

## FUTURE TESTS OF PARTON DISTRIBUTIONS\* \*\*

JUAN CRUZ-MARTINEZ, STEFANO FORTE

Tif Lab, Dipartimento di Fisica, Università di Milano  
and

INFN, Sezione di Milano, Via Celoria 16, 20133 Milano, Italy

EMANUELE R. NOCERA

The Higgs Centre for Theoretical Physics, University of Edinburgh  
JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland*(Received March 15, 2021; accepted March 29, 2021)*

We discuss a test of the generalization power of the methodology used in the determination of parton distribution functions (PDFs). The “future test” checks whether the uncertainty on PDFs, in regions in which they are not constrained by current data, is compatible with future data. The test is performed by using the current optimized methodology for PDF determination, but with a limited dataset, as available in the past, and by checking whether results are compatible within uncertainty with the result found using a current more extensive dataset. We use the future test to assess the generalization power of the NNPDF4.0 unpolarized PDF and the NNPDFpol1.1 polarized PDF methodology. Specifically, we investigate whether the former would predict the rise of the unpolarized proton structure function  $F_2$  at small  $x$  using only pre-HERA data, and whether the latter would predict the so-called “proton spin crisis” using only pre-EMC data.

DOI:10.5506/APhysPolB.52.243

“But how can I use a method to discredit that very method,  
if the method is discreditable?”<sup>1</sup>

Stanisław Lem, *The Futurological Congress* [1]**1. Parton distributions in the era of precision**

Knowledge of parton distribution functions (PDFs), or lack thereof, is currently one of the dominant sources of uncertainty in the computation

\* Funded by SCOAP<sup>3</sup> under Creative Commons License, CC-BY 4.0.\*\* Prepared for the 60<sup>th</sup> anniversary of the Cracow School of Theoretical Physics.<sup>1</sup> Jakże jednak użyć metody, która dzięki sobie samej ma zostać podana w niesławę?

of LHC processes. Over the last several years, as the precision of PDF determinations has been gradually improving, a major issue has been that of making sure that this increased precision is matched by a correspondingly high accuracy. Specifically, the question is whether uncertainties due to the methodology used in PDF extraction are estimated accurately. As it is well known, the problem in determining parton distributions is the lack of knowledge of their underlying functional form. Since one is extracting a probability distribution for a set of functions from a discrete set of data points, it is necessary to make assumptions in order to make the problem solvable. The question is then how to reliably estimate the uncertainty coming from these assumptions.

Over the last several years, it has been shown that this problem can be successfully handled using machine-learning techniques, by essentially viewing it as a pattern recognition problem (see Ref. [2] for a recent review). The basic idea is that standard tools, such as neural networks, can be used to infer an underlying pattern, *i.e.* the shape of the PDF, without having to assume a functional form. Redundancy then ensures that no bias is introduced, while quality control tools, such as cross-validation, ensure that the most general result which is compatible with the data is obtained, without also reproducing statistical noise. Uncertainties are obtained as the spread of a population of equally good best-fits. The reader is referred to more detailed discussions [2–5] for a treatment of this machine-learning-based methodology and how it relates to other approaches used for PDF determination.

This paper deals with the issue of validating PDFs, specifically those determined using machine-learning methods. As we shall see, this validation problem really consists of two different problems: the first has to do with the ability of machine-learning tools to learn an unknown underlying law, and the second has to do with their generalization power. Whereas there are now relatively well-established tools for the validation of the former, here we discuss a new tool for the validation of the latter: the “future test”. We will first discuss in Section 2 the issue of PDF validation. In order to make the discussion self-contained, we briefly review, in Section 2.1, the relation between PDFs and the underlying data, and we then turn, in Section 2.2, to an explanation of the reason why a new validation method is needed, specifically in order to address generalization, and to a presentation of the idea of the future test. In the remaining two sections, we see the future test at work, by applying it to unpolarized (Section 3) and polarized (Section 4) PDFs. Specifically, we discuss the small- $x$  rise of the unpolarized structure function  $F_2$  at HERA, and the “spin crisis”, related to the first measurements of the polarized structure function  $g_1$ . In both cases, we address the issue whether these then-surprising discoveries could have been predicted, had modern-day methodology been available.

## 2. Validation of PDFs

The original motivation for introducing machine-learning tools in PDF determination [6] is to avoid the sources of bias involved in other methodologies, such as the need to pick a particular functional form, or an insufficiently flexible choice of model. The claim that a certain methodology is unbiased, or less biased, however, immediately raises the question: how can we make sure that results, and specifically uncertainties, are faithful? The choice of using machine-learning tools suggests that validation should be performed *a posteriori*, given the lack of direct control on the underlying model. In searching for an *a posteriori* validation methodology, there are then two rather distinct issues, corresponding to the kinematic regions we consider.

### 2.1. Factorization and kinematic regions

In order to understand the different issues related to PDF validation it is necessary to briefly recall the way PDFs are related to the observable cross sections that are used in order to determine them. Let us specifically consider a hadronic collision. A typical cross section  $\sigma(Q^2, \tau, \{k\})$  is expressed in terms of PDFs as

$$\sigma(Q^2, \tau, \{k\}) = \sum_{ij} \int_{\tau}^1 \frac{dz}{z} \mathcal{L}_{ij}(z, Q^2) \hat{\sigma}_{ij}\left(\frac{\tau}{z}, \alpha_s(Q^2), \{k\}\right), \quad (1)$$

where  $\hat{\sigma}_{ij}$  is the (“partonic”) cross section for producing the desired final state with incoming partons  $i, j$ ,  $Q^2$  is a hard scale (for example, for  $Z$  production, the  $Z$  mass  $Q^2 = M_Z^2$ ),  $\tau$  is a dimensionless scaling variable (for example, for the total inclusive  $Z$ -production cross section, the ratio  $\tau = M_Z^2/s$ , with  $s$  the center-of-mass energy), and  $\{k\}$  denotes the set of other kinematic variables that the (generally differential) cross section  $\sigma$  may depend upon (for example, the rapidity of the dilepton pair into which the  $Z$  decays). The dependence on the PDFs is contained in the luminosity

$$\mathcal{L}_{ij}(z, Q^2) = \int_x^1 \frac{dx}{x} f_i^{h_1}(x, Q^2) f_j^{h_2}\left(\frac{z}{x}, Q^2\right), \quad (2)$$

where  $f_i^{h_k}(x, Q^2)$  is the PDF for extracting a parton of species  $i$  from hadron  $h_k$ , carrying a fraction  $x$  of its momentum at scale  $Q^2$ .

Now, the structure of Eq. (2) trivially implies that only PDFs in a certain range are probed by any given physical process: specifically, the PDF for  $x < \tau$  completely decouples from an observable measured at  $\tau$ . Furthermore, the dominant contribution to any given process typically comes from

a restricted kinematic region: in the simplest case, from a single kinematic point. For example, if the  $Z$ -rapidity distribution is evaluated at leading order, the partonic cross section  $\hat{\sigma}$  is proportional to a two-dimensional Dirac delta which removes the convolution integrals from Eq. (2), and only PDFs evaluated for the fixed values  $x_1 = \tau \exp(\frac{1}{2}\eta)$ ,  $x_2 = \tau \exp(-\frac{1}{2}\eta)$ ,  $Q^2 = M_Z^2$  contribute, with  $\eta$  the rapidity of the  $Z$  boson. Even when the kinematics is completely fixed, it is only fixed at leading perturbative order; at higher perturbative orders, the cross section is no longer a Dirac delta, and all values of  $x$  which are larger than that fixed by leading-order kinematics are accessible, with the smaller  $x$  region remaining inaccessible. This is easy to understand physically: at higher perturbative order, a parton can always radiate other partons, so all partons that carry a momentum fraction of their parent hadron which is larger than the minimal value needed in order to produce the desired final state can contribute. However, they must at least carry the minimal required momentum fraction. Therefore, for any given process, there is a minimal value of  $x$ , though, in principle, no maximal value.

In Fig. 1, we show a scatter plot of data used for a current determination of unpolarized PDFs: each point corresponds to the value of  $x$  and  $Q^2$  at which the PDF contributes to each datapoint, determined using leading-order kinematics, whenever the kinematics of the leading-order process completely fixes them, and the minimum value of  $x$ , whenever it does not fix them completely (such as for single-inclusive jets). Note that, as already mentioned, the region of  $x$  smaller than the smallest  $x$  datapoints of Fig. 1 is completely inaccessible. The larger  $x$  region is, in principle, accessible at higher orders, though in practice this is hardly the case for sufficiently large  $x$  (say  $x \gtrsim 0.5$ ). Indeed, PDFs vanish at  $x = 1$  and drop quite fast as  $x$  increases (typically as a power of  $(1-x)$ ) so, for many processes, not only the contribution from larger  $x$  arises at higher perturbative orders, but also, it rapidly becomes much smaller than that from the smallest value of  $x$ , thus being in practice unobservable.

Inspection of Fig. 1 immediately shows that the PDF determination, and its validation, entails two rather distinct issues, related to the fact that data points are localized in a well-defined, connected kinematic region — the “data region”, henceforth. On the one hand, the problem of determining PDFs in the region probed by the data remains mathematically ill-posed, because one is determining a function from a piece of discrete information. Yet, because of the convolution integral, PDFs are not probed at a single point, but rather in a region, and also, even in the polarized case, where data are more scarce, datapoints are typically dense in the experimentally accessible kinematic region, because experimental collaborations strive to measure physical observables with as fine a binning as possible. Conse-

quently, whereas lack of knowledge of the PDF functional form entails that a population of equally likely best-fits to any given set of data is possible, assumptions of smoothness, which are physically justified, and built in standard machine-learning interpolants (such as neural networks) ensure that a reasonably narrow set of results follow.

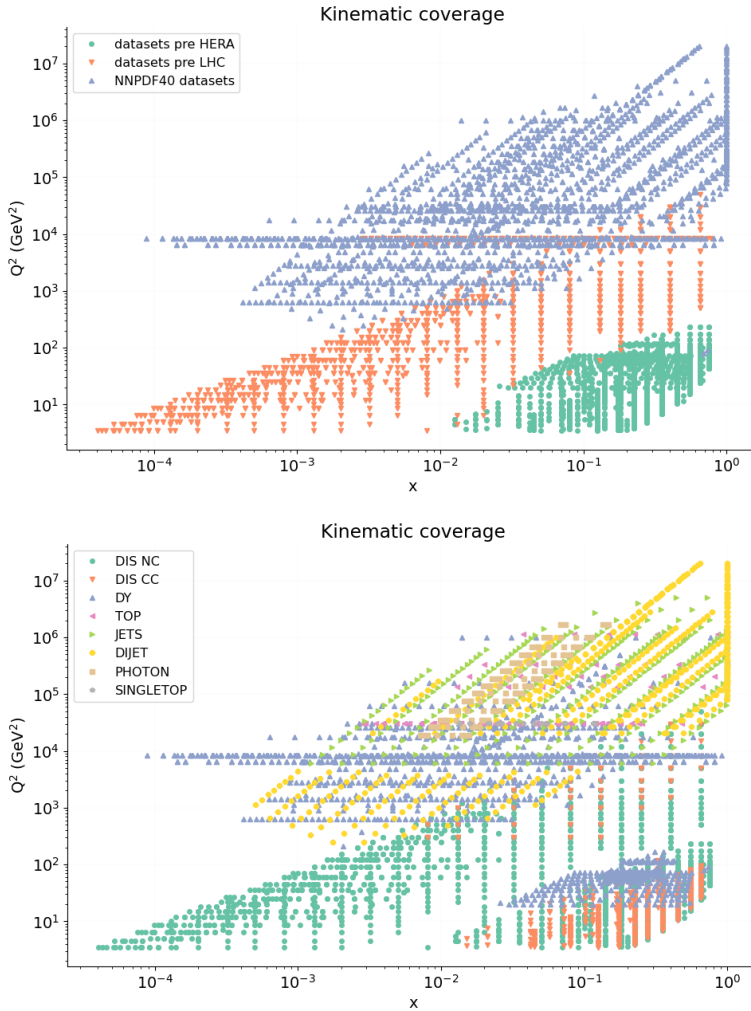


Fig. 1. (Color online) Scatter plot for the datapoints entering the NNPdf4.0 PDF determination and future tests in the unpolarized case, listed in Table I. The data are shown color-coded by future test grouping (top), and the full dataset is also shown color-coded by physics process (bottom).

TABLE I

The pre-HERA, pre-LHC and NNPDF4.0 datasets: for each experiment, the reference to the original publication and the number of datapoints are given. Pre-HERA PDFs are fitted to pre-HERA data, pre-LHC PDFs are fitted to the pre-HERA and pre-LHC data, and NNPDF4.0 PDFs are fitted to the union of the three datasets.

