fast QCD Evolution and Convolution," *Comput. Phys. Commun.* **182** (2011) 490–532, arXiv:1005.1481 [hep-ph].

[16] S. Carrazza, A. Ferrara, D. Palazzo, and J. Rojo, "APFEL Web," *J. Phys.* **G42** (2015) no. 5, 057001, arXiv:1410.5456 [hep-ph].

[17] F. Hautmann, H. Jung, M. Krmer, P. J. Mulders, E. R. Nocera, T. C. Rogers, and A. Signori, "TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions," *Eur. Phys. J.* **C74** (2014) 3220, arXiv:1408.3015 [hep-ph].

[18] F. Hautmann, H. Jung, and S. T. Monfared, "The CCFM uPDF evolution uPDFevolv Version 1.0.00," *Eur. Phys. J.* **C74** (2014) 3082, arXiv:1407.5935 [hep-ph].

[19] **Particle Data Group** Collaboration, K. A. Olive *et al.*, "Review of Particle Physics," *Chin. Phys.* **C38** (2014) 090001.

[20] A. Kusina, F. I. Olness, I. Schienbein, T. Jezo, K. Kovarik, T. Stavreva, and J. Y. Yu, "Hybrid scheme for heavy flavors: Merging the fixed flavor number scheme and variable flavor number scheme," *Phys. Rev.* **D88** (2013) no. 7, 074032, arXiv:1306.6553 [hep-ph].

[21] V. Bertone, S. Carrazza, and J. Rojo, "APFEL: A PDF Evolution Library with QED corrections," *Comput. Phys. Commun.* **185** (2014) 1647–1668, arXiv:1310.1394 [hep-ph].

[22] S. Lomatch, F. I. Olness, and J. C. Collins, "Mini - Jets and B Pairs: Analytic and Numeric Results for Small X Dominated Processes," *Nucl. Phys.* **B317** (1989) 617–634.

[23] C. A. Argelles, F. Halzen, L. Wille, M. Kroll, and M. H. Reno, "High-energy behavior of photon, neutrino, and proton cross sections," *Phys. Rev.* **D92** (2015) no. 7, 074040, arXiv:1504.06639 [hep-ph].

[24] A. Bhattacharya, R. Enberg, M. H. Reno, I. Sarcevic, and A. Stasto, "Perturbative charm production and the prompt atmospheric neutrino flux in light of RHIC and LHC," *JHEP* **06** (2015) 110, arXiv:1502.01076 [hep-ph].

[25] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC," *Eur. Phys. J.* **C63** (2009) 189–285, `arXiv:0901.0002 [hep-ph]`.

[26] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, "New parton distribution functions from a global analysis of quantum chromodynamics," *Phys. Rev.* **D93** (2016) no. 3, 033006, `arXiv:1506.07443 [hep-ph]`.

[27] **NNPDF** Collaboration, R. D. Ball *et al.*, "Parton distributions for the LHC Run II," *JHEP* **04** (2015) 040, `arXiv:1410.8849 [hep-ph]`.

[28] **ZEUS, H1** Collaboration, H. Abramowicz *et al.*, "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data," *Eur. Phys. J.* **C75** (2015) no. 12, 580, `arXiv:1506.06042 [hep-ex]`.

[29] S. Alekhin, J. Blumlein, and S. Moch, "The ABM parton distributions tuned to LHC data," *Phys. Rev.* **D89** (2014) no. 5, 054028, `arXiv:1310.3059 [hep-ph]`.

[30] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato, "Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data," *Phys. Rev.* **D93** (2016) no. 11, 114017, `arXiv:1602.03154 [hep-ph]`.

[31] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, "New parton distributions for collider physics," *Phys. Rev.* **D82** (2010) 074024, `arXiv:1007.2241 [hep-ph]`.

[32] J. M. Campbell, J. W. Huston, and W. J. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics," *Rept. Prog. Phys.* **70** (2007) 89, `arXiv:hep-ph/0611148 [hep-ph]`.

[33] A. Kusina, T. Stavreva, S. Berge, F. I. Olness, I. Schienbein, K. Kovarik, T. Jezo, J. Y. Yu, and K. Park, "Strange Quark PDFs and Implications for Drell-Yan Boson Production at the LHC," *Phys. Rev.* **D85** (2012) 094028, `arXiv:1203.1290 [hep-ph]`.

