Averaging of DIS Cross Section Data

S. Glazov (DESY)

Introduction

- Why averaging ?
- The method
- Features of the averaging program
- Average of all published H1/Zeus data
- Some first results using the average data
- Next steps

Steps in a QCD analysis of experimental data

Standard QCD analysis to extract proton PDFs uses individual datasets from various experiments. All modern fits use both central values of F_2, xF_3, etc as reported by experiments as well as information about correlation between experimental points. These data are used directly to extract PDFs in a global QCD fit.

Unfortunately this "direct" procedure has a set of drawbacks:

- Even just for F_2 structure function the complete world dataset (including correlations) is large and difficult to obtain. Some of the correlations between experiments (e.g. H1 and Zeus) are not completely documented. Handling of the experimental data without additional "expert" knowledge became difficult.
- The treatment of the systematic errors is not unique. In "Lagrange multipliers" method the systematic uncertainties are floated in the fit and thus "fitted" to QCD. In "offset" method they are fixed. Both methods have advantages and disadvantages, it is difficult to select the standard one.
- Some global QCD fits use non-statistical $\Delta\chi^2 > 1$ criteria to estimate PDF uncertainties. Without model independent consistency check of the data it is maybe the safest method.

Motivation for the averaging of the data

The mentioned above drawbacks can be significantly reduced by *averaging* of the world structure function data:

- One combined world structure function dataset is much easier to handle.
- The averaging procedure is unique (will be discussed next), it removes the drawback of the offset method – systematic errors are floated (reduced) in the averaging procedure.
- χ^2/dof of the average allows model independent consistency check between experiments.

X-sections averaging procedure

Standard F_2 representation:

$$\chi^2(\{F_2^{true}\},\{\alpha\}) = \sum_i \frac{\left[F_2^{i,true} - \left(F_2^i + \sum_j \frac{\partial F_2^i}{\partial \alpha_j}\alpha_j\right)\right]^2}{\sigma_{F_2}^2} + \sum_j \frac{\alpha_j^2}{\sigma_{\alpha_j}^2}. \quad (1)$$

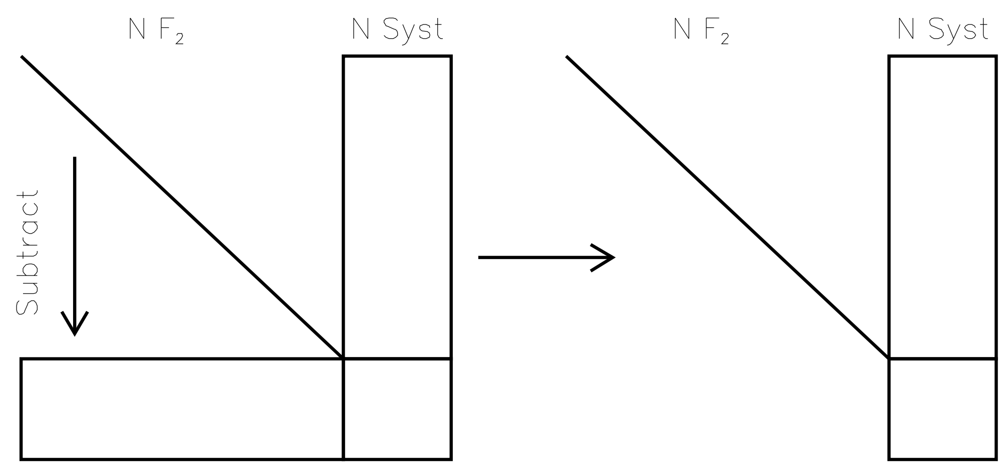
Here α_j — are correlated systematic uncertainty sources.

For several experiments, $\chi^2_{tot} = \sum_{exp} \chi^2_{exp}$. This χ^2 is normally used in QCD fits where $F_2^{true} = F_2^{theory}(glue, quarks)$.

