Parton Distributions and their Uncertainties

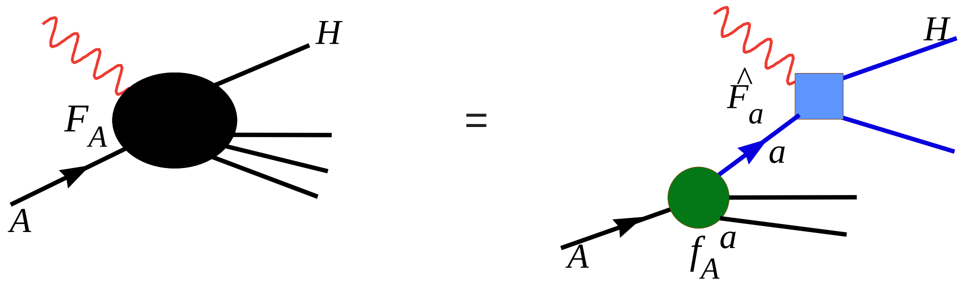
Jon Pumplin

DPF2002 – *Williamfburg* 5/25/02

CTEQ6 PDF analysis (J. Pumplin, D. Stump, W.K. Tung, J. Huston, H. Lai, P. Nadolsky [hep-ph/0201195])

- include new data sets
- include correlated systematic experimental errors
- evaluate uncertainties of the result:
  - Eigenvector PDF sets to map uncertainties
  - Lagrange multiplier results
- Universal PDF interface: Les Houches Accord
- Results:
  - W and Z production
  - parton-parton luminosities
  - gluon and quark distributions
- Measures of uncertainty
  - Measurement of α_s
  - Statistical bootstraps
- Outlook

Overview of QCD Global Analysis



Experimental
Input

Parton Distributions:
Nonperturbative parametrization at Q_0
DGLAP Evolution to Q

$$F_A^\lambda(x, \frac{m}{Q}, \frac{M}{Q}) = \sum_a f_A^a(x, \frac{m}{\mu}) \otimes \hat{F}_a^\lambda(x, \frac{Q}{\mu}, \frac{M}{Q}) + \mathcal{O}((\frac{\Lambda}{Q})^2)$$

Sources of uncertainty:

1. Experimental errors included in χ^2
2. Unknown experimental errors
3. Parametrization dependence
4. Higher-order corrections & Large Logarithms
5. Power Law corrections ("higher twist")

Fundamental difficulties:

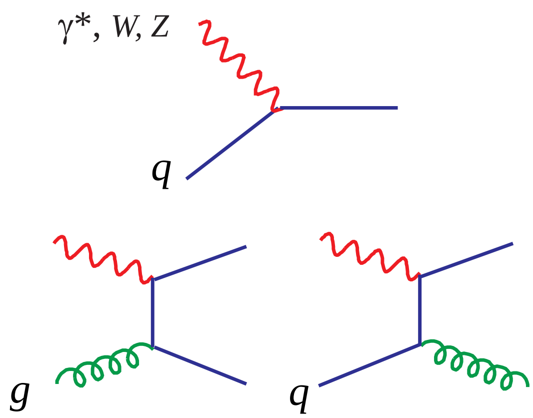
1. Good experiments run until systematic errors dominate; and the magnitude of systematic errors involves guesswork.
2. Systematic errors of the theory and their correlations cannot even be guessed.

Experimental Input

DIS

e N
μ N

ν N
$\bar{\nu}$ N

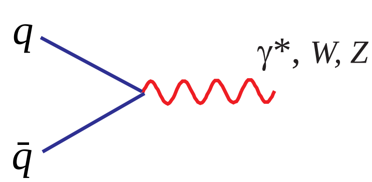


SLAC
BCDMS
NMC, E665
H1, ZEUS

CDHS, CHARM
CCFR
CHORUS

DY

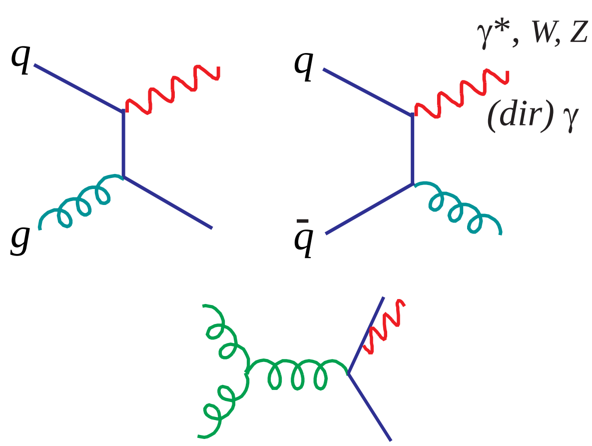
p N
π N
k N
$\bar{p}$ N



E605, E772
NA51
E866

Dir.Ph.

p N
π N
k N
$\bar{p}$ N



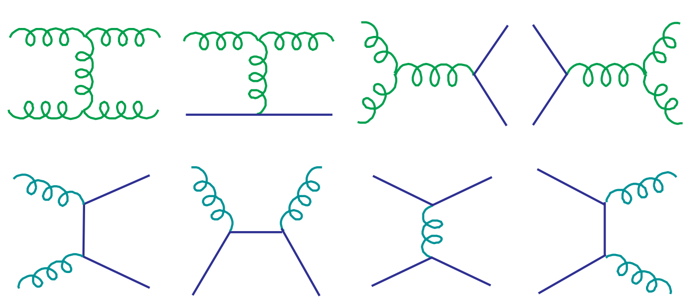
CDF, D0

WA70, UA6
E706

CDF, D0

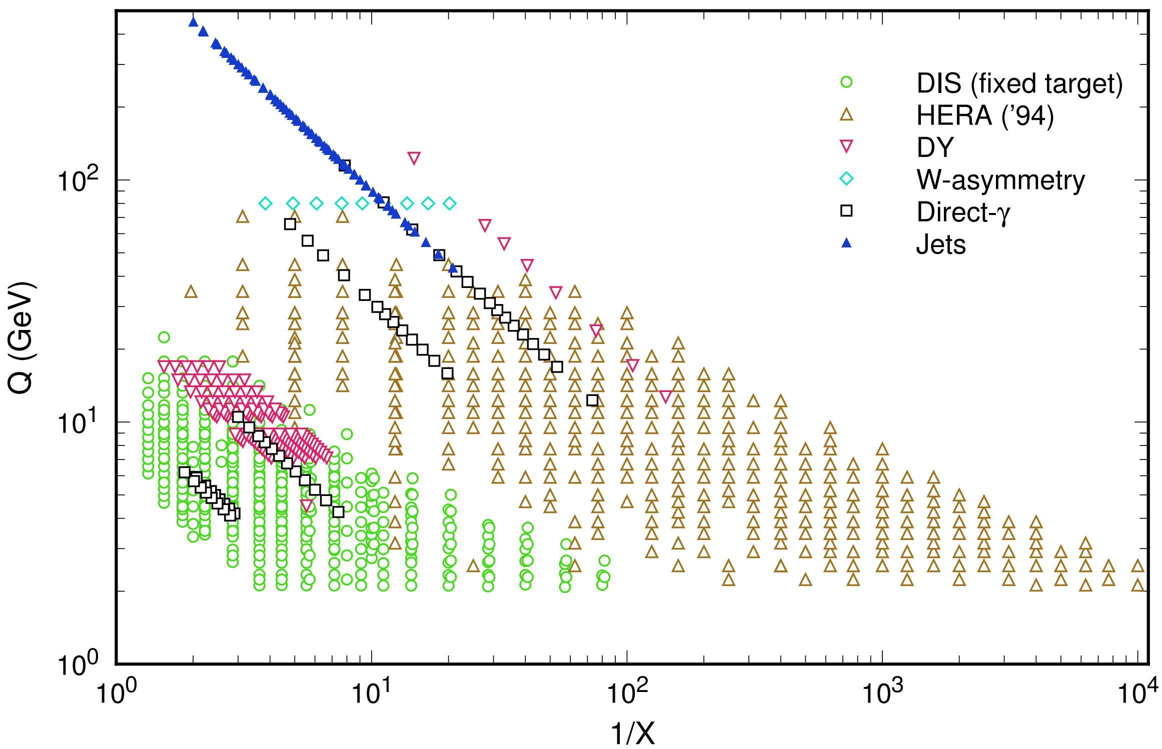
Jet Inc.

$\bar{p}$ p



CDF, D0

Kinematic region covered by data



A wide variety of data are tied together by the DGLAP renormalization group evolution equation.

Consistency – or lack thereof – between the experiments can be observed only by applying QCD to tie them together in a global fit.

