Unbiased determination of polarized parton distributions and their uncertainties

Unbiased determination of polarized parton distributions and their uncertainties

NNPDF Collaboration

Richard D. Ball [a], Stefano Forte [b,*], Alberto Guffanti [c],
Emanuele R. Nocera [b], Giovanni Ridolfi [d], Juan Rojo [e]

[a] *Tait Institute, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland, UK*
[b] *Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[c] *The Niels Bohr International Academy and Discovery Center, The Niels Bohr Institute, Blegdamsvej 17,
DK-2100 Copenhagen, Denmark*
[d] *Dipartimento di Fisica, Università di Genova and INFN, Sezione di Genova, Genova, Italy*
[e] *PH Department, TH Unit, CERN, CH-1211 Geneva 23, Switzerland*

Abstract

We present a determination of a set of polarized parton distributions (PDFs) of the nucleon, at next-to-leading order, from a global set of longitudinally polarized deep-inelastic scattering data: NNPDFpol1.0. The determination is based on the NNPDF methodology: a Monte Carlo approach, with neural networks used as unbiased interpolants, previously applied to the determination of unpolarized parton distributions, and designed to provide a faithful and statistically sound representation of PDF uncertainties. We present our dataset, its statistical features, and its Monte Carlo representation. We summarize the technique used to solve the polarized evolution equations and its benchmarking, and the method used to compute physical observables. We review the NNPDF methodology for parametrization and fitting of neural networks, the algorithm used to determine the optimal fit, and its adaptation to the polarized case. We finally present our set of polarized parton distributions. We discuss its statistical properties, test for its stability upon various modifications of the fitting procedure, and compare it to other recent polarized parton sets, and in particular obtain predictions for polarized first moments of PDFs based on it. We find that the uncertainties on the gluon, and to a lesser extent the strange PDF, were substantially underestimated in previous determinations.

---

* Corresponding author.
 E-mail address: forte@mi.infn.it (S. Forte).

1. Introduction

The interest in the determination of polarized parton distributions (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 [1,2], soon followed by the theoretical realization [3,4] that the perturbative behavior of polarized PDFs deviates from parton model expectations, according to which gluons decouple in the asymptotic limit. The theoretical interpretation of these results has spawned a huge literature, while at the same time experimental information on polarized PDFs from deep-inelastic scattering but also from a variety of other processes has been accumulating over the years (see e.g. [5] and references therein).

First studies of the polarized structure of the nucleon were aimed at an accurate determination of polarized first moments (including detailed uncertainty estimates) [6–8], but did not attempt a determination of a full PDF set, which was first proposed in Ref. [9], but without uncertainty estimation. More recently, polarized PDF sets with uncertainties have been constructed by at least four groups (BB [10,11], AAC [12], LSS [13,14] and DSSV [15,68]). These PDF sets slightly differ in the choice of datasets, the form of PDF parametrization, and in several details of the QCD analysis (such as the treatment of higher-twist corrections), but they are all based on the standard Hessian methodology for PDF fitting and uncertainty determination, which has been widely used in the unpolarized case (see [16,17] and references therein). This methodology is known [16] to run into difficulties especially when information is scarce, because of the intrinsic bias of the Hessian method based on a 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 [18–29]. The NNPDF technique uses a robust set of statistical tools, which include Monte Carlo methods for error propagation, neural networks for PDF parametrization, and genetic algorithms for their training. The NNPDF sets are now routinely used by the Tevatron and LHC collaborations in their data analysis and for data-theory comparisons. In this work we extend the application of the NNPDF methodology to the determination of polarized parton distributions of the nucleon. As we shall see, some PDF uncertainties will turn out to be underestimated in existing PDF determinations: in particular those of the polarized gluon distribution, but also those of the strange distribution.

The outline of this paper is as follows. In Section 2 we present the dataset used to determine polarized PDFs, and we review the relationship between measured asymmetries and structure functions. In Section 3 we discuss the parametrization of polarized PDFs in terms of neural networks, and the construction of polarized structure functions. Then in Section 4 we discuss the minimization strategy. The results for the `NNPDFpol1.0` polarized partons are presented in Section 5, and in Section 6 we discuss the phenomenological implications for the spin content of the proton and the test of the Bjorken sum rule. Finally in Section 7 we summarize our results and outline future developments. Some details on the benchmarking of polarized PDF evolution are given in Appendix A.

2. Experimental data

The bulk of the experimental information on (longitudinal) polarized proton structure comes from inclusive polarized deep-inelastic scattering with charged lepton beams. Deep-inelastic scattering with longitudinally polarized beams and targets allows a determination of the longitudinal structure function $g_1(x, Q^2)$, which in turn admits a factorized expression in terms of polarized PDFs. Neutral-current deep-inelastic scattering does not allow us to disentangle the contribution of quarks and antiquarks. Using both proton and neutron (deuteron or ^{3}He) targets it is possible to separate the isospin singlet and triplet quark contributions to structure functions, with the gluon determined from scaling violations. A weak control on the separation of the isospin singlet quark contribution into its SU(3) octet and singlet component is possible using baryon decays to fix the respective normalization of these contributions, with in principle their different scale dependence providing some constraint on their shape.

Only charged-current deep-inelastic scattering would allow for full flavor separation [30]: this could be feasible with neutrino beams (such as available at a neutrino factory [31]), or perhaps very high-energy polarized charged lepton beams (such as available at an Electron–Ion Collider [32]). Therefore, current constraints on flavor separation are only provided by semi-inclusive deep-inelastic scattering data or by polarized hadron collider processes, such as polarized Drell–Yan production in fixed target collisions and polarized W production at the Relativistic Heavy Ion Collider (RHIC). Likewise, direct constraints on the medium- and large-x polarized gluon require hadron and jet production either in fixed target experiments or at RHIC, while the small-x gluon can only be probed by going to higher energy, such as at a polarized Electron–Ion Collider.

In this paper we will concentrate on inclusive longitudinally polarized DIS data, and thus we will only determine a subset of PDF combinations. This first polarized PDF set based on NNPDF methodology will then be available for inclusion of other datasets through the reweighting technique of Refs. [24,28].

We will first review the experimental observables which we use for the determination of polarized structure functions, and the information which various experiments provide on them. Then, we will summarize the features of the data we use, and finally the construction and validation of the Monte Carlo pseudo-data sample from the input experimental data.

2.1. Experimental observables and longitudinal polarized structure functions

Standard perturbative factorization provides predictions for the polarized structure functions $g_1(x, Q^2)$. However, experiments measure cross section asymmetries, defined by considering longitudinally polarized leptons scattering off a hadronic target, polarized either longitudinally or transversely with respect to the collision axis, from which the longitudinal ($A_\parallel$) and transverse ($A_\perp$) asymmetries are determined as

$$A_\parallel = \frac{d\sigma^{\rightarrow\Rightarrow} - d\sigma^{\rightarrow\Leftarrow}}{d\sigma^{\rightarrow\Rightarrow} + d\sigma^{\rightarrow\Leftarrow}}, \qquad A_\perp = \frac{d\sigma^{\rightarrow\Uparrow} - d\sigma^{\rightarrow\Downarrow}}{d\sigma^{\rightarrow\Uparrow} + d\sigma^{\rightarrow\Downarrow}}. \tag{1}$$

The hadronic tensor for polarized, parity conserving deep-inelastic scattering can be parametrized in terms of four structure functions: two of them, $F_1(x, Q^2)$ and $F_2(x, Q^2)$, characterize spin-averaged deep-inelastic scattering, while $g_1(x, Q^2)$ and $g_2(x, Q^2)$ appear when both the lepton beam and the nucleon target are in definite polarization states. For the conventional definition of the hadronic tensor in terms of structure functions, see e.g. [33].

The two polarized structure functions are related to the measurable asymmetries (1) by

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

$$g_2(x, Q^2) = \frac{F_1(x, Q^2)}{(1 + \gamma^2)(1 + \eta\zeta)} \left[\left(\frac{\zeta}{\gamma} - 1 \right) \frac{A_\parallel}{D} + \left(\eta + \frac{1}{\gamma} \right) \frac{A_\perp}{d} \right]. \tag{3}$$

In Eqs. (2)–(3) the dependence on the nucleon mass m is taken into account through the factor

$$\gamma^2 \equiv \frac{4m^2 x^2}{Q^2}, \tag{4}$$

which also appears in the definitions of the other kinematic factors in Eqs. (2)–(3):

$$d = \frac{D\sqrt{1 - y - \gamma^2 y^2/4}}{1 - y/2}, \tag{5}$$

$$D = \frac{1 - (1 - y)\epsilon}{1 + \epsilon R(x, Q^2)}, \tag{6}$$

$$\eta = \frac{\epsilon \gamma y}{1 - \epsilon(1 - y)}, \tag{7}$$

$$\zeta = \frac{\gamma(1 - y/2)}{1 + \gamma^2 y/2}, \tag{8}$$

$$\epsilon = \frac{4(1 - y) - \gamma^2 y^2}{2y^2 + 4(1 - y) + \gamma^2 y^2}. \tag{9}$$

Here y is the standard lepton scaling variable, given by

$$y = \frac{p \cdot q}{p \cdot k} = \frac{Q^2}{2xmE} \tag{10}$$

in terms of the nucleon, lepton and virtual photon momenta, p, k and q, or, in the target rest frame, in terms of the energy E of the incoming lepton beam.

The unpolarized structure function F_1 and unpolarized structure function ratio R which enter the definition (2)–(3) of the polarized structure function may be expressed in terms of F_2 and F_L by

$$F_1(x, Q^2) \equiv \frac{F_2(x, Q^2)}{2x[1 + R(x, Q^2)]}(1 + \gamma^2), \tag{11}$$

$$R(x, Q^2) \equiv \frac{F_L(x, Q^2)}{F_2(x, Q^2) - F_L(x, Q^2)}. \tag{12}$$

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

$$A_\parallel = D(A_1 + \eta A_2), \qquad A_\perp = d(A_2 - \zeta A_1), \tag{13}$$

where

$$A_1(x, Q^2) \equiv \frac{\sigma_{1/2}^T - \sigma_{3/2}^T}{\sigma_{1/2}^T + \sigma_{3/2}^T}, \qquad A_2(x, Q^2) \equiv \frac{2\sigma^{TL}}{\sigma_{1/2}^T + \sigma_{3/2}^T}. \tag{14}$$

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. (13) 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. (13) in Eqs. (2)–(3) we may express the structure functions in terms of A_1 and A_2 instead:

$$g_1(x, Q^2) = \frac{F_1(x, Q^2)}{1 + \gamma^2} \big[A_1(x, Q^2) + \gamma A_2(x, Q^2) \big],$$ (15)

$$g_2(x, Q^2) = \frac{F_1(x, Q^2)}{1 + \gamma^2} \left[\frac{A_2}{\gamma} - A_1 \right].$$ (16)

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)–(3) we may obtain an expression of it in terms of the two asymmetries $A_\parallel$, $A_\perp$, or, using Eqs. (15)–(16), in terms of the two asymmetries A_1, A_2. Clearly, up to corrections of $\mathcal{O}(\frac{m}{Q})$, g_1 is fully determined by $A_\parallel$, which coincides with A_1 up to $\mathcal{O}(\frac{m}{Q})$ 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:

$$g_1(x, Q^2) = \frac{F_1(x, Q^2)}{1 + \gamma \eta} \frac{A_\parallel}{D} + \frac{\gamma(\gamma - \eta)}{\gamma \eta + 1} g_2(x, Q^2),$$ (17)

or, equivalently, in terms of A_1 and g_2:

$$g_1(x, Q^2) = A_1(x, Q^2) F_1(x, Q^2) + \gamma^2 g_2(x, Q^2).$$ (18)

It is then possible to use Eq. (17) or Eq. (18) 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 [34–36]. 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 [37]:

$$g_2(x, Q^2) = g_2^{t2}(x, Q^2) + g_2^{t3}(x, Q^2).$$ (19)

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

$$g_2^{t2}(x, Q^2) = -g_1(x, Q^2) + \int_x^1 \frac{dy}{y} g_1(y, Q^2)$$ (20)

which in Mellin space becomes

$$g_2^{t2}(N, Q^2) = -\frac{N-1}{N} g_1(N, Q^2).$$ (21)

It is important to note that g_2^{t3} is not suppressed by a power of $\frac{m}{Q}$ in comparison to g_2^{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^{t3} is compatible with zero at low scale $Q^2 \sim m^2$. Fits to g_2^{t3} [38,39], as well as theoretical estimates of it [38,40] support the conclusion that

$$g_2(x, Q^2) \approx g_2^{t2}(x, Q^2) \equiv g_2^{\mathrm{WW}}(x, Q^2),\qquad(22)$$

which is known as the Wandzura–Wilczek [37] relation.

We will thus determine g_1, using Eq. (17) or Eq. (18), from an experimental determination of the longitudinal asymmetry, and use the approximate Wandzura–Wilczek form (22) for g_2. In order to test the dependence of results on this approximation, we will also consider the opposite assumption that $g_2 = 0$ identically.

2.2. The dataset: Observables, kinematic cuts, uncertainties and correlations

We use deep-inelastic lepton–nucleon scattering (DIS) data coming from all relevant experiments [2,34,35,41–48] performed at CERN, SLAC and DESY. The experiments use different nucleon targets (protons, neutrons or deuterons). The main features of these datasets are summarized in Table 1, where we show, for each experiment, the number of available data points, the kinematic range covered by the experiment, and the quantity which is published and which we use for the extraction of g_1. This quantity is not the same for all experiments: the primary observable can be one of the many asymmetries or structure functions discussed in Section 2.1, as we now summarize (individual experiments are labeled as in Table 1).

- **EMC, SMC, SMClowx, COMPASS, HERMES97**
  All these experiments have performed a measurement of $A_{\parallel}$. They have then determined A_1 from it using Eq. (13), under the assumption $\eta \approx 0$. Therefore, what these experiments actually publish is a measurement of $\frac{A_{\parallel}}{D}$. We determine g_1 from $\frac{A_{\parallel}}{D}$ using Eq. (17). This is possible because D is completely fixed by Eq. (6) in terms of the unpolarized structure function ratio (12). We determine the unpolarized structure function ratio using as primary inputs F_2, for which we use the parametrization of Refs. [18,49], and F_L, which we determine from its expression in terms of parton distributions, using the `NNPDF2.1 NNLO` parton set [26].
- **HERMES**
  This experiment has performed a measurement of $A_{\parallel}$, and it publishes both $A_{\parallel}$ and A_1 (which is determined using Eq. (13) 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. (17).
- **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. (13); for the two lowest energies only $A_{\parallel}$ is measured and A_1 is extracted from it using Eqs. (15)–(16) while assuming the form (22) for g_2. The values of A_1 obtained with the three beam energies are combined into a single determination of A_1; radiative corrections are applied at this combination stage. Because of this, we must use this combined value of A_1, from which we then determine g_1 using Eq. (18). In order to determine y Eq. (10), which depends on the beam energy, we use the mean of the three energies.

Table 1
Experimental datasets included in the present analysis. For each experiment we show the number of points before and after (in parenthesis) applying kinematic cuts, the kinematic range and the measured observable.

Experiment	Set	N_{dat}	x_{min}	x_{max}	Q^2_{min} [GeV^2]	Q^2_{max} [GeV^2]	F	Ref.
EMC								
	EMC-A1P	10	0.0150	0.4660	3.5	29.5	$A^p_{\parallel}/D$	[2]
SMC								
	SMC-A1P	12	0.0050	4.0800	1.3	58.0	$A^p_{\parallel}/D$	[41]
	SMC-A1D	12	0.0050	0.4790	1.3	54.8	$A^d_{\parallel}/D$	[41]
SMClowx								
	SMClx-A1P	15 (8)	0.0001 (0.0043)	0.1210	0.02 (1.09)	23.1	$A^p_{\parallel}/D$	[42]
	SMClx-A1D	15 (8)	0.0001 (0.0043)	0.1210	0.02 (1.09)	22.9	$A^d_{\parallel}/D$	[42]
E143								
	E143-A1P	28 (25)	0.0310	0.5260	1.27	9.52 (7.72)	A^p_1	[34]
	E143-A1D	28 (25)	0.0310	0.5260	1.27	9.52 (7.72)	A^d_1	[34]
E154								
	E154-A1N	11	0.0170	0.5640	1.2	15.0	A^n_1	[43]
E155								
	E155-G1P	22 (20)	0.0150	0.7500 (0.5000)	1.22	34.72 (26.86)	g^p_1/F^p_1	[44]
	E155-G1N	22 (20)	0.0150	0.7500 (0.5000)	1.22	34.72 (26.86)	g^n_1/F^n_1	[44]
COMPASS-D								
	CMP07-A1D	15	0.0046	0.5660	1.10	55.3	$A^d_{\parallel}/D$	[45]
COMPASS-P								
	CMP10-A1P	15	0.0046	0.5680	1.10	62.1	$A^p_{\parallel}/D$	[46]
HERMES97								
	HER97-A1N	9 (8)	0.0330	0.4640 (0.3420)	1.22	5.25 (3.86)	$A^n_{\parallel}/D$	[47]
HERMES								
	HER-A1P	38 (28)	0.0264	0.7311 (0.5823)	1.12	14.29 (11.36)	$A^p_{\parallel}$	[48]
	HER-A1D	38 (28)	0.0264	0.7311 (0.5823)	1.12	14.29 (11.36)	$A^d_{\parallel}$	[48]
Total		290 (245)						

- **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. (18).