Fit vs F_2, α values $\rightarrow$ *average* F_2

Some Technical Details

- Many more free parameters (all F_2 points !) vs QCD fit
- Data points from different experiments must be quoted at about the same Q^2, x.
- χ^2 has simple quadratic form $\rightarrow$ minimum is obtain by solving $N_{F_2} + N_{Syst}$ system of linear equations.
- The solution can be obtain using technique similar to simultaneous Z-vertex fit in H1 reconstruction:



(requires $\sim N_{F_2} \times N^2_{syst}$ operations).

Status of cross section averaging program

- Written in FORTRAN, under CVS, uses cernlib.
- Can calculate simultaneous average for different data types with correlated systematic sources (e.g. NC and CC cross sections which depend on hadronic energy scale)
- All data points are interpolated to the grid points defined by H1/Zeus grid, this interpolation uses NC/CC cross section parametrization obtained in H1 QCD fit (normally small correction factor).
- The cross section data points can be adjusted to the same center of mass energy using H1 cross section parametrizations.
- Output format directly suitable for H1 QCD fitting program.

Output data format

Three options for the output data format:

1. Complete *covariance matrix* of all $X-$section measurements.
2. Dependence of the average X-section on each systematic source + correlation matrix for the systematic sources.
3. Same as 2) but systematic error matrix is diagonalized

The (dis?)advantage of the first approach that the systematic uncertainties are frozen, they can not be modified by an external user (similar to Zeus offset method). The second-third approaches are very similar to the standard representations of the individual experiments, both "offset" and "lagrange multiplier" methods can be used.

Cross checks of the program

- Reasonable behaviour for toy dataset
- Passes trivial checks – no change of systematic uncertainties if same dataset is averaged to itself.
- Average of H1/Zeus data separately:

$$\chi^2/ndf_{\mathrm{H1\ only}} = 113.4/154$$
$$\chi^2/ndf_{\mathrm{Zeus\ only}} = 101.7/119$$

- For a set of random σ, α points χ^2 is calculated using the original data vs the average data:

Check Chi2 for several points

Che	Std Chi2	Ave1 Chi2	Ave2 Chi2	Ave1/Std-1	Ave2/Std-1
0	0.195525E+04	0.195525E+04	0.195525E+04	0.677409E-07	0.677409E-07
1	0.138108E+11	0.138108E+11	0.138108E+11	-0.156344E-08	-0.156344E-08
2	0.137381E+11	0.137381E+11	0.137381E+11	-0.113673E-08	-0.113673E-08
3	0.127106E+11	0.127106E+11	0.127106E+11	0.363016E-09	0.363015E-09
4	0.129243E+11	0.129243E+11	0.129243E+11	-0.593676E-09	-0.593678E-09
5	0.136068E+11	0.136068E+11	0.136068E+11	-0.998215E-10	-0.998215E-10
6	0.132560E+11	0.132560E+11	0.132560E+11	-0.113940E-09	-0.113939E-09
7	0.134828E+11	0.134828E+11	0.134828E+11	-0.458048E-09	-0.458048E-09
8	0.142298E+11	0.142298E+11	0.142298E+11	-0.151956E-08	-0.151956E-08

$\rightarrow$ OK

Average of all HERA data

Changes in systematic uncertainties:

