LHC precision EW working group

Proposal for PDF benchmarking exercise using LHC precision EW data and pseudodata

ABSTRACT

Over the past few years, a number of very precise measurements from the LHC experiments have been produced concerning inclusive Drell-Yan production of W/Z bosons. In parallel, the global fits to parton distribution functions have improved their accuracy through the use of the wealth of available LHC data, through the inclusion of the full HERA-2 data and of legacy precision measurements from the Tevatron, and through further developments of the underlying theory. Because of the intrinsic uncertainties related to the proton parton distribution functions, the precision of recent measurements of some of the fundamental parameters of the theory, such as the W-boson mass and the weak mixing angle, is limited in large part by the uncertainties in the parton distribution functions themselves. In addition, clear limitations have appeared in the use of purely perturbative QCD to describe the inclusive production of vector bosons at the LHC when performing global fits of the parton distribution functions.

This short note is a follow-up on a previous note sent to the PDF4LHC forum and lays out a skeleton of a new benchmarking exercise for the global PDF fits, which would be based on a common set of data used for global PDF fits, including a specific set of LHC run-1 data of relevance to precision Drell-Yan measurements. This set would then be augmented with pseudodata from the LHC experiments based on the full run-2 dataset, with the goal of evaluating the fit results obtained with the baseline dataset alone compared to that with the additional pseudodata. The benchmarking exercise would also include the production of a significant number of toys based on the LHC run-1 dataset with the goal of directly evaluating the correlations between the PDF fits obtained by the various groups. The large number of PDF fits required to fulfill the goals of this exercise may possibly only be realised by pooling computing resources between experiments and theory. How to do this in practice is one aspect of the discussions planned for the next PDF4LHC meeting on 13th of December 2018, where all the aspects of this proposal will be discussed with the experts in the community.

Proposal for discussion: **December 3, 2018**.

1 Introduction

In an initial note sent to the experts of the PDF4LHC forum in view of the November LHC precision EW working group meeting, we had summarised the main parton distribution function (PDF) issues which have been discussed in the context of various LHC run-1 precision analyses using the Drell-Yan process at the LHC. We had also provided specific requests to the experts from the PDF4LHC forum, with the goal of progressing sufficiently in our mutual understanding of how to deal with PDFs and their uncertainties for the combinations of precision EW measurements to be performed between the different LHC experiments with the full run-2 dataset over the next years.

At the November meeting itself, we had a nice and comprehensive experimental overview from the experiments in the morning, followed in the afternoon by very interesting presentations from the MMHT, NNPDF, CTEQ and ABM groups. We also enjoyed some lively discussions on a number of topics relevant to precision Drell-Yan measurements at the LHC. In a wrap-up session held at the end of the workshop, we emerged with a proposal to perform a second-generation benchmarking exercise after the very useful first benchmarking exercise which led to the birth of the PDF4LHC15 combined PDF set (this did not include ABM for a number of reasons).

This short note is a first attempt at specifying this second benchmarking exercise. In contrast to the first one, the experiments propose to participate in this exercise by using their computing resources to help provide run-2 pseudo-data, produce toys based on different datasets, and fit them in collaboration with each PDF group. The results of these fits would then be evaluated by all participants to assess more quantitatively the correlations between the different PDF fits when performed ideally on identical data using the same theory.

2 Overview of specifications for the benchmarking

The specifications for this benchmarking exercise can be divided into four broad categories:

1. The data to be used, where the word "data" covers both real data and pseudo-data provided by the experiments.

   Three sets of data are envisioned:

   - the first one (labelled "HERA") would correspond to the data from HERA (and possibly from lower-energy experiments if needed) which was used for the first benchmarking exercise.
   - the second one (labelled "LHC1") would correspond to the first one augmented with published precision Drell-Yan run-1 data from the LHC experiments (and from Tevatron). This LHC1 dataset is very important since it is the one from which reference PDF predictions would be based for future run-2 precision EW measurements. These reference PDF predictions would be presumably close to NNPDF3.1 and ABM16, which both use a significant fraction of the published precision run-1 data from ATLAS, CMS and LHCb.
   - the third one (labelled "LHC2") would correspond to the second one augmented with pseudodata (specified in more detail in the second item below) provided by the experiments, based on expectations of the most sensitive measurements with the full run-2 statistics, assumed to correspond to an integrated luminosity of 140 fb^{-1}.