<b>pre-HERA</b>	Ref.	$N_{\text{dat}}$	<b>NNPDF4.0 (cont.'d)</b>	Ref.	$N_{\text{dat}}$
NMC $d/p$	[8]	121	ATLAS $W^- + \text{jet}$ 8 TeV	[28]	16
NMC $p$	[9]	204	ATLAS $Z$ $p_{\text{T}}$ 8 TeV ( $p_{\text{T}}^H, M_{ll}$ )	[29]	44
SLAC $p$	[10]	33	ATLAS $Z$ $p_{\text{T}}$ 8 TeV ( $p_{\text{T}}^H, y_{ll}$ )	[29]	48
SLAC $d$	[10]	34	ATLAS $\sigma_{t\bar{t}}^{\text{tot}}$ 7, 8, 13 TeV	[30, 31]	3
BCDMS $p$	[11]	333	ATLAS $t\bar{t}$ $y_t$ normalized 8 TeV	[32]	4
BCDMS $d$	[12]	248	ATLAS $t\bar{t}$ $y_{t\bar{t}}$ normalized 8 TeV	[32]	4
CHORUS $\sigma_{CC}^{\nu}$	[13]	416	ATLAS $t\bar{t}$ normalized $ y_t $ dilepton 8 TeV	[33]	5
CHORUS $\sigma_{CC}^p$	[13]	416	ATLAS jets 8 TeV, $R = 0.6$	[34]	171
NuTeV $\sigma_c^{\nu}$	[14]	39	ATLAS dijets 7 TeV, $R = 0.6$	[35]	90
NuTeV $\sigma_c^p$	[14]	37	ATLAS direct photon production 13 TeV	[36]	53
DYE 866 $\sigma_{\text{DY}}^d / \sigma_{\text{DY}}^p$	[15]	15	ATLAS single top $R_t$ 7 TeV	[37]	1
DY E886 $\sigma_{\text{DY}}^p$	[16]	89	ATLAS single top $R_t$ 13 TeV	[38]	1
DY E605 $\sigma_{\text{DY}}^p$	[17]	85	ATLAS single top $y_t$ norm. 7, 8 TeV	[37, 39]	6
Total pre-HERA		2070	ATLAS single antitop $y$ norm. 7, 8 TeV	[37, 39]	6
<b>pre-LHC</b>	Ref.	$N_{\text{dat}}$	CMS $W$ electron asymmetry 7 TeV	[40]	11
HERA I+II inclusive NC $e^- p$	[18]	159	CMS $W$ muon asymmetry 7 TeV	[41]	11
HERA I+II inclusive NC $e^+ p$ 460 GeV	[18]	204	CMS Drell-Yan 2D 7 TeV	[42]	110
HERA I+II inclusive NC $e^+ p$ 575 GeV	[18]	254	CMS $W$ rapidity 8 TeV	[43]	22
HERA I+II inclusive NC $e^+ p$ 820 GeV	[18]	70	CMS $Z$ $p_{\text{T}}$ ( $p_{\text{T}}^H, y_{ll}$ ) 8 TeV	[44]	28
HERA I+II inclusive NC $e^+ p$ 920 GeV	[18]	377	CMS dijets 7 TeV	[45]	54
HERA I+II inclusive CC $e^- p$	[18]	42	CMS 3D dijets 8 TeV	[46]	122
HERA I+II inclusive CC $e^+ p$	[18]	39	CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 7, 8, 13 TeV	[47, 48]	3
HERA comb. $\sigma_{c\bar{c}}^{\text{red}}$	[19]	37	CMS $t\bar{t}$ rapidity $y_{t\bar{t}}$ $\ell + \text{jet}$ 8 TeV	[49]	9
HERA comb. $\sigma_{b\bar{b}}^{\text{red}}$	[19]	26	CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 5 TeV	[50]	1
CDF $Z$ rapidity	[20]	28	CMS $t\bar{t}$ 2D $2\ell$ ( $m_{t\bar{t}}, y_t$ ) 8 TeV	[51]	16
D0 $Z$ rapidity	[21]	28	CMS $t\bar{t}$ absolute $y_t$ $\ell + \text{jet}$ 13 TeV	[52]	10
D0 $W \rightarrow \mu\nu$ asymmetry	[21]	9	CMS $t\bar{t}$ absolute $ y_t $ $2\ell$ 13 TeV	[53]	11
Total pre-LHC		1273	CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	[54]	1
<b>NNPDF4.0</b>	Ref.	$N_{\text{dat}}$	CMS single top $R_t$ 8 TeV	[55]	1
ATLAS $W, Z$ 7 TeV	[22]	30	CMS single top $R_t$ 13 TeV	[56]	1
ATLAS HM DY 7 TeV	[23]	5	LHCb $Z$ 940 pb	[57]	9
ATLAS low-mass DY 7 TeV	[24]	6	LHCb $Z \rightarrow ee$ 2 fb	[58]	17
ATLAS $W, Z$ 7 TeV central selection	[25]	46	LHCb $W, Z \rightarrow \mu$ 7 TeV	[59]	29
ATLAS $W, Z$ 7 TeV forward selection	[25]	15	LHCb $W, Z \rightarrow \mu$ 8 TeV	[60]	30
ATLAS DY 2D 8 TeV	[26]	48	LHCb $Z \rightarrow \mu\mu$ 13 TeV	[61]	16
ATLAS $W, Z$ inclusive 13 TeV	[27]	3	LHCb $Z \rightarrow ee$ 13 TeV	[61]	15
ATLAS $W^+ + \text{jet}$ 8 TeV	[28]	16	Total NNPDF4.0		1148
			Grand total		4491

On the other hand, in the small- and large- $x$  regions which are not probed by the data — the “extrapolation region”, henceforth — the uncertainty is *a priori* infinite, unless one introduces some assumptions allowing for generalization of the behavior observed in the region covered by the data. Infinite uncertainties would imply the impossibility of predicting any hadronic process, such as Higgs production. Indeed, the bulk of the PDF information which is necessary to predict a given LHC process can be estimated using suitable tools [7], but a certain amount of extrapolation is always needed. The only way to avoid this unpalatable conclusion is to introduce some assumption — motivated by physical considerations, or perhaps based on extrapolating, or generalizing a behavior seen in the data region.

## 2.2. Validation methods: closure tests and future tests

We can now discuss validation in the data and extrapolation region, and why they differ. The validation of PDF uncertainties in the data region has been addressed in Ref. [62] (see also Refs. [2, 63, 64] for an update): it can be performed using the closure test methodology. The idea is to assume a particular “true” underlying form of the PDFs, generate data based on this assumption, with some statistical distribution corresponding to a particular covariance matrix (typically, the same as that of the true experimental data), and then perform the PDF determination on this data. The faithfulness of the uncertainties can be checked by comparing the distribution of answers when the procedure is repeated several times with the true value, which is now known. Since both the experimental uncertainties and the theory are now completely correct by construction — they are used in the data generation — the procedure tests the accuracy and reliability of the methodology.

The validation of the methodology in the extrapolation region is an open issue: it is the problem that we would like to address here. There is a fundamental reason why this cannot be approached in the same way as the validation in the data region, so in particular, a closure test is not advisable. This is related to the fact that the current methodology, if correctly implemented, incorporates the constraints which are present (even if sometimes not explicitly known) in the current data. In order to understand this, consider a simple example. It is known that PDFs satisfy sum rules (see *e.g.* Ref. [3]): for example, the momentum sum rule

$$\sum_i \int_0^1 dx x f_i(x, Q^2) = 1, \quad (3)$$

where the sum runs over all parton species, which holds for all values of  $Q^2$ .

Equation (3) expresses the fact that the total momentum carried by all partons adds up to the momentum of their parent hadron. Since this is a known property of QCD, it is usually imposed as a hard-wired constraint; however, it can be checked that if it is not imposed, it is actually reproduced by the PDFs extracted from the data [65].

Now, the machine-learning methodology for PDF determination is actually optimized on existing data. This optimization was done by trial and error in the past, and it is currently performed, at least in part, using automatic hyperoptimization techniques [2, 66]. Clearly, this optimization is performed in the space of possible solutions, which is a subspace of all possible PDFs, incorporating both known constraints, such as the momentum sum rule Eq. (3), but also possible constraints that we do not know of.

Let us assume now for the sake of argument that we did not know the momentum sum rule. A methodology optimized on current data would work well on data that respect it. In order to closure-test the methodology, we have to assume an underlying functional form. Typically, we would assume that a true PDF is one PDF as determined now, and so even if we did not know the momentum sum rule, we would generate data that respect it to good approximation, and use them to closure-test a methodology that unknown to us incorporates this constraint. So the test would be meaningful.

Let us, however, assume now that we generated data *outside* the currently measured region. We would then have to make some assumption on the underlying PDF there, and this might not incorporate true constraints that hold there, or it might actually include some constraints that in actual fact do not really hold. In our example, if we did not know the momentum sum rule, we might generate pseudodata that violate it. A validation based on this procedure would then be unreliable because there is no way to know whether there is a mismatch between constraints in the generated data and the methodology. In other words, closure testing is only effective if the generated data carry an amount of information that is smaller, or at most comparable, to real data. If the generated data carry more information — as it must be the case if closure-testing the extrapolation region — there is no way to know whether a bias in the methodology will go undetected, or whether the test would force a bias upon the methodology itself.

The idea of the future test is to verify whether a methodology optimized on current data generalizes correctly the features contained in a subset of this data. So, for example, a future test of the methodology used for the NNPDF3.1 determination of PDFs [67] consists of using the same methodology to extract PDFs from, say, only the deep-inelastic scattering (DIS) subset of the NNPDF3.1 dataset. The future test then amounts to comparing these PDFs to the full NNPDF3.1 dataset. Indeed, DIS data cannot constrain the large- $x$  gluon PDF, which is instead well-determined by the



LHC data in the NNPDF3.1 dataset [68]. So, if one compares these data to PDFs determined from DIS only, one is testing whether the methodology manages to generalize the gluon PDF outside the data region.

The simplest way to perform the test is to view the computation of all data not included in the PDF fit as a prediction, and compare this prediction, with its PDF uncertainty, with the actual value. The future test is then successful if the  $\chi^2$  per data point is of the order of one, because this means that the PDF uncertainties have been estimated faithfully in the extrapolation region. More quantitatively, a criterion for the future test to be successful can be spelled out as

$$\left| \chi_{\text{not fitted}}^{2\text{PDF}} - \chi_{\text{fitted}}^2 \right| \ll \left| \chi_{\text{not fitted}}^{2\text{PDF}} - \chi_{\text{not fitted}}^2 \right|, \quad (4)$$

where we denote by  $\chi^{2\text{PDF}}$  and  $\chi^2$ , respectively, the  $\chi^2$  per datapoint with and without PDF uncertainties included. In other words, the test is successful if the inclusion of the PDF uncertainties accounts for the missing information which leads the predictions for data which are not fitted to deviate from the measured values.

The test can be thought of as the comparison of a prediction of future data with the future data themselves — hence the name. However, this is a manner of speaking: it is important to understand what the future test does and does not test. The future test does not test whether current data could have been predicted in the past, because the future test tests the *current* methodology. Indeed, the methodology that we are future-testing is optimized to the current dataset: our goal is to test the generalization power of our current best methodology. To this purpose, we test the capability of this methodology of extracting features of the full dataset from a subset, *i.e.* its generalization power. Also, the future test relies on all the information which is available on the full present-day dataset, both from an experimental and a theoretical point of view. Thus, for example, we now know that some data must be treated at the highest perturbative accuracy which is currently available, and these data are, therefore, excluded from PDF determinations which are now performed at a lower perturbative order — but higher-order corrections may have been unavailable when these data were originally published.

In this sense, the future test is weaker than a closure test. Indeed, a successful future test is necessary but not sufficient in order to guarantee that the extrapolation of current PDFs outside their data region is reliable. It could be that a reliable extrapolation requires information that is currently inaccessible: however, this is the best we can do now. This is akin to the situation that arises when estimating uncertainties related to missing higher-order terms in a perturbative computation. The best we can do is to base

the estimate on the known terms, but it cannot be excluded that some new piece of information is needed, which only arises at some higher perturbative order, in which case the estimate is unreliable. Even so, the future test poses stringent requirements on PDFs, as we will show in explicit examples in the next sections.

### 3. Unpolarized PDFs and the rise of structure functions

We start by presenting a future test of the NNPDF4.0 methodology. This is a methodology based, for the first time, on automatic hyperoptimization [66] (see [2] for a review). PDFs constructed with this methodology are currently in a testing phase and will be released soon [63]. This methodology is the state-of-the-art in PDF determination based on machine learning, and it is used for a PDF determination based on a dataset whose size exceeds any previous determination, specifically including a large number of LHC Run 2 data. Consequently, PDF uncertainties are smaller than in any previous PDF determination, and validation issues are particularly important. The validation of NNPDF4.0 PDFs in the data region, performed using an updated version [2] of the closure test methodology of Ref. [62], will be discussed in Ref. [63].