[34] K. Kovarik, T. Stavreva, A. Kusina, T. Jezo, F. I. Olness, I. Schienbein, and J. Y. Yu, "A Survey of Heavy Quark Theory for PDF Analyses," *Nucl. Phys. Proc. Suppl.* **222-224** (2012) 52–60, `arXiv:1201.1946 [hep-ph]`.

[35] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP* **07** (2002) 012, `arXiv:hep-ph/0201195 [hep-ph]`.

[36] S. Alekhin *et al.*, "The PDF4LHC Working Group Interim Report," `arXiv:1101.0536 [hep-ph]`.

[37] J. Gao and P. Nadolsky, "A meta-analysis of parton distribution functions," *JHEP* **07** (2014) 035, `arXiv:1401.0013 [hep-ph]`.

[38] P. M. Nadolsky and Z. Sullivan, "PDF uncertainties in WH production at Tevatron," *eConf* **C010630** (2001) P510, `arXiv:hep-ph/0110378 [hep-ph]`.

[39] **ATLAS** Collaboration, G. Aad *et al.*, "Z boson production in p+Pb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV measured with the ATLAS detector," *Phys. Rev.* **C92** (2015) no. 4, 044915, `arXiv:1507.06232 [hep-ex]`.

[40] T. A. collaboration, "Measurement of $W \to \mu\nu$ production in p+Pb collision at $\sqrt{s_{NN}}$ = 5.02 TeV with ATLAS detector at the LHC,".

[41] **CMS** Collaboration, V. Khachatryan *et al.*, "Study of Z boson production in pPb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV," *Phys. Lett.* **B759** (2016) 36–57, `arXiv:1512.06461 [hep-ex]`.

[42] **CMS** Collaboration, V. Khachatryan *et al.*, "Study of W boson production in pPb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV," *Phys. Lett.* **B750** (2015) 565–586, `arXiv:1503.05825 [nucl-ex]`.

[43] **LHCb** Collaboration, R. Aaij *et al.*, "Observation of Z production in proton-lead collisions at LHCb," *JHEP* **09** (2014) 030, `arXiv:1406.2885 [hep-ex]`.

[44] **ALICE** Collaboration, K. Senosi, "Measurement of W-boson production in p-Pb collisions at the LHC with ALICE," *PoS* **Bormio2015** (2015) 042, `arXiv:1511.06398 [hep-ex]`.

[45] **ATLAS** Collaboration, G. Aad *et al.*, "Measurement of Z boson Production in Pb+Pb Collisions at $\sqrt{s_{NN}}$ = 2.76 TeV with the ATLAS Detector," *Phys. Rev. Lett.* **110** (2013) no. 2, 022301, `arXiv:1210.6486 [hep-ex]`.

[46] **ATLAS** Collaboration, G. Aad *et al.*, "Measurement of the production and lepton charge asymmetry of W bosons in Pb+Pb collisions at $\sqrt{\mathbf{s_{NN}}}$ = **2.76 TeV** with the ATLAS detector," *Eur. Phys. J.* **C75** (2015) no. 1, 23, `arXiv:1408.4674 [hep-ex]`.

[47] **CMS** Collaboration, S. Chatrchyan *et al.*, "Study of Z production in PbPb and pp collisions at $\sqrt{s_{NN}}$ = 2.76 TeV in the dimuon and dielectron decay channels," *JHEP* **03** (2015) 022, `arXiv:1410.4825 [nucl-ex]`.

[48] **CMS** Collaboration, S. Chatrchyan *et al.*, "Study of W boson production in PbPb and pp collisions at $\sqrt{s_{NN}}$ = 2.76 TeV," *Phys. Lett.* **B715** (2012) 66–87, `arXiv:1205.6334 [nucl-ex]`.

[49] H. Paukkunen and C. A. Salgado, "Constraints for the nuclear parton distributions from Z and W production at the LHC," *JHEP* **03** (2011) 071, `arXiv:1010.5392 [hep-ph]`.

[50] N. Armesto, H. Paukkunen, J. M. Penn, C. A. Salgado, and P. Zurita, "An analysis of the impact of LHC Run I protonlead data on nuclear parton densities," *Eur. Phys. J.* **C76** (2016) no. 4, 218, `arXiv:1512.01528 [hep-ph]`.

