Bundesministerium für Bildung und Forschung

HERAFitter

DESY

HERAFitter: PDF determination using DIS data

Hayk Pirumov

Proton Structure in the LHC Era, DESY, 29.09.2014

Outline

- Introduction
- Fit Setup
- DIS: formalism and hf-schemes
- Experimenting with cuts
- Information on correlations
- Propagating uncertainties to PDFs
- Adding data

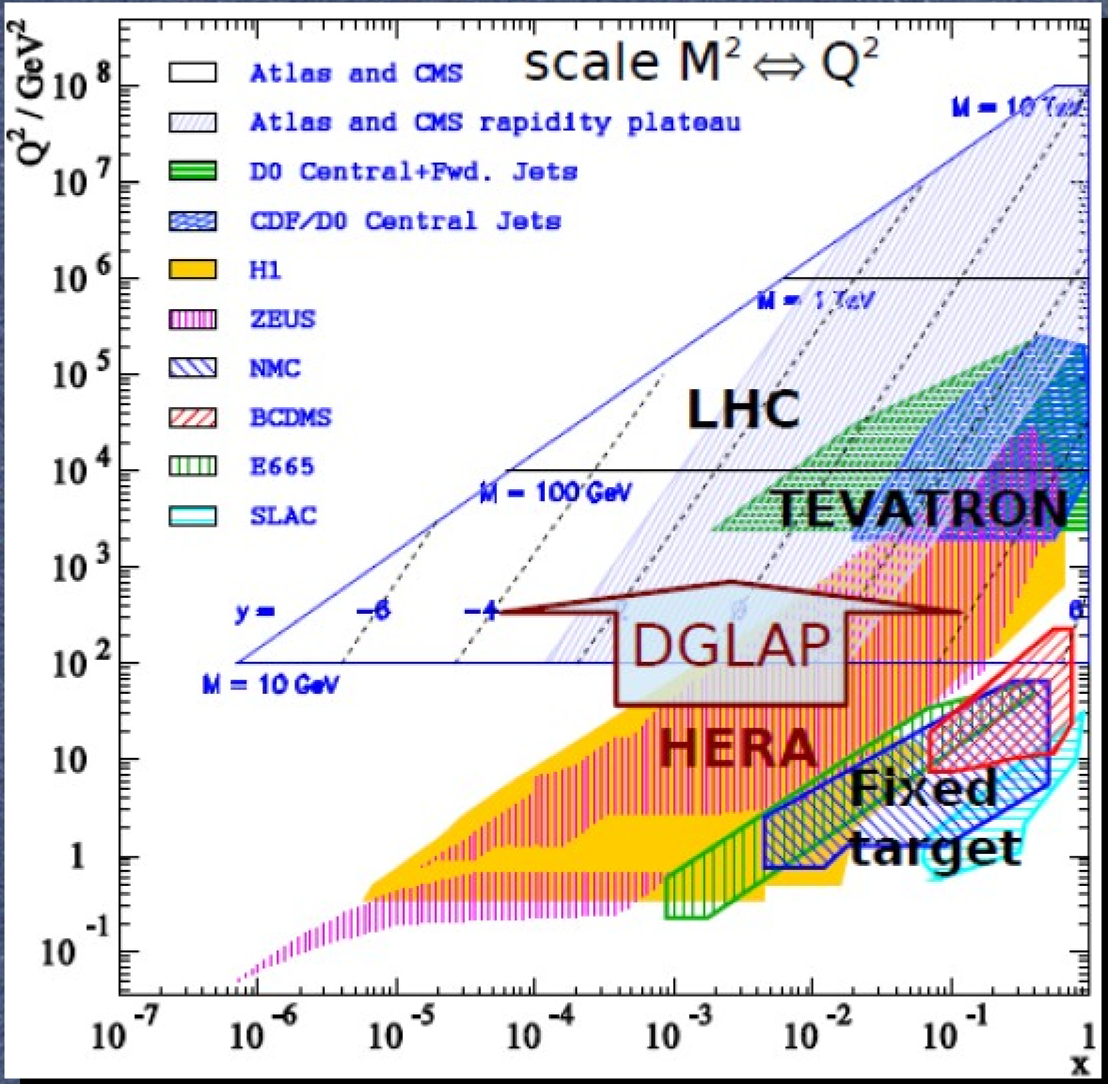
Introduction

According to the factorisation theorem:

cross section: PDFs ⊗ hard-scattering coefficients

Main information on PDFs comes from DIS data at HERA

- Sensitive to light quarks and gluons
- No information on flavour decomposition of the sea
- Various LHC processes provide information on sea decomposition *(see talk by Voica Radescu)*



In this tutorial we will use inclusive NC and CC DIS from HERA to determine PDFs as used in HERAPDF1.0 *[JHEP01(2010)109]*

Setup

For the tutorial HERAFitter-1.1.0 version will be used:

```
./configure
make clean
make
make install;
```

Copy a tar file containing some files for this tutorial:

```
wget www.desy.de/~pirumov/pirumov_tutorial.tgz
tar xvfz pirumov_tutorial.tgz
```

Copy steering and minuit card to the herafitter directory:

```
cp pirumov_tutorial/steering.txt.rtfast steering.txt
cp pirumov_tutorial/minuit.in.txt.basic minuit.in.txt
```

We'll use HERA I data and 'RT FAST' heavy flavor scheme

Run a fit:

```
./bin/FitPDF
```

DIS: formalism

Neutral Current cross section:

$$\frac{d^2\sigma_{NC}^{e^\pm p}}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}[Y_+\tilde{F}_2^\pm \mp Y_- x\tilde{F}_3^\pm - y^2\tilde{F}_L^\pm]$$

Charged Current cross section:

$$\frac{d^2\sigma_{CC}^{e^\pm p}}{dxdQ^2} = \frac{G_F^2}{4\pi x}\left(\frac{M_W^2}{Q^2+M_W^2}\right)^2[Y_+W_2^\pm + y^2W_L^\pm \mp Y_- xW_3^\pm]$$

Relation to PDFs in QCD (LO):

$$[F_2, F_2^{\gamma Z}, F_2^Z] = x\sum_q [e_q^2, 2e_q v_q, v_q^2 + a_q^2]\{q(x,Q^2) + \overline{q}(x,Q^2)\}$$
$$[xF_3^{\gamma Z}, xF_3^Z] = x\sum_q [e_q^2 a_q, 2v_q a_q]\{q(x,Q^2) - \overline{q}(x,Q^2)\}$$

$$\tilde{\sigma}_{CC}^{e^+p} = x[\overline{u}+\overline{c}] + (1-y)^2 x[d+s]$$
$$\tilde{\sigma}_{CC}^{e^-p} = x[u+c] + (1-y)^2 x[\overline{d}+\overline{s}].$$

A lot more details in Amanda's talk

DIS: heavy flavour schemes

Fixed-Flavour number scheme (FFN)

- Heavy quark contributions included explicitly in hard scattering coefficients
  - Available in HERAFitter via interface OPENQCDRAD (as used in ABM) or via QCDNUM.

