Published for SISSA by Springer

Received: *July 20, 2024*
Accepted: *August 27, 2024*
Published: *September 17, 2024*

LO, NLO, and NNLO parton distributions for LHC event generators

Juan Cruz-Martinez,[a] **Stefano Forte**,[b] **Niccolò Laurenti**,[b] **Tanjona R. Rabemananjara**[c,d] **and Juan Rojo**[c,d]

[a]*Theoretical Physics Department, CERN, Geneva 23 CH-1211, Switzerland*
[b]*Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, Milano I-20133, Italy*
[c]*Department of Physics and Astronomy, Vrije Universiteit, Amsterdam NL-1081 HV, The Netherlands*
[d]*Nikhef Theory Group, Science Park 105, Amsterdam 1098 XG, The Netherlands*

E-mail: juan.manuel.cruz.martinez@cern.ch, forte@mi.infn.it, niccolo.laurenti@unimi.it, rrabeman@nikhef.nl, j.rojo@vu.nl

Abstract: We present NNPDF4.0MC, a variant of the NNPDF4.0 set of parton distributions (PDFs) at LO, NLO and NNLO, with and without inclusion of the photon PDF, suitable for use with Monte Carlo (MC) event generators, which require PDFs to satisfy additional constraints in comparison to standard PDF sets. These requirements include PDF positivity down to a low scale $Q \sim 1\,\text{GeV}$, smooth extrapolation in the very small and large x regions, and numerically stable results even in extreme regions of phase space for all PDFs. We compare the NNPDF4.0MC PDFs to their baseline NNPDF4.0 counterparts, and to the NNPDF2.3LO set entering the Monash tune of the Pythia8 event generator. We briefly assess the phenomenological impact of these PDFs on the cross-sections for hard and soft QCD processes at the LHC.

Keywords: Parton Distributions, Parton Shower, Specific QCD Phenomenology

ArXiv ePrint: 2406.12961

https://doi.org/10.1007/JHEP09(2024)088

Contents

1 Introduction

Monte Carlo (MC) event generators [1–4] provide a complete description of the final state in high-energy particle collisions, and, as such, are an essential ingredient in the interpretation of particle physics experiments. Widely used event generators for LHC physics include Pythia8 [5, 6], Herwig7 [7, 8], Sherpa [9, 10], Powheg [11], MG5_aMC@NLO [12], and more recently PanScales [13–16].

Within a MC event generator, parton distributions (PDFs) [17, 18] are used not only in the evaluation of hadronic cross-section through their convolution with partonic matrix elements, but also for the initial-state backwards parton shower, and as inputs to the modeling of non-perturbative phenomena [19] such as the underlying event (UE), multiple parton interactions (MPI), and related soft QCD processes. For these latter aspects, PDFs should respect some additional constraints in comparison to default PDFs. First, their usage in initial-state showers requires that they be non-negative down to the perturbative cutoff of $Q \simeq 1\,\text{GeV}$. Furthermore, their application to models of the UE, MPI, and other low-energy QCD phenomena demands a very smooth extrapolation down to very small x and very small Q^2 values, and a gluon PDF that grows sufficiently fast in the small x region. In order to prevent numerical problems associated to Monte Carlo integration and sampling, PDFs should be numerically stable even in extreme regions of phase space which may be irrelevant for phenomenology. Finally, in order to match to standard parton showers, the charm PDF must be generated perturbatively (i.e. an intrinsic component is not allowed), and in order to account for electroweak corrections, the possibility of including a photon PDF $\gamma(x, Q^2)$ and QED splittings in perturbative evolution should be allowed.

Several groups [20–24] have presented variants of their LO PDF sets, aimed to usage in MC event generators. For instance, the NNPDF2.3QED LO PDFs developed in [25–27] were integrated in Pythia8, and used as one of the inputs for its popular Monash tune [28] of non-perturbative QCD physics. Beyond LO, BFKL-resummed variants of the NNPDF3.1 PDF set including the constraints on the small-x gluon from D-meson production at LHCb presented in [29–31] also satisfy the above requirements, and are available in Pythia8 as a stand-alone PDF set.

JHEP09(2024)088

Here we present variants of NNPDF4.0 [32–35] at LO and, for the first time, NLO and NNLO, tailored to their usage in modern MC event generators. The main goal of these NNPDF4.0MC sets is to satisfy the requirements discussed above, while at the same time providing the best possible description of the NNPDF4.0 dataset, in particular at NLO and NNLO.

2 Methodology

Unless otherwise specified, we adopt the same experimental dataset, theory calculations, and methodology used in the construction of the recent MHOU, QED, and aN^{3}LO NNPDF4.0 PDF sets [34–36]. In particular, we exploit the new NNPDF theory pipeline [37] built upon the EKO [38] evolution code, YADISM DIS module [39], and PINEAPPL fast grid interface [40]. The same values of the input SM parameters are used, in particular $\alpha_s(m_Z) = 0.118$ for the LO, NLO, and NNLO fits. We only provide a central PDF, instead of a set of PDF replicas representing the PDF probability distribution, because in the presence of extra constraints uncertainties might become unreliable, and they are anyway not relevant for applications to MC event generators.

Positivity and perturbative charm. Positivity of MC PDFs is required both for their usage in the initial-state shower as well as for the modeling of soft QCD phenomena. At LO, PDFs can be identified with physical cross-sections and hence are positive-definite. This is not necessarily true at NLO and beyond, where PDFs become scheme dependent and may be negative in certain regions of the phase space. Whereas in the commonly used $\overline{\text{MS}}$ scheme PDFs are positive also at NLO and beyond, this only holds in the perturbative region, i.e. at high enough scale [41–43], and correspondingly PDF positivity may fail when extrapolating to low Q values.

In the baseline NNPDF4.0 analysis, PDF positivity is imposed at the initial parametrization scale ($Q_0 = 1.65\,\text{GeV}$) at LO and at a higher scale, $Q^2_{\text{pos}} = 5\,\text{GeV}^2$, at NLO and beyond, following the prescription of [41, 43]. In addition, positivity of a set of physical observables at Q^2_{pos} is also imposed. Therefore, within the NNPDF4.0 methodology, the NLO and NNLO PDFs may be negative at low values of Q^2 as long as, upon evolution, they become positive at $Q^2 \geq Q^2_{\text{pos}}$. Even though this may happen in regions of phase space for which there are no direct experimental constraints, or such that a fixed-order leading-twist approximation breaks down, positivity is nevertheless required by MC generators. Furthermore, in the default NNPDF4.0 sets the charm PDF is parametrized and determined from the data on the same footing as all other PDFs [44], with its behavior for $Q < Q_0$ determined by backwards QCD evolution together with the matching from the $n_{\text{f}} = 4$ to the $n_{\text{f}} = 3$ flavor scheme [45]. However, a variant of NNPDF4.0 in which charm vanishes in the $n_{\text{f}} = 3$ flavor scheme and is determined by perturbative matching conditions in the $n_{\text{f}} = 4$ scheme is also available; in this case PDFs are parametrized at $Q_0 = 1\,\text{GeV}$, hence below the matching scale, set at $\mu_{\text{c}} = m_{\text{c}} = 1.51\,\text{GeV}$.

We consequently start from this perturbative charm variant of NNPDF4.0, with perturbative matching conditions used to determine charm at the matching scale $\mu_{\text{c}} = m_{\text{c}}$. We then impose the positivity of $g(x, Q_0)$ and $\Sigma(x, Q_0)$ at $Q_0 = 1\,\text{GeV}$ by squaring the corresponding

JHEP09(2024)088

neural network outputs. This ensures positivity of the gluon and the quark singlet PDFs at $Q_0 = 1\,\text{GeV}$ and consequently also for $Q > Q_0$ thanks to their rise at small x induced by perturbative QCD evolution as the scale is increased. Positivity of individual quark and antiquark PDFs is imposed at $Q^2_{\text{pos}} = 5\,\text{GeV}^2$ as in the default. This is sufficient to guarantee positivity down to Q_0 both at large x, where perturbative evolution is moderate even at low scale, and also at small x, where nonsinglet PDFs vanish. This strategy leads to positive-definite PDFs in the full range of (x, Q^2) probed by MC generators at LO and NLO.

