

PARTON DISTRIBUTION FUNCTIONS AND THEIR IMPACT ON PRECISION OF THE CURRENT THEORY CALCULATIONS*

MARIA UBIALI

DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK

*Received 9 April 2024, accepted 29 May 2024,
published online 5 August 2024*

The unprecedented precision of experimental measurements at the Large Hadron Collider (LHC) and the increased statistics that will be reached in the High-Luminosity phase of the LHC (HL-LHC) are pushing the phenomenology community to a new precision frontier, in which new challenges present themselves and new questions arise. Key ingredients of theoretical predictions at hadron colliders are the Parton Distribution Functions (PDFs) of the proton. This contribution highlights some of the new developments in the determination of PDFs from a global set of experimental data, from approximate N³LO PDFs and the inclusion of theory uncertainties in PDF fits, to the realisation of the non-trivial interplay between parton densities at large- x and possible signals of new physics in high-energy tails of the distributions, which highlights the synergy between high-energy and low-energy experimental programs.

DOI:10.5506/APhysPolBSupp.17.5-A6

1. Introduction

Precision phenomenology at hadron colliders and searches for deviations from the Standard Model (SM) predictions rely upon a precise and accurate estimate of the uncertainty in the SM predictions. Most theory predictions for hadron colliders require Parton Distribution Functions (PDFs) [1–9], the non-perturbative functions parametrising the subnuclear structure of the protons in terms of their partonic constituents for high-energy scattering processes.

PDFs cannot be obtained by perturbative methods, hence they must be fitted from experimental data. The space of possible parton distributions is an infinite-dimensional space of functions obeying the DGLAP evolution equations; as such, given only finite amounts of data, it is an ill-posed

* Presented at the 30th Cracow Epiphany Conference on *Precision Physics at High Energy Colliders*, Cracow, Poland, 8–12 January, 2024.

problem to determine the PDFs. Therefore, any PDF fitting attempt must initially restrict the infinite-dimensional PDF space to a finite-dimensional space instead, by assuming a functional form for the PDFs at some initial scale Q_0 ¹, and then DGLAP evolution is used to obtain the PDF at all scales.

A faithful determination of PDFs and of their uncertainty is a highly non-trivial problem and the precision of experimental data is presenting new challenges to PDF fitting collaborations. Accurate and precise PDFs are crucial at the LHC. For example, despite the progress in the precise determination of PDFs, PDF uncertainty is still one of the largest sources of theoretical uncertainty affecting the predictions for Higgs boson production [10–12]. The experimental PDF uncertainty is not the only component of the total uncertainty, as there is another component representing the mismatch in the perturbative order of the PDFs, evaluated at N²LO, and the perturbative QCD cross sections evaluated at N³LO [13, 14]. The recent publication of approximate N³LO PDF sets [15–17], which will be discussed in Section 2.1, is a crucial step in reducing this source of uncertainty. PDFs do also contribute to the precise extraction of the SM parameters from the LHC data, such as, for example, the strong coupling constant [18] and the W -boson mass [19], for which different PDF sets yield results that are significantly different as compared to the size of statistical and systematic uncertainties.

Finally, PDF uncertainties become sizeable at large- x due to the lack of precise experimental constraints in that region, and this yields a large uncertainty in the predictions for the high-energy tails of the measured distributions, where programs of indirect new physics searches focus, see, for example, the discussion in Ref. [20]. A precise and accurate description of the large- x content of the proton has always been crucial. As an illustration, the discrepancy that was observed by the CDF Collaboration in 1995 between the SM predictions for a jet inclusive cross section at large p_T [21] was mostly due to the lack of an estimate of the uncertainty of the gluon at large- x , and indeed the discrepancy disappeared once the CDF jet distributions were included by CTEQ in their global analysis of PDFs [22]. The story taught us the importance of propagating the experimental uncertainty in the space of PDFs. However, now that uncertainties are much more carefully estimated, how can we make sure that potential deviations from the SM predictions in the high-energy tails of the distributions are genuinely due to new physics rather than to our poor understanding of the proton structure at large- x ? This question will be the focus of the discussion in Section 2.3.

