Global Analysis of Parton Distributions
Discussion Topics

Santa Barbara Workshop on Collider Physics

Issue: Uncertainties on PDFs

- The statistical principles and methods for uncertainty analyses are well established:
  Likelihood, χ^2, … etc.---all textbook stuff, nothing extraordinary in principle.

- The devil is, not mainly in the details, rather:
  - Unknown theoretical uncertainties
  - Unknown experimental uncertainties
- What's needed?

Reality #1 : compatibility of experiments

	H1	BCDMS	E665	ZEUS	NMC	LEP
H1-MRST set	-	67%	21%	0.5%	<0.1%	31%
BCDMS-MRST set	85%	-	23%	1.5%	<0.1%	0.5%
E665-MRST set	30%	82%	-	1.6%	1.0%	99%
ZEUS-MRST set	22%	<0.1%	5.0%	-	<0.1%	24%
NMC-MRST set	<0.1%	28%	1.5%	<0.1%	-	3.2%

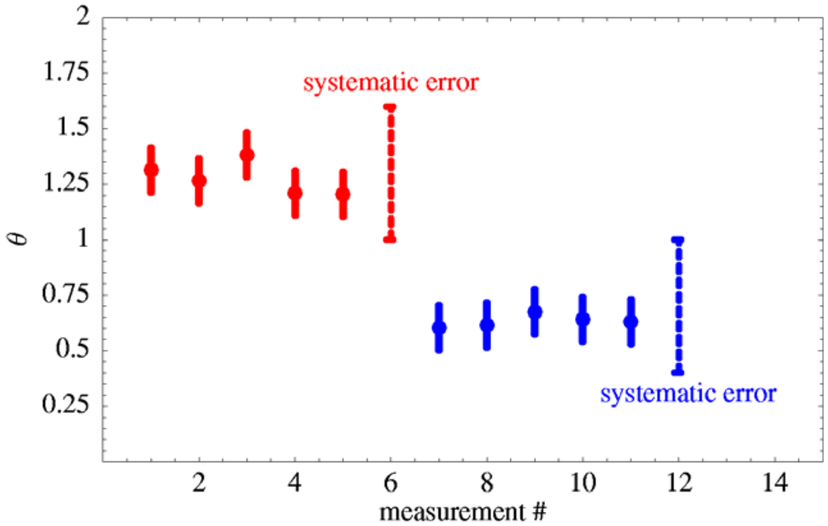
Table 2: The confidence level of each experiment given the different sets. The name of the set is composed of all included experiments and the PDF parameterization choice.

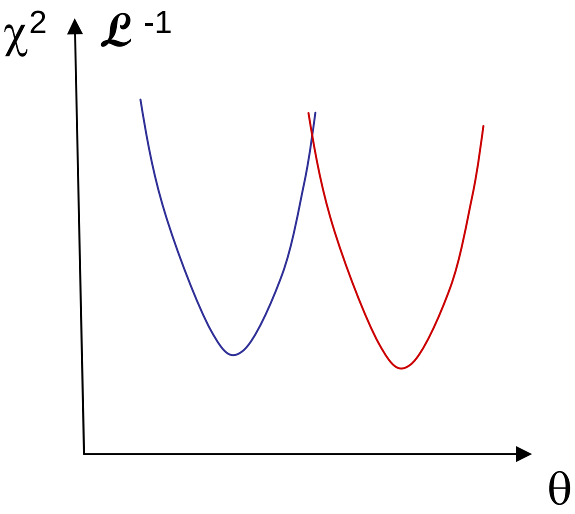
(Giele etal, 2001)

Basic dilemma:

What is the *real uncertainty* on a measured quantity due to incompatible experimental results?

Imagine that two experimental groups have measured a quantity θ, with the results shown.





What is the value of θ? What do confidence levels mean?

(This is common occurrence in the real world.)

Are all experimental errors understood? Should the errors be taken at face value?

Case study: consequences on α_s analysis in the GKK approach (likelihood)

	α_S-distribution	ΔL^2-distribution			
set	CL interval	$\min(\Delta L^2)$	$\langle \Delta L^2 \rangle$	$\frac{1}{2}\sigma^2(\Delta L^2)$	NDP/DOF
ZEUS-MRST	(0.113,(0.114,(0.115),0.116),0.117)	466.1	481.2	42.9	187/23
NMC-MRST	(0.098,(0.102,(0.108),0.112),0.117)	185.4	196.4	10.1	127/23
H1-MRST	(0.108,(0.112,(0.115),0.117),0.119)	166.6	175.9	7.8	188/28
H1+LEP-MRST	(0.114,(0.116,(0.118),0.119),0.121)	167.3	176.0	8.3	189/28
BCDMS-MRST	(0.104,(0.106,(0.108),0.110),0.112)	317.2	328.1	12.1	344/23
BCDMS+LEP-MRST	(0.112,(0.113,(0.116),0.117),0.119)	325.0	335.8	15.7	345/23
E665-MRST	(0.106,(0.112,(0.116),0.127),0.133)	57.9	65.5	4.9	53/23
E665+LEP-MRST	(0.114,(0.117,(0.120),0.123),0.126)	59.1	66.5	6.0	54/23
H1+BCDMS-MRST	(0.109,(0.110,(0.112),0.114),0.115)	510.9	525.8	11.4	532/28
H1+BCDMS+LEP-MRST	(0.110,(0.111,(0.112),0.114),0.115)	511.5	521.8	10.0	533/28
H1+BCDMS+E665-MRST	(0.109,(0.111,(0.112),0.114),0.115)	580.3	596.2	12.3	585/28
H1+BCDMS+E665+LEP-MRST	(0.110,(0.112,(0.113),0.114),0.115)	579.7	592.3	10.4	586/28

Table 3: The relevant properties of the α_S and ΔL^2 distributions for the optimized sets. The confidence level intervals are for a CL of 4.55%, 31.73% and 100%. The bin width used to calculate the confidence level intervals is 0.005 using the 1,000 PDF's

Uncertainties of Physical Predictions: What is the true uncertainty? (GKK)

