HERAFitter tutorial

PDF determination using LHC data

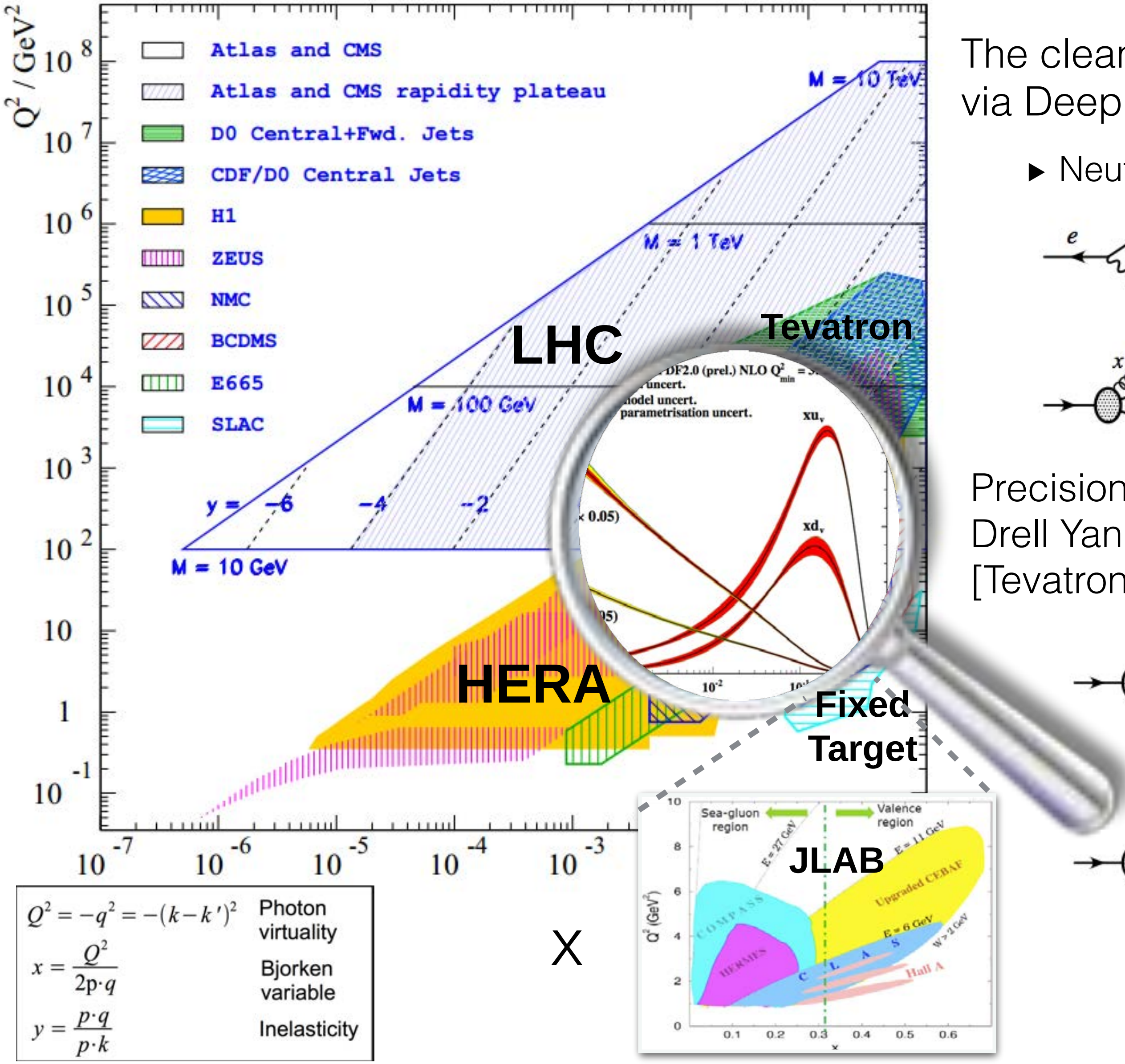
Voica Radescu

This tutorial:

- how to use HERAFitter with external predictions
- how to use HERAFitter with in-built theory predictions
- how to check impact of new data on PDFs:
  - valence quarks: W asymmetry
  - light quarks: W, Z data
  - gluon: jet, top data
- Summary

Prerequisites:
-> herafitter, qcdnum, (lhapdf) and applgrid installed
-> all export paths defined (usually stored in setup.txt)

Proton Structure Measurements



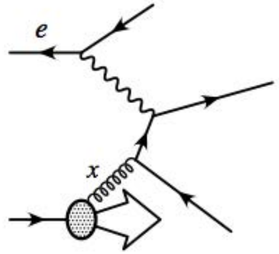
$$Q^2 = -q^2 = -(k-k')^2$$ Photon virtuality

$$x = \frac{Q^2}{2p\cdot q}$$ Bjorken variable

$$y = \frac{p\cdot q}{p\cdot k}$$ Inelasticity

The cleanest way to probe Proton Structure is via Deep Inelastic Scattering:

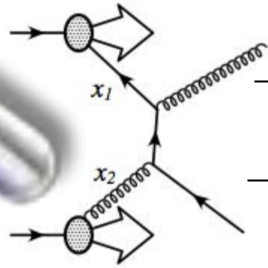
- Neutrinos, muons, electrons



—> probes linear combination of quarks

—> see tutorial by H. Pirumov

Precision of PDFs can be complemented by the Drell Yan processes at the collider experiments - [Tevatron and LHC]



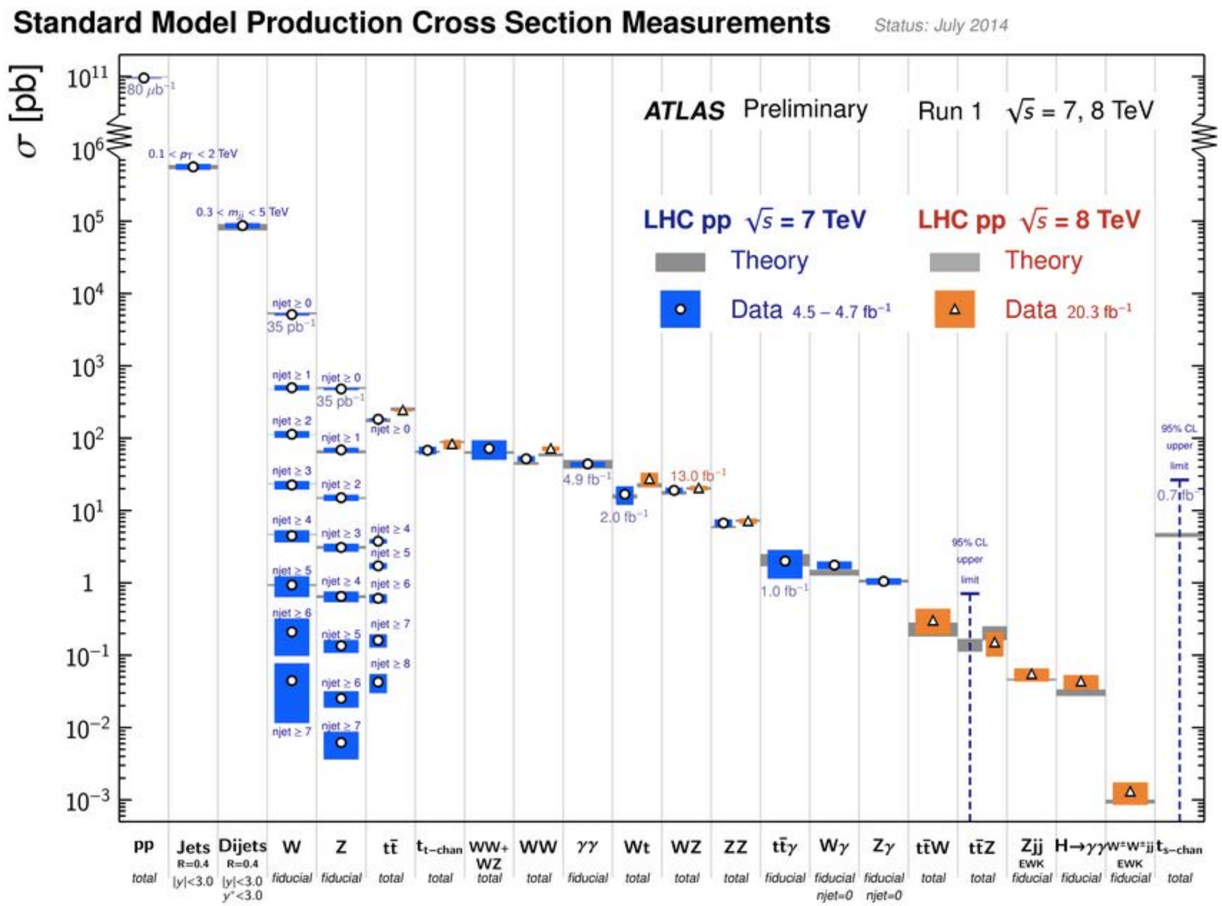
—> can provide flavour separation and more insight into gluons

—> probes bilinear combination of quarks

Different data constrain different parton combinations at different x, evolution with the scale is predicted by pQCD:

LHC measurements from RUN1

- Successful run in 2010 - 2012 at the LHC confirmed and tested SM



LHC can provide with its multitude of new measurements:

