UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata

Dipartimento di Fisica

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata

Ciclo XXVI

Unbiased spin-dependent Parton Distribution Functions

Settore Scientifico Disciplinare FIS/02

Supervisore: Professor Stefano FORTE

Co-Supervisore: Dottor Juan ROJO

Coordinatore: Professor Marco BERSANELLI

Tesi di Dottorato di:

Emanuele R. NOCERA

Anno Accademico 2012-2013

Commission of the final examination:

External Referee:
Prof. Richard D. BALL

External Members:
Prof. Mauro ANSELMINO
Prof. Giovanni RIDOLFI

Final examination:

February, 28 2014
Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

MIUR subjects:

FIS/02 - Fisica Teorica, Modelli e Metodi Matematici

PACS:

02.70.Uu, 07.05.Mh, 12.38.-t, 13.60.Hb, 13.88.+e

Keywords:

Spin, Parton Distribution Functions (PDF), Neural Networks, NNPDF, High-energy Physics

Cover illustration:

Emanuele R. Nocera

Internal illustrations:

Emanuele R. Nocera, made with ROOT v5.34/09 and `feynMF` v1.5
Fig. 2.2: courtesy of Mauro Anselmino
Fig. 3.1: taken from Ref. [1]
Fig. 3.2: courtesy of Stefano Carrazza

Template design:

Anna Lisa Varri, Ph.D.
Typeset by Emanuele R. Nocera using TEXLive 2013 on Ubuntu 12.04 LTS

To my family

Abstract

We present the first unbiased determination of spin-dependent, or polarized, parton distribution functions of the proton. A statistically sound representation of the corresponding uncertainties is achieved by means of the NNPDF methodology, formerly developed for unpolarized distributions and generalized to the polarized here for the first time. The features of the procedure, based on robust statistical tools - Monte Carlo sampling for error propagation, neural networks for PDF parametrization, genetic algorithm for their minimization and possibly reweighting for including new data samples without refitting - and their implementation are illustrated in detail. Sets of polarized parton distributions are obtained at next-to-leding order accuracy in perturbative quantum chromodynamics, based on both fixed-target inclusive deeply-inelastic scattering and the most recent polarized hadron collider data. A quantitative appraisal on the potential role of future measurements at an Electron-Ion Collider is also presented. We study the stability of our results upon the variation of several theoretical and methodological assumptions and we present a detailed investigation of the first moments of our polarized parton distributions, compared to other recent analyses. We find that the uncertainty on the gluon distribution from available data was substantially underestimated in previous determinations; in particular, we emphasize that a large contribution to the gluon may arise from the unmeasured small-x region, against the common belief that this is actually rather small. We demonstrate that an Electron-Ion Collider would provide evidence of a possible large gluon contribution to the nucleon spin, though with a sizable residual uncertainty.

List of Publications

Refereed publications

- R. D. Ball et al., *Polarized Parton Distributions at an Electron-Ion Collider*, Phys. Lett. **B728** (2014) 524
  [arXiv:1310.0461] DOI: 10.1016/j.physletb.2013.12.023

- R. D. Ball et al., *Unbiased determination of polarized parton distributions and their uncertainites*, Nucl. Phys. **B874** (2013) 36
  [arXiv:1303.7236] DOI: 10.1016/j.nuclphysb.2013.05.007

- M. Anselmino, M. Boglione, U. D'Alesio, S. Melis, F. Murgia, E. R. Nocera and A. Prokudin, *General Helicity Formalism for Polarized Semi-Inclusive Deep Inelastic Scattering*, Phys. Rev. **D83** (2011) 114019
  [arXiv:1101.1011] DOI: 10.1103/PhysRevD.83.114019

Publications in preparation

- R. D. Ball et al., *A first unbiased global extraction of polarized parton distributions*

Publications in conference proceedings

- E. R. Nocera, *Constraints on polarized parton distributions from open charm and W production data*, **PoS DIS2013** (2013) 211
  [arXiv:1307.0146]

- E. R. Nocera, *Inclusion of $W^{\pm}$ single-spin asymmetry data in a polarized PDF determination via Bayesian reweighting*, Nuovo Cim. **C36** (2013) 143-147
  [arXiv:1302.6409] DOI: 10.1393/ncc/i2013-11592-4

- N. P. Hartland and E. R. Nocera, *A Mathematica interface to NNPDFs*, Nucl. Phys. Proc. Suppl. **224** (2013) 54-57
  [arXiv:1209.2585] DOI: 10.1016/j.nuclphysbps.2012.11.013

- E. R. Nocera, S. Forte, G. Ridolfi and J. Rojo, *Unbiased Polarized Parton Distributions and their Uncertainties*, Proceedings of 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012), p.937-942
  [arXiv:1206.0201] DOI:10.3204/DESY-PROC-2012-02/273

Acknowledgements

I would like to thank all those people who have supported me in the course of my work as a graduate student towards my Ph.D., particularly during the completion of this Thesis. I apologize in advance for missing some of them.

First of all, I am grateful to my supervisor, prof. Stefano Forte, who set an example to me of a scientist and a teacher. In particular, I acknowledge his wide expertise in physics, his patient willingness, his inexhaustible enthusiasm and his illuminating ideas, from which I had the opportunity to benefit (and learn) almost every day in the last three years.

Second, I am in debt to dr. Juan Rojo, my co-supervisor, for having introduced me to the NNPDF methodology, for continuous assistance with issues about code writing and for encouragement in pursuing my research with dedication. I owe my gratitude to him also for the opportunity he will give me to spend a couple of months in Oxford soon. To both Stefano and Juan I give my thanks for their training not only in undertaking scientific research, but also in presenting results: in particular, I very much appreciated (and benefitted from) their advices on talks and proceedings prepared for conferences in which I took part as a speaker.

Besides, I would like to thank prof. Richard D. Ball, who accepted to be the referee of this Thesis, and had carefully reviewed the manuscript: it has much improved thanks to his corrections and suggetions. I thank as well the other two external members of the final examination committee, prof. Mauro Anselmino and prof. Giovanni Ridolfi. On the one hand, I acknowledge Mauro and his group in Torino for having introduced me to spin physics some years ago; also, I am grateful to him for financial support to attend the CLXXX course of the Enrico Fermi summer school in Varenna. On the other hand, I am in debt to Giovanni for the forthcoming opportunity to spend a year in Genoa.

Additional thanks are due to: the Ph.D. school board, for extra financial support provided for housing costs during my stay in Milan; the *Laboratorio di Calcolo & Multimedia* (LCM) staff, for the computational resources provided to undertake the numerical analyses presented in this Thesis; dr. Francesco Caravaglios, who has allowed for sharing with me not only his office, but also his personal ideas on physics, mathematics, metaphysics, biology, chemistry: though looking weird at first, they finally reveal to me his genius.

I thank all those people I met during the last three years who also taught me a lot with their experience and knowledge. I would like to mention the members of the Department of Physics at the University of Milan, particularly those of the Theoretical Division who participated in the weekly journal club: Daniele Bettinelli, Giuseppe Bozzi, Giancarlo Ferrera, Alessandro Vicini. I also thank Mario Raciti for giving me the opportunity to work as a teaching assistant to his course in General Physics for bachelor students in *Comunicazione Digitale*.

I aknowledge several physicists I met many times at conferences and workshops, who raised me

in the spin physics community: in particular Alessandro Bacchetta, Elena Boglione, Maarten Buffing, Aurore Courtoy, Isabella Garzia, Francesca Giordano, Delia Hasch, Cédric Lorcé, Stefano Melis, Alexei Prokudin, Marco Radici, Ignazio Scimemi.

I very much appreciated the friendly and cheerful environment in the Milan Ph.D. school, thanks to all my colleagues fellow students. I just mention Alberto, Alice, Elena, Elisa, Rosa, Sofia. Special thanks are deserved by Stefano Carrazza, who has rapidly become for me an example of efficiency and hard work and a friend. I really enjoyed our discussions not only about physics and NNPDF code, but also about motorbikes and trekking. I will never forget our journey from Milan to Marseille to attend DIS2013 nor our excursion to Montenvers, *Grand Balcon Nord* and *Plan de l'Aiguille*.

During my stay in Milan I was housed in *Centro Giovanile Pavoniano*: I would like to thank the directorate, in particular Fr. Giorgio, for his kind hospitality, as well as all other guests who contributed to a friendly and enjoyable atmosphere.

Finally, I thank my family, in particular my parents, for their support, even though they stopped understanding what I am doing long ago: they have given me the chance to look at the world with curiosity and love.

Emanuele R. Nocera

Contents

1 Introduction

The investigation of the internal structure of nucleons is an old and intriguing problem which dates back to almost fifty years ago. For the past few decades, physicists have been able to describe with increasing details the fundamental particles that constitute protons and neutrons, which actually make up all nuclei and hence most of the visible matter in the Universe. This understanding is encapsulated in the Standard Model, supplemented with pertubartive Quantum Chromodynamics (QCD), the field theory which currently describes the strong interaction between the nucleon's fundamental constituents, quarks and gluons. It is a remarkable property of QCD, known as *confinement*, that these are not seen in isolation, but only bound to singlet states of the their respective strong *color* charge.

Protons and neutrons are spin one-half bound states. Spin is one of the most fundamental concepts in physics, deeply rooted in Poincaré invariance and hence in the structure of space-time itself. The elementary constituents of the nucleon carry spin, quarks are spin one-half particles and gluons are spin-one particles. It is worth recalling that the discovery of the fact that the proton has structure - and hence really the birth of strong interaction physics - was due to spin, through the measurement of a very unexpected *anomalous* magnetic moment of the proton by O. Stern and collaborators in 1933 [2]. After decades of ever more detailed studies of nucleon structure, the understanding of the observed spin of the nucleon in terms of their constituents is a major challenge, far from being succesfully achieved.

1.1 Historical overview on the nucleon structure

The current picture of the nucleon structure is the result of more than half a century of theoretical and experimental efforts from physicists around the world. Even though a detailed historical overview is beyond the scope of this introduction, we find it useful to summarize the main steps in the building of our knowledge on the proton structure, with some emphasis on its spin.

Quarks were originally introduced in 1963 by Gell-Mann, Ne'eman and Zweig, simply based on symmetry considerations [3–6], in an attempt to bring order into the large array of strongly-interacting particles observed in experiment. In a few words, they recognised that the known hadrons could be associated to some representations of the special unitary $SU(3)$ group. This led to the concept of quarks as the building block of hadrons. Mesons were expected to be

quark-antiquark bound states, while baryons were interpreted as bound states of three quarks. In Nature there are no indications of the existence of other multiquark states: in order to explain this fundamental evidence and to satisfy the Pauli exclusion principle for baryons, such as the Δ^{++} or the Ω^{-} which are made up of three quarks of the same flavor, the spin-one-half quarks had to carry a new quantum number [7], later termed *colour*. The modern version of this *constituent* quark model still successfully describes most of the qualitative features of the baryon spectroscopy.

A modern realization of Rutherford's experiment has shown us that quarks are real. This experiment is the deeply-inelastic scattering (DIS) of electrons (and, later, other leptons, including positrons, muons and neutrinos) off the nucleon, a program that was started in the late 1960's at SLAC [8] (for a review see also Ref. [9]). A high-energy electron interacts with the nucleon, via exchange of a highly virtual gauge boson. For a virtuality of $Q^2 > 1$ GeV^2, distances shorter than 0.2 fm are probed in the proton. The early DIS results led to an interpretation as elastic scattering of the electron off pointlike, spin-one-half, constituents of the nucleon [10–13], called *partons*. At first, this was understood in the so-called *parton model*: in this model, the nucleon is observed in the so-called *infinite momentum frame*, a Lorentz frame in which it is moving with large four-momentum: partons are assumed to move collinearly to the parent hadron, hence their transverse momenta and masses can be neglected. Lepton-nucleon scattering is then described in the *impulse approximation*, *i.e.* leptons interact with a parton in the nucleon via a fast exchange of a virtual photon. Partons are treated as free particles and all partons' self-interactions are neglected. In the impulse approximation, lepton-nucleon scattering is simply the incoherent sum of lepton interactions with the individual partons in the nucleon, which are carrying a fraction x of its four-momentum. These interactions can be computed in perturbation theory, and have to be weighted with the probability that the nucleon contains a parton with the proper value of x. This probability, denoted as $f_{q/p}(x)$, encodes the momentum density of any parton species q, with longitudinal fraction x, in a nucleon p, and is called Parton Distribution Function (PDF). This cannot be computed using perturbative theory, since it depends on the non-perturbative process that determines the structure of the nucleon; rather, it is a non-perturbative object to be determined from the experiment.

Partons carrying fractional electric charge were subsequently identified with the quarks. The existence of gluons was proved indirectly from a missing ($\sim 50\%$) contribution [14, 15] to the proton momentum not accounted for by the quarks. Later on, direct evidence for gluons was found in three-jet production in electron-positron annihilation [16–18]. From the observed angular distributions of the jets it became clear that gluons have spin one [19, 20].

The successful parton interpretation of DIS assumed that partons are almost free (*i.e.*, non-interacting) on the short time scales set by the high virtuality of the exchanged photon. This implied that the underlying theory of the strong interactions must actually be relatively weak on short time or, equivalently, distance scales [21]. In a groundbreaking development, Gross, Wilczek and Politzer showed in 1973 that the non-abelian theory of quarks and gluons, QCD, which had just been developed a few months earlier [22–24], possessed this remarkable feature of *asymptotic freedom* [25, 26], a discovery for which they were awarded the 2004 Nobel Prize for Physics. The interactions of partons at short distances, while weak in QCD, were then predicted to lead to visible effects in the experimentally measured DIS structure functions known as *scaling violations* [27, 28]. These essentially describe the response of the partonic structure of the proton to the resolving power of the virtual photon, set by its virtuality Q^2. The greatest triumph of QCD is arguably the prediction of scaling violations, which have been observed experimentally and verified with great precision. Deeply-inelastic scattering

thus paved the way for QCD.

Over the following two decades or so, studies of nucleon structure became ever more detailed and precise. This was partly due to increased luminosities and energies of lepton machines, eventually culminating in the HERA electron-proton collider [29]. Also, hadron colliders entered the scene. It was realized, again thanks to asymptotic freedom and factorization, which follows from it, that the partonic structure of the nucleon seen in DIS is universal, in the sense that a variety of sufficiently inclusive hadron collider processes, characterized by a large scale, admit a factorized description [30–41]. This offered the possibility of learning about other aspects of nucleon structure (and hence, QCD), for instance about its gluon content which is not primarily accessed in DIS. Being known with more precision, nucleon structure also allowed for new physics studies at hadron colliders, the outstanding example perhaps being the discovery of the $W^{\pm}$ and Z^0 bosons at CERN's $Sp\bar{p}S$ collider [42–44]. The Tevatron and the Large Hadron Collider (LHC) are the most recent continuations of this line of research, which has culminated in the discovery of the Higgs boson [45, 46], announced by ATLAS and CMS collaborations on the 4$^{\text{th}}$ July 2012.

Concerning spin physics, a milestone in the study of the nucleon was the advent of polarized electron beams in the early seventies [47]. This later on allowed for DIS measurements with polarized lepton beams and nucleon targets [48], and offered the possibility of studying whether quarks and antiquarks show on average preferred spin directions inside a polarized nucleon. The program of polarized DIS has been continuing ever since and it is now a successful branch of particle physics. Its most important result is the finding that quark and antiquark spins provide an anomalously small - only about $20\% - 30\%$ - amount of the proton spin [49, 50], firstly observed by the EMC experiment in the late 1980's. This finding, which opened a *spin crisis* in the understanding of the nucleon structure [51], has raised the interest of physicists in clarifying the potential role played by new candidates to the nucleon's spin, like gluons' polarizations and partons' orbital angular momenta. In parallel, there also was a very important line of research on polarization phenomena in hadron-hadron reactions in fixed-target kinematics. In particular, unexpectedly large single-transverse spin asymmetries were seen [52–56].

In the last decade, the advent of the Relativistic Heavy Ion Collider (RHIC), the first machine to collide polarized proton beams, started to probe the proton spin in new profound ways [57], complementary, but independent, to polarized DIS. In particular, more knowledge on the polarization of gluons in the proton and details of the flavor structure of the polarized quarks and antiquarks has been recently achieved, as we will discuss in detail in this Thesis. However, despite a flurry of experimental and theoretical activity, a complete and satisfactory udnerstanding of the so-called *spin puzzle* is still lacking.

1.2 Compelling questions in spin physics

The information on the proton spin structure is encoded in spin-dependent, or polarized, Parton Distribution Functions (PDFs) of quarks, antiquarks and gluons

$$\Delta f_{q/p}(x,Q^2) = f^{\uparrow}_{q/p}(x,Q^2) - f^{\downarrow}_{q/p}(x,Q^2)\,, \tag{1.1}$$

which are the momentum densities of partons q with helicity along ($\uparrow$) or opposite ($\downarrow$) the polarization direction of the parent nucleon p. The Q^2 dependence of the parton distributions, known as Q^2 evolution [58], is quantitatively predictable in QCD perturbation theory, thanks to asymptotic freedom. Physically, it may be thought of the consequence of the fact that partons

are observed with higher resolution when they are probed at higher scales; hence it is more likely that a struck quark has radiated one or more gluons so that it is effectively resolved into several partons, each with lower momentum fraction. Similarly, a struck quark may have originated from a gluon splitting into a quark-antiquark pair.

Polarized inclusive, neutral-current, DIS allows one to only access the flavor combinations $\Delta q^+ \equiv \Delta q + \Delta\bar{q}$, and the gluon polarization, though the latter is mostly determined indirectly by scaling violations. Of particular interest is the singlet quark antiquark combination $\Delta\Sigma = \sum_{q=u,d,s} \Delta q^+$, since its integral, known as the *axial charge*, yields the average of all quark and antiquark contributions to the proton spin:

$$\langle S_q \rangle \sim \frac{1}{2} \int_0^1 dx \Delta\Sigma(x, Q^2) \,. \tag{1.2}$$

The anomalously small value observed experimentally for this quantity, from almost three decades of DIS measurement after the EMC result, strenghten the common belief that only about a quarter of the proton spin is carried by quarks and antiquarks. The EMC result was followed by an intense scrutiny of the basis of the corresponding theoretical framework, which led to the realization [59, 60] that the perturbative behavior of polarized PDFs deviates from parton model expectations, according to which gluons decouple at large energy scale. The almost vanishing value measured by EMC for the singlet axial charge, Eq. (1.2), can be explained as a cancellation between a reasonably large quark spin contribution, *e.g.* $\Delta\Sigma \simeq 0.6 - 0.7$, as expected intuitively, and an anomalous gluon contribution. A large value of the gluon contribution to the proton spin is required to achieve such a cancellation, and QCD predicts that this contribution grows with the energy scale. Despite some experimental evidence has suggested that the gluon polarization in the nucleon may be rather small, we emphasize that it is instead still largely uncertain, as we will carefully demonstrate in this Thesis. Other candidates for carrying the nucleon spin can come from quark and gluon orbital angular momenta [61–63] (for a recent discussion on the spin decomposition see also Ref. [64]).

In any case, the results from polarized inclusive DIS clearly call for further investigations: we summarize in the following some of the outstanding questions to be aswered in spin physics.

With which accuracy do we know spin-dependent parton distributions? The assessment of singlet and gluon contributions to the proton spin requires in turn a determination of polarized parton distributions from available experimental data. In the last decade, several such determinations have been performed at next-to-leading order (NLO) in QCD [65–80], which is the current state-of-the-art accuracy for polarized fits. Some of them also include a significant amount of data other than DIS, namely from semi-inclusive DIS with identified hadrons in final states [75, 77, 80] or from polarized proton-proton collisions [75].

However, we notice that they are all based on the *standard* Hessian methodology for PDF fitting and uncertainty estimation. This approach is known [81] to potentially lead to an underestimation of PDF uncertainties, due to the limitations in the linear propagation of errors and, more importantly, to PDF parametrization in terms of fixed functional forms, both assumed in the *standard* methodology. These issues are especially delicate when the experimental information is scarce, like in the case of polarized data. In particular, in this Thesis we will clearly demonstrate that a more flexible PDF parametrization is better suited to analyse polarized experimental data without prejudice. This will lead to larger, but more faithful, estimates of PDF uncertainties with respect to the results of all available analyses. At least, one should conclude that our knowledge of parton's contribution to the nucleon spin is

much more uncertain than commonly believed, unless one makes strong *a priori* assumptions on their behavior in the unmeasured kinematic regions. The two following questions address in much detail some issues related to quarks and gluons separately.

How do gluons contribute to the proton spin? The interest in an accurate determination of the gluon polarization $\Delta g(x,Q^2)$ is of particular interest for both phenomenological and theoretical reasons.

On the phenomenological side, inclusive DIS allows for an indirect determination of the gluon distribution, through scaling violations. Since experimental data have a rather limited Q^2 lever arm, it follows that it is only weakly constrained. Processes other than inclusive DIS, which receive leading contributions from gluon initiated parton-level subprocesses, are better suited to provide direct information on the gluon distribution. In particular, these include open-charm production in fixed-target experiments and jet or semi-inclusive production in proton-proton collisions. However, the kinematic coverage of these data is limited: hence the integral of the gluon distribution can receive large contributions from the unmeasured region.

On the theoretical side, it is a remarkable feature of QCD that the gluon contribution to nucleon spin may well be significant even at large momentum scales. The reason is that the integral of $\Delta g(x,Q^2)$ evolves as $1/\alpha_s(Q^2)$ [59], that is, it rises logarithmically with Q. This peculiar evolution pattern is a very deep prediction of QCD, related to its so-called axial anomaly. It has inspired ideas that a reason for the smallness of the quark spin contribution should be sought in a *shielding* of the quark spins due to a particular perturbative part of the DIS process $\gamma^* g \to q\bar{q}$ [59]. The associated contributions arise only at order $\alpha_s(Q^2)$; however, the peculiar evolution of the first moment of the polarized gluon distribution would compensate this suppression. To be of any practical relevance, such models would require a large positive gluon spin contribution, $\langle \Delta g \rangle > 1$, even at low *hadronic* scales of a GeV or so. A very large polarization of the confining fields inside a nucleon, even though suggested by some nucleon models [82–85], would be a very puzzling phenomenon and would once again challenge our picture of the nucleon.

What are the patterns of up, down, and strange quark and antiquark polarizations? Inclusive DIS provides information only on the total flavor combinations Δq^+, $q = u,d,s$. Nevertheless, in order to understand the proton helicity structure in detail, one needs to learn about the various quark and antiquark densities, Δu, $\Delta\bar{u}$, Δd, $\Delta\bar{d}$ and Δs, $\Delta\bar{s}$ separately. This also provides an important additional test of the smallness of the quark spin contribution, and could reveal genuine flavor asymmetry $\Delta\bar{u} - \Delta\bar{d}$ in the proton sea, claimed by some models of nucleon structure [86, 87]. These predictions are often related to fundamental concepts such as the Pauli principle: since the proton has two valence-u quarks which primarily spin along with the proton spin direction, $u\bar{u}$ pairs in the sea will tend to have the u quark polarized opposite to the proton. Hence, if such pairs are in a spin singlet, one expects $\Delta\bar{u} > 0$ and, by the same reasoning, $\Delta\bar{d} < 0$. Such questions become all the more exciting due to the fact that rather large unpolarized asymmetries $\bar{u} - \bar{d} \neq 0$ have been observed in DIS and Drell-Yan measurements [88–90]. Further fundamental questions concern the strange quark polarization. The polarized DIS measurements point to a sizable negative polarization of strange quarks, in line with other observations of significant strange quark effects in nucleon structure.

What orbital angular momenta do partons carry? Quark and gluon orbital angular momenta are the other candidates for the carriers of the proton spin. Consequently, theoretical

work focused also on these in the years following *spin crisis*. A conceptual breakthrough was made in the mid 1990s when it was realized [62] that a particular class of *off-forward* nucleon matrix elements, in which the nucleon has different momentum in the initial and final states, measure total parton angular momentum. Simply stated, orbital angular momentum is $\overrightarrow{r} \times \overrightarrow{p}$, where the operator r can be viewed in Quantum Mechanics as a derivative with respect to momentum transfer. Thus, in analogy with the measurement of the Pauli form factor it takes a finite momentum transfer on the nucleon to access matrix elements with operators containing a factor r. It was also shown how these *off-forward* distributions, referred to as generalized parton distribution functions (GPDs), may be experimentally determined from certain exclusive processes in lepton-nucleon scattering, the prime example being deeply-virtual Compton scattering (DVCS) $\gamma^* p \to \gamma p$ [62]. A major emphasis in current and future experimental activities in lepton scattering is on the DVCS and related reactions.

What is the role of transverse spin in QCD? So far, we have only considered the helicity structure of the nucleon, that is, the partonic structure we find when we probe the nucleon when its spin is aligned with its momentum. Experimental probes with transversely polarized nucleons could also be studied, both at fixed-target and collider facilities, and it has been known for a long time now that very interesting spin effects are associated with this in QCD. Partly, this is known from theoretical studies which revealed that besides the helicity distributions Δf discussed above, for transverse polarization there is a new set of parton densities, called transversity [91, 92]. They are defined analogously to Eq. (1.1), but now for transversely polarized partons polarized along or opposite to the transversely polarized proton. Furthermore, if we allow quarks to have an intrinsic Fermi motion in the nucleon, they can be interpreted in the light of more fundamental objects, the so-called Transverse Momentum Dependent parton distribution functions (TMDs) [93], in which the dependence on the intrinsic transverse momentum $k_\perp$ is made explicit. We refer to [94] for a comprehensive review on TMDs and the transverse spin structure of the proton. Here, we only mention that the present knowledge of TMDs is comparable to that of PDFs in the early 1970's and very little is known about transversity. Both the COMPASS and HERMES experiments are taking data in polarized semi-inclusive DIS (SIDIS) to provide a better determination of those distributions [95–98].

1.3 Outline of the thesis

This Thesis addresses the three first questions in the above list, presenting a determination of spin-dependent parton distributions for the proton. In particular, two sets are obtained, the first based on inclusive DIS data only, the second also including the most recent data from polarized proton-proton collider. In comparison to other recent analyses, our study is performed within the NNPDF methodology, which makes use of robust statistical tools, including Monte Carlo sampling for error propagation and parametrization of PDFs in terms of neural networks. The methodology has been succesfully applied to the unpolarized case [99–111] with the goal of providing a faithful representation of the PDF underlying probability distribution. This is particularly relevant with polarized data, which are rather scarce and, in general, affected by larger uncertainties with respect to their unpolarized counterparts. Hence, our parton sets do not suffer from the theoretical bias introduced by a fixed functional form for PDF parametrization or by quadratic approximation in the Hessian propagation of errors, as in all other *standard* fits. For this reason, we consider our unbiased determination to be crucial for investigating to which extent the common belief, only partly supported by experiment, that about a quarter of

the nucleon spin is carried by quarks and antiquarks, while the gluon contribution is even much smaller, actually holds.

Our parton determinations are publicly released together with computational tools to access them, including `FORTRAN`, `C++` and *Mathematica* interfaces. Hence they could be used for any phenomenological study of hard scattering processes involving polarized hadrons in initial states. We should notice that, in addition to the investigation of the nucleon spin structure, such studies have recently included probes of different beyond-standard-model (BSM) scenarios [112] and the determination of the Higgs boson spin in the diphoton decay channel, by means of the linear polarization of gluons in an unpolarized proton [113].

In this Thesis, we will not address the study of either TMDs or GPDs, but we notice that the methods ilustrated here apply to the determination of any non-perturbative object from experimental data: hence they could be used to assess the determination of such new distributions in the future, as experimental data will reach more and more abundance and accuracy. Also, we do not describe either the apparatus which had to be developed to carry out spin physics experiments or the related technical challenges which had to be faced. A complete survey on these aspects can be found in Refs. [114, 115].

The outline of this Thesis is as follows.

Chapter 2. Polarized Deeply-Inelastic Scattering. We review the theoretical formalism for the description of Deeply-Inelastic Scattering (DIS) with both polarized lepton beams and nuclear targets. In particular we derive the expressions for the differential cross-section of the process in terms of polarized structure functions. We will restrict our discussion to the contribution arising from the exchange of a virtual photon between the lepton and the nucleon. This is indeed sufficient to describe currently available esperimental data, whose energy does not exceed a few hundreds of GeV and none of which come from neutrino beams: hence we do not include either the suppressed contribution to neutral-current DIS mediated by a Z^0 boson or charged-current DIS mediated by a $W^{\pm}$ boson. Then, we present the parton model expectations for DIS spin asymmetries introducing helicity-dependent, or polarized, Parton Distribution Functions (PDFs); we discuss how they are modified in the framework of perturbative Quantum Chromodynamics. We complete our theoretical overview on polarized DIS with a sketch of spin sum rules, a summary of the relevant phenomenological relations between structure functions and measured observables, and the formalism adopted to take into account kinematic target mass corrections.

Chapter 3. Phenomenology of polarized Parton Distributions. We review how a set of PDFs is usually determined from a global fit to experimental data. First, we sketch the general strategy for PDF determination and its main theoretical and methodological issues, focusing on those which are peculiar to the polarized case. Second, we summarize how some of these problems are addressed within the NNPDF methodology, which was developed in recent years with the goal of providing a statistically sound determination of PDFs and their uncertainties. Finally, we provide an overview on available polarized PDF sets.

Chapter 4. Unbiased polarized PDFs from inclusive DIS. We present the first determination of polarized PDFs based on the NNPDF methodology, `NNPDFpol1.0`. This analysis includes all available data from inclusive, neutral-current, polarized DIS and aims at an unbiased extraction of total quark-antiquark and gluon distributions at NLO accuracy. We discuss how the statistical distribution of experimental data is sampled with Monte Carlo generation of

pseudodata. We provide details of the QCD analysis and discuss the PDF parametrization in terms of neural networks; we emphasize the minimization strategy and the peculiarities in the polarized case. We present the NNPDFpol1.0 parton set, illustrating its statistical features and its stability upon the variation of several theoretical and methodological assumptions. We also compare our results to other recent polarized PDF sets. Finally, we discuss phenomenological implications for the spin content of the proton and the test of the Bjorken sum rule. The analysis presented in this Chapter has been published by the NNPDF collaboration as a refereed paper [116].

Chapter 5. Polarized Parton Distributions at an Electron-Ion Collider. We investigate the potential impact of inclusive DIS data from a future Electron-Ion Collider (EIC) on the determination of polarized PDFs. After briefly motivating our study, we illustrate which EIC pseudodata sets we use in our analysis and how the fitting procedure needs to be optimized. Resulting PDFs are compared to NNPDFpol1.0 throughout. Finally, we reassess their first moments and we give an estimate of the charm contribution to the g_1 structure function of the proton at an EIC. The analysis presented in this Chapter has been published by the NNPDF Collaboration as a refereed paper [117].

Chapter 6. A first global determination of unbiased polarized PDFs. We extend the analysis presented in previous Chapter in order to include in our parton set, on top of inclusive DIS data, also recent measurements of open-charm production in fixed-target DIS, and of jet and W production in polarized proton-proton collisions. Hence, we present the first global determination of polarized PDFs based on the NNPDF methodology: NNPDFpol1.1. After motivating our analysis, we review the theoretical description of the new processes and present the features of the relative experimental data we include in our study. We then turn to a detailed discussion of the way the NNPDFpol1.1 parton set is obtained via Bayesian reweighting of prior PDF Monte Carlo ensembles, followed by unweighting. We also present its main features in comparison to NNPDFpol1.0. Finally, we discuss some phenomenological implications for the spin content of the proton, based on our new polarized parton set. The analysis discussed in this Chapter has been presented in preliminary form in Refs. [118, 119].

Chapter 7. Conclusions and outlook. We will draw our conclusions, highlighting the main results presented in this Thesis. We also provide an outlook on future possible developments in the determination of polarized parton distributions within the NNPDF methodology.

Appendix A. Statistical estimators. We collect the definitions of the statistical estimators used in the NNPDF analyses presented in Chaps. 4-5-6. Despite they were already described in Refs. [99, 104, 120], we find useful to give them for completeness and ease of reference here.

Appendix B. A *Mathematica* interface to NNPDF parton sets. We present a package for handling both unpolarized and polarized NNPDF parton sets within .nb *Mathematica* notebook file, which appeared as a contribution to conference proceedings in Ref. [121]. This allows for performig PDF manipulations easily and quickly, thanks to the powerful features of the *Mathematica* software. The package was tailored to the users who are not familiar with FORTRAN or C++ programming codes, on which the standard available PDF interface, LHAPDF [122, 123], is based. However, since our *Mathematica* package includes all the

features available in the LHAPDF interface, any user can benefit from the interactive usage of PDFs within *Mathematica*.

Appendix C. The FONLL scheme for $g_1(x,Q^2)$ up to $\mathcal{O}(\alpha_s)$. We collect the relevant explicit formulae for the practical computation of the polarized DIS structure function $g_1(x,Q^2)$ within the FONLL approach [124] up to $\mathcal{O}(\alpha_s)$. In particular, we will restrict to the heavy charm quark contribution g_{1_c} to the polarized proton structure function g_1, which might be of interest for studies at an Electron-Ion Collider in the future, as mentioned in Chap. 5.

2 Polarized Deeply-Inelastic Scattering

This Chapter is devoted to a detailed discussion of Deeply-Inelastic Scattering (DIS) with both polarized lepton beams and nuclear targets. In particular, we will focus on neutral-current DIS, limited to the kinematic regime in which the exchange of a virtual photon between the lepton and the nucleon provides the leading contribution to the process. In Sec. 2.1, we rederive the expression for the differential cross-section of polarized DIS in terms of polarized structure functions. We present the naive parton model expectations for spin asymmetries in Sec. 2.2.1 and we discuss how they should be modified in the framework of QCD in Sec. 2.2.2. We complete our theoretical overview on polarized DIS with a sketch of spin sum rules in Sec. 2.3. Finally, we summarize the relevant phenomenological relations between structure functions and measured observables in Sec. 2.4, and the formalism adopted to take into account kinematic target mass corrections in Sec. 2.5.

2.1 General formalism

Let us consider the inclusive, neutral-current, inelastic scattering of a polarized lepton (electron or muon) beam off a polarized nucleon target,

$$l(\ell) + N(P) \to l'(\ell') + X(P_X)\,, \tag{2.1}$$

where the four-momenta of the incoming (outgoing) lepton l (l'), the nucleon target N and the undetected final hadronic system X are labelled as ℓ (ℓ'), P, and P_X respectively. If the momentum transfer involved in the reaction is much smaller than the Z^0 boson mass, as it is customary at past and present polarized DIS facilities, only virtual photon exchange provides a sizable contribution to the process, as depicted in Fig. 2.1.

In order to work out the kinematics, we denote the nucleon mass, M, the lepton mass, m_ℓ, the covariant spin four-vector of the incoming (outgoing) lepton s_ℓ ($s_{\ell'}$) and the spin four-vector

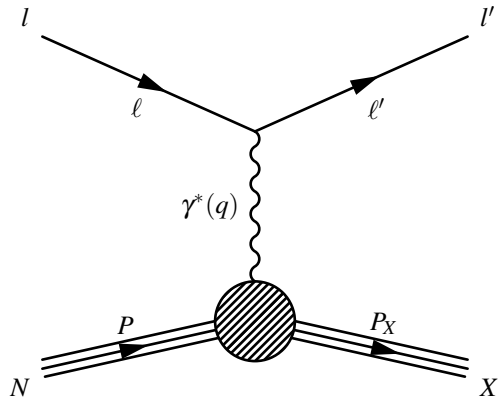

Figure 2.1: Neutral-current DIS mediated by a virtual photon.

of the nucleon, S. In the target rest frame, we define the four-momenta to be

$$\begin{array}{rcll} \ell & = & (E,\ell) & \text{incoming lepton,} \\ \ell' & = & (E',\ell') & \text{outgoing lepton,} \\ P & = & (M,\mathbf{0}) & \text{proton.} \end{array} \tag{2.2}$$

The deeply inelastic regime is identified by the invariant mass W of the final hadronic system to be much larger than the nucleon mass, namely

$$W^2 = M^2 + Q^2\frac{1-x}{x} \gg M^2\,. \tag{2.3}$$

This allows us to neglect all masses and to use the approximation

$$\ell^2 = \ell'^2 \approx 0, \quad E \approx |\ell|, \quad E' \approx |\ell'|. \tag{2.4}$$

Based on these assumptions, only two kinematical variables (besides the centre of mass energy $s = (\ell + P)^2$ or, alternatively, the lepton beam energy E) are needed to describe the process in Eq. (2.1). They can be chosen among the following invariants:

$$Q^2 = -q^2 = (\ell - \ell')^2 = 2EE'(1-\cos\theta) = 4EE'\sin^2\left(\frac{\theta}{2}\right) \tag{2.5}$$

the laboratory-frame photon square momentum,

$$\nu = E - E' = \frac{P\cdot q}{M} \tag{2.6}$$

the laboratory-frame photon energy,

$$x = \frac{Q^2}{2P\cdot q} = \frac{Q^2}{2M\nu} \tag{2.7}$$

the Bjorken scaling variable,

$$y = \frac{P\cdot q}{P\cdot \ell} = \frac{\nu}{E} \tag{2.8}$$

the energy fraction lost by the incoming lepton ℓ,

where θ is the scattering angle between the incoming and the outgoing lepton beams.

The differential cross-section for lepton-nucleon scattering then reads

$$d^3\sigma(\ell N \to \ell' X) = \frac{1}{2s}\frac{d^3\ell'}{(2\pi)^3 2E'}\sum_{s_{\ell'}}\sum_X \int d\Pi_X |\mathscr{M}(\ell N \to \ell' X)|^2, \tag{2.9}$$

where

$$\int d\Pi_X = \int \frac{d^3\mathbf{P}_X}{(2\pi)^3 2E_X}(2\pi)^4\delta^4(P+q-P_X) \tag{2.10}$$

is the phase space factor for the unmeasured hadronic system and

$$\begin{aligned} |\mathscr{M}(\ell N \to \ell' X)|^2 &= \frac{e^4}{q^4}\left[\bar{u}(\ell,s_\ell)\gamma^\nu u(\ell',s_{\ell'})\bar{u}(\ell',s_{\ell'})\gamma^\mu u(\ell,s_\ell)\right] \\ &\times \left[\langle P,S|J_\nu^\dagger(q)|P_X\rangle\langle P_X|J_\mu(q)|P,S\rangle\right] \end{aligned} \tag{2.11}$$

is the squared amplitude including the Fourier transform of the quark electromagnetic current $J^\mu(q)$ flowing through the hadronic vertex. Since we are describing the scattering of polarized leptons on a polarized target, with no measurement of the outgoing lepton polarization nor of the final hadronic system, in Eq. (2.9) we must sum over the final lepton spin $s_{\ell'}$ and over all final hadrons X, but must not average over the initial lepton spin, nor sum over the nucleon spin.

It is customary to define the leptonic tensor

$$L^{\mu\nu} = \sum_{s_{\ell'}}\left[\bar{u}(\ell,s_\ell)\gamma^\nu u(\ell',s_{\ell'})\bar{u}(\ell',s_{\ell'})\gamma^\mu u(\ell,s_\ell)\right] \tag{2.12}$$

and the hadronic tensor

$$W_{\mu\nu} = \frac{1}{2\pi}\sum_X \int d\Pi_X \left[\langle P,S|J_\nu^\dagger(q)|P_X\rangle\langle P_X|J_\mu(q)|P,S\rangle\right] \tag{2.13}$$

in order to rewrite Eq. (2.9) as

$$d^3\sigma = \frac{1}{2s}\frac{e^4}{Q^4}2\pi L^{\mu\nu}W_{\mu\nu}\frac{d^3\ell'}{(2\pi)^3 2E'} \tag{2.14}$$

or, in the target rest frame, where $s = 2ME$, and considering $d^3\ell' = E'^2 dE' d\Omega$, $d\Omega = d\cos\theta d\varphi$,

$$\frac{d^3\sigma}{d\Omega dE'} = \frac{\alpha_{em}^2}{2MQ^4}\frac{E'}{E}L^{\mu\nu}W_{\mu\nu}\,. \tag{2.15}$$

This is the differential cross-section to find the scattered lepton in solid angle $d\Omega$ with energy $(E', E'+dE')$ usually quoted in the literature (see for example Ref. [125]). In Eq. (2.15), α_{em} is the fine-structure electromagnetic constant, while φ is the azimuthal angle of the outgoing lepton. The variables E' and θ are natural ones, in that they are measured in the laboratory frame, by detecting the scattered lepton. However, it is more convenient to perform the variable transformation $(E',\theta) \to (x,y)$ and to express the differential cross-section in terms of the latter quantities as

$$\frac{d^3\sigma}{dxdyd\varphi} = \frac{\alpha_{em}^2 y}{2Q^4}L^{\mu\nu}W_{\mu\nu}\,, \tag{2.16}$$

since these are gauge-invariant and dimensionless.

2.1.1 Leptonic tensor

In a completely general way, the leptonic tensor $L^{\mu\nu}$ can be decomposed into a symmetric and an antisymmetric part under $\mu \leftrightarrow \nu$ interchange

$$L_{\mu\nu} = L^{(S)}_{\mu\nu}(\ell,\ell') + iL^{(A)}_{\mu\nu}(\ell,s_\ell,\ell')\,. \tag{2.17}$$

Recalling the identity satisfied by the spinor $u(p,s)$, for a fermion with polarization vector s^μ,

$$u(p,s)\bar{u}(p,s) = (\not{p}+m)\frac{1}{2}(1+\gamma_5\not{s})\,, \tag{2.18}$$

and summing only on $s_{\ell'}$, the leptonic tensor reads

$$L_{\mu\nu} = \mathrm{Tr}\left\{\gamma_\mu(\not{\ell}'+m_\ell)\gamma_\nu(\not{\ell}+m_\ell)\frac{1}{2}(1+\gamma_5\not{s}_\ell)\right\}. \tag{2.19}$$

Trace computation via Dirac algebra finally leads to (retaining lepton masses)

$$L^{(S)}_{\mu\nu} = 2\left[\ell_\mu\ell'_\nu + \ell_\nu\ell'_\mu - g_{\mu\nu}(\ell\cdot\ell' - m_\ell^2)\right], \tag{2.20}$$

$$L^{(A)}_{\mu\nu} = 2m_\ell\varepsilon_{\mu\nu\rho\sigma}s_\ell^\rho(\ell-\ell')^\sigma\,. \tag{2.21}$$

If the incoming lepton is longitudinally polarized, its spin vector can be expressed as

$$s_\ell^\mu = \frac{\lambda_\ell}{m_\ell}(|\ell|,\hat{\ell}E),\;\; \hat{\ell} = \frac{\ell}{|\ell|}\,, \tag{2.22}$$

i.e. it is parallel ($\lambda_\ell = +1$) or antiparallel ($\lambda_\ell = -1$) to the direction of motion ($\lambda_\ell = \pm 1$ is twice the lepton helicity). Eq. (2.21) then reads

$$L^{(A)}_{\mu\nu} = 2\lambda_\ell\varepsilon_{\mu\nu\rho\sigma}\ell^\rho(\ell-\ell')^\sigma = 2\lambda_\ell\varepsilon_{\mu\nu\rho\sigma}\ell^\rho q^\sigma\,. \tag{2.23}$$

Notice that the lepton mass m_ℓ appearing in Eq. (2.21) has been cancelled by the denominator in Eq. (2.22), which refers to a longitudinally polarized lepton. In contrast, if it is transversely polarized, that is, $s_\ell^\mu = s_{\ell\perp}^\mu$, no such cancellation occurs and the corresponding contribution is suppressed by a factor m_ℓ/E.

2.1.2 Hadronic tensor

The hadronic tensor $W_{\mu\nu}$ allows for a decomposition analogous to Eq. (2.17), that is

$$W_{\mu\nu} = W^{(S)}_{\mu\nu}(q,P) + iW^{(A)}_{\mu\nu}(q;P,S)\,, \tag{2.24}$$

where the symmetric and antisymmetric parts can be expressed in terms of two pairs of structure functions, W_1, W_2 and G_1, G_2, as

$$\begin{aligned}\frac{1}{2M}W^{(S)}_{\mu\nu} &= \left(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2}\right)W_1(P\cdot q,q^2)\\ &+ \frac{1}{M^2}\left(P_\mu - \frac{P\cdot q}{q^2}q_\mu\right)\left(P_\nu - \frac{P\cdot q}{q^2}q_\nu\right)W_2(P\cdot q,q^2)\,,\end{aligned} \tag{2.25}$$

$$\frac{1}{2M}W^{(A)}_{\mu\nu} = \varepsilon_{\mu\nu\rho\sigma}q^{\rho}\Big\{MS^{\sigma}G_1(P\cdot q,q^2) + \frac{1}{M}\left[P\cdot qS^{\sigma}-S\cdot qP^{\sigma}\right]G_2(P\cdot q,q^2)\Big\}. \tag{2.26}$$

It is customary to introduce the dimensionless structure functions

$$F_1(x,Q^2)\equiv MW_1(\nu,Q^2), \qquad F_2(x,Q^2)\equiv \nu W_2(\nu,Q^2), \tag{2.27}$$

$$g_1(x,Q^2)\equiv M^2\nu G_1(\nu,Q^2), \qquad g_2(x,Q^2)\equiv M\nu^2 G_2(\nu,Q^2), \tag{2.28}$$

and to rewrite the symmetric and antisymmetric parts of the hadronic tensor as

$$W^{(S)}_{\mu\nu} = 2\left(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2}\right)F_1(x,Q^2) + \frac{2}{P\cdot q}\left(P_\mu-\frac{P\cdot q}{q^2}q_\mu\right)\left(P_\nu-\frac{P\cdot q}{q^2}q_\nu\right)F_2(x,Q^2), \tag{2.29}$$

$$W^{(A)}_{\mu\nu} = \frac{2M\varepsilon_{\mu\nu\rho\sigma}q^{\rho}}{P\cdot q}\left\{S^{\sigma}g_1(x,Q^2)+\left[S^{\sigma}-\frac{S\cdot q}{P\cdot q}P^{\sigma}\right]g_2(x,Q^2)\right\}. \tag{2.30}$$

These expressions give the most general gauge invariant decompositions of the hadronic tensor for pure electromagnetic, parity conserving, interaction, see *e.g* [126] and references therein. A theoretical description of both neutral and charged current DIS at energies of the order of the weak boson masses (or higher) must include parity violating terms in the decomposition of the hadronic tensor. Because of them, one no longer has the correspondence that its symmetric part, Eq. (2.29), is spin independent and its antisymmetric part, Eq. (2.30), is spin dependent. Actually, the spin-dependent part of the hadronic tensor becomes a superposition of symmetric and antisymmetric pieces. Four more independent structure functions appear in this case, usually called F_3, g_3, g_4 and g_5: the first multiplies a term independent from the lepton or nucleon spin four-vector, while the three latter do not. Such a general decomposition of the hadronic tensor in DIS, with particular emphasis on the polarized case, can be found in Ref. [127].

Since experimental data on polarized DIS are taken with electron or muon beams at momentum transfer values not exceeding $Q^2 \sim 100$ GeV^2, we can safely neglect the contribution from weak boson exchange to describe them properly. However, the most general decomposition will be needed in the future to handle neutral-current DIS at high energies, as it may be performed at an Electron-Ion Collider [128, 129], or charged-current DIS with neutrino beams, as it might be available at a neutrino factory [130].

2.1.3 Polarized cross-sections differences

Insertion of Eqs. (2.17) and (2.24) into Eq. (2.16) yields the expression

$$\frac{d^3\sigma^{s_\ell,S}}{dxdyd\varphi} = \frac{\alpha^2_{em}y}{2Q^4}\left[L^{(S)}_{\mu\nu}W^{\mu\nu(S)}-L^{(A)}_{\mu\nu}W^{\mu\nu(A)}\right], \tag{2.31}$$

and differences of cross-sections with opposite target spin helicity states single out the tensor antisymmetric parts

$$\frac{d^3\sigma^{\lambda_\ell,+S}}{dxdyd\varphi}-\frac{d^3\sigma^{\lambda_\ell,-S}}{dxdyd\varphi} = -\frac{\alpha^2_{em}y}{Q^4}L^{(A)}_{\mu\nu}W^{\mu\nu(A)}. \tag{2.32}$$

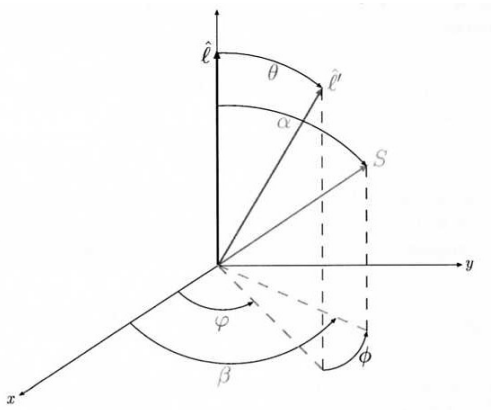

Figure 2.2: Azimuthal and polar angles of the final lepton momentum, ℓ', and the nucleon polarization vector, **S**. The initial lepton moves along the positive z-axis. Often one defines the $(\hat{\ell},\hat{\ell}')$ lepton plane as the $\varphi = 0$ plane.

In the target rest frame, using the notation of Fig. 2.2, we parametrize the nucleon spin four vector as

$$S^\mu = (0,\hat{\mathbf{S}}) = (0,\sin\alpha\cos\beta,\sin\alpha\sin\beta,\cos\alpha)\,, \tag{2.33}$$

where we have assumed $|\mathbf{S}| = 1$. Taking the direction of the incoming lepton to be along the z-axis, we also have

$$\ell^\mu = E(1,0,0,1), \tag{2.34}$$
$$\ell'^\mu = E'(1,\sin\theta\cos\varphi,\sin\theta\sin\varphi,\cos\varphi)\,. \tag{2.35}$$

Supposing now that the incoming lepton is polarized collinearly to its direction of motion, *i.e.* $\lambda_\ell = +1$ and $s_\ell^\mu = \frac{\ell^\mu}{m_\ell}$, we have

$$\begin{aligned} L^{(A)}_{\mu\nu}W^{\mu\nu(A)} &= -\frac{8}{\nu}\Big\{\left[(\ell\cdot q)(S\cdot q) - q^2(S\cdot q)\right]g_1(x,Q^2) \\ &- q^2\left[S\cdot\ell - \frac{(P\cdot\ell)(S\cdot q)}{P\cdot q}\right]g_2(x,Q^2)\Big\}\,. \end{aligned} \tag{2.36}$$

Note that, owing to currrent conservation, we have $L_{\mu\nu}q^\mu = L_{\mu\nu}q^\nu = 0$ and the terms proportional to q^μ and q^ν in Eq. (2.29) do not contribute when contracted with the leptonic tensor. Explicit computation of four-momentum products in this equation yields to a new expression for the differential asymmetry (2.32)

$$\begin{aligned} &\frac{d^3\sigma^{+;+S}}{dxdyd\phi} - \frac{d^3\sigma^{+;-S}}{dxdyd\phi} \\ &= -\frac{4\alpha_{em}^2}{Q^2}y\Big\{\cos\alpha\left[\left(\frac{E}{\nu}+\frac{E'}{\nu}\cos\theta\right)g_1(x,Q^2) + \frac{2EE'}{\nu^2}(\cos\theta-1)g_2(x,Q^2)\right] \\ &+ \sin\alpha\cos\phi\left[\frac{E'}{\nu}\sin\theta g_1(x,Q^2) + \frac{2EE'}{\nu^2}\sin\theta g_2(x,Q^2)\right]\Big\}\,, \end{aligned} \tag{2.37}$$

where $\phi = \beta - \varphi$ is the azimuthal angle between the lepton plane and the $(\hat{\ell}, \hat{\mathbf{S}})$ plane. Notice that the r.h.s. of this equation is not expressed in terms of the usual invariants x and y; to this purpose, let us define the $\mathscr{O}(1/Q)$ quantity

$$\gamma \equiv \frac{Mx}{Q} \tag{2.38}$$

and work out a little algebra to obtain the final expression for the differential polarized cross-section difference (2.32)

$$\begin{aligned} & \frac{d^3\sigma^{+;+S}}{dxdyd\phi} - \frac{d^3\sigma^{+;-S}}{dxdyd\phi} \\ = & -\frac{4\alpha_{em}^2}{Q^2}\left\{\left[\left(2-y-\frac{\gamma^2y^2}{4}\right)g_1(x,Q^2) - \gamma^2 y g_2(x,Q^2)\right]\cos\alpha \right. \\ + & \left.\sqrt{1-y-\frac{\gamma^2y^2}{4}}\left[y g_1(x,Q^2) + 2g_2(x,Q^2)\right]\sin\alpha\cos\phi\right\}. \end{aligned} \tag{2.39}$$

Results obtained so far need a few comments.

1. The terms *longitudinal* and *transverse*, when speaking about the nucleon polarization, are somewhat ambiguous, insofar as a reference axis is not specified. From an experimental point of view, the *longitudinal* or *transverse* nucleon polarizations are defined with respect to the lepton beam axis, thus *longitudinal* (*transverse*) indicates the direction parallel (orthogonal) to this axis. We will use the large arrows $\Rightarrow$ ($\Uparrow$) to denote these two cases respectively.

2. Eq. (2.39) refers to the scattering of longitudinally polarized (positive helicity) leptons off a nucleon with positive or negative polarization along an arbitrary direction $\hat{\mathbf{S}}$. The cross-section difference is proportional to $L_{\mu\nu}^{(A)}$, according to Eqs. (2.21)-(2.32), which contains a small factor m_ℓ; as already noticed, this small factor is cancelled by the $1/m_\ell$ factor appearing in the lepton-helicity four-vector, Eq. (2.22). This would not be the case with transversely polarized leptons, for which one would have $s^\mu = (0,\hat{\mathbf{s}})$, with $\hat{\mathbf{s}}\cdot\ell = 0$. Then, transversely polarized leptons lead to tiny cross-section asymmetries of order of m_ℓ/E.

3. Eq. (2.39) can be specialized to particular cases of the nucleon polarization. For longitudinally polarized nucleons, that is $\hat{\mathbf{S}} \parallel \ell$, one has $\alpha = 0$ and the differential cross-section, integrated over the azimuthal angle ϕ, reads

$$\frac{d^3\sigma^{+;\Rightarrow}}{dxdy} - \frac{d^3\sigma^{+;\Leftarrow}}{dxdy} = -\frac{8\pi\alpha_{em}^2}{Q^2}\left[\left(2-y-\frac{\gamma^2y^2}{2}\right)g_1(x,Q^2) - \gamma^2 y g_2(x,Q^2)\right]; \tag{2.40}$$

for nucleons polarized transversely to the lepton direction, one has $\alpha = \pi/2$ and the differential cross-section is

$$\frac{d^3\sigma^{+;\Uparrow}}{dxdy} - \frac{d^3\sigma^{+;\Downarrow}}{dxdy} = -\frac{8\pi\alpha_{em}^2}{Q^2}\gamma\sqrt{1-y-\frac{\gamma^2y^2}{4}}\left[y g_1(x,Q^2) + 2g_2(x,Q^2)\right]\cos\varphi\,. \tag{2.41}$$

In general, the term proportional to g_2 is suppressed by a factor γ, Eq. (2.38), with respect to the one proportional to g_1: in the Bjorken limit, Eqs. (2.40)-(2.41) decouple and only g_1 is asymptotically relevant. We emphasize that, in the case of transverse polarizations, both g_1 and g_2 structure functions equally contribute, but the whole cross-section difference is suppressed by the overall factor Eq. (2.38) of order $1/Q$. In the following, we will mostly concentrate on the longitudinally polarized cross-section difference, Eq. (2.40).

4. From Eq. (2.31), it is straightforward to obtain the unpolarized cross-section for inclusive DIS by averaging over spins of the incoming lepton (s_ℓ) and of the nucleon (S) and by integrating over the azimuthal angle φ. It reads

$$\frac{d^2\sigma^{unp}}{dxdy} = 2\pi\frac{1}{2}\sum_{s_\ell}\frac{1}{2}\sum_{S}\frac{d^2\sigma^{s_\ell,S}}{dxdy} = \frac{\alpha_{em}^2 y}{2Q^4}L_{\mu\nu}^{(S)}W^{\mu\nu(S)}\,. \tag{2.42}$$

Finally, the unpolarized cross-section, expressed in terms of the usual unpolarized structure functions F_1 and F_2, when neglecting contributions of order M^2/Q^2, is

$$\frac{d^2\sigma^{unp}}{dxdy} = \frac{4\pi\alpha_{em}^2 s}{Q^4}\left[xy^2F_1(x,Q^2) + (1-y)F_2(x,Q^2)\right]\,. \tag{2.43}$$

2.2 Factorization of structure functions

In the previous Section, we have parametrized the hadronic tensor which describes the coupling of the virtual photon to the composite nucleon in terms of four structure functions, namely F_1, F_2 and g_1, g_2, see Eqs. (2.29)-(2.30). We have then derived the expression for the differential cross-section asymmetries of longitudinally and transversely polarized nucleons in terms of g_1 and g_2, Eqs. (2.40)-(2.41). In principle, by performing DIS experiments with nucleons polarized both longitudinally and transversely, one should learn about the structure functions g_1 and g_2, as we will discuss in detail in Sec. 2.4 below. In this Section, we would like to provide a description of DIS in the framework of QCD, and in particular give a *factorized* expression for the structure function g_1. Actually, even though QCD is asymptotically free, the computation of any cross-section does involve non-perturbative contributions, since the initial and final states are not the fundamental degrees of freedom of the theory, but compound states of quarks and gluons. As we shall see, the factorization theorem allows for the separation of a hard, perturbative and process-dependent part from a low energy, process-independent contribution. The latter is given by the Parton Distribution Functions (PDFs), which parametrize our ignorance on the inner structure of the proton. In order to deal with the factorized expression for the structure function g_1, we will first provide the leading-order (LO) QCD description of polarized DIS as borrowed from the naive parton model; we will then give a heuristic development of the next-to-leading order (NLO) perturbative QCD corrections to polarized DIS, focusing on their effects on the g_1 structure function.

2.2.1 Naive parton-model expectations

The information on the *a priori* unknown structure of a polarized nucleon is carried by the structure functions g_1 and g_2. As discussed in Sec. 2.1.2, they can only be functions of x and Q^2. In the naive parton model [10–13], they allow for simple expressions, since the cross-section for

lepton-nucleon scattering is regarded as the incoherent sum of point-like interactions between the lepton and a free, massless parton

$$\frac{d^2\sigma}{dxdy} = \sum_q e_q^2 f_{q/p}(x) \frac{d\hat{\sigma}}{dy}\,. \tag{2.44}$$

In this expression, e_q is the fractional charge carried by the parton q, $\frac{d\hat{\sigma}}{dy}$ is the cross-section for the elementary QED process $\ell q \to \ell' q$ and $f_{q/p}$ is the parton distribution function (PDF), the probability density distribution for the momentum fraction x of any parton q in a nucleon p. In the simple picture provided by the parton model, PDFs do not depend on the scale Q^2 and the structure functions are observed to obey the scaling law $g_{1,2}(x,Q^2) \to g_{1,2}(x)$ [10]. This property is related to the assumption that the transverse momentum of the partons is small. In the framework of QCD, however, the radiation of hard gluon from the quarks violates this assumption beyond leading order in pertubation theory, as we will discuss below. Of course, the naive parton model predates QCD, but we consider it is of great value for its intuitive nature.

If we specialize Eq. (2.44) to polarized cross-section asymmetries, we should write

$$\frac{d^2\sigma^{+;\Rightarrow}}{dxdy} - \frac{d^2\sigma^{+;\Leftarrow}}{dxdy} = \sum_q e_q^2 \Delta f_{q/p}(x) \left[\frac{d\hat{\sigma}^{+;+}}{dy} - \frac{d\hat{\sigma}^{+;-}}{dy} \right], \tag{2.45}$$

where $\frac{d\hat{\sigma}^{\lambda_\ell,\lambda_q}}{dy}$ denotes the elementary cross-section retaining helicity states of both the lepton (λ_ℓ) and the struck parton (λ_q). We also introduced helicity-dependent, or polarized PDFs, $\Delta f_{q/p}$, defined as the momentum densities of partons with spin aligned parallel or antiparallel to the longitudinally polarized parent nucleon:

$$\Delta f_{q/p}(x) \equiv f^{\uparrow}_{q/p}(x) - f^{\downarrow}_{q/p}(x)\,. \tag{2.46}$$

Explicit expressions for the elementary cross-sections appearing in the r.h.s of Eq. (2.45) are easily computed at the lowest order in QED:

$$\frac{d\hat{\sigma}^{+,+}}{dy} = \frac{4\pi\alpha_{em}^2}{Q^2}\frac{1}{y}, \quad \frac{d\hat{\sigma}^{+,-}}{dy} = \frac{4\pi\alpha_{em}^2}{Q^2}\frac{(1-y)^2}{y}\,. \tag{2.47}$$

Replacing these elementary cross-sections in Eq. (2.45) leads to the expression

$$\frac{d^2\sigma^{+;\Rightarrow}}{dxdy} - \frac{d^2\sigma^{+;\Leftarrow}}{dxdy} = \frac{4\pi\alpha_{em}^2}{Q^2}\left[\sum_q e_q^2 \Delta q(x)(2-y) \right], \tag{2.48}$$

which can be directly compared to Eq. (2.40). Neglecting terms of order $\mathcal{O}(\gamma^2)$, we finally obtain the naive parton model relations between structure functions $g_1(x)$, $g_2(x)$ and the polarized distribution $\Delta f_{q/p}(x)$:

$$g_1(x) = \frac{1}{2}\sum_q e_q^2 \Delta f_{q/p}(x)\,, \tag{2.49}$$

$$g_2(x) = 0\,. \tag{2.50}$$

These results require a few comments.

1. The structure function g_1, Eq. (2.49), can be unambigously expressed in terms of quark and antiquark polarized parton distributions. Assuming the number of flavors $n_f = 3$, we can define the singlet, $\Delta\Sigma$, and nonsinglet triplet, ΔT_3, and octet, ΔT_8, combinations of polarized quark densities

$$\begin{aligned} \Delta\Sigma &= \Delta u^+ + \Delta d^+ + \Delta s^+ , && (2.51)\\ \Delta T_3 &= \Delta u^+ - \Delta d^+ , && (2.52)\\ \Delta T_8 &= \Delta u^+ + \Delta d^+ - 2\Delta s^+ , && (2.53) \end{aligned}$$

where $\Delta q^+ = \Delta q + \Delta\bar{q}$, $q = u,d,s$ are the total parton densities. Then the structure function g_1, Eq (2.49), can be cast into the form

$$g_1(x) = \frac{1}{9}\Delta\Sigma(x) + \frac{1}{12}\Delta T_3(x) + \frac{1}{36}\Delta T_8(x) . \tag{2.54}$$

The structure function g_1 does not receive any contribution from gluons, yet we shall see in Sec. 2.2.2 that it is not true in the framework of QCD beyond Born approximation.