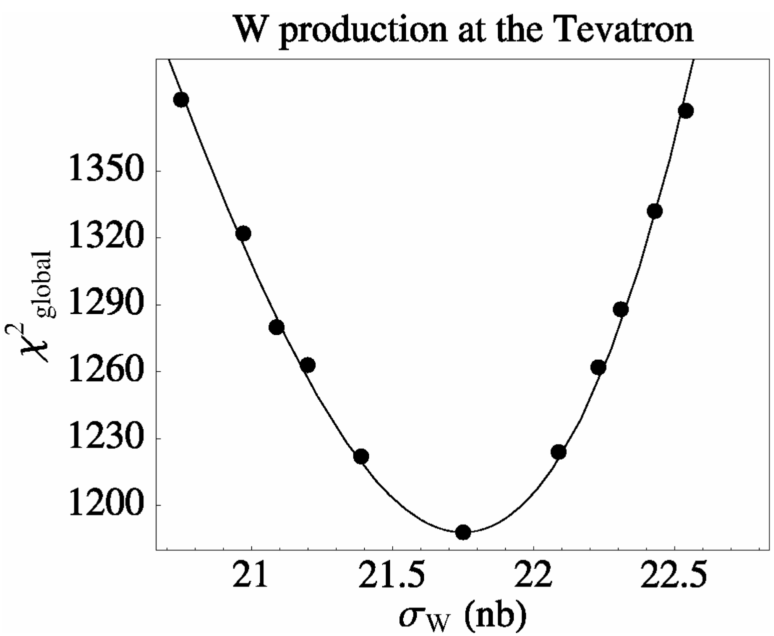
	$\sigma(W)$ (nb)	$\sigma(Z)$ (nb)
D0 1a	2.36+0.15-0.15	0.218+0.016-0.016
D0 1b	2.31+0.11-0.11	0.221+0.011-0.011
MRS99	2.49	0.218
CTEQ5M	2.55	0.222
ZEUS-MRST	2.45+0.06-0.06	0.227+0.007-0.007
NMC-MRST	2.35+0.11-0.09	0.231+0.008-0.011
H1-MRST	2.10+0.18-0.13	0.195+0.013-0.013
H1+LEP-MRST	2.12+0.13-0.16	0.194+0.013-0.013
BCDMS-MRST	2.41+0.12-0.08	0.231+0.011-0.007
BCDMS+LEP-MRST	2.50+0.09-0.11	0.237+0.008-0.011
E665-MRST	2.34+0.09-0.16	0.227+0.005-0.019
E665+LEP-MRST	2.41+0.06-0.18	0.227+0.025-0.015
H1+BCDMS-MRST	2.48+0.09-0.05	0.234+0.004-0.008
H1+BCDMS+LEP-MRST	2.22+0.06-0.12	0.208+0.006-0.007
H1+BCDMS+E665-MRST	2.44+0.07-0.07	0.232+0.007-0.006
H1+BCDMS+E665+LEP-MRST	2.35+0.04-0.03	0.220+0.006-0.004

$\Delta\sigma \sim$ 0.4

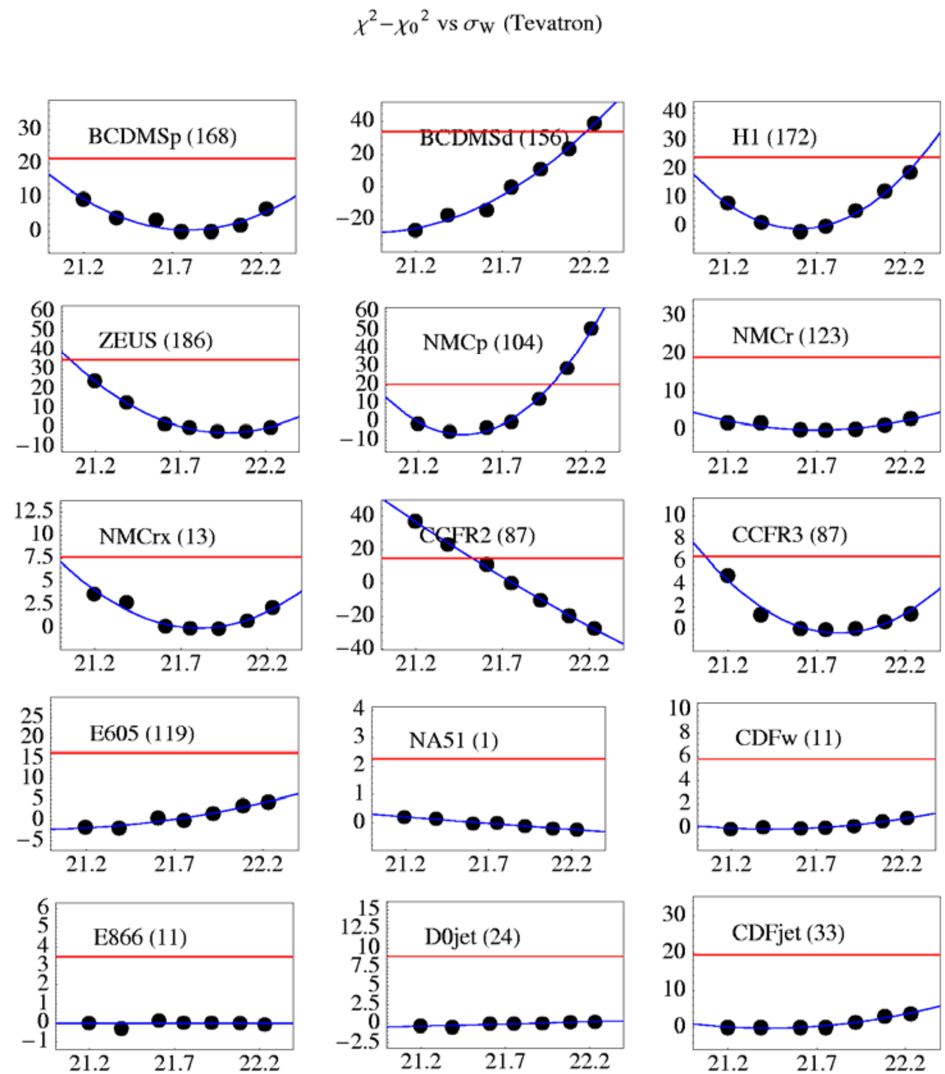
$\Delta\sigma \sim$ 0.1

Case study: CTEQ global analysis of σ_W (χ^2 method)

Estimate the uncertainty on the predicted cross section for $pp_{bar} \rightarrow W+X$ at the Tevatron collider.

