



# NNPDF Reweighting Update to LHAPDF6

Kristin Lohwasser, Alberto Guffanti

---

DESY, NBI-Kopenhagen





# Reminder: Reweighting

... workflow of the (NNPDF) reweighting approach  
for random replica PDF sets

## 1) Choose PDF sets to be reweighted

- if this is NNPDF, then they are already in the right format
- if they are Hessian PDFs, you need to  
**create a random replica version** of the PDF set

## 2) Run the reweighting

- feed in theory/experimental data and calculate  $\chi^2$

## 3) Evaluate results

- automatic printouts, draw PDFs

(some words on alternatives → *Profiling*)



# Pre-requisites

## 1) Install LHAPDF 6.1.4 at least

(should get error message if version is too small)

## 2) configure HeraFitter to use LHAPDF and NNPDF-reweighting:

```
./configure --enable-nnpdfWeight --enable_lhapdf  
make  
make install
```

## 3) Adapt the steering file

(see next slide.... )



# Steering file: Basics

\* **At the bottom of steering.txt**

\* (Optional) reweighting steering cards

\*

&reweighting

<b>FLAGRW</b>	<b>= True</b>	<b>! Should reweighting be done?</b>
RWPDFSET	= 'HERAPDF15NLO_EIG'	! LHAPDF grid file
RWDATA	= 'test'	! arbitrary name for new datasample to be put in
<b>RWMETHOD</b>	<b>= 1</b>	<b>! either 1=chi2 or 2=data</b>
<b>DORWONLY</b>	<b>= True</b>	<b>! do / do not run usual HERA fit</b>
RWREPLICAS	= 100	! Number of input replicas used to build the PDF
probability distributions from the Hessian input PDFs		
RWOUTREPLICAS	= 10	! how many output replica of the NNPDF should be kept?

&End

The red settings ***need to be changed*** (“default” for the reweighting)

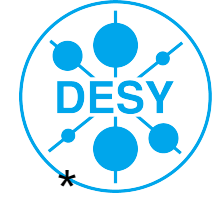
1) **FLAGRW**: → switched on PDF reweighting

2) **RWMETHOD**: → we can plug in either data+theory

(or by hand-calculated chi2 values, which is not yet implemented)

3) **DORWONLY**: → switched off usual HERAFit

(switching on runs the usual HERA fit independently,  
directly after the reweighting)



# Steering file: Choose PDF sets

\* (Optional) reweighting steering cards

\*

&reweighting

FLAGRW	= True	! Should reweighting be done?
<b>RWPDFSET</b>	<b>= 'HERAPDF15NLO_EIG'</b>	<b>! LHAPDF grid file</b>
RWDATA	= 'test'	! arbitrary name for new datasample to be put in
RWMETHOD	= 1	! either 1=chi2 or 2=data
DORWONLY	= True	! do / do not run usual HERA fit
<b>RWREPLICAS</b>	<b>= 100</b>	<b>! Number of input replicas used to build the PDF probability distributions from the Hessian input PDFs</b>
RWOUTREPLICAS	= 10	! how many output replica of the NNPDF should be kept?

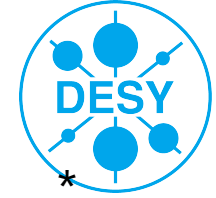
Now need to actually steer what we want to reweight

1) **RWPDFSET**: → Which PDF set do you want to start from?

(if this is a NNPDF set → creating of random sets will be skipped)

2) **RWREPLICAS**: → number should be *~large (500-1000)*

(to allow for a large enough sample from which to get the reweightings)



# Steering file: Output

\* (Optional) reweighting steering cards

\*

&reweighting

FLAGRW	= True	! Should reweighting be done?
RWPDFSET	= 'HERAPDF15NLO_EIG'	! LHAPDF grid file
<b>RWDATA</b>	<b>= 'test'</b>	<b>! arbitrary name for new datasample to be put in</b>
RWMETHOD	= 1	! either 1=chi2 or 2=data
DORWONLY	= True	! do / do not run usual HERA fit
RWREPLICAS	= 100	! Number of input replicas used to build the PDF

probability distributions from the Hessian input PDFs

**RWOUTREPLICAS = 10**      **! how many output replica of the NNPDF should be kept?**

What will our output look like?