2. The structure function g_2 is zero, Eq. (2.50). However, non-zero values of g_2 can be obtained by allowing the quarks to have an intrinsic Fermi motion inside the nucleon. In this case, there is no unambiguous way to calculate g_2 in the naive parton model. We will not further investigate this issue in this Thesis, a detailed discussion of the problem can be found in Ref. [126].

3. The expression of the hadronic tensor $W_{\mu\nu}$ in terms of the Fourier transform of the nucleon matrix elements of the elctromagnetic current $J_\mu(x)$, Eq. (2.13), allows for the use of Wilson Operator Product Expansion (OPE) to give the moments of the structure functions g_1 and g_2 in terms of hadronic matrix elements of certain operators multiplied by perturbatively calculable Wilson coefficient functions. In this framework, it can be shown (see *e.g.* [126]) that the first moment of the singlet quark density

$$a_0 = \int_0^1 dx \Delta\Sigma(x) \tag{2.55}$$

is related to the matrix element of the flavor singlet axial current. Hence, a_0 can be interpreted as the contribution of quarks to the proton's spin and one expects it should be twice the expectation value of the sum of the z-components of quark and antiquark spins

$$a_0 = 2\langle S_z^{\text{quarks+antiquarks}} \rangle . \tag{2.56}$$

Uncritically, one should expect $a_0 \approx 1$, while in a more realistic relativistic model one finds $a_0 \approx 0.6$ [61]. In the late 80s, this expectation was found to be in contrast with the anomalously small value measured by the European Muon Collaboration at CERN [49, 50]. This result could be argued to imply that the sum of the spins carried by the quarks in a proton, $\langle S_z^{\text{quarks+antiquarks}} \rangle$, was consistent with zero rather than $1/2$, suggesting a *spin crisis* in the parton model [51]. This led to an intense scrutiny of the basis of the theoretical calculation of the structure function g_1 and the spin crisis was immediately recognized not to be a fundamental problem, but rather an interesting property of spin structure functions to be understood in terms of QCD. We will give a summary of such a description in the following Section.

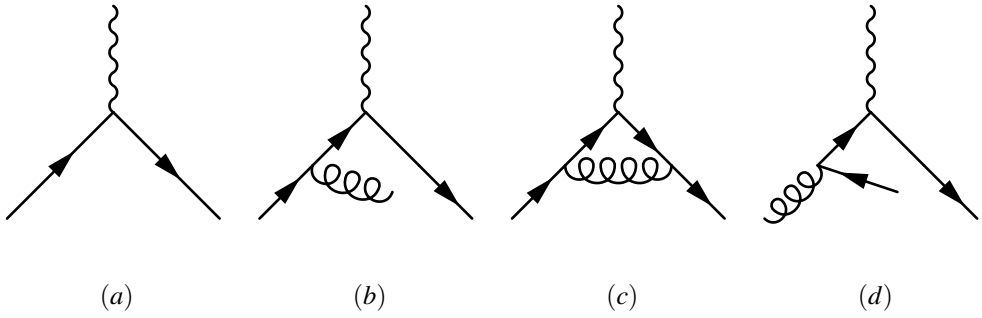

Figure 2.3: Leading contribution (a) and next-to-leading order corrections $(b),(c),(d)$ to DIS.

2.2.2 QCD corrections and evolution

The parton model predates the formulation of QCD. As soon as QCD is accepted as the theory of strong interactions, with quark and gluon fields as the fundamental fields, one should describe the lepton scattering off partons in the nucleon perturbatively. At Born level, the interaction is described by the Feynman diagram in Fig. 2.3-(a) as the tree-level scattering of a quark (or antiquark) off the virtual photon γ^*. In this case, quarks are free partons and one recovers the parton model expressions for the structure functions, Eqs. (2.49)-(2.50). At order $\mathcal{O}(\alpha_s)$, several new contributions appear: the one-loop correction, Fig. 2.3-(b), the emission of a gluon, Fig. 2.3-(c), and the process initiated by a gluon which then splits into a quark-antiquark pair, the so-called photon-gluon fusion process, Fig. 2.3-(d). The main impact of the QCD interactions is twofold: first, they introduce a mild, calculable, logarithmic Q^2 dependence in the parton distributions; second, the correction in Fig. 2.3-(d) generate a contribution to the structure function g_1 arising from the polarization of the gluons in the nucleon. We shall describe both these effects in the following.

Scale dependence of parton distributions

When including NLO corrections, Fig. 2.3-(b)-(c), problems arise from the so-called collinear singularities linked to the effective masslessness of quarks. The factorization theorem is probed [131] to allow for a separation of the process into a hard and a soft part and for the absorption of the infinity into the soft part (the PDF), which in any case cannot be calculated and has to be determined from experimental data. The scale at which the separation is made is called factorization scale μ^2. Schematically, one finds terms of the form $\alpha_s \ln(Q^2/M)$, which one splits as follows

$$\alpha_s \ln \frac{Q^2}{M} = \alpha_s \ln \frac{Q^2}{\mu^2} + \alpha_s \ln \frac{\mu^2}{M^2}\,; \tag{2.57}$$

one then absorbs the first term on the r.h.s. of Eq. (2.57) into the hard part of the process and the second term into the soft part. The factorization scale μ^2 can be chosen arbitrarily and, in exact calculations, physical results must not depend on it. In practice, since we never calculate to all orders in perturbation theory, it can make a difference what value we choose, but it turns out that an optimal choice is $\mu^2 = Q^2$. Consequently, parton distributions no longer obey exact Bjorken scaling, but develop a slow logarithmic dependence on Q^2. Actually, if one keeps

only the leading log terms (proportional to $\alpha_s \ln(Q^2/\mu^2)$), one finds that the parton model expression, Eqs. (2.49)-(2.50), still hold, provided the replacement

$$\Delta f_{q/p}(x) \longrightarrow \Delta f_{q/p}(x, Q^2) \tag{2.58}$$

to the Q^2-dependent PDF is made. We can think of the scale dependence of PDFs within the following picture. As the scale increases the photon starts to *see* evidence for the point-like valence quarks within the proton. If the quarks were non-interacting, no further structure would be resolved increasing the resolving scale: the Bjorken scaling would set in, and the naive parton model would be satisfactory. For this reason, we can consider the naive parton model as the approximation to Born-level QCD. However, QCD predicts that on increasing the resolution, one should see that each quark is itself surrounded by a cloud of partons. The number of resolved partons which share the proton's momentum increases with the scale.

The perturbative dependence of the polarized PDFs on the scale Q^2 is given by the Altarelli-Parisi evolution equations [58], a set of $(2n_f+1)$ coupled integro-differential equations. It is customary to write them in the *evolution basis*, *i.e.* in terms of linear combinations of the individual quark distributions such that the $(2n_f+1)$ equations maximally decouple from each other. To this purpose, we define the polarized gluon distribution Δg as in Eq. (2.46) and the singlet and nonsinglet quark PDF combinations as

$$\Delta q_{\rm NS}(x, Q^2) \equiv \sum_{i=1}^{n_f} \left(\frac{e_i^2}{\langle e^2 \rangle} - 1 \right) \left[\Delta q_i(x, Q^2) + \Delta \bar{q}_i(x, Q^2) \right], \tag{2.59}$$

$$\Delta \Sigma(x, Q^2) \equiv \sum_{i=1}^{n_f} \left[\Delta q_i(x, Q^2) + \Delta \bar{q}_i(x, Q^2) \right], \tag{2.60}$$

where $\Delta q_i(x, Q^2)$ and $\Delta \bar{q}_i(x, Q^2)$ are the scale-dependent quark and antiquark polarized densities of flavor i, also defined according to Eq. (2.46). The evolution equations are coupled for the singlet quark-antiquark combination and the gluon distribution

$$\mu^2 \frac{\partial}{\partial \mu^2} \begin{pmatrix} \Delta\Sigma(x,\mu^2) \\ \Delta g(x,\mu^2) \end{pmatrix} = \frac{\alpha_s(\mu^2)}{2\pi} \begin{pmatrix} \Delta P_{qq}^{\rm S} & 2n_f \Delta P_{qg}^{\rm S} \\ \Delta P_{qg}^{\rm S} & \Delta P_{gg}^{\rm S} \end{pmatrix} \otimes \begin{pmatrix} \Delta\Sigma(x,\mu^2) \\ \Delta g(x,\mu^2) \end{pmatrix}, \tag{2.61}$$

while the nonsinglet quark-antiquark combination evolves independently as

$$\mu^2 \frac{\partial}{\partial \ln \mu^2} \Delta q_{\rm NS}(x, \mu^2) = \frac{\alpha_s(\mu^2)}{2\pi} \Delta P_{qq}^{\rm NS} \otimes \Delta q_{\rm NS}(x, \mu^2). \tag{2.62}$$

In Eqs. (2.61)-(2.62), $\Delta P_{ij}^{\rm S/NS}$, $i, j = q, g$, denotes the singlet/nonsinglet spin-dependent splitting functions for quarks and gluons and $\otimes$ is the shorthand notation for the convolution product with respect to x

$$f \otimes g = \int_x^1 \frac{dy}{y} f\left(\frac{x}{y}\right) g(y). \tag{2.63}$$

We notice that Eqs. (2.61)-(2.62) hold to all orders in perturbative theory, hence splitting functions may be expanded in powers of the strong coupling α_s:

$$\Delta P_{ij}^{p} = \Delta P_{ij}^{p(0)}(x) + \frac{\alpha_s(\mu^2)}{2\pi} \Delta P_{ij}^{p(1)}(x) + \mathscr{O}(\alpha_s^2), \tag{2.64}$$

where $p =$S, NS, g and $i, j = q, g$. The splitting functions for polarized PDFs were computed at LO in Ref. [58] for the first time, the computation was then extended to NLO in Refs. [132,133], while only partial results are available at NNLO so far [134].

The solution of the Altareli-Parisi equations may be written as

$$\Delta f_i(x,Q^2) = \sum_j \Gamma_{ij}(x,\alpha_s,\alpha_s^0) \otimes \Delta f_j(x,Q_0^2)\,, \tag{2.65}$$

where $f_i = \Sigma, q_{\rm NS}, g$, $\Delta f_j(x,Q_0^2)$ are the input PDFs parametrized at initial scale Q_0^2 to be determined from experimental data, $\Gamma_{ij}(x,\alpha_s,\alpha_s^0)$ are the evolution factors, and we have used the shorthand notation

$$\alpha_s \equiv \alpha_s(Q^2) \qquad \alpha_s^0 \equiv \alpha_s(Q_0^2)\,. \tag{2.66}$$

The evolution factors also satisfy evolution equations

$$\mu^2 \frac{\partial}{\partial \mu^2} \Gamma_{ij}(x,\alpha_s,\alpha_s^0) = \sum_k P_{ik}(x,\alpha_s) \otimes \Gamma_{kj}(x,\alpha_s,\alpha_s^0)\,, \tag{2.67}$$

with boundary conditions $\Gamma_{ij}(x,\alpha_s^0,\alpha_s^0) = \delta_{ij}\delta(1-x)$. The QCD evolution equations are most easily solved using Mellin moments since then all convolutions become simple products, and the equations can be solved in a closed form. We refer to [135–137] for a comprehensive discussion about details concerning such a technique.

The gluon contribution to the g_1 structure function

Another important consequence of QCD corrections is the rise of a contribution to the g_1 structure function from the polarization of gluons in the nucleon, see for instance Fig. 2.3-(d). For this reason, the factorized leading-twist expression for the structure function g_1 reads, instead of Eq. (2.54),

$$g_1(x,Q^2) = \frac{\langle e^2 \rangle}{2} \left[C_{\rm NS} \otimes \Delta q_{\rm NS} + C_{\rm S} \otimes \Delta\Sigma + 2 n_f C_g \otimes \Delta g \right], \tag{2.68}$$

where $\langle e^2 \rangle = n_f^{-1} \sum_{i=1}^{n_f} e_i^2$ is the average charge, with n_f the number of active flavors and e_i their electric charge, and $\otimes$ denotes the convolution product with respect to x, Eq. (2.63). Besides, $\Delta q_{\rm NS}$ and $\Delta\Sigma$ are the scale-dependent nonsinglet and singlet quark PDF combinations, Eqs. (2.59)-(2.60) and Δg is the gluon PDF. Finally, $C_{\rm NS}$, $C_{\rm S}$ and C_g are the corresponding coefficient functions related to calculable short-distance cross-sections, for hard photon-quark and photon-gluon cross-sections respectively. Coefficient functions are perturbative objects and may be expanded in powers of the strong coupling α_s

$$C_p(x,\alpha_s) = C_p^{(0)}(x) + \frac{\alpha_s(\mu^2)}{2\pi} C_p^{(1)}(x) + \mathscr{O}(\alpha_s)\,, \tag{2.69}$$

with $p =$S, NS, g and $i, j = q, g$. At the lowest order in α_s, $C_{\rm NS}^{(0)} = C_{\rm S}^{(0)} = \delta(1-x)$ and $C_g^{(0)} = 0$ [58], hence the structure function g_1 decouples from the gluon contribution, see Eq. (2.68), and the parton model prediction, Eq. (2.54) is recovered. Polarized coefficient functions have been computed up to $\mathscr{O}(\alpha_s^2)$ so far [138].

Notice incidentally that the substitution of Eq. (2.65) into Eq. (2.68) lead to the relation

$$g_1(x,Q^2) = \frac{\langle e^2 \rangle}{2} \sum_{j=\mathrm{NS},\Sigma,g} K_j(x,\alpha_s,\alpha_s^0) \otimes \Delta f_j(x,Q_0^2)\,, \tag{2.70}$$

where the hard kernel defined as

$$K_j(z,\alpha_s,\alpha_s^0) = \sum_{k=\mathrm{NS},\mathrm{S},g} C_k(x,\alpha_s) \otimes \Gamma_{ij}(x,\alpha_s,\alpha_s^0) \tag{2.71}$$

is completely computable in perturbation theory. Hence, in Eq. (2.70) we have fully separated the perturbative and the non-perturbative parts entering the structure function g_1. Besides, the hard kernels in Eq. (2.71) are independent of the particular set of input PDFs, and may thus be computed separately once and for all, suitably interpolated and stored. This is of uttermost importance while performing a fit of PDFs to experimental data, as we will further delineate in Sec. 4.2, since this involves the evaluation of only the one set of convolutions Eq. (2.70), which is amenable to computational optimization.

The gluonic term in the expression of the g_1 structure function, see for instance Eq. (2.68), can be shown [59, 139, 140] to entail an additional contribution to the flavor singlet axial charge of the form

$$a_0^{\mathrm{gluons}} = -n_f \frac{\alpha_s(Q^2)}{2\pi} \int_0^1 dx \Delta g(x,Q^2)\,. \tag{2.72}$$

We remind that n_f is the number of light flavors, u, d, s, and heavy flavors are assumed not to contribute. Hence, the naive parton model expectation for the axial current a_0, Eq. (2.55), should be replaced by

$$a_0 = \int_0^1 dx \Delta\Sigma(x,Q^2) + a_0^{\mathrm{gluons}}\,. \tag{2.73}$$

At first sight, one should expect that the gluonic term in Eq. (2.73) would not survive at large Q^2, since it looks like an α_s correction which would disappear at large Q^2 as the running coupling α_s vanishes. However, the first moment of the gluon contribution $\int_0^1 dx \Delta g(x,Q^2)$ grows as $[\alpha_s(Q^2)]^{-1}$ for large values of Q^2, as dictated by Altarelli-Parisi evolution equations, see Eq. (2.61).* Hence, the gluon does not decouple from g_1 asymptotically and the parton model expression for the structure function g_1, Eq. (2.49), is not recovered in perturbative QCD even in the limit $\alpha_s \to 0$. The Q^2 behavior of the first moment of the polarized gluon density was originally derived when the QCD evolution equations were first written down in x space [58]. In fact, the first moment of the polarized gluon splitting function is finite and proportional to the first coefficient of the QCD beta function, which establishes the quoted relation with the running coupling $\alpha_s(Q^2)$. This relation between the Q^2 evolution of the first moment of the polarized gluon density and the running coupling is induced by the axial anomaly [139] corresponding to the QCD version of the anomalous *triangle diagram* [141, 142].

As a consequence, the definition of the singlet quark first moment becomes totally ambiguous, because two generic definitions differ by terms of order $\alpha_s(Q^2) \int_0^1 dx \Delta g(x,Q^2)$. For the first moment, what is formally a NLO correction is potentially of the same size. Owing to Eq. (2.73), the singlet quark first moment, Eq. (2.55), defined directly from the structure function g_1 and used by the EMC experiment when the *spin crisis* [51] was announced, does

*Higher moments are instead decreasing functions of $\log Q^2$, falling at a faster rate than for the unpolarized gluon density.

not have to coincide with the constituent quark value, *i.e.* the total fraction of the spin carried by quarks. Only for exactly conserved quantities do the corresponding values for constituent and parton quarks have to coincide. The first moments of the quark densities are in general only conserved at LO by the QCD evolution, but, due to the axial anomaly, the singlet quark first moment defined from g_1 is not conserved in higher orders. We conclude that a definition of the singlet quark density $\Delta\Sigma(x,Q^2)$ must be carefully specified, as further discussed in Sec. 2.2.3 below.

The result quoted in Eqs. (2.72)-(2.73) was advocated to reconcile the EMC result with the theoretical expectation for the proton spin content. As explained above, what the EMC experiment actually observed was the singlet axial charge a_0, Eq. (2.73). The almost vanishing value measured for this quantity can be explained as a cancellation between a reasonably large quark spin contribution, *e.g.* $\Delta\Sigma \simeq 0.6-0.7$, as expected intuitively, and the anomalous gluon contribution. However, in order to accomplish this cancellation, one should require a large gluon spin contribution at $\langle Q^2\rangle \simeq 10$ GeV^2, *e.g.* $\int_0^1 dx\Delta g(x,Q^2) \simeq 4$. As we have explained, the latter momentum grows indefinitely as Q^2 increases, so that such a large value cannot be ruled out. Hence, we must carefully investigate with which accuracy we are able to determine each term in Eq. (2.73), particularly the gluon contribution, by scrutinizing both the available experimental data and the methodology we use to determine parton distributions from them. This is exactly the goal of this Thesis: we will find that the gluon is still largely uncertain, in contrast to somewhat common belief, and that its determination is still a challenge in the field of spin physics.

2.2.3 Scheme dependence of parton distribution moments

Beyond leading order, coefficient and splitting functions are no longer universal, hence even though the scale dependence of the structure function g_1 is determined uniquely, at least up to higher order corrections, its separation into contributions due to quarks and gluons is scheme dependent (and thus essentially arbitrary). The NLO coefficient functions may be modified by a change of the factorization scheme which is partially compensated by a corresponding change in the NLO splitting functions, hence both are required for a consistent NLO computation. A comprehensive discussion of scheme dependence can be found in Ref. [114]. Here, we summarize the main features of the schemes which are commonly used in the analysis of polarized DIS.

1. The most popular renormalization scheme is the so-called $\overline{\rm MS}$ scheme [132, 133], in which the first moment of the gluon coefficient function vanishes. In this scheme, the gluon density does not contribute to the first moment of the structure function g_1 and the scale-dependent singlet axial charge is equal to the singlet quark first moment:

$$a_0(Q^2) = \int_0^1 dx\Delta\Sigma(x,Q^2)\Big|_{\overline{\rm MS}}. \tag{2.74}$$

   Also, the first moments of the nonsinglet triplet and octet PDF combinations

$$a_3 = \int_0^1 dx\Delta T_3(x,Q^2) \qquad a_8 = \int_0^1 dx\Delta T_8(x,Q^2) \tag{2.75}$$

   are independent of Q^2.

2. An alternatively scheme is the so-called Adler-Bardeen scheme [143] defined such that the first moment of the singlet PDF combination is independent from Q^2, thus it can be identified with the total quark helicity. The polarized gluon density is defined as in the $\overline{\rm MS}$ scheme, but directly contributes to the singlet axial charge, and consequently to the first moment of g_1, which now reads

$$a_0(Q^2) = \int_0^1 \Delta\Sigma(x,Q^2)\Big|_{\rm AB} - n_f \frac{\alpha_s(Q^2)}{2\pi}\int_0^1 \Delta g(x,Q^2)\,. \tag{2.76}$$

The relation between the first moments of the singlet quark combination in AB and $\overline{\rm MS}$ renormalization schemes is then simply obtained by comparing Eqs. (2.74)-(2.76). As discussed above, the difference is proportional to $\alpha_s(Q^2)\int_0^1 dx\Delta g(x,Q^2)$ and is due to the anomalous nonconservation of the singlet axial current. At LO, it is scale-invariant: this implies that the first moment of the polarized gluon distribution increases as $1/\alpha_s(Q^2)$ with Q^2, hence the gluon contribution in Eq. (2.76) is not asymptotically suppressed by powers of α_s. As a consequence, this scheme dependence does not vanish at large Q^2, and the definition of the singlet quark first moment is therefore maximally ambiguous.

2.3 Sum rules

Moments of structure functions are a powerful tool to study fundamental properties of nucleon structure, like the total momentum fraction carried by quarks or the total contribution of quark helicities to the spin of the nucleon. While a complete description of structure functions based on fundamental QCD principles may be unattainable for now, moments of structure functions can be directly compared to rigorous theoretical results, like sum rules, lattice QCD calculations and chiral perturbation theory.

The light-cone expansion of the current product in Eq. (2.13) implies that the n-th moments of the structure functions g_1 and g_2, at leading twist are given by [114]

$$\int_0^1 dx x^{n-1} g_1(x,Q^2) = \frac{1}{2}\sum_i \delta_i a_n^i C_{1,i}^n(Q^2,\alpha_s) \qquad n=1,3,5,\ldots \tag{2.77}$$

$$\int_0^1 dx x^{n-1} g_2(x,Q^2) = \frac{1-n}{2n}\sum_i \delta_i \left[a_n^i C_{1,i}^n(Q^2,\alpha_s) - d_n^i C_{2,i}^n(Q^2,\alpha_s)\right] \qquad n=3,5,7,\ldots \tag{2.78}$$

where the δ_i are numerical coefficients, the $C_i^n(Q^2,\alpha_s)$ are the coefficient functions and the a_n^i and d_n^i are related to the hadronic matrix elements of the local operator. The label i indicates what kind of operator is contributing: for flavor-nonsinglet operators, only quark fields an their covariant derivatives occur.

In the case of the first moment of the g_1 structure function, Eq. (2.77) can be recast as

$$\Gamma_1^{p,n} \equiv \int_0^1 dx g_1(x,Q^2) = \frac{1}{12}\left[C_{\rm NS}(Q^2)\left(\pm a_3 + \frac{1}{3}a_8\right) + \frac{4}{3}C_{\rm S}(Q^2)a_0\right], \tag{2.79}$$

where the plus (minus) sign refers to a proton (neutron) target. In the above, a_3 and a_8 are measures of the proton matrix elements of an $SU(3)$ flavor octet of quark axial-vector currents.

The octet of axial-vector currents is precisely the set of currents that controls the weak β-decays of the neutron and of the spin-$1/2$ hyperons. Consequently, a_3 and a_8 can be expressed

in terms of two parameters F and D measured in hyperon β decays [144]

$$a_3 = F + D = 1.2701 \pm 0.0025\,, \qquad a_8 = \frac{1}{\sqrt{3}}(3F - D) = 0.585 \pm 0.025\,. \tag{2.80}$$

It follows that a measurement of $\Gamma_1^p(Q^2)$ in polarzed DIS can be interpreted as a measurement of $a_0(Q^2)$. Indeed, Eq. (2.79) can be rewritten as

$$C_{\rm S}(Q^2)a_0(Q^2) = 9\Gamma_1^p(Q^2) - \frac{1}{2}C_{\rm NS}(Q^2)(3F + D)\,. \tag{2.81}$$

Since the two terms on the r.h.s. are roughly of the same order, the value of a_0 arises from a large cancellation between them (see *e.g.* Ref. [145]). The present measured value for a_0 is still disturbingly small, as briefly noted at the end of previous Sec. 2.2.2.

Finally, in going from the case of a proton to a neutron, a_0 and a_8 in Eq. (2.79) remain unchanged, whereas a_3 reverses its sign. One thus finds the Bjorken sum rule [146, 147]

$$\Gamma_1^p(Q^2) - \Gamma_1^n(Q^2) = \frac{1}{6}C_{\rm NS}(Q^2)a_3\,, \tag{2.82}$$

which was originally derived from current algebra and isospin asymmetry. A comparison with experimental data, thus allows for a direct test of isospin, as well as of the predicted scale dependence. Furthermore, since the nonsiglet coefficient function $C_{\rm NS}$ is known up to three loops [148], Eq. (2.82) potentially provides a theoretically very accurate handle on the strong coupling α_s [67]. This feature will be discussed in Sec. 4.5.2 in the framework of a determination of an unbiased parton set from inclusive polarized DIS data.

Finally, we notice that relations like Eq. (2.79) can be obtained also for the g_2 structure function. These include the Burkhardt-Cottingham sum rule [149] and the Efremov-Leader-Teryaev sum rule [150], for a discussion of which we refer to [126, 151]. Here we only notice that data on g_2 are not yet accurate enough for a significant test on them.

2.4 Phenomenology of polarized structure functions

Experimental information on the structure functions $g_1(x,Q^2)$ and $g_2(x,Q^2)$ is extracted from measured cross-section asymmetries, both longitudinal, $A_\parallel$, and transeverse, $A_\perp$. These are defined by considering longitudinally polarized leptons scattering off a hadronic target, polarized either longitudinally or transversely with respect to the collision axis, and read

$$A_\parallel = \frac{d\sigma^{\rightarrow\Rightarrow} - d\sigma^{\rightarrow\Leftarrow}}{d\sigma^{\rightarrow\Rightarrow} + d\sigma^{\rightarrow\Leftarrow}}\,; \quad A_\perp = \frac{d\sigma^{\rightarrow\Uparrow} - d\sigma^{\rightarrow\Downarrow}}{d\sigma^{\rightarrow\Uparrow} + d\sigma^{\rightarrow\Downarrow}}\,. \tag{2.83}$$

The numerator of these expressions is given by Eqs. (2.40)-(2.41), while the denominator is twice the unpolarized cross-section, Eq. (2.43).

Inversion of Eqs. (2.40)-(2.41) gives the explicit relation between the polarized structure functions and the measurable asymmetries Eq. (2.83)

$$g_1(x,Q^2) = \frac{F_1(x,Q^2)}{(1+\gamma^2)(1+\eta\zeta)}\left[(1+\gamma\zeta)\frac{A_\parallel}{D} - (\eta-\gamma)\frac{A_\perp}{d}\right], \tag{2.84}$$

$$g_2(x,Q^2) = \frac{F_1(x,Q^2)}{(1+\gamma^2)(1+\eta\zeta)}\left[\left(\frac{\zeta}{\gamma}-1\right)\frac{A_\parallel}{D} + \left(\eta+\frac{1}{\gamma}\right)\frac{A_\perp}{d}\right], \tag{2.85}$$

where we have defined the kinematic factors

$$d = \frac{D\sqrt{1-y-\gamma^2y^2/4}}{1-y/2}, \tag{2.86}$$

$$D = \frac{1-(1-y)\varepsilon}{1+\varepsilon R(x,Q^2)}, \tag{2.87}$$

$$\eta = \frac{\varepsilon\gamma y}{1-\varepsilon(1-y)}, \tag{2.88}$$

$$\zeta = \frac{\gamma(1-y/2)}{1+\gamma^2y/2}, \tag{2.89}$$

$$\varepsilon = \frac{4(1-y)-\gamma^2y^2}{2y^2+4(1-y)+\gamma^2y^2}. \tag{2.90}$$

The unpolarized structure function F_1 and unpolarized structure function ratio R which enter the definition Eq. (2.84-2.85) of the asymmetries may be expressed in terms of F_2 and F_L by

$$F_1(x,Q^2) \equiv \frac{F_2(x,Q^2)}{2x[1+R(x,Q^2)]}\left(1+\gamma^2\right), \tag{2.91}$$

$$R(x,Q^2) \equiv \frac{F_L(x,Q^2)}{F_2(x,Q^2)-F_L(x,Q^2)}. \tag{2.92}$$

The longitudinal and transverse asymmetries are sometimes expressed in terms of the virtual photo-absorption asymmetries A_1 and A_2 according to

$$A_\parallel = D(A_1+\eta A_2), \qquad A_\perp = d(A_2-\zeta A_1), \tag{2.93}$$

where

$$A_1(x,Q^2) \equiv \frac{\sigma^T_{1/2}-\sigma^T_{3/2}}{\sigma^T_{1/2}+\sigma^T_{3/2}}, \qquad A_2(x,Q^2) \equiv \frac{2\sigma^{TL}}{\sigma^T_{1/2}+\sigma^T_{3/2}}. \tag{2.94}$$

Recall that $\sigma^T_{1/2}$ and $\sigma^T_{3/2}$ are cross-sections for the scattering of virtual transversely polarized photons (corresponding to longitudinal lepton polarization) with helicity of the photon-nucleon system equal to 1/2 and 3/2 respectively, and σ^{TL} denotes the interference term between the transverse and longitudinal photon-nucleon amplitudes. In the limit $M^2 \ll Q^2$ Eqs. (2.93) reduce to $D = A_\parallel/A_1$, $d = A_\perp/A_2$, thereby providing a physical interpretation of d and D as depolarization factors.

Using Eqs. (2.93) in Eqs. (2.84-2.85) we may express the structure functions in terms of A_1 and A_2 instead:

$$g_1(x,Q^2) = \frac{F_1(x,Q^2)}{1+\gamma^2}\left[A_1(x,Q^2)+\gamma A_2(x,Q^2)\right], \tag{2.95}$$

$$g_2(x,Q^2) = \frac{F_1(x,Q^2)}{1+\gamma^2}\left[\frac{A_2}{\gamma}-A_1\right]. \tag{2.96}$$

We are interested in the structure function $g_1(x,Q^2)$, whose moments are proportional to nucleon matrix elements of twist-two longitudinally polarized quark and gluon operators, and therefore can be expressed in terms of longitudinally polarized quark and gluon distributions.

Using Eqs. (2.84-2.85), we may obtain an expression of it in terms of the two asymmetries $A_\parallel$, $A_\perp$, or, using Eqs. (2.95-2.96), in terms of the two asymmetries A_1, A_2. Clearly, up to corrections of $\mathcal{O}(M/Q)$, g_1 is fully determined by $A_\parallel$, which coincides with A_1 up to $\mathcal{O}(M/Q)$ terms, while g_2 is determined by $A_\perp$ or A_2. It follows that, even though in principle a measurement of both asymmetries is necessary for the determination of g_1, in practice most of the information comes from $A_\parallel$ or A_1, with the other asymmetry only providing a relatively small correction unless Q^2 is very small.

It may thus be convenient to express g_1 in terms of $A_\parallel$ and g_2

$$g_1(x,Q^2) = \frac{F_1(x,Q^2)}{1+\gamma\eta}\frac{A_\parallel}{D} + \frac{\gamma(\gamma-\eta)}{\gamma\eta+1}g_2(x,Q^2), \tag{2.97}$$

or, equivalently, in terms of A_1 and g_2

$$g_1(x,Q^2) = A_1(x,Q^2)F_1(x,Q^2) + \gamma^2 g_2(x,Q^2). \tag{2.98}$$

It is then possible to use Eq. (2.97) or Eq. (2.98) to determine $g_1(x,Q^2)$ from a dedicated measurement of the longitudinal asymmetry, and an independent determination of $g_2(x,Q^2)$.

In practice, experimental information on the transverse asymmetry and structure function g_2 is scarce [152–154]. However, the Wilson expansion for polarized DIS implies that the structure function g_2 can be written as the sum of a twist-two and a twist-three contribution [155]:

$$g_2(x,Q^2) = g_2^{\rm t2}(x,Q^2) + g_2^{\rm t3}(x,Q^2). \tag{2.99}$$

The twist-two contribution to g_2 is simply related to g_1. One finds

$$g_2^{\rm t2}(x,Q^2) = -g_1(x,Q^2) + \int_x^1 \frac{dy}{y} g_1(y,Q^2) \tag{2.100}$$

which in Mellin space becomes

$$g_2^{\rm t2}(N,Q^2) = -\frac{N-1}{N} g_1(N,Q^2). \tag{2.101}$$

It is important to note that $g_2^{\rm t3}$ is not suppressed by a power of M/Q in comparison to $g_2^{\rm t2}$, because in the polarized case the availability of the spin vector allows the construction of an extra scalar invariant. Nevertheless, experimental evidence suggests that $g_2^{\rm t3}$ is compatible with zero at low scale $Q^2 \sim M^2$. Fits to $g_2^{\rm t3}$ [156,157], as well as theoretical estimates of it [156,158] support the conclusion that

$$g_2(x,Q^2) \approx g_2^{\rm t2}(x,Q^2) \equiv g_2^{\rm WW}(x,Q^2), \tag{2.102}$$

which is known as the Wandzura-Wilczek [155] relation. The effect of such an assumption on a determination of parton distributions from experimental data could be tested to compare to results obtained with the opposite assumption, *i.e.* $g_2(x,Q^2) = 0$. We will follow this strategy in our extraction of unbiased polarized PDFs presented in Chap. 4.

2.5 Target mass corrections

A large part of experimental data in polarized DIS are taken at relatively low values of Q^2, typically a few GeV^2, and at medium-to-large x values. In this kinematical region, the target-mass factor γ, Eq. (2.38), is of order unity with the finite value of the nucleon mass M, hence

contributions which are elsewhere suppressed may play a relevant role. These corrections are usually referred to as kinematic higher-twist terms. Another source of terms suppressed by inverse powers of Q^2 arises from the Wilson expansion of the hadronic tensor, Eqs. (2.29)-(2.30), *i.e.* from matrix elements of operators of non-leading twist. These corrections are referred to as dynamical higher-twist terms. A detailed analysis of the effect of twist-3 and twist-4 corrections to g_1, as well as a twist-3 correction to the g_2 structure function on the determiantion of a set of polarized PDFs from inclusive DIS data has been recently presented [78]. They find that they could have a sizable effect, particularly in the high-x and low-Q^2 kinematic region.

In order to include exactly the effect of kinematic target mass corrections, TMCs henceforth, one has to deal with Eqs. (2.97)-(2.98), supplemented by some model assumption on g_2, as discussed in Sec. 2.4. This is required since experimental data on g_2 are restricted to a limited range in (x,Q^2) and are affected by large uncertainties. Following Ref. [159], target mass corrections assume simple expressions in Mellin space, where they read

$$\begin{aligned}\tilde{g}_1(N,Q^2) &= g_1(N,Q^2)+\frac{M^2}{Q^2}\frac{N(N+1)}{(N+2)^2} \\ &\times \left[(N+4)g_1(N+2,Q^2)+4\frac{N+2}{N+1}g_2(N+2,Q^2)\right]+\mathscr{O}\left(\frac{M^2}{Q^2}\right)^2 \qquad (2.103)\\ \tilde{g}_2(N,Q^2) &= g_2(N,Q^2)+\frac{M^2}{Q^2}\frac{N(N-1)}{(N+2)^2} \\ &\times \left[N\frac{N+2}{N+1}g_2(N+2,Q^2)-g_1(N+2,Q^2)\right]+\mathscr{O}\left(\frac{M^2}{Q^2}\right)^2 . \qquad (2.104)\end{aligned}$$

Here, we have denoted by $\tilde{g}_{1,2}(N,Q^2)$ the Mellin space structure functions with TMCs included, while $g_{1,2}(N,Q^2)$ are the structure functions determined in the $M=0$ limit. These expressions can be specialized under assumptions for g_2. In particular, in the Wandzura-Wilczek case, substituting Eq. (2.101) in Eq. (2.104) and taking the inverse Mellin transform, we get

$$\tilde{g}_1(x,Q^2)=\frac{1}{2\pi i}\int dN\,x^{-N}\left[1+\frac{M^2x^2}{Q^2}\frac{(N-2)^2(N-1)}{N^2}\right]g_1(N,Q^2), \qquad (2.105)$$

where we have shifted $N\to N-2$ in the term proportional to M^2. Inverting the Mellin transform we then obtain

$$\begin{aligned}\tilde{g}_1(x,Q^2) &= g_1(x,Q^2)+\frac{M^2x^2}{Q^2}\left[-5g_1(x,Q^2)-x\frac{\partial g_1(x,Q^2)}{\partial x}\right. \\ &\left.+\int_x^1\frac{dy}{y}\left(8g_1(y,Q^2)+4g_1(y,Q^2)\log\frac{x}{y}\right)\right]. \qquad (2.106)\end{aligned}$$

Conversely, if we simply set $g_2=0$, we have

$$\tilde{g}_1(x,Q^2)=\frac{1}{2\pi i}\int dN\,x^{-N}\left[1+\frac{M^2x^2}{Q^2}\frac{(N^2-4)(N-1)}{N^2}\right]g_1(N,Q^2), \qquad (2.107)$$

whence

$$\begin{aligned}\tilde{g}_1(x,Q^2) &= g_1(x,Q^2)+\frac{m^2x^2}{Q^2}\left[-g_1(x,Q^2)-x\frac{\partial g_1(x,Q^2)}{\partial x}\right. \\ &\left.-\int_x^1\frac{dy}{y}\left(4g_1(y,Q^2)+4g_1(y,Q^2)\log\frac{x}{y}\right)\right]. \qquad (2.108)\end{aligned}$$

The usefulness of these realtions will be apparent in Chap. 4, where we will use them in an unbiased determination of a polarized parton set based on inclusive polarized DIS data.

3 Phenomenology of polarized Parton Distributions

In this Chapter, we review how a set of parton distribution functions is usually determined from a global fit to experimental data. In Sec. 3.1, we delineate the general strategy for PDF determination and its main theoretical and methodological issues, focusing on those which are peculiar to the polarized case. In Sec. 3.2, we summarize how some of these problems are addressed within the NNPDF methodology, which was developed in recent years to provide a statistically sound determination of parton distributions and their uncertainties. In Sec. 3.3, we finally conclude with an overview on available polarized PDF sets.

3.1 General strategy for *standard* PDF determination

In principle, the general strategy for parton fitting can be simply stated. Thanks to the factorization theorem, theoretical predictions for the various measured observables are expressed as the convolution between coefficient functions and parton distributions. The former are perturbative quantities, computed within field theory at desired accuracy, but different for each partonic subrocess contributing to the observable. The latter are non-perturbative objects, but universal, *i.e.* they do not depend on the observable under investigation. In order to determine parton distributions from experimental data, they have to be parametrized, usually at an initial energy scale Q_0^2, and then randomly initialized. They then need to be evolved up to the energy scale Q^2, relevant for the measurement under investigation, by solving Altarelli-Parisi equations.

The agreement between measured observables ($O_i^{(\mathrm{exp})}$) and corresponding theoretical predictions ($O_i^{(\mathrm{th})}$) is quantified by a figure of merit, usually chosen as the χ^2 function,

$$\chi^2 = \sum_{i,j}^{N_{\mathrm{dat}}} (O_i^{(\mathrm{exp})} - O_i^{(\mathrm{th})})[\mathrm{cov}_{ij}](O_j^{(\mathrm{exp})} - O_j^{(\mathrm{th})}), \tag{3.1}$$

where cov_{ij} is the experimental covariance matrix. If data are provided with no correlated systematics, as the case in most polarized measurements, this is the diagonal matrix of the

experimental uncertainties. The best fit is obtained by minimizing the figure of merit, Eq. (3.1), and the corresponding best-fit parameters will finally fix the PDF shape.

Despite the apparent simplicity of this strategy, the determination of parton distributions from a global set of experimental data, possibly coming from different processes, is a challenging exercise. This requires to face several issues, some of which are peculiar to the polarized case. We summarize them as follows.

Lack of experimental data. Each observable has its own definition in terms of parton distributions and for this reason a specific observable can constrain or disentangle some distributions and not all of them. In principle, several different observables are needed to determine all the $2n_f + 1$ independent parton components (quark, antiquark and gluon) inside the nucleon. If such an information is lacking, either a subset of parton distributions is determined, or general assumptions on the unconstrained PDFs have to be made.

For instance, the bulk of experimental data to constrain polarized PDFs consists of polarized inclusive DIS data. However, this process only allows for a determination of the total quark distributions $\Delta u^+ = \Delta u + \Delta\bar{u}$, $\Delta d^+ = \Delta d + \Delta\bar{d}$, $\Delta s^+ = \Delta s + \Delta\bar{s}$, and of the gluon Δg (for details, see Chap. 2). Information on light sea antiquarks is provided either by semi-inclusive DIS with identified hadrons in the final state (mostly pions or kaons) or W production in proton-proton collisions. These reactions actually receive leading contributions from partonic subrocesses initiated by $\bar{u}$ and $\bar{d}$ antiquarks, hence they will be able to provide information on the corresponding distributions $\Delta\bar{u}$ and $\Delta\bar{d}$.

Besides, it is worth noticing that inclusive DIS indirectly constrains the polarized gluon, Δg, through scaling violations. Unfortunately, these turn out to have a mild effect for its determination, due to the small Q^2 lever arm of experimental data. Observables receiving leading contribution from gluon-initiated partonic subprocesses will be more suited to probe Δg directly. They include asymmetries for jet and pion production in proton-proton collisions and for one- or two-hadron and open-charm production in fixed-target lepton-nucleon scattering. The theory and phenomenology of these processes will be discussed in more detail in Chap. 6.

Finally, we notice that polarized data are less abundant and less accurate than their unpolarized counterparts and have a rather limited kinematic coverage in the (x, Q^2) plane. The various experimental processes that provide information on polarized PDFs, together with the corresponding leading partonic subprocesses, PDFs that are being probed and covered kinematic ranges are summarized in Tab. 3.1.

Theoretical issues. Many theoretical subtleties may affect the determination of parton distributions. In particular, we distinguish between issues related to the QCD analysis and issues related to the methodology adopted for the fitting of parton distributions. The latter especially include the choice of the functional form to parametrize PDFs and of the formalism to propagate uncertainties: they will be addressed separately in the next paragraphs.

Theoretical details in the QCD analysis concern for instance the treatment of heavy quark mass effects. So far, they are almost completely unaddress in the analysis of polarized PDFs. This is because such effects have been shown to be relatively small on the scale of present-day unpolarized PDF uncertainties [106], which are rather smaller than those of their polarized counterparts. Hence, the effects of heavy quark masses cannot clearly emerge when comparing to experimental data.

Another QCD theoretical issue is related to the treatment of higher-twist and nuclear corrections. As mentioned in Sec. 2.5, a large part of experimental data in polarized DIS are

REACTION	PARTONIC SUBPROCESS	PROBED PDF	x	Q^2 [GeV^2]
$\ell^\pm\{p,d,n\} \to \ell^\pm X$	$\gamma^* q \to q$	$\Delta q + \Delta\bar{q}$ Δg	$0.003 \lesssim x \lesssim 0.8$	$1 \lesssim Q^2 \lesssim 70$
$\overrightarrow{p}\overrightarrow{p} \to jet(s)X$	$gg \to qg$ $qg \to qg$	Δg	$0.05 \lesssim x \lesssim 0.2$	$30 \lesssim p_T^2 \lesssim 800$
$\overrightarrow{p}p \to W^\pm X$	$u_L\bar{d}_R \to W^+$ $d_L\bar{u}_R \to W^-$	$\Delta u\ \Delta\bar{u}$ $\Delta d\ \Delta\bar{d}$	$0.05 \lesssim x \lesssim 0.4$	$\sim M_W^2$
$\ell^\pm\{p,d\} \to \ell^\pm DX$	$\gamma^* g \to c\bar{c}$	Δg	$0.06 \lesssim x \lesssim 0.2$	$0.04 \lesssim p_T^2 \lesssim 4$
$\ell^\pm\{p,d\} \to \ell^\pm hX$	$\gamma^* q \to q$	$\Delta u\ \Delta\bar{u}$ $\Delta d\ \Delta\bar{d}$ Δg	$0.005 \lesssim x \lesssim 0.5$	$1 \lesssim Q^2 \lesssim 60$
$\overrightarrow{p}\overrightarrow{p} \to \pi X$	$gg \to qg$ $qg \to qg$	Δg	$0.05 \lesssim x \lesssim 0.4$	$1 \lesssim p_T^2 \lesssim 200$

Table 3.1: Summary of polarized processes to determine polarized PDFs. For each of them, we show the leading partonic subprocesses, probed polarized PDFs, and the ranges of x and Q^2 that become accessible. Processes are separated according to the need to use fragmentation functions for their analysis: processe in the upper part of the Table do not depend on fragmentation, while those in the lower part do.

taken at relatively low values of Q^2, typically a few GeV^2, and at medium-to-large x values. In this kinematic region, such corrections could play a relevant role, as demonstrated in Ref. [78].

Furthermore, assuming exact $SU(3)$ symmetry, the first moments of the nonsinglet quark combinations can be related to hyperon octet decay constants, see Eqs. (2.80). However, the violation of $SU(3)$ flavor symmetry is debated in the literature [160], even though a detailed phenomenological analysis seems to support it [161]. It should be clear that, due to the lack of experimental information, theoretical constraints, such these sum rules, positivity of measured cross-sections and integrability of parton distributions, provide the major input for determining the shape of PDFs in some kinematic regions, as we will discuss in Sec. 4.4.

Finally, we notice that the inclusion of processes involving identified hadrons in final states requires the usage of poorly known fragmentation functions. Recent work has emphasized the troubles for all available fragmentation function sets to describe the most updated inclusive charged-particle spectra data at the LHC [162]. Hence, the inclusion of semi-inclusive DIS and collider pion production data in a global determination of polarized PDFs is likely to introduce, via fragmentation functions, an uncertainty which is diffucult to estimate (though it is usually neglected).

Functional parametrization. The choice of the PDF parametrization is a crucial point. In principle, since PDFs represent our ignorance of the non-perturbative nucleon structure, there should be complete freedom in choosing their parametric form. However, in order to carry out a parton fit, one needs to choose a particular functional form for the PDFs at the initial scale, usually

$$x\Delta f_i(x,Q_0^2) = \eta_i A_i x^{a_i}(1-x)^{b_i}\left(1+\rho_i x^{\frac{1}{2}}+\gamma_i x\right). \tag{3.2}$$

Some of the parameters in Eq. (3.2) can be constrained by barion octet decay constants, by Regge interpretation at small-x, and by constraining to zero parton distributions at $x = 1$. Arbitrary assumptions are often made on the parameters to make the fit minimization to converge. Of course, as the number of free parameters decreases, the PDF parametrization turns out to be more rigid, thus introducing a bias in the final PDF determination.

Error estimates. The Hessian formalism [163, 164] is the most commonly used method for PDF error determination. The χ^2 function, Eq. (3.1), is quadratically expanded about its global minimum. It can be written as

$$\Delta\chi^2 = \chi^2 - \chi^2_0 = \sum_{i=1}^{N_{\rm par}} \sum_{j=1}^{N_{\rm par}} H_{ij}(a_i - a_i^0)(a_j - a_j^0), \tag{3.3}$$

with $\chi^2_0 = \chi^2(S_0)$ and $\{\mathbf{a}^0\}$ the χ^2 and the set of $N_{\rm par}$ parameters corresponding to the best estimate S_0 for the PDF set $\{\Delta f\}$ respectively; H_{ij} is the Hessian matrix element defined as

$$H_{ij} = \frac{\partial^2 \chi^2(\{\mathbf{a}\})}{\partial a_i \partial a_j}. \tag{3.4}$$

The Hessian matrix, Eq. (3.4), has a complete set of $N_{\rm par}$ orthonormal eigenvectors v_{ik} with eigenvalues ε_k defined by

$$\sum_{j=1}^{N_{\rm par}} H_{ij}\{\mathbf{a}^0\} v_{jk} = \varepsilon_k v_{ik}, \qquad \sum_{i=1}^{N_{\rm par}} v_{il} v_{jk} = \delta_{lk}. \tag{3.5}$$

Moving the parameters around their best value, a shift is observed in the χ^2 function, $\Delta\chi^2$. Each eigenvector determines a direction in the parameter space along which the χ^2 variation about the minimum can be expressed in a *natural* way. However, the fit quality typically deteriorates far more quickly upon variations in some directions than others, hence the eigenvalues ε_k are distributed over a wide range that covers many orders of magnitude. In terms of the diagonalized set of parameters defined with respect to the eigenvectors v_{ij}

$$z_i = \sqrt{\frac{\varepsilon_i}{2}} \sum_{j=1}^{N_{\rm par}} (a_j - a_j^0) v_{ij} \tag{3.6}$$

one obtains $\Delta\chi^2 = \sum_{i=1}^{N_{\rm par}} z_i^2$, *i.e.* the region of acceptable fits around the global minimum is contained inside a hypersphere of radius $\sqrt{\Delta\chi^2}$. The scheme for the diagonalization procedure is shown in Fig. 3.1.

The observable O, which depends on the set of parameters $\{\mathbf{a}\}$ via the PDF parametrization, is assumed to be rather well approximated in the neighbourhood of the global minimum by the first term of its Taylor-series expansion. The deviation of O from its best estimate O_0 is then given by $\Delta O = O - O_0 \approx \sum_{i=1}^{N_{\rm par}} O_i z_i$, with $O_i \equiv \partial O/\partial z_i|_{\mathbf{z}=\mathbf{0}}$. For a given variation $\Delta\chi^2$ of the χ^2, the error estimate on the observable O is then evaluated as

$$\Delta O = \left\{ \sum_{i=1}^{N_{\rm par}} \left[O(S_i^+) - O(S_i^-) \right]^2 \right\}^{\frac{1}{2}}, \tag{3.7}$$

where $S_i^{\pm}$ are $2N_{\rm rep}$ sets of PDFs computed at the two points defined by

$$z_i^{\pm} = \pm \frac{\sqrt{\Delta\chi^2}}{2} \tag{3.8}$$

on the edge of the $N_{\rm rep}$-dimensional hypersphere in the z parameter space. Together with S_0 they form $2N_{\rm sets} + 1$ sets of PDFs, that are the ones needed to compute PDFs errors on S_0, besides observable errors from Eq. (3.7).

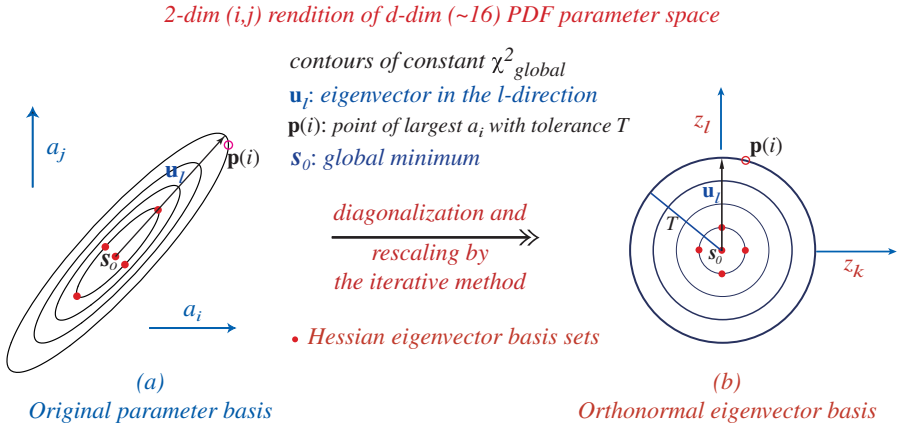

Figure 3.1: A schematic representation of the transformation from the PDF parameter basis to the orthonormal eigenvector basis as defined by Eq. (3.6). The figure is taken from Ref. [1].

It is worth noticing that the propagation of PDF uncertainties in the Hessian method has been derived under the assumption that a first order, linear approximation is adequate. Of course, due to the complicate nature of a global fit, deviations, also from the simple quadratic behavior, Eq. (3.3), are inevitable. Textbook statistics implies that, if the measurements belong to experimental data sets which are compatible with each others, and linear error propagation holds, one should have $\Delta\chi^2 = 1$. However, it has been argued that somewhat larger *tolerance* value $T = \Delta\chi^2$ [165] should be adopted in order for the distribution of χ^2 values between different experiments in a global fit to be reasonable. The reasonable value of the tolerance T can be determined, for instance, by estimating the range of overall χ^2 along each of the eigenvector directions within which a fit to all data sets can be obtained and then averaging the ranges over the $N_{\rm par}$ eigenvector directions. This method has been investigated in detail in Ref. [166] by demanding that indeed 90% of experiments approximately fall within the 90% confidence level. More refined methods involve the determination of a different tolerance [167] along each Hessian eigenvector (the so-called *dynamical* tolerance). Nevertheless, we should notice that the use of $T > 1$ is somewhat controversial, given that there is no rigorous statistical proof for the criterions adopted to estimate it. For all these reasons, error estimates based on the Hessian method are not necessarily always accurate.

A way to estimate PDF uncertainties avoiding quadratic approximation of χ^2 is provided by the Lagrange multiplier method [168, 169]. This is implemented by minimizing a function

$$\Psi(\{a_i\},\{\lambda_i\}) = \chi^2(\{a_i\}) + \sum_j \lambda_j O_j(\{a_i\}) \tag{3.9}$$

with respect to the set of PDF parameters $\{a_i\}$ for fixed values of the Lagrange multipliers $\{\lambda_i\}$. Each multiplier is related to one specific observable O_j, and the choice $\lambda_j = 0$ corresponds to the best fit S_0. By repeating this minimization procedure for many values of λ_j, one can map out precisely how the fit to data deteriorates as the expectation for the observable O_j is forced to change. Unlike the Hessian method, the Lagrange multiplier technique does not rely on any assumption regarding the dependence of the χ^2 on the fit parameters $\{a_i\}$.

3.2 The NNPDF approach to parton fitting

In recent years, several polarized PDF sets with uncertainties have been released. They slightly differ in the choice of data sets, the form of PDF parametrization, and in several details of the QCD analysis, like the treatment of higher-twist corrections, as will be reviewed with some more detail in Sec. 3.3 below. However, they are all based on the *standard* methodology for PDF fitting, based on fixed functional parametrization of PDFs and Hessian error estimate. This methodology is known [81] to run into difficulties especially when information is scarce, because of the intrinsic bias of fixed parton parametrization. This is likely to be particularly the case for polarized PDFs, which rely on data both less abundant and less accurate than their unpolarized counterparts.

In order to overcome these difficulties, the NNPDF collaboration has proposed and developed a new methodology for PDF determination [99–111]. So far, the NNPDF methodology has been successfully adopted to determine unpolarized PDF sets with increasing accuracy, which are now routinely used by the LHC collaborations in their data analysis and for data-theory comparisons. The method is based on a Monte Carlo approach, with neural networks used as unbiased interpolants. Monte Carlo sampling allows one to evaluate all quantities, such as the uncertainty or the correlation of PDFs, in a statistically sound way, while the use of neural networks provides a robust and flexible parametrization of the parton distributions at the initial scale. In the following, we will describe in detail the main features of the NNPDF methodology.

3.2.1 Monte Carlo sampling of the probability density distribution

Given a PDF or an observable depending on (polarized) PDFs, $O(\{\Delta f_i\})$, its average is given by the integration - in the functional space $V(\{\Delta f_i\})$ spanned by the parton distributions and weighted by a suitably defined probability measure - of all possible functions describing PDFs at a reference scale:

$$\langle O[\Delta f]\rangle = \int_V \mathscr{D}\Delta f\, \mathscr{P}[\Delta f] O[\Delta f]\,. \tag{3.10}$$

In the NNPDF approach, the probability measure is represented by a Monte Carlo sample in the space of PDFs. An ensemble of replicas of the original data set is generated, such that it reproduces the statistical distribution of the experimental data, followed by its projection into the space of PDFs through the fitting procedure. Notice that all theoretical assumptions represent a prior for the determination of such probability measure. The ensemble in the space of data has to contain all the available experimental information. In practice, most data are given with multi-Gaussian probability distributions of statistical and systematic errors, described by a covariance matrix. In such cases this is the distribution that will be used to generate the pseudodata. However, any other probability distribution can be used if and when required by the experimental data. Each replica of the experimental data is a member of the Monte Carlo ensemble and contains as many data points as are originally available. Whether the given ensemble has the desired statistical features can be verified by means of statistical standard tests by comparing quantities calculated from it with the original properties of the data. Such tests, together with pseudodata generation, will be explicitely discussed in Sec. 4.1 for the case of polarized DIS.

Here we should notice that the sampling of the underlying probability density distribution of data allows one to circumvent the non-trivial issues related to Hessian propagation of errors. Indeed, an ensemble of parton distributions is fitted to pseudodata: this means that,

at the end of the fitting procedure, one obtains as many PDFs as the number of replicas $N_{\rm rep}$ of the data that were generated. The experimental values in each replica will fluctuate according to their distribution in the Monte Carlo ensemble and the best fit PDFs will fluctuate accordingly for each replica. Even though individual PDF replicas might fluctuate significantly, averaged quantities like central value and 1σ error bands are smooth inasmuch as the size of the ensemble increases. The advantages of the Monte Carlo methodology are then apparent. First, the expectation value for any observable depending on the PDFs or the PDFs themselves can be easily computed as the Monte Carlo average over the PDF ensemble: Eq. (3.10) is then replaced by

$$\langle O[\Delta f]\rangle = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}} O[\Delta f_k]\,, \tag{3.11}$$

and similarly uncertainties can be obtained as standard deviations, and so forth. Second, non-Gaussian behavior of uncertainties can be tested either at the level of experimental data, by sampling them according to an arbitrary distribution, or at the level of the results in the best fit ensemble of PDFs, by defining proper confidence levels. In any case, we do not have to rely on the quadratic assumption, Eq. (3.3), made in the Hessian approach. Finally, the stability of results upon a change of parametrisation can be verified by standard statistical tools, for instance by computing the distance between results in units of their standard deviation. Likewise, it is possible to verify that fits performed by removing data from the set have wider error bands but remain compatible within these enlarged uncertainties, or to address how results change within different theoretical assumptions. The reliability of the results can thus be assessed directly.

3.2.2 Neural network parametrization

The Monte Carlo technique adopted to propagate the experimental error into the space of PDFs is completely independent of the method used to parametrize parton distributions; it might well be used along with standard parametrization, Eq. (3.2) [170]. On the other hand, in order to get a faithful determination of parton distributions, one ought to make sure that the chosen functional form is redundant enough not to introduce a theoretical bias which would artificially reduce parton uncertainty in regions where data do not constrain enough PDFs. There are several ways for obtaining such a redundant parametrisation. One may use some clever polynomial basis, or more refined tools such as self-organising maps [171].

Within the NNPDF methodology, each parton distribution is parametrized by a neural network, which provide a redundant and minimally biased parametrization. The only theoretical assumption is smoothness, a presumed feature of PDFs, which is ensured by the flexibility and adaptability of neural networks. In particular, we use feed-forward neural networks [120]. They are made of a set of interconnected units, called neurons, eventually organized in groups, called layers. The state or activation of a given neuron i in a given layer l, $\xi_i^{(l)}$, is a real number, determined as a function of the activation of the neurons connected to it, namely those in the previous, $l-1$, layer. Each pair of neurons (i,j) is then connected by a synapsis, characterized by a real number $\omega_{ij}^{(l-1)}$, called weight. The activation of each neuron i in a given layer l is a function g of the difference between a weighted average of input from neurons in the preceding

layer and a threshold $\theta_i^{(l)}$:

$$\xi_i^{(l)} = g\left(\sum_{j=1}^{N_{l-1}} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)}\right), \tag{3.12}$$

where N_{l-1} is the number of neurons in the $(l-1)^{\text{th}}$ layer. The input and output vectors are labeled as $\xi^{(1)}$ and $\xi^{(L)}$ respectively, with L the number of layers in the network.

The activation function g is in general non-linear. The simplest example of activation function $g(x)$ is the step function $g(x) = \Theta(x)$, which produces binary activation only. However, it turns out to be advantageous to use an activation function with two distinct regimes, linear and non-linear, such as the sigmoid

$$g(x) = \frac{1}{1-e^{-\beta x}}. \tag{3.13}$$

This function approaches the step function at large β; without loss of generality, in the NNPDF methodology we usually take $\beta = 1$. The sigmoid activation function has a linear response when $x \approx 0$, and it saturates for large positive or negative arguments. If weights and thresholds are such that the sigmoids work on the crossover between linear and saturation regimes, the neural network behaves in a non-linear way. Thanks to this non-linear behavior, the neural network is able to reproduce nontrivial functions.

Basically, multilayer feed-forward neural networks provide a non-linear map between some input $\xi_i^{(1)}$ and output $\xi_j^{(L)}$ variables, parametrized by weights, thresholds and activation function,

$$\xi^{(L)} = F\left[\xi^{(1)}; \{\omega_{ij}^{(l)}\}, \{\theta_i^{(l)}\}; g\right]. \tag{3.14}$$

For given activation function, the parameters can be tuned in such a way that the neural network reproduces any continuous function. The behavior of a neural network is determined by the joint behavior of all its connections and thresholds, and it can thus be built to be redundant, in the sense that modifying, adding or removing a neuron has little impact on the final output. Because of these reasons, neural networks can be considered to be robust, unbiased universal approximants.

In Sec. 4.3, we will explicitely discuss how polarized PDFs can be parametrized in terms of neural networks. In particular, we will discuss their architecture and preprocessing of data to enforce the asymptotic small- and large-x behavior of PDFs.

3.2.3 Minimization and stopping

Once each independent PDF is parametrized in terms of neural networks at an inital reference scale Q_0^2, physical observables are computed by convolving hard kernels with PDFs evolved to the scale of the experimental measurements by Altarelli-Parisi evolution. The best fit set of parton distribution is determined by comparing the theoretical computation of the observable for a given PDF set with their replica experimental values by evaluating a suitable figure of merit, *e.g.* Eq. (3.1).

Both the minimization and the determination of the best-fit in the wide, non-local, space of parameters spanned by the neural network parameters are delicate issues. To minimize the error function, a genetic algorithm is used. The main advantage of such an algorithm is that it works on a population of solutions, rather than tracing the progress of one point through

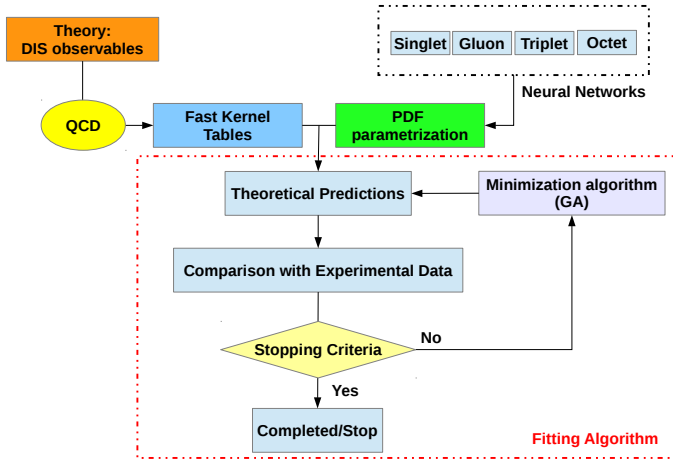

Figure 3.2: Scheme of the NNPDF methodology for parton fitting.

parameter space. Thus, many regions of parameter space are explored simultaneously, thereby lowering the possibility of getting trapped in local minima.