We choose to use the most recent NNPDF4.0 methodology as a first illustration of the future test because PDF uncertainties found with this methodology are relatively smaller, and thus the issue of accurately estimating PDF uncertainties in extrapolation is particularly serious. However, the methodology that we will discuss in Section 4 in the polarized case is essentially the same as the unpolarized NNPDF3.1 methodology: hence, this methodology will also be implicitly future-tested there. A direct comparison between future tests of subsequent NNPDF methodologies is an interesting topic that will be left for future studies [63]. A relevant observation, in this respect, is that all published NNPDF PDF sets turned out to be forward–backward compatible, in the sense that more recent PDFs have smaller uncertainties, but are compatible within uncertainties with previously published ones.

#### 3.1. Gluons and HERA, quarks and the LHC

We performed future tests of the NNPDF4.0 methodology based on two datasets of increasing size: a pre-HERA dataset, which only includes fixed-target DIS and Drell–Yan production data; and a dataset obtained by combining this with a pre-LHC dataset, which also includes HERA DIS data and Tevatron collider  $W$ ,  $Z$  production and single-inclusive jet data. The NNPDF4.0 PDFs are determined by adding to these also an NNPDF4.0 dataset, which includes LHC data for  $W$  and  $Z$  production, single-inclusive jets and dijets,  $Z$  transverse momentum distributions, single top and top

pair production, and prompt photon production. Note that PDFs are thus determined from strictly hierarchical datasets: pre-HERA PDFs are determined from the pre-HERA data, pre-LHC PDFs are determined combining pre-HERA and pre-LHC data, and NNPDF4.0 PDFs are determined combining pre-HERA, pre-LHC, and NNPDF4.0 data. Although, as explained in Section 2.2, the historical nature of these datasets is somewhat besides the point, since the future test is a test of current PDF methodology, it might be useful to think about these datasets roughly as those on which PDF determinations were respectively based circa 1993, such as the CTEQ1 [69] or MRS [70] PDFs, for the pre-HERA PDFs, and circa 2010, such as NNPDF2.1 [71] for the pre-LHC PDFs.

The three datasets are listed in Table I and displayed in Fig. 1, where the breakdown of the full dataset into single processes is also shown. The choice of these two future test datasets is motivated by aspects having to do both with physics and methodology. In terms of physics, the pre-LHC dataset only imposes very weak constraints on the flavor separation of quarks and antiquarks, and also on the large- $x$  gluon. In the pre-LHC dataset, the small- $x$  gluon is also unconstrained so, in fact, the gluon is largely unconstrained, except in a very small region around  $x \sim 0.1$ . In terms of methodology, a pre-LHC fit tests whether relatively small PDF uncertainties (of the order of say 5–10%) are compatible with later more precise data: *i.e.*, it tests the ability of the methodology to perform a near extrapolation. On the contrary, the pre-HERA fit tests whether the methodology can capture broad features of the results even in the presence of a minimal amount of information: *i.e.* it tests the far extrapolation.

A particularly intriguing aspect of the pre-HERA future test in this respect is the small- $x$  behavior of the gluon PDF and, correspondingly, of the  $F_2$  proton structure function. Indeed, when this structure function was first measured at HERA, the observed rise of the structure function at small  $x$  (see *e.g.* [72]) came as a surprise, and indeed standard pre-HERA PDF sets displayed a wide variety of small- $x$  gluon behaviors, with a flat gluon taken as a baseline option in the absence of New Physics effects [70]. A rather steeper rise of the gluon at small  $x$  is a common feature of all post-HERA PDF sets (see [73] for a comparative discussion). It is thus interesting to wonder how a contemporary methodology, presented with pre-HERA data, would behave in this respect.

### 3.2. Future testing NNPDF4.0

We now turn to the results of the future test of the NNPDF4.0 methodology. The reader is referred to the forthcoming Ref. [63] for full details on the NNPDF4.0 PDF determination theory and methodology; here, we pro-

vide some basic information. The machine-learning methodology which is being future-tested has been constructed through a hyperoptimization procedure [66] and its general features are reviewed in Ref. [2]. It uses the same approach, based on a Monte Carlo representation of PDF uncertainties, and neural networks as underlying interpolants as previous NNPDF determinations, such as the most recent published NNPDF3.1 [67]. It differs from the previous NNPDF methodology in the architecture of neural networks and minimization algorithm and, more importantly, due to the fact that architecture and algorithms have been selected through a hyperoptimization procedure tuned using  $K$ -foldings, see [2]. Preprocessing is treated according to the most recent NNPDF methodology [74].

The figure of merit which is being minimized is the same as in previous NNPDF determinations and, specifically, it includes all available information on statistical and systematic experimental uncertainties and their correlations through the  $t_0$  method [75, 76]. The NNLO QCD theory is used throughout, with charm quark mass effects included through the FONLL method [77] and a parametrized charm PDF [78, 79]. Uncertainties due to deuterium and heavier nuclear targets are included [80–82] as a theory uncertainty contribution to the covariance matrix [83].

An identical methodology is used to perform the future tests. The NNPDF methodology calls for the  $t_0$  covariance matrix [75] and the preprocessing exponents to be determined iteratively and self-consistently: this has been done for each of the two future tests. Results are collected in Table II, where we show  $\chi^2$  values for the three PDF sets: namely, the baseline NNPDF4.0, the pre-HERA and the pre-LHC future tests. Values are shown for all data, both included and not included in each fit. We show both standard  $\chi^2$  values computed using the experimental covariance matrix, and values computed also including PDF uncertainties. Note that the experimental covariance matrix is not the same as the  $t_0$  covariance matrix used for minimization in order to obtain unbiased results, but just the standard covariance matrix as published by experimental collaborations (see Ref. [76] for a comparative discussion). The  $\chi^2$  computed using it measures the goodness of fit provided by the PDFs fitted to data.

In the  $\chi^2$  including PDF uncertainties, these are accounted for by adding to the experimental covariance matrix the PDF uncertainty covariance matrix, in turn, computed as the covariance over the NNPDF replica set [84]. The total covariance matrix is then inverted and used in the standard definition of the  $\chi^2$ . Addition of two contributions to the covariance matrix is justified if they correspond to uncorrelated uncertainties [83]. This is surely the case when considering the PDF uncertainty and the uncertainty on new data not used for the determination of those PDF. Consequently, the  $\chi^2$  values shown in Table II with PDF uncertainties included, for the

future test PDFs when compared to data not used to fit, indicate whether the PDF uncertainty estimate on these future test PDFs is reliable: it tests the generalization power of the methodology.

TABLE II

The  $\chi^2$  per datapoint for the NNPDF4.0 PDF determination and its pre-HERA and pre-LHC future tests. Values are shown for the three datasets displayed in Fig. 1. In the left table, all values are computed using only the experimental covariance matrix, while in the right table, all values are computed by also including PDF uncertainties. Values with PDF uncertainties included computed for data included in the fit, shown in parenthesis, do not have a strict statistical meaning (see the text). All numbers in italic (without PDF uncertainty included) and in boldface (with PDF uncertainty included) are predictions.

	$n_{\text{dat}}$	NNPDF4.0	pre-LHC	pre-HERA	NNPDF4.0	pre-HERA	pre-LHC
pre-HERA dataset	2070	1.09	1.01	0.90	(1.02)	(0.95)	(0.86)
pre-LHC dataset	1273	1.21	1.20	<i>23.1</i>	(1.18)	(1.17)	<b>1.22</b>
NNPDF4.0 dataset	1148	1.29	<i>3.30</i>	<i>23.1</i>	(1.23)	<b>1.30</b>	<b>1.38</b>
Total dataset	4491	1.17	1.65	12.9	1.12	1.10	1.10

Note that, of course, the PDF uncertainty is not uncorrelated to the uncertainty on data used for the determination of the same PDFs, since the latter propagates into the former. Consequently,  $\chi^2$  values computed including PDF uncertainties for datasets included in any PDF determination, shown in parenthesis in Table II, are only indicative and do not have a strict statistical meaning.

It is important to stress that both the experimental covariance matrix and the PDF covariance matrix consistently include correlations. In fact, these are often quite large, and  $\chi^2$  values found not including correlations would be substantially different. For the experimental covariance matrix, this is because for many LHC data uncorrelated statistical uncertainties are very small: sometimes, for example, for  $Z$ -production data, at the per mille level. Dominant uncertainties are then systematic, and these are typically highly or fully correlated: an example is the luminosity uncertainty, which is common to all data in a given experiment. For the PDF covariance matrix, this is due to the fact that many data points are kinematically very close to each other, and of course PDF values and uncertainties on neighboring values of  $x$  and  $Q^2$  are highly correlated.

The relevant information from Table II is contained in the  $\chi^2$  values corresponding to predictions, *i.e.* computed from data not included in the determination of the corresponding PDF sets. These are shown in italic

(without PDF uncertainties) and boldface (with PDF uncertainties). These values show that the future test is impressively successful, and the criterion Eq. (4) is clearly satisfied. Before inclusion of PDF uncertainties (values in *italic*), the  $\chi^2$  of future test PDFs is of the order of 20, but after their inclusion (values in **boldface**), it becomes of the order of one and, in fact, almost identical to that of fitted PDFs (without uncertainties included). This means that PDF uncertainties increase on average by more than two orders of magnitude in the extrapolation region and that this increase is exactly of the size required in order to account for the missing information contained in the excluded data. This is especially remarkable for the  $\chi^2$  of the pre-HERA PDFs when compared to the post-HERA dataset, whose value is hardly larger than that found in the NNPDF4.0 fit itself. In fact, inspection of  $\chi^2$  values for individual experiments [63] shows that the  $\chi^2$  values after inclusion of PDF uncertainties are all very close to the values found in the NNPDF4.0 fit (without PDF uncertainties), while, especially for the pre-HERA PDFs, some of the values found without PDF uncertainties, especially for experiments in far extrapolation regions, can be up to two orders of magnitude higher, as we will discuss below for some specific examples. This means that uncertainties are quite reliable even in far extrapolation regions.

Note that the  $\chi^2$  values for the data included in the fit (which, as mentioned, do not have a statistical meaning strictly) are almost unchanged when PDF uncertainties are included. This is due to the fact that, for data included in the fit, the PDF uncertainty is generally rather smaller than that of data used for its determination, because PDFs combine the information coming from many datapoints. Hence for these data, the experimental uncertainty is dominant, and this explains why  $\chi^2$  values are almost unaffected by the inclusion of PDF uncertainties.

A more detailed picture can be obtained by looking at individual PDFs. Specifically, in Fig. 2 the up, antidown, strange quark and gluon PDFs found in the two future test fits are compared to NNPDF4.0. It is clear that PDFs are generally compatible within errors, with differences exceeding the one sigma level in a limited set of cases (as expected given that one sigma corresponds to a 68% confidence level). This is true both for PDFs determined relatively accurately (such as the up quark in the pre-LHC fit), or those affected by large uncertainties (such as the gluon in the pre-HERA fit).

The case of the pre-HERA gluon is especially remarkable: despite the very large uncertainty, the future test correctly extrapolates the small- $x$  rising trend. In fact, with hindsight, this trend can be seen in some pre-HERA DIS data (especially from NMC), though historically it was overlooked. Of course, the methodology that we are future-testing has been hyperoptimized on a dataset that does contain small- $x$  data, so the pre-HERA future test cannot be taken as evidence that the rise of the structure function could

have been predicted without the HERA data. However, it does suggest that an unbiased inspection of this data should have at least suggested this possibility.

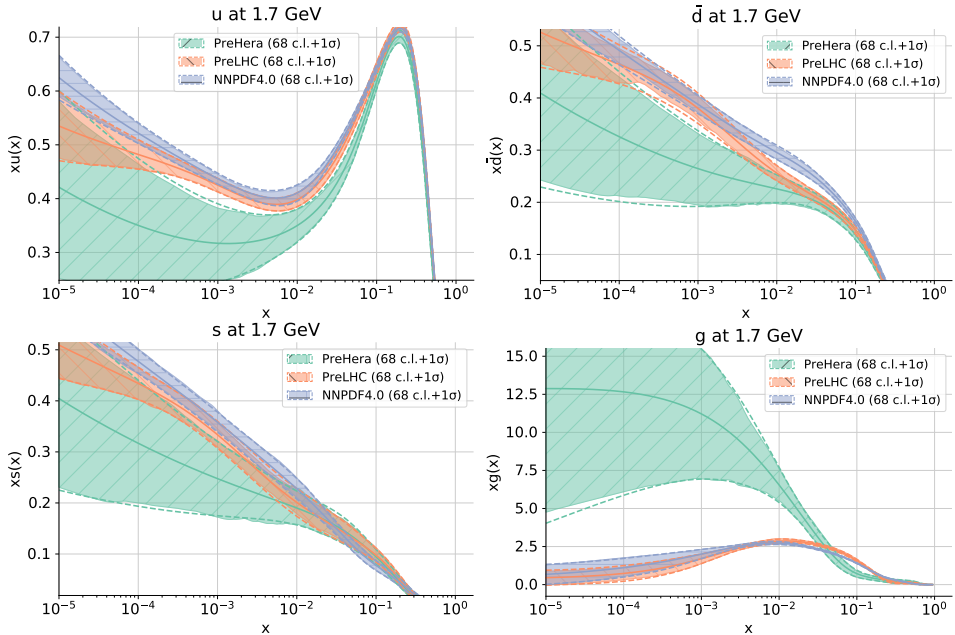


Fig. 2. Comparison of the pre-HERA, pre-LHC and NNPDF4.0 PDFs. The up (top left), antidown (top right), strange (bottom left) and gluon (bottom right) PDFs are shown at the parametrization scale  $Q = 1.7$  GeV.