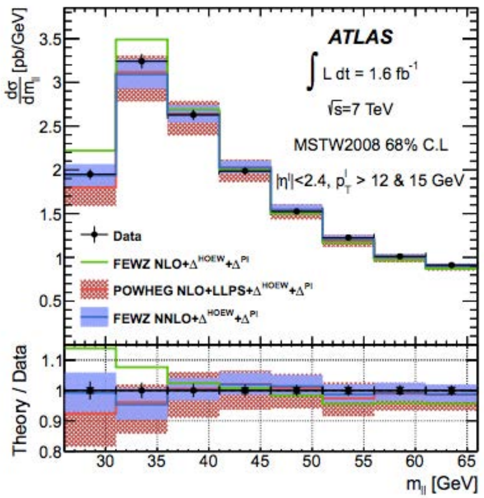
• PDF discrimination by confronting theory with data
• PDF improvement by using LHC data in QCD fit

1. W and Z production
2. W+c production
3. Drell-Yan: low and high invariant mass
4. Inclusive Jet, Di-Jet and Tri-jet production
5. Prompt Photon + Jets
6. Top, ttbar

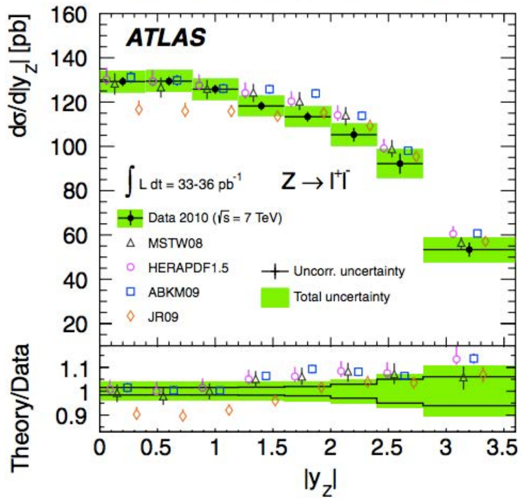
—> see A. Cooper-Sarkar's lecture

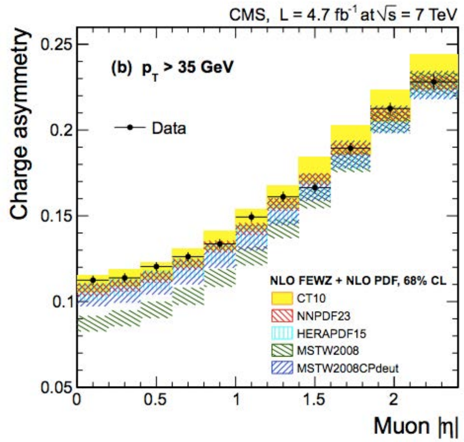
Data vs QCD predictions

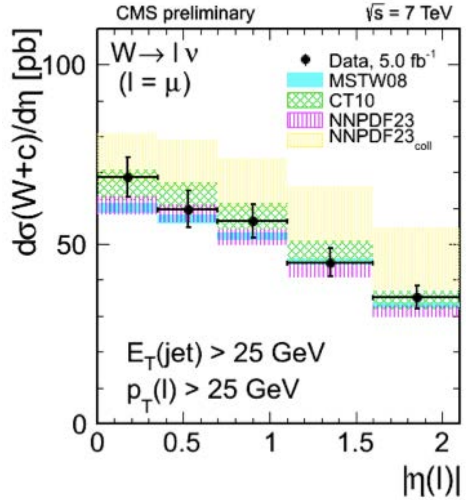
- We measure in order to confront theory with measurement:
  - example: Low and High Mass DY (ATLAS), W,Z inclusive (ATLAS), W asymmetry (CMS), inclusive jets …
  - data is compared to various theoretical predictions, most of the time very much computationally time consuming (FEWZ, DYNNLO, POWHEG, MCFM)

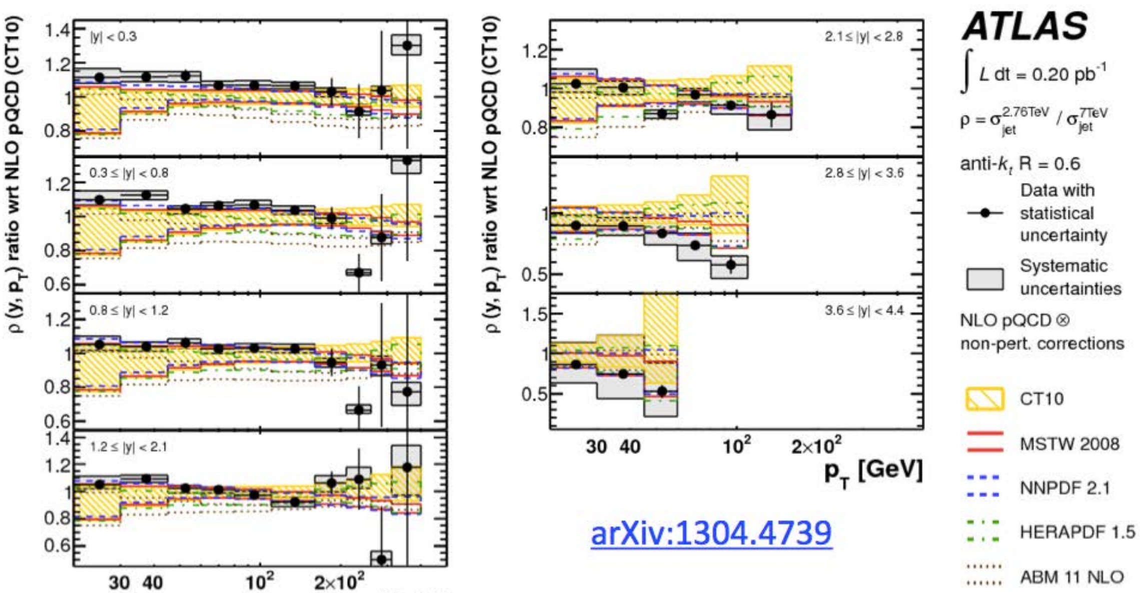


PHYSICAL REVIEW D **85**, 072004 (2012)



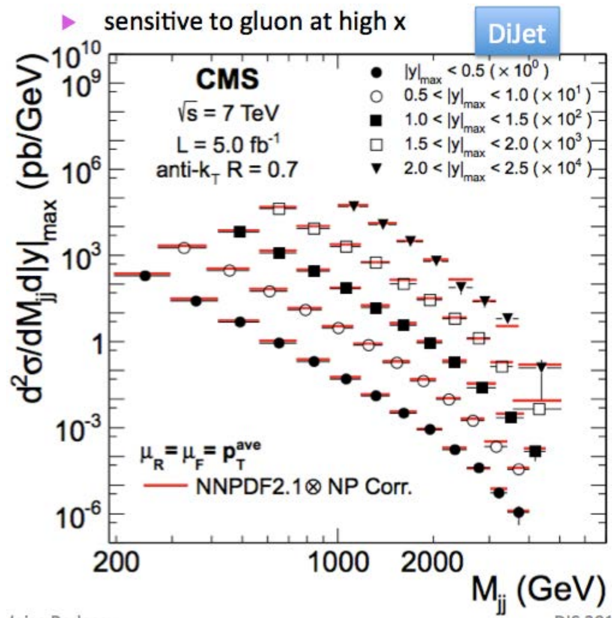






arXiv:1304.4739

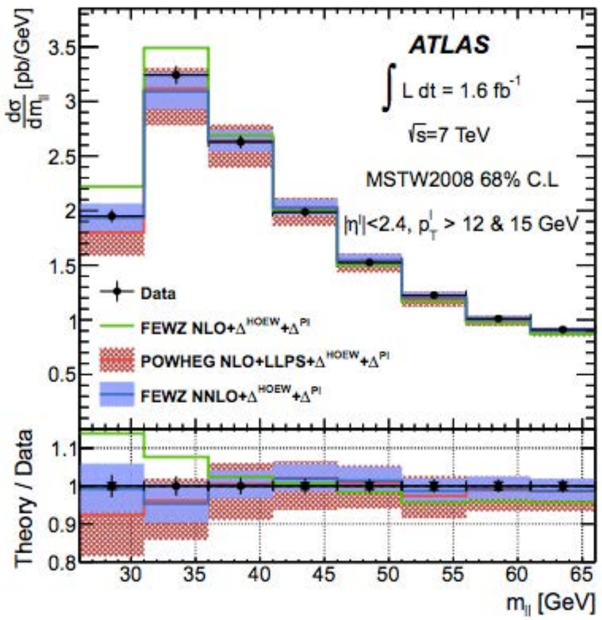
▸ sensitive to gluon at high x



Use of External Predictions

Data vs QCD predictions: hands on example

- We measure in order to confront theory with measurement:
  - **example: Low Mass DY ATLAS data http://arxiv.org/abs/arXiv:1404.1212**



Experiment: **CERN-LHC-ATLAS (ATLAS)**
Published in **JHEP 1406,112 (2014)** (DOI:10.1007/JHEP06(2014)112)
Preprinted as **CERN-PH-EP-2014-020**
Archived as: **ARXIV:1404.1212**