Variable-Flavour nember scheme (VFN):

- Number of active flavours changes, when the scale crosses the threshold for a heavy quark production.
  - Available in HERAfitter RT (as used in MSTW), ACOT (as used by CTEQ)

HF-Schemes in HERAFitter:

```
! --- Scheme for heavy flavors :
! ---  HF_SCHEME = 'ZMVFNS'    : ZM-VFNS (massless),
! ---  HF_SCHEME = 'RT'        :  Thorne-Roberts VFNS (massive)
! ---  HF_SCHEME = 'RT FAST'   : Fast approximate TR VFNS scheme, usign k-factor
! ---  HF_SCHEME = 'RT OPT'        :  Thorne-Roberts VFNS (massive)
! ---  HF_SCHEME = 'RT OPT FAST'   : Fast approximate TR VFNS scheme, usign k-factor
! ---  HF_SCHEME = 'ACOT Full' : ACOT - F.Olness Version (massive), using k-factors
! ---  HF_SCHEME = 'ACOT Chi' : ACOT - F.Olness Version (massive), using k-factors
! ---  HF_SCHEME = 'ACOT ZM'   : ACOT - F.Olness Version (massless), using k-factors
! ---  HF_SCHEME = 'FF'        : Fixed Flavour Number Scheme (qcdnum)
! ---  HF_SCHEME = 'FF ABM'    : Fixed Flavour Number Scheme (ABM)
! ---  HF_SCHEME = 'FF ABM RUNM' : Fixed Flavour Number Scheme (ABM) using run mass def

HF_SCHEME = 'RT FAST'
```

→ Scheme we are using

k-factors

To speed up calculations k-factor technique can be used

k-factors can be defined:

- As the ratio between massless and massive scheme in the same QCD order (RT-FAST implementation in HERAFitter).
- As the ratio between massless LO and massive NLO (ACOT implementation in HERAFitter)

NOTE: k-factors are calculated for the input minuit parameters and are not updated with fit iterations. If the resulting PDFs significantly different from the input values, calculation should be repeated using the final PDF parameter as an input to the fit.

Results: different hf-schemes

Store the output of our RT-FAST fit: cp -r output output.rtfast
Copy PDFs from ACOT and ZM-VFNS: cp -r pirumov_tutorial/output.zmvfns .
cp -r pirumov_tutorial/output.acot .
Draw PDFs from different hf-schemes on one plot:

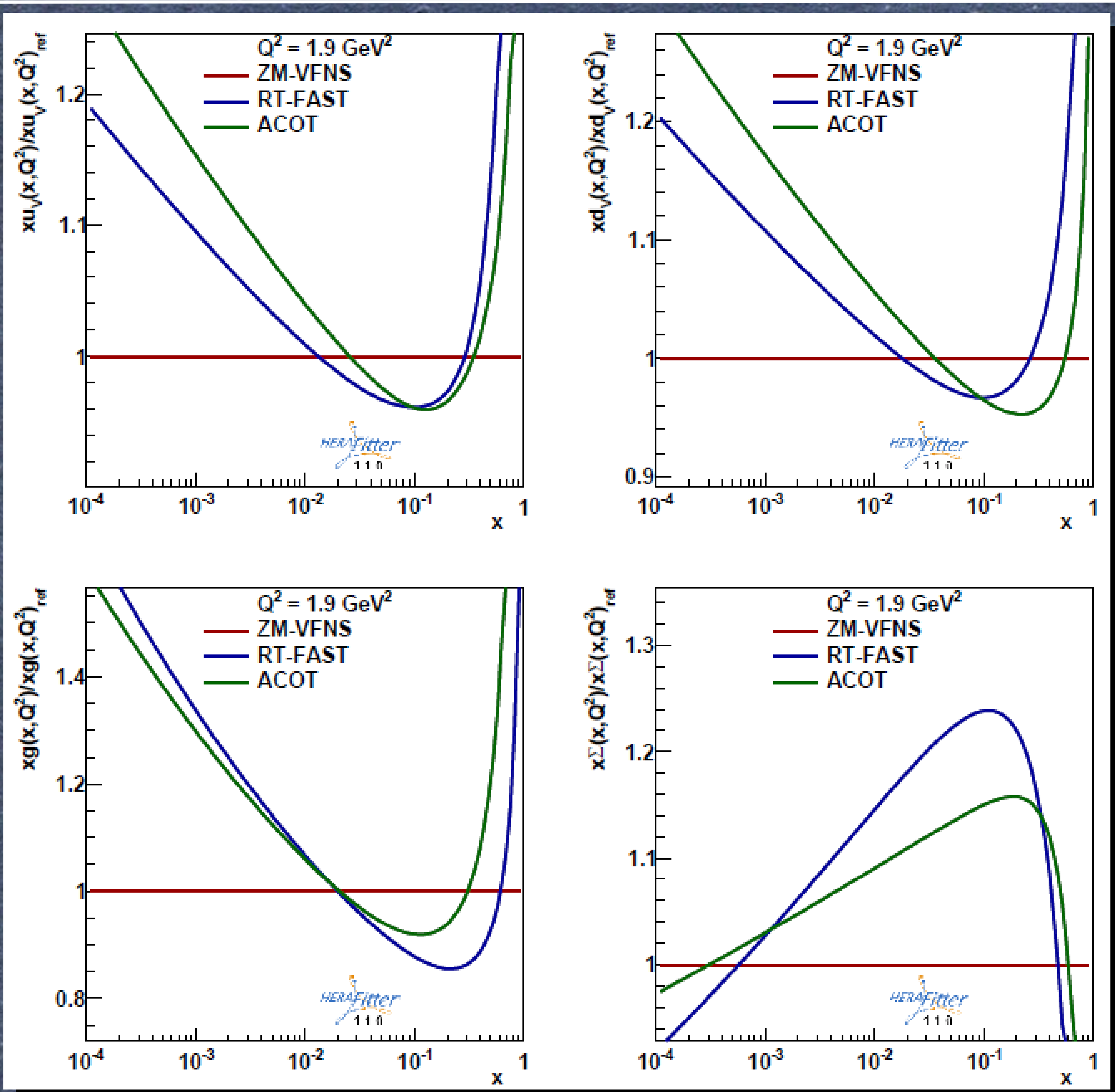
```
./bin/DrawPdfs output.zmvfns:'ZM-VFNS' output.rtfast:'RT-FAST' output.acot:'ACOT'
```

** Specify the output directories as arguments of ./bin/DrawPdfs*
One can specify legend in the plots by adding :'LEGEND' after each of the output directory name, otherwise names of output directories will appear as a legend

