ePump
(Error PDF Updating Method Package)

Tie-Jiun Hou
Xinjiang University
In Collaboration with Jon Pumplin, Carl Schmidt,
C.-P. Yuan, Sayipjamal Dulat, Ibrahim Sitiwaldi, Zhite Yu
April 18, 2018 DIS2018

Motivation for ePump

- **UpdatePDFs**:

  With many data sets and NNLO calculations, global fitting can be time consuming.

  - Need for fast and efficient method to estimate effects of new data before doing global fit.
  - Can estimate effects of different data set choices in real time.

- **OptimizePDFs**:

  Experimental analyses may require many MC calculations, using PDF error sets. Again, it's time consuming.

  - Optimize Hessian error PDFs to the observables, so irrelevant error PDFs may be discarded, while PDF-dependence is still maintained to desired precision.

Hessian Updating

- PDF parametrization $f(x, Q, \mathbf{Z})$: (parameters $\mathbf{Z}$)

  best-fit: $f^0 = f(x, Q, \mathbf{O})$, error PDFs: $f^{\pm i} = f(x, Q; \pm \mathbf{e^i})$

- Updated Chi-square function :

$$\Delta\chi^2(\mathbf{Z}) = \Delta\chi^2_{old}(\mathbf{Z}) + \left(X^E_\alpha - X_\alpha(\mathbf{Z})\right) C^{-1}_{\alpha\beta} \left(X^E_\beta - X_\beta(\mathbf{Z})\right)$$

- Hessian approximation :

$$\Delta\chi^2_{old}(\mathbf{Z}) = T^2 \mathbf{Z^2} \quad \text{(T = tolerance parameter)}$$

$$X_\alpha(\mathbf{Z}) = X_\alpha(\mathbf{0}) + \Delta X_\alpha \cdot \mathbf{Z} \quad \text{with} \quad \Delta X^i_\alpha = \frac{1}{2}\left(X_\alpha(+\mathbf{e})^i - X_\alpha(-\mathbf{e})^i\right)$$

- Minimize to find new best fit:

$$Z^2_{new} = (\mathbf{1} + \mathbf{M})^{-1} \mathbf{A} \quad \text{with} \quad A^i = \frac{1}{T^2}\left(X^E_\alpha - X_\alpha(\mathbf{0})\right) C^{-1}_{\alpha\beta} \Delta X^i_\beta$$

$$M^{ij} = \frac{1}{T^2} \Delta X^i_\alpha C^{-1}_{\alpha\beta} \Delta X^j_\beta$$

Updated PDF set

- New best-fit PDF : $f^0_{new} = f^0 + \Delta \cdot \boldsymbol{Z}$
- New error PDFs : $f^{\pm(r)} = f^0_{new} \pm \frac{1}{\sqrt{1+\lambda^{(r)}}} \Delta \boldsymbol{f} \cdot \boldsymbol{U}^{(\boldsymbol{r})}$

  where $\lambda^{(r)}$ and $\boldsymbol{U}^{(\boldsymbol{r})}$ are the eigenvalues and eigenvectors of matrix $\boldsymbol{M}$

Example:

- Best choices for $\Delta \boldsymbol{f}$ within the linear approximation
- Dynamical tolerances : $\pm \boldsymbol{e}^{\boldsymbol{i}} \rightarrow \left(T^{\pm i}/T\right) \boldsymbol{e}^{\boldsymbol{i}}$
- Inclusion of diagonal quadratic terms in expansion of $X_\alpha(\boldsymbol{Z})$
- Direct update of other observables :

$$Y^0_{new} = Y^0 + \Delta \boldsymbol{Y} \cdot \boldsymbol{Z} \qquad |\Delta Y| = \boldsymbol{\Delta Y} \cdot (1 + \boldsymbol{M})^{-1} \cdot \boldsymbol{\Delta Y}$$

How to use ePump

FullCT14HERA2.in — Edited

```
+++ N(EV pairs)                    N(Data Sets)    PDFtype(C/L/N)  Dyn_Tol?(Y/N)  Tol_squared
        27                              33               C               Y           100.0
+++ ObservableFile                              N(Observables)  Data?(Y/N)      Error_type      Weight
  CT14HERA2ex/tabs/E160.If1363                          1120           Y              1           1.0
  CT14HERA2ex/tabs/E101.If1363                           337           Y              1           1.0
  CT14HERA2ex/tabs/E102.If1363                           250           Y              1           1.0
  CT14HERA2ex/tabs/E104.If1363                           123           Y              1           1.0
  CT14HERA2ex/tabs/E108.If1363                            85           Y              1           1.0
  CT14HERA2ex/tabs/E109.If1363                            96           Y              1           1.0
  CT14HERA2ex/tabs/E110.If1363                            69           Y              1           1.0
  CT14HERA2ex/tabs/E111.If1363                            86           Y              1           1.0
  CT14HERA2ex/tabs/E124.If1363                            38           Y              1           1.0
  CT14HERA2ex/tabs/E125.If1363                            33           Y              1           1.0
  CT14HERA2ex/tabs/E126.If1363                            40           Y              1           1.0
  CT14HERA2ex/tabs/E127.If1363                            38           Y              1           1.0
  CT14HERA2ex/tabs/E147.If1363                            47           Y              1           1.0
  CT14HERA2ex/tabs/E145.If1363                            10           Y              1           1.0
  CT14HERA2ex/tabs/E169.If1363                             9           Y              1           1.0
  CT14HERA2ex/tabs/E201.If1363                           119           Y              1           1.0
  CT14HERA2ex/tabs/E203.If1363                            15           Y              1           1.0
  CT14HERA2ex/tabs/E204.If1363                           184           Y              1           1.0
  CT14HERA2ex/tabs/E225.If1363                            11           Y              1           1.0
  CT14HERA2ex/tabs/E227.If1363                            11           Y              1           1.0
  CT14HERA2ex/tabs/E234.If1363                             9           Y              1           1.0
  CT14HERA2ex/tabs/E260.If1363                            28           Y              1           1.0
  CT14HERA2ex/tabs/E261.If1363                            29           Y              1           1.0
  CT14HERA2ex/tabs/E267.If1363                            11           Y              1           1.0
  CT14HERA2ex/tabs/E268.If1363                            41           Y              1           1.0
  CT14HERA2ex/tabs/E240.If1363                            14           Y              1           1.0
  CT14HERA2ex/tabs/E241.If1363                             5           Y              1           1.0
  CT14HERA2ex/tabs/E281.If1363                            13           Y              1           1.0
  CT14HERA2ex/tabs/E266.If1363                            11           Y              1           1.0
  CT14HERA2ex/tabs/E504.If1363                            72           Y              1           1.0
  CT14HERA2ex/tabs/E514.If1363                           110           Y              1           1.0
  CT14HERA2ex/tabs/E535.If1363                            90           Y              1           1.0
  CT14HERA2ex/tabs/E538.If1363                           133           Y              1           1.0
+++       PDFin                 PDFout
  PDFs/CT14HERA2ex/If1363   CT14HERA2ex/PDFtmp/If1363
```