- **E155**

  This experiment only measures $A_{\parallel}$, from which $\frac{g_1}{F_1}$ is extracted using Eq. (15) with the Wandzura–Wilczek form of g_2 Eq. (22). In this case, we use these values of $\frac{g_1}{F_1}$, and we extract g_1 using Eq. (11) for F_1, together with the parametrization of Refs. [18,49] for F_2 and the expression in terms of parton distributions and the `NNPDF2.1 NNLO` parton set [26] for F_L, as in the other cases.

We have excluded from our analysis all data points with $Q^2 \leqslant Q^2_{cut} = 1$ GeV^2, since below such energy scale perturbative QCD cannot be considered reliable. A similar choice of cut was
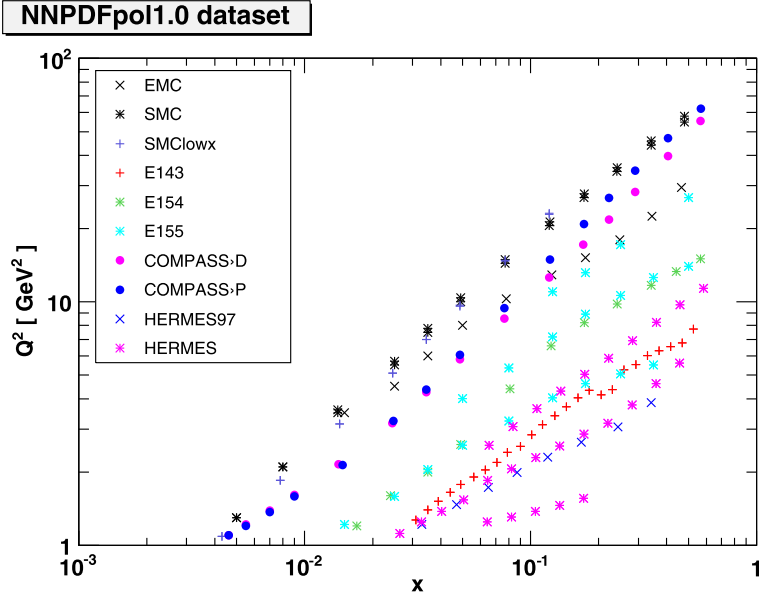
Fig. 1. Experimental data in the (x, Q^2) plane (after kinematic cuts).

made in Refs. [6–8,11,12,15]. 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 higher-twist corrections. The cut is chosen based on a study presented in Ref. [50], where higher-twist terms were added to the observables, with a coefficient fitted to the data, and it was shown that the higher-twist contribution becomes compatible with zero if one imposes the cut $W^2 \geqslant W^2_{\text{cut}} = 6.25 \text{ GeV}^2$. 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. 1: in particular, it excludes all available JLAB data [51–53]. The number of data points surviving the kinematic cuts for each dataset is given in parenthesis in Table 1.

As can be seen from the scatter plot in Fig. 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 $1 \text{ GeV}^2 \leqslant Q^2 \lesssim 60 \text{ GeV}^2$. In recent years, the coverage of the low-x region has been improved by a complementary set of SMC data [42] and by the more recent COMPASS data [45,46]. 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 [48]. In comparison to the dataset used in Refs. [6–8] several new datasets are being used, in particular the SMC [42], HERMES [48] and COMPASS [45,46] data. The dataset used in this paper is the same as that of Ref. [11], and also the same as the DIS data of the fit of Ref. [15], which however has a wider dataset which extends beyond inclusive DIS.

Each experimental collaboration provides uncertainties on the measured quantities listed in the next-to-last column of Table 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 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. We include all available correlated systematics. These are provided by the experimental collaboration 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 Section 2.2 (which, of course, is very close to the value determined by the experimental collaboration).

We then construct a covariance matrix

$$\text{cov}_{pq} = \left(\sum_i \sigma_{i,p}^{(c)} \sigma_{i,q}^{(c)} + \delta_{pq} \sigma_p^{(u)} \sigma_q^{(u)} \right) g_{1,p} g_{1,q}, \tag{23}$$

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 uncertainty, and $\sigma_p^{(u)}$ the uncorrelated uncertainties, which are in turn found as a sum in quadrature of all uncorrelated sources of statistical $\sigma_{i,p}^{(\text{stat})}$ and systematic $\sigma_{i,p}^{(\text{syst})}$ uncertainty on each point:

$$\left(\sigma_p^{(u)} \right)^2 = \sum_i \left(\sigma_{i,p}^{(\text{stat})} \right)^2 + \sum_j \left(\sigma_{j,p}^{(\text{syst})} \right)^2. \tag{24}$$

The correlation matrix is defined as

$$\rho_{pq} = \frac{\text{cov}_{pq}}{\sigma_p^{(\text{tot})} \sigma_q^{(\text{tot})} g_{1,p} g_{1,q}}, \tag{25}$$

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

$$\left(\sigma_p^{(\text{tot})} \right)^2 = \left(\sigma_p^{(u)} \right)^2 + \sum_i \left(\sigma_{i,p}^{(c)} \right)^2. \tag{26}$$

We show in Table 2 the average experimental uncertainties for each dataset, 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 Table 1, we provide the values before and after kinematic cuts (if different).

In Table 1, we distinguish between experiments, defined as groups of data which cannot be correlated to each other, and datasets 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 Table 2), this distinction will be useful in the minimization strategy, see Section 4 below.

2.3. 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. The statistical sample is obtained by generating N_{rep} pseudo-data 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 $g_{1,p}^{(\text{exp})} \equiv g_1(x_p, Q_p^2)$, we generate $k = 1, \ldots, N_{\text{rep}}$ artificial points $g_{1,p}^{(\text{art}),(k)}$ according to

Table 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 dataset are provided in Table 1.

Experiment	Set	$\langle \delta g_{1s} \rangle$	$\langle \delta g_{1c} \rangle$	$\langle \delta g_{1\text{tot}} \rangle$
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)

$$g_{1,p}^{(\text{art}),(k)}(x, Q^2) = \left[1 + \sum_i r_{(c),p}^{(k)} \sigma_{i,p}^{(c)} + r_{(u),p}^{(k)} \sigma_p^{(u)} \right] g_{1,p}^{(\text{exp})}(x, Q^2), \qquad (27)$$

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. (27) 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. 2, where we display scatter-plots of the central values and errors for samples of $N_{\text{rep}} = 10, 100$ and 1000 replicas.
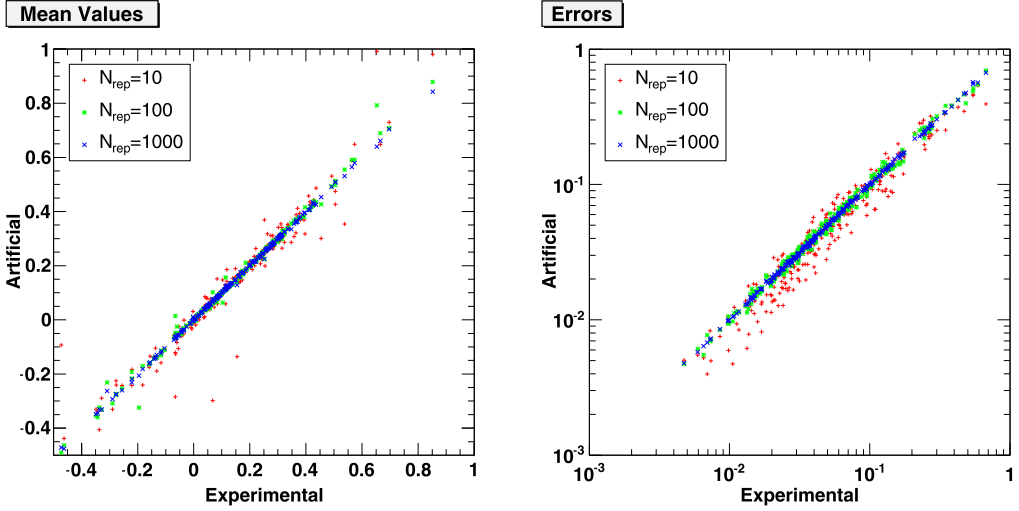
Fig. 2. Scatter-plot of experimental versus artificial Monte Carlo mean central values and absolute uncertainties of polarized structure functions computed from ensembles made of $N_{rep} = 10, 100, 1000$ replicas.

Table 3
Table of statistical estimators for the mean value computed from the Monte Carlo sample with $N_{rep} = 10, 100, 1000$ replicas. Estimators refer to individual experiments and are defined in Appendix B of Ref. [18].

	Estimator	$\langle PE[\langle g_1^{(art)} \rangle] \rangle$ [%]			$r[g_1^{(art)}]$		
	N_{rep}	10	100	1000	10	100	1000
Experiment	EMC	23.7	3.5	2.9	0.76037	0.99547	0.99712
	SMC	19.4	5.6	1.2	0.94789	0.99908	0.99993
	SMClowx	183	25.8	15.4	0.80370	0.99239	0.99960
	E143	18.5	5.7	2.1	0.99159	0.99860	0.99984
	E154	239	44.0	21.9	0.99635	0.99981	0.99994
	E155	37.3	13.4	4.3	0.99798	0.99993	0.99998
	COMPASS-D	26.4	8.6	3.2	0.96016	0.98774	0.99917
	COMPASS-P	16.4	1.9	1.5	0.91942	0.99829	0.99902
	HERMES97	22.5	6.2	2.2	0.96168	0.99762	0.99979
	HERMES	10.5	5.8	1.2	0.98564	0.99916	0.99973

A more quantitative comparison can be performed by defining suitable statistical estimators (see, for example, Appendix B of Ref. [18]).

We show in Tables 3–4 the percentage error and the scatter correlation r (which is crudely speaking the correlation between the input value and the value computed from the replica sample) for central values and errors respectively. We do not compute values for correlations, as these, as seen in Table 2, are only available for a very small number of data points from two experiments. Note that some large values of the percentage uncertainty are due to the fact that g_1 for some experiments 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 pseudo-data with $N_{rep} = 100$ 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 $N_{rep} = 1000$ is moderate. Therefore, we will henceforth use a $N_{rep} = 100$ replica sample as a default in the remainder of this paper.

Table 4

Table of statistical estimators for the errors computed from the Monte Carlo sample with $N_{rep} = 10, 100, 1000$ replicas. Estimators refer to individual experiments and are defined in Appendix B of Ref. [18].

	Estimator	$\langle PE[\langle \delta g_1^{(art)} \rangle] \rangle$ [%]			$r[\delta g_1^{(art)}]$		
	N_{rep}	10	100	1000	10	100	1000
Experiment	EMC	12.8	4.9	2.0	0.97397	0.99521	0.99876
	SMC	22.4	5.4	1.7	0.96585	0.99489	0.99980
	SMClowx	16.9	6.2	2.1	0.97959	0.99490	0.99905
	E143	16.0	7.4	2.0	0.95646	0.98684	0.99946
	E154	19.1	3.7	1.3	0.99410	0.99871	0.99992
	E155	21.2	5.6	1.8	0.99428	0.99971	0.99997
	COMPASS-D	15.5	5.2	1.6	0.99375	0.99687	0.99993
	COMPASS-P	18.4	7.4	1.5	0.99499	0.99005	0.99988
	HERMES97	17.9	6.4	1.6	0.89065	0.97318	0.99894
	HERMES	19.5	6.0	1.6	0.91523	0.99237	0.99942

3. From polarized PDFs to observables

3.1. Leading-twist factorization of the structure functions

At leading twist, the polarized structure function g_1 for neutral-current virtual photon DIS is given in terms of the polarized quark and gluon distributions by

$$g_1(x, Q^2) = \frac{\langle e^2 \rangle}{2} [C_{NS} \otimes \Delta q_{NS} + C_S \otimes \Delta \Sigma + 2n_f C_g \otimes \Delta g]. \tag{28}$$

Here n_f is the number of active flavors, the average charge is given by $\langle e^2 \rangle = n_f^{-1} \sum_{i=1}^{n_f} e_i^2$ in terms of the electric charge e_i of the i-th quark flavor, $\otimes$ denotes the convolution with respect to x, and the nonsinglet and singlet quark distributions are defined as

$$\Delta q_{NS} \equiv \sum_{i=1}^{n_f} \left(\frac{e_i^2}{\langle e^2 \rangle} - 1 \right) (\Delta q_i + \Delta \bar{q}_i), \qquad \Delta \Sigma \equiv \sum_{i=1}^{n_f} (\Delta q_i + \Delta \bar{q}_i), \tag{29}$$

where Δq_i and $\Delta \bar{q}_i$ are the polarized quark and antiquark distributions of flavor i and Δg is the polarized gluon PDF.

In the parton model, Eq. (28) reduces to

$$g_1(x, Q^2) = \frac{1}{2} \sum_{i=1}^{n_f} e_i^2 (\Delta q_i(x, Q^2) + \Delta \bar{q}_i(x, Q^2)), \tag{30}$$

but in perturbative QCD the parton model expression is not recovered even when $\alpha_s \to 0$ because at large Q^2 the first moment of the gluon distribution $\int_0^1 dx \, \Delta g \sim (\alpha_s(Q^2))^{-1}$, so the gluon does not decouple from g_1 asymptotically. Be that as it may, below charm threshold, with $n_f = 3$, Eq. (30) can be rewritten as

$$g_1(x, Q^2) = \frac{1}{9} \Delta \Sigma(x, Q^2) + \frac{1}{12} \Delta T_3(x, Q^2) + \frac{1}{36} \Delta T_8(x, Q^2), \tag{31}$$

in terms of the singlet quark–antiquark distribution $\Delta \Sigma(x, Q^2)$, defined in Eq. (29), the isospin triplet combination

$$\Delta T_3(x, Q_0^2) = \Delta u(x, Q_0^2) + \Delta \bar{u}(x, Q_0^2) - [\Delta d(x, Q_0^2) + \Delta \bar{d}(x, Q_0^2)], \tag{32}$$

and the SU(3) octet combination

$$\Delta T_8(x, Q_0^2) = \Delta u(x, Q_0^2) + \Delta \bar{u}(x, Q_0^2) + \Delta d(x, Q_0^2) + \Delta \bar{d}(x, Q_0^2)$$
$$- 2[\Delta s(x, Q_0^2) + \Delta \bar{s}(x, Q_0^2)]. \tag{33}$$

It is clear from Eqs. (28)–(31) that neutral current g_1 data only allow for a direct determination of the four polarized PDF combinations Δg, $\Delta \Sigma$, ΔT_3 and ΔT_8. In principle, an intrinsic polarized component could also be present for each heavy flavor. 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 (ZM-VFNS) scheme. In such a scheme all heavy quark mass effects are neglected. While they can be introduced for instance through the FONLL method [54], 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.

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

$$\Delta u^p = \Delta d^n, \qquad \Delta d^p = \Delta u^n, \qquad \Delta s^p = \Delta s^n, \tag{34}$$

and likewise for the polarized antiquarks. In the following we will always assume that PDFs refer to the proton. The first moment of all nonsinglet combinations of quark and antiquark distributions are scale-independent because of axial current conservation, while the first moment of the singlet quark distribution is not. Because of the axial anomaly, the first moment of the singlet quark distribution is scale-dependent in the $\overline{\text{MS}}$ scheme. However, it may be convenient to choose a factorization scheme in which the first moment of the singlet quark distribution is also scale-independent so that all the individual quark and antiquark spin fractions are scale-independent. Several such schemes, including the so-called Adler–Bardeen (AB) scheme, were discussed in Ref. [6], where the transformation connecting them to the $\overline{\text{MS}}$ scheme was constructed explicitly.

