being ecumenical

PDF tools and uncertainties

J. Huston

Michigan State University

W theory workshop 11/6/03

Global pdf fits

- Calculation of production cross sections at the Tevatron and LHC relies upon knowledge of pdfs in relevant kinematic range
- pdfs are determined by global analyses of data from DIS, DY and jet production
- Two major groups that provide semi-regular updates to parton distributions when new data/theory becomes available
  - MRS->MRST98->MRST99->MRST2001->MRST2002
  - CTEQ->CTEQ5->CTEQ5(1)->CTEQ6->CTEQ6.1
- CTEQ6 is based on series of previous CTEQ distributions, but represents more than an evolutionary advance
  - update to new data sets
  - incorporation of correlated systematic errors for all experiments in the fit
  - new methodology enables full characterization of parton parametrization space in neighborhood of global minimum
    - Hessian method
    - Largrange Multiplier
  - results available both in conventional formalism and in Les Houches accord format (more on this later)

Uncertainties in pdf fits

- Two sources
  - Experimental errors
    - Hessian/Lagrange multiplier techniques designed to address estimate of these effects
      - question is what $\Delta\chi^2$ change best represents estimate of uncertainty
        - a strict fundamentalist would say $\Delta\chi^2$ of 1 (for 1 σ error)
        - CTEQ uses $\Delta\chi^2$ of 100 (out of 2000) for something like a 90% CL limit
        - MRST uses $\Delta\chi^2$ of 50 for 90% CL limit
  - Theoretical
    - higher twist/non-perturbative effects
      - choose Q^2 and W cuts to try to avoid
    - higher order effects
      - is NNLO necessary yet?
    - small x
      - $\alpha_s^n \ln^{n-1}(1/x)$ terms
    - large x
      - $\alpha_s^n \ln^{2n-1}(1-x)$ terms
    - edge of phase space effects
      - see, for example, in hep-ph/0303013 for inclusive jets at the Tevatron

See for example discussion in hep-ph/0308087; MRST paper, another ecumenical gesture

Nuts/bolts of fits

- Functional form for CTEQ fits used is:
  - $xf(x,Q_o) = A_o\, x^{A1}\, (1-x)^{A2}\, e^{A3x}\, (1 + A_4 x)^{A5}$
    - Q_o = 1.3 GeV (below any data used in fit)
      - easier to do forward evolution than backward
    - functional form arrived at by adding a 1:1 Pade expansion to quantity d(log xf)/dx
    - more versatile than form used in CTEQ5 or MRST
    - there are 20 free parameters used in the global fit
- Light quarks treated as massless; evolution kernels of PDFs are mass-independent
- Zero mass Wilson coefficients used in DIS structure functions

PDF Uncertainties

- What's unknown about PDF's
  - the gluon distribution
  - strange and anti-strange quarks
  - details in the {u,d} quark sector; up/down differences and ratios
  - heavy quark distributions

- □ of quark distributions (q + qbar) is well-determined over wide range of x and Q^2
  - Quark distributions primarily determined from DIS and DY data sets which have large statistics and systematic errors in few percent range (±3% for $10^{-4}<x<0.75$)
  - Individual quark flavors, though may have uncertainties larger than that on the sum; important, for example, for W asymmetry
- information on dbar and ubar comes at small x from HERA and at medium x from fixed target DY production on H_2 and D_2 targets
  - note dbar≠ubar
- strange quark sea determined from dimuon production in □ DIS (CCFR)
- d/u at large x comes from FT DY production on H_2 and D_2 and lepton asymmetry in W production

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,

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

The fitting parameters will be $\{a_\lambda\}$ (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 kth (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

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

where

$$A_{kk'} = \delta_{kk'} + \sum_{i=1}^{N} \frac{\beta_{ki}\beta_{k'i}}{\alpha_i^2}$$

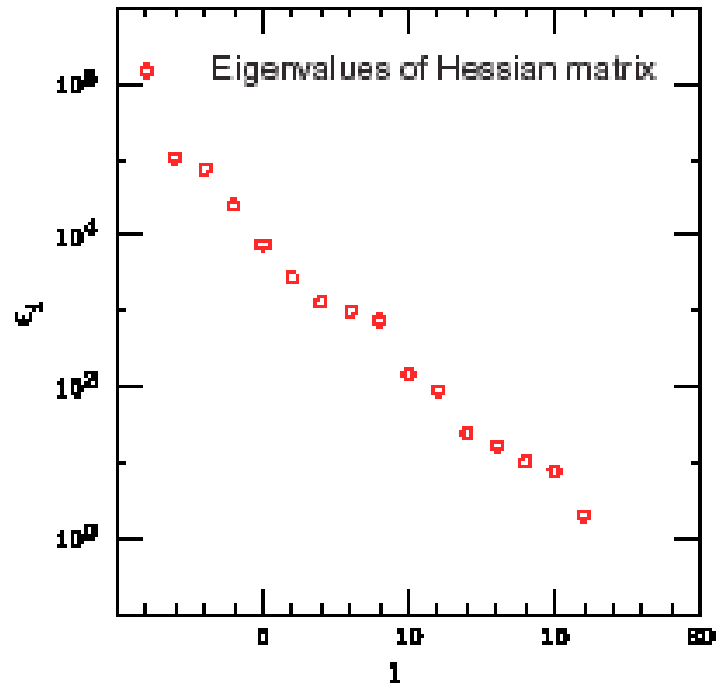
$$B_k = \sum_{i=1}^{N} \frac{\beta_{ki}(D_i - T_i)}{\alpha_i^2}.$$