Open pdf file with plots: acroread plots/plots.pdf
Drawing tool also provides tables with chi2 values (p27) and PDF parameters (p28) from specified directories side by side:

Dataset	ZM-VFNS	RT-FAST	ACOT
NC cross section HERA-I H1-ZEUS combined e-p.	106 / 145	107 / 145	108 / 145
NC cross section HERA-I H1-ZEUS combined e+p.	438 / 379	419 / 379	406 / 379
CC cross section HERA-I H1-ZEUS combined e-p.	22 / 34	20 / 34	21 / 34
CC cross section HERA-I H1-ZEUS combined e+p.	34 / 34	29 / 34	31 / 34
Correlated χ^2	0	0	0
Log penalty χ^2	+0.00	+0.00	+0.00
Total χ^2 / dof	599 / 582	575 / 582	565 / 582
χ^2 p-value	0.30	0.57	0.69

Results: different hf-schemes



Open pdf file with plots:
acroread plots/plots.pdf

PDFs are presented at the starting scale Q^2 = 1.9 GeV^2

Differences get smaller with the rise of Q^2

Cuts on data

Cuts on various kinematic variables can be set in the steering.
Change the Q^2 for NC and CC data: emacs steering.txt
edit line 174 (NC) and line 187 (CC), save changes: ctrl+x+s

```
*
* Process dependent cuts
*
&Cuts

  !--------------------- NC ep  -----

  ! Rule #1: Q2 cuts
   ProcessName(1)      = 'NC e+-p'
   Variable(1)         = 'Q2'
   CutValueMin(1)      = 3.5
   CutValueMax(1)      = 1000000.0

  ! Rule #2: x cuts
   ProcessName(2)      = 'NC e+-p'
   Variable(2)         = 'x'
   CutValueMin(2)      = 0.000001
   CutValueMax(2)      = 1.0

  !---------------------  CC ep  ----

   ProcessName(3)      = 'CC e+-p'
   Variable(3)         = 'Q2'
   CutValueMin(3)      = 3.5
   CutValueMax(3)      = 1000000.0

   ProcessName(4)      = 'CC e+-p'
   Variable(4)         = 'x'
   CutValueMin(4)      = 0.000001
   CutValueMax(4)      = 1.0
```

```
*
* Process dependent cuts
*
&Cuts

  !--------------------- NC ep  -----

  ! Rule #1: Q2 cuts
   ProcessName(1)      = 'NC e+-p'
   Variable(1)         = 'Q2'
   CutValueMin(1)      = 10.
   CutValueMax(1)      = 1000000.0

  ! Rule #2: x cuts
   ProcessName(2)      = 'NC e+-p'
   Variable(2)         = 'x'
   CutValueMin(2)      = 0.000001
   CutValueMax(2)      = 1.0

  !---------------------  CC ep  ----

   ProcessName(3)      = 'CC e+-p'
   Variable(3)         = 'Q2'
   CutValueMin(3)      = 10.
   CutValueMax(3)      = 1000000.0

   ProcessName(4)      = 'CC e+-p'
   Variable(4)         = 'x'
   CutValueMin(4)      = 0.000001
   CutValueMax(4)      = 1.0
```

Run the fit:
./bin/FitPDF

Copy the fit output:
cp -r output output.q2cut

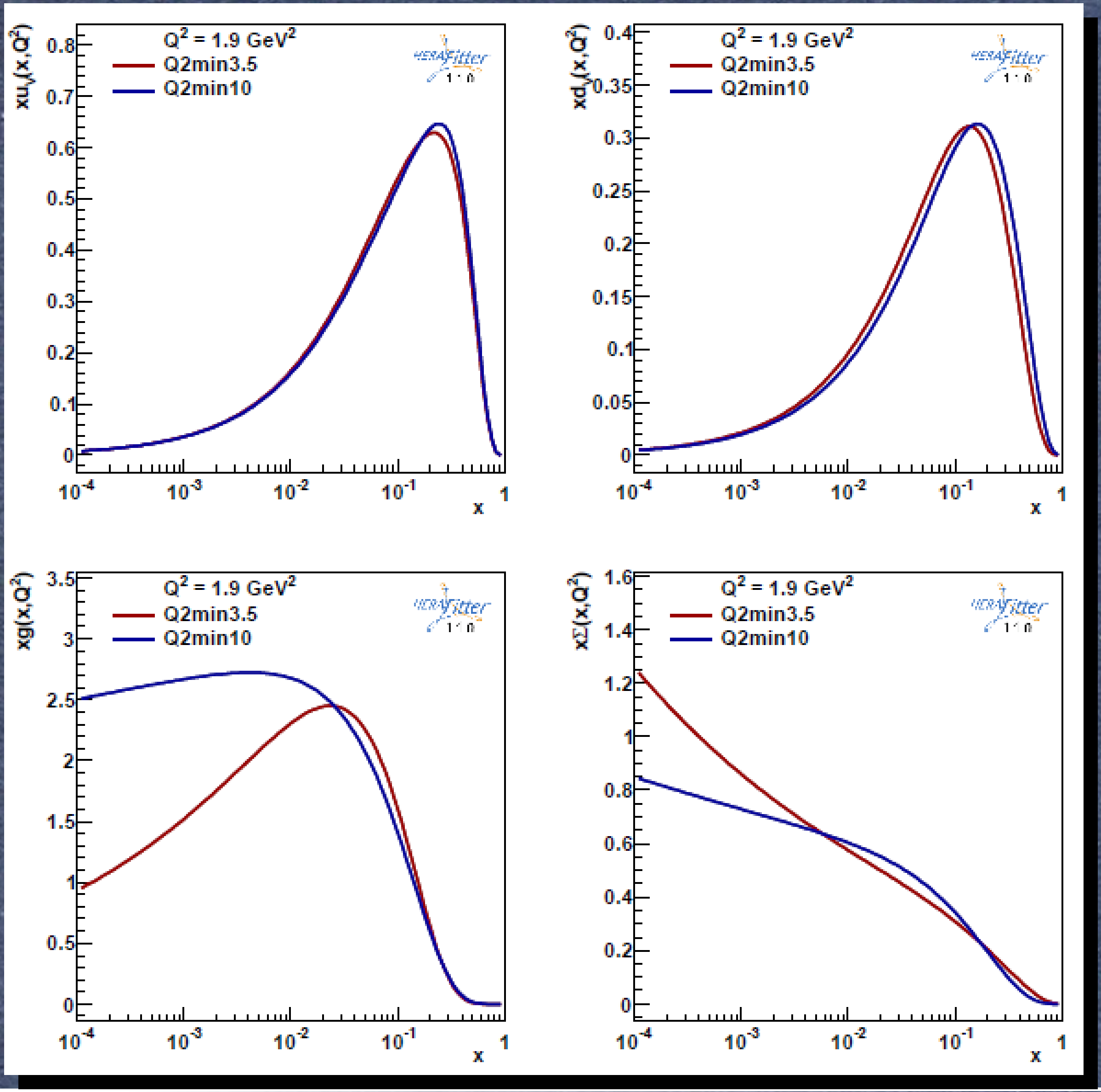
Steering with Q^2 cut is available:
pirumov_tutorial/steering.txt.q2cut

Output of the fit with Q^2 > 10 cut:
pirumov_tutorial/output.rtfast.q2cut

Results of varying Q^2 cut

Draw PDFs with different Q^2 cut:

./bin/DrawPdfs –no-data output.rtfast:'Q2min3.5' output.q2cut:'Q2min10'



** If –no-data option is specified, only PDFs are plotted*

Big difference in low x gluon –> gluon is constrained by low Q2 data

Taking into account correlated systematics

Data are reported with their central values and experimental uncertainties.
Systematic uncertainties can be correlated bin-to-bin and also between measurements.

