Parton Distribution Functions and their Uncertainties

Jon Pumplin – APS Meeting Philadelphia 4/8/03

High energy hadrons interact through their quark and gluon constituents, which Feynman collectively called partons. Hadronic interactions become weak at short distances because of the asymptotic freedom property of Quantum Chromodynamics, allowing perturbation theory to be applied to a rich variety of experiments.

The nonperturbative nature of the proton is characterized by Parton Distribution Functions of momentum scale Q and light-cone momentum fraction x. Evolution in Q is determined perturbatively by QCD renormalization group equations, so the non-perturbative physics is specified by functions of x at a fixed Q_0. Those functions are to be measured and applied.

- Introduction to PDFs
- Including correlated experimental errors
- Estimating uncertainties
  - Eigenvector PDF sets
  - Lagrange multipliers
  - reweighting experiments
  - Bootstrap methods
- Applications

Collaborators

J.P., D. Stump, W.K. Tung, J. Huston, H. Lai, P. Nadolsky, F. Olness, S. Kuhlmann, J. Owens; J. Collins

Selected References

- INCLUSIVE JET PRODUCTION, PARTON DISTRIBUTIONS, AND THE SEARCH FOR NEW PHYSICS, D. Stump, J. Huston, J. Pumplin, Wu-Ki Tung, H.L. Lai, S. Kuhlmann, J.F. Owens (hep-ph/0303013)

- NEW GENERATION OF PARTON DISTRIBUTIONS WITH UNCERTAINTIES FROM GLOBAL QCD ANALYSIS, J. Pumplin, D. Stump, J. Huston, H.L. Lai, P. Nadolsky, W.K. Tung JHEP 0207:012,2002 (hep-ph/0201195)

- TESTS OF GOODNESS OF FIT TO MULTIPLE DATA SETS. J. Collins, J. Pumplin (hep-ph/0105207)

- UNCERTAINTIES OF PREDICTIONS FROM PARTON DISTRIBUTION FUNCTIONS 1: THE LAGRANGE MULTIPLIER METHOD. D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H.L. Lai, W.K. Tung, Phys.Rev.D65:014012,2002 (hep-ph/0101051)

- UNCERTAINTIES OF PREDICTIONS FROM PARTON DISTRIBUTION FUNCTIONS 2: THE HESSIAN METHOD. J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.L. Lai, W.K. Tung Phys.Rev.D65:014013,2002 (hep-ph/0101032)

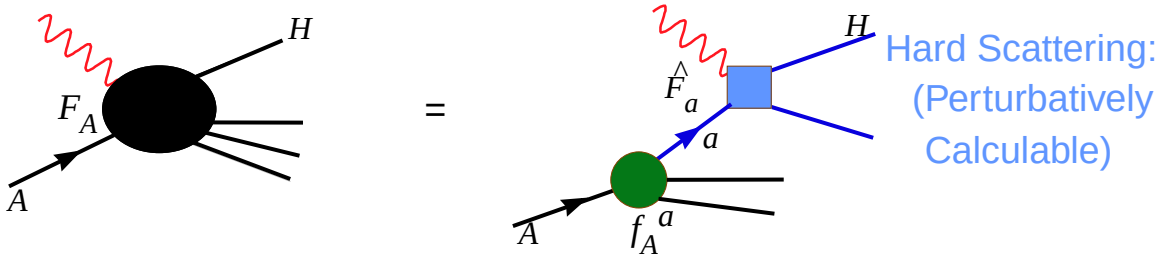
- MULTIVARIATE FITTING AND THE ERROR MATRIX IN GLOBAL ANALYSIS OF DATA. J. Pumplin, D.R. Stump, W.K. Tung Phys.Rev.D65:014011,2002 (hep-ph/0008191)

Global QCD analysis

- Extract universal non-perturbative (large distance scale) features of proton or nucleus from a large variety of experiments with perturbatively calculable (short distance) hard scattering using
  - Factorization
  - Asymptotic Freedom
    (Hard scattering is perturbative)
  - Renormalization Group Evolution in scale Q
    (PDFs characterize by functions of x at Q_0)
- Test consistency of QCD – globally and with individual experiments
- Make results available in convenient form for applications
- Explore the possible range of uncertainties

Factorization Theorem

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



Experimental Input

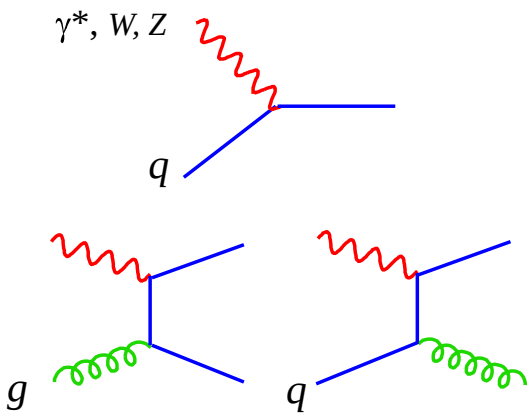
Parton Distributions:
Nonperturbative parametrization at Q_0
DGLAP Evolution to Q

Experimental Input

DIS

e N
μ N

ν N
ν̄ N

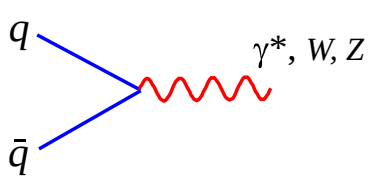


SLAC
BCDMS
NMC, E665
H1, ZEUS

CDHS, CHARM
CCFR
CHORUS

DY

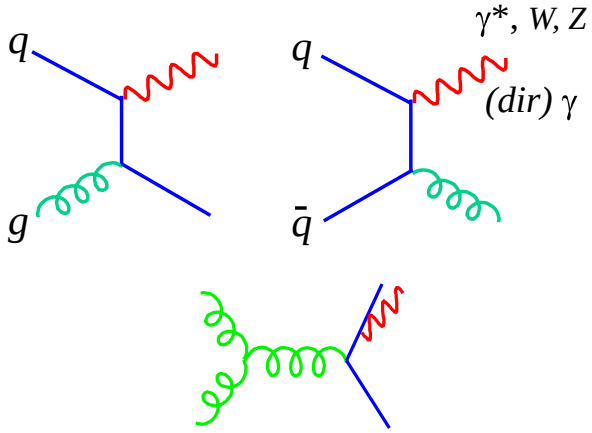
p N
π N
k N
p̄ N



E605, E772
NA51
E866

Dir.Ph.

