Averaging of the structure functions

- Why do we want to average structure functions data ?
- Some technical details
- First results – average of all HERA cross sections 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 (or even χ^2 function with complete systematic uncertainties) is much easier to handle. No more mainstream global QCD fits only, hard-core low-x theorist can also become experts in QCD fitting !
- 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.

The averaging procedure 1

Suppose we have a measurement A_{meas} with an uncertainty σ_A. Assuming Gaussian shape of the uncertainty, that measurement is equivalent to probability distribution for A:

$$\mathcal{P}\left(A_{true}\right) \sim \exp\left(\frac{-(A_{true} - A_{meas})^2}{\sigma_A^2}\right) \tag{1}$$

This can be re-written as a χ^2:

$$\chi^2(A_{true}) = (A_{true} - A_{meas})^2/\sigma_A^2 \tag{2}$$

Two measurements of A, A_1, σ_1 and A_2, σ_2 correspond to a χ^2 which is the sum of the two: $\chi^2_{sum} = \chi^2_1 + \chi^2_2$

Of course, χ^2_{sum} can be re-written in the form of Eq. 2. In this case A_{meas} is replaced by *Average* A_{ave} and σ_A is replaced by the error on this average. This form is obtained by minimizing χ^2_{sum}; it is easy to show that this leads to a usual averaging rule ($1/\sigma^2$ weights).

The averaging procedure 2

For the HERA measurements of F_2, we have both statistical (assumed uncorrelated) and systematic (often strongly correlated) uncertainties. A now standard way to represent these measurements is:

$$\chi^2(\{F_2^{true}\},\{\alpha\}) = \sum_i \frac{\left[F_2^{i,true} - \left(F_2^i + \sum_j \frac{\partial F_2^i}{\partial \alpha_j}\alpha_j\right)\right]^2}{\sigma^2_{F_2}} + \sum_j \frac{\alpha_j^2}{\sigma^2_{\alpha_j}}. \tag{3}$$

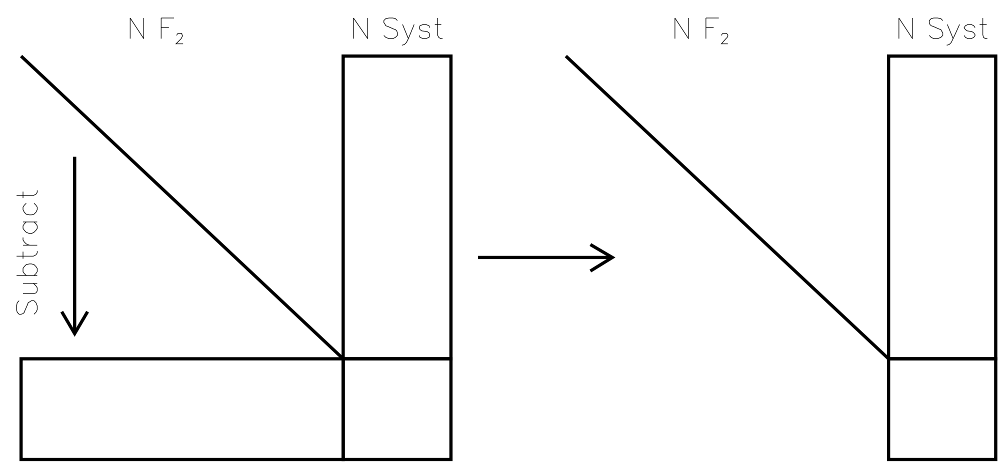
Here α_j — corresponds to a correlated systematic uncertainty source j (e.g. calorimeter electromagnetic energy scale).

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

Before performing QCD fit one can *average* F_2 values to get model independent world measurements of 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 vertex fit in H1 reconstruction:



(requires $\sim N_{F_2} \times N_{syst}^2$ 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 currently uses NC/CC cross section parametrization obtained in H1 QCD fit (normally small correction factor).
- First "All HERA" average for CC, NC e^+p and e^-p data.
- Output format directly suitable for H1 QCD fitting program.