Note that the $\hat{r}_k$'s depend on the PDF model parameters $\{a_\lambda\}$. Then

$$\boxed{\chi^2(a_\lambda) = \min_{\{r_k\}} \chi'^2(a_\lambda, r_k)}$$

Computational methods: Hessian

- Hessian matrix method
  - explore the variation of $\chi^2(a)$ in the neighborhood of the global minimum in the n-dimensional parameter space
  - 20 free parameters in the fit so 20-dimensional space and 20 eigenvectors
    - MRST has 15 free parameters
  - largest eigenvalues correspond to best-determined directions



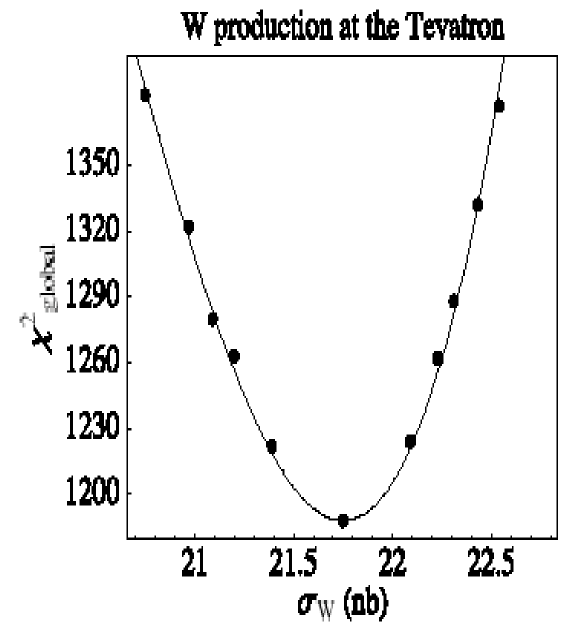
$$\chi^2 - \chi^2_{min} \equiv \Delta\chi^2 = \sum_{i,j} H_{ij}(a_i - a_i^{(0)})(a_j - a_j^{(0)})$$

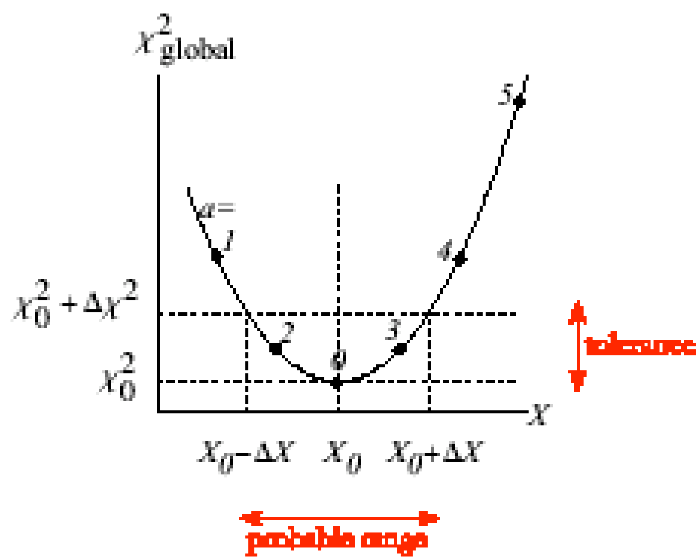
We can then use the standard formula for linear error propagation.

