New frontiers in PDF analyses in the HL-LHC era

Maria Ubiali (DAMTP, University of Cambridge, UK), M.Ubiali@damtp.cam.ac.uk

Precision phenomenology at the Large Hadron Collider (LHC) relies upon an accurate estimate of the uncertainty in Standard Model (SM) predictions. Two dominant sources of theoretical uncertainties at hadron colliders are missing higher order uncertainty in perturbative QCD calculations and parton distribution function (PDF) uncertainties. A global fit of PDFs involves several ingredients, which are all essential in order to achieve a robust determination of the proton subnuclear structure: the experimental data and an account of all their statistical, systematical errors and experimental correlations; the accuracy of the theoretical predictions used in the fit; and finally the methodology, which encompasses both the functional forms adopted to parametrise PDFs, the propagation of the experimental uncertainties onto the error associated with the fitted functions, the various minimisation techniques, the way to enforce positivity of physical observables, function integrability and sum rules. In this contribution I focus on what I believe to be the most fascinating new frontiers that the HL-LHC era will open in the context of PDF determination.

Recent PDF analyses clearly indicate that the LHC data is increasingly crucial in pinning down the parton densities, and its constraining power will become even more crucial in the High Luminosity (HL) run [1]. While in the NNPDF3.1 analysis [2] a reasonable consistency among different datasets is found (for example among the measurements sensitive to the large-x gluon [3], namely inclusive jet cross sections, transverse momentum distributions of the Z-boson and top-pair rapidity distributions) other analyses, such as the recent CT18 [4] display stronger tensions. Furthermore, preliminary comparisons of PDF sets suggests that differences among sets increased compared to PDF4LHC15 combination [5]. A thorough benchmark study to understand whether there is genuine tension among datasets or whether differences are due to different fitting methodologies is ongoing in the PDF4LHC working group.

Irrespectively of possible tensions among datasets, which may strengthen or weaken the current HL-LHC projections, we are in a situation such that, after years trying to reduce PDF uncertainties, the parton luminosity uncertainties are down to about 1% in the central rapidity region and for scales about the Z pole. Uncertainties may go down to 0.3-0.5% in a decade. Can we really trust PDFs to that level of precision? In such a situation, the precision versus accuracy challenge becomes crucial. Very often, when a new PDF analysis including new data is released by a PDF fitting collaboration, shifts between the previous and the PDF set may be larger than PDF uncertainties themselves. This does not undermine the accuracy of a PDF determination *per se*, as long as the origin of such shifts can be identified and all aspects of the fit are kept under control.

As far as *experimental data* are concerned, the main challenges have to do with datasets which, as the luminosity increases, are more and more dominated by correlated systematics. These highly-correlated datasets makes fits very hard, because small changes in the data covariance matrix lead to dramatic changes in the fully correlated χ^2 to the data. Studies on covariance matrix stabilisation and on the effects of the de-correlating systematics are ongoing and will become increasingly vital. They require a strong synergy between theorists and experimentalists. As far as the *fitting methodology* is concerned, several aspects are at play. Big shifts of PDFs are often observed as a result of an "a posteriori" data-driven parametrisation updates: fixed and somewhat restricted PDF parametrisation needs to be enlarged in order to include a new measurement in the fit, see for example the change in the down quark valence of Ref. [6]. A further source of shifts in the fitted PDFs has to do with improvements in the fitting methodology itself. For example in the NNPDF3.0 set [7] an improvement in the genetic algorithm minimisation originated a 1σ-level shift with respect to the previous NNPDF2.3 set [8]. As a result, a validation of the fitting methodology via statistical closure tests was introduced: PDFs determined from pseudo-data generated from a known underlying law must correctly reproduce the statistical distributions expected on the basis of the assumed experimental uncertainties. This kind of closure tests ensures that methodological uncertainties are negligible in comparison to the generic theoretical and experimental uncertainties of PDF determination. It is crucial that all modern PDF sets undergo similar tests. A further promising avenue that has been pursued in a series of recent studies [9] is the idea of a methodology fitting itself via hyper-parameter scan, to let the computer decide automatically the best methodology by scanning over thousands of hyper-parameter combinations and define a reward function to grade the methodology. Finally, the third crucial element is the *theory framework*. A large amount of work has been devoted to improve it in the past months. First of all, the theory error associated with the perturbative truncation of the theoretical predictions for the processes that enter PDF fits, that were believed to be generally less important than the experimental uncertainties, have become significant at the present level of precision. They have been included for the first time in a NLO PDF fit as an extra contribution to the covariance matrix when determining PDFs from data [10, 11]. The proposed approach is particularly interesting as other kinds of theory uncertainties, such as nuclear uncertainties [12], can be included in the same fashion. The natural expansion of this seminal work is the determination of the NNLO theory uncertainties and work towards N3LO PDFs. Other sources of uncertainties will become crucial in the future. For example, as even more high-energy data from the LHC will be included in PDF fits, the tails of the distributions that are used in PDF determination are potentially affected by new physics effects. To make sure that new physics is not absorbed or "fitted away" in PDFs one would either have to exclude this data, thus losing potentially important constraints, or thoroughly disentangle new physics effects. In [13] a systematic analysis of the interplay between PDFs and new physics effects, using the DIS structure functions, was given as a proof of concept and more work is ongoing.

