ManeParse: a Mathematica reader for Parton Distribution Functions

D. B. Clark[a,1]**, E. Godat**[b,1]**, F. I. Olness**[c,1]

[1]Department of Physics, Southern Methodist University, Dallas, TX 75275, USA

arXiv:1605.08012v1 [hep-ph] 25 May 2016

Abstract Parton Distribution Functions (PDFs) are essential non-perturbative inputs for calculation of any observable with hadronic initial states. These PDFs are released by individual groups as discrete grids as a function of the Bjorken-x and energy scale Q. The LHAPDF project at HepForge maintains a repository of PDFs from various groups in a new standardized LHAPDF6 format, as well as older formats such as the CTEQ PDS grid format. ManeParse is a package that provides PDFs within the Mathematica framework to facilitate calculating and plotting. The program is self-contained so there are no external links to any FORTRAN, C or C++ programs. The package includes the option to use the built-in Mathematica interpolation or a custom cubic Lagrange interpolation routine which allows for flexibility in the extrapolation (particularly at small x values). ManeParse is fast enough to enable simple calculations (involving even one or two integrations) to be easily computed in the Mathematica framework.

Contents

[a]e-mail: dbclark@smu.edu

[b]e-mail: egodat@smu.edu
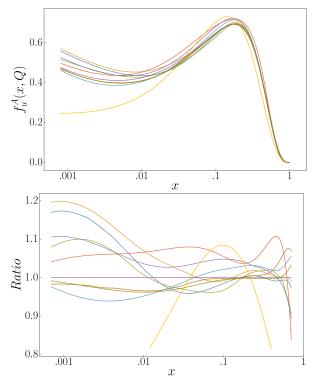
[c]e-mail: olness@smu.edu

Fig. 1 a) We display $x f_u^A(x, Q)$ for the up-quark at $Q = 2$ GeV as a function of x for the 10 PDFs listed in Table 1. b) We display the ratio of the PDFs in a) compared to CT10 proton PDF ($A = 1$) as a function of x. While we don't identify them individually, the one curve (yellow) that distinctly deviates from the others is the nuclear PDF for lead $A = 208$.

1 What is ManeParse?

Parton Distribution Functions (PDFs) are essential elements for making predictions involving hadrons (protons and nuclei) in the initial state. For example, at the LHC, we can compute the Higgs production cross section (σ) using the formula $\sigma = f_{a/P} \otimes f_{b/P} \otimes \omega_{ab \to H}$ where PDFs $f_{a/P}$ and $f_{b/P}$ give the probability density for finding partons "a" and "b" in the two proton beams, and the hard cross section, $\omega_{ab \to H}$, gives the probability density for partons a and b producing the Higgs, H. The PDFs cannot be computed from first principles at this time[1], so they must be extracted using fits to experimental data. This analysis is performed by a number of collaborations, and the PDFs are generally distributed as a grid of values in x and Q which must be interpolated to generate the PDF $f_a(x, Q)$ for flavor "a" at momentum fraction x and energy scale Q.

---

[1]Lattice QCD has made great strides in computing PDFs in recent years. [5, 19]

ManeParse[2] is a flexible, modular, lightweight, standalone package used to provide access to a wide variety of PDFs within Mathematica. To illustrate the flexibility, in Fig. 1 we show how ManeParse can work simultaneously with different PDF sets from a variety of groups. This figure[3]displays the selected PDF sets listed in Table 1. Some of the sets are in the LHAPDF6 grid format[10], and others are in the older CTEQ PDS grid format.[21] These sets also have different numbers of active flavors, N_F, different values for the initial evolution scale, Q_0, different values for the heavy quark masses, $\{m_c, m_b, m_t\}$, and they can represent either protons or nuclei. Nevertheless, ManeParse is able to easily compare and contrast sets from different groups in a common framework.

We describe the key features of ManeParse available to the user. In Section 2, we sketch a minimal example of how the program is used. In Section 3, we provide some details of how the PDFs are parsed, stored and interpolated. In Section 4, we display some example plots that are easily constructed using ManeParse. In Section 5, we provide examples of the functions in the pdfError module. In the Appendix, we discuss files provided by ManeParse and how to obtain the external PDF files.

2 A simple example

We begin by outlining a simple example of how ManeParse may be used. After loading the ManeParse packages into Mathematica, the user can enter the following commands:

```
Get[pdfParseLHA.m]
iSet1=pdfParseLHA[LHA_file.info,LHA_file.dat]
pdfFunction[iSet1,iParton,x,Q]

Get[pdfParseCTEQ.m]
iSet2=pdfParseCTEQ[PDS_file.pds]
pdfFunction[iSet2,iParton,x,Q]
```

The first and fourth line load the parsing subpackages included in ManeParse. Loading either of these, causes the pdfCalc package to be loaded as well. The second line reads an LHAPDF6 formatted external data file (LHA_File.dat) and its associated information file (LHA_File.info), and generates an internal PDF set that is referenced by the integer iSet1. The fifth line reads a PDS formatted external data file[4] (PDS_File.pds) and

---

[2]The ManeParse program was originally developed to run on the SMU computing cluster "ManeFrame" which is a play on words inspired by the school mascot, Peruna the pony.

[3]All plots presented here have been generated in Mathematica.

[4]Note that the LHAPDF6 files have both a data file and an info file whereas the older CTEQ PDS files have only a data file.

PDF Set			Total	$\overline{u}$	$\overline{d}$	g	d	u	s	c	b
MSTW2008nnlo68cl [20]			99.87	3.3	3.8	43.5	14.6	29.3	2.0	0.7	0
CT14nnlo [12]			100.01	3.1	3.7	43.4	14.6	29.7	2.0	0.8	0
NNPDF30_nnlo_as_0118_nf_6 [7]			99.98	3.2	3.7	43.6	14.6	29.4	2.2	0.8	0
HERAPDF20_NLO_VAR [1]			99.98	3.9	3.0	41.7	14.6	31.2	2.2	0.6	0
abm12lhc_5_nnlo [3]			100.14	2.9	3.5	43.4	14.8	30.4	2.0	0.7	0
CJ15nlo [2]			99.96	3.0	3.7	43.3	15.1	29.8	1.8	0.7	0
nCTEQ15_1_1 [14]			100.10	3.1	3.8	43.0	15.0	30.2	1.8	0.7	0
nCTEQ15_208_82 [14]			99.99	2.7	3.4	44.6	17.0	27.2	1.8	0.8	0
ct10.pds [18]			99.97	3.0	3.7	43.4	14.6	29.6	2.2	0.7	0
ctq66m.pds [21]			99.98	2.9	3.6	43.6	14.5	29.4	2.3	0.7	0

Table 1 We compute the momentum sum rule, Eq. 1, (in percent) for the individual partons at $Q = 3$ GeV. Partons $\{\bar{s}, \bar{c}, \bar{b}\}$ are not shown, but are equal to $\{s, c, b\}$. The totals sum to 100% within uncertainties of integration and interpolation. Here the colors matched with each set correspond to that set in Fig. 1 and Fig. 7

generates an internal PDF set that is referenced by the integer `iSet2`.