¹ This can be a set of polynomials that typically involves 30–50 parameters or a redundant parametrization such as a deep neural network, involving hundreds of parameters and a regularisation or minimisation stopping criterion, depending on the methodology.

2. New frontiers and challenges

As data precision increases, so does the challenge of balancing precision and accuracy. New PDF analyses sometimes lead to significant shifts from previous sets, potentially larger than nominal uncertainties, yet this does not undermine accuracy if the source of shifts is identified and the fitting process is controlled. Growing luminosity in experimental data poses challenges, especially with increasingly correlated systematics dominating datasets. Stabilizing covariance matrices and addressing systematic decorrelation are ongoing tasks crucial for assessing the goodness of fit [23, 24]. Continuous improvements and rigorous tests, ensure robustness, particularly important for modern PDF sets, see, for example, discussion on closure tests in Ref. [25] and future tests in Ref. [26].

In what follows, we will select three key aspects: the theory framework advancements that are behind approximate N³LO PDFs (discussed in Section 2.1), the inclusion of theoretical uncertainty in PDF fits (discussed in Section 2.2), and finally the interplay between PDF and new physics (discussed in Section 2.3).

2.1. Approximate N³LO PDFs

Calculations of partonic cross sections at N³LO order have been available for a long time for massless deep-inelastic scattering (DIS), and have become recently available for an increasingly large set of hadron processes, see Ref. [27] for an overview. To obtain theoretical predictions for hadronic observables at the same level of accuracy, the N³LO partonic cross sections must be convoluted with PDFs determined at the same perturbative order. The main bottleneck in carrying out this programme is the lack of exact expressions for the N³LO splitting functions that govern the scale dependence of the PDFs: for these, only partial information is available, see Ref. [17] for a review. However, a combination of the available partial results is possible, and an approximate determination of the N³LO splitting functions, see, for example, Refs. [28, 29].

Currently, two approximate N³LO PDF sets are available, with the MSHT set [29] available for a couple of years and the combined QED and N³LO MSHT set [15] and the NNPDF4.0 set [17] became available only recently. There are several differences in the approaches, as the MSHTaN3LO set includes all available theory input at the time of publication, while the NNPDF40aN3LO includes more theory inputs that were published in between, in particular 6(1) extra moments for the P_{qg} , P_{qq} (P_{qg} , P_{gg}) splitting functions, some terms in the large- n_f limit, and several sub-leading small- x and large- x terms. A benchmark between the two approaches at the level of splitting functions is in progress in a Les Houches benchmark exercise. An-

other important difference is that in the MSHT approach, the incomplete $N^3\text{LO}$ terms are added as variation in the prior and estimated by fitting nuisance parameters to the data, hence the posterior is determined by fitting the data, while the NNPDF approximation incorporates only theory inputs and their variations are added via an additional theory covariance matrix associated with incomplete missing higher orders (IHOU). Moreover, in the NNPDF approach, the missing higher-order uncertainties (MHOUs) associated with the NNLO contributions are added via an MHOUs theory covariance matrix [30, 31].

While the differences in the approach yield slightly discrepant predictions, overall a good perturbative convergence is observed, with differences decreasing as the perturbative order increases, and the approximate $N^3\text{LO}$ result is always compatible with the NNLO within uncertainties. Also, while for quark PDFs the difference between NNLO and $N^3\text{LO}$ results is extremely small, for the gluon PDF, a more significant shift is observed between NNLO and $N^3\text{LO}$ (this is much more marked in the MSHT set than in the NNPDF set), thus making the inclusion of $N^3\text{LO}$ crucial for precision phenomenology.