[51] P. Ru, S. A. Kulagin, R. Petti, and B.-W. Zhang, "Study of $W^{\pm}$ and Z boson production in proton-lead collisions at the LHC with Kulagin-Petti nuclear parton distributions," *Phys. Rev.* **D94** (2016) no. 11, 113013, `arXiv:1608.06835 [nucl-th]`.

[52] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, "Global Analysis of Nuclear Parton Distributions," *Phys. Rev.* **D85** (2012) 074028, `arXiv:1112.6324 [hep-ph]`.

[53] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, "EPPS16: Nuclear parton distributions with LHC data," *Eur. Phys. J.* **C77** (2017) no. 3, 163, `arXiv:1612.05741 [hep-ph]`.

[54] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, "Parton distributions in the LHC era: MMHT 2014 PDFs," *Eur. Phys. J.* **C75** (2015) no. 5, 204, `arXiv:1412.3989 [hep-ph]`.

[55] H. Khanpour and S. Atashbar Tehrani, "Global Analysis of Nuclear Parton Distribution Functions and Their Uncertainties at Next-to-Next-to-Leading Order," *Phys. Rev.* **D93** (2016) no. 1, 014026, `arXiv:1601.00939 [hep-ph]`.

[56] K. Kovarik, I. Schienbein, F. I. Olness, J. Y. Yu, C. Keppel, J. G. Morfin, J. F. Owens, and T. Stavreva, "Nuclear corrections in neutrino-nucleus DIS and their compatibility with global NPDF analyses," *Phys. Rev. Lett.* **106** (2011) 122301, `arXiv:1012.0286 [hep-ph]`.

[57] S. X. Nakamura *et al.*, "Towards a Unified Model of Neutrino-Nucleus Reactions for Neutrino Oscillation Experiments," *Rept. Prog. Phys.* **80** (2017) no. 5, 056301, `arXiv:1610.01464 [nucl-th]`.

[58] **NuTeV** Collaboration, D. Mason *et al.*, "Measurement of the Nucleon Strange-Antistrange Asymmetry at Next-to-Leading Order in QCD from NuTeV Dimuon Data," *Phys. Rev. Lett.* **99** (2007) 192001.

[59] J.-w. Qiu, "QCD factorization and rescattering in proton nucleus collisions," `arXiv:hep-ph/0305161 [hep-ph]`.

[60] W. T. Giele and S. Keller, "Implications of hadron collider observables on parton distribution function uncertainties," *Phys. Rev.* **D58** (1998) 094023, `arXiv:hep-ph/9803393 [hep-ph]`.

[61] **NNPDF** Collaboration, R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, "Reweighting NNPDFs: the W lepton asymmetry," *Nucl. Phys.* **B849** (2011) 112–143, `arXiv:1012.0836 [hep-ph]`. [Erratum: Nucl. Phys.B855,927(2012)].

[62] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, "Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data," *Nucl. Phys.* **B855** (2012) 608–638, `arXiv:1108.1758 [hep-ph]`.

[63] N. Sato, J. F. Owens, and H. Prosper, "Bayesian Reweighting for Global Fits," *Phys. Rev.* **D89** (2014) no. 11, 114020, `arXiv:1310.1089 [hep-ph]`.

[64] H. Paukkunen and P. Zurita, "PDF reweighting in the Hessian matrix approach," *JHEP* **12** (2014) 100, `arXiv:1402.6623 [hep-ph]`.

[65] G. Watt and R. S. Thorne, "Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs," *JHEP* **08** (2012) 052, `arXiv:1205.4024 [hep-ph]`.

[66] T.-J. Hou *et al.*, "Reconstruction of Monte Carlo replicas from Hessian parton distributions," *JHEP* **03** (2017) 099, `arXiv:1607.06066 [hep-ph]`.

[67] F. Arleo, . Chapon, and H. Paukkunen, "Scaling properties of inclusive $W^{\pm}$ production at hadron colliders," *Eur. Phys. J.* **C76** (2016) no. 4, 214, `arXiv:1509.03993 [hep-ph]`.

[68] A. Kusina, F. Lyonnet, D. B. Clark, E. Godat, T. Jezo, K. Kovarik, F. I. Olness, I. Schienbein, and J. Y. Yu, "LHC lead data and nuclear PDFs," *Acta Phys. Polon.* **B48** (2017) 1035, `arXiv:1705.06704 [hep-ph]`.

[69] **CHORUS** Collaboration, G. Onengut *et al.*, "Measurement of nucleon structure functions in neutrino scattering," *Phys. Lett.* **B632** (2006) 65–75.