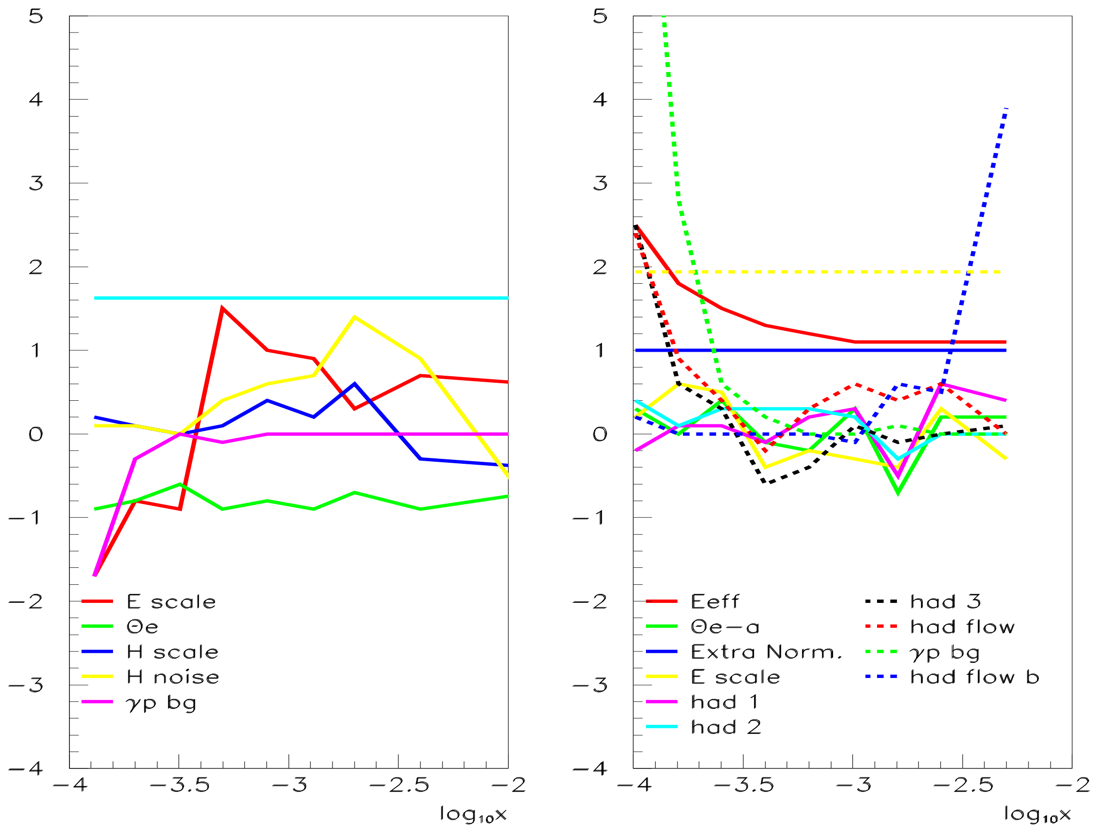
Fitted systematics:

		shift	uncertainty
1	zlumi1_zncepl	-1.2841	0.5836
2	h2_Ee_Spacal	0.6440	0.3281
3	h3_Ee_Lar_00	-0.8265	0.4435
4	h4_ThetaE_spacal	-0.2569	0.6566
5	h5_ThetaE_94-97	-0.1756	0.7802
6	h6_ThetaE_00	-0.3027	0.5288
7	h7_H_Scale_Spacal	0.3750	0.4813
8	h8_H_Scale_Lar	-0.8554	0.5353
9	h9_Noise_Hcal	-0.6404	0.3591
10	h10_GP_BG_Spacal	-0.1805	0.8260
11	h11_GP_BG_LAr	1.0769	0.8560
12	h12_BG_CC_94-97	0.2680	0.7883
13	h13_BG_CC_98-00	-1.0295	0.8589
14	h14_ChargeAsym	0.0246	0.9993
15	h1lumi1_SPACAL_bulk	-0.0696	0.5612
16	h1lumi2_SPACAL_MB	1.0815	0.6271
17	h1lumi3_LAr_94-97_e+p	-2.7111	0.6103
18	h1lumi4_LAr_e-p	-0.6585	0.7737
19	h1lumi5_LAr_2000	-2.5156	0.5885

- Good global $\chi^2/ndf = 533.9/601$
- Most of the changes are within 1σ
- Several systematic sources are reduced by factor 2 and more

