

Studies of Parton Distributions in Nuclei

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von

Marina Walt, geb. Holzmann

aus Ust-Kamenogorsk/Kasachstan

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	12.03.2020
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Werner Vogelsang
2. Berichterstatter:	Prof. Dr. Thomas Gutsche

„MAN MUSS NOCH CHAOS IN SICH HABEN,
UM EINEN TANZENDEN STERN GEBÄREN ZU
KÖNNEN.“

- *Friedrich Nietzsche*

Abstract (engl.)

The subject of this thesis is the global analysis of nuclear parton distribution functions (nPDFs). The main goal of the project is to determine a new set of nPDFs at next-to-next-to-leading order (NNLO) in perturbative quantum chromodynamics (pQCD) and to make it publicly available for the scientific community. The purpose of nPDFs is to describe the collinear momentum distribution of the partons (quarks and gluons) inside nucleons which are bound to a nucleus. The increase in perturbative precision in the nPDF determination enables the application of the most accurate calculations for the partonic scattering processes with nuclei. Furthermore, the improved accuracy of nPDFs is important for the analysis of the heavy-ion experiments at the LHC and RHIC, and also to make predictions for future research projects such as, for example, the Electron-Ion Collider (EIC).

Neither free proton PDFs, nor nuclear PDFs can be calculated from first principles, but can be derived in a global analysis by applying suitable data, i.e. from the scattering experiments with nucleons (protons and neutrons) bound in an atom's nucleus in case of nPDFs. According to the collinear factorization theorem, the perturbatively calculable partonic scattering processes can be factorized from the non-perturbative PDFs, which require a non-perturbative input. The necessary scale evolution of this initial non-perturbative input is given by pQCD.

In this work two new nPDF analyses, performed at next-to-leading order (NLO) and NNLO in pQCD are introduced. The resulting PDF sets, referred to as TUJU19, are released in a standardized LHAPDF format. The presented framework is based on an open-source tool for proton PDF analysis, xFITTER, which has been modified to be applicable also for a nuclear PDF analysis. The required extensions of the code have been published, providing a first open-source tool to analyze nuclear PDFs. The framework

of the analysis, including the form of the initial parameterization of PDFs as well as the included data sets from deeply inelastic scattering (DIS) experiments are described. The performed uncertainty analysis and the results of an alternative method of error analysis applied for the obtained nPDFs are discussed. The results of the QCD analysis in this thesis are compared to the existing nPDF sets and to the fitted data. A very good agreement with most of the applied data is found and the resulting nPDFs are consistent with previous works. Finally, a study on further processes, i.e. Drell-Yan and W^\pm, Z boson production, is presented and an outlook is given.

Kurzfassung (dt.)

Der Gegenstand dieser Arbeit ist die globale Analyse von nuklearen Partonverteilungsfunktionen (engl. nPDFs). Das Ziel des Projektes ist es ein neues Set von nPDFs bis zur zweiten Ordnung der Störungstheorie (engl. NNLO) zu bestimmen und diese der wissenschaftlichen Öffentlichkeit zur Verfügung zu stellen. Die nPDFs beschreiben die kollineare Impulsverteilung eines im Kern gebundenen Nukleons auf seine Bestandteile, die Partonen (Quarks und Gluonen). Das Erhöhen der störungstheoretischen Genauigkeit bei der Ermittlung der nPDFs erlaubt eine Anwendung für hoch-präzise Berechnungen der partonischen Streuprozesse mit Kernen. Darüber hinaus ist die Verbesserung der Genauigkeit wichtig für die Analyse der Schwerionen-Experimente am LHC und am RHIC, und auch um Voraussagen für die zukünftigen Forschungsprojekte, wie zum Beispiel das Electron-Ion Collider (EIC), zu machen.

PDFs können nicht anhand der grundlegenden theoretischen Prinzipien berechnet werden. Stattdessen werden diese in einer globalen Analyse durch den Vergleich mit experimentellen Daten, aus Streuexperimenten mit den im Atomkern gebundenen Nukleonen (Protonen und Neutronen) im Fall von nPDFs, ermittelt. Gemäß dem kollinearen Faktorisierungstheorem kann der störungstheoretisch berechenbare partonische Streuprozess separat von den nicht-perturbativen PDFs betrachtet werden, die einen nicht-perturbativen Input benötigen. Die notwendige Skalen-Entwicklung dieses initialen, nicht-perturbativen Inputs kann mit den Methoden der störungstheoretischen Quanten-Chromodynamik (engl. pQCD) erfolgen.

In dieser Dissertation werden zwei nPDF Analysen, in erster (NLO) und zweiter (NNLO) Ordnung der Störungstheorie vorgestellt. Die resultierenden PDF Sets, die als TUJU19 bezeichnet werden, wurden in einem standardisierten LHAPDF Format veröffentlicht. Diese Arbeit basiert auf einem Open-Source-Tool für Proton PDF Analyse, xFITTER,

welches für die Zwecke der nuklearen PDF Analyse modifiziert wurde. Die implementierten Erweiterungen des numerischen Programmcodes wurden veröffentlicht und stellen somit das erste Open-Source-Tool für eine nPDF Analyse dar. Der theoretische und numerische Rahmen dieser Arbeit, d. h. sowohl die Form der initialen Parametrisierung der Verteilungsfunktionen, als auch die verwendeten Daten aus Experimenten der tiefinelastischen Streuung (engl. DIS), sind hier beschrieben. Die durchgeführte Analyse der experimentellen Ungenauigkeiten und die Ergebnisse einer alternativen Methode der Fehleranalyse, angewendet auf nPDFs, werden diskutiert. Die Ergebnisse dieser QCD Analyse werden mit den bestehenden nPDF Sets und den experimentellen Daten verglichen. Die Übereinstimmung mit den meisten verwendeten Messdaten ist sehr gut und die resultierenden nPDF Sets sind im Einklang mit den vorangegangenen Arbeiten. Abschließend wird eine Anwendung auf weitere Prozesse, wie z. B. Drell-Yan und die Produktion der W^\pm , Z -Bosonen, präsentiert.

Contents

Abstract (engl.)	v
Kurzfassung (dt.)	vii
List of Abbreviations	xiii
List of Publications	xvii
1. Introduction	1
1.1. Motivation	1
1.2. Nuclear and Particle Physics	2
1.3. Global nPDF Analysis	6
1.4. Structure of the Thesis	8
2. QCD in a Nutshell	11
2.1. Lagrangian	11
2.2. Perturbative QCD	14
3. Theoretical Principles	19
3.1. Nucleon Structure and Nuclear Effects	19
3.2. Deeply Inelastic Scattering	25
3.3. Parton Model and Scaling Violations	30
3.4. Factorization Theorem and Parton Distribution Functions	34
4. DGLAP Evolution	37
4.1. DGLAP Equations and Splitting Functions	37
4.2. Solving DGLAP Equations	40
4.2.1. Solution in N -space	41

4.2.2.	Solution in x -space	46
4.2.3.	Numerical Implementations	47
4.3.	Higher-Order Corrections	47
4.4.	Limitations of DGLAP Equations	48
5.	Framework of the Global Analysis	51
5.1.	Theoretical Framework	52
5.1.1.	PDF Parameterization	52
5.1.2.	Approximations and Constraints	53
5.1.3.	Fitting Procedure	55
5.2.	Numerical Framework	57
5.3.	Summary: Proton vs. Nucleon	61
6.	Study I: Comparison to DIS Data	65
6.1.	Experimental DIS Data	66
6.2.	Corrections and Experimental Uncertainties	68
6.2.1.	Isoscalar Modifications	68
6.2.2.	Correlated Uncertainties	70
6.3.	Results NC and CC DIS	71
6.4.	Study of Neutrino-Nucleus DIS	79
7.	Study II: Uncertainty Analysis	85
7.1.	Hessian Method	86
7.2.	Monte Carlo Method	89
7.3.	Results: Hessian vs. MC Method	92
7.3.1.	Comparison for Proton Baseline	92
7.3.2.	Comparison for nPDFs	93
8.	Study III: Parton Distribution Functions TUJU19	101
8.1.	Proton Baseline	102
8.2.	Nuclear PDFs	103
8.3.	Comparison to nPDF Sets	106
8.3.1.	Comparison at NLO	106
8.3.2.	Comparison at NNLO	109
8.3.3.	Summary	112

9. Study IV: Application to Further Processes	115
9.1. Drell-Yan and W^\pm , Z Boson Production Processes	116
9.2. Application of TUJU19 for W^\pm Boson Production	119
9.3. Updated Free Proton Baseline TUJU20	122
9.4. Outlook	129
10. Summary and Outlook (engl.)	131
10.1. Summary	131
10.2. Outlook	134
11. Zusammenfassung und Ausblick (dt.)	137
11.1. Zusammenfassung	137
11.2. Ausblick	143
A. Final PDF Parameters	145
B. User Guide for nPDFs in xFitter	147
B.1. Introduction	147
B.2. nPDF Parameterization	147
B.3. Nuclear Data	152
B.4. Output and Error Analysis	153
B.5. Change Log	154
C. Calculation of \sqrt{s} from Polarization	157
D. MC Uncertainty Analysis: Scaling Properties	159
List of Figures	165
List of Tables	177
Bibliography	179
Acknowledgements / Danksagung	203

List of Abbreviations

ABM	Alekhin-Blümlein-Moch
APFEL	A Parton Distribution Function Evolution Library
ATLAS	A Toroidal LHC ApparatuS
BCDMS	A Collaboration at CERN
BFKL	Balitsky-Fadin-Kuraev-Lipatov
BK	Balitsky-Kovchegov
CC	Charged Current
CDHSW	A Collaboration at CERN
CERN	Conseil Européen pour la Recherche Nucléaire
CGC	Color Glass Condensate
CHORUS	A Collaboration at CERN
CMS	Compact Muon Solenoid
CSV	Charge Symmetry Violation
CTEQ	Coordinated Theoretical-Experimental project on QCD
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi
DIS	Deeply Inelastic Scattering
DLLA	Double-Leading-Logarithm Approximation

List of Abbreviations

DSSZ	de Florian, Sassot, Stratmann, Zurita
DY	Drell-Yan
EMC	European Muon Collaboration
EPS16	Eskola, Paakkinen, Paukkunen, Salgado 2016
FF	Fragmentation Function
FNAL	Fermi National Accelerator Laboratory
FONLL	Fixed-Order plus Next-to-Leading Logs
GM-VFNS	General-Mass Variable-Flavour-Number-Scheme
GPD	Generalized Parton Distribution
GR	General Relativity
GSF	Gesellschaft für Schwerionenforschung
HERA	Hadron-Elektron-Ringanlage (dt.), Hadron-Electron Ring Accelerator (engl.)
HOPPET	Higher Order Perturbative Parton Evolution Toolkit
IMF	Infinite Momentum Frame
IR	Infrared
JIMWLK	Jalilian-Marian, Iancu, McLerran, Weigert, Leonidov, Kovner
JR09	Jimenez-Delgado, Reya 2009
KA15	Khanpour, Atashbar Tehrani 2015
LHAPDF	Les Houches accord PDFs
LHC	Large Hadron Collider
LO	Leading Order
MC	Monte Carlo

MELA	Mellin Evolution LibrAry
MRT	Magnetic Resonance Tomography
MS ($\overline{\text{MS}}$)	(Modified) Minimal Subtraction
nCTEQ15	Nuclear CTEQ 2015
NC	Neutral Current
NLO	Next-to-Leading Order
NMC	New Muon Collaboration
NNLO	Next-to-Next-to-Leading Order
nNNPDF1.0	Nuclear NNPDF 1.0
NNPDF	Neural Networks Parton Distribution Function
nPDF	Nuclear Parton Distribution Function
NuTeV	A Collaboration at Fermilab Tevatron
PDF	Parton Distribution Function
PEGASUS	Parton Evolution Generated Applying Symbolic U -matrix Solutions
pQCD	Perturbative QCD
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QFT	Quantum Field Theory
RT	Thorne-Roberts
SACOT	Simplified Aivazis-Collins-Olness-Tung
SLAC	Stanford Linear Accelerator Center
SRC	Short Range Correlations

List of Abbreviations

TMD	Transverse-Momentum Dependent PDF
TUJU19	Tübingen University, Jyväskylä University 2019
UV	Ultraviolet
ZM-VFNS	Zero-Mass Variable-Flavour-Number-Scheme

List of Publications

This thesis consists of an introductory part and the results mostly presented in publications [1] and [2].

Journal Papers

- [1] M. Walt, I. Helenius, and W. Vogelsang. “Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO”. in: *Phys. Rev. D* 100.9 (2019), p. 096015. arXiv: 1908.03355 [hep-ph]

Conference Proceedings

- [2] M. Walt, I. Helenius, and W. Vogelsang. “A QCD analysis for nuclear PDFs at NNLO”. in: *PoS DIS2019* (2019), p. 039. arXiv: 1908.04983 [hep-ph]

The author has worked on the concept creation, numerical implementation and execution of the calculations for both papers [1] and [2]. The experimental proton data applied for the analysis were used as provided by xFITTER developers team, whereas the experimental nuclear data have been collected from different public sources and transferred into the proper format by the author. All numerical results presented in these articles have been obtained by the programs largely modified by the author. The performed modifications of the open-source code have been documented for the user guidance by the author and made publicly available. The author has also prepared the released sets of nuclear parton distribution functions TUJU19 in a standard format. The author has written the original drafts of the manuscript (except introduction) for publication [1]. The author has written the complete original draft version for publication [2], which is based on a conference talk contributed, and prepared, by the author.

1. Introduction

1.1. Motivation

A precise knowledge of the inner structure of the visible matter is one of the key questions for the scientific particle physics society. The information on the building blocks of the nuclei inside the atoms is important in order to analyse the experiments, understand the nature and to make predictions for future research projects. Although it is so important, the existing knowledge is quite limited due to the complication that nuclei are bound objects with highly non-trivial dependencies between the single building blocks. The high-energy physics community requires a description how the momentum of a bound nucleon is distributed along its constituents. These constituents (quarks and gluons) are called "partons" in high-energy particle physics and the required information is provided in a form of so-called parton distribution functions (PDFs). These functions are rather well known for free protons, but are less well known for protons and neutrons which are bound in an atom's nucleus. The need to have dedicated nuclear PDFs is caused by the experimentally detected nuclear effects in the measured observables as demonstrated in figure 1.1. The lack of precision in nuclear PDFs (nPDFs) results as a theoretical uncertainty in the baseline calculations for different experiments, such as heavy-ion programme of the Large Hadron Collider (LHC) at CERN.

In addition to the academic and scientific relevance, the knowledge of the inner structure of nuclei can lead to many applications, e.g. in medicine, like for instance the well-established magnetic resonance tomography (MRT) for imaging the inside of living human bodies, or the method of cancer therapy with heavy ions currently being used in studies at GSI in Darmstadt. A recent example of an application based on insights from nuclear physics is the nuclear clock [3] developed at GSI.

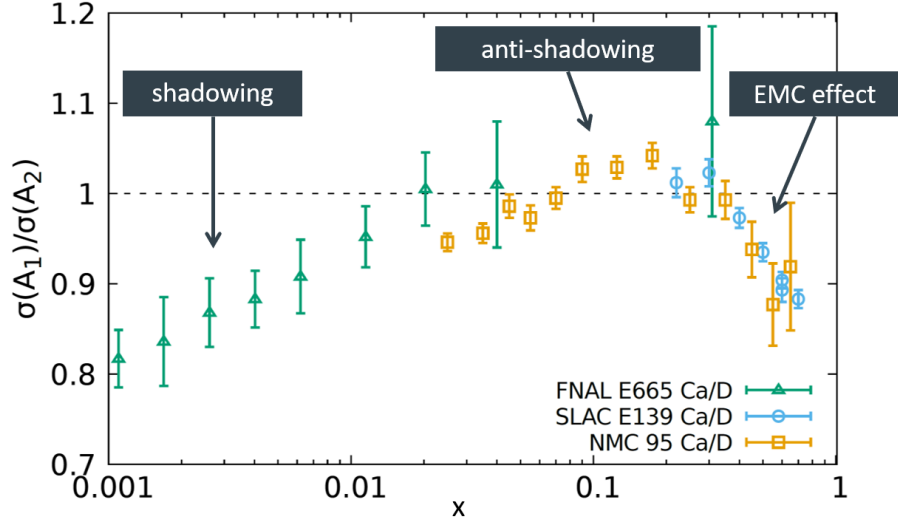


Figure 1.1.: Illustration of sample data points measured by the experimental collaborations FNAL [4], SLAC [5] and NMC [6]. The functional form of the ratio $\sigma(A_1)/\sigma(A_2)$ along momentum fraction x for nucleons in two different nuclei with mass numbers A_1, A_2 is characterized by the nuclear effects named shadowing, anti-shadowing and EMC effect. The dashed line symbolizes the ratio $\sigma(A_1)/\sigma(A_2) \approx 1$ which one would expect without nuclear effects. The observed deviation of $\sigma(A_1)/\sigma(A_2)$ from ≈ 1 due to the nuclear effects, that will be discussed later (cf. sec. 3.1), causes the need for nuclear parton distribution functions.

1.2. Nuclear and Particle Physics

As argued in the first section, the understanding and the scientific description of the inner structure of the visible matter is one of the key aspects of the fundamental physics research. Nowadays, it is known that all visible matter consists of atoms with a comparably small but massive nucleus in the center of the atom and electrons orbiting around it, as shown in figure 1.2. Furthermore, it has been observed that nuclei are formed by bound nucleons - protons and neutrons. Each nucleon itself consists of the so-called partons - quarks and gluons. As per the underlying quantum field theory (QFT) gluons are fields mediating the interaction between the quarks. As per the Standard Model of particle physics, quarks are fermionic particles having spin quantum number $1/2$, which the hadronic part of all visible matter of the universe is supposed to be made of. In the context of PDFs we differentiate between valence quarks and sea quarks. Valence quarks are constituents of the nucleon at any energy scale and carry its mass effectively. They

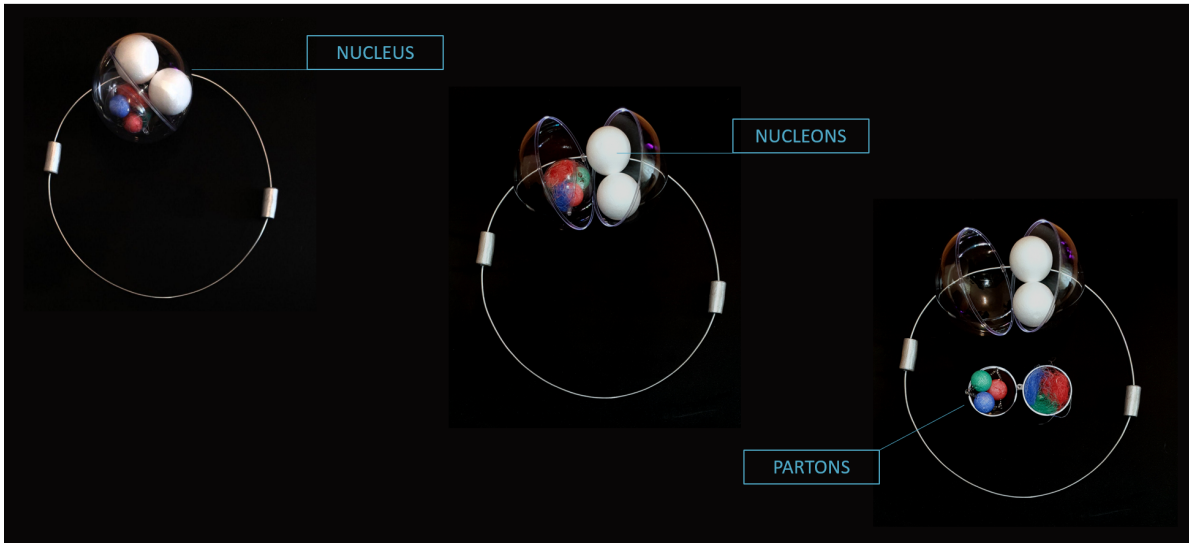


Figure 1.2.: A visual model of the atomic substructure. On the left-hand side an atom consisting of a nucleus and orbiting electrons is shown. Inside the nucleus there are nucleons - protons and neutrons. On the right-hand side the inner structure of a nucleon is indicated. It is made up of partons - quarks, i.e. sea quarks and valence quarks, and gluons.

are supposed to determine the properties of the hadron, e.g. the electric charge. For example, a proton p consists of three valence quarks $p = (uud)$, two up quarks (u) and one down quark (d). Apart from valence quarks there are also sea quarks inside the hadron.

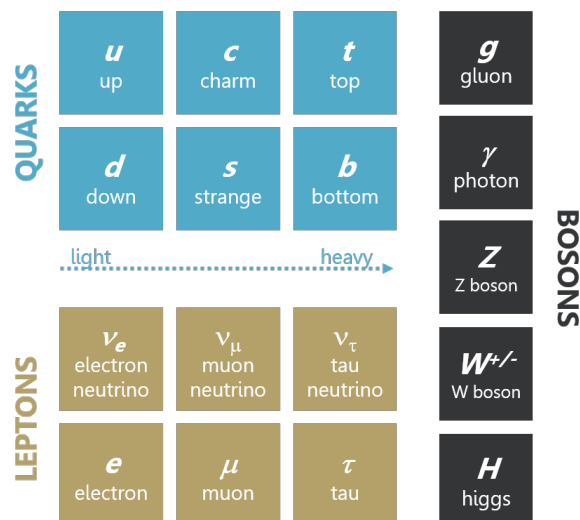


Figure 1.3.: Particles in the Standard Model of particle physics.

1. Introduction

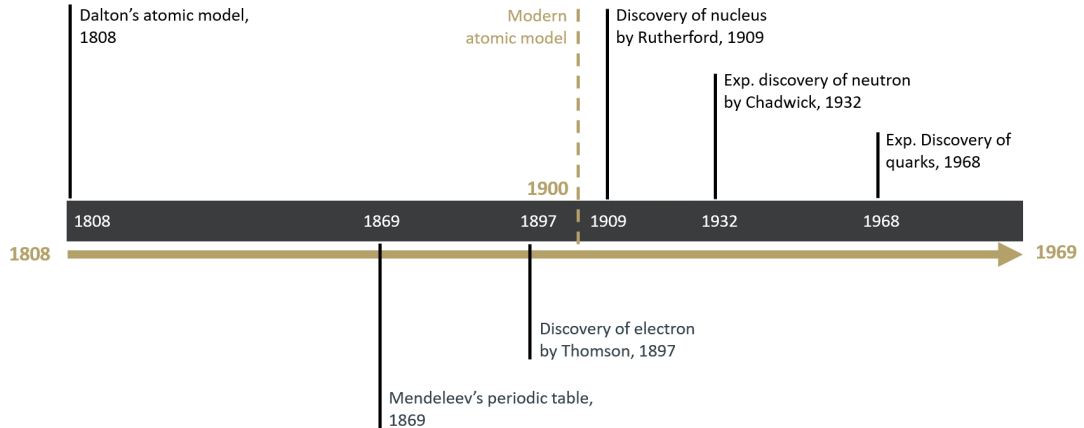


Figure 1.4.: Timeline from 1808 to 1969. Milestones on the way to quarks and leptons. To be continued in figure 1.5.

One of the important unsolved problems in physics, the so-called spin puzzle [7–10], is related to the role of valence and sea quarks inside a hadron. One question of the spin puzzle is how the spin $1/2$ of the proton is built up from the quantum numbers of the quarks and gluons inside. However, as part of a global PDF analysis the heavy flavours of the sea quarks, e.g. charm (c), are radiated at higher energy scales used to probe the nucleons and the necessary scale evolution can be explained by the QCD dynamics, which is described in the following chapters. An overview of all particles predicted by the Standard Model [11] and confirmed by the experiments is shown in figure 1.3.

As per the Standard Model (fig. 1.3) there are six quark flavours with different masses. The up, down and strange quarks are called light, whereas charm, top and bottom are considered as heavy quarks. Only quarks and gluons interact strongly. The other elementary fermionic particles, leptons, are objects to electromagnetic and weak interactions, but *not* the strong interaction. Leptons, e.g. an electron, are scattered off the nucleus, or more precisely the partons, in the deeply inelastic scattering (DIS) experiments considered as part of this analysis. As can be seen in figure 1.3, beside gluons there are further bosons, for example photon γ which is the exchange particle of the electromagnetic interaction, and W^\pm, Z bosons mediating weak interaction. These bosons are involved in the Drell-Yan (DY) and W^\pm, Z boson production processes discussed in the last chapter of this work.

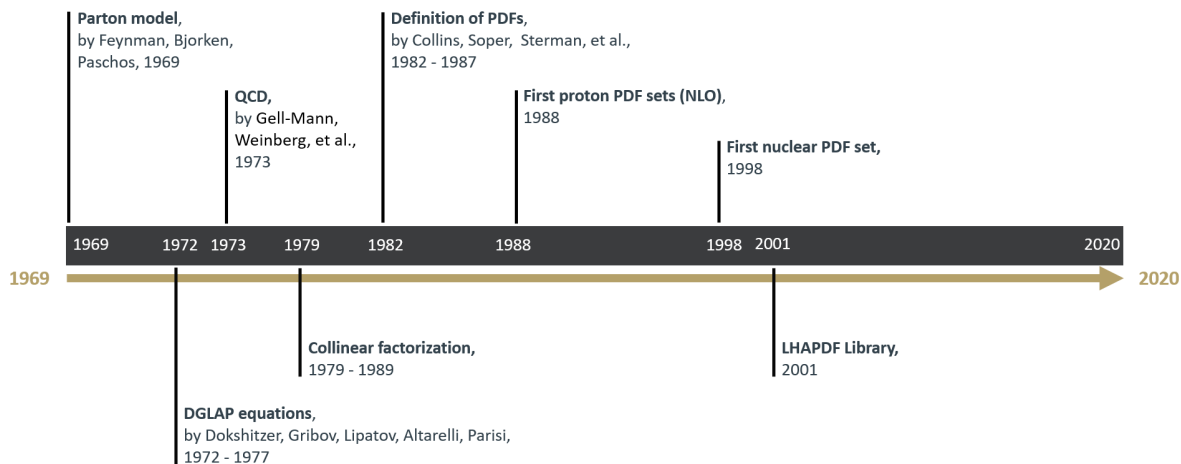


Figure 1.5.: Timeline from 1969 to 2020. Summary of concepts and developments in theoretical particle physics in particular relevant for this work.

For the modern understanding of the substructure of an atom shown in figure 1.2 a series of developments and discoveries over the past 200 years was necessary. Figure 1.4 shows the milestones on the way to quarks and leptons, and figure 1.5 summarizes concepts and developments in theoretical particle physics which are relevant for this work.

Although the building blocks of an atom are known nowadays, it was not obvious just 200 years ago. As per DALTON's atomic model [12, 13], an atom was supposed to be the smallest elementary, non-dividable particle. This theory was disproved and revised in the early 20th century by BOHR et al. However, without an exact knowledge of the atomic substructure MENDELEEV was able to arrange the chemical elements in a periodic table [14] in 1869. Later, in 1897 THOMSON's discovery of the electron [15] contributed to the formulation of the modern atomic model. The experimental evidence for the nucleus followed in 1909 by RUTHERFORD and GEIGER [16]. Finally, the discovery of the neutron by CHADWICK in 1932 [17, 18] and the experimental proof of the existence of quarks, e.g. as per references [19, 20], lead to the basis for the analysis performed in this work.

Beside the understanding of the inner structure of the visible matter, several concepts and developments in theoretical particle physics (cf. fig. 1.5) were necessary to allow a

performance of a global analysis and the determination of parton distribution functions. The most important concepts, like parton model [21, 22], collinear factorization [23, 24] and Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [25–28] will be described in details in the next chapters. In 1973 Quantum Chromodynamics (QCD), the theory that describes the interactions between quarks and gluons, was proposed in several works by FRITSCH, GELL-MANN, LEUTWYLER, GROSS, WILCZEK, and WEINBERG in 1973 [29–32]. Another important contribution was the definition of PDFs, e.g. by COLLINS et al. [33], leading to the first PDF sets appearing in the 1980s. As mentioned earlier, the main goal of this study of parton distributions in nuclei is to determine a new nPDF set at NNLO which is made available for the scientific community through the LHAPDF library [34–36]. This library was invented in 2001, although first proton PDF sets at NLO were already available in 1988, e.g. [37], and first nuclear PDFs were published in 1998, e.g. [38–40], as summarized in figure 1.5.

This section was meant to provide a short introduction into the current Standard Model of particle physics and an overview of historical developments. The relevant theoretical concepts and further details of the performed nPDF analysis are described in the following chapters.

1.3. Global nPDF Analysis

As mentioned in the previous section, the fundamental theory that describes the interactions between quarks and gluons is QCD. Due to the complicated nature of the theory, PDFs can not be directly computed from first principles. Therefore, in addition to theoretical calculations, different experimental data need to be used to determine these distributions. Typically PDFs are determined in a global analysis where they are parameterized at an initial energy scale and the parameters are fixed by comparing to a wide range of experimental data at different energies. This procedure (cf. figure 1.6) requires computation of the theoretical predictions and iterative adaptation of the free parameters. That means, the comparison procedure is a multidimensional minimization problem. During a global fit the iterative procedure has to be repeated thousands of times until the optimal set of parameters is found. After the optimal set of parameters is found, an error analysis is performed to study the constraining power of the applied data.

The method to calculate theoretical predictions is based on perturbative expansion of QCD (pQCD) [41]. Given a set of parameters at the initial scale pQCD allows to calculate the PDFs at any higher scale with so called DGLAP equations. Furthermore, pQCD can be applied to calculate probabilities for the partonic scattering processes. This framework relies on the perturbative expansion in the strong coupling constant and thus the accuracy depends on how many terms (orders) are included in the calculations. The calculations become quickly complicated when higher order corrections are included. The first nPDFs were determined at leading order (LO) and the most established state-of-the-art fits are performed at next-to-leading order (NLO). The main goal of this project is to determine a new set of nuclear PDFs at next-to-next-to-leading order (NNLO) and to make it publicly available for the scientific community. The increase in perturbative precision in the nPDF determination enables to apply the more accurate calculations for the partonic scattering processes also in nuclear collisions in a consistent manner.

For an accurate nuclear PDF fit it is important to use experimental data from different kind of experimental setups in order to have constraints for different parton flavours, to achieve a wide coverage of the kinematic range, but also to prove the validity of the factorization theorem and to serve the universality of PDFs, i.e. their applicability for different processes. Experimentally and theoretically the cleanest process to study the nuclear modifications of PDFs is the deeply inelastic scattering (DIS) process in lepton-nucleus collisions. Experimentally, the measurement can be done accurately since only one incoming hadron is involved and the required kinematic variables can be derived by measuring the scattering angle and the energy of the outgoing lepton, without taking care about the outgoing hadrons produced in the scattering process. For the same reason, namely that only one incoming hadron is involved in the DIS process, the theoretical cross section is directly proportional to a single PDF, and *not* to a combination of two PDFs like for example in proton-proton collisions. In this work data from charged lepton-nucleus DIS and neutrino-nucleus DIS processes have been applied successfully in the performed nPDF analysis and the obtained results have been published [1, 2].

The impact of the experimental uncertainties on the nPDF precision can be estimated

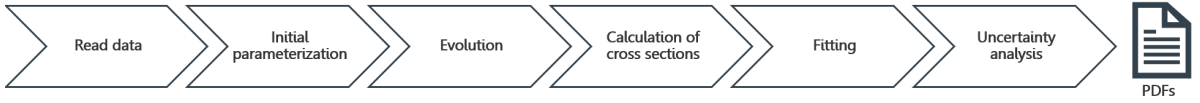


Figure 1.6.: High-level procedure of the nPDF analysis. The single steps will be used to visualize the relevance of the particular chapter in regard to the overall fitting procedure through the manuscript.

by performing an error analysis. At first, Hessian uncertainty analysis with underlying quadratic approximation has been performed. Investigations to perform a less-biased Monte Carlo error analysis are also presented in this thesis. However, the released set of nuclear parton distribution functions TUJU19 in the standard LHAPDF format [36] is based on Hessian uncertainty analysis.

1.4. Structure of the Thesis

The thesis is organized as follows. In the first chapter, after a short motivation and presenting the aim of this work, the reader is briefly introduced to the nuclear particle physics and the global nPDF analysis process. The relevance of the following chapters in regard to the overall PDF analysis is indicated by highlighting the particular step in figure 1.6 throughout the document. Thus, also the structure of the manuscript is described here for overview purposes and to guide the reader.

In preparation to the actual consideration of the performed global analysis the basics of QCD, i.e. the perturbative QCD, are introduced in chapter 2 since the theoretical methods for a global analysis are based on QCD.

Next, the relevant theoretical principles which are fundamental for the calculation of cross sections, but also later for the scale evolution of PDFs, are described in chapter 3. Also, important theoretical concepts, like collinear factorization theorem, universality and the concept of parton distribution functions (PDFs) are explained there. The energy scale evolution by the DGLAP equations is described in chapter 4.

After that, the theoretical and the numerical framework is presented in chapter 5. First, the selected form of the initial parameterization, the applied approximations and con-

strains are described. Also the definition of the quality of a global fit is provided there. Second, an overview of the applied numerical analysis framework and the necessary modifications to accommodate a nuclear PDF analysis are given. Last, a short summary of a comparison between the free proton PDF analysis and the nuclear PDF analysis is provided in the last section of chapter 5.

According to the title of this work, studies of parton distributions in nuclei are being performed. First, comparison of the computed theoretical predictions to the incorporated experimental data is discussed in chapter 6. Also the isoscalar modifications applied on the experimental data by the different collaborations are summarized there. Furthermore, the results of a study specific to the neutrino-nucleus scattering experiments are evaluated in chapter 6. Next, two methods which can be used for the analysis of experimental uncertainties in a global nPDF fit are described and the results of the comparison are presented in chapter 7. After that, the obtained parton distribution functions for the nuclei, but also for the free proton baseline, are presented in chapter 8. Naturally, a comparison to the existing nPDF sets is provided there as well. Finally, the application of the obtained PDF sets and the implemented framework on further collision systems and observables is studied in chapter 9.

All abbreviations and acronyms used in this thesis are summarized in the "List of abbreviations" at the beginning of the manuscript. At the end, an appendix is attached to provide more detailed information on the executed calculations and further reading for the performed studies. They are referenced in the related passage in the text in chapters 5-8. A summary of this thesis and an outlook on further investigations can be found in chapter 10 in English, and in chapter 11 in German.

2. QCD in a Nutshell

The Standard Model of particle physics accounts for three interactions: the electromagnetic, the weak and the strong interaction. The fourth fundamental interaction, gravity, is described by the theory of general relativity (GR) and is not part of the Standard Model, as no renormalizable quantum field theoretical description has been found so far. The electroweak, i.e. electromagnetic and weak, interactions are covered by the electroweak theory of which the electromagnetic part is contained within the field theory called Quantum Electrodynamics (QED). Lastly, the strong interaction of coloured quarks and gluons is described by the gauge field theory called Quantum Chromodynamics (QCD). The colour property of quarks and gluons is specific to QCD and leads to the phenomenon called **confinement**. Due to confinement freely propagating quarks are never observed but only bound objects of multiple quarks, called hadrons, which are necessarily white in regard to the colour charge. This is why the parton distribution functions need to be determined as part of a global analysis where hadrons (and not single partons) are observed by the experiment. As for the global analysis, the theoretical principles and the methods for the calculation of cross sections are based on QCD. Therefore, the most important concepts of QCD, such as QCD Lagrangian, perturbative QCD, regularisation, renormalisation and asymptotic freedom, are introductorily outlined in this chapter. For more details the reader is referred to the standard text books, e.g., [41–43].

2.1. Lagrangian

As per the Standard Model and QCD the hadronic part of all visible matter is made of quarks and gluons. The quarks q are fermionic particles with spin quantum number $1/2$, carrying a fractional electric charge of $-1/3$ or $+2/3$ and a colour charge a (usually called red, green or blue). The force between quarks is mediated by gluons g ,

which are massless bosonic particles with spin 1. However, colour confinement implies that freely propagating quarks or gluons are never observed but only colour-neutral (or white) objects can exist in nature. This colour-neutral state is referred to as a colour singlet state of the $SU(3)$ group. The antiquarks \bar{q} carry the opposite electric charge as the quarks and the anti-colours (anti-red, anti-green, anti-blue). The possible hadronic colour singlet states are mesons $q_a\bar{q}^a$ and baryons $\epsilon^{abc}q_aq_bq_c$ with the totally asymmetric Levi-Civita tensor ϵ^{abc} . Other states such as tetra- or pentaquarks [44–47], consisting of four or more quarks, are typically considered as exotic hadrons. The evidence for the existence of partons was demonstrated by the deeply inelastic scattering experiments in the late 1960s [48–50]. First the constituents inside hadrons were assumed to be two or three approximately point-like quarks. However, the experimentally measured shape of the observable (here: structure function F_2) indicated a more complicated substructure, as described in the next chapter.

The two experimentally confirmed assumptions - that all hadrons consist of quarks and gluons and that partons cannot be observed as free particles - are reflected in the form of the QCD Lagrangian, which is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^A F^{A\mu\nu} + \sum_{\text{flavours } f} \bar{q}_a^f(x) [i\gamma^\mu D_\mu - m_f]_{ab} q_b^f(x) \quad (2.1)$$

where γ^μ are Dirac matrices and m_f is the mass of a quark of flavour f . The colour index a runs from $a = 1$ to 3, since there are $N_C = 3$ quark colours. The covariant derivative D_μ in equation (2.1) is given by

$$(D_\mu)_{ab} = \partial_\mu \delta_{ab} - ig_S (t^C \mathcal{A}_\mu^C)_{ab}, \quad (2.2)$$

where g_S is the coupling constant of the strong interaction and $t^C = \lambda^C/2$ are the Hermitian and traceless generators of $SU(3)$ represented by the eight Gell-Mann matrices λ^C , with $A, B, C = 1, \dots, 8$ according to the eight colour degrees of freedom of the gluon field \mathcal{A}_μ^C . Finally, the non-Abelian gluon field strength tensor $F_{\mu\nu}^A$ is defined by

$$F_{\mu\nu}^A = \partial_\mu \mathcal{A}_\nu^A - \partial_\nu \mathcal{A}_\mu^A + g_S f^{ABC} \mathcal{A}_\mu^B \mathcal{A}_\nu^C \quad (2.3)$$

with the totally antisymmetric structure constants f^{ABC} of the $SU(3)$ colour group, and

the color indices $A, B, C = 1, \dots, 8$ as introduced above^[1].

The form of the QCD Lagrangian in equation (2.1) was proposed in several works by FRITSCH, GELL-MANN, LEUTWYLER, GROSS, WILCZEK, and WEINBERG in 1973 [29–32]. The Lagrangian satisfies the gauge symmetry invariance required to protect the zero mass of the gluon from higher-order corrections [43]. Furthermore, as mentioned above, the field strength tensor is non-Abelian which is due to the last term on the right-hand side in equation (2.3). This is an important property of QCD, which distinguishes it from the Abelian field theories like e.g., QED, that causes gluon self-interactions and leads to the property of asymptotic freedom described in the following section. As scattering processes within QCD are not exactly solvable in an analytical calculation, methods of perturbative quantum chromodynamics (pQCD) need to be applied as outlined in the next section.

In order to properly define the building blocks of the perturbation theory such as a propagator, which describes the propagation of physical particles, a so-called gauge-fixing term

$$\mathcal{L}_{gauge-fixing} = -\frac{1}{2\lambda} (\partial^\mu \mathcal{A}_\mu^A)^2, \quad (2.4)$$

with gauge parameter λ , and the supplemented ghost term

$$\mathcal{L}_{ghost} = \partial_\mu \eta^{A\dagger} (D_{AB}^\mu \eta^B) \quad (2.5)$$

with the complex scalar fields $\eta^{A,B}$, which obey Fermi statistics, must be added to the QCD Lagrangian. The term in equation (2.4) is needed to define the propagator for the gluon field which requires a choice of gauge [23, 43]. The ghost fields introduced in equation (2.5) have the functionality to cancel the unphysical degrees of freedom of the gluon which would appear otherwise. The Feynman rules for the pQCD Lagrangian and the corresponding Feynman diagrams to visualise the process at the particular perturbative order are provided for reader's reference in [43, 51].

[1] The nomenclature used in this work implies the choice of natural units $\hbar = c = 1$.

2.2. Perturbative QCD

Processes described by the QCD Lagrangian (2.1) from the previous section can not be solved exactly, but only approximately with the methods of perturbation theory. A perturbation theory provides mathematical methods for an approximate solution of the problem. For that matter the exact expression is expanded in terms of a power series in a necessarily small parameter. The summands of the power series are referred to as order in regard to the small parameter. The first term is called leading-order (LO), the next one is next-to-leading order (NLO) etc, where the contribution to the overall result by the higher-order terms decreases continually. Though, the accuracy of the obtained results is supposed to increase with increasing perturbation order. In case of pQCD the perturbation order is defined by the order of the coupling constant α_S in the expanded series. Current state-of-the-art calculations are usually at least of NLO.

The perturbative calculus leads to divergent integrals appearing at higher orders. However, the physical observables must be finite. For that purpose, two principles - *regularisation* and *renormalisation* - are described further in this section. These two principles are required to cover different classes of divergences, so-called ultraviolet (UV) divergences, and infrared (IR) and collinear divergences.

- **Collinear divergencies** are generated by the collinear emission of particles. This type of divergences is sometimes also referred to as mass singularities as they occur when the momenta of two massless particles become parallel. In this case the propagator of a massless particle (gluon or quark considered as massless) gives rise to a singularity in the according integral.
- **Infrared divergences** are generated when so-called soft particles, characterised by the small momentum $p^\mu \rightarrow 0$, are emitted from the parent parton.
- **Ultraviolet divergences** appear in the so-called loop-integrals, e.g. coming from the virtual correction diagrams, at large momenta $p^\mu \rightarrow \infty$. They are taken into account by the renormalisation of the strong coupling constant α_S .

First we discuss the treatment of the divergences by the regularisation. **Regularisation** is the mechanism to regularize integrals which otherwise would be divergent. The choice of the regularisation method provides the instructions how to treat the affected integrals

in order to solve them analytically. One possible choice is dimensional regularisation, where the dimension of the actually 4-dimensional momentum integral is substituted by

$$\int \frac{d^4q}{(2\pi)^4} \rightarrow \mu^{2\epsilon} \int \frac{d^{4-2\epsilon}q}{(2\pi)^{4-2\epsilon}}, \quad (2.6)$$

with the new dimension $D \equiv 4 - 2\epsilon$ and term $\mu^{2\epsilon}$ to account for the proper mass dimension. In anticipation of the discussion of the renormalisation we mention here that μ is the so-called renormalisation scale that is introduced to retain the strong coupling constant dimensionless in any dimension, which is required for a renormalizable field theory like QCD. In regard to the regularisation procedure, the transition from 4 to D dimensions renders the integral finite for some $\epsilon \neq 0$ so that it can be calculated analytically. The original divergences of the integral will show up as terms proportional to $1/\epsilon$. Once the calculation of the observable is completed, the physical quantity may not depend on the regulator (here ϵ) and therefore $\epsilon \rightarrow 0$ needs to be applied. Due to the poles at $\epsilon = 0$ this limit cannot be taken right away. The removal of the UV divergences to obtain finite physical quantities is done through the techniques of renormalisation, as outlined in the next paragraph.

The QCD, and also QED, are renormalizable theories since the respective coupling constants have a mass dimension zero, which is a prerequisite for the renormalisation. The discussion of the renormalisation naturally leads to the principle of *asymptotic freedom*, described in the next paragraph, which is an important property of QCD providing a mandatory condition for perturbative QCD to be applicable in high-energy processes.

Renormalisation is required to ensure that calculated physical quantities are finite. For that purpose the so-called "bare" quantities in the Lagrangian, which can be infinite, e.g., fields ψ_0 , are replaced by their renormalised substitutes ψ_R , which are obligatorily finite, together with a renormalisation constant Z_ψ that absorbs the infinite terms of the bare quantity:

$$\psi_0 = \sqrt{Z_\psi} \psi_R. \quad (2.7)$$

Equation (2.7) represents the renormalisation of a field. Also physical parameters (e.g. the electric charge and the electron mass in QED) and gauge parameters have to be renormalised. In general, the normalisation constant Z_ψ can be expanded perturbatively

in form of a Taylor series (implicitly in α_S)

$$\sqrt{Z_\psi} \approx 1 + \frac{1}{2}\delta Z_\psi + \mathcal{O}(\alpha_S^2). \quad (2.8)$$

The renormalisation procedure introduces so-called counter-terms leading to additional Feynman diagrams which need to be taken into account during the computation. The renormalisation condition demands that the sum of the originally infinite diagram together with the counter-terms is finite. This condition allows to define the renormalisation constant at an arbitrary order. However, there is some freedom how to define the finite terms of the renormalisation constant. The corresponding choice is described by the renormalisation scheme. Commonly in QCD calculations, the so-called modified minimal subtraction ($\overline{\text{MS}}$) scheme is used. Contrary to other schemes such as the on-shell scheme, which requires experimental measurement for the masses and couplings, the $\overline{\text{MS}}$ scheme at a fixed order introduces an unphysical renormalisation scale μ . However, taking into account all orders in perturbation theory, any physical observable o_R must be independent of the renormalisation scale μ :

$$\mu^2 \frac{d}{d\mu^2} o_R(Q^2/\mu^2, \alpha_S) \equiv \left[\mu^2 \frac{\partial}{\partial \mu^2} + \mu^2 \frac{\partial \alpha_S}{\partial \mu^2} \frac{\partial}{\partial \alpha_S} \right] o_R = 0, \quad (2.9)$$

with a typical energy scale Q of the process. This physical requirement, specifically by defining the beta function

$$\mu^2 \frac{\partial \alpha_S}{\partial \mu^2} \equiv \beta(\alpha_S) \quad (2.10)$$

currently known up to four-loop ($\mathcal{O}(\alpha_S^5)$) [52], leads to the phenomenon referred to as the running coupling constant $\alpha_S(\mu^2)$. The functional form of $\alpha_S(\mu^2)$ usually provided in the literature (cf. figure 2.1) is specific to the $\overline{\text{MS}}$ scheme.

As can be seen from figure 2.1 the strong coupling constant $\alpha_S(Q^2)$ is not constantly strong, but decreases with increasing energy Q . This behaviour allows the application of pQCD methods for sufficiently large energies and is referred to as **asymptotic freedom**. As per definition [23, 43], asymptotic freedom states that the otherwise strong coupling of quarks and gluons is small at short distances so that the partons behave as quasi-free particles at asymptotic, i.e. high, energies.

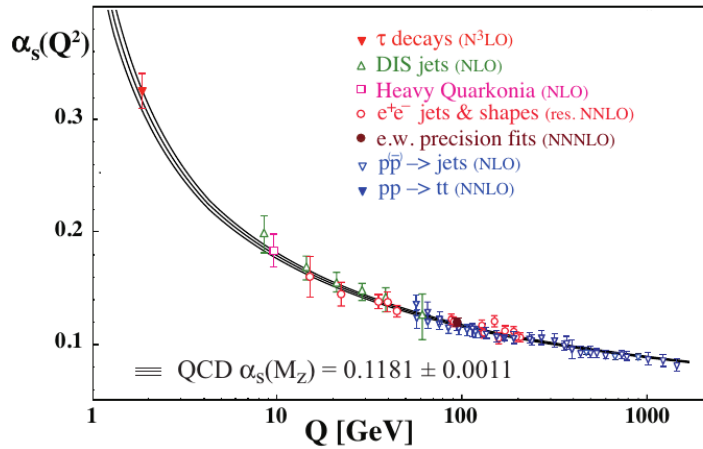


Figure 2.1.: The running coupling constant $\alpha_S(Q^2)$ as a function of the energy scale Q (here $\mu^2 = Q^2$). The function values of $\alpha_S(Q^2)$ at certain scale Q are fixed by the measurements, summarized here. Figure credited from [11].

In the global nPDF analysis performed as part of this work the perturbative QCD methods are used. An alternative approach to perform QCD calculations uses the numerical principles of Lattice QCD, which is mentioned here for the reason of completeness. The interested reader is referred to [53] for further details.

3. Theoretical Principles

The calculation of partonic cross sections at NLO and NNLO is one of the key steps of the global nPDF analysis performed as part of this work. The relevance in regard to the overall analysis process is shown in figure 3.1 for a better overview. Section 3.2 of this chapter covers the calculation of cross sections for the deeply inelastic scattering (DIS) process considered in the performed analysis. Furthermore, the structure of the nucleon, as well as the important theoretical principles for a global PDF analysis, such as parton model, factorization theorem, and the parton distribution functions, are also described in this chapter.



Figure 3.1.: Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: calculation of cross sections with the underlying theoretical principles are described.

3.1. Nucleon Structure and Nuclear Effects

As introduced in chapter 1, the purpose of PDFs is to describe the collinear momentum distribution of the partons (quarks and gluons) inside a nucleon. A nucleon can be either a proton or a neutron. Both particles have very similar masses: in natural units the proton mass is $m_p \approx 0.938 \text{ GeV}$ and neutron mass is $m_n \approx 0.939 \text{ GeV}$ [11]. Due to that fact, in nuclear physics proton and neutron are considered as two manifestations of the same particle called nucleon [42]. In order to describe such a system, a mathematical structure named **isospin** has been introduced, defined in a similar way as the spin quantum number of an electron also being a two-states system. However, the nuclear force is supposed to be invariant under isospin transformations [42]. And also we are using

the isospin symmetry to derive the nuclear PDFs in this analysis, as will be discussed in section 5.1.

The inner structure of the nucleons is described by the dimensionless **structure functions** $F_1(x, Q^2)$, $F_2(x, Q^2)$ and $F_3(x, Q^2)$ that depend on the Bjorken- x and the energy scale used to probe the nucleon Q^2 . The structure functions, that are needed to calculate the differential cross sections of lepton-proton scattering processes analysed in this work, and the variables will be formally defined in section 3.2. Here we anticipate the discussion in order to introduce the basic idea. Sometimes the structure function $F_1(x, Q^2)$ is also called the magnetic form factor and $F_2(x, Q^2)$ is referred to as the electric form factor. In the simple parton model, as described in section 3.3, they satisfy the Callan-Gross relation [54]

$$F_2(x) = 2xF_1(x) \quad (3.1)$$

that is valid for fermions, e.g. quarks, the spin-1/2 constituents of the nucleon. In the simple parton model (cf. sec. 3.3) the electromagnetic structure functions in terms of the quark and antiquark distributions $q_f(x), \bar{q}_f(x)$ of flavour f , which will be formally defined in chapter 3.4, are

$$F_2(x) = 2xF_1(x) = \sum_f e_f^2 [xq_f(x) + x\bar{q}_f(x)] \quad (3.2)$$

with e_f the electric charge of the quark of flavour f . However, for the considerations in this PDF analysis, the structure functions of the QCD-improved parton model are used (cf. sec. 3.2). For example, for the so-called neutral current (NC) lepton-proton scattering processes, with γ and/or Z -boson exchange, the structure functions at LO are

$$F_2^{\text{NC}}(x, Q^2) = 2xF_1^{\text{NC}}(x, Q^2) = \sum_f A_f(Q^2) [xq_f(x) + x\bar{q}_f(x)] , \quad (3.3)$$

$$F_3^{\text{NC}}(x, Q^2) = \sum_f B_f(Q^2) [q_f(x) - \bar{q}_f(x)] , \quad (3.4)$$

where the factors $A_f(Q^2)$ and $B_f(Q^2)$ in equations (3.3), (3.4) are the quark couplings to γ and Z boson as given in reference [55]. The contribution to the lepton-proton scattering cross section by the parity-violating structure function $F_3(x, Q^2)$ becomes relevant at very high energies in NC processes, where Z boson exchange mechanism needs to be

taken into account, and in the charged current (CC) reactions. The expressions given in this section are meant for demonstrative purposes only. For a complete set of structure functions for neutral current and charged current lepton-proton scattering please refer to references [43, 55, 56].

In regard to the nucleon structure and the analysis performed in this work, we know that due to confinement free propagating single quarks are never observed, but only bound objects of multiple quarks (e.g. proton). So, before clarifying the details of the momentum distribution of the partons inside a hadron, one should recall how we know there are partons inside a nucleon and why we can assume a non-trivial momentum distribution.

If nucleons would have no substructure and would be point-like particles themselves,

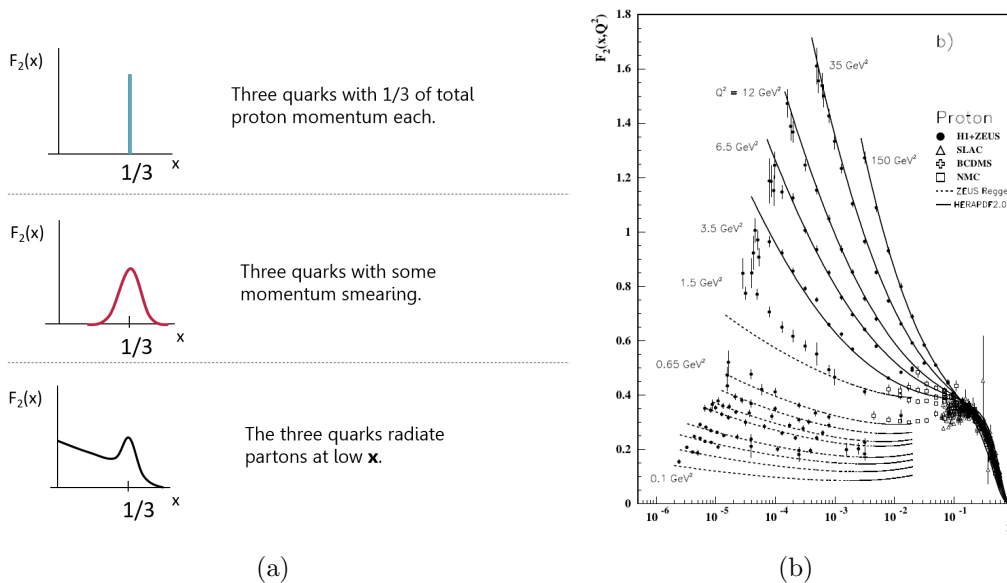


Figure 3.2.: a) The expected functional form of structure function $F_2(x)$ along momentum fraction x dependent on the inner structure of a proton. For instance, in case a proton consists of three valence quarks only each of them carries a momentum fraction $1/3$ and therefore one would expect the contribution for $F_2(x) \neq 0$ only at $x = 1/3$. b) The actual functional form of structure function $F_2(x)$ as measured by different experiments (same as figure 3.5b, credited from [11]). The measurement corresponds to the lowest panel in a) and therefore shows that the three valence quarks must radiate partons at low x .

one would expect a δ -like structure function $F_2(x)$, i.e. with discrete function values as shown in the upper panel in figure 3.2(a). In reality the experimental results from DIS experiments show a different form of $F_2(x)$ (cf. fig. 3.2(b)) clearly indicating the existence of a substructure in a hadron. However, for the momentum distribution of the partons there are several possible options which would satisfy confinement, as shown in figure 3.2(a). Fortunately, the experimental measurements provide a more detailed picture by favouring the form of theoretical prediction which corresponds to the situation where three valence quarks radiate partons at low x . This knowledge paves the way for the following studies of the collinear momentum distributions of the partons inside bound nucleons.

The need to derive dedicated PDFs for nucleons bound in nuclei results from the fact that **nuclear effects**, such as shadowing [57, 58], anti-shadowing [59] and EMC-effect [60], are observed when the cross sections are measured with nuclear targets, as shown in figure 3.3. Nuclear effects cause a modification of quark and gluon distributions in bound nucleons and are described in the following.

Shadowing is the effect in the small x region $x \lesssim 0.1$ implying a reduction of the measured lepton-nucleus cross section (or structure function) compared to the expected value based on an observable constructed from free nucleon PDFs, or $\sigma(A_1)/\sigma(A_2) < 1$ for the ratio of observables measured for nucleons in two different nuclei with mass numbers A_1, A_2 as shown in figure 3.3. The existing explanations for the shadowing are based on the idea that the inner nucleons of nuclei are being shielded by the surface nucleons. In particular, the photon could interact with different partons in several nucleons of the nucleus experiencing multiple scattering. An overview of models to explain shadowing based on multiple scattering is provided in reference [57]. Some representative models are listed here:

- Glauber-like rescattering [61] suggests that a virtual photon has a hadronic component with a partonic structure like in the dipole model [62]. The Glauber-Gribov theory [63, 64] describes the multiple scattering of these hadronic components of the photon having a fixed size, e.g. $\bar{q}q$ -pairs, with the nucleons inside a nucleus by neglecting the binding energy.
- Gribov inelastic scattering is based on the relativistic Gribov theory [63, 64], where

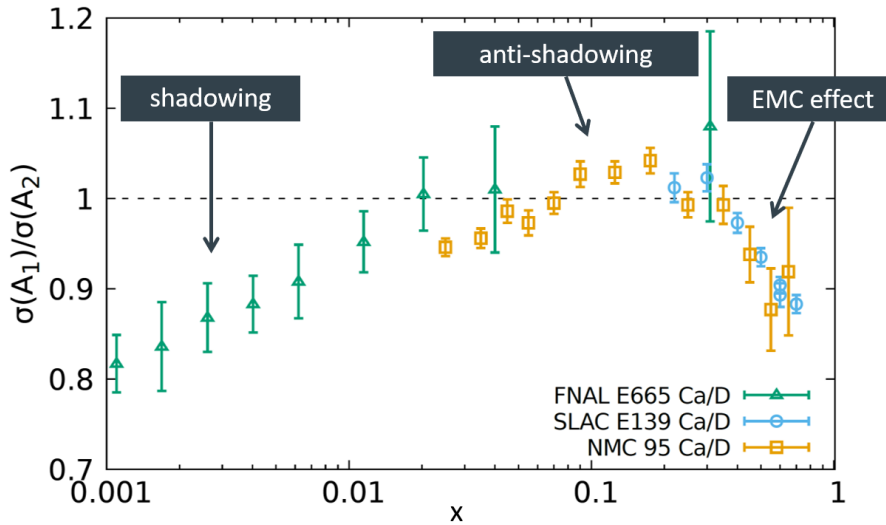


Figure 3.3.: Illustration of sample data points measured by the experimental collaborations FNAL [4], SLAC [5] and NMC [6]. The functional form of the ratio $\sigma(A_1)/\sigma(A_2)$ along momentum fraction x for nucleons in two different nuclei with mass numbers A_1, A_2 is characterized by the nuclear effects named shadowing, anti-shadowing and EMC effect. The dashed line symbolizes the ratio $\sigma(A_1)/\sigma(A_2) \approx 1$ which one would expect without nuclear effects. The observed deviation of $\sigma(A_1)/\sigma(A_2)$ from ≈ 1 due to the nuclear effects causes the need for nuclear parton distribution functions.

subsequent interactions of the projectile with the nucleons in the nucleus are suppressed at high energies. The collision proceeds through simultaneous interactions, where the intermediate states are not the same as the initial state, which is called inelastic.