In Fig. 3, we also show the comparison of predictions obtained using either NNPDF4.0, or the two future test PDF sets, to data for some physical observables: the HERA  $F_2$  structure function [18], the CMS double-differential distribution of top pairs [51], the LHCb  $W/Z$  rapidity distribution [85] and the ATLAS  $Z$   $p_T$  distribution [29]. These observables are sensitive to a wide array of PDF features which are poorly constrained by pre-HERA data. Specifically, the HERA structure function probes the small- $x$  gluon, the top pair and  $Z$   $p_T$  distributions depend on the large- $x$  gluon and the LHCb large-rapidity  $W$  and  $Z$  cross sections probe the flavor separation at large  $x$ . In each case, we clearly see that the PDF uncertainty on the prediction using future test PDF correctly accounts for the deviation from the data, despite the very substantial amount of extrapolation which is required in all cases shown. This is manifestly seen when inspecting the corresponding  $\chi^2$  values. For instance, the  $\chi^2$  values per datapoint for the pre-HERA fit are, for the HERA I+II inclusive NC  $e^+p$  920 GeV

data,  $\chi^2 = 44.2$  without PDF uncertainties and  $\chi^{2\text{PDF}} = 1.29$  when PDF uncertainties are included (fitted NNPDF4.0 value  $\chi^2 = 1.31$ ); for CMS  $t\bar{t}$  2D  $2\ell$  ( $m_{t\bar{t}}, y_t$ ) 8 TeV, they are respectively  $\chi^2 = 5.125$  and  $\chi^{2\text{PDF}} = 1.15$  (fitted NNPDF4.0 value  $\chi^2 = 0.89$ ); for LHCb  $W, Z \rightarrow \mu$  7 TeV,  $\chi^2 = 15.37$  and  $\chi^{2\text{PDF}} = 1.71$  (fitted NNPDF4.0 value  $\chi^2 = 1.96$ ); and for ATLAS  $Z$   $p_T$  8 TeV ( $p_T^l, y_l$ ),  $\chi^2 = 28.69$  and  $\chi^{2\text{PDF}} = 0.997$  (fitted NNPDF4.0 value  $\chi^2 = 0.89$ ). In all cases, the inclusion of the PDF uncertainties brings down the  $\chi^2$  to a value that is close to that found when the data are fitted, and the criterion Eq. (4) is very well satisfied, as the figures show graphically.

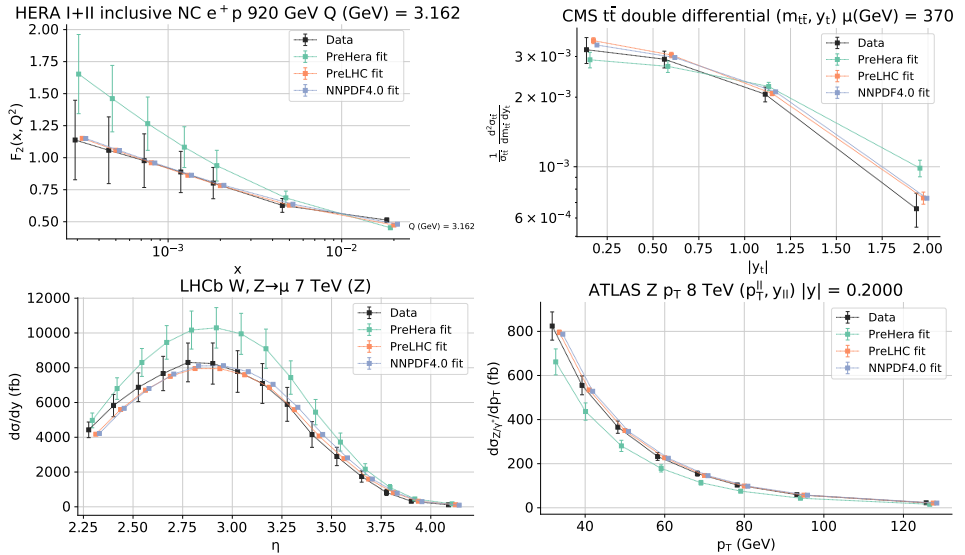


Fig. 3. Comparison to experimental data of predictions obtained using the pre-HERA, pre-LHC and NNPDF4.0 PDF sets. The diagonal PDF uncertainties and data uncertainties are shown. Illustrative results are displayed for the HERA  $F_2$  proton structure function [18] (top left), the CMS double differential top pair distribution [51] (top right), the LHCb  $Z$  rapidity distribution [85] (bottom left), and the ATLAS  $Z$   $p_T$  distribution [29] (bottom right).

#### 4. Polarized PDFs and the “spin crisis”

We now turn to longitudinally polarized parton distributions (polarized PDFs, henceforth), defined as

$$\Delta f_i(x, Q^2) = f_i^{\uparrow\uparrow}(x, Q^2) - f_i^{\downarrow\downarrow}(x, Q^2), \quad (5)$$



where  $f_i^{\uparrow\uparrow}$  ( $f_i^{\uparrow\downarrow}$ ) is the  $i^{\text{th}}$  species of PDF when the parton's spin is parallel (antiparallel) to that of its parent hadron, so that unpolarized PDFs are  $f_i(x, Q^2) = f_i^{\uparrow\uparrow}(x, Q^2) + f_i^{\uparrow\downarrow}(x, Q^2)$ .

Polarized PDFs are based on rather more scarce experimental information than their unpolarized counterparts. Not only a much smaller number of data and processes is available, but also these are typically affected by significantly larger uncertainties. Furthermore, fully correlated systematics are often not available, so statistical and systematic uncertainties must be added in quadrature, thereby leading to even larger uncertainties. Extrapolation problems are accordingly more serious for polarized PDFs.

#### 4.1. Polarized PDFs and data

Until quite recently, polarized PDFs were determined purely using neutral-current deep-inelastic scattering on proton, deuteron and neutron (*i.e.* helium) targets from the polarized structure function  $g_1$  directly, or the polarized asymmetries from which the structure function is extracted (see [86] for a review). More recently, these were supplemented by open charm and semi-inclusive hadron production and, more importantly, by precious little data on  $W$ , inclusive jet and pion production from RHIC, see [5] and references therein. The dataset for the most recent determination of polarized PDFs based on the NNPDF methodology, NNPDFpol1.1 [87], is shown in Fig. 4. We have performed a three-fold future test of this PDF set and methodology by comparing it to pre-EMC and pre-RHIC datasets, and also to a post-NNPDF1.1 dataset. This dataset includes some more recent data, which appeared after the original NNPDFpol1.1 PDF determination. Some of these data were discussed, in the context of NNPDF fits, in Refs. [88, 89] without leading to a fully updated PDF release. The datasets are all shown in Fig. 4, and also listed in Table III. Since the NNPDFpol1.1 methodology is essentially the same as the NNPDF3.1 methodology, the future tests presented here also effectively test the NNPDF3.1 methodology.

A PDF determination based purely on neutral-current DIS cannot disentangle quark and antiquark PDFs, and only allows for a determination of their sum

$$\Delta q_i^+(x, Q^2) = \Delta q_i(x, Q^2) + \bar{q}_i(x, Q^2) \quad (6)$$

for each flavor  $i$ . Furthermore, DIS on proton targets determines only a fixed linear combination of quark flavors, *i.e.*, in practice, a fixed linear combination of up, down and strange contributions since the contribution of heavy quarks is negligible, at least within present-day accuracy. If also deuteron or neutron targets are available, it is then possible to determine separately the triplet combination

$$\Delta T_3(x, Q^2) = \Delta u^+(x, Q^2) - \Delta d^+(x, Q^2) . \quad (7)$$

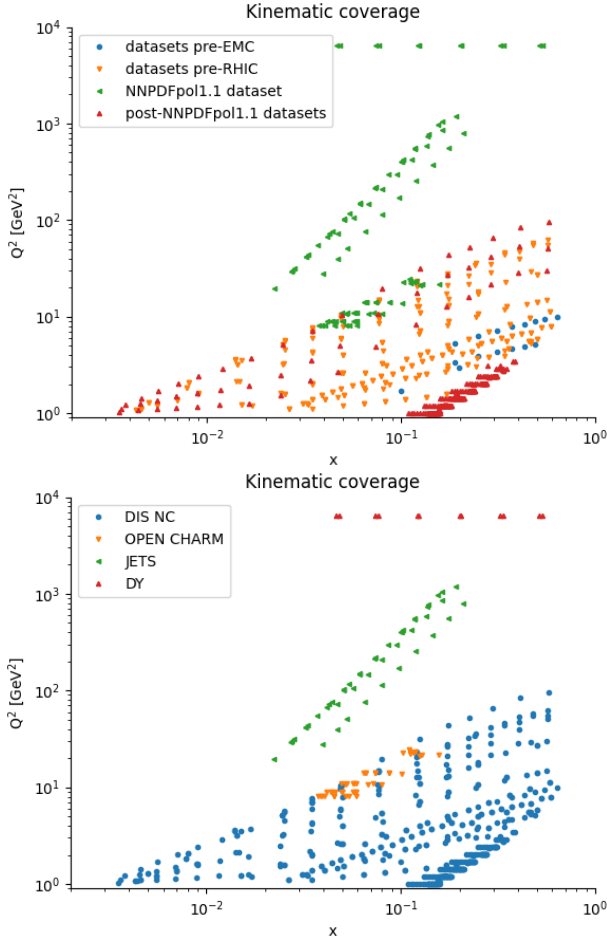


Fig. 4. The same as Fig. 1 but now for polarized PDFs.

However, it is not *a priori* possible to disentangle the two linear combinations orthogonal to it, namely the singlet

$$\Delta\Sigma(x, Q^2) = \Delta u^+(x, Q^2) + \Delta d^+(x, Q^2) + \Delta s^+(x, Q^2) \quad (8)$$

and octet

$$\Delta T_8(x, Q^2) = \Delta u^+(x, Q^2) + \Delta d^+(x, Q^2) - 2\Delta s^+(x, Q^2). \quad (9)$$

However, the first moments (*i.e.* integrals) of triplet and octet can be extracted from weak decays. If only DIS on proton targets is used, this is then the only information available on flavor separation.

TABLE III

The pre-EMC, pre-RHIC, NNPDFpol1.1 and post-NNPDFpol1.1 datasets: for each experiment the reference to the original publication and the number of datapoints is given. Pre-EMC polarized PDFs are fitted to pre-NMC, pre-RHIC PDFs are fitted to the pre-EMC and pre-RHIC data, and NNPDFpol1.1 PDFs are fitted to the union of these three datasets. Post-NNPDFpol1.1 PDFs are only used for future testing.