“.in” file

How to use ePump

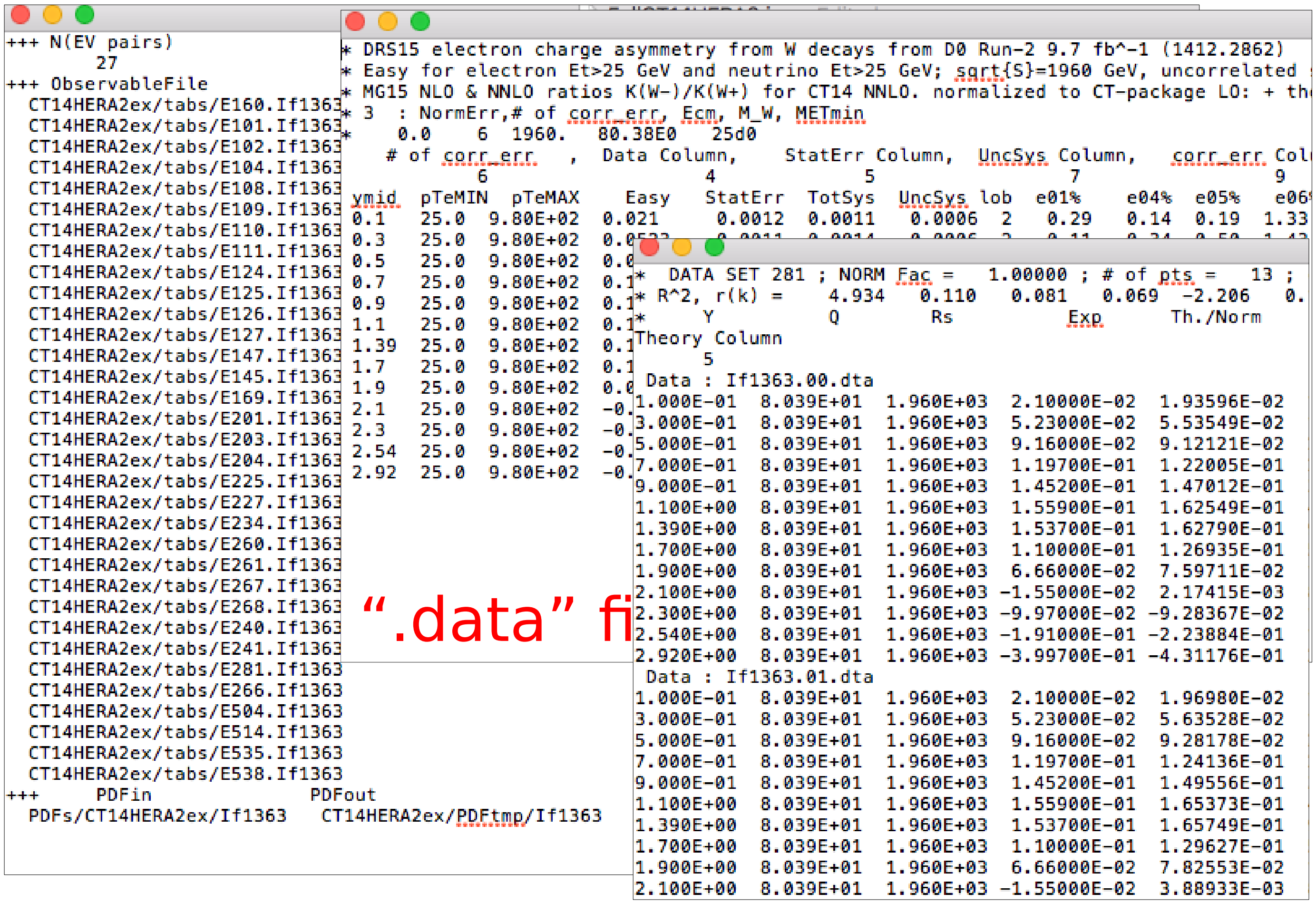
```
* DRS15 electron charge asymmetry from W decays from D0 Run-2 9.7 fb^-1 (1412.2862)
* Easy for electron Et>25 GeV and neutrino Et>25 GeV; sqrt{S}=1960 GeV, uncorrelated
* MG15 NLO & NNLO ratios K(W-)/K(W+) for CT14 NNLO. normalized to CT-package LO: + th
* 3  : NormErr,# of corr_err, Ecm, M_W, METmin
*    0.0    6  1960.   80.38E0   25d0
   # of corr_err   ,  Data Column,    StatErr Column,  UncSys Column,   corr_err Col
          6                  4                5               7               9
 ymid  pTeMIN  pTeMAX    Easy    StatErr  TotSys  UncSys lob  e01%    e04%  e05%   e06
 0.1   25.0  9.80E+02  0.021     0.0012  0.0011   0.0006  2   0.29    0.14  0.19  1.33
 0.3   25.0  9.80E+02  0.0523    0.0011  0.0014   0.0006  2   0.11    0.34  0.50  1.43
 0.5   25.0  9.80E+02  0.0916    0.0011  0.0018   0.0007  2   0.13    0.27  0.38  1.15
 0.7   25.0  9.80E+02  0.1197    0.0011  0.0025   0.0007  2   0.06    0.28  0.41  1.1
 0.9   25.0  9.80E+02  0.1452    0.0012  0.0032   0.0008  2   0.08    0.25  0.37  1.18
 1.1   25.0  9.80E+02  0.1559    0.0018  0.0041   0.0013  2   0.06    0.24  0.35  1.55
 1.39  25.0  9.80E+02  0.1537    0.0067  0.0061   0.0018  2   0.02    0.25  0.38  2.67
 1.7   25.0  9.80E+02  0.1100    0.0031  0.0049   0.0016  2   0.10    0.2   0.31  3.86
 1.9   25.0  9.80E+02  0.0666    0.0120  0.0053   0.0027  2   0.51    0.08  0.15  6.32
 2.1   25.0  9.80E+02  -0.0155   0.0053  0.0061   0.0046  2   2.39    0.26  0.45 21.48
 2.3   25.0  9.80E+02  -0.0997   0.0071  0.0088   0.0077  2   0.19    0.02  0.05  3.41
 2.54  25.0  9.80E+02  -0.1910   0.0041  0.0116   0.0103  2   0.11    0.15  0.25  2.22
 2.92  25.0  9.80E+02  -0.3997   0.0090  0.0223   0.0210  2   0.01    0.2   0.33  0.87
```

".data" file

```
+++ N(EV pairs)
        27
+++ ObservableFile
  CT14HERA2ex/tabs/E160.If1363
  CT14HERA2ex/tabs/E101.If1363
  CT14HERA2ex/tabs/E102.If1363
  CT14HERA2ex/tabs/E104.If1363
  CT14HERA2ex/tabs/E108.If1363
  CT14HERA2ex/tabs/E109.If1363
  CT14HERA2ex/tabs/E110.If1363
  CT14HERA2ex/tabs/E111.If1363
  CT14HERA2ex/tabs/E124.If1363
  CT14HERA2ex/tabs/E125.If1363
  CT14HERA2ex/tabs/E126.If1363
  CT14HERA2ex/tabs/E127.If1363
  CT14HERA2ex/tabs/E147.If1363
  CT14HERA2ex/tabs/E145.If1363
  CT14HERA2ex/tabs/E169.If1363
  CT14HERA2ex/tabs/E201.If1363
  CT14HERA2ex/tabs/E203.If1363
  CT14HERA2ex/tabs/E204.If1363
  CT14HERA2ex/tabs/E225.If1363
  CT14HERA2ex/tabs/E227.If1363
  CT14HERA2ex/tabs/E234.If1363
  CT14HERA2ex/tabs/E260.If1363
  CT14HERA2ex/tabs/E261.If1363
  CT14HERA2ex/tabs/E267.If1363
  CT14HERA2ex/tabs/E268.If1363
  CT14HERA2ex/tabs/E240.If1363
  CT14HERA2ex/tabs/E241.If1363
  CT14HERA2ex/tabs/E281.If1363                          13            Y             1           1.0
  CT14HERA2ex/tabs/E266.If1363                          11            Y             1           1.0
  CT14HERA2ex/tabs/E504.If1363                          72            Y             1           1.0
  CT14HERA2ex/tabs/E514.If1363                         110            Y             1           1.0
  CT14HERA2ex/tabs/E535.If1363                          90            Y             1           1.0
  CT14HERA2ex/tabs/E538.If1363                         133            Y             1           1.0
+++      PDFin                 PDFout
  PDFs/CT14HERA2ex/If1363   CT14HERA2ex/PDFtmp/If1363
```

".in" file

How to use ePump



```
+++ N(EV pairs)
        27
+++ ObservableFile
  CT14HERA2ex/tabs/E160.If1363
  CT14HERA2ex/tabs/E101.If1363
  CT14HERA2ex/tabs/E102.If1363
  CT14HERA2ex/tabs/E104.If1363
  CT14HERA2ex/tabs/E108.If1363
  CT14HERA2ex/tabs/E109.If1363
  CT14HERA2ex/tabs/E110.If1363
  CT14HERA2ex/tabs/E111.If1363
  CT14HERA2ex/tabs/E124.If1363
  CT14HERA2ex/tabs/E125.If1363
  CT14HERA2ex/tabs/E126.If1363
  CT14HERA2ex/tabs/E127.If1363
  CT14HERA2ex/tabs/E147.If1363
  CT14HERA2ex/tabs/E145.If1363
  CT14HERA2ex/tabs/E169.If1363
  CT14HERA2ex/tabs/E201.If1363
  CT14HERA2ex/tabs/E203.If1363
  CT14HERA2ex/tabs/E204.If1363
  CT14HERA2ex/tabs/E225.If1363
  CT14HERA2ex/tabs/E227.If1363
  CT14HERA2ex/tabs/E234.If1363
  CT14HERA2ex/tabs/E260.If1363
  CT14HERA2ex/tabs/E261.If1363
  CT14HERA2ex/tabs/E267.If1363
  CT14HERA2ex/tabs/E268.If1363
  CT14HERA2ex/tabs/E240.If1363
  CT14HERA2ex/tabs/E241.If1363
  CT14HERA2ex/tabs/E281.If1363
  CT14HERA2ex/tabs/E266.If1363
  CT14HERA2ex/tabs/E504.If1363
  CT14HERA2ex/tabs/E514.If1363
  CT14HERA2ex/tabs/E535.If1363
  CT14HERA2ex/tabs/E538.If1363
+++      PDFin                  PDFout
  PDFs/CT14HERA2ex/If1363    CT14HERA2ex/PDFtmp/If1363
```

```
* DRS15 electron charge asymmetry from W decays from D0 Run-2 9.7 fb^-1 (1412.2862)
* Easy for electron Et>25 GeV and neutrino Et>25 GeV; sqrt{S}=1960 GeV, uncorrelated
* MG15 NLO & NNLO ratios K(W-)/K(W+) for CT14 NNLO. normalized to CT-package LO: + th
* 3  : NormErr,# of corr_err, Ecm, M_W, METmin
*    0.0     6  1960.    80.38E0    25d0
   # of corr_err    ,  Data Column,     StatErr Column,  UncSys Column,    corr_err Col
            6              4              5              7                9
 ymid   pTeMIN   pTeMAX    Easy    StatErr  TotSys   UncSys  lob   e01%    e04%   e05%    e06
 0.1    25.0   9.80E+02  0.021     0.0012   0.0011   0.0006   2    0.29    0.14   0.19   1.33
 0.3    25.0   9.80E+02
 0.5    25.0   9.80E+02  0.0
 0.7    25.0   9.80E+02  0.1
 0.9    25.0   9.80E+02  0.1
 1.1    25.0   9.80E+02  0.1
 1.39   25.0   9.80E+02  0.1
 1.7    25.0   9.80E+02  0.1
 1.9    25.0   9.80E+02  0.0
 2.1    25.0   9.80E+02  -0.
 2.3    25.0   9.80E+02  -0.
 2.54   25.0   9.80E+02  -0.
 2.92   25.0   9.80E+02  -0.
```