1) **RWDATA**: → Defines name for the output PDF sets

2) **RWOUTREPLICAS**: → How many members to keep after **reweighting** (should be large enough to get a decent mean +/- RMS error, smooth distributions → 50-100)



# Exercise 1 – Creating replicas

1) Adapt the steering file to run the following:

- use MSTW2008nlo68cl as input set
- output dir: test
- 10 replicas (for input) – 5 output replicas (after reweighting)

2) Run: `./bin/FitPDF`

3) Look at the output:

```
[..... lots of HeraFitter initialization, reading in of data sets.....]
```

```
Initialize theory for datasets
```

```
isSymmetricErrors  F
isSymmetricErrors  0
nnpdfrw isSymmetric  0
*****MSTW2008nlo68cl_test_chi2
*****MSTW2008nlo68cl
```

Some useful  
expert printouts,  
debug-info

```
Writing out replica LHAPDF grid: MSTW2008nlo68cl_InputReplicas (.LHgrid)
```

```
creating 10 random replicas from PDF with asymmetric errors
```

```
Using LHAPDF version: 6.1.4
```

```
Creating random replica input LHgrid file
```

```
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_10InputReplicas/MSTW2008nlo68cl_10InputReplicas This is our output LHAPDF file / directory
```



# Exercise 1 – More printouts

Writing replica: 1/10

```
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_10InputReplicas/MSTW2008nlo68cl_10InputReplicas_0001.dat
```

```
Generating x-Q2 grid for replica 1/10 -- 100 % done
```

Writing replica: 2/10

```
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_10InputReplicas/MSTW2008nlo68cl_10InputReplicas_0002.dat
```

```
Generating x-Q2 grid for replica 2/10 -- 100 % done
```

Number of replicas to be made is always shown with progress

**Note: 1000 replicas take ~4 hours (but we can re-use sets)**

Automatic skipping (also for NNPDF inputs)

→ *extra Exercise: Test it*

Output file

```
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_10InputReplicas/MSTW2008nlo68cl_10InputReplicas_0000.dat already exists. Skipping generation of random replicas!
```

MSTW2008nlo68cl

**Random replica file exists → skip replica creation**

Initialize theory for datasets

isSymmetricErrors T

isSymmetricErrors 1

PDF set is already NNPDF set

**NNPDF set → skip replica creation**

proceed directly to reweighting





# Exercise 1 – Pitfalls, Tipps

## 1) Need to setup LHAPDF

```
./bin/FitPDF: error while loading shared libraries: libAPPLgrid.so.0:  
cannot open shared object file: No such file or directory
```

```
export PATH=$PATH://home/kristin/PDFGenerator/LHAPDF/bin/  
LD_LIBRARY_PATH=$LD_LIBRARY_PATH://home/kristin/PDFGenerator/LHAPDF/lib/
```

## 2) Mistyped PDF set name

```
Read      4 data files  
Fitting for PROTON PDFs, PDFType=proton  
terminate called after throwing an instance of 'LHAPDF::UserError'  
what():  Can't find a valid PDF MSTW208nlo68cl/0
```

## 3) Things to note:

HERA VAR and EIG sets are treated separately at the moment  
(no possibility to create combined random replica set)



# Exercise 1 – Solution

\* (Optional) reweighting steering cards

\*

&reweighting

FLAGRW = True ! Should reweighting be done?

RWPDFSET = 'MSTW2008nlo68cl' ! LHAPDF grid file

RWDATA = 'test' ! arbitrary name for new datasample to be put in

RWMETHOD = 1 ! either 1=chi2 or 2=data

DORWONLY = True ! do / do not run usual HERA fit

RWREPLICAS = 10 ! Number of input replicas used to build the PDF probability

distributions from the Hessian input PDFs

RWOUTREPLICAS = 5 ! how many output replica of the NNPDF should be kept?

