Study on PDF parametrisation uncertainties using Monte Carlo technique

Voica Radescu
(Physikalisches Institut Heidelberg)
A. Glazov, S. Moch
(DESY)

DIS2010 Firenze
20 April 2010

Outline

- Introduction
- Method
    - Chebyshev polynomials
    - Monte Carlo Technique
- Results
- Summary

Introduction

- PDFs are crucial inputs for studies at the LHC, therefore precise knowledge and understanding of them is essential.

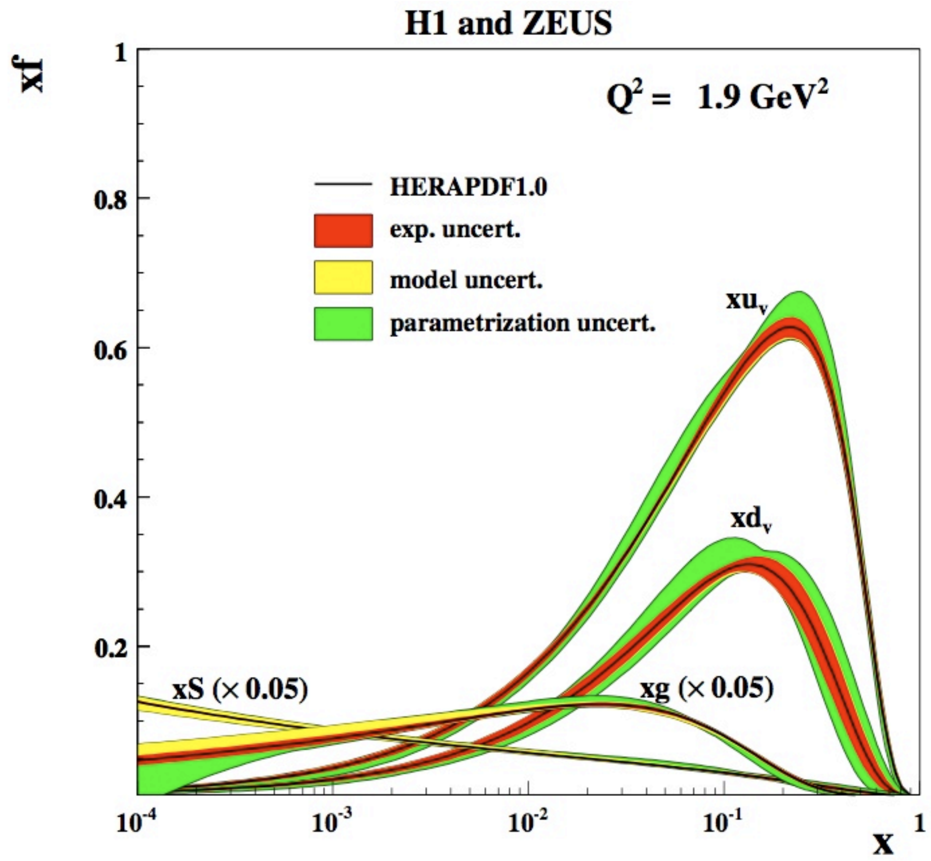
- PDFs are parametrised and extracted from fits, however the parametrisation uncertainty needs to be understood.

  - H1 and ZEUS collaborations made an effort to estimate uncertainty on the HERAPDF parametrisation by scanning the parameter space affecting especially the high x region
    [see S.Habib's presentation]
  - Standard Parametrisation Form:

$$Ax^B(1-x)^C(1+Dx+..)$$

    - describes the shape of PDFs with few input parameters
    - difficult to study systematically both the low and high x regions
    - multiple similar solutions for $x>x_{min}$
      - equivalent solutions for D~0 and $Dx_{min}>>1$

$$Ax^B(1-x)^C(1+Dx) = DAx^{B+1}(1-x)^C\left(\frac{1}{Dx}+1\right)$$



  - Neural Network PDF group uses Neural Nets to study PDF param. biases

Chebyshev Polynomials

Another method to study parametrisation biases, is to use orthogonal polynomials to parametrise PDFs: Chebyshev Polynomials of the first kind

- Orthogonally defined in the [-1,1] interval and given by the recurrence relation:

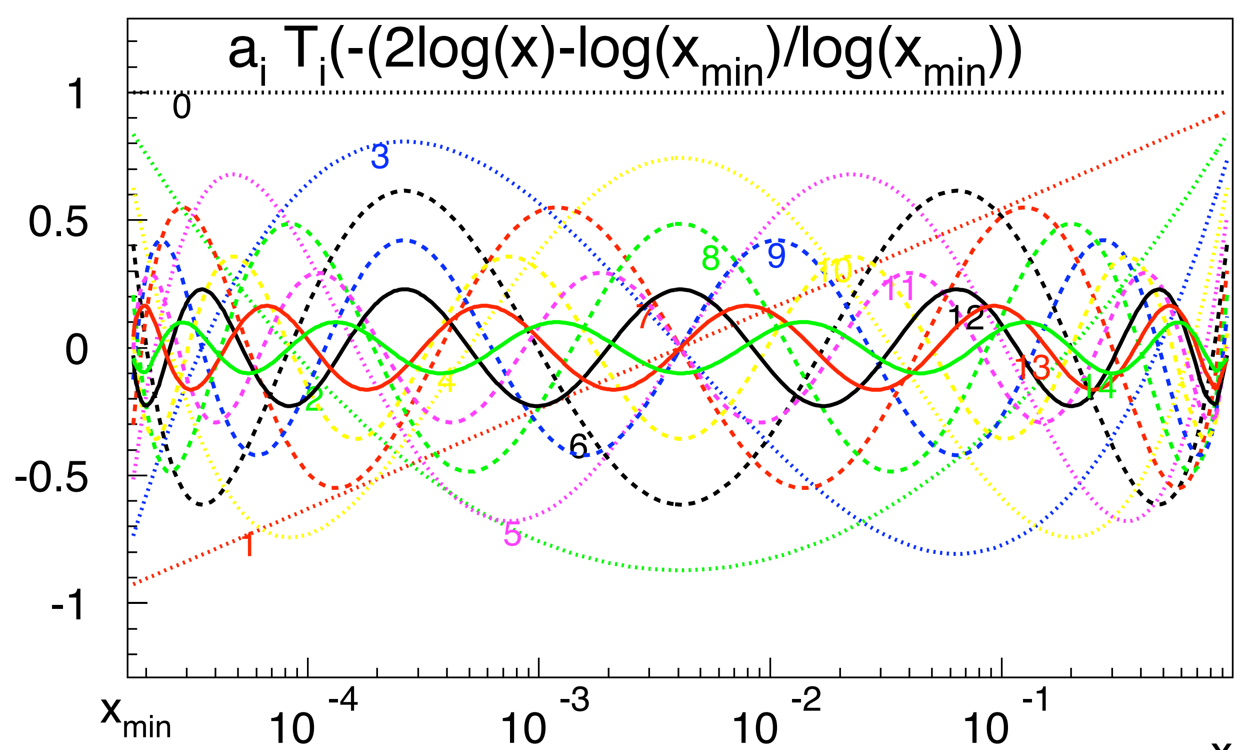
$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$$

  - To approximate PDFs, change variable $x \rightarrow \frac{-(2\log x - \log x_{min})}{\log x_{min}}$ such that [log(x_{min}),0] interval is mapped to [-1,1]
  - This allows to approximate PDF with few parameters:

$$xf(x) = \sum_{i=0}^{N-1} a_i T_i\left(\frac{-(2\log x - \log x_{min})}{\log x_{min}}\right)(1-x)$$

    - (1-x) term is to force xf(x)=0 for x=1
- Momentum Sum Rule leads to simple finite integrals

$$T_0(x) = 1$$
$$T_1(x) = x$$
$$T_2(x) = 2x^2 - 1$$
$$T_3(x) = 4x^3 - 3x$$
$$T_4(x) = 8x^4 - 8x^2 + 1$$
$$T_5(x) = 16x^5 - 20x^3 + 5x$$
$$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$$
$$T_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x$$
$$T_8(x) = 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1$$
$$T_9(x) = 256x^9 - 576x^7 + 432x^5 - 120x^3 + 9x.$$



Settings

The study is performed using:

- Published Combined HERA I data of NC and CC $e^{\pm}p$ scattering cross sections following HEP01 (2010) 109
- Fit program HERAPDF based on QCDNUM implementation at NLO [ref. M. Botje]
  - $\overline{MS}$ renormalisation scheme, DGLAP evolution at NLO, massless quarks (ZMVFNS)
  - starting scale Q_0^2= 1.9 GeV^2
- PDFs are parametrised using ZEUS Parametrisation [EPJ C42,1(2005)hep-ph/0503274]
  - $\mathbf{xd_v(x), xu_v(x), x\Delta = x\bar{u}(x) - x\bar{d}(x)}$ with xΔ fixed

$$xu_v(x) = A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} (1 + D_{u_v} x)$$

$$xd_v(x) = A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}}$$

  - $\mathbf{xS(x) = 2x\left(\bar{u}(x) + \bar{d}(x) + \bar{s}(x) + \bar{c}(x)\right), xG(x)}$ using Chebyshev Polynomials
    - ☆ up to the 15th order in Chebyshev series expansion
    - ☆ Chebyshev Polynomials can reproduce the shape of the standard parametrisation with few parameters

- Errors are estimated using Monte Carlo technique [DESY-PROC-2009-02]

Monte Carlo Method

- Method consists in preparing replicas of data sets allowing the central values of the cross sections to fluctuate within their systematic and statistical uncertainties taking into account all point to point correlations.

