How to run ePump

Yao Fu

Preparation for running ePump

➢ Input file

xxx.in

To specify how many data files we have, the weight of each data sets, the number of eigenvector, tolerance setting, etc…

Data file

xxx.data

This file puts all the information of data set, which include the data point, statistical uncertainty, uncorrelated systematic uncertainty, correlated systematic uncertainty, etc…

Theory file

xxx.theory

This file puts the theoretical prediction corresponding to our data set, which includes all the prediction for each PDF eigenvector.

The PDF grid files are also needed. (Either CTEQ or LHAPDF format is OK.)

Data table

➢ Data Type 1 (using the A_{FB} spectrum as an example)

Once we have data with statistical uncertainty, we have to convert our data into a certain format.

```
root [8] for(int i=1;i<=AFB_CF->GetNbinsX();i++){cout<<fixed<<setprecision(5)<<setw(8)<<AFB_CF->GetBinContent(i)<<" +- "<<AFB_CF->GetBinError(i)<<endl;}
-0.13677 +- 0.00014
-0.22259 +- 0.00017
-0.00138 +- 0.00007
 0.06824 +- 0.00006
 0.26173 +- 0.00024
 0.37121 +- 0.00032
 0.44523 +- 0.00044
 0.49261 +- 0.00074
 0.54854 +- 0.00122
 0.53148 +- 0.00231
 0.60488 +- 0.00419
 0.64189 +- 0.00811
```

Line starts with " * " is comment line. We can put anything whatever we want to write. No limit for the number of lines.

```
* 2 NormError #_corr_sys     Ecm       |# grids|# KF|# col|
*  0.0          0         13000.0          888231
# of corr_err   ,  Data Column,    StatErr Column,  UncSys Column,   corr_err Column
      2                  3              4                6                7
left right DataPoint StatErr TotSys UncSys  Corr1 Corr2
40  66   0.13677  0.000137037  0.000137037  0  0  0
66  80  -0.222589  0.000172035  0.000172035  0  0  0
80  91  -0.00137505  6.56798e-05  6.56798e-05  0  0  0
91  102  0.0682376  5.80559e-05  5.80559e-05  0  0  0
102  116  0.261725  0.000243483  0.000243483  0  0  0
116  145  0.371207  0.000320109  0.000320109  0  0  0
145  200  0.445225  0.000436122  0.000436122  0  0  0
200  275  0.492607  0.000737345  0.000737345  0  0  0
275  381  0.548535  0.0012181  0.0012181  0  0  0
381  525  0.531484  0.00231314  0.00231314  0  0  0
525  725  0.604882  0.00418944  0.00418944  0  0  0
725  1000  0.641893  0.00810986  0.00810986  0  0  0
```

From left to right:
The number of correlated systematic error
Data column
Statistical error column (absolute value)
Uncorrelated systematic column (absolute value)
Correlated systematic column starts with (percentage value)

Note that there is additional line which is for explaining what each column means.

(In this example, it means we have 2 sources of correlated systematic error, they start from the column 7. The data point is in the third column, the statistical error is in the fourth column, the uncorrelated systematic error is in the sixth column. Here, the fifth column fills the "Total uncertainty", but ePump doesn't read this line. The first and second column fill the binning information, ePump also doesn't read those lines. They are all additional information that helps people know the information of this data set.)

Data table

➢ Data Type 2

Same as Data Type 1

Line starts with " * " is comment line.
The end line should be

#a #b CorrCoeff

Then ePump will start to read the correlation coefficient.

```
*: table_doi: 10.17182/hepdata.71419.v1/t1
*: name: Table 1
*: descrip systemati beam and luminosity.
*: data_file: Table1.yaml
*: keyword reactions: P P --> W+ X | P P --> W- X
*: keyword observables: DSIG/DETARAP
*: keyword phrases: Inclusive | Single Differential Cross Section | Proton-Proton Scattering | W Production | Muon production
*: keyword cmenergies: 8000.0
*: RE                        P P --> W- < MU- NUBAR > X
*: PT(MU) [GEV]              > 20
*: ETARAP(MU)                2.0-4.5
*: SQRT(S) [GeV]                  8000
*  Note that correlated beam and lumi errors are absolute (not percent of xsec)
# of corr_err   ,  Data Column,    StatErr Column,  UncSys Column,   corr_err Column
    2                4                5                6                7
       #    eta_low   eta_high     xsec      stat       sys      beam      lumi
       1      2.00      2.25     134.0       0.9       1.8       1.2       1.6
       2      2.25      2.50     119.8       0.7       1.4       1.0       1.4
       3      2.50      2.75     110.6       0.6       1.2       1.0       1.3
       4      2.75      3.00     102.4       0.6       1.2       0.9       1.2
       5      3.00      3.25      92.5       0.6       1.1       0.8       1.1
       6      3.25      3.50      79.9       0.5       0.9       0.7       0.9
       7      3.50      4.00     119.3       0.6       1.5       1.0       1.4
       8      4.00      4.50      60.0       0.7       1.6       0.5       0.7
*: table_doi: 10.17182/hepdata.71419.v1/t13
*: name: Table 13
*: descrip which are are excluded. This table lists bin-to-bin correlations for $W^{-}$.
*: data_file: Table13.yaml
*: keyword reactions: P P --> W- X
*: keyword observables: DSIG/DETARAP | CORR
*: keyword phrases: Inclusive | Single Differential Cross Section | Proton-Proton Scattering | W Production
*: keyword cmenergies: 8000.0
*: PT(MU) [GEV]                                  > 20
*: ETARAP(MU)                                    2.0-4.5
*: RE(-)                                         P P --> W- < MU- NUBAR > X
*: SQRT(S) [GeV]                                       8000
       #a       #b    CorrCoeff
       1        1         1.00
       2        1         0.21
       3        1         0.18
       4        1         0.21
       5        1         0.22
       6        1         0.23
       7        1        -0.06
       8        1        -0.09
       2        2         1.00
       3        2         0.22
       4        2         0.25
       5        2         0.28
```

Correlation coefficient, given by experiment group.

