Practical implementation of the PDF4LHC recipe [1]

1 General remarks

Let us call $\mathcal{O}$ a generic observable, the total cross section or the value in one bin of a distribution. We want to evaluate its central value and the associated α_s+PDF uncertainty: $\mathcal{O}_0{}^{+\sigma(\alpha_s+PDF,+)}_{-\sigma(\alpha_s+PDF,-)}$, where the total uncertainty can be, in general, not symmetric.

- All the uncertainties, due to the PDFs and to α_s, are evaluated at 68% C.L. .
- It has been chosen to consider $\Delta\alpha_s = 0.0012$ as the 68% C.L. variation of the strong coupling constant.
- Since only some confidence-level intervals are practically available for the different quantities, we need to apply rescaling factors to obtain the 68% C.L. intervals, assuming gaussian scaling of the uncertainties. Let us call C_X the factor by which we have to divide an uncertainty interval with a confidence level equal to X%, to obtain the corresponding 68% C.L. value. The following values will be used:
  $C_{90} = 1.64485...$ (PDF uncertainty with CTEQ);
  $C_{59} = 5/6$ (α_s uncertainty with CTEQ);
  $C_{79} = 5/4$ (lower α_s uncertainty with MSTW).
- The PDF+α_s uncertainty band is a smooth function of the Higgs mass. In order to evaluate it in the whole Higgs mass spectrum, it is sufficient to compute the uncertainty with a spacing of 20 GeV for $100 \leq m_H \leq 500$ (GeV) and with a spacing of 100 GeV for $500 \leq m_H \leq 1000$ (GeV).
- Given the present knowledge of the PDFs, it is sufficient to determine the precise value of the total uncertainty band with a 10% accuracy.

[1]prepared by Alessandro Vicini, alessandro.vicini@mi.infn.it
based on the prescription by the PDF4LHC working group, cfr: *The PDF4LHC Working Group Interim Report*, the updated version can be downloaded from https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSectionsCalc

2 CTEQ6.6, NLO-QCD

2.1 PDF uncertainty

The file `cteq66.LHgrid` contains 45 members: `mem=0` is the best fit, `mem=1,...,44` are the 44 sets that describe the PDF uncertainty.

- Use `mem=0` to evaluate the central value $\mathcal{O}_0$.
- Use the $N^{CTEQ} = 44$ members `mem=1,...,44` to evaluate the 68% C.L. PDF uncertainty[2], according to the formula [3]

$$\sigma^{CTEQ}(PDF,+) = \frac{1}{C_{90}}\sqrt{\sum_{i=1}^{N^{CTEQ}/2}\left(\max\{\mathcal{O}[\{q^{(2i-1)}\}]-\mathcal{O}[\{q^{(0)}\}],\mathcal{O}[\{q^{(2i)}\}]-\mathcal{O}[\{q^{(0)}\}],0\}\right)^2}$$

$$\sigma^{CTEQ}(PDF,-) = \frac{1}{C_{90}}\sqrt{\sum_{i=1}^{N^{CTEQ}/2}\left(\max\{\mathcal{O}[\{q^{(0)}\}]-\mathcal{O}[\{q^{(2i-1)}\}],\mathcal{O}[\{q^{(0)}\}]-\mathcal{O}[\{q^{(2i)}\}],0\}\right)^2}$$

where $\mathcal{O}[\{q^{(i)}\}]$ is the value of the observable evaluated with the PDF set `mem=`i.

2.2 α_s uncertainty

The best-fit PDF set has been extracted by CTEQ using $\alpha_s(M_Z) = 0.118$ as input parameter. The file `cteq66alphas.LHgrid` contains 5 members (`mem=0,1,...,4`) that correspond to parton densities extracted using as input parameter $\alpha_s(M_Z) = 0.116, 0.117, 0.118, 0.119, 0.120$ respectively.

- Compute the observable $\mathcal{O}^{(+)}_{\alpha_s}$ using `mem=3` as PDF set and using the corresponding value $\alpha_s(M_Z) = 0.119$ in the matrix element (the α_s value should be automatically available after the initialization of the PDF member, if one uses the LHAPDF interface).
- Compute the observable $\mathcal{O}^{(-)}_{\alpha_s}$ using `mem=1` as PDF set and using the corresponding value $\alpha_s(M_Z) = 0.117$ in the matrix element.
- Call

$$\sigma^{CTEQ}(\alpha_s,\pm) = \left(\mathcal{O}^{(\pm)}_{\alpha_s} - \mathcal{O}_0\right)\frac{1}{C_{59}} \quad (1)$$

the 68% C.L. uncertainties due to an α_s variation of $\Delta\alpha_s$ [4].