  - Various assumptions can be considered for the error distributions: Gauss, Log-Normal, etc. ...

  - Shift central values randomly within their uncorrelated errors assuming Gauss distributions of the errors:

$$\sigma_i = \sigma_i(1 + \delta_i^{uncorr} RAND_i)$$

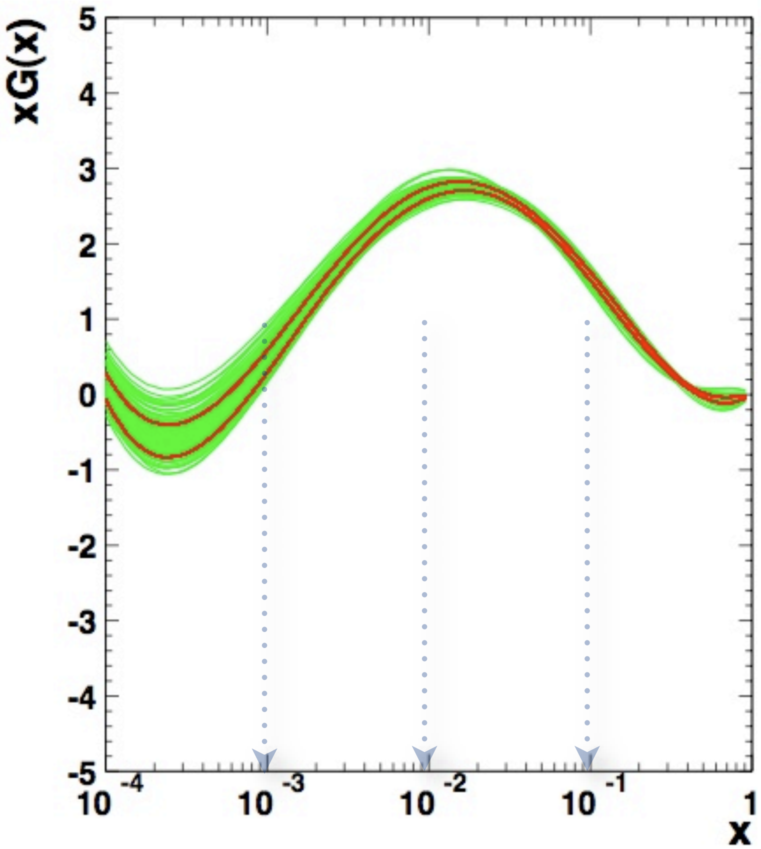
  - Shift central values with the same probability of the corresponding correlated systematic shift assuming Gauss distribution of the errors:

$$\sigma_i = \sigma_i(1 + \delta_i^{uncorr} RAND_i + \sum_j^{N_{sys}} \delta_{ij}^{corr} RAND_j)$$

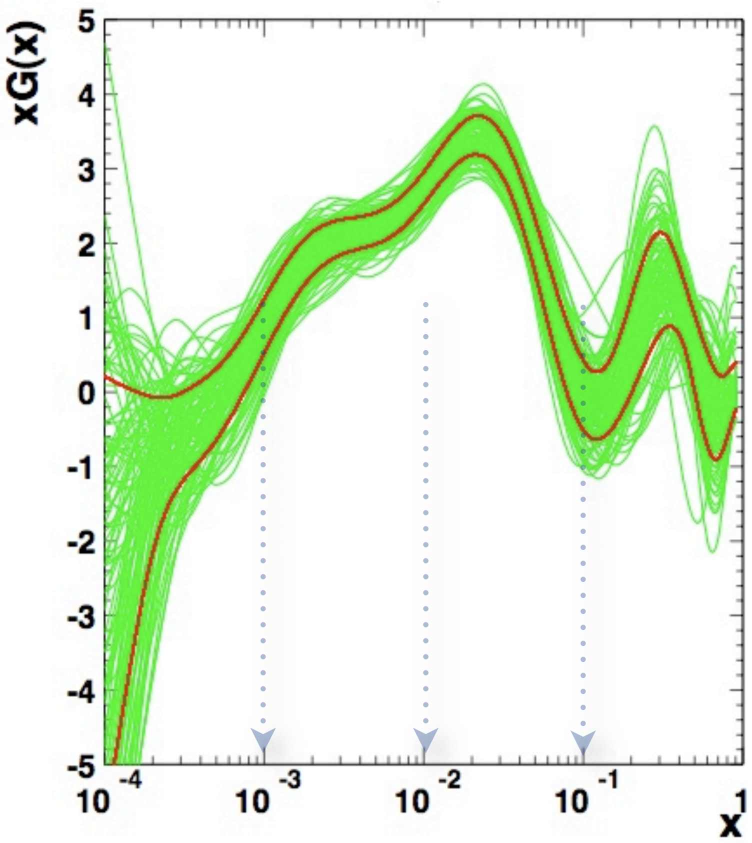
- Preparation of the data is repeated for N times (N>100):
  - For each replicas NLO QCD fit is performed to extract the PDF set