Data table

Comment lines

- **Data Type 3**

```
* 2 NormError #_corr_sys     Ecm       |# grids|# KF|# col|
*  0.0           9          1960.0           888231
Experimental Data Values
0.111763
0.0551275
*
* More comments here
*
Inverse Covariance Matrix:
a=1, b=1,2
8.95999e+04   5.75107e+04
a=2, b=2,2
4.33579e+04
```

Data point.

There is no statistical error and systematic error. All the error information are in the covariance matrix

Inverse covariance matrix.

Input the inverse covariance matrix. Note that the covariance matrix should include both statistical uncertainty, uncorrelated systematic uncertainty and correlated systematic uncertainties.

For the statistical uncertainty and uncorrelated systematic uncertainty, the correlation coefficient is unit matrix. There is only square of uncertainty in the diagonal term. Off-diagonal terms are zero. Then we should add all the covariance matrix together, and calculate the inverse covariance matrix.

$$\mathrm{cov}_{ij} = \delta_{ij}\Delta_i^{\mathrm{stat}}\Delta_j^{\mathrm{stat}} + \mathrm{Corr}_{ij}\Delta_i^{\mathrm{sys}}\Delta_j^{\mathrm{sys}}$$

Theory table

➢ Theory template (using the A_{FB} spectrum as an example)

We also should have the theory prediction for every data set. The theory prediction should include the PDF central set and all the PDF error sets.

```
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.00/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.01/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.02/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.03/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.04/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.05/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.06/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.07/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.08/MC_results.root
/msu/data/t3work5/yfu/MainCode/run/ResBosHighMassZ/i2Tn3.09/MC_results.root
```

For example, CT18NNLO has one central set + 58 error sets.

…

…

```
Theory Column
      1
 Data : i2Tn3.00.dta
-0.13677
-0.222589
-0.00137505
0.0682376
0.261725
0.371207
0.445225
0.492607
0.548535
0.531484
0.604882
0.641893
 Data : i2Tn3.01.dta
-0.137781
-0.22463
-0.00136634
0.0689802
0.264281
0.374426
0.448354
0.494891
0.550563
0.533074
0.606429
```

This number tells ePump which column is the needed information. Here there is only one column. In fact, ePump only need one column to read. We can also write some other information to other columns, like the binning information. Then we should just tell ePump which column should be read.

Splitting line for separating each prediction of different error PDF

Theory prediction for each PDF set.

The theory prediction must be matched with the Data file. They should have the same bin number. There is no error input for the ePump. The number of theory prediction should also match to the number of EV set, which will be introduced in how to write the input card file.

Input card

```
+++ N(EV pairs)                    N(Data Sets)   PDFtype(C/L/N)   Dyn_Tol?(Y/N)   Tol_squared
        29                               16              C               Y             100.0
+++ ObservableFile                 N(Observables)  Data?(Y/N)      Error type      Weight          PS:
test/Xsec_2D_CC                          144           Y               1             0.0
test/Xsec_2D_CF                           52           Y               1             0.0
test/AFB_Mass_ZY_CC                      144           Y               1             0.0
test/AFB_Mass_ZY_CF                       52           Y               1             0.0
test/PositiveLeptonEta                    20           Y               1             0.0
test/NegativeLeptonEta                    20           Y               1             0.0
test/PositiveLeptonPt                     40           Y               1             0.0
test/NegativeLeptonPt                     40           Y               1             0.0
test/PositiveLeptonPtEta                 100           Y               1             1.0
test/NegativeLeptonPtEta                 100           Y               1             1.0
test/WPlusMT                              40           Y               1             0.0
test/WMinusMT                             40           Y               1             0.0
test/WLeptonEtaAsymmetry                  20           Y               1             0.0
test/WHighMassMT                          25           Y               1             0.0
test/ZHighMass_CC                         25           Y               1             0.0
test/ZHighMass_CF                         25           Y               1             0.0
+++      PDFname                      PDFout
PDFs/i2Tn3.58/i2Tn3                outPDFs/i2Tn3



# .in for ePump_v20171221 and after
```

Weight of each data set.

If it is set to be 0.0, it has the same function as setting the "Data?" to "N", but ePump will output the values after PDF updating for this data set. If we want to turn off some data set, setting weight to zero is better.

Error Type

Name of date/theory file. For example, there should be two files called test/Xsec_2D_CC.data and test/Xsec_2D_CC.theory

This number of data points for a certain file. It should match to the real number of data points of your .data file, otherwise the results will be wrong and no any warning information output.