By means of the SU(2) or SU(3) flavor symmetry it is possible to relate the first moments of the nonsinglet C-even combinations (ΔT_3 and ΔT_8) to the baryon octet decay constants a_3 and a_8:

$$a_3 = \int_0^1 dx \, \Delta T_3(x, Q^2), \tag{35}$$

$$a_8 = \int_0^1 dx \, \Delta T_8(x, Q^2), \tag{36}$$

whose current experimental values are [55]

$$a_3 = g_A = 1.2701 \pm 0.0025, \tag{37}$$
$$a_8 = 0.585 \pm 0.025. \tag{38}$$

A much larger uncertainty on the octet axial charge, up to about 30%, is found if SU(3) symmetry is violated [56]. Even though a detailed phenomenological analysis does not seem to support this conclusion [57], we will take as default this more conservative uncertainty estimation

$$a_8 = 0.585 \pm 0.176. \tag{39}$$

The impact of replacing this with the more aggressive determination given in Eq. (38) will be studied in Section 5.3.2.

Structure functions will be computed in terms of polarized parton distributions using the so-called NNPDF `FastKernel` method, introduced in Ref. [23]. 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. 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 we have

$$g_1^p = \left\{ K_{g1,\Delta\Sigma} \otimes \Delta\Sigma_0 + K_{g1,\Delta g} \otimes \Delta g_0 + K_{g1,+} \otimes \left(\Delta T_{3,0} + \frac{1}{3}\Delta T_{8,0} \right) \right\}, \tag{40}$$

where the kernels $K_{g1,\Delta\Sigma}$, $K_{g1,\Delta g}$, $K_{g1,+}$ 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

$$g_1^n = 2\frac{g_1^d}{1 - 1.5\omega_D} - g_1^p, \tag{41}$$

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. (40) by interchanging the up and down quark PDFs, which amounts to changing the sign of ΔT_3.

The implementation of the polarized PDF evolution up to NLO has been benchmarked against the `HOPPET` evolution code [58] using the settings of the Les Houches PDF evolution benchmark tables [59]. This benchmarking is discussed in more detail in Appendix A. 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.

3.2. Target-mass corrections to g_1

The leading-twist expressions of structure functions given in Section 3.1 are corrected both by dynamical and kinematic higher-twist terms. The former are related to the contribution of higher-twist operators to the Wilson expansion, and are generally expected to be small. The latter are related to target-mass corrections (TMCs), and because of their kinematical origin they can be included exactly: we do this following Ref. [60]. As discussed in Section 2.1, we thus consistently include all nucleon mass effects, both in the relation between measured asymmetries and structure functions, and in the relation between the latter and parton distributions.

The target-mass corrections are especially simple in Mellin space, where they take the form [60]

$$\tilde{g}_1(N, Q^2) = g_1(N, Q^2) + \frac{m^2}{Q^2} \frac{N(N+1)}{(N+2)^2} \left[(N+4)\, g_1(N+2, Q^2) \right.$$
$$\left. + 4 \frac{N+2}{N+1}\, g_2(N+2, Q^2) \right] + \mathcal{O}\left(\frac{m^2}{Q^2}\right)^2, \tag{42}$$

$$\tilde{g}_2(N, Q^2) = g_2(N, Q^2) + \frac{m^2}{Q^2} \frac{N(N-1)}{(N+2)^2} \left[N \frac{N+2}{N+1} g_2(N+2, Q^2) \right.$$
$$\left. - g_1(N+2, Q^2) \right] + \mathcal{O}\left(\frac{m^2}{Q^2}\right)^2. \tag{43}$$

We denote 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.

As discussed in Section 2.1, in the absence of precise data on the structure function g_2, we will either determine it using the Wandzura–Wilczek approximation (22) (which is uncorrected by target-mass effects [60]), or, as a cross-check, simply setting it to zero. In either case, we may then determine $\tilde{g}_1$, Eq. (42), in terms of g_1.

In the former (Wandzura–Wilczek) case, substituting Eq. (21) in Eq. (42) and taking the inverse Mellin transform, we get

$$\tilde{g}_1(x, Q^2) = \frac{1}{2\pi i} \int dN\, x^{-N} \left[1 + \frac{m^2 x^2}{Q^2} \frac{(N-2)^2(N-1)}{N^2} \right] g_1(N, Q^2), \tag{44}$$

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

$$\tilde{g}_1(x, Q^2) = g_1(x, Q^2) + \frac{m^2 x^2}{Q^2} \left[-5 g_1(x, Q^2) - x \frac{dg_1(x, Q^2)}{dx} \right.$$
$$\left. + \int_x^1 \frac{dy}{y} \left(8 g_1(y, Q^2) + 4 g_1(y, Q^2) \log \frac{x}{y} \right) \right]. \tag{45}$$

If instead $g_2 = 0$,

$$\tilde{g}_1(x, Q^2) = \frac{1}{2\pi i} \int dN\, x^{-N} \left[1 + \frac{m^2 x^2}{Q^2} \frac{(N^2 - 4)(N - 1)}{N^2} \right] g_1(N, Q^2), \tag{46}$$

whence

$$\tilde{g}_1(x, Q^2) = g_1(x, Q^2) + \frac{m^2 x^2}{Q^2} \left[-g_1(x, Q^2) - x \frac{dg_1(x, Q^2)}{dx} \right.$$
$$\left. - \int_x^1 \frac{dy}{y} \left(4 g_1(y, Q^2) + 4 g_1(y, Q^2) \log \frac{x}{y} \right) \right]. \tag{47}$$

The numerical implementation of Eq. (45) or Eq. (47) is difficult, because of the presence of the first derivative of g_1 in the correction term. Therefore, 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 the target mass corrected g_1 structure function is computed by means of Eqs. (45)–(47) using the g_1 obtained in the previous minimization step.

4. Neural networks and fitting strategy

We will now briefly review the NNPDF methodology for parton parametrization in terms of neural networks, and their optimization (fitting) through a genetic algorithm. The details of the procedure have been discussed in previous NNPDF papers, in particular Refs. [20,23,61]. Here we summarize the main steps of the whole strategy, and discuss in greater detail some points which are specific to the polarized case.

4.1. Neural network parametrization

Each of the independent polarized PDFs in the evolution basis introduced in Section 3.1, $\Delta\Sigma$, Δg, ΔT_3 and ΔT_8, is parametrized using a multi-layer feed-forward neural network [27]. 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. 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 [27]. Given that polarized data are much less abundant and affected by much larger uncertainties than unpolarized ones, this architecture is adequate also in the polarized case.

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 $NN_{\Delta f}$ as follows

$$\Delta\Sigma\left(x, Q_0^2\right) = (1-x)^{m_1} x^{-n_1} NN_{\Delta\Sigma}(x),$$
$$\Delta T_3\left(x, Q_0^2\right) = A_3 (1-x)^{m_3} x^{-n_3} NN_{\Delta T_3}(x),$$
$$\Delta T_8\left(x, Q_0^2\right) = A_8 (1-x)^{m_8} x^{-n_{\Delta T_8}} NN_{\Delta T_3}(x),$$
$$\Delta g\left(x, Q_0^2\right) = (1-x)^{m_g} x^{-n_g} NN_{\Delta g}(x). \tag{48}$$

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

$$m_{\text{eff}}\left(Q^2\right) \equiv \lim_{x\to 1} \frac{\ln \Delta f(x, Q^2)}{\ln(1-x)}, \tag{49}$$

$$n_{\text{eff}}\left(Q^2\right) \equiv \lim_{x\to 0} \frac{\ln \Delta f(x, Q^2)}{\ln \frac{1}{x}}, \tag{50}$$

where $\Delta f = \Delta\Sigma, \Delta T_3, \Delta T_8, \Delta g$. 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

Table 5
Ranges for the small- and large-x preprocessing exponents (48).

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]

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 Table 5, while the effective exponents obtained in our fit will be discussed in Section 5.5. It is apparent from Table 5 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. It is enough to perform this check at the input evolution scale, $Q_0^2 = 1 \text{ GeV}^2$.

Two of the PDFs in the parametrization basis (48), namely the nonsinglet triplet and octet ΔT_3 and ΔT_8, are supplemented by a prefactor. This is because these PDFs must satisfy the sum rules (35), (36), which are enforced by letting

$$
A_3 = \frac{a_3}{\int_0^1 dx \, (1-x)^{m_3} x^{-n_3} \text{NN}_{\Delta T_3}(x)},
$$
$$
A_8 = \frac{a_8}{\int_0^1 dx \, (1-x)^{m_8} x^{-n_8} \text{NN}_{\Delta T_8}(x)}.
$$
(51)

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. (37), (39).

4.2. Genetic algorithm minimization

As discussed at length in Ref. [20], 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 [20],

$$
E^{(k)} = \frac{1}{N_{\text{dat}}} \sum_{I,J=1}^{N_{\text{dat}}} \left(g_I^{(\text{art})(k)} - g_I^{(\text{net})(k)} \right) \left((\text{cov}^{-1}) \right)_{IJ} \left(g_J^{(\text{art})(k)} - g_J^{(\text{net})(k)} \right).
$$
(52)

Here $g_I^{(\text{art})(k)}$ is the value of the observable g_I at the kinematical point I corresponding to the Monte Carlo replica k, and $g_I^{(\text{net})(k)}$ is the same observable computed from the neural network PDFs; the covariance matrix $(\text{cov})_{IJ}$ is defined in Eq. (23).

The minimization procedure we adopt follows closely that of Ref. [19], to which we refer for a more general discussion. Minimization is performed by means of a genetic algorithm, which minimizes the figure of merit, Eq. (52) by creating, 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.

Table 6

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

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

Table 7

Values of the parameters of the genetic algorithm.

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

The parameters which characterize the behavior of the genetic algorithm are tuned in order to optimize the efficiency of the minimization procedure: here, we rely on previous experience of the development of unpolarized NNPDF sets. In particular, the algorithm is characterized by a mutation rate, which we take to decrease as a function of the number of iterations N_{ite} of the algorithm according to [20]

$$\eta_{i,j} = \eta^{(0)}_{i,j}/N^{r_\eta}_{\text{ite}}, \tag{53}$$

so that 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 Table 6. 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. [23].

Furthermore, following Ref. [23], 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^{mut}, we use a large population of mutants, $N^{a}_{\text{mut}} \gg 1$, so a larger space of mutations is being explored. At later stages of the minimization, as the minimum is approached, a smaller number of mutations $N^{b}_{\text{mut}} \ll N^{a}_{\text{mut}}$ is used. The values of the parameters $N^{\text{mut}}_{\text{gen}}$, N^{a}_{mut} and N^{b}_{mut} are collected in Table 7.

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 datasets 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. [23]. 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^{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 training experiments which have already reached the target value. This is done by minimizing, instead of the figure of merit (52), the weighted figure of merit

$$E_{\text{wt}}^{(k)} = \frac{1}{N_{\text{dat}}} \sum_{j=1}^{N_{\text{sets}}} p_j^{(k)} N_{\text{dat},j} E_j^{(k)}, \tag{54}$$

where $E_j^{(k)}$ is the error function for the j-th dataset with $N_{\text{dat},j}$ points, and the weights $p_j^{(k)}$ are given by

1. If $E_i^{(k)} \geqslant E_i^{\text{targ}}$, then $p_i^{(k)} = (E_i^{(k)}/E_i^{\text{targ}})^n$,
2. If $E_i^{(k)} < E_i^{\text{targ}}$, then $p_i^{(k)} = 0$,

with n a free parameter which essentially determines the amount of weighting. In the unpolarized fits of Refs. [23,25,26,29] 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_i^{targ} 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 Section 4.3 below). 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 $N_{\text{gen}}^{\text{wt}}$ generations, unless the total error function (52) is above some threshold $E^{(k)} \geqslant E^{\text{sw}}$. If it is, weighted training continues until $E^{(k)}$ falls below the threshold value. Afterwards, the error function is just the unweighted error function (52) computed on experiments. This ensures that the figure of merit behaves smoothly in the last stages of training. The values for the parameters $N_{\text{gen}}^{\text{wt}}$ and E^{sw} are also given in Table 7.

4.3. Determination of the optimal fit

Because the neural network parametrization is very 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 (52). We thus determine the best fit, as in Refs. [19,20], using a cross-validation method [62]: for each replica, the data are randomly divided in two sets, training and validation, which include a fraction $f_{\text{tr}}^{(j)}$ and $f_{\text{val}}^{(j)} = 1 - f_{\text{tr}}^{(j)}$ of the data points respectively. The figure of merit (52) 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. [19,20,23,25,26,29] equal training and validation fractions were uniformly chosen, $f_{\text{tr}}^{(j)} = f_{\text{val}}^{(j)} = 1/2$. 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 Table 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. $f_{\text{tr}}^{(j)} = 0.8$ and $f_{\text{val}}^{(j)} = 0.2$. 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 $f_{\text{tr}}^{(j)} = 1.0$ and $f_{\text{val}}^{(j)} = 0.0$), 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 $N_{\text{gen}}^{\text{wt}}$ minimization steps. Furthermore, we check that all experiments have reached a value of the figure of merit below a minimal threshold E_{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^{sw}, the total figure of merit must be below this value in order for stopping to be possible.

We then compute moving averages

$$\langle E_{\text{tr,val}}(i) \rangle \equiv \frac{1}{N_{\text{smear}}} \sum_{l=i-N_{\text{smear}}+1}^{i} E_{\text{wt;tr,val}}(l), \tag{55}$$

of the figure of merit (54) for either the training or the validation set at the l-th genetic minimization step. The fit is then stopped if

$$r_{\text{tr}} < 1 - \delta_{\text{tr}} \quad \text{and} \quad r_{\text{val}} > 1 + \delta_{\text{val}}, \tag{56}$$

where

$$r_{\text{tr}} \equiv \frac{\langle E_{\text{tr}}(i) \rangle}{\langle E_{\text{tr}}(i - \Delta_{\text{smear}}) \rangle}, \tag{57}$$

$$r_{\text{val}} \equiv \frac{\langle E_{\text{val}}(i) \rangle}{\langle E_{\text{val}}(i - \Delta_{\text{smear}}) \rangle}. \tag{58}$$

The parameter N_{smear} determines the width of the moving average; the parameter Δ_{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 δ_{tr}, δ_{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. [23], this is done studying the profiles of the error functions for individual dataset and for individual replicas. In order to avoid unacceptably long fits, training is stopped anyway when a maximum number of iterations $N_{\text{gen}}^{\text{max}}$ is reached, even though the stopping conditions (56) 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 $N_{\text{gen}}^{\text{max}}$ without the stopping criterion having been satisfied, we also check that the total figure of merit is below the value E^{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 Table 8.

An example of how the stopping criterium works in practice is shown in Fig. 3. We display the moving averages Eq. (55) of the training and validation error functions $\langle E_{\text{tr,val}}^{(k)} \rangle$, computed with the parameter settings of Table 8, 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

Table 8
Parameters for the stopping criterium.

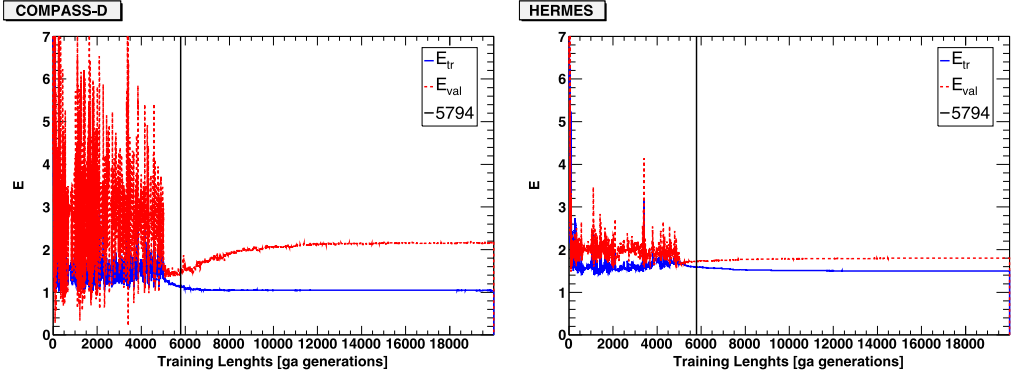| $N_{\text{gen}}^{\max}$	E_{thr}	N_{smear}	Δ_{smear}	δ_{tr}	δ_{val}
20 000	8	100	100	$5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$

Fig. 3. Behavior of the moving average Eq. (55) of the training and validation figure of merit for two different datasets 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 Table 8. The weighted training is switched off at $N_{\text{gen}}^{\text{wt}} = 5000$.

the fit. The wide fluctuations which are observed in the first part of training, up to the $N_{\text{gen}}^{\text{wt}}$-th generation, are due to the fact that the weights which enter the definition of the figure of merit (54) 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 see clearly that the minimization has entered an overlearning regime, in which the validation error function $E_{\text{val}}^{(k)}$ is rising while the training $E_{\text{tr}}^{(k)}$ is still decreasing. Note that the stopping point, which in this particular case occurs at $N_{\text{gen}}^{\text{stop}} = 5794$, 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 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.