After reading these data files, the user is provided with the core function for computing the PDFs: `pdfFunction[iSet,iParton,x,Q]`. Here, `iSet` selects the individual PDF set, `iParton` selects the parton flavor as shown in Table 2, and `{x,Q}` specify the momentum fraction, x, and the energy scale, Q, in GeV.

`pdfFunction` performs the bulk of the work for the `ManeParse` program, so the package has been optimized for speed to make it practical to perform single or double integrals in a reasonable amount of time; specifically, the `pdfFunction` call generally takes less than 1ms per core on a standard laptop or desktop.

Additionally, `ManeParse` can handle an arbitrary number of PDF sets and can switch between sets without delay. When the external PDF file is parsed, the data is stored internally (about 1Mb per PDF set) and the `iSet` variable essentially functions as a pointer to the set; thus, it is trivial to loop over many PDF sets as was done in Fig. 1. This feature contrasts to some of the older `FORTRAN` programs, which could only store a fixed number of sets in memory and often had to re-read the data files.

These are the key elements of the package, however, we also provide many auxiliary functions described below. Consistent with the `Mathematica` convention, all our public functions begin with the prefix "`pdf`". One can obtain a complete list with the command `?pdf*`.

The usage message for individual functions is displayed in a similar manner to:

> `?pdfFunction`

pdfFunction[setNumber, flavor, x, Q]
− This function returns the interpolated value of the PDF for the `.pds`/`.dat` file specified by *setNumber*, for the given *flavor* and value of Bjorken x and scale Q.
− *Warning*: The results of this function are only reliable between the maximum and minimum values of x and Q in the `.pds`/`.dat` file[5].

3 Inside the `ManeParse` Package

3.1 Overview of package

`ManeParse` internally consists of four modules (or subpackages) as illustrated in Fig. 2. The modular structure of `ManeParse` allows for separate parsers for the `LHAPDF6` (`pdfParseLHA`) and PDS (`pdfParseCTEQ`) grids which read the individual file types and pass the information on to a common calculation (`pdfCalc`) module.

---

[5]If interpolation outside the given grid is requested by the user, `ManeParse` is equipped to handle this. The `Mathematica` interpolator will throw a warning message and proceed to use built-in extrapolation techniques. The `ManeParse` interpolator will extrapolate using the behavior defined with `pdfSetXpower`.
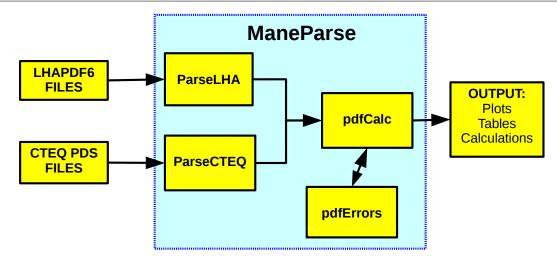
Fig. 2 A schematic overview of the `ManeParse` package and the individual modules.

flavor #	0 or 21	± 1	± 2	± 3	± 4	± 5	± 6
parton	gluon	down/dbar	up/ubar	strange/sbar	charm/cbar	bottom/bbar	top/tbar

Table 2 The standard Monte Carlo (MC) flavor numbering convention[23] used within `ManeParse`. This differs from the mass-ordered convention used in many older CTEQ releases. `ManeParse` converts these releases into the MC ordering.

The new `LHAPDF6` format is intended as a standard that all groups can use to release their results; additionally, many older PDF sets have been converted into this format.

The `ManeParse` modular structure provides flexibility, as the user can use both `LHAPDF6` and PDS format, or even write a custom parser to read a set that is not in one of these formats.

The error PDFs module (`pdfError`) uses `pdfCalc` to construct PDF uncertainties, luminosities, and correlations as illustrated in Sec. 5.

The key elements of each PDF set include the 3-dimensional $\{x, Q, N_F\}$ grid and the associated information, which is stored as a set of `Mathematica` rules. We now describe the features and some details of these structures.

3.2 The PDF $\{x, Q, N_F\}$ grid

The parsing routines `pdfParseLHA` and `pdfParseCTEQ` read the external files and assemble the PDF sets into a common data structure that is used by the `pdfCalc` module. The central structure is a 3-dimensional grid of PDF values in $\{x, Q, N_F\}$ space, which uses vectors $\{x_{vec}, Q_{vec}\}$ to specify the grid points. The spacing of $\{x_{vec}, Q_{vec}\}$ need not be uniform; typically, Q_{vec} uses logarithmic spacing, and x_{vec} is commonly logarithmic at small x and linear at large x. Different spacings in

x_{vec} and Q_{vec} do not pose a problem for the `pdfCalc` package, as the grid points are simply interpolated to provide the PDF at a particular point in $\{x, Q, N_F\}$. The user is agnostic to the specific grid spacing chosen in a PDF release.

3.2.1 N_F Convention

The N_F flavor dimension is determined by the `iSet` value passed to `pdfFunction`. The association between the grid slice in N_F and `iSet` is specified in the `LHAPDF6` info file using the "`key:data`" format such as "`Flavors: [−5, −4, −3, −2, −1, 1, 2, 3, 4, 5, 21]`". This tells us which partons are in the grid, and their proper order.[6] Note: we use the standard Monte Carlo (MC) convention[7] throughout `ManeParse` where $d = 1$ and $u = 2$ rather than the mass-ordered convention (see Table 2).[8] The standard MC convention also labels the gluon as `iParton = 21`; for compatibility, the gluon in `ManeParse` can be identified with either `iParton = 21` or `iParton = 0`.

---

[6] For the PDS files, this information is contained in the header of the data file so there is not a separate info file; `pdfParseCTEQ` extracts the proper association.

[7] See Ref. [23] "Review of Particle Physics," Chapter 34 entitled "Monte Carlo particle numbering scheme."