Going back to basic steering with NC and CC data only:

cp pirumov_tutorial/steering.txt.rtfast steering.txt

So far when fitting NC and CC data we assumed all systematic uncertainties to be uncorrelated:

emacs datafiles/hera/H1ZEUS_CC_e-p_HERA1.0.dat

Uncomment line 24

```
!   To treat error uncorrelately, then: first  is uncor, then the sys_i(i=1,114) -> 115 sources
             !     Bins       x-sec                                      Errors
  ColumnName = 'x','Q2','y','x-section','stat','uncor',110*'uncor',4*'ignore'

!   To take into account the correlations then set SystScales to 1. and uncomment below:
!   ColumnName = 'x', 'Q2', 'y', 'Sigma',
!                    'stat', 'uncor',  'HERA_I_Source_1', 'HERA_I_Source_2', 'HERA_I_Source_3',
rce_6', 'HERA_I_Source_7', 'HERA_I_Source_8', 'HERA_I_Source_9', 'HERA_I_Source_10',
```

Comment line 21

```
!   To treat error uncorrelately, then: first  is uncor, then the sys_i(i=1,114) -> 115 sources
             !     Bins       x-sec                                      Errors
! ColumnName = 'x','Q2','y','x-section','stat','uncor',110*'uncor',4*'ignore'

!   To take into account the correlations then set SystScales to 1. and uncomment below:
   ColumnName = 'x', 'Q2', 'y', 'Sigma',
                    'stat', 'uncor',  'HERA_I_Source_1', 'HERA_I_Source_2', 'HERA_I_Source_3', '
ce_6', 'HERA_I_Source_7', 'HERA_I_Source_8', 'HERA_I_Source_9', 'HERA_I_Source_10',
```

* If the correlation column has the same name in different data files it is also considered correlated between that measurements

Results: taking correlations into account

Copy already modified data files to save time:

./pirumov_tutorial/copy_data.sh

Run a fit:

./bin/FitPDF

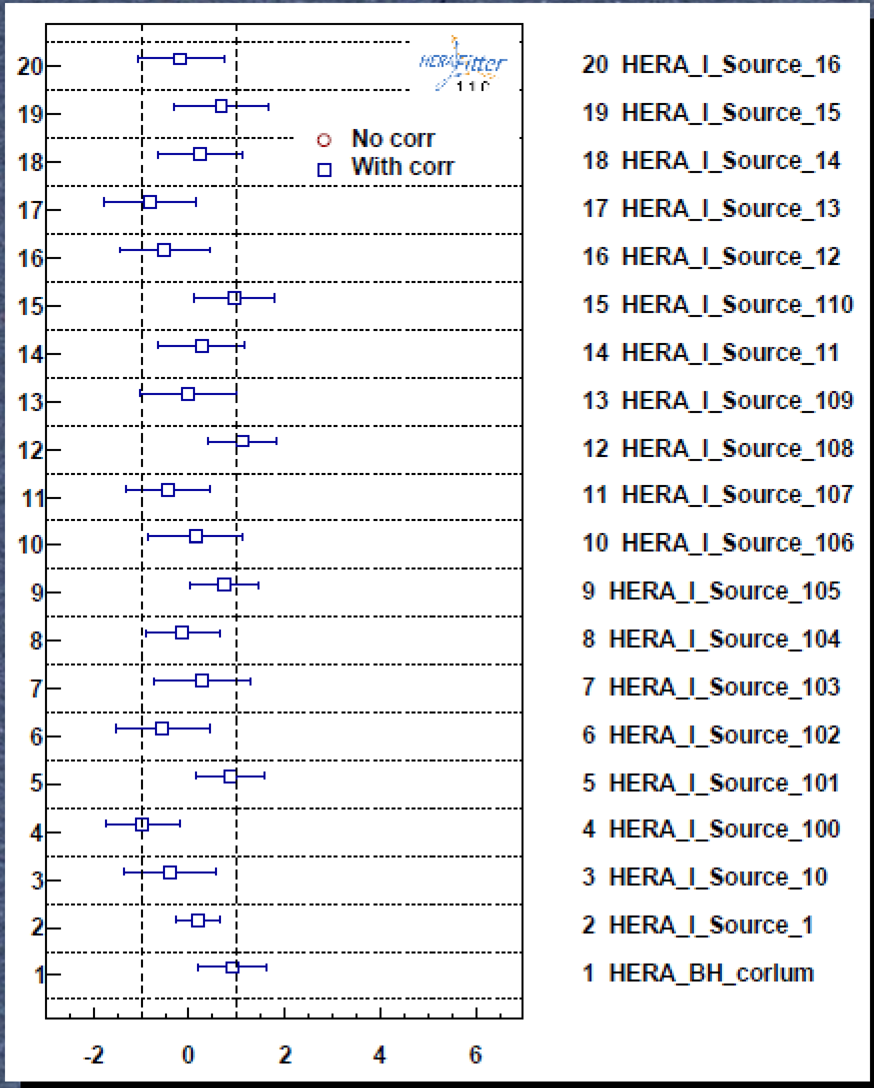
Copy fit results:

cp -r output output.corr

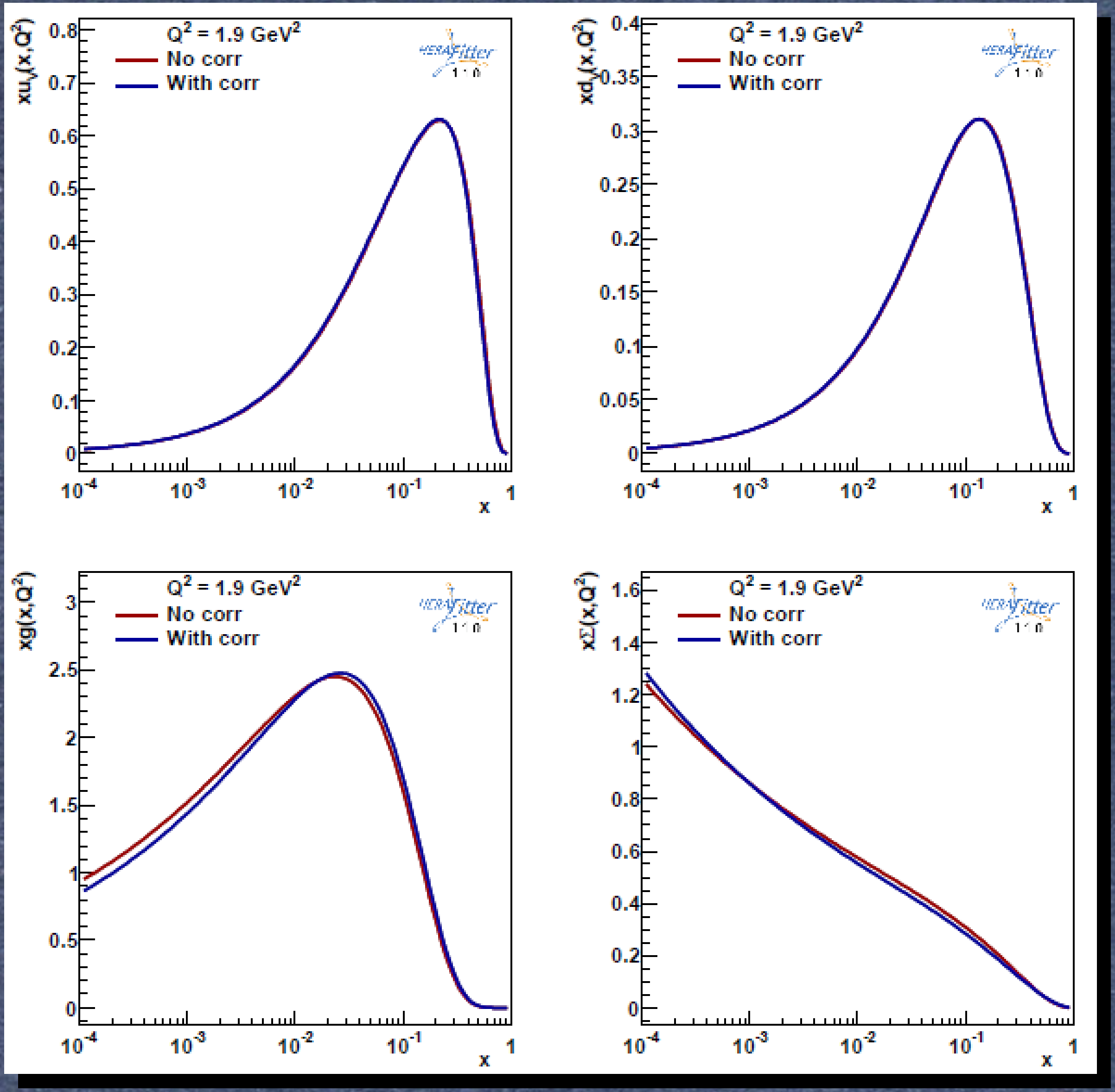
Draw results:

./bin/DrawPdfs output.rtfast:'No corr' output.corr:'With corr'

Drawing tool also plots shifts due to correlated systematics

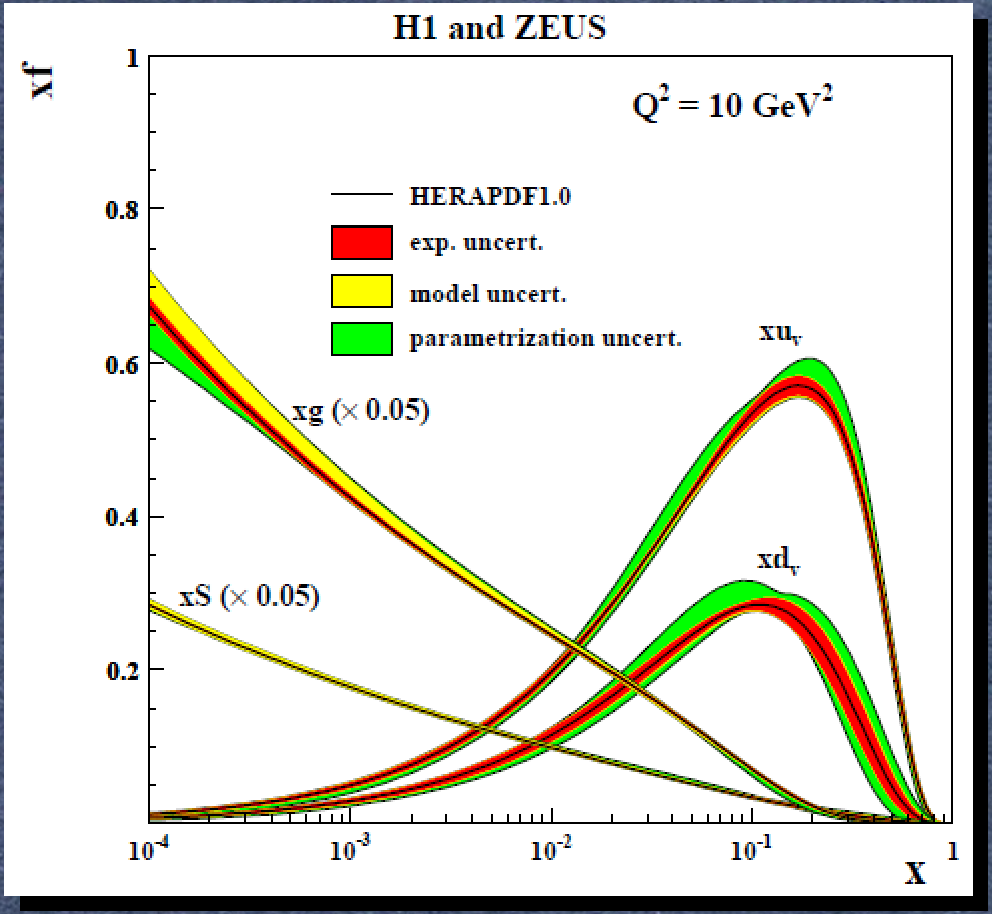


Results: taking correlations into account



Comparison of PDFs using information on correlated systematics and treating all systematics uncorrelated.

PDF uncertainties



HERAPDF fits determine valence, sea and gluon PDFs.
The sources of uncertainty include:

- experimental accuracy,
- model or theoretical uncertainties,
- uncertainties due to the choice of the parametrization form for PDFs.