All experiments that use hadrons in the initial state – Tevatron, LHC, and non-accelerator experiments – require the parton distributions for their analysis.

Selection of Data

CTEQ5			CTEQ6		
	#	sys		#	sys
BCDMS μp	168	no	BCDMS μp	339	yes
BCDMS μd	156	no	BCDMS μd	251	yes
H1 ep	172	no	H1a ep •	104	yes
			H1b ep •	126	yes
ZEUS ep	186	no	ZEUS ep •	229	yes
NMC μp	104	no	NMC μp	201	yes
NMC $\mu p/\mu n$	123	no	NMC $\mu p/\mu n$	123	yes
CCFR F_2 νN	87	no	CCFR F_2 νN	159	yes
CCFR F_3 νN	87	no	CCFR F_3 νN	87	no
E605 pp DY	119	no	E605 pp	119	no
NA51 pd/pp DY	1	no	NA51 pd/pp	1	no
E866 pd/pp DY	15	no	E866 pd/pp	15	no
CDF W	11	no	CDF W	11	no
CDF jet	33	yes	CDF jet	33	yes
DØjet	24	yes	DØJet •	90	yes

• New Data

(Direct photon data are not used because of uncontrolled systematic "k_T" effects, which need resummation)

CTEQ6 Global analysis

Input from Experiment:

- $\sim$ 2000 data points with $Q > 2\,\mathrm{GeV}$ from e, μ, ν DIS; lepton pair production (DY); lepton asymmetry in W production; high p_T inclusive jets; $\alpha_s(M_Z)$ from LEP

Input from Theory:

- NLO QCD evolution and hard scattering
- Parametrize at Q_0: $A_0\,x^{A_1}\,(1-x)^{A_2}\,(1+A_3 x^{A_4})$
- $s = \bar{s} = 0.4\,(\bar{u}+\bar{d})/2$ at Q_0; no intrinsic b or c

Construct effective $\chi^2_{\text{global}} = \sum_{\text{expts}} \chi^2_n$:

- χ^2_{global} includes the known systematic errors
- Minimizing χ^2_{global} yields "Best Fit" PDFs.
- Variation of χ^2_{global} in neighborhood of the minimum defines uncertainty limits.
- Estimate uncertainty as region of parameter space where $\chi^2 < \chi^2(\text{BestFit}) + T^2$ with $T \approx 10$.

(Quite different from Gaussian statistics because of unknown correlated systematic errors in theory and experiments – as measured by inconsistency between experiments).

Comment on Parametrization

For d_{val}, u_{val}, or g, we use

$$xf(x,Q_0) = A_0\, x^{A_1}\,(1-x)^{A_2}\, e^{A_3 x}(1+e^{A_4}x)^{A_5}$$

This corresponds to

$$\frac{d}{dx}\ln{(xf)} = \frac{A_1}{x} - \frac{A_2}{1-x} + \frac{c_3 + c_4 x}{1 + c_5 x}$$

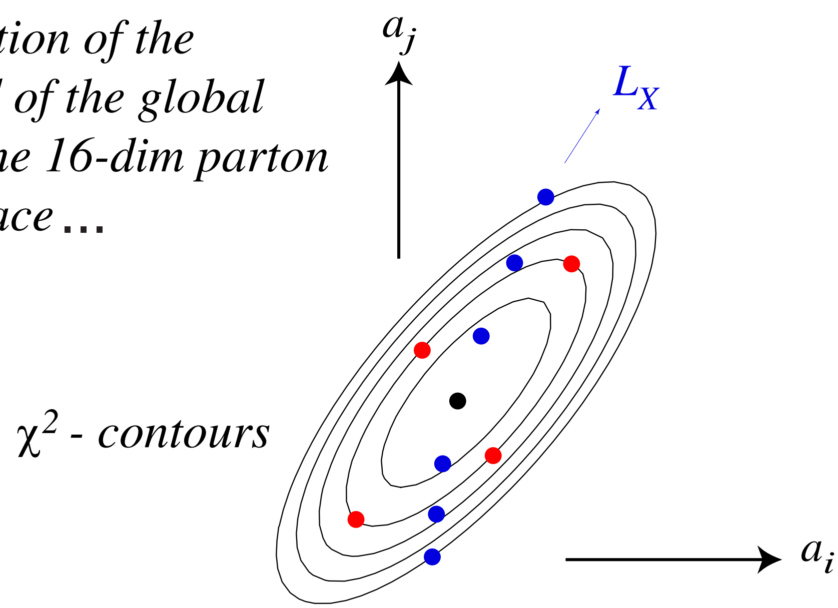
i.e., we add a 1:1 Padé form to the singular terms of the traditional $A_0\, x^{A_1}\,(1-x)^{A_2}$ parametrization.

A sufficiently flexible parametrization is important; but for convergence, there must not be too many "flat directions." For that reason, some of the parameters are frozen for some flavors.

(To measure a set of continuous PDF functions at Q_0 on the basis of a finite set of data points would appear to be an ill-posed mathematical problem. However, this difficulty is not so severe as might be expected since the actual predictions of interest that are based on the PDFs are discrete quantities. In particular, fine-scale structure in x in the PDFs at Q_0 tend to be smoothed out by evolution in Q. They correspond to flat directions in χ^2 space, so they are not accurately measured; but they have little effect on the applications of interest.)

MSU/CTEQ uncertainty methods

2-dim illustration of the neighborhood of the global minimum in the 16-dim parton parameter space ...



- Hessian Matrix Method: eigenvectors of error matrix yield 40 sets $\{S_i^{\pm}\}$ that are displaced "up" or "down" by $\Delta\chi^2 = 100$ from the best fit. Get error by sum of squares and construct extreme PDFs for any observable; or simply look at extremes from the 40 sets.

- Lagrange Multiplier Method: Track χ^2 as function of F (e.g. σ_W) by minimizing $\chi^2 + \lambda F$. Yields special-purpose PDFs that give extremes of σ_W, or $\langle y \rangle$ for rapidity distribution of W, or σ for $t\bar{t}$ production; or $\sigma_{t\bar{t}}(\sqrt{s} = 14\,\text{TeV})/\sigma_{t\bar{t}}(\sqrt{s} = 2\,\text{TeV})$, or M_W mass measurement error,. . .

Hessian (Error Matrix) method

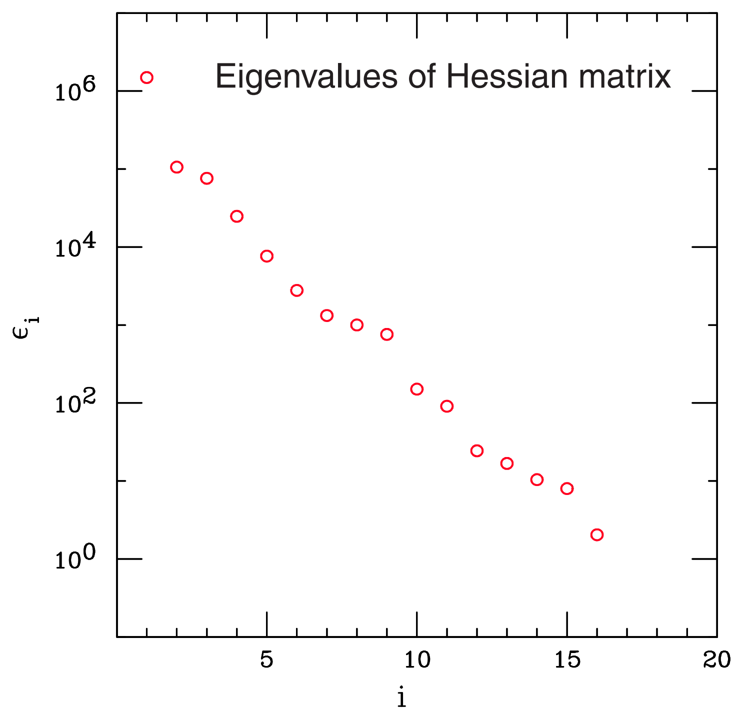
Classical error formulae

$$\Delta\chi^2 = \sum_{ij}(a_i - a_i^{(0)})(H)_{ij}(a_j - a_j^{(0)})$$

$$(\Delta F)^2 = \Delta\chi^2 \sum_{ij} \frac{\partial F}{\partial a_i}(H^{-1})_{ij}\frac{\partial F}{\partial a_j}$$

Hessian matrix H is inverse of error matrix.