IsData.
If it is set to N, it will not read the .data file. But meanwhile, ePump will also not output the values after PDF updating

Input card

Number of eigenvector. For example, CT18NNLO has 0+58 PDF set, NEV is 29

Number of data sets. It should match to how many data sets written below

Dynamical tolerance: If set it to N, it will read the Tol_squared, which is the delta chi squared. If set it to Y, it will read the .tol file.

```
+++ N(EV pairs)                 N(Data Sets)  PDFtype(C/L/N)  Dyn_Tol?(Y/N)  Tol_squared
        29                          16              C               Y            100.0
+++ ObservableFile              N(Observables)  Data?(Y/N)      Error_type     Weight          PS:
test/Xsec_2D_CC                      144            Y              1            0.0
test/Xsec_2D_CF                       52            Y              1            0.0
test/AFB_Mass_ZY_CC                  144            Y              1            0.0
test/AFB_Mass_ZY_CF                   52            Y              1            0.0
test/PositiveLeptonEta                20            Y              1            0.0
test/NegativeLeptonEta                20            Y              1            0.0
test/PositiveLeptonPt                 40            Y              1            0.0
test/NegativeLeptonPt                 40            Y              1            0.0
test/PositiveLeptonPtEta             100            Y              1            1.0
test/NegativeLeptonPtEta             100            Y              1            1.0
test/WPlusMT                          40            Y              1            0.0
test/WMinusMT                         40            Y              1            0.0
test/WLeptonEtaAsymmetry              20            Y              1            0.0
test/WHighMassMT                      25            Y              1            0.0
test/ZHighMass_CC                     25            Y              1            0.0
test/ZHighMass_CF                     25            Y              1            0.0
+++      PDFname                  PDFout
PDFs/i2Tn3.58/i2Tn3              outPDFs/i2Tn3


# .in for ePump_v20171221 and after
```

PDF type input/output.
C: CTEQ format,
L: LHAPDF format,
N: no PDF files output

The path to the input PDF files and output directory. Here we should create the "outPDFs" by ourselves, ePump will not help us create.
Note that there should not be an "enter" between data file lines and PDF file lines, otherwise the program will interrupt.

Running

➢ File structure

```
[yfu@pandeia Training]$ ls
outPDFs  PDFs  test  test.in  UpdatePDFs
[yfu@pandeia Training]$ ls test
NegativeLeptonPtEta.data    PositiveLeptonPtEta.data    WHighMassMT.data    WLeptonEtaAsymmetry.data    ZHighMass_CC.data    ZHighMass_CF.data
NegativeLeptonPtEta.theory  PositiveLeptonPtEta.theory  WHighMassMT.theory  WLeptonEtaAsymmetry.theory  ZHighMass_CC.theory  ZHighMass_CF.theory
[yfu@pandeia Training]$ ls PDFs/i2Tn3.58/
i2Tn3.00.pds  i2Tn3.06.pds  i2Tn3.12.pds  i2Tn3.18.pds  i2Tn3.24.pds  i2Tn3.30.pds  i2Tn3.36.pds  i2Tn3.42.pds  i2Tn3.48.pds  i2Tn3.54.pds
i2Tn3.01.pds  i2Tn3.07.pds  i2Tn3.13.pds  i2Tn3.19.pds  i2Tn3.25.pds  i2Tn3.31.pds  i2Tn3.37.pds  i2Tn3.43.pds  i2Tn3.49.pds  i2Tn3.55.pds
i2Tn3.02.pds  i2Tn3.08.pds  i2Tn3.14.pds  i2Tn3.20.pds  i2Tn3.26.pds  i2Tn3.32.pds  i2Tn3.38.pds  i2Tn3.44.pds  i2Tn3.50.pds  i2Tn3.56.pds
i2Tn3.03.pds  i2Tn3.09.pds  i2Tn3.15.pds  i2Tn3.21.pds  i2Tn3.27.pds  i2Tn3.33.pds  i2Tn3.39.pds  i2Tn3.45.pds  i2Tn3.51.pds  i2Tn3.57.pds
i2Tn3.04.pds  i2Tn3.10.pds  i2Tn3.16.pds  i2Tn3.22.pds  i2Tn3.28.pds  i2Tn3.34.pds  i2Tn3.40.pds  i2Tn3.46.pds  i2Tn3.52.pds  i2Tn3.58.pds
i2Tn3.05.pds  i2Tn3.11.pds  i2Tn3.17.pds  i2Tn3.23.pds  i2Tn3.29.pds  i2Tn3.35.pds  i2Tn3.41.pds  i2Tn3.47.pds  i2Tn3.53.pds  i2Tn3.tol
```

./UpdatePDFs test

Or

./UpdatePDFs test [global weight]

```
+++ N(EV pairs)                    N(Data Sets)   PDFtype(C/L/N)   Dyn_Tol?(Y/N)  Tol_squared
        29                               6              C              Y             100.0
+++ ObservableFile                 N(Observables)  Data?(Y/N)      Error_type     Weight          PS:
test/PositiveLeptonPtEta                100          Y               1            1.0
test/NegativeLeptonPtEta                100          Y               1            1.0
test/WLeptonEtaAsymmetry                 20          Y               1            0.0
test/WHighMassMT                         25          Y               1            0.0
test/ZHighMass_CC                        25          Y               1            0.0
test/ZHighMass_CF                        25          Y               1            0.0
+++     PDFname                    PDFout
PDFs/i2Tn3.58/i2Tn3               outPDFs/i2Tn3



# .in for ePump_v20171221 and after
```

Example: Take neutral-current high-mass Drell-Yan, use data from below 1 TeV to update the uncertainty at high mass