&End



# Exercise 1 – Check Replicas: Plot

You can use the random replica PDFs just like a regular LHAPDF grid (albeit of the NNPDF style)

But: you need to copy it into the LHAPDF directory:

```
cp -r
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_10InputReplicas/
directory-to-LHAPDF/share/LHAPDF/.
```

A quick preparation to check the output (standalone code):

```
tar -zxf plot_LHAPDF.tar.gz
cd plot_LHAPDF/
make
```

```
emacs plot.in.1
```

And adapt the name of the PDF files

```
./plot_pdf plot.in.1
```

```
0 # SUBSET
10000 # NPOINTS
2 # nset
0.00001 # xmin
0.15 # xmax
80.385 # Q 80.385
91.2 # mz
downquark #
```

```
MSTW2008nlo68cl_10InputReplicas
MSTW2008nlo68cl
```



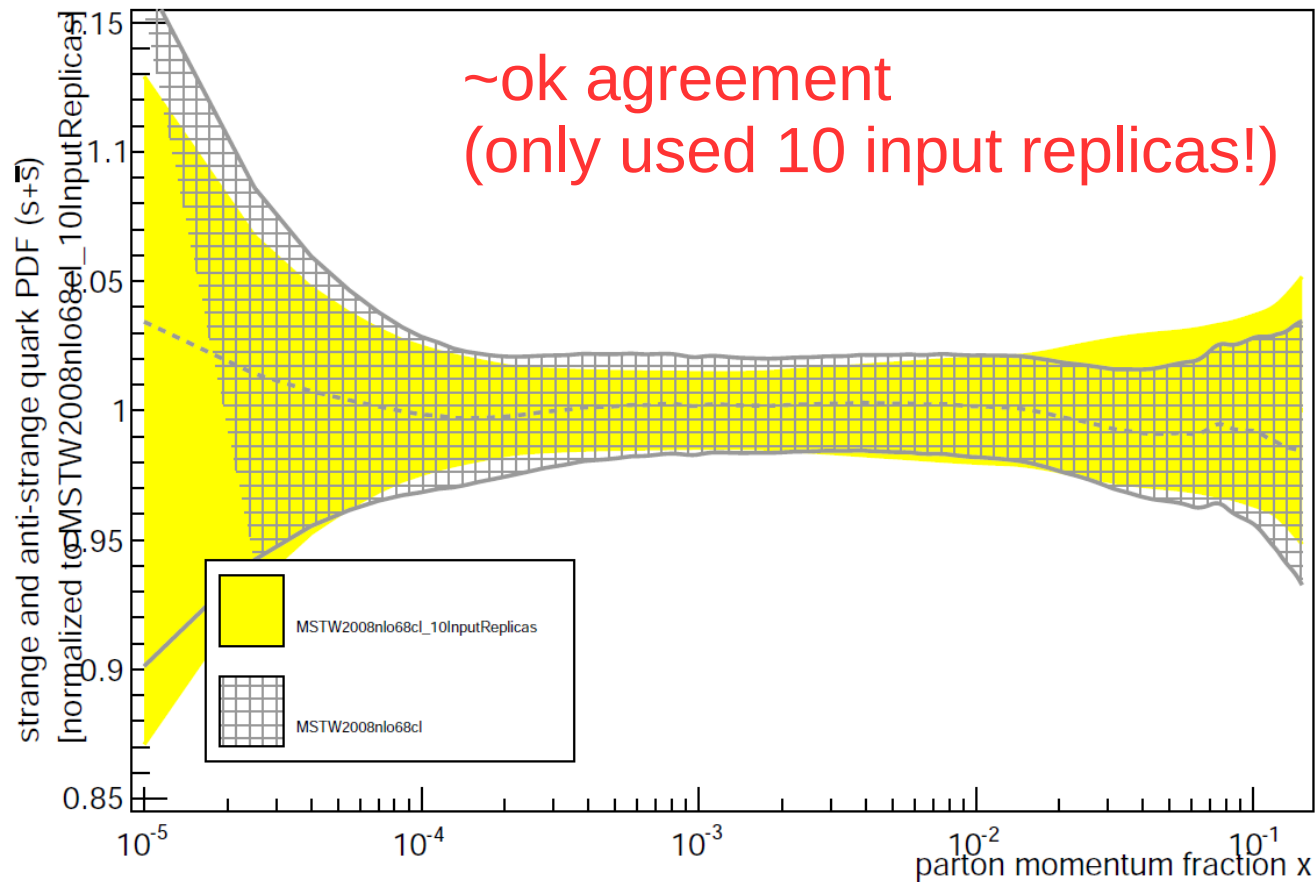
# Exercise 1 – Output

You get 3x4 files:

- XXX.pdf – parton distribution for parton XXX
- RatioXXX.pdf – Ratio of PDFs to their respective central values
- RatiorefXXX.pdf – Ratio of PDFs for parton XXX to reference PDF

Plotted are up, down and  $s+\bar{s}$

(note: gluon files are buggy)





# Steering file: Data samples

## At the start of steering.txt

→ **the same as for usual HeraFitting...**

\* Namelist to control input data

\*

&InFiles

! Number of input files

NInputFiles = 4

! Input files:

InputFileNames(1) = 'datafiles/hera/H1ZEUS\_NC\_e-p\_HERA1.0.dat'

InputFileNames(2) = 'datafiles/hera/H1ZEUS\_NC\_e+p\_HERA1.0.dat'

InputFileNames(3) = 'datafiles/hera/H1ZEUS\_CC\_e-p\_HERA1.0.dat'

InputFileNames(4) = 'datafiles/hera/H1ZEUS\_CC\_e+p\_HERA1.0.dat'