Direct application fails because of extreme differences in variation of χ^2 for different directions in the space of fitting parameters ("steep" and "flat" directions), as shown by a huge range of eigenvalues of H:



Convergence problems are solved by an iterative method that finds and rescales the eigenvectors of H, leading to a diagonal form

$$\Delta\chi^2 = \sum_i z_i^2$$

$$(\Delta F)^2 = \sum_i \left(F(S_i^{(+)}) - F(S_i^{(-)}) \right)^2$$

where $S_i^{(+)}$ and $S_i^{(+)}$ are PDF sets that are displaced along the eigenvector directions. The iterative

procedure is available in FORTRAN at http://www.pa.msu.edu/~pumplin/iterate/

χ^2 and Systematic Errors

The simplest definition

$$\chi_0^2 = \sum_{i=1}^{N} \frac{(D_i - T_i)^2}{\sigma_i^2} \qquad \begin{cases} D_i = & \text{data} \\ T_i = & \text{theory} \\ \sigma_i = & \text{``expt. error''} \end{cases}$$

is optimal for random Gaussian errors,

$$D_i = T_i + \sigma_i r_i \qquad \text{with} \qquad P(r) = \frac{e^{-r^2/2}}{\sqrt{2\pi}}.$$

With systematic errors,

$$D_i = T_i(a) + \alpha_i r_{\text{stat},i} + \sum_{k=1}^{K} r_k \beta_{ki}\,.$$

The fitting parameters are $\{a_\lambda\}$ (theoretical model) and $\{r_k\}$ (corrections for systematic errors).

Published experimental errors:

- α_i is the ‘standard deviation’ of the random uncorrelated error.

- β_{ki} is the ‘standard deviation’ of the k th (completely correlated!) systematic error on D_i.

To take into account the systematic errors, we define

$$\chi'^2(a_\lambda, r_k) = \sum_{i=1}^{N} \frac{\left(D_i - \sum_k r_k \beta_{ki} - T_i\right)^2}{\alpha_i^2} + \sum_k r_k^2,$$

and minimize with respect to $\{r_k\}$. The result is

$$\widehat{r}_k = \sum_{k'} \left(A^{-1}\right)_{kk'} B_{k'}, \qquad \text{(systematic shift)}$$

where

$$\begin{aligned} A_{kk'} &= \delta_{kk'} + \sum_{i=1}^{N} \frac{\beta_{ki}\beta_{k'i}}{\alpha_i^2} \\ B_k &= \sum_{i=1}^{N} \frac{\beta_{ki}\left(D_i - T_i\right)}{\alpha_i^2}. \end{aligned}$$

The $\widehat{r}_k$'s depend on the PDF model parameters $\{a_\lambda\}$. We can solve for them explicitly since the dependence is quadratic.

We then minimize the remaining $\chi^2(a)$ with respect to the model parameters $\{a_\lambda\}$.

- $\{a_\lambda\}$ determine $f_i(x, Q_0^2)$.

- $\{\widehat{r}_k\}$ are are the optimal "corrections" for systematic errors; i.e., systematic shifts to be applied to the data points to bring the data from different experiments into compatibility, within the framework of the theoretical model.

Comparison to Data

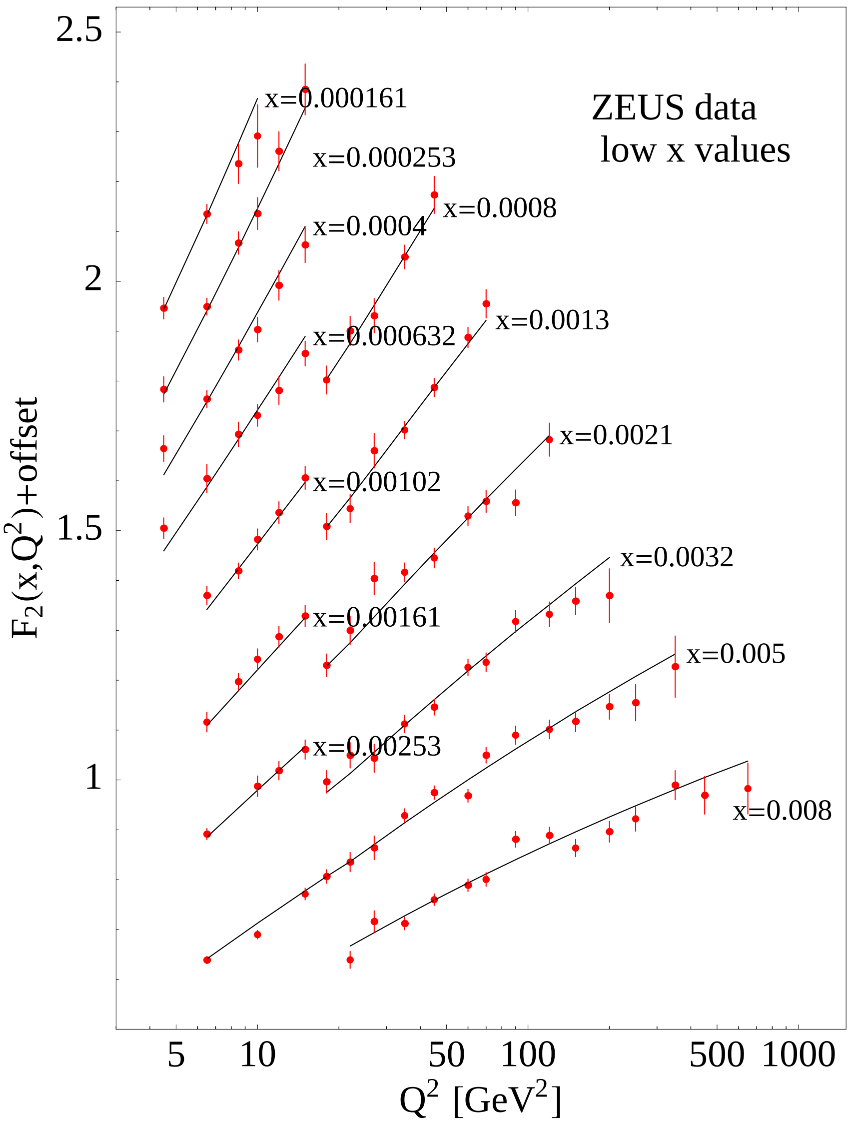
Comparison of the CTEQ6M fit to data with correlated systematic errors.

data set	N_e	χ_e^2	χ_e^2/N_e
BCDMS p	339	377.6	1.114
BCDMS d	251	279.7	1.114
H1a	104	98.59	0.948
H1b	126	129.1	1.024
ZEUS	229	262.6	1.147
NMC F2p	201	304.9	1.517
NMC F2d/p	123	111.8	0.909
DØ jet	90	69.0	0.766
CDF jet	33	48.57	1.472

Other data sets:

CCFR	ν DIS	(150/156)
E605	Drell-Yan	(95/119)
E866	Drell-Yan	(6/15)
CDF	W-lepton asymmetry	(10/11)

