Fall 12-21-2019

Constraints on Parton Distribution Functions Imposed by Hadronic Experiments

Boting Wang
Southern Methodist University, botingw@smu.edu

CONSTRAINTS ON PARTON DISTRIBUTION FUNCTIONS IMPOSED BY HADRONIC EXPERIMENTS

Approved by:

Dr. Pavel Nadolsky
Associate Professor, Ph. D. Advisor

Dr. Fredrick I. Olness
Professor, Ph. D. Advisor

Dr. Jingbo Ye
Professor

Dr. Daniel Stump
Professor - Michigan State University

CONSTRAINTS ON PARTON DISTRIBUTION FUNCTIONS IMPOSED BY HADRONIC EXPERIMENTS

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

Boting Wang

B.S., Physics, National Tsin Hua University
M.S., Physics, National Taiwan University

December 21, 2019

Wang, Boting

B.S., Physics, National Tsin Hua University
M.S., Physics, National Taiwan University

Constraints on Parton Distribution Functions Imposed by Hadronic Experiments

Advisor: Dr. Pavel Nadolsky, Dr. Fredrick I. Olness
Doctor of Philosophy degree conferred December 21, 2019
Dissertation completed October 10, 2019

The theoretical uncertainties of the Large Hadron Collider (LHC) observables are decreasing with the increasing statistics of the LHC experiments, and it is becoming more and more important to reduce the uncertainties in the knowledge of the nucleon structure. The latest LHC high-energy experiments, future experimental proposals, and computational tools are expected to enhance the knowledge of the nucleon structure. However, the global analysis that assesses their impact on Parton Distribution Functions (PDFs) knowledge is computationally expensive due to the corresponding large size of data. I developed a new approach that can make a quick preliminary evaluation to help the analysis process. It quantifies the impact of hadronic experiments on PDFs based on Hessian correlation and quality of measurements. This approach is accessible through an open source software called PDFSense. I used PDFSense and other statistical methods to evaluate the impact of the latest LHC datasets on PDFs. In addition, I used this software to investigate the synergy of lattice calculations and $\overline{MS}$ PDFs on improving the picture of the nucleon's collinear structure. The assessment of the impact of PDFs knowledge imposed by high luminosity upgrade to the LHC (HL-LHC), Large Hadron-electron Collider (LHeC), and Electron Ion Collider (EIC), some future high-energy experimental proposals, is also implemented.

TABLE OF CONTENTS

CHAPTER

APPENDIX

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

In ancient times, the atom was thought to be the fundamental, indivisible element of matter from which the universe was constructed. During the last century, we have split the atom and developed a whole field of sub-atomic physics studying the components. We now know the atom is comprised of a cloud of electrons surrounding a dense nuclear core consisting of protons and neutrons. My research is conducted in the sub-nuclear domain where we are investigating the constituents that make up the protons and neutrons as well as the very strong forces that bind these constituents inside the nucleus. The past work on these issues have won several Nobel Prizes. A deeper understanding of them is of great help in exploring other phenomena in the universe. In this thesis, I will present detailed studies of three projects that I participated in.

1. We developed a statistical framework (and the corresponding software) to efficiently evaluate the impact of recent experimental data on our theoretical models.

2. We used these tools to consider the impact of proposed future experiments on our field.

3. We assessed the synergy between our approach driven by experiments and perturbative calculations with the non-perturbative Lattice QCD calculations.

Together, these projects represent important advances within our field, and these will enable us to advance our knowledge of the underlying structure of the nuclei which lie at the heart of all matter. In the remainder of this chapter, I will outline the background knowledge related to the structure of the nuclei, and give a more in-depth look at my research in the last section of this chapter.

1.1. The Standard Model

For a long time, physicists have tried to find out the fundamental building blocks of matter and the fundamental forces in the universe. The Standard Model (SM) [1,2] describes 17 kinds of elementary particles and three of the four currently known interactions in the universe, namely, the strong force, the weak force, and the electromagnetic force.[1]

The SM contains two types of particles: bosons and fermions. Bosons are particles with integer spins. Bosons in the SM include four gauge bosons and a special Higgs boson. Fermions are particles with half-integer spins. The table in Fig. 1.1 summarize the particles and their properties in the SM.

The gauge bosons in the SM are particles of spin-1 and are responsible for transmitting the three forces previously discussed. Photons are responsible for transmitting the electromagnetic force. Since photons are massless, the electromagnetic force can propagate over long distances. W and Z bosons are responsible for transmitting the weak force. W and Z bosons can only propagate within a short distance. Gluons are responsible for transmitting the strong force. Although gluons have no mass, like photons, due to the fact that QCD is self interacting (I will discuss it in the QCD section), the effective range of the strong force is limited to the size of nucleons.

In addition, the spin-0 scalar boson in the SM is the Higgs particle, which is responsible for the spontaneous symmetry breaking of the electroweak field and giving W and Z gauge bosons mass. The Higgs particle discovered in the Large Hadron Collider (LHC) experiment is a powerful validation of the SM.

The fermions in the SM are spin-1/2 particles that make up most of the matter we know. Fermions can be divided into two broad categories: leptons and quarks. There are two

[1] The gravitational force has not yet been successfully integrated into the SM framework. It describes the rules of celestial body movement and many phenomena on the ground, such as free fall motion and projectile motion.

groups of leptons. The first group contains electrons, muons, and taus. The first group of leptons carry electrical charge and weak isospin, which are affected by the electromagnetic force and the weak force. The second group contains electron neutrinos, muon neutrinos, and tau neutrinos. The second group of leptons carry weak isospin, which are affected by the weak force. Quarks carry electrical charge, weak isospin, and color, which are affected by the electromagnetic force, the weak force, and the strong force.

There are three generations of fermions in the SM. The three generations are the three columns to the left of Fig. 1.1. Each generation has two types of leptons and two types of quarks, as shown in the four rows in Fig. 1.1. For each type of lepton and quark except for neutrinos, the mass of the particles increases with generation. The fermions in the first generation (the left-most column) are the lightest, and most of the matter in the universe is made up of them because heavy particles tend to decay into light particles.

The SM is a quantum field theory (QFT) with the symmetry of $SU(3) \otimes SU(2) \otimes U(1)$. Each symmetry corresponds to a fundamental force accompanied by some conserved quantities, namely color, weak isospin, and hypercharge. The electromagnetic force is described by quantum electrodynamics (QED), which describes the interaction between charged particles. QED corresponds to $U(1)$ gauge symmetry. The weak force describes the phenomenon of the radioactive decay of atoms. The symmetry corresponding to the weak force is $SU(2)$. The strong force can be understood by quantum chromodynamics (QCD), describing the matter and its dynamics of quarks. QCD corresponds to $SU(3)$.

1.2. Quantum Chromodynamics

1.2.1. Overview

Quantum chromodynamics (QCD) is a non-Abelian gauge field theory based on the $SU(3)$ symmetry. This theoretical framework contains two kinds of particles (i.e. quarks

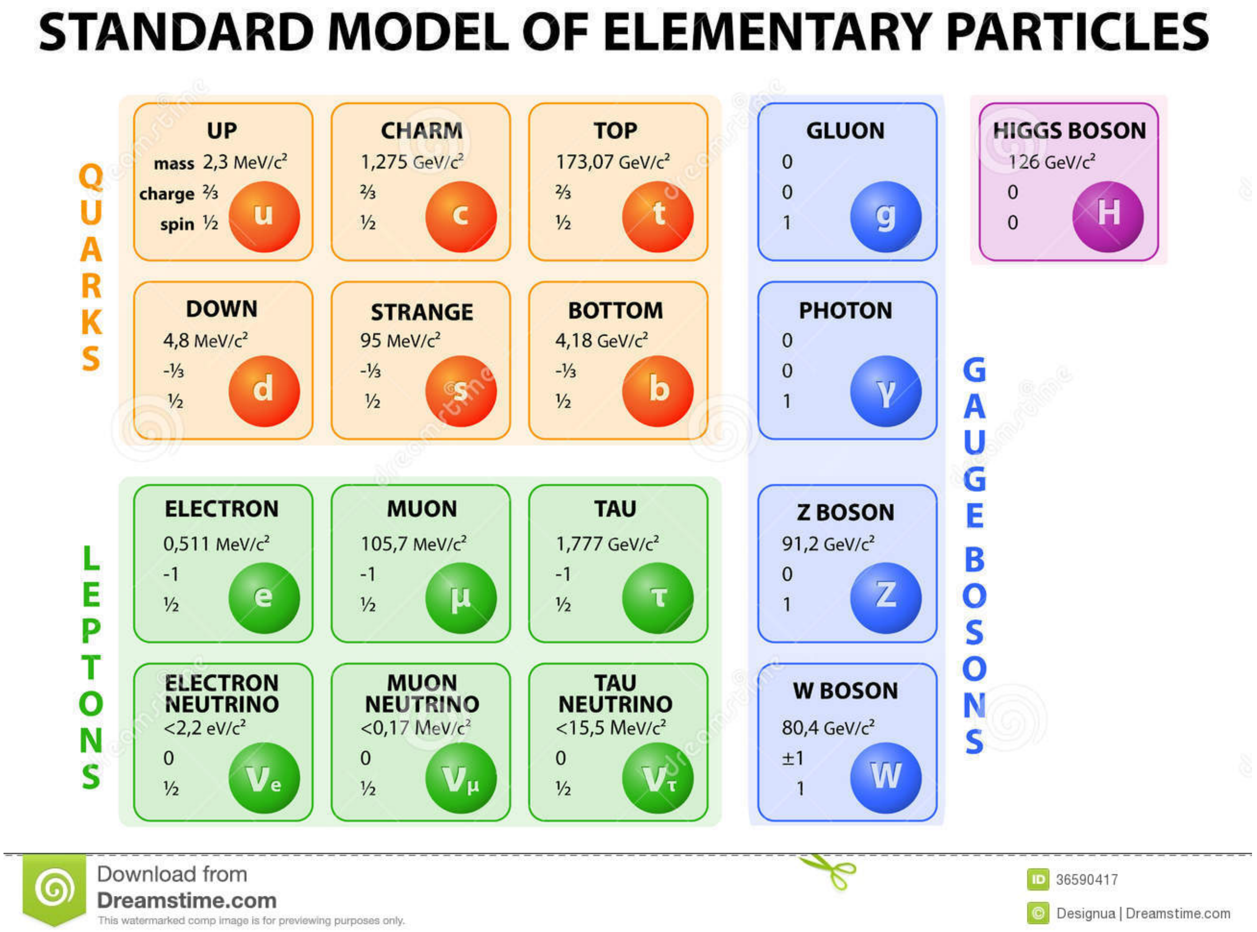


Figure 1.1: The representation of elementary particles included in the Standard Model.

and gluons). Both quarks and gluons carry color. There are three kinds of colors (red, blue, green or r, b, g) and three corresponding anti-colors (anti-red, anti-blue, and anti-green or $\bar{r}$, $\bar{b}$, $\bar{g}$). Each quark carries one color, and each anti-quark carries one anti-color. Each gluon carries both one color and one anti-color. Gluons are responsible for mediating the strong force between quarks. Since gluons can carry color charge, gluons are self-interacting, unlike photons in QED. This leads to some important features of QCD phenomenology, such as asymptotic freedom and color confinement.

1.2.2. Color Confinement

Color confinement is a very important feature of QCD. It means we are unable to see the individual color-charged particles in experiments. Thus, the color-charged quarks are confined in color-neutral composite particles called hadrons. There are two main types of hadrons: the mesons (one quark and one anti-quark (e.g. $r\bar{r}$)) and the baryons (three quarks with rgb or three anti-quarks with $\bar{r}\bar{g}\bar{b}$). rgb or $\bar{r}\bar{g}\bar{b}$ are also color-neutral. The phenomenon of color confinement can be explained by asymptotic freedom.

1.2.3. Asymptotic Freedom

Asymptotic freedom means that the strong coupling weakens as the energy scale of the interaction rises, which is opposite from QED. The coupling constant in QED increases with increasing energy scale. Conversely, QCD's coupling constant becomes weak at high energy scales. That is to say, quarks and gluons are "free" (without interaction) at very high energy scales. This phenomenon in which coupling constants change with energy is one of the important predictions given by QFT.

To understand asymptotic freedom in QCD, let us first explain why the coupling constant in QED increases with energy. In classical electromagnetism, the Coulomb repulsive force between electrons is kq_1q_2/r^2. However, in QED, the vacuum between charged particles can produce virtual electron–positron pairs, which causes the magnitude of charge perceived by charged particles to change with the distance between each other. For example, in the process of electron-electron scattering (QED), the Coulomb force is produced by the exchanged photons between two electrons. In this process, photons can create electron–positron pairs. Similar to the polarization of a dielectric medium, these pairs will be polarized, causing the screening effect. The screening effect caused by the vacuum between the two electrons increases with the distance between them. In other words, the "effective charge" felt by electrons will become stronger at high energy scales (short distances) and weaken at low energy scales (long distances).

In QCD, gluon self-interactions will enhance the strength of the gluon field, offsetting the screening effect caused by color dipoles, which we usually call the anti-screening effect. According to the calculation, as long as the number of flavors is less than 16, the anti-screening effect caused by gluons will be stronger than the screening effect caused by quarks, so that the coupling constant becomes smaller as the energy scale increases. Since QCD has only six kinds of quarks, the coupling of strong interactions between quarks is close to infinite at low energy scales and becomes weaker at high energy scales.

Fig. 1.2 [3] is the strength change of strong coupling on various energy scales: it causes a number of results including color confinement. The binding force between quarks become very strong at very low energy scales, which qualitatively explains why we can't see individual quarks and gluons in hadronic scattering experiments: since the energy between the two quarks that are far apart (low energy scale) is much higher than the two pairs of color-neutral quark pairs that are far apart, the vacuum creates extra quarks and turns the two single quarks into two quark pairs to reduce energy of systems. In addition, because the strong coupling is too large in the low energy regime, we can only believe the results of QCD perturbative theory above a scale $\sim$ 1 GeV. According to perturbative theory, QFT scattering processes can be described by an expansion in powers of the coupling constants $c_0 + c_1(\frac{Q^2}{\bar{\mu}^2})\alpha(\bar{\mu}^2) + c_2(\frac{Q^2}{\bar{\mu}^2})\alpha^2(\bar{\mu}^2) + ...$ at arbitrary renormalization scales $\bar{\mu}$. Because we want the series to converge in finite terms, usually we will select $\bar{\mu} = Q$ for the physical energy scale Q. However, when the coupling constants $\alpha(Q^2) \gtrsim 1$, physical observables cannot be accurately approximated by a finite number of terms, so the perturbation theory fails. Therefore, we must use other methods to understand the hadronic states below ~ 1 GeV. PDFs research is the main means for understanding it.

1.3. PDFs and Factorization

1.3.1. Hadronic Phenomenology

Because of asymptotic freedom, quarks and gluons are confined in hadrons and we can only see hadrons in experimental detectors. The means of studying quarks and gluons in hadrons are therefore strongly dependent on high-energy hadronic scattering experiments. This type of experiments will implement particle collisions involving hadrons in the initial state and detect debris in the collision processes to study the state inside the hadrons.

In scattering experiments, the resolution of a detecting particle is approximately equal to its wavelength. Therefore, if we want detections to be sensitive to the interior of hadrons, the wavelengths of detecting particles should be smaller than the scales of hadrons. According to the de Broglie's equation for wavelength, we can relate the wavelength of a detecting particle and its momentum by $\lambda = h/p$. The size of a proton is about $\sim 10^{-15}$ m, so we should explore its structure with momenta of detecting particles above 1 GeV.

1.3.2. DIS Processes and Parton Distribution Functions

Let us take Deep Inelastic Scattering (DIS) as an example. This type of processes contains a hadron h with the momentum p and a lepton particle ℓ with the momentum p_ℓ in the initial

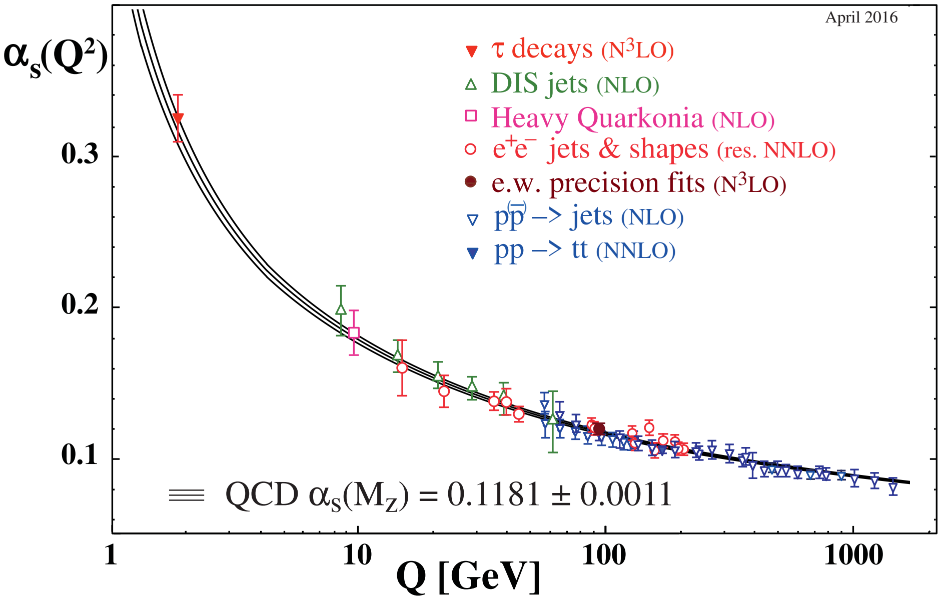


Figure 1.2: Summary of measurements of α_s as a function of the renormalization scale, here denoted as Q. This figure is in Figure 9.3 of the Review of Particle Physics [3].

state ($h(p)\ell(p_\ell) \to \ell'(p_{\ell'})X$). We can factorize the entire collision process σ at a given energy scale μ (also called the factorization scale μ) into two parts, one part is a hard scattering process $\hat{\sigma}$ with the collision between the lepton and a quark or gluon inside the hadron, and another part is collinear Parton Distribution Functions $f_a(\xi, \mu)$ (PDFs), representing the probability that a parton (quarks or gluon) of flavor "a" inside the hadron with a momentum fraction ξ of the hadron momentum participates in the interaction. The relationship between the experimental results and PDFs can be written as a convolution equation:

$$\frac{\mathrm{d}^2\sigma}{dxdQ^2} = \sum_a \int_x^1 \frac{\mathrm{d}\xi}{\xi}\, f_a(\xi, \mu^2) \frac{\mathrm{d}^2\hat{\sigma}^a}{dxdQ^2}(\frac{x}{\xi}, \frac{Q^2}{\mu^2}, \alpha_s(\bar{\mu}^2), \frac{Q^2}{\bar{\mu}^2}) \tag{1.1}$$

or in a concise representation

$$\sigma(x, Q) = \sum_a [f_a \otimes \hat{\sigma}^a](x, Q, \mu, \bar{\mu}), \tag{1.2}$$

where kinematic variables Q^2 and x are $Q^2 = -q^\mu q_\mu$ and $x = Q^2/(p \cdot q)$. q^μ is the four-momentum of the exchanged photon p_ℓ^μ-$p_{\ell'}^\mu$. For an event in DIS scattering experiments, we can measure the energy $E_{\ell'}$ of the lepton ℓ' and the angle θ between ℓ and ℓ' to reconstruct $\{x, Q^2\}$ as we can prove that

$$x = \frac{2E_\ell E_{\ell'} \sin^2(\theta/2)}{M_h(E_\ell - E_{\ell'})}, \tag{1.3}$$

$$Q^2 = 4E_\ell E_{\ell'} \sin^2(\theta/2). \tag{1.4}$$

The LHS in Eq. 1.2 is the experimental cross section σ, and the RHS in Eq. 1.2 contains the PDF $f_a(\xi, \mu)$ and the hard scattering cross section $\hat{\sigma}^a$ corresponding to the lepton-quark/gluon scattering process for parton "a". The entire cross section σ (without a hat) is the sum of the cross sections corresponding to all possible hard scattering processes. $\otimes$ stands for the convolution, which means for two functions f and g, we have

$$[f \otimes g](z) = \int \frac{\mathrm{d}x}{x}\, f(x) g(\frac{z}{x}). \tag{1.5}$$

In Eq. 1.1 and Eq. 1.2, the the hard scattering cross section $\hat{\sigma}^a$ is perturbatively calculable and the experimental cross section σ (without a hat) is measurable, which allows us to acquire the knowledge of PDFs through hadronic scattering experiments. These means of obtaining knowledge are bi-directional: we can also use the known PDFs to predict observables of other hadronic scattering processes. It should be noted that the factorization scale μ and the renormalization scale $\bar{\mu}$ can be arbitrarily selected, but in order for the high-order terms in the perturbation expansion to be ignorable, we should select specific energy scales μ and $\bar{\mu}$. So μ and $\bar{\mu}$ are usually set equal to the scale of the hard scattering process (e.g. the scale is Q for DIS processes).[2]

1.3.3. Factorization

Based on this factorization assumption of PDFs, the probabilities of partons inside hadrons in Eq. 1.2 are independent of the hard processes $\hat{\sigma}$. The factorization assumption indicates that the partons in hadrons are free particles.

Under this assumption, in a hadronic interaction, an external particle randomly collides with one of the free partons inside the hadron. This assumption is reasonable at a high energy scale because the asymptotic freedom guarantees the small coupling between partons. Another way to see this is to estimate the reaction time between the quarks and gluons in the hadrons as $t \sim 1/\Lambda$ for the mass scale of a proton $\Lambda \sim 1$ GeV. If the energy scale of hard scattering Q in a quark or gluon collision process is much larger than Λ ($Q \gg \Lambda$), the reaction time $1/Q$ will be much less than t, which will allow the reaction to be completed before the quarks and gluons feel the force between each other. Therefore, at a collision energy far greater than Λ, our assumption is reasonable.

[2] All orders beyond LO of $\hat{\sigma}^a(\mu)$ includes $\ln\frac{Q^2}{\mu^2}$ and $\ln\frac{Q^2}{\bar{\mu}^2}$, so $\mu = \bar{\mu} = Q$ can eliminate $\ln\frac{Q^2}{\mu^2}$ and $\ln\frac{Q^2}{\bar{\mu}^2}$.

1.3.4. μ-dependence of PDFs

However, the μ-dependence of $f_a(\xi, \mu)$ in the convolution equation shows that during collision processes, the parton distributions seen by external particles will also be related to the collision energy. For example, we can observe that $f_a(\xi, \mu)$ change by the energy scale of hard scattering Q in DIS processes at HERA [4]. How do we understand the μ-dependence in PDFs? As we mentioned earlier, the convolution equation indicates that we can divide the hadronic scattering processes into perturbative parts $\hat{\sigma}$ at high-energy scales and non-perturbative parts (PDFs) at low energy scales. However, choosing which factorization scale μ to separate high energy and low energy parts is arbitrary above Λ. This means that the chosen factorization scale μ will change the (energy) range covered by hadronic physics, which in turn affects PDFs. The derivative of the convolution equation with respect to μ makes us know more clearly how PDFs change with μ. Because μ is unphysical, hadronic cross sections should not depend on μ ($\frac{d\sigma}{d\mu} = 0$). Therefore, according to the chain rule, we get

$$\frac{d\sigma}{d\mu} = \frac{df}{d\mu} \otimes \hat{\sigma} + f \otimes \frac{d\hat{\sigma}}{d\mu} = 0, \tag{1.6}$$

which suggests the change of hard scattering contribution by μ must be balanced by the change of PDFs. This equation will derive the DGLAP equations in the next section.

1.3.5. DGLAP Equations

The insight in Eq. 1.6 is encoded with the DGLAP equations [5–7] to describe the rate of change of PDFs withthe factorization scale μ. Using Mellin transformation, we can transform the convolution of two functions $f \otimes g$ into the product of another two functions $\tilde{f}\tilde{g}$. Thus, we can separate $\frac{df}{d\ln\mu^2}$ from Eq. 1.6 with the following steps:

$$\frac{d\sigma}{d\ln\mu^2} = \frac{d\tilde{f}}{d\ln\mu^2}\tilde{\hat{\sigma}} + \tilde{f}\frac{d\tilde{\hat{\sigma}}}{d\ln\mu^2} = 0, \tag{1.7}$$

$$\frac{d\tilde{f}}{d\ln\mu^2} = -\tilde{f}\frac{1}{\tilde{\hat{\sigma}}}\frac{d\tilde{\hat{\sigma}}}{d\ln\mu^2} = -\tilde{f}\gamma. \tag{1.8}$$

Then we do the inverse Mellin transformation and get the DGLAP equations

$$\frac{df}{d\ln\mu^2} = -f\otimes\gamma. \tag{1.9}$$

The DGLAP equations are written as

$$\frac{df_a(\xi,\mu^2)}{d\ln\mu^2} = \sum_b [P_{ab}\otimes f_b(\xi,\mu^2)] \tag{1.10}$$

or

$$\frac{df_a(\xi,\mu^2)}{d\ln\mu^2} = \sum_b \int_\xi^1 \frac{\mathrm{d}z}{z}\, P_{ab}(z) f_b(\xi/z,\mu^2). \tag{1.11}$$

The splitting functions P_{ab} describe the probability that a parton "b" with a momentum fraction x radiates a soft or collinear parton "a". The interpretation of this formula is that the simplest case for a parton changing its μ before the hard scattering is splitting to two partons ($q\to qg$, $g\to q\bar{q}$, and $g\to gg$). The probability of this splitting, according to the Feynman rules, is proportional to the order one[3] of the strong coupling constant $\alpha_s(\bar{\mu}^2)$, and the parton "a" carries the momentum fraction z of the parton "b". Since the partons' momentum fractions will become smaller during the splitting processes, it is expected that there will be more small-x partons in nucleons at high-μ scales. The Fig. 1.3 (the Fig. 3 in [8]) is a comparison of two different scales of PDFs. We found that small-x PDFs became very large at high μ^2, which came from a large amount of gluon radiation. The DGLAP equations tell us that the evolution of PDFs in any $\mu > Q_0$ region is related to the convolution of PDFs and the splitting functions, where we only extract the PDFs above an initial scale $Q_0 \sim 1$ GeV as we want to make sure that the perturbative assumption works well.

[3] The lowest order splitting functions are proportional to $\alpha_s(\mu^2)$. When considering higher-order effects, there will be high-order terms of $\alpha_s(\mu^2)$ in splitting functions. For instance, if the split parton splits again, its probability will have a $\alpha_s^2(\mu^2)$ term.

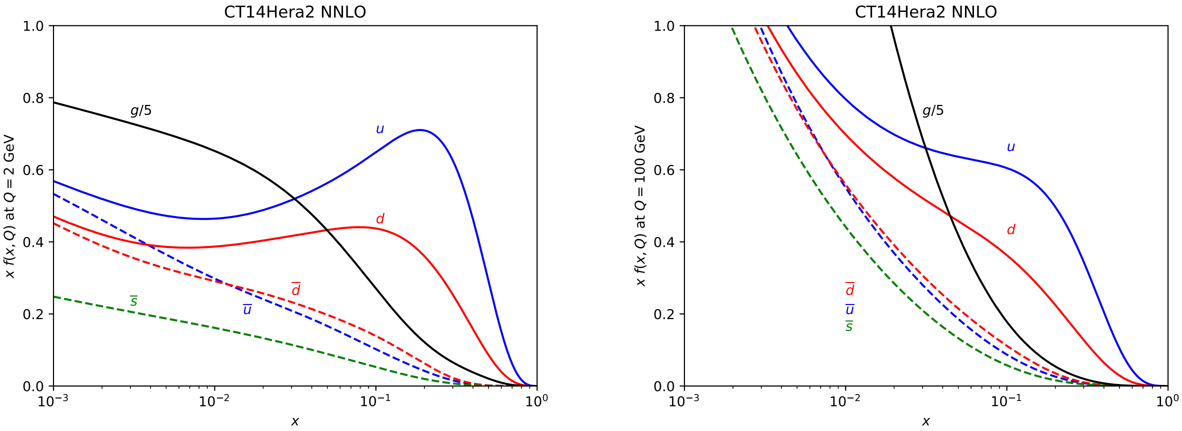


Figure 1.3: CT14HERA2 NNLO parton distribution functions [9] at 2 GeV (the left panel) and 100 GeV (the right panel).

1.4. Global QCD Analysis

1.4.1. Overview

Numerically computing analytical solutions to PDFs via first principles is still impossible. The most reliable way to get knowledge of PDFs is a global QCD analysis [10,11] that applies the global fit method to extract PDFs from experimental data. The global fit is often applied in the field of high energy physics. Given a physical model with many parameters, this method estimates the best parameters that describe experimental data. In addition, this method also allows us to evaluate which models fit the data best. In the case of the global QCD analysis, we hope to explore PDFs of various flavors and $\{\xi, \mu\}$ ranges. Therefore, the data we choose must be as sensitive as possible to PDFs in a broad range of $\{\xi, \mu\}$.

Currently, the data used in global QCD analyses include the following types: Deep Inelastic processes (DIS) ($h\ell \to \ell' X$), Drell-Yan processes ($h_1 h_2 \to W/Z/\gamma^* + X$), inclusive jet processes ($h_1 h_2 \to jets + X$), and $t\bar{t}$ processes ($h_1 h_2 \to t\bar{t}X$). Among them, DIS directly measures the structure functions of the nucleon. Therefore, its results are related to various PDFs of flavors. The most common channels in Drell-Yan processes are W and Z production processes. These W and Z bosons have a high probability of being produced by u and d, so Drell-Yan processes have a great influence on u and d PDFs. The asymmetry of W_+ and W_- processes is also sensitive to u/d ratio of PDFs. The main channels of jet processes are gluon interactions, so jet processes are very sensitive to the gluon PDF.

1.4.2. Parametrization Functions

In the global QCD analysis, we describe PDFs with parametrized functions :

$$f_a(\xi, Q_0) = c_0 x^{c_1} (1-\xi)^{c_2} F(c_3, c_4, ...), \tag{1.12}$$

where "a" is the flavor of the parton and c_n are the parameters of the functions. The ξ^{c_1} term that controls the shape of the PDFs as $\xi \to 0$ is guided by Regge theory. The $(1-\xi)^{c_2}$ term that controls the shape of the PDFs as $\xi \to 1$ is guided by quark counting rules. Each PDF set has its own $F(c_3, c_4, ...)$, which are smooth functions. $F(c_3, c_4, ...)$ must be flexible enough to describe the underlying structure of the data and avoid using too many parameters.

To extract PDF parameters, we use the convolution equation in Eq. 1.1, which relates the experimental results (i.e. hadronic cross sections) and PDFs. Because the hadronic cross sections and the hard scattering cross sections in Eq. 1.1 can be obtained from experiments and perturbative expansion, PDFs can therefore be fitted. Utilizing the DGLAP equations, we fit PDFs at Q_0 and evolve PDFs to the energy scales of hard scattering processes in the convolution equation.

1.4.3. The χ^2 Test in Global QCD Analyses

A global QCD analysis performs minimization of the χ^2 function, a measure of goodness of fit of theoretical predictions to experimental data, to determine the best-fit parameters

$\{c_n\}$ in parametrization functions. χ^2 summarizes the discrepancy between experimental data and the predictions of theoretical models. The residual (r_i) of a data point i with correlated systematic errors $\beta_{1i}, \beta_{2i}, ..., \beta_{Ki}$ could be defined as follows:

$$r_i = \frac{T_i - D_{i,sh}}{s_i}. \tag{1.13}$$

The residual r_i estimates the goodness of the fit of each point i. r_i is defined as the the deviation between the theoretical value T_i and the shifted experimental value $D_{i,sh}$ scaled to the uncorrelated experimental error s_i. The shifted experimental value

$$D_{i,sh} = D_i - \sum_{k}^{N_k} \bar{\lambda}_k \beta_{ki} \tag{1.14}$$

represents the data points D_i shifted by systematic errors that give rise to the best fit. And χ^2 can be written as

$$\chi^2 = \sum_E \chi_E^2 + \chi_{th}^2, \tag{1.15}$$

$$\chi_E^2 = \sum_{i}^{N_{pt}} r_i^2 + \sum_{k}^{N_k} \bar{\lambda}_k^2. \tag{1.16}$$

To measure the goodness of fit of all data points, we define χ^2 as the sum of χ_E^2 over all experimental data sets E plus χ_{th}^2 that imposes the theoretical constraints. χ_E^2 for an experimental data set E with N_{pt} data points is defined as the summation of all r_i^2 for all data points i plus the contribution of N_k parameters $\bar{\lambda}_k$ associated with the sources of correlated systematic errors. The complete formulas for χ_E^2 and χ_{th}^2 can be found in [12]. The detailed discussion of χ_E^2 and the analytical solutions of optimal parameters $\bar{\lambda}_k$ are in Appendix B of [13].

We vary the parameters $\{c_n\}$ to find the minimum χ^2, and the parameters corresponding to the minimum χ^2 represent the best-fit PDF set. There are many numerical methods for finding extreme values. For example, one approach that is often applied to χ^2 minimization is the gradient descent method. The gradient descent is a method that iteratively approaches the global minimum of a function by moving along the direction of the steepest descent of the value of χ^2. This direction is defined by the negative gradient at the point. The step size of each iteration can be continually adjusted to optimize the rate of convergence.

1.4.4. PDF Uncertainties

The estimates of the uncertainties of PDFs and PDFs-dependent physical observables are Hessian method [14] and Monte Carlo (MC) [15] Method.

Hessian Uncertainties The Hessian method uses the second-order Taylor expansion to explore the neighborhood of the χ^2 global minimum, and quantify the uncertainties of PDFs and PDFs-dependent physical observables based on this quadratic approximation. We assume that the χ^2 function near the minimum can be described by the quadratic expansion: $\chi^2(\vec{c}) = \chi^2(\vec{c}_0) + \sum\limits_{i,j} H_{i,j}\, \delta c_i\, \delta c_j$ for the displacement $\delta c_i = (\vec{c}_i - \vec{c}_{0,i})$ from the best-fit PDF parameters $\vec{c}_0$ to $\vec{c}$, where the Hessian matrix $H_{i,j}$ contains the contribution of the second-order derivative at $\vec{c}_0$. According to the approximation of the quadratic expansion, we will get an elliptical shape around the best-fit parameters $\vec{c}_0$ as the region of the tolerable uncertainties for a given tolerance parameter $\chi^2_{tolerance}$ satisfying $\chi^2(\vec{c}) < \chi^2(\vec{c}_0) + \chi^2_{torelance}$.

We diagonalize the Hessian matrix to get N sets of eigenvectors representing each principal axis of the ellipse, and take the PDFs on two points along the i-th principal axis of the ellipse as PDF error replicas in the ± directions of the i-th parameter dimension. That is to say, the χ^2 at two points we take on the i-th principal axis are $\chi^2(\vec{c}_0) + \chi^2_{torelance}$. Then the uncertainty on an PDF-dependent observable X can be estimated with these $2N$ error replicas. According to propagation of uncertainty, the uncertainty on X along the i-th direction can be approximated by $(X_i^+ - X_i^-)/2$ for X_i^+ and X_i^- depending on the ± error replicas corresponding to i. Therefore, under the assumption that uncertainties on various dimensions are uncorrelated to each other, the X uncertainty is described by the master equation: $\Delta X = \frac{1}{2}\sqrt{\sum_i^N (X_i^+ - X_i^-)^2}$ [13].

The ensemble, which estimates PDF uncertainties, also carries a lot of statistical information, enabling us to explore the relationship between PDF-dependent observables (e.g. one approach of the exploration is the Hessian correlation [16]). My work for quantifying the impact of data on PDF knowledge is based on the statistical analysis of Hessian ensembles (see chapter 2).

MC Uncertainties The Monte Carlo (MC) method uses statistical random sampling to perform statistical experiments to obtain numerical results. The Monte Carlo method can handle problems involving many random variables or problems that are difficult to solve analytically.

When we estimate PDF uncertainties with the Monte Carlo method, we can sample the data or sample the PDF parameters. For example, NNPDF Collaboration generates N sets of Monte Carlo replicas of the “artificial experimental data” based on the probability density in the space of data [17], then fit PDFs with these replicas. Then we extract the mean and statistical uncertainties in PDFs from N sets of best-fit PDF replicas. For an observable F that depends on PDFs, its mean is $\langle F \rangle = \frac{1}{N}\sum_i^N F^i$. The MC uncertainty in F is quantified by a standard deviation σ with $\sigma^2[F] = \frac{1}{N-1}\sum_i^N (F^i - \langle F \rangle)^2$.

1.4.5. Sum Rules

We determine the parametrizations of the PDFs at the initial scale Q_0 based on known physical constraints. One fundamental physical restriction is the momentum sum rule

$$\int_0^1 \xi \sum_a f_a(\xi)\, dx = 1. \tag{1.17}$$

The momentum sum rule represents the momentum conservation law of partons in hadrons. It states that the sum of the momentum of the partons inside a hadron should be the momentum of this hadron.

The quark number sum rules are other fundamental physical restrictions that represent quantum numbers of hadrons. For instance, the quark number sum rules of protons are

$$\int_0^1 [u(\xi) - \bar{u}(\xi)]\, dx = 2, \tag{1.18}$$

$$\int_0^1 [d(\xi) - \bar{d}(\xi)]\, dx = 1, \tag{1.19}$$

$$\int_0^1 [q(\xi) - \bar{q}(\xi)]\, dx = 0,\ q = s,\ c,\ b. \tag{1.20}$$

The above rules represent that each proton consists of two u quarks, one d quark, and zero s, c, and b quarks (i.e. uud). In addition, we often use other restrictions. For example, the global QCD analyses of the CTEQ group also include the following restrictions: $s = \overline{s}$ (see Appendix in [9]) and no intrinsic charm and bottom quarks.[4]

1.5. Lattice QCD

Lattice QCD is a tool for studying non-perturbative QCD based on first principles. It has been successfully applied to study some hadronic properties [3]. In the study of the nucleon

[4] The assumption means that charm and bottom are not in nucleons at the initial scale Q_0 and therefore charm and bottom are all produced by splitting processes. Whether charm quarks exist in nucleons at Q_0 is still an issue. [18] reviews the recent progress on this topic.

structure, it is complementary to the global QCD analysis. The main idea of Lattice QCD is to formulate QCD on a discrete Euclidean space-time grid. This discrete QCD theory approaches the continuum QCD when the lattice comes to the continuum limit. It allows us to numerically simulate non-perturbative QCD with first principles. There are some means to extract PDF-related quantities and information with lattice QCD techniques [19].

PDF Mellin Moments is one of the quantities that can be extracted from lattice QCD techniques. They are similar to multipole moments in classical electromagnetism: the distribution of charge can be expanded into a series of multiple moments, where the first few terms (monopole, dipole) retain the most important information about the charge distribution. Theoretically, as long as we know enough moments, we can reconstruct the PDFs. Although the current Lattice QCD techniques can only compute the first few Mellin moments [20–22], they can still provide useful information for constraining PDFs.

On the other hand, parton *quasi-distribution* functions (qPDFs) [23] is another new approach that has attracted attention in recent years. This approach can extract ξ-dependent PDFs. While the global QCD analysis is still the most mature method for the determination of PDFs, the methods provided by Lattice QCD are expected to contribute more to PDF knowledge in the future [19].

Therefore, I discuss my work on guiding the synergy between Lattice QCD and PDF phenomenologist communities to improve the PDF knowledge in chapter 3.

1.6. Future High-Energy Experimental Facilities

At present, the kinematic parameter space in PDFs constrained by hadron experiments is approximately between $\xi \sim [10^{-5}, 0.5]$ and $\mu \sim [1, 5000]$ GeV. These hadronic experiments include DIS processes, Drell-Yan processes, inclusive jet processes, and $t\bar{t}$ processes. We have mentioned these four processes in the section: global QCD analysis 1.4.

Since PDFs are used to make many predictions for LHC cross sections [24], we need to reduce the PDF uncertainties and expand the $\{\xi, \mu\}$ range of the exploration of PDFs.

In addition to the current LHC data, some LHC upgrade programs, such as the High-Luminosity LHC (HL-LHC) [25] and the Large Hadron Electron Collider (LHeC) [26], have the potential to further constrain PDFs. The HL-LHC is a program to upgrade the LHC to produce higher luminosity. The proposal of the LHeC is to install an electron accelerator in the LHC. This project allows us to obtain electron-proton and electron-ion collisions at higher energy scales compared with the present DIS datasets (far higher than the energy scale of the current highest-energy DIS datasets from HERA, the electron (positron)-proton collider at DESY, Hamburg). In addition, there are some electron-ion collider (EIC) projects [27] that have the potential to constrain PDFs. Finally, the higher collision energies and luminosity of Future Circular Collider (FCC) [28] can explore PDF knowledge that complements the previously mentioned facilities. Therefore, I will talk about my survey about the impact of future experimental plans to PDF knowledge in chapter 4. Below I will overview these accelerators.

The Large Hadron Collider (LHC) is by far the world's highest energy accelerator. It mainly performs proton-proton collisions. The (center of mass) energy of a collision event measures the highest mass of particles the event can produce. Therefore, increasing the center of mass energies of collision events is often critical to the search for new physics. After the upgrade during 2013-2015, the LHC can reach 6.5 TeV per beam of protons. The relationship between the beam energies E_1 and E_2 and the center of mass energy is $\sqrt{s} = 2\sqrt{E_1 E_2}$. Thus, its center of mass energy is 13 TeV. By 2017, its luminosity has reached $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{sec}^{-1}$ and its integrated luminosity[5] has reached $40\,\mathrm{fb}^{-1}$ (or $4 \times 10^{40}\,\mathrm{cm}^{-2}$) [29]. Where the luminosity is the number of collisions occurring per unit time and unit cross-section, which measures

[5] The conversions between the SI unit and the "Barn" unit used by particle physicists are as following: $1\,pb^{-1} = 10^{36}\,\mathrm{cm}^{-2}$; $1\,fb^{-1} = 10^{39}\,\mathrm{cm}^{-2}$; $1\,ab^{-1} = 10^{42}\,\mathrm{cm}^{-2}$

the event rate of an accelerator. The integrated luminosity, another useful concept, measures the number of collisions that an accelerator has accumulated over time.

The High Luminosity Large Hadron Collider (HL-LHC) is an upgrade to the LHC with the strategy of leveraging the potential of the LHC, the world's largest accelerator. It is expected to begin operations in 2026 with the goal of increasing the integrated luminosity of LHC (expected to be $0.3\,\mathrm{fb}^{-1}$ or $3 \times 10^{41}\,\mathrm{cm}^{-2}$) by a factor of 10 ($3\,\mathrm{ab}^{-1}$ or $3 \times 10^{42}\,\mathrm{cm}^{-2}$) during the operational period (to the late 2030s) [28] to collect more data and increase the potential for discovery. Its center of mass energy is 14 TeV, similar to the LHC.

The Large Hadron electron Collider (LHeC) is an upgrade program that adds an electron accelerator to the LHC, which makes full use of the LHC for electron-proton and electron-ion collisions at high energies. With $10 \sim 60$ GeV electrons and 7 TeV protons, the center of mass energy of the LHeC is range from $\sqrt{s} = 0.2 \sim 1.3$ TeV [30], which is much higher than the center of mass energy of HERA ($\sqrt{s} = 0.32$ TeV [31]) that has finished running in 2007. The expected luminosity of the LHeC is $10^{34}\,\mathrm{cm}^{-2}\mathrm{sec}^{-1}$ and the expected integrated luminosity of the LHeC is $1\,\mathrm{ab}^{-1}$ for HL-LHC [30], which is roughly 1000 times higher than the integrated luminosity of HERA ($\sim 700\,\mathrm{pb}^{-1}$ [31]).

The Electron-Ion Collider (EIC) is a type of accelerators that collides electrons with protons and ions. Its main purpose is to observe the structure of protons and neutrons. There are currently several such proposals worldwide. In addition to LHeC in Europe, the Brookhaven National Laboratory and Thomas Jefferson National Accelerator Facility in the United States also have proposals. The two programs are called JLEIC and eRHIC. For the baseline design of the JLEIC, the luminosity is 0.5 to $1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{sec}^{-1}$ and the center of mass energy is $\sqrt{s} = 15 \sim 65$ GeV with $3 \sim 10$ GeV electrons and $20 \sim 100$ GeV protons [32]. The peak luminosity of eRHIC is $10^{34}\,\mathrm{cm}^{-2}\mathrm{sec}^{-1}$ and the center of mass energy of eRHIC is $\sqrt{s} = 20 \sim 140$ GeV with 275 GeV protons and $5 \sim 18$ GeV electrons [33].

The Future Circular Collider (FCC) project aims to follow the exploration of the LHC. The goal is to greatly enhance the collision energy and luminosity of the survey to help us answer some observations that the SM cannot explain. Such as the matter-antimatter asymmetry problem and the source of dark matter. The FCC study emphasizes proton/proton (hadron) and electron/positron colliders, as well as the hadron/lepton scenario. The ultimate goal of the FCC is to reach collision energies of 100 TeV and an integrated luminosity of $20\,\mathrm{fb}^{-1}$ (or $2 \times 10^{40}\,\mathrm{cm}^{-2}$) for hh collisions (FCC-hh) in 25 years of operation [28]. For electron/positron collisions (FCC-ee), the setup of luminosities and center of mass energies of four modes in a 14-years life-circle are different. The four modes are the Z-pole ($20\,\mathrm{ab}^{-1}$ and 88, 91, 94 GeV), the W_+/W_- threshold ($10\,\mathrm{ab}^{-1}$ and ~ 161 GeV), the ZH maximum ($5\,\mathrm{ab}^{-1}$ and ~ 240 GeV), and the $t\bar{t}$ threshold ($1.5\,\mathrm{ab}^{-1}$ and $340 \sim 350$ GeV, and the remainder at 365 GeV).

1.7. Overview of My Work

My thesis focuses on techniques for improving our PDF knowledge, and assesses the future development of PDF phenomenology with these techniques. I will discuss three topics that have been published.

1.7.1. PDFSense

In chapter 2, I discuss a method to statistically quantify and map the sensitivities to PDFs imposed by hadronic experimental datasets. The approaches developed in this research can help us to preliminarily evaluate the impact of data on PDF knowledge.

Determinations of the proton's collinear parton distribution functions (PDFs) are emerging with growing precision due to increased experimental activity at facilities like the Large Hadron Collider. While this copious information is valuable, the speed at which it is released makes it difficult to quickly assess its impact on the PDFs, short of performing computationally expensive global fits. As an alternative, we explore new methods for quantifying the potential impact of experimental data on the extraction of proton PDFs. Our approach relies crucially on the Hessian correlation between theory-data residuals and the PDFs themselves, as well as on a newly defined quantity --- the *sensitivity* --- which represents an extension of the correlation and reflects both PDF-driven and experimental uncertainties. This approach is realized in a new, publicly available analysis package PDFSense, which operates with these statistical measures to identify particularly sensitive experiments, weigh their relative or potential impact on PDFs, and visualize their detailed distributions in a space of the parton momentum fraction x and factorization scale μ. This tool offers a new means of understanding the influence of individual measurements in existing fits, as well as a predictive device for directing future fits toward the highest impact data and assumptions. Along the way, many new physics insights can be gained or reinforced. As one of many examples, PDFSense is employed to rank the projected impact of new LHC measurements in jet, vector boson, and $t\bar{t}$ production and leads us to the conclusion that inclusive jet production at the LHC has a potential for playing an indispensable role in future PDF fits. These conclusions are independently verified by preliminarily fitting this experimental information and investigating the constraints they supply using the Lagrange multiplier technique.

1.7.2. Synergy Between Lattice QCD and Phenomenological PDFs

In chapter 3, I present the analysis of the coming synergy between lattice QCD (a field that solves non-perturbative phenomena in QCD by computational approaches) and high-energy PDF phenomenology. In this work, I also used some techniques developed in PDFSense.

Building upon the `PDFSense` framework developed in Ref. [34], we perform a comprehensive analysis of the sensitivity of present and future high-energy data to a number of "standard candle" quantities typically evaluated in lattice gauge theory, with a particular focus on the integrated Mellin moments of nucleon parton distribution functions (PDFs), such as $\langle x \rangle_{u^+-d^+}$ and $\langle x \rangle_g$, as well as x-dependent quark quasi-distributions --- in particular, that of the isovector combination. Our results demonstrate the potential for lattice calculations and phenomenological quark distributions informed by high-energy data to co-operatively improve the picture of the nucleon's collinear structure. This will increasingly be the case as computational resources for lattice calculations further exponentiate and QCD global analyses continue to grow in sophistication. Our sensitivity analysis suggests that a future lepton-hadron collider would be especially instrumental in providing phenomenological constraints to lattice observables.

1.7.3. Assessing the Potential of Future Experimental Facilities

In chapter 4, I explore the potential impact of possible future experimental facilities on the knowledge of PDFs with approaches developed in PDFSense.

Particle and nuclear physics are moving toward a new generation of experiments to stress-test the Standard Model (SM), search for novel degrees of freedom, and comprehensively map the internal structure of hadrons. Due to the complex nature of QCD and wide array of past, present, and possible future experiments, measurements taken at these next-generation facilities will inhabit an expansive space of high-energy data. Maximizing the impact of each future collider program will depend on identifying its place within this sprawling landscape. As an initial exploration, we use the recently-developed `PDFSense` framework to assess the PDF sensitivity of two future high-energy facilities --- the high-luminosity upgrade to the LHC (HL-LHC) and the Large Hadron-electron Collider (LHeC) proposal --- as well as the electron-ion collider (EIC) proposed to map the few-GeV quark-hadron transition region. We report that each of the experimental facilities considered occupies a unique place in the kinematical parameter space with specialized pulls on particular collinear quantities. As such, there is a clear complementarity among these programs, with an opportunity for each to mutually reinforce and inform the others.

In chapter 5, I present my results and conclusions.

Chapter 2

Mapping the Sensitivity of Hadronic Experiments to Nucleon Structure

2.1. Introduction

The determination of collinear parton distribution functions (PDFs) of the nucleon is becoming an increasingly precise discipline with the advent of high-luminosity experiments at both colliders and fixed-target facilities. Several research groups are involved in the rich research domain of the modern PDF analysis [9, 35–40]. By quantifying the distribution of a parent hadron's longitudinal momentum among its constituent quarks and gluons, PDFs offer both a description of the hadronic structure and an essential ingredient of perturbative QCD computations. PDFs enjoy a symbiotic relationship with high-energy experimental data, in the sense that they are crucial for understanding hadronic collisions in the Standard Model (SM) and beyond, while reciprocally benefiting from a wealth of high-energy data that constrain the PDFs. In fact, since the start of the Large Hadron Collider Run II (LHC Run II), the volume of experimental data pertinent to the PDFs is growing with such speed that keeping pace with the rapidly expanding datasets and isolating measurements of greatest impact presents a significant challenge for PDF fitters. This paper intends to meet this challenge by presenting a method for identifying high-value experiments which constrain the PDFs and the resulting SM predictions that depend on them.

That such expansive datasets can constrain the PDFs is a consequence of the latter's universality — a feature which relies upon QCD factorization theorems to separate the inherently nonperturbative PDFs (at long distances) from process-dependent, short-distance matrix elements. For instance, the cross section for inclusive single-particle hadroproduction (of, *e.g.*, a weak gauge boson W/Z) in proton-proton collisions at the LHC is directly sensitive to the nucleon PDFs via an expression of the form

$$\begin{aligned} \sigma(AB \to W/Z{+}X) \;&=\; \sum_n \alpha_s^n(\mu_R^2) \sum_{a,b} \int dx_a dx_b \\ &\times\; f_{a/A}(x_a, \mu^2)\, \hat{\sigma}^{(n)}_{ab\to W/Z+X}(\hat{s},\, \mu^2, \mu_R^2)\, f_{b/B}(x_b, \mu^2)\ , \end{aligned} \tag{2.1}$$

in which $f_{a/A}(x_a, \mu^2)$ represents the PDF for a parton of flavor f_a carrying a fraction x_a of the 4-momentum of proton p_A at a factorization scale μ; the n^{th}-order hard matrix element is denoted by $\hat{\sigma}^{(n)}_{ab\to W/Z+X}(\hat{s},\, \mu^2, \mu_R^2)$ and is dependent upon the partonic center-of-mass energy $\hat{s} = x_a x_b s$, in which s in the center-of-mass energy of the initial hadronic system; and μ_R is the renormalization scale in the QCD coupling strength $\alpha_s(\mu_R)$. In Eq. (2.1), subleading corrections $\sim\Lambda^2/M^4_{W/Z}$ have been omitted, and we emphasize that factorization theorems like Eq. (2.1) have been proved to arbitrary order in α_s for essential observables in the global PDF analysis, such as the inclusive cross sections in DIS and Drell-Yan processes. For compactness and generality, we shall refer henceforth to a PDF for the parton of flavor f simply as $f(x, \mu)$.

Given this formalism, one is confronted with the problem of finding those experiments that provide reliable new information about the PDF behavior. With the proliferation of potentially informative new data, incorporating them all into a global QCD fit inevitably incurs significant cost both in terms of computational resources and required fitting time. Indeed, tremendous progress in the precision of PDFs and robustness of SM predictions is driven by the technology for performing global analysis that has vastly grown in complexity and sophistication. Nowadays, the state-of-the-art in perturbative QCD (pQCD) treatments are done at NNLO (and increasingly even N^{3}LO), and advanced statistical techniques are commonly employed in PDF error estimation. The magnitude of this subject is vast, and we refer the interested reader to Refs. [11, 24] for comprehensive reviews. The tradeoff of this progress is that the impact of an experiment on the ultimate PDF uncertainty is often hard to foresee without doing a complicated fit. Various publications claim sensitivity of new experiments to the PDFs. In this paper, we look into these claims using statistical techniques that bypass doing the fits, and with an eye on theoretical, experimental, and methodological components relevant at the NNLO precision.

The potential cost is steepened by the large size of the global datasets usually involved. This point can be seen in Fig. 2.1, which plots the default dataset considered in the present analysis in a space of partonic momentum fraction x and factorization scale μ. We label these data as the "CTEQ-TEA set," given that it is an extension of the 3287 raw data points (given by the sum over N_{pt} in Tables B.1 and B.2) treated in the NNLO CT14HERA2 analysis of Ref. [41], now augmented by the inclusion of 734 raw data points (given by the sum over N_{pt} in Table B.3) from more recent LHC data. These raw measurements can ultimately be mapped to 5227 typical $\{x, \mu\}$ values in Fig. 2.1, such that each symbol corresponds to a data point from an experiment shown in the legend, at the approximate x and μ values characterizing the data point as described in Appendix A. The experiments are labeled by a short-hand name which includes the year of final publication (*e.g.*, "HERAI+II'15" — corresponding to the 2015 combined HERA Run I and Run II data), following the translation key also given in Tables B.1–B.3 of App. B.1. The experiments included in the CT14HERA2 analysis are listed in the left column and upper part of the right column of the legend, while the newer LHC data considered for the upcoming CTEQ-TEA analysis are the last 14 entries of the right column.