To conclude, the precision physics frontier at HL-LHC opens up new fascinating challenges also in the field of PDF determination. The HL-LHC projections are extremely encouraging, with a foreseen reduction of PDF uncertainties by factor 2-3, but to achieve this goal benchmark among PDF sets and thorough scrutiny of each PDF analysis is necessary. A robust methodology based on statistical closure tests is vital, as well as an increased precision in theoretical predictions in PDF fits (N3LO PDFs, an estimate of missing higher order uncertainties, the inclusion of EW corrections, photon [14] and lepton PDFs [15]). A more general and ambitious aim is to make steps towards global fits of all parameters that enter QCD analyses (PDFs + α_s, PDFs + EW parameters, PDFs + BSM EFT parametrisation). Clearly a broad effort from the phenology community and a cross-talk with the experimentalists are essential to face the fascinating challenges that we are preparing for.

References

[1] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, "Towards Ultimate Parton Distributions at the High-Luminosity LHC," Eur. Phys. J. C, vol. 78, no. 11, p. 962, 2018.

[2] R. D. Ball et al., "Parton distributions from high-precision collider data," Eur. Phys. J. C, vol. 77, no. 10, p. 663, 2017.

[3] E. R. Nocera and M. Ubiali, "Constraining the gluon PDF at large x with LHC data," PoS, vol. DIS2017, p. 008, 2018.

[4] T.-J. Hou et al., "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC," 12 2019.

[5] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II," J. Phys. G, vol. 43, p. 023001, 2016.

[6] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. Martin, and R. Nathvani, "Updates of PDFs using the MMHT framework," PoS, vol. DIS2019, p. 036, 2019.

[7] R. D. Ball et al., "Parton distributions for the LHC Run II," JHEP, vol. 04, p. 040, 2015.

[8] R. D. Ball et al., "Parton distributions with LHC data," Nucl. Phys. B, vol. 867, pp. 244–289, 2013.

[9] S. Carrazza and J. Cruz-Martinez, "Towards a new generation of parton densities with deep learning models," Eur. Phys. J. C, vol. 79, no. 8, p. 676, 2019.

[10] R. Abdul Khalek et al., "A first determination of parton distributions with theoretical uncertainties," Eur. Phys. J., vol. C, p. 79:838, 2019.

[11] R. Abdul Khalek et al., "Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies," Eur. Phys. J. C, vol. 79, no. 11, p. 931, 2019.

[12] R. Pearson, R. Ball, and E. R. Nocera, "Uncertainties due to Nuclear Data in Proton PDF Fits," PoS, vol. DIS2019, p. 027, 2019.

[13] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo, and M. Ubiali, "Can New Physics hide inside the proton?," Phys. Rev. Lett., vol. 123, no. 13, p. 132001, 2019.

[14] A. V. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, "The Photon Content of the Proton," JHEP, vol. 12, p. 046, 2017.

[15] L. Buonocore, P. Nason, F. Tramontano, and G. Zanderighi, "Leptons in the proton," JHEP, vol. 08, no. 08, p. 019, 2020.