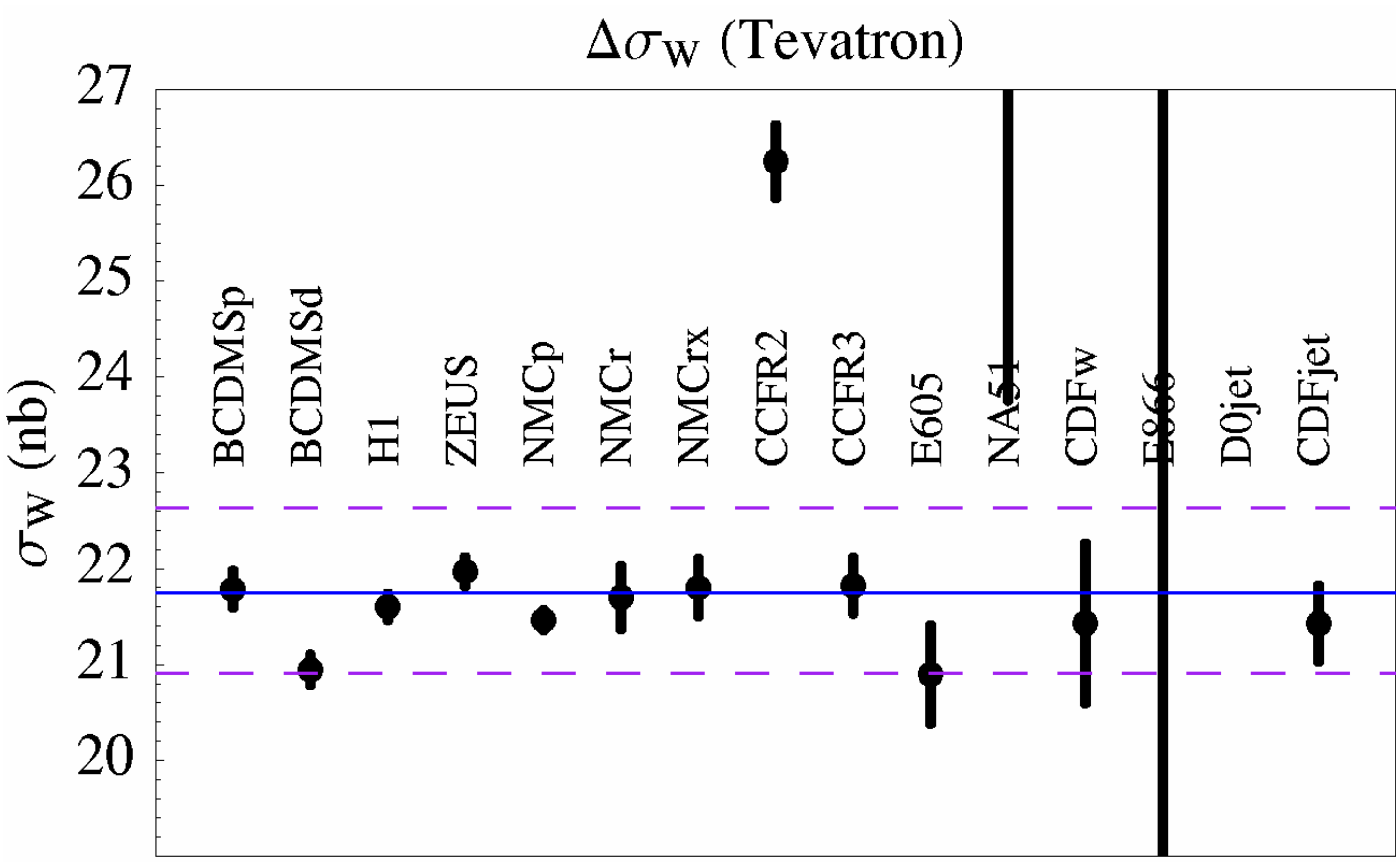


global χ^2

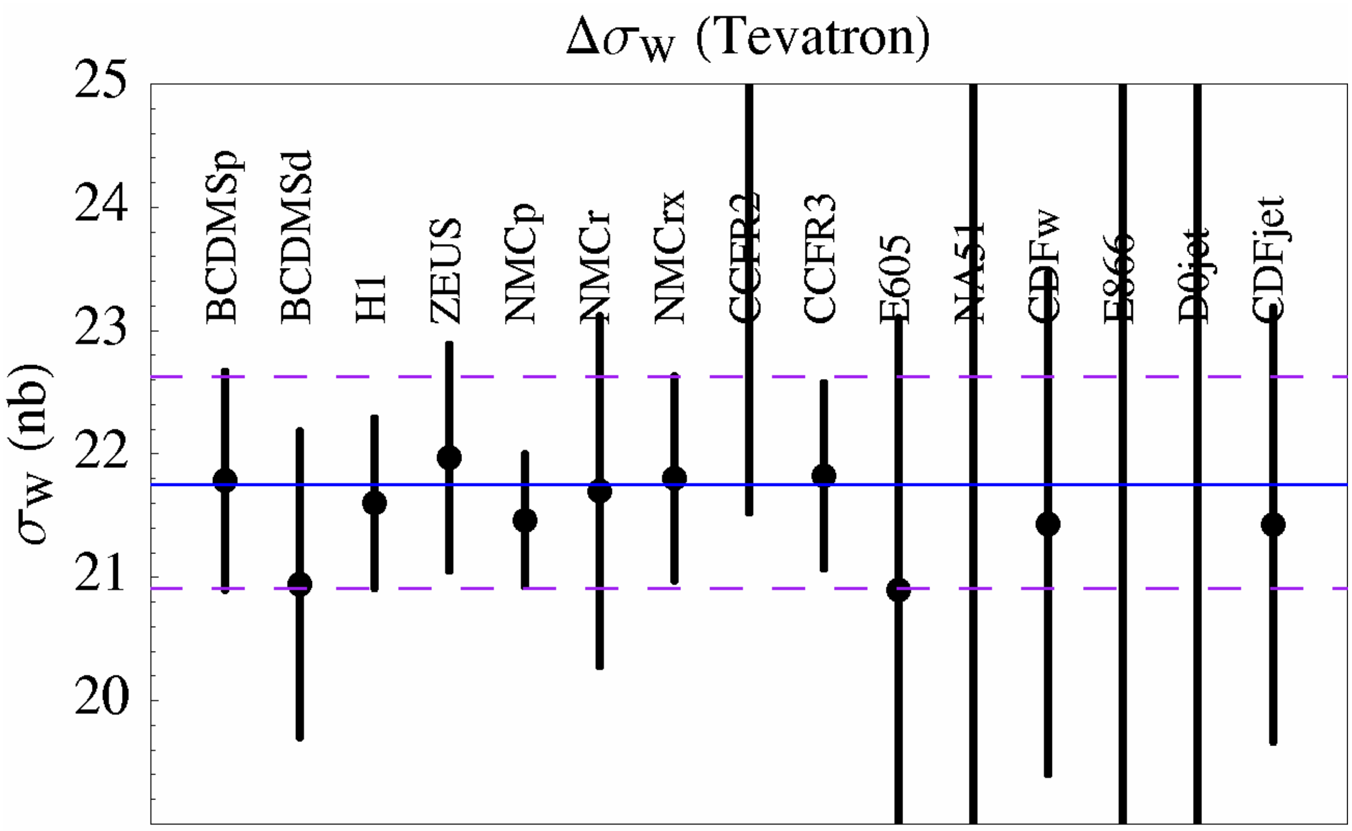


local χ^2's

Each experiment defines a “prediction” and a “range”.
This figure shows the $\Delta\chi^2 = 1$ ranges.

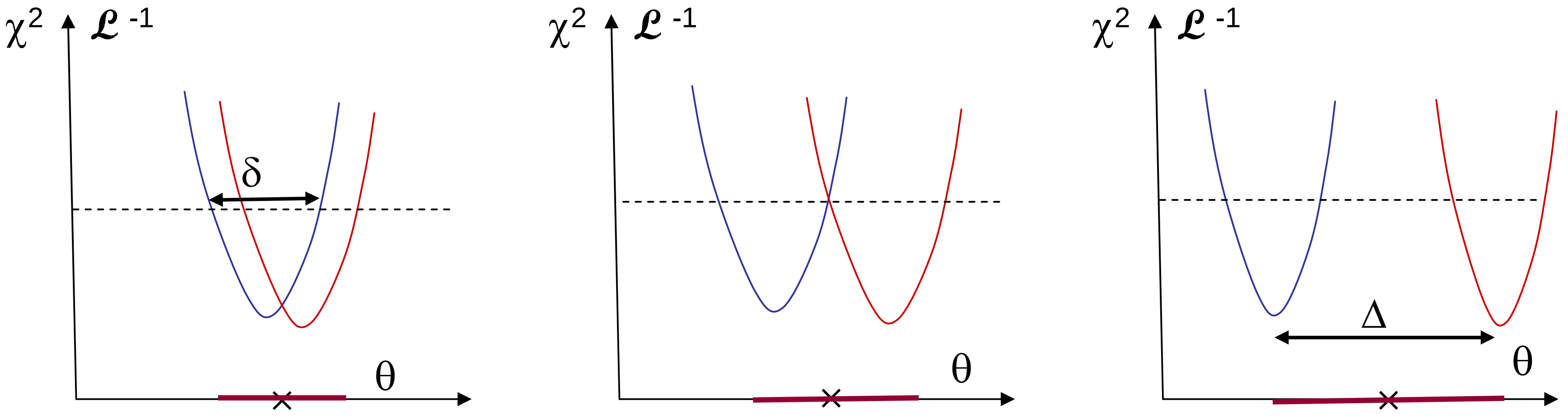


This figure shows broader ranges for each experiment based on the “90% confidence level” (cumulative distribution function of the rescaled χ^2).



"Uncertainty" in 3 scenarios

(either directly measured or indirectly inferred physical quantity θ)



Uncertainty dominated by:

δ $\Delta \otimes \delta$ Δ

- Only case I is textbook safe; but II and III are "real".
- There are commonly used prescriptions for dealing with II and III; but none can be rigorously justified.
- Over time, inconsistencies are eliminated by refined experiments and analyses

This is the Source of large "tolerance", $\Delta\chi^2$

Mimi-Summary

- The important issue is not about methodology: likelihood vs. χ^2; or Monte Carlo sampling or Hessian approximation, …