$m_{\ell\ell}$ [GeV]	POWHEG $\frac{d\sigma}{dm_{\ell\ell}}$ [pb/GeV]	POWHEG δ^{pdf} [%]	POWHEG δ^{scale} [%]	FEWZ NLO $\frac{d\sigma}{dm_{\ell\ell}}$ [pb/GeV]	FEWZ NLO δ^{pdf} [%]	FEWZ NLO δ^{scale} [%]	FEWZ NNLO $\frac{d\sigma}{dm_{\ell\ell}}$ [pb/GeV]	FEWZ NNLO δ^{pdf} [%]	FEWZ NNLO δ^{scale} [%]
26–31	1.80	2.5	$^{+7.3}_{-11.4}$	2.22	2.7	$^{+4.9}_{-7.9}$	1.93	$^{+3.5}_{-2.7}$	5.7
31–36	3.12	2.4	$^{+5.3}_{-10.0}$	3.49	2.7	$^{+4.7}_{-6.3}$	3.04	$^{+3.2}_{-2.5}$	4.5
36–41	2.64	2.3	$^{+4.6}_{-8.8}$	2.69	2.6	$^{+4.1}_{-5.0}$	2.58	$^{+3.1}_{-2.4}$	2.3
41–46	2.03	2.2	$^{+3.5}_{-7.5}$	2.00	2.6	$^{+3.6}_{-4.2}$	1.98	$^{+3.1}_{-2.3}$	2.1
46–51	1.54	1.9	$^{+3.7}_{-6.1}$	1.50	2.5	$^{+3.2}_{-3.5}$	1.51	$^{+3.0}_{-2.2}$	1.7
51–56	1.19	2.4	$^{+4.5}_{-5.1}$	1.17	2.4	$^{+2.8}_{-2.9}$	1.18	$^{+2.9}_{-2.2}$	1.3
56–61	1.00	2.4	$^{+2.3}_{-4.7}$	0.97	2.4	$^{+2.6}_{-2.6}$	0.98	$^{+2.9}_{-2.1}$	1.3
61–66	0.90	2.1	$^{+2.0}_{-4.5}$	0.87	2.3	$^{+2.3}_{-2.3}$	0.88	$^{+2.8}_{-2.1}$	1.2

Table 8. Theory predictions for NLO+LLPS and for fixed-order calculations at NLO and NNLO including higher-order electroweak corrections, for the nominal analysis of the differential cross section $\frac{d\sigma}{dm_{\ell\ell}}$ as a function of the invariant mass $m_{\ell\ell}$. The scale uncertainty is defined as the envelope of variations for $0.5 \leq \mu_R, \mu_F \leq 2$ for POWHEG. For FEWZ the scale uncertainty is defined by the variation $0.5 \leq \mu_R = \mu_F \leq 2$.

Prediction	χ^2 (8 points) Nominal
POWHEG NLO+LLPS	22.4 (19.8)
FEWZ NLO	48.7 (28.6)
FEWZ NNLO	13.9 (12.9)

One needs NNLO calculation to describe data

Exercise: Use of external predictions

- HERAFitter can estimate quantitatively level of agreement between data and theory using external to HERAFitter predictions, but taking properly into account all the correlations (experimental or theoretical):

in **herafitter** directory:

```
> wget http://www.desy.de/~voica/LM_PDFschool.tgz
> tar -xzvf LM_PDFschool.tgz.tgz
              dy_lowmass_nominal.dat

              theo_powheg_nlo_nll.dat
              theo_fewz_nlo.dat
              theo_fewz_nnlo.dat


              kf.nominal.NNLO-NLOEW.txt
              kf.nominal.PI.txt
              kf.nominal.scheme.txt
              low_fidu.root
> cp LowMass_school/dy_lowmass_nominal.dat datafiles/.
> cp LowMass_school/theo* .
> cp LowMass_school/kf* theoryfiles/.
> cp LowMass_school/low_fidu.root theoryfiles/.
```

—> data format: bins, sigma, correlations

—> external theory predictions

```
&Data
Name = "Theory for ATLAS LM DY nominal"
NData = 8
NColumn = 6
ColumnType =  2*"Bin","Theory", 3*"Error"
ColumnName = "eta1","eta2","theory","PDF","scale-", "scale+"
Percent = 28*True
&End
 26 31  1.80    2.5     -11.4   7.3
 31 36  3.12    2.4     -10.0   5.3
 36 41  2.64    2.3     -8.8    4.6
 41 46  2.03    2.2     -7.5    3.5
 46 51  1.54    1.9     -6.1    3.7
 51 56  1.19    2.4     -5.1    4.5
 56 61  1.00    2.4     -4.7    2.3
 61 66  0.90    2.1     -4.5    2.0
```

$$\chi^2 = \sum_i \left(\frac{\mu_i - m_i\left[1+\sum_j b_j^{\mathrm{exp}}\gamma_{ji}^{\mathrm{exp}} + \sum_j b_j^{\mathrm{theo}}\gamma_{ji}^{\mathrm{theo}}\right]}{\Delta_i}\right)^2 + \sum_j (b_j^{\mathrm{exp}})^2 + \sum_j (b_j^{\mathrm{theo}})^2$$

-> accounts also for theory uncertainties

Use of external predictions:

- HERAFitter can estimate quantitatively level of agreement between data and theory using external to HERAFitter predictions, but taking properly into account all the correlations (experimental or theoretical):

```
> wget http://www.desy.de/~voica/LM_PDFschool.tgz
> tar -xzvf LM_PDFschool.tgz
           dy_lowmass_nominal.dat

           theo_powheg_nlo_nll.dat
           theo_fewz_nlo.dat
           theo_fewz_nnlo.dat
```

—> data format: bins, sigma, correlations

```
&DATA
 Name = 'ATLAS low mass DY 2011'
 Ndata  = 8
 NColumn   = 21
 ColumnType = 'Flag', 2*'Bin','Sigma',17*'Error'
 ColumnName = 'binFlag', 'mass1','mass2','Sigma', 'stat','uncor','ignore', 'LMDYcorr1','LMDYcorr2','LMDYcorr3','LMDYcorr4', 'LMDY
corr5', 'LMDYcorr6', 'LMDYcorr7', 'LMDYcorr8', 'LMDYcorr9', 'LMDYcorr10', 'LMDYcorr11', 'LMDYcorr12', 'LMDYcorr13', 'ALumi2011'

 NInfo    = 1
 CInfo    = 'theoryunit'
 DataInfo = 100000.

 IndexDataset = 177
 Reaction = 'NC pp'
 TheoryType     = 'expression'
 TermName = 'A1', 'K1', 'K2', 'K3'
 TermType = 'applgrid','kfactor', 'kfactor',  'kfactor'
 TermSource = 'theoryfiles/low_fidu.root' ,
              'theoryfiles/kf.nominal.NNLO-NLOEW.txt',
              'theortfiles/kf.nominal.PI.txt'
              'theoryfiles/kf.nominal.scheme.txt'
! TheorExpr= 'K1*K3*A1+K2'
 TheorExpr= 'A1'
 PERCENT = 3*F, 32*T
&END
&PlotDesc
   PlotN = 1
   PlotDefColumn = 'mass2'
   PlotDefValue = 0., 100.
   PlotOptions(1)  = 'Experiment:ATLAS@ExtraLabel:pp #rightarrow l^{+}l^{-}; #sqrt{s} = 7 TeV; #int L = 4.7 fb^{-1}@XTitle: M_{ll
} @YTitle: d#sigma/dM_{ll} [pb]'
&End
*M_ll        CS[pb/GeV]   Stat[pb/GeV]   Ucorr[pb/GeV]  Total[pb/Gev]  corr1[%] corr2[%] corr3[%] corr4[%] corr5[%] corr6[%] corr7
[%]  corr8[%]  corr9[%]  corr10[%] corr11[%] corr12[%] corr13[%]
1 26 31  0.1949922E+01  0.1745244E-01  0.3118785E-01  0.5818630E-01   0.0927   0.3564  -1.2121   0.6657  -0.3506  -0.5873   0.361
4    0.4552   -1.2861   -0.0180   -0.6285   -0.2746    0.7527  1.8
1 31 36  0.3242065E+01  0.2299554E-01  0.4433022E-01  0.8363228E-01   0.1164   0.3139  -1.1332   0.5640  -0.2858  -0.4449   0.248
2    0.2257   -1.1202   -0.3932   -0.4162   -0.3867    0.6656  1.8
1 36 41  0.2628920E+01  0.2020543E-01  0.3110829E-01  0.5847444E-01   0.1653   0.2288  -1.0318   0.4736  -0.2074  -0.2181   0.277
3    0.2716   -0.7756   -0.5580   -0.1696   -0.3466    0.5337  1.8
1 41 46  0.1986611E+01  0.1732874E-01  0.2202702E-01  0.3942961E-01   0.1693   0.2256  -0.9876   0.4139  -0.1937  -0.0176   0.286
8    0.3539   -0.4607   -0.2136   -0.1968   -0.0423    0.4232  1.8
1 46 51  0.1524534E+01  0.1434589E-01  0.1632578E-01  0.2855119E-01   0.2313   0.3030  -0.8225   0.3729  -0.1196   0.0744   0.202
9    0.2516   -0.3645   -0.3176   -0.0342   -0.1558    0.3616  1.8
1 51 56  0.1225001E+01  0.1228500E-01  0.1270887E-01  0.2195444E-01   0.2222   0.2937  -0.7972   0.3107  -0.1351   0.1366   0.157
6    0.2067   -0.2168   -0.0022   -0.1906    0.0683    0.2643  1.8
1 56 61  0.1010638E+01  0.1055659E-01  0.9674996E-02  0.1736092E-01   0.2550   0.2660  -0.7256   0.2731  -0.1059   0.1520   0.156
7    0.2105   -0.1650   -0.1355   -0.0563   -0.0509    0.2279  1.8
1 61 66  0.9106037E+00  0.9330309E-02  0.5891908E-02  0.1492615E-01   0.3085   0.2687  -0.6217   0.2550  -0.0458   0.1620   0.052
```