&End

InputFileNames(5) = 'datafiles/lhc/atlas/WZ2010/WP\_applgrid\_nnlo.dat'

InputFileNames(2) = 'datafiles/lhc/atlas/WZ2010/WM\_applgrid\_nnlo.dat'

InputFileNames(3) = 'datafiles/lhc/atlas/WZ2010/Z0\_applgrid\_nnlo.dat'

Let's try fitting some ATLAS data (for real)



# Exercise 2 – Reweighting

1) Adapt the steering file as shown in previous slide

- RWREPLICAS = 200
- RWOUTREPLICAS = 10

2) Run, random replica creation will be skipped, what else happens?

3) Look at the output:

reading in the random replicas

```
LHAPDF 6.1.4 loading
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_200InputReplicas/MSTW2008nlo68cl_200InputReplica
s_0001.dat
MSTW2008nlo68cl_200InputReplicas PDF set, member #1, version 6; LHAPDF ID = -1
  NDiag=      30 NCoavar=      0
Got MC set=   1 chi2=   82.6 ndf=   30
LHAPDF 6.1.4 loading
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_200InputReplicas/MSTW2008nlo68cl_200InputReplica
s_0002.dat
MSTW2008nlo68cl_200InputReplicas PDF set, member #2, version 6; LHAPDF ID = -1
Got MC set=   2 chi2=   58.0 ndf=   30
```

chi2 between theory calculate with  
the replica PDF set  
and the data to be fitted



# Exercise 2 – Reweighting

## More output

```
READING IN NNPDF steering FILE: input_steering/MSTW2008nlo68cl_test_chi2.in
RWMODE: Chi2 File
Found chi2 data file: NNPDF/data/test/MSTW2008nlo68cl/chi2/test_MSTW2008nlo68cl-
chi2.res
Reweighting with n_dat: 30
Data file: NNPDF/data/test/MSTW2008nlo68cl/chi2/test_MSTW2008nlo68cl-chi2.res
```