- Errors on the PDFs are estimated from the RMS of the spread of the N curves corresponding to the N individual extracted PDFs.

Fit Results



Npar=5
for xG, xS



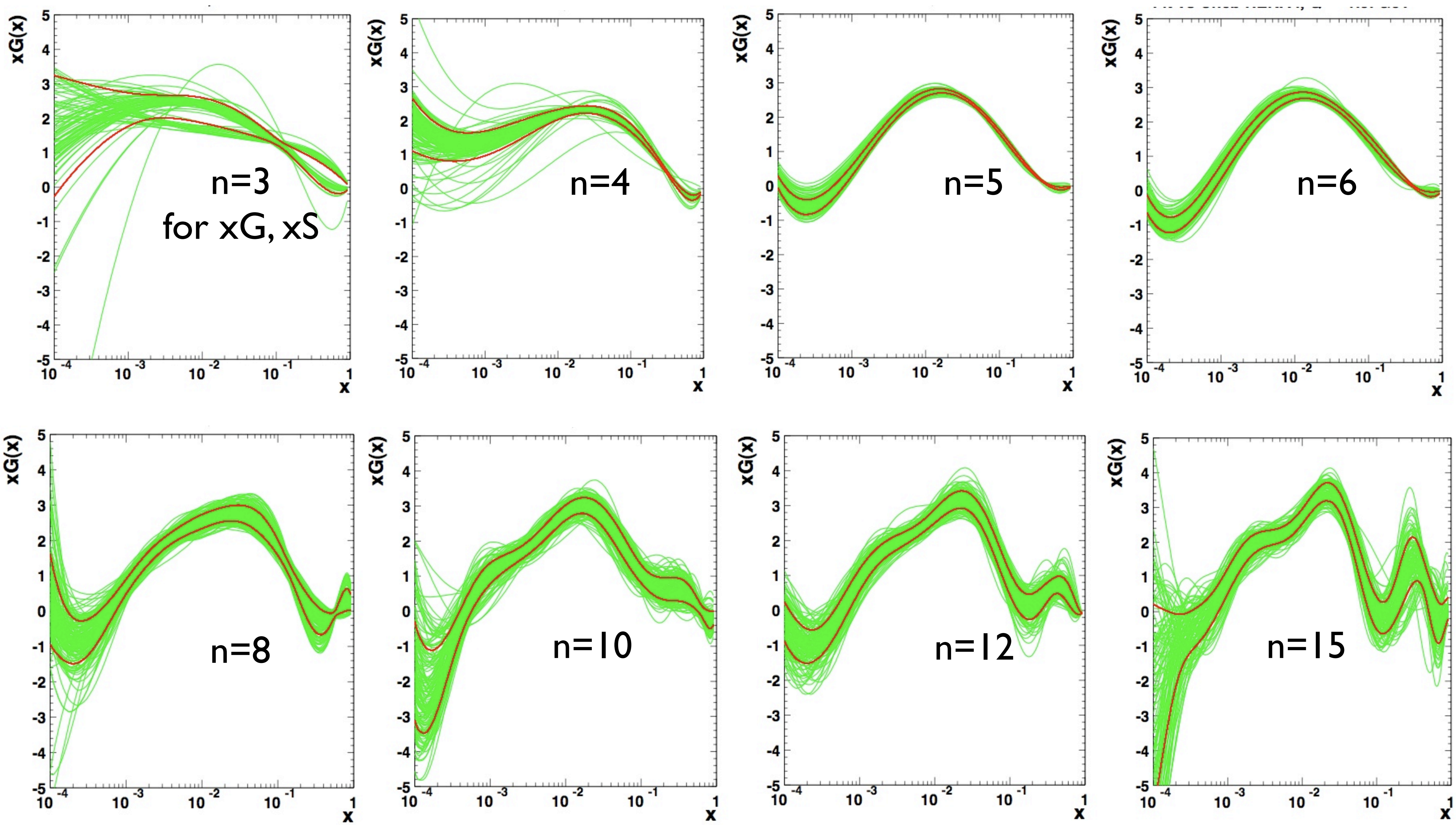
Npar=15
for xG, xS

- All Plots are shown for gluon distribution at Q^2=1.9 GeV^2
- MC replicas are shown in green lines (N>100)
- The uncertainty is estimated as the RMS of the spread and is shown in red
- Study in more details for fixed x points
  - x=0.0001
  - x=0.001
  - x=0.01
  - x=0.1

  at the edges of sensitivity and for the bulk of precision.