The basic idea underlying the genetic algorithm is the following. Starting from a randomly chosen set of parameters, a pool of possible new sets is generated by mutation of one or more parameters at a time. Each new set that has undergone mutation is a mutant. To each set of parameters corresponds a value of the figure of merit that is being minimized, and so those configurations that fall far away from the minimum can be discarded. This procedure is iterated over a sufficiently large number of generations. Of course, one has to be sure that the final set of parameters corresponds to acceptable PDFs, *i.e.* they must satisfy theoretical constraints like positivity of cross-sections and sum rules. Usually, this requirement is fulfilled by penalizing unacceptable replicas during the minimization by arbitrarily increasing their figure of merit. In Sec. 4.3, we will discuss technical details related to how this issue is faced in the determination of polarized distributions.

The redundancy of the parametrization also implies another subtle problem, known as *overlearning*, which happens when neural networks start to fit statistical fluctuations of data, rather than their underlying physical law. The solution to this problem is achieved using a cross-validation method to determine a criterion for the fit to stop before entering the overlearning regime. Technical details about the specific implemetation of stopping in the polarized case will be extensively discussed in Sec. 4.3.

The main ingredients of the NNPDF methodology described above, namely Monte Carlo sampling of data distribution, neural network parametrization of PDFs, minimization and stopping, are finally sketched in Fig. 3.2.

3.2.4 Bayesian reweighting of Monte Carlo PDF ensembles

Monte Carlo sampling of the underlying probability density distribution of data used in the NNPDF methodology allows for applying standard statistical tools to the resulting PDF

ensembles. Most importantly, Bayesian inference can be exploited to determine the impact on PDFs of new data sets that were not included in the fit. The methodology, referred to as reweighting, was presented in detail in Refs. [105, 109]. In short, the main idea underlying reweighting is to assign to each replica in the PDF ensemble a weight which assesses the probability that this replica agrees with new data. The expectation value for an observable O taking into account the new data is then given by the weighted average

$$\langle O[\Delta f]\rangle_{\rm new} = \int_V \mathscr{D}\Delta f\, \mathscr{P}_{\rm new}[\Delta f] O[\Delta f] = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} w_k O[\Delta f_k]\,. \tag{3.15}$$

The weights w_k are computed from the χ^2 of the new data to the prediction obtained using a given replica, according to the formula

$$w_k = \frac{(\chi^2_k)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2_k}}{\frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}} (\chi^2_k)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2_k}}\,, \tag{3.16}$$

where n is the number of new data. The formula, Eq. (3.16), is derived under the assumption that new data have Gaussian errors and that they are statistically independent of the old data. By the law of multiplication of probabilities it then follows that

$$\mathscr{P}_{\rm new} = \mathscr{N}_\chi \mathscr{P}(\chi|\Delta f)\mathscr{P}_{\rm old}(\Delta f)\,, \tag{3.17}$$

where $\mathscr{N}_\chi$ is a normalization factor chosen such that $\sum_{k=1}^{N_{\rm rep}} w_k = N_{\rm rep}$ and

$$w_k = \mathscr{N}_\chi \mathscr{P}(\chi|\Delta f_k) \propto (\chi^2_k)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2_k}\,. \tag{3.18}$$

Notice that, after reweighting a given PDF ensemble of $N_{\rm rep}$ replicas, the efficiency in describing the distribution of PDFs is no longer the same. In fact, the weights give the relative importance of the different replicas, and the replicas with very small weights will become almost irrelevant in ensemble averages. The reweighted replicas will thus no longer be as efficient as the old: for a given $N_{\rm rep}$, the accuracy of the representation of the underlying distribution $\mathscr{P}_{\rm new}(\Delta f)$ will be less than it would be in a new fit. This loss of efficiency can be quantified using the Shannon entropy to determine the effective number of replicas left after reweighting:

$$N_{\rm eff} = exp\left\{\frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}} w_k \ln\left(\frac{N_{\rm rep}}{w_k}\right)\right\}. \tag{3.19}$$

Clearly $0 < N_{\rm eff} < N_{\rm rep}$: the reweighted fit has the same accuracy as a refit with $N_{\rm eff}$ replicas. Hence, if $N_{\rm eff}$ becomes too low, the reweighting procedure will no longer be reliable, either because the new data contain a lot of information on the PDFs, neccessitating a full refitting, or because the new data are inconsistent with the old.

Bayesian reweighting allows for the inclusion of new pieces of experimental information in an ensemble of PDFs without performing a new fit. This is desirable, particularly when dealing with observables for the computation of which no fast code is available. Also notice that, from a conceptual point of view, a determination of a PDF set might be performed by including all data through reweighting of a first reasonable guess, called *prior*, as suggested in Refs. [172, 173]. Typically, this is the result of a previous PDF fit to data (different from

those with which the parton set will be reweighted) or a model based on some theoretical assumptions. The reweighting method works fine provided the prior is unbiased, so that one can check that the final results do not depend on the choice of the initial guess. Chap. 6 will be mostly devoted to the discussion of this aspect in a particular case of physical interest: we will reweight a DIS-based parton set with jet and W production data from polarized proton-proton collisions.

Results obtained via reweighting of existing Monte Carlo PDF sets were shown to be statistically equivalent to those obtained via global refitting: in particular, the method was validated in the unpolarized case by considering inclusive jet and LHC W lepton asymmetry data [105, 109]. As a further illustration of possible applications of the method, reweighting was used to quantify the impact on unpolarized PDFs of direct photon data [174] and of top quark production data [175], and to study the constraints on nuclear PDFs from LHC pPb data [176]. Also, notice that the same technique can be straightforwardly applied in the case of Hessian PDF sets [177].

Once a reweighted PDF set has been determined, it would be interesting to be able to produce a new PDF ensemble with the same probability distribution as a reweighted set, but without the need to include the weight information. A method of unweighting has therefore been developed, whereby the new set is constructed by deterministically sampling with replacement the weighted probability distribution [109]. This means that replicas with a very small weight will no longer appear in the final unweighted set while replicas with large weight will occur repeatedly.

If the probability for each replica and the probability cumulants are defined as

$$p_k = \frac{w_k}{N_{\rm rep}} \qquad P_k \equiv P_{k-1} + p_k = \sum_{j=0}^{k} p_j \,, \tag{3.20}$$

it is possible to quantitatively describe the unweighting procedure. Starting with $N_{\rm rep}$ replicas with weights w_k, $N_{\rm rep}$ new weights w'_k are determined:

$$w'_k = \sum_{j=1}^{N'_{\rm rep}} \theta\left(\frac{j}{N'_{\rm rep}} - P_{k-1}\right) \theta\left(P_k - \frac{j}{N'_{\rm rep}}\right) . \tag{3.21}$$

These weights are therefore either zero or a positive integer. By construction they satisfy

$$N'_{\rm rep} \equiv \sum_{k=1}^{N_{\rm rep}} w'_k \,, \tag{3.22}$$

i.e. the new unweighted set consists of $N_{\rm rep}$ replicas, simply constructed by taking w_k copies of the k-th replica, for all $k = 1, \ldots, N_{\rm rep}$.

In Chap. 6, Bayesian reweighting of polarized parton distributions, determined within the NNPDF methodology from inclusive DIS data, will be used to assess the impact of open-charm production data in fixed-target DIS as well as W and jet production data in proton-proton collisions. Finally a new polarized parton set will be determined via unweighting.

3.3 Overview of available polarized parton sets

First studies of the polarized structure of the nucleon were aimed at an accurate determination of polarized first moments, including detailed uncertainty estimates [66, 67, 143], but did not

attempt a determination of a full PDF set. This was first proposed in Ref. [65], but without uncertainty estimation. Several polarized parton sets have been delivered in recent years, all at NLO accuracy, usually updated with the inclusion of new data, theoretical or statistical features. In this Section, we review the main features of the presently available polarized PDF sets; the discussion about recent developments in the determination of a polarized parton set based on the NNPDF methodology will be addressed in Chaps. 4-5-6.

3.3.1 DIS-based fits

The bulk of experimental information on longitudinally polarized nucleon structure consists of inclusive, photon induced, deep-inelastic scattering with both polarized charged lepton beams and nucleon targets. Actually, only data sets coming from this process are included in most of the polarized PDF determinations, as follows.

BB10. The `BB10` parton set [76] includes world-available data on the asymmetry A_1 or the structure function g_1 from polarized inclusive DIS. Four independent polarized PDFs are determined there, namely the valence combinations $\Delta u^- = \Delta u - \Delta\bar{u}$ and $\Delta d^- = \Delta d - \Delta\bar{d}$, the gluon, Δg and the quark sea, $\Delta\bar{q}$, assuming symmetric sea, $\Delta\bar{u} = \Delta\bar{d} = \Delta\bar{s} = \Delta s$. Each PDF is parametrized in terms of a fixed functional form like that of Eq. (3.2), but only a subset of them are actually taken to be free, depending on the PDF. Also the QCD scale $\Lambda_{\rm QCD}$ is a parameter to be determined in the fit. Errors are determined within the standard Hessian approach without any tolerance criterion. The analysis is supplemented by the inclusion of experimental systematic uncertainties from different sources, namely data, NMC parametrization of F_2 [178] and R parametrization [179] both entering Eq. (2.84). Also theoretical systematic uncertainties were estimated from variation of factorization and renormalization scale and input parametrization scale. Finally, higher twist contributions from heavy flavor Wilson coefficients in the fixed-flavor number scheme were included in the theoretical QCD analysis.

The `BB10` parton set is publicly available together with a `FORTRAN` program which allows for computing PDF central values and errors.

JAM13. The `JAM13` parton set [78], instead of the asymmetry A_1 or the structure function g_1, directly fits the measured longitudinal and transverse asymmetries $A_\parallel$ and $A_\perp$, Eqs. (2.83). This makes a differencee with most of other analyses, which usually include the information from the g_1 structure function extracted from observed asymmetries by each experimental collaboration within different assumptions. Besides, a large number of preliminary inclusive DIS data from JLAB are included in the `JAM13` study. Since these data lie in the large-x and small-Q^2 kinematic region, they are expected to be particularly sensitive to target mass and higher-twist effects. For this reason, both twist-3 and twist-4 corrections to g_1, as well as a twist-3 correction to the g_2 structure function are taken into account. Moreover, they consistently apply the nuclear smearing corrections to both the g_1 and g_2 structure functions for both deuterium and ^{3}He, within the framework of the weak binding approximation [180–182].

Six independent polarized PDFs are determined, namely the total quark combinations, $\Delta u^+ = \Delta u + \Delta\bar{u}$ and $\Delta d^+ = \Delta d + \Delta\bar{d}$, the gluon, Δg, and the quark sea, $\Delta\bar{q}$ ($q = u, d, s$) and symmetric strangeness $\Delta s = \Delta\bar{s}$ is assumed. These are parametrized with the functional form Eq. (3.2), however, since the experimental piece of information does not allow for a complete determination of all the six PDFs above, several additional contraints are adopted. In particular, the $\Delta\bar{u}$ and $\Delta\bar{d}$ parton distributions, which do not contribute directly to the description of

inclusive DIS data included in the analysis, are fixed by requiring

$$\lim_{x\to 0} \Delta\bar{q}(x,Q_0^2) = \frac{1}{2}\lim_{x\to 0}\Delta q^+(x,Q_0^2)\,, \tag{3.23}$$

with $q = u,d$. In addition, to avoid unphysical results and provide *reasonable* values for all distributions, the following constraint is imposed

$$\frac{1}{2}\left(\left|\frac{\Delta\bar{q}^{(2)}}{\Delta\bar{s}^{(2)}}\right| + \left|\frac{\Delta\bar{s}^{(2)}}{\Delta\bar{q}^{(2)}}\right|\right) = 1 \pm 0.25\,. \tag{3.24}$$

Both Eq. (3.23) and Eq. (3.24 entail the decrease in the number of free parameters whose values are actually fitted to data. The choice of limiting the flexibility of the parametrization given by Eq. (3.2) is also made for the gluon, for which the values of the exponents a_g and b_g are somewhat arbitrarily fixed. The propagation of uncertainties is performed within the Hessian approach without tolerance criterion.

ABFR98. The `ABFR98` parton set [67] is based on a less updated set of inclusive DIS experimental data and provides four polarized PDFs, both total combinations Δq^+ and the gluon Δg, within fixed functional parametrization and Hessian error estimate. Despite the `ABFR98` is less recent than other analyses discussed here, it should be worth mentioning it at least for two reasons. First, this provides polarized PDFs that are fitted in the AB renormalization scheme (see Sec. 2.2.2) instead of $\overline{\text{MS}}$ used in all other PDF determinations. Second, it includes a detailed discussion of theoretical uncertainties originated by neglected higher orders, higher twists, position of heavy quark thresholds, value of the strong coupling, violation of SU(3) flavor symetry and finally uncertainties related to the choice of functional form.

3.3.2 Global PDF fits

In recent years, the knowledge on longitudinally polarized nucleon structure has been supplemented by data coming from processes different from polarized inclusive DIS. As outlined in Sec. 3.1, these include fixed-target SIDIS and hadron or jet production in polarized proton-proton collisions. Including this piece of experimental information in a global NLO analysis is a challenging task, both theoretical and computational, because of the need to deal with hadronic observables. They may also depend on the fragmentation of quarks in the final measured hadrons and hence their analysis require the usage of poorly known fragmentation functions. Several such global parton sets have been determined in very recent years, as summarized as follows.

AAC08. Besides polarized inclusive DIS data, the `AAC08` analysis [73] also includes π^0 production data at RHIC, via a K-factor approximation for the NLO corrections. Because these data only provide constraints on the gluon polarization, only four idependent PDFs are determined: they are the same as in the `BB10` analysis (symmetric sea is also assumed), but the functional form they use reads

$$\Delta f(x,Q_0^2) = [\delta x^\nu - \kappa(x^\nu - x^\mu)]\, f(x,Q_0^2)\,, \tag{3.25}$$

where δ, μ, ν, κ are free parameters to be determined in the fit and $f(x,Q_0^2)$ is the corresponding unpolarized PDF which was taken from Ref. [183]. For the description of the fragmentation

Fit	Ref.	Data set(s)	Scheme	Parton Distributions	Uncertainties
BB10	[76]	DIS	$\overline{\text{MS}}$	Δu^-, Δd^-, $\Delta\bar{q}$, Δg	Hessian $\Delta\chi^2 = 1$
JAM13	[78]	DIS	$\overline{\text{MS}}$	Δu^+, Δd^+, $\Delta\bar{u}$, $\Delta\bar{d}$, $\Delta\bar{s}$, Δg	Hessian $\Delta\chi^2 = 1$
ABFR98	[67]	DIS	AB	Δu^+, Δd^+, Δs^+, Δg	Hessian $\Delta\chi^2 = 1$
AAC08	[73]	DIS, π^0	$\overline{\text{MS}}$	Δu^+, Δd^+, Δs^+, Δg	Hessian $\Delta\chi^2 = 12.95$
LSS10	[73]	DIS, SIDIS	$\overline{\text{MS}}$	Δu^+, Δd^+, $\Delta\bar{u}$, $\Delta\bar{d}$, $\Delta\bar{s}$, Δg	Hessian $\Delta\chi^2 = 1$
DSSV08	[75]	DIS, SIDIS, π^0, Jets	$\overline{\text{MS}}$	Δu^+, Δd^+, $\Delta\bar{u}$, $\Delta\bar{d}$, $\Delta\bar{s}$, Δg	Hessian $\Delta\chi^2 = 1$ Lagr. mult. $\Delta\chi^2/\chi^2 = 2\%$

Table 3.2: Main features of the available polarized PDF fits.

into a pion, the HKNS07 set [184] is used. Error estimates are handled via standard Hessian approach, but a tolerance $\Delta\chi^2 = 12.95$ is assumed in order for the distribution of χ^2 values between different experiments in the global fit to be reasonable.

The AAC08 parton set is publicly available together with a FORTRAN program which allows for computing PDF central values and uncertainties.

LSS10. The LSS10 parton set [77] is based on all world-available data from incluisve and semi-inclusive DIS. These allow for a determination of light antiquarks. Six polarized PDFs are parametrized according to Eq. (3.2), namely the total PDF combinations Δu^+ and Δd^-, the antiquarks $\Delta\bar{u}$, $\Delta\bar{d}$ and $\Delta\bar{s}$ and the gluon Δg (for the four latter $\rho_i = 0$ in Eq. (3.2)). The assumption $\Delta s = \Delta\bar{s}$ is also made. Hessian error propagation is performed assuming $\Delta\chi^2 = 1$. Fragmentations functions are taken from the DSS07 analysis [185,186]. Similarly to the JAM13 fit, the theoretical QCD analysis takes into account the $1/Q^2$ terms, arising from kinematic target mass corrections and dynamic higher twist corrections, in the expression of the nucleon spin structure function g_1.

The LSS10 parton set is publicly available together with a FORTRAN program which allows for computing PDF central values.

DSSV. Different PDF sets belong to the DSSV family, due to the remarkable effort put by this collaboration in updating their polarized parton sets with new available data. The first global analysis to include polarized collider measurements at RHIC, besides inclusive and semi-inclusive polarized DIS data, was DSSV08 [74,75]. Semi-inclusive pion and single-inclusive jet production data were considered in their study. This was recently updated by the DSSV+ and DSSV++ fits [187,188]; in particular, the latter includes the most recent jet production data at RHIC, which were found to constrain the gluon shape in the mid-to-large x region with unprecedented accuracy.

As in the LSS10 analysis, six independent polarized PDFs are determined from data, and the DSS07 fragmentation functions are used. The uncertainty estimates are provided through the Lagrange multiplier method described in Sec. 3.1 with the conservative assumption $\Delta\chi^2/\chi^2 = 2\%$, even though the standard Hessian approach, with $\Delta\chi^2 = 1$, is also used for comparison. The DSSV08 set is publicly available as 38 hessian eigenvector sets (a set for each minimized parameter for each direction of variation) plus a central value set.

In Tab. 3.2, we summarize the features of the available polarized parton sets; a detailed comparison of polarized PDF between them and the NNPDF determination will be discussed in Chap. 4.

4 Unbiased polarized PDFs from inclusive DIS

In this Chapter, we present the first determination of polarized parton distributions based on the NNPDF methodology, `NNPDFpol1.0`. This analysis includes all available data from inclusive, neutral-current, polarized DIS and aims at an unbiased extraction of total quark-antiquark and gluon distributions at NLO accuracy. In Sec. 4.1 we present the data sets used in the present analysis, and we discuss how their statistical distribution is sampled with Monte Carlo generation of pseudodata. We provide details of the QCD analysis in Sec. 4.2, then we discuss the PDF parametrization in terms of neural networks in Sec. 4.3; we give particular emphasis on minimization strategy and the peculiarities of the polarized case. The `NNPDFpol1.0` parton set is presented in Sec. 4.4, where we illustrate its statistical features, and its stability upon the variation of several theoretical and methodological assumptions. We also compare our results to other recent polarized PDF sets reviewed in Sec. 3.3. Finally, we discuss phenomenological implications for the spin content of the proton and the test of the Bjorken sum rule in Sec. 4.5. The analysis presented in this Chapter mostly reproduces Ref. [116].

4.1 Experimental input

We present the features of the experimental data sets included in the `NNPDFpol1.0` analysis and we discuss in detail which piece of information they provide on the polarized structure functions. Then, we summarize the construction and the validation of the Monte Carlo pseudodata sample from the input experimental data.

4.1.1 The data set: observables, kinematic cuts, uncertainties and correlations

We consider inclusive, neutral current, lepton-nucleon DIS data coming from-all-over-the-world experiments performed at CERN [50, 189–192], SLAC [152, 193, 194] and DESY [195, 196]. These experiments use different lepton beams, electrons or muons, and nucleon targets, protons and neutrons (deuteron or ^{3}He). The main features of the data sets included in our analysis are summarized in Tab. 4.1, where we show, for each of them, the number of available data

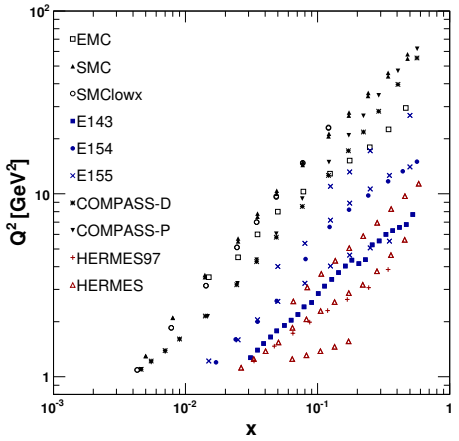

Figure 4.1: Experimental data in the (x, Q^2) plane (after kinematic cuts): black points are from CERN experiments [50, 189–192], blue from SLAC [152, 193, 194] and red from DESY [195, 196].

points, the covered kinematic range, and the published observable we use to reconstruct the g_1 structure function. Their kinematic coverage in the (x, Q^2) plane is also shown in Fig. 4.1. This quantity differs experiment by experiment, since the primary observable can be one of the asymmetries or structure functions discussed in Sec. 2.4. In the following, we summarize how we reconstruct the g_1 structure function from the published experimental observables for individual experiments (labelled as in Tab. 4.1).

EMC, SMC, SMClowx, COMPASS, HERMES97. All these experiments have performed a measurement of $A_\parallel$, then they determined A_1 using Eq. (2.93), under the assumption $\eta \approx 0$. Therefore, the observable published by these experiments actually corresponds to a measurement of $\frac{A_\parallel}{D}$. We determine g_1 from $\frac{A_\parallel}{D}$ using Eq. (2.97). This is possible because D is completely fixed by Eq. (2.87) in terms of the unpolarized structure function ratio Eq. (2.92) and of the kinematics. We determine the unpolarized structure function ratio using as primary inputs F_2, for which we use the parametrization of Ref. [99, 120], and F_L, which we determine from its expression in terms of parton distributions, using the NLO `NNPDF2.1` parton set [107].

HERMES. This experiment has performed a measurement of $A_\parallel$, and it published both $A_\parallel$ and A_1 (which is determined using Eq. (2.93) and a parametrization of A_2). We use the published values of $A_\parallel$, which are closer to the experimentally measured quantity, to determine g_1 through Eq. (2.97).

E143. This experiment has taken data with three different beam energies, $E_1 = 29.1$ GeV, $E_2 = 16.2$ GeV, $E_3 = 9.7$ GeV. For the highest energy both $A_\parallel$ and $A_\perp$ are independently measured and A_1 is extracted from them using Eq. (2.93); for the two lowest energies only $A_\parallel$ is measured and A_1 is extracted from it using Eqs. (2.95-2.96), while assuming the form Eq. (2.102) for g_2. The values of A_1 obtained with the three beam energies are combined into a

Experiment	Set	$N_{\rm dat}$	$x_{\rm min}$	$x_{\rm max}$	$Q^2_{\rm min}$ [GeV^2]	$Q^2_{\rm max}$ [GeV^2]	F	Ref.
EMC								
	EMC-A1P	10	.0150	.4660	3.5	29.5	$A_\parallel^p/D$	[50]
SMC								
	SMC-A1P	12	.0050	.4800	1.3	58.0	$A_\parallel^p/D$	[189]
	SMC-A1D	12	.0050	.4790	1.3	54.8	$A_\parallel^d/D$	[189]
SMClowx								
	SMClx-A1P	15 (8)	.0001 (.0043)	.1210	0.02 (1.09)	23.1	$A_\parallel^p/D$	[190]
	SMClx-A1D	15 (8)	.0001 (.0043)	.1210	0.02 (1.09)	22.9	$A_\parallel^d/D$	[190]
E143								
	E143-A1P	28 (25)	.0310	.5260	1.27	9.52 (7.72)	A_1^p	[152]
	E143-A1D	28 (25)	.0310	.5260	1.27	9.52 (7.72)	A_1^d	[152]
E154								
	E154-A1N	11	.0170	.5640	1.2	15.0	A_1^n	[193]
E155								
	E155-G1P	22 (20)	.0150	.7500 (.5000)	1.22	34.72 (26.86)	g_1^p/F_1^p	[194]
	E155-G1N	22 (20)	.0150	.7500 (.5000)	1.22	34.72 (26.86)	g_1^n/F_1^n	[194]
COMPASS-D								
	CMP07-A1D	15	.0046	.5660	1.10	55.3	$A_\parallel^d/D$	[191]
COMPASS-P								
	CMP10-A1P	15	.0046	.5680	1.10	62.1	$A_\parallel^p/D$	[192]
HERMES97								
	HER97-A1N	9 (8)	.0330	.4640 (.3420)	1.22	5.25 (3.86)	$A_\parallel^n/D$	[195]
HERMES								
	HER-A1P	38 (28)	.0264	.7311 (.5823)	1.12	14.29 (11.36)	$A_\parallel^p$	[196]
	HER-A1D	38 (28)	.0264	.7311 (.5823)	1.12	14.29 (11.36)	$A_\parallel^d$	[196]
Total		290 (245)						

Table 4.1: Experimental data sets included in the `NNPDFpol1.0` analysis. For each experiment we show the number of data points before and after (in parenthesis) applying kinematic cuts, the covered kinematic range and the measured observable.

single determination of A_1; radiative corrections are applied at this combination stage. Because of this, we should use this combined value of A_1, from which we then determine g_1 using Eq. (2.98). In order to determine y, Eq. (2.8), which depends on the beam energy, we use the mean of the three energies.

E154. This experiment measures $A_\parallel$ and $A_\perp$ independently, and then extracts a determination of A_1. We use these values of A_1 to determine g_1 by means of Eq. (2.98).

E155. This experiment only measures the longitudinal asymmetry $A_\parallel$, from which the ratio g_1/F_1 is extracted using Eq. (2.95) with the Wandzura-Wilczek form of g_2, Eq. (2.102). In this case, we use these values of g_1/F_1, and we extract g_1 using Eq. (2.91) for F_1, together with the parametrization of Ref. [99, 120] for F_2 and the expression in terms of parton distributions and the NLO `NNPDF2.1` parton set [107] for F_L.

All these experiments also provide an extraction of the g_1 structure function in their specific framework, based on different assumptions on g_2. For this reason, we preferred to use the experimental asymmetries, instead of the corresponding structure functions, which instead we reconstructed on our own. We checked that they are consistent with those provided by the experimental collaboration themselves within their uncertainties.

We have excluded from our analysis all data points with $Q^2 \leq Q^2_{\rm cut} = 1$ GeV^2, since below such energy scale perturbative QCD cannot be considered reliable. A similar choice of cut is usually made in all polarized analyses, specifically in Refs. [66, 67, 73, 75, 76, 143]. Notice that this value of $Q^2_{\rm cut}$ is somewhat lower than that adopted in unpolarized fits, where $Q^2_{\rm cut} = 2$ GeV^2. This difference arises not to exclude a large piece of experimental information which, in the polarized case, lies in the small-Q^2 region (see Fig. 4.1).

We further impose a cut on the squared invariant mass of the hadronic final state $W^2 = Q^2(1-x)/x$ in order to remove points which may be affected by sizable dynamical higher-twist corrections. The cut is chosen based on a study presented in Ref. [197], where higher twist terms were added to the observables, with a coefficient fitted to the data: it was shown there that the higher twist contributions become compatible with zero if one imposes the cut $W^2 \geq W^2_{\rm cut} = 6.25$ GeV^2. We will follow this choice, which excludes data points with large Bjorken-x at moderate values of the squared momentum transfer Q^2, roughly corresponding to the bottom-right corner of the (x, Q^2)-plane, see Fig. 4.1: in particular, it excludes all available JLAB data [198–200]. The number of data points surviving the kinematic cuts for each data set is given in parenthesis in Tab. 4.1.

As can be seen from Fig. 4.1, the region of the (x, Q^2)-plane where data are available after kinematic cuts is roughly restricted to $4 \cdot 10^{-3} \lesssim x \lesssim 0.6$ and 1 GeV$^2 \leq Q^2 \lesssim 60$ GeV^2. In recent years, the coverage of the low-x region has been improved by a complementary set of SMC data [190] and by the more recent COMPASS data [191, 192], both included in the present analysis. In the large-x region, information is provided at rather high Q^2 by the same COMPASS data and at lower energy by the latest HERMES measurements [196]. In the near future, additional polarized inclusive DIS measurements are expected from an update of COMPASS data [201] and from JLAB spin program [78]. However, the latter will cover the large-x and small Q^2 corner, hence their inclusion in a global fit will require a careful treatment of higher-twist corrections, as performed in Ref. [78]. The data set used in this paper is the same as that of Ref. [76], and also the same as the DIS data of the fit of Ref. [75], which however has a wider data set beyond inclusive DIS.

Each experimental collaboration provides uncertainties on the measured quantities listed in the next-to-last column of Tab. 4.1. Correlated systematics are only provided by EMC and E143, which give the values of the systematics due to the uncertainty in the beam and target polarizations, while all other experiments do not provide any piece of information on the covariance matrix. For each experiment, we determine the uncorrelated uncertainty on g_1 by combining the uncertainty on the experimental observable with that of the unpolarized structure function using standard error propagation. For EMC and E143 experiments, we also include all available correlated systematics. These are provided by both experimental collaborations as a percentage correction to g_1 (or, alternatively, to the asymmetry A_1): we apply the percentage uncertainty on g_1 to the structure function determined by us as discussed in Sect. 2.4.

We then construct the covariance matrix

$$\mathrm{cov}_{pq} = \left(\sum_i \sigma_{i,p}^{(c)} \sigma_{i,q}^{(c)} + \delta_{pq} \sigma_p^{(u)} \sigma_q^{(u)} \right) g_{1,p} g_{1,q} \,, \tag{4.1}$$

where p and q run over the experimental data points, $g_{1,p} \equiv g_1(x_p, Q_p^2)$ ($g_{1,q} \equiv g_1(x_q, Q_q^2)$), $\sigma_{i,p}^{(c)}$ are the various sources of correlated uncertainties, and $\sigma_p^{(u)}$ are the uncorrelated uncertainties, which are in turn found as a sum in quadrature of all uncorrelated sources of statistical $\sigma_{i,p}^{(\mathrm{stat})}$ and systematic $\sigma_{i,p}^{(\mathrm{syst})}$ uncertainties on each point:

$$\left(\sigma_p^{(u)}\right)^2 = \sum_i \left(\sigma_{i,p}^{(\mathrm{stat})}\right)^2 + \sum_j \left(\sigma_{j,p}^{(\mathrm{syst})}\right)^2 . \tag{4.2}$$

The correlation matrix is defined as

$$\rho_{pq} = \frac{\mathrm{cov}_{pq}}{\sigma_p^{(\mathrm{tot})} \sigma_q^{(\mathrm{tot})} g_{1,p} g_{1,q}} \,, \tag{4.3}$$

where the total uncertainty $\sigma_p^{(\mathrm{tot})}$ on the p-th data point is

$$\left(\sigma_p^{(\mathrm{tot})}\right)^2 = (\sigma_p^{(u)})^2 + \sum_i \left(\sigma_{i,p}^{(c)}\right)^2 . \tag{4.4}$$

In Tab. 4.2, we show the average experimental uncertainties for each data set, with uncertainties separated into statistical and correlated systematics. All values are given as absolute uncertainties and refer to the structure function g_1, which has been reconstructed for each experiment as discussed above. As in the case of Tab. 4.1, we provide the values before and after kinematic cuts (if different).

Finally, notice that in both Tabs. 4.1-4.2 we distinguish between experiments, defined as groups of data which cannot be correlated to each other, and data sets within a given experiment, which could in principle be correlated with each other, as they correspond to measurements of different observables in the same experiment, or measurements of the same observable in different years. Even though, in practice, only two experiments provide such correlated systematics (see Tab. 4.2), this distinction will be useful in the minimization strategy, see Sect. 4.3 below.

Experiment	Set	$\langle\delta g_{1\mathrm{s}}\rangle$	$\langle\delta g_{1\mathrm{c}}\rangle$	$\langle\delta g_{1\mathrm{tot}}\rangle$
EMC				
	EMC-A1P	0.144	0.037	0.150
SMC				
	SMC-A1P	0.098	–	0.098
	SMC-A1D	0.116	–	0.116
SMClowx				
	SMClx-A1P	18.379 (0.291)	– (–)	18.379 (0.291)
	SMClx-A1D	22.536 (0.649)	– (–)	22.536 (0.649)
E143				
	E143-A1P	0.042 (0.046)	0.009 (0.009)	0.043 (0.047)
	E143-A1D	0.053 (0.058)	0.004 (0.005)	0.054 (0.059)
E154				
	E154-A1N	0.044	–	0.044
E155				
	E155-G1P	0.040 (0.043)	– (–)	0.040 (0.043)
	E155-G1N	0.124 (0.135)	– (–)	0.124 (0.135)
COMPASS-D				
	CMP07-A1D	0.061	–	0.061
COMPASS-P				
	CMP10-A1P	0.101	–	0.101
HERMES97				
	HER97-A1N	0.087 (0.093)	– (–)	0.087 (0.093)
HERMES				
	HER-A1P	0.067 (0.062)	– (–)	0.067 (0.062)
	HER-A1D	0.040 (0.034)	– (–)	0.040 (0.034)

Table 4.2: Averaged statistical, correlated systematic and total uncertainties before and after (in parenthesis) kinematic cuts for each of the experimental sets included in the present analysis. Uncorrelated systematic uncertainties are considered as part of the statistical uncertainty and they are added in quadrature. All values are absolute uncertainties and refer to the structure function g_1, which has been reconstructed for each experiment as discussed in the text. Details on the number of points and the kinematics of each data set are provided in Tab. 4.1.

4.1.2 Monte-Carlo generation of the pseudo-data sample

Error propagation from experimental data to the fit is handled by a Monte Carlo sampling of the probability distribution defined by data, as discussed in Sec. 3.2. The statistical sample is obtained by generating N_{rep} pseudodata replicas, according to a multi-Gaussian distribution centered at the data points and with a covariance equal to that of the original data. Explicitly, given an experimental data point $g_{1,p}^{(\mathrm{exp})} \equiv g_1(x_p, Q_p^2)$, we generate $k = 1, \ldots, N_{\mathrm{rep}}$ artificial points $g_{1,p}^{(\mathrm{art}),(k)}$ according to

$$g_{1,p}^{(\mathrm{art}),(k)}(x,Q^2) = \left[1 + \sum_i r_{(c),p}^{(k)} \sigma_{i,p}^{(c)} + r_{(u),p}^{(k)} \sigma_p^{(u)}\right] g_{1,p}^{(\mathrm{exp})}(x,Q^2)\,, \tag{4.5}$$

where $r_{(c),p}^{(k)}$, $r_{(u),p}^{(k)}$ are univariate Gaussianly distributed random numbers, and $\sigma_{i,p}^{(c)}$ and $\sigma_p^{(u)}$ are respectively the relative correlated systematic and statistical uncertainty. Unlike in the

unpolarized case, Eq. (4.5) receives no contribution from normalization uncertainties, given that all polarized observables are obtained as cross-section asymmetries.

The number of Monte Carlo replicas of the data is determined by requiring that the central values, uncertainties and correlations of the original experimental data can be reproduced to a given accuracy by taking averages, variances and covariances over the replica sample. A comparison between expectation values and variances of the Monte Carlo set and the corresponding input experimental values as a function of the number of replicas is shown in Fig. 4.2, where we display scatter plots of the central values and errors for samples of $N_{\rm rep} = 10, 100$ and 1000 replicas. A more quantitative comparison can be performed by defining suitable statistical estimators (see, for example, Appendix B of Ref. [99]).

In Tabs. 4.3-4.4 we show the percentage error and the scatter correlation r for central values and errors respectively, whose definition is recalled in appendix A. The scatter correlation r is, crudely speaking, the correlation between the input value and the value computed from the replica sample. We do not compute values for correlations, as these are available for a small number of data points from only two experiments, see Tab. 4.2. Some large values of the percentage uncertainty are due to the fact that, for some experiments, g_1 can take values which are very close to zero. It is clear from both the tables and the plots that a Monte Carlo sample of pseudodata with $N_{\rm rep} = 100$ is sufficient to reproduce the mean values and the errors of experimental data to an accuracy which is better than 5%, while the improvement in going up to $N_{\rm rep} = 1000$ is moderate. Therefore, we will henceforth use a $N_{\rm rep} = 100$ replica sample as a default for our reference fit.

4.2 Details of the QCD analysis

We will now briefly outline some details of the QCD analysis of polarized structure functions. The observable with which we fit experimental data is the g_1 structure function, Eq. (2.68), expressed in terms of the PDF combinations in Eqs. (2.59)-(2.60). From these relations, supplemented with Eqs. (2.49)-(2.54), it is clear that neutral-current g_1 data only allow for a direct determination of the four polarized PDF combinations $\Delta\Sigma$, ΔT_3, ΔT_8 and Δg: these will form the basis of polarized PDFs to be determined in our analysis. In principle, an intrinsic polarized component could also be present for each heavy flavor, as observed in Sec. 3.1. However, we will neglect it here and assume that heavy quark PDFs are dynamically generated above threshold by (massless) Altarelli-Parisi evolution, in a zero-mass variable-flavor number scheme (ZM-VFNS). In such a scheme all heavy quark mass effects are neglected. While they can be introduced for instance through the `FONLL` method [124], these effects have been shown to be relatively small already on the scale of present-day unpolarized PDF uncertainties, and thus are most likely negligible in the polarized case where uncertainties are rather larger. We will further comment on this issue in Sec. 5.5.2 and Appendix C, where we will sketch how to handle intrinsic charm contribution via the `FONLL` scheme.

The proton and neutron PDFs are related to each other by isospin, which we will assume to be exact, thus yielding

$$\Delta u^p = \Delta d^n, \quad \Delta d^p = \Delta u^n, \quad \Delta s^p = \Delta s^n, \tag{4.6}$$

and likewise for the polarized anti-quarks. In the following we will always assume that PDFs refer to the proton.

As discussed at length in Sec. 2.2.2, beyond leading order in QCD the first moment of all non-singlet combinations of quark and antiquark distributions are scale independent

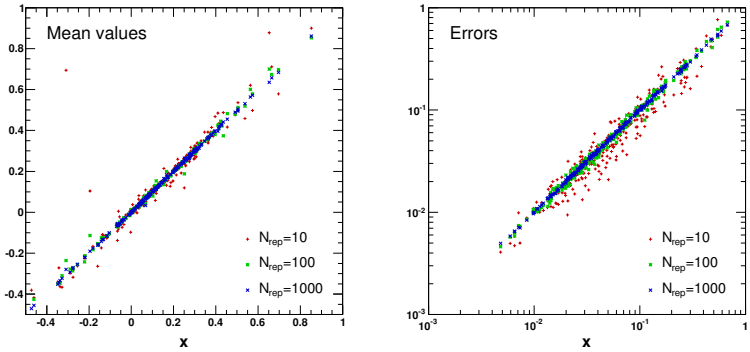

Figure 4.2: Scatter plot of experimental versus artificial Monte Carlo mean central values and absolute uncertainties of polarized structure functions computed from ensembles made of $N_{\rm rep} = 10, 100, 1000$ replicas.

	Estimator	$\left\langle \text{PE}\left[\langle g_1^{(\text{art})}\rangle\right]\right\rangle$ [%]			$r\left[g_1^{(\text{art})}\right]$		
	$N_{\rm rep}$	10	100	1000	10	100	1000
Experiment	EMC	23.7	3.5	2.9	.76037	.99547	.99712
	SMC	19.4	5.6	1.2	.94789	.99908	.99993
	SMClowx	183	25.8	15.4	.80370	.99239	.99960
	E143	18.5	5.7	2.1	.99159	.99860	.99984
	E154	239	44.0	21.9	.99635	.99981	.99994
	E155	37.3	13.4	4.3	.99798	.99993	.99998
	COMPASS-D	26.4	8.6	3.2	.96016	.98774	.99917
	COMPASS-P	16.4	1.9	1.5	.91942	.99829	.99902
	HERMES97	22.5	6.2	2.2	.96168	.99762	.99979
	HERMES	10.5	5.8	1.2	.98564	.99916	.99973

Table 4.3: Table of statistical estimators for the mean value computed from the Monte Carlo sample with $N_{\rm rep} = 10, 100, 1000$ replicas. Estimators refer to individual experiments and are defined in Appendix B of Ref. [99].

	Estimator	$\left\langle \text{PE}\left[\langle \delta g_1^{(\text{art})}\rangle\right]\right\rangle$ [%]			$r\left[\delta g_1^{(\text{art})}\right]$		
	$N_{\rm rep}$	10	100	1000	10	100	1000
Experiment	EMC	12.8	4.9	2.0	.97397	.99521	.99876
	SMC	22.4	5.4	1.7	.96585	.99489	.99980
	SMClowx	16.9	6.2	2.1	.97959	.99490	.99905
	E143	16.0	7.4	2.0	.95646	.98684	.99946
	E154	19.1	3.7	1.3	.99410	.99871	.99992
	E155	21.2	5.6	1.8	.99428	.99971	.99997
	COMPASS-D	15.5	5.2	1.6	.99375	.99687	.99993
	COMPASS-P	18.4	7.4	1.5	.99499	.99005	.99988
	HERMES97	17.9	6.4	1.6	.89065	.97318	.99894
	HERMES	19.5	6.0	1.6	.91523	.99237	.99942

Table 4.4: Same as Tab. 4.3, but for errors.

due to axial current conservation. Besides, we enforce $SU(2)$ and $SU(3)$ flavor asymmetry by requiring the first moments of the non-singlet, C-even, combinations, Eqs. (2.75) to be fixed by the experimental values of baryon octet decay constants, Eqs. (2.80). Actually, a much larger uncertainty on the octet axial charge, up to about 30%, is found if SU(3) symmetry is violated [160] with respect to that quoted in Eqs. (2.80). Even though a detailed phenomenological analysis does not seem to support this conclusion [161], we will take as default this more conservative uncertainty estimation

$$a_8 = 0.585 \pm 0.176\,. \tag{4.7}$$

The impact of replacing this with the more aggressive determination given in Eq. (2.80) will be studied in Sect. 4.4.2.

Structure functions will be computed in terms of polarized parton distributions using the so-called NNPDF `FastKernel` method, introduced in Ref. [104]. In short, in this method the PDFs at scale Q^2 are obtained by convoluting the parton distributions at the parametrization scale Q_0^2 with a set of Green's functions, which are in turn obtained by solving the QCD evolution equations in Mellin space (below also denoted as N-space). These Green's functions are then convoluted with coefficient functions, so that the structure function can be directly expressed in terms of the PDFs at the parametrization scale through suitable kernels K. In terms of the polarized PDFs at the input scale (labelled with the subscript 0) we have

$$g_1^p = \left\{ K_{g_1,\Delta\Sigma} \otimes \Delta\Sigma_0 + K_{g_1,\Delta g} \otimes \Delta g_0 + K_{g_1,+} \otimes \left(\Delta T_{3,0} + \frac{1}{3}\Delta T_{8,0} \right) \right\}, \tag{4.8}$$

where the kernels $K_{g_1,\Delta\Sigma}$, $K_{g_1,\Delta g}$, $K_{g_1,+}$ take into account both the coefficient functions and Q^2 evolution. This way of expressing structure functions is amenable to numerical optimization, because all kernels can then be precomputed and stored, and convolutions may be reduced to matrix multiplications by projecting onto a set of suitable basis functions.

The neutron polarized structure function g_1^n is given in terms of the proton and deuteron ones as

$$g_1^n = 2\frac{g_1^d}{1 - 1.5\omega_D} - g_1^p\,, \tag{4.9}$$

with $\omega_D = 0.05$ the probability that the deuteron is found in a D state. Under the assumption of exact isospin symmetry, the expression of g_1^n in terms of parton densities is obtained from Eq. (4.6) by interchanging the up and down quark PDFs, which amounts to changing the sign of ΔT_3. We will assume the values $\alpha_s(M_Z^2) = 0.119$ for the strong coupling constant and $m_c = 1.4$ GeV and $m_b = 4.75$ GeV for the charm and bottom quark masses respectively.

We have benchmarked our implementation of the evolution of polarized parton densities up to NLO by cross-checking against the Les Houches polarized PDF evolution benchmark tables [202].* These tables were obtained from a comparison of the `HOPPET` [203] and `PEGASUS` [134] evolution codes, which are $x-$space and $N-$space codes respectively. In order to perform a meaningful comparison, we use the so-called iterated solution of the $N-$space evolution equations and use the same initial PDFs and running coupling as in [202]. The relative difference $\varepsilon_{\rm rel}$ between our PDF evolution and the benchmark tables of Refs. [202] at NLO in the ZM-VFNS are tabulated in Tab. 4.5 for various combinations of polarized PDFs: the accuracy of our code is $\mathcal{O}(10^{-5})$ for all relevant values of x, which is the nominal accuracy of the agreement between `HOPPET` and `PEGASUS`. Therefore, we can conclude that the accuracy

*Note that in Ref. [202] the polarized sea PDFs are given incorrectly, and should be $x\Delta\bar{u} = -0.045x^{0.3}(1-x)^7$ and $x\Delta\bar{d} = -0.055x^{0.3}(1-x)^7$.

x	$\varepsilon_{\rm rel}(\Delta u^-)$	$\varepsilon_{\rm rel}(\Delta d^-)$	$\varepsilon_{\rm rel}(\Delta\Sigma)$	$\varepsilon_{\rm rel}(\Delta g)$
10^{-3}	$1.1\,10^{-4}$	$9.2\,10^{-5}$	$9.9\,10^{-5}$	$1.1\,10^{-4}$
10^{-2}	$1.4\,10^{-4}$	$1.9\,10^{-4}$	$3.5\,10^{-4}$	$9.3\,10^{-5}$
0.1	$1.2\,10^{-4}$	$1.6\,10^{-4}$	$5.4\,10^{-6}$	$1.7\,10^{-4}$
0.3	$2.3\,10^{-6}$	$1.1\,10^{-5}$	$7.5\,10^{-6}$	$1.7\,10^{-5}$
0.5	$5.6\,10^{-6}$	$9.6\,10^{-6}$	$1.6\,10^{-5}$	$2.5\,10^{-5}$
0.7	$1.2\,10^{-4}$	$9.2\,10^{-7}$	$1.6\,10^{-4}$	$7.8\,10^{-5}$
0.9	$3.5\,10^{-3}$	$1.1\,10^{-2}$	$4.1\,10^{-3}$	$7.8\,10^{-3}$

Table 4.5: Percentage difference between FastKernel perturbative evolution of polarized PDFs and the Les Houches benchmark tables [202] for different polarized PDF combinations at NLO in the ZM-VFNS.

of the polarized PDF evolution in the `FastKernel` framework is satisfactory for precision phenomenology.

Finally, we include exactly all kinematic target mass corrections within the formalism presented in Sec. 2.5. However, we notice that the numerical implementation of Eqs. (2.106) or Eq. (2.108) is difficult, because of the presence of the first derivative of g_1 in the correction term. For this reason, we do not factorize TMCs into the hard kernels, as it was done in the unpolarized case, where the first derivative of F_1 does not appear [101]. Rather, we will include target mass effects in an iterative way: we start by performing a fit in which we set $M = 0$ and at each iteration of the minimization procedure the target mass corrected g_1 structure function is computed by means of Eqs. (2.106-2.108) using the g_1 obtained in the previous minimization step. We found that this strategy allows for convergence in a few minimization steps, hence TMCs are properly included when the fit is stopped.

4.3 Fitting strategy

We will now discuss some details of the fitting strategy adopted in the `NNPDFpol1.0` analysis. In particular, we describe how PDFs are parametrized in terms of neural networks and how they are trained to experimental data to obtain the optimal fit. The main steps of the minimization strategy have been summarized in Sec. 3.2, and can be found in Ref. [104]. Here, we emphasize some specific features which were introduced to deal with issues peculiar to the polarized case, in particular with respect to the implementation of theoretical constraints.

4.3.1 Neural network parametrization

The four independent polarized PDF flavor combinations in the evolution basis, $\Delta\Sigma$, ΔT_3 and ΔT_8, and the gluon Δg are separately parametrized using a multi-layer feed-forward neural network [108]. All neural networks have the same architecture, namely 2-5-3-1, which corresponds to 37 free parameters for each PDF, and thus a total of 148 free parameters. This is to be compared to about 10-15 free parameters for all other available determinations of polarized PDFs within the *standard* methodology, see Sec. 3.1. This parametrization has been explicitly shown to be redundant in the unpolarized case, in that results are unchanged when a smaller neural network architecture is adopted: this ensures that results do not depend on the architecture [108]. Given that polarized data are much less abundant and affected by much

PDF	m	n
$\Delta\Sigma(x,Q_0^2)$	$[1.5,3.5]$	$[0.2,0.7]$
$\Delta g(x,Q_0^2)$	$[2.5,5.0]$	$[0.4,0.9]$
$\Delta T_3(x,Q_0^2)$	$[1.5,3.5]$	$[0.4,0.7]$
$\Delta T_8(x,Q_0^2)$	$[1.5,3.0]$	$[0.1,0.6]$

Table 4.6: Ranges for the small and large x preprocessing exponents Eq. (4.11).

larger uncertainties than their unpolarized counterparts, this architecture is surely adequate in the polarized case too.

The neural network parametrization is supplemented with a preprocessing function. In principle, large enough neural networks can reproduce any functional form given sufficient training time. However, the training can be made more efficient by adding a preprocessing step, *i.e.* by multiplying the output of the neural networks by a fixed function. The neural network then only fits the deviation from this function, which improves the speed of the minimization procedure if the preprocessing function is suitably chosen. We thus write the input PDF basis in terms of preprocessing functions and neural networks $\mathrm{NN}_{\Delta\mathrm{pdf}}$ as follows

$$\begin{aligned}
\Delta\Sigma(x,Q_0^2) &= (1-x)^{m_1}x^{-n_1}\mathrm{NN}_{\Delta\Sigma}(x)\,,\\
\Delta T_3(x,Q_0^2) &= A_3(1-x)^{m_3}x^{-n_3}\mathrm{NN}_{\Delta T_3}(x)\,,\\
\Delta T_8(x,Q_0^2) &= A_8(1-x)^{m_8}x^{-n_{\Delta T_8}}\mathrm{NN}_{\Delta T_3}(x)\,,\\
\Delta g(x,Q_0^2) &= (1-x)^{m_g}x^{-n_g}\mathrm{NN}_{\Delta g}(x)\,.
\end{aligned} \tag{4.10}$$

Of course, one should check that no bias is introduced in the choice of preprocessing functions. To this purpose, we first select a reasonable range of values for the large and small-x preprocessing exponents m and n, and produce a PDF determination by choosing for each replica a value of the exponents at random with uniform distribution within this range. We then determine effective exponents for each replica, defined as

$$m_{\mathrm{eff}}(Q^2) \equiv \lim_{x\to 1}\frac{\ln\Delta f(x,Q^2)}{\ln(1-x)}\,, \tag{4.11}$$

$$n_{\mathrm{eff}}(Q^2) \equiv \lim_{x\to 0}\frac{\ln\Delta f(x,Q^2)}{\ln\frac{1}{x}}\,, \tag{4.12}$$

where $\Delta f = \Delta\Sigma$, ΔT_3, ΔT_8, Δg. Finally, we check that the range of variation of the preprocessing exponents is wider than the range of effective exponents for each PDF. If it is not, we enlarge the range of variation of preprocessing, then repeat the PDF determination, and iterate until the condition is satisfied. This ensures that the range of effective large- and small-x exponents found in the fit is not biased, and in particular not restricted, by the range of preprocessing exponents. Our final values for the preprocessing exponents are summarized in Tab. 4.6, while the effective exponents obtained in our fit will be discussed in Sect. 4.4.4. It is apparent from Tab. 4.6 that the allowed range of preprocessing exponents is rather wider than in the unpolarized case, as a consequence of the limited amount of experimental information.

The nonsinglet triplet and octet PDF combinations in the parametrization basis, Eq. (4.11), ΔT_3 and ΔT_8, are supplemented by a prefactor. This is because these PDFs must satisfy the

sum rules Eqs. (2.75), which are enforced by letting

$$\begin{aligned} A_3 &= \frac{a_3}{\int_0^1 dx\,(1-x)^{m_3}x^{-n_3}\mathrm{NN}_{\Delta T_3}(x)}\,, \\ A_8 &= \frac{a_8}{\int_0^1 dx\,(1-x)^{m_8}x^{-n_8}\mathrm{NN}_{\Delta T_8}(x)}\,. \end{aligned} \tag{4.13}$$

The integrals are computed numerically each time the parameters of the PDF set are modified. The values of a_3 and a_8 are chosen for each replica as Gaussianly distributed numbers, with central value and width given by the corresponding experimental values, Eqs. (2.80)-(4.7).

4.3.2 Genetic algorithm minimization

As discussed at length in Ref. [101] and summarized in Sec. 3.2, minimization with a neural network parametrization of PDFs must be performed through an algorithm which explores the very wide functional space efficiently. This is done by means of a genetic algorithm, which is used to minimize a suitably defined figure of merit, namely the error function [101],

$$E^{(k)} = \frac{1}{N_{\rm dat}} \sum_{I,J=1}^{N_{\rm dat}} \left(g_I^{(\rm art)(k)} - g_I^{(\rm net)(k)}\right) \left((\rm cov)^{-1}\right)_{IJ} \left(g_J^{(\rm art)(k)} - g_J^{(\rm net)(k)}\right) . \tag{4.14}$$

Here $g_I^{(\rm art)(k)}$ is the value of the observable g_I at the kinematical point I corresponding to the Monte Carlo replica k, and $g_I^{(\rm net)(k)}$ is the same observable computed from the neural network PDFs; the covariance matrix $(\rm cov)_{IJ}$ is defined in Eq. (4.1).

The minimization procedure we adopt follows closely that of Ref. [100], to which we refer for a more general discussion. Minimization is perfomed by means of a genetic algorithm, which minimizes the figure of merit, Eq. (4.14) by generating, at each minimization step, a pool of new neural nets, obtained by randomly mutating the parameters of the starting set, and retaining the configuration which corresponds to the lowest value of the figure of merit.

The parameters which characterize the behavior of the genetic algorithm are tuned in order to optimize the efficiency of the minimization procedure. We essentially rely on previous experience of the development of unpolarized NNPDF sets: in particular, the algorithm is characterized by a mutation rate, which decreases as a function of the number of the algorithm iterations $N_{\rm ite}$ according to the law [101]

$$\eta_{i,j} = \eta_{i,j}^{(0)} / N_{\rm ite}^{r_\eta}\,. \tag{4.15}$$

This way, in the early stages of the training large mutations are allowed, while they become less likely as one approaches the minimum. The starting mutation rates are chosen to be larger for PDFs which contain more information. We perform two mutations per PDF at each step, with the starting rates given in Tab. 4.7. The exponent r_η has been introduced in order to optimally span the whole range of possible beneficial mutations and it is randomized between 0 and 1 at each iteration of the genetic algorithm, as in Ref. [104].

Furthermore, following Ref. [104], we let the number of new candidate solutions depend on the stage of the minimization. At earlier stages of the minimization, when the number of generations is smaller than $N^{\rm mut}$, we use a large population of mutants, $N_{\rm mut}^a \gg 1$, so a larger space of mutations is being explored. At later stages of the minimization, as the minimum is

$\eta^{(0)}_{i,\Delta\Sigma}$	$\eta^{(0)}_{i,\Delta g}$	$\eta^{(0)}_{i,\Delta T_3}$	$\eta^{(0)}_{i,\Delta T_8}$
5, 0.5	5, 0.5	2, 0.2	2, 0.2

Table 4.7: The initial values of the mutation rates for the two mutations of each PDF.

$N^{\rm mut}_{\rm gen}$	$N^{a}_{\rm mut}$	$N^{b}_{\rm mut}$	$N^{\rm wt}_{\rm gen}$	$E^{\rm sw}$
200	50	10	5000	2.5

Table 4.8: Values of the parameters of the genetic algorithm.

approached, a smaller number of mutations $N^{b}_{\rm mut} \ll N^{a}_{\rm mut}$ is used. The values of the parameters $N^{\rm mut}_{\rm gen}$, $N^{a}_{\rm mut}$ and $N^{b}_{\rm mut}$ are collected in Tab. 4.8.

Because the minimization procedure stops the fit to all experiments at once, we must make sure that the quality of the fit to different experiments is approximately the same. This is nontrivial, because of the variety of experiments and data sets included in the fit. Therefore, the figure of merit per data point for a given set is not necessarily a reliable indicator of the quality of the fit to that set, because some experiments may have systematically underestimated or overestimated uncertainties. Furthermore, unlike for unpolarized PDF fits, information on the experimental covariance matrix is only available for a small subset of experiments, so for most experiments statistical and systematic errors must be added in quadrature, thereby leading to an overestimate of uncertainties: this leads to a wide spread of values of the figure of merit, whose value depends on the size of the correlated uncertainties which are being treated as uncorrelated.

A methodology to deal with this situation was developed in Ref. [104]. The idea is to first determine the optimal value of the figure of merit for each experiment, *i.e.* a set of target values $E^{\rm targ}_i$ for each of the i experiments, then during the fit give more weight to experiments for which the figure of merit is further away from its target value, and stop to train experiments which have already reached the target value. This is done by minimizing, instead of the figure of merit Eq. (4.14), the weighted figure of merit

$$E^{(k)}_{\rm wt} = \frac{1}{N_{\rm dat}} \sum_{j=1}^{N_{\rm sets}} p^{(k)}_j N_{{\rm dat},j} E^{(k)}_j \,, \tag{4.16}$$

where $E^{(k)}_j$ is the error function for the j-th data set with $N_{{\rm dat},j}$ points, and the weights $p^{(k)}_j$ are given by

1. If $E^{(k)}_i \geq E^{\rm targ}_i$, then $p^{(k)}_i = \left(E^{(k)}_i / E^{\rm targ}_i\right)^n$,
2. If $E^{(k)}_i < E^{\rm targ}_i$, then $p^{(k)}_i = 0$,

with n a free parameter which essentially determines the amount of weighting. In the unpolarized fits of Refs. [104, 106, 107, 110] the value $n = 2$ was used. Here instead we will choose $n = 3$. This larger value, determined by trial and error, is justified by the wider spread of figures of merit in the polarized case, which in turn is related to the absence of correlated systematics for most experiments.

The target values $E^{\rm targ}_i$ are determined through an iterative procedure: they are set to one at first, then a very long fixed-length fit is run, and the values of E_i are taken as targets for a

new fit, which is performed until stopping (according to the criterion to be discussed in the following Section). The values of E_i at the end of this fit are then taken as new targets until convergence is reached, usually after a couple iterations.

Weighted training stops after the first $N_{\rm gen}^{\rm wt}$ generations, unless the total error function Eq. (4.14) is above some threshold $E^{(k)} \geq E^{\rm sw}$. If it is, weighted training continues until $E^{(k)}$ falls below the threshold value. Afterwards, the error function is just the unweighted error function Eq. (4.14) computed on experiments. This ensures that the figure of merit behaves smoothly in the last stages of training. The values for the parameters $N_{\rm gen}^{\rm wt}$ and $E^{\rm sw}$ are also given in Tab. 4.8.

4.3.3 Determination of the optimal fit

Because the neural network parametrization is extremely redundant, it may be able to fit not only the underlying behavior of the PDFs, but also the statistical noise in the data. Therefore, the best fit does not necessarily coincide with the absolute minimum of the figure of merit Eq. (4.14). We thus determine the best fit, as in Refs. [100, 101], using a cross-validation method [204]: for each replica, the data are randomly divided in two sets, training and validation, which include a fraction $f_{\rm tr}^{(j)}$ and $f_{\rm val}^{(j)} = 1 - f_{\rm tr}^{(j)}$ of the data points respectively. The figure of merit Eq. (4.14) is then computed for both sets. The training figure of merit function is minimized through the genetic algorithm, while the validation figure of merit is monitored: when the latter starts increasing while the former still decreases, the fit is stopped. This means that the fit is stopped as soon as the neural network is starting to learn the statistical fluctuations of the points, which are different in the training and validation sets, rather than the underlying law which they share.

In the unpolarized fits of Refs. [100, 101, 104, 106, 107, 110] equal training and validation fractions were uniformly chosen, $f_{\rm tr}^{(j)} = f_{\rm val}^{(j)} = 1/2$. However, in this case we have to face the problem that the number of data points is quite small: most experiments include about ten data points (see Tab. 4.1). Hence, it is difficult to achieve a stable minimization if only half of them are actually used for minimization, as we have explicitly verified. Therefore, we have chosen to include 80% of the data in the training set, *i.e.* $f_{\rm tr}^{(j)} = 0.8$ and $f_{\rm val}^{(j)} = 0.2$. We have explicitly verified that the fit quality which is obtained in this case is comparable to the one achieved when including all data in the training set (*i.e.* with $f_{\rm tr}^{(j)} = 1.0$ and $f_{\rm val}^{(j)} = 0.0$), but the presence of a nonzero validation set allows for a satisfactory stopping, as we have checked by explicit inspection of the profiles of the figure of merit as a function of training time.

In practice, in order to implement cross-validation we must determine a stopping criterion, namely, give conditions which must be satisfied in order for the minimization to stop. First, we require that the weighted training stage has been completed, *i.e.*, that the genetic algorithm has been run for at least $N_{\rm gen}^{\rm wt}$ minimization steps. Furthermore, we check that all experiments have reached a value of the figure of merit below a minimal threshold $E_{\rm thr}$. Note that because stopping can occur only after weighted training has been switched off, and this in turn only happens when the figure of merit falls below the value $E^{\rm sw}$, the total figure of merit must be below this value in order for stopping to be possible.

We then compute moving averages

$$\langle E_{\rm tr,val}(i)\rangle \equiv \frac{1}{N_{\rm smear}} \sum_{l=i-N_{\rm smear}+1}^{i} E_{\rm wt;tr,val}(l)\,, \tag{4.17}$$

of the figure of merit Eq. (4.16) for either the training or the validation set at the l-th genetic

$N_{\rm gen}^{\rm max}$	$E_{\rm thr}$	$N_{\rm smear}$	$\Delta_{\rm smear}$	$\delta_{\rm tr}$	$\delta_{\rm val}$
20000	8	100	100	$5\cdot10^{-4}$	$5\cdot10^{-4}$

Table 4.9: Parameters for the stopping criterion.

minimzation step. The fit is then stopped if

$$r_{\rm tr} < 1-\delta_{\rm tr} \quad \text{and} \quad r_{\rm val} > 1+\delta_{\rm val}\,, \tag{4.18}$$

where

$$r_{\rm tr} \equiv \frac{\langle E_{\rm tr}(i)\rangle}{\langle E_{\rm tr}(i-\Delta_{\rm smear})\rangle}\,, \tag{4.19}$$

$$r_{\rm val} \equiv \frac{\langle E_{\rm val}(i)\rangle}{\langle E_{\rm val}(i-\Delta_{\rm smear})\rangle}\,. \tag{4.20}$$

The parameter $N_{\rm smear}$ determines the width of the moving average; the parameter $\Delta_{\rm smear}$ determines the distance between the two points along the minimization path which are compared in order to determine whether the figure of merit is increasing or decreasing; and the parameters $\delta_{\rm tr}$, $\delta_{\rm val}$ are the threshold values for the decrease of the training and increase of the validation figure of merit to be deemed significant. The optimal value of these parameters should be chosen in such a way that the fit does not stop on a statistical fluctuation, yet it does stop before the fit starts overlearning (*i.e.* learning statistical fluctuation). As explained in Ref. [104], this is done studying the profiles of the error functions for individual data set and for individual replicas. In order to avoid unacceptably long fits, training is stopped anyway when a maximum number of iterations $N_{\rm gen}^{\rm max}$ is reached, even though the stopping conditions Eqs. (4.18) are not satisfied. This leads to a small loss of accuracy of the corresponding fits: this is acceptable provided it only happens for a small enough fraction of replicas. If a fit stops at $N_{\rm gen}^{\rm max}$ without the stopping criterion having been satisfied, we also check that the total figure of merit is below the value $E^{\rm sw}$ at which weighted training is switched off. If it hasn't, we conclude that the specific fit has not converged, and we retrain the same replica, *i.e.*, we perform a new fit to the same data starting with a different random seed. This only occurs in about one or two percent of cases. The full set of parameters which determine the stopping criterion is given in Tab. 4.9.