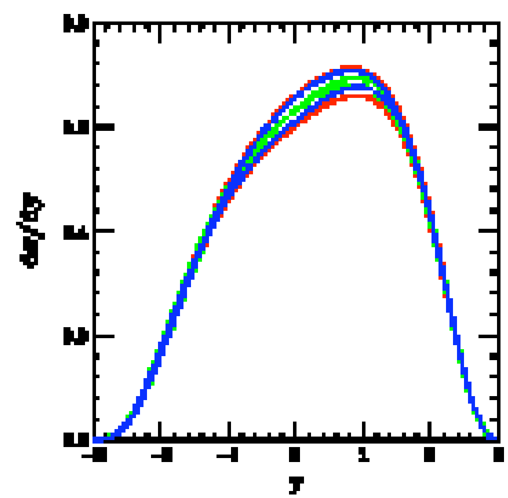
$$(\Delta F)^2 = \Delta\chi^2 \sum_{i,j} \frac{\partial F}{\partial a_i}(H)^{-1}_{ij}\frac{\partial F}{\partial a_j},$$

Computational methods: Lagrange multiplier

- Lagrange multiplier
  - constrained minimization to obtain the best fit as a function of X
    - X can be W cross section, W rapidity, W asymmetry, high E_T jet cross section
  - T^2=max allowed change in χ^2