Pseudodata: Drell-Yan kinematics below 1TeV

Target observable: High mass Drell-Yan

```
+++ N(EV pairs)                    N(Data Sets)    PDFtype(C/L/N)   Dyn_Tol?(Y/N)  Tol_squared
        29                               11               C              Y            100.0
+++ ObservableFile                 N(Observables)  Data?(Y/N)      Error_type     Weight          PS:
test/Xsec_2D_CC                          144           Y               1             1.0
test/Xsec_2D_CF                           52           Y               1             1.0
test/AFB_Mass_ZY_CC                      144           Y               1             1.0
test/AFB_Mass_ZY_CF                       52           Y               1             1.0
test/PositiveLeptonEta                    20           Y               1             0.0
test/NegativeLeptonEta                    20           Y               1             0.0
test/WLeptonEtaAsymmetry                  20           Y               1             0.0
test/ZHighMass_CC                         25           Y               1             0.0
test/ZHighMass_CF                         25           Y               1             0.0
test/ZHighMass_utype                      25           Y               1             0.0
test/ZHighMass_dtype                      25           Y               1             0.0
+++     PDFname                    PDFout
PDFs/i2Tn3.58/i2Tn3                outPDFs/i2Tn3



# .in for ePump_v20171221 and after
```

Pseudodata used for updating PDF:
Neutral current 2D Xsec below 1TeV
AFB vs Mass, ZY below 1TeV

Observable:
Neutral current high-mass Drell-Yan, frow 1TeV to 5TeV

How to read output file

test.out

Before update
After update

For Observables X[a] in Data Set: test/ZHighMass_CF

	a	X[a]	DXsym[a]	DXasym[a]	
Old	1	5.7871E+02	2.9156E+01	-3.4919E+01	+2.4594E+01
New	1	5.7871E+02	7.5069E+00	-9.8213E+00	+6.5443E+00
Old	2	4.1307E+02	2.1042E+01	-2.4529E+01	+1.8415E+01
New	2	4.1307E+02	5.5090E+00	-7.1928E+00	+4.9757E+00
Old	3	2.9330E+02	1.5416E+01	-1.7851E+01	+1.3638E+01
New	3	2.9330E+02	4.0937E+00	-5.3376E+00	+3.7384E+00

Theory prediction

$$\Delta X = \frac{1}{2}\left(\sum_{i=1}^{N_p}\left[X(\{z_i^+\}) - X(\{z_i^-\})\right]^2\right)^{1/2}$$

$$\delta^+ X = \sqrt{\sum_{i=1}^{N_a}\left[\max\left(X_i^{(+)} - X_0, X_i^{(-)} - X_0, 0\right)\right]^2},$$

$$\delta^- X = \sqrt{\sum_{i=1}^{N_a}\left[\max\left(X_0 - X_i^{(+)}, X_0 - X_i^{(-)}, 0\right)\right]^2},$$

Correlation cosine

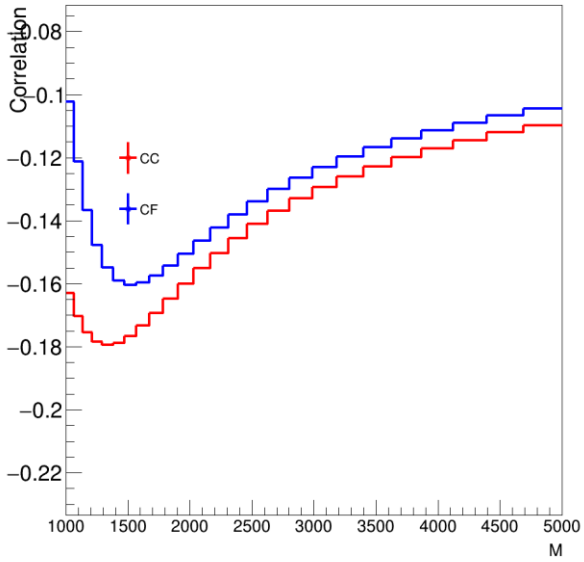
test.out

```
Original (Old) matrix of Correlation Cosines between X[a] (test/Xsec_2D_CC) and Y[b] (test/ZHighMass_CF)

        Y[01]     Y[02]     Y[03]     Y[04]     Y[05]     Y[06]     Y[07]     Y[08]     Y[09]     Y[10]     Y[11]     Y[12]     Y[13]     Y[14]     Y[15]
  Y[16]     Y[17]     Y[18]     Y[19]     Y[20]     Y[21]     Y[22]     Y[23]     Y[24]     Y[25]
X[01]   0.44799   0.40644   0.37760   0.35740   0.33341   0.31208   0.29510   0.27248   0.25994   0.24030   0.22482   0.21510   0.21775   0.19525   0.19206
 0.17979   0.17722   0.17607   0.16691   0.15943   0.15093   0.13713   0.13173   0.12721   0.10288
X[02]   0.45121   0.40989   0.38120   0.36105   0.33702   0.31565   0.29861   0.27559   0.26290   0.24271   0.22697   0.21704   0.21967   0.19689   0.19360
 0.18137   0.17889   0.17779   0.16882   0.16143   0.15322   0.13927   0.13392   0.12916   0.10463
X[03]   0.46504   0.42306   0.39373   0.37295   0.34814   0.32592   0.30824   0.28395   0.27059   0.24902   0.23258   0.22208   0.22431   0.20134   0.19785
 0.18588   0.18363   0.18266   0.17428   0.16714   0.15956   0.14543   0.14022   0.13501   0.10999
X[04]   0.48887   0.44547   0.41475   0.39267   0.36644   0.34259   0.32365   0.29734   0.28281   0.25911   0.24154   0.23010   0.23144   0.20847   0.20468
 0.19316   0.19124   0.19042   0.18303   0.17634   0.16975   0.15552   0.15056   0.14481   0.11916
X[05]   0.50954   0.46601   0.43481   0.41210   0.38504   0.36008   0.34026   0.31205   0.29648   0.27046   0.25163   0.23908   0.23951   0.21611   0.21177
 0.20065   0.19894   0.19825   0.19201   0.18591   0.18065   0.16652   0.16202   0.15581   0.12998
X[06]   0.52646   0.48345   0.45225   0.42922   0.40167   0.37586   0.35535   0.32540   0.30886   0.28054   0.26043   0.24666   0.24605   0.22216   0.21715
 0.20648   0.20490   0.20438   0.19953   0.19423   0.19062   0.17703   0.17323   0.16689   0.14138
```

$$\cos\phi = \frac{X_i' Y_i'}{\Delta X \Delta Y} = \frac{(X_i^+ - X_i^-)(Y_i^+ - Y_i^-)}{4\Delta X \Delta Y}$$