At NNLO, the perturbative matching conditions lead to a charm PDF that at $Q = m_c$ is negative at small $x \lesssim 10^{-2}$, though it is already positive at all x for $Q^2 \gtrsim 5\,\text{GeV}^2$. Hence at NNLO it is not possible to simultaneously satisfy at $\mu_c = m_c$ the requirements that charm be positive and determined by perturbative matching. As we will discuss in section 3, the low-scale positivity of the gluon at small x is disfavored by the data and consequently imposing it leads to some deterioration of the fit quality.

Extrapolation in x and Q^2. General-purpose MC event generators should provide reliable results for the broadest possible region of phase space. This requires input PDFs with a smooth behavior in a wide Q range, from $Q \simeq 1\,\text{GeV}$ (initial-state showers, non-perturbative QCD modeling) up to $Q \sim 100\,\text{TeV}$ (relevant for future particle colliders and for applications to astroparticle physics) and from $x \simeq 10^{-9}$ (forward particle production) all the way up to large-x values close to the elastic limit $x = 1$ (required for high-mass new physics searches). Since these regions extend beyond the coverage of available data, a robust extrapolation procedure is necessary.

While PDF extrapolation in Q^2 is fixed by perturbative QCD evolution, extrapolation in x depends on assumptions. In the NNPDF4.0 approach, extrapolation to the small x and large x regions is provided by the output of a preprocessed neural network, and thus controlled by the behavior of both the neural net and the preprocessing function. This extrapolation to low Q^2 and large x values might be affected by numerical instabilities, both native, and related to their storage as LHAPDF grids. Specifically, the low Q^2 behavior is controlled by evolution from higher scales, that may amplify small differences in the initial condition, due to the growing value of $\alpha_s(Q)$, while at large x PDFs become very small and thus particularly sensitive to numerical instabilities. These two issues are intertwined, since even small $\mathcal{O}\left(10^{-5}\right)$ numerical differences in the solution of evolution equations may be enough to distort the PDFs in the large x region where they are almost vanishing. While such instabilities are innocuous for phenomenological applications, they may lead to numerical issues when PDFs are used in MC generators.

In order to prevent these instabilities and ensure that the MC PDFs are everywhere smooth and well-behaved, the NNPDF4.0MC PDFs are delivered as an LHAPDF grid with a finer coverage in x for the region $x \in [0.7, 0.95]$. For $x \gtrsim 0.95$, PDFs essentially vanish and any residual oscillations can be safely set to zero. In addition, instabilities of the order of the accuracy of the LHAPDF interpolation are averaged out by means of a dedicated Gaussian filter. Possible issues related to backward evolution are prevented by parametrizing PDFs at $Q_0 = 1\,\text{GeV}$, so no backward evolution is needed. We thus deliver LHAPDF grids that provides an interpolated output for all $x \in \left[10^{-9}, 1\right]$ and $Q \in \left[1, 10^6\right]\,\text{GeV}$.

JHEP09(2024)088

ID	ref.	evolution (Q_0)	Positivity ($Q_{\rm pos}$)	$\alpha_s(m_Z)$	Charm
NNPDF23_lo_as_0130_qed	[27]	$\text{QCD}_{\text{LO}} \otimes \text{QED}_{\text{LO}}$ TRN (1.0 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	0.130	pert.
NNPDF40_lo_as_01180	[32]	QCD_{LO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ (1.65 GeV)	0.118	fitted
NNPDF40_lo_pch_as_01180	[32]	QCD_{LO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	0.118	pert.
NNPDF40MC_lo_as_01180	t.w.	QCD_{LO} TRN (1.0 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	0.118	pert.
NNPDF40MC_lo_as_01180_qed	t.w.	$\text{QCD}_{\text{LO}} \otimes \text{QED}_{\text{LO}}$ EXA (1.0 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	0.118	pert.
NNPDF40_nlo_as_01180	[32]	QCD_{NLO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	fitted
NNPDF40_nlo_pch_as_01180	[32]	QCD_{NLO} TRN (1 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	pert.
NNPDF40MC_nlo_as_01180	t.w.	QCD_{NLO} TRN (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	pert.
NNPDF40_nlo_as_01180_qed	[34]	$\text{QCD}_{\text{NLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	fitted
NNPDF40MC_nlo_as_01180_qed	t.w.	$\text{QCD}_{\text{NLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	pert.
NNPDF40_nnlo_as_01180	[32]	QCD_{NNLO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	fitted
NNPDF40_nnlo_pch_as_01180	[32]	QCD_{NNLO} TRN (1 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	pert.
NNPDF40MC_nnlo_as_01180	t.w.	QCD_{NNLO} TRN (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	pert.
NNPDF40_nnlo_as_01180_qed	[34]	$\text{QCD}_{\text{NNLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	fitted
NNPDF40MC_nnlo_as_01180_qed	t.w.	$\text{QCD}_{\text{NNLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	0.118	pert.

Table 1. The NNPDF4.0MC PDFs presented in this work (t.w.) and their baseline counterparts.

QED evolution and the photon PDF. As shown in [31, 34, 46–48] and related studies, the impact of inclusion of a photon PDF alongside quark and gluon PDFs is moderate, its main effect being a reduction of the gluon momentum fraction by up to around 0.5% in favor of the photon. Here we take the photon PDF $\gamma(x, Q^2)$ at $Q = 1$ GeV from the NNPDF4.0 QED NNLO PDF set [34], we include it as boundary condition to the QCD $\otimes$ QED evolution of the LO, NLO, and NNLO NNPDF4.0MC PDFs, and impose a momentum sum rule that now also includes a photon contribution. We adopt the so-called exact-iterated (EXA) solution of the QCD $\otimes$ QED evolution equations, as implemented in EKO [38], as in ref. [34] to which we refer for more details. For pure QCD evolution we use instead the truncated (TRN) solution as in ref. [32], so that in each case the PDF sets presented here are based on the same form of the solution of the evolution equations as their default counterparts.

NNPDF4.0MC overview. In table 1 we summarize the settings adopted for the NNPDF4.0MC PDFs, compared to those of their baseline counterparts: LHAPDF naming ID, publication reference, PDF parametrization scale and solution of the evolution equations, positivity scale, value of $\alpha_s(m_Z)$, and treatment of charm (data-driven, or determined from perturbative matching). In this table q_i, $\bar{q}_i$ denote light (up, down, and strange) quarks and antiquark PDFs, as, following [41, 43], positivity of the charm PDF is never imposed.