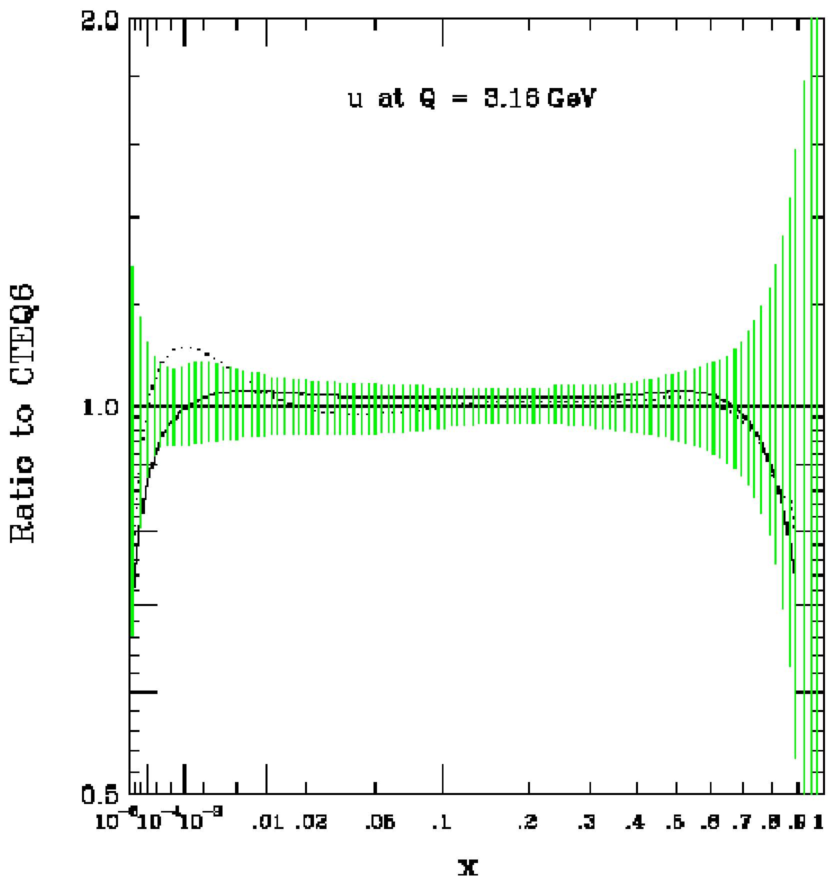


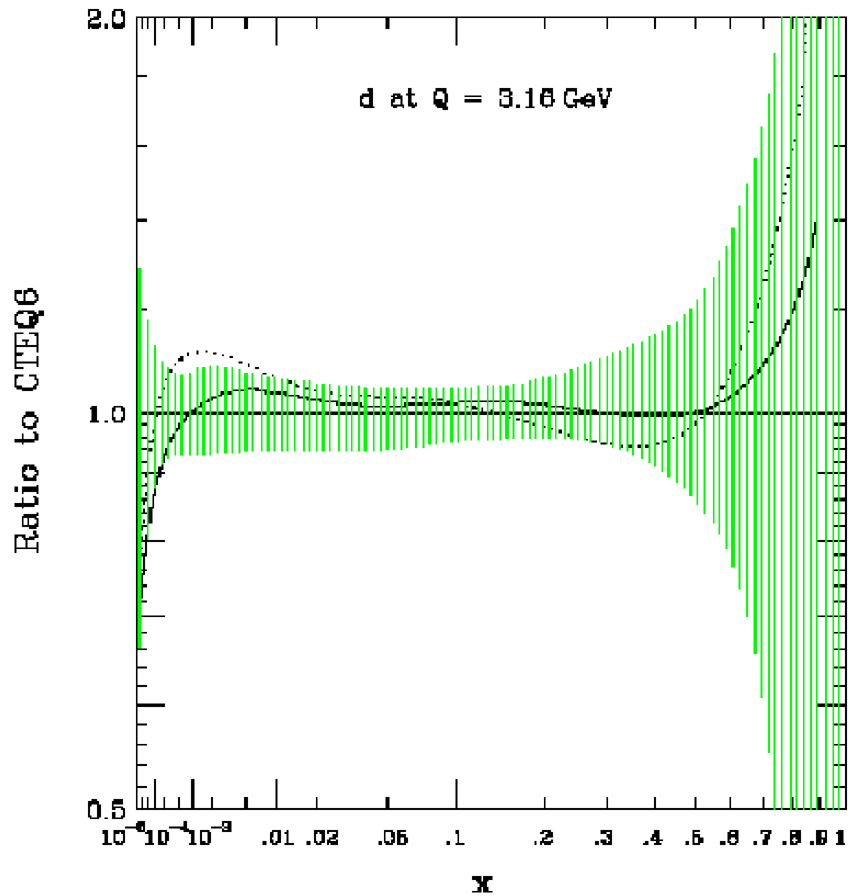




PDF Uncertainties

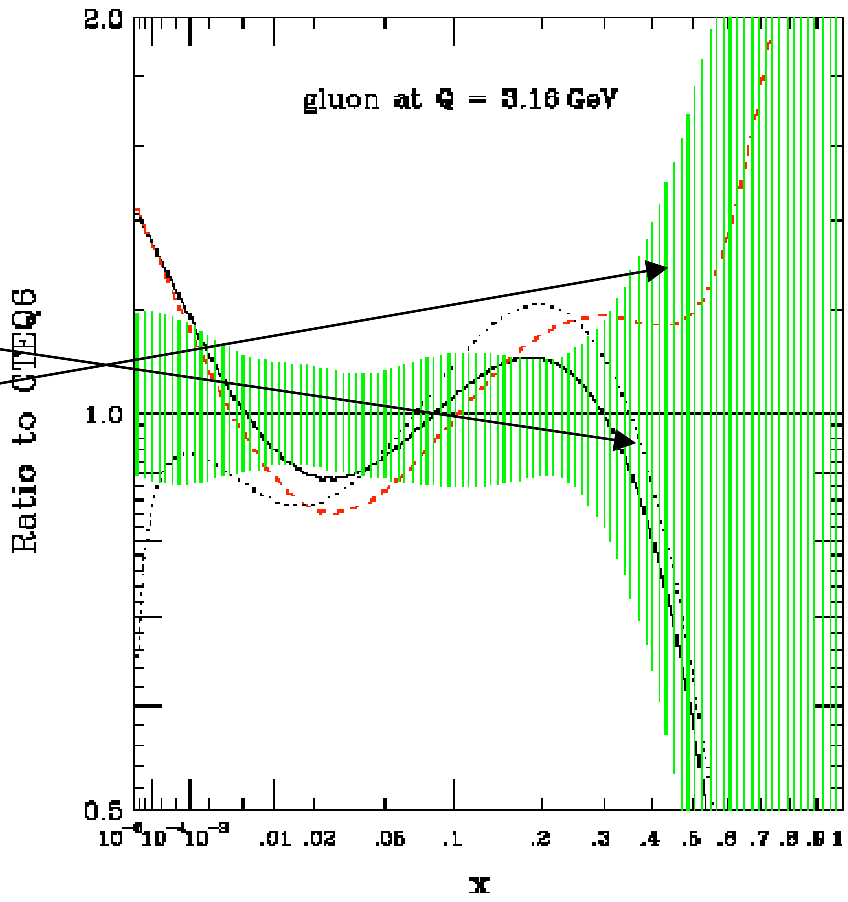
- Use Hessian technique (T=10; $\Delta\chi^2$=100)