An example of how the stopping criterion works in practice is shown in Fig. 4.3. We display the moving averages Eq. (4.17) of the training and validation error functions $\langle E_{\rm tr,val}^{(k)}\rangle$, computed with the parameter settings of Tab. 4.9, and plotted as a function of the number of iterations of the genetic algorithm, for a particular replica and for two of the experiments included in the fit. The wide fluctuations which are observed in the first part of training, up to the $N_{\rm gen}^{\rm wt}$-th generation, are due to the fact that the weights which enter the definition of the figure of merit Eq. (4.16) are frequently adjusted. Nevertheless, the downwards trend of the figure of merit is clearly visible. Once the weighted training is switched off, minimization proceeds smoothly. The vertical line denotes the point at which the stopping criterion is satisfied. Here, we have let the minimization go on beyond this point, and we clearly see that the minimization has entered an overlearning regime, in which the validation error function $E_{\rm val}^{(k)}$ is rising while the training $E_{\rm tr}^{(k)}$ is still decreasing. Note that the stopping point, which in this particular case occurs at $N_{\rm gen}^{\rm stop} = 5794$, is determined by verifying that the stopping criteria are satisfied by the total figure of merit, not that of individual experiments shown here. The fact that the two different

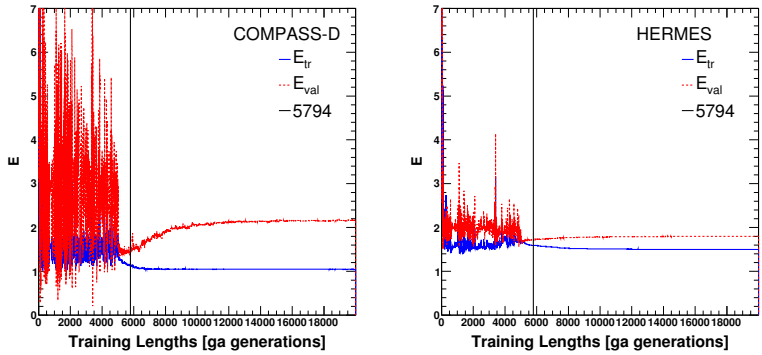

Figure 4.3: Behaviour of the moving average Eq. (4.17) of the training and validation figure of merit for two different data sets included in a global fit (`COMPASS-P` and `HERMES`) as a function of training length. The straight vertical line indicates the point at which the fit stops with the stopping parameters of Tab. 4.9. The weighted training is switched off at $N_{\rm gen}^{\rm wt} = 5000$.

experiments considered here both start overlearning at the same point shows that the weighted training has been effective in synchronizing the fit quality for different experiments.

4.3.4 Theoretical constraints

Polarized PDFs are only loosely constrained by data, which are scarce and not very accurate. Theoretical constraints are thus especially important in reducing the uncertainty on the PDFs. We consider in particular positivity and integrability.

Positivity of the individual cross-sections which enter the polarized asymmetries Eq. (2.83) implies that, up to power-suppressed corrections, longitudinal polarized structure functions are bounded by their unpolarized counterparts, *i.e.*

$$|g_1(x,Q^2)| \leq F_1(x,Q^2)\,. \tag{4.21}$$

At leading order, structure functions are proportional to parton distributions, so imposing Eq. (4.21) for any process (and a similar condition on an asymmetry which is sensitive to polarized gluons [205]), would imply

$$|\Delta f_i(x,Q^2)| \leq f_i(x,Q^2) \tag{4.22}$$

for any pair of unpolarized and polarized PDFs f and Δf, for all quark flavors and gluon i, for all x, and for all Q^2. Beyond leading order, the condition Eq. (4.21) must still hold, but it does not necessarily imply Eq. (4.22). Rather, one should then impose at least a number of conditions of the form of Eq. (4.21) on physically measurable cross-sections which is equal to the number of independent polarized PDFs. For example, in principle one may require that the condition Eq. (4.21) is separately satisfied for each flavor, *i.e.* when only contributions from the i-th flavor are included in the polarized and unpolarized structure function: this corresponds to requiring positivity of semi-inclusive structure functions which could in principle be measured (and that fragmentation effects cancel in the ratio). A condition on the gluon can be obtained

by imposing positivity of the polarized and unpolarized cross-sections for inclusive Higgs production in gluon-proton scattering [205], again measurable in principle if not in practice.

Because $g_1/F_1 \sim x$ as $x \to 0$ [206], the positivity bound Eq. (4.21) is only significant at large enough $x \gtrsim 10^{-2}$. On the other hand, at very large x the NLO corrections to the LO positivity bound become negligible [205,207]. Therefore, the NLO positivity bound in practice only differs from its LO counterpart Eq. (4.22) in a small region $10^{-2} \lesssim x \lesssim 0.3$, and even there by an amount of rather less that 10% [205], which is negligible in comparison to the size of PDF uncertainties, as we shall see explicitly in Sec. 4.4.

Therefore, we will impose the leading-order positivity bound Eq. (4.22) on each flavor combination $\Delta q_i + \Delta\bar{q}_i$ and on the gluon Δg (denoted as Δf_i below). We do this by requiring

$$|\Delta f_i(x,Q^2)| \leq f_i(x,Q^2) + \sigma_i(x,Q^2), \tag{4.23}$$

where $\sigma_i(x,Q^2)$ is the uncertainty on the corresponding unpolarized PDF combination $f_i(x,Q^2)$ at the kinematic point (x,Q^2). This choice is motivated by two considerations. First, it is clearly meaningless to impose positivity of the polarized PDF to an accuracy which is greater than that with which the unpolarized PDF has been determined. Second, because the unpolarized PDFs satisfy NLO positivity, they can become negative and thus they may have nodes. As a consequence, the LO bound Eq. (4.22) would imply that the polarized PDF must vanish at the same point, which would be clearly meaningless.

As in Ref. [104] positivity is imposed during the minimization procedure, thereby guaranteeing that the genetic algorithm only explores the subspace of acceptable physical solutions. This is done through a Lagrange multiplier $\lambda_{\rm pos}$, *i.e.* by computing the polarized PDF at $N_{\rm dat,pos}$ fixed kinematic points (x_p,Q_0^2) and then adding to the error function Eq. (4.14) a contribution

$$E_{\rm pos}^{(k)} = \lambda_{\rm pos} \sum_{p=1}^{N_{\rm dat,pos}} \left\{ \sum_{j=u+\bar{u},d+\bar{d},s+\bar{s},g} \Theta\left[\left| \Delta f_j^{({\rm net})(k)}(x_p,Q_0^2) \right| - \left(f_j+\sigma_j\right)(x_p,Q_0^2) \right] \right.$$
$$\left. \times \left[\left| \Delta f_j^{({\rm net})(k)}(x_p,Q_0^2) \right| - \left(f_j+\sigma_j\right)(x_p,Q_0^2) \right] \right\}. \tag{4.24}$$

This provides a penalty, proportional to the violation of positivity, which enforces Eq. (4.23) separately for all the non-zero quark-antiquark combinations. The values of the unpolarized PDF combination $f_j(x,Q^2)$ and its uncertainty $\sigma_j(x,Q^2)$ are computed using the `NNPDF2.1` PDF set at NLO [106], while $\Delta f_j^{({\rm net})(k)}$ is the corresponding polarized PDF computed from the neural network parametrization for the k-th replica. The polarized and unpolarized PDFs are evaluated at $N_{\rm dat,pos} = 20$ points with x equally spaced in the interval

$$x \in [10^{-2}, 0.9]. \tag{4.25}$$

Positivity is imposed at the initial scale $Q_0^2 = 1$ GeV^2 since once positivity is enforced at low scales, it is automatically satisfied at larger scales [205,207]. After stopping, we finally test the positivity condition Eq. (4.23) is satisfied on a grid of $N_{\rm dat,pos} = 40$ points in the same intervals. Replicas for which positivity is violated in one or more points are discarded and retrained.

In the unpolarized case, in which positivity only played a minor role in constraining PDFs, a fixed value of the Lagrange multiplier $\lambda_{\rm pos}$ was chosen. In the polarized case it turns out to be necessary to vary the Lagrange multiplier along the minimization. Specifically, we let

$$\left\{ \begin{array}{lcll} \lambda_{\rm pos} & = & \lambda_{\rm max}^{(N_{\rm gen}-1)/(N_{\lambda_{\rm max}}-1)} & N_{\rm gen} < N_{\lambda_{\rm max}} \\ \lambda_{\rm pos} & = & \lambda_{\rm max} & N_{\rm gen} \geq N_{\lambda_{\rm max}}. \end{array} \right. \tag{4.26}$$

This means that the Lagrange multiplier increases as the minimization proceeds, starting from $\lambda_{\rm pos} = 1$, at the first minimization step, $N_{\rm gen} = 1$, up to $\lambda_{\rm pos} = \lambda_{\rm max} \gg 1$ when $N_{\rm gen} = N_{\lambda_{\rm max}}$. After $N_{\lambda_{\rm max}}$ generations $\lambda_{\rm pos}$ is then kept constant to $\lambda_{\rm max}$. The rationale behind this choice is that the genetic algorithm can thus learn experimental data and positivity at different stages of minimization. During the early stages, the contribution coming from the modified error function Eq. (4.24) is negligible, due to the moderate value of the Lagrange multiplier; hence, the genetic algorithm will mostly learn the basic shape of the PDF driven by experimental data. As soon as the minimization proceeds, the contribution coming from the Lagrange multiplier increases, thus ensuring the proper learning of positivity: at this stage, most of the replicas which will not fulfill the positivity bound will be discarded.

The final values of $N_{\lambda_{\rm max}} = 2000$ and $\lambda_{\rm max} = 10$ have been determined as follows. First of all, we have performed a fit without any positivity constraint and we have observed that data were mostly learnt in about 2000 generations: hence we have taken this value for $N_{\lambda_{\rm max}}$. Then we have tried different values for $\lambda_{\rm max}$ until we managed to reproduce the same χ^2 obtained in the previous, positivity unconstrained, fit. This ensures that positivity is not learnt to the detriment of the global fit quality.

Notice that the value of $\lambda_{\rm max}$ is rather small if compared to the analogous Lagrange multiplier used in the unpolarized case [106]. This depends on the fact that, in this latter case, positivity is learnt at the early stages of minimization, when the error function can be much larger than its asymptotic value: a large Lagrange multiplier is then needed to select the best replicas. Also, unpolarized PDFs are quite well constrained by data and positivity is almost automatically fulfilled, except in some restricted kinematic regions; only a few replicas violate positivity and need to be penalized. This means that the behavior of the error function Eq. (4.14), which governs the fitting procedure, is essentially dominated by data instead of positivity.

In the polarized case, instead, positivity starts to be effectively implemented only after some minimizaton steps, when the error function has already decreased to a value of a few units. Furthermore, we have checked that, at this stage, most of replicas slightly violate the positivity condition Eq. (4.23): thus, a too large value of the Lagrange multiplier on the one hand would penalize replicas which are good in reproducing experimental data and only slightly worse in reproducing positivity; on the other, it would promote replicas which fulfill positivity but whose fit to data is quite bad. As a consequence of this behavior, the convergence of the minimization algorithm would be harder to reach. We also verified that, using a value for the Lagrange multiplier up to $\lambda_{\rm pos} = 100$ leads to no significant improvement neither in the fulfillment of positivity requirement nor in the fit quality. We will show in detail the effects of the positivity bound Eq. (4.23) on the fitted replicas and on polarized PDFs in Sect. 4.4.

Finally, we impose that PDFs are integrable, *i.e.* that they have finite first moments. This corresponds to the assumption that the nucleon matrix element of the axial current for the i-th flavor is finite. The integrability condition is imposed by computing at each minimization step the integral of each of the polarized PDFs in a given interval,

$$I(x_1, x_2) = \int_{x_1}^{x_2} dx\, \Delta q_i(x, Q_0^2) \qquad \Delta q_i = \Delta\Sigma, \Delta g, \Delta T_3, \Delta T_8 \tag{4.27}$$

with x_1 and x_2 chosen in the small x region, well below the data points, and verifying that in this region the growth of the integral as x_1 decreases for fixed x_2 is less than logarithmic. In

NNPDFpol1.0	
$\chi^2_{\rm tot}$	0.77
$\langle E\rangle \pm \sigma_E$	1.82 ± 0.18
$\langle E_{\rm tr}\rangle \pm \sigma_{E_{\rm tr}}$	1.66 ± 0.49
$\langle E_{\rm val}\rangle \pm \sigma_{E_{\rm val}}$	1.88 ± 0.67
$\langle {\rm TL}\rangle \pm \sigma_{\rm TL}$	6927 ± 3839
$\langle \chi^{2(k)}\rangle \pm \sigma_{\chi^2}$	0.91 ± 0.12

Table 4.10: Statistical estimators for NNPDFpol1.0 with $N_{\rm rep} = 100$ replicas.

practice, we test for the condition

$$\frac{I(x_1,x_2)}{I(x_1',x_2)} < \frac{\ln\frac{x_2}{x_1}}{\ln\frac{x_2}{x_1'}}, \tag{4.28}$$

with $x_1 < x_1'$. Mutations which do not satisfy the condition are rejected during the minimization procedure. In our default fit, we chose $x_1 = 10^{-5}$, $x_1' = 2\cdot 10^{-5}$ and $x_2 = 10^{-4}$.

4.4 Results

In this Section, we present the first determination of a polarized PDF set based on the NNPDF methodology, NNPDFpol1.0. We will first illustrate the statistical features of our PDF fit, then compare the PDFs in our set to those from other recent determinations introduced in Sec. 3.3. We will also discuss the stability of our results upon the variation of several theoretical and methodological assumptions, namely the treatment of target mass corrections, the use of sum rules to fix the nonsinglet axial charges, the effect of positivity constraints on polarized PDFs, and impact of preprocessing of neural networks on small- and large-x PDF behavior.

4.4.1 Statistical features

The statistical features of the NNPDFpol1.0 analysis are summarized in Tabs. 4.10-4.11, for the full data set and for individual experiments and sets respectively. The mean value of the error function, Eq. (4.14), $\langle E\rangle$, shown in the tables both for the total, training and validation data sets is the figure of merit for the quality of the fit of each PDF replica to the corresponding data replica. The quantity which is actually minimized during the neural network training is this figure of merit for the training set, supplemented by weighting in the early stages of training according to Eq. (4.16) and by a Lagrange multiplier to enforce positivity according to Eq. (4.24). In the table we also show the average over all replicas, $\langle \chi^{2(k)}_{\rm tot}\rangle$, of $\chi^{2(k)}_{\rm tot}$ computed for the k-th replica, which coincides with the figure of merit Eq. (4.16), but with the data replica $g_I^{({\rm art})(k)}$ replaced by the experimental data $g_I^{({\rm dat})}$. We finally show $\chi^2_{\rm tot}$, which coincides with the figure of merit Eq. (4.16), but again with $g_I^{({\rm art})(k)}$ replaced by $g_I^{({\rm dat})}$, and also with $g_I^{({\rm net})(k)}$ replaced by $\langle g_I^{({\rm net})(k)}\rangle$, *i.e.* the average of the observable over replicas, which provides our best prediction. The average number of iterations of the genetic algorithm at stopping, $\langle {\rm TL}\rangle$, is also given in this table.

The distribution of $\chi^{2(k)}$, $E_{\rm tr}^{(k)}$, and training lengths among the $N_{\rm rep} = 100$ replicas are shown in Fig. 4.4 and Fig. 4.5 respectively. Note that the latter has a long tail which causes

Experiment	Set	$\chi^2_{\rm tot}$	$\langle E\rangle \pm \sigma_E$
EMC		0.44	1.54 ± 0.64
	EMC-A1P	0.44	1.54 ± 0.64
SMC		0.93	1.93 ± 0.51
	SMC-A1P	0.40	1.44 ± 0.54
	SMC-A1D	1.46	2.42 ± 0.82
SMClowx		0.97	1.90 ± 0.67
	SMClx-A1P	1.40	2.32 ± 1.13
	SMClx-A1D	0.53	1.48 ± 0.69
E143		0.64	1.68 ± 0.29
	E143-A1P	0.43	1.49 ± 0.34
	E143-A1D	0.88	1.90 ± 0.45
E154		0.40	1.69 ± 0.61
	E154-A1N	0.40	1.69 ± 0.61
E155		0.89	1.96 ± 0.36
	E155-G1P	0.89	2.00 ± 0.51
	E155-G1N	0.88	1.93 ± 0.47
COMPASS-D		0.65	1.72 ± 0.53
	CMP07-A1D	0.65	1.72 ± 0.53
COMPASS-P		1.31	2.38 ± 0.72
	CMP10-A1P	1.31	2.38 ± 0.72
HERMES97		0.34	1.37 ± 0.69
	HER97-A1N	0.34	1.37 ± 0.69
HERMES		0.79	1.79 ± 0.30
	HER-A1P	0.44	1.49 ± 0.39
	HER-A1D	1.13	2.09 ± 0.50

Table 4.11: Same as Tab. 4.10 but for individual experiments.

an accumulation of points at the maximum training length, $N_{\rm gen}^{\rm max}$. This means that there is a fraction of replicas that do not fulfill the stopping criterion. This may cause a loss in accuracy in outlier fits, which however make up fewer than 10% of the total sample.

The features of the fit can be summarized as follows:

- The quality of the central fit, as measured by its $\chi^2_{\rm tot} = 0.77$, is good. However, this value should be taken with care in view of the fact that uncertainties for all experiments but two are overestimated because the covariance matrix is not available and thus correlations between systematics cannot be properly accounted for. This explains the value lower than one for this quantity, which would be very unlikely if it had included correlations.

- The values of $\chi^2_{\rm tot}$ and $\langle E\rangle$ differ by approximately one unit. This is due to the fact that replicas fluctuate within their uncertainty about the experimental data, which in turn are Gaussianly distributed about a true value [120]: it shows that the neural network is correctly reproducing the underlying law thus being closer to the true value. This is confirmed by the fact that $\langle \chi^{2(k)}\rangle$ is of order one.

- The distribution of χ^2 for different experiments (also shown as a histogram in Fig. 4.6) shows sizable differences, and indeed the standard deviation (shown as a dashed line in the plot) about the mean (shown as a solid line) is very large. This can be understood as

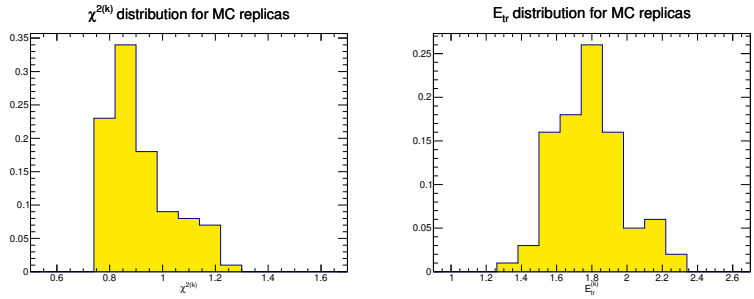

Figure 4.4: Distribution of $\chi^{2(k)}$ and $E_{\rm tr}^{(k)}$ over the sample of $N_{\rm rep} = 100$ replicas.

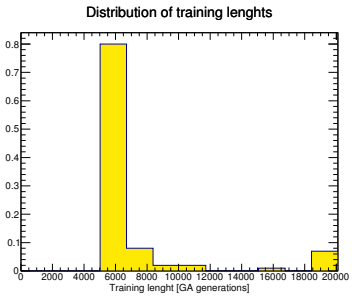

Figure 4.5: Distribution of training lengths over the sample of $N_{\rm rep} = 100$ replicas.

a consequence of the lack of information on the covariance matrix: experiments where large correlated uncertainties are treated as uncorrelated will necessarily have a smaller value of the χ^2.

The `NNPDFpol1.0` parton distributions, computed from a set of $N_{\rm rep} = 100$ replicas, are displayed in Fig. 4.7 at the input scale $Q_0^2 = 1$ GeV^2, in the PDF parametrization basis as a function of x both on a logarithmic and linear scale. In Figs. 4.8-4.9 the same PDFs are plotted in the flavor basis, and compared to other available NLO PDF sets: `BB10` [76] and `AAC08` [73] in Fig. 4.8, and `DSSV08` [75] in Fig. 4.9. We do not show a direct comparison to the `LSS10` [77] nor `JAM13` [78] PDF sets because they are not publicly available. We remind from Sec. 3.3 that all these parton determinations are based on somewhat different data sets. For instance, `BB10` contains purely DIS data and `AAC08` contains DIS data supplemented by a few high-p_T-π^0 production data from RHIC: hence they are closely comparable to our PDF determination. Instead, the `DSSV08` determination includes, on top of DIS data, polarized jet production data, and, more importantly, a large amount of semi-inclusive DIS data which in particular allow for quark-antiquark separation and a more direct handle on strangeness. In these plots, NNPDF uncertainties correspond to the nominal 1-σ error bands, while for other PDF sets they are Hessian uncertainties corresponding to the default value assumed in each analysis. We remind from Sec. 3.3 that it is assumed to be $T = 12.95$ for `AAC08`, while $T = 1$ for all other PDF sets.

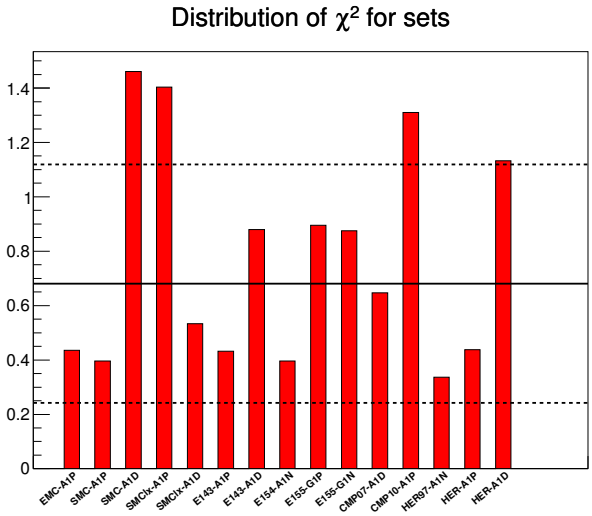

Figure 4.6: Value of the χ^2 per data point for the data sets included in the `NNPDFpol1.0` reference fit, listed in Tab. 4.11. The horizontal line is the unweighted average of these χ^2 over the data sets and the black dashed lines give the one-sigma interval about it.

The main conclusions of this comparison can be summarized as follows.

- The central values of the $\Delta u + \Delta\bar{u}$ and the $\Delta d + \Delta\bar{d}$ PDF combinations are in reasonable agreement with those of other parton sets. The `NNPDFpol1.0` results are in best agreement with `DSSV08`, in slightly worse agreement with `AAC08`, and in worst agreement with `BB10`. Uncertainties on these PDFs are generally slightly larger for `NNPDFpol1.0` than for other sets, especially `DSSV08`, which however is based on a much wider data set.

- The `NNPDFpol1.0` determination of $\Delta s + \Delta\bar{s}$ is affected by a much larger uncertainty than `BB10` and `AAC08`, for almost all values of x. Overall, the `AAC08` and `BB10` total strange distributions fall well within the `NNPDFpol1.0` uncertainty band.

- The `NNPDFpol1.0` determination of total strangeness, $\Delta s + \Delta\bar{s}$ is inconsistent at the two sigma level in the medium-to-small $x \sim 0.1$ region with `DSSV08`, which is also rather more accurate. However, we notice that total strangeness is constrained in the two analyses by rather different experimental information. In `NNPDFpol1.0`, it is determined through its Q^2 evolution at different scales together with fixing the first moments of the nonsinglet PDF combinations to the baryonic octet decay constants. Conversely, in `DSSV08` the total strangeness is mostly determined from semi-inclusive data with strange hadrons in the final states. Hence, the flavor combination $\Delta s + \Delta\bar{s}$ is also sensitive to the corresponding fragmentations functions. Since these are poorly known, especially for strange hadrons (namely kaons), the result obtained in the `DSSV08` analysis is likely to be biased by the form assumed for the fragmentation funcions.

- The gluon PDF is affected by a large uncertainty, rather larger than any other set, especially at small x. In particular, the `NNPDFpol1.0` polarized gluon distribution is

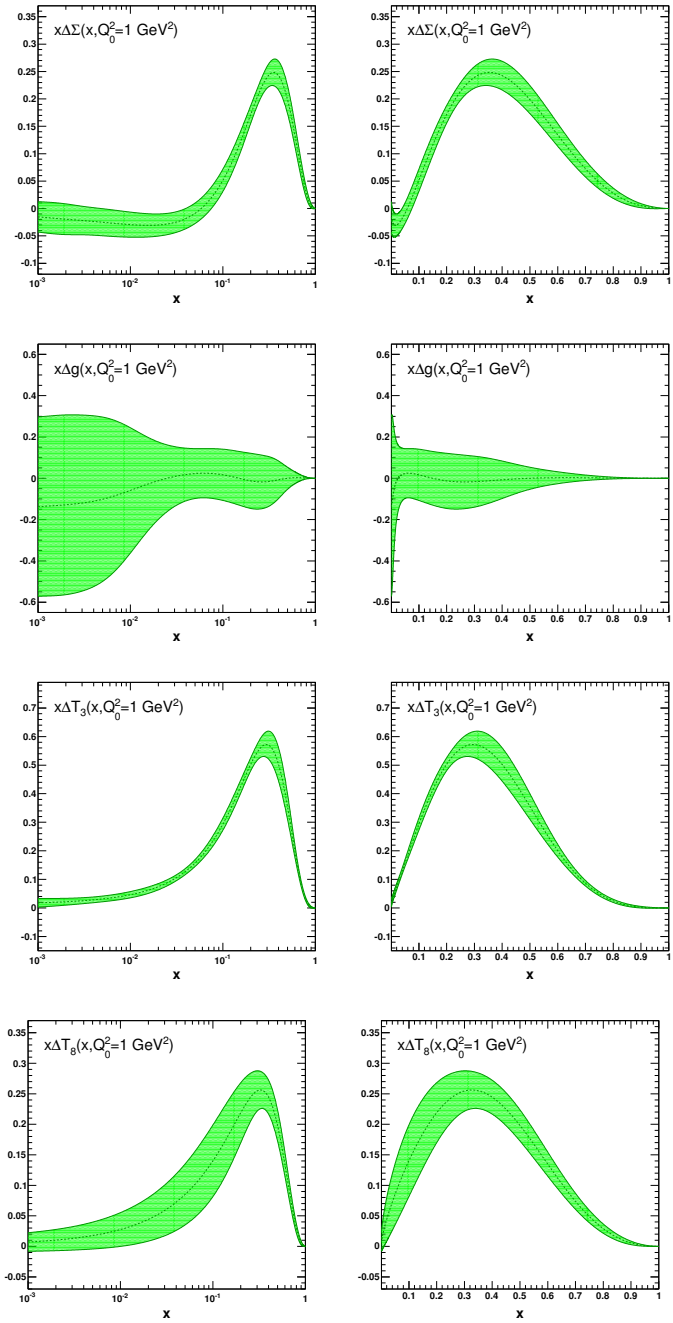

Figure 4.7: The `NNPDFpol1.0` parton distributions at $Q_0^2 = 1$ GeV^2 in the parametrization basis plotted as a function of x, on a logarithmic (left) and linear (right) scale.

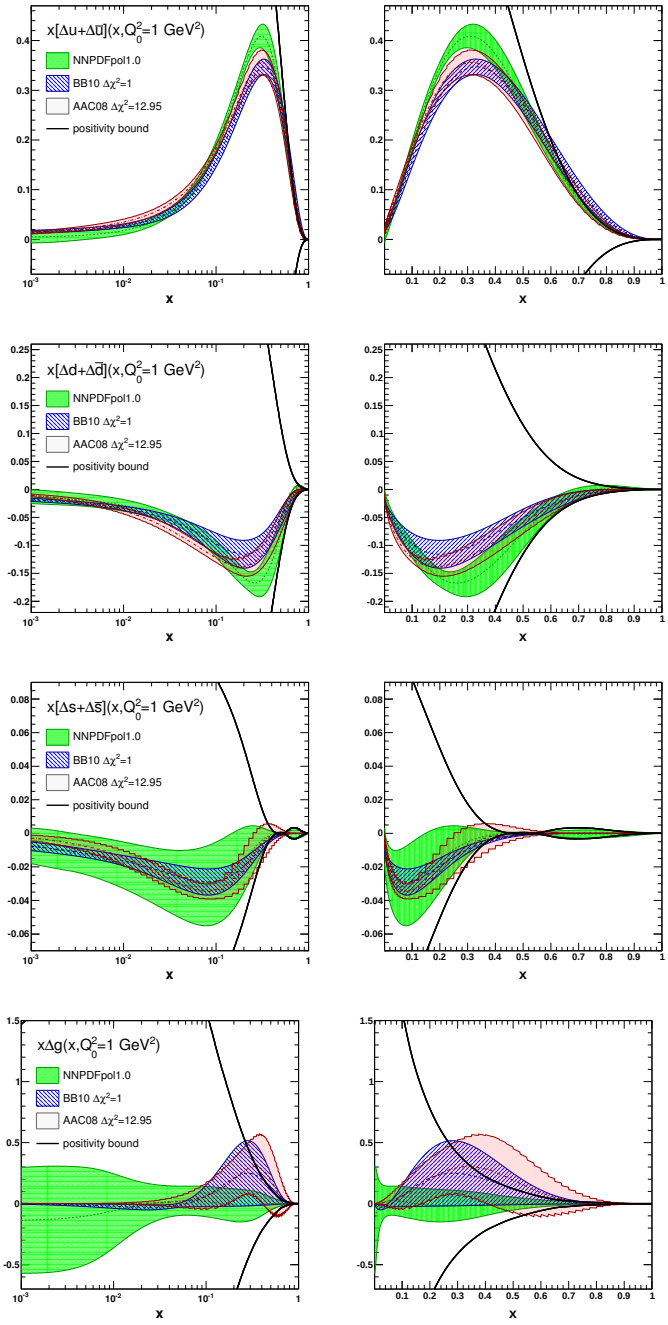

Figure 4.8: he `NNPDFpol1.0` parton distributions at $Q_0^2 = 1$ GeV^2 in the flavor basis plotted as a function of x, on a logarithmic (left) and linear (right) scale and compared to `BB10` and `AAC08` parton sets.

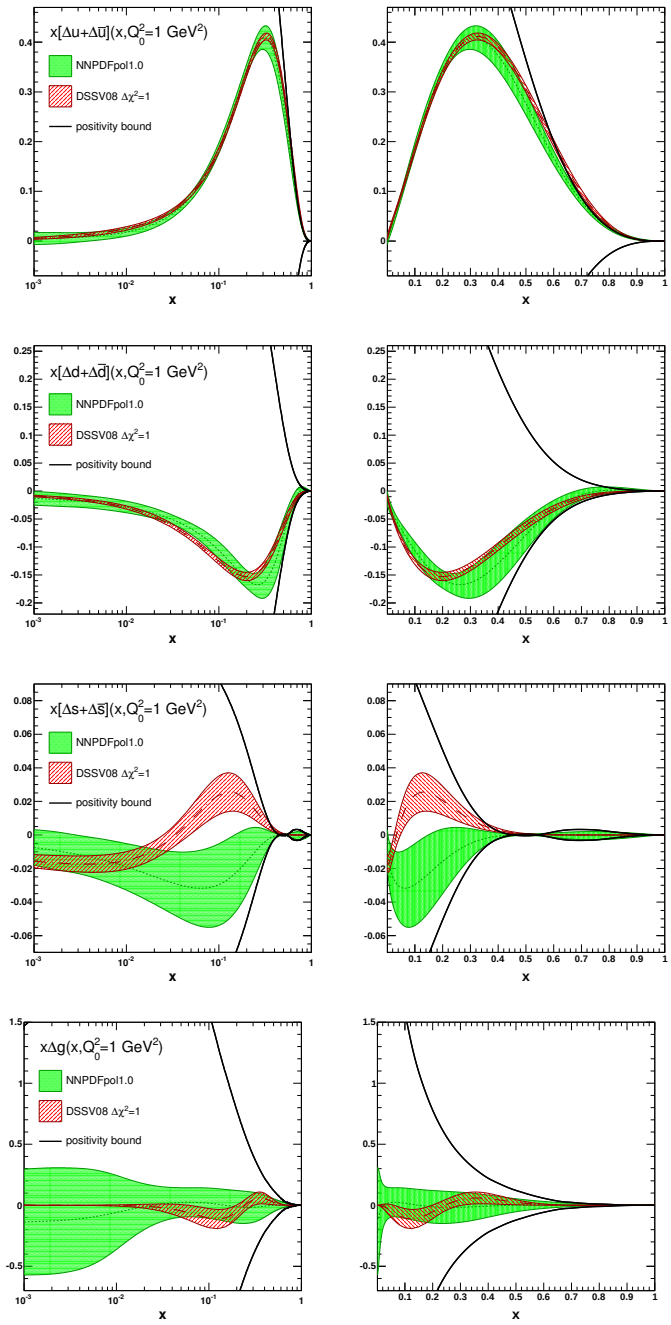

Figure 4.9: Same as Fig. 4.8, but compared to DSSV08 parton set.

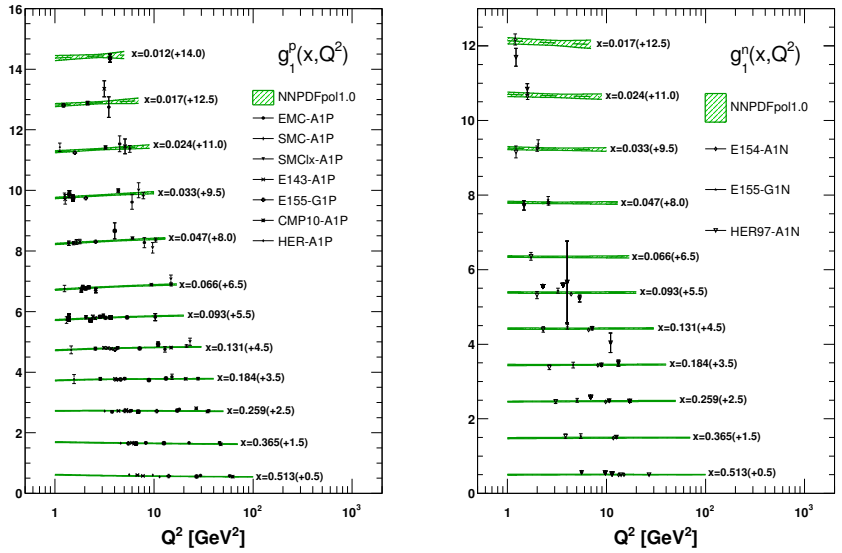

Figure 4.10: The proton, neutron and deuteron polarized structure function $g_1(x,Q^2)$ as functions of Q^2 in different bins of x compared to experimental data. Experimental data are grouped in bins of x, while `NNPDFpol1.0` results are given at the center of each bin, whose value is given next to each curve. In order to improve legibility, the values of $g_1(x,Q^2)$ have been shifted by the amount given next to each curve.

compatible with zero for all values of x. At $0.04 \lesssim x \lesssim 0.2$, the gluon determination in the `DSSV08` parton set benefits from sensitivity to pion and jet production data, which are not included in the other determinations.

- Uncertainties on the PDFs in the regions where no data are available tend to be larger than those of other sets. At very large values of x the PDF uncertainty band is largely determined by the positivity constraint, while at small values of x it is prevented to blow up arbitrarily by demanding the integrability of its first moment.

In Fig. 4.10 we compare the structure function $g_1(x,Q^2)$ for proton and neutron, computed using `NNPDFpol1.0` (with its one-sigma uncertainty band) to the experimental data included in the fit. Experimental data are grouped in bins of x with a logarithmic spacing, while the theoretical prediction and its uncertainty are computed at the central value of each bin. The uncertainty band in the `NNPDFpol1.0` result is typically smaller than the experimental errors, except at small-x where a much more restricted data set is available; in that region, the uncertainties are comparable. Scaling violations of the polarized structure functions are clearly visible, especially for g_1^p, despite the limited range in Q^2.

4.4.2 Stability of the results

Our results have been obtained with a number of theoretical and methodological assumptions, discussed in Sects. 4.2-4.3. We will now test their stability upon variation of these assumptions.

Fit	NNPDFpol1.0 $g_2 = g_2^{\rm WW}$	NNPDFpol1.0 $M = 0$	NNPDFpol1.0 $g_2 = 0$
$\chi^2_{\rm tot}$	0.77	0.78	0.75
$\langle E\rangle \pm \sigma_E$	1.82 ± 0.18	1.81 ± 0.16	1.83 ± 0.15
$\langle E_{\rm tr}\rangle \pm \sigma_{E_{\rm tr}}$	1.66 ± 0.49	1.62 ± 0.50	1.70 ± 0.38
$\langle E_{\rm val}\rangle \pm \sigma_{E_{\rm val}}$	1.88 ± 0.67	1.84 ± 0.70	1.96 ± 0.56
$\langle \chi^{2(k)}\rangle \pm \sigma_{\chi^2}$	0.91 ± 0.12	0.90 ± 0.09	0.86 ± 0.09

Table 4.12: The statistical estimators of Tab. 4.10 (obtained assuming $g_2 = g_2^{\rm WW}$) compared to a fit with $M = 0$ or with $g_2 = 0$.

Target-mass corrections and g_2.

We have consistently included in our determination of g_1 corrections suppressed by powers of the nucleon mass which are of kinematic origin. Thus in particular, we have included target-mass corrections (TMCs) up to first order in M^2/Q^2. Furthermore, both TMCs and the relation between the measured asymmetries and the structure function g_1 involve contributions to the structure function g_2 proportional to powers of M^2/Q^2 which we include according to Eq. (2.97) or Eq. (2.98) (see the discussion in Sect. 2.5). Our default PDF set is obtained assuming that g_2 is given by the Wandzura-Wilczek relation, Eq. (2.102).

In order to assess the impact of these assumptions on our results, we have performed two more PDF determinations. In the first, we set $M = 0$ consistently everywhere, both in the extraction of the structure functions from the asymmetry data and in our computation of structure functions. This thus removes TMCs, and also contributions proportional to g_2. In the second, we retain mass effects, but we assume $g_2 = 0$.

The statistical estimators for each of these three fits over the full data set are shown in Tab. 4.12. Clearly, all fits are of comparable quality.

Furthermore, in Fig. 4.11 we compare the PDFs at the initial scale Q_0^2 determined in these fits to our default set: differences are hardly visible. This comparison can be made more quantitative by using the distance $d(x,Q^2)$ between different fits, as defined in Appendix A (see also Appendix A of Ref. [104]). The distance is defined in such a way that if we compare two different samples of $N_{\rm rep}$ replicas each extracted from the same distribution, then on average $d = 1$, while if the two samples are extracted from two distributions which differ by one standard deviation, then on average $d = \sqrt{N_{\rm rep}}$ (the difference being due to the fact that the standard deviation of the mean scales as $1/\sqrt{N_{\rm rep}}$).

The distances $d(x,Q^2)$ between central values and uncertainties of the three fits of Tab. 4.12 are shown in Fig. 4.12. They never exceed $d = 4$, which means less than half a standard deviation for $N_{\rm rep} = 100$. It is interesting to observe that distances tend to be larger in the large-x region, where the expansion in powers of M^2/Q^2 is less accurate, and the effects of dynamical higher twists can become relevant. It is reassuring that even in this region the distances are reasonably small.

We conclude that inclusive DIS data, with our kinematic cuts, do not show sensitivity to finite nucleon mass effects, neither in terms of fit quality, nor in terms of the effect on PDFs.

Sum rules

Our default PDF fit is obtained by assuming that the triplet axial charge a_3 is fixed to its value extracted from β decay, Eq. (2.80), and that the octet axial charge a_8 is fixed to the value of

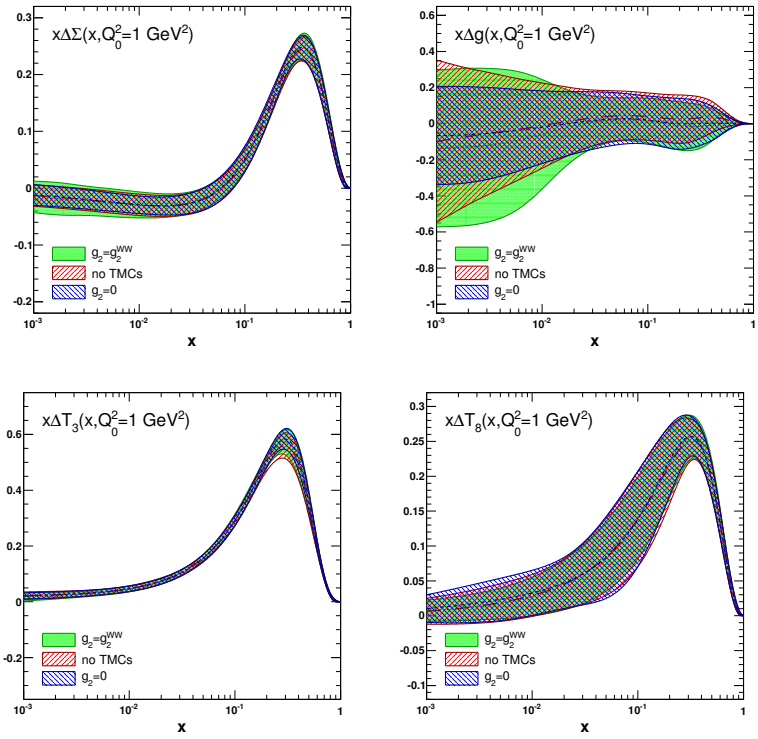

Figure 4.11: Comparison between the default `NNPDFpol1.0` PDFs (labeled as $g_2 = g_2^{\rm WW}$ in the plot), PDFs with $M = 0$ (labeled as noTMCs in the plot) and PDFs with $g_2 = 0$; each corresponds to the statistical estimators of Tab. 4.12.

Fit	free a_3	a_8 Eq. (2.80)
$\chi^2_{\rm tot}$	0.79	0.77
$\langle E\rangle \pm \sigma_E$	1.84 ± 0.19	1.86 ± 0.19
$\langle E_{\rm tr}\rangle \pm \sigma_{E_{\rm tr}}$	1.73 ± 0.41	1.66 ± 0.53
$\langle E_{\rm val}\rangle \pm \sigma_{E_{\rm val}}$	1.93 ± 0.58	1.87 ± 0.71
$\langle \chi^{2(k)}\rangle \pm \sigma_{\chi^2}$	0.93 ± 0.12	0.92 ± 0.15

Table 4.13: The statistical estimators of Tab. 4.10, but for fits in which the triplet sum rule is not imposed (free a_3) or in which the octet sum rule is imposed with the smaller uncertainty Eq. (2.80).

a_8 determined from baryon octet decays, but with an inflated uncertainty in order to allow for $SU(3)$ violation, Eq. (4.7). As discussed after Eq. (4.13) uncertainties on them are included by randomizing their values among replicas.

In order to test the impact of these assumptions, we have produced two more PDF determinations. In the first, we have not imposed the triplet sum rule, so in particular a_3 is free and determined by the data, instead of being fixed to the value Eq. (2.80). In the second, we have assumed that the uncertainty on a_8 is given by the much smaller value of Eq. (2.80).

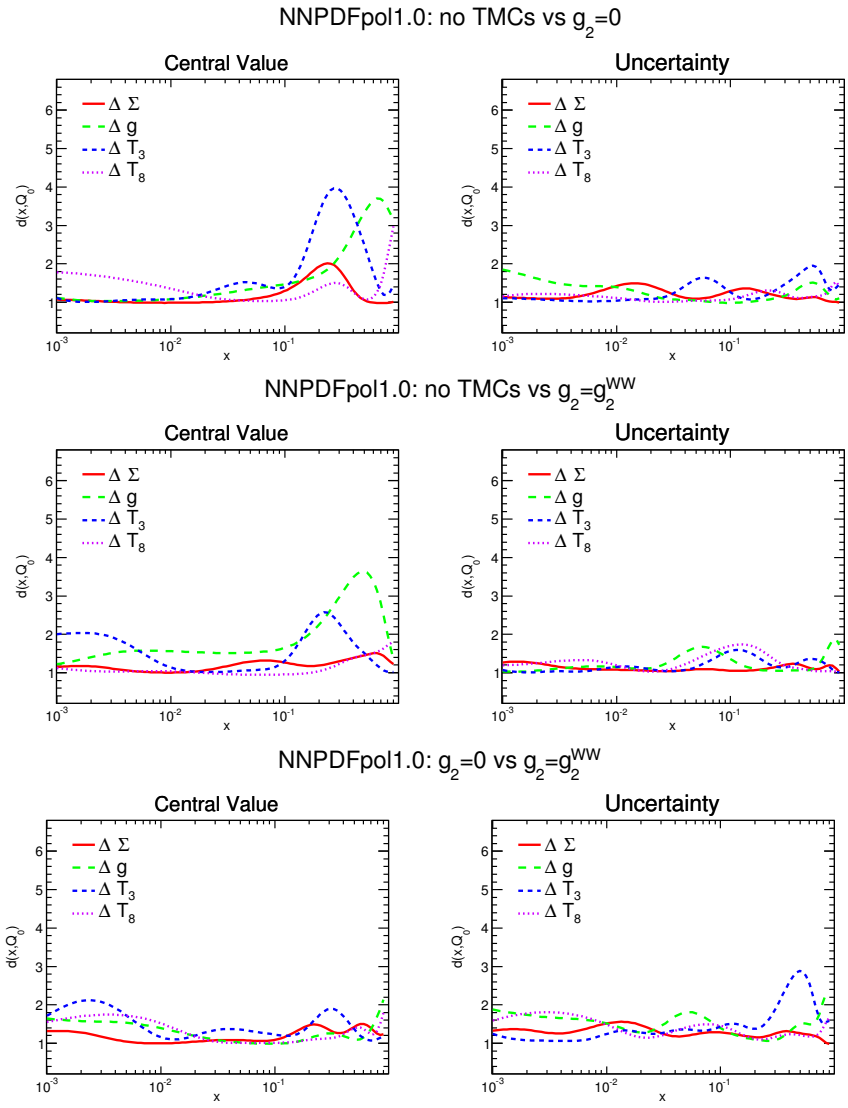

Figure 4.12: Distances between each pair of the three sets of PDFs shown in Fig. 4.11.

The statistical estimators for the total data set for each of these fits are shown in Tab. 4.13. Here too, there is no significant difference in fit quality between these fits and the default. The distances between PDFs in the default and the free a_3 fits are displayed in Fig. 4.13. As one may expect, only the triplet is affected significantly: the central value is shifted by about $d \sim 5$, *i.e.* about half-σ, in the region $x \sim 0.3$, where $x\Delta T_3$ has a maximum, and also around $x \sim 0.01$. The uncertainties on the PDFs are very similar in both cases for all PDFs, except ΔT_3 at small-x: in this case, removing the a_3 sum rule results in a moderate increase of the uncertainties; the effect of removing a_3 is otherwise negligible. The singlet and triplet PDFs for these two fits

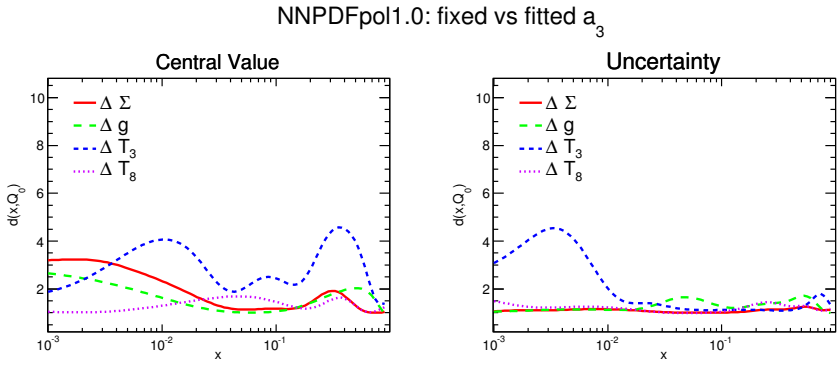

Figure 4.13: Distances between PDFs (central values and uncertainties) for the default fit, with a_3 fixed, and the fit with free a_3, computed using $N_{\rm rep} = 100$ replicas from each set.

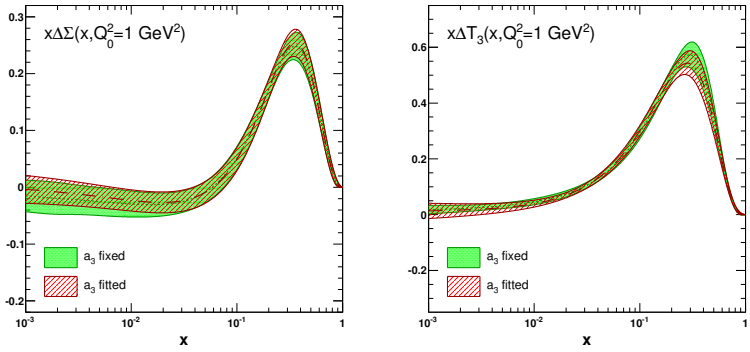

Figure 4.14: Comparison of the singlet and triplet PDFs for the default fit, with a_3 fixed, and the fit with free a_3.

are compared in Fig. 4.14.

The distances between the default and the fit with the smaller uncertainty on a_8 are shown in Fig. 4.15. In this case, again as expected, the only effect is on the ΔT_8 uncertainty, which changes in the region $10^{-2} \lesssim x \lesssim 10^{-1}$ by up to $d \sim 6$ (about half a standard deviation): if a more accurate value of a_8 is assumed, the determined ΔT_8 is correspondingly more accurate. Central values are unaffected. The singlet and octet PDFs for this fit are compared to the default in Fig. 4.16. We conclude that the size of the uncertainty on ΔT_8 has a moderate effect on our fit; on the other hand it is clear that if the octet sum rule were not imposed at all, the uncertainty on the octet and thus on strangeness would increase very significantly, as we have checked explicitly. We conclude that our fit results are quite stable upon variations of our treatment of both the triplet and the octet sum rules.

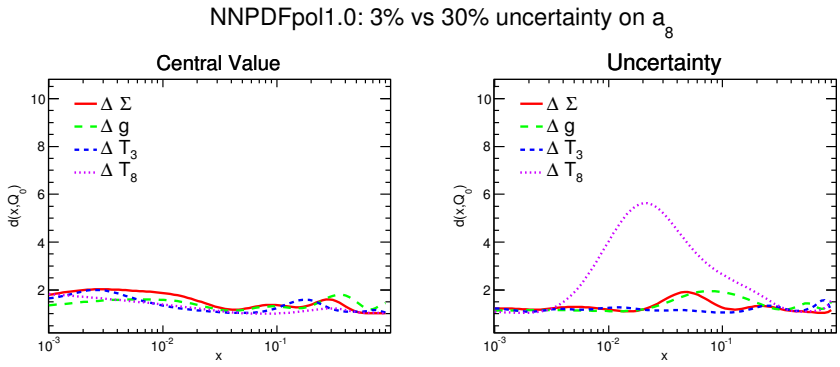

Figure 4.15: Distances between PDFs (central values and uncertainties) for the default fit, with a_8 Eq. (4.7), and the fit with the value of a_8 with smaller uncertainty, Eq. (2.80).

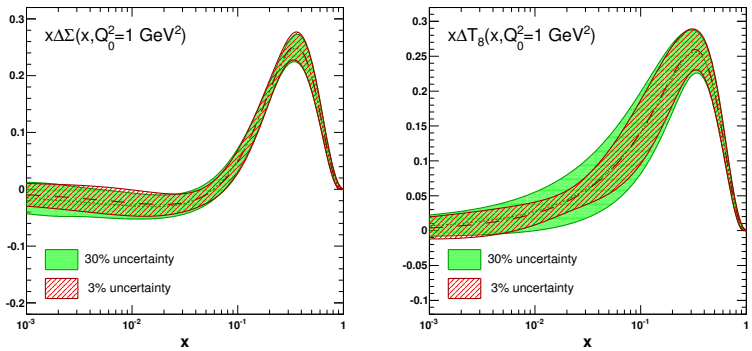

Figure 4.16: Comparison of the singlet and octet PDFs for the default fit, with a_8 Eq. (4.7), and the fit with the value of a_8 with smaller uncertainty, Eq. (2.80).

4.4.3 Positivity

As discussed in Sect. 4.3, positivity of the individual cross-sections entering the polarized asymmetries Eq. (2.83) has been imposed at leading order according to Eq. (4.23), using the NLO `NNPDF2.1` PDF set [106], separately for the lightest polarized quark PDF combinations $\Delta u + \Delta\bar{u}$, $\Delta d + \Delta\bar{d}$, $\Delta s + \Delta\bar{s}$ and for the polarized gluon PDF, by means of a Lagrange multiplier Eq. (4.24). After stopping, positivity is checked a posteriori and replicas which do not satisfy it are discarded and retrained.

In Fig. 4.17 we compare to the positivity bound for the up, down, strange PDF combinations and gluon PDF a set of $N_{\rm rep} = 100$ replicas obtained by enforcing positivity through a Lagrange multiplier, but before the final, *a posteriori* check. Almost all replicas satisfy the constraint, but at least one replica which clearly violates it for the total strangeness combination (and thus will be discarded) is seen.

In order to assess the effect of the positivity constraints, we have performed a fit without

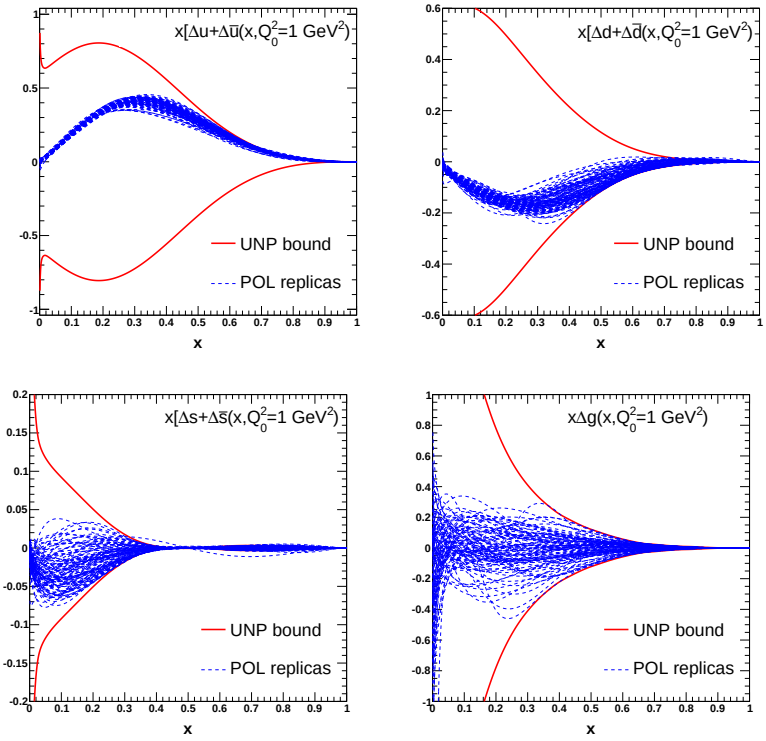

Figure 4.17: The positivity bound Eq. (4.23), compared to a set of $N_{\rm rep} = 100$ replicas (dashed lines).

PDF	m
$\Delta\Sigma(x,Q_0^2)$	$[0.5,5.0]$
$\Delta g(x,Q_0^2)$	$[0.5,5.0]$
$\Delta T_3(x,Q_0^2)$	$[0.5,4.0]$
$\Delta T_8(x,Q_0^2)$	$[0.5,6.0]$

Table 4.14: Ranges for the large-x preprocessing exponents Eq. (4.11) for the fit in which no positivity is imposed. The small-x exponents are the same as in the baseline fit Tab. 4.6.

imposing positivity. Because positivity significantly affects PDFs in the region where no data are available, and thus in particular their large-x behavior, preprocessing exponents for this PDF determination had to be determined again using the procedure described in Sect. 4.3. The values of the large x preprocessing exponents used in the fit without positivity are shown in Tab. 4.14. The small x exponents are the same as in the baseline fit, Tab. 4.6.

The corresponding estimators are shown in Tab. 4.15. Also in this case, we see no significant change in fit quality, with only a slight improvement in $\chi^2_{\rm tot}$ when the constraint is removed. This shows that our PDF parametrization is flexible enough to easily accommodate positivity. On the other hand, clearly the positivity bound has a significant impact on PDFs, especially

Fit	NNPDFpol1.0 no positivity
$\chi^2_{\rm tot}$	0.72
$\langle E\rangle \pm \sigma_E$	1.84 ± 0.22
$\langle E_{\rm tr}\rangle \pm \sigma_{E_{\rm tr}}$	1.60 ± 0.20
$\langle E_{\rm val}\rangle \pm \sigma_{E_{\rm val}}$	2.07 ± 0.39
$\langle \chi^{2(k)}\rangle \pm \sigma_{\chi^2}$	0.95 ± 0.16

Table 4.15: The statistical estimators of Tab. 4.10 for a fit without positivity constraints.

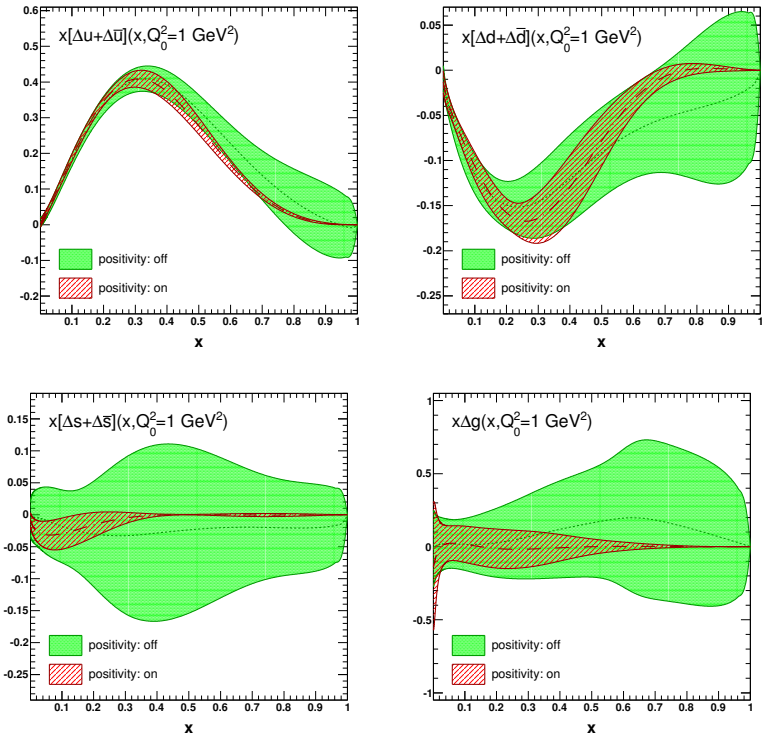

Figure 4.18: The NNPDFpol1.0 PDFs with and without positivity constraints compared at the initial parametrization scale $Q_0^2 = 1$ GeV^2 in the flavor basis.

in the large-x region, as shown in Fig. 4.18, where PDFs obtained from this fit are compared to the baseline. At small x, instead, the impact of positivity is moderate, because $g_1/F_1 \sim x$ as $x \to 0$ [206] so there is no constraint in the limit. This in particular implies that there is no significant loss of accuracy in imposing the LO positivity bound, because in the small $x \lesssim 10^{-2}$ region, where the LO and NLO positivity bounds differ significantly [207] the bound is not significant.

4.4.4 Small- and large-x behavior and preprocessing

The asymptotic behavior of both polarized and unpolarized PDFs for x close to 0 or 1 is not controlled by perturbation theory, because powers of $\ln\frac{1}{x}$ and $\ln(1-x)$ respectively appear in the perturbative coefficients, thereby spoiling the reliability of the perturbative expansion close to the endpoints. Non-perturbative effects are also expected to set in eventually (see *e.g.* [206]). For this reason, our fitting procedure makes no assumptions on the large- and small-x behaviors of PDFs, apart from the positivity and integrability constraints discussed in the previous Section.

It is however necessary to check that no bias is introduced by the preprocessing. We do this following the iterative method described in Sect. 4.3. The outcome of the procedure is the set of exponents Eq. (4.11), listed in Tab. 4.6. The lack of bias with these choices is explicitly demonstrated in Fig. 4.19, where we plot the 68% confidence level of the distribution of

$$\alpha[\Delta q(x,Q^2)] = \frac{\ln \Delta q(x,Q^2)}{\ln\frac{1}{x}}\,, \tag{4.29}$$

$$\beta[\Delta q(x,Q^2)] = \frac{\ln \Delta q(x,Q^2)}{\ln(1-x)}\,, \tag{4.30}$$

$\Delta q = \Delta\Sigma$, Δg, ΔT_3, ΔT_8, for the default `NNPDFpol1.0` $N_{\rm rep} = 100$ replica set, at $Q^2 = Q_0^2 = 1$ GeV^2, and compare them to the ranges of Tab. 4.6. It is apparent that as the endpoints $x = 0$ and $x = 1$ are approached, the uncertainties on both the small-x and the large-x exponents lie well within the range of the preprocessing exponents for all PDFs, thus confirming that the latter do not introduce any bias.

4.5 Polarized nucleon structure

We use the `NNPDFpol1.0` parton set to compute the the first moments of the polarized PDFs. These are the quantities of greatest physical interest, in that they are directly related to the spin structure of the nucleon, as discussed in Chap. 2. We also assess whether the isotriplet first moment determined within our parton set could provide an unbiased handle on the strong coupling α_s, via the Bjorken sum rule.

