A meta-analysis of parton distribution functions

Jun Gao and Pavel Nadolsky

Department of Physics, Southern Methodist University,
Dallas, TX 75275-0181, U.S.A.

E-mail: jung@smu.edu, nadolsky@smu.edu

ABSTRACT: A "meta-analysis" is a method for comparison and combination of nonperturbative parton distribution functions (PDFs) in a nucleon obtained with heterogeneous procedures and assumptions. Each input parton distribution set is converted into a "meta-parametrization" based on a common functional form. By analyzing parameters of the meta-parametrizations from all input PDF ensembles, a combined PDF ensemble can be produced that has a smaller total number of PDF member sets than the original ensembles. The meta-parametrizations simplify the computation of the PDF uncertainty in theoretical predictions and provide an alternative to the 2010 PDF4LHC convention for combination of PDF uncertainties. As a practical example, we construct a META ensemble for computation of QCD observables at the Large Hadron Collider using the next-to-next-to-leading order PDF sets from CTEQ, MSTW, and NNPDF groups as the input. The META ensemble includes a central set that reproduces the average of LHC predictions based on the three input PDF ensembles and Hessian eigenvector sets for computing the combined PDF+α_s uncertainty at a common QCD coupling strength of 0.118.

Contents

1 Introduction

Parton distribution functions (PDFs) in the nucleon describe long-distance nonperturbative effects in high-energy QCD processes. Their detailed understanding is vital for the physics program at the Large Hadron Collider (LHC). First years of the LHC operation have culminated in the discovery of a Higgs scalar particle [1, 2] and tested various Standard Model processes at the highest precision ever attained. No direct indications for physics beyond the Standard Model have been found so far, yet future plans [3] envision reaching much higher accuracy in new physics searches at the raised LHC energy in the next decade. Toward this goal, many theoretical predictions for the LHC must include next-to-next-to-leading order (NNLO) QCD and next-to-leading order (NLO) electroweak contributions, as well as parametrizations for nonperturbative PDFs of comparable accuracy. Modern PDFs are determined by several groups from a statistical analysis of multiple hadronic experiments utilizing diverse methods. In this paper, we discuss estimation of the full uncertainty in QCD predictions based on the combination of inputs from PDFs by several groups. The combination of PDF uncertainties is non-trivial, given a variety of considerations that go into their estimation, and has important implications for those LHC measurements where the PDF uncertainty results in a leading systematic error.

Recent NNLO PDF parametrizations include CT10 [4], MSTW 2008 [5], and NNPDF 2.3 [6], determined from the global analysis of hadronic data; ABM11 [7], ABM12 [8], and JR09 [9, 10], determined from the analysis of deep-inelastic scattering (DIS) and vector boson production; and HERAPDF 1.5 [11], determined solely based on DIS. Every group provides an ensemble of error PDF sets, besides the central PDF set, to compute the PDF uncertainty according to the Hessian [12, 13] or Monte-Carlo sampling methods [14, 15].[1] The central PDFs and PDF error bands at the 68% confidence level (c.l.) from CT10, HER-APDF1.5, MSTW'08, and NNPDF2.3 ensembles are similar, but not exactly identical [16]. These PDF ensembles may share some common ingredients, like using the same experimental data sets or analogous statistical methods. They also exhibit substantial differences, notably in the choice of the QCD coupling strength, heavy-quark masses and heavy-quark schemes, PDF functional forms, and error propagation. There are also many other specialized PDFs, e.g., CJ12 [17], PDFs with intrinsic charm [18, 19] and photons [20, 21], SOM PDFs [22], and nuclear PDFs [23–26].

When predicting a new QCD observable X, a user may estimate the full uncertainty due to the PDFs by repeatedly computing X for every provided error PDF set and combining all such predictions according to a prescribed formula. Such estimation may cause a practical bottleneck if X is a complex observable that must be recomputed with hundreds of PDF parametrizations.

A related difficulty concerns combination of estimated PDF uncertainties on X using PDFs from several ensembles. Various groups follow incongruous procedures when defining their PDF uncertainties, which must be somehow reconciled in order for the combination to proceed. The PDF4LHC study recommended in 2010 [27, 28] to find the combined 68% c.l. uncertainty for X using NLO PDFs from three groups (CT, MSTW, and NNPDF) by first calculating the 68% c.l. interval independently for each group and in accordance with the definition that the group had adopted; and then taking an envelope of the three 68% c.l. intervals as the total uncertainty. Later, this procedure was applied to the NNLO PDFs [16]. The resulting estimate of the PDF uncertainty spans all individual confidence intervals obtained under non-identical definitions.

A related combination method [29, 30] (which will be referred to as a 'replica combination method') estimates the total PDF uncertainty by combining Monte-Carlo replicas (random PDF error sets) generated from the input PDF ensembles. To carry out this approach, each input ensemble utilizing the Hessian method is first converted into a secondary Monte-Carlo ensemble with about 100 member sets that is then merged with the rest of the Monte-Carlo ensembles. When CT10, MSTW'08, and NNPDF2.3 ensembles are combined, the total PDF uncertainty in the replica combination method is close to that derived according to the PDF4LHC recommendation. In either the PDF4LHC or replica combination method, the number of the final PDF error sets turns out to be the same or even larger than the number of the input sets, which is in the range of hundreds

---

[1]Throughout the paper, we draw a distinction between the "PDF sets" and "PDF ensembles". A "PDF set" is a group of parametrizations of PDFs for various flavors ($u(x,Q)$, $d(x,Q)$, etc.) corresponding to one point in the PDF parameter space. The "PDF ensemble" is a group of PDF sets from several points in the parameter space, such as in the "CT10 ensemble of 51 PDF error sets".

when predictions from many groups are combined. Both methods estimate X for multiply redundant PDF sets, many of which predict the X values that are close to the central prediction and contribute little to the total uncertainty.

We will now outline an alternative approach called a "meta-analysis" in which the input PDF sets are combined *before* the observable X is computed. Each member set of the initial PDF ensembles is approximated at an energy scale Q_0 by an intermediary functional form dependent on PDF parameters $\{a\}$. The distribution of the member sets over the parameters $\{a\}$ is analyzed, and a hypervolume spanning parameters of all PDF error sets is determined. Finally, using the shared parametrization form, we can construct a relatively small number of PDF eigenvector sets spanning all 68% c.l. error sets using the Hessian or Monte-Carlo sampling method. When the input PDF ensembles are statistically consistent with one another, the new PDF error sets reproduce the average and total PDF uncertainty of the initial PDF sets, but the number of the independent sets can be considerably reduced in the new ensemble.

The final ensemble of META PDFs constructed this way serves the same purpose as the PDF4LHC and replica combination methods, yet it combines the PDFs directly in the PDF parameter space while at the same time minimizing numerical computation efforts. In general it is difficult to bring all diverse PDFs into a common framework, for example because distinct heavy-quark schemes are used at momentum scales of a few GeV. But for purposes of the LHC studies, and focusing only the x and Q ranges corresponding to the LHC kinematics, most of the difficulties can be circumvented by requiring the lowest scale Q_0 for the PDF evolution to be above the bottom quark mass, where all PDF sets evolve assuming the same number of active flavors. The effects of choosing different heavy-quark schemes are then taken into account through the boundary conditions at the low scale Q_0. A variety of checks is performed to guarantee that the physics inputs and statistical features of the input PDF sets are preserved by the META PDF ensemble.

In the next sections, we build a combined ensemble of NNLO META PDFs for LHC studies from CT10, MSTW 2008 and NNPDF2.3 ensembles with 5 active quark flavors. These ensembles use compatible central values of the QCD coupling (α_s), similar data sets and procedures, and are in quantitative agreement [16]. The combination of such compatible sets is the simplest one to carry through, but the framework of the meta-analysis does not limit the number of the PDF ensembles that can be combined. We also derive and examine the meta-parametrizations for the ABM11 and HERAPDF1.5 sets that were not included in the PDF4LHC combination. The ABM and HERAPDF ensembles are quite distinct from the three global ensembles, hence we cannot include them into the META ensemble yet using the combination procedure that was chosen. We comment on the comparison of the ABM and HERAPDF ensembles with the META PDFs in later sections.

The paper is organized as follows. Section 2 introduces the meta-parametrizations to approximate the input PDFs at energies relevant for the LHC. Section 3 constructs a META PDF ensemble from CT10, MSTW'08, and NNPDF2.3 NNLO ensembles. In section 4 we briefly discuss phenomenological applications of the META PDFs. Section 5 contains a short summary.

Input PDF ensemble	Ref.	Initial scale	$\alpha_s(M_Z)$	Analysis type	Error type	Free parameters	Error sets	Included in META PDFs
CT10	[4]	1.3 GeV	0.118	Global	Hessian	25	50	Yes
MSTW'08	[5]	1.0 GeV	0.11706	Global	Hessian	20	40	Yes
NNPDF2.3	[6]	1.414 GeV	0.118	Global	MC	259	100	Yes
ABM11($n_f=5$)	[7]	3.0 GeV	0.118 (0.1134)	DIS+DY	Hessian	29	28	No
HERAPDF1.5	[11]	1.378 GeV	0.1176	DIS	Hessian	14	28	No

Table 1. Input NNLO PDF ensembles considered in the meta-analysis. In the ABM analysis, 28 eigenvector sets are provided with $\alpha_s(M_Z)$ varied around a best-fit value of 0.1134. The meta-parametrization is obtained for the ABM set for $\alpha_s(M_Z) = 0.118$ and compared to the other PDF ensembles at close $\alpha_s(M_Z)$ values.

2 Meta-parametrizations for input PDFs

2.1 Selection of input PDF ensembles

In a general-purpose PDF ensemble, parton distributions are parametrized by certain functions $\Phi_f(x,Q)$ of the partonic momentum fraction x at an initial scale Q_0 of order $1\,\text{GeV}$ and found numerically at higher scales Q by solving QCD evolution equations. The PDF depends on the flavor f of the probed parton, but we omit the index f in $\Phi(x,Q)$ to simplify the notation, assuming that all considerations apply to every parton flavor.

At low scales comparable to the charm and bottom masses, the PDFs from various groups cannot be directly combined because of different implementations of heavy-quark mass contributions. In typical LHC applications, on the other hand, the hard factorization scale Q tends to be well above the bottom quark mass $m_b \approx 4.8\,\text{GeV}$. At such scales the PDFs from all groups evolve in the same way according to the DGLAP evolution equations with 5 active quark flavors. At even higher scales, $Q^2 \gg m_b^2$, masses of charm and bottom quarks can be neglected in hard-scattering contributions. We therefore can circumvent explicit treatment of heavy-quark effects by approximating the PDFs $\Phi(x,Q_0)$ by flexible meta-parametrizations $f(x,Q_0;\{a\})$ at a scale Q_0 above m_b. For definiteness we choose $Q_0 = 8\,\text{GeV}$ and assume 9 independent PDF parametrizations at this scale, for g, u, $\overline{u}$, d, $\overline{d}$, s, $\overline{s}$, $c = \overline{c}$, and $b = \overline{b}$. We neglect small differences between c and $\overline{c}$, and b and $\overline{b}$ PDFs.

The input PDF ensembles that will be considered in this analysis are summarized in table 1. They are obtained at the next-to-next-to-leading order (NNLO) in the QCD coupling strength α_s. Every ensemble except ABM11 includes 28-100 PDF error sets probing the uncertainty due to the PDF parametrization at a fixed $\alpha_s(M_Z)$ close to 0.118, as well as additional best-fit PDF parametrizations for alternative $\alpha_s(M_Z)$ to examine the magnitude of the α_s uncertainty. The 0.118 value is compatible with the world-average QCD coupling 0.1184 ± 0.0007 [31] and will be used as a common $\alpha_s(M_Z)$ value in most comparisons. In the ABM11 ensemble, the error PDFs are given for a lower central value of $\alpha_s(M_Z) = 0.1134$ and include the $\alpha_s(M_Z)$ variation in the covariance matrix. An update of the ABM11 analysis tuned to the LHC data, called ABM12, has been released very recently [8]. Only the ABM11 ensemble provides a member set for $\alpha_s(M_Z) = 0.118$ from their α_s series (but not the PDF error sets at this α_s). Because the ABM error sets

correspond to a lower $\alpha_s(M_Z)$ value, they should not be compared on the same footing with the error PDFs for a fixed $\alpha_s(M_Z) \approx 0.118$ from the other groups. For this comparison, we will examine the meta-parametrization for the ABM11 member set at $\alpha_s(M_Z) = 0.118$, as well as LHC predictions for ABM12 using the central $\alpha_s(M_Z) = 0.1132$. Note that, for making theoretical predictions, the ABM group recommends to use their PDF sets corresponding to their best-fit $\alpha_s(M_Z) \approx 0.113$.

When combining the PDF ensembles, one follows two common methods used for estimating the PDF uncertainty, the Hessian method [12, 13] and the Monte Carlo (MC) sampling method [14, 15]. We summarize the core relations of the two methods for completeness. Generally speaking, the Hessian approach provides $2N_{\text{eig}}$ PDF error sets (called "eigenvector sets") that only estimate a given confidence interval for the PDFs without specifying the actual probability distribution. The MC method returns N_{rep} PDF error sets called "replicas" to estimate the probability density for the desired QCD observable.

The following simplest statistics can be computed with either method.

1. *The central prediction* for a QCD observable X, such as a cross section or the PDF itself, is the value X_C^H given by the most probable PDF in the Hessian approach (corresponding to the lowest log-likelihood χ^2 in the global fit); or the mean value $X_C^M = \langle X \rangle_{\text{replicas}}$ of predictions from all the replicas in the MC method.

2. *The PDF uncertainty* in the Hessian method estimates the boundaries of the confidence interval on X, usually at the 68% or 90% c.l. Both the symmetric [12, 13] uncertainty, δ^H, and asymmetric uncertainties $\delta_\pm^H$ [32] can be estimated:

$$\delta^H(X) = \frac{1}{2}\sqrt{\sum_{i=1}^{N_{\text{eig}}}[X_i^+ - X_i^-]^2}, \tag{2.1}$$

$$\delta_+^H(X) = \sqrt{\sum_{i=1}^{N_{\text{eig}}}[\max(X_i^+ - X_C^H, \ X_i^- - X_C^H, \ 0)]^2}, \tag{2.2}$$

$$\delta_-^H(X) = \sqrt{\sum_{i=1}^{N_{\text{eig}}}[\max(X_C^H - X_i^+, \ X_C^H - X_i^-, \ 0)]^2}, \tag{2.3}$$

where $X_i^\pm$ are the values of X corresponding to the upper and lower boundary of the confidence interval for the parameter a_i. These formulas are used by the CTEQ, HERA, and MSTW groups, while the ABM group uses a variant of the Hessian master formula that estimates the symmetric uncertainty from the differences between the predictions based on the central PDF and one error set for each eigenvector direction [7].

In the MC approach, the symmetric 68% c.l. PDF uncertainty is given by a standard deviation δ^M:

$$\delta^M(X) = \sqrt{\frac{1}{N_{\text{rep}-1}}\sum_{i=1}^{N_{\text{rep}}}[X_i - X_C^M]^2}. \tag{2.4}$$

3. *The tolerance (error) ellipse* is introduced to study correlations between two observables, e.g., X and Y. In the Hessian approach, we first define the correlation angle $\Delta\varphi$ as [12, 32, 33]

$$\cos(\Delta\varphi) = \frac{1}{4\delta^H(X)\delta^H(Y)} \sum_{i=1}^{N_{\text{eig}}} (X_i^+ - X_i^-)(Y_i^+ - Y_i^-) \,. \tag{2.5}$$

The tolerance ellipse is the projection onto the XY plane of the Hessian hypersphere spanned by the eigenvector sets in the linear approximation for PDF dependence of X and Y. The ellipse can be found from the parametric equations

$$X(\theta) = X_C^H + \delta^H(X)\cos\theta, \quad Y(\theta) = Y_C^H + \delta^H(Y)\cos(\theta + \Delta\varphi) \tag{2.6}$$

for $0 \leq \theta < 2\pi$.

In the MC approach we can also define the error ellipse by an equation

$$\frac{(X - X_C^M)^2}{\delta^M(X)^2} + \frac{(Y - Y_C^M)^2}{\delta^M(Y)^2} - \frac{2\rho(X - X_C^M)(Y - Y_C^M)}{\delta^M(X)\delta^M(Y)} = p_0^2(1 - \rho^2), \tag{2.7}$$

where $\rho \equiv \text{cov}(X,\,Y)/\left(\delta^M(X)\,\delta^M(Y)\right)$ is the correlation of X and Y, and

$$\text{cov}(X,Y) = \frac{1}{N_{\text{rep}} - 1} \sum_{i=1}^{N_{\text{rep}}} [X_i - X_C] \cdot [Y_i - Y_C]. \tag{2.8}$$

Projections of the ellipse on the X and Y axes coincide with the confidence intervals on X and Y determined by p_0, for instance $p_0 = 1$ (1.64) for the 68 (90)% c.l. Eqs. (2.5) and (2.7) delineate the regions of the prescribed confidence when the distributions of X and Y are close to Gaussian ones. The equivalence of the Hessian and MC error ellipses in the Gaussian case can be demonstrated by identifying $\rho = \cos(\Delta\varphi)$.