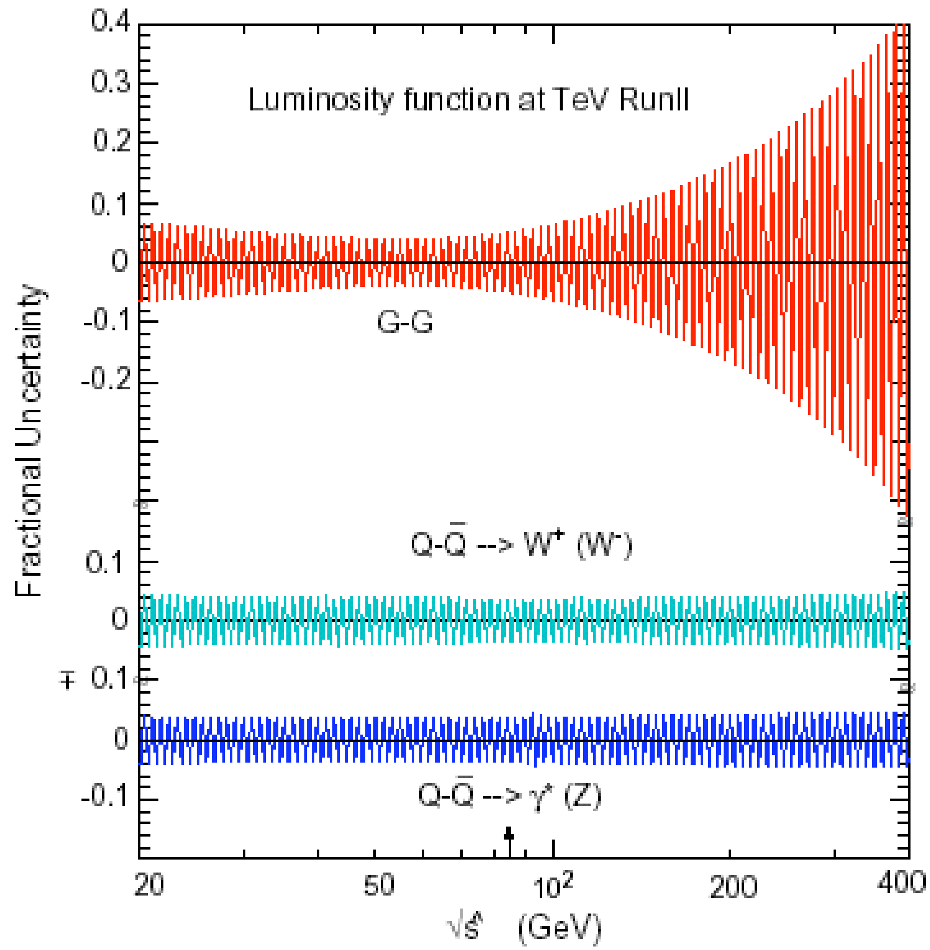
Gluon Uncertainty

- Gluon is fairly well-constrained up to an x-value of 0.3
- CTEQ6.1 gluon is stiffer than CTEQ5M
- Not quite as stiff as CTEQ5HJ



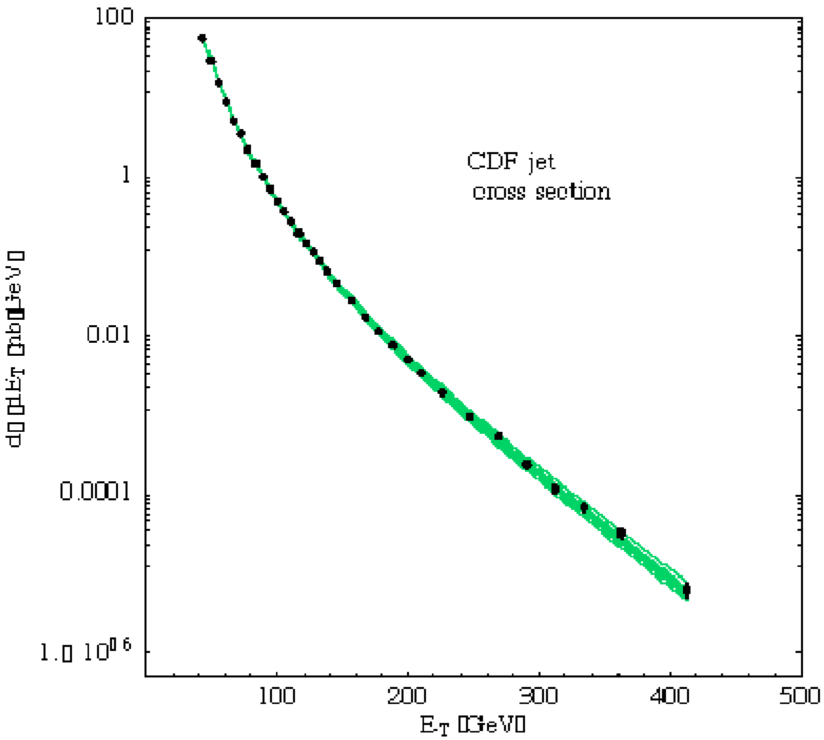
Luminosity function uncertainties at the Tevatron

- Uncertainties are larger for processes involving gluons (as expected)
- Quark processes well-constrained