- Vector meson dominance is a model that proposes a hadron-like behaviour of the photon [65–69]. There the hadronic component of the virtual photon is modeled as a superposition of hadronic states with photon quantum numbers. The idea was developed based on the picture that a nucleon is surrounded by a pion cloud and the measurements of the form-factors were interpreted as the evidence for an isoscalar vector meson. Subsequently, the photon-hadron interactions were assumed to be described by this picture.
- Parton recombination is a model that suggests that the shadowing in the small x region occurs due to the gluon recombination into gluons or into quarks [70, 71].

As can be seen in figure 3.3, along the x -axis the nuclear shadowing is followed up by the **anti-shadowing**, which is often considered as a consequence of the shadowing effect. For example, when using the explanation of multiple scattering, shadowing and anti-shadowing are supposed to arise due to destructive or respectively constructive interference of amplitudes resulting from the multiple scattering of quarks in the nucleus [59]. Another explanation is provided by the fact that anti-shadowing results from the application of momentum and baryon number sum rules incorporated into the global analysis (cf. sec. 5.1.2).

Probably the most intriguing effect for the high-energy particle physics community is the **EMC effect** [72] observed in the x region $0.3 \lesssim x \lesssim 0.8$ measured for the first time by the EMC Collaboration [73, 74]. Several explanations exist for the EMC region, for example short range correlations, nuclear binding, pion exchange, change in the nucleon radius, etc. For an overview of the present status of the EMC effect please refer e.g. to [75]. Some representative models are listed here:

- Nucleon-nucleon short range correlations (SRC) [76–82], that appear for pairs of nucleons close to each other, generate attractive or repulsive potentials leading to nucleons with high momenta. Studies [83, 84] showed that the nuclear structure functions for different nuclei become universal in the EMC region if they are rescaled by the number of SRC pairs. This explanation is especially favoured by the research efforts at Jefferson Lab [85–88].
- Nuclear binding [89–96] is the explanation which considers the so-called binding correction that is the difference between the calculated results with bound off-mass-shell (virtual) nucleons and with free on-mass-shell (real) nucleons [90]. In [89] the authors claim that the nuclear binding effect causes a violation of the sum rules.
- Pion based approaches [97–102] suggest that pionic contributions to the nuclear structure function need to be taken into account by considering pions as effective nuclear constituents.
- Rescaling models [103, 104] propose that by changing, i.e. rescaling, either the Q^2 or the x scale for the nuclear structure function the nuclear observable becomes comparable to the one of a free nucleon. In this scenario the EMC effect is supposed to be related to a change of confinement size inside the nucleus.

In addition to the nuclear effects shown in figure 3.3, in literature one finds the Fermi motion region for $x \gtrsim 0.8$. The effect leading to $\sigma(A_1)/\sigma(A_2) > 1$ is supposed to be caused by the Fermi motion of the nucleons in the nucleus.

Of course one expects the strong, non-perturbative nuclear binding to have an effect on the parton substructure of bound nucleons. Nonetheless, due to the validity of the factorization theorem, which is supposed to be verified in a global QCD analysis to an extended degree, the non-perturbative effects are absorbed by the nPDFs and the partons at high Q^2 satisfy the same QCD dynamics as their counterparts in a free proton. In the global nPDF analysis performed in this work we do not use any of the models described above to moderate or rescale the nuclear effects in the particular x -region. The modifications of the structure functions by the nuclear effects are absorbed by the non-perturbative nPDFs, which distinguishes the nuclear PDFs from the free proton PDFs. However, it is important to understand the nuclear effects and thus to improve the precision level of the nPDF analysis (here: NNLO) in order to provide a good basis of comparison to verify the particular models with less uncertainties.

3.2. Deeply Inelastic Scattering

One part of the global PDF analysis is the calculation of cross sections, which are later compared to the experimental data in order to determine the optimal free parameters. The calculation of cross sections is specific to the considered process. In this work the relevant process is the deeply inelastic scattering (DIS) that is described in this section based on references [11, 43, 51, 55, 56], and some paragraphs are credited from the own publication [1]. Deeply inelastic scattering is a scattering process between a high-energy lepton and a hadron - here a nucleon. A Feynman diagram for charged lepton-nucleon DIS is shown in Figure 3.4.

First, one needs to introduce the standard **DIS variables**. As per the diagram in figure 3.4 the incoming lepton four-momentum is labeled k^μ and the outgoing one is k'^μ . The momentum transfer is given by $q^\mu = k^\mu - k'^\mu$ and the momentum of the target

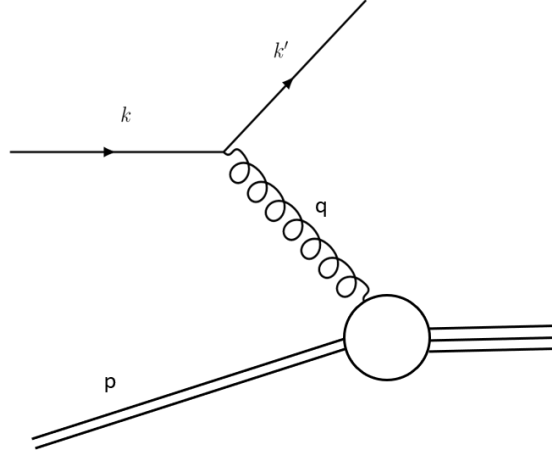


Figure 3.4.: Feynman diagram for deeply inelastic charged lepton-nucleon scattering at leading order.

hadron is p^μ . In the target rest frame we have

$$\begin{aligned}
 Q^2 &\equiv -q^2 = 4EE' \sin^2(\theta/2) && \text{(Virtuality),} \\
 x &\equiv \frac{Q^2}{2q \cdot p} = \frac{Q^2}{2M\nu} && \text{(Bjorken } x), \\
 y &\equiv \frac{q \cdot p}{k \cdot p} = 1 - E'/E && \text{(Inelasticity),}
 \end{aligned} \tag{3.5}$$

with the energy of the incoming or outgoing lepton E or E' respectively, and the proton (hadron) mass M , as well as

$$\begin{aligned}
 M^2 &= p^2 && \text{and,} \\
 \nu &= \frac{q \cdot p}{M} = (E' - E) .
 \end{aligned} \tag{3.6}$$

All three DIS variables in equation (3.5) are necessary for the calculation of DIS cross sections, as described in the following. Additionally, Q^2 and x will be important when presenting and discussing the results. Based on the invariant variables introduced in relations (3.5) and (3.6) one defines the invariant mass W of the final hadronic state

$$W^2 \equiv (p + q)^2 = M^2 + Q^2 \frac{1 - x}{x} \approx Q^2 \left(\frac{1}{x} - 1 \right) . \tag{3.7}$$

The factorization of the partonic scattering process and the non-perturbative PDFs [24],

described in section 3.4, as well as the perturbative QCD framework are valid only at sufficiently high energy scales. Therefore, we had to exclude some data points from the analysis by the application of so-called kinematic cuts. In this work we have selected the kinematic cuts $Q^2 > 3.5 \text{ GeV}^2$, $x < 0.7$, and $W^2 > 12 \text{ GeV}^2$.

Next, to calculate the DIS cross section required for the fitting procedure during a global PDF analysis, we write the differential, spin-independent (averaged) cross-section as

$$d\sigma = \frac{4\pi M e^4}{2s} \frac{d^3\mathbf{k}'}{q^4 (2\pi)^3 2E'} L^{\mu\nu} W_{\mu\nu} \quad (3.8)$$

where e is the electric coupling constant, and $L^{\mu\nu}$, $W_{\mu\nu}$ are the leptonic and hadronic tensors. The leptonic tensor is defined as

$$L^{\mu\nu} \equiv \frac{1}{2} \text{Tr} [\not{k}' \gamma^\mu \not{k} \gamma^\nu] = 2 \left[k'^\mu k^\nu + k^\mu k'^\nu - (k \cdot k') g_{\mu\nu} \right], \quad (3.9)$$

and the hadronic tensor is given by

$$W^{\mu\nu} \equiv \frac{1}{4\pi M} \frac{1}{2} \sum_{\sigma=\pm 1} \sum_X \langle p, \sigma | J^\mu(0) | X \rangle \langle X | J^\nu(0) | p, \sigma \rangle \cdot (2\pi)^4 \delta^4(p + q + p_X) \quad (3.10)$$

averaging over spin σ and summing over all final hadronic states X , with the quark electromagnetic current J^μ . Consideration of all final hadronic states is called inclusive DIS. The hadronic tensor $W^{\mu\nu}$ can not be computed perturbatively due to the non-perturbative nature of QCD. However, since the leptonic tensor $L^{\mu\nu}$ is real and symmetric under interchange of indices, it provides some information on the hadronic tensor which should satisfy $W_{\mu\nu} = W_{\mu\nu}^* = W_{\nu\mu}$ taking into account the parity conservation. The first equality sign is postulated since the overall cross section is a real quantity. Additionally, due to the conservation of the electromagnetic current a further restriction is provided by the Ward identity $q^\mu W_{\mu\nu} = q^\nu W_{\mu\nu} = 0$, leading to the form

$$W_{\mu\nu} = -W_1 \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2} \right) + \frac{W_2}{M^2} \left(p_\mu - \frac{p \cdot q}{q^2} q_\mu \right) \left(p_\nu - \frac{p \cdot q}{q^2} q_\nu \right) + \frac{W_3}{2M^2} i p^\gamma q^\delta \epsilon_{\mu\nu\gamma\delta}, \quad (3.11)$$

with the totally asymmetric Levi-Civita tensor $\epsilon_{\mu\nu\gamma\delta}$, and where W_i are unknown, scalar functions of x and Q^2 , called **structure functions** characterizing the structure of the hadron. Terms which vanish in the contraction with the leptonic tensor have been omitted in equation (3.11). After contracting the hadronic tensor with the leptonic tensor (3.9) one can write the differential cross section in terms of incoming and outgoing energy E, E' and the scattering angle θ in the target rest frame [56]

$$\begin{aligned} \frac{d^2\sigma}{dx dy} = N^l & \left[W_1(x, Q^2) \sin^2(\theta/2) + W_2(x, Q^2) \cos^2(\theta/2) \right. \\ & \left. \pm W_3(x, Q^2) \frac{E + E'}{M} \sin^2(\theta/2) \right] \end{aligned} \quad (3.12)$$

with the normalization factor N^l as per equation (3.16) or (3.17) respectively. In the following description, and as usually used in the literature, e.g. [56], one defines:

$$\begin{aligned} F_1(x, Q^2) &\equiv W_1, \\ F_2(x, Q^2) &\equiv \frac{\nu}{M} W_2 \\ F_3(x, Q^2) &\equiv \frac{\nu}{M} W_3, \end{aligned} \quad (3.13)$$

so that the differential cross section for deeply inelastic scattering can be expressed in terms of Lorentz invariant variables from equation (3.5):

$$\frac{d^2\sigma}{dx dy} = N^l \left[y^2 x F_1^l(x, Q^2) + (1 - y) F_2^l(x, Q^2) \mp \left(y - \frac{y^2}{2} \right) x F_3^l(x, Q^2) \right]. \quad (3.14)$$

As can be seen from equation (3.14), all kinematic variables defined in equation (3.5) are required to calculate the differential cross section. Nonetheless, some of the experimental data sets do not specify the invariant y , but when the center-of-mass energy \sqrt{s} is known, the invariants can be derived from the relation

$$Q^2 \approx y x s, \quad (3.15)$$

which is provided here for the reason of completeness.

In this analysis, neutral-current (NC) and charged-current (CC) DIS processes are con-

sidered, NC in case of electron(positron)-nucleus (eA) and CC for (anti)neutrino-nucleus (νA) scattering. Going back to equation (3.14), the coupling factor N^l depends on these scattering types and $F_{1,2,3}^l$ denote the structure functions [11, 55, 56] for scattering of lepton l . In equation (3.14), the index l covers different beams including $l = \nu, \bar{\nu}, e^+, e^-, \mu^+, \mu^-$. For the nuclear data used in this work two normalization factors are relevant – one for NC DIS in case of unpolarized leptons and one for CC DIS of incoming (anti-)neutrinos. In case of unpolarized leptons the normalization factor $N^{l,NC}$ for NC DIS is [11]

$$N^{l,NC} = \frac{4\pi\alpha^2}{xyQ^2}, \quad (3.16)$$

and in case of incoming (anti-)neutrinos the normalization factor $N^{\nu,CC}$ for CC DIS is

$$N^{\nu,CC} = \frac{G_F^2 M_W^4 Q^2}{4\pi xy (Q^2 + M_W^2)^2}, \quad (3.17)$$

where Q^2 is the virtuality of the intermediate boson that provides the scale at which the nucleons are probed. For the CC processes G_F is the Fermi coupling constant and M_W is the mass of the W^\pm bosons. When combining the structure functions into the differential cross sections in alignment with (3.14), the sign before F_3 is positive for ν and e^+ , and negative for $\bar{\nu}$ and e^- .

In QCD the structure functions F_i , as introduced in eq. (3.13), are related to the scale-dependent parton distribution functions $f_j(x, Q^2)$, with $j = g$ or $j = q, \bar{q}$, via

$$F_i(x, Q^2) = \sum_j C_i^j(x, \alpha_s(\mu^2), \mu^2/Q^2) \otimes f_j(x, \mu^2), \quad (3.18)$$

where one typically chooses $\mu = Q$. The symbol \otimes in equation (3.18) denotes a convolution between the parton distribution functions and the Wilson coefficients C_i^j (see e.g. refs. [105–107] for C_2^j at NLO and NNLO). For example, the structure functions for neutrinos and anti-neutrinos are given at leading order (LO) by [11]

$$\begin{aligned} F_1^\nu &= d + s + b + \bar{u} + \bar{c} + \bar{t}, \\ F_2^\nu &= 2x(d + s + b + \bar{u} + \bar{c} + \bar{t}), \\ F_3^\nu &= 2(d + s + b - \bar{u} - \bar{c} - \bar{t}), \end{aligned} \quad (3.19)$$

and

$$\begin{aligned}
 F_1^{\bar{\nu}} &= u + c + t + \bar{d} + \bar{s} + \bar{b}, \\
 F_2^{\bar{\nu}} &= 2x(u + c + t + \bar{d} + \bar{s} + \bar{b}), \\
 F_3^{\bar{\nu}} &= 2(u + c + t - \bar{d} - \bar{s} - \bar{b}).
 \end{aligned}
 \tag{3.20}$$

The expressions given in this section are meant for demonstrative purposes only. For a complete set of structure functions for neutral-current and charged-current lepton-proton scattering please refer to references [43, 55, 56].

3.3. Parton Model and Scaling Violations

As introduced in the previous section, the DIS process describes lepton-hadron scattering. A hadron, i.e. proton, is not elementary but a bound object consisting of partons - quarks and gluons. The fundamental assumption of the original parton model is that the interactions of hadrons are due to the interactions of partons [43]. The parton model [21, 22] is defined as

$$d\sigma = \sum_q \int_0^1 d\xi d\hat{\sigma}_0^q(\xi P) f_q(\xi)
 \tag{3.21}$$

with proton's momentum P , momentum fraction $0 \leq \xi \leq 1$ and parton's momentum $p = \xi P$. Further quantities in equation (3.21) are the leading order (Born) cross section for the lepton parton scattering $\hat{\sigma}_0^q(\xi P)$, and the so-called parton distribution functions $f_q(\xi)$ which represent the number density of partons of flavour q in the hadron.

The DIS variables in the previous section have been introduced in the target-rest frame, whereas one needs to mention that the parton model is motivated in the lepton-proton center-of-mass frame where the proton is ultrarelativistic. This frame is referred to as Bjorken frame or the infinite momentum frame (IMF). In such a frame, the photon scatters off a pointlike parton which is moving parallel (longitudinal) with the parent proton. In the parton model the partons inside of a proton are assumed to be non-interacting. This approximation is reasonable since in IMF the typical time scale of the parton interactions inside a hadron (proton, nucleon) is much longer than the time scale of DIS [51].

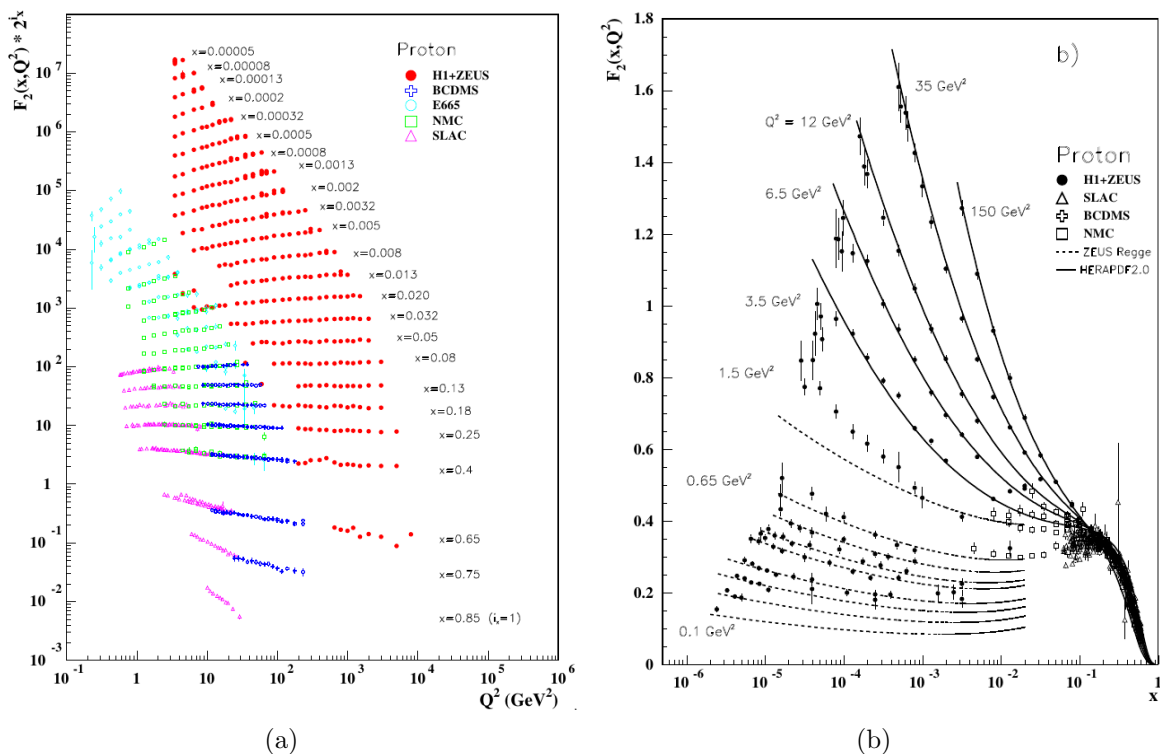


Figure 3.5.: Experimental data for the structure function $F_2(x, Q^2)$, measured by the collaborations HERA (H1+ZEUS), BCDMS, FNAL (E665) and NMC. The observed Q^2 -dependence of the structure function F_2 is called Bjorken-scale violation. Figure credited from [11]. a) Proton structure function $F_2(x, Q^2)$ along Q^2 -axis at different x values. The structure function for small and large x values is *not* constant along Q^2 , but shows a more or less linear dependence in $\log(Q^2)$. b) Proton structure function $F_2(x, Q^2)$ along x -axis at different Q^2 values. The scaling violations are particularly large at small x .

The originally introduced parton model predicts that the structure functions $F_{1,2}$ are only functions of x , and are not supposed to depend on Q^2 in the $Q^2 \gg M^2$ limit. More specifically, the structure functions should obey the so-called **Bjorken-scaling** law in the Bjorken limit defined as $Q^2, \nu \rightarrow \infty$ with x being fixed. This property is related to the assumption that parton's transverse momenta are small in IMF [11]. Contradictory to the naive parton model, the modern experimental data show a clear Q^2 -dependence of the measured structure function $F_2(x, Q^2)$ as shown in figure 3.5. In the literature

this observation is called **scaling violations** and is nowadays explained by the QCD dynamics. As indicated on the timeline in figure 1.5 the parton model was invented in 1969, which is few years before QCD was formulated. Therefore it is plausible that naive the parton model provides only an approximate description.

The origin of the scaling violations is explained by the higher-order QCD corrections to the DIS cross sections caused by higher-order perturbative processes [108–110] such as

- virtual corrections,
- gluon emission,
- initial state gluon,

described in the following. The Feynman graphs for QCD corrections to DIS processes at NLO are shown in figure 3.6. A sample selection of contributing corrections at NNLO is shown in figure 3.7 for demonstrative purpose. For a complete and detailed description of all second order QCD corrections to DIS processes please refer to [110]. **Virtual corrections** occur in processes that contain one quark in the initial state and one quark in the final state (cf. figures 3.6(a), 3.7(a)). For example, these processes include one-loop corrections such as the vertex correction and the quark self-energy correction. Another type of processes are **radiative corrections** where a real gluon is emitted either from the initial or from the final state quark (cf. figures 3.6(b), 3.7(b)). The third category of higher-order corrections considers an **initial state gluon**, that is when an incoming gluon is present in the initial state (cf. figures 3.6(c), 3.7(c)). Since the gauge boson cannot interact with a gluon, a quark-antiquark pair is created in such processes leading to two final state quarks.

Similar processes can also appear as higher-order QCD corrections in the so-called DGLAP evolution equations (cf. chapter 4). As shown in Figure 3.5 the dependence of the structure function $F_2(x, Q^2)$ on Q^2 is particularly large at small x [11], which is also described by the DGLAP equations. These DGLAP equations, which we will apply in our global PDF analysis, are introduced in the next chapter.

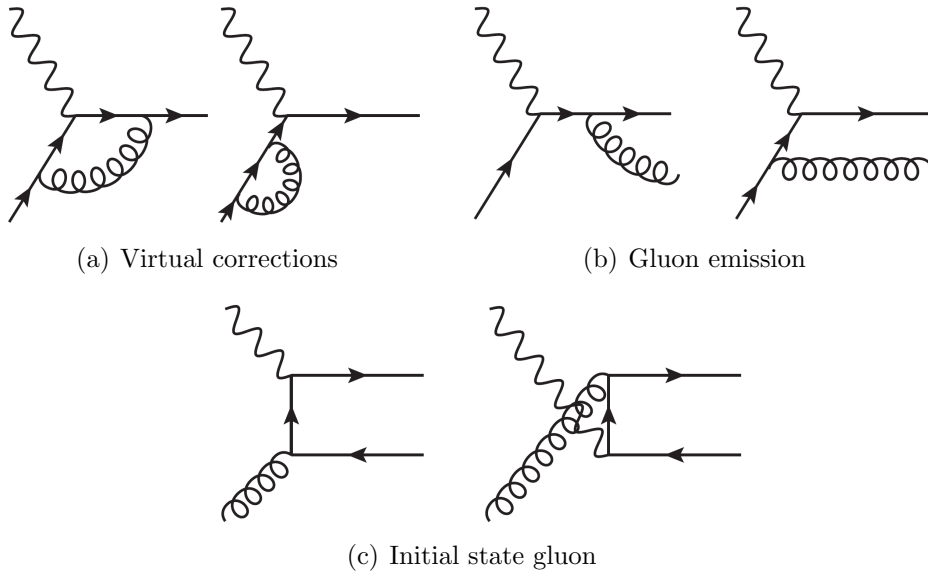


Figure 3.6.: Diagrams for the QCD corrections at NLO. Quarks are represented by arrows, spiral lines symbolize gluons, wavy line stands for a gauge boson (e.g. γ). a) processes with one quark in the initial state and one quark in the final state. b) processes with a real emitted gluon. c) processes with an initial state gluon. Graphs credited from [108].

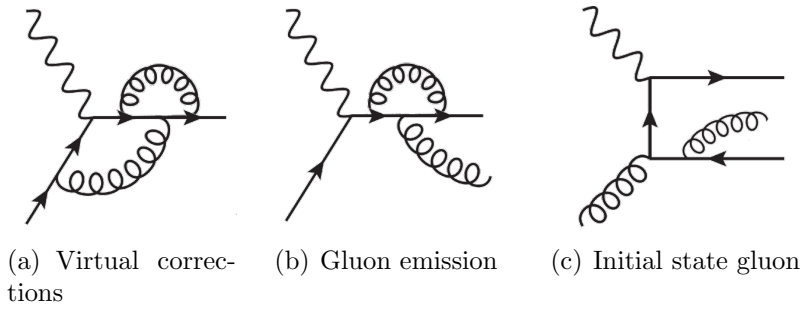


Figure 3.7.: A selection of sample diagrams for the QCD corrections at NNLO. Quarks are represented by arrows, spiral lines symbolize gluons, wavy line stands for a gauge boson (e.g. γ). a) processes with one quark in the initial state and one quark in the final state. b) processes with a real emitted gluon. c) processes with an initial state gluon. Graphs credited from [108], supplementary modified for NNLO contributions by M. Walt.

3.4. Factorization Theorem and Parton Distribution Functions

Collinear Factorization Theorem: „ The field theory realization of the parton model is the theorem of *factorization* of long-distance from short-distance dependence for deeply inelastic scattering [111]. This theorem states that the sum of all diagrammatic contributions to the structure functions is a direct generalization of the parton model results [...].“ [56]. That is, the collinear factorization [23, 24] allows to factorize, i.e. to disjoint, the *process-dependent* partonic, perturbative scattering cross section and the *universal*, i.e. process-independent, non-perturbative parton distribution functions. In the framework of the collinear factorization the partons' momenta are assumed to be parallel, i.e. collinear, to the total momentum of the parent hadron.

For the general case of hadron-hadron scattering between h_1 and h_2 with the final states k, X the factorization formula is

$$d\sigma^{h_1+h_2 \rightarrow k+X} = \sum_{i,j,X'} f_i(x_1, Q^2) \otimes f_j(x_2, Q^2) \otimes d\hat{\sigma}^{ij \rightarrow k+X'} + \mathcal{O}(1/Q^2) \quad (3.22)$$

with the parton distribution functions $f_{i,j}(x_{1,2}, Q^2)$ of the involved hadrons h_1, h_2 , and the partonic cross section $d\hat{\sigma}$. The factorization scale μ_f (hier $\mu_f^2 = Q^2$) serves to define the separation of short-distance from long-distance effects [56]. The so-called power corrections indicated by $\mathcal{O}(1/Q^2)$ in equation (3.22), but of the form $\mathcal{O}(\Lambda_{\text{QCD}}^2/Q^2)^n$ with the QCD scale parameter Λ_{QCD} in general, are neglected in this analysis since Q^2 is assumed to be sufficiently high. However, these terms might become important for small Q^2 . The power corrections might appear, e.g. if the partons of the nucleon are not exactly collinear but carry some transverse momentum p_T . Such terms can arise from multi-parton interactions, which are potentially different in a free proton compared to a bound nucleon. Although the treatment of power corrections might be different in the free proton and in the nuclear PDF analyses, the collinear factorization theorem is valid for both. However, the global nPDF analysis is of essence to show its validity.

The formal proofs of the collinear factorization theorem are mathematically demanding. Therefore, such proofs exist only for few processes involving hadrons in the initial

state, like for example DIS. For all other hard hadronic interactions, i.e. hadronic scattering processes at a large invariant scale, it is typically assumed to be valid. For the special case of DIS, where only one hadron is involved in the scattering process, we write

$$\frac{d^2\sigma}{dx dQ^2} = \sum_i \frac{d^2\hat{\sigma}_i(x, Q^2)}{dx dQ^2} \otimes f_i(x, Q^2) . \quad (3.23)$$

At LO the **parton distribution function (PDF)** $f_i(x, Q^2)$ for the parton flavour $i = q, \bar{q}, g$ can be interpreted as the probability density for the parton i to carry a fraction x of the total longitudinal momentum of the proton. With the evolution description by the DGLAP equations the parton distribution function depends on the energy scale Q^2 , as described by QCD and observed by the experiments.

One important property of the PDFs is **universality**. The term universal implies that the same distributions are valid in different processes, e.g. in lepton-nucleus and hadron-nucleus collisions, but also for different measurements of the same scattering process. The global PDF analysis is an important verification of this property since different processes and data from different measurements are considered there.

The PDFs describe the *longitudinal* momentum distribution for single partons with respect to the total momentum of the *initial*-state hadron, i.e. a free proton in case of free proton PDFs or a bound nucleon in case of nuclear PDFs. They provide no information on the final-state hadron. As mentioned earlier, the DIS processes considered in this work are fully inclusive, i.e. there is no differentiation in regards to the final hadronic state. However, if one wants to consider semi-inclusive processes the fragmentation functions (FFs) become relevant. The fragmentation functions describe the propagation of the partons involved in the considered scattering reaction into measurable hadrons in the final state. Usually the FFs, pursuant to PDFs, are analyzed in a stand-alone analysis [112–118]. Nowadays, there are attempts for a simultaneous global analysis of PDFs and FFs, e.g., [119].

As described above, the PDFs depend on the longitudinal momentum fraction x and are based on the collinear factorization. Therefore they do not provide any information about the transverse structure of hadrons. The transverse momentum distribution is

3. *Theoretical Principles*

described by TMDs (transverse-momentum dependent PDFs). The correlation between the parton's transverse position in the nucleon and its longitudinal momentum is provided by GPDs (generalized parton distributions). GPDs and TMDs are not considered as part of this work and are mentioned here for the reason of completeness. A nice introduction to GPDs and TMDs can be found in [120], whereas an overview of the nucleon tomography is provided in [121].

4. DGLAP Evolution

The scale evolution at NLO and NNLO of parton distribution functions from the initial energy scale Q_0 to any higher energy scale Q is one of the key steps of the global nPDF analysis performed as part of this work. The relevance in regard to the overall analysis process is shown in figure 4.1 for a better overview. The scale evolution is based on the so-called Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [25–28] which are described in this chapter. The procedure to solve these equations is outlined. Furthermore, the effects of the higher-order QCD corrections and the limitations of the DGLAP equations are discussed here.



Figure 4.1.: Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: scale evolution.

4.1. DGLAP Equations and Splitting Functions

As introduced above, the idea of a global PDF analysis includes a process step called scale evolution. First, one starts with initial parton distributions, e.g. for gluon (g), valence-like quarks (u_v, d_v) and the light sea quarks (u, d, s), at some very low energy scale Q_0^2 . Then, the DGLAP evolution equations are applied to evolve the initial distributions to a higher energy scale. The application of DGLAP equations leads to a radiative generation of gluons and sea quarks, which typically cause increased gluon and sea-quark distributions at higher energy scales (here Q^2) especially in the kinematic region of the low momentum fraction x . At sufficiently large energies, i.e. above the mass threshold of the heavy quarks, also distributions for the heavy sea quarks, e.g. charm (c), are generated by the DGLAP evolution. For the reason of completeness, it needs to

be mentioned that there are investigations by other groups performing PDF analyses to include so-called intrinsic charm, that is non-vanishing c distributions at the initial energy scale, cf. e.g. references [122, 123]. In this chapter, the DGLAP evolution equations and the corresponding splitting functions are presented based on references [43, 51].

In general, the **DGLAP equations** are $(2n_f+1)$ -dimensional matrix equations (n_f number of quark flavours) [43]:

$$t \frac{\partial}{\partial t} \begin{pmatrix} q_i(x,t) \\ g(x,t) \end{pmatrix} = \frac{\alpha_S(t)}{2\pi} \sum_j \int_x^1 \frac{d\xi}{\xi} \begin{pmatrix} P_{q_i q_j} \left(\frac{x}{\xi}, \alpha_S(t) \right) & P_{q_i g} \left(\frac{x}{\xi}, \alpha_S(t) \right) \\ P_{g q_j} \left(\frac{x}{\xi}, \alpha_S(t) \right) & P_{g g} \left(\frac{x}{\xi}, \alpha_S(t) \right) \end{pmatrix} \begin{pmatrix} q_j(\xi,t) \\ g(\xi,t) \end{pmatrix}. \quad (4.1)$$

The DGLAP equations (4.1) are applied to evolve the distribution functions of quarks $q_i(x,t)$ and gluons $g(x,t)$ from an initial energy scale t_0 to a higher scale t , as introduced above. Usually variable t is set to the scale μ^2 ($t = \mu^2$) or as in this work to $t = \mu^2 = Q^2$. The momentum fraction x is fixed during the scale evolution. The matrix in equation (4.1) is the so-called evolution kernel with **splitting functions** $P_{ab}(x/\xi, \alpha_S)$ for the different parton flavours $a, b = q, \bar{q}, g$. In the framework of perturbative QCD the splitting functions are provided in form of a perturbative expansion in $\alpha_S(t)$ which defines the perturbative order (LO, NLO or NNLO) of the DGLAP equations, i.e.

$$P_{ab}(z, \alpha_S) = P_{ab}^{(0)}(z) + \frac{\alpha_S}{2\pi} P_{ab}^{(1)}(z) + \dots, \quad (4.2)$$

with $z = x/\xi$ as usually used in the literature. At the leading order (LO) a splitting function $P_{ab}^{(0)}(z)$ describes the probability of finding a parton of type a in a parton of type b with a fraction z of the longitudinal momentum of the parent parton. This interpretation of $P_{ab}^{(0)}(z)$ implies

- positive definiteness for $z < 1$,
- quark number conservation and momentum conservation in the splittings of quarks and gluons:

$$\int_0^1 dz P_{qq}^{(0)}(z) = 0, \quad (4.3)$$

$$\int_0^1 dz z [P_{qq}^{(0)}(z) + P_{gq}^{(0)}(z)] = 0, \quad (4.4)$$

$$\int_0^1 dz z [2n_f P_{qg}^{(0)}(z) + P_{gg}^{(0)}(z)] = 0 \quad (4.5)$$

at LO. In general, the splitting function $P_{ab}(z, \alpha_S)$ satisfies a series of symmetries and physical properties [43]:

- charge conjugation invariance, i.e. P_{qg}, P_{gq} are the same for quark and antiquark,
- $SU(n_f)$ flavour symmetry, i.e. P_{qq} is defined as independent of the quark flavour.

The visualization of the splitting functions at LO is shown in figure 4.2. The described properties and the representation in figure 4.2 are valid for the unpolarized case, as relevant for this work. As can be seen from figure 4.2, during the scale evolution, a quark can generate (radiate) a gluon, and a gluon can split into quarks, which leads to radiatively generated sea quarks and gluons.

$$C_F = 4/3, \quad T_R = 1/2, \quad C_G = 3$$

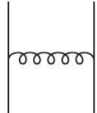
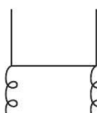
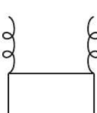
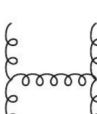
	$P_{qq}^{(0)}(z) = C_F \left[\frac{1+z^2}{1-z} + \frac{3}{2} \delta(1-z) \right]$
	$P_{qg}^{(0)}(z) = T_R [z^2 + (1-z)^2]$
	$P_{gq}^{(0)}(z) = C_F \left[\frac{1+(1-z)^2}{z} \right]$
	$P_{gg}^{(0)}(z) = 2C_G \left[\frac{z}{1-z} + \frac{1-z}{z} + z(1-z) \right] + \delta(1-z) \frac{11C_G - 4n_f T_R}{6}$

Figure 4.2.: Feynman diagrams and expressions for the unpolarized splitting functions at LO. Figure credited from [109], supplementary modified by M. Walt.

In this work, the global PDF analysis is performed at NLO and NNLO. The splitting functions at NLO can be found in references [27, 124–133] and those at NNLO (so-called three-loop splitting functions) are published in [134, 135].

4.2. Solving DGLAP Equations

In this section the approach for solving the DGLAP equations given in (4.1) is described. As part of the global PDF analysis the DGLAP equations are solved in every iteration step. The methods of solving DGLAP evolution equations can be assigned into two categories by differentiating if they are solved directly in x -space or transformed in the so-called Mellin momentum space or N -space.

The advantage of solving DGLAP equations in the N -space is that the Mellin transform turns the convolution in equation (4.1) into a simple multiplication, which is numerically more efficient. The disadvantage of the Mellin method is that the functional form of the initial parameterization is somewhat limited as it needs to be transformable into Mellin space. Therefore, one needs to know the Mellin transform of the initial conditions, but also of the splitting functions. In this regard the x -space is supposed to be more flexible since the input is provided directly in x -space, but it is numerically less efficient by performing the necessary convolution in equation (4.1). Both solution methods are outlined in this section. A summary of available numerical routines is provided in subsection 4.2.3.

The full set of evolution equations for quark distributions q , anti-quark distributions \bar{q} and gluon distribution g can be written as

$$\begin{aligned}
 Q^2 \frac{\partial q_i}{\partial Q^2} &= \frac{\alpha_S}{2\pi} \left[\sum_j P_{q_i q_j} \otimes q_j + \sum_j P_{q_i \bar{q}_j} \otimes \bar{q}_j + P_{qg} \otimes g \right], \\
 Q^2 \frac{\partial \bar{q}_i}{\partial Q^2} &= \frac{\alpha_S}{2\pi} \left[\sum_j P_{\bar{q}_i q_j} \otimes q_j + \sum_j P_{\bar{q}_i \bar{q}_j} \otimes \bar{q}_j + P_{\bar{q}g} \otimes g \right], \\
 Q^2 \frac{\partial g}{\partial Q^2} &= \frac{\alpha_S}{2\pi} \left[\sum_j P_{gq_j} \otimes q_j + \sum_j P_{g\bar{q}_j} \otimes \bar{q}_j + P_{gg} \otimes g \right], \tag{4.6}
 \end{aligned}$$

with i, j running over the different quark flavours. The arguments of the parton densities, the splitting functions and the strong coupling constant are omitted in equation (4.6) for a compact representation. The symbol \otimes in equation (4.6) denotes a convolution between the splitting functions and the parton distribution functions.

4.2.1. Solution in N -space

The DGLAP equations (4.6) are integro-differential equations. In order to reduce the convolution, that is an integration, into a simple multiplication a Mellin transform into the N -space is performed. The approach described in this section and the used notation are based on references [136–138]. First, the *approach* to solve the DGLAP equations in N -space will be described based on a representative example, i.e. the simpler case of a non-singlet distribution at NLO, for demonstrative purpose. Next, the *expressions* for the obtained parton distributions for the general case, i.e. up to order NNLO or N^mLO respectively, will be provided based on reference [138] for the reason of completeness.

The **Mellin transform** of a function $f_i(x)$ is defined as

$$\tilde{f}_i(N) \equiv \int_0^1 dx x^{N-1} f_i(x), \quad (4.7)$$

with the inverse transform from N -space to x -space

$$f_i(x) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} dN x^{-N} \tilde{f}_i(N). \quad (4.8)$$

The integral over N -space in equation (4.8) runs along a contour that is parallel to the imaginary axis and lies on the right of all singularities of $\tilde{f}_i(N)$, which are located on the real axis in the present case. The arbitrary real number a can be adjusted to chose the proper contour.

The transformation (4.8) from N -space to x -space needs to be performed as the final step in order to obtain Bjorken- x dependent PDFs which can be compared to the experimental data that are provided dependent on Bjorken- x , as well. However, for the solution of the DGLAP equation, one first performs the Mellin transform (4.7) to obtain DGLAP equations which are not integro-differential equations anymore, but are usual,

coupled differential equations. Next, a different basis of distribution functions needs to be selected in order to decouple the equations as far as possible. For that purpose, first we define the so-called **non-singlet** quark distributions

$$\begin{aligned} v_{-,i} &\equiv q_i - \bar{q}_i, \\ v_{+,l} &\equiv \sum_{i=1}^k (q_i + \bar{q}_i) - k(q_k + \bar{q}_k) \end{aligned} \quad (4.9)$$

where $v_{+,0} \equiv 0$, and q, \bar{q} are the quark distributions (meant to be Mellin transformed in this section), with $i = 1, \dots, f$, $k = 1, \dots, f$, $l = k^2 - 1$ and f the number of active quark flavours. Second, the **singlet vector** \vec{q} containing the quark singlet Σ and the gluon distribution g is defined as

$$\vec{q} = \begin{pmatrix} \Sigma \\ g \end{pmatrix}, \quad \text{and} \quad \Sigma = \sum_{i=1}^f (q_i + \bar{q}_i). \quad (4.10)$$

Now, one can write the DGLAP evolution equations for the non-singlet combinations v_-, v_+ and the singlet vector \vec{q} , which will decouple the differential equations to a certain extent. For the decomposition of the splitting functions P_{qq} and $P_{q\bar{q}}$ one chooses the following notation to separate the flavour-preserving valence part and the possibly-flavour-changing sea quarks part:

$$P_{q_i q_k} = \delta_{ik} P_{qq}^V + P_{qq}^S \quad \text{and} \quad P_{q_i \bar{q}_k} = \delta_{ik} P_{q\bar{q}}^V + P_{q\bar{q}}^S. \quad (4.11)$$

All splitting functions in equation (4.11) are supposed to be Mellin transformed in this section.

After performing the first two steps:

1. the Mellin transform into N -space (cf. eq. (4.7)), and
2. decoupling the differential equations by the definition of a new basis (cf. eq. (4.9), (4.9)),

the simplest resulting equation is the one for the non-singlet v_- which is given here for

demonstrative purpose

$$\frac{d}{d \ln Q^2} v_{-,i} = \frac{\alpha_S}{2\pi} P_- v_{-,i} \quad (4.12)$$

with $P_- \equiv P_{qq}^V - P_{q\bar{q}}^V$. For a complete set of resulting DGLAP equations in N -space please refer to [136, 137].

Furthermore, as mentioned in the beginning, the evolution step as part of the global analysis is based on perturbative QCD and so is object of perturbative expansion in α_S . Therefore, the DGLAP evolution equations are rewritten in terms of α_S . For that one uses the expanded form of equation (2.10)

$$\frac{\partial \alpha_S}{\partial \ln \mu^2} = -\beta_0 \alpha_S^2 - \beta_1 \alpha_S^3 - \beta_2 \alpha_S^4 + \mathcal{O}(\alpha_S^5), \quad (4.13)$$

with β_i ($i = 0 \dots 2$) as given in [52]:

$$\begin{aligned} \beta_0 &= \frac{11}{3} C_A - \frac{4}{3} T_F n_f, \\ \beta_1 &= \frac{34}{3} C_A^2 - 4 C_F T_F n_f - \frac{20}{3} C_A T_F n_f \\ \beta_2 &= \frac{2857}{54} C_A^3 + 2 C_F^2 T_F n_f - \frac{205}{9} C_F C_A T_F n_f \\ &\quad - \frac{1415}{27} C_A^2 T_F n_f + \frac{44}{9} C_F T_F^2 n_f^2 + \frac{158}{27} C_A T_F^2 n_f^2, \end{aligned} \quad (4.14)$$

where $C_F = 4/3$, $T_F = 1/2$, $C_A = 3$ and n_f is the number of quark flavours. In addition to the perturbative expansion in terms of α_S , also the splitting functions are expressed in form of a perturbative series (cf. eq. (4.2)). So, after

3. performing the perturbative expansion in terms of α_S , and
4. using the perturbative series of the splitting functions,

one receives the DGLAP equations in N -space. Since the NNLO terms become very long, we provide the NLO expression for demonstrative purpose. For our representative example, the evolution equations for the non-singlet densities $v_{\pm,i}$ at NLO are

$$\frac{d}{d \alpha_S} v_{\pm,i} = -\frac{-2}{\beta_0 \alpha_S} \left\{ P_{qq}^{(0)} + \frac{\alpha_S}{2\pi} \left(P_{\pm}^{(1)} - \frac{\beta_1}{2\beta_0} P_{qq}^{(0)} \right) + \mathcal{O}(\alpha_S^2) \right\} v_{\pm,i} \quad (4.15)$$

with the splitting functions P_{\pm}, P_{qq} in Mellin space. Now, the ordinary differential equations (4.15) can be solved by separation of the variables. The straightforward solution is

$$v_{\pm,i}(Q^2) = \left\{ 1 - \frac{\alpha_S(Q^2) - \alpha_S(Q_0^2)}{\pi\beta_0} \left(P_{\pm}^{(1)} - \frac{\beta_1}{2\beta_0} P_{qq}^{(0)} \right) + \mathcal{O}(\alpha_S^2) \right\} \cdot [\alpha_S(Q^2)/\alpha_S(Q_0^2)]^{-(2/\beta_0)P_{qq}^{(0)}} v_{\pm,i}(Q_0^2), \quad (4.16)$$

here at NLO. For the expressions of three-loop splitting functions for the singlet and non-singlet cases the reader is referred to [134, 135]. The input distribution $v_{\pm,i}(Q_0^2)$ at the initial scale Q_0^2 in equation (4.16) results from the boundary condition selected here. The singlet case is more complicated since the resulting DGLAP equations are coupled for the quarks and gluon densities. A detailed description can be found in references [136–138].

The described *approach* to solve the DGLAP equations in N -space has been demonstrated based on a representative example of the non-singlet distribution at NLO. Now, we want to provide the *expressions* for the general case based on reference [138]. First, we write the scale dependence of parton distribution functions f_j of quarks, antiquarks and gluons as given by the evolution equations

$$\frac{\partial}{\partial \ln \mu^2} f_i(x, \mu^2) = \sum_j P_{ij}(x, \mu^2) \otimes f_j(x, \mu^2) \quad (4.17)$$

with the factorization scale μ (here $\mu = \mu_r$) and the momentum fraction x . Here \otimes denotes Mellin convolution, i.e. a simple product in N -space. The splitting functions $P_{ij}(x, \mu^2)$ up to N^mLO are approximately given by

$$P_{ij}^{\text{N}^m\text{LO}}(x, \mu^2) = \sum_{k=0}^m a_S^{k+1}(\mu^2) P_{ij}^{(k)}(x), \quad (4.18)$$

so that $P_{ij}(x, \mu^2)$ depends on the scale μ^2 only via the coupling $a_S(\mu^2)$. Following the notation from [138], $a_S(\mu^2)$ is the strong coupling constant normalized to $a_S \equiv \alpha_S/(4\pi)$

that satisfies the relation for the β -function to the order N^mLO

$$\frac{da_S}{d \ln \mu_r^2} = \beta_{\text{N}^m\text{LO}}(a_S) = - \sum_{k=0}^m a_S^{k+2} \beta_k, \quad (4.19)$$

where μ_r is the renormalisation scale, and β_k are the expansion coefficients of the β -function. For the general case $\mu \neq \mu_r$ the splitting functions (4.18) at NNLO, which is the relevant order for this analysis, obtained by Taylor expanding $a_S(\mu^2)$ in terms of $a_S(\mu_r^2)$, read

$$\begin{aligned} P_{ij}^{\text{NNLO}}(x, \mu, \mu_r) = & a_S(\mu_r^2) P_{ij}^{(0)}(x) \\ & + a_S^2(\mu_r^2) \left[P_{ij}^{(1)}(x) - \beta_0 L P_{ij}^{(0)}(x) \right] \\ & + a_S^3(\mu_r^2) \left[P_{ij}^{(2)}(x) - 2\beta_0 L P_{ij}^{(1)}(x) - (\beta_1 L - \beta_0^2 L) P_{ij}^{(0)}(x) \right] \end{aligned} \quad (4.20)$$

with $L \equiv \ln(\mu^2/\mu_r^2)$.

For the evolution of the distribution functions we select an alternative basis. The singlet-vector \vec{q} remains as given in equation (4.10), so the sum of quarks and antiquarks in the first component and the gluon distribution in the second component. For the non-singlet case we define

$$\begin{aligned} q_{\text{ns},ik}^{\pm} &= q_i \pm \bar{q}_i - (q_k \pm \bar{q}_k) && \text{(flavour asymmetries),} \\ q_{\text{ns}}^{\text{V}} &= \sum_{r=1}^{n_f} (q_r - \bar{q}_r) && \text{(total valence distribution).} \end{aligned} \quad (4.21)$$

In the selected basis the general DGLAP equations using the matrix notation, i.e. the evolution kernel containing the splitting functions, for the singlet case up to any perturbative order are

$$\begin{aligned} \frac{\partial \vec{q}(N, a_S)}{\partial a_S} &= \{\beta_{\text{N}^m\text{LO}}(a_S)\}^{-1} P^{\text{N}^m\text{LO}} \vec{q}(N, a_S) \\ &= - \frac{1}{a_S} \left[\mathbf{R}_0(N) + \sum_{k=1}^{\infty} a_S^k \mathbf{R}_k(N) \right] \vec{q}(N, a_S) \end{aligned} \quad (4.22)$$

with the rekursive relations

$$\begin{aligned} \mathbf{R}_0(N) &\equiv \frac{1}{\beta_0} \mathbf{P}^{(0)} && \text{and} \\ \mathbf{R}_k(N) &\equiv \frac{1}{\beta_0} \mathbf{P}^{(k)} - \sum_{i=1}^k b_i \mathbf{R}_{k-i}(N) && \text{with } b_k \equiv \beta_k/\beta_0, \end{aligned} \quad (4.23)$$

where $\mathbf{P}^{(k)}$ is the coefficient of a_S^{k+1} in equation (4.20). For the NNLO case, relevant in this PDF analysis, one obtains the singlet solution

$$\begin{aligned} \vec{q}_{\text{NNLO}}(a_S) = & \left[\mathbf{L} + a_S \mathbf{U}_1 \mathbf{L} - a_0 \mathbf{L} \mathbf{U}_1 \right. \\ & \left. a_S^2 \mathbf{U}_2 \mathbf{L} - a_S a_0 \mathbf{U}_1 \mathbf{L} \mathbf{U}_1 + a_0^2 \mathbf{L} (\mathbf{U}_1^2 - \mathbf{U}_2) \right] \vec{q}(a_0) \end{aligned} \quad (4.24)$$

with $a_0 \equiv a_S(\mu_{r,0}^2)$, $\mathbf{L} \equiv (a/a_0)^{-\mathbf{R}_0(N)}$ and the evolution matrices \mathbf{U}_k constructed from the splitting function combinations $\mathbf{R}_{l \leq k}$ as can be found in [138]. For the simpler case of the non-singlet distributions (cf. eq. 4.21) we provide the non-truncated solution at N^mLO

$$q_{\text{ns}}^{\pm, \text{V}}(a_S) = \left[1 + \sum_{k=1}^m a_S^k U_k^{\pm, \text{V}} \right] \left[1 + \sum_{k=1}^m a_0^k U_k^{\pm, \text{V}} \right]^{-1} L^{\text{ns}} q_{\text{ns}}^{\pm, \text{V}}(a_0) \quad (4.25)$$

where $L^{\text{ns}} \equiv (a/a_0)^{-\mathbf{R}_0^{\text{ns}}(N)}$ and $U_k^{\pm, \text{V}}$ the evolution matrices for the non-singlet case (cf. ref. [138]).

Finally, all solutions obtained in the N -space need to be transformed to x -space (cf. eq. (4.8)) for the purposes of a QCD analysis in order to provide the experimentally relevant Bjorken- x dependent parton distribution functions $f(x, Q^2)$. Alternatively, the DGLAP equations (4.6) can be solved directly in x -space as roughly outlined in the next subsection.

4.2.2. Solution in x -space

The integro-differential equations (4.6) can not be solved analytically in x -space, but by numerical routines, e.g. [139]. That is, the DGLAP evolution equations are solved numerically on a discrete $n \times m$ grid in x and μ^2 (here: Q^2). The total number of

operations is quadratic in n and linear in m [139]. The numerical method used in [139] is based on the polynomial spline interpolation [140] on an equidistant logarithmic grid in x and a logarithmic grid in μ^2 . The convolution integrals are evaluated as a weighted sum with the according weights computed at the initialization of the program. Please refer to [141–146] for further details on the numerical features.

4.2.3. Numerical Implementations

All high-level steps of the global analysis performed in this work (cf. fig. 5.2 or 5.1) have been solved numerically within the framework described in chapter 5. For the calculation of the DIS cross sections the corresponding integrals can be evaluated by existing programs, like e.g. QCDNUM [139]. Even though, the actual purpose of QCDNUM is solving the integro-differential DGLAP equations numerically. Also APFEL [147] provides the functionality of solving DGLAP equations. A recent comparison between QCDNUM and APFEL is provided in reference [148]. A selection of solution methods in x -space can be found in references [149, 150] and also in [141, 151], which is described for another numerical tool HOPPET. A routine implemented in Mellin space is provided by QCD-PEGASUS [138], and by MELA [152]. An approach which combines advantages of the N -space and x -space methods has been presented in [153].

4.3. Higher-Order Corrections

The origin of the scaling violations, described in section 3.3, is explained by the higher-order QCD corrections to the DIS cross sections caused by higher-order perturbative processes shown in figures 3.6 and 3.7. Similar processes can also occur as higher-order QCD corrections in the DGLAP evolution equations leading to interesting effects.

For example, the NLO evolution breaks the flavour symmetry $\bar{u} = \bar{d}$ of the sea quarks densities if the valence distributions are not identical, which is the case for proton. Even if the provided initial condition $\bar{u}(\mu_0^2) = \bar{d}(\mu_0^2)$ describes a symmetric sea, the DGLAP evolution at NLO leads to $\bar{u} \neq \bar{d}$ since the recursive relations for the splitting functions (cf. eq. (4.23)) are different, i.e. $R_1^+ \neq R_1^-$. For the symmetry of the strange sea quarks $s = \bar{s}$ the DGLAP evolution at NNLO results into $s \neq \bar{s}$ even if $s(\mu_0^2) = \bar{s}(\mu_0^2)$, whereas such an asymmetry at LO and NLO can be produced only by the according asymmetry

in the initial input parameters [154]. Both effects are very small. As shown in [154], the perturbatively generated asymmetry of strange quarks at scale Q^2 will be the larger the lower the initial scale Q_0^2 . However, the effect is recognizable especially at the low x -values. This kinematic region is not well constrained by the experimental data available for the performed nuclear PDF analysis (cf. section 6.1).

In order to study such higher-order effects our NNLO analysis can be an important step in this context, especially when the initially applied constraint $\bar{u} = \bar{d} = s = \bar{s}$ (cf. section 5.1.2) would be released in the next project phase by considering more processes sensitive to the sea quark distributions and by incorporating experimental data covering a larger kinematic region.

4.4. Limitations of DGLAP Equations

Equation (4.2) demonstrates that DGLAP equations include splitting functions perturbatively expanded in $\alpha_S(t)$. Due to the QCD property of a running coupling constant $\alpha_S(t)$, a perturbative expansion is valid only at sufficiently large energy scales $t = Q^2$. Furthermore, the higher-order corrections, e.g. gluon emission lead to summands with leading powers of $\alpha_S \ln(Q^2)$ [43]. In the limit of small Bjorken- x , which corresponds to the high energy of virtual photon-proton scattering, summands of the form [51, 155]

$$\alpha_S \ln \frac{Q^2}{Q_0^2} \ln \frac{1}{x} \tag{4.26}$$

appear. Thus, the perturbative series even if expanded in the small coupling parameter $\alpha_S \ll 1$ can have large contributions in case of large Q^2 and small x , so that $\ln(1/x) \gg 1$. Such large logarithmic terms which violate the principles of a perturbative expansion are treated by the techniques of resummation, where the logarithms are summed up to all orders in perturbation theory. For example, the resummation of logarithmic terms in the evolution kernels for the non-singlet case at small x is presented in [156]. Due to the terms of the form as given in (4.26) the resummation in $\alpha_S \ln(Q^2)$ is valid for $\ln(Q^2) \gg \ln(1/x)$, i.e. *not* for small x . In the *asymptotic limit* where $x \rightarrow 0$ but $Q^2 \rightarrow \infty$ the leading $\ln(Q^2)$ resummation can be supplemented by the double-leading-logarithm approximation (DLA), where only leading $\ln(1/x)$ terms are kept [43].

In a general, *non-asymptotic case*, the DGLAP evolution equations might not be valid in the small- x region. There, an alternative treatment can be performed based on the Balitsky-Fadin-Kuraev-Lipatov (BFKL) equation [157]. Another framework applicable in the small- x region is the Color Glass Condensate (CGC) [158–160] using the Jalilian-Marian, Iancu, McLerran, Weigert, Leonidov, Kovner (JIMWLK) equations [158–164]. Unlike DGLAP equations, JIMWLK equations are nonlinear equations, based on the nonlinear Balitsky-Kovchegov (BK) evolution equation [165, 166]. However, for the global analysis of nuclear PDFs the small- x region is not covered by the available experimental data, so that DGLAP equations keep their validity.

5. Framework of the Global Analysis

In this chapter the theoretical and numerical framework of the overall process is presented. The initial parameterization and the incorporated assumptions are described. Furthermore, the definition of the optimal fit is documented. The discussed numerical framework is based on an open-source tool, xFITTER, which has been modified to be applicable also for a nuclear PDF analysis. The required extensions of the code are explained here to illustrate the modifications required for nPDF analysis and so the differences to a free proton PDF analysis. The implemented extensions of the program code have been published providing a first open-source tool to analyse nuclear PDFs [167]. A short summary of a comparison between the free proton PDF analysis and the nuclear PDF analysis is provided in the last section of this chapter.

As described in the introduction, the PDFs cannot be calculated from first principles, but can be derived in a global analysis (cf. fig. 5.1). In addition to theoretical calculations, different experimental data need to be used to determine these distributions. Typically the PDFs are determined in a global analysis where the PDFs are first parameterized at an initial energy scale Q_0^2 . Next, the PDFs are evolved from the initial to a higher energy scale Q^2 by the DGLAP equations. Then, by using the evolved PDFs and the factorization theorem the cross sections are calculated. These theoretical quantities are compared to a wide range of experimental data at different energies in order to fix

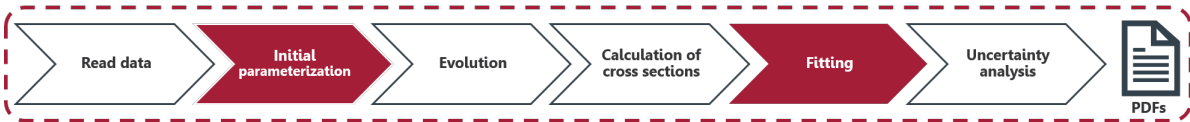


Figure 5.1.: Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: the framework of the overall process of global analysis, incl. initial parameterization, and the fitting part are explained.

the free parameters. There, the initial parameterization of the PDFs is varied until the optimal fit result is achieved, which is referred to as the fitting procedure in this work. After the optimal set of parameters is found, an error analysis is performed, that is described in chapter 7.