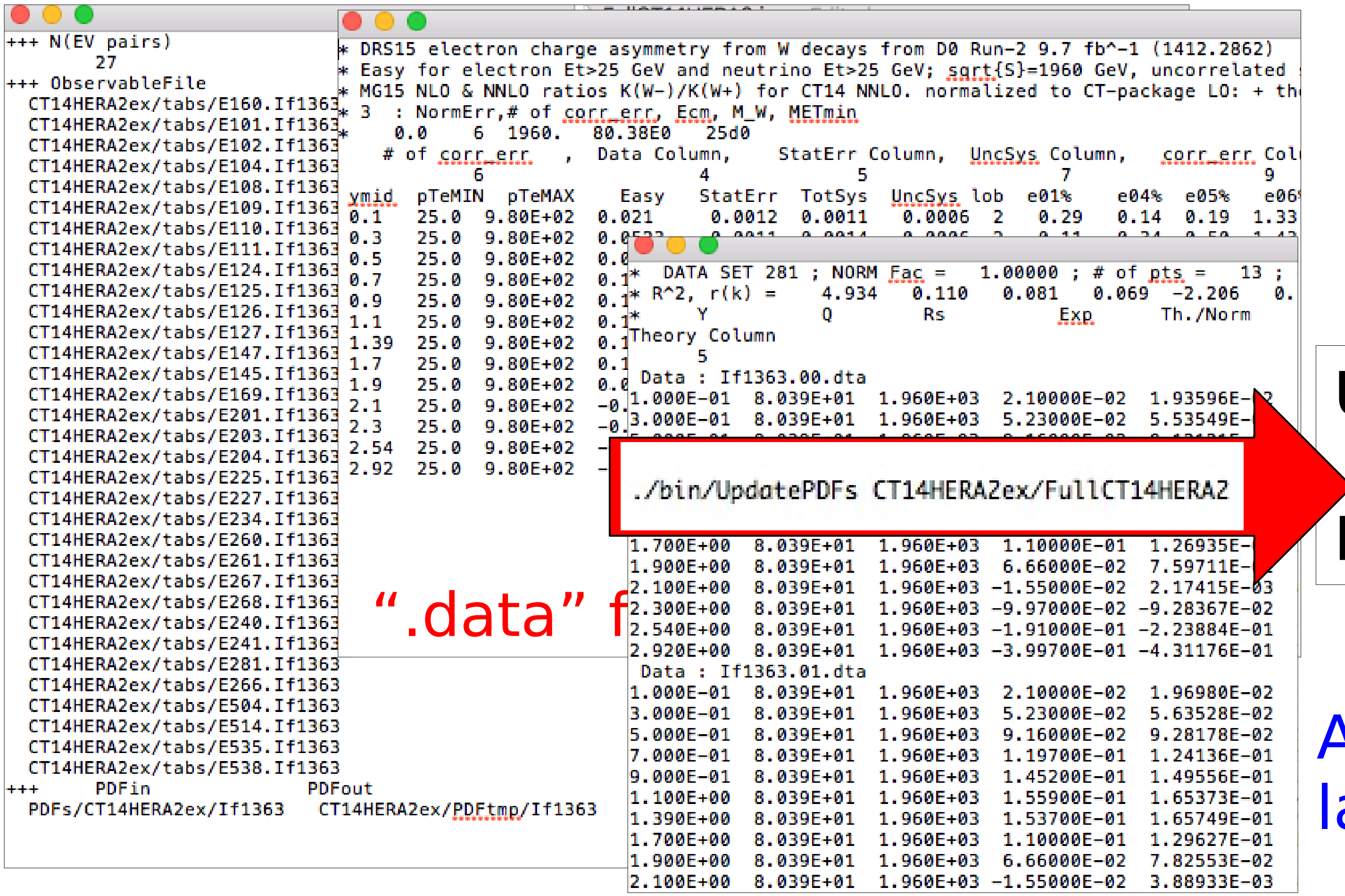
```
*  DATA SET 281 ; NORM Fac =    1.00000 ; # of pts =    13 ;
* R^2, r(k) =     4.934     0.110     0.081     0.069   -2.206     0.
*         Y           Q          Rs          Exp          Th./Norm
Theory Column
      5
 Data : If1363.00.dta
1.000E-01   8.039E+01   1.960E+03   2.10000E-02   1.93596E-02
3.000E-01   8.039E+01   1.960E+03   5.23000E-02   5.53549E-02
5.000E-01   8.039E+01   1.960E+03   9.16000E-02   9.12121E-02
7.000E-01   8.039E+01   1.960E+03   1.19700E-01   1.22005E-01
9.000E-01   8.039E+01   1.960E+03   1.45200E-01   1.47012E-01
1.100E+00   8.039E+01   1.960E+03   1.55900E-01   1.62549E-01
1.390E+00   8.039E+01   1.960E+03   1.53700E-01   1.62790E-01
1.700E+00   8.039E+01   1.960E+03   1.10000E-01   1.26935E-01
1.900E+00   8.039E+01   1.960E+03   6.66000E-02   7.59711E-02
2.100E+00   8.039E+01   1.960E+03  -1.55000E-02   2.17415E-03
2.300E+00   8.039E+01   1.960E+03  -9.97000E-02  -9.28367E-02
2.540E+00   8.039E+01   1.960E+03  -1.91000E-01  -2.23884E-01
2.920E+00   8.039E+01   1.960E+03  -3.99700E-01  -4.31176E-01
 Data : If1363.01.dta
1.000E-01   8.039E+01   1.960E+03   2.10000E-02   1.96980E-02
3.000E-01   8.039E+01   1.960E+03   5.23000E-02   5.63528E-02
5.000E-01   8.039E+01   1.960E+03   9.16000E-02   9.28178E-02
7.000E-01   8.039E+01   1.960E+03   1.19700E-01   1.24136E-01
9.000E-01   8.039E+01   1.960E+03   1.45200E-01   1.49556E-01
1.100E+00   8.039E+01   1.960E+03   1.55900E-01   1.65373E-01
1.390E+00   8.039E+01   1.960E+03   1.53700E-01   1.65749E-01
1.700E+00   8.039E+01   1.960E+03   1.10000E-01   1.29627E-01
1.900E+00   8.039E+01   1.960E+03   6.66000E-02   7.82553E-02
2.100E+00   8.039E+01   1.960E+03  -1.55000E-02   3.88933E-03
```