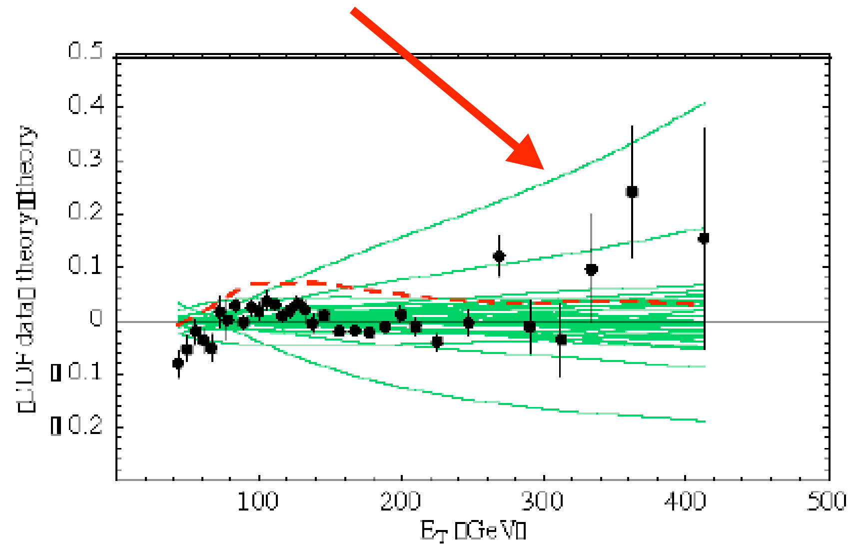
Example: PDF uncertainties for Run 1 jet cross section

- 20 free parameters in the fit
- In the Hessian method, a 20X20 matrix is diagonalized and 20 orthogonal eigenvector directions in parameter space are determined



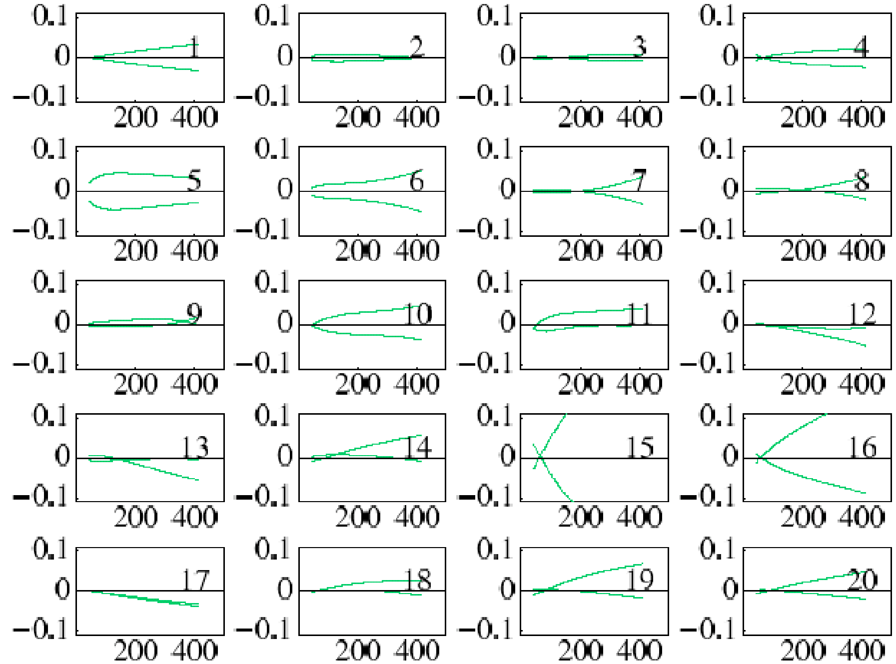
•Each eigenvector direction corresponds to some linear combination of pdf parameters
•Large eigenvalues correspond to highly determined directions (e.g. valence quarks)
•Small eigenvalues correspond to poorly determined directions (high x gluon)
•Result is 40 pdf's (go along + and - direction $\Delta\chi^2$ of 100 for each eigenvalue)