2.3 Combination of PDF and α_s uncertainties

The combination of the PDF and α_s uncertainties can be obtained by summing in quadrature the two values.

$$\sigma^{CTEQ}(\alpha_s + PDF,\pm) = \sqrt{\left(\sigma^{CTEQ}(PDF,\pm)\right)^2 + \left(\sigma^{CTEQ}(\alpha_s,\pm)\right)^2} \quad (2)$$

[2]The PDF uncertainty provided by CTEQ represents a 90% C.L. interval and it has to be divided by C_{90} to obtain the corresponding 68% C.L. value.

[3]The use of symmetric uncertainties is also possible; in many cases the difference with the asymmetric treatment tends to be small.

[4]It is possible to use `mem=0` and `mem=4` to compute $\mathcal{O}^{(\pm)}_{\alpha_s}$, but then it is necessary to divide by C_{90} instead of C_{59} to obtain the 68% C.L. interval.

3 MSTW2008, NLO-QCD

3.1 PDF uncertainty

The file `MSTW2008nlo68cl.LHgrid` contains 41 members: `mem=0` is the best fit, `mem=1,...,40` are the 40 sets that describe the PDF, at 68% C.L., uncertainty.

- Use `mem=0` to evaluate the central value $\mathcal{O}_0$.
- Use the $N^{MSTW} = 40$ members `mem=1,...,40` to evaluate the asymmetric 68% C.L. PDF uncertainty, according to the formulae

$$\sigma^{MSTW}(PDF,+) = \sqrt{\sum_{i=1}^{N^{MSTW}/2} \left(\max\{\mathcal{O}[\{q^{(2i-1)}\}] - \mathcal{O}[\{q^{(0)}\}], \mathcal{O}[\{q^{(2i)}\}] - \mathcal{O}[\{q^{(0)}\}], 0\}\right)^2}$$

$$\sigma^{MSTW}(PDF,-) = \sqrt{\sum_{i=1}^{N^{MSTW}/2} \left(\max\{\mathcal{O}[\{q^{(0)}\}] - \mathcal{O}[\{q^{(2i-1)}\}], \mathcal{O}[\{q^{(0)}\}] - \mathcal{O}[\{q^{(2i)}\}], 0\}\right)^2}$$

where $\mathcal{O}[\{q^{(i)}\}]$ is the value of the observable evaluated with the PDF set `mem=`i.

3.2 α_s uncertainty

The best-fit PDF set has been extracted by MSTW together with a fit of the strong coupling constant, whose best value is, at NLO-QCD, $\alpha_s(M_Z) = 0.12018$.
The file `MSTW2008nlo68cl_asmz+68cl.LHgrid` contains 41 members (`mem=0,1,...,41`) that correspond to parton densities extracted using as input parameter $\alpha_s(M_Z) = 0.12018 + 0.0012$.
The file `MSTW2008nlo68cl_asmz-68cl.LHgrid` contains 41 members (`mem=0,1,...,41`) that correspond to parton densities extracted using as input parameter $\alpha_s(M_Z) = 0.12018 - 0.0015$.

- Compute the observable $\mathcal{O}^{(+)}_{\alpha_s}$ using `mem=0` of the file `MSTW2008nlo68cl_asmz+68cl.LHgrid` as PDF set and using the corresponding value $\alpha_s(M_Z) = 0.12138$ in the matrix element (the α_s value should be automatically available after the initialization of the PDF member, if one uses the LHAPDF interface).
- Compute the observable $\mathcal{O}^{(-)}_{\alpha_s}$ using `mem=0` of the file `MSTW2008nlo68cl_asmz-68cl.LHgrid` as PDF set and using the corresponding value $\alpha_s(M_Z) = 0.11868$ in the matrix element.
- Call

$$\sigma^{MSTW}(\alpha_s,+) = \left(\mathcal{O}^{(+)}_{\alpha_s} - \mathcal{O}_0\right) \qquad \sigma^{MSTW}(\alpha_s,-) = \left(\mathcal{O}^{(-)}_{\alpha_s} - \mathcal{O}_0\right)\frac{1}{C_{79}}. \tag{3}$$

the 68% C.L. uncertainties due to an α_s variation of $\Delta\alpha_s$ [5].