The growing complexity of PDF fitting stimulates development of less computationally involved approaches to estimate the impact of new experimental data on full global fits, such as Hessian profiling techniques [42] and Bayesian reweighting [43, 44] of PDFs. Although these approaches do simulate the expansion of a particular global fit by including theretofore absent dataset(s), they are also limited in that the interpretation of their outcomes is married to the specific PDF parametrization and definition of PDF errors. For example, conclusions

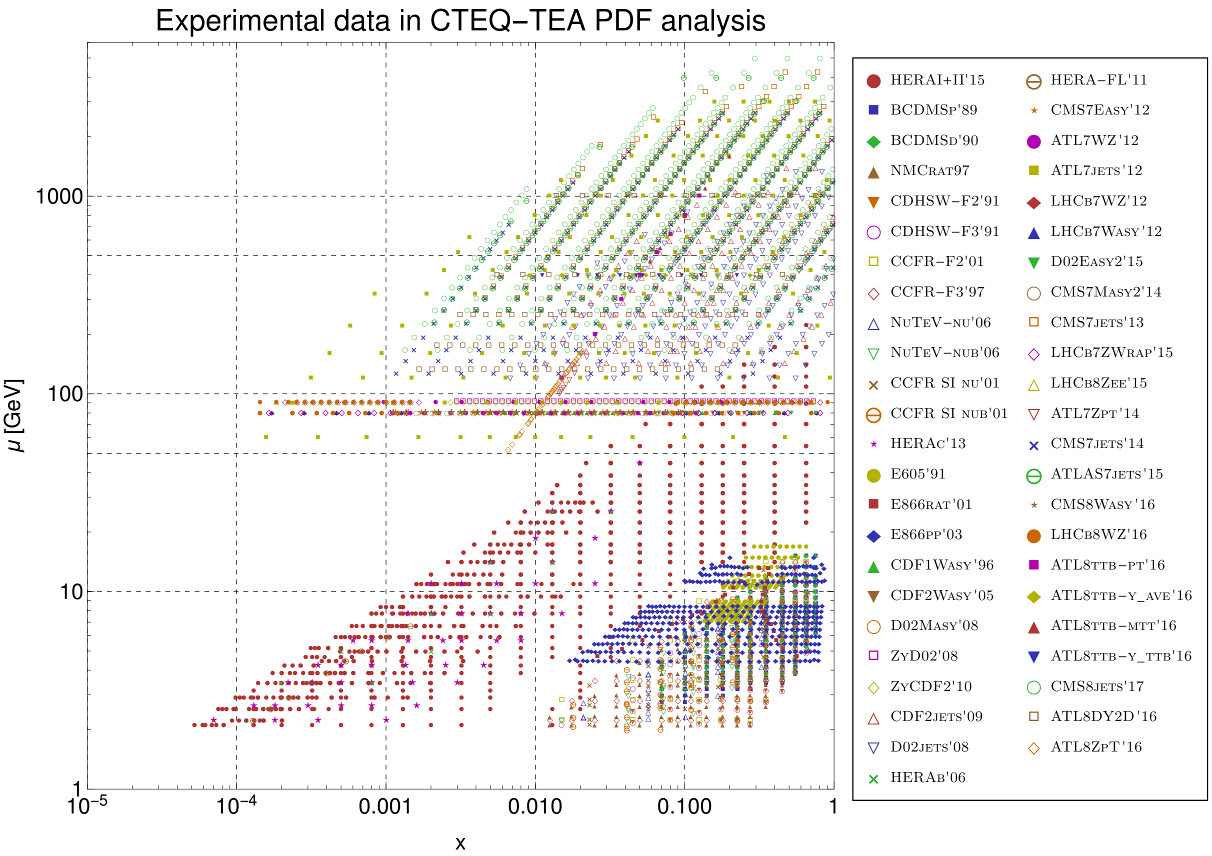


Figure 2.1: A graphical representation of the space of $\{x, \mu\}$ points probed by the full dataset treated in the present analysis, designated as "CTEQ-TEA". It represents an expansion to include newer LHC data of the CT14HERA2 dataset [41] fitted in the most recent CT14 framework [9], which involved measurements from Run II of HERA [39]. Details of the datasets corresponding to the short-hand names given in the legend may be found in Tables B.1–B.3.

obtained by PDF reweighting regarding the importance of a given data set strongly depend on the assumed statistical tolerance or the choice of reweighting factors [45, 46].

Parallel to these efforts, the notion of using correlations between the PDF uncertainties of two physical observables was proposed in Refs. [14, 47] as a means of quantifying the degree to which these quantities were related based upon their underlying PDFs. The PDF-mediated correlation C_f in this case, which we define in Sec. 2.3.1, embodies the Pearson correlation coefficient computed by a generalization of the "master formula" [13] for the Hessian PDF uncertainty. The Hessian correlation was deployed extensively in Ref. [16] to explore implications of the CTEQ6.6 PDFs for envisioned LHC observables. It proved to be instrumental for identifying the specific PDF flavors and x ranges most correlated with the PDF uncertainties for W, Z, H, and $t\bar{t}$ production cross sections as well as other processes. The Pearson correlation coefficient has also proven to be informative in the approach based on Monte-Carlo PDF replicas, see, e.g., Refs. [17, 48]. However, the PDF-mediated correlation with a theoretical cross section is only partly indicative of the sensitivity of the experiment. The constraining power of the experiment also depends on the size of experimental errors that were not normally considered in correlation studies, as well as on correlated systematic effects that are increasingly important.

As a remedy to these limitations, we introduce a new format for the output of CTEQ-TEA fits and a natural extension of the correlation technique to quantify the sensitivity of any given experimental data point to a PDF-dependent observable of the user's choice. In this approach, we work with *statistical residuals* quantifying the goodness-of-fit to individual data points. We demonstrate that the complete set of residuals computed for Hessian PDF sets characterizes the CTEQ-TEA fit well enough to permit a means of gauging the influence of empirical information on PDFs in a fashion that does not require complete refits.

A generalization of the PDF-mediated correlations called the *sensitivity* S_f — to be characterized in detail in Sec. 2.3.2 — better identifies those experimental data points that tightly constrain PDFs both by merit of their inherent precision and their ability to discriminate among PDF error fluctuations. Such an approach aids in identifying regions of $\{x, \mu\}$ for which PDFs are particularly constrained by physical observables.

In fact, in the numerical approach presented in the forthcoming sections, the user can quantify the sensitivity of data not only to individual PDF flavors, but even to specific physical observables, including the modifications due to correlated systematic uncertainties in every experiment of the CT14HERA2 analysis. For example, for Higgs boson production via gluon fusion ($gg \to H$) at the LHC 14 TeV, the short-distance cross sections are known up to N^{3}LO with a scale uncertainty of about 3% [49]. It has been suggested that $t\bar{t}$ production and high-p_T Z boson production on their own constrain the gluon PDF in the x region sensitive to the LHC Higgs production, and that these are comparable to the constraints from LHC and Tevatron data [50, 51]. Verifying the degree to which this hypothesis is true has been difficult without actually including all these data in a fit.

As an alternative to doing a full global fit, we can critically assess this supposition in the context of the entire global dataset of Fig. 2.1 using the Hessian correlations and sensitivities, $|C_f|$ and $|S_f|$. The detailed procedure is explained in Secs. 2.3.1 and 2.3.2. In the example at hand, we could rely on the established wisdom that the theoretical cross sections that have an especially large correlation with σ_{H^0} may constrain the PDF dependence of σ_{H^0}; say, when $|C_f| \gtrsim 0.7$ [16]. Along this reasoning, the left frame in Fig. 2.2 illustrates 310 experimental data points in $\{x, \mu\}$ space that have the highest absolute correlation, $|C_f|$, between the point's statistical residual defined in Sec. 2.3.1 and the cross section σ_{H^0} at 14 TeV via the CT14HERA2 NNLO PDFs. To locate such points in the figure, we highlighted them with color according to the convention shown on the color scale to the right. The respective $|C_f|$ for the highlighted data points ranges between 0.42 and 1. The rest of the data points have smaller correlations and are shown in gray.

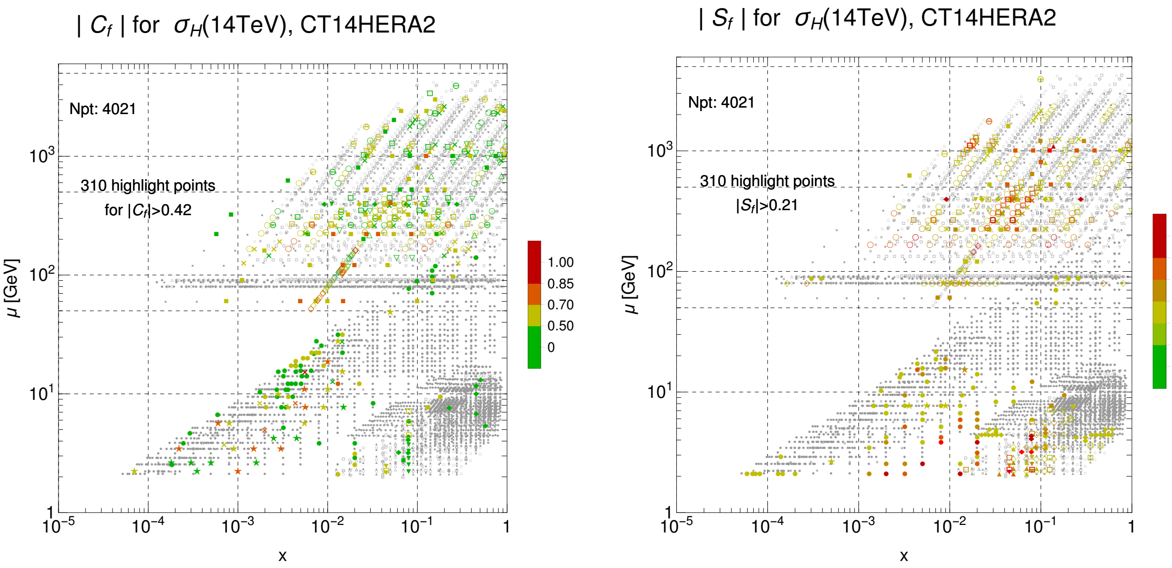


Figure 2.2: For the full CTEQ-TEA dataset of Fig. 2.1, we show the absolute correlation $|C_f|$ and sensitivity $|S_f|$ associated with the 14 TeV Higgs production cross section $\sigma_{H^0}(14\,\mathrm{TeV})$. 310 input data points with most significant magnitudes of $|C_f|$ and $|S_f|$ are highlighted with color. When only the $|C_f|$ plot is considered, only a very small subpopulation of jet production data (diagonal open circles and closed squares with $\mu \gtrsim 100$ GeV) exhibits significant correlations with $|C_f| > 0.7$ (orange and red colors), as well as some HERA DIS, high-p_T Z boson, and $t\bar{t}$ production data points. Our novel definition for the sensitivity in the right panel, on the other hand, reveals more points that have comparable potency for constraining the Higgs cross section. In this case, a larger fraction of the jet production points is important (especially CMS measurements of CMS8jets'17 and CMS7jets'14), as well as a number of other processes at smaller μ, particularly DIS data from HERA, BCDMS, NMC, CDHSW, and CCFR (experiments HERAI+II'15, BCDMSd'90, NMCrat'97, CDHSW-F2'91, CCFR-F2'01, CCFR-F3'97). Although its cumulative impact is comparatively modest, ATLAS $t\bar{t}$ production data (ATL8ttb-pt'16, ATL8ttb-y_ave'16, ATL8ttb-mtt'16, ATL8ttb-y_ttb'16) register significant per-point sensitivities, as do E866 pp Drell-Yan pair production (E866pp'03), LHCb W,Z production (LHCb7ZWrap'15, LHCb8WZ'16), and charge lepton asymmetries at D0 and CMS (D02Masy'08, CMS7Masy2'14, CMS7Easy'12). Similarly, some of the high-p_T Z production information (ATL7ZpT'14, ATL8ZpT'16) from ATLAS provide modest constraints.

We find that the 310 data points with the highest correlation for σ_{H^0} belong to 20 experiments. Nearly all of them are contributed by HERA Neutral Current (NC) DIS, LHC and Tevatron jet production, and HERA charm production. Some of the data points with highest $|C_f|$ come from high-p_T Z boson and even $t\bar{t}$ production.

The correlations C_f, however, do not reflect the experimental uncertainties, which vary widely across the experiments. In the left panel of Fig. 2.2, fewer than 30 points have a strong correlation of 0.7 or more; but more data points impose relevant constraints in the global fit. To include the information about the uncorrelated and correlated experimental errors, in the right panel of Fig. 2.2, we plot the distributions of 310 data points with the highest sensitivity parameter S_f, which more faithfully reproduces the actual constraints during the fitting. In general, we find substantial differences between the C_f and S_f distributions. Even the most significant correlations, of order $|C_f|\sim 0.7$ and above, do not guarantee a significant contribution of the experimental point to the log-likelihood χ^2 if the errors are large. On the other hand, we argue that $|S_f|$ is closely related to a contribution of the data point to χ^2. According to the distribution in the right figure, the 310 data points with the highest sensitivity $|S_f|$ to $\sigma_{H^0}(14\text{ TeV})$ arise from 27 experiments. Among these data points, only some have a large correlation $|C_f|$ with $\sigma_{H^0}(14\text{ TeV})$. Nonetheless, they have medium-to-large sensitivity, $|S_f|> 0.21$, according to the criterion developed in Sec. 2.3.2. We stress that, while one might suggests plausible dynamical reasons why certain experiments might be particularly sensitive to Higgs production via the gluon PDF, (*e.g.*, via the leading-order qg and gg hard cross sections in jet production and DGLAP scaling violations in inclusive DIS), this reasoning alone does not predict the actual sensitivity revealed by S_f in the presence of multiple experimental constraints.

As one noticeable difference from the $|C_f|$ figure, while inclusive DIS at HERA continues to contribute a large number of data points (about 80) with a high $|S_f|$, also the fixed-target DIS experiments (BCDMS, NMC, CDHSW, CCFR) contribute about the same number of sensitive points in the right panel that were not identified by large correlations. Other sensitive points belong to the jet production data sets from ATLAS and CMS and some vector boson production experiments (muon charge asymmetries at D0, CMS; E866 low-energy Drell-Yan production; LHCb 7 TeV W and Z cross sections).

On the other hand, HERA charm production, ATLAS 7/8 TeV high-p_T Z production, have suppressed sensitivities despite their large correlations, reflecting the larger experimental uncertainties in these measurements. While the LHC $t\bar{t}$ production experiments have large per-point sensitivities, they contribute relatively little to χ^2 because of their small total number of data points. From this comparison, one finds, perhaps somewhat unexpectedly, that fixed-target DIS experiments impose important constraints on $\sigma_{H^0}(14 \text{ TeV})$, thus complementing the HERA inclusive DIS data. One would conclude that efforts to constrain PDF-based SM predictions for Higgs production by relying only on a few points of $t\bar{t}$ data, but to the neglect of high-energy jet production points, would be significantly handicapped by the absence of the latter. We will return to this example in Sec. 2.4.

The discriminating power of a sensitivity-based analysis therefore forms the primary motivation for this work, and we present the attendant details below. To assess information about the PDFs encapsulated in the residuals for large collections of hadronic data implemented in the CTEQ-TEA global analysis, we make available a new statistical package PDFSense to map the regions of partonic momentum fractions x and QCD factorization scales μ where the experiments impose strong constraints on the PDFs. In companion studies, we have applied PDFSense to select new data sets for the next generation of the CTEQ-TEA global analysis, to quantitatively explore the physics potential for constraining the PDFs at a future Electron-Ion Collider (EIC) [27,52–54] and Large Hadron-Electron Collider (LHeC) [26], and to investigate the potential of high-energy data to inform lattice-calculable quantities [19] like the Mellin moments of structure functions [55] and quark quasi-distributions [23]. We reserve many instructive results for follow-up publications currently in preparation, while presenting select calculations in this article to demonstrate the power of the method. We find that the sensitivity technique generally agrees with the preliminary CTEQ-TEA fits and Hessian reweighting realized in the ePump program [56]. However, assessing the sensitivity is much simpler than doing the global fit. It does not require access to a fitting program or the application of (potentially subtle) PDF reweighting techniques.

The remainder of the article proceeds as follows. Pertinent aspects of the PDFs and their standard determination via QCD global analyses are summarized in Sec. 2.2. Then, we introduce *normalized residual variations* to extract, visualize, and quantify statistical information about the global QCD fit. In Sec. 2.3, we construct a number of statistical quantities that characterize the PDF constraints in the global analysis using the residual variations. In Sec. 2.4, we apply the thus constructed sensitivity parameter to examine the impact of various CTEQ-TEA datasets on extractions of the gluon PDF $g(x, \mu)$. In this section and in the conclusion contained in Sec.2.5, we emphasize a number of *physics insights* that we obtained by applying our sensitivity analysis techniques. Additional aspects of the technique and supplementary tables are reserved for Apps. A, B.1, and B.2.

2.2. PDF Preliminaries

2.2.1. Data Residuals in a Global QCD Analysis

This section is a further discussion of the content in Sec. 1.4.3. While various theoretical models exist for computing nucleon PDFs [57–59], unambiguous evaluation of the PDFs entirely in terms of QCD theory is not yet possible due to the fact that the PDFs can in general receive substantial nonperturbative contributions at infrared momenta. For this reason, precise PDF determination has proceeded mainly through the technique of the QCD global analysis — a method enabled by QCD factorization and PDF universality.

In this approach, a highly flexible parametric form is ascribed for the various flavors in a given analysis at a relatively low scale Q_0^2. For example, one might take the input PDF for a given quark flavor f to be a parametric form

$$f(x,\mu^2 = Q_0^2) = A_{f,0}\, x^{A_{f,1}} (1-x)^{A_{f,2}}\, F(x;\, A_{f,3}, \dots)\ , \tag{2.2}$$

in which $F(x;\, A_{f,3}, \dots)$ can be a suitable polynomial function, *e.g.*, a Chebyshev or Bernstein polynomial, or replaced with a feed-forward neural network $\mathrm{NN}_f(x)$ as in the NNPDF approach. While the full statistical theory for PDF determination and error quantification is beyond the intended range of this analysis, roughly speaking, a best fit is found for a vector $\vec{A}$ of N PDF parameters A_l by minimizing a goodness-of-fit function χ^2 describing agreement of the QCD data and physical observables computed in terms of the PDFs. Based on the behavior of χ^2 in the neighborhood of the global minimum, it is then possible to construct an ensemble of error PDFs to quantify uncertainties of PDFs at a predetermined probability level.

There are various ways to evaluate uncertainties on PDFs, *e.g.*, the Hessian [13, 14], the Monte Carlo [15, 60], and the Lagrange Multiplier approaches [61]. In this analysis our default PDF input set is CT14HERA2, which uses the Hessian method to estimate uncertainties and is therefore based on the quadratic assumption for $\chi^2(\vec{A})$ in the vicinity of the global minimum. In the Hessian method, an orthonormal basis of PDF parameters $\vec{a}$ is derived from the input PDF parameters $\vec{A}$ by the diagonalization of a Hessian matrix H, which encodes the second-order derivatives of χ^2 with respect to A_l. The eigenvector PDF combinations $\vec{a}_l^{\pm}$ are found for two extreme variations from the best-fit vector $\vec{a}_0$ along the direction of the l^{th} eigenvector of H allowed at a given probability level. The uncertainty on a QCD observable X can then be estimated with one of the available "master formulas" [13,47], the “symmetric” variety of which is

$$\Delta X = \frac{1}{2}\sqrt{\sum_{l=1}^{N}(X_l^+ - X_l^-)^2} \ . \tag{2.3}$$

In the CTEQ-TEA global analysis, the χ^2 function accounts for multiple sources of experimental uncertainties, as well as for some prior theoretical constraints on the a_l parameters. Consequently, the global χ^2 function takes the form

$$\chi^2_{global} = \sum_E \chi^2_E + \chi^2_{th} \ , \tag{2.4}$$

where the sum runs over all experimental datasets (E); and χ^2_{th} imposes theoretical constraints. The complete formulas for χ^2_E and χ^2_{th} can be found in Ref. [12]. For the purposes of this paper, we express χ^2_E for each experiment E in a compact form as a sum of squared *shifted residuals* $r_i^2(\vec{a})$, which are summed over N_{pt} individual data points i in this experiment, as well as the contributions of N_λ best-fit nuisance parameters $\overline{\lambda}_\alpha$ associated with correlated systematic errors:

$$\chi^2_E(\vec{a}) = \sum_{i=1}^{N_{pt}} r_i^2(\vec{a}) + \sum_{\alpha=1}^{N_\lambda} \overline{\lambda}_\alpha^2(\vec{a}) \ . \tag{2.5}$$

In turn, $r_i(\vec{a})$ for the i^{th} data point is constructed from the theoretical prediction $T_i(\vec{a})$ evaluated in terms of PDFs, total uncorrelated uncertainty s_i, and the shifted central data value $D_{i,sh}(\vec{a})$:

$$r_i(\vec{a}) = \frac{1}{s_i}\left(T_i(\vec{a}) - D_{i,sh}(\vec{a})\right) \ . \tag{2.6}$$

This representation arises in the Hessian formalism due to the presence of correlated systematic errors in many experimental datasets, which require χ^2_E to depend on nuisance parameters λ_α. This is in addition to the dependence of χ^2_E on the PDF parameters $\vec{a}$ and theoretical parameters such as $\alpha_s(M_Z)$ and particle masses. The λ_α parameters are optimized for each $\vec{a}$ according to the analytic solution derived in Appendix B of Ref. [13]. Optimization effectively shifts the central value D_i of the data point by an amount determined by the optimal nuisance parameters $\overline{\lambda}_\alpha(\vec{a})$ and the correlated systematic errors $\beta_{i\alpha}$:

$$D_i \to D_{i,sh}(\vec{a}) = D_i - \sum_{\alpha=1}^{N_\lambda} \beta_{i\alpha}\overline{\lambda}_\alpha(\vec{a}) \ . \tag{2.7}$$

It should be noted that the contribution of the squared best-fit nuisance parameters to χ^2_E in Eq. (2.5) is dominated in general by the first term involving the shifted residuals, which tends to be much larger — especially for more sizable datasets.

We point out also that some alternative representations for χ^2 include the correlated systematic errors via a covariance matrix $(\text{cov})_{ij}$, rather than the above mentioned CTEQ-preferred form that explicitly operates with λ_α. Various χ^2 definitions in use are reviewed in [62], as well as in [40]. Crucially, however, the representations based upon operating with λ_α and $(\text{cov})_{ij}$ are derivable from each other [12]. From an extension of the derivation in Ref. [13], we may relate the shifted residual to the covariance matrix at an i^{th} point and optimal nuisance parameters as

$$r_i(\vec{a}) \ = \ s_i \sum_{j=1}^{N_{pt}} (\text{cov}^{-1})_{ij} \ (T_j(\vec{a}) - D_j) \, , \tag{2.8}$$

$$\overline{\lambda}_\alpha(\vec{a}) = \sum_{i,j=1}^{N_{pt}} (\text{cov}^{-1})_{ij} \frac{\beta_{i\alpha}}{s_i} \frac{(T_j(\vec{a}) - D_j)}{s_j}, \tag{2.9}$$

where

$$(\text{cov}^{-1})_{ij} \ = \ \left[\frac{\delta_{ij}}{s_i^2} - \sum_{\alpha,\beta=1}^{N_\lambda} \frac{\beta_{i\alpha}}{s_i^2} A^{-1}_{\alpha\beta} \frac{\beta_{j\beta}}{s_j^2} \right] \ , \tag{2.10}$$

and

$$A_{\alpha\beta} \;=\; \delta_{\alpha\beta} + \sum_{k=1}^{N_{pt}} \frac{\beta_{k\alpha}\beta_{k\beta}}{s_k^2}\;. \tag{2.11}$$

Thus, even for those PDF analyses which operate with the covariance matrix one is still able to determine the shifted residuals r_i from $\left(\text{cov}^{-1}\right)_{ij}$ using Eq. (2.8). In this article, we conveniently follow the CTEQ methodology and obtain $r_i(\vec{a})$ directly from the CTEQ-TEA fitting program, together with the optimal nuisance parameters $\overline{\lambda}_\alpha(\vec{a})$ and shifted central data values $D_{i,sh}(\vec{a})$.

2.2.2. Visualization of the Global Fit with the Help of Residuals

The shifted residuals r_i draw our interest because, in consequence of the definitions in Eqs. (2.5)-(2.6), they contain substantial low-level information about the agreement of PDFs with every data point in the global QCD fit in the presence of systematic shifts. The response of $r_i(\vec{a})$ to the variations in PDFs depends on the experiment type and kinematic range associated with the i^{th} data point, and the totality of these responses can be examined with modern data-analytical methods. The sum of squared residuals over all points of the global dataset renders the bulk of the log-likelihood, or experimental, component χ^2_E of the global χ^2. In turn, the root-mean-squared residual $\langle r_0\rangle_E$ for experiment E and the central PDF set $\vec{a}_0$ is tied to $\chi^2_E(\vec{a}_0)/N_{pt}$, the standard measure of agreement with experiment E at the best fit:

$$\langle r_0\rangle_E \equiv \sqrt{\frac{1}{N_{pt}}\sum_{i=1}^{N_{pt}} r_i^2(\vec{a}_0)} = \sqrt{\frac{1}{N_{pt}}\left(\chi^2_E(\vec{a}_0) - \sum_{\alpha=1}^{N_\lambda}\overline{\lambda}_\alpha^2(\vec{a}_0)\right)} \approx \sqrt{\frac{\chi^2_E(\vec{a}_0)}{N_{pt}}}. \tag{2.12}$$

Notice that $\langle r_0\rangle_E \approx 1$ when the fit to the experimental data set E is good.

We will now invoke the Hessian formalism to first organize the analysis of the PDF dependence of individual residuals, and then introduce a framework to evaluate sensitivity of individual data points to PDF-dependent physical observables. To test the effectiveness of the proposed method, we study constraints using CT14HERA2 parton distributions [41] fitted to datasets from DIS processes, $Z \to l^+l^-$, $d\sigma/dy_l$, $W \to l\nu$, and jet production $(p_1 p_2 \to jjX)$. We include both the experiments that were used to construct the CT14HERA2 dataset, as well as a number of LHC experiments that may be fitted in the future. The experimental data sets are summarized in Tables B.1-B.3.

Given the urgency in improving constraints on the gluon PDF for investigations of the Higgs sector, we focus attention on several candidate experiments that may probe $g(x, \mu)$: high-p_T Z-boson production (ATL8ZpT'16, ATL7ZpT'14), $t\bar{t}$ production (ATL8ttb-pt'16, ATL8ttb-y_ave'16, ATL8ttb-mtt'16, ATL8ttb-y_ttb'16), as well as high-luminosity or alternative data sets for jet production, such as the high-luminosity ATLAS 7 TeV jet data (ATLAS7jets'15) that is to replace the counterpart low-luminosity set ATL7jets'12, or the CMS 7 TeV jet data set (CMS7jets'14) that extends to lower jet p_T and higher rapidity, $2.5 < |y_j| < 3$, than the previously fitted CMS 7 TeV jet data set (CMS7jets'13).[1] The dependence of such experiments on $g(x, \mu)$ is scrutinized in a number of ways. We examine their statistical properties using both the PDFs from the CT14HERA2 NNLO analysis, which already impose significant constraints on the large-x gluon using the Tevatron inclusive jet data sets, CDF2jets'09 and D02jets'08; and in some comparisons using a special version of the NNLO PDFs that are fitted to the same CT14HERA2 data set, except without including the above jet data sets. As yet another aspect, we investigate a range of measurements of Drell-Yan pair production cross sections and charge lepton asymmetries with the goal to understand their sensitivity predominantly to the (anti)quark sector.

To parametrize the response of a residual $\vec{r}_i$, we evaluate it for every eigenvector PDF $\vec{a}_l^{\pm}$ of the CT14HERA2 PDF set with $N = 28$ PDF parameters. Then, given the *normalized*

[1] As a result, a small number of data points that contributes to both the data sets CMS7jets'14 and CMS7jets'13 is double-counted in the histograms, without affecting the conclusions.

residual variations

$$\delta^{\pm}_{i,l} \equiv \left(r_i(\vec{a}^{\pm}_l) - r_i(\vec{a}_0)\right)/\langle r_0\rangle_E \tag{2.13}$$

between the residuals for the PDF eigenvectors $\vec{a}^{\pm}_l$ and for the CT14HERA2 central PDF $\vec{a}_0$, we construct a $2N$-dimensional vector

$$\vec{\delta}_i = \left\{\delta^{+}_{i,1},\, \delta^{-}_{i,1},\, ...,\delta^{+}_{i,N},\, \delta^{-}_{i,N}\right\} \tag{2.14}$$

for each data point of the global dataset.

The components of $\vec{\delta}_i$ parametrize responses of r_i to PDF variations along the independent directions given by $\vec{a}^{\pm}_l$. The differences are normalized to the central root-mean-square (r.m.s.) residual $\langle r_0\rangle_E$ of experiment E [see Eq. (2.12)] so that the normalized residual variations do not significantly depend on $\chi^2(\vec{a}_0)/N_{pt}$, the quality of fit to experiment E. Recall that a substantial spread over the fitted experiments is generally obtained for χ^2_E/N_{pt}. Moreover, it is reasonable to expect significantly larger values for χ^2_E/N_{pt} for the experiments that have not been yet fitted, but are included in the analysis of the residuals, *e.g.*, the new LHC experiments shown in Fig. 2.1. With the definitions in Eqs. (2.13) and (2.14), however, $\vec{\delta}_i$ is only weakly sensitive to χ^2_E/N_{pt}.

Thus, we represent the PDF-driven variations of the residuals of a global dataset by a bundle of vectors $\vec{\delta}_i$ in a $2N$-dimensional space.[2] This mapping opens the door to applying various data-analytical methods for classification of the data points and identifying the data points of the utmost utility for PDF fits. As the length of $\vec{\delta}_i$ is equal to the PDF-induced fractional error on r_i as compared to the average residual at the best fit, it can be argued that important PDF constraints arise from new data points that either have a large $|\vec{\delta}_i|$ or are otherwise distinct from the existing data points. Conversely, new data points with a

[2] In this section, we consider separate variations along $\vec{a}_l$ in the positive and negative directions. Alternatively, it is possible to work with a vector of N symmetric differences $\delta_{i,l} \equiv \left(r_i(\vec{a}^{+}_l) - r(\vec{a}^{-}_l)\right)/(2\langle r_0\rangle_E)$ and arrive at similar conclusions. Symmetric differences will be employed to construct correlations and sensitivities in Sec. 2.3.

small $|\vec{\delta}_i|$, or the ones that are embedded in the preexisting clusters of points, are not likely to improve constraints on the PDFs.

2.2.3. Manifold Learning and Dimensionality Reduction

2.2.3.1. PCA and t-SNE Visualizations

We illustrate a possible analysis technique carried out with the help of the TensorFlow Embedding Projector software for the visualization of high-dimensional data [63]. A table of 4021 vectors $\vec{\delta}_i$ for the CTEQ-TEA dataset (corresponding to our total number of raw data points) is generated by our package PDFSense and uploaded to the Embedding Projector website. As variations along many eigenvector directions result only in small changes to the PDFs, the 56-dimensional $\vec{\delta}_i$ vectors can in fact be projected onto an effective manifold spanned by fewer dimensions. Specifically, the Embedding Projector approximates the 56-dimensional manifold by a 10-dimensional manifold using principal component analysis (PCA). In practice, this 10-dimensional manifold is constructed out of the 10 components of greatest variance in the effective space, such that the most variable combinations of $\delta_{i,l}$ are retained, while the remaining 46 components needed to fully reconstruct the original 56-dimensional $\vec{\delta}_i$ are discarded. However, because the 10 PCA-selected components describe the bulk of the variance of $\delta_{i,l}$, the loss of these 46 components results in only a minimal relinquishment of information, and in fact provides a more efficient basis to study $\delta_{i,l}$ variations.

We encourage the reader to download the table of the normalized residual variations $\vec{\delta}_i$ for CT14HERA2 NNLO from the PDFSense website [64] and explore it for themselves using the Embedding Projector [63] or another program for multidimensional data visualization such as a tour [65]. These tools help to understand the detailed PDF dependence of individual data sets *without doing the global fit*. Performing such task has been challenging for nonexperts, if not for the PDF fitters themselves. With the proposed method, we can visually examine the PDF dependence of the residuals from the diverse data sets before quantitatively characterizing these distributions using the estimators developed in the next sections. In the future, a computer algorithm can be written to select the experimental data for PDF fits, based on the residual variations, and with minimal involvement from humans.

To offer an illustration, while grasping the full PDF dependence of the data points in the original 56-parameter space is daunting, in the 10-dimensional representation obtained via PCA, some directions result in efficient separation of the data points of different types according to their residual variations. The left panel of Fig. 2.3 shows one such 3-dimensional projection of $\vec{\delta}_i$ that separates clusters of residual variations arising from data for DIS, vector boson production, and jet/$t\bar{t}$ production. In this example, the jet/$t\bar{t}$ cluster, shown in red, is roughly orthogonal to the blue DIS cluster and intersects it. This separation is quite remarkable, as it is based only on numerical properties of the $\vec{\delta}_i$ vectors, and not on the meta-data about the types of experiments that is entered only after the PCA is completed; in other projections, the data types are not separated. The underlying reasons for this separation, namely, dependence on independent PDF combinations, will be quantified by the sensitivities in the next section.

As an alternative, the Embedding Projector can organize the $\vec{\delta}_i$ vectors into clusters according to their similarity using t-distributed stochastic neighbor embedding (t-SNE) [66]. A representative 3-dimensional distribution of the vectors obtained by t-SNE is displayed in the right panel of Fig. 2.3. In the figure, we show that the t-SNE method is able to identify and separate the clusters of data according to the experimental process (DIS, vector production, or jet production). In fact, the reader can perform the t-SNE analysis on the Embedding Projector website themselves and verify that it actually sorts the $\vec{\delta}_i$ vectors into the clusters according to their values of x and μ, and even the experiment itself. This exercise demonstrates, yet again, that the statistical residuals provided in PDFSense reflect the key properties of the global fit. Information can be extracted from them and examined in a number of ways.

The breakdown of the vectors over experiments in the PCA representation is illustrated by Fig. 2.4. Here, we see that the bulk of the DIS cluster from the left Fig. 2.3 originates with the combined HERA1+2 DIS data [HERAI+II'15]. The jet cluster in Fig. 2.3 will be dominated by ATLAS and CMS inclusive jet datasets [CMS7jets'14, ATLAS7jets'15, and CMS8jets'17], which add dramatically more points across a wider kinematical range on top of the CDF Run-2 and D0 Run-2 jet production datasets (CDF2jets'09) and (D02jets'08).

In contrast, although the $t\bar{t}$ production experiments (ATL8ttb-pt'16, ATL8ttb-y_ave'16, ATL8ttb-mtt'16, ATL8ttb-y_ttb'16) are generally characterized by large $\vec{\delta}_i$ vectors, they contribute only a few data points lying within the jet cluster of Fig. 2.4 and, by themselves, will not make much difference in a global fit. The same conclusion applies to data from high-p_T Z production, which has too few points to stand out in a fit with significant inclusive jet data samples. We return to this point in the discussion of reciprocated distances below.

It is also interesting to note that semi-inclusive charm production at HERA [HERAc'13] lies between, and partly overlaps with, the DIS and jet clusters. Finally, CCFR/NuTeV dimuon semi-inclusive DIS [SIDIS] (CCFR-F2'01, CCFR-F3'97, NuTeV-nu'06, NuTeV-nub'06) extends in an orthogonal direction, not well separated from the other datasets in the selected three-dimensional projection.

2.2.3.2. Reciprocated Distances

As a complement to the visualization methods based on PCA and t-SNE just presented, it is also possible to evaluate another similarity measure based on the distances between the vectors of the residual variations. For example, rather than applying the PCA to an ensemble of $\vec{\delta}_i$ vectors to perform dimensionality reduction, we might instead compute over

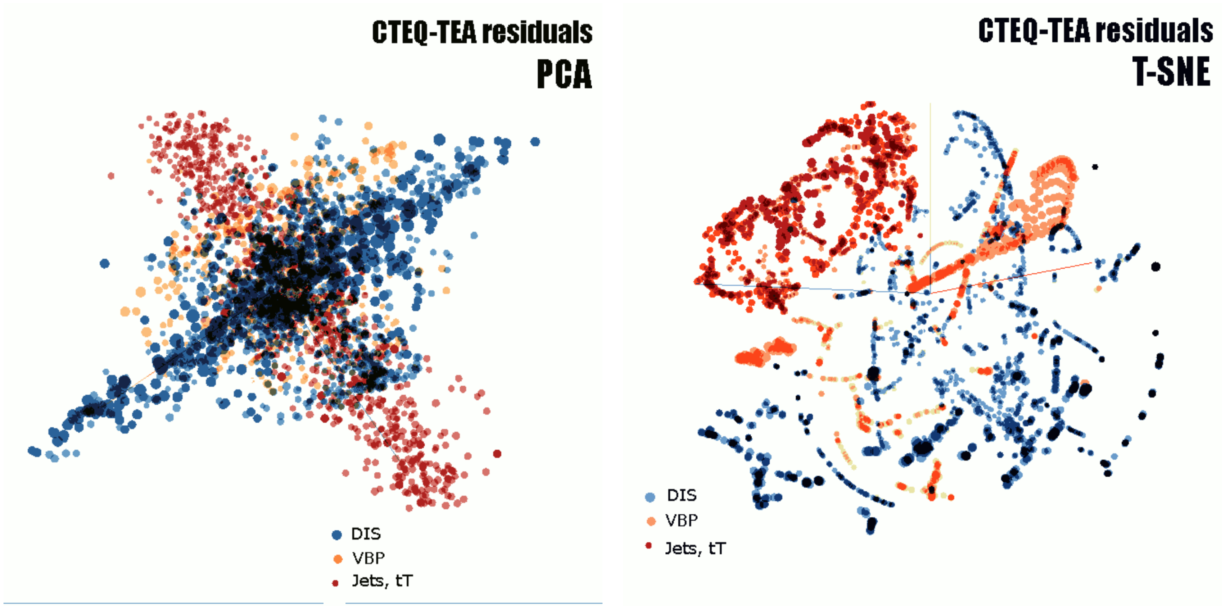


Figure 2.3: Distributions of residual variations $\vec{\delta}_i$ from the CTEQ-TEA analysis obtained by dimensionality reduction methods. Left: a 3-dimensional projection of a 10-dimensional manifold constructed by principal component analysis (PCA). Right: a distribution from the 3-dimensional t-SNE clustering method. Blue, orange, and red colors indicate data points from DIS, vector boson production, and jet/$t\bar{t}$ production processes.

the vector space a pair-wise *reciprocated distance* measure, which we define as

$$\mathcal{D}_i \equiv \left(\sum_{j \neq i}^{N_{all}} \frac{1}{|\vec{\delta}_j - \vec{\delta}_i|} \right)^{-1} , \tag{2.15}$$

and evaluate for the i points in each experimental dataset. We allow the sum over j in Eq. (2.15) to run over all the data points in the CTEQ-TEA set regardless of experiment (denoted by N_{all}). The distances can be computed either in the 56-dimensional space or in the reduced dimensionality space.[3] We plot the result of applying Eq. (2.15) to the 56-dimensional residual variations of the full CTEQ-TEA dataset computed using two PDF ensembles: CT14HERA2 fitted to all data in the left panel, and CT14HERA2 fitted only to the DIS and vector boson production data (excluding jet production data) in the right panel.

[3] Alternative definitions for the reciprocated distance can be also used, with qualitatively similar conclusions. For example, we could sum over all experimental data, but excluding those points belonging to the same experiment as point i, and normalizing $\mathcal{D}_i$ by $(N_{pt} - N_{all})/N_{pt}$ to compensate for different numbers of points in the experiment.

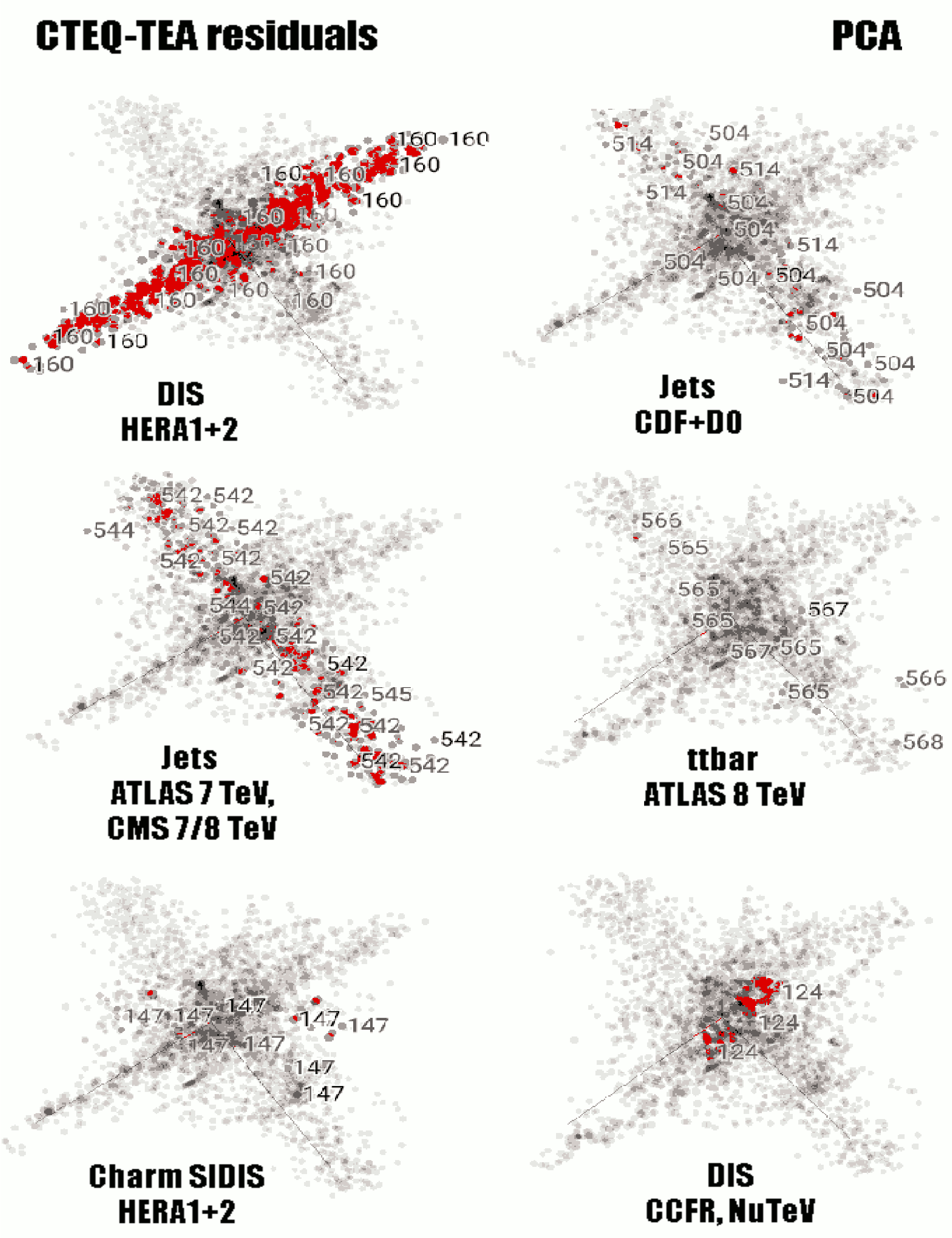


Figure 2.4: The PCA distribution from Fig. 2.3, indicating distributions of points from classes of experiments. In the numbering scheme used here, points labeled 1XX correspond to fixed-target measurements and 5XX to jet and $t\bar{t}$ production as given in Tables B.1–B.3. The specific experiments are noted in the plots.

Fig. 2.5 represents the distribution of the reciprocated distances over individual experiments of the CTEQ-TEA dataset. The CT Experiment ID # is shown on the abscissa, and the $\mathcal{D}_i$ values for every point of the experiment are indicated by the scatter points.

The advantage of the definition in Eq. (2.15) is that it enables a quantitative measure of the degree to which separate experiments broadly differ in terms of their residual variations, and therefore provides information analogous to that found in Figs. 2.3–2.4. For example, by inspection of Eq. (2.15) it can be seen that those experimental measurements which are widely separated from the rest of the CTEQ-TEA dataset in space of $\vec{\delta}_i$ vectors will correspond to comparatively large values of $\mathcal{D}_i$, and experiments that systematically differ from the rest of the total dataset are thus expected to have especially tall distributions in the panels of Fig. 2.5. On this basis, it can be seen that information yielded by W asymmetry measurements (D02Masy'08, CMS7Masy2'14, D02Easy2'15) are particularly distinct, as well as the combined HERA DIS data (HERAI+II'15) and fixed-target Drell-Yan measurements, such as E605 (E605'91) and E866 data (E866rat'01 and E866pp'03). Similarly, direct comparison of the $\mathcal{D}_i$ distributions in the panels of Fig. 2.5 allows one to compare constraints with and without the jet data. We note that the 7 and 8 TeV ATLAS high-p_T Z production (ATL7ZpT'14 and ATL8ZpT'16) and $t\bar{t}$ production (ATL8ttb-pt'16) provide a number of "remote" points and hence are potentially useful in the fits sensitive to the gluon. On the other hand, new jet production experiments (CMS7jets'14, ATLAS7jets'15, CMS8jets'17) all include large numbers of points characterized by significant reciprocated distances.

2.3. Quantifying Distributions of Residual Variations

We have demonstrated that the multi-dimensional distribution of the $\vec{\delta}_i$ vectors reflects the PDF dependence of individual data points. In this section, we will focus on numerical metrics to assess the emerging geometrical picture associated with the $\vec{\delta}_i$ distribution, and to visualize the regions of partonic momentum fractions x and QCD factorization scales μ where the experiments impose strong constraints on a given PDF-dependent observable X.

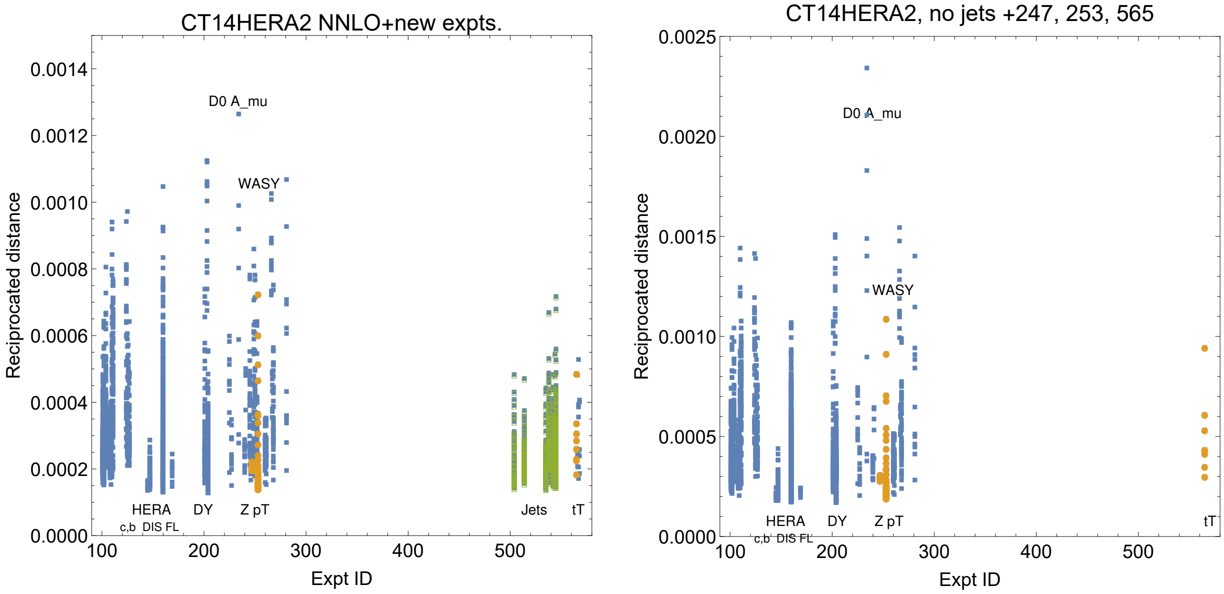


Figure 2.5: A plot of the reciprocated distances $\mathcal{D}_i$ obtained from the PDFs fitted to the full CT14HERA2 dataset [left] and to the CT14HERA2 dataset without jet production experiments [right]. The horizontal axis displays numerical experimental CT IDs of the constituent CTEQ-TEA datasets, for each of which is shown a column of values of the reciprocated distance. We highlight columns corresponding to Expt. IDs ATL7ZpT'14 [247], ATL8ZpT'16 [253], and ATL8ttb-pt'16 [565] as discussed in text.

Gradients of r_i in a space of Hessian eigenvector PDF parameters $\vec{a}$ are naturally related to the PDF uncertainty. Recall that in the Hessian method the PDF uncertainty on $X(\vec{a})$ is found as

$$\Delta X(\vec{a}) = X(\vec{a}) - X(\vec{a}_0) = \vec{\nabla} X|_{\vec{a}_0} \cdot \Delta\vec{a}, \tag{2.16}$$

where $\vec{a}_0$ is the best-fit combination of PDF parameters, and $\Delta\vec{a}$ is the maximal displacement along the gradient that is allowed within the tolerance hypersphere of radius T centered on the best fit [13, 14]. The standard master formula

$$\Delta X = \left|\vec{\nabla} X\right| = \frac{1}{2}\sqrt{\sum_{l=1}^{N}\left(X_l^{+} - X_i^{-}\right)^2} \tag{2.17}$$

is obtained by representing the components of $\vec{\nabla} X$ by a finite-difference formula

$$\frac{\partial X}{\partial a_i} = \frac{1}{2}(X_i^+ - X_i^-), \tag{2.18}$$

in terms of the values $X_l^{\pm}$ for extreme displacements of $\vec{a}$ within the tolerance hypersphere along the l-th direction.

In this setup, a dot product between the gradients provides a convenient measure of the degree of similarity between PDF dependence of two quantities [16]. A dot product $\vec{\nabla} r_i \cdot \vec{\nabla} f$ between the gradients of a shifted residual r_i and another QCD variable f, such as the PDF at some $\{x, \mu\}$ or a cross section, can be cast in a number of useful forms.

2.3.1. Correlation Cosine

The correlation for the i^{th} $\{x, \mu\}$ point, which we define following Refs. [11, 14, 16, 47] as

$$C_f \;\equiv\; \mathrm{Corr}[f, r_i] = \frac{\vec{\nabla} f \cdot \vec{\nabla} r_i}{\Delta f \, \Delta r_i}, \tag{2.19}$$

can determine whether there *may* exist a predictive relationship between f and goodness of fit to the i^{th} point. The correlation function $\mathrm{Corr}[X, Y]$ for the quantities X, Y in Eq. (2.19) represents the realization in the Hessian formalism of Pearson's correlation coefficient, which we express as

$$\mathrm{Corr}[X, Y] = \frac{1}{4 \Delta X \Delta Y} \sum_{j=1}^{N} (X_j^+ - X_j^-)(Y_j^+ - Y_j^-) \ , \tag{2.20}$$

with the sum in these expressions being over the j parameters of the full PDF model space. Geometrically, $\mathrm{Corr}[X, Y]$ represents the cosine of the angle that determines the eccentricity of an ellipse satisfying $\chi^2(\vec{a}) < \chi^2(\vec{a}_0) + T^2$ in the $\{X, Y\}$ plane. This latter point follows from the fact that the mapping of the tolerance hypersphere onto the $\{X, Y\}$ plane is an ellipse with an eccentricity that depends on the correlation of X and Y, which is given in turn by Eq. (2.20) above.

$\mathrm{Corr}[f, r_i]$ does not indicate how constraining the residual is, but it may indicate a predictive relation between r_i and f. On the basis of previous work [16], we say that the (anti-)correlation between X and Y is significant roughly if $|\mathrm{Corr}[X, Y]| \gtrsim 0.7$, while smaller (anti-)correlation values are less robust or predictive. Following this rule-of-thumb, correlations have been used successfully to identify PDF combinations that dominate PDF uncertainties of complicated observables, for instance to show that the gluon uncertainty dominates the total uncertainty on LHC W and Z production, or that the uncertainty on the ratio σ_W/σ_Z of $W^{\pm}$ and Z^0 boson cross sections at the LHC is dominated by the strangeness PDF, rather than u and d (anti-)quark PDFs [16].

2.3.2. Sensitivity in the Hessian Method

The correlation C_f alone does not fully encode the potential impact of separate or new measurements on improving PDF determinations in terms of the uncertainty reduction. Rather, we employ $\vec{\nabla} f \cdot \vec{\nabla} r_i$ again to define the *sensitivity* S_f to f of the i^{th} point in experiment E:

$$S_f \equiv \frac{\vec{\nabla} f \cdot \vec{\nabla} r_i}{\Delta f \, \langle r_0 \rangle_E} = \frac{\Delta r_i}{\langle r_0 \rangle_E} C_f \ , \tag{2.21}$$

where Δr_i and $\langle r_0 \rangle_E$ are computed according to Eqs. (2.3) and (2.12), respectively. In other words, Δr_i again represents the variation of the residuals across the set of Hessian error PDFs, and we normalize it to the r.m.s. residual for the whole dataset E to reduce the impact of random fluctuations in the data values $D_{i,sh}$. This definition has the benefit of encoding not only the correlated relationship of f with r_i, but also the comparative size of the experimental uncertainty with respect to the PDF uncertainty. In consequence, for example, if new experimental data have reported uncertainties that are much tighter than the present PDF errors, these data would then register as high-sensitivity points by the definition in Eq. (2.21).

Geometrically, S_f represents a projection onto the direction of the gradient $\vec{\nabla} f$ of the residual variation $\vec{\delta}_i$, defined in Sec. 2.3 using the symmetrized formula for $\delta_{i,l}$ noted in footnote 2, namely,

$$\delta_{i,l} \equiv \left(r_i(\vec{a}_l^{+}) - r(\vec{a}_l^{-}) \right) / (2\langle r_0 \rangle_E) \ . \tag{2.22}$$

Figure 2.6 shows a pictorial illustration of this interpretation. This interpretation suggests that the total strength of constraints along the direction of $\vec{\nabla} f$ can be quantified by summing projections S_f onto this direction of all individual vectors $\vec{\delta}_i$.

As with correlations, only a sufficiently large absolute magnitude of $|S_f|$ is indicative of a predictive constraint of the i^{th} point on f. Recall that r_i^2 is the contribution of the i^{th} point to χ^2, and that only residuals with a large enough Δr_i as compared to the r.m.s. residual $\langle r_0 \rangle_E$ are sensitive to PDF variations. The S_f magnitude is of order $\Delta r_i / \langle r_0 \rangle_E$, which suggests an estimate of a minimal value of S_f that would be deemed sensitive according to the respective χ^2 contribution. For the numerical comparisons in this study, we assume that $|S_f|$ must be no less than 0.25 to indicate a predictive constraint, as the PDF uncertainty of the i^{th} residual contributes no less than r_i^2 =0.0625 to the variation in the global χ^2. The reader can choose a different minimal value in the PDFSense figures depending on the desired accuracy. The cumulative sensitivities that we obtain in later sections are independent of this choice.