2.2 Parametrization and sum rules

The functional forms for $\Phi(x, Q_0)$ are often assumed to behave as $\sim x^{a_2}(1-x)^{a_3}$ in the $x \to 0$ and $x \to 1$ limits on the basis of Regge and quark counting arguments. They must satisfy the valence quark number and momentum sum rules and predict positive cross sections. The free parameters of $\Phi(x, Q_0)$ are found from a fit, their total number is chosen so as to provide very flexible functional forms agreeing with the data, but without overfitting the data. NNPDF2.3 uses a neural network with 259 parameters to minimize the parametrization bias, while the CT10 and MSTW'2008 NNLO ensembles keep 25 and 20 free parameters, respectively. Another convenient form utilizes Chebyshev orthogonal polynomials $T_i(y(x))$ [34–36]. This form is particularly suited for constructing the meta-parametrizations and will be employed in our analysis.

In those x ranges where the experimental data impose tight constraints, the number of effective degrees of freedom is smaller than in the full PDF ensemble, so a smaller number of parameters is needed to approximate the acceptable PDF shapes. We will therefore
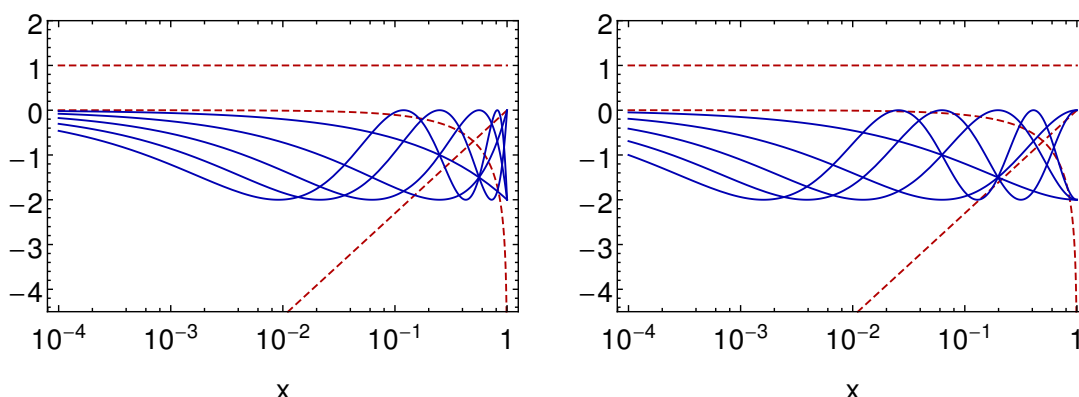
Figure 1. $d\ln(f)/da_i$ for $a_{1,2,3}$ (red dashed curves) and $a_{4,5,6,7,8}$ (blue solid curves) with $y(x) = 1 - 2x^{1/2}$ (left inset) and $y = \cos(\pi x^{1/4})$ (right inset).

fit the effective parametrizations to the input PDFs only in the x ranges where sufficient experimental constraints are available. The boundaries of these ranges are taken to satisfy $x > 3 \cdot 10^{-5}$ for all quark flavors; $x < 0.8$ for the gluon and u, d, c, b quarks; $x < 0.4$ for $\bar{u}, \bar{d}$ quarks; and $x < 0.3$ for $s, \bar{s}$ quarks. By construction, the uncertainty bands of the input PDFs and their meta-parametrizations agree well in the fitted x regions. Outside these ranges, the meta-parametrizations are determined by extrapolation and span a wide uncertainty band, which is close, although not identical, to the original PDF uncertainty (which has a large uncertainty of its own at such x). The PDFs in the outside (unfitted) regions are hardly constrained at the moment and have negligible contributions to most LHC observables even at $14\,\text{TeV}$.

The specific effective form that we choose at the scale Q_0 for each flavor is

$$f(x, Q_0; \{a\}) = e^{a_1} x^{a_2} (1 - x)^{a_3} e^{\sum_{i \geq 4} a_i \left[T_{i-3}(y(x)) - 1 \right]}. \qquad (2.9)$$

It satisfies the above asymptotic behaviors, $\sim x^{a_2}$ at $x \to 0$ and $\sim (1-x)^{a_3}$ at $x \to 1$, while the detailed shape is regulated by the Chebyshev polynomials, $T_j(y(x))$ with $j \geq 1$, bound to lie between -1 to 1 for $-1 \leq y \leq 1$. The positivity condition is automatically satisfied for $0 \leq x \leq 1$, in accord with the input PDFs, which are positive for all groups at the x and Q_0 values we chose. The function $y(x)$ maps the $0 \leq x \leq 1$ interval onto the $-1 \leq y \leq 1$ interval. The form of $y(x)$ is selected so as to avoid large cancellations between the coefficients of Chebyshev polynomials in the $y \to \pm 1$ limits, hence to reduce the number of $T_i(y(x))$ needed for approximating $\Phi(x, Q_0)$.

We select $y = \cos(\pi x^\beta)$ with $\beta = 1/4$ as a mapping function that generally requires fewer Chebyshev polynomials than the other tried form $y = 1 - 2x^\alpha$ with $\alpha = 1/2$ suggested by [34]. The choice of $y(x)$ is illustrated in figure 1, where the logarithmic derivatives $d(\ln f)/da_i$ are compared for $y = 1 - 2x^{1/2}$ in the left subfigure and $y = \cos(\pi x^{1/4})$ in the right subfigure. From eq. (2.9) we have

$$\frac{d\ln(f)}{da_i} = \left\{ 1, \ln x, \ln(1-x), T_{i-3}(y(x)) - 1 \right\}, \qquad (2.10)$$

thus $\ln f$ is linear in the a_i parameters. The coefficients $d \ln f/da_i = \{T_1\left(y(x)\right) - 1, T_2\left(y(x)\right) - 1, \ldots\}$ for $i \geq 4$ are shown by blue solid lines. With the choice $y = \cos(\pi x^{1/4})$ in the right inset, the oscillations of the polynomials are stretched across a wider span of x, resulting in a better approximation of the PDF shapes.

We evaluate $\Phi(x_k, Q_0)$ and $f(x_k, Q_0; \{a\})$ on a lattice of momentum fractions $\{x_k\}$ for each flavor and fit f to Φ by minimizing a metric function

$$E\left[\Phi,\ f(a)\right] = \sum_{\text{flavors}, x\,\text{grid}} \left[\frac{\ln \Phi(x_k, Q_0) - \ln f(x_k, Q_0; \{a\})}{\delta(\ln \Phi(x_k, Q_0))}\right]^2, \qquad (2.11)$$

where $\delta(\ln \Phi(x_k, Q_0)) \equiv \delta(\Phi(x_k, Q_0))/\Phi(x_k, Q_0)$, and $\delta(\Phi(x_k, Q_0))$ is the symmetric PDF uncertainty of $\Phi(x_k, Q_0)$. For each of the 9 flavors, $f(x, Q_0; a)$ depends on at least three important parameters, $a_{1,2,3}$. The number of additional Chebyshev polynomials varies depending on the complexity of the PDF shape and is especially large for the NNPDF parametrizations that oscillate. By trial and error, we found that all current NNLO PDFs can be approximated by including up to order-5 polynomials T_j for the gluon, u and d quarks; and up to order-4 polynomials for other flavors. This will be our default choice, rendering a total of 66 PDF parameters. The differences between the input and fitted PDF uncertainty bands are much smaller than the uncertainty bands in this case. The order of Chebyshev polynomials is still low enough to avoid large cancellations between their coefficients.

We checked that the sum rules are automatically preserved by the approximate parametrizations. The integrals of the meta-parametrizations must obey

$$\Sigma_m \equiv \sum_{\text{flavors}} \int_0^1 x\, f(x, Q)\, dx = 1, \qquad (2.12)$$

$$\Sigma_{uv} \equiv \int_0^1 \left(u(x, Q) - \overline{u}(x, Q)\right)\, dx = 2, \qquad (2.13)$$

$$\Sigma_{dv} \equiv \int_0^1 \left(d(x, Q) - \overline{d}(x, Q)\right)\, dx = 1. \qquad (2.14)$$

In figure 2, we evaluate partial contributions to these integrals over the fitted x regions at the scale Q_0. For all input PDFs, the partial integrals render an average of 0.996 for the momentum sum, 1.99 for the u-valence sum, and 0.997 for the d-valence sum. Thus the sum rules are satisfied by the approximate parametrizations even if they are not enforced by an explicit condition, and if integration is only over the selected partial x regions.

In figures 3 and 4, we compare the input NNLO PDFs (solid lines) for 5 PDF ensembles to their respective meta-parametrizations, referred to as "fitted PDFs" (dashed lines). The comparisons are made at scales $Q = 8$ and $85\,\text{GeV}$ after evolution of the fitted PDFs that will be explained a bit later. Due to the limited space, we only show results for the gluon PDF $g(x, Q)$ as a representative example. Analogous plots for all flavors and PDF ensembles are available at [37].

In each subfigure, we see the ratio of the central fitted PDF to the central input PDF for each ensemble, indicated by the dashed line at the center. The bands of the 90% c.l.
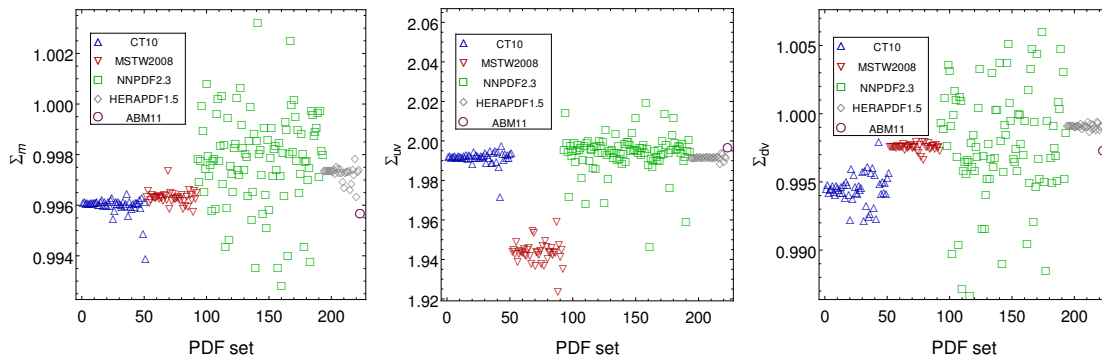
Figure 2. Partial integrals over the fitted x ranges for the momentum sum rule, u-valence sum rule, and d-valence sum rule at $Q_0 = 8\,\text{GeV}$. In each inset, the points from left to right are found from CT10 (up triangles), MSTW2008 (down triangles), NNPDF2.3 (squares), HERAPDF1.5 (diamonds), and ABM11 (circle) PDF sets.

PDF uncertainties for the input and fitted PDFs are also shown, normalized to the central input PDF. When the input ensemble provides only 68% c.l. error sets, the 90% c.l. errors are obtained by multiplying the 68% c.l. ones by 1.64. For ABM11, only the best-fit PDF with $\alpha_s(M_Z) = 0.118$ is presented, since their PDF error sets correspond to a lower central $\alpha_s(M_Z)$ and include the $\alpha_s(M_Z)$ uncertainty in the covariance matrix.

Apart from not too consequential differences at very small or large x values, the agreement between the input and fitted PDFs is good, especially for CT10. For MSTW at $Q = 8\,\text{GeV}$ and, similarly, for HERAPDF, only for x below 10^{-4} the fitted PDFs show a systematically upward shift due to the double pole structure of the MSTW gluon PDF [5] that is absent in our functional form. It is even more interesting to examine the fit to NNPDF2.3, which originally has a much more flexible parametrization with a total of 259 PDF parameters. From figure 3, we find that the fitted PDFs show almost the same statistical features as the original NNPDF except for the PDF uncertainties in the region with $x < 10^{-4}$. The differences between the input and fitted PDF errors are further reduced at higher scales, such as $Q = 85\,\text{GeV}$ in figure 4. These small differences outside of the typical $x - Q$ region of the LHC will hardly matter in practice.

In principle, above the initial scale $Q_0 = 8$ GeV, the selected NNLO PDFs and associated QCD coupling strengths α_s follow identical evolution equations based on 3-loop QCD splitting functions with 5 active flavors [38, 39], since Q_0 is above the PDF transition threshold from 4 to 5 flavors used in the variable flavor number schemes. On the other hand, some numerical differences can't be excluded initially in numerical implementation of these equations by various groups. To benchmark the α_s running and PDF evolution above Q_0, we compared the tabulated Q dependence of the PDFs and $\alpha_s(Q)$ from the five ensembles available in the LHAPDF library [40] to the explicit evolution of their respective input parametrizations done with the computer code HOPPET [41]. The benchmarking comparison is summarized in the appendix. Acceptable agreement between the tabulated and HOPPET evolution at better than 1–2% is observed for all PDF ensembles and at practically all x and Q.
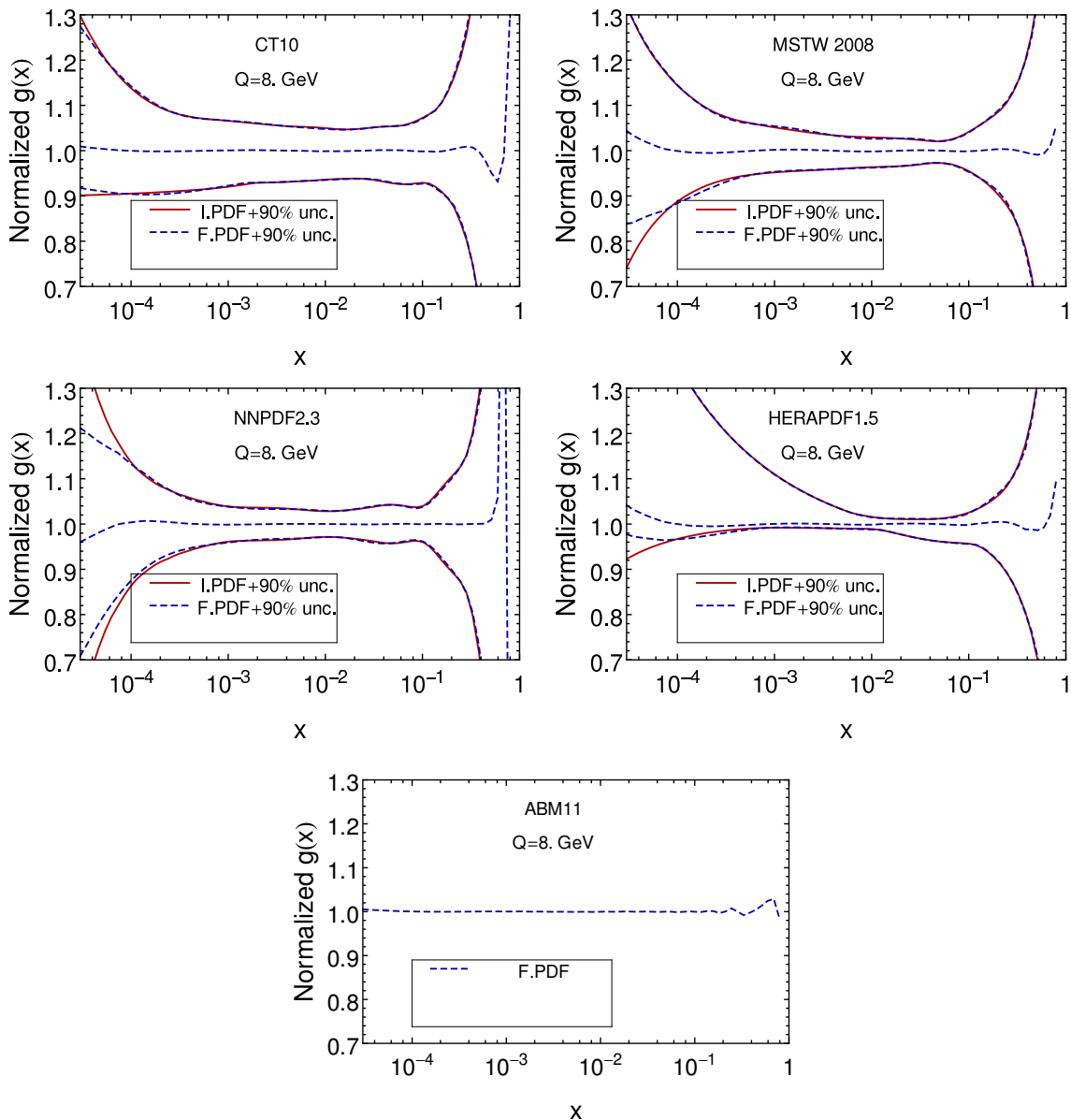
Figure 3. Comparison of the input (I) and fitted (F) gluon PDFs and their 90% c.l. uncertainties at the initial scale $Q = 8\,\mathrm{GeV}$, normalized to the best-fit gluon PDF of the input ensemble.