p N
π N
k N
p̄ N



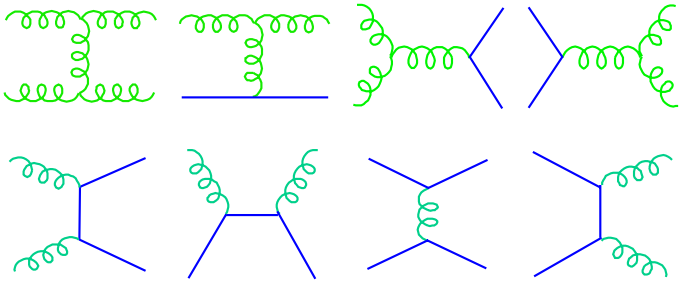
CDF, D0

WA70, UA6
E706

CDF, D0

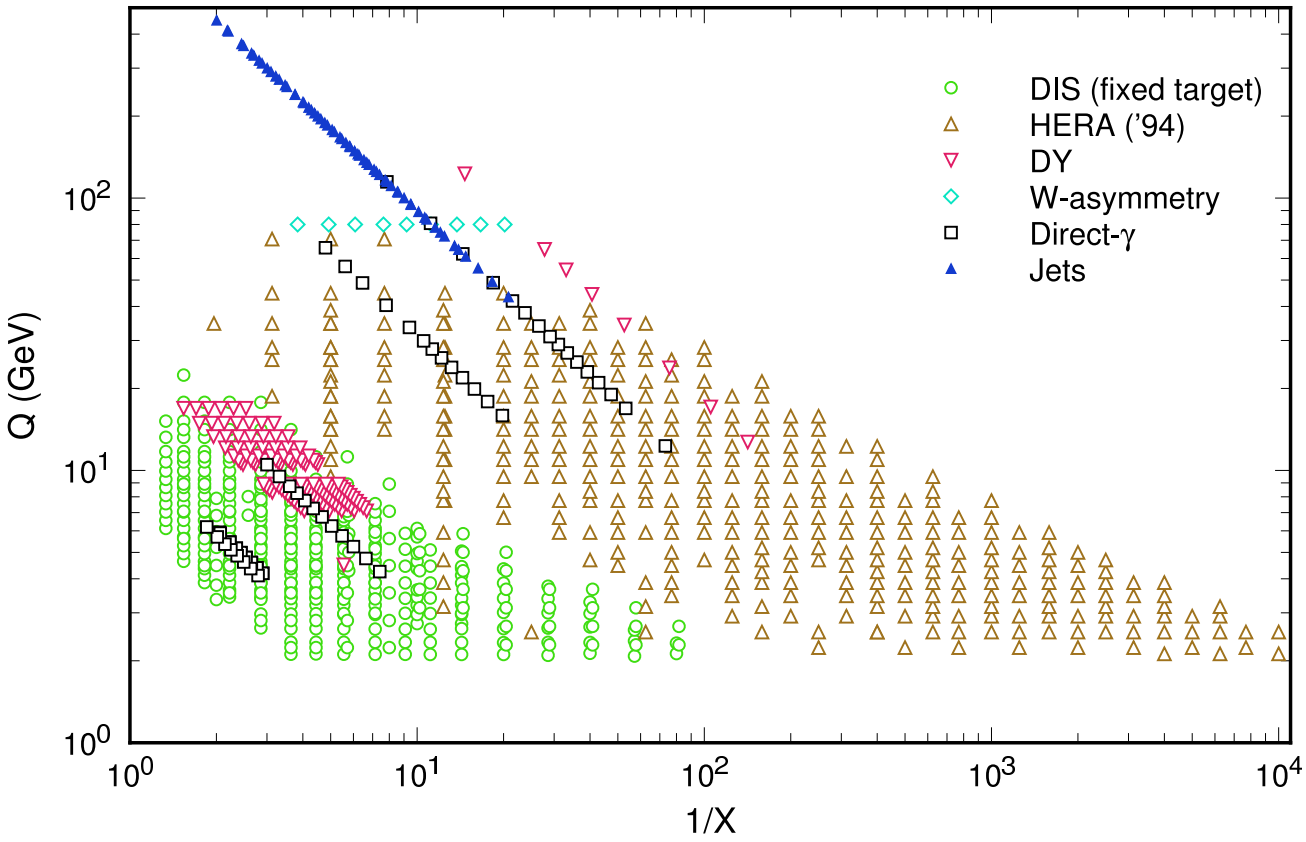
Jet Inc.

p̄ p



CDF, D0

Kinematic region covered by data



Data with a wide range of scales are tied together by the DGLAP renormalization group evolution equation.

Consistency or inconsistency between the different processes can be observed only by applying QCD to tie them together in a global fit.

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

CTEQ6 Global analysis

Input from Experiment:

- ~ 2000 data points with $Q > 2\,\text{GeV}$ 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:

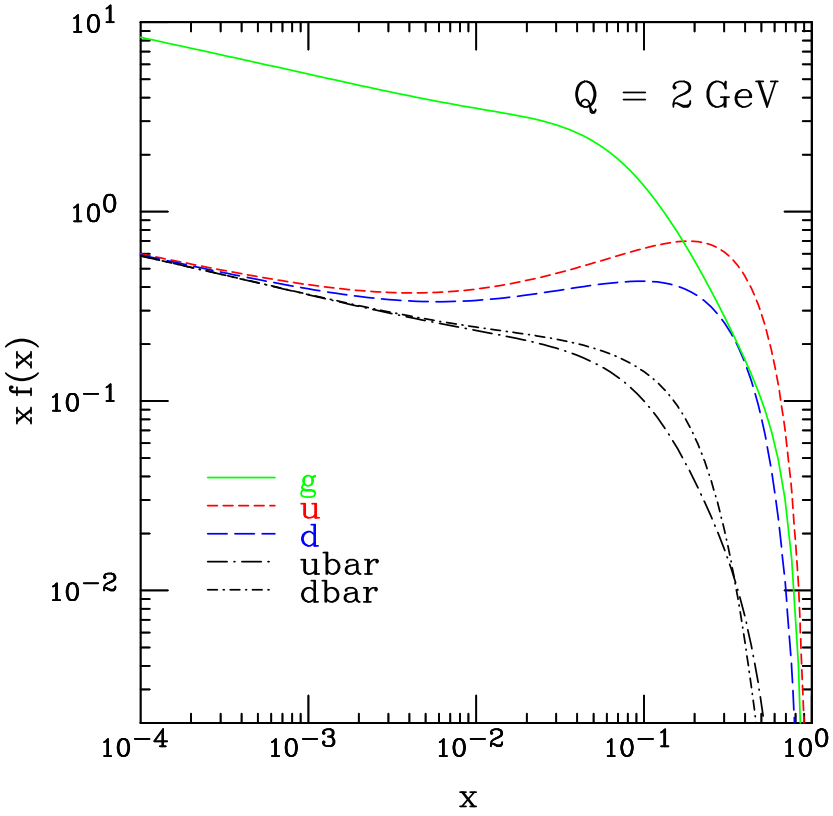
- NLO QCD evolution and hard scattering
- Parametrize at Q_0: $A_0\, x^{A_1}\,(1-x)^{A_2}\, e^{A_3 x}(1+A_4 x)^{A_5}$
- $s = \bar{s} = 0.4\,(\bar{u}+\bar{d})/2$ at Q_0; no intrinsic b or c

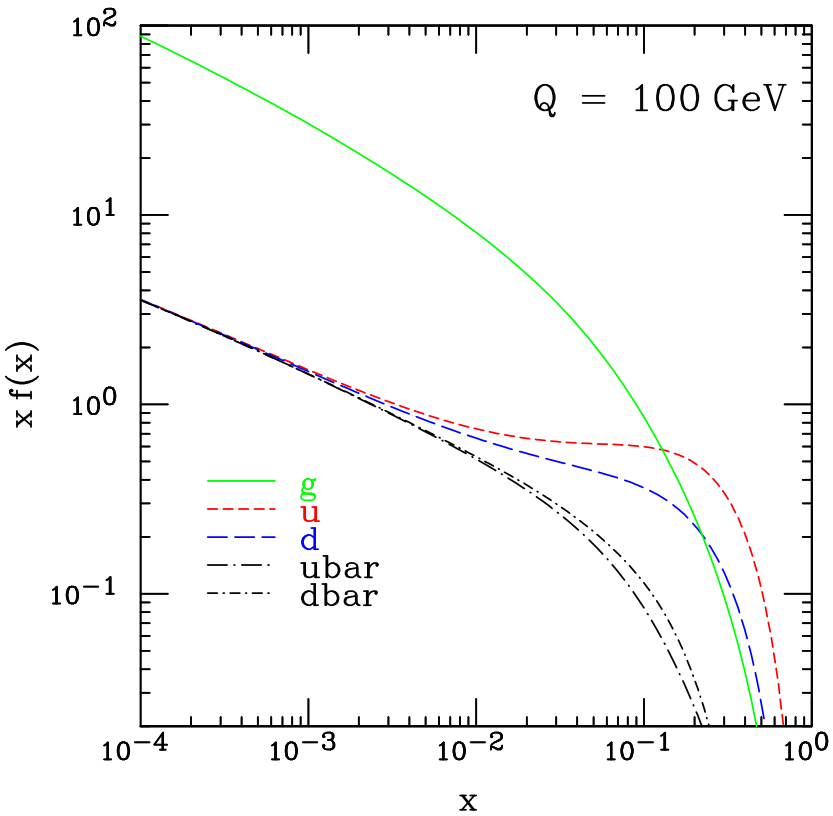
Construct effective $\chi^2_{\text{global}} = \sum_{\text{expts}} \chi^2_n$:

- 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).