Output data format

Three options for the output data format:

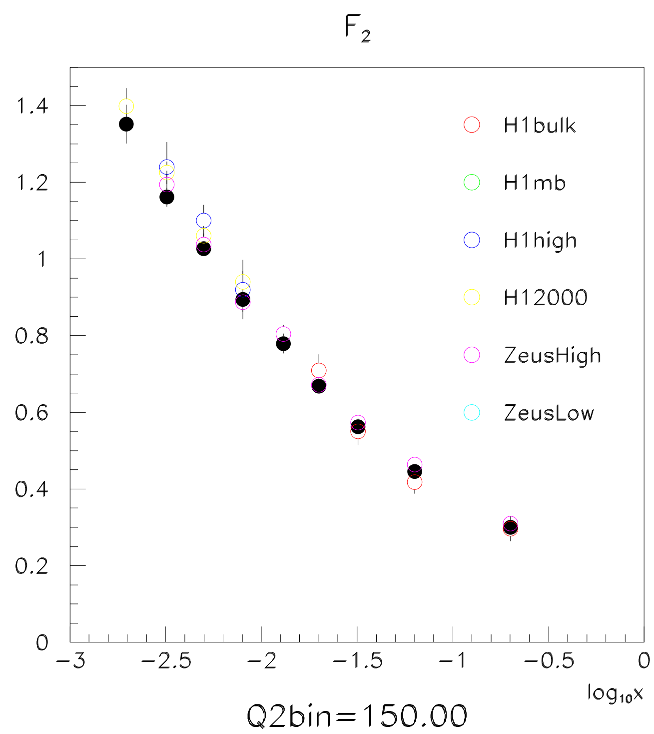
1. Complete *covariance matrix* of all $X-$section measurements.
2. Dependence of the average X-section on each systematic source + correlation matrix for the systematic sources.
3. Same as 2) but systematic error matrix is diagonalized

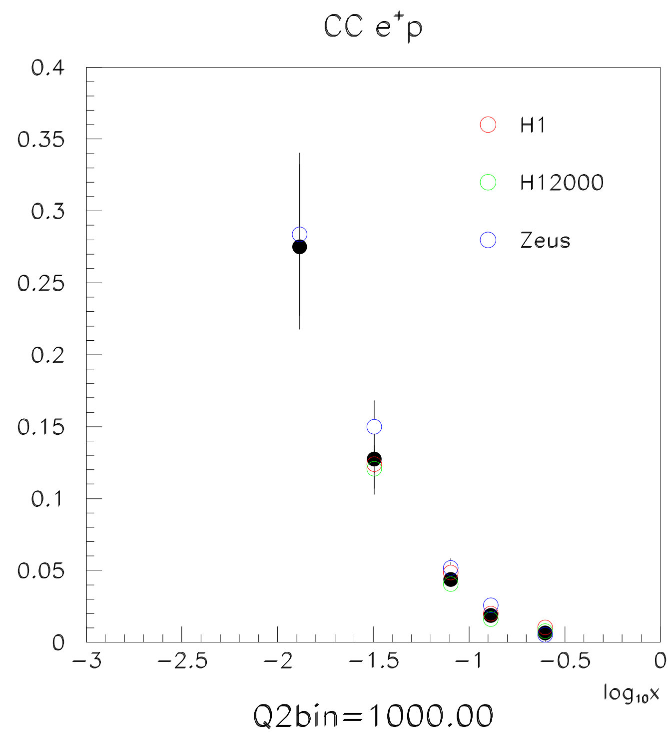
The (dis?)advantage of the first approach that the systematic uncertainties are frozen, they can not be modified by an external user (similar to Zeus offset method). The second-third approaches are very similar to the standard representations of the individual experiments, both "offset" and "lagrange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

Average of all published HERA NC/CC data

Input data sets (separate for e^+p and e^-p):

- H1: low Q^2 96-97, NC/CC 94-97, NC/CC 98 NC/CC 00
- Zeus: NC 96-97, CC/NC e^-p 98-99, e^+p 99-00