“.data” fi

“.in” file

“.theory” file

How to use ePump



Updated best-fit
and
Hessian Error PDFs

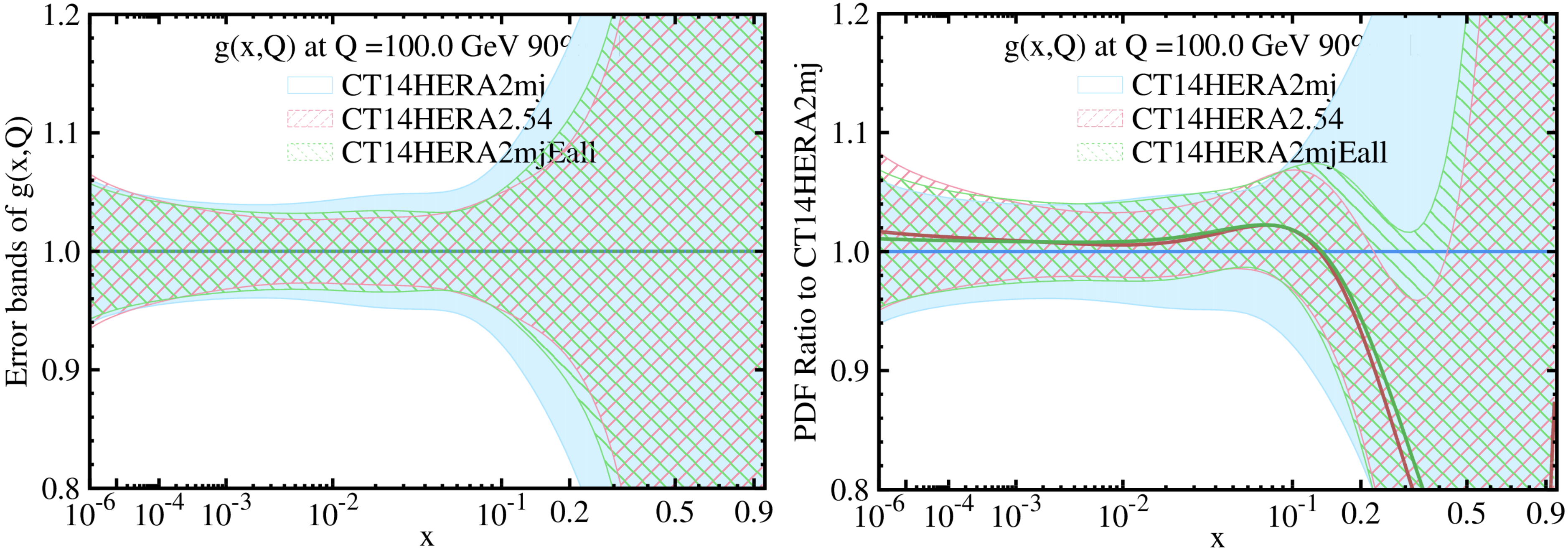
A few seconds later…

".in" file

".theory" file

Test 1: CT14HERA2 minus Jets

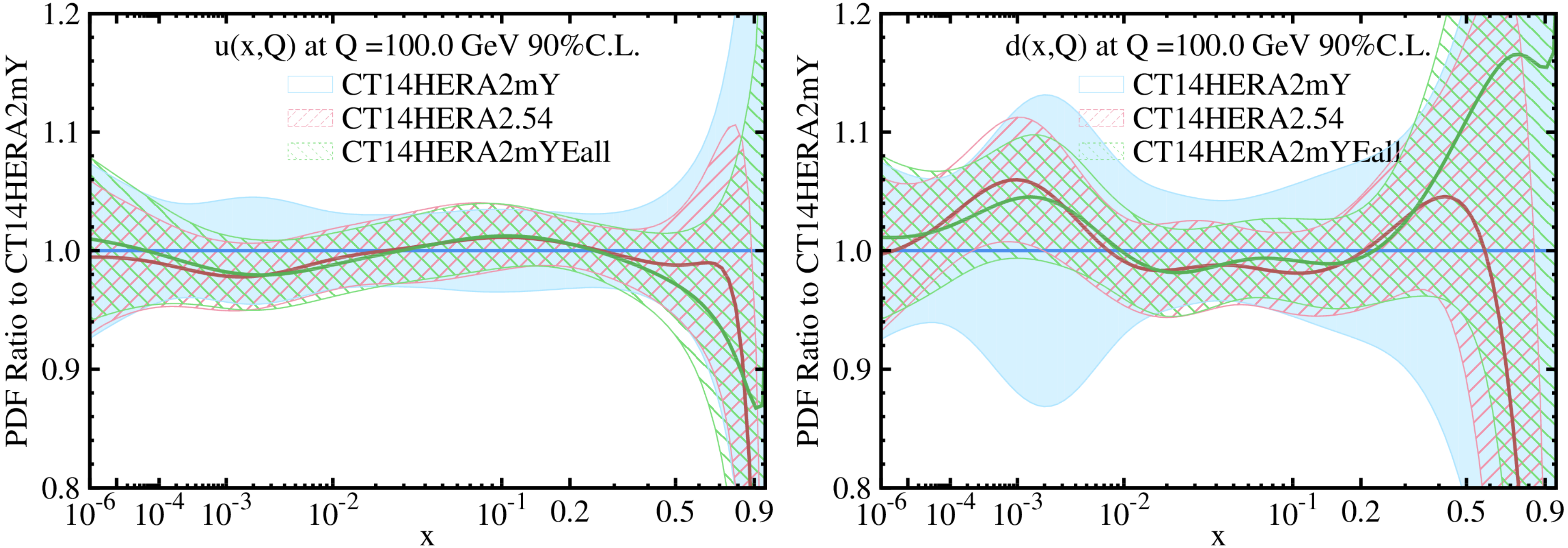
- Remove all CDF, D0, Atlas7, CMS7 jet data from CT14HERA2 and refit => CT14HERA2mj.
- Add back the 4 data sets to CT14HERA2mj by ePump and compare with CT14HERA2.



gluon-PDF

Test 2: CT14HERA2 minus Drell-Yan

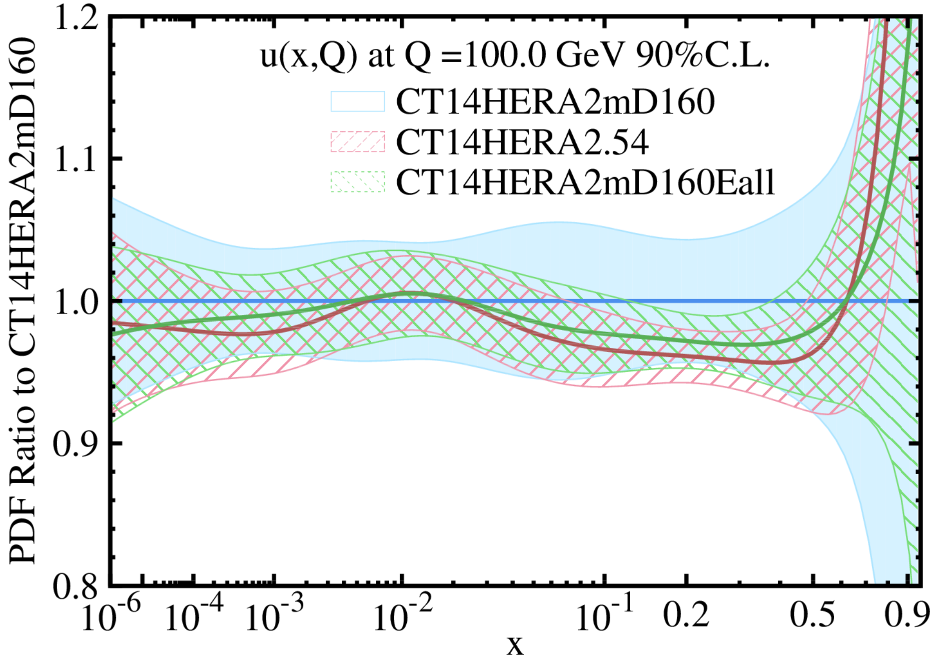
- Remove 19 Drell-Yan data sets (fixed target exp., Tevatron and LHC) from CT14HERA2 and refit => CT14HERA2mY.
- Add back the 19 data sets to CT14HERA2mY by ePump and compare with CT14HERA2.



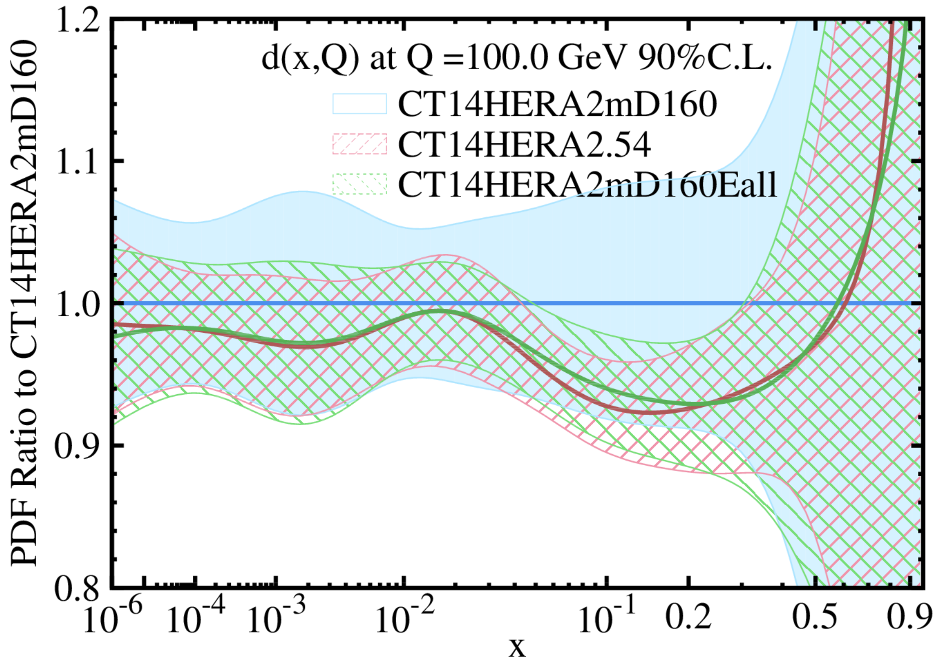
u-PDF d-PDF

Test 3: CT14HERA2 minus DIS

- Remove 14 DIS data sets (excluding HERA1+2) from CT14HERA2 and refit => CT14HERA2mD160.
- Add back the 14 data sets to CT14HEA2mD160 by ePump and compare with CT14HERA2.



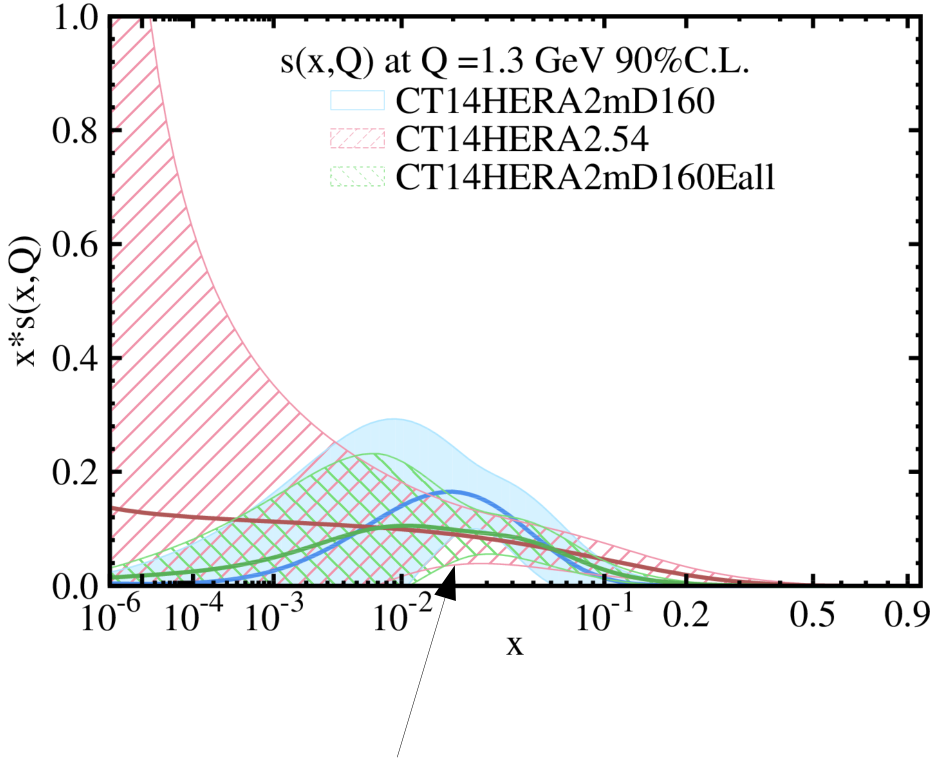
u-PDF



d-PDF

Test 3: CT14HERA2 minus DIS (2)

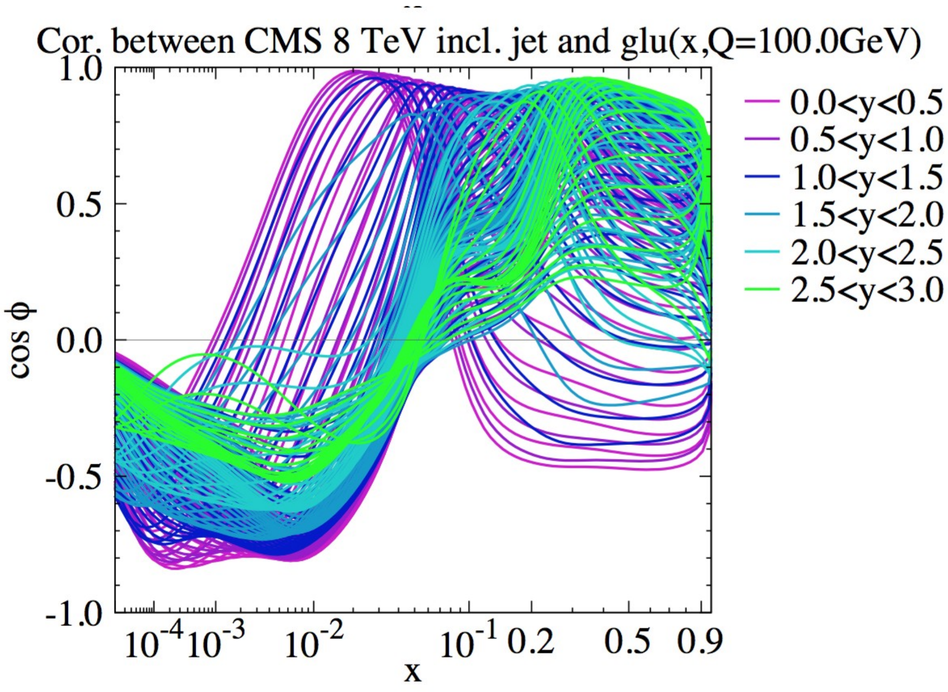
- Almost all of constraints on s-PDF are from NuTeV and CCFR neutrino dimuon data.