Note 1 eigenvector(15+) leads to noticeably larger prediction than the others



CDF jet cross section uncertainties

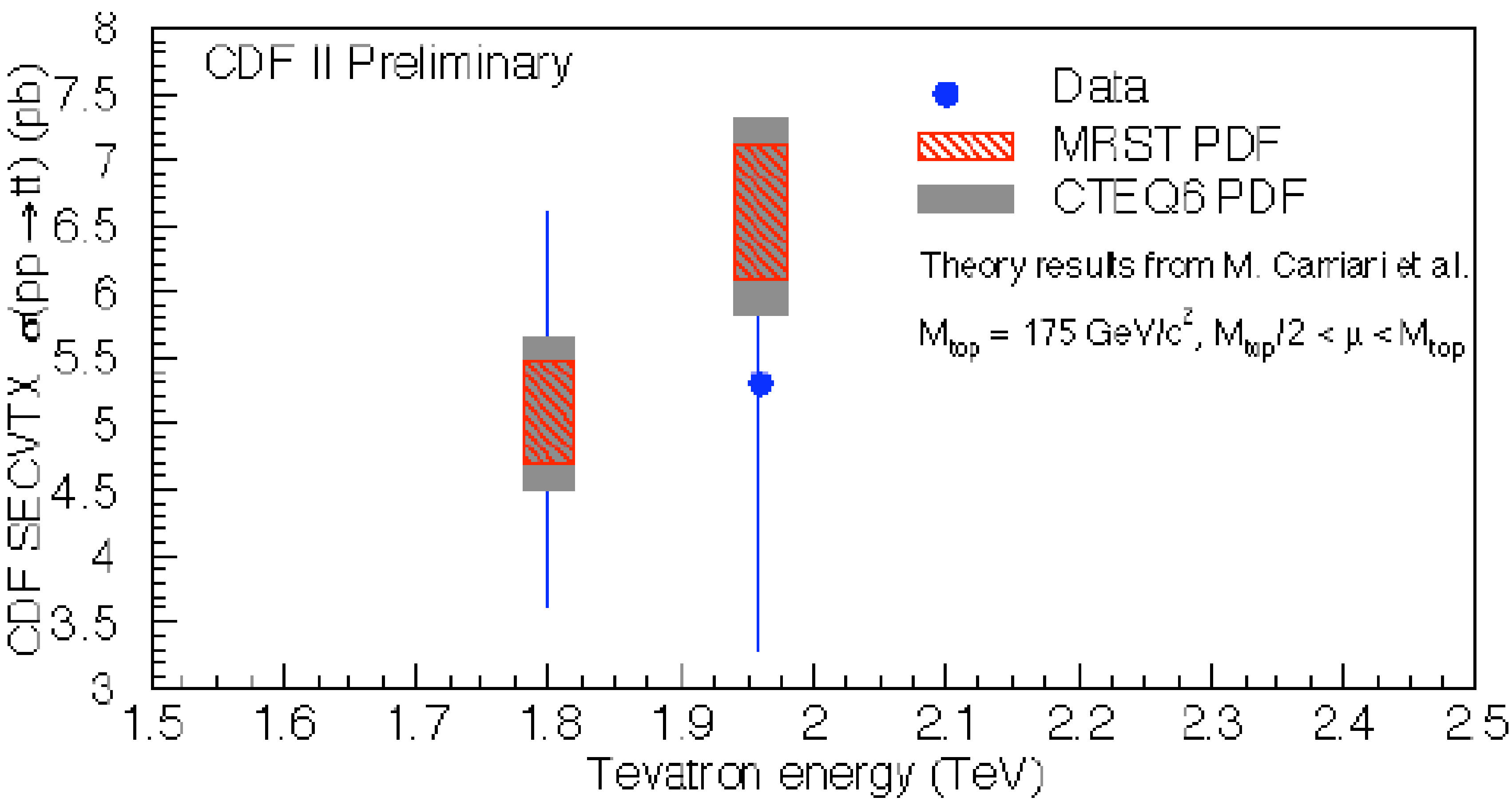
- On the right are shown the uncertainties for the CDF jet cross section along each eigenvector (□□^{2} = 100)
  - jet cross section most sensitive to eigenvector 15
    - which mainly contains parameters relating to behavior of high x gluon



NB: MRST will not have the equivalent of the last 5 eigenvectors

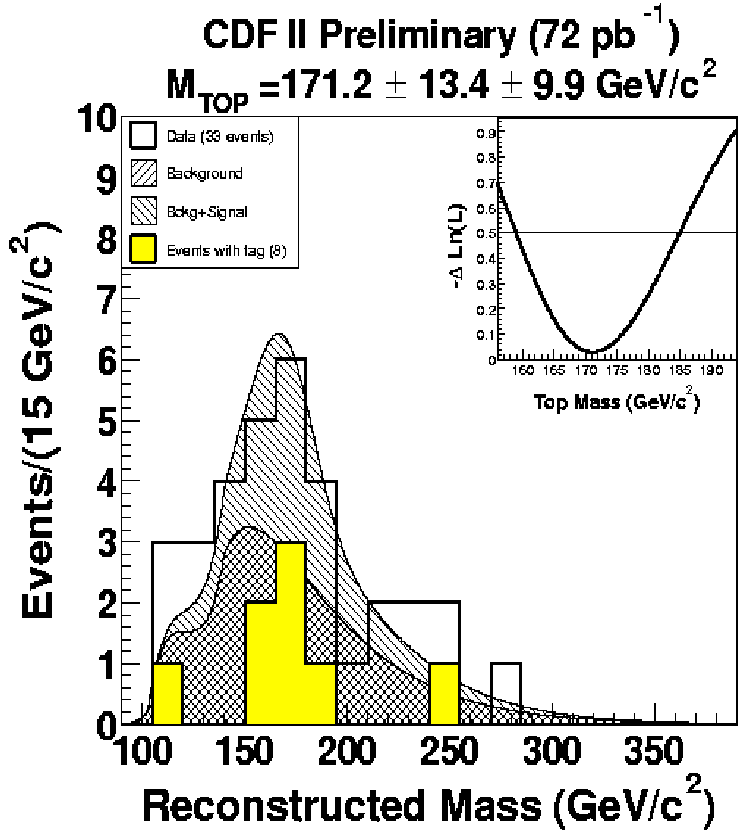
pdf uncertainties for top cross sections