Note: since the implemented extensions of the program code have been made publicly available providing a first open-source tool to analyse nuclear PDFs, parts of the text in sections 5.1 and 5.2 have been credited from the own publication [1], for which the author has written the original draft of the manuscript (except introduction). The author has worked on the concept creation, numerical implementation and execution of the calculations published there. All numerical results presented in [1] have been obtained by the programs largely modified by the author. The performed modifications of the open-source code have been documented for the user guidance (cf. appendix B) by the author and made publicly available. The author has also prepared the released sets of nuclear parton distribution functions TUJU19 in a standard format [168].

5.1. Theoretical Framework

5.1.1. PDF Parameterization

A global DGLAP-based analysis requires a non-perturbative input for the PDFs at the initial scale of the fit. In this analysis parton distributions of a free proton and of a proton bound in a nucleus are parameterized as

$$x f_i^{p/A}(x, Q_0^2) = c_0 x^{c_1} (1-x)^{c_2} (1 + c_3 x + c_4 x^2) \quad (5.1)$$

for parton flavour $i = g, d_v, u_v, \bar{u}, \bar{d}, s$, at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. This form of PDF parameterization is similar to the functional form used in the HERAPDF2.0 analysis [169] and is motivated by the fact that the main constraints for the free proton PDF baseline come from the same DIS data. In order to keep the framework consistent we use the same form for the nuclear PDFs.

The main focus of this work is on the nuclear PDFs for which the fit parameters c_k in equation (5.1) are re-parameterized to be dependent on the nuclear mass number A

as

$$c_k \rightarrow c_k(A) = c_{k,0} + c_{k,1} (1 - A^{-c_{k,2}}) \quad (5.2)$$

where $k = 0, \dots, 4$. A similar form was also successfully used in the nCTEQ15 analysis [170]. At the same time, if $A = 1$ the A -dependent right hand part of equation (5.2) becomes zero and the free proton PDFs are recovered by default. The explicit A -dependence of the nuclear PDFs allows us to make predictions also for nuclei which were not part of the actual analysis, but are possibly interesting for future experiments.

As discussed, $f_i^{p/A}$ given in equation (5.1) defines the parton distribution of a proton bound to a nucleus with mass number A . In addition there are also neutrons in a nucleus which we denote by $f_i^{n/A}$. The full PDF for a nucleon inside a nucleus $f_i^{N/A}$ can be obtained by averaging over the number of protons and neutrons in nuclei:

$$f_i^{N/A}(x, Q^2) = \frac{Z \cdot f_i^{p/A} + (A - Z) \cdot f_i^{n/A}}{A}. \quad (5.3)$$

The PDFs of neutrons are not separately fitted, but are determined from the proton PDFs based on isospin symmetry.

5.1.2. Approximations and Constraints

A global nPDF analysis requires a series of approximations and constraints to overcome limitations on the part of either theoretical access or experimental data. The approximations, assumptions and constraints applied in this nPDF analysis are described in this section.

As mentioned above, the PDFs of bound neutrons are determined from the proton PDFs based on SU(2) symmetry, the so-called **isospin symmetry**. In particular, the distributions of u and d quarks are exchanged: $f_u^{p/A} \rightarrow f_d^{n/A}$ and $f_d^{p/A} \rightarrow f_u^{n/A}$, which is valid for valence and sea quarks. For completeness we mention that this interchange requires the validity of charge symmetry, and in Ref. [171] it has been suggested that some charge symmetry violation (CSV) could take place in the small- x region. However, in the x -region covered by the nuclear DIS data we use, such effects should be negligible. Besides the DIS experiments, CSV effects can be studied in experiments measuring

asymmetries in W boson production. Further experiments and tests of CSV in parton distributions are suggested in refs. [172, 173]. Similarly, possible isospin symmetry violations have been studied in refs. [174, 175]. In this work, however, we assume that the charge and isospin symmetries hold.

In addition to the isospin symmetry, we have assumed $s = \bar{s}$ and $s = \bar{s} = \bar{u} = \bar{d}$ as the included DIS data are not sensitive enough to constrain the strange-quark content, nor the sea quark flavour decomposition. In particular, even though the neutrino DIS data are sensitive to the separation of up- and down-type quarks, the kinematic region covered by the incorporated data ($x \gtrsim 0.01$) is where the valence quarks dominate the cross section.

Next, as part of the PDF analysis gluons and quarks are considered massless. Since this is not true for heavy quarks like charm and bottom a **mass scheme** for the treatment of **heavy quarks** needs to be chosen as part of the QCD analysis. Below the mass threshold $Q^2 < m_q^2$ (m_q quark mass) the heavy quarks can be considered as non-contributing because the energy is not sufficiently large for a creation of a quark anti-quark pair. Above the threshold $Q^2 > m_q^2$ the heavy quarks are created and can be treated as massless at high scales, i.e. $Q^2 \gg m_q^2$. This is called zero-mass approximation and can be used as zero-mass variable-flavour-number-scheme (ZM-VFNS) in a QCD analysis. However, near the threshold $Q^2 \sim m_q^2$ the zero-mass approximation is not sufficient anymore because the heavy quarks might be created by the radiation process but can not be considered as massless in this energy region. A consistent approach for the treatment of heavy quarks is provided by the so-called general-mass variable-flavour-number-scheme (GM-VFNS), see Ref. [176] for a recent overview. There are several options of GM-VFNS implemented in xFITTER^[1], including (S)ACOT schemes [179–182], RT and RT optimal schemes [183–185], as well as FONLL scheme [186, 187]. In this work we have applied the FONLL-A scheme for the NLO analysis and the FONLL-C at NNLO, implemented in the APFEL package [147]. The heavy quark masses are fixed to $m_{\text{charm}} = 1.43$ GeV and $m_{\text{bottom}} = 4.50$ GeV. The strong coupling constant α_S is set to $\alpha_S(M_Z) = 0.118$ for both the NLO and the NNLO fits.

[1] Also fixed-flavour mass schemes, like e.g. the ABM scheme [177, 178], are available in xFITTER.

For the parton distribution functions $x f_i^{p/A}$ as defined in equation (5.1), we assume the **baryon number sum rules** and the **momentum sum rule** to be satisfied by every nucleon in the nucleus,

$$\int_0^1 dx f_{u_v}^{p/A}(x, Q_0^2) = 2, \quad (5.4)$$

$$\int_0^1 dx f_{d_v}^{p/A}(x, Q_0^2) = 1, \quad (5.5)$$

$$\int_0^1 dx \sum_i x f_i^{p/A}(x, Q_0^2) = 1. \quad (5.6)$$

Strictly speaking, for nuclear parton distribution functions the sum rules are approximations that might not hold for individual nucleons in a nucleus in general, but which are reasonable at the available level of precision in regard to the experimental uncertainties. In this work, equation (5.4) is used to fix the normalization of u_v quarks in a proton and equation (5.5) defines the normalization of d_v quarks in a proton, nucleus per nucleus. The momentum sum rule (5.6) is used to constrain the normalization of the sea quarks. The remaining unconstrained normalization coefficient c_0^g in the gluon PDF is treated as a regular free parameter during the fitting procedure. Alternatively, the gluon normalization could have been fixed by the momentum sum rule as done in many earlier analyses, e.g. [188].

5.1.3. Fitting Procedure

The optimal values for the PDF parameters are obtained by minimizing χ^2 defined as

$$\chi^2 = \sum_i \frac{(\mu_i - \hat{m}_i)^2}{\Delta_i^2} + \sum_\alpha b_\alpha^2 \quad (5.7)$$

with

$$\hat{m}_i = m_i + \sum_\alpha \Gamma_{i\alpha} b_\alpha. \quad (5.8)$$

Here, μ_i is the value of the measured data point for a given observable, Δ_i is the uncorrelated experimental error, whereas the sum over correlated systematic errors is given

by the term $\sum_{\alpha} b_{\alpha}^2$ in equation (5.7). The theoretical predictions for each data point i are represented by \hat{m}_i , defined in equation (5.8). There, m_i is the actual theoretical value calculated using DGLAP-evolved PDFs with given parameters $\{c_k\}$, $\Gamma_{i\alpha}$ are the correlated errors and b_{α} are the so-called nuisance parameters. A nuisance parameter quantifies the strength of the correlated error source α , whereas $\Gamma_{i\alpha}$ quantifies the sensitivity of the i^{th} measurement to the correlated systematic error source α . The quality of the fit can be estimated from the resulting χ^2/N_{dp} ratio, where N_{dp} is the number of data points. A value $\chi^2/N_{\text{dp}} \approx 1$ indicates that the agreement between the theoretical prediction and the measured observable is on average at the level of the experimental uncertainties. Although the fitting procedure is a minimization problem, achievement of values $\chi^2/N_{\text{dp}} \ll 1$ is considered as overfitting which arises when an unreasonable amount of free parameters, compared to the constraining power of the framework and the experimental data, is used.

There are several ways to take into account the correlated and uncorrelated uncertainties and to combine the statistical and uncorrelated systematic uncertainties in the χ^2 definition in xFITTER which is used in this analysis. Here we use the following definition

$$\chi^2(\mathbf{m}, \mathbf{b}) = \sum_i \frac{[m^i - \sum_{\alpha} \gamma_{\alpha}^i \mu^i b_{\alpha} - \mu^i]^2}{\left(\delta_{i,\text{stat}} \sqrt{\mu^i m^i}\right)^2 + (\delta_{i,\text{uncorr}} m^i)^2} + \sum_{\alpha} b_{\alpha}^2. \quad (5.9)$$

The variables have been introduced in (5.7) with $\Gamma_{\alpha}^i = \gamma_{\alpha}^i \mu^i$, and $\delta_{i,\text{stat}}$ and $\delta_{i,\text{uncorr}}$ are the relative statistical and uncorrelated systematic uncertainties, respectively.

In general, if not absolute but relative errors are used in the analysis, different scaling options exist which have been studied and are described below for reference purpose. The numerical implementation in xFITTER, which was the tool used for this analysis, foresees an algorithm which calculates relative values (percentage) even if absolute errors are provided. These calculated relative errors need to be turned into absolute values during the fitting procedure. For that, the percentage values are multiplied either with the measured cross sections or the theoretical predictions, or the combination of both (cf. table 5.1). This is referred to as "scaling" in this section.

Scaling name	Scaling rule
noscale	μ_i
linear	m_i
poisson	$\sqrt{m_i \mu_i}$

Table 5.1.: Properties that can be applied for the scaling of different kinds of uncertainties. The above form of χ^2 in equation (5.9) which was used in this analysis corresponds to a Poisson-like scaling for the statistical experimental uncertainties, whereas the systematic uncorrelated and correlated uncertainties are scaled linearly.

The simplest form of combining different experimental uncertainties is without any rescaling by defining the total uncorrelated experimental error as $\Delta_i = \sqrt{\delta_{i,\text{stat}}^2 + \delta_{i,\text{syst}}^2} \mu_i$. We want to call this option "noscale" in table 5.1. Alternative scaling schemes are "linear" $\Delta_i = \sqrt{\delta_{i,\text{stat}}^2 + \delta_{i,\text{syst}}^2} m_i$ or "poisson" $\Delta_i = \sqrt{\delta_{i,\text{stat}}^2 + \delta_{i,\text{syst}}^2} \sqrt{m_i \mu_i}$.

The form of χ^2 used in this analysis (cf. eq. (5.9)) corresponds to a Poisson-like scaling for the statistical experimental uncertainties, whereas the systematic uncorrelated and correlated uncertainties are scaled linearly. This choice for χ^2 is similar to the one used in the HERAPDF2.0 analysis incorporating the combined H1 and ZEUS DIS data [169], on which our proton PDF baseline is mainly based. For consistency the same form has been used also for the nuclear PDF analysis.

The different scaling options will be followed up in the study of the Monte Carlo (MC) error analysis as described in chapter 7 section 7.2, and are listed here for the reason of completeness.

5.2. Numerical Framework

The global analyses of the baseline proton and nuclear PDFs are performed with the xFITTER [190, 191] tool. The main goal of the xFITTER project is to provide an open-

source tool to fit proton PDFs with different theoretical assumptions. A schematic view of the fitting procedure and relations to different external programs are shown in figure 5.2. Being an open-source tool it is available to everyone and makes the research process fully transparent, which is important in order to establish a common knowledge base and a deep understanding of the opportunities and limitations. The released version covers various options like different PDF parameterization forms, mass schemes, etc. Furthermore, xFITTER provides interfaces programmed in Fortran or C++ to the commonly used tools like MINUIT [192, 193], QCDNUM [139], APPLGRID [194] or APFEL [147], etc. The DGLAP evolution routine and the calculation of DIS cross sections are implemented at NNLO. Further functionalities in regards to the future po-

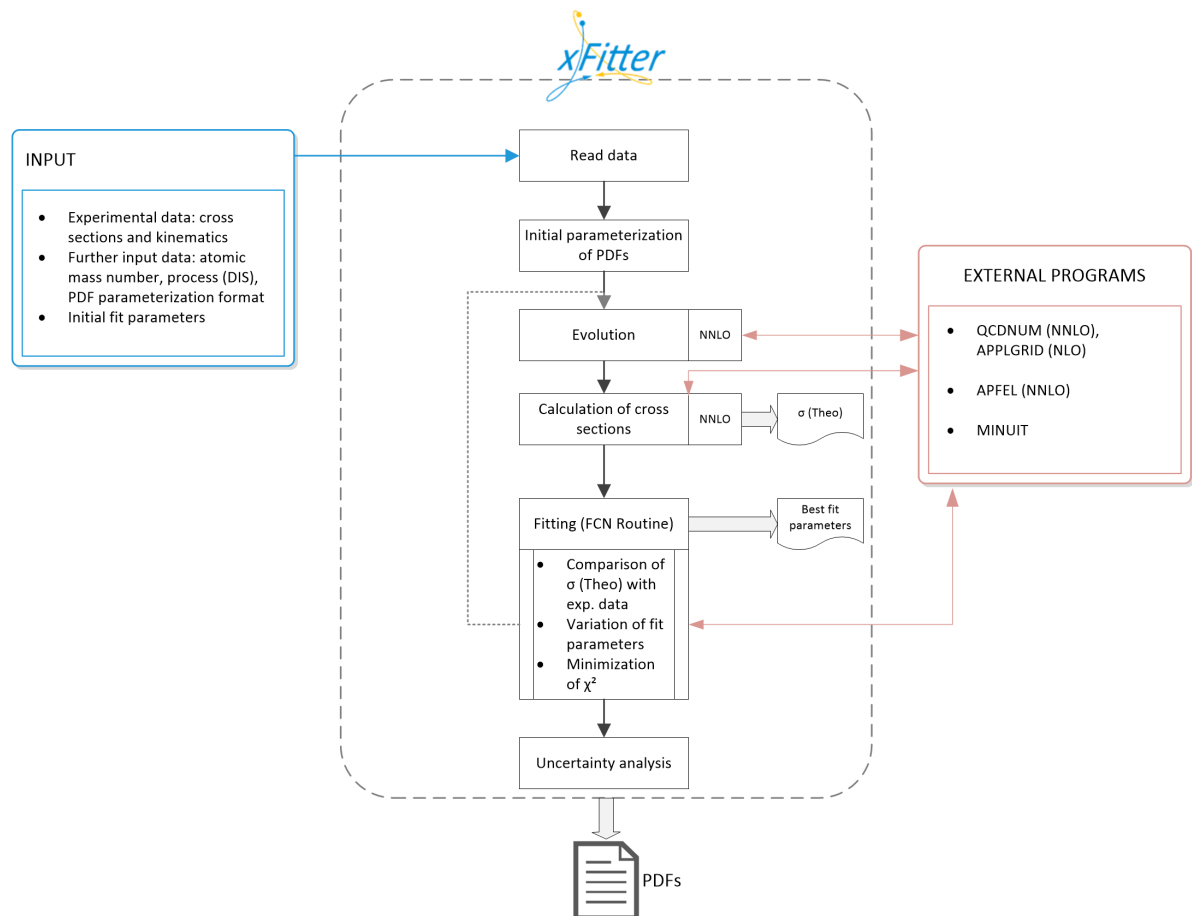


Figure 5.2.: Schematic view of the high-level xFITTER functionalities. xFITTER logo credited from [189]. Figure from [1] modified by M.Walt.

tential and alternative fitting approaches, including dipole models [195, 196] and small- x resummation [197, 198], are available in the released version.

In order to perform a nuclear PDF analysis several modifications of the code were required. First, the PDF parameterization had to be adapted for the purpose of nuclear PDFs. Thus, new parameters $c_k(A)$ dependent on the nuclear mass number A as per equation (5.2) were introduced. In order to reflect the new nPDF parameterization, the form of the steering file, as well as the file containing the initial parameters for MINUIT, and the according interpretation routine were adapted. As the next step, the mass number A and proton number Z of a given nucleus for the nucleon decomposition (cf. eq. (5.3)) of the up and down quarks were included. The possible combinations of data sets for different mass numbers A and proton numbers Z were kept flexible in order to deal with data for ratios between different nuclei. The information on A and Z depends on the data set and thus needs to be provided inside the data files. Therefore, the form and the routine to read the experimental data files were extended accordingly. Additionally, the overall minimization routine FCN has been modified so that the DGLAP evolution can evolve nuclear PDFs covering different combinations of A and Z individually. Next, the calculation of sum rules (equations 5.4, 5.5, 5.6) had to be adapted in order to reflect the flexibility of an A -dependent normalization. Additionally, to keep the form of the PDF parameterization flexible, a new numerical integration routine for the calculation of sum rules was implemented.

Besides that, the cross section calculation routine was enhanced for the treatment of various isoscalar modifications, as described in section 6.2. Three flags identifying the 'NMC', 'EMC' or 'SLAC' forms of the correction (cf. sec. 6.2) have been implemented. Furthermore, experimental nuclear data is often provided in terms of ratios $\sigma(A_1)/\sigma(A_2)$ or $F_2(A_1)/F_2(A_2)$ for two different nuclei with mass numbers A_1 and A_2 . Thus, we had to extend the xFITTER mechanisms for the consideration of these ratios by the implementation of a two-step loop. The underlying PDF flavour decomposition for a proton was modified so that in case of a nucleus the PDF decomposition is applied for a nucleon of the form (5.3).

Moreover, charged-current (CC) processes for the neutrino DIS data were incorporated

5. Framework of the Global Analysis

in xFITTER according to the differential cross section described in subsection 3.2. Originally, the implemented calculation of the reduced CC DIS cross section by using the FONLL scheme in xFITTER was customized to the HERA framework, e.g. factor 1/2 in equation (8) in [199], for the lepton-proton DIS data. For the neutrino-nucleus DIS data, we are using the non-reduced differential cross section with the prefactor from [11], and especially with the structure functions defined in equations (3.19) and (3.20). Thus, for the calculation of the CC DIS cross sections, a division by factor of 2 was removed from xFITTER in case of neutrino beams.

Finally, the uncertainty analysis routine [200, 201] has been modified so that scaling of the error bands (cf. eq. 7.5) can be performed also for $\Delta\chi^2 > 1.0$. In this work the option 'DoBands' has been used to generate asymmetric error bands, which is based on the 'iterate' method by John Pumplin [201]. Its advantage is that, if necessary, the iteration routine will add a positive value X to all eigenvalues to force the matrix to be

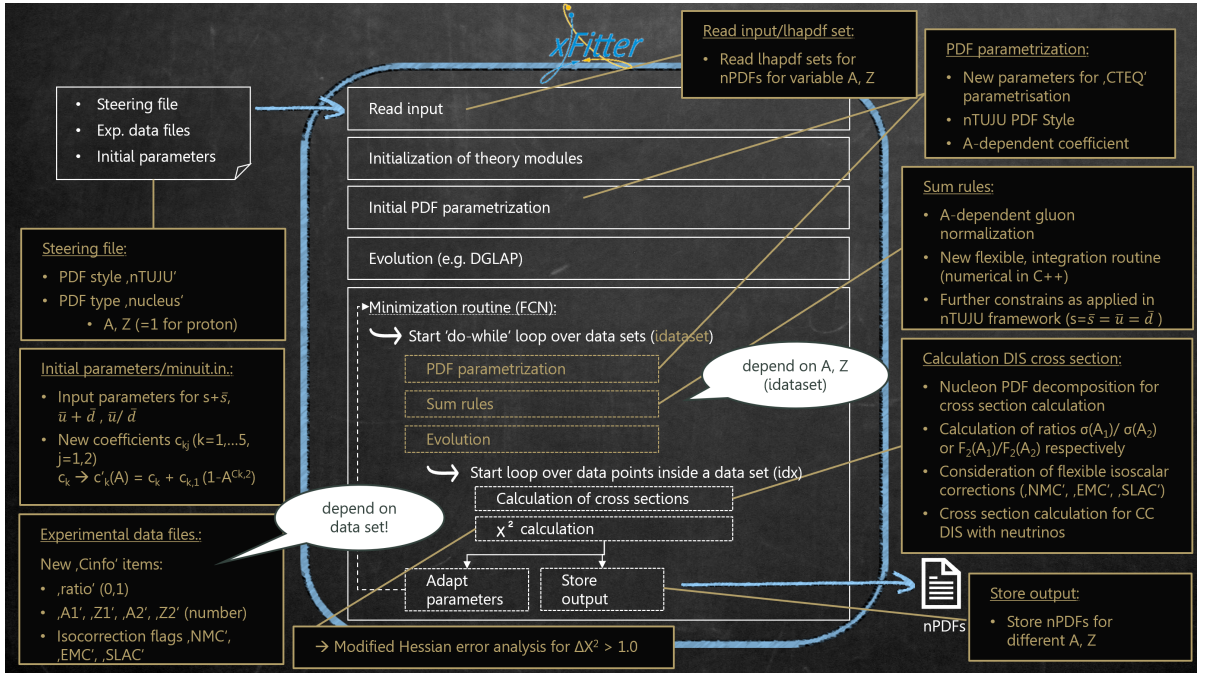


Figure 5.3.: A graphical overview of modifications of the xFITTER program code performed for the purpose of nuclear PDF analysis. The implemented extensions are summarized in black boxes with the golden frame. For a schematic view of the original high-level xFITTER functionalities please refer to figure 5.2. xFITTER logo credited from [189].

positive definite, which is as close as possible to the actual Hessian matrix \hat{H} . The positive definiteness of the Hessian matrix relies on the second derivatives, which is a difficult computation and is numerically often approximate. One reason is that the minimized function χ^2 is not exact, but given by a second order polynomial in the space of the fit parameters. Thus, if some fit parameters are not well constrained by the data, higher-order polynomial terms of χ^2 might become relevant [202]. Another point is that the function χ^2 might not be as smooth as necessary due to the limited numerical precision at which the DGLAP equations are solved and due to the finite accuracy of the integrals.

All modifications of the xFITTER program code performed for the purpose of nuclear PDF analysis are summarized in figure 5.3 for a graphical overview. The extensions of the code described in this section have been published providing a first open-source tool to analyse nuclear PDFs. For that purpose, a short user guide has been created which can be integrated into the overall xFITTER manual, and is provided in appendix B for reference purpose.

5.3. Summary: Proton vs. Nucleon

In this section the differences and the common aspects of the free proton PDFs versus the nuclear PDFs of a nucleon bound in a nucleus are summarized. An overview of some selected key criteria with respect to the theoretical basics or in regard to the framework is provided in table 5.2.

Common to both analyses are the theoretical basics, like the calculation of partonic cross sections (here: for DIS processes), the universality of PDFs and the collinear factorization theorem. Though, the power correction terms to the form of the factorization theorem given in equation (3.23) might be potentially different for a free proton compared to a bound nucleon, as described in section 3.4. However, these power corrections are neglected in this analysis since Q^2 is assumed to be sufficiently high.

Furthermore, the approximations and constraints selected in this work, e.g. parton flavour decomposition, sum rules, as well as heavy quarks mass scheme, are common to

free proton PDF and nPDF analyses. Also the applied kinematic cuts have been chosen likewise. The differences appear in regard to nuclear nature of nPDFs. For instance, the initial parameterization of nPDFs includes parameters dependent on the nuclear mass number A . Furthermore, a nucleon consisting of protons and neutrons has to be constructed by the usage of isospin symmetry. The main difference results from the nuclear effects which are absorbed in the nuclear PDFs.

As a small anticipation to the next two chapters we also mention, that the observables (e.g. cross section σ or structure function F_2) used in the nPDF analysis are often provided by the experiments in form of ratios measured for two different nuclei with mass numbers A_1 , A_2 . Furthermore, isoscalar modifications are sometimes applied on the measured data and need to be taken into account during the global analysis. For further details please refer to section 6.2. Finally, as described and argued in section 7.1, the tolerance criterion $\Delta\chi^2$, chosen for the Hessian uncertainty analysis performed as part of this work, is different for the free proton PDFs compared to the nPDFs of a bound nucleon.

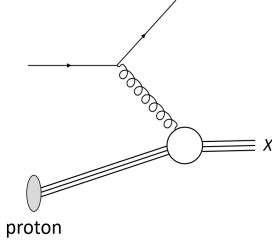
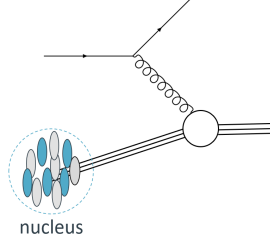
	Free proton PDF	Bound nucleon nPDF	
			cf. fig. 3.4
Nuclear effects	—	✓	sec. 3.1
Factorization theorem	✓	✓ (different power corrections)	sec. 3.4
Universality	✓	✓	sec. 3.4
Parameterization	eq. (5.1)	dependent on A , i.e. eqs. (5.1) & (5.2) & (5.3)	sec. 5.1.1
Approximations	$\bar{u} = \bar{d} = s = \bar{s}$		sec. 5.1.2
	sum rules	sum rules, satisfied by each nucleon per nucleus	eqs. (5.4), (5.5), (5.6)
	—	isospin symmetry	
	heavy quarks mass scheme FONLL		
Kinematic cuts	$Q^2 > 3.5 \text{ GeV}^2$, $x < 0.7$, and $W^2 > 12 \text{ GeV}^2$		sec. 3.2
Isoscalar modifications	—	✓	sec. 6.2
Observables	absolute observables (e.g. σ , F_2)	absolute observables (e.g. $\sigma(A)$, $F_2(A)$) or ratios of observables (e.g. $\sigma(A_1)/\sigma(A_2)$, $F_2(A_1)/F_2(A_2)$)	sec. 6.2
Uncertainty analysis	$\Delta\chi^2 = 20$	$\Delta\chi^2 = 50$	sec. 7.1

Table 5.2.: Overview of some selected criteria, with respect to the theoretical basics or in regard to the framework, compared for the free proton PDFs versus the nPDFs of a nucleon bound in a nucleus with the mass number A . The compared criteria are given in the left column. The reference in this manuscript, i.e. section, equation, figure, is provided in the right column.

6. Study I: Comparison to DIS Data

The comparison between the calculated theoretical predictions and the incorporated experimental data controls the quality of the fit and constrains the free PDF parameters. The obtained results depend on two steps in the overall fitting procedure (cf. fig. 6.1) - calculation of the cross sections as described in section 3.2 and the processing of the incorporated experimental data. The results of this comparison are presented and discussed in this chapter. Also, the selection of suitable experimental data is one of the many choices to make during a global PDF analysis. Furthermore, the results of a study specific to the neutrino-nucleus scattering experiments are evaluated in section 6.4. Therefore, some important aspects of the selections applied for this work are described in this chapter.

Note: sections 6.1-6.3 are largely based on the own publication [1] for which the author has written the original draft of the manuscript (except introduction). The author has worked on the concept creation, numerical implementation and execution of the calculations for publication [1]. The experimental proton data applied for the analysis were used as provided by xFITTER developers team, whereas the experimental nuclear data have been collected from different public sources and transferred into the proper format by the author. All numerical results presented in [1] have been obtained by the programs largely modified by the author.

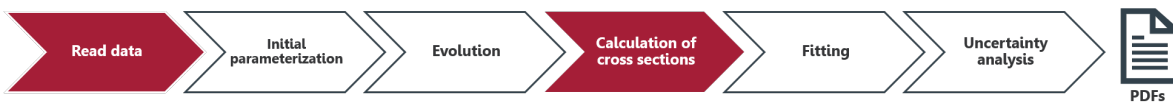


Figure 6.1.: Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: the experimental data and the comparison to the theoretically calculated cross sections are documented.

6.1. Experimental DIS Data

The QCD analyses presented here have been performed by including the experimental data from DIS measurements. The free proton baseline was fitted using data from HERA [169], BCDMS [204] and NMC [204] experiments, as listed in table 6.1. The fixed-target DIS data with lepton and neutrino beams used to determine the nuclear parton distribution functions are summarized in table 6.2. For the used neutrino data we would like to emphasize that the measured cross sections, rather than structure functions, were used in this analysis. The advantage compared to the isospin-averaged structure functions F_2 and F_3 , e.g. as utilized in DSSZ [205], is that the sensitivity to the flavour decomposition is retained in the cross sections. Another approach was used in EPPS16 [188] where normalized (anti-)neutrino cross sections were considered. This increases the sensitivity to the shape of nuclear modifications. In order to extract complete, i.e. without isospin-averaging, information from the incorporated neutrino data sets, the absolute cross sections are exploited here. The choice of neutrino-nucleus DIS data is discussed in section 6.4. The quality requirements and kinematic cuts for the selected experimental NC DIS and CC DIS data are summarized in the following.

Exp.	Dataset	Year	Ref.	N_{dp}	χ^2 NLO	χ^2 NNLO
BCDMS	F2p 100GeV	1996	[203]	83	88.88	90.98
	F2p 120GeV			90	69.97	67.75
	F2p 200GeV			79	89.46	85.91
	F2p 280GeV			75	66.97	68.73
HERA 1+2	NCep 920	2015	[169]	377	455.15	475.14
	NCep 820			70	72.47	73.84
	NCep 575			254	225.24	228.97
	NCep 460			204	223.23	223.95
	NCem			159	233.55	229.42
	CCep			39	42.19	44.41
	CCem			42	65.94	68.99
NMC-97	NCep	1997	[204]	100	124.56	111.64
In total:				1559	1845.99	1909.08

Table 6.1.: Summary of experimental DIS data used to determine proton PDFs. In the last two columns the resulting χ^2 values at NLO and NNLO obtained in our analysis are provided. The overall agreement is found to be sufficiently good, and the total χ^2/N_{dp} is 1.184 at NLO and 1.225 at NNLO. Table from [1].

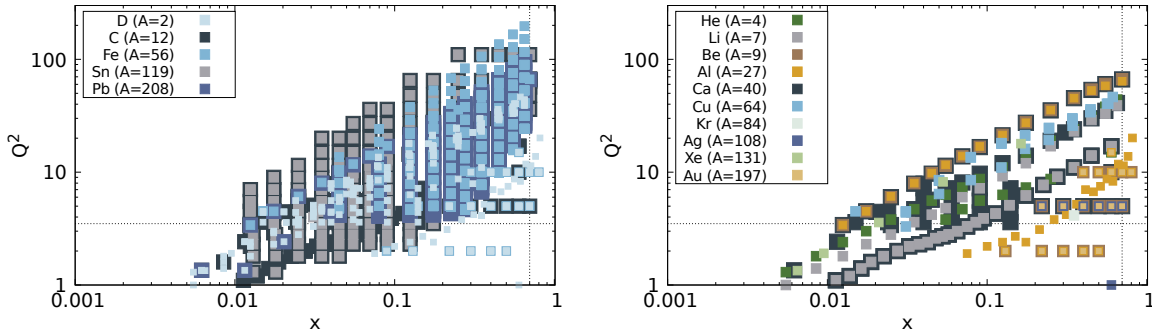


Figure 6.2.: Kinematic reach of experimental DIS data in the (x, Q^2) plane used to constrain the nuclear PDFs. Figure from [1].

At first, the factorization of the partonic scattering process and the non-perturbative PDFs, as well as the perturbative QCD framework, described in chapter 3, are valid at sufficiently high energy scales. For the measurements listed in tables 6.1 and 6.2, there are experimental data points published that lie outside this validity region. Therefore, we had to exclude such data points from the analysis by the application of so-called kinematic cuts. In this work we have selected the kinematic cuts $Q^2 > 3.5 \text{ GeV}^2$, $x < 0.7$, and $W^2 > 12 \text{ GeV}^2$. The kinematic reach of the included experimental DIS data in the (x, Q^2) plane is shown in figure 6.2 for the different nuclear targets, with applied kinematic cuts illustrated by the dotted line.

As demonstrated in figure 6.2, the number of available data points varies for different nuclei. A large number of data points are available for deuteron (D) and the heavier nuclei of carbon (C), iron (Fe), tin (Sn) and lead (Pb), as shown in the left panel of figure 6.2. These data points are provided either in the form of absolute cross sections, or as ratios where D is usually used as the reference (denominator). Also calcium (Ca) has been intensively used in the relevant experiments. For the other nuclei (right panel in figure 6.2), only a few data points are available. Therefore the nuclei predominantly present in the included data are expected to be better constrained than the nuclei with fewer data points.

Eventually, as can be seen from equation (3.14), the knowledge of all kinematic variables x, y, Q^2 is required to calculate the differential DIS cross section. Therefore, only mea-

surements which published these variables could be considered for the analysis. Nonetheless, some of the experimental data sets do not specify the invariant y , but when the center-of-mass energy \sqrt{s} is known, it can be derived from the relation (3.15). However, for the experimental SLAC data [5, 211, 216] neither y nor \sqrt{s} is known. Thus, the quantity ϵ provided by the experiment is used to derive the required information as described in appendix C.

6.2. Corrections and Experimental Uncertainties

6.2.1. Isoscalar Modifications

Some experimental analyses of charged-lepton DIS have modified the measured structure functions to achieve isospin-symmetry also for non-symmetric nuclei such as iron or lead. According to the relations summarized in Ref. [188], an *isoscalar* structure function of a nucleus with the mass number A is defined as

$$\hat{F}_2^A \equiv \frac{1}{2} F_2^{p,A} + \frac{1}{2} F_2^{n,A} \quad (6.1)$$

with $F_2^{p,A}$ and $F_2^{n,A}$ representing the structure functions of the bound protons and neutrons. By definition, the isoscalar structure function contains an equal number of protons and neutrons, which holds only for specific nuclei. A *general* structure function for a nucleus with Z protons and $N = A - Z$ neutrons can be written as

$$F_2^A = \frac{Z}{A} F_2^{p,A} + \frac{N}{A} F_2^{n,A}, \quad (6.2)$$

which is *not* isoscalar if $Z \neq N \neq A/2$. The relation between the isoscalar structure function (6.1) and the general structure function (6.2) is given by

$$\hat{F}_2^A = \beta F_2^A, \quad (6.3)$$

where

$$\beta = \frac{A}{2} \left(1 + \frac{F_2^{n,A}}{F_2^{p,A}} \right) / \left(Z + N \frac{F_2^{n,A}}{F_2^{p,A}} \right). \quad (6.4)$$

Furthermore, it is assumed that the ratio $F_2^{n,A}/F_2^{p,A}$ for any nucleus is unchanged compared to that for unbound nucleons, so that the relation

$$\frac{F_2^{n,A}}{F_2^{p,A}} = \frac{F_2^n}{F_2^p} \quad (6.5)$$

can be used in (6.4) to describe the isoscalar modifications. The ratio F_2^n/F_2^p for the isoscalar "correction" is parameterized in a different way by each experiment,

- EMC [60]:

$$\frac{F_2^n}{F_2^p} = 0.92 - 0.86x \quad (6.6)$$

- SLAC [5]:

$$\frac{F_2^n}{F_2^p} = 1 - 0.8x \quad (6.7)$$

- NMC [217]:

$$\frac{F_2^n}{F_2^p} = A(x) (Q^2/20 \text{ GeV}^2)^{B(x)} (1 + x^2 \text{ GeV}^2/Q^2) \quad (6.8)$$

with $A(x) = 0.979 - 1.692x + 2.979x^2 - 4.313x^3 + 3.075x^4$
and $B(x) = -0.171x + 0.244x^2$.

A graphical visualisation of the ratio F_2^n/F_2^p for the different experiments as summarized above is provided in figure 6.3. Common to all parameterizations selected by the different collaborations is that firstly the ratio incorporated into the isoscalar modifications decreases with increasing x values up to $x \approx 0.6$. After that, the ratio applied by SLAC and EMC collaborations continue to decrease linearly, whereas the ratio selected by the NMC collaboration starts to grow rapidly. And also, only the form chosen by the NMC collaboration has an additional dependency on Q^2 . Consequently, a common physical law cannot be recognized from the form of these isoscalar "corrections" so that it does not seem to be a fundamental manifestation of any effect, but rather a pure choice made during the data analysis. Therefore, it is definitely consistent and necessary to apply the same, individual form of the isoscalar modification on the calculated cross sections in order to equalize the quantities which are compared to each other. In this work the general form of the structure function (6.2) is used to calculate the theoretical predictions. In case isoscalar modifications were applied to the measured quantities, for

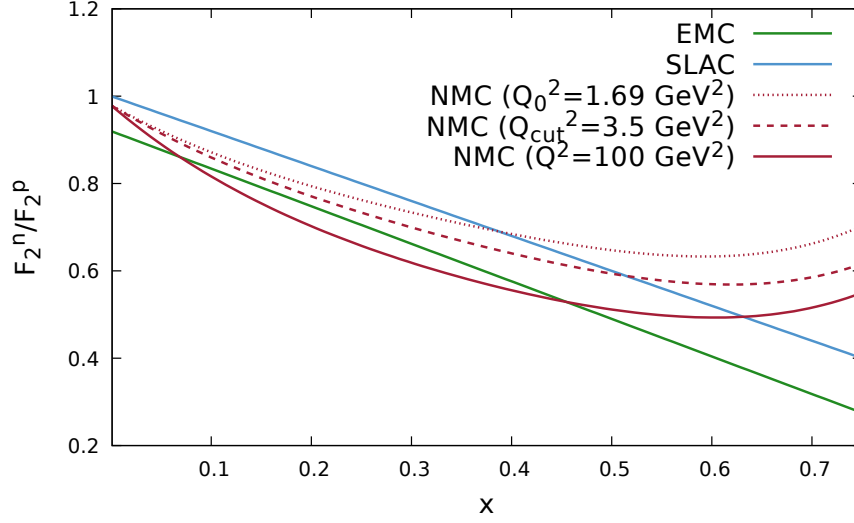


Figure 6.3.: Ratio F_2^n/F_2^p along x axis as parameterized by different experiments for the isoscalar modifications applied on the measured data. The form of these isoscalar modifications used by the NMC collaboration additionally depends on Q^2 , whereas it is constant in Q^2 for SLAC and EMC. Thus, the NMC F_2^n/F_2^p ratio is shown for some sample Q^2 values: the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ used in this work, the value of $Q_{\text{cut}}^2 = 3.5 \text{ GeV}^2$ at which kinematic cuts have been applied as part of the performed PDF analysis, and $Q^2 = 100 \text{ GeV}^2$.

consistency the same modifications are applied to the obtained theoretical results by using relation (6.3). However, for the future measurements a good proof of quality could be established if additionally also the unmodified experimental data are published.

6.2.2. Correlated Uncertainties

Some of the experiments provide normalization uncertainties on top of the systematic and statistical errors. In this work normalization uncertainties have been treated as *correlated* errors as discussed in [200, 201, 218]. The correlated uncertainties are treated as implemented in xFITTER, described further in refs. [218–220]. The same procedure applies if any *overall* uncertainties are provided in addition to the point-to-point uncertainties, e.g. for the SLAC data [5, 211, 216].

6.3. Results NC and CC DIS

The optimal set of nPDF parameters is derived by minimizing the χ^2 as defined in eq. (5.7) by comparing to the measured data presented in table 6.2. The resulting cross sections, structure functions and ratios are compared to the data we used in figures 6.5–6.11 for neutral current DIS processes and in figures 6.12 and 6.13 for charged current DIS processes with neutrinos for a subset of that data. The graphs demonstrate the central results having the best agreement with the data and the according error bands, obtained during the uncertainty analysis described in chapter 7.

An overview of the resulting χ^2 values, divided by the number of data points N_{dp} , is shown in figure 6.4 for NLO and NNLO. Values above $\chi^2/N_{\text{dp}} > 3.0$ have been truncated in this graph for better representation, but the actual numbers are shown in table 6.2. Figure 6.4 demonstrates that the agreement between the theoretical predictions and the experimental measurements varies between different data sets. For example, the agreement for most of the data published by the NMC collaboration is excellent, whereas the agreement to the HERMES data is clearly not optimal. In this particular example one needs to point out that the number of data points by the HERMES experiment is much smaller than the number of NMC data points, so that the contribution to the

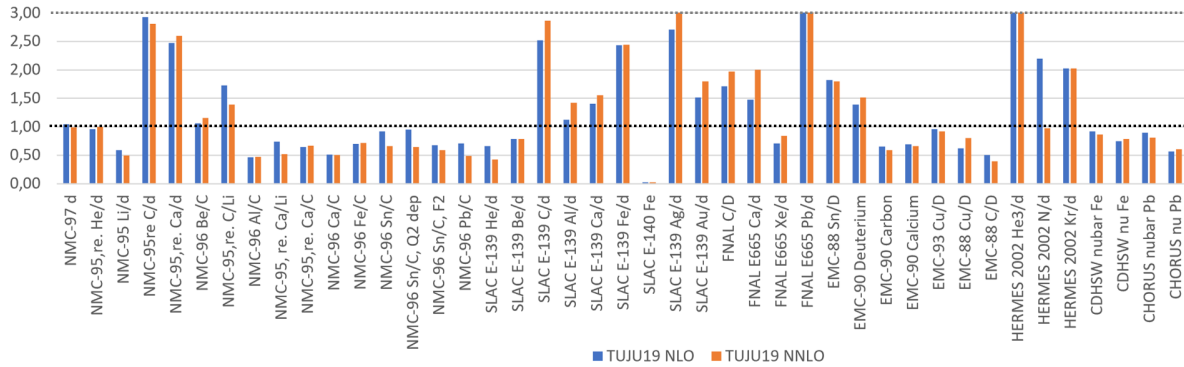


Figure 6.4.: Comparison of χ^2 values divided by the individual number of data points per dataset N_{dp} at NLO and NNLO. The ”ideal” value $\chi^2/N_{\text{dp}} = 1.0$ is marked by the horizontal black dotted line. The bars in the diagram corresponding to $\chi^2/N_{\text{dp}} > 3.0$ have been truncated for the purpose of a clearer representation, which is symbolised by the dashed light-grey line. Figure from [1].

6. Study I: Comparison to DIS Data

total χ^2 is relatively small for the HERMES data. Apart from the few outlying data sets, the overall agreement is found to be very good, and the total χ^2/N_{dp} is 0.887 at NLO and 0.862 at NNLO. Even though some of the data sets are better described at NLO and some at NNLO, the total χ^2 values are very close at the different orders. The good agreement is apparent also in figures 6.5–6.13. Interestingly, a very good agreement is also achieved for the neutrino data, even though some earlier studies observed difficulties when incorporating these data in a global nPDF analysis [221]. However, as concluded in Ref. [222], this likely follows due to the tension caused by the NuTeV data which we have not included as discussed in section 6.4.

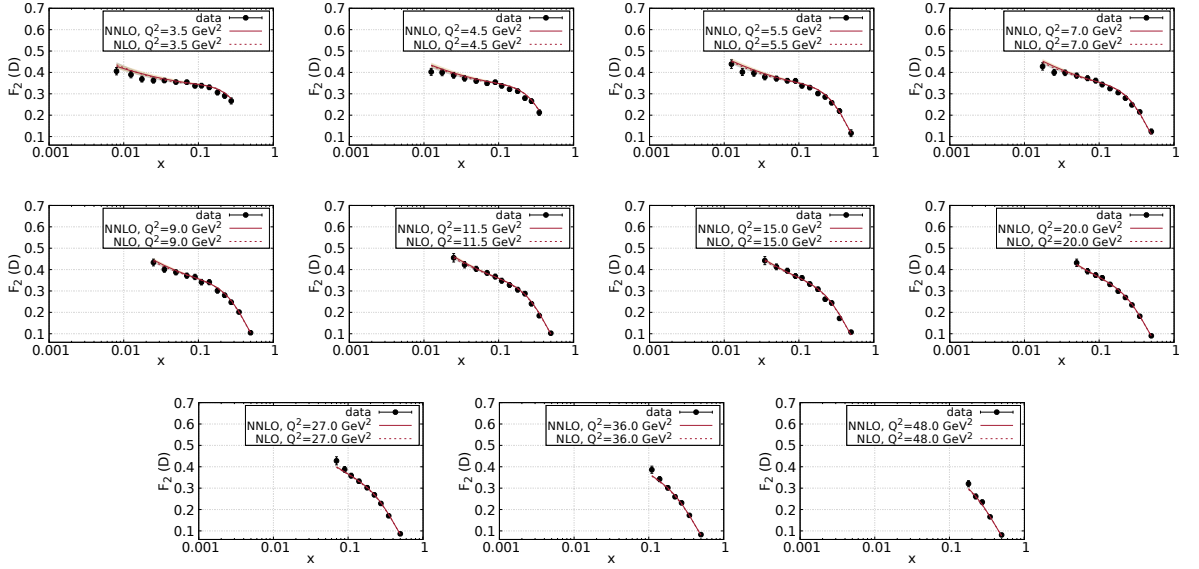


Figure 6.5.: Comparison to NMC $F_2(D)$ data at different values of Q^2 at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

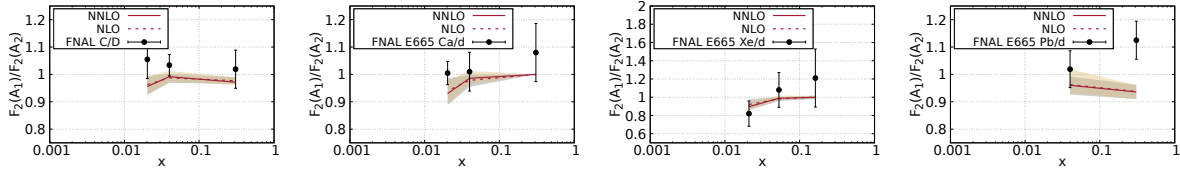


Figure 6.6.: Comparison to FNAL E665 data for different ratios $F_2(A_1)/F_2(A_2)$ for nuclei with mass numbers A_1 and A_2 , at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

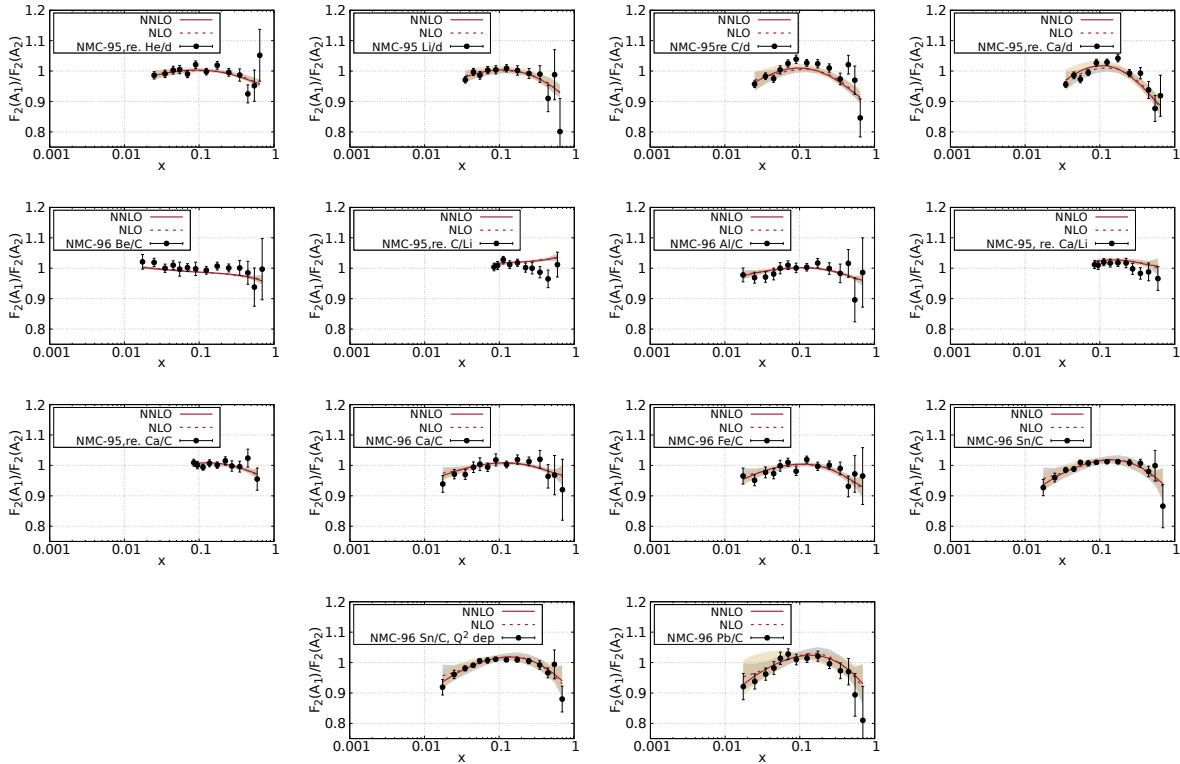


Figure 6.7.: Comparison to NMC $F_2(A_1)/F_2(A_2)$ data measured for different combinations of nuclei with mass numbers A_1 and A_2 , at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

6. Study I: Comparison to DIS Data

Nucleus	Exp.	Year	Ref.	N_{dp}	χ^2 NLO	χ^2 NNLO
D	NMC 97	1996	[204]	120	124.85	118.66
	EMC 90	1989	[206]	21	29.23	31.73
He/D	HERMES	2002	[207]	7	54.64	37.99
	NMC 95, re.	1995	[6]	13	12.44	12.98
	SLAC E139	1994	[5]	11	7.21	4.68
Li/D	NMC 95	1995	[208]	12	7.06	5.93
Be/D	SLAC E139	1994	[5]	10	7.84	7.83
Be/C	NMC 96	1996	[209]	14	14.80	16.19
C	EMC 90	1989	[206]	17	11.01	10.05
C/D	FNAL E665	1995	[4]	3	5.12	5.91
	SLAC E139	1994	[5]	6	15.12	17.16
	EMC 88	1988	[210]	9	4.49	3.50
	NMC 95, re.	1995	[6]	13	38.08	36.52
C/Li	NMC 95, re.	1995	[6]	10	17.27	13.90
N/D	HERMES	2002	[207]	1	2.20	0.97
Al/D	SLAC E139	1994	[5]	10	11.20	14.22
Al/C	NMC 96	1996	[209]	14	6.51	6.55
Ca	EMC 90	1989	[206]	19	13.17	12.56
Ca/D	NMC 95, re.	1995	[6]	12	29.61	31.12
	FNAL E665	1995	[4]	3	4431	6.01
	SLAC E139	1994	[5]	6	8.44	9.34
Ca/Li	NMC 95, re.	1995	[6]	10	7.36	5.16
Ca/C	NMC 95, re.	1995	[6]	10	6.47	6.70
	NMC 96	1996	[209]	14	7.14	6.99
Fe	SLAC E140	1993	[211]	2	0.05	0.05
Fe/D	SLAC E139	1994	[5]	14	34.08	34.18
Fe/C	NMC 96	1996	[209]	14	9.82	9.96
ν Fe	CDHSW	1991	[212]	464	347.74	365.14
$\bar{\nu}$ Fe	CDHSW	1991	[212]	462	423.06	398.25
Cu/D	EMC 93	1993	[60]	19	18.12	17.45
	EMC 88	1988	[210]	9	5.59	7.22
Kr/D	HERMES	2002	[207]	1	2.02	2.02
Ag/D	SLAC E139	1994	[5]	6	16.24	18.81
Sn/D	EMC 88	1988	[210]	8	14.56	9.24
Sn/C	NMC 96	1996	[209]	14	12.90	7.61
	NMC 96, Q^2 dep.	1996	[213]	134	94.7	79.85
Xe/D	FNAL E665	1992	[214]	3	2.13	2.53
Au/D	SLAC E139	1994	[5]	11	16.64	19.80
Pb/D	FNAL E665	1995	[4]	2	12.24	13.32
Pb/C	NMC 96	1996	[209]	14	9.94	6.77
ν Pb	CHORUS	2005	[215]	405	229.11	243.85
$\bar{\nu}$ Pb	CHORUS	2005	[215]	405	361.35	328.28
In total:				2336	2072.29	2014.02

Table 6.2.: Summary of experimental DIS data used to determine the nuclear PDFs. In the last two columns the resulting χ^2 values at NLO and NNLO obtained in our analysis are provided. Table from [1].