We can also examine the quality of fits to individual error sets from each group, as is illustrated in figures 5–8 for some PDF flavors. For instance, in figure 5 the green long-dashed lines indicate the ratios of 50 PDF eigenvector sets in CT10 to the central set. The blue short-dashed lines indicate the ratio of the meta-parametrization to the input parametrization for each eigenvector set. The differences between the input and meta-parametrizations are below 2% in most cases, much smaller than the spread of the error PDFs. Similar level of agreement is reached for the MSTW and HERA PDFs in figures 6 and 7, where slightly larger deviations (up to about 5%) are observed at $x < 10^{-4}$. Finally, for NNPDF in figure 8, the meta-parametrizations tend to have fewer oscillations than the
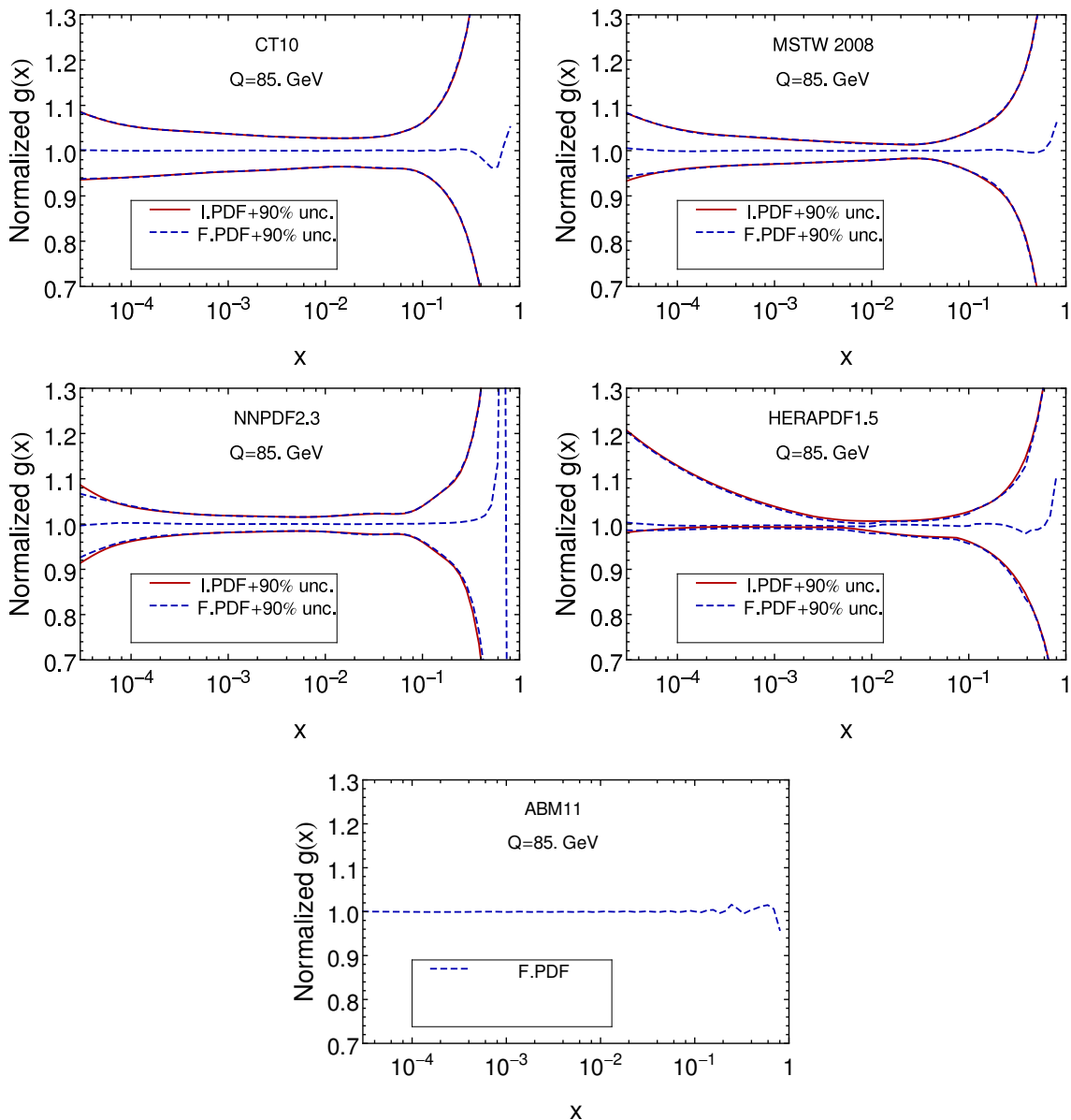
Figure 4. Same as figure 3, at scale $Q = 85\,\text{GeV}$ after evolution of the fitted PDFs.

input error PDF sets because of the functional form we chose. The differences between the individual error PDFs of the input and fitted NNPDF ensembles are larger in this case, but cancel well in the full PDF uncertainty, which comes to be about the same in the input and fitted ensembles as discussed above.

2.3 Comparison of meta-PDF parameters from various ensembles

Once all input ensembles are converted into a shared functional form, it becomes possible to directly examine and compare their meta-parameters as an alternative to the usual comparisons of the PDF shapes in the x space. The distributions of the meta-parameters
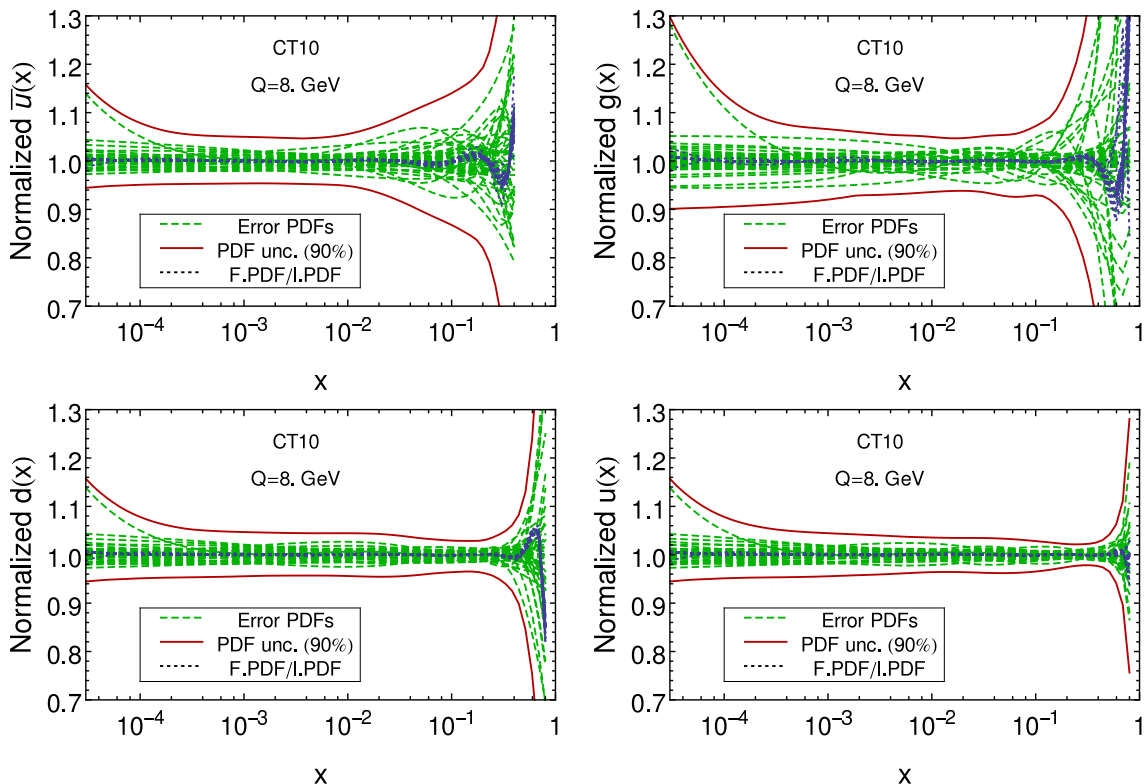
Figure 5. The CT10 NNLO error band at 90% c.l. (outer solid lines) is compared against individual error sets of the CT10 NNLO ensemble (long-dashed lines), all normalized to the central CT10 NNLO set, as well as ratios of the input and fitted parametrizations for each error set (short-dashed lines). The scale $Q = 8\,\mathrm{GeV}$ is assumed.

follow a variety of trends [37], some of each are illustrated in figures 9 and 10. Here we show the probability distributions for select pairs of the meta-parameters from the five fitted ensembles, with $f(i)$ in the axis labels indicating the parameter a_i in the meta-parametrization (2.9) for flavor $f = \bar{s}, \bar{u}$, etc. The discrete markers and lines indicate the central values and 90% c.l. error ellipses computed according to eqs. (2.5)–(2.7) for each fitted ensemble. Blue, red, green, gray, and magenta colors correspond to the CT, MSTW, NNPDF, HERAPDF, and ABM ensembles, respectively.

In general, the figures indicate good agreement between the recent NNLO PDFs. While the CT, MSTW, and NNPDF parameters are the most compatible among themselves, the ABM and HERAPDF parameters are more likely to deviate from the averages of the three global sets, as is illustrated by some insets in figures 9 and 10, as well as later in figures 14 and 15.

Figure 9 shows distributions of the parameter a_3 that controls the PDFs in the $x \to 1$ limit. The values of a_3 are well constrained for u and d quarks, but the spread of a_3 for sea flavors is much wider. [The axis ranges are chosen to coincide in most subfigures to visually compare the spread of the a_3 parameters for various flavors.] For the HERA ensemble, the central a_3 value for u is outside of the 90% error ellipses of the global ensemble, and
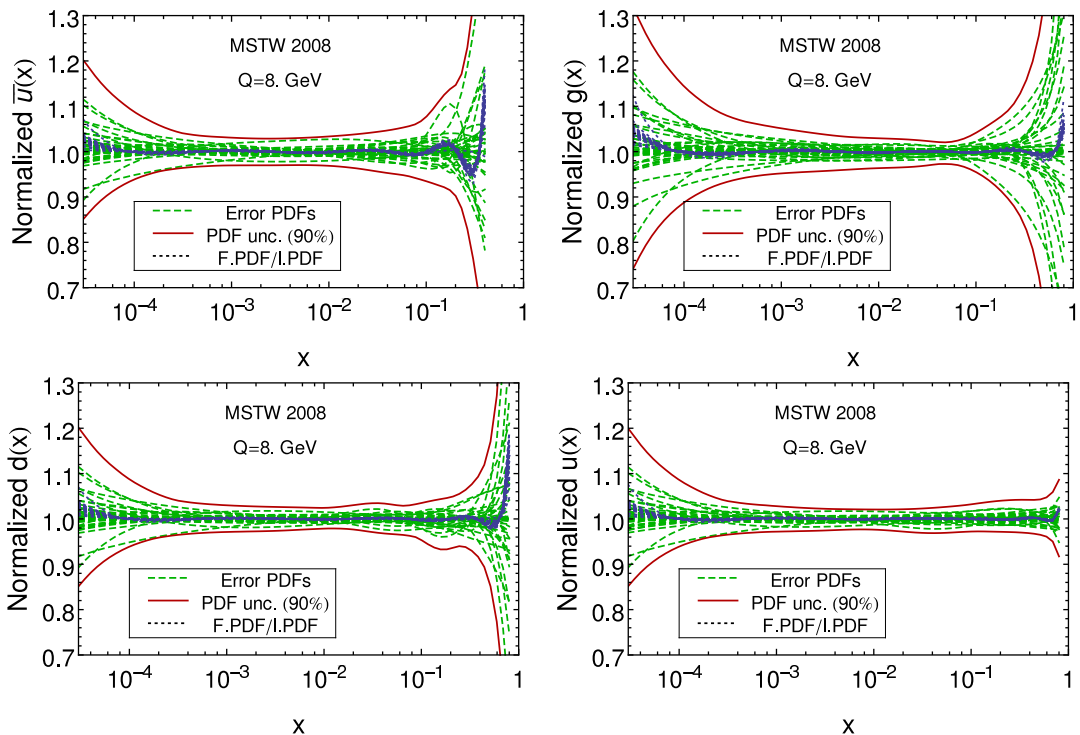
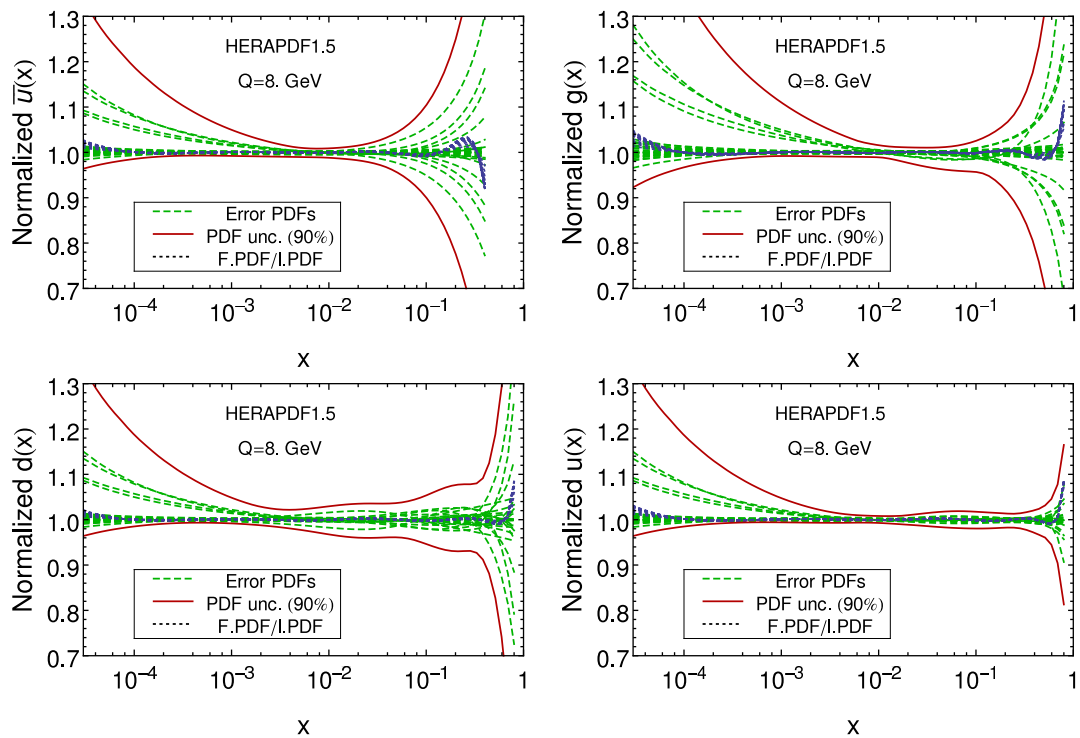**Figure 6**. Same as figure 5, for the MSTW'08 NNLO ensemble.

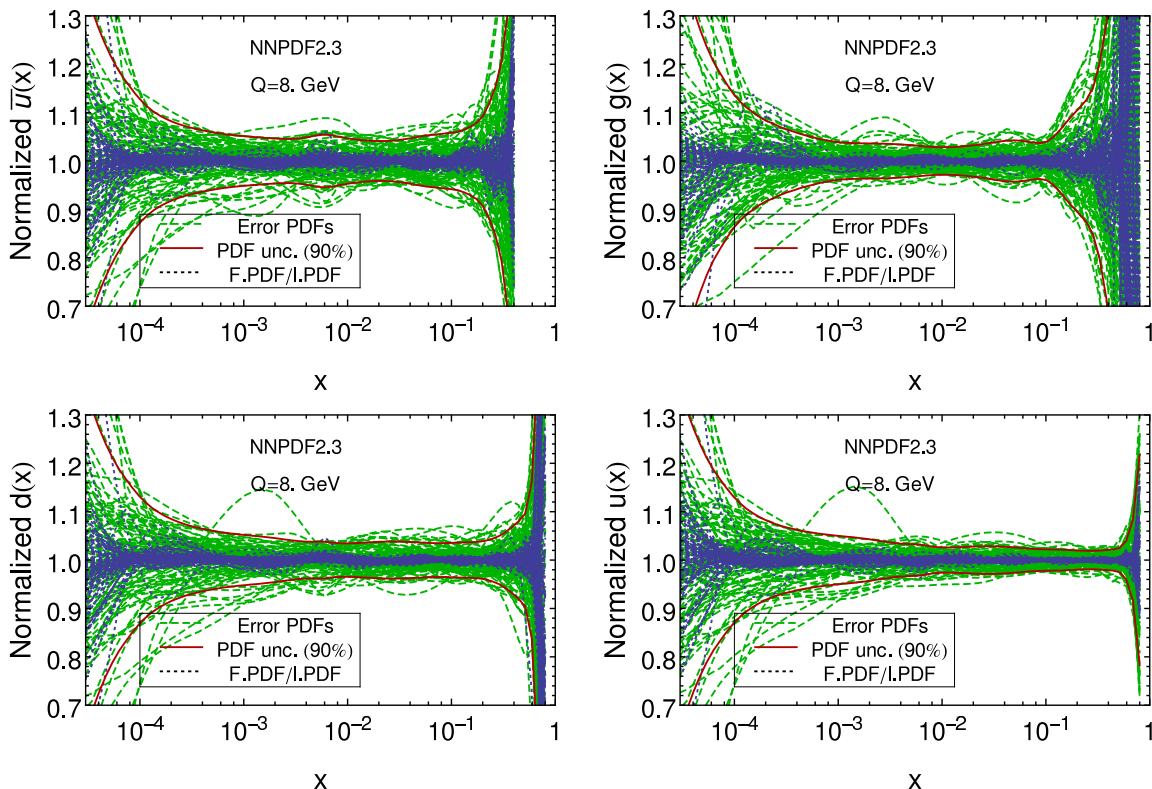**Figure 7**. Same as figure 5, for the HERAPDF1.5 NNLO ensemble.