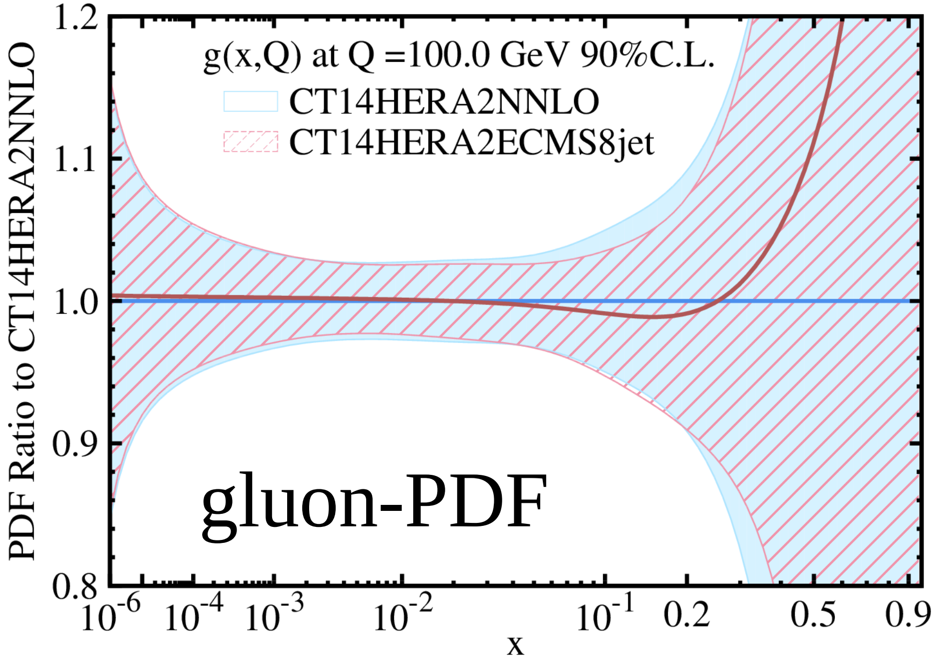


ePump work well in the x region constrained by data.

Adding New Data with ePump

- One of the most sensitive new data from the LHC is CMS 8 TeV jet data.



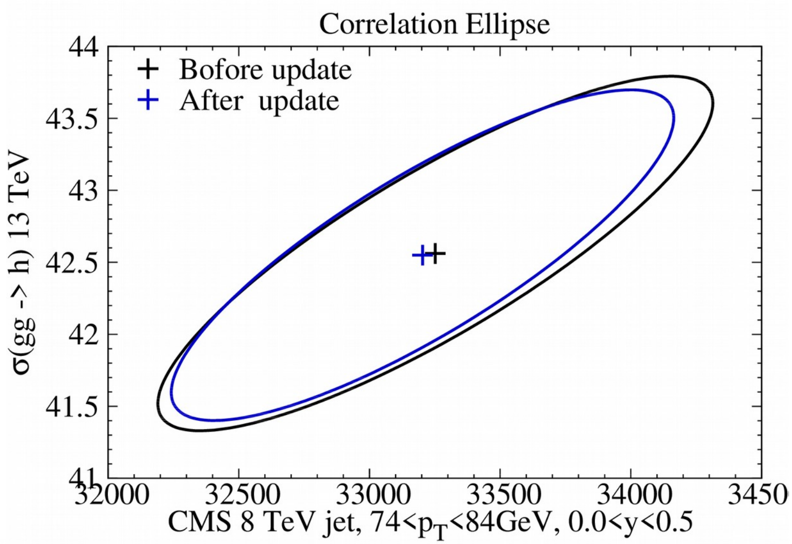


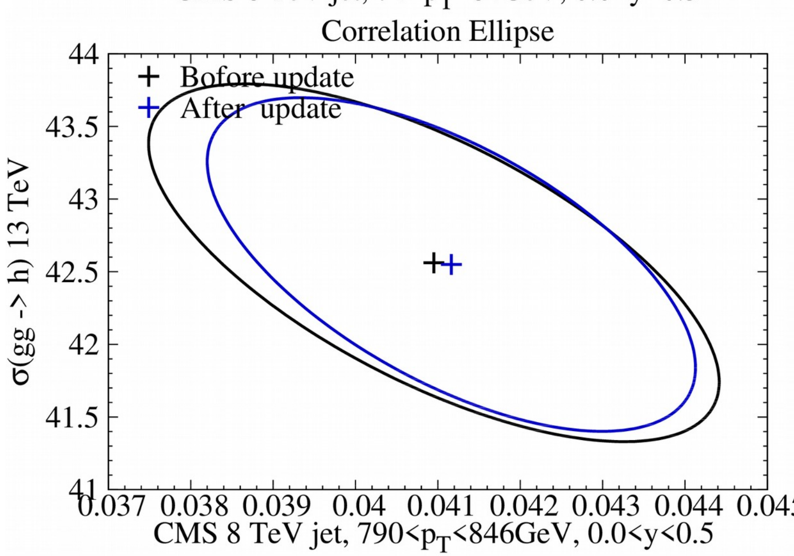
Correlation with g(x,Q)

Updated with ePump

Effect of Updating on $\sigma(gg \to H)$ due to the CMS 8TeV jet data

- Include ggH theory files to update predictions with ePump





The updated PDFs reduce the uncertainty by about 10%

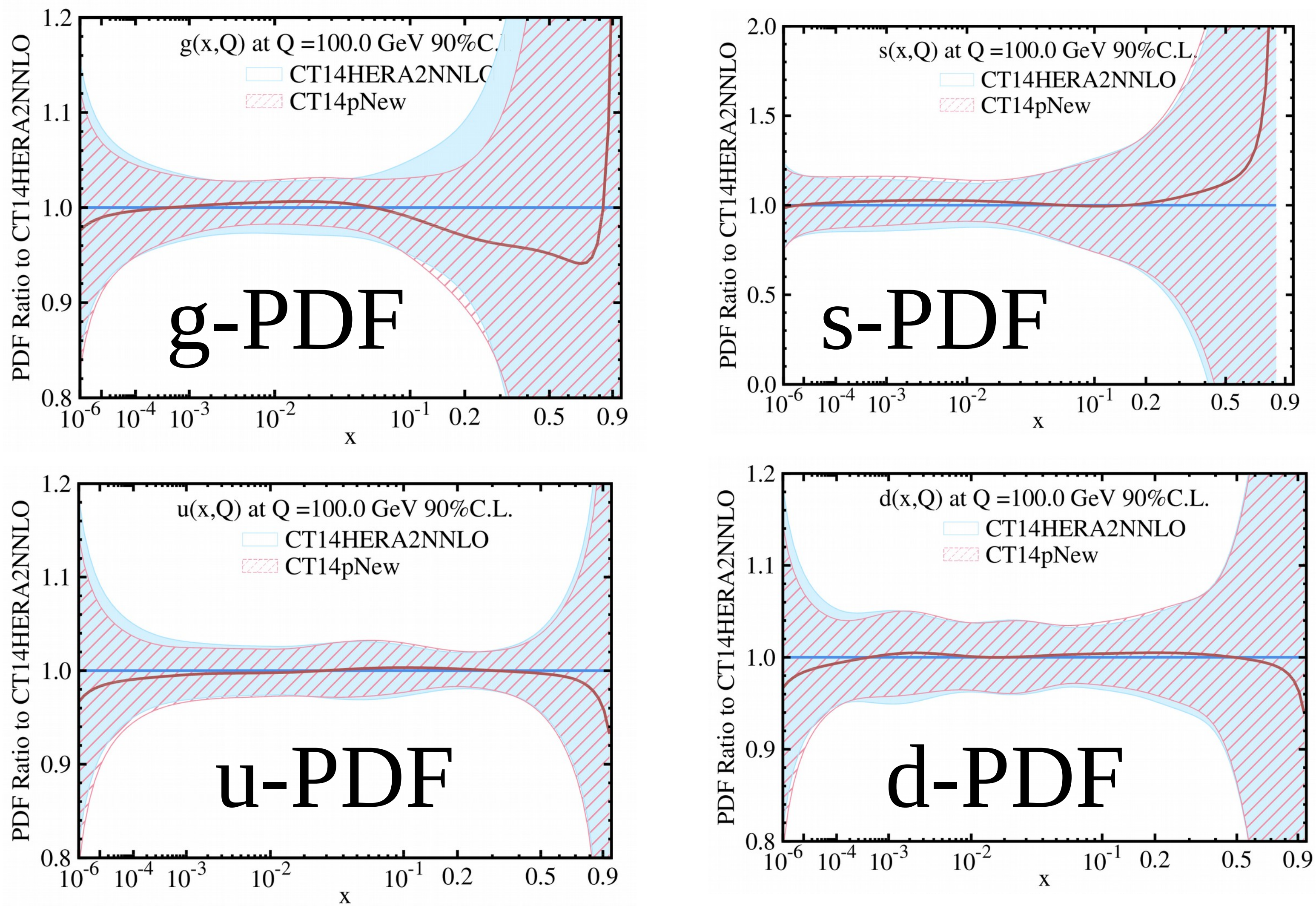
7 TeV		
	σ(gg-h)	$\delta\sigma$
CT14HERA2	14.60	0.44
updated	14.57	0.42