Yet another possible definition, which we list for completeness, is to further normalize the sensitivity as

$$S_f' \equiv \frac{\vec{\nabla} f \cdot \vec{\nabla} r_i}{f_0 \, \langle r_0 \rangle_E} = \frac{\Delta f}{f_0} \, S_f \ . \tag{2.23}$$

For instance, if f is the PDF $f(x_i, \mu_i)$ or parton luminosity evaluated at the $\{x_i, \mu_i\}$ points extracted according to the data, the definition of S_f' in Eq. (2.23) de-emphasizes those points where the PDF uncertainty $\Delta f(x_i, \mu_i)$ is small compared to the best-fit PDF value $f_0(x_i, \mu_i)$

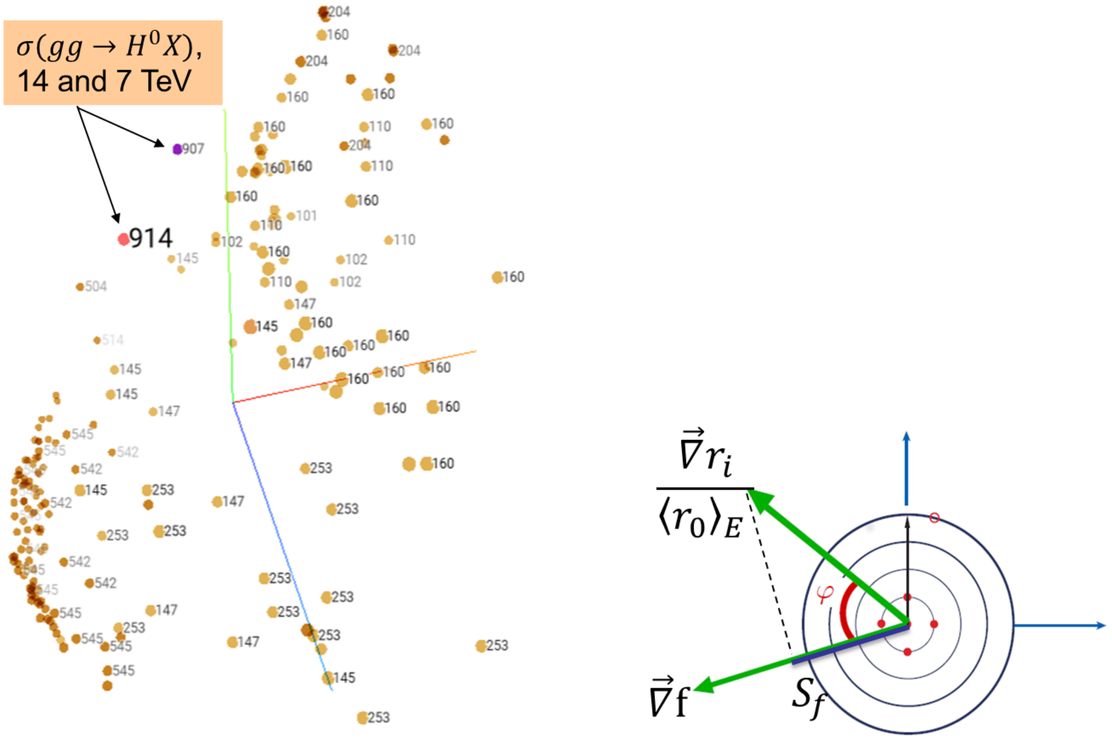


Figure 2.6: Left: A PDF-dependent quantity f defines a direction in space of (2)N PDF parameters. The direction is specified by the gradient $\vec{\nabla} f$ in the symmetric convention. Here, the Embedding Projector [63] visualizes the vectors $\vec{\delta}_{907}$ and $\vec{\delta}_{914}$ for NNLO cross sections for Higgs boson production at 7 and 14 TeV, and vectors $\vec{\delta}_i$ for CT14HERA2 NNLO data points from [64] (brown circles), showing only $\vec{\delta}_i$ with the smallest angular distances to $\vec{\delta}_{914}$. These points impose the strongest constraints on the PDF dependence of the Higgs cross sections in the CT14HERA2 analysis, if they have large enough $|\vec{\delta}_i|$. Again, in the numbering scheme used here, points labeled 1XX correspond to fixed-target measurements, 2XX to Drell-Yan processes and boson production, and 5XX to jet and $t\bar{t}$ production as given in Tables B.1–B.3. Right: the sensitivity S_f of the i-th data residual can be interpreted as the projection of $\vec{\delta}_i \equiv \vec{\nabla} r_i / \langle r_0 \rangle_E$ onto the direction of $\vec{\nabla} f$.

— analogously to how S_f de-emphasizes (relative to the correlation C_f) those data points whose normalized residual variations $\Delta r_i/\langle r_0\rangle_E$ have already been more tightly constrained.

2.3.3. Sensitivity in the Monte-Carlo Method

The above statistical measures are general enough and can be extended to other representations for the PDF uncertainties, such as the representation based on Monte-Carlo replica PDFs [15, 17, 60] of the kind employed, e.g., in the NNPDF framework. A family of Monte-Carlo PDFs consists of N_{rep} member PDF sets $q_a^{(k)}(x,\mu) \equiv \{q^{(k)}\}$, with $k = 1,\ \ldots,\ N_{\mathrm{rep}}$, and those are used to determine an expectation value $\langle X\rangle$ for a PDF-dependent quantity $X[\{q\}]$ such as a high-energy cross section:

$$\langle X\rangle = \frac{1}{N_{\mathrm{rep}}}\sum_{k=1}^{N_{\mathrm{rep}}} X[\{q^{(k)}\}]\ . \tag{2.24}$$

The resulting Monte-Carlo uncertainty on X can be extracted from the ensemble as

$$\Delta_{\mathrm{MC}}X \ = \left(\frac{1}{N_{\mathrm{rep}}-1}\sum_{k=1}^{N_{\mathrm{rep}}}\left(X[\{q^{(k)}\}] - \langle X\rangle\right)^2\right)^{1/2} \ . \tag{2.25}$$

In consequence of these definitions, the central value of a particular PDF itself in the NNPDF framework is specified as

$$q_{(0)} \equiv \langle q\rangle = \frac{1}{N_{\mathrm{rep}}}\sum_{k=1}^{N_{\mathrm{rep}}} q^{(k)}\ . \tag{2.26}$$

Akin to the Pearson correlation defined in Eq. (2.19) of Sec. 2.3.1, statistical correlations between two PDF-dependent quantities $X[\{q\}]$ and $Y[\{q\}]$ can be constructed from the PDF replica language above in terms of ensemble averages [17]:

$$\mathrm{Corr}_{\mathrm{MC}}\left[X,Y\right] = \frac{\langle XY\rangle - \langle X\rangle\langle Y\rangle}{\Delta_{\mathrm{MC}}X\Delta_{\mathrm{MC}}Y}\ . \tag{2.27}$$

Then, using our definitions in Eqs. (2.19) and (2.21), we immediately construct the realizations of the correlation and sensitivity for a PDF-dependent quantity f in the Monte-Carlo method:

$$C_{f,\ \mathrm{MC}} = \mathrm{Corr}_{\mathrm{MC}}[f, r_i]\ , \tag{2.28}$$

$$S_{f,\ \mathrm{MC}} = \frac{\Delta_{\mathrm{MC}} r_i}{\langle r_0 \rangle_E} \mathrm{Corr}_{\mathrm{MC}}\left[f, r_i\right]. \tag{2.29}$$

2.4. Case study: CTEQ-TEA Global Data

2.4.1. Maps of Correlations and Sensitivities

We will now discuss a number of practical examples of using C_f or S_f to quickly evaluate the impact of various hadronic data sets upon the knowledge of the PDFs in a fashion that does not require a full QCD analysis of the type described in Sec. 2.2. For this demonstration, we will continue to study the dataset shown in Fig. 2.1 of the CT14HERA2 analysis [41] augmented by the candidate LHC data.

We have already noted the extent of this dataset in the $\{x, \mu\}$ plane in Fig. 2.1, where it is decomposed into constituent experiments labeled according to the conventions in Tables B.1-B.3. It is instructive to create similar maps in the $\{x, \mu\}$ plane showing the C_f or S_f values for each data point. Such maps are readily produced by the PDFSense program for a variety of PDF flavors and for user-defined observables, such as the Higgs cross section. For demonstration we have collected a large number of these maps at the companion website [64]. We invite the reader to review these additional figures while reading the paper to validate the conclusions that will be summarized below.

Thus, we obtain scatter plots of $C_f(x_i, \mu_i)$ or $S_f(x_i, \mu_i)$ for a given QCD observable $f = \sigma$, such as the LHC Higgs production cross section shown in Fig. 2.2, or with a PDF f evalu-

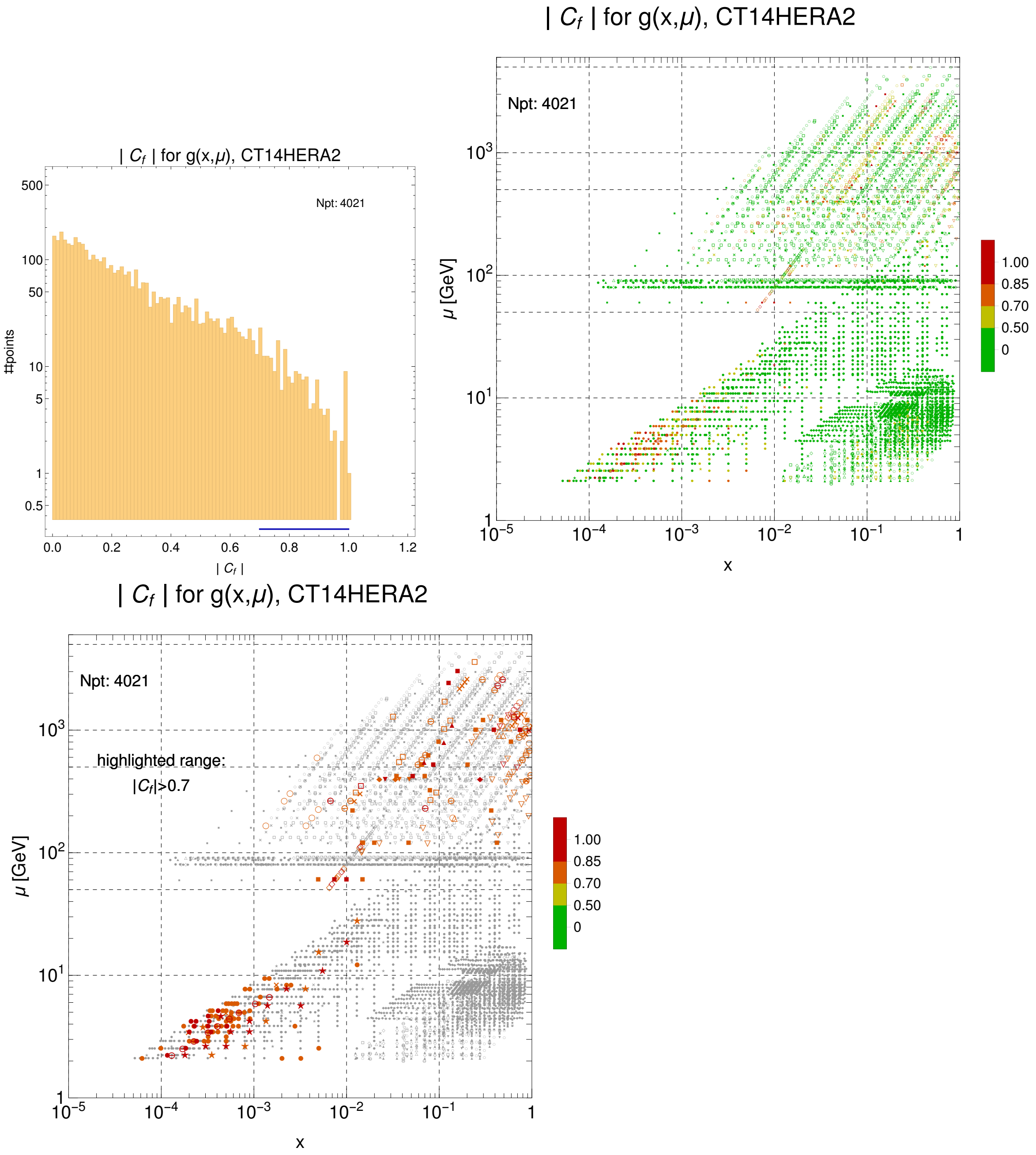


Figure 2.7: Representations of the correlation $|C_g|(x_i, \mu_i)$ of the gluon PDF $g(x, \mu)$ with the point-wise residual r_i of the augmented CT14HERA2 analysis. In the first panel, we plot a histogram showing the distribution of correlations for 4021 physical measurements. In the second panel we show the 5227-point $\{x_i, \mu_i\}$ map corresponding to these data within the full dataset, generated as in Appendix A. To adjust for the fact that some measurements of rapidity dependent quantities match to two distinct points in $\{x_i, \mu_i\}$ space using the rules of Appendix A, we assign weights of 0.5 to these complementary $\{x_i, \mu_i\}$ points in computing the $N_{pt} = 4021$-count histogram at left. The third figure is the same as the second one, but only the data points satisfying $|C_f|> 0.7$ are highlighted.

ated at the same $\{x_i, \mu_i\}$ determined by the data points, with examples shown for $g(x_i, \mu_i)$ in Figs. 2.7 and 2.8. The typical $\{x_i, \mu_i\}$ values characterizing the data points are found according to Born-level approximations appropriate for each scattering process included in the CTEQ-TEA dataset, with the formulas to compute these kinematic matchings summarized in App. A. Here and in general, we find it preferable to consider the absolute values $|C_f|$ and $|S_f|$ on the grounds that the signs of C_f and S_f flip when the data points randomly overshoot or undershoot their theory predictions.

Together with the map in the $\{x, \mu\}$ plane, PDFSense also returns a histogram of the values for each quantity it plots. An example is shown for $|C_g|\,(x_i, \mu_i)$ in the first panel of Fig. 2.7. One would judge that stronger constraints are in general provided to those PDFs for which the $|C_f|$ histogram has many entries comparatively closely to $|C_f| \sim 1$. In the first panel of Fig. 2.7, we can see that, while the distribution peaks at low correlations, $|C_g| \sim 0$, the distribution has an extended tail in the region $0.7 \lesssim |C_g| \lesssim 1$. This feature shows that, of the 4021 experimental data points within the augmented CT14HERA2 set in Fig. 2.1, nearly two-hundred — specifically, 192 — have especially strong ($|C_f| \geq 0.7$) correlations (or anti-correlations) with the gluon PDF. This region of such strong correlations within the histogram is indicated by the horizontal blue bar that runs along the abscissa.

To identify these points, we plot complementary information in the second panel of the same figure – specifically, a map in $\{x, \mu\}$ space of each of the data points shown in Fig. 2.1. As before, they are colorized according to the magnitude of $|C_g|$ following the color palette in the "rainbow strip" on the right. "Cooler" colors (green/yellow) correspond to weaker correlation strengths, while "hotter" colors (orange/red) represent comparatively stronger correlations, as indicated. To reveal the data points with the highest correlations, we reproduce the same figure in the third panel, but showing in color only the data points satisfying $|C_f| > 0.7$. Thus, we obtain two maps in the $\{x, \mu\}$ plane that look similar to the $|C_f|$ map in the left panel of Fig. 2.2, apart from the differences that (a) Fig. 2.7 shows the

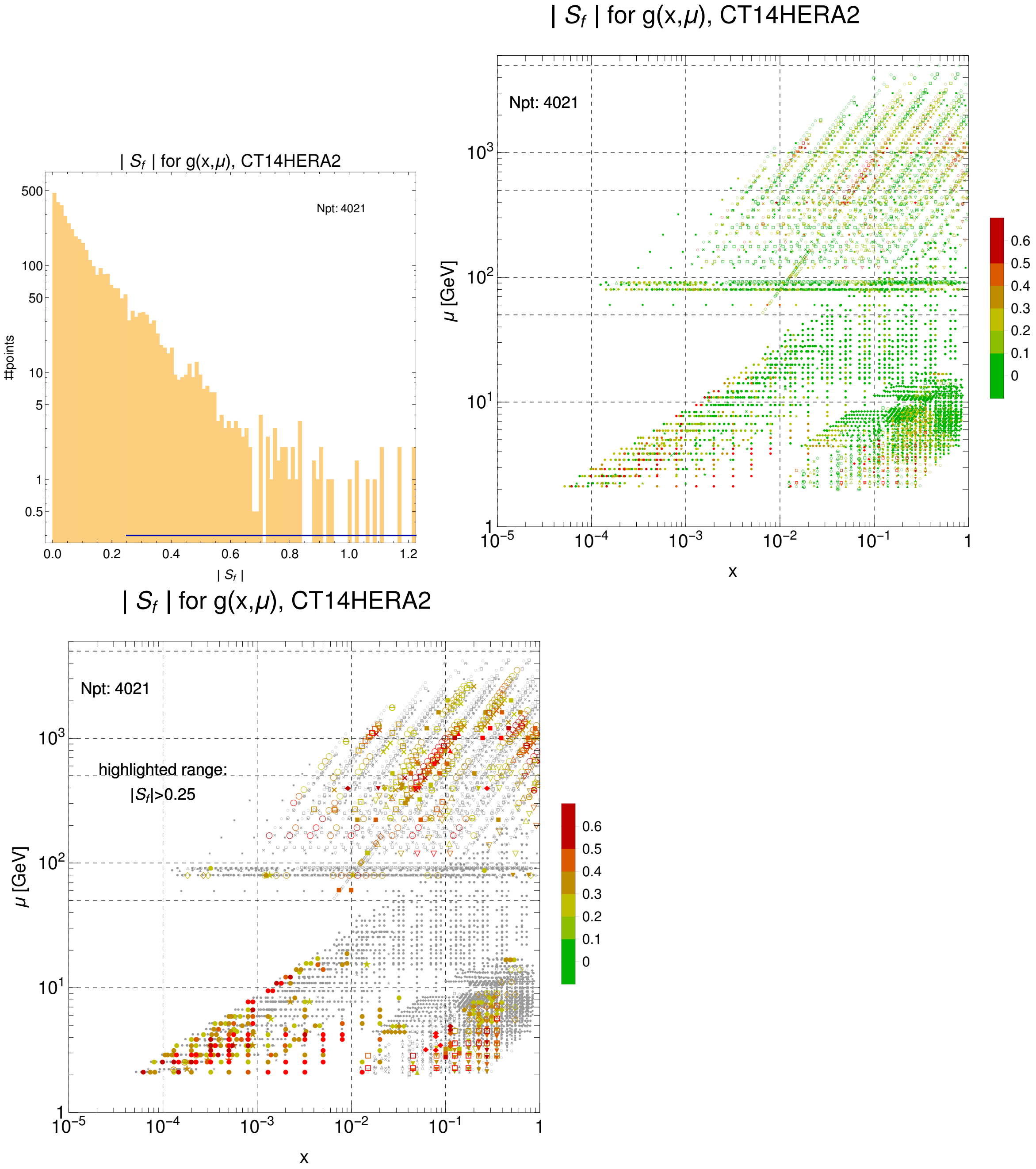


Figure 2.8: Like Fig. 2.7, but for the gluon sensitivity $|S_g|(x_i, \mu_i)$ as defined in Eq. (2.21). In the third figure, only the data points satisfying $|S_f|> 0.25$ are highlighted.

correlation $|C_g|$ for $g(x_i, \mu_i)$ at the same typical values $\{x_i, \mu_i\}$ as in the data, rather than $|C_{\sigma_{H^0}}|$ for Higgs production cross section in Fig. 2.2; and (b) Fig. 2.2 highlights 310 points with the highest $|C_{\sigma_{H^0}}|$.

The correlations for the LHC Higgs production cross section trace those for $g(x_i, \mu_i)$, but not entirely, as we will see in a moment. Large magnitudes of $|C_g|$ in Fig. 2.7 are found for inclusive jet production measurements, especially those recently obtained by CMS at 8 TeV [67] (Expt. CMS8jets'17, inverted triangles) with $|C_g|(x_i, \mu_i)$ as high as 0.85, including at the highest values of x and μ. Beyond these, a sizable cluster of HERA (HERAI+II'15) data points at the lowest values of x are also seen to have large correlations with the gluon PDF, consistent with the common wisdom that HERA DIS constrains the gluon PDF at small x via DGLAP scaling violations. Under the jet production cluster, high-p_T Z production (ATL7ZpT'14, ATL8ZpT'16) and $t\bar{t}$ production (ATL8ttb-pt'16, ATL8ttb-y_ave'16, ATL8ttb-mtt'16, ATL8ttb-y_ttb'16) at the LHC show a high $|C_g|(x_i, \mu_i)$ correlation. At the same time, many other measurements, including fixed-target data at large x and W asymmetry data near $\mu \sim 100$ GeV, have feeble correlations with $g(x_i, \mu_i)$ and would therefore be less emphasized by an analysis based solely upon the PDF-residual correlations.

We can also consider the analogous plots for the sensitivity $|S_g|\,(x_i, \mu_i)$ as defined in Eq. (2.21), which we plot in Fig. 2.8. In the first panel, we again consider the histogram, here for the magnitudes of the gluon sensitivity $|S_g|(x_i, \mu_i)$, in which the correlations $|C_g|$ are now weighted by the relative size of the PDF uncertainty Δr_i in the residual. As discussed in Sec. 2.3.2, this additional weighting emphasizes those data points for which the PDF-driven fluctuations in the residuals are comparatively large relatively to experimental uncertainties. This leads to a redistribution of the data points shown in the $|C_g|$ histogram of Fig. 2.7, with the result being a considerably longer-tailed histogram for $|S_g|$ such that, in this instance, there are 546 raw data points with larger sensitivities, $|S_f| \geq 0.25$, indicated by the horizontal blue bar. Unlike the correlation, $|S_g|$ can be arbitrarily large, depending on the Δr_i value.

It is suppressed at the data points with large uncertainties or smeared over the regions of data points with correlated systematic uncertainties.

In the second and third panels, we show the respective $\{x, \mu\}$ maps for $|S_g|$, with color highlighting given either for all points or only those with high sensitivities $|S_f|> 0.25$, respectively. $|S_g|$ places additional emphasis on the combined HERA dataset (HERAI+II'15) constraining $g(x_i, \mu_i)$ at lowest x. In contrast to the $|C_g|$ plot, we observe increased sensitivity in the precise fixed-target DIS data from BCDMS (BCDMSp'89, BCDMSd'90) and CCFR (CCFR-F2'01, CCFR-F3'97), which are sensitive to the gluon via scaling violations despite only moderate correlation values. Similarly, we observe heightened sensitivities at highest x for the LHC (CMS7jets'14, ATLAS7jets'15, CMS8jets'17) and Tevatron (D02jets'08) jet production data, which have both large correlations with $g(x_i, \mu_i)$ and small experimental uncertainties. Sensitivity $|S_g|$ of LHC jet experiments, CMS7jets'14, ATLAS7jets'15, CMS8jets'17, varies in a large range, and can significantly improve, depending on the implementation of experimental systematic uncertainties in the analysis, cf. the discussion of the jet data in the next section.

We also observe enhanced sensitivity for *individual points* in a large number of experiments, including CDHSW DIS (CDHSW-F2'91); HERA F_L (HERA-FL'11); the Drell-Yan process (E605'91, E866pp'03); CDF 8 TeV W charge asymmetry (CMS7Masy2'14); HERA charm SIDIS (HERAc'13); ATLAS high-p_T Z production (ATL7ZpT'14, ATL8ZpT'16); and especially strongly sensitive points in $t\bar{t}$ production (ATL8ttb-pt'16, ATL8ttb-y_ave'16, ATL8ttb-mtt'16, ATL8ttb-y_ttb'16). However, since the latter category includes fewer points per each experiment, it constrains the gluon less than the high-statistics DIS and jet production data.

These findings comport with the idea that the gluon PDF remains dominated by substantial uncertainties at both $x \sim 0$ and in the elastic limit $x \to 1$, a fact which has driven an intense focus upon production of hadronic jets, $t\bar{t}$ pairs, and high-p_T Z bosons, which themselves are measured at large center-of-mass energies $\sqrt{s}$ and are expected to be sensitive to the gluon PDF across a wide interval of x, including $x \sim 0.01$ typical for Higgs boson production via gluon fusion at the LHC. Turning back to the distributions of $|C_{\sigma_H}|(x_i, \mu_i)$ and $|S_{\sigma_H}|(x_i, \mu_i)$ for the Higgs cross section σ_H at $\sqrt{s} = 14$ TeV in Fig. 2.2, we notice that they largely reflect the distributions of $|C_g|(x_i, \mu_i)$ and $|S_g|(x_i, \mu_i)$ around $x \sim M_H/\sqrt{s} = 125/14000 = 0.009$ and $\mu = M_H = 125$ GeV. We also see some differences: although the average x and μ are fixed in σ_H, it is nonetheless sensitive to some constraints at much lower x values as a result of the momentum sum rule.

The reader is welcome to examine the plots of sensitivities and correlations available on the PDFSense website for a large collection of PDF flavors and PDF ratios, such as d/u, $\overline{d}/\overline{u}$, and $(s+\overline{s})/\big(\overline{u}+\overline{d}\big)$. Sensitivities for other PDF combinations and hadronic cross sections can be computed and plotted in a matter of minutes using the PDFSense program. We will now turn to another aspect of this analysis: summarizing the abundant information contained in the sensitivity plots. For this purpose, we will introduce numerical indicators and propose a practical procedure to rank the experimental data sets according to their sensitivities to the PDFs or PDF-dependent observables of interest.

2.4.2. Experiment Rankings According to Cumulative Sensitivities

Being one-dimensional projections of normalized residual variations $\vec{\delta}_i$ on a given direction in the PDF parameter space, sensitivities can be linearly added to construct a number of useful estimators. By summing absolute sensitivities $|S_f^{(i)}|$ over the data points i of a given data set E, we find the maximal cumulative sensitivity of E to the PDF dependence of a QCD observable f.

Alternatively, from the examination of multiple $\{x, \mu\}$ maps for $|S_f|$ of various PDF flavors collected on the website [64], we find that the most precise experiments constrain several flavors at the same time; most notably, the combined HERA data. For the purpose of identifying such experiments, we can compute an overall sensitivity statistic for each experiment E to the parton distributions $f_a(x_i, \mu_i)$ evaluated at the same kinematic parameters $\{x_i, \mu_i\}$ as the data. Furthermore, to obtain one overall ranking, we can add up sensitivity measures as an unweighted sum over the "basis PDF" flavors, such as the six light flavors $(\overline{d}, \overline{u}, g, u, d, s)$. To obtain these measures, we say that an experiment E consisting of N_{pt} physical measurements can be characterized by its mean sensitivity per raw data point[4] to a PDF of given flavor $f_a(x, \mu)$: $\langle |S_f^E| \rangle \equiv (N_{pt})^{-1} \sum_{i=1}^{N_{pt}} |S_f| (x_i, \mu_i)$, from which we derive several additional statistical measures of experimental sensitivity. For each experiment and flavor we then determine a cumulative sensitivity measure, numerically adjusted to the size of each experimental dataset E, according to $|S_f^E| \equiv N_{pt} \langle |S_f^E| \rangle$. In addition, we also track cumulative, flavor-summed sensitivity measures $\sum_f |S_f^E|$ and $\langle \sum_f |S_f^E| \rangle$, with f running over $\overline{d}$, $\overline{u}$, g, u, d, s.

We list the corresponding values of these four types of sensitivities for each experiment of the CTEQ-TEA dataset in summary tables in App. B.1 as well as extensive Supplementary Material in App. B.2. This is also detailed for categories of experiments from the CTEQ-TEA dataset.

With the above estimators, we *quantify* and *compare* the cumulative sensitivities of each experiment to the basis 6 parton flavors. In fact, based on the various trials that we performed, we find that the cumulative sensitivity to the 6 basic flavors is a good measure of the overall sensitivity to a large range of PDF combinations. Recall that the $N_f = 5$ CT14HERA2 PDFs (with up to 11 independent parton species) are obtained by DGLAP evolution of the 6 basic parton flavors from the initial scale of order 1 GeV. There exist alternative approaches for measuring the importance of a given experiment in a global fit,

[4] For those circumstances in which an individual measurement, *e.g.*, obtained via the Drell-Yan process, maps to two sensitivity values in $\{x, \mu\}$ space, we compute the average of these and assign the result to that specific measurement.

for example, by counting the numbers of eigenvector parameters [68] or eigenvector directions [35] that the experiment constrains. Those other methods, however, require access to the full machinery of the global fit, while the sensitivities allow the reader to rank the experiments according to much the same information, for a variety of PDF-dependent observables, with the help of PDFSense, and at a fraction of computational cost.

In fact, in a companion study we use the above sensitivity estimators to select the new LHC experiments for the inclusion in the next generation of the CTEQ-TEA PDF analysis. Full tables given in App. B.1 and in the Supplementary Material of App. B.2 provide detailed information about the PDF sensitivities of every experiment of the CTEQ-TEA data set. For a non-expert reader, along the full tables, we provide their simplified versions in Tables B.4-B.5, where we rank the experimental sensitivities according to a reward system described in the caption of Table B.4. In each table, experiments are listed in descending order according to the cumulative sensitivity measure $\sum_f |S_f^E|$ to the six light-parton flavors. For each PDF flavor, the experiments with especially high overall flavor-specific sensitivities receive an "**A**" rating (shown in bold), per the convention in the caption of Table B.4. Successively weaker overall sensitivities receive marks of "B" and "C," while those falling below a lower limit $|S_f^E|= 20$ are left unscored.

We similarly evaluate each experimental dataset based on its point-averaged sensitivity, in this case scoring according to a complementary scheme in which the highest score is "**1**". The short-hand names of the candidate experiments that were *not* included in the CT14HERA2 NNLO fit, that is, the new LHC experiments, are also shown in bold to facilitate their recognition in the tables.

Not only do the sensitivity rankings confirm findings known by applying other methods, they also provide new insights. According to this ranking system in Tables B.4-B.5, we find that the expanded HERA dataset (HERAI+II'15) tallies the highest overall sensitivity to the PDFs, with enhanced sensitivity to the distributions of the u- and $\bar{u}$-quarks, as well as that of the gluon. On similar footing, but with slightly weaker overall sensitivities, are a number of other fixed-target measurements, including structure function measurements from BCDMS for $F_2^{p,d}$ (BCDMSp'89, BCDMSd'90) and CCFR extractions of xF_3^p (CCFR-F3'97) — as well as several other DIS datasets. Among the LHC experiments, the inclusive jet measurements have the highest cumulative sensitivities, with CMS jets at 8 TeV (CMS8jets'17), 7 TeV (CMS7jets'13, CMS7jets'14), and ATLAS 7 TeV (ATLAS7jets'15) occupying positions 10, 12/13, and 16 in the total sensitivity rankings. They demonstrate the strongest sensitivities among the candidate LHC experiments, and at the same time are not precise enough and fall behind the top fixed-target DIS and Drell-Yan experiments: BCDMS, CCFR, E605, E866, and NMC. The two versions CMS7jets'13 and CMS7jets'14 of the CMS 7 TeV jet data that largely overlap have very close sensitivities and rankings in Tables B.4-B.5. The set CMS7jets'13 that extends to higher p_{Tj} has a slightly better overall sensitivity, surpassing the larger data set CMS7jets'14 that includes the extra data points at $p_{Tj} < 100$ GeV or $|y_j| > 2.5$, yet cannot beat CMS7jets'13 except for in the overall sensitivity to the Higgs cross section at 7 TeV.

Going beyond the rankings based upon overall sensitivities, which are more closely tied to the impact of an entire experimental dataset in aggregate, it is useful to consider the point-averaged sensitivity as well, which quantifies how sensitive each individual point is. [Some experiments with very high point-averaged sensitivity have a small cumulative sensitivity because of a small number of points.] Based on their high point-averaged sensitivity, CMS μ asymmetry measurements at 8 and 7 TeV (CMS8Wasy'16 and CMS7Masy2'14) especially stand out, despite their small number of individual points, $N_{pt} = 11$); this is especially true again for the gluon, $\bar{d}$-, and u-quark PDFs, for which this set of measurements is particularly highly rated in Table B.4. Another "small-size" data set with the exceptional point-average sensitivity is the $\sigma_{pd}/(2\sigma_{pp})$ ratio from the E866 lepton pair production experiment (E866rat'01). The average sensitivity of this data set to $\bar{u}$ and $\bar{d}$ PDFs is 0.8, making it extremely valuable for constraining the ratio $\overline{d}/\overline{u}$ at $x \sim 0.1$, in spite of its small size (15 data points).

Aside from the quark- and gluon-specific rankings of specific measurements, we can also assess experiments based upon the constraints they impose on various interesting flavor combinations and observables as presented in Table B.5. As was the case with Table B.4, a considerable amount of information resides in Table B.5 of which we only highlight several notable features here. Among these features are the sharp sensitivities to the Higgs cross section (*e.g.*, $|S|_{H7}$, $\langle|S_{H7}|\rangle$, *etc.*) found for Run I+II HERA data, as well as the tier-C overall sensitivities of the BCDMS $F_2^{p,d}$ and CMS jet production measurements, corresponding to Exps. BCDMSd'90, BCDMSp'89, CMS8jets'17 and CMS7jets'14. While their overall sensitivity is small, the corresponding ATLAS $t\bar{t}$ data also possesses significant point-averaged sensitivity. On the other hand, measurements of p_T-dependent Z production (ATL7ZpT'14, ATL8ZpT'16) appear to have somewhat less pronounced sensitivity to the gluon and other PDF flavor combinations. The total and mean sensitivities of high-p_T Z boson production experiment ATL8ZpT'16 at 8 TeV is on par with HERA charm SIDIS data (HERAc'13) and provides comparable constraints to charm DIS production, albeit in a different $\{x, \mu\}$ region.

For the light-quark PDF combinations like u_v, d_v, d/u, and $\overline{d}/\overline{u}$, the various DIS datasets — led by Run II of HERA and CCFR measurements of the proton structure function — demonstrate the greatest sensitivity. At the same time, however, Run-2 Tevatron data from D0 on the μ asymmetry (D02Easy2'15) and Run-1 CDF measurements for the corresponding $A_e(\eta^e)$ asymmetry (CDF1Wasy'96) also exhibit substantial point-wise sensitivity as well. We collect a number of other observations in the conclusion below, Sec. 2.5.

2.4.3. Estimating the Impact of LHC Datasets on CTEQ-TEA Fits

The presented rankings suggest that including the candidate LHC data sets will produce mild improvements in the uncertainties of the CT14 HERA2 PDFs. This projection may appear underwhelming, but keep in mind that the CT14HERA2 NNLO analysis already includes significant experimental constraints, for example, imposed on the gluon PDF at $x > 0.01$ by the Tevatron and LHC jet experiments, CDF2jets'09, D02jets'08, ATL7jets'12, CMS7jets'13. If all jet experiments are eliminated from the PDF fit, as illustrated in the Supplementary Material tables of App. B.2, the candidate LHC experiments will be promoted to higher rankings, with the CMS 8 and 7 TeV jet experiments (CMS8jets'17 and CMS7jets'13/CMS7jets'14) elevated to positions 4 and 7/8 in the overall sensitivity rankings, respectively.

Our investigations also find that the sensitivities of CMS jet experiments may improve considerably if the current correlated systematic effects are moderately reduced compared to the published values. For instance, by requiring a full correlation of the JEC2 correlation error over all rapidity bins in the CMS 7 TeV jet data set CMS7jets'14, instead of its partial decorrelation implemented according to the CMS recommendation [69], we obtain a very strong sensitivity of the data set CMS7jets'14 to g over the full $\{x, \mu\}$ region; but also strong sensitivities to $\overline{u}, \overline{d}$, and even $\overline{s}$ PDFs.[5] The overall sensitivity of the data set CMS7jets'14 in this case is elevated to the 4th position from the 13th position in the CT14HERA2 NNLO analysis in Tables B.4 and B.5. Similarly, for the CMS 8 TeV jet data set CMS8jets'17, the sensitivity to the above flavors can increase under moderate reduction of systematic uncertainties, easily surpassing the sensitivity of CMS7jets'14 because of the larger number of points in CMS8jets'17.

[5] With the fully correlated jet energy correction JEC2 source, the data set CMS7jets'14 would provide a strong overall constraint on $s(x, \mu)$ comparable to one of the NuTeV or neutrino CCFR experimental data sets.

2.4.4. PDFSense Predictions versus Lagrange Multiplier Scans

How do the surveys based on PDFSense compare against the actual fits? As we noted, the PDFSense method is designed to provide a fast large-scope estimation of the impact of the existing and future data sets in conjunction with other tools, such as the ePump [56] program for PDF reweighting. It works the best in the quadratic (Hessian) approximation near the best fit, and when the new experiments are compatible with the old ones. When detailed understanding of the experimental constraints is necessary, the PDFSense approach must be supplemented by other techniques, such as Lagrange multiplier (LM) scans [61, 70, 71].

As an illustration of the scope of the differences between the PDFSense predictions before and after the fit, the left panels in Figs. 2.9 and 2.10 show the PDFSense maps for $d/u(x = 0.1, \mu = 1.3\ \text{GeV})$ and $g(x = 0.01, \mu = 125\ \text{GeV})$ evaluated using a preliminary CT18 NNLO fit (designated as "CT18pre") that includes 11 new LHC experimental data sets, namely CMS8jets'17, CMS7jets'14, ATLAS7jets'15, LHCb8WZ'16, CMS8Wasy'16, LHCb8Zee'15, LHCb7ZWrap'15, ATL8ZpT'16, ATL8ttb-pt'16, ATL8ttb-mtt'16, and 8 TeV $t\bar{t}$ production at CMS ('CMS8 ttb pTtyt') [72] in addition to the experiments included in the CT14HERA2 fit. The full details of the CT18 fit will be presented in an upcoming publication [73]. Some modifications were made in the methodology adopted in CT18, as compared to CT14HERA2; notably the PDF parametrization forms and treatment of NNLO radiative contributions have been changed, while some shown curves are also subject to a theoretical uncertainty associated with the QCD scale choices. In accord with the PDFSense predictions based on the CT14HERA2 NNLO PDFs, we find that including the above LHC experiments into the fit produces only mild differences between the CT18pre and CT14HERA2 NNLO PDFs. Consequently the PDFSense $\{x, \mu\}$ maps based on CT18pre NNLO PDFs are similar to the CT14HERA2 ones [64]. One noticeable difference is that the sensitivity of the new experiments decreases after they are included in the CT18pre fit, be-

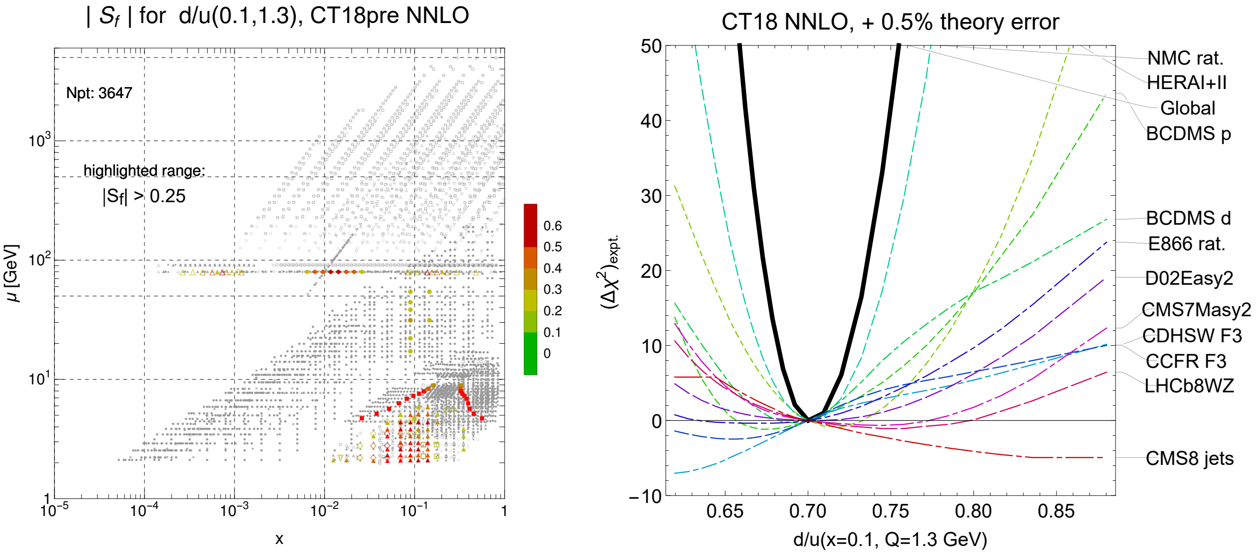


Figure 2.9: Left: the PDFSense map for the sensitivity of the fitted dataset of the CT18pre NNLO analysis to the d/u PDF ratio, $d/u(x=0.1, \mu=1.3\ \text{GeV})$. Right: Dependence of χ^2 for the individual and all experiments of the CT18pre dataset on the value of $d/u(x=0.1, \mu=1.3\ \text{GeV})$ obtained with the LM scan technique. The curves show the deviations $\Delta\chi^2_{\text{expt.}} \equiv \chi^2_{\text{expt.}}(\vec{a}) - \chi^2_{\text{expt.}}(\vec{a}_0)$ from the best-fit values in χ^2 for the indicated experiments, as well as for the totality of all experiments.

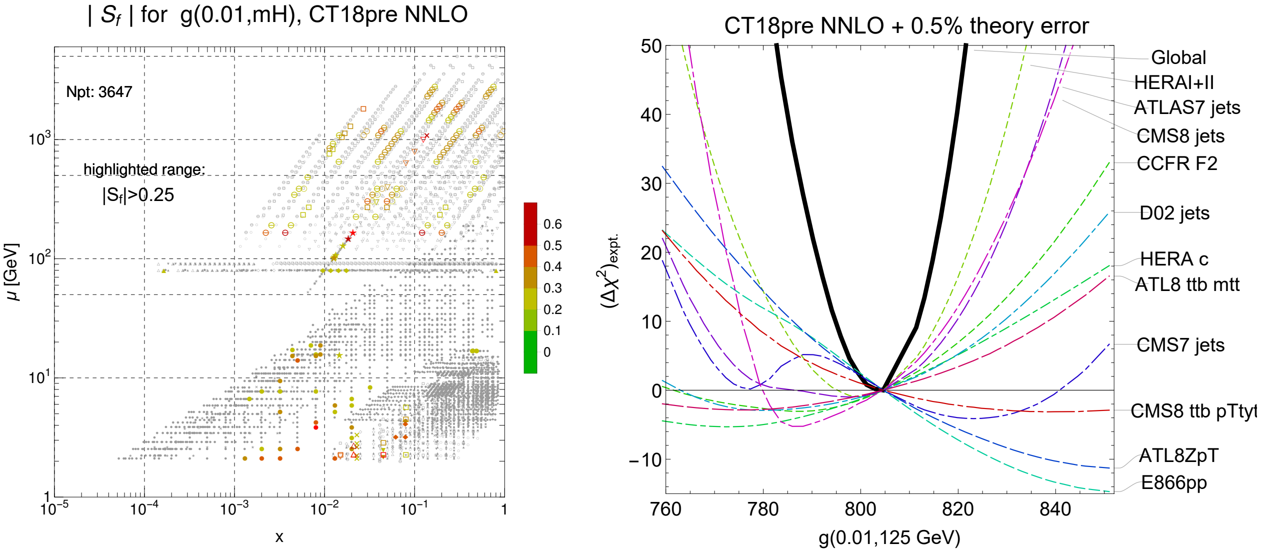


Figure 2.10: Like Fig. 2.9, but comparing the PDFSense map (left) and LM scan (right) for the gluon PDF $g(x=0.01, \mu=m_H)$ in the Higgs boson production region.

cause the new information from the newly added experiments suppresses PDF uncertainties of data residuals.

In the right panels of Figs. 2.9 and 2.10, we illustrate the constraints on the same quantities, $d/u(0.1, 1.3\text{ GeV})$ and $g(0.01, 125\text{ GeV})$ in the candidate CT18pre NNLO fit, now obtained with the help of LM scans. A LM scan [61,70,71] is a powerful technique that elicits detailed information about a PDF-dependent quantity $X(\vec{a})$, such as a PDF or cross section, from a constrained global fit in which the value of $X(\vec{a})$ is fixed by an imposed condition. By minimizing a modified goodness-of-fit function $\chi^2_{LM}(\lambda, \vec{a})$ that includes a 'generalized-force' term equal to $X(\vec{a})$ with weight λ, in addition to the global χ^2_{global} in Eq. (2.4), a LM scan reveals the parametric relationship between $X(\vec{a})$ and χ^2_{global} or $\chi^2_{\text{expt.}}$ contributions from individual experiments, including any non-Gaussian dependence. In the LM scans at hand, the modified fitted function takes the form

$$\chi^2_{LM}(\lambda, \vec{a}) = \chi^2_{\text{global}}(\vec{a}) + \lambda X(\vec{a}), \tag{2.30}$$

and $X(\vec{a})$ are $d/u(x, \mu)$ or $g(x, \mu)$ at a specific location in $\{x, \mu\}$ space. For the optimal parameter combination $\vec{a} \equiv \vec{a}_0$ at which $\chi^2_{\text{global}}(\vec{a})$ is minimized, we find in Fig. 2.9 that $d/u(0.1, 1.3\text{ GeV}) \approx 0.7$. The LM scan for the d/u then consists of a series of refits of the parameters $\vec{a}_k$, as the multiplier parameter λ is dialed along a set of discrete values λ_k, effectively pulling d/u away from the value ~ 0.7 at $\vec{a} = \vec{a}_0$ preferred by the global fit. The right panel of Fig. 2.9 shows the relationship between $d/u(0.1, 1.3\text{ GeV})$ and χ^2_{global} that is quantified this way; and similarly for $g(0.01, 125\text{ GeV})$.

We can also examine how the χ^2 changes for the individual experiments. Figs. 2.9 and 2.10 show the curves for 11 experiments with the largest variations $\max(\chi^2) - \min(\chi^2)$ in the shown ranges of d/u and g, i.e., the most constraining experiments. We notice that, while the $\Delta\chi^2$ dependence is nearly Gaussian for the total χ^2, it is sometimes less so for the individual experiments. Some experiments may be inconsistent when they have a large best-fit $\chi^2(\vec{a}_0)$ or prefer an incompatible X value. Figure 2.9 is an example of a good agreement between the experiments, when the individual $\Delta\chi^2_{expt.}$ curves are approximately quadratic and minimized at about the same location. Figure 2.10 shows more pronounced inconsistencies, notably in the case of the E866pp and ATL8ZpT curves that prefer a significantly larger $g(0.01, 125\text{ GeV})$ than in the rest of the experiments.

The LM procedure thus allows a systematic exploration of the exact constraints from the experiments on X without relying on the Gaussian assumption that is inherent to the PDFSense method. Both PDFSense and LM scans successfully identify the experiments with the strongest sensitivity to X, while their specific rankings of such experiments are not strictly identical and reflect the chosen ranking prescription and settings of the global fit. We emphasize that, though informative, the LM scans are computationally intensive, with a typical 30-point scan at NNLO requiring ∼6500 CPU core-hours on a high-performance cluster. This is in contrast to the PDFSense analysis, which can be run for our entire 4021-point dataset on a single CPU core of a modern workstation in ∼5 minutes, representing a $\sim 0.8 \times 10^5$ savings in computational cost.

Let us further illustrate these observations by referring again to Figs. 2.9 and 2.10, as well as to Table 2.1 that displays the top 10 experiments with the largest cumulative sensitivity to $d/u(0.1, 1.3\text{ GeV})$ and $g(0.01,\ 125\text{ GeV})$ according to PDFSense and LM scans, with either CT14HERA2 or CT18pre PDFs used to construct the PDFSense rankings. In the PDFSense columns, the experiments are ranked in order of descending cumulative sensitivities $\sum_{i=1}^{N_{pt}} |S_f|(x_i, \mu_i)$ according to the same prescription as in Sec. 2.4.2. For the LM scans, the table shows the experiments that have the largest variations $\max(\chi^2) - \min(\chi^2)$ in the range of X corresponding to $\Delta\chi^2_{global} \leq 100$, that is, within approximately the 90% probability level interval of the CT18pre NNLO PDFs. As the residual uncertainties Δr_i in the sensitivities S_f are normalized to the root-mean-squared residuals $\langle r_0 \rangle_E$ at the best

$d/u(x\!=\!0.1, \mu\!=\!1.3\ \text{GeV})$			$g(x\!=\!0.01, \mu\!=\!125\ \text{GeV})$		
PDFSense		LM scan	PDFSense		LM scan
CT14HERA2	CT18pre	CT18pre	CT14HERA2	CT18pre	CT18pre
HERAI+II’15	NMCrat’97	NMCrat’97	HERAI+II’15	HERAI+II’15	HERAI+II’15
BCDMSp’89	HERAI+II’15	CCFR-F3’97	CMS8jets’17	CMS8jets’17	CMS8jets’17
NMCrat’97	BCDMSp’89	HERAI+II’15	CMS7jets’14	CMS7jets’14	ATL8ZpT’16
CCFR-F3’97	CCFR-F3’97	BCDMSd’90	ATLAS7jets’15	E866pp’03	E866pp’03
E866pp’03	BCDMSd’90	BCDMSp’89	E866pp’03	ATLAS7jets’15	ATLAS7jets’15
BCDMSd’90	E605’91	CDHSW-F3’91	BCDMSd’90	BCDMSd’90	CCFR-F2’01
CDHSW-F3’91	E866pp’03	E866rat’01	CCFR-F3’97	BCDMSp’89	D02jets’08
CMS8jets’17	E866rat’01	CMS7Masy2’14	D02jets’08	D02jets’08	HERAc’13
E866rat’01	CMS8jets’17	NuTeV-nu’06	NMCrat’97	NMCrat’97	NuTeV-nub’06
LHCb8WZ’16	CDHSW-F3’91	CMS8jets’17	BCDMSp’89	CDHSW-F2’91	CCFR-F3’97

Table 2.1: We list the top 10 experiments predicted to drive knowledge of the d/u PDF ratio and of the gluon distribution in the Higgs region according to PDFSense and LM scans. For both, we list the PDFSense evaluations based both on the CT14HERA2 fit and on a preliminary CT18pre fit in the first and second columns on either side of the double-line partition.

fit, cf. Eq. (2.21), we similarly divide $\max(\chi^2) - \min(\chi^2)$ by the best-fit $\chi^2(\vec{a}_0)/N_{pt}$ of the experiment in the rankings for the LM scans in Table 2.1.

From the side-by-side examination of the figures and the table, we can draw a broad conclusion that both the pre-fit PDFSense and post-fit LM scan approaches agree in identifying the most constraining experiments, even though they may result in different orderings of these experiments. This agreement is especially impressive in the instance of $d/u(x = 0.1, \mu = 1.3\ \text{GeV})$, when the rankings agree on 8 out of 10 leading experiments, confirming the dominance of the NMC p/d ratio, HERAI+II, CCFR F_3, and BCDMS p and d measurements. For $g(x\!=\!0.01, \mu\!=\!m_H)$, for which we see more tension and non-Gaussian behavior in Fig. 2.10, both PDFSense and LM scans concur on the crucial role played by the top 5-6 experiments, namely, HERAI+II, E866pp, and inclusive jet production data from CMS, ATLAS, and D0 Run-2. The upward pull on g from the incompatible ATL8ZpT data set seen in Fig. 2.10 modifies the rankings of the trailing experiments, such as CMS7 jets or BCDMS. Based upon an extended battery of LM scans we have performed, including the two examples presented here, we conclude that the `PDFSense` surveys perform as intended.

Lastly, we reiterate that a number of subtleties exists in comparing the results of LM scans and PDFSense sensitivity plots. Most importantly, PDFSense is intended by conception as a tool to quantify the anticipated *average* impact of potentially unfitted data based upon their precision in comparison to the PDF uncertainties. We discussed simplifying assumptions made in PDFSense in order to bypass certain complexities of the full fit and obtain quick estimates. LM scans, on the other hand, provide post-fit assessments of the contributions of specific data to the global χ^2 function, as specific quantities predicted by the QCD analysis are dialed away from their optimal values. In the comparisons we made, the detailed pictures produced by both PDFSense and the LM scans depend on a variety of theoretical settings like pQCD scale choices, as well as upon the specific implementation of correlated experimental uncertainties [from up to ∼100 different sources in some experiments] and the parametric forms chosen for the nonperturbative parametrizations at the starting scale $\mu = Q_0$. The inclusion of additional theory uncertainties and decorrelation of some experimental correlated errors are necessitated in a few experiments by the relatively large χ^2 values that would otherwise be obtained. All these have some peripheral effect on the specific orderings of experiments shown in Table 2.1. Thus, rather than anticipating an exact point-to-point matching between the PDFSense and LM methods, we instead expect, and indeed find, the general congruity between the most important experiments identified by the two approaches illustrated in this section.

2.5. Summary

In the foregoing analysis, we have confronted the modern challenge of a rapidly growing set of global QCD data with new statistical methodologies for quantifying and exploring the impact of this information. These novel methodologies are realized in a new analysis tool PDFSense [64], which allows the rapid exploration of the impact of both existing and potential data on PDF determinations, thus providing a means of weighing the impact of measurements of QCD processes in a way that allows meaningful conclusions to be drawn without the cost of a full global analysis. We expect this approach to guide future PDF fitting efforts by allowing fitters to examine the world's data *a priori,* so as to concentrate analyses on the highest impact datasets. In particular, this work builds upon the existing CT framework with its reliance on the Hessian formalism and assumed quasi-Gaussianity, but these features do not impact the validity of our analysis and conclusions. Our approach provides a means to carry out a detailed study of data residuals, for which we explored novel visualizations in several ways, including the PCA, t-SNE, and reciprocated distance approaches discussed in Sec. 2.2.3. These techniques show promise for moving forward by providing useful insights into the numerical relationships among datasets and experimental processes.

Crucial to this analysis is the leveraging of both the existing and proposed statistical measures laid out in Secs. 2.3.1 and 2.3.2. Of these, the flavor-specific sensitivity S_f of Eq. (2.21) for a data point to the PDF serves as a particularly powerful discriminator, and we deployed it and the correlation C_f of Eq. (2.19) to map PDF constraints provided by data over a wide range in $\{x, \mu\}$. This was facilitated by the fact that the sensitivity and correlation are readily computable over the extent of the global dataset. The companion website collects a large number of figures illustrating the sensitivities to various flavors as a function of x and μ.

To quantify the abundant information contained in the maps of sensitivities, in Sec. 2.4.2 we presented statistical estimators to systematically rank and assess subsidiary datasets within the world's data according to their potential to be influential in constraining PDFs. We note that one is allowed some freedom in choosing a specific ranking prescription, but we find our conclusions to be stable against variations among these possible choices. In this context, we reaffirmed the unique advantage of DIS and jet production for determination of the PDFs.

Many intriguing physics results can be established using our sensitivity methods, and the specific results in the previous sections are only illustrative examples. We stress that these results take the complementary form of sensitivity tables (for example, Table B.4) and $\{x, \mu\}$ plots (such as Fig. 2.2), which respectively offer global categorizations of the experimental landscape and detailed mappings of the placements of PDF constraints in $\{x, \mu\}$ space. In totality, the full range of physics insights from this method is beyond the scope of the present article, but the interested user can explore them using our PDFSense package at [64]. We mention only a representative sample of these to motivate the reader:

- A wide range of experimental processes possess sensitivity to the nucleon's quark sea distributions; for example, for the distribution $\overline{d}(x, \mu)$, the σ_{pd} DY measurements of E866 (E866rat'01) exhibit strong sensitivity, but so do DY data from E605 (E605'91) as well as (at larger μ) information on the μ-production asymmetry $A_\mu(\eta)$ from CMS at 7 TeV (CMS7Masy2'14); at high x and μ, CMS inclusive jet data (CMS8jets'17, CMS7jets'14) also acquire some sensitivity to $\bar{u}$ and $\bar{d}$. Still, however, the recent HERA data (HERAI+II'15) registers the greatest overall sensitivity.