Figure 8. Same as figure 5, for the NNPDF2.3 NNLO ensemble.

similarly, the central a_3 value of ABM for c quark is barely on the boundary of the 90% ellipses of the global ensembles. It is also remarkable that the a_3 parameters for many flavors are not correlated: the axes of the corresponding error ellipses are oriented along the axes of the plot, rather than along the diagonals. On the other hand, in figure 10 we observe strong correlations between the PDF parameters of different flavors that are due to the basic physical properties of the PDFs, e.g., close connections between the heavy quarks and gluon, strangeness and anti-strangeness, $\bar{u}$ and $\bar{d}$.

2.4 Predictions based on the input and fitted PDFs for the LHC

As another test of consistency between the input and fitted PDF ensembles, we compared their predictions for several common observables at the LHC. We examined a group of observables that probe various combinations of PDFs in the typical LHC measurements, as discussed in [4]. They include the total cross sections for W and Z boson production (with a decay into a lepton pair) at NNLO ([42], Vrap); SM Higgs boson production via gg and $b\bar{b}$ fusion, computed at NNLO ([43], iHixs1.3); $t\bar{t}$ production with $m_t = 173.5\,\mathrm{GeV}$ at partial NNLO [44–47] and including resummed contributions [48–50], implemented in TOP++1.5; as well as differential distributions of single-inclusive jet production at NLO ([51], FastNLO2.0) for ATLAS kinematic bins [52].

Figures 11–12 illustrate the level of agreement observed in these comparisons. Here the central predictions and their PDF uncertainties using the input (fitted) PDFs are
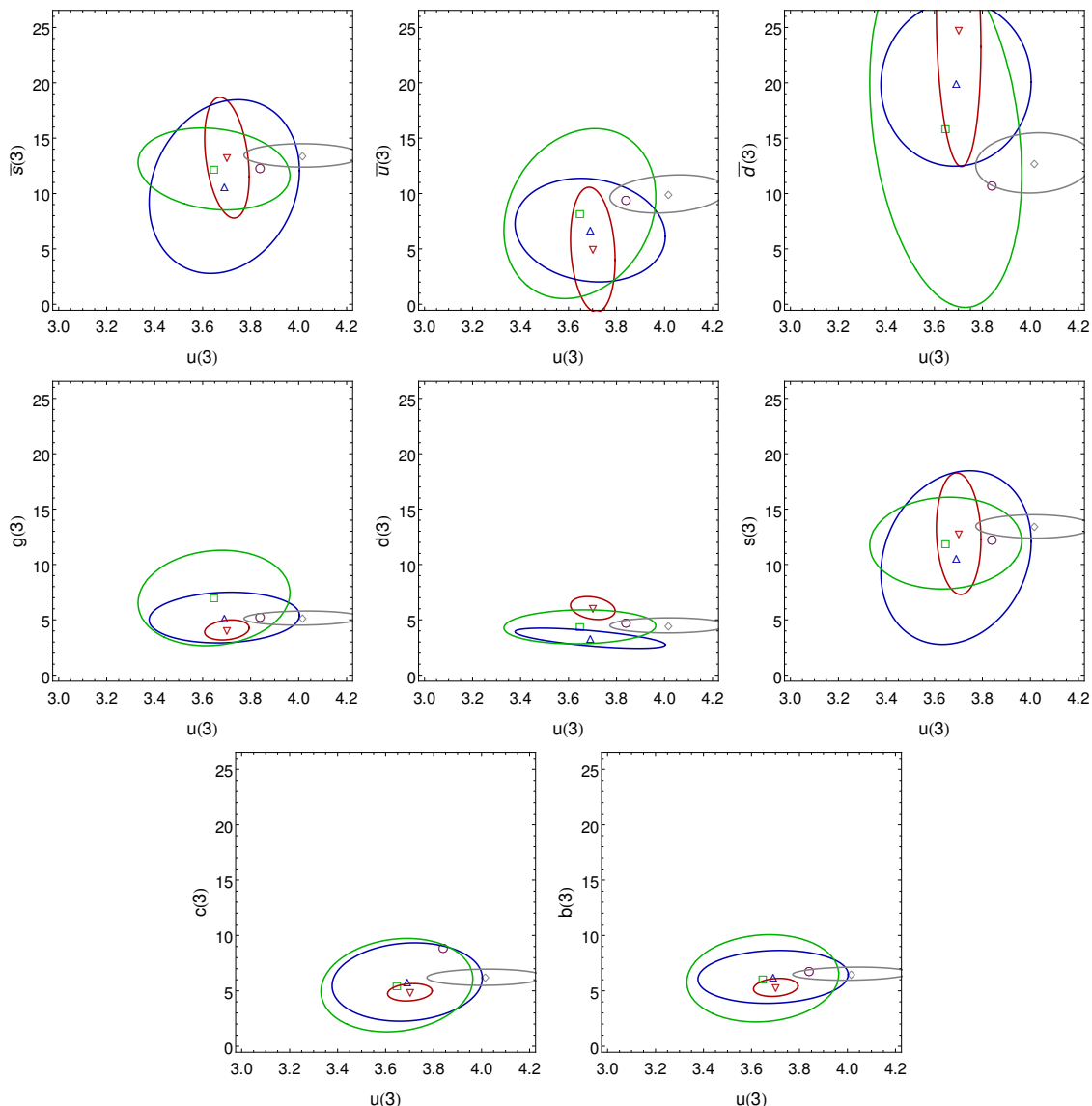
Figure 9. Fitted PDF parameters and 90% c.l. ellipses for CT10 (blue up triangle), MSTW08 (red down triangle), NNPDF2.3 (green square), HERAPDF1.5 (gray diamond) and ABM11 (magenta circle).

indicated by solid (dotted) error bars, for CT10, MSTW'08, and NNPDF2.3 at the LHC 14 TeV. Similar results for other processes and the LHC 8 TeV are available at [37]. The PDF uncertainties at 90% c.l. are computed according to eqs. (2.2)–(2.3) for the CT10 and MSTW ensembles, and eq. (2.4) for the NNPDF2.3 ensemble. The central predictions are for the default $\alpha_s(M_Z)$ values of each ensemble listed in table 1. In each group of four predictions, the first pair of error bars corresponds to the PDF uncertainties, while the second pair is the α_s uncertainty obtained with a central value of 0.118 for $\alpha_s(M_Z)$ and variations of ± 0.002.[2]

---

[2]To compute the α_s uncertainty, we also fit the α_s PDF series of all the groups.
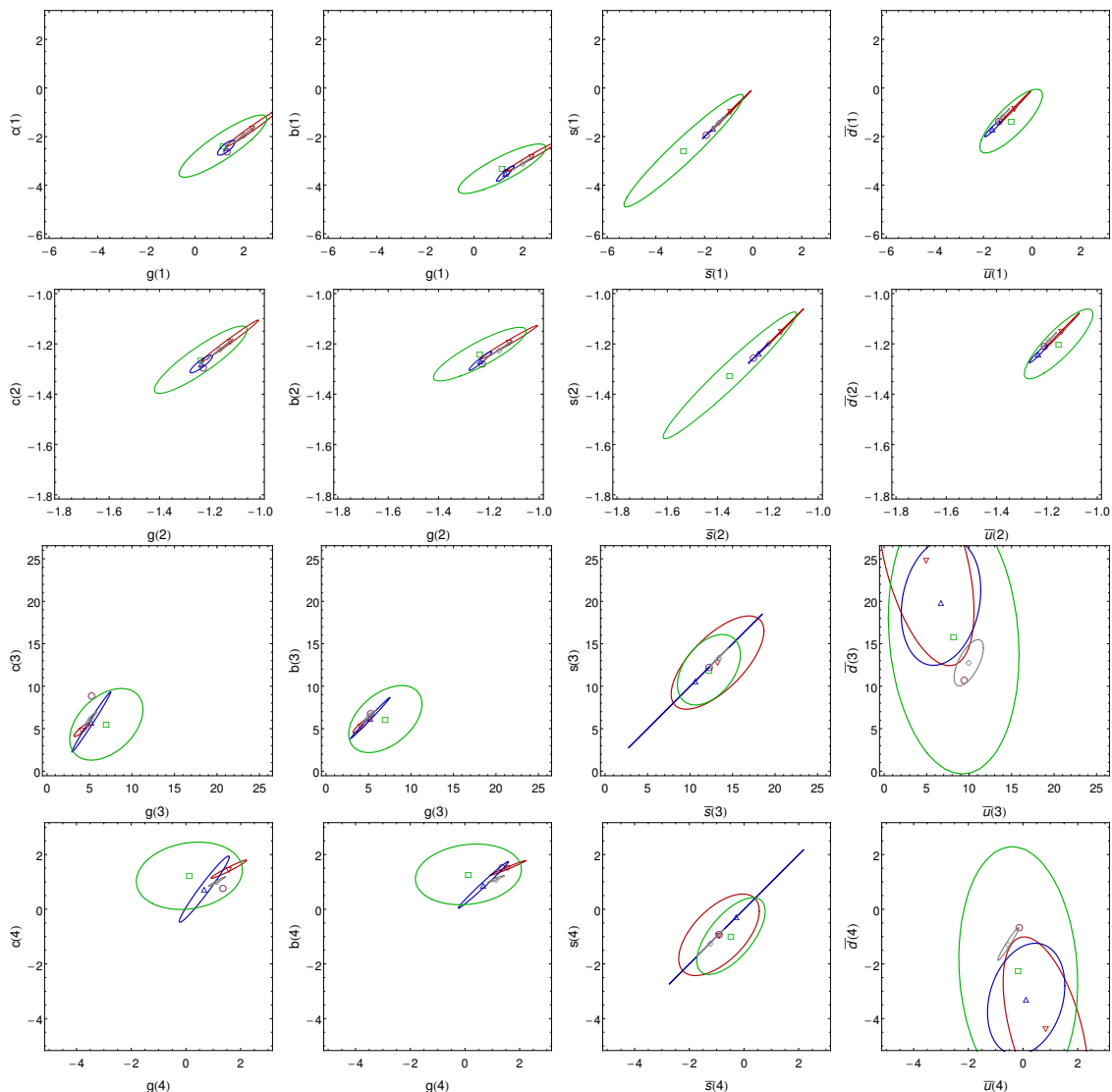
Figure 10. Correlations between representative pairs of meta-PDF parameters, for the same PDF ensembles as in figure 9.
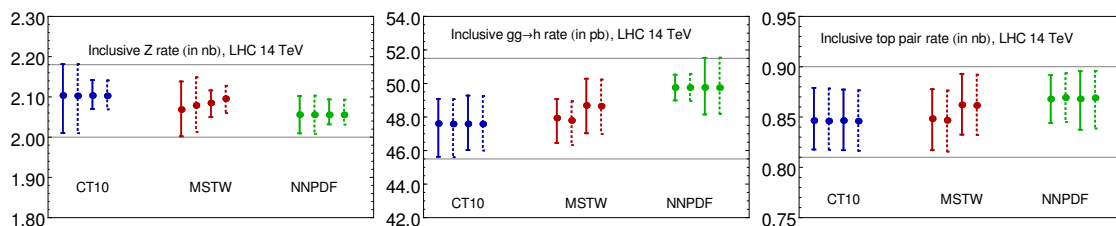

Figure 11. Comparison of inclusive Z, SM Higgs, and $t\bar{t}$ production cross sections at NNLO using the input (solid) and fitted (dotted) PDFs. The first (second) pair of error bars for each ensemble indicates the PDF (α_s) uncertainty. The renormalization and factorization scales are set equal to M_Z, M_H, and m_t, respectively.
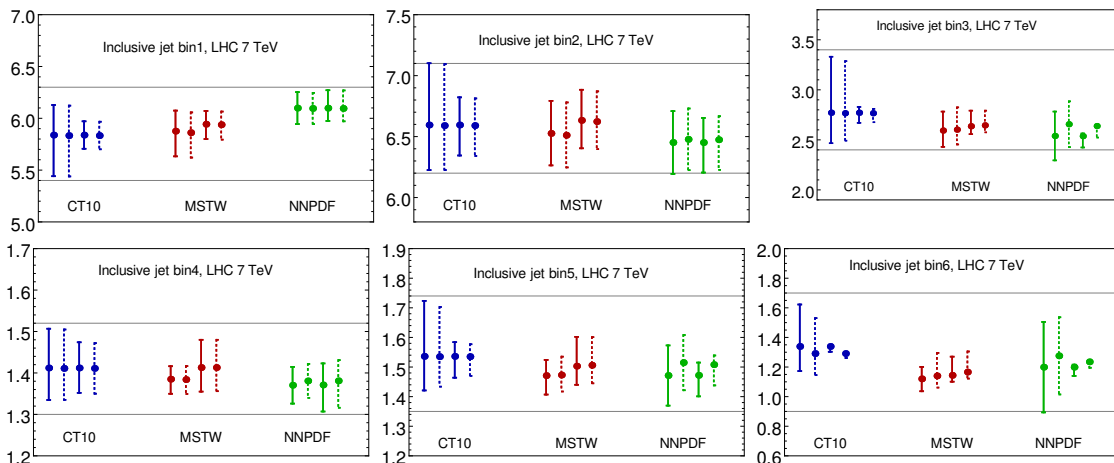
Figure 12. Same as figure 11, for inclusive jet production cross sections (in arbitrary unit) at NLO at the LHC 7 TeV. The upper row corresponds, from left to right, to the central jet rapidity region [0, 0.3] in p_T bins of [20, 30], [310, 400], and [1200, 1500] GeV. The lower row corresponds to the rapidity [4.0, 4.4] and p_T bins of [30, 45], [60, 80], and [110, 160] GeV. The renormalization and factorization scales are equal to the p_T of the individual jet.
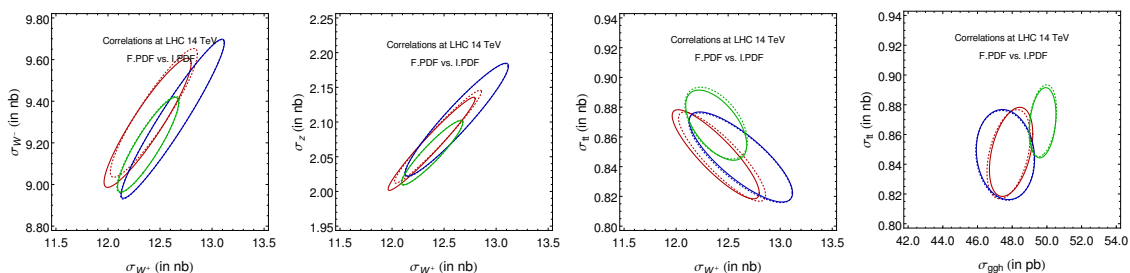

Figure 13. Comparison of correlations of NNLO cross sections using the input (solid) and fitted (dotted) PDFs.

The agreement between the input and fitted PDFs (both for $\alpha_s(M_Z) = 0.118$ and the α_s series) is nearly perfect for CT10 and NNPDF2.3. Minor differences can be noticed upon closer examination, but they are always smaller than the PDF uncertainties. For some MSTW2008 predictions, there is an overall shift ($< 1\%$) due to the small difference between their tabulated and the HOPPET numerical DGLAP evolution mentioned in the appendix.

The 90% c.l. error ellipses, computed according to eqs. (2.5)–(2.7) for the above LHC cross sections, are plotted in figure 13. As before, the solid (dotted) ellipses represent predictions based on the input (fitted) PDFs for CT10 (in blue), MSTW (in red), and NNPDF (in green). The locations and shapes of the original error ellipses are well preserved by the meta-parametrizations. From these comparisons we conclude that the chosen meta-parametrization form with 66 parameters in eq. (2.9) reproduces well the central LHC cross sections, their PDF uncertainties, and correlated flavor dependence of all input PDF sets. We can now proceed to the combination of the error PDFs into a META ensemble.