8 TeV		
	(gg-h)	sym(90%C.L.)
CT14HERA2	18.60	0.55
updated	18.57	0.52

13 TeV		
	(gg-h)	sym(90%C.L.)
CT14HERA2	42.56	1.23
updated	42.55	1.15

14 TeV		
	(gg-h)	sym(90%C.L.)
CT14HERA2	47.99	1.39
updated	47.98	1.30

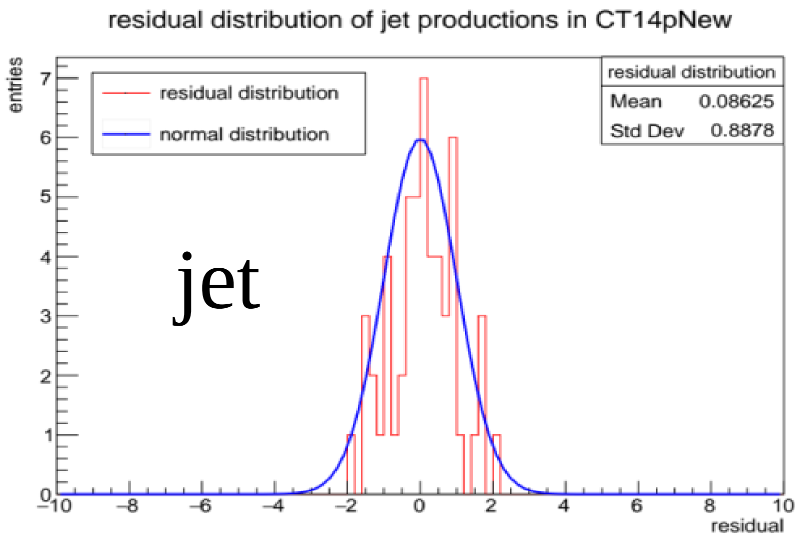
Impact of some LHC 8 TeV (jet, LHCbWZ, DY, ttbar)data

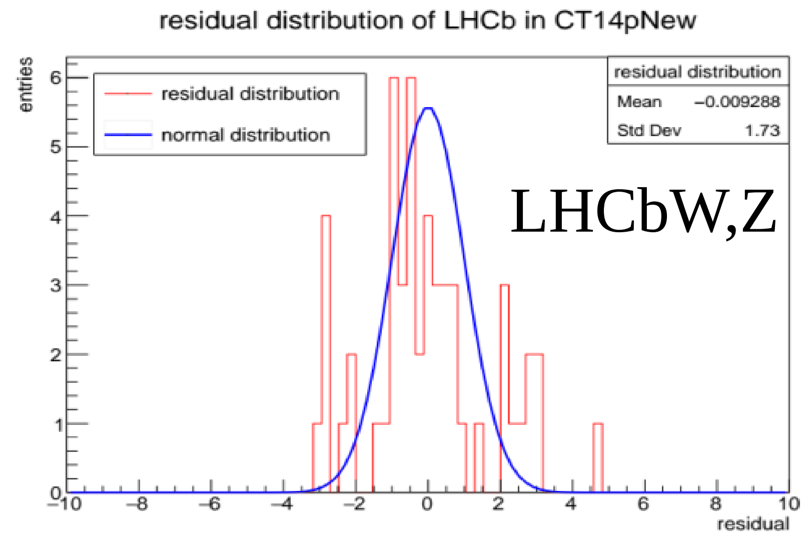


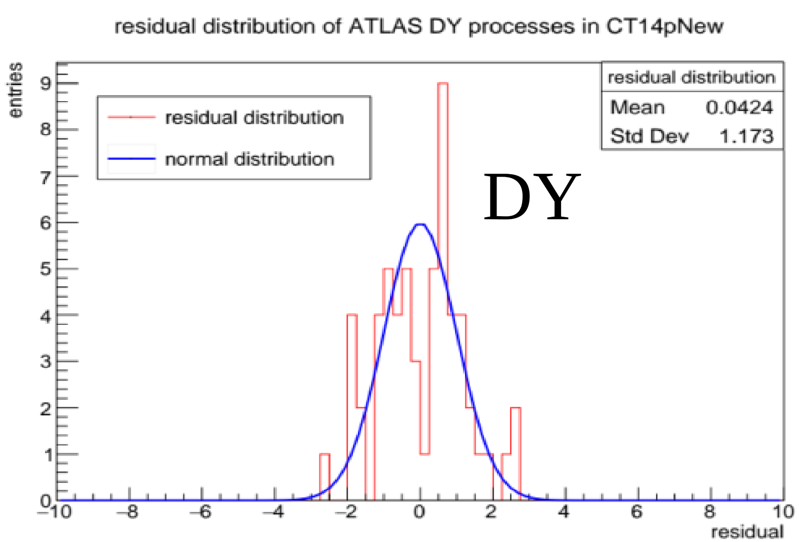
- CT14pNew includes some LHC 8 TeV (jet, LHCbWZ, DY, ttbar)data

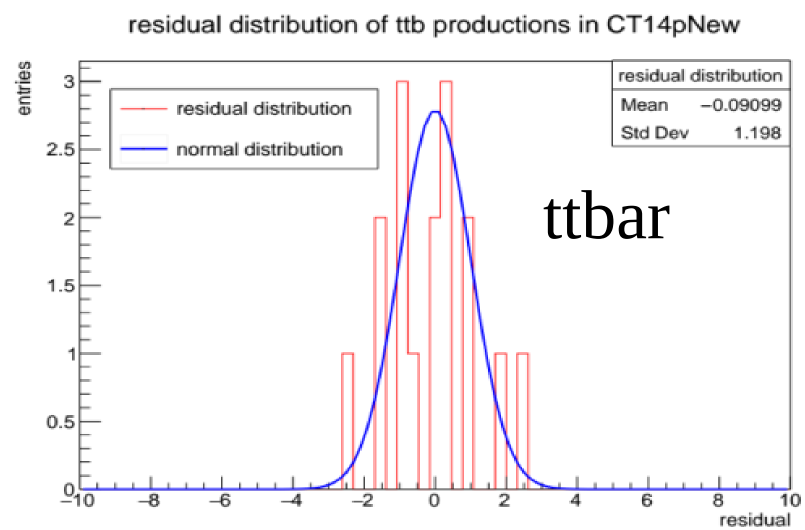
Goodness of CT14pNew

(including some jet, LHCbWZ, DY, ttbar data at 8TeV LHC)





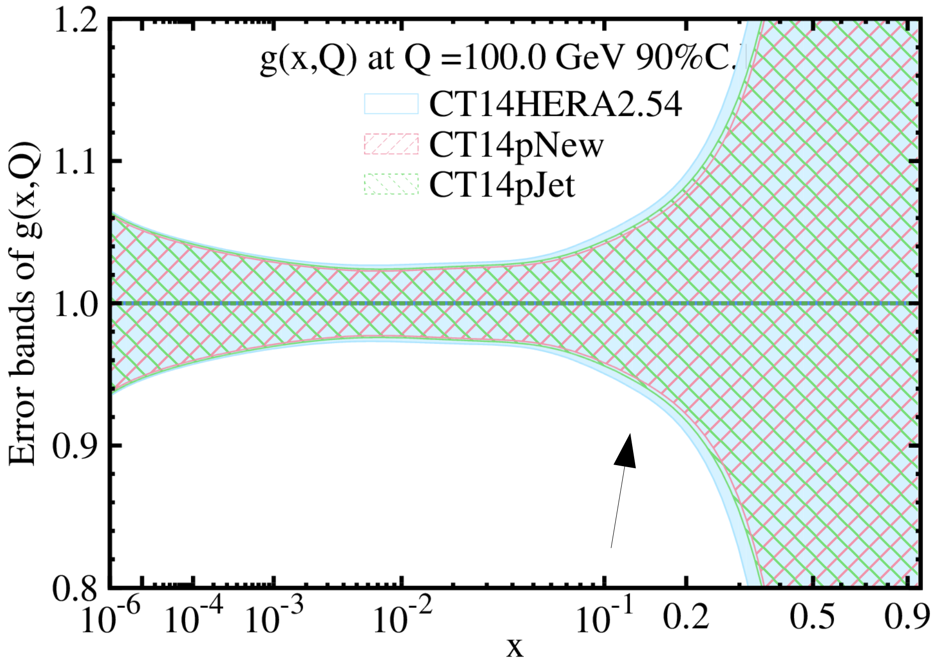


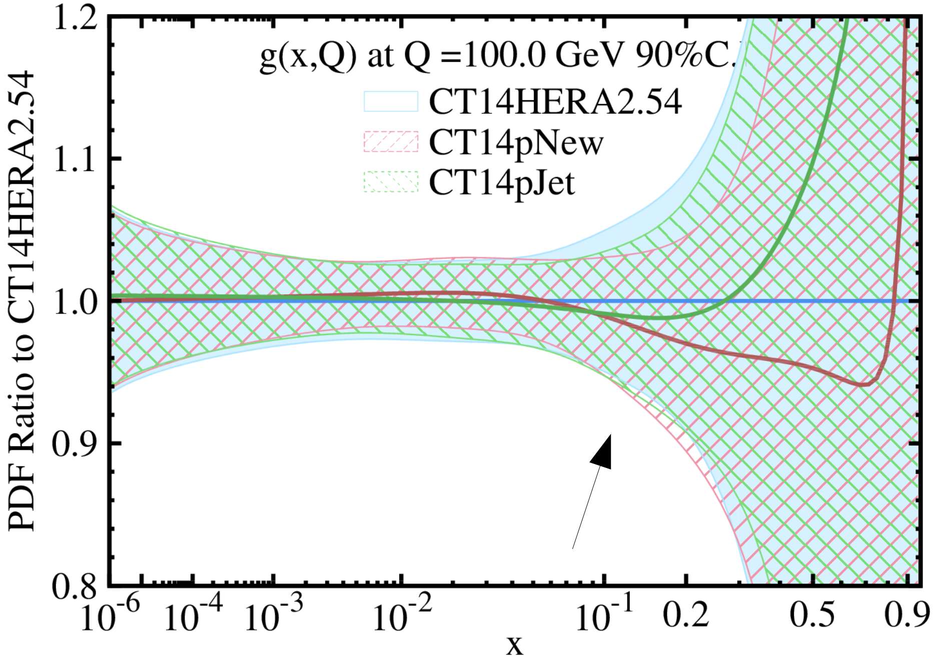


$$Residual \equiv \left(\frac{T - D_{shifted}}{\sqrt{\sigma_{stat}^2 + \sigma_{unc.sys}^2}} \right)$$

Impact from CMS 8TeV incl. jet

- Small reduction of gluon uncertainty in large x region, and reduction of PDF for x around 0.2.