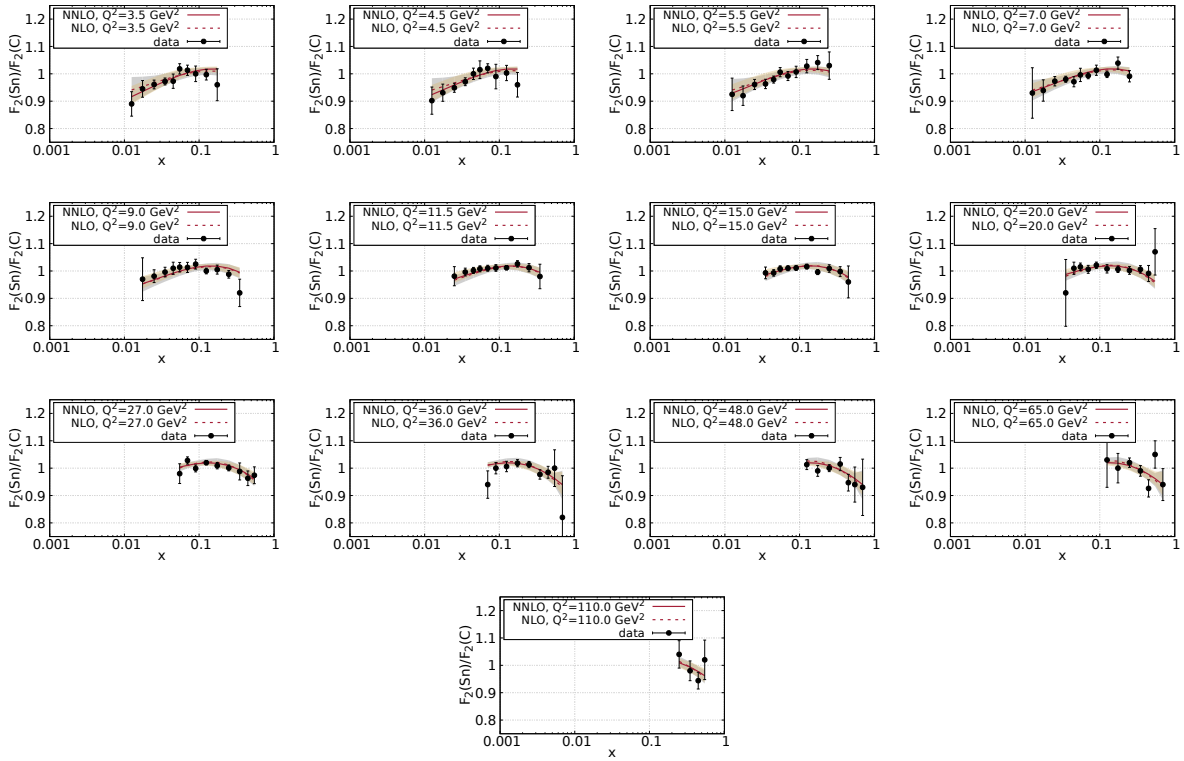


Figure 6.8.: Comparison to NMC $F_2(\text{Sn})/F_2(\text{C})$ data at different values of Q^2 at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

6. Study I: Comparison to DIS Data

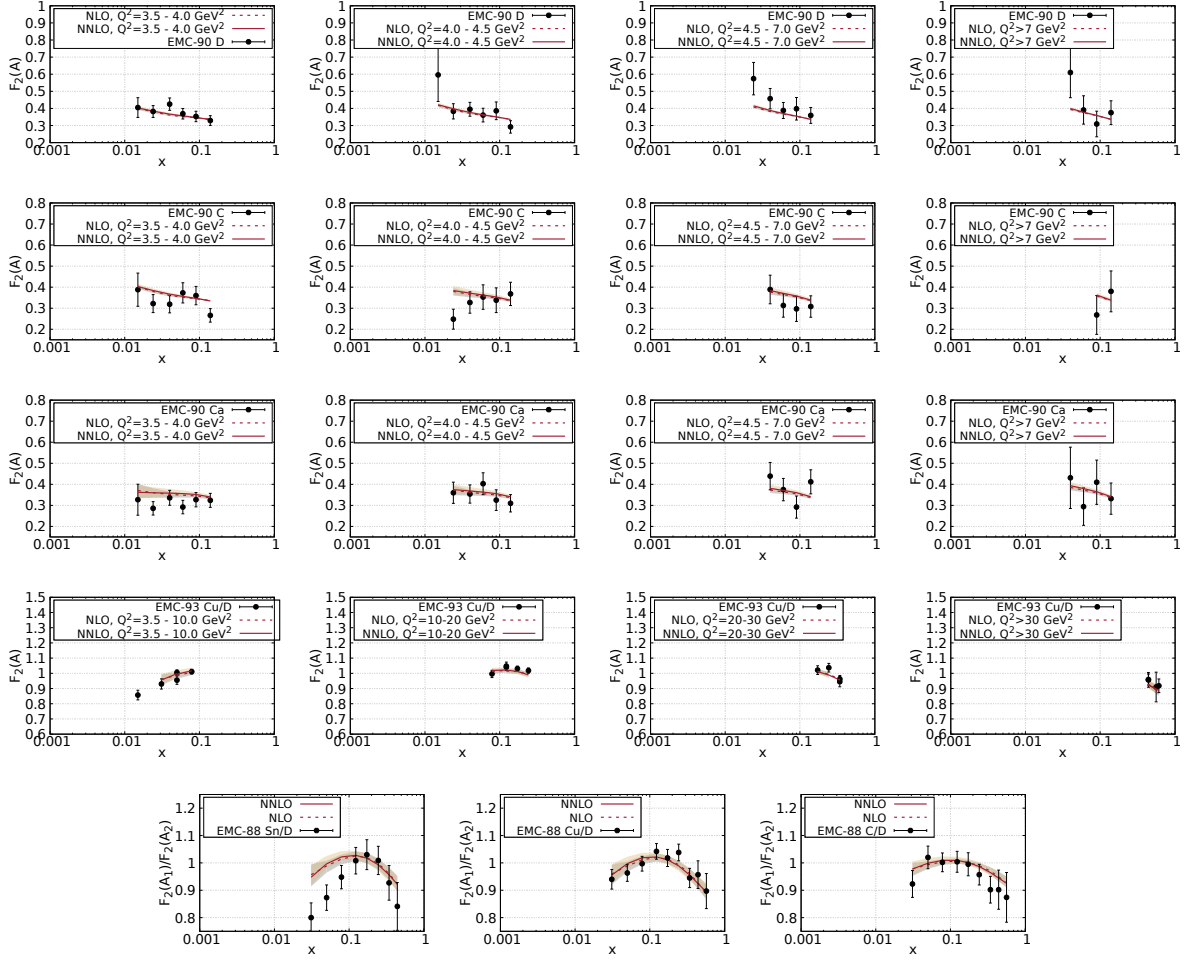


Figure 6.9.: Comparison to EMC data, first for the structure function F_2 at different Q^2 , and then for different ratios $F_2(A_1)/F_2(A_2)$ measured for nuclei with mass numbers A_1 and A_2 . The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

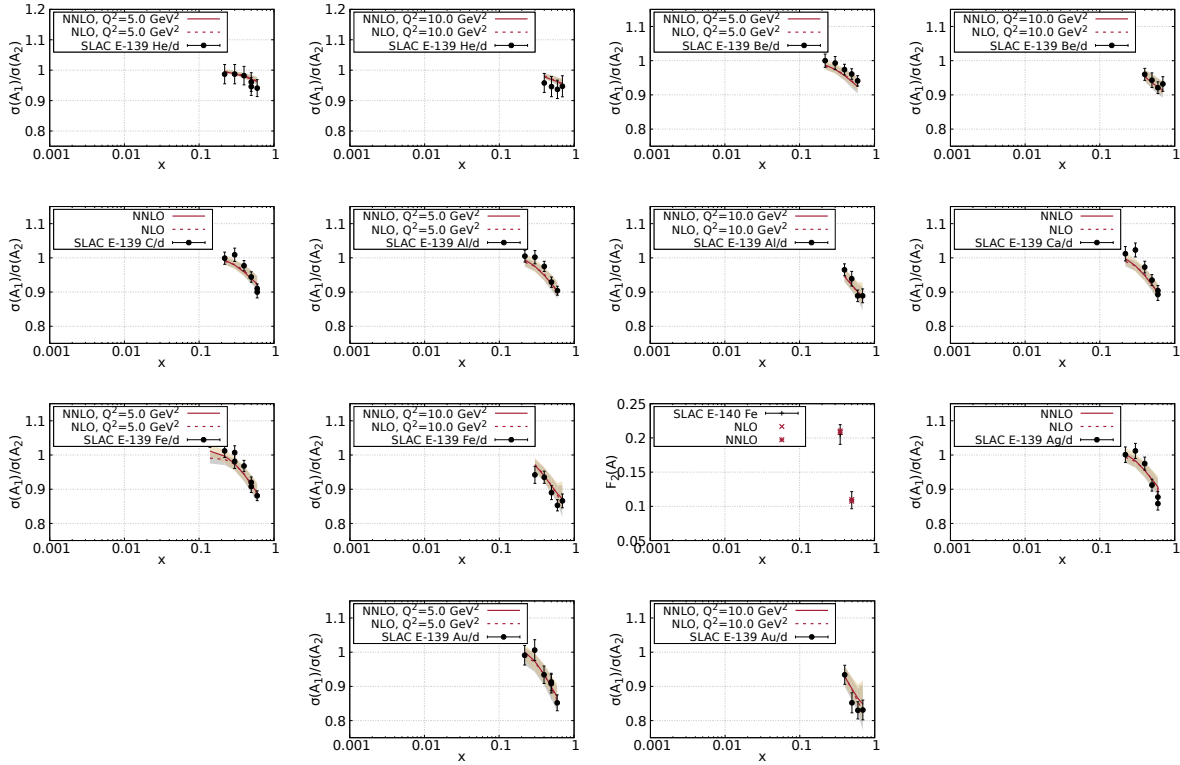


Figure 6.10.: Comparison to SLAC data for different ratios of reduced differential cross sections $\sigma(A_1)/\sigma(A_2)$ for nuclei with mass numbers A_1 and A_2 , at different values of Q^2 at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

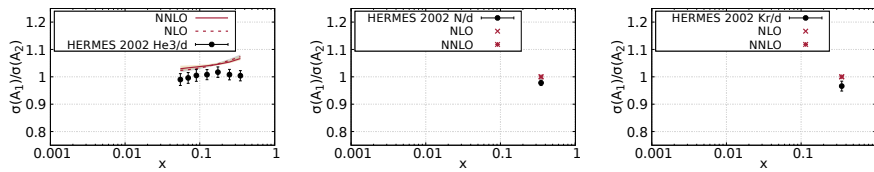


Figure 6.11.: Comparison to HERMES data for different ratios of reduced differential cross sections $\sigma(A_1)/\sigma(A_2)$ for nuclei with mass numbers A_1 and A_2 , at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

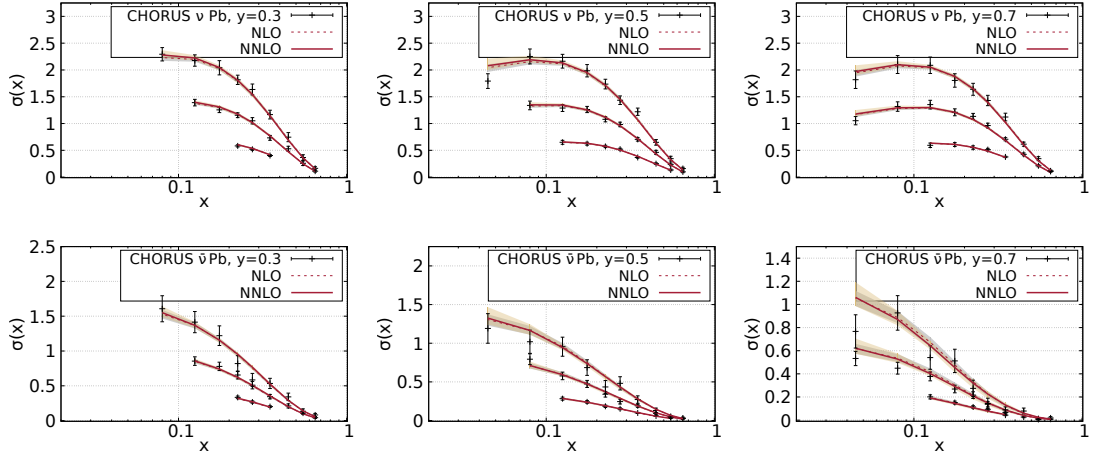


Figure 6.12.: Sample comparison to selected CHORUS data for CC deeply inelastic scattering on lead (Pb). We show the results for either neutrinos (ν) or anti-neutrinos ($\bar{\nu}$), for one y value (cf. legend) each at different beam energies (35, 70, 110 GeV). The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

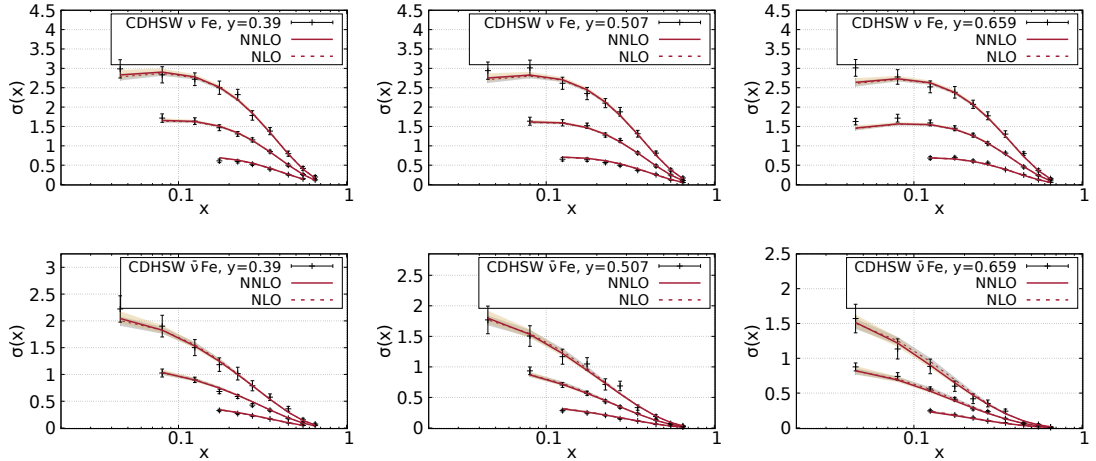


Figure 6.13.: Sample comparison to selected CDHSW data for CC deeply inelastic scattering on iron (Fe). We show the results for either neutrinos (ν) or anti-neutrinos ($\bar{\nu}$), for one y value (cf. legend) each at different beam energies (38.9, 85.4, 144.3 GeV). The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].

6.4. Study of Neutrino-Nucleus DIS

For the results presented in section 6.3, neutrino data from the CDHSW νFe experiment [212] and the CHORUS νPb experiment [215] were included. In addition there are more neutrino scattering data available, e.g. measured cross sections with Fe target by the NuTeV collaboration [223], and also data from the CCFRR collaboration [224]. The data from the CCFRR experiment were excluded from our analysis for two reasons. Firstly, the quantities x and Q^2 , required for the analysis procedure, were not publicly available for the cross sections. Secondly, only the averaged structure functions F_2 and F_3 for νFe and $\bar{\nu}\text{Fe}$ were available, which lose the sensitivity to flavour decomposition.

In regard to NuTeV data, an early study in Ref. [225] found that these data could be accommodated together with the CHORUS neutrino DIS data when constraining d/u ratio but with the applied nuclear corrections some pull against other DIS and DY data was observed. Later on, the analyses in refs. [226, 227] found some unresolved tension between the NuTeV neutrino DIS data and lepton-nucleus data. In a following work, a similar tension was found also when taking into account neutrino DIS data from CHORUS and CCFRR experiments in Ref. [221]. Simultaneously, a study presented in Ref. [228] concluded that the pull against other data was specifically due to the NuTeV data at certain energies, whereas CDHSW and CHORUS data were well compatible with the existing lepton-nucleus DIS data. This tension was further studied in Ref. [222] where again the NuTeV data were found incompatible with the other considered data. Only by normalizing the differential data with the integrated cross section at each energy bin an acceptable agreement was achieved. Due to the demonstrated tension, we have not included the NuTeV neutrino DIS data in the published TUJU19 analysis. However, as part of this work, the NuTeV data have been included in a separate fit to study the reported tension. The according results and insights are presented here.

First of all, the comparison between the calculated cross sections, computed as part of the fit including NuTeV data, and the measured NuTeV data is presented in figure 6.14 for some selected data samples. The lowest energy shown there corresponds to the lowermost line. There, the agreement between the data points and the predictions is reasonable. When going to higher energies some deviations appear and the data uncertainties grow due to the limited statistics. For instance, one can see that there is

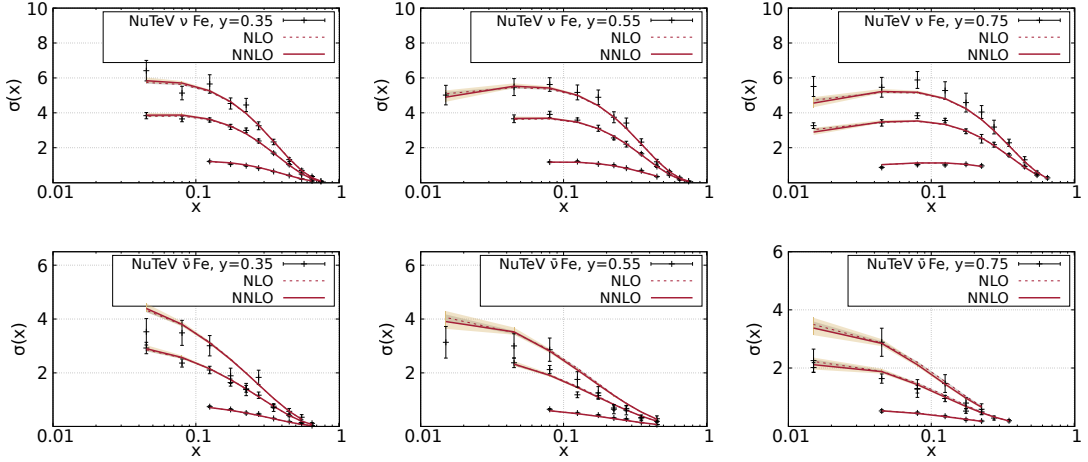


Figure 6.14.: Sample comparison to selected NuTeV data for CC deeply inelastic scattering on iron (Fe). We show the results for either neutrinos (ν) or anti-neutrinos ($\bar{\nu}$), for one y value (cf. legend) each at different beam energies (65, 190, 275 GeV). The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands).

a fluctuation in the data collected at high energies and the measured shape sometimes (e.g. for $y = 0.35$) does not follow any smooth trend, nor the predicted form. At the same time it becomes apparent that the data points published by the NuTeV collaboration are available at much higher energies (up to $E \approx 300$ GeV) than those published by CDHSW and CHORUS ($E < 200$ GeV). Therefore, when saying that NuTeV data is incompatible with the other considered data, it could be specific to the data points corresponding to the high energies. Although, one cannot identify if this observation is something related to the extraction process of the data or has another origin.

Next, we study the obtained χ^2/N_{dp} values to discuss if there are hints for a tension between NuTeV and other neutrino data sets, or neutrino-nucleus DIS data and lepton-nucleus DIS data in general. For that purpose we perform three fits:

- A)** one fit that corresponds to the setting of the published nPDFs, including NCDIS data, as well as CHORUS and CDHSW neutrino data (referred to as TUJU19),
- B)** one with lepton-nucleus DIS data only (referred to as NCDIS),
- C)** and one with lepton-nucleus DIS data and all neutrino data matching our requirements, i.e. CHORUS, CDHSW and NuTeV (referred to as NuTeV).

The obtained χ^2/N_{dp} values are summarized in figure 6.15 at NLO and figure 6.16 at NNLO. Columns 'A', 'B' and 'C' show the χ^2 values divided by the individual number of data points per dataset N_{dp} . The blue bar indicates the size of the number which corresponds to the representation shown in figure 6.4. The columns '=B-A' and '=C-A' demonstrate the differences observed between the additional fits compared to the reference analysis TUJU19. There, the overview is supported by the background colours of the rows - green colour signalizes that the result is better than in TUJU19, and red stands for a worse outcome.

When evaluating the results in figures 6.15 and 6.16 we first compare TUJU19 analysis to the fit with lepton-nucleus NCDIS data only. Column '=B-A' shows that the conclusion, whether there is an improvement or not, varies between different data sets. Also, one needs to keep in mind that the number of data points differs between the considered data sets, so that the contribution to the total χ^2 is relatively small for the NCDIS data compared to the neutrino data. Though, the overall agreement is found to be very good in the reference TUJU19 fit. As can be concluded from figures 6.15 and 6.16 the quality of the fit is very similar between TUJU19 and NCDIS.

Now, when looking at the comparison between TUJU19 fit and NuTeV fit in column '=C-A' several χ^2/N_{dp} values obtained for the lepton-nucleus data sets show a worse agreement, and only few cases are described better. For the CHORUS and CDHSW neutrino data sets included in TUJU19 analysis the impact is marginal. Even though the number of NuTeV data points (2121 in total) clearly dominates the contribution to the overall χ^2 value, the obtained $\chi_{\text{NuTeV}}^2/N_{\text{dp}}$ values are a bit worse than those of the other two neutrino experiments. However, this is consistent with the above discussion for the comparison between the computed cross sections and the experimental measurements for the NuTeV data. Therefore we conclude that lepton-nucleus and neutrino-nucleus DIS data can be accommodated in a global analysis without creating tensions between different sets, apart from the discussed points related to the NuTeV data. However, the inclusion of the measurements by the NuTeV collaboration does not seem to provide an additional benefit but leaves some doubts in regard to the bad agreement between the theoretical predictions and the experimental data. In all cases presented in figures 6.15 and 6.16 the NLO and NNLO analyses provide very similar agreement.

6. Study I: Comparison to DIS Data

Dataset	N_{dp}	A	B	=B-A	C	=C-A
		TUJU19 NLO	NCDIS NLO		NuTeV NLO	
NMC-97 d	120	1,04	1,02	-0,02	1,11	0,07
NMC-95, re, He/d	13	0,96	0,76	-0,20	1,17	0,21
NMC-95 Li/d	12	0,59	0,77	0,18	0,95	0,36
NMC-95 re C/d	13	2,93	1,23	-1,70	4,70	1,77
NMC-95, re, Ca/d	12	2,47	1,01	-1,46	4,80	2,33
NMC-96 Be/C	14	1,06	0,82	-0,24	0,95	-0,11
NMC-95, re, C/Li	10	1,73	1,37	-0,36	1,68	-0,04
NMC-96 Al/C	14	0,47	0,36	-0,10	0,56	0,09
NMC-95, re, Ca/Li	10	0,74	0,44	-0,30	0,77	0,03
NMC-95, re, Ca/C	10	0,65	0,73	0,08	0,77	0,12
NMC-96 Ca/C	14	0,51	0,40	-0,11	0,71	0,20
NMC-96 Fe/C	14	0,70	0,65	-0,06	0,84	0,14
NMC-96 Sn/C	14	0,92	0,33	-0,59	1,23	0,31
NMC-96 Sn/C , Q ² dep.	14	0,95	0,31	-0,64	1,30	0,35
NMC-96 Sn/C , F2	120	0,68	0,53	-0,15	0,74	0,06
NMC-96 Pb/C	14	0,71	0,69	-0,02	1,48	0,77
SLAC E-139 He/d	11	0,66	0,30	-0,36	0,73	0,07
SLAC E-139 Be/d	10	0,78	0,78	-0,01	1,05	0,27
SLAC E-139 C/d	6	2,52	1,42	-1,10	3,95	1,43
SLAC E-139 Al/d	10	1,12	0,80	-0,33	2,31	1,19
SLAC E-139 Ca/d	6	1,41	1,12	-0,29	3,04	1,64
SLAC E-139 Fe/d	14	2,43	1,33	-1,11	4,27	1,84
SLAC E-140 Fe	2	0,03	0,10	0,08	0,05	0,02
SLAC E-139 Ag/d	6	2,71	1,27	-1,43	4,53	1,82
SLAC E-139 Au/d	11	1,51	1,21	-0,30	2,73	1,21
FNAL C/D	3	1,71	2,22	0,51	1,38	-0,33
FNAL E665 Ca/d	3	1,48	3,02	1,55	0,81	-0,67
FNAL E665 Xe/d	3	0,71	0,84	0,13	0,75	0,04
FNAL E665 Pb/d	2	6,12	5,08	-1,04	6,26	0,14
EMC-88 Sn/D	8	1,82	1,92	0,10	1,91	0,09
EMC-90 Deuterium	21	1,39	1,61	0,22	1,23	-0,17
EMC-90 Carbon	17	0,65	0,55	-0,10	0,73	0,09
EMC-90 Calcium	19	0,69	0,63	-0,06	0,77	0,07
EMC-93 Cu/D	19	0,95	0,52	-0,43	1,80	0,85
EMC-88 Cu/D	9	0,62	0,40	-0,22	1,00	0,37
EMC-88 C/d	9	0,50	0,65	0,15	0,46	-0,04
HERMES 2002 He3/d	7	7,81	4,29	-3,51	7,45	-0,36
HERMES 2002 N/d	1	2,20	0,74	-1,46	3,10	0,90
HERMES 2002 Kr/d	1	2,02	2,02	0,00	2,02	0,00
CHORUS nu Pb	405	0,57			0,60	0,04
CHORUS nubar Pb	405	0,89			0,87	-0,02
CDHSW nu Fe	464	0,75			0,71	-0,04
CDHSW nubar Fe	462	0,92			0,81	-0,11
NuTeV nu Fe	958				1,01	
NuTeV nubar Fe	1163				1,29	
TOTAL χ^2		0,887	0,905		1,043	

Figure 6.15.: Comparison of χ^2 values divided by the individual number of data points per dataset N_{dp} at NLO. The obtained results are shown for the published nPDF set TUJU19 (column 'A'), but also for the separately performed fits with NC DIS data only (column 'B') and the one including NuTeV data (column 'C'). The blue bar visualizes the χ^2/N_{dp} value per data set. The columns '=B-A' and '=C-A' demonstrate the differences observed between the additional fits compared to the originally published results, where green background colour of the row means that the result was better, and red stands for a worse outcome.

6.4. Study of Neutrino-Nucleus DIS

Dataset	N _{dp}	A		B		C	
		TUJU19 NNLO	NCDIS NNLO	=B-A	NuTeV NNLO	=C-A	
NMC-97 d	120	0,99	0,93	-0,06	0,98	0,00	
NMC-95, re, He/d	13	1,00	0,82	-0,18	1,22	0,22	
NMC-95 Li/d	12	0,49	0,54	0,05	0,79	0,30	
NMC-95 re C/d	13	2,81	1,39	-1,42	4,24	1,43	
NMC-95, re, Ca/d	12	2,59	1,38	-1,22	4,96	2,37	
NMC-96 Be/C	14	1,16	1,07	-0,09	1,12	-0,04	
NMC-95, re, C/Li	10	1,39	1,21	-0,19	1,30	-0,09	
NMC-96 Al/C	14	0,47	0,40	-0,07	0,57	0,10	
NMC-95, re, Ca/Li	10	0,52	0,29	-0,23	0,47	-0,05	
NMC-95, re, Ca/C	10	0,67	0,75	0,08	0,80	0,13	
NMC-96 Ca/C	14	0,50	0,42	-0,08	0,73	0,23	
NMC-96 Fe/C	14	0,71	0,69	-0,02	0,83	0,11	
NMC-96 Sn/C	14	0,66	0,35	-0,31	0,84	0,18	
NMC-96 Sn/C , Q ² dep.	14	0,65	0,32	-0,33	0,97	0,33	
NMC-96 Sn/C , F2	120	0,59	0,53	-0,06	0,63	0,04	
NMC-96 Pb/C	14	0,48	0,71	0,23	0,72	0,24	
SLAC E-139 He/d	11	0,43	0,23	-0,20	0,53	0,10	
SLAC E-139 Be/d	10	0,78	0,77	-0,02	1,32	0,54	
SLAC E-139 C/d	6	2,86	1,34	-1,52	4,95	2,09	
SLAC E-139 Al/d	10	1,42	0,90	-0,53	3,44	2,02	
SLAC E-139 Ca/d	6	1,56	0,97	-0,59	3,76	2,20	
SLAC E-139 Fe/d	14	2,44	1,25	-1,19	5,04	2,60	
SLAC E-140 Fe	2	0,03	0,23	0,21	0,01	-0,02	
SLAC E-139 Ag/d	6	3,14	1,25	-1,89	5,58	2,45	
SLAC E-139 Au/d	11	1,80	1,06	-0,74	3,83	2,03	
FNAL C/D	3	1,97	2,59	0,62	1,72	-0,25	
FNAL E665 Ca/d	3	2,00	3,68	1,68	1,34	-0,67	
FNAL E665 Xe/d	3	0,84	1,13	0,29	0,72	-0,13	
FNAL E665 Pb/d	2	6,66	5,47	-1,20	6,98	0,32	
EMC-88 Sn/D	8	1,79	1,85	0,06	1,90	0,10	
EMC-90 Deuterium	21	1,51	1,81	0,30	1,38	-0,13	
EMC-90 Carbon	17	0,59	0,52	-0,07	0,65	0,05	
EMC-90 Calcium	19	0,66	0,66	-0,01	0,69	0,03	
EMC-93 Cu/D	19	0,92	0,59	-0,33	1,63	0,71	
EMC-88 Cu/D	9	0,80	0,48	-0,33	1,26	0,46	
EMC-88 C/d	9	0,39	0,53	0,14	0,34	-0,05	
HERMES 2002 He3/d	7	5,43	4,03	-1,40	5,65	0,22	
HERMES 2002 N/d	1	0,97	0,16	-0,81	1,82	0,85	
HERMES 2002 Kr/d	1	2,02	2,02	0,00	2,02	0,00	
CHORUS nu Pb	405	0,60			0,68	0,07	
CHORUS nubar Pb	405	0,81			0,82	0,01	
CDHSW nu Fe	464	0,79			0,73	-0,06	
CDHSW nubar Fe	462	0,86			0,74	-0,13	
NuTeV nu Fe	958				1,05		
NuTeV nubar Fe	1163				1,29		
TOTAL χ^2		0,862	0,908		1,045		

Figure 6.16.: Same as for figure 6.15, but at NNLO

7. Study II: Uncertainty Analysis

During the fitting procedure of the PDF analysis a central set of distribution functions is obtained first. The central set of PDFs is the set from the best fit, having the best agreement with the data. As described in the previous chapters, the experimental data are provided with given statistical, systematic and correlated uncertainties, like e.g. normalization errors. Therefore, a variance from the central PDF fit needs to be analysed to reflect the fact that there is a certain flexibility within the given experimental error bands. Or more precise, an error analysis needs to be performed to study how well the experimental data actually constrain the fitted distributions. The distribution functions resulting from the uncertainty analysis are provided as part of the PDF set.

There are two established methods which can be used for the error analysis, the Hessian method [200, 201] or the Monte Carlo (MC) method [229–231]. The Hessian method is based on the approximation that the fitted function has a quadratic dependence on the varied parameter in the region small enough from the minimum. The user has to select an allowed deviation from the minimum $\Delta\chi^2$. This method is using the Hessian error matrix and its eigenvalues in order to obtain the variations of distribution functions. The form of the parameterization (sec. 5.1.1) may bias the resulting uncertainty bands. Alternatively, the Monte Carlo method can be used to analyse the uncertainties. In this case, the data is varied within the given uncertainties, and for each variation a PDF replica set is fitted. The resulting PDF set consists of several PDF replicas. This method is supposed to be less biased in regard to the allowed deviation, but requires



Figure 7.1.: Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: uncertainty analysis.

the execution of a time-consuming fit for each replica set. The quality of the obtained uncertainty bands is limited by the amount of the performed PDF replicas. Both methods have been studied as part of this work, although the published PDF sets TUJU19 are based on the Hessian approach. For the reason of completeness we mention that also Lagrange multiplier method [201, 218] has been used in some works to analyse the uncertainties (see e.g. [232]). However, in this work only Hessian and MC methods have been studied. Therefore, further details of these methods are described in the following sections.

Note: since the published PDF sets TUJU19 are based on the Hessian approach, large parts of the text in section 7.1 have been credited from the own publication [1], for which the author has written the original draft of the manuscript (except introduction). The author has worked on the concept creation, numerical implementation and execution of the calculations for publication [1]. All numerical results presented in [1] have been obtained by the programs largely modified by the author. The author has also prepared the released sets of nuclear parton distribution functions TUJU19 in a standard format.

7.1. Hessian Method

The Hessian error analysis is performed assuming a quadratic expansion of the function $\chi^2 = \chi_0^2 + \Delta\chi^2$ around its global minimum. Here, χ_0^2 is the value of the function at the global minimum and $\Delta\chi^2$ is the displacement from the minimum [200, 201]. The Hessian matrix \hat{H} is constructed by the second derivatives of χ^2 at the minimum. The matrix elements H_{ij} are defined as

$$H_{ij} = \frac{1}{2} \left(\frac{\delta^2 \chi^2}{\delta y_i \delta y_j} \right), \quad (7.1)$$

with y_i being the displacement of the parameter a_i from its value a_0 at the minimum. For the analysed function χ^2 one writes

$$\chi^2 = \chi_0^2 + \sum_{i,j} H_{ij} y_i y_j. \quad (7.2)$$

The Hessian matrix is symmetric and thus has a complete set of orthonormal eigenvectors v_{ij} . The eigenvectors and the eigenvalues ϵ_j of the Hessian matrix are used to transform the displacements y_i into a new set of parameters z_i

$$y_i = \sum_j v_{ij} \sqrt{\frac{1}{\epsilon_j}} z_j, \quad (7.3)$$

leading to a simplified relation

$$\Delta\chi^2 = \chi^2 - \chi_0^2 = \sum_i z_i^2. \quad (7.4)$$

This representation has the advantage that the surfaces of constant χ^2 are spheres in z_i space with $\Delta\chi^2$ being the squared distance from the minimum. The varied parameters a_i from which the resulting error sets are defined can be written as

$$a_i = a_0 \pm \Delta a_i = a_0 \pm \Delta\chi^2 \sum_j \frac{v_{ij}^2}{\epsilon_j}, \quad (7.5)$$

where $\Delta\chi^2$ defines the tolerance criterion determining the allowed growth of χ^2 . The resulting error bands for three different values of $\Delta\chi^2$ are shown in figure 7.3 for the proton baseline at NLO for demonstrative purpose. The relation in equation (7.5) shows that the parameters which correspond to the eigenvectors of the Hessian matrix with large eigenvalues are well determined since their Δa_i is small, whereas the weakly determined parameters correspond to small eigenvalues.

The uncertainties for a given observable X can be calculated via [188, 233]

$$(\Delta X^\pm)^2 = \sum_{i=1}^{n_{\text{param}}} \left\{ \max_{\min} \left[X(S_i^+) - X(S_0), X(S_i^-) - X(S_0), 0 \right] \right\}^2, \quad (7.6)$$

where $X(S_0)$ is the observable calculated with the central parameter set and the S_i^\pm correspond to the error sets in the positive and negative direction determined from the diagonalized parameter z_i , with n_{param} number of the fit parameters.

The nPDFs obtained in this analysis are published considering the nPDF uncertainties for 16 fit parameters and the free proton uncertainties for 13 fit parameters. In total, there is one central set S_0 and 58 error sets S_j ($j = 1, \dots, 58$) per nucleus. Thus, applying equation (7.6) on our framework the error bands are given by

$$(\Delta X^\pm)^2 = \sum_{j=1}^{29} \left\{ \max_{\min} \left[X(S_{2j-1}) - X(S_0), X(S_{2j}) - X(S_0), 0 \right] \right\}^2, \quad (7.7)$$

with $n_{\text{param}} \equiv n_{\text{param,nucleus}} + n_{\text{param,proton}} = 16 + 13 = 29$.

In an ideal case one would choose the tolerance criterion so that $\Delta\chi^2 = 1$. However, since we consider several data sets which are not necessarily in a mutual agreement with one another, such a choice would underestimate the underlying uncertainty. In this work the tolerance for $\Delta\chi^2$ is based on the statistically motivated method as discussed e.g. in refs. [188, 234]. For the proton baseline with 13 free fit parameters (cf. tab. A.1, A.2) it becomes $\Delta\chi^2 = 20$ with 90% confidence level. This choice has been validated also by comparing to the error bands generated with the MC method, though the $\Delta\chi^2$ value preferred by the MC method is quite flavour- and kinematics-dependent, as discussed in section 7.3. Previous nPDF studies have shown that such a statistically motivated method would not fully cover the experimental uncertainties in the nuclear data [170, 188]. Thus for the nuclear PDF error analysis we increase the tolerance from the statistically motivated value and choose $\Delta\chi^2 = 50$ for our 16 free parameters (cf. tab. A.1, A.2).

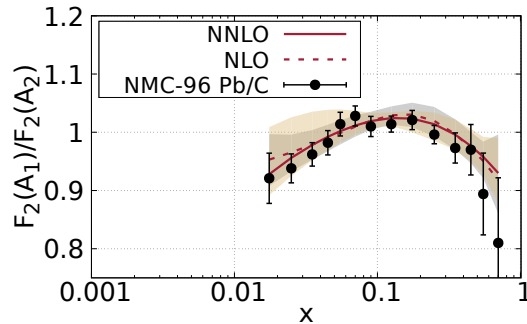


Figure 7.2.: Sample DIS data set for the ratio of the structure functions $F_2(Pb)/F_2(C)$ measured by the NMC collaboration. The actual data, as well as the best-fit results at NLO and NNLO including **Hessian** error bands from [1] are shown here for reference purpose.

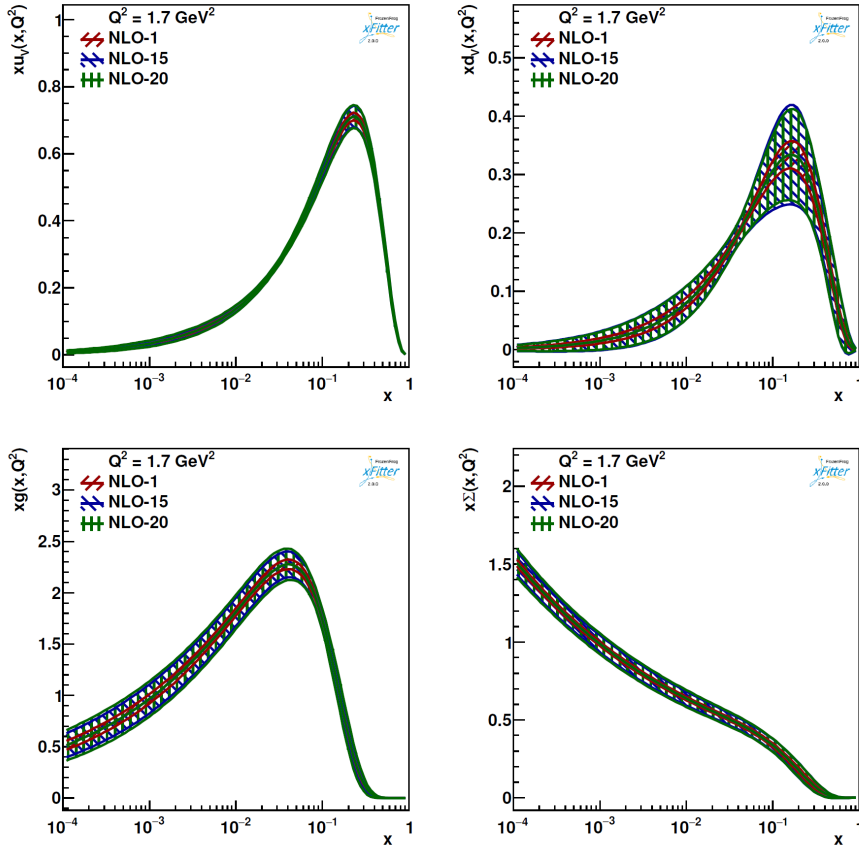


Figure 7.3.: Comparison of proton baseline PDFs at NLO with error bands obtained by the Hessian uncertainty analysis with different values of $\Delta\chi^2$ (1, 15, 20), shown for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$ at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. As expected, due to the quadratic nature of the tolerance criterion the error bands for $\Delta\chi^2 = 15$ and $\Delta\chi^2 = 20$ are very close to each other. With the method described above, for the proton baseline with 13 free fit parameters we select $\Delta\chi^2 = 20$ with 90% confidence level, $\Delta\chi^2 = 15$ would correspond to 68% confidence level.

7.2. Monte Carlo Method

The Monte Carlo (MC) uncertainty analysis is based on the idea to use so-called MC replicated data which are not measured in a real experiment, but are generated artificially by varying the actual data points randomly within their uncertainty bars. Figure 7.4(a)

shows the actual experimental data compared to one set of MC replicated data. It needs to be emphasized that multiple MC replicas are required to be used as a basis for the new fits delivering new results. The combination of the best fit and all obtained MC replicas is used to create a full PDF set with MC uncertainty bands.

Usually, the actual, measured data points are used to be varied in order to generate a new, MC data set. Technically, there is also a possibility to use the theoretical predictions (calculated cross sections) to be varied within the order of magnitude of the given experimental uncertainties. In xFITTER this option is controlled by setting 'IRANDDATA = True' (for data) or 'IRANDDATA=False' (for theory). For actual studies, of course, the actual experimental data are supposed to be used to generate new MC replicated data sets. However, in cases where pseudo data or theoretical predictions from other analyses are used instead of the data, the option 'IRANDDATA=False' can be used for enhanced studies, e.g. like impact of (simulated) EIC data. Furthermore, the existence of this option provides a possibility to get deeper insights into the Monte Carlo method. Therefore, this option has been studied and is described here.

For demonstrative purpose one sample DIS data set for the ratio of the structure functions $F_2(Pb)/F_2(C)$ will be used for the following discussion. For this data set, the actual data, as well as the best-fit results at NLO and NNLO including Hessian error bands are shown in figure 7.2 for reference purpose.

Obviously, if the measured data and the calculated cross sections coincide, the choice which one to be varied is redundant and the 'IRANDDATA' should not impact the resulting Monte Carlo uncertainties. In the normal case, i.e. 'IRANDDATA=True' the experimental data are used to be varied within the experimental uncertainties. Therefore, the resulting, varied MC replicated data point is expected to lie mainly within the error bars (cf. figure 7.4(a)). However, if 'IRANDDATA=False' is set, and so the theoretical values are varied, two cases can appear in respect to the generated data points. First, if the calculated cross section lies completely outside of the error bands (causing $\chi^2 > 1$), it is likely to happen that the MC replicated data point will also lie outside the actual error bands. Second, whereas the error bars in one direction will be exceeded, the other direction is most probably to be underestimated. Both cases lead to

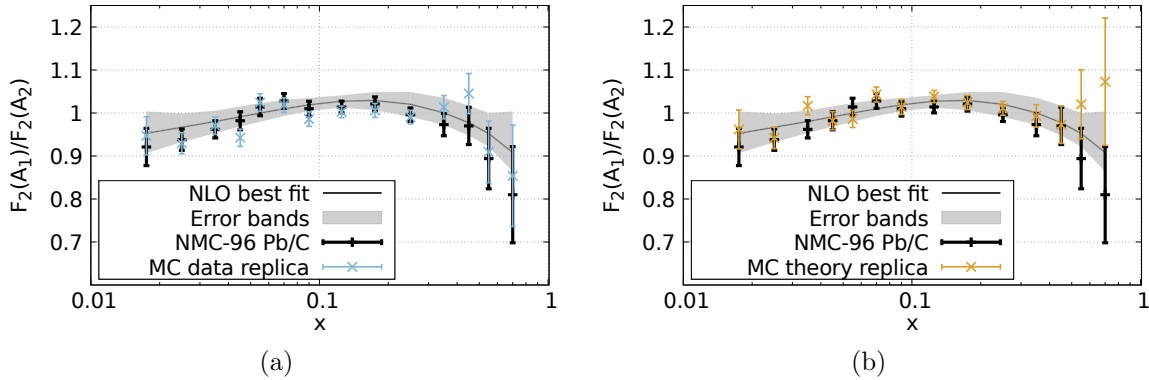


Figure 7.4.: **Monte Carlo** replicated data shown with the actual data points with error bars (black) and the best fit NLO result (solid line) with corresponding error bands (grey) obtained by the Hessian method. In 7.4(a) the actual data points (light-blue) have been varied. In 7.4(b) the calculated theoretical predictions (copper-coloured) have been varied.

a Monte Carlo data set which is not representing the actual data set properly. An example of MC data points generated based on the theoretical values is shown in figure 7.4(b).

As expected, some data points generated with 'IRANDDATA=False' lie outside the original error bars (cf. 7.4(b)). In the small- x region (e.g. third data point from the left) it might effectively pretend a reduction of the shadowing effect. Whereas the coexistent anti-shadowing behaviour remains unaffected. In parallel, the deviation of the first data point on the right-hand side from the allowed uncertainty region, might potentially lead to a suppression or even reversion of the EMC-effect. So, especially for the ratios of cross sections or structure functions with large uncertainties, the usage of 'IRANDDATA=False' might lead to anomalous MC data sets. Though, this could be another study to validate the constraining power of a certain data set. For example, if the data (e.g. from a DIS measurement) are claimed *not* to be sensitive to the sea quarks or gluon distributions, the resulting PDFs and the corresponding error bands obtained for the sea quarks and gluons should remain the same despite the fact which starting position has been used.

Another option having the impact on the quality of the MC error method is the choice of the scaling behaviour for the experimental uncertainties as described in section 5.1.3.

Typically, the scaling properties should be consistent with the chosen form of χ^2 , in this work given in equation (5.9). Nonetheless, in principle this choice can be varied. For an overview of the different combinations of the available scaling properties and the corresponding results (error bands), please refer to appendix D.

7.3. Results: Hessian vs. MC Method

7.3.1. Comparison for Proton Baseline

As described in section 7.1, for the proton baseline with 13 free fit parameters the selected tolerance value for the performed Hessian uncertainty analysis becomes $\Delta\chi^2 = 20$ with 90% confidence level based on the statistically motivated method. As shown in figure 7.3, the error bands for $\Delta\chi^2 = 15$ and $\Delta\chi^2 = 20$ are very close to each other. With the method described above, $\Delta\chi^2 = 15$ would correspond to 68% confidence level. Since the uncertainty bands obtained with $\Delta\chi^2 = 15$ and $\Delta\chi^2 = 20$ are very similar, the value of $\Delta\chi^2 = 15$ is selected here to demonstrate the comparison of our choice to the error bands generated with the MC method, shown in figures 7.5 – 7.8.

As can be seen in figures 7.5 – 7.8, the Hessian error bands with a choice $\Delta\chi^2 = 1$ would clearly underestimate proton PDF uncertainties compared to the Monte Carlo method, which is in line with the statistically motivated method. However, going towards large $\Delta\chi^2$ the result becomes flavour- and kinematics-dependent. For example, at NLO MC method would prefer $\Delta\chi^2 > 15$ for u_v quark distribution at initial and higher scales, whereas for the d_v also values $\Delta\chi^2 > 4$ seem to be comparable to the MC error bands. At NNLO the MC error analysis suggests uncertainty bands which are comparable to the Hessian results with values $\Delta\chi^2 > 15$ for both valence quark flavours. At both pQCD orders, NLO and NNLO, the conclusion for gluon and sea quarks is not that clear. There, high $\Delta\chi^2$ values seem to overestimate the uncertainty bands compared to MC results. However, the PDFs of this pQCD analysis have been published with the error bands obtained with the Hessian method. The comparison presented here was performed to study the agreement between the two approaches and the impact of our choice of $\Delta\chi^2$, though the $\Delta\chi^2$ value preferred by the MC method is quite flavour- and kinematics-dependent, as shown in figures 7.5-7.8.

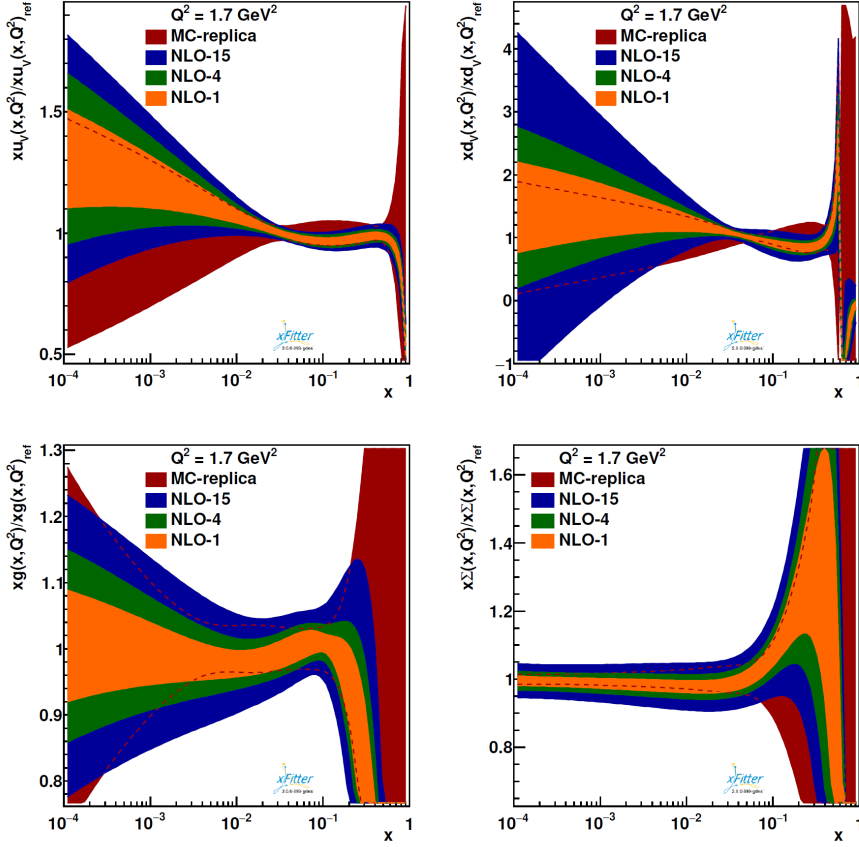
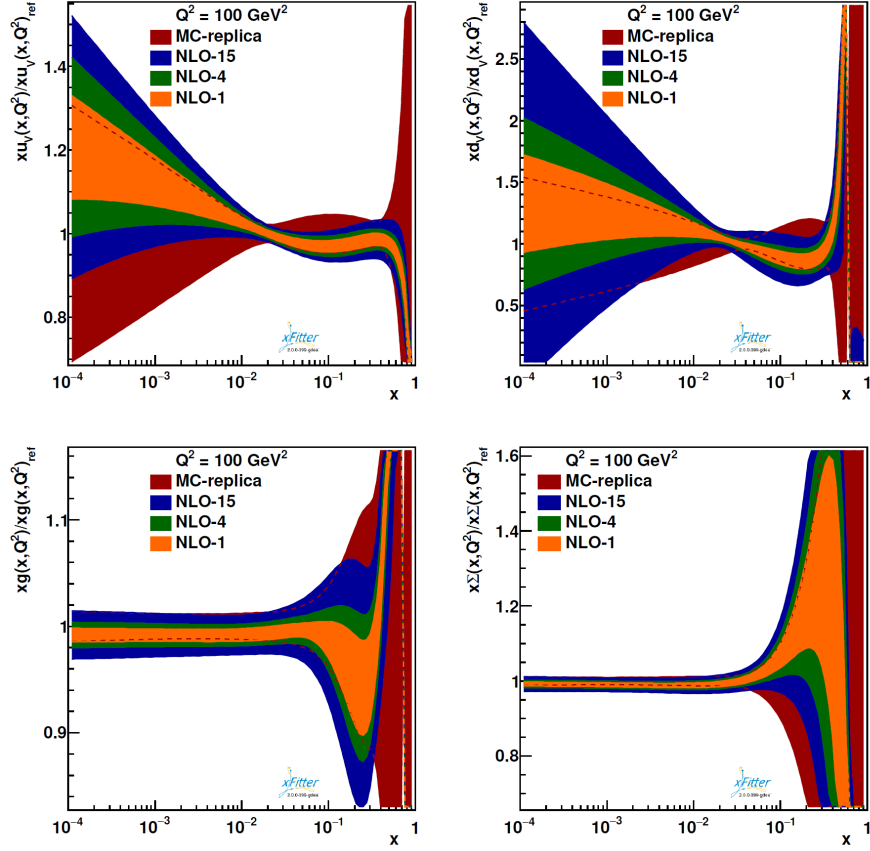


Figure 7.5.: Comparison of error bands for the proton baseline PDFs at NLO computed based on different error methods at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. The ratios $xf(x, Q^2)/xf(x, Q^2)_{ref}$ are shown for the error bands obtained by the Hessian uncertainty analysis with different values of $\Delta\chi^2$ (1, 4, 15) compared to the uncertainty bands obtained by the Monte Carlo method ('ref'). The representation is provided for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$.

7.3.2. Comparison for nPDFs

The nuclear parton distribution functions obtained as part of this work have been published with the error sets computed based on the Hessian uncertainty analysis with the statistically motivated tolerance value $\Delta\chi^2 = 50$ as described above. The main results are presented in chapter 8. The discussion here is provided for demonstrative purpose. For the comparison of the uncertainty bands for nPDFs generated with Hessian method vs. MC method 100 converged MC replicas were used. As mentioned above, the MC method requires a lot of high-performance computing resources. Additionally, since

Figure 7.6.: Same as for figure 7.5, but at the higher scale $Q^2 = 100 \text{ GeV}^2$.

many replicas need to be created the generated output files (i.e. pdfs in LHAPDF6 format) allocate a lot of storage space. For example, for 20 different nuclei and 100 MC replicas one needs 8-10 GB storage space, each at NLO and NNLO. Other nPDF analyses, e.g. nNNPDF1.0 [235], published their nPDF LHAPDF sets with 200 MC replicas, whereas they claim that those 200 published error set members were selected from the overall amount of 1000 created MC replicas. In order to reproduce that particular quality at least 100 GB storage space would be required for each order of pQCD. Therefore, in this demonstrative study only 100 converged MC replicas were used which implicates somewhat limited statistics and potentially limited validity of the obtained MC error bands. When running the fits with the different MC data replicas for the nuclear analysis approximately 40% of the fits failed due to numerical limitations of the optimization procedure and only 60% converged, that is the estimated distance to minimum was smaller than the tolerance criterion. However, for the preparation of the

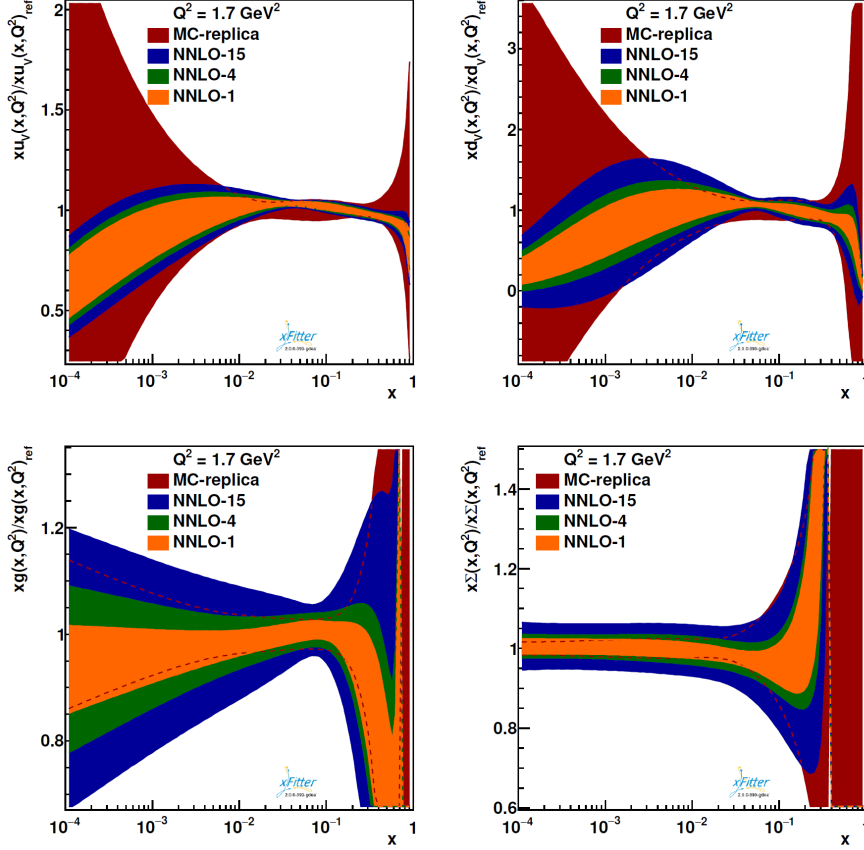


Figure 7.7.: Same as for figure 7.5, but at NNLO.

figures 7.9 - 7.12 only converged MC replicas were used.

Our results at NLO are shown in figures 7.9, 7.10 and the NNLO ones in figures 7.11, 7.12. Surprisingly, the uncertainty bands obtained with MC method are much more narrow than the Hessian error bands. Whereas the observed difference becomes more moderate for the gluon and sea quark distributions at higher energy scales at NNLO (cf. fig. 7.12), it remains noticeable for the valence quarks. However, the error bands generated with both methods are placed along very similar central values which indicates a good agreement for the best fit result. With the limited amount of the MC replicas and consideration of DIS process only in the initial phase of the project we conclude that the MC error method for nuclear PDFs requires creation of more MC replicas and the consideration of several processes to unfold the full benefit.

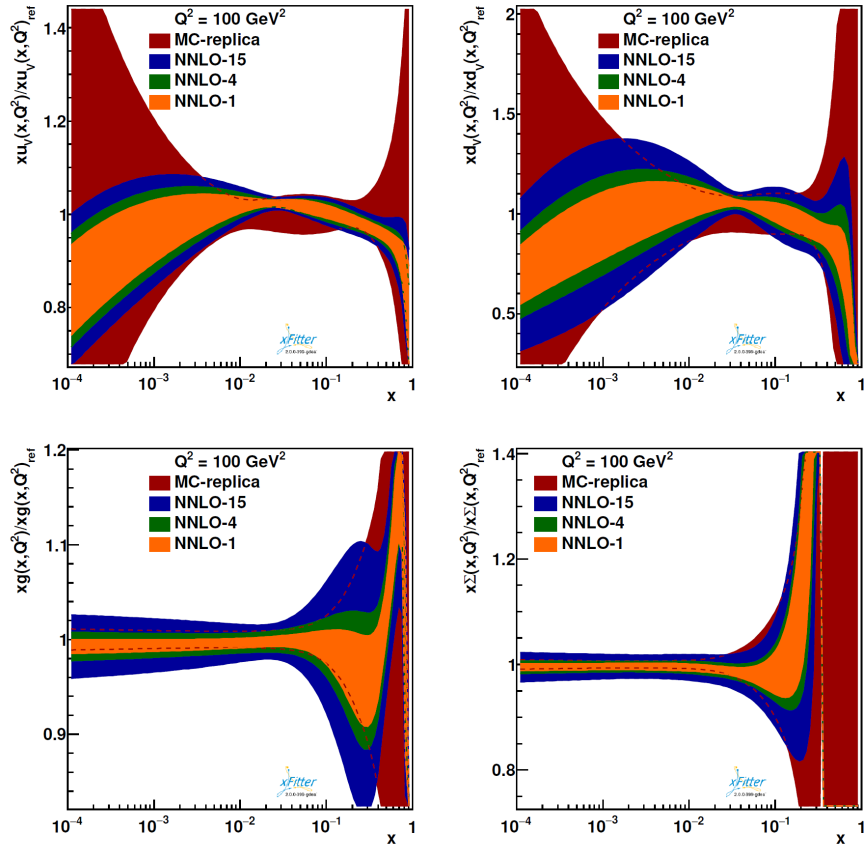


Figure 7.8.: Same as for figure 7.7, but at the higher scale $Q^2 = 100 \text{ GeV}^2$.

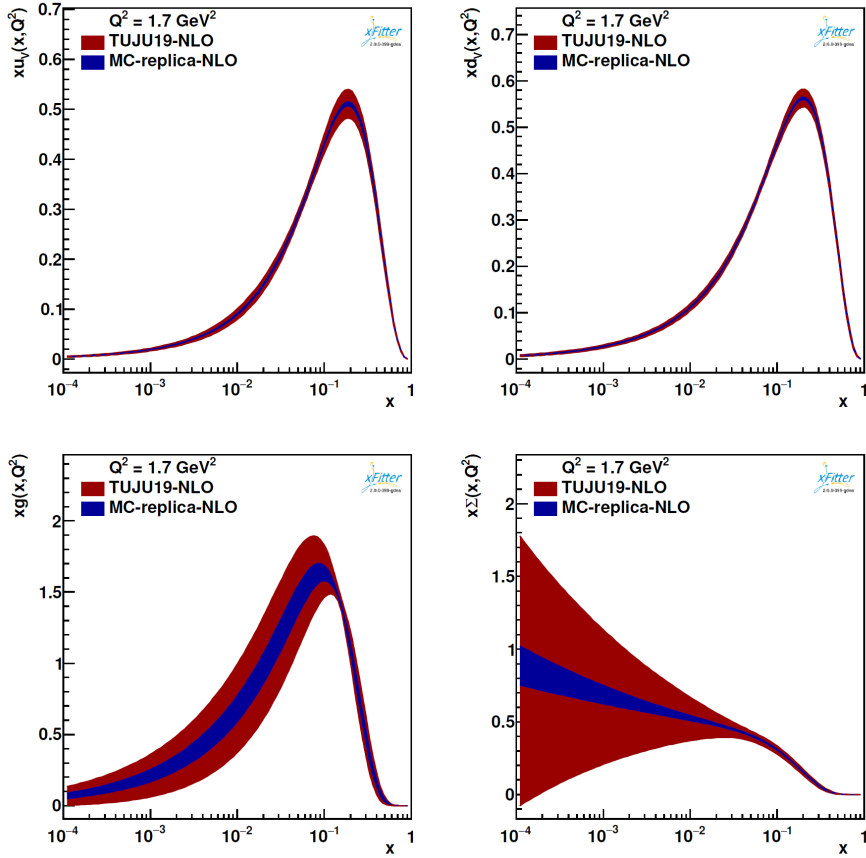


Figure 7.9.: Uncertainty bands for nPDFs at NLO obtained with Hessian method compared to MC replicas at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. The Hessian error bands (red) were created with statistically motivated value of $\Delta\chi^2 = 50$ with 90% confidence level. For the MC error bands (blue) 100 converged MC replicas were used. The comparison is shown for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$.