<b>pre-EMC</b>	Ref.	$N_{\text{dat}}$	<b>NNPDFpol1.1</b>	Ref.	$N_{\text{dat}}$
SLAC (1978) $A_1^p$	[90]	4	COMPASS open charm $A_{LL}$	[91]	45
SLAC (1983) $A_1^p$	[92]	14	STAR 1-jet (2005) $A_{LL}$	[93]	10
Total pre-EMC		18	STAR 1-jet (2006) $A_{LL}$	[93]	9
<b>pre-RHIC</b>	Ref.	$N_{\text{dat}}$	STAR 1-jet (2009) $A_{LL}$	[94]	22
EMC $A_1^p$	[95]	10	PHENIX 1-jet $A_{LL}$	[96]	6
SMC $A_1^p$	[97]	12	Total NNPDFpol1.1		92
SMC $A_1^d$	[97]	12	<b>post-NNPDFpol1.1</b>	Ref.	$N_{\text{dat}}$
SMC (low- $x$ ) $A_1^p$	[98]	8	COMPASS (2015) $A_1^p$	[99]	51
SMC (low- $x$ ) $A_1^d$	[98]	8	COMPASS (2016) $A_1^d$	[100]	43
E143 $A_1^p$	[101]	25	JLAB-EG1-DVCS $g_1^p$	[102]	9
E143 $A_1^d$	[101]	25	JLAB-EG1-DVCS $g_1^d$	[102]	9
E154 $A_1^n$	[103]	11	JLAB-E93-009 $A_1^p$	[104]	62
E155 $g_1^p$	[105]	20	JLAB-E93-009 $A_1^d$	[104]	86
E155 $g_1^d$	[105]	20	JLAB-E06-014 $A_1^n$	[106]	2
COMPASS (2007) $A_1^d$	[107]	15	STAR 2-jet (2009) $A_{LL}$	[108]	14
COMPASS (2010) $A_1^p$	[109]	15	STAR 1-jet (2012) $A_{LL}$	[110]	14
HERMES (1997) $A_1^n$	[111]	8	STAR 2-jet (2012) $A_{LL}$	[110]	42
HERMES $A_1^p$	[112]	28	Total postNNPDFpol1.1		332
HERMES $A_1^d$	[112]	28	Grand Total		687
Total pre-RHIC		245			

The pre-EMC dataset, which only includes proton DIS data, only allows for a determination of quark and gluon (the latter from the  $Q^2$  dependence) with minimal information on flavor separation from first moments. The pre-RHIC dataset, including proton, deuteron and neutron DIS data allows for a more detailed flavor separation, and a reasonable determination of the triplet. The NNPDFpol1.1 allows for a determination of individual flavor and antiflavours (from  $W$  production) with jet production providing a direct handle on the gluon. These datasets are therefore hierarchical in terms of the flavor separation they allow: pre-EMC and pre-RHIC correspond respectively to far and near extrapolation in flavor space.

As it is apparent from Fig. 4, these datasets are also hierarchical in terms of coverage in  $(x, Q^2)$  space. More coverage in  $Q^2$  allows for a better extraction of the gluon, which is determined mainly by the scale dependence, while more coverage in  $x$  allows for better determination of the first moment of individual PDFs and, in particular, the first moment of the singlet combination. Unlike the triplet and octet, the singlet is not determined by weak decays and has a simple physical interpretation as the fraction of the parent hadron’s spin carried by quarks (up to field theoretical complications related to the axial anomaly [113]). So also in terms of  $x$ -coverage, gluon determination, and determination of the first moments, the pre-EMC and pre-RHIC datasets correspond to far and near extrapolation.

In fact, this is the reason why specifically a pre-EMC dataset was chosen for future testing. Indeed, when the EMC data [95] were originally published, a first determination of the singlet first moment was possible. The result turned out to be surprising: the first moment was found to be rather smaller than expected and, in fact, compatible with zero within uncertainties — which would mean (at least naively) that the proton spin is not carried by quarks. This has been often referred to as the “proton spin crisis” [114]. Hence, just like in the case of the rise of the structure function at HERA, it is natural to wonder what a contemporary polarized methodology would make of pre-EMC data.

Finally, we have also performed a future test with post-NNPDFpol1.1 data because these significantly extend the precision and also somewhat enlarge the kinematic range of the NNPDFpol1.1 dataset: in a sense, these are the first polarized data that provide a first step towards the direction of precision polarized PDFs. However, we have future-tested them using the NNPDFpol1.1 methodology because this is the most recent polarized methodology that we have. In this sense, this future test confirms (or disproves) that the NNPDFpol1.1 methodology is still adequate, even in view of this more recent data.

#### 4.2. Future-testing polarized PDFs

The polarized future test tests the NNPDFpol1.1 methodology [87]. This is quite similar to the methodology used for unpolarized NNPDF sets until the most recent published NNPDF3.1 [67]. It differs from the methodology whose future test was discussed in Section 3.2, essentially because it is not obtained from an automatic hyperoptimization procedure, so methodological details, such as the choice of neural network architecture, are based on previous experience rather than being optimized automatically. A peculiar aspect of the polarized fitting methodology is the need to assume an underlying set of unpolarized PDFs, since experimental data are usually presented

for polarized asymmetries. Polarized PDF determinations presented in this section assume the published NNPDF3.1 as an underlying unpolarized set.

As in the unpolarized case, we have computed  $\chi^2$  values for all datasets and PDF sets, both with and without PDF uncertainties included. Results are collected in Table IV. An important caveat in the polarized case is that correlated systematic uncertainties are mostly not available, except for the most recent experiments. Hence, statistical and systematic uncertainties have to be added in quadrature, and the covariance matrix must be taken to be diagonal. We will come back to this point shortly. PDF uncertainties, on the other hand, are always included with all correlations fully accounted for.

TABLE IV

The  $\chi^2$  per datapoint for the NNPDFpol1.1 PDF determination and its pre-EMC and pre-RHIC future tests. Values are shown for the four datasets displayed in Fig. 4. In the left table, all values are computed using only the experimental covariance matrix, while in the right table, all values are computed by also including PDF uncertainties. Values with PDF uncertainties included computed for data included in the fit, shown in parenthesis, do not have a strict statistical meaning (see the text). All numbers in italic (without PDF uncertainty included) and in boldface (with PDF uncertainty included) are predictions.

	$n_{\text{dat}}$	NNPDFpol1.1 pre-RHIC pre-EMC			NNPDFpol1.1 pre-RHIC pre-EMC		
pre-EMC dataset	18	0.53	0.53	1.09	(0.53)	(0.52)	(0.77)
pre-RHIC dataset	245	0.75	0.77	<i>20.4</i>	(0.64)	(0.67)	<b>0.51</b>
NNPDFpol1.1 dataset	92	1.37	<i>1.66</i>	<i>7.36</i>	(1.35)	<b>1.38</b>	<b>1.40</b>
post-NNPDFpol1.1 dataset	332	<i>2.58</i>	<i>2.54</i>	<i>8.46</i>	<b>0.97</b>	<b>0.96</b>	<b>0.90</b>
Total dataset	687	1.54	1.58	13.2	(0.88)	(0.89)	(0.81)

As in the unpolarized case, the relevant comparison is between  $\chi^2$  values for data not included in the fit, before (in italic) and after (in boldface) the inclusion of PDF uncertainties. As in the unpolarized case, we find that all  $\chi^2$  values with PDF uncertainties included are of the order of one, thus showing that the future test is successful. The pre-EMC case is especially remarkable, with a  $\chi^2$  value that, for data not used for fitting, is of order thirteen per datapoint, and decreases to a value of the order of one once the PDF uncertainty is included. This shows that with only pre-EMC data fitted, uncertainties increase by one and a half order of magnitude, and the increase is exactly the right size to account for the missing information, *i.e.* the deviation of results from the true value.

When comparing NNPDFpol1.1 data to the pre-RHIC PDFs, and post-NNPDFpol1.1 data to NNPDF1.1pol PDFs,  $\chi^2$  values are of the order of three–four per datapoint, which means that the increase of uncertainties in the region of the data not included is not so dramatic, though still noticeable. The data not included only provide a relatively minor extension of kinematic reach, without adding any new process or target. Moreover, in these cases, the  $\chi^2$  value after inclusion of PDF uncertainties decreases to a value around one, thus showing that uncertainties are correctly estimated also in near extrapolation.

Inspection of the results of Table IV shows that  $\chi^2$  values without PDF uncertainties, for datasets which are fitted, is in most cases smaller, and often rather smaller than one. This is due to the fact that, as already mentioned, statistical and systematic experimental uncertainties are added in quadrature, because of a lack of information on their correlation. This leads to an overestimate of the experimental uncertainties, and thus to an underestimate of the  $\chi^2$ . Moreover, as mentioned, correlations between PDF uncertainties instead are always consistently included, so when a dataset is not fitted and PDF uncertainties are dominant, the expected correct result has a  $\chi^2$  of order one per datapoint, even if the  $\chi^2$  of the same dataset when fitted is much smaller than one. The results shown in Table IV confirm this expectation.

We now turn to a comparison of PDFs. Since, as mentioned, only the three combinations, Eqs. (7)–(9), of PDFs are accessible, it is more convenient to look directly at these, and the gluon. They are compared in figure 5 for the pre-EMC, pre-RHIC PDF and NNPDFpol1.1 sets. The pre-EMC PDFs are perfectly compatible with NNPDFpol1.1, within their extremely large uncertainties. The pre-RHIC and NNPDFpol1.1 polarized quarks are almost identical, while the pre-RHIC polarized gluon is somewhat more uncertain than the NNPDFpol1.1 gluon, especially at large  $x \gtrsim 0.1$ , but in very good agreement with it.

As discussed in Section 4.1, much of the interest in polarized PDFs is related to the EMC discovery that the first moment of the polarized quark distribution is unusually small and, in fact, compatible with zero within uncertainties. This also raised interest in the value of the first moment of the polarized gluon distribution, both because of its possible role in carrying a sizable fraction of the proton spin, and also, in explaining the smallness of the quark first moment at low scale due to its mixing with it driven by the axial anomaly [113]. Note that, as mentioned, the first moment of the triplet and octet combinations, Eqs. (7)–(9), are instead fixed by weak meson decay constants.

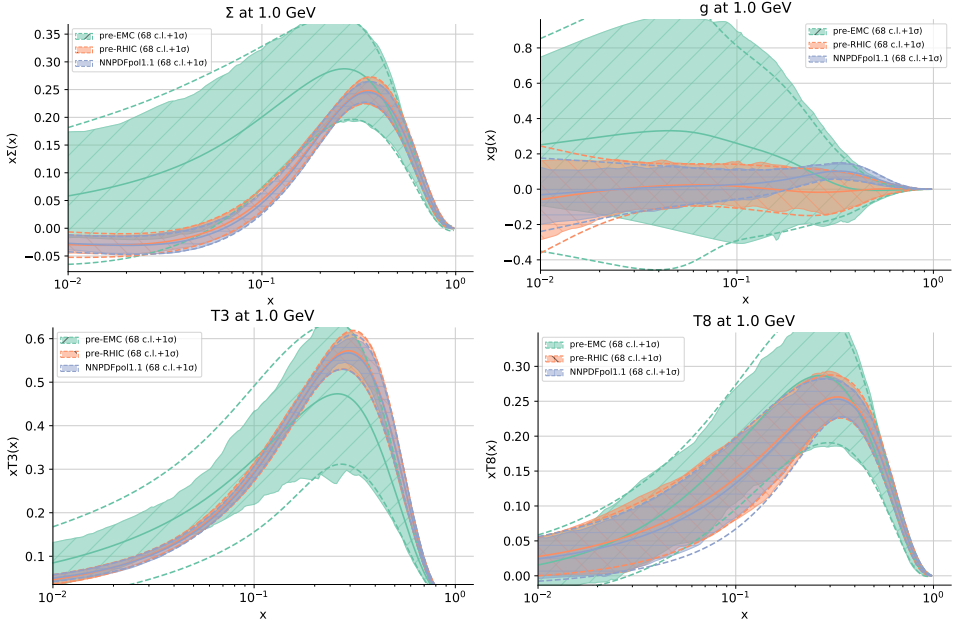


Fig. 5. Comparison of the pre-EMC, pre-RHIC and NNPDFpol1.1 PDFs. The singlet (top left), gluon (top right), triplet (bottom left) and octet (bottom right) PDF combinations (see the text) are shown at the parametrization scale  $Q = 1$  GeV.

The values of the polarized quark singlet and gluon first moments are collected in Table V for the NNPDFpol1.1, pre-EMC and pre-RHIC polarized PDF sets. Of course, the qualitative behavior of the first moments is the same as that of the PDFs seen in figure 5: they are all compatible within uncertainties, with pre-EMC uncertainties much larger, and pre-RHIC uncertainties somewhat larger for the quark and rather larger for the gluon in comparison to NNPDFpol1.1. The “spin crisis”, with the polarized singlet quark first moment compatible with zero, and significantly different from a value of the order of one, is clearly seen already in the pre-RHIC PDF set. On the other hand, no conclusion on the polarized quark first moment can be drawn from pre-EMC data: the result is compatible both with zero and with one, within its large uncertainty. Even with contemporary fitting methodology, no hint of the “spin crisis” can be seen in pre-EMC data. The spin crisis was a real experimental surprise. Interestingly, knowledge of the gluon first moment has not improved much over the years: both a vanishing value or a large value (a possible explanation of the “spin crisis”) are compatible with both EMC and contemporary data.

TABLE V

The first moment of the polarized quark singlet and gluon distributions, computed using the NNPDFpol1.1 PDF sets and the two future test pre-EMC and pre-RHIC PDF sets, evaluated at the scale  $Q^2 = 1 \text{ GeV}^2$ .