[70] M. A. G. Aivazis, F. I. Olness, and W.-K. Tung, "QCD formulation of charm production in deep inelastic scattering and the sea quark - gluon dichotomy," *Phys. Rev. Lett.* **65** (1990) 2339–2342.

[71] M. A. G. Aivazis, F. I. Olness, and W.-K. Tung, "Leptoproduction of heavy quarks. 1. General formalism and kinematics of charged current and neutral current production processes," *Phys. Rev.* **D50** (1994) 3085–3101, `arXiv:hep-ph/9312318 [hep-ph]`.

[72] M. A. G. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, "Leptoproduction of heavy quarks. 2. A Unified QCD formulation of charged and neutral current processes from fixed target to collider energies," *Phys. Rev.* **D50** (1994) 3102–3118, `arXiv:hep-ph/9312319 [hep-ph]`.

[73] R. S. Thorne and R. G. Roberts, "A Variable number flavor scheme for charged current heavy flavor structure functions," *Eur. Phys. J.* **C19** (2001) 339–349, `arXiv:hep-ph/0010344 [hep-ph]`.

[74] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Heavy-quark mass dependence in global PDF analyses and 3- and 4-flavour parton distributions," *Eur. Phys. J.* **C70** (2010) 51–72, `arXiv:1007.2624 [hep-ph]`.

[75] S. Forte, E. Laenen, P. Nason, and J. Rojo, "Heavy quarks in deep-inelastic scattering," *Nucl. Phys.* **B834** (2010) 116–162, `arXiv:1001.2312 [hep-ph]`.

[76] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology," *Nucl. Phys.* **B849** (2011) 296–363, `arXiv:1101.1300 [hep-ph]`.

[77] S. Alekhin, J. Blumlein, S. Klein, and S. Moch, "The 3, 4, and 5-flavor NNLO Parton from Deep-Inelastic-Scattering Data and at Hadron Colliders," *Phys. Rev.* **D81** (2010) 014032, `arXiv:0908.2766 [hep-ph]`.

[78] S. Alekhin, J. Blumlein, and S. Moch, "Parton Distribution Functions and Benchmark Cross Sections at NNLO," *Phys. Rev.* **D86** (2012) 054009, `arXiv:1202.2281 [hep-ph]`.

[79] T. Stavreva, F. I. Olness, I. Schienbein, T. Jezo, A. Kusina, K. Kovarik, and J. Y. Yu, "Heavy Quark Production in the ACOT Scheme at NNLO and N3LO," *Phys. Rev.* **D85** (2012) 114014, `arXiv:1203.0282 [hep-ph]`.

[80] S. Qian, *THE CWZ SUBTRACTION SCHEME (A NEW RENORMALIZATION PRESCRIPTION FOR QCD) AND ITS APPLICATION.* PhD thesis, IIT, Chicago, 1985. http://wwwlib.umi.com/dissertations/fullcit?p8517585.

[81] J. C. Collins and W.-K. Tung, "Calculating Heavy Quark Distributions," *Nucl. Phys.* **B278** (1986) 934.

[82] M. Buza, Y. Matiounine, J. Smith, R. Migneron, and W. L. van Neerven, "Heavy quark coefficient functions at asymptotic values Q**2 ¿¿ m**2," *Nucl. Phys.* **B472** (1996) 611–658, `arXiv:hep-ph/9601302 [hep-ph]`.

[83] **Particle Data Group** Collaboration, C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys.* **C40** (2016) no. 10, 100001.

[84] S. Alekhin *et al.*, "HERAFitter," *Eur. Phys. J.* **C75** (2015) no. 7, 304, `arXiv:1410.4412 [hep-ph]`.

[85] **xFitter team** Collaboration, O. Zenaiev, "xFitter project," *PoS* **DIS2016** (2016) 033.

[86] V. Bertone, A. Glazov, A. Mitov, A. Papanastasiou, and M. Ubiali, "Heavy-flavor parton distributions without heavy-flavor matching prescriptions," `arXiv:1711.03355 [hep-ph]`.

[87] J. C. Collins, F. Wilczek, and A. Zee, "Low-Energy Manifestations of Heavy Particles: Application to the Neutral Current," *Phys. Rev.* **D18** (1978) 242.