Corr(A_{FB}, high mass DY)



Root format output

test.root

```
KEY: TH1D     ZHighMass_CC_old;1      ZHighMass_CC_old
KEY: TH1D     ZHighMass_CC_new;1      ZHighMass_CC_new
KEY: TH1D     ZHighMass_CC_oldPDFError;1      ZHighMass_CC_oldPDFError
KEY: TH1D     ZHighMass_CC_newPDFError;1      ZHighMass_CC_newPDFError
KEY: TH1D     ZHighMass_CF_old;1      ZHighMass_CF_old
KEY: TH1D     ZHighMass_CF_new;1      ZHighMass_CF_new
KEY: TH1D     ZHighMass_CF_oldPDFError;1      ZHighMass_CF_oldPDFError
KEY: TH1D     ZHighMass_CF_newPDFError;1      ZHighMass_CF_newPDFError
```

Theory prediction and PDF error before and after update

```
KEY: TH2D     Correlation_old_1_1;1   Correlation_old_1_1
KEY: TH2D     Correlation_new_1_1;1   Correlation_new_1_1
KEY: TH2D     Correlation_old_1_2;1   Correlation_old_1_2
KEY: TH2D     Correlation_new_1_2;1   Correlation_new_1_2
KEY: TH2D     Correlation_old_1_3;1   Correlation_old_1_3
KEY: TH2D     Correlation_new_1_3;1   Correlation_new_1_3
KEY: TH2D     Correlation_old_1_4;1   Correlation_old_1_4
KEY: TH2D     Correlation_new_1_4;1   Correlation_new_1_4
```

Correlation cosine

How to draw PDF plots

PDFs after update are saving in the directory “outPDFs”. We can use CPP package which is developed by Tie-Jiun Hou to draw the PDF plots.

```
[yfu@pandeia CPP-v2019.1021.2001.+0800]$ ls
bin  CT18NNLO  doc  genLatex.sh  Makefile  run.sh  setup.sh  src  updated-by-Xsec-2D-AFB-2D
```

Renamed by “outPDFs”

./bin/cpp pdf 20s68all 1.3 100 ches90 CT18NNLO ches90 updated-by-Xsec-2D-AFB-2D

20s68all: Compare two PDFs

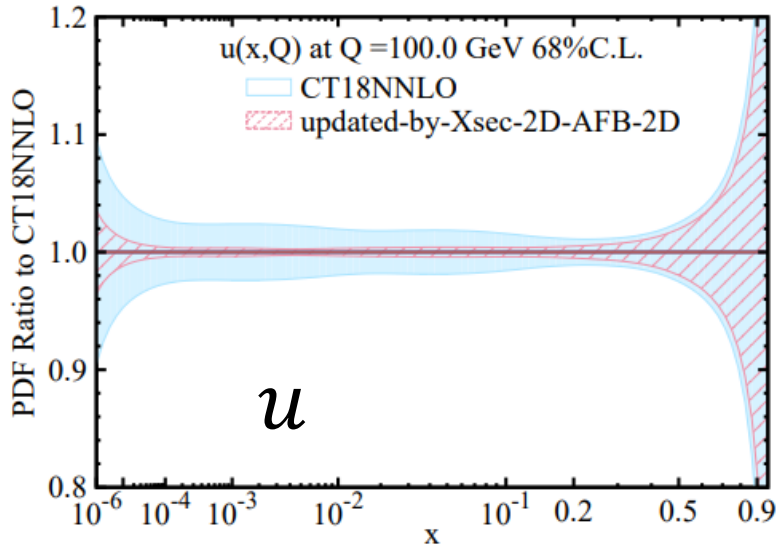
1.3 100: Two Q values

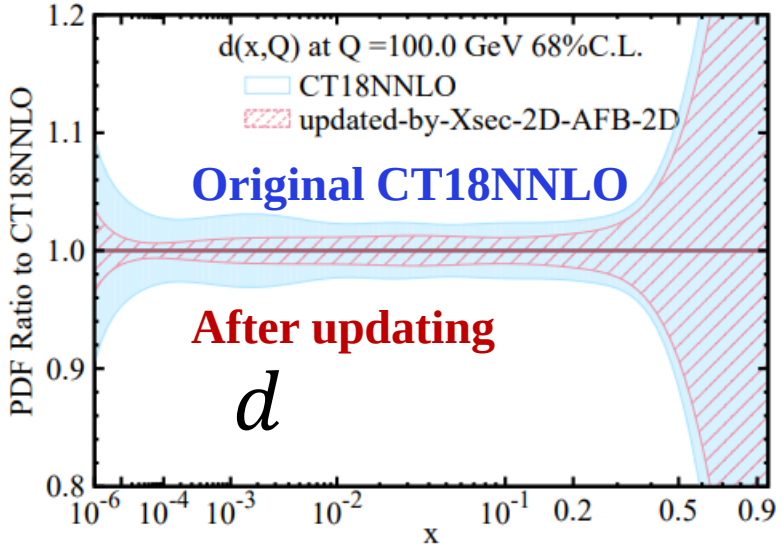
ches90:
c: CTEQ format PDFs
h: Hessian set
e: Error
s: Symmetry
90: C.L.

