Quantifying the interplay of experimental constraints in analyses of parton distributions

Xiaoxian Jing[1], Amanda Cooper-Sarkar[2], Aurore Courtoy[3], Thomas Cridge[4], Francesco Giuli[5],
Lucian Harland-Lang,[6] T. J. Hobbs[7], Joey Huston[8], Pavel Nadolsky[1,*],
Robert S. Thorne[6], Keping Xie[9], and C.-P. Yuan[8]

[1]*Department of Physics, Southern Methodist University, Dallas, Texas 75275-0181, USA*
[2]*Department of Physics, University of Oxford, Oxford OX1 3RH, United Kingdom*
[3]*Instituto de Física, Universidad Nacional Autónoma de México,*
Apartado Postal 20-364, 01000 Ciudad de México, Mexico
[4]*Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, Hamburg 22607, Germany*
[5]*CERN, CH-1211 Geneva, Switzerland*
[6]*Department of Physics and Astronomy, University College, London WC1E 6BT, United Kingdom*
[7]*High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[8]*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*
[9]*Pittsburgh Particle Physics, Astrophysics, and Cosmology Center, Department of Physics and Astronomy,*
University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

Parton distribution functions (PDFs) play a central role in calculations for the LHC. To gain a deeper understanding of the emergence and interplay of constraints on the PDFs in the global QCD analyses, it is important to examine the relative significance and mutual compatibility of the experimental datasets included in the PDF fits. Toward this goal, we discuss the L_2 sensitivity, a convenient statistical indicator for exploring the statistical pulls of individual datasets on the best-fit PDFs and identifying tensions between competing datasets. Unlike the Lagrange multiplier method, the L_2 sensitivity can be quickly computed for a range of PDFs and momentum fractions using the published Hessian error sets. We employ the L_2 sensitivity as a common metric to study the relative importance of datasets in the recent ATLAS, CTEQ-TEA, MSHT, and reduced PDF4LHC21 PDF analyses at next-to-next-to-leading-order and approximate next-to-next-to-next-to-leading-order. We illustrate how this method can aid the users of PDFs to identify datasets that are important for a PDF at a given kinematic point, to study quark flavor composition and other detailed features of the PDFs, and to compare the data pulls on the PDFs for various perturbative orders and functional forms. We also address the feasibility of computing the sensitivities using Monte Carlo error PDFs. Together with the article, we present a companion interactive website with a large collection of plotted L_2 sensitivities for eight recent PDF releases and a C++ program to plot the L_2 sensitivities.

I. INTRODUCTION

Parton distribution functions (PDFs) used for predictions at the Large Hadron Collider (LHC) and elsewhere are determined by multivariate fits to a selection of precise experimental measurements from deeply inelastic scattering, vector boson, jet, top-quark production, and other

processes, including data from the LHC. The number of experimental datasets included in the recent PDF fits ranges from a few most precise ones to several tens in the most comprehensive global fits. The contemporary next-to-next-to-leading-order (NNLO) PDF fits may either use the Hessian method [1,2] or the Monte Carlo method [3] to both determine the central PDF and estimate the uncertainty on the PDF parameters. For data from an experimental measurement to influence the PDF fit in a particular region of x and Q^2, two conditions usually must be met: (i) the parton-level dynamics underlying the measurement must substantially depend on a particular PDF (e.g., that of the gluon), as manifest via a statistical correlation between the PDF in that kinematic region and the experimental observable [4]; and (ii) the measurement must have

sufficient resolving power to nontrivially contribute to the likelihood function of a QCD analysis. The latter depends on the experimental errors of the dataset, both statistical and systematic.

In general, all experimental datasets included in a PDF fit have some influence on the determination of a given PDF flavor for a given x, Q^2, with the amount of influence typically varying over a wide range. In this article we focus on the examination of the statistical pulls that the experimental datasets impose on the PDFs at the best fit, determined from the dependence of the goodness-of-fit function χ^2 on the PDF values. In numerical efforts to quantify the PDF pulls of commonly fitted experiments through figures of merit [5,6], some datasets are found to be particularly influential, with the aggregated PDF sensitivity of the full dataset typically dominated by a handful of measurements and a power-law-like falloff in the pulls of less sensitive experiments. Detailed investigations of these sensitivities may indicate ways for increasing the collective impact of the full dataset, either for currently available measurements, from the perspective of alternative implementations of new measurements, or in anticipation of possible data that might be recorded at upcoming facilities like the high-luminosity LHC, Electron-Ion Collider, or neutrino experiments.

For many experiments, their pulls on the PDFs are indistinguishable from those due to statistical fluctuations in the data samples. Such pulls may either reflect a good agreement of the experiment with the best-fit PDF model (in which case such an experiment may nevertheless impose essential constraints on the PDF uncertainty); or it may be that the constraints from the experiment are just weak. In addition, due to imperfections in the data measurements and/ or theory, there can be tensions among datasets and even within a single experiment or dataset (e.g., among differing rapidity bins in hadroproduction experiments), resulting in significant opposing pulls on the PDFs. The existence of disagreements between some available datasets has been noticed since the early days of PDF fitting (see, e.g., [7]). Tensions among the experiments may lead to a smaller reduction in the PDF uncertainty than might have been expected based on the nominal constraining power of the individual datasets and have motivated introduction of tolerance [1,8,9] on the final uncertainty.

The complex inner workings of a PDF fit may leave an impression of a black box, especially to end users of the PDFs. For this reason, there is substantial interest in numerical methods to quantify the pulls among fitted experiments inside the full fit in terms of the respective log-likelihood variations for these experiments under systematic shifts in the PDFs. Such tools can thereby clarify the influences on the extracted PDFs by the various datasets as well as by variations in the assumed theoretical formalism, such as the perturbative order or deployment of nonperturbative corrections. One such technique is the Lagrange multiplier (LM) scan [10], which provides robust (not dependent on the Gaussian approximation) information on the constraints on a particular PDF or observable due to the datasets. This technique can, in practice, only be applied by the PDF authors, and in addition it is computationally expensive and limited to fixed values of the parton momentum fraction, x, and the factorization scale, Q. Other popular approximate techniques include Hessian profiling and updating [11–13] as well as Monte Carlo PDF reweighting [3,14–16], which however depend either on the choice of statistical weights or of tolerance, as well are limited to using static parametric forms.

Another technique without the above drawbacks is based on the L_2 sensitivity measure [6], employed together with the LM scans in the recent CT18 global analysis [17] and in the PDF4LHC21 benchmarking study [18]. The L_2 sensitivity technique maps the influences of the datasets on a given PDF or PDF-dependent quantity by taking into account both the correlation of each dataset with the PDF and the degree to which the dataset is influencing the determination of the PDF. The L_2 sensitivity can be plotted against x in a PDF for a given Q or, in the case of the parton-parton luminosities, as a function of the final-state invariant mass M_X for a given $\sqrt{s}$. In this form, the L_2 sensitivity quantifies the pull of each experiment on the PDF at a given x and Q.

The L_2 sensitivity is calculated using the log-likelihood (χ^2) values for the fitted experiments and the error PDF sets—the readily available outputs of the PDF fits. As such, it streamlines comparisons among independent PDF analyses. A few years ago, the CTEQ-JLab and CTEQ-TEA groups performed such comparisons at the NLO accuracy in the QCD coupling strength with the goal to understand the role of large-x and nuclear experimental measurements [19].

In this article, we expand the L_2 sensitivity comparisons to the NNLO PDFs by the ATLAS [20], CTEQ-TEA [13], and MSHT [21] groups. We include the recent CT18As_Lat NNLO analysis with lattice QCD constraints [22] and the approximate N3LO analysis by the MSHT group, MSHT20aN3LO [23].

Aside from the utility of quantifying the statistical pulls of fitted data on the PDFs for the sake of practical phenomenology, there is another fundamental motivation for applying the LM or L_2 sensitivity methods. They both explore the parametric dependence of χ^2 in the immediate vicinity of the global minimum of χ^2, i.e., the best fit. They therefore contain rich information regarding the multidimensional geometry of the likelihood function, which is closely connected to the ultimate PDF uncertainty and its interpretation. For example, the L_2 method can elucidate the complicated correlations among PDF flavors or regions of x or Q. Interpretation of the L_2 or similar methods, especially when contrasting distinct fits, invokes a range of subtleties in the precise definition of the PDF uncertainty, use of tolerance criteria, and relationships between methods

based on Hessian or Monte Carlo uncertainties. We discuss these more formal aspects of uncertainty quantification and their relation to the L_2 method before proceeding to the numerical comparisons of the fits by three groups.

The outline of this paper is as follows. In Sec. II, we first give a general description of the L_2 sensitivity and explain how it can be calculated for both Hessian-based and Monte Carlo replica-based PDF fits. In Sec. III we summarize the PDF sets that will be considered in this analysis, discussing the data and theory settings, as well as χ^2 and PDF error definitions. In Sec. IV the sensitivities for the nonglobal PDF sets, namely dedicated ATLAS and reduced benchmarking fits, are presented. In Sec. V the sensitivities for the global CT family of PDF sets are presented. In Sec. VI the sensitivities for the MSHT20 fits are presented. In Sec. VII direct comparisons are made of the sensitivities for a range experiments and parton flavors between the different PDF sets. Finally, in Sec. VIII we conclude. The Appendix summarizes a computation of the sensitivities using Monte Carlo replicas that results in a close agreement with the Hessian approach.

II. DEFINITIONS AND BASIC PROPERTIES OF THE L_2 SENSITIVITY

A. The Hessian method

Error PDFs are widely used to estimate probability distributions for PDF-dependent quantities according to two common methods.

The ATLAS, CTEQ-TEA, and MSHT groups adopt the Hessian format [1,2] as the default to publish their PDF error sets. An ensemble of D Hessian error PDFs estimates the uncertainty by assuming that the probability distribution is approximately Gaussian. In a notation adopted from Ref. [24], a function $X(\vec{R})$ of the parameters R_i in the vicinity of the minimum of the global χ^2 corresponding to $\vec{R} = \vec{0}$ and $X(\vec{0}) \equiv X_0$ can be estimated as the Taylor series expansion,

$$X(\vec{R}) = X_0 + \sum_{i=1}^{D} \frac{\partial X}{\partial R_i}\bigg|_{\vec{R}=\vec{0}} R_i + \frac{1}{2} \sum_{i,j=1}^{D} \frac{\partial^2 X}{\partial R_i \partial R_j}\bigg|_{\vec{R}=\vec{0}} R_i R_j + \cdots.$$

$$(1)$$

Given $X_{\pm i} \equiv X(0, \ldots, R_i = \pm 1, 0, \ldots)$ for a pair of PDF displacements $R_{\pm i} \equiv \pm 1$ at the 68% confidence level (C.L.) along the eigenvector direction i, the first-order derivative in this direction is estimated by a symmetrized finite difference,

$$\frac{\partial X}{\partial R_i}\bigg|_{\vec{R}=\vec{0}} \approx \frac{X_{+i} - X_{-i}}{2}.$$

$$(2)$$

A symmetric estimate of the 68% C.L. PDF uncertainty [2] then follows as the maximal variation of $X(\vec{R})$ within a

hypersphere of unit radius centered at the global minimum, called the "tolerance hypersphere":

$$\delta_{\mathrm{H}} X = |\vec{\nabla} X| = \frac{1}{2} \sqrt{\sum_{i=1}^{D} [X_{+i} - X_{-i}]^2}.$$

$$(3)$$

The second-order Taylor terms in Eq. (1) are important when the probability distribution is asymmetric. The full description of the second-order terms, while possible in principle [24], would require having additional Hessian eigenvector sets that are not provided in the published PDF ensembles. Contributions from diagonal second-order derivatives, $\partial^2 X/\partial R_i^2$, can be estimated by using the asymmetric PDF uncertainties [25] with the usual Hessian PDFs. The linear approximation captures the essential features of the uncertainties, while the complete description of the nonlinear terms involves many subtleties [24].

We will thus restrict ourselves entirely to the linear approximations and will use symmetrized finite-difference formulas like in Eq. (2) to minimize nonlinear terms in subsequent derivations. In this spirit, the Pearson correlation between two quantities $X(\vec{R})$ and $Y(\vec{R})$, interpreted as the cosine of the correlation angle for X and Y in the PDF parameter space, can be computed as [1,4]

$$C_{\mathrm{H}}(X, Y) = \frac{1}{4\delta_{\mathrm{H}} X \, \delta_{\mathrm{H}} Y} \sum_{i=1}^{D} (X_{+i} - X_{-i})(Y_{+i} - Y_{-i}).$$

$$(4)$$

The PDF sensitivity is a statistical indicator that visualizes constraints from the included experiments on the PDFs. In the Hessian representation, the L_2 sensitivity for some $f(\vec{R})$ reads [6]

$$S_{f,L2}^{\mathrm{H}}(E) \equiv \frac{\vec{\nabla}\chi_E^2 \cdot \vec{\nabla} f}{\delta_{\mathrm{H}} f},$$

$$= (\delta_{\mathrm{H}}\chi_E^2) \, C_{\mathrm{H}}(f, \chi_E^2),$$

$$(5)$$

where $C_{\mathrm{H}}(f, \chi_E^2)$ represents the cosine of the correlation angle between f and the χ^2 for experiment E, evaluated over the $2D$ Hessian eigenvector sets. Thus, if the direction of decreasing χ_E^2 of dataset E is also the direction of decreasing values of the PDF f (at a given x value), the two quantities are positively correlated, and $S_{f,L2}^{\mathrm{H}}(E)$ is positive. This indicates that the data from this dataset would like to pull the PDF downward. If the two quantities are anti-correlated, then $S_{f,L2}^{\mathrm{H}}(E)$ is negative, and the data would like to pull the PDF upwards.

The name "L_2 sensitivity" reflects its reliance on the χ^2 to quantify the pulls of experimental data on the PDFs, i.e., on the quadratic, or L_2, norm of the vectors of the statistical residuals between theoretical predictions and experimental measurements. This emphasizes its distinction from the

alternative definitions of the PDF sensitivity that are also possible. One alternative definition operates with the absolute values of the residuals, i.e., the L_1 norm, and was employed in the first practical studies of the sensitivities using the PDFSENSE program [5,6].

The other connotation is that the L_2 sensitivity serves as the "second Lagrangian technique" that complements the classical conditional optimization with multiplier terms invented by J.-L. Lagrange. This technique is a fast approximation to the Lagrange multiplier scan [10], now realized using the published eigenvector sets outside of the PDF fit. The LM scan and L_2 sensitivity both visualize the probability in the multidimensional PDF parameter space.

B. The Monte Carlo method

The Monte Carlo (MC) method [3] for PDF uncertainties is adopted by default by NNPDF, while all groups can convert between the Hessian and MC error PDFs in both directions [26–29]. By analogy with the Hessian approach, the L_2 sensitivity can be introduced in the MC method [5]. The MC case involves subtleties that are absent in the Hessian method. The MC method provides an ensemble of PDF replicas $f_a^{(k)}(x, Q) \equiv f^{(k)}$ with $k \in [1, ..., N_{\rm rep}]$, i.e., stochastically generated error PDF sets. By evaluating a PDF-dependent quantity $X(f_a^{(k)}(x, Q)) \equiv X_k$ with these replica PDFs, the probability distribution for X might in principle be reconstructed with arbitrary accuracy for a sufficiently large $N_{\rm rep}$, including the non-Gaussian features. However, high dimensionality of the PDF parameter space presents a salient impediment and requires careful implementation of the sensitivity.

As in the Hessian case, we find that accounting for the asymmetries of the probability, while possible in principle, substantially complicates the analysis. We will therefore work with the MC formulas that average over the asymmetries in the probability, such as the standard formulas for the central value and PDF uncertainty of X given, respectively, by the expectation value and standard deviation:

$$\langle X \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} X_k, \qquad (6)$$

$$\delta_{\rm MC} X = \sqrt{\frac{1}{N_{\rm rep} - 1} \sum_{k=1}^{N_{\rm rep}} (X_k - \langle X \rangle)^2}. \qquad (7)$$

Similarly, the Pearson correlation is represented as

$$C_{\rm MC}(X, Y) = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{\delta_{\rm MC} X \cdot \delta_{\rm MC} Y}. \qquad (8)$$

The MC analog of the Hessian L_2 sensitivity in Eq. (5) is written as

$$S_{f,L2}^{\rm MC}(E) = \eta(D) \, (\delta_{\rm MC} \chi_E^2) \, C_{\rm MC}(f, \chi_E^2), \qquad (9)$$

in terms of the MC estimates (7) and (8) for f and χ_E^2. We have also introduced a normalization constant $\eta(D)$ that depends on the number of PDF parameters D and may differ from unity, depending on how the MC ensemble samples the space of PDF solutions.

The Hessian and Monte Carlo definitions of the L_2 sensitivity are equivalent in the linear approximation, which in turn is justified when displacements of the PDF error sets from the global minimum are small. In practice, the linearity condition for χ_E^2 does not hold with the standard MC replicas from NNPDF or other groups, and hence the direct estimation of $S_{f,L2}^{\rm MC}(E)$ according to Eq. (9) may be vulnerable to errors. The reason is that, in a typical MC ensemble, the majority of PDF replicas lie several standard deviations away from the global minimum as a consequence of the high dimensionality of the parameter space [24]. As χ_E^2 includes quadratic and higher powers of the PDF parameters, such far displacements introduce large nonlinearities, which lead to the accuracy loss in $S_{f,L2}^{\rm MC}(E)$.

Thus, to obtain a numerically stable estimate of $S_{f,L2}^{\rm MC}(E)$ using the MC replicas, one must avoid large parameter displacements from the global minimum, which in many dimensions requires unconventional sampling. The Appendix presents such an example, in which a Hessian ensemble is converted into a MC one so that the MC replicas are distributed uniformly on the surface of a hypersphere rather than over the whole parameter space. With this ensemble, probability integrations in the angular (radial) directions are performed numerically (analytically), and the radial integration contributes a normalization constant $\eta(D) = \sqrt{D}$ in $S_{f,L2}^{\rm MC}(E)$ in Eq. (9). With this procedure, the MC sensitivity in Eq. (9) closely agrees with the sensitivity (5) of the progenitor Hessian PDFs, as demonstrated in the Appendix.

In this article, we primarily compare the Hessian sensitivities, which are simpler to compute. In the next subsection we illustrate the meanings of the L_2 sensitivities in the Hessian and MC frameworks using a toy example.

C. A one-dimensional example

Let us take f in Eqs. (5) and (9) to be a PDF, $f_a(x, Q)$, at some value of $\{x, Q\}$. The minimum χ_0^2 of the total χ^2 is obtained at $f = f_0$. In the Hessian representation, we have $\overrightarrow{\nabla} f = \delta_{\rm H} f \hat{e}_f$, with $\delta_{\rm H} f$ being the one-sigma uncertainty on the PDF f, and $\hat{e}_f$ the unit vector along $\overrightarrow{\nabla} f$ [1,2]. Similarly, $\overrightarrow{\nabla} \chi_E^2 \cdot \hat{e}_f = \partial \chi_E^2 / \partial f$, and hence $S_{f,L_2}^H = \delta_{\rm H} \chi_E^2(\hat{e}_f)$ is the variation of χ_E^2 from the best-fit $\chi_{E,0}^2$ along direction $\hat{e}_f$. We can approximate S_{f,L_2}^H in Eq. (5) using a symmetric finite-difference derivative for $\Delta \chi_E^2(f) \equiv \chi_E^2(f) - \chi_{E,0}^2$:
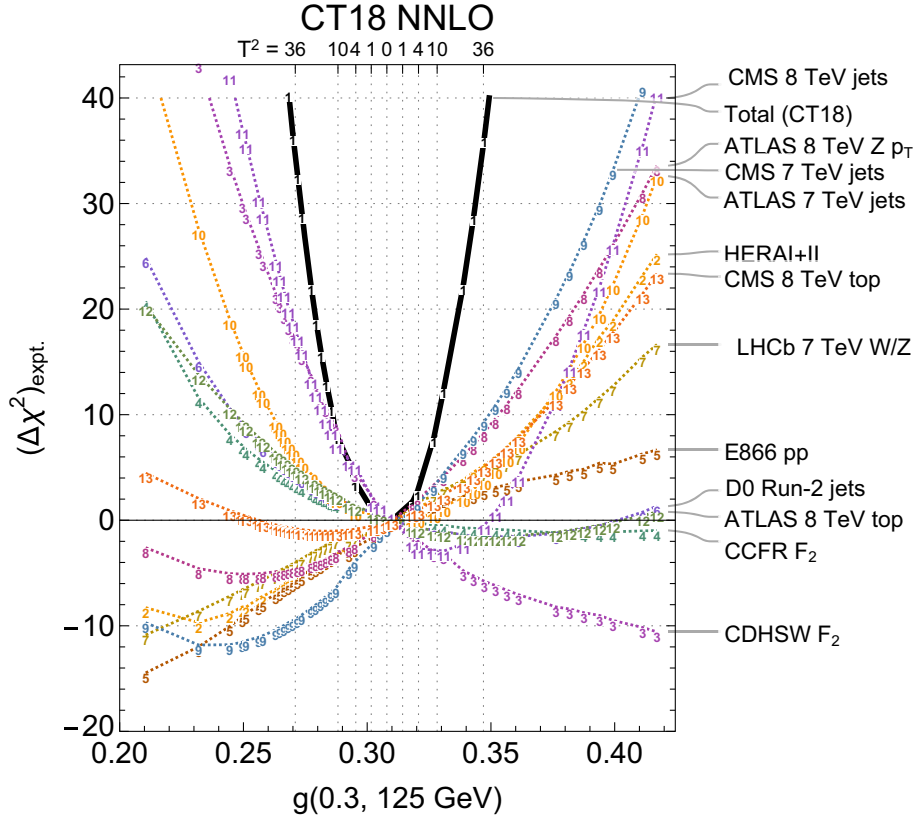
FIG. 1.   A Lagrange multiplier scan for the gluon PDF, $g(0.3, 125 \text{ GeV})$, in the CT18 NNLO analysis. The tolerance T^2, computed from the total $\Delta\chi^2$ curve, is displayed on the top of the plot. The total chi-squared distribution corresponds to the thick black curve. Each curve is marked by a unique numerical ID.

$$S_{f,L_2}^H = \left.\frac{\partial\chi_E^2(f)}{\partial f}\right|_{f=f_0} \delta_H f,$$

$$\approx \frac{\Delta\chi_E^2(f+\delta_H f) - \Delta\chi_E^2(f-\delta_H f)}{2}. \quad (10)$$

Taking $f = g(0.3, 125 \text{ GeV})$ (the gluon PDF) from CT18 NNLO as an example, in Fig. 1 we plot $\Delta\chi^2(f)$ versus f for several leading experimental datasets and for all datasets together. We perform a series of fits with a condition $f = f_i$ for a set of discrete f_i and interpolate $\chi^2(f)$ in between the f_i values. In this LM scan, we find that the minimum of the total χ^2 corresponds to $f = f_0 \approx 0.31$. This minimum optimizes the total χ^2 among competing pulls of the individual experiments.

The total χ^2 enters the likelihood probability,

$$P(f) \propto e^{-(\chi^2(f)-\chi_0^2)/(2T^2)}, \quad (11)$$

which also depends on the chosen tolerance T^2. The latter in turn fixes the value of $\delta_H f$, as illustrated by the vertical dashed lines in Fig. 1. The tolerance should be such that the linear approximation dominates for the leading χ_E^2 curves in the interval $-\delta_H f \le f - f_0 \le +\delta_H f$. In this example, we choose $T^2 = 10$.

In the MC approach, we first generate N_{rep} instances of f by stochastically sampling them according to $P(f)$. Then, we compute the L_2 sensitivity through Eq. (9), where $\eta(D) = 1$ in one dimension.

Figure 2 shows histograms of the Hessian and MC L_2 sensitivities computed for the χ_E^2 curves in Fig. 1. The histograms agree with one another, confirming that both the Hessian and MC methods are compatible in the neighborhood of the minimum. The trend of the MC histogram does not critically depend on the number of replicas, N_{rep}, as long as $N_{\text{rep}} > 100$. The experiments "CDHSW F_2" and "CMS 7 TeV jets," which have the largest negative and positive $\partial\Delta\chi_E^2(f)/\partial f$ at $f = f_0$ in Fig. 1, contribute to S_{f,L_2}^H values of about -5 and 4.5 in Fig. 2, respectively.

On the other hand, since the first derivative of the total χ^2 vanishes at the global minimum, the sum of S_{f,L_2}^H over all experiments must be zero within uncertainties. For both the Hessian and MC representations, we find in this example that

$$\sum_E S_{f,L_2} \ll T^2 < \sum_E |S_{f,L_2}|, \quad (12)$$
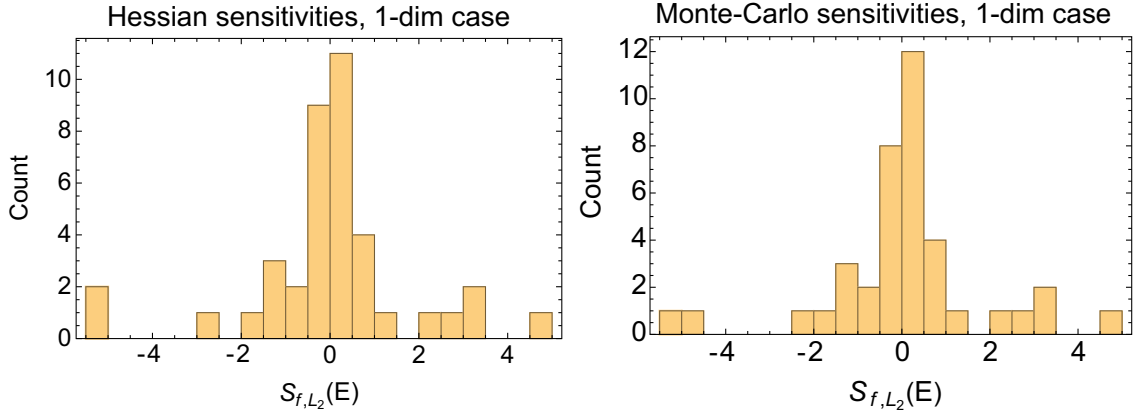
with $\sum S_{f,L_2}^H \lesssim \sum S_{f,L_2}^{MC}$.

FIG. 2. The S_{f,L_2} counts for the N_E experiments for (a) the Hessian case and (b) $N_{rep} = 1000$ MC replicas generated from the distribution of the total χ^2 from the LM scan.