2.2. Missing higher-order uncertainties

As the overall precision in the determination grows, it is paramount to consistently treat the theoretical uncertainties of the input predictions in PDF fits. Until very recently, PDF uncertainties in NNLO PDF fits included only experimental component. However, such fits are performed by comparing fixed-order theoretical predictions to experimental data, and theoretical predictions are affected by MHOUs, which is usually estimated by means of scale variation. The inclusion of scale variations in a state-of-the-art NLO PDF fit was presented for the first time in Refs. [30, 31]. The approach, built upon the construction of a theory covariance matrix that describes the scale variations of the processes included in a PDF fit and models their theoretical correlations, has now been generalised to an NNLO PDF set [32]. The theory covariance matrix so determined can be added to the experimental covariance matrix, and the data are fitted using a total covariance matrix, which is the sum of the two contributions. The inclusion of MHOUs generally improves perturbative convergence and represents an important step towards achieving accuracy and improving perturbative convergence. A similar method has also been applied to nuclear uncertainties, which are relevant in the description of data based on nuclear targets [33, 34].

With the availability of PDF sets including MHOUs, the total uncertainty of any theory prediction would combine PDF uncertainty (now including an MHOUs component), with the MHOUs uncertainty on the partonic cross-section computations. In order to keep into account the correlation be-

tween the scale variations in the process used for PDF determination and the scale variation on the partonic cross section that an external user wishes to compute, one should compute the cross-correlation of MHO uncertainties between the predicted process and those used for PDF determination [35, 36]. An alternative option for the inclusion of MHOUs on predictions that accounts for MHOUs on PDFs and their full correlation to MHOUs on the partonic cross sections, is to include the scale variation in the Monte Carlo sampling, as proposed in the MCscales approach [37]. Basically, the Monte Carlo sampling method used to propagate the experimental uncertainty of the data into the PDFs is extended by adding scale fluctuations for each theoretical prediction on top of the fluctuations in the input data that are used to propagate experimental uncertainties. The development and the comparison of the various ways of including MHOUs in the estimate of PDF uncertainties will become increasingly crucial as the experimental component of PDF uncertainties is set to decrease thanks to the increased statistics and better understanding of the experimental systematics in Run 3 and 4, as well as in the HL-LHC phase.

2.3. PDFs and new physics interplay

Recent PDF analyses indicate that the LHC data are increasingly crucial in pinning down the parton densities, especially for the gluon and sea quarks the intermediate to large- x regions and its constraining power will become even more crucial in the HL-LHC phase [38]. As an increasing number of high-energy data from the LHC are included in PDF fits, the tails of the distributions that are used in PDF determination are potentially affected by new physics effects, which are usually parametrised by means of a model-independent EFT expansion [39]. To make sure that new physics is not absorbed or “fitted away” in the PDFs, one would either have to exclude these data, thus losing potentially important constraints, or carefully disentangle the SM and new physics effects.

Several studies have been performed in this direction by both experimental collaborations and theory groups [40–49]. In particular, it has been shown that the determination of SMEFT Wilson coefficients from a fit of the LHC data, like the determination of SM precision parameters from the LHC data [50], might display a non-negligible interplay with the input set of PDFs used to compute theory predictions. For example, in [41], it was shown that, while the effect of four-fermion operators on Deep Inelastic Scattering (DIS) data can be non-negligible if DIS data were fitted while taking the effect of such operators into account, the fit quality would deteriorate proportionally to the energy scale of the data included in the determination. This makes a fit based on DIS-only data “safe” from new physics effects.

On the other hand, in the context of high-mass Drell–Yan, especially in the HL-LHC scenario, neglecting the cross-talk between the large- x PDFs and the SMEFT effects in the tails could potentially miss new physics manifestations or misinterpret them [43, 47, 51], as the bounds on SMEFT operators are significantly broader if the PDFs are fitted by including the effect of the \hat{W} and \hat{Y} universal operators [52] in the high-energy tails of the data. In Ref. [47], it was shown that the large- x antiquark distributions can absorb the effect of universal new physics in the tails of the Drell–Yan distributions by leading to significantly softer antiquark distributions in the large- x region, far beyond the nominal PDF uncertainties. The study shows that the only way to avoid this kind of BSM contamination would be to include more low-energy observables in a PDF fit, thus highlighting the synergy between low- and high-energy experiments.