  - They are essentially equivalent, given consistent theoretical and experimental input.
- The challenges concern:
  - Catalog, define, and quantify theoretical uncertainties;
  - Learn to live and work with imperfect and incompatible data sets---there is no unique procedure, only intuition;
  - Learn to "agree to disagree";
  - Learn to compromise, forge consensus (e.g. choice of sensible schemes), while also emphasize distinctiveness, hence diversity and integrity of the physics results.

“Tension” between different physical processes and experiments?

- Intra-process tension:
  - BCDMS / NMC / HERA ?   (cf. GKK; α_s analyses)
  - CC (CCFR) / NC ? (nuclear vs. nucleon targets, ..)
  - CDF / D0 (both prefer large-x gluons; but there are more subtle tensions)
- Inter-process tension:
  - DIS / Jets ? (MRST2003)
  - DY / Jets ? (MRST2001 ?)

How do we systematically address these potential incompatibilities?

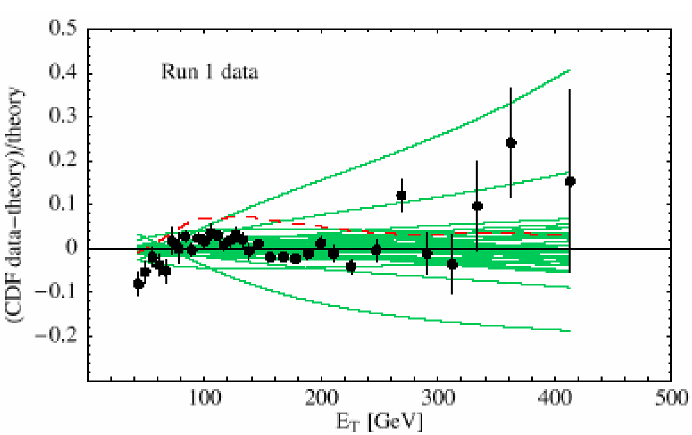
Likelihood method of GKK; Collins and Pumplin