Parton distributions at $Q = 2$ and 100 GeV





- Valence quarks dominate for $x \longrightarrow 1$
- Gluon dominates for $x \longrightarrow 0$, especially at large Q

Comment on Parametrization

For $d_{\rm val}$, $u_{\rm val}$, or g, we use

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

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.)

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.

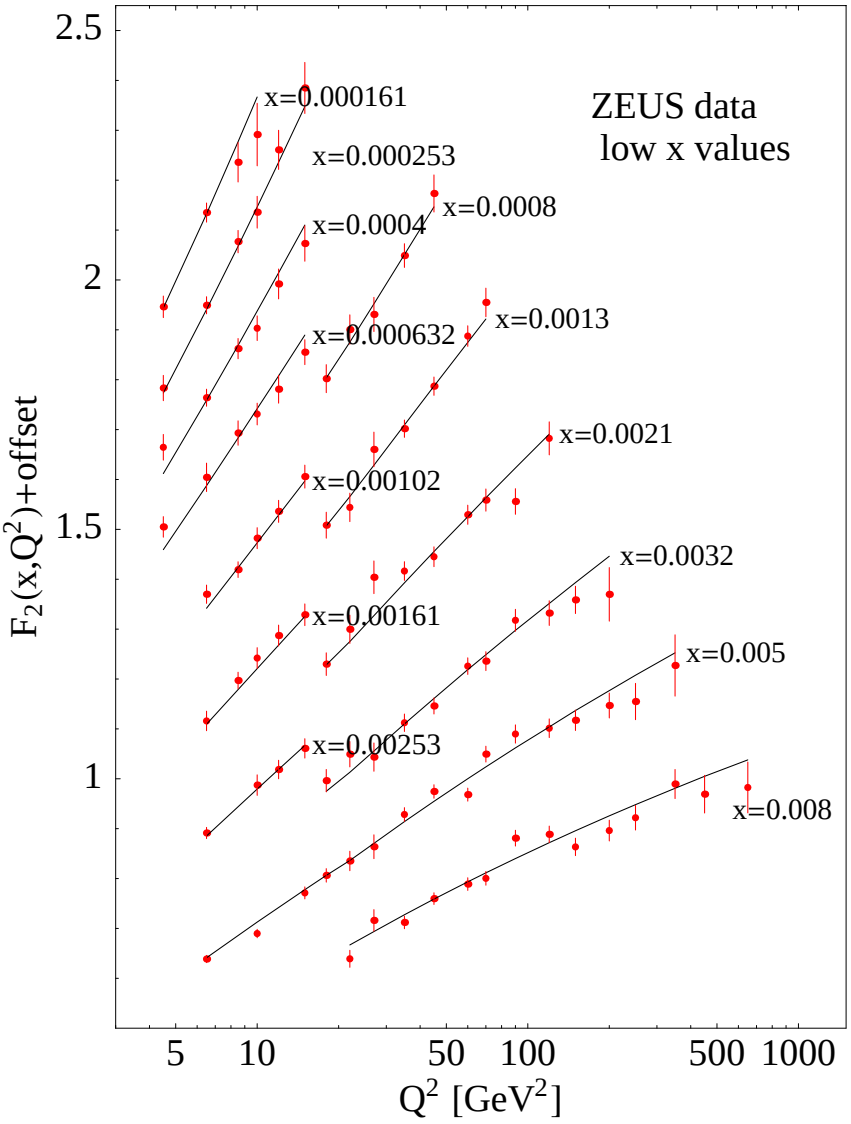
We then minimize the remaining $\chi^2(a)$ with respect to the model parameters $\{a_\lambda\}$.

- $\{a_\lambda\}$ 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 of CTEQ6M fit to data sets with correlated systematic errors

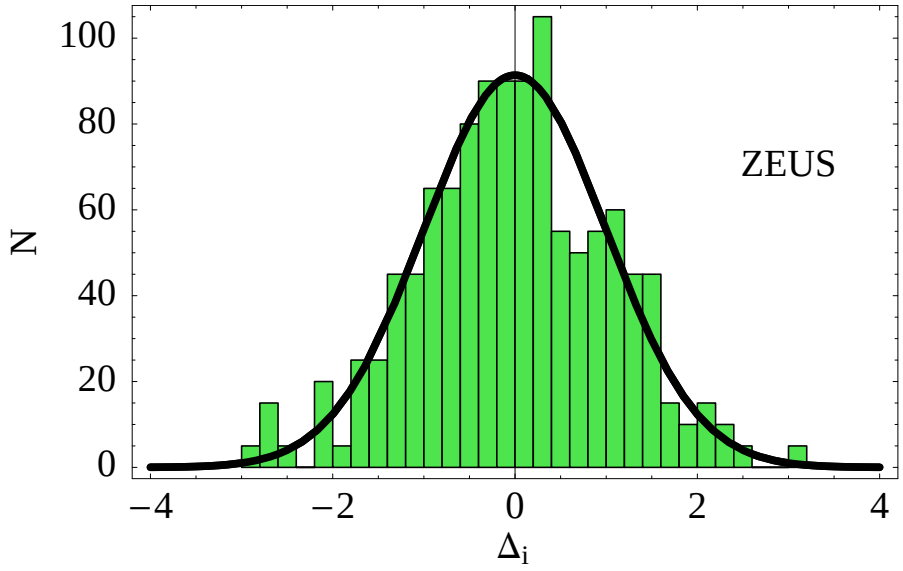
data set	N_e	χ^2_e	χ^2_e/N_e
BCDMS p	339	377.6	1.114
BCDMS d	251	279.7	1.114
H1a	104	98.59	0.948
H1b	126	129.1	1.024
ZEUS	229	262.6	1.147
NMC F2p	201	304.9	1.517
NMC F2d/p	123	111.8	0.909
DØ jet	90	69.0	0.766
CDF jet	33	48.57	1.472

CTEQ6M fit to ZEUS data at low x

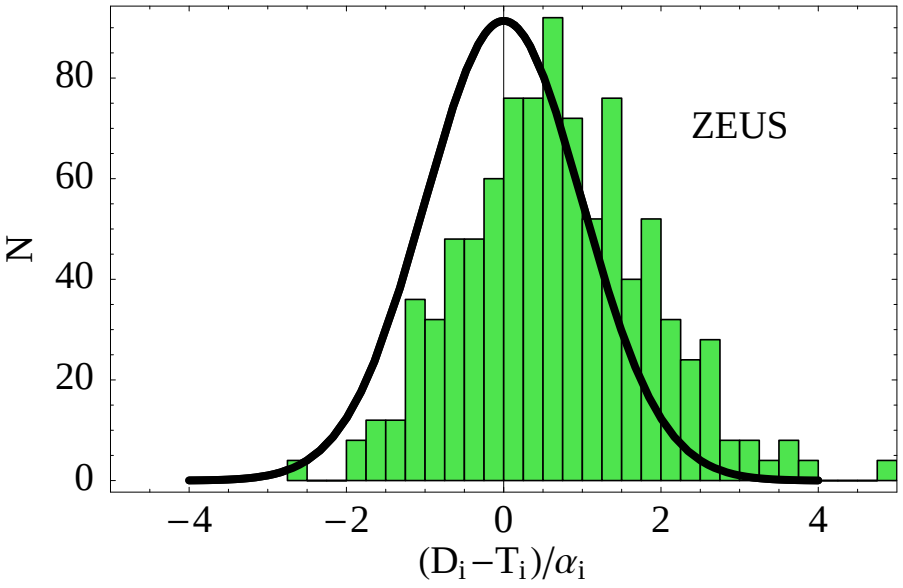


The data points include the estimated corrections for systematic errors. That is to say, the central values plotted have been shifted by an amount that is consistent with the estimated systematic errors, where the systematic error parameters are determined using other experiments via the global fit.

The error bars are statistical errors only.