2. The measured observables to be provided by the experiments for the run-2 pseudo-data in LHC2.

   These would include:

   - observables used for the extraction of the weak mixing angle, namely the A_4 angular coefficient in the full phase space of the decay leptons and the forward-backward asymmetry A_{FB} in the fiducial phase space of the decay leptons appropriate for ATLAS/CMS on one side and for LHCb on the other side. Both observables would be binned in dilepton mass and rapidity.
   - observables used for the extraction of the W-boson mass m_W, namely the lepton transverse momentum distribution, possibly binned in lepton pseudorapidity and in W-boson transverse momentum.
   - differential fiducial cross sections, e.g. the dilepton rapidity distribution binned in dilepton mass and the dilepton transverse momentum distribution in the region well described by perturbative QCD for Z-boson decays, together with the pseudo-rapidity distributions of the positively and negatively charged

leptons for W-boson decays. These might be augmented in terms of their sensitivity to PDFs and to the underlying theoretical description by providing two-dimensional distributions in lepton transverse momentum and pseudorapidity.

3. The theory used for Drell-Yan predictions and its uncertainties.

   Presumably, the exact same theory at NNLO in QCD and at LO in EW could be used, for example DYNNLO or FEWZ. If ready to be used for PDF fits, perhaps the most state-of-the-art tool for precision Drell-Yan predictions in perturbative QCD, namely NNLOJET could be considered. The EW scheme should be specified together with the input parameters used and those calculated internally. To account for QCD scale variations, possibly each PDF fit would be provided for a fixed list of scale choices (seven per process?). The uncertainties/biases (from a few per mil to a few per cent depending on the observable and the phase space explored) related to the use of purely perturbative QCD could be accounted for by a function provided by the experiments which would provide a correction for each relevant observable based on the ratio of resummed to perturbative predictions obtained eg by DYTurbo at NNLO+NNLL.

4. The PDF uncertainties and the correlations between them.

   The main goal of the benchmarking exercise would be to derive a mathematically corrected weighted average of precision EW measurements such as those of the weak mixing angle and the W-boson mass between the four main PDF sets, where the correctness of the weighted average requires a precise knowledge of the correlated and uncorrelated uncertainty components between the different PDF fits.

   Possibly the only way to obtain these without embarking on a restructuring of the internals of the PDF fits themselves would be to produce O(10000) toy datasets (from real data of the HERA/LHC1 dataset and from these augmented by the pseudodata for the run-2 LHC data of the LHC2 dataset). Each toy dataset would then be fitted by each PDF group and the output of each set of 10000 fits would be used to construct the correlations between each pair of PDF groups for each parton flavour of interest (starting for example with the valence u and d quarks). Note that the uncertainties need not be recomputed for each fit and that one set of uncertainties therefore covers in principle a specific set of 10000 toys.

In the following sections, some more details on the items above are given together with more specific questions to the PDF experts from each group.

3 HERA and LHC1 datasets

Questions to each PDF group:

- Do you believe there is added value to do the benchmarking exercise starting with the HERA dataset and then moving to the LHC1 dataset?
- From past experience would you agree that it is quite straightforward to define these two datasets? If yes, can you specify which LHC1 data you would recommend to use from each experiment? If you exclude certain data from the LHC experiments, can you explain on what basis this is done?
- Would you be happy if the LHC1 data were initially specified by the experiments themselves?
- Do you have specific requests on the format of the data and its information content? It may not be possible for certain published data to change things for practical reasons, but your input here is crucial to provide the information as uniformly and comprehensively as possible in the future.

4 LHC2 dataset and observables

The LHC full run-2 pseudodata will be provided in a HEPDATA format very similar to that provided by each experiment in their run-1 publications. These pseudodata will be produced in several flavours, depending on the input PDFs used:

- Based on the four LHC1 PDF fits obtained by each group, labelled eg as ABM1, CT1, MMHT1, and NNPDF1, each experiment would produce four sets of LHC2 pseudodata.

- Other pseudodata could be produced using alternate PDF fits, incorporating eg charrn or photon PDFs (the impact of such contributions is currently under study and was partially covered in the presentations of the November workshop).
- Each set of pseudodata would be incorporated into one specific instance of an LHC2 dataset which would then go through the toy generation procedure and the global PDF fit per toy procedure outlined above.
- The QED ISR effects if considered important might lead us to produce pseudodata with or without including QED ISR. This point needs further study to assess the impact quantitatively as well as possible before considering the inclusion of these additional effects in the benchmarking exercise.

Questions to each PDF group:

- Given the large potential for inflation of datasets if one varies too many parameters in the configuration of the LHC run-2 pseudodata, which are the ones you consider most critical to the benchmarking exercise?
- Do you have any additional suggestions for observable on top of the ones listed above for the run-2 pseudo-data?