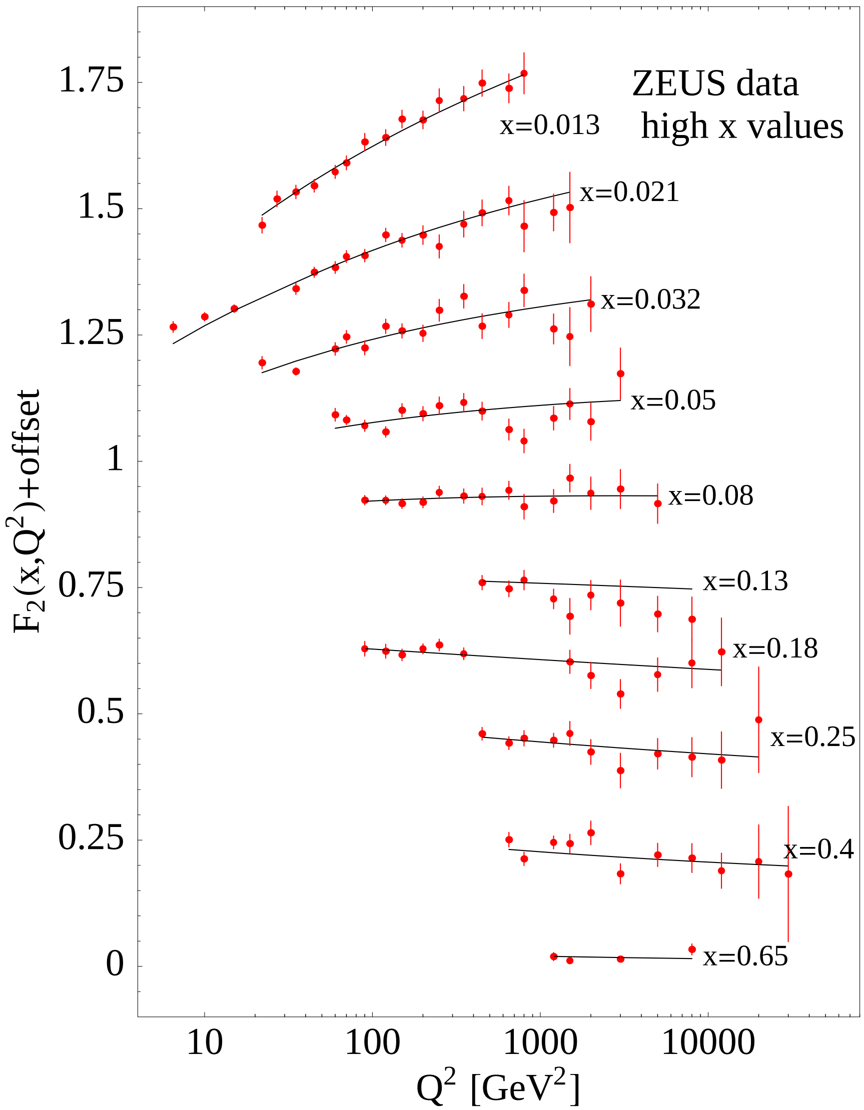
CTEQ6M fit to ZEUS data at low x



The data points include the estimated corrections for systematic errors. That is to say, the central values plotted have been shifted by an amount that is consistent with the estimated systematic errors, where the systematic error parameters are determined using other experiments via the global fit.

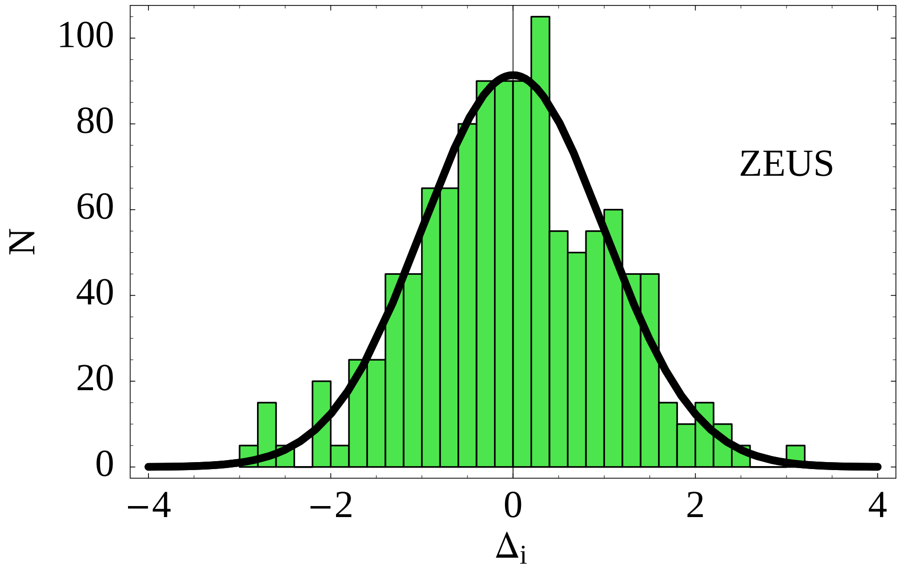
The error bars are statistical errors only.

CTEQ6M fit to ZEUS data at high x

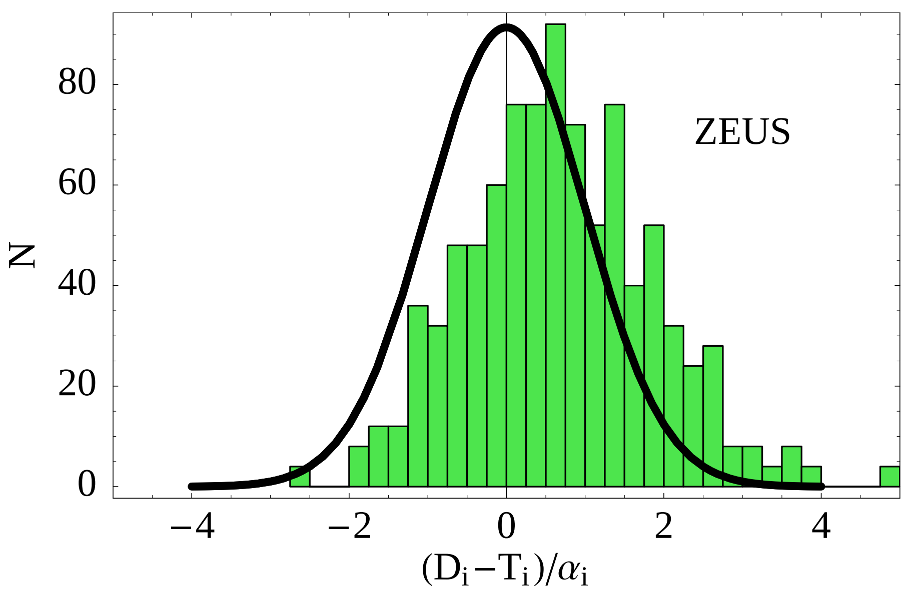


The data points include the estimated corrections for systematic errors.

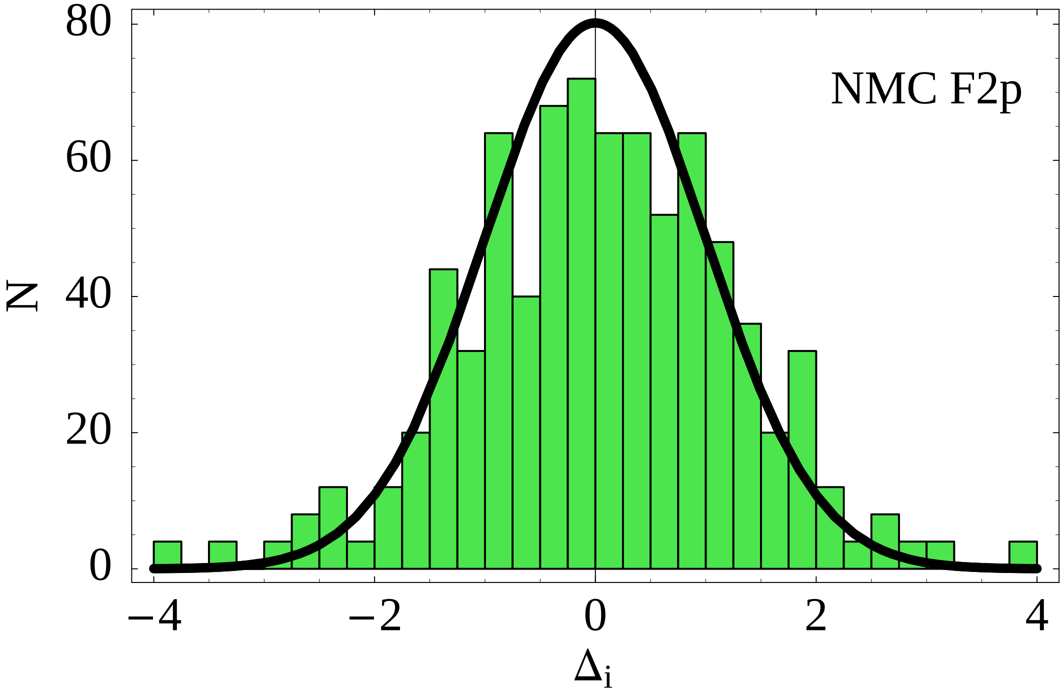
The error bars are statistical errors only.