# of data points

each correlated systematic source with its identifier tag name

A: applgrid (here interfaced to MCFM, NLO)
K: kfactors
TheoExpr: ->adjusted to match the theory one wants to build

how to read the uncertainties: in % (T) or absolute (F)

Use of external predictions:

- HERAFitter can estimate quantitatively level of agreement between data and theory using external to HERAFitter predictions, but taking properly into account all the correlations (experimental or theoretical):
- In `steering.txt`, add the following lines:

```
&InFiles
 ! Number of intput files
   NInputFiles = 1
   InputFileNames(1) = ‘datafiles/dy_lowmass_nominal.dat’
 !  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

 &InTheory
    InputTheoNames(1) = 'theo_powheg_nlo_nll.dat'
 !InputTheoNames(1) = 'theo_fewz_nlo.dat'
 !InputTheoNames(1) = 'theo_fewz_nnlo.dat'
 &End

  CHI2SettingsName = 'StatScale', 'UncorSysScale', 'CorSysScale', 'UncorChi2Type', 'CorChi2Type'
  Chi2Settings     = 'Poisson'  , 'Linear',        'Linear'     , 'Diagonal'     , 'Hessian'
  Chi2ExtraParam = 'PoissonCorr'
```

$$\chi^2_{tot}(\mathbf{m},\mathbf{b}) = \sum_i \frac{[\mu^i - m^i(1-\sum_j \gamma^i_j b_j)]^2}{\delta^2_{i,stat}\mu^i m^i(1-\sum_j \gamma^i_j b_j) + (\delta_{i,unc} m^i)^2} + \sum_j b_j^2 + \sum_i \ln \frac{\delta^2_{i,unc} m_i^2 + \delta^2_{i,stat}\mu^i m^i}{\delta^2_{i,unc}\mu_i^2 + \delta^2_{i,stat}\mu_i^2}$$

modified χ^2 definition includes ln term to account for likelihood transition to χ^2 after error scaling

- `make clean`
- `./configure --enable-applgrid  ! not necessary to have it for this exercise`
- `make && make install`
- `bin/FitPDF > powheg.log`

Use of external predictions: output

> bin/FitPDF

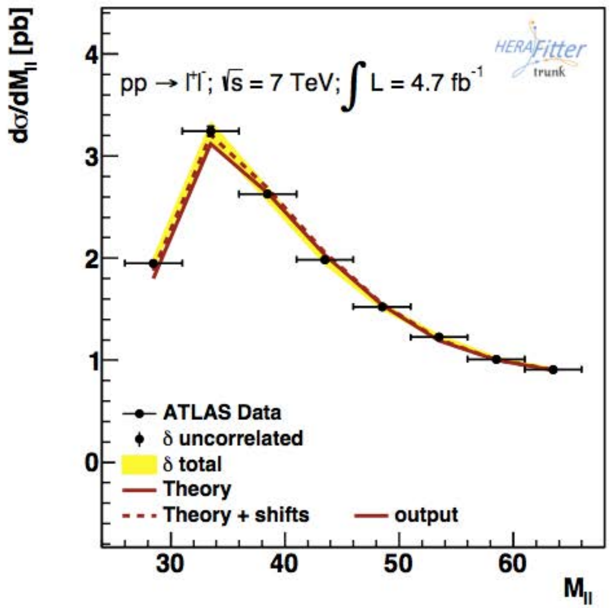
Ignore these, in this case dof = number of points, however HERAFitter sees the minuit file and counts these as free parameters

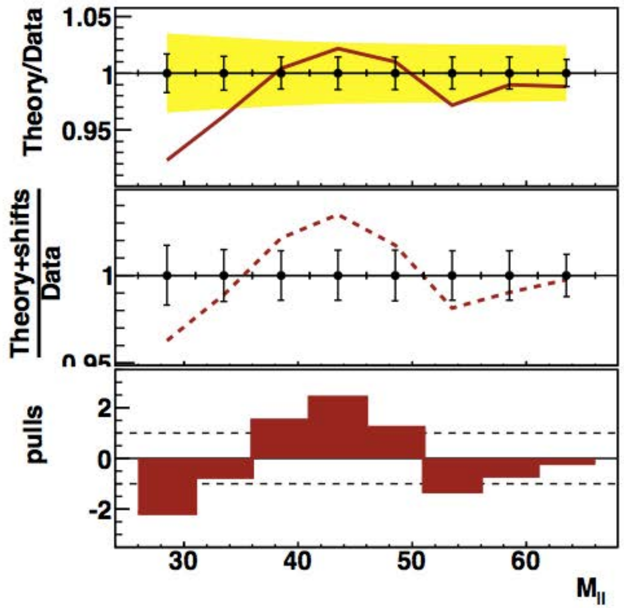
```
After minimisation       19.76     -2     -9.878

Dataset     1      17.01( -0.23)      8  ATLAS low mass DY 2011
 Correlated Chi2     2.9752322813730916
 Log penalty Chi2  -0.22516399388497274
 --------- in store-pdfs -------
cpu_time        0.57        0.92        0.35
```

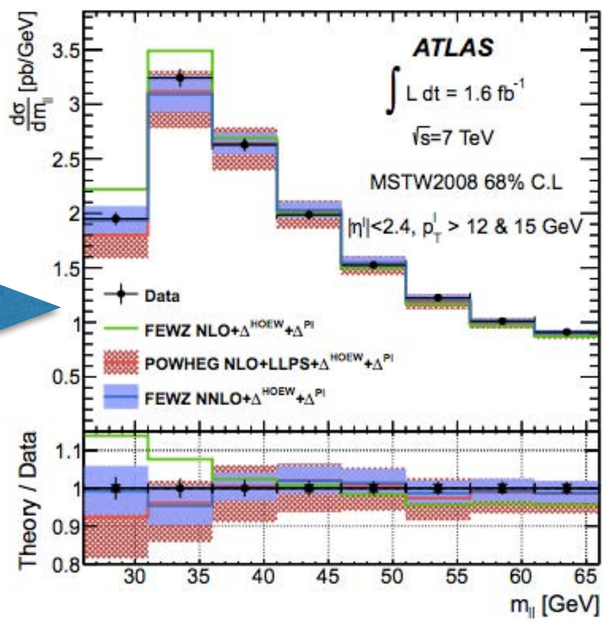
```
>ls output
Results.txt        lhapdf.block.txt  parsout_0          pdfs_q2val_02.txt  pdfs_q2val_05.txt  pulls.last.txt
fittedresults.txt  minuit.out.txt    parsout_1          pdfs_q2val_03.txt  pdfs_q2val_06.txt
herapdf            minuit.save.txt   pdfs_q2val_01.txt  pdfs_q2val_04.txt  pulls.first.txt
```

```
> bin/DrawPdfs --no-pdfs --3panels output
```

For more options: bin/DrawPdfs - - help







Use of external predictions: output

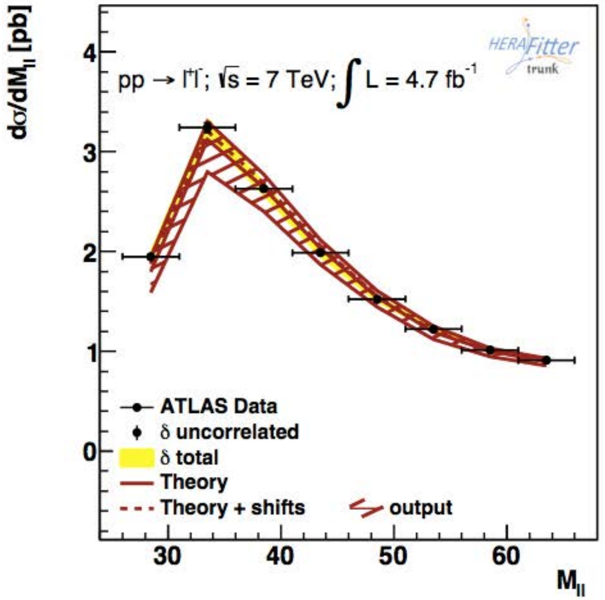
```
> bin/FitPDF
```

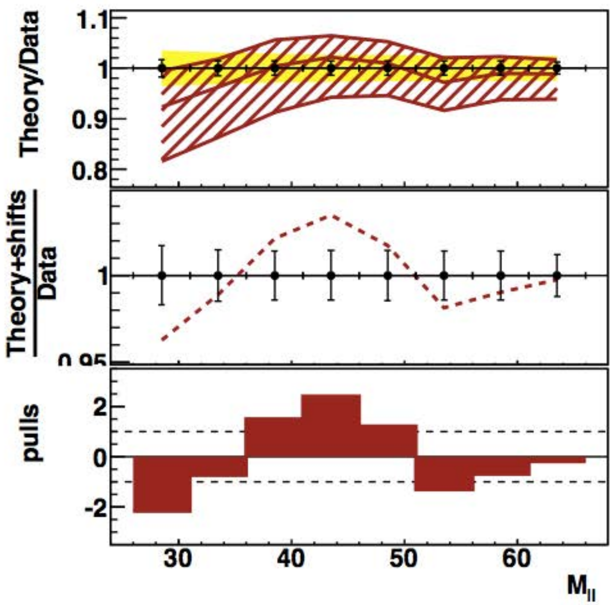
```
After minimisation      19.76    -2     -9.878

Dataset     1      17.01( -0.23)      8  ATLAS low mass DY 2011
 Correlated Chi2     2.9752322813730916
 Log penalty Chi2  -0.22516399388497274
 --------- in store-pdfs --------
cpu_time       0.57        0.92       0.35
```