Reduction of the systematic errors

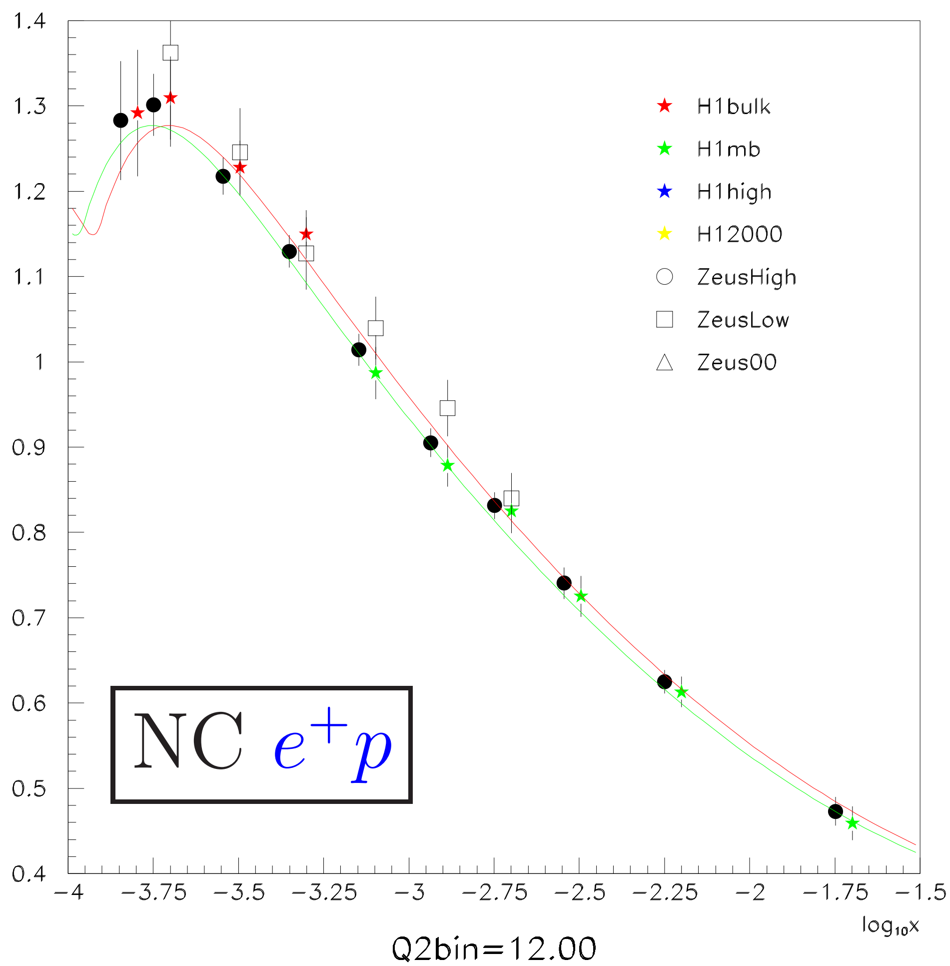
H1 and Zeus use different kinematic reconstruction methods – different shape vs x, Q^2.