Dependence on Chebyshev Expansion

- Observe shape differences with the increased number of Chebyshev parameters:

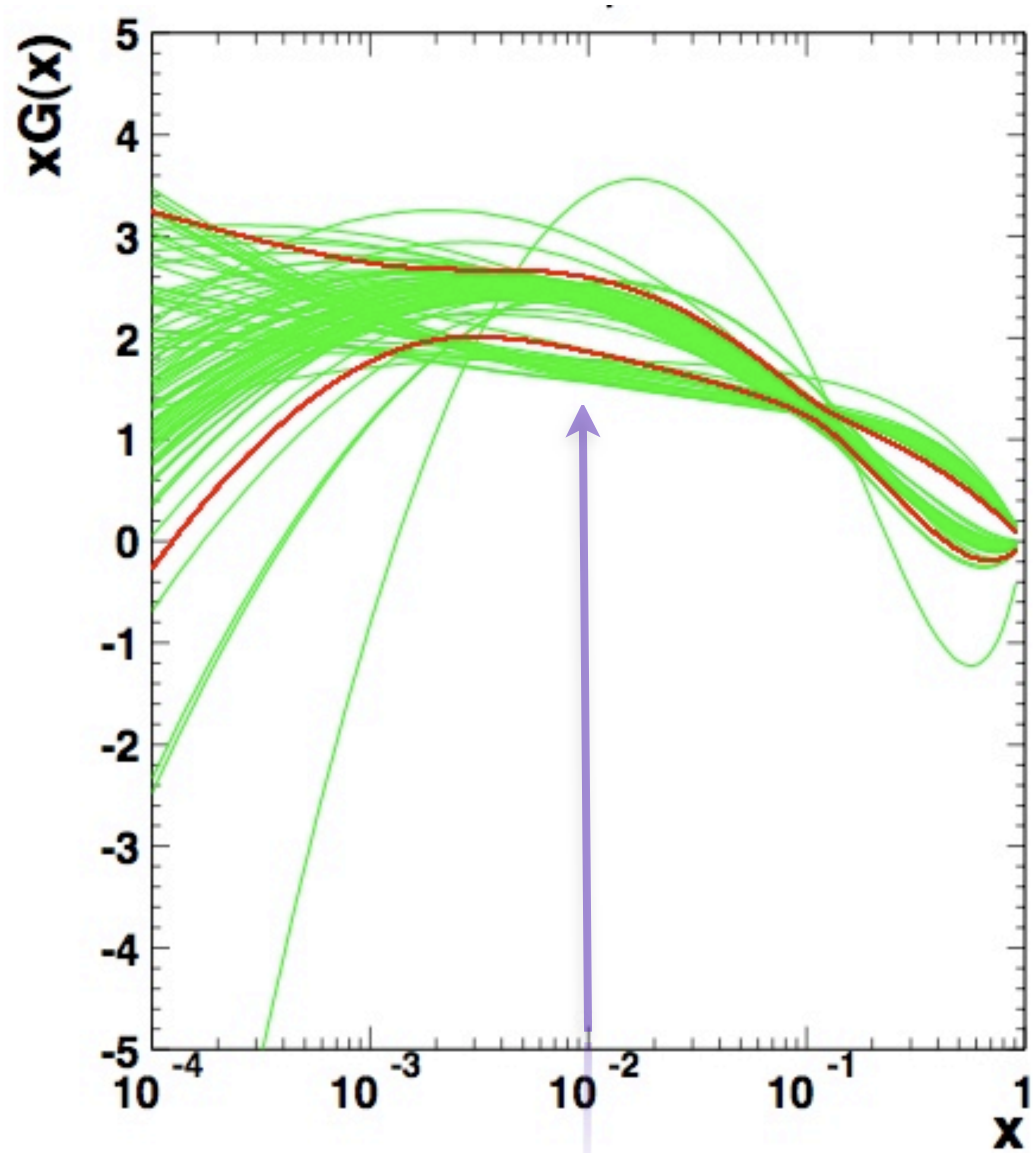


Double minima for low Npar

- A feature is observed for low number of Chebyshev parameters.
- Two solutions are preferred by the minimisation procedure for n=3 Chebyshev parameters, clearly observed at x=0.01