Experimental uncertainties

```
*
* Output steering cards
*
&Output
  ! -- Error bands on parton distributions
  DoBands = False

  ! -- Q2 values at which the pdfs & errors are done (up to 20)
  Q2VAL = 1.9, 4., 10., 100., 6464, 8317

  ! How many x points to write (standard = 101)
  OUTNX = 101

  ! x-range of output (standard = 1E-4 1.0)
  OUTXRANGE = 1E-4, 0.9999
&End
```

Experimental uncertainties can be propagated to PDFs using *DoBands* = *True* option.

The method of quantifying PDF uncertainties associated with experimental uncertainties is described in details *[Phys.Rev.D65:014011,2001]*

Copying steering:

cp -r pirumov_tutorial/steering.txt.bands steering.txt

Running the fit:

./bin/FitPDF

Plot the bands

Copy fit output:

mv output output.bands

Make a directory for output:

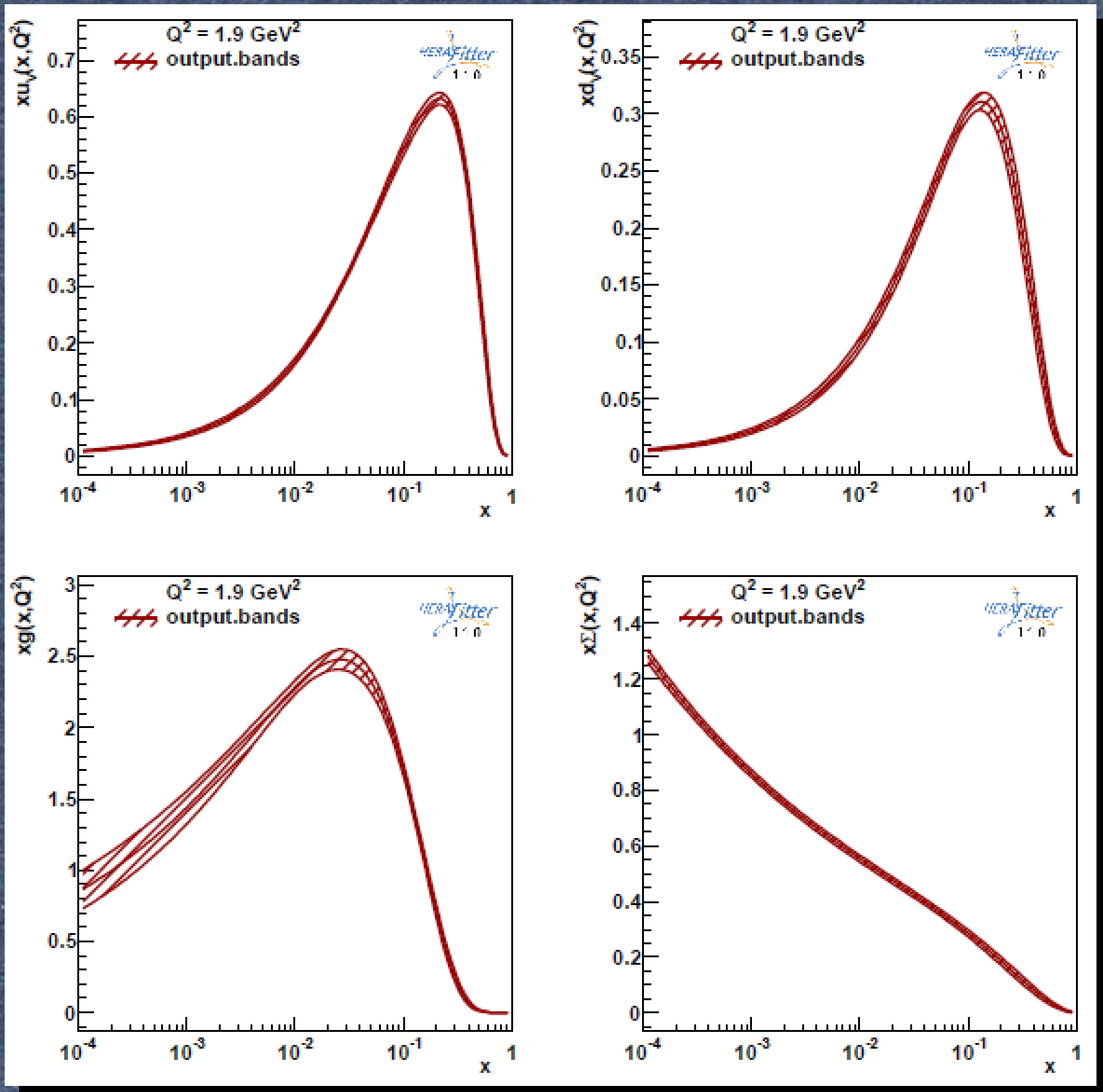
mkdir output

Plotting with the bands:

./bin/DrawPdfs output.bands

NOTE: we used mv this time not to carry along the output files of the DoBands=True option. A nicer alternative to copy the output directories every time is to specify the name of output directory in the steering:

```
&OutDir
  ! Name of the directory
    OutDirName = 'output'
&End
```



Monte Carlo method

Alternative way to propagate experimental uncertainties is to use Toy MC method. The input data are randomly fluctuated around the central values according to their experimental and systematic uncertainties.

```
*
* (Optional) MC errors steering cards
*
&MCErrors
  ! Activate MC method for error estimation if lRand = True
  lRAND    = True
  ! Use data (true, default) or theory (false) for the central values of the MC replica
  lRANDDATA = True

  ! MC method Seed
  ISeedMC = 1

  ! --- Choose what distribution for the random number generator
  ! STATYPE (SYS_TYPE)  =   1  gauss
  ! STATYPE (SYS_TYPE)  =   2  uniform
  ! STATYPE (SYS_TYPE)  =   3  lognormal
  ! STATYPE (SYS_TYPE)  =   4  poisson (only for lRANDDATA = False !)
  STATYPE =  1
  SYSTYPE =  1
&End
```

