Suggestions on calculating the PDF4LHC prescription

Stefano Forte, Joey Huston, Robert Thorne, Alessandro Vicini

September 21, 2010

Suggestions on calculating the PDF4LHC prescription for cross-sections and the PDF and α_s uncertainty as quickly as possible.

A general comment is that we do not consider any approximation which changes the uncertainty by 10% or less to be important as the uncertainty on the uncertainty is of this order. We also note that all the PDFs required and a general interface for using them is available at http://projects.hepforge.org/lhapdf/ and very widely used.

The use of MSTW, NNPDF and CTEQ PDFs in the recipe leads to a factor, of three extra in time, not a large factor, and other than to some extent the case of NNLO, see below, there is no way to improve upon this. However, this seems unlikely to be the main reason for excessive computation time.

To calculate the PDF uncertainty for each set requires summing over 40 or more eigenvectors for CTEQ or MSTW, or a similar minimum number of replicas for NNPDF. There is nothing that can be done to avoid this, but it leads to a great saving in time if one avoids calculating the cross-section anew for each eigenvector by using PDF re-weighting.

The majority of the CPU time in a NLO (NNLO) calculation is due to the evaluation of the NLO (NNLO) matrix elements. To form the full cross section, the matrix elements have to be convoluted with PDFs determined at the same order of perturbative QCD. There is a factorization present in the determination of a cross section at NLO (NNLO) such that a re-evaluation of the matrix elements is not necessary when calculating the PDF uncertainty of that cross section using error PDFs. The relative weights of the error PDFs, with respect to the central PDF of a global PDF set, are easy to evaluate and to store. This is straightforward, in most cases, to implement in any NLO (NNLO) program. For example, MCFM [1], which has a large collection of NLO processes, will automatically calculate the PDF uncertainty for any observable, using the error PDF sets based on a central PDF. The information on the error PDF weights can also be stored for each event, for example in a ROOT ntuple. MCFM uses the FROOT formalism developed by Pavel Nadolsky. The format allows for the storage of all parton 4-vectors, the total event weight, the event weight by the initial state partons, and the event weights for the

parton distribution error PDFs. A generalization of the FROOT formalism was developed in Les Houches 2009 [2]. The storage of the error PDF information in a ROOT ntuple does increase the storage requirements, but has little impact on the CPU necessary for evaluation of the cross section. Thus, the provisions for (1) automatic evaluation of error PDFs in any NLO (NNLO) program and (2) the ability to store event information in a ROOT ntuple are strongly recommended.

Most experiments also have the capability of performing the re-weighting themselves, as long as the information for each event for the kinematics, scales and flavors are available. The PDF re-weighting should work exactly for the case of fixed order calculations. There are complications when used with parton shower Monte Carlos, as the PDFs are also used in the weighting for the initial state showering. However, it has been shown that any resulting error introduced into the Sudakov form factor is negligible [3].

Specific comments which may help for each of the individual PDF sets are:

- For CTEQ the recommended procedure of adding the α_s uncertainty in quadrature makes the process as streamlined as possible once PDF reweighting has been applied. Within the quadratic approximation for χ^2, this procedure is exact.

- For MSTW calculating the addition α_s uncertainty involves repeating the uncertainty calculation for the central α_s value for four additional values, i.e. +/- half sigma and +/- one sigma α_s eigenvector sets, and taking the envelope. This is a factor of 5 compared to just taking one eigenvector set. A quicker approximation for MSTW is to follow the CTEQ procedure of treating the central PDF sets for the + and - one sigma α_s variation as one more eigenvector pair to be added in quadrature with the 20 eigenvector pairs for the default α_s set, i.e. $\Delta_+^{PDF+\alpha_s} = \sqrt{\left(\Delta_+^{\alpha_s}\right)^2 + \left((\Delta_{PDF}^{\alpha_s^0})_+\right)^2}$ and $\Delta_-^{PDF+\alpha_s} = \sqrt{\left(\Delta_-^{\alpha_s}\right)^2 + \left((\Delta_{PDF}^{\alpha_s^0})_-\right)^2}$ This is then very nearly equal in time to calculating PDF uncertainty only. It tends to underestimate the total α_s uncertainty for MSTW a little, most notably in the gluon-gluon fusion process, but is a reasonable approximation, particularly when used in the envelope when this is only a small contribution to the total uncertainty.