In fact, an informative validation test of the Hessian approximation also in a D-dimensional case consists in checking that Eq. (12) is satisfied. The equation reflects the existence of the global minimum of the total χ^2 for any dimensionality of the fit. In our studies, we have observed that a non-negligible number of published ensembles in the LHAPDF library do not automatically satisfy this condition, which can happen with older PDF sets or poorly

constrained eigenvector sets that don't comply with the stated T^2, especially when T^2 is of order 10 or less. In these cases, the $\sum S_{f,L_2}^H$ estimates may be biased; one can correct the deficient eigenvector sets by rescaling their displacements along the respective EV directions, as summarized in the appendix of the CT-CJ NLO comparative study [19], in which such deviation was first observed and corrected. We apply the test in our study as well to validate the accuracy of
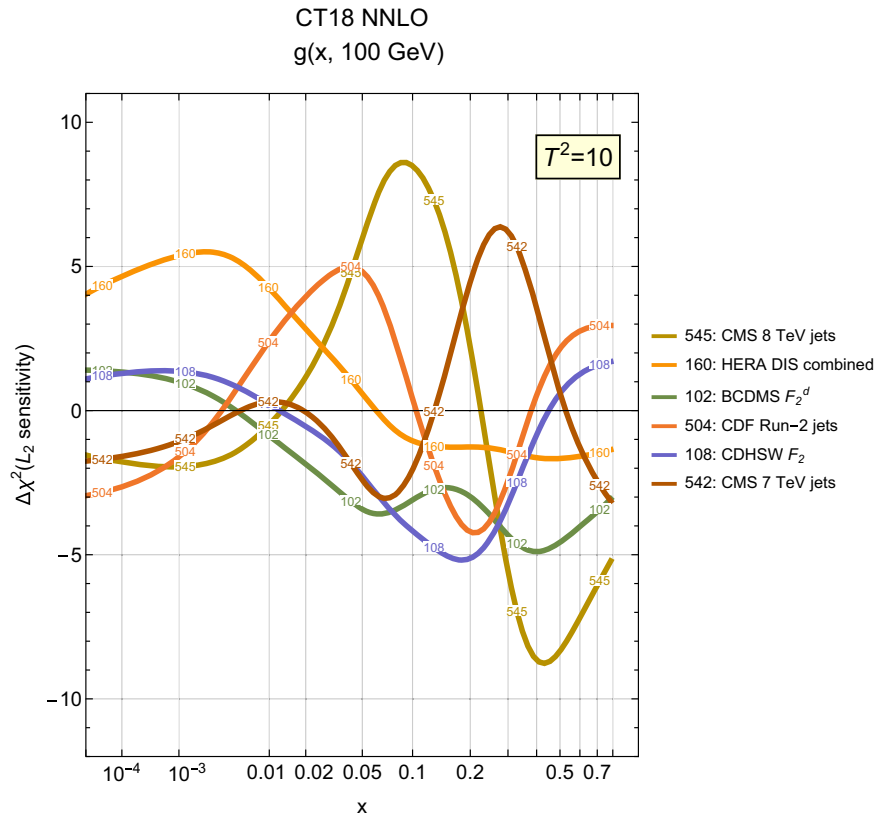

FIG. 3. The L_2 sensitivity to the CT18 NNLO gluon distribution at $Q = 100$ GeV of the six experiments with the greatest pulls, taking $T^2 = 10$. Each curve is marked by the numerical ID of the corresponding experiment provided in Tables II–IV.
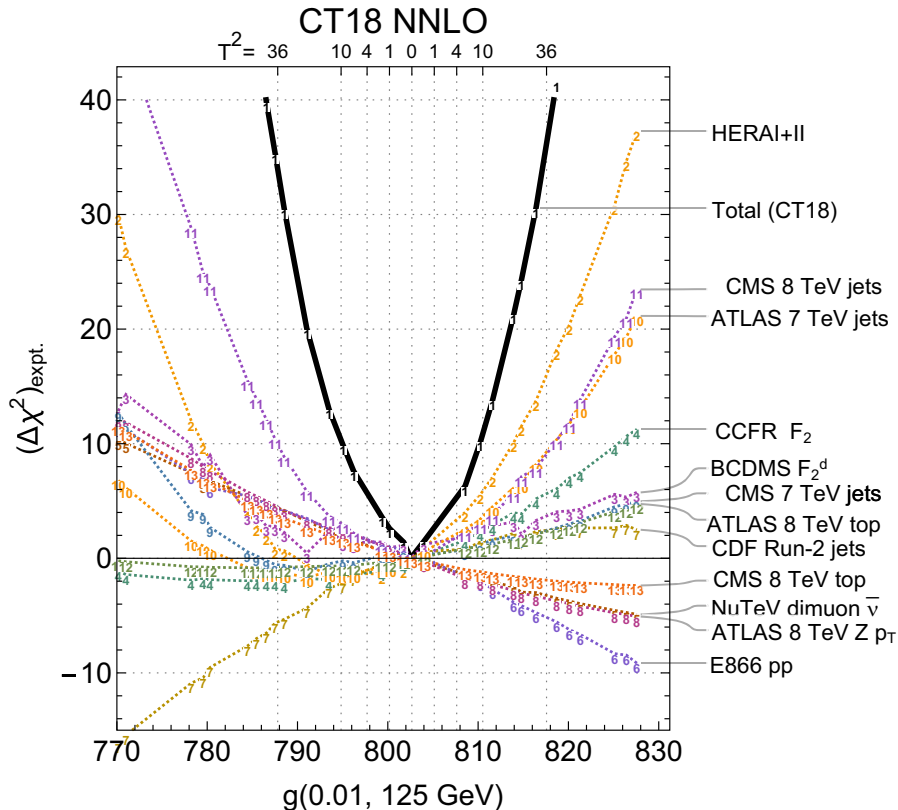
FIG. 4.　A LM scan for the gluon, $g(x = 0.01, Q = 125 \text{ GeV})$, in the CT18 NNLO fit.

the examined Hessian PDFs. The plots of the summed sensitivities can be viewed on our website [30].

D. How to interpret L_2 sensitivities

This article presents the L_2 sensitivities in two forms, chosen to illustrate either the leading sensitivities of the experiments to a given PDF or the sensitivities of a given experiment to a collection of PDF flavors or PDF combinations.

1. Cumulative sensitivities to PDF flavors

As an example from the first category, Fig. 3 shows the L_2 sensitivities to the gluon distribution $g(x, Q)$ for CT18 NNLO, as a function of the partonic momentum fraction x at a Q value of 100 GeV. Only the L_2 sensitivities from the six most significant experiments are plotted, for purposes of clarity.[1]

One way to understand more intuitively the meaning of the L_2 sensitivity is to compare it to the more familiar LM scans, either the already discussed case of the LM scan on the high-x gluon, $g(0.3, 125 \text{ GeV})$, in Fig. 1 or an analogous scan on $g(0.01, 125 \text{ GeV})$ as relevant for $gg \rightarrow$ Higgs

production, shown in Fig. 4. In a LM scan, the strength of the constraint provided by a dataset determines the narrowness of its corresponding χ_E^2 parabola. For example, in Fig. 4, the CMS experiment on jet production at 8 TeV (curve 11) imposes such a constraint, similar in magnitude to that provided by the HERA I + II data. The central gluon at this x agrees with the value preferred by the CMS 8 TeV jet data, so there is no pull from these data. In contrast, the LM scan at $x = 0.3$, in Fig. 1, indicates that the CMS 8 TeV jet data prefers a gluon value of 0.33, larger than the best-fit value of 0.31. These features of the CMS 8 TeV jet data at $x = 0.01$ and $x = 0.3$ can be seen in the L_2 sensitivity plot of Fig. 3, as explained below.[2]

At either x value, a number of experiments prefer somewhat different values of the gluon than at the best fit. We can quantify these pulls on the gluon by computing the change, $\Delta\chi_E^2$, for each experiment E when the considered PDF increases by one standard deviation from the total χ_0^2, which in the following comparisons was chosen to correspond to $\Delta\chi^2 = 10$. We note that $\Delta\chi_E^2$ were defined above Eq. (10) in Sec. II C. The L_2 sensitivity $S_{f,L2}$ estimates these $\Delta\chi_E^2$ in the linear approximation for the whole range of x. The magnitude of $S_{f,L2}$ depends on the

---

[1]On the companion website [30], such plots can show 4, 6, or 8 most sensitive experiments, or alternatively all experiments that have $|S_{f,L2}| > 3$ in some range of x.

[2]Note that sensitivities change very slowly with the scale Q, as illustrated in Sec. IV.
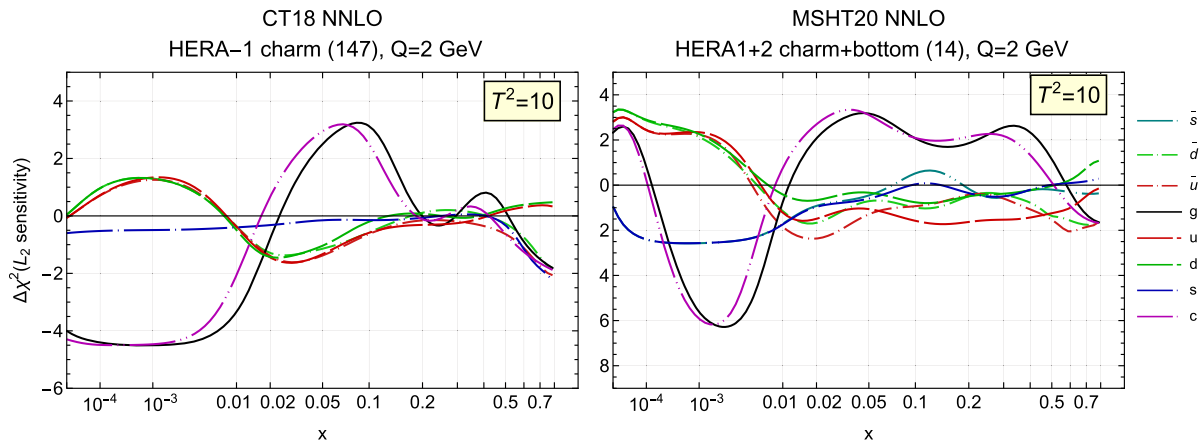
FIG. 5.   Sensitivities for the HERA I charm data set in the CT18 NNLO fit and HERA I+II charm+bottom data set in the MSHT20 NNLO fit at $Q = 2$ GeV .

correlation of the χ_E^2 with a particular PDF, the constraining power of the data on that PDF, and the difference between the PDF preferred by that dataset and the PDF determined by the full fit. A positive value of $S_{f,L2}$ indicates the preference for a lower value of the PDF at the specified x and Q, and vice versa.

From Fig. 3, we conclude following this rule that the HERA I + II deep inelastic scattering (DIS) data (ID = 160) has the largest sensitivity at $x < 0.01$, preferring a lower gluon here than the CDF and CMS jet data sets (504, 542, 545). The CMS 8 TeV jet dataset has little sensitivity at $x = 0.01$, even though it is well constraining at this $\{x, Q\}$ point, in accord with the behavior in the LM scan in Fig. 4 noted above. By contrast, we see that at $x = 0.3$ the sensitivity is negative, so that this dataset prefers a larger gluon at this x, as noted in Fig. 1. The CMS 7 and 8 TeV inclusive jet data have the largest $|L_2|$ sensitivity at $x > 0.05$.[3] We note that no single experiment dominates over the entire x range, and the L_2 sensitivity for many experiments will switch sign as a function of x.

Again, these trends are consistent with the LM scans at two fixed x values shown in Figs. 1 and 4. If an experiment has a major influence on the best-fit PDF at a particular x value, it will normally have a large absolute value of S_{f,L_2}. But, if the reference PDF value already agrees with that preferred by the dataset, it may have only a relatively small value of S_{f,L_2} (since the gradient of the χ^2 is small).[4]

It is also possible that an experiment is constraining enough for a particular PDF over a wide x region and has a low value of S_{f,L_2} across that entire x region, as the PDF is forced to be close to that preferred by that experiment. In reality, this rarely occurs: the combined weight of the other experiments tends to provide a counterconstraint to the dominant experiment. See, for example, that the HERA DIS combined data in Fig. 3—the most dominant dataset in all global PDF fits—strongly prefers a smaller gluon distribution at $x < 0.02$, but even then is counteracted by the CDF and CMS inclusive jet data.

The sum of the S_{f,L_2} values for all experiments for a given kinematic point should be close to zero, as the pulls have to balance out to produce the central PDF at that point. One can also sum up all of the positive values of S_{f,L_2}, as well as all of the negative values, for each parton x value. The sums (both positive and negative) tend to be relatively flat as a function of x, and are roughly equal to the number of experiments; i.e., each experiment contributes on the order of 1 to the sum. There is no obvious correlation between the value of the sum and the size of the PDF uncertainty at that point.

2. Cumulative sensitivities of individual datasets

Figure 5 is an example of the second form of comparisons, in which sensitivities for the indicated PDF flavors are plotted for a given dataset indicated in the plot label. In the figure, we compare the sensitivities of the HERA I combined charm dataset [31] included in the CT18 NNLO analysis in the left subpanel, and the combined HERA I + II charm + bottom dataset [32] included in the MSHT20 NNLO analysis in the right one. While the more recent HERA I + II dataset [32] covers an extended kinematic range and has smaller uncertainties, the CTEQ-TEA group found it difficult to accommodate these data, with its χ^2/N_{pts} remaining high (> 1.7) under a variety of explored assumptions [33]. The HERA I + II dataset is included in

---

[3]It is interesting to note that, at high x, these jet data at the two energies have large L_2 sensitivity values and opposite signs. Even though the two measurements are carried out by the same experiment at very similar energies, there is a tension between them, representative of the situations encountered in a global PDF fit. Reference [9] discusses consequences of such tensions for the interpretation of PDF uncertainties.

[4]This agreement may be an accident or an indication of the constraining power of that experiment.

the MSHT20 analysis, and it is interesting to compare its impact with that of the HERA I charm data in the CT18 NNLO fit. We find that both datasets prefer higher (lower) gluon and charm PDFs at $10^{-4} < x < 0.01$ ($0.02 < x < 0.2$). In the MSHT20 NNLO case, the preferences for a lower gluon extends to higher x of up to 0.5. In both PDF fits, these heavy-quark data prefer lower u and d (anti)quark PDFs at $x < 0.01$. There is mild preference for a higher strangeness PDF at $x < 0.1$. This preference is more pronounced in the MSHT20 case. Overall, according to the sensitivities that stay within a few units, the HERA charm data (at Q of tens of GeV) impose moderate constraints on the gluon and other PDFs.

E. L_2 sensitivity and the likelihood-ratio test

The likelihood-ratio test is a classical Bayesian test, together with the closely related Lagrange multiplier test and Wald test, that discriminates between two theoretical hypotheses, T_1 and T_2, based on their agreement with a set of observational data D. In the context of PDF fits [9,34], it is more common to formulate the likelihood-ratio test as a comparison of log-likelihood functions χ^2 for two PDF models, T_1 and T_2, related to the augmented likelihood as $P(D|T_i) = \text{const} \cdot \exp(-\chi^2(D, T_i)/2)$. When there is no strong prior preference for either T_1 or T_2, but the data D strongly favor one of them, the ratio of posterior probabilities is dominated by the ratio of the likelihoods that is related to the difference of χ^2 for T_1 and T_2:

$$\left(\frac{P(T_2)}{P(T_1)}\right)_{\text{posterior}} = \frac{P(D|T_2)}{P(D|T_1)} \left(\frac{P(T_2)}{P(T_1)}\right)_{\text{prior}}, \quad (13)$$

where

$$\frac{P(D|T_2)}{P(D|T_1)} = \exp\left(-\frac{\chi^2(D, T_2) - \chi^2(D, T_1)}{2}\right). \quad (14)$$

Based on this ratio, the PDFs rendering the lowest χ^2 are the most likely ones according to the empirical data. The PDFs with a low, but not the lowest, χ^2 can

be acceptable with some probability determined by the tolerance prescription. A LM scan examines the change in χ^2 as a function of a PDF parameter or PDF-dependent observable, hence it realizes the ratio test between the best-fit PDF and nearby PDF solutions. The L_2 sensitivity serves the same purpose. These tests do not need to know how the PDFs are found. Section III B reviews the practical implementations of the likelihoods in the fits by three groups.

III. PARTICULARS OF THE COMPARED FITS

A. Overview of the global analyses

One of the goals of this study is to explore how the constraints on PDFs emerge as one successively adds new experiments into the PDF analysis. We compare sensitivities in two categories of PDFs:

(1) Nonglobal fits, which include a small(er) number of sensitive experiments. Examples include the ATLASpdf21 and PDF4LHC21 benchmark reduced fits.

(2) Global fits, which include several tens of datasets of varied sensitivity. The examples considered here include NNLO ensembles CT18, CT18As, CT18As_Lat, MSHT20 as well as the approximate N3LO ensemble MSHT20aN3LO.

This section provides the background for the comparisons of the L_2 sensitivities that will be presented in the following sections. The reader broadly familiar with the selection of experiments and methodologies of three groups can skip much of this section. The ensuing discussions of the L_2 sensitivities will extensively refer to the selections of the experimental datasets in the fits. Table I lists the datasets included in the ATLASpdf21 analysis, while Tables II–IV list the datasets included in the CT18 and MSHT fits. Next to each dataset in the tables, we list its numerical ID adopted to mark the corresponding L_2 sensitivity curves in the figures. We will see that, while there are large differences among these PDF analyses, the L_2 sensitivity elucidates their comparisons as a common metric.

The next three subsections briefly review the selections of experiments and theoretical computations in ATLASpdf21,

TABLE I. Summary of all the input datasets considered in the ATLASpdf21 fit.

ID	Data set	$\sqrt{s}$ [TeV]	Luminosity [fb^{-1}]	Decay channel	Observables entering the fit
160	HERA inclusive DIS [35]	Varied	Varied		Reduced cross sections
68	Inclusive $W, Z/\gamma^*$ [36]	7	4.6	e, μ combined	η_ℓ (W), y_Z (Z)
89	Inclusive Z/γ^* [37]	8	20.2	e, μ combined	$\cos\theta^*$ in bins of $y_{\ell\ell}$, $m_{\ell\ell}$
86	Inclusive W [38]	8	20.2	μ	η_μ
56	$W^\pm +$ jets [39]	8	20.2	e	p_T^W
	$Z +$ jets [40]	8	20.2	e	p_T^{jet} in bins of $|y^{\text{jet}}|$
7	$t\bar{t}$ [41,42]	8	20.2	lepton + jets, dilepton	$m_{t\bar{t}}$, p_T^t, $y_{t\bar{t}}$
8	$t\bar{t}$ [43]	13	36	lepton + jets	$m_{t\bar{t}}$, p_T^t, y_t, $y_{t\bar{t}}^b$
9	Inclusive isolated γ [44]	8, 13	20.2, 3.2	$\cdots$	E_T^γ in bins of η^γ
10	Inclusive jets [45]	8	20.2	$\cdots$	p_T^{jet} in bins of $|y^{\text{jet}}|$

CTEQ-TEA, and MSHT global fits, then moving on to the comparisons of CT18 and MSHT20 datasets in Sec. III A 4, and then to additional discussion of the PDF4LHC21 reduced fits in Sec. III A 5. While the CT18 and MSHT20 NNLO global fits share many key data sets, there are important differences among them. To compare the CT18, MSHT20, and NNPDF3.1 methodologies using (nearly) the same datasets, reduced fits were performed by the PDF4LHC group in the course of 2021 benchmarking study [18] by including only 12 shared datasets indicated in the rightmost column in Tables II–IV. The findings from these comparisons guided the construction of the PDF4LHC21 combination of the PDFs from the three NNLO global analyses.

The value of χ_E^2 generally depends on the approximations made in the likelihood, or "χ^2 definition." The specific implementations are reviewed in Sec. III B. The tolerance conventions for the published PDFs, which determine both the size of PDF uncertainties and L_2 sensitivities, are compared for the three groups in Sec. III C.

1. Summary of ATLAS fits

ATLAS PDF fits concentrate on the impact of ATLAS data on PDFs. However, it is not possible to make an accurate PDF fit to ATLAS data alone. The HERA DIS combined data [35] are used as the backbone of the ATLAS PDF fits, to which ATLAS data are added. The HERA experiments cover a very broad range of Q^2, the absolute four-momentum transfer squared, from near 1 GeV^2 to above 10^4 GeV^2, and of Bjorken x from ~0.6 down to 10^{-4}.

ATLASpdf21 [20]—the most comprehensive fit from this series—uses the HERA inclusive DIS data and a broad variety of ATLAS data, while accounting for correlations between the various ATLAS measurements. We also consider several intermediate ATLAS fits leading to ATLASpdf21. Table I lists these fits together with the included datasets, their numerical IDs, center-of-mass energies, luminosities, decay channels, and observables entering each fit.

The L_2 sensitivities for the intermediate fits can be used as a pedagogical example of the effect of adding in sensitive datasets: all these PDFs are obtained starting from the PDF set HERAPDF2.0 that was determined staring with the HERA data alone [35] and by successively adding new ATLAS datasets. First in this series is the ATLASepWZ16 PDF set [36], in which the ATLAS precision measurements of the inclusive differential $W^\pm$ and Z/γ^* boson cross sections at 7 TeV were added to HERA data. This PDF set improved on the HERAPDF2.0 set in various respects. First, the strange content of the sea could be fitted rather than assumed to be a fixed fraction of the light sea. Indeed, the strange sea was found to be enhanced at low x, $x \lesssim 0.05$ compared to previous determinations. Second, the accuracy of the valence quark distributions was considerably improved.

Further improvement was achieved by adding $t\bar{t}$ differential cross sections at 8 TeV from both the lepton + jets and the dilepton channels. This fit was called ATLASepWZtop18 [64]. In the lepton + jets channel, the mass of the $t\bar{t}$ pair, m_{tt}, and the average top-quark transverse momentum, p_T^t, were fitted simultaneously. In the dilepton channel, the rapidity of the $t\bar{t}$ pair was fitted. Care was taken to include correlated systematic and statistical uncertainties. This fit improved the uncertainties of the high-x gluon.

The final ATLASpdf21 ensemble discussed here was obtained by including all the previous ATLAS datasets mentioned above and adding to them more data constraining the valence quarks and the flavor of the sea (from ATLAS 8 TeV inclusive W and Z data), more data constraining the high-x strange sea and the gluon PDFs (from ATLAS 8 TeV W and Z boson data + jets[5]), and more data constraining the gluon (from ATLAS 8 TeV data on inclusive jets, ATLAS 13 TeV data on top-antitop distributions, and ratios of ATLAS 13 and 8 TeV direct photon data).

The increase in the number of datasets from HERAPDF2.0 to ATLASpdf21 has facilitated an increase in the freedom of the parametrizations from 14 to 21 parameters (and in the corresponding number of eigenvectors). Extra freedom has been added in the high-x valence and gluon PDFs and in the sea parametrization at low x, whereby the low x $\bar{u}$, $\bar{d}$, and $\bar{s}$ were all independently parametrized. Extra variations due to model assumptions and additional parameters were also considered. Tension between datasets has also led to a consideration of appropriate χ^2 tolerance as $T = 3$ as well as $T = 1$. See Sec. III C for an explanation of the choice of tolerance for the present L_2 study.

Uniquely, the ATLASpdf21 fit included full information both on correlated systematic sources of uncertainty of all datasets and on the correlations between the ATLAS datasets, in contrast to the global fits. The largest sources of such correlations among the data sets come from the measurements involving jets: not only from inclusive jet production but also boson + jets and $t\bar{t}$ in the lepton + jets channel. The L_2 sensitivities in our comparisons account for the correlations among the experimental datasets.

The theoretical predictions for the ATLASpdf21 fit were computed at NNLO in QCD using programs DYNNLO [66,67], FEWZ [68–70], NNLOJET [71,72] and results from [73–76], and at NLO in electroweak theory using DYNNLO, FEWZ, SHERPA and results from [77–79]. A full description of the theoretical treatment is given in Ref. [20].

2. Summary of CT18 NNLO

The CT18 NNLO analysis [17] constituted a major new release of the CTEQ-TEA family of PDF studies, having

---

[5]In fact, the ATLASepWZjets20 [65] PDF fit also uses these data, but it is not further discussed here.

TABLE II. DIS and SIDIS datasets in the CT18 and MSHT20 ensembles. For each dataset, we indicate the publication reference, number of data points, and the numerical ID in the figures.

Data set	CT18 NNLO			MSHT20 NNLO			In the PDF4LHC21 reduced fit?
	Ref.	N_{pts}	ID	Ref.	N_{pts}	ID	
BCDMS F_2^p	[46]	337	101	[46]	163	1	Yes
BCDMS F_2^d	[46]	250	102	[46]	151	2	Yes
NMC F_2^p				[47]	123	3	
NMC F_2^d				[47]	123	4	
NMC ratio	[47]	123	104	[48]	148	11	Yes
SLAC ep F_2				[49,50]	37	5	
SLAC ed F_2				[49,50]	38	6	
E665 μd F_2				[51]	53	7	
E665 μp F_2				[51]	53	8	
HERAI + II DIS combined	[35]	1120	160	[35]	1185	160	Yes
HERA e^+p CC				[52]	39	22	
HERA e^-p CC				[52]	42	23	
HERA e^+p NC 820 GeV				[52]	75	24	
HERA e^+p NC 920 GeV				[52]	402	25	
HERA e^-p NC 460 GeV				[52]	209	26	
HERA e^-p NC 575 GeV				[52]	259	27	
HERA e^-p NC 920 GeV				[52]	159	28	
HERA charm + bottom				[32]	79	14	
HERA I	[31]	47	147				
H1 bottom	[53]	10	145				
H1 F_L	[54]	9	169				
NMC/BCDMS/SLAC/HERA F_L				[46,47,50,54–56]	57	15	
CDHSW F_2^p	[57]	85	108				
CDHSW $x_BF_3^p$	[57]	96	109				
CCFR F_2^p	[58]	69	110				
CCFR $x_BF_3^p$	[59]	86	111				
CHORUS νN F_2				[60]	42	19	
CHORUS νN x_BF_3				[60]	28	20	
NuTeV νN F_2				[61]	53	9	
NuTeV νN x_BF_3				[61]	42	10	
CCFR dimuon combined				[62]	86	16	
ν_μ	[62]	40	126				
$\bar\nu_\mu$	[62]	38	127				
NuTeV dimuon combined		81	593	[62]	84	17	Yes
ν_μ	[63]	38	124				
$\bar\nu_\mu$	[63]	33	125				

included ~700 LHC data points on top of the baseline fit in the previous main release, CT14HERA2 [80]. A detailed discussion of the theory used in CT18, selected datasets, and other statistical or methodological choices is presented in Ref. [17]; these aspects were further summarized for the purpose of the recent PDF4LHC21 benchmarking study in Ref. [18].