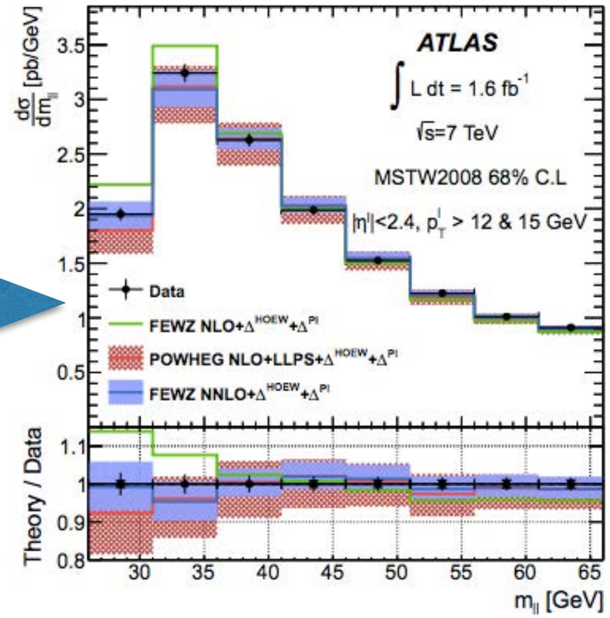
```
>ls output
Results.txt          lhapdf.block.txt   parsout_0          pdfs_q2val_02.txt  pdfs_q2val_05.txt  pulls.last.txt
fittedresults.txt    minuit.out.txt     parsout_1          pdfs_q2val_03.txt  pdfs_q2val_06.txt
herapdf              minuit.save.txt    pdfs_q2val_01.txt  pdfs_q2val_04.txt  pulls.first.txt
```

```
> bin/DrawPdfs --no-pdfs --therr --3panels output
```

For more options: bin/DrawPdfs —help







Use of external predictions: output

```
> bin/FitPDF
```

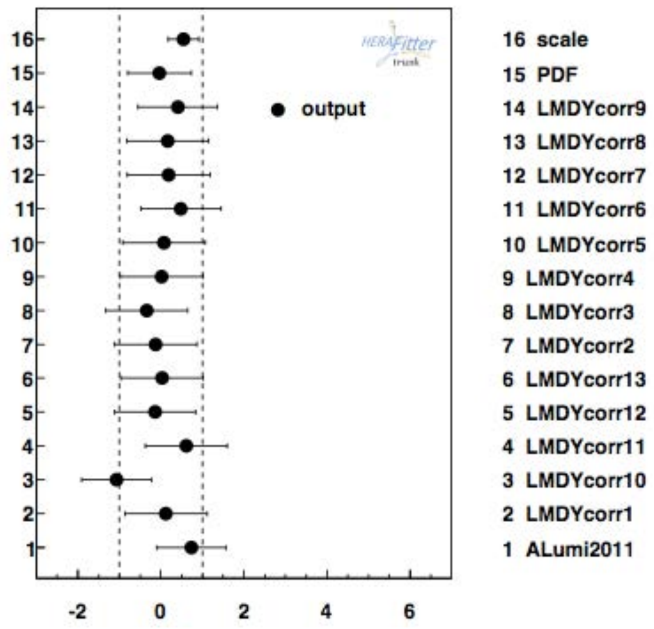
```
After minimisation       19.76    -2      -9.878

Dataset     1      17.01( -0.23)      8  ATLAS low mass DY 2011
 Correlated Chi2     2.9752322813730916
 Log penalty Chi2  -0.22516399388497274
 --------- in store-pdfs --------
cpu_time       0.57       0.92       0.35
```

```
>ls output
Results.txt          lhapdf.block.txt   parsout_0            pdfs_q2val_02.txt   pdfs_q2val_05.txt   pulls.last.txt
fittedresults.txt    minuit.out.txt     parsout_1            pdfs_q2val_03.txt   pdfs_q2val_06.txt
herapdf              minuit.save.txt    pdfs_q2val_01.txt    pdfs_q2val_04.txt   pulls.first.txt
```

```
> bin/DrawPdfs --no-pdfs --therr --3panels output
```

For more options: bin/DrawPdfs —help

Dataset	output
ATLAS low mass DY 2011	17 / 8
Correlated χ^2	3.0
Log penalty χ^2	-0.23
Total χ^2 / dof	20 / -2
χ^2 p-value	0.00



Use of external predictions: output

```
> bin/FitPDF > powheg.log
```

```
After minimisation       19.76     -2     -9.878

Dataset     1      17.01( -0.23)      8  ATLAS low mass DY 2011
 Correlated Chi2     2.9752322813730916
 Log penalty Chi2  -0.22516399388497274
 --------- in store-pdfs -------
cpu_time        0.57       0.92       0.35
```

```
>ls output
Results.txt         lhapdf.block.txt  parsout_0          pdfs_q2val_02.txt  pdfs_q2val_05.txt  pulls.last.txt
fittedresults.txt   minuit.out.txt    parsout_1          pdfs_q2val_03.txt  pdfs_q2val_06.txt
herapdf             minuit.save.txt   pdfs_q2val_01.txt  pdfs_q2val_04.txt  pulls.first.txt

> bin/DrawPdfs --no-pdfs --therr --3panels output
```

Homework:

- re-run now for FEWZ NLO and NNLO and check if you can reach the paper conclusions
- Hint: adjust the steering.txt and save your outputs into separate directories, draw all 3 cases with

```
>bin/DrawPdfs --no-pdfs --therr output1:POWHEG-NLO output2:FEWZ_NLO output3:FEWZ_NNLO
```

PDF from QCD fits

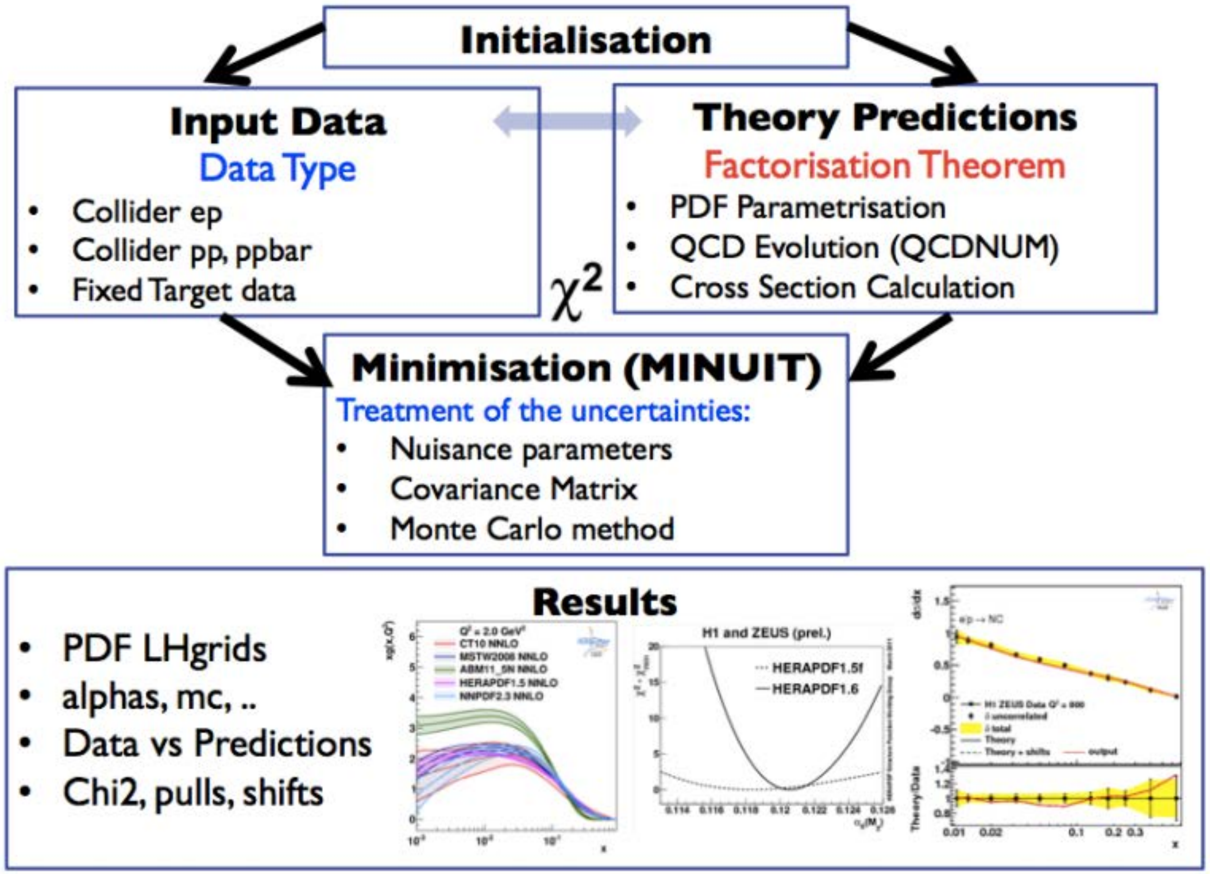
Extraction of PDFs through QCD fits

Typical measurements sensitive to PDFs are precise, with statistical uncertainties < 10%, so they follow normal distribution which allows the use of chi square minimisation for PDF extraction.