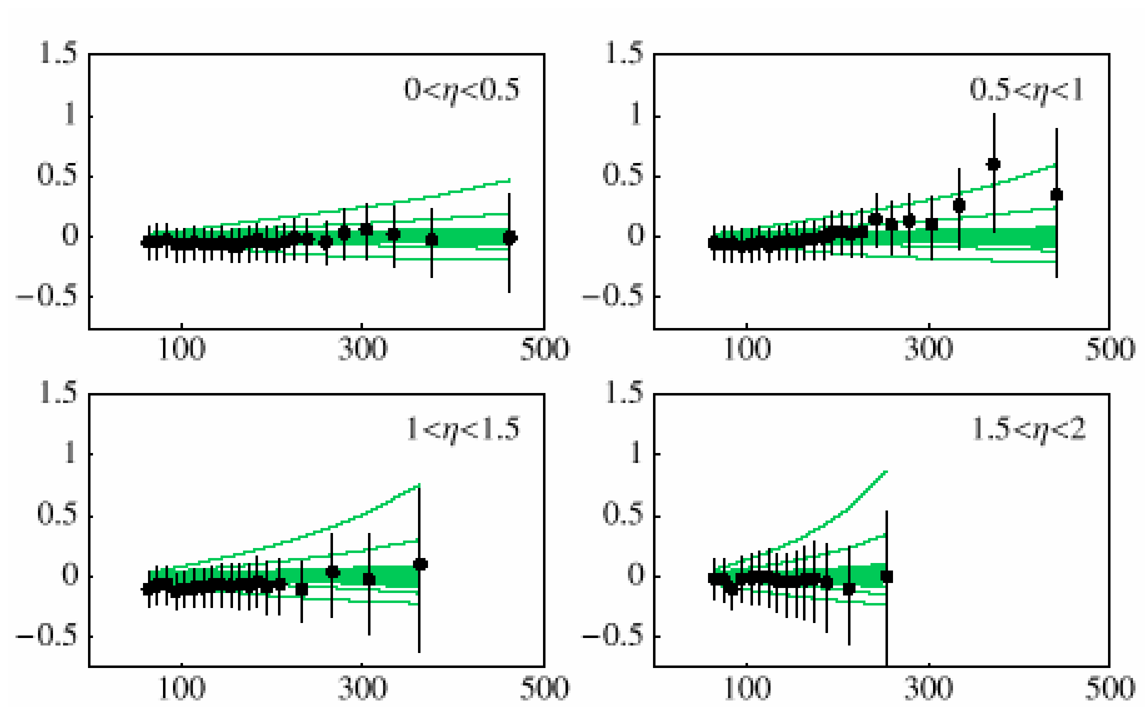
Tension between CDF/D0 data sets?

- CTEQ6 Analysis: Eigenvector 15 in the Hessian approach is particularly sensitive to jet data:

+ direction: D0=1.24 CDF=1.60

- direction: 0.435 2.04





New ways to measure consistency of fit

Pumplin – Ringberg03: (Work in progress with John Collins)

Key idea: In addition to the

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

use the stronger

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

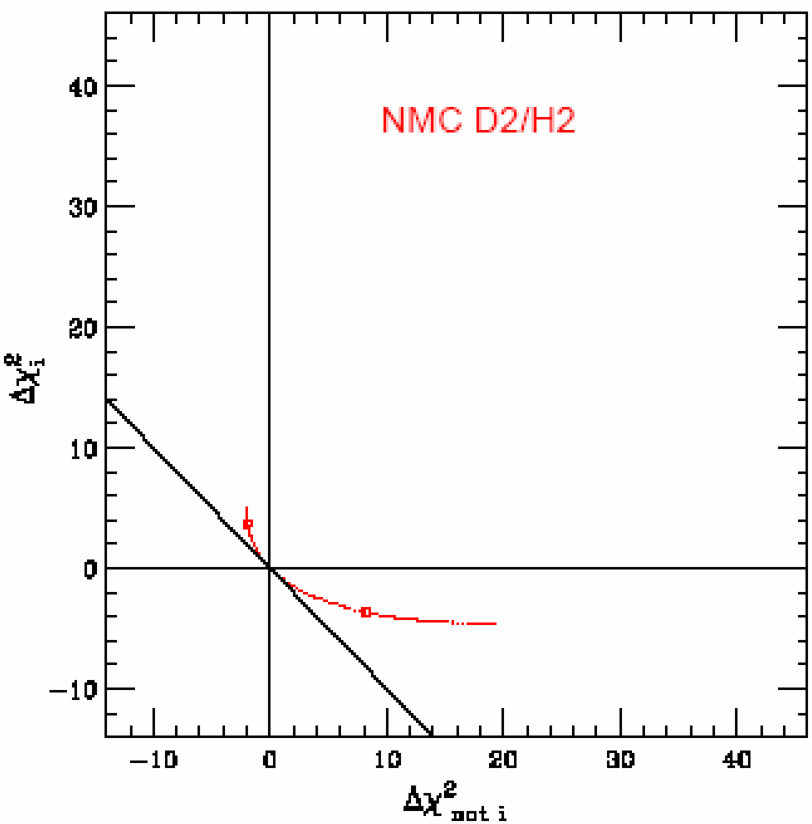
Parameters here are relative weights assigned to