4.4. 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 (1) implies that, up to power-suppressed corrections, longitudinal polarized structure functions are bounded by their unpolarized counterparts, i.e.

$$\left| g_1\left(x, Q^2\right) \right| \leqslant F_1\left(x, Q^2\right). \tag{59}$$

At leading order, structure functions are proportional to parton distributions, so imposing Eq. (59) for any process (and a similar condition on an asymmetry which is sensitive to polarized gluons [63]), would imply

$$\left|\Delta f_i(x, Q^2)\right| \leqslant f_i(x, Q^2) \tag{60}$$

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 (59) must still hold, but it does not necessarily imply Eq. (60). Rather, one should then impose at least a number of conditions of the form of Eq. (59) 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 (59) 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 [63], again measurable in principle if not in practice.

Because $g_1/F_1 \sim x$ as $x \to 0$ [64], the positivity bound (59) 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 [63,65]. Therefore, the NLO positivity bound in practice only differs from its LO counterpart (60) in a small region $10^{-2} \lesssim x \lesssim 0.3$, and even there by an amount of rather less that 10% [63], which is negligible in comparison to the size of PDF uncertainties, as we shall see explicitly in Section 5.

Therefore, we will impose the leading-order positivity bound (60) 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

$$\left|\Delta f_i(x, Q^2)\right| \leqslant f_i(x, Q^2) + \sigma_i(x, Q^2), \tag{61}$$

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 (60) would imply that the polarized PDF must vanish at the same point, which would be clearly meaningless.

As in Ref. [23] 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 λ_{pos}, i.e. by computing the polarized PDF at $N_{\text{dat,pos}}$ fixed kinematic points (x_p, Q_0^2) and then adding to the error function (52) a contribution

$$E_{\text{pos}}^{(k)} = \lambda_{\text{pos}} \sum_{p=1}^{N_{\text{dat,pos}}} \left\{ \sum_{j=u+\bar{u}, d+\bar{d}, s+\bar{s}, g} \Theta\left[\left|\Delta f_j^{(\text{net})(k)}(x_p, Q_0^2)\right| - (f_j + \sigma_j)(x_p, Q_0^2)\right] \right.$$
$$\left. \times \left[\left|\Delta f_j^{(\text{net})(k)}(x_p, Q_0^2)\right| - (f_j + \sigma_j)(x_p, Q_0^2)\right] \right\}. \tag{62}$$

This provides a penalty, proportional to the violation of positivity, which enforces Eq. (61) separately for all the nonzero 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 NNLO

PDF set [25], while $\Delta f_j^{(\text{net})(k)}$ is the corresponding polarized PDF computed from the neural network parametrization for the k-th replica. The polarized and unpolarized PDFs are evaluated at $N_{\text{dat,pos}} = 20$ points with x equally spaced in the interval

$$x \in [10^{-2}, 0.9]. \tag{63}$$

Positivity is imposed at the initial scale $Q_0^2 = 1 \text{ GeV}^2$ since once positivity is enforced at low scales, it is automatically satisfied at larger scales [63,65]. After stopping, we finally test the positivity condition (61) is satisfied on a grid of $N_{\text{dat,pos}} = 40$ 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 λ_{pos} was chosen. In the polarized case it turns out to be necessary to vary the Lagrange multiplier along the minimization. Specifically, we let

$$\begin{cases} \lambda_{\text{pos}} = \lambda_{\max}^{(N_{\text{gen}}-1)/(N_{\lambda_{\max}}-1)}, & N_{\text{gen}} < N_{\lambda_{\max}}, \\ \lambda_{\text{pos}} = \lambda_{\max}, & N_{\text{gen}} \geqslant N_{\lambda_{\max}}. \end{cases} \tag{64}$$

This means that the Lagrange multiplier increases as the minimization proceeds, starting from $\lambda_{\text{pos}} = 1$, at the first minimization step, $N_{\text{gen}} = 1$, up to $\lambda_{\text{pos}} = \lambda_{\max} \gg 1$ when $N_{\text{gen}} = N_{\lambda_{\max}}$. After $N_{\lambda_{\max}}$ generations λ_{pos} is then kept constant to $\lambda_{\max}$. 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 (62) 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 $N_{\lambda_{\max}} = 2000$ and $\lambda_{\max} = 10$ 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 $N_{\lambda_{\max}}$. Then we have tried different values for $\lambda_{\max}$ 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 $\lambda_{\max}$ is rather small if compared to the analogous Lagrange multiplier used in the unpolarized case [25]. 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 (52), 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 minimization 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 (61): 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 $\lambda_{\text{pos}} = 100$ 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 (61) on the fitted replicas and on polarized PDFs in Section 5.

Finally, as already mentioned, we impose an integrability constraint. The requirement that polarized PDFs be integrable, i.e. that they have finite first moments, 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,

$$I(x_1, x_2) = \int_{x_1}^{x_2} dx \, \Delta q_i\left(x, Q_0^2\right), \quad \Delta q_i = \Delta \Sigma, \Delta g, \Delta T_3, \Delta T_8 \tag{65}$$

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 practice, we test for the condition

$$\frac{I(x_1, x_2)}{I(x_1', x_2)} < \frac{\ln \frac{x_2}{x_1}}{\ln \frac{x_2}{x_1'}}, \tag{66}$$

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 $x_2 = 10^{-4}$.

5. Results

We now present the main result of this paper, namely 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 NNPDFpol1.0 PDFs to other recent polarized parton sets [11,12,14,15]. We will finally discuss the stability of our results upon the variation of several theoretical and methodological assumptions: the treatment of target-mass corrections, the use of sum rules to fix the triplet and octet axial charges, the implementation of positivity of PDFs, and preprocessing of neural networks and its impact on small- and large-x behavior.

We will not discuss here the way predictions for PDFs and uncertainties are obtained from NNPDF replica sets, for which we refer to general reviews, such as Ref. [66].

5.1. Statistical features

The statistical features of the NNPDFpol1.0 analysis are summarized in Tables 9–10, for the full dataset and for individual experiments and sets respectively. The error function $\langle E \rangle$ Eq. (52) shown in the tables both for the total, training and validation datasets 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. (54) and by a Lagrange multiplier to enforce positivity according to Eq. (62). In the table we also show the average over all replicas $\langle \chi_{\text{tot}}^{2(k)} \rangle$ of $\chi_{\text{tot}}^{2(k)}$ computed for the k-th replica, which coincides with the figure of merit (54), but with the data replica $g_I^{(\text{art})(k)}$ replaced by the experimental data $g_I^{(\text{dat})}$. We finally show χ_{tot}^2, which coincides with the figure of merit (54), but again with $g_I^{(\text{art})(k)}$ replaced

Table 9

Statistical estimators for `NNPDFpol1.0` with $N_{rep} = 100$ replicas.

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

Table 10

Same as Table 9 but for individual experiments.

Experiment	Set	χ^2_{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

by $g_I^{(dat)}$, and also with $g_I^{(net)(k)}$ replaced by $\langle g_I^{(net)(k)} \rangle$, 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, $\langle TL \rangle$, is also given in this table.

The distribution of $\chi^{2(k)}$, $E_{tr}^{(k)}$, and training lengths among the $N_{rep} = 100$ replicas are shown in Fig. 4 and Fig. 5 respectively. Note that the latter has a long tail which causes an accumulation
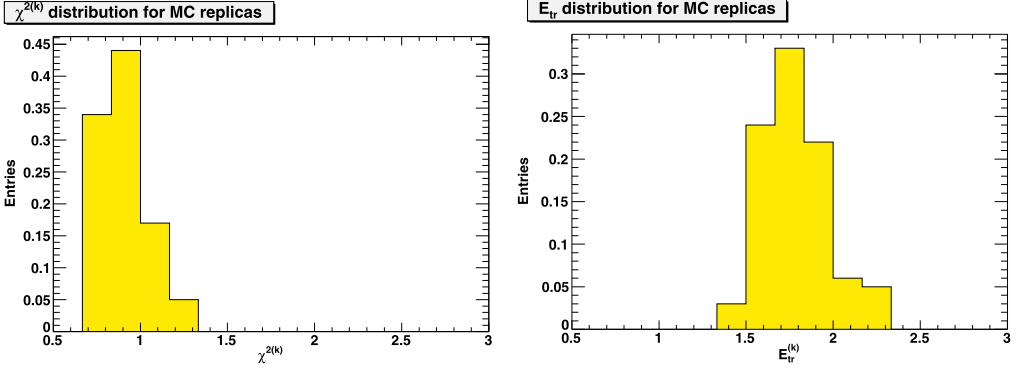
Fig. 4. Distribution of $\chi^{2(k)}$ and $E_{\text{tr}}^{(k)}$ over the sample of $N_{\text{rep}} = 100$ replicas.
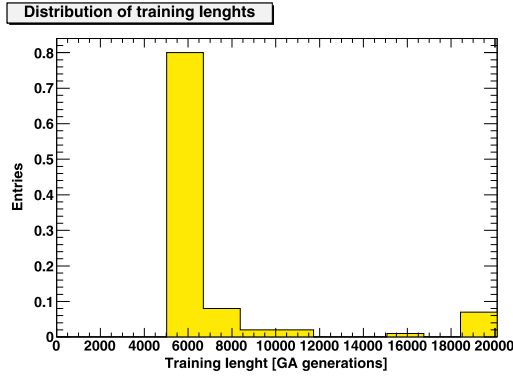
Fig. 5. Distribution of training lengths over the sample of $N_{\text{rep}} = 100$ replicas.

of points at the maximum training length, $N_{\text{gen}}^{\text{max}}$. 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 $\chi_{\text{tot}}^2 = 0.77$, 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 χ_{tot}^2 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 [49]: it shows that the neural net 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. 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 a con-
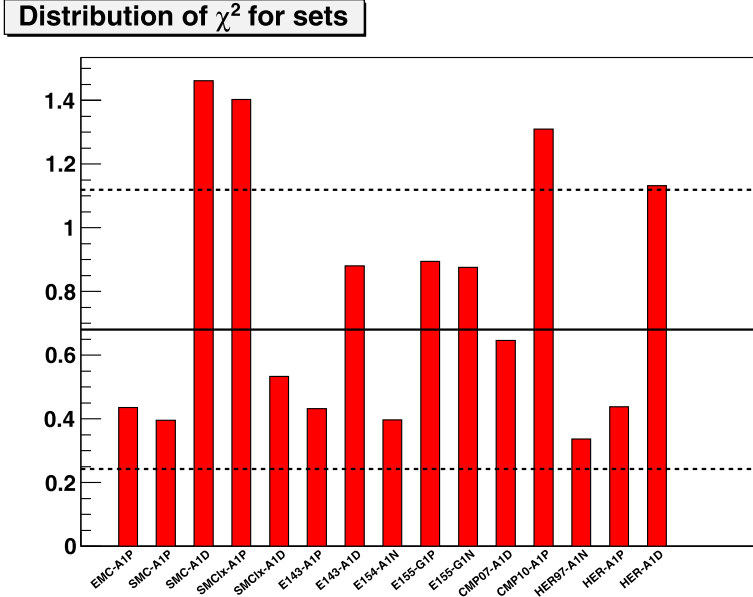
Fig. 6. Value of the χ^2 per data point for the datasets included in the `NNPDFpol1.0` reference fit, listed in Table 10. The horizontal line is the unweighted average of these χ^2 over the datasets and the black dashed lines give the one-sigma interval about it.

sequence 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.

5.2. Parton distributions

The `NNPDFpol1.0` parton distributions, computed from a set of $N_{rep} = 100$ replicas, are displayed in Fig. 7 at the input scale $Q_0^2 = 1 \text{ GeV}^2$, in the PDF parametrization basis (48) as a function of x both on a logarithmic and linear scale. In Figs. 8–9 the same PDFs are plotted in the flavor basis, and compared to other available NLO PDF sets: BB10 [11] and AAC08 [12] in Fig. 8, and DSSV08 [15] in Fig. 9. We do not show a direct comparison to the LSS polarized PDFs [14] because there are no publicly available routines for the computation of PDF uncertainties for this set. Note that the dataset used for the BB10 determination contains purely DIS data, and that for AAC contains DIS supplemented by some high-p_T RHIC pion production data: hence they are directly comparable to our PDF determination. The DSSV08 determination instead 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 flavor–antiflavor separation and a more direct handle on strangeness. All uncertainties in these plots correspond to the nominal 1-σ error bands.

The main conclusions of this comparison are the following:

- The central values of the $\Delta u + \Delta \bar{u}$ and the $\Delta d + \Delta \bar{d}$ 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
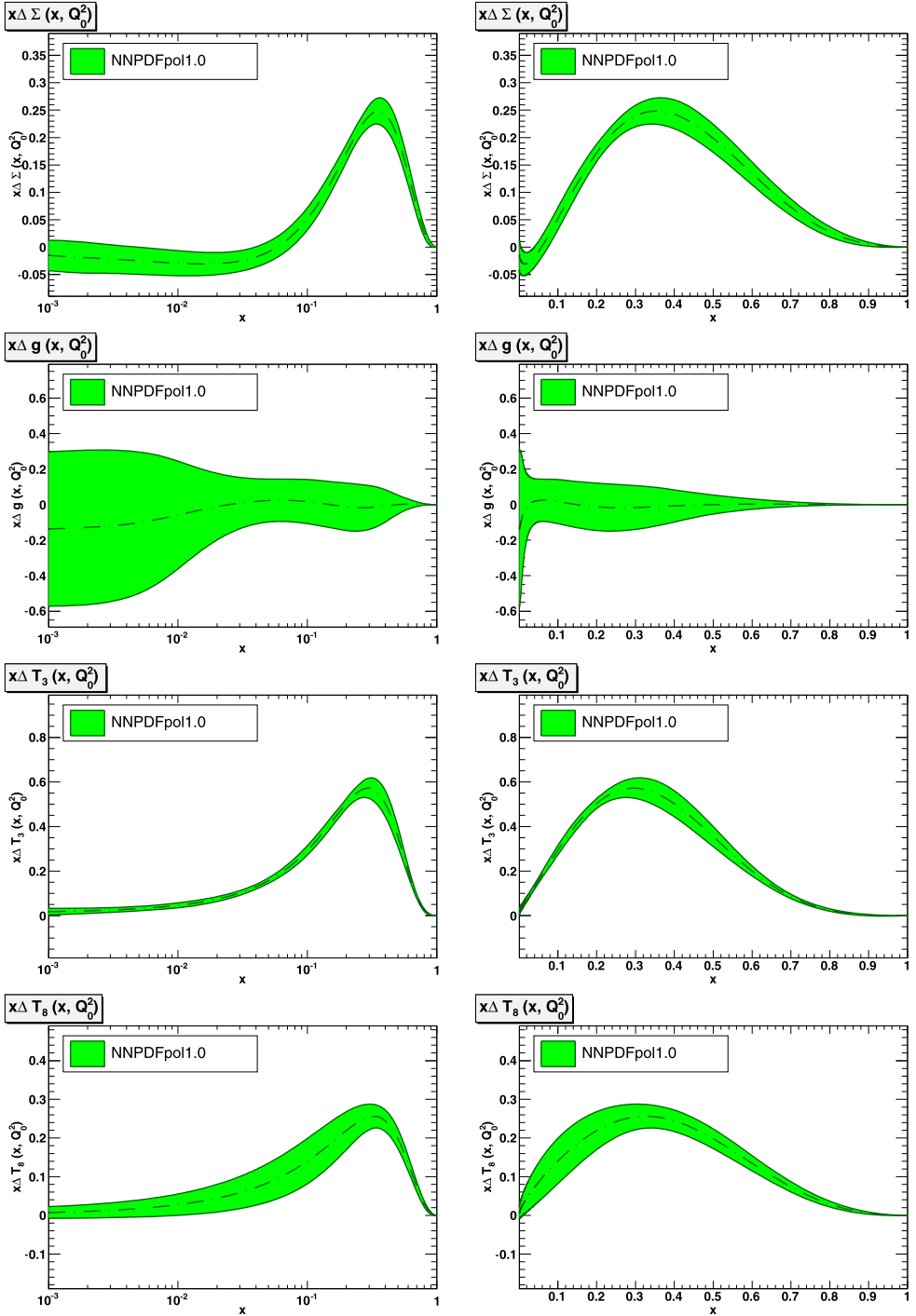
Fig. 7. The `NNPDFpol1.0` polarized parton distributions at $Q_0^2 = 1$ GeV^2 in the parametrization basis plotted as a function of x, on a logarithmic (left) and linear (right) scale.
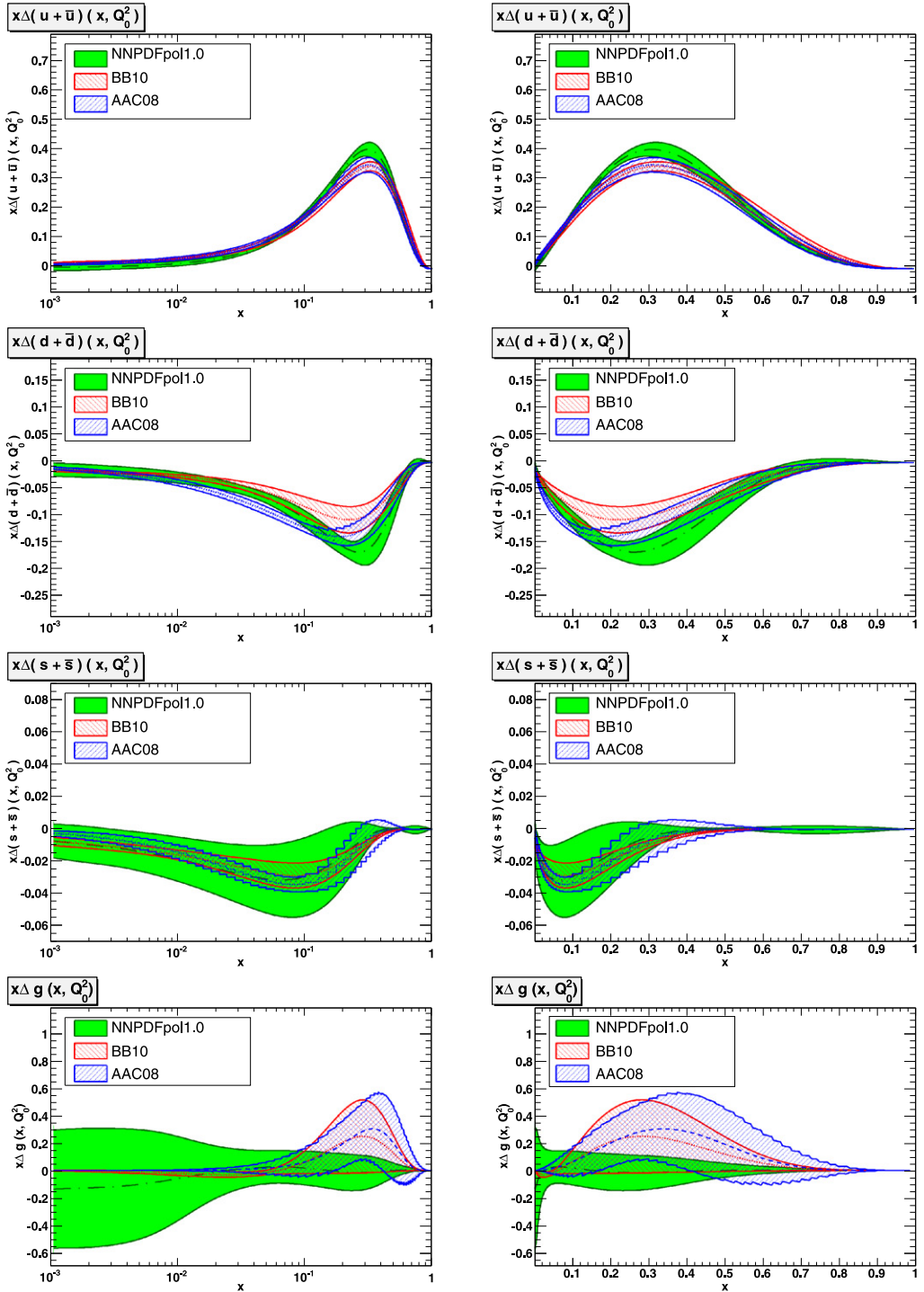
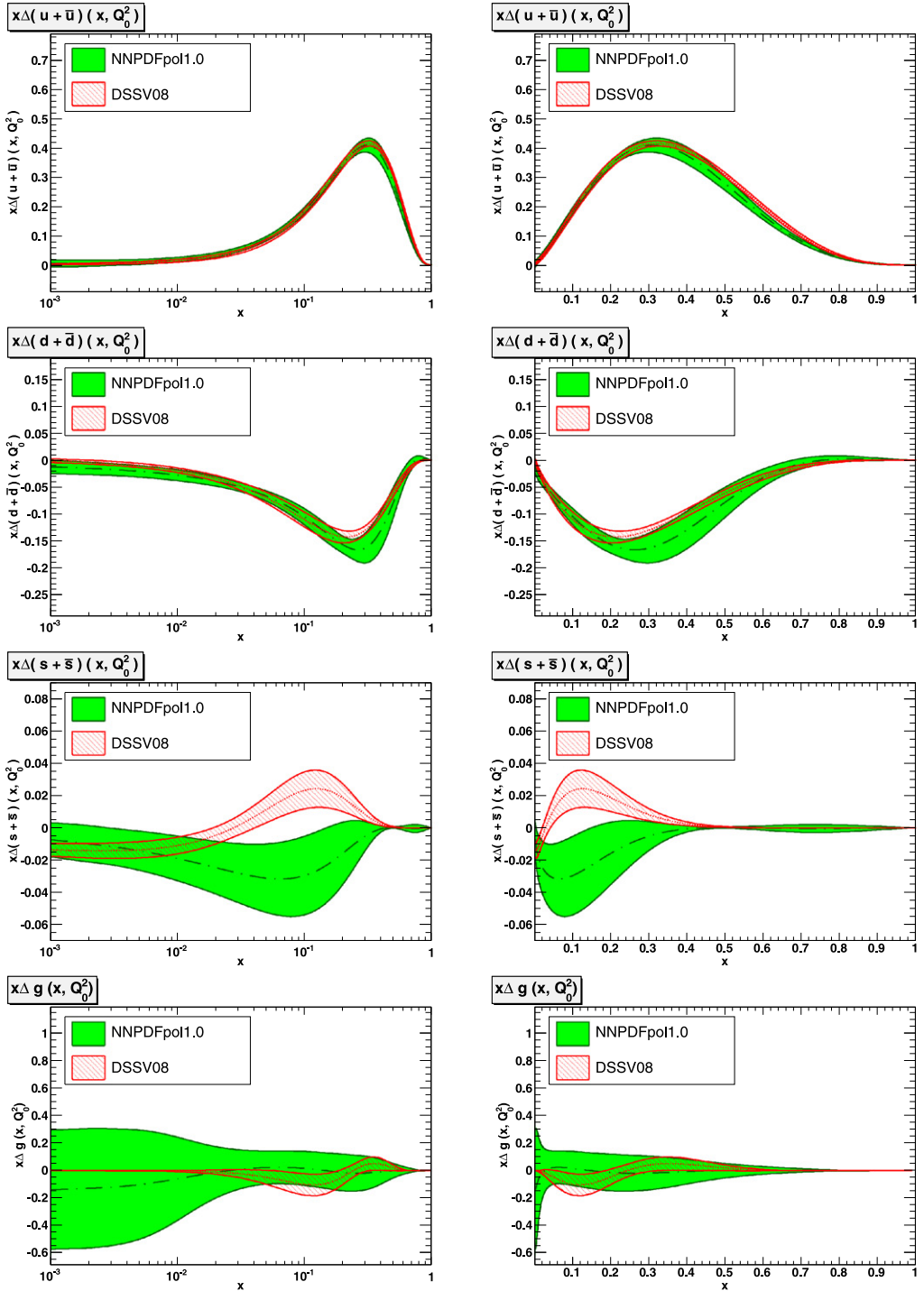Fig. 8. Comparison of the `NNPDFpol1.0` PDFs (in the flavor basis) and the BB10 [11] and AAC08 [12] PDFs.

Fig. 9. Comparison of the `NNPDFpol1.0` PDFs (in the flavor basis) and the DSSV08 PDFs [15].

these PDFs are generally slightly larger for NNPDF than for other sets, especially DSSV, which however is based on a much wider dataset.

- 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. The AAC08 and BB10 strange PDFs fall well within the `NNPDFpol1.0` uncertainty band.

- The `NNPDFpol1.0` determination of $\Delta s + \Delta \bar{s}$ is inconsistent at the two sigma level in the medium-small $x \sim 0.1$ region with DSSV08, which is also rather more accurate, as one would expect as it includes semi-inclusive data (in particular for production of hadrons with strangeness). This suggests a tension between the inclusive analysis data and the semi-inclusive analysis.

- 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 compatible with zero for all values of x.

- 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.

Finally, in Fig. 10 we compare the structure function $g_1(x, Q^2)$ for proton, deuteron and neutron, computed using `NNPDFpol1.0` (with its one-σ uncertainty band) to the experimental data included in the fit. Experimental data are grouped in bins of x with a logarithmic spacing, while the NNPDF 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 dataset 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.

5.3. Stability of the results

Our results have been obtained with a number of theoretical and methodological assumptions, discussed in Sections 3–4. We will now test their stability upon variation of these assumptions.

5.3.1. 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, as explained in Section 3.2, 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. (17) or Eq. (18) (see the discussion in Section 2.2). Our default PDF set is obtained assuming that g_2 is given by the Wandzura–Wilczek relation, Eq. (22).

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 $g_2 = 0$.
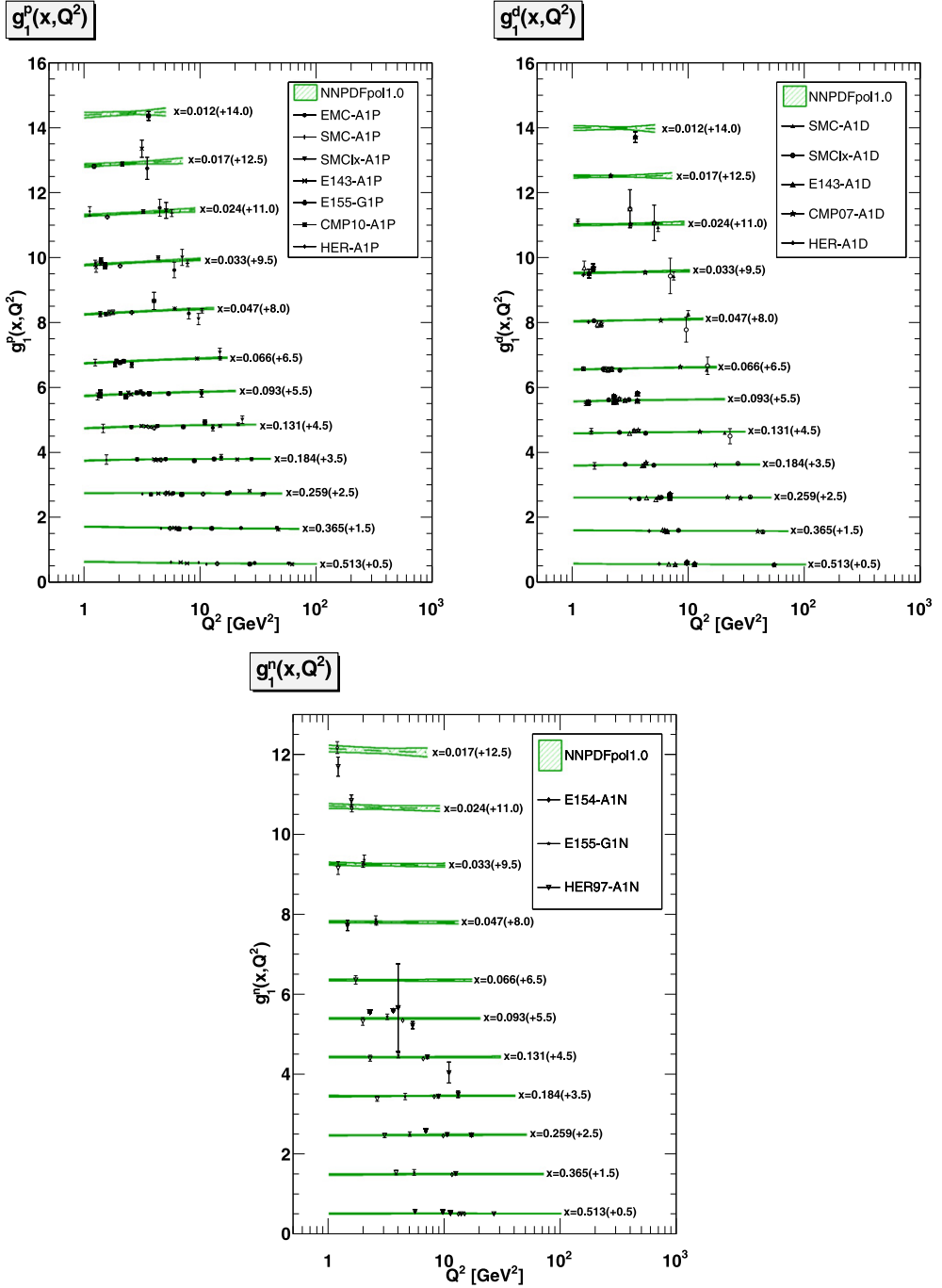
Fig. 10. The proton, neutron and deuteron polarized structure function $g_1(x, Q^2)$ as functions of Q^2 in different bins of x compared to experimental data. Experimental data are grouped in bins of x, while NNPDFpol1.0 results are given at the center of each bin, whose value is given next to each curve. In order to improve legibility, the values of $g_1(x, Q^2)$ have been shifted by the amount given next to each curve.

Table 11
The statistical estimators of Table 9 (obtained assuming $g_2 = g_2^{\mathrm{WW}}$) compared to a fit with $m = 0$ or with $g_2 = 0$.

Fit	NNPDFpol1.0 $g_2 = g_2^{\mathrm{WW}}$	NNPDFpol1.0 $m = 0$	NNPDFpol1.0 $g_2 = 0$
χ^2_{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
$\langle E_{\mathrm{tr}} \rangle \pm \sigma_{E_{\mathrm{tr}}}$	1.66 ± 0.49	1.62 ± 0.50	1.70 ± 0.38
$\langle E_{\mathrm{val}} \rangle \pm \sigma_{E_{\mathrm{val}}}$	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
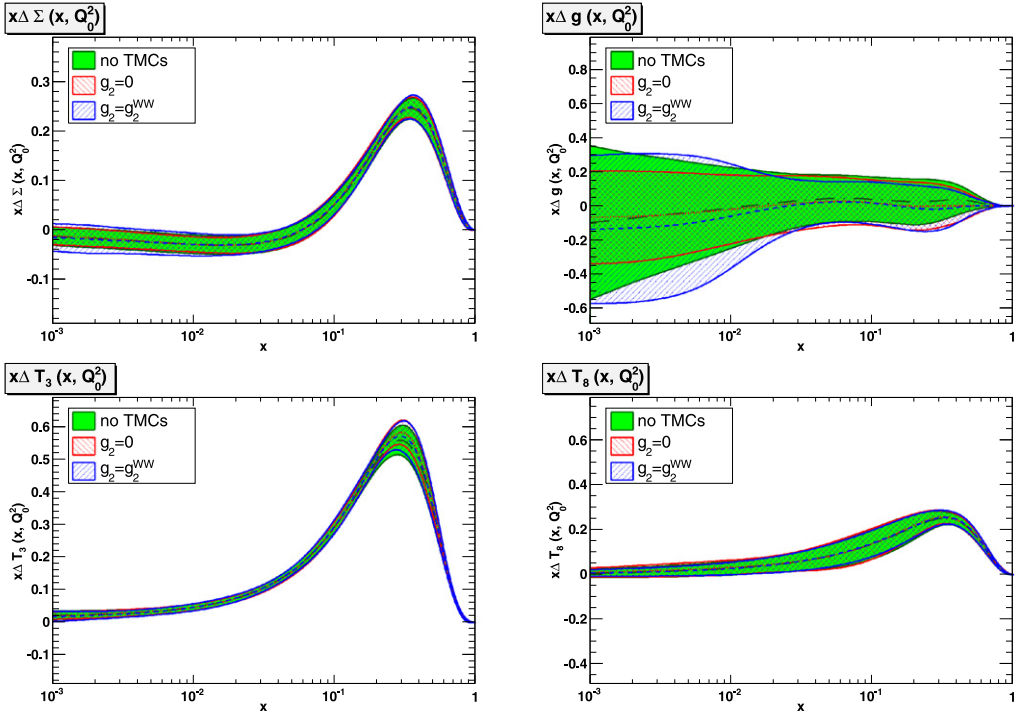
Fig. 11. Comparison between the default NNPDFpol1.0 PDFs (labeled as $g_2 = g_2^{\mathrm{WW}}$ in the plot), PDFs with $m = 0$ (labeled as noTMCs in the plot) and PDFs with $g_2 = 0$; each corresponds to the statistical estimators of Table 11.