Put to true to activate MC method

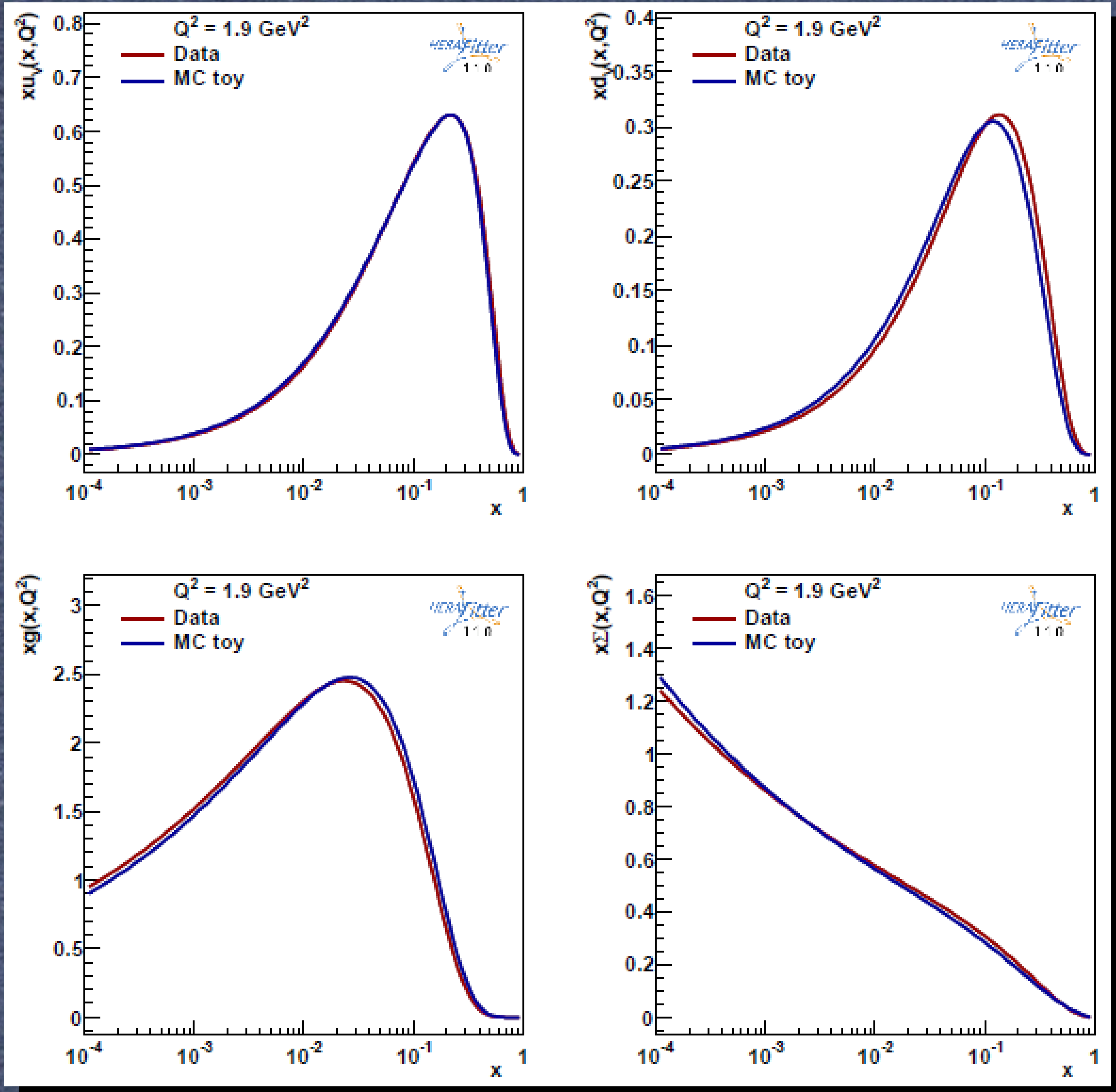
Change seed to create a different replica

Copy the steering: cp pirumov_tutorial/steering.txt.mc_toys
Run the fit: ./bin/FitPDF
Copy fit output: cp -r output output.mc_toy

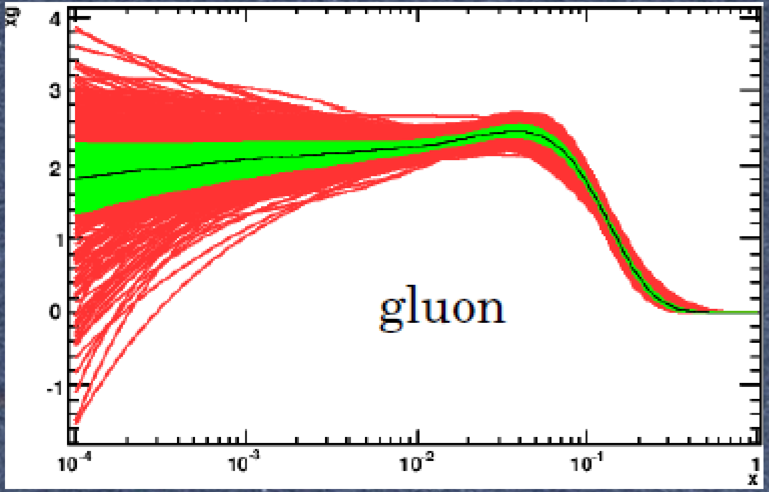
MC toy vs data

Compare PDFs from fit to data and fit to toy MC:

```
./bin/DrawPdfs output.rtfast:'Data' output.mc_toy:'MC toy'
```



After preparing a large number of MC toys and fitting them, one can construct the experimental uncertainty by taking RMS of the parton distributions from toy fits:



Model uncertainties

Model assumptions (such as quark mass, strong coupling constant, etc.) made for fit can be varied to estimate model uncertainties.
Model variations for HERAPDF1.0 is listed in the table:

Variation	Standard Value	Lower Limit	Upper Limit
f_s	0.31	0.23	0.38
m_c [GeV]	1.4	1.35 ($Q_0^2 = 1.8$)	1.65
m_b [GeV]	4.75	4.3	5.0
Q^2_{min} [GeV^2]	3.5	2.5	5.0
Q_0^2 [GeV^2]	1.9	1.5 ($f_s = 0.29$)	2.5 ($m_c = 1.6, f_s = 0.34$)

Quark masses are defined in *ewparam.txt*.
pirumov_tutorial/ewparam.txt.mb_var file can be used as an input to the fit instead of standard ew-card as an example of mb variation.

Parametrization uncertainty

In case of HERAPDF sets, parton distributions are parametrized using the following functional form:

$$xf(x) = Ax^B(1-x)^C P_i(x),$$

The functional form for the central HERAPDF1.0 set (see Ringaile's talk) has 10 free parameters.

The parametrization uncertainties in case of HERAPDF sets are estimated by adding free parameter to the central functional form one at a time and performing a fit.

An envelope is constructed from largest deviations of the new fits from the central one, forming parametrization uncertainty.

This can be technically done by modifying minuit.in.txt, releasing one of the parameters:

1	'Ag'	0.0000	0.
2	'Bg'	0.213846	0.010000
3	'Cg'	9.013846	0.500000
4	'Dg'	0.0000	0.
5	'Eg'	0.0000	0.
7	'Aprig'	0.0000	0.
8	'Bprig'	0.0000	0.
9	'Cprig'	0.0000	0.
11	'Auv'	0.0000	0.
12	'Buv'	0.665589	0.010000
13	'Cuv'	4.652237	0.500000
14	'Duv'	0.0000	0.