[88] J. C. Collins, "Hard scattering factorization with heavy quarks: A General treatment," *Phys. Rev.* **D58** (1998) 094002, `arXiv:hep-ph/9806259 [hep-ph]`.

[89] F. Maltoni, G. Ridolfi, and M. Ubiali, "b-initiated processes at the LHC: a reappraisal," *JHEP* **07** (2012) 022, `arXiv:1203.6393 [hep-ph]`. [Erratum: JHEP04,095(2013)].

[90] M. Lim, F. Maltoni, G. Ridolfi, and M. Ubiali, "Anatomy of double heavy-quark initiated processes," *JHEP* **09** (2016) 132, `arXiv:1605.09411 [hep-ph]`.

[91] **NNPDF** Collaboration, R. D. Ball, V. Bertone, L. Del Debbio, S. Forte, A. Guffanti, J. Rojo, and M. Ubiali, "Theoretical issues in PDF determination and associated uncertainties," *Phys. Lett.* **B723** (2013) 330–339, `arXiv:1303.1189 [hep-ph]`.

[92] M. Cacciari, M. Greco, and P. Nason, "The P(T) spectrum in heavy flavor hadroproduction," *JHEP* **05** (1998) 007, `arXiv:hep-ph/9803400 [hep-ph]`.

[93] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, "Inclusive B-meson production at small p_T in the general-mass variable-flavor-number scheme," *Eur. Phys. J.* **C75** (2015) no. 3, 140, `arXiv:1502.01001 [hep-ph]`.

[94] M. Bonvini, A. S. Papanastasiou, and F. J. Tackmann, "Resummation and matching of b-quark mass effects in $b\bar{b}H$ production," *JHEP* **11** (2015) 196, `arXiv:1508.03288 [hep-ph]`.

[95] M. Bonvini, A. S. Papanastasiou, and F. J. Tackmann, "Matched predictions for the $b\bar{b}H$ cross section at the 13 TeV LHC," *JHEP* **10** (2016) 053, `arXiv:1605.01733 [hep-ph]`.

[96] **NNPDF** Collaboration, R. D. Ball, V. Bertone, M. Bonvini, S. Carrazza, S. Forte, A. Guffanti, N. P. Hartland, J. Rojo, and L. Rottoli, "A Determination of the Charm Content of the Proton," *Eur. Phys. J.* **C76** (2016) no. 11, 647, `arXiv:1605.06515 [hep-ph]`.

[97] R. D. Ball, V. Bertone, M. Bonvini, S. Forte, P. Groth Merrild, J. Rojo, and L. Rottoli, "Intrinsic charm in a matched general-mass scheme," *Phys. Lett.* **B754** (2016) 49–58, `arXiv:1510.00009 [hep-ph]`.

[98] R. D. Ball, M. Bonvini, and L. Rottoli, "Charm in Deep-Inelastic Scattering," *JHEP* **11** (2015) 122, `arXiv:1510.02491 [hep-ph]`.

[99] F. Lyonnet, A. Kusina, T. Jeo, K. Kovark, F. Olness, I. Schienbein, and J.-Y. Yu, "On the intrinsic bottom content of the nucleon and its impact on heavy new physics at the LHC," *JHEP* **07** (2015) 141, `arXiv:1504.05156 [hep-ph]`.

[100] **H1** Collaboration, F. D. Aaron *et al.*, "Measurement of the Charm and Beauty Structure Functions using the H1 Vertex Detector at HERA," *Eur. Phys. J.* **C65** (2010) 89–109, `arXiv:0907.2643 [hep-ex]`.

[101] **ZEUS, H1** Collaboration, H. Abramowicz *et al.*, "Combination and QCD Analysis of Charm Production Cross Section Measurements in Deep-Inelastic ep Scattering at HERA," *Eur. Phys. J.* **C73** (2013) no. 2, 2311, `arXiv:1211.1182 [hep-ex]`.

[102] **ZEUS** Collaboration, H. Abramowicz *et al.*, "Measurement of beauty and charm production in deep inelastic scattering at HERA and measurement of the beauty-quark mass," *JHEP* **09** (2014) 127, `arXiv:1405.6915 [hep-ex]`.

[103] V. Bertone, S. Carrazza, and N. P. Hartland, "APFELgrid: a high performance tool for parton density determinations," *Comput. Phys. Commun.* **212** (2017) 205–209, `arXiv:1605.02070 [hep-ph]`.