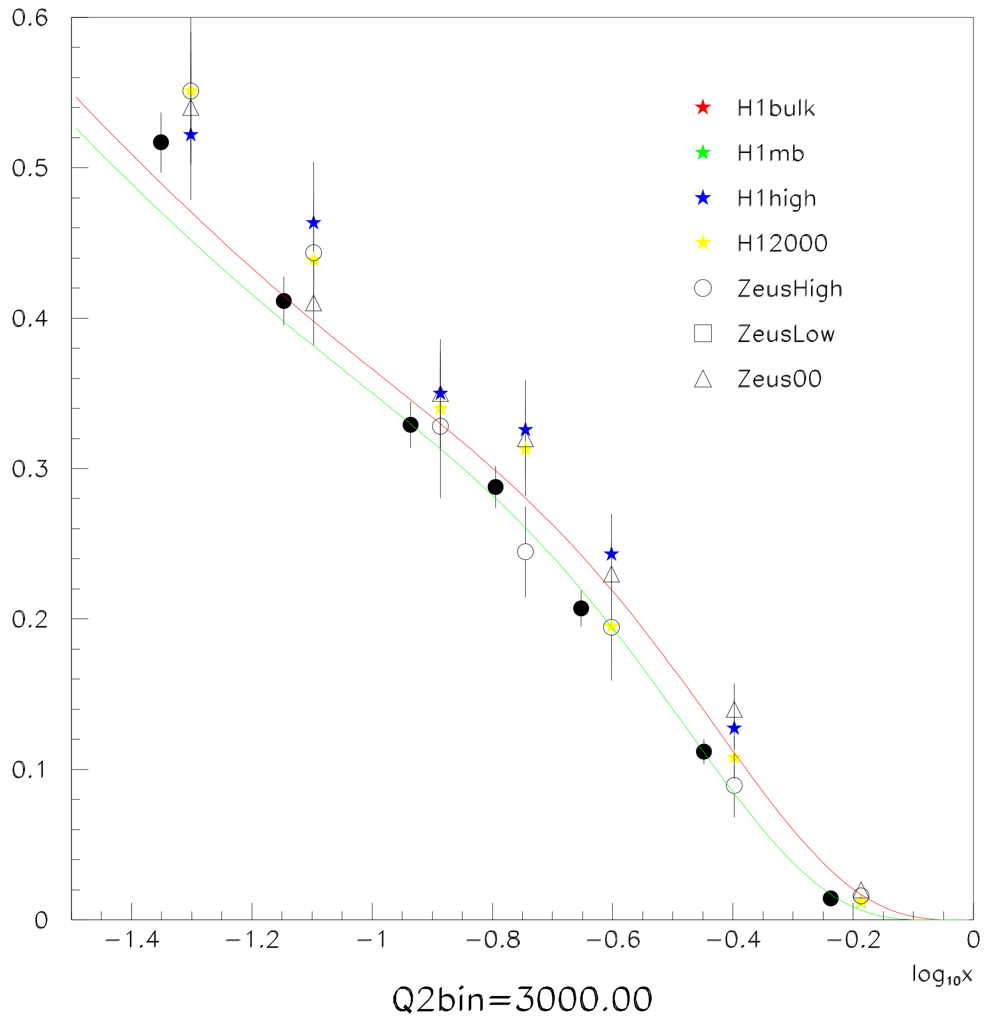
Sensitivity to syst. errors for $Q^2 = 6.5$ GeV^2

Average of all published HERA NC/CC data

16 individual data sets of NC/CC data published by H1 and Zeus collaborations. Examples for some Q^2 bins:

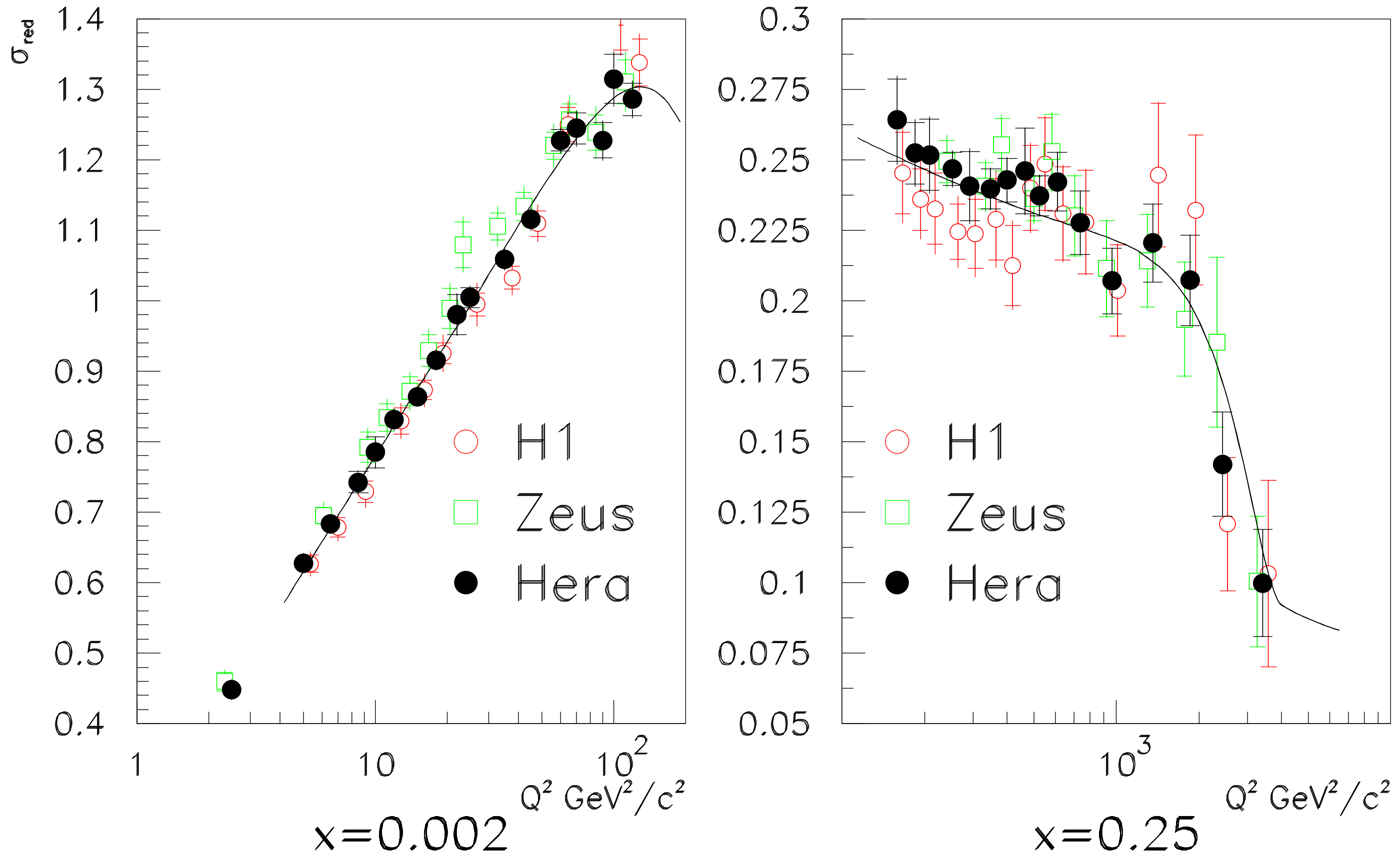


Average result for high Q^2 NC e^+p data



Factor of ~ 2 improvement in errors. For low $Q^2 \sim 10$ GeV^2 data reaches $< 2.0\%$ precision. Bins at $Q^2 = 3000.$ GeV^2 have 4% precision (plus 0.5% overall luminosity uncertainty).