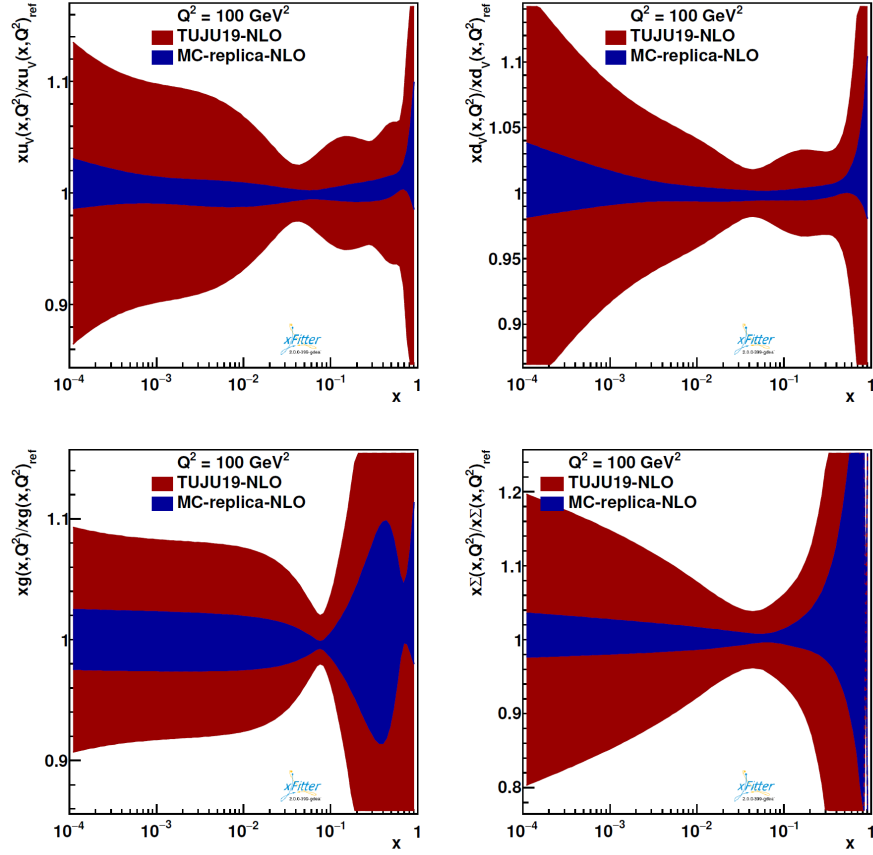


Figure 7.10.: Uncertainty bands for nPDFs at NLO obtained with Hessian method compared to MC replicas at the energy scale $Q^2 = 100 \text{ GeV}^2$. The ratios $xf(x, Q^2)/xf(x, Q^2)_{\text{ref}}$ are shown for the error bands obtained by 100 MC replicas (blue) compared to the Hessian uncertainty analysis (red) with $\Delta\chi^2 = 50$. The comparison is shown for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$.

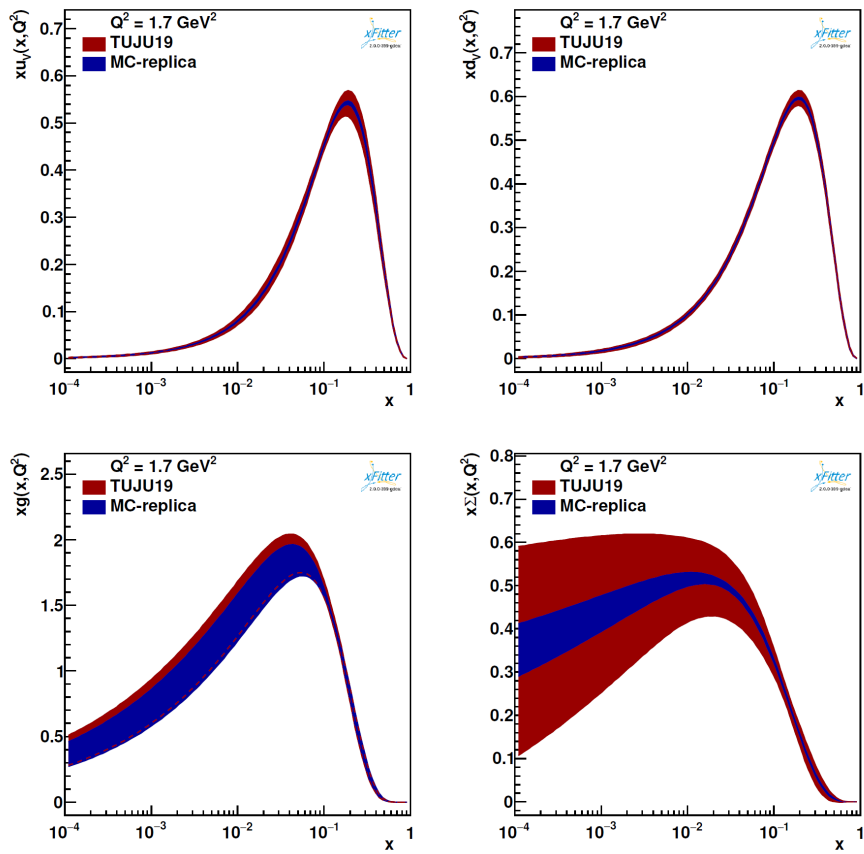


Figure 7.11.: Same as for figure 7.9, but at NNLO.

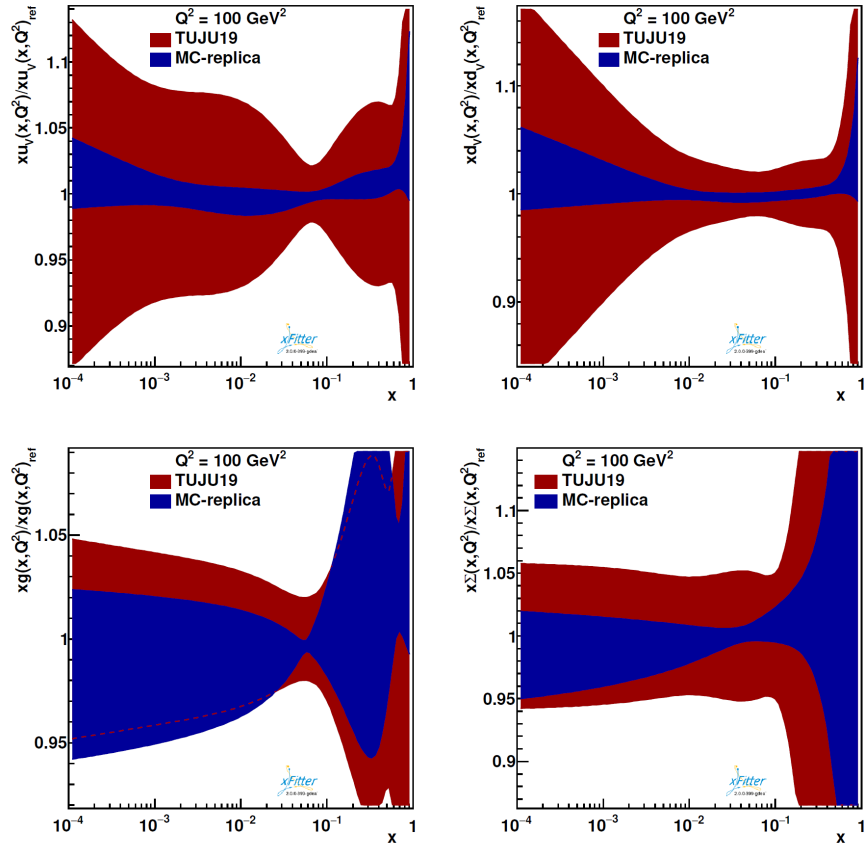


Figure 7.12.: Same as for figure 7.10, but at NNLO.

8. Study III: Parton Distribution Functions TUJU19

In this work two new nPDF analyses, performed at NLO and NNLO in pQCD have been introduced. The results of the presented global analysis are described in this chapter. First, the parton distribution functions for the free proton baseline are demonstrated. Next, the obtained nuclear PDFs are presented and a comparison to the existing nPDF sets is provided.

Note: the resulting PDF sets, referred to as TUJU19 [1], have been published in a standardized LHAPDF format. Therefore, this chapter, except section 8.3.3, is largely based on the own publication [1], for which the author has written the original draft of the manuscript (except introduction). The author has worked on the concept creation, numerical implementation and execution of the calculations for publication [1]. All numerical results presented in [1] have been obtained by the programs largely modified by the author. The performed modifications of the open-source code have been documented for the user guidance by the author and made publicly available. The author has also prepared the released sets of nuclear parton distribution functions TUJU19 in a standard format.

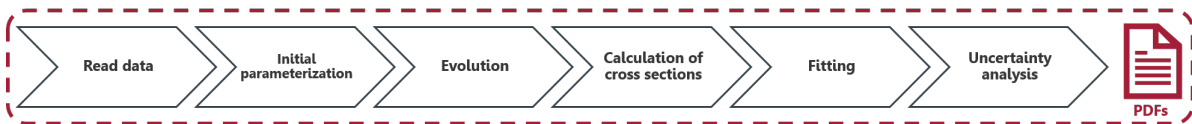


Figure 8.1.: Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: the overall results of the global analysis - PDFs.

8.1. Proton Baseline

Analyses of nuclear PDFs have often been performed by using an existing proton PDF set as a baseline for the nuclear modifications. In this work, however, we have fitted the proton PDFs using the same setup as for the nuclear PDFs. This ensures that all assumptions like sum rules, parton flavour decomposition, etc., as well as all parameters like coupling constants and quark masses, and also further settings like e.g. the heavy flavour mass scheme, are applied in a consistent way. Furthermore, this paves the way for a future combined proton and nuclear PDF analysis.

As we use xFITTER as our analysis framework, the baseline proton PDFs are derived with a very similar setup as for the HERA2.0 PDFs [169]. However, in addition to the combined HERA DIS data we also include data from other experiments (cf. table 6.1). Another difference is that we use the parameterization in eq. (5.1) whereas the HERA2.0 analysis includes additional terms for the gluon at the initial scale of the analysis. The obtained parton distribution functions are compared to the HERA2.0 PDFs [169] in figure 8.2 at NLO and in figure 8.3 at NNLO. As shown in Ref. [169], the HERA2.0 PDFs are well compatible with other state-of-the-art proton PDFs in the kinematic region considered in this work, and since the main focus of this work is on nuclear PDFs we do not present further comparisons to other proton PDF sets. As expected, because of our use of the same fitting framework with similar data and definition of χ^2 , the agreement with

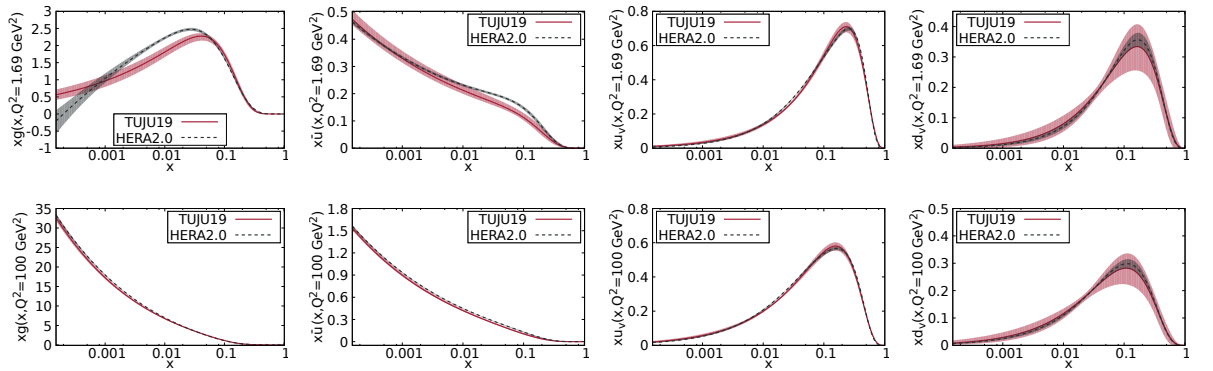


Figure 8.2.: Proton baseline PDFs TUJU19 at NLO compared to the HERA2.0 results, shown at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at $Q^2 = 100 \text{ GeV}^2$ after DGLAP evolution. Figure from [1].

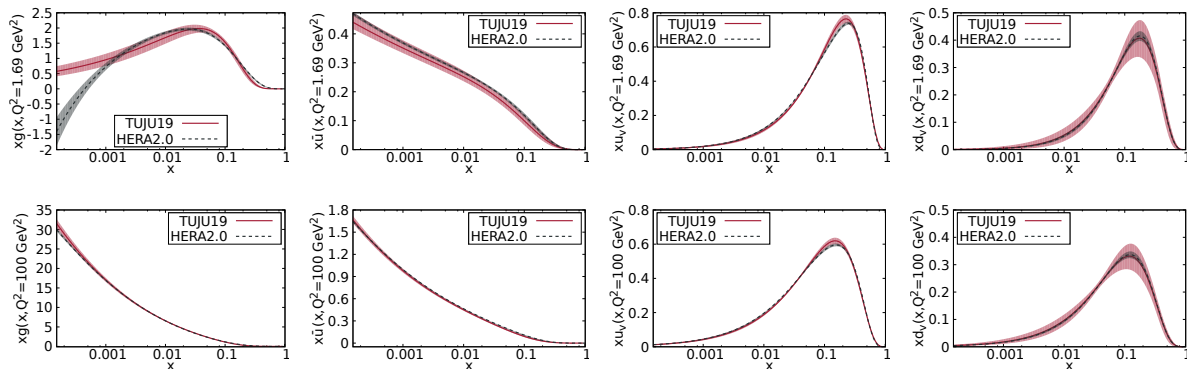


Figure 8.3.: Same as for figure 8.2, but at NNLO. Figure from [1].

the HERA2.0 PDFs is very good both at NLO and NNLO. The observed difference of the gluon PDFs at small x can be traced back to the different parameterization applied. Since we include also data from experiments other than HERA, we have used a larger $\Delta\chi^2$ value, namely $\Delta\chi^2 = 20$ (see section 7.1), which results in larger uncertainties than quoted for the HERA2.0 PDFs.

8.2. Nuclear PDFs

The resulting nuclear PDFs, referred to as TUJU19, are presented in figure 8.4 at NLO and in figure 8.5 at NNLO for a few different nuclei, together with the fitted proton baseline PDFs. As the sea-quark nPDFs have been assumed flavour-independent, i.e. $s = \bar{s} = \bar{u} = \bar{d}$, the $x\bar{u}(x, Q^2)$ distribution represents all sea quarks in the plots. Many earlier analyses have assumed that the nuclear modifications for the deuteron are negligible and constructed its PDFs from the free proton ones using isospin symmetry. In this work we, instead, treat the deuteron as a nucleus in the fitting procedure. Small deviations from the proton PDFs are found for the proton in a deuteron, as shown in figures 8.4 and 8.5. The deviation from the proton PDFs becomes larger with increasing mass number A , and significant effects are found in heavy nuclei such as iron and lead. The optimal parameters according to the chosen parameterizations in eqs. (5.1) and (5.2) are listed in appendix A.

8. Study III: Parton Distribution Functions TUJU19

Whereas figures 8.4 and 8.5 show the absolute PDFs, the nuclear modifications of the PDFs for the lead nucleus, defined as

$$R_i^{\text{p/Pb}} = \frac{f_i^{\text{p/Pb}}(x, Q^2)}{f_i^{\text{p}}(x, Q^2)}, \quad (8.1)$$

where $f_i^{\text{p/Pb}}(x, Q^2)$ and $f_i^{\text{p}}(x, Q^2)$ are the PDFs for the bound and the free proton, respectively, are shown in figure 8.6. The NLO and NNLO modifications are compared at the initial scale of the analysis ($Q^2 = 1.69 \text{ GeV}^2$) and at a higher scale ($Q^2 = 100 \text{ GeV}^2$) after DGLAP evolution. In both cases, the ratio of gluon PDFs shows some low- x

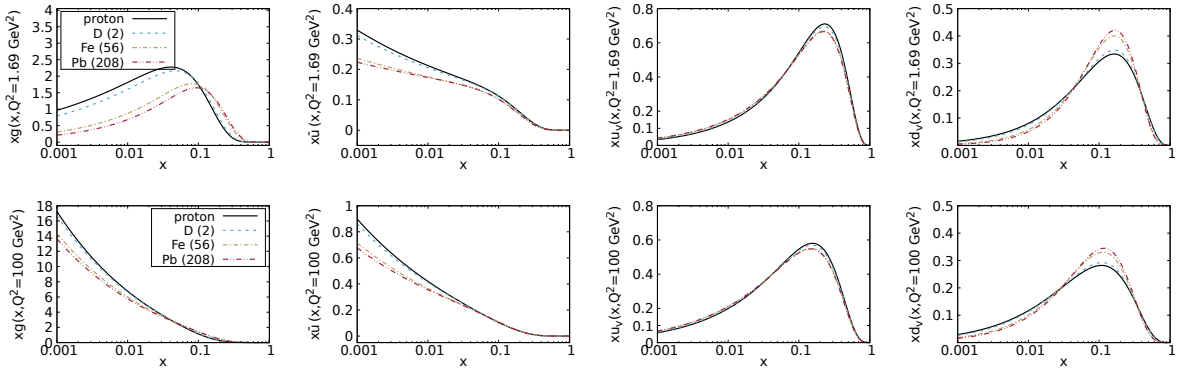


Figure 8.4.: Nuclear parton distributions functions TUJU19 in different nuclei with the mass number A at NLO, shown at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at $Q^2 = 100 \text{ GeV}^2$ after DGLAP evolution. Figure from [1].

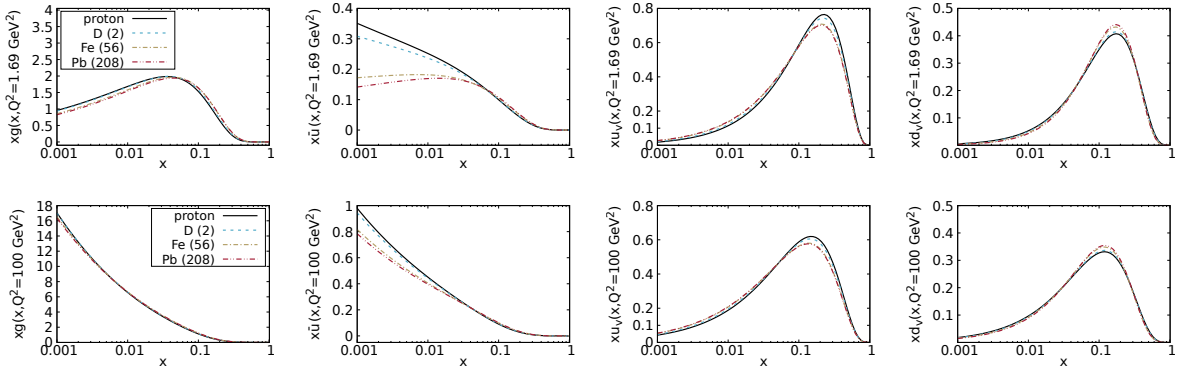


Figure 8.5.: Same as for figure 8.4, but at NNLO. Figure from [1].

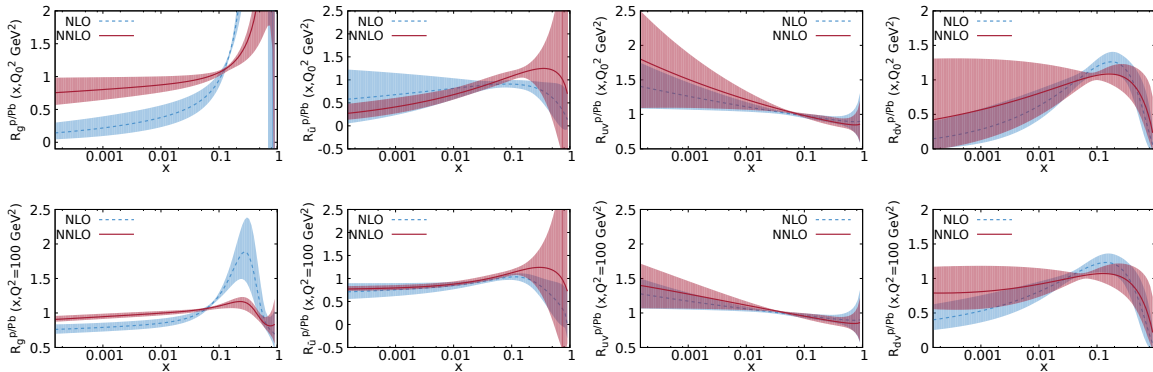


Figure 8.6.: Ratios $R_i^{p/Pb}$ of parton distribution functions per parton flavour i in a proton bound in lead compared to a free proton p . The obtained ratios are shown at NLO and NNLO, both at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at a higher scale $Q^2 = 100 \text{ GeV}^2$. Figure from [1].

shadowing and a rapid rise with increasing x . This behaviour is similar to what was observed in the HKN07 analysis [234], but in our case the enhancements are moderated at higher scales and a recognizable anti-shadowing peak develops around $x \sim 0.3$. For the sea quarks the typical nuclear modifications, shadowing, anti-shadowing and EMC suppression are visible already at the initial scale. However, especially the magnitude of the small- x shadowing differs at different perturbative orders, NNLO favoring a stronger effect. Since for gluons the behaviour is opposite we conclude that these differences arise due to the fact that the sea-quark and gluon evolution are coupled and the applied DIS data is not sensitive enough to fully separate the contributions. At higher scales the sea quark modifications come to a better mutual agreement though some difference still persists at large x . The valence PDFs were allowed to be flavour-dependent and the resulting nuclear effects indeed become rather different for bound valence up- and down-quark distributions. For d_v again the typical features including shadowing and anti-shadowing are well visible but for u_v we find that some amount of low- x enhancement is preferred. We note that a similar behaviour was observed in the nCTEQ15 analysis [170], although no neutrino DIS data were included there that would provide additional flavour sensitivity especially for the valence sector. One should keep in mind that the full nPDF for an average nucleon will be the sum of those for protons and neutrons, so the opposite behaviour will cancel out to a certain extent.

The uncertainty bands for the nPDFs provided in this work have been generated with $\Delta\chi^2 = 50$ as described in subsection 7.1. The resulting uncertainty bands do, however, depend also on the flexibility of the applied parameterization. Due to the limited sensitivity of the applied data to the gluon and sea quark nPDFs, we had to limit the number of A -dependent parameters in order to achieve numerical convergence of the fits. Therefore the provided uncertainty bands for the gluon distribution likely underestimate the true uncertainty to some extent, which should be kept in mind when comparing to previous works. In future, by adding more data providing further constraints one could consider admitting more parameters and therefore allowing larger flexibility of the parameterization.

8.3. Comparison to nPDF Sets

Each nPDF analysis is based on a set of assumptions, e.g. the form of the non-perturbative input at the initial scale, the choice of the proton baseline and the kinematic cuts. Therefore, even when based on the same set of data it is not guaranteed that the results will be equivalent. However, some level of agreement – within the estimated uncertainties – is expected.

8.3.1. Comparison at NLO

In figure 8.7 we compare our obtained nPDFs to those of other recent NLO nPDF analyses, nCTEQ15 and EPPS16, at our initial scale $Q^2 = 1.69 \text{ GeV}^2$ and at $Q^2 = 100 \text{ GeV}^2$. The comparisons are shown for g , \bar{u} , u_v , d_v and $V = u_v + d_v$ in a proton bound in lead. For gluons at the initial scale the agreement is not very good, though still well within the uncertainties. Towards higher scales a much better mutual agreement is observed. For sea quarks (here represented by \bar{u}) the agreement with the previous analyses is better already at the initial scale, and at $Q^2 = 100 \text{ GeV}^2$ our result is between EPPS16 and nCTEQ15. For valence quarks we find that u_v tends to stay below (above) the EPPS16 and nCTEQ15 results at $x \gtrsim 0.03$ ($x \lesssim 0.03$) whereas the opposite behaviour is found for d_v . This can be explained by the fact that in case of nuclear data only a combination of u_v and d_v is probed, and even with the included neutrino data the flavour dependence of valence quarks is not well constrained. Indeed, we find a very good agreement between the three analyses for the sum of valence quarks V .

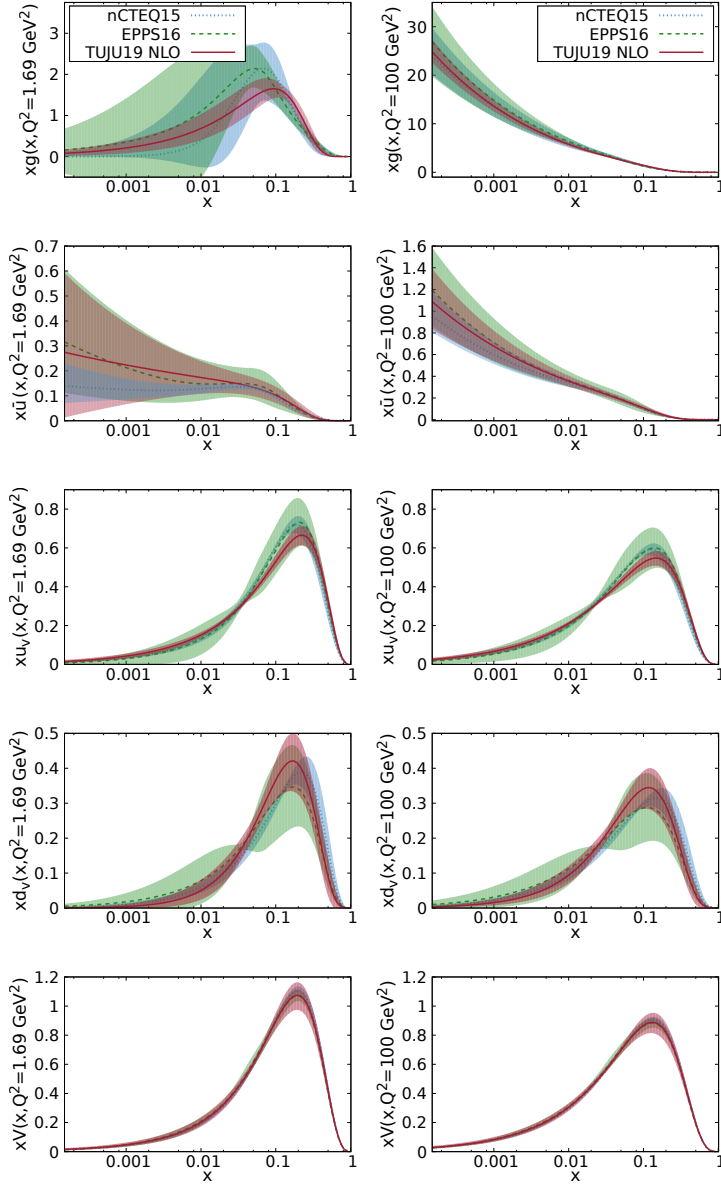


Figure 8.7.: Nuclear parton distribution functions TUJU19 (carmine) in lead at NLO compared to nCTEQ15 [170] (blue) and EPPS16 [188] (green), shown at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at a higher scale $Q^2 = 100 \text{ GeV}^2$. The comparison is presented for the distribution functions $x f_i(x, Q^2)$ per parton flavour $i = g, \bar{u}, u_v, d_v, V$, where V is the sum of valence quarks, in a bound proton in lead. Figure from [1].

The uncertainty bands in our NLO fit are similar to those obtained in the earlier analyses for sea quarks, but for gluons the resulting uncertainties are somewhat smaller. Since both EPPS16 and nCTEQ15 include additional data with some sensitivity to gluons, we conclude that our reduced uncertainties are likely due to the limited number of free parameters in the gluon nPDFs, and the uncertainty due to the lack of data constraints is underestimated. One should note that nCTEQ15 does not provide error sets for the baseline proton PDFs, which partly explains why their uncertainties for the sea quarks (at the initial scale) and the valence quarks tend to be smaller than in EPPS16 and this analysis. The comparisons were generated by using the LHAPDF6 library [36] and the published grids.

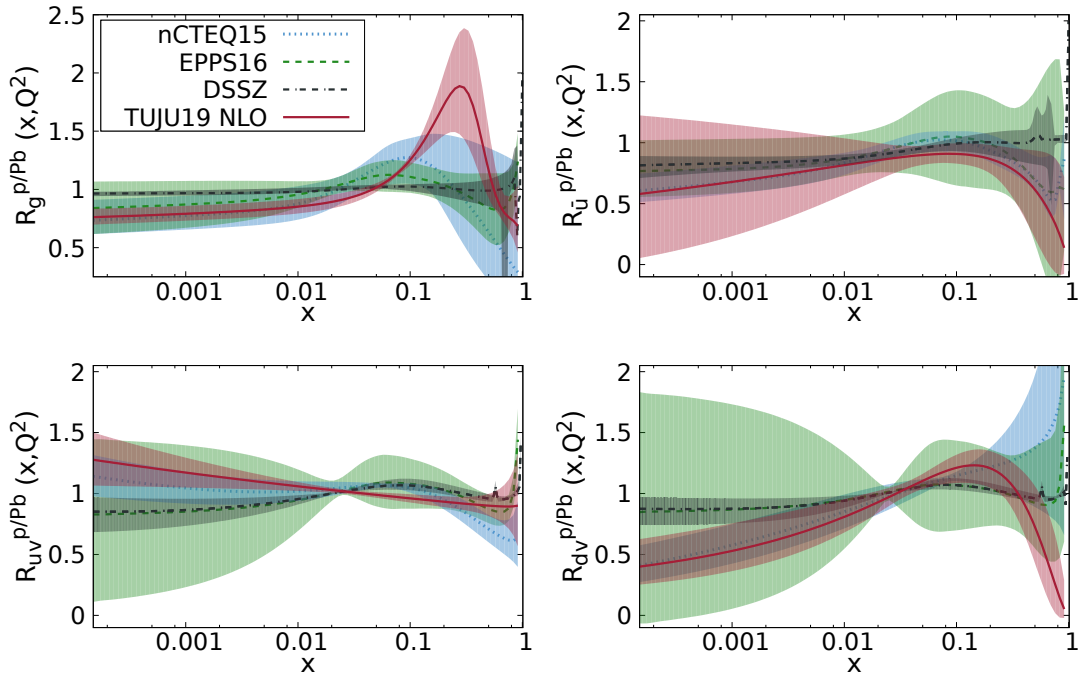


Figure 8.8.: Nuclear parton distribution functions TUJU19 (carmine) in lead at NLO compared to the nPDF sets nCTEQ15 [170] (blue), EPPS16 [188] (green), and DSSZ [205] (dark grey) shown at the higher scale $Q^2 = 100 \text{ GeV}^2$. The comparison is presented per parton flavour i for the ratios $R_i^{p/Pb}$ of PDFs in a proton bound in lead compared to the PDFs in a free proton. Figure from [1].

In figure 8.8 we compare the nuclear modifications of the PDFs as defined in eq. (8.1) at $Q^2 = 100 \text{ GeV}^2$. Also comparison to the DSSZ analysis is included, for which only ratios $R_i^{\text{p/Pb}}$ (eq. 8.1) were available with error bands^[1]. In most cases the results are compatible within the estimated uncertainties, though some features stand out. A rather prominent feature of our NLO gluons is the large antishadowing around $x \sim 0.3$. Such a large enhancement is not supported by other analyses which include data sensitive to gluon antishadowing and underlines the need for further data sensitive to such effects. However, the obtained gluon shadowing is in good agreement with EPPS16 and nCTEQ15 results, though with somewhat reduced uncertainty estimates. Only the DSSZ $R_g^{\text{p/Pb}}$ with very mild shadowing is outside the uncertainty bands in this region. For the flavour dependence of the valence quarks we find a similar behaviour as nCTEQ15 where some small- x enhancement and large- x suppression were observed for u_v , along with opposite behaviour for d_v . However, when calculating the total valence distribution for a complete nucleon, as shown in figure 8.7, we find a good agreement with the other analyses.

8.3.2. Comparison at NNLO

The comparison of the TUJU19 NNLO nPDF fit to other NNLO nPDF analyses is shown in figures 8.9, 8.10 for nNNPDF1.0 [235] and in figure 8.11 for KA15 [236]. The comparison at NNLO is separated into multiple figures since different information is available for nNNPDF1.0^[2] and KA15^[3] analyses. In case of nNNPDF1.0 we consider lead nPDFs for gluon g and the quark singlet $\Sigma = u + \bar{u} + d + \bar{d} + s + \bar{s}$ (as per [235]) in figure 8.9. Additionally, we show the equivalent comparison for the deuteron nPDFs (cf. fig. 8.10), since it was treated as a nucleus without neglecting nuclear effects in our analysis and nNNPDF1.0 is the only other group providing a dedicated LHAPDF set for the deuteron. In several earlier analyses the deuteron has been constructed as a pure composition of free proton and neutron PDFs. In our analysis the framework is flexible enough to treat the deuteron as a regular nucleus, even if the nuclear effects might or might not be small.

[1] No LHAPDF6 grids are available for DSSZ.

[2] The nNNPDF1.0 LHAPDF6 set is provided with the assumption that $u = d$ and that $\bar{u} = \bar{d} = \bar{s} = s$ to comply with the LHAPDF format, i.e. to provide individual quark flavours. For our comparison, we have used the provided PDFs and LHAPDF uncertainties at 90% confidence level, keeping in mind that only the quark multiplets Σ , T_3 , T_8 were determined in the nNNPDF1.0 analysis.

[3] For the comparison at NNLO also the KA15 nPDFs, provided on request by the authors, were included. The ratios $R^{\text{p/Pb}}$ are not available for this analysis.

8. Study III: Parton Distribution Functions TUJU19

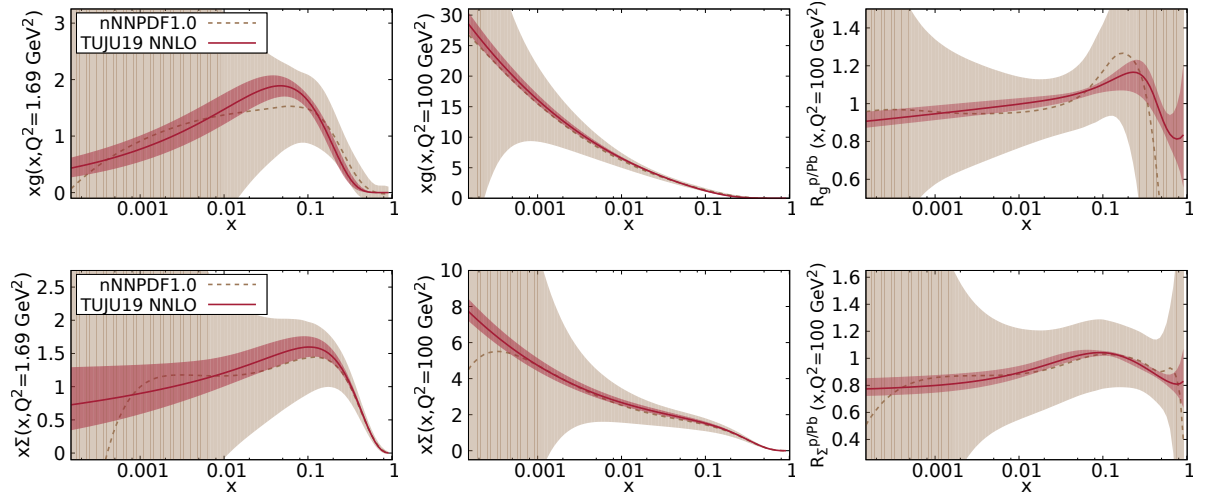


Figure 8.9.: Nuclear parton distribution functions TUJU19 (carmine) in lead at NNLO compared to the LHAPDF set nNNPDF1.0 [235] (light brown), shown at our initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at a higher scale $Q^2 = 100 \text{ GeV}^2$ for distribution functions xf_i , and at higher scale $Q^2 = 100 \text{ GeV}^2$ for the ratios $R_i^{p/Pb}$ of PDFs in a proton bound in lead compared to PDFs in a free proton. The comparison is presented for the gluon g and for the quark singlet $\Sigma = u + \bar{u} + d + \bar{d} + s + \bar{s}$ in a bound proton in lead. Figure from [1].

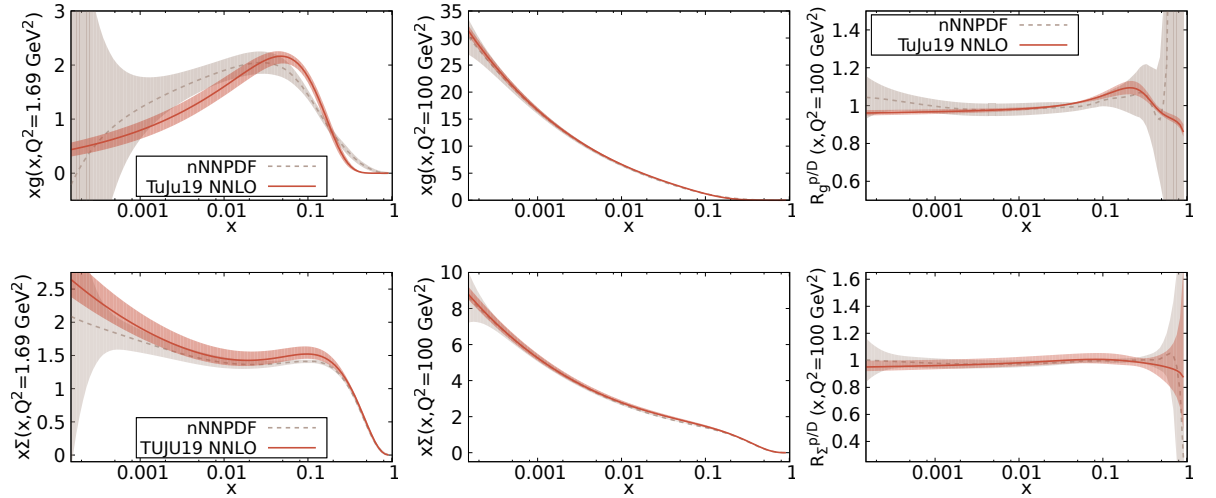


Figure 8.10.: Same as for figure 8.9, but for deuteron.

The comparisons to nNNPDF1.0 are shown at two scales, at $Q_0^2 = 1.69 \text{ GeV}^2$ and at $Q^2 = 100 \text{ GeV}^2$ for the distribution functions, and at the higher scale $Q^2 = 100 \text{ GeV}^2$ for the ratios $R_i^{p/A}$ of PDFs in a proton bound in nucleus compared to PDFs in a free proton. Even though none of the analyses includes data directly sensitive to the gluon distribution, a reasonable behaviour is found for the gluon at the initial scale of this analysis, keeping in mind that our uncertainty bands for gluon are potentially underestimated, as discussed above. The congruence is improved when going towards higher scales (e.g. $Q^2 = 100 \text{ GeV}^2$). Furthermore, a very good agreement is observed for the quark singlet Σ shown in the lower panels of figures 8.9 and 8.10. The values of $x\Sigma$ are lying within the error bands at the initial scale, and become even more consistent at the higher scale. The observed deviation in the low- x region ($x < 0.0005$) reflects the lack of low- x constraints by the available nuclear DIS data. We also show the ratios $R^{p/A}$ for TUJU19 compared to nNNPDF1.0 at NNLO in figures 8.9 and 8.10. Again a reasonable shape is found for the gluon nuclear modification, and a very good agreement is visible for the quark singlet, as it is well constrained by the incorporated experimental data. The uncertainties of the nNNPDF1.0 distributions are considerably larger compared to the ones found in our analysis, especially in case of lead, whereas the error bands in nNNPDF1.0 set for deuteron are more moderate. Indeed, as shown in Ref. [235], the uncertainty bands for a combination of singlet and octet contributions become comparable to those in the earlier works where direct data constraints exist. Still, the nNNPDF1.0 uncertainties grow rapidly towards small- x which can be accounted for the applied neural network framework that is not as prone to parameterization bias as the traditional Hessian error analysis applied here.

The nuclear parton distribution functions TUJU19 at NNLO compared to the results by the KA15 group are presented in figure 8.11. Here we consider lead nPDFs for gluons, sea quarks (here \bar{s}) and the sum of valence quarks, $V = u_v + d_v$. Again the comparisons are shown at two scales, at the initial scale of KA15, $Q_0^2 = 2.0 \text{ GeV}^2$, and at $Q^2 = 100 \text{ GeV}^2$. Considering the fact that neither of the analyses includes data directly sensitive to the gluon distribution, a fair agreement is found for the gluon at the initial scales. However, while the agreement between TUJU19 and nNNPDF1.0 remains at higher scales (fig. 8.9), the gluon distribution from KA15 falls below the other two (fig. 8.11) at $Q^2 = 100 \text{ GeV}^2$. The \bar{s} distributions in turn are in a reasonable agreement

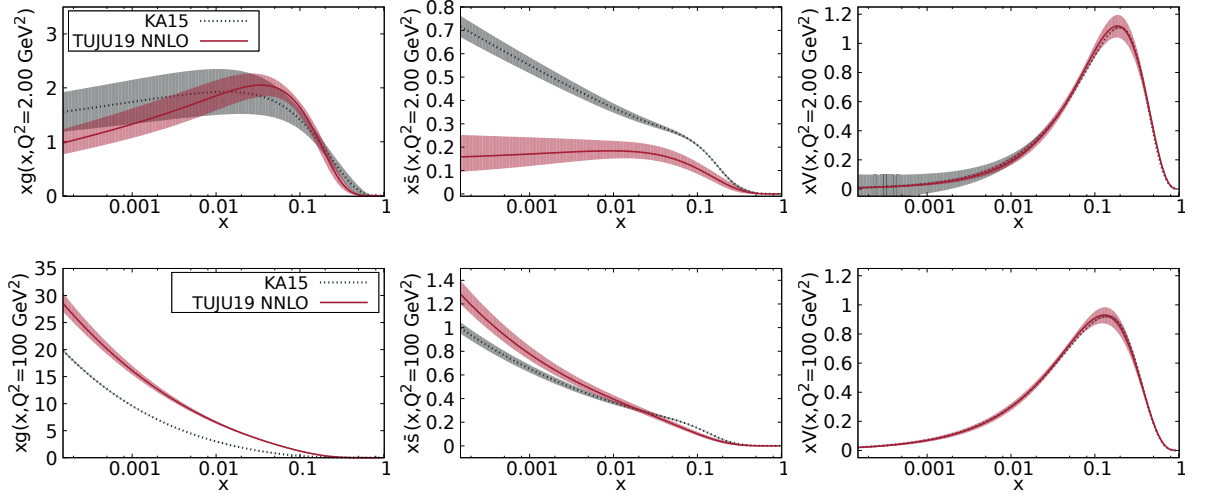


Figure 8.11.: Nuclear parton distribution functions TUJU19 (carmine) in lead at NNLO compared to the results by the KA15 group [236] (dark grey), shown at the initial scale of KA15, $Q_0^2 = 2.0 \text{ GeV}^2$, and at a higher scale $Q^2 = 100 \text{ GeV}^2$. The comparison is presented for the distribution functions $xf_i(x, Q^2)$ with $i = g, \bar{s}, V$, where V is the sum of valence quarks, in a bound proton in lead. Figure from [1].

at higher scales, although at the initial scale the KA15 result is considerably above TUJU19. The total valence distributions from TUJU19 are found to be in very good agreement with those of the KA15 analysis. Apart from the gluon nPDFs at the initial scale, the KA15 uncertainties tend to be very small. This may partly follow from the rather rigid parameterization applied, but also be due to the chosen low error tolerance $\Delta\chi^2 = 1$.

8.3.3. Summary

All nPDF analyses used for the comparison in the previous subsections have been reviewed with respect to the different criteria like considered processes, parameterization and further analysis details. The comparison of these criteria for our analysis TUJU19 and the previous nPDF analyses is summarized in table 8.1 providing an overview on common and different aspects. Common to all nPDF sets is that neutral current DIS data are used. Additionally, DSSZ, EPPS16 and TUJU19 (this analysis) were able to include the neutrino nucleus DIS measurements. The nPDF sets produced by the groups

DSSZ, nCTEQ15, KA15 and EPPS16 also include the Drell-Yan process into their analysis, which can be part of the next project phase for TUJU19. Furthermore, the analysis performed by EPPS16 also contains the information from the latest LHC experiments.

Another aspect is that the nuclear PDFs are usually determined for a specific *free* proton PDF set. As described in the previous subsections, in this analysis not an existing proton PDF set is used, but an own PDF set for a *free* proton is obtained first. The analysis is performed at next-to-leading order (NLO) and next-to-next-to-leading order (NNLO), for which LHAPDF sets are provided by this analysis and the nPDF set recently created by the NNPDF group.

As part of the presented framework, a new form of parameterization, a combination of equations (5.1) and (5.2), has been used to determine the nuclear PDFs. For that matter deuteron has been considered being a nucleus with non-negligible nuclear effects. The complete analysis framework was implemented in an open source tool set by modifying the tool called xFITTER.

	DSSZ [205]	nCTEQ15 [170]	KA15 [236]	EPPS16 [188]	nNNPDF1.0 [235]	TUJU19 [1]
Highest order in α_s	NLO	NLO	NNLO	NLO	NNLO	NNLO
Lepton-nucleus DIS	✓	✓	✓	✓	✓	✓
Neutrino-nucleus DIS	✓	-	-	✓	-	✓
Drell-Yan	✓	✓	✓	✓	-	-
LHC $p + Pb$ W, Z data	-	-	-	✓	-	-
Number of data points	1579	740	1479	1811	451	2336
Q^2 cut in DIS	1.0 GeV ²	4.0 GeV ²	1.0 GeV ²	1.69 GeV ²	3.5 GeV ²	3.5 GeV ²
Initial scale Q_0^2	1.0 GeV ²	1.69 GeV ²	2.0 GeV ²	1.69 GeV ²	1.0 GeV ²	1.69 GeV ²
Free parameters	25	17	16	20	184	16
Flavours	$g, u_v = d_v,$ $s = \bar{s} = \bar{u} = \bar{d}$	$g, u_v, d_v,$ $s = \bar{s} = \kappa/2(\bar{u} + \bar{d})$	$g, u_v, d_v,$ $s = \bar{s} = 1/4(\bar{u} + \bar{d})$	$g, u_v, d_v,$ $s = \bar{s}, \bar{u}, \bar{d}$	g, Σ, T_8	$g, u_v, d_v,$ $s = \bar{s} = \bar{u} = \bar{d}$
Free proton baseline	MSTW2008 [237]	CTEQ6.1-like [225]	JR09 [238]	CT14NLO [239]	NNPDF3.1 [240]	own proton fit
Heavy quark scheme	GM-VFNS	GM-VFNS (ACOT)	ZM-VFNS	GM-VFNS (SACOT)	GM-VFNS (FONLL)	GM-VFNS (FONLL)
Error analysis	Hessian	Hessian	Hessian	Hessian	Monte Carlo	Hessian
Error tolerance $\Delta\chi^2$	30	35	1	52	-	50
LHAPDF set	-	✓	-	✓	✓	✓

Table 8.1.: Comparison of the most recent previous nPDF analyses and this work TUJU19. Different criteria like considered processes, parameterization and analysis details are listed in the left column and compared between the nPDF sets specified with the original publication in the headline.

9. Study IV: Application to Further Processes

In this work, the numerical setup for studying deeply inelastic scattering (DIS) processes was completed successfully and the obtained results were published [1]. However, the DIS process is sensitive only to certain types of partons and the kinematic coverage of the current data is limited. In the future, we are planning to successively add data also for other measured processes for which theoretical calculations at NNLO exist, for example data from dilepton production in fixed-target proton-nucleus collisions, the so-called Drell-Yan (DY) processes [241–244] and W^\pm, Z boson production data from the proton-lead (p+Pb) collisions at the LHC [245–248]. We expect the DY data and Z boson production data to provide further constraints for the sea quark distributions. Additionally, it will increase the kinematic coverage of the applied data and therefore result in a nuclear PDF set with higher precision and wider applicability range. The W^\pm boson production data from p+Pb collisions at the LHC are sensitive to the flavour decomposition and could help to disentangle observed differences in the nuclear modifications of the valence quarks.

In this chapter we present some results from the calculation of the theoretical predictions for W^\pm boson production processes by using the published nPDF sets TUJU19. This application of the obtained nPDFs to further processes motivates the outlook to continue the work with the implemented framework. The extensions of the set-up necessary for inclusion of further processes are outlined here. Finally, the updated *free* proton PDF sets based on Drell-Yan and W^\pm, Z boson production data, referred to as TUJU20, and an outlook for nPDFs are described.

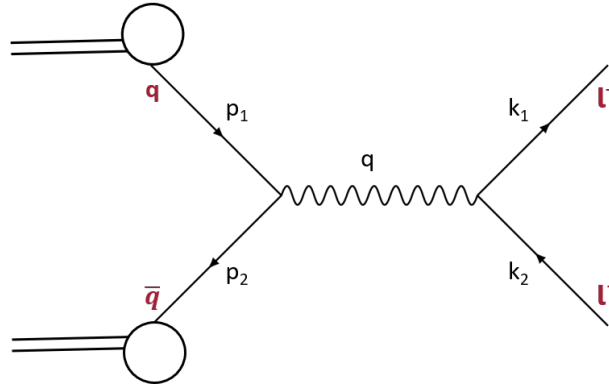


Figure 9.1.: Feynman diagram for Drell-Yan process at leading order. The scattering partons (quarks) are coming from two different hadrons (symbolized by circles).

9.1. Drell-Yan and W^\pm , Z Boson Production Processes

With the completion of the project phase by considering nuclear DIS data for the scattering reactions with charged leptons and neutrinos, the basic framework for the analysis of further processes, e.g., so-called Drell-Yan processes and W^\pm , Z boson production processes, is prepared. However, there are differences in the fitting procedure dependent on the considered process. The differences arise at the time when the cross sections need to be calculated. In principle one could implement the calculations of cross sections for DY and W^\pm , Z boson production processes in a similar manner as for DIS. The problem is that the computations become too slow for an iterative fitting procedure to be effective, especially at NNLO. Therefore, further tools, such as APPLGRID [194] and MCFM [249–251], are needed here. The schematic overview of a possible fitting routine and the required tool set (apart from xFITTER) is visualized in figure 9.2.

In contrast to DIS processes, not a lepton is scattered off a hadron, but two hadrons are involved in the scattering process (cf. figure 9.1). The production of a lepton pair ($l\bar{l}$ with $l = e, \mu, \tau$) of the same flavour with large invariant mass in hadronic collisions is referred to as Drell-Yan (DY) production, first proposed in refs. [252–256] by DRELL and YAN.

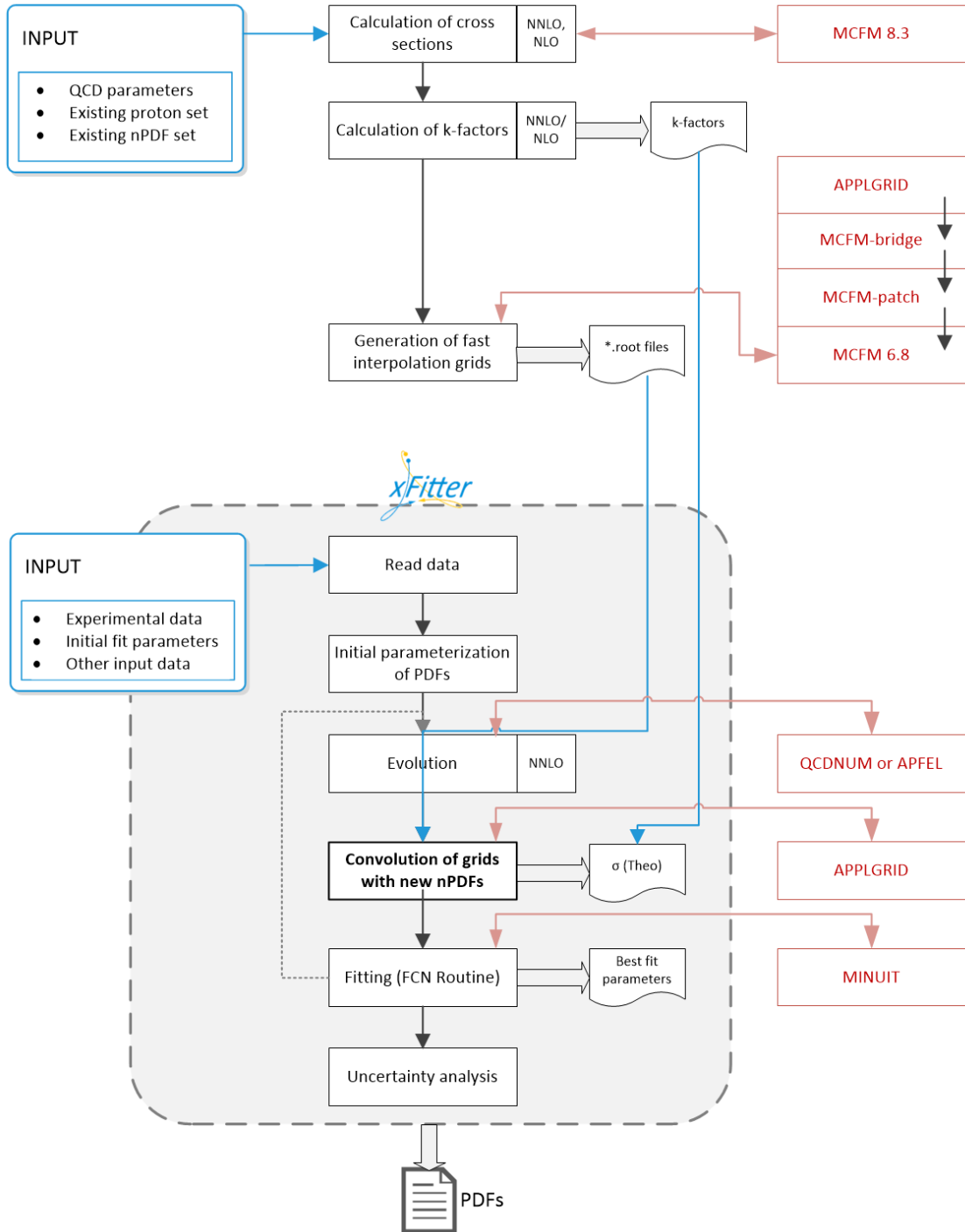


Figure 9.2.: Schematic view of the high-level xFITTER functionalities considering Drell-Yan and W^\pm , Z boson production processes. xFITTER logo credited from [189].

The nomenclature to declare the scattering reaction for Drell-Yan is

$$q\bar{q} \rightarrow \gamma \rightarrow l\bar{l}. \quad (9.1)$$

The Feynman diagram for the Drell-Yan process at leading order is shown in figure 9.1, where the scattering partons (quarks) are coming from two different hadrons. Similarly, the reaction for so-called W^\pm or Z boson production at leading order is

$$\begin{aligned} q\bar{q} &\rightarrow W^\pm \rightarrow l\bar{l}, \quad \text{and} \\ q\bar{q} &\rightarrow Z \rightarrow l\bar{l}. \end{aligned} \quad (9.2)$$

In the W^\pm boson production, pair of a charged-lepton and an uncharged-lepton (neutrino ν) is created in the final state, whereas for Z boson production lepton pair $l\bar{l}$ of the same flavour ($l = e, \mu, \tau$) occurs.

The validity of the factorization theorem for DY is shown e.g., in reference [257]. For DY one uses the factorization theorem of the form given in equation (3.22), since two hadrons are involved in the scattering process. The double differential cross section at leading order (LO) is calculated as per [43, 56]

$$\frac{d^2\sigma}{dM^2 dy} = \frac{4\pi\alpha^2\tau}{3N_C M^4} \sum_q e_q^2 [q(x_1)\bar{q}(x_2) + \bar{q}(x_1)q(x_2)] \quad (9.3)$$

with the number of colors N_C , $\tau = M_{ll}^2/s$ and the kinematic variables

$$\begin{aligned} M^2 &= (p_1 + p_2)^2 \equiv \hat{s} = x_1 x_2 s && \text{(Invariant mass),} \\ y &= \frac{1}{2} \log \frac{x_1}{x_2} && \text{(Rapidity)} \end{aligned} \quad (9.4)$$

with $p_{1,2}$ momenta of the incoming partons carrying momentum fraction $x_{1,2}$ of the total hadron (proton) momentum, and the center-of-mass energy \sqrt{s} , leading to

$$x_1 = \frac{M}{\sqrt{s}} \exp(y) \quad \text{and} \quad x_2 = \frac{M}{\sqrt{s}} \exp(-y). \quad (9.5)$$

Experimentally it might be challenging to measure the rapidity for highly relativistic particles since in order to provide the rapidity one needs to identify the particle, so to

know its mass. Therefore, often not the rapidity y but the so-called pseudo-rapidity η , that describes the angle between the particle and the beam axis, is provided by the experiments. For instance, also for the discussion of the application in the next section the variable η is used. However, for highly-relativistic particles it is $\eta \simeq y$.

9.2. Application of TUJU19 for W^\pm Boson Production

In this section we present some sample results from the calculation of the theoretical predictions for W^\pm boson production processes by using the published nPDF sets TUJU19 in order to validate the obtained nPDF sets and to prepare the extension of the global analysis framework. The calculated cross sections for W^\pm boson production compared to the experimental data from LHC Run II measured by the CMS collaboration [248] are presented in figures 9.3 and 9.4. In figure 9.3 the results are shown for the calculations performed with free proton PDF set and with nPDF set for the reason of comparison. Furthermore, the same approach has been repeated with PDF sets from previous analyses, here EPPS16, as they included LHC data from Run I in the analysis, and their proton baseline CT14, for reference purpose. As expected, due to the nuclear effects in p+Pb collisions the experimental data cannot be represented properly by the usage of

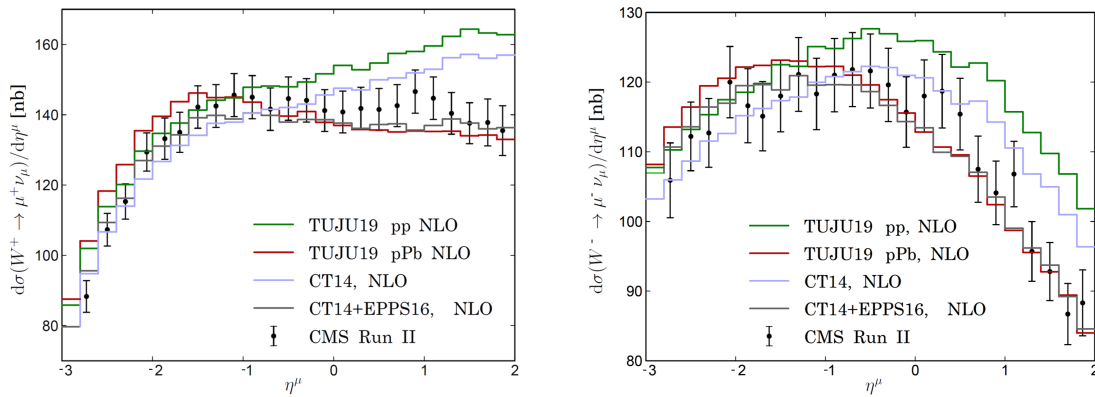


Figure 9.3.: Calculated cross sections for W^\pm boson production at NLO compared to the experimental data from LHC Run II [248] presented along pseudorapidity η^μ . The comparison is shown for the calculation with free proton PDF sets TUJU19 (green) and CT14 (purple), and with nPDF sets TUJU19 (red) and EPPS16 (grey). Figures prepared by I. Helenius, edited by M. Walt.

free proton PDF sets, which is valid for both, our TUJU19 set and the reference CT14 set. Although the shape of the calculated predictions seems to be in a good agreement, the normalization preferred by the free proton PDF set TUJU19 shows an offset compared to the line based on CT14. This can be interpreted as an indication that the overall quality of the global pQCD analysis can be improved by inclusion of DY and W^\pm, Z boson production data from $p+p$ collisions into the free proton fit. The numbers obtained with nPDF sets show a better agreement with the experimental data, and consequently a better agreement between the figures obtained with the different nPDF sets, TUJU19 and EPPS16. Only in the kinematic region $\eta^\mu < 0$ a deviation is noticeable.