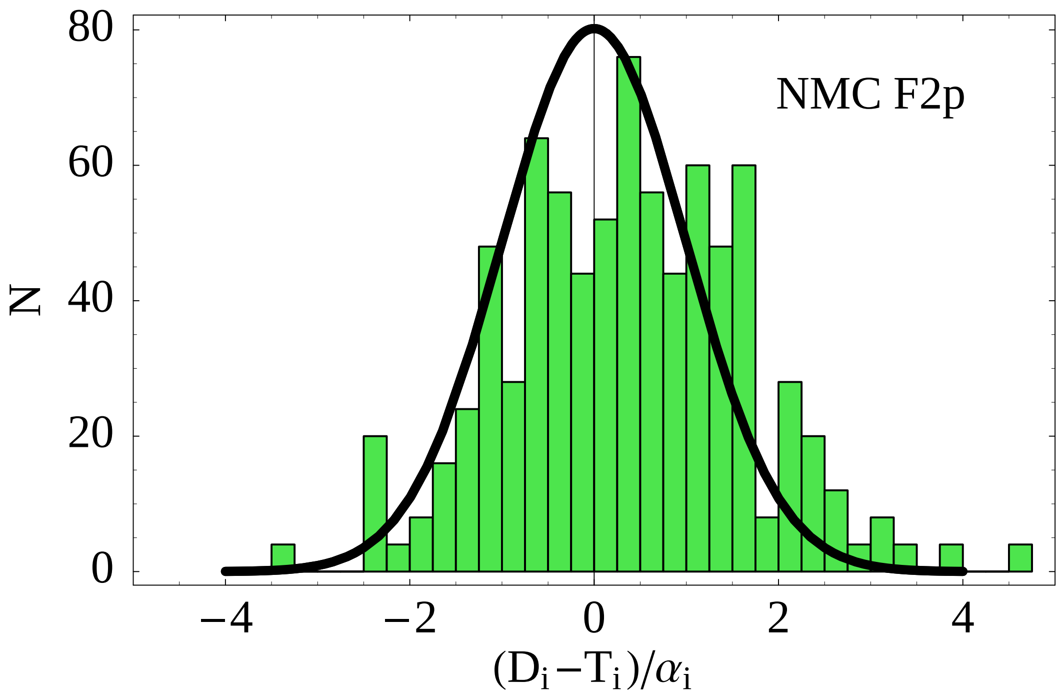
(a) Histogram of residuals for the ZEUS data. The curve is a Gaussian of width 1.



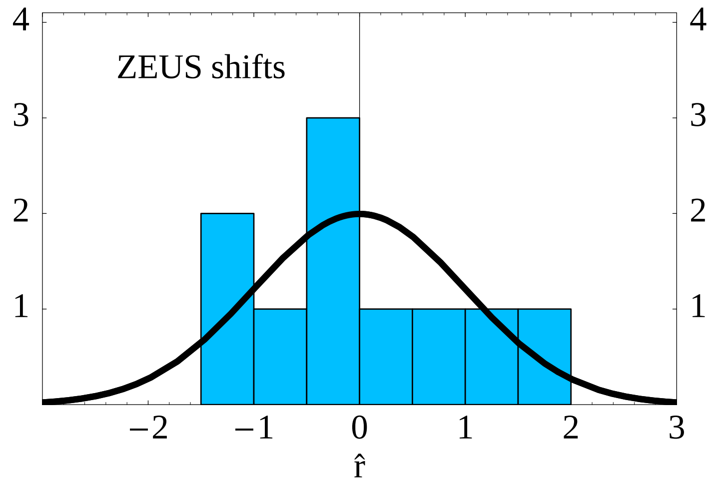
(b) A similar comparison but without the corrections for systematic errors on the data points.



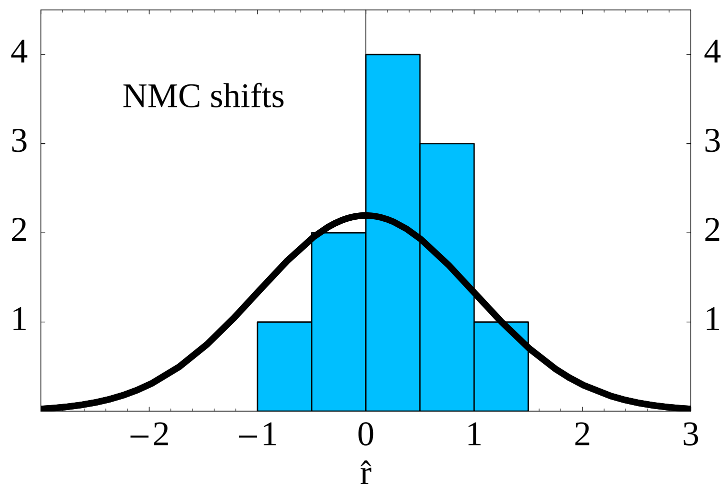
(a) Histogram of residuals for the NMC data.



(b) A similar comparison but without the corrections for systematic errors on the data points.

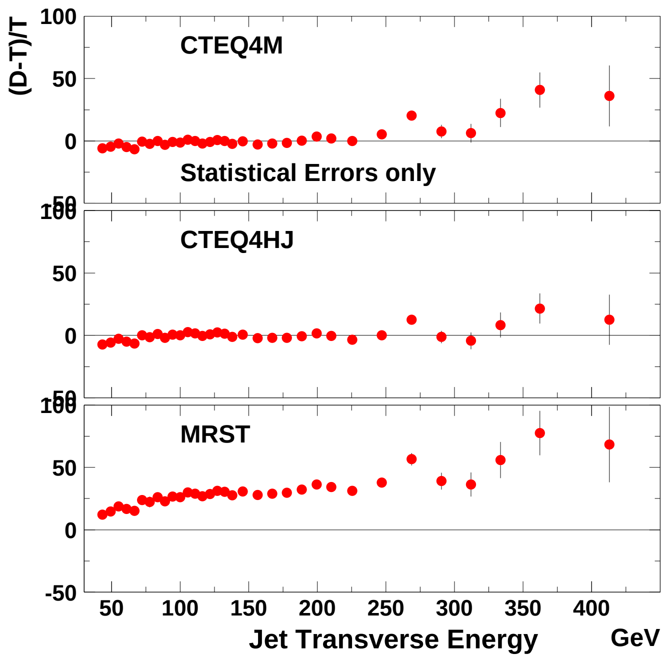


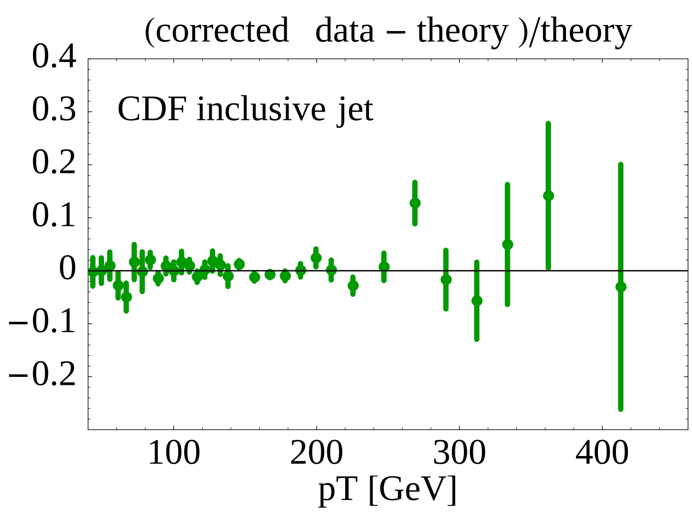
Systematic shifts for the ZEUS data
(10 systematic errors)



Systematic shifts for the NMC data
(11 systematic errors)

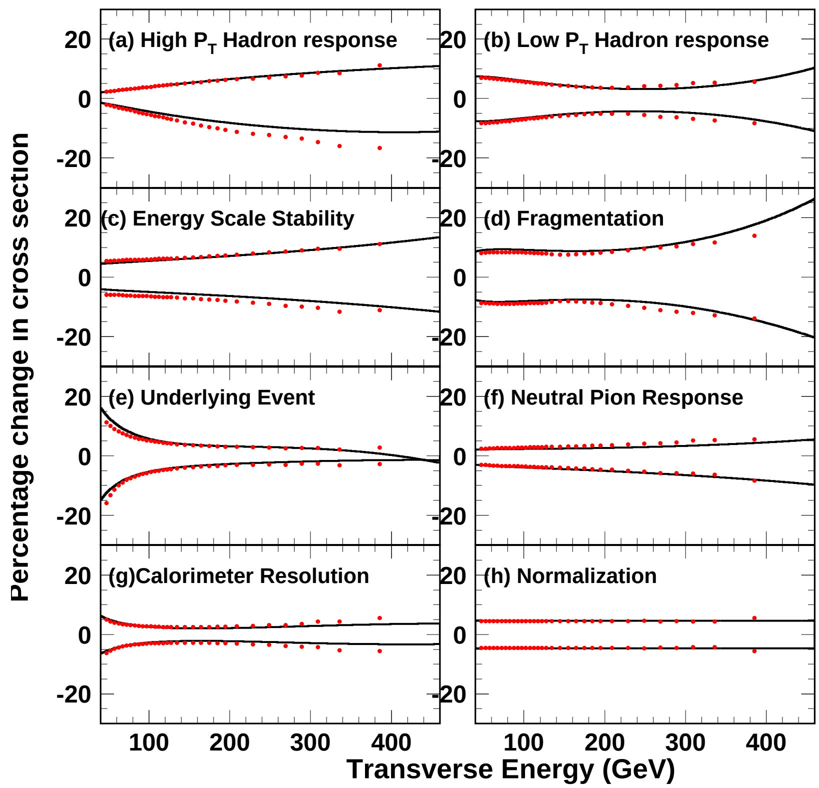
CDF inclusive jet cross section





Recall that these inclusive jet cross section measurements provided the first major stimulus to the study of PDF uncertainties – in particular, the uncertainties associated with choices made in the form of parametrizations at Q_0.