3 The NNPDF4.0MC PDFs

We now compare the NNPDF4.0MC PDF sets to the baseline NNPDF4.0 fits and to the NNPDF2.3QED LO PDFs used for the MONASH tune [28] of PYTHIA8. Here we only present

JHEP09(2024)088

Dataset by process group	NLO					NNLO				
	$n_{\rm dat}$	QCD		QCD+QED		$n_{\rm dat}$	QCD		QCD+QED	
		BL	MC	BL	MC		BL	MC	BL	MC
DIS NC	1953	1.35	1.37	1.38	1.54	2110	1.22	1.30	1.22	1.29
DIS CC	988	0.91	0.92	0.94	0.95	989	0.90	0.89	0.90	0.89
DY NC	669	1.58	1.84	1.67	2.04	736	1.20	1.30	1.22	1.33
DY CC	197	1.38	1.56	1.40	1.61	157	1.45	1.55	1.47	1.57
Top pairs	66	2.40	2.14	2.51	2.47	64	1.27	1.16	1.31	1.27
Single-inclusive jets	356	0.82	0.88	0.83	0.93	356	0.94	1.01	0.93	1.00
Dijets	144	1.51	1.55	1.56	1.62	144	2.01	2.01	1.94	1.93
Photon	53	0.57	0.60	0.64	0.74	53	0.76	0.67	0.74	0.68
Single top	17	0.36	0.36	0.38	0.36	17	0.37	0.38	0.39	0.40
Total	4443	1.28	1.30	1.30	1.44	4626	1.16	1.22	1.17	1.22

Table 2. The number of data points and the χ^2 per data point for the NLO and NNLO baseline NNPDF4.0 fits (BL), compared to their NNPDF4.0MC counterparts (MC), with the same process categorisation as in ref. [36]. The χ^2 values are provided for the QCD-only (`NNPDF40(MC)_<order>_as_01180`) and for the QCD$\otimes$QED (`NNPDF40(MC)_<order>_as_01180_qed`) fits of table 1.

some representative results; an extensive set of comparisons is available online.[1] In all comparisons below, unless otherwise stated, NNPDF4.0 refers to the default sets, and indeed the purpose of the comparison is to illustrate the difference in phenomenology to be expected if the MC sets instead of the default are used, for instance in applications to experimental analysis. In particular, a comparison to the perturbative charm variants of NNPDF4.0 listed in table 1 will only be shown in figures 3–4, for the sake of assessing the impact of this particular assumption among the others that characterize the NNPDF4.0MC sets.

The fit quality for the NLO and NNLO PDF sets of table 1 is summarized in table 2, where we show the number of data points and the χ^2 per data point; LO χ^2 values are not shown since fit quality at LO is generally poor and the specific value of the χ^2 is not significant. When comparing fit quality, the MC PDFs constructed here should be viewed as PDFs that include some additional theory assumptions: for instance, the positive small x behavior of the gluon at low scale can be justified based on non-perturbative physics arguments (see e.g. [49]). Because extra constraints are introduced, the agreement with the data of the MC PDFs will be either unchanged, or possibly worse than that of the default, i.e. the fit quality will deteriorate (or remain unchanged). The purpose of the comparison is then to check that the deterioration in fit quality is not such as to rule out these extra assumptions.

For pure QCD PDFs, we find that at NLO (NNLO) the total χ^2 per data point of the baseline fit increases from 1.28 (1.16) to 1.30 (1.22), an effect of about 1σ (3σ) in units of the statistical variance of the χ^2 distribution for $n_{\rm dat} = 4443$ (4626) data points. Therefore, imposing the MC PDF conditions at NLO cannot be distinguished from a change in χ^2 value due to a random fluctuation of the data. At NNLO the MC conditions do lead to a mild

[1] https://data.nnpdf.science/vp-public/NNPDF40MC_comparisons/.

JHEP09(2024)088

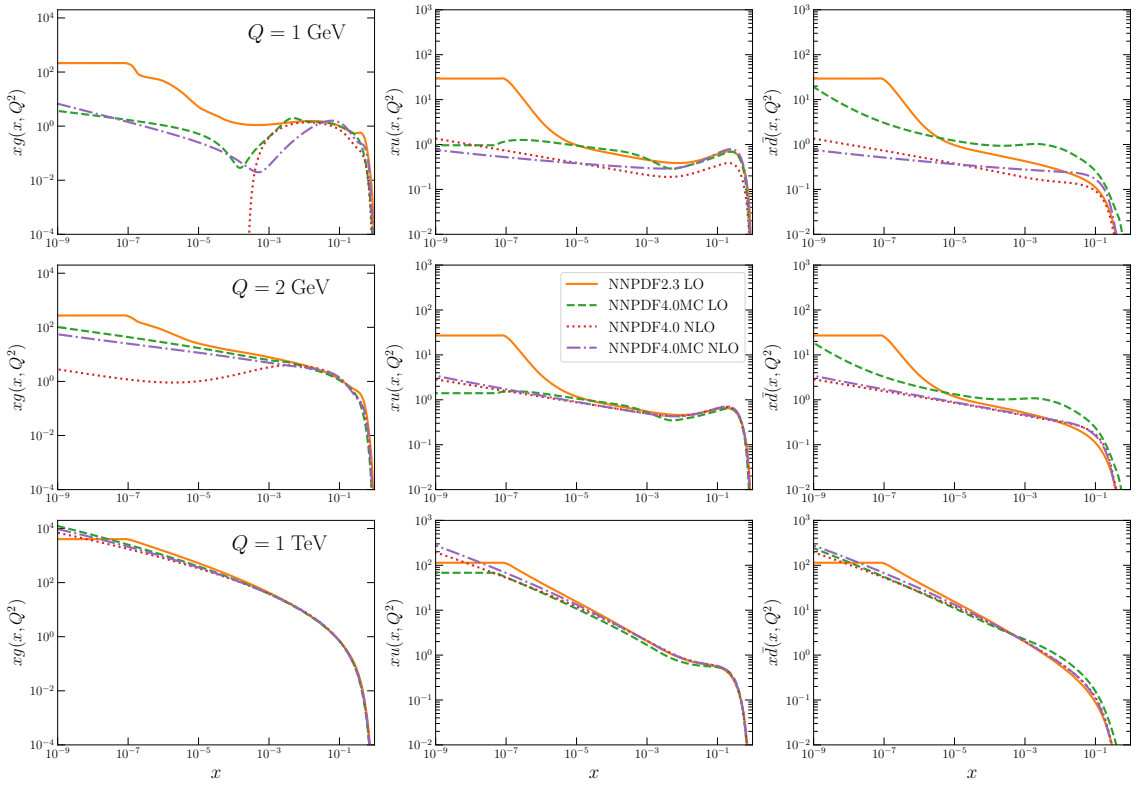

Figure 1. The NNPDF4.0MC LO and NLO gluon, up, and antidown PDFs (from left to right) compared to NNPDF2.3LO and NNPDF4.0 NLO, at three scales: $Q = 1\,\text{GeV}$, $2\,\text{GeV}$, and $1\,\text{TeV}$ (from top to bottom). Only central values are shown, in the region for which PDFs are provided via LHAPDF.

deterioration of fit quality, related to the fact that the rapid rise of the gluon at small x as the scale increases tends to lead in turn to a negative gluon at scales $Q^2 \lesssim$ few GeV^2 [43]. This rise is stronger at NNLO, and at low scale NNLO corrections become large; consequently at NNLO a low-scale positive gluon is more difficult to accommodate, though again it cannot be excluded. For the QCD $\otimes$ QED sets, the same behavior is observed at NNLO, while now at NLO a more significant deterioration of fit quality is seen. This can be traced to the fact that subleading terms included in the EXA solution of the evolution equations lead to perturbative evolution that is faster than for the TRN solution, especially when the anomalous dimension is large, which then makes the problem with low-scale gluon positivity more serious at NLO. The difference between the pure QCD and QCD $\otimes$ QED cases at NLO should thus be viewed as driven by missing NNLO QCD corrections.