The application of TUJU19 nPDF sets for the calculation of the cross sections of W^\pm boson production at NLO and NNLO is shown in figure 9.4. The uncertainty bands pre-

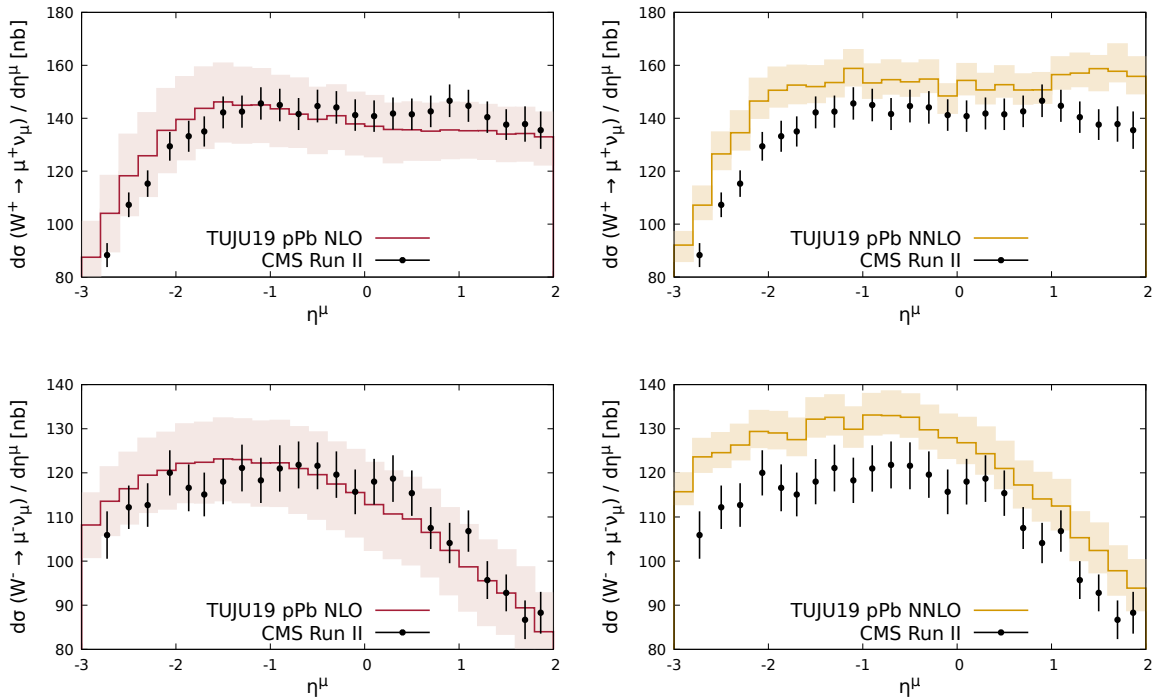


Figure 9.4.: Cross sections for W^\pm boson production calculated with TUJU19 PDF sets at NLO and NNLO compared to the experimental data from LHC Run II [248] presented along pseudorapidity η^μ , with uncertainty bands. Figures with central values prepared by I. Helenius, edited by M. Walt to include error bands.

sented in figure 9.4 have been calculated by considering all 58 nPDF members published within the LHAPDF set, i.e. including nPDF error sets and proton's error sets. Due to the long computation time at NNLO, which rapidly increases when using 58 nPDF members for the calculation, the uncertainty bands at NNLO have been obtained based on the so-called k -factors. That is, the calculation of the observable $d\sigma/d\eta$ has been performed at NLO by using the TUJU19 nPDFs at NNLO. Afterwards, the $k_{\text{NNLO/NLO}}$ -factor has been used to scale the uncertainties to the NNLO results. The $k_{\text{NNLO/NLO}}$ -factor is defined as

$$k_{\text{NNLO/NLO}} = \frac{d\sigma^{\text{NNLO}}/d\eta}{d\sigma^{\text{NLO}}/d\eta}, \quad (9.6)$$

for the observables $d\sigma^{(\text{N})\text{NLO}}/d\eta$ each calculated with the particular central nPDF set.

The shape of the obtained predictions (cf. fig. 9.4) shows a valid behaviour, only the normalization at NNLO is off. By considering the error bands, the measured data points are in line with the calculated results at NLO within the obtained uncertainties. The NNLO calculations with the TUJU19 PDFs are somewhat above the data. One possible reason might be a too large proton baseline that should be improved by including the W^\pm, Z boson production and DY data. Furthermore, as can be seen in figure 9.4 the uncertainties at NNLO are reduced compared to NLO. When looking at the TUJU19 nPDFs (cf. fig. 8.6) it seems that sea quarks have reduced error bands at NNLO compared to NLO. Though, one cannot clearly say whether this is a real NNLO effect or a coincidence. Nonetheless, the observation of reduced uncertainty bands at NNLO in figure 9.4 could be interpreted as a hint that the NNLO uncertainties are possibly underestimated. However, the observation at NLO provides a good validation of the nPDF sets obtained with DIS data only, and at the same time the NNLO results support the intention to include DY and W^\pm, Z boson production data in the next phase of the project. For that we expect that the updated nPDF set at NNLO will lead to the reduced differential cross sections $d\sigma^{\text{NNLO}}/d\eta(W^\pm \rightarrow \mu^\pm\nu_\mu)$ calculated at NNLO. Though, due to the interplay of two PDFs based on the form of the factorization theorem (cf. eq. (3.22)) one cannot make clear predictions how the distributions of the single partons (quarks or gluons) will change. Summing up, the main conclusion of this section is that W^\pm, Z boson production data and DY data will provide important constraints for the future analysis and should be included into the fit.

9.3. Updated Free Proton Baseline TUJU20

In this section the updated free proton PDF sets are compared to the published free proton PDF sets determined based on DIS data only. The free proton PDFs used as the baseline for the nuclear part of the QCD analysis were updated by considering experimental data for DY, W^\pm , Z boson production processes measured by the ATLAS and CMS collaborations at the LHC [258–262]. The new free proton PDF sets were created at NLO and NNLO and are referred to as TUJU20. As shown in figure 9.2 the main difference in the analysis procedure compared to the consideration of DIS processes is the calculation of cross sections, or more precise the generation of fast interpolation grids, which requires the setup of new tools and is demanding in terms of computational resources. However, for the experimental proton data applied for the new step of the analysis the fast interpolation grids were publicly available and could be used within the implemented framework. The obtained values for χ^2/N_{dp} are 1.306 at NLO and 1.241 at NNLO with 1,693 data points in total, which is similar to the free proton values from the TUJU19 analysis, especially at NNLO. The results, in form of a comparison to the TUJU19 free proton PDFs, are presented in figures 9.5 - 9.10.

The impact at NLO (e.g. fig. 9.5) is rather mild and the uncertainties are slightly smaller in some cases, especially for valence quarks. At NNLO (e.g. fig. 9.8) the resulting distributions for valence quarks, and especially for gluons are a bit lower. This could bring down the NNLO cross sections also for nuclear PDFs, since the changes in the free proton baseline will propagate into the nPDF analysis planned in the next project phase. Of course, only the final results of the updated nPDF analysis can show what the net effect will be.

The results obtained by the updated free proton PDF baseline confirm the motivation to include further processes, like e.g. DY, and W^\pm , Z boson production, into the global nPDF analysis. Also, it illustrates the advantage of having a consistent common framework for proton PDF and nuclear PDF analyses providing a flexibility to consider the same processes in both steps.

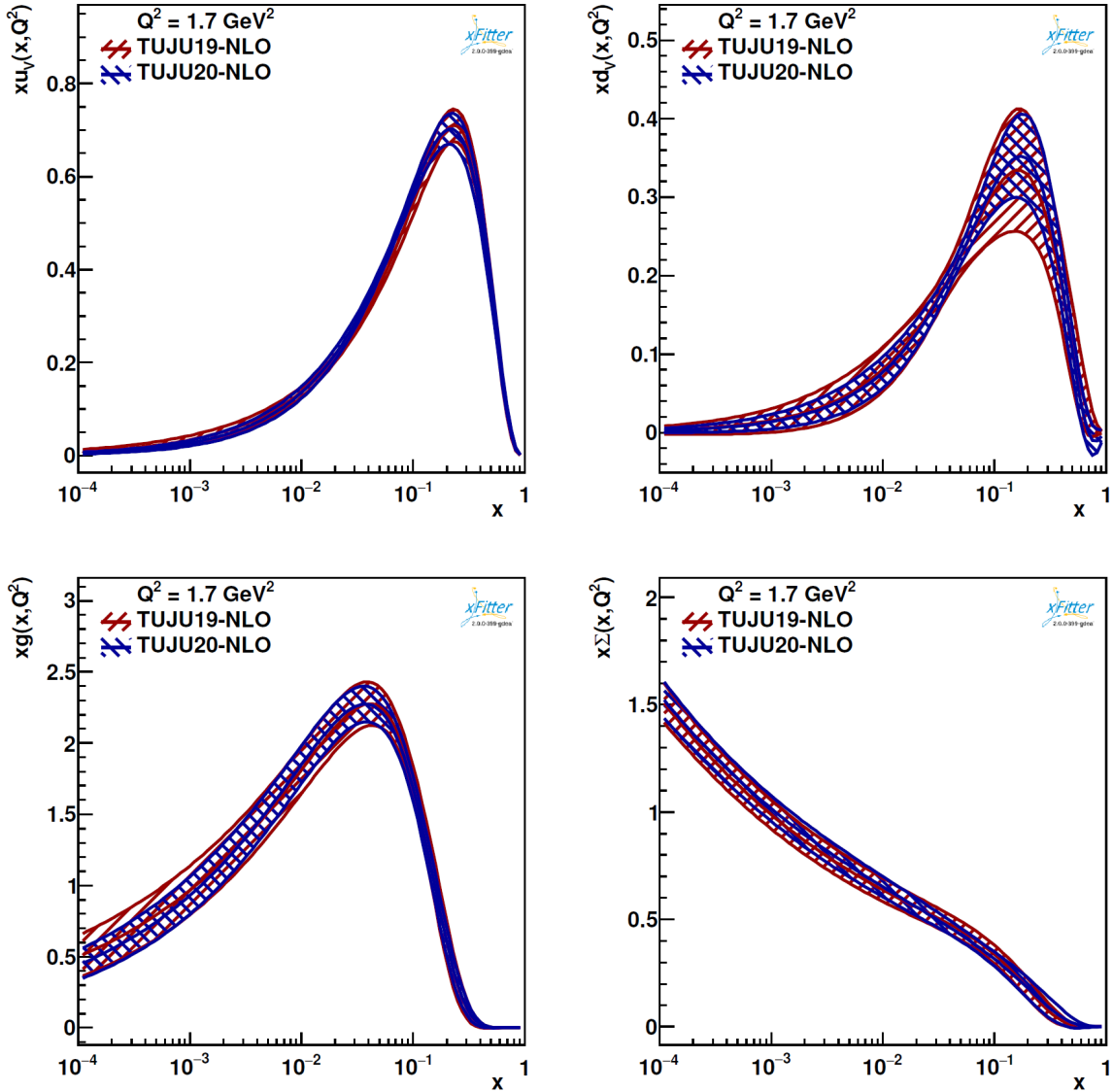


Figure 9.5.: Free proton PDF set obtained with DIS data only (TUJU19) compared to the updated free proton PDF set based on DIS and DY, W^\pm , Z boson production data (TUJU20) at NLO. The representation is provided for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$ at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$.

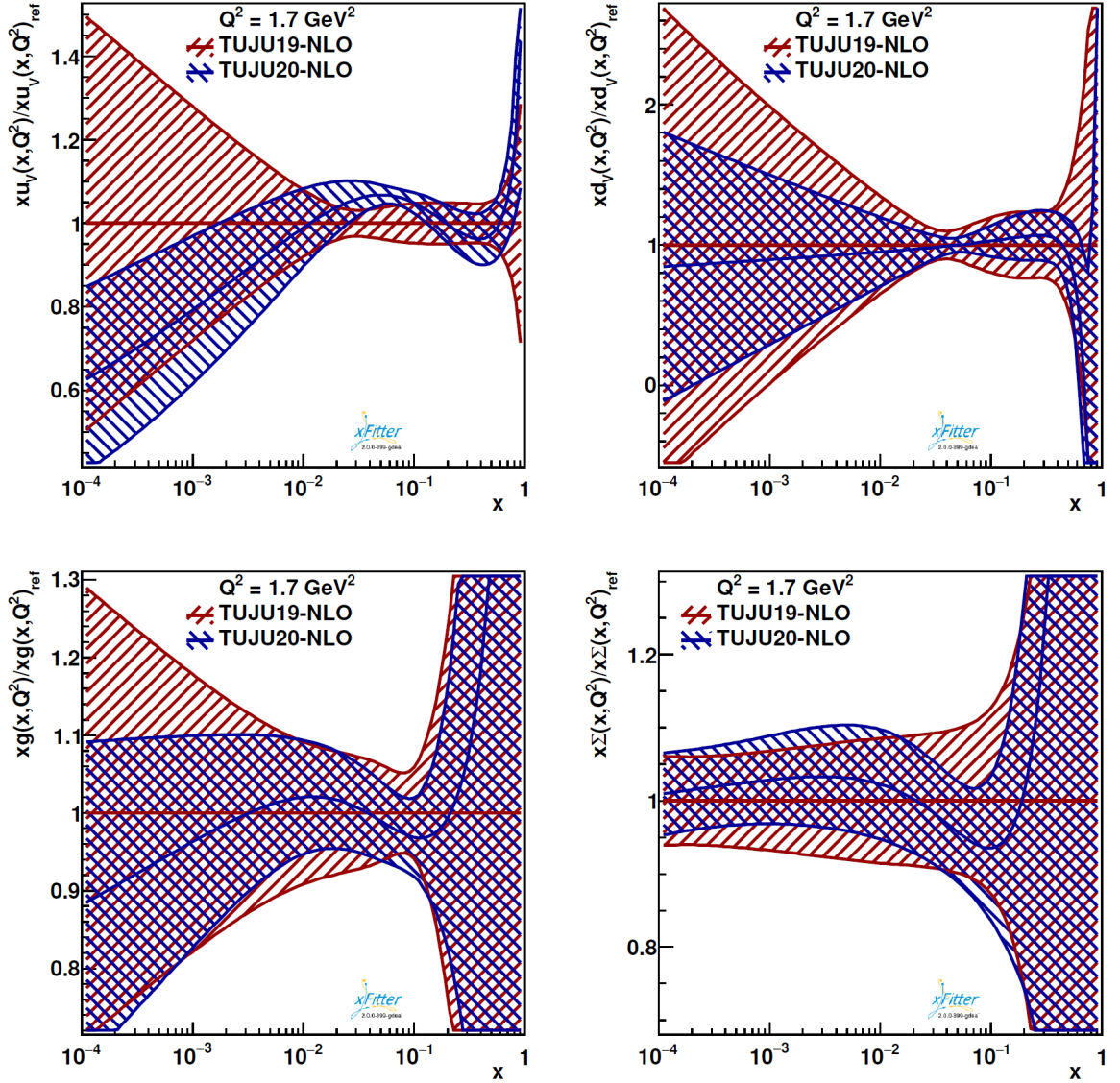


Figure 9.6.: Same content as in figure 9.5, but different visualization. The ratios $xf(x, Q^2)/xf(x, Q^2)_{\text{ref}}$ are shown for the free proton PDF set obtained with DIS data only referred to as TUJU19 ('ref') compared to the updated free proton PDF set based on DIS and DY, W^\pm , Z boson production data (TUJU20) at NLO.

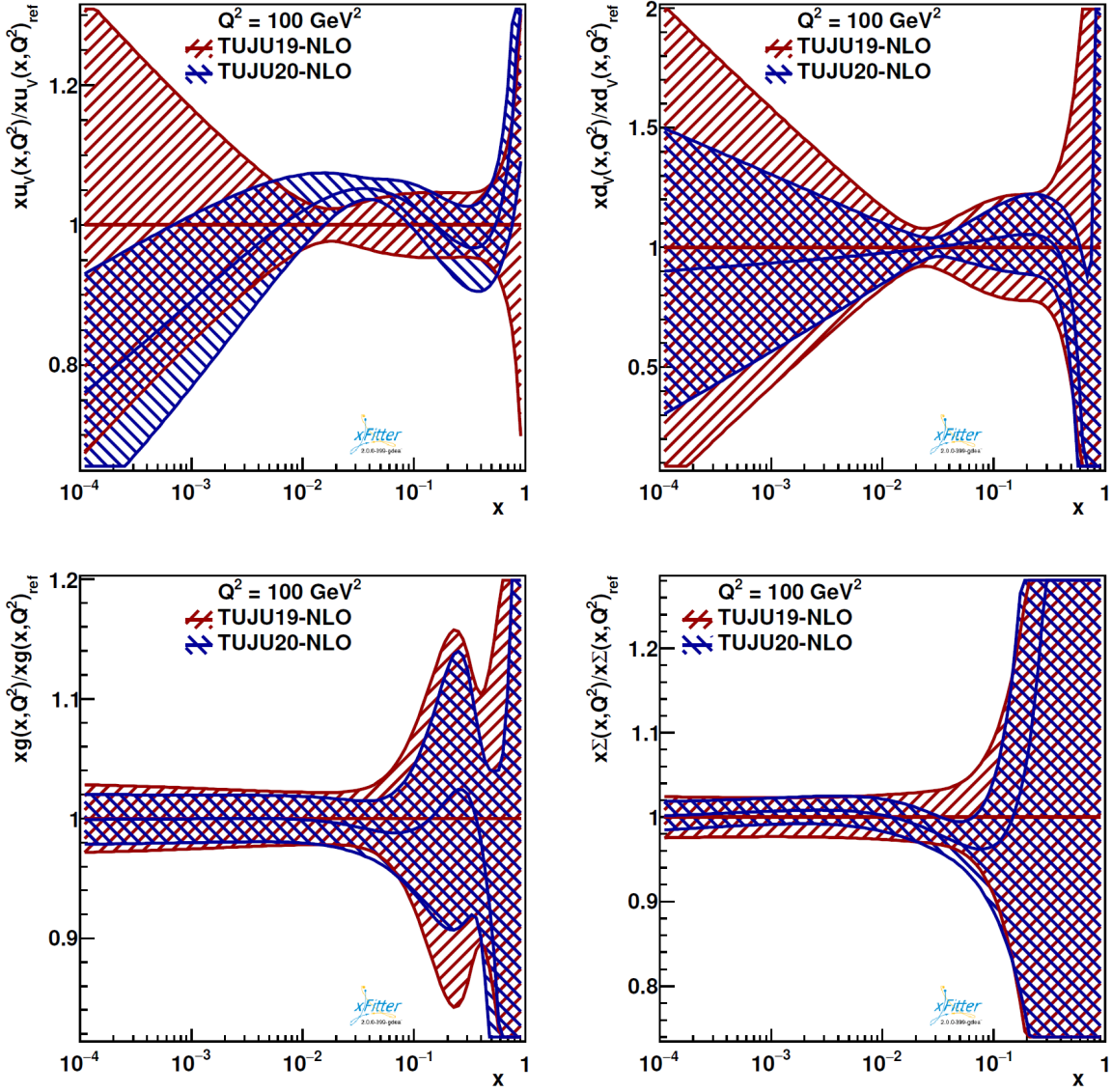


Figure 9.7.: Same as for figure 9.6, but at the energy scale $Q^2 = 100 \text{ GeV}^2$.

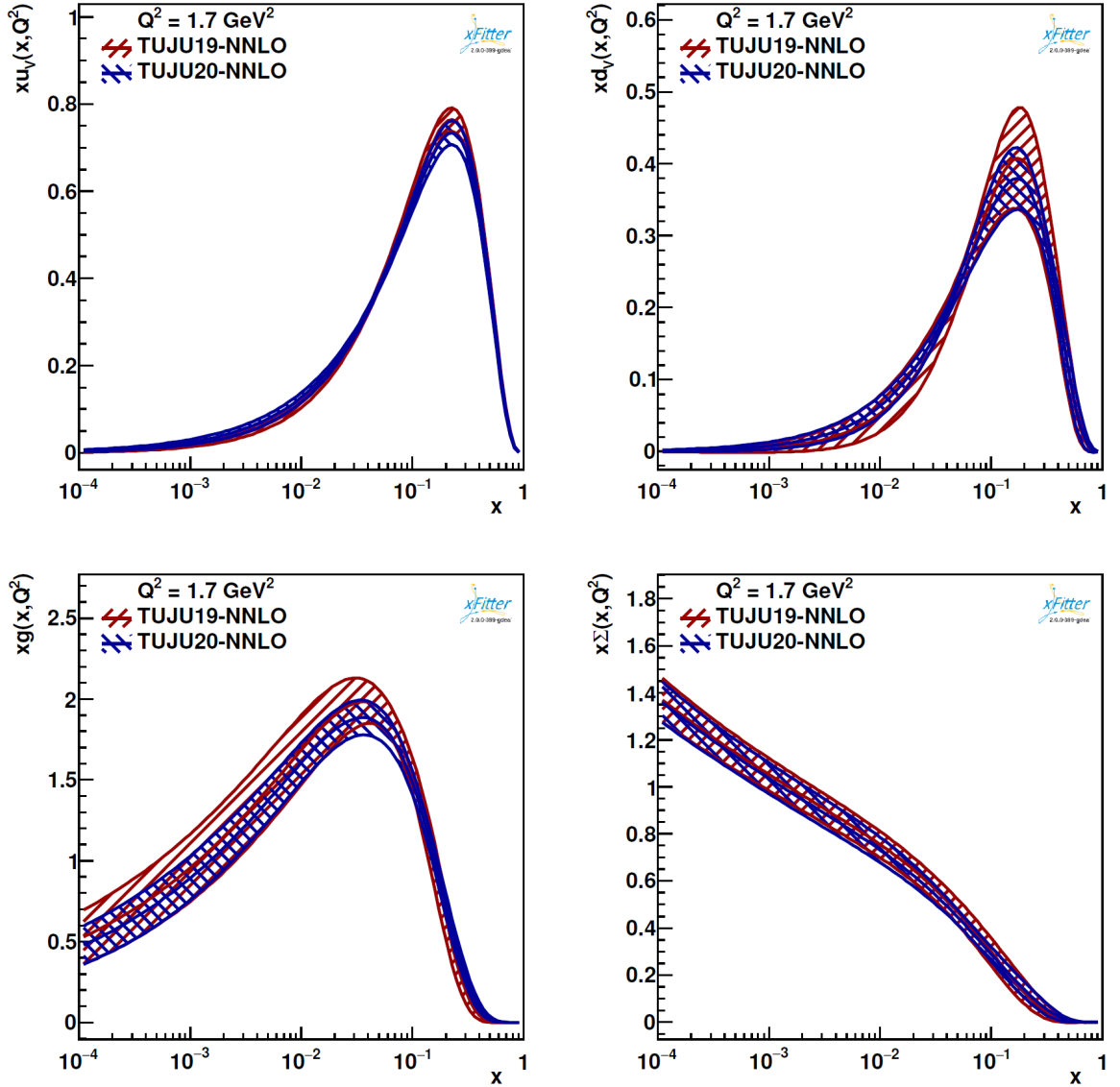


Figure 9.8.: Free proton PDF set obtained with DIS data only (TUJU19) compared to the updated free proton PDF set based on DIS and DY, W^\pm , Z boson production data (TUJU20) at NNLO. The representation is provided for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$ at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$.

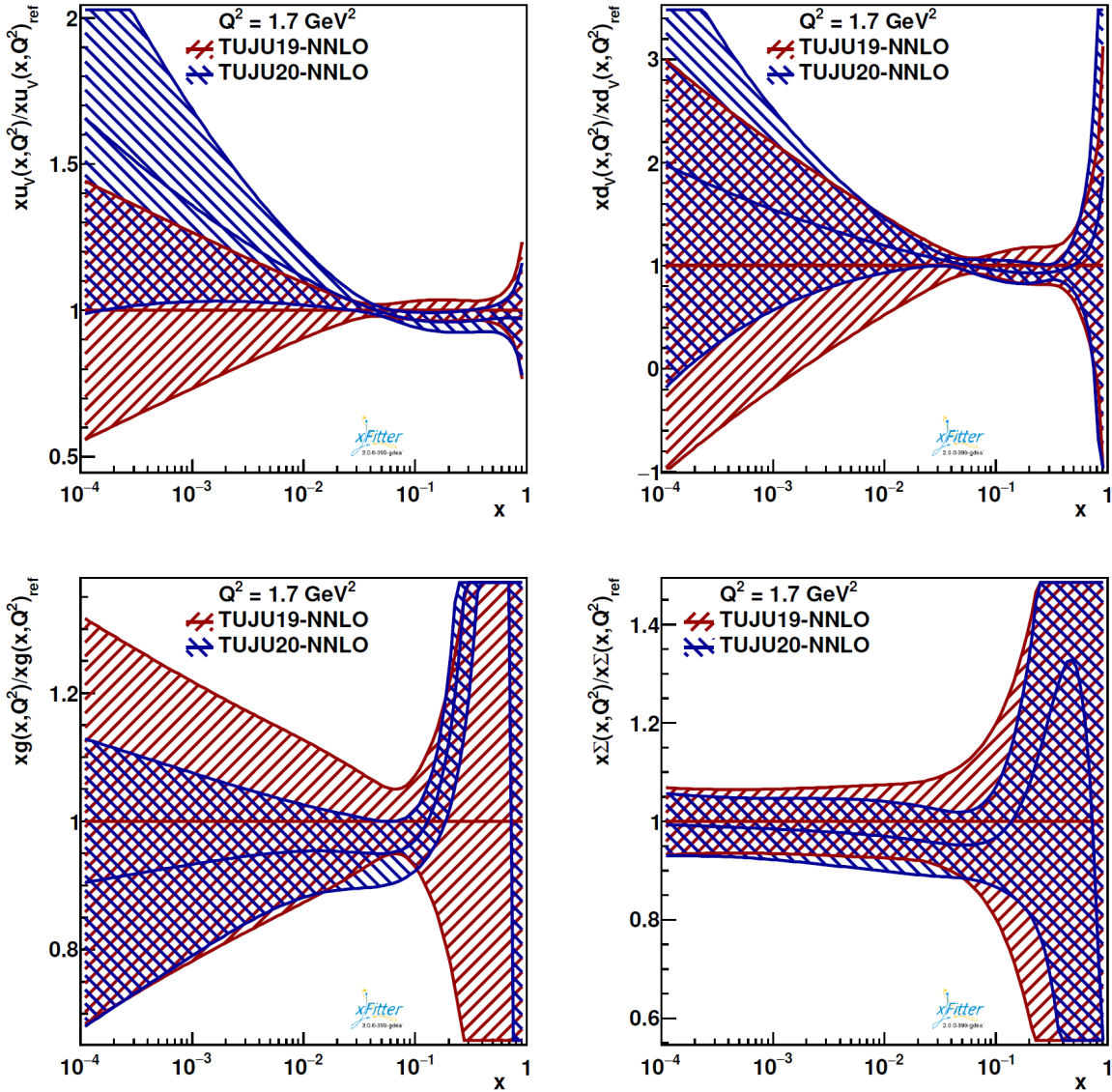


Figure 9.9.: Same content as in figure 9.8, but different visualization. The ratios $xf(x, Q^2)/xf(x, Q^2)_{\text{ref}}$ are shown for the free proton PDF set obtained with DIS data only referred to as TUJU19 ('ref') compared to the updated free proton PDF set based on DIS and DY, W^\pm , Z boson production data (TUJU20) at NNLO.

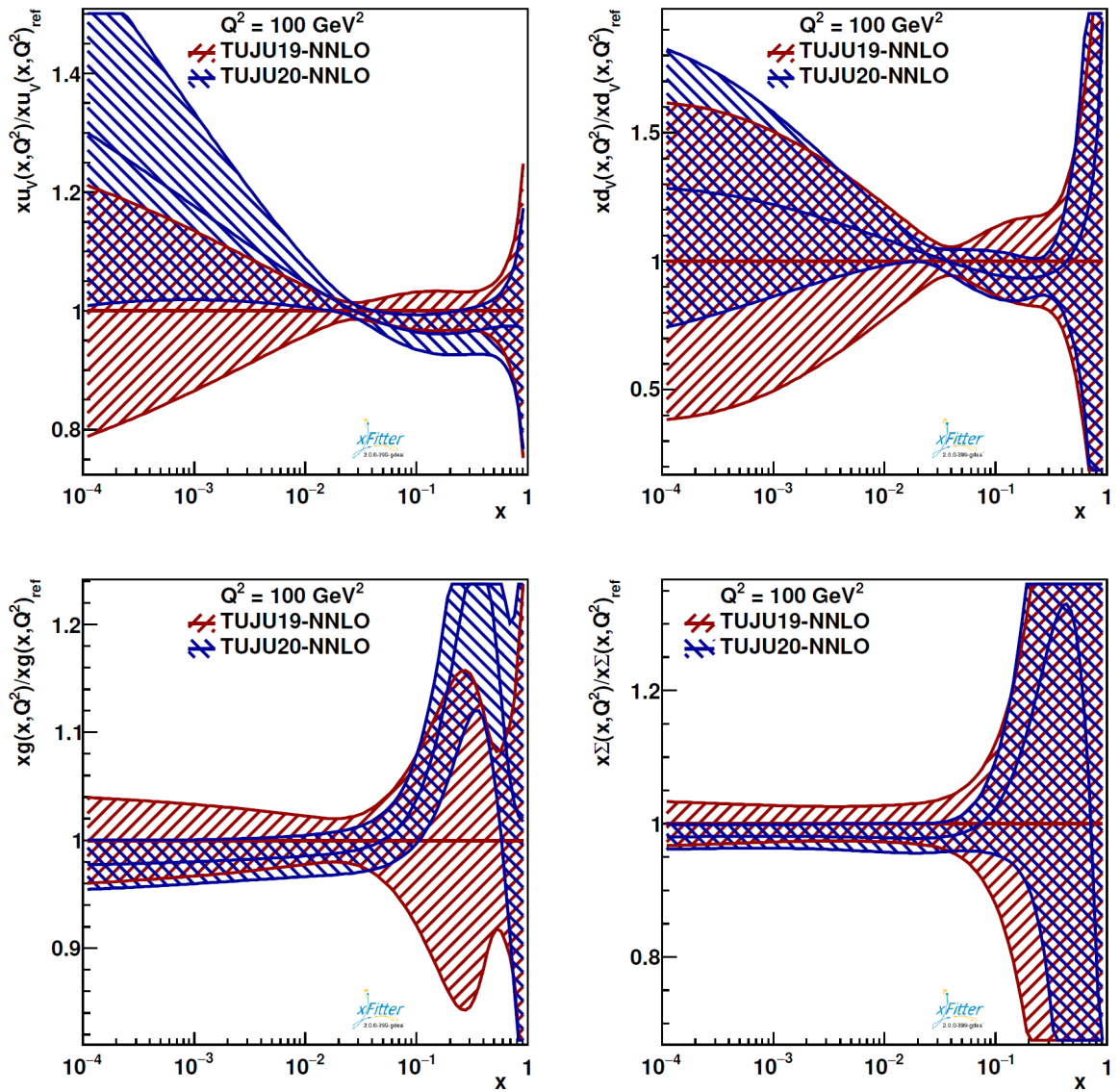


Figure 9.10.: Same as for figure 9.9, but at the energy scale $Q^2 = 100 \text{ GeV}^2$.

9.4. Outlook

As this is the first NNLO nPDF fit within the developed framework, only fixed-target DIS data with lepton and neutrino beams were included. In this chapter we have demonstrated that the implemented setup is capable to include Drell-Yan and $W^\pm Z$ boson production data in the next project phase. The application of TUJU19 nPDFs discussed in section 9.2 and the updated proton PDFs presented in section 9.3 confirm the motivation for this proposal. Furthermore, the W^\pm boson production data from p+Pb collisions at the LHC [245–248] are sensitive to the flavour decomposition and could therefore help to disentangle observed differences in valence quark nuclear modifications. Also, the Z boson production data and the fixed-target proton-nucleus DY dilepton data [241–244] could provide further constraints for the sea quark distributions. With the added flavour dependence some assumptions, e.g. $\bar{u} = \bar{d}$ could be released in the next project phase.

10. Summary and Outlook (engl.)

10.1. Summary

The subject of this thesis is the global analysis of nuclear parton distribution functions (nPDFs). The main goal of the project is to determine a new set of nPDFs at next-to-next-to-leading order (NNLO) and to make it publicly available for the scientific community. The purpose of nPDFs is to describe the collinear momentum distribution of the partons (quarks and gluons) inside a proton which is bound to a nucleus. Although this information is so important for the high-energy physics community, the existing knowledge is quite limited due to the complication that nuclei are bound objects with highly non-trivial dependencies between the single building blocks. The PDFs are rather well known for free protons, but are less well known for protons and neutrons which are bound in an atom's nucleus. The increase in perturbative precision in the nPDF determination enables the application of the most accurate calculations for the partonic scattering processes. Furthermore, the improved accuracy of nPDFs is important for the analysis of the heavy-ion experiments at the LHC and RHIC, and also to make predictions for future research projects like for example the EIC.

Since the PDFs cannot be calculated from first principles they are derived in a global analysis by incorporating suitable data for bound nucleons (protons and neutrons). The so-called pQCD analysis is performed based on the collinear factorization theorem and with the scale evolution described by the DGLAP equations. The nPDF analysis is an important proof for the validity of the factorization theorem and the universality of nuclear PDFs, which implies that the same parton distribution functions are valid for all partonic scattering processes.

The need to derive dedicated PDFs for nucleons bound in nuclei results from the fact that

nuclear effects, such as shadowing, anti-shadowing and EMC-effect, are observed when the cross sections are measured with nuclear targets (cf. fig. 3.3). The modifications of the measured observables by the nuclear effects are absorbed by the non-perturbative nPDFs, which distinguishes the nuclear PDFs from the free proton PDFs. The understanding of nuclear effects, based on several models usually valid for one particular effect, is an important subject of current research in high-energy particle physics. The nPDFs provide a way to compare the model predictions to nuclear modifications extracted from the experimental data via the global analysis. Furthermore, the nPDF analyses bring information on universality of such nuclear effects, i.e. if a particular effect is specific to a scattering system or observable. This can also constrain some models for nuclear effects, which is another motivation to improve the precision level of the nPDF analysis (here: NNLO) in order to provide a good basis of comparison.

There are joint efforts by several groups to derive nuclear parton distribution functions, as discussed in section 8.3. A summarizing overview of different selected criteria for the available LHAPDF sets for nPDFs is shown in figure 10.1. Common to all nPDF sets is that neutral current DIS data are used. Additionally, EPPS16 and TUJU19 (this analysis) were able to include the neutrino nucleus DIS measurements. As part of this analysis, the neutrino DIS data were implemented for the first time in a global pQCD analysis at NNLO. Here, the published nPDF sets are based on the neutrino data from the CDHSW and CHORUS experiments. Additionally, a fit including neutrino measurements by the NuTeV collaboration was performed as part of the presented studies. However, the inclusion of the measurements by the NuTeV collaboration does not seem to provide an additional benefit but leaves some doubts in regard to the bad agreement between the theoretical predictions and the experimental data. From the overall study of neutrino-nucleus DIS we conclude that lepton-nucleus and neutrino-nucleus DIS data can be accommodated in a global analysis without creating tensions between different sets, apart from the points related to the NuTeV data discussed in section 6.4.

A further aspect shown in figure 8.3 is that the existing LHAPDF nPDF sets produced by the groups nCTEQ15 and EPPS16 also include the Drell-Yan process into their analysis. Furthermore, the analysis performed by EPPS16 contains the information from the latest LHC experiments. In this work we have shown that the implemented setup is capable to include Drell-Yan and W^\pm, Z boson production data in the next project phase, as

nCTEQ15	EPPS16	nNNPDF1.0	TUJU19
Neutral current DIS			
	Neutrino nucleus DIS		Neutrino nucleus DIS
Drell-Yan data			
	LHC data		
CTEQ6	CTEQ14	NNPDF3.1	own proton fit
NLO		NLO, NNLO	NLO, NNLO
$f(x)$	$R_{A/p}$	$f'(x)$	$f^*(x)$
			Deuteron fitted as nucleus
			Open source toolset

Figure 10.1.: Comparison of the published LHAPDF sets from previous nPDF analyses to this work (TUJU19). Some selected criteria, like considered processes, free proton PDF baseline, form of parameterization etc. are shown in this overview. For a more detailed comparison please refer to table 8.1.

demonstrated based on the obtained results for the new proton PDF baseline (TUJU20).

As discussed earlier, the nuclear PDFs are usually determined for a specific *free* proton PDF set. Contrary to most previous analyses, our nPDF sets are based on a proton baseline fitted within the same framework, which guarantees consistency throughout the analysis concerning the series of choices on parameter values, assumptions, constraints and kinematic cuts that need to be made when performing a global analysis. The analysis is performed at next-to-leading order (NLO) and NNLO, which is provided by this analysis and the nPDF set recently created by the NNPDF group.

As part of the presented framework, a new form of parameterization has been used to determine the nuclear PDFs. For that matter deuteron has been considered being a nucleus with non-negligible nuclear effects. The complete analysis framework was implemented in an open source tool set by modifying the tool called xFITTER. The source code with all applied modifications required for the treatment of nuclear PDFs has been published, providing a first open-source tool for a global nuclear PDF analysis. A short user guide for nPDFs in xFITTER has been prepared and shared with the community.

The common framework paves the way for a simultaneous proton and nuclear PDF analysis in the future.

The resulting cross sections show very good agreement with the included experimental data, both for neutral-current and charged-current DIS processes, as confirmed by the resulting $\chi^2/N_{\text{dp}} \sim 1.0$ for the nuclear part of the analysis. The comparisons to the existing nPDF sets demonstrate a reasonable agreement within the error bands. The obtained results are consistent with the expectation that due to the consideration of DIS data only the nPDFs for valence quarks are well constrained by the experimental data, whereas gluon and sea quarks are constrained only indirectly by the included data and mostly by DGLAP evolution. The resulting PDFs for the free proton baseline, as well as the nPDFs have been published in the LHAPDF6 format including uncertainties for both, the proton baseline and the nuclear PDF analysis, derived with a Hessian uncertainty analysis. Additionally, the Monte Carlo method has been studied for the error analysis of nPDFs. The results of the Monte Carlo analysis and the specific options have been presented in this work. It could be shown that it can be applied as part of the implemented setup, but requires more replica sets and ideally experimental data from different processes, to achieve a better numerical convergence of the fits.

10.2. Outlook

As this is the first NNLO nPDF fit within the developed framework, only fixed-target DIS data with lepton and neutrino beams were included. In the future we plan to add data also for other observables for which theoretical calculations at NNLO exist. The W^\pm boson production from proton-lead (p+Pb) collisions at the LHC are sensitive to the flavour decomposition and could therefore help to disentangle observed differences in valence quark nuclear modifications. Furthermore, Z boson production data from the same experiments and the fixed-target proton-nucleus DY dilepton data would provide further constraints for the sea quark distributions. Even after these, direct gluon constraints will remain sparse. Recently it has been shown that such constraints could be obtained from the existing data for dijet and charmed-meson production in p+Pb collisions at the LHC, e.g. in [188]. Further in the future, an electron-ion collider (EIC) [121] would provide precision data for nPDF analyses. In order to get the best information

from these data the highest possible perturbative precision will be required, and our NNLO analysis is an important step in this context.

Furthermore, the implemented framework is common to free proton PDF and nuclear PDF analyses and therefore paves the way for a simultaneous proton and nuclear PDF fit in the future, supported by the fact that xFITTER already includes a lot of collected data for pp collisions.

11. Zusammenfassung und Ausblick (dt.)

11.1. Zusammenfassung

Präzise Kenntnis der inneren Struktur der sichtbaren Materie ist wichtig für die Forschungsarbeit in der Physik. Die Information über die Bestandteile der Kerne innerhalb eines Atoms ist maßgeblich für die Analyse der Experimente, das Verständnis der Natur, und um zuverlässige Voraussagen für die zukünftigen Forschungsprojekte machen zu können. Obwohl dieses Wissen so wichtig ist, ist dessen Verfügbarkeit sehr beschränkt, was auf die nicht-triviale Abhängigkeiten zwischen den einzelnen Bestandteilen in gebundenen Objekten, so wie Atomkernen, zurück zu führen ist. Die Hochenergie-Physik benötigt eine Beschreibung, wie der Gesamtimpuls eines Nukleons (z.B. Protons) auf seine Komponenten verteilt ist. Diese Komponenten (Quarks und Gluonen) werden auch "Partonen" genannt. Die benötigte Information über die Impulsverteilung wird in Form von Partonverteilungsfunktionen (PDFs, engl. parton distribution functions) angegeben. Die PDFs für *freie* Protonen sind relativ präzise bekannt, während die Beschreibung für Protonen, die in einem Atomkern gebunden sind, noch ausbaufähig ist. Die fehlende Präzision in diesen sogenannten *nuklearen* PDFs (nPDFs) führt zu theoretischen Ungenauigkeiten in den zugrundeliegenden Berechnungen für verschiedene Experimente, wie z. B. die Schwerionen-Programme des LHC (engl. Large Hadron Collider) am CERN.

Die Notwendigkeit dedizierter nuklearen PDFs für gebundene Protonen resultiert aus den experimentell beobachteten nuklearen Effekten (siehe Abb. 11.1). Die Modifizierung der Messgrößen durch die nuklearen Effekten wird durch die nicht-perturbativen nPDFs absorbiert, was diese von den PDFs eines freien Protons unterscheidet. Das Verständnis dieser nuklearen Effekte, basierend auf Modellen, die meist nur einen Effekt

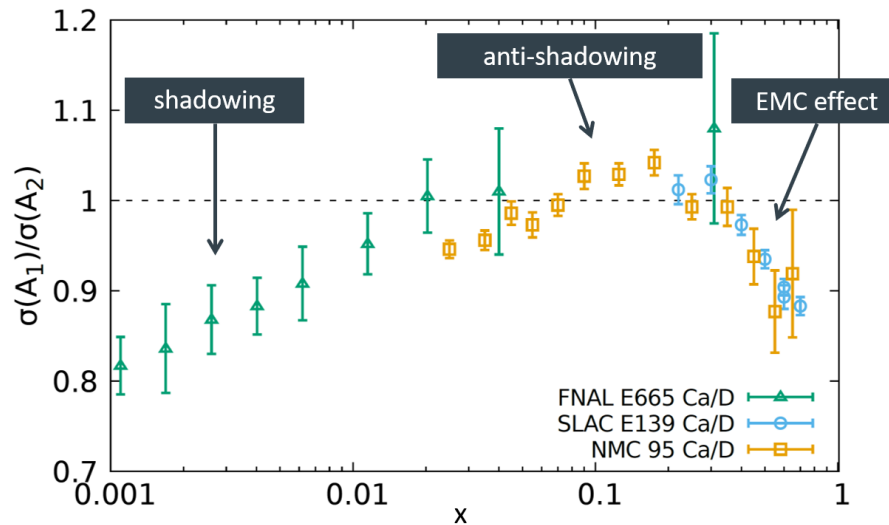


Abb. 11.1.: Beispielhafte Illustration der Datenpunkte gemessen von den FNAL, SLAC und NMC Kollaborationen. Der Verlauf des Verhältnisses $\sigma(A_1)/\sigma(A_2)$ zweier Streuquerschnitte für zwei verschiedene Kerne mit den Massenzahlen A_1, A_2 entlang der x -Achse ist charakterisiert durch die nuklearen Effekte benannt 'shadowing' (dt. Abschattung), 'anti-shadowing' (gegenläufiger Effekt zu 'shadowing') und 'EMC effect' (benannt nach der EMC Kollaboration, als es zum ersten Mal experimentell beobachtet wurde). Die gestrichelte Linie symbolisiert das Verhältnis $\sigma(A_1)/\sigma(A_2) \approx 1$, welches man ohne Modifizierung durch die nuklearen Effekte erwarten würde. (dt.)

erklären können, ist ein aktuelles Forschungsfeld in der Hochenergie-Physik. Die nPDFs bieten eine Möglichkeit die Modell-Vorhersagen mit den nuklearen Modifizierungen, die im Rahmen einer globalen Analyse aus den Daten extrahiert werden, zu vergleichen. Dies liefert eine weitere Motivation die Genauigkeit der nPDFs zu verbessern, um so eine präzisere Vergleichsbasis zu gewährleisten.

Neben dem akademischen und wissenschaftlichen Nutzen, hat die Kenntnis der inneren Struktur der Atomkerne bereits zu vielen Anwendungen in der Medizin geführt, wie zum Beispiel das bildgebenden Verfahren der Magnetresonanztomographie (MRT) oder die Methode der Krebsbehandlung mit schweren Ionen, die derzeit in Studien der Gesellschaft für Schwerionenforschung (GSI) in Darmstadt entwickelt und erbrobt wird. Ein anderes aktuelles Beispiel für eine Anwendung basierend auf den Kenntnissen der Kernphysik bietet die an der GSI entwickelte Kernuhr [3].

Für die Weiterentwicklung unserer Kenntnisse der innersten Struktur der Bestandteile eines Atomkerns werden nPDFs in globalen Analysen untersucht. Die fundamentale Theorie, welche die Wechselwirkung der Quarks und Gluonen beschreibt, ist die Quantenchromodynamik (QCD), die in dieser Arbeit einleitend vorgestellt wurde. Aufgrund der komplizierten Eigenschaften dieser Theorie können die PDFs *nicht* direkt mithilfe von grundlegenden Prinzipien berechnet werden. Stattdessen, werden in globalen Analysen zusätzlich zu theoretischen Berechnungen auch verschiedene experimentelle Daten genutzt. Typischerweise wählt man für die Parametrisierung der funktionalen Form der PDFs einen Ansatz auf einer anfänglichen, niedrigen Energieskala. Die freien Parameter werden dann durch den Vergleich mit experimentellen Daten aus vielen unterschiedlichen Energiebereichen bestimmt. Diese Prozedur erfordert die Berechnung der theoretischen Vorhersagen und eine iterative Anpassung der freien Parameter. Im mathematischen Sinne stellt dieser Vergleichsvorgang ein mehrdimensionales Minimierungsproblem dar. Während einer globalen Analyse, auch globaler Fit genannt, wird der Vorgang mehrere Tausend Mal wiederholt, um ein optimales Set an Parametern zu finden. Abschließend, bevor ein neues PDF-Set erstellt werden kann, wird eine Fehleranalyse durchgeführt.

Die Berechnung der theoretischen Vorhersagen basiert auf störungstheoretischen QCD (engl. pQCD). Der initiale Ansatz für die Parametrisierung der Funktion auf einer anfänglichen Energieskala wird mithilfe der sogenannten DGLAP Gleichungen zu höheren Energieskalen entwickelt. Diese Entwicklung und die dazugehörigen DGLAP Gleichungen sind ebenfalls als Teil dieser Arbeit beschrieben. Zusätzlich basiert die globale PDF Analyse auf dem kollinearen Faktorisierungstheorem, das es erlaubt die störungstheoretisch zugänglichen, partonischen Streuquerschnitte und die nicht-perturbativen PDFs separat zu ermitteln. Die globale Analyse der nPDFs bietet eine wichtige Überprüfung für die Gültigkeit des Faktorisierungstheorems und die Universalität der nPDFs, welche besagt, dass die selben PDFs für alle partonischen Streuprozesse gültig sind. Daher basiert die durchgeführte globale Analyse der nPDFs auf der störungstheoretischen Entwicklung und somit hängt die Genauigkeit der Ergebnisse von der gewählten Ordnung in Störungstheorie ab.

Dabei werden die Berechnungen umso schneller kompliziert, je mehr Korrekturen höherer Ordnung berücksichtigt werden. Es gibt einige nukleare Partonverteilungsfunktionen in

führender Ordnung (LO, engl. leading order), und die meisten etablierten modernen Fits existieren bis zur nächst-führender Ordnung NLO (engl. next-to-leading order). Das Hauptziel dieses Projekts ist es ein neues nPDF-Set bis zur zweiten Ordnung der Störungstheorie (NNLO, engl. next-to-next-to-leading order) zu ermitteln und für wissenschaftliche Zwecke öffentlich zugänglich zu machen. Die Verbesserung der störungstheoretischen Genauigkeit in der nPDF Analyse ermöglicht akkuratere Berechnungen der Streuquerschnitte auch in nuklearen Streuprozessen in einer konsistenten Art und Weise.

Für einen akkuraten nPDF Fit ist es wichtig experimentelle Daten aus verschiedenen Experimenten zu verwenden. Konzeptionell und numerisch ist es naheliegend die Studie der nuklearen Effekte in PDFs mit den Prozessen der tief-inelastischen Streuung (DIS, engl. deeply inelastic scattering) zu beginnen. In dieser Arbeit wurden Messdaten aus Lepton-Kern- und Neutrino-Kern-Streuexperimenten erfolgreich studiert und die finalen Ergebnisse wurden publiziert. Die Genauigkeit einer globalen nPDF Analyse hängt auch von den experimentellen Messunsicherheiten ab, welche analysiert werden, bevor ein neuen nPDF-Set veröffentlicht wird. Die veröffentlichten Ergebnisse wurden mit der der Hesse-Methode der Fehleranalyse, die auf einer quadratischen Näherung basiert, erzielt. Untersuchungen zu einer alternativen Vorgehensweise, Monte Carlo-Methode, wurden in diese Arbeit ebenfalls vorgestellt. Die neuen nPDF-Sets wurden in einem standardisierten LHAPDF Format veröffentlicht.

Bemühungen die nuklearen Partonverteilungsfunktionen zu ermitteln, werden derzeit von mehreren Gruppen unternommen (vgl. Abschnitt 8.3). Ein zusammenfassender Vergleich einiger ausgewählter Kriterien für die verfügbaren nuklearen LHAPDF Sets ist in Abbildung 11.2 gezeigt. Allen nPDF-Sets gemein ist, dass Daten aus DIS Prozessen mit neutralen Strömen verwendet wurden. Zusätzlich, konnten die Gruppen EPPS16 und TUJU19 (diese Arbeit) die Messdaten aus Neutrino-Kern-Streuexperimenten miteinbeziehen. Als Bestandteil dieser Arbeit wurden die Neutrino-DIS-Daten zum ersten Mal in einem globalen Fit bis zur Ordnung NNLO analysiert. Dabei basieren die publizierten nPDF Sets auf den Neutrino-Daten gemessen in den CDHSW und CHORUS Experimenten. Zusätzlich wurde eine Analyse unter der Berücksichtigung der Messdaten von der NuTeV Kollaboration durchgeführt. Die Einbeziehung dieser Daten scheint keinen zusätzlichen Nutzen für die Qualität der Analyse zu bieten und lässt gleichzeitig Zweifel

nCTEQ15	EPPS16	nNNPDF1.0	TUJU19
DIS Prozesse mit neutralen Strömen			
	Neutrino-Kern DIS		Neutrino-Kern DIS
Drell-Yan Daten			
	LHC Daten		
CTEQ6	CTEQ14	NNPDF3.1	Eigene Proton-PDFs
NLO		NLO, NNLO	NLO, NNLO
$f(x)$	$R_{A/p}$	$f'(x)$	$f^*(x)$
			Deuteron als Kern
			Open-Source-Programm

Abb. 11.2.: Vergleich der publizierten LHAPDF Sets aus vorhergehenden Analysen der nPDFs und dieser Arbeit (TUJU19). Die Übersicht zeigt einige ausgewählte Kriterien, wie z. B. betrachtete Prozesse, das Basis-PDF-Set für freie Protonen, Form der initialen Parametrisierung, usw. Eine detailliertere und umfangreichere Übersicht aller verglichenen Kriterien ist in Tabelle 8.1 enthalten. (dt.)

bezüglich der schlechten Übereinstimmung zwischen den theoretischen Vorhersagen und den experimentellen Daten offen. Mit Ausnahme der Einbeziehung der Datenpunkte aus dem NuTeV Experiment, gemäß der Diskussion im Abschnitt 6.4, folgern wir aus der durchgeführten Studie der DIS Prozesse mit Neutrino-Daten, dass Lepton-Kern und Neutrino-Kern DIS Daten in einer globalen Analyse vereinbar sind, ohne einen Widerspruch zwischen den verschiedenen Daten zu offenbaren.

Ein weiterer Punkt (vgl. Abb.11.2) ist, dass die LHAPDF-Sets nCTEQ15 und EPPS16 zusätzlich auf den Daten aus Drell-Yan Experimenten basieren. Darüber hinaus beinhaltet die Analyse der Gruppe EPPS16 die aktuellen Informationen aus den Experimenten am LHC. In dieser Dissertation wurde am Beispiel des überarbeiteten Basis-PDF-Sets für freie Protonen (TUJU20) demonstriert, dass mit der entwickelten Analyse-Systemlandschaft weitere Daten aus Drell-Yan Experimenten und Produktionsprozessen von W^\pm , Z -Bosonen in der nächsten Projektphase aufgenommen werden können.

Wie in Abbildung 11.2 gezeigt, basieren die nuklearen PDFs üblicherweise auf einem existierenden Basis-PDF-Set für *freie* Protonen. Im Gegensatz zu den früheren globalen Analysen, basiert unser nPDF-Set auf einem eigenen Basis-PDF-Set für freie Protonen, was für ein durchgehend konsistentes Vorgehen sorgt, insbesondere bezüglich der Auswahl von Parameter-Werten, Annahmen, beschränkenden Bedingungen, und der Auswahl der kinematischen Region der gültigen Messdaten. Die in dieser Arbeit präsentierte Analyse wurde bis zur ersten und zweiten Ordnungen in Störungstheorie (NLO und NNLO) durchgeführt, was nur in dieser Arbeit und in der kürzlich veröffentlichten Analyse durch NNPDF Gruppe gegeben ist.

Für die in dieser Arbeit vorgestellte Analyse wurde eine neue Form der initialen Parametrisierung verwendet, um die nuklearen PDFs herzuleiten. Auch das Deuteron wurde als ein Kern mit nicht-vernachlässigbaren nuklearen Effekten behandelt, was in den vorhergehenden Analysen nicht immer der Fall war. Die verwendete Systemlandschaft wurde für die Zwecke der nuklearen PDF Analyse erweitert und die implementierten Änderungen an dem Open-Source-Programm xFITTER wurden veröffentlicht. Damit ist es das erste öffentlich zugängliche Programm für eine globale Analyse von nPDFs. Ein kurzes Benutzerhandbuch für nPDFs in xFITTER wurde ebenfalls erstellt.

Die resultierenden Streuquerschnitte zeigen eine gute Vergleichbarkeit mit den verwendeten experimentellen Daten, für beide DIS-Prozesse, mit neutralen und geladenen Strömen, was anhand der Zahl $\chi^2/N_{\text{dp}} \sim 1.0$ für den nuklearen Teil der Analyse festgestellt wird. Der Vergleich mit den vorhergehenden nPDF-Sets weist eine angemessene Übereinstimmung innerhalb des zulässigen Fehlerbereichs auf. Die gewonnenen Ergebnisse stimmen mit der Erwartung überein, dass die verwendeten DIS-Daten nur für die Valenz-Quarks gute Zwangsbedingungen liefern, während die See-Quarks und die Gluonen nur indirekt, durch die DGLAP-Entwicklung entlang der Energieskala bestimmt sind. Die Basis-PDF-Sets für die freien Protonen, und die neuen nPDF-Sets wurden im standardisierten LHAPDF6 Format veröffentlicht.

11.2. Ausblick

Die in dieser Arbeit vorgestellte Analyse der nPDFs wurde zum ersten Mal bis zur zweiten Ordnung der Störungstheorie (NNLO) in der hier verwendeten und dafür erweiterten Systemlandschaft durchgeführt. Zunächst wurden nur experimentelle Daten aus der tief-inelastischen Streuung (DIS) von Leptonen und Neutrinos an Kernen betrachtet. Für die Zukunft ist es geplant weitere Messdaten, für welche theoretische Berechnungen bis zur Ordnung NNLO existieren, in die Analyse aufzunehmen. Die Dilepton-Daten aus Proton-Kern-Streuung in sogenannten Drell-Yan (DY) Experimenten, sowie die Messpunkte von Z -Bosonen-Produktion würden weitere Zwangsbedingungen für die Verteilung der See-Quarks liefern. Darüber hinaus sind die Messdaten für W^\pm -Bosonen-Produktion aus den Proton-Blei-Streuexperimenten am LHC sensibel für die Quark-Flavour-Zusammensetzung und könnten dazu beitragen die beobachteten Unterschiede in den nuklearen Effekten zwischen den Valenz-Quarks zu beschreiben. Es ist zu beachten, dass die bestimmenden Bedingungen für die Gluonen auch dann mangelhaft bleiben. In aktuellen Untersuchungen, zum Beispiel [188], wurde gezeigt, dass solche Bedingungen in den vorhandenen Daten aus den Proton-Blei-Streuexperimenten am LHC zu sogenannten Di-Jet-Prozessen und Charm-Meson-Produktionsprozessen enthalten sein könnten. Langfristig, wird das EIC Projekt (Elektronen-Ionen-Beschleuniger, engl. electron-ion collider) [121] präzise Messdaten für Untersuchungen der nPDFs liefern. Hierfür werden hoch-präzise Berechnungen bis zu höheren Ordnungen in Störungstheorie benötigt, um die bestmögliche Information aus den akkuraten Daten extrahieren zu können. Mit dieser Arbeit haben wir einen wichtigen Schritt in diese Richtung vollbracht.