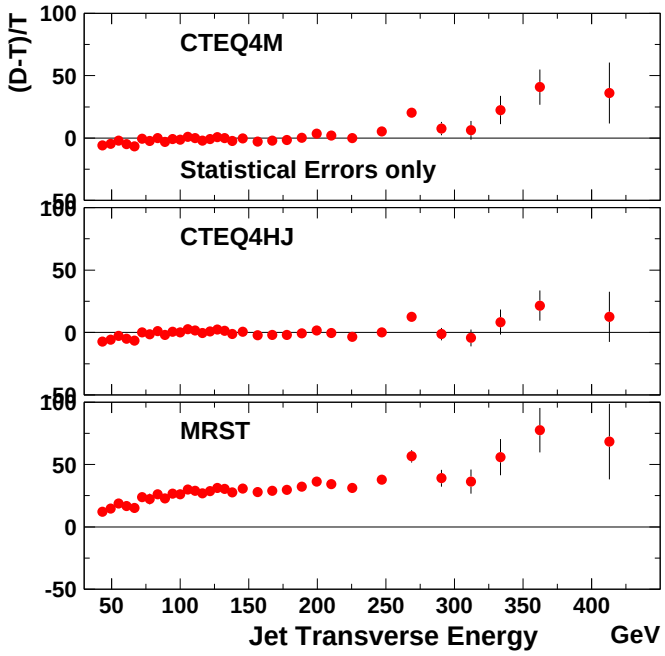


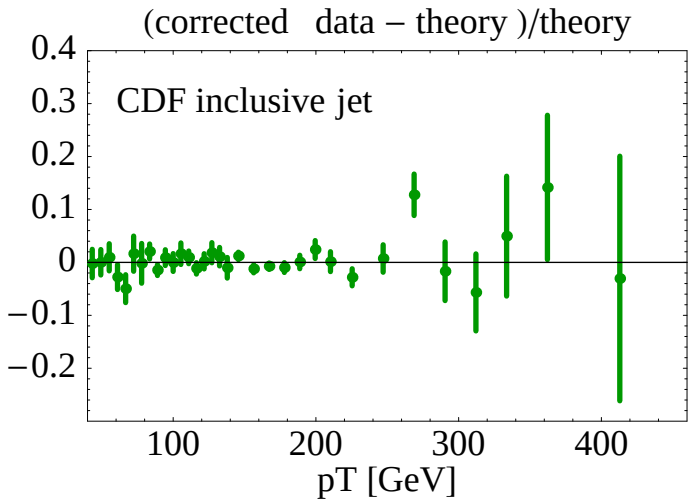
(a) Histogram of residuals for the ZEUS data. The curve is a Gaussian of width 1.



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

CDF inclusive jet cross section





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.

Values of the fitted systematic error parameters for CDF Inclusive jet cross section:

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

All parameters are $\lesssim 1$ as they should be.

Sources of uncertainty:

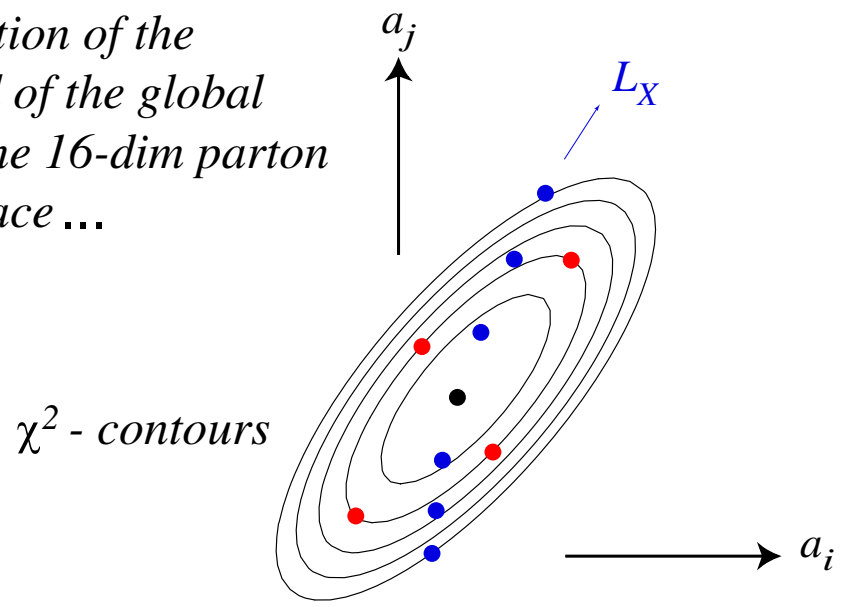
1. Experimental errors included in χ^2
2. Unknown experimental errors
3. Parametrization dependence
4. Higher-order corrections & Large Logarithms
5. Power Law corrections ("higher twist")

Fundamental difficulties:

1. Good experiments run until systematic errors dominate: the magnitude of remaining systematic errors involves guesswork.
2. Systematic errors of the theory and their correlations are even harder to guess.
3. Quasi–ill-posed problem: determine continuous functions from discrete data set
4. Some combinations of variables are unconstrained, e.g., $s - \bar{s}$ before NuTeV data.

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

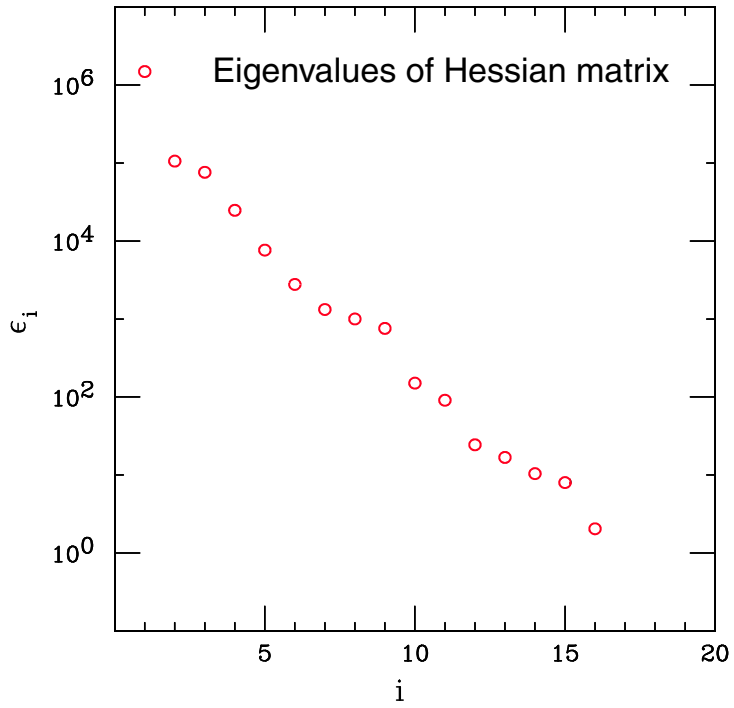
Classical error formulae

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

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

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 in the minimization 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/

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 want to 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^2_i vs. $\chi^2_{\rm tot} - \chi^2_i$, where χ^2_i is one of the experiments, or all data on nuclei, or all data at low Q^2,...

or

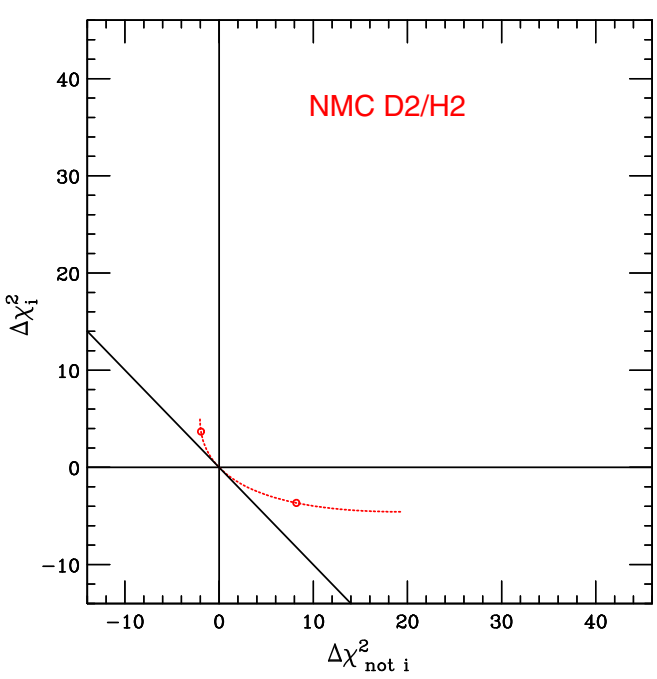
- Plot both as function of Lagrange multiplier u where $(1-u)\chi^2_i + (1+u)(\chi^2_{\rm tot} - \chi^2_i)$ is the quantity minimized.