The statistical estimators for each of these three fits over the full dataset are shown in Table 11. Clearly, all fits are of comparable quality.

Furthermore, in Fig. 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 of Ref. [23]. The distance is defined in such a way that if we compare two different samples of N_{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 $d = \sqrt{N_{\mathrm{rep}}}$ (the difference being due to the fact that the standard deviation of the mean scales as $1/\sqrt{N_{\mathrm{rep}}}$).
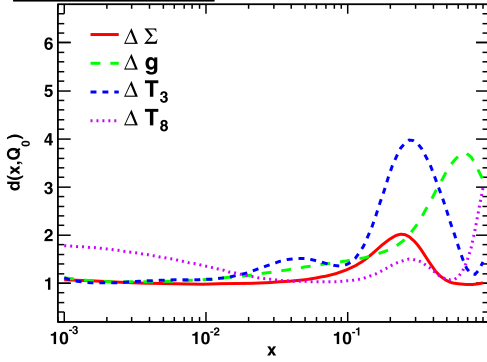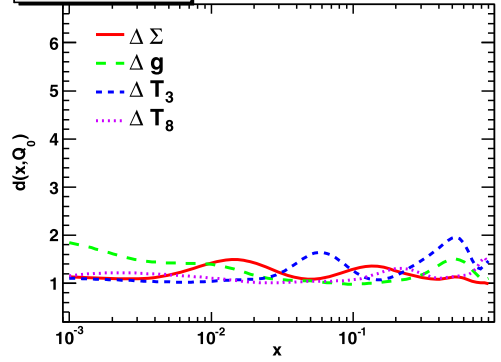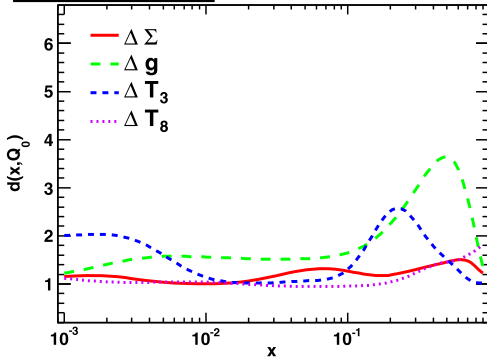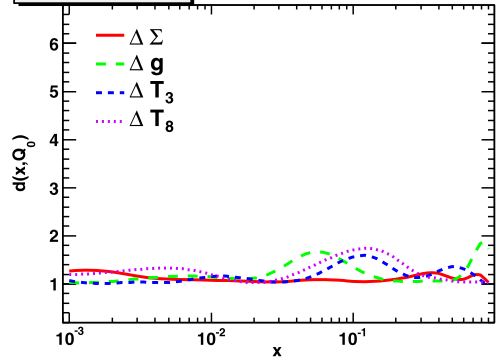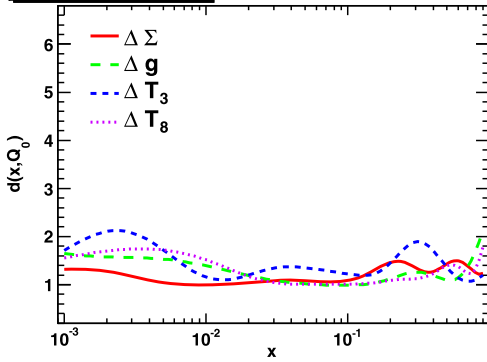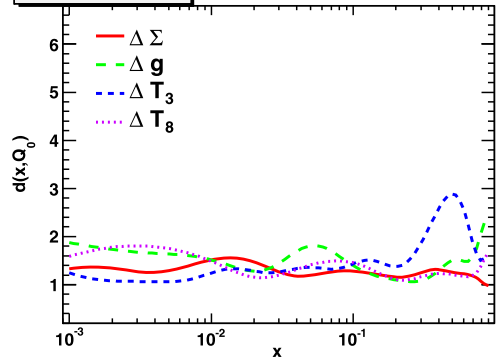
Fig. 12. Distances between each pair of the three sets of PDFs shown in Fig. 11.

Table 12
The statistical estimators of Table 9, 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 (38).

Fit	free a_3	a_8 Eq. (38)
χ^2_{tot}	0.79	0.77
$\langle E \rangle \pm \sigma_E$	1.84 ± 0.19	1.86 ± 0.19
$\langle E_{\text{tr}} \rangle \pm \sigma_{E_{\text{tr}}}$	1.73 ± 0.41	1.66 ± 0.53
$\langle E_{\text{val}} \rangle \pm \sigma_{E_{\text{val}}}$	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

The distances $d(x, Q^2)$ between central values and uncertainties of the three fits of Table 11 are shown in Fig. 12. They never exceed $d = 4$, which means less than half a standard deviation for $N_{\text{rep}} = 100$. 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.

5.3.2. 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. (37), and that the octet axial charge a_8 is fixed to the value of a_8 determined from baryon octet decays, but with an inflated uncertainty in order to allow for SU(3) violation, Eq. (39). As discussed after Eq. (51) 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 (35), so in particular a_3 is free and determined by the data, instead of being fixed to the value (37). In the second, we have assumed that the uncertainty on a_8 is given by the much smaller value of Eq. (38).

The statistical estimators for the total dataset for each of these fits are shown in Table 12. 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. 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 are compared in Fig. 14.

The distances between the default and the fit with the smaller uncertainty on a_8, Eq. (38), are shown in Fig. 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. 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.
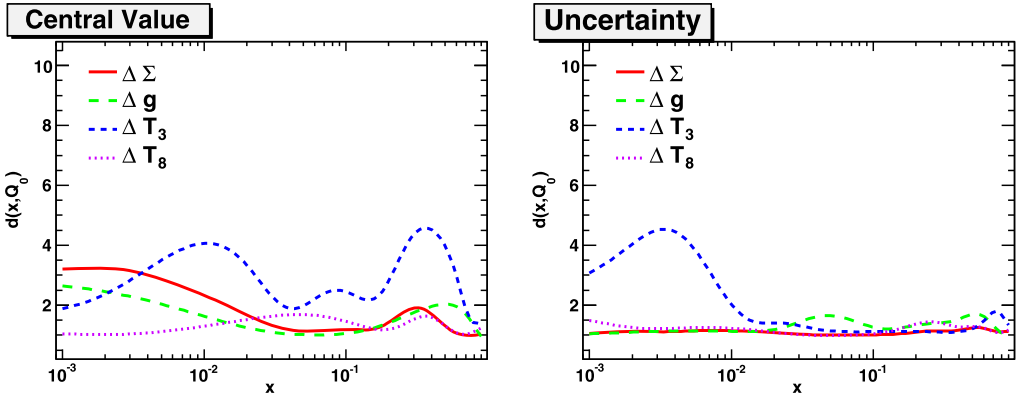
Fig. 13. Distances between PDFs (central values and uncertainties) for the default fit, with a_3 fixed, and the fit with free a_3, computed using $N_{rep} = 100$ replicas from each set.
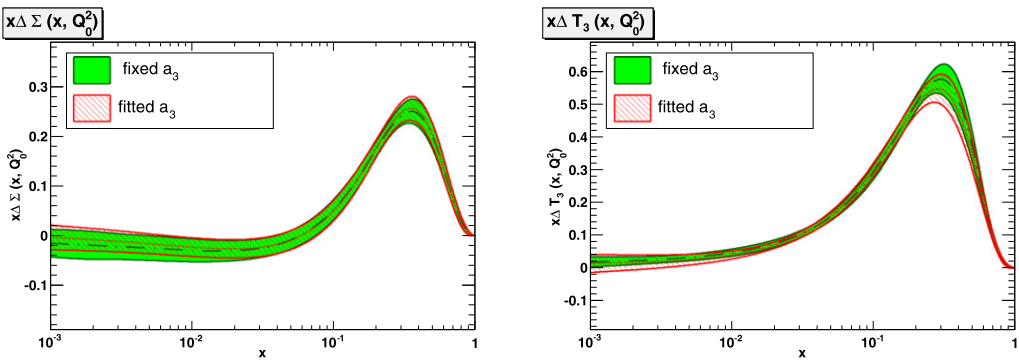

Fig. 14. Comparison of the singlet and triplet PDFs for the default fit, with a_3 fixed, and the fit with free a_3.
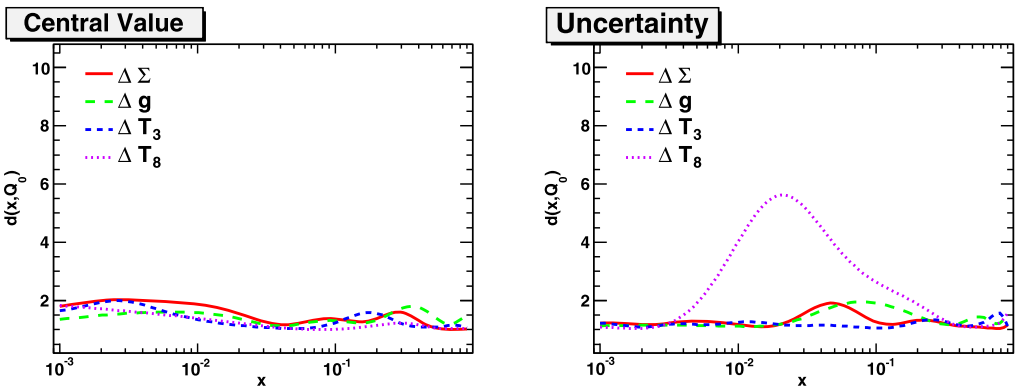

Fig. 15. Distances between PDFs (central values and uncertainties) for the default fit, with a_8 Eq. (39), and the fit with the value of a_8 with smaller uncertainty, Eq. (38).
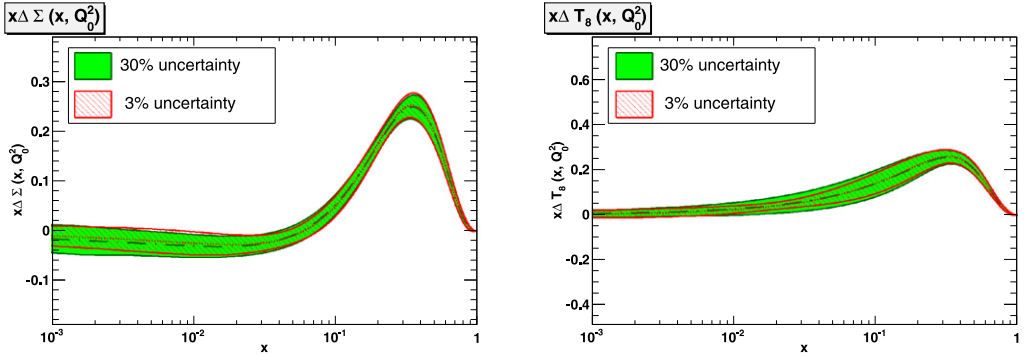
Fig. 16. Comparison of the singlet and octet PDFs for the default fit, with a_8 Eq. (39), and the fit with the value of a_8 with smaller uncertainty, Eq. (38).

We conclude that our fit results are quite stable upon variations of our treatment of both the triplet and the octet sum rules.

5.4. Positivity

As discussed in Section 4, positivity of the individual cross sections entering the polarization asymmetries (1) has been imposed at leading order according to Eq. (61), using the NNPDF2.1 NNLO PDF set [25], 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 (62). After stopping, positivity is checked a posteriori and replicas which do not satisfy it are discarded and retrained.

In Fig. 17 we compare to the positivity bound for the up, down, strange PDF combinations and gluon PDF a set of $N_{\text{rep}} = 100$ 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 $s + \bar{s}$ combination (and thus will be discarded) is seen.

In order to assess the effect of the positivity constraints we have performed a fit without 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 Section 4.1. The values of the large-x preprocessing exponents used in the fit without positivity are shown in Table 13. The small-x exponents are the same as in the baseline fit, Table 5.

The corresponding estimators are shown in Table 14. Also in this case, we see no significant change in fit quality, with only a slight improvement in χ^2_{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 in the large-x region, as shown in Fig. 18, where PDFs obtained from this fit are compared to the baseline. At small x, instead, the impact of positivity is moderate, because, as discussed in Section 4.4, $g_1/F_1 \sim x$ as $x \to 0$ [64] 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 [65] the bound is not significant.
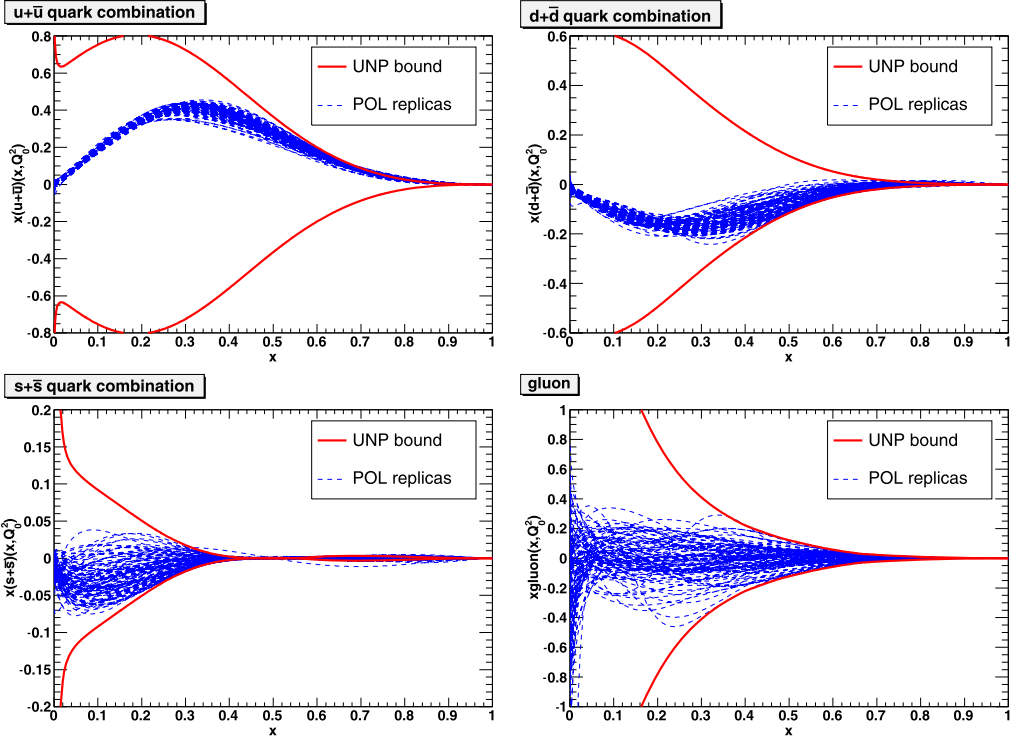
Fig. 17. The positivity bound (61), compared to a set of $N_{\text{rep}} = 100$ replicas (dashed lines).

Table 13

Ranges for the large-x preprocessing exponents (48) for the fit in which no positivity is imposed. The small-x exponents are the same as in the baseline fit Table 5.