- For NNPDF the recommended procedure is to take a gaussian distribution of replicas. For example with the recommended range $\alpha_s = 0.119 \pm 0.0012$, and the recommended $N_{rep} = 100$ replicas for the central value, take $N_{rep} = 1, 5, 27, 72, 100, 72, 27, 5, 1$ replicas with $\alpha_s = 0.115, 0.116, 0.117, 0.118, 0.119, 0.120, 0.121, 0.122, 0.123$. This would corresponds to 310 replicas, i.e. a factor 3 more than the recommended 100 replicas for the central value. However, because the total α_s+PDF uncertainty is of the same order of magnitude as the PDF uncertainty, this increase in number of replicas is not needed, and the above number of replicas can be cut by a factor three. Hence, it is sufficient to take $N_{rep} = 2, 9, 24, 33, 24, 9, 2$ for $\alpha_s = 0.116, 0.117, 0.118, 0.119, 0.120, 0.121, 0.122.$, corresponding to a total number of $N_{rep} = 103$ in order to determine the α_s+PDF uncertainty with roughly the same accuracy

as that obtained for the PDF only uncertainty when using $N_{rep} = 100$ replicas with fixed α_s. Cutting by a further factor two the number of replicas in each bin would lead to results approximately as accurate as those obtained with $N_{rep} = 50$ replicas with fixed α_s. This is probably still acceptable in most cases, though a further reduction in the number of replicas is usually not advisable. In case of limitations of computing time, it is recommended to perform the calculation with $N_{rep} = 50$ for the selected values of the Higgs mass, and only increase the number of replicas if results do not appear to behave smoothly as the mass is varied.

Regarding the number of bins in m_H for which calculations need be done. The variation of the PDF+α_s uncertainty with Higgs mass is very smooth as can be seen from figure 13 of arXiv:1004.0962, plots on http://projects.hepforge.org/mstwpdf/pdf4lhc/ggHiggs7 TeV.html and various LHC-Higgs results so far produced. Hence, the uncertainty can be calculated to a good approximation by consideration of a relatively small number of mass points. Gaps of 20 GeV should be plenty, and the evidence suggests that even gaps of 50 GeV (particularly for higher masses) should lead to little loss of accuracy if linear interpolation is used between the points. Of course for any process a check should be made that there is no rapid variation in uncertainty for a given range of mass (rapid being changes of much more than 10% in uncertainty between mass points). Particular regions where one might be more careful are the lowest masses, i.e. near 100 GeV, and in the vicinity of the top pair resonance.

Regarding the NNLO prescription, we note that NNLO calculations only need be done in this case for MSTW. If necessary, to save time the approximation for the α_s uncertainty described above can be used. For the multiplicative factor for the uncertainty which is obtained at NLO from MSTW, NNPDF and CTEQ we particularly note that this is only an approximation intended to mimic the expected result if NNLO NNPDF and CTEQ sets were available, so there is even less reason to be very precise in calculating this factor and an accuracy of 10% is easily good enough. We note that for $gg \to H$ this factor has already been found to be roughly 2 with the variation with m_H being only about 10%, and at worst about 20%, for 7 TeV and 14 TeV. The factor may not be the same for other processes, indeed it appears to be less for VBF, but it can be obtained from a small number of Higgs masses - the separation in m_H being sufficient only that the variation in the expansion factor (i.e. the ratio of the envelope to the MSTW uncertainty alone), between points is not significantly above 10%. This may well be possible with gaps of 50 GeV or even larger. One the scaling factor has been found with sufficient accuracy (i.e. at the level of 10%) for a given process once, a fit to a parametric form can be made and used for subsequent studies.

Finally we note that if any group has specific questions regarding the use of the PDFs and appropriate procedures they are free to contact us with questions. We have not refused or ignored any requests of this type so far.

References

[1] J. M. Campbell and R. K. Ellis, MCFM home page, http://mcfm.fnal.gov

[2] J. R. Andersen *et al.* [SM and NLO Multileg Working Group], arXiv:1003.1241 [hep-ph].

[3] S. Gieseke, JHEP **0501**, 058 (2005) [arXiv:hep-ph/0412342].