Q^2=1.9 GeV^2

n=3
for xG, xS

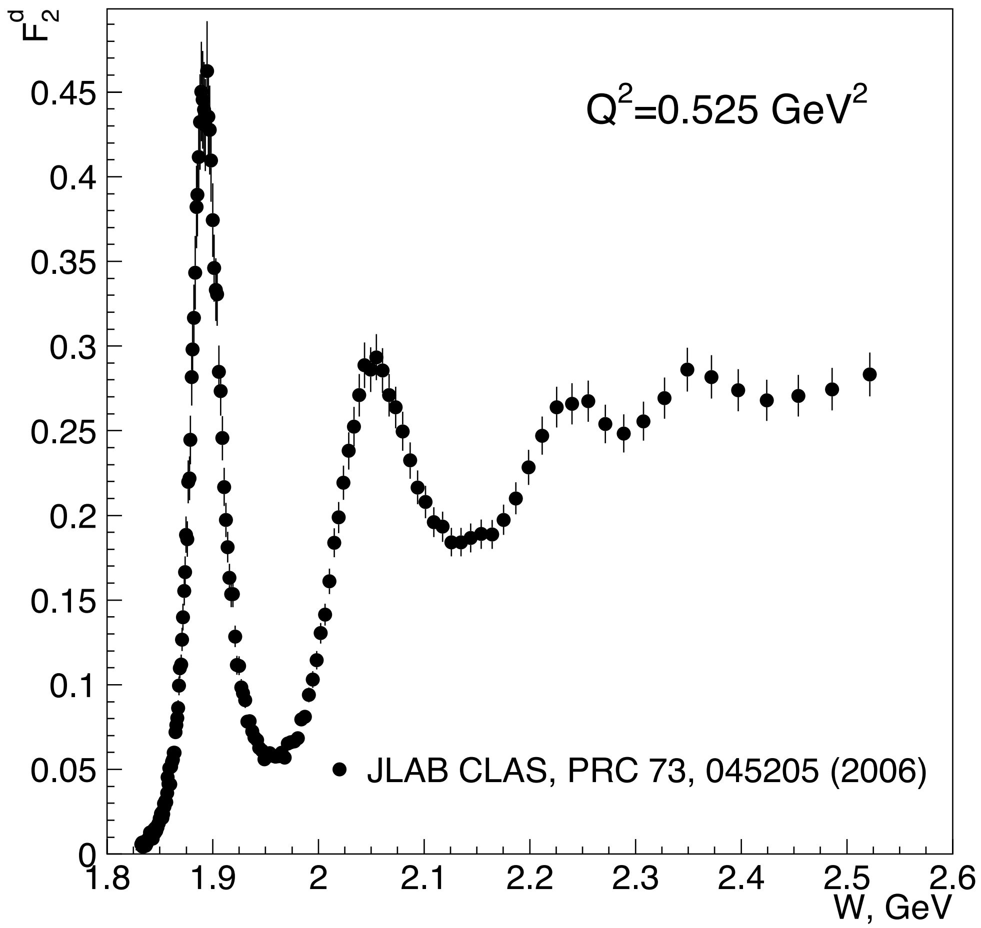


Constraining the shape of PDFs

- Humpy shapes in x can be correlated with the humps in the W:

$$W \approx Q\sqrt{\frac{1-x}{x}}$$

  - Resonances are observed at low W but they disappear for high W
    - Resonances at low W, as depicted by the JLAB CLAS experiment



- Idea: use "length" as an extra constraint [Ref: W. Giele]:

  - prefer solutions which are smoother in W
  - apply length penalty to χ^2

$$L = \int_{W_{min}}^{W_{max}} \sqrt{1+\left(\frac{dxf(W)}{dW}\right)^2}dW$$

$$\Delta\chi^2 = P\cdot(L - L_{min})$$

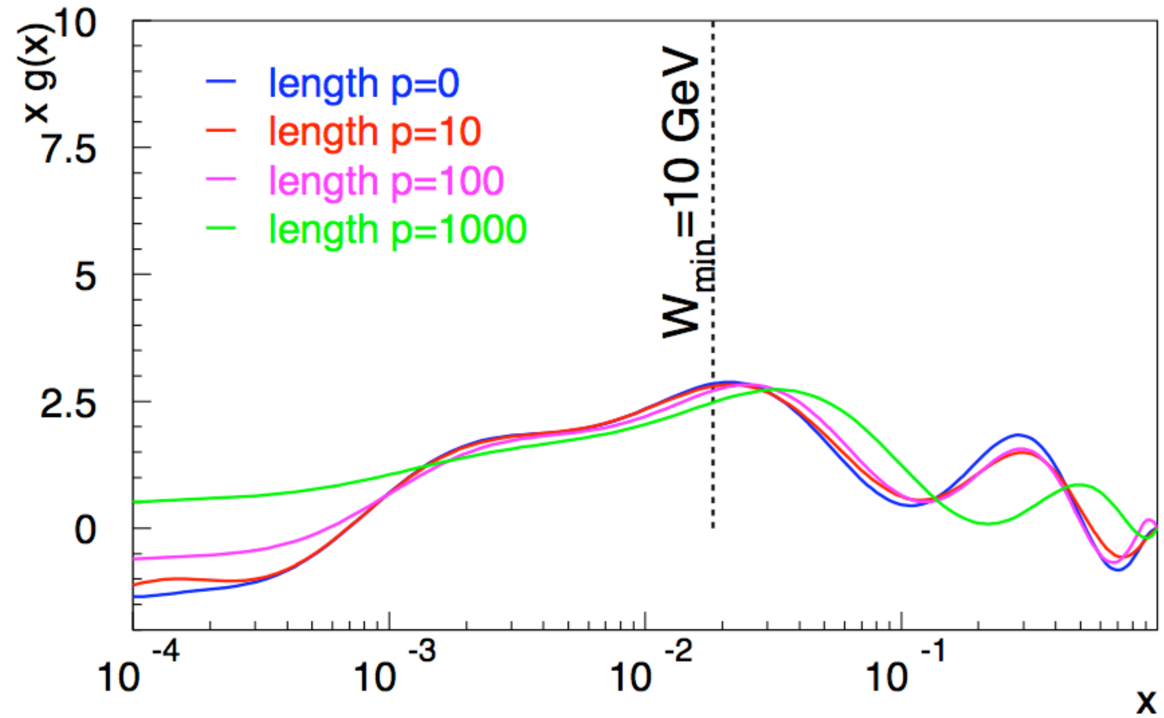
$$L_{min} = W_{max} - W_{min}$$

Stabilising the fit using Length Constraint

- Focus on the low x region:
  - Wmin=10 GeV
- Apply length penalty at the starting scale for each MC replica
  - p=0, 10, 100, 1000
- The Constraint is efficient (χ^2/ndf)

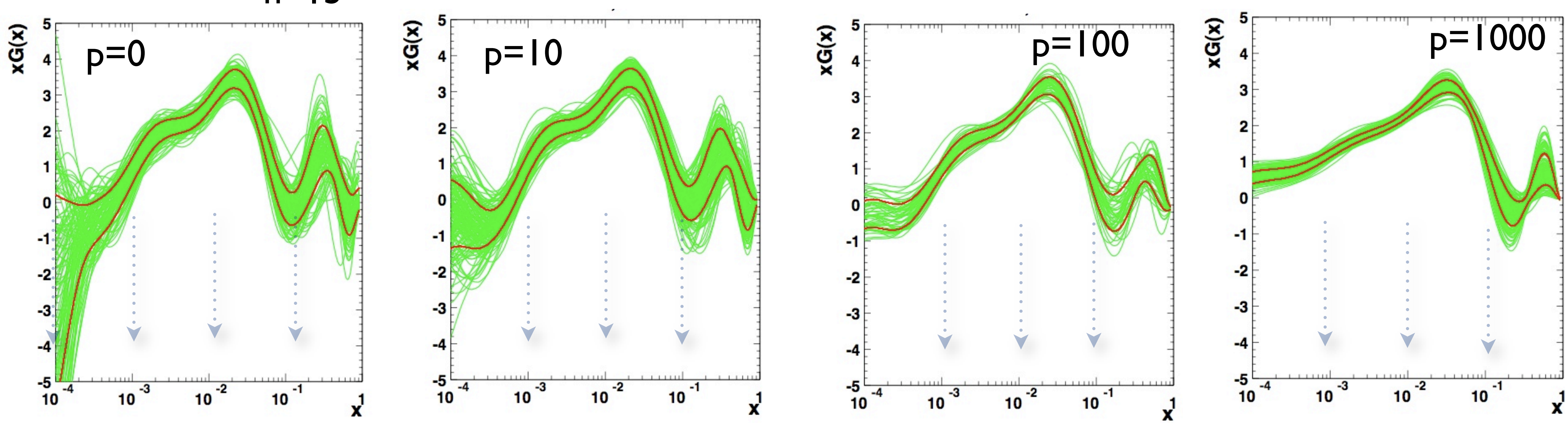
α, GeV^{-1}	0	10	100	1000
χ^2/n_{df}	560/557	561/557	572/557	626/557