3 Combining PDFs into a META ensemble

3.1 Overview

In the previous section, we have shown that, by using a shared PDF parametrization form, we can closely approximate NNLO error PDFs from all groups at the initial scale of $Q_0 = 8\,\text{GeV}$. Even more, if numerical evolution by HOPPET is adopted, the fitted PDFs also agree with the input PDFs at all scales above Q_0.

Using these meta-parametrizations, we can create a META ensemble of PDFs from all groups that would serve similar purposes as the PDF4LHC recommendation [27]. For the combination to proceed, the PDFs must use compatible values of α_s, heavy-quark masses, number of active flavors, and other physics inputs. Furthermore, it is desirable to have about the same number of experimental points fitted in each input ensemble to be allowed to use unweighted averages when combining them, as we do below.

Thus we choose to combine the meta-parametrizations from CT10, MSTW2008, and NNPDF2.3 with $\alpha_s(M_Z) = 0.118$, 0.11707, and 0.118, which satisfy these requirements. In the META PDF ensemble we will assume a common $\alpha_s(M_Z) = 0.118$. The available error sets from MSTW correspond to a lower $\alpha_s(M_Z) = 0.11707$, but MSTW also provides an additional central set corresponding to $\alpha_s(M_Z) = 0.118$. We estimate the meta-parameters of all MSTW error PDF sets shifted to 0.118 by adding the differences between the parameters of MSTW central sets for 0.118 and 0.11707 to the parameters of the MSTW error PDF sets for 0.11707. In the future, we hope that the different groups will make their PDF ensembles available for the same $\alpha_s(M_Z)$ to facilitate their combination into META PDFs.

Before proceeding, we should warn that the definitions of the PDF uncertainties, as well as the procedures for implementing them, differ significantly from group to group. It is beyond the scope of our study to reconcile all pertinent differences among the individual input PDF sets. For now, we will assume that the PDF error sets that *nominally* correspond to the same confidence level can be combined.

In the Hessian approach [12, 13] adopted by CT10 and MSTW, a small number of independent PDF eigenvector sets spans only the boundary of the 68% or 90% c.l. hypervolume in their respective PDF parameter space. In the Monte-Carlo (MC) approach of NNPDF [15], the same hypervolume is populated by discrete unweighted PDF replicas that can be used to reconstruct the probability density in space of PDF parameters. To combine the Hessian and MC error sets, we first generate MC replicas from the CT10 and MSTW2008 meta-sets using a prescription that is close to the one in [53]. The distribution of replicas from all ensembles will approximate the *a priori* unknown probability distribution in space of 66 meta-PDF parameters. With a sufficient number of replicas generated from the original sets, this probability distribution can be reconstructed with increasing accuracy.

We will follow a simplistic but not unreasonable thinking that the parameters of the META ensemble roughly obey a 66-dimensional Gaussian distribution. We will also symmetrize the PDF errors and include all PDF ensembles with the same weights to simplify many considerations.

The combined META PDF uncertainties can be estimated using either the Hessian or MC approach. The latter could follow the strategy that is similar to the combination of Monte-Carlo replicas from various ensembles discussed in [29, 30]. The Hessian approximation is at least as attractive, as we found that only about 100 eigenvector sets are required to estimate the combined confidence levels regardless of the number of the input error sets that we tried. Furthermore, if we are interested in the uncertainty of a specified observable, the number of needed Hessian META sets can be reduced to a few by the procedure dubbed 'data set diagonalization' [54], the possibility that we investigated in a follow-up study [55].

3.2 Generation of MC replicas from Hessian ensembles

The first step in the derivation of the META ensemble consists in generating the MC replicas for the CT and MSTW ensembles. Initially each of them contains $2N_{\mathrm{eig}}$ Hessian eigenvector sets distributed *on the boundary of* their respective hypervolume, corresponding to 90% c.l. for both CT and MSTW. The MC replicas for each ensemble are produced by generating groups of N_{eig} random numbers R_j, each sampled independently from a standard normal distribution in the interval $-\infty < R_j < +\infty$. For every group of R_j, we construct a pair of MC replicas $f_{\mathrm{MC}}(x, Q_0)$ for each flavor by

$$f_{\mathrm{MC}}(x, Q_0) = f_0(x, Q_{,0}) \pm \sum_{j=1}^{N_{\mathrm{eig}}} \frac{R_j}{r} \cdot \frac{f_{j,+}(x, Q_0) - f_{j,-}(x, Q_0)}{2}, \tag{3.1}$$

where $f_0(x, Q_0)$ indicates the central PDF, and $f_{j,\pm}(x, Q_0)$ are two Hessian error sets associated with the positive and negative displacements along the j-th eigenvector. The factor r is 1.64 to convert from the 90% to 68% c.l., assuming the Gaussian distribution of the PDF parameters.[3] This formula symmetrizes the displacements of the MC replicas from the start, as indicated by "$\pm$". It is equivalent to the generation of MC replicas for Hessian PDF sets in [53], but operates with symmetrically distributed replicas.[4]

By doing this we can convert any Hessian eigenvector set into an arbitrary number N_{rep} of MC replicas representing the probability density in $\{a\}$ space. The underlying assumption (which holds well in the $\{x, Q\}$ regions with sufficient kinematic constraints) is that the input probability distributions are close to the Gaussian ones, not too asymmetric, and compatible with one another. By trial and error, we found that $N_{\mathrm{rep}} = 100$ of replicas per ensemble is sufficient for the combination of errors. This is the number of MC replicas that was included from each PDF ensemble.

3.3 Constructing META PDFs

Given N_{rep} values $a_i(k)$ of a parameter a_i on a group g of PDF sets, we can find the expectation value $\langle a_i \rangle_g$, sample covariance $\mathrm{cov}(a_i, a_j)_g$, and standard deviation $(\delta a_i)_g$ of

---

[3]If the MC replicas are generated from a 68% c.l. eigenvector set, $r = 1$.

[4]In ref. [53], the MC replicas are generated either for PDF sets or for QCD observables. We follow the first option, by generating the MC replicas for the PDFs themselves.

a_i on g:

$$\langle a_i \rangle_g = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} a_i(k), \tag{3.2}$$

$$\text{cov}(a_i, a_j)_g = \frac{N_{\text{rep}}}{N_{\text{rep}} - 1} \langle (a_i - \langle a_i \rangle_g) \cdot (a_j - \langle a_j \rangle_g) \rangle_g, \tag{3.3}$$

$$(\delta a_i)_g = \sqrt{\text{cov}(a_i, a_i)_g}. \tag{3.4}$$

When N_g groups of PDF replicas ($N_g = 3$: CT10, MSTW'08, and NNPDF2.3) are combined into a META ensemble, the expectation value $\langle a \rangle_{\text{META}}$ of a on the whole ensemble of $N_{\text{rep}} \cdot N_g$ replicas can be found by averaging the expectation values on each group:

$$\langle a_i \rangle_{\text{META}} = \frac{1}{N_g} \sum_g \langle a_i \rangle_g. \tag{3.5}$$

The summation over "g" runs over all PDF groups. The sample covariance and standard deviation on the META ensemble are similarly related to their values on the individual groups by (assuming $N_{\text{rep}} \gg 1$)

$$\text{cov}(a_i, a_j)_{\text{META}} = \frac{1}{N_g} \sum_g \text{cov}(a_i, a_j)_g + \frac{1}{N_g} \sum_g \langle a_i \rangle_g \langle a_j \rangle_g - \langle a_i \rangle_{\text{META}} \langle a_j \rangle_{\text{META}}, \tag{3.6}$$

and

$$(\delta a_i)^2_{\text{META}} = \frac{1}{N_g} \sum_g (\delta a_i)^2_g + \frac{1}{N_g} \sum_g \langle a \rangle^2_g - \langle a \rangle^2_{\text{META}}. \tag{3.7}$$

Given these equations, the expectation value and standard deviation for any parameter or observable a on the META ensemble can be derived either by averaging on the full META ensemble according to eqs. (3.2)–(3.4) (and identifying the group g with the whole META ensemble); or by averaging the expectation values and standard deviations on the individual input groups with the help of eqs. (3.5) and (3.7). However, since much of the information contained in 300 replicas is redundant, we can reduce the number of needed error PDFs by constructing a smaller Hessian eigenvector ensemble that will still be sufficient for estimating the confidence intervals.

Toward this goal, we select a new basis $a'_i = O_{ij} \cdot (a_j - \langle a_j \rangle_{\text{META}})$ that diagonalizes the covariance matrix, $O_{ki} \cdot \text{cov}(a_i, a_j)_{\text{META}} \cdot O^T_{jl} = \lambda_k \delta_{kl}$. Each new parameter a'_i follows an independent one-dimensional Gaussian distribution with an expectation value 0 and standard deviation $\sqrt{\lambda_i}$. Here O_{ij} is an orthogonal matrix, and λ_i are the eigenvalues of the covariance matrix. The eigenvalues λ_i in the PDF analysis span many orders of magnitude [12]. Those directions in $\{a'\}$ space that are associated with very small λ_i have negligible contributions to the PDF uncertainties, so that the corresponding parameters a'_i can be fixed at their central values, thus reducing the number of independent eigenvectors. For each remaining eigenvector direction a'_i, we construct two meta-PDF sets corresponding to the lower and upper boundaries of the respective 68% c.l. interval.

In the end, we obtain a set of Hessian eigenvector PDFs that spans the 68% c.l. hypersurface on the ensemble of META PDF replicas. We can repeat the same procedure to construct the 90% c.l. META PDF replicas.

In the current analysis, we found that a Hessian set with 50 eigenvectors (or 100 eigenvector sets) is sufficient for estimation of the combined PDF uncertainty. The PDF parameter associated with each eigenvector direction follows an independent Gaussian distribution, thus we can compute the PDF uncertainties and correlation angles for arbitrary observables according to the master formulas of the Hessian formalism:

$$\delta^H(X) = \frac{1}{2}\sqrt{\sum_{i=1}^{N_{\rm eig}}[X_i^+ - X_i^-]^2},$$

$$\delta_+^H(X) = \sqrt{\sum_{i=1}^{N_{\rm eig}}[\max(X_i^+ - X_C^H, \ X_i^- - X_C^H, \ 0)]^2},$$

$$\delta_-^H(X) = \sqrt{\sum_{i=1}^{N_{\rm eig}}[\max(X_C^H - X_i^+, \ X_C^H - X_i^-, \ 0)]^2}, \qquad (3.8)$$

and

$$\cos(\Delta\varphi) = \frac{1}{4\delta(X)\delta(Y)}\sum_{i=1}^{N_{\rm eig}}(X_i^+ - X_i^-)(Y_i^+ - Y_i^-). \qquad (3.9)$$

The 68% or 90% c.l. hypervolumes of the META ensemble enclose the central predictions of all PDF groups, as is illustrated on the example of some gluon or u quark PDF parameters in figures 14 or 15. As before, these comparisons are obtained with a common $\alpha_s(M_Z)$ value of 0.118. The smaller (larger) ellipse corresponds to the 68% (90%) confidence region. The markers indicate parameter combinations for the best-fit PDFs from different groups. They are localized inside the 90% error ellipse of the META prescription in general. Almost all the best-fit PDFs are within the 90% c.l. ellipse of the META ensemble, except for some parameters of the ABM and HERA meta-PDFs.

Going back to the x space representation, in figures 16 and 17 we compare the 90% c.l. error band of META PDFs with the central PDFs with $\alpha_s(M_Z) = 0.118$ from five groups, all normalized to the central META PDF. Again, the central PDFs of three global ensembles lie within the 90% c.l. uncertainty bands of the META PDF. The ABM and HERA PDFs for u, $\bar{u}$, $\bar{d}$, g flavors can be outside of the META bands in some x regions. The largest deviation is observed in the bottom quark PDF for ABM11, which is very different at the scale 8 GeV compared to the other ensembles due to the different treatment of heavy quarks. This difference of the bottom distribution is largely reduced as going to higher scales like 85 GeV.

3.4 Combination with the α_s uncertainty

To estimate the PDF+α_s uncertainty using the META PDF ensemble, we provide a series of additional meta-sets for $\alpha_s(M_Z) \neq 0.118$, obtained by taking the average of the corresponding PDFs from the three groups. Depending on the PDF analysis, $\alpha_s(M_Z)$ is either
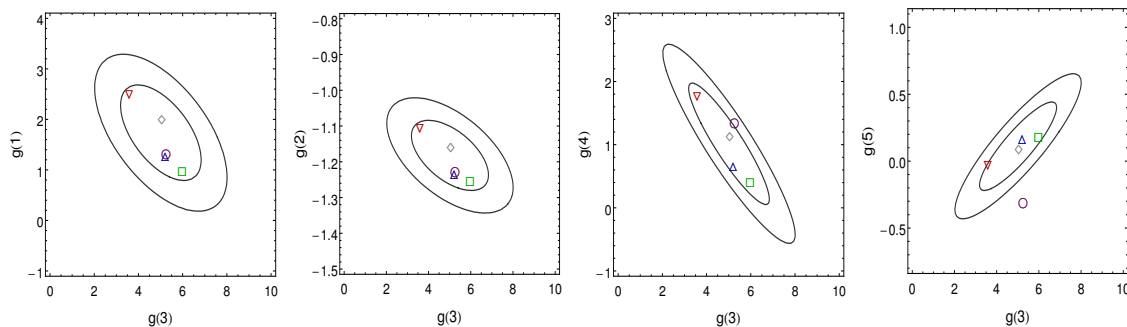
Figure 14. Comparison of META PDF confidence intervals with central NNLO PDFs of the input PDF ensembles in space of meta-parameters a_{1-5} for the gluon PDF. Up triangle, down triangle, square, diamond, and circle correspond to the best-fit PDFs from CT10, MSTW, NNPDF, HERA-PDF, and ABM respectively. The ellipses correspond to 68 and 90% c.l. ellipses of META PDFs.
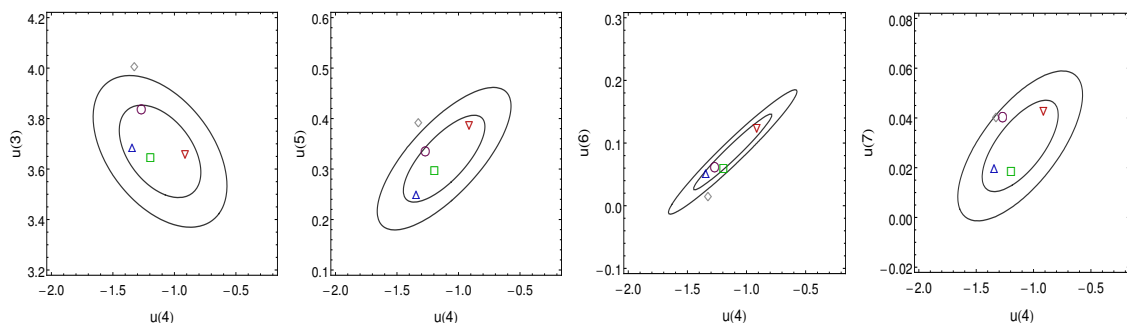

Figure 15. Same as figure 14, for a_{3-7} of the u quark PDF.

treated as an external parameter determined by its world average (as in CT10 [56]) or fitted together with the PDFs (as in MSTW [57] or NNPDF [58]). In the latter case, the best-fit value of $\alpha_s(M_Z)$ from the PDF fit at NNLO (0.1171 from MSTW, 0.1173 from NNPDF, and 0.1132 from ABM12) tends to be lower than the latest world average of 0.1184 ± 0.0007 at 68% c.l. [31]. The PDF4LHC recommendation [27, 28] suggests a slightly larger error of 0.0012 at 68% c.l. than in the world average. Note that the PDF4LHC recommendation also takes different central values of $\alpha_s(M_Z)$ for different PDF ensembles, which enlarges the α_s uncertainty in the end.

Given these options, one can come up with several ways for calculating the combined PDF+α_s uncertainty, depending on the interpretation and statistical confidence assigned to the α_s input. The combined uncertainty may vary significantly depending on the prescription, especially for the QCD observables that are directly sensitive to α_s or the gluon PDF. Here we suggest several practical possibilities for estimating the PDF+α_s uncertainty using the META PDFs.