4.5.1 First moments

We have computed the first moments

$$\langle \Delta f(Q^2)\rangle \equiv \int_0^1 dx\, \Delta f(x,Q^2) \tag{4.31}$$

of each light polarized quark-antiquark, $\Delta u + \Delta\bar{u}$, $\Delta d + \Delta\bar{d}$, $\Delta s + \Delta\bar{s}$, and gluon, Δg, distribution using a sample of $N_{\rm rep} = 100$ `NNPDFpol1.0` PDF replicas. The histogram of the distribution of first moments over the replica sample at $Q_0^2 = 1$ GeV^2 are displayed in Fig. 4.20: they appear to be reasonably approximated by a Gaussian.

The central value and one-sigma uncertainties of the quark-antiquark combination first moments are listed in Tab. 4.16, while those of the singlet quark combination and the gluon are given in Tab. 4.17. Results are compared to those from other parton sets, namely `ABFR98` [67], `DSSV08` [75], `AAC08` [73], `BB10` [76] and `LSS10` [77]. Results from other PDF sets are not

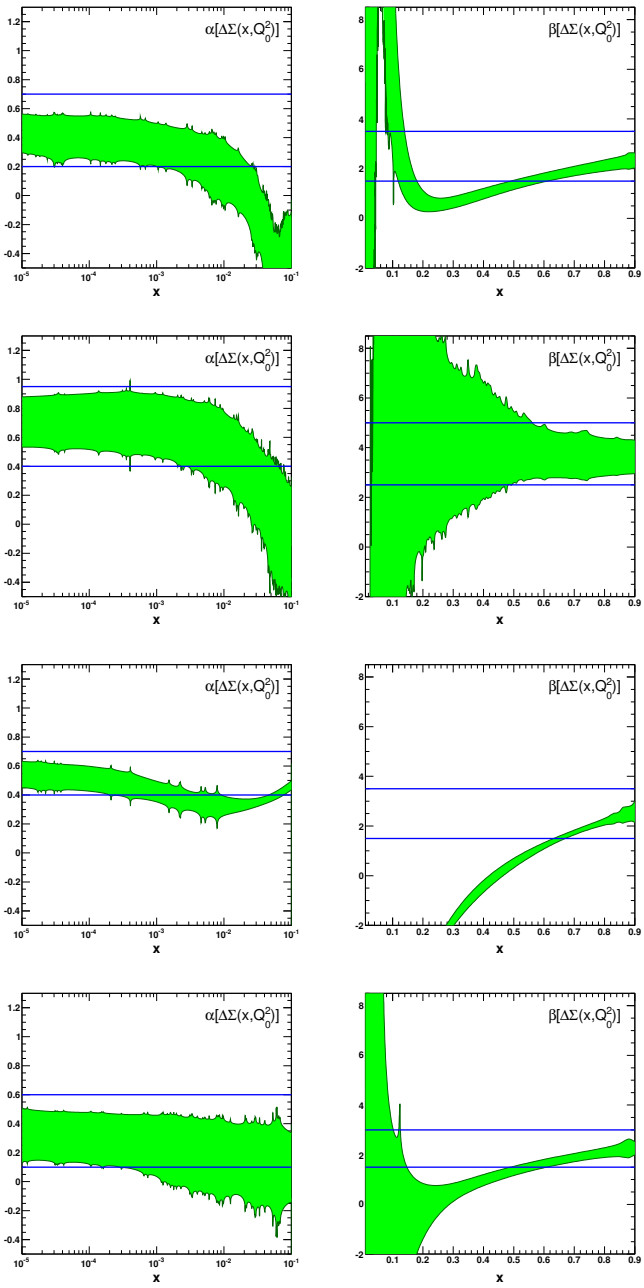

Figure 4.19: The 68% confidence level of the distribution of effective small- and large-x exponents Eqs. (4.29)-(4.30) for the default $N_{\rm rep} = 100$ replica `NNPDFpol1.0` set at $Q_0^2 = 1$ GeV^2, plotted as a functions of x. The range of variation of the preprocessing exponents of Tab. 4.6 is also shown in each case (solid lines).

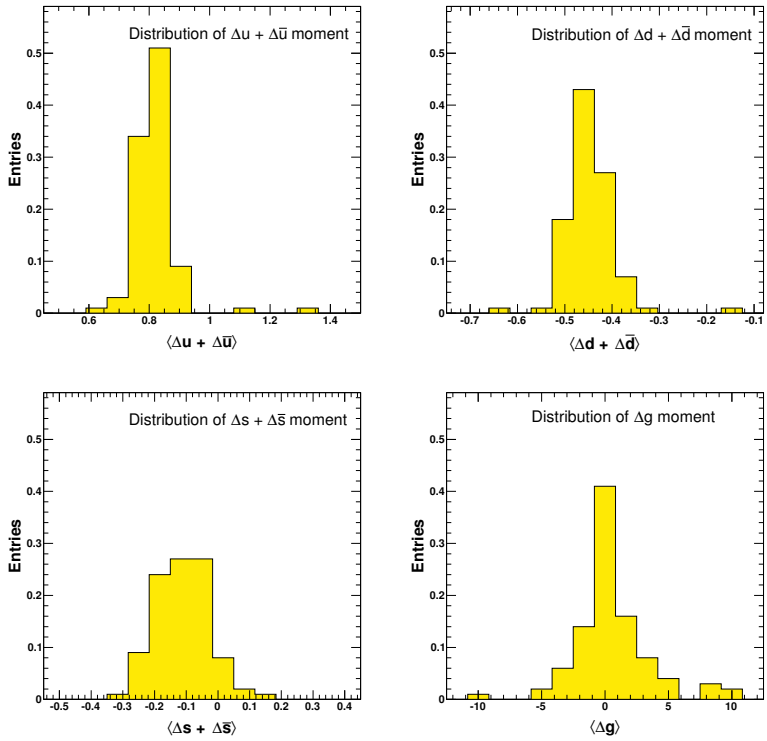

Figure 4.20: Distribution of the first moments of $\Delta u + \Delta\bar{u}$ (top left), $\Delta d + \Delta\bar{d}$ (top right), $\Delta s + \Delta\bar{s}$ (bottom left) and Δg (bottom right) over a set of $N_{\rm rep} = 100$ `NNPDFpol1.0` PDF replicas.

	$\langle\Delta u+\Delta\bar{u}\rangle$				$\langle\Delta d+\Delta\bar{d}\rangle$				$\langle\Delta s+\Delta\bar{s}\rangle$			
	cv	*exp*	*th*	*tot*	*cv*	*exp*	*th*	*tot*	*cv*	*exp*	*th*	*tot*
`NNPDFpol1.0`	0.80	0.08	—	0.08	-0.46	0.08	—	0.08	-0.13	0.09	—	0.09
`DSSV08` [74]	0.82	0.01	0.01	0.02	-0.45	0.01	0.04	0.04	-0.11	0.02	0.10	0.10

Table 4.16: First moments of the polarized quark distributions at $Q_0^2 = 1$ GeV^2; *cv* denotes the central value, while *exp* and *th* denote uncertainties (see text) whose sum in quadrature is given by *tot*.

available for all combinations and scales, because public codes only allow for the computation of first moments in a limited x range, in particular down to a minimum value of x: hence we must rely on published values for the first moments. In particular, the `DSSV08` and `AAC08` results are shown at $Q_0^2 = 1$ GeV^2, while the `BB10` and `LSS10` results are shown at $Q^2 = 4$ GeV^2. For ease of reference, in Tab. 4.17 the `NNPDFpol1.0` values for both scales are shown.

In order to compare the results for first moments shown in Tabs. 4.16-4.17, it should be understood that the uncertainties shown, and sometimes also the central values, have somewhat different meanings.

		$\langle\Delta\Sigma\rangle$				$\langle\Delta g\rangle$			
		cv	*exp*	*th*	*tot*	*cv*	*exp*	*th*	*tot*
NNPDFpol1.0	(1GeV^2)	0.22	0.20	—	0.20	-1.2	4.2	—	4.2
	(4GeV^2)	0.18	0.20	—	0.20	-0.9	3.9	—	4.2
ABFR98 [67]		0.12	0.05	$^{+0.19}_{-0.12}$	$^{+0.19}_{-0.13}$	1.6	0.4	0.8	0.9
DSSV08 [74]		0.26	0.02	0.13	0.13	-0.12	0.12	0.06	0.13
AAC08 [73]	(*positive*)	0.26	0.06	—	0.06	0.40	0.28	—	0.28
	(*node*)	0.25	0.07	—	0.07	-0.12	1.78	—	1.78
BB10 [76]		0.19	0.08	0.23	0.24	0.46	0.43	0.004	0.43
LSS10 [77]	(*positive*)	0.21	0.03	—	0.03	0.32	0.19	—	0.19
	(*node*)	0.25	0.04	—	0.04	-0.34	0.46	—	0.46

Table 4.17: Same as Tab. 4.16, but for the total singlet quark distribution and the gluon distribution. The NNPDFpol1.0 results are shown both at $Q_0^2 = 1$ GeV^2 and $Q^2 = 4$ GeV^2, the ABFR98, DSSV08 and AAC08 results are shown at $Q_0^2 = 1$ GeV^2, and the BB10 and LSS10 are shown at $Q^2 = 4$ GeV^2.

NNPDFpol1.0**.** The *exp* uncertainty, determined as the standard deviation of the replica sample, is a pure PDF uncertainty: it includes the propagation of the experimental data uncertainties and the uncertainty due to the interpolation and extrapolation.

ABFR98**.** The central values were obtained in the AB factorization scheme discussed in Sec. 2.2.2. In this scheme, the first moment of the gluon coincides with that in the $\overline{\text{MS}}$ scheme used in all other PDF fits presented here, and thus the corresponding value from Ref. [67] is shown in Tab. 4.17. Conversely, the singlet first moments in the two schemes are different, but are related by the simple relation Eq. (2.76). In Ref. [67] a value of the singlet axial charge a_0 in the limit of infinite Q^2 was also given. In the $\overline{\text{MS}}$, the singlet axial charge and the first moment of $\Delta\Sigma$ coincide (see Sec. 2.2.2), hence we have determined $\langle\Delta\Sigma\rangle$ for ABFR98 by evolving down to $Q^2 = 1$ GeV^2 the value of $a_0(\infty)$ given in Ref. [67], at NLO and with $\alpha_s(M_z) = 0.118$ [208] (the impact of the α_s uncertainty is negligible). We have checked that the same result is obtained if a_0 is computed as the appropriate linear combination of $\langle\Delta\Sigma\rangle$ in the AB scheme and the first moment of Δg, Eq. (2.72). In the ABFR98 study, the *exp* uncertainty is the Hessian uncertainty on the best fit, and it thus includes the propagated data uncertainty. The *th* uncertainty includes the uncertainty originated by neglected higher orders (estimated by renormalization and factorization scale variations), higher twists, position of heavy quark thresholds, value of the strong coupling, violation of SU(3), and finally uncertainties related to the choice of functional form, estimated by varying the functional form. This latter source of theoretical uncertainty corresponds to interpolation and extrapolation uncertainties which are included in the *exp* for NNPDFpol1.0.

DSSV08**,** BB10**.** The central value is obtained by computing the first moment integral of the best-fit with a fixed functional form restricted to the data region, and then supplementing it with a contribution due to the extrapolation in the unmeasured, small-x, region. The *exp* uncertainty in the table is the Hessian uncertainty given by DSSV08 or BB10 on the moment in the measured region, and it thus includes the propagated data uncertainty. In both cases, we have determined the *th* uncertainty shown in the table as the difference between the full first moment quoted by DSSV08 or BB10, and the first moment in the measured region. It is thus the contribution

from the extrapolation region, which we assume to be 100% uncertain. In both cases, we have computed the truncated first moment in the measured region using publicly available codes, and checked that it coincides with the values quoted by DSSV08 and BB10.

AAC08. The central value is obtained by computing the first moment integral of the best-fit with a fixed functional form, and the *exp* uncertainty is the Hessian uncertainty on it. However, AAC08 uses the tolerance [165] criterion for the determination of Hessian uncertainties, which rescales the $\Delta\chi^2 = 1$ region by a suitable factor, in order to effectively keep into account also interpolation errors. Hence, the *exp* uncertainties include propagated data uncertainties, as well as uncertainties on the PDF shape.

LSS10. The central value is obtained by computing the first moment of the best fit with a fixed functional form, and the *exp* uncertainty is the Hessian uncertainty on it. Hence it includes the propagated data uncertainty.

In all cases, the total uncertainty is computed as the sum in quadrature of the *exp* and *th* uncertainties. Roughly speaking, for LSS10 this includes only the data uncertainties; for DSSV08, and BB10 it also includes extrapolation uncertainties; for AAC08 interpolation uncertainties; for NNPDFpol1.0 both extrapolation and interpolation uncertainties; and for ABFR98 all of the above, but also theoretical (QCD) uncertainties. For LSS10 and AAC08, we quote the results obtained from two different fits, both assuming positive- or node-gluon PDF: their spread gives a feeling for the missing uncertainty due to the choice of functional form. Remind that the AAC08 results correspond to their Set B which includes, besides DIS data, also RHIC π^0 production data; the DSSV08 fit also includes, on top of these, RHIC jet data and semi-inclusive DIS data; LSS10 includes, beside DIS, also semi-inclusive DIS data. All other sets are based on DIS data only.

Coming now to a comparison of results, we see that for the singlet first moment $\langle\Delta\Sigma\rangle$ the NNPDFpol1.0 result is consistent within uncertainties with that of other groups. The uncertainty on the NNPDFpol1.0 result is comparable, if somewhat larger, to that found whenever the extrapolation uncertainty has been included. For individual quark flavors we find excellent agreement in the central values obtained between NNPDFpol1.0 and DSSV08, see Tab. 4.16; the NNPDFpol1.0 uncertainties are rather larger, but this could also be due to the fact that the data set included in DSSV08 is sensitive to quark-antiquark separation.

For the gluon first moment $\langle\Delta g\rangle$, the NNPDFpol1.0 result is characterized by an uncertainty which is much larger than that of any other determination: a factor of three or four larger than ABFR98 and AAC08, ten times larger than BB10, and twenty times larger than DSSV08 and LSS10. It is compatible with zero within this large uncertainty. We have seen that for the quark singlet, the NNPDFpol1.0 uncertainty is similar to that of groups which include an estimate of extrapolation uncertainties. In order to assess the impact of the extrapolation uncertainty for the gluon, we have computed the gluon first moment truncated in the region $x \in [10^{-3}, 1]$:

$$\int_{10^{-3}}^{1} dx \Delta g(x, Q^2 = 1\text{GeV}^2) = -0.26 \pm 1.19\,, \tag{4.32}$$

to be compared with the result of Tab. 4.17, which is larger by almost a factor four.

We must conclude that the experimental status of the gluon first moment is still completely uncertain, unless one is willing to make strong theoretical assumptions on the behavior of the polarized gluon at small x, and that previous different conclusions were affected by a

significant underestimate of the impact of the bias in the choice of functional form, in the data and especially in the extrapolation region. Because of the large uncertainty related to the extrapolation region, only low-x data can improve this situation, such as those which could be collected at a high energy Electron-Ion Collider [128, 209], as we will show in Chap. 5.

4.5.2 The Bjorken sum rule

The Bjorken sum rule presented in Sec. 2.3, Eq. (2.82)

$$\Gamma_1^{\rm NS} \equiv \Gamma_1^p\left(Q^2\right) - \Gamma_1^n\left(Q^2\right) = \frac{1}{6}\Delta C_{\rm NS}(\alpha_s(Q^2))a_3 \tag{4.33}$$

potentially provides a theoretically very accurate handle on the strong coupling α_s. We recall that we have defined the first moment of the proton (neutron) structure function $\Gamma_1^{p,n}(Q^2)$ in Eq. (2.79), the first moment of the nonsinglet triplet PDF combination in the first relation of Eqs. (2.75) and $\Delta C_{\rm NS}(\alpha_s(Q^2))$ is the first moment of the corresponding coefficient function, which is known up to three loops. In principle, the truncated isotriplet first moment

$$\Gamma_1^{\rm NS}\left(Q^2, x_{\rm min}\right) \equiv \int_{x_{\rm min}}^1 dx\left[g_1^p\left(x,Q^2\right) - g_1^n\left(x,Q^2\right)\right] \tag{4.34}$$

can be extracted from data without any theoretical assumption. Given a measurement of $\Gamma_1^{\rm NS}\left(Q^2,0\right)$ at a certain scale, the strong coupling can then be extracted from Eq. (4.33) using the value of a_3 from β decays, while given a measurement of $\Gamma_1^{\rm NS}\left(Q^2,0\right)$ at two scales, both a_3 and the value of α_s can be extracted simultaneously.

In Ref. [210], a_3 and α_s where simultaneously determined from a set of nonsinglet truncated moments, both the first and higher moments, by exploiting the scale dependence of the latter [211], with the result $a_3 = 1.04 \pm 0.13$ and $\alpha_s(M_z) = 0.126^{+0.006}_{-0.014}$, where the uncertainty is dominated by the data, interpolation and extrapolation, but also includes theoretical QCD uncertainties. In this reference, truncated moments were determined from a neural network interpolation of existing data, sufficient for a computation of moments at any scale. However, because the small-x behavior of the structure function is only weakly constrained by data, the $x \to 0$ extrapolation was done by assuming a powerlike Regge behavior [212].

The situation within `NNPDFpol1.0` can be understood by exploiting the PDF determination in which a_3 is not fixed by the triplet sum rule, discussed in Sec. 4.4.2. Using the results of this determination, we find

$$a_3 = \int_0^1 dx \Delta T_3(x,Q^2) = 1.19 \pm 0.22\,. \tag{4.35}$$

The uncertainty is about twice that of the determination of Ref. [210]. As mentioned, the latter was obtained from a neural network parametrization of the data with no theoretical assumptions, and based on a methodology which is quite close to that of the `NNPDFpol1.0` PDF determination discussed here, the only difference being the assumption of Regge behavior in order to perform the small-x extrapolation. This strongly suggests that, as in the case of the gluon distribution discussed above, the uncertainty on the value Eq. (4.35) is dominated by the small-x extrapolation.

To study this effect, in Fig. 4.21 we plot the value of the truncated Bjorken sum rule $\Gamma_1^{\rm NS}\left(Q^2, x_{\rm min}\right)$ Eq. (4.34) as a function of the lower limit of integration $x_{\rm min}$ at $Q_0^2 = 1$ GeV^2, along with the asymptotic value

$$\Gamma_1^{\rm NS}\left(1{\rm GeV}^2, 0\right) = 0.16 \pm 0.03 \tag{4.36}$$

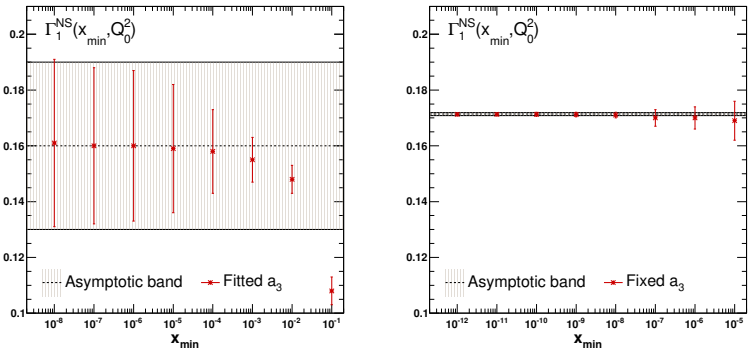

Figure 4.21: The truncated Bjorken sum rule $\Gamma_1^{\rm NS}\left(Q^2,x\right)$ Eq. (4.34) plotted as a function of x for $Q^2 = 1$ GeV^2, for the fit with free a_3 (left) and for the reference fit with a_3 fixed to the value Eq. (2.80) (right). In the left plot, the shaded band corresponds to the asymptotic value of the truncated sum rule, Eq. (4.36), while in the right plot it corresponds to the experimental value Eq. (2.80).

which at NLO corresponds to the value of a_3 given by Eq. (4.35). As a consistency check, we also show the same plot for our baseline fit, in which a_3 is fixed by the sum rule to the value Eq. (2.80). It is clear that indeed the uncertainty is completely dominated by the small x extrapolation.

We conclude that a determination of α_s from the Bjorken sum rule is not competitive unless one is willing to make assumptions on the small x behavior of the nonsinglet structure function in the unmeasured region. Indeed, it is clear that a determination based on `NNPDFpol1.0` would be affected by an uncertainty which is necessarily larger than that found in Ref. [210], which is already not competitive. The fact that a determination of α_s from the Bjorken sum rule is not competitive due to small x extrapolation ambiguities was already pointed out in Ref. [67], where values of a_3 and α_s similar to those of Ref. [210] were obtained.

5 Polarized Parton Distributions at an Electron-Ion Collider

In this Chapter we investigate the potential impact of inclusive DIS data from a future Electron-Ion Collider (EIC) on the determination of polarized parton distributions. After briefly motivating our study in Sec. 5.1, we illustrate in Sec. 5.2 which EIC pseudodata sets we use in our analysis and in Sec. 5.3 how the fitting procedure described in Sec. 4.3 needs to be optimized. Resulting PDFs are presented in Sec. 5.4, and they are compared to `NNPDFpol1.0` throughout. Finally, in Sec. 5.5 we reassess the computation of their first moments and we give an estimate of the charm contribution to the g_1 structure function. The analysis presented in this Chapter is mostly based on Ref. [117].

5.1 Motivation

As already noticed several times in this Thesis, the bulk of experimental information on longitudinally polarized proton structure comes from inclusive neutral-current DIS, which allows one to obtain information on the light quark-antiquark combinations $\Delta u + \Delta\bar{u}$, $\Delta d + \Delta\bar{d}$, $\Delta s + \Delta\bar{s}$ and on the gluon distribution Δg. However, presently available DIS data covers only a small kinematic region of momentum fractions and energies (x, Q^2), as shown in fig. 4.1. On the one hand, the lack of experimental information for $x \lesssim 10^{-3}$ prevents a reliable determination of polarized PDFs at small-x. Hence, their first moments will strongly depend on the functional form one assumes for PDF extrapolation to the unmeasured x region. On the other hand, the gluon PDF, which is determined by scaling violations, is only weakly constrained, due to the small lever-arm in Q^2 of the experimental data. Both these limitations were emphasized in Chap. 4, when we have presented the first unbiased set of polarized PDFs, `NNPDFpol1.0`.

For these reasons, despite many efforts, both experimental and theoretical, the size of the polarized gluon contribution to the nucleon spin is still largely uncertain, as demonstrated in Sec. 4.5 and in Ref. [187]. In Sec. 3.1, we mentioned that other processes, receiving leading partonic contributions from gluon-initiated suprocesses, may provide direct information on the polarized gluon PDF. They include open-charm photoproduction data from COMPASS [213] and polarized hadron collider measurements from RHIC [214–218], specifically semi-inclusive

particle and jet production data. We explicitely assess their impact on the determination of Δg in Chap. 6, but we note here that all these data are restricted to the medium- and large-x region.

An Electron-Ion Collider (EIC) [128, 209, 219], with polarized lepton and hadron beams, would allow for a widening of the kinematic region comparable to the one achieved in the unpolarized case with the DESY-HERA experiments H1 and ZEUS [220]. Note that a Large Hadron-electron Collider (LHeC) [221] would not have the option of polarizing the hadron beam. The potential impact of the EIC on the knowledge of the nucleon longitudinal spin structure has been quantitatively assessed in a recent study [222], in which projected neutral-current inclusive DIS and semi-inclusive DIS (SIDIS) artificial data were added to the `DSSV+` polarized PDF determination [187]; this study was then extended by also providing an estimate of the impact of charged-current inclusive DIS pseudo-data on the polarized quark-antiquark separation in Ref. [223]. In view of the fact that a substantially larger gluon uncertainty is found in `NNPDFpol1.0` in comparison to previous PDF determinations [73, 75–77], it is worth repeating the study of the impact of EIC data, but now using NNPDF methodology. This is the goal of the study in the present Chapter.

5.2 Inclusive DIS pseudodata from an Electron-Ion Collider

The realization of an EIC has been proposed for two independent designs so far: the electron Relativistic Heavy Ion Collider (eRHIC) at Brookhaven National Laboratory (BNL) [224] and the Electron Light Ion Collider (ELIC) at Jefferson Laboratory (JLab) [225]. In both cases, a staged upgrade of the existing facilities has been planned [128, 209, 219], so that an increased center-of-mass energy would be available at each stage. Concerning the eRHIC option of an EIC [224], first measurements would be taken by colliding the present RHIC proton beam of energy $E_p = 100 - 250$ GeV with an electron beam of energy $E_e = 5$ GeV, while a later stage envisages electron beams with energy up to $E_e = 20$ GeV.

In order to quantitatively assess the impact of future EIC measurements on the determination of polarized PDFs, we have supplemented our QCD analysis presented in Chap. 4 and Ref. [116] with DIS pseudodata from Ref. [222]. They consist of three sets of data points at different possible eRHIC electron and proton beam energies, as discussed above. These pseudodata were produced by running the `PEPSI` Monte Carlo (MC) generator [226], assuming momentum transfer $Q^2 > 1$ GeV^2, squared invariant mass of the virtual photon-proton system $W^2 > 10$ GeV^2 and fractional energy of the virtual photon $0.01 \leq y \leq 0.95$; they are provided in five (four) bins per logarithmic decade in x (Q^2).

For each data set, the Q^2 range spans the values from $Q^2_{\rm min} = 1.39$ GeV^2 to $Q^2_{\rm max} = 781.2$ GeV^2, while the accessible values of momentum fraction $x = Q^2/(sy)$ depend on the available center-of-mass energy, $\sqrt{s}$. In Tab. 5.1, we summarize, for each data set, the number of pseudodata $N_{\rm dat}$, the electron and proton beam energies E_e, E_p, the corresponding center-of-mass energies $\sqrt{s}$, and the smallest and largest accessible value in the momentum fraction range, $x_{\rm min}$ and $x_{\rm max}$ respectively.

The kinematic coverage of the EIC pseudodata is displayed in Fig. 5.1 together with the fixed-target DIS data points dscussed in Chap. 4. The dashed regions show the overall kinematic reach of the EIC data with the two electron beam energies $E_e = 5$ GeV or $E_e = 20$ GeV, corresponding to each of the two stages at eRHIC. It is apparent from Fig. 5.1 that EIC data will extend the kinematic coverage significantly, even for the lowest center-of-mass energy. In particular, hitherto unreachable small-x values, down to 10^{-4}, will be attained, thereby leading to a significant reduction of the uncertainty in the low-x extrapolation region.

Experiment	Set	$N_{\rm dat}$	$E_e \times E_p$ [GeV]	$\sqrt{s}$ [GeV]	$x_{\rm min}$	$x_{\rm max}$	$\langle \delta g_1 \rangle$
EIC	EIC-G1P-1	56	5×100	44.7	8.2×10^{-4}	0.51	0.010
	EIC-G1P-2	63	5×250	70.7	3.2×10^{-4}	0.51	0.032
	EIC-G1P-3	61	20×250	141	8.2×10^{-5}	0.32	0.042

Table 5.1: The three EIC pseudodata sets [222]. For each set we show the number of points $N_{\rm dat}$, the electron and proton beam energies E_e and E_p, the center-of-mass energy $\sqrt{s}$, the kinematic coverage in the momentum fraction x, and the average absolute statistical uncertainty $\langle \delta g_1 \rangle$.

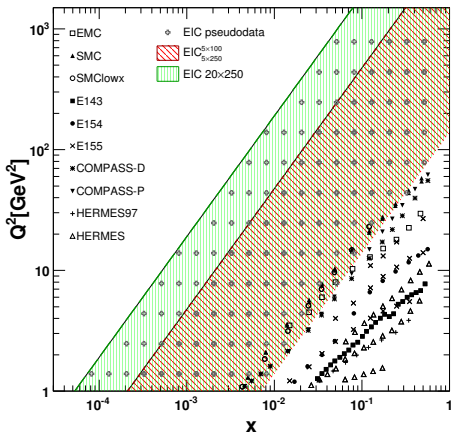

Figure 5.1: Kinematic coverage in the (x, Q^2) plane for the fixed-target experimental data included in the `NNPDFpol1.0` polarized parton fit [116] and the EIC pseudodata from [222]. The shaded bands show the expected kinematic reach of each of the two EIC scenarios discussed in the text.

Furthermore, the increased lever-arm in Q^2, for almost all values of x should allow for much more stringent constraints on $\Delta g(x, Q^2)$ from scaling violations.

The ratio $g_1(x, Q^2)/F_1(x, Q^2)$ is provided in Ref. [222] as the inclusive DIS observable, whose relation with the experimentally measured asymmetries was discussed in Sec. 2.4. The generation of pseudodata assumes a "true" underlying set of parton distributions: in Ref. [222] these are taken to be `DSSV+` [187] and `MRST` [163] polarized and unpolarized PDFs respectively. Uncertainties are then determined assuming an integrated luminosity of 10 fb^{-1}, which corresponds to a few months operations for the anticipated luminosities for eRHIC [224], and a 70% beam polarization. Because the `DSSV+` polarized gluon has rather more structure than that of `NNPDFpol1.0`, which is largely compatible with zero, assuming this input shape will allow us to test whether the EIC data are sufficiently accurate to determine the shape of the gluon distribution.

We reconstruct the g_1 polarized structure function from the pseudodata following the same procedure used in Sec. 4.1 for the E155 experiment. We provide its average statistical uncertainty in the last column of Tab. 5.1. A comparison of these values with the analogous quantities for fixed-target experiments (see Tab. 4.2 in Sec. 4.1) clearly shows that EIC data are expected to be far more precise, with uncertainties reduced up to one order of magnitude. No

Fit	$N_{\rm gen}^{\rm max}$	$N_{\rm mut}^{\rm gen}$	$N_{\rm mut}^{\rm a}$	$N_{\rm mut}^{\rm b}$	$N_{\rm gen}^{\rm wt}$	$N_{\rm smear}$	$\Delta_{\rm smear}$
NNPDFpolEC-A	20000	2000	80	30	5000	200	200
NNPDFpolEC-B	50000	2000	80	30	10000	200	200

Table 5.2: Values of the minimization and stopping parameters entering the fitting algorithm. The corresponding values used in the NNPDFpol1.0 analysis are quoted in Tabs. 4.8-4.9.

information on the expected systematic uncertainties is available, hence we will ignore them in our present analysis. However, we notice that the projected statistical uncertainties set the scale at which one needs to control systematics, which arise from luminosity and polarization measurements, detector acceptance and resolution, and QED radiative corrections.

We will perform two different fits, corresponding to the two stages envisaged for the eRHIC option of an EIC [224] discussed above, which will be referred to as NNPDFpolEIC-A and NNPDFpolEIC-B. The former includes the first two sets of pseudodata listed in Tab. 5.1, while the latter also includes the third set.

5.3 Fit optimization

The methodology for the determination of polarized PDFs, including their parametrization in terms of neural networks and their minimization through a genetic algorithm, follows the one discussed in detail in Sec. 4.3. However, due to the accuracy and the kinematic coverage of EIC pseudodata, which are respectively higher and wider in comparison to their fixed-target counterparts, the parameters entering the genetic algorithm and determining its stopping had to be re-tuned.

In particular, in order to allow the genetic algorithm to explore the space of parameters more efficiently, we have used a large population of mutants and increased the number of weighted training generations to ensure all data sets are learnt with comparable accuracy. The target values of the figure of merit used in the weighted training formula, Eq. (4.16), were consistently determined for each one of the three EIC pseudodata sets, following the iterative procedure discussed in Sec. 4.3. As for the stopping criterion, we have modified the values of the width of the moving average $N_{\rm smear}$ and the smearing parameter $\Delta_{\rm smear}$; for the NNPDFpolEIC-B fit, we have also increased the maximum number of genetic algorithm generations at which the minimization stops if the stopping criterion is not fulfilled. Also, equal training and validation fractions were chosen for pseudodata sets, unlike their fixed-target DIS counterparts. Indeed, we checked that the EIC pseudodata set size (about fifty points per set) is large enough to ensure fit stability. The values of the minimization and stopping parameters used in the NNPDFpolEIC-A and NNPDFpolEIC-B determinations are collected in Tab. 5.2: they can be compared straightforwardly with those used in the NNPDFpol1.0 analysis, see Tabs. 4.8-4.9.

Furthermore, we have redetermined the range in which preprocessing exponents are randomized, since the new information from EIC pseudodata may modify the large- and small-x PDF behavior. In Tab. 5.3, we show the values we use for the present fit, which can be compared to NNPDFpol1.0 from Tab. 4.6. We have checked that our choice of preprocessing exponents does not bias our fit, according to the procedure discussed in Sec. 4.3.

PDF	m	n
$\Delta\Sigma(x,Q_0^2)$	$[1.5,3.5]$	$[0.1,0.7]$
$\Delta g(x,Q_0^2)$	$[2.0,4.0]$	$[0.1,0.8]$
$\Delta T_3(x,Q_0^2)$	$[1.5,3.0]$	$[0.1,0.6]$
$\Delta T_8(x,Q_0^2)$	$[1.5,3.0]$	$[0.1,0.6]$

Table 5.3: Ranges for the small- and large-x preprocessing exponents.

	NNPDFpolEIC-A	NNPDFpolEIC-B
$\chi^2_{\rm tot}$	0.79	0.86
$\langle E\rangle \pm \sigma_E$	2.24 ± 0.34	2.44 ± 0.31
$\langle E_{\rm tr}\rangle \pm \sigma_{E_{\rm tr}}$	1.87 ± 0.54	1.81 ± 0.79
$\langle E_{\rm val}\rangle \pm \sigma_{E_{\rm val}}$	2.61 ± 1.05	2.47 ± 1.17
$\langle \chi^{2(k)}\rangle \pm \sigma_{\chi^2}$	1.30 ± 0.31	1.50 ± 0.30
$\langle {\rm TL}\rangle \pm \sigma_{\rm TL}$	7467 ± 3678	19320 ± 14625

Table 5.4: Statistical estimators and average training length for the NNPDFpolEIC-A and NNPDFpolEIC-B with $N_{\rm rep} = 100$ replicas. The corresponding estimators for NNPDFpol1.0 are quoted in Tab. 4.10.

5.4 Results

We now present our polarized parton sets based on inclusive DIS pseudodata at an EIC discussed in Sec. 5.2, NNPDFpolEIC-A and NNPDFpolEIC-B. First, we discuss their statistical features, then we show the corresponding parton distributons, compared to NNPDFpol1.0. All results presented in this section are obtained out of PDF ensembles of $N_{\rm rep} = 100$ replicas.

5.4.1 Statistical features

Various general features of the NNPDFpolEIC-A and NNPDFpolEIC-B PDF determinations are summarized in Tab. 5.4, and can be straightforwardly compared to NNPDFpol1.0, see Tab. 4.10. These include: the χ^2 per data point of the final best-fit PDF set compared to data, (denoted as $\chi^2_{\rm tot}$); the average and standard deviation over the replica sample of the same figure of merit for each replica when compared to the corresponding data replica (denoted as $\langle E\rangle \pm \sigma_E$) computed for the total, training and validation sets; the average and standard deviation of the χ^2 of each replica when compared to data (denoted as $\langle \chi^{2(k)}\rangle$); and the average number of iterations of the genetic algorithm at stopping $\langle {\rm TL}\rangle$ and its standard deviation over the replica sample. All these estimators were introduced in Sec. 4.4.1 and are discussed in detail in Refs. [101, 104]. The distributions of $\chi^{2(k)}$, $E_{\rm tr}^{(k)}$ and training lenghts among the $N_{\rm rep} = 100$ replicas are shown in Fig. 5.2 and Fig. 5.3. respectively. As for the training lenghts, notice the different scale on the horizontal axis for the NNPDFpolEIC-A and NNPDFpolEIC-B fits, consistent with the different maximum number of training generations $N_{\rm gen}^{\rm max}$ (see Tab. 5.2).

The fit quality, as measured by $\chi^2_{\rm tot}$, is comparable to that of NNPDFpol1.0 ($\chi^2_{\rm tot} = 0.77$) for both the NNPDFpolEIC-A ($\chi^2_{\rm tot} = 0.79$) and the NNPDFpolEIC-B ($\chi^2_{\rm tot} = 0.86$) fits. This shows that our fitting procedure can easily accommodate EIC pseudodata. The histogram of χ^2 values for each data set included in our fits is shown in Fig. 5.4, together with the NNPDFpol1.0 result; the unweighted average $\langle \chi^2\rangle_{\rm set} \equiv \frac{1}{N_{\rm set}} \sum_{j=1}^{N_{\rm set}} \chi^2_{\rm set,j}$ and standard deviation over data sets

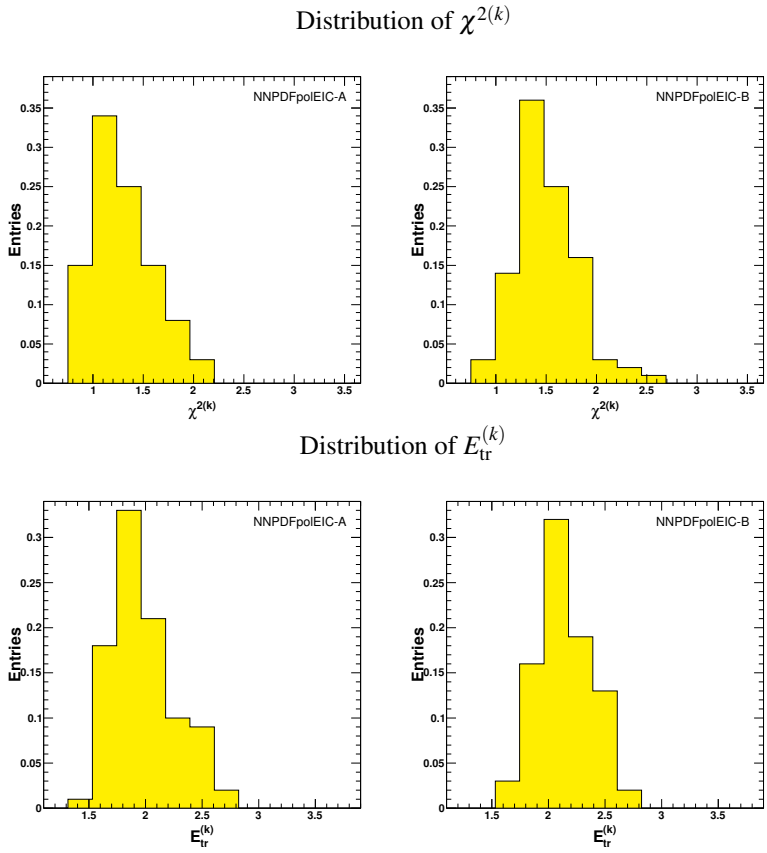

Figure 5.2: Distribution of $\chi^{2(k)}$ (upper plots) and $E_{\rm tr}^{(k)}$ (lower plots) from a sample of $N_{\rm rep} = 100$ replicas, for the `NNPDFpolEIC-A` (left plots) and `NNPDFpolEIC-B` (right plots) parton determinations.

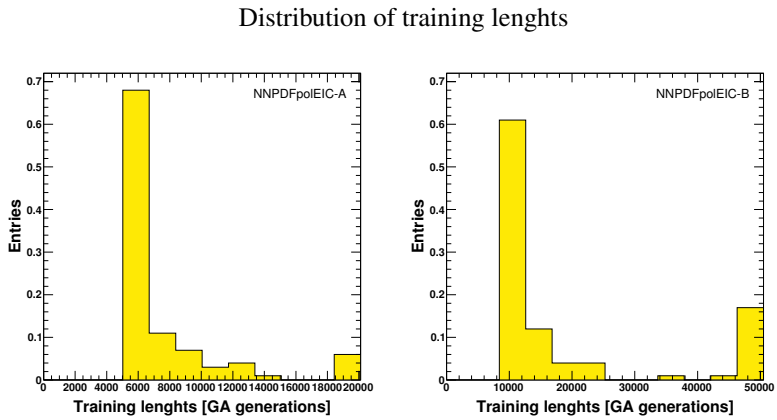

Figure 5.3: Distribution of training lenghts from a sample of $N_{\rm rep} = 100$ replicas, for the `NNPDFpolEIC-A` (left plots) and `NNPDFpolEIC-B` (right plots) parton determinations.

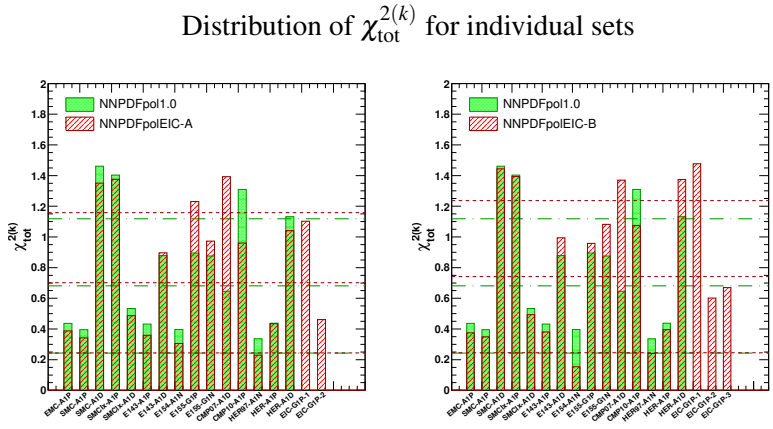

Figure 5.4: Value of the χ^2 per data point for the data sets included in the `NNPDFpolEIC-A` (left) and in the `NNPDFpolEIC-B` (right) fits, compared to `NNPDFpol1.0` [116]. The horizontal lines correspond to the unweighted average of the χ^2 values shown, and the one-sigma interval about it. The dashed lines refer to `NNPDFpolEIC-A` (left plot) or `NNPDFpolEIC-B` (right plot) fits, while the dot-dashed lines refer to `NNPDFpol1.0` [116].

are also shown. As already pointed out in Sec. 4.4, χ^2 values significantly below one are found as a consequence of the fact that information on correlated systematics is not available for most experiments, and thus statistical and systematic errors are added in quadrature. Note that this is not the case for the EIC pseudodata, for which, as mentioned, no systematic uncertainty was included; this may explain the somewhat larger (closer to one) value of the χ^2 per data point which is found when the pseudodata are included.

We notice that EIC pseudodata, which are expected to be rather more precise than fixed-target DIS experimental data, require more training to be properly learned by the neural network. This is apparent in the increase in $\langle TL \rangle$ in Tab. 5.4 when going from `NNPDFpol1.0` to `NNPDFpolEIC-A` and then `NNPDFpolEIC-B`. We checked that the statistical features discussed above do not improve if we run very long fits, up to $N_{\rm gen}^{\rm max} = 50000$ generations, without dynamical stopping. In particular, we do not observe a decrease of the χ^2 for those experiments whose value exceeds the average by more than one sigma. This ensures that these deviations are not due to underlearning, *i.e.* insufficiently long minimization.

5.4.2 Parton Distributions

Parton distributions from the `NNPDFpolEIC-A` and `NNPDFpolEIC-B` fits are compared to `NNPDFpol1.0` in Figs. 5.5-5.6 respectively. In these plots, PDFs are displayed at $Q_0^2 = 1$ GeV^2 as a function of x on a logarithmic scale; all uncertainties shown here are one-sigma bands. The positivity bound, obtained from the `NNPDF2.3` NLO unpolarized set [110] as discussed in Sec. 4.3, is also drawn.

The most visible impact of inclusive EIC pseudodata in both our fits is the reduction of PDF uncertainties in the low-x region ($x \lesssim 10^{-3}$) for light flavors and the gluon. The size of the effects is different for different PDFs. As expected, the most dramatic improvement is seen for the gluon, while uncertainties on light quarks are only reduced by a significant factor in

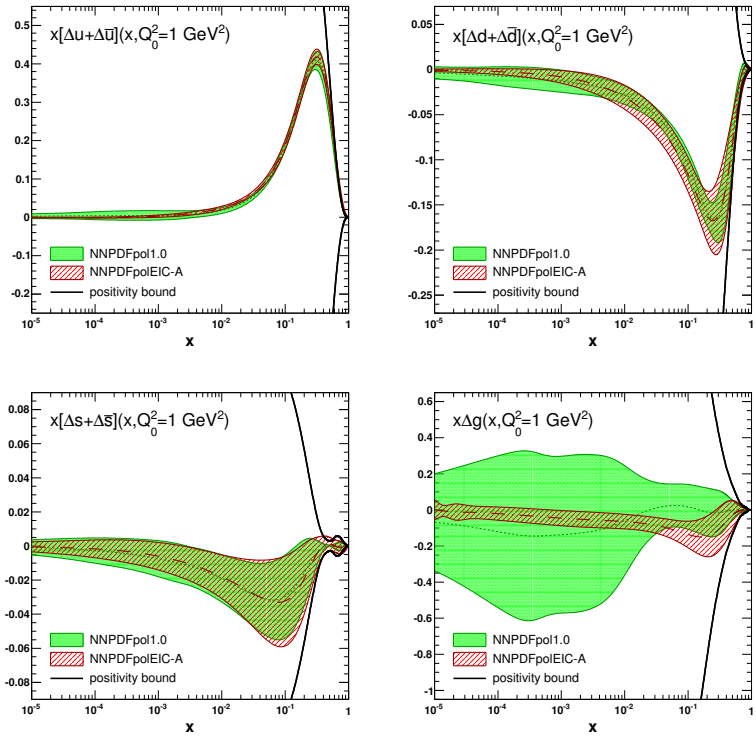

Figure 5.5: The `NNPDFpolEIC-A` parton distributions at $Q_0^2 = 1$ GeV^2 plotted as a function of x on a logarithmic scale, compared to `NNPDFpol1.0`.

the small x region. The uncertainty on the strange distribution is essentially unaffected: unlike in Ref. [222], we find no improvement on strangeness, due to the fact that we do not include semi-inclusive kaon production data, contrary to what was done there. When moving from `NNPDFpolEIC-A` to `NNPDFpolEIC-B` the gluon uncertainty decreases further, while other PDF uncertainties are basically unchanged.

In Fig. 5.7, we compare the polarized gluon PDF in our EIC fits to the `DSSV08` [75] and `NNPDFpol1.0` parton determinations, both at $Q_0^2 = 1$ GeV^2 and $Q^2 = 10$ GeV^2. The `DSSV08` uncertainty is the Hessian uncertainty computed assuming $\Delta\chi^2 = 1$, which corresponds to the default uncertainty estimate in Ref. [75]. This choice may lead to somewhat underestimated uncertainties, as discussed at length in Sec. 3.1.

It is clear that the gluon PDF from our fits including EIC pseudodata is approaching the `DSSV08` PDF shape, especially at a lower scale where the corresponding gluon does have some structure, despite the fact that at higher scales, where much of the data is located, perturbative evolution tends to wash out this shape. Also, this is more pronounced as more EIC pseudodata are included in our fit, *i.e.* moving from `NNPDFpolEIC-A` to `NNPDFpolEIC-B`. This means that EIC data would be sufficiently accurate to reveal the polarized gluon structure, if any.

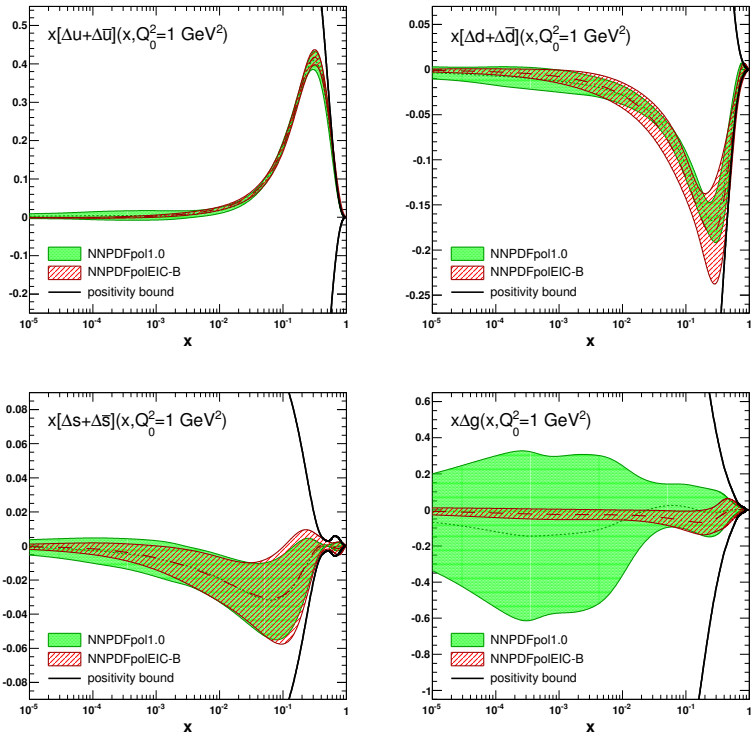

Figure 5.6: Same as Fig. 5.5, but for `NNPDFpolEIC-B`, compared to `NNPDFpol1.0`.

5.5 Phenomenological implications of EIC pseudodata

In this Section, we use our `NNPDFpolEIC-A` and `NNPDFpolEIC-B` parton determinations to reassess the spin content of the proton in the light of future EIC data. We also determine the expected contribution of the charm quark to the polarized structure function g_1, focusing on its potential to further pin down the uncertainty of the gluon distribution.

5.5.1 The spin content of the proton

It is particularly interesting to examine how the EIC data affect the determination of the first moments of the polarized PDFs $\Delta f(x,Q^2)$, Eq. (4.31) as they are directly related to the nucleon spin structure. We have computed the first moments, Eq. (4.31), of the singlet, lightest quark-antiquark combinations and gluon for the `NNPDFpolEIC-A` and `NNPDFpolEIC-B` PDF sets. The corresponding central values and one-sigma uncertainties at $Q_0^2 = 1$ GeV^2 are shown in Tab. 5.5, compared to `NNPDFpol1.0`.

It is clear that EIC pseudodata reduce all uncertainties significantly. Note that moving from `NNPDFpolEIC-A` to `NNPDFpolEIC-B` does not improve significantly the uncertainty on quark-antiquark first moments, but it reduces the uncertainty on the gluon first moment by a factor two. However, it is worth noticing that, despite a reduction of the uncertainty on

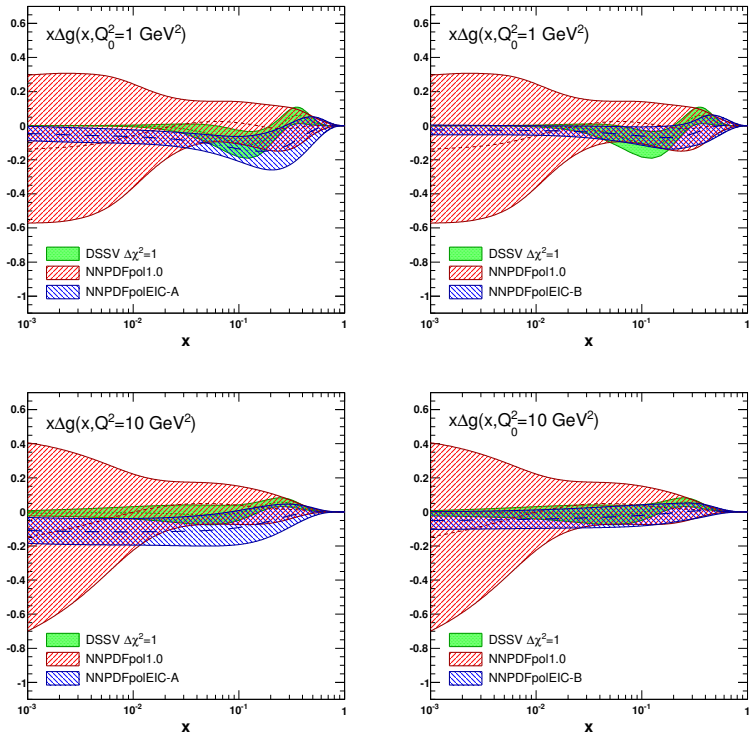

Figure 5.7: The polarized gluon PDF $\Delta g(x,Q_0^2)$, at $Q_0^2 = 1$ GeV^2 (upper panels) and at $Q^2 = 10$ GeV^2 (lower panels), in the `NNPDFpolEIC` PDF sets, compared to DSSV [75] and to `NNPDFpol1.0`.

Fit	$\langle\Delta\Sigma\rangle$	$\langle\Delta u+\Delta\bar{u}\rangle$	$\langle\Delta d+\Delta\bar{d}\rangle$	$\langle\Delta s+\Delta\bar{s}\rangle$	$\langle\Delta g\rangle$
`NNPDFpolEIC-A`	0.24 ± 0.08	0.82 ± 0.02	-0.45 ± 0.02	-0.13 ± 0.07	-0.59 ± 0.86
`NNPDFpolEIC-B`	0.21 ± 0.06	0.81 ± 0.02	-0.47 ± 0.02	-0.12 ± 0.07	-0.33 ± 0.43

Table 5.5: First moments of the polarized quark distributions at $Q_0^2 = 1$ GeV^2 for the fits in the present analysis. The corresponding values for `NNPDFpol1.0` are quoted in Tabs. 4.16-4.17.

the gluon first moment, even for the most accurate `NNPDFpolEIC-B` fit, the value remains compatible with zero even though the central value is sizable (and negative).

In order to assess the residual extrapolation uncertainty on the singlet and gluon first moments, we determine the contribution to them from the data range $x \in [10^{-3}, 1]$, *i.e.*

$$\langle\Delta\Sigma(Q^2)\rangle_{\rm TR} \equiv \int_{10^{-3}}^{1} dx\,\Delta\Sigma(x,Q^2)\,, \qquad \langle\Delta g(Q^2)\rangle_{\rm TR} \equiv \int_{10^{-3}}^{1} dx\,\Delta g(x,Q^2)\,. \tag{5.1}$$

The first moments Eq. (5.1) are given in Tab. 5.6 at $Q_0^2 = 1$ GeV^2 and $Q^2 = 10$ GeV^2, where results for central values, uncertainties, and correlation coefficients between the gluon and quark are collected.

	$Q^2 = 1$ GeV^2		$Q^2 = 10$ GeV^2		
	$\langle\Delta\Sigma(Q^2)\rangle_{\rm TR}$	$\langle\Delta g(Q^2)\rangle_{\rm TR}$	$\langle\Delta\Sigma(Q^2)\rangle_{\rm TR}$	$\langle\Delta g(Q^2)\rangle_{\rm TR}$	$\rho(Q^2)$
NNPDFpol1.0	0.25 ± 0.09	-0.26 ± 1.19	0.23 ± 0.16	-0.06 ± 1.12	$+0.861$
NNPDFpolEIC-A	0.27 ± 0.06	-0.53 ± 0.37	0.23 ± 0.05	-0.59 ± 0.50	-0.186
NNPDFpolEIC-B	0.24 ± 0.05	-0.23 ± 0.25	0.22 ± 0.04	-0.19 ± 0.32	-0.103

Table 5.6: The singlet and gluon truncated first moments and their one-sigma uncertainties at $Q^2 = 1$ GeV^2 and $Q^2 = 10$ GeV^2 for the NNPDFpolEIC PDF sets, compared to NNPDFpol1.0 [116]. The correlation coefficient ρ at $Q^2 = 10$ GeV^2 is also provided.

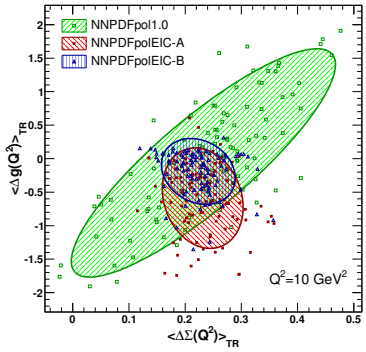

Figure 5.8: One-sigma confidence region for the quark singlet and gluon first moments in the measured region, Eq. (5.1). The values for individual replicas are also shown.

Comparing the results at $Q^2 = 1$ GeV^2 of Tab. 5.5 and Tab. 5.6 with those in Tabs. 4.16-4.17, we see that in the NNPDFpol1.0 PDF determination for the quark singlet combination the uncertainty on the full first moment is about twice as large as that from the measured region, and for the gluon it is about four times as large. The difference is due to the extra uncertainty coming from the extrapolation. In NNPDFpolEIC-B the corresponding increases are by 20% for the quark and 30% for the gluon, which shows that thanks to EIC data the extrapolation uncertainties would be largely under control. The correlation coefficient ρ significantly decreases upon inclusion of the EIC data: this means that the extra information contained in these data allows for an independent determination of the quark and gluon first moments.

In Fig. 5.8, we plot the one-sigma confidence region in the $(\langle\Delta\Sigma(Q^2)\rangle_{\rm TR}, \langle\Delta g(Q^2)\rangle_{\rm TR})$ plane at $Q^2 = 10$ GeV^2, for NNPDFpolEIC-A, NNPDFpolEIC-B and NNPDFpol1.0. Confidence regions are elliptical, since we have assumed that the truncated moments are Gaussianly distributed among the $N_{\rm rep} = 100$ replicas in the PDF ensemble. This is a reasonable assumption for all the three parton sets we are considering here, as shown in Fig. 5.9. The main result of our analysis, Fig. 5.8, can be directly compared to Fig. 8 of Ref. [222], which was based on the DSSV framework and is comparable to our NNPDFpolEIC-B results. In both analyses EIC pseudodata determine the singlet first moment in the measured region with an uncertainty of about ± 0.05.

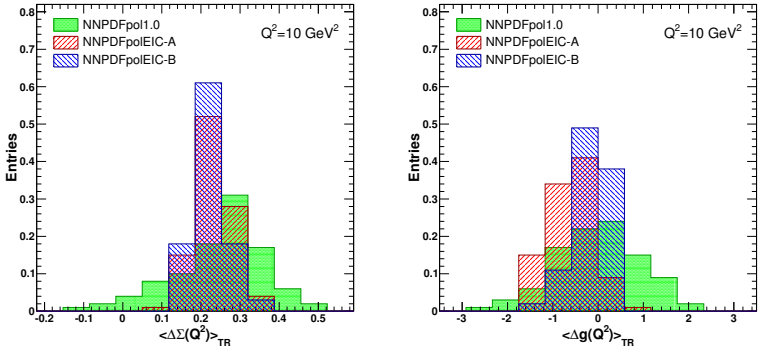

Figure 5.9: Distributions of the singlet (left) and gluon (right) truncated first moments at $Q^2 = 10$ GeV^2 from a set of $N_{\rm rep} = 100$ replicas in the `NNPDFpol1.0`, `NNPDFpolEIC-A` and `NNPDFpolEIC-B` parton ensembles.

On the other hand, in Ref. [222] the uncertainty on the gluon was found to be about ± 0.02, while we get a much larger result of ± 0.30. One may wonder whether this difference may be due at least in part to the fact that the DSSV fit on which the result of Ref. [222] is based also includes jet production and pion production data from RHIC, which may reduce the gluon uncertainty. To answer this, we have computed the contribution to the gluon first moment (again at $Q^2 = 10$ GeV^2) from the reduced region $0.05 \leq x \leq 0.2$, where the RHIC data are located. We find that the uncertainty on the contribution to the gluon first moment in this restricted range is ± 0.083 using `NNPDFpolEIC-B`, while it is ± 0.147 with `NNPDFpol1.0` and $^{+0.129}_{-0.164}$ with `DSSV+` [188]. We conclude that before the EIC data are added, the uncertainties in `NNPDFpol1.0` and `DSSV+` are quite similar despite the fact that `DSSV+` also includes RHIC data. Hence, the larger gluon uncertainty we find for the `NNPDFpolEIC-B` fit in comparison to Ref. [222] is likely to be due to our more flexible PDF parametrization, though some difference might also come from the fact that the SIDIS pseudodata included in Ref. [222] provide additional information on the gluon through scaling violations of the fragmentation structure function g_1^h. Of course this also introduces an uncertainty related to the fragmentation functions which is difficult to quantify.

5.5.2 Charm contribution to the g_1 structure function

In the QCD analysis of presently available DIS data, the contribution of heavy quarks to the polarized structure function g_1 is usually neglected. In both `NNPDFpol1.0` and `NNPDFpolEIC` polarized PDF determinations, heavy quarks are dynamically generated above threshold by (massless) Altarelli-Parisi evolution in the ZM-VFN scheme (see Sec. 4.2). An exception to this treatment of heavy quark masses is Ref. [76], where charm quark production in photon-gluon fusion process is treated at LO [227] in the fixed-flavor number (FFN) scheme.

Intrinsic heavy quark effects are neglected in most analyses of polarized PDFs since they have been shown to be relatively small already on the scale of present-day unpolarized PDF uncertainties [106], which are rather smaller than their polarized counterparts. However, as EIC data are expected to be far more accurate than those available so far, effects of the finite heavy quark masses could be at least non-negligible. The treatment of these effects requires

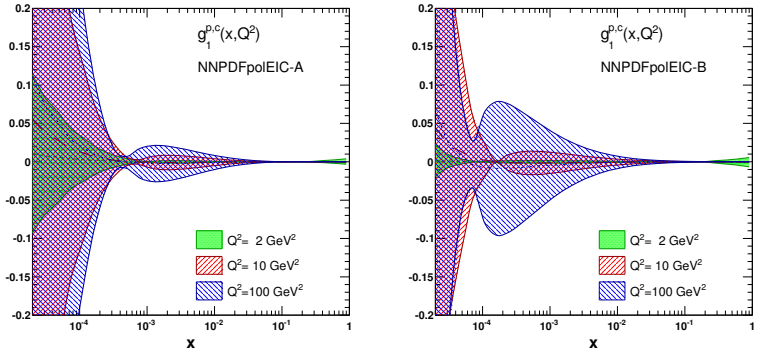

Figure 5.10: The charm contribution $g_1^{p,c}(x,Q^2)$ to the DIS proton polarized structure function g_1^p as a function of x at three different energy scales Q^2. Results are shown for both the `NNPDFpolEIC-A` (left) and `NNPDFpolEIC-B` (right) parton determinations.

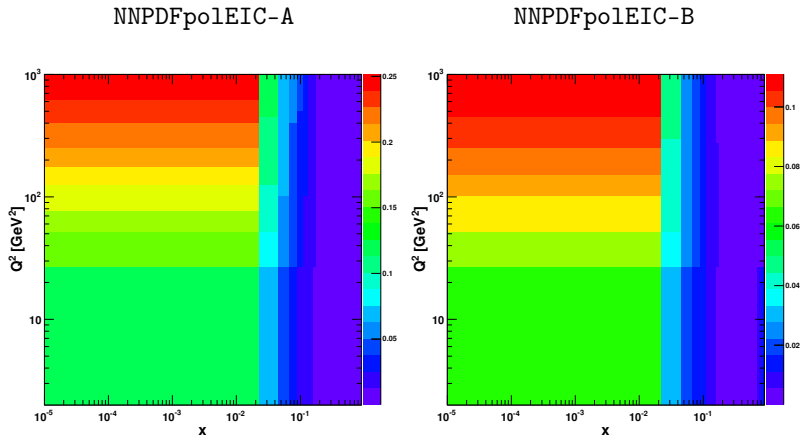

Figure 5.11: The contour plots for the ratio $g_1^{p,c}(x,Q^2)/g_1^p(x,Q^2)$ from both the `NNPDFpolEIC-A` (left) and `NNPDFpolEIC-B` (right) parton determinations.

a proper quark mass scheme, for instance the `FONLL` scheme, firstly introduced in Ref. [228] and explicitely extended to DIS in Ref. [124]. The method is based upon the idea of looking at both the massless and massive scheme calculations as power expansions in the strong coupling constant, and replacing the coefficient of the expansion in the former with their exact massive counterpart in the latter, when available. In order to suppress higher order contributions arising in the subtraction term near the threshold region, two prescriptions are proposed in Ref. [124]. One consists in damping the subtraction term by a threshold factor which differs from unity by power-suppressed terms; the other consists in using a rescaling variable. Both prescriptions introduce terms which are formally subleading with respect to the order of the calculation, therefore they do not change its nominal accuracy, but they may in practice improve the perturbative stability and smoothness of the results.