The MC and baseline LO and NLO PDFs are compared in figure 1, where we display the gluon, up and antidown PDFs at $Q = 1\,\text{GeV}$, $2\,\text{GeV}$, and $1\,\text{TeV}$. Recall that the small-x behavior of all quark and antiquark PDFs is the same, and dominated by that of the singlet quark distribution. We show the full x region in which the NNPDF4.0MC PDFs are provided via the LHAPDF interpolation, i.e. $10^{-9} \leq x \leq 1$. Note that the NNPDF2.3LO set was only provided for $x \geq 10^{-7}$, while for smaller x values PDFs are frozen to their value at $x = 10^{-7}$. Apart from this trivial difference, the main difference between the 2.3 and 4.0 LO sets is that for NNPDF4.0MC the rise of the small-x gluon is qualitatively similar at

JHEP09(2024)088

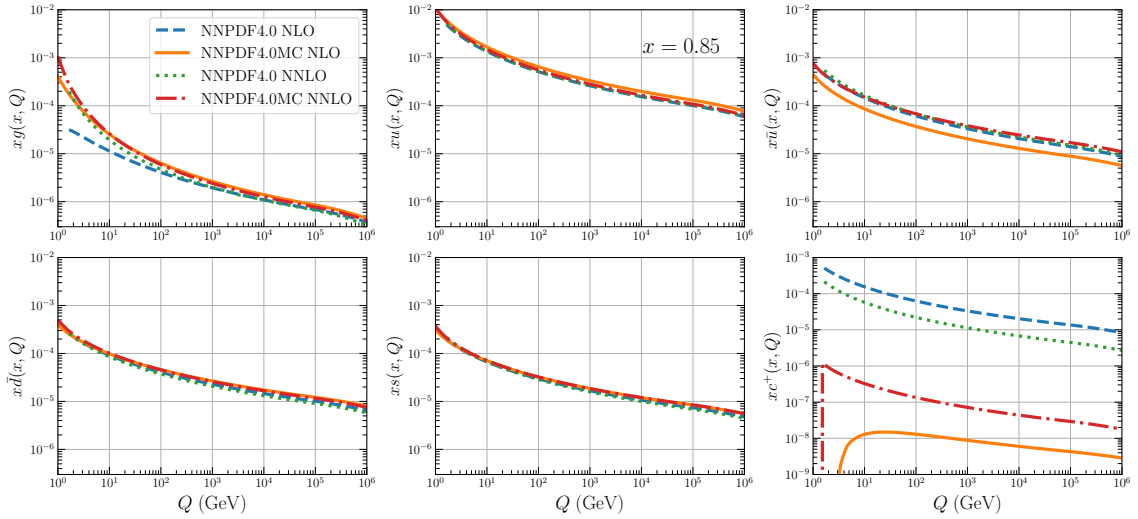

Figure 2. The NNPDF4.0MC NLO and NNLO gluon, up, antiup, antidown, strange and total charm PDFs (from left to right and from top to bottom), compared to their baseline counterparts as a function of scale for a fixed large $x = 0.85$ value. The range $1 \leq Q \leq 10^6$ GeV shown corresponds to the full interpolation range provided by the LHAPDF grids that we deliver.

LO and NLO, a feature facilitating the tuning of soft QCD models in MC event generators. This is due to the greater theoretical consistency of assumptions between LO and NLO in the NNPDF4.0MC sets, specifically the choice of the same value of α_s. The main difference between the MC and default NLO PDFs is related to the small x positivity of the gluon at low scale. As the scale Q is increased, relative differences between the various PDF sets are washed out by perturbative evolution.

In order to demonstrate smoothness of the NNPDF4.0MC sets in the large-x extrapolation region, we display in figure 2 the NLO and NNLO NNPDF4.0MC PDFs for $x = 0.85$ as a function of scale, compared to the central value of their baseline counterparts. The Q range shown corresponds to the full interpolation range in the LHAPDF grids that we provide. All PDFs displayed exhibit a satisfactory level of smoothness.

In order to fully assess the difference between the MC sets and their baseline counterparts, in figure 3 we display the ratio of the NNPDF4.0MC NLO PDFs to the baseline, also showing the 68% CL PDF uncertainties on the latter. In order to trace the origin of differences, the NNPDF4.0 NLO set with perturbative charm of table 1 is also shown. In the region $x \gtrsim 10^{-3}$, where the bulk of experimental data is located, the quark MC PDFs are mostly contained within the uncertainty band of the baseline. Larger differences, that can be traced to the requirement of low-scale positivity, are observed for the gluon PDF, especially at small $x \lesssim 10^{-2}$. These in turn propagate onto the other PDFs at small x, all of which display a stronger small-x rise in comparison to the baseline in the extrapolation region $x \lesssim 10^{-3}$.

The results of figure 3 imply that the additional model assumptions entering the MC PDFs do not distort the baseline PDFs in the bulk of the data region beyond the 1σ level, indicating that most LHC cross-sections obtained with the NNPDF4.0MC sets will be consistent with those derived using the baseline PDFs. In fact, it is clear from figure 3 that for most PDFs,

JHEP09(2024)088

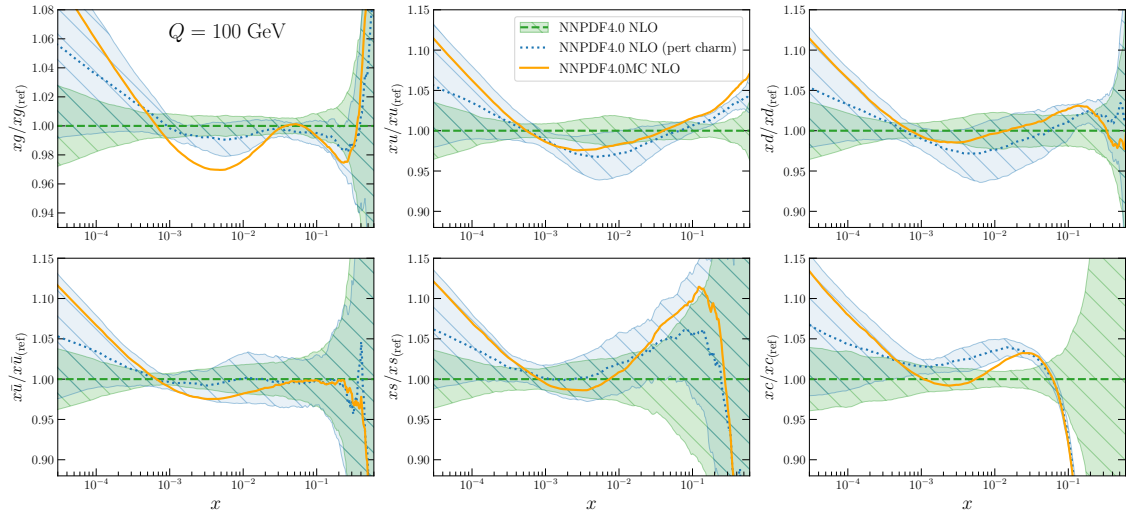

Figure 3. The NLO NNPDF4.0MC gluon, up, down, antiup, strange and charm PDFs at $Q = 100\,\mathrm{GeV}$ (from left to right and from top to bottom), shown as a ratio to their baseline counterpart. The uncertainty shown is the 68% CL on the baseline. The baseline variant with perturbative charm is also shown.

especially for the sea quark PDFs, a large part of the difference between the MC PDFs and the default is due to having adopted perturbative charm.

4 Impact on LHC physics

We now carry out a brief assessment of the phenomenological impact of similarities and differences between the MC PDFs and their baseline counterparts shown in figures 1–3.