[8] Caution is required here as many of the older CTEQ releases use the mass-ordered convention with $u = 1$ and $d = 2$. `ManeParse` converts these mass-ordered sets into the MC ordering.
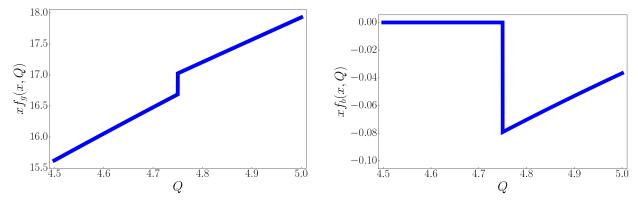
Fig. 3 The discontinuity of the gluon (left) and b-quark (right) PDFs across the $m_b = 4.75$ GeV flavor threshold; the horizontal axis is Q (in GeV), and the vertical axis is $x\,f(x,Q)$. The curves are for the MSTW2008nnlo68cl PDF with $x = 10^{-4}$. Note that the gluon and b-quark shift in opposite directions to ensure the momentum sum rule is satisfied.
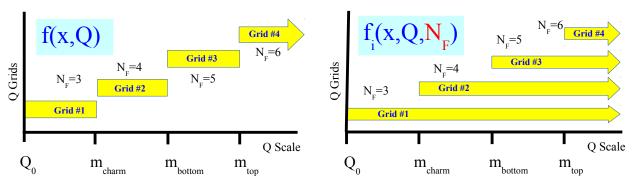

Fig. 4 A schematic representation of the Q sub-grids used to handle discontinuities across the heavy quark thresholds at $\{m_c, m_b, m_t\}$. Fig. a) shows the conventional arrangement for $f(x, Q)$ with non-overlapping sub-grids; for a given Q, the N_F flavor dimension is uniquely determined. Fig. b) shows a flavor-number-dependent PDF $f(x, Q, N_F)$ where the user has the freedom to choose the N_F flavor dimension value (and hence the sub-grid).

ManeParse is able to work with PDF sets with different numbers of flavors. For example, in Fig. 1, the NNPDF set includes $N_F = 6$ where iParton $= \{\bar{t}, \ldots, t\}$, while most of the other sets have $N_F = 5$. If a flavor, iParton, is not defined, pdfFunction will return zero. This feature allows the user to write a sum over all quarks $\sum f_i(x, Q)$ for $i = \{-6, \ldots, 6\}$ without worrying whether some PDF sets might have less than 6 active flavors.

Additionally, the ManeParse framework has the flexibility to handle new particles such as a 4th generation of quarks with iParton $= \{b', t'\} = \{7, 8\}$ or a light gluino with iParton $= \tilde{g} = 1000021$ PDF by identifying the flavor index, iParton, with the appropriate grid position in the LHAPDF6 info file.

3.2.2 Q Sub-Grids

At NNLO and beyond, the PDFs can become discontinuous across the mass flavor thresholds. This is illustrated using the NNLO MSTW set in Fig. 3 where we observe a discontinuity of both the gluon and b-quark PDF across the b-quark threshold at $m_b = 4.75$ GeV.

ManeParse accommodates this by using sub-grids in Q as illustrated in Fig. 4-a); for example, we use separate grids below and above the threshold at $Q = m_b = 4.75$ GeV. When we call the PDF at a specific Q value, ManeParse looks up the relevant heavy quark thresholds, $\{m_c, m_b, m_t\}$, to determine which sub-grid to use for the interpolation. For $Q < m_b$, sub-grid #2 ($N_F = 4$) is used, and for $Q \geq m_b$, sub-grid #3 ($N_F = 5$) is chosen.

Note that for the x value (10^{-4}) displayed in Fig. 3, the b-quark PDF is negative for Q just above m_b; this is the correct higher-order result and justifies (in part) why we do not force the PDFs to be positive definite. This behavior also makes sense in terms of the momentum sum rule, which we will discuss in Sec. 4.3.

3.2.3 An N_F-dependent PDF: $f(x, Q, N_F)$

Note, the use of sub-grids in Q also enables the use of overlapping N_F ranges as in a hybrid scheme as described in Ref. [16]; in this case, we generalize the PDF so that it also becomes a function of the number of flavors: $f(x, Q, N_F)$. This feature is useful if, for example,

YAML	Mathematica
key: "data"	"key" → "data"
SetDesc: "nCTEQ15 ..."	"SetDesc" → "nCTEQ15 ..."
NumFlavors: 5	"NumFlavors" → 5
Flavors: [-5,-4,-3,-2,-1,1,2,3,4,5,21]	"Flavors"→{-5,-4,-3,-2,-1,1,2,3,4,5,21}
AlphaS_Qs: [1.299999e+00, ...]	"AlphaS_Qs"→ $\{1.299999 \times 10^{+00}, ...\}$
UnknownKey: data	"UnknownKey"→"data"

Table 3 Sample YAML entries contained in the `LHAPDF6` info file, and the corresponding rules passed to `Mathematica`. The rules for a specific PDF set are obtained using the `pdfGetInfo[iSet]`function.

we are performing a fit to data in the region $Q \sim m_b$; we can perform a consistent $N_F = 4$ flavor fit even if some of the data are above the $N_f = 5$ threshold ($Q > m_b$) by selecting $f(x, Q, N_F = 4)$; thus, we avoid encountering any discontinuities in the region of the data.[9] We illustrate this generalized case for $f(x, Q, N_F)$ in Fig. 4-b). Here, the user has the freedom to choose the active number of flavors, N_F, rather than being forced to transition at the quark mass values as in Fig. 4-a).

3.3 The `LHAPDF6` Info File

In addition to the 3-dimensional $\{x, Q, N_F\}$ grid, there is auxiliary material associated with each PDF set. In the LHA format, each PDF collection has an associated "info" file which contains the additional data in a YAML format,[10] whereas in the CTEQ PDS format files, the auxiliary information is contained at the top of each PDS data file. Each parser interprets this information and builds a list of `Mathematica` rules.

The basic syntax of YAML is [`key`: "`data`"], and the LHA parser converts this into a `Mathematica` rule as {"`key`"→"`data`"}. This can be viewed within `ManeParse` using the function `pdfGetInfo[iSet]`, and Table 3 demonstrates the some sample mappings between the two.

If "`key`" is known to be a number, "`data`" is converted from a string into a number. This behavior applies to values such as {NumFlavors, QMin, MTop, ...}. If "`key`" is known to be a list such as {Flavors, AlphaS_Qs}, "`data`" is converted from a string into a `Mathematica` list. If "`key`" is unknown, "`data`" is left as a string. This means that `ManeParse` can handle any

unknown "`key`", and the user can modify these rules after the fact, or introduce a custom modification by identifying "`key`" to the parser.

3.4 Interpolation

Once the 3-dimensional {x,Q,N_F} grid and auxiliary rules are given to the `pdfCalc` module, we are ready to interact with the PDFs. When the user calls for $f_i(x, Q)$, the `pdfCalc` module will determine the appropriate N_F index and Q grid and do a 4-point interpolation in the 2-dimensional $\{x, Q\}$ space. Generally, `pdfCalc` will interpolate $\{x, Q\}$ values with 2 grid points on each side, but at the edges of the grid, it will use a 3-1 split. It also will extrapolate beyond the limits of the grid and will return a number, even if it is unphysical. Except for setting $f_i(x, Q) = 0$ for $x > 1$, we do not check bounds, as this would slow the computation; in the sample files, we do provide examples of how the user can implement particular boundaries if desired.