File which records the Chi2 values

```
*****
Nweights: 200
Sum of Weights: 200
Average weight: 1
Max Weight: 47.1943
Min Weight: 4.06443e-12
Shannon: 23.876
N_rep,eff: 10.8737
*****
```

Info output from NNPDF reweighting

```
LHAPDF 6.1.4 loading all 41 PDFs in set MSTW2008nlo68cl
MSTW2008nlo68cl, version 2; 41 PDF members
```

```
Writing out LHAPDF grid: MSTW2008nlo68cl_test_chi2_nRep20 Output PDF set
```

```
Using LHAPDF version: 6.1.4
```

```
output/MSTW2008nlo68cl_test_chi2/MSTW2008nlo68cl_test_chi2_nRep20/MSTW2008nl
o68cl_test_chi2_nRep20_0001.dat
```

```
Writing replica: 1/20
```



# Exercise 2 – Reweighting

## More output

LHAPDF Writeout successful

Info in <TCanvas::Print>: eps file **output/MSTW2008nlo68cl\_test\_chi2/whist-rw.eps** has been created

output/MSTW2008nlo68cl\_test\_chi2/palpha-rw.eps

0x71d393a0

Info in <TCanvas::Print>: eps file **output/MSTW2008nlo68cl\_test\_chi2/palpha-rw.eps** has been created

Check files for the reweighting





# Exercise 2 – Pitfalls, Tipps

## 1) Chi2-file already there

```
Data file: NNPDF/data/test/CT10nlo/chi2/test_CT10nlo-chi2.res  
Error: chi2 file has 200 entries  
PDF consists of 100 replicas.
```

## 2) no reweighting possible

happens, if there are too little sets leftover after reweighting  
no special output message → **but check plots are empty or skewed**



# Alternatives: Profiling

Using  $\chi^2$  minimization procedure developed for HERA measurements and also used in ATLAS analysis → for combinations between measurements

Code adapted to also take into account uncertainty on theoretical predictions and their correlations

$$\chi^2 = \sum_{k,i} w_k^i \frac{\left[ \mu_k^i - m^i \left( 1 + \sum_j \gamma_{j,k}^i b_j + \sum_j (\gamma_{j,k}^{\text{theo}})^i b_j^{\text{theo}} \right) \right]^2}{(\delta_{\text{sta},k}^i)^2 \mu_k^i (m^i - \sum_j \gamma_{j,k}^i m^i b_j - \sum_j (\gamma_{j,k}^{\text{theo}})^i m^i b_j^{\text{theo}}) + (\delta_{\text{unc},k}^i m^i)^2} + \sum_j b_j^2 + \sum_j (b_j^{\text{theo}})^2$$

→ Profiling!

Minimization allows Eigenvectors to be shifted such that difference between data and MC becomes minimal (with punishment factor, if shift is too big)

After the minimization, the eigenvectors are correlated (because the data are not necessarily affecting only single eigenvectors) → need to uncorrelate to be able to relate shifts in the eigenvectors to parton distributions

Successfully used in ATLAS W+c measurement

<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2012-14/>



# Switch on Profiling in HeraFitter

- \* Option in the steering card:
- \* (Optional) LHAPDF steering card
- \*

&lhapdf

**LHAPDFErrors = true**

LHAPDFSET = 'CT10nlo' ! LHAPDF grid file

ILHAPDFSET = 0 ! Set number withing PdfSet

&End

Now run:

```
./bin/FitPDF
```

Output: Going through the PDF eigenvectors one-by-one

-----  
Start LHAPDFErrors chi2 evaluation  
-----

LHAPDF 6.1.4 loading /home/kristin/PDFGenerator/share/LHAPDF/CT10nlo/CT10nlo\_0000.dat

CT10nlo PDF set, member #0, version 4; LHAPDF ID = 11000

Number of PDF members for this set: 52

NDiag= 30 NCovar= 0

PDF set number: 0 chi2=33.61 ndf=30

----- in store-pdfs -----

LHAPDF 6.1.4 loading /home/kristin/PDFGenerator/share/LHAPDF/CT10nlo/CT10nlo\_0001.dat