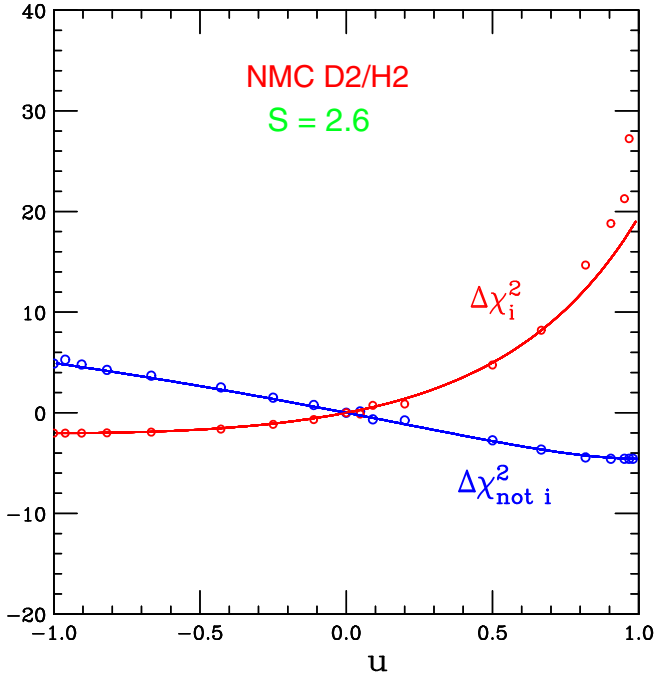
Can obtain quantitative results by fitting to a model with a single common parameter p:

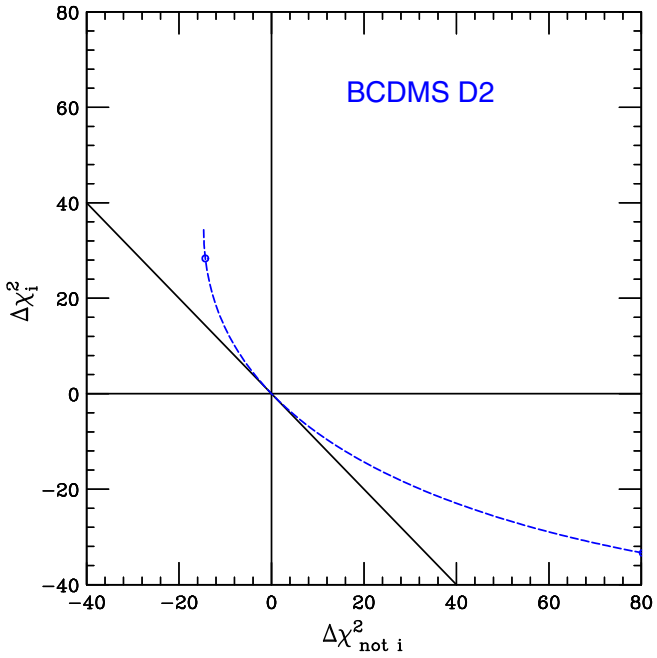
$$\chi^2_i = A + \left(\frac{p}{\sin\theta}\right)^2 \quad \Rightarrow \quad p = 0 \pm \sin\theta$$

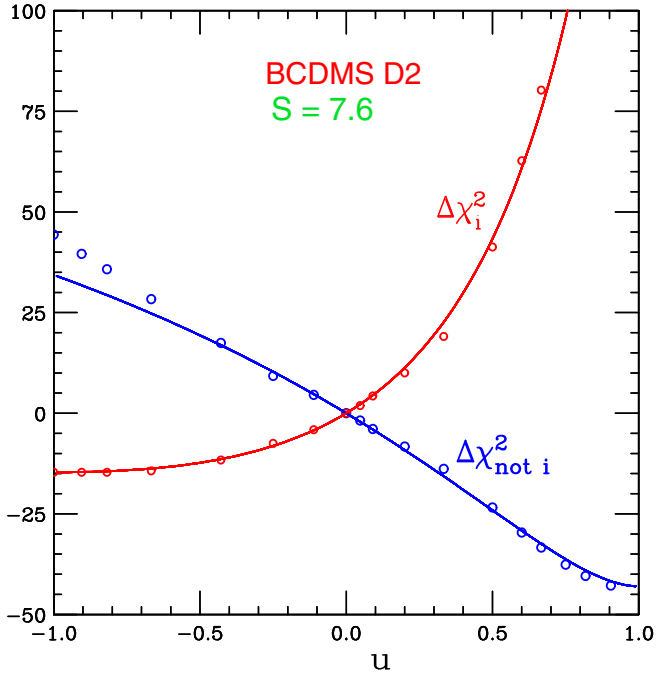
$$\chi^2_{{\rm not}\ i} = B + \left(\frac{p-S}{\cos\theta}\right)^2 \quad \Rightarrow \quad p = S \pm \cos\theta$$

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





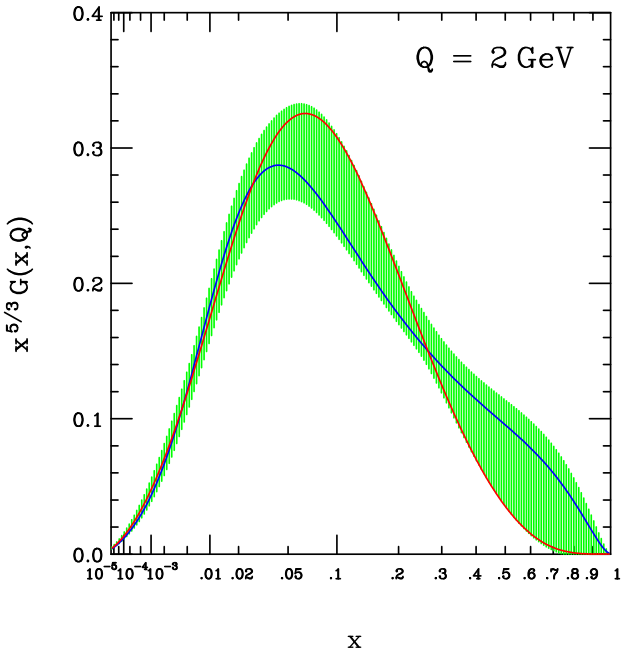




Fits to 8 of the experiments in the CTEQ5 analysis

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

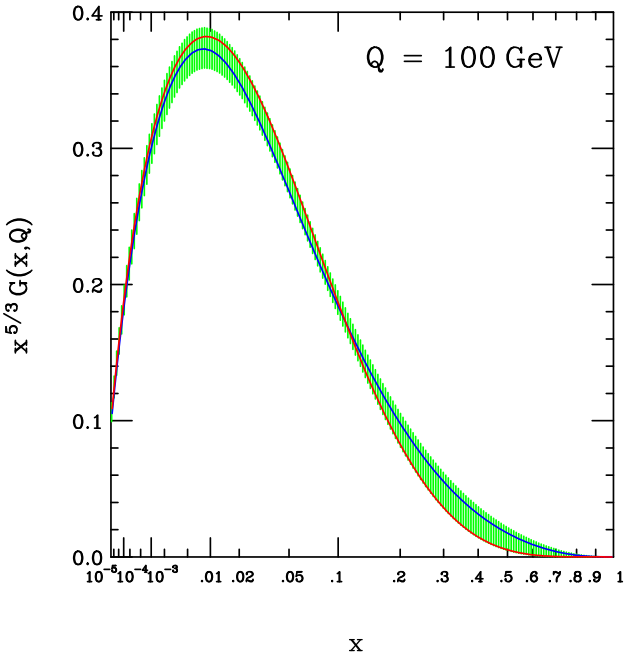
Uncertainty of Gluon distribution



Red: Weight 50 for CDF Jet Blue: Weight 50 for DØ Jet

Consistency check: Estimated uncertainty is comparable to the difference between nominally similar experiments.