Additionally, we allow the interpolated PDF to be negative. At very large x this can happen due to numerical uncertainty, but there are also instances where a negative PDF is the physical result, such as at NNLO (illustrated in Fig. 3). Within `Mathematica`, it is easy for the user to impose particular limits (*i.e.* positivity) if desired. The interpolation can be performed either with the `Mathematica Interpolate` function (default) or a custom 4-point Lagrange interpolator and is set with the `pdfSetInterpolator` function. We set the `Mathematica Interpolate` function as the default, as it is slightly faster, but the custom 4-point Lagrange interpolator often will provide better extrapolation of the PDFs beyond the grid boundaries and has some adjustable parameters which are useful in the low x region.

---

[9]Note that the APFEL PDF evolution library[8] is in the process of implementing these features.
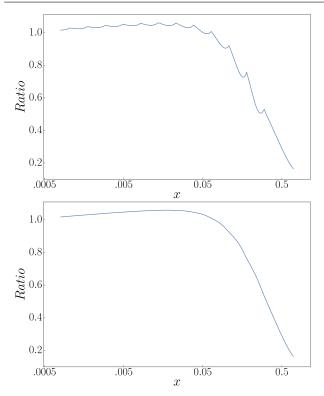
[10]"YAML Ain't Markup Language" http://yaml.org/

Fig. 5 The ratio of PDFs can sometimes lead to interpolation problems; we display the ratio of two gluon PDFs at $Q = 100$ GeV. Fig.-a) on the left was generated with the default Mathematica interpolator, and Fig.-b) on the right was generated with the custom 4-point Lagrange interpolation with the default scaling of $a = 1$.

Fig. 6 $\alpha_S(Q)$ vs. Q in GeV from NNPDF. Note the discontinuity across the $m_b = 4.18$ GeV threshold which is enlarged in Fig. b).

Because the PDFs typically scale as $1/x$ for small x, the grids actually store $x\, f_i(x, Q)$, as this generally improves the interpolation. To return $f_i(x, Q)$ we divide by x, but to avoid dividing by zero we internally impose a default minimum x value of $x_{min} = 10^{-30}$.

In principle, the interpolation of $f_i(x, Q)$ can be generalized using a scaling factor $1/x^a$ with a variable a. This is one of the adjustable parameters in the custom 4-point Lagrange interpolation, and can be set with the pdfSetXpower[a] function.

We have bench-marked many of the PDF sets to ensure our interpolations are accurate across the defined grid in $\{x, Q\}$ space. For the PDS files, our interpolation (with xPower = 1.0) essentially uses the same algorithm so our results match better than one part in 10^3. When bench-marking with LHAPDF6 interpolation, our interpolation is generally matches to 2 parts in 10^3, and this may increase at very high or low x values. We also compared our interpolator while varying xPower from 0.5 to 1.5 as well as the Mathematica interpolator to LHAPDF6 interpolation; these agree better than 1 part in 10^3 for $x > 3 \times 10^{-3}$, but can increase for small x. If necessary, the user can supply a custom interpolation routine.
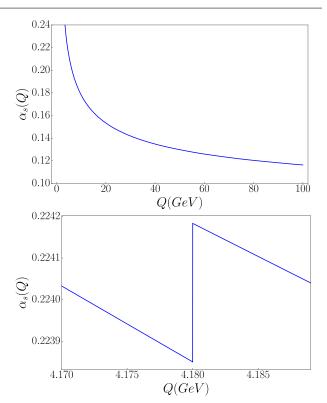
We find that ratios of PDFs are more sensitive to the interpolation than the PDFs themselves. For illustrative purposes, in Fig. 5, we show an example of a poor interpolation generated with the Mathematica interpolator compared to a good interpolation by the custom 4-point Lagrange interpolation with the default $a = 1$ scaling; in general, we find the custom 4-point Lagrange interpolation computes smoother ratios and provides better extrapolation beyond the grid limits.

3.5 α_S Function

For some of the PDF sets, the value of $\alpha_S(Q)$ is provided as a list of points associated with Q_{vec}. For these sets, we interpolate[11] $\alpha_S(Q)$ to provide a matched function called pdfAlphaS[iSet,Q]; this is displayed in Fig. 6 for a sample PDF set. The pdfGetInfo[iSet] function will display the information associated with the corresponding PDF set (including any α_S values). If the PDF set does not have any α_S information, the pdfAlpha function will return Null. In Fig. 6-a) we display $\alpha_S(Q)$ for the NNPDF set, and in Fig. 6-b) we enlarge the region near $m_b = 4.18$ GeV to display the disconti-

---

[11] Since at Leading Order (LO), $\alpha_S(Q) = 1/[\beta_0 \ln(Q^2/\Lambda^2)]$, we obtained improved results by interpolating in $1/\alpha_S(Q)$.

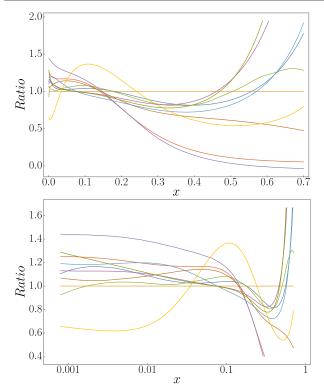**Fig. 8** Small x extrapolation of the gluon PDF from the nCTEQ15 proton at $Q = 100$ GeV using `pdfLowFunction`. Here, $x_{min} = 5 \times 10^{-6}$, and the extrapolation exponent $1/x^a$ is set to $a = \{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$.

Fig. 9 The integrated momentum fraction Eq. 1 of the PDF flavors vs. Q in GeV for the NNPDF set. At large Q the curves are (in descending order) $\{g, u, d, \bar{u}, \bar{d}, s, c, b, t\}$.

Fig. 7 Sample linear and log ratio plots of the gluon PDFs from Table 1 compared to CT14 as a function of x at $Q = 2.0$ GeV.

nuity. In general, $\alpha_S(Q)$ will be discontinuous at NNLO and higher and at all mass thresholds, $\{m_c, m_b, m_t\}$.

4 Sample Plots & Calculations

The advantage of importing the PDF sets into Mathematica is that we have the complete set of built-in tools that we can use for calculating and graphing. We illustrate some of these features here.

4.1 Graphical Examples

To highlight the graphical capabilities, in Figure 7 we display a selection of PDFs using both linear (top) and log (bottom) scale. Using the flexible graphics capabilities of `Mathematica` it is easy to automatically generate such plots for different PDF sets.

4.2 Small x Extrapolation

Sometimes it is useful to extrapolate to low x values beyond the limits of the PDF grid; for example, the study of high energy cosmic ray experiments that use very small x extrapolations. [6, 9] We provide `pdfLowFunction[iSet,iParton,x,Q,power]` which allows the user to
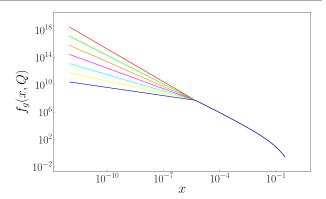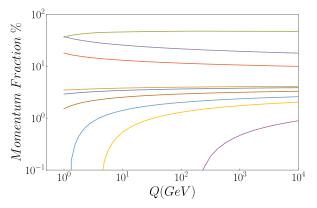
choose the extrapolation power in the small x region.[12] An example is displayed in Fig. 8 for the nCTEQ15 proton PDF. The minimum x value for this set for the grid is $x_{min} = 5 \times 10^{-6}$; beyond this limit `pdfLowFunction` will extrapolate using the form $1/x^a$. In this example, we vary the power from 0.4 to 1.6; using the `Mathematica` integration routines it is easy to find that this range of variation in the small x behavior will only change the momentum fraction of the gluon by 1/2%.