	NNPDFpol1.1	pre-EMC	pre-RHIC
$\Delta\Sigma$	$0.18 \pm 0.21$	$0.83 \pm 0.74$	$0.16 \pm 0.30$
$\Delta g$	$0.02 \pm 3.24$	$1.71 \pm 4.80$	$0.95 \pm 3.87$

Finally, in figure 6, we compare predictions for selected pre-EMC and pre-RHIC data, and in figure 7 for selected post-NNPDFpol1.1 data, obtained using the three polarized PDF sets. The excellent compatibility of the three

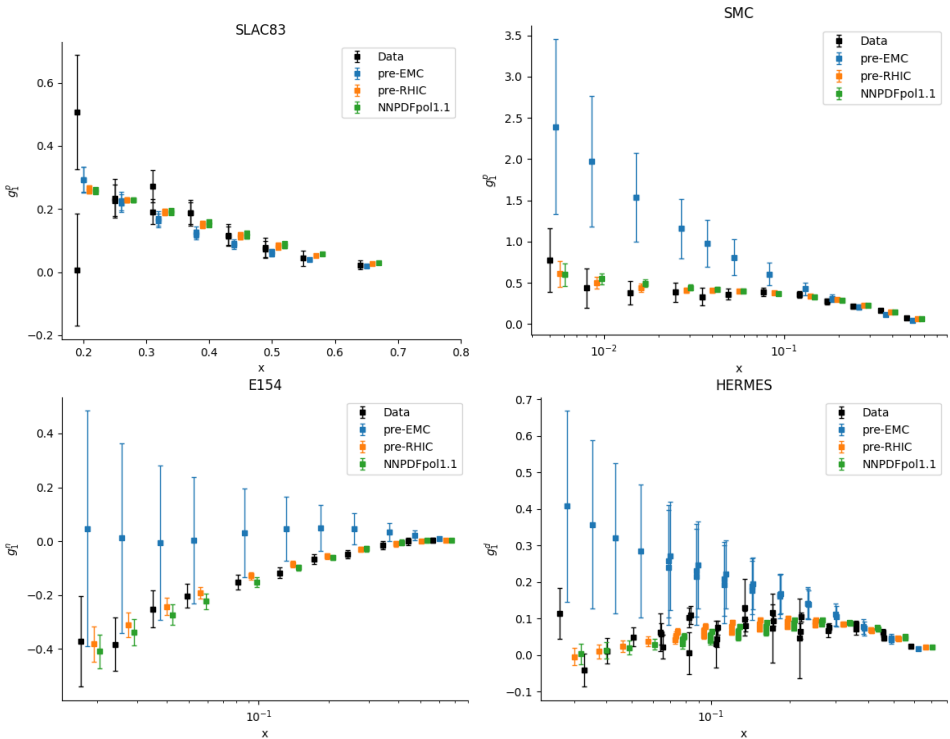


Fig. 6. Comparison to experimental data of predictions obtained using the pre-EMC, pre-RHIC and NNPDFpol1.1 PDF sets. The diagonal PDF uncertainties and data uncertainties are shown. Illustrative results are displayed for the proton structure function  $g_1^p$  measured at SLAC pre-EMC (top left) [92] and measured by SMC (top right) [97], as well as for the E154 measurement of the neutron structure function  $g_1^n$  (bottom left) [103] and the HERMES measurement of the deuteron structure function  $g_1^d$  (bottom right) [112].



PDF sets with each other and with all the NNPDFpol1.1 data is clear, with the uncertainties of the pre-EMC PDFs very large, and of the right amount to account for the missing information from the subsequent data. The post-NNPDFpol1.1 data pose more stringent requirements on the NNPDFpol1.1 PDFs, whose uncertainties are just large enough to accommodate them, as we had seen already from the  $\chi^2$  values of Table IV. It is apparent from these plots how the PDF uncertainty correctly accounts for the deviation between data, and predictions obtained using future test PDFs.

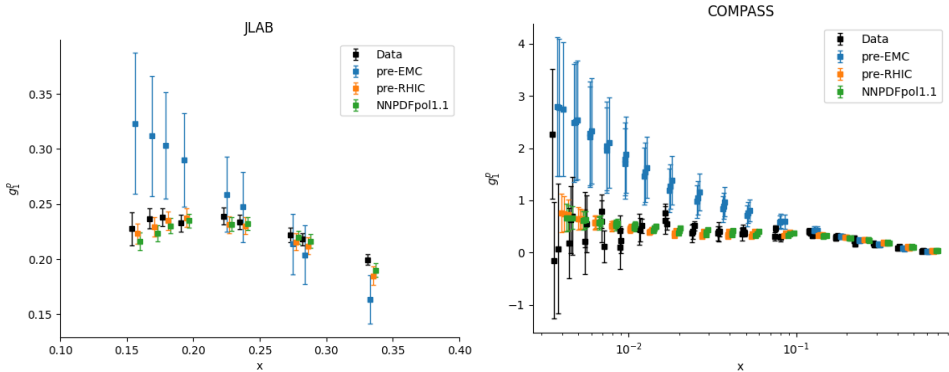


Fig. 7. The same as Fig. 6, but now comparing to post-NNPDFpol1.1 data. Results are shown for  $g_1^P$  measured at JLAB (left) [102] and COMPASS (right) [99].

## 5. Conclusions and outlook

Testing and validating the generalization power of machine-learning tools is one of the most difficult and important problems of artificial intelligence. In the context of PDF determination, this is an especially delicate issue since it is the generalization of PDFs outside the data region that makes it possible to obtain physics predictions at hadron colliders with finite uncertainties. This issue manifests itself as the problem of forward–backward compatibility of PDFs. Indeed, when PDFs are assumed to have some fixed underlying functional form, it is a not uncommon occurrence that subsequent data require enlarging the functional form, with the paradoxical consequence that adding new data results in larger, rather than smaller uncertainties (see *e.g.* Ref. [115]). PDFs determined using NNPDF methodology so far have been free of this problem, with later PDF sets always compatible within uncertainties with previous ones.

Here, we have suggested the use of forward–backward compatibility as a criterion for validating a PDF methodology. Using forward compatibility as a criterion for validation of a current methodology is actually more stringent than simply checking the forward compatibility of existing PDF sets. This is

because as the methodology improves, methodological uncertainties become smaller with fixed underlying data, as a consequence of the fact that more recent methodologies are more efficient in extracting information from the data.

We have seen that the current most recent NNPDF methodologies in the unpolarized and polarized case pass the future test, thereby showing that the NNPDF methodology correctly extrapolates outside the data region. It is interesting to ask whether the method can also be used for a comparative assessment of different PDF fitting methodology, possibly not only to validate *a posteriori* the extrapolation methodology, but also to actually optimize it. This question will be left to future investigations.

We thank all the members of the NNPDF Collaboration and the N<sup>3</sup>PDF team for innumerable discussions on PDF determination and validation, and especially Christopher Schwan for a critical reading of the manuscript and many insightful comments and Juan Rojo for several useful suggestions on the draft. S.F. is very grateful to Michał Przaszałowicz for managing to organize the LX Zakopane school as an online meeting in the current difficult circumstances, and to Marek Karliner for suggesting during the meeting to investigate the proton spin crisis using future tests. This work is supported by the European Research Council under the European Union's Horizon 2020 research and innovation Programme (grant agreement No. 740006). E.R.N. is supported by the UK STFC grant ST/T000600/1.

## REFERENCES

- [1] S. Lem, «Kongres Futurologiczny. 1<sup>st</sup> English edition», *Seabury Press*, 1974, translated by Michael Kandel.
- [2] S. Forte, S. Carrazza, «Parton distribution functions», [arXiv:2008.12305](https://arxiv.org/abs/2008.12305) [[hep-ph](#)].
- [3] S. Forte, «Parton Distributions at the Dawn of the LHC», *Acta Phys. Pol. B* **41**, 2859 (2010), [arXiv:1011.5247](https://arxiv.org/abs/1011.5247) [[hep-ph](#)].
- [4] J. Gao, L. Harland-Lang, J. Rojo, «The structure of the proton in the LHC precision era», *Phys. Rep.* **742**, 1 (2018), [arXiv:1709.04922](https://arxiv.org/abs/1709.04922) [[hep-ph](#)].
- [5] J.J. Ethier, E.R. Nocera, «Parton Distributions in Nucleons and Nuclei», *Annu. Rev. Nucl. Part. Sci.* **70**, 43 (2020), [arXiv:2001.07722](https://arxiv.org/abs/2001.07722) [[hep-ph](#)].
- [6] S. Forte, L. Garrido, J.I. Latorre, A. Piccione, «Neural network parametrization of deep-inelastic structure functions», *J. High Energy Phys.* **0205**, 062 (2002), [arXiv:hep-ph/0204232](https://arxiv.org/abs/hep-ph/0204232).
- [7] S. Carrazza, S. Forte, Z. Kassabov, J. Rojo, «Specialized minimal PDFs for optimized LHC calculations», *Eur. Phys. J. C* **76**, 205 (2016), [arXiv:1602.00005](https://arxiv.org/abs/1602.00005) [[hep-ph](#)].

- [8] New Muon Collaboration (M. Arneodo *et al.*), «Accurate measurement of  $F_2^d/F_2^p$  and  $R_d - R_p$ », *Nucl. Phys. B* **487**, 3 (1997), [arXiv:hep-ex/9611022](#).
- [9] New Muon Collaboration (M. Arneodo *et al.*), «Measurement of the proton and deuteron structure functions,  $F_2^p$  and  $F_2^d$ , and of the ratio  $\sigma_L/\sigma_T$ », *Nucl. Phys. B* **483**, 3 (1997), [arXiv:hep-ph/9610231](#).
- [10] L.W. Whitlow *et al.*, «Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross sections», *Phys. Lett. B* **282**, 475 (1992).
- [11] BCDMS Collaboration (A.C. Benvenuti *et al.*), «A high statistics measurement of the proton structure functions  $F_2(x, Q^2)$  and  $R$  from deep inelastic muon scattering at high  $Q^2$ », *Phys. Lett. B* **223**, 485 (1989).
- [12] BCDMS Collaboration (A.C. Benvenuti *et al.*), «A high statistics measurement of the Deuteron structure functions  $F_2(x, Q^2)$  and  $R$  from deep inelastic muon scattering at high  $Q^2$ », *Phys. Lett. B* **237**, 592 (1990).
- [13] CHORUS Collaboration (G. Onengut *et al.*), «Measurement of nucleon structure functions in neutrino scattering», *Phys. Lett. B* **632**, 65 (2006).
- [14] NuTeV Collaboration (M. Goncharov *et al.*), «Precise measurement of dimuon production cross-sections in  $\nu_\mu\text{Fe}$  and  $\bar{\nu}_\mu\text{Fe}$  deep inelastic scattering at the Tevatron», *Phys. Rev. D* **64**, 112006 (2001), [arXiv:hep-ex/0102049](#).
- [15] FNAL E866/NuSea Collaboration (R.S. Towell *et al.*), «Improved measurement of the  $\bar{d}/\bar{u}$  asymmetry in the nucleon sea», *Phys. Rev. D* **64**, 052002 (2001), [arXiv:hep-ex/0103030](#).
- [16] NuSea Collaboration (J.C. Webb *et al.*), «Absolute Drell–Yan dimuon cross sections in 800 GeV/ $c$   $pp$  and  $pd$  collisions», [arXiv:hep-ex/0302019](#).
- [17] G. Moreno *et al.*, «Dimuon production in proton–copper collisions at  $\sqrt{s} = 38.8$  GeV», *Phys. Rev. D* **43**, 2815 (1991).
- [18] ZEUS, H1 collaborations (H. Abramowicz *et al.*), «Combination of measurements of inclusive deep inelastic  $e^\pm p$  scattering cross sections and QCD analysis of HERA data», *Eur. Phys. J. C* **75**, 580 (2015), [arXiv:1506.06042 \[hep-ex\]](#).
- [19] ZEUS, H1 collaborations (H. Abramowicz *et al.*), «Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic  $ep$  scattering at HERA», *Eur. Phys. J. C* **78**, 473 (2018), [arXiv:1804.01019 \[hep-ex\]](#).
- [20] CDF Collaboration (T. Aaltonen *et al.*), «Measurement of the inclusive jet cross section at the Fermilab Tevatron  $p\bar{p}$  collider using a cone-based jet algorithm», *Phys. Rev. D* **78**, 052006 (2008); *Erratum ibid.* **79**, 119902 (2009).
- [21] D0 Collaboration (V.M. Abazov *et al.*), «Measurement of the muon charge asymmetry in  $p\bar{p} \rightarrow W + X \rightarrow \mu\nu + X$  events at  $\sqrt{s} = 1.96$  TeV», *Phys. Rev. D* **88**, 091102 (2013), [arXiv:1309.2591 \[hep-ex\]](#).