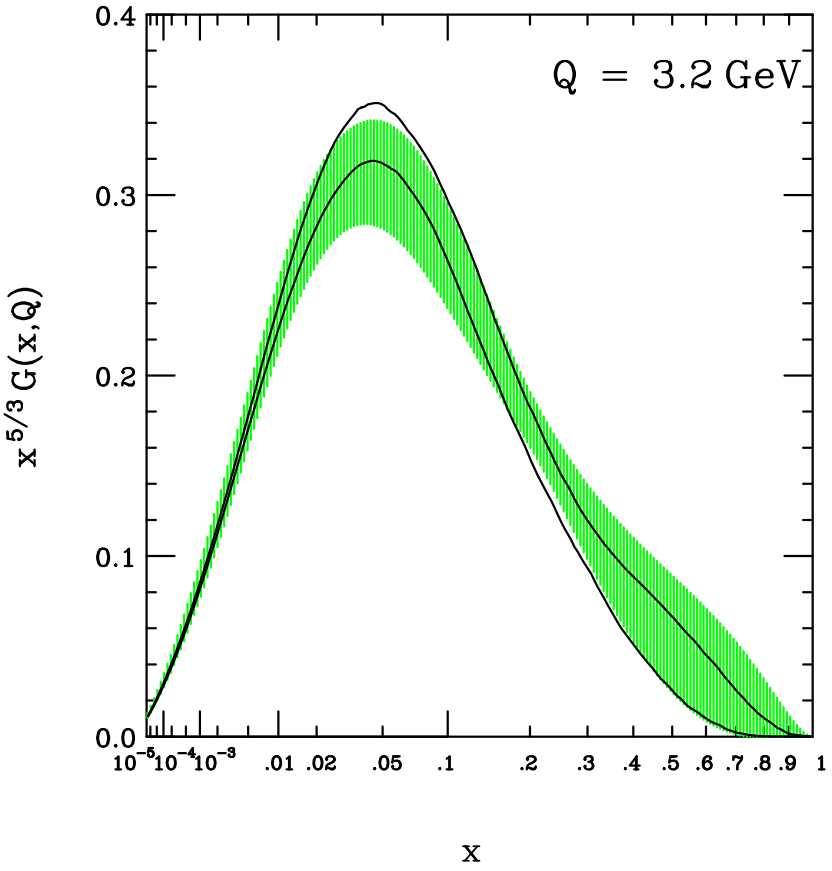
Area under curve is proportional to momentum fraction carried by gluon – strongly constrained by DIS data. Hence the envelope itself is not an allowed solution.



Convergent Evolution: Uncertainty smaller at large Q

Statistical Bootstrap method

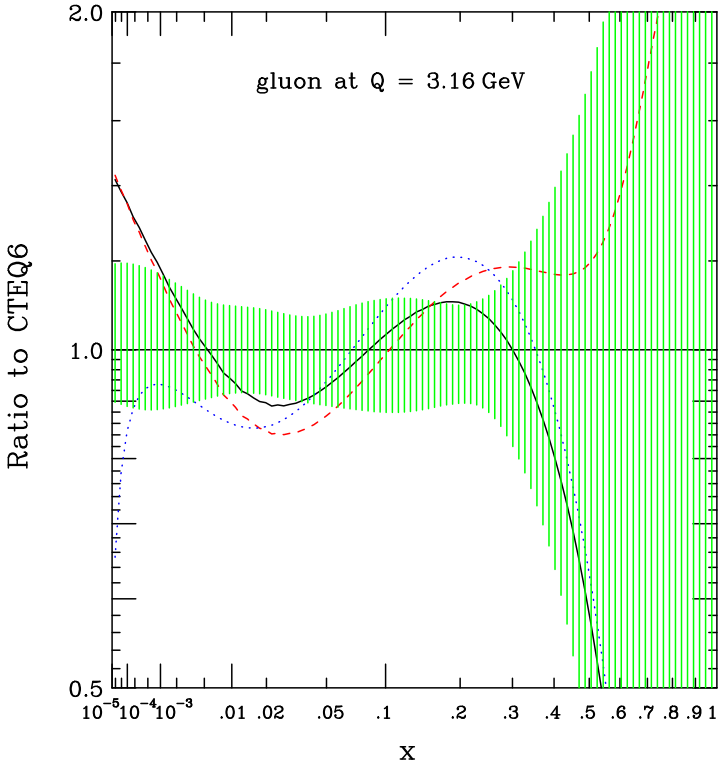
Generate random weights for each of the 16 experiments in global fit by $\frac{dP}{dW_i} = e^{-W_i}$. Find best fit for each set of weights. Repeat 200 times and take the central 90 % at each x as the measure of uncertainty range. Shows a sizable uncertainty with no *ad hoc* assumption such as $\Delta\chi^2 = 100$.



Traditional statistical bootstrap uses integer weights $0 - 16$ defined by random selection. This continuum method is similar but avoids zero weights. Traditional method:

- Efron and Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall 1993.
- M. Chernick, *Bootstrap Methods: A Practitioner's Guide*, John Wiley & Sons 1999.

Fractional uncertainty of gluon



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

Curves show

CTEQ5M1 (solid)

CTEQ5HJ (dashed)

MRST2001 (dotted)

The differences between these are comparable to the estimated uncertainty.

Uncertainties of quark distributions (not shown) are smaller than the gluon uncertainty, because extensive DIS measurements are sensitive to the square of the quark charge in leading order.

Summary of Uncertainty Methods

Consistent estimates of the uncertainty ranges are found using several different methods:

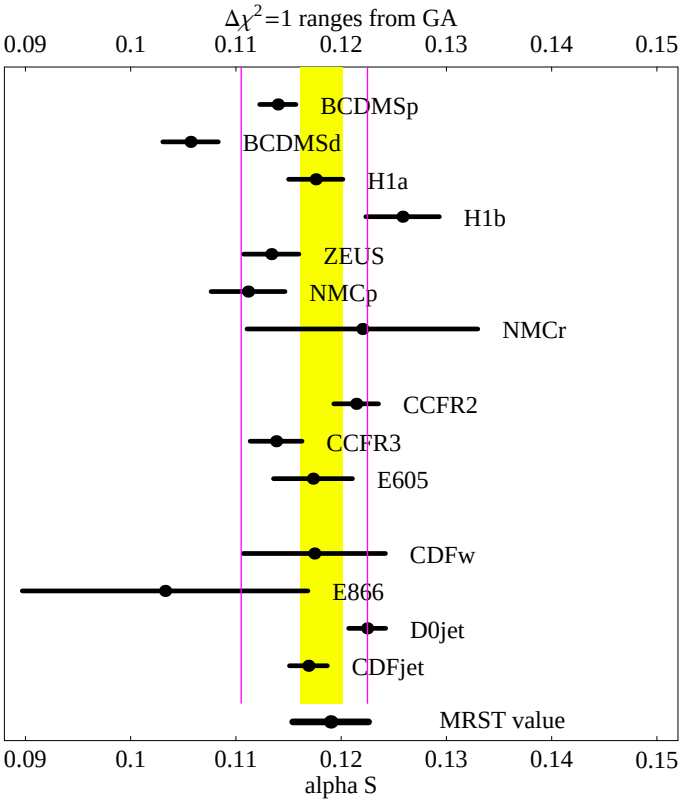
- "Hessian Method" – eigenvectors of the error matrix
- "Lagrange Multiplier Method" – variation of χ^2
- systematic reweighting of experiments
- random reweighting (statistical bootstrap)