3.3 Combination of PDF and α_s uncertainties

The combination of the PDF and α_s uncertainties can be obtained by summing in quadrature the two values.

$$\sigma^{MSTW}(\alpha_s + PDF, \pm) = \sqrt{\left(\sigma^{MSTW}(PDF,+)\right)^2 + \left(\sigma^{MSTW}(\alpha_s,\pm)\right)^2} \tag{4}$$

[5] The rescaling by $1/C_{79}$ of the downwards α_s uncertainty is not the MSTW recommendation, but it is done to obtain a uniform treatment of the α_s variation among the different collaborations.

4 NNPDF2.0, NLO-QCD

4.1 Combination of PDF and α_s uncertainties

- The files `NNPDF20_as_0114_100.LHgrid`, `NNPDF20_as_0115_100.LHgrid`, ... , `NNPDF20_as_0124_100.LHgrid` contain each 100 replicas of PDF sets, which have been extracted by setting in input $\alpha_s(M_z) = 0.114, 0.115, ..., 0.124$ respectively. The file `NNPDF20_as_0119_100.LHgrid`≡ `NNPDF20_100.LHgrid` coincides with the best-fit ensemble of replicas.
  In each of the above files, the 100 replicas give a representation of the PDF uncertainty.

- The combined PDF+α_s uncertainty can be estimated combining in a new ensemble N_{rep} replicas that have been extracted with different values of $\alpha_s(M_Z)$ in input.

- Assuming the probability distribution of values of α_s to be gaussian and peaked around $\alpha_s^{(0)} = 0.119$, the number of replicas corresponding to a given value $\alpha_s = \alpha_s^{(j)}$ is

$$N_{\rm rep}^{\alpha_s^{(j)}} \propto \exp\left(-\frac{\left(\alpha_s^{(j)} - \alpha_s^{(0)}\right)^2}{2\left(\Delta\alpha_s\right)^2}\right) \qquad\qquad N_{\rm rep} = \sum_{j=1}^{N_{\alpha_s}} N_{\rm rep}^{\alpha_s^{(j)}} . \tag{5}$$

  with the normalization constraint given by N_{rep} as the total number of replicas used in the evaluation [6].

- Setting $N_{rep} = 50$, the numbers of replicas to be evaluated is $(1, 4, 12, 16, 12, 4, 1)$ with $\alpha_s(M_Z) = 0.116, 0.117, 0.118, 0.119, 0.120, 0.121, 0.122$ respectively.
  From the practical point of view, it means that one has to initialize `NNPDF20_as_0116_100.LHgrid` and evaluate the observable $\mathcal{O}$ with one member of this set,
  then initialize `NNPDF20_as_0117_100.LHgrid` and evaluate $\mathcal{O}$ four times with four members of this set,
  and so forth with the other files obtained with the different values of $\alpha_s(M_Z)$.

- The central value and PDF+α_s uncertainty are obtained by computing the mean value,

$$\mathcal{O}_0 \;=\; \langle\mathcal{O}\rangle_{rep} = \frac{1}{N_{rep}} \sum_{j=1}^{N_\alpha} \sum_{k_j=1}^{N_{\rm rep}^{\alpha_s^{(j)}}} \mathcal{O}\left(\mathrm{PDF}^{(k_j,j)}, \alpha_s^{(j)}\right) , \tag{6}$$

  and the standard deviation

$$\sigma^{NNPDF}(\alpha_s + PDF) = \left[\frac{1}{N_{rep}-1} \sum_{j=1}^{N_\alpha} \sum_{k_j=1}^{N_{\rm rep}^{\alpha_s^{(j)}}} \left(\mathcal{O}\left(\mathrm{PDF}^{(k_j,j)}, \alpha_s^{(j)}\right) - \mathcal{O}_0\right)^2\right]^{1/2} \tag{7}$$

  of the N_{rep} results obtained so far. In eqs.(6, 7) $\mathcal{O}\left(\mathrm{PDF}^{(k_j,j)}, \alpha_s^{(j)}\right)$ indicates the observable $\mathcal{O}$ evaluated with `mem=`k_j extracted with $\alpha_s^{(j)}$.

- To achieve a better statistical accuracy, a larger number of replicas N_{rep} can be used, multiplying by the same factor the number of replicas indicated above for the various α_s values.