First, in figure 4 we display the gluon-gluon, quark-antiquark, and quark-quark parton luminosities at the LHC with $\sqrt{s} = 13.6\,\mathrm{TeV}$ as a function of the invariant mass of the final state m_X, computed from the same PDFs shown in figure 3, and shown as a ratio to the NNPDF4.0 baseline. The luminosities are integrated over the full rapidity range and are thus dominated by the PDF behavior in the central rapidity region, where $x_1 \sim x_2 \sim m_X/\sqrt{s}$. For $50\,\mathrm{GeV} \lesssim m_X \lesssim 1\,\mathrm{TeV}$ this is a medium-small x region, where differences between the MC PDFs and the baseline are generally moderate and only noticeable for the gluon. Indeed, in the case of the gluon-gluon luminosity MC PDFs lead to a suppression of around 2% in comparison to the baseline for $100\,\mathrm{GeV} \lesssim m_X \lesssim 3\,\mathrm{TeV}$, while otherwise differences between NNPDF4.0 NLO and its MC variant are at the 1% level, and only become larger, though well within uncertainties, for $m_X \lesssim 100\,\mathrm{GeV}$ due to stronger small x rise of the MC PDFs.

We then consider representative inclusive hard cross-sections: Higgs and gauge boson production at the LHC with $\sqrt{s} = 13.6\,\mathrm{TeV}$, computed using the ggHiggs [50], n3loxs [51] and proVBFH [52, 53] codes. In figure 5 we compare results obtained at NLO and NNLO (both for PDF and the matrix element) with the MC sets and their baseline counterparts, and for the latter also aN^{3}LO, using the settings of ref. [35]. The uncertainty shown is for the MC sets only that related to missing higher orders in the matrix element, evaluated from standard 7-point scale variation, while for the baseline sets it also includes the PDF

JHEP09(2024)088

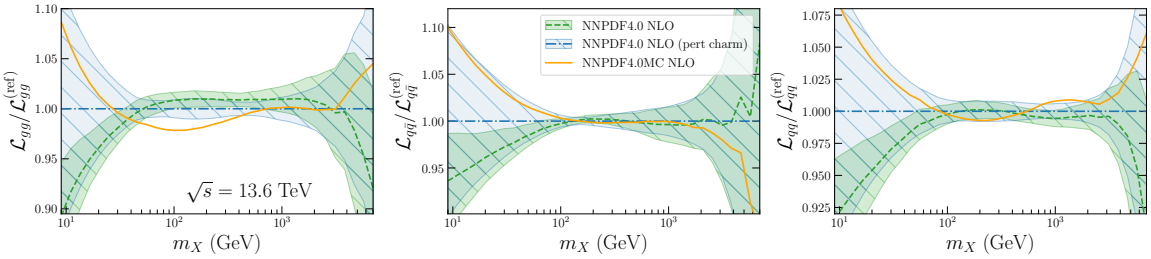

Figure 4. The gluon-gluon, quark-antiquark, and quark-quark parton luminosities at the LHC with $\sqrt{s} = 13.6\,\text{TeV}$ as a function of the invariant mass m_X for the same PDFs as in figure 3, shown as a ratio to the NNPDF4.0 baseline.

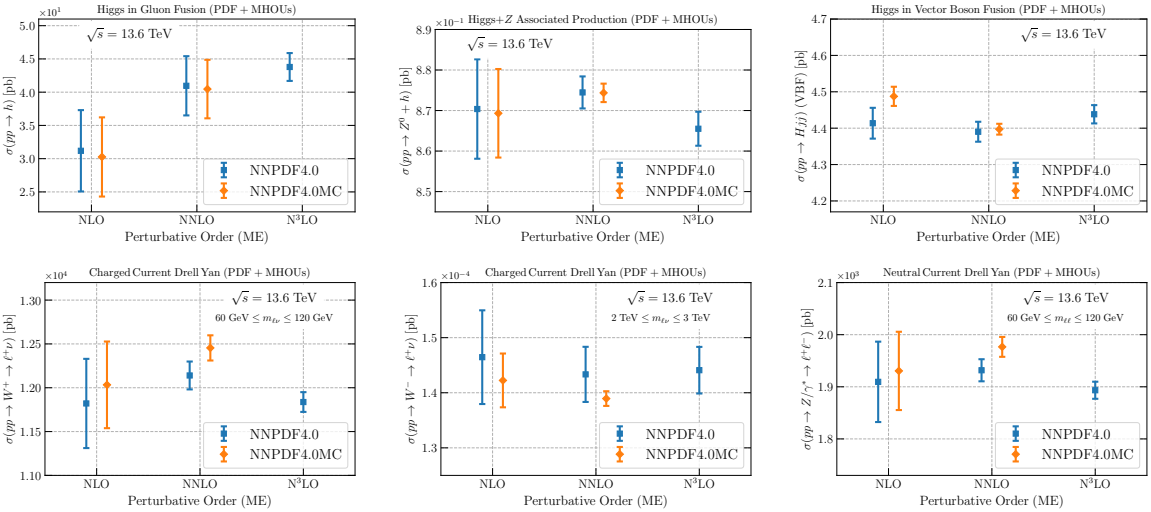

Figure 5. The inclusive NLO and NNLO cross-sections for Higgs production in gluon fusion, in association with a Z boson, and in vector boson fusion (top), and on-shell and high-mass W and on-shell Z production at the LHC $\sqrt{s} = 13.6\,\text{TeV}$ (bottom), comparing NNPDF4.0MC PDFs and the baseline. For the baseline NNPDF4.0, the aN^{3}LO result is also shown. The uncertainty shown is scale variation with 7-point prescription only for the MC PDFs, combined in quadrature with the PDF uncertainty for the baseline sets.

uncertainty, combined in quadrature with it. The corresponding uncertainty bands always overlap, reflecting the differences seen in parton luminosities.

We turn next to processes that are also sensitive to soft physics. We show results for LHC differential distributions at leading order obtained from Pythia8 simulations interfaced to the Rivet analysis toolkit [54]. We neglect PDF uncertainties and only display the central values found using NNPDF2.3LO, NNPDF4.0 NLO, and NNPDF4.0MC NLO PDFs. We first consider the normalized Z boson transverse momentum distribution, reconstructed from bare dilepton events, either electrons or muons, which is sensitive to both soft and hard QCD. In figure 6 the Pythia8 LO predictions for $1\,\text{GeV} \leq p_\perp(\ell\ell) \leq 300\,\text{GeV}$ are compared to ATLAS data at 7 TeV from ref. [55]. The low and high $p_\perp$ regions respectively probe soft and hard QCD radiation. For the normalized distributions shown, higher-order QCD corrections partially cancel out. The difference between PDF sets is negligible, and good agreement with the data is found using all PDF sets except at very small $p_\perp$ in the electron channel.

JHEP09(2024)088

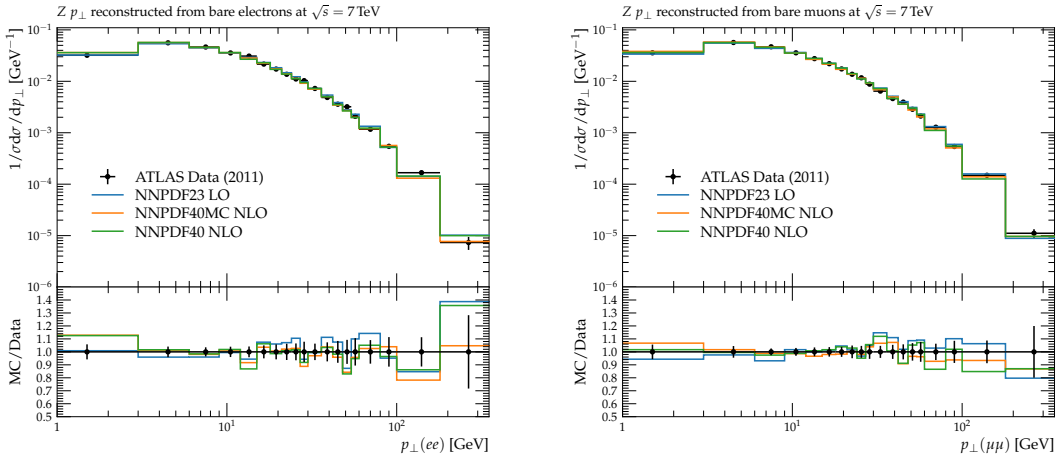

Figure 6. The normalized Z boson p_T distribution computed at LO using PYTHIA8 and RIVET using NNPDF2.3LO, NNPDF4.0 NLO, and NNPDF4.0MC NLO PDFs. Predictions are compared to the ATLAS [55] data at $\sqrt{s} = 7\,\text{TeV}$ using bare electron (left) or muon (right) pairs; error bars on the data include statistical and systematic uncertainties. Both the absolute distribution (top) and the ratio of the theory prediction to the data (bottom) are shown.