4.3 Momentum Sum Rules

The PDFs satisfy a number of momentum and number sum rules, and this provides a useful cross check on the results. The momentum sum rule:

$$\sum_i \int_0^1 dx\, x\, f_i(x, Q) = 1, \tag{1}$$

---

[12]The "a" argument is optional; the default power is 1.0. We use a separate function `pdfLowFunction` so as not to slow the computation of `pdfFunction`.
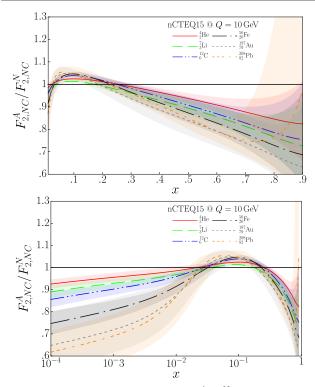
Fig. 11 The differential parton-parton luminosity $dL_{a\bar{a}}/dM_X^2$ vs. M_x in GeV at $\sqrt{s} = 14$ TeV for (in descending order) $a = \{g, u, d, s, c, b\}$.

$F_2^A(x, Q) = x \sum_q e_q^2 f_{q/A}(x, Q)$ at leading order where F_2^N is an isoscalar, and F_2^A is the scaled structure function[14] for nuclei A. We have also superimposed the uncertainty bands; we will discuss this more in the following Section.

4.5 Luminosity

Using the integration capabilities of Mathematica it is easy to compute the differential parton-parton luminosity [11] for partons a and b:

$$\frac{d\mathcal{L}_{ab}}{d\hat{s}} = \frac{1}{s} \frac{1}{1 + \delta_{ab}} \int_\tau^1 \frac{dx}{x} f_a(x, \sqrt{\hat{s}}) f_b(\frac{\tau}{x}, \sqrt{\hat{s}}) + (a \leftrightarrow b),$$

(2)

where $\tau = \hat{s}/s$, and the cross section is

$$\sigma = \sum_{a,b} \int \left(\frac{d\hat{s}}{\hat{s}}\right) \left(\frac{d\mathcal{L}_{ab}}{d\hat{s}}\right) (\hat{s}\,\hat{\sigma}_{ab}).$$

(3)

Note, the luminosity definition[15] of Eq.2 has dimensions of a cross section $(1/\hat{s})$, and in Eq.3 we multiply by a scaled (dimensionless) cross section $(\hat{s}\,\hat{\sigma}_{ab})$.

We define the pdfLuminosity function to compute Eq. 2. The hadron-hadron production cross section for producing particle of mass $\sqrt{\hat{s}} = M_X$ is proportional to the luminosity times the scaled partonic cross section $\hat{s}\sigma$ as in Eq. 3. In Fig. 11 we display the differential luminosity $dL_{a\bar{a}}/dM_X^2$ for parton– anti-parton ($a\bar{a}$) combinations; this luminosity would be appropriate if we

Fig. 10 Nuclear correction ratios F_2^A/F_2^N vs. x for $Q = 10$ GeV for the nCTEQ15 PDFs over an iso-scalar target. The left plot is on a linear scale, and the right plot is a log scale. This figure is comparable to Fig. 1 of Ref. [25].

says that the total momentum fraction of the partons must sum to 100%. If any single parton flavor were not imported correctly, this cross-check would be violated; hence, this provides a powerful "sanity check" on our implementation. In Table 1 we display the partonic momentum fractions (in percent) and the total; for each PDF set the momentum sum rule checks within numerical accuracy.[13]

While Table 1 presented the momentum fraction for a single Q value (3 GeV), it is interesting to see how these values change with the energy scale. In Fig. 9 we show the momentum carried by each PDF flavor (in percent) as a function of Q in GeV. We can see the heavy quarks, $\{c, b, t\}$ enter as we cross the flavor mass thresholds. In the limit of large Q, the $\{\bar{u}, \bar{d}, \bar{s}\}$ PDFs approach each other asymptotically.

4.4 Nuclear Correction Factors

Given the PDFs, it is then trivial to build up simple calculations. In Fig. 10 we display the nuclear correction factors F_2^A/F_2^N for a variety of nuclei. Here, the F_2 structure functions are related to the PDFs via
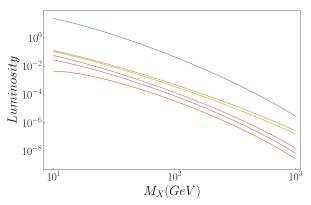
---

[13]Numerical uncertainties arise from the extrapolation down to $x \to 0$, the interpolation, and the integration precision.

[14]More specifically, F_2^N is the average of the proton and neutron $(p+n)/2$ and F_2^A is composed of Z protons, (A-Z) neutrons, and scaled by A to a make it "per nucleon:" $[Z\,p + (A - Z)\,n]/A$.
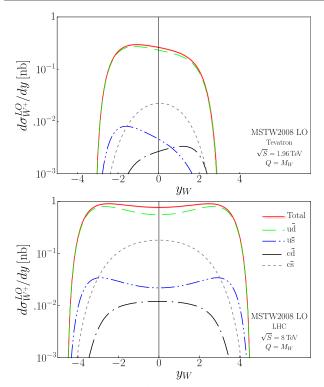
[15]There are other definitions of the luminosity in the literature which are dimensionless such as $\mathcal{L} = f_a \otimes f_b$.

Fig. 12 Leading-Order W^+ production cross section, $d\sigma/dy$ at the Tevatron ($p\bar{p}$, 1.96 TeV) and the LHC (pp, 8 TeV). We display the total cross section and the individual partonic contributions.

were interested in estimating the size of the cross section for the process of quark–anti-quark annihilation into a Higgs boson, $b\bar{b} \to H$ for example [16].