Furthermore, the analysis of the top-quark sector of Ref. [46] demonstrates that in that case, the bounds of the operators are not broadened by the interplay with the PDFs, however the correlation between the top sector and the gluon is manifest in the gluon PDF itself, which becomes softer in the large- x region if the PDF fit is augmented by the top data and PDFs are fitted simultaneously along the Wilson coefficients that determine the top-quark sector. This hints to a non-negligible interplay between large- x gluons and hints at new physics signals in the large invariant mass tails of the $t\bar{t}$ distributions, and possibly in the high- E_T jets distributions.

The recently publicly released tool **SIMUnet** [48] allows users to assess the interplay between PDFs and new physics, either by performing simultaneous fits of SMEFT operators (or operators in any other EFT expansion) alongside the PDFs of the proton, or by injecting any new physics model in the data and checking whether a global fit of PDFs can absorb the effects induced by such a model in the data. **SIMUnet**, based on the first tagged version of the public **NNPDF** code [53], augments it with a number of new features that allow for the exploration of the correlation between a fit of PDFs and BSM degrees of freedom. The tool allows for a global analysis of the SMEFT combining the Higgs, top, diboson, and electroweak sectors, and more data can be added in a rather straightforward way to combine the sectors presented here alongside Drell–Yan and flavour observables, for example. Further exploration of the SMEFT and PDF interplay and of the possible signal of new physics that might be fitted away by PDFs especially in the top and jets sector would add important pieces of the mosaic that has just started to be unveiled.

3. Conclusions

The success of the ambitious programme of the upcoming Run 3 at the LHC and its subsequent High- Luminosity (HL-LHC) upgrade relies not

only on achieving the highest possible accuracy in the experimental measurements and in the corresponding theoretical predictions, but also on the availability of statistically robust tools capable of yielding global interpretations of all subtle deviations from the Standard Model (SM) that the data might indicate. In this contribution, we discussed how the careful inspection and use of novel theoretical advances such as the recently developed approximate N³LO PDFs, the inclusion of MHOU and of QED effects are important to achieve the best accuracy at the LHC. Thanks to the number of newly released data, a powerful testing ground of the generalisation and extrapolation of the various PDF sets is available to be explored. Moreover, the new light shed on the interplay between PDFs and new physics can further advance the synergy between new experiments of the deeply-inelastic scattering (DIS), for example at the Electron–Ion Collider [54], or the perspective Large Hadron–Electron Collider [55], or even at the future Forward Physics Facilities planned at CERN [56]. The challenges that we are presented with are certainly a great opportunity to refine our tools and advance in our understanding of fundamental physics.

REFERENCES

- [1] PDF4LHC Working Group (R.D. Ball *et al.*), *J. Phys. G: Nucl. Part. Phys.* **49**, 080501 (2022), [arXiv:2203.05506 \[hep-ph\]](#).
- [2] NNPDF Collaboration (R.D. Ball *et al.*), *Eur. Phys. J. C* **77**, 663 (2017), [arXiv:1706.00428 \[hep-ph\]](#).
- [3] T.-J. Hou *et al.*, *Phys. Rev. D* **103**, 014013 (2021), [arXiv:1912.10053 \[hep-ph\]](#).
- [4] S. Bailey *et al.*, *Eur. Phys. J. C* **81**, 341 (2021), [arXiv:2012.04684 \[hep-ph\]](#).
- [5] NNPDF Collaboration (R.D. Ball *et al.*), *Eur. Phys. J. C* **82**, 428 (2022), [arXiv:2109.02653 \[hep-ph\]](#).
- [6] ATLAS Collaboration (G. Aad *et al.*), *Eur. Phys. J. C* **82**, 438 (2022), [arXiv:2112.11266 \[hep-ex\]](#).
- [7] A. Accardi *et al.*, *Phys. Rev. D* **93**, 114017 (2016), [arXiv:1602.03154 \[hep-ph\]](#).
- [8] S. Alekhin, J. Blümlein, S. Moch, R. Placakyte, *Phys. Rev. D* **96**, 014011 (2017), [arXiv:1701.05838 \[hep-ph\]](#).
- [9] J. Butterworth *et al.*, *J. Phys. G: Nucl. Part. Phys.* **43**, 023001 (2016), [arXiv:1510.03865 \[hep-ph\]](#).
- [10] LHC Higgs Cross Section Working Group (D. de Florian *et al.*), [arXiv:1610.07922 \[hep-ph\]](#).
- [11] S. Amoroso *et al.*, *Acta Phys. Pol. B* **53**, 12-A1 (2022), [arXiv:2203.13923 \[hep-ph\]](#).