- [22] ATLAS Collaboration (G. Aad *et al.*), «Measurement of the inclusive  $W^\pm$  and  $Z/\gamma^*$  cross sections in the electron and muon decay channels in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector», *Phys. Rev. D* **85**, 072004 (2012), [arXiv:1109.5141 \[hep-ex\]](#).
- [23] ATLAS Collaboration (G. Aad *et al.*), «Measurement of the high-mass Drell–Yan differential cross-section in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector», *Phys. Lett. B* **725**, 223 (2013), [arXiv:1305.4192 \[hep-ex\]](#).
- [24] ATLAS Collaboration (G. Aad *et al.*), «Measurement of the low-mass Drell–Yan differential cross section at  $\sqrt{s} = 7$  TeV using the ATLAS detector», *J. High Energy Phys.* **1406**, 112 (2014), [arXiv:1404.1212 \[hep-ex\]](#).
- [25] ATLAS Collaboration (M. Aaboud *et al.*), «Precision measurement and interpretation of inclusive  $W^+$ ,  $W^-$  and  $Z/\gamma^*$  production cross sections with the ATLAS detector», *Eur. Phys. J. C* **77**, 367 (2017), [arXiv:1612.03016 \[hep-ex\]](#).
- [26] ATLAS Collaboration (G. Aad *et al.*), «Measurement of the double-differential high-mass Drell–Yan cross section in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector», *J. High Energy Phys.* **1608**, 009 (2016), [arXiv:1606.01736 \[hep-ex\]](#).
- [27] ATLAS Collaboration (G. Aad *et al.*), «Measurement of  $W^\pm$  and  $Z$ -boson production cross sections in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector», *Phys. Lett. B* **759**, 601 (2016), [arXiv:1603.09222 \[hep-ex\]](#).
- [28] ATLAS Collaboration (M. Aaboud *et al.*), «Measurement of differential cross sections and  $W^+/W^-$  cross-section ratios for  $W$  boson production in association with jets at  $\sqrt{s} = 8$  TeV with the ATLAS detector», *J. High Energy Phys.* **1805**, 077 (2018); *Erratum ibid.* **2010**, 048 (2020).
- [29] ATLAS Collaboration (G. Aad *et al.*), «Measurement of the transverse momentum and  $\phi_\eta^*$  distributions of Drell–Yan lepton pairs in proton–proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector», *Eur. Phys. J. C* **76**, 291 (2016), [arXiv:1512.02192 \[hep-ex\]](#).
- [30] ATLAS Collaboration (G. Aad *et al.*), «Measurement of the  $t\bar{t}$  production cross-section using  $e\mu$  events with  $b$ -tagged jets in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS detector», *Eur. Phys. J. C* **74**, 3109 (2014); *Addendum, ibid.* **76**, 642 (2016), [arXiv:1406.5375 \[hep-ex\]](#).
- [31] ATLAS Collaboration (G. Aad *et al.*), «Measurement of the  $t\bar{t}$  production cross-section in the lepton+jets channel at  $\sqrt{s} = 13$  TeV with the ATLAS experiment», *Phys. Lett. B* **810**, 135797 (2020), [arXiv:2006.13076 \[hep-ex\]](#).
- [32] ATLAS Collaboration (G. Aad *et al.*), «Measurements of top-quark pair differential cross-sections in the lepton+jets channel in  $pp$  collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector», *Eur. Phys. J. C* **76**, 538 (2016), [arXiv:1511.04716 \[hep-ex\]](#).

- [33] ATLAS Collaboration (M. Aaboud *et al.*), «Measurement of top quark pair differential cross-sections in the dilepton channel in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV with ATLAS», *Phys. Rev. D* **94**, 092003 (2016); *Addendum ibid.* **101**, 119901 (2020), [arXiv:1607.07281 \[hep-ex\]](#).
- [34] ATLAS Collaboration (M. Aaboud *et al.*), «Measurement of the inclusive jet cross-sections in proton–proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector», *J. High Energy Phys.* **1709**, 020 (2017), [arXiv:1706.03192 \[hep-ex\]](#).
- [35] ATLAS Collaboration (G. Aad *et al.*), «Measurement of dijet cross sections in  $pp$  collisions at 7 TeV centre-of-mass energy using the ATLAS detector», *J. High Energy Phys.* **1405**, 059 (2014), [arXiv:1312.3524 \[hep-ex\]](#).
- [36] ATLAS Collaboration (M. Aaboud *et al.*), «Measurement of the cross section for inclusive isolated-photon production in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector», *Phys. Lett. B* **770**, 473 (2017), [arXiv:1701.06882 \[hep-ex\]](#).
- [37] ATLAS Collaboration (G. Aad *et al.*), «Comprehensive measurements of  $t$ -channel single top-quark production cross sections at  $\sqrt{s} = 7$  TeV with the ATLAS detector», *Phys. Rev. D* **90**, 112006 (2014), [arXiv:1406.7844 \[hep-ex\]](#).
- [38] ATLAS Collaboration (M. Aaboud *et al.*), «Measurement of the inclusive cross-sections of single top-quark and top-antiquark  $t$ -channel production in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector», *J. High Energy Phys.* **1704**, 086 (2017), [arXiv:1609.03920 \[hep-ex\]](#).
- [39] ATLAS Collaboration (M. Aaboud *et al.*), «Fiducial, total and differential cross-section measurements of  $t$ -channel single top-quark production in  $pp$  collisions at 8 TeV using data collected by the ATLAS detector», *Eur. Phys. J. C* **77**, 531 (2017), [arXiv:1702.02859 \[hep-ex\]](#).
- [40] CMS Collaboration (S. Chatrchyan *et al.*), «Measurement of the Electron Charge Asymmetry in Inclusive  $W$  Production in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV», *Phys. Rev. Lett.* **109**, 111806 (2012), [arXiv:1206.2598 \[hep-ex\]](#).
- [41] CMS Collaboration (S. Chatrchyan *et al.*), «Measurement of the muon charge asymmetry in inclusive  $pp$  to  $WX$  production at  $\sqrt{s} = 7$  TeV and an improved determination of light parton distribution functions», *Phys. Rev. D* **90**, 032004 (2014), [arXiv:1312.6283 \[hep-ex\]](#).
- [42] CMS Collaboration (S. Chatrchyan *et al.*), «Measurement of the differential and double-differential Drell–Yan cross sections in proton–proton collisions at  $\sqrt{s} = 7$  TeV», *J. High Energy Phys.* **1312**, 030 (2013), [arXiv:1310.7291 \[hep-ex\]](#).
- [43] CMS Collaboration (V. Khachatryan *et al.*), «Measurement of the differential cross section and charge asymmetry for inclusive  $pp \rightarrow W^\pm + X$  production at  $\sqrt{s} = 8$  TeV», *Eur. Phys. J. C* **76**, 469 (2016), [arXiv:1603.01803 \[hep-ex\]](#).

- [44] CMS Collaboration (V. Khachatryan *et al.*), «Measurement of the  $Z$  boson differential cross section in transverse momentum and rapidity in proton–proton collisions at 8 TeV», *Phys. Lett. B* **749**, 187 (2015), [arXiv:1504.03511 \[hep-ex\]](#).
- [45] CMS Collaboration (S. Chatrchyan *et al.*), «Measurements of differential jet cross sections in proton–proton collisions at  $\sqrt{s} = 7$  TeV with the CMS detector», *Phys. Rev. D* **87**, 112002 (2013), [arXiv:1212.6660 \[hep-ex\]](#).
- [46] CMS Collaboration (A.M. Sirunyan *et al.*), «Measurement of the triple-differential dijet cross section in proton–proton collisions at  $\sqrt{s} = 8$  TeV and constraints on parton distribution functions», *Eur. Phys. J. C* **77**, 746 (2017), [arXiv:1705.02628 \[hep-ex\]](#).
- [47] S. Spannagel, «Top quark mass measurements with the CMS experiment at the LHC», *PoS DIS2016*, 150 (2016), [arXiv:1607.04972 \[hep-ex\]](#).
- [48] CMS Collaboration (V. Khachatryan *et al.*), «Measurement of the Top Quark Pair Production Cross Section in Proton–Proton Collisions at  $\sqrt{s} = 13$  TeV», *Phys. Rev. Lett.* **116**, 052002 (2016), [arXiv:1510.05302 \[hep-ex\]](#).
- [49] CMS Collaboration (V. Khachatryan *et al.*), «Measurement of the differential cross section for top quark pair production in  $pp$  collisions at  $\sqrt{s} = 8$  TeV», *Eur. Phys. J. C* **75**, 542 (2015), [arXiv:1505.04480 \[hep-ex\]](#).
- [50] CMS Collaboration (A.M. Sirunyan *et al.*), «Measurement of the inclusive  $t\bar{t}$  cross section in  $pp$  collisions at  $\sqrt{s} = 5.02$  TeV using final states with at least one charged lepton», *J. High Energy Phys.* **1803**, 115 (2018), [arXiv:1711.03143 \[hep-ex\]](#).
- [51] CMS Collaboration (A.M. Sirunyan *et al.*), «Measurement of double-differential cross sections for top quark pair production in  $pp$  collisions at  $\sqrt{s} = 8$  TeV and impact on parton distribution functions», *Eur. Phys. J. C* **77**, 459 (2017), [arXiv:1703.01630 \[hep-ex\]](#).
- [52] CMS Collaboration (A.M. Sirunyan *et al.*), «Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from  $pp$  collisions at  $\sqrt{s} = 13$  TeV», *Phys. Rev. D* **97**, 112003 (2018), [arXiv:1803.08856 \[hep-ex\]](#).
- [53] CMS Collaboration (A.M. Sirunyan *et al.*), «Measurements of  $t\bar{t}$  differential cross sections in proton–proton collisions at  $\sqrt{s} = 13$  TeV using events containing two leptons», *J. High Energy Phys.* **1902**, 149 (2019), [arXiv:1811.06625 \[hep-ex\]](#).
- [54] CMS Collaboration (S. Chatrchyan *et al.*), «Measurement of the single-top-quark  $t$ -channel cross section in  $pp$  collisions at  $\sqrt{s} = 7$  TeV», *J. High Energy Phys.* **1212**, 035 (2012), [arXiv:1209.4533 \[hep-ex\]](#).
- [55] CMS Collaboration (V. Khachatryan *et al.*), «Measurement of the  $t$ -channel single-top-quark production cross section and of the  $|V_{tb}|$  CKM matrix element in  $pp$  collisions at  $\sqrt{s} = 8$  TeV», *J. High Energy Phys.* **1406**, 090 (2014), [arXiv:1403.7366 \[hep-ex\]](#).

- [56] CMS Collaboration (A.M. Sirunyan *et al.*), «Cross section measurement of  $t$ -channel single top quark production in  $pp$  collisions at  $\sqrt{s} = 13$  TeV», *Phys. Lett. B* **772**, 752 (2017), [arXiv:1610.00678 \[hep-ex\]](#).
- [57] LHCb Collaboration (R. Aaij *et al.*), «Measurement of the cross-section for  $Z \rightarrow e^+e^-$  production in  $pp$  collisions at  $\sqrt{s} = 7$  TeV», *J. High Energy Phys.* **1302**, 106 (2013), [arXiv:1212.4620 \[hep-ex\]](#).
- [58] LHCb Collaboration (R. Aaij *et al.*), «Measurement of forward  $Z \rightarrow e^+e^-$  production at  $\sqrt{s} = 8$  TeV», *J. High Energy Phys.* **1505**, 109 (2015), [arXiv:1503.00963 \[hep-ex\]](#).
- [59] LHCb Collaboration (R. Aaij *et al.*), «Measurement of the forward  $W$  boson cross-section in  $pp$  collisions at  $\sqrt{s} = 7$  TeV», *J. High Energy Phys.* **1412**, 079 (2014), [arXiv:1408.4354 \[hep-ex\]](#).
- [60] LHCb Collaboration (R. Aaij *et al.*), «Measurement of forward  $W$  and  $Z$  boson production in  $pp$  collisions at  $\sqrt{s} = 8$  TeV», *J. High Energy Phys.* **1601**, 155 (2016), [arXiv:1511.08039 \[hep-ex\]](#).
- [61] LHCb Collaboration (R. Aaij *et al.*), «Measurement of the forward  $Z$  boson production cross-section in  $pp$  collisions at  $\sqrt{s} = 13$  TeV», *J. High Energy Phys.* **1609**, 136 (2016), [arXiv:1607.06495 \[hep-ex\]](#).
- [62] NNPDF Collaboration (R.D. Ball *et al.*), «Parton distributions for the LHC run II», *J. High Energy Phys.* **1504**, 040 (2015), [arXiv:1410.8849 \[hep-ph\]](#).
- [63] NNPDF Collaboration (R.D. Ball *et al.*), «The structure of the proton to one-percent accuracy», in preparation, 2021.
- [64] L. Del Debbio, M. Wilson, in preparation, 2021.
- [65] NNPDF Collaboration (R.D. Ball *et al.*), «Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO», *Nucl. Phys. B* **855**, 153 (2012), [arXiv:1107.2652 \[hep-ph\]](#).
- [66] S. Carrazza, J. Cruz-Martinez, «Towards a new generation of parton densities with deep learning models», *Eur. Phys. J. C* **79**, 676 (2019), [arXiv:1907.05075 \[hep-ph\]](#).
- [67] NNPDF Collaboration (R.D. Ball *et al.*), «Parton distributions from high-precision collider data», *Eur. Phys. J. C* **77**, 663 (2017), [arXiv:1706.00428 \[hep-ph\]](#).
- [68] E.R. Nocera, M. Ubiali, «Constraining the gluon PDF at large  $x$  with LHC data», *PoS DIS2017*, 008 (2018), [arXiv:1709.09690 \[hep-ph\]](#).
- [69] CTEQ Collaboration (J. Botts *et al.*), «CTEQ parton distributions and flavor dependence of sea quarks», *Phys. Lett. B* **304**, 159 (1993), [arXiv:hep-ph/9303255](#).
- [70] A.D. Martin, W.J. Stirling, R.G. Roberts, «New information on parton distributions», *Phys. Rev. D* **47**, 867 (1993).
- [71] R.D. Ball *et al.*, «A first unbiased global NLO determination of parton distributions and their uncertainties», *Nucl. Phys. B* **838**, 136 (2010), [arXiv:1002.4407 \[hep-ph\]](#).
- [72] A. De Roeck, «Proton structure function data and search for BFKL signatures at HERA», *Acta Phys. Pol. B* **27**, 1175 (1996).