Thanks to its simplicity, the `FONLL` scheme has been implemented in the `FastKernel`

framework for the determination of unpolarized PDFs based on the NNPDF methodology [106]. The generalization of the FONLL scheme to the polarized structure function g_1 is quite simple: details are given in Appendix C. For spin-dependent DIS the charm contribution to the structure function g_1, generated through photon-gluon fusion process, $\gamma g \to c\bar{c}$, will very much depend on the currently unknown size of $\Delta g(x,Q^2)$ at small x. For instance, we plot in Fig. 5.10 the expectations for the charm contribution $g_1^{p,c}$ to the proton structure function g_1^p, computed from the NNPDFpolEIC-A and NNPDFpolEIC-B parton determinations. Results are displayed at three different energy scales, namely $Q^2 = 2, 10, 100$ GeV^2. In Fig. 5.11, we also show the contour plot for the ratio $g_1^{p,c}(x,Q^2)/g_1^p(x,Q^2)$ for both these fits. We conclude that the charm contribution to the proton structure function $g_1^{p,c}$, though being small, could be as much larger as 10-20% of the total g_1^p in the kinematic region probed by an EIC. Hence, in order to further pin down the gluon uncertainty from intrinsic charm effects, one should be able to measure its corresponding contribution to the g_1 structure function within this accuracy. Our result is consistent with that obtained in the DSSV framework presented in Refs. [128].

In summary, the EIC data would entail a considerable reduction in the uncertainty on the polarized gluon PDF, they would provide first evidence for its possible nontrivial x shape and for its possible large contribution to the nucleon spin. However, this goal would be reached with sizable residual uncertainty: hence, the measurement of the charm contribution to the proton structure function $g_1^{p,c}$, which is directly sensitive to the gluon, might provide more information on the Δg distribution.

6 A first global determination of unbiased polarized PDFs

In this Chapter, we present a first global determination of polarized parton distributions based on the NNPDF methodology: `NNPDFpol1.1`. Compared to `NNPDFpol1.0`, the parton set determined in Chap. 4, `NNPDFpol1.1` is obtained using, on top of inclusive DIS data, also data from recent measurements of open-charm production in fixed-target DIS, and of jet and W production in proton-proton collisions. After motivating our analysis in Sec. 6.1, we will review the theoretical description of these processes in Sec. 6.2. The features of the experimental data included in our analysis are then presented in Sec. 6.3. In Sec. 6.4, we discuss how the `NNPDFpol1.1` parton set is obtained via Bayesian reweighting of prior PDF Monte Carlo ensembles, followed by unweighting, as outlined in Sec. 3.2.4. We also present its main features in comparison to `NNPDFpol1.0` and `DSSV08`. Finally, in Sec. 6.5, we discuss some phenomenological implications of our new polarized parton set with respect to the spin content of the proton. Some of the results presented in this Chapter have appeared in preliminary form in Refs. [118, 119].

6.1 Motivation

The `NNPDFpol1.0` parton set presented in Chap. 4 is the first determination of polarized parton distributions based on the NNPDF methodology. However, this is based on inclusive, neutral-current, DIS data only, which have two major drawbacks, as pointed out several times in this Thesis. First, they do not allow for quark-antiquark separation; second, their kinematic coverage is rather limited, both at small-x and high-Q^2 values. For this reason, all PDFs are affected by large uncertainties where experimental data are not available. Besides, the polarized gluon PDF, determined through scaling violations in DIS, is almost unconstrained because of the rather short Q^2 lever arm provided by data.

As discussed in Sec. 3.1, one has to resort to processes other than inclusive DIS to obtain further knowledge of polarized parton distributions. An impressive set of experimental data have become available in the last years: these include semi-inclusive DIS (SIDIS) data in fixed-target experiments [95, 96, 229–231], one- or two-hadron and open-charm production data in

lepton-nucleon scattering [213,232–235], and semi-inclusive particle production [214,216,217], high-p_T jet production [215, 218] and parity-violating $W^{\pm}$ boson production [236, 237] data in polarized proton-proton collisions at RHIC. As already summarized in Tab. 3.1, all these data are expected to probe different aspects of polarized PDFs: semi-inclusive DIS and $W^{\pm}$ production data allow one to determine the light quark-antiquark separation, while jet and pion production data in polarized proton-proton collisions, as well as hadron or open-charm electroproduction data in fixed-target experiments, give a handle on the size and the shape of the polarized gluon distribution. A theoretical description of the processes corresponding to these data will be given in Sec. 6.2 below.

Nevertheless, all these measurements fall in the x region already covered by DIS data. Given that the uncertainties on the first moments of polarized PDFs, which eventually determine the contribution of each parton to the total proton spin, are already limited by the extrapolation into the unconstrained small-x region (see the discussion in Sec. 4.5.1), it is clear that only moderate improvements in these are expected from the addition of new data other than DIS data. Only a future high-energy polarized Electron-Ion Collider (EIC) [128, 209, 219] would be likely to probe the small-x regime of PDFs, and thus improve our knowledge of the polarized PDFs first moments, as we demonstrated in Chap. 5.

Since no further progress is expected by the time an EIC will start to operate, some effort has been devoted to perform *global* determinations of polarized parton sets, including all available experimental data. Presently, only two such sets are available: `LSS10` [77], which also includes SIDIS beside inclusive DIS data, and those from the DSSV family [75, 187, 188], which include, on top of these, also inclusive jet and identified hadron production measurements from polarized proton-proton collisions at RHIC. The goal of the analysis in this Chapter is to incorporate in the NNPDF determination of polarized parton distributions the new experimental information provided by some of the processes mentioned above, and thus release the first global polarized PDF set based on the NNPDF methodology: `NNPDFpol1.1`.

6.2 Theoretical overview of polarized processes other than DIS

Before addressing our global determination of a set of polarized parton distributions, we provide a theoretical overview of polarized processes other than DIS, available from both fixed-target and collider experiments.

6.2.1 Semi-inclusive lepton-nucleon scattering in fixed-target experiments

The role of fixed-target lepton-nucleon scattering in determining the spin structure of the nucleon is not restricted to inclusive reactions, as those we have considered so far. Actually, more exclusive processes, in which one measures one or more outgoing final state particles, can be used to constrain polarized sea quark distributions and to gain some more knowledge on the polarized gluon distribution. In the following, we discuss in turn the potential of each of these processes, namely semi-inclusive DIS and heavy flavor hadron production.

Semi-inclusive DIS. Semi-inclusive DIS (SIDIS) is a DIS process in which a hadron h, originated by the fragmentation of the struck quark, is detected in the final state:

$$l(\ell) + N(P) \to l'(\ell') + h(P_h) + X(P_X)\,, \tag{6.1}$$

where we use the same notation adopted in Eq. (2.1) supplemented with the four-momentum of the final hadron, P_h. Due to the statistical correlation between the flavor of the struck quark and the type of the hadron formed in the fragmentation process, semi-inclusive DIS with identified pions or kaons may provide a handle on the $\Delta\bar{u}$, $\Delta\bar{d}$ and $\Delta\bar{s}$ parton distributions, respectively [95]. For example, roughly speaking the presence of a π^+ in the final state indicates that it is likely that a u-quark or a $\bar{d}$-antiquark was struck in the scattering, because the π^+ is a $(u\bar{d})$ bound state.

The theoretical description of SIDIS, with longitudinally polarized lepton beams, closely follows that of inclusive DIS given in Chap. 2. In analogy to Eq. (2.16), the SIDIS differential cross-section can be written as

$$\frac{d^6\sigma}{dxdyd\phi dzdp_T^2 d\Phi} \propto L_{\mu\nu} W_h^{\mu\nu}; \tag{6.2}$$

here the leptonic tensor, $L_{\mu\nu}$, has exactly the form given in Eq. (2.17) with the symmetric and antisymmetric parts of Eqs. (2.20)-(2.23), while the hadronic tensor, $W_h^{\mu\nu}$, now contains additional degrees of freedom corresponding to the fractional energy z of the final state hadron, the p_T component of the final hadron three momentum, transverse to that of the virtual photon, and the azimuthal angle Φ of the hadron production plane relative to the lepton scattering plane. Integration over Φ and p_T^2 produces the cross-section relevant for the experimental observables. These are the longitudinal and transverse asymmetries $A_\parallel^h$ and $A_\perp^h$, defined in analogy to Eq. (2.83).

Thanks to factorization, hadronic and leptonic degrees of freedom are separated, hence kinematic factors depending only on x and y (or Q^2) are carried over directly from inclusive scattering in relating the measured asymmetries to their virtual photo-absorption counterparts, Eq. (2.93). In particular, A_1 and A_2, Eq. (2.94), should now read A_1^h and A_2^h in terms of the SIDIS cross-sections $\sigma_{1/2,3/2}^h$ of produced hadrons of type h. In particular, we obtain

$$A_1^h(x,Q^2,z) = \frac{\sum_q e_q^2 \Delta q(x,Q^2) \otimes D_q^h(z,Q^2)}{\sum_q e_{q'}^2 q'(x,Q^2) \otimes D_{q'}^h(z,Q^2)}, \tag{6.3}$$

where e_q are the quark electric charges, Δq (q) are the helicity-dependent (-averaged) PDFs, D_q^h is the fragmentation function for the quark q to fragment into a hadron h, and $\otimes$ denotes the convolution product, Eq. (2.63).

Measurements of spin-dependent asymmetries in SIDIS have been performed by several experimental collaborations, namely SMC [230], HERMES [95] and COMPASS [96–98], and have been included in some global determinations of polarized parton distributions [75,77]. We should notice that the analysis of these data requires the usage of fragmentation functions, which have to be determined in turn from experimental data. Usually, they are extracted, independently of PDFs, from electron-positron annihilation, proton-proton collisions and possibly SIDIS data (see Ref. [238] for a review). Despite the considerable experimental and theoretical effort which has gone into the determination of sets of fragmentation functions [184–186, 239–243], their knowledge is still rather poor. In particular, recent work [162] has emphasized the failure of all available sets of fragmentation functions in describing the most updated inclusive charged-particle spectra data at the LHC. For these reasons, when used in a global determination of polarized PDFs including SIDIS data, they are likely to introduce an uncertainty which is difficult to quantify.

Heavy flavor hadron production. Heavy flavor hadron production is clearly sensitive to the shape and the size of the spin-dependent gluon distribution [244, 245]. In this case, the gluon polarization can be accessed via the photon-gluon fusion (PGF) mechanism, which results in the production of a quark-antiquark pair, see Fig. 2.3-(d). Experimental signatures to tag this partonic subprocess are the production of one or two hadrons with high p_T in the final state, and open-charm events: in this case, the $q\bar{q}$ pair is required to be a $c\bar{c}$ pair and an outgoing charmed meson is reconstructed. At LO, the virtual photon-nucleon asymmetry for open-charm production, $A_{LL}^{\gamma N\to D^0 X}$, is expressed as

$$A_{LL}^{\gamma N\to D^0 X} \equiv \frac{\Delta\sigma_{\gamma N}}{\sigma_{\gamma N}} = \frac{\Delta\hat{\sigma}_{\gamma g}\otimes\Delta g\otimes D_c^{D^0}}{\hat{\sigma}_{\gamma g}\otimes g\otimes D_c^{D^0}}\,, \tag{6.4}$$

where $\Delta\hat{\sigma}_{\gamma g}$ ($\hat{\sigma}_{\gamma g}$) is the spin-dependent (-averaged) partonic cross-section for PGF, $\gamma^* g\to c\bar{c}$, Δg (g) is the polarized (unpolarized) gluon PDF, and $D_c^{D^0}$ is the non-perturbative fragmentation function of a produced charm quark into the observed D^0 meson, which is assumed to be spin independent. In principle, a measurement of the asymmetry in Eq. (6.4) can then provide a direct handle on Δg.

Measurements of longitudinal spin asymmetries in high-p_T hadron production were performed by HERMES [232, 233] at DESY and by SMC [234] and COMPASS [235] at CERN. However, in such a reaction, the measured asymmetries receive contributions not only by pure PGF events, but also by a significant fraction of background events, mainly due to the two competing processes of gluon radiation by QCD Compton scattering ($\gamma^* q\to qg$) and photon absorption at the lowest order of DIS ($\gamma^* q\to q$). At variance with hadron-pair production, open-charm production is free of background, since the PGF subprocess is the main mechanism for producing charm quarks in polarized DIS.

The proper theoretical description of these processes depends on the virtuality of the probing photon. In case of photoproduction, where a quasi-real photon is exchanged, on top of direct contributions [246–249], one has to include also *resolved* contributions [250, 251], where the photon fluctuates into a vector meson of the same quantum numbers before the hard scattering with partons in the proton takes place. If the virtuality Q of the photon is of $\mathcal{O}(1\text{ GeV})$ or higher, resolved processes are sufficiently suppressed but the additional momentum scale Q greatly complicates the calculations of phase-space and loop integrals.

Only few calculations are available for heavy flavor hadron production in case of polarized beams and targets at NLO accuracy [244–247]. A complete phenomenological study of charm quark photoproduction in longitudinally polarized lepton-hadron collisions at NLO accuracy has been presented only very recently [252]. For the first time, both direct and resolved photon contributions have been included there to compute the relevant cross-sections for spin-dependent heavy flavor hadroproduction. For this reason, the available data on hadron and open-charm production [213, 232–235] have not been included in global QCD analyses of polarized parton distributions so far. Experimental collaborations have analyzed their data only in terms of the gluon polarization, $\Delta g(x,Q^2)/g(x,Q^2)$, under certain simplifying assumptions and based on leading order matrix elements. Nonetheless, the results of these exercises, illustrated in Fig. 6.1, are in fairly good agreement with the NLO prediction obtained from our `NNPDFpol1.0` analysis. We will explicitly show to which extent COMPASS open-charm data can further pin down the polarized gluon uncertainty in Sec. 6.4.2.

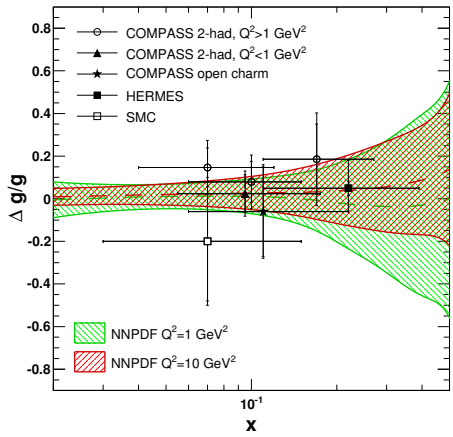

Figure 6.1: The theoretical prediction for the ratio $\Delta g(x,Q^2)/g(x,Q^2)$ computed from the polarized (unpolarized) `NNPDFpol1.0` (`NNPDF2.3`) parton sets at NLO, compared to LO determinations from one- or two-hadron and open-charm production data in fixed-target DIS experiments [213, 232–235].

6.2.2 Spin asymmetries in proton-proton collisions

High-energy collisions from longitudinally polarized proton beams, as available at the Relativistic Heavy-Ion Collider (RHIC), provide a unique way to probe proton spin structure and dynamics [57, 253]. Typically, the quantities measured at RHIC are spin asymmetries. As an example, for collisions of longitudinally polarized proton beams, one defines the double-spin asymmetry for a given process as

$$A_{LL} = \frac{\sigma^{++} - \sigma^{+-}}{\sigma^{++} + \sigma^{+-}} \equiv \frac{\Delta\sigma}{\sigma}, \tag{6.5}$$

where σ^{++} (σ^{+-}) is the cross-section for the process with equal (opposite) proton beam polarizations. As for collisions of unpolarized proton beams, spin-dependent inelastic cross-sections factorize into convolutions of polarized parton distribution functions of the proton and hard-scattering cross-sections describing the spin-dependent interactions of partons. Similarly to inclusive DIS, one can write schematically

$$\sigma = \sum_{a,b,(c)=q,\bar{q},g} f_a \otimes f_b (\otimes D_c^H) \otimes \hat{\sigma}_{ab}^{(c)}, \tag{6.6}$$

$$\Delta\sigma = \sum_{a,b,(c)=q,\bar{q},g} \Delta f_a \otimes \Delta f_b (\otimes D_c^H) \otimes \Delta\hat{\sigma}_{ab}^{(c)}, \tag{6.7}$$

for the denominator and the numerator in Eq. (6.5) respectively. Here, the sum is over all contributing partonic channels $a+b \to c(\to H)+X$ producing the desired high-p_T, large-invariant mass final state (or detected hadron H). As usual, $\otimes$ denotes the convolution, Eq. (2.63), between unpolarized (polarized) parton distributions, $f_{a,b}$ ($\Delta f_{a,b}$), parton-to-hadron fragmentation function, D_c^H (only for those processes with identified hadrons in final states), and

elementary spin-averaged (-dependent) hard partonic cross-section $\hat{\sigma}_{ab}^{(c)}$ ($\Delta\hat{\sigma}_{ab}^{(c)}$). In particular, the spin-dependent partonic cross-section is defined as

$$\Delta\hat{\sigma}_{ab}^{(c)} \equiv \frac{1}{2}\left[\hat{\sigma}_{ab}^{++} - \hat{\sigma}_{ab}^{+-}\right], \tag{6.8}$$

the signs denoting the helicity states of the initial partons a, b.

At RHIC, there are a number of processes which allow for the measurement of spin asymmetries like that in Eq. (6.5). Depending on the dominant partonic subrocess, they can probe different aspects of the nucleon spin structure. For instance, some of them allow for a clean determination of gluon polarizations, while others are more sensitive to quark and antiquark helicity states. We will discuss below the main processes for which measurements of spin asymmetries at RHIC are presently available. We do not give here any details about other measurements possible at RHIC, but not yet performed, such as the observation of high-p_T or *prompt* photon production, heavy-flavor production and Drell-Yan production of lepton pairs. We refer to [57] for a theoretical review on them.

High-p_T inclusive jet production double-spin asymmetry. A clean and theoretically robust process to probe the polarized gluon PDF, Δg, is inclusive jet production, thanks to the dominance of the gg and qg initiated subprocesses in the accessible kinematic range [254, 255] (see also Tab. 3.1). The situation is analogous to the unpolarized case, where inclusive jet production data from the Tevatron and the LHC [256–259] are instrumental in pinning down the medium- and large-x gluon behavior.

In polarized collisions, the relevant experimental observable in jet production is the longitudinal double-spin asymmetry defined along Eq. (6.5)

$$A_{LL}^{1\,jet} = \frac{\sigma^{++} - \sigma^{+-}}{\sigma^{++} + \sigma^{+-}}. \tag{6.9}$$

For dijet production, at LO, the parton kinematics is given by

$$x_1 = \frac{p_T}{2\sqrt{s}}\left(e^{\eta_3} + e^{\eta_4}\right), \qquad x_2 = \frac{p_T}{2\sqrt{s}}\left(e^{-\eta_3} + e^{-\eta_4}\right), \tag{6.10}$$

where p_T is the transverse jet momentum, $\eta_{3,4}$ are the rapidities of the two jets and $\sqrt{s}$ is the center-of-mass energy. In single-inclusive jet production, the underlying Born kinematics is not uniquely determined because the second jet is being integrated out. For the illustrative purposes of Fig. 6.2, we will use the following expression to characterize the Born kinematics

$$x_{1,2} = \frac{p_T}{\sqrt{s}} e^{\pm\eta}, \tag{6.11}$$

with η the rapidity of the leading jet, which corresponds to Eq. (6.10), provided the incoming partons carry an equal amount of longitudinal momentum and thus $\eta_3 = -\eta_4$. This is a good approximation at RHIC due to the limited coverage in rapidity as compared to unpolarized hadron colliders.

The calculation of both the numerator and the denominator in Eq. (6.9) requires the definition of a suitable jet algorithm. Also, notice that corrections up to NLO accuracy should be included in the computation of jet cross-sections, through the algorithm for jet reconstruction, since it is only at NLO that the QCD structure of the jet starts to play a role in the theoretical

description. This would also provide for the first time the possibility of realistically matching the procedures used in experiment to group final-state particles into jets.

At NLO, a large number of infrared divergencies are found in the computation of virtual and real diagram contributions to the jet cross-section, due to the large number of color-interacting, massless partons involved in the hard-scattering processes. It is then necessary to devise a procedure to perform the calculation of the divergent parts and to show their cancellation in the sum which defines any infrared-safe physical observable. Several independent methods to calculate any infrared-safe quantity in any kind of hard unpolarized collision are available in the literature [260–262]. In particular, the subtraction method of Ref. [261] allows for organizing the computation is such a way that the singularities are extracted and canceled by hand, while the remainder may be integrated numerically over phase space. This approach has the advantage of being very flexible; it may be used for any infrared-safe observable, with any experimental cut. On the other hand, the numerical integration involved turns out to be rather delicate and time-consuming. The subtraction method has been used in Ref. [263] to develop a computer code that generates partonic events and outputs the momenta of the final-state partons which can be eventually used to define the physical observables for one ore more jet production in proton-proton collisions. Such a *parton generator* is not equivalent to the usual Monte Carlo parton shower programs, since it is the result of a fixed-order QCD calculation. The subtraction method and the computer code of Refs. [261, 263], supplemented with the proper matrix elements [264, 265], was then extended to the case of polarized proton-proton collisions in Ref. [266].

The spin-dependent (and spin-averaged) cross-section for single-inclusive high-p_T jet production is also available from Ref. [267]. In comparison to Ref. [266], the approach of Ref. [267] uses a largely analytic technique for deriving the relevant partonic cross-sections, which becomes possible if one assumes the jet to be a rather narrow object. This assumption is equivalent to the approximation that the cone opening R of the jet is not too large, and hence was termed *small-cone approximation* (SCA) in [267]. In the SCA, one systematically expands the partonic cross-sections around $R = 0$. The dependence on R is of the form $\mathscr{A}\log R + \mathscr{B} + \mathscr{O}(R^2)$: the coefficients $\mathscr{A}$ and $\mathscr{B}$ are retained and calculated analytically, whereas the remaining terms $\mathscr{O}(R^2)$ and beyond are neglected. The advantage of this procedure is that it leads to much faster and more efficient computer code, since all singularities arising in intermediate steps have explicitly canceled and are not subject to delicate numerical treatments. The SCA was shown [267] to produce results comparable to those of Ref. [266] (which is available for arbitrary cone openings), in the kinematic range where RHIC data are available. For these reasons, the code provided in Ref. [267] is better suited than that of Ref. [266] for the intensive computations required to include jet data in a global QCD fit.

Semi-inclusive π^0 production double-spin asymmetry. In order to constrain the polarized gluon distribution, one can look for high-p_T leading hadrons such as π^-, π^0, π^+, whose production proceeds through the same partonic subprocesses involved in jet production, in particular $qg \to qg$ and $gg \to qg$ (see Tab. 3.1). However, the hadronization of the struck parton into the final, measured, pion is described by a non-perturbative fragmentation function, D_c^π: this enters the theoretical description of the corresponding double-spin asymmetry, as encompassed in Eq. (6.5).

Next-to-leading order QCD corrections to the spin-dependent cross-section for single-inclusive hadron production in proton-proton collisions have been computed for the first time in Refs. [268, 269]. As dictated by Eqs. (6.6)-(6.7), we need to sum over all possible final

states in each channel $ab \to cX$, in compliance with the requirement of single-inclusiveness of the cross section. For instance, in case of $qg \to qX$ one needs, besides the virtual corrections to $qg \to qg$, three different $2 \to 3$ reactions: $qg \to q(gg)$, $qg \to q(q\bar{q})$, $qg \to q(q'\bar{q}')$ (where brackets indicate the unobserved parton pair). The combination of all three processes together will allow for obtaining a finite result.

The two computations presented in Refs. [268, 269] differ from each other in the way the integration over the entire phase space of the two unobserved partons, in the $2 \to 3$ contributions, is carried out. In Ref. [268], this calculation is performed numerically, by extending the already mentioned subtraction method of Refs. [261, 263] to the case of single-hadron production observables. A computer code customized to compute any infrared-safe quantity corresponding to one-hadron production at NLO accuracy is then presented as a result. Conversely, in Ref. [269], the phase-space integration of the $2 \to 3$ contributions is performed analitically. As already noticed in the case of jet production, this has the main advantage to obtain much faster and more efficient computer code, which is better suited for the intensive computations required by a global QCD fit of polarized parton distributions.

Small-p_T single-spin $W^\pm$ production asymmetry. Production of $W^\pm$ bosons in high energy collisions from longitudinally polarized proton beams provides an ideal tool for the study of individual helicity states of quarks and antiquarks inside the proton, complementary to, but independent of, SIDIS [57]. Within the stadard model, the process $\overrightarrow{p}\, p \to W^\pm X$ (the arrow denotes the polarized proton beam) is driven by a purely weak interaction which couples left-handed quarks with right-handed antiquarks only ($u_L \bar{d}_R \to W^+$ and $d_L \bar{u}_R \to W^-$, with some contamination from s, c, $\bar{s}$ and $\bar{c}$, mostly through quark mixing), thus giving rise to a W parity-violating longitudinal single-spin asymmetry, sensitive to Δq and $\Delta\bar{q}$ flavor dependence. This asymmetry is defined as

$$A_L \equiv \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-} = \frac{\Delta\sigma}{\sigma}, \tag{6.12}$$

where $\sigma^{+(-)}$ denotes the cross-section for colliding of positive (negative) longitudinally polarized protons off unpolarized protons. Notice that this definition differs from that provided in Eq. (6.5), since only one of the two proton beams is polarized.

If we consider the simplest parton-level process $u\bar{d} \to W^+$ at LO, Eq. (6.12) will read [57]

$$A_L^{W^+} \approx \frac{\Delta u(x_1)\bar{d}(x_2) - \Delta\bar{d}(x_1)u(x_2)}{u(x_1)\bar{d}(x_2) - \bar{d}(x_1)u(x_2)}, \tag{6.13}$$

where x_1 and x_2 are the momentum fractions, carried by quarks and antiquarks, related to y_W, the W boson rapidity relative to the polarized proton, and to $\sqrt{s}$, the hadronic center-of-mass energy, by the relation

$$x_{1,2} = \frac{M_W}{\sqrt{s}} e^{\pm y_W}. \tag{6.14}$$

The measurement of the rapidity distribution of the W bosons thus provide a direct handle on the flavor-separated polarized quark and antiquark distributions. Indeed, at large rapidities, $y_W \gg 0$, where sea distributions are suppressed because x_1 is in the valence region, the asymmetry $A_L^{W^+}$ is given by $\Delta u/u$, whereas it approaches $\Delta\bar{d}/\bar{d}$ in the opposite limit, $y_W \ll 0$. The situation is similar for W^- production, now interchanging the roles of u and d flavors.

The naive Born-level picture given above needs to be modified to account for additional aspects, both theoretical and experimental [57]. The former include higher-order perturbative

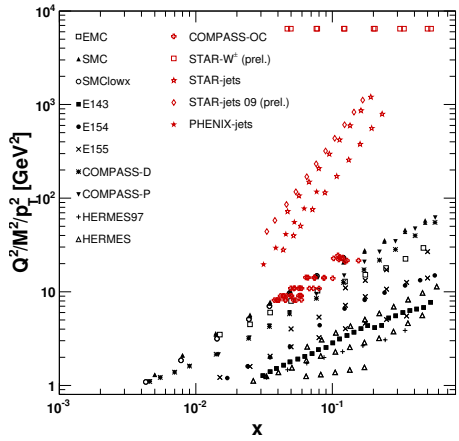

Figure 6.2: Kinematic coverage in the (x, Q^2) plane of experimental data included in the `NNPDFpol1.1` parton set. New hadronic data are listed in the second column.

corrections and other allowed initial states, including Cabibbo-suppressed channels. Concerning the latter, W bosons are reconstructed through their leptonic decays $W^{\pm} \to e^{\pm}\nu$ at RHIC, therefore the observed process is actually $pp \to \ell^{\pm} X$, with the neutrino escaping undetected. One will then measure the rapidity distributions of the charged leptons rather than of the W bosons themselves. All these issues have been taken into account in the NLO calculation of the cross-section and longitudinal single-spin asymmetry, Eq. (6.12), presented in Ref. [270] as a computer program: it may be readily used to include experimental spin asymmetry data in a global analysis of helicity parton densities.

6.3 Experimental input

Among the processes described in Sec. 6.2, in the present analysis we only consider open-charm production from COMPASS and single-inclusive high-p_T jet and W production from RHIC. Actually, the precise knowledge of fragmentation functions plays a minor role in the theoretical description of these processes. On the one hand, in the kinematic regime accessed by COMPASS, open-charm production shows only a slight dependence on the fragmentation of the charm quark into a D meson; on the other hand, the theoretical description of jet and W production asymmetries in proton-proton collisions does not involve fragmentation into identified hadrons in the final state. We prefer not to include data whose analysis requires the usage of fragmentation functions since these are poorly known objects: hence, they are likely to affect our unbiased PDFs by an uncertainty difficult to quantify, as discussed in Sec. 6.2.1.

In Fig. 6.2, we plot the new data points considered in the present analysis, together with inclusive DIS data already included in the fit presented in Chap. 4. More details about leading partonic subprocesses, probed polarized PDFs, and the ranges of x and Q^2 that become accessible were already summarized in Tab. 3.1. We will discuss below the main features of the new data sets, separately for each process.

Data set	$N_{\rm dat}$	jet-algorithm	R	$[\eta_{\rm min},\eta_{\rm max}]$	$\mathscr{L}$ [pb^{-1}]	Ref.
STAR 1j-05	10	midpoint-cone	0.4	$[+0.20,+0.80]$	2.1	[215]
STAR 1j-06	9	midpoint-cone	0.7	$[-0.70,+0.90]$	5.5	[215]
STAR 1j-09 (prel.)	11	midpoint-cone	0.7	$[-1.00,+1.00]$	25	[273]
PHENIX 1j	6	seed-cone	0.3	$[-0.35,+0.35]$	2.1	[218]

Table 6.1: Some features of the jet data included in the present analysis: the number of available data points, $N_{\rm dat}$, the algortihm used for jet reconstruction, the range over which the rapidity η is integrated and the integrated luminosity, $\mathscr{L}$.

Open-charm production at COMPASS. The COMPASS collaboration has recently presented experimental results for the photon-nucleon asymmetry $A_{LL}^{\gamma N\to D^0X}$, Eq. (6.4), obtained by scattering polarized muons of energy $E_\mu = 160$ GeV^2 (center-of-mass energy roughly $\sqrt{s}\sim 18$ GeV^2) off longitudinally polarized protons or deuterons from a ^{6}LiD or NH_3 targets, in the photoproduction regime (photon virtuality roughly $Q^2\sim 0$ GeV^2) [213]. A detailed description of the experimental setup can be found in Ref. [271]. Three different data sets, each one of $N_{\rm dat}=15$ data points, were presented there, depending on which D^0 decay mode was assumed to reconstruct the charmed hadron from the observed final states: $D^0\to K^-\pi^+$, $D^0\to K^-\pi^+\pi^0$ or $D^0\to K^-\pi^+\pi^+\pi^-$. In the following, they will be referred to as COMPASS $K1\pi$, COMPASS $K2\pi$ and COMPASS $K3\pi$ respectively. Experimental correlations between systematic uncertainties are not provided. Assuming LO kinematics, the experiment may probe the polarized gluon distribution at medium momentum fraction values, $0.06\lesssim x\lesssim 0.22$, and at energy scale $Q^2=4(m_c^2+p_T^2)\sim 13$ GeV^2, where m_c is the charm quark mass and p_T is the transverse momentum of the produced charmed hadron, see Fig. 6.2.

High-p_T jet production at STAR and PHENIX. Both the STAR and PHENIX experiments at RHIC provided their measurements of the longitudinal double-spin asymmetry for inclusive jet production, Eq. (6.9). This is obtained by colliding two polarized proton beams at center-of-mass energy $\sqrt{s}=200$ GeV. We refer to [272] and references therein for a detailed description of the RHIC experimental setup. Data from STAR are available for the 2005 and 2006 runs and, in a preliminary form, also for the most recent 2009 run; only one set of data is available for PHENIX, corresponding to data taken in 2005. In the following, they will be referred to as STAR 1j-05, STAR 1j-06, STAR 1j-09 and PHENIX 1j respectively. The features of these data sets, including the number of data points $N_{\rm dat}$, jet-finding algorithm and corresponding cone radius R used for data reconstruction, covered range in integrated rapidity η and intergrated luminosity $\mathscr{L}$, are summarized in Tab. 6.1. Experimental correlations between systematic uncertainties are not provided. Within LO kinematics, jet data roughly cover the range $0.05\lesssim x\lesssim 0.2$ and $30\lesssim p_T^2\lesssim 800$ GeV^2, see Fig. 6.2.

W boson production at STAR. Both the STAR and PHENIX collaborations at RHIC presented first measurements of the parity-violating spin asymmetry $A_L^{W^\pm}$, Eq. (6.12), based on the 2009 run at $\sqrt{s}=500$ GeV. [236, 237]. Unfortunately, due to the low integrated luminosity ($\mathscr{L}=12$ pb^{-1} and $\mathscr{L}=8.6$ pb^{-1} for STAR and PHENIX respectively), these data will have little impact in determining antiquark flavors, once included in a polarized parton set. More interestingly, the STAR collaboration has recently presented preliminary results for the asymmetry $A_L^{W^\pm}$, based on data collected in 2012 at $\sqrt{s}=510$ GeV and with an integrated luminosity

$\mathscr{L} = 72$ pb^{-1} [274]. Two data sets are provided in six bins of the lepton rapidity η and at integrated lepton transverse momentum $25 < p_T < 50$ GeV, separately for W^+ and W^-: they will be referred to as STAR-W^+ and STAR-W^- henceforth. Given STAR kinematics, these data sets are likely to constrain light antiquark PDFs roughly in the momentum fraction interval $0.05 \lesssim x \lesssim 0.4$ and at the energy scale of the W mass (see Fig. 6.2). Besides uncorrelated statistical uncertainties, the measured asymmetries are provided with a correlated systematic uncertainty related to the uncertainty in the beam polarization, given as the 3.4% of the measured asymmetry. Other uncorrelated systematics, due to background and relative luminosity, are estimated to be less than 10% of statistical errors in the preliminar STAR analysis [274]. Even though these uncertainties are expected to become smaller in the final data released by STAR, we conservately assume them to be as large as 10%.

6.4 Determination of parton distributions

In this section, we illustrate how the `NNPDFpol1.0` parton set determined in Chap. 4 is supplemented with the piece of experimental information discussed above. Instead of a global refit of parton distributions including the new data, we will use PDF reweighting, followed by unweighting: this methodology was presented in Refs. [105, 109] and its main features were summarized in Sec. 3.2.4.

We recall that parton sets determined trough the NNPDF methodology are provided as Monte Carlo ensembles made of equally probable PDF replicas, each fitted to a data replica generated according to the uncertainties and the corresponding correlations measured in the experiments. The number of replicas in a given ensemble, $N_{\rm rep}$, is determined by requiring that the central values, uncertainties and correlations of the original experimental data can be reproduced to a given accuracy by taking averages, variances and covariances over the replica sample. For the case of polarized PDFs, in Sec. 4.1.2 we determined that a Monte Carlo sample of pseudodata with $N_{\rm rep} = 100$ replicas is sufficient to reproduce the mean values and the errors of experimental data to an accuracy which is better than 5%. Only a moderate improvement was observed in going up to $N_{\rm rep} = 1000$, hence our default choice was $N_{\rm rep} = 100$. The PDF ensemble forms an accurate representation of the underlying probability distribution of PDFs, conditional on the input data and the particular assumptions (such as the details of the QCD analysis) used in the fit. Based on statistical inference and Bayes theorem, PDF reweighting cosists in assigning to each replica in a Monte Carlo ensemble of PDFs, which is referred to as the *prior* ensemble, a weight which assesses the likelihood that this particular replica agrees with the new data. We refer to Sec. 3.2.4 for details about the way the weights are determined from the χ^2 of the new data to the prediction obtained using a given replica in the prior. The theoretical bases of the reweighting methodology were carefully checked in Refs. [105, 109] and it was also shown that results obtained via global refitting or reweighting with new data are statistically equivalent between each other.

The main limitation of the reweighting method is that the information brought in by new data should be only a moderate correction as compared to the information already included in the prior PDF ensemble. This is precisely our present case, since we will add to our fit a few dozens of new hadronic data points, as we discussed in Sec. 6.3. On the other hand, the reweighting methodology has the main advantage to avoid global refitting: in particular, for each PDF replica in the prior ensemble, the lenghty computation of observables has to be performed only once, instead of at each minimization step required in the case of refitting. This is of particular relevance for the polarized case, since the `FastKernel` method [104], used

in Sec. 4.2 to perform fast evolution and fast computation of inclusive DIS observables, has not yet been implemented for the polarized hadronic processes considered here. Alternatively, exact NLO calculations could be directly used in a global PDF fit by extending to polarized observables the `FastNLO` framework [275] and the general-purpose interface `APPLgrid` [276], but this too is not yet available.

The main steps of the procedure we follow to determine the `NNPDFpol1.1` parton set are described below.

6.4.1 Construction of the prior PDF ensemble

Our goal is to include the piece of new exerimental information discussed in Sec. 6.3 into the determination of polarized parton distributions presented in Chap. 4 via Bayesian reweighting. To this purpose, we have to compute the theoretical predictions for the measured observables, *i.e.* the longitudinal spin asymmetries Eqs. 6.4)-(6.9)-(6.12), based on PDF replicas in `NNPDFpol1.0`, which is then the prior ensemble. The χ^2 of the new data to the prediction obtained using a given replica in the prior is then used to compute the weight corresponding to each replica according to Eq. (3.16). Unfortunately, the `NNPDFpol1.0` parton set determined in Chap. 4 cannot be used to compute the theoretical predictions for the asymmetries straightforwardly: at NLO accuracy, this computation requires the knowledge of quark and antiquark distributions separately, which are not provided by the `NNPDFpol1.0` parton set, because it was determined from a fit to inclusive DIS data only.

A separation of the polarized quark and antiquark distributions can be achieved in a global fit including SIDIS data, as done in the `DSSV08` [75] and LSS10 [77] analyses. In principle, we could perform a new, global fit to inclusive and semi-inclusive DIS data and then use the corresponding Monte Carlo ensemble of parton distributions as the prior for the reweighting with collider data presented in Sec. 6.3. However, as already noticed many times in this thesis, the analysis of SIDIS data requires the usage of poorly known fragmentation functions, which may significantly affect the accuracy of our results. For consistency, we should determine a set of fragmentation functions based on the NNPDF methodology and then use this set to perform a fit including SIDIS data, but this is beyond the scope of the present analysis.

Be that as it may, we circumvent the issue related to the quark-antiquark separation in the prior using a different approach, which will be described here for the first time. The idea is to supplement the information available in the `NNPDFpol1.0` parton set for the Δu^+, Δd^+, Δs^+ and Δg distributions with some assumptions on $\Delta\bar{u}$ and $\Delta\bar{d}$ in order to construct a suitable prior ensemble. This is a sensible approach since different ansatz for the $\Delta\bar{u}$ and $\Delta\bar{d}$ prior distributions will lead to the same final result after including the new data by reweighting. Indeed, if new data brings in a sufficient amount of information, the final reweighted PDFs will be independent of the original choice of prior [172, 173]. Of course, it will be essential to show explicitly that this is what happens in our particular situation.

In principle, we could think of making the quark-antiquark separation for u and d flavors in the prior arbitrarily: for example, we could assign to each replica random values for its Δu, Δd, $\Delta\bar{u}$ and $\Delta\bar{d}$ distributions, provided their sum reproduces the corresponding total distributions determined in `NNPDFpol1.0` and they separately satisfy theoretical constraints (see Sec. 4.3.4). However, such a prior will lead to highly unefficient reweighting, in the sense that the new underlying PDF probability distribution after reweighting will be sampled by a small number of effective replicas (for details, see Sec. 3.2.4). In order to avoid this loss of efficiency, prior ensembles with a huge number of replicas should be produced, but this is extremely demanding

in terms of computational resources.

Alternatively, the additional information on quark-antiquark separation needed to construct suitable prior ensembles can be obtained from one of the aforementioned fits to SIDIS data. Furthermore, we can allow for deviations from the corresponding best-fit $\Delta\bar{u}$ and $\Delta\bar{d}$ determinations, by supplementing them with additional statistical noise and uncertainties, until the independence of the reweighted results from the prior is achieved. The loss of consistency of these priors will be under control, thus not requiring to be huge, as we will demonstrate below.

We now discuss in practice how we construct suitable prior ensembles to be afterwards reweighted with the new data discussed in Sec. 6.3. We supplement the `NNPDFpol1.0` parton set with the information on $\Delta\bar{u}$ and $\Delta\bar{d}$ distributions from the `DSSV08` parton fit [75], which includes all available SIDIS data, and we construct a collection of independent prior PDF ensembles. First of all, we sample the `DSSV08` $\Delta\bar{u}$ and $\Delta\bar{d}$ parton distributions at a fixed reference scale $Q_0^2 = 1$ GeV^2. We select ten points, half logaritmically and half linearly spaced in the interval of momentum fraction $10^{-3} \lesssim x \lesssim 0.4$, which roughly corresponds to the range covered by SIDIS experimental data relevant for separating quark-antiquark contributions. We sample four independent sets of data points assuming the `DSSV08` best fit plus one, two, three or four times its nominal $\Delta\chi^2 = 1$ Hessian uncertainty. Separate prior PDF ensembles, labelled as 1σ, 2σ, 3σ and 4σ henceforth, will then be constructed for each one of these data sets. Of course, a different ansatz could be made on data: the rationale we followed was to increase their uncertainty until independence of the reweighted results from the prior was reached. We will explicitly show that this requirement is fulfilled at least by the 3σ and 4σ prior ensembles at the end of this Section.

Data points sampled from the `DSSV08` fit are then treated, separately for $\Delta\bar{u}$ and for $\Delta\bar{d}$, as sets of *experimental* pseudo-observables. Henceforth, they will be labelled as $\texttt{DSSV08}_U$ and $\texttt{DSSV08}_D$ respectively. More precisely, we generate $N_{\rm rep} = 1000$ replicas of of the original `DSSV08` pseudodata, following the procedure described in Sec. 4.1, and then for each individual replica we perform a neural network fit to them: the result gives the $\Delta\bar{u}$ and $\Delta\bar{d}$ distributions with wich we supplement the `NNPDFpol1.0` parton set. In order to meaningfully fit the $\Delta\bar{u}$ and $\Delta\bar{d}$ pseudodata, we need to supplement the input PDF basis given in Sec. 4.3, namely $\Delta\Sigma$, ΔT_3, ΔT_8 and Δg, with two new linearly independent light quark combinations; we choose them to be the total valence, ΔV, and the valence isotriplet, ΔV_3,

$$\Delta V(x,Q_0^2) = \Delta u^-(x,Q_0^2) + \Delta d^-(x,Q_0^2), \tag{6.15}$$

$$\Delta V_3(x,Q_0^2) = \Delta u^-(x,Q_0^2) - \Delta d^-(x,Q_0^2), \tag{6.16}$$

where $\Delta q^- = \Delta q - \Delta\bar{q}$, $q = u,d$. In addition, in Eqs. (6.15)-(6.16) we have assumed that $\Delta s = \Delta\bar{s}$. Even though data are presently insufficient to discriminate between any guess on strage-antistrange distributions and there are actually no theoretical motivations to support a symmetric polarized strangeness, we adopt the choice $\Delta s = \Delta\bar{s}$ as it is usual in all polarized PDF analyses. We emphasize that the distribution which is physically meaningful is instead the total strange combination Δs^+ which was already determined from inclusive DIS data in `NNPDFpol1.0`, see Chap 4.

Each of the new PDF combinations in Eqs. (6.15)-(6.16) is parametrized as usual by means of a neural network supplemented with a preprocessing polynomial,

$$\Delta V(x,Q_0^2) = (1-x)^{m_{\Delta V}} x^{n_{\Delta V}} {\rm NN}_{\Delta V}(x), \tag{6.17}$$

$$\Delta V_3(x,Q_0^2) = (1-x)^{m_{\Delta V_3}} x^{n_{\Delta V_3}} {\rm NN}_{\Delta V_3}(x), \tag{6.18}$$

PDF	m	n
$\Delta V(x,Q_0^2)$	[1.5,3.0]	[0.05,0.60]
$\Delta V_3(x,Q_0^2)$	[1.5,3.0]	[0.01,0.60]

Table 6.2: Ranges for the small and large x preprocessing exponents in Eqs. (6.17)-(6.18).

where $\mathrm{NN}_{\Delta\mathrm{pdf}}$, pdf $= V, V_3$, is the output of the neural network and the preprocessing exponents m, n are linearly randomized for each Monte Carlo replica within the ranges given in Tab. 6.2. We have checked that our choice of preprocessing exponents does not bias the fit, according to the procedure discussed in Sec. 4.3. The neural network architecture is the same as in `NNPDFpol1.0`, namely 2-5-3-1, see Sec. 3.2.2. Note that in terms of the quark PDF input basis, the $\Delta\bar{u}$ and $\Delta\bar{d}$ distributions which here play the role of pseudo-data are given by the following combinations:

$$\Delta\bar{u}(x,Q_0^2) = \frac{1}{12}\left(2\Delta\Sigma + 3\Delta T_3 + \Delta T_8 - 3\Delta V - 3\Delta V_3\right)(x,Q_0^2), \tag{6.19}$$

$$\Delta\bar{d}(x,Q_0^2) = \frac{1}{12}\left(2\Delta\Sigma - 3\Delta T_3 + \Delta T_8 - 3\Delta V + 3\Delta V_3\right)(x,Q_0^2). \tag{6.20}$$

For each pseudodata replica $\Delta\bar{u}^{(k)}$, $\Delta\bar{d}^{(k)}$ in Eqs. (6.19)-(6.20), $k = 1,\ldots,N_{\mathrm{rep}}$, we supplement the neural networks for ΔV and ΔV_3 that are being fitted with random replicas from `NNPDFpol1.0`. All the prior PDF ensembles are composed of $N_{\mathrm{rep}} = 1000$ replicas; this larger number of PDF members, in comparison to that used in the analysis presented in Chap. 4, where $N_{\mathrm{rep}} = 100$, is required to ensure that replicas left after reweighting still describe the underlying PDF probability distribution with sufficient accuracy; $N_{\mathrm{rep}} = 1000$ replicas of `NNPDFpol1.0` were generated with this purpose.

In these fits to pseudodata, minimization is performed by means of a genetic algorithm, as discussed in Secs.3.2.3-4.3.2. Implementation of theoretical constraints, both positivity and integrability, also follow consistently the procedure from Ref. [116] in order to take care of flavor and antiflavor separation. In particular, Eqs. (4.23)-(4.24) have been enforced by letting $f = u, \bar{u}, d, \bar{d}$ separately. Note that no additional sum rules affect ΔV and ΔV_3.

Following this procedure, we end up with four separate prior PDF ensembles, labeled as 1σ, 2σ, 3σ and 4σ, corresponding to the different factors by which the `DSSV08` nominal PDF uncertainty has been enlarged. The goodness of the pseudodata fits is quantitatively assessed by the χ^2 values per data point quoted in Tab. 6.3, which are close to one for both separate and combined $\mathtt{DSSV08}_U$ and $\mathtt{DSSV08}_D$ data sets. In Fig. 6.3, we show the $x\Delta\bar{u}(x,Q_0^2)$ and $x\Delta\bar{d}(x,Q_0^2)$ PDFs at the initial energy scale $Q_0^2 = 1$ GeV^2 from all the prior ensembles we have contructed. In these plots the positivity bound discussed in Sec. 4.3 and data points sampled from the `DSSV08` parton set [75] are also shown.

Monte Carlo ensembles of polarized parton distributions obtained in this way are equivalent to `NNPDFpol1.0` in the $\Delta q^+ = \Delta q + \Delta\bar{q}$ ($q = u, d, s$) and Δg sectors, but they are supplemented with quark-antiquark separation, suitable as a starting point for the reweighting procedure. Notice that they have exactly the same gluon distribution, since this is not affected by construction. In the next sections we quantify the impact on this prior of the new hadronic data, and show that results are independent of the specific choice of prior starting from the 3σ case.

			$\chi^2_{\rm tot}$			
Experiment	Set	$N_{\rm dat}$	1σ	2σ	3σ	4σ
DSSV08		20	1.04	1.10	1.09	0.97
	DSSV08$_U$	10	1.13	1.09	1.08	0.97
	DSSV08$_D$	10	0.96	1.10	1.09	0.96

Table 6.3: The value of the $\chi^2_{\rm tot}$ per data point for both separate and combined $\Delta\bar{u}$ and $\Delta\bar{d}$ data sets after the neural network fit to pseudodata sampled from DSSV08 parton set.

6.4.2 Reweighting with new data sets

We would like to reweight the prior PDF ensembles determined in Sec. 6.4.1 with the data described in Sec. 6.3. To this purpose, we should compute theoretical predictions for the observables measured in each process under investigation and then compare them to experimental data. As explained in Sec. 3.2, we will then assign to each replica in the PDF ensembles a weight proportional to the χ^2 of the new data to the corresponding prediction, given by Eq. (3.16).

Before discussing the impact of each new data set in turn, we summarize a few methodological aspects which apply to all experiments.

- For each of the new processess considered in our analysis, the experimental observable is a longitudinal spin asymmetry, *i.e.* the ratio between cross-sections depending on polarized PDFs (in the numerator) and on unpolarized PDFs (in the denominator), see Eqs. (6.4)-(6.9)-(6.12). In these expressions, the numerator is computed for each polarized replica from the ensembles determined in Sec. 6.4.1, while the denominator is evaluated only once using the mean value from the unpolarized NNPDF2.3 parton set [110] at NLO. This strategy accounts for the fact that the uncertainty affecting the observed asymmetries is mostly driven by the uncertainty of the polarized parton distributions rather than by that of their unpolarized counterparts, which is actually negligible.

- For each data set we will have to check the effectiveness of the reweighting procedure. To this purpose, we will look at the distribution of χ^2 per data point among replicas and at its mean value, which is expected to decrease after reweighting. Also, we will have to keep under control the PDF loss of accuracy in the description of the underlying probability distribution, by ensuring that the number of replicas left after reweighting, $N_{\rm eff}$, does not become too low. In particular, we require that $N_{\rm eff}$ should be comparable with the number of replicas, $N_{\rm rep}$, in NNPDFpol1.0, *i.e.* $N_{\rm eff} \sim N_{\rm rep} = 100$. Indeed, we determined in Sec. 4.1.2 that a Monte Carlo sample of $N_{\rm rep} = 100$ replicas is sufficient to reproduce the mean values and the errors of experimental data within percent accuracy.

- We may be interested in determining whether the new collider data are consistent with the old incluisve DIS data. To this purpose, for each data set we will evaluate the $\mathscr{P}(\alpha)$ distribution, defined by Eq. (12) of Ref. [105]. The parameter α measures the consistency of the data which are used for reweighting with those included in the prior PDF sets, by providing the factor by which the uncertainty on the new data must be rescaled in order the two sets to be consistent. Hence, if the probability density for the parameter α,

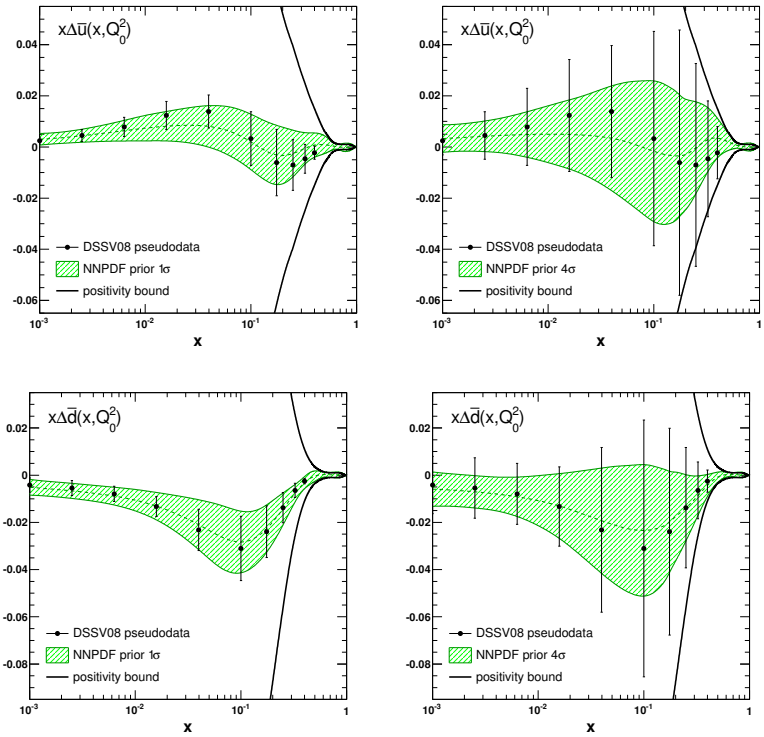

Figure 6.3: The polarized sea quark distributions $x\Delta\bar{u}(x,Q_0^2)$ (upper plots) and $x\Delta\bar{d}(x,Q_0^2)$ (lower plots) at the initial energy scale $Q_0^2 = 1$ GeV^2 from the neural network fit (full band) to pseudodata sampled from `DSSV08` parton set (points with uncertainties). Results are shown for the 1σ (left plots) and 4σ (right plots) prior ensembles. The positivity bound from the corresponding unpolarized `NNPDF2.3` parton set is also shown.

$\mathscr{P}(\alpha)$, peaks close to one, one can conclude that new and old data are consistent to each other.

Open-charm production at COMPASS

Predictions for the photon-nucleon asymmetry $A_{LL}^{\gamma N\to D^0 X}$ are computed at LO accuracy, using the expressions in Ref. [197], based in turn on Ref. [244], for both the numerator and the denominator in Eq. (6.4). Results are compared in Fig. 6.4 to COMPASS experimental data, separated into individual decay channels and into three bins of the charmed hadron energy E_{D^0}. The curves labelled as DSSV, AAC and BB are obtained using the corresponding polarized parton sets [73, 75, 76] and either the CTEQ6 [165] (for DSSV) or the MRST2004 [277] (for AAC and BB) unpolarized PDF sets. The curve labelled as NNPDF is instead computed using the `NNPDFpol1.0` parton set. Notice that we do not need the prior Monte Carlo ensembles constructed in Sec. 6.4.1, since we compute the observable at LO and only the polarized gluon appears in Eq. (6.4). By construction, this distribution is exactly the same in both the parton

Experiment	Set	$N_{\rm dat}$	$\chi^2/N_{\rm dat}$ NNPDF	DSSV	AAC	BB
COMPASS		45	1.23	1.23	1.27	1.25
	COMPASS $K1\pi$	15	1.27	1.27	1.43	1.38
	COMPASS $K2\pi$	15	0.51	0.51	0.56	0.55
	COMPASS $K3\pi$	15	1.90	1.90	1.81	1.82

Table 6.4: Values of $\chi^2/N_{\rm dat}$ before reweighting for different polarized parton sets.

sets discussed in Sec. 6.4.1 and in `NNPDFpol1.0`, hence they all give the same prediction for the photon-nucleon asymmetry $A_{LL}^{\gamma N\to D^0X}$. For all the curves shown in Fig. 6.4, we have used the Peterson parametrization of the fragmentation function $D_c^{D^0}$ [278]; we checked that results are unaffected by the choice of other, slightly different, available parametrizations [279], as noticed in Ref. [252].

The agreement between predictions from different PDF sets and experimental data can be quantified by the value of the χ^2 per data point, which we compute for both separate and combined COMPASS data sets, see Tab. 6.4. The corresponding distribution among NNPDF replicas is shown in the first panel of Fig. 6.5 only for the combined data set. Notice that, lacking the experimental covariance matrix, the uncertainties which enter the χ^2 definition are taken as the sum in quadrature of statistical and systematic uncertainties. From Tab. 6.4, it is clear that, even before reweighting, predictions are already in good agreement with experimental data, which however are affected by large errors in comparison to the uncertainty estimate of the asymmetry itself.

In Tab. 6.5, we quote the χ^2 per data point after reweighting, $\chi^2_{\rm rw}/N_{\rm dat}$, while the number of effective replicas left after reweighting, $N_{\rm eff}$, and the modal value of the $\mathscr{P}(\alpha)$ distribution, $\langle\alpha\rangle$, for both separate and combined data sets are collected in Tab. 6.6. In Fig. 6.5, we also plot the distribution of χ^2 per data point before and after reweighting, and of $\mathscr{P}(\alpha)$ for all data sets combined together.

From Tabs. 6.5-6.6 and Fig.6.5, it is evident that reweighting with COMPASS open-charm data leaves the prior parton set almost unaffected: the χ^2 value per data point and its distribution are essentially unchanged after reweighting. Also, almost all replicas in the prior ensemble are preserved: this further demonstrates the mild constraining power of COMPASS data sets. Furthermore, the observable $A_{LL}^{\gamma N\to D^0X}$, Eq. (6.4), is compared before and after reweighting in Fig. 6.6, showing unnoticeable differences. Finally, the reweighted polarized gluon PDF is drawn in Fig. 6.7 together with its one-sigma absolute error and compared to `NNPDFpol1.0` [116] at $Q_0^2 = 1$ GeV^2. Again, we notice that the differences are mild.

From these results we can conclude that COMPASS open-charm data lead to a moderate improvement in our knowledge of the polarized gluon PDF, due to the large experimental uncertainties. One may wonder whether this conclusion still holds once NLO corrections are taken into account in the computation of the photon-nucleon asymmetry, $A_{LL}^{\gamma N\to D^0X}$. Such a computation has been recently completed [252] and it was shown that contributions beyond Born-level may significantly affect this asymmetry. However, COMPASS experimental data are not only affected by large uncertainties with respect to the corresponding theoretical predictions (both at LO and NLO), but also do not show a clear and stable trend over the covered range of $p_T^{D^0}$, see Figs. 6.4-6.6. For this reason, the NLO computation of the photon-nucleon spin asymmetry will not allow for determining the polarized gluon more precisely than its LO

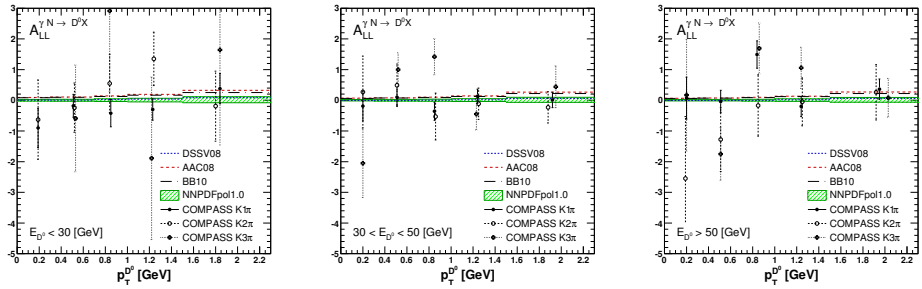

Figure 6.4: Experimental double-spin asymmetry for D^0 meson photoproduction $A_{LL}^{\gamma N \to D^0 X}$ measured by COMPASS [213] from three decay channels compared to its LO prediction, Eq. (6.4), computed for different PDF sets in three bins of the charmed hadron energy E_{D^0} and in five bins of its transverse momentum $p_T^{D_0}$.

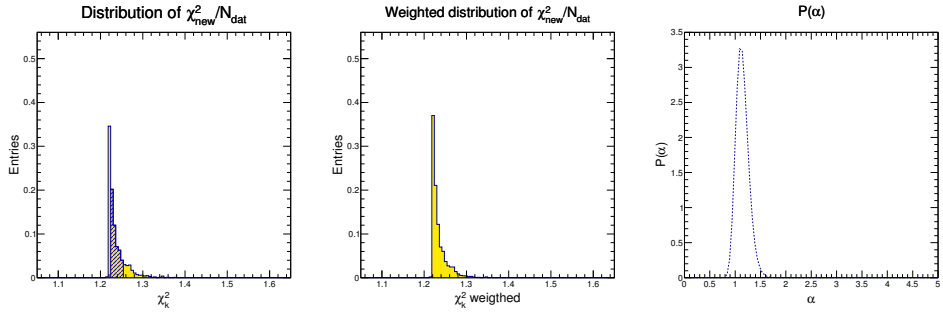

Figure 6.5: Distribution of χ^2/N_{dat} for individual replicas before (first panel) and after (second panel) reweighting with COMPASS open-charm production data [213]. The shaded region in the first panel corresponds to the central 68% of the distribution. The $\mathscr{P}(\alpha)$ distribution (third panel) is also shown. All plots refer to the three COMPASS data set combined together.

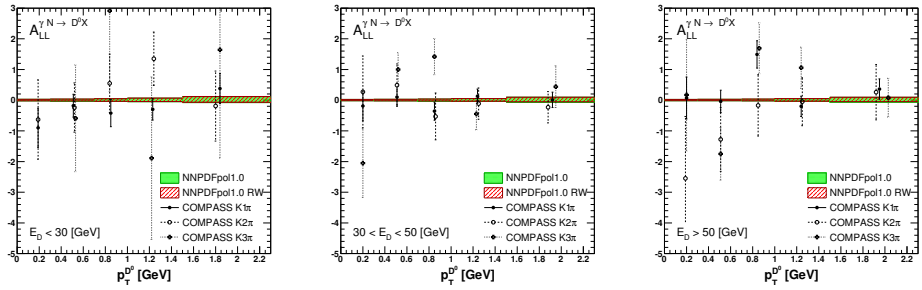

Figure 6.6: Comparison of the double-spin asymmetry $A_{LL}^{\gamma N \to D^0 X}$, Eq. (6.4), before and after reweighting with COMPASS open-charm data [213]. Experimental points are also shown.