Darüber hinaus ist das implementierte Setup gleichermaßen für die PDF Analyse der freien Protonen als auch für die der gebundenen Nukleonen gültig und anwendbar. Zusammen mit der Tatsache, dass in xFITTER bereits eine große Fülle an gesammelten experimentellen Daten zu Proton-Proton-Kollisionen zur Verfügung steht, bietet das implementierte Setup die Möglichkeit in der Zukunft eine simultane PDF Analyse für freie Protonen und für gebundene Nukleonen durchzuführen.

A. Final PDF Parameters

This chapter is taken over from the own publication [1]. Here we show the final parameters obtained for the proton and nuclear parton distribution functions. The naming convention corresponds to the PDF parameterization given in equations (5.1) and (5.2). Table A.1 provides the NLO parameters, while table A.2 presents the NNLO ones.

Some of the parameters were deliberately excluded from the fit. In most cases this means that in the initial version of the analysis procedure those parameters were used, but the obtained parameter value turned out to be very close to zero with very large uncertainty. Thus, that parameter was considered as not required. Alternatively, some of the nuclear parameters were never included as free parameters since the best fit criterion for nuclear PDFs, $\chi^2 \leq 1.0$, could be satisfied by the selected subset of the free parameters.

g	value	u_v	value	d_v	value	\bar{u}	value
$c_{0,0}^g$	7.0352	$c_{0,0}^{u_v}$	(SR)	$c_{0,0}^{d_v}$	(SR)	$c_{0,0}^{\bar{u}}$	(SR)
$c_{1,0}^g$	0.2871	$c_{1,0}^{u_v}$	0.6046	$c_{1,0}^{d_v}$	0.7376	$c_{1,0}^{\bar{u}}$	-0.1915
$c_{2,0}^g$	14.243	$c_{2,0}^{u_v}$	3.7064	$c_{2,0}^{d_v}$	2.9225	$c_{2,0}^{\bar{u}}$	7.5403
$c_{3,0}^g$	11.459	$c_{3,0}^{u_v}$	4.6595	$c_{3,0}^{d_v}$	-0.8736	$c_{3,0}^{\bar{u}}$	8.2448
$c_{4,0}^g$	-	$c_{4,0}^{u_v}$	-	$c_{4,0}^{d_v}$	-	$c_{4,0}^{\bar{u}}$	-
$c_{1,1}^g$	-50.064	$c_{1,1}^{u_v}$	-0.0616	$c_{1,1}^{d_v}$	-52.218	$c_{1,1}^{\bar{u}}$	-7.4250
$c_{1,2}^g$	-0.0008	$c_{1,2}^{u_v}$	0.4455	$c_{1,2}^{d_v}$	-0.1002	$c_{1,2}^{\bar{u}}$	-0.0021
$c_{2,1}^g$	-6.5209	$c_{2,1}^{u_v}$	-11.643	$c_{2,1}^{d_v}$	3.1722	$c_{2,1}^{\bar{u}}$	-0.2658
$c_{2,2}^g$	0.2039	$c_{2,2}^{u_v}$	0.0002	$c_{2,2}^{d_v}$	0.1336	$c_{2,2}^{\bar{u}}$	-0.2754

Table A.1.: Values of the NLO fit parameters at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. (SR) means that the normalization for that particular parton is fixed by the sum rules. A dash symbolizes that this parameter was excluded from the fit. Parameter values for the sea quarks, apart from \bar{u} , were derived from the applied constraints $\bar{s} = s = \bar{d} = \bar{u}$. Table from [1].

A. Final PDF Parameters

g	value	u_v	value	d_v	value	\bar{u}	value
$c_{0,0}^g$	6.2654	$c_{0,0}^{u_v}$	(SR)	$c_{0,0}^{dv}$	(SR)	$c_{0,0}^{\bar{u}}$	(SR)
$c_{1,0}^g$	0.2712	$c_{1,0}^{u_v}$	0.8060	$c_{1,0}^{dv}$	1.0227	$c_{1,0}^{\bar{u}}$	-0.1162
$c_{2,0}^g$	11.334	$c_{2,0}^{u_v}$	3.6897	$c_{2,0}^{dv}$	4.2717	$c_{2,0}^{\bar{u}}$	7.1632
$c_{3,0}^g$	5.0606	$c_{3,0}^{u_v}$	1.6388	$c_{3,0}^{dv}$	-0.6035	$c_{3,0}^{\bar{u}}$	-
$c_{4,0}^g$	-	$c_{4,0}^{u_v}$	-	$c_{4,0}^{dv}$	-	$c_{4,0}^{\bar{u}}$	2.4190
$c_{1,1}^g$	-2.4627	$c_{1,1}^{u_v}$	-0.1080	$c_{1,1}^{dv}$	-2.8603	$c_{1,1}^{\bar{u}}$	-3.2213
$c_{1,2}^g$	-0.0024	$c_{1,2}^{u_v}$	0.3766	$c_{1,2}^{dv}$	-0.0099	$c_{1,2}^{\bar{u}}$	-0.0123
$c_{2,1}^g$	-1.4764	$c_{2,1}^{u_v}$	74.620	$c_{2,1}^{dv}$	1.1235	$c_{2,1}^{\bar{u}}$	-0.0028
$c_{2,2}^g$	0.3704	$c_{2,2}^{u_v}$	-0.0001	$c_{2,2}^{dv}$	0.2357	$c_{2,2}^{\bar{u}}$	-0.9263

Table A.2.: Same as table A.1, but at NNLO. Table from [1].

B. User Guide for nPDFs in xFitter

B.1. Introduction

A new QCD analysis for nuclear parton distribution functions (nPDFs) at next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) was published in [1]. The framework of the analysis, including the form of the parameterization as well as the included DIS datasets, are discussed there. Also the results of that QCD analysis are compared to the existing nPDF sets and to the fitted data in reference [1]. The presented framework is based on xFITTER [190, 191] which has been modified to be applicable also for a nuclear PDF analysis. The purpose of this documentation is to provide a short user guide for nPDFs in xFITTER. For the general xFitter manual please refer to [263]. A summary of the required modifications can be found in [1].

B.2. nPDF Parameterization

Usually, the information on the mass number A and the proton number Z is individual per dataset since different nuclei can be involved in the measured scattering reaction. Therefore, it is recommended to provide A and Z for the particular nucleus inside the dataset file. This option is described in section B.3.

In order to activate the nPDF analysis one needs to set the following parameters in the **steering.txt** file:

1. PDFType = 'nucleus'
2. Anucleus = 1.0 [Mass number A , =1 for proton]
3. Znucleus = 1.0 [Proton number Z , =1 for proton]

Please note: there are several nuclei, and not a single one, involved in a global analysis. Therefore, it is strongly recommended to provide the information on the individual combinations of A and Z inside the data files. Here, in the steering.txt file, the parameters Anucleus, Znucleus are used only in case that this information is *not* provided in the data files.

As per the current implementation, the nPDFs are available only for the following combination of parameterization:

4. PDFStyle = 'CTEQ'
5. nTUJU = True [Flag to activate constraints, like e.g. sum rules, in the TUJU19 style [1]]

The extension 'n' in the name of the flag 'nTUJU' symbolizes that 'TUJU' framework has been developed for *nuclear* PDFs. However, the flag 'nTUJU' can be used for both PDFTypes proton and nucleus. If the user does not set the flag 'nTUJU = True' the program will run, but the results might be potentially inconsistent.

By selecting PDFStyle = 'CTEQ' the following parameterization is applied:

$$x f_i^{p/A}(x, Q_0^2) = c_0 x^{c_1} (1-x)^{c_2} (1 + c_3 x + c_4 x^2) \quad (\text{B.1})$$

with $i = g, u_v, d_v, \bar{u}, \bar{d}, s$ as per TUJU19 framework. A similar ansatz has been used to derive [169]. Note: For 'nTUJU=True' the number of parameters per parton flavour is limited to 5 (c_i with $i = 0, \dots, 4$).

The same form of the parameterization (B.1) is valid for both, proton and nuclear PDFs. The difference appears in regards to the parameters c_i ($i = 0, \dots, 4$). For nuclear PDFs the coefficients in equation (B.1) are further parameterized to be dependent on the nuclear mass number A as

$$c_k \rightarrow c_k(A) = c_{k,0} + c_{k,1} (1 - A^{-c_{k,2}}) \quad (\text{B.2})$$

with $k = 0, \dots, 4$. This form of A -dependent coefficients was used in the nCTEQ15 analysis [170]. This A -dependent parameterization has the advantage that in case of a free

proton ($A = 1$) the term $(1 - A^{-c_{k,2}})$ in equation (B.2) becomes zero and the functional form of a free proton is automatically retained. Here, again $k = 0, \dots, 4$ is valid only for 'nTUJU=True' setting.

As per the current implementation, nuclear, i.e. A -dependent coefficients are implemented only for the PDFStyle='CTEQ' as described above. For each coefficient in equation (B.1) one can have two additional A -dependent parameters as per equation (B.2).

In order to provide the initial parameters the **minuit.in.txt** file is used. The input format of coefficients $c_k \equiv c_{k,0}$ remains unchanged, i.e.:

Parameters	Parton
1 – 9	g
11 – 19	u_v
21 – 29	d_v
31 – 39	\bar{u}
41 – 49	\bar{d}
81 – 89	s
90 – 100	others

These parameters are what one could call coefficients of the PDF for the *free* proton. What is new for nPDFs are the A -dependent coefficients $c_{k,1}$ and $c_{k,2}$. For those the following parameter numbers have been reserved in the **minuit.in.txt** file:

B. User Guide for nPDFs in xFitter

Parameter #	Parameter name	Parton
111, 112	$c_{0,1}, c_{0,2}$	g
113, 114	$c_{1,1}, c_{1,2}$	
115, 116	$c_{2,1}, c_{2,2}$	
117, 118	$c_{3,1}, c_{3,2}$	
119, 120	$c_{4,1}, c_{4,2}$	
.....		
131, 132	$c_{0,1}, c_{0,2}$	u_v
133, 134	$c_{1,1}, c_{1,2}$	
135 - 140	...	
.....		
151, 152	$c_{0,1}, c_{0,2}$	d_v
153, 154	$c_{1,1}, c_{1,2}$	
155 - 160	...	
.....		
171, 172	$c_{0,1}, c_{0,2}$	$\bar{u} = \bar{d}$
173, 174	$c_{1,1}, c_{1,2}$	
175 - 180	...	

The $c_{k,1}$ and $c_{k,2}$ parameters for \bar{d} are determined by the constrain $\bar{u} = \bar{d}$. Due to the limited maximum number of parameter allowed in **minuit.in.txt** file, the $c_{k,1}$ and $c_{k,2}$ coefficients for s have been assigned to the following parameter space:

Parameter #	Parameter name	Parton
129, 130	$c_{0,1}, c_{0,2}$	$s = \bar{s}$
149, 150	$c_{1,1}, c_{1,2}$	
169, 170	$c_{2,1}, c_{2,2}$	
184, 185	$c_{3,1}, c_{3,2}$	
186, 187	$c_{4,1}, c_{4,2}$	

Please note, that the implemented routine has been validated only for 'nTUJU=True' and 'PDFStyle =CTEQ' with $c_{k,\max} = 4$. In the general case ('nTUJU=False') the according assignment might change or overlap (especially for s quarks).

Furthermore, please note, that with 'nTUJU=True' the number sum rule and the momentum sum rule are used to constrain the normalizations of d_v, u_v and \bar{u} . Furthermore, the constraint $\bar{u} = \bar{d} = s = \bar{s}$ is applied.

The nuclear parton distribution function $f_i^{N/A}$ for a *bound* nucleon inside a nucleus with mass number A is constructed from the *bound* proton's PDF $f_i^{p/A}$ (not from a free proton's PDF f^p). In particular for the distribution of partons in a bound nucleon we write

$$f_i^{N/A}(x, Q^2) = \frac{Z \cdot f_i^{p/A} + (A - Z) \cdot f_i^{n/A}}{A}, \quad (\text{B.3})$$

where Z is the number of protons in the nucleus. The PDF of the *bound* neutron $f_i^{n/A}$ is determined from the fitted proton's PDF using the isospin symmetry.

This nucleon decomposition has been implemented in xFitter for the calculation of DIS cross sections by using the following heavy-quark schemes only: 'ZMVFNS' scheme and 'FONLL' scheme. For the other schemes available in xFitter, the nucleon decomposition has not been implemented/validated.

B.3. Nuclear Data

For nuclear data, the experimental measurements are often published for a ratio of a cross section measured on one nucleus with mass number A_1 to the cross section of the other nuclear target A_2 , i.e. $\sigma(A_1)/\sigma(A_2)$ for cross sections or $F_2(A_1)/F_2(A_2)$ for structure functions. In such a case, inside the data file one would not provide a single observable, but a ratio of two observables. As part of the fitting routine, the quantities inside the data files are compared to the calculated theoretical values. In order to have a consistent comparison one needs to identify inside the data file if the bin 'Sigma' is a ratio or not. For a ratio it needs to be set:

```
CInfo = 'ratio'
```

```
DataInfo = 1.0 [and 0.0 for an absolute cross section].
```

The information on Z_1, A_1 and Z_2, A_2 (if applicable) is also provided inside the data file:

```
CInfo = 'A1', 'Z1', 'A2', 'Z2'
```

```
DataInfo = 56., 26.0, 12.0, 6.0 [an example for  $\sigma(Fe)/\sigma(C)$ ].
```

Besides that, some experiments apply isoscalar corrections to the measured data and publish only the modified information. Thus, the analysis procedure has been adapted so that the theoretically calculated quantities are consistent with the iso-corrected experimental data. For this purpose, different flags were introduced in xFITTER for the different forms of isoscalar corrections, which are specific to the corresponding experiments (CInfo='NMC', 'EMC', 'SLAC'). For the explicit form of the isoscalar correction please refer to the TUJU19 publication [1]. The information if or if not an isoscalar flag needs to be applied is provided by the particular experimental publication. In order to set the corresponding flag CInfo and DataInfo are used:

```
CInfo = 'NMC' [or 'EMC' or 'SLAC' respectively]
```

```
DataInfo = 1.0 [if True, and 0.0 if False].
```

Eventually, another modification on xFITTER was necessary for the treatment of charged current DIS processes measured in neutrino-nucleus scattering reactions. As part of this framework, the differential cross sections $d\sigma^2/dy dQ$ were used for the analysis. In order to identify the corresponding reaction one needs to specify inside the data file:

Reaction = 'antineutrino+p CC' [or Reaction = 'neutrino+p CC'].

Please note that these additional processes have been implemented only for the two heavy-quark schemes 'ZMVFNS' and 'FONLL'. When using another scheme available in xFitter an error message might probably occur.

B.4. Output and Error Analysis

When running the analysis routine for nPDFs in xFitter the file `minuit.out.txt` will automatically contain all final parameters (proton and nuclear parameters). Files containing PDFs, originally named `'pdfs_q2val_0i.txt'`, are named `'A-XXX_pdfs_q2val_0i.txt'` where XXX is the nuclear mass number (e.g. 2 for D, and 208 for Pb). During the Hessian error analysis, there are also PDF files generated, those do not carry an A -dependent extension, but are valid for the nucleus used last in the series of data files. This is due to the reason, that one would need an additional loop over all nuclei in the error analysis routine which is not implemented yet. A workaround exists by running the error analysis part ('DoBands=True') changing the series of data files, so that every time another nucleus is listed last. The same is valid for the generated output in the LHAPDF format. The central PDF members and the *.info files are generated for all nuclei and are stored in folders named `'A-XXX_xfitter_pdf'`, whereas the error set members are created only for the last nucleus on the list. This is valid for the Hessian error analysis only. If the Monte Carlo (MC) error analysis method is used, all nuclei are covered simultaneously. The text files containing PDFs and the grids in LHAPDF format are generated for the partons in a nucleon (*not* proton).

For nuclear PDFs the Hessian error analysis is usually performed with $\Delta\chi^2 > 1$. The error bands routine has been modified accordingly so that scaling based on the quadratic approximation is applied (please refer to [1] for more details). In order to set the $\Delta\chi^2$ parameter the command

```
set errdef 10.0 [for example]
```

can be used in the `minuit.in.txt` file.

B.5. Change Log

The modifications described here have been applied on xFitter versoin 2.0.1.

Modified header and include files:

- include/c_interface.inc
- include/dimensions.h
- include/indata.inc
- include/pdfparam.inc
- include/qcdnumhelper.inc
- include/steering.inc
- include/theo.inc
- include/theoexpr.inc
- include/xfitter_cpp.h

Modified source code files (bold items have been modified largely):

- interfaces/src/hf_pdf_calls.f
- src/c_interface.f
- src/chi2scan.cc
- src/dataset_tools.f
- **src/dis_sigma.f**
- src/error_bands_pumplin.f
- src/evolution.f
- **src/fcn.f**

- src/lhapdf6_output.c
- src/lhapdferrors.cc
- **src/pdf_param.f**
- src/read_data.f
- src/read_steer.f
- **src/sumrules.f**
- src/theory_dispatcher.f

New source code files:

- src/nucl_pdf.f
- src/nucl_pdfcc.cc (optional, requires gsl libraries)

Additionally, files Makefile.in and Makefile.am in the src/ folder have been modified in order to add new source code files.

C. Calculation of \sqrt{s} from Polarization

The experimental DIS data used for the analysis are summarised in section 6.1. Some of the available measurements require a special treatment due to the fact that different collaborations publish different types of information. The data published by the SLAC collaboration [5, 211, 216] are published for the differential cross sections $\frac{d^2\sigma}{dx dQ^2}(x, y, Q^2)$. In order to calculate the differential cross sections, which are then used for the comparison with the data, the exact knowledge of kinematic variables x , y and Q^2 applicable for the experiment is a prerequisite. For example, if the quantity y itself is not published but the information on the center of mass energy \sqrt{s} is available, the value of y can be calculated with relation (3.15). However, for the experimental SLAC data [5, 211, 216] neither y nor \sqrt{s} is given. Thus, the quantity ϵ provided by the experiment is used to derive the required information.

As per [5], the polarization of the virtual photon ϵ is given by relation

$$\epsilon = \left\{ 1 + 2 \left(1 + \left[\frac{\nu^2}{Q^2} \right] \right) \tan^2 \left(\frac{\theta}{2} \right) \right\}^{-1} \quad (\text{C.1})$$

which completely specifies the kinematics if used along with x and Q^2 .

The idea is to use relation (C.1) to derive the beam energy E which allows to calculate the center of mass energy \sqrt{s} by

$$\sqrt{s} = \sqrt{2EM} \quad (\text{C.2})$$

with the nucleon mass M .

In order to derive the beam energy E from relation (C.1) the terms ν^2/Q^2 and $\tan^2(\theta/2)$ need to be expressed by the variables x , Q^2 and E .

It is:

$$\nu = \frac{Q^2}{2Mx} \quad (\text{C.3})$$

and

$$E' = E + \nu. \quad (\text{C.4})$$

Furthermore we have

$$Q^2 = 2 E E' (1 - \cos \theta). \quad (\text{C.5})$$

Equation (C.5) can be used to eliminate the term $\tan^2(\theta/2)$ in relation (C.1) since in general one can write

$$\tan \frac{x}{2} = \sqrt{\frac{1 - \cos x}{1 + \cos x}}, \quad (\text{C.6})$$

which leads to

$$\tan^2 \left(\frac{\theta}{2} \right) = \frac{1 - \cos \theta}{1 + \cos \theta} = \left(\frac{4 E E'}{Q^2} - 1 \right)^{-1}. \quad (\text{C.7})$$

Finally, with (C.7) and (C.3) the polarization of the virtual photon ϵ given in (C.1) can be written as

$$\epsilon = \left\{ 1 + 2 \cdot \left(1 + \frac{Q^2}{4M^2x^2} \right) \cdot \left(\frac{4 E E'}{Q^2} - 1 \right)^{-1} \right\}^{-1} \quad (\text{C.8})$$

which now allows to determine the required beam energy

$$E = -\frac{Q^2}{4Mx} \pm \frac{1}{2} \left\{ \left(\frac{Q^2}{2Mx} \right)^2 + Q^2 \left[\frac{2}{(\epsilon^{-1} - 1)} \cdot \left(1 + \frac{Q^2}{4M^2x^2} \right) + 1 \right] \right\}^{1/2} \quad (\text{C.9})$$

where

$$E E' = E^2 + \frac{EQ^2}{2Mx} \quad (\text{C.10})$$

has been used as per equations (C.3) and (C.4).

D. MC Uncertainty Analysis: Scaling Properties

There are several options for the fine tuning when performing the Monte Carlo analysis with xFITTER. Setting an inadequate option or a combination of options can lead to invalid results. As part of the underlying PDF analysis, first all combinations of the available options had to be tested with proton PDFs in order to find the optimal set up.

The scaling properties summarised in table 5.1 can be applied independently for the statistical errors, uncorrelated systematic errors and the correlated systematic uncertainties. In order to find the optimal combination of the settings given in table 5.1 for the Monte Carlo method all allowed options have been tested, first for the proton PDFs. Figures D.1 - D.3 show some sample results considered as valid or invalid. In table D.1 the evaluated results are summarised and classified as "positive" or "negative" based on the visual examination of the results.

Since the scaling properties are directly connected to the fit quality through the definition of χ^2 their influence on the results is plausible in general. Specifically, the provided experimental errors are scaled (or not) dependent on the selected property. Thus, the error bands change in regard to the particular setting. Still, such a dramatic deviation of error bands from the central value of the best fit (like e.g. shown in figure D.2) is surprising. Thus, we conclude that the quality of the MC error method strongly depends on the amount of generated replicas and that considerably more replicas might be required to evaluate the final results of the MC error method.

StatScale	UncorSysScale	CorSysScale	
poisson	poisson	linear	ok
poisson	poisson	norescale	-
poisson	linear	linear	ok
poisson	linear	norescale	-
poisson	norescale	linear	ok
poisson	norescale	norescale	-
norescale	poisson	linear	positive
norescale	poisson	norescale	-
norescale	linear	linear	positive
norescale	linear	norescale	-
norescale	norescale	linear	positive
norescale	norescale	norescale	-

Table D.1.: Results of the examination of different scaling options for the Monte Carlo method, if the measured data (*not* the theoretical predictions) are varied. 'StatScale' is the scaling applied on statistical uncertainties, 'UncorSysScale' is the one for the uncorrelated systematic uncertainties and 'CorSysScale' is the scaling property for the correlated systematic errors.

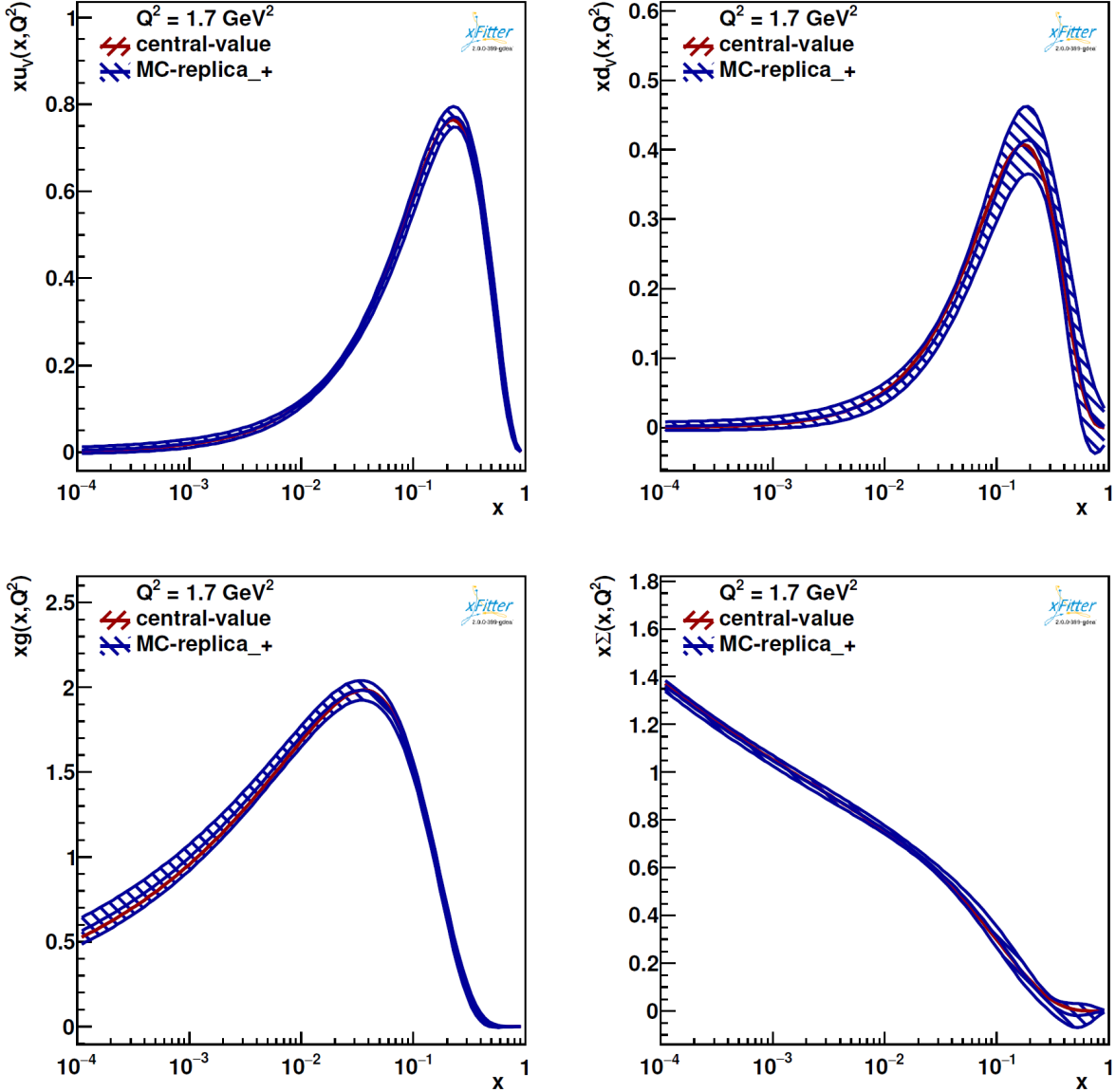


Figure D.1.: Error bands generated by Monte Carlo replicas for the combination `STATSCALE = 'norescale'`, `UNCORSYSSCALE = 'norescale'`, `CORSYSSCALE = 'linear'`. The best fit parton distributions (red) lie within the error bands, as expected. This representation has been classified to be 'positive', i.e. reasonable combination of options for the Monte Carlo error analysis.

D. MC Uncertainty Analysis: Scaling Properties

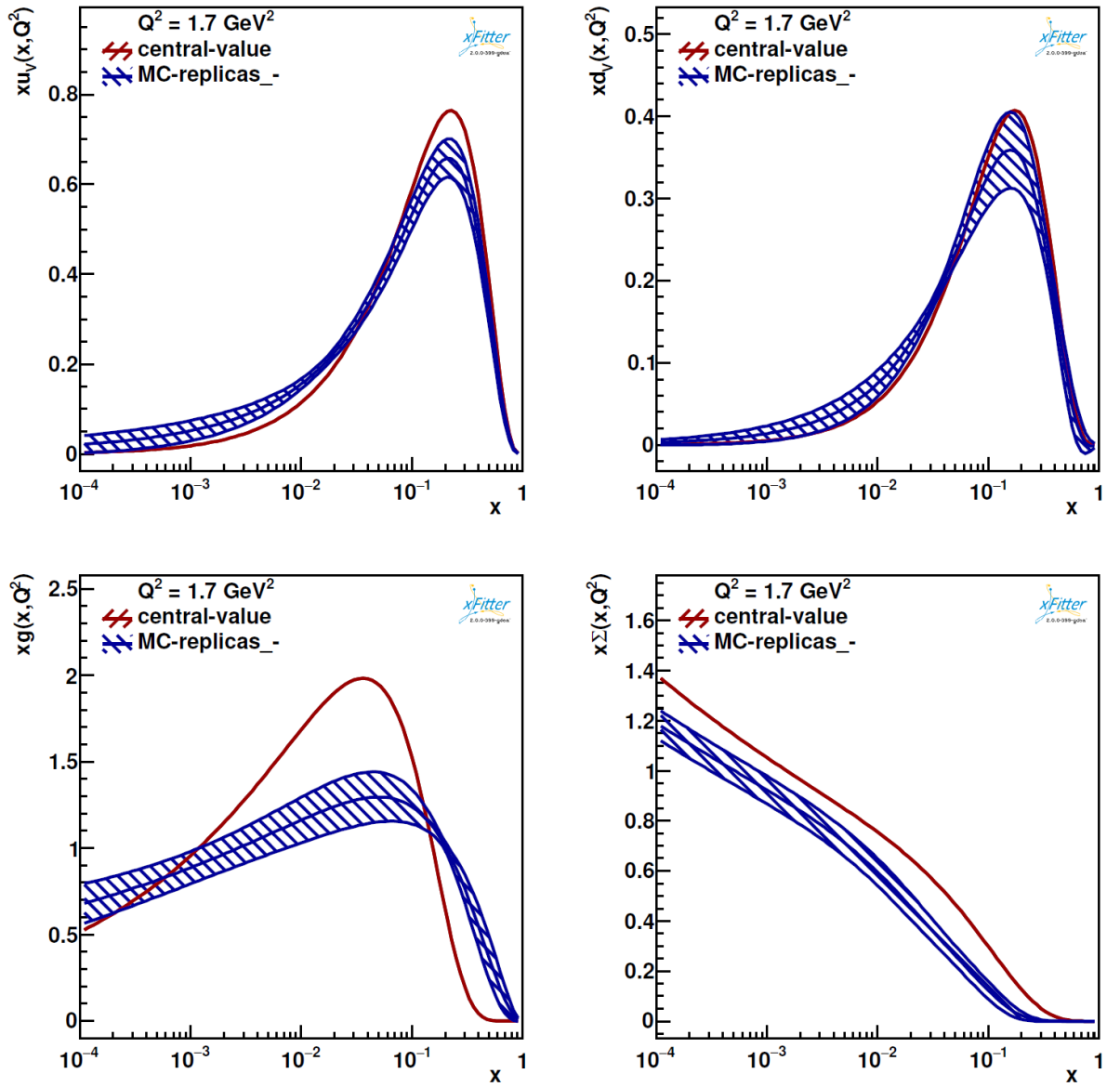


Figure D.2.: Error bands generated by Monte Carlo replicas for the combination `STATSCALE = 'norescale'`, `UNCORSYSSCALE = 'norescale'`, `CORSYSSCALE = 'norescale'`. This representation has been classified as 'negative', since the best fit parton distributions (red) lie outside the error bands, which is contradicting the definition of the central pdf set. This combination of options for the Monte Carlo error analysis is not recommended to be used for the underlying analysis.

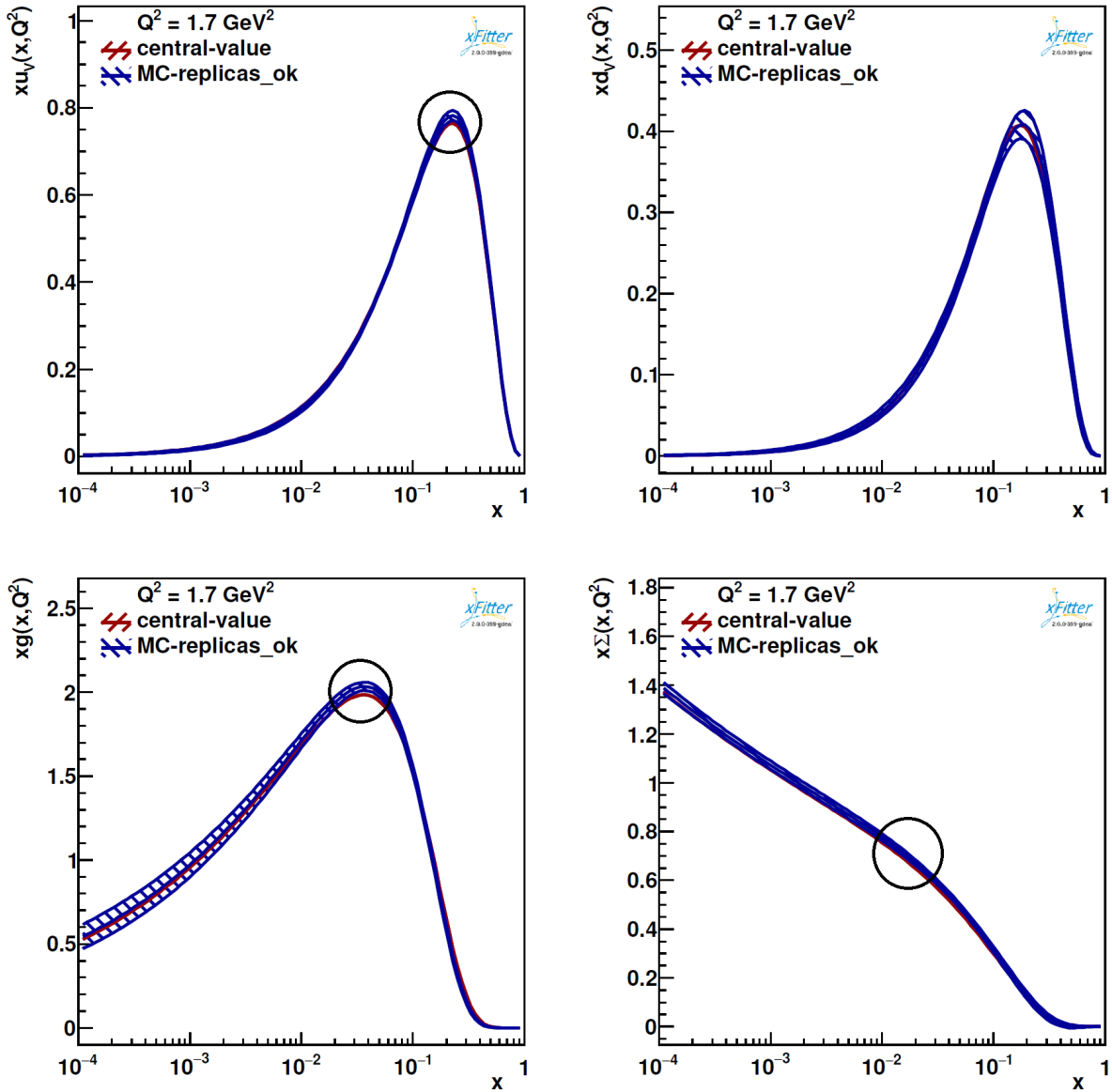


Figure D.3.: Error bands generated by Monte Carlo replicas for the combination `STATSCALE = 'poisson'`, `UNCORSYSSCALE = 'norescale'`, `CORSYSSCALE = 'linear'`. The best fit parton distributions (red) lie mostly within the error bands. Mismatching areas are highlighted by the black circles. This representation has been classified to be 'ok', i.e. potentially a valid combination of options for the Monte Carlo error analysis, since an improvement is expected if a higher number of MC replicas is used.

List of Figures

1.1.	Illustration of sample data points measured by the experimental collaborations FNAL [4], SLAC [5] and NMC [6]. The functional form of the ratio $\sigma(A_1)/\sigma(A_2)$ along momentum fraction x for nucleons in two different nuclei with mass numbers A_1, A_2 is characterized by the nuclear effects named shadowing, anti-shadowing and EMC effect. The dashed line symbolizes the ratio $\sigma(A_1)/\sigma(A_2) \approx 1$ which one would expect without nuclear effects. The observed deviation of $\sigma(A_1)/\sigma(A_2)$ from ≈ 1 due to the nuclear effects, that will be discussed later (cf. sec. 3.1), causes the need for nuclear parton distribution functions.	2
1.2.	A visual model of the atomic substructure. On the left-hand side an atom consisting of a nucleus and orbiting electrons is shown. Inside the nucleus there are nucleons - protons and neutrons. On the right-hand side the inner structure of a nucleon is indicated. It is made up of partons - quarks, i.e. sea quarks and valence quarks, and gluons.	3
1.3.	Particles in the Standard Model of particle physics.	3
1.4.	Timeline from 1808 to 1969. Milestones on the way to quarks and leptons. To be continued in figure 1.5.	4
1.5.	Timeline from 1969 to 2020. Summary of concepts and developments in theoretical particle physics in particular relevant for this work.	5
1.6.	High-level procedure of the nPDF analysis. The single steps will be used to visualize the relevance of the particular chapter in regard to the overall fitting procedure through the manuscript.	8
2.1.	The running coupling constant $\alpha_S(Q^2)$ as a function of the energy scale Q (here $\mu^2 = Q^2$). The function values of $\alpha_S(Q^2)$ at certain scale Q are fixed by the measurements, summarized here. Figure credited from [11]. . .	17

3.1.	Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: calculation of cross sections with the underlying theoretical principles are described.	19
3.2.	a) The expected functional form of structure function $F_2(x)$ along momentum fraction x dependent on the inner structure of a proton. For instance, in case a proton consists of three valence quarks only each of them carries a momentum fraction $1/3$ and therefore one would expect the contribution for $F_2(x) \neq 0$ only at $x = 1/3$. b) The actual functional form of structure function $F_2(x)$ as measured by different experiments (same as figure 3.5b, credited from [11]). The measurement corresponds to the lowest panel in a) and therefore shows that the three valence quarks must radiate partons at low x	21
3.3.	Illustration of sample data points measured by the experimental collaborations FNAL [4], SLAC [5] and NMC [6]. The functional form of the ratio $\sigma(A_1)/\sigma(A_2)$ along momentum fraction x for nucleons in two different nuclei with mass numbers A_1, A_2 is characterized by the nuclear effects named shadowing, anti-shadowing and EMC effect. The dashed line symbolizes the ratio $\sigma(A_1)/\sigma(A_2) \approx 1$ which one would expect without nuclear effects. The observed deviation of $\sigma(A_1)/\sigma(A_2)$ from ≈ 1 due to the nuclear effects causes the need for nuclear parton distribution functions.	23
3.4.	Feynman diagram for deeply inelastic charged lepton-nucleon scattering at leading order.	26
3.5.	Experimental data for the structure function $F_2(x, Q^2)$, measured by the collaborations HERA (H1+ZEUS), BCDMS, FNAL (E665) and NMC. The observed Q^2 -dependence of the structure function F_2 is called Bjorken-scale violation. Figure credited from [11]. a) Proton structure function $F_2(x, Q^2)$ along Q^2 -axis at different x values. The structure function for small and large x values is <i>not</i> constant along Q^2 , but shows a more or less linear dependence in $\log(Q^2)$. b) Proton structure function $F_2(x, Q^2)$ along x -axis at different Q^2 values. The scaling violations are particularly large at small x	31

3.6. Diagrams for the QCD corrections at NLO. Quarks are represented by arrows, spiral lines symbolize gluons, wavy line stands for a gauge boson (e.g. γ). a) processes with one quark in the initial state and one quark in the final state. b) processes with a real emitted gluon. c) processes with an initial state gluon. Graphs credited from [108]. 33

3.7. A selection of sample diagrams for the QCD corrections at NNLO. Quarks are represented by arrows, spiral lines symbolize gluons, wavy line stands for a gauge boson (e.g. γ). a) processes with one quark in the initial state and one quark in the final state. b) processes with a real emitted gluon. c) processes with an initial state gluon. Graphs credited from [108], supplementary modified for NNLO contributions by M. Walt. . . . 33

4.1. Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: scale evolution. 37

4.2. Feynman diagrams and expressions for the unpolarized splitting functions at LO. Figure credited from [109], supplementary modified by M. Walt. . . 39

5.1. Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: the framework of the overall process of global analysis, incl. initial parameterization, and the fitting part are explained. 51

5.2. Schematic view of the high-level xFITTER functionalities. xFITTER logo credited from [189]. Figure from [1] modified by M.Walt. 58

5.3. A graphical overview of modifications of the xFITTER program code performed for the purpose of nuclear PDF analysis. The implemented extensions are summarized in black boxes with the golden frame. For a schematic view of the original high-level xFITTER functionalities please refer to figure 5.2. xFITTER logo credited from [189]. 60

6.1. Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: the experimental data and the comparison to the theoretically calculated cross sections are documented. 65

6.2. Kinematic reach of experimental DIS data in the (x, Q^2) plane used to constrain the nuclear PDFs. Figure from [1]. 67

6.3.	Ratio F_2^n/F_2^p along x axis as parameterized by different experiments for the isoscalar modifications applied on the measured data. The form of these isoscalar modifications used by the NMC collaboration additionally depends on Q^2 , whereas it is constant in Q^2 for SLAC and EMC. Thus, the NMC F_2^n/F_2^p ratio is shown for some sample Q^2 values: the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ used in this work, the value of $Q_{\text{cut}}^2 = 3.5 \text{ GeV}^2$ at which kinematic cuts have been applied as part of the performed PDF analysis, and $Q^2 = 100 \text{ GeV}^2$	70
6.4.	Comparison of χ^2 values divided by the individual number of data points per dataset N_{dp} at NLO and NNLO. The "ideal" value $\chi^2/N_{\text{dp}} = 1.0$ is marked by the horizontal black dotted line. The bars in the diagram corresponding to $\chi^2/N_{\text{dp}} > 3.0$ have been truncated for the purpose of a clearer representation, which is symbolised by the dashed light-grey line. Figure from [1].	71
6.5.	Comparison to NMC $F_2(\text{D})$ data at different values of Q^2 at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].	72
6.6.	Comparison to FNAL E665 data for different ratios $F_2(A_1)/F_2(A_2)$ for nuclei with mass numbers A_1 and A_2 , at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].	73
6.7.	Comparison to NMC $F_2(A_1)/F_2(A_2)$ data measured for different combinations of nuclei with mass numbers A_1 and A_2 , at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].	73
6.8.	Comparison to NMC $F_2(\text{Sn})/F_2(\text{C})$ data at different values of Q^2 at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].	75
6.9.	Comparison to EMC data, first for the structure function F_2 at different Q^2 , and then for different ratios $F_2(A_1)/F_2(A_2)$ measured for nuclei with mass numbers A_1 and A_2 . The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1].	76

- 6.10. Comparison to SLAC data for different ratios of reduced differential cross sections $\sigma(A_1)/\sigma(A_2)$ for nuclei with mass numbers A_1 and A_2 , at different values of Q^2 at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1]. 77
- 6.11. Comparison to HERMES data for different ratios of reduced differential cross sections $\sigma(A_1)/\sigma(A_2)$ for nuclei with mass numbers A_1 and A_2 , at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1]. 77
- 6.12. Sample comparison to selected CHORUS data for CC deeply inelastic scattering on lead (Pb). We show the results for either neutrinos (ν) or anti-neutrinos ($\bar{\nu}$), for one y value (cf. legend) each at different beam energies (35, 70, 110 GeV). The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1]. 78
- 6.13. Sample comparison to selected CDHSW data for CC deeply inelastic scattering on iron (Fe). We show the results for either neutrinos (ν) or anti-neutrinos ($\bar{\nu}$), for one y value (cf. legend) each at different beam energies (38.9, 85.4, 144.3 GeV). The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). Figure from [1]. 78
- 6.14. Sample comparison to selected NuTeV data for CC deeply inelastic scattering on iron (Fe). We show the results for either neutrinos (ν) or anti-neutrinos ($\bar{\nu}$), for one y value (cf. legend) each at different beam energies (65, 190, 275 GeV). The calculated quantities are shown at NLO (dashed line, grey error bands) and NNLO (solid line, golden-coloured error bands). 80

6.15.	Comparison of χ^2 values divided by the individual number of data points per dataset N_{dp} at NLO. The obtained results are shown for the published nPDF set TUJU19 (column 'A'), but also for the separately performed fits with NC DIS data only (column 'B') and the one including NuTeV data (column 'C'). The blue bar visualizes the χ^2/N_{dp} value per data set. The columns '=B-A' and '=C-A' demonstrate the differences observed between the additional fits compared to the originally published results, where green background colour of the row means that the result was better, and red stands for a worse outcome.	82
6.16.	Same as for figure 6.15, but at NNLO	83
7.1.	Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: uncertainty analysis. . . .	85
7.2.	Sample DIS data set for the ratio of the structure functions $F_2(Pb)/F_2(C)$ measured by the NMC collaboration. The actual data, as well as the best-fit results at NLO and NNLO including Hessian error bands from [1] are shown here for reference purpose.	88
7.3.	Comparison of proton baseline PDFs at NLO with error bands obtained by the Hessian uncertainty analysis with different values of $\Delta\chi^2$ (1, 15, 20), shown for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$ at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. As expected, due to the quadratic nature of the tolerance criterion the error bands for $\Delta\chi^2 = 15$ and $\Delta\chi^2 = 20$ are very close to each other. With the method described above, for the proton baseline with 13 free fit parameters we select $\Delta\chi^2 = 20$ with 90% confidence level, $\Delta\chi^2 = 15$ would correspond to 68% confidence level. . . .	89
7.4.	Monte Carlo replicated data shown with the actual data points with error bars (black) and the best fit NLO result (solid line) with corresponding error bands (grey) obtained by the Hessian method. In 7.4(a) the actual data points (light-blue) have been varied. In 7.4(b) the calculated theoretical predictions (copper-coloured) have been varied.	91

7.5.	Comparison of error bands for the proton baseline PDFs at NLO computed based on different error methods at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. The ratios $xf(x, Q^2)/xf(x, Q^2)_{\text{ref}}$ are shown for the error bands obtained by the Hessian uncertainty analysis with different values of $\Delta\chi^2$ (1, 4, 15) compared to the uncertainty bands obtained by the Monte Carlo method ('ref'). The representation is provided for valence quarks u_v, d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$	93
7.6.	Same as for figure 7.5, but at the higher scale $Q^2 = 100 \text{ GeV}^2$	94
7.7.	Same as for figure 7.5, but at NNLO.	95
7.8.	Same as for figure 7.7, but at the higher scale $Q^2 = 100 \text{ GeV}^2$	96
7.9.	Uncertainty bands for nPDFs at NLO obtained with Hessian method compared to MC replicas at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. The Hessian error bands (red) were created with statistically motivated value of $\Delta\chi^2 = 50$ with 90% confidence level. For the MC error bands (blue) 100 converged MC replicas were used. The comparison is shown for valence quarks u_v, d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$	97
7.10.	Uncertainty bands for nPDFs at NLO obtained with Hessian method compared to MC replicas at the energy scale $Q^2 = 100 \text{ GeV}^2$. The ratios $xf(x, Q^2)/xf(x, Q^2)_{\text{ref}}$ are shown for the error bands obtained by 100 MC replicas (blue) compared to the Hessian uncertainty analysis (red) with $\Delta\chi^2 = 50$. The comparison is shown for valence quarks u_v, d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$	98
7.11.	Same as for figure 7.9, but at NNLO.	99
7.12.	Same as for figure 7.10, but at NNLO.	100
8.1.	Relevance of this chapter (highlighted in carmine) related to the particular steps in the high-level fitting procedure. Here: the overall results of the global analysis - PDFs.	101
8.2.	Proton baseline PDFs TUJU19 at NLO compared to the HERA2.0 results, shown at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at $Q^2 = 100 \text{ GeV}^2$ after DGLAP evolution. Figure from [1].	102
8.3.	Same as for figure 8.2, but at NNLO. Figure from [1].	103

8.4.	Nuclear parton distributions functions TUJU19 in different nuclei with the mass number A at NLO, shown at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at $Q^2 = 100 \text{ GeV}^2$ after DGLAP evolution. Figure from [1].	104
8.5.	Same as for figure 8.4, but at NNLO. Figure from [1].	104
8.6.	Ratios $R_i^{p/Pb}$ of parton distribution functions per parton flavour i in a proton bound in lead compared to a free proton p . The obtained ratios are shown at NLO and NNLO, both at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at a higher scale $Q^2 = 100 \text{ GeV}^2$. Figure from [1].	105
8.7.	Nuclear parton distribution functions TUJU19 (carmine) in lead at NLO compared to nCTEQ15 [170] (blue) and EPPS16 [188] (green), shown at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at a higher scale $Q^2 = 100 \text{ GeV}^2$. The comparison is presented for the distribution functions $xf_i(x, Q^2)$ per parton flavour $i = g, \bar{u}, u_v, d_v, V$, where V is the sum of valence quarks, in a bound proton in lead. Figure from [1].	107
8.8.	Nuclear parton distribution functions TUJU19 (carmine) in lead at NLO compared to the nPDF sets nCTEQ15 [170] (blue), EPPS16 [188] (green), and DSSZ [205] (dark grey) shown at the higher scale $Q^2 = 100 \text{ GeV}^2$. The comparison is presented per parton flavour i for the ratios $R_i^{p/Pb}$ of PDFs in a proton bound in lead compared to the PDFs in a free proton. Figure from [1].	108
8.9.	Nuclear parton distribution functions TUJU19 (carmine) in lead at NNLO compared to the LHAPDF set nNNPDF1.0 [235] (light brown), shown at our initial scale $Q_0^2 = 1.69 \text{ GeV}^2$ and at a higher scale $Q^2 = 100 \text{ GeV}^2$ for distribution functions xf_i , and at higher scale $Q^2 = 100 \text{ GeV}^2$ for the ratios $R_i^{p/Pb}$ of PDFs in a proton bound in lead compared to PDFs in a free proton. The comparison is presented for the gluon g and for the quark singlet $\Sigma = u + \bar{u} + d + \bar{d} + s + \bar{s}$ in a bound proton in lead. Figure from [1].	110
8.10.	Same as for figure 8.9, but for deuteron.	110

8.11. Nuclear parton distribution functions TUJU19 (carmine) in lead at NNLO compared to the results by the KA15 group [236] (dark grey), shown at the initial scale of KA15, $Q_0^2 = 2.0 \text{ GeV}^2$, and at a higher scale $Q^2 = 100 \text{ GeV}^2$. The comparison is presented for the distribution functions $xf_i(x, Q^2)$ with $i = g, \bar{s}, V$, where V is the sum of valence quarks, in a bound proton in lead. Figure from [1].	112
9.1. Feynman diagram for Drell-Yan process at leading order. The scattering partons (quarks) are coming from two different hadrons (symbolized by circles).	116
9.2. Schematic view of the high-level xFITTER functionalities considering Drell-Yan and W^\pm, Z boson production processes. xFITTER logo credited from [189].	117
9.3. Calculated cross sections for W^\pm boson production at NLO compared to the experimental data from LHC Run II [248] presented along pseudorapidity η^μ . The comparison is shown for the calculation with free proton PDF sets TUJU19 (green) and CT14 (purple), and with nPDF sets TUJU19 (red) and EPPS16 (grey). Figures prepared by I. Helenius, edited by M. Walt.	119
9.4. Cross sections for W^\pm boson production calculated with TUJU19 PDF sets at NLO and NNLO compared to the experimental data from LHC Run II [248] presented along pseudorapidity η^μ , with uncertainty bands. Figures with central values prepared by I. Helenius, edited by M. Walt to include error bands.	120
9.5. Free proton PDF set obtained with DIS data only (TUJU19) compared to the updated free proton PDF set based on DIS and DY, W^\pm, Z boson production data (TUJU20) at NLO. The representation is provided for valence quarks u_v, d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$ at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$	123
9.6. Same content as in figure 9.5, but different visualization. The ratios $xf(x, Q^2)/xf(x, Q^2)_{\text{ref}}$ are shown for the free proton PDF set obtained with DIS data only referred to as TUJU19 ('ref') compared to the updated free proton PDF set based on DIS and DY, W^\pm, Z boson production data (TUJU20) at NLO.	124

9.7.	Same as for figure 9.6, but at the energy scale $Q^2 = 100 \text{ GeV}^2$	125
9.8.	Free proton PDF set obtained with DIS data only (TUJU19) compared to the updated free proton PDF set based on DIS and DY, W^\pm , Z boson production data (TUJU20) at NNLO. The representation is provided for valence quarks u_v , d_v , gluon g and $\Sigma \equiv \bar{u} + \bar{d} + s$ at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$	126
9.9.	Same content as in figure 9.8, but different visualization. The ratios $xf(x, Q^2)/xf(x, Q^2)_{\text{ref}}$ are shown for the free proton PDF set obtained with DIS data only referred to as TUJU19 ('ref') compared to the updated free proton PDF set based on DIS and DY, W^\pm , Z boson production data (TUJU20) at NNLO.	127
9.10.	Same as for figure 9.9, but at the energy scale $Q^2 = 100 \text{ GeV}^2$	128
10.1.	Comparison of the published LHAPDF sets from previous nPDF analyses to this work (TUJU19). Some selected criteria, like considered processes, free proton PDF baseline, form of parameterization etc. are shown in this overview. For a more detailed comparison please refer to table 8.1.	133
11.1.	Beispielhafte Illustration der Datenpunkte gemessen von den FNAL, SLAC und NMC Kollaborationen. Der Verlauf des Verhältnisses $\sigma(A_1)/\sigma(A_2)$ zweier Streuquerschnitte für zwei verschiedene Kerne mit den Massenzahlen A_1, A_2 entlang der x -Achse ist charakterisiert durch die nuklearen Effekte benannt 'shadowing' (dt. Abschattung), 'anti-shadowing' (gegenläufiger Effekt zu 'shadowing') und 'EMC effect' (benannt nach der EMC Kollaboration, als es zum ersten Mal experimentell beobachtet wurde). Die gestrichelte Linie symbolisiert das Verhältnis $\sigma(A_1)/\sigma(A_2) \approx 1$, welches man ohne Modifizierung durch die nuklearen Effekte erwarten würde. (dt.)	138
11.2.	Vergleich der publizierten LHAPDF Sets aus vorhergehenden Analysen der nPDFs und dieser Arbeit (TUJU19). Die Übersicht zeigt einige ausgewählte Kriterien, wie z. B. betrachtete Prozesse, das Basis-PDF-Set für freie Protonen, Form der initialen Parametrisierung, usw. Eine detailliertere und umfangreichere Übersicht aller verglichenen Kriterien ist in Tabelle 8.1 enthalten. (dt.)	141

- D.1. Error bands generated by Monte Carlo replicas for the combination `STATSCALE = 'norescale'`, `UNCORSYSSCALE = 'norescale'`, `CORSYSSCALE = 'linear'`. The best fit parton distributions (red) lie within the error bands, as expected. This representation has been classified to be 'positive', i.e. reasonable combination of options for the Monte Carlo error analysis. . . . 161
- D.2. Error bands generated by Monte Carlo replicas for the combination `STATSCALE = 'norescale'`, `UNCORSYSSCALE = 'norescale'`, `CORSYSSCALE = 'norescale'`. This representation has been classified as 'negative', since the best fit parton distributions (red) lie outside the error bands, which is contradicting the definition of the central pdf set. This combination of options for the Monte Carlo error analysis is not recommended to be used for the underlying analysis. 162
- D.3. Error bands generated by Monte Carlo replicas for the combination `STATSCALE = 'poisson'`, `UNCORSYSSCALE = 'norescale'`, `CORSYSSCALE = 'linear'`. The best fit parton distributions (red) lie mostly within the error bands. Mismatching areas are highlighted by the black circles. This representation has been classified to be 'ok', i.e. potentially a valid combination of options for the Monte Carlo error analysis, since an improvement is expected if a higher number of MC replicas is used. 163

List of Tables

5.1. Properties that can be applied for the scaling of different kinds of uncertainties. The above form of χ^2 in equation (5.9) which was used in this analysis corresponds to a Poisson-like scaling for the statistical experimental uncertainties, whereas the systematic uncorrelated and correlated uncertainties are scaled linearly.	57
5.2. Overview of some selected criteria, with respect to the theoretical basics or in regard to the framework, compared for the free proton PDFs versus the nPDFs of a nucleon bound in a nucleus with the mass number A . The compared criteria are given in the left column. The reference in this manuscript, i.e. section, equation, figure, is provided in the right column.	63
6.1. Summary of experimental DIS data used to determine proton PDFs. In the last two columns the resulting χ^2 values at NLO and NNLO obtained in our analysis are provided. The overall agreement is found to be sufficiently good, and the total χ^2/N_{dp} is 1.184 at NLO and 1.225 at NNLO. Table from [1].	66
6.2. Summary of experimental DIS data used to determine the nuclear PDFs. In the last two columns the resulting χ^2 values at NLO and NNLO obtained in our analysis are provided. Table from [1].	74
8.1. Comparison of the most recent previous nPDF analyses and this work TUJU19. Different criteria like considered processes, parameterization and analysis details are listed in the left column and compared between the nPDF sets specified with the original publication in the headline. . .	114

A.1. Values of the NLO fit parameters at the initial scale $Q_0^2 = 1.69 \text{ GeV}^2$. (SR) means that the normalization for that particular parton is fixed by the sum rules. A dash symbolizes that this parameter was excluded from the fit. Parameter values for the sea quarks, apart from \bar{u} , were derived from the applied constraints $\bar{s} = s = \bar{d} = \bar{u}$. Table from [1].	145
A.2. Same as table A.1, but at NNLO. Table from [1].	146
D.1. Results of the examination of different scaling options for the Monte Carlo method, if the measured data (<i>not</i> the theoretical predictions) are varied. 'StatScale' is the scaling applied on statistical uncertainties, 'UncorSysScale' is the one for the uncorrelated systematic uncertainties and 'CorSysScale' is the scaling property for the correlated systematic errors.	160

Bibliography

- [1] M. Walt, I. Helenius, and W. Vogelsang. “Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO”. In: *Phys. Rev.* D100.9 (2019), p. 096015. arXiv: 1908.03355 [hep-ph].
- [2] M. Walt, I. Helenius, and W. Vogelsang. “A QCD analysis for nuclear PDFs at NNLO”. In: *PoS DIS2019* (2019), p. 039. arXiv: 1908.04983 [hep-ph].
- [3] J. Thielking et al. “Laser spectroscopic characterization of the nuclear-clock isomer ^{229m}Th ”. In: *Nature* 556.7701 (2018), pp. 321–325. arXiv: 1709.05325 [nucl-ex].
- [4] M. R. Adams et al. “Shadowing in inelastic scattering of muons on carbon, calcium and lead at low $x(\text{Bj})$ ”. In: *Z. Phys.* C67 (1995), pp. 403–410. arXiv: hep-ex/9505006 [hep-ex].
- [5] J. Gomez et al. “Measurement of the A-dependence of deep inelastic electron scattering”. In: *Phys. Rev.* D49 (1994), pp. 4348–4372.
- [6] P. Amaudruz et al. “A Reevaluation of the nuclear structure function ratios for D, He, Li-6, C and Ca”. In: *Nucl. Phys.* B441 (1995), pp. 3–11. arXiv: hep-ph/9503291 [hep-ph].
- [7] J. Ashman et al. “A Measurement of the Spin Asymmetry and Determination of the Structure Function $g(1)$ in Deep Inelastic Muon-Proton Scattering”. In: *Phys. Lett.* B206 (1988). [,340(1987)], p. 364.
- [8] J. T. Londergan. “Nucleon Resonances and Quark Structure”. In: *Int. J. Mod. Phys.* E18 (2009), pp. 1135–1165. arXiv: 0907.3431 [hep-ph].
- [9] S. D. Bass. “The Proton spin puzzle: Where are we today?” In: *Mod. Phys. Lett.* A24 (2009), pp. 1087–1101. arXiv: 0905.4619 [hep-ph].