The strategy of the CT18 analysis was to first examine a large group of datasets using preliminary fits and fast Hessian techniques in order to select an ensemble of constraining and maximally consistent datasets. The LM scans and L_2 sensitivities were extensively employed to identify such datasets, as documented in Ref. [17] and on

the CT18 website [81]. The published CT18 NNLO fit was performed to this final data set. The selected experiments are shown in Tables II–IV and include neutral-current and charged-current DIS, as well as production of vector bosons, jets, and top quark pairs.

When validating the fit, the mutual agreement of the datasets was examined based on the strong goodness-of-fit criteria [9], in addition to requiring a good total χ^2. Out of the newly included LHC datasets, the ATLAS 7 TeV W, Z production dataset [36] (ID = 248) was found to be both precise and showing a tension with the NuTeV dimuon and HERA DIS datasets as a result of its preference for a larger strangeness PDF at $x \sim 0.02$. This dataset was included in

the alternative fits, CT18A and CT18Z, and not in the nominal CT18.

The size and complexity of the CT18 analysis made it worthwhile to consider a number of variations on the assumptions within the analysis. Together with the default CT18, several complementary PDF ensembles were released to quantify the impact of these assumptions, including different selections in the fitted datasets (e.g., CT18A, which fitted the 2016 7 TeV ATLAS inclusive W, Z production data); alternative factorization scales (e.g., CT18X, which assigned a different x-dependent scale, μ_{DIS}, to DIS data to mimic the effects of low-x resummation); and an amalgamation of these choices (CT18Z, which also took a slightly enlarged charm mass, $m_c = 1.4$ GeV, relative to the default of $m_c = 1.3$ GeV used in the other CT18 fits).

In subsequent years, a number of follow-up studies have expanded the CT18 framework by investigating various physics issues relevant for both high- and medium-energy datasets fitted in CT. These have included the following: the introduction of an explicit photon PDF and associated electroweak corrections in the CT18QED analysis for a proton [82] and neutron [83]; evaluation of the S-ACOT-χ NNLO theory with heavy-quark mass effects with phenomenology for the Electron-Ion Collider (EIC) and neutrino DIS [84], including, in a later study, scattering of cosmic neutrinos at ultra-high energies [85]; an investigation of the PDF impact of light nuclear corrections, in particular, those connected to nucleon off-shellness in deuterium [19]; an analysis of the high-x PDF behavior in light of power-counting arguments [86]; a study of fitted charm in light of the full CT18 dataset with a range of nonperturbative models allowing $c \neq \bar{c}$ [87]; inclusion of high-x lattice QCD constraints for the strange quark and antiquark PDFs in an analysis allowing for $s \neq \bar{s}$ [22]; and an investigation of the impact on the PDFs by the post-CT18 data on vector boson production at the LHC at 8 and 13 TeV [88].

In CT18 and follow-up studies, the perturbative QCD theory for the LHC data is evaluated at NNLO accuracy, typically through the use of fast interpolation tables provided by FASTNLO [89–91] and APPLGRID [92] and calculated at NLO using MCFM [93–96], NLOJET++ [97], and AMCFAST [98]; these calculations are then corrected to NNLO via point-by-point NNLO/NLO K factors based on DYNNLO [66,67], FEWZ [68–70], MCFM, and NNLOJET [71,72]; an exception to this procedure applies to top-quark production, which is computed at NNLO directly using FASTNNLO [74,77,99] grids. We note that scale-choice and other theoretical uncertainties were explored in the main CT18 publication; while these were not quantified systematically in a specialized error treatment, the final CT18 uncertainty was determined to ensure coverage of variations associated with these uncertainties. The ultimate CT18 parametrization (given explicitly in Appendix C of Ref. [17]) resembles that used in CT14HERA2 and is

similarly formulated in terms of Bernstein polynomials, but with slightly more flexibility accorded to the light-quark sea. In addition, CT18 considered $\mathcal{O}(250)$ alternative parametrization forms. The nominal CT18 PDF uncertainty was also assessed to encompass the corresponding variations driven by these alternative nonperturbative forms. Like earlier studies, CT18 deployed an NNLO implementation of the S-ACOT-χ scheme to treat heavy-flavor production, partonic thresholds, and related dynamics.

Computing the L_2 sensitivities provided powerful insights into many post-CT18 studies and was an essential feature of Refs. [19,22,86]. In the current article, we present the L_2 sensitivities for CT18 NNLO in the uniform format that facilitates comparisons with ATLASpdf21 and MSHT20. We also apply the sensitivity method to two NNLO PDF fits CT18As and CT18As_Lat [22] that determined the magnitude of a nonzero strangeness asymmetry, $s_-(x, Q) \equiv s(x, Q) - \bar{s}(x, Q)$, by releasing the assumption of $s = \bar{s}$ made in the CT18 and CT18Z analyses. These fits follow the setup of the CT18A analysis and, in particular, include the high-luminosity ATLAS 7 TeV W, Z dataset 248 [36]. They also examine agreement with the E906/SeaQuest dataset on the Drell-Yan pd/pp ratio without fitting it. To constrain the $s_-(x, Q)$ combination at $0.3 < x < 0.8$, where no relevant experimental sensitivity currently exists, the CT18As_Lat analysis includes constraints from quasi-PDF matrix elements (extrapolated to physical pion mass) computed in lattice QCD [100]. The lattice QCD input significantly reduces the allowed magnitude of $s_-(x, Q)/s_+(x, Q)$ at $x \to 1$, which otherwise can be very large (approaching 100%) if only the extrapolations of experimental constraints are included. In the CT18As_Lat analysis, the lattice QCD input is implemented with the help of Lagrange multipliers. The fitting code reports its contribution to χ^2 in one category with the contributions from the normalization shifts for BCDMS, CDHSW, and CCFR data on the DIS structure functions (ID = 101, 102, 108, 109, 110, 111). We compute $S_{f,L2}$ for this "Lattice + DIS normalizations" category under ID = 701, keeping in mind that the lattice constraints dominate between the two.[6]

3. Summary of MSHT20 NNLO and approximate N3LO

The baseline MSHT20 PDF sets [21] represent the latest in the MRST/MSTW/MMHT line of PDF fits, with substantial improvements made on top of the previous MMHT2014 study [101] on the experimental, methodological, and theoretical fronts.

On the experimental side, these include additional LHC datasets on vector boson, top and jet production among others, as well as somewhat augmented and different selections of the fixed-target and Tevatron datasets

---

[6]The DIS normalizations are mildly correlated with the gluon PDF in the fixed-target region.

TABLE III. Vector boson production datasets in the CT18 and MSHT20 ensembles. The ATLAS 7 TeV 35 pb^{-1} W, Z dataset, marked by ‡‡, is replaced by the high-luminosity dataset (4.6 fb^{-1}), marked by ‡, in the CT18As and CT18As_Lat fits. The E906/SeaQuest Drell-Yan ratio (marked by $\star$) is included in the CT18As and CT18As_Lat comparisons, but not fitted.

Data set	CT18 NNLO			MSHT20 NNLO			In the PDF4LHC21 reduced fit?
	Ref.	N_{pts}	ID	Ref.	N_{pts}	ID	
ATLAS 7 TeV 35 pb^{-1} W, Z cross sec., A_{ch}	[112]	41	268‡‡	[112]	30	52	Yes
ATLAS 7 TeV high-mass Drell-Yan				[113]	13	58	
ATLAS 7 TeV high precision W, Z	[36]	34	248‡	[36]	61	68	
ATLAS 8 TeV high-mass Drell-Yan				[114]	48	82	
ATLAS 8 TeV W				[38]	22	86	
ATLAS 8 TeV Z p_T	[115]	27	253	[115]	104	71	
ATLAS 8 TeV double differential Z				[37]	59	89	
CDF Run-1 lepton A_{ch}, $p_{T\ell} > 25$ GeV	[116]	11	225				
CDF Run-2 electron A_{ch}, $p_{T\ell} > 25$ GeV	[117]	11	227				
CDF Run-2 W asymmetry				[118]	13	43	
CDF Run-2 Z rapidity	[119]	29	261	[119]	28	37	
CMS 7 TeV lepton asymmetry							
36 pb^{-1}				[120]	24	54	
840 pb^{-1}	[121]	11	267	[121]	11	53	Yes
4.7 fb^{-1}, muon	[122]	11	266				
CMS 7 TeV $Z \to e^+e^-$				[123]	35	57	
CMS 7 TeV double diff. Drell-Yan				[124]	132	60	
CMS 8 TeV muon asymmetry	[125]	11	249	[125]	22	64	
E605 Drell-Yan	[126]	119	201				
E866/NuSea pd/pp Drell-Yan ratio	[127]	15	203	[127]	15	13	Yes
E866/NuSea pp Drell-Yan	[128]	184	204	[129]	184	12	
E906/SeaQuest pd/pp Drell-Yan ratio	[130]	6	206$\star$				
DØ Run-2 $W \to \nu\mu$ asymmetry	[131]	9	234	[132]	10	38	
DØ Run-2 $W \to \nu e$ asymmetry	[133]	13	281	[134]	12	44	
DØ Run-2 Z rapidity	[135]	28	260	[135]	28	36	Yes
DØ Run-2 W asymmetry				[136]	14	70	
LHCb 7 TeV W asymmetry $p_{T\ell} > 20$ GeV				[137]	10	56	
LHCb 7 TeV $Z \to e^+e^-$				[138]	9	55	
LHCb 7 and 8 TeV W and Z	[139,140]	50	258	[139,140]	67	61	Yes
7 TeV	[139]	33	245	[139]			
8 TeV	[140]	17	246				
LHCb 8 TeV $Z \to ee$	[141]	34	250	[141]	17	62	Yes

relative to the previous MMHT2014 PDFs, as can be seen from Tables II–IV. Simultaneously with the increase in the number and precision of included experimental constraints on the PDFs, methodological improvements were also made. Foremost among these was an extension of the PDF parametrization at the input scale $Q_0 = 1$ GeV. In MSHT the PDFs are parametrized in terms of orthogonal Chebyshev polynomials, with the number of parameters now increased from 4 to 6 for each of the input PDF flavor combinations. This followed previous work [102], in which it was shown that this would allow a fit of the PDFs to data of better than 1% accuracy over the entire x range, provided the data and theory allowed. As a result, the number of PDF parameters has increased from 37 in

MMHT to 52 in MSHT, with a consequent increase in the number of eigenvectors from 25 to 32 to allow a more detailed reflection of the PDF uncertainties. Further details of these improvements are given in [103–106] as well as the MSHT20 paper [21]. The MSHT PDFs also allow a nonzero strangeness asymmetry by default, although in this case the constraints are weaker, so fewer parameters are used, while the charm PDF is generated perturbatively.

In our baseline MSHT20 study and subsequent work, the theoretical predictions for hadronic cross sections are made using NNLO QCD theory via grids provided at NLO by FASTNLO [89–91] and APPLGRID [92] using calculations from MCFM [93–96] and NLOJET++ [97]. These are supplemented by NNLO/NLO K factors from a variety of

TABLE IV. Inclusive jet, boson + jet, and top-quark pair production datasets in the CT18 and MSHT20 ensembles.

Dataset	CT18 NNLO			MSHT20 NNLO			In the PDF4LHC21 reduced fit?
	Ref.	N_{pts}	ID	Ref.	N_{pts}	ID	
ATLAS 7 TeV incl. jets	[142]	140	544	[142]	140	66	
CDF Run-2 $p\bar{p}$ incl. jets	[143]	72	504	[144]	76	35	
CMS 2.76 TeV incl. jets				[145]	81	87	
CMS 7 TeV incl. jets	[146]	158	542	[146]	158	69	
CMS 8 TeV incl. jets	[147]	185	545	[147]	174	73	Yes
DØ Run-2 $p\bar{p}$ incl. jets	[148]	110	514	[149]	110	48	
ATLAS 8 TeV W + jets				[39]	30	83	
CMS 7 TeV $W + c$				[150]	10	67	
ATLAS 8 TeV single diff $t\bar{t}$	[41]	15	580	[41]	25	74	
ATLAS 8 TeV single diff $t\bar{t}$ dilepton				[42]	5	81	
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$				[151,152]	17	59	
CMS 8 TeV double differential $t\bar{t}$	[153]	16	573	[153]	15	84	
CMS 8 TeV single differential $t\bar{t}$ ($y_{t\bar{t}}$)				[154]	9	88	

sources including DYNNLO [66,67], FEWZ [68–70], MCFM, N_{jetti} [73,107], and NNLOJET [71,72]; with the K factors smoothed and an uncertainty applied to account for this—as described further in [21]. Top production at the LHC is however computed directly at NNLO using provided FASTNNLO grids [74,77,99]. These are then supplemented by NLO electroweak corrections where relevant, as described further in [108].

A variety of follow-up studies have extended the MSHT20 investigations since its publication. A study of the dependence of the PDFs on the strong coupling α_s and the heavy-quark masses was presented in [109], while an update to include QED effects in the DGLAP evolution and a corresponding photon PDF was presented in [108], the latter representing a further theoretical improvement and building on previous work [110]. The MSHT20 NNLO PDF set was also the PDF set which contributed to the PDF4LHC21 combination [18] and the reduced fit studies leading up to it also described in [111]. Most recently, the first global PDF analysis at approximate N3LO (aN3LO) was presented in [23], extending beyond the current highest NNLO order in QCD of theoretical predictions achieved in contemporary PDF fits. This MSHT20aN3LO study includes for the first time already known N3LO ingredients, as well as the estimates of uncertainties due to the yet unavailable pieces and implicitly some higher-order contributions beyond N3LO. In particular, the splitting functions, transition matrix elements, coefficient functions and K factors for multiple processes are approximated using the exact partial N3LO results and constrained to be consistent with the wide range of already available information about this order [23]. Theoretical nuisance parameters are then used to include uncertainties from the missing pieces into the overall PDF uncertainties.

For the present study, we compute the L_2 sensitivities using the χ^2 values and MSHT20 Hessian eigenvector sets presented in the NNLO [21] and the aN3LO [23] analyses. For this study, we note minor modifications in the precise details from the above publications. In the NNLO case, a correction has been made in the treatment of photon-initiated production for the ATLAS 8 TeV double differential Z [37] data, as described in [108]. For the aN3LO fit, an incorrect application of the N3LO K factor was corrected in the final version of [23] for two Drell-Yan datasets.[7]

The MSHT datasets included in the aN3LO study are identical to the MSHT20 NNLO baseline. In [23], it was found that the addition of aN3LO theory leads in some places to significant changes and improvements in the fit quality, with evidence that in some cases this is due to the tensions between datasets being alleviated. This therefore motivates studying the pulls of the datasets in the fit in both the MSHT20 NNLO and the aN3LO fit using the L_2 sensitivity approach.

4. CT18 and MSHT20 datasets side by side

Tables II–IV list the CT18 and MSHT20 datasets side by side together with the references, numbers of points, and numerical IDs. Sometimes one PDF analysis fits several datasets independently, while the other analysis fits them as a combined dataset. There are several such examples in Tables II–IV. To facilitate the comparisons in such cases, we occasionally compute the sensitivity using

---

[7]Since the K factors are very largely decorrelated from PDF parameters, the publicly available PDFs did not require a refit. A full refit has, however, been performed for the PDFs used in this article and results in a very small change in the PDFs and a reduction in the total χ^2 by a few units.

the χ^2 for the combination in both fits and then plot the sensitivities both for the constituent and combined datasets. Such datasets include the following:

(1) HERA DIS (a combined dataset 160 in CT18; independent sets 22–28 in MSHT);

(2) HERA charm and bottom production (independent HERA-I charm and H1 bottom datasets, 147 and 145, in CT18; a combined data HERA I+II charm and bottom set 14 in MSHT);

(3) CCFR dimuon SIDIS (independent neutrino and antineutrino datasets, 126 and 127, in CT18; a combined dataset 16 in MSHT);

(4) NuTeV dimuon SIDIS (independent neutrino and antineutrino datasets, 124 and 125, as well as their *a posteriori* combination 593, in CT18; a combined dataset 17 in MSHT);

(5) LHCb W and Z production at forward rapidities (independent datasets at 7 and 8 TeV, 245 and 246, as well as their *a posteriori* combination 258, in CT18; a combined dataset 61 in MSHT).

The numbers of the data points selected by the groups from a shared dataset can be different in reflection of the trade-off between the total number of points and accuracy of the individual points and theoretical predictions. One example are the BCDMS datasets on proton and deuteron DIS structure functions, which include more points in the CT18 fit (as datasets 101 and 102) than in MSHT20 (as datasets 1 and 2). This difference, however, reflects a different presentation of the same data, with the same information and constraints encoded. In particular in MSHT20 the data are averaged over energy runs [46]. In the CT18 analysis, the BCDMS datasets have a particularly pronounced sensitivity to the large-x behavior of the up and down PDFs [86]. Of these, the extracted high-x behavior of the d-PDF can be especially influenced by model-dependent prescriptions to correct the deuteron structure function to that of an isoscalar nucleon target [19]. In the MSHT20 analysis, the BCDMS sensitivities differ from CT18 in reflection of the differences in the obtained PDFs in terms of central values and uncertainties, and from the treatment of nuclear effects.

Another example are the ATLAS 8 TeV datasets on the transverse momentum of Z boson, labeled as 253 and 71 by the two groups, respectively. The CT18 analysis selected 27 data points presented single differentially in the interval $45 \leq p_{T,Z} < 150$ GeV, where the fixed-order NNLO theory has best convergence. The MSHT analyses, on the other hand, include 104 points and additional intervals of $30 \leq p_{T,Z} < 45$ GeV and $p_{T,Z} > 150$ GeV, while using the cross sections that are double differential in $p_{T,Z}$ and y_Z in the Z mass bin. At NNLO, the Z p_T dataset demonstrates a stronger sensitivity to the gluon than in CT18, which is consistent with the larger number of data points and

extended $p_{T,Z}$ range in the MSHT20 fit. Both CT and MSHT groups choose the factorization scale to be the transverse mass of the vector boson. On the other hand, some potential differences in their theoretical cross sections (including electroweak contributions at high p_{TZ}) and prescriptions for systematic uncertainties need to be further explored. MSHT20 observe further changes when going to aN3LO, as discussed in Sec. VI.

5. PDF4LHC21 reduced fits

The PDF4LHC21 reduced fits are discussed here only for the two participating Hessian-based analyses, the modified CT18 and MSHT20 NNLO. In that exercise, an agreed-upon list of datasets, common to all groups and marked in Tables II–IV, was fitted. The CTEQ-TEA collaboration contributed CT18′ NNLO, which included the ATLAS 7 TeV high-precision W, Z data (ID = 68 for ATLAS and MSHT, or 248 for CT) as in CT18A, and differed slightly from the default CT18 in the choice of $m_c^{\text{pole}} = 1.4$ GeV and small differences in the other datasets. Besides this baseline choice of datasets, common theory settings, like coinciding QCD parameters or setting $s = \bar{s}$, were adapted to minimize discrepancies between all fitting groups. All differences found in the reduced fits were hence mostly due to methodological choices. In reflection of this, the L_2 sensitivities for the CT18′ and MSHT20 reduced fits discussed in Sec. IV B quantify the differences due to their methodologies while fitting (practically) the same dataset.

B. Conventions for log-likelihoods

The probabilities $P(D|T)$ in the likelihood ratio tests (13) and (14) are in fact augmented likelihoods [155] in the sense that they include prior probability contributions associated with the nuisance parameters describing correlated systematic effects. Normally these prior contributions are included as quadratic sums of the nuisance parameters that are assumed to obey the standard normal distribution $\mathcal{N}(0, 1)$. Since the experiments provide only a partial information about correlations among systematic factors, the PDF analyses use approximate correlation or covariance matrices to estimate these factors according to a number of prescriptions. The choice of these χ^2 definitions leads to non-negligible differences among the resulting PDFs, with proper modeling of systematic effects presenting a central issue for NNLO PDF determinations (see Sec. V.A in [156]). For example, the pulls on the large-x gluon by jet production experiments depend on the χ^2 definition.

For completeness, here we list the functional forms of the χ^2 that were used to compute the sensitivities.

In the ATLASpdf21 NNLO analysis [20], the definition of the χ^2 is

$$
\chi^2 = \sum_{i,k=1}^{N_{\mathrm{pts}}} \left(D_i - T_i + \sum_{\alpha=1}^{N_{\mathrm{corr}}} X_{i,\alpha} \beta_{i\alpha} \lambda_\alpha \right) C^{-1}_{\mathrm{stat,uncor},ik}(D_i, D_k)
$$
$$
\times \left(D_k - T_k + \sum_{\alpha'=1}^{N_{\mathrm{corr}}} X_{k,\alpha} \beta_{k\alpha'} \lambda_{\alpha'} \right)
$$
$$
+ \sum_{i=1}^{N_{\mathrm{pts}}} \log \frac{\delta_{i,\mathrm{uncor}}^2 T_i^2 + \delta_{i,\mathrm{stat}}^2 D_i T_i}{\delta_{i,\mathrm{uncor}}^2 D_i^2 + \delta_{i,\mathrm{stat}}^2 D_i^2} + \sum_{\alpha=1}^{N_{\mathrm{corr}}} \lambda_\alpha^2, \quad (15)
$$

where D_i are the central values of the measured data, T_i are the corresponding theoretical predictions, $\delta_{i,\mathrm{uncor}}$ and $\delta_{i,\mathrm{stat}}$ are the uncorrelated systematic uncertainties and the statistical uncertainties of D_i, respectively, and the fractional uncorrelated systematic and statistical uncertainties described by the correlation matrix $\beta_{i\alpha}$ are accounted for using the nuisance parameters λ_α. The quantity $C_{\mathrm{stat,uncor},ik}$ is a covariance matrix for both the statistical and uncorrelated systematic uncertainties. Summations over i and k run over all N_{pts} data points, and summations over α and α' run over all N_{corr} sources of correlated systematic uncertainty. For each dataset, the first term gives the main contribution to the partial χ^2 of the dataset. The second term is a small bias correction term, referred to as the log penalty, which arises because the diagonal term of the matrix, C, is given by $C_{ii} = \delta_{i,\mathrm{uncor}}^2 T_i^2 + \delta_{i,\mathrm{stat}}^2 D_i T_i$, with different weighting for statistical uncertainties and uncorrelated systematic uncertainties. This form of the χ^2 is used as standard in HERA and ATLAS PDF fits [35,36,64].

The CTEQ-TEA and MSHT analyses use

$$
\chi^2 = \sum_{i=1}^{N_{\mathrm{pts}}} \left(\frac{D_i - T_i + \sum_{\alpha=1}^{N_{\mathrm{corr}}} X_{i,\alpha} \beta_{i,\alpha} \lambda_\alpha}{\sqrt{\delta_{i,\mathrm{uncor}}^2 + \delta_{i,\mathrm{stat}}^2 D_i}} \right)^2 + \sum_{\alpha=1}^{N_{\mathrm{corr}}} \lambda_\alpha^2, \quad (16)
$$

when the individual correlated sources are provided. When only the final covariance matrix C_{ij} is available, they instead compute

$$
\chi^2 = \sum_{i=1}^{N_{\mathrm{pts}}} \sum_{j=1}^{N_{\mathrm{pts}}} (D_i - T_i)(C^{-1})_{ij}(D_j - T_j). \quad (17)
$$

In all these fits, $\beta_{i,\alpha}$ are published as percentage correlation matrices that must be multiplied by a reference value X_i for each data point when included in Eqs. (15) and (16). In addition, the MSHT20aN3LO fit includes nuisance parameters and the associated correlation matrices to estimate theoretical uncertainty sources from the modeled aN3LO ingredients—see [23] for details. The three groups treat all correlated systematic uncertainties as multiplicative ones, which amounts to setting $X_{i,\alpha} = T_i$ for all α in the above χ^2 definitions. While such a prescription is not

unique, it helps to reduce the bias when the uncertainties are dominated by statistical fluctuations [157–162]. It should be kept in mind, however, that the true $X_{i,\alpha}$ are generally unknown, reflecting the broad challenges in the modeling of the LHC systematic uncertainties (see Sec. 5.A in [156]), and other biases can be present, especially when the statistical errors are small compared to the systematic ones. See the discussion of this point in Ref. [162].

C. Tolerance conventions

The L_2 sensitivity S^H_{f,L_2} in Eq. (5) is proportional to the PDF uncertainty $\delta_H f$, and hence its magnitude reflects the tolerance prescription for constructing the Hessian eigenvector sets. These prescriptions get elaborate and nonuniform in the published PDF ensembles and may reflect the asymmetric behavior of the uncertainties, parametrization and scale dependence, and other nontrivial features. For the comparisons of the L_2 sensitivities, it is desirable to follow a simple and identical tolerance prescription for all compared fits.

In this regard, we choose the global tolerance with T^2 of about 10 for the presented comparisons, as it amounts to generating all eigenvector sets with $\Delta\chi^2 \approx 10$ and can be easily implemented in the fit and related to the LM scans, as discussed in Sec. II C. The value of $\Delta\chi^2 = 10$ is low enough to suppress nonlinear deviations from the L_2 sensitivity formula.

Furthermore, the actual tolerances in the ATLASpdf21 and MSHT20 fits are close on average to the global $\Delta\chi^2 = 10$ tolerance. From the ATLASpdf21 NNLO analysis, we take the Hessian eigenvector sets with $t^2 = 9$. We include only the error sets corresponding to the experimental uncertainties and not to the model and parametrization uncertainties, because when tolerance $t = 3$ is used, the model and parametrization uncertainties are relatively small. Here and in the following, the lowercase "t" indicates that the displacements of the eigenvector sets from the best-fit set were computed using the idealized Gaussian χ^2 arising in the diagonalization of the Hessian matrix. The uppercase "T" indicates that the displacements are computed using the actual χ^2, which includes some nonquadratic terms and is mildly asymmetric. The T and t criteria are generally close but not identical.