CDF Inclusive jets – systematic errors

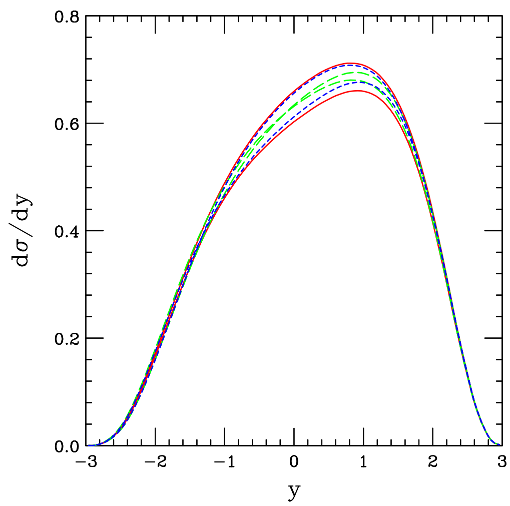


k	$\widehat{r}_k$
1	-0.511
2	0.816
3	0.022
4	1.347
5	-1.307
6	0.089
7	-0.222
8	xxx

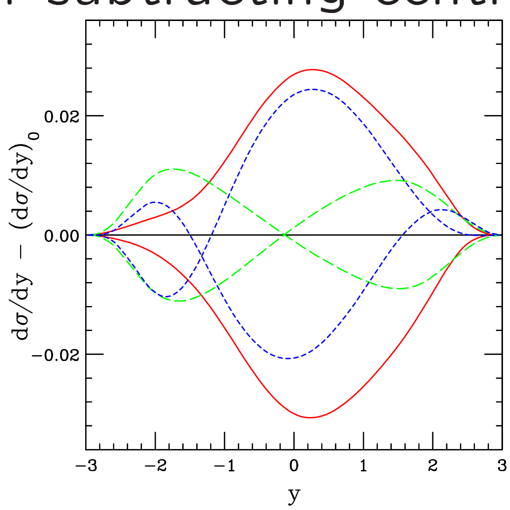
W rapidity distributions

Our methods allow us to calculate the extreme predictions due to PDF uncertainty for whatever quantity is of experimental interest.

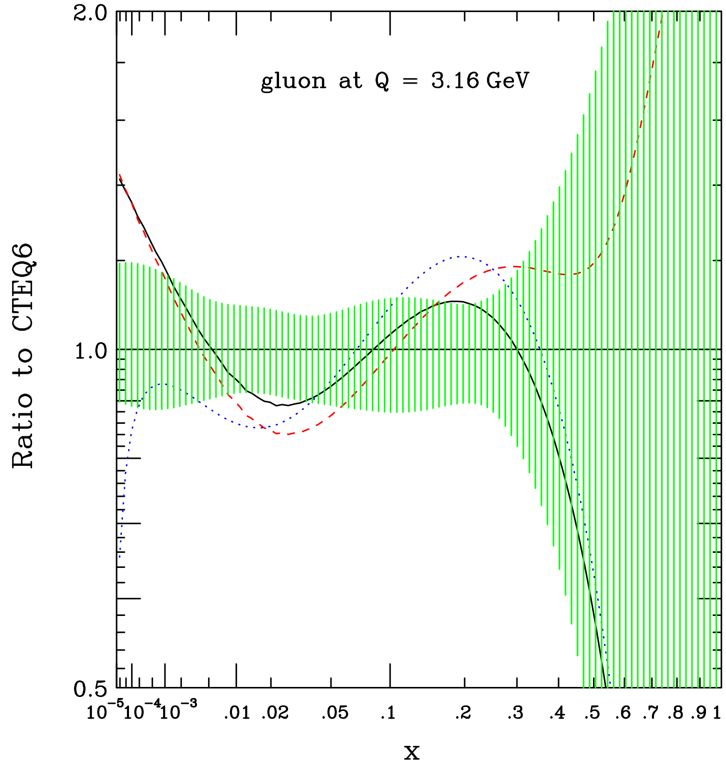
For example, extremes of σ_W, $\langle y \rangle$, $\langle y^2 \rangle$ for W production at FNAL – relevant for M_W measurement:



Same curves after subtracting central values. . .



Uncertainty of the gluon distribution



Uncertainty bands (envelope of possible fits) for the gluon distribution at $Q^2 = 10\,\mathrm{GeV}^2$.

The curves correspond to

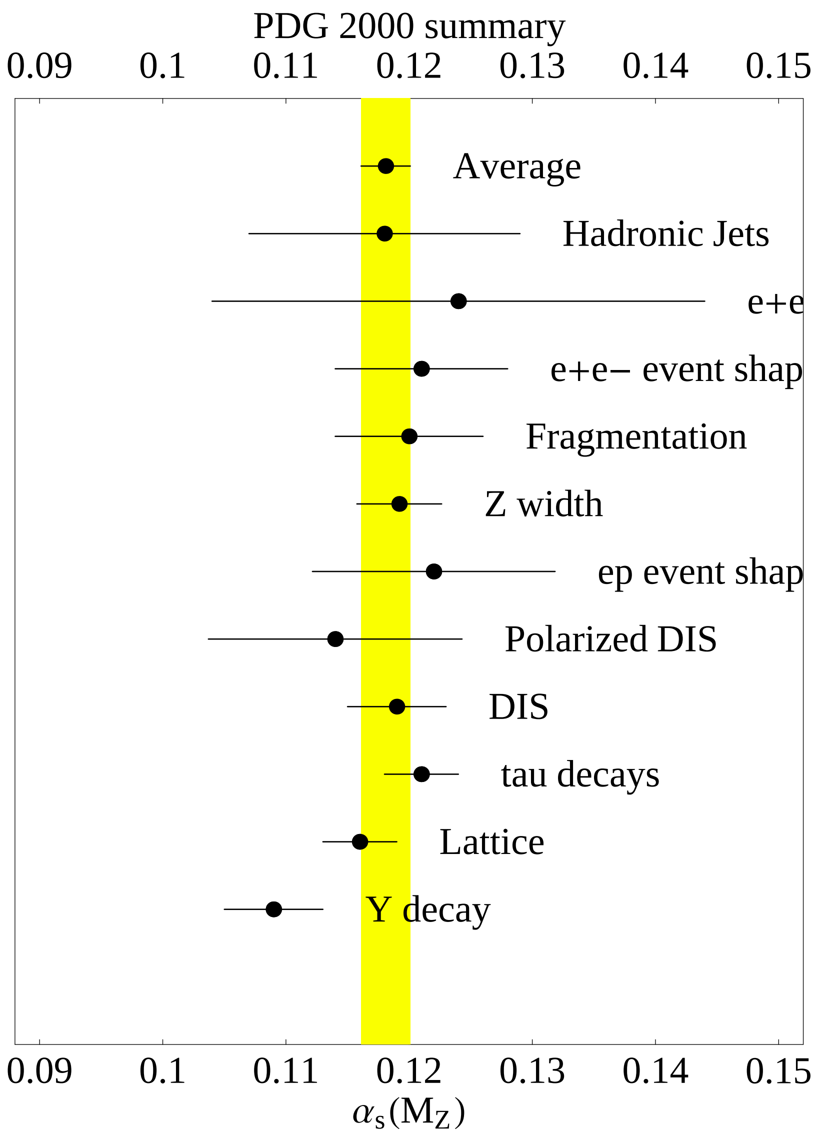
CTEQ5M1 (solid)
CTEQ5HJ (dashed)
MRST2001 (dotted)

Ironically, the differences between these is comparable to the estimated uncertainty!