Dominant error in top cross section arises from eigenvector 15

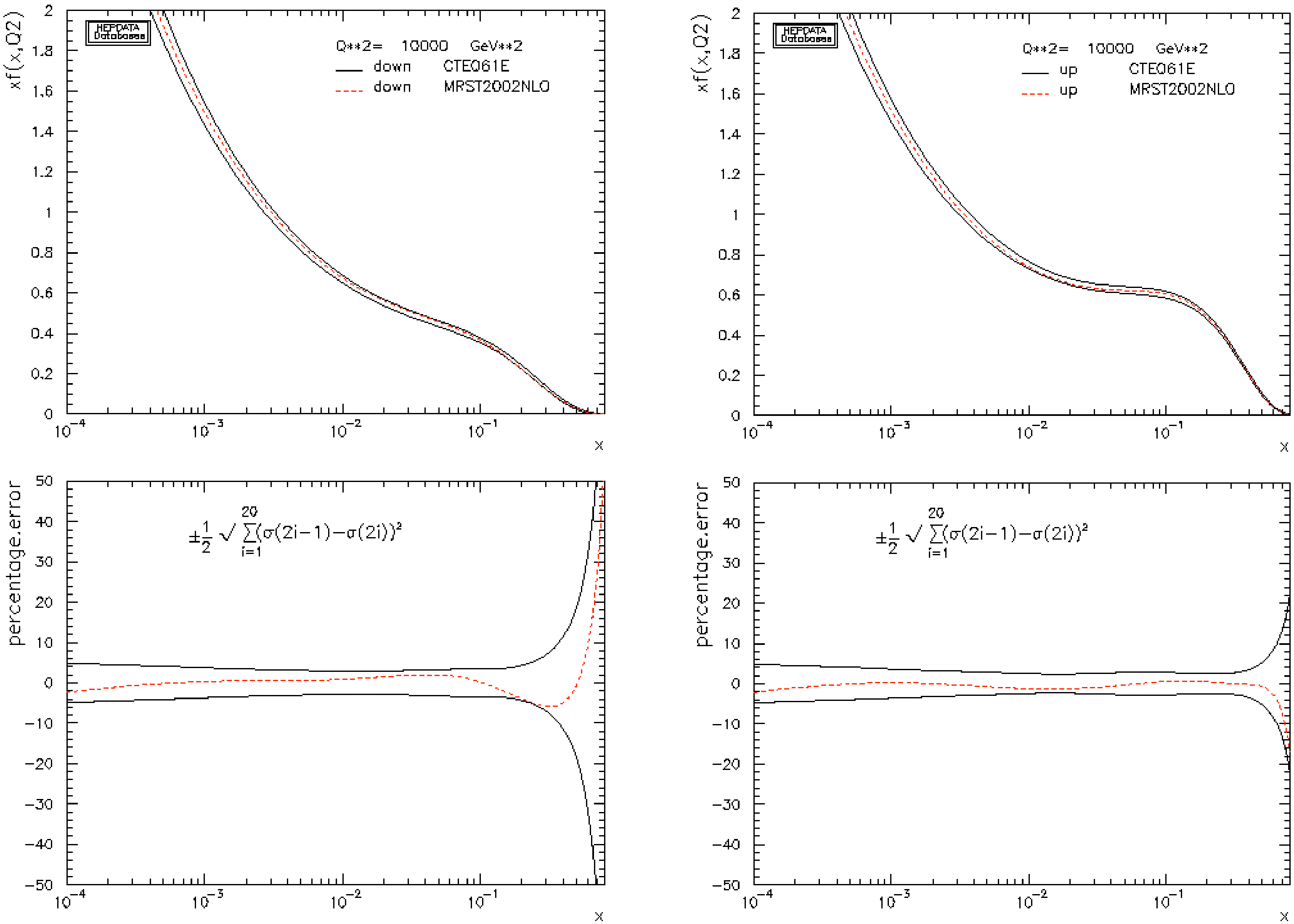


Uncertainty in top mass