# Output of Profiling

All output is in the output directory:

```
.....  
-rw-rw-r-- 1 kristin kristin 20K Sep 30 10:19 pdfs_q2val_s26p_06.txt  
-rw-rw-r-- 1 kristin kristin 20K Sep 30 10:19 pdfs_q2val_s26p_05.txt  
-rw-rw-r-- 1 kristin kristin 6,7K Sep 30 10:19 theo_03.dat  
-rw-rw-r-- 1 kristin kristin 8,7K Sep 30 10:19 theo_02.dat  
-rw-rw-r-- 1 kristin kristin 8,7K Sep 30 10:19 theo_01.dat  
-rw-rw-r-- 1 kristin kristin 3,0K Sep 30 10:19 Results_00.txt  
-rw-rw-r-- 1 kristin kristin 5,1K Sep 30 10:19 Results.txt  
-rw-rw-r-- 1 kristin kristin 5,4K Sep 30 10:19 pdf_vector_cor.dat  
-rw-rw-r-- 1 kristin kristin 5,7K Sep 30 10:19 fittedresults.txt
```

Let's have a look at the most important ones:

```
less output/Results.txt
```



# Output of Profiling

After minimisation 31.03 30 1.034

Partial chi2s

Dataset 1 15.78 11 ATLAS W+ lepton pseudorapidity, 2010 data  
 Dataset 2 9.92 11 ATLAS W- lepton pseudorapidity, 2010 data  
 Dataset 3 2.61 8 ATLAS Z rapidity, 2010 data

Correlated Chi2 2.7130499552987168  
 Systematic shifts 57

	Name	Shift	+/-	Error	Type
1	ATL_WZ2010_Source_1		0.1268	+/- 0.9937	:N:M
2	ATL_WZ2010_Source_2		-0.1333	+/- 0.9830	:N:M
3	ATL_WZ2010_Source_3		-0.1300	+/- 0.9947	:N:M

.....

31	ATLAS_lumi2010		0.0069	+/- 0.7265	:N:M
32	PDF_nuisance_param_01		-0.1498	+/- 0.9229	:N:M
33	PDF_nuisance_param_02		0.0950	+/- 0.96	
34	PDF_nuisance_param_03		-0.0550	+/- 0.95	
35	PDF_nuisance_param_04		-0.1259	+/- 0.78	
36	PDF_nuisance_param_05		0.2021	+/- 0.96	

	CT10	MSTW2008	HERAPDF1.5
$W^+c\text{-jet } (\chi^2/\text{ndof})$	3.8/11	6.1/11	3.5/11
$W^-c\text{-jet } (\chi^2/\text{ndof})$	9.0/11	10.3/11	8.3/11
$W^+D^- (\chi^2/\text{ndof})$	3.6/4	3.7/4	3.7/4
$W^-D^+ (\chi^2/\text{ndof})$	3.7/4	4.6/4	3.3/4
$W^+D^{*-} (\chi^2/\text{ndof})$	2.9/4	6.0/4	2.2/4
$W^-D^{*+} (\chi^2/\text{ndof})$	3.0/4	4.4/4	2.4/4
$N_{\text{exp}}$	114	114	114
$N_{\text{theo}}$	28	22	16
Correlated $\chi^2$ (exp)	0.8	1.8	0.9
Correlated $\chi^2$ (theo)	6.2	1.9	2.6
Correlated $\chi^2$ (scale)	0.6	2.5	1.1
Total $\chi^2/\text{ndof}$	33.6/38	41.3/38	28.0/38