The MSHT20 fits by default apply the "dynamic tolerance" procedure, based on a weaker hypothesis-testing criterion and described in more detail in [8,21,101]. This enlarges the uncertainties beyond the $\Delta\chi^2 = 1$ definition to account for dataset tensions, as well as potential mismatch of data and theory due to imprecision in the theory or parametrization and experimental measurements. This procedure may be too complicated when calculating the L_2 sensitivity, so for this study the MSHT group provides additional Hessian sets that are computed with a fixed tolerance of $T^2 = 10$. The resulting PDF errors are rather

close to the ones obtained by applying the dynamic tolerance of MSHT.

While the aN3LO central PDFs include the estimated N3LO corrections, in the plotted aN3LO uncertainties and sensitivities, we exclude the contributions from the eigenvector sets due to the nuisance parameters that correspond to the theoretical uncertainty from the remaining ambiguities in the N3LO (and implicitly higher-order) ingredients. This is done to provide the most like-for-like comparison with the other PDF ensembles, which do not include such theoretical uncertainty. As observed in [23], the neglected error sets are in general subleading and do not significantly modify the outcomes for the purposes of the current study.

Finally, the CT18 error sets are determined using a two-tier (global + dynamic) tolerance that accounts for experimental, parametrization, and methodological uncertainties and results in error bands that are wider than with the $T^2 = 10$ tolerance. It was easy, however, to obtain a special set of CT18 eigenvector sets corresponding to the global $T^2 = 10$.

IV. SENSITIVITIES FOR NONGLOBAL FITS

With the background given in the previous sections, we can now analyze the Hessian sensitivities of the selected PDF ensembles and even compare L_2 sensitivity patterns obtained with ensembles from various groups. In this section, the sensitivities will highlight the prominent features of the fits that are based on specific datasets, like the ATLASpdf21 fit [20] that is dedicated to HERA + ATLAS data, Sec. IV A. Another class of fits that we consider here is the so-called reduced fits, Sec. IV B. Examples of such fits were proposed in the context of the PDF4LHC21 benchmarking exercise [18,111].

The full collection of the figures is available at the companion website [30], showing L_2 sensitivities to PDFs and PDF combinations [163] and for individual fitted experiments [164].

A. Sensitivities for ATLAS fits

We begin by comparing the χ_E^2 values and the L_2 sensitivity variable for the three classes of data, HERA, ATLAS W, Z 7 TeV and ATLAS $t\bar{t}$ 8 TeV, within the three fits (HERAPDF2.0NNLO, ATLASepWZ16 and ATLASepWZtop18) summarized in Sec. III A 1. Table V gives the χ_E^2 values and number of data points, N_{pts}, for data sets entering each of the ATLAS fits. Note that not all data sets enter every fit. If a data set is not fitted, its values are printed in italics. Note also that the number of data points for a particular data set may differ according to the PDF fit under consideration. Comparing the χ^2 values among the fits shows how fitting improves the description of these data. Inclusion of additional data sets may also degrade the

TABLE V. χ_E^2 values, and number of data points, N_{pts}, for the HERA combined, ATLAS W, Z 7 TeV and ATLAS $t\bar{t}$ datasets as predicted by each of the PDFs HERAPDF2.0NNLO, ATLAS-epWZ16, ATLASepWZtop18. If a dataset is not used in the PDF fit, then its values are in italics.

PDF	Dataset	χ_E^2	N_{pts}
HERAPDF2.0	HERA I + II combined	1363	1145
ATLASepWZ16	HERA I + II combined	1213	1056
ATLASepWZtop18	HERA I + II combined	1149	1016
HERAPDF2.0	ATLAS W,Z 7 TeV	*384*	*61*
ATLASepWZ16	ATLAS W,Z 7 TeV	108	61
ATLASepWZtop18	ATLAS W,Z 7 TeV	79	55
HERAPDF2.0	ATLAS $t\bar{t}$ 8 TeV	*31*	*20*
ATLASepWZ16	ATLAS $t\bar{t}$ 8 TeV	*19*	*20*
ATLASepWZtop18	ATLAS $t\bar{t}$ 8 TeV	17	20

description of data already in the fit, revealing tensions among the data sets.

The χ_E^2 value of the HERA data (160) within the HERAPDF2.0NNLO fit is higher than ideal, but it is similar in value to that in the CT18 and MSHT20 fits. However, it is significantly reduced in the ATLASepWZ16 and ATLASepWZtop18 fits because of the harder Q^2 cuts imposed on these HERA data: $Q^2 > 7.5$ GeV^2 for the ATLASepWZ16 PDFs and $Q^2 > 10$ GeV^2 for the ATLASepWZtop18 PDFs. These cuts avoid data for which low-x higher-twist effects or $\ln(1/x)$ resummation may be important.

The χ_E^2 value of the ATLAS W, Z 7 TeV data (68) is very high for the HERAPDF2.0NNLO fit, reduces in the ATLASepWZ16 fit (as the W, Z dataset is now a part of the fit), and reduces further in the ATLASepWZtop18 fit, in which the six data points from the low-mass Drell-Yan data are removed due to poor understanding of their electroweak radiative contributions.

The χ_E^2 values of the ATLAS $t\bar{t}$ 8 TeV data (7) are not very high, even for PDFs that are not fitted to the data. There is no tension between the $t\bar{t}$ data included in the ATLASepWZtop18 fit and the HERA data or the W, Z data.

The features of the fits revealed above using the χ_E^2 values are much more clearly seen in the L_2 sensitivity plots. Figure 6 compares the L_2 sensitivities for the gluon PDF in the HERAPDF2.0NNLO, ATLASepWZ16, and ATLASepWZtop18 fits for the three datasets: HERA combined inclusive DIS data (ID = 160); ATLAS W,Z 7 TeV data (68); ATLAS $t\bar{t}$ 8 TeV data in lepton + jets and dilepton channels (7). Note that the sensitivity to the ATLAS W, Z data for the HERAPDF2.0NNLO fit is very large and has been divided by a factor of 10 for display. These large sensitivities are dramatically reduced when the W, Z data are fitted. Then one can observe that the HERA data and the $t\bar{t}$ data pull against each other in the
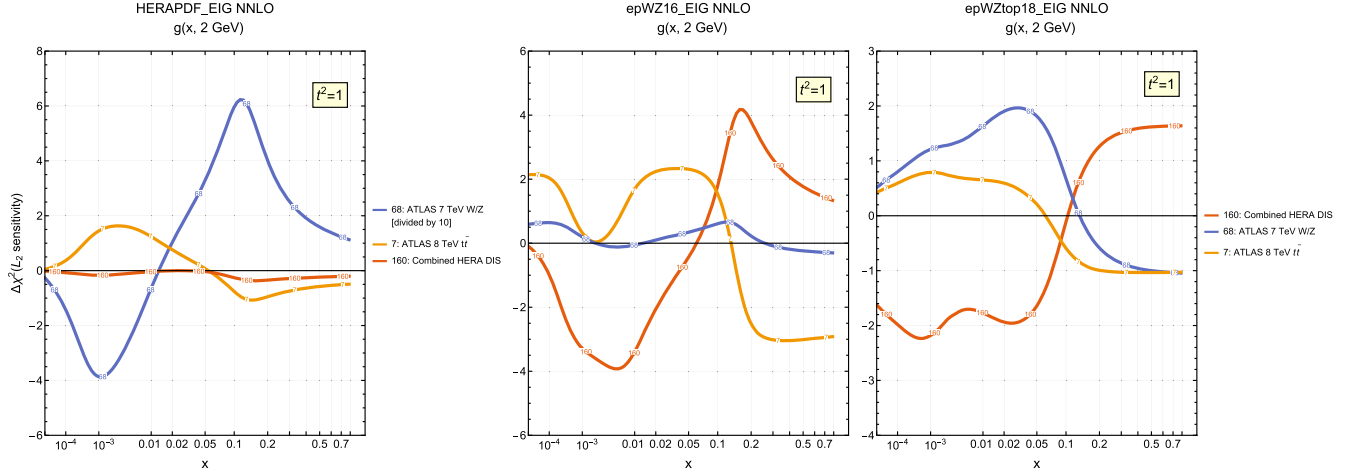
FIG. 6. L_2 sensitivity for HERA, ATLAS W, Z 7 TeV and $t\bar{t}$ 8 TeV datasets for the gluon at $Q = 2$ GeV in the fits: left, HERAPDF2.0; middle, ATLASepWZ16; right, ATLASepWZtop18. In the left panel for HERAPDF2.0, the ATLAS W, Z sensitivity is divided by ten.

ATLASepWZ16 fit, with the $t\bar{t}$ data favoring a harder gluon at high x. Once the $t\bar{t}$ data are fitted in the ATLASepWZ18 fit, the gluon becomes harder, and the magnitudes of sensitivity for these data decrease. One can also observe the change in sensitivity to the HERA data between these two fits, where sensitivity to the gluon at low x is decreased when the harder Q^2 cut is imposed in the ATLASepWZtop18 fit. Although there are subtle changes in the pulls of the datasets between the two ATLAS fits, no significant tensions between the ATLAS datasets and the HERA dataset are evident. As we commented earlier, the sensitivity quantifies such tensions, while the HERA DIS dataset is in a good agreement with the PDFs in this case and imposes strong constraints on them.

Figure 7 compares the L_2 sensitivities for the strangeness PDF in the HERAPDF2.0NNLO, ATLASepWZ16 and ATLASepWZtop18 PDFs for the same datasets. Again the sensitivity to the ATLAS W, Z data in the HERAPDF2.0NNLO fit is reduced by a factor of 10 for

display. Clearly the ATLAS W, Z data show preference to increase the strangeness PDF of HERAPDF2.0, particularly at $x \sim 0.01$. This was confirmed when the data were fitted. There are no large sensitivities remaining in the ATLAS PDFs, although there are subtle changes of sensitivity shape and sign between the two fits.

Now we consider the L_2 sensitivity for the most complete ATLAS fit to date, ATLASpdf21, which includes all the datasets listed in Table I. We first plot the sensitivities to the gluon and strange PDFs for six most sensitive experiments in Fig. 8. It turns out that, whatever the PDF flavor (including u and d flavors, not shown), the most sensitive experiments are the ATLAS 7 TeV W, Z, the ATLAS 8 TeV Z3D data, HERA combined data, the ATLAS 8 TeV W data, the ATLAS 8 TeV V + jets data, and the ATLAS 8 TeV inclusive jet data. The $t\bar{t}$ data and the direct photon data have smaller sensitivities and so are not shown. For the gluon, the ATLAS 7 TeV W, Z data are the most sensitive, and we observe the opposing, though weaker, tendency of
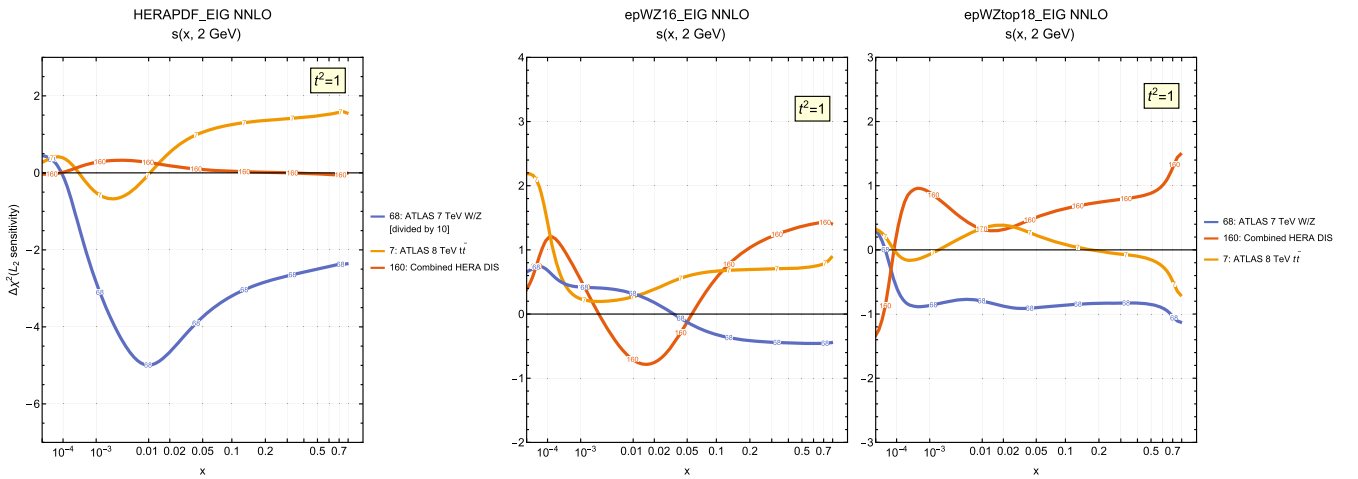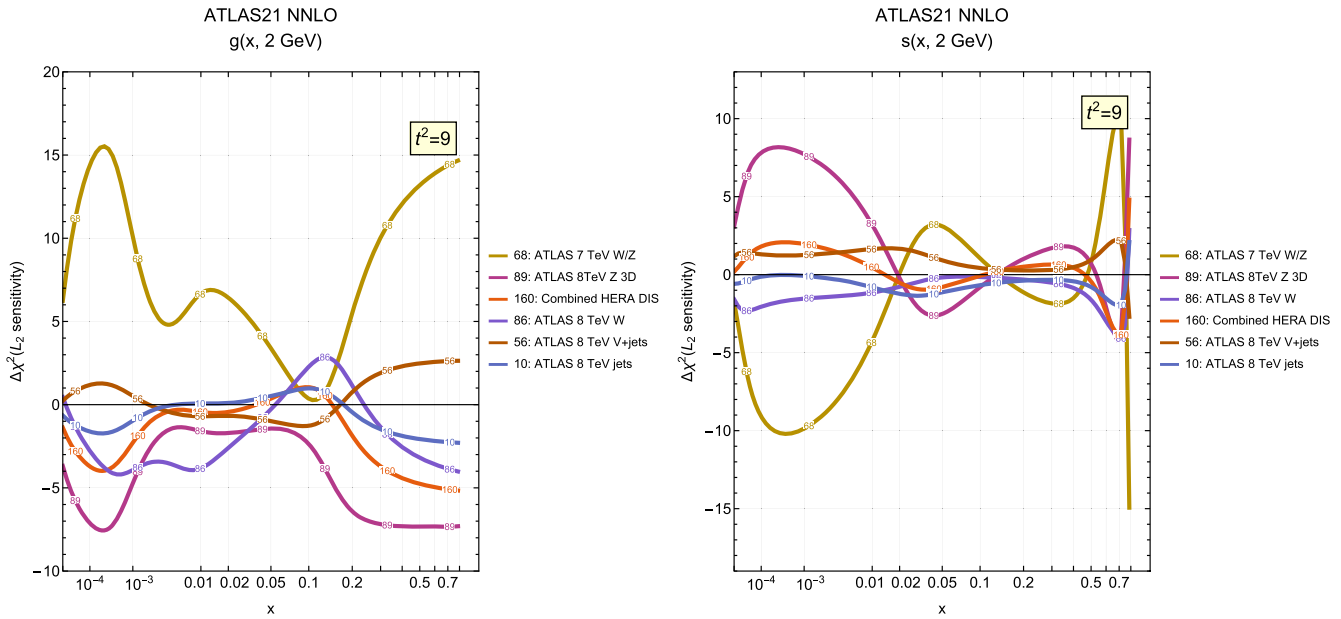

FIG. 7.   Same as Fig. 6, for the strangeness PDF.

FIG. 8. L_2 sensitivity for all datasets in the ATLASpdf21 fit: gluon (left), strange (right).

the ATLAS 8 TeV Z3D data. For the strangeness, we again see the opposing tendencies of ATLAS 7 TeV W, Z and ATLAS 8 TeV Z3D data, with ATLAS 7 TeV W, Z favoring more strangeness at low x, and ATLAS 8 TeV Z3D data favoring less with almost equal weight. It is interesting that these data are in agreement at $x \sim 0.02$ for $Q = 2$ GeV, since this is approximately the x and Q^2 values at which ATLAS chose to illustrate the ratio of strangeness to light quarks. Above this x value, the opposition persists but is much weaker until $x > 0.5$, where ATLAS 7 TeV W, Z data

and ATLAS 8 TeV V + jets data favor less strangeness but are opposed by HERA combined data, ATLAS 8 TeV W data and ATLAS 8 TeV inclusive jets data.

In Fig. 9, we show the sensitivities to the strangeness ratio $2s(x, Q)/(\bar{u}(x, Q) + \bar{d}(x, Q))$ at $Q = 2$ GeV and $Q = 100$ GeV. We again see the opposing tendencies of ATLAS 7 TeV W, Z and ATLAS 8 TeV Z3D data, with ATLAS 7 TeV W, Z favoring a higher strangeness ratio at low x, and ATLAS 8 TeV Z3D data favoring a smaller ratio with almost equal weight. As we saw for the strangeness
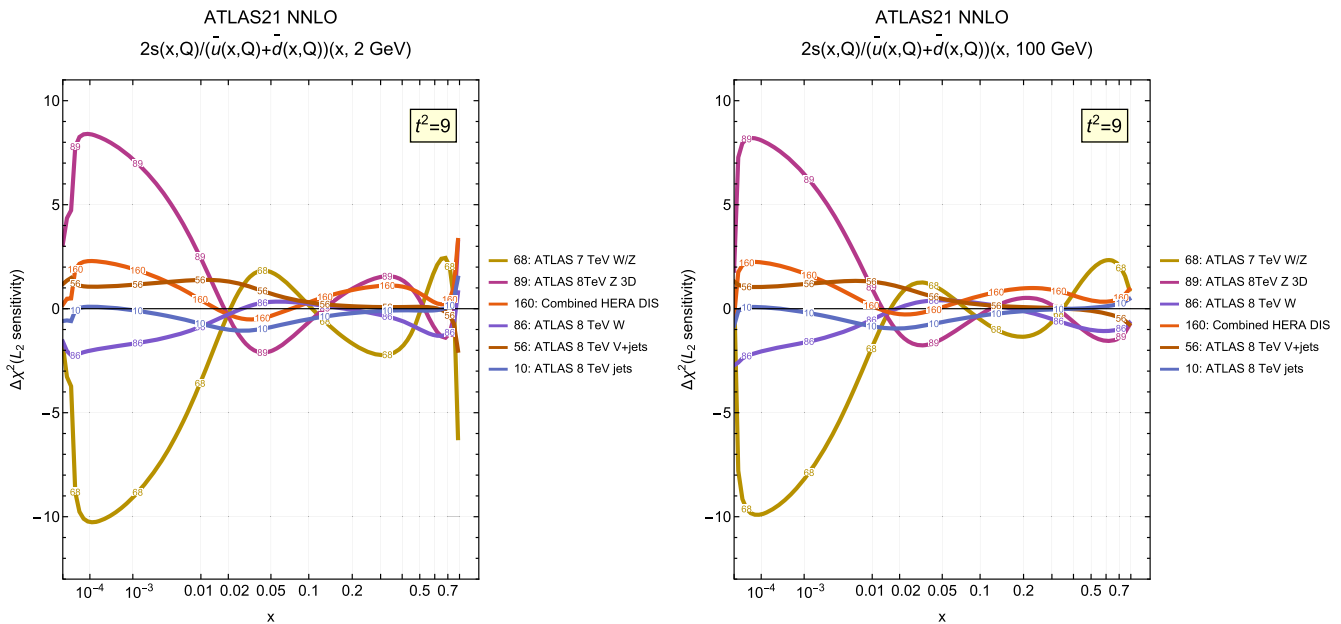
FIG. 9. L_2 sensitivity for all datasets in the ATLASpdf21 fit for the strangeness ratio $2s(x, Q)/(\bar{u}(x, Q) + \bar{d}(x, Q))$: $Q = 2$ GeV (left), $Q = 100$ GeV (right).
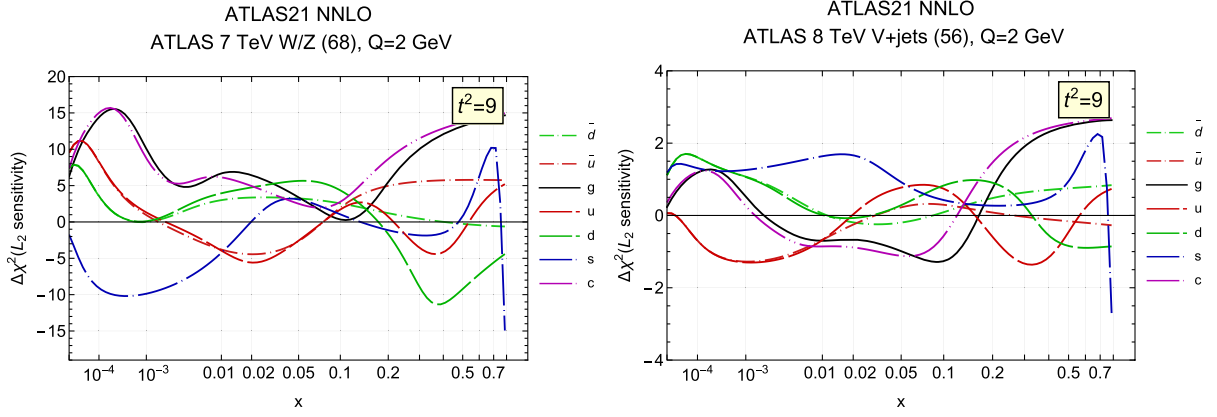
FIG. 10. The L_2 sensitivities to all PDF flavors in the ATLASpdf21 fit for the ATLAS 7 TeV W, Z data (left), ATLAS 8 TeV V + jets data (right).

PDF, these experiments are in agreement for $x \sim 0.02$. The picture at low scale $Q = 2$ and at high scale $Q = 100$ GeV is very similar, with the shapes moving to lower x, as expected by QCD evolution.

An alternative way to look at the sensitivities is to plot the sensitivity to each experiment for all the PDF flavors. This is shown in Fig. 10 for the ATLAS 7 TeV W, Z data and the ATLAS 8 TeV V + jets data.
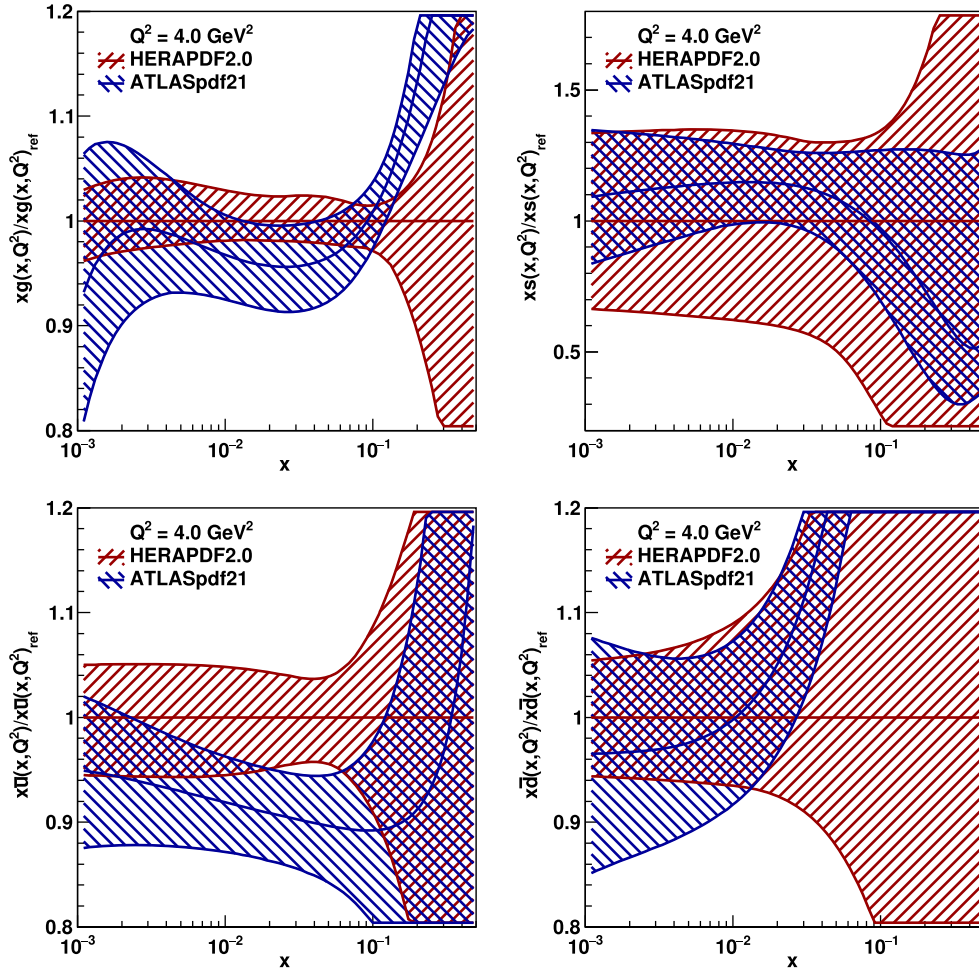
FIG. 11. The ratio of HERAPDF2.0 (red) and ATLASpdf21 (blue) taken to the central PDF of HERAPDF2.0 for g, s, $\bar{u}$, and $\bar{d}$ PDFs at $Q^2 = 4$ GeV^2. The HERAPDF band represents the full uncertainty (EIG + VAR) for $T = 1$. ATLASpdf21 is given for the tolerance of $T = 3$.

The ATLAS 7 TeV W, Z data are quite strongly sensitive to the gluon as well as to the strange PDF at $x < 0.05$, to d and (to a lesser extent) u at high x. The ATLAS 8 TeV $V + $ jets data are sensitive both to the gluon (mostly due to $Z + $ jets data) and to the strange PDF (mostly due to $W + $ jets data) at high x, wanting both smaller high-x gluon and smaller high-x strangeness. A consequence of the smaller high-x strangeness is also a larger high-x $\bar{d}$, as discussed in Ref. [65].