- [10] A. Ekstedt, H. Ghaderi, G. Ingelman, and S. Leupold. “Towards solving the proton spin puzzle”. In: (2018). arXiv: 1808.06631 [hep-ph].
- [11] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98.3 (2018), p. 030001.
- [12] F. W. Clarke. “The Atomic Theory”. In: *Science* 18, Issue 460 (1903), pp. 513–529.
- [13] H.-W. Kirchhoff. *Vorstellungen vom Atom 1800 - 1934, von Dalton bis Heisenberg : zur Geschichte und Didaktik*. Köln: Aulis-Verl. Deubner, 2001, p. 133.
- [14] D. I. Mendeleev. *Grundlagen der Chemie (Aus. d. Russ. übers)*. St. Petersburg; Leipzig: Ricker, 1892, p. 1126.
- [15] J. J. Thomson. “Cathode rays”. In: *Phil. Mag. Ser.5* 44 (1897), pp. 293–316.
- [16] E. Rutherford and H. Geiger. “Die Streuung von alpha und beta Teilchen und die Struktur des Atoms”. In: *Phys. Zeits* 10 (1909), pp. 1–6.
- [17] J. Chadwick. “Possible Existence of a Neutron”. In: *Nature* 129 (1932), p. 312.
- [18] J. Chadwick. “The Existence of a Neutron”. In: *Proc. Roy. Soc. Lond.* A136.830 (1932), pp. 692–708.
- [19] J. G. Taylor. “Hunting the quark”. In: *Science J.* 4N12 (1968), pp. 61–66.
- [20] E. M. Riordan. “The Discovery of quarks”. In: *Science* 256 (1992), pp. 1287–1293.
- [21] R. P. Feynman. “Very high-energy collisions of hadrons”. In: *Phys. Rev. Lett.* 23 (1969). [494(1969)], pp. 1415–1417.
- [22] J. D. Bjorken and E. A. Paschos. “Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon”. In: *Phys. Rev.* 185 (1969), pp. 1975–1982.
- [23] R. K. Ellis, H. Georgi, M. Machacek, H. D. Politzer, and G. G. Ross. “Perturbation Theory and the Parton Model in QCD”. In: *Nucl. Phys.* B152 (1979), pp. 285–329.
- [24] J. C. Collins, D. E. Soper, and G. F. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. arXiv: hep-ph/0409313 [hep-ph].

-
- [25] L. N. Lipatov. “The parton model and perturbation theory”. In: *Sov. J. Nucl. Phys.* 20 (1975). [*Yad. Fiz.*20,181(1974)], pp. 94–102.
- [26] V. N. Gribov and L. N. Lipatov. “Deep inelastic e p scattering in perturbation theory”. In: *Sov. J. Nucl. Phys.* 15 (1972). [*Yad. Fiz.*15,781(1972)], pp. 438–450.
- [27] G. Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. In: *Nucl. Phys.* B126 (1977), pp. 298–318.
- [28] Y. L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov. Phys. JETP* 46 (1977). [*Zh. Eksp. Teor. Fiz.*73,1216(1977)], pp. 641–653.
- [29] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. “Advantages of the Color Octet Gluon Picture”. In: *Phys. Lett.* 47B (1973), pp. 365–368.
- [30] D. J. Gross and F. Wilczek. “Ultraviolet Behavior of Nonabelian Gauge Theories”. In: *Phys. Rev. Lett.* 30 (1973). [,271(1973)], pp. 1343–1346.
- [31] D. J. Gross and F. Wilczek. “ASYMPTOTICALLY FREE GAUGE THEORIES. 2.” In: *Phys. Rev.* D9 (1974), pp. 980–993.
- [32] S. Weinberg. “Nonabelian Gauge Theories of the Strong Interactions”. In: *Phys. Rev. Lett.* 31 (1973), pp. 494–497.
- [33] J. C. Collins and D. E. Soper. “Parton Distribution and Decay Functions”. In: *Nucl. Phys.* B194 (1982), pp. 445–492.
- [34] M. R. Whalley, D. Bourilkov, and R. C. Group. “The Les Houches accord PDFs (LHAPDF) and LHAGLUE”. In: *HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B.* 2005, pp. 575–581. arXiv: hep-ph/0508110 [hep-ph].
- [35] D Bourilkov, R. C. Group, and M. R. Whalley. “LHAPDF: PDF use from the Tevatron to the LHC”. In: *TeV4LHC Workshop - 4th meeting Batavia, Illinois, October 20-22, 2005.* 2006. arXiv: hep-ph/0605240 [hep-ph].
- [36] A. Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *Eur. Phys. J.* C75 (2015), p. 132. arXiv: 1412.7420 [hep-ph].

- [37] M. Diemoz, F. Ferroni, E. Longo, and G. Martinelli. “Parton Densities from Deep Inelastic Scattering to Hadronic Processes at Super Collider Energies”. In: *Z. Phys. C* 39 (1988), p. 21.
- [38] K. J. Eskola, V. J. Kolhinen, and P. V. Ruuskanen. “Scale evolution of nuclear parton distributions”. In: *Nucl. Phys. B* 535 (1998), pp. 351–371. arXiv: hep-ph/9802350 [hep-ph].
- [39] K. J. Eskola, V. J. Kolhinen, and C. A. Salgado. “The Scale dependent nuclear effects in parton distributions for practical applications”. In: *Eur. Phys. J. C* 9 (1999), pp. 61–68. arXiv: hep-ph/9807297 [hep-ph].
- [40] M. Vanttinen, G. Piller, L. Mankiewicz, W. Weise, and K. J. Eskola. “Nuclear quark and gluon distributions in coordinate space”. In: *Eur. Phys. J. A* 3 (1998), pp. 351–359. arXiv: hep-ph/9808330 [hep-ph].
- [41] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 9780201503975, 0201503972.
- [42] F. Halzen and A. D. Martin. *Quarks and Leptons: An Introductory Course In Modern Particle Physics*. New York, USA: Wiley, 1984, 396p. ISBN: 0471887412, 9780471887416.
- [43] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. Cambridge University Press, 1996.
- [44] R. Aaij et al. “Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays”. In: *Phys. Rev. Lett.* 115 (2015), p. 072001. arXiv: 1507.03414 [hep-ex].
- [45] R. Aaij et al. “Model-independent evidence for $J/\psi p$ contributions to $\Lambda_b^0 \rightarrow J/\psi p K^-$ decays”. In: *Phys. Rev. Lett.* 117.8 (2016), p. 082002. arXiv: 1604.05708 [hep-ex].
- [46] R. Aaij et al. “Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$ ”. In: *Phys. Rev. Lett.* 122.22 (2019), p. 222001. arXiv: 1904.03947 [hep-ex].
- [47] A. Ali et al. “First Measurement of Near-Threshold J/Ψ Exclusive Photoproduction off the Proton”. In: *Phys. Rev. Lett.* 123.7 (2019), p. 072001. arXiv: 1905.10811 [nucl-ex].

-
- [48] W. K. H. Panofsky. “ELECTROMAGNETIC INTERACTIONS: LOW q^{*2} ELECTRODYNAMICS: ELASTIC AND INELASTIC ELECTRON (AND MUON) SCATTERING”. In: *Proc. of Int. Conf. on High-Energy Physics, Vienna, 1968. Geneva, CERN, 1968. p. 23-39.* 1968, pp. 23–39.
- [49] E. D. Bloom et al. “High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees”. In: *Phys. Rev. Lett.* 23 (1969), pp. 930–934.
- [50] M. Breidenbach et al. “Observed Behavior of Highly Inelastic electron-Proton Scattering”. In: *Phys. Rev. Lett.* 23 (1969), pp. 935–939.
- [51] Y. V. Kovchegov and E. Levin. “Quantum chromodynamics at high energy”. In: *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 33 (2012), pp. 1–350.
- [52] T. van Ritbergen, J. A. M. Vermaseren, and S. A. Larin. “The Four loop beta function in quantum chromodynamics”. In: *Phys. Lett.* B400 (1997), pp. 379–384. arXiv: hep-ph/9701390 [hep-ph].
- [53] H. Meyer. *Lattice QCD: A brief Introduction. (eds) Lattice QCD for Nuclear Physics. Lecture Notes in Physics.* Springer, Cham, 2015.
- [54] C. G. Callan Jr. and D. J. Gross. “High-energy electroproduction and the constitution of the electric current”. In: *Phys. Rev. Lett.* 22 (1969), pp. 156–159.
- [55] R. G. Roberts. *The Structure of the proton: Deep inelastic scattering.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1994. ISBN: 9780521449441, 9781139242448.
- [56] G. Sterman et al. *Handbook of Perturbative QCD.* College Park, Maryland: CTEQ, 2001.
- [57] N. Armesto. “Nuclear shadowing”. In: *J. Phys.* G32 (2006), R367–R394. arXiv: hep-ph/0604108 [hep-ph].
- [58] G. Piller, W. Ratzka, and W. Weise. “Phenomenology of nuclear shadowing in deep inelastic scattering”. In: *Z. Phys.* A352 (1995), pp. 427–439. arXiv: hep-ph/9504407 [hep-ph].
- [59] S. J. Brodsky, I. Schmidt, and J.-J. Yang. “Nuclear antishadowing in neutrino deep inelastic scattering”. In: *Phys. Rev.* D70 (2004), p. 116003. arXiv: hep-ph/0409279 [hep-ph].

- [60] J. Ashman et al. “A Measurement of the ratio of the nucleon structure function in copper and deuterium”. In: *Z. Phys.* C57 (1993), pp. 211–218.
- [61] N. Armesto. “A Simple model for nuclear structure functions at small x in the dipole picture”. In: *Eur. Phys. J.* C26 (2002), pp. 35–43. arXiv: hep-ph/0206017 [hep-ph].
- [62] N. N. Nikolaev and B. G. Zakharov. “Color transparency and scaling properties of nuclear shadowing in deep inelastic scattering”. In: *Z. Phys.* C49 (1991). [733(1990)], pp. 607–618.
- [63] V. N. Gribov. “Glauber corrections and the interaction between high-energy hadrons and nuclei”. In: *Sov. Phys. JETP* 29 (1969). [Zh. Eksp. Teor. Fiz.56, 892(1969)], pp. 483–487.
- [64] V. N. Gribov. “Interaction of gamma quanta and electrons with nuclei at high-energies”. In: *Sov. Phys. JETP* 30 (1970). [Zh. Eksp. Teor. Fiz.57,1306(1969)], pp. 709–717.
- [65] L. Stodolsky. “Hadron-like behavior of gamma, neutrino nuclear cross-sections”. In: *Phys. Rev. Lett.* 18 (1967), pp. 135–137.
- [66] D. Schildknecht. “ON SHADOWING IN PHOTOPRODUCTION AND INELASTIC ELECTRON SCATTERING FROM COMPLEX NUCLEI”. In: *Nucl. Phys.* B66 (1973), pp. 398–411.
- [67] G. Shaw. “Shadowing in the Scaling Region”. In: *Phys. Lett.* B228 (1989), pp. 125–128.
- [68] S. A. Kulagin, G. Piller, and W. Weise. “Shadowing, binding and off-shell effects in nuclear deep inelastic scattering”. In: *Phys. Rev.* C50 (1994), pp. 1154–1169. arXiv: nucl-th/9402015 [nucl-th].
- [69] J. Kwiecinski and B. M. Badelek. “Unified Description of Nuclear Shadowing of Virtual Photons”. In: *Phys. Lett.* B208 (1988), pp. 508–512.
- [70] L. V. Gribov, E. M. Levin, and M. G. Ryskin. “Semihard Processes in QCD”. In: *Phys. Rept.* 100 (1983), pp. 1–150.
- [71] A. H. Mueller and J.-w. Qiu. “Gluon Recombination and Shadowing at Small Values of x ”. In: *Nucl. Phys.* B268 (1986), pp. 427–452.

-
- [72] N. Armesto and M. A. Braun. “The EMC effect at low x in perturbative QCD”. In: *Z. Phys.* C76 (1997), pp. 81–90. arXiv: hep-ph/9603360 [hep-ph].
- [73] J. J. Aubert et al. “The ratio of the nucleon structure functions F_2^n for iron and deuterium”. In: *Phys. Lett.* 123B (1983), pp. 275–278.
- [74] J. J. Aubert et al. “Measurements of the nucleon structure functions F_2^n in deep inelastic muon scattering from deuterium and comparison with those from hydrogen and iron”. In: *Nucl. Phys.* B293 (1987), pp. 740–786.
- [75] K. Rith. “Present Status of the EMC effect”. In: *Subnucl. Ser.* 51 (2015), pp. 431–449. arXiv: 1402.5000 [hep-ex].
- [76] L. Frankfurt, M. Sargsian, and M. Strikman. “Recent observation of short range nucleon correlations in nuclei and their implications for the structure of nuclei and neutron stars”. In: *Int. J. Mod. Phys.* A23 (2008), pp. 2991–3055. arXiv: 0806.4412 [nucl-th].
- [77] J. Arrington, D. W. Higinbotham, G. Rosner, and M. Sargsian. “Hard probes of short-range nucleon-nucleon correlations”. In: *Prog. Part. Nucl. Phys.* 67 (2012), pp. 898–938. arXiv: 1104.1196 [nucl-ex].
- [78] O. Hen, D. W. Higinbotham, G. A. Miller, E. Piassetzky, and L. B. Weinstein. “The EMC Effect and High Momentum Nucleons in Nuclei”. In: *Int. J. Mod. Phys.* E22 (2013), p. 1330017. arXiv: 1304.2813 [nucl-th].
- [79] O. Hen, G. A. Miller, E. Piassetzky, and L. B. Weinstein. “Nucleon-Nucleon Correlations, Short-lived Excitations, and the Quarks Within”. In: *Rev. Mod. Phys.* 89.4 (2017), p. 045002. arXiv: 1611.09748 [nucl-ex].
- [80] N. Fomin, D. Higinbotham, M. Sargsian, and P. Solvignon. “New Results on Short-Range Correlations in Nuclei”. In: *Ann. Rev. Nucl. Part. Sci.* 67 (2017), pp. 129–159. arXiv: 1708.08581 [nucl-th].
- [81] I. C. Cloet et al. “Exposing Novel Quark and Gluon Effects in Nuclei”. In: *J. Phys.* G46.9 (2019), p. 093001. arXiv: 1902.10572 [nucl-ex].
- [82] J. Xu and F. Yuan. “Gluonic Probe for the Short Range Correlation in Nucleus”. In: *Phys. Lett.* B801 (2020), p. 135187. arXiv: 1908.10413 [hep-ph].
- [83] B. Schmookler et al. “Modified structure of protons and neutrons in correlated pairs”. In: *Nature* 566.7744 (2019), pp. 354–358.

- [84] E. P. Segarra et al. “Flavor dependence of the nucleon valence structure from nuclear deep inelastic scattering data”. In: (2019). arXiv: 1908.02223 [nucl-th].
- [85] K. S. Egiyan et al. “Measurement of 2- and 3-nucleon short range correlation probabilities in nuclei”. In: *Phys. Rev. Lett.* 96 (2006), p. 082501. arXiv: nucl-ex/0508026 [nucl-ex].
- [86] L. B. Weinstein et al. “Short Range Correlations and the EMC Effect”. In: *Phys. Rev. Lett.* 106 (2011), p. 052301. arXiv: 1009.5666 [hep-ph].
- [87] N. Fomin et al. “New measurements of high-momentum nucleons and short-range structures in nuclei”. In: *Phys. Rev. Lett.* 108 (2012), p. 092502. arXiv: 1107.3583 [nucl-ex].
- [88] J. Arrington et al. “A detailed study of the nuclear dependence of the EMC effect and short-range correlations”. In: *Phys. Rev.* C86 (2012), p. 065204. arXiv: 1206.6343 [nucl-ex].
- [89] S. V. Akulinichev, S. Shlomo, S. A. Kulagin, and G. M. Vagrado. “LEPTON NUCLEUS DEEP INELASTIC SCATTERING”. In: *Phys. Rev. Lett.* 55 (1985), pp. 2239–2241.
- [90] S. V. Akulinichev and S. Shlomo. “Nuclear Binding Effect in Deep Inelastic Lepton Scattering”. In: *Phys. Lett.* B234 (1990), pp. 170–174.
- [91] O. Benhar, V. R. Pandharipande, and I. Sick. “Nuclear binding and deep inelastic scattering”. In: *Phys. Lett.* B410 (1997), pp. 79–85.
- [92] O. Benhar, V. R. Pandharipande, and I. Sick. “Density dependence of the EMC effect”. In: *Phys. Lett.* B469 (1999), pp. 19–24.
- [93] O. Benhar and I. Sick. “Nuclear binding, correlations and the origin of EMC effect”. In: (2012). arXiv: 1207.4595 [nucl-th].
- [94] C. Ciofi Degli Atti and S. Liuti. “On the Effects of Nucleon Binding and Correlations in Deep Inelastic Electron Scattering by Nuclei”. In: *Phys. Lett.* B225 (1989), pp. 215–221.
- [95] A. E. L. Dieperink and G. A. Miller. “Nucleon binding corrections to lepton nucleus deep inelastic scattering: Use of a realistic spectral function”. In: *Phys. Rev.* C44 (1991), pp. 866–869.

-
- [96] E. Marco, E. Oset, and P. Fernandez de Cordoba. “Mesonic and binding contributions to the EMC effect in a relativistic many body approach”. In: *Nucl. Phys.* A611 (1996), pp. 484–513. arXiv: nucl-th/9510060 [nucl-th].
- [97] C. H. Llewellyn Smith. “A Possible Explanation of the Difference Between the Structure Functions of Iron and Deuterium”. In: *Phys. Lett.* 128B (1983), pp. 107–111.
- [98] M. Ericson and A. W. Thomas. “Pionic Corrections and the EMC Enhancement of the Sea in Iron”. In: *Phys. Lett.* 128B (1983), pp. 112–116.
- [99] V. Sanjose, V. Vento, and S. Noguera. “PIONIC EFFECTS IN INELASTIC ELECTRON - NUCLEUS SCATTERING”. In: *Nucl. Phys.* A470 (1987), pp. 509–522.
- [100] H. Jung and G. A. Miller. “NUCLEONIC CONTRIBUTION TO LEPTON NUCLEUS DEEP INELASTIC SCATTERING”. In: *Phys. Lett.* B200 (1988), pp. 351–356.
- [101] H. Jung and G. A. Miller. “Pionic contributions to deep inelastic nuclear structure functions”. In: *Phys. Rev.* C41 (1990), pp. 659–664.
- [102] P. Gonzalez and V. Vento. “Pionic effects in deep inelastic scattering off nuclei”. In: *Mod. Phys. Lett.* A8 (1993), pp. 1563–1572. arXiv: nucl-th/9211001 [nucl-th].
- [103] F. E. Close, R. G. Roberts, and G. G. Ross. “The Effect of Confinement Size on Nuclear Structure Functions”. In: *Phys. Lett.* 129B (1983), pp. 346–350.
- [104] O. Nachtmann and H. J. Pirner. “Color Conductivity in Nuclei and the EMC Effect”. In: *Z. Phys.* C21 (1984), p. 277.
- [105] W. L. van Neerven and A. Vogt. “NNLO evolution of deep inelastic structure functions: The Nonsinglet case”. In: *Nucl. Phys.* B568 (2000), pp. 263–286. arXiv: hep-ph/9907472 [hep-ph].
- [106] W. L. van Neerven and A. Vogt. “NNLO evolution of deep inelastic structure functions: The Singlet case”. In: *Nucl. Phys.* B588 (2000), pp. 345–373. arXiv: hep-ph/0006154 [hep-ph].

- [107] J. A. M. Vermaseren, A. Vogt, and S. Moch. “The Third-order QCD corrections to deep-inelastic scattering by photon exchange”. In: *Nucl. Phys.* B724 (2005), pp. 3–182. arXiv: hep-ph/0504242 [hep-ph].
- [108] I. Helenius. “The Structure Functions to the Order α_S in Deep Inelastic Neutrino Scattering Processes”. Master Thesis. Jyväskylä, Finland: University of Jyväskylä, 2010.
- [109] H. Paukkunen. “Global analysis of nuclear parton distribution functions at leading and next-to-leading order perturbative QCD”. PhD thesis. Jyväskylä U., 2009. arXiv: 0906.2529 [hep-ph].
- [110] E. B. Zijlstra. “Second order QCD corrections to deep inelastic processes”. PhD thesis. Leiden U., 1993.
- [111] J. C. Collins, D. E. Soper, and G. F. Sterman. *PERTURBATIVE QUANTUM CHROMODYNAMICS*. Ed. by A. H. Mueller. Vol. 5. Singapore: WSP, 1989.
- [112] D. de Florian, R. Sassot, M. Epele, R. J. Hernández-Pinto, and M. Stratmann. “Parton-to-Pion Fragmentation Reloaded”. In: *Phys. Rev.* D91.1 (2015), p. 014035. arXiv: 1410.6027 [hep-ph].
- [113] D. P. Anderle, F. Ringer, and M. Stratmann. “Fragmentation Functions at Next-to-Next-to-Leading Order Accuracy”. In: *Phys. Rev.* D92.11 (2015), p. 114017. arXiv: 1510.05845 [hep-ph].
- [114] T. Kaufmann, A. Mukherjee, and W. Vogelsang. “Hadron Fragmentation Inside Jets in Hadronic Collisions”. In: *Phys. Rev.* D92.5 (2015), p. 054015. arXiv: 1506.01415 [hep-ph].
- [115] N. Sato et al. “First Monte Carlo analysis of fragmentation functions from single-inclusive e^+e^- annihilation”. In: *Phys. Rev.* D94.11 (2016), p. 114004. arXiv: 1609.00899 [hep-ph].
- [116] D. P. Anderle, T. Kaufmann, M. Stratmann, and F. Ringer. “Fragmentation Functions Beyond Fixed Order Accuracy”. In: *Phys. Rev.* D95.5 (2017), p. 054003. arXiv: 1611.03371 [hep-ph].
- [117] D. de Florian, M. Epele, R. J. Hernandez-Pinto, R. Sassot, and M. Stratmann. “Parton-to-Kaon Fragmentation Revisited”. In: *Phys. Rev.* D95.9 (2017), p. 094019. arXiv: 1702.06353 [hep-ph].

-
- [118] V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo. “A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties”. In: *Eur. Phys. J. C* 77.8 (2017), p. 516. arXiv: 1706.07049 [hep-ph].
- [119] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk. “Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions”. In: (2019). arXiv: 1905.03788 [hep-ph].
- [120] M. Diehl. “Introduction to GPDs and TMDs”. In: *Eur. Phys. J. A* 52.6 (2016), p. 149. arXiv: 1512.01328 [hep-ph].
- [121] A. Dainese et al. “Heavy ions at the Future Circular Collider”. In: *CERN Yellow Rep.* 3 (2017), pp. 635–692. arXiv: 1605.01389 [hep-ph].
- [122] T.-J. Hou et al. “CT14 Intrinsic Charm Parton Distribution Functions from CTEQ-TEA Global Analysis”. In: *JHEP* 02 (2018), p. 059. arXiv: 1707.00657 [hep-ph].
- [123] M. Guzzi et al. “CTEQ-TEA parton distribution functions with intrinsic charm”. In: *EPJ Web Conf.* 192 (2018), p. 00003. arXiv: 1810.00264 [hep-ph].
- [124] D. J. Gross and F. Wilczek. “Asymptotically Free Gauge Theories - I”. In: *Phys. Rev.* D8 (1973), pp. 3633–3652.
- [125] H. Georgi and H. D. Politzer. “Electroproduction scaling in an asymptotically free theory of strong interactions”. In: *Phys. Rev.* D9 (1974), pp. 416–420.
- [126] E. G. Floratos, D. A. Ross, and C. T. Sachrajda. “Higher Order Effects in Asymptotically Free Gauge Theories: The Anomalous Dimensions of Wilson Operators”. In: *Nucl. Phys.* B129 (1977). [Erratum: *Nucl. Phys.*B139,545(1978)], pp. 66–88.
- [127] E. G. Floratos, D. A. Ross, and C. T. Sachrajda. “Higher Order Effects in Asymptotically Free Gauge Theories. 2. Flavor Singlet Wilson Operators and Coefficient Functions”. In: *Nucl. Phys.* B152 (1979), pp. 493–520.
- [128] A. Gonzalez-Arroyo, C. Lopez, and F. J. Yndurain. “Second Order Contributions to the Structure Functions in Deep Inelastic Scattering. 1. Theoretical Calculations”. In: *Nucl. Phys.* B153 (1979), pp. 161–186.

- [129] A. Gonzalez-Arroyo and C. Lopez. “Second Order Contributions to the Structure Functions in Deep Inelastic Scattering. 3. The Singlet Case”. In: *Nucl. Phys.* B166 (1980), pp. 429–459.
- [130] G. Curci, W. Furmanski, and R. Petronzio. “Evolution of Parton Densities Beyond Leading Order: The Nonsinglet Case”. In: *Nucl. Phys.* B175 (1980), pp. 27–92.
- [131] W. Furmanski and R. Petronzio. “Singlet Parton Densities Beyond Leading Order”. In: *Phys. Lett.* 97B (1980), pp. 437–442.
- [132] E. G. Floratos, C. Kounnas, and R. Lacaze. “Higher Order QCD Effects in Inclusive Annihilation and Deep Inelastic Scattering”. In: *Nucl. Phys.* B192 (1981), pp. 417–462.
- [133] R. Hamberg and W. L. van Neerven. “The Correct renormalization of the gluon operator in a covariant gauge”. In: *Nucl. Phys.* B379 (1992), pp. 143–171.
- [134] S. Moch, J. A. M. Vermaseren, and A. Vogt. “The Three loop splitting functions in QCD: The Nonsinglet case”. In: *Nucl. Phys.* B688 (2004), pp. 101–134. arXiv: hep-ph/0403192 [hep-ph].
- [135] A. Vogt, S. Moch, and J. A. M. Vermaseren. “The Three-loop splitting functions in QCD: The Singlet case”. In: *Nucl. Phys.* B691 (2004), pp. 129–181. arXiv: hep-ph/0404111 [hep-ph].
- [136] W. Furmanski and R. Petronzio. “Lepton - Hadron Processes Beyond Leading Order in Quantum Chromodynamics”. In: *Z. Phys.* C11 (1982), p. 293.
- [137] A. Vogt. “The Partonic structure of hadrons and photons”. PhD thesis. Dortmund U., 1992.
- [138] A. Vogt. “Efficient evolution of unpolarized and polarized parton distributions with QCD-PEGASUS”. In: *Comput. Phys. Commun.* 170 (2005), pp. 65–92. arXiv: hep-ph/0408244 [hep-ph].
- [139] M. Botje. “QCDNUM: Fast QCD Evolution and Convolution”. In: *Comput. Phys. Commun.* 182 (2011), pp. 490–532. arXiv: 1005.1481 [hep-ph].
- [140] C. de Boor. *A Practical Guide to Splines. Applied Mathematical Sciences*. New York, USA: Springer, 1978, pp. XVIII, 348. ISBN: 978-0-387-95366-3.

-
- [141] G. P. Salam and J. Rojo. “A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)”. In: *Comput. Phys. Commun.* 180 (2009), pp. 120–156. arXiv: 0804.3755 [hep-ph].
- [142] M. Miyama and S. Kumano. “Numerical solution of Q^{*2} evolution equations in a brute force method”. In: *Comput. Phys. Commun.* 94 (1996), pp. 185–215. arXiv: hep-ph/9508246 [hep-ph].
- [143] P. G. Ratcliffe. “A matrix approach to numerical solution of the DGLAP evolution equations”. In: *Phys. Rev.* D63 (2001), p. 116004. arXiv: hep-ph/0012376 [hep-ph].
- [144] C. Pascaud and F. Zomer. “A Fast and precise method to solve the Altarelli-Parisi equations in x space”. In: (2001). arXiv: hep-ph/0104013 [hep-ph].
- [145] A. Cafarella and C. Coriano. “Direct solution of renormalization group equations of QCD in x space: NLO implementations at leading twist”. In: *Comput. Phys. Commun.* 160 (2004), pp. 213–242. arXiv: hep-ph/0311313 [hep-ph].
- [146] A. Cafarella, C. Coriano, and M. Guzzi. “Precision Studies of the NNLO DGLAP Evolution at the LHC with CANDIA”. In: *Comput. Phys. Commun.* 179 (2008), pp. 665–684. arXiv: 0803.0462 [hep-ph].
- [147] V. Bertone, S. Carrazza, and J. Rojo. “APFEL: A PDF Evolution Library with QED corrections”. In: *Comput. Phys. Commun.* 185 (2014), pp. 1647–1668. arXiv: 1310.1394 [hep-ph].
- [148] M. Botje. “Erratum for the time-like evolution in QCDNUM”. In: (2016). arXiv: 1602.08383 [hep-ph].
- [149] S. Kumano and T. H. Nagai. “Comparison of numerical solutions for Q^{*2} evolution equations”. In: *J. Comput. Phys.* 201 (2004), pp. 651–664. arXiv: hep-ph/0405160 [hep-ph].
- [150] P. Santorelli and E. Scrimieri. “A Semianalytical method to evolve parton distributions”. In: *Phys. Lett.* B459 (1999), pp. 599–606. arXiv: hep-ph/9807572 [hep-ph].

- [151] G. Salam and J. Rojo. “The HOPPET NNLO parton evolution package”. In: *Proceedings, 16th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2008): London, UK, April 7-11, 2008*. 2008, p. 42. arXiv: 0807.0198 [hep-ph].
- [152] V. Bertone, S. Carrazza, and E. R. Nocera. “Reference results for time-like evolution up to $\mathcal{O}(\alpha_s^3)$ ”. In: *JHEP* 03 (2015), p. 046. arXiv: 1501.00494 [hep-ph].
- [153] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo. “Neural network determination of parton distributions: The Nonsinglet case”. In: *JHEP* 03 (2007), p. 039. arXiv: hep-ph/0701127 [hep-ph].
- [154] S. Catani, D. de Florian, G. Rodrigo, and W. Vogelsang. “Perturbative generation of a strange-quark asymmetry in the nucleon”. In: *Phys. Rev. Lett.* 93 (2004), p. 152003. arXiv: hep-ph/0404240 [hep-ph].
- [155] Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller, and S. I. Troian. *Basics of perturbative QCD*. Gif-sur-Yvette, France: Ed. Frontieres, 1991, p. 274.
- [156] J. Blumlein and A. Vogt. “On the behavior of nonsinglet structure functions at small x ”. In: *Phys. Lett.* B370 (1996), pp. 149–155. arXiv: hep-ph/9510410 [hep-ph].
- [157] E. A. Kuraev, L. N. Lipatov, and V. S. Fadin. “The Pomernanchuk Singularity in Nonabelian Gauge Theories”. In: *Sov. Phys. JETP* 45 (1977). [*Zh. Eksp. Teor. Fiz.* 72,377(1977)], pp. 199–204.
- [158] E. Iancu, A. Leonidov, and L. D. McLerran. “The Renormalization group equation for the color glass condensate”. In: *Phys. Lett.* B510 (2001), pp. 133–144. arXiv: hep-ph/0102009 [hep-ph].
- [159] E. Iancu, A. Leonidov, and L. D. McLerran. “Nonlinear gluon evolution in the color glass condensate. 1.” In: *Nucl. Phys.* A692 (2001), pp. 583–645. arXiv: hep-ph/0011241 [hep-ph].
- [160] E. Ferreiro, E. Iancu, A. Leonidov, and L. McLerran. “Nonlinear gluon evolution in the color glass condensate. 2.” In: *Nucl. Phys.* A703 (2002), pp. 489–538. arXiv: hep-ph/0109115 [hep-ph].

-
- [161] J. Jalilian-Marian, A. Kovner, A. Leonidov, and H. Weigert. “The BFKL equation from the Wilson renormalization group”. In: *Nucl. Phys.* B504 (1997), pp. 415–431. arXiv: hep-ph/9701284 [hep-ph].
- [162] J. Jalilian-Marian, A. Kovner, and H. Weigert. “The Wilson renormalization group for low x physics: Gluon evolution at finite parton density”. In: *Phys. Rev.* D59 (1998), p. 014015. arXiv: hep-ph/9709432 [hep-ph].
- [163] J. Jalilian-Marian, A. Kovner, A. Leonidov, and H. Weigert. “The Wilson renormalization group for low x physics: Towards the high density regime”. In: *Phys. Rev.* D59 (1998), p. 014014. arXiv: hep-ph/9706377 [hep-ph].
- [164] H. Weigert. “Unitarity at small Bjorken x”. In: *Nucl. Phys.* A703 (2002), pp. 823–860. arXiv: hep-ph/0004044 [hep-ph].
- [165] I. Balitsky. “Operator expansion for high-energy scattering”. In: *Nucl. Phys.* B463 (1996), pp. 99–160. arXiv: hep-ph/9509348 [hep-ph].
- [166] Y. V. Kovchegov. “Small x $F(2)$ structure function of a nucleus including multiple pomeron exchanges”. In: *Phys. Rev.* D60 (1999), p. 034008. arXiv: hep-ph/9901281 [hep-ph].
- [167] *Download xFitter source code extensions for nPDFs:*
https://gitlab.com/fitters/xfitter/tree/NuclearPDFs_2.0.1 (2019-12-04 11:10).
- [168] *Download LHAPDF6 sets:*
<https://lhapdf.hepforge.org/pdfsets.html> (2020-01-13 10:55).
- [169] H. Abramowicz et al. “Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data”. In: *Eur. Phys. J.* C75.12 (2015), p. 580. arXiv: 1506.06042 [hep-ex].
- [170] K. Kovarik et al. “nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework”. In: *Phys. Rev.* D93.8 (2016), p. 085037. arXiv: 1509.00792 [hep-ph].
- [171] C. Boros, J. T. Londergan, and A. W. Thomas. “Evidence for charge symmetry violation in parton distributions”. In: *Phys. Rev.* D59 (1999), p. 074021. arXiv: hep-ph/9810220 [hep-ph].
- [172] K. Saito et al. “Deep inelastic scattering on asymmetric nuclei”. In: *Phys. Lett.* B493 (2000), pp. 288–292. arXiv: nucl-th/0008017 [nucl-th].

- [173] J. T. Londergan, D. P. Murdock, and A. W. Thomas. “Experimental tests of charge symmetry violation in parton distributions”. In: *Phys. Rev. D* 72 (2005), p. 036010. arXiv: hep-ph/0507029 [hep-ph].
- [174] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne. “Uncertainties of predictions from parton distributions. 2. Theoretical errors”. In: *Eur. Phys. J. C* 35 (2004), pp. 325–348. arXiv: hep-ph/0308087 [hep-ph].
- [175] X. G. Wang, A. W. Thomas, and R. D. Young. “Electromagnetic contribution to charge symmetry violation in parton distributions”. In: *Phys. Lett. B* 753 (2016), pp. 595–599. arXiv: 1512.04139 [nucl-th].
- [176] A. Accardi et al. “A Critical Appraisal and Evaluation of Modern PDFs”. In: *Eur. Phys. J. C* 76.8 (2016), p. 471. arXiv: 1603.08906 [hep-ph].
- [177] S. Alekhin. *OPENQCDRAD, a program description and the code are available via: <http://www.zeuthen.desy.de/~alekhin/OPENQCDRAD>*.
- [178] S. Alekhin and S. Moch. “Heavy-quark deep-inelastic scattering with a running mass”. In: *Phys. Lett. B* 699 (2011), pp. 345–353. arXiv: 1011.5790 [hep-ph].
- [179] M. A. G. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung. “Leptoproduction of heavy quarks. 2. A Unified QCD formulation of charged and neutral current processes from fixed target to collider energies”. In: *Phys. Rev. D* 50 (1994), pp. 3102–3118. arXiv: hep-ph/9312319 [hep-ph].
- [180] M. Krämer, F. I. Olness, and D. E. Soper. “Treatment of heavy quarks in deeply inelastic scattering”. In: *Phys. Rev. D* 62 (2000), p. 096007. arXiv: hep-ph/0003035 [hep-ph].
- [181] W.-K. Tung, S. Kretzer, and C. Schmidt. “Open heavy flavor production in QCD: Conceptual framework and implementation issues”. In: *J. Phys. G* 28 (2002), pp. 983–996. arXiv: hep-ph/0110247 [hep-ph].
- [182] S. Kretzer, H. L. Lai, F. I. Olness, and W. K. Tung. “Cteq6 parton distributions with heavy quark mass effects”. In: *Phys. Rev. D* 69 (2004), p. 114005. arXiv: hep-ph/0307022 [hep-ph].
- [183] R. S. Thorne and R. G. Roberts. “An Ordered analysis of heavy flavor production in deep inelastic scattering”. In: *Phys. Rev. D* 57 (1998), pp. 6871–6898. arXiv: hep-ph/9709442 [hep-ph].

-
- [184] R. S. Thorne. “A Variable-flavor number scheme for NNLO”. In: *Phys. Rev.* D73 (2006), p. 054019. arXiv: hep-ph/0601245 [hep-ph].
- [185] R. S. Thorne. “Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections”. In: *Phys. Rev.* D86 (2012), p. 074017. arXiv: 1201.6180 [hep-ph].
- [186] M. Cacciari, M. Greco, and P. Nason. “The P(T) spectrum in heavy flavor hadroproduction”. In: *JHEP* 05 (1998), p. 007. arXiv: hep-ph/9803400 [hep-ph].
- [187] S. Forte, E. Laenen, P. Nason, and J. Rojo. “Heavy quarks in deep-inelastic scattering”. In: *Nucl. Phys.* B834 (2010), pp. 116–162. arXiv: 1001.2312 [hep-ph].
- [188] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado. “EPPS16: Nuclear parton distributions with LHC data”. In: *Eur. Phys. J.* C77.3 (2017), p. 163. arXiv: 1612.05741 [hep-ph].
- [189] *xFitter*: <https://www.xfitter.org/xFitter>.
- [190] O. Zenaiev. “xFitter project”. In: *PoS DIS2016* (2016), p. 033.
- [191] V. Bertone et al. “xFitter 2.0.0: An Open Source QCD Fit Framework”. In: *PoS DIS2017* (2018), p. 203. arXiv: 1709.01151 [hep-ph].
- [192] F. James and M. Roos. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”. In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367.
- [193] A. Lazzaro and L. Moneta. “MINUIT package parallelization and applications using the RooFit package”. In: *J. Phys. Conf. Ser.* 219 (2010), p. 042044.
- [194] T. Carli et al. “A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project”. In: *Eur. Phys. J.* C66 (2010), pp. 503–524. arXiv: 0911.2985 [hep-ph].
- [195] A. Luszczak and H. Kowalski. “Dipole model analysis of high precision HERA data”. In: *Phys. Rev.* D89.7 (2014), p. 074051. arXiv: 1312.4060 [hep-ph].
- [196] A. Luszczak and H. Kowalski. “Dipole model analysis of highest precision HERA data, including very low Q^2 ’s”. In: *Phys. Rev.* D95.1 (2017), p. 014030. arXiv: 1611.10100 [hep-ph].

- [197] R. D. Ball et al. “Parton distributions with small- x resummation: evidence for BFKL dynamics in HERA data”. In: *Eur. Phys. J. C* 78.4 (2018), p. 321. arXiv: 1710.05935 [hep-ph].
- [198] H. Abdolmaleki et al. “Impact of low- x resummation on QCD analysis of HERA data”. In: *Eur. Phys. J. C* 78.8 (2018), p. 621. arXiv: 1802.00064 [hep-ph].
- [199] C. Adloff et al. “Measurement and QCD analysis of neutral and charged current cross-sections at HERA”. In: *Eur. Phys. J. C* 30 (2003), pp. 1–32. arXiv: hep-ex/0304003 [hep-ex].
- [200] J. Pumplin et al. “Uncertainties of predictions from parton distribution functions. 2. The Hessian method”. In: *Phys. Rev. D* 65 (2001), p. 014013. arXiv: hep-ph/0101032 [hep-ph].
- [201] J. Pumplin, D. R. Stump, and W. K. Tung. “Multivariate fitting and the error matrix in global analysis of data”. In: *Phys. Rev. D* 65 (2001), p. 014011. arXiv: hep-ph/0008191 [hep-ph].
- [202] K. J. Eskola, P. Paakkinen, and H. Paukkunen. “Non-quadratic improved Hessian PDF reweighting and application to CMS dijet measurements at 5.02 TeV”. In: *Eur. Phys. J. C* 79.6 (2019), p. 511. arXiv: 1903.09832 [hep-ph].
- [203] 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 ”. In: *Phys. Lett. B* 223 (1989), pp. 485–489.
- [204] 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 ”. In: *Nucl. Phys. B* 483 (1997), pp. 3–43. arXiv: hep-ph/9610231 [hep-ph].
- [205] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann. “Global Analysis of Nuclear Parton Distributions”. In: *Phys. Rev. D* 85 (2012), p. 074028. arXiv: 1112.6324 [hep-ph].
- [206] M. Arneodo et al. “Measurements of the nucleon structure function in the range $0.002 - \text{GeV}^2 < x < 0.17 - \text{GeV}^2$ and $0.2 - \text{GeV}^2 < q^2 < 8 - \text{GeV}^2$ in deuterium, carbon and calcium”. In: *Nucl. Phys. B* 333 (1990), pp. 1–47.
- [207] A. Airapetian et al. “Measurement of $R = \sigma(L) / \sigma(T)$ in deep inelastic scattering on nuclei”. In: (2002). arXiv: hep-ex/0210068 [hep-ex].

-
- [208] M. Arneodo et al. “The Structure Function ratios $F_2(\text{Li}) / F_2(\text{D})$ and $F_2(\text{C}) / F_2(\text{D})$ at small x ”. In: *Nucl. Phys.* B441 (1995), pp. 12–30. arXiv: hep-ex/9504002 [hep-ex].
- [209] M. Arneodo et al. “The A dependence of the nuclear structure function ratios”. In: *Nucl. Phys.* B481 (1996), pp. 3–22.
- [210] J. Ashman et al. “Measurement of the Ratios of Deep Inelastic Muon - Nucleus Cross-Sections on Various Nuclei Compared to Deuterium”. In: *Phys. Lett.* B202 (1988), pp. 603–610.
- [211] S. Dasu et al. “Measurement of kinematic and nuclear dependence of $R = \sigma_L / \sigma_T$ in deep inelastic electron scattering”. In: *Phys. Rev.* D49 (1994), pp. 5641–5670.
- [212] J. P. Berge et al. “A Measurement of Differential Cross-Sections and Nucleon Structure Functions in Charged Current Neutrino Interactions on Iron”. In: *Z. Phys.* C49 (1991), pp. 187–224.
- [213] M. Arneodo et al. “The Q^2 dependence of the structure function ratio $F_2 \text{ Sn} / F_2 \text{ C}$ and the difference $R \text{ Sn} - R \text{ C}$ in deep inelastic muon scattering”. In: *Nucl. Phys.* B481 (1996), pp. 23–39.
- [214] M. R. Adams et al. “Saturation of shadowing at very low x_{BJ} ”. In: *Phys. Rev. Lett.* 68 (1992), pp. 3266–3269.
- [215] G. Onengut et al. “Measurement of nucleon structure functions in neutrino scattering”. In: *Phys. Lett.* B632 (2006), pp. 65–75.
- [216] A. Bodek et al. “A Comparison of the Deep Inelastic Structure Functions of Deuterium and Aluminum Nuclei”. In: *Phys. Rev. Lett.* 51 (1983), p. 534.
- [217] P. Amaudruz et al. “The ratio $F_2(\text{n}) / F_2(\text{p})$ in deep inelastic muon scattering”. In: *Nucl. Phys.* B371 (1992), pp. 3–31.
- [218] D. Stump et al. “Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method”. In: *Phys. Rev.* D65 (2001), p. 014012. arXiv: hep-ph/0101051 [hep-ph].
- [219] M. Botje. “Error estimates on parton density distributions”. In: *J. Phys.* G28 (2002), pp. 779–790. arXiv: hep-ph/0110123 [hep-ph].

- [220] F. D. Aaron et al. “Measurement of the Inclusive ep Scattering Cross Section at Low Q^2 and x at HERA”. In: *Eur. Phys. J. C* 63 (2009), pp. 625–678. arXiv: 0904.0929 [hep-ex].
- [221] K. Kovarik et al. “Nuclear Corrections in Neutrino-Nucleus DIS and Their Compatibility with Global NPDF Analyses”. In: *Phys. Rev. Lett.* 106 (2011), p. 122301. arXiv: 1012.0286 [hep-ph].
- [222] H. Paukkunen and C. A. Salgado. “Agreement of Neutrino Deep Inelastic Scattering Data with Global Fits of Parton Distributions”. In: *Phys. Rev. Lett.* 110.21 (2013), p. 212301. arXiv: 1302.2001 [hep-ph].
- [223] M. Tzanov et al. “Precise measurement of neutrino and anti-neutrino differential cross sections”. In: *Phys. Rev. D* 74 (2006), p. 012008. arXiv: hep-ex/0509010 [hep-ex].
- [224] D. MacFarlane et al. “Nucleon Structure Functions from High-Energy Neutrino Interactions with Iron and QCD Results”. In: *Z. Phys. C* 26 (1984), pp. 1–12.
- [225] J. F. Owens et al. “The Impact of new neutrino DIS and Drell-Yan data on large-x parton distributions”. In: *Phys. Rev. D* 75 (2007), p. 054030. arXiv: hep-ph/0702159 [HEP-PH].
- [226] I. Schienbein et al. “Nuclear parton distribution functions from neutrino deep inelastic scattering”. In: *Phys. Rev. D* 77 (2008), p. 054013. arXiv: 0710.4897 [hep-ph].
- [227] I. Schienbein et al. “PDF Nuclear Corrections for Charged and Neutral Current Processes”. In: *Phys. Rev. D* 80 (2009), p. 094004. arXiv: 0907.2357 [hep-ph].
- [228] H. Paukkunen and C. A. Salgado. “Compatibility of neutrino DIS data and global analyses of parton distribution functions”. In: *JHEP* 07 (2010), p. 032. arXiv: 1004.3140 [hep-ph].
- [229] W. T. Giele and S. Keller. “Implications of hadron collider observables on parton distribution function uncertainties”. In: *Phys. Rev. D* 58 (1998), p. 094023. arXiv: hep-ph/9803393 [hep-ph].
- [230] W. T. Giele, S. A. Keller, and D. A. Kosower. “Parton Distribution Function Uncertainties”. In: (2001). arXiv: hep-ph/0104052 [hep-ph].

-
- [231] N. P. Hartland et al. “A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector”. In: *JHEP* 04 (2019), p. 100. arXiv: 1901.05965 [hep-ph].
- [232] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne. “Uncertainties of predictions from parton distributions. 1: Experimental errors”. In: *Eur. Phys. J. C* 28 (2003), pp. 455–473. arXiv: hep-ph/0211080 [hep-ph].
- [233] P. M. Nadolsky and Z. Sullivan. “PDF Uncertainties in WH Production at Tevatron”. In: *eConf* C010630 (2001), P510. arXiv: hep-ph/0110378 [hep-ph].
- [234] M. Hirai, S. Kumano, and T. H. Nagai. “Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order”. In: *Phys. Rev. C* 76 (2007), p. 065207. arXiv: 0709.3038 [hep-ph].
- [235] R. Abdul Khalek, J. J. Ethier, and J. Rojo. “Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider”. In: *Eur. Phys. J. C* 79.6 (2019), p. 471. arXiv: 1904.00018 [hep-ph].
- [236] H. Khanpour and S. Atashbar Tehrani. “Global Analysis of Nuclear Parton Distribution Functions and Their Uncertainties at Next-to-Next-to-Leading Order”. In: *Phys. Rev. D* 93.1 (2016), p. 014026. arXiv: 1601.00939 [hep-ph].
- [237] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (2009), pp. 189–285. arXiv: 0901.0002 [hep-ph].
- [238] P. Jimenez-Delgado and E. Reya. “Dynamical NNLO parton distributions”. In: *Phys. Rev. D* 79 (2009), p. 074023. arXiv: 0810.4274 [hep-ph].
- [239] S. Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Phys. Rev. D* 93.3 (2016), p. 033006. arXiv: 1506.07443 [hep-ph].
- [240] R. D. Ball et al. “Parton distributions from high-precision collider data”. In: *Eur. Phys. J. C* 77.10 (2017), p. 663. arXiv: 1706.00428 [hep-ph].
- [241] D. M. Alde et al. “Nuclear dependence of dimuon production at 800-GeV. FNAL-772 experiment”. In: *Phys. Rev. Lett.* 64 (1990), pp. 2479–2482.

- [242] M. A. Vasilev et al. “Parton energy loss limits and shadowing in Drell-Yan dimuon production”. In: *Phys. Rev. Lett.* 83 (1999), pp. 2304–2307. arXiv: [hep-ex/9906010](#) [[hep-ex](#)].
- [243] P. Bordalo et al. “Nuclear Effects on the Nucleon Structure Functions in Hadronic High Mass Dimuon Production”. In: *Phys. Lett.* B193 (1987), p. 368.
- [244] J. Badier et al. “Test of Nuclear Effects in Hadronic Dimuon Production”. In: *Phys. Lett.* 104B (1981). [[807\(1981\)](#)], p. 335.
- [245] G. Aad et al. “Z boson production in p +Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV measured with the ATLAS detector”. In: *Phys. Rev.* C92.4 (2015), p. 044915. arXiv: [1507.06232](#) [[hep-ex](#)].
- [246] V. Khachatryan et al. “Study of Z boson production in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett.* B759 (2016), pp. 36–57. arXiv: [1512.06461](#) [[hep-ex](#)].
- [247] V. Khachatryan et al. “Study of W boson production in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett.* B750 (2015), pp. 565–586. arXiv: [1503.05825](#) [[nucl-ex](#)].
- [248] A. M. Sirunyan et al. “Observation of nuclear modifications in W^\pm boson production in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: (2019). arXiv: [1905.01486](#) [[hep-ex](#)].
- [249] A. Falkowski, M. L. Mangano, A. Martin, G. Perez, and J. Winter. “Data driving the top quark forward–backward asymmetry with a lepton-based handle”. In: *Phys. Rev.* D87.3 (2013), p. 034039. arXiv: [1212.4003](#) [[hep-ph](#)].
- [250] J. M. Campbell, R. K. Ellis, and W. T. Giele. “A Multi-Threaded Version of MCFM”. In: *Eur. Phys. J.* C75.6 (2015), p. 246. arXiv: [1503.06182](#).
- [251] R. Boughezal et al. “Color singlet production at NNLO in MCFM”. In: *Eur. Phys. J.* C77.1 (2017), p. 7. arXiv: [1605.08011](#) [[hep-ph](#)].
- [252] S. D. Drell, D. J. Levy, and T.-M. Yan. “A Field Theoretic Model for electron-Nucleon Deep Inelastic Scattering”. In: *Phys. Rev. Lett.* 22 (1969), pp. 744–748.
- [253] S. D. Drell, D. J. Levy, and T.-M. Yan. “A Theory of Deep Inelastic Lepton-Nucleon Scattering and Lepton Pair Annihilation Processes. 1.” In: *Phys. Rev.* 187 (1969), pp. 2159–2171.

-
- [254] S. D. Drell, D. J. Levy, and T.-M. Yan. “A Theory of Deep Inelastic Lepton Nucleon Scattering and Lepton Pair Annihilation Processes. 2. Deep Inelastic electron Scattering”. In: *Phys. Rev. D* 1 (1970), pp. 1035–1068.
- [255] S. D. Drell and T.-M. Yan. “Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies”. In: *Phys. Rev. Lett.* 25 (1970). [Erratum: *Phys. Rev. Lett.* 25,902(1970)], pp. 316–320.
- [256] S. D. Drell and T.-M. Yan. “Partons and their Applications at High-Energies”. In: *Annals Phys.* 66 (1971). [*Annals Phys.* 281,450(2000)], p. 578.
- [257] G. T. Bodwin. “Factorization of the Drell-Yan Cross-Section in Perturbation Theory”. In: *Phys. Rev. D* 31 (1985). [Erratum: *Phys. Rev. D* 34,3932(1986)], p. 2616.
- [258] G. Aad et al. “Determination of the strange quark density of the proton from ATLAS measurements of the $W \rightarrow \ell\nu$ and $Z \rightarrow \ell\ell$ cross sections”. In: *Phys. Rev. Lett.* 109 (2012), p. 012001. arXiv: 1203.4051 [hep-ex].
- [259] G. Aad et al. “Measurement of the high-mass Drell-Yan differential cross-section in pp collisions at $\sqrt{s}=7$ TeV with the ATLAS detector”. In: *Phys. Lett. B* 725 (2013), pp. 223–242. arXiv: 1305.4192 [hep-ex].
- [260] G. Aad et al. “Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector”. In: *JHEP* 06 (2014), p. 112. arXiv: 1404.1212 [hep-ex].
- [261] V. Khachatryan et al. “Measurement of the differential cross section and charge asymmetry for inclusive $pp \rightarrow W^\pm + X$ production at $\sqrt{s} = 8$ TeV”. In: *Eur. Phys. J. C* 76.8 (2016), p. 469. arXiv: 1603.01803 [hep-ex].
- [262] M. Aaboud et al. “Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector”. In: *Eur. Phys. J. C* 77.6 (2017), p. 367. arXiv: 1612.03016 [hep-ex].
- [263] *xFitter User Manual*: <https://www.xfitter.org/xFitter/xFitter/DownloadPage?action=AttachFile&do=view&target=manual.pdf>.

Acknowledgements / Danksagung

I thank my supervisor Prof. Dr. Werner Vogelsang for the possibility to write this thesis and introducing me into the field of high energy particle physics. I am particularly thankful for the mixture of the supervising discussions and the opportunities to work independently. Furthermore, many thanks for supporting the travelling activities for the advanced physics schools, workshops and conferences and so allowing a smooth integration into the scientific community.

I thank Prof. Dr. Thomas Gutsche for providing the second appraisal of this thesis, and supporting the application procedure by writing letters of recommendation.

I thank Dr. Ilkka Helenius for the collaboration, i.e. technical discussions, help to set up the numerical tools and batch jobs and of course help in writing our publications, as well as for the review of the relevant parts of this manuscript, and the support in writing grant proposals. Many thanks for all advices and conversations.

I thank Dr. Marco Stratmann for sharing his knowledge, having technical discussions and for the contributed information, especially all the useful references.

I thank all members of the research group of Prof. Werner Vogelsang and the research group of Prof. Barbara Jäger for the interesting seminars, the nice atmosphere and the refreshing conversations during the lunch time. I thank Dr. Christoph Borschen-sky for the patient discussions on QCD and reviewing the according chapter of this manuscript. I thank all my office mates, I had the pleasure to share my office with in the past years, for a welcoming and friendly working atmosphere. Especially, I thank Dr. Felix Hekhorn for the nice and helpful conversations, sharing the experiences and reviewing this manuscript.

The author thanks xFITTER development team for providing operative support. The author acknowledges cooperative discussions and scientific exchange with other nPDF groups. Furthermore, the author acknowledges discussions and technical input by Dr. Hannu Paukkunen. In particular, many thanks for the provided insights during the lectures at the 28th Summer School at the University of Jyväskylä.

The author acknowledges support by the state of Baden-Württemberg through bwHPC providing the possibility to run the computational calculations on the high-performance cluster. This work was supported in part by the Bundesministerium für Bildung und Forschung (BMBF) grant 05P18VTCA1.

Ich danke Prof. Dr. Holger Cartarius für seine Hilfsbereitschaft und sein Engagement in der Arbeit mit den Studenten seit dem ersten Semester an der Universität Stuttgart. Ich bedanke mich dafür, dass er es verstanden hat, jeden, der motiviert war, in die wissenschaftliche Community aufzunehmen und das Interesse für Physik zu teilen. Ich bedanke mich auch für die Unterstützung beim Anfertigen der Empfehlungsschreiben während der Bewerbungsphase.

Ich bedanke mich bei meinen Kommilitonen/-innen und Studienfreunden/-innen von der Universität Stuttgart, dafür, dass wir so eine produktive und freudige Studienzeit hatten. Ich bin froh über die Unterstützung und den Austausch während der Studien- und Promotionszeit. Ich danke auch meinen Kommilitonen/-innen und Studienfreunden/-innen, die ich an der Universität Tübingen getroffen habe, für ihre tatkräftige Unterstützung und das Teilen der wissenschaftlichen Ambitionen. An dieser Stelle möchte ich mich besonders bei Gela Hämmerling bedanken, mit wem ich das Glück hatte sowohl in Stuttgart als auch in Tübingen gemeinsam zu studieren. Ihre Gesellschaft war mir stets eine Stütze und eine Freude.

Ich danke meinen lieben Freunden/-innen außerhalb der Physik, für ihre unermüdliche Ermunterung und Unterstützung in allen Lebenslagen, so wie auch während der Promotion. Ich bedanke mich bei Lena Gutjahr zusätzlich für ihre wertvolle Hilfe als Gesprächspartnerin beim Konzipieren der Fachvorträge und ihre Unterstützung bei der graphischen Vortragsgestaltung.

Mein besonderer Dank gilt meiner Familie für ihre Unterstützung während der Promotion, aber auch für all die Zeit und die Anstrengungen, die dahin geführt haben. Mein Dank gilt meiner Familie hier in Deutschland, aber auch den Familienmitgliedern in Kasachstan. Ich danke von ganzem Herzen meinem Mann Eugen Walt dafür, dass er mich so geduldig und liebevoll auf dem Lebensweg begleitet, und mich auch bei den verrücktesten Vorhaben unterstützt.

Danke!