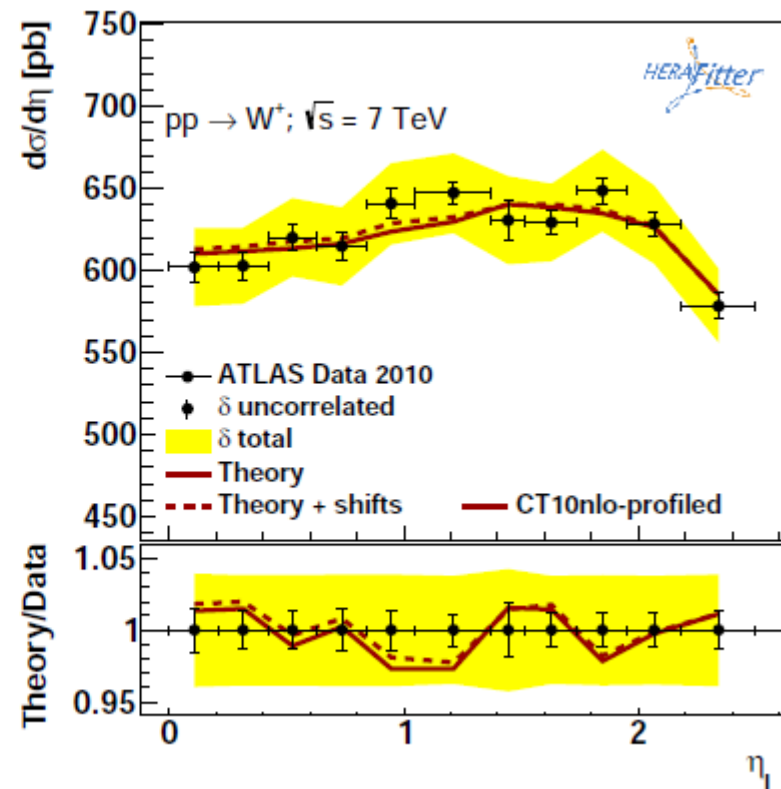
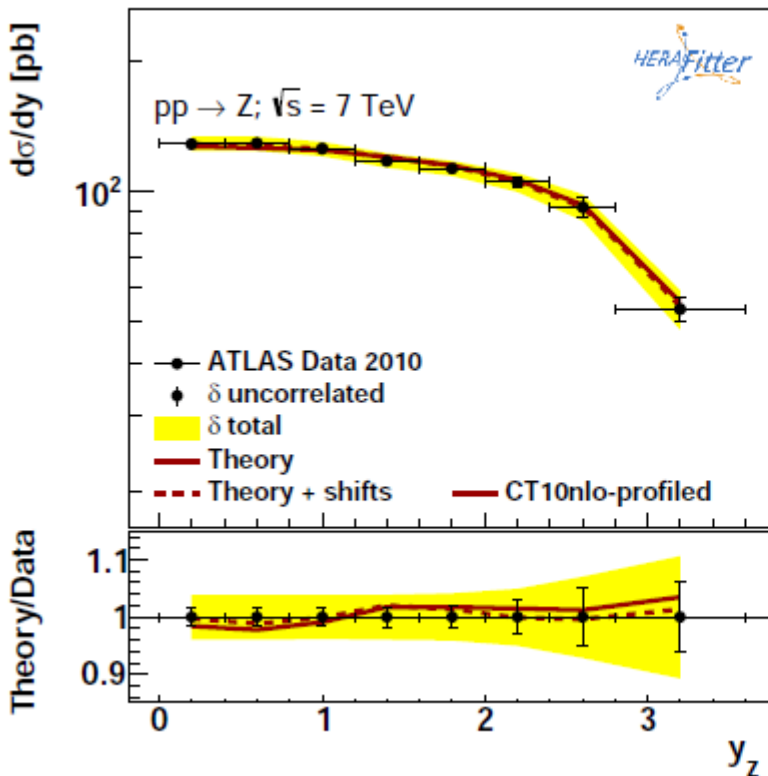
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2012-14/>



# Plotting of Profiling

For plotting can use normal HeraFitter Drawing:

```
./bin/DrawPdfs output/:CT10nnlo profiled:output/:CT10nnlo-profiled --asym --bands --therr --no-version
```





# Discussion

---