E.g. set the step for Dg parameter different from 0

Adding new data: H1 jets

Adding H1 jet data:

cp tutorial/steering.txt.jet_data steering.txt

```
*  Namelist to control input data
*
&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/hera/H1_NormInclJets_HighQ2_99-07.dat'
&End

&InCorr
  ! Number of correlation (statistical, systematical or full) files
    NCorrFiles = 1

  ! Correlation files:
    CorrFileNames(1) = 'datafiles/hera/H1_NormInclJets_HighQ2_99-07___H1_NormInclJets_HighQ2_99-07.corr'
&End
```

Increase number of input files

Add data file

Add correlation matrix file

H1 jet data provides additional information on statistical correlations (more details later):

datafiles/hera/H1_NormInclJets_HighQ2_99-07___H1_NormInclJets_HighQ2_99-07.corr

Jets and alpha_s

Jet cross sections are sensitive to alpha_s, let's try to fit it also:

emacs steering.txt

Modify line 142 :
Set *step* value that
corresponds to alphas to 0.01

```
* Add extra to minuit parameters. These MUST include alpha_S and fs
*
&ExtraMinimisationParameters
   name  = 'alphas',   'fs',    'fcharm'
   value =  0.1176 ,   0.31,       0.
   step  =  0.01      ,    0.0 ,       0.      ! set to 0 to avoid minimisation
&End
```

We will also need FastNLO theory table (technique will be discussed by Daniel Britzger):

cp tutorial/theoryfiles/fnh5401_I818707.tab theoryfiles/

Run the fit:

./bin/FitPDF

Correlation matrix

H1 jet data provides additional information on statistical correlations:

datafiles/hera/H1_NormInclJets_HighQ2_99-07___H1_NormInclJets_HighQ2_99-07.corr

```
&StatCorr
   Name1 = 'H1 normalised inclusive jet 99-07 data'
   Name2 = 'H1 normalised inclusive jet 99-07 data'

   NIdColumns1 = 2
   NIdColumns2 = 2

   IdColumns1 = 'q2min','ptmin'
   IdColumns2 = 'q2min','ptmin'

   NCorr = 18

  ! 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

   MatrixType = 'Statistical correlations'
&End
```

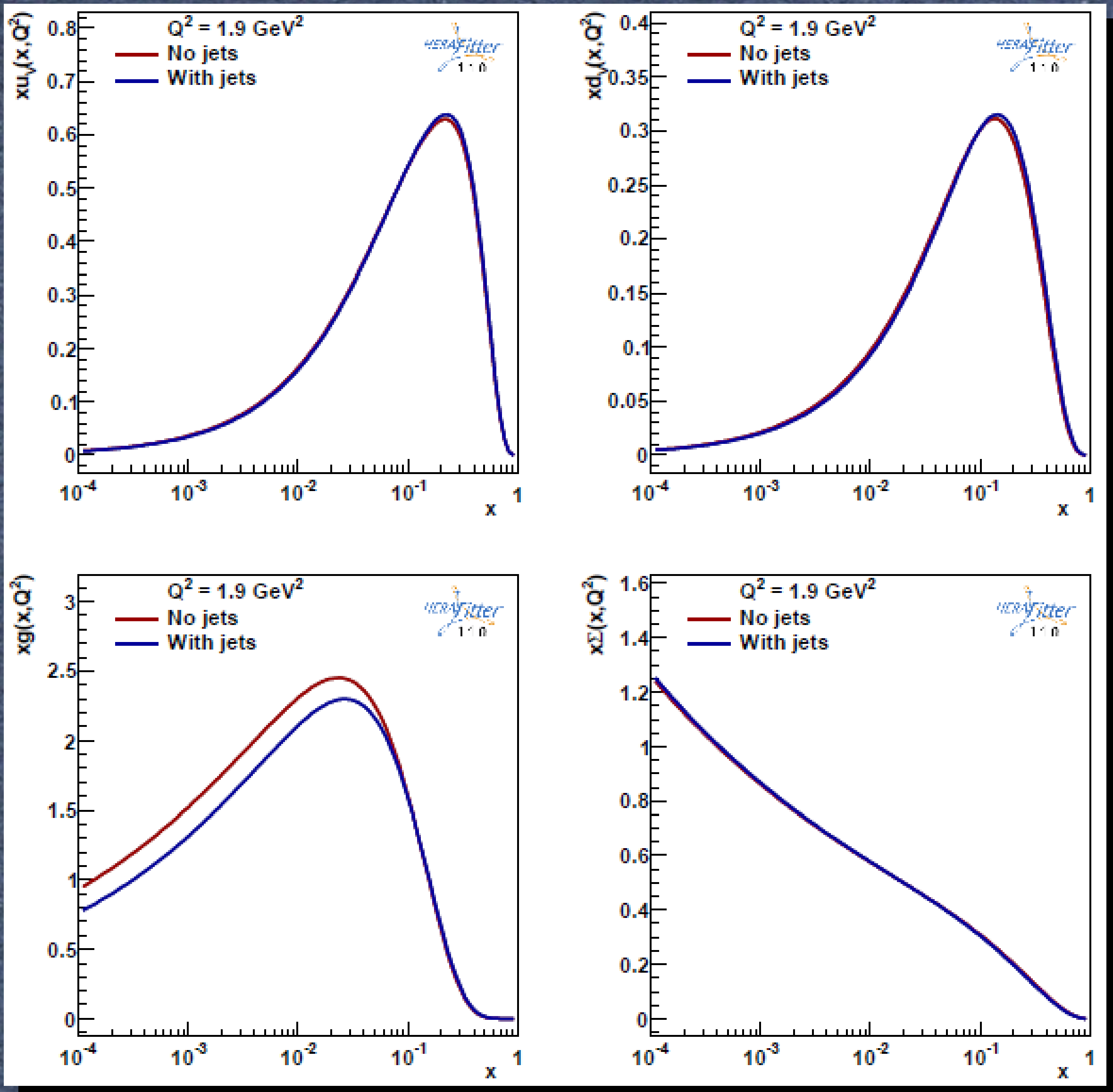
Names should match those of the data files, correlation between which is provided

Type of the correlation matrix

Results: using jet data

Compare PDFs with and without information from jets:

./bin/DrawPdfs output.rtfast:'No jets' output.jet_data:'With jets'



Copying fit output:

cp -r output output.jet_data

Looking into minuit output:

less output.jet_data/minuit.out.txt

```
101     alphas      0.11861         0.15140E-02
```

Fitted value

Uncertainty from minuit

Summary

What we know by now

How to use HERAFitter with DIS data as an example:

- different hf-schemes,
- taking into account correlated uncertainties,
- propagating experimental uncertainties to PDFs
- estimating model and parametrization uncertainties

How to add new data with H1 jets data as an example:

- use correlation information provided as a correlation matrix
- fit parameters other PDFs: alpha_s

The steerings and outputs for most of the exercises can be found in:

www.desy.de/~pirumov/pirumov_tutorial.tgz