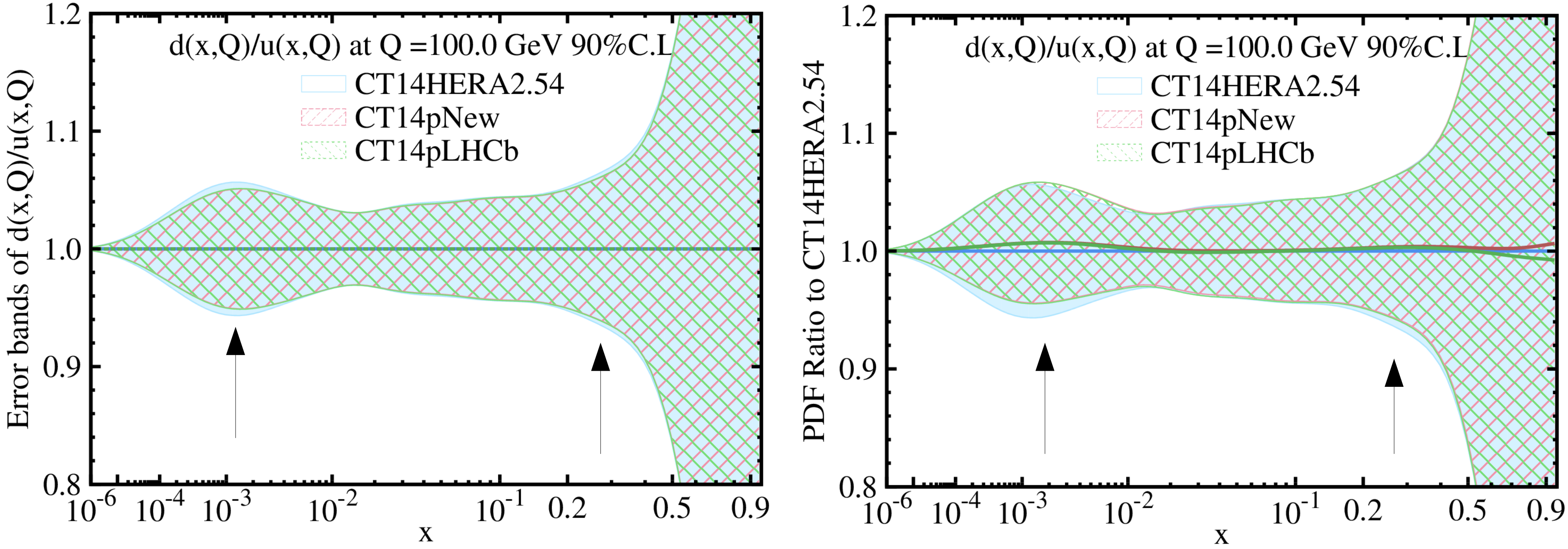




g-PDF

Impact from LHCb 8TeV (W,Z) data

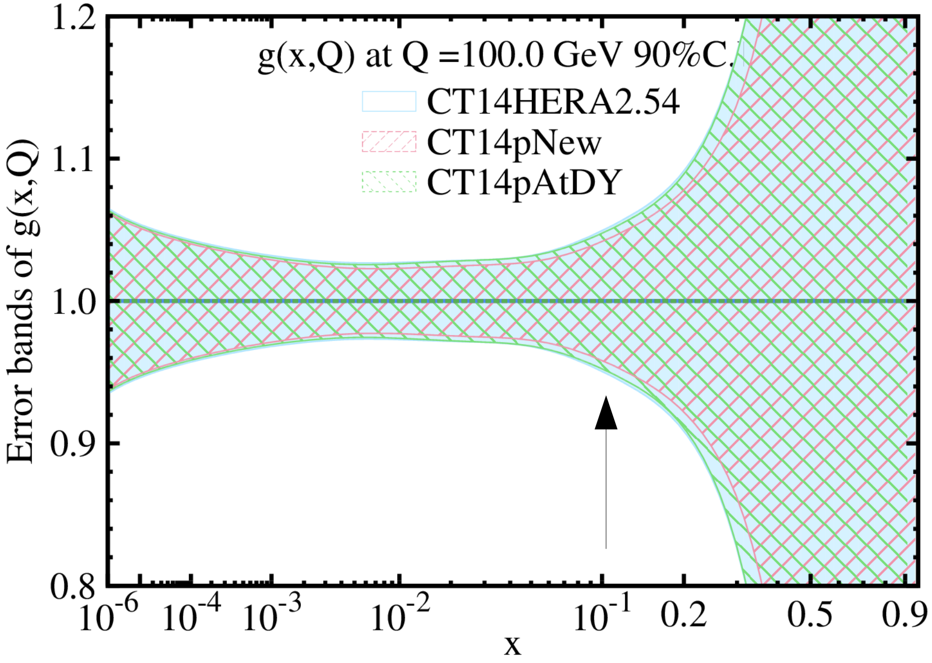
- Mild improvement of the uncertainty in d/u PDF ration from LHCb (W,Z) data

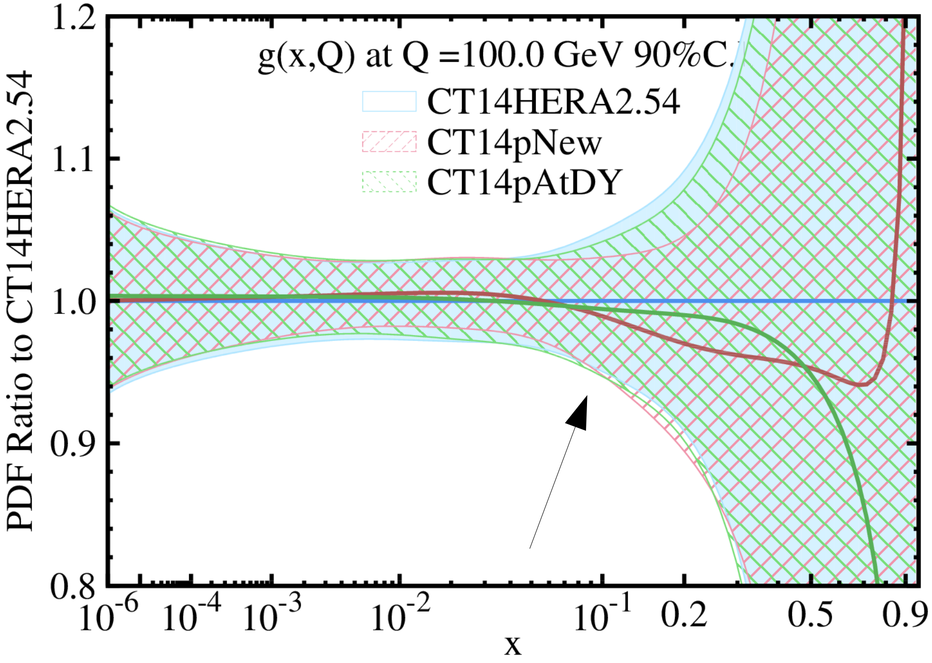


d/u PDF ratio

Impact from ALTAS 8TeV Drell-Yan data
(dSig/dpTdM and dSig/dM)