[6] At most it is possible to use 100 replicas in the central bin with $\alpha_s(M_Z) = 0.119$

5 PDF4LHC envelope and correction at NNLO-QCD

- The NLO-QCD envelope has to be computed, with (i=CTEQ,MSTW,NNPDF),

$$
\begin{aligned}
U &= \max_i \{\mathcal{O}_0^i + \sigma^{(i)}(\alpha_s + PDF, +)\} \\
L &= \min_i \{\mathcal{O}_0^i - \sigma^{(i)}(\alpha_s + PDF, -)\} \\
M &= \frac{U+L}{2}
\end{aligned}
\tag{8}
$$

  where U, L are the upper and lower edges of the envelope and M is its mid-point.

- The percentual width of the envelope and the percentual width of the MSTW NLO-QCD PDF+α_s uncertainty bands

$$
\delta_{env} = \frac{U-M}{M}, \qquad \delta^{\pm}_{MSTW,NLO} = \frac{\sigma^{MSTW,NLO}(PDF + \alpha_s, \pm)}{\mathcal{O}_0^{MSTW,NLO}} \tag{9}
$$

  have to be compared to form the rescaling factor

$$
R^{\pm} = \frac{\delta_{env}}{\delta^{\pm}_{MSTW,NLO}} \tag{10}
$$

- The MSTW α_s+PDF uncertainty band can be obtained at NNLO-QCD, or in presence of NNLO-QCD+resummation, by means of the `MSTW2008nnlo` PDF sets, according to a procedure identical to the one described in Section 3, with the replacement of `nlo` with `nnlo`.

- The total envelope at NNLO-QCD is given by the MSTW-NNLO α_s+PDF band, multiplied by the rescaling factor $R^{\pm}$ (upper edge of the MSTW band times R^+, lower edge with R^-).

6 Reweighting

In any code where the partonic cross section of a process is already subtracted of initial state collinear divergences, it is possible to speed up the evaluation of all the different PDF sets that parametrize the PDF uncertainty (only the PDF uncertainty), by a reweigthing procedure. Let us say that any integration over phase space is discretized, so that the value of any observable, evaluated with PDF set i, can be written as a sum

$$\mathcal{O}^{(i)} = \sum_j \mathcal{O}_j^{(i)} = \sum_j \mathcal{L}_j^{(i)} \hat{\sigma}_j \tag{11}$$

where j runs over all relevant phase-space points, $\mathcal{L}_j^{(i)}$ are the corresponding PDF values and $\hat{\sigma}_j$ the partonic cross section.

- The mean value of the observable over N_{rep} PDF replicas is

$$\mathcal{O}_0 = \frac{1}{N_{rep}} \sum_{i=1}^{N_{rep}} \sum_j \mathcal{O}_j^{(i)} = \sum_j \left(\frac{1}{N_{rep}} \sum_{i=1}^{N_{rep}} \mathcal{L}_j^{(i)} \right) \hat{\sigma}_j \tag{12}$$

  It is evident that for every phase-space point j it is possible to evaluate in one shot the contribution of all the N_{rep} PDF replicas.

- The standard deviation is

$$\begin{aligned} \sigma_{\mathcal{O}} &= \left\{ \frac{1}{N_{rep}-1} \sum_{i=1}^{N_{rep}} \left[\mathcal{O}_i - \mathcal{O}_0 \right]^2 \right\}^{1/2} \\ &= \left\{ \frac{1}{N_{rep}-1} \sum_{i=1}^{N_{rep}} \left[\sum_j \mathcal{L}_j^{(i)} \hat{\sigma}_j - \sum_j \left(\frac{1}{N_{rep}} \sum_{l=1}^{N_{rep}} \mathcal{L}_j^{(l)} \right) \hat{\sigma}_j \right]^2 \right\}^{1/2} \\ &= \left\{ \sum_j \frac{1}{N_{rep}-1} \left[\sum_{i=1}^{N_{rep}} \mathcal{L}_j^{(i)} - \left(\frac{1}{N_{rep}} \sum_{l=1}^{N_{rep}} \mathcal{L}_j^{(l)} \right) \right]^2 \hat{\sigma}_j \right\}^{1/2} \end{aligned} \tag{13}$$

  For every space point, it is possible to compute in one shot all the N_{rep} contributions in square brackets.

- The implementation of eqs.(12,13) can be obtained by using the LHAPDF interface.