Central, no MC



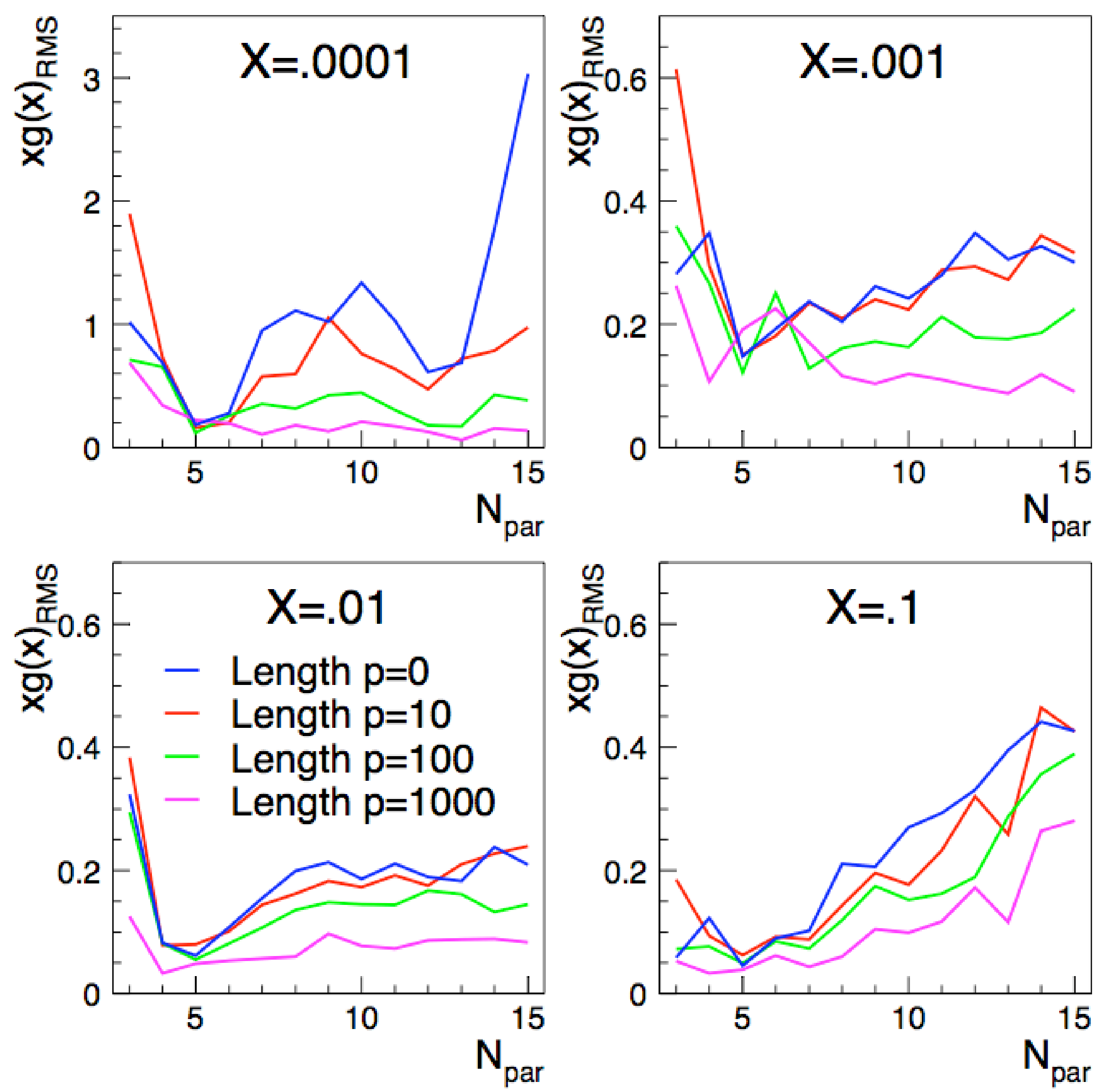
Q^2=1.9 GeV^2

n=15



Effect of the Length Constraint

- Even soft constraint against extra minima reduces uncertainty at low X
  - p=10 (red line)
- Tighter constraints limit the uncertainty better than the data.
- For x=0.1 constraint does not do much
- For the bulk of the data constraint does not do much for RMS but it becomes more constant vs Npar.



Summary

- Presented a study on the PDF parametrisation using Chebyshev polynomials with the emphasis on the low x region for gluon and sea quarks.

- Presented a method to constrain PDFs using simple, physically motivated penalty term against extra min/max vs W.

- The data are stable vs parameterisation change in the bulk region x=0.001-0.01.

- Minimal constraint improve precision for smallest x.