1. We may take the world-average α_s value as an external input for the PDF analysis and assume that the α_s uncertainty is decoupled from the systematic errors in the fit due to missing higher-order effects. In this approach, the 68% c.l. error on $\alpha_s(M_Z)$ is
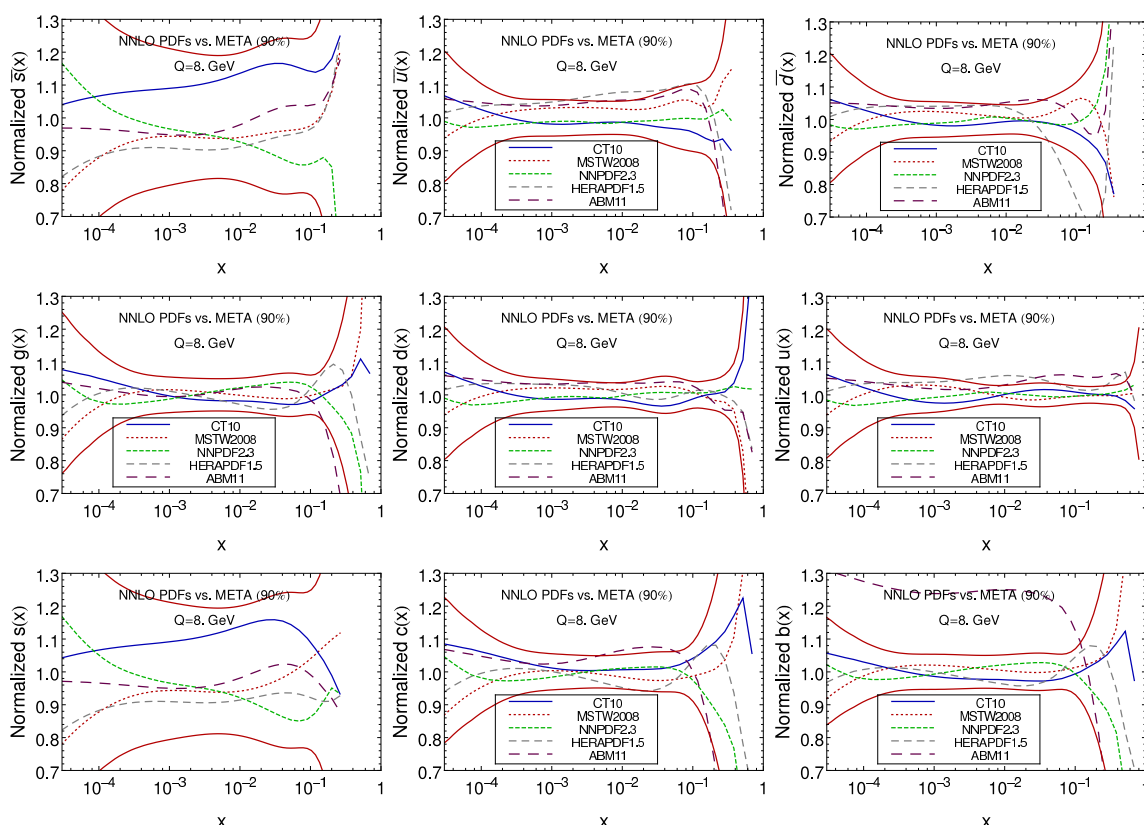
Figure 16. Comparison of the META PDFs and all the best-fit NNLO PDFs with $\alpha_s(M_Z) = 0.118$ at a common scale $Q = 8\,\text{GeV}$.

given by the world average of ± 0.0007. We vary α_s in this range when estimating the α_s uncertainty on a QCD observable and add it to the PDF uncertainty in quadrature, which produces the PDF+α_s uncertainty with full correlations, as has been discussed by CTEQ group [56].

2. An alternative prescription is to replace the input $\alpha_s(M_Z)$ error by 0.0012 at 68% c.l. as in the PDF4LHC recommendation [27, 28], and add the α_s and PDF uncertainties in quadrature. The resulting PDF+α_s uncertainty will be smaller than the one from the PDF4LHC recommendation, as a single value of 0.118, rather than the PDF4LHC envelope of central α_s values from all ensembles, is used in the META case.

3. Lastly, we can enlarge the input $\alpha_s(M_Z)$ error to 0.002 at 68% c.l. and add the α_s and PDF uncertainties linearly. This prescription fully covers the preferred α_s values by different groups and is numerically close to the PDF+α_s uncertainty based on the PDF4LHC envelope.

The first two conventions predict smaller α_s and PDF+α_s uncertainties than the envelope prescription used by the 2010 PDF4LHC study [27, 28] and 2012 PDF benchmarking
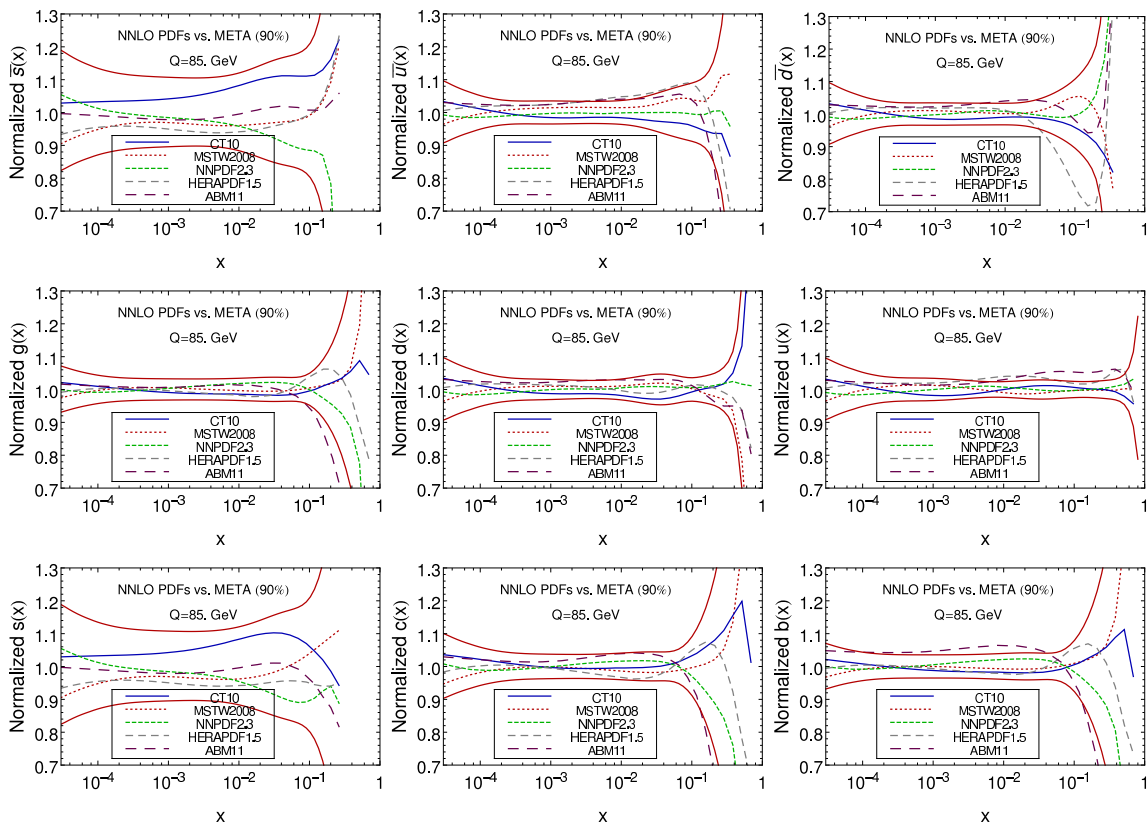
Figure 17. Same as figure 16, for $Q = 85$ GeV.

study [16]. In these cases the PDF uncertainty always dominates in the combined uncertainty. The third convention turns out to be numerically close to the envelope prescription. In our remaining comparisons the PDF+α_s uncertainties are estimated based on the third convention.

4 Phenomenological applications

The META PDF ensemble incorporates measurements from HERA, fixed-target experiments, and Tevatron. Although the NNPDF2.3 ensemble used here includes some of the early LHC data in the fit, like jet production cross sections [52], W and Z rapidity distributions [59, 60], and W electron charge asymmetry [61], we do not expect these relatively weak LHC constraints to impact strongly the META PDFs' behavior. Thus the META ensemble can be used for any LHC predictions based on non-LHC measurements.

Before discussing the LHC applications, we would like to review two important features of the META PDFs. First, as mentioned earlier, the sum rule constraints are not enforced directly in the meta-fit. After obtaining our final META PDFs, we examined the partial integrals for the momentum and valence sums in the same fashion as in figure 2. The sum rules are obeyed by the META PDFs at the accuracy that is similar, or better than, the individual PDF ensembles [37].
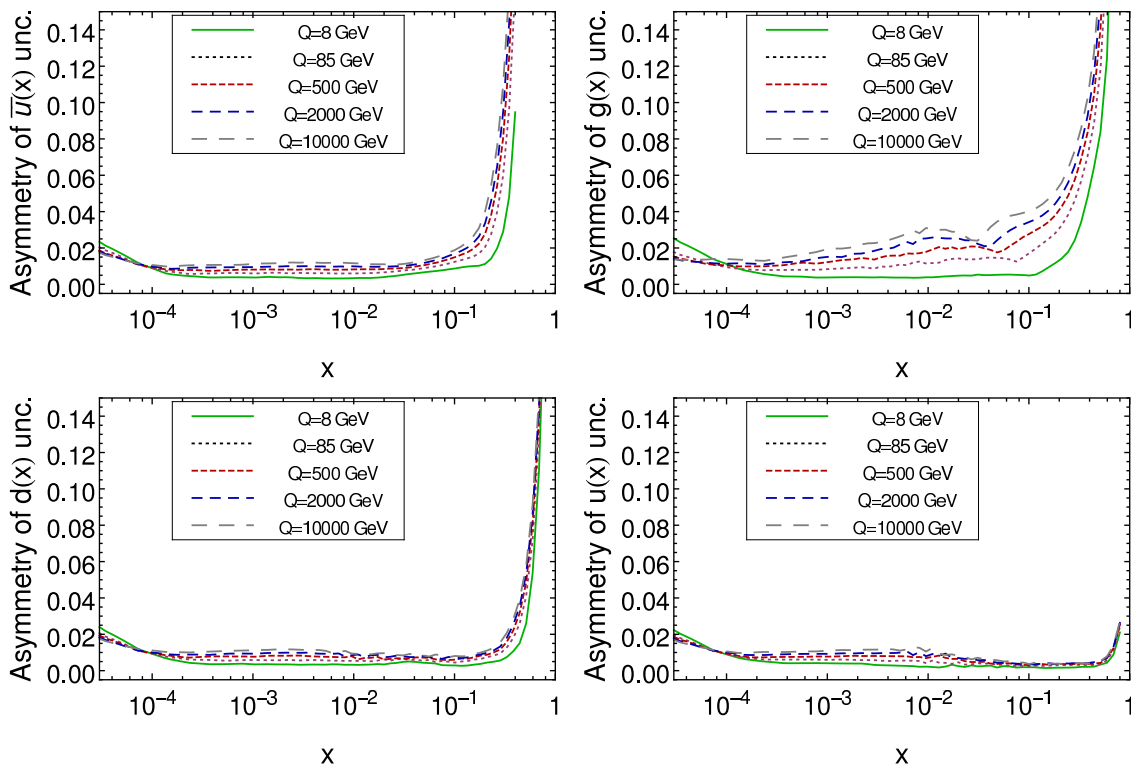
Figure 18. Asymmetry of PDF errors of different flavors for the META PDFs at different Q values.

The Hessian method of error propagation relies on the assumption of the approximately linear dependence of QCD observables on the PDF parameters in the vicinity of the central fit. Uncertainties of the META PDF parameters are symmetric by definition, but the uncertainties of the PDF functional forms are generally not. We verified the validity of the linear approximation by first checking the asymmetry $|\delta_+^H - \delta_-^H|/(\delta_+^H + \delta_-^H)$ of the uncertainties on the PDFs and PDF luminosities, where $\delta_\pm^H$ indicate the asymmetric errors according to eq. (3.8). For example, in figure 18 we plot the asymmetry of PDF errors of different flavors, which gives an estimation of the non-linear behavior of the PDFs due to both the parametrization and the PDF evolution. The asymmetries observed in the figure are very small, at the level of a few percent in the region with $x < 0.1$. They can exceed 15% in the large x region, where the PDFs are not well constrained. As Q increases from 8 to 10000 GeV, the asymmetry slowly spreads from large toward smaller x as a consequence of the PDF evolution. In figure 19 we further show the asymmetry of the PDF uncertainty of the parton luminosities in different kinematic regions at the LHC 8 TeV [62], as functions of the invariant mass and rapidity of the produced final state. These asymmetries are also small except for very large invariant masses or rapidities that correspond to the x values close to 1. By investigating the asymmetry of the PDFs as well as of sample LHC cross sections, we conclude that the linear dependence of most LHC observables on small variations of the PDF parameters is well satisfied.

Using the eigenvector sets and α_s series of the META ensemble, we can calculate central predictions and uncertainties of the cross sections at the LHC and compare them with the
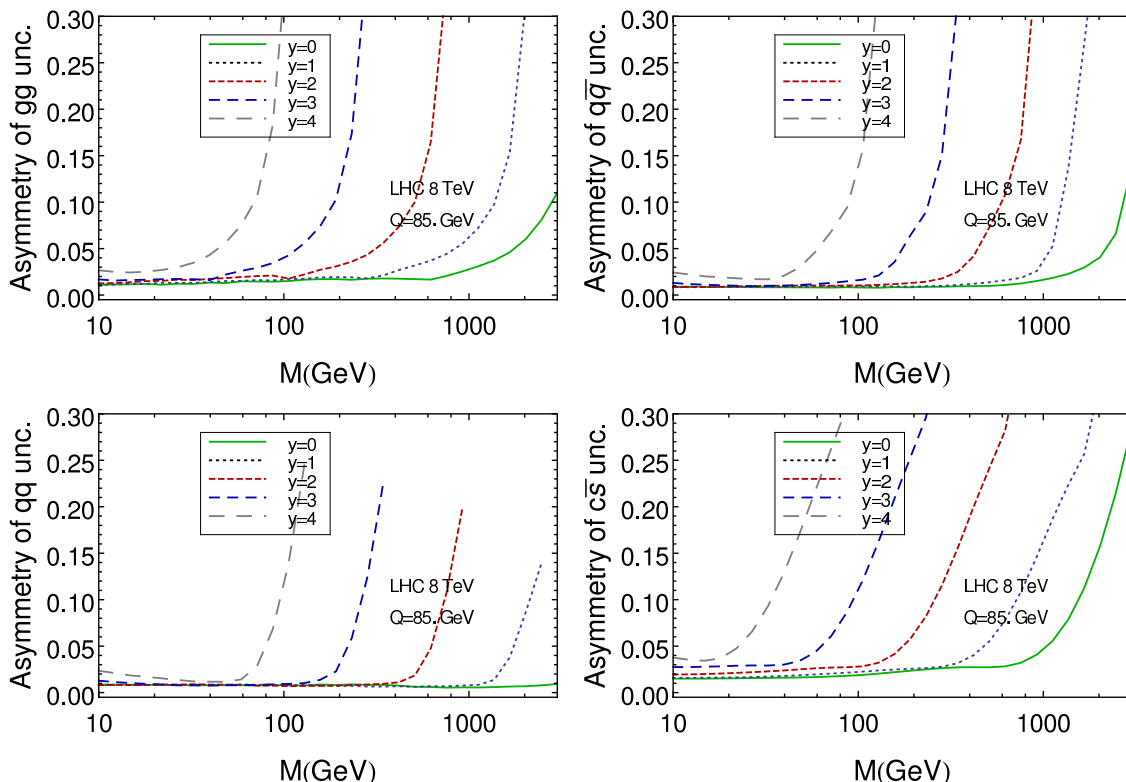
Figure 19. Asymmetry of PDF errors of parton luminosities for the META PDFs.

results from individual groups. As an example, figures 20–22 show NNLO predictions for W, Z, SM Higgs, $t\bar{t}$, and NLO predictions for inclusive jet production, using the same settings as in figures 11–12. The solid horizontal lines indicate the central prediction from the Hessian META PDF ensemble, the dotted and dashed ones correspond to the 68% c.l. PDF uncertainty and PDF+α_s uncertainty of the META ensemble, respectively. For every group except ABM, we show the central predictions with $\alpha_s(M_Z) = 0.118$ and the 68% c.l. PDF (PDF+α_s) uncertainties as the left (right) error bar. The rightmost points in each inset correspond to the central predictions and 68% c.l. PDF+α_s uncertainties for the ABM12 PDF ensemble with 5 active flavors [8], computed according to their convention for $\alpha_s(M_Z) = 0.1132$ and top quark pole mass of 171 GeV. As we expect, the META PDFs work as an average of CT10, MSTW2008, and NNPDF2.3 PDFs, as can be noticed by comparing the central predictions. The PDF or PDF+α_s uncertainties of the META ensemble are slightly smaller than the envelope prescription in the benchmarking study [16]. For instance, the NNLO Higgs cross section through gluon fusion is 18.75 ± 1.24 pb for LHC 8 TeV according to the envelope prescription, while it is 18.78 ± 1.15 pb for the META PDF.