- Were they taken cumulatively together as a single dataset, CMS jet production at 7 and 8 TeV (CMS7jets'14 and CMS8jets'17) would provide a total sensitivity $|S_s^E| = 11.9 + 8.11$ to $s(x, \mu)$ that is comparable to one of the NuTeV (NuTeV-nu'06) or CCFR (CCFR SI nu'01, CCFR SI nub'01) dimuon SIDIS experiments, which have very strong average sensitivity to the strange distribution. Still, the strongest constraint is contributed by a mix of the DIS measurements, including $\nu\mu\mu$ data from NuTeV (NuTeV-nu'06), data on $\nu(\overline{\nu})\mu\mu$ processes from SIDIS at CCFR (CCFR SI nu'01 and CCFR SI nub'01), as well as the inclusive DIS data at lower x from HERA1+2 (HERAI+II'15) that actually has the strongest cumulative sensitivity. Similarly, various vector boson production data sets have a rank-3 point-averaged sensitivity to the strangeness, including the $A_\mu(\eta^\mu)$ data from D0 (D02Masy'08) and CMS (CMS8Wasy'16, CMS7Masy2'14), as well ATLAS W/Z production (ATL8DY2D'16, ATL7WZ'12) and high-p_T Z production (ATL8ZpT'16) cross sections. Although each of the individual vector boson production data set has a weak cumulative sensitivity to $s(x, \mu)$ because of a small number of data points, in totality a group of *mutually consistent* LHC experiments on vector boson production can provide a competing constraint on $s(x, \mu)$ that confronts the low-energy CCFR/NuTeV constraints.

- Knowledge of the charm distribution $c(x, \mu)$ is most influenced by a number of datasets, with HERA (HERAI+II'15) at low x especially important. Fixed target measurements, particularly those of CDHSW on the proton's F_2^p structure function (CDHSW-F2'91) have strong sensitivity at slightly higher $x \sim 10^{-1}$, while a wide range of jet measurements, including 7 TeV data from ATLAS (ATLAS7jets'15) and CMS (CMS7jets'14), and 8 TeV CMS (CMS8jets'17) points are also sensitive. This pattern of sensitive measurements broadly follows the corresponding plot for $|S_g|(x_i, \mu_i)$ [as well as $|S_b|(x_i, \mu_i)$] due to the dominance of boson fusion graphs in heavy quark production. The datasets of importance we identify are broadly consistent with the conclusions of the recent CT14 analysis [74] of the nucleon's intrinsic charm [59].

- One can also study the correlations and sensitivities for various derived PDF combinations. For instance, for the $\overline{d}/\overline{u}$ ratio representing deviations from flavor symmetry in the nucleon sea, the E866 experiment (E866rat'01) shows exceptional point-averaged sensitivity, $\langle|S_{\bar{d}/\bar{u}}|\rangle = 1.67$ such that its "C" ranking for its overall sensitivity to $\bar{d}/\bar{u}$ places it in the company of only a few other DIS and DY experiments, despite their much larger number of measurements, $N_{pt} = 15$. At somewhat lower $x \gtrsim 0.01$, NMC data on the structure function ratio F_2^d/F_2^p (NMCrat'97) show sensitivity in the range $0.8 < |S_{\overline{d}/\overline{u}}| < 2$. At still lower x, the CMS 8 and 7 TeV A_μ points (CMS8Wasy'16, CMS7Masy2'14) and W/Z data from LHCb (LHCb8WZ'16) show strong pull, corresponding to point-averaged rankings of "2," "**1**," and "2," respectively.

- We also consider the PDF ratio $d/u(x, \mu)$, which often serves as a discriminant among various nucleon structure models, especially at high x. For $x > 0.1$ an amalgam of fixed-target experiments, including the NMC F_2^d/F_2^p data (NMCrat'97) particularly, but also F_2^p measurements from BCDMS (BCDMSp'89) and CCFR (CCFR-F2'01) as well as xF_3^p data from CCFR drive the current status. At higher μ, however, the LHCb W/Z data (LHCb8WZ'16) and $A_e(\eta)$ measurements from Run-2 of D0 (D02Easy2'15) also constrain the high x behavior of d/u together with $A_\mu(\eta)$ points from CMS at 7 TeV (CMS7Masy2'14).

- More generally, we note that, among the new LHC experiments to be considered for future global fits, the datasets for inclusive jet production are expected to have the greatest impact, followed by a group of vector boson production experiments at ATLAS, CMS, and LHCb. We find that the constraints from jet production at the LHC depend significantly on the treatment of experimental systematic uncertainties — especially the correlated systematic errors. It is conceivable that, with the full implementation of NNLO theoretical cross sections and modest reduction in the experimental systematic uncertainties, the constraints from the LHC jet production will catch up in strength to the effect of adding a large fixed-target DIS dataset, such as BCDMS F_2^p (BCDMSp'89). Meanwhile, the magnitude of the constraint on the gluon PDF from high-p_T Z production (ATL8ZpT'16) is comparable to those from the combined HERA SIDIS charm dataset (HERAc'13) or inclusive jet production from CDF Run-2 (CDF2jets'09); that is, the high-p_T Z data are significant in the event that the jet datasets are not included, in overall consistency with the findings in Ref. [51]. The smaller ATLAS $t\bar{t}$ production data sets (ATL8ttb-pt'16, ATL8ttb-y_ave'16, ATL8ttb-mtt'16, ATL8ttb-y_ttb'16) have strong point-by-point sensitivity to the gluon, but will have a more diminished role when combined with other, larger data sets. HERA DIS (HERAI+II'15), BCDMS F_2^d (BCDMSd'90), and CMS inclusive jets at 8 TeV

(CMS8jets'17) render the strongest overall constraints on the Higgs production cross section at the LHC according to the rankings in Table B.5.

Quantifying correlations and sensitivities thus provides a comprehensive means of evaluating the ability of a global dataset to constrain our knowledge of nucleon structure. It must be emphasized, however, that this analysis is not a substitute for actually performing a QCD global analysis, which remains the single most robust means of determining the nucleon PDFs themselves. Rather, the method presented in the paper is a guiding tool to both supplement and direct fits by gauging the potential for improving PDFs with the incorporation of new datasets.

The essential ingredients of this study are the PDF-residual correlation and sensitivity $|C_f|$ and $|S_f|$, with the latter representing an extension of the correlation used elsewhere in the modern PDF literature. These definitions are robust enough that we can exhaustively score the data points in an arbitrary global dataset to construct and map the resulting distributions, as shown in Figs. 2.7 and 2.8. Accordingly, we found it possible to impose cuts on these distributions to identify points of especially strong correlation ($|C_f|> 0.7$) or sensitivity ($|S_f|> 0.25$); we stress that these cuts are chosen as approximate indicators, and any user can adjust them freely. On the other hand, the distributions themselves, as shown in the second panels of Figs. 2.7 and 2.8, are not subject to such cut choices. Although the conclusions of this analysis are resistant to alterations in the basic approach, it is worth noting that other formats are possible for evaluating experimental sensitivities and performing the rankings of measurements. For example, one might use somewhat different matchings than those outlined in App. A to extract $\{x, \mu\}$ points from the experimental data, but we expect the resulting impact on the overall picture to be minor. Similarly, while the ordering inside ranking tables like Table B.4 was decided according to the total sensitivity to serve our specific goal of identifying the most valuable experiments for the CTEQ-TEA fit, for other purposes one might produce alternative tables ranked according to point-averaged sensitivities, or sensitivities to specific flavors. Such alternate conventions would also yield important information, and PDFSense allows the user to do this. It should be stressed that these elections for the form of our presentation can always be recovered from the more fundamental information — the numerical values of the sensitivities detailed in the Supplementary Material of App. B.2.

While we have demonstrated these techniques in the context of the CT14 family of global fits, they are of sufficient generality that one could readily repeat our analysis using alternative PDF sets. For the sake of testing this point and validating our predictions for the most decisive experiments in the CTEQ-TEA dataset, we performed a preliminary fit including the CT14HERA2 and the candidate LHC experiments ('CT18pre'), and directly compared PDFSense predictions against Lagrange multiplier scans quantifying the constraints these fitted measurements imposed on select quantities. This provided a demonstration of the robustness of our sensitivity-based analysis, which identified the same sets of high-impact measurements *before fitting*. The results of this study can be expected to vary somewhat depending on the specifics of the PDF sets used to compute $|C_f|$ and $|S_f|$, but we see this as an advantage of PDFSense. One could imagine exploiting them to undertake a systematic analysis of the impact of various theoretical assumptions implemented in competing global fits (*e.g.*, the choice of input PDF parametrization or the status of the perturbative QCD treatment implemented in various processes). The sensitivity S_f can be constructed either from the Hessian or Monte-Carlo PDF uncertainties, as prescribed by Eqs. (2.21) and (2.29), while the shifted residuals that are crucial to our analysis can be recovered from any type of covariance matrix, as argued in relation to Eq. (2.8). In the same spirit but on the side of the data, PDFSense empowers the user to evaluate the combined impact of multiple experimental datasets — for example, to evaluate the extent to which the impact of a proposed experiment might be diminished by the constraints already imposed by existing measurements. These various functions collectively suggest a number of possible avenues to use the presented approach and the PDFSense tool to advance PDF knowledge in the coming years.

Chapter 3

The Coming Synergy Between Lattice QCD and High-Energy Phenomenology

3.1. Introduction

Owing to steady theoretical progress and the growing availability of computational resources, the ability of perturbative QCD (pQCD) to predict parton-level processes at high energies has continued to improve in recent years, with accuracies now reaching next-to-next-to-leading order (N^{2}LO) in many circumstances. Inevitably, however, predictions for experiments involving hadronic collisions or final states require precise knowledge of QCD bound state structure at comparatively small energy scales similar to the nucleon mass, $\Lambda \sim M$, at which $\alpha_s(\Lambda) \sim 1$ is too large to permit a converging diagrammatic expansion of the relevant amplitudes. This general consequence of the negative β-function of QCD is realized in the theory of spin-averaged deeply-inelastic lepton-nucleon scattering, for example, in the factorization of physical cross sections into perturbatively calculable short-distance matrix elements and inherently nonperturbative long-distance parton distribution functions (PDFs), $q(x, \mu)$, of the quark-to-hadron light-front momentum fraction $x = k^+/p^+$ and factorization scale μ.

Given the nonperturbative nature of the latter long-distance parton distribution functions (PDFs), the prevailing recourse has traditionally been either to fit them in comprehensive analyses of global data using flexible parametric forms [9, 35, 36, 38, 41], or to calculate them in the context of models or effective theories [58, 59, 75–82] that aim to capture specific aspects of QCD — *e.g.*, its pattern of dynamical chiral symmetry breaking. Parallel to these efforts, the past couple of decades have seen a complementary effort founded in the use of lattice gauge theory techniques to either indirectly compute the x dependence of the PDFs themselves, or, at minimum, determine the integrated moments of the parton distributions in Mellin space. (For a comprehensive review, we refer the reader to the recent white paper, Ref. [19].)

By definition, the PDFs are intrinsically nonlocal correlation functions constructed between parton fields with lightlike spacetime separation — *viz.* $\sim \langle p|\bar{q}(x^+)\hat{\mathcal{O}}q(0)|p\rangle$; however, dynamically simulating such matrix elements on a hypercubic lattice is numerically problematic, given the fact that $x^2 = x^+x^- - x_\perp^2 = 0$ can only trivially hold at the origin in a Euclidean spacetime, for which $x_\mathrm{E}^2 = x_1^2 + \cdots + x_4^2$. In contrast, the integrated *Mellin moments* of the quark distributions have a direct interpretation in terms of the matrix elements of local operators and can be accessed on a Euclidean lattice via an operator product expansion (OPE). Moments computed in this fashion are informative in the sense that they encapsulate aspects of the nonperturbative dynamics responsible for a hadron's low energy structure — for instance, the magnitude of the nucleon's collinear magnitude carried by its total u-quark content,

$$\langle x\rangle_{u^+} = \int_0^1 dx\, x[u+\overline{u}](x,\mu_F)\ . \tag{3.1}$$

Lattice calculations generally evaluate moments like Eq. (3.1) using a scheme and renormalization scale μ^lat chosen to match the $\overline{\mathrm{MS}}$ scheme usually employed by phenomenologists. Most often in the literature, this scale is taken to be $\mu^\mathrm{lat} = 2$ GeV, and in this analysis we shall for consistency compute moments at a matching factorization scale, $\mu_F = \mu^\mathrm{lat} = 2$ GeV, unless otherwise indicated. Various attempts have been made to determine the x dependence of the PDFs by computing a sufficient number of moments in Mellin space that the transform into PDF space can be determined (typically with the help of some parametrization). In practice, however, the mixing among operators of successively higher spin and the resulting signal-to-noise issues become less controlled as additional covariant derivatives are inserted to obtain PDF moments of higher order. In effect, only a small number of moments can be accessed on the lattice — presently, up to the quark distributions' third moment, $\langle x^3\rangle_{q^+}$

(although there are recent suggestions that perhaps several more may become available in the near future). It should be noted that the uncertainties of the lattice moments typically grow with increasing order.

Still, the ostensible ability of lattice gauge theory to access even several moments of the PDFs has long presented the possibility of determining (or at least constraining) the parton distributions directly from a first-principles QCD calculation. Indeed, with a sufficiently restrictive parametric form for the quark distribution of a given flavor, the latter can be fully determined given enough moments [21]; for example, if the PDF of the u-quark distribution is taken to have a very simple x dependence given by $u(x, Q_0) = \alpha\, x^{\beta}\, (1-x)^{\gamma}$, knowledge of 3 distinct moments would in principle be adequate to parametrically determine (up to some uncertainty) the above-noted distribution. At the same time, however, both the diversity of the experimental data inputs and sophistication of modern QCD analyses are such that much more flexible parametric forms are required, and lattice calculations remain far below the requisite level of precision across the many flavors and moment orders needed to be competitive in a complete determination of the PDFs according to such a procedure.

More recently, a promising method which may allow the calculation of the PDFs' x dependence on the QCD lattice in terms of parton *quasi-distribution* functions (qPDFs) has been introduced by Ji [23]. Extracting information from quasi-distributions requires an accompanying large momentum effective theory (LaMET) for performing the necessary ultraviolet matchings that are realized as convolutional relations of the form

$$\widetilde{q}(x, P_z, \widetilde{\mu}) \;=\; \int dy\, Z\left(\frac{x}{y}, \frac{\Lambda}{P_z}, \frac{\mu}{P_z}\right) q(y, \mu) \;+\; \mathcal{O}\left(\frac{\Lambda^2}{P_z^2}\, , \frac{M^2}{P_z^2}\right)\;\; , \tag{3.2}$$

which relate the quasi-distribution $\widetilde{q}$ to the traditional phenomenological PDF q with the usual support over $x \in [0, 1]$; this matching depends critically upon the pQCD-calculable ultraviolet matching function, Z. In practice, the quasi-distribution $\widetilde{q}(x, P_z, \widetilde{\mu})$ [*i.e.*, the left-hand side of Eq. (3.2)] may be evaluated on the lattice for a specific choice of the longitudinal hadron momentum P_z, and the usual PDF extracted by numerical inversion of Eq. (3.2). This method therefore has the potential to yield information on the x dependence of the PDFs themselves, up to knowledge of dynamical and mass-dependent corrections, the perturbative order of the matching kernel Z, and technical details of the actual lattice calculation — for instance, artifacts arising from the finite lattice spacing or signal-to-noise problems. In addition, it should be pointed out that limitations to this procedure remain, especially given the fact that the lattice calculations and LaMET procedure are in a relatively early stage of theoretical development — much as there are limitations to the lattice computed PDF moments as well.

For the reasons noted above, as computational resources continue to grow, it will be necessary to reconcile the output of lattice-based methods (especially, concerning the PDF moments and quasi-distributions) with work in the context of QCD global analyses. This will necessarily go both directions: benchmarking the lattice calculations with knowledge of the PDFs from phenomenological analyses, and constraining QCD analyses with the output of the lattice. Laying the groundwork for this synergy will require a comprehensive understanding of the relation between phenomenological constraints placed on the PDFs determined in fits (and, by extension, the x-weighted moments computed therefrom) and information obtained from the lattice.

In this analysis we systematically canvass these issues, using the recently developed `PDFSense` framework of chapter 2 (or Ref. [34]) to present a comprehensive summary of the differential impact modern data have upon knowledge of the PDF Mellin moments evaluated from phenomenological fits — as well as which data give the leading contributions to present understanding of one of the typical quasi-distributions, that of the isovector combination, $u-d$. The remainder of the paper is therefore as follows: after a brief review of the formalism and a description of the `PDFSense` methodology, in Sec. 3.2 we review the constraints HEP data place on the lowest moments for several light quark $q^{\pm}$ distributions and the gluon; in Sec. 3.3 we illustrate the constraints data place on the $u{-}d$ quasi-distributions at several choices of the momentum fraction, $\pm x$, and hadronic boost momentum, P_z, while Sec. 3.4 demonstrates the sizable potential impact future measurements at a high-luminosity lepton-hadron collider will have on these quantities. Sec. 3.5 contains a number of conclusions drawn from our analysis of the PDF moments and qPDFs regarding expected consequences of implementing lattice information into future global fits. In Sec. 3.6 we provide a number of closing observations, focusing on points that will allow this work to be leveraged in the future. Lastly, Appendix C collects several tables — counterpart to Figs. 3.7 and 3.8 — summarizing the aggregated impact on lattice QCD observables of the HEP experiments considered in this work.

3.2. The Sensitivity of HEP Data to PDF Mellin Moments

3.2.1. Theory of PDF Mellin Moments

The x-weighted moments of the PDFs have long been of interest to practitioners of QCD analyses on the logic that they may provide the necessary input to either reconstruct or at least constrain the PDFs determined in global fits. The accessibility of these moments to lattice gauge techniques is facilitated by the OPE [83–88], which allows an expansion of the PDF moments in terms of matrix elements of well-defined, local twist-2 operators which can be evaluated in a discretized Euclidean spacetime. Subsequently, the Mellin moments themselves may be derived via algebraic relations from the matrix elements of these twist-2 operators. In principle, it is possible to reconstruct a given PDF's x dependence via an inverse Mellin transform *if* its moments $\langle x^n \rangle_q$ are known to all orders, as noted in Sec. 3.1.

The crucial relation that connects x-dependent parton distributions $q(x)$ to an n-dependent tower of integrals in Mellin space is the inverse Mellin transform, which enables one in principle to reconstruct integrated Mellin moments of the PDFs. These PDF Mellin moments have the general form

$$\langle x^n \rangle_q = \int_0^1 dx\ x^n \left[q(x) + (-1)^{n+1}\, \overline{q}(x) \right] . \tag{3.3}$$

Using Eq. (3.3), it is possible to define a collection of PDF moments $\langle x^n \rangle_{q^\pm}$ which are actually calculable on the QCD lattice, such as $\langle x \rangle_u^+$ of Eq. (3.1). These are

$$\langle x^n \rangle_{q^+} = \langle x^n \rangle_q \quad \text{for} \quad n = 2\ell - 1 \tag{3.4a}$$

$$\langle x^n \rangle_{q^-} = \langle x^n \rangle_q \quad \text{for} \quad n = 2\ell \tag{3.4b}$$

where $\ell \in \mathbb{Z}^+$ such that the lattice may provide, for instance, $\langle x \rangle_{u^+}$, $\langle x^2 \rangle_{u^-}$, $\langle x^3 \rangle_{u^+}$, *etc.* Moreover, the PDFs themselves can be unfolded from a complete set of Mellin moments via the inverse Mellin transform,

$$q(x) + (-1)^{n+1}\, \overline{q}(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dn\, x^{-n-1} \langle x^n \rangle_q \ . \tag{3.5}$$

In practice, however, lattice QCD techniques currently enable the calculation of the few lowest PDF moments. Once the accuracy of these computations improves, theoretical constraints in the form of lattice-calculated PDF moments (or x-dependent information unfolded from parton quasi-distributions, discussed in Sec. 3.3) might eventually be implemented as χ^2 penalties in global QCD analyses — essentially, taking lattice data as theoretical priors to constrain the likelihood function of a global fit. For example, exploratory studies based on Bayesian profiling have suggested that lattice calculations even with somewhat sizable uncertainties can still provide powerful constraints to PDFs in regions that are relatively unconstrained by experimental data (see Ref. [19] and references therein). In the remainder of this subsection, we review the essential theory for accessing the integrated moments of the PDFs, with a special focus on unpolarized distributions, given their importance to high-energy phenomenology. (Although it is worth noting that, given the comparative paucity of spin-dependent data, it is reasonable to expect that lattice calculations for the moments of helicity distributions may more quickly become competitive against fitted spin-dependent PDFs; we defer such considerations, however, to future work.)

The OPE expands hadronic matrix elements of non-local products of field operators in terms of local operator matrix elements weighted by Wilson coefficients that obey renormalization group evolution [84, 89]. It is then possible to calculate these coefficients in the context of QCD perturbatively (*i.e.*, they embody the relevant short-distance dynamics), while the local operator matrix elements are nonperturbative (depending on the details of the long-distance physics). In the case relevant for the present study, the matrix elements of twist-2 operators may be expanded by the OPE as

$$\frac{1}{2}\sum_{s}\langle p,s|\mathcal{O}^{q}_{\{\mu_1,\cdots,\mu_{n+1}\}}|p,s\rangle = 2v_q^{n+1}\left[p_{\mu_1}\cdots p_{\mu_{n+1}} - \text{traces}\right] , \tag{3.6}$$

$$\langle p,s|\mathcal{O}^{5\,q}_{\{\sigma\mu_1,\cdots,\mu_{n+1}\}}|p,s\rangle = \frac{1}{n+2}a_q^{n+1}\left[s_\sigma p_{\mu_1}\cdots p_{\mu_{n+1}} - \text{traces}\right] , \tag{3.7}$$

where p and s represent the nucleon 4-momentum and spin, respectively, q indicates the flavor of the relevant quark field, and $\{\}$ stands for index symmetrization. Higher-twist terms enter as power suppressed corrections in $1/Q^2$ [90], so here we only consider the contribution of Mellin moments from twist-2 operators. For the quark fields, the twist-2 operators occurring in the OPE expressions above are constructed from the usual bilinears as

$$\mathcal{O}^{q}_{\{\mu_1,\cdots,\mu_{n+1}\}} = \left(\frac{i}{2}\right)^{n}\bar{q}(x)\gamma_{\mu_1}\overleftrightarrow{D}_{\mu_2}\ldots\overleftrightarrow{D}_{\mu_{n+1}}q(x) \tag{3.8}$$

and

$$\mathcal{O}^{5\,q}_{\{\sigma\mu_1,\cdots,\mu_{n+1}\}} = \left(\frac{i}{2}\right)^{n+1}\bar{q}(x)\gamma_\sigma\gamma_5\overleftrightarrow{D}_{\mu_1}\ldots\overleftrightarrow{D}_{\mu_{n+1}}q(x) \tag{3.9}$$

where

$$\overleftrightarrow{\mathcal{D}}_\mu = \frac{1}{2}\left(\overrightarrow{\mathcal{D}}_\mu - \overleftarrow{\mathcal{D}}_\mu\right), \qquad \overrightarrow{\mathcal{D}}_\mu = \overrightarrow{\partial}_\mu - ig\, t_a A^a_\mu(z), \qquad \overleftarrow{\mathcal{D}}_\mu = \overleftarrow{\partial}_\mu + ig\, t_a A^a_\mu(z) \tag{3.10}$$

are the gauge covariant derivatives, $A^a_\mu(z)$ denotes gluon fields, and t_a represents the 8 standard generators of the SU(3) color group. In Eqs. (3.6) and (3.7), v_q^{n+1} and a_q^n are identifiable with the n^{th}-order moments of the twist-2 PDFs of unpolarized and longitudinally polarized nucleons, respectively [19, 55]:

$$v_q^{n+1} = \int_0^1 dx\, x^n\, q(x,\, \mu)\ , \tag{3.11a}$$

$$a_q^n = \int_0^1 dx\, x^n\, \Delta q(x,\, \mu)\ . \tag{3.11b}$$

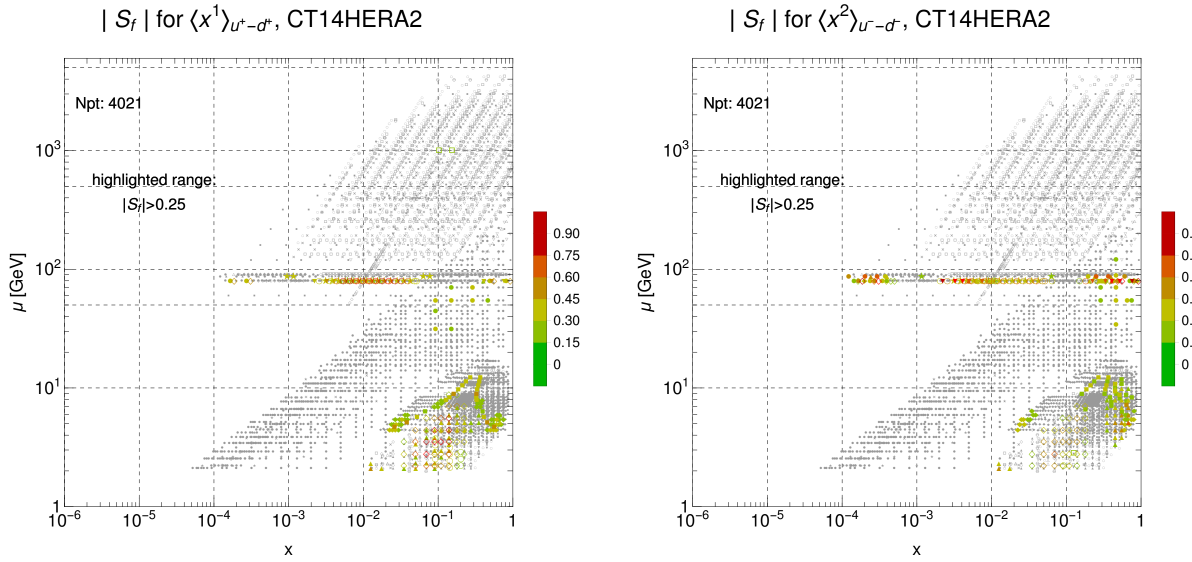
Lastly, we note that analogous matrix elements are responsible for moments of the gluon distribution, with the lowest lattice-accessible moment $\langle x\rangle_g$ for the total gluon momentum fraction given by the insertion of a twist-2 operator constructed from the gluon field strength as $\mathcal{O}^g_{\mu_1\mu_2} = -\mathrm{Tr}(\mathcal{G}_{\mu_1\alpha}\mathcal{G}_{\mu_2\alpha})$.

Insofar as the LHS expressions of Eqs. (3.6) and (3.7) can be formulated in terms of lattice gauge theory and evaluated on a discretized Euclidean spacetime, the PDF moments in Eq. (3.11) are themselves directly accessible on the QCD lattice through the direct evaluation of nucleonic matrix elements of twist-2 operators. For reasons that we sketch below, however, the extraction of higher moments is complicated by operator-mixing effects, and modern lattice calculations can reliably extract only the lowest few moments in practice [20–22]. Various systematic errors in generic lattice calculations are reviewed and assessed in Ref. [91], with the dominant systematic errors in evaluations of Mellin moments arising from power-divergent operator mixing and renormalization effects. Power-divergent mixing is associated with an $O(4)$ symmetry breaking inherent to the Euclidean spacetime discretization of lattice calculations: the lattice regulator breaks Lorentz symmetry, causing radiative divergences in operators of different mass dimensions mix together [92]. The renormalization of nonlocal operators on a discretized lattice induces another error: the renormalized fields that are nonperturbatively determined on the lattice are power-divergent. In addition to these, a number of other systematic effects generally enter lattice QCD calculations, including corrections from the unphysically large quark (or pion) masses often used as well as the associated chiral extrapolations to the physical pion mass. Moreover, details involved in the extraction of lattice signals as a function of lattice time contribute to the landscape of systematic uncertainties. The effort to control these and other lattice artifacts partially depends upon the ability of lattice practitioners to benchmark their calculations against phenomenological knowledge. Exploring the capacity of high-energy data to tighten these benchmarks is a primary motivation for the present study.

3.2.2. Analysis Procedure

To explore the sensitivity of high-energy data to the PDF Mellin moments and qPDFs accessible in lattice QCD, in this work we extend the analysis of chapter 2 (or Ref. [34]) that applied the recently developed PDFSense framework to weigh the impact of an extended amalgam of HEP experimental data under consideration for the upcoming CT18 PDF global analysis (the ‘CTEQ-TEA’ data, plotted in Fig. 2.1). In this case, special emphasis was placed on the impact these data might have on the unpolarized collinear PDFs themselves and on observables derived directly therefrom, including the 14 TeV inclusive Higgs production cross section, σ_H. Given the fact that a number of lattice QCD observables are calculable from phenomenological PDFs along the lines described in Secs. 3.1, 3.2.1, and 3.3 below, we repeat our analysis to illustrate the constraints a typical experimental data set can impose on our phenomenological knowledge of such lattice observables.

We refer the reader to Secs. II and III of chapter 2 (or Ref. [34]) for a systematic presentation of the details of the PDFSense framework. Still, it is worthwhile to summarize


| S_f | for $\langle x^1\rangle_{u^+-d^+}$, CT14HERA2

| S_f | for $\langle x^2\rangle_{u^--d^-}$, CT14HERA2

Figure 3.1: Sensitivity of the CTEQ-TEA data sets to $\langle x\rangle_{u^+-d^+}$ and $\langle x^2\rangle_{u^--d^-}$. The factorization scales of Mellin moments and PDFs are $\mu_F = \mu^{\rm lat} = 2$ GeV.

the particulars of a sensitivity analysis dedicated to the PDF moments $\langle x^n\rangle_{q^\pm}$. Whereas in chapter 2 (or Ref. [34]) we primarily concentrated on the sensitivities of data to the local values of the collinear PDFs $q(x_i, \mu_i)$ at the typical x_i and μ_i of the high-energy data points (see Appendix A), here we are chiefly concerned with the sensitivity to Mellin moments for which the x dependence has been integrated away, and in general at a scale $\mu_F = \mu^{\rm lat} = 2$ GeV at which moments are typically computed in lattice QCD. Whether by a Hessian or Monte Carlo error procedure, a PDF global analysis typically produces a central PDF set and a finite ensemble replicas of the error PDFs, $q^{j\in\{2N\}}(x, \mu_F)$. Given this ensemble, it is then possible to evaluate a respective error set for values of integrated PDF moments. In the case of the CT fitting approach, from the underlying Hessian error sets — of which there are $2N$ for an N-parameter global fit [leading to 1(56) central (error) PDFs in the 28-dimensional CT14HERA2 NNLO fit] — we directly compute error replicas for the moments by integrating over the CT fitted distributions. Namely,

$$q^{j\in\{2N\}}(x, \mu^{\rm lat}) \;\longrightarrow \langle x^n\rangle^{j\in\{2N\}}_{q^\pm,\,\mu^{\rm lat}} = \int_0^1 dx\, x^n \left(q(x, \mu^{\rm lat}) \pm \overline{q}(x, \mu^{\rm lat})\right)_{j\in\{2N\}} . \tag{3.12}$$

With replica sets for lattice observables like the PDF Mellin moments as in Eq. (3.12), we may deploy the statistical framework of chapter 2 (or Ref. [34]), computing the Pearson correlation between the residual $r_i(\vec{a})$ of the i^{th} measurement of our CTEQ-TEA set (again, evaluated over the 1[56] central [error] sets of CT14HERA2 NNLO) and the corresponding ensemble for $\langle x^n \rangle^{j\in\{2N\}}_{q^\pm,\, \mu^{\rm lat}}$. In practice, the correlation is computed using

$$\begin{aligned} C_f(x_i, \mu_i) &= \mathrm{Corr}[f, r_i(x_i, \mu_i)] \ , \\ \mathrm{Corr}[X, Y] &= \frac{1}{4\Delta X \Delta Y} \sum_{l=1}^{N} (X_l^+ - X_l^-)(Y_l^+ - Y_l^-) \ , \end{aligned} \tag{3.13}$$

in which f is a generic function of the PDFs (*e.g.*, a PDF of given flavor at the matched (x_i, μ_i) of the i^{th} data point as in chapter 2 (or Ref. [34]), or a PDF moment computed from the PDFs), and the N $l^\pm$ pairs of Eq. (3.13) may be identified with the $2N$ Hessian error sets just described; the uncertainty quantities in the denominator of Eq. (3.13) are evaluated from the Hessian error sets as

$$\Delta X = \frac{1}{2} \sqrt{\sum_{l=1}^{N} \left(X_l^+ - X_l^- \right)^2} \ . \tag{3.14}$$

One of the principal results of chapter 2 (or Ref. [34]) was the demonstration that the Pearson correlation of Eq. (3.13) cannot fully capture the phenomenological weight of individual measurements, given the fact that it does not explicitly depend upon the *magnitudes* of the PDF or experimental uncertainties. For this reason, we introduced a generalization of the correlation we call the *sensitivity*, S_f, of the i^{th} point in experiment E to PDF flavor (or PDF-derived quantity) f:

$$S_f \ = \ \frac{\Delta r_i}{\langle r_0 \rangle_E} C_f \ , \tag{3.15}$$

in which the residual of the i^{th} data point, $r_i = (T_i - D_i^{sh})/s_i$, is the difference between the theoretical prediction T_i and data value D_i^{sh} shifted to accommodate systematic error correlations; this difference is weighted by the total uncorrelated uncertainty, s_i. Δr_i is calculated using Eq. (3.14) and $\langle r_0 \rangle_E$ represents the point-averaged residuals of each of the points of experiment E computed with the central PDF set.

With the family of PDF error sets of CT14HERA2 NNLO and the statistical formalism and metric embodied by Eq. (3.15), the sensitivities of data points shown in Fig. 2.1 may be assessed and mapped in the plane of the typical momentum fraction x_i and factorization scale μ_i. We may thereby isolate physical processes and individual data sets with high sensitivity to lattice observables, but also understand the most constraining regions of (x, μ). The size of the absolute sensitivities $|S_{\langle x^n \rangle_{q^\pm}}|$ for highlighted points ($|S_{\langle x^n \rangle_{q^\pm}}|> 0.25$) are identified by the "rainbow stripe" color palette: hot colors (red, orange) correspond to strong sensitivities, and cool colors (yellow, green) correspond to weak sensitivities. Unhighlighted points — *i.e.*, those with relatively minimal expected impact, ($|S_{\langle x^n \rangle_{q^\pm}}|< 0.25$), are represented with gray colors.

We will compute and investigate the sensitivity of the CTEQ-TEA data set to the lowest moments of the unpolarized light quark and gluon distributions, $|S_{\langle x^n \rangle_{q^\pm,g}}|$, in the (x, μ) plane with the `PDFSense` package introduced in chapter 2 (or Ref. [34]) We use the CT14HERA2 NNLO PDF set [41] in the theoretical predictions and residuals of experimental data. Our data sets measurements in the CT14HERA2 fit and the latest LHC jet, $t\bar{t}$, W/Z production data sets. In addition, due to the fact that the scientific program of an EIC or EIC-like machine is anticipated to significantly complement lattice QCD, we also include a preliminary impact study in Sec. 3.4 for a set of EIC-like pseudodata for inclusive neutral-current (NC) and charge-current (CC)-mediated $e^\pm p$ deeply inelastic scattering (DIS). The procedure in this case is broadly similar to that for the CTEQ-TEA data, but in this case based upon pseudodata obtained by generating Gaussian fluctuations about the CT14HERA2 NNLO theoretical prediction for the reduced cross sections according to an assumed precision.

The flavor combinations primarily discussed in this paper are $u^\pm - d^\pm, u^\pm, d^\pm, s^+$ and g in which q^+ here refers to the C-even combination of (anti-)quark distributions, $q + \bar{q}$;

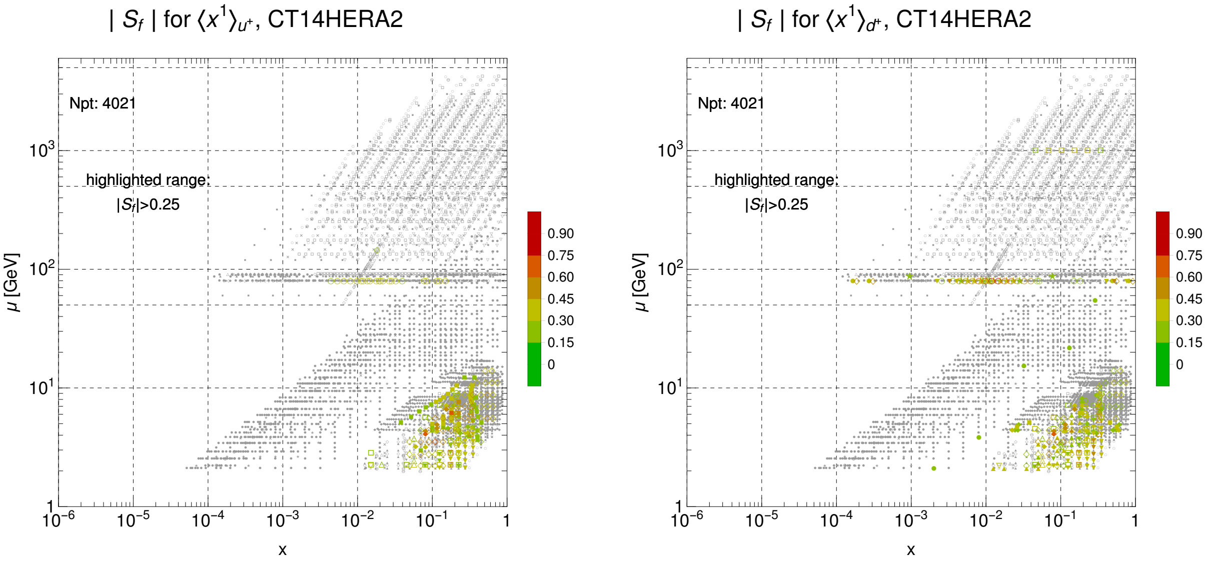


Figure 3.2: Sensitivity of the CTEQ-TEA data sets to $\langle x\rangle_{u^+}$ (left panel) and $\langle x\rangle_{d^+}$ (right panel). As in Fig. 3.1, sensitivities are with respect to moments evaluated at $\mu_F = \mu^{\text{lat}} = 2$ GeV.

correspondingly, we also consider C-odd quantities, $q^- = q - \bar{q}$, as defined in Eqs. (3.3)-(3.4). The PDF moment(s) for the light flavor combinations $u^\pm - d^\pm$, $u^\pm$, $d^\pm$, and s^+ are computed on the lattice by the $\mathcal{O}^F_{\{\mu_1,\cdots,\mu_n\}}$ operators noted above in Eq. (3.8), whereas for the gluon distribution g, the operator noted immediately after Eq. (3.11) is required. The present status of the lattice QCD calculations of these parton moments is widely varied, with some moments (especially for the isovector combination $u-d$) evaluated by multiple groups with various systematic treatments; on the other hand, lattice information on the second moments of the individual light quark flavors $\langle x^2\rangle_{q^-}$, for instance, is comparatively sparse. At the same time, the corpus of lattice computations is growing with enormous rapidity, and the availability of calculations for the moments considered here (and beyond) will quickly increase.

We note that many numerical results for PDF moments computed both on the lattice and from different QCD global analyses are detailed in Appendices B and C of the recent review in Ref. [19].

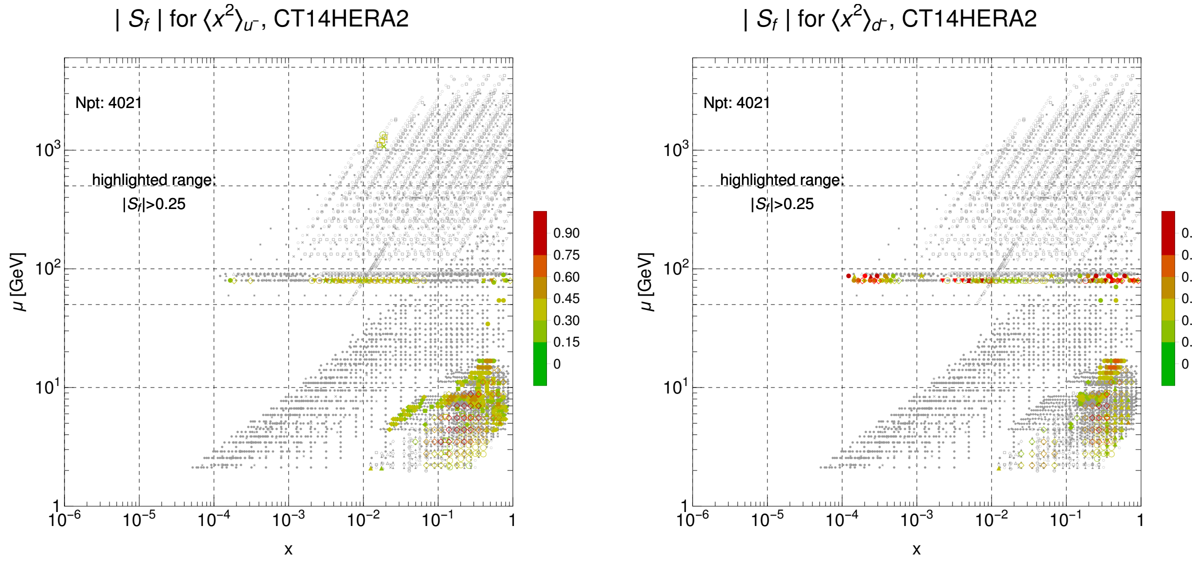


Figure 3.3: Sensitivity of the CTEQ-TEA data to the $v^3_{u,d}$ moments $\langle x^2 \rangle_{u^-}$ (left) and $\langle x^2 \rangle_{d^-}$ (right), computed for $\mu_F = 2$ GeV.

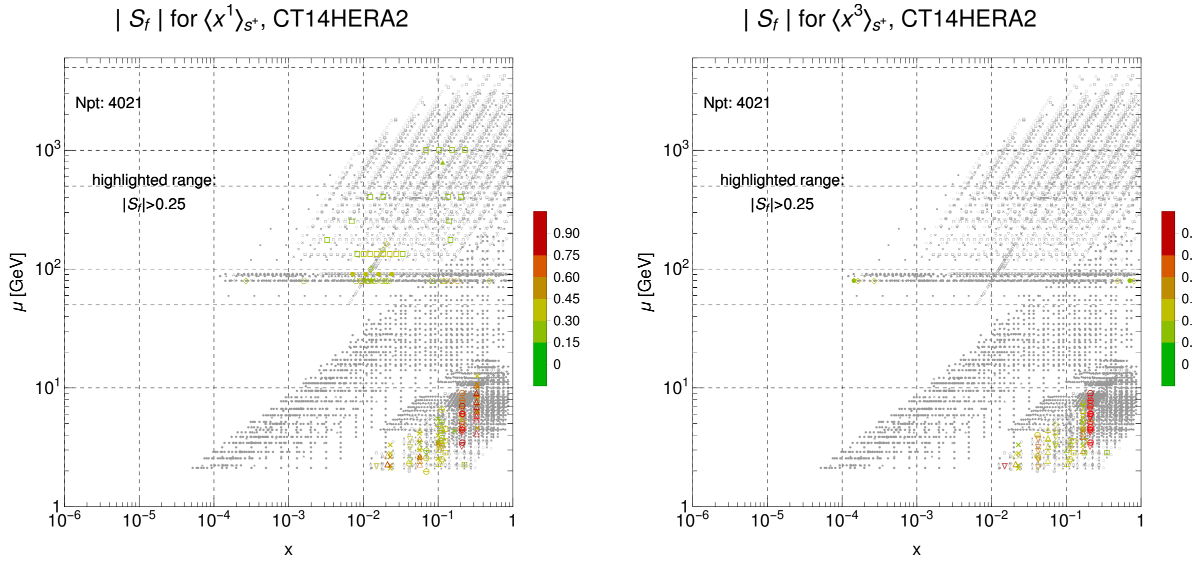


Figure 3.4: CTEQ-TEA sensitivity to the 1^{st}- and 3^{rd}-order Mellin moments of the s^+ distribution. As elsewhere, the factorization scales of Mellin moments and PDFs are $\mu_F = 2$ GeV.

3.2.3. Numerical Results

We now present the calculated sensitivity maps for the CTEQ-TEA data to each of the lowest moments of the light quark (u–d, u, d, s) distributions and the gluon. We also examine the aggregated impact of the experiments in the CTEQ-TEA set on each of these quantities and consider the implications for unraveling the nucleon's flavor structure and benchmarking lattice QCD output of the same objects.

3.2.3.1. Moments of Nucleon Quark Distributions

3.2.3.1.1 Moments of Isovector Flavor Distributions. Historically, computation of isovector PDF combinations in SU(2) isospin space has represented an especially fertile proving ground for lattice gauge methods — particularly given that gluon and singlet quark densities mix evenly with u^+ and d^+ distributions under DGLAP evolution, such that $u^+ - d^+$ has a nonsinglet scale dependence. A consequence specific to lattice QCD is the fact that contributions from disconnected insertions vanish in the difference (assuming parton-level charge symmetry), and a much less computationally costly calculation based purely on connected insertions is generally adequate. For this main reason, the isovector combinations have been a focus of lattice calculations for both the PDF Mellin moments — and, more recently, the quasi-distributions in both the nucleon [93–98] and pion [99–102].

We plot the sensitivity map of the CTEQ-TEA data to two of the lower moments of the nucleon isovector distribution in Fig. 3.1 — namely, the sensitivities to the total isovector momentum $\langle x \rangle_{u^+ - d^+}$ (left panel) and the second-order asymmetry moment $\langle x^2 \rangle_{u^- - d^-}$ (right panel). These plots have the same basic configuration as developed in chapter 2 (or Ref. [34]), with emphasis placed on specific measurements of individual experimental data sets with particularly strong pulls in the global analysis. The predicted pull of these measurements as quantified by the sensitivity $|S_f|$ is represented by the color scheme shown in the offset to the right for each panel in Fig. 3.1. As in chapter 2 (or Ref. [34]), we draw attention to the most impactful data and physical processes by imposing a highlighting cut $|S_f|> 0.25$, and selecting a coloration scheme which suitably reveals the dependence of data sensitivities on the kinematical matching parameters.

The sensitivity maps of Fig. 3.1 and throughout this analysis are complemented by the information contained in the companion plots shown in Figs. 3.7 and 3.8, which essentially integrate the information displayed in the sensitivity maps like Fig. 3.1 experiment-by-experiment, thereby summarizing the point-averaged sensitivities $\langle|S_f|\rangle$ for each CTEQ-TEA experiment, as well as the corresponding aggregated sensitivities $\sum_{i\in N_{pt}}|S_f^i|$. These companion plots encapsulate the related information summarized in Tables B.1–B.3 of Appendix B.1. It is worth noting that there is often a closer correspondence between experiments highlighted in sensitivity maps like Fig. 3.1 and those identified in the grid plots for the point-averaged sensitivities $\langle|S_f|\rangle$, *i.e.*, the upper panels of Figs. 3.7 and 3.8. For this reason, in discussing our numerical results as illustrated by the following sensitivity maps, we summarize the highest impact experiments according to complementary considerations of those data sets that enjoy sizable per-datum sensitivities to the PDF moments, and those that may not in general possess high-impact points taken in isolation, but are nonetheless predicted to have a large aggregated impact — often by merit of the large number of experimental data points N_{pt} they have. On the basis of these complementary assessments, we are in a position to make a number of observations regarding the empirical information that drives the current knowledge of the lower $\langle x^n\rangle_{u-d}$ Mellin moments.

As noted above, the highlighted points emphasized in the panels of Fig. 3.1 often rather closely correspond to the CTEQ-TEA experiments which enjoy the highest per-datum sensitivities; in decreasing order, these are CMS7Masy2'14 (0.557), E866rat'01 (0.365), CMS7Easy'12 (0.333), CCFR-F3'97 (0.307), and NMCrat'97 (0.212), where the quantity in parentheses is the computed average of each experiment *per measured point*. For the purpose of enumerating this information, we include only those experimental measurements with point-averaged sensitivities exceeding lower bound $\langle|S_f|\rangle > 0.2$. On the other hand, by the total sensitivity metric $\sum_{i\in N_{pt}}|S_f^i|$, we identify a somewhat different collection of experiments with leading impact on the first isovector moment; *viz.* HERAI+II'15 (37.8); CCFR-F3'97 (26.4); NMCrat'97 (26.1); E866pp'03 (20.7); BCDMSp'89 (19.3); In this context, there is a pronounced influence of the combined HERAI+II experiment due to the extremely large number of measurements ($N_{pt} = 1120$) taken — and despite the fact that only a minimal number of these exhibit per-point sensitivities that exceed the highlighting cut $|S_f|> 0.25$ imposed on the impact maps in this analysis.

Continuing, the right panel of Fig. 3.1 also shows the corresponding distribution of CTEQ-TEA sensitivities in $(x,\,\mu)$ space for the second isovector moment $\langle x^2\rangle_{u^- - d^-}$, for which the constraints arising from individual experiments fitted by CT are somewhat different. In this instance, we find the distribution of point-averaged sensitivities within the CTEQ-TEA data set to be driven primarily by electroweak boson production measurements: CMS7Masy2'14 (0.492), D02Easy2'15 (0.416), CMS7Easy'12 (0.282), LHCb7Wasy'12 (0.250), CCFR-F3'97 (0.224), and D02Masy'08 (0.211). In contrast to the total isovector momentum considered above, we therefore again find a leading role for the 7 TeV CMS lepton asymmetry measurements of $A_\mu(\eta)$ [CMS7Masy2'14] and $A_e(\eta)$ [CMS7Easy'12], although the size of the sets ($N_{pt} = 11$) is such that their aggregated pull on $\langle x^2\rangle_{u^- - d^-}$ is dominated by larger fixed-target data sets identified by an analysis of the summed sensitivities, as we point out below. In addition to the CMS measurements, a number of other electroweak boson sets evidently have stronger pull on the $\langle x^2\rangle$ isovector moment, including the corresponding D∅ measurement of $A_e(\eta)$ (D02Easy2'15), which probes higher x, as well as LHCb.

The evaluation according to the aggregated sensitivities reveals a different hierarchy. In this case, fixed-target measurements of DIS cross sections and structure functions — as well as a couple Drelly-Yan sets — are dominant, namely, HERAI+II'15 (36.5), BCDMSp'89

(33.1), E866pp'03 (22.2), CCFR-F3'97 (19.3), and NMCrat'97 (18.4), with a rapid falloff in the aggregated sensitivity below $\sum|S_f|$ beyond these experiments. It should be noted, however, that were the boson production data sets with especially strong per-datum sensitivities indicated above combined into a single collection, the resulting aggregated impact of this collection would approach $\sum|S_f|\sim 34$, placing this combination of 150 data points just beyond the BCDMS F_2^p data ($N_{pt} = 337$) and only behind the HERAI+II'15 set ($N_{pt} = 1120$) in total sensitivity.

3.2.3.1.2 Moments of q^+ Distributions. As we pointed out in the discussion of the $u - d$ moments at the start of Sec. 3.2.3.1.1 above, the fact that the disconnected insertions contribute equally to u and d-type distributions implies their vanishing for isovector (τ_3) charges. Unlike these combinations, the moments of flavor-separated distributions like $u^+(x, \mu^{\mathrm{lat}})$ and $d^+(x, \mu^{\mathrm{lat}})$ receive contributions from both connected and disconnected insertions. The disconnected insertions arise from Wick contractions of quark fields not explicitly present in interpolation operators used to construct the 2-point function associated with the nucleon propagator; disconnected insertions are therefore essentially equally present in both u-type and d-type flavor-separated moments. Unfortunately, evaluating disconnected insertions on the lattice is computationally expensive and, historically, has proved challenging.

In the case of the higher moments, they are generally quite small — *e.g.*, Ref. [103] found $\langle x^2\rangle_{u^-,d^-}$ to be consistent with zero, and, along these lines, the disconnected contributions in these instances will themselves be fairly small. In fact, even for the larger first moments $\langle x\rangle_q^+$, the differences between calculations with and without disconnected insertions are within uncertainties, suggesting that these contributions may not be so large for the u and d-type distribution moments. Nucleon strangeness, on the other hand, necessarily originates exclusively with disconnected insertions, since the proton possesses no valence strange content, and, consequently, no strange quark fields are explicitly present in the nucleon interpolation operators from which two-point correlation functions are evaluated. Precise lattice data involving each of these flavors and for multiple Mellin moment orders would be instrumental in disentangling the interplay of connected vs. disconnected insertions and helping to resolve the underlying dynamics. This observation also motivates a comprehensive assessment of the same moments as computed from phenomenological PDFs as well as a reckoning of the the various pulls from experimental data that act upon them.

3.2.3.1.3 u^+-quark Moments. For $\langle x \rangle_{u+}$ we consider the CTEQ-TEA sensitivity contained in the map of the LHS panel of Fig. 3.2; as is the case fairly generically for the the leading moments of the light quark distributions, the most concentrated locus of high-sensitivity data are found in the fixed-target sector in the lower right quadrant of the (x, μ) plot — particularly, for $x \gtrsim 0.01$ and $\mu \lesssim 10$ GeV. Upon inspection, these points arise from measurements at BCDMS (on the proton — BCDMSp'89 — as well as the deuteron, BCDMSd'90) and the E866 data. Empirical information with especially larger per-datum sensitivities can again be identified by listing the leading experiments in descending order of their point-averaged sensitivities. These are CCFR-F3'97 (0.337), E866rat'01 (0.277), D02Masy'08 (0.250), CMS7Masy2'14 (0.248), and NuTeV-nu'06 (0.221). While for the total sensitivities we find HERAI+II'15 (40.8), BCDMSp'89 (39.5), CCFR-F3'97 (29.0), BCDMSd'90 (24.8), CDHSW-F2'91 (16.5), CDHSW-F3'91 (15.1), E866pp'03 (10.3).