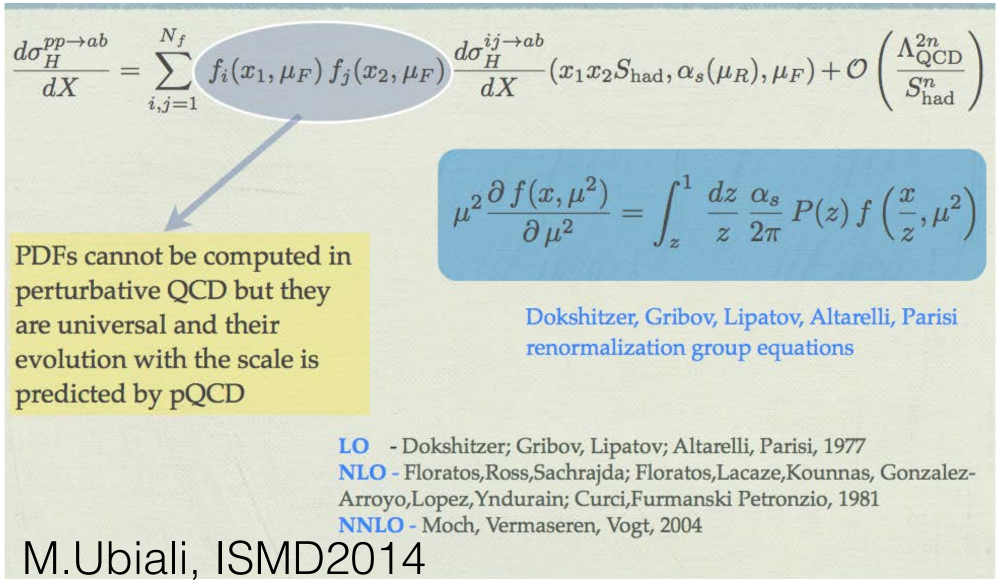
HERAFitter www.herafitter.org

Main Steps:

- Parametrise PDFs at starting scale
- Evolve to the scale corresponding to data point
- Calculate the cross section
- Compare with data via chi2
- Minimize chi2 with respect to PDF parameters



Building Internal Predictions
-> rely on factorisation theorem



HERAFitter contains already some ready-to-use files for QCD Fit analyses (datafiles and theoryfiles):

-> ATLAS: WZ2010, Jets2010_7TeV, Jets2010_2p76TeV, ttbar_7TeV, ..

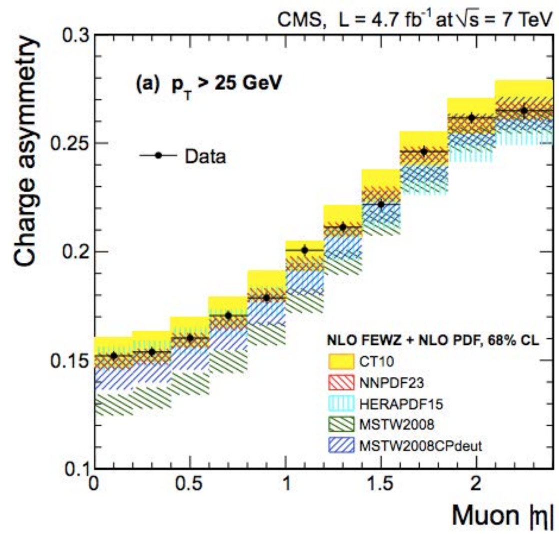
-> CMS: CMS_mAsymmetry, CMS_eAsymmetry, CMS_Z_boson_Rapidity, …

Building internal predictions for QCD fits @NLO

- We learned that we rely on factorisation theorem to simplify our problems:
  - Fast techniques developed: fastNLO, APPLGRID, k-factors … that help speeding up the calculation process:
    - in the fit we deal with ~2000 iterations
    - HERAFitter can provide within 15min first feedback on the general tendency/impact of data
- **For NLO QCD predictions: we often rely on APPLGRID or fastNLO (a matter of choice)**
  - EX: CMS W asymmetry:

```
> less datafiles/lhc/cms/CMS_mAsymmetry_SMP_12_021.dat
```

```
* The CMS W muon assymetry data (CMS_SMP_12_021)
* Publication: arXiv:1312.6283, submitted to PRD
*
*
&Data
   Name = 'CMS W muon asymmetry'

 ! Description of the data table
   NDATA = 11
   NColumn = 19
   ColumnType = 'Flag',2*'Bin','Sigma',15*'Error'  ! 2 bins, Asymmetry, errors
   ColumnName = 'binFlag','eta1','eta2','Asymmetry','stat const',14*'uncor const'
 ! ColumnName = 'binFlag','eta1','eta2','Asymmetry','stat',14*'ignore'
 ! ColumnName = 'binFlag','eta1','eta2','Asymmetry','stat','CMS_WmuAsym_eff','CMS_WmuAsym_QCDpm','CMS_WmuAsym_QC
shape','CMS_WmuAsym_muScale','CMS_WmuAsym_FSR','CMS_WmuAsym_PDF','CMS_WmuAsym_DYbkg','CMS_WmuAsym_etaPhiMod','CM
_WmuAsym_recoil','CMS_WmuAsym_PU','CMS_WmuAsym_lumi', 'CMS_WmuAsym_ttbar', 'CMS_WmuAsym_wtau', 'CMS_WmuAsym_qt'

   NInfo    = 5
   DataInfo = 7000.,  1., 25., 0., 1
   CInfo    = 'sqrt(S)', 'asymmetry', 'ptm cut', 'ptnu cut', 'theoryunit'

   IndexDataset = 245

   Reaction  = 'CC pp'

   TheoryType     = 'expression'
   TermName = 'A1', 'A2'
   TermType = 'applgrid','applgrid'
   TermSource = 'theoryfiles/cms/SMP_12_021/Wplus_applgrid_Pt25.root',
                'theoryfiles/cms/SMP_12_021/Wminus_applgrid_Pt25.root'
   TheorExpr= '(A1-A2)/(A1+A2)'

   Percent   = 15*F

&End
```



$$\mathcal{A}(\eta) = \frac{\frac{d\sigma}{d\eta}(W^+ \to \ell^+\nu) - \frac{d\sigma}{d\eta}(W^- \to \ell^-\bar{\nu})}{\frac{d\sigma}{d\eta}(W^+ \to \ell^+\nu) + \frac{d\sigma}{d\eta}(W^- \to \ell^-\bar{\nu})},$$

Theoretically, the prediction is build from 2 applgrid files

Building internal predictions for QCD fits @NLO

- We learned that we rely on factorisation theorem to simplify our problems:
  - Fast techniques developed: fastNLO, APPLGRID, k-factors … that help speeding up the calculation process:
    - in the fit we deal with ~2000 iterations
    - HERAFitter can provide within 15min first feedback on the general tendency/impact of data
- **For NLO QCD predictions: we often rely on APPLGRID or fastNLO (a matter of choice)**
  - EX: CMS W asymmetry:
  - Adjust the **`steering.txt`** file:

```
&InFiles
  ! Number of intput files
    NInputFiles = 5

  ! 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'
    InputFileNames(5) = 'datafiles/lhc/cms/CMS_mAsymmetry_SMP_12_021.dat'
&End

&InCorr
  ! Number of correlation (statistical, systematical or full) files
    NCorrFiles = 1
    CorrFileNames(1)= 'datafiles/lhc/cms/CMS_mAsymmetry_SMP_12_021___CMS_mAsymmetry_SMP_12_021.corr'
&End
```

The file comes with correlated matrix information

  - Adjust the **minuit.in.txt** file:

```
*set print 3
*call fcn 3
migrad 200000
*hesse
set print 3
```