The L_2 sensitivities for the ATLAS fits thus illustrate how large sensitivity to a dataset is evident before it is fitted, how sensitivity is reduced if a good fit is made to these data, and how smaller residual tensions between fitted datasets can be made evident. More information can be gained by the use of the online plotter [30]. The trends described above, of adding ATLAS data to the HERA data, can be observed in the comparison plot for the HERA and ATLASpdf21 ensembles themselves, Fig. 11.
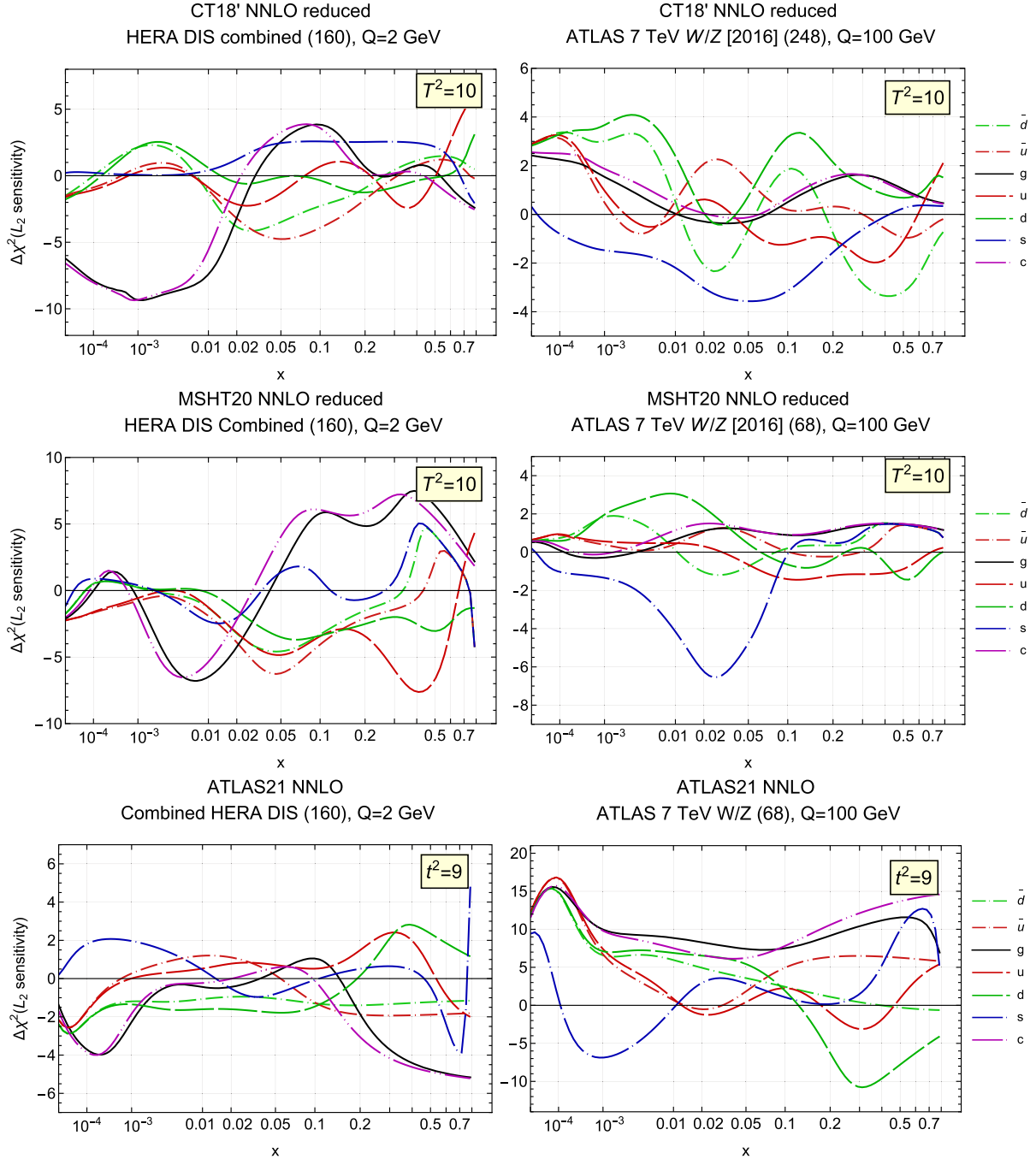

FIG. 12.　Sensitivities for the CT18′, MSHT20, and ATLASpdf21 nonglobal fits for HERA DIS dataset (left column, $Q = 2$ GeV in S_{f,L_2}) and ATLAS 7 TeV W, Z dataset (right column, $Q = 100$ GeV).
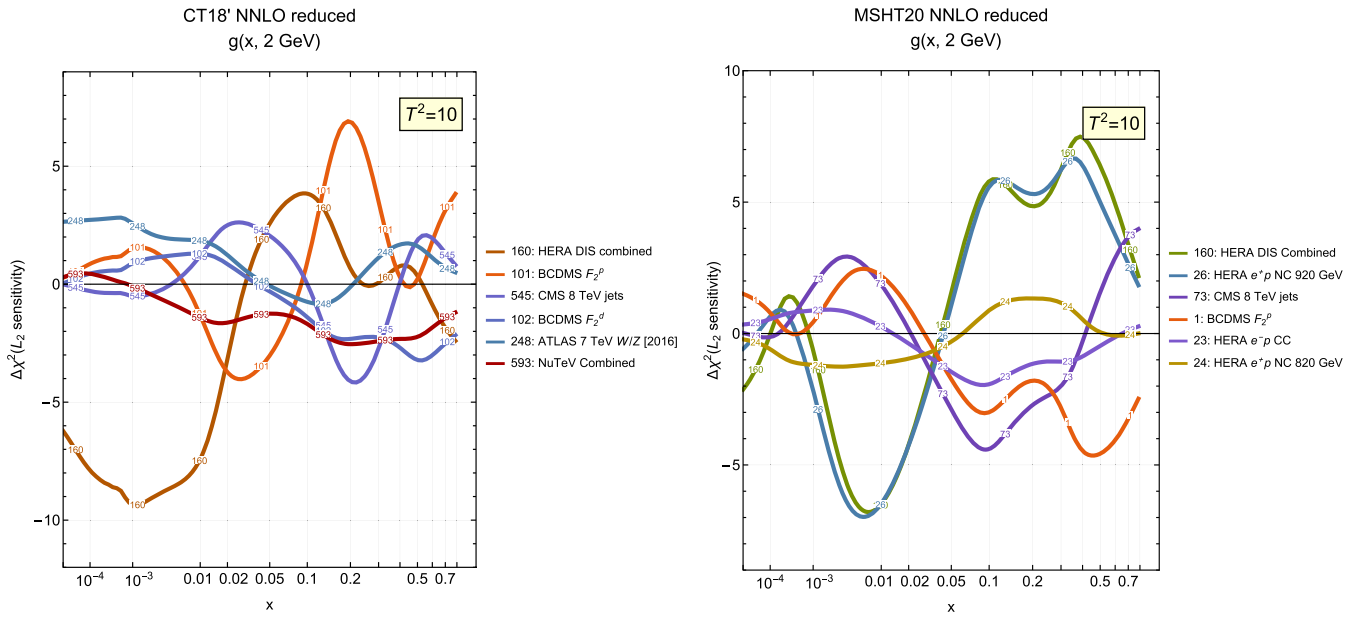
FIG. 13. Sensitivities to the gluon PDF at $Q = 2$ GeV in the CT18′ and MSHT20 reduced fits. The equivalent for ATLASpdf21 is given in Fig. 8.

B. Sensitivities for reduced benchmarking fits

Section III A 5 summarized the CT18′ and MSHT20 reduced fits that led to the PDF4LHC21 combination [18,111]. These two analyses based on the Hessian methodology show little differences in their central values as well as in the magnitude of their uncertainties, as can be appreciated in, e.g., Figs. 3.4, 3.5 of Ref. [18]. The L_2 sensitivity technique was used to aid selection of the optimal common dataset for the benchmarking exercise, and it motivated dedicated studies for specific data, see Appendices C and D of Ref. [18]. In this section, we illustrate the insights that can be gleaned about the reduced fits from the plots of their sensitivities collected on the online plotter [30].

Figure 12 shows the L_2 sensitivities for the HERA combined DIS and ATLAS 7 TeV W, Z [2016] datasets in the CT18′, MSHT20 reduced fits and the ATLASpdf21 fit. The PDF factorization scales in S_{f,L_2} are set to $Q = 2$ and 100 GeV for the HERA DIS and ATLAS W, Z datasets, respectively. Qualitatively, the shapes of the sensitivities are more similar between the CT18′ and MSHT20 reduced fits than in the full fits. The differences in the magnitudes of sensitivities are more pronounced. Attention must be paid to the ranges of sensitivities on the vertical scales that are not the same across the figures.

The HERA data (lhs) impose a pronounced preference for a higher gluon at $x < 0.02$ in all three nonglobal fits. At higher x, the ATLASpdf21 fit still mostly prefers a larger gluon, in reflection of having less data sensitive to the gluon at high x. In both the CT18′ and MSHT20 reduced fits, on the other hand, the HERA data prefer a suppressed gluon at $x = 0.02$–0.2, and in the MSHT20 reduced fit a smaller gluon is preferred above $x = 0.2$ as well, which is in the

pattern that is not dissimilar to the differences in the sensitivities of the HERA heavy-quark cross sections in the full fits in Fig. 5. As the reduced analyses fit the same HERA data, the behavior of sensitivities may thus reflect methodological differences among the CT18′ and MSHT20 reduced fits, such as their different heavy-quark schemes and gluon parametrization forms.

The pulls on the strangeness from ATLAS 7 TeV W, Z data (rhs) favor larger strangeness at intermediate x for both CT18′ and MSHT20 reduced fits. This pull is reduced in the ATLAS fit, in which these data exert the dominant influence on strangeness, and the strangeness PDF has maximally increased to reflect this. For the reduced fits, this strong pull on the strangeness is opposed by the NuTeV data, as shown in Appendix D of Ref. [18]. The variations between the pulls for the large-x gluon and strangeness go hand in hand.

In Fig. 13, we show the L_2 sensitivities to the gluon at $Q = 2$ GeV for the CT18′ and MSHT20 reduced fits. The sensitivity to the ATLASpdf21 gluon PDF is shown in Fig. 8. The orders of magnitude of $S^{\mathrm{H}}_{f,L2}$ are comparable for the two reduced ensembles. However, ATLASpdf21 shows slightly larger values for the single sensitivities, the dominant one coming from the high precision ATLAS 7 TeV W, Z measurement. Global features of the pulls are similar for the three ensembles for, e.g., the HERA combined data, for which the negative pull at small x, favoring a larger gluon PDF, is larger for the reduced sets than for ATLASpdf21, because of the relative dominance of the HERA data in the ATLAS fit, which has already adjusted to reflect this. Here, both CT18′ and MSHT20 reduced fits show counterbalancing pulls on the gluon from the CMS 8 TeV jet data (labels 73 and 545), with the

ATLAS W, Z dataset 248 exerting an additional downward pull in the case of CT18′.

On the other hand, at $x > 0.1$, MSHT20red is still pulled downward mostly by the HERA combined data (160), while BCDMS data on $F_{2,p}$ (1) and CMS 8 TeV jets exert the main opposing upward pulls. In the CT18′ case, however, the BCDMS $F_{2,p}$ pull (101) at large x is downward and dominates over the weakened HERA DIS pull (160) of the same sign. The pull of the CMS 8 TeV jets continues to be consistently upward for CT18′.

The PDF4LHC21 document [18] noted this pronounced agreement of the pulls by the CMS 8 TeV jets between two reduced fits. The respective plot for the L_2 sensitivity was shown at 2 GeV in Fig. D1 of [18]. Already the early PDFSENSE study [5] pointed out the prominence of constraints from the CMS jet production revealed by the sensitivity method. Namely, this dataset dominates the pulls for the gluon and charm PDFs, resulting in smaller values of both PDFs at $x \sim 10^{-3}$ as well as in the valence region, but larger gluon and charm PDFs at $x \sim 0.05$–0.1. The corresponding plots for the full fit are shown in Sec. VII (see Fig. 21), where we observe the opposite trend between reduced and full fits for both MSHT20 and CT18, nonetheless its pulls remain in excellent agreement also between the full fits. The difference between the reduced and the full fits shows that, for the CMS 8 TeV jet dataset, the L_2 sensitivity is most influenced by the competing pulls of the other input datasets rather than the functional form. A counterexample is seen for in the gluon sensitivities of Fig. 13, where the differing contributions of the HERA combined data to the MSHT20 and CT18′ reduced fits (which use the same datasets) highlights the remaining differences between the reduced fits, including, for example, the functional form and heavy-quark scheme.

V. SENSITIVITIES FOR CT18, CT18AS, AND CT18AS_LAT FITS

As we summarized in Sec. III A 2, the comprehensive CT18 NNLO study [17], published by the CTEQ-TEA collaboration in early 2021, has extensively employed the L_2 sensitivities in the prefit and postfit investigations of the constraints from the experiments. The CT18 website [81] collects 340 plots of L_2 sensitivities for the individual experiments, PDF flavors, PDF combinations, and parton luminosities at the LHC in the CT18 and C18Z NNLO analyses. Many of these results are discussed in the CT18 publication. Complementary insights were gained from visualizations of L_1 sensitivities of individual data points using the PDFSENSE program [5] and the CT14HERA2 NNLO ensemble (the immediate predecessor of CT18), with the resulting figures collected online at [165]. The PDFSENSE results were especially helpful for charting the global map of sensitivities of the fitted experiments to the PDF flavors [5] and to PDF combinations and Mellin moments that can be computed on the lattice [6].

This article and its companion website [30] reproduce many $S_{f,L2}$ plots from the CT18 NNLO analysis [17,81] in the format that facilitates the comparisons against the other PDF ensembles. Examples include Fig. 5 for HERA I charm production, Fig. 21 for CMS 8 TeV inclusive jet production, Fig. 22 for ATLAS and CMS 8 TeV $t\bar{t}$ production (IDs = 580 and 573), Fig. 23 for the combined HERA I + II DIS (ID = 160), Fig. 25 for ATLAS 7 TeV W, Z production (ID = 248), and Fig. 27 for E866 pd/pp ratio; as well as plots for the LHCb W, Z production, and DØ charge asymmetry in the Supplemental Material [166]. These figures are discussed later in their respective sections.

In addition to the experiment-by-experiment plots of L_2 sensitivities, we provide cumulative figures of sensitivities of the leading experiments to specific PDF flavors or PDF combinations, such as the valence PDFs $q_V(x, Q) \equiv q(x, Q) - \bar{q}(x, Q)$ and PDF ratios $f_a(x, Q)/f_b(x, Q)$. We already discussed such Fig. 3 for the CT18 NNLO sensitivities to the gluon. As another illustration, consider Figs. 14–16 comparing $S_{f,L2}$ for quark and antiquark PDFs between the CT18, CT18As, and CT18As_Lat PDF ensembles. We expect, and the sensitivities reflect it, that the primary differences between CT18 and CT18As_Lat arise in the (anti)quark flavor separation. The CT18A PDF ensemble is identical to CT18 but includes ATLAS 7 TeV W, Z precision data, which show sizable tension with other datasets, such as the HERA I + II DIS combined data and NuTeV dimuon data [17]. On top of that, the CT18As fit also adds degrees of freedom in the strangeness parametrization to allow s to be not equal to $\bar{s}$ at the initial scale $Q_0 = 1.3$ GeV. This choice slightly reduces the above-mentioned tension [22]: the $s(\bar{s})$ PDF error band becomes larger in comparison with the analogous CT18A PDFs. Through a feedback loop in the fit, the additional freedom in CT18As also enlarges the $d(\bar{d})$ error bands. The PDF ratio $\bar{d}/\bar{u}$ is also modified.

The CT18As_Lat fit [22] further includes lattice data [100] to constrain the strangeness asymmetry in the large-x region, $0.3 < x < 0.8$. The lattice s_- data modifies both the central fit and the error bands of R_s and $s_-(x)$. The R_s of CT18As_Lat gets higher (closer to that of MSHT20 and NNPDF4.0). The strangeness asymmetry s_- of CT18As_Lat is thus closer to zero at large x than those of CT18As, MSHT20, and NNPDF4.0.

But how exactly are these changes distributed among the flavors, pulls of datasets, and as a function of partonic x? This question is answered by plotting the L_2 sensitivities. In the upper right panel in Fig. 14, we see that the ATLAS 7 TeV $W\,Z$ dataset (248) in the CT18As_Lat fit prefers a higher s PDF than the other datasets at $x \approx 0.02$ and actually along the whole x range. This preference is discussed at length in Ref. [17]. We further see that the pattern of leading sensitivities in CT18 NNLO shown in the upper left panel has been modified in CT18As_Lat. For example, while in CT18 NNLO the NuTeV dimuon

FIG. 14. Upper: sensitivities to the strangeness PDF for CT18 and CT18As_Lat NNLO fits with $T^2 = 10$ at $Q = 2$ GeV, shown for the six most sensitive experiments. Lower: sensitivities to the strangeness asymmetry $(s - \bar{s})/(s + \bar{s})$ for CT18As and CT18As_Lat NNLO. (Notice that the color code and the ranges of the vertical scale vary between the plots.)

experiments (124 and 125) both prefer lower strangeness, but at somewhat different x, and the E866 pp Drell-Yan production moderately prefers to increase it, in the CT18As_Lat case the pulls of the NuTeV neutrino and antineutrino datasets become more alike at $x < 0.1$; at $x > 0.1$, the NuTeV ν dimuon dataset (124) develops a preference for a higher s that is largely canceled by the opposite pull by CCFR $\bar{\nu}$ dimuon set (127).

The CT18As and CT18As_Lat sensitivities available from the website clarify how these modifications emerge through the triple effect of including the ATLAS W, Z dataset, releasing the $s \neq \bar{s}$ condition (which substantially

modifies the flavor composition in CT18As), and including a lattice constraint on the magnitude of $s - \bar{s}$ (which partly offsets the previous effect). The lower two figures in Fig. 14 compare $S_{f,L2}$ for the $(s - \bar{s})/(s + \bar{s})$ asymmetry of CT18As and CT18As_Lat sets. In the CT18As case, without the lattice constraints on s_-, the pulls on the strangeness asymmetry by CCFR inclusive (110) and $\bar{\nu}$ dimuon (127) datasets, together with the same-sign pulls of E866 pp (204) and LHCb 7 and 8 TeV W, Z (258), are counteracted by the strong pull by the NuTeV $\bar{\nu}$ dimuon data (125). For CT18As_Lat, these sensitivities are substantially rebalanced by the lattice constraints on $s - \bar{s}$,

FIG. 15.   Sensitivities to the valence PDF $u_V = u - \bar{u}$ for CT18 NNLO (left) and CT18As_Lat NNLO (right) at $Q = 2$ GeV.

which is included in the figures together with the contributions of normalizations for fixed-target experiments (ID = 701). While the lattice + normalizations prefer a smaller asymmetry at $x > 0.1$, they have to compensate by preferring a larger asymmetry at $x < 0.1$ to satisfy the net absence of strangeness in the proton, $\int_0^1 s_-(x, Q)dx = 0$. The pull of the lattice s_- data is offset by the ATLAS W, Z, NuTeV, and BCDMS F_2^p curves over a large span of x. (The sensitivity of the ATLAS 7 TeV W, Z data to s_- is small in the CT18As case without the lattice data.)

Figure 15 illustrates the six strongest L_2 pulls on the up-valence PDF at $Q = 2$ GeV. In the CT18 NNLO fit

(left panel), over most of the x range we see the interplay of opposing pulls from the BCDMS F_2^p (101) and E866 Drell–Yan pp cross section (204), joined by a tug of war between CCFR xF_3 (111) and the LHCb W, Z production (245 or 258) datasets over a wide x range. When ATLAS 7 TeV W, Z data are added in the CT18As_Lat in the right panel, it partly offsets the pulls of LHCb (258) at $x < 0.2$, with HERA DIS (160) and NMC d/p ratio (104) offering additional opposing pulls. These figures can be compared with the L_2 sensitivities for MSHT20 u_V PDFs in Fig. 19, showing a similar behavior of the L_2 sensitivity for HERA DIS as in CT18As_Lat, as well as more prominent roles of

FIG. 16.   Sensitivities to the ratio $\bar{d}/\bar{u}$ for CT18 (left) and CT18As_Lat (right) NNLO fits at $Q = 2$ GeV.

BCDMS F_2^p and F_2^d datasets (101 and 102), and of the DØ Run-2 W boson-level asymmetry that is not included in CT18 fits.

In the same vein, the $\bar{d}/\bar{u}$ sensitivities for the CT18 and CT18As_Lat fits shown in Fig. 16 can be compared against the MSHT20 ones in Fig. 20. Overall, the ratio of pd and pp Drell-Yan cross sections in the E866/NuSea experiment (ID = 203 for CT and 13 for MSHT) prefers a lower $\bar{d}/\bar{u}$ ratio across the majority of the x range. The exception is the pronounced preference for a higher $\bar{d}/\bar{u}$ at $x = 0.04$–0.2 in CT18 ($x = 0.07$–0.3 in CT18As_Lat). This E866 pull is balanced by a combination of HERA DIS, NMC d/p ratio, BCDMS F_2^p, and LHC 8 TeV W, Z pulls in the CT18 case. For CT18As_Lat, the E866 pull toward smaller $\bar{d}/\bar{u}$ at $0.01 < x < 0.07$ is amplified and nearly completely opposed by an upward preference for $\bar{d}/\bar{u}$ arising from the ATLAS 7 TeV W, Z dataset. In both the CT and MSHT cases, a variety of vector boson production experiments at the Tevatron and LHC have sensitivity to $\bar{d}/\bar{u}$. The (not fitted) E906/SeaQuest experiment included in the CT18As_Lat analysis largely opposes the preference of the E866 ratio for diminished $\bar{d}/\bar{u}$ at $x > 0.2$.

VI. SENSITIVITIES FOR THE MSHT20 FITS

In this section, we present some results comparing the L_2 sensitivities for the MSHT20 NNLO and aN3LO PDFs, showing the improvements that are obtained in some cases at aN3LO. First, in Fig. 17 we compare the sensitivities for the gluon. At NNLO we see this is dominated by the ATLAS Z p_T data, which are also comparatively poorly fit according to the χ^2 value presented in [21]. A similar pattern in the shape of the sensitivities is observed at

aN3LO, although their amplitudes are substantially reduced, typically being less than a half of that at NNLO. It is also observed in [23] that it has a substantially reduced χ^2, corresponding to a better fit quality at aN3LO. At NNLO, there are large corrections for these precise ATLAS Z p_T data so significant changes at aN3LO may well be expected. The sensitivity to the HERA combined structure function data at NNLO is also very much reduced at aN3LO, with this set no longer in the top six datasets shown in Fig. 17. Tensions between these two datasets at NNLO are largely eliminated at aN3LO. At the PDF level, the better agreement between the two experiments is accompanied by some rearrangement of the gluon and quark PDFs at aN3LO in the relevant x region of about 0.01. The sensitivities of the CMS 8 TeV jet data and NMC F_2^d data do not change much and are even enhanced slightly when going to aN3LO (note the different axes scales), showing that inclusion of the approximate N3LO corrections do not help with the tensions for these sets. However, overall, at NNLO the tension is overwhelmingly between ATLAS Z p_T data and other datasets, whereas at aN3LO the overall tension between different datasets is much reduced and more evenly spread, though ATLAS Z p_T data and CMS 8 TeV jet data are still quite directly in conflict.