[104] **xFitter Developers' Team** Collaboration, V. Bertone *et al.*, "A determination of $m_c(m_c)$ from HERA data using a matched heavy-flavor scheme," *JHEP* **08** (2016) 050, `arXiv:1605.01946 [hep-ph]`.

[105] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations," *Comput. Phys. Commun.* **10** (1975) 343–367.

[106] **H1** Collaboration, F. D. Aaron *et al.*, "Inclusive Deep Inelastic Scattering at High Q^2 with Longitudinally Polarised Lepton Beams at HERA," *JHEP* **09** (2012) 061, `arXiv:1206.7007 [hep-ex]`.

[107] M. Bonvini, S. Marzani, and T. Peraro, "Small-x resummation from HELL," *Eur. Phys. J.* **C76** (2016) no. 11, 597, `arXiv:1607.02153 [hep-ph]`.

[108] D. B. Clark, E. Godat, and F. I. Olness, "ManeParse : A Mathematica reader for Parton Distribution Functions," *Comput. Phys. Commun.* **216** (2017) 126–137, `arXiv:1605.08012 [hep-ph]`.

[109] M. Botje, "Erratum for the time-like evolution in QCDNUM," `arXiv:1602.08383 [hep-ph]`.

[110] F. Caola, S. Forte, and J. Rojo, "HERA data and DGLAP evolution: Theory and phenomenology," *Nucl. Phys.* **A854** (2011) 32–44, `arXiv:1007.5405 [hep-ph]`.

[111] M. Krmer, F. I. Olness, and D. E. Soper, "Treatment of heavy quarks in deeply inelastic scattering," *Phys. Rev.* **D62** (2000) 096007, `arXiv:hep-ph/0003035 [hep-ph]`.

[112] T. Stavreva, I. Schienbein, F. Arleo, K. Kovarik, F. Olness, J. Y. Yu, and J. F. Owens, "Probing gluon and heavy-quark nuclear PDFs with gamma + Q production in pA collisions," *JHEP* **01** (2011) 152, `arXiv:1012.1178 [hep-ph]`.

[113] J. F. Owens, J. Huston, C. E. Keppel, S. Kuhlmann, J. G. Morfin, F. Olness, J. Pumplin, and D. Stump, "The Impact of new neutrino DIS and Drell-Yan data on large-x parton distributions," *Phys. Rev.* **D75** (2007) 054030, `arXiv:hep-ph/0702159 [HEP-PH]`.

[114] D. Stump, J. Huston, J. Pumplin, W.-K. Tung, H. L. Lai, S. Kuhlmann, and J. F. Owens, "Inclusive jet production, parton distributions, and the search for new physics," *JHEP* **10** (2003) 046, `arXiv:hep-ph/0303013 [hep-ph]`.

[115] D. de Florian and R. Sassot, "Nuclear parton distributions at next-to-leading order," *Phys. Rev.* **D69** (2004) 074028, `arXiv:hep-ph/0311227 [hep-ph]`.

[116] **New Muon** Collaboration, M. Arneodo *et al.*, "Measurement of the proton and deuteron structure functions, F2(p) and F2(d), and of the ratio sigma-L / sigma-T," *Nucl. Phys.* **B483** (1997) 3–43, `arXiv:hep-ph/9610231 [hep-ph]`.

[117] **HERMES** Collaboration, A. Airapetian *et al.*, "Measurement of R = sigma(L) / sigma(T) in deep inelastic scattering on nuclei," `arXiv:hep-ex/0210068 [hep-ex]`.

[118] **New Muon** Collaboration, P. Amaudruz *et al.*, "A Reevaluation of the nuclear structure function ratios for D, He, Li-6, C and Ca," *Nucl. Phys.* **B441** (1995) 3–11, `arXiv:hep-ph/9503291 [hep-ph]`.

[119] J. Gomez *et al.*, "Measurement of the A-dependence of deep inelastic electron scattering," *Phys. Rev.* **D49** (1994) 4348–4372.

[120] **New Muon** Collaboration, M. Arneodo *et al.*, "The Structure Function ratios F2(li) / F2(D) and F2(C) / F2(D) at small x," *Nucl. Phys.* **B441** (1995) 12–30, `arXiv:hep-ex/9504002 [hep-ex]`.