4.6 W Boson Production

Next, we compute a simple leading-order (LO) cross section for W^+ boson production at the Tevatron proton–anti-proton collider (1.96 TeV) and the LHC proton-proton collider (8 TeV). Schematically, the cross section is $\sigma(W^+) = f_a \otimes f_b \otimes \omega_{ab \to W^+}$. There are two convolution integrals, but the constraint that the partonic energies sum to the boson mass W^+ eliminates one.[15, 17] Hence, this can easily be performed inside of `Mathematica`, and the results are displayed in Fig. 12. It is interesting to note the much larger width of the rapidity distribution at the LHC as well as the increased relative contribution of the heavier quark channels (such as $c\bar{s}$ and $u\bar{s}$).
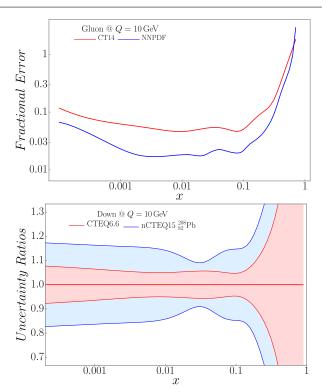
---

[16]`ManeParse` also has the capability to handle custom PDFs. This allows the user to explore a wide variety of phenomena, such as intrinsic heavy quarks, as long as the custom PDFs are written in either `LHAPDF6` or CTEQ format

Fig. 13 The fractional PDF uncertainty vs. x at $Q = 10$ GeV. a) The upper (red) curve is CT14 using the `pdfHessianError` function, and the lower (blue) curve is the NNPDF using the `pdfMCError` function for the gluon. (Note, these curves do not necessarily represent the same confidence level.) b) The down quark PDF uncertainty band for the CTEQ6.6 PDFs (inner, red) and the `nCTEQ15` lead 208 (outer, blue);

5 Error PDFs & Correlations

5.1 PDF Uncertainties

We now examine some of the added features provided by the `pdfError` module. To accommodate the PDF errors, it is common for the PDF groups to release a set of grids to characterize the uncertainties; the number of PDFs in each error is typically in the range 40 to 100, but can in principle be as many as 1000.

As `Mathematica` handles lists naturally, we can exploit this feature to manipulate the error PDFs. The `pdfFamilyParseLHA` and `pdfFamilyParseCTEQ` functions will read an entire directory of PDFs and return the associated set numbers as a list; this list can then be used to manipulate the entire group of error PDFs.

For example, we can use this feature to read the 100 PDFs of the NNPDF set displayed in Fig. 13, capture the returned list of `iSet` values, and pass this to the plotting function; we'll describe this more in the following.

When working with the error PDFs, the first step is to take the list of `iSet` values and obtain a list of the

PDF values. Constructing the PDF error depends on whether the set is based on the Hessian or the Monte Carlo method.

The Hessian PDF error sets can be organized as follows $\{X_0, X_1^+, X_1^-, X_2^+, X_2^-, ..., X_N^+, X_N^-\}$ where X_0 represents the central set, $\{X_1^+, X_1^-\}$ represent the plus and minus directions along eigenvector #1, and so on up to eigenvector N. For the Hessian PDF sets, there should be an odd number equal to $2N+1$ where N is the number of eigenvector directions. The PDF errors can then be constructed using symmetric, plus, or minus definitions:[11, 24]

$$\Delta X_{sym}^{Hess} = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left[X_i^+ - X_i^-\right]^2} \qquad (4)$$

$$\Delta X_{plus}^{Hess} = \sqrt{\sum_{i=1}^{N}\left[\max\left\{X_i^+ - X_0, X_i^- - X_0, 0\right\}\right]^2} \quad (5)$$

$$\Delta X_{minus}^{Hess} = \sqrt{\sum_{i=1}^{N}\left[\max\left\{X_0 - X_i^+, X_0 - X_i^-, 0\right\}\right]^2} \quad (6)$$

These can be computed using the function `pdfHessianError[iSet,(method)]`, and can take an optional "`method`" argument, $\{$"sym","plus","minus"$\}$, to specify which formula is used to compute the error; the default being "sym".

We next turn to the Monte Carlo sets. For example, the NNPDF set (#3 in Table 1) has 101 elements; the "zeroth" set is the central set, and the remaining 100 replica sets span the PDF uncertainty space. The central set is the average of all the sets, and the PDF error is the standard deviation of the 100 replica sets. For these sets, `pdfMCCentral` will return the central PDF value. `pdfMCError[iSet,(method)]` will return the associated error. This function can also take an optional "method" argument, $\{$"sym", "plus","minus"$\}$, defined by Eqs. 7, 8, 9.[4][13]

The modification from the Hessian case is due to the MC error PDFs using replica sets, not eigenvector pairs.[17] The formula for ΔX_{sym}^{MC} is a straightforward extension of the Hessian case:

$$\Delta X_{sym}^{MC} = \sqrt{\frac{1}{N_{rep}}\sum_{i=1}^{N}\left[X_i - X_0\right]^2} \quad . \qquad (7)$$

where N_{rep} counts the 100 replica sets not including the "zeroth" central set. This quantity is simply the standard deviation of the values. The $1/\sqrt{N_{rep}}$ factor compensates for the fact that Monte Carlo sets can have an arbitrary number of replicas, in contrast to the Hessian sets which have a fixed number of eigenvector sets.

---

[17]See the LHAPDF6 reference[10] for a more complete description of the error definitions and calculation.

It is possible to define extensions for Monte Carlo "plus" and "minus" uncertainties as:[22]

$$\Delta X_{plus}^{MC} = \sqrt{\frac{1}{N_{rep}^+}\sum_{i=1}^{N}\left[\max\left\{X_i - X_0, 0\right\}\right]^2} \qquad (8)$$

$$\Delta X_{minus}^{MC} = \sqrt{\frac{1}{N_{rep}^-}\sum_{i=1}^{N}\left[\max\left\{X_0 - X_i, 0\right\}\right]^2} \quad , \qquad (9)$$

where $N_{rep}^{\pm}$ are the number of replicas above/below the mean.

In Fig. 13-a), we compute the fractional PDF error for the CT14 PDF gluon using the `pdfHessianError` function with the "sym" formula of Eq. 4. The same is done for the NNPDF set `pdfMCError` function, using Eq. 7. As expected, we see the uncertainty increase both as $x \to 1$ and at very small x values.

In Fig. 13-b), we compute the error bands for the down quark in the CTEQ6.6 proton PDF and also the `nCTEQ15` lead-208 PDF; as expected, we see the uncertainties on the nuclear PDF are larger than the proton PDF uncertainties.

5.2 Correlation Angle

Finally, we can compute the correlation cosines via the relation: [21]

$$\cos\varphi = \frac{\overrightarrow{\nabla}X \cdot \overrightarrow{\nabla}Y}{\triangle X \triangle Y}$$

$$= \frac{1}{4\triangle X \triangle Y}\sum_{i=1}^{N}\left(X_i^+ - X_i^-\right)\left(Y_i^+ - Y_i^-\right) . \quad (10)$$

We have implemented separate functions `pdfHessianCorrelation` and `pdfMCCorrelation` as the computation of the uncertainty in the denominator $\triangle X \triangle Y$ could depend on Eqs. 4, 5, 6 or 7 8 9.

In Fig. 14 we display an example where we show the correlation cosine between the W^+ cross section and the partonic flavors for both the Tevatron and LHC. We observe the behavior of the flavors is quite similar except for the u and d quarks which stand out at large x.