- [12] J. Baglio, C. Duhr, B. Mistlberger, R. Szafron, *J. High Energy Phys.* **2022**, 066 (2022), [arXiv:2209.06138 \[hep-ph\]](#).
- [13] C. Anastasiou *et al.*, *J. High Energy Phys.* **2016**, 058 (2016), [arXiv:1602.00695 \[hep-ph\]](#).
- [14] B. Mistlberger, *J. High Energy Phys.* **2018**, 028 (2018), [arXiv:1802.00833 \[hep-ph\]](#).
- [15] T. Cridge, L.A. Harland-Lang, R.S. Thorne, [arXiv:2312.07665 \[hep-ph\]](#).
- [16] T. Cridge, L.A. Harland-Lang, R.S. Thorne, [arXiv:2404.02964 \[hep-ph\]](#).
- [17] NNPDF Collaboration (R.D. Ball *et al.*), *Eur. Phys. J. C* **84**, 659 (2024), [arXiv:2402.18635 \[hep-ph\]](#).
- [18] ATLAS Collaboration (G. Aad *et al.*), [arXiv:2309.12986 \[hep-ex\]](#).
- [19] ATLAS Collaboration (G. Aad *et al.*), [arXiv:2403.15085 \[hep-ex\]](#).
- [20] R.D. Ball *et al.*, *Eur. Phys. J. C* **82**, 1160 (2022), [arXiv:2209.08115 \[hep-ph\]](#).
- [21] CDF Collaboration (F. Abe *et al.*), *Phys. Rev. Lett.* **75**, 608 (1995).
- [22] J. Huston *et al.*, *Phys. Rev. Lett.* **77**, 444 (1996), [arXiv:hep-ph/9511386](#).
- [23] S. Bailey, L. Harland-Lang, *Eur. Phys. J. C* **80**, 60 (2020), [arXiv:1909.10541 \[hep-ph\]](#).
- [24] Z. Kassabov, E.R. Nocera, M. Wilson, *Eur. Phys. J. C* **82**, 956 (2022), [arXiv:2207.00690 \[hep-ph\]](#).
- [25] L. Del Debbio, T. Giani, M. Wilson, *Eur. Phys. J. C* **82**, 330 (2022), [arXiv:2111.05787 \[hep-ph\]](#).
- [26] J. Cruz-Martinez, S. Forte, E.R. Nocera, *Acta Phys. Pol. B* **52**, 243 (2021), [arXiv:2103.08606 \[hep-ph\]](#).
- [27] F. Caola *et al.*, [arXiv:2203.06730 \[hep-ph\]](#).
- [28] S. Moch *et al.*, *Phys. Lett. B* **849**, 138468 (2024), [arXiv:2310.05744 \[hep-ph\]](#).
- [29] T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, *Eur. Phys. J. C* **82**, 90 (2022), [arXiv:2111.05357 \[hep-ph\]](#).
- [30] R. Abdul Khalek *et al.*, *SciPost Phys.* **7**, 051 (2019), [arXiv:1906.10127 \[hep-ph\]](#).
- [31] NNPDF Collaboration (R. Abdul Khalek, J.J. Ethier, J. Rojo), *Eur. Phys. J. C* **79**, 471 (2019), [arXiv:1904.00018 \[hep-ph\]](#).
- [32] NNPDF Collaboration (R.D. Ball *et al.*), *Eur. Phys. J. C* **84**, 517 (2024), [arXiv:2401.10319 \[hep-ph\]](#).
- [33] R.D. Ball, E.R. Nocera, R.L. Pearson, *Eur. Phys. J. C* **81**, 37 (2021), [arXiv:2011.00009 \[hep-ph\]](#).
- [34] NNPDF Collaboration (R.D. Ball, E.R. Nocera, R.L. Pearson), *Eur. Phys. J. C* **79**, 282 (2019), [arXiv:1812.09074 \[hep-ph\]](#).
- [35] L.A. Harland-Lang, R.S. Thorne, *Eur. Phys. J. C* **79**, 225 (2019), [arXiv:1811.08434 \[hep-ph\]](#).