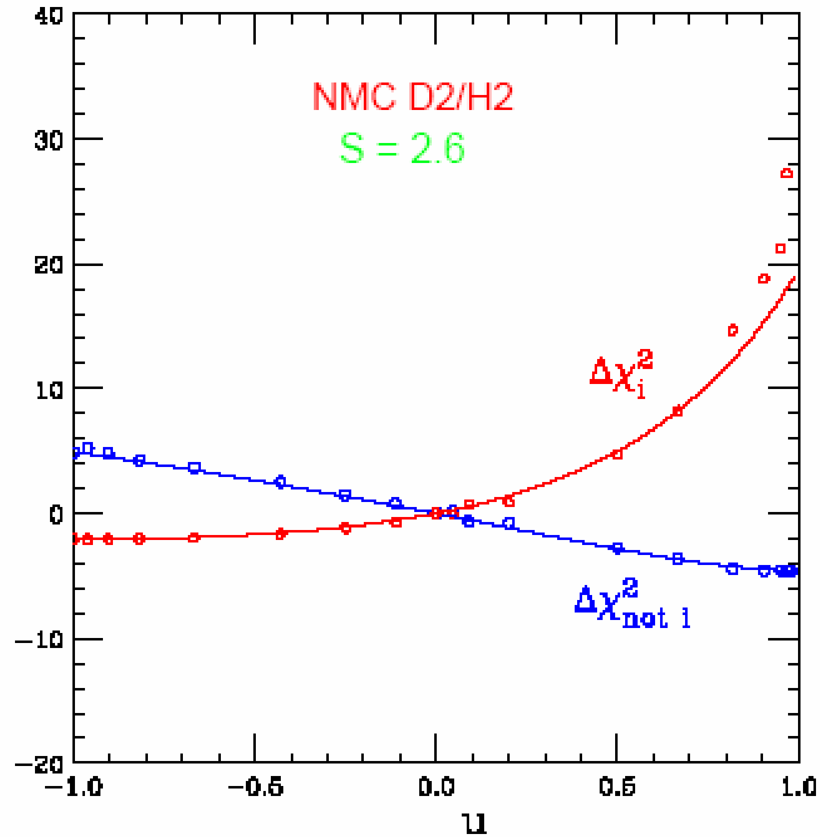
various experiments, or to results obtained using various experimental methods. Examples:

- 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





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

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

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

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

Lessons learned, so far, are not surprising:

- The scale of acceptable changes of χ^2 must be large. Adding a new data set and refitting may increase the χ^2's of other data sets by amounts >> 1.
- Global analysis requires compromises – the PDF model that gives the best fit to one set of data does not give the best fit to others.

But it provides a systematic way of investigating the relevant problems, and quantifying the "incompatibilities".

A critical technical advance in the Hessian approach which enabled the CTEQ uncertainty studies

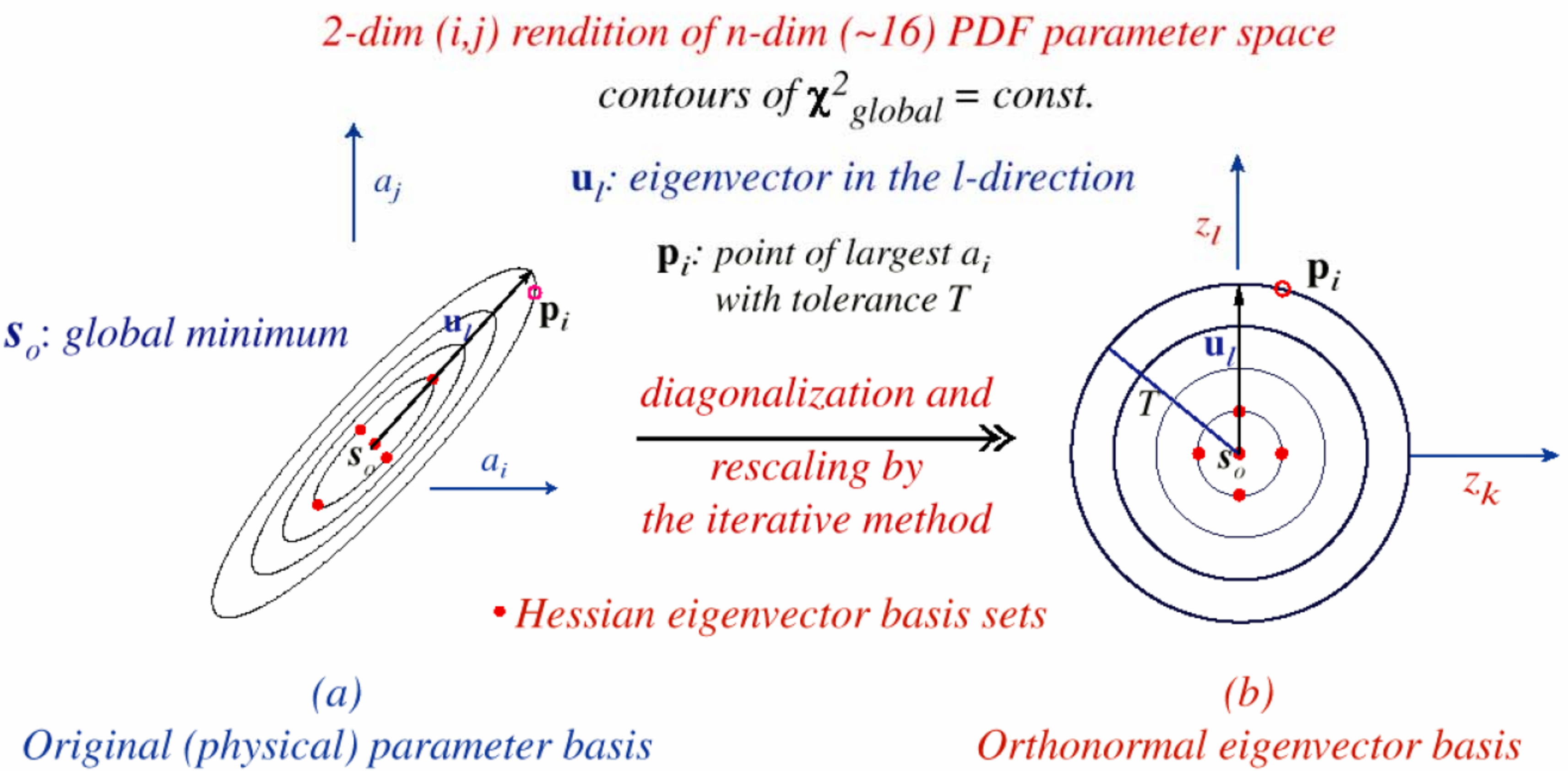
The Hessian method for χ^2 analysis has always been the standard, but uncertainty estimates in global QCD analysis by standard tools had been known to be extremely unreliable due to two practical problems:

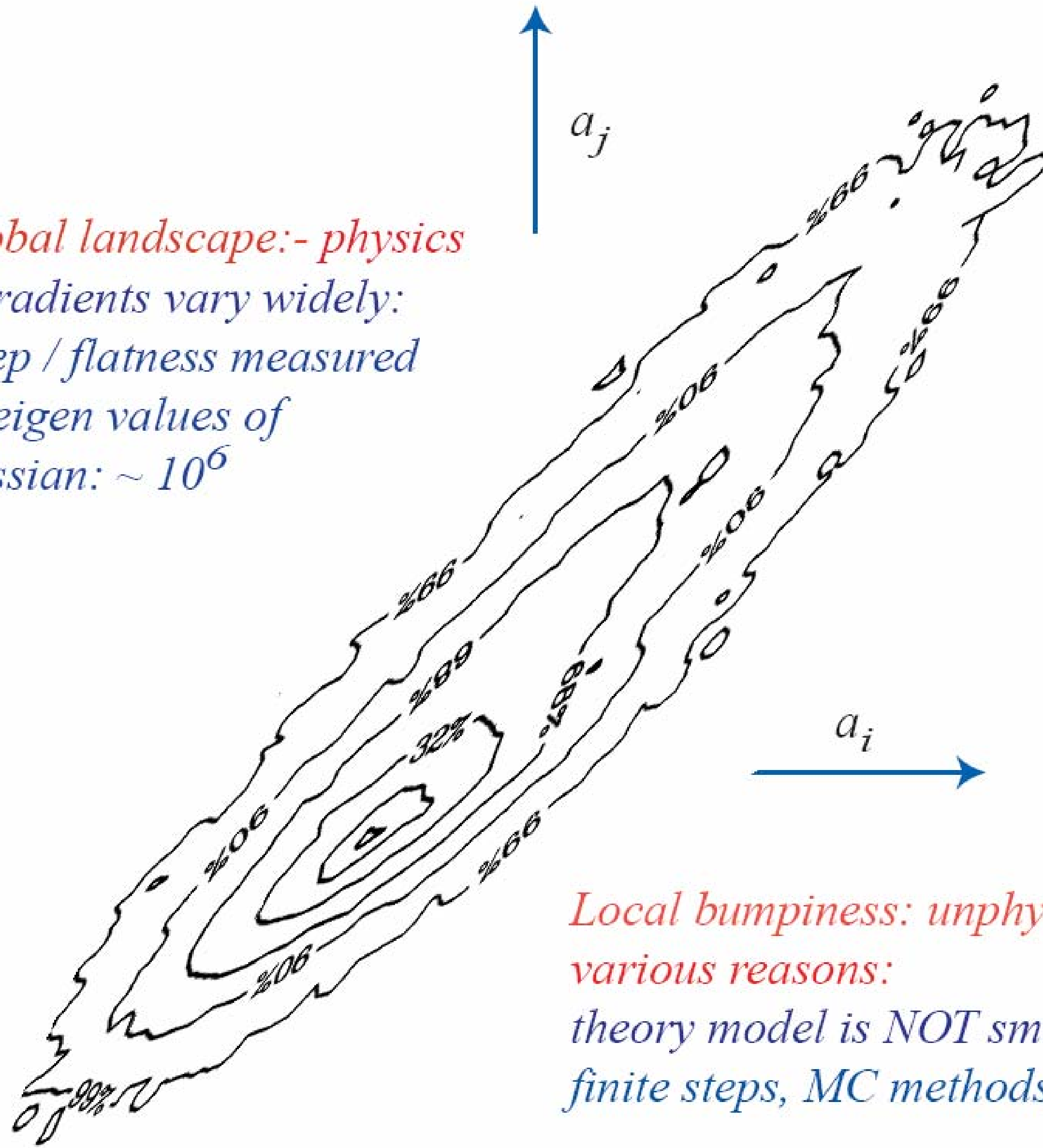
•.extreme range of eigenvalues (flat vs. steep)

• numerical fluctuations of theory predictions

An iterative method by Jon Pumplin solved both of these technical difficulties, provided the means to generate reliable eigenvectors in parton parameter space, hence allow the systematic exploration of this space, particularly the a priori unknown “flat directions”

The Hessian Method of quantifying uncertainties by a complete set of orthonormal eigenvector PDFs





Global landscape:- physics

* gradients vary widely: steep / flatness measured by eigen values of Hessian: ~ 10^6

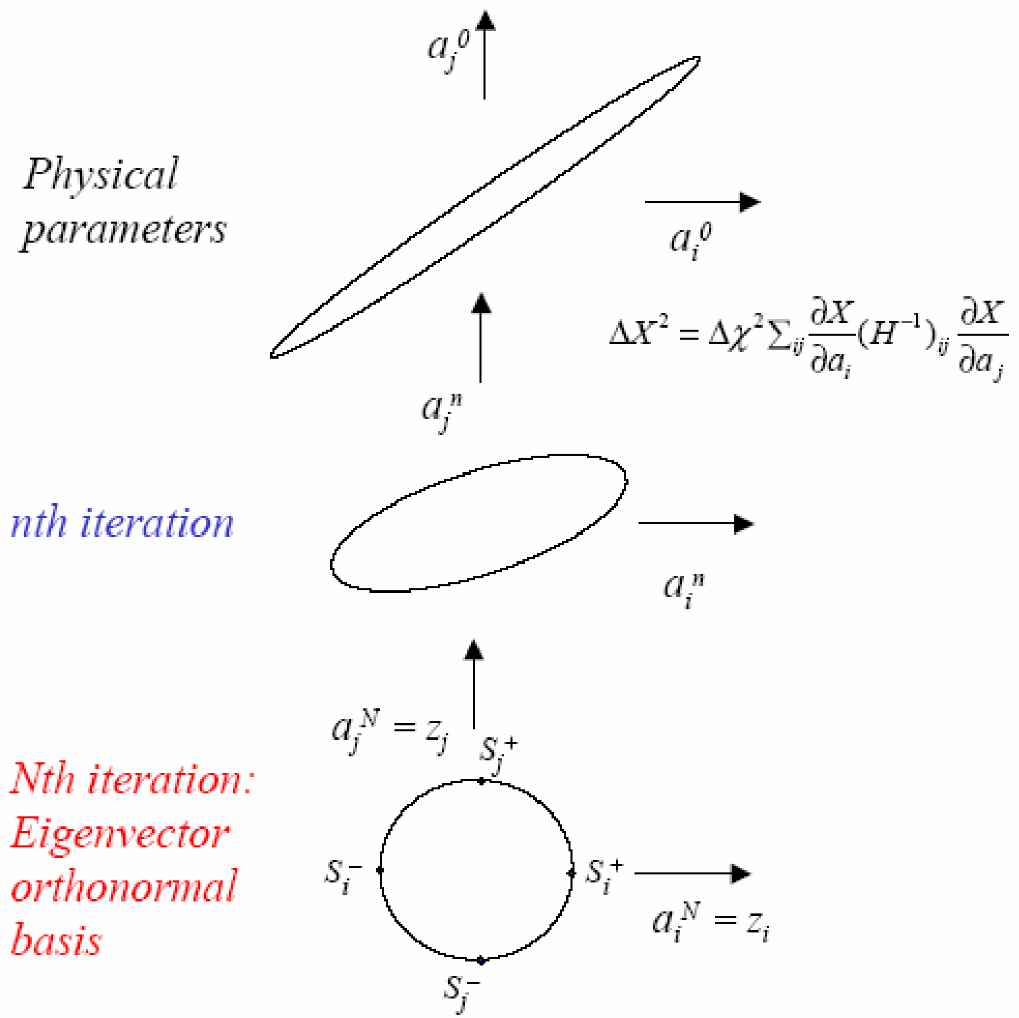
Local bumpiness: unphysical various reasons:

theory model is NOT smooth!; finite steps, MC methods, ...

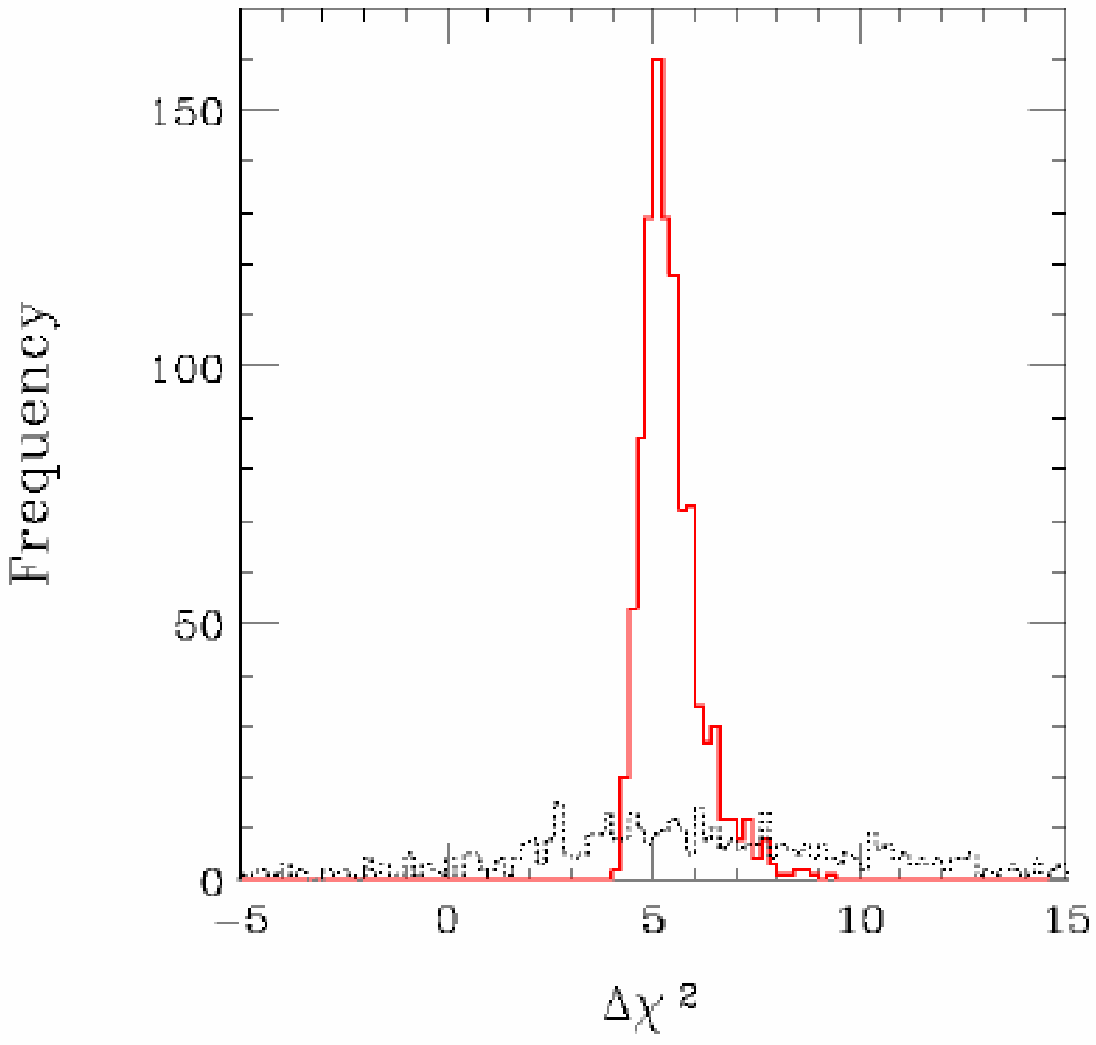
Iterative Method to generate Eigenvectors:

(and dramatically improve numerical reliability)

the $\chi^2 = const.$ ellipsoid



$$\Delta X^2 = \Delta\chi^2 \Sigma_i \left(\frac{\partial X}{\partial z_i}\right)^2 = \Sigma_i [X(S_i^+) - X(S_i^-)]^2$$



Frequency distribution of $\Delta\chi^2$ *according to the Hessian approximation for displacements in random directions for which the true value is* $\Delta\chi^2 = 5$*:*

Solid histogram: results from our iterative method;

Dotted histogram: results obtained from MINUIT.

CTEQ agenda for studying Nucleon Structure and Collider Physics

- Large x behavior of *G(x,Q)*, *u(x,Q)* and *d(x,Q)*;

New frontiers on detailed flavor structure of the nucleon:

- Pinning down the strangeness sector of nucleon structure;
- Understanding the charm content of the nucleon;

Precision W/Z phenomenology at the Tevatron and LHC

- Predictions by and feedback to global analysis
- Transverse momentum, resummation and W-mass
- NNLO analysis
- Higgs, Top, and Beyond SM Phenomenology