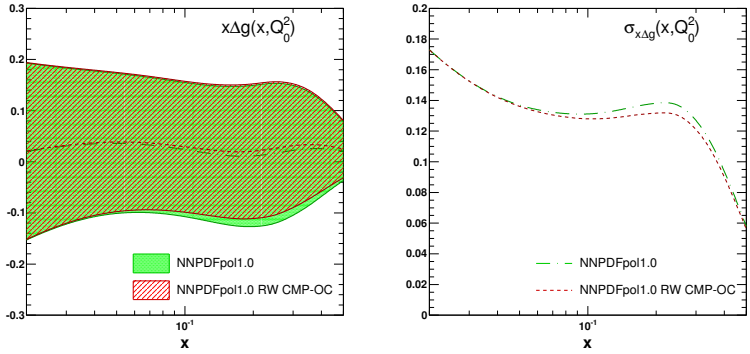

Figure 6.7: Comparison between the unweighted and the reweighted polarized gluon distribution at $Q_0^2 = 1$ GeV^2 (left panel) and the improvement in its absolute error (right panel).

counterpart, though in principle the former contains more QCD structure than the latter. Hence, we expect that the impact of COMPASS open-charm data on the determination of the polarized gluon PDF will be comparable to that found in our LO analysis, once they will be included in a global NLO QCD fit of parton distributions, as also anticipated in Ref. [252]. *

High-p_T jet production at STAR and PHENIX

Predictions for the longitudinal double-spin asymmetry from single-inclusive jet production, $A_{LL}^{1\,jet}$, are shown in Fig. 6.8. They are plotted as a function of the transverse jet momentum, p_T, and compared to each experimental data set in Tab. 6.1. Predictions are computed at NLO using the code of Ref. [267], which was modified to handle NNPDF parton sets. We use the same jet-algorithm, cone radius and kinematic cuts adopted in the experiment (see Tab. 6.1). In Fig. 6.8, results are shown only for the 1σ prior PDF ensemble. We have explicitly checked their stability upon the choice of any prior PDF ensemble discussed in Sec. 6.4.1: the asymmetry shows hardly noticeable differences among different priors, thus proving it is not sensitive to quark-antiquark separation, as expected.

The χ^2 per data point before reweighting for separate and combined data sets is quoted in Tab. 6.5. Neither STAR nor PHENIX jet data are provided with experimental covariance matrix, hence we will assume the systematics to be uncorrelated and then sum in quadrature with statistical errors. Also, we have to account for the fact that data are taken in bins of p_T, whereas the corresponding theoretical predictions are computed for the center of each bin. We estimate the corresponding uncertainty as the maximal variation of the observable within each bin and take that value as a further uncorrelated systematic uncertainty. Since STAR data are provided with asymmetric systematic uncertainties, we must take care of symmetrizing them, according to Eqs. (7)-(8) in Ref. [99].

*We were not able to compute the photon-nucleon asymmetry, $A_{LL}^{\gamma N\to D^0 X}$, at NLO accuracy because the code developed to this purpose in Ref. [252] is not publicly available. Qualitatively, the impact of the NLO corrections on this asymmetry can be inferred by comparing our results in Fig. 6.4 (at LO) with those in Fig. 8 of Ref. [252] (at NLO): differences are actually hardly noticeable.

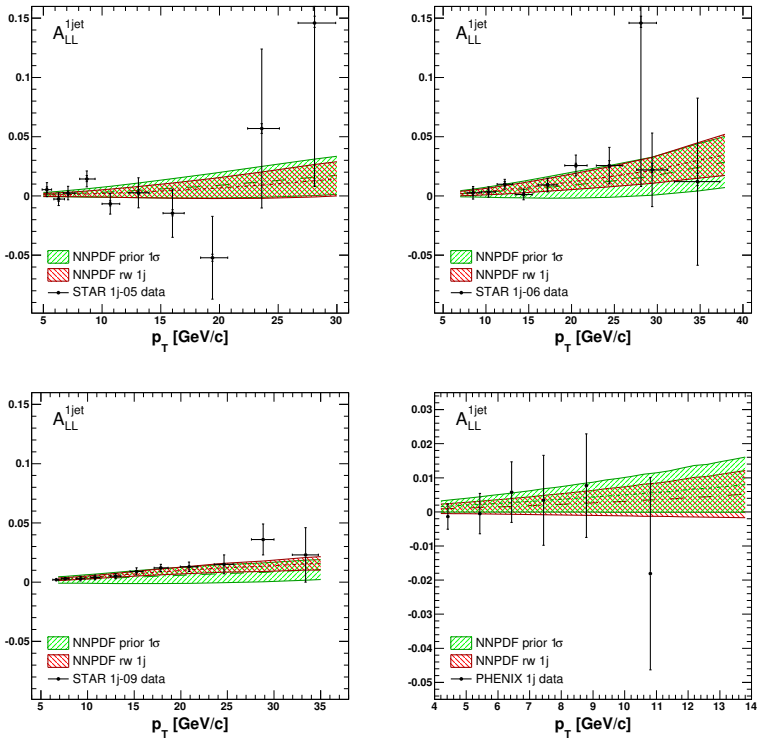

Figure 6.8: Predictions for the double longitudinal spin asymmetry for single-inclusive jet production, $A_{LL}^{1\,jet}$, before and after reweighting with RHIC data. Results are obtained from the 1σ prior PDF ensemble discussed in Sec. 6.4.1. Experimental data points are also drawn. Notice the different scale of vertical axis for PHENIX.

We observe that our predictions are in good agreement with experimental data, as $\chi^2/N_{\rm dat} \sim 1$ for all data sets. Nevertheless, notice that, except for STAR 1j-09 data set, experimental data points are affected by rather large errors in comparison to the uncertainty on the observable due to the PDFs: hence, this will limit their potential in constraining the polarized gluon distribution. Furthermore, similar results are obtained with different choices of the prior PDF ensemble: this strengthens the conclusion that we are looking at a process which is almost insensitive to the quark content of the proton.

Next, we reweight the various prior ensembles with both separate and combined RHIC inclusive jet production data. In Tab. 6.5, we quote the values for the χ^2 per data point after reweighting, $\chi^2_{\rm rw}/N_{\rm dat}$, while the number of effective replicas, $N_{\rm eff}$, and the modal value of the $\mathscr{P}(\alpha)$ distribution, $\langle\alpha\rangle$, are collected in Tab. 6.6. In Fig. 6.9, we display, for combined jet data sets, the unweighted distribution of $\chi^2/N_{\rm dat}$, the corresponding weighted distribution of $\chi^2_{\rm rw}/N_{\rm dat}$, and the $\mathscr{P}(\alpha)$ distribution. In Fig. 6.8 we finally compare the asymmetry before and after reweighting.

It is clear from Tab. 6.5 and Figs. 6.8-6.9 that jet data from RHIC carry an important piece of experimental information. In particular, we observe substantial improvement in the description of the high precision STAR 1j-09 data set, for which the value of the $\chi^2/N_{\rm dat}$ decreases from

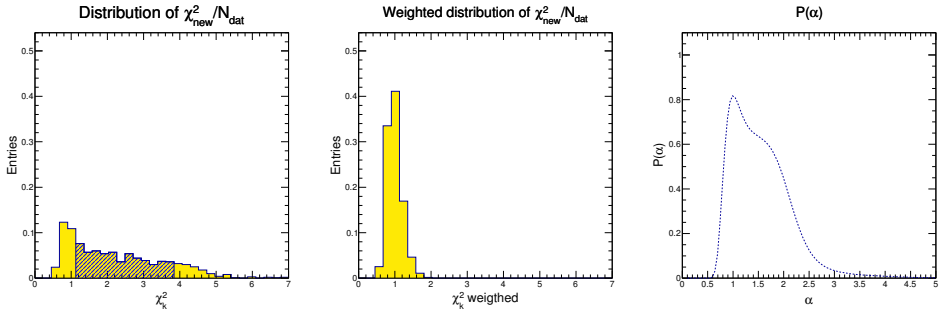

Figure 6.9: Same as Fig. 6.5, but for combined RHIC jet data and for the 1σ prior discussed in Sec. 6.4.1.

1.69 to 1.02. As we will show below, this improvement translates into significant constraints on the polarized gluon PDF. Moreover, the comparison between the χ^2 distributions before and after reweighting shows that their peak moves close to one, with a slight narrowing due to the increase in the total number of data points. The reweighted observable nicely agrees with experimental data and its uncertainty is reduced with respect to the prior. Also, notice that the PDF set loss of accuracy in describing the underlying probability distribution is well under control: the number of replicas left after reweighting, $N_{\rm eff}$, is always larger than $N_{\rm rep} = 100$, the value roughly required for `NNPDFpol1.0` to reproduce data and errors with percent accuracy. Finally, new data sets are consistent with the experimental information already included in the prior, as shown by the $\mathscr{P}(\alpha)$ distribution, which is clearly peaked at one (see Fig. 6.9).

Nevertheless, we notice that different data sets have different power in improving the fit: the more accurate they are, the more effective they are at constraining the theoretical prediction for the asymmetry A_{LL}^{1jet} and, in turn, the polarized gluon distribution. It is clear from Tabs. 6.5-6.6 that a substantial amount of experimental information comes from the STAR 1j-09 data set. We also observe that PHENIX data have a fair impact in improving the fit, due to their large errors: it is likely that they were overestimated, since the modal value of the α parameter is far below one (see Tab. 6.6).

Finally, in Fig. 6.10 we compare the polarized gluon PDF from the `NNPDFpol1.0` parton set [116] with its counterpart from the 1σ prior reweighted with RHIC jet data. We also show the corresponding one-sigma absolute error and a comparison between the reweighted results obtained with extremum prior PDF sets, namely 1σ and 4σ. We observe that, in the kinematic range probed at RHIC, the polarized gluon PDF becomes positive and its uncertainty is reduced. This effect is consistent with what was found in Ref. [188], where these data were included in the `DSSV++` parton set. Again, we stress that results are independent of the choice of any prior PDF ensemble discussed in Sec. 6.4.1: the stability of our results is clearly visible from both Tabs. 6.5-6.6 and from the third panel of Fig. 6.10. Hence, results obtained from the reweighting of any prior are equivalent, and can be used indifferently to describe the polarized gluon PDF. We have explicitly checked that other PDFs, not shown in Fig. 6.10 are unaffected by jet data, as expected.

W boson production at STAR

Predictions for the longitudinal single-spin asymmetry, Eq. (6.12), are computed at NLO using the program of Ref. [270], which was consistently modified to handle NNPDF parton

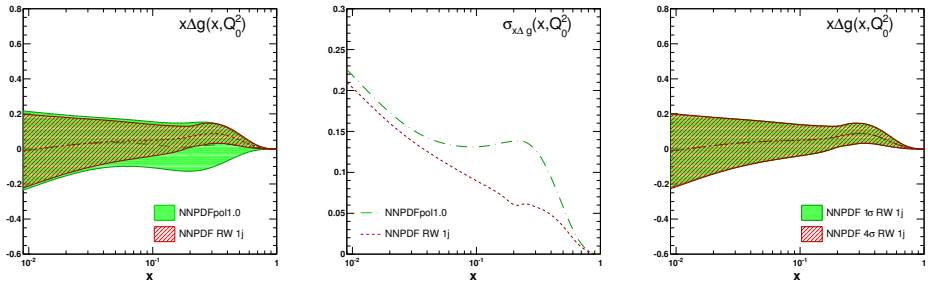

Figure 6.10: (First panel) Comparison between the polarized gluon PDF from `NNPDFpol1.0` parton set [116] and the result obtained by reweighting with RHIC jet data. (Second panel) the same comparison, but for absolute PDF uncertainty. (Third panel) Comparison between the reweighted gluon PDF obtained from the 1σ and the 4σ prior PDF ensembles.

sets. In Fig. 6.11, we show our predictions for the longitudinal positron (electron) single spin asymmetry $A_L^{e^+}$ ($A_L^{e^-}$) from production and decay of $W^{+(-)}$ bosons in bins of the rapidity η, compared to STAR data [274]. Results are diplayed for the 1σ and 4σ prior PDF ensembles, and we have checked that intermediate results are obtained from the other priors constructed in Sec. 6.4.1. Unlike open-charm and jet production observables discussed above, the longitudinal single spin asymmetry, Eq. (6.12), is sensitive to separate quark and antiquark PDFs. For this reason, the corresponding theoretical predictions made from the 4σ prior are affected by larger uncertainties than those obtained from the 1σ prior.

In Tab. 6.5, we summarize the agreement between experimental data and parton sets before reweighting, as quantified by the value of $\chi^2/N_{\rm dat}$: as we can see, it is not very good, and all prior PDF sets fail to describe these data sets with sufficient accuracy. The discrepancy between data and corresponding theoretical predictions is particularly noticeable for the W^- set. This may reflect some tensions between semi-inclusive data, included in the `DSSV08` global fit, and W production data. Indeed, the experimental information included in `DSSV08` was inherited by NNPDF prior ensembles, due to the way they were constructed. After reweighting with STAR data, we should demonstrate that this dependence has been removed and check that reweighted PDFs properly describe $W^\pm$ data sets.

As for the other observables, we proceed to reweight the different prior ensembles with STAR W production data. Results are collected in Tabs. 6.5-6.6 for both separate and combined STAR data sets, and compared to the analogous quantity before reweighting. In Fig. 6.12 we display, for each prior PDF ensemble, and for combined W^+ and W^- STAR data sets, the unweighted distribution of $\chi^2/N_{\rm dat}$, the corresponding weighted distribution of $\chi^2_{\rm rw}/N_{\rm dat}$, and the $\mathscr{P}(\alpha)$ distribution. Finally, in Fig. 6.11 we compare the asymmetry before and after reweighting.

We see that after the reweighting, our predictions are in perfect agreement with the STAR data, both for the W^+ and for the W^- final states. In particular, we observe that the $\chi^2/N_{\rm dat}$ value decreases after reweighting (see Tab. 6.5), and that the χ^2 distribution tends to properly peak at one, being more narrow than before reweighting (compare, for each prior PDF ensemble, plots in the first and second column of Fig. 6.12). The value of the effective number of replicas after reweighting, $N_{\rm eff}$, is always larger than $N_{\rm rep} = 100$, thus the size of the initial prior sample was large enough. The modal value of the $\mathscr{P}(\alpha)$ distribution, $\langle\alpha\rangle$ is close to one for the

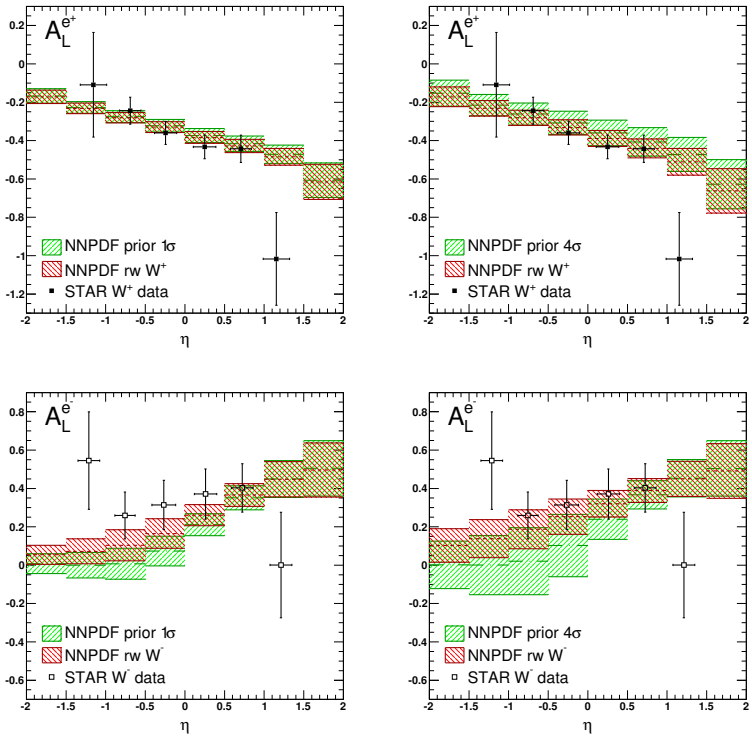

Figure 6.11: Predictions for the longitudinal positron (upper plots) and electron (lower plots) single spin asymmetry $A_L^{e^+}$ and $A_L^{e^-}$ before and after reweighting with STAR data [274]. Results from 1σ, 2σ, 3σ and 4σ prior PDF ensembles are shown from left to right. Curves are obtained at NLO with the CHE code [270]. Experimental points are also shown, uncertainties are statistical only.

W^+ data but almost 1.4 for the W^- data, suggesting that in the latter case some experimental uncertainties might be somewhat underestimated. This might also explain why the STAR-W^- data set has a much smaller value of $N_{\rm eff}$ than STAR-W^+, while we expected the two sets to have a similar impact. Finally, in Fig. 6.11 we show the comparison of the $A_L^{e^+}$ and $A_L^{e^-}$ asymmetries before and after the reweighting. Finally, the reweighted observable nicely agrees with experimental data and its uncertainty is reduced with respect to the prior, as clearly shown in Fig. 6.11.

As emphasized in Sec. 6.4.1, in order to get reliable results we must require the reweighted PDFs to be independent of the prior PDF ensemble. In other words, we should discard results which are not stable upon the choice of different prior PDF ensembles and, if needed, we should construct new priors, assuming a different ansatz on the quark-antiquark PDF separation, until this independence is effectively achieved. In this case, we can verify that both the 3σ and 4σ prior PDF ensembles lead to fully equivalent results, as can be seen at the level of the χ^2 value per data point after reweighting (see Tab. 6.5) and single-spin asymmetries (see Fig. 6.11). Then, we can conclude that results from both 3σ and 4σ prior PDF ensembles are statistically indistinguishable and that they can both be used to provide a robust determination of the light

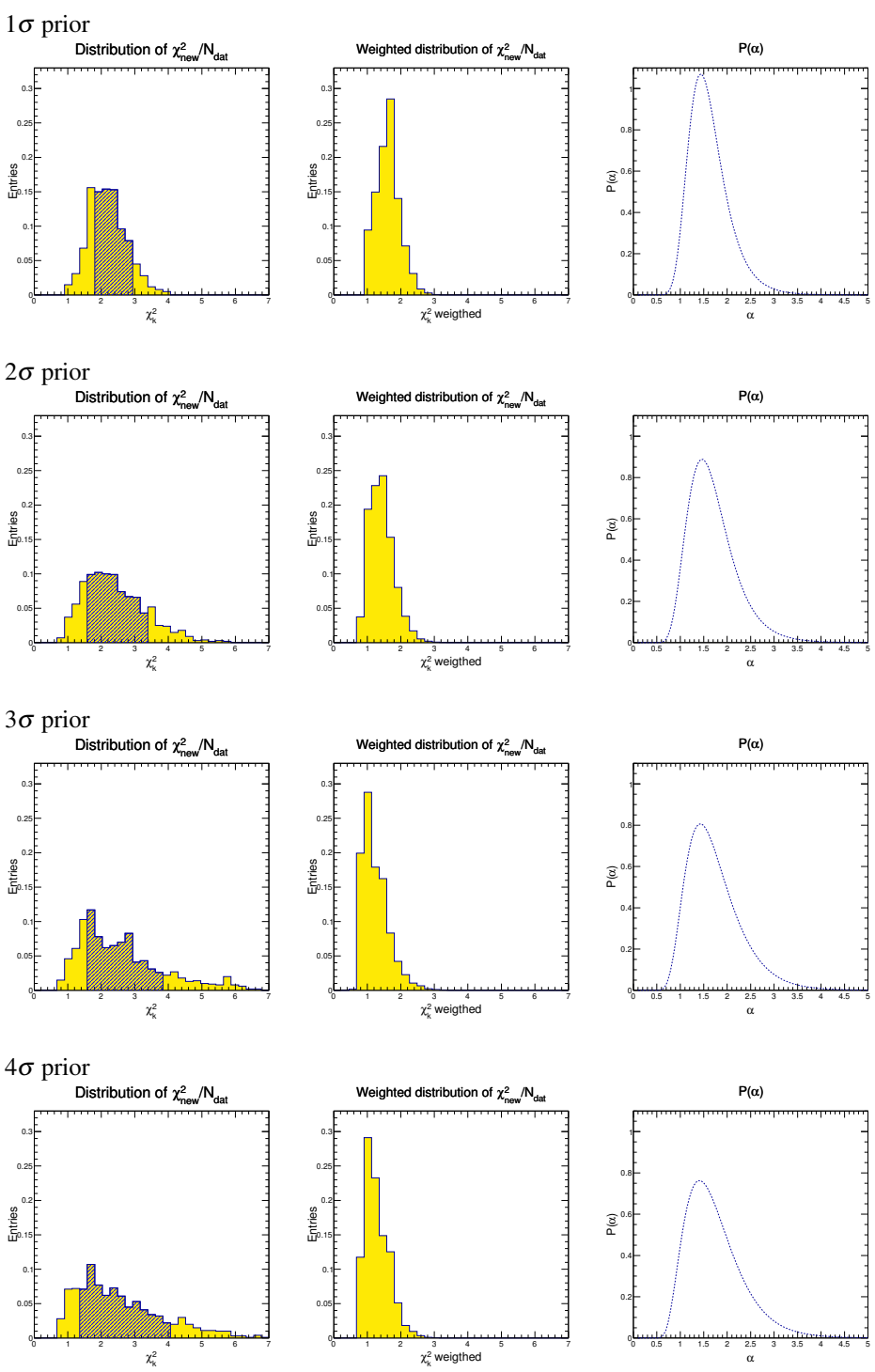

Figure 6.12: Same as Fig. 6.5, but for combined W^+ and W^- STAR data sets [274] and for each prior.

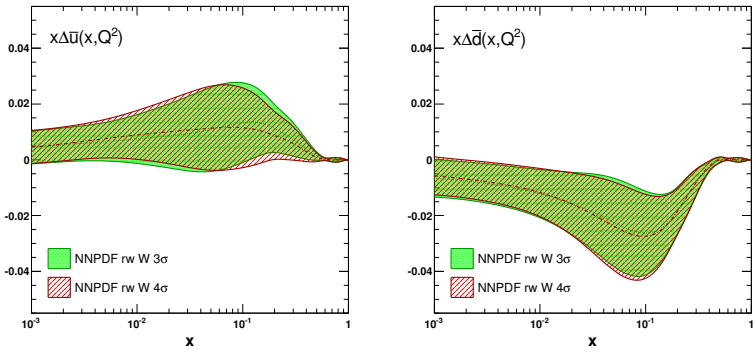

Figure 6.13: Comparison between the $\Delta\bar{u}$ ($\Delta\bar{d}$) PDF obtained from the reweighting of the 3σ and 4σ prior ensembles with STAR W data at $Q^2 = 10$ GeV^2.

antiquark PDFs, $\Delta\bar{u}$ and $\Delta\bar{d}$. The explicit comparison of these two distributions from the 3σ and 4σ samples is shown in Fig. 6.13 at $Q_0^2 = 1$ GeV^2.

Results displayed in Fig. 6.13 refer to simultaneous reweighting with both W^+ and W^- data sets; we have explicitly checked that reweighting with W^+ (W^-) data set separately probe $\Delta\bar{d}$ ($\Delta\bar{u}$) PDFs. Parton distributions not shown in Fig. 6.13, including strangeness, are not affected by reweighting with W data, as we have explicitly checked. The situation is rather different from the unpolarized case, where, instead, W production data also provide some information on strangeness. However, we notice that in the unpolarized case Drell-Yan data from both fixed-target experiments and colliders are available: in particular, the former come from E605 and E688 experiments at Fermilab [280–283], while the latter come from CDF and D0 experiments at the Tevatron [284–286] and from ATLAS, CMS and LHCb at the LHC [287–289]. These data sets span about three orders of magnitude in the energy scale Q^2, from $Q^2 \sim 20 - 250$ GeV^2 for fixed-target experiments, up to the W and Z masses for collider experiments. Hence, the effects of the evolution, enhanced by the rather wide Q^2 lever-arm of experimental data, and of the contributions from the Cabibbo-favored partonic subprocesses, initiated by a $s\bar{c}$ or $\bar{s}c$ pair, combine together in order to provide some constraints on the strangeness. Of course, since in the polarized case only data from collider are available, we cannot observe this effect.

Combining COMPASS, STAR and PHENIX data

The goal of the present analysis is to deliver a polarized parton set including the experimental information coming from the complete piece of information provided by data discussed in Sec. 6.3. To this purpose, we perform a global reweighting of our prior polarized PDF ensembles generated in Sec. 6.4.1 with all the relevant data from the COMPASS, STAR and PHENIX experiments simultaneously. In Tab. 6.5, we show the values of the χ^2 per data point before ($\chi^2/N_{\rm dat}$) and after ($\chi^2_{\rm rw}/N_{\rm dat}$) reweighting for each prior PDF ensemble. In Tab. 6.6, we also quote the number of effective replicas left after reweighting, $N_{\rm eff}$, and the modal value of the $\mathscr{P}(\alpha)$ distribution. In both Tabs. 6.5-6.6 the corresponding values for separate experiments and data sets are also provided.

As discussed in Sec. 6.4.1, we must retain only the results which are stable upon the

Experiment	Set	$N_{\rm dat}$	$\chi^2/N_{\rm dat}$				$\chi^2_{\rm rw}/N_{\rm dat}$			
			1σ	2σ	3σ	4σ	1σ	2σ	3σ	4σ
COMPASS		45	1.23	1.23	**1.23**	1.23	1.23	1.23	**1.23**	1.23
	COMPASS $K1\pi$	15	1.27	1.27	**1.27**	1.27	1.27	1.27	**1.27**	1.27
	COMPASS $K2\pi$	15	0.51	0.51	**0.51**	0.51	0.51	0.51	**0.51**	0.51
	COMPASS $K3\pi$	15	1.90	1.90	**1.90**	1.90	1.89	1.89	**1.89**	1.89
STAR		30	1.19	1.19	**1.20**	1.20	0.79	0.79	**0.79**	0.79
	STAR 1j-05	10	1.04	1.04	**1.04**	1.04	1.01	1.02	**1.02**	1.01
	STAR 1j-06	9	0.75	0.75	**0.75**	0.76	0.59	0.59	**0.59**	0.59
	STAR 1j-09	11	1.69	1.69	**1.70**	1.71	1.02	1.02	**1.03**	1.04
PHENIX										
	PHENIX 1j	6	0.24	0.24	**0.24**	0.24	0.24	0.24	**0.24**	0.24
STAR		12	1.93	1.97	**1.97**	1.91	1.43	1.12	**1.03**	1.02
	STAR-W^+	6	1.15	1.14	**1.22**	1.25	1.04	1.01	**1.02**	1.01
	STAR-W^-	6	2.71	2.80	**2.59**	2.43	1.78	1.25	**1.04**	1.02
GLOBAL REWEIGHTING		63	1.22	1.25	**1.25**	1.24	1.07	1.04	**1.02**	1.02

Table 6.5: The value of the χ^2 per data point $\chi^2/N_{\rm dat}$ ($\chi^2_{\rm rw}/N_{\rm dat}$) before (after) global reweighting with all data sets and for each prior PDF ensemble discussed in the text.

Experiment	Set	$N_{\rm dat}$	$N_{\rm eff}$				$\langle\alpha\rangle$			
			1σ	2σ	3σ	4σ	1σ	2σ	3σ	4σ
COMPASS		45	980	980	**980**	980	1.07	1.07	**1.07**	1.07
	COMPASS $K1\pi$	15	990	990	**990**	990	1.15	1.15	**1.15**	1.15
	COMPASS $K2\pi$	15	990	990	**990**	990	0.72	0.72	**0.72**	0.72
	COMPASS $K3\pi$	15	970	970	**970**	970	1.36	1.36	**1.36**	1.36
STAR		19	299	301	**297**	300	1.01	1.10	**1.02**	1.01
	STAR 1j-05	10	931	931	**930**	931	1.03	1.02	**1.05**	1.06
	STAR 1j-06	9	621	623	**622**	621	0.99	1.20	**0.92**	0.98
	STAR 1j-09	11	233	235	**235**	234	1.12	1.14	**1.10**	1.11
PHENIX										
	PHENIX 1j	6	740	740	**740**	741	0.55	0.50	**0.50**	0.55
STAR		12	395	335	**346**	350	1.45	1.43	**1.40**	1.39
	STAR-W^+	6	896	850	**739**	715	1.11	1.13	**1.15**	1.14
	STAR-W^-	6	337	298	**376**	412	1.60	1.47	**1.38**	1.35
GLOBAL REWEIGHTING		63	224	197	**177**	176	1.22	1.21	**1.23**	1.22

Table 6.6: The fraction of replicas left after reweighting, $N_{\rm eff}$, and the modal value of the $\mathscr{P}(\alpha)$ distribution. Results refer to global reweighting with all data sets and for each prior PDF ensemble discussed in the text.

choice of different prior PDF ensembles. Looking at the values of the χ^2 per data point after reweighting, we argue that this stability is achieved for results obtained starting from the 3σ and 4σ priors, see Tab. 6.5. Conversely, the results obtained from the reweighting of the 1σ and 2σ prior ensembles must be discarded, since they actually depend on the assumptions on quark-antiquark separation we made for constructing the corresponding priors. In order to quantitatively check that the results of the global reweighting of the 3σ and 4σ samples are indeed statistically equivalent, we compute the distances $d(x,Q^2)$ between the respective PDFs. We recall that the distance is a statistical estimator which has the value $d \sim 1$ when the

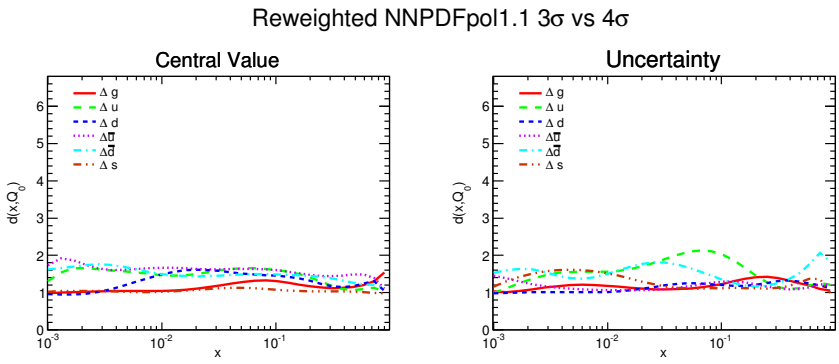

Figure 6.14: Distances between parton sets obtained via global reweighting of 3σ and 4σ prior PDF ensembles at $Q^2 = 10\ \mathrm{GeV}^2$.

two samples of $N_{\rm rep}$ replicas are extracted from the same underlying distribution, while it is $d = \sqrt{N_{\rm rep}}$ when the two samples are extracted from two distributions which differ on average by one standard deviation (see Sec. 4.4.2 and Appendix A for further details). The distances are plotted at $Q^2 = 10\ \mathrm{GeV}^2$ in Fig. 6.14. As $d \sim 2$, we conclude that the two ensembles obtained from the reweighting of the 3σ or 4σ priors describe the same underlying PDF probbaility distribution. We choose the reweighted set obtained from the 3σ prior for reference.

Furthermore, the overall agreement between new data and the corresponding theoretical predictions obtained with the reweighted PDF sets is very good, as quantified by the value of the χ^2 per data point, $\chi^2_{\rm rw}/N_{\rm dat} = 1.02$. The effective number of replicas is $N_{\rm eff} \sim 180$, so the size of the initial prior ($N_{\rm rep} = 1000$) was large enough even when the information from all data sets is simultaneously combined. The modal value of the $\mathscr{P}(\alpha)$ distribution, is $\langle\alpha\rangle \sim 1.2$, thus quantifying the good agreement between inclusive DIS data in `NNPDFpol1.0` and the new data included in `NNPDFpol1.1`.

6.4.3 Unweighting: the `NNPDFpol1.1` parton set

After global reweighting of the 3σ prior PDF ensemble, the unweighting procedure described in Sec. 3.2 is used to produce a polarized PDF set made of $N_{\rm rep} = 100$ replicas, `NNPDFpol1.1`, statistically equivalent to the reweighted PDF ensemble, but in which all PDF replicas are equally probable and hence do not need to be used with corresponding weights. In comparison to `NNPDFpol1.0`, the new polarized parton set provides a meaningful determination of sea flavor PDFs $\Delta\bar{u}$ and $\Delta\bar{d}$, though based on a small set of W boson production data (but with the advantage to be free of any bias, including poorly known fragmentation functions) and a determination of the gluon PDF Δg which is improved by open-charm and, particularly, jet data. In order to study the compatibility of the new data with the DIS sets, included in `NNPDFpol1.0`, in Tab. 6.7 we show the χ^2 of each of the experiments included in the `NNPDFpol1.0` analysis evaluated with both the old `NNPDFpol1.0` and the new `NNPDFpol1.1` parton sets. We observe that DIS data are described by the two parton sets with comparable accuracy, as we already noticed from the modal value of the $\mathscr{P}(\alpha)$ distribution, $\langle\alpha\rangle \sim 1.2$ (see Tab. 6.6).

In Fig. 6.15, we compare total PDF combinations $\Delta u + \Delta\bar{u}$, $\Delta d + \Delta\bar{d}$, $\Delta s + \Delta\bar{s}$ and gluon

Experiment	NNPDFpol1.0	NNPDFpol1.1
EMC	0.44	0.43
SMC	0.93	0.90
SMClowx	0.97	0.97
E143	0.64	0.67
E154	0.40	0.45
E155	0.89	0.85
COMPASS-D	0.65	0.70
COMPASS-P	1.31	1.38
HERMES97	0.34	0.34
HERMES	0.79	0.82

Table 6.7: The χ^2 per data point of all the experiments included in the NNPDFPol1.0 analysis evaluated with NNPDFpol1.0 and NNPDFpol1.1 parton sets

PDF Δg from NNPDFpol1.1 with NNPDFpol1.0 at $Q^2 = 10$ GeV^2. Since the latter is a fit to inclusive DIS data, only these PDFs can be compared meaningfully between the two parton sets. In order to quantitatively asses the difference between them, we plot the coreesponding distances $d(x,Q)$ at $Q^2 = 10$ GeV^2 in Fig. 6.16.

We observe that the total quark PDF combinations from the two determinations are statistically equivalent, since the distance for both their central value and uncertainty is not larger than two (Fig. 6.16) and differences between them are hardly noticeable (Fig. 6.15). On the other hand, the gluon PDF shows significant differences between the two NNDPF parton determinations, in particular in the x region probed by STAR jet data, $0.05 \lesssim x \lesssim 0.2$. In this region, the gluon from the NNPDFpol1.1 parton set is definitely positive and has a much reduced error band with respect to its NNPDFpol1.0 counterpart. The polarized gluon PDF in the two determinations actually samples different underlying probability distributions, which may differ up to one sigma, as the distance grows up to $d \sim 10$. At lower values of x, where no new data are included, the NNPDFpol1.0 and NNPDFpol1.1 are again statistically equivalent, as expected.

In Fig. 6.17, we compare PDFs from NNPDFpol1.1 with those from the global DSSV08 fit [75]: we display $x\Delta u$, $x\Delta d$, $x\Delta\bar{u}$, $x\Delta\bar{d}$, $x\Delta s$ and $x\Delta g$ at $Q^2 = 10$ GeV^2. Uncertainties are nominal one-sigma error bands for NNPDF, while they are Hessian uncertainties ($\Delta\chi^2 = 1$) for DSSV08. This choice was made in Ref. [75] with some caution, since it may lead to underestimated PDF ucertainties in some (x,Q^2) regions, particularly where constraints from experimental data are weak, see also Sec. 3.1.

The main conclusions of the comparison in Fig. 6.17 are the following.

- Consistent results are found in the two parton determinations for Δu and Δd PDFs, though the NNPDF uncertainties are slightly larger, especially at small-x values, where experimental data are lacking.

- The NNPDF polarized gluon PDF is in perfect agreement with its DSSV counterpart in the large-x region, $x \gtrsim 0.2$, where they show similar uncertainties. However, for $x < 0.2$, Δg has a node in the DSSV08 determination, while it is clearly positive from NNPDFpol1.1. This result is driven in particular by the most recent and precise jet production data from STAR (labelled as STAR 1j-09 above), which were not available at the time of the original DSSV08 analysis [75] shown in Fig. 6.17. Actually, only the data sets labelled as STAR 1j-05 and STAR 1j-06 were included there. An update of

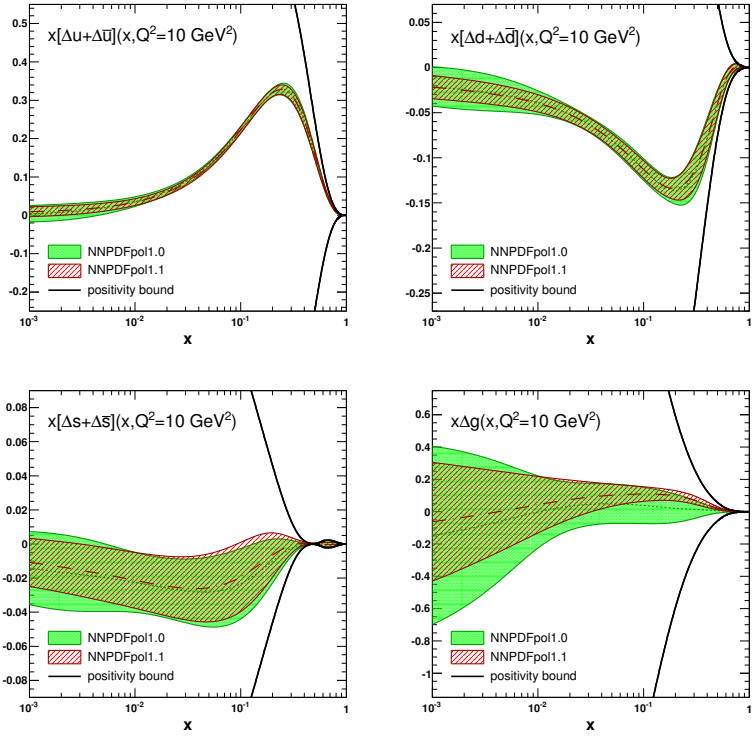

Figure 6.15: Comparison between `NNPDFpol1.0` and `NNPDFpol1.1` parton sets at $Q^2 = 10$ GeV^2.

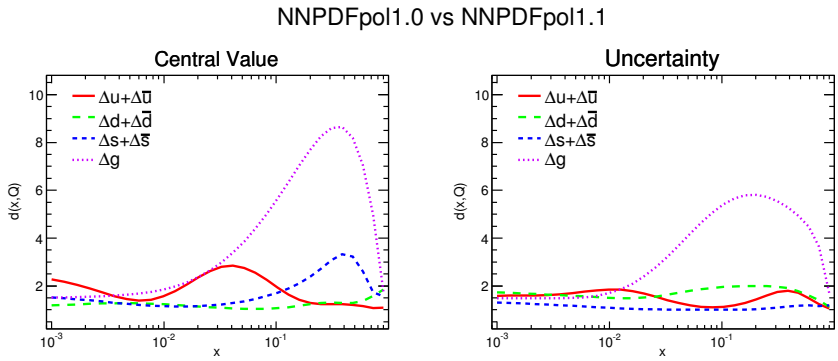

Figure 6.16: Distances between `NNPDFpol1.0` and `NNPDFpol1.1` parton determinations at $Q^2 = 10$ GeV^2.

the `DSSV08` fit including also preliminar STAR 1j-09, called `DSSV++` [188], pointed to a positive Δg consistent with the result of our analysis.

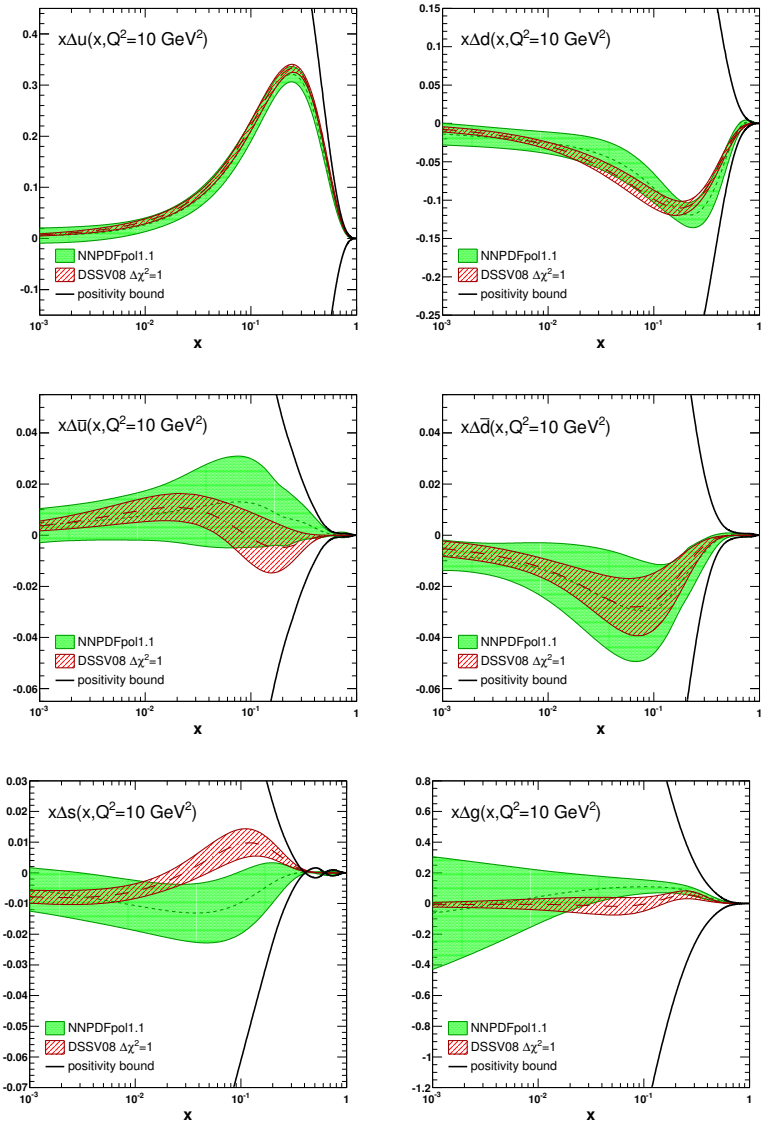

Figure 6.17: The `NNPDFpol1.1` parton set compared to `DSSV08` [75] at $Q^2 = 10\,\mathrm{GeV}^2$.

- Related to the polarized sea quarks, a slight discrepancy is clearly noticeable for the $\Delta\bar{u}$ distribution above $x \sim 3 \cdot 10^{-2}$ between the two parton sets. We recall that $W^{\pm}$ data were not included in the `DSSV08` global fit [75] shown in Fig. 6.17, hence the differences in the $\Delta\bar{u}$ distribution may suggest some tension between $W^{\pm}$ and SIDIS data. This discrepancy may be explained by our poor knowledge of fragmentation functions. A shift away the $\Delta\bar{u}$ central curve towards positive values, as observed in our analysis, was also found in

a preliminary global fit including STAR data in the DSSV framework [290].

- Since W boson production data in the kinematic regime probed by STAR are not sensitive to strangeness, the discrepancy between the NNPDF and DSSV determinations of Δs, already found in `NNPDFpol1.0`, is still present. As discussed in Sec. 4.4, in the NNPDF analysis the polarized strange PDF is obtained from inclusive DIS data through its Q^2 evolution and assumptions about flavor symmetry of the proton sea enforced by experimentally measured baryon octet decay constants, see Sec. 2.3. On the other hand, the `DSSV08` determination of polarized PDFs also includes semi-inclusive data with identified kaons in final states, which are directly sensitive to strangeness, but are likely to introduce an uncertainty difficult to quantify due to poor knowledge of kaon fragmentation function. Finally, we made the choice $\Delta s = \Delta\bar{s}$, like all other polarized analyses [73, 75–78]: actually, this is not justified by any physical reason, but experimental data do not allow for a determination of Δs and $\Delta\bar{s}$ separately.

6.5 Phenomenology of the nucleon spin structure

In this Section, we use the `NNPDFpol1.1` parton set to reevalute the first moments of the polarized PDFs, separately for each quark flavor-antiflavor and for the gluon, in light of the new data sets included in this analysis. Then, we produce some predictions for single-hadron production spin asymmetries at RHIC; we compare them with available measured results in order to qualitatively gauge their potential in pinning down the gluon uncertainty.

6.5.1 The spin content of the proton revisited

The first moments of the polarized PDFs can be directly related to the fraction of the proton spin carried by partons, as explained in Sec. 2.2. This reason has especially motivated our efforts in an accurate and unbiased determination of polarized PDF uncertainties. In Sec. 4.5.1, we presented a detailed analysis of the first moments of the total quark PDF combinations, Δu^+, Δd^+, Δs^+ and of the gluon PDF, Δg, from a fit to polarized inclusive DIS data only. Their potential improvement at a future Electron-Ion Collider was then studied in Sec. 5.5.1. Now, we use the `NNPDFpol1.1` parton set to reassess the determination of the first moments of the polarized parton distributions in order to quantify the impact of new data on the proton's spin content. We recall thet the (truncated) first moments of the polarized PDFs $\Delta f(x,Q^2)$ in the region $[x_{\rm min}, x_{\rm max}]$ are defined as

$$\langle \Delta f(Q^2) \rangle^{[x_{\rm min}, x_{\rm max}]} = \int_{x_{\rm min}}^{x_{\rm max}} dx \Delta f(x,Q^2) \,. \tag{6.21}$$

We consider both *full* moments, *i.e.* $\langle \Delta f(Q^2) \rangle^{[0,1]}$, and *truncated* moments in the x region covered by experimental data, *i.e.* $\langle \Delta f(Q^2) \rangle^{[10^{-3},1]}$.

Let us begin by looking at polarized quarks and antiquarks. We compute Eq. (6.21) for the total quark-antiquark combinations, *i.e.* $\Delta f = \Delta u^+, \Delta d^+$, for sea quarks, *i.e.* $\Delta f = \Delta\bar{u}, \Delta\bar{d}$, for the polarized strangeness, *i.e.* $\Delta f = \Delta s$, and for the singlet PDF combination, *i.e.* $\Delta f = \Delta\Sigma = \sum_{q=u,d,s} \Delta q^+$. The corresponding central values and one-sigma PDF uncertainties obtained from the $N_{\rm rep} = 100$ replicas of the `NNPDFpol1.1` parton set at $Q^2 = 10$ GeV^2 are collected in Tab. 6.8. We compare our results to both `NNPDFpol1.0` and `DSSV08`. In the latter case, we quote the conservative uncertainty estimate obtained in Ref. [75] using the Lagrange

	$\langle\Delta f(Q^2)\rangle^{[0,1]}$		$\langle\Delta f(Q^2)\rangle^{[10^{-3},1]}$		
Δf	NNDPFpol1.0	NNPDFpol1.1	NNDPFpol1.0	NNPDFpol1.1	DSSV08 [75]
$\Delta u+\Delta\bar{u}$	$+0.77\pm0.10$	$+0.79\pm0.06$	$+0.76\pm0.06$	$+0.76\pm0.03$	$+0.793^{+0.028}_{-0.034}(\pm0.020)$
$\Delta d+\Delta\bar{d}$	-0.46 ± 0.10	-0.47 ± 0.06	-0.41 ± 0.06	-0.41 ± 0.04	$-0.416^{+0.035}_{-0.025}(\pm0.042)$
$\Delta\bar{u}$	—	$+0.06\pm0.05$	—	$+0.05\pm0.05$	$+0.028^{+0.059}_{-0.059}(\pm0.008)$
$\Delta\bar{d}$	—	-0.12 ± 0.07	—	-0.10 ± 0.05	$-0.089^{+0.090}_{-0.080}(\pm0.026)$
$\Delta\bar{s}$	-0.07 ± 0.06	-0.06 ± 0.05	-0.06 ± 0.04	-0.05 ± 0.04	$-0.006^{+0.028}_{-0.031}(\pm0.051)$
$\Delta\Sigma$	$+0.16\pm0.30$	$+0.20\pm0.18$	$+0.23\pm0.15$	$+0.25\pm0.10$	$+0.366^{+0.042}_{-0.062}(\pm0.124)$

Table 6.8: Full and truncated first moments of the polarized quark distributions at $Q^2 = 10$ GeV^2 for the NNPDFpol1.1 set compared to NNPDFpol1.0 and DSSV08 [75]. The uncertainty quoted in parenthesis for DSSV08 is due to the extrapolation in the unintegrated region as discussed in the text.

multiplier method with $\Delta\chi^2/\chi^2 = 2\%$. In parenthesis, we also show the uncertainty due to the extrapolation outside the region covered by experimnetal data, estimated as the difference between the full first moment and its truncated counterpart in the region $[10^{-3},1]$ quoted in Ref. [75].

Results from Tab. 6.8 clearly show that first moments obtained with NNPDFpol1.1 and NNPDFpol1.0 are perfectly consistent with each other, as we already knew from the corresponding agreement at the level of polarized PDF, see Fig. 6.15. Besides, the sensitivity to quark-antiquark separation introduced by W data allows for a reduction of the uncertainty up to 50% in the NNPDFpol1.1 determination with respect to NNPDFpol1.0. The comparison between NNPDF full and truncated moments shows that the relative contribution to the total PDF uncertainty from the small-x extrapolation region is roughly the same in both NNPDFpol1.0 and NNPDFpol1.1, and it is about two times larger than the uncertainty in the measured x region. Therefore, we can conclude that the contribution to the total uncertainty from the extrapolation region has reduced by almost a half in NNPDFpol1.1 with respect to NNPDFpol1.0. This is likely because the new data, supplemented by the smoothness provided by the neural network parametrization, decrease the number of acceptable small-x behaviors of the polarized quark PDFs. Nevertheless, the uncertainty from the small-x extrapolation region is still dominant and it could be finally pinned down only by accurate measurements in this region. These may be performed at a future Electron-Ion Collider and they were demonstrated to largely keep under control the extrapolation uncertainties in Sec. 5.5.1.

Coming now to the comparison between NNPDFpol1.1 and DSSV08 [75], we notice that truncated first moments are in perfect agreement, both central value and uncertainties. Slight differences are found for $\Delta\bar{u}$ and $\Delta\bar{s}$, due to the different shape of the corresponding PDFs (see Fig. 6.17). On the other hand, when considering full first moments the NNPDFpol1.1 uncertainties are somewhat larger than those found in the DSSV08 analysis whenever the extrapolation uncertainty is included. This is the major effect of our more flexible PDF parametrization, as already discussed in Sec. 4.5.1.

Let us now move on to discuss the first moment of the polarized gluon Δg. Results for full and truncated moments at $Q^2 = 10$ GeV^2 are presented in Tab. 6.9. There, we compare the predictions from NNPDFpol1.1 with those from NNPDFpol1.0, and two fits from the DSSV family: we consider both the original DSSV08 parton set [75] and its update, DSSV++ [188], which includes the same jet production data in NNPDFpol1.1. As for quarks, we compute both the full and the truncated moments in the measured region $[10^{-3},1]$. In order to quantify the impact of the RHIC inclusive jet data, we also provide results for the truncated first moment

	$\langle\Delta g(Q^2)\rangle^{[0,1]}$	$\langle\Delta g(Q^2)\rangle^{[10^{-3},1]}$	$\langle\Delta g(Q^2)\rangle^{[0.05,0.2]}$
NNPDFpol1.0	-0.95 ± 3.87	-0.06 ± 1.12	$+0.05 \pm 0.15$
NNPDFpol1.1	-0.13 ± 2.60	$+0.31 \pm 0.77$	$+0.15 \pm 0.06$
DSSV08 [75]	—	$0.013^{+0.702}_{-0.314}(\pm 0.097)$	$0.005^{+0.129}_{-0.164}$
DSSV++ [188]	—	—	$0.10^{+0.06}_{-0.07}$

Table 6.9: Full and truncated first moments of the polarized gluon distributions at $Q^2 = 10$ GeV^2 for the NNPDFpol1.1 set compared to NNPDFpol1.0 and various fits of the DSSV family. The uncertainty quoted in parenthesis for DSSV08 is due to the extrapolation in the unintegrated region as discussed in the text.

restricted to the region $x \in [0.05, 0.2]$, which corresponds to the range covered by these data, see Fig. 6.2.

The results quoted in Tab. 6.9 show the substantial improvement in the PDF uncertainties of the gluon first moment in NNPDFpol1.1 as compared to NNPDFpol1.0, due to the constraints on Δg provided by RHIC jet data, see Fig. 6.15. This is further illustrated by the truncated first moment in the region covered by these data, $\langle\Delta g(Q^2)\rangle^{[0.05,0.2]}$, where the PDF uncertainty is reduced by a factor close to three, and where its central value is clearly positive, almost three sigma away from zero. It is clear that the RHIC jet data strongly suggest a positive polarized gluon first moment in the region $x \in [0.05, 0.2]$, unfortunately the absence of other direct constraints outside this region still lead to a quite large value of the gluon first moment when integrating outside this region. In addition, our results for $\langle\Delta g(Q^2)\rangle^{[0.05,0.2]}$, in terms of both central value and uncertainty, turn out to be very close to those obtained in the DSSV++ analysis, which is based on the same set of inclusive jet data [188].

As in the previous NNPDFpol1.0 analysis, the uncertainty due to the extrapolation outside the region covered by experimental data is substantial and dominates the total uncertainty of the full first moment $\langle\Delta g(Q^2)\rangle^{[0,1]}$. The only way to further reduce this uncertainty is to provide measurements which probe Δg at smaller values of x than those that are available now. In this respect, additional jet data from RHIC taken at higher center-of-mass energy, up to $\sqrt{s} = 500$ GeV^2, may be helpful. However, to really pin down the small-x behavior of the polarized PDFs and thus be able to finally reach an accurate determination of $\langle\Delta g(Q^2)\rangle^{[0,1]}$, one will have to resort to the Electron-Ion Collider, as quantified in detail in Chap. 5.

Our main conclusion on the partons' contribution to the proton's spin is then twofold. On the one hand, we have found that their first moments are rather well determined in the kinematic region covered by experimental data and in good agreement with the values obtained in the most updated analysis including collider data, DSSV++ [188]. In particular, the singlet full first moment is less than a half of the proton spin within its uncertainty (see tab. 6.8); the gluon first moment is definitely positive, though rather small, in the region constrained by recent STAR jet data, roughly $0.05 \lesssim x \lesssim 0.2$. On the other hand, we emphasize that the uncertainty on both the singlet and the gluon full first moments coming from the extrapolation to the unmeasured, small-x, region dominates their total uncertainty. For this reason, large values of the gluon first moment are not completely ruled out: within our accurate determination of uncertainties, the almost vanishing value for the singlet axial charge observed in the experiment may still be explained as a cancellation between a rather large quark and the anomalous gluon contributions, as discussed in Sec. 2.2.2. Of course, more experimental data, such as those available at a future Electron-ion Collider, are needed to discriminate which is the behavior of the polarized

gluon in the unmeasured small-x region, in particular whether it is actually small as available experimental data seem to point to and is commonly believed.

6.5.2 Predictions for single-hadron production asymmetries at RHIC

The parton set presented in Sec. 6.4.3 does not include information from semi-inclusive hadron production spin asymmetries at colliders. As discussed in Sec. 6.2.2, the analysis of these data requires the usage of fragmentation functions, whose poor knowledge entails an additional source of theoretical uncertainty, which is difficult to quantify, on the extracted PDFs (see Sec. 6.2.1). Nevertheless, in view of the experimental program which is ongoing at RHIC, it is interesting to compare some predictions for these spin asymmetries with the experimental data, which are now available with significant statistics. The size of uncertainties of theoretical predictions will then fix, at least qualitatively, the experimental precision required for data to further pin down the uncertainties on PDFs, once included in a global fit.

In Fig. 6.18, we show the double-spin asymmetry for single-hadron production in polarized proton-proton collisions, Eq. (6.5), compared to experimental measurements from PHENIX. In particular, we provide predictions for neutral-pion production at center-of-mass energy $\sqrt{s} = 200$ GeV [217] and $\sqrt{s} = 62.4$ GeV [214], and mid-rapidity ($|\eta| < 0.35$) charged hadron production at $\sqrt{s} = 62.4$ GeV [291]. Earlier measurments with neutral-pions [292–294], with significantly larger uncertainties, are not considered here. In Fig. 6.19, we compare the double-spin asymmetry for neutral-pion production at forward rapidity ($0.8 < \eta < 2.0$) with recent STAR data at $\sqrt{s} = 200$ GeV [216]. In Fig. 6.19, we also show the predictions for neutral- and charged-pion production spin asymmetries for which data from the PHENIX experiment will be soon available.

The asymmetries are computed as illustrated in Sec. 6.4.2: the (polarized) numerator in Eq. (6.5) is computed for each replica in the `NNPDFpol1.1` parton set ($N_{\rm rep} = 100$), while the (unpolarized) denominator is computed only once taking the average PDFs from the `NNPDF2.3` [110] parton set at NLO. In both the numerator and the denominator, we use the best fit fragmentation functions from the `DSS07` set [185]. The central value and the uncertainty of the prediction are then obtained as the mean and the standard deviations computed from the $N_{\rm rep} = 100$ results for each replica in the polarized PDF set. Hence, the estimated uncertainty on our prediction only takes into account the uncertainty of the polarized PDFs. While we expect the uncertainty on unpolarized PDFs has a negligible impact, conversely the additional theoretical uncertainty due to the choice of a set of fragmentation functions may have a rather large impact. We finally note that our predictions are made using the code presented in Ref. [269], which we have modified to handle NNPDF parton sets.

Results in Figs. 6.18-6.19 show that the asymmetry remains very small in the measured p_T range. Experimental data are in good agreement with predictions and seems to reinforce the conclusion that the polarized gluon is small in the measured kinematic range. However, we notice that, except for PHENIX neutral-pion production at small p_T, experimental uncertainties are rather large in comparison to those of the corresponding theoretical predictions: this is particularly evident for STAR data in Fig. 6.19, which, however, cover a rapidity range larger than that measured by PHENIX. The mutual size of experimental and theoretical uncertainties is similar to that observed talking of COMPASS open-charm data in Sec. 6.4.2. However in the present the data show a more definite trend towards a growing asymmetry as p_T increases, followed by a consistent behavior of the corresponding theoretical prediction. For this reason, we expect that data on semi-inclusive particle production presented here will have a moderate

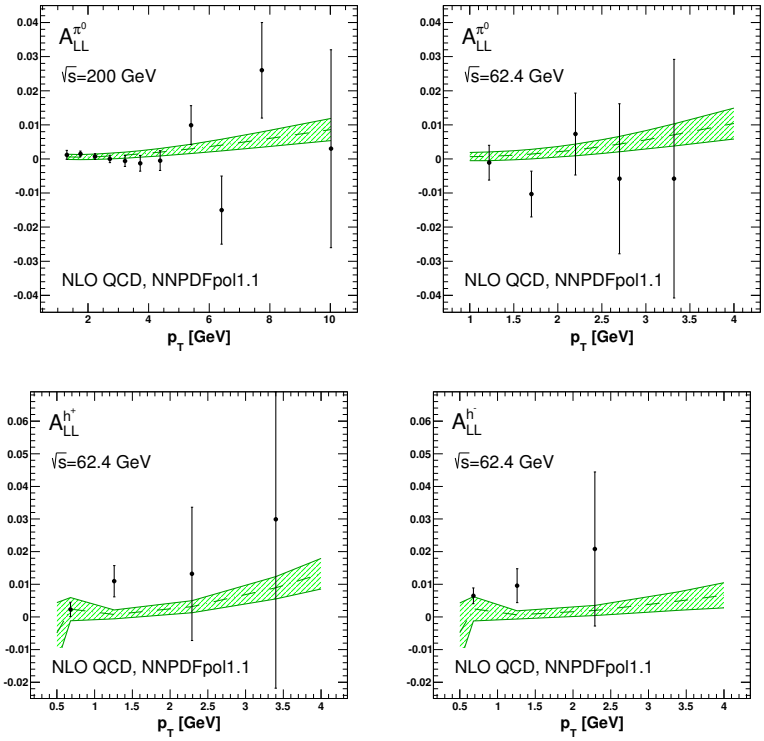

Figure 6.18: Predictions for the neutral-pion (upper plots) and charged hadron (lower plots) spin asymmetries computed at NLO accuracy with the `NNPDFpol1.1` and `NNPDF2.3` PDFs, compared to measured data from PHENIX [214, 217, 291].

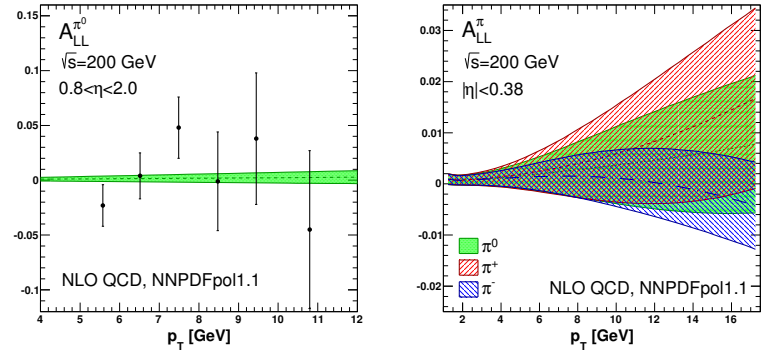

Figure 6.19: (Left panel) Predictions for the neutral-pion spin asymmetry compared to data measured by STAR [216]. (Right panel) Prediction for the neutral- and charged-pion spin asymmetries in the kinematic range accessed by upcoming PHENIX measurements. All theoretical predictions are obtained from the `NNPDFpol1.1` and `NNPDF2.3` PDFs at NLO accuracy.

impact on pinning down the size of the polarized gluon uncertainty, once included in a global PDF determination.

7 Conclusions and outlook

In this Thesis, we presented the first unbiased determination of spin-dependent, or polarized, Parton Distribution Functions (PDFs) of the proton. These distributions are defined as the momentum densities of partons polarized along or opposite the direction of the parent nucleon and are usually denoted as $\Delta f(x, Q^2)$, where f refers to a quark or antiquark flavor or to the gluon. Parton distributions depend on both the Bjorken scaling variable x, the fraction of the proton momentum carried by the parton, and on the energy scale Q^2 with which the proton is probed. While the first is a non-perturbative dependence to be determined from experimental data, the latter is fully predictable in perturbative Quantum Chromodynamics (QCD), the theory of strong interaction.

In the framework of perturbative QCD, polarized parton distributions are essential ingredients for any phenomenological study of hard scattering processes involving polarized hadrons in initial states. The description of these processes in terms of expressions in which the perturbative and the non-perturbative parts are factorized, is a powerful success of QCD itself. In such a picture, parton distributions are the fundamental objects encoding the information on the inner structure of the nucleon; in particular, polarized parton distributions are related to its spin structure, since their integrals over the Bjorken scaling variable are interpreted as fractions of the proton spin.

The interest in the determination of polarized PDFs of the nucleon is largely related to the experimental discovery, in the late 80s, that the singlet axial charge of the proton is anomalously small [49, 50], soon followed by the theoretical realization [59, 60] that the perturbative behavior of polarized PDFs deviates from parton model expectations, according to which gluons decouple in the asymptotic limit. An accurate determination of polarized PDFs is then needed to precisely assess which fraction of the nucleon spin is carried by quark and gluon spins. The residual part of the nucleon spin which possibly would not be accounted for by quarks and gluons may be explained by resorting to their intrinsic Fermi motion and orbital angular momenta [61–63] (for a recent discussion on the spin decomposition see also Ref. [64]). In addition to the investigation of the nucleon's spin structure, polarized PDFs have been recently shown to play a role in the probe of different beyond-standard-model scenarios [112] and in the determination of the Higgs boson spin in the diphoton decay channel, by means of the linear polarization of gluons in an unpolarized proton [113].