[121] **E665** Collaboration, M. R. Adams *et al.*, "Shadowing in inelastic scattering of muons on carbon, calcium and lead at low x(Bj)," *Z. Phys.* **C67** (1995) 403–410, `arXiv:hep-ex/9505006 [hep-ex]`.

[122] **European Muon** Collaboration, J. Ashman *et al.*, "Measurement of the Ratios of Deep Inelastic Muon - Nucleus Cross-Sections on Various Nuclei Compared to Deuterium," *Phys. Lett.* **B202** (1988) 603–610.

[123] **European Muon** Collaboration, M. Arneodo *et al.*, "Measurements of the nucleon structure function in the range $0.002 - \mathrm{GeV}^2 < x < 0.17 - \mathrm{GeV}^2$ and $0.2 - GeV^2 < q^2 < 8 - GeV^2$ in deuterium, carbon and calcium," *Nucl. Phys.* **B333** (1990) 1–47.

[124] **BCDMS** Collaboration, G. Bari *et al.*, "A Measurement of Nuclear Effects in Deep Inelastic Muon Scattering on Deuterium, Nitrogen and Iron Targets," *Phys. Lett.* **163B** (1985) 282.

[125] A. Bodek *et al.*, "A Comparison of the Deep Inelastic Structure Functions of Deuterium and Aluminum Nuclei," *Phys. Rev. Lett.* **51** (1983) 534.

[126] A. Bodek *et al.*, "Electron Scattering from Nuclear Targets and Quark Distributions in Nuclei," *Phys. Rev. Lett.* **50** (1983) 1431.

[127] S. Dasu *et al.*, "Measurement of kinematic and nuclear dependence of R = sigma-L / sigma-t in deep inelastic electron scattering," *Phys. Rev.* **D49** (1994) 5641–5670.

[128] **BCDMS** Collaboration, A. C. Benvenuti *et al.*, "Nuclear Effects in Deep Inelastic Muon Scattering on Deuterium and Iron Targets," *Phys. Lett.* **B189** (1987) 483–487.

[129] **European Muon** Collaboration, J. Ashman *et al.*, "A Measurement of the ratio of the nucleon structure function in copper and deuterium," *Z. Phys.* **C57** (1993) 211–218.

[130] **E665** Collaboration, M. R. Adams *et al.*, "Saturation of shadowing at very low x_{BJ}," *Phys. Rev. Lett.* **68** (1992) 3266–3269.

[131] **New Muon** Collaboration, M. Arneodo *et al.*, "The A dependence of the nuclear structure function ratios," *Nucl. Phys.* **B481** (1996) 3–22.

[132] **New Muon** Collaboration, M. Arneodo *et al.*, "The Q**2 dependence of the structure function ratio F2 Sn / F2 C and the difference R Sn - R C in deep inelastic muon scattering," *Nucl. Phys.* **B481** (1996) 23–39.

[133] D. M. Alde *et al.*, "Nuclear dependence of dimuon production at 800-GeV. FNAL-772 experiment," *Phys. Rev. Lett.* **64** (1990) 2479–2482.

[134] **NuSea** Collaboration, M. A. Vasilev *et al.*, "Parton energy loss limits and shadowing in Drell-Yan dimuon production," *Phys. Rev. Lett.* **83** (1999) 2304–2307, arXiv:hep-ex/9906010 [hep-ex].

[135] A. Accardi *et al.*, "Hard probes in heavy ion collisions at the lhc: pdfs, shadowing and pa collisions," in *3rd Workshop on Hard Probes in Heavy Ion Collisions: 3rd Plenary Meeting Geneva, Switzerland, October 7-11, 2002.* 2004. arXiv:hep-ph/0308248 [hep-ph]. https://inspirehep.net/record/642052/files/arXiv:hep-ph_0308248.pdf.

[136] J. Arrington, R. Ent, C. E. Keppel, J. Mammei, and I. Niculescu, "Low Q scaling, duality, and the EMC effect," *Phys. Rev.* **C73** (2006) 035205, arXiv:nucl-ex/0307012 [nucl-ex].

[137] S. Malace, D. Gaskell, D. W. Higinbotham, and I. Cloet, "The Challenge of the EMC Effect: existing data and future directions," *Int. J. Mod. Phys.* **E23** (2014) no. 08, 1430013, arXiv:1405.1270 [nucl-ex].

[138] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC," *JHEP* **07** (2011) 018, arXiv:1105.0020 [hep-ph].