Next, in Fig. 18 we compare the results for the strange quark. Here, despite no very obvious direct reason for sensitivity, we see the NNLO fit displays significant sensitivity again to the ATLAS Z p_T data, showing the general issues in fitting these data at NNLO. This is absent at aN3LO. Similar sensitivity is seen in both PDFs for the NuTeV dimuon data and the ATLAS 7 TeV W, Z data, which pull in the same direction at high x but oppose at lower x. However, at NNLO these are in tension with ATLAS Z p_T data, while
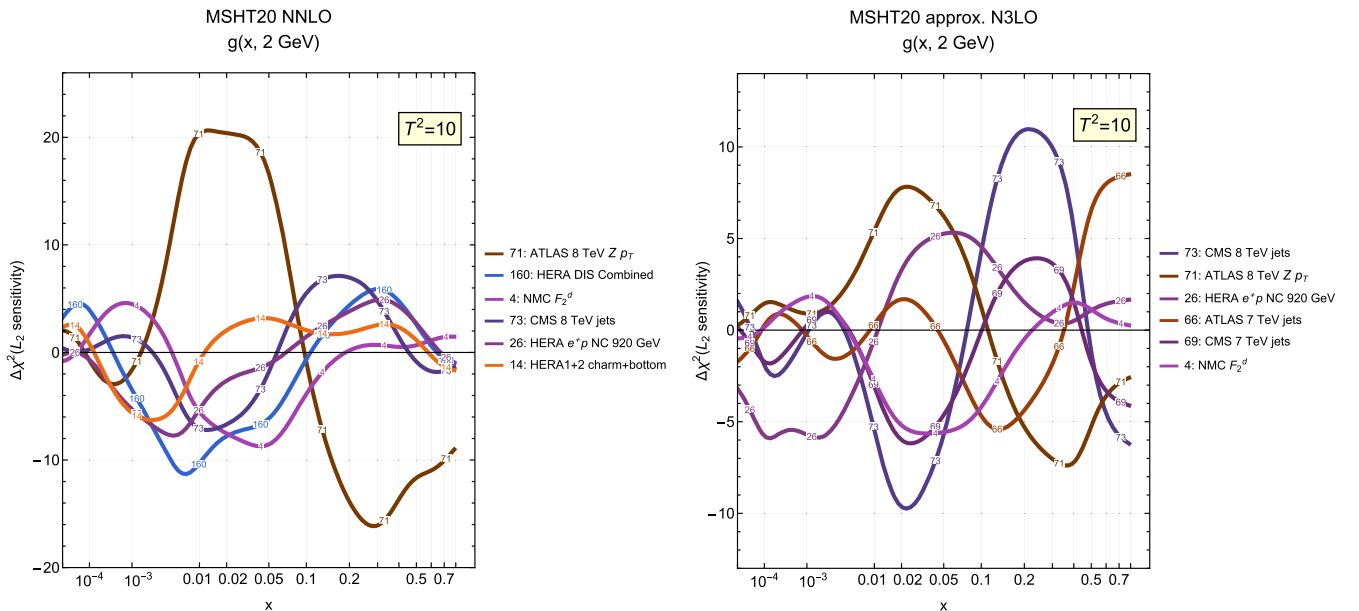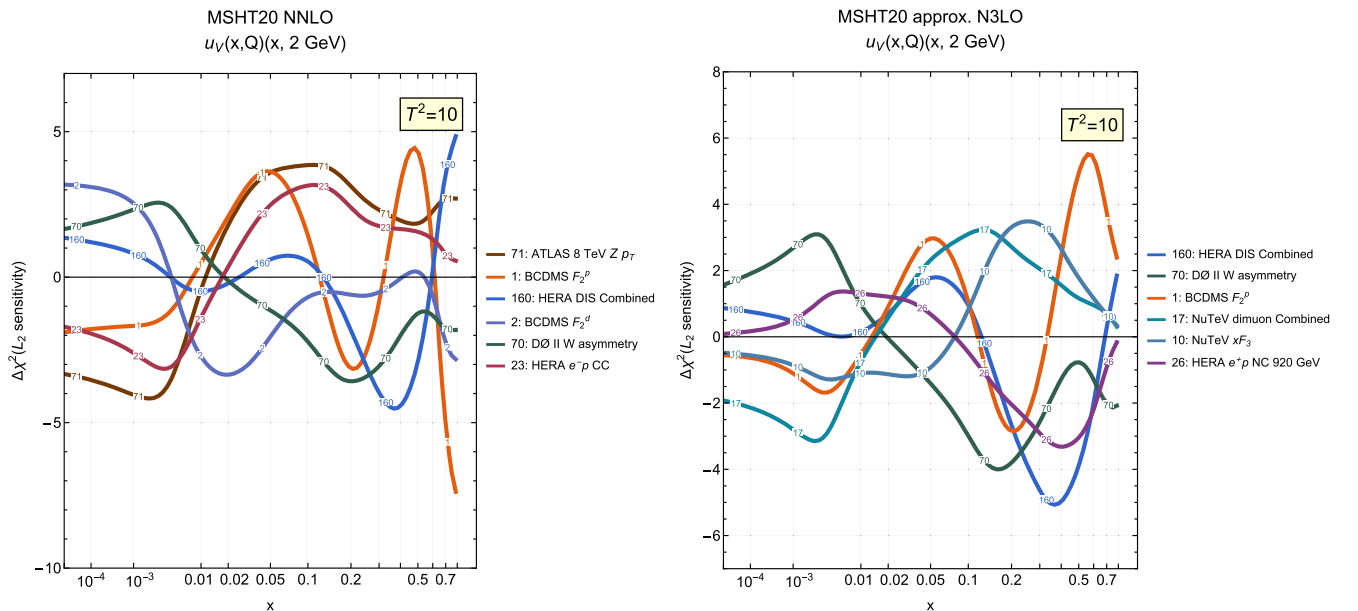

FIG. 17. L_2 sensitivities showing the six most sensitive datasets for the MSHT20 NNLO and MSHT20aN3LO gluon PDFs.
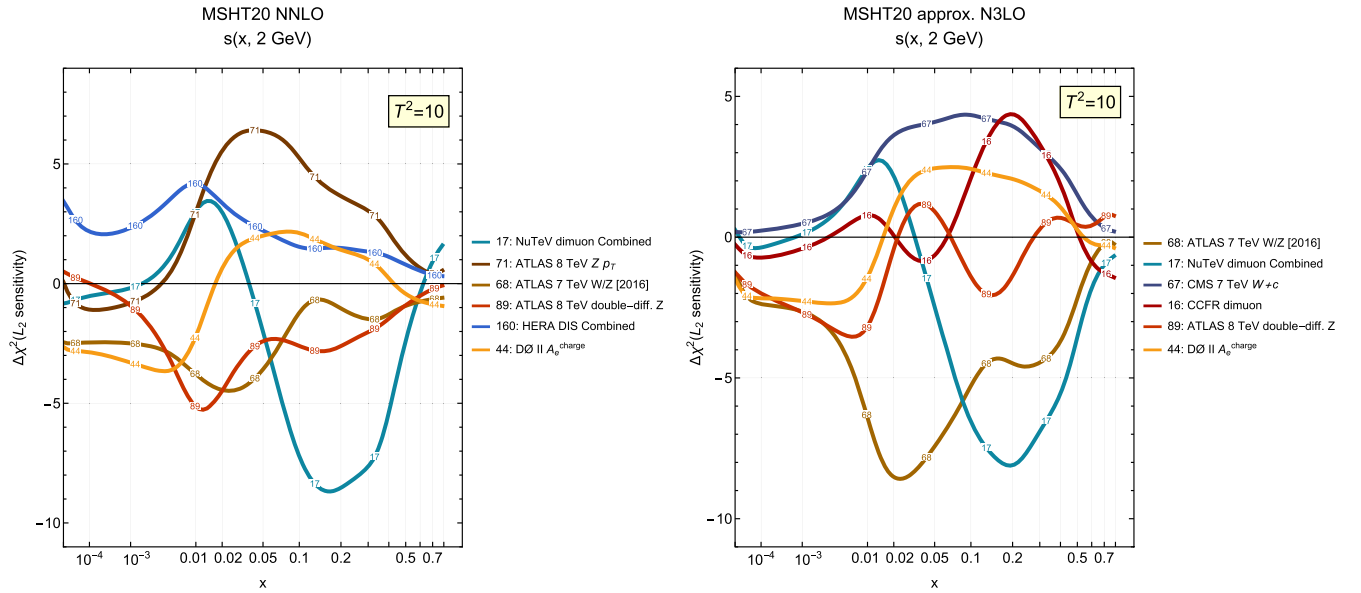
FIG. 18. L_2 sensitivities showing the six most sensitive datasets for the MSHT20 NNLO and aN3LO strange PDFs.

at aN3LO it is largely with CMS $W + c$ data, which favors a slightly reduced strangeness over most of the x range. This may be caused by minor issues in the aN3LO fit due to relatively little knowledge of the theory cross section beyond NLO, and also likely due to the larger strange quark at aN3LO.

In Fig. 19, we compare the results for the up valence quark. Again we see that the NNLO fit displays some significant sensitivity to the ATLAS Z p_T data, which disappears at aN3LO. Beyond this, there are similar behaviors at NNLO and aN3LO for the HERA combined structure function data and the BCDMS F_2^p data, which are two of the most sensitive

sets. This is because the aN3LO fit has smaller impact at high x with respect to the NNLO fit and does not significantly alter the details of flavor decomposition. Overall the u_V quark is well constrained, with no very large tensions between datasets. Indeed, all sensitivities across the whole range of x are $\lesssim 5$ units. The degree of tension is not obviously further reduced at aN3LO.

Finally, in Fig. 20 we compare the results for the $\bar{d}/\bar{u}$ antiquark ratio. Here we might expect little impact from aN3LO corrections. Indeed this is largely the case. In both cases the E866 Drell-Yan ratio data and the ATLAS 8 TeV W data provide constraints at primarily large and small x,

FIG. 19. L_2 sensitivities showing the six most sensitive datasets for the MSHT20 NNLO and aN3LO u_V PDFs.
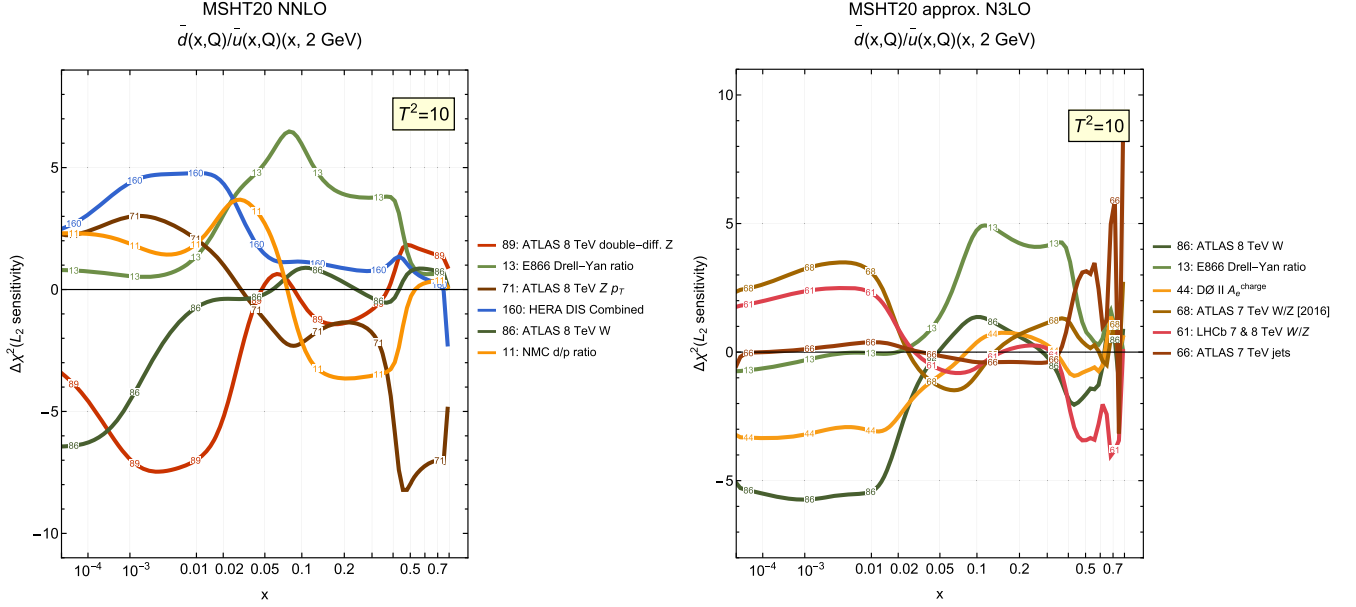
FIG. 20. L_2 sensitivities showing the six most sensitive datasets for the MSHT20 NNLO and aN3LO $\bar{d}/\bar{u}$ PDF ratios.

respectively. The former favors a reduced $\bar{d}/\bar{u}$ at large x (i.e., a larger asymmetry with $\bar{d} < \bar{u}$ at large x), while the latter favors an enhanced $\bar{d}/\bar{u}$ at small x. This differs slightly from the observations made for CT18 in Fig. 16 where it was observed that the E866 Drell-Yan ratio data had sensitivity of -9 units (positive pull on the isospin asymmetry ratio) in the region $0.04 < x < 0.2$. These differences partly reflect differences in parametrization choices, with MSHT employing a more flexible parametrization at high x and in turn obtaining a smaller $\bar{d}/\bar{u}$; see Fig. 28 and its discussion in the next section. However, a tension exists between these E866/NuSea Drell-Yan data and the (not fitted) E906 Seaquest data in pulls on $\bar{d}/\bar{u}$ at high x, which is visible in Fig. 16 (right), with the latter favoring instead an increased $\bar{d}/\bar{u}$. In MSHT20, the remaining datasets showing sensitivity to this PDF ratio include, once again, some ATLAS Z p_T sensitivity at NNLO, which is absent at aN3LO. At NNLO, there is also some (small) tension between HERA structure function data and ATLAS 8 TeV Z data at small x, with each showing pulls in opposite directions on the $\bar{d}/\bar{u}$ with sensitivities of around five units. This is absent at aN3LO, but replaced, to a lesser degree, by a smaller ATLAS 7 TeV W, Z and DØ electron asymmetry data tension. In general, only small tensions ($\lesssim 5$ units) exist at each order, and at both orders the E866 Drell-Yan ratio data provides the dominant constraint at $x > 0.1$.

VII. COMPARING SENSITIVITIES OF EXPERIMENTS AMONG DIFFERENT PDF FITS

Sections IV–VI provided examples of sensitivities computed for the PDF sets of each participating group. We saw

that, among the PDFs of one group, the similarities were maintained for the most influential datasets. At the same time, various differences were also noted and could be readily explained in some cases. We also saw that the differences are even more pronounced when comparing sensitivities to the same PDF flavor among the different groups, such as those in Secs. V and VI. These differences partly reflect nonidentical choices in the data selection and fitting methodologies reviewed in Sec. III. In this section, we will further elaborate on representative examples of differences and similarities among the PDFs of different groups, this time focusing on the plots of sensitivities to multiple PDF flavors for the shared experimental datasets.

To start, we return to the example of the CMS 8 TeV jet dataset, now showing L_2 sensitivities for CT18 and MSHT20 PDFs with $T^2 = 10$ at $Q = 100$ GeV side by side in Fig. 21.[8] In the figure, one readily observes that this dataset plays an important role for the gluon distribution (the black solid curve) and the charm distribution (the magenta dot-dot-dashed curve) that closely follows the former. There is less sensitivity to (anti)quark PDFs, as reflected by much weaker magnitudes of the other curves. We already remarked in Sec. II D that the CMS 8 TeV jet dataset plays an important role in constraining the gluon distribution in the CT18 analysis. More remarkably, by comparing the $S_{f,L2}$ patterns for the CT18As_Lat, MSHT20 NNLO, and aN3LO PDFs, we now confirm a very similar picture for these other PDF sets as well.

---

[8]The CT18 NNLO and MSHT20 NNLO sensitivities at 100 GeV, here compared against CT18As_Lat NNLO and aN3LO, have been also compared against the reduced fits in Appendix D of Ref. [18].
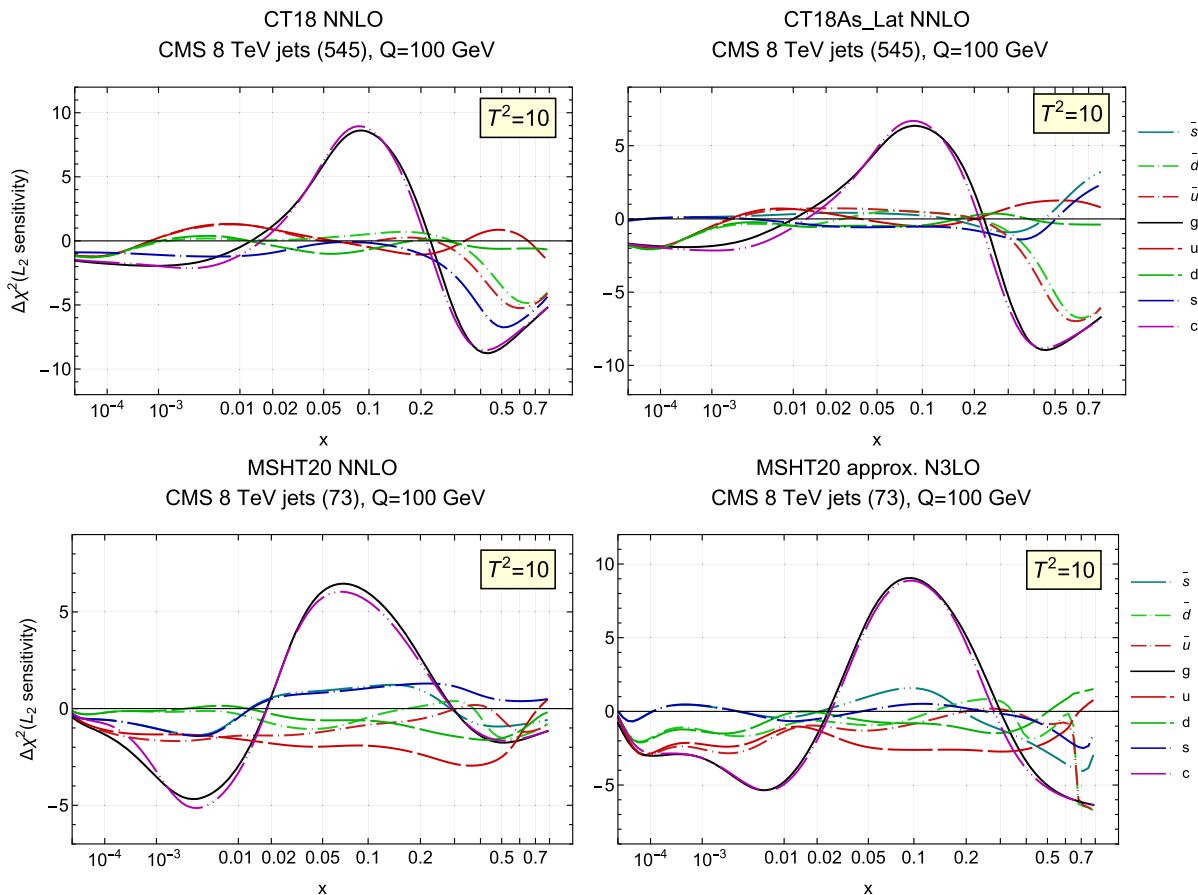
FIG. 21. Sensitivities of the CMS 8 TeV inclusive jet dataset for the PDFs with $T^2 = 10$ at $Q = 100$ GeV. Top: CT18 and CT18As_Lat at NNLO. Bottom: MSHT20 at NNLO and aN3LO.

The sensitivities tell us that, for all four PDF ensembles, at $x = 0.1$ an increase of the value of the gluon PDF by one sigma would be disfavored by up to ten units of χ^2. A smaller upward preference for the gluon at low x is another common feature.

This is one of the rare patterns where very good agreement is so apparent between the fitting groups. Note, however, that, for both CT and MSHT, the pull on the gluon from the CMS 8 TeV jet dataset is actually very different to that seen in the reduced fits. For the latter, the pulls on the gluon by the CMS 8 TeV jets can be seen in Fig. 13: the reduced fits both display the opposite pattern in the shape of the sensitivities across x, which in turn is tied to a different magnitude of the gluon PDF. The magnitude of the L_2 pulls is slightly smaller in the reduced sets, suggesting an improved agreement with the other datasets. This example illustrates that the tension can be enhanced in a global fit upon the addition of datasets to the reduced fit, and the patterns of the pulls can change, too.

Other interesting observations about the gluon come from the ATLAS and CMS $t\bar{t}$ data, displayed in Fig. 22. The top and middle plots concern the lepton + jets $t\bar{t}$ data and reveal sensitivity to the gluon (and charm), at small and valencelike regions in x, though at a smaller magnitude with respect to the CMS 8 TeV jet dataset. The patterns found by both CT18 and MSHT20 again agree between each other but differ significantly between ATLAS and CMS: their pulls on the gluon distribution at $x > 0.05$ are opposite (see the upper and middle rows). These $t\bar{t}$ data also affect the $\bar{u}$ and $\bar{d}$ distributions at large x, but with weaker pulls than other datasets. On the other hand, consistency of the pulls by the dilepton $t\bar{t}$ data from ATLAS and CMS at 8 TeV on the MSHT20 NNLO PDFs in the bottom row is quite good: the pulls are similar in shape, albeit with different magnitudes. Finally, the CT18 sensitivities to the gluon at $x \approx 0.01$ in the upper row of Fig. 22 are in agreement, both in their magnitudes and signs, with the weak preferences for a lower (higher) gluon exhibited by the ATLAS (CMS) $t\bar{t}$ datasets in the LM scan in Fig. 4, where the respective curves have IDs 12 and 13.

The final dataset influencing the gluon PDF that we will consider is the HERA combined DIS cross section data in Fig. 23. As it happened with the 8 TeV CMS jet data, the pulls by HERA combined data on the gluon distribution are inverted when going from the reduced to the full CT18 NNLO set, whereas for MSHT20 they remain quite similar.
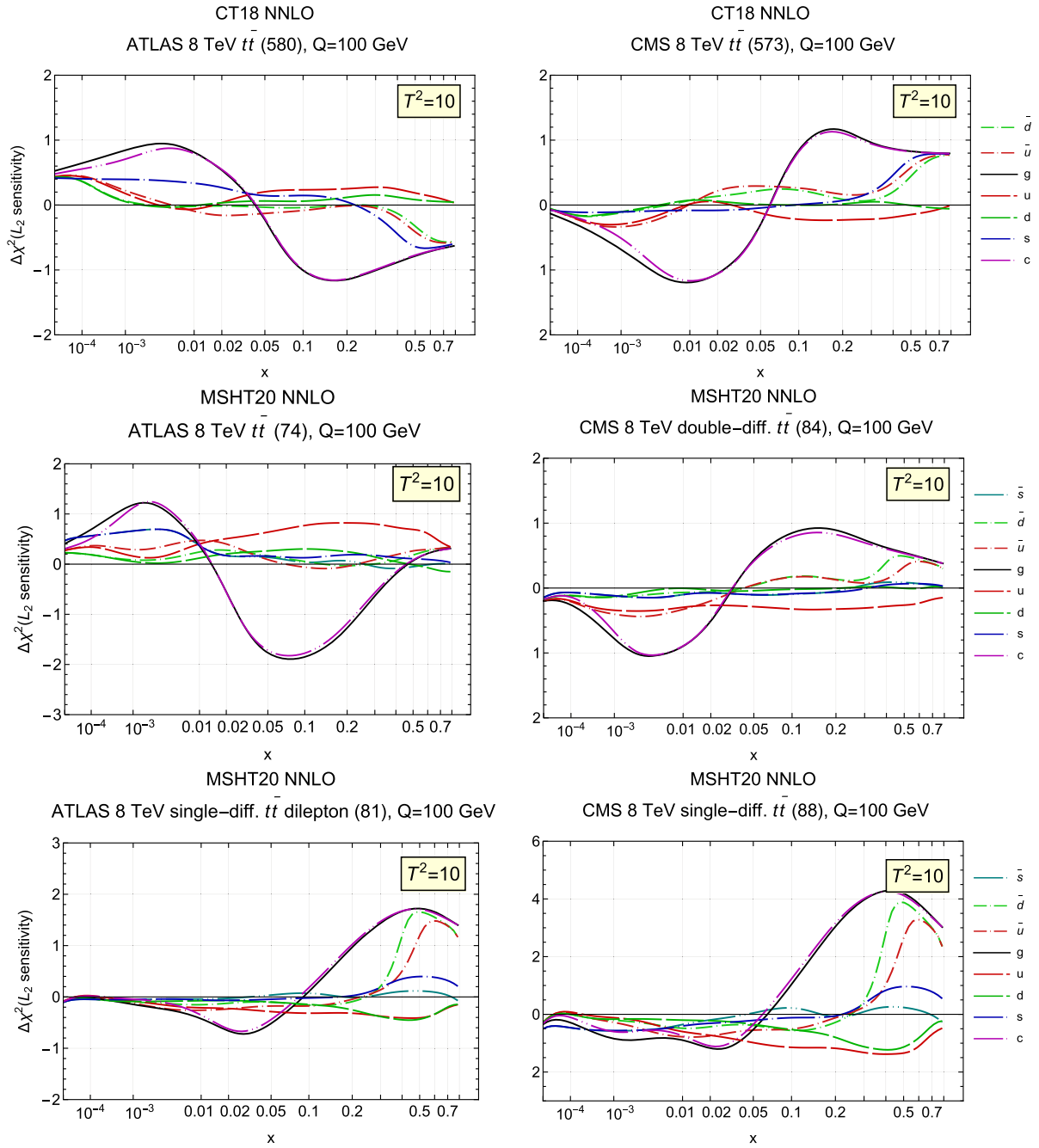
FIG. 22.   Sensitivities for ATLAS 8 TeV (left column) and CMS 8 TeV (right column) $t\bar{t}$ datasets in the CT18 (upper row) and MSHT20 (middle and lower rows) analyses at $Q = 100$ GeV.

Indeed, for the global fits the trends of the pull on the gluon from this dataset for CT18 and MSHT20 are distinctly different, particularly at small x values. As already noted, for MSHT aN3LO the trend is different to the NNLO fit and smaller in magnitude, due to the different and reduced tensions, while for the ATLASpdf21 NNLO set the pulls on the gluon from the HERA data are smaller, reflecting the dominant weight these data have in the ATLAS dataset.

Ultimately, the patterns of sensitivities to the gluon reflect several factors, e.g., the shapes and flexibility of

the relevant PDFs, implementation of the theoretical cross sections, and implementation of systematic uncertainties for the considered experiment. Take, for instance, the gluon PDFs for the corresponding PDF sets, shown in Fig. 24 at $Q = 2$ GeV in the left panel and as the ratios to the MSHT20 NNLO gluon PDF at $Q = 100$ GeV in the right panel. Going back to the CMS 8 TeV jet data in Fig. 21, we observed that NNLO PDFs predict very similar gluon sensitivities at $x > 0.02$. In Fig. 24, we observe that the NNLO gluon PDFs agree well at $x > 0.02$, and hence it is
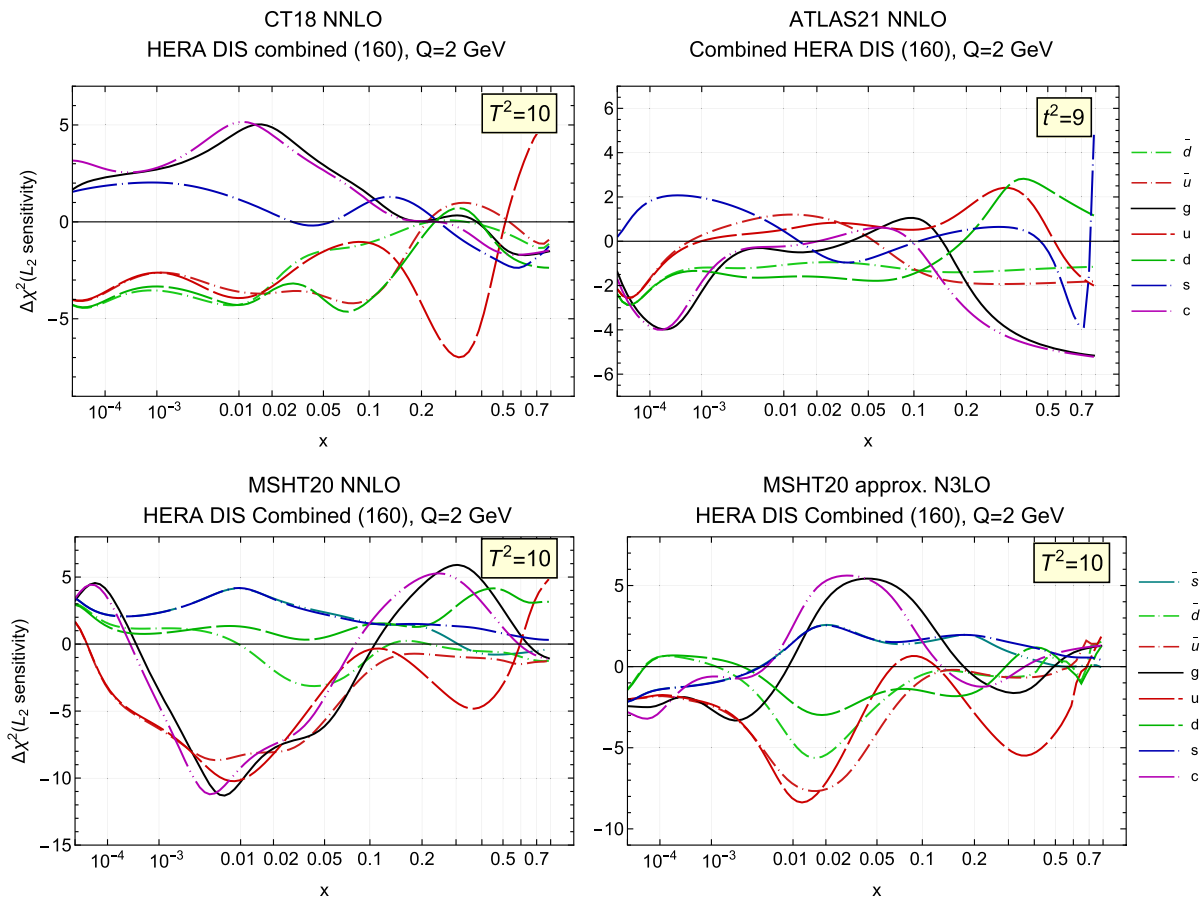
FIG. 23. Sensitivities of the combined HERA DIS data. Top left: CT18 NNLO, top right: ATLASpdf21 NNLO, bottom left: MSHT20 NNLO, bottom right: MSHT20aN3LO.

not surprising that the sensitivities for the CMS 8 TeV jet data in the respective CT18 and MSHT20 analyses (which use close implementations of theoretical cross sections and systematic errors in this case) also agree.