- [73] W.-K. Tung, «Status of global QCD analysis and the parton structure of the nucleon», in: «12<sup>th</sup> International Workshop on Deep Inelastic Scattering (DIS 2004)», 2004, pp. 218–230, [arXiv:hep-ph/0409145](#).
- [74] R.D. Ball, «Global Parton Distributions for the LHC Run II», *Nuovo Cim. C* **38**, 127 (2016), [arXiv:1507.07891 \[hep-ph\]](#).
- [75] The NNPDF Collaboration (R.D. Ball *et al.*), «Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties», *J. High Energy Phys.* **1005**, 075 (2010), [arXiv:0912.2276 \[hep-ph\]](#).
- [76] R.D. Ball *et al.*, «Parton Distribution Benchmarking with LHC Data», *J. High Energy Phys.* **1304**, 125 (2013), [arXiv:1211.5142 \[hep-ph\]](#).
- [77] S. Forte, E. Laenen, P. Nason, J. Rojo, «Heavy quarks in deep-inelastic scattering», *Nucl. Phys. B* **834**, 116 (2010), [arXiv:1001.2312 \[hep-ph\]](#).
- [78] R.D. Ball *et al.*, «Intrinsic charm in a matched general-mass scheme», *Phys. Lett. B* **754**, 49 (2016), [arXiv:1510.00009 \[hep-ph\]](#).
- [79] R.D. Ball, M. Bonvini, L. Rottoli, «Charm in Deep-Inelastic Scattering», *J. High Energy Phys.* **1511**, 122 (2015), [arXiv:1510.02491 \[hep-ph\]](#).
- [80] NNPDF Collaboration (R.D. Ball, E.R. Nocera, R.L. Pearson), «Nuclear Uncertainties in the Determination of Proton PDFs», *Eur. Phys. J. C* **79**, 282 (2019), [arXiv:1812.09074 \[hep-ph\]](#).
- [81] NNPDF Collaboration (R. Abdul Khalek, J.J. Ethier, J. Rojo), «Nuclear parton distributions from lepton–nucleus scattering and the impact of an electron–ion collider», *Eur. Phys. J. C* **79**, 471 (2019), [arXiv:1904.00018 \[hep-ph\]](#).
- [82] R.D. Ball, E.R. Nocera, R.L. Pearson, «Deuteron uncertainties in the determination of proton PDFs», *Eur. Phys. J. C* **81**, 37 (2021), [arXiv:2011.00009 \[hep-ph\]](#).
- [83] NNPDF Collaboration (R. Abdul Khalek *et al.*), «Parton distributions with theory uncertainties: general formalism and first phenomenological studies», *Eur. Phys. J. C* **79**, 931 (2019), [arXiv:1906.10698 \[hep-ph\]](#).
- [84] F. Demartin *et al.*, «Impact of PDF and  $\alpha_s$  uncertainties on Higgs Production in gluon fusion at hadron colliders», *Phys. Rev. D* **82**, 014002 (2010), [arXiv:1004.0962 \[hep-ph\]](#).
- [85] LHCb Collaboration (R. Aaij *et al.*), «Measurement of the forward Z boson production cross-section in  $pp$  collisions at  $\sqrt{s} = 7$  TeV», *J. High Energy Phys.* **1508**, 039 (2015), [arXiv:1505.07024 \[hep-ex\]](#).
- [86] C.A. Aidala, S.D. Bass, D. Hasch, G.K. Mallot, «The spin structure of the nucleon», *Rev. Mod. Phys.* **85**, 655 (2013), [arXiv:1209.2803 \[hep-ph\]](#).
- [87] NNPDF Collaboration (E.R. Nocera *et al.*), «A first unbiased global determination of polarized PDFs and their uncertainties», *Nucl. Phys. B* **887**, 276 (2014), [arXiv:1406.5539 \[hep-ph\]](#).
- [88] E.R. Nocera, «Unbiased polarized PDFs upgraded with new inclusive DIS data», *J. Phys.: Conf. Ser.* **678**, 012030 (2016), [arXiv:1510.04248 \[hep-ph\]](#).



- [89] E.R. Nocera, «Impact of Recent RHIC Data on Helicity-Dependent Parton Distribution Functions», in: «22<sup>nd</sup> International Symposium on Spin Physics», 2017, [arXiv:1702.05077 \[hep-ph\]](#).
- [90] M.J. Alguard *et al.*, «Deep Inelastic *ep* Asymmetry Measurements and Comparison with the Bjorken Sum Rule and Models of Proton Spin Structure», *Phys. Rev. Lett.* **41**, 70 (1978).
- [91] COMPASS Collaboration (C. Adolph *et al.*), «Leading and next-to-leading order gluon polarization in the nucleon and longitudinal double spin asymmetries from open charm muoproduction», *Phys. Rev. D* **87**, 052018 (2013), [arXiv:1211.6849 \[hep-ex\]](#).
- [92] G. Baum *et al.*, «A New Measurement of Deep Inelastic *ep* Asymmetries», *Phys. Rev. Lett.* **51**, 1135 (1983).
- [93] STAR Collaboration (L. Adamczyk *et al.*), «Longitudinal and transverse spin asymmetries for inclusive jet production at mid-rapidity in polarized *p + p* collisions at  $\sqrt{s} = 200$  GeV», *Phys. Rev. D* **86**, 032006 (2012), [arXiv:1205.2735 \[nucl-ex\]](#).
- [94] STAR Collaboration (L. Adamczyk *et al.*), «Precision Measurement of the Longitudinal Double-spin Asymmetry for Inclusive Jet Production in Polarized Proton Collisions at  $\sqrt{s} = 200$  GeV», *Phys. Rev. Lett.* **115**, 092002 (2015), [arXiv:1405.5134 \[hep-ex\]](#).
- [95] European Muon Collaboration (J. Ashman *et al.*), «An Investigation of the Spin Structure of the Proton in Deep Inelastic Scattering of Polarized Muons on Polarized Protons», *Nucl. Phys. B* **328**, 1 (1989).
- [96] PHENIX Collaboration (A. Adare *et al.*), «Event structure and double helicity asymmetry in jet production from polarized *p + p* collisions at  $\sqrt{s} = 200$  GeV», *Phys. Rev. D* **84**, 012006 (2011), [arXiv:1009.4921 \[hep-ex\]](#).
- [97] Spin Muon Collaboration (B. Adeva *et al.*), «Spin asymmetries  $A_1$  and structure functions  $g_1$  of the proton and the deuteron from polarized high-energy muon scattering», *Phys. Rev. D* **58**, 112001 (1998).
- [98] Spin Muon Collaboration (B. Adeva *et al.*), «Spin asymmetries  $A_1$  of the proton and the deuteron in the low  $x$  and low  $Q^2$  region from polarized high-energy muon scattering», *Phys. Rev. D* **60**, 072004 (1999); *Erratum ibid.* **62**, 079902 (2000).
- [99] COMPASS Collaboration (C. Adolph *et al.*), «The spin structure function  $g_1^p$  of the proton and a test of the Bjorken sum rule», *Phys. Lett. B* **753**, 18 (2016), [arXiv:1503.08935 \[hep-ex\]](#).
- [100] COMPASS Collaboration (C. Adolph *et al.*), «Final COMPASS results on the deuteron spin-dependent structure function  $g_1^d$  and the Bjorken sum rule», *Phys. Lett. B* **769**, 34 (2017), [arXiv:1612.00620 \[hep-ex\]](#).
- [101] E143 Collaboration (K. Abe *et al.*), «Measurements of the proton and deuteron spin structure functions  $g_1$  and  $g_2$ », *Phys. Rev. D* **58**, 112003 (1998), [arXiv:hep-ph/9802357](#).

- [102] CLAS Collaboration (Y. Prok *et al.*), «Precision measurements of  $g_1$  of the proton and the deuteron with 6 GeV electrons», *Phys. Rev. C* **90**, 025212 (2014), [arXiv:1404.6231 \[nucl-ex\]](#).
- [103] E154 Collaboration (K. Abe *et al.*), «Precision Determination of the Neutron Spin Structure Function  $g_1^n$ », *Phys. Rev. Lett.* **79**, 26 (1997), [arXiv:hep-ex/9705012](#).
- [104] CLAS Collaboration (N. Guler *et al.*), «Precise determination of the deuteron spin structure at low to moderate  $Q^2$  with CLAS and extraction of the neutron contribution», *Phys. Rev. C* **92**, 055201 (2015), [arXiv:1505.07877 \[nucl-ex\]](#).
- [105] E155 Collaboration (P.L. Anthony *et al.*), «Measurements of the  $Q^2$ -dependence of the proton and neutron spin structure functions  $g_p^1$  and  $g_n^1$ », *Phys. Lett. B* **493**, 19 (2000), [arXiv:hep-ph/0007248](#).
- [106] Jefferson Lab Hall A Collaboration (D.S. Parno *et al.*), «Precision measurements of  $A_1^n$  in the deep inelastic regime», *Phys. Lett. B* **744**, 309 (2015), [arXiv:1406.1207 \[nucl-ex\]](#).
- [107] COMPASS Collaboration (V.Y. Alexakhin *et al.*), «The deuteron spin-dependent structure function  $g_1^d$  and its first moment», *Phys. Lett. B* **647**, 8 (2007), [arXiv:hep-ex/0609038](#).
- [108] STAR Collaboration (L. Adamczyk *et al.*), «Measurement of the cross section and longitudinal double-spin asymmetry for di-jet production in polarized  $pp$  collisions at  $\sqrt{s} = 200$  GeV», *Phys. Rev. D* **95**, 071103 (2017), [arXiv:1610.06616 \[hep-ex\]](#).
- [109] COMPASS Collaboration (M.G. Alekseev *et al.*), «The spin-dependent structure function of the proton  $g_1^p$  and a test of the Bjorken sum rule», *Phys. Lett. B* **690**, 466 (2010), [arXiv:1001.4654 \[hep-ex\]](#).
- [110] STAR Collaboration (J. Adam *et al.*), «Longitudinal double-spin asymmetry for inclusive jet and dijet production in  $pp$  collisions at  $\sqrt{s} = 510$  GeV», *Phys. Rev. D* **100**, 052005 (2019), [arXiv:1906.02740 \[hep-ex\]](#).
- [111] HERMES Collaboration (K. Ackerstaff *et al.*), «Measurement of the neutron spin structure function  $g_1^n$  with a polarized  $^3\text{He}$  internal target», *Phys. Lett. B* **404**, 383 (1997), [arXiv:hep-ex/9703005](#).
- [112] HERMES Collaboration (A. Airapetian *et al.*), «Precise determination of the spin structure function  $g_1$  of the proton, deuteron and neutron», *Phys. Rev. D* **75**, 012007 (2007), [arXiv:hep-ex/0609039](#).
- [113] G. Altarelli, G.G. Ross, «The anomalous gluon contribution to polarized lepton production», *Phys. Lett. B* **212**, 391 (1988).
- [114] E. Leader, M. Anselmino, «A crisis in the parton model: Where, oh where is the proton's spin?», *AIP Conf. Proc.* **187**, 764 (1989).
- [115] T.-J. Hou *et al.*, «New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC», *Phys. Rev. D* **103**, 014013 (2021), [arXiv:1912.10053 \[hep-ph\]](#).