We next consider the fiducial cross-sections for Higgs production in the $H \to ZZ^\star \to 4\ell$ ($\ell = e, \mu$) decay channel. In figure 7 we compare predictions to the ATLAS data collected at $\sqrt{s} = 13\,\text{TeV}$ with an integrated luminosity of $\mathcal{L} = 139\,\text{fb}^{-1}$ [56]. Results are shown for the transverse momentum distribution of the four hardest leptons in the event, $p_T^{4\ell}$, in the rapidity range $1.0 < |y_{4\ell}| < 1.5$, and for the transverse momentum of the leading jet in the invariant mass range $115 < m_{4\ell} < 130\,\text{GeV}$. Also in this case, differences between different PDF sets are negligible, and good agreement with the data is found.

We then turn to the energy flow, defined as

$$\frac{dE}{d\eta} = \frac{1}{|\eta_{\text{max}} - \eta_{\text{min}}|} \left(\frac{1}{N_{\text{inel}}} \sum_{i=1}^{n_{\text{part}}} E_i \theta(\eta_i > \eta_{\text{min}}) \theta(\eta_i < \eta_{\text{max}}) \right) , \tag{4.1}$$

where η is the midpoint of the rapidity interval, $[\eta_{\text{min}}, \eta_{\text{max}}]$, N_{inel} is the number of inelastic pp collisions, and n_{part} is the number of stable particles in the event whose energy is equal to E_i. The energy flow in dijet events and in minimum-bias events at $\sqrt{s} = 7\,\text{TeV}$ in the forward $3.2 \leq \eta \leq 4.9$ ranges is shown in figure 8, compared to the CMS data of [57]. For the dijet sample, a $p_\perp^{\text{jet}} > 20\,\text{GeV}$ cut is imposed. For both dijet and minimum-bias events, the simulations based on NNPDF2.3LO display good agreement with the data, while those obtained using NNPDF4.0 NLO sets (both MC and baseline) tend to undershoot the experimental measurements, which suggests the need for a dedicated tune of soft QCD physics.

Finally, in figure 9 we show the charged-hadron multiplicity distribution, differential in pseudorapidity and in transverse momentum, $d^2N_{\text{chapter}}/d\eta dp_\perp$, as a function of $p_\perp$ at fixed rapidity $|\eta| = 0.3$ and as a function of η integrated over the full $p_\perp$ range. Predictions are compared to the CMS measurements of [58], for events that satisfy both $p_\perp \leq 2\,\text{GeV}$ and $|\eta| < 2.5$ in order to highlight the sensitivity to the modeling of nonperturbative QCD dynamics. As in the case of the energy flow, the NNPDF2.3LO set provides the best description of the

JHEP09(2024)088

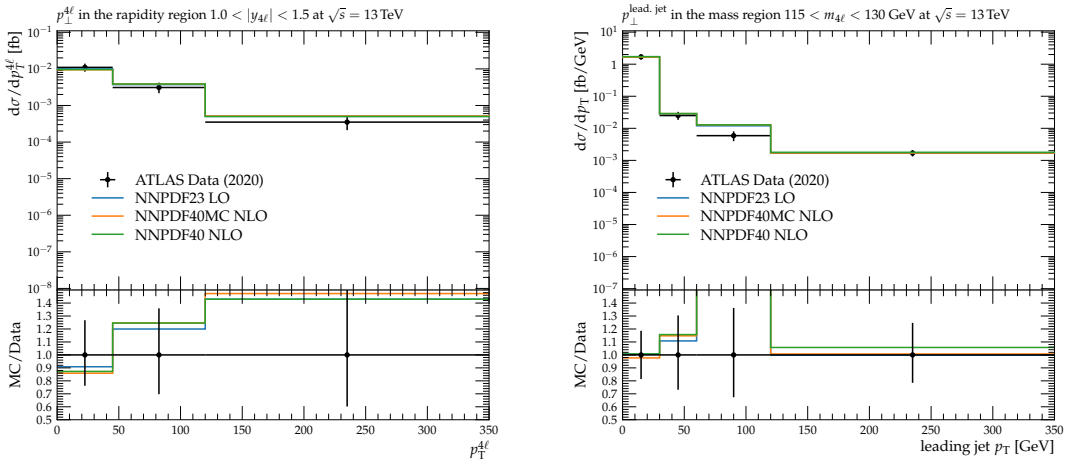

Figure 7. Same as figure 6 for the fiducial cross-section for Higgs production at $\sqrt{s} = 13\,\text{TeV}$ in the $H \to ZZ^{\star} \to 4\ell$ ($\ell = e, \mu$) decay channel. The four-lepton p_T distribution for $1.0 < |y_{4\ell}| < 1.5$ (left) and the p_T of the leading jet in events with ≥ 1 jet for $115 < m_{4\ell} < 130\,\text{GeV}$ (right) are shown, compared to ATLAS data [56].

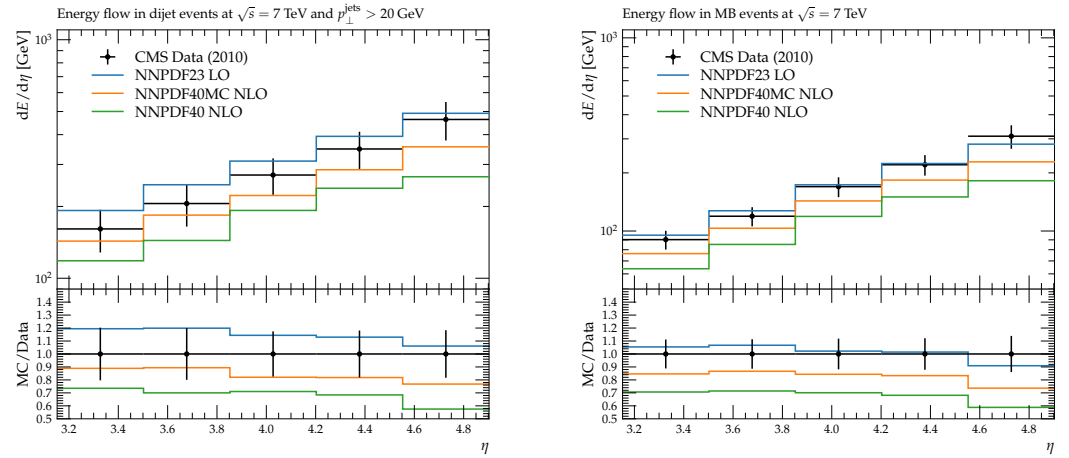

Figure 8. Same as figure 6, now for the energy flow in dijet (left) and minimum-bias events with $\sqrt{s} = 7\,\text{TeV}$ and $3.2 \leq \eta \leq 4.9$ compared to CMS data [57].

experimental data, while the NNPDF4.0 sets undershoot the CMS measurements. Indeed, both the energy flow of figure 8 and the charged-hadron differential distributions of figure 9 are sensitive to non-perturbative QCD processes. It follows that achieving a good description requires a dedicated tune of soft QCD, and differences seen in figures 8–9 do not have a simple physical interpretation, and are simply a manifestation of the fact that the NNPDF4.0 sets have not been used in the Monte Carlo tune. The Monash 2013 tune of Pythia8 used here is based on NNPDF2.3LO, explaining the good agreement found for this set.