Figure 23 shows the 90% c.l. tolerance ellipses for pairs of cross sections computed using either the META PDFs or original PDF ensembles with $\alpha_s(M_Z) = 0.118$. The ellipses are computed according to eq. (2.5) and correspond to the probability density contours of 90%. The ellipses of the META PDFs preserve the PDF-induced (anti-)correlations observed in the three input PDF ensembles.
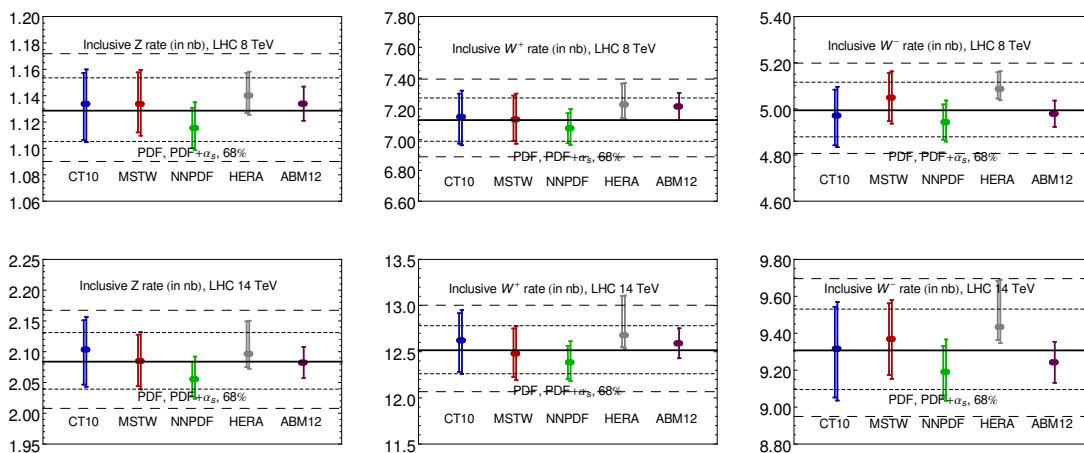
Figure 20. Comparison of inclusive $W^{\pm}$ and Z cross sections at NNLO for various PDF ensembles. The PDF and PDF+α_s errors of META PDFs (at 68% c.l.) are shown by the short-dashed and long-dashed lines.
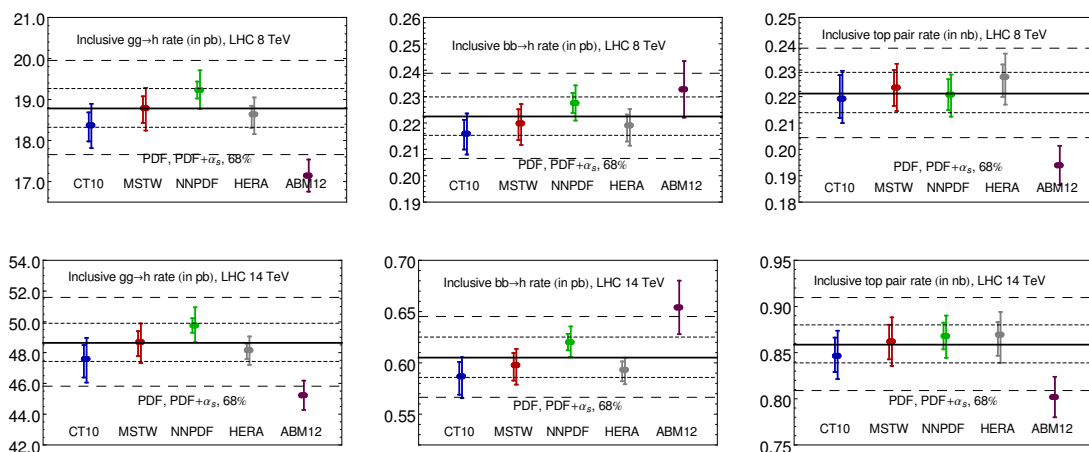

Figure 21. Comparison of inclusive Higgs boson and top quark pair production cross sections at NNLO for various PDF ensembles. The PDF and PDF+α_s errors of META PDFs (at 68% c.l.) are shown by the short-dashed and long-dashed lines.

In a follow-up paper, we demonstrate that the META PDF approach performs as well for various differential observables that we explored. The META PDFs provide an average of three input PDF ensembles for various distributions, such as NNLO rapidity distributions of W and Z bosons produced at the LHC in figure 24, obtained with Vrap [42] at the NNLO. The rapidity distributions are normalized to the central predictions of the META PDFs, the bands indicate 90% c.l. PDF uncertainties. The central predictions of the META PDFs represent the average of the three global sets across a wide rapidity range. The META uncertainties cover the spread of the input PDF ensembles.
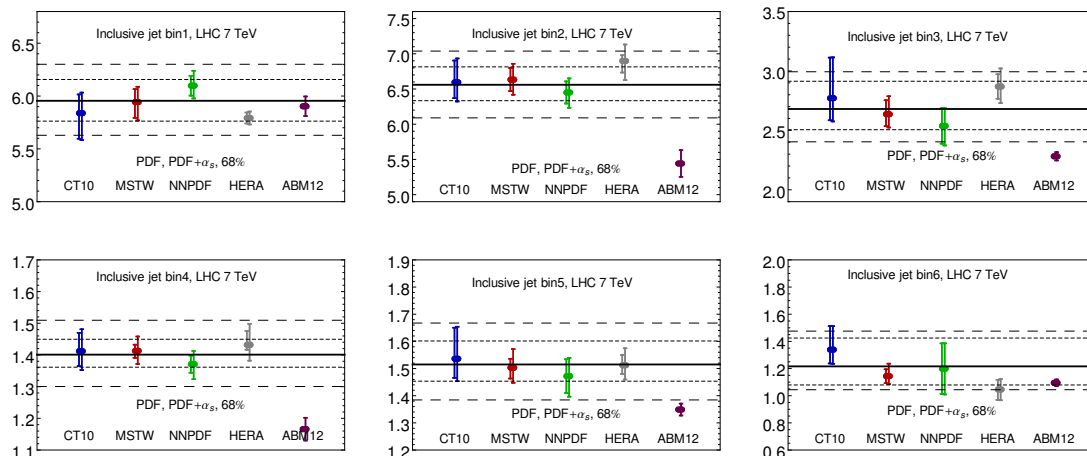
Figure 22. Comparison of inclusive jet production cross sections (in arbitrary unit) at NLO for various PDF ensembles. The PDF and PDF+α_s errors of META PDFs (at 68% c.l.) are shown by the short-dashed and long-dashed lines.
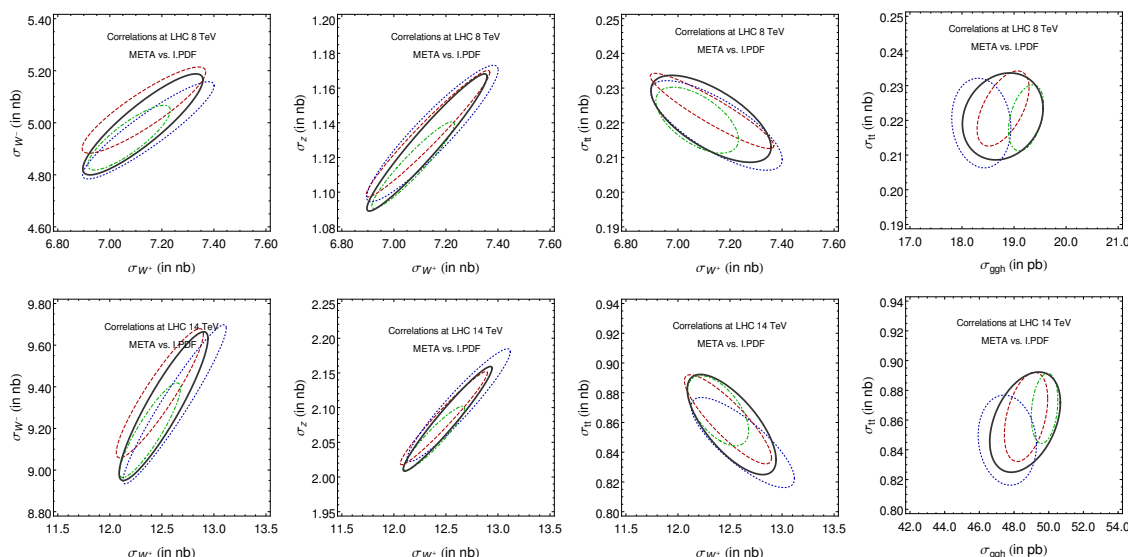

Figure 23. Comparison of 90% error ellipses of NNLO cross sections using various ensembles: META PDF(black solid), CT10(blue dotted), MSTW(red dashed), NNPDF(green dot-dashed).

5   Summary

Faithful estimation of the PDF dependence in theoretical predictions is relevant practically and challenging conceptually. Diversity of PDF ensembles provided by several groups reflects a variety of factors impacting the PDFs in modern QCD calculations. Comparisons of PDFs deal with their miscellaneous assumptions and disparate formats, as well as with large numbers of member sets included in the PDF ensembles. To estimate the net uncertainty in QCD cross sections due to the PDF inputs, theoretical predictions are currently
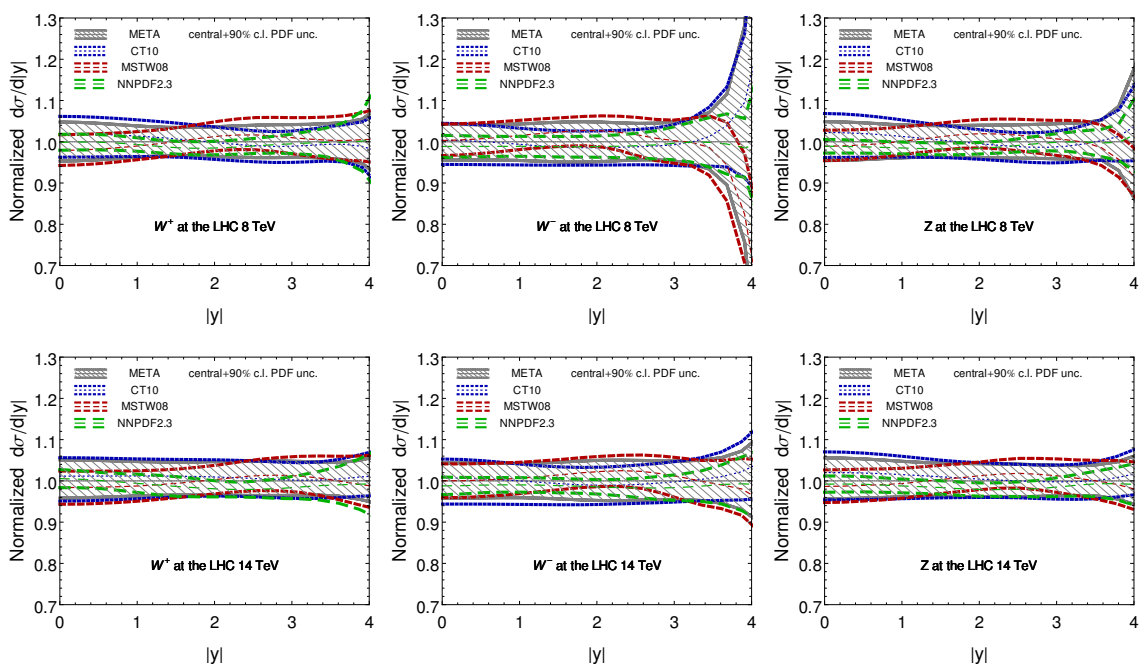
Figure 24. Comparison of rapidity distributions of the W and Z boson produced at the LHC. All distributions are normalized to the central predictions of the META PDFs.

calculated on a process-by-process basis for multiple PDF sets and combined at the final stage based on a certain prescription, e.g., the PDF4LHC recommendation [16, 27, 28] or the replica combination method [29, 30]. These approaches, while providing a reasonable uncertainty estimate, handle the PDFs inefficiently and promote lengthy computations. Much of the information contained in the input PDF sets is discarded in the final combination. Development of fast interpolation interfaces for (N)NLO cross sections, such as ApplGrid [63] or FastNLO [51, 64], speeds up the computation, but does not eliminate the key hurdle of inefficient information processing, especially when simulations involve many scattering processes.

The meta-analysis described in this study follows an alternative approach in which a variety of PDFs are combined into a single PDF ensemble *before* the individual theoretical predictions are computed. Such ensemble is suitable for most LHC applications and consists of a relatively small number of error PDFs that reproduce the total PDF+α_s uncertainty provided by an arbitrary number of the input PDF ensembles. In the specific example that we constructed, a META ensemble is comprised of 100 Hessian eigenvector sets for evaluating the 68% c.l. combined uncertainty of CT10, MSTW'2008, and NNPDF2.3 at NNLO. The three ensembles are selected because they can be combined with minimal obstacles. They correspond to close values of the QCD coupling strength $\alpha_s(M_Z)$ (compatible with the world average of 0.118) and, although they follow different heavy-quark schemes, these differences can be absorbed into their parametrizations at the initial evolution scale of META PDFs chosen to be above the b quark mass (at 8 GeV).

Each error set of the input ensembles is approximated by an auxiliary functional form (a meta-parametrization) in the kinematic range typical for LHC studies, taken to be $x > 10^{-5}$ and $8 < Q < 10000\,\mathrm{GeV}$. As the meta-parametrizations of all error sets share the same functional form, distributions of their free parameters (66 in total) can be easily compared. We determine these parameters at a common coupling $\alpha_s(M_Z)$ of 0.118. Small differences in the input α_s values are compensated for by minor shifts of the respective meta-parameters found using the α_s series of the input PDF sets.

Then we combine the meta ensembles from CT10, MSTW2008 and NNPDF2.3 by generating Monte-Carlo replicas for each Hessian input ensemble and diagonalizing the covariance matrix constructed from the Monte-Carlo replicas. The final result consists of a central META PDF set, equivalent to the unweighted average of three input central sets, and 100 eigenvector META sets obtained in the Hessian method [12, 13]. The PDF uncertainty for the META ensemble is computed according to the master formulas in eq. (3.8). We also provide a PDF α_s series of the META PDFs for calculating the PDF+α_s uncertainty, including the correlations, by adding the PDF and α_s uncertainties in quadrature [56]. The META PDFs are stored in a tabulated format together with an interface for their numerical interpolation at [37].

The META ensemble predicts about the same PDF+α_s uncertainty in the key LHC observables as the PDF4LHC recommendation, and in addition provides a natural way to estimate correlations of different observables. In comparison to massive computation efforts characteristic of LHC simulations, the construction of the META ensemble is relatively simple and was carried out entirely by using standard functions for statistics and data analysis in *Mathematica 8*.

Our results present an initial attempt to combine LHC predictions from different PDF groups independently of the processes studied. We note several avenues for further developments of this approach. The simple combination procedure that was tried (equivalent to using an unweighted average for finding expectation values) would be inappropriate for including too disparate PDF ensembles or for non-Gaussian parameter distributions. In our case, the three global ensembles that were combined are similar in their fitted experimental samples, $\alpha_s(M_Z)$, central PDFs, and nominal PDF uncertainties. We also examined the meta-parametrizations for ABM'11 and HERAPDF1.5 ensembles and observed that their central PDF sets for $\alpha_s(M_Z) = 0.118$ sometimes lie outside of the 90% c.l. error bands of the META ensemble, see figures 16 and 17. We could not fully consider the ABM'11 error sets in the current combination at a fixed $\alpha_s(M_Z) = 0.118$, as $\alpha_s(M_Z)$ in the ABM error sets varies around a low central value of $\alpha_s(M_Z) = 0.1134$. For HERAPDF1.5, a part of the difference with the global sets can be attributed to a larger PDF uncertainty on some combinations of the HERA PDF parameters, given their smaller experimental data set. Such an ensemble could be added by averaging the META parameters with statistical weights that account for different sizes of the experimental data samples.

The error PDFs of the future META ensembles can be obtained using either the Monte Carlo (MC) sampling or Hessian methods. In the MC approach [29, 30], probability distributions of arbitrary complexity can be in principle described, while selection of replicas is simplified via using a shared meta-parametrization form. Constraints from new exper-
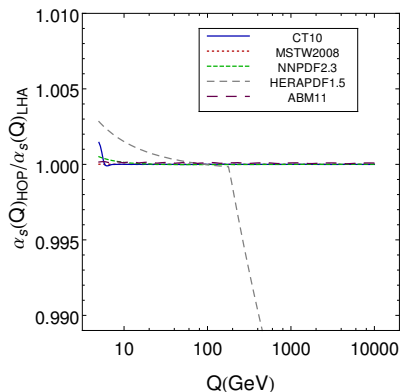
Figure 25. Comparison of the QCD coupling strengths $\alpha_s(Q)$ from five PDF ensembles as tabulated in LHAPDF (LHA) or evolved with $N_f = 5$ by HOPPET (HOP).

iments can be imposed using the PDF reweighting technique [14, 65, 66] in either MC or Hessian approaches. The method of data set diagonalization [54] can be applied in the case of the Hessian representation to find a small number of eigenvector sets that dominate the PDF uncertainty of given observables. While the 100 META eigenvector sets are designated for a wide range of LHC theoretical predictions, a much smaller number (6–10) of eigenvector sets dominates the uncertainty in various important processes, such as massive electroweak boson production [55]. The framework of the meta-analysis is well-suited for exploring these possibilities.