Polarized parton distributions are presently known with much less accuracy than their unpolarized counterparts. As pointed out several times in this Thesis, this is mostly due to the experimental data they rely on, which are both less abundant and less accurate than those available in the unpolarized case. Several polarized PDF sets have been determined in the last few years [73–80, 116], but they are all based on the *standard* Hessian methodology for PDF fitting and uncertainty estimation. This approach is known [81] to potentially lead to an underestimation of PDF uncertainties, due to the limitations in the linear propagation of errors and to PDF parametrization in terms of fixed functional forms, both assumed in the *standard* methodology. These issues are especially delicate when the experimental information is scarce, like in the case of polarized data.

In light of these considerations, an unbiased determination of polarized PDFs is crucial in order to provide an adequate estimate of the uncertainty with which quark and gluon spins can be determined from available experimental data. In particular, such a determination allows for scrutinizing the common belief that the anomalous gluon contribution is too small to compensate a reasonably large quark and antiquark spin contribution into the almost vanishing axial charge observed in experiments. Providing the first unbiased determination of polarized parton distributions has precisely been the goal of the present Thesis.

7.1 Summary of the main results

In this Thesis, the determination of polarized parton sets has been carried out within the NNPDF methodology. This uses a robust set of statistical tools, devised for a statistically sound determination of PDFs and their uncertainties, which include Monte Carlo methods for error propagation, neural networks for PDF parametrization and genetic algorithms for their minimization. This methodology has already been successfully applied in the unpolarized case [99–111], where the NNPDF sets are routinely used by the LHC collaborations in their data analysis and data-theory comparison. It has been extended here to polarized PDFs for the first time. In more detail, the main achievements presented in this Thesis are summarized below.

- Based on world-available data from polarized inclusive Deep-Inelastic Scattering, we determined a first polarized parton set at next-to-leading order accuracy, `NNPDFpol1.0`. We reviewed in detail the theory and phenomenology of polarized DIS, in particular focusing on the features of the data included in our analysis. We discussed how the NNPDF methodology has been adapted to the polarized case and which strategies have been devised to face some issues, like the proper implementation of target mass corrections and positivity constraints in the fitting algorithm.

  Our analysis showed that some PDF uncertainties are likely to be underestimated in other existing determinations, based on the standard methodology, due to their less flexible parametrization. This is particularly the case of the gluon, which is left almost unconstrained by inclusive DIS data: hence, its contribution to the nucleon spin is still largely uncertain, unless one makes strong assumptions on the PDF functional form in the small-x ($x \lesssim 10^{-3}$) extrapolation region, where experimental data are presently lacking. For the same reason, we also showed that a determination of the strong coupling α_s from the Bjorken sum rule is not competitive, again because the nonsinglet structure function in the unmeasured small-x region is largely uncertain.

These conclusions were supported by a careful analysis of the stability of our results upon the variation of a number of theoretical and methodological assumptions, in particular related to the effects of target mass corrections, sum rules, and positivity constraints. First, we found that inclusive DIS data, with our kinematic cuts, do not show sensitivity to finite nucleon mass effects, neither in terms of fit quality, nor in terms of the effect on PDFs. Second, we concluded that our fit results are quite stable upon variations of the treatment of sum rules dictated by hyperon decays. Finally, we emphasized that positivity significantly affects PDFs in the region where no data are available, in particular their large-x behavior.

- The proposed Electron-Ion Collider (EIC) [128, 209, 219] is expected to enlarge the kinematic coverage of data, which is presently rather limited, by at least two orders of magnitude in both x and Q^2. This will reduce the uncertainty due to PDF extrapolation to small-x values and will allow for a better determination of the polarized gluon PDF through scaling violations, thanks to a larger Q^2 lever arm. Using simulated pseudodata for two realistic scenarios at an EIC, with increasing energy of both the lepton and hadron beams, we have studied its potential impact on the determination of polarized PDFs. We found that inclusive DIS data at an EIC would entail a considerable reduction in the gluon PDF uncertainty and also provide evidence of a possible large gluon contribution to the nucleon spin, though the latter goal would still be reached with a sizable residual uncertainty.

  The measurement of the charm contribution to the proton structure function $g_1^{p,c}$, which is directly sensitive to the gluon, might provide more information on the corresponding distribution. We showed that $g_1^{p,c}$, though being small, could be as much larger as $10-20\%$ of the total structure function g_1^p in the kinematic region probed by an EIC: hence, in order to further pin down the gluon uncertainty from intrinsic charm effects, one should be able to measure the corresponding contribution to the g_1 structure function within this accuracy.

- The Relativistic Heavy Ion Collider (RHIC) is the first facility in the world to collide polarized proton beams. Measurements on inclusive jet and W boson production asymmetries have been recently presented: we studied their potential in constraining the polarized gluon and in separating light quark and antiquark PDFs, respectively. This new piece of experimental information was included in our polarized parton set by means of Bayesian reweighting of suitable Monte Carlo PDF ensembles [105, 109]. This method, which consist of updating the underlying PDF probability distribution of a prior ensemble according to the conditional probability for the old PDFs with respect to new data, allows for the inclusion of new data in a PDF set without the need of a global refitting; hence, it could be used to quickly update the current PDF set with any new piece of experimental information. This way, we were able to provide the first global polarized PDF set obtained within the NNPDF framework, `NNPDFpol1.1`. In comparison to `NNPDFpol1.0`, the new polarized parton set provides a meaningful determination of sea flavor PDFs $\Delta\bar{u}$ and $\Delta\bar{d}$, based on W boson production data (otherwise not determined by inclusive DIS data or determined in SIDIS, but with the bias introduced by poorly known fragmentation functions), and a determination of the gluon PDF Δg which is improved by open-charm and, particularly, jet data.

  We should also notice that, from a conceptual point of view, the methodology we followed to determine this parton set is in itself particularly valuable. Indeed, we have

explicitely shown how a PDF set can be succesfully obtained by including all data through reweighting of a first unbiased guess, as originally proposed in Refs. [172, 173].

The main conclusion on the partons' contributions to the nucleon's spin based on the NNPDFpol1.1 parton set is twofold. On the one hand, we have found that their first moments are rather well determined in the kinematic region covered by experimental data and are in good agreement with the values obtained in the only available analysis including the same collider data [188]. In particular, in the region constrained by data, the singlet full first moment is less than a half of the proton spin within its uncertainty, while the gluon first moment is definitely positive, though rather small. The determination of the gluon is more accurate in NNPDFpol1.1 than in NNPDFpol1.0, mostly thanks to jet data, located in the region $0.05 \lesssim x \lesssim 0.2$. On the other hand, we emphasize that the uncertainty on both the singlet and the gluon full first moments coming from the extrapolation to the unmeasured, small-x, region dominates their total uncertainty. For this reason, large values of the gluon first moment are not completely ruled out: within our accurate determination of uncertainties, the almost vanishing value for the singlet axial charge observed in the experiment may still be completely explained as a cancellation between a rather large quark and the anomalous gluon contributions.

- We developed a *Mathematica* package which allows for fast and interactive usage of any available NNPDF parton set, both unpolarized and polarized, see Appendix B. This interface includes all the features already available through LHAPDF [122, 123], but they can be profitably combined together with those provided by *Mathematica*. The software we developed was tailored to the users who are not familiar with Fortran or C++ languages used by the LHAPDF interface and who can benefit from the more direct usage of PDFs within a *Mathematica* notebook.

The NNPDFpol1.0 and NNPDFpol1.1 polarized PDF sets, with $N_{\rm rep} = 100$ replicas, are publicly available from the NNPDF website

http://nnpdf.hepforge.org/ .

The *Mathematica* interface as well as FORTRAN and C++ stand-alone codes for handling these parton distributions are also available from the same source.

7.2 Future directions

The NNPDFpol1.1 parton set is based on all the relevant and up-to-date experimental information from deep-inelastic scattering and proton-proton collision observables which do not depend on the fragmentation of the struck quark into final observed hadrons. Further data are expected from PHENIX and STAR in the upcoming years, which will further improve the accuracy of polarized PDF determinations. As further refinements of polarized PDFs will be achieved, they will become more and more appealing for the experimental collaborations to be used in their analysis and for data-theory comparison. In this sense, efforts will be devoted to make the NNPDF polarized parton sets the *gold standard*, as their unpolarized counterparts are quickly becoming.

In order to obtain additional information on the spin structure of the proton, it will be certainly beneficial to include a wide range of semi-inclusive measurements, namely semi-inclusive DIS in fixed-target experiments [95–98, 230], and semi-inclusive particle production

in polarized collisions at RHIC [214, 216, 217, 291]. However, a consistent inclusion of these data in a *global* fit requires first of all the corresponding determination of fragmentation functions using the NNPDF methodology. Indeed, available fragmentation function sets [184–186, 239–243] suffer from several limitations due to their too rigid parametrization. It was recently shown that none of these sets can describe the most updated inclusive charged-particle spectra data at the LHC satisfactorily [162]. Therefore, a determination of fragmentation functions using the NNPDF methodology is highly desirable by itself, and may be important in various areas of phenomenology [238]; in particular, it will pave the way to use a large data set of semi-inclusive polarized data in future NNPDF analyses.

Finally, we notice that the methods illustrated here apply to the determination of any non-perturbative object from experimental data. Hence, even though a phenomenological study of either TMDs or GPDs was beyond the scope of this Thesis, the NNPDF methodology may be used as well to provide the determination of such distributions in the future, when relevant experimental data will reach more and more abundance and accuracy.

In summary, in this Thesis not only we have extended the NNPDF framework to the determination of spin-dependent parton sets, but we have also reached the state-of-the-art in our unbiased understanding of the proton's spin content, as allowed by available experimental data. Further constraints will be provided by a variety of semi-inclusive measurements, which in turn will require the development of a set of parton fragmentation functions using the NNPDF methodology. In the long term, the final word on the spin content of the proton will require brand new facilities such as an Electron-Ion Collider, as we have also extensively discussed in this Thesis. Indeed, it could finally bring polarized PDF determinations to a similar level of accuracy as the one reached for their unpolarized counterparts. We hope that the NNPDF collaboration will play a leading role in this exciting game.

A Statistical estimators

In this Appendix, we collect the definitions of the statistical estimators used in the NNPDF analyses presented in Chaps. 4-5-6. Despite they were already described in Refs. [99, 104, 120], we find useful to give them for completeness and ease of reference here. In the following, we denote with $\mathscr{O}$ a generic quantity depending on replicas in a Monte Carlo ensemble of PDFs; it may be a PDF, a linear combination of PDFs, or a physical observable. We also denote as $\langle\mathscr{O}\rangle_{\rm rep}$ the mean computed over the $N_{\rm rep}$ replicas in the ensemble and as $\langle\mathscr{O}\rangle_{\rm dat}$ the mean computed over the $N_{\rm dat}$ experimental data for a fixed replica in the ensemble.

- Central value
$$\langle\mathscr{O}\rangle_{\rm rep} = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}\mathscr{O}^{(k)}\,. \tag{A.1}$$

- Variance
$$\sigma = \sqrt{\langle\mathscr{O}^2\rangle_{\rm rep} - \langle\mathscr{O}\rangle_{\rm rep}^2}\,. \tag{A.2}$$

- Elements of the correlaton matrix
$$\rho_{ij} = \frac{\left\langle\mathscr{O}_i\mathscr{O}_j\right\rangle_{\rm rep} - \langle\mathscr{O}_i\rangle_{\rm rep}\left\langle\mathscr{O}_j\right\rangle_{\rm rep}}{\sigma_i\sigma_j}\,. \tag{A.3}$$

- Elements of the covariance matrix
$$\mathrm{cov}_{ij} = \rho_{ij}\sigma_i\sigma_j\,. \tag{A.4}$$

- Percentage error over the $N_{\rm dat}$ data points
$$\left\langle \mathrm{PE}\left[\langle\mathscr{O}\rangle_{\rm rep}\right]\right\rangle = \frac{1}{N_{\rm dat}}\sum_{i=1}^{N_{\rm dat}}\left[\frac{\langle\mathscr{O}_i\rangle_{\rm rep}-\mathscr{O}_i}{\mathscr{O}_i}\right]\,. \tag{A.5}$$

- Scatter correlation between two quantities
$$r(\mathscr{O}_1,\mathscr{O}_2) = \frac{\langle\mathscr{O}_1\mathscr{O}_2\rangle_{\rm dat} - \langle\mathscr{O}_1\rangle_{\rm dat}\langle\mathscr{O}_2\rangle_{\rm dat}}{\sigma_1\sigma_2}\,, \tag{A.6}$$
where $\mathscr{O}_{1,2}$ may be obtained as averages over Monte Carlo replicas.

- Square distance between central value estimates from two PDF ensembles

$$d^2\left(\langle\mathcal{O}^{(1)}\rangle,\langle\mathcal{O}^{(2)}\rangle\right)=\frac{[\langle\mathcal{O}^{(1)}\rangle-\langle\mathcal{O}^{(2)}\rangle]^2}{\sigma^2\left[\langle\mathcal{O}^{(1)}\rangle\right]+\sigma^2\left[\langle\mathcal{O}^{(2)}\rangle\right]}, \tag{A.7}$$

  where the variance of the mean is given by

$$\sigma^2[\langle\mathcal{O}^{(i)}\rangle]=\frac{1}{N_{\rm rep}^{(i)}}\sigma^2[\mathcal{O}^{(i)}] \tag{A.8}$$

  in terms of the variance $\sigma[\mathcal{O}^{(i)}]$ of the variables $\mathcal{O}^{(i)}$, estimated as the variance of the replica sample, Eq. (A.2). In our notation $i=1,2$.

- Square distance between square uncertainty estimates from two PDF ensembles

$$d^2\left(\sigma^2[\mathcal{O}^{(1)}],\sigma^2[\mathcal{O}^{(2)}]\right)=\frac{\left(\bar{\sigma}^2_{(1)}-\bar{\sigma}^2_{(2)}\right)^2}{\sigma^2[\bar{\sigma}^2_{(1)}]+\sigma^2[\bar{\sigma}^2_{(2)}]}, \tag{A.9}$$

  where we have defined $\bar{\sigma}^2_{(i)}\equiv\sigma^2[\mathcal{O}^{(i)}]$, $i=1,2$. In practice, for small-size replica samples the distances defined in Eqs. (A.9)-(A.8) display sizable statistical fluctuations. In order to stabilize the result, all distances computed in this thesis are determined as follows: we randomly pick $N_{\rm rep}^{(i)}/2$ out of the $N_{\rm rep}^{(i)}$ replicas for each of the two subsets. The computation of the square distance Eq. (A.9) or Eq. (A.8) is then repeated for $N_{\rm part}=100$ (randomly generated) choices of $N_{\rm rep}^{(i)}/2$ replicas, and the result is averaged: this is sufficient to bring the statistical fluctuations of the distance at the level of a few percent.

B

A *Mathematica* interface to NNPDF parton sets

In this Appendix, we present a package for handling both unpolarized and polarized NNPDF parton sets within `.nb` *Mathematica* notebook files [121]. This allows for performig PDF manipulations easily and quickly, thanks to the powerful features of the *Mathematica* software. The package was tailored to the users who are not familiar with `FORTRAN` or `C++` programming codes, on which the standard available PDF interface, LHAPDF [122, 123], is based. However, since our *Mathematica* package includes all the features available in the LHAPDF interface, any user can benefit from the interactive usage of PDFs within *Mathematica*.

The NNPDF *Mathematica* package can be downloaded from the NNPDF web page

`http://nnpdf.hepforge.org/`

together with sample notebooks containing a step by step explanation of the NNPDF usage within *Mathematica*, as well as a variety of examples. The procedure to download and run our *Mathematica* package is rather simple:

1. **Download**
   `wget http://nnpdf.hepforge.org/math_package.tgz`

2. **Unpack**
   `tar -xvzf math_package.tgz`

3. **Run the tutorial**
   `mathematica Demo-unpol.nb`
   `mathematica Demo-pol.nb`

The input for our *Mathematica* package is any `.LHgrid` file delivered by the NNPDF Collaboration as the final result of a fit. These files are publicly available from the NNPDF hepforge website or from the LHAPDF library and should be downloaded separately from the *Mathematica* package.

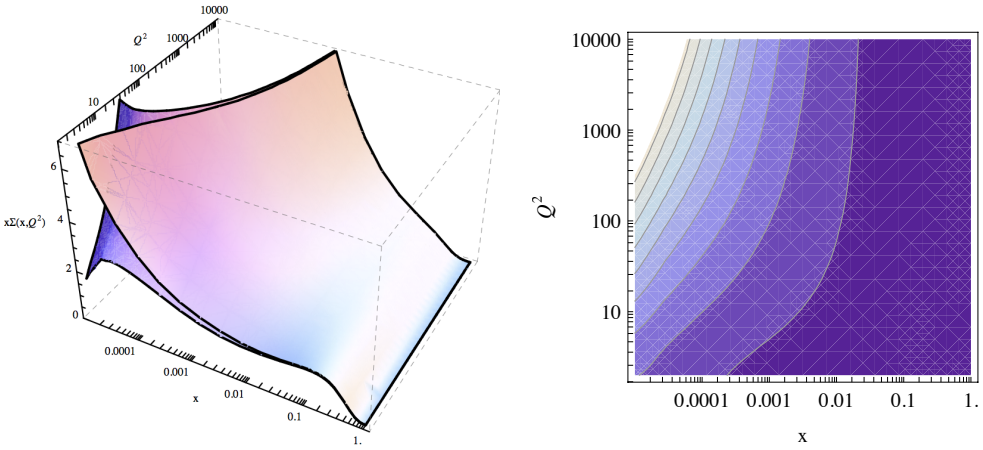

Figure B.1: (Left) Simultaneous (x, Q^2) dependence of the singlet PDF and its 1σ error band from the NLO `NNPDF2.3` parton set. (Right) Contour plot of the of the singlet PDF from the `NNPDF2.3` parton set at NLO in the (x, Q^2) plane.

The functions implemented in the package are summarized in Tab. B.1. In the following, we briefly demonstrate the NNDPF *Mathematica* package by examining the `NNPDF2.3` parton determination at NLO [110].

- **Compute PDF central value and variance** We have defined proper functions to keep the computation of PDF central value and variance very easy. These built-in functions only need x, Q^2 and PDF flavour as input. The user can also specify the confidence level to which central value and variance should be computed.

- **Make PDF plots** *Mathematica* enables a wide range of plotting options. As a few examples, we show the 3D plot and the contour plot of the singlet PDF combination from the `NNPDF2.3` parton set at NLO (see Fig. B.1).

- **Perform computations involving PDFs** PDF manipulation can be carried out straighforwardly since we have defined functions which handle either single replicas or the whole Monte Carlo ensemble. The user can then easily perform any computation which involves PDFs. For example, we show in Fig. B.2 a snapshot of a typical *Mathematica* notebook in which we use our interface to check the momentum and valence sum rules from the `NNPDF2.3` parton set at NLO.

Function	Description
`InitializePDFGrid[path, namegrid]`	It reads the `.LHgrid` file into memory specified by `namegrid` (both unpolarized or polarized) from the location specified by `path`. It also performs the PDF interpolation in the (x-Q^2) space by means of built-in *Mathematica* interpolation algorithm.
`xPDFcv[x,Q2,f]`	It returns `x` times the central value of the PDF with flavor `f` at a given momentum fraction `x` and scale `Q2` (in GeV^2). Note that `f` must be an integer, and `x` and `Q2` must be numeric quantities. For the unpolarized case, and polarized `NNPDFpol1.1` onwards, the LHAPDF convention is used for the flavor `f`, that is, `f`=-6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6 corresponds to $\bar{t}$, $\bar{b}$, $\bar{c}$, $\bar{s}$, $\bar{u}$, $\bar{d}$, g, d, u, s, c, b, t. For the polarized case, `NNPDFpol1.0`, the following convention is used for the flavor `f`: `f`=0, 1, 2, 3, 4 corresponds to Δg, $\Delta u+\Delta\bar{u}$, $\Delta d+\Delta\bar{d}$, $\Delta s+\Delta\bar{s}$.
`xPDFEnsemble[x,Q2,f]`	It returns `x` times the vector of PDF replicas of flavor `f` at a given momentum fraction `x` and scale `Q2` (in GeV^2).
`xPDFRep[x,Q2,f,irep]`	It returns `x` times the `irep` PDF replica of flavor `f` at a given momentum fraction `x` and scale `Q2` (in GeV^2). For `irep`=0, the mean value over the PDF ensemble is returned.
`xPDF[x,Q2,f]`	It returns `x` times the value of the PDF of flavor `f` at a given momentum fraction `x` and scale `Q2` (in GeV^2) with its standard deviation.
`xPDFCL[ensemble,x,Q2,f,CL]`	It returns `x` times the value of the PDF of flavor `f` at a given momentum fraction `x` and scale `Q2` (in GeV^2) with its standard deviation. These values are computed over the PDF ensemble specified by `ensemble` at a confidence level specified by `CL`. The variable `ensemble` must be a function of `x`, `Q2` and `f`: it can be the function `xPDFEnsemble` or any PDF ensemble defined by the user. The variable `CL` must be a real number between 0 and 1.
`NumberPDF[]`	It returns the number of PDF members in the set.
`Infoalphas[]`	It returns information on α_s evolution used in the QCD analysis.
`alphas[Q2,ipo,imodev]`	It returns the QCD strong coupling constant α_s. The inputs are: • `Q2`: the energy scale Q^2, in GeV^2, at which α_s is computed; • `ipo`: the perturbative order at which α_s is computed; 0: LO; 1: NLO; 2: NNLO; • `imodev`: the evolution mode with which α_s is computed; 0: α_s is computed as a function of α_s at the Z mass as given in the `.LHgrid` file; 1: exact solution of the QCD β-function equation using Runge-Kutta algorithm.
`alphasMZ[]`	It returns the list of α_s values at the Z mass used in the QCD analysis for each replica.
`mCharm[], mBottom[], mTop[], mZ[]`	They return the charm, bottom, top quark mass (in GeV) and the Z boson mass (in GeV) used in the QCD analysis
`Lam4[], Lam5[]`	They return the $\Lambda_{\rm QCD,4}$, $\Lambda_{\rm QCD,5}$ used in the QCD analysis
`xMin[], xMax[], Q2Min[], Q2Max[]`	They return the minimum or maximum value for x or Q^2 (in GeV^2) used in the input `.LHgrid` grid.

Table B.1: Description of the functions available in the *Mathematica* interface to NNPDF partons sets.

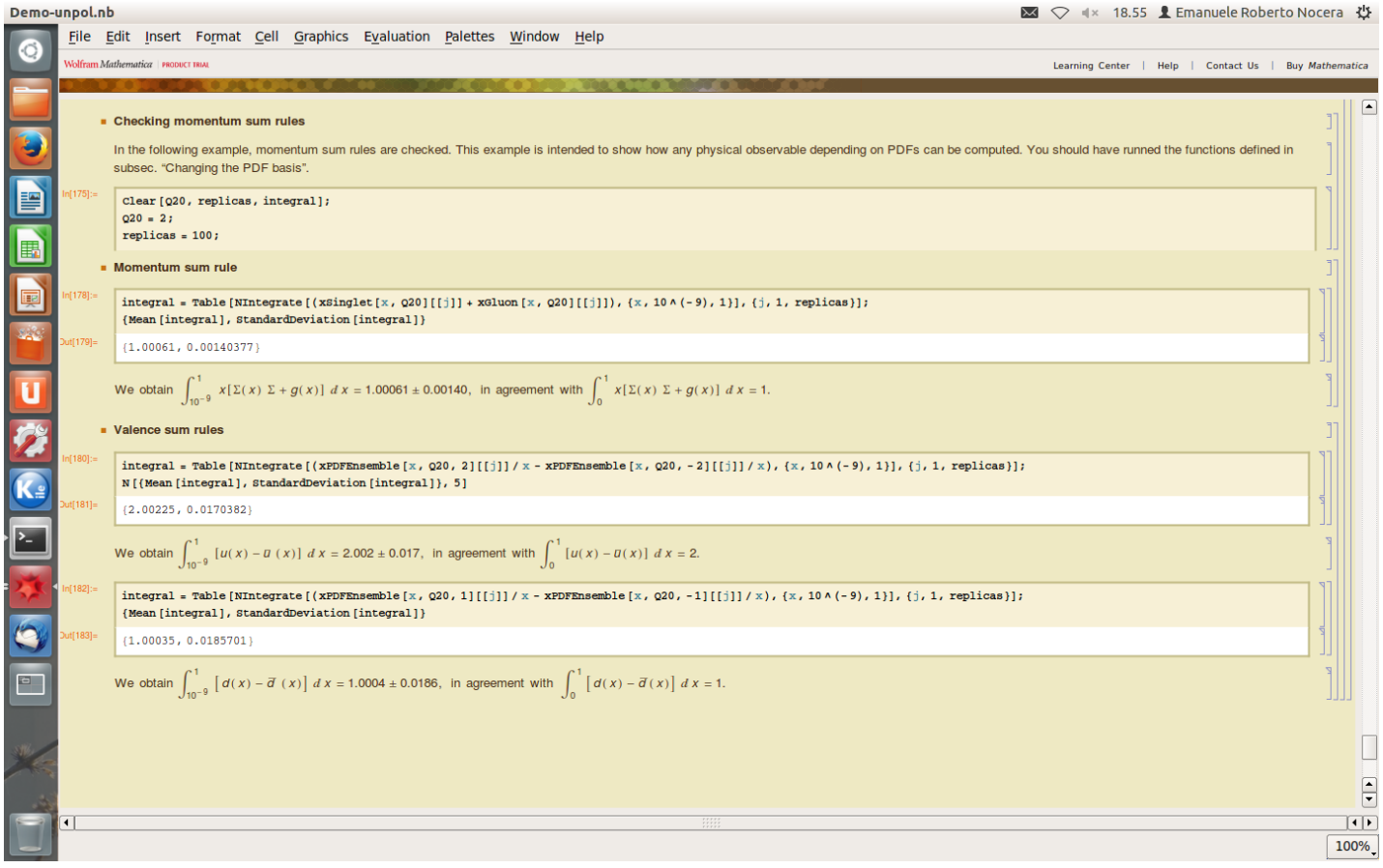

Figure B.2: A snapshot from a *Mathematica* notebook in which the NNPDF interface is used: here momentum and valence sum rules are checked from the `NNPDF2.3` parton set at NLO.

C The FONLL scheme for $g_1(x,Q^2)$ up to $\mathscr{O}(\alpha_s)$

In this Appendix, we collect the relevant explicit formulae for the practical computation of the polarized DIS structure function $g_1(x,Q^2)$ within the FONLL approach [124] up to $\mathscr{O}(\alpha_s)$. In particular, we will restrict to the heavy charm quark contribution g_{1_c} to the polarized proton structure function g_1, which will be of interest for studies at an Electron-Ion Collider, see Sec. 5.5.2. The formulae below extend to the polarized case those collected in Appendix A of Ref. [124].

For g_{1_c}, up to $\mathscr{O}(\alpha_s)$, the relevant equations, to be compared to Eqs. (88-92) in Ref. [124], are expressed in terms of $n_l = 3$ light flavours:

$$g_{1_c}^{\rm FONLL}(x,Q^2) = g_{1_c}^{(d)}(x,Q^2) + g_{1_c}^{n_l}(x,Q^2), \tag{C.1}$$

$$g_{1_c}^{(d)}(x,Q^2) = g_{1_c}^{(n_l+1)}(x,Q^2) - g_{1_c}^{n_l,0}(x,Q^2), \tag{C.2}$$

$$g_{1_c}^{(n_l+1)}(x,Q^2) = e_c^2 \int_x^1 \frac{dz}{z} \left\{ C_q(z,\alpha_s(Q^2)) \Delta c^+ \left(\frac{x}{z},Q^2\right) \right. \tag{C.3}$$

$$\left. + \frac{\alpha_s(Q^2)}{4\pi} C_g^{(1)}(z) \Delta g \left(\frac{x}{z},Q^2\right) \right\}, \tag{C.4}$$

$$g_{1_c}^{(n_l)}(x,Q^2) = e_c^2 \frac{\alpha_s(Q^2)}{4\pi} \int_{ax}^1 \frac{dz}{z} H_g^{(1)}(Q^2,m_c^2,z) \Delta g \left(\frac{x}{z},Q^2\right), \tag{C.5}$$

$$g_{1_c}^{(n_l,0)}(x,Q^2) = e_c^2 \frac{\alpha_s(Q^2)}{4\pi} \int_x^1 \frac{dz}{z} H_g^{(1),0} \left(\ln\left(\frac{Q^2}{m_c^2}\right),z\right) \Delta g \left(\frac{x}{z},Q^2\right), \tag{C.6}$$

where the strong coupling $\alpha_s(Q^2)$, the polarized gluon distribution $\Delta g(x,Q^2)$ and the charm combination $\Delta c^+(x,Q^2) = \Delta c(x,Q^2) + \Delta \bar{c}(x,Q^2)$ are expressed in the same decoupling $n_f = n_l = 3$ scheme. The charm fractional charge squared is $e_c^2 = 4/9$ and $a = 1 + 4m_c^2/Q^2$. The massive coefficient functions and their massless limits (labelled with the superscript 0) in Eqs. (C.4-C.6) are taken from Ref. [138, 295, 296]* and read:

*We notice two misprintings in Ref. [295]: Eq. (8) should read $sq = \sqrt{1 - 4\frac{z}{1-z}\frac{m_c^2}{Q^2}}$ and the first term in the second

- for the gluon

$$C_g^{(1)}(z) = T_f\left[4(2z-1)\ln\left(\frac{1-z}{z}\right)+4(3-4z)\right] \quad \text{(C.7)}$$

$$H_g^{(1)}(Q^2,m_c^2,z) = T_f\left[4(2z-1)\ln\left(\frac{1+\beta}{1-\beta}\right)+4(3-4z)\beta\right] \quad \text{(C.8)}$$

$$H_g^{(1),0}\left(\ln\left(\frac{Q^2}{m_c^2}\right),z\right) = T_f\left[4(2z-1)\ln\left(\frac{1-z}{z}\frac{Q^2}{m_c^2}\right)+4(3-4z)\right] \quad \text{(C.9)}$$

with $T_f = 1/2$ and

$$\beta = \sqrt{1-4\frac{z}{1-z}\frac{m_c^2}{Q^2}} \quad \text{(C.10)}$$

- for the quark

$$C_q(z,\alpha_s(Q^2)) = \delta(1-z)+\frac{\alpha_s(Q^2)}{4\pi}C_q^{(1)}(z) \quad \text{(C.11)}$$

with

$$\begin{aligned} C_q^{(1)}(z) = C_F\Bigg\{&4\left[\frac{\ln(1-z)}{1-z}\right]_+ - 3\left[\frac{1}{1-z}\right]_+ - 2(1+z)\ln(1-z) \\ &-2\frac{1+z^2}{1-z}\ln z + 4 + 2z + \delta(1-z)[-4\zeta(2)-9]\Bigg\} \end{aligned} \quad \text{(C.12)}$$

and $C_F = 4/3$.

The formulae listed in this Appendix, should be implemented in the `FastKernel` framework [104], as succesfully performed in the unpolarized case [106], in order to gauge the impact of heavy charm flavour on polarized PDFs. As discussed in Sec. 5.5.2 this would be of particular interest in future studies at an Electron-Ion Collider.

line of Eq. (12) should read $-2\frac{1+z^2}{1-z}\ln z$.

Bibliography

[1] J.M. Campbell, J. Huston and W. Stirling, Rept.Prog.Phys. 70 (2007) 89, hep-ph/0611148.

[2] I. Estermann, R. Frisch and S. O., Nature 132 (1933) 169.

[3] M. Gell-Mann, Phys.Rev. 125 (1962) 1067.

[4] G. Zweig, CERN-TH-412 .

[5] M. Gell-Mann, Phys.Lett. 8 (1964) 214.

[6] Y. Dothan, M. Gell-Mann and Y. Ne'eman, Phys.Lett. 17 (1965) 148.

[7] O. Greenberg, Phys.Rev.Lett. 13 (1964) 598.

[8] E.D. Bloom et al., Phys.Rev.Lett. 23 (1969) 930.

[9] J.I. Friedman and H.W. Kendall, Ann.Rev.Nucl.Part.Sci. 22 (1972) 203.

[10] J. Bjorken, Phys.Rev. 179 (1969) 1547.

[11] R.P. Feynman, Phys.Rev.Lett. 23 (1969) 1415.

[12] J. Bjorken and E.A. Paschos, Phys.Rev. 185 (1969) 1975.

[13] J. Callan, Curtis G. and D.J. Gross, Phys.Rev.Lett. 22 (1969) 156.

[14] J. de Groot et al., Phys.Lett. B82 (1979) 292.

[15] J. de Groot et al., Z.Phys. C1 (1979) 143.

[16] TASSO Collaboration, R. Brandelik et al., Phys.Lett. B86 (1979) 243.

[17] PLUTO Collaboration, C. Berger et al., Phys.Lett. B86 (1979) 418.

[18] JADE Collaboration, W. Bartel et al., Phys.Lett. B91 (1980) 142.

[19] TASSO Collaboration, R. Brandelik et al., Phys.Lett. B97 (1980) 453.

[20] J.R. Ellis and I. Karliner, Nucl.Phys. B148 (1979) 141.

[21] J. Callan, Curtis G. and D.J. Gross, Phys.Rev. D8 (1973) 4383.

[22] H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys.Lett. B47 (1973) 365.

[23] D. Gross and F. Wilczek, Phys.Rev. D8 (1973) 3633.

[24] S. Weinberg, Phys.Rev.Lett. 31 (1973) 494.

[25] D.J. Gross and F. Wilczek, Phys.Rev.Lett. 30 (1973) 1343.

[26] H.D. Politzer, Phys.Rev.Lett. 30 (1973) 1346.

[27] H. Georgi and H.D. Politzer, Phys.Rev. D9 (1974) 416.

[28] H. Georgi and H.D. Politzer, Phys.Rev. D14 (1976) 1829.

[29] preprint DESY-HERA-81-10 (1981).

[30] S. Drell and T.M. Yan, Phys.Rev.Lett. 25 (1970) 316.

[31] S. Berman and M. Jacob, Phys.Rev.Lett. 25 (1970) 1683.

[32] S. Berman, J. Bjorken and J.B. Kogut, Phys.Rev. D4 (1971) 3388.

[33] R.L. Jaffe, Phys.Rev. D5 (1972) 2622.

[34] R. Feynman, R. Field and G. Fox, Phys.Rev. D18 (1978) 3320.

[35] R.K. Ellis et al., Phys.Lett. B78 (1978) 281.

[36] R.K. Ellis et al., Nucl.Phys. B152 (1979) 285.

[37] D. Amati, R. Petronzio and G. Veneziano, Nucl.Phys. B140 (1978) 54.

[38] D. Amati, R. Petronzio and G. Veneziano, Nucl.Phys. B146 (1978) 29.

[39] J.C. Collins, D.E. Soper and G.F. Sterman, Phys.Lett. B134 (1984) 263.

[40] J.C. Collins, D.E. Soper and G.F. Sterman, Nucl.Phys. B261 (1985) 104.

[41] J.C. Collins, Nucl.Phys. B394 (1993) 169, hep-ph/9207265.

[42] UA1 Collaboration, G. Arnison et al., Phys.Lett. B122 (1983) 103.

[43] UA1 Collaboration, G. Arnison et al., Phys.Lett. B126 (1983) 398.

[44] UA2 Collaboration, M. Banner et al., Phys.Lett. B122 (1983) 476.

[45] ATLAS Collaboration, G. Aad et al., Phys.Lett. B716 (2012) 1, 1207.7214.

[46] CMS Collaboration, S. Chatrchyan et al., Phys.Lett. B716 (2012) 30, 1207.7235.

[47] C.Y. Prescott et al., SLAC-PROPOSAL-E-095 .

[48] M. Alguard et al., Phys.Rev.Lett. 37 (1976) 1261.

[49] European Muon Collaboration, J. Ashman et al., Phys.Lett. B206 (1988) 364.

[50] European Muon Collaboration, J. Ashman et al., Nucl.Phys. B328 (1989) 1.

[51] E. Leader and M. Anselmino, Z.Phys. C41 (1988) 239.

[52] E581 Collaboration, E704 Collaboration, D. Adams et al., Phys.Lett. B261 (1991) 201.

[53] K. Krueger et al., Phys.Lett. B459 (1999) 412.

[54] G. Bunce et al., Phys.Rev.Lett. 36 (1976) 1113.

[55] K.J. Heller et al., Phys.Lett. B68 (1977) 480.

[56] S. Gourlay et al., Phys.Rev.Lett. 56 (1986) 2244.

[57] G. Bunce et al., Ann.Rev.Nucl.Part.Sci. 50 (2000) 525, hep-ph/0007218.

[58] G. Altarelli and G. Parisi, Nucl.Phys. B126 (1977) 298.

[59] G. Altarelli and G.G. Ross, Phys.Lett. B212 (1988) 391.

[60] G. Altarelli and B. Lampe, Z.Phys. C47 (1990) 315.

[61] R. Jaffe and A. Manohar, Nucl.Phys. B337 (1990) 509.

[62] X.D. Ji, Phys.Rev.Lett. 78 (1997) 610, hep-ph/9603249.

[63] B. Bakker, E. Leader and T. Trueman, Phys.Rev. D70 (2004) 114001, hep-ph/0406139.

[64] E. Leader and C. Lorce, Phys.Rept. (2013), 1309.4235.

[65] T. Gehrmann and W.J. Stirling, Phys.Rev. D53 (1996) 6100, hep-ph/9512406.

[66] G. Altarelli et al., Nucl.Phys. B496 (1997) 337, hep-ph/9701289.

[67] G. Altarelli et al., Acta Phys.Polon. B29 (1998) 1145, hep-ph/9803237.

[68] M. Gluck et al., Phys.Rev. D63 (2001) 094005, hep-ph/0011215.

[69] Asymmetry Analysis Collaboration, M. Hirai, S. Kumano and N. Saito, Phys.Rev. D69 (2004) 054021, hep-ph/0312112.

[70] D. de Florian and R. Sassot, Phys.Rev. D62 (2000) 094025, hep-ph/0007068.

[71] E. Leader, A.V. Sidorov and D.B. Stamenov, Eur.Phys.J. C23 (2002) 479, hep-ph/0111267.

[72] J. Blumlein and H. Bottcher, Nucl.Phys. B636 (2002) 225, hep-ph/0203155.

[73] Asymmetry Analysis Collaboration, M. Hirai and S. Kumano, Nucl.Phys. B813 (2009) 106, 0808.0413.

[74] D. de Florian et al., Phys.Rev.Lett. 101 (2008) 072001, 0804.0422.

[75] D. de Florian et al., Phys.Rev. D80 (2009) 034030, 0904.3821.

[76] J. Blumlein and H. Bottcher, Nucl.Phys. B841 (2010) 205, 1005.3113.

[77] E. Leader, A.V. Sidorov and D.B. Stamenov, Phys.Rev. D82 (2010) 114018, 1010.0574.

[78] P. Jimenez-Delgado, A. Accardi and W. Melnitchouk, (2013), 1310.3734.

[79] A.N. Khorramian et al., Phys.Rev. D83 (2011) 054017, 1011.4873.

[80] F. Arbabifar, A.N. Khorramian and M. Soleymaninia, (2013), 1311.1830.

[81] S. Forte, Acta Phys.Polon. B41 (2010) 2859, 1011.5247.

[82] R. Jaffe, Phys.Lett. B365 (1996) 359, hep-ph/9509279.

[83] L. Mankiewicz, G. Piller and A. Saalfeld, Phys.Lett. B395 (1997) 318, hep-ph/9611326.

[84] V. Barone, T. Calarco and A. Drago, Phys.Lett. B431 (1998) 405, hep-ph/9801281.

[85] H.J. Lee et al., Phys.Lett. B491 (2000) 257, hep-ph/0006004.

[86] B. Dressler et al., Eur.Phys.J. C14 (2000) 147, hep-ph/9909541.

[87] F.G. Cao and A. Signal, Phys.Rev. D68 (2003) 074002, hep-ph/0306033.

[88] New Muon Collaboration, P. Amaudruz et al., Phys.Rev.Lett. 66 (1991) 2712.

[89] NA51 Collaboration, A. Baldit et al., Phys.Lett. B332 (1994) 244.

[90] NuSea Collaboration, E. Hawker et al., Phys.Rev.Lett. 80 (1998) 3715, hep-ex/9803011.

[91] J.P. Ralston and D.E. Soper, Nucl.Phys. B152 (1979) 109.

[92] R. Jaffe and X.D. Ji, Phys.Rev.Lett. 67 (1991) 552.

[93] J.C. Collins and D.E. Soper, Nucl.Phys. B193 (1981) 381.

[94] Proceedings of the International School of Physics "Enrico Fermi", Course CLXXX - Three-dimensional Partonic Structure of the Nucleon, 2012.

[95] HERMES Collaboration, A. Airapetian et al., Phys.Rev. D71 (2005) 012003, hep-ex/0407032.

[96] COMPASS Collaboration, M. Alekseev et al., Phys.Lett. B660 (2008) 458, 0707.4077.

[97] COMPASS Collaboration, M. Alekseev et al., Phys.Lett. B680 (2009) 217, 0905.2828.

[98] COMPASS Collaboration, M. Alekseev et al., Phys.Lett. B693 (2010) 227, 1007.4061.

[99] NNPDF Collaboration, L. Del Debbio et al., JHEP 0503 (2005) 080, hep-ph/0501067.

[100] NNPDF Collaboration, L. Del Debbio et al., JHEP 0703 (2007) 039, hep-ph/0701127.

[101] NNPDF Collaboration, R.D. Ball et al., Nucl.Phys. B809 (2009) 1, 0808.1231.

[102] NNPDF Collaboration, R.D. Ball et al., JHEP 1005 (2010) 075, 0912.2276.

[103] NNPDF Collaboration, R.D. Ball et al., Nucl.Phys. B823 (2009) 195, 0906.1958.

[104] R.D. Ball et al., Nucl.Phys. B838 (2010) 136, 1002.4407.

[105] NNPDF Collaboration, R.D. Ball et al., Nucl.Phys. B849 (2011) 112, 1012.0836.

[106] R.D. Ball et al., Nucl.Phys. B849 (2011) 296, 1101.1300.

[107] NNPDF Collaboration, R.D. Ball et al., Nucl.Phys. B855 (2012) 153, 1107.2652.

[108] NNPDF Collaboration, R.D. Ball et al., (2011), 1110.1863.

[109] R.D. Ball et al., Nucl.Phys. B855 (2012) 608, 1108.1758.

[110] R.D. Ball et al., Nucl.Phys. B867 (2013) 244, 1207.1303.

[111] NNPDF, R.D. Ball et al., Nucl.Phys. B877 (2013) 290, 1308.0598.

[112] G. Bozzi, B. Fuks and M. Klasen, Phys.Lett. B609 (2005) 339, hep-ph/0411318.

[113] D. Boer et al., Phys.Rev.Lett. 111 (2013) 032002, 1304.2654.

[114] E. Leader, Spin in Particle Physics (Cambridge University Press, 2001).

[115] C.A. Aidala et al., Rev.Mod.Phys. 85 (2013) 655, 1209.2803.

[116] The NNPDF Collaboration, R.D. Ball et al., Nucl.Phys. B874 (2013) 36, 1303.7236.

[117] The NNPDF Collaboration, R.D. Ball et al., Phys.Lett. B728 (2013) 524, 1310.0461.

[118] E.R. Nocera, Nuovo Cim. C36 (2013) 143, 1302.6409.

[119] E.R. Nocera, PoS DIS2013 (2013) 211, 1307.0146.

[120] S. Forte et al., JHEP 0205 (2002) 062, hep-ph/0204232.

[121] N.P. Hartland and E.R. Nocera, Nucl.Phys.Proc.Suppl. 234 (2013) 54, 1209.2585.

[122] M. Whalley, D. Bourilkov and R. Group, (2005), hep-ph/0508110.

[123] `https://lhapdf.hepforge.org/`.

[124] S. Forte et al., Nucl.Phys. B834 (2010) 116, 1001.2312.

[125] E. Leader and E. Predazzi, An Introduction to Gauge Theories and the New Physics (Cambridge University Press, 1985).

[126] M. Anselmino, A. Efremov and E. Leader, Phys.Rept. 261 (1995) 1, hep-ph/9501369.

[127] M. Anselmino, P. Gambino and J. Kalinowski, Z.Phys. C64 (1994) 267, hep-ph/9401264.

[128] D. Boer et al., (2011), 1108.1713.

[129] A. Accardi et al., (2012), 1212.1701.

[130] M.L. Mangano et al., (2001), hep-ph/0105155.

[131] J.C. Collins, D.E. Soper and G.F. Sterman, Adv.Ser.Direct.High Energy Phys. 5 (1988) 1, hep-ph/0409313.

[132] R. Mertig and W. van Neerven, Z.Phys. C70 (1996) 637, hep-ph/9506451.

[133] W. Vogelsang, Phys.Rev. D54 (1996) 2023, hep-ph/9512218.

[134] A. Vogt et al., Nucl.Phys.Proc.Suppl. 183 (2008) 155, 0807.1238.

[135] M. Diemoz et al., Z.Phys. C39 (1988) 21.

[136] A. Vogt, Comput.Phys.Commun. 170 (2005) 65, hep-ph/0408244.

[137] J. Blumlein, H. Bottcher and A. Guffanti, Nucl.Phys. B774 (2007) 182, hep-ph/0607200.

[138] E. Zijlstra and W. van Neerven, Nucl.Phys. B417 (1994) 61.

[139] G. Altarelli and W.J. Stirling, Part.World 1 (1989) 40.

[140] R.D. Carlitz, J.C. Collins and A.H. Mueller, Phys.Lett. B214 (1988) 229.

[141] S.L. Adler, Phys.Rev. 177 (1969) 2426.

[142] J. Bell and R. Jackiw, Nuovo Cim. A60 (1969) 47.

[143] R.D. Ball, S. Forte and G. Ridolfi, Phys.Lett. B378 (1996) 255, hep-ph/9510449.

[144] Particle Data Group, K. Nakamura et al., J.Phys. G37 (2010) 075021.

[145] S. Forte, (1994), hep-ph/9409416.

[146] J. Bjorken, Phys.Rev. 148 (1966) 1467.

[147] J. Bjorken, Phys.Rev. D1 (1970) 1376.

[148] S. Larin and J. Vermaseren, Phys.Lett. B259 (1991) 345.

[149] H. Burkhardt and W. Cottingham, Annals Phys. 56 (1970) 453.

[150] A. Efremov, O. Teryaev and E. Leader, Phys.Rev. D55 (1997) 4307, hep-ph/9607217.

[151] S. Kuhn, J.P. Chen and E. Leader, Prog.Part.Nucl.Phys. 63 (2009) 1, 0812.3535.

[152] E143 collaboration, K. Abe et al., Phys.Rev. D58 (1998) 112003, hep-ph/9802357.

[153] E155 Collaboration, P. Anthony et al., Phys.Lett. B553 (2003) 18, hep-ex/0204028.

[154] The HERMES Collaboration, A. Airapetian et al., Eur.Phys.J. C72 (2012) 1921, 1112.5584.

[155] S. Wandzura and F. Wilczek, Phys.Lett. B72 (1977) 195.

[156] A. Accardi et al., JHEP 0911 (2009) 093, 0907.2942.

[157] J. Blumlein and H. Bottcher, (2012) 237, 1207.3170.

[158] V. Braun et al., Phys.Rev. D83 (2011) 094023, 1103.1269.

[159] A. Piccione and G. Ridolfi, Nucl.Phys. B513 (1998) 301, hep-ph/9707478.

[160] R. Flores-Mendieta, E.E. Jenkins and A.V. Manohar, Phys.Rev. D58 (1998) 094028, hep-ph/9805416.

[161] N. Cabibbo, E.C. Swallow and R. Winston, Ann.Rev.Nucl.Part.Sci. 53 (2003) 39, hep-ph/0307298.

[162] D. d'Enterria et al., (2013), 1311.1415.

[163] A. Martin et al., Eur.Phys.J. C28 (2003) 455, hep-ph/0211080.

[164] J. Pumplin et al., Phys.Rev. D65 (2001) 014013, hep-ph/0101032.

[165] J. Pumplin et al., JHEP 0207 (2002) 012, hep-ph/0201195.

[166] J.C. Collins and J. Pumplin, (2001), hep-ph/0105207.

[167] A. Martin et al., Eur.Phys.J. C63 (2009) 189, 0901.0002.

[168] J. Pumplin, D. Stump and W. Tung, Phys.Rev. D65 (2001) 014011, hep-ph/0008191.

[169] D. Stump et al., Phys.Rev. D65 (2001) 014012, hep-ph/0101051.

[170] A. Signori et al., JHEP 1311 (2013) 194, 1309.3507.

[171] E. Askanazi, K. Holcomb and S. Liuti, (2013), 1309.7085.

[172] W.T. Giele and S. Keller, Phys.Rev. D58 (1998) 094023, hep-ph/9803393.

[173] W.T. Giele, S.A. Keller and D.A. Kosower, (2001), hep-ph/0104052.

[174] D. d'Enterria and J. Rojo, Nucl.Phys. B860 (2012) 311, 1202.1762.

[175] M. Czakon et al., JHEP 1307 (2013) 167, 1303.7215.

[176] N. Armesto et al., JHEP 1311 (2013) 015, 1309.5371.

[177] G. Watt and R. Thorne, JHEP 1208 (2012) 052, 1205.4024.

[178] New Muon Collaboration., M. Arneodo et al., Phys.Lett. B364 (1995) 107, hep-ph/9509406.

[179] L. Whitlow et al., Phys.Lett. B250 (1990) 193.

[180] J. Ethier and W. Melnitchouk, Phys.Rev. C88 (2013) 054001, 1308.3723.

[181] S.A. Kulagin and W. Melnitchouk, Phys.Rev. C77 (2008) 015210, 0710.1101.

[182] S.A. Kulagin and W. Melnitchouk, Phys.Rev. C78 (2008) 065203, 0809.3998.

[183] M. Gluck, E. Reya and A. Vogt, Eur.Phys.J. C5 (1998) 461, hep-ph/9806404.

[184] M. Hirai et al., Phys.Rev. D75 (2007) 094009, hep-ph/0702250.

[185] D. de Florian, R. Sassot and M. Stratmann, Phys.Rev. D75 (2007) 114010, hep-ph/0703242.

[186] D. de Florian, R. Sassot and M. Stratmann, Phys.Rev. D76 (2007) 074033, 0707.1506.

[187] D. de Florian et al., Prog.Part.Nucl.Phys. 67 (2012) 251, 1112.0904.

[188] E. Aschenauer et al., (2013), 1304.0079.

[189] Spin Muon Collaboration, B. Adeva et al., Phys.Rev. D58 (1998) 112001.

[190] Spin Muon Collaboration, B. Adeva et al., Phys.Rev. D60 (1999) 072004.

[191] COMPASS Collaboration, V.Y. Alexakhin et al., Phys.Lett. B647 (2007) 8, hep-ex/0609038.

[192] COMPASS Collaboration, M. Alekseev et al., Phys.Lett. B690 (2010) 466, 1001.4654.

[193] E154 Collaboration, K. Abe et al., Phys.Rev.Lett. 79 (1997) 26, hep-ex/9705012.

[194] E155 Collaboration, P. Anthony et al., Phys.Lett. B493 (2000) 19, hep-ph/0007248.

[195] HERMES Collaboration, K. Ackerstaff et al., Phys.Lett. B404 (1997) 383, hep-ex/9703005.

[196] HERMES Collaboration, A. Airapetian et al., Phys.Rev. D75 (2007) 012007, hep-ex/0609039.

[197] C. Simolo, Extraction of characteristic constants in QCD with perturbative and nonperturbative methods, PhD thesis, 2006, 0807.1501.

[198] Jefferson Lab Hall A Collaboration, X. Zheng et al., Phys.Rev. C70 (2004) 065207, nucl-ex/0405006.

[199] CLAS Collaboration, R. Fatemi et al., Phys.Rev.Lett. 91 (2003) 222002, nucl-ex/0306019.

[200] CLAS Collaboration, K. Dharmawardane et al., Phys.Lett. B641 (2006) 11, nucl-ex/0605028.

[201] V. Andrieux, New compass results on the proton spin-dependent structure function g_1^p, Talk presented at DIS2013, Parc Chanot Marseille.

[202] M. Dittmar et al., (2005), hep-ph/0511119.

[203] G.P. Salam and J. Rojo, Comput.Phys.Commun. 180 (2009) 120, 0804.3755.

[204] C.B. Bishop, Neural Networks for Pattern Recognition (Oxford University Press, 1995).

[205] G. Altarelli, S. Forte and G. Ridolfi, Nucl.Phys. B534 (1998) 277, hep-ph/9806345.

[206] R.D. Ball, S. Forte and G. Ridolfi, Nucl.Phys. B444 (1995) 287, hep-ph/9502340.

[207] S. Forte, G. Altarelli and G. Ridolfi, Nucl.Phys.Proc.Suppl. 74 (1999) 138, hep-ph/9808462.

[208] Particle Data Group, J. Beringer et al., Phys.Rev. D86 (2012) 010001.

[209] A. Deshpande et al., Ann.Rev.Nucl.Part.Sci. 55 (2005) 165, hep-ph/0506148.

[210] L. Del Debbio, A. Guffanti and A. Piccione, JHEP 0911 (2009) 060, 0907.2506.

[211] S. Forte and L. Magnea, Phys.Lett. B448 (1999) 295, hep-ph/9812479.

[212] F. Close and R. Roberts, Phys.Lett. B336 (1994) 257, hep-ph/9407204.

[213] COMPASS Collaboration, C. Adolph et al., Phys.Rev. D87 (2013) 052018, 1211.6849.

[214] PHENIX Collaboration, A. Adare et al., Phys.Rev. D79 (2009) 012003, 0810.0701.

[215] STAR Collaboration, L. Adamczyk et al., Phys.Rev. D86 (2012) 032006, 1205.2735.

[216] STAR Collaboration, L. Adamczyk et al., (2013), 1309.1800.

[217] PHENIX Collaboration, A. Adare et al., Phys.Rev.Lett. 103 (2009) 012003, 0810.0694.

[218] PHENIX Collaboration, A. Adare et al., Phys.Rev. D84 (2011) 012006, 1009.4921.

[219] A. Accardi et al., (2012), 1212.1701.

[220] H1 and ZEUS Collaboration, F. Aaron et al., JHEP 1001 (2010) 109, 0911.0884.

[221] LHeC Study Group, J. Abelleira Fernandez et al., J.Phys. G39 (2012) 075001, 1206.2913.

[222] E.C. Aschenauer, R. Sassot and M. Stratmann, Phys.Rev. D86 (2012) 054020, 1206.6014.

[223] E.C. Aschenauer et al., Phys.Rev. D88 (2013) 114025, 1309.5327.

[224] See for example `https://wiki.bnl.gov/eic/`.

[225] See for example http://www.jlab.org/elic/.

[226] L. Mankiewicz, A. Schafer and M. Veltri, Comput.Phys.Commun. 71 (1992) 305.

[227] A. Watson, Z.Phys. C12 (1982) 123.

[228] M. Cacciari, M. Greco and P. Nason, JHEP 9805 (1998) 007, hep-ph/9803400.

[229] HERMES Collaboration, K. Ackerstaff et al., Phys.Lett. B464 (1999) 123, hep-ex/9906035.

[230] Spin Muon Collaboration, B. Adeva et al., Phys.Lett. B420 (1998) 180, hep-ex/9711008.

[231] COMPASS Collaboration, M. Alekseev et al., Eur.Phys.J. C64 (2009) 171, 0907.0388.

[232] HERMES Collaboration, A. Airapetian et al., Phys.Rev.Lett. 84 (2000) 2584, hep-ex/9907020.

[233] HERMES Collaboration, A. Airapetian et al., JHEP 1008 (2010) 130, 1002.3921.

[234] Spin Muon Collaboration (SMC), B. Adeva et al., Phys.Rev. D70 (2004) 012002, hep-ex/0402010.

[235] COMPASS Collaboration, C. Adolph et al., Phys.Lett. B718 (2013) 922, 1202.4064.

[236] STAR Collaboration, M. Aggarwal et al., Phys.Rev.Lett. 106 (2011) 062002, 1009.0326.

[237] PHENIX Collaboration, A. Adare et al., Phys.Rev.Lett. 106 (2011) 062001, 1009.0505.

[238] S. Albino, Rev.Mod.Phys. 82 (2010) 2489, 0810.4255.

[239] S. Kretzer, Phys.Rev. D62 (2000) 054001, hep-ph/0003177.

[240] B.A. Kniehl, G. Kramer and B. Potter, Nucl.Phys. B582 (2000) 514, hep-ph/0010289.

[241] L. Bourhis et al., Eur.Phys.J. C19 (2001) 89, hep-ph/0009101.

[242] S. Albino, B. Kniehl and G. Kramer, Nucl.Phys. B725 (2005) 181, hep-ph/0502188.

[243] S. Albino, B. Kniehl and G. Kramer, Nucl.Phys. B803 (2008) 42, 0803.2768.

[244] S. Frixione and G. Ridolfi, Phys.Lett. B383 (1996) 227, hep-ph/9605209.

[245] M. Stratmann and W. Vogelsang, Z.Phys. C74 (1997) 641, hep-ph/9605330.

[246] I. Bojak and M. Stratmann, Phys.Lett. B433 (1998) 411, hep-ph/9804353.

[247] I. Bojak and M. Stratmann, Nucl.Phys. B540 (1999) 345, hep-ph/9807405.

[248] Z. Merebashvili, A. Contogouris and G. Grispos, Phys.Rev. D62 (2000) 114509, hep-ph/0007050.

[249] A. Contogouris, Z. Merebashvili and G. Grispos, Phys.Lett. B482 (2000) 93, hep-ph/0003204.

[250] J. Riedl, A. Schafer and M. Stratmann, Phys.Rev. D80 (2009) 114020, 0911.2146.

[251] I. Bojak and M. Stratmann, Phys.Rev. D67 (2003) 034010, hep-ph/0112276.

[252] J. Riedl, M. Stratmann and A. Schafer, Eur.Phys.J. C73 (2013) 2360, 1212.1319.

[253] C. Bourrely and J. Soffer, Phys.Lett. B314 (1993) 132.

[254] C. Bourrely, J. Guillet and J. Soffer, Nucl.Phys. B361 (1991) 72.

[255] P. Chiappetta et al., Z.Phys. C59 (1993) 629.

[256] CDF Collaboration, T. Aaltonen et al., Phys.Rev. D78 (2008) 052006, 0807.2204.

[257] D0 Collaboration, V. Abazov et al., Phys.Rev.Lett. 101 (2008) 062001, 0802.2400.

[258] ATLAS Collaboration, G. Aad et al., Phys.Rev. D86 (2012) 014022, 1112.6297.

[259] CMS Collaboration, S. Chatrchyan et al., Phys.Rev. D87 (2013) 112002, 1212.6660.

[260] W. Giele, E.N. Glover and D.A. Kosower, Nucl.Phys. B403 (1993) 633, hep-ph/9302225.

[261] S. Frixione, Z. Kunszt and A. Signer, Nucl.Phys. B467 (1996) 399, hep-ph/9512328.

[262] S. Catani and M. Seymour, Nucl.Phys. B485 (1997) 291, hep-ph/9605323.

[263] S. Frixione, Nucl.Phys. B507 (1997) 295, hep-ph/9706545.

[264] Z. Bern and D.A. Kosower, Nucl.Phys. B379 (1992) 451.

[265] Z. Kunszt, A. Signer and Z. Trocsanyi, Nucl.Phys. B411 (1994) 397, hep-ph/9305239.

[266] D. de Florian et al., Nucl.Phys. B539 (1999) 455, hep-ph/9808262.

[267] B. Jager, M. Stratmann and W. Vogelsang, Phys.Rev. D70 (2004) 034010, hep-ph/0404057.

[268] D. de Florian, Phys.Rev. D67 (2003) 054004, hep-ph/0210442.

[269] B. Jager et al., Phys.Rev. D67 (2003) 054005, hep-ph/0211007.

[270] D. de Florian and W. Vogelsang, Phys.Rev. D81 (2010) 094020, 1003.4533.

[271] COMPASS Collaboration, P. Abbon et al., Nucl.Instrum.Meth. A577 (2007) 455, hep-ex/0703049.

[272] STAR Collaboration, K. Ackermann et al., Nucl.Instrum.Meth. A499 (2003) 624.

[273] P. Djawotho, (2013), 1303.0543.

[274] Collaboration STAR, J.R. Stevens, (2013), 1302.6639.

[275] T. Kluge, K. Rabbertz and M. Wobisch, (2006) 483, hep-ph/0609285.

[276] T. Carli et al., Eur.Phys.J. C66 (2010) 503, 0911.2985.

[277] A. Martin, W. Stirling and R. Thorne, Phys.Lett. B636 (2006) 259, hep-ph/0603143.

[278] C. Peterson et al., Phys.Rev. D27 (1983) 105.

[279] G. Colangelo and P. Nason, Phys.Lett. B285 (1992) 167.

[280] G. Moreno et al., Phys.Rev. D43 (1991) 2815.

[281] NuSea Collaboration, J. Webb et al., (2003), hep-ex/0302019.

[282] J.C. Webb, (2003), hep-ex/0301031.

[283] NuSea Collaboration, R. Towell et al., Phys.Rev. D64 (2001) 052002, hep-ex/0103030.

[284] CDF Collaboration, T. Aaltonen et al., Phys.Rev.Lett. 102 (2009) 181801, 0901.2169.

[285] CDF Collaboration, T.A. Aaltonen et al., Phys.Lett. B692 (2010) 232, 0908.3914.

[286] D0 Collaboration, V. Abazov et al., Phys.Rev. D76 (2007) 012003, hep-ex/0702025.

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

[288] CMS Collaboration, S. Chatrchyan et al., Phys.Rev.Lett. 109 (2012) 111806, 1206.2598.

[289] LHCb Collaboration, R. Aaij et al., JHEP 1206 (2012) 058, 1204.1620.

[290] M. Stratmann, Polarized nucleon structure. status and path forward., Plenary talk at DIS2013, Parc Chanot Marseille.

[291] PHENIX Collaboration, A. Adare et al., Phys.Rev. D86 (2012) 092006, 1202.4020.

[292] PHENIX Collaboration, S. Adler et al., Phys.Rev.Lett. 93 (2004) 202002, hep-ex/0404027.

[293] PHENIX Collaboration, S. Adler et al., Phys.Rev. D73 (2006) 091102, hep-ex/0602004.

[294] PHENIX Collaboration, A. Adare et al., Phys.Rev. D76 (2007) 051106, 0704.3599.

[295] F. Steffens, Nucl.Phys. B523 (1998) 487, hep-ph/9708430.

[296] M. Gluck, E. Reya and W. Vogelsang, Nucl.Phys. B351 (1991) 579.

List of Figures

List of Tables