- [36] R.D. Ball, R.L. Pearson, *Eur. Phys. J. C* **81**, 830 (2021), [arXiv:2105.05114 \[hep-ph\]](#).
- [37] Z. Kassabov, M. Ubiali, C. Voisey, *J. High Energy Phys.* **2023**, 148 (2023), [arXiv:2207.07616 \[hep-ph\]](#).
- [38] R. Abdul Khalek *et al.*, *Eur. Phys. J. C* **78**, 962 (2018), [arXiv:1810.03639 \[hep-ph\]](#).
- [39] I. Brivio, M. Trott, *Phys. Rep.* **793**, 1 (2019), [arXiv:1706.08945 \[hep-ph\]](#).
- [40] ZEUS Collaboration (H. Abramowicz *et al.*), *Phys. Rev. D* **99**, 092006 (2019), [arXiv:1902.03048 \[hep-ex\]](#).
- [41] S. Carrazza *et al.*, *Phys. Rev. Lett.* **123**, 132001 (2019), [arXiv:1905.05215 \[hep-ph\]](#).
- [42] CMS Collaboration (A. Tumasyan *et al.*), *J. High Energy Phys.* **2022**, 142 (2022), [arXiv:2111.10431 \[hep-ex\]](#); *Addendum ibid.* **2022**, 035 (2022).
- [43] A. Greljo *et al.*, *J. High Energy Phys.* **2021**, 122 (2021), [arXiv:2104.02723 \[hep-ph\]](#).
- [44] M. McCullough, J. Moore, M. Ubiali, *J. High Energy Phys.* **2022**, 019 (2022), [arXiv:2203.12628 \[hep-ph\]](#).
- [45] J. Gao *et al.*, *J. High Energy Phys.* **2023**, 003 (2023), [arXiv:2211.01094 \[hep-ph\]](#).
- [46] Z. Kassabov *et al.*, *J. High Energy Phys.* **2023**, 205 (2023), [arXiv:2303.06159 \[hep-ph\]](#).
- [47] E. Hammou *et al.*, *J. High Energy Phys.* **2023**, 090 (2023), [arXiv:2307.10370 \[hep-ph\]](#).
- [48] M.N. Costantini *et al.*, [arXiv:2402.03308 \[hep-ph\]](#).
- [49] X.-M. Shen *et al.*, *PoS EPS-HEP2023*, 291 (2024).
- [50] S. Forte, Z. Kassabov, *Eur. Phys. J. C* **80**, 182 (2020), [arXiv:2001.04986 \[hep-ph\]](#).
- [51] S. Iranipour, M. Ubiali, *J. High Energy Phys.* **2022**, 032 (2022), [arXiv:2201.07240 \[hep-ph\]](#).
- [52] R. Torre, L. Ricci, A. Wulzer, *J. High Energy Phys.* **2021**, 144 (2021), [arXiv:2008.12978 \[hep-ph\]](#).
- [53] NNPDF Collaboration (R.D. Ball *et al.*), *Eur. Phys. J. C* **81**, 958 (2021), [arXiv:2109.02671 \[hep-ph\]](#).
- [54] R. Abdul Khalek *et al.*, *Nucl. Phys. A* **1026**, 122447 (2022), [arXiv:2103.05419 \[physics.ins-det\]](#).
- [55] LHeC, FCC-he Study Group (P. Agostini *et al.*), *J. Phys. G: Nucl. Part. Phys.* **48**, 110501 (2021), [arXiv:2007.14491 \[hep-ex\]](#).
- [56] J.M. Cruz-Martinez *et al.*, *Eur. Phys. J. C* **84**, 369 (2024), [arXiv:2309.09581 \[hep-ph\]](#).