For the combined HERA DIS data in Fig. 23, we pointed out large differences in the gluon sensitivities at $x < 0.1$. For this experiment, the groups use different heavy-quark schemes and χ^2 definitions, while the interplay of the pulls from various x regions depends on the assumed parametrization forms. (Recall that the MSHT parametrization for the gluon is more flexible than the CT one.) As influential are the shapes of the central PDFs, which are distinctly apart at $x \lesssim 10^{-3}$ in Fig. 24. In the ATLASpdf21 case, the small-x region is affected by a cut of $Q^2 > 10$ GeV^2 imposed on the fitted HERA data because of doubts about the adequacy of NNLO DGLAP to describe the HERA data at low x and Q below this cut. The ATLASpdf21 PDFs are designed for use at higher x, $x > 10^{-4}$, and the absence of the low-x data explains the reduction of the gluon at very small x and its compensating enhancement at moderate x. The deviation of the ATLASpdf21 from the others at the lowest x values demarcates the region where low-x physics effects may need to be considered.

Now turning to the sensitivities to (anti)quarks, it can be seen in Fig. 23 that the HERA DIS data also have significant pulls on PDFs other than the gluon. In particular, they affect the up and down distributions. For CT18, the data favors larger u and d PDFs at small- and mid/valence-x values at $Q = 2$ GeV, and smaller up distribution at very large x. While the trend for the u PDF is also negative for MSHT20 NNLO (bottom left), the pulls on the d distribution are smaller and positive, though there is now more pull on the strange quark. The magnitude of the pulls from HERA DIS data for the u and $\bar{u}$ of the MSHT20 L_2 plots at small x is similar to that of the gluon. The signs of pulls of MSHT20aN3LO align quite well to those of CT18 NNLO and MSHT20 NNLO, but are generally a little reduced. Overall, as with the gluon, ATLASpdf21 shows less tension with the HERA DIS data for the quarks, a consequence of its smaller sample of datasets.

Another dataset that is known to be strongly constraining the PDFs is the 7 TeV ATLAS W, Z precision data. Its L_2 sensitivity plots are shown in Fig. 25. Since the rapidity dependence of these very precise data constrain the details of the x dependence of all flavors of quarks, and hence, via evolution, also the gluon, there are significant pulls on a wide variety of PDFs. Indeed, the indirect gluon constraint
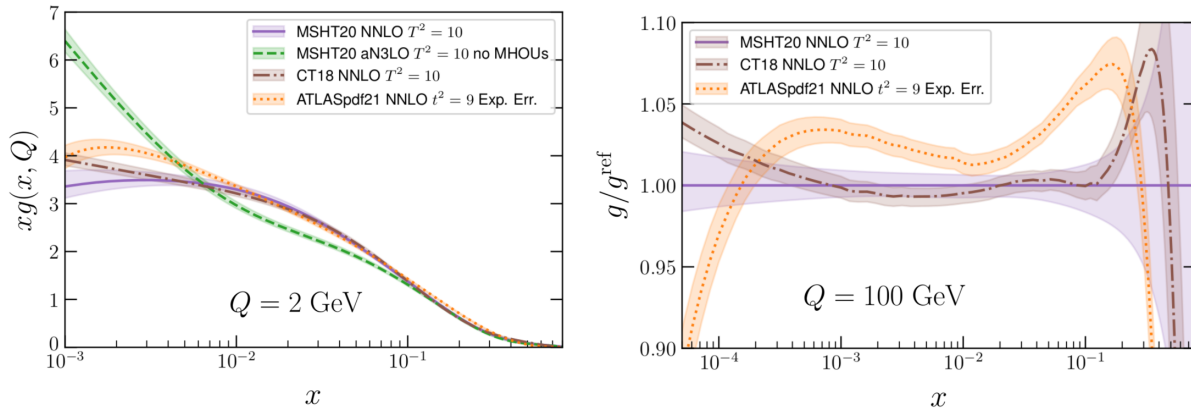
FIG. 24. Comparison of the gluon PDFs. Left: absolute PDFs at $Q = 2$ GeV shown for MSHT20 NNLO (purple, solid), MSHT20aN3LO (green, dashed), CT18 NNLO (brown, dot-dashed) and ATLASpdf21 NNLO (orange, dotted). Right: ratios to the central MSHT20 NNLO at $Q = 100$ GeV. Notice that MSHT20aN3LO and CT18As_Lat are not displayed. All ensembles are given for $T^2 = 10$ ($t^2 = 9$ for ATLAS).

does result in a noticeable set of pulls on this PDF, generally opposing that from HERA data at small x, particularly for ATLASpdf21. The very specific constraint this dataset brings, though, is on the strange quark, which is not directly constrained by most of the other datasets in a

global fit. In particular, it has long been known that NuTeV and CCFR combined dimuon data favor a smaller strangeness at $x \sim 0.02$, and historically this combination of data has been the main constraint on the strange quark PDF. However, this is now compensated by an upwards pull on
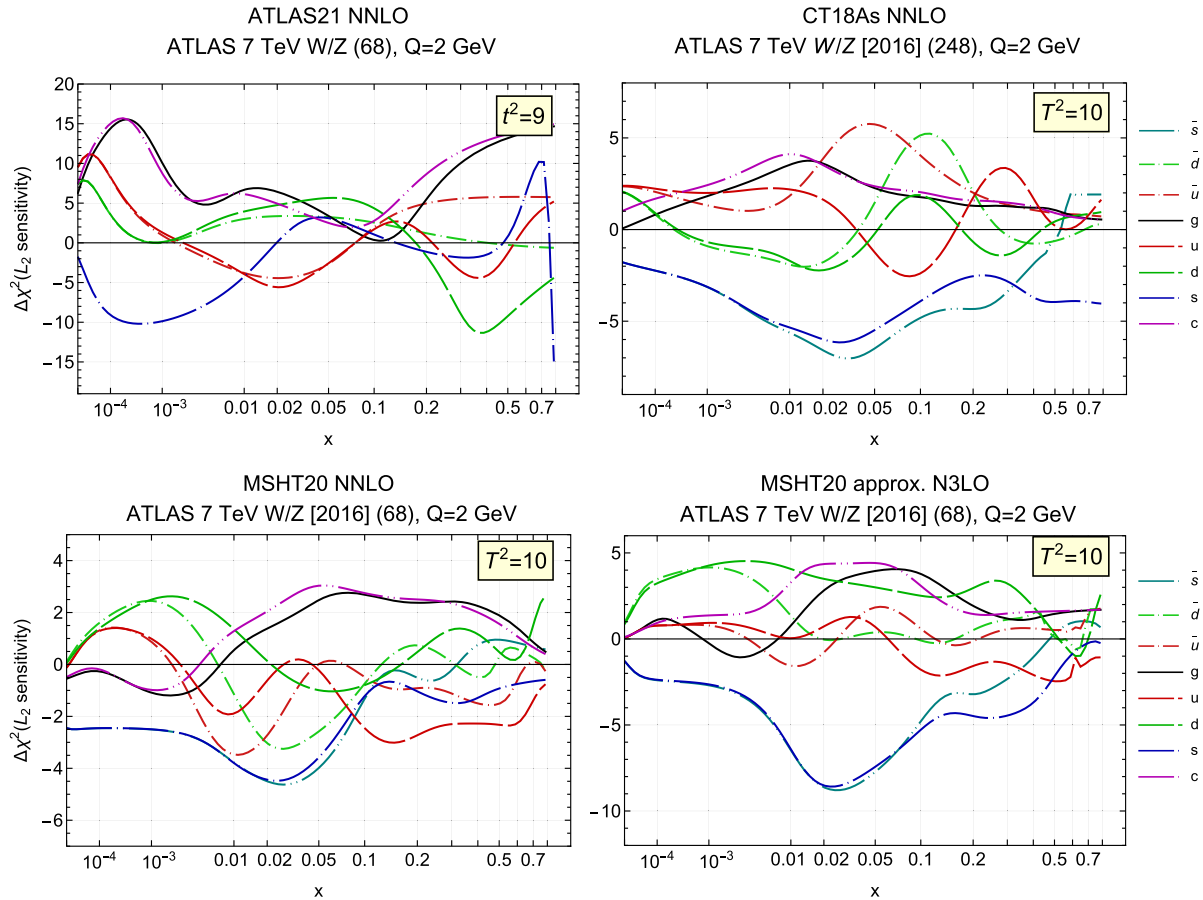
FIG. 25. Sensitivities of ATLAS 7 TeV W/Z data at $Q = 2$ GeV: ATLASpdf21 NNLO (top left), CT18As NNLO (top right), MSHT20 NNLO (bottom left), MSHT20aN3LO (bottom right).
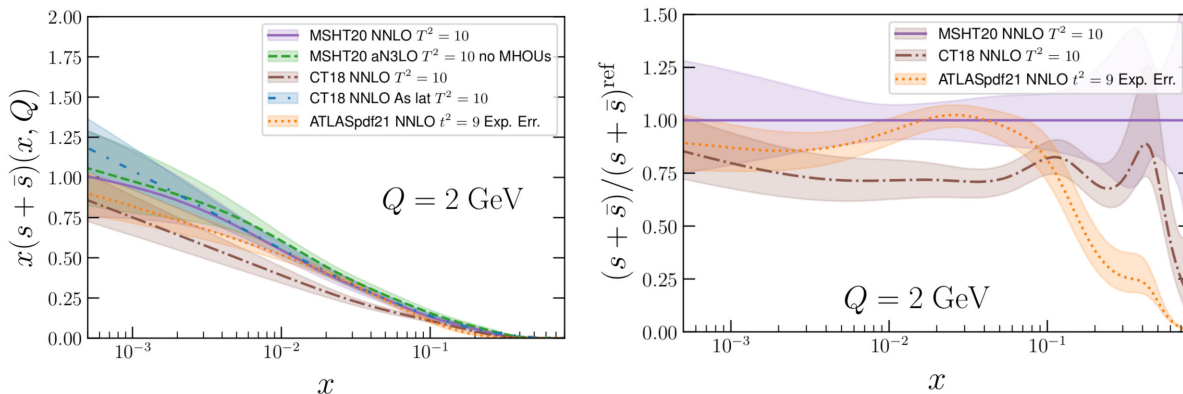
FIG. 26. Comparison of the strangeness PDFs. Left: $s + \bar{s}$ PDFs shown for MSHT20 NNLO (purple, solid), MSHT20aN3LO (green, dashed), CT18 NNLO (brown, dot-dashed), CT18As_Lat NNLO (blue, large dot-dashed), and ATLASpdf21 NNLO (orange, dotted). Right: ratios to the central MSHT20 NNLO (MSHT20aN3LO and CT18As_Lat are not displayed, as their uncertainties are similar to their corresponding nominal sets). All ensembles are given for $T^2 = 10$ ($t^2 = 9$ for ATLAS).

the strangeness (corresponding to negative L_2 sensitivity) from ATLAS 7 TeV W, Z over a generally wide range of x values. This is evident for both the CT18As PDF set (the default CT18 PDFs do not fit to this dataset—as apparent from its reduced strangeness seen in Fig. 26) and the MSHT20 sets (where there is also an opposing pull from CMS 7 TeV $W + c$ data for MSHT20aN3LO). The pull in the ATLASpdf21 sensitivity plot is much smaller at relatively high x values, though is also negative at very small x values, where these data are opposed by the 8 TeV Z3D data, which has opposite tendencies.

A comparison of the distributions for total strangeness is given in Fig. 26. All PDF sets are strongly influenced by the ATLAS W, Z data, and hence agree quite well in the region covered by data. Noticeably there is a smaller strangeness PDF for CT18 at $x \sim 0.01$–0.1, mostly due to the absence of the ATLAS W, Z dataset in the CT18 nominal fit, as opposed to the CT18A(Z) alternative sets.

The final data we consider are the E866 fixed-target Drell-Yan cross section ratios, σ^{pd}/σ^{pp}, for which the

corresponding sensitivity plots are shown in Fig. 27. The CTEQ-TEA publications singled out the E866 ratio as a small experiment that provides a particularly sensitive probe of flavor-symmetry breaking in the quark sea at high x, specifically by constraining $\bar{d}(x, Q)/\bar{u}(x, Q)$ [or, alternatively, $\bar{d}(x, Q) - \bar{u}(x, Q)$] above $x = 0.01$ and at Q of a few GeV. Indeed, the PDFSENSE analyses in Refs. [5,6] noted that the E866 ratio is a top experiment in terms of the sensitivity per data point, and it is especially valuable for constraining a sea PDF combination that can be computed on the lattice [6].

These observations remain partially true with respect to the total sensitivity of the E866 ratio to individual PDFs presented in Fig. 27. At $x > 0.05$—the region of kinematic coverage by the experiment—the pulls in CT18 are in opposing directions for the $\bar{u}$ and $\bar{d}$ flavors, while MSHT20 very predominantly prefers lower values of both d and $\bar{d}$, with weaker sensitivity to u and $\bar{u}$. (The overall pulls are not large, but this is because this dataset has few points and imposes a strong constraint on the central PDF, so little
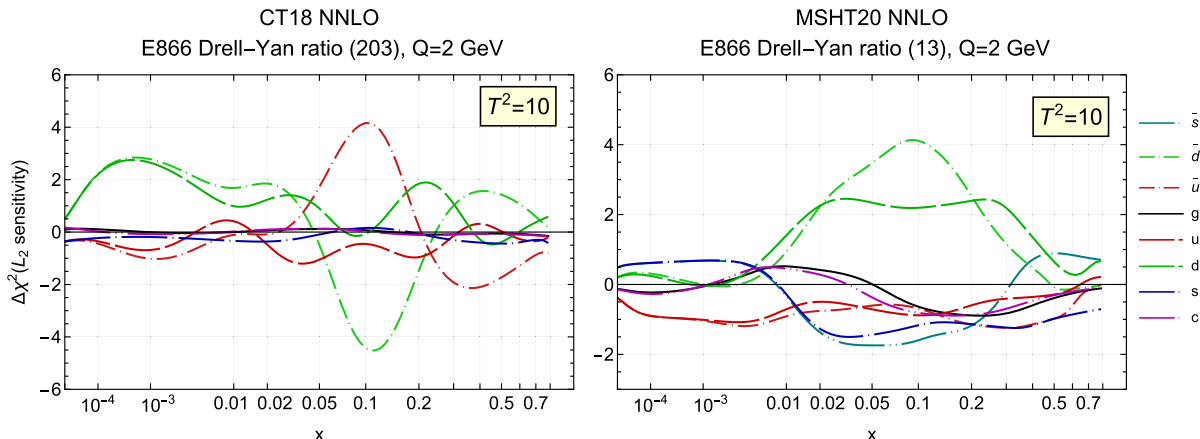

FIG. 27. Sensitivities of the E866 pd/pp Drell-Yan ratio at $Q = 2$ GeV. Left: CT18 NNLO, right: MSHT20 NNLO.
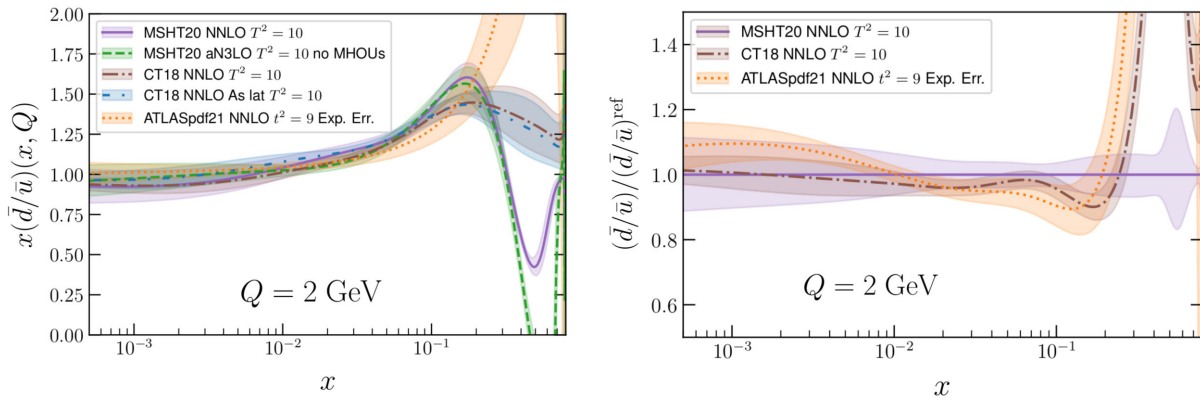
FIG. 28. Comparison of the $\bar{d}/\bar{u}$ PDF ratios. Left: absolute ratios of PDFs. Right: ratios to central MSHT20 NNLO. Color code and legend as in Fig. 24.

tension exists.) At least some of this difference is likely due to the PDF parametrizations. Different flavor combinations are parametrized at input by CT and MSHT, with the former parametrizing $\bar{u}$ and $\bar{d}$ separately and requiring the asymptotic $x \to 0$ or $x \to 1$ behavior of $\bar{u}$ and $\bar{d}$ to coincide (see Appendix C in Ref. [17]). This places some constraint on $\bar{d}/\bar{u}$ at high x. Instead, MSHT effectively parametrizes the sum and ratio $\bar{d}/\bar{u}$, with the latter having a somewhat freer high-x behavior as determined by the PDF fit. The plots of $\bar{d}/\bar{u}$ in Fig. 28 show that consequently there is close agreement between all PDFs below $x = 0.2$, while substantial differences persist at larger x with poor experimental constraints. The E906/SeaQuest ratio largely compensates the pulls of the E866 ratio in the CT18As_Lat fit; see Fig. 16 and Fig. SM.3 in the Supplemental Material [166]. Finally, we note that the E866 sensitivities also propagate to $x < 0.01$ (beyond the actual kinematic reach), most likely through extrapolation according to the given functional form. Here, both CT18 and MSHT PDF sets show a pull towards higher $\bar{u}$ compared to $\bar{d}$. However, this derives mainly from pull on $\bar{d}$ (and d) for CT18 and on $\bar{u}$ (and u) for MSHT20.

The examples above illustrate how the sensitivities can reveal subtle dynamics among the experimental constraints in PDF fits. In the Supplemental Material [166], we collect such comparisons for several other interesting experiments: LHCb production of W, Z bosons at 8 TeV; DØ Run-2 charge asymmetry at the lepton and vector boson level; additional plots of PDF ratios for the E866 and E906 measurements of σ_{pd}/σ_{pp} in Drell-Yan process; and $t\bar{t}$ production datasets in the ATLASpdf21 analysis.

VIII. CONCLUSIONS

In this article, we explored a new statistical technique to "look inside" the global fits that determine parton distribution functions for a variety of applications. Such examinations are important both for the PDF fitter and the PDF user, for instance, to understand the role of each fitted dataset and emergence of new constraints as more datasets are added.

Tools such as the LM method have been extensively used by the PDF fitters for this purpose, but are not accessible to the PDF user. In addition, the LM scan is CPU-intensive and is limited to a single kinematic point at a time.

To elucidate these issues, authors of the ATLASpdf, CT, and MSHT NNLO PDF ensembles have presented a study using the L_2 sensitivity method to explore the relative importance of each dataset used in their PDF fits. The L_2 sensitivity, or $S_{f,L2}$, can be easily computed using the Hessian error PDFs published along with each central PDF set. These error PDF sets are already extensively used in calculations and comparisons at the LHC.

$S_{f,L2}$ estimates statistical pulls of one or several datasets on the best-fit PDF at a particular x and Q point, parton-parton luminosity, or another PDF-dependent quantity. It reflects both the power of that dataset to constrain the PDF, i.e., its statistical power in the fit, as well as the correlation between the PDF and log-likelihood of the dataset. Unlike the LM scan, the L_2 sensitivity covers the entire partonic x range in one (quick) evaluation. An examination of $S_{f,L2}$ indicates which experiments are important in a given x range, and which datasets are competing or in "tension" with each other, i.e., one preferring to pull the PDF up (resulting in a negative $S_{f,L2}$), and another preferring to pull it down (a positive $S_{f,L2}$). Opposite pulls on the best-fit PDFs by the subsets of global data are present in any analysis. Excessive pulls indicate systematic disagreements, and $S_{f,L2}$ estimates the degree of such tensions. The sum of all pulls, i.e., the values of $S_{f,L2}$, should be close to zero according to the inequality in Eq. (12). Checking this inequality in fact serves as a validation test of the Hessian error set. While in this study we focused on special PDFs obtained with the simplified global tolerance of about 10, the method works with the standard error definitions as well.

The L_2 sensitivity method is fast because it relies on the linear approximation for the χ^2 functions of individual experiments in the close neighborhood of the global minimum. For the experiments that are not perfectly

consistent, the linear pieces dominate their χ^2 near the global minimum. That said, it is not uncommon that a highly constraining experiment agrees well with the best-fit PDF and renders a low $S_{f,L2}$ value in some x region while still imposing strong constraints on the net PDF uncertainty via quadratic and higher-order terms. This has been illustrated by the influence of the CMS 8 TeV jets dataset on the gluon PDF at $x = 0.01$ vs $x = 0.3$ at $Q = 125$ GeV in Sec. II D. The L_2 sensitivity, in its role of the linearized indicator of the local agreement among the experiments, thus offers a tool that complements the methods testing quadratic dependence.

The nonlinear χ^2 contributions become numerically important when some error sets substantially deviate from the global minimum. This happens when the L_2 sensitivity is computed using Monte Carlo replicas, as discussed in Sec. II B. The Appendix shows how to organize sampling of Monte Carlo replicas in many parameter dimensions to prevent large deviations from the global minimum and obtain meaningful sensitivities.

In this work and on the companion website [30], we compare L_2 sensitivities from the non-global and global Hessian fits of three groups in the formats showing either the sensitivities of leading experiments to a given PDF flavor or PDF combination, or for all PDF flavors and a particular experiment. We demonstrate how the L_2 sensitivity confirms and expands conclusions made based on the traditional examination of χ^2 for experiments and LM scans. The L_2 sensitivity can provide information on the impact of PDF parametrization choices and different perturbative orders. In this paper, this was examined in the context of comparing CT18, CT18As, and CT18As_Lat NNLO PDFs (with the latter one including independent parametrizations of strange quark and antiquark PDFs together with the lattice QCD prediction for the large-x strangeness asymmetry), as well as the MSHT20 NNLO and aN3LO PDFs. One interesting result is the examination of pulls on strangeness and other sea PDFs in the various fits using the same L_2 metric. Another is the evidence of reduced tensions in the MSHT aN3LO PDFs, compared to the corresponding NNLO ones; this is most clearly seen in the strong decrease of the value of $S_{f,L2}$ for the gluon distribution for the ATLAS Z p_T distribution from NNLO to aN3LO, perhaps indicating a release of tension by transitioning to the (approximate) higher order in the PDF fit.

To conclude, the L_2 sensitivities offer any user of Hessian PDFs valuable insights using the tools outside of the PDF fit. Additional comparisons are provided on the website [30] and in the Supplemental material [166]. The website also provides a C++ program that computes or plots sensitivities using tabulated vectors of χ^2 and error PDFs in either LHAPDF or tabulated format.

ACKNOWLEDGMENTS

APPENDIX: FROM A HESSIAN TO A MONTE CARLO PDF ENSEMBLE

In this article, we focused on the exploration of the L_2 sensitivity using error PDF sets in the Hessian approximation. We pointed out in Sec. II B that the L_2 sensitivity can also be defined in the MC method, while its calculation in this method requires reducing the impact of nonlinearities.

If $X(\vec{R})$ is a nonlinear function of the PDF parameters $\vec{R}$, the cumulative probability $\mathcal{P}(X(\vec{R}))$ is generally not the same as $\mathcal{P}(\vec{R})$, implying that the PDF parameters at a certain confidence level for $\vec{R}$ correspond to a different confidence level for $X(\vec{R})$ [24]. As a flip side, a linear function $X(\vec{R})$ of the normally distributed $\vec{R}$ is also normally distributed. Consequently, the error PDFs work the best when the variations of $X(\vec{R})$ from the best-fit value X_0 are approximately linear. An MC replica ensemble that is generated from a Hessian eigenvector ensemble by following one of the available methods [24,26] then closely reproduces the Hessian PDF uncertainty and correlations.

In some situations the normally sampled MC replicas lead to unphysical results, for example, by predicting negative cross sections with a nonzero probability at large Bjorken x. In this case, MC sampling according to an alternative, e.g., log-normal, probability distribution is needed to get the physical behavior [24].

The log-likelihood function χ_E^2, with its quadratic and higher-order terms, is particularly prone to nonlinear effects. Indeed, nearly all MC replicas $f^{(k)}$ generated with a standard method deviate from the best fit by several standard deviations in some directions because of the large number of PDF parameters $D \sim 30$ in a typical fit [24]. For these replicas, the $\chi_E^2(f^{(k)})$ functions are large and nonlinear. To demonstrate the equivalence of the Hessian and MC sensitivities, one must resort to an alternative method that stays close enough to the global minimum to suppress the quadratic component.