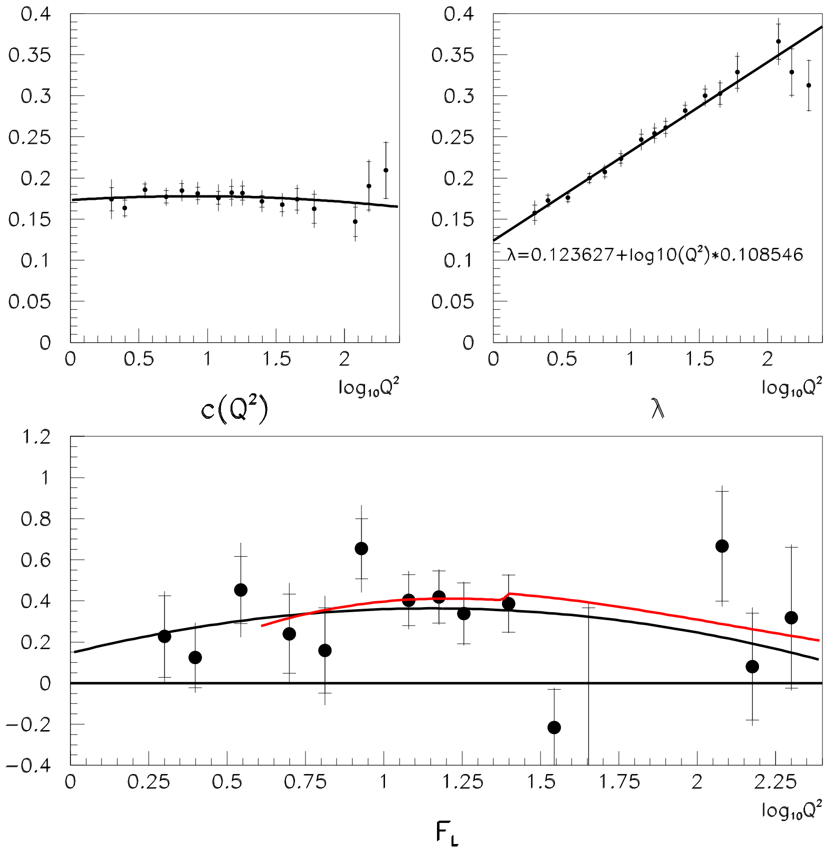
Q^2 dependence of NC e^+p X-section



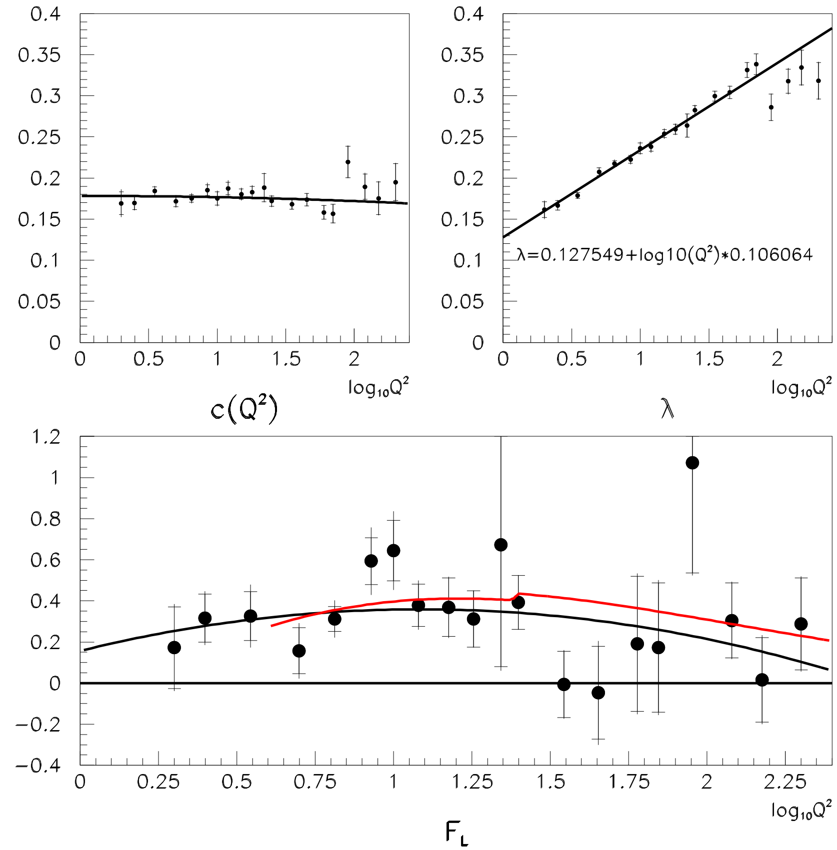
Here H1/Zeus data sets are *average* of all H1/Zeus published data.