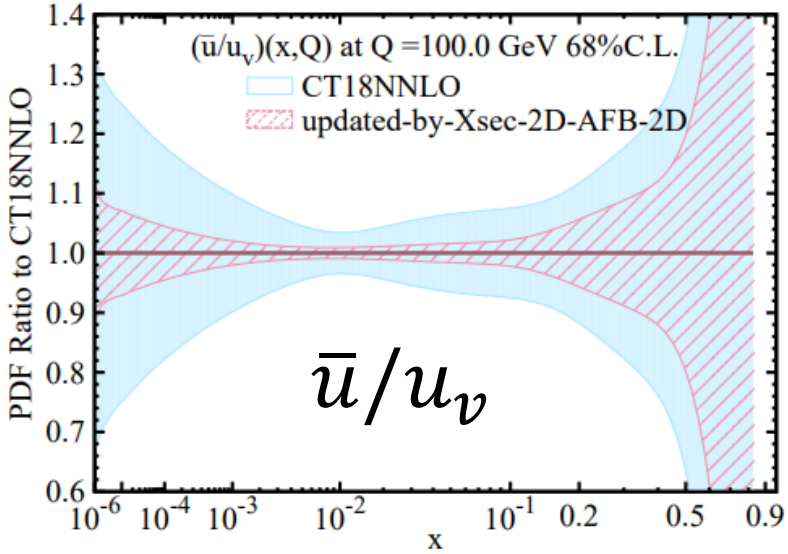
More information are written in doc/0pdf

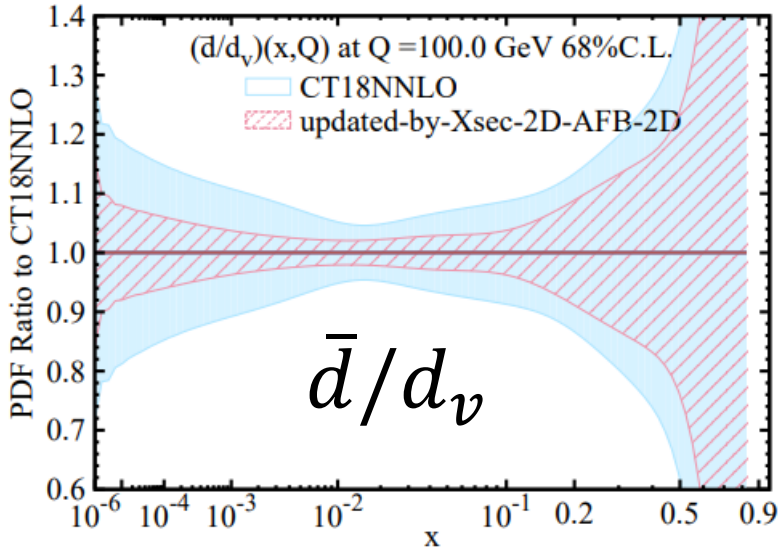
PDF plots

PDFs-20a_S68CL_1.3-100.0GeV_CT18NNLO_updated-by-Xsec-2D-AFB-2D_plots









ePump-optimization

Before study which experimental observable is sensitive to the high mass DY, we can use the high mass DY theory template to perform ePump-optimization, to see which flavor contributes most important information on the high mass DY.

```
[yfu@pandeia Rotation]$ ls
OptimizePDFs  outPDFs  PDFs  test  test.in
[yfu@pandeia Rotation]$ ls test
ZHighMass_CF.theory
[yfu@pandeia Rotation]$
```

Only .theory file is needed.

./OptimizePDFs test

```
-------------------------------------------------------------------------------
 Fractional contribution of Optimized EV i (column) to variance of observable a (row)
-------------------------------------------------------------------------------

 DataSet Obs        EV01        EV02        EV03        EV04        EV05        EV06        EV07        EV08        EV09        EV10        EV11        EV12
      EV13        EV14        EV15        EV16        EV17        EV18        EV19        EV20        EV21        EV22        EV23        EV24        EV25
      EV26        EV27        EV28        EV29
    1     1     0.71782344  0.16123257  0.09868187  0.02063947  0.00127702  0.00023380  0.00000184  0.00005924  0.00004235  0.00000174  0.00000038  0.0000000
2  0.00000142  0.00000266  0.00000142  0.00000048  0.00000010  0.00000017  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
  0.00000000  0.00000000  0.00000000  0.00000000
    1     2     0.76880343  0.15628819  0.06824580  0.00647763  0.00014921  0.00000618  0.00000563  0.00000842  0.00000256  0.00000034  0.00000000  0.0000000
8  0.00000384  0.00000096  0.00000502  0.00000151  0.00000006  0.00000053  0.00000021  0.00000020  0.00000009  0.00000001  0.00000003  0.00000005  0.00000000
  0.00000000  0.00000000  0.00000000  0.00000000
```

How many PDF uncertainty that eigenvectors after ePump-optimization contribute for each bin.

```
Totals          21.13867651  3.00474840  0.73344090
```