—> this way we allow to have up to 200000 calls for fit to converge

Building internal predictions for QCD fits @NLO

- We learned that we rely on factorisation theorem to simplify our problems:
  - Fast techniques developed: fastNLO, APPLGRID, k-factors … that help speeding up the calculation process:
    - in the fit we deal with ~2000 iterations
    - HERAFitter can provide within 15min first feedback on the general tendency/impact of data
- **For NLO QCD predictions: we often rely on APPLGRID or fastNLO (a matter of choice)**
  - EX: CMS W asymmetry:
  - Adjust the **`steering.txt`** file:

```
&InFiles
  ! Number of intput files
    NInputFiles = 5

  ! Input files:
    InputFileNames(1) = 'datafiles/hera/H1
    InputFileNames(2) = 'datafiles/hera/H1
    InputFileNames(3) = 'datafiles/hera/H1
    InputFileNames(4) = 'datafiles/hera/H1
    InputFileNames(5) = 'datafiles/lhc/cms
&End

&InCorr
  ! Number of correlation (statistical, sy
    NCorrFiles = 1
    CorrFileNames(1)= 'datafiles/lhc/cms/C
&End
```

```
&StatCorr
   Name1 = 'CMS W muon asymmetry'
   Name2 = 'CMS W muon asymmetry'

   NIdColumns1 = 2
   NIdColumns2 = 2

   IdColumns1 = 'eta1','eta2'
   IdColumns2 = 'eta1','eta2'

   NBins1 = 11
   NBins2 = 11

!   NCorr = 121

    MatrixFormatIsTable = true

 ! Matrix Type:
 !   'Statistical correlations':     Given are correlation factors that need to be applied to statistical errors
 !                                    needs 'stat' column in data files
 !   'Systematic correlations' :     Given are correlation factors that need to be applied to systematic errors
 !                                    needs 'uncor' column in data files
 !   'Systematic covariance matrix': Given is a systematic covariance matrix
 !
 !   'Full covariance matrix':       Given is a full covariance matrix including stat and syst parts
 !
 !   'Full correlation matrix':      Given is a full correlation matrix including stat and syst parts

 !  For uncertainties provided in matrix format use the following format and options:
 !  bin1_min         bin1_max    [delta1_i]   c_11      c_12   c_13    ...
 !  bin2_min         bin2_max    [delta2_i]   c_21      c_22   c_23    ...
 !  ....
 !
 ! If full correlation matrix, provide also the delta_i

   MatrixType = 'Systematic correlations'

&End
0    0       0.0    0.2    0.4    0.6    0.8    1.0    1.2    1.4    1.6   1.85    2.1
0    0       0.2    0.4    0.6    0.8    1.0    1.2    1.4    1.6   1.85    2.1    2.4
0.0 0.2        1  0.281  0.324  0.329  0.271  0.290  0.295  0.280  0.305  0.261  0.167
0.2 0.4    0.281      1  0.307  0.314  0.256  0.275  0.279  0.263  0.289  0.245  0.158
0.4 0.6    0.324  0.307      1  0.374  0.309  0.338  0.345  0.321  0.361  0.303  0.193
0.6 0.8    0.329  0.314  0.374      1  0.311  0.340  0.344  0.320  0.363  0.304  0.200
0.8 1.0    0.271  0.256  0.309  0.311      1  0.285  0.295  0.280  0.312  0.269  0.173
1.0 1.2    0.290  0.275  0.338  0.340  0.285      1  0.326  0.311  0.348  0.302  0.193
1.2 1.4    0.295  0.279  0.345  0.344  0.295  0.326      1  0.328  0.369  0.322  0.208
1.4 1.6    0.280  0.263  0.321  0.320  0.280  0.311  0.328      1  0.360  0.327  0.213
1.6 1.85   0.305  0.289  0.361  0.363  0.312  0.348  0.369  0.360      1  0.371  0.249
1.85 2.1   0.261  0.245  0.303  0.304  0.269  0.302  0.322  0.327  0.371      1  0.244
2.1 2.4    0.167  0.158  0.193  0.200  0.173  0.193  0.208  0.213  0.249  0.244      1
```

The file comes with correlated matrix information

see R. Placakyte's tutorial for more info

Exercise:

- Save your previous output into a reference file:
  - for today we use RT FAST scheme:

```
> mv output Ref
> mkdir output
> bin/FitPDF
> less output/Results.txt
```

```
 First iteration     749.73335051215508            593    1.2643058187388787
After minimisation      608.79    593     1.027

  Partial chi2s
Dataset     1     105.71     145   NC cross section HERA-I H1-ZEUS combined e-p.
Dataset     2     429.99     379   NC cross section HERA-I H1-ZEUS combined e+p.
Dataset     3      19.84      34   CC cross section HERA-I H1-ZEUS combined e-p.
Dataset     4      37.39      34   CC cross section HERA-I H1-ZEUS combined e+p.
Dataset     5      15.85      11   CMS W muon asymmetry

 Correlated Chi2    0.0000000000000000
 Systematic shifts            0
```

  - **Homework:**

    A. Can you obtain a better fit by adding more free parameters. Hint: adjust this in the **`minuit.in.txt`** by releasing extra gluon, Ddv, Duv..

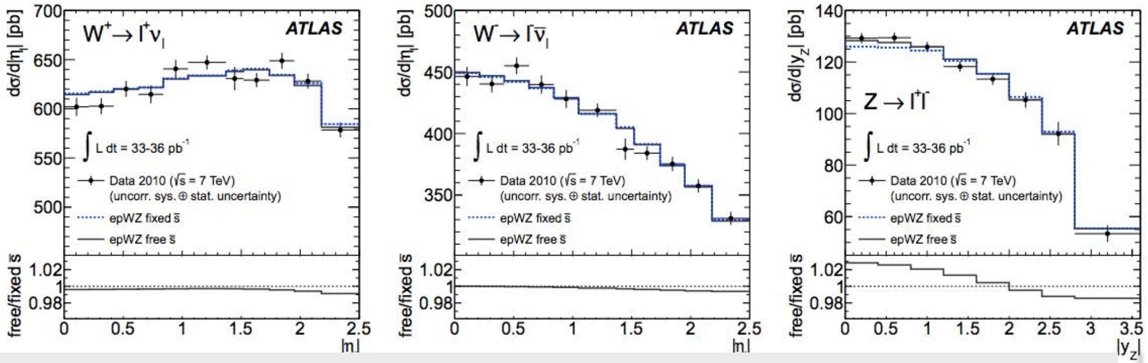
! be aware RT FAST is a shortcut option only to guide your intuition

Building internal predictions for QCD fits @ NNLO

- For NNLO QCD predictions: we apply **k-factors = [NNLO_fewz/NLO_fewz] * NLO_GRID(i)**
  - caveat: NLO EW corrections are also non-negligible:

kfactors_EW = NLO_EW/NLO_noEW or NNLO_EW_fewz/NLO_noEW_fewz

- EX: ATLAS W+, W-, Z inclusive differential measurements (2010)




```
> less datafiles/lhc/atlas/WZ2010/WP_applgrid_nnlo.dat:
> pico -w steering.txt:

&InFiles
  ! Number of input files
    NInputFiles = 7

  ! 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'
    InputFileNames(5) = 'datafiles/lhc/atlas/WZ2010/WP_applgrid_nnlo.dat'
    InputFileNames(6) = 'datafiles/lhc/atlas/WZ2010/WM_applgrid_nnlo.dat'
    InputFileNames(7) = 'datafiles/lhc/atlas/WZ2010/Z0_applgrid_nnlo.dat'

&End

&Scales
    DataSetMuR  = 4*1.0     ! Set muR scale to 1 for all 4 datasets
    DataSetMuF  = 4*1.0     ! Set muF scale to 1 for all 4 datasets
    DataSetIOrder = 4*3, 5*2
&End

  Order   = 'NNLO'
 HF_SCHEME = 'RT FAST'
```

```
TheoryType      = 'expression'
TermName = 'A1', 'K'
TermType = 'applgrid','kfactor'
TermSource = 'theoryfiles/atlas/WZ2010/WP-applgrid.root' ,
             'theoryfiles/atlas/WZ2010/KF-Wp-nnlo2nlo-ew.txt'
TheorExpr= 'K*A1'
```

bin1	bin2	kfactor (NNLO/NLO)
0.00	0.21	0.980972
0.21	0.42	0.979131
0.42	0.63	0.977933
0.63	0.84	0.973563
0.84	1.05	0.977653
1.05	1.37	0.975987
1.37	1.52	0.984766
1.52	1.74	0.973862
1.74	1.95	0.976351
1.95	2.18	0.982303
2.18	2.50	0.976012

theoryfiles/atlas/WZ2010/KF-Wp-nnlo2nlo-ew.txt

Releasing assumptions:

- In addition …
- HERA data can not provide flavour decomposition of the sea quarks, hence some assumptions are imposed in the default settings of the HERAFitter:
- check your **`steering.txt:`**

```
&ExtraMinimisationParameters
   name  = 'alphas',  'fs',   'fcharm'
   value =  0.1176 ,  0.31,     0.
   step  =  0.0    ,  0.0 ,     0.
! set to 0 to avoid minimisation
&End
```

LHC data →

$$r_s = \frac{f_s}{1 - f_s}.$$

```
&ExtraMinimisationParameters
   name  = 'alphas',  'rs',   'fcharm'
   value =  0.1176 ,  0.5,      0.
   step  =  0.0    ,  0.1 ,     0.
! set to 0 to avoid minimisation
&End
```

- This is what we parametrise and assume:

H1-14-042 / ZEUS-prel-14-007

$$xf(x,Q_0^2) = A \cdot x^B \cdot (1-x)^C \cdot (1 + Dx + Ex^2)$$

xg	xg	
xu_v	$xU = xu + xc$	
xd_v	$xD = xd + xs$	
$x\bar{U}$	$x\bar{U} = x\bar{u} + x\bar{c}$	
$x\bar{D}$	$x\bar{D} = x\bar{d} + x\bar{s}$	

→

$$\begin{aligned}
xg(x) &= A_g x^{B_g}(1-x)^{C_g}, \\
xu_v(x) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}\left(1 + E_{u_v}x^2\right), \\
xd_v(x) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}}, \\
x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}, \\
x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}}.
\end{aligned}$$

HERAPDF1.0 & NLO HERAPDF1.5

$$\begin{aligned}
xg(x) &= A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{C'_g}, \\
xu_v(x) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}\left(1 + D_{u_v}x + E_{u_v}x^2\right), \\
xd_v(x) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}}, \\
x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}(1 + D_{\bar{U}}x), \\
x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}}.
\end{aligned}$$

HERAPDF2.0

$x\bar{s} = f_s x\bar{D}$ strange sea is a fixed fraction f_s of $\bar{D}$ at Q_0^2