- Some reduction of gluon PDF at x around 0.2

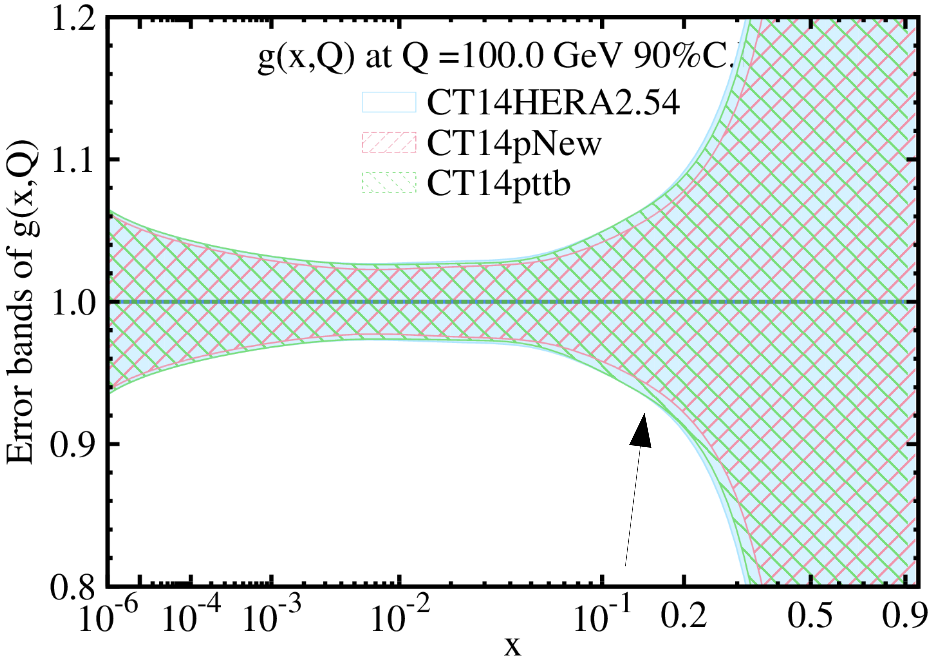


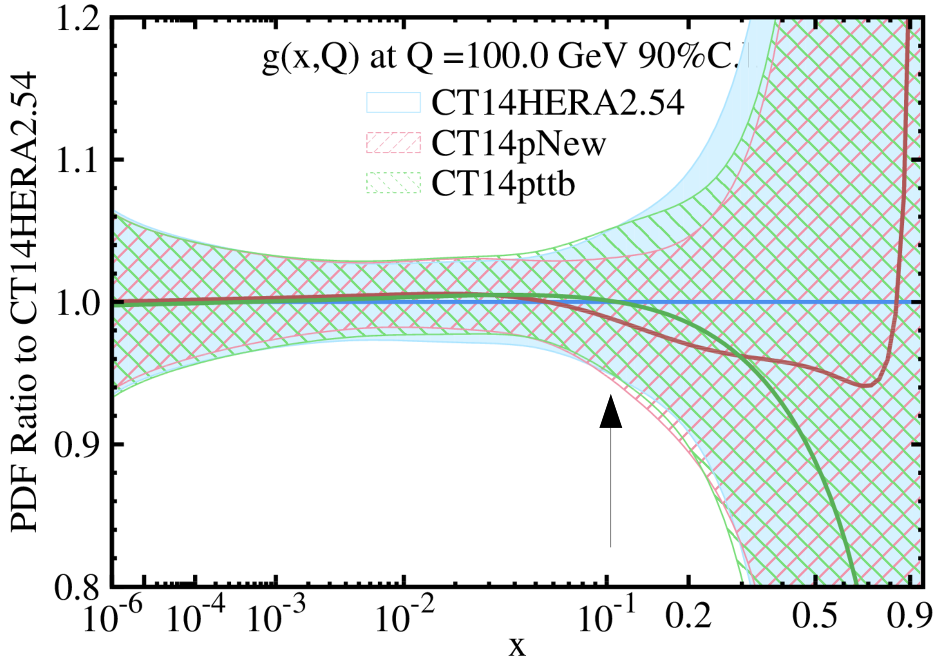


g-PDF

Impact from 8TeV ttbar data
(CMS yttbar and ATLAS ytave)

- Mild improvement of gluon uncertainty and reduction of gluon PDF for x larger than 0.1.





g-PDF

Optimized PDFs

- Based on Data Set Diagonalization

Pumplin, PRD 80 (2009) 034002

Maximize: $\Sigma\left(X_\alpha(\boldsymbol{Z})-X_\alpha(\boldsymbol{0})\right)^2/|\Delta X_\alpha|^2$ subject to constraint $\boldsymbol{Z}^2=1$

Using Hessian approximation, $X_\alpha(\boldsymbol{Z})=X_\alpha(\boldsymbol{0})+\Delta X_\alpha\cdot\boldsymbol{Z}$

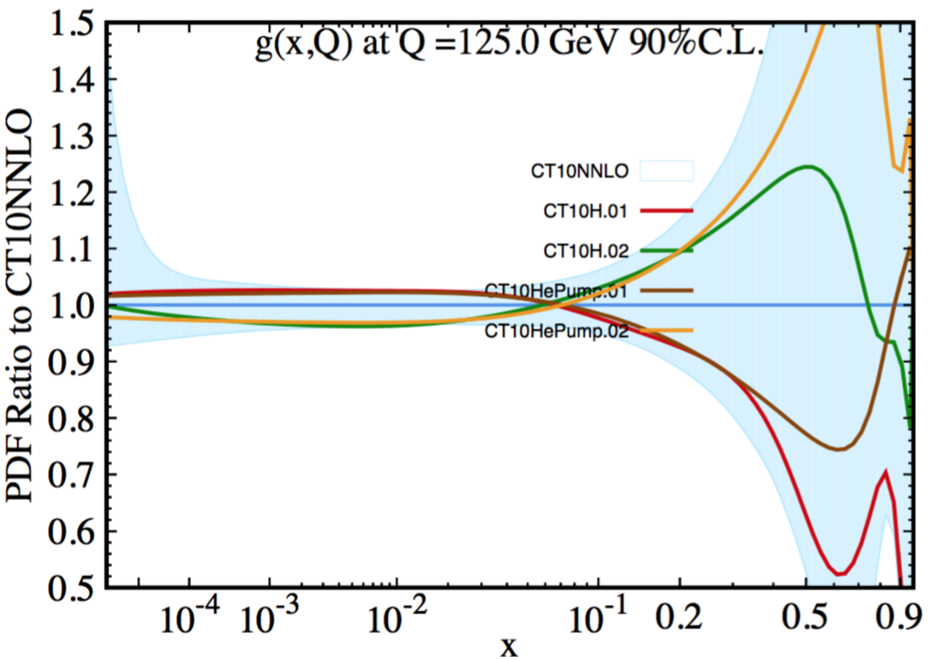
leads to matrix $M^{ij}=\Sigma\Delta X_\alpha^i\Delta X_\alpha^j/|\Delta X_\alpha|^2$ with eigenvalues/vectors, $\lambda^{(r)}$ and $\boldsymbol{U}^{(\boldsymbol{r})}$

New error PDFs: $f^{\pm(r)}=f^0\pm\boldsymbol{\Delta f}\cdot\boldsymbol{U}^{(\boldsymbol{r})}$

  - Order PDFs by eigenvectors (note : $\sum\lambda^{(r)}$ number of observables)
  - Full set of optimized PDFs reproduces Hessian symmetric errors
  - Eigenvalue $\lambda^{(r)}$ gives (sum of) fractional contribution of $f^{\pm(r)}$ to variances
  - Depending on precision required, keep reduced set of error PDFs, based on eigenvalues.

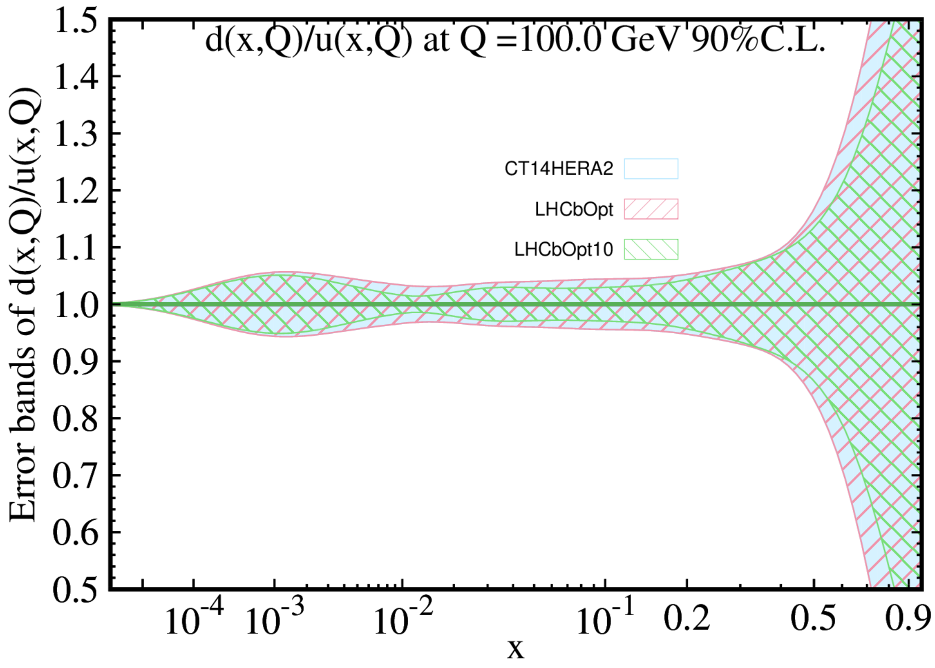
Examples for Optimized PDFs

- One observable: gg->H cross section at 14 TeV: Optimize CT10NNLO
  - Eigenvalues: 1,0,0,0,0,…
  - Cross section only depends on first optimized eigenvector pair
  - Equivalent to CT10H extreme sets determined exactly by Lagrange Multiplier method
- 8 observables (H & ttbar, each at 7,8,13,14 TeV)
  - Eigenvalues: 4.45, 3.46, 0.089, 0.0017, $1\text{x}10^{-4}$, …
  - First 4 eigenvector pairs pick up 99.99% of PDF dependence for the observables
- Input given by same ".theory" files as for updating.
  - Any number of observables or sets of observables can be used for optimizing.



Optimize PDF to LHCb 8 TeV WZ data

- It has a total of 51 data points
- LHCbOpt10 include the first 10 pairs of Eigenvector sets of the Optimized PDFset, which covers most of the uncertainty in d/u ratio over wide range of x.



CT14HERA2 contains 28 pairs of error PDF sets

Summary

- The ePump package contains two functionalities

  UpdatePDFs is a fast and efficient method to estimate the effect of new data on the a current set of best-fit and Hessian error PDFs.

  OptimizePDFs can be used to find optimized set of Hessian error PDFs for specialized experimental analyses. It gives a simple method for reducing the number of optimized error PDFs, while maintaining a specified precision.

- Extensively cross-checked against CT14 global fits
- ePump paper is in preparation, and ePump code will be made public available at http://hep.pa.msu.edu/epump/