Final PDF uncertainty fraction
e.g.: EV01: 21.13/25 (# of bins)

ePump-optimization

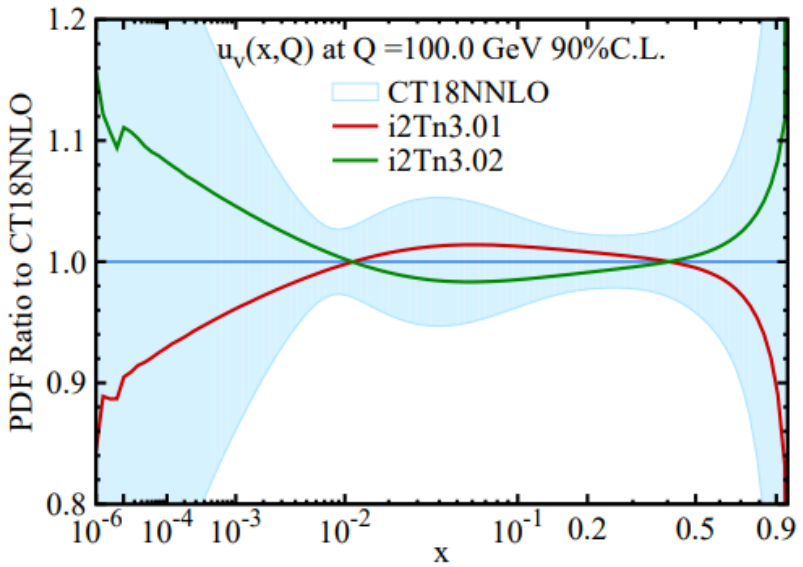
First eigenvector contributes on each flavor combination.

./bin/cpp pdf 21s90all 1.3 100 ches90 CT18NNLO c rotated-by-ZHighMass_CF_Eigen1

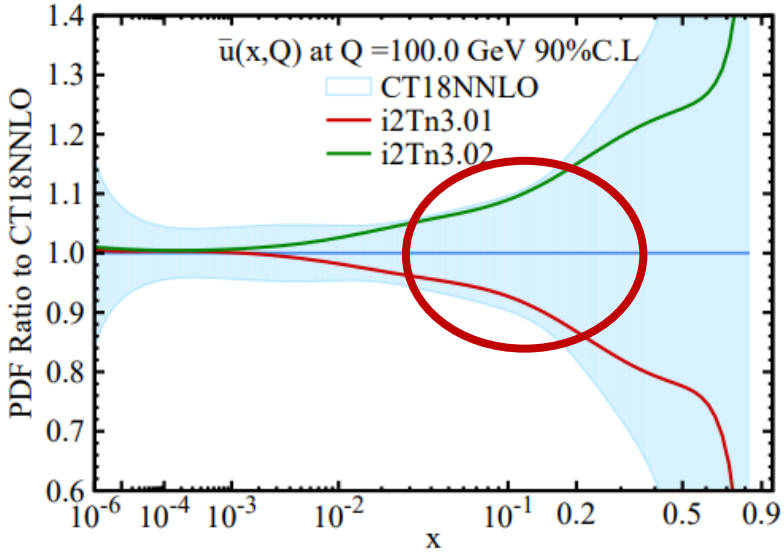
```
[yfu@pandeia CPP-v2019.1021.2001.+0800]$ ls rotated-by-ZHighMass_CF_Eigen1/
i2Tn3.01.pds  i2Tn3.02.pds
```

Results after ePump-optimization.
Put the first eigenvector into this directory

Less sensitive to u valence



Sensitive to ubar



We should find observables which are also sensitive to ubar in large x region to constrain the PDF uncertainty of high mass DY.

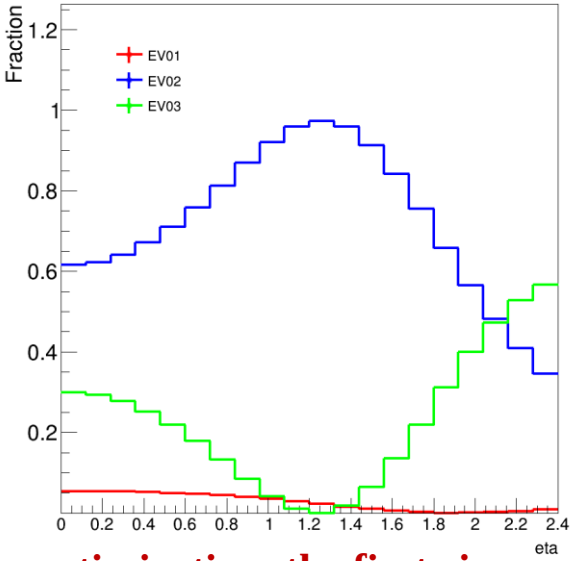
Correlation between two different observables

An application of ePump-optimization

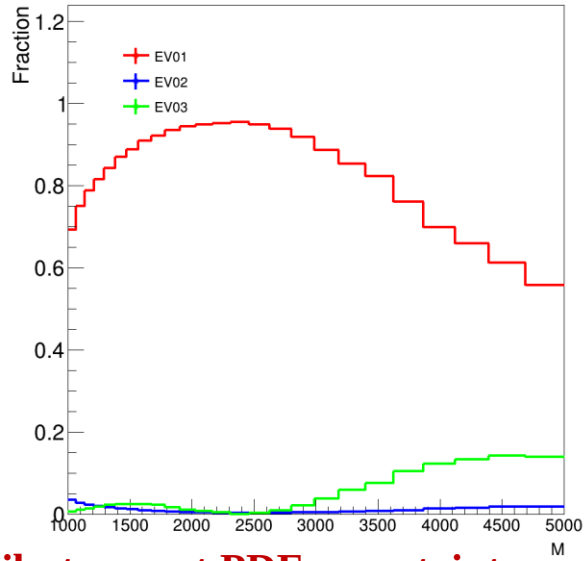
```
+++ N(EV pairs)   N(Data Sets)   PDFtype(C/L)
        29              2             C
+++ ObservableFile                  N(Observables)
test/WLeptonEtaAsymmetry                  20
test/ZHighMass_CF                         25
+++     PDFin                     PDFout
PDFs/i2Tn3.58/i2Tn3          outPDFs/i2Tn3
```

Perform ePump-optimization using two different observables at the same time.