[139] G. P. Salam and J. Rojo, "A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)," *Comput. Phys. Commun.* **180** (2009) 120–156, arXiv:0804.3755 [hep-ph].

[140] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, "A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project," *Eur. Phys. J.* **C66** (2010) 503–524, arXiv:0911.2985 [hep-ph].

[141] W. Giele *et al.*, "The QCD / SM working group: Summary report," in *Physics at TeV colliders. Proceedings, Euro Summer School, Les Houches, France, May 21-June 1, 2001*, pp. 275–426. 2002. arXiv:hep-ph/0204316 [hep-ph]. http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-02-410.

[142] M. Dittmar *et al.*, "Working Group I: Parton distributions: Summary report for the HERA LHC Workshop Proceedings," arXiv:hep-ph/0511119 [hep-ph].

[143] J. Rojo, "PDF4LHC recommendations for Run II," *PoS* **DIS2016** (2016) 018, arXiv:1606.08243 [hep-ph].

[144] J. M. Campbell, R. K. Ellis, and W. T. Giele, "A Multi-Threaded Version of MCFM," *Eur. Phys. J.* **C75** (2015) no. 6, 246, arXiv:1503.06182 [physics.comp-ph].

[145] R. M. Chatterjee, M. Guchait, and R. Placakyte, "QCD analysis of the CMS inclusive differential Z production data at $\sqrt{s} = 8$TeV," *Phys. Rev.* **D94** (2016) no. 3, 034035, arXiv:1603.09619 [hep-ph].

[146] G. Balossini, G. Montagna, C. M. Carloni Calame, M. Moretti, O. Nicrosini, F. Piccinini, M. Treccani, and A. Vicini, "Combination of electroweak and QCD corrections to single W production at the Fermilab Tevatron and the CERN LHC," *JHEP* **01** (2010) 013, arXiv:0907.0276 [hep-ph].

[147] J. M. Campbell, J. Rojo, E. Slade, and C. Williams, "Direct photon production and PDF fits reloaded," arXiv:1802.03021 [hep-ph].

[148] D. S. J. W. F. Zimmermann, M. Benedikt, "Challenges for Highest Energy Circular Colliders,".

[149] **CMS** Collaboration, S. Chatrchyan *et al.*, "Studies of dijet transverse momentum balance and pseudorapidity distributions in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\ TeV$," *Eur. Phys. J.* **C74** (2014) no. 7, 2951, arXiv:1401.4433 [nucl-ex].

[150] **NuTeV** Collaboration, V. A. Radescu, "Differential cross section results from NuTeV," in *Proceedings, 12th International Workshop on Deep Inelastic Scattering (DIS 2004): Strbske Pleso, Slovakia, April 14-18, 2004*, pp. 386–391. 2004. arXiv:hep-ex/0408006 [hep-ex]. http://www.saske.sk/dis04/proceedings/A/radescu.ps.gz.

[151] K. Kovarik, "Update and Comparison of Nuclear Parton Distribution Functions and Neutrino DIS," in *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012): Bonn, Germany, March 26-30, 2012*, pp. 425–428. 2012. arXiv:1206.2178 [hep-ph]. https://inspirehep.net/record/1117866/files/arXiv:1206.2178.pdf. [,425(2012)].

[152] **MINERvA** Collaboration, J. Mousseau *et al.*, "Measurement of Partonic Nuclear Effects in Deep-Inelastic Neutrino Scattering using MINERvA," *Phys. Rev.* **D93** (2016) no. 7, 071101, arXiv:1601.06313 [hep-ex].

[153] A. Kurepin and N. Topilskaya, "Heavy-ion collisions in a fixed target mode at the LHC beams," *EPJ Web Conf.* **138** (2017) 03009.

[154] **LHCb** Collaboration, E. Maurice, "Fixed-target physics at LHCb," in *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. arXiv:1708.05184 [hep-ex]. https://inspirehep.net/record/1616496/files/arXiv:1708.05184.pdf.

[155] B.-T. Wang, T. J. Hobbs, S. Doyle, J. Gao, T.-J. Hou, P. M. Nadolsky, and F. I. Olness, "Visualizing the sensitivity of hadronic experiments to nucleon structure," arXiv:1803.02777 [hep-ph].

[156] R. D. Ball *et al.*, "Parton distributions with LHC data," *Nucl. Phys.* **B867** (2013) 244–289, arXiv:1207.1303 [hep-ph].