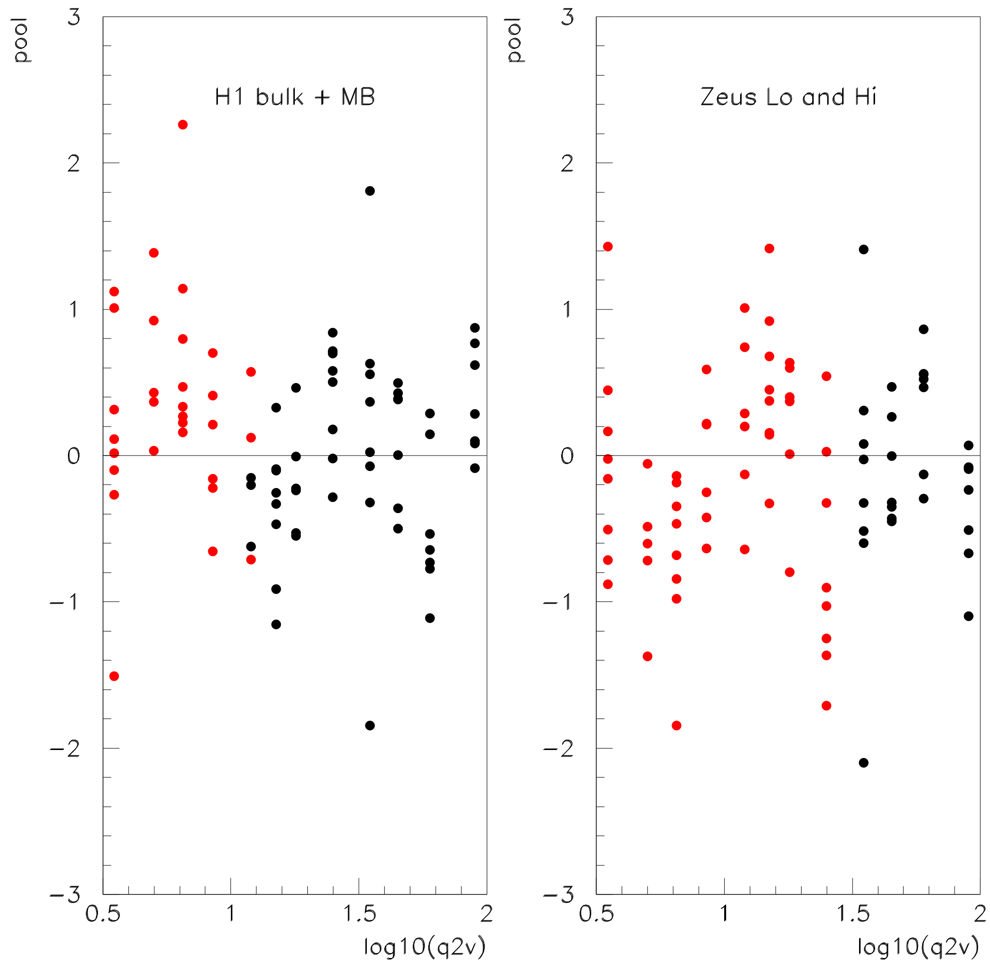
(Too) Good global $\chi^2/dof = 394/491$

Systematic uncertainties

For most of the 32 systematic sources (including 5 H1 + 3 Zeus normalizations) the change is reasonable. The larger changes are:

- H1 luminosity for e^+p 94-97 high Q^2 data: -2.4σ
- H1 luminosity for e^+p 2000 high Q^2 data: -2.3σ
- H1 luminosity for e^+p min. bias low Q^2 data: $+1.3\sigma$
- H1 hadronic energy scale in high Q^2 analysis: -1.5σ
- Zeus luminosity for e^+p data: -0.8σ
- Zeus electron energy scale for e^+p data: $+1.1\sigma$

Low Q^2 pool distribution



Indication of some differences at low Q^2.
New H1 result for low Q^2 will be published soon.

Goals

- Official average of HERA data.
  - Understanding of common correlated uncertainties (luminosity theory uncertainty 0.5%), what else ?
  - Understanding of correlations between datasets.
- Maintain updates of the average for the coming data
- Public release of the averaging code, documentation