Application: Measurement of α_s

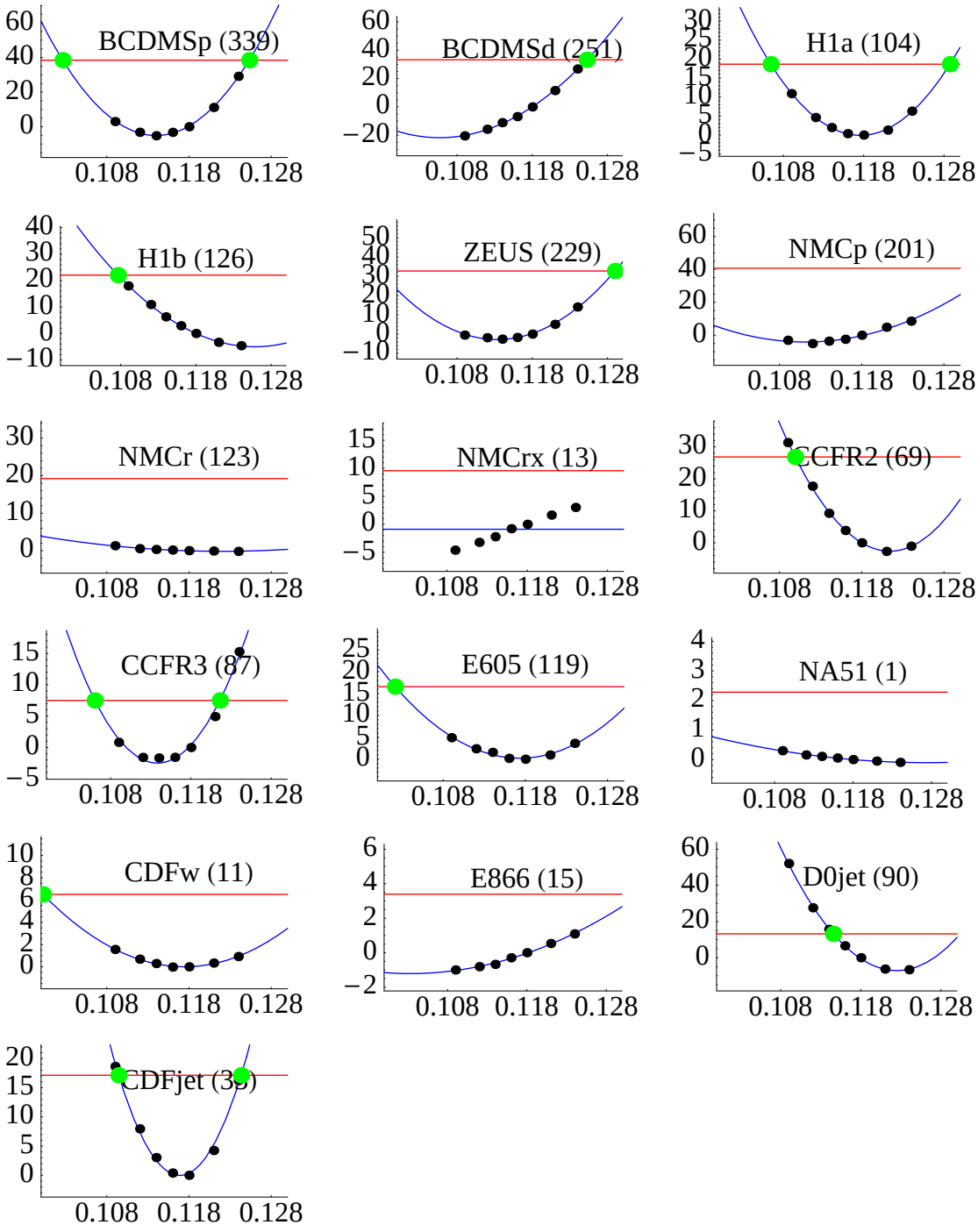
The CTEQ6 analysis gives

$$\alpha_S(M_Z) = 0.1165 \pm 0.0065$$

This is nicely consistent with the World Average, but not precise enough to improve on it.



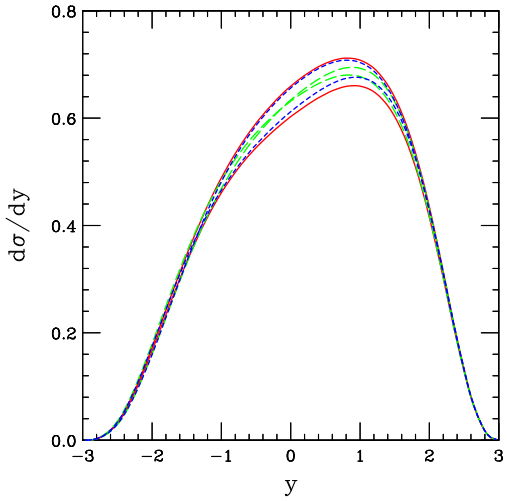
χ^2 versus $\alpha_S(M_Z)$ for individual data sets in CTEQ6



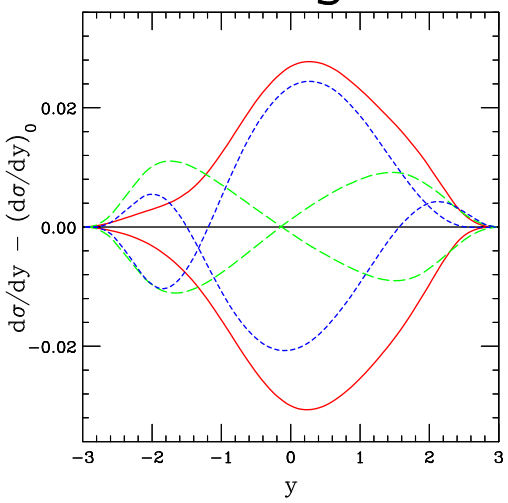
Application: W rapidity distribution

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:

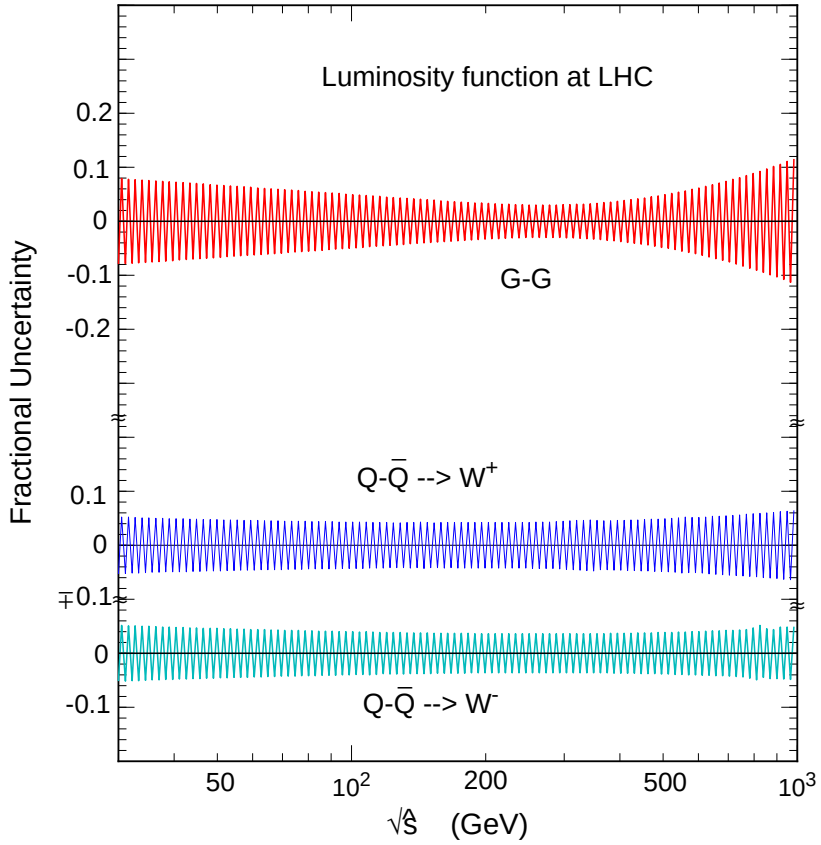


Same curves after subtracting central values. . .