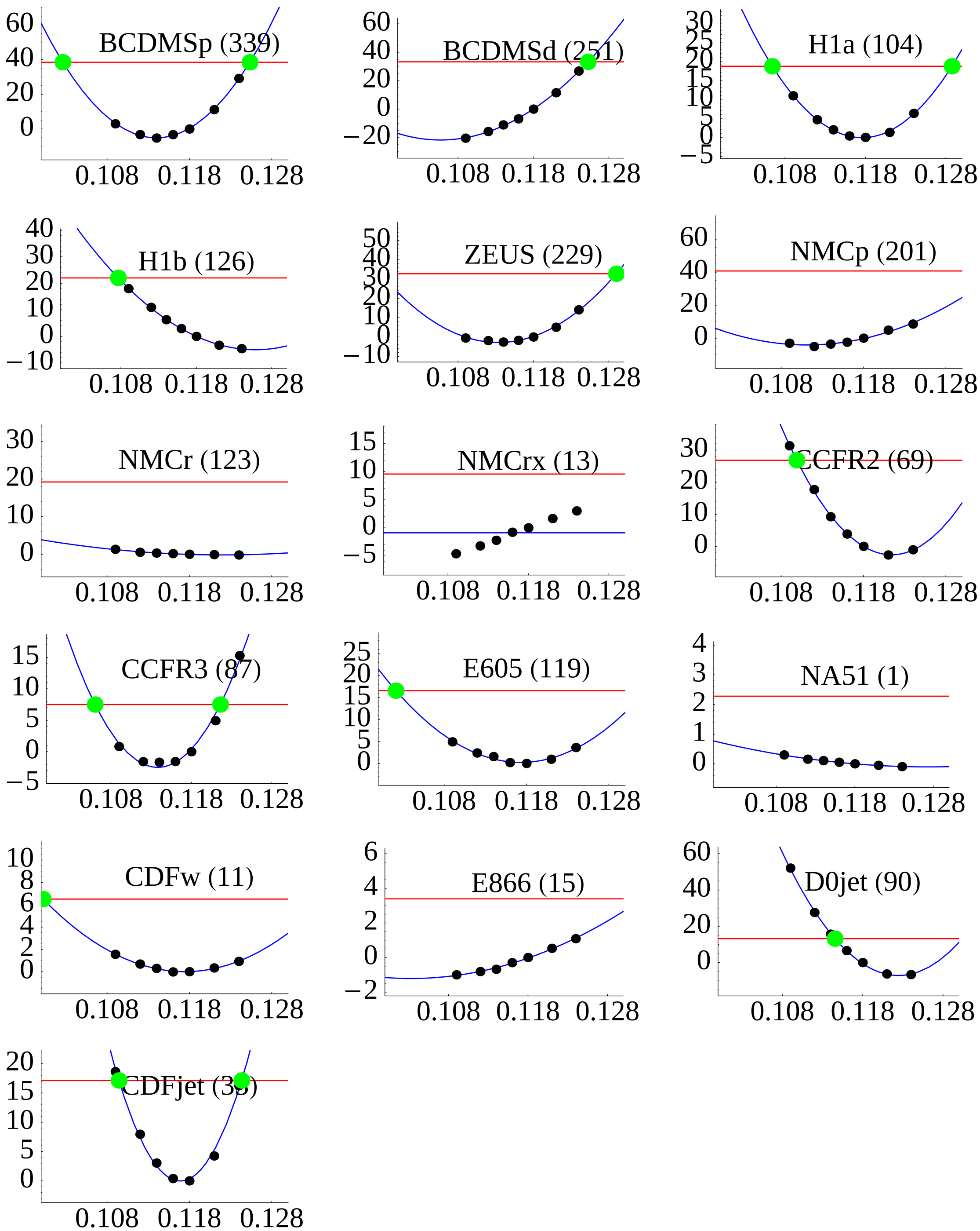
The uncertainties of quark distributions (not shown) are smaller than this gluon uncertainty, because the DIS measurements are sensitive to the square of the quark charge in leading order. The uncertainties of all PDFs decrease with increasing Q – "convergent evolution"

Measurement of α_s

We find that the CTEQ6 analysis is nicely consistent with the World Average determination of $\alpha_s(M_Z)$. But it is not precise enough to improve that value.



χ^2 versus $\alpha_S(M_Z)$ for individual data sets in CTEQ6

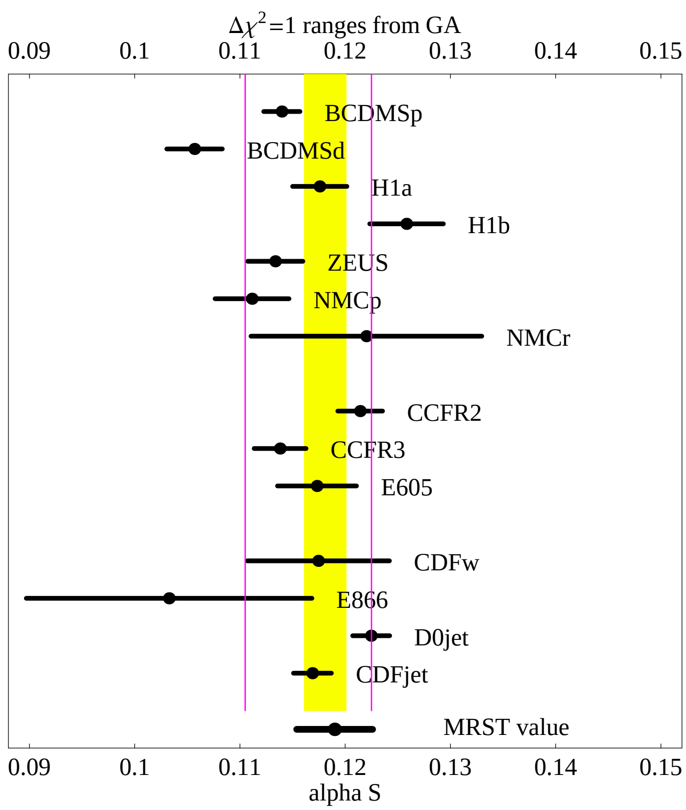


Measurment of $\alpha_S(M_Z)$:

If assume $\Delta\chi^2 = 1$ criterion in each experiment, the experiments are inconsistent.

Our error estimate ($T = 10$) is
$\alpha_S(M_Z) = 0.1165 \pm 0.0065$

This corresponds to somewhat conservative assumptions – perhaps to be thought of as an effective "2 σ" limit. Hence it is comparable to the MRST limit based on $T = 5$.

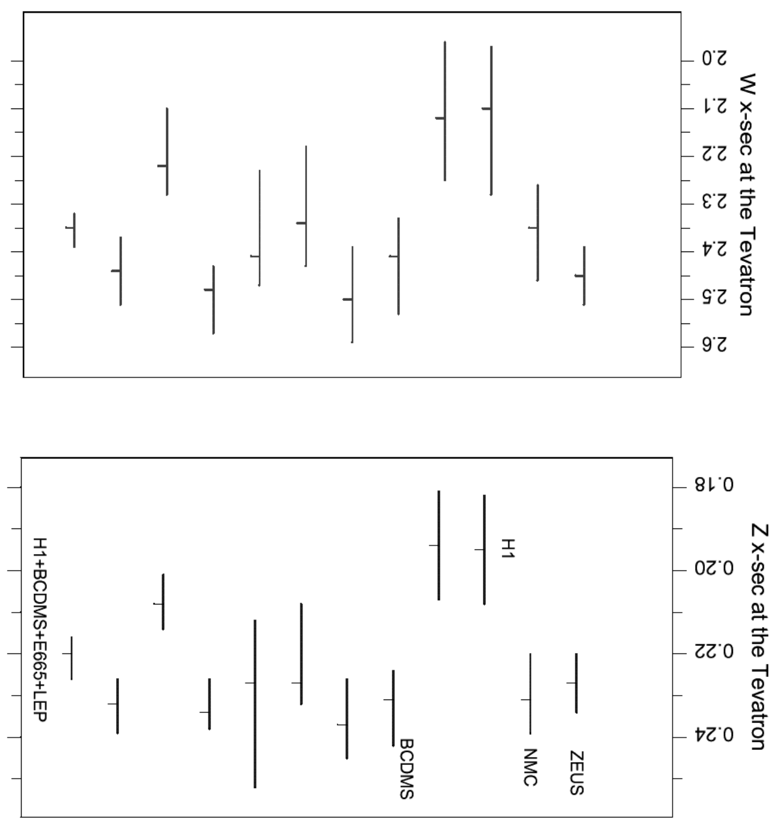


Similar situation for W and Z cross sections

When a strict $\Delta\chi^2 = 1$ criterion was applied to self-consistent subsets of the experiments, the subsets were not consistent with each other.

The true error is therefore considerably larger than $\Delta\chi^2 = 1$ would imply.