$c \cdot x^{-\lambda}$ and F_L with the “shape” method

$$\sigma_{red} = C(Q^2) \cdot x^{-\lambda} - y^2/Y_+ \cdot F_L(Q^2)$$



H1



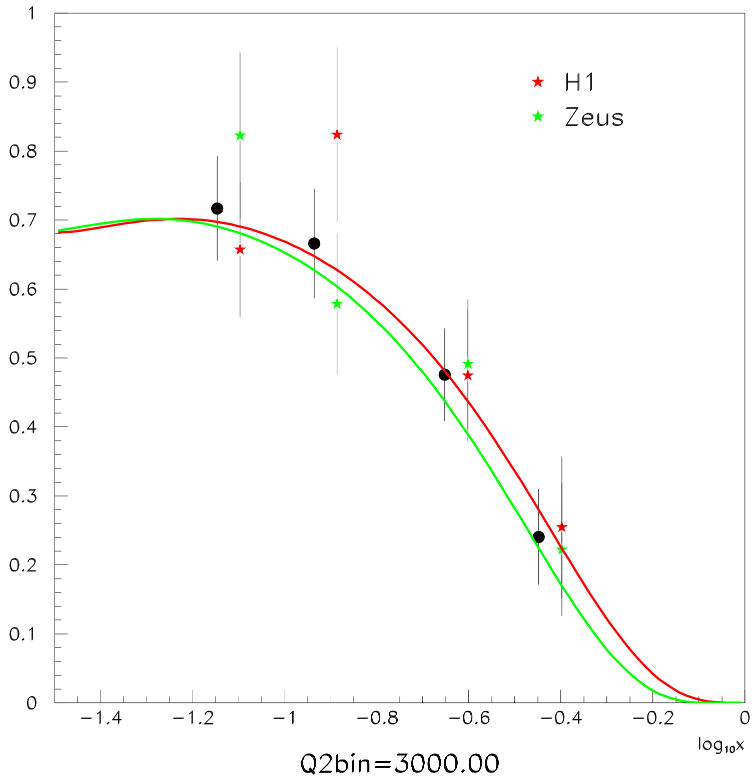
Hera

The errors for λ are reduced twice vs H1.

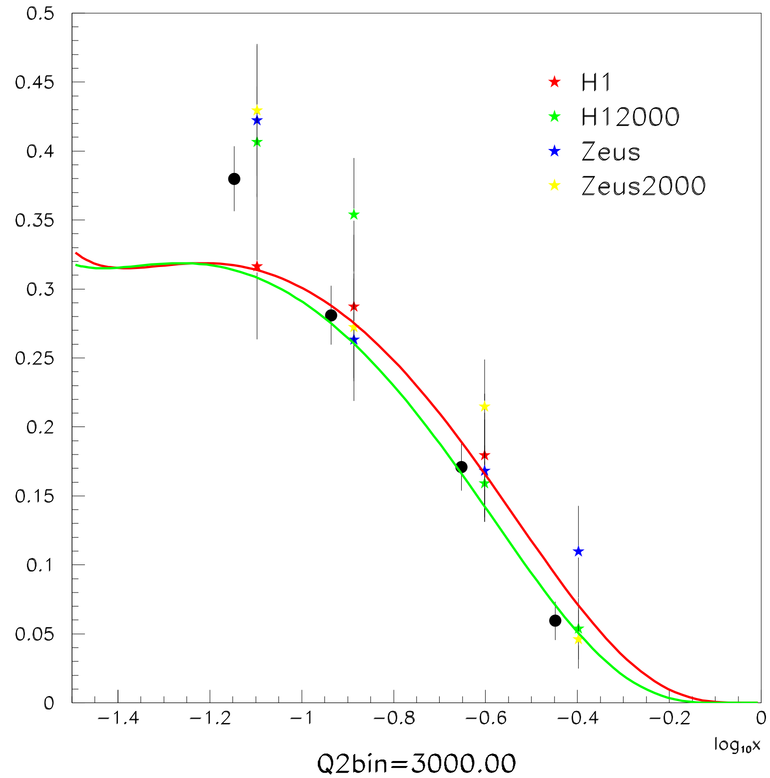
Conclusions/next steps

- Good agreement between H1 and Zeus data.
- Remarkable reduction of the total errors which needs further investigations → H1/Zeus joint averaging group → official HERA average.
- Model independent analysis of the average data
- QCD analysis of the average data.
  Joint H1/Zeus QCD fit ?
- Additional options for the averaging, for example xF_3 can be further improved if F_2 measured with e^-, e^+ data is averaged. F_L exrtraction can be also optimized.
- Stat. errors are significant across the kinematic plane → more X-section data is wellcome.

Extra: CC e^-p and e^+p



e^-p



e^+p

For e^+p, errors are reduced by about factor of 2 vs each dataset.