The cosine of the correlation angle indicates the degree to which the error on a particular parton's PDF contributes to the uncertainty on some function of the PDFs, usually a physical observable. A value close to one for some parton indicates that the PDF error on the observable is being driven by the error on that parton's PDF. Similarly, a value close to zero indicates that the error on the parton's PDF does not contribute significantly to the error on the observable. More details can be found in Ref. [21].
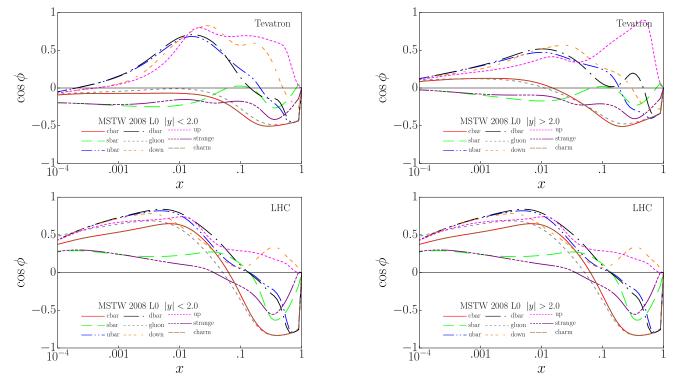
Fig. 14 The cosine of the correlation angle, $\cos\phi$, as in Eq. 10, as a function of Bjorken-x for the leading-order W^+ cross section and each of the partonic flavors. Differences between Tevatron (top) $p\bar{p}$ collisions at $\sqrt{S} = 1.96$ TeV and LHC (bottom) pp collisions at $\sqrt{S} = 8$TeV are visible in the high-x region in both the central ($|y| < 2$) and high absolute rapidity ($|y| > 2$) regions.

6 Conclusions

We have presented the `ManeParse` package which provides PDFs within the `Mathematica` framework. This is designed to work with any of the `LHAPDF6` format PDFs, and is extensible to other formats such as the CTEQ PDS format. `ManeParse` can also work with nuclear PDFs such as the nCTEQ15 sets.

The `ManeParse` package implements a number of novel features. It adapts YAML relations into `Mathematica` rules including unknown keys, and can handle discontinuities in both the PDFs and $\alpha_s(Q^2)$. We have implemented a flexible interpolation with a tuneable parameter, and it can extrapolate to small x with a variable power. Additionally, we have implemented functions to facilitate the calculation of PDF uncertainties for both Hessian and Monte Carlo PDF sets.

`ManeParse` provides many tools to simplify calculations involving PDFs, and is fast enough such that even one or two convolutions can easily be computed within the `Mathematica` framework. We illustrated these features with examples of W production, luminosity calculations, nuclear correction factors, and N_F-dependent PDFs.

In summary, the `ManeParse` package is a versatile, flexible, user-extensible tool that can be used by be-

ginning users to make simple PDF plots, as well as by advanced users investigating subtle features of higher-order discontinuities and PDF uncertainty calculations.

Appendix A: Appendices

Appendix A.1: `ManeParse` Distribution Files

The `ManeParse` package is distributed as a gzipped tar file (about 2.6Mb), and this is available at `cteq.org` or `ncteq.HepForge.org`.

When this is unpacked, the `ManeParse` modules {pdfCalc, pdfErrors, pdfParseCTEQ, pdfParseLHA} will be in the `./MP_Packages/` directory.

There is a `Demo.nb` `Mathematica` notebook which will illustrate the basic functionality of the program; we also include a `Demo.pdf` file so the user can see examples of the correct output.

We do not distribute any PDF files, so these must be obtained from the LHAPDF6 website[18] or the CTEQ website.[19] The `README` file will explain how to run the `MakeDemo.py` python script to download and set up the necessary directories for the PDF files.[20]

The `MakeDemo.py` script will also run the Perl script `noe2.perl` on the CT10 data files. Older versions of these files use a two digit exponent (e.g. 1.23456E-12), but occasionally three digits are required in which case the value is written as 1.23456-123 instead of 1.23456E-123. While the GNU compiler writes and reads this properly, other programs (including `Mathematica`) do not, so the `noe2.perl` script fixes this. This script can also be run interactively, in which case it will print out any lines that are modified.

There is a manual in both `Mathematica` format (manual_v1.nb) and PDF format (manual_v1.pdf); this allows the user to execute the notebook directly, but also see how the output should look. The manual provides examples of all the functions of `ManeParse`.

There is also a glossary file `User.pdf` which provides a list and usage of all the commands.

Appendix A.2: A Simple Example

First we define some directory paths. You should adjust for your particular machine. Note, for LHAPDF6, the individual "dat" and "info" files are stored in subdirectories.

> **pacDir="../ManeParse/Demo/packs"**
> **pdfDir="../LHAPDF"**
> **subDir1=pdfDir<>"/MSTW2008nnlo68cl"**
> **subDir2=pdfDir<>"NNPDF30_nnlo_as_0118_nf_6"**
> **ctqDir="../ManeParse/Demo/PDF_Sets/PDS"**

---

[18]http://lhapdf.hepforge.org/
[19]http://cteq.org/
[20]Python is not essential to `ManeParse` as the files can be setup manually.

Next, we load the `ManeParse` packages. The `pdfCalc` package is automatically loaded by both `pdfParseLHA` and `pdfParseCTEQ`, so we do not need to do this separately.

> **Get[pacDir<>"/pdfParseLHA.m"];**
> **Get[pacDir<>"/pdfParseCTEQ.m"];**
> **Get[pacDir<>"/pdfErrors.m"];**

`pdfParseLHA` will read the PDF set and assign an "iSet" number, which in this case is 1.

> **iSetMSTW=**
> **pdfParseLHA[**
>   **subDir1<>"/MSTW2008nnlo68cl.info",**
>   **subDir1<>"/MSTW2008nnlo68cl_0000.dat"]**
> Out[...]:=1

The "iSet" numbers are assigned sequentially, and are returned by `pdfParseLHA` which we use to define the variable iSetMSTW (=1 in this example). We can then evaluate the PDF values.

> **iParton=0; (* Gluon *)**
> **x=0.1;**
> **q=10.;**
> **pdfFunction[iSetMSTW,iParton,x,q]**
> Out[...]:=11.714

Next, we can read in an NNPDF PDF set.

> **iSetNNPDF=**
> **pdfParseLHA[**
>   **subDir2<>"/NNPDF30_nnlo_as_0118_nf_6.info",**
>   **subDir2 <>"/NNPDF30_nnlo_as_0118_nf_6_0000.dat"]**
> Out[...]:=2

We can then evaluate this PDF. We find it is similar (but not identical) to the value above.

> **pdfFunction[iSetNNPDF,iParton,x,q]**
> Out[...]:=11.8288

Finally, we load a `ctq66` PDF file in the older "pds" format using the `pdfParseCTEQ` function; note this only takes a single file as the "info" details are contained in the "pds" file header.