Giele-Keller 2001 (hep-ph/0104053)



New ways to measure consistency of fit

(Work in progress with John Collins)

Key idea: In addition to the

Hypothesis-testing criterion $\Delta\chi^2 \sim \sqrt{2N}$

we use the stronger

Parameter-fitting criterion $\Delta\chi^2 \sim 1$

The parameters here are relative weights assigned to various experiments, or to results obtained using various experimental methods. Examples:

- Plot minimum χ_i^2 vs. $\chi_{\text{tot}}^2 - \chi_i^2$, where χ_i^2 is one of the experiments, or all data on nuclei, or all data at low Q^2,. . .

or

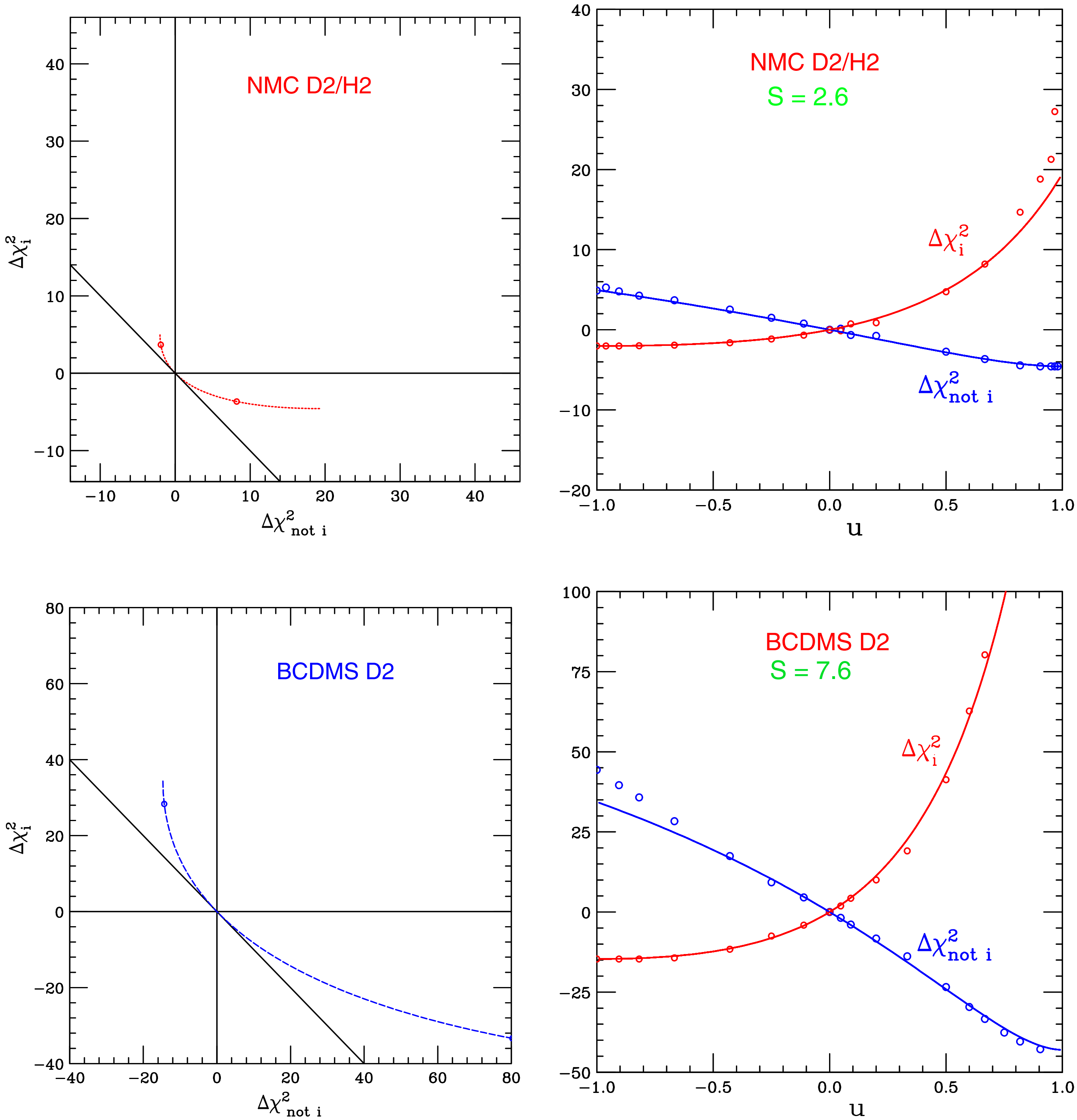
- Plot both as function of Lagrange multiplier u where $(1-u)\chi_i^2 + (1+u)(\chi_{\text{tot}}^2 - \chi_i^2)$ is the quantity minimized.

Can obtain quantitative results by fitting to a model with a single common parameter p:

$$\chi_i^2 = A + \left(\frac{p}{\sin\theta}\right)^2 \quad \Rightarrow \quad p = 0 \pm \sin\theta$$

$$\chi_{\text{not } i}^2 = B + \left(\frac{p-S}{\cos\theta}\right)^2 \quad \Rightarrow \quad p = S \pm \cos\theta$$

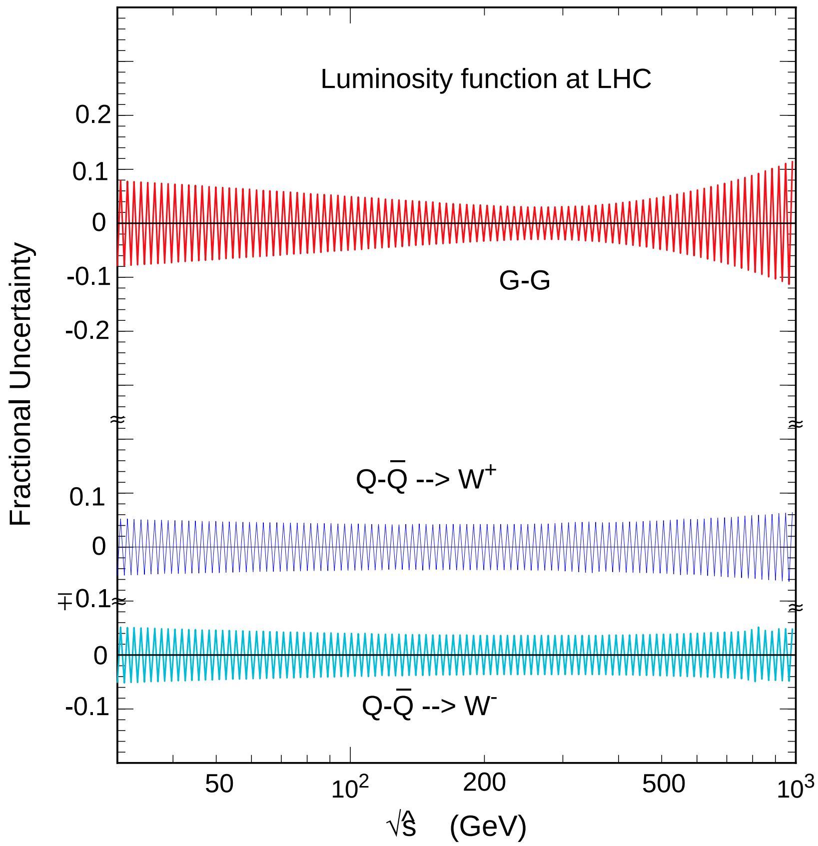
These differ by $S \pm 1$, i.e., by S "standard deviations"



Fits to 8 of the experiments in the CTEQ5 analysis

Expt	1	2	3	4	5	6	7	8
S	2.7	3.3	3.3	4.2	5.3	7.6	7.4	8.3
$\tan\phi$	0.56	0.54	0.99	0.86	0.71	1.14	0.65	0.39

Application: Uncertainties of luminosity functions at LHC



Note that one component of the uncertainty in predicting the Higgs production cross section at LHC is an uncertainty of $\sim 8\,\%$ due to PDF uncertainty.

Outlook

- Parton distributions of the proton are increasingly well measured.
- Useful tools are in place to estimate the uncertainty of PDFs and to propagate those uncertainties to physical predictions.
- The Les Houches Accord interface makes it easy to handle the large number of PDF solutions that are needed to characterize uncertainties. (hep-ph/0204316)
- Work on refining the knowledge of the "Tolerance Parameter" T is underway
  - Collins & Pumplin [hep-ph/0105207]
  - Statistical bootstrap methods
- Improvements in the treatment of heavy quark effects are in progress.
- Fermilab run II data and HERA II data will provide the next major experimental steps forward.

Parton Distribution Functions are a major avenue toward understanding the fundamental nonperturbative physics of the proton. They are also a crucial prerequisite for precision Standard Model studies and New Physics searches at hadron colliders and experiments with hadron targets.