5 Theoretical tools and uncertainties

From a benchmarking point of view, it would be highly desirable to use a common set of predictions for all global PDF fits. These predictions might be provided by the experiments (benefitting thus from their pooled CPU resources on the computing grid) and perhaps a state-of-the-art tool such as NNLOJET could be considered for this if ready to be technically used for PDF fits.

Questions to each PDF group for Drell-Yan process calculations:

- Which theory tools do you use in pQCD? Would you object to using another tool with the same accuracy in pQCD? If yes, why?
- What is the exact configuration with which you run your tool? Electroweak scheme, input parameters chosen, other SM parameters calculated, please provide a table of values. Would you object to using the same set of values and scheme for all PDF groups for this exercise? If yes, why?
- Do you consider it important to run with a fit including the photon PDF from LuxQED? If yes, why?
- Would you accept to use a correction function provided externally to account for small but significant biases arising from impact of resummed versus perturbative calculation of fiducial cuts? If not, how would you propose to account for these effects?
- Would you accept to produce an envelope of uncertainties arising from unknown higher-order corrections in QCD by producing PDF fits with different renormalisation and factorisation scale choices? Do you agree that these scale variations should be produced independently for each physics process considered? If not, which approach would you advocate to properly account for the theory uncertainties in PDF fits given that these uncertainties are not accounted for at the moment? If yes, which variations of scales would you consider reasonable in terms of minimising the amount of computing time devoted to this exercise? Please note also here that only the results of the PDF fit with different scale choices are important, the uncertainties need not be reevaluated for each fit.
- Are there other uncertainties which even though small are not accounted for by your theory model?

6 Determination of correlations between different PDF fits

Questions to each PDF group concerning the determination of the correlations between global PDF fits:

- Do you agree that the proposed way of determining the correlations between global PDF fits would be feasible without any significant restructuring of the way the fits are performed and their uncertainties determined?

- Would you agree with the O(10000) magnitude of the toys required to establish precisely the correlations between the fitted PDFs obtained by the different groups? If yes, would you agree that it would make sense to use the pooled resources of our full community (experiments and theory) to produce and fit these toys? Assuming that most of the CPU time is required for the PDF fit itself, how much time does your specific PDF fit require?

- In practice, do you already produce such toy datasets internally based on the output of your PDF fits? If yes, what would be in your opinion the additional information one might learn in terms of closure by also producing the toys directly from the fit result?

- Would you agree that the XFITTER framework would be an appropriate tool in practice to accomplish our goals? If not, could you specify what might be missing or inadequate in XFITTER? In any event, the benchmarking exercise will be organised such that the data and the theoretical predictions would be made available to all PDF groups to perform fits in their own framework, especially in the first step of the benchmarking exercise which will require significant validation effort of the technical details.

- The specific resources of each PDF group would certainly be useful but might be insufficient to perform the global fits themselves (one global fit per toy yielding a set of PDFs but not their uncertainties. Can you confirm that this is indeed the case and could you help us estimate the timescale required to produce such benchmarking results? A not very educated guess at what the overall timescale might be is given below to provide a concrete template for this discussion:

  1. Preparation of tools and data (possibly with the help of XFITTER) and first trial runs: January-March
  2. Preparation, validation and running of NNLO QCD predictions: January-May
  3. Production of toys and of fit results: March-June
  4. Second iteration based on first set of results: June-October
  5. Documentation of results for Yellow Report: June-December

- To summarise, could you conclude by stating to what extent your group would be willing to contribute to such an exercise? If you perceive that there could be significant resource and timescale problems, which parts of it might be curtailed to achieve our goals within available resources and the target timescale of one year? If you perceive that the scope is incomplete or skewed or oversimplified, which modifications would you propose concretely?

7 Summary

This note hopefully lays out with sufficient detail and clarity how a second-generation benchmarking test of the global PDF sets could be carried out with the specific goal of defining in an unambiguous and clear way how to account properly for the spread observed between different PDF sets for precision EW measurements at the LHC. We believe that the half-day discussion in the next PDF4LHC forum on 13th of December will be very beneficial to converge all together on the definition of this exercise, such that it can be launched immediately after the New Year. A first fairly complete set of results could hopefully be available after a few months of preparations and test runs by summer 2019, such that a comprehensive report of our joint findings could be written in the second half of 2019. This would be synchronised quite well with the time when full run-2 measurements from the LHC experiments would become mature for publication and soon after for combinations with a well understood and agreed upon methodology for dealing with combined uncertainties from PDFs. We emphasise again here how important this benchmarking exercise is, since we know from the run-1 measurements that the PDF uncertainties are among the main limiting factors in the ultimate achievable precision which LHC experiments could provide to the global EW fit.

References