Apply momentum/counting sum rules:

$$\int_0^1 dx \cdot (xu_v + xd_v + x\bar{U} + x\bar{D} + xg) = 1$$

$$\int_0^1 dx \cdot u_v = 2 \qquad \int_0^1 dx \cdot d_v = 1$$

$B_{\bar{U}} = B_{\bar{D}}$

$Sea = 2(\bar{U} + \bar{D})$

$A_{\bar{U}} = A_{\bar{D}}(1 - f_s)$

ensures $x\bar{u} \to x\bar{d}$ as $x \to 0$

$B_{u_v} = B_{d_v}$ constraint removed since HERAPDF1.5

$Q_0^2 = 1.9$

$Q^2_{\min} = 3.5$ or $10\ GeV^2$

$\alpha_s(M_Z^2) = 0.118$

$2 \cdot 10^{-4} \le x \le 0.65$

Eram Rizvi   QCD@LHC - Suzdal, Russia - 2014   18

$$\begin{aligned}
x\bar{D}(X) &= x\bar{d}(x) + xs(x), \\
x\bar{d}(x) &= A_{\bar{d}} x^{B_{\bar{d}}}(1-x)^{C_{\bar{d}}}, \\
xs(x) &= r_s A_{\bar{d}} x^{B_s}(1-x)^{C_s}.
\end{aligned}$$

```
in minuit.in.txt:
  83   'Cstr' 1.  0.1
```

Releasing assumptions exercise:

```
> cp input_steering/minuit.in.txt.13pHERAPDF minuit.in.txt
```

- Adjust minuit so it can minimise (for today use RT FAST)
- Run a job with rs fixed

```
    &ExtraMinimisationParameters
       name  = 'alphas',    'rs',    'fcharm'
       value =  0.1176 ,    0.5,       0.
       step  =  0.0    ,    0.0 ,       0.
    ! set to 0 to avoid minimisation
    &End

> bin/FitPDF
> mv output output.rsfixed
```

Run a job with rs free

```
    &ExtraMinimisationParameters
       name  = 'alphas',    'rs',    'fcharm'
       value =  0.1176 ,    0.5,       0.
       step  =  0.0    ,    0.1 ,       0.
    ! set to 0 to avoid minimisation
    &End

> bin/FitPDF
> mv output output.rsfree
```

Releasing assumptions exercise: Results

```
> cp input_steering/minuit.in.txt.13pHERAPDF minuit.in.txt
```

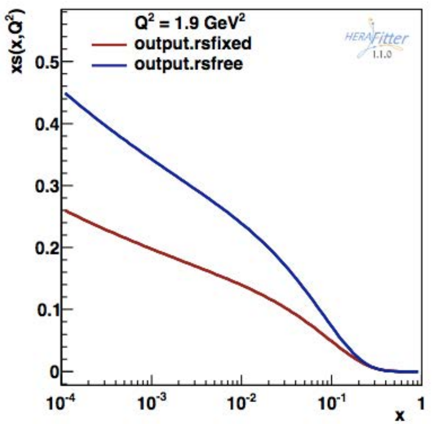
- Adjust minuit so it can minimise (for today use RT FAST)
- Run a job with rs fixed

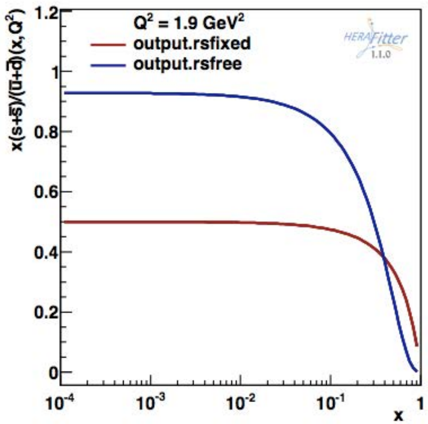
Run a job with rs free

```
bin/DrawPdfs output.rsfixed output.rsfree
```

Dataset	output.rsfixed	output.rsfree
NC cross section HERA-I H1-ZEUS combined e-p.	107 / 145	108 / 145
NC cross section HERA-I H1-ZEUS combined e+p.	446 / 379	448 / 379
CC cross section HERA-I H1-ZEUS combined e-p.	20 / 34	20 / 34
CC cross section HERA-I H1-ZEUS combined e+p.	32 / 34	32 / 34
ATLAS Z rapidity, 2010 data	8.3 / 8	3.2 / 8
ATLAS W+ lepton pseudorapidity, 2010 data	19 / 11	18 / 11
ATLAS W- lepton pseudorapidity, 2010 data	11 / 11	10 / 11
Correlated χ^2	5.8	3.7
Log penalty χ^2	+0.00	+0.00
Total χ^2 / dof	647 / 609	642 / 608
χ^2 p-value	0.14	0.16

Parameter	output.rsfixed	output.rsfree
'Bg'	-0.014 ± 0.052	-0.000 ± 0.041
'Cg'	4.98 ± 0.21	5.27 ± 0.46
'Aprig'	0.065 ± 0.092	0.085 ± 0.054
'Bprig'	-0.37 ± 0.10	-0.329 ± 0.059
'Cprig'	25.00	25.00
'Buv'	0.778 ± 0.027	0.799 ± 0.028
'Cuv'	4.76 ± 0.13	4.71 ± 0.14
'Euv'	10.1 ± 1.7	8.8 ± 1.7
'Bdv'	0.910 ± 0.074	0.969 ± 0.074
'Cdv'	4.08 ± 0.37	4.29 ± 0.38
'CUbar'	7.15 ± 0.65	6.86 ± 0.67
'ADbar'	0.2609 ± 0.0078	0.316 ± 0.022
'BDbar'	-0.1201 ± 0.0036	-0.1188 ± 0.0042
'CDbar'	8.1 ± 1.1	9.6 ± 1.4
'alphas'	0.1176	0.1176
'rs'	0.5000	0.93 ± 0.16





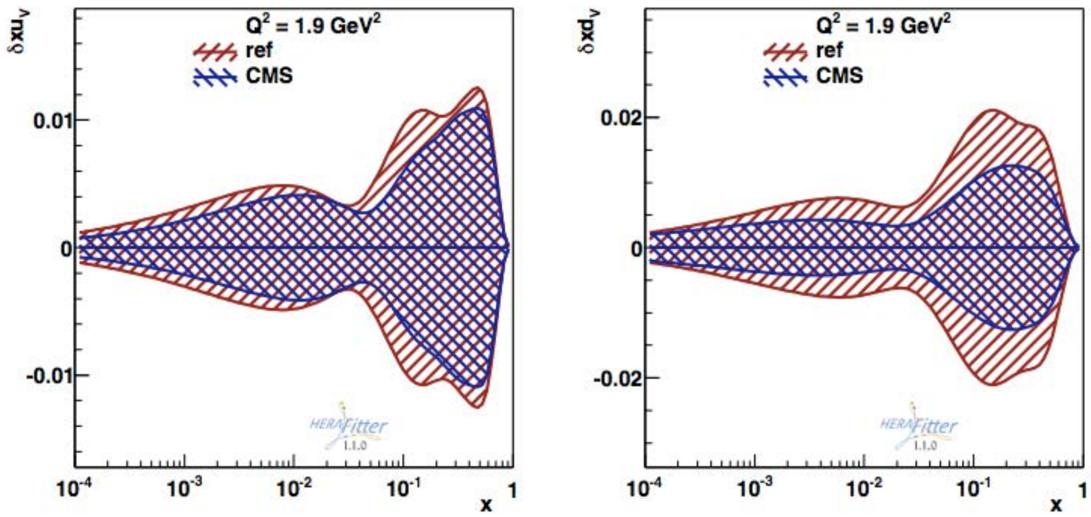
How to assess the impact of a new data

- Check your data against predictions based on different PDFs (CT10, MSTW08_CPdeut, NNPDF3.0, HERAPDF1.5, ABM12), if any deviation/tension with some of the set can be observed, then data could bring constraining power to help PDFs to converge.

- Decide on your reference set:
  - usually we choose HERA data —> consistent data set, free of corrections
  - run fits with `DOBANDS=TRUE` —> Hessian Errors
  - save the output directory (remember the settings!)
- Add your new data files:
  - run under exactly same conditions with `DOBANDS=TRUE`
  - save the output directory
- use Drawing tools to visualise the results:

```
> bin/DrawPdfs --bands --absolute-errors ref CMS
```

Here i saved the outputs from CMS and HERA I



Parameter	ref	CMS
'Bg'	−0.344 ± 0.080	−0.353 ± 0.067
'Cg'	5.96 ± 0.67	6.11 ± 0.74
'Aprig'	0.82 ± 0.15	0.88 ± 0.17
'Bprig'	−0.415 ± 0.059	−0.415 ± 0.050
'Cprig'	25.00	25.00
'Buv'	0.644 ± 0.027	0.681 ± 0.020
'Cuv'	4.72 ± 0.21	4.68 ± 0.16
'Euv'	10.6 ± 2.2	9.5 ± 1.5
'Bdv'	0.635 ± 0.075	0.513 ± 0.042
'Cdv'	4.29 ± 0.61	3.57 ± 0.44
'CUbar'	2.28 ± 0.43	2.73 ± 0.44
'ADbar'	0.1623 ± 0.0073	0.1674 ± 0.0081
'BDbar'	−0.1638 ± 0.0063	−0.1603 ± 0.0067
'CDbar'	2.51 ± 0.56	2.38 ± 0.64
'alphas'	0.1176	0.1176
'fs'	0.3100	0.3100

- one can also use MC method (see H. Pirumov's tutorial)

Summary …

- HERAFitter is a versatile package
    - once installed and linked to all external dependencies if needed
- It can be used with external predictions to quantitatively assess level of agreement between data and theory by taking into account all sources of provided uncertainties

- It can be used to test the impact of new measurements and study PDF discrimination and PDF improvement
- It can fit or test sensitivity to other QCD parameters: alphas, heavy quark masses

- More features can be added, for that your feedback is needed … you have a break to think of what we could further improve (physics cases, user friendliness)..

- for any question, feel free to contact your tutors, or just send us e-mail at

herafitter-help@desy.de

THANK YOU