Acknowledgments

A   A test of QCD evolution

In this appendix, we verify that the dependence of $\alpha_s(Q)$ and PDFs $\Phi(x, Q)$ on the scale Q in the input ensembles is compatible and allows their combination. We compared the tabulated evolution of $\alpha_s(Q)$ and $\Phi(x, Q)$ for five ensembles from the LHAPDF library [40] with the numerical evolution by the HOPPET program [41] from the initial scale $Q_0 = 8\,\text{GeV}$ up to $10000\,\text{GeV}$. In all cases, we assume 3 QCD loops and 5 flavors in the QCD beta-function and splitting kernels.

Figure 25 shows the ratio of $\alpha_s(Q)$ that we get from LHAPDF to the corresponding values from HOPPET using the original $\alpha_s(M_Z)$ values of the input ensembles listed in table 1. The agreement of the tabulated and evolved $\alpha_s(Q)$ is perfect for all ensembles except for HERAPDF, where a kink at the top quark mass occurs because HERAPDF uses the 6-flavor $\alpha_s(Q)$ above the top quark threshold.
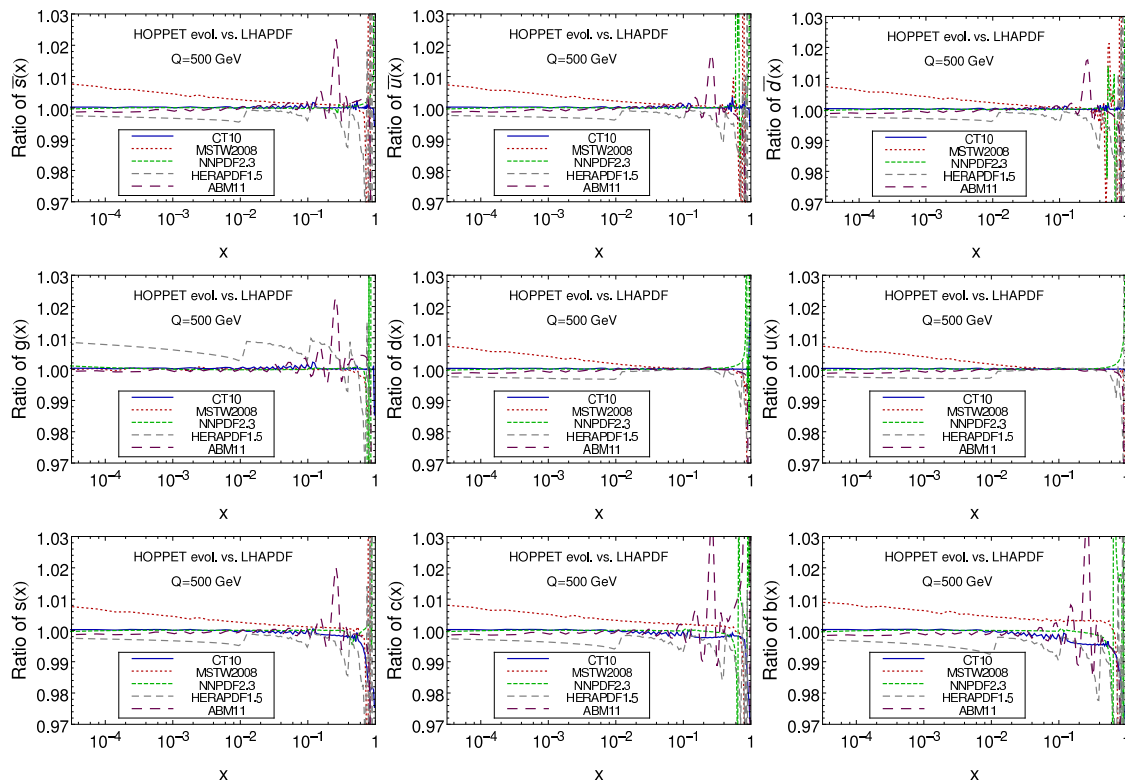
Figure 26. Comparison of the PDF evolution from different best-fit PDF sets with HOPPET for $Q = 500$ GeV.

The PDF evolution is examined by computing ratios of $\Phi(x, Q)$ returned by LHAPDF and HOPPET at several Q values above 8 GeV. The x-dependent ratios at $Q = 500$ GeV shown in figure 26 are representative of the general trend at other Q values. We observe excellent agreement at the level of 0.1% between LHAPDF and HOPPET for both CT10 and NNPDF almost across the entire shown region of $10^{-5} \leq x < 1$, except for $x \gtrsim 0.5$, where differences in the Q dependence can be a few percent for all sea PDFs. In the case of ABM, fluctuations for all sea PDFs are observed at somewhat smaller values of x of about 0.3, but again they happen in the region where the sea PDFs are small, and the accuracy of LHAPDF interpolation is sufficient compared to the PDF uncertainties. The tabulated and numerical evolution of quark PDFs from MSTW shows a systematic discrepancy reaching 1% at $x < 0.05$, while the tabulated and numerical evolution of the MSTW gluon is in a perfect agreement. The tabulated and evolved HERA PDFs differ by $\approx 0.5\%$ across most x, which is expected at $Q = 500$ GeV given the 5-flavor evolution in HOPPET and 6-flavor evolution of the HERAPDF's tabulated $\alpha_s(Q)$.

References

[1] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B 716** (2012) 1 [arXiv:1207.7214] [INSPIRE].

[2] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B 716** (2012) 30 [arXiv:1207.7235] [INSPIRE].

[3] Higgs Working Group Report, Snowmass'2013 Community Planning study, http://www.snowmass2013.org/tiki-index.php?page=The+Higgs+Boson.

[4] J. Gao et al., *The CT10 NNLO Global Analysis of QCD*, *Phys. Rev.* **D 89** (2014) 033009 [arXiv:1302.6246] [INSPIRE].

[5] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C 63** (2009) 189 [arXiv:0901.0002] [INSPIRE].

[6] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B 867** (2013) 244 [arXiv:1207.1303] [INSPIRE].

[7] S. Alekhin, J. Blumlein and S. Moch, *Parton Distribution Functions and Benchmark Cross Sections at NNLO*, *Phys. Rev.* **D 86** (2012) 054009 [arXiv:1202.2281] [INSPIRE].

[8] S. Alekhin, J. Bluemlein and S. Moch, *The ABM parton distributions tuned to LHC data*, *Phys. Rev.* **D 89** (2014) 054028 [arXiv:1310.3059] [INSPIRE].

[9] P. Jimenez-Delgado and E. Reya, *Dynamical NNLO parton distributions*, *Phys. Rev.* **D 79** (2009) 074023 [arXiv:0810.4274] [INSPIRE].

[10] P. Jimenez-Delgado and E. Reya, *Variable Flavor Number Parton Distributions and Weak Gauge and Higgs Boson Production at Hadron Colliders at NNLO of QCD*, *Phys. Rev.* **D 80** (2009) 114011 [arXiv:0909.1711] [INSPIRE].

[11] ZEUS and H1 collaborations, A.M. Cooper-Sarkar, *PDF Fits at HERA*, PoS(EPS-HEP2011)320 [arXiv:1112.2107] [INSPIRE].

[12] J. Pumplin et al., *Uncertainties of predictions from parton distribution functions. 2. The Hessian method*, *Phys. Rev.* **D 65** (2001) 014013 [hep-ph/0101032] [INSPIRE].

[13] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012 [hep-ph/0201195] [INSPIRE].

[14] W.T. Giele and S. Keller, *Implications of hadron collider observables on parton distribution function uncertainties*, *Phys. Rev.* **D 58** (1998) 094023 [hep-ph/9803393] [INSPIRE].

[15] NNPDF collaboration, L. Del Debbio, S. Forte, J.I. Latorre, A. Piccione and J. Rojo, *Neural network determination of parton distributions: the nonsinglet case*, *JHEP* **03** (2007) 039 [hep-ph/0701127] [INSPIRE].

[16] R.D. Ball et al., *Parton distribution benchmarking with LHC Data*, *JHEP* **04** (2013) 125 [arXiv:1211.5142] [INSPIRE].

[17] J.F. Owens, A. Accardi and W. Melnitchouk, *Global parton distributions with nuclear and finite-Q^2 corrections*, *Phys. Rev.* **D 87** (2013) 094012 [arXiv:1212.1702] [INSPIRE].

[18] J. Pumplin, H.L. Lai and W.K. Tung, *The Charm Parton Content of the Nucleon*, *Phys. Rev.* **D 75** (2007) 054029 [hep-ph/0701220] [INSPIRE].

[19] S. Dulat et al., *Intrinsic Charm Parton Distribution Functions from CTEQ-TEA Global Analysis*, *Phys. Rev.* **D 89** (2014) 073004 [arXiv:1309.0025] [INSPIRE].

[20] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, *Parton distributions incorporating QED contributions*, *Eur. Phys. J.* **C 39** (2005) 155 [hep-ph/0411040] [INSPIRE].

[21] NNPDF collaboration, R.D. Ball et al., *Parton distributions with QED corrections*, *Nucl. Phys.* **B 877** (2013) 290 [arXiv:1308.0598] [INSPIRE].

[22] E.M. Askanazi, K.A. Holcomb and S. Liuti, *Self-Organizing Maps Parametrization of Deep Inelastic Structure Functions with Error Determination*, arXiv:1309.7085 [INSPIRE].

[23] K. Kovarik et al., *CTEQ nuclear parton distribution functions*, PoS(DIS 2013)274 [arXiv:1307.3454] [INSPIRE].

[24] M. Hirai, S. Kumano and T.-H. Nagai, *Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order*, *Phys. Rev.* **C 76** (2007) 065207 [arXiv:0709.3038] [INSPIRE].

[25] K.J. Eskola, H. Paukkunen and C.A. Salgado, *EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions*, *JHEP* **04** (2009) 065 [arXiv:0902.4154] [INSPIRE].

[26] D. de Florian, R. Sassot, P. Zurita and M. Stratmann, *Global Analysis of Nuclear Parton Distributions*, *Phys. Rev.* **D 85** (2012) 074028 [arXiv:1112.6324] [INSPIRE].

[27] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, arXiv:1101.0538 [INSPIRE].

[28] S. Alekhin et al., *The PDF4LHC Working Group Interim Report*, arXiv:1101.0536 [INSPIRE].

[29] S. Forte and G. Watt, *Progress in the Determination of the Partonic Structure of the Proton*, *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 291 [arXiv:1301.6754] [INSPIRE].

[30] S. Forte, *Parton distributions at the dawn of the LHC*, *Acta Phys. Polon.* **B 41** (2010) 2859 [arXiv:1011.5247] [INSPIRE].

[31] PARTICLE DATA GROUP collaboration, J. Beringer et al., *Review of Particle Physics (RPP)*, *Phys. Rev.* **D 86** (2012) 010001 [INSPIRE].

[32] P.M. Nadolsky and Z. Sullivan, *PDF uncertainties in WH production at Tevatron*, *eConf* **C 010630** (2001) P510 [hep-ph/0110378] [INSPIRE].

[33] P.M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, *Phys. Rev.* **D 78** (2008) 013004 [arXiv:0802.0007] [INSPIRE].

[34] J. Pumplin, *Parametrization dependence and $\Delta\chi^2$ in parton distribution fitting*, *Phys. Rev.* **D 82** (2010) 114020 [arXiv:0909.5176] [INSPIRE].

[35] A. Glazov, S. Moch and V. Radescu, *Parton Distribution Uncertainties using Smoothness Prior*, *Phys. Lett.* **B 695** (2011) 238 [arXiv:1009.6170] [INSPIRE].

[36] A.D. Martin et al., *Extended Parameterisations for MSTW PDFs and their effect on Lepton Charge Asymmetry from W Decays*, *Eur. Phys. J.* **C 73** (2013) 2318 [arXiv:1211.1215] [INSPIRE].

[37] http://metapdf.hepforge.org/.

[38] S. Moch, J.A.M. Vermaseren and A. Vogt, *The Three loop splitting functions in QCD: The Nonsinglet case*, *Nucl. Phys.* **B 688** (2004) 101 [hep-ph/0403192] [INSPIRE].

[39] A. Vogt, S. Moch and J.A.M. Vermaseren, *The Three-loop splitting functions in QCD: The Singlet case*, *Nucl. Phys.* **B 691** (2004) 129 [hep-ph/0404111] [INSPIRE].

[40] http://lhapdf.hepforge.org/.

[41] G.P. Salam and J. Rojo, *A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)*, *Comput. Phys. Commun.* **180** (2009) 120 [arXiv:0804.3755] [INSPIRE].

[42] C. Anastasiou, L.J. Dixon, K. Melnikov and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, *Phys. Rev.* **D 69** (2004) 094008 [hep-ph/0312266] [INSPIRE].

[43] C. Anastasiou, S. Buehler, F. Herzog and A. Lazopoulos, *Total cross-section for Higgs boson hadroproduction with anomalous Standard Model interactions*, *JHEP* **12** (2011) 058 [arXiv:1107.0683] [INSPIRE].

[44] P. Bärnreuther, M. Czakon and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, *Phys. Rev. Lett.* **109** (2012) 132001 [arXiv:1204.5201] [INSPIRE].

[45] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, *JHEP* **12** (2012) 054 [arXiv:1207.0236] [INSPIRE].

[46] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, *JHEP* **01** (2013) 080 [arXiv:1210.6832] [INSPIRE].

[47] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, arXiv:1112.5675 [INSPIRE].

[48] M. Beneke, P. Falgari and C. Schwinn, *Soft radiation in heavy-particle pair production: all-order colour structure and two-loop anomalous dimension*, *Nucl. Phys.* **B 828** (2010) 69 [arXiv:0907.1443] [INSPIRE].

[49] M. Czakon, A. Mitov and G.F. Sterman, *Threshold Resummation for Top-Pair Hadroproduction to Next-to-Next-to-Leading Log*, *Phys. Rev.* **D 80** (2009) 074017 [arXiv:0907.1790] [INSPIRE].

[50] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, *Phys. Lett.* **B 710** (2012) 612 [arXiv:1111.5869] [INSPIRE].

[51] FASTNLO collaboration, M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz and F. Stober, *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*, arXiv:1109.1310 [INSPIRE].

[52] ATLAS collaboration, *Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *Phys. Rev.* **D 86** (2012) 014022 [arXiv:1112.6297] [INSPIRE].

[53] G. Watt and R.S. Thorne, *Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs*, *JHEP* **08** (2012) 052 [arXiv:1205.4024] [INSPIRE].

[54] J. Pumplin et al., *Collider Inclusive Jet Data and the Gluon Distribution*, *Phys. Rev.* **D 80** (2009) 014019 [arXiv:0904.2424] [INSPIRE].

[55] J. Gao, J. Huston and P.M. Nadolsky, in preparation.

[56] H.-L. Lai et al., *Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions*, *Phys. Rev.* **D 82** (2010) 054021 [arXiv:1004.4624] [INSPIRE].

[57] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Uncertainties on α_s in global PDF analyses and implications for predicted hadronic cross sections*, *Eur. Phys. J.* **C 64** (2009) 653 [arXiv:0905.3531] [INSPIRE].

[58] R.D. Ball et al., *Precision NNLO determination of $\alpha_s(M_Z)$ using an unbiased global parton set*, *Phys. Lett.* **B 707** (2012) 66 [arXiv:1110.2483] [INSPIRE].

[59] ATLAS collaboration, *Measurement of the inclusive $W^{\pm}$ and Z/γ^* cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev.* **D 85** (2012) 072004 [arXiv:1109.5141] [INSPIRE].

[60] LHCb collaboration, *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, *JHEP* **06** (2012) 058 [arXiv:1204.1620] [INSPIRE].

[61] CMS collaboration, *Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **109** (2012) 111806 [arXiv:1206.2598] [INSPIRE].

[62] J.M. Campbell, J.W. Huston and W.J. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, *Rept. Prog. Phys.* **70** (2007) 89 [hep-ph/0611148] [INSPIRE].

[63] M. Sutton et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, PoS(DIS 2010)051.

[64] T. Kluge, K. Rabbertz and M. Wobisch, *FastNLO: Fast pQCD calculations for PDF fits*, hep-ph/0609285 [INSPIRE].

[65] NNPDF collaboration, R.D. Ball et al., *Reweighting NNPDFs: the W lepton asymmetry*, *Nucl. Phys.* **B 849** (2011) 112 [*Erratum ibid.* **B 854** (2012) 926] [arXiv:1012.0836] [INSPIRE].

[66] N. Sato, J.F. Owens and H. Prosper, *Bayesian Reweighting for Global Fits*, arXiv:1310.1089 [INSPIRE].