5 Summary and outlook

The NNPDF4.0MC PDFs presented in this work satisfy the requirements of event generators not only at LO but also at NLO and NNLO accuracy, while the NLO and NNLO sets provide

JHEP09(2024)088

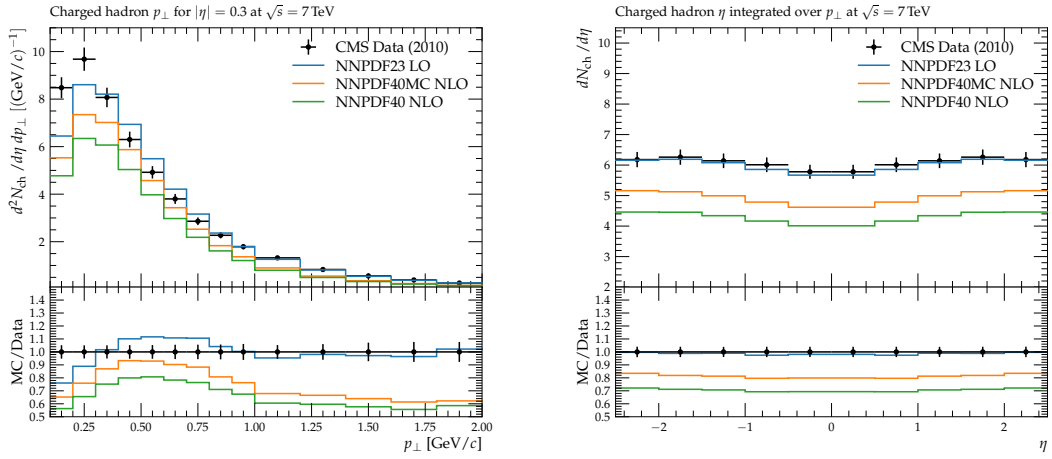

Figure 9. Same as figure 8, now for the charged-hadron transverse momentum (left) and pseudorapidity (right) distributions in proton-proton collisions at $\sqrt{s} = 7\,\mathrm{TeV}$, comparing to the CMS measurements of [58].

a satisfactory description of the global dataset and minimize differences in comparison to the baseline sets, ensuring their reliability to evaluate hard cross-sections at the LHC and elsewhere. It thus becomes possible to combine the precision and accuracy enjoyed by global PDF sets at NLO and NNLO without compromising the usability of these PDFs in generators for initial-state radiation and the modeling of soft QCD processes.

In order to also achieve agreement with the data for non-perturbative processes such as the underlying event, pileup, and low-p_T radiation, the soft QCD models specific to each event generator will need to be tuned to the data using as input these new NNPDF4.0MC PDFs, since their behavior, especially for low-x physics, becomes a component of the tuning model. Such dedicated tunes will be needed in order for the NNPDF4.0MC PDF sets to become instrumental in the development of a next generation of Monte Carlo codes that reaches higher perturbative accuracy. To this purpose, we aim to collaborate with event generator developers in order to integrate NNPDF4.0MC in their frameworks and produce dedicated tunes of soft QCD physics such that the whole palette of LHC processes, from the soft to the perturbative region, can be satisfactory described within a single physics simulation.

The NNPDF4.0 MC sets are made available through the LHAPDF interface [59] and the NNPDF collaboration website.[2]

Acknowledgments

We are very grateful to Gavin Salam and Peter Skands for many useful discussions, productive feedback, and extensive benchmarking concerning the MC PDF sets presented in this work. We also thank Melissa van Beekveld, Silvia Ferrario Ravasio, Christian Gutschow, Max Knobbe, Frank Krauss, and Oliver Mattelaer for assistance with the validation of the NNPDF4.0 MC PDFs. J.R. and T.R. are partially supported by NWO, the Dutch Research Council, and by The Netherlands eScience Center (NLeSC).

[2] https://nnpdf.mi.infn.it/nnpdf4-0-mc/.

JHEP09(2024)088

References

[1] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rep.* **504** (2011) 145 [arXiv:1101.2599] [INSPIRE].

[2] P. Nason and B. Webber, *Next-to-Leading-Order Event Generators*, *Annu. Rev. Nucl. Part. Sci.* **62** (2012) 187 [arXiv:1202.1251] [INSPIRE].

[3] S. Höche, *Introduction to parton-shower event generators*, in the proceedings of the *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, Boulder, CO, U.S.A., 2–27 June 2014, pp. 235–295 [DOI:10.1142/9789814678766_0005] [arXiv:1411.4085] [INSPIRE].

[4] J.M. Campbell et al., *Event generators for high-energy physics experiments*, *SciPost Phys.* **16** (2024) 130 [arXiv:2203.11110] [INSPIRE].

[5] T. Sjostrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[6] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codebases* **2022** (2022) 8 [arXiv:2203.11601] [INSPIRE].

[7] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196 [arXiv:1512.01178] [INSPIRE].

[8] J. Bellm et al., *Herwig 7.2 release note*, *Eur. Phys. J. C* **80** (2020) 452 [arXiv:1912.06509] [INSPIRE].

[9] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007 [arXiv:0811.4622] [INSPIRE].

[10] SHERPA collaboration, *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034 [arXiv:1905.09127] [INSPIRE].

[11] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [arXiv:1002.2581] [INSPIRE].

[12] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[13] M. van Beekveld et al., *Introduction to the PanScales framework, version 0.1*, arXiv:2312.13275 [DOI:10.21468/SciPostPhysCodeb.31] [INSPIRE].

[14] M. van Beekveld et al., *PanScales showers for hadron collisions: all-order validation*, *JHEP* **11** (2022) 020 [arXiv:2207.09467] [INSPIRE].

[15] M. van Beekveld, S. Ferrario Ravasio, G.P. Salam, A. Soto-Ontoso, G. Soyez and R. Verheyen, *PanScales parton showers for hadron collisions: formulation and fixed-order studies*, *JHEP* **11** (2022) 019 [arXiv:2205.02237] [INSPIRE].

[16] M. van Beekveld et al., *A new standard for the logarithmic accuracy of parton showers*, arXiv:2406.02661 [INSPIRE].

JHEP09(2024)088

[17] J. Gao, L. Harland-Lang and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, *Phys. Rep.* **742** (2018) 1 [arXiv:1709.04922] [INSPIRE].

[18] K. Kovařík, P.M. Nadolsky and D.E. Soper, *Hadronic structure in high-energy collisions*, *Rev. Mod. Phys.* **92** (2020) 045003 [arXiv:1905.06957] [INSPIRE].

[19] P.Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, *Phys. Rev. D* **82** (2010) 074018 [arXiv:1005.3457] [INSPIRE].

[20] H.-L. Lai et al., *Parton Distributions for Event Generators*, *JHEP* **04** (2010) 035 [arXiv:0910.4183] [INSPIRE].

[21] M. Yan, T.-J. Hou, P. Nadolsky and C.-P. Yuan, *CT18 global PDF fit at leading order in QCD*, *Phys. Rev. D* **107** (2023) 116001 [arXiv:2205.00137] [INSPIRE].

[22] NNPDF collaboration, *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, *Nucl. Phys. B* **855** (2012) 153 [arXiv:1107.2652] [INSPIRE].

[23] A. Sherstnev and R.S. Thorne, *Different PDF approximations useful for LO Monte Carlo generators*, in the proceedings of the *16th International Workshop on Deep Inelastic Scattering and Related Subjects*, London, U.K., 7–11 April 2008, p. 149 [DOI:10.3360/dis.2008.149] [arXiv:0807.2132] [INSPIRE].