When we use the normal sampling, we consider N_{rep} sets of the PDF vectors $\vec{R}^{(k)} \equiv \{R_1^{(k)}, \ldots, R_D^{(k)}\}$ that are randomly sampled from the D-dimensional standard normal distribution $\mathcal{N}_D(0, 1)$ associated with the probability density $\mathcal{P}(\vec{R}) = (2\pi)^{-D/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{D} R_i^2\right)$. For a PDF $f(\vec{R})$, the MC replica can be constructed as[9]

$$f^{(k)} \equiv f_0 + \sum_{i=1}^{D} \frac{f_{+i} - f_{-i}}{2} R_i^{(k)}, \tag{A1}$$

with only the linear term kept on the right-hand side. For $N_{\text{rep}} \to \infty$, the MC estimators for the mean PDF $\langle f \rangle$, the PDF uncertainty $\delta_{\text{MC}} f$, and PDF-PDF correlation $C_{\text{MC}}(f_1, f_2)$, given respectively in Eqs. (6)–(8), coincide with the Hessian f_0, $\delta_H f$ and $C_H(f_1, f_2)$ in Eqs. (3) and (4) [167]. To demonstrate this, we use

$$\lim_{N_{\text{rep}} \to \infty} \langle f \rangle = \lim_{N_{\text{rep}} \to \infty} \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} f^{(k)} = \int d^D R f(\vec{R}) \mathcal{P}(\vec{R}) \tag{A2}$$

and

$$\langle R_i \rangle = 0, \qquad \langle R_i^2 \rangle = 1, \qquad \langle \vec{R}^2 \rangle = \sum_{i=1}^{D} \langle R_i^2 \rangle = D$$

$$\text{for } \vec{R} \sim \mathcal{N}_D(0, 1). \tag{A3}$$

The χ_E^2 function for such replicas contains a nonlinear component $\mathcal{O}(R^2)$ that is at least comparable to the linear term if D is large:

---

[9]In some replica generation prescriptions [24], the random variation for the MC replica also contains a constant shift from the central value that vanishes at $N_{\text{rep}} \to \infty$. This shift does not modify our conclusions.

$$\chi_E^{2(k)} \equiv \chi_E^2(f^{(k)}) \approx \chi_E^2(f_0) + \sum_{i=1}^{D} \frac{\chi_{E,+i}^2 - \chi_{E,-i}^2}{2} R_i^{(k)} + \mathcal{O}(R^2). \tag{A4}$$

We wish to suppress the spurious nonlinear component for the reasons discussed above. To stay in the close vicinity of the global minimum, where the $\mathcal{O}(R^2)$ component is small, we can generate the MC replicas according to the uniform distribution $\mathcal{U}_D(R^2 = 1)$ on the tolerance hypersphere instead of $\mathcal{N}_D(0, 1)$. The probability density associated with $\mathcal{U}_D(R^2 = 1)$ is $\mathcal{P} = 1/\Omega_D$, where $\Omega_D = 2\pi^{D/2}/\Gamma(D/2)$ is the full solid angle in D dimensions. We also have

$$\langle R_i \rangle = 0, \quad \langle R_i^2 \rangle = 1/D, \quad \langle \vec{R}^2 \rangle = 1 \quad \text{for } \vec{R} \sim \mathcal{U}_D(R^2 = 1), \tag{A5}$$

from which it follows that the uncertainty values on the two probability distributions are related as

$$\delta_{\text{MC}} f|_{\mathcal{N}_D(0,1)} = \sqrt{D} \delta_{\text{MC}} f|_{\mathcal{U}_D(R^2=1)} + \mathcal{O}(R^2). \tag{A6}$$

We now discard the $\mathcal{O}(R^2)$ terms on the right-hand side, retaining only the sought linear part of the expectation value. This prescription gives us a trustworthy MC formula for the L_2 sensitivity,

$$S_{f,L2}^{\text{MC}}(E) = \sqrt{D} (\delta_{\text{MC}} \chi_E^2) C_{\text{MC}}(f, \chi_E^2), \tag{A7}$$

where the MC replicas are generated from the uniform distribution $\mathcal{U}_D(R^2 = 1)$ on the tolerance hypersphere.

The MC sensitivity in Eq. (A7) is close to the Hessian one because (i) the means of the linear terms in Eq. (A7) are related by a constant factor, since both $\mathcal{N}_D(0, 1)$ and $\mathcal{U}_D(R^2 = 1)$ distributions are spherically symmetric, and (ii) the nonlinear terms can be suppressed on the tolerance hypersphere by choosing a small T^2.

To illustrate that this prescription is successful, we compare the Hessian L_2 sensitivities for CT18 NNLO and those computed with the MC replicas that were generated using either $\mathcal{U}_D(R^2 = 1)$ or $\mathcal{N}_D(0, 1)$. Figure 29 presents such comparison for the BCDMS $F_2^p(x, Q^2)$ dataset [ID = 101]. The MC representation obtained with $\mathcal{U}_D(R^2 = 1)$ in Fig. 29(b) closely reproduces the Hessian sensitivity in Fig. 29(a). Similar level of agreement has been observed for the L_2 sensitivities of the other experiments.

In contrast, the sensitivities for the MC representation based on $\mathcal{N}_D(0, 1)$ in Fig. 29(c) show larger deviations from the Hessian result, with especially manifest differences present in the d- and u-quark sensitivities at large x. These discrepancies reflect the large total number D of the PDF parameters, which should be contrasted with
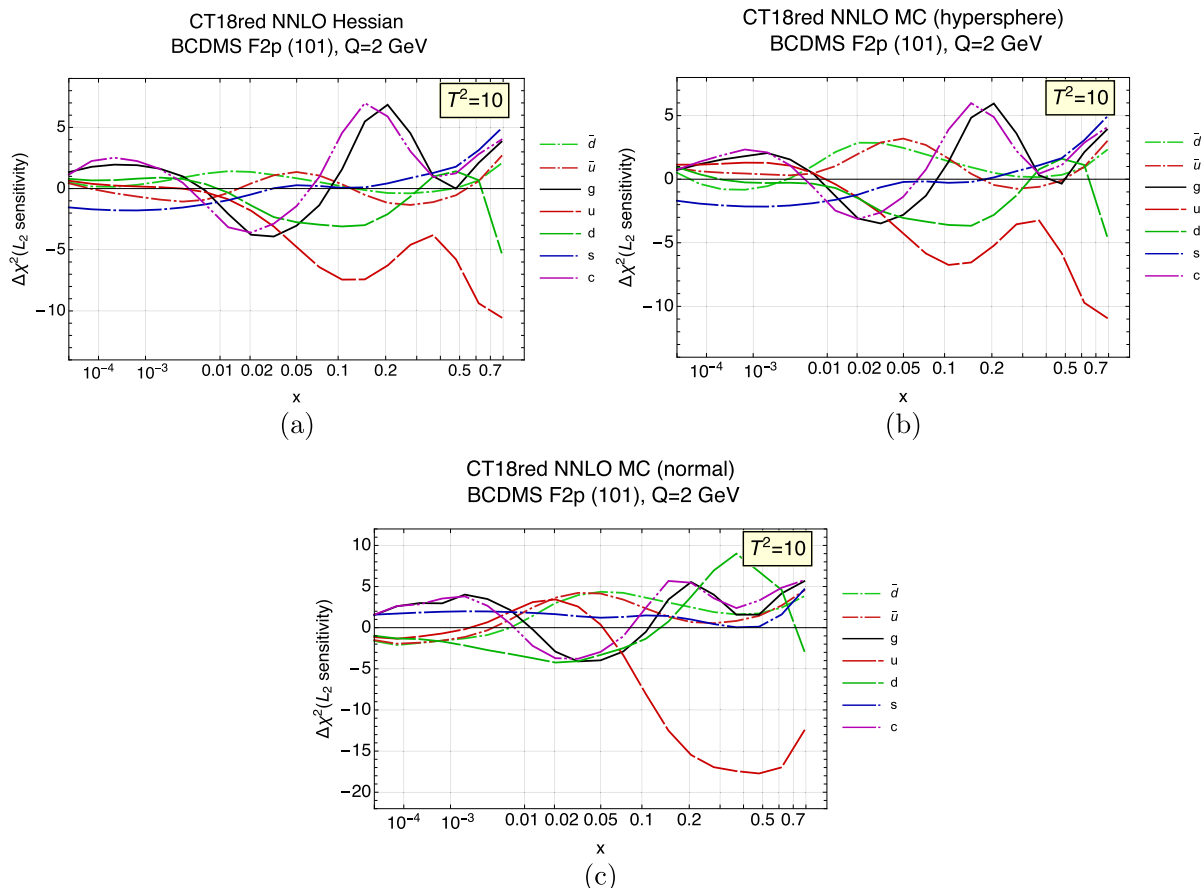
FIG. 29. L_2 sensitivities for the BCDMS $F_2^p(x, Q^2)$ dataset and CT18 NNLO PDFs with $T^2 = 10$ in (a) the Hessian representation; (b),(c) MC representations generated with the uniform distribution $\mathcal{U}_D(R^2 = 1)$ on the tolerance hypersphere and the standard normal distribution $\mathcal{N}_D(0, 1)$.

the good agreement between the Hessian and MC sensitivities in the one-dimensional case demonstrated in Sec. II C.

Estimation of the L_2 sensitivity with the MC replicas therefore requires a sampling algorithm that populates well the close vicinity of the global minimum and follows an isotropic probability distribution that can be related to $\mathcal{N}_D(0, 1)$ in the linear approximation. These conditions are not automatic when using the commonly available MC replicas. We have shown how to realize them when converting a Hessian error set into an MC one. The form of the χ^2 or cost function during training of MC replicas may differ from the one used in the postfit computations of the L_2 sensitivity. The dependence on the χ^2 definition is not insignificant [162].

[1] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, and W.-K. Tung, Phys. Rev. D **65**, 014013 (2001).

[2] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W.-K. Tung, J. High Energy Phys. 07 (2002) 012.

[3] W. T. Giele and S. Keller, Phys. Rev. D **58**, 094023 (1998).

[4] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, Phys. Rev. D **78**, 013004 (2008).

[5] B.-T. Wang, T. J. Hobbs, S. Doyle, J. Gao, T.-J. Hou, P. M. Nadolsky, and F. I. Olness, Phys. Rev. D **98**, 094030 (2018).

[6] T. J. Hobbs, B.-T. Wang, P. M. Nadolsky, and F. I. Olness, Phys. Rev. D **100**, 094040 (2019).

[7] J. C. Collins and J. Pumplin, arXiv:hep-ph/0105207.

[8] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. C **63**, 189 (2009).

[9] K. Kovařík, P. M. Nadolsky, and D. E. Soper, Rev. Mod. Phys. **92**, 045003 (2020).

[10] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, and W.-K. Tung, Phys. Rev. D **65**, 014012 (2001).

[11] H. Paukkunen and P. Zurita, J. High Energy Phys. 12 (2014) 100.

[12] C. Schmidt, J. Pumplin, and C.-P. Yuan, Phys. Rev. D **98**, 094005 (2018).

[13] T.-J. Hou, Z. Yu, S. Dulat, C. Schmidt, and C.-P. Yuan, Phys. Rev. D **100**, 114024 (2019).

[14] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali (NNPDF Collaboration), Nucl. Phys. **B849**, 112 (2011); **B854**, 926 (E) (2012); **B855**, 927(E) (2012).

[15] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, Nucl. Phys. **B855**, 608 (2012).

[16] N. Sato, J. Owens, and H. Prosper, Phys. Rev. D **89**, 114020 (2014).

[17] T.-J. Hou *et al.*, Phys. Rev. D **103**, 014013 (2021).

[18] R. D. Ball *et al.* (PDF4LHC Working Group), J. Phys. G **49**, 080501 (2022).

[19] A. Accardi, T. J. Hobbs, X. Jing, and P. M. Nadolsky, Eur. Phys. J. C **81**, 603 (2021).

[20] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **82**, 438 (2022).

[21] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 341 (2021).

[22] T.-J. Hou, H.-W. Lin, M. Yan, and C.-P. Yuan, Phys. Rev. D **107**, 076018 (2023).

[23] J. McGowan, T. Cridge, L. A. Harland-Lang, and R. S. Thorne, Eur. Phys. J. C **83**, 185 (2023).

[24] T.-J. Hou *et al.*, J. High Energy Phys. 03 (2017) 099.

[25] P. M. Nadolsky and Z. Sullivan, eConf **C010630**, P510 (2001).

[26] G. Watt and R. S. Thorne, J. High Energy Phys. 08 (2012) 052.

[27] J. Gao and P. Nadolsky, J. High Energy Phys. 07 (2014) 035.

[28] S. Carrazza, S. Forte, Z. Kassabov, J. I. Latorre, and J. Rojo, Eur. Phys. J. C **75**, 369 (2015).

[29] S. Carrazza, S. Forte, Z. Kassabov, and J. Rojo, Eur. Phys. J. C **76**, 205 (2016).

[30] For the online plotter of ATLAS21, CT18, and MSHT20 sensitivities, see https://metapdf.hepforge.org/L2.

[31] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Eur. Phys. J. C **73**, 2311 (2013).

[32] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Eur. Phys. J. C **78**, 473 (2018).

[33] M. Guzzi, K. Xie, T.-J. Hou, P. Nadolsky, C. Schmidt, M. Yan, and C.-P. Yuan, Proc. Sci. EPS-HEP2021 (**2022**) 370.

[34] D. E. Soper and J. C. Collins, arXiv:hep-ph/9411214.

[35] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Eur. Phys. J. C **75**, 580 (2015).

[36] M. Aaboud *et al.* (ATLAS Collaboration), Eur. Phys. J. C **77**, 367 (2017).

[37] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 12 (2017) 059.

[38] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **79**, 760 (2019).

[39] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 05 (2018) 077; 10 (2020) 48.

[40] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **79**, 847 (2019).

[41] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **76**, 538 (2016).

[42] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Rev. D **94**, 092003 (2016); **101**, 119901(A) (2020).

[43] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **79**, 1028 (2019); **80**, 1092(E) (2020).

[44] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 04 (2019) 093.

[45] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 09 (2017) 020.

[46] A. C. Benvenuti *et al.* (BCDMS Collaboration), Phys. Lett. B **223**, 485 (1989).

[47] M. Arneodo *et al.* (New Muon Collaboration), Nucl. Phys. **B483**, 3 (1997).

[48] M. Arneodo *et al.* (New Muon Collaboration), Nucl. Phys. **B487**, 3 (1997).

[49] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock, and A. Bodek, Phys. Lett. B **282**, 475 (1992).

[50] L. W. Whitlow, S. Rock, A. Bodek, E. M. Riordan, and S. Dasu, Phys. Lett. B **250**, 193 (1990).

[51] M. R. Adams *et al.* (E665 Collaboration), Phys. Rev. D **54**, 3006 (1996).

[52] F. D. Aaron *et al.* (H1 and ZEUS Collaborations), J. High Energy Phys. 01 (2010) 109.

[53] A. Aktas *et al.* (H1 Collaboration), Eur. Phys. J. C **40**, 349 (2005).

[54] F. D. Aaron *et al.* (H1 Collaboration), Eur. Phys. J. C **71**, 1579 (2011).

[55] F. D. Aaron *et al.* (H1 Collaboration), Phys. Lett. B **665**, 139 (2008).

[56] S. Chekanov *et al.* (ZEUS Collaboration), Phys. Lett. B **682**, 8 (2009).

[57] J. P. Berge *et al.*, Z. Phys. C **49**, 187 (1991).

[58] U.-K. Yang *et al.* (CCFR/NuTeV Collaboration), Phys. Rev. Lett. **86**, 2742 (2001).

[59] W. G. Seligman *et al.*, Phys. Rev. Lett. **79**, 1213 (1997).

[60] G. Onengut *et al.* (CHORUS Collaboration), Phys. Lett. B **632**, 65 (2006).

[61] M. Tzanov *et al.* (NuTeV Collaboration), Phys. Rev. D **74**, 012008 (2006).

[62] M. Goncharov *et al.* (NuTeV Collaboration), Phys. Rev. D **64**, 112006 (2001).

[63] D. A. Mason, Ph.D. thesis, Oregon University, 2006.

[64] M. Aaboud *et al.* (ATLAS Collaboration), CERN Report No. ATL-PHYS-PUB-2018-017, https://cds.cern.ch/record/2633819.

[65] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 07 (2021) 223.

[66] S. Catani and M. Grazzini, Phys. Rev. Lett. **98**, 222002 (2007).

[67] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, Phys. Rev. Lett. **103**, 082001 (2009).

[68] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, Comput. Phys. Commun. **182**, 2388 (2011).

[69] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, Comput. Phys. Commun. **184**, 208 (2013).

[70] Y. Li and F. Petriello, Phys. Rev. D **86**, 094034 (2012).

[71] J. Currie, E. W. N. Glover, and J. Pires, Phys. Rev. Lett. **118**, 072002 (2017).

[72] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, and D. M. Walker, Phys. Rev. Lett. **120**, 122001 (2018).

[73] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, and T. A. Morgan, Phys. Rev. Lett. **117**, 022001 (2016).

[74] M. Czakon, D. Heymes, and A. Mitov, arXiv:1704.08551.

[75] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera, and J. Rojo, J. High Energy Phys. 04 (2017) 044.

[76] J. M. Campbell, R. K. Ellis, and C. Williams, Phys. Rev. Lett. **118**, 222001 (2017); **124**, 259901(E) (2020).

[77] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, and M. Zaro, J. High Energy Phys. 10 (2017) 186.

[78] T. Becher and X. Garcia i Tormo, Phys. Rev. D **88**, 013009 (2013).

[79] S. Dittmaier, A. Huss, and C. Speckner, J. High Energy Phys. 11 (2012) 095.

[80] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C.-P. Yuan, Phys. Rev. D **95**, 034003 (2017).

[81] CTEQ-TEA Collaboration, https://ct.hepforge.org/PDFs/ct18/figures/L2Sensitivity/.

[82] K. Xie, T. J. Hobbs, T.-J. Hou, C. Schmidt, M. Yan, and C.-P. Yuan (CTEQ-TEA Collaboration), Phys. Rev. D **105**, 054006 (2022).

[83] K. Xie, B. Zhou, and T. J. Hobbs, arXiv:2305.10497.

[84] J. Gao, T. J. Hobbs, P. M. Nadolsky, C. Sun, and C.-P. Yuan, Phys. Rev. D **105**, L011503 (2022).

[85] K. Xie, J. Gao, T. J. Hobbs, D. R. Stump, and C.-P. Yuan, arXiv:2303.13607.

[86] A. Courtoy and P. M. Nadolsky, Phys. Rev. D **103**, 054029 (2021).

[87] M. Guzzi, T. J. Hobbs, K. Xie, J. Huston, P. Nadolsky, and C.-P. Yuan, Phys. Lett. B **843**, 137975 (2023).

[88] I. Sitiwaldi, K. Xie, A. Ablat, S. Dulat, T.-J. Hou, and C.-P. Yuan, arXiv:2305.10733.

[89] T. Kluge, K. Rabbertz, and M. Wobisch, *Deep Inelastic Scattering DIS 2006* (2006), pp. 483–486, 10.1142/9789812706706_0110.

[90] M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober (fastNLO Collaboration), arXiv:1109.1310.

[91] D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch (fastNLO Collaboration), arXiv:1208.3641.

[92] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, Eur. Phys. J. C **66**, 503 (2010).

[93] J. M. Campbell and R. K. Ellis, Nucl. Phys. B, Proc. Suppl. **205**–**206**, 10 (2010).

[94] R. Boughezal, J. M. Campbell, R. K. Ellis, C. Focke, W. Giele, X. Liu, F. Petriello, and C. Williams, Eur. Phys. J. C **77**, 7 (2017).

[95] J. M. Campbell, R. K. Ellis, and C. Williams, J. High Energy Phys. 07 (2011) 018.

[96] J. M. Campbell, R. K. Ellis, and W. T. Giele, Eur. Phys. J. C **75**, 246 (2015).

[97] Z. Nagy, Phys. Rev. D **68**, 094002 (2003).

[98] V. Bertone, R. Frederix, S. Frixione, J. Rojo, and M. Sutton, J. High Energy Phys. 08 (2014) 166.

[99] http://www.precision.hep.phy.cam.ac.uk/results/ttbar-fastnlo/.

[100] R. Zhang, H.-W. Lin, and B. Yoon, Phys. Rev. D **104**, 094511 (2021).

[101] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, Eur. Phys. J. C **75**, 204 (2015).

[102] A. D. Martin, A. J. T. M. Mathijssen, W. J. Stirling, R. S. Thorne, B. J. A. Watt, and G. Watt, Eur. Phys. J. C **73**, 2318 (2013).

[103] L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **78**, 248 (2018).

[104] S. Bailey and L. Harland-Lang, Eur. Phys. J. C **80**, 60 (2020).

[105] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. Nathvani, *Proc. Sci.*, DIS2019 (2019) 036 [arXiv:1907.08147].

[106] R. Thorne, S. Bailey, T. Cridge, L. Harland-Lang, and A. D. Martin, SciPost Phys. Proc. **8**, 018 (2022).

[107] R. Boughezal, C. Focke, X. Liu, and F. Petriello, Phys. Rev. Lett. **115**, 062002 (2015).

[108] T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **82**, 90 (2022).

[109] T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 744 (2021).

[110] R. Nathvani, R. Thorne, L. Harland-Lang, and A. Martin, *Proc. Sci.*, DIS2018 (2018) 029 [arXiv:1807.07846].

[111] T. Cridge (PDF4LHC21 Combination Group), SciPost Phys. Proc. **8**, 101 (2022).

[112] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. D **85**, 072004 (2012).

[113] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **725**, 223 (2013).

[114] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 08 (2016) 009.

[115] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **76**, 291 (2016).

[116] F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. **81**, 5754 (1998).

[117] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 051104 (2005).

[118] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **102**, 181801 (2009).

[119] T. A. Aaltonen *et al.* (CDF Collaboration), Phys. Lett. B **692**, 232 (2010).

[120] S. Chatrchyan *et al.* (CMS Collaboration), J. High Energy Phys. 04 (2011) 050.

[121] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **109**, 111806 (2012).

[122] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. D **90**, 032004 (2014).

[123] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. D **85**, 032002 (2012).

[124] S. Chatrchyan *et al.* (CMS Collaboration), J. High Energy Phys. 12 (2013) 030.

[125] V. Khachatryan *et al.* (CMS Collaboration), Eur. Phys. J. C **76**, 469 (2016).

[126] G. Moreno *et al.*, Phys. Rev. D **43**, 2815 (1991).

[127] R. S. Towell *et al.* (NuSea Collaboration), Phys. Rev. D **64**, 052002 (2001).

[128] J. C. Webb *et al.* (NuSea Collaboration), arXiv:hep-ex/0302019.

[129] J. C. Webb, Ph.D. thesis, New Mexico State University, 2003, arXiv:hep-ex/0301031.

[130] J. Dove *et al.* (SeaQuest Collaboration), Nature (London) **590**, 561 (2021); **604**, E26 (2022).

[131] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. D **77**, 011106 (2008).

[132] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. D **88**, 091102 (2013).

[133] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. D **91**, 032007 (2015); **91**, 079901(E) (2015).

[134] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. Lett. **101**, 211801 (2008).

[135] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. D **76**, 012003 (2007).

[136] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. Lett. **112**, 151803 (2014); **114**, 049901(E) (2015).

[137] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 06 (2012) 058.

[138] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 02 (2013) 106.

[139] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 08 (2015) 039.

[140] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 05 (2015) 109.

[141] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 01 (2016) 155.

[142] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 02 (2015) 153; 09 (2015) 141(E).

[143] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **78**, 052006 (2008); **79**, 119902(E) (2009).

[144] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. D **75**, 092006 (2007); **75**, 119901(E) (2007).

[145] V. Khachatryan *et al.* (CMS Collaboration), Eur. Phys. J. C **76**, 265 (2016).

[146] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. D **90**, 072006 (2014).

[147] V. Khachatryan *et al.* (CMS Collaboration), J. High Energy Phys. 03 (2017) 156.

[148] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. Lett. **101**, 062001 (2008).

[149] V. M. Abazov *et al.* (DØ Collaboration), Phys. Rev. D **85**, 052006 (2012).

[150] S. Chatrchyan *et al.* (CMS Collaboration), J. High Energy Phys. 02 (2014) 013.

[151] T. A. Aaltonen *et al.* (CDF and DØ Collaborations), Phys. Rev. D **89**, 072001 (2014).

[152] S. Chatrchyan *et al.* (CMS Collaboration), J. High Energy Phys. 02 (2014) 024; 02 (2014) 102(E).

[153] A. M. Sirunyan *et al.* (CMS Collaboration), Eur. Phys. J. C **77**, 459 (2017).

[154] V. Khachatryan *et al.* (CMS Collaboration), Eur. Phys. J. C **75**, 542 (2015).

[155] G. P. Lepage, B. Clark, C. T. H. Davies, K. Hornbostel, P. B. Mackenzie, C. Morningstar, and H. Trottier (HPQCD Collaboration), Nucl. Phys. B, Proc. Suppl. **106**, 12 (2002).

[156] S. Amoroso *et al.*, Acta Phys. Pol. B **53**, A1 (2022).

[157] G. D'Agostini, Nucl. Instrum. Methods Phys. Res., Sect. A **346**, 306 (1994).

[158] G. D'Agostini, Bayesian reasoning in high-energy physics: Principles and applications, CERN Reports No. CERN-99-03, No. CERN-YELLOW-99-03, 1999.

[159] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali (NNPDF Collaboration), J. High Energy Phys. 05 (2010) 075.

[160] R. D. Ball *et al.*, J. High Energy Phys. 04 (2013) 125.

[161] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, Phys. Rev. D **89**, 033009 (2014).

[162] A. Courtoy, J. Huston, P. Nadolsky, K. Xie, M. Yan, and C.-P. Yuan, Phys. Rev. D **107**, 034008 (2023).

[163] http://metapdf.hepforge.org/L2/2023_T210/index2.html.

[164] http://metapdf.hepforge.org/L2/2023_T210/index3.html.

[165] PDFSense: visualization of experimental constraints on the nucleon structure, https://metapdf.hepforge.org/PDFSense/.

[166] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevD.108.034029 for additional figures listed at the end of Sec. VII.

[167] X. Jing, Ph.D. thesis, Southern Methodist University, Dallas, TX, 2021.