3.2.3.1.4 d^+-quark Moments. As an illustration of the flavor dependence of the PDF moments, we compare in the right panel of Fig. 3.2 with the corresponding sensitivities for $\langle x \rangle_{d+}$, shown in the right panel. Here we find according to the leading per-datum sensitivities a strong role again for charged-current processes, namely, from CMS lepton-charge asymmetry data and νA DIS: CMS7Masy2'14 (0.419), NuTeV-nu'06 (0.238), CMS7Easy'12 (0.228), CCFR-F3'97 (0.227), CDHSW-F2'91 (0.225). On the basis of the total sensitivities of these

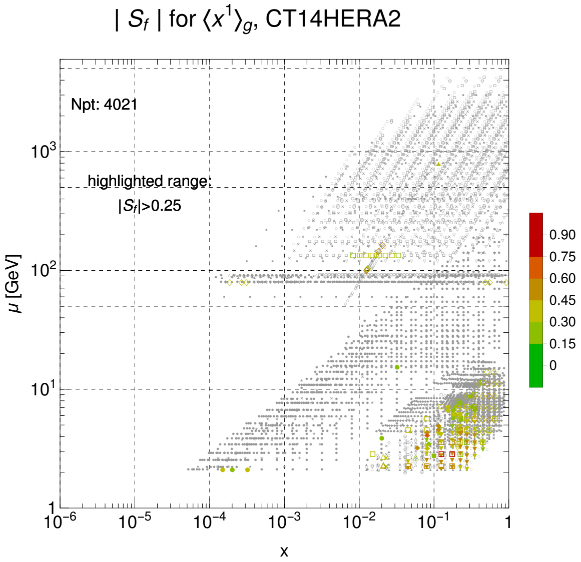


Figure 3.5: Sensitivity of the CTEQ-TEA data sets to $\langle x \rangle_g$. The factorization scale at which the moment is evaluated is $\mu_F = 2$ GeV. Here we have only a single panel for $\langle x \rangle_g$ given the fact that lattice computations thus far only exist for $\langle \mathcal{G}_{\mu\nu} \mathcal{G}^{\mu\nu} \rangle$.

experiments, however, we again find a hierarchy dominated by the combined HERA data set, for which the charge-current (CC) $e^{\pm}p$ channels show somewhat enhanced sensitivity to moments of $d(x)$ relative to $u(x)$ according to both the $\langle |S_f| \rangle$ and $\sum |S_f|$ metrics illustrated in Fig. 3.7; this is particularly true of the CC e^+p HERAI+II information, for which the LO reduced cross section $\sigma_r(x, Q^2)$ is closely driven by the behavior of d-type quark distributions, especially at larger x. Beyond the HERA measurements, the descending list of high total sensitivity experiments has a trailing collection of fixed-target measurements, namely, HERAI+II'15 (54.2), BCDMSd'90 (26.5), NMCrat'97 (22.6), CCFR-F3'97 (19.5), CDHSW-F2'91 (19.1), BCDMSp'89 (18.5), E866pp'03 (14.8). In this instance, the second most influential measurement is to be found in the deuteron target structure functions extractions from BCDMS (BCDMSd'90) — a fact consistent with the traditional importance ascribed to deuteron measurements for performing nucleon flavor separations.

3.2.3.1.5 s^+-quark Moments.

For the sensitivities to the moments of the s^+ distribution, we find for $\langle x \rangle_{s^+}$ the measurements with leading point-averaged sensitivities to be

NuTeV-nu'06 (0.429), CCFR SI nub'01 (0.344), CCFR SI nu'01 (0.313), NuTeV-nub'06 (0.302), D02Masy'08 (0.274); while those with the highest predicted total impact based on aggregated sensitivity are HERAI+II'15 (31.4), NuTeV-nu'06 (16.3), CCFR SI nub'01 (13.1), CCFR SI nu'01 (12.5), NuTeV-nub'06 (10.0). Across both aggregated and average per-point sensitivities, the decisive role of neutrino scattering data is evident, despite the still leading role of the combined HERA measurements — especially noting the fact that the summed sensitivity of the 4 leading ν experiments mentioned above is $\sum|S^{\nu}_{f}|= 51.9$, exceeding the HERA accumulated impact by $\sim$65%.

In the CT14HERA2 PDF set, strangeness was parametrized symmetrically (*i.e.*, under the assumption $s(x) = \bar{s}(x)$; as a result, the moments of the s^--type distributions, including $\langle x^2 \rangle_{s^-}$, are identically zero. For that reason, we instead consider here the next highest moment of the strangeness distribution, *i.e.*, the third moment $\langle x^3 \rangle_{s^+}$; for which we find the point-averaged sensitivities of the leading experiments (again, cutting at $\langle |S_f| \rangle > 0.2$) to be NuTeV-nub'06 (0.568), CCFR SI nub'01 (0.387), and NuTeV-nu'06 (0.269), clearly suggesting the very important role of the NuTeV $\bar{\nu}$ dimuon production measurements (NuTeV-nub'06), which show especially enhanced sensitivity to the higher $\langle x^3 \rangle$ moment than was seen for the total strange momentum $\langle x \rangle_{s^+}$. For the total sensitivities, the constraints imposed by the CTEQ-TEA data set come primarily from several experiments HERAI+II'15 (20.3), NuTeV-nub'06 (18.7), CCFR SI nub'01 (14.7), and NuTeV-nu'06 (10.2). Thus, for both Mellin moments of the s^+ distribution, there is a clear advantage enjoyed by the fixed-target ν DIS experiments.

3.2.3.1.6 Moments of q^- Distributions. At present, lattice determinations for the next highest $\langle x^2 \rangle_q$ moments of the light quark distributions have not matured to the level of extant calculations of the first moments $\langle x \rangle_q$, particularly in the sense that these have been computed thus far only in Ref. [103] in the quenched approximation (*i.e.*, excluding dynamical quark loops). Nonetheless, such determinations are likely forthcoming, and can yield vital information regarding asymmetric x dependence in the light quark distributions.

We plot the sensitivity maps to the $\langle x^2 \rangle_{q^-}$ moments of the u and d quark distributions in the left and right panels of Fig. 3.3, respectively. As elsewhere, these panels examine the sensitivity of the CTEQ-TEA set to moments evaluated at the typical lattice scale $\mu = \mu^{\rm lat} = 2$ GeV.

3.2.3.1.7 u^--quark Moment. For the second moment of the u^- distribution, the leading point-averaged sensitivities are due to fixed-target DIS experiments and the 7 TeV CMS lepton charge asymmetries, led by CCFR-F3'97 (0.503); beyond this, experiments with $\langle |S_f| \rangle > 0.2$ are CMS7Masy2'14 (0.413), CDHSW-F3'91 (0.248), and CMS7Easy'12 (0.244). In this context, the fact that information on the parity-odd structure function F_3^p — especially as provided by CCFR-F3'97 — shows such sizable influence over $\langle x^2 \rangle_{u^-}$ is consistent with the leading-order $\sim q - \bar{q}$ behavior of F_3^p in the quark-parton model. As such, thorough knowledge of the x dependence of xF_3 facilitates an unraveling of the C-odd distributions of the q^- type, and constrains their higher moments. As was the case, however, for the $\langle x \rangle_{q^+}$ moments, consideration of the aggregated sensitivities reveals a larger spread of experiments with strongest pulls belonging again to the combined HERA data set HERAI+II'15 (43.9), the νDIS measurements of xF_3 from CCFR [CCFR-F3'97 (43.2)] identified by the point-averaged ranking above, BCDMSp'89 (39.2), and E866pp'03 (32.7). Having somewhat diminished but still significant pulls are several of the other fixed-target experiments involving both neutrino and μ DIS as well as the Drell-Yan process; namely, these are CDHSW-F3'91 (23.8), E605'91 (18.4), BCDMSd'90 (13.6), and NMCrat'97 (13.4).

3.2.3.1.8 d^--quark Moment. As observed above for the lower $\langle x \rangle_{q^+}$ moments imaged in Fig. 3.2, there are evident differences between the sensitivity maps for d- vs. u-quark moments, and this holds again for the explicit comparison of $\langle x^2 \rangle_{d^-,u^-}$ illustrated in Fig. 3.3. In fact, these systematic differences are especially marked for the $\langle x^2 \rangle$ moments, as inspection of Fig. 3.3 attests. Especially notable in the right panel of Fig. 3.3 is the very strong sensitivity $|S_f| \gtrsim 0.75$ for a select subset of the gauge production data, particularly for $x \gtrsim 10^{-4}$ and separately for $x \gtrsim 0.1$. These especially strong constraints to $\langle x^2 \rangle_{d^-}$ originate from an amalgam of electroweak data sets, among which we find the 8 TeV forward $W^\pm, Z$ production cross section data of LHCb (LHCb8WZ'16), the analogous information at 7 TeV (LHCb7ZWrap'15), as well as the forward-backward e^+e^- asymmetry in $W^\pm, Z$ production at Runs 1 and 2 of CDF, CDF1Wasy'96 and CDF2Wasy'05. Compared with $\langle x^2 \rangle_{u^-}$ on the other hand, for the second moment of $d^-(x)$ we find a substantially more restricted outlay of individually high-impact measurements in the fixed-target region, with significantly fewer data belonging to very high $x \gtrsim 0.4$ or $x \lesssim 0.2$ identified. Of these, the E605, NMCrat, and CCFR-F3 points enjoy special prominence. Many of these trends revealed by the sensitivity map in the right panel of Fig. 3.3 are further confirmed by quantitative ranking of the CTEQ-TEA experiments, especially based on the per-point sensitivities. For the second moment of the d^- distribution, the point-averaged sensitivity ranked experiments are D02Easy2'15 (0.519), CCFR-F3'97 (0.381), LHCb7Wasy'12 (0.362), CMS7Masy2'14 (0.328), D02Masy'08 (0.293), CDF1Wasy'96 (0.252), LHCb8WZ'16 (0.217), LHCb7ZWrap'15 (0.214), and E605'91 (0.207).

For the aggregated sensitivities, here also we find knowledge of the d^- second moment to be driven foremost by xF_3 data from CCFR and the combined HERA data, CCFR-F3'97 (32.8) and HERAI+II'15 (32.3), respectively. We note, however, that the total sensitivity of these leading experiments to the d^- distribution is reduced roughly $\sim 30\%$ relative to what was found for the corresponding u-quark sensitivities. Beyond these leading measurements, an assortment of μ and νDIS and Drell-Yan experiments have the tightest pulls. Again in descending order, these are 605'91 (24.7), CDHSW-F3'91 (18.5), BCDMSd'90 (15.2), NMCrat'97 (14.8), BCDMSp'89 (14.4), E866pp'03 (13.8), and CDHSW-F2'91 (11.7).

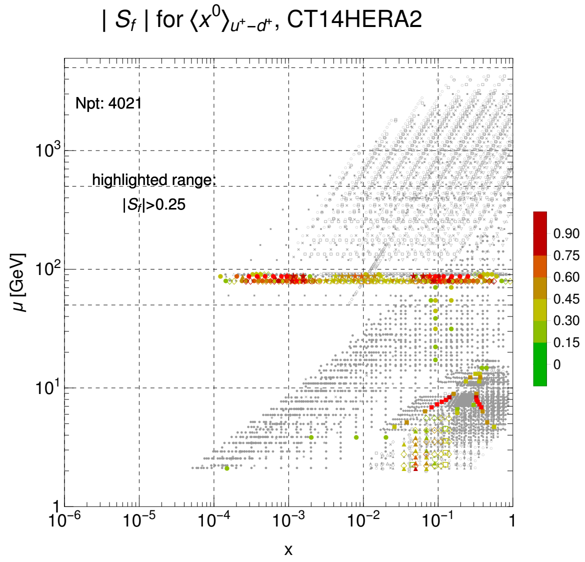


Figure 3.6: Sensitivity of the CTEQ-TEA data to the first moment $\langle 1 \rangle_{(u^+-d^+)}$. The factorization scale taken for the Mellin moment is $\mu = 2$ GeV. We stress that, while this combination is not directly calculable by the usual lattice methods, its appearance in the Gottfried Sum Rule motivates its study, as well as a focus upon higher moments.

3.2.3.2. The Gluon Momentum Fraction

We can extend this program to the gluonic sector, considering the total nucleon momentum carried by gluons as characterized by the first moment of the gluon distribution, $\langle x \rangle_g$; for the time being, this is the only moment of the gluon PDF which has been evaluated by multiple lattice groups, and we therefore concentrate on it primarily. Fig. 3.5 illustrates the sensitivity to $\langle x \rangle_g$ of the CTEQ-TEA data considered in the plots of the preceding section.

Unlike what was generally observed for the quark distribution moments reported above, only two experiments within the CTEQ-TEA set lie above the $\langle |S_f| \rangle > 0.2$ ranking cut used previously. Based on their point-averaged sensitivities to $\langle x \rangle_g$, these are both measurements of F_2^p (albeit extracted from nuclear data), specifically, CDHSW-F2'91 (0.312) and CCFR-F2'01 (0.237). Immediately beyond these most valuable 'per-point' measurements of F_2, several other experiments fall immediately below the cut with slightly weaker averaged sensitivities, including the νDIS measurement of $F_3(x, Q^2)$ recorded by CCFR-F3'97 (0.188), the 7 TeV ATLAS high-p_T Z production data of ATL7ZpT'14 (0.184), and the 8 TeV $t\bar{t}$ measurements from ATLAS, ATL8ttb-mtt'16 (0.172).

As we found for many of the light quark moments studied above, in terms of the the aggregated sensitivities, we observe a distinctly important role for the combined HERA data set — HERAI+II'15 (49.2) — a result consistent with the significant precision and very wide coverage over x and Q^2 of these cross section data. This wide coverage acts as a crucial lever arm to constrain the QCD evolution in the CT (or indeed any) parametrization, and thereby restricts the phenomenological behavior of the singlet and gluon distributions. After the reduced cross section measurements of HERA, a cascading series of nucleon or deuteron structure function $F_2^{p,d}$ measurements obtained on either hydrogen or nuclear targets contain the greatest share of information on the integrated gluon distribution. In descending order, these are the $\nu - \mathrm{Fe}$ DIS data of CDHSW-F2'91 (26.5), followed by μ scattering data from BCDMS, first on the deuteron, BCDMSd'90 (25.8), as well as on a hydrogen target, BCDMSp'89 (24.9). Lastly, neutrino data from CCFR on F_2 [CCFR-F2'01 (16.3)] and xF_3 [CCFR-F3'97 (16.2)] have comparable pull between these two structure function measurement channels, and important influence in the wider fit. It is intriguing to notice that, while the aggregated pull of HERAI+II'15 (49.2) strongly dominates the spread of CTEQ-TEA experiments considered in isolation, were the leading $\nu - \mathrm{Fe}$ experiments above regarded as a single experiment and their accumulated sensitivities simply combined directly, the result ($59 \gtrsim 49.2$) surpasses the very large combined HERA data set, which is based on $N_{pt} = 1120$ cross section measurements. A similar observation holds for the BCDMS data. We therefore again stress the observation made above in the context of the aggregated CTEQ-TEA sensitivities to, *e.g.*, $\langle x^2 \rangle_{u^- - d^-}$: while the great extent of the combined HERA data set's kinematical coverage frequently awards it a leading role in terms of its aggregated effect, agglomerations of much smaller, targeted data sets can quickly have a comparable or greater combined effect, in principle.

Although they do not appear among the core of most decisive experiments detailed above, some of the newer LHC data sets canvassed in chapter 2 (or Ref. [34]) are nonetheless among the top $\sim$10 most sensitive experiments to $\langle x \rangle_g$ — particularly the inclusive jet data found in chapter 2 (or Ref. [34]) to provide important constraints to the gluon distribution overall. Specifically, these are the 8 and 7 TeV CMS inclusive jet production data, CMS8jets'17 (7.1) and CMS7jets'14 (6.1), respectively.

3.2.3.3. Flavor Asymmetries of the Nucleon Sea

As a final consideration in our present study of the PDF moment sensitivities of high-energy data, we demonstrate the importance of disentangling the various flavor-dependent moments analyzed above to the phenomenology of the nucleon's light quark sea. In particular, the flavor structure of the proton's quark sea has for decades attracted sustained focus, especially regarding the dynamical origin of the observed charge-flavor asymmetry embodied by the breaking of the SU(3) PDF relation $\bar{u}(x) = \bar{d}(x) = s(x) = \bar{s}(x)$ often assumed in the earliest phenomenological QCD global fits. Much formal interest in this topic attends to the fact that the x-dependent breaking of the SU(2) symmetry relation $\bar{d}(x,\mu) - \bar{u}(x,\mu) = 0$ at low scales is principally understood as a feature of nonperturbative QCD [104, 105]; for instance, patterns of dynamical chiral symmetry breaking in QCD favor hadronic dissociations of the proton having the form $p \to \pi^+ n$ at low energies which are thought to produce generic excesses of $\bar{d}$ over $\bar{u}$ in contributing to the nucleon's flavor structure [106–108]. It should be noted, however, that accounting for the detailed x dependence of $\bar{d}(x) - \bar{u}(x)$ (or, equivalently, of deviations of the flavor ratio from $\bar{d}/\bar{u} = 1$) in the context of meson-cloud models informed by this physical picture has been challenging.

Historically, much of the empirical information on parton-level flavor symmetry violation in the nucleon sea has been garnered through examinations of the unpolarized DIS structure functions. Formally, structure functions can be described using well-established factorization theorems in terms of which they may be separated via convolution integrals over the long-distance PDFs and the perturbative coefficient functions. While analyses extended to higher orders in α_s entail many complications, even leading-order decompositions of the DIS structure functions made using the quark-parton model (QPM) can illustrate the connection to quark-level flavor symmetry breaking.

In this context, a crucial observable first measured systematically by NMC [109, 110] is the Gottfried Sum Rule [111], which is sensitive to nonperturbative dynamics leading to the SU(2) flavor asymmetries in the light quark sea mentioned above. The canonical expression of the sum rule can be obtained by applying the leading-order QPM to the isovector structure function difference:

$$\begin{aligned}\int_0^1 \frac{dx}{x}(F_2^p - F_2^n)|_{\mathrm{QPM}} = \frac{1}{3}\int_0^1 dx\,(u^+ - d^+) &\equiv \frac{1}{3}\langle 1\rangle_{u^+-d^+} \\ &= \frac{1}{3} - \frac{2}{3}\int_0^1 dx\,(\bar{d} - \bar{u})\ , \end{aligned} \tag{3.16}$$

where we have used isospin and the identities $q^+ = q^- + 2\bar{q}$ and $\int dx(u^- - d^-) = 1$ to rearrange the first line into the standard statement of the sum rule on the second. Most importantly, we highlight the fact that the zeroth moment of the isovector PDF, $\langle 1\rangle_{(u^+-d^+)}$ [the RHS of the first line of Eq. (3.16)], is directly related to the behavior of $\bar{d}-\bar{u}$, deviating from unity when $\langle 1\rangle_{\bar{d}-\bar{u}} \neq 0$. While this latter sea quark PDF moment appearing on the far RHS of Eq. (3.16) is not directly accessible on the lattice as a zeroth unpolarized moment, we are nonetheless able to compute the sensitivity of the CTEQ-TEA set to $\langle 1\rangle_{(u^+-d^+)}$ and the related violation of the symmetric sea $\bar{u} = \bar{d}$ scenario formulated in terms of Mellin moments; this connection crucially motivates lattice measurement of the higher isovector moments $\langle x^{1,3}\rangle_{u^+-d^+}$ treated in Secs. 3.2.3.1 and 3.5, which would constrain the behavior of the phenomenological isovector distribution and inform its zeroth moment and analyses of the Gottfried Sum Rule. Moreover, the $x < 0$ region of the isovector quasi-distribution presented in Sec. 3.3 is immediately related to $\bar{d}(x)-\bar{u}(x)$, again implying a complementary avenue for lattice sensitivity to the light quark sea.

In Fig. 3.6 we diagram the calculated sensitivities of the CTEQ-TEA high-energy data set to the zeroth isovector moment $\langle 1 \rangle_{(u^+-d^+)}$. While the general pattern of sensitivities in Fig. 3.6 is consistent with what we observed for the higher isovector moments illustrated in Fig. 3.1, the very substantial magnitude of the sensitivities here, especially of the $W^\pm$ and Z production information and E866 cross section ratios (E866rat'01), represents an especially pronounced effect. To a lesser extent, we find in the fixed-target regime an assembly of measurements with notable pull, including several DIS experiments: the NMC structure function ratio information (NMCrat'97), the CCFR measurements of xF_3^p (CCFR-F3'97), and the BCDMS μ-H data (BCDMSp'89).

These visible features of the `PDFSense` sensitivity map are largely borne out by the point-averaged CTEQ-TEA sensitivities to the zeroth isovector moment; like the moments of the higher isovector moments and $d^\pm$ distributions explored above, the list of leading experiments ranked by this metric is again led by the 7 TeV μ asymmetry data recorded by CMS [CMS7Masy2'14 (0.645)], followed closely by the deuteron-proton cross section ratios measured by E866 [E866rat'01 (0.600)]; for the latter, this strong pull is notably consistent with E866's aim of probing the x dependence of $\bar{d}(x)/\bar{u}(x)$ — a topic which continues to motivate modern experiments like SeaQuest. Following these, the per-datum sensitivities of the CTEQ-TEA data are dominated by an amalgam of electroweak experiments represented by the rows of gauge boson data shown in Fig. 3.6. Again in descending order, these include LHCb7Wasy'12 (0.546), LHCb8WZ'16 (0.432), CMS7Easy'12 (0.381), CMS8Wasy'16 (0.370), LHCb7ZWrap'15 (0.351), D02Easy2'15 (0.323), D02Masy'08 (0.252), and ATL7WZ'12 (0.219). Ordered according to their aggregated impact, on the other hand, only 9 experiments exceed $\sum|S_f|> 10$. These now include the usual DIS information from HERA and fixed-target data from NMC, CCFR, and BCDMS, as well as the E866 *pp* Drell-Yan cross section data — again owing to the aggregated pull of these enlarged data sets. In order of total sensitivity, these most decisive experiments are HERAI+II'15 (51.0), BCDMSp'89 (21.2), LHCb8WZ'16 (18.1), CCFR-F3'97 (15.7), NMCrat'97 (15.2), E866pp'03 (14.8), CMS8Wasy'16 (12.2), LHCb7ZWrap'15 (11.6), and BCDMSd'90 (10.2). Of these, there is again a pronounced effect from DIS experiments led by the combined HERA data which contribute by merit of their marginal per-datum sensitivity $\sim 0.05 - 0.1$ and the magnitude N_{pt} of the data sets to which they belong, much as we observed for many of the other light quark moments above.

3.3. Sensitivities to Quark Quasi-Distributions

In addition to the PDF Mellin moments we analyzed in Sec. 3.2, it has recently been proposed [23] that lattice QCD may evaluate parton "quasi-distributions" (qPDFs) over the quark-hadron longitudinal momentum fraction $x = k_z/P_z$ by evaluating matrix elements of the form

$$\widetilde{q}(x, P_z, \widetilde{\mu}) = \int_{-\infty}^{\infty} \frac{dz}{4\pi} e^{ixP_z z} \langle P|\overline{\psi}(z)\gamma^z U(z,0)\psi(0)|P\rangle \ , \tag{3.17}$$

where $U(z,0)$ is a gauge link along the longitudinal z direction, and the argument $\widetilde{\mu}$ represents the scales in the RI/MOM scheme; in practice, this involves the introduction of the parameters p_z^R and μ_R, which, for the purpose of this analysis, we fix to the values given in Ref. [97], $p_z^R = 2.2$ GeV and $\mu_R = 3.7$ GeV. Given its status as a matrix element of correlation functions along a spacelike longitudinal direction (unlike the ordinary $\overline{\text{MS}}$ PDFs), the quasi-distribution of Eq. (3.17) can be computed on the lattice, and ultimately matched to traditional phenomenological PDFs via an inversion of the expression given in Eq. (3.2). On the other hand, rather than inverting Eq. (3.2) to obtain the $\overline{\text{MS}}$ PDF from lattice qPDF output, it is also possible to use Eq. (3.2) to compute the P_z-dependent qPDF from a phenomenological $\overline{\text{MS}}$ PDF; we show the result of doing this to evaluate the quasi-distribution matched from the CT14HERA2 NNLO PDFs in Fig. 3.9. While the expression appearing in Eq. (3.17) is standard in the quasi-PDF literature, we clarify that in practice it can be advantageous to compute matrix elements with the replacement $\gamma^z \to \gamma^t$ as described in Ref. [97]. While quasi-distributions computed with γ^t have similar limiting behavior for $P_z \to \infty$ as those determined using γ^z, lattice calculations carried out with γ^t enjoy greater stability against operator mixing [112], and the numerical results shown in this section therefore assume this procedure. Analogously to the calculations for the PDF moments in Sec. 3.2 using Eq. (3.12), Hessian sets for the qPDFs can be obtained algorithmically by applying Eq. (3.2) to the collection of CT14HERA2 NNLO PDFs at a given choice of P_z, $\overline{\text{MS}}$, and RI/MOM factorization and regularization scales to similarly obtain an error ensemble $\widetilde{q}_{j\in\{2N\}}(x, P_z, \widetilde{\mu})$ in addition to a central value. This Hessian set may then be used to compute the sensitivities of the CTEQ-TEA set to the quark quasi-distributions along the lines described in Sec. 3.2.2

Before lattice output matures to a sufficient level to help specify the x dependence of PDFs through the combination of qPDF calculations and LaMET, it will be crucial to benchmark lattice calculations against our current knowledge of the fitted PDFs. For this purpose, we can again deploy `PDFSense` in a proof-of-principle demonstration showing the constraints from the present data in the CTEQ-TEA on the P_z-dependent quasi-distributions computed according to Eq. (3.2) from the underlying phenomenological PDFs, given current knowledge of the perturbative matching coefficient Z in Eq. (3.17); in the present Section, we assume an $\overline{\text{MS}}$ factorization scale of $\mu_F = 3$ GeV to agree with Ref. [97].

We wish to highlight both the dependence upon x and P_z of the quasi-distribution of the CTEQ-TEA sensitivities, and we therefore plot in this section four panels in Fig. 3.10 showing the behavior of the quasi-distribution $[\widetilde{u} - \widetilde{d}](x, P_z, \widetilde{\mu})$ at two representative values at relatively large $|x|$: $x = -0.5$, 0.85 for $P_z = 1.5$ and 3 GeV. For the quasi-distributions evaluated for $x < 0$, we note the implementation of the canonical relation $\overline{q}(x) = -q(-x)$, such that the negative x region of the quasi-distribution is related to the x dependence of the phenomenological anti-quark PDFs (on the logic that backward-moving quarks with longitudinal momenta $k_z = xP_z < 0$ are identifiable with forward-moving anti-quarks).

The essential point that emerges from Fig. 3.10 is the fact that a common cluster of experiments, mostly of higher x fixed-target and $W^{\pm}$ production and asymmetry measurements, represent the primary constraint to the $\overline{u} - \overline{d}$ quasi-distribution in a fashion that is largely independent of the boosted hadron's momentum P_z. Some intriguing P_z dependence does begin to emerge, however, for the CTEQ-TEA sensitivities to the highest x region of the isovector quasi-distribution, evident in Fig. 3.10 by comparing the $x = 0.85$ maps obtained for $P_z = 3.0$ and 1.5 GeV in the upper-right (b) and lower-right(d) panels. In particular, the P_z dependence appearing in the $|S_f|(x, \mu)$ distributions of Fig. 3.10 is signaled by the enhancement in the sensitivity to $[\widetilde{u} - \widetilde{d}](x = 0.85)$ of the highest $x \gtrsim 0.5$ μp DIS points of BCDMSp'89 and NMCrat'97 found for the $P_z = 1.5$ GeV [Panel (d)] compared to the analogous calculation, for the sensitivities to the $P_z = 3$ GeV quasi-distribution [Panel (b)]. This relative increase the sensitivity of the high-x DIS information is offset by an accompanying relative reduction in the general sensitivity of the $W^{\pm}, Z$ production data, which for $P_z = 3$ GeV exhibited significant pulls on $[\widetilde{u} - \widetilde{d}](x = 0.85)$, especially for the 7 TeV $A_{\mu}(\eta)$ asymmetry data taken by CMS, CMS7Masy2'14. The implication of these observations is the fact that a careful exploration of the nucleon structure function at high x may provide crucial information for constraining the P_z dependence of the quasi-distributions required for a robust application of LaMET.

3.4. Motivation for Future Experiments

An important motivation for numerous planned or proposed future measurements is precise unfolding of the nucleon's collinear PDFs, which are integral to tomographic maps of the proton's structure and crucial inputs for new physics searches on the energy frontier. While we have already observed the close connection between knowledge of the x dependence of phenomenological PDFs and their integrated Mellin moments, we emphasize that these future measurements themselves can potentially play an important constraining role *vis-à-vis* the PDF moments and other lattice QCD observables.

These future experiments will also play an important role in recording data that can constrain QCD analyses to benchmark improving lattice calculations along the lines highlighted above for the CTEQ-TEA set of high-energy data. A number of futuristic machines have either been proposed or planned with a stated aim (among other physics motivations) of disentangling the collinear structure of QCD bound states, including various futuristic hadron-collider experiments like the HL-LHC [113] and, *e.g.*, the AFTER@CERN proposal [114]. In addition to these, however, a number of lepton-hadron colliders have also been advocated, especially a future US-based electron-ion collider (EIC) [27, 52–54] and a lepton-nucleon/nucleus variant of the LHC, the Large Hadron-Electron Collider (LHeC) [26]. Among these various proposals, an EIC, given its high-luminosity coverage of the crucial few-GeV quark-hadron transition region in the kinematical parameter space, is most likely to serve the dedicated role of a hadron tomography machine. An EIC would therefore enjoy unprecedented facility in unfolding the nucleon's collinear and transverse structure at scales adjacent to the nucleon mass, $\gtrsim M$, such that the science output of an EIC would greatly build upon the JLab12 program [115] and impel next-generation Lattice QCD calculations.

As a simple illustration, we compute the analogous sensitivity maps that result from implementing a set of pseudodata into the `PDFSense` framework and examining our impact metrics for some of the primary quantities analyzed in this study — the first moment of the isovector distribution $\langle x \rangle_{u^+-d^+}$, and the high-$x$ behavior of the $P_z = 1.5$ GeV isovector quark quasi-distribution $[\widetilde{u} - \widetilde{d}](x, P_z, \widetilde{\mu})$. To avoid marrying our predictions to the specifics of a particular experimental proposal, we instead consider a hypothetical machine as a typical example, measuring the reduced cross section $\sigma_r(x, Q^2)$ via inclusive $e^\pm$ scattering on an unpolarized proton target. For this generic example, pseudodata are produced by randomly generating cross sections about the CT14HERA2 NNLO theoretical prediction with a Gaussian smearing function of standard deviation equal to the an assumed uncorrelated error. Theoretical predictions are for the reduced cross section measured in $e^\pm p$ scattering at $\sqrt{s} = 100$ GeV in both neutral- and charge-current interactions. For this illustration, statistical uncertainties are based upon assumed integrated luminosities of $\mathcal{L} = 100\,\mathrm{fb}^{-1}$ in e^-p scattering and $\mathcal{L} = 10\,\mathrm{fb}^{-1}$ for e^+p events.

Fig. 3.12 estimates the potential impact such a lepton-nucleon collider might have on the above-noted lattice computable quantities: in the left panel, the first moment, $\langle x \rangle_{u^+-d^+}$, of the isovector quark distribution, and, in the right panel, the large-x quasi-PDF matched from the CT14HERA2 NNLO PDF set according to Eq. (3.2). In both panels, physical channels for the inclusive DIS process are explicitly represented by unique symbols; these are NC e^-p (disks); NC e^+p (diamonds); CC e^-p (squares); CC e^+p (triangles).

Fig. 3.12 indicates that measurements at a high-luminosity lepton-nucleon collider can considerably improve the constraints on both quantities considered. In particular, they supply very substantial sensitivities across the range of x of the data set, with especially large predicted impacts for $x \gtrsim 0.1$ as well as the $x \lesssim 0.01$ regions. A notable feature of this information is the separation that emerges illustrating the crucial role of both electron and positron probes: once separated among channels, a prominent effect is the important role of the charged current (CC)-mediated positron-nucleon scattering (e^+p); this impact is very pronounced at large $x \gtrsim 0.1$. and EIC, $x > 0.01$ sensitivities mainly come from CC channel and EIC NC channel, and $x < 0.001$ sensitivities mainly come from NC e^+p and e^-p channels.

A recurring observation in the previous section(s) has been the significant impact of data involving nuclear targets, which aford critical, and, in many cases, leading, information on essentially all PDF moments analyzed in Sec. 3.2. This is similarly true of the isovector qPDF examined in Sec. 3.3. Details of the nuclear binding at work in the deuteron, for instance, are relevant for a number of the CTEQ-TEA sets, including BCDMSd'90, NMCrat'97, and E866rat'01. On the other hand, heavier nuclear systems were probed in several other fixed-target experiments, especially those involving νDIS; these include CDHSW (both F_2 and F_3 sets, measured on Fe), the inclusive CCFR and semi-inclusive dimuon data from NuTeV and CCFR (all also measured on Fe), and the E605 fixed-target pA Drell-Yan measurements (Cu target). In multiple instances — for example, in the impact plots for the strangeness moments $\langle x^{1,3}\rangle_{s^+}$, the C-odd combinations $\langle x^2\rangle_{u^-,d^-}$, and even the gluon total momentum $\langle x\rangle_g$ — these experiments represent the first, second, or third most influential information by the aggregated or point-averaged sensitivity, or both. Present phenomenological constraints, particularly at large x, are therefore strongly dependent on data for which nuclear corrections are an important consideration. These corrections are imperfectly known, and often dependent on model treatments or an assumption that that nuclear correction effects are simply absorbed into extracted PDF uncertainties. An EIC would be well-poised to address these issues by performing detailed studies of nuclear medium effects.

3.5. Implementation of Lattice Data in QCD Analyses

In the foregoing sections we have analyzed various empirical constraints upon *individual* lattice QCD observables which are either presently accessible or expected to be in the foreseeable future. These experimental data were taken either from the CTEQ-TEA high-energy data set or generated as hypothetical pseudodata recorded at an EIC-like $e^{\pm}p$ DIS collider. The main purpose of this exploration is a thorough understanding of the processes and measurements that impose the strongest constraints on PDF-based predictions of lattice-calculable quantities and which will be required to improve future phenomenological benchmarks. As stressed in Sec. 3.1, stringent benchmarks supplied by global fits constrained by the most incisive experimental information will be necessary to ensure lattice calculations achieve a sufficient level of maturity that they can be reliably incorporated as prior constraints on QCD analyses.

But to this latter point, it is possible to proceed in a direction converse to that taken in the sections above by asking: how might the multifaceted results of lattice QCD constrain the x dependence of collinear PDFs fitted in future global analyses? The constraining power and robustness of the lattice information incorporated into these future analyses will derive from the inclusion of outputs computed across a variety of flavors using a mix of operators and techniques to evaluate various observables, including the Mellin moments and qPDFs considered above. Given the complexity of the multichannel information furnished by the lattice, its inclusion in upcoming PDF fits can be expected to produce a complicated set of pulls on the underlying parametrizations. It will therefore be very difficult to predict *a priori* how information from the lattice might collectively impact a particular global analysis. Still, the `PDFSense` framework employed in this study can help illuminate this issue by quantifying and mapping the subtle relationships that exist among the PDFs and various observables the lattice may calculate.

Rather than attempting to disentagle the many complications that would accompany the implementation of lattice data into a global fit, we instead illustrate with a simple example. We again consider the moments of the SU(2) isovector distribution $u - d$, in this case, contrasting the two lowest moments of the q^+-type distribution, $\langle x \rangle_{u^+ - d^+}$ and $\langle x^3 \rangle_{u^+ - d^+}$, for which we plot the $|S_f|(x, \mu)$ sensitivity map in the left and right panels of Fig. 3.13, respectively. We clarify that in Fig. 3.1 of Sec. 3.2.3.1, we examined $\langle x \rangle_{u^+ - d^+}$ and $\langle x^2 \rangle_{u^- - d^-}$, but here we directly examine the effect of increasing the Mellin moment order on a specific flavor/charge combination for the purpose of showcasing the relationship between the order and the associated x dependence of the constraints imposed by data. In moving from the first moment characterized in Fig. 3.1 and surrounding text to the third, we find a notable reduction in the point-averaged sensitivity, $\langle |S_f| \rangle$, of the leading experiment according to this metric, which remains CMS7Masy2'14 (0.342), but is now immediately succeeded by D02Easy2'15 (0.307). The other leading experiments by per-datum sensitivity also remain E866rat'01 (0.225), CMS7Easy'12 (0.203), and CCFR-F3'97 (0.187), but with significant decreases in their values of $\langle |S_f| \rangle$ for $\langle x^3 \rangle_{u^+ - d^+}$. By inspecting the total sensitivities, an important reordering of the experimental hierarchy becomes evident. In this case, BCDMSp'89 (39.4), with its large share of measurements at high x and $\mu \sim 10$ GeV, displaces the combined HERA data, HERAI+II'15 (34.9), in terms of total pull. The important experiments identified by their aggregated sensitivities to $\langle x\rangle_{u^+-d^+}$ continue to place strong constraints upon $\langle x^3\rangle_{u^+-d^+}$, with these being E866pp'03 (24.4), CCFR-F3'97 (16.1), and NMCrat'97 (15.5). For these data, however, there is a salient rightward shift toward higher x in the (x,μ) space displayed in Fig. 3.13. As a straightforward metric to quantify the distribution over x of the sensitivities $|S_f|(x,\mu)$ plotted in Fig. 3.13, we may evaluate an ensemble average

$$\overline{x}_{|S_f|} = \frac{\sum_i x_i |S^i_f|}{\sum_i |S^i_f|} \ , \tag{3.18}$$

where the sum i runs over data points in the CTEQ-TEA set. On the basis of this metric and the panels of Fig. 3.13, a relationship emerges between the order n of the PDF moment $\langle x^n\rangle_{q^\pm}$ and the kinematics of the most constraining data in the global analysis, with PDF moments of higher order being constrained more strongly by data recorded at higher x. For the first and third isovector moments plotted in Fig. 3.13, we obtain a systematic increase in $\overline{x}_{|S_f|}$ as the order of the moment is enlarged, finding a shift from $\overline{x}_{|S_f|} = 0.193$ for $\langle x\rangle_{u^+-d^+}$ to $\overline{x}_{|S_f|} = 0.286$ for $\langle x^3\rangle_{u^+-d^+}$. Similar relationships are observed between $\overline{x}_{|S_f|}$ as given by Eq. (3.18) and the Mellin moments of other PDF flavors and combinations.

It is possible to further unravel the observed x dependence in $|S_f|(x,\mu)$ by considering the correlation defined in Eq. (3.13). In Fig. 3.14, we plot the x-dependent correlation between the PDF and its corresponding Mellin moment for two examples — the lowest three lattice-accessible moments of the d^+ distribution (left panel) and the same information for the isovector $u^+ - d^+$. Across both panels, we observe the same qualitative x dependence in the correlation as the order n of the Mellin moment $\langle x^n\rangle_{q^+}$ is increased. Specifically, while the lowest $n=1$ moment is significantly correlated with its PDF's x depedence over a wide range of x, peaking near $x \sim 0.1$, this correlation vanishes rapidly at highest x. On the other hand, the PDF correlations with higher moments are rather different, in this case being quite modest, especially for the highest $n=5$ moment, over most of the plotted range before becoming very large, $C_f \sim 1$, at $x \gtrsim 0.3$. In fact, this behavior was reflected in Fig. 3.13, which demonstrated the sensitivity shift in $|S_f|(x,\mu)$ to favor many of the large-x data as the Mellin moment is increased. Taken in conjunction with the correlation results shown in Fig. 3.14, we may infer that the sensivity of high-x data to higher moments follows from an underlying reciprocal relation that connects the high-x behavior of PDFs to their higher-order Mellin moments. The observation that moments $\langle x^n \rangle_{q^\pm}$ of successively higher order ($n > 1$) are increasingly sensitive to the PDFs' large-x behavior provides an impetus to seek alternative moment-weighting functions which may be sensitive to low x. One possible choice would be successively higher moments of distributions smeared with polynomials in the difference, $(1-x)^n$, *i.e.*, $\langle (1-x)^n \rangle_{q^\pm}$. In principle, information on the $\langle (1-x)^n \rangle_{q^\pm}$ moments may be of use for constraining PDFs in the region of small x, where they must be integrable in the limit $x \to 0$ to ensure $\langle (1-x)^n \rangle_q$ are well-defined. In fact, since the polynomial expansion of $(1-x)^n$ is a linear combination in x^n, *i.e.*

$$\begin{aligned} \langle (1-x)^n \rangle_q &= \sum_{k=0}^{n} C_k^n \; (-1)^k \; \langle x^k \rangle_q \\ C_k^n &\equiv \frac{n!}{k!\,(n-k)!} \end{aligned} \tag{3.19}$$

results for select moments $\langle x^n \rangle$ as might be provided by the lattice could constrain PDF behavior not only at successively larger x, but also provide information at the lower portion of the PDF's support.

We therefore stress that it will be crucial for future phenomenological analyses to leverage a plurality of lattice results to most strongly constrain their likelihood functions or parameter spaces. The necessity of doing so is apparent from our results on the moments of $\langle x^n \rangle_{u^+ - d^+}$, in which higher moments ($n > 1$) will be useful in coordination with the leading total momentum fractions $\langle x \rangle_{u^+ - d^+}$ to maximize constraints over wide reaches of x. We also emphasize the fact that the $|S_f|(x,\mu)$ plots imply a synergistic relationship between high-energy and lattice data, with, for example, higher-order moments being especially valuable in informing fitted PDFs' x dependence near $x \approx 1$, where high-precision data can be technically difficult to obtain. Similar logic will place a high premium on lattice results obtained using the quasi-PDF approach.

3.6. Summary

As the number of observables which are reliably accessible to lattice QCD continues to grow, the necessity for PDF phenomenologists to grapple with the resulting output will be increasingly unavoidable. Rather than being a problem, the chief message of this article is that this reality promises a substantial boon to both the PDF and lattice communities. The burgeoning lattice-PDF connection is a particularly exciting development due to the potentially synergistic complementarity between the two approaches. Fundamentally, this synergy is grounded in the ability of PDF phenomenologists to push improvements in lattice calculations with benchmarks informed by high-energy data, while the lattice provides informative constraints in kinematical regions that are otherwise challenging to constrain empirically.

Before the envisioned functional relationship between the lattice and QCD analyses is fully realized, however, both communities face the serious task of establishing a common basis for comparing results from lattice QCD and global fits. The challenge implicit in this task derives from the complexity of the current landscape of available lattice calculations and global fits, which have been undertaken with a patchwork of theoretical settings, systematic assumptions, and — in the case of QCD analyses — empirical data sets. The `PDFSense` technology deployed in this article provides a standardized analysis framework in which apples-to-apples evaluations of the pulls of experimental information on lattice-calculable quantities are possible. This fact suggests one avenue for comprehensively assessing the empirical origins of phenomenological predictions of lattice data, as well as a path forward for improving them.

Thus, while some studies have usefully investigated the result of selecting some assortment of lattice data for inclusion into a global analysis framework, in this work we have essentially gone the other direction and examined the constraints data place on a collection of quantities which have been computed on the lattice. In the process, we have established several banner results:

- We have demonstrated the correspondence between phenomenological predictions for specific physical measurements and the importance of experimental information for benchmarking lattice calculations. Conversely, our impact study in the form of the sensitivity maps for the various lattice observables illustrates those regions of parameter space where improved lattice data can be expected to have a driving impact on PDF studies.

- Sec. 3.2 explored in detail the primary sources of phenomenological knowledge for the main PDF Mellin moments that are either presently accessible to the lattice or expected in the near future. In doing so, we have generally found most moments are particularly constrained by a small collection of high-impact experiments — for many of the light quark Mellin moments, for instance, a combination of HERA and fixed-target DIS data are especially decisive. We have also observed systematic tendencies in the sensitivities of high-energy data to Mellin moments, including a robust connection between the order n of the Mellin moments $\langle x^n \rangle_{q^\pm}$ of quark distributions and regions of x of the PDFs. This connection is 'bidirectional' in the sense that experimental information at higher x are likely to exert stronger pulls on higher-order Mellin moments, while lattice information on the higher-order moments may potentially constrain the high-x behavior of fitted PDFs. These observations suggest that the eventual implementation of lattice data into QCD analyses will benefit from the inclusion of information involving Mellin moments of various order for multiple parton flavors as well as knowledge gained from quasi-PDFs (qPDFs) to constrain PDFs' x dependence as widely as possible.

- Expanding this program to a novel domain, we have for the first time studied in Sec. 3.3 the driving constraints from high-energy data on calculations of the P_z-dependent qPDFs. In so doing, we have illustrated the direct link between qPDFs as theoretical quantities and the empirical information upon which calculations of the matched qPDFs from phenomenological distributions depend. One intriguing consequence of this is the possibility of more thoroughly constraining the P_z dependence of the phenomenologically-matched qPDFs with, *e.g.*, DIS measurements concentrated at large x — an undertaking that could help drive formal improvements in LaMET.

- We have constructed a comparative basis to weigh the potential impact of future experiments — for instance, the high-luminosity lepton-nucleon collider sensitivity shown in Fig. 3.12. Precise measurements that might be supplied by a high-luminosity DIS collider would be extremely beneficial for advancing phenomenological knowledge of both the moments and matched qPDFs. In particular, our analysis indicates that the main phenomenological constraints on the relevant Mellin moments arise from DIS data on fixed *nuclear* targets. Thus, the Mellin moments derived from phenomenological PDFs may be affected by nuclear corrections. As one example, Sec. 3.2.3.2 showed that the gluon Mellin moment $\langle x \rangle_g$ receives its largest constraints from DIS data measured on iron nuclei (CCFR, CDHSW), which are known to have a different preference for the large-x gluon than HERA DIS information on the proton. A future lepton-nucleus collider would enlighten such dependence of $\langle x \rangle_g$ and other moments on the nuclear environment.

As gains continue to be made on the complementary fronts of lattice theory and QCD analysis in the coming years, the analysis carried out in this work will be of value to guide phenomenologists and lattice practitioners in fully leveraging the synergy between their fields to improve our knowledge of hadron structure.

Finally, we note that a comprehensive set of results will be collected at the public URL in Ref. [116]. While this collection includes many of the calculations shown in this manuscript, a range of other sensitivity maps and related computations omitted here for brevity are also shown.

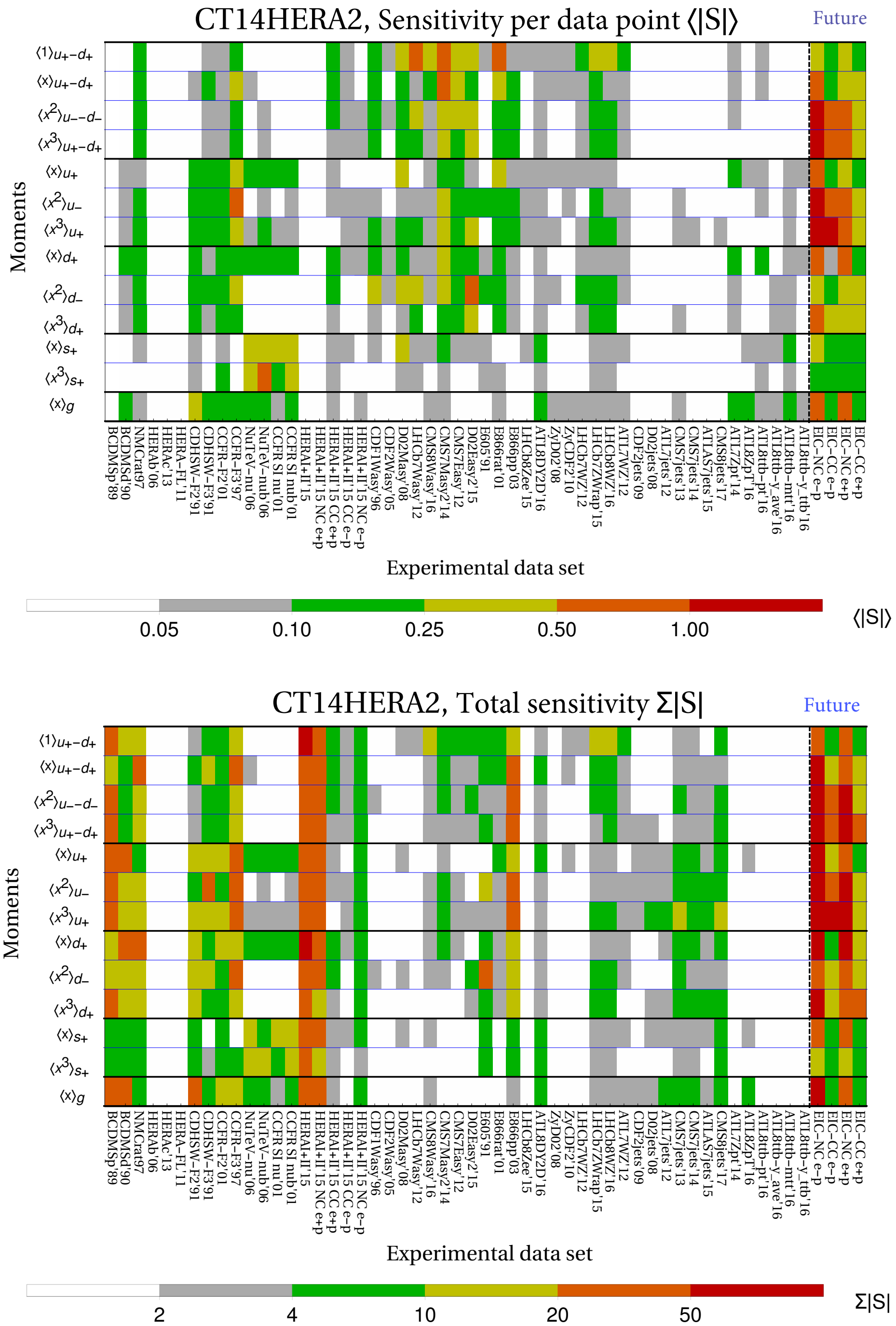


Figure 3.7: A graphical representation of the sensitivity of each of the constituent experiments contributing to the CTEQ-TEA data set. The grids summarize the point-averaged (upper panel) and summed or total (lower panel) sensitivities of each experimental data set to each moment for several flavor combinations of strong interest as indicated at the left; the color of the cell encodes the magnitude of the combined sensitivity for that particular moment. In addition, we also include in the rightmost columns the sensitivities obtained for pseudodata consistent with a future EIC-like lepton-nucleon collider experiment in the inclusive, unpolarized sector.

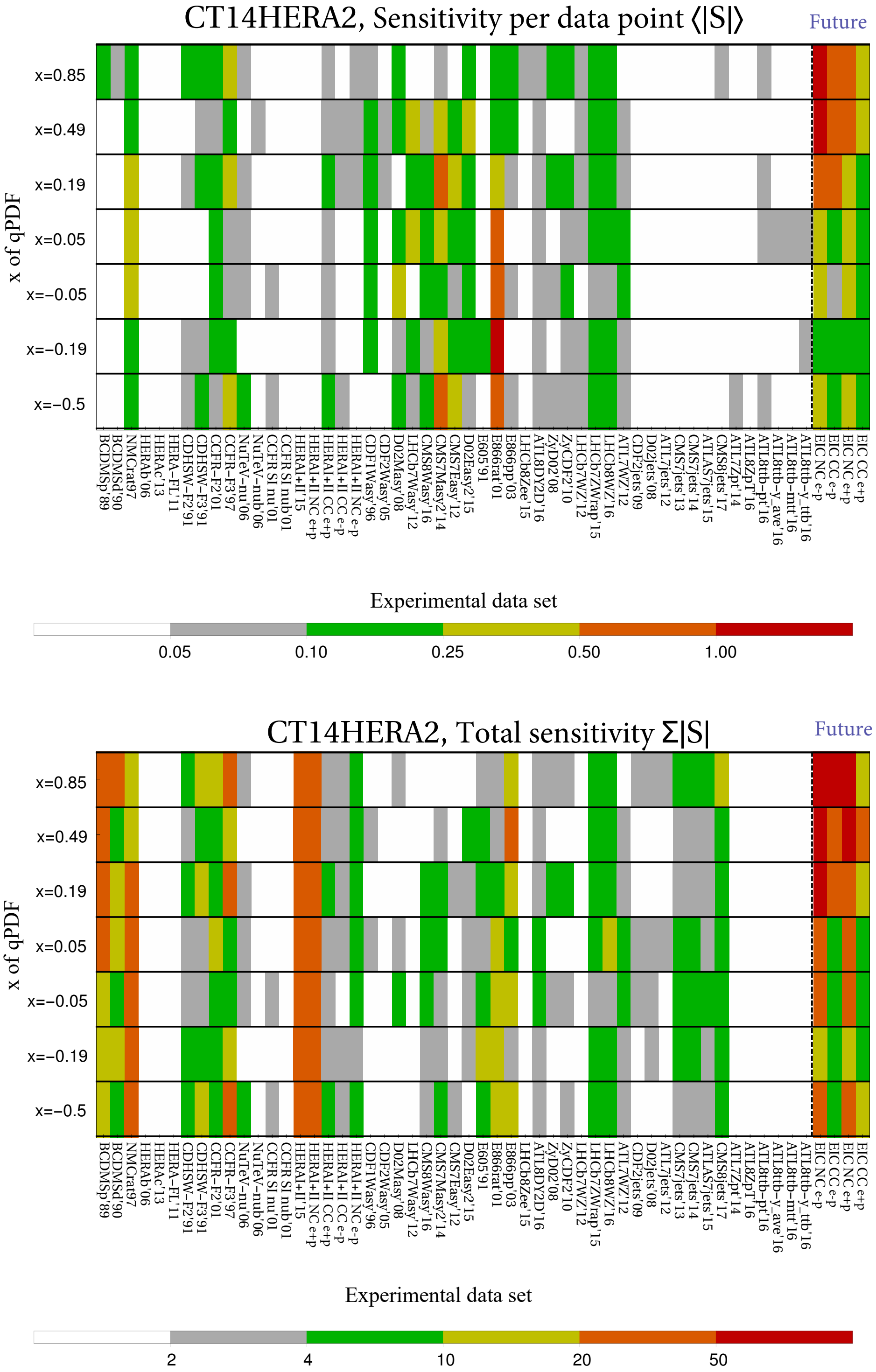


Figure 3.8: Like Fig. 3.7, but in this case illustrating the per-datum (upper panel) and aggregated (lower panel) sensitivities of the experiments within the CTEQ-TEA set to specific x regions (indicated at the left hand side of each row) of the isovector quasi-distribution at $P_z = 1.5$ GeV. As in Fig. 3.7, we again show the evaluations for pseudodata corresponding to a future EIC-like machine at the far right.