[24] A. Sherstnev and R.S. Thorne, *Parton Distributions for LO Generators*, *Eur. Phys. J. C* **55** (2008) 553 [arXiv:0711.2473] [INSPIRE].

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

[26] NNPDF collaboration, *Parton distributions with QED corrections*, *Nucl. Phys. B* **877** (2013) 290 [arXiv:1308.0598] [INSPIRE].

[27] S. Carrazza, S. Forte and J. Rojo, *Parton Distributions and Event Generators*, in the proceedings of the *43rd International Symposium on Multiparticle Dynamics*, Chicago, IL, U.S.A., 15–20 September 2013, pp. 89–96 [arXiv:1311.5887] [INSPIRE].

[28] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur. Phys. J. C* **74** (2014) 3024 [arXiv:1404.5630] [INSPIRE].

[29] R.D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo and L. Rottoli, *Parton distributions with small-x resummation: evidence for BFKL dynamics in HERA data*, *Eur. Phys. J. C* **78** (2018) 321 [arXiv:1710.05935] [INSPIRE].

[30] V. Bertone, R. Gauld and J. Rojo, *Neutrino Telescopes as QCD Microscopes*, *JHEP* **01** (2019) 217 [arXiv:1808.02034] [INSPIRE].

[31] NNPDF collaboration, *Illuminating the photon content of the proton within a global PDF analysis*, *SciPost Phys.* **5** (2018) 008 [arXiv:1712.07053] [INSPIRE].

[32] NNPDF collaboration, *The path to proton structure at 1% accuracy*, *Eur. Phys. J. C* **82** (2022) 428 [arXiv:2109.02653] [INSPIRE].

[33] NNPDF collaboration, *An open-source machine learning framework for global analyses of parton distributions*, *Eur. Phys. J. C* **81** (2021) 958 [arXiv:2109.02671] [INSPIRE].

[34] NNPDF collaboration, *Photons in the proton: implications for the LHC*, *Eur. Phys. J. C* **84** (2024) 540 [arXiv:2401.08749] [INSPIRE].

[35] NNPDF collaboration, *The path to N^3LO parton distributions*, *Eur. Phys. J. C* **84** (2024) 659 [arXiv:2402.18635] [INSPIRE].

JHEP09(2024)088

[36] NNPDF collaboration, *Determination of the theory uncertainties from missing higher orders on NNLO parton distributions with percent accuracy*, *Eur. Phys. J. C* **84** (2024) 517 [arXiv:2401.10319] [INSPIRE].

[37] A. Barontini, A. Candido, J.M. Cruz-Martinez, F. Hekhorn and C. Schwan, *Pineline: Industrialization of high-energy theory predictions*, *Comput. Phys. Commun.* **297** (2024) 109061 [arXiv:2302.12124] [INSPIRE].

[38] A. Candido, F. Hekhorn and G. Magni, *EKO: evolution kernel operators*, *Eur. Phys. J. C* **82** (2022) 976 [arXiv:2202.02338] [INSPIRE].

[39] A. Candido, F. Hekhorn, G. Magni, T.R. Rabemananjara and R. Stegeman, *Yadism: yet another deep-inelastic scattering module*, *Eur. Phys. J. C* **84** (2024) 698 [arXiv:2401.15187] [INSPIRE].

[40] S. Carrazza, E.R. Nocera, C. Schwan and M. Zaro, *PineAPPL: combining EW and QCD corrections for fast evaluation of LHC processes*, *JHEP* **12** (2020) 108 [arXiv:2008.12789] [INSPIRE].

[41] A. Candido, S. Forte and F. Hekhorn, *Can $\overline{MS}$ parton distributions be negative?*, *JHEP* **11** (2020) 129 [arXiv:2006.07377] [INSPIRE].

[42] J. Collins, T.C. Rogers and N. Sato, *Positivity and renormalization of parton densities*, *Phys. Rev. D* **105** (2022) 076010 [arXiv:2111.01170] [INSPIRE].

[43] A. Candido, S. Forte, T. Giani and F. Hekhorn, *On the positivity of $\overline{MS}$ parton distributions*, *Eur. Phys. J. C* **84** (2024) 335 [arXiv:2308.00025] [INSPIRE].

[44] NNPDF collaboration, *A Determination of the Charm Content of the Proton*, *Eur. Phys. J. C* **76** (2016) 647 [arXiv:1605.06515] [INSPIRE].

[45] NNPDF collaboration, *Evidence for intrinsic charm quarks in the proton*, *Nature* **608** (2022) 483 [arXiv:2208.08372] [INSPIRE].

[46] A. Manohar, P. Nason, G.P. Salam and G. Zanderighi, *How bright is the proton? A precise determination of the photon parton distribution function*, *Phys. Rev. Lett.* **117** (2016) 242002 [arXiv:1607.04266] [INSPIRE].

[47] T. Cridge, L.A. Harland-Lang, A.D. Martin and R.S. Thorne, *QED parton distribution functions in the MSHT20 fit*, *Eur. Phys. J. C* **82** (2022) 90 [arXiv:2111.05357] [INSPIRE].

[48] CTEQ-TEA collaboration, *Photon PDF within the CT18 global analysis*, *Phys. Rev. D* **105** (2022) 054006 [arXiv:2106.10299] [INSPIRE].

[49] T. Sjöstrand and M. Utheim, *Hadron interactions for arbitrary energies and species, with applications to cosmic rays*, *Eur. Phys. J. C* **82** (2022) 21 [arXiv:2108.03481] [INSPIRE].

[50] M. Bonvini, R.D. Ball, S. Forte, S. Marzani and G. Ridolfi, *Updated Higgs cross section at approximate N^3LO*, *J. Phys. G* **41** (2014) 095002 [arXiv:1404.3204] [INSPIRE].

[51] J. Baglio, C. Duhr, B. Mistlberger and R. Szafron, *Inclusive production cross sections at N^3LO*, *JHEP* **12** (2022) 066 [arXiv:2209.06138] [INSPIRE].

[52] M. Cacciari, F.A. Dreyer, A. Karlberg, G.P. Salam and G. Zanderighi, *Fully Differential Vector-Boson-Fusion Higgs Production at Next-to-Next-to-Leading Order*, *Phys. Rev. Lett.* **115** (2015) 082002 [*Erratum ibid.* **120** (2018) 139901] [arXiv:1506.02660] [INSPIRE].

[53] F.A. Dreyer and A. Karlberg, *Vector-Boson Fusion Higgs Production at Three Loops in QCD*, *Phys. Rev. Lett.* **117** (2016) 072001 [arXiv:1606.00840] [INSPIRE].

[54] C. Bierlich et al., *Robust Independent Validation of Experiment and Theory: Rivet version 3*, *SciPost Phys.* **8** (2020) 026 [arXiv:1912.05451] [INSPIRE].

JHEP09(2024)088

[55] ATLAS collaboration, *Measurement of the transverse momentum distribution of Z/γ^* bosons in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Lett. B* **705** (2011) 415 [arXiv:1107.2381] [INSPIRE].

[56] ATLAS collaboration, *Measurements of the Higgs boson inclusive and differential fiducial cross sections in the 4ℓ decay channel at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **80** (2020) 942 [arXiv:2004.03969] [INSPIRE].

[57] CMS collaboration, *Measurement of energy flow at large pseudorapidities in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **11** (2011) 148 [*Erratum ibid.* **02** (2012) 055] [arXiv:1110.0211] [INSPIRE].

[58] CMS collaboration, *Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **105** (2010) 022002 [arXiv:1005.3299] [INSPIRE].

[59] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132 [arXiv:1412.7420] [INSPIRE].

JHEP09(2024)088