> **iSetC66=pdfParseCTEQ[**
>   **ctqDir<>"/ctq66.00.pds"];**
> Out[...]:= 3

> **pdfFunction[iSetC66,iParton,x,q]**
> Out[...]:= 11.0883

Now that we have these functions defined inside of Mathematica, we can make use of all the numerical and graphical functions. Detailed working examples are provided in the auxiliary files.
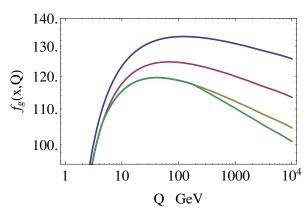
Fig. 15 We display the gluon PDF $f_g(x, Q)$ at $x = 0.03$ vs. Q for $N_F = \{3, 4, 5, 6\}$; $N_F = 3$ is the largest, and $N_F = 6$ is the smallest curve.

Appendix A.3: N_F-Dependent PDF Example

We provide an example of implementing the N_F-dependent PDFs within the `ManeParse` framework using the matched set of PDFs[21] with $N_F = \{3, 4, 5, 6\}$ from Ref. [16]. We load the `ManeParse` packages as above, and then read in the grid files which are in "pds" format.

```
pdfDir="../vfnsnf";
iSetNF3=pdfParseCTEQ[pdfDir<>"/nf3_q1.2.pds"]
iSetNF4=pdfParseCTEQ[pdfDir<>"/nf4_q1.2.pds"]
iSetNF5=pdfParseCTEQ[pdfDir<>"/nf5_q1.2.pds"]
iSetNF6=pdfParseCTEQ[pdfDir<>"/nf6_q1.2.pds"]
```

`pdfParseCTEQ` returns the "iSet" number and we store these in {iSetNF3, ...}. The below function `pdfNF` allows the user to choose N_F, and then returns the appropriate PDF.

```
Clear[pdfNF,nf,iParton,x,q];
pdfNF[nf_,iParton_,x_,q_]:=Module[{iSet=0},
    If[nf==3,iSet=iSetNF3];
    If[nf==4,iSet=iSetNF4];
    If[nf==5,iSet=iSetNF5];
    If[nf==6,iSet=iSetNF6];
    If[iSet==0,Return[Null]];
    Return[pdfFunction[iSet,iParton,x,q]]
]
```

The the `pdfNF` function, the "iSet" variable is local to the `Module`. We now compute some sample values.

```
iParton=0; (* Gluon *)
x=0.03;
q=10.;
```

```
{pdfNF[3,iParton,x,q], pdfNF[4,iParton,x,q],
 pdfNF[5,iParton,x,q], pdfNF[6,iParton,x,q]}
Out[...]:={123.288, 117.694, 115.331, 115.341}
```

As we have taken $Q = 10$ GeV, we are above the charm and bottom transition, but below the top transition; hence the $N_F = \{5, 6\}$ results are the same, but the $N_F = \{3, 4\}$ values differ.

In Fig. 15 we display the gluon PDF vs. Q for $N_F = \{3, 4, 5, 6\}$. We observe as we activate more flavors in the PDF evolution the gluon is reduced as a function of N_F. This decrease in the gluon PDF will be (partially) compensated by the new N_F channels.

---

[21] These PDF sets are available at `http://ncteq.hepforge.org/`.

References

1. H. Abramowicz et al. Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. *Eur. Phys. J.*, C75(12):580, 2015.
2. A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato. Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data. 2016.
3. S. Alekhin, J. Blumlein, and S. Moch. The ABM parton distributions tuned to LHC data. *Phys. Rev.*, D89(5):054028, 2014.
4. S. Alekhin et al. The PDF4LHC Working Group Interim Report. 2011.
5. C. Alexandrou. Parton distribution functions from Lattice QCD. In *Theory and Experiment for Hadrons on the Light-Front (Light Cone 2015) Frascati , Italy, September 21-25, 2015*, 2016.
6. C. A. Argüelles, F. Halzen, L. Wille, M. Kroll, and M. H. Reno. High-energy behavior of photon, neutrino, and proton cross sections. *Phys. Rev.*, D92(7):074040, 2015.
7. R. D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.
8. V. Bertone, S. Carrazza, and J. Rojo. APFEL: A PDF Evolution Library with QED corrections. *Comput. Phys. Commun.*, 185:1647–1668, 2014.
9. A. Bhattacharya, R. Enberg, M. H. Reno, I. Sarcevic, and A. Stasto. Perturbative charm production and the prompt atmospheric neutrino flux in light of RHIC and LHC. *JHEP*, 06:110, 2015.
10. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt. LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J.*, C75:132, 2015.
11. J. M. Campbell, J. W. Huston, and W. J. Stirling. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept. Prog. Phys.*, 70:89, 2007.
12. S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev.*, D93(3):033006, 2016.
13. J. Gao and P. Nadolsky. A meta-analysis of parton distribution functions. *JHEP*, 07:035, 2014.
14. K. Kovarik et al. nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework. *Phys. Rev.*, D93(8):085037, 2016.
15. K. Kovarik, T. Stavreva, A. Kusina, T. Jezo, F. I. Olness, I. Schienbein, and J. Y. Yu. A Survey of Heavy Quark Theory for PDF Analyses. *Nucl. Phys. Proc. Suppl.*, 222-224:52–60, 2012.
16. A. Kusina, F. I. Olness, I. Schienbein, T. Jezo, K. Kovarik, T. Stavreva, and J. Y. Yu. Hybrid scheme for heavy flavors: Merging the fixed flavor number scheme and variable flavor number scheme. *Phys. Rev.*, D88(7):074032, 2013.
17. A. Kusina, T. Stavreva, S. Berge, F. I. Olness, I. Schienbein, K. Kovarik, T. Jezo, J. Y. Yu, and K. Park. Strange Quark PDFs and Implications for Drell-Yan Boson Production at the LHC. *Phys. Rev.*, D85:094028, 2012.
18. H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev.*, D82:074024, 2010.
19. Y.-Q. Ma and J.-W. Qiu. QCD Factorization and PDFs from Lattice QCD Calculation. *Int. J. Mod. Phys. Conf. Ser.*, 37:1560041, 2015.
20. A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J.*, C63:189–285, 2009.
21. P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan. Implications of CTEQ global analysis for collider observables. *Phys. Rev.*, D78:013004, 2008.
22. P. M. Nadolsky and Z. Sullivan. PDF uncertainties in WH production at Tevatron. *eConf*, C010630:P510, 2001.
23. K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.
24. J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.
25. I. Schienbein, J. Y. Yu, K. Kovarik, C. Keppel, J. G. Morfin, F. Olness, and J. F. Owens. PDF Nuclear Corrections for Charged and Neutral Current Processes. *Phys. Rev.*, D80:094004, 2009.