Fractional contribution to the first observable



Fractional contribution to the second observable



After ePump-optimization, the first eigenvector contributes most PDF uncertainty on the high mass DY, but contributes less PDF uncertainty on the W asymmetry. However, the second eigenvector contributes most PDF uncertainty on the W asymmetry, but not for high mass DY. Therefore, these two observables are independent on PDF.

Appendix

How to generate data and theory files

```
FileName = "AFB_CF_40_1000";
file = new FileForEPUMP(FileName);
file->InputData(DataName, "AFB_CF");
file->InputTheory(TheoryTemplates, "AFB_CF");
file->writedata();
file->writetheory();
delete file;
```

FileForEPUMP::InputData(const char* rootName, const char* histName)
FileForEPUMP::InputTheory(vector<TString> rootNames, const char* histName)

Or

FileForEPUMP::InputData(TH1D *h1)
FileForEPUMP::InputTheory(vector<TH1D *> histos)

2D and 3D plots are also implemented.

ePumpHelper (A c++ interface code to ePump)

```
ePumpHelper* EU = new ePumpHelper(argv[1]);
EU->Update();
```

Same function as the default ePump.

```
ePumpHelper* EU = new ePumpHelper(argv[1]);
EU->Update();
EU->LinkFile(file);

TH1D* OldPDFError;
EU->ConstructOldPDFErrorHist("test/WHighMassMT", "", OldPDFError, "Relative");
TH1D* OldPDFErrorWPlus;
EU->ConstructOldPDFErrorHist("test/WPlusHighMassMT", "", OldPDFErrorWPlus, "Relative");
TH1D* OldPDFErrorWMinus;
EU->ConstructOldPDFErrorHist("test/WMinusHighMassMT", "", OldPDFErrorWMinus, "Relative");


EU->ResetUpdate();
EU->SetDataWeight("test/PositiveLeptonEta", 1.0);
EU->SetDataWeight("test/NegativeLeptonEta", 1.0);
EU->Update();
TH1D* NewPDFError_WEta;
EU->ConstructNewPDFErrorHist("test/WHighMassMT", "_updated_by_WEta", NewPDFError_WEta, "Relative");
TH1D* NewPDFErrorWPlus_WEta;
EU->ConstructNewPDFErrorHist("test/WPlusHighMassMT", "_updated_by_WEta", NewPDFErrorWPlus_WEta, "Relative");
TH1D* NewPDFErrorWMinus_WEta;
EU->ConstructNewPDFErrorHist("test/WMinusHighMassMT", "_updated_by_WEta", NewPDFErrorWMinus_WEta, "Relative");

EU->ResetUpdate();
EU->SetDataWeight("test/PositiveLeptonPt", 1.0);
EU->SetDataWeight("test/NegativeLeptonPt", 1.0);
EU->Update();
```

One can reset the status of ePump, and change the input setting inside the code, then run update again, by “EU->ResetUpdate(); EU->SetDataWeight(); EU->Update();”

ePumpHelper (A c++ interface code to ePump)

```
inline double GetTheoryTemplate(TString DataName, int i, int iPDF);
inline double GetOldTheory(TString DataName, int i);
inline double GetNewTheory(TString DataName, int i);
inline double GetOldPDFError(TString DataName, int i);
inline double GetNewPDFError(TString DataName, int i);
inline double GetOldRelativePDFError(TString DataName, int i);
inline double GetNewRelativePDFError(TString DataName, int i);
inline double GetOldPDFCorrelation(TString DataName1, TString DataName2, int i, int j);
inline double GetNewPDFCorrelation(TString DataName1, TString DataName2, int i, int j);
inline double GetOldChi2(TString DataName);
inline double GetNewChi2(TString DataName);

inline vector<double> GetNewTheoryWeightFunction(TString DataName, int i);
inline vector<double> GetNewPDFErrorWeightFunction(TString DataName, int i);

virtual void ConstructDataHist(TString DataName, TH1D* &Data);
virtual void ConstructTheoryHist(TString DataName, vector<TH1D *> &Theory);
virtual void ConstructOldPDFErrorHist(TString DataName, TString subName, TH1D* &Hist, TString type = "");
virtual void ConstructNewPDFErrorHist(TString DataName, TString subName, TH1D* &Hist, TString type = "");
```

All the information can be obtained by ePumpHelper between two updating.

How to compile codes

Path of code: /msu/data/t3work5/yfu/ePumpTraining

Setup environment:
source /msu/data/t3work5/yfu/ePumpTraining/setup.sh

Compile code:

ePump:

make –j
make OptimizePDFs

CPP:

make

gnuplot:

export PATH=/home/yfu/gnuplot/bin:${PATH}

ePumpHelper:

mkdir build
cd build
cmake ..
make –j install

export PATH=/path to ePumpHelper/build/bin:${PATH}
export LD_LIBRARY_PATH=/path to ePumpHelper/build/lib:${LD_LIBRARY_PATH}

Back up