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 14

The statistical estimators of Table 9 for a fit without positivity constraints.

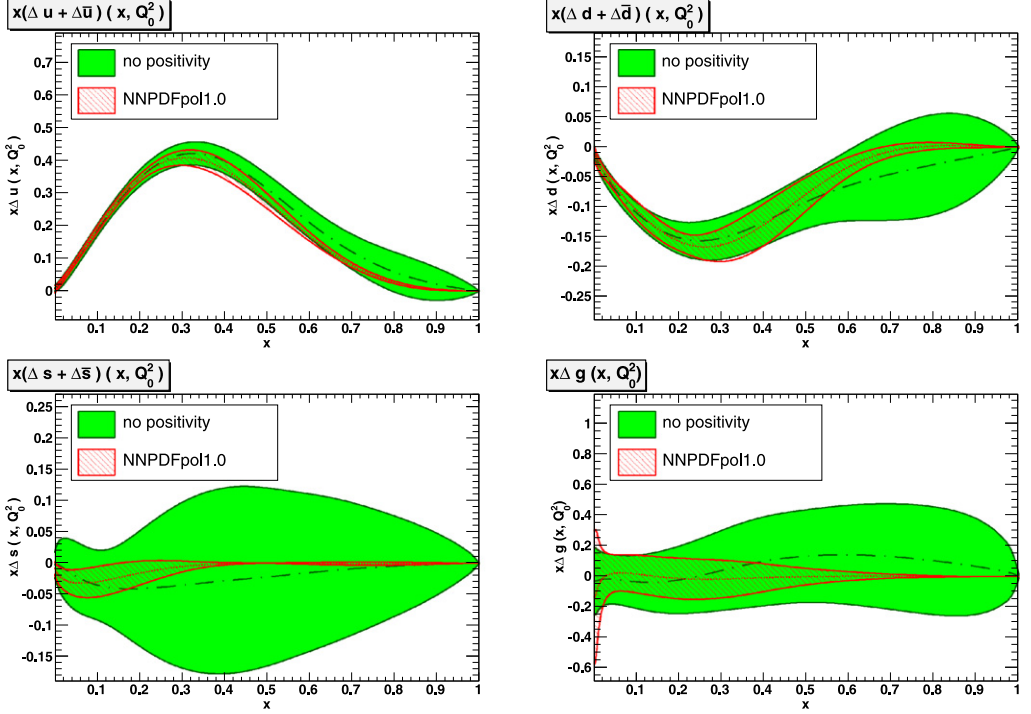| Fit	NNPDFpol1.0 no positivity
χ^2_{tot}	0.72
$\langle E \rangle \pm \sigma_E$	1.84 ± 0.22
$\langle E_{\text{tr}} \rangle \pm \sigma_{E_{\text{tr}}}$	1.60 ± 0.20
$\langle E_{\text{val}} \rangle \pm \sigma_{E_{\text{val}}}$	2.07 ± 0.39
$\langle \chi^{2(k)} \rangle \pm \sigma_{\chi^2}$	0.95 ± 0.16

Fig. 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.

5.5. 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. [64,67]). 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 Section 4.1. The outcome of the procedure is the set of exponents (48), listed in Table 5. The lack of bias with these choices is explicitly demonstrated in Fig. 19, where we plot the 68% confidence level of the distribution of

$$\alpha\big[\Delta q\big(x, Q^2\big)\big] = \frac{\ln \Delta f(x, Q^2)}{\ln \frac{1}{x}}, \tag{67}$$

$$\beta\big[\Delta q\big(x, Q^2\big)\big] = \frac{\ln \Delta f(x, Q^2)}{\ln(1-x)}, \tag{68}$$

$\Delta f = \Delta\Sigma$, 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 Table 5. 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

Table 15

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

	$\langle \Delta u + \Delta \bar{u} \rangle$				$\langle \Delta d + \Delta \bar{d} \rangle$				$\langle \Delta s + \Delta \bar{s} \rangle$			
	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 [68]	0.817	0.013	0.008	0.015	−0.453	0.011	0.036	0.038	−0.110	0.023	0.098	0.101

lie well within the range of the preprocessing exponents for all PDFs, thus confirming that the latter do not introduce any bias.

6. Polarized nucleon structure

The NNPDFpol1.0 PDF set may be used for a determination of the first moments of polarized parton distributions. As briefly summarized in the introduction, these are the quantities of greatest physical interest in that they are directly related to the spin structure of the nucleon, and indeed their determination, in particular the determination of the first moments of the quark and gluon distributions, has been the main motivation for the experimental campaign of g_1 measurements. The determination of the isotriplet first moment, because of the Bjorken sum rule, provides a potentially accurate and unbiased handle on the strong coupling α_s.

6.1. First moments

We have computed the first moments

$$\langle \Delta f (Q^2) \rangle \equiv \int_0^1 dx \, \Delta f(x, Q^2) \tag{69}$$

of each light polarized quark–antiquark and gluon distribution using a sample of $N_{\text{rep}} = 100$ NNPDFpol1.0 PDF replicas. The histogram of the distribution of first moments over the replica sample at $Q_0^2 = 1 \text{ GeV}^2$ are displayed in Fig. 20: they appear to be reasonably approximated by a Gaussian.

The central value and one-σ uncertainties of the quark first moments are listed in Table 15, while those of the singlet quark combination (29) and the gluon are given in Table 16. Results are compared to those from other parton sets, namely ABFR98 [8], DSSV10 [15], AAC08 [12], BB10 [11] and LSS10 [14]. Results from other PDF sets are not 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 DSSV and AAC results are shown at $Q_0^2 = 1 \text{ GeV}^2$, while the BB and LSS results are shown at $Q^2 = 4 \text{ GeV}^2$. For ease of reference, the NNPDF values for both scales are shown in Table 16.

In order to compare the results for first moments shown in Tables 15–16, it should be understood that the uncertainties shown, and sometimes also the central values, have somewhat different meanings. In particular:
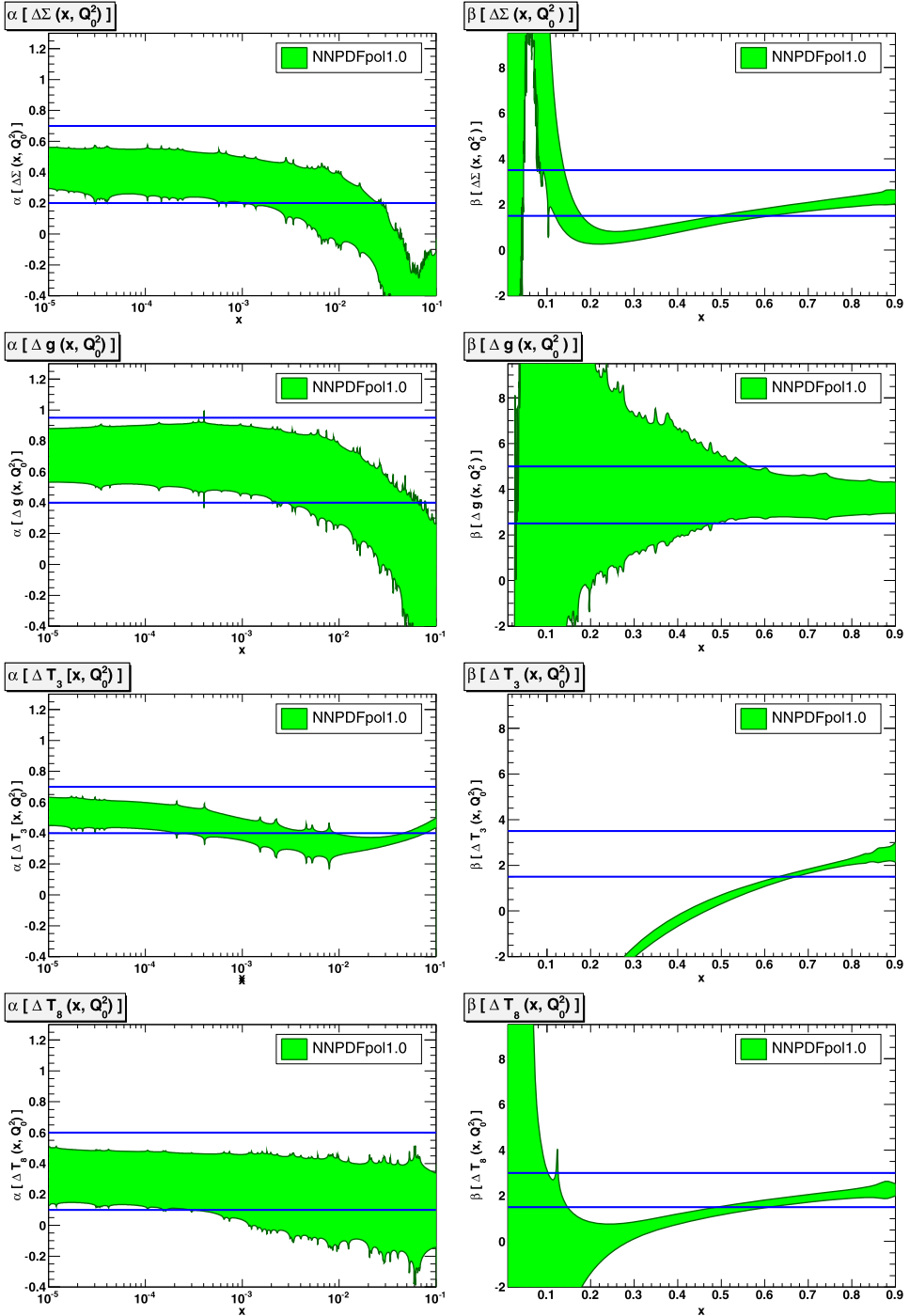
Fig. 19. The 68% confidence level of the distribution of effective small- and large-x exponents (67)–(68) for the default $N_{\rm rep} = 100$ replica NNPDFpol1.0 set at $Q_0^2 = 1$ GeV^2, plotted as functions of x. The range of variation of the preprocessing exponents of Table 5 is also shown in each case (solid lines).
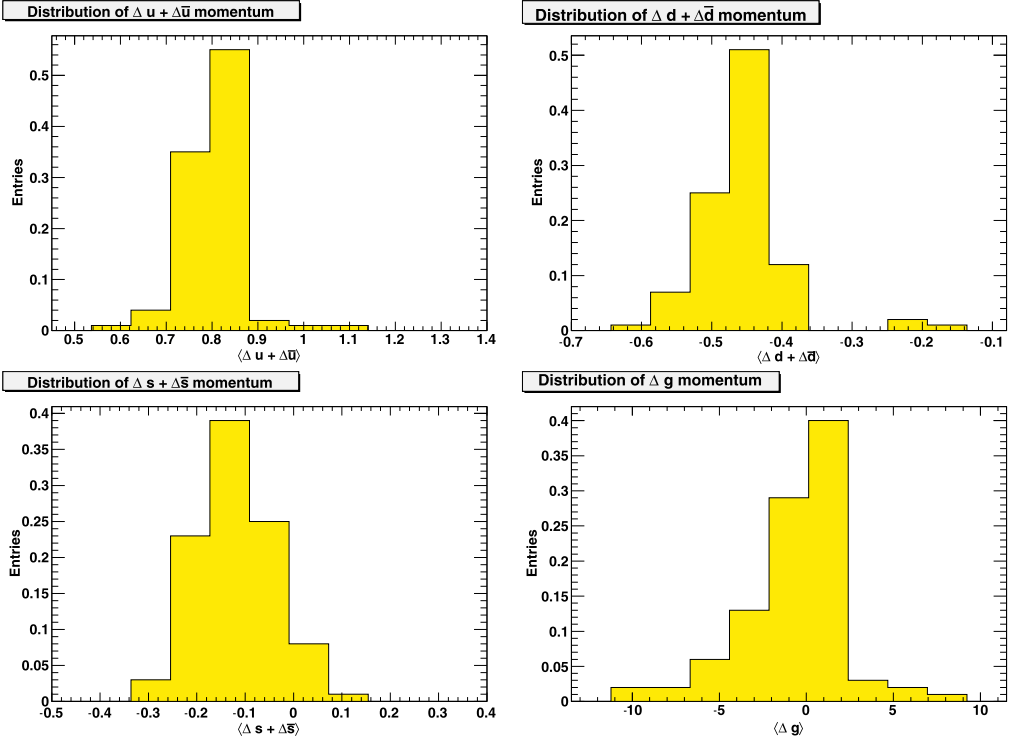
Fig. 20. Distribution of the first moments of $\Delta u + \Delta \bar{u}$ (top left), $\Delta d + \Delta \bar{d}$ (top right), $\Delta s + \Delta \bar{s}$ (bottom left) over a set of $N_{\rm rep} = 100$ `NNPDFpol1.0` PDF replicas.

- For `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.

- In the ABFR98 study, the central values were obtained in the so-called AB factorization scheme [6]. While the gluon in this scheme coincides with the gluon in the $\overline{\rm MS}$ scheme used here (and thus the value from Ref. [8] for the gluon is shown in Table 16), the quark singlet differs from it. However, in Ref. [8] a value of the singlet axial charge a_0 in the limit of infinite Q^2 was also given. In the $\overline{\rm MS}$ scheme, the singlet axial charge and the first moment of $\Delta \Sigma$ coincide [6], 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. [8], at NLO and with $\alpha_s(M_z) = 0.118$ [69] (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. 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 due to 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) (uncertainty on a_8 Eq. (36)), 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 NNPDF.

Table 16

Same as Table 15, but for the total singlet quark distribution and the gluon distribution. The NNPDF results are shown both at $Q_0^2 = 1$ GeV^2 and $Q^2 = 4$ GeV^2, the ABFR, DSSV and AAC results are shown at $Q_0^2 = 1$ GeV^2, and the BB10 and LSS10 are shown at $Q^2 = 4$ GeV^2.

		$\langle \Delta \Sigma \rangle$				$\langle \Delta g \rangle$			
		cv	exp	th	tot	cv	exp	th	tot
NNPDFpol1.0 (1 GeV^2)		0.22	0.20	–	0.20	−1.2	4.2	–	4.2
NNPDFpol1.0 (4 GeV^2)		0.18	0.20	–	0.20	−0.9	3.9	–	4.2
ABFR98 [8]		0.12	0.05	$+0.19$ -0.12	$+0.19$ -0.13	1.6	0.4	0.8	0.9
DSSV08 [68]		0.255	0.019	0.126	0.127	−0.12	0.12	0.06	0.13
AAC08 [12]	(*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 [11]		0.19	0.08	0.23	0.24	0.46	0.43	0.004	0.43
LSS10 [14]	(*positive*)	0.207	0.034	–	0.034	0.316	0.190	–	0.190
	(*node*)	0.254	0.042	–	0.042	−0.34	0.46	–	0.46

- For DSSV08 and BB10 PDFs, the central value is obtained by computing the first moment 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.
- For 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 a so-called tolerance [70] 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.
- For LSS10, 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. 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. Note 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 (Table 15) we find excellent agreement in the central values obtained between NNPDFpol1.0 and DSSV08; the NNPDF uncertainties are rather larger, but this could also be due to the fact that the DSSV08 dataset is sensitive to flavor 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 truncated moment in the region $x \in [10^{-3}, 1]$:

$$\int_{10^{-3}}^{1} dx\, \Delta g\left(x, Q^2 = 1\ \text{GeV}^2\right) = -0.26 \pm 1.19, \tag{70}$$

to be compared with the result of Table 16, 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 results 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 [32,71].

6.2. The Bjorken sum rule

Perturbative factorization, expressed in this context by Eq. (28) for the structure function $g_1(x, Q^2)$, and the assumption of exact isospin symmetry, immediately lead to the so-called Bjorken sum rule (originally derived [72,73] using current algebra):

$$\Gamma_1^p\left(Q^2\right) - \Gamma_1^n\left(Q^2\right) = \frac{1}{6}\Delta C_{\text{NS}}\left(\alpha_s\left(Q^2\right)\right)a_3, \tag{71}$$

where

$$\Gamma_1^{p,n}\left(Q^2\right) \equiv \int_0^1 dx\, g_1^{p,n}\left(x, Q^2\right), \tag{72}$$

and $\Delta C_{\text{NS}}(\alpha_s(Q^2))$ is the first moment of the nonsinglet coefficient function, while a_3 is defined in Eq. (35).