These extremes can be important for the measurement of W mass.

Application: Uncertainties of luminosity functions at LHC



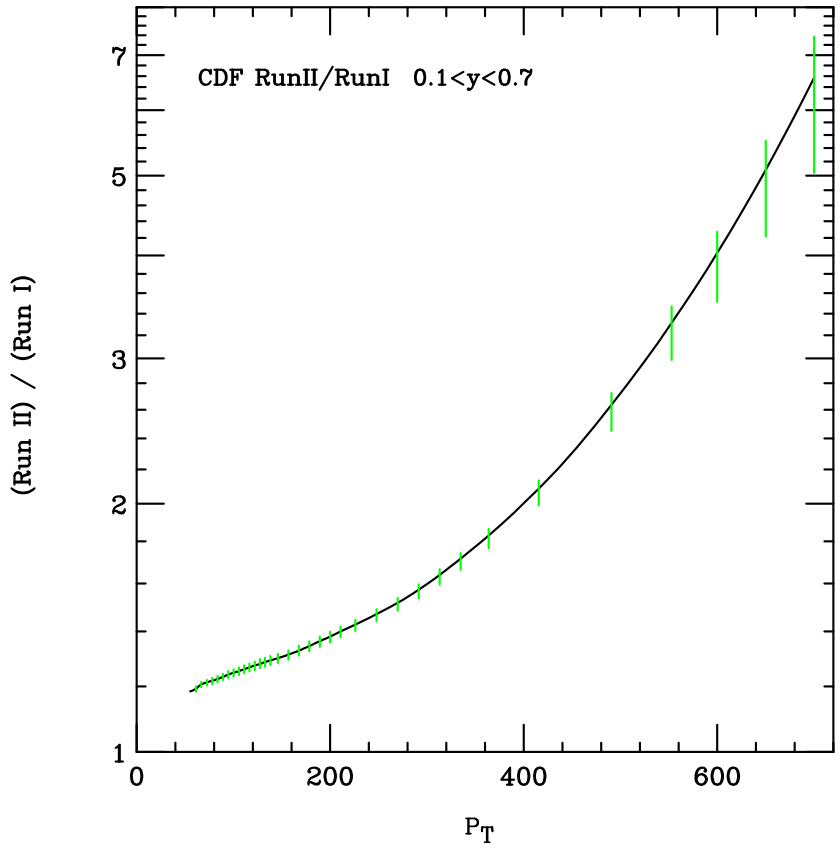
- One component of the uncertainty in predicting the Higgs production cross section at LHC is an uncertainty of 8% due to PDF uncertainty.

Application: Inclusive jet ratio

Inclusive jet energy dependence

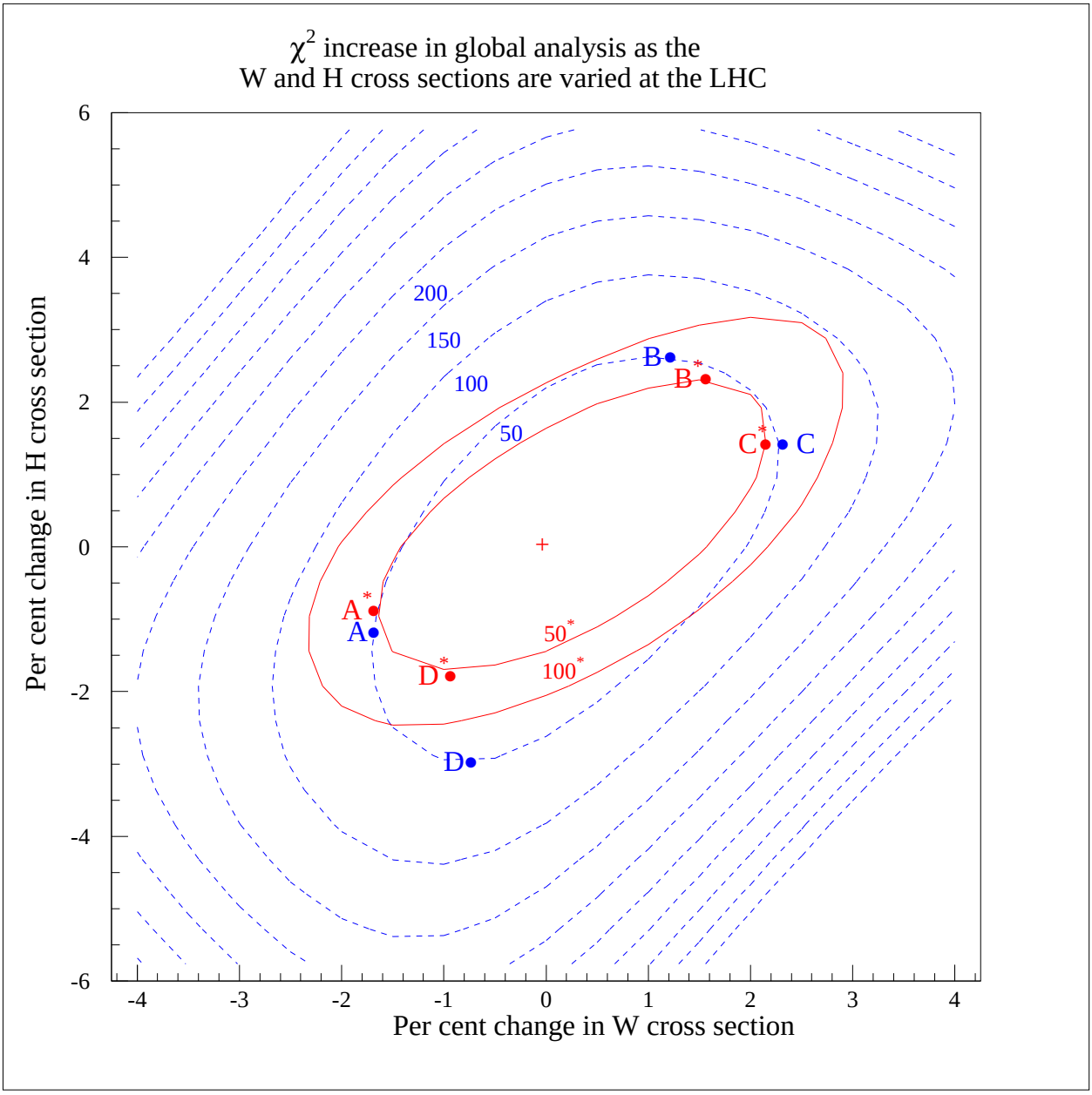
$$\frac{\frac{d\sigma}{dP_T}(1.96\,\text{TeV})}{\frac{d\sigma}{dP_T}(1.80\,\text{TeV})}$$

between Tevatron Run I and Run II offers a sensitive test of QCD and a probe for quark substructure, because many systematic errors cancel. Right now it is an important check on the experimental jet "energy scale" calibration.



Prediction and uncertainty range from CTEQ6.1

Correlated Predictions



Contours in $\Delta\chi^2$ show the correlation in PDF uncertainty for predictions of cross sections for W and Higgs production at Fermilab (MRST)

Outlook

- Parton Distribution Functions are a necessary infrastructure for precision Standard Model studies and New Physics searches at hadron colliders and experiments using hadron targets.
- PDFs 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. There is adequate agreement between various methods for estimating the uncertainty.
- 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, http://vircol.fnal.gov)
- PDFs summarize fundamental nonperturbative physics of the proton – a challenge to be computed!
- Improvements in the treatment of heavy quark effects are in progress; will allow improved flavor differentiation
- HERA and Fermilab run II data will provide the next major experimental steps forward, followed by LHC.