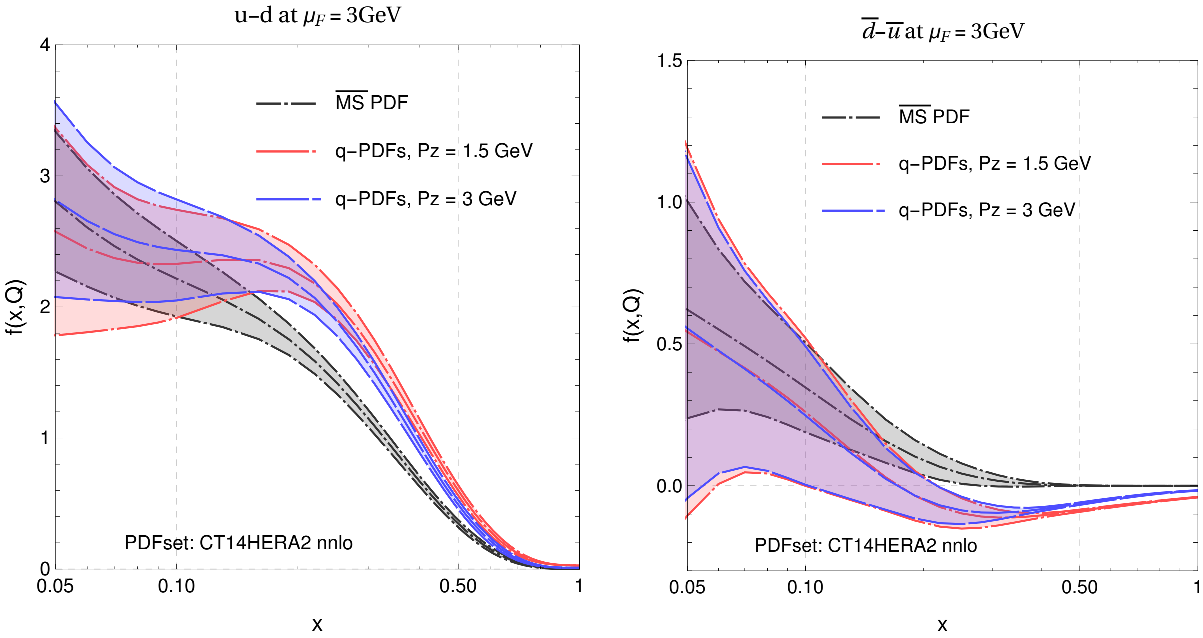
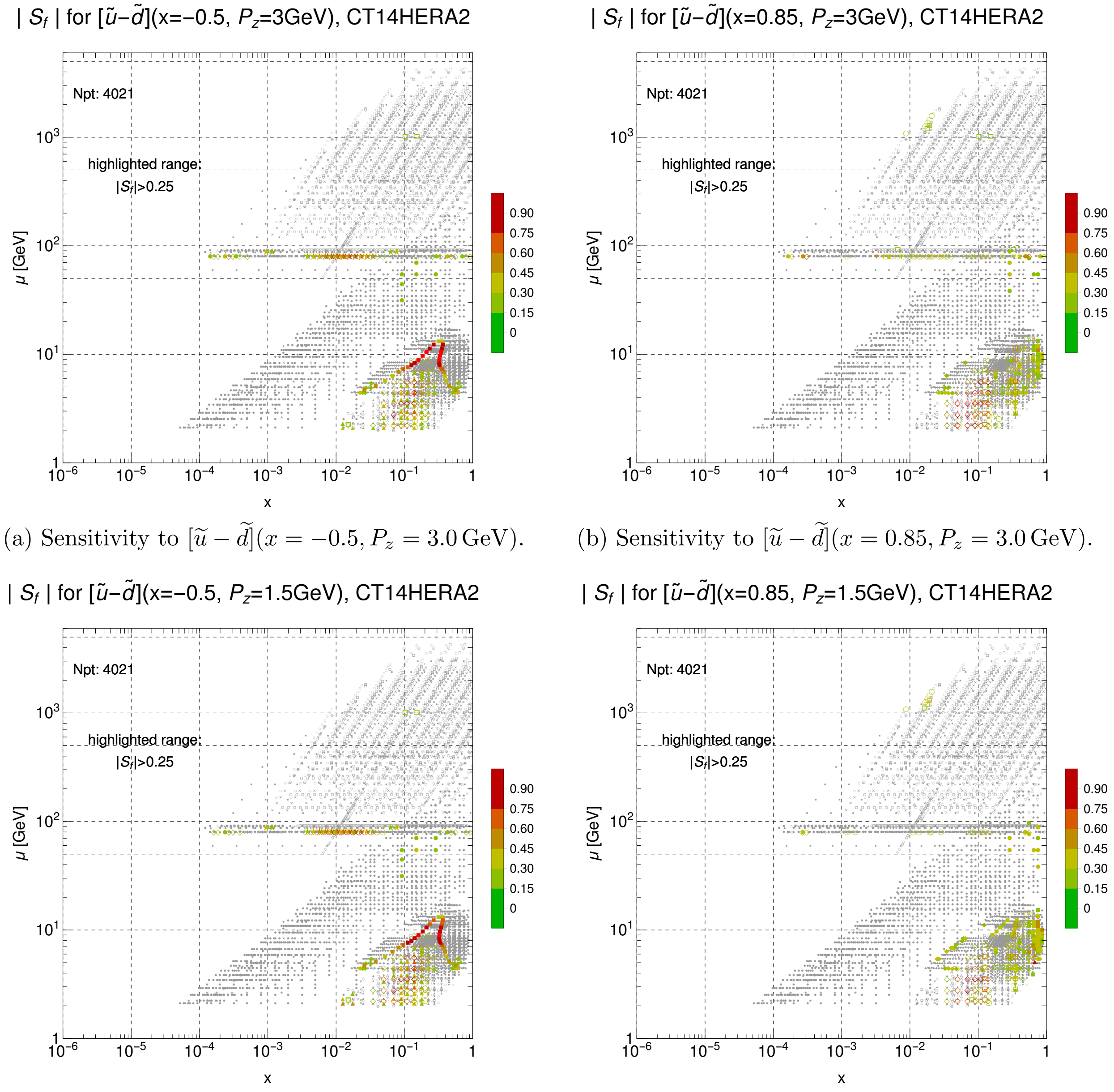


Figure 3.9: The parton quasi-distribution function $[\widetilde{u} - \widetilde{d}](x, P_z, \widetilde{\mu})$ for $x > 0$ (left panel) and $x < 0$ (right panel), with the latter computed from the $\overline{\text{MS}}$ PDF $\bar{u} - \bar{d}$ as given by CT14HERA2 NNLO.



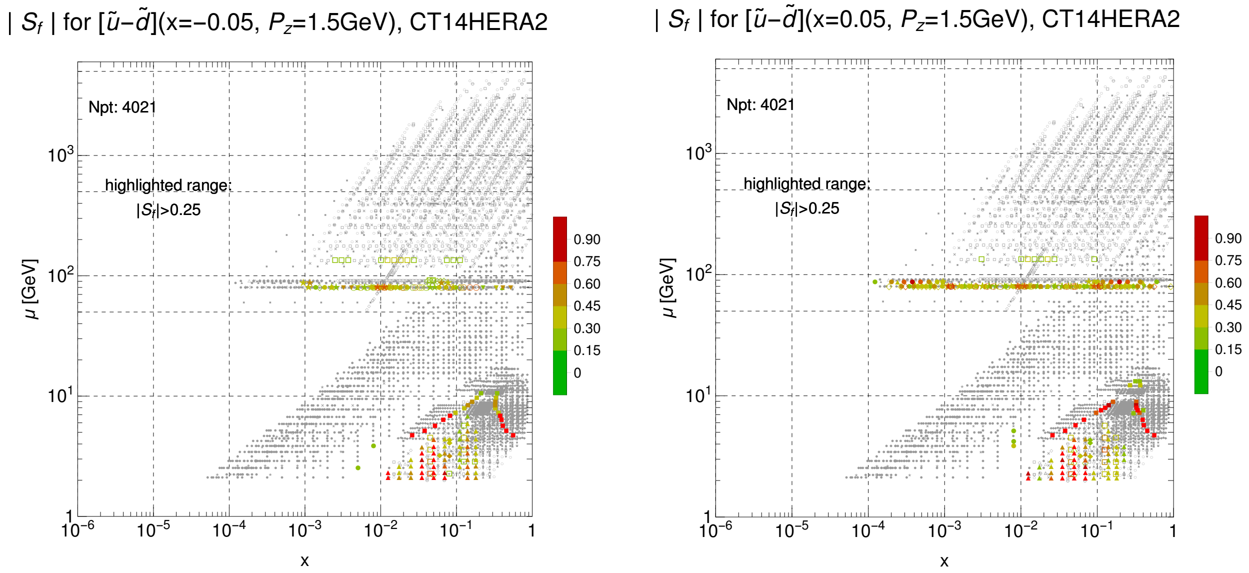
(a) Sensitivity to $[\widetilde{u}-\widetilde{d}](x=-0.5, P_z=3.0\,\mathrm{GeV})$.

(b) Sensitivity to $[\widetilde{u}-\widetilde{d}](x=0.85, P_z=3.0\,\mathrm{GeV})$.

(c) Sensitivity to $[\widetilde{u}-\widetilde{d}](x=-0.5, P_z=1.5\,\mathrm{GeV})$.

(d) Sensitivity to $[\widetilde{u}-\widetilde{d}](x=0.85, P_z=1.5\,\mathrm{GeV})$.

Figure 3.10: The isovector quark quasi-distribution at large values of $\pm x = k_z/P_z$, *i.e.*, $x = -0.5$ (left panels) and $x = 0.85$ (right panels) for a relatively fast moving proton, boosted to $P_z = 3$ GeV (top panels), and comparatively slow protons boosted to 1/2 this momentum, $P_z = 1.5$ GeV (lower panels).



(a) Sensitivity to $[\widetilde{u}-\widetilde{d}](x=-0.05, P_z=1.5\,\text{GeV})$. (b) Sensitivity to $[\widetilde{u}-\widetilde{d}](x=0.05, P_z=1.5\,\text{GeV})$.

Figure 3.11: Like Fig. 3.10 for the isovector quark quasi-distribution, but now for comparatively small values of $|x|$, in this case, $x=-0.05$ (left panel) and $x=0.05$ (right panel). Here, we plot $|S_f|$ maps only for the smaller boost scale, $P_z=1.5$ GeV, as we find the P_z dependence of the sensitivities at these smaller values of $|x|$ to be very mild.

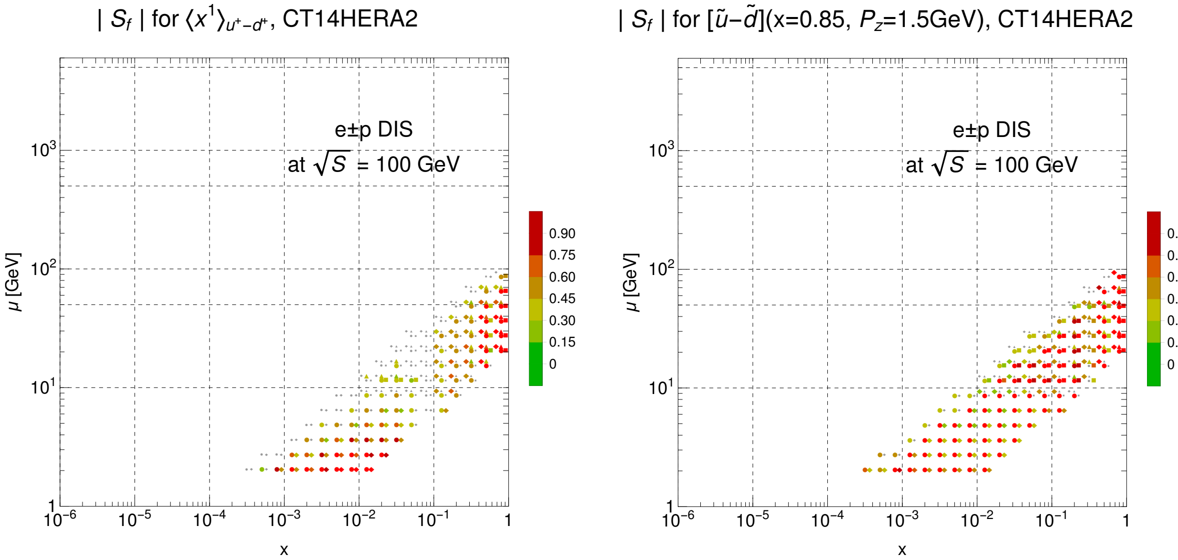


Figure 3.12: Sensitivity of pseudodata for the inclusive DIS of $e^{\pm}$ on unpolarized protons at $\sqrt{s}=100$ GeV to the first Mellin moment of the isovector PDF combination $\langle x\rangle_{u^+-d^+}$ (left) at an $\overline{\text{MS}}$ scale of $\mu=2$ GeV. The right panel shows the sensitivity to the high-x behavior of the quasi-distribution for the same isovector PDF, $[\widetilde{u}-\widetilde{d}](x,P_z,\widetilde{\mu})$ for $P_z=1.5$ GeV, $\mu=3$ GeV, and $\widetilde{\mu}$ taken from Ref. [97] computed according to Eq. (3.2). The plotted symbols characterize the specific channel as NC e^-p (disks); NC e^+p (diamonds); CC e^-p (squares); CC e^+p (triangles).

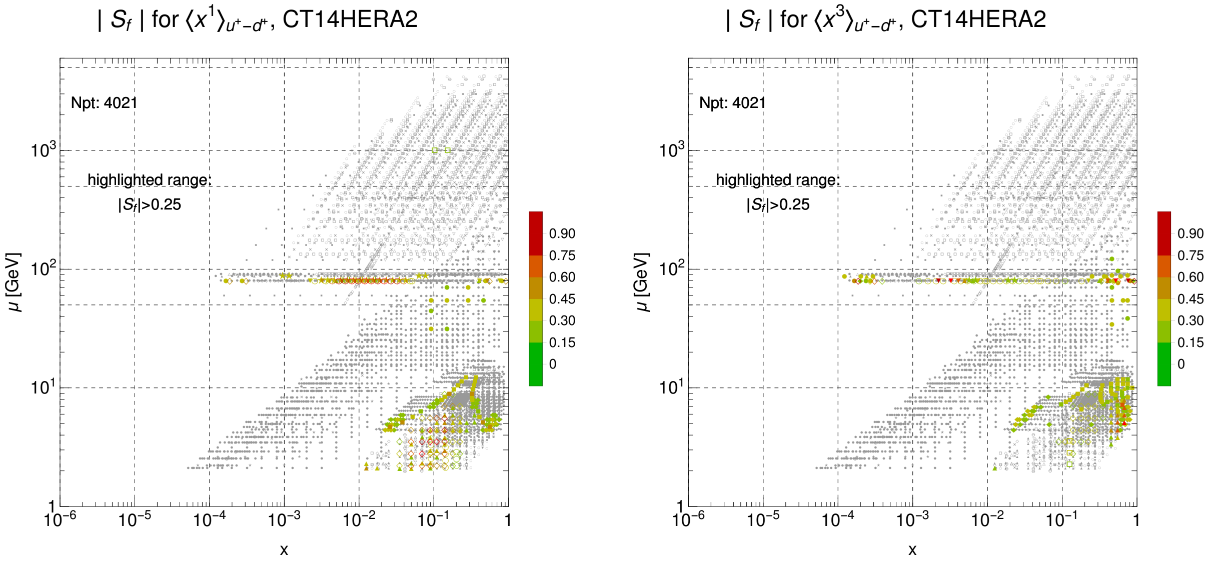


Figure 3.13: The two lowest lattice-accessible moments of the isovector PDF combination $u^+ - d^+$.

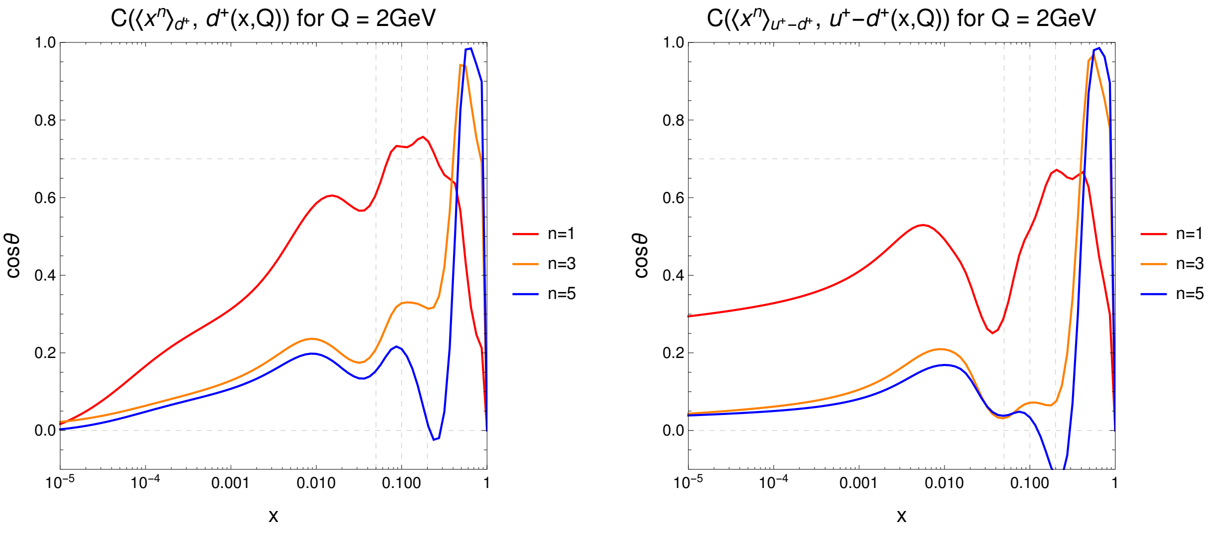


Figure 3.14: The correlation between the Mellin moments and their corresponding PDFs for the d^+ (left panel) as well as the isovector $u^+ - d^+$ (right panel) moments plotted in Fig. 3.13. As done elsewhere for the Mellin moment calculations, these are shown for $\mu_F = 2$ GeV. For each flavor, we plot the x-dependent correlation between the PDF and its integrated moment $\langle x^n \rangle_{q^+}$, for $n = 1, 3$, and 5.

Chapter 4

Collinear PDFs in the Era of HL-LHC, LHeC, and EIC

4.1. Introduction

Particle and nuclear physics today find themselves at an important crossroads. At high energies, the LHC has made tremendous progress in completing the Standard Model (SM) with the recent discovery of the Higgs boson and ongoing tests of the SM. Meanwhile, at more intermediate energies, experimental programs at JLab, RHIC, and a number of other facilities have made strides in refining our understanding of hadronic bound states and of the properties of nuclei. Despite these advances, numerous questions about fundamental physics remain. Among these are the quest to unravel the interactions of the Higgs with other SM particles, the origin of the matter-antimatter asymmetry of the universe, and the the fact that the exact nature of dark matter remains unidentified. Meanwhile, our ability to systematically relate the bulk properties of strongly-interacting matter with the quark-gluon dynamics of the underlying theory (QCD) remains limited. For these reasons, a next generation of experiments has been proposed with the objective of making decisive advances on these complementary fronts.

On the HEP phenomenology side, the planned High-Luminosity LHC (HL-LHC) [25] and a possible high-energy DIS collider, the Large Hadron-electron Collider (LHeC) [26] are expected to clarify our understanding of the electroweak sector, high-energy QCD, and perform more sensitive collider searches for new physics. In particular, the HL-LHC is expected to achieve percent-level precision in measurements of the Higgs couplings, as well as very precise determinations of electroweak observables like the W-boson mass, M_W, and weak-mixing angle, $\sin^2\theta_W$. At the same time, at medium energies near the quark-hadron transition region at $Q\sim$few GeV, an Electron-Ion Collider (EIC) [27] has been proposed for tomographic exploration of QCD bound states by essentially probing the multi-dimensional *wave function* of the nucleon and other hadrons or nuclei, a goal which necessitates the collection of enormous amount of data.

Progress along the "energy frontier" and in the quark-hadron transition region therefore entails the accumulation of very large, $\mathcal{O}(1\,\mathrm{ab}^{-1})$, data sets. This enormous quantity of data will have important impact on phenomenology at various scales. For example, high-precision measurements at an EIC will have important phenomenological implications for high-energy measurements, while explorations at very high scales carried out at HL-LHC or LHeC can provide constraints to nucleon structure observables.

Due to the possible syntheses and complementary impacts among these different programs spanning the energy spectrum, detailed accounting of possible overlaps and different pulls originating with each proposal is an urgent necessity. We explore this issue for unpolarized proton PDFs — the essential nonperturbative input into theoretical predictions of HEP observables at the LHC. In particular, we concentrate on the pulls of pseudodata representative of each of the programs noted above. To perform this analysis, we use the `PDFSense` framework developed in chapter 2 (or Ref. [34]), following the conventions established therein. Additional details, including possible synergies between precise EIC measurements and lattice QCD calculations — discussed in passing here — can be found in chapter 3 (or Ref. [117]). These proceedings may be read in parallel with the talk upon which they are based, given as Ref. [118].

4.2. Future HEP Programs and the Nucleon's Collinear Structure

As of the time of this writing, the HL-LHC represents a planned upgrade to the LHC, with the ultimate goal being a factor $\sim$5 improvement upon the LHC's instantaneous luminosity, generating an integrated luminosity as large as $\mathcal{L} = 3\,\mathrm{ab}^{-1}$. Like the LHC Run 2, the HL-LHC will measure pp events at $\sqrt{s} = 14$ TeV, probing a wide range of parton momentum fractions, $5 \cdot 10^{-5} < x < 1$, at large factorization scales μ. The LHeC proposal, on the other hand, envisions the incorporation of a charged lepton beam to scatter $E_{e^\pm} = 60$ GeV $e^\pm$ off protons in the LHC main hadron ring. This scenario would result in a high-luminosity collider capable of reaching unprecedented (for DIS) energies of $\sqrt{s} = \sqrt{4E_{e^\pm}E_p} = 1.3$ TeV; in turn, this would afford access to very low $x \gtrsim 5 \cdot 10^{-6}$. The HL-LHC and LHeC programs are differentiated by the physics they would be adapted to explore. We elucidate their differences by implementing an assumed $\mathcal{L} = 100\,\mathrm{fb}^{-1}$ of unpolarized electron/positron scattering pseudodata from Ref. [119], as well as NNLO theory predictions for HL-LHC pseudodata [120] at $\mathcal{L} = 3\,\mathrm{ab}^{-1}$ collected over a varied set of typical LHC processes and experiments; the resulting sensitivities can then be directly compared in the `PDFSense` framework in chapter 2 (or [34]). The complementary, mostly non-overlapping regions of leading kinematic sensitivity for the HL-LHC vs. LHeC can be visualized by comparing kinematic distributions of experimental data points in the (x, μ) plane. Figure 4.1 plots the sensitivity $|S_f|$ of the LHeC (left panel) and HL-LHC (right panel) pseudodata to the d-quark distribution, $d(x, \mu)$, as estimated using PDF4LHC15 NNLO PDFs [24]. We show this figure as a representative example of the differentiated pulls on the PDFs by the LHeC and HL-LHC pseudodata. Similar comparisons can be made between HL-LHC and LHeC for PDFs of the gluon and other quark flavors; a subset of these are shown in Slides 14-17 of Ref. [118], and we refer the interested reader to those plots.

For the HL-LHC information in the right panel, the key strength is the coverage at the highest energy scales, $\mu \sim 10$ TeV, for intermediate $x \gtrsim 0.01$. Meanwhile, the LHeC pseudodata shown on the left probe PDFs at $x \lesssim 10^{-5}$, beyond the lower reach of HERA, primarily via neutral-current (NC) exchanges. In addition, at $x > 0.1$, the charged-current (CC) DIS reduced cross section has the form

$$\sigma^{e^+p}_{r,\,\mathrm{CC}} = \frac{Y_+}{2} W_2^+ \mp \frac{Y_-}{2} x W_3^+ - \frac{y^2}{2} W_L^+ \simeq [1-y]^2\, x(d+s) + x(\overline{u} + \overline{c})\ , \qquad (4.1)$$

where $Y_{\pm} = 1 \pm (1-y)^2$, and from which one may infer $\sigma^{e^+p}_{r,\,\mathrm{CC}} \simeq xd\,[1-y]^2$ in the $x \to 1$ limit. The combination of NC and CC DIS at the LHeC would enjoy extensive coverage to perform the u and d separation in the high $x > 0.1$ region over two decades of scales, $10^1 < \mu < 10^3$ GeV. We thus see the complimentarity of the LHeC and the HL-LHC kinematic coverages for $d(x, \mu)$ in Fig. 4.1. It should be stressed that the sensitivity to the high-x d-quark distribution at the LHeC comes without the potential ambiguities of a nuclear target (in contrast to typical extractions of F_2^n from DIS on the deuteron) and via a "cleaner" electroweak probe furnished by lepton-nucleon DIS (rather than pp collisions). The HL-LHC pseudodata in the right panel would probe the intervening regions between the highest and lowest x constrained by the LHeC, with the leading input especially coming from high-luminosity jet production at $\mu > 10^2$ GeV and $10^{-3} < x < 10^{-1}$. Precise data on $W^{\pm}$ hadroproduction extends the HL-LHC d-PDF sensitivity to high and low x, albeit over a comparatively narrower range of factorization scales.

The NC LHeC pseudodata at $x < 10^{-3}$ exhibits strong sensitivities to the gluon distribution (via Bjorken scaling violation) and the singlet quark PDF. Like the HERA program, the breadth of x and Q^2 over which DIS observables would be accessible at the LHeC offers a strong empirical "lever-arm" to test QCD evolution and the parameters of the theory in the perturbative region.

It can also be instructive to compare the *aggregated* sensitivity, $\mathbf{S}_f \equiv \sum_i |S^i_f|$, of each set of data to various PDFs to give an approximate sense for the total pull the pseudodata can be expected to have in QCD analyses. By this figure of merit, the potential for mutual complementarity between the flavor sensitivities of the HL-LHC and LHeC programs is strongly suggested [118]. For the total pull on the PDF4LHC15 gluon PDF, $\mathbf{S}_g$, we find $\mathbf{S}_g^{\mathrm{HL-LHC}} = 245 > \mathbf{S}_g^{\mathrm{LHeC}} = 151$, implying a potential advantage for the HL-LHC in constraining the gluon. For the d-quark PDF plotted in Fig. 4.1, this balance is reversed, with a larger aggregated sensitivity for LHeC, $\mathbf{S}_g^{\mathrm{LHeC}} = 214 > \mathbf{S}_g^{\mathrm{HL-LHC}} = 171$. LHeC

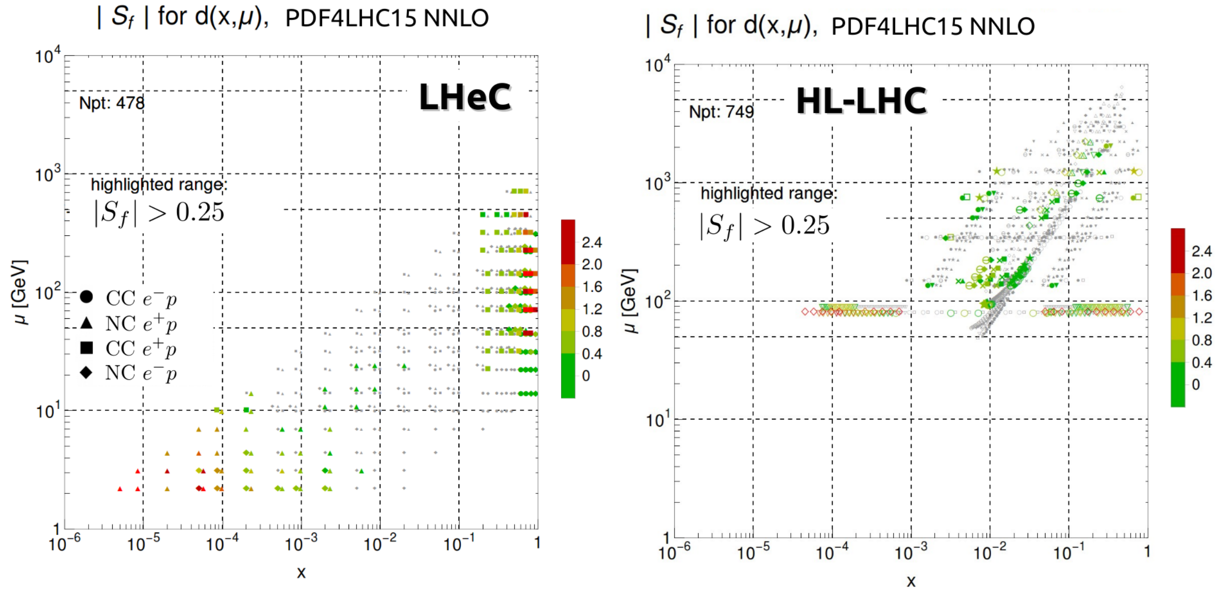


Figure 4.1: Future HEP experiments like the HL-LHC and possible LHeC proposal can have substantial PDF sensitivity, as shown here for the PDF4LHC15 $d(x, \mu)$ distribution computed according to the conventions in chapter 2 (or Ref. [34]). The panels display the x- and μ-dependent sensitivity of pseudodata for LHeC (left panel) and HL-LHC (right panel).

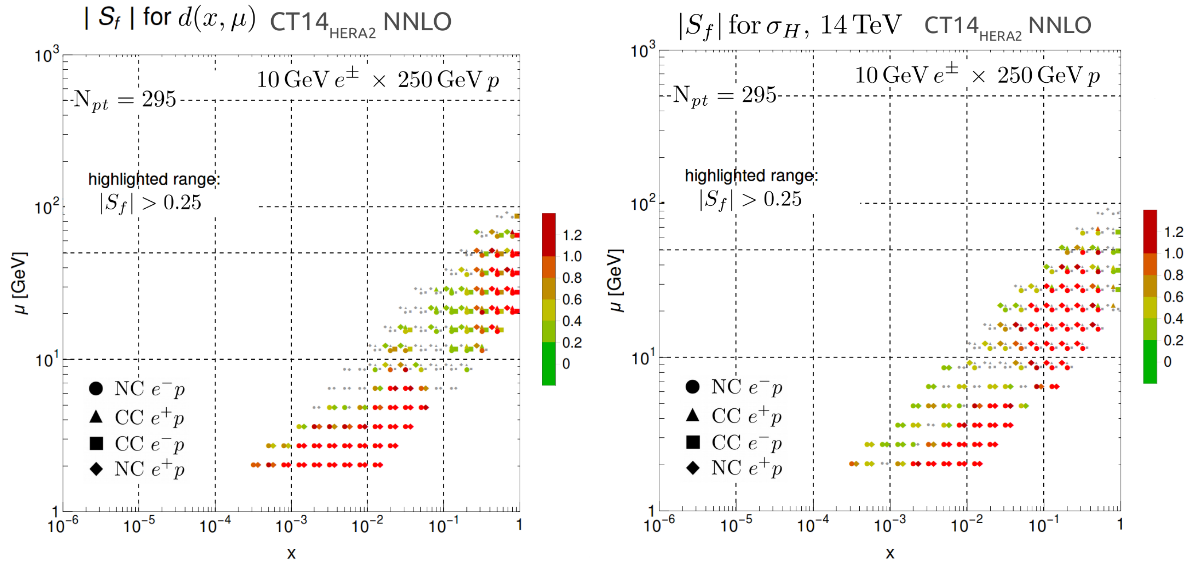


Figure 4.2: As a counterpart to the results plotted in Fig. 4.1, we show the EIC-like pseudodata sensitivity to $d(x, \mu)$ [left panel] as well as to $\sigma_H(14\,\text{TeV})$, the total cross section for Higgs production at 14 TeV [right panel]. The pseudodata assume an integrated luminosity of $\mathcal{L} = 100\,\text{fb}^{-1}$.

similarly enjoys a modest advantage in unfolding nucleon strangeness, while there is close parity in the sensitivity to the $\bar{d}$, $\bar{u}$ distributions. It should be emphasized, however, that the integrated luminosity assumed here for the LHeC pseudodata ($\mathcal{L} = 100\,\mathrm{fb}^{-1}$) is a fraction of what a full LHeC data-collection campaign could ultimately achieve. To the extent that LHeC errors might be statistics-limited, these projections should be considered lower bounds.

4.3. The HEP Implication of an EIC

The scope of measurements to be undertaken at an EIC will have powerful implications for PDFs, and, by extension, it will impact upon HEP phenomenology at the LHC and elsewhere. We highlight the impact on the unpolarized nucleon PDFs that one might expect with a generic, $\sqrt{s} = 100$ GeV DIS collider that is broadly consistent with the proposed design profiles for either the JLab-based (JLEIC) or BNL-based (eRHIC) incarnations of the EIC. In Fig. 3.12, we plot the sensitivity of EIC pseudodata to the CT14HERA2 NNLO PDF [41] for the d quark in the left panel, as well as the PDF-driven sensitivity to the high-energy Higgs production cross section, $\sigma_H(14\,\mathrm{TeV})$, in the right panel. Here, an upper scale choice of $|S_f|= 1.2$ for highlighted points is chosen. By the panels of Fig. 3.12, the powerful sensitivity of the EIC pseudodata generally surpasses that of the fixed-target experiments that currently dominate the constraints on high-x PDFs. The EIC will therefore strongly constrain PDF dependence of HEP observables at moderate and large x, including several in the Higgs and electroweak sectors like M_W and $\sin^2\theta_W$. As an emblematic example of this, the right panel of Fig. 3.12 shows the substantial EIC sensitivity to the total Higgs production cross section at the LHC, $\sigma_H(14\,\mathrm{TeV})$. The constraints that a medium-energy machine like an EIC would place on Higgs phenomenology stem from the predominance of the $gg \to H$ fusion channel in $\sigma_H(14\,\mathrm{TeV})$. In particular, $\sigma_H(14\,\mathrm{TeV}) = 62.1\,\mathrm{pb}$, of which $88 \pm 4\%$ emanates from gluon fusion. While the leading sensitivity to Higgs production at the LHC is expected to originate from the "Higgs region" at $\mu = m_H = 125$ GeV and

$x \sim m_H/(14\,\text{TeV}) \sim 0.01$, QCD evolution connects the gluon PDF behavior at such x and μ to the behavior at the lower μ and higher x that will be probed by the EIC.

Just as the EIC, *lattice QCD* calculations aim at detailed understanding of the structure, spectrum, and interactions of the nucleon, lighter hadrons, and their excitations. The recent analysis of chapter 3 (or Ref. [117]), briefly summarized in Slides 18-25 of Ref. [118], has explored the possibility of a future synergy between phenomenology informed by the EIC data and lattice calculations of integrated PDF moments and quasi-PDFs, $\widetilde{q}(x, \mu, P_z)$. Cooperation between lattice studies and an EIC tomography program would constrain nucleon PDFs necessary for HEP phenomenology. Advances in lattice QCD techniques driven by EIC-improved benchmarks could similarly feed-forward into lattice calculations of, *e.g.*, branching ratios or quantities sensitive to CP-violation.

4.4. Summary

We reiterate our principle finding: the HL-LHC, EIC, and LHeC each have unique and complementary access to a broad range of physics as embodied by their PDF pulls. The HL-LHC will be distinguished by its reach to the highest μ scales in a variety of pp processes with strong sensitivity to the gluon and $\bar{u}$, $\bar{d}$ distributions. On the other hand, the LHeC would leverage its combination of LHC energies with the DIS collider process to access very low and high x, with especially strong impact on the g, u, d, and s parton densities. The combination of the HL-LHC and LHeC constraints on PDFs across a wide range of x and μ will be vital for the high-precision energy-frontier physics program that the HL-LHC will pursue. The complementarity we find between the HL-LHC and LHeC is broadly consistent with the results recently reported in Ref. [121]. Lastly, by pursuing hadron tomography with extremely high precision using polarized beams, an EIC would supersede the bulk of fixed-target DIS experiments, providing critical information needed to disentangle the nucleon's nonperturbative structure.

Chapter 5

Conclusion

From the historical experience, the progress of fundamental science often leads to technological breakthroughs. There is no exception to the exploration of the basic structure of matter. The understanding of atom-related knowledge promotes the development of chemistry. The understanding of the nuclei inside atoms brings about nuclear and nuclear magnetic resonance techniques. Therefore, we have reason to believe that PDF-related knowledge will bring about a technological revolution in the future. Also, PDF research and other important areas of fundamental science are closely related. These areas include particle physics, astronomy, and cosmology. Therefore, the field of PDFs, as an infrastructure for science and technology, deserves the manpower and funding to invest in research.

In Chapter 2, we have developed several statistical approaches to promote QCD phenomenological analysis and complement the classic global QCD analysis. Everyone can use these approaches from the open source software I developed. In addition, in Chapter 3, we have developed a framework to help the cross-disciplinary integration between phenemenological PDFs and Lattice QCD. This framework provides a bi-directional assessment of how two areas can improve each other's progress. Finally, in Chapter 4, we have provided the analysis and comparison needed to plan and design future high-energy experimental facilities (HL-LHC, LHeC, and EIC). The analysis tells us, through the constraints on the knowledge of the nucleon structure, what theoretical issues we can solve and what opportunities for discovery these experimental facilities can provide.

The various statistical tools presented in Chapter 2 can provide different perspectives on the relationship among physical issues related to QCD. In particular, we borrowed several statistical tools in data science, and they brought various ways to explore the relationship between various information carried by data (e.g. I demonstrate how to learn, with PCA and T-SNE, which data points are related to similar PDF parameter space). Therefore, the statistical tools that have been explored and not yet explored can not only help the analysis of sub-nuclear physics, but they are also suitable for integrating various physical fields, such as Higgs physics, the BSMs, and QCD. My work in Chapter 3 demonstrates a case study of multidisciplinary integration. This work opens up new avenues to find out what phenomenological knowledge and theoretical calculations can learn from each other. This kind of analyses leverage the improvement of knowledge in experimental activities.

In addition, these statistical tools can help others to look at data in new ways. People will be able to gain a broader view of data, which maximizes the value that data can provide. It is worth mentioning that the ability of methods and services in data science (open source softwares or online services) to mine high level features will bring more benefits, but we have not fully exploited the potential of these methods and services. Further exploring the applications of these methods and services in high-energy experimental data will greatly promote the progress in areas including but not limited to physics. For example, the analyses of experimental data or data with large sizes can also draw on the experience of high-energy experiments.

Finally, the evaluation of future experimental programs based on these statistical tools (in Chapter 4) will help us to predict the potential of the experimental programs to discover new physics and the issues the experimental activities will be able to clarify. Therefore, the results of these analyses will contribute to the strategic planning of future high-energy experiments. These strategic plans include an assessment of the expected return on investment in these experimental facilities, determining which experimental facilities we should build in the future, and how to design these experimental facilities and prioritize these designs to maximize the benefits of the facilities.

5.1. My Main Contribution

In chapter 2, I developed most of the features in PDFSense, which includes statistical tools and Mathematica scripts that can be executed from the command line. In addition, I implemented a sensitivity analysis for CTEQ-TEA datasets. Finally, I demonstrated several examples of using statistical methods to obtain insights into relationships between experimental processes. In chapter 3, I compared $\overline{\text{MS}}$ PDFs and P_z-dependent qPDFs, analyzed sensitivities of CTEQ-TEA datasets and pseudodata of several future experimental proposals to Mellin Moments and qPDFs, and specified the x-ranges in which Lattice Mellin moments can constrain the most on PDFs. In chapter 4, I surveyed the sensitivities of pseudodata of future experimental facilities to PDFs. In addition to the work mentioned in my thesis, I also participated in several other projects. By comparing LHC cross sections that depend on PDF4LHC15 and other various PDF sets, I helped the benchmark study of PDF4LHC15 PDFs [122]. These efforts can contribute to the advancement of PDFs and high-energy physics phenomenology.

5.2. Outlook

PDFSense currently only analyzes collinear nucleon PDFs. We can expand the types of PDFs it can analyze, and nuclear PDFs [123] [124] is a suitable object. The first is that both collinear nuclear PDFs and collinear nucleon PDFs are functions of $\{x, \mu\}$, so our method is easy to transport to nuclear PDFs. In addition, Nuclear effects are not only important for theoretical predictions involving nuclei (like electron-ion collisions of EIC and pPb collisions of LHC), but also for the advancement of knowledge of nulceon PDFs [124].

Another object that can be extended is the transverse momentum dependent distribution functions (TMDs) [125]. TMDs provide important information about the three-dimensional structure of hadrons. The knowledge of TMDs can be applied to many collision processes that require QCD factorization theorems. Promoting the development of TMDs will be very helpful for LHC phenomenology and spin physics.

Using PDFSense to explore high-energy experiments to help many of the physical issues related to PDFs is another direction that we can develop. An example of this is the Higgs cross sections I have analyzed in Chapter 2. Higgs physics is the main research direction of LHC, and it is the key to verifying the SM and searching for new physics [126] [127]. Other physical quantities we can explore are physical quantities that may benefit from more accurate PDFs, such as weak mixing angles and some of the BSM's observables [128] [129].

Appendix A

Approximate Kinematical Variables

In this section, we describe in detail our method for identifying the values of $\{x_i, \mu_i\}$ that correspond to experimental data.

For each experimental data point i, we can establish an approximate relation between the kinematical quantities for that data point, and unobserved quantities specifying the PDFs: the partonic momentum fraction x and QCD factorization scale μ. For example, in DIS, x and μ are approximately equal to Bjorken x_B and momentum transfer Q according to the Born-level kinematic relation. Although this relation is violated by higher-order radiative contributions, it will approximately hold in most scattering events. The same overall logic can be followed to relate the kinematical quantities in every process of the CTEQ-TEA global set to the *approximate* unobserved quantities x and μ in the PDFs. These relations vary by process and are used to assign approximate pairs $\{x_i, \mu_i\}$ for each data point.[1]

Specifically, for DIS, which primarily measures the differential cross sections of the form $d^2\sigma/(dx_B dQ^2)$, we simply take

$$\mu_i \approx Q|_i\,,\ x_i \approx x_B|_i \tag{A.1}$$

[1] It should be pointed out that, while there are 5227 $\{x, \mu\}$ points generated by the 4021 physical measurements in the default CTEQ-TEA dataset of this study, occasionally there are instances in which $|C_f|$ and $|S_f|$ cannot be meaningfully computed for select flavors. For example, since the bottom quark PDF $b(x, \mu)$ has no sensible definition below its partonic threshold (*i.e.*, for $\mu < m_b = 4.75$ GeV), it is not possible to evaluate $|S_b|$ for data points extracted at μ scales below the b-quark mass. Similarly, there are situations when the extracted parton fraction $x_i \approx 1$, such that some PDF flavors $f(x_i, \mu_i) \approx 0$, and the Hessian procedures described in this paper do not yield a well-defined correlation or sensitivity. In these cases, we simply redact the associated $\{x_i, \mu_i\}$ points.

as mentioned above, where the kinematical variables inside "$|_i$" are evaluated at their experimentally measured values for the i^{th} data point. The above approximate relations hold even when (N)NLO radiative contributions are included.

For one-particle-inclusive particle production in hadron-hadron scattering of the form $AB \to CX$, we plot two x values if the rapidity y_C is known:

$$\mu_i \approx Q|_i\,,\ x_i^{\pm} \approx \frac{Q}{\sqrt{s}} \exp(\pm y_C)\bigg|_i\,. \tag{A.2}$$

We set $y_C = 0$ if the rapidity is integrated away. We point out that for processes of this type, Eq. (A.2) implies that a measurement in a single rapidity bin can in fact probe two distinct values of x; for this and other potential reasons, the number of raw data points in such an experiment (N_{pt}) should not be expected to match the number of extracted $\{x, \mu\}$ points in the figures.

In vector boson production, $AB \to (\gamma^*, Z \to \ell\bar{\ell})X$ or $AB \to (W \to \ell\nu_\ell)X$, we set $Q = m_{\ell\bar{\ell}}$ (invariant mass of the lepton pair), and $y_C = y_\ell$ if a single-lepton rapidity is provided or $y_C = y_{\ell\bar{\ell}}$ if the lepton-pair rapidity is provided. If the rapidity y_ℓ of the lepton is known, yet $y_{\ell\bar{\ell}}$ of the pair is unknown, we use the fact that $y_\ell \sim y_{\ell\bar{\ell}} \pm 1$ for most events because of the shape of the decay leptonic tensor. Thus, the momentum fractions $x_i^{\pm}$ can still be estimated as $x_i^{\pm} \approx (Q/\sqrt{s}) \exp(\pm y)|_i$, where $y \sim y_\ell$ (up to an error of less than 1 unit)

In single-inclusive jet production, $AB \to j + X$, we set $Q = 2p_{Tj}$, $y_C = y_j$.

In single-inclusive $t\bar{t}$ pair production, $AB \to t\bar{t}X$, we set $Q = m_{t\bar{t}}$, $y = y_{t\bar{t}}$ if known, or 0 otherwise.

In single-inclusive top (anti-)quark production, $AB \to (\bar{t})tX$, we take $Q = 2p_{Tt}$, $y = 0$ for $d\sigma/dp_{T_t}$ (as in Expt. ATL8ttb-pt'16). On the other hand, for $d\sigma/d\langle y_t\rangle$ or $d\sigma/dy_{t\bar{t}}$, in which the $t\bar{t}$ invariant mass is integrated out (Expts. ATL8ttb-y_ave'16 and ATL8ttb-y_ttb'16), we take an average mass scale $\mu_i = 400$ GeV that is slightly above the observed peak of $d\sigma/dm_{t\bar{t}}$ at $m_{t\bar{t}} \approx 2m_t$.

Lastly, for the $d\sigma/dp_T^Z$ measurements from $AB \to (\gamma^*, Z \to \ell\bar{\ell})X$ in Expts. ATL7ZpT'14 and ATL8ZpT'16, we take $Q = \sqrt{(p_T^Z)^2 + (M_Z)^2}$, $y_C = y_Z$. [Here Q denotes the boson's transverse mass, not the invariant mass.]

Appendix B

Tabulated Results

B.1. Main Material

In Tables B.1–B.3 we provide a detailed key for the individual experiments mapped in Fig. 2.1, including the physical process, number of points, and luminosities, where available. We group these tables broadly according to subprocess — Table B.1 corresponds to DIS experiments, while Tables B.2 and B.3 collect various measurements for the hadroproduction of, *e.g.*, gauge boson, jet, and $t\bar{t}$ pairs — and thus provide a translation key for the experimental short-hand names given in Fig. 2.1.

In Tables B.4 and B.5, we collect the flavor-specific ($|S_f^E|$) and overall ($\sum_f |S_f^E|$) sensitivities for the experimental datasets contained in this analysis. In Table B.4 we list the total and point-averaged sensitivities for each main flavor ($\bar{d}, \bar{u}, g, u, d, s$), while Table B.5 gives the corresponding information for a number of quantities derived from these, as explained in the associated captions.

Table B.1: Experimental datasets considered as part of CT14HERA2 and included in this analysis: deep-inelastic scattering. We point out that the numbering scheme (CT ID#) included in this and subsequent tables follows the standard CTEQ labeling system with, *e.g.*, Expt. IDs of the form 1XX representing DIS experiments, *etc.* The HERA combined data set HERAI+II'15 consists of both neutral-current (NC) and charge-current (CC) scattering events.

Experiment name	**CT ID#**	**Dataset details**		N_{pt}
BCDMSp'89	101	BCDMS F_2^p	[130]	337
BCDMSd'90	102	BCDMS F_2^d	[131]	250

Experiment name	**CT ID#**	**Dataset details**	N_{pt}
NMCrat'97	104	NMC F_2^d/F_2^p [132]	123
CDHSW-F2'91	108	CDHSW F_2^p [133]	85
CDHSW-F3'91	109	CDHSW F_3^p [133]	96
CCFR-F2'01	110	CCFR F_2^p [134]	69
CCFR-F3'97	111	CCFR xF_3^p [135]	86
NuTeV-nu'06	124	NuTeV $\nu\mu\mu$ SIDIS [136]	38
NuTeV-nub'06	125	NuTeV $\bar{\nu}\mu\mu$ SIDIS [136]	33
CCFR SI nu'01	126	CCFR $\nu\mu\mu$ SIDIS [137]	40
CCFR SI nub'01	127	CCFR $\bar{\nu}\mu\mu$ SIDIS [137]	38
HERAb'06	145	H1 σ_r^b (57.4 pb^{-1}) [138] [139]	10
HERAc'13	147	Combined HERA charm production (1.504 fb^{-1}) [140]	47
HERAI+II'15	160	HERA1+2 Combined NC and CC DIS (1 fb^{-1}) [39]	1120
HERA-FL'11	169	H1 F_L (121.6 pb^{-1}) [141]	9

Table B.2: Same as Table B.1, showing experimental datasets for production of vector bosons, single-inclusive jets, and $t\bar{t}$ pairs.

Experiment name	**CT ID#**	**Dataset details**	N_{pt}
E605'91	201	E605 DY [142]	119
E866rat'01	203	E866 DY, $\sigma_{pd}/(2\sigma_{pp})$ [143]	15
E866pp'03	204	E866 DY, $Q^3 d^2\sigma_{pp}/(dQdx_F)$ [144]	184
CDF1Wasy'96	225	CDF Run-1 $A_e(\eta^e)$ (110 pb^{-1}) [145]	11
CDF2Wasy'05	227	CDF Run-2 $A_e(\eta^e)$ (170 pb^{-1}) [146]	11
D02Masy'08	234	D$\emptyset$ Run-2 $A_\mu(\eta^\mu)$ (0.3 fb^{-1}) [147]	9
LHCb7WZ'12	240	LHCb 7 TeV W/Z muon forward-η Xsec (35 pb^{-1}) [148]	14

Experiment name	**CT ID#**	**Dataset details**	N_{pt}
LHCb7Wasy'12	241	LHCb 7 TeV W $A_\mu(\eta^\mu)$ (35 pb^{-1}) [148]	5
ZyD02'08	260	DØ Run-2 Z $d\sigma/dy_Z$ (0.4 fb^{-1}) [149]	28
ZyCDF2'10	261	CDF Run-2 Z $d\sigma/dy_Z$ (2.1 fb^{-1}) [150]	29
CMS7Masy2'14	266	CMS 7 TeV $A_\mu(\eta)$ (4.7 fb^{-1}) [151]	11
CMS7Easy'12	267	CMS 7 TeV $A_e(\eta)$ (0.840 fb^{-1}) [152]	11
ATL7WZ'12	268	ATLAS 7 TeV W/Z Xsec, $A_\mu(\eta)$ (35 pb^{-1}) [153]	41
D02Easy2'15	281	DØ Run-2 $A_e(\eta)$ (9.7 fb^{-1}) [154]	13
CDF2jets'09	504	CDF Run-2 incl. jet ($d^2\sigma/dp_T^j dy_j$) (1.13 fb^{-1}) [155]	72
D02jets'08	514	DØ Run-2 incl. jet ($d^2\sigma/dp_T^j dy_j$) (0.7 fb^{-1}) [156]	110
ATL7jets'12	535	ATLAS 7 TeV incl. jet ($d^2\sigma/dp_T^j dy_j$) (35 pb^{-1}) [157]	90
CMS7jets'13	538	CMS 7 TeV incl. jet ($d^2\sigma/dp_T^j dy_j$) (5 fb^{-1}) [158]	133

Table B.3: Same as Table B.1, showing experimental datasets for production of vector bosons, single-inclusive jets, and $t\bar{t}$ pairs that were not incorporated in the CT14HERA2 fit but included in our augmented CTEQ-TEA set.

Experiment name	**CT ID#**	**Dataset details**	N_{pt}
LHCb7ZWrap'15	**245**	LHCb 7 TeV Z/W muon forward-η Xsec (1.0 fb^{-1}) [159]	33
LHCb8Zee'15	**246**	LHCb 8 TeV Z electron forward-η $d\sigma/dy_Z$ (2.0 fb^{-1}) [160]	17
ATL7ZpT'14	**247**	ATLAS 7 TeV $d\sigma/dp_T^Z$ (4.7 fb^{-1}) [161]	8
CMS8Wasy'16	**249**	CMS 8 TeV W muon, Xsec, $A_\mu(\eta^\mu)$ (18.8 fb^{-1}) [162]	33
LHCb8WZ'16	**250**	LHCb 8 TeV W/Z muon, Xsec, $A_\mu(\eta^\mu)$ (2.0 fb^{-1}) [163]	42
ATL8DY2D'16	**252**	ATLAS 8 TeV Z ($d^2\sigma/d\|y\|_{ll}dm_{ll}$) (20.3 fb^{-1}) [164]	48
ATL8ZpT'16	**253**	ATLAS 8 TeV ($d^2\sigma/dp_T^Z dm_{ll}$) (20.3 fb^{-1}) [165]	45
CMS7jets'14	**542**	CMS 7 TeV incl. jet, R=0.7, ($d^2\sigma/dp_T^j dy_j$) (5 fb^{-1}) [166]	158

Experiment name	**CT ID#**	**Dataset details**	N_{pt}
ATLAS7jets'15	**544**	ATLAS 7 TeV incl. jet, R=0.6, $(d^2\sigma/dp_T^j dy_j)$ (4.5 fb^{-1}) [167]	140
CMS8jets'17	**545**	CMS 8 TeV incl. jet, R=0.7, $(d^2\sigma/dp_T^j dy_j)$ (19.7 fb^{-1}) [67]	185
ATL8ttb-pt'16	**565**	ATLAS 8 TeV $t\bar{t}\, d\sigma/dp_T^t$ (20.3 fb^{-1}) [168]	8
ATL8ttb-y_ave'16	**566**	ATLAS 8 TeV $t\bar{t}\, d\sigma/dy_{<t/\bar{t}>}$ (20.3 fb^{-1}) [168]	5
ATL8ttb-mtt'16	**567**	ATLAS 8 TeV $t\bar{t}\, d\sigma/dm_{t\bar{t}}$ (20.3 fb^{-1}) [168]	7
ATL8ttb-y_ttb'16	**568**	ATLAS 8 TeV $t\bar{t}\, d\sigma/dy_{t\bar{t}}$ (20.3 fb^{-1}) [168]	5

Table B.4: For each experiment E we have defined its flavor-specific sensitivity $|S_f^E|$ and its point-averaged counterpart $\langle|S_f^E|\rangle$ in Sec. 2.4.2. Using these quantities, we tabulate the total overall (*i.e.*, flavor-summed) sensitivity and a flavor-dependent sensitivity for the various experiments in our dataset, ordering the table in descending magnitude for the overall sensitivity. Thus, row 1 for the combined HERA Run I + Run 2 dataset has the greatest overall sensitivity, while row 47 for the H1 σ_r^b reduced cross section has the least overall sensitivity according to that metric. For each flavor, we award particularly sensitive experiments a rank $\mathbf{A}, \mathrm{B}, \mathrm{C}$ or $\mathbf{1*}, \mathbf{1}, 2, 3$ based on their total and point-averaged sensitivities, respectively. These ranks are decided using the criteria: $C \iff |S_f^E| \in [20, 50]$, $B \iff |S_f^E| \in [50, 100]$, and $\mathbf{A} \iff |S_f^E| > 100$ according to the total sensitivities for each flavor; and, analogously, $3 \iff \langle|S_f^E|\rangle \in [0.1, 0.25]$, $2 \iff \langle|S_f^E|\rangle \in [0.25, 0.5]$, $\mathbf{1} \iff \langle|S_f^E|\rangle \in [0.5, 1]$, and $\mathbf{1*} \iff \langle|S_f^E|\rangle > 1$ according to the point-averaged sensitivities. Experiments with sensitivities falling below the lowest ranks (that is, with $|S_f^E| < 20$ or $\langle|S_f^E|\rangle < 0.1$) are not awarded a rank for that category/flavor. Note that we sum over the light quark + gluon flavors to compute $\langle\sum_f |S_f^E|\rangle$ within this and subsequent tables. Also, new experimental datasets not originally included in CT14HERA2 are indicated by **bold** Expt. names in the second column.

			Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E\rvert$	$\langle\sum_f \lvert S_f^E\rvert\rangle$	$\lvert S_{\bar{d}}^E\rvert$	$\langle\lvert S_{\bar{d}}^E\rvert\rangle$	$\lvert S_{\bar{u}}^E\rvert$	$\langle\lvert S_{\bar{u}}^E\rvert\rangle$	$\lvert S_g^E\rvert$	$\langle\lvert S_g^E\rvert\rangle$	$\lvert S_u^E\rvert$	$\langle\lvert S_u^E\rvert\rangle$	$\lvert S_d^E\rvert$	$\langle\lvert S_d^E\rvert\rangle$	$\lvert S_s^E\rvert$	$\langle\lvert S_s^E\rvert\rangle$
1	HERAI+II'15	1120.	620.	0.0922	B		**A**	3	**A**	3	**A**	3	B		C	
2	CCFR-F3'97	86	218.	0.423	C	**1**	C	**1**		3	B	**1**	C	2		
3	BCDMSp'89	337	184.	0.0908			C		C		B	3	C			
4	NMCrat'97	123	169.	0.229	C	2					C	2	B	2		
5	BCDMSd'90	250	141.	0.0939	C				C	3	C	3	C	3		
6	CDHSW-F3'91	96	115.	0.199	C	2	C	2		3	C	2	C	3		
7	E605'91	119	113.	0.158	C	2	C	2				3				

			Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
8	E866pp'03	184	103.	0.0935		3	C	3			C	3				
9	CCFR-F2'01	69	89.1	0.215		3		3	C	2		3		2		3
10	**CMS8jets'17**	185	87.6	0.0789					C	3						
11	CDHSW-F2'91	85	82.4	0.162		3		3		3		3	C	3		
12	CMS7jets'13	133	63.8	0.0799					C	3						
13	NuTeV-nu'06	38	58.9	0.259		3		3				3		3	C	**1**
14	**CMS7jets'14**	158	57.5	0.0606					C	3						
15	CCFR SI nub'01	38	49.4	0.217		3		3				3		3	C	**1**
16	**ATLAS7jets'15**	140	48.2	0.0574						3						
17	CCFR SI nu'01	40	48.	0.2		3		3				3		3	C	**1**
18	**LHCb8WZ'16**	42	41.4	0.164		3		3		3		3		2		
19	ATL7WZ'12	41	39.6	0.161		3		3				3		3		3
20	**CMS8Wasy'16**	33	39.2	0.198		2		3				3		2		3
21	D02jets'08	110	37.5	0.0568						3						
22	NuTeV-nub'06	33	36.7	0.185		3		3				3		3		2

			Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
23	**ATL8DY2D’16**	48	34.7	0.121		3		3				3				3
24	E866rat’01	15	33.3	0.37		**1**		**1**				3		2		
25	ATL7jets’12	90	30.4	0.0563						3						
26	**LHCb7ZWrap’15**	33	30.2	0.152		3		3		3		3		3		
27	CMS7Masy2’14	11	29.4	0.446		**1**		2		2		2		**1**		3
28	CDF2jets’09	72	21.5	0.0497						3						
29	**ATL8ZpT’16**	45	17.2	0.0638						3						3
30	HERAc’13	47	15.1	0.0537						3						
31	D02Masy’08	9	15.	0.278		3		3				2		2		2
32	CMS7Easy’12	11	14.3	0.216		2		3		3		3		2		
33	D02Easy2’15	13	14.	0.18		3		3				3		2		
34	ZyD02’08	28	11.6	0.0693								3		3		
35	ZyCDF2’10	29	11.2	0.0647								3				
36	CDF1Wasy’96	11	8.83	0.134		3		3				3		2		
37	LHCb7WZ’12	14	7.27	0.0866		3								3		

			Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
38	**LHCb8Zee'15**	17	7.1	0.0696								3				
39	**ATL8ttb-pt'16**	8	6.2	0.129		3		3		2						
40	LHCb7Wasy'12	5	6.11	0.204		2		3				3		2		3
41	**ATL7ZpT'14**	8	5.84	0.122		3		3		3		3		3		
42	HERA-FL'11	9	3.99	0.0739						2						
43	**ATL8ttb-mtt'16**	7	3.81	0.0907						2						
44	CDF2Wasy'05	11	3.7	0.056										3		
45	**ATL8ttb-y_ttb'16**	5	3.37	0.112						2						
46	**ATL8ttb-y_ave'16**	5	3.2	0.107						2						
47	HERAb'06	10	1.14	0.0191												

Table B.5: A horizontal continuation of the information in Table B.4, containing the flavor-dependent total and mean sensitivities of a number of derived quantities, as opposed to the individual flavors given in Table B.4. Going across, the total and mean sensitivities are tabulated for valence distributions of the u and d quarks, the partonic flavor ratios $\bar{d}/\bar{u}$ and d/u, and the Higgs production cross section $\sigma_{pp\to H^0X}$ at 7, 8, and 14 TeV, respectively. The ranking criteria, ordering, and other conventions are again as described in Table B.4.

		Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
1	HERAI+II'15	B		C		C		B		B		B		B	
2	CCFR-F3'97	B	**1**	B	**1**			C	2		3		3		3
3	BCDMSp'89	B	3	C		C		C	3	C					
4	NMCrat'97	C	2	C	3	C	2	B	**1**						
5	BCDMSd'90	C		C	3			C		C		C			
6	CDHSW-F3'91	C	2	C	2				3						
7	E605'91	C	3	C	3										
8	E866pp'03	C	3												
9	CCFR-F2'01		3		3		3		3		3		3		3
10	**CMS8jets'17**										3	C	3	C	3
11	CDHSW-F2'91		3		3				3		3		3		

		Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
12	CMS7jets'13										3		3		3
13	NuTeV-nu'06														
14	**CMS7jets'14**										3		3		3
15	CCFR SI nub'01														
16	**ATLAS7jets'15**														
17	CCFR SI nu'01														
18	**LHCb8WZ'16**		3		3		2		2		3		3		
19	ATL7WZ'12		3				3		3						
20	**CMS8Wasy'16**		3		3		2		2						
21	D02jets'08														
22	NuTeV-nub'06														
23	**ATL8DY2D'16**		3				3		3						
24	E866rat'01		2		2	C	**1***		2		3		3		
25	ATL7jets'12										3		3		3
26	**LHCb7ZWrap'15**		3		3		2		2		3		3		

		Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	$\vert S^E_{u_v}\vert$	$\langle\vert S^E_{u_v}\vert\rangle$	$\vert S^E_{d_v}\vert$	$\langle\vert S^E_{d_v}\vert\rangle$	$\vert S^E_{\bar{d}/\bar{u}}\vert$	$\langle\vert S^E_{\bar{d}/\bar{u}}\vert\rangle$	$\vert S^E_{d/u}\vert$	$\langle\vert S^E_{d/u}\vert\rangle$	$\vert S^E_{H7}\vert$	$\langle\vert S^E_{H7}\vert\rangle$	$\vert S^E_{H8}\vert$	$\langle\vert S^E_{H8}\vert\rangle$	$\vert S^E_{H14}\vert$	$\langle\vert S^E_{H14}\vert\rangle$
27	CMS7Masy2'14		2		2		**1**		**1**		3		3		3
28	CDF2jets'09														
29	**ATL8ZpT'16**														3
30	HERAc'13										3		3		3
31	D02Masy'08		2		2		2		2						3
32	CMS7Easy'12		3		3		2		2						
33	D02Easy2'15		3		2		3		2						
34	ZyD02'08		3												
35	ZyCDF2'10		3												
36	CDF1Wasy'96		3		2		3		2						
37	LHCb7WZ'12						3		3						
38	**LHCb8Zee'15**														
39	**ATL8ttb-pt'16**		3								2		2		2
40	LHCb7Wasy'12		3		3		2		2		3		3		3
41	**ATL7ZpT'14**								3		3		3		3

		Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
42	HERA-FL'11										3		3		
43	**ATL8ttb-mtt'16**										3		3		3
44	CDF2Wasy'05				3				3						
45	**ATL8ttb-y_ttb'16**										2		2		3
46	**ATL8ttb-y_ave'16**										2		2		3
47	HERAb'06														

B.2. Supplementary Material As Supplementary Material, we enclose in this Appendix a series of additional tables that further illustrate the details of our sensitivity analysis. These include a detailed breakdown of the various CTEQ-TEA experiments according to physical process (Table B.6) and associated sensitivity rankings, both for individual PDF flavors (Table B.7) and for various derived quantities (Table B.8). In addition, in Tables B.9 and B.10, we give numerical values of sensitivities corresponding to the rankings shown in Tables B.4 and B.5. In Tables B.11 and B.12, numerical values of sensitivities corresponding to Tables B.7 and B.8 are also given. Lastly, in Tables B.13 and B.14, sensitivity ranking tables of the CTEQ-TEA dataset based upon a companion fit that excluded jet data are given, and corresponding numerical values are shown in Tables B.15 and B.16.

Table B.6: The experimental IDs of the datasets making up the process types considered in this analysis; we identify these various processes by abbreviated labels: charge current DIS (DISCC), neutral current DIS (DISNC), NC/CC DIS (DISNCCC), and all DIS; Vector Boson Production (VBP) of the W (VBPW), Z (VBPZ), and W/Z processes (VBPWZ); $p_T^{W/Z}$ of Z (VBPZpT); jet production (JP) and $t\bar{t}$. "Old" sets were in CT14HERA2, but the "New" only in CTEQ-TEA.

Process	Experiment Names
DIS Old	BCDMSp'89, BCDMSd'90, NMCrat'97, CDHSW-F2'91, CDHSW-F3'91, CCFR-F2'01, CCFR-F3'97, NuTeV-nu'06, NuTeV-nub'06, CCFR SI nu'01, CCFR SI nub'01, HERAb'06, HERAc'13, HERAI+II'15, 169
DISCC Old	CDHSW-F2'91, CDHSW-F3'91, CCFR-F2'01, CCFR-F3'97, NuTeV-nu'06, NuTeV-nub'06, CCFR SI nu'01, CCFR SI nub'01
JP New	**CMS7jets'14**, **ATLAS7jets'15**, **CMS8jets'17**
DISNCCC	HERAI+II'15
VBPZ Old	E605'91, E866rat'01, E866pp'03, ZyD02'08, ZyCDF2'10
DISNC Old	BCDMSp'89, BCDMSd'90, NMCrat'97, HERAb'06, HERAc'13, HERA-FL'11
JP Old	CDF2jets'09, D02jets'08, ATL7jets'12, CMS7jets'13
VBPW Old	CDF1Wasy'96, CDF2Wasy'05, D02Masy'08, LHCb7Wasy'12, CMS7Masy2'14, CMS7Easy'12, D02Easy2'15
VBPWZ New	**LHCb7ZWrap'15**, **LHCb8WZ'16**

Process	Experiment Names
VBPZ New	**LHCb8Zee'15**, **ATL8DY2D'16**
VBPWZ Old	LHCb7WZ'12, ATL7WZ'12
VBPW New	**CMS8Wasy'16**
VBPZpT	**ATL7ZpT'14**, **ATL8ZpT'16**
$t\bar{t}$	**ATL8ttb-pt'16**, **ATL8ttb-y_ave'16**, **ATL8ttb-mtt'16**, **ATL8ttb-y_ttb'16**

Table B.7: Similar to Table B.4, yet we tabulate the various types of processes rather than the separate experiments in our global CTEQ-TEA dataset. The process labels are explained in the caption of Table B.6, which summarized the constituent experiments contributing to each process type.

			Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Process	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar d}^E \rvert$	$\langle \lvert S_{\bar d}^E \rvert \rangle$	$\lvert S_{\bar u}^E \rvert$	$\langle \lvert S_{\bar u}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
1	DIS Old	2381	1.83E3	0.128	**A**	3	**A**	3	**A**	3	**A**	3	**A**	3	**A**	
2	DISCC Old	485	698.	0.24	**A**	2	**A**	3	B	3	**A**	2	**A**	2	**A**	3
3	DISNCCC	1120	620.	0.0922	B		**A**	3	**A**	3	**A**	3	B		C	
4	DISNC Old	776	514.	0.11	B	3	B		B		**A**	3	**A**	3	C	
5	VBPZ Old	375	272.	0.121	B	3	B	3	C		C	3	C			
6	JP New	483	193.	0.0667	C		C		B	3					C	
7	JP Old	405	153.	0.063			C		B	3					C	
8	VBPW Old	71	91.4	0.215	C	2		3				3	C	2		3
9	VBPWZ New	75	71.6	0.159		3		3		3		3		3		
10	VBPWZ Old	55	46.9	0.142		3		3				3		3		3
11	VBPZ New	65	41.8	0.107		3		3				3				3
12	VBPW New	33	39.2	0.198		2		3				3		2		3
13	VBPZpT	53	23.1	0.0725						3						3

No.	Process	N_{pt}	Rankings, **CT14 HERA2 NNLO PDFs**													
			$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
14	$t\bar{t}$	25	16.6	0.111				3		2						

Table B.8: Continuation of Table B.7, listing the PDF combinations and Higgs production cross sections similarly to Table B.5 of the main paper.

		Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Process	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
1	DIS Old	**A**	3	**A**	3	**A**		**A**	3	**A**		**A**		**A**	
2	DISCC Old	**A**	2	**A**	2	C		B	3	B	3	B	3	C	
3	DISNCCC	B		C		C		B		B		B		B	
4	DISNC Old	**A**	3	B		B	3	**A**	3	B		B		C	
5	VBPZ Old	B	3	B	3	B	3	C							
6	JP New									C		B	3	B	3
7	JP Old									C	3	C	3	C	3
8	VBPW Old		3		2	C	2	C	2						
9	VBPWZ New		3		3	C	2	C	2		3		3		
10	VBPWZ Old		3				3		3						
11	VBPZ New								3						
12	VBPW New		3		3		2		2						
13	VBPZpT														3

		Rankings, **CT14 HERA2 NNLO PDFs**													
No.	Process	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
14	$t\bar{t}$										2		2		3

Table B.9: Here we separately collect the numerical values of the total and point-averaged sensitivities used to determine the rankings in Table B.4 of the main paper.

			Values, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
1	HERAI+II'15	1120	620.	0.0922	78.3	0.0699	124.	0.111	143.	0.128	146.	0.13	83.	0.0741	45.	0.0402
2	CCFR-F3'97	86	218.	0.423	46.2	0.537	49.8	0.579	16.8	0.195	63.7	0.741	34.7	0.404	6.85	0.0797
3	BCDMSp'89	337	184.	0.0908	16.6	0.0492	31.1	0.0922	25.8	0.0765	81.2	0.241	20.3	0.0601	8.77	0.026
4	NMCrat'97	123	169.	0.229	44.4	0.361	9.87	0.0802	9.7	0.0789	36.4	0.296	61.4	0.499	7.44	0.0605
5	BCDMSd'90	250	141.	0.0939	24.3	0.0971	17.7	0.0707	26.5	0.106	26.9	0.108	36.2	0.145	9.19	0.0368
6	CDHSW-F3'91	96	115.	0.199	24.2	0.253	25.3	0.264	9.81	0.102	31.6	0.329	20.2	0.21	3.76	0.0392
7	E605'91	119	113.	0.158	42.	0.353	36.9	0.31	5.41	0.0455	13.3	0.112	7.81	0.0656	7.1	0.0596
8	E866pp'03	184	103.	0.0935	19.3	0.105	31.3	0.17	12.8	0.0693	23.2	0.126	9.63	0.0523	6.98	0.0379
9	CCFR-F2'01	69	89.1	0.215	15.2	0.22	9.32	0.135	21.1	0.306	14.6	0.211	19.6	0.284	9.3	0.135
10	**CMS8jets'17**	185	87.6	0.0789	10.4	0.0564	12.5	0.0676	36.7	0.198	9.08	0.0491	6.9	0.0373	11.9	0.0645
11	CDHSW-F2'91	85	82.4	0.162	11.7	0.138	9.14	0.107	19.8	0.233	15.7	0.185	20.5	0.241	5.55	0.0653
12	CMS7jets'13	133	63.8	0.0799	8.21	0.0617	8.87	0.0667	25.5	0.192	6.75	0.0508	6.23	0.0468	8.18	0.0615
13	NuTeV-nu'06	38	58.9	0.259	7.77	0.204	8.53	0.224	2.57	0.0676	6.41	0.169	8.45	0.222	25.2	0.664

			Values, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E\rvert$	$\langle\sum_f \lvert S_f^E\rvert\rangle$	$\lvert S_{\bar d}^E\rvert$	$\langle\lvert S_{\bar d}^E\rvert\rangle$	$\lvert S_{\bar u}^E\rvert$	$\langle\lvert S_{\bar u}^E\rvert\rangle$	$\lvert S_g^E\rvert$	$\langle\lvert S_g^E\rvert\rangle$	$\lvert S_u^E\rvert$	$\langle\lvert S_u^E\rvert\rangle$	$\lvert S_d^E\rvert$	$\langle\lvert S_d^E\rvert\rangle$	$\lvert S_s^E\rvert$	$\langle\lvert S_s^E\rvert\rangle$
14	**CMS7jets'14**	158	57.5	0.0606	7.45	0.0472	8.38	0.0531	23.7	0.15	4.76	0.0301	5.	0.0316	8.11	0.0513
15	CCFR SI nub'01	38	49.4	0.217	6.06	0.16	5.82	0.153	2.48	0.0652	6.39	0.168	7.37	0.194	21.3	0.559
16	**ATLAS7jets'15**	140	48.2	0.0574	6.62	0.0473	6.98	0.0499	19.8	0.141	3.94	0.0281	3.55	0.0253	7.34	0.0524
17	CCFR SI nu'01	40	48.	0.2	6.52	0.163	6.99	0.175	1.87	0.0468	5.33	0.133	6.96	0.174	20.4	0.509
18	**LHCb8WZ'16**	42	41.4	0.164	10.2	0.244	5.75	0.137	4.58	0.109	7.27	0.173	11.3	0.269	2.32	0.0552
19	ATL7WZ'12	41	39.6	0.161	8.09	0.197	5.14	0.125	3.1	0.0757	5.91	0.144	8.25	0.201	9.13	0.223
20	**CMS8Wasy'16**	33	39.2	0.198	9.04	0.274	5.08	0.154	3.1	0.0938	5.05	0.153	9.81	0.297	7.09	0.215
21	D02jets'08	110	37.5	0.0568	5.36	0.0487	5.68	0.0517	15.8	0.144	2.91	0.0265	2.33	0.0212	5.4	0.0491
22	NuTeV-nub'06	33	36.7	0.185	3.9	0.118	3.84	0.116	2.62	0.0795	4.75	0.144	5.33	0.161	16.3	0.493
23	**ATL8DY2D'16**	48	34.7	0.121	5.78	0.12	6.85	0.143	3.23	0.0673	5.8	0.121	4.7	0.098	8.36	0.174
24	E866rat'01	15	33.3	0.37	11.8	0.789	12.1	0.804	0.697	0.0464	3.59	0.239	3.98	0.265	1.16	0.0773
25	ATL7jets'12	90	30.4	0.0563	3.33	0.037	4.24	0.0471	13.	0.144	2.6	0.0289	2.35	0.0261	4.86	0.054
26	**LHCb7ZWrap'15**	33	30.2	0.152	6.15	0.186	4.71	0.143	3.71	0.112	6.24	0.189	7.02	0.213	2.34	0.0709
27	CMS7Masy2'14	11	29.4	0.446	8.43	0.767	3.24	0.294	2.91	0.265	3.35	0.305	8.86	0.805	2.62	0.238
28	CDF2jets'09	72	21.5	0.0497	2.92	0.0406	2.9	0.0403	9.06	0.126	2.77	0.0384	1.31	0.0182	2.5	0.0347

			Values, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar d}^E \rvert$	$\langle \lvert S_{\bar d}^E \rvert \rangle$	$\lvert S_{\bar u}^E \rvert$	$\langle \lvert S_{\bar u}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
29	**ATL8ZpT'16**	45	17.2	0.0638	1.32	0.0293	2.18	0.0485	5.38	0.12	1.66	0.0368	1.29	0.0286	5.4	0.12
30	HERAc'13	47	15.1	0.0537	1.85	0.0393	1.69	0.036	6.53	0.139	1.97	0.0419	1.91	0.0406	1.2	0.0255
31	D02Masy'08	9	15.	0.278	1.85	0.205	2.08	0.231	0.638	0.0709	4.05	0.449	3.86	0.428	2.55	0.284
32	CMS7Easy'12	11	14.3	0.216	4.4	0.4	1.66	0.151	1.25	0.114	1.55	0.141	4.32	0.393	1.1	0.1
33	D02Easy2'15	13	14.	0.18	2.69	0.207	2.69	0.207	0.896	0.0689	2.8	0.215	4.1	0.315	0.865	0.0666
34	ZyD02'08	28	11.6	0.0693	1.69	0.0602	0.881	0.0315	1.49	0.0532	3.49	0.125	2.81	0.1	1.28	0.0457
35	ZyCDF2'10	29	11.2	0.0647	1.75	0.0603	1.07	0.0369	1.36	0.0467	3.15	0.109	2.81	0.0968	1.12	0.0386
36	CDF1Wasy'96	11	8.83	0.134	1.49	0.135	1.2	0.109	0.712	0.0647	1.26	0.115	3.34	0.304	0.828	0.0753
37	LHCb7WZ'12	14	7.27	0.0866	1.49	0.107	1.23	0.0882	0.973	0.0695	1.37	0.0979	1.67	0.119	0.535	0.0382
38	**LHCb8Zee'15**	17	7.1	0.0696	0.885	0.0521	1.48	0.0869	1.27	0.0747	1.75	0.103	1.23	0.0724	0.483	0.0284
39	**ATL8ttb-pt'16**	8	6.2	0.129	1.11	0.139	1.41	0.176	2.14	0.268	0.579	0.0724	0.514	0.0643	0.447	0.0559
40	LHCb7Wasy'12	5	6.11	0.204	1.67	0.334	0.856	0.171	0.466	0.0933	0.97	0.194	1.64	0.327	0.51	0.102
41	**ATL7ZpT'14**	8	5.84	0.122	1.31	0.163	0.939	0.117	1.18	0.148	0.891	0.111	1.24	0.155	0.286	0.0357
42	HERA-FL'11	9	3.99	0.0739	0.238	0.0265	0.364	0.0405	2.3	0.256	0.37	0.0411	0.238	0.0265	0.477	0.053
43	**ATL8ttb-mtt'16**	7	3.81	0.0907	0.177	0.0253	0.338	0.0482	2.26	0.322	0.408	0.0583	0.46	0.0657	0.167	0.0239

No.	Expt.	N_{pt}	Values, **CT14 HERA2 NNLO PDFs**													
			$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
44	CDF2Wasy'05	11	3.7	0.056	0.57	0.0518	0.635	0.0578	0.151	0.0138	0.593	0.0539	1.33	0.121	0.422	0.0384
45	**ATL8ttb-y_ttb'16**	5	3.37	0.112	0.407	0.0814	0.461	0.0921	1.56	0.311	0.191	0.0382	0.45	0.09	0.308	0.0616
46	**ATL8ttb-y_ave'16**	5	3.2	0.107	0.273	0.0546	0.339	0.0678	1.57	0.314	0.194	0.0388	0.368	0.0736	0.455	0.0911
47	HERAb'06	10	1.14	0.0191	0.16	0.016	0.107	0.0107	0.463	0.0463	0.121	0.0121	0.165	0.0165	0.128	0.0128

Table B.10: Sensitivity values for Table B.5 of the main paper.

No.	Expt.	Values, **CT14 HERA2 NNLO PDFs**													
		$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
1	HERAI+II'15	64.6	0.0577	36.7	0.0328	39.7	0.0354	63.7	0.0569	76.2	0.0681	74.3	0.0663	78.6	0.0701
2	CCFR-F3'97	77.5	0.901	65.4	0.761	5.87	0.0682	25.5	0.297	11.3	0.131	11.8	0.137	13.3	0.154
3	BCDMSp'89	70.8	0.21	21.3	0.0632	20.2	0.0598	40.4	0.12	21.4	0.0634	19.3	0.0573	11.7	0.0349
4	NMCrat'97	34.6	0.281	25.6	0.208	48.7	0.396	76.6	0.623	9.47	0.077	9.94	0.0808	10.5	0.0852
5	BCDMSd'90	20.3	0.0813	27.	0.108	12.6	0.0503	22.5	0.09	23.9	0.0958	23.1	0.0922	18.1	0.0724
6	CDHSW-F3'91	36.7	0.382	32.2	0.336	2.56	0.0266	12.3	0.128	5.45	0.0568	5.48	0.0571	5.88	0.0613
7	E605'91	21.6	0.181	23.7	0.199	11.1	0.0932	3.7	0.0311	1.67	0.014	1.82	0.0153	2.19	0.0184
8	E866pp'03	29.6	0.161	16.2	0.0882	17.7	0.0959	12.4	0.0675	9.6	0.0522	11.	0.0596	13.8	0.0749
9	CCFR-F2'01	10.6	0.154	12.	0.173	9.64	0.14	12.6	0.183	12.2	0.176	11.3	0.164	8.58	0.124
10	**CMS8jets'17**	11.2	0.0604	6.68	0.0361	6.48	0.035	6.65	0.036	19.3	0.105	20.8	0.112	26.6	0.144
11	CDHSW-F2'91	10.5	0.123	12.6	0.148	4.26	0.0501	12.2	0.143	16.7	0.196	15.	0.177	8.48	0.0998
12	CMS7jets'13	6.	0.0451	4.92	0.037	4.14	0.0311	3.66	0.0275	18.	0.135	18.3	0.137	18.9	0.142
13	NuTeV-nu'06	0.959	0.0252	1.12	0.0294	2.23	0.0587	3.68	0.0968	2.32	0.061	1.99	0.0524	1.8	0.0473
14	**CMS7jets'14**	4.56	0.0289	4.02	0.0254	4.64	0.0293	3.33	0.0211	18.5	0.117	18.3	0.116	17.2	0.109

		Values, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	$\|S^E_{u_v}\|$	$\langle\|S^E_{u_v}\|\rangle$	$\|S^E_{d_v}\|$	$\langle\|S^E_{d_v}\|\rangle$	$\|S^E_{\bar{d}/\bar{u}}\|$	$\langle\|S^E_{\bar{d}/\bar{u}}\|\rangle$	$\|S^E_{d/u}\|$	$\langle\|S^E_{d/u}\|\rangle$	$\|S^E_{H7}\|$	$\langle\|S^E_{H7}\|\rangle$	$\|S^E_{H8}\|$	$\langle\|S^E_{H8}\|\rangle$	$\|S^E_{H14}\|$	$\langle\|S^E_{H14}\|\rangle$
15	CCFR SI nub'01	2.39	0.0629	1.92	0.0506	1.25	0.033	3.19	0.0839	3.3	0.0868	3.14	0.0827	2.69	0.0707
16	**ATLAS7jets'15**	4.17	0.0298	3.69	0.0263	3.55	0.0254	2.48	0.0177	10.1	0.072	11.	0.0787	13.4	0.0958
17	CCFR SI nu'01	0.594	0.0148	0.829	0.0207	1.3	0.0325	2.52	0.0629	1.37	0.0343	1.19	0.0297	1.34	0.0335
18	**LHCb8WZ'16**	6.35	0.151	5.98	0.142	12.7	0.303	13.2	0.313	4.53	0.108	4.37	0.104	3.62	0.0861
19	ATL7WZ'12	6.52	0.159	3.95	0.0963	9.07	0.221	8.54	0.208	1.83	0.0446	1.47	0.0358	1.29	0.0314
20	**CMS8Wasy'16**	6.84	0.207	4.5	0.136	10.4	0.315	9.68	0.293	1.58	0.0478	1.18	0.0357	1.73	0.0523
21	D02jets'08	2.51	0.0228	1.74	0.0158	2.91	0.0265	1.88	0.0171	8.03	0.073	8.39	0.0763	8.96	0.0814
22	NuTeV-nub'06	2.39	0.0723	2.1	0.0637	1.41	0.0427	2.93	0.0888	3.05	0.0925	2.85	0.0865	2.68	0.0811
23	**ATL8DY2D'16**	5.14	0.107	2.18	0.0455	5.56	0.116	6.07	0.126	4.13	0.086	3.93	0.0819	3.35	0.0699
24	E866rat'01	7.03	0.469	6.41	0.427	25.	1.67	4.78	0.319	2.01	0.134	1.83	0.122	1.17	0.0782
25	ATL7jets'12	2.65	0.0295	1.59	0.0177	1.77	0.0196	1.71	0.019	9.5	0.106	9.56	0.106	9.63	0.107
26	**LHCb7ZWrap'15**	4.95	0.15	4.73	0.143	8.36	0.253	8.55	0.259	4.34	0.131	4.11	0.125	2.96	0.0898
27	CMS7Masy2'14	5.21	0.474	2.88	0.262	9.87	0.897	10.8	0.981	1.59	0.144	1.63	0.148	2.15	0.196
28	CDF2jets'09	2.52	0.035	1.19	0.0165	1.95	0.027	1.43	0.0198	5.11	0.0709	4.72	0.0655	3.87	0.0538
29	**ATL8ZpT'16**	1.53	0.0339	0.86	0.0191	2.29	0.0508	1.63	0.0363	2.1	0.0468	2.09	0.0464	5.18	0.115

		Values, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	$\vert S^E_{u_v}\vert$	$\langle\vert S^E_{u_v}\vert\rangle$	$\vert S^E_{d_v}\vert$	$\langle\vert S^E_{d_v}\vert\rangle$	$\vert S^E_{\bar{d}/\bar{u}}\vert$	$\langle\vert S^E_{\bar{d}/\bar{u}}\vert\rangle$	$\vert S^E_{d/u}\vert$	$\langle\vert S^E_{d/u}\vert\rangle$	$\vert S^E_{H7}\vert$	$\langle\vert S^E_{H7}\vert\rangle$	$\vert S^E_{H8}\vert$	$\langle\vert S^E_{H8}\vert\rangle$	$\vert S^E_{H14}\vert$	$\langle\vert S^E_{H14}\vert\rangle$
30	HERAc'13	0.985	0.021	0.619	0.0132	0.752	0.016	0.937	0.0199	5.52	0.117	5.42	0.115	4.81	0.102
31	D02Masy'08	3.85	0.428	4.14	0.46	2.7	0.3	3.78	0.42	0.512	0.0569	0.633	0.0704	0.986	0.11
32	CMS7Easy'12	2.24	0.203	1.21	0.11	5.18	0.471	5.23	0.475	0.958	0.0871	1.01	0.0918	1.05	0.0954
33	D02Easy2'15	3.02	0.232	4.47	0.343	2.51	0.193	4.77	0.367	0.378	0.0291	0.438	0.0337	0.545	0.0419
34	ZyD02'08	3.45	0.123	2.06	0.0735	1.6	0.057	1.74	0.0623	1.22	0.0436	0.971	0.0347	0.518	0.0185
35	ZyCDF2'10	3.08	0.106	2.42	0.0836	1.71	0.0591	2.26	0.0779	1.12	0.0386	0.887	0.0306	0.739	0.0255
36	CDF1Wasy'96	1.2	0.109	3.5	0.318	1.64	0.149	3.78	0.343	0.357	0.0325	0.442	0.0402	0.66	0.06
37	LHCb7WZ'12	0.871	0.0622	1.01	0.0725	1.54	0.11	1.7	0.121	1.07	0.0765	0.961	0.0686	0.572	0.0409
38	**LHCb8Zee'15**	0.537	0.0316	0.64	0.0377	0.704	0.0414	1.05	0.0617	0.714	0.042	0.579	0.0341	0.37	0.0218
39	**ATL8ttb-pt'16**	0.893	0.112	0.589	0.0737	0.234	0.0292	0.536	0.067	2.4	0.3	2.56	0.32	2.68	0.335
40	LHCb7Wasy'12	0.942	0.188	0.886	0.177	1.97	0.393	1.98	0.395	0.675	0.135	0.677	0.135	0.554	0.111
41	**ATL7ZpT'14**	0.0843	0.0105	0.189	0.0237	0.718	0.0898	0.809	0.101	1.63	0.203	1.56	0.195	1.35	0.169
42	HERA-FL'11	0.0644	0.00716	0.0624	0.00693	0.0874	0.00971	0.0828	0.0092	1.15	0.128	1.03	0.114	0.422	0.0469
43	**ATL8ttb-mtt'16**	0.402	0.0575	0.42	0.06	0.179	0.0256	0.259	0.037	1.08	0.154	1.05	0.15	1.22	0.174
44	CDF2Wasy'05	0.529	0.0481	1.35	0.123	0.67	0.0609	1.57	0.143	0.0668	0.00608	0.0794	0.00722	0.11	0.00998

		Values, **CT14 HERA2 NNLO PDFs**													
No.	Expt.	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
45	**ATL8ttb-y_ttb'16**	0.13	0.0259	0.0977	0.0195	0.166	0.0333	0.293	0.0585	1.49	0.298	1.46	0.291	1.11	0.222
46	**ATL8ttb-y_ave'16**	0.121	0.0242	0.193	0.0386	0.177	0.0354	0.222	0.0444	1.31	0.262	1.34	0.268	1.21	0.243
47	HERAb'06	0.0873	0.00873	0.0543	0.00543	0.0927	0.00927	0.105	0.0105	0.491	0.0491	0.465	0.0465	0.349	0.0349

Table B.11: Sensitivity values for Table B.7.

			Values, **CT14 HERA2 NNLO PDFs**													
No.	Process	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
1	DIS Old	2381	1.83E3	0.128	287.	0.121	304.	0.128	292.	0.122	441.	0.185	326.	0.137	181.	0.0759
2	DISCC Old	485	698.	0.24	122.	0.251	119.	0.245	77.	0.159	148.	0.306	123.	0.254	109.	0.224
3	DISNCCC	1120	620.	0.0922	78.3	0.0699	124.	0.111	143.	0.128	146.	0.13	83.	0.0741	45.	0.0402
4	DISNC Old	776	514.	0.11	87.5	0.113	60.8	0.0783	71.3	0.0919	147.	0.189	120.	0.155	27.2	0.0351
5	VBPZ Old	375	272.	0.121	76.6	0.204	82.2	0.219	21.7	0.0579	46.7	0.125	27.	0.0721	17.6	0.047
6	JP New	483	193.	0.0667	24.5	0.0508	27.9	0.0577	80.3	0.166	17.8	0.0368	15.4	0.032	27.4	0.0567
7	JP Old	405	153.	0.063	19.8	0.0489	21.7	0.0536	63.4	0.157	15.	0.0371	12.2	0.0302	20.9	0.0517
8	VBPW Old	71	91.4	0.215	21.1	0.297	12.4	0.174	7.03	0.099	14.6	0.205	27.4	0.386	8.89	0.125
9	VBPWZ New	75	71.6	0.159	16.4	0.218	10.5	0.14	8.29	0.111	13.5	0.18	18.3	0.244	4.66	0.0621
10	VBPWZ Old	55	46.9	0.142	9.59	0.174	6.38	0.116	4.08	0.0741	7.29	0.132	9.91	0.18	9.67	0.176
11	VBPZ New	65	41.8	0.107	6.67	0.103	8.33	0.128	4.5	0.0692	7.55	0.116	5.94	0.0913	8.84	0.136
12	VBPW New	33	39.2	0.198	9.04	0.274	5.08	0.154	3.1	0.0938	5.05	0.153	9.81	0.297	7.09	0.215
13	VBPZpT	53	23.1	0.0725	2.63	0.0496	3.12	0.0589	6.56	0.124	2.55	0.0481	2.52	0.0476	5.68	0.107
14	$t\bar{t}$	25	16.6	0.111	1.97	0.0787	2.55	0.102	7.52	0.301	1.37	0.0549	1.79	0.0717	1.38	0.0551

Table B.12: Sensitivity values for Table B.8.

No.	Process	Values, **CT14 HERA2 NNLO PDFs**													
		$\|S^E_{u_v}\|$	$\langle\|S^E_{u_v}\|\rangle$	$\|S^E_{d_v}\|$	$\langle\|S^E_{d_v}\|\rangle$	$\|S^E_{\bar{d}/\bar{u}}\|$	$\langle\|S^E_{\bar{d}/\bar{u}}\|\rangle$	$\|S^E_{d/u}\|$	$\langle\|S^E_{d/u}\|\rangle$	$\|S^E_{H7}\|$	$\langle\|S^E_{H7}\|\rangle$	$\|S^E_{H8}\|$	$\langle\|S^E_{H8}\|\rangle$	$\|S^E_{H14}\|$	$\langle\|S^E_{H14}\|\rangle$
1	DIS Old	333.	0.14	240.	0.101	151.	0.0632	279.	0.117	194.	0.0814	186.	0.0782	169.	0.0711
2	DISCC Old	142.	0.292	128.	0.264	28.5	0.0588	74.9	0.154	55.6	0.115	52.7	0.109	44.7	0.0922
3	DISNCCC	64.6	0.0577	36.7	0.0328	39.7	0.0354	63.7	0.0569	76.2	0.0681	74.3	0.0663	78.6	0.0701
4	DISNC Old	127.	0.163	74.6	0.0961	82.4	0.106	141.	0.181	61.9	0.0798	59.2	0.0763	45.9	0.0592
5	VBPZ Old	64.7	0.173	50.8	0.135	57.1	0.152	24.9	0.0664	15.6	0.0417	16.5	0.0439	18.4	0.0491
6	JP New	19.9	0.0412	14.4	0.0298	14.7	0.0304	12.5	0.0258	47.9	0.0991	50.1	0.104	57.2	0.118
7	JP Old	13.7	0.0338	9.45	0.0233	10.8	0.0266	8.68	0.0214	40.6	0.1	40.9	0.101	41.3	0.102
8	VBPW Old	17.	0.239	18.4	0.26	24.5	0.346	31.9	0.449	4.53	0.0639	4.91	0.0691	6.06	0.0853
9	VBPWZ New	11.3	0.151	10.7	0.143	21.1	0.281	21.7	0.289	8.86	0.118	8.49	0.113	6.58	0.0877
10	VBPWZ Old	7.39	0.134	4.96	0.0902	10.6	0.193	10.2	0.186	2.9	0.0527	2.43	0.0441	1.86	0.0338
11	VBPZ New	5.67	0.0873	2.82	0.0434	6.27	0.0964	7.12	0.109	4.84	0.0745	4.51	0.0694	3.72	0.0573
12	VBPW New	6.84	0.207	4.5	0.136	10.4	0.315	9.68	0.293	1.58	0.0478	1.18	0.0357	1.73	0.0523
13	VBPZpT	1.61	0.0304	1.05	0.0198	3.	0.0567	2.44	0.0461	3.73	0.0704	3.65	0.0689	6.53	0.123
14	$t\bar{t}$	1.55	0.0618	1.3	0.052	0.756	0.0302	1.31	0.0524	6.28	0.251	6.4	0.256	6.22	0.249

Table B.13: Experiment rankings as in Table B.4 of the main paper, for the PDFs obtained using the CT14HERA2 NNLO data sets (Tables II, III of the main paper), with the exclusion of Tevatron and LHC jet production experiments. The **bold** font indicates those experiments which were *not* fitted as constraints leading to the PDF parametrization *CT14 HERA2 NNLO PDFs fitted without jet data* used to evaluate the sensitivities in this table.

			Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
1	HERAI+II'15	1120	599.	0.0892	B		**A**	3	**A**	3	**A**	3	B		B	
2	BCDMSp'89	337	230.	0.114	C		C	3	C	3	B	3	C		C	
3	CCFR-F3'97	86	223.	0.433	B	**1**	B	**1**		3	B	**1**	C	2		3
4	**CMS8jets'17**	185	205.	0.185	C	3	C	3	B	2					C	3
5	NMCrat'97	123	168.	0.228	C	2		3			C	2	B	2		
6	BCDMSd'90	250	160.	0.107	C	3	C		C	3	C	3	C	3		
7	**CMS7jets'13**	133	131.	0.165	C	3	C	3	B	2						3
8	**CMS7jets'14**	158	126.	0.133	C	3	C	3	C	2						3
9	CDHSW-F3'91	96	117.	0.203	C	2	C	2		3	C	2		3		
10	E605'91	119	112.	0.157	C	2	C	2				3				
11	**ATLAS7jets'15**	140	109.	0.129		3	C	3	C	2						3
12	E866pp'03	184	107.	0.0966		3	C	3			C	3				

			Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E\rvert$	$\langle\sum_f \lvert S_f^E\rvert\rangle$	$\lvert S_{\bar{d}}^E\rvert$	$\langle\lvert S_{\bar{d}}^E\rvert\rangle$	$\lvert S_{\bar{u}}^E\rvert$	$\langle\lvert S_{\bar{u}}^E\rvert\rangle$	$\lvert S_g^E\rvert$	$\langle\lvert S_g^E\rvert\rangle$	$\lvert S_u^E\rvert$	$\langle\lvert S_u^E\rvert\rangle$	$\lvert S_d^E\rvert$	$\langle\lvert S_d^E\rvert\rangle$	$\lvert S_s^E\rvert$	$\langle\lvert S_s^E\rvert\rangle$
13	CCFR-F2'01	69	103.	0.248		2		3	C	2		3		2		3
14	CDHSW-F2'91	85	90.6	0.178		3		3		3		3	C	3		3
15	**D02jets'08**	110	90.4	0.137		3		3	C	2						3
16	NuTeV-nu'06	38	67.2	0.295		3		2		3		3		2	C	**1**
17	**CDF2jets'09**	72	61.9	0.143		3		3	C	2						3
18	**ATL7jets'12**	90	61.2	0.113		3		3	C	2						3
19	CCFR SI nu'01	40	54.7	0.228		3		3				3		3	C	**1**
20	CCFR SI nub'01	38	54.2	0.238		3		3				3		3	C	**1**
21	**ATL8DY2D'16**	48	40.9	0.142		3		3		3		3		3		3
22	NuTeV-nub'06	33	40.3	0.204		3		3		3		3		3		**1**
23	ATL7WZ'12	41	39.3	0.16		3		3				3		3		3
24	**LHCb8WZ'16**	42	38.8	0.154		3		3				3		2		
25	**CMS8Wasy'16**	33	36.9	0.187		2		3				3		2		3
26	E866rat'01	15	34.9	0.388		**1**		**1**		3		3		2		3
27	CMS7Masy2'14	11	30.1	0.457		**1**		2		3		2		**1**		3

			Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
28	**LHCb7ZWrap'15**	33	27.9	0.141		3		3		3		3		3		
29	**ATL8ZpT'16**	45	23.4	0.0865						3						3
30	HERAc'13	47	17.3	0.0614						3						
31	D02Masy'08	9	14.5	0.269		2		3		2		2		2		3
32	CMS7Easy'12	11	13.9	0.211		2		3				3		2		
33	ZyD02'08	28	13.3	0.0792								3		3		
34	ZyCDF2'10	29	12.6	0.0722										3		
35	D02Easy2'15	13	12.3	0.158		3		3				3		2		
36	CDF1Wasy'96	11	7.39	0.112		3						3		3		
37	LHCb7WZ'12	14	7.27	0.0865		3								3		
38	**ATL8ttb-pt'16**	8	6.91	0.144		3		3		2						
39	**LHCb8Zee'15**	17	6.31	0.0619												
40	**ATL8ttb-mtt'16**	7	5.94	0.141				3		2						
41	**ATL7ZpT'14**	8	5.73	0.119		3		3		3		3		3		
42	LHCb7Wasy'12	5	5.4	0.18		2		3				3		2		

			Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E\rvert$	$\langle\sum_f \lvert S_f^E\rvert\rangle$	$\lvert S_{\bar{d}}^E\rvert$	$\langle\lvert S_{\bar{d}}^E\rvert\rangle$	$\lvert S_{\bar{u}}^E\rvert$	$\langle\lvert S_{\bar{u}}^E\rvert\rangle$	$\lvert S_g^E\rvert$	$\langle\lvert S_g^E\rvert\rangle$	$\lvert S_u^E\rvert$	$\langle\lvert S_u^E\rvert\rangle$	$\lvert S_d^E\rvert$	$\langle\lvert S_d^E\rvert\rangle$	$\lvert S_s^E\rvert$	$\langle\lvert S_s^E\rvert\rangle$
43	**ATL8ttb-y_ave'16**	5	4.95	0.165		3		3		2						3
44	**ATL8ttb-y_ttb'16**	5	4.35	0.145		3		3		2						
45	CDF2Wasy'05	11	3.57	0.0541												
46	HERA-FL'11	9	1.82	0.0337						3						
47	HERAb'06	10	1.47	0.0244												

Table B.14: Continuation of Table B.13 for the PDFs obtained without imposing constraints from jet production experiments. The PDF combinations and Higgs production cross sections are the same as in Table B.5 of the main paper.

		Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
1	HERAI+II'15	B		C		C		B		B		B		**A**	
2	BCDMSp'89	B	3	C		C		C	3	C		C		C	
3	CCFR-F3'97	B	**1**	B	**1**		3	C	2		3		3		3
4	**CMS8jets'17**						3			C	3	C	3	C	3
5	NMCrat'97	C	2	C	3	C	2	B	**1**						
6	BCDMSd'90			C	3			C		C	3	C	3	C	
7	**CMS7jets'13**									C	2	C	2	C	2
8	**CMS7jets'14**									C	2	C	2	C	3
9	CDHSW-F3'91	C	2	C	2				3						
10	E605'91		3	C	3		3								
11	**ATLAS7jets'15**									C	3	C	3	C	3
12	E866pp'03	C	3											C	3
13	CCFR-F2'01		3		3		3		3		2		3		3

No.	Expt.	Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
		$\vert S^E_{u_v}\vert$	$\langle\vert S^E_{u_v}\vert\rangle$	$\vert S^E_{d_v}\vert$	$\langle\vert S^E_{d_v}\vert\rangle$	$\vert S^E_{\bar{d}/\bar{u}}\vert$	$\langle\vert S^E_{\bar{d}/\bar{u}}\vert\rangle$	$\vert S^E_{d/u}\vert$	$\langle\vert S^E_{d/u}\vert\rangle$	$\vert S^E_{H7}\vert$	$\langle\vert S^E_{H7}\vert\rangle$	$\vert S^E_{H8}\vert$	$\langle\vert S^E_{H8}\vert\rangle$	$\vert S^E_{H14}\vert$	$\langle\vert S^E_{H14}\vert\rangle$
14	CDHSW-F2'91		3		3				3		3		3		
15	**D02jets'08**										3		3		3
16	NuTeV-nu'06						3		3		3		3		3
17	**CDF2jets'09**										3		3		3
18	**ATL7jets'12**										3		3		3
19	CCFR SI nu'01						3				3		3		
20	CCFR SI nub'01										3		3		3
21	**ATL8DY2D'16**		3				3		3						
22	NuTeV-nub'06														
23	ATL7WZ'12		3		3		3		3		3		3		3
24	**LHCb8WZ'16**		3		3		2		2						
25	**CMS8Wasy'16**		3		3		2		2						
26	E866rat'01		2		2	C	**1***		2		3		3		3
27	CMS7Masy2'14		2		2		**1**		**1***						
28	**LHCb7ZWrap'15**		3		3		3		2		3				

No.	Expt.	Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
		$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
29	**ATL8ZpT'16**										3		3		3
30	HERAc'13										3		3		3
31	D02Masy'08		2		2		2		2		2		2		2
32	CMS7Easy'12		3		3		2		2						
33	ZyD02'08														
34	ZyCDF2'10														
35	D02Easy2'15		3		2				2						
36	CDF1Wasy'96		3		3		3		2						
37	LHCb7WZ'12								3						
38	**ATL8ttb-pt'16**		3		3						**1**		**1**		**1**
39	**LHCb8Zee'15**														
40	**ATL8ttb-mtt'16**				3						2		2		2
41	**ATL7ZpT'14**										3		3		3
42	LHCb7Wasy'12		3		3		2		2		3		3		3
43	**ATL8ttb-y_ave'16**				3						2		2		2

No.	Expt.	Rankings, **CT14 HERA2 NNLO PDFs fitted without jet data**													
		$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
44	**ATL8ttb-y_ttb'16**										2		2		2
45	CDF2Wasy'05								3						
46	HERA-FL'11														
47	HERAb'06														

Table B.15: Sensitivity values for Table B.13.

			Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
1	HERAI+II'15	1120	599.	0.0892	73.8	0.0659	122.	0.109	143.	0.128	136.	0.121	73.8	0.0659	51.1	0.0457
2	BCDMSp'89	337	230.	0.114	25.	0.0742	38.8	0.115	45.6	0.135	76.9	0.228	22.3	0.0662	21.	0.0622
3	CCFR-F3'97	86	223.	0.433	50.8	0.591	51.3	0.597	19.5	0.227	61.3	0.713	30.2	0.352	10.1	0.117
4	**CMS8jets'17**	185	205.	0.185	30.2	0.163	39.1	0.211	72.3	0.391	15.2	0.0824	14.5	0.0784	33.6	0.181
5	NMCrat'97	123	168.	0.228	45.7	0.371	12.9	0.105	6.69	0.0544	37.8	0.308	58.6	0.476	6.77	0.0551
6	BCDMSd'90	250	160.	0.107	28.7	0.115	23.5	0.0942	33.1	0.132	25.2	0.101	35.1	0.14	14.7	0.059
7	**CMS7jets'13**	133	131.	0.165	20.1	0.151	24.1	0.181	51.1	0.384	8.87	0.0667	7.44	0.0559	19.8	0.149
8	**CMS7jets'14**	158	126.	0.133	20.	0.127	23.6	0.149	46.	0.291	9.59	0.0607	8.09	0.0512	19.2	0.121
9	CDHSW-F3'91	96	117.	0.203	26.	0.271	25.5	0.266	10.9	0.114	30.	0.312	18.5	0.192	5.94	0.0619
10	E605'91	119	112.	0.157	42.5	0.357	30.1	0.253	9.58	0.0805	12.6	0.106	9.99	0.084	7.22	0.0607
11	**ATLAS7jets'15**	140	109.	0.129	17.5	0.125	20.1	0.144	37.8	0.27	8.94	0.0638	6.93	0.0495	17.5	0.125
12	E866pp'03	184	107.	0.0966	19.	0.103	30.	0.163	16.8	0.0911	21.3	0.116	12.	0.065	7.72	0.042
13	CCFR-F2'01	69	103.	0.248	17.8	0.258	14.2	0.205	24.5	0.355	15.	0.218	19.2	0.279	11.9	0.172
14	CDHSW-F2'91	85	90.6	0.178	13.3	0.157	11.6	0.137	18.7	0.22	16.1	0.19	20.6	0.243	10.2	0.12

			Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
15	**D02jets'08**	110	90.4	0.137	14.8	0.135	17.7	0.161	31.2	0.284	6.96	0.0633	5.15	0.0468	14.5	0.132
16	NuTeV-nu'06	38	67.2	0.295	8.5	0.224	10.1	0.266	3.81	0.1	7.18	0.189	10.4	0.274	27.2	0.716
17	**CDF2jets'09**	72	61.9	0.143	9.85	0.137	12.1	0.168	21.	0.292	4.58	0.0636	3.44	0.0478	10.9	0.151
18	**ATL7jets'12**	90	61.2	0.113	9.07	0.101	11.4	0.127	23.9	0.266	3.92	0.0435	2.88	0.032	9.97	0.111
19	CCFR SI nu'01	40	54.7	0.228	6.92	0.173	8.46	0.212	2.44	0.061	6.19	0.155	8.11	0.203	22.5	0.564
20	CCFR SI nub'01	38	54.2	0.238	6.03	0.159	6.58	0.173	3.34	0.0878	6.79	0.179	6.55	0.172	24.9	0.656
21	**ATL8DY2D'16**	48	40.9	0.142	6.66	0.139	7.92	0.165	5.96	0.124	6.	0.125	5.4	0.113	9.	0.188
22	NuTeV-nub'06	33	40.3	0.204	4.72	0.143	3.58	0.109	3.54	0.107	4.55	0.138	5.43	0.165	18.5	0.561
23	ATL7WZ'12	41	39.3	0.16	8.88	0.216	5.28	0.129	3.67	0.0895	4.6	0.112	9.14	0.223	7.78	0.19
24	**LHCb8WZ'16**	42	38.8	0.154	9.68	0.23	5.78	0.138	3.92	0.0932	6.03	0.144	10.8	0.258	2.57	0.0613
25	**CMS8Wasy'16**	33	36.9	0.187	9.62	0.292	5.5	0.167	2.05	0.062	3.92	0.119	9.97	0.302	5.87	0.178
26	E866rat'01	15	34.9	0.388	11.4	0.759	11.2	0.744	2.52	0.168	3.29	0.219	3.92	0.262	2.63	0.175
27	CMS7Masy2'14	11	30.1	0.457	8.91	0.81	5.23	0.475	1.13	0.102	4.71	0.428	8.39	0.762	1.79	0.163
28	**LHCb7ZWrap'15**	33	27.9	0.141	5.61	0.17	4.55	0.138	3.68	0.112	4.68	0.142	6.95	0.211	2.47	0.0749
29	**ATL8ZpT'16**	45	23.4	0.0865	1.58	0.0351	2.8	0.0622	9.71	0.216	1.99	0.0442	1.84	0.0409	5.45	0.121

			Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
30	HERAc'13	47	17.3	0.0614	2.2	0.0468	2.15	0.0457	7.33	0.156	2.41	0.0512	2.28	0.0484	0.955	0.0203
31	D02Masy'08	9	14.5	0.269	2.33	0.259	1.29	0.143	2.92	0.324	3.77	0.418	2.94	0.327	1.3	0.144
32	CMS7Easy'12	11	13.9	0.211	4.24	0.385	2.6	0.236	0.416	0.0378	2.18	0.198	3.76	0.341	0.747	0.0679
33	ZyD02'08	28	13.3	0.0792	2.18	0.0777	1.16	0.0414	1.54	0.055	2.99	0.107	3.42	0.122	2.03	0.0725
34	ZyCDF2'10	29	12.6	0.0722	1.96	0.0677	1.13	0.039	1.49	0.0514	2.73	0.0942	3.15	0.109	2.1	0.0723
35	D02Easy2'15	13	12.3	0.158	2.32	0.178	1.86	0.143	1.17	0.0896	2.19	0.168	3.83	0.295	0.979	0.0753
36	CDF1Wasy'96	11	7.39	0.112	1.37	0.125	0.752	0.0684	0.82	0.0745	1.23	0.112	2.75	0.25	0.461	0.042
37	LHCb7WZ'12	14	7.27	0.0865	1.45	0.104	1.29	0.092	1.03	0.0736	1.15	0.0824	1.72	0.123	0.629	0.0449
38	**ATL8ttb-pt'16**	8	6.91	0.144	1.48	0.185	1.71	0.213	2.78	0.347	0.271	0.0339	0.255	0.0319	0.423	0.0529
39	**LHCb8Zee'15**	17	6.31	0.0619	0.897	0.0528	1.26	0.0743	1.3	0.0765	0.948	0.0558	1.13	0.0665	0.772	0.0454
40	**ATL8ttb-mtt'16**	7	5.94	0.141	0.599	0.0856	0.86	0.123	3.32	0.474	0.345	0.0492	0.382	0.0546	0.433	0.0618
41	**ATL7ZpT'14**	8	5.73	0.119	1.1	0.138	0.989	0.124	1.46	0.183	0.809	0.101	1.06	0.132	0.315	0.0394
42	LHCb7Wasy'12	5	5.4	0.18	1.45	0.291	0.892	0.178	0.269	0.0537	0.866	0.173	1.55	0.31	0.37	0.074
43	**ATL8ttb-y_ave'16**	5	4.95	0.165	0.617	0.123	0.887	0.177	2.23	0.446	0.399	0.0799	0.28	0.0559	0.538	0.108
44	**ATL8ttb-y_ttb'16**	5	4.35	0.145	0.623	0.125	0.716	0.143	2.01	0.403	0.355	0.0709	0.289	0.0579	0.351	0.0702

No.	Expt.	N_{pt}	Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
			$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
45	CDF2Wasy'05	11	3.57	0.0541	0.614	0.0558	0.504	0.0459	0.521	0.0474	0.606	0.055	1.09	0.0991	0.235	0.0214
46	HERA-FL'11	9	1.82	0.0337	0.119	0.0132	0.0199	0.00221	1.37	0.152	0.0239	0.00266	0.118	0.0132	0.174	0.0193
47	HERAb'06	10	1.47	0.0244	0.203	0.0203	0.141	0.0141	0.674	0.0674	0.153	0.0153	0.208	0.0208	0.088	0.0088

Table B.16: Sensitivity values for Table B.14.

No.	Expt.	Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
		$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
1	HERAI+II'15	56.6	0.0505	36.	0.0321	44.9	0.0401	58.8	0.0525	90.6	0.0809	95.4	0.0852	111.	0.0993
2	BCDMSp'89	66.3	0.197	26.4	0.0784	22.5	0.0667	38.	0.113	32.9	0.0975	31.4	0.0932	25.2	0.0748
3	CCFR-F3'97	77.1	0.896	67.2	0.781	10.4	0.121	31.6	0.367	10.1	0.117	11.3	0.131	14.4	0.167
4	**CMS8jets'17**	17.8	0.0962	14.9	0.0806	18.8	0.102	10.3	0.0559	39.9	0.216	40.5	0.219	41.4	0.224
5	NMCrat'97	36.9	0.3	27.8	0.226	49.	0.398	76.	0.618	8.66	0.0704	8.53	0.0693	8.13	0.0661
6	BCDMSd'90	18.7	0.0746	27.7	0.111	14.	0.056	21.8	0.0871	29.6	0.118	28.5	0.114	23.3	0.0933
7	**CMS7jets'13**	9.71	0.073	11.1	0.0831	10.5	0.0789	5.37	0.0404	46.	0.346	46.3	0.348	44.3	0.333
8	**CMS7jets'14**	8.55	0.0541	11.	0.0696	10.1	0.0637	4.64	0.0294	41.3	0.261	40.8	0.258	36.8	0.233
9	CDHSW-F3'91	35.7	0.372	32.4	0.337	4.86	0.0507	14.1	0.147	4.15	0.0433	4.81	0.0501	6.58	0.0686
10	E605'91	18.7	0.157	26.9	0.226	17.5	0.147	4.69	0.0394	9.25	0.0777	9.53	0.0801	9.73	0.0818
11	**ATLAS7jets'15**	6.46	0.0462	7.63	0.0545	10.6	0.0755	3.88	0.0277	22.9	0.164	23.	0.164	22.2	0.159
12	E866pp'03	25.9	0.141	15.1	0.0821	17.1	0.0931	13.5	0.0736	15.	0.0815	16.8	0.0912	20.7	0.112
13	CCFR-F2'01	8.94	0.13	13.5	0.196	8.88	0.129	10.8	0.156	17.4	0.252	16.5	0.239	13.	0.189
14	CDHSW-F2'91	10.6	0.125	13.1	0.154	3.2	0.0377	11.7	0.138	13.5	0.159	11.9	0.14	7.92	0.0931

No.	Expt.	Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
		$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
15	**D02jets'08**	7.5	0.0682	6.58	0.0598	8.85	0.0805	3.82	0.0347	19.1	0.173	18.1	0.165	14.8	0.134
16	NuTeV-nu'06	1.77	0.0465	2.41	0.0634	6.75	0.178	4.4	0.116	8.23	0.217	7.87	0.207	6.62	0.174
17	**CDF2jets'09**	4.12	0.0572	3.74	0.052	6.35	0.0882	2.55	0.0354	14.2	0.198	13.4	0.187	10.6	0.148
18	**ATL7jets'12**	4.56	0.0507	4.96	0.0552	5.28	0.0587	2.2	0.0244	19.9	0.221	19.4	0.216	17.7	0.197
19	CCFR SI nu'01	1.2	0.0299	1.63	0.0408	4.11	0.103	2.92	0.0729	4.69	0.117	4.44	0.111	3.65	0.0913
20	CCFR SI nub'01	2.21	0.0581	1.03	0.0272	2.24	0.059	1.97	0.0519	5.15	0.136	4.95	0.13	4.27	0.112
21	**ATL8DY2D'16**	5.37	0.112	3.39	0.0707	6.03	0.126	7.15	0.149	3.53	0.0736	3.75	0.0781	4.29	0.0894
22	NuTeV-nub'06	2.22	0.0673	1.09	0.033	2.68	0.0813	3.03	0.092	2.62	0.0793	2.52	0.0763	2.2	0.0666
23	ATL7WZ'12	6.95	0.17	4.74	0.116	10.1	0.247	10.1	0.248	4.69	0.114	4.79	0.117	4.7	0.115
24	**LHCb8WZ'16**	4.91	0.117	5.9	0.141	11.7	0.28	12.7	0.302	3.58	0.0853	3.39	0.0807	2.6	0.062
25	**CMS8Wasy'16**	7.02	0.213	5.12	0.155	11.9	0.362	11.3	0.342	3.09	0.0937	3.1	0.0941	2.94	0.0891
26	E866rat'01	4.51	0.301	5.01	0.334	24.8	1.65	4.6	0.307	2.53	0.168	2.52	0.168	2.23	0.149
27	CMS7Masy2'14	4.92	0.447	3.67	0.334	10.7	0.976	11.5	1.04	0.609	0.0554	0.696	0.0633	0.856	0.0778
28	**LHCb7ZWrap'15**	3.66	0.111	5.08	0.154	7.41	0.225	8.33	0.252	3.37	0.102	3.13	0.0949	2.21	0.0669
29	**ATL8ZpT'16**	1.76	0.0391	1.13	0.0252	3.36	0.0746	2.67	0.0593	4.88	0.108	6.04	0.134	9.06	0.201

		Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
No.	Expt.	$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
30	HERAc'13	0.902	0.0192	0.985	0.021	0.614	0.0131	0.84	0.0179	6.79	0.144	6.73	0.143	6.32	0.135
31	D02Masy'08	4.	0.445	2.6	0.289	2.85	0.316	4.36	0.485	2.32	0.257	2.41	0.268	2.4	0.266
32	CMS7Easy'12	1.99	0.181	2.15	0.196	5.2	0.473	5.24	0.477	0.227	0.0207	0.251	0.0228	0.294	0.0268
33	ZyD02'08	2.76	0.0986	2.68	0.0956	1.89	0.0675	2.	0.0714	1.63	0.0583	1.49	0.0531	0.925	0.033
34	ZyCDF2'10	2.54	0.0877	2.6	0.0897	2.15	0.0741	2.38	0.0821	1.35	0.0465	1.17	0.0404	0.679	0.0234
35	D02Easy2'15	2.8	0.215	4.46	0.343	1.26	0.097	4.08	0.314	0.621	0.0478	0.668	0.0514	0.823	0.0633
36	CDF1Wasy'96	1.29	0.118	2.35	0.214	1.24	0.113	3.21	0.292	0.493	0.0448	0.527	0.0479	0.589	0.0535
37	LHCb7WZ'12	0.76	0.0543	1.14	0.0815	1.36	0.0971	1.75	0.125	0.845	0.0604	0.745	0.0532	0.548	0.0391
38	**ATL8ttb-pt'16**	0.96	0.12	1.28	0.16	0.213	0.0266	0.227	0.0284	4.7	0.587	4.75	0.594	4.54	0.568
39	**LHCb8Zee'15**	0.349	0.0205	0.719	0.0423	0.884	0.052	1.14	0.0673	0.707	0.0416	0.586	0.0345	0.208	0.0122
40	**ATL8ttb-mtt'16**	0.629	0.0898	0.844	0.121	0.387	0.0553	0.231	0.033	2.23	0.319	2.34	0.334	2.5	0.357
41	**ATL7ZpT'14**	0.242	0.0302	0.193	0.0241	0.349	0.0436	0.547	0.0683	1.81	0.226	1.79	0.223	1.59	0.199
42	LHCb7Wasy'12	0.726	0.145	1.02	0.204	1.68	0.336	1.88	0.376	0.588	0.118	0.594	0.119	0.568	0.114
43	**ATL8ttb-y_ave'16**	0.324	0.0648	0.529	0.106	0.254	0.0508	0.151	0.0302	2.17	0.433	2.16	0.431	1.94	0.389
44	**ATL8ttb-y_ttb'16**	0.15	0.0301	0.499	0.0998	0.139	0.0278	0.141	0.0283	2.05	0.41	2.01	0.401	1.71	0.341

No.	Expt.	Values, **CT14 HERA2 NNLO PDFs fitted without jet data**													
		$\lvert S^E_{u_v}\rvert$	$\langle\lvert S^E_{u_v}\rvert\rangle$	$\lvert S^E_{d_v}\rvert$	$\langle\lvert S^E_{d_v}\rvert\rangle$	$\lvert S^E_{\bar{d}/\bar{u}}\rvert$	$\langle\lvert S^E_{\bar{d}/\bar{u}}\rvert\rangle$	$\lvert S^E_{d/u}\rvert$	$\langle\lvert S^E_{d/u}\rvert\rangle$	$\lvert S^E_{H7}\rvert$	$\langle\lvert S^E_{H7}\rvert\rangle$	$\lvert S^E_{H8}\rvert$	$\langle\lvert S^E_{H8}\rvert\rangle$	$\lvert S^E_{H14}\rvert$	$\langle\lvert S^E_{H14}\rvert\rangle$
45	CDF2Wasy'05	0.638	0.058	0.768	0.0698	0.614	0.0558	1.36	0.124	0.413	0.0375	0.438	0.0398	0.468	0.0426
46	HERA-FL'11	0.0789	0.00876	0.043	0.00478	0.122	0.0136	0.119	0.0132	0.662	0.0736	0.582	0.0647	0.314	0.0349
47	HERAb'06	0.0953	0.00953	0.0979	0.00979	0.0933	0.00933	0.111	0.0111	0.676	0.0676	0.658	0.0658	0.564	0.0564

Appendix C

Tabulated Sensitivities

The appendix gives tables similar to Table VI of Appendix B.1 (or Table VI of the first PDFSense paper), in this case consisting of ranking tables of the Mellin moments $\langle x^n \rangle$ and for the isovector quasi-PDF at several values of x and P_z.

Table C.1: The aggregated sensitivities to moments of the $u^{\pm}$ quark distributions of the experiments in the CTEQ-TEA set. Here and in the subsequent tables, we arrange the CTEQ-TEA experiments in descending order based on their summed sensitivity $\sum|S^E|$ to each of the three moments displayed in the rightmost columns.

No.	ID	N_{pt}	$\sum\|S^E\|$	$\langle\sum\|S^E\|\rangle$	$\|S_{\langle x^1\rangle_{u^+}}\|$	$\langle\|S_{\langle x^1\rangle_{u^+}}\|\rangle$	$\|S_{\langle x^2\rangle_{u^-}}\|$	$\langle\|S_{\langle x^2\rangle_{u^-}}\|\rangle$	$\|S_{\langle x^3\rangle_{u^+}}\|$	$\langle\|S_{\langle x^3\rangle_{u^+}}\|\rangle$
1	BCDMSp'89	337	125.	0.123	A	3	A	3	A	3
2	HERAI+II'15	1120.	122.	0.0363	A		A		A	
3	CCFR-F3'97	86	95.8	0.371	A	2	A	1	A	2
4	E866pp'03	184	69.3	0.125	B		A	3	A	3
5	BCDMSd'90	250	55.3	0.0737	A		B		B	
6	CDHSW-F3'91	96	52.7	0.183	B	3	A	3	B	3
7	CDHSW-F2'91	85	35.7	0.14	B	3	C	3	B	3
8	NMCrat'97	123	34.5	0.0934	C		B	3	B	3
9	CCFR-F2'01	69	31.5	0.152	C	3	C	3	B	3
10	CMS7jets'13	133	27.5	0.0689	C		C		B	
11	CMS8jets'17	185	25.3	0.0456	C		C		B	
12	E605'91	119	24.	0.0671			B	3		
13	CMS7jets'14	158	19.5	0.0411	C		C		C	
14	ATLAS7jets'15	140	16.2	0.0387			C		C	
15	NuTeV-nu'06	38	12.9	0.113	C	3				
16	LHCb8WZ'16	42	12.5	0.0989					C	3
17	CCFR SI nub'01	38	11.9	0.104	C	3				
18	NuTeV-nub'06	33	11.8	0.119	C	3				3
19	CMS7Masy2'14	11	11.5	0.349		2	C	2	C	2
20	LHCb7ZWrap'15	33	11.4	0.115				3	C	3
21	ATL7jets'12	90	11.1	0.0412					C	
22	ATL8DY2D'16	48	10.2	0.0706	C					
23	D02jets'08	110	10.2	0.0308					C	
24	E866rat'01	15	10.	0.223	C	2		3		3
25	CCFR SI nu'01	40	9.71	0.0809	C	3				
26	CDF2jets'09	72	8.51	0.0394						
27	CMS8Wasy'16	33	7.18	0.0726						
28	ATL7WZ'12	41	6.97	0.0567						
29	CMS7Easy'12	11	6.81	0.206		3		3		3
30	D02Easy2'15	13	6.46	0.166				3		2
31	ATL8ZpT'16	45	5.21	0.0386						

No.	ID	N_{pt}	$\sum\lvert S^E\rvert$	$\langle\sum\lvert S^E\rvert\rangle$	$\lvert S_{\langle x^1\rangle_{u^+}}\rvert$	$\langle\lvert S_{\langle x^1\rangle_{u^+}}\rvert\rangle$	$\lvert S_{\langle x^2\rangle_{u^-}}\rvert$	$\langle\lvert S_{\langle x^2\rangle_{u^-}}\rvert\rangle$	$\lvert S_{\langle x^3\rangle_{u^+}}\rvert$	$\langle\lvert S_{\langle x^3\rangle_{u^+}}\rvert\rangle$
32	ZyCDF2'10	29	4.88	0.0561						
33	ZyD02'08	28	4.61	0.0549						
34	D02Masy'08	9	4.15	0.154		2				3
35	LHCb8Zee'15	17	3.32	0.065						
36	HERAc'13	47	3.15	0.0224						
37	CDF1Wasy'96	11	2.75	0.0835						3
38	LHCb7WZ'12	14	2.34	0.0557						
39	ATL8ttb-mtt'16	7	1.2	0.0573						
40	ATL8ttb-pt'16	8	1.17	0.0489						
41	LHCb7Wasy'12	5	1.17	0.0781						3
42	ATL7ZpT'14	8	1.04	0.0435		3				
43	CDF2Wasy'05	11	1.02	0.0311						
44	ATL8ttb-y_ave'16	5	0.548	0.0365						
45	HERA-FL'11	9	0.508	0.0188						
46	ATL8ttb-y_ttb'16	5	0.495	0.033						
47	HERAb'06	10	0.328	0.0109						

Table C.2: The sensitivities to moments of the $d^{\pm}$ quark distributions of the CTEQ-TEA experiments in the CTEQ-TEA set.

No.	ID	N_{pt}	$\sum\lvert S^E\rvert$	$\langle\sum\lvert S^E\rvert\rangle$	$\lvert S_{\langle x^1\rangle_{d^+}}\rvert$	$\langle\lvert S_{\langle x^1\rangle_{d^+}}\rvert\rangle$	$\lvert S_{\langle x^2\rangle_{d^-}}\rvert$	$\langle\lvert S_{\langle x^2\rangle_{d^-}}\rvert\rangle$	$\lvert S_{\langle x^3\rangle_{d^+}}\rvert$	$\langle\lvert S_{\langle x^3\rangle_{d^+}}\rvert\rangle$
1	HERAI+II'15	1120.	116.	0.0346	A*		A		A	
2	CCFR-F3'97	86	70.3	0.272	B	3	A	2	B	3
3	BCDMSd'90	250	58.8	0.0784	A	3	B		B	
4	NMCrat'97	123	55.6	0.151	A	3	B	3	B	3
5	BCDMSp'89	337	53.8	0.0532	B		B		A	
6	E866pp'03	184	44.2	0.08	B		B		B	
7	CDHSW-F2'91	85	43.5	0.171	B	3	B	3	B	3
8	E605'91	119	37.3	0.105	C		A	3	C	
9	CDHSW-F3'91	96	36.2	0.126	C		B	3	C	
10	CCFR-F2'01	69	27.4	0.133	B	3	C	3	C	3
11	LHCb8WZ'16	42	16.4	0.13			C	3	C	3
12	CMS7jets'13	133	16.2	0.0405	C		C		C	
13	CMS8jets'17	185	15.2	0.0273	C				C	
14	LHCb7ZWrap'15	33	14.	0.141			C	3	C	3
15	CMS7jets'14	158	13.4	0.0283	C				C	
16	D02Easy2'15	13	12.	0.309		3	C	1		2
17	NuTeV-nu'06	38	11.7	0.103	C	3				
18	CMS7Masy2'14	11	10.2	0.308	C	2		2		3
19	ATLAS7jets'15	140	9.11	0.0217					C	
20	ATL8DY2D'16	48	8.82	0.0612						
21	CCFR SI nu'01	40	8.61	0.0718	C	3				
22	E866rat'01	15	8.48	0.188		3		3		3
23	ATL7WZ'12	41	8.38	0.0681						
24	CCFR SI nub'01	38	7.62	0.0668	C	3				
25	CMS8Wasy'16	33	6.99	0.0706						
26	NuTeV-nub'06	33	6.68	0.0674	C	3				
27	ATL7jets'12	90	6.61	0.0245						
28	D02jets'08	110	6.14	0.0186						
29	CMS7Easy'12	11	5.72	0.173		3		3		3
30	CDF1Wasy'96	11	5.12	0.155		3		2		
31	ATL8ZpT'16	45	4.99	0.037						
32	D02Masy'08	9	4.35	0.161		3		2		
33	ZyD02'08	28	3.8	0.0452						
34	LHCb7WZ'12	14	3.77	0.0898				3		

No.	ID	N_{pt}	$\sum\lvert S^E\rvert$	$\langle\sum\lvert S^E\rvert\rangle$	$\lvert S_{\langle x^1\rangle_{d+}}\rvert$	$\langle\lvert S_{\langle x^1\rangle_{d+}}\rvert\rangle$	$\lvert S_{\langle x^2\rangle_{d-}}\rvert$	$\langle\lvert S_{\langle x^2\rangle_{d-}}\rvert\rangle$	$\lvert S_{\langle x^3\rangle_{d+}}\rvert$	$\langle\lvert S_{\langle x^3\rangle_{d+}}\rvert\rangle$
35	CDF2jets'09	72	3.58	0.0166						
36	ZyCDF2'10	29	3.06	0.0352						
37	LHCb7Wasy'12	5	2.78	0.185				2		3
38	ATL7ZpT'14	8	2.34	0.0976		3				
39	HERAc'13	47	1.97	0.014						
40	CDF2Wasy'05	11	1.86	0.0565						
41	LHCb8Zee'15	17	1.63	0.032						
42	ATL8ttb-pt'16	8	1.47	0.0614		3				
43	ATL8ttb-mtt'16	7	1.37	0.0654						
44	ATL8ttb-y_ave'16	5	0.776	0.0517						
45	ATL8ttb-y_ttb'16	5	0.461	0.0307						
46	HERA-FL'11	9	0.407	0.0151						
47	HERAb'06	10	0.252	0.0084						

Table C.3: The sensitivities to moments of the $u^{\pm}-d^{\pm}$ isovector quark distributions of the CTEQ-TEA experiments in the CTEQ-TEA set.

No.	ID	N_{pt}	$\sum\lvert S^E\rvert$	$\langle\sum\lvert S^E\rvert\rangle$	$\lvert S_{\langle x^1\rangle_{u^+-d^+}}\rvert$	$\langle\lvert S_{\langle x^1\rangle_{u^+-d^+}}\rvert\rangle$	$\lvert S_{\langle x^2\rangle_{u^--d^-}}\rvert$	$\langle\lvert S_{\langle x^2\rangle_{u^--d^-}}\rvert\rangle$	$\lvert S_{\langle x^3\rangle_{u^+-d^+}}\rvert$	$\langle\lvert S_{\langle x^3\rangle_{u^+-d^+}}\rvert\rangle$
1	HERAI+II'15	1120.	109.	0.0325	A		A		A	
2	BCDMSp'89	337	91.9	0.0909	B		A		A	3
3	E866pp'03	184	67.3	0.122	A	3	A	3	A	3
4	CCFR-F3'97	86	61.9	0.24	A	2	B	3	B	3
5	NMCrat'97	123	60.	0.163	A	3	B	3	B	3
6	CDHSW-F3'91	96	27.4	0.0953	B	3	C		C	
7	BCDMSd'90	250	25.5	0.034	C		C		C	
8	LHCb8WZ'16	42	15.3	0.122	C		C	3	C	3
9	CMS7Masy2'14	11	15.3	0.463	C	1	C	2		2
10	CCFR-F2'01	69	15.2	0.0733	C		C		C	
11	E605'91	119	15.1	0.0423	C				C	
12	CMS8jets'17	185	14.1	0.0254			C		C	
13	LHCb7ZWrap'15	33	12.9	0.13	C	3	C	3		3
14	D02Easy2'15	13	11.6	0.298		3	C	2		2
15	CDHSW-F2'91	85	11.4	0.0449	C					
16	E866rat'01	15	11.3	0.251	C	2		3		3
17	CMS7jets'13	133	11.2	0.0282			C			
18	ATL8DY2D'16	48	11.1	0.0772	C					
19	CMS7Easy'12	11	8.98	0.272		2		2		3
20	CMS8Wasy'16	33	8.77	0.0886		3				
21	ATL7WZ'12	41	8.64	0.0703						
22	CMS7jets'14	158	8.54	0.018						
23	ATLAS7jets'15	140	8.07	0.0192						
24	CDF2jets'09	72	5.22	0.0241						
25	D02jets'08	110	5.	0.0151						
26	NuTeV-nu'06	38	4.99	0.0437						
27	ATL7jets'12	90	4.84	0.0179						
28	CDF1Wasy'96	11	4.83	0.146		3		3		3
29	NuTeV-nub'06	33	4.28	0.0432						
30	D02Masy'08	9	4.2	0.155		3		3		3
31	CCFR SI nub'01	38	3.71	0.0325						

No.	ID	N_{pt}	$\sum\lvert S^E\rvert$	$\langle\sum\lvert S^E\rvert\rangle$	$\lvert S_{\langle x^1\rangle_{u^+-d^+}}\rvert$	$\langle\lvert S_{\langle x^1\rangle_{u^+-d^+}}\rvert\rangle$	$\lvert S_{\langle x^2\rangle_{u^--d^-}}\rvert$	$\langle\lvert S_{\langle x^2\rangle_{u^--d^-}}\rvert\rangle$	$\lvert S_{\langle x^3\rangle_{u^+-d^+}}\rvert$	$\langle\lvert S_{\langle x^3\rangle_{u^+-d^+}}\rvert\rangle$
32	CCFR SI nu'01	40	3.61	0.0301						
33	ZyCDF2'10	29	3.58	0.0411						
34	ATL8ZpT'16	45	3.29	0.0243						
35	ZyD02'08	28	3.11	0.0371						
36	LHCb7WZ'12	14	2.88	0.0685						
37	LHCb7Wasy'12	5	2.47	0.165				2		3
38	HERAc'13	47	1.79	0.0127						
39	LHCb8Zee'15	17	1.74	0.034						
40	CDF2Wasy'05	11	1.73	0.0524						
41	ATL7ZpT'14	8	1.2	0.05						
42	ATL8ttb-pt'16	8	0.951	0.0396						
43	ATL8ttb-mtt'16	7	0.517	0.0246						
44	ATL8ttb-y_ave'16	5	0.236	0.0157						
45	ATL8ttb-y_ttb'16	5	0.164	0.0109						
46	HERAb'06	10	0.162	0.0054						
47	HERA-FL'11	9	0.15	0.00554						

BIBLIOGRAPHY

[1] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008. 2

[2] D. H. Perkins, *Introduction to High Energy Physics*. Cambridge University Press, 4 ed., 2000, 10.1017/CBO9780511809040. 2

[3] Particle Data Group collaboration, M. Tanabashi et al., *Review of Particle Physics*, *Phys. Rev.* **D98** (2018) 030001. ix, 6, 8, 18

[4] ZEUS collaboration, M. Derrick et al., *Measurement of the F_2 structure function in deep inelastic e^+p scattering using 1994 data from the ZEUS detector at HERA*, *Z. Phys.* **C72** (1996) 399–424, [hep-ex/9607002]. 11

[5] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys.* **B126** (1977) 298–318. 11

[6] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, *Sov. Phys. JETP* **46** (1977) 641–653. 11

[7] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438–450. 11

[8] C. Willis, R. Brock, D. Hayden, T.-J. Hou, J. Isaacson, C. Schmidt et al., *New method for reducing parton distribution function uncertainties in the high-mass Drell-Yan spectrum*, *Phys. Rev.* **D99** (2019) 054004, [1809.09481]. 12

[9] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.* **D93** (2016) 033006, [1506.07443]. ix, 13, 18, 26, 29, 79

[10] K. Kovařík, P. M. Nadolsky and D. E. Soper, *Hadron structure in high-energy collisions*, 1905.06957. 13

[11] J. Gao, L. Harland-Lang and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, *Phys. Rept.* **742** (2018) 1–121, [1709.04922]. 13, 27, 49

[12] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky et al., *CT10 next-to-next-to-leading order global analysis of QCD*, *Phys. Rev.* **D89** (2014) 033009, [1302.6246]. 15, 37, 38

[13] 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, [hep-ph/0201195]. 15, 17, 30, 36, 38, 48

[14] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk et al., *Uncertainties of predictions from parton distribution functions. 2. The Hessian method*, *Phys. Rev.* **D65** (2001) 014013, [hep-ph/0101032]. 16, 30, 36, 48, 49

[15] W. T. Giele, S. A. Keller and D. A. Kosower, *Parton Distribution Function Uncertainties*, hep-ph/0104052. 16, 36, 53

[16] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump et al., *Implications of CTEQ global analysis for collider observables*, *Phys. Rev.* **D78** (2008) 013004, [0802.0007]. 17, 30, 31, 49, 50

[17] NNPDF collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione et al., *A Determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* **B809** (2009) 1–63, [0808.1231]. 17, 30, 53

[18] T. J. Hobbs, *Recent progress on intrinsic charm*, *EPJ Web Conf.* **137** (2017) 08008, [1612.05686]. 18

[19] H.-W. Lin et al., *Parton distributions and lattice QCD calculations: a community white paper*, *Prog. Part. Nucl. Phys.* **100** (2018) 107–160, [1711.07916]. 19, 34, 80, 84, 86, 91

[20] W. Detmold, W. Melnitchouk and A. W. Thomas, *Parton distributions from lattice QCD*, *Eur. Phys. J.direct* **3** (2001) 13, [hep-lat/0108002]. 19, 86

[21] W. Detmold, W. Melnitchouk and A. W. Thomas, *Extraction of parton distributions from lattice QCD*, *Mod. Phys. Lett.* **A18** (2003) 2681–2698, [hep-lat/0310003]. 19, 81, 86

[22] W. Detmold, W. Melnitchouk and A. W. Thomas, *Parton distribution functions in the pion from lattice QCD*, *Phys. Rev.* **D68** (2003) 034025, [hep-lat/0303015]. 19, 86

[23] X. Ji, *Parton Physics on a Euclidean Lattice*, *Phys. Rev. Lett.* **110** (2013) 262002, [1305.1539]. 19, 34, 81, 107

[24] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* **G43** (2016) 023001, [1510.03865]. 20, 27, 128

[25] G. Apollinari, O. Brüning, T. Nakamoto and L. Rossi, *High Luminosity Large Hadron Collider HL-LHC*, *CERN Yellow Rep.* (2015) 1–19, [1705.08830]. 20, 126

[26] LHeC Study Group collaboration, J. L. Abelleira Fernandez et al., *A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector*, *J. Phys.* **G39** (2012) 075001, [1206.2913]. 20, 34, 110, 126

[27] A. Accardi et al., *Electron Ion Collider: The Next QCD Frontier*, *Eur. Phys. J.* **A52** (2016) 268, [1212.1701]. 20, 34, 110, 127

[28] F. Bordry, M. Benedikt, O. Brüning, J. Jowett, L. Rossi, D. Schulte et al., *Machine Parameters and Projected Luminosity Performance of Proposed Future Colliders at CERN*, 1810.13022. 20, 21, 22

[29] "https://home.cern/news/news/accelerators/record-luminosity-well-done-lhc." 20

[30] "https://indico.cern.ch/event/656250/contributions/2939331/attachments/1634262/-2606713/caseLHeCmaxklein.pdf." 21

[31] M. Klein and R. Yoshida, *Collider Physics at HERA*, Prog. Part. Nucl. Phys. **61** (2008) 343–393, [0805.3334]. 21

[32] "https://www.jlab.org/conferences/elba/talks/monday/morning_session/Yoshida.pdf." 21

[33] "https://www.bnl.gov/aum2017/content/plenary/pdf/eRHIC_RHIC-AGS-Users-meetig-2017_vs2.pdf." 21

[34] B.-T. Wang, T. J. Hobbs, S. Doyle, J. Gao, T.-J. Hou, P. M. Nadolsky et al., *Mapping the sensitivity of hadronic experiments to nucleon structure*, Phys. Rev. **D98** (2018) 094030, [1803.02777]. xiii, 23, 82, 87, 88, 89, 90, 93, 94, 104, 127, 128, 130

[35] 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) 204, [1412.3989]. 26, 62, 79

[36] NNPDF collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, Eur. Phys. J. **C77** (2017) 663, [1706.00428]. 26, 79

[37] S. Alekhin, J. Blümlein, S. Moch and R. Placakyte, *Parton distribution functions, α_s, and heavy-quark masses for LHC Run II*, Phys. Rev. **D96** (2017) 014011, [1701.05838]. 26

[38] 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) 114017, [1602.03154]. 26, 79

[39] H1, ZEUS 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) 580, [1506.06042]. ix, 26, 29, 141

[40] S. Alekhin et al., *HERAFitter*, Eur. Phys. J. **C75** (2015) 304, [1410.4412]. 26, 38

[41] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *CTEQ-TEA parton distribution functions and HERA Run I and II combined data*, Phys. Rev. **D95** (2017) 034003, [1609.07968]. ix, 28, 29, 40, 54, 79, 90, 131

[42] HERAFitter developers' Team collaboration, S. Camarda et al., *QCD analysis of W- and Z-boson production at Tevatron*, Eur. Phys. J. **C75** (2015) 458, [1503.05221]. 28

[43] NNPDF collaboration, R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti et al., *Reweighting NNPDFs: the W lepton asymmetry*, Nucl. Phys. **B849** (2011) 112–143, [1012.0836]. 28

[44] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti et al., *Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data*, Nucl. Phys. **B855** (2012) 608–638, [1108.1758]. 28

[45] N. Sato, J. F. Owens and H. Prosper, *Bayesian Reweighting for Global Fits*, *Phys. Rev.* **D89** (2014) 114020, [1310.1089]. 30

[46] H. Paukkunen and P. Zurita, *PDF reweighting in the Hessian matrix approach*, *JHEP* **12** (2014) 100, [1402.6623]. 30

[47] P. M. Nadolsky and Z. Sullivan, *PDF Uncertainties in WH Production at Tevatron*, *eConf* **C010630** (2001) P510, [hep-ph/0110378]. 30, 36, 49

[48] S. Carrazza, S. Forte, Z. Kassabov and J. Rojo, *Specialized minimal PDFs for optimized LHC calculations*, *Eur. Phys. J.* **C76** (2016) 205, [1602.00005]. 30

[49] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog et al., *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, *JHEP* **05** (2016) 058, [1602.00695]. 31

[50] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera and J. Rojo, *Pinning down the large-x gluon with NNLO top-quark pair differential distributions*, *JHEP* **04** (2017) 044, [1611.08609]. 31

[51] R. Boughezal, A. Guffanti, F. Petriello and M. Ubiali, *The impact of the LHC Z-boson transverse momentum data on PDF determinations*, *JHEP* **07** (2017) 130, [1705.00343]. 31, 75

[52] D. Boer et al., *Gluons and the quark sea at high energies: Distributions, polarization, tomography*, 1108.1713. 34, 110

[53] S. Abeyratne et al., *Science Requirements and Conceptual Design for a Polarized Medium Energy Electron-Ion Collider at Jefferson Lab*, 1209.0757. 34, 110

[54] E. C. Aschenauer et al., *eRHIC Design Study: An Electron-Ion Collider at BNL*, 1409.1633. 34, 110

[55] M. Gockeler, R. Horsley, E.-M. Ilgenfritz, H. Perlt, P. E. L. Rakow, G. Schierholz et al., *Polarized and unpolarized nucleon structure functions from lattice QCD*, *Phys. Rev.* **D53** (1996) 2317–2325, [hep-lat/9508004]. 34, 86

[56] C. Schmidt, J. Pumplin, C. P. Yuan and P. Yuan, *Updating and optimizing error parton distribution function sets in the Hessian approach*, *Phys. Rev.* **D98** (2018) 094005, [1806.07950]. 35, 66

[57] G. R. Farrar and D. R. Jackson, *Pion and Nucleon Structure Functions Near x=1*, *Phys. Rev. Lett.* **35** (1975) 1416. 35

[58] T. J. Hobbs, M. Alberg and G. A. Miller, *Constraining nucleon strangeness*, *Phys. Rev.* **C91** (2015) 035205, [1412.4871]. 35, 79

[59] T. J. Hobbs, J. T. Londergan and W. Melnitchouk, *Phenomenology of nonperturbative charm in the nucleon*, *Phys. Rev.* **D89** (2014) 074008, [1311.1578]. 35, 74, 79

[60] W. T. Giele and S. Keller, *Implications of hadron collider observables on parton distribution function uncertainties*, *Phys. Rev.* **D58** (1998) 094023, [hep-ph/9803393]. 36, 53

[61] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk et al., *Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method*, *Phys. Rev.* **D65** (2001) 014012, [hep-ph/0101051]. 36, 66, 68

[62] R. D. Ball et al., *Parton Distribution Benchmarking with LHC Data*, *JHEP* **04** (2013) 125, [1211.5142]. 38

[63] "http://projector.tensorflow.org." xi, 42, 52

[64] "PDFSensewebsite:http://metapdf.hepforge.org/PDFSense/." xi, 42, 52, 54, 60, 66, 71, 73

[65] D. Cook, U. Laa and G. Valencia, *Dynamical projections for the visualization of PDFSense data*, *Eur. Phys. J.* **C78** (2018) 742, [1806.09742]. 42

[66] L. van der Maaten and G. Hinton, , *Journel of Machine Learning Research* **9** (2008) 2579. 43

[67] CMS collaboration, V. Khachatryan et al., *Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV*, *JHEP* **03** (2017) 156, [1609.05331]. 58, 143

[68] J. Pumplin, *Experimental consistency in parton distribution fitting*, *Phys. Rev.* **D81** (2010) 074010, [0909.0268]. 62

[69] CMS collaboration, V. Khachatryan et al., *Constraints on parton distribution functions and extraction of the strong coupling constant from the inclusive jet cross section in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C75** (2015) 288, [1410.6765]. 65

[70] J. Pumplin, D. R. Stump and W. K. Tung, *Multivariate fitting and the error matrix in global analysis of data*, *Phys. Rev.* **D65** (2001) 014011, [hep-ph/0008191]. 66, 68

[71] R. Brock, D. Casey, J. Huston, J. Kalk, J. Pumplin, D. Stump et al., *Uncertainties of parton distribution functions and their implications on physical predictions*, in *Workshop on B Physics at the Tevatron: Run II and Beyond Batavia, Illinois, September 23-25, 1999*, pp. 159–161, 2000. hep-ph/0006148. 66, 68

[72] CMS collaboration, A. M. Sirunyan et al., *Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions*, *Eur. Phys. J.* **C77** (2017) 459, [1703.01630]. 66

[73] T.-J. H. et al., , *in preparation* (2018) . 66

[74] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *CT14 Intrinsic Charm Parton Distribution Functions from CTEQ-TEA Global Analysis*, *JHEP* **02** (2018) 059, [1707.00657]. 74

[75] J. Speth and A. W. Thomas, *Mesonic contributions to the spin and flavor structure of the nucleon*, *Adv. Nucl. Phys.* **24** (1997) 83–149. 79

[76] S. Kumano, *Flavor asymmetry of anti-quark distributions in the nucleon*, *Phys. Rept.* **303** (1998) 183–257, [hep-ph/9702367]. 79

[77] C. Bourrely, J. Soffer and F. Buccella, *A Statistical approach for polarized parton distributions*, *Eur. Phys. J.* **C23** (2002) 487–501, [hep-ph/0109160]. 79

[78] R. J. Holt and C. D. Roberts, *Distribution Functions of the Nucleon and Pion in the Valence Region*, Rev. Mod. Phys. **82** (2010) 2991–3044, [1002.4666]. 79

[79] H. Avakian, A. V. Efremov, P. Schweitzer and F. Yuan, *The transverse momentum dependent distribution functions in the bag model*, Phys. Rev. **D81** (2010) 074035, [1001.5467]. 79

[80] K. D. Bednar, I. C. Cloët and P. C. Tandy, *Nucleon quark distribution functions from the Dyson–Schwinger equations*, Phys. Lett. **B782** (2018) 675–681, [1803.03656]. 79

[81] W. Melnitchouk and M. Malheiro, *Strange asymmetries in the nucleon sea*, Phys. Lett. **B451** (1999) 224–232, [hep-ph/9901321]. 79

[82] T. J. Hobbs, M. Alberg and G. A. Miller, *Bayesian analysis of light-front models and the nucleon's charmed sigma term*, Phys. Rev. **D96** (2017) 074023, [1707.06711]. 79

[83] W. Zimmermann, *Normal products and the short distance expansion in the perturbation theory of renormalizable interactions*, Annals Phys. **77** (1973) 570–601. 83

[84] K. G. Wilson, *Nonlagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499–1512. 83, 85

[85] N. H. Christ, B. Hasslacher and A. H. Mueller, *Light cone behavior of perturbation theory*, Phys. Rev. **D6** (1972) 3543. 83

[86] D. J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories - I*, Phys. Rev. **D8** (1973) 3633–3652. 83

[87] D. J. Gross and F. Wilczek, *ASYMPTOTICALLY FREE GAUGE THEORIES. 2.*, Phys. Rev. **D9** (1974) 980–993. 83

[88] H. Georgi and H. D. Politzer, *Electroproduction scaling in an asymptotically free theory of strong interactions*, Phys. Rev. **D9** (1974) 416–420. 83

[89] H. Georgi and H. D. Politzer, *Freedom at Moderate Energies: Masses in Color Dynamics*, Phys. Rev. **D14** (1976) 1829. 85

[90] J. Blumlein and H. Bottcher, *Higher Twist Contributions to the Structure Functions $F_2^p(x, Q^2)$ and $F_2^d(x, Q^2)$ at Large x and Higher Orders*, Phys. Lett. **B662** (2008) 336–340, [0802.0408]. 85

[91] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, Eur. Phys. J. **C77** (2017) 112, [1607.00299]. 86

[92] M. Gockeler, R. Horsley, E.-M. Ilgenfritz, H. Perlt, P. E. L. Rakow, G. Schierholz et al., *Lattice operators for moments of the structure functions and their transformation under the hypercubic group*, Phys. Rev. **D54** (1996) 5705–5714, [hep-lat/9602029]. 87

[93] Y. Aoki, T. Blum, H.-W. Lin, S. Ohta, S. Sasaki, R. Tweedie et al., *Nucleon isovector structure functions in (2+1)-flavor QCD with domain wall fermions*, Phys. Rev. **D82** (2010) 014501, [1003.3387]. 93

[94] H.-W. Lin, J.-W. Chen, S. D. Cohen and X. Ji, *Flavor Structure of the Nucleon Sea from Lattice QCD*, Phys. Rev. **D91** (2015) 054510, [1402.1462]. 93

[95] J.-W. Chen, T. Ishikawa, L. Jin, H.-W. Lin, Y.-B. Yang, J.-H. Zhang et al., *Parton distribution function with nonperturbative renormalization from lattice QCD*, *Phys. Rev.* **D97** (2018) 014505, [1706.01295]. 93

[96] R. Gupta, Y.-C. Jang, B. Yoon, H.-W. Lin, V. Cirigliano and T. Bhattacharya, *Isovector Charges of the Nucleon from 2+1+1-flavor Lattice QCD*, *Phys. Rev.* **D98** (2018) 034503, [1806.09006]. 93

[97] Y.-S. Liu, J.-W. Chen, L. Jin, H.-W. Lin, Y.-B. Yang, J.-H. Zhang et al., *Unpolarized quark distribution from lattice QCD: A systematic analysis of renormalization and matching*, 1807.06566. xiii, 93, 107, 108, 124

[98] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato and F. Steffens, *Light-Cone Parton Distribution Functions from Lattice QCD*, *Phys. Rev. Lett.* **121** (2018) 112001, [1803.02685]. 93

[99] J.-H. Zhang, J.-W. Chen, L. Jin, H.-W. Lin, A. Schäfer and Y. Zhao, *First direct lattice-QCD calculation of the x-dependence of the pion parton distribution function*, *Phys. Rev.* **D100** (2019) 034505, [1804.01483]. 93

[100] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu and D. G. Richards, *Pion Valence Quark Distribution from Matrix Element Calculated in Lattice QCD*, *Phys. Rev.* **D99** (2019) 074507, [1901.03921]. 93

[101] M. Oehm, C. Alexandrou, M. Constantinou, K. Jansen, G. Koutsou, B. Kostrzewa et al., *$\langle x \rangle$ and $\langle x^2 \rangle$ of the pion PDF from lattice QCD with $N_f = 2+1+1$ dynamical quark flavors*, *Phys. Rev.* **D99** (2019) 014508, [1810.09743]. 93

[102] T. J. Hobbs, *Quantifying finite-momentum effects in the quark quasidistribution functions of mesons*, *Phys. Rev.* **D97** (2018) 054028, [1708.05463]. 93

[103] M. Deka, T. Streuer, T. Doi, S. J. Dong, T. Draper, K. F. Liu et al., *Moments of Nucleon's Parton Distribution for the Sea and Valence Quarks from Lattice QCD*, *Phys. Rev.* **D79** (2009) 094502, [0811.1779]. 96, 99

[104] A. W. Thomas, *A Limit on the Pionic Component of the Nucleon Through SU(3) Flavor Breaking in the Sea*, *Phys. Lett.* **126B** (1983) 97–100. 104

[105] A. I. Signal, A. W. Schreiber and A. W. Thomas, *Flavor SU(2) symmetry breaking in deep inelastic scattering*, *Mod. Phys. Lett.* **A6** (1991) 271–276. 104

[106] A. W. Schreiber, P. J. Mulders, A. I. Signal and A. W. Thomas, *The Pion cloud of the nucleon and its effect on deep inelastic structure*, *Phys. Rev.* **D45** (1992) 3069–3078. 104

[107] M. Alberg and G. A. Miller, *Taming the Pion Cloud of the Nucleon*, *Phys. Rev. Lett.* **108** (2012) 172001, [1201.4184]. 104

[108] Y. Salamu, C.-R. Ji, W. Melnitchouk and P. Wang, *$\bar{d} - \bar{u}$ asymmetry in the proton in chiral effective theory*, *Phys. Rev. Lett.* **114** (2015) 122001, [1409.5885]. 104

[109] New Muon collaboration, P. Amaudruz et al., *The Gottfried sum from the ratio F_2^n/F_2^p*, *Phys. Rev. Lett.* **66** (1991) 2712–2715. 105

[110] New Muon collaboration, M. Arneodo et al., *A Reevaluation of the Gottfried sum*, *Phys. Rev.* **D50** (1994) R1–R3. 105

[111] K. Gottfried, *Sum rule for high-energy electron - proton scattering*, *Phys. Rev. Lett.* **18** (1967) 1174. 105

[112] M. Constantinou and H. Panagopoulos, *Perturbative renormalization of quasi-parton distribution functions*, *Phys. Rev.* **D96** (2017) 054506, [1705.11193]. 108

[113] G. Apollinari, I. Béjar Alonso, O. Brüning, M. Lamont and L. Rossi, *High-Luminosity Large Hadron Collider (HL-LHC) : Preliminary Design Report*, . 110

[114] S. J. Brodsky, F. Fleuret, C. Hadjidakis and J. P. Lansberg, *Physics Opportunities of a Fixed-Target Experiment using the LHC Beams*, *Phys. Rept.* **522** (2013) 239–255, [1202.6585]. 110

[115] J. Dudek et al., *Physics Opportunities with the 12 GeV Upgrade at Jefferson Lab*, *Eur. Phys. J.* **A48** (2012) 187, [1208.1244]. 110

[116] "http://metapdf.hepforge.org/PDFSense/Lattice/." 119

[117] T. J. Hobbs, B.-T. Wang, P. M. Nadolsky and F. I. Olness, *The coming synergy between lattice QCD and high-energy phenomenology*, 1904.00022. 127, 132

[118] "https://tinyurl.com/FutureExpts-DIS19." 127, 128, 129, 132

[119] M. Klein, "LHeC pseudodata." http://hep.ph.liv.ac.uk/~mklein/lhecdata/, 2017. 128

[120] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang and J. Rojo, *Towards Ultimate Parton Distributions at the High-Luminosity LHC*, *Eur. Phys. J.* **C78** (2018) 962, [1810.03639]. 128

[121] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang and J. Rojo, *Probing Proton Structure at the Large Hadron electron Collider*, 1906.10127. 132

[122] J. R. Andersen et al., *Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report*, in *9th Les Houches Workshop on Physics at TeV Colliders (PhysTeV 2015) Les Houches, France, June 1-19, 2015*, 2016. 1605.04692. 135

[123] S. Atashbar Tehrani, *Nuclear parton distribution functions (nPDFs) and their uncertainties in the LHC Era*, in *Proceedings, 3rd International Conference on Particle Physics and Astrophysics (ICPPA 2017): Moscow, Russia, October 2-5, 2017*, 2017. 1712.02153. 135

[124] K. Kovarik, T. Jezo, A. Kusina, F. I. Olness, I. Schienbein, T. Stavreva et al., *CTEQ nuclear parton distribution functions*, *PoS* **DIS2013** (2013) 274, [1307.3454]. 135

[125] R. Angeles-Martinez et al., *Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects*, *Acta Phys. Polon.* **B46** (2015) 2501–2534, [1507.05267]. 135

[126] LHC Higgs Cross Section Working Group collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, 1101.0593. 136

[127] S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, 1201.3084. 136

[128] CMS collaboration, A. M. Sirunyan et al., *Measurement of the weak mixing angle using the forward-backward asymmetry of Drell-Yan events in pp collisions at 8 TeV*, Eur. Phys. J. **C78** (2018) 701, [1806.00863]. 136

[129] W. Beenakker, C. Borschensky, M. Krämer, A. Kulesza, E. Laenen, S. Marzani et al., *NLO+NLL squark and gluino production cross-sections with threshold-improved parton distributions*, Eur. Phys. J. **C76** (2016) 53, [1510.00375]. 136

[130] BCDMS collaboration, A. C. Benvenuti et al., *A High Statistics Measurement of the Proton Structure Functions $F_2(x, Q^2)$ and R from Deep Inelastic Muon Scattering at High Q^2*, Phys. Lett. **B223** (1989) 485–489. 140

[131] BCDMS collaboration, A. C. Benvenuti et al., *A High Statistics Measurement of the Deuteron Structure Functions $F_2(x, Q^2)$ and R From Deep Inelastic Muon Scattering at High Q^2*, Phys. Lett. **B237** (1990) 592–598. 140

[132] New Muon collaboration, M. Arneodo et al., *Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T*, Nucl. Phys. **B483** (1997) 3–43, [hep-ph/9610231]. 141

[133] J. P. Berge et al., *A Measurement of Differential Cross-Sections and Nucleon Structure Functions in Charged Current Neutrino Interactions on Iron*, Z. Phys. **C49** (1991) 187–224. 141

[134] CCFR/NuTeV collaboration, U.-K. Yang et al., *Measurements of F_2 and $xF_3^\nu - xF_3^{\bar{\nu}}$ from CCFR ν_μ−Fe and $\bar{\nu}_\mu$−Fe data in a physics model independent way*, Phys. Rev. Lett. **86** (2001) 2742–2745, [hep-ex/0009041]. 141

[135] W. G. Seligman et al., *Improved determination of $\alpha(s)$ from neutrino nucleon scattering*, Phys. Rev. Lett. **79** (1997) 1213–1216, [hep-ex/9701017]. 141

[136] D. A. Mason, *Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data*. PhD thesis, Oregon U., 2006. 10.2172/879078. 141

[137] NuTeV collaboration, M. Goncharov et al., *Precise Measurement of Dimuon Production Cross-Sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe Deep Inelastic Scattering at the Tevatron.*, Phys. Rev. **D64** (2001) 112006, [hep-ex/0102049]. 141

[138] H1 collaboration, A. Aktas et al., *Measurement of F2($c\bar{c}$) and F2($b\bar{b}$) at high Q^2 using the H1 vertex detector at HERA*, Eur. Phys. J. **C40** (2005) 349–359, [hep-ex/0411046]. 141

[139] H1 collaboration, A. Aktas et al., *Measurement of $F_2^{c\bar{c}}$ and $F_2^{b\bar{b}}$ at low Q^2 and x using the H1 vertex detector at HERA*, Eur. Phys. J. **C45** (2006) 23–33, [hep-ex/0507081]. 141

[140] H1, ZEUS 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) 2311, [1211.1182]. 141

[141] H1 collaboration, F. D. Aaron et al., *Measurement of the Inclusive e±p Scattering Cross Section at High Inelasticity y and of the Structure Function F_L*, Eur. Phys. J. **C71** (2011) 1579, [1012.4355]. 141

[142] G. Moreno et al., *Dimuon production in proton - copper collisions at $\sqrt{s}$ = 38.8-GeV*, Phys. Rev. **D43** (1991) 2815–2836. 141

[143] NuSea collaboration, R. S. Towell et al., *Improved measurement of the $\bar{d}/\bar{u}$ asymmetry in the nucleon sea*, Phys. Rev. **D64** (2001) 052002, [hep-ex/0103030]. 141

[144] NuSea collaboration, J. C. Webb et al., *Absolute Drell-Yan dimuon cross-sections in 800 GeV / c pp and pd collisions*, hep-ex/0302019. 141

[145] CDF collaboration, F. Abe et al., *Forward-backward charge asymmetry of electron pairs above the Z^0 pole*, Phys. Rev. Lett. **77** (1996) 2616–2621. 141

[146] CDF collaboration, D. Acosta et al., *Measurement of the forward-backward charge asymmetry from $W \to e\nu$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. **D71** (2005) 051104, [hep-ex/0501023]. 141

[147] D0 collaboration, V. M. Abazov et al., *Measurement of the muon charge asymmetry from W boson decays*, Phys. Rev. **D77** (2008) 011106, [0709.4254]. 141

[148] LHCb collaboration, R. Aaij et al., *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, JHEP **06** (2012) 058, [1204.1620]. 141, 142

[149] D0 collaboration, V. M. Abazov et al., *Measurement of the ratios of the $Z/\gamma^* + >= n$ jet production cross sections to the total inclusive Z/γ^* cross section in $p\bar{p}$ collisions at $s^(1/2) = 1.96$ TeV*, Phys. Lett. **B658** (2008) 112–119, [hep-ex/0608052]. 142

[150] CDF collaboration, T. A. Aaltonen et al., *Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Lett. **B692** (2010) 232–239, [0908.3914]. 142

[151] CMS collaboration, S. Chatrchyan et al., *Measurement of the muon charge asymmetry in inclusive $pp \to W + X$ production at $\sqrt{s}$ = 7 TeV and an improved determination of light parton distribution functions*, Phys. Rev. **D90** (2014) 032004, [1312.6283]. 142

[152] CMS collaboration, S. Chatrchyan et al., *Measurement of the Electron Charge Asymmetry in Inclusive W Production in pp Collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **109** (2012) 111806, [1206.2598]. 142

[153] ATLAS collaboration, G. Aad et al., *Measurement of the inclusive $W^\pm$ and Z/γ cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. **D85** (2012) 072004, [1109.5141]. 142

[154] D0 collaboration, V. M. Abazov et al., *Measurement of the electron charge asymmetry in $\boldsymbol{p\bar{p}\to W+X\to e\nu+X}$ decays in $\boldsymbol{p\bar{p}}$ collisions at $\boldsymbol{\sqrt{s}=1.96}$ TeV*, Phys. Rev. **D91** (2015) 032007, [1412.2862]. 142

[155] CDF collaboration, T. Aaltonen et al., *Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron $p\bar{p}$ Collider Using a Cone-Based Jet Algorithm*, Phys. Rev. **D78** (2008) 052006, [0807.2204]. 142

[156] D0 collaboration, V. M. Abazov et al., *Measurement of the inclusive jet cross-section in $p\bar{p}$ collisions at $s^{(1/2)} = 1.96$ TeV*, Phys. Rev. Lett. **101** (2008) 062001, [0802.2400]. 142

[157] ATLAS collaboration, G. Aad et al., *Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. **D86** (2012) 014022, [1112.6297]. 142

[158] CMS collaboration, S. Chatrchyan et al., *Measurements of Differential Jet Cross Sections in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV with the CMS Detector*, Phys. Rev. **D87** (2013) 112002, [1212.6660]. 142

[159] LHCb collaboration, R. Aaij et al., *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **08** (2015) 039, [1505.07024]. 142

[160] LHCb collaboration, R. Aaij et al., *Measurement of forward $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ production at $\sqrt{s} = 8$ TeV*, JHEP **05** (2015) 109, [1503.00963]. 142

[161] ATLAS collaboration, G. Aad et al., *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **09** (2014) 145, [1406.3660]. 142

[162] CMS collaboration, V. Khachatryan et al., *Measurement of the differential cross section and charge asymmetry for inclusive $\mathrm{pp} \to \mathrm{W}^\pm + X$ production at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. **C76** (2016) 469, [1603.01803]. 142

[163] LHCb collaboration, R. Aaij et al., *Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **01** (2016) 155, [1511.08039]. 142

[164] ATLAS collaboration, G. Aad et al., *Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **08** (2016) 009, [1606.01736]. 142

[165] ATLAS collaboration, G. Aad et al., *Measurement of the transverse momentum and ϕ^*_η distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. **C76** (2016) 291, [1512.02192]. 142

[166] CMS collaboration, S. Chatrchyan et al., *Measurement of the Ratio of Inclusive Jet Cross Sections using the Anti-k_T Algorithm with Radius Parameters R=0.5 and 0.7 in pp Collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. **D90** (2014) 072006, [1406.0324]. 142

[167] ATLAS collaboration, G. Aad et al., *Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using 4.5 fb^{-1} of data with the ATLAS detector*, JHEP **02** (2015) 153, [1410.8857]. 143

[168] ATLAS collaboration, G. Aad et al., *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) 538, [1511.04716]. 143