Because the first moment of the nonsinglet coefficient function ΔC_{NS} is known up to three loops [74] and isospin symmetry is expected to hold to high accuracy, the Bjorken sum rule
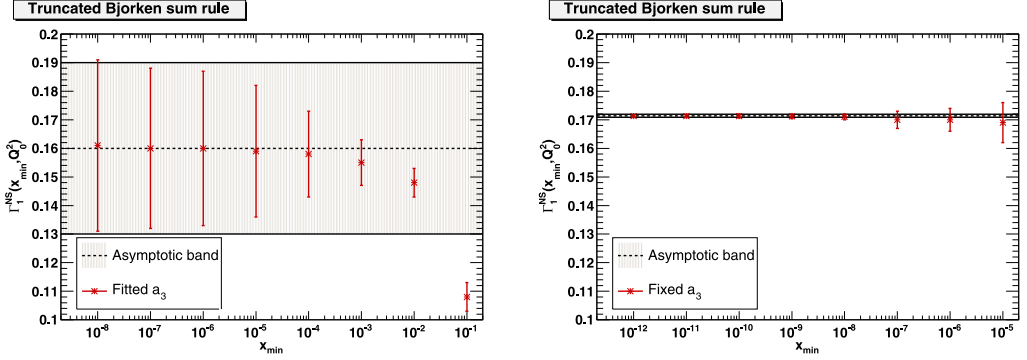
Fig. 21. The truncated Bjorken sum rule $\Gamma_1^{\mathrm{NS}}(Q^2, x)$ Eq. (73) plotted as a function of x for $Q^2 = 1$ GeV^2, for the fit with free a_3 (left) and for the reference fit with a_3 fixed to the value (37), (right panel). In the left plot, the shaded band corresponds to the asymptotic value of the truncated sum rule, Eq. (75), while in the right plot it corresponds to the experimental value (37).

Eq. (71) potentially provides a theoretically very accurate handle on the strong coupling constant: in principle, the truncated isotriplet first moment

$$\Gamma_1^{\mathrm{NS}}(Q^2, x_{\min}) \equiv \int_{x_{\min}}^{1} dx \left[g_1^p(x, Q^2) - g_1^n(x, Q^2) \right] \tag{73}$$

can be extracted from the data without any theoretical assumption. Given a measurement of $\Gamma_1^{\mathrm{NS}}(Q^2, 0)$ at one scale the strong coupling can then be extracted from Eq. (71) using the value of a_3 from β decays, while given a measurement of $\Gamma_1^{\mathrm{NS}}(Q^2, 0)$ at two scales both a_3 and the value of α_s can be extracted simultaneously.

In Ref. [75], a_3 and α_s were simultaneously determined from a set of nonsinglet truncated moments (both the first and higher moments), by exploiting the scale dependence of the latter [76], with the result $g_A = 1.04 \pm 0.13$ 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 [77].

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 Section 5.3.2. Using the results of this determination, we find

$$a_3 = \int_{0}^{1} dx \, \Delta T_3(x, Q^2) = 1.19 \pm 0.22. \tag{74}$$

The uncertainty is about twice that of the determination of Ref. [75]. 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 (74) is dominated by the small-x extrapolation.

To study this, in Fig. 21 we plot the value of the truncated Bjorken sum rule $\Gamma_1^{\text{NS}}(Q^2, x_{\min})$ Eq. (73) as a function of the lower limit of integration $x_{\min}$ at $Q_0^2 = 1 \text{ GeV}^2$, along with the asymptotic value

$$\Gamma_1^{\text{NS}}\big(1 \text{ GeV}^2, 0\big) = 0.16 \pm 0.03 \tag{75}$$

which at NLO corresponds to the value of a_3 given by Eq. (74). 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 (37). It is clear that indeed the uncertainty is completely dominated by the small-x extrapolation.

This suggests 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. [75], 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. [8], where values of a_3 and α_s similar to those of Ref. [75] were obtained.

7. Conclusions and outlook

We have presented a first determination of polarized parton distributions based on the NNPDF methodology: `NNPDFpol1.0`. We have determined polarized PDFs from the most recent inclusive data on proton, deuteron and neutron deep-inelastic polarized asymmetries and structure functions. Our main result is that the uncertainty in the gluon distribution, and to a lesser extent the strange distribution, and in the small-x extrapolation for all parton distributions, is rather larger than in previous polarized PDF determinations. Also, there seems to be some tension between strangeness determined in deep-inelastic scattering and using semi-inclusive data.

In particular, we find that the role of the gluon distribution in the spin structure of the nucleon is essentially unknown, as the first moment of the gluon distribution is compatible with zero, but with an uncertainty which is compatible with a very large positive or negative gluon spin fraction. Likewise, the contribution from the small-x region to the Bjorken sum rule makes its use as a means to determine α_s essentially impossible. Different conclusions can be reached only if one is willing to make strong theoretical assumptions on the small-x behavior of polarized PDFs.

Future experiments, in particular open charm and hadron production in fixed target experiments, [78,86] inclusive jet production [79,80] and W boson production [81–83] from the RHIC collider may improve the knowledge on individual polarized flavors and antiflavors and on the gluon distribution in the valence region. However, only a high-energy electron–ion collider [32, 71] might provide information on polarized PDFs at small x and thus reduce the uncertainty on first moments in a significant way.

The `NNPDFpol1.0` polarized PDFs, with $N_{\text{rep}} = 100$ replicas, are available from the NNPDF HEPFORGE web site, http://nnpdf.hepforge.org/. A Mathematica driver code is also available from the same source.

Acknowledgements

Table 17

Percentage difference between FastKernel perturbative evolution of polarized PDFs and the Les Houches benchmark tables [84] for different polarized PDF combinations at NLO in the ZM-VFNS scheme.

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

Appendix A. Benchmarking of polarized PDF evolution

We have benchmarked our implementation of the evolution of polarized parton densities by cross-checking against the Les Houches polarized PDF evolution benchmark tables [84]. Note that in Ref. [84] the polarized sea PDFs are given incorrectly, and should be

$$
\begin{aligned}
x\Delta\bar{u} &= -0.045 x^{0.3}(1-x)^7, \\
x\Delta\bar{d} &= -0.055 x^{0.3}(1-x)^7.
\end{aligned}
\tag{A.1}
$$

These tables were obtained from a comparison of the HOPPET [58] and PEGASUS [85] 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 [84]. The relative difference $\epsilon_{\rm rel}$ between our PDF evolution and the benchmark tables of Refs. [84] at NLO in the ZM-VFNS scheme are tabulated in Table 17 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 of the polarized PDF evolution in the FastKernel framework is satisfactory for precision phenomenology.

References

 [1] European Muon Collaboration, J. Ashman, et al., Phys. Lett. B 206 (1988) 364.
 [2] European Muon, J. Ashman, et al., Nucl. Phys. B 328 (1989) 1.
 [3] G. Altarelli, G.G. Ross, Phys. Lett. B 212 (1988) 391.
 [4] G. Altarelli, B. Lampe, Z. Phys. C 47 (1990) 315.
 [5] D. de Florian, et al., Prog. Part. Nucl. Phys. 67 (2012) 251, arXiv:1112.0904.
 [6] R.D. Ball, S. Forte, G. Ridolfi, Phys. Lett. B 378 (1996) 255, arXiv:hep-ph/9510449.
 [7] G. Altarelli, et al., Nucl. Phys. B 496 (1997) 337, arXiv:hep-ph/9701289.
 [8] G. Altarelli, et al., Acta Phys. Polon. B 29 (1998) 1145, arXiv:hep-ph/9803237.
 [9] T. Gehrmann, W.J. Stirling, Phys. Rev. D 53 (1996) 6100, arXiv:hep-ph/9512406.
[10] J. Blumlein, H. Bottcher, Nucl. Phys. B 636 (2002) 225, arXiv:hep-ph/0203155.

[11] J. Blumlein, H. Bottcher, Nucl. Phys. B 841 (2010) 205, arXiv:1005.3113.
[12] Asymmetry Analysis Collaboration, M. Hirai, S. Kumano, Nucl. Phys. B 813 (2009) 106, arXiv:0808.0413.
[13] E. Leader, A.V. Sidorov, D.B. Stamenov, Phys. Rev. D 75 (2007) 074027, arXiv:hep-ph/0612360.
[14] E. Leader, A.V. Sidorov, D.B. Stamenov, Phys. Rev. D 82 (2010) 114018, arXiv:1010.0574.
[15] D. de Florian, et al., Phys. Rev. D 80 (2009) 034030, arXiv:0904.3821.
[16] S. Forte, Acta Phys. Polon. B 41 (2010) 2859, arXiv:1011.5247.
[17] S. Forte, G. Watt, arXiv:1301.6754, 2013.
[18] NNPDF Collaboration, L. Del Debbio, et al., JHEP 0503 (2005) 080, arXiv:hep-ph/0501067.
[19] NNPDF Collaboration, L. Del Debbio, et al., JHEP 0703 (2007) 039, arXiv:hep-ph/0701127.
[20] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 809 (2009) 1, arXiv:0808.1231.
[21] NNPDF Collaboration, R.D. Ball, et al., JHEP 1005 (2010) 075, arXiv:0912.2276.
[22] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 823 (2009) 195, arXiv:0906.1958.
[23] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 838 (2010) 136, arXiv:1002.4407.
[24] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 849 (2011) 112, arXiv:1012.0836.
[25] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 849 (2011) 296, arXiv:1101.1300.
[26] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 855 (2012) 153, arXiv:1107.2652.
[27] NNPDF Collaboration, R.D. Ball, et al., arXiv:1110.1863, 2011.
[28] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 855 (2012) 608, arXiv:1108.1758.
[29] NNPDF Collaboration, R.D. Ball, et al., Nucl. Phys. B 867 (2013) 244, arXiv:1207.1303.
[30] S. Forte, M.L. Mangano, G. Ridolfi, Nucl. Phys. B 602 (2001) 585, arXiv:hep-ph/0101192.
[31] M.L. Mangano, et al., arXiv:hep-ph/0105155, 2001.
[32] D. Boer, et al., arXiv:1108.1713, 2011.
[33] R.K. Ellis, W.J. Stirling, B. Webber, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. 8 (1996) 1.
[34] E143 Collaboration, K. Abe, et al., Phys. Rev. D 58 (1998) 112003, arXiv:hep-ph/9802357.
[35] E155 Collaboration, P. Anthony, et al., Phys. Lett. B 553 (2003) 18, arXiv:hep-ex/0204028.
[36] A. Airapetian, et al., Eur. Phys. J. C 72 (2012) 1921, arXiv:1112.5584.
[37] S. Wandzura, F. Wilczek, Phys. Lett. B 72 (1977) 195.
[38] A. Accardi, et al., JHEP 0911 (2009) 093, arXiv:0907.2942.
[39] J. Blumlein, H. Bottcher, arXiv:1207.3170, 2012.
[40] V. Braun, et al., Phys. Rev. D 83 (2011) 094023, arXiv:1103.1269.
[41] Spin Muon Collaboration, B. Adeva, et al., Phys. Rev. D 58 (1998) 112001.
[42] Spin Muon Collaboration, B. Adeva, et al., Phys. Rev. D 60 (1999) 072004.
[43] E154 Collaboration, K. Abe, et al., Phys. Rev. Lett. 79 (1997) 26, arXiv:hep-ex/9705012.
[44] E155 Collaboration, P.L. Anthony, et al., Phys. Lett. B 493 (2000) 19, arXiv:hep-ph/0007248.
[45] COMPASS Collaboration, V.Y. Alexakhin, et al., Phys. Lett. B 647 (2007) 8, arXiv:hep-ex/0609038.
[46] COMPASS Collaboration, M.G. Alekseev, et al., Phys. Lett. B 690 (2010) 466, arXiv:1001.4654.
[47] HERMES Collaboration, K. Ackerstaff, et al., Phys. Lett. B 404 (1997) 383, arXiv:hep-ex/9703005.
[48] HERMES Collaboration, A. Airapetian, et al., Phys. Rev. D 75 (2007) 012007, arXiv:hep-ex/0609039.
[49] S. Forte, et al., JHEP 0205 (2002) 062, arXiv:hep-ph/0204232.
[50] C. Simolo, arXiv:0807.1501, 2006.
[51] Jefferson Lab Hall A, X. Zheng, et al., Phys. Rev. C 70 (2004) 065207, arXiv:nucl-ex/0405006.
[52] CLAS Collaboration, R. Fatemi, et al., Phys. Rev. Lett. 91 (2003) 222002, arXiv:nucl-ex/0306019.
[53] CLAS Collaboration, K.V. Dharmawardane, et al., Phys. Lett. B 641 (2006) 11, arXiv:nucl-ex/0605028.
[54] S. Forte, et al., Nucl. Phys. B 834 (2010) 116, arXiv:1001.2312.
[55] Particle Data Group, K. Nakamura, et al., J. Phys. G 37 (2010) 075021.
[56] R. Flores-Mendieta, E.E. Jenkins, A.V. Manohar, Phys. Rev. D 58 (1998) 094028, arXiv:hep-ph/9805416.
[57] N. Cabibbo, E.C. Swallow, R. Winston, Ann. Rev. Nucl. Part. Sci. 53 (2003) 39, arXiv:hep-ph/0307298.
[58] G.P. Salam, J. Rojo, Comput. Phys. Commun. 180 (2009) 120, arXiv:0804.3755.
[59] Z.J. Ajaltouni, et al., arXiv:0903.3861, 2009.
[60] A. Piccione, G. Ridolfi, Nucl. Phys. B 513 (1998) 301, arXiv:hep-ph/9707478.
[61] J. Rojo, J.I. Latorre, JHEP 0401 (2004) 055, arXiv:hep-ph/0401047.
[62] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.
[63] G. Altarelli, S. Forte, G. Ridolfi, Nucl. Phys. B 534 (1998) 277, arXiv:hep-ph/9806345.
[64] R.D. Ball, S. Forte, G. Ridolfi, Nucl. Phys. B 444 (1995) 287, arXiv:hep-ph/9502340.
[65] S. Forte, G. Altarelli, G. Ridolfi, Nucl. Phys. B (Proc. Suppl.) 74 (1999) 138, arXiv:hep-ph/9808462.
[66] S. Alekhin, et al., arXiv:1101.0536, 2011.

[67] R.G. Roberts, The Structure of the Proton: Deep-Inelastic Scattering, Cambridge University Press, Cambridge, UK, 1990.
[68] D. de Florian, et al., Phys. Rev. Lett. 101 (2008) 072001, arXiv:0804.0422.
[69] Particle Data Group, J. Beringer, et al., Phys. Rev. D 86 (2012) 010001.
[70] J. Pumplin, et al., JHEP 0207 (2002) 012, arXiv:hep-ph/0201195.
[71] A. Deshpande, et al., Ann. Rev. Nucl. Part. Sci. 55 (2005) 165, arXiv:hep-ph/0506148.
[72] J. Bjorken, Phys. Rev. 148 (1966) 1467.
[73] J. Bjorken, Phys. Rev. D 1 (1970) 1376.
[74] S.A. Larin, J.A.M. Vermaseren, Phys. Lett. B 259 (1991) 345.
[75] L. Del Debbio, A. Guffanti, A. Piccione, JHEP 0911 (2009) 060, arXiv:0907.2506.
[76] S. Forte, L. Magnea, Phys. Lett. B 448 (1999) 295, arXiv:hep-ph/9812479.
[77] F. Close, R. Roberts, Phys. Lett. B 336 (1994) 257, arXiv:hep-ph/9407204.
[78] COMPASS Collaboration, C. Adolph, et al., arXiv:1202.4064, 2012.
[79] PHENIX Collaboration, A. Adare, et al., Phys. Rev. D 84 (2011) 012006, arXiv:1009.4921.
[80] STAR Collaboration, L. Adamczyk, et al., arXiv:1205.2735, 2012.
[81] STAR Collaboration, M.M. Aggarwal, et al., arXiv:1009.0326, 2010.
[82] PHENIX Collaboration, A. Adare, et al., arXiv:1009.0505, 2010.
[83] STAR Collaboration, J.R. Stevens, arXiv:1302.6639, 2013.
[84] M. Dittmar, et al., arXiv:hep-ph/0511119, 2005.
[85] A. Vogt, et al., Nucl. Phys. B (Proc. Suppl.) 183 (2008) 155, arXiv:0807.1238.
[86] COMPASS Collaboration, C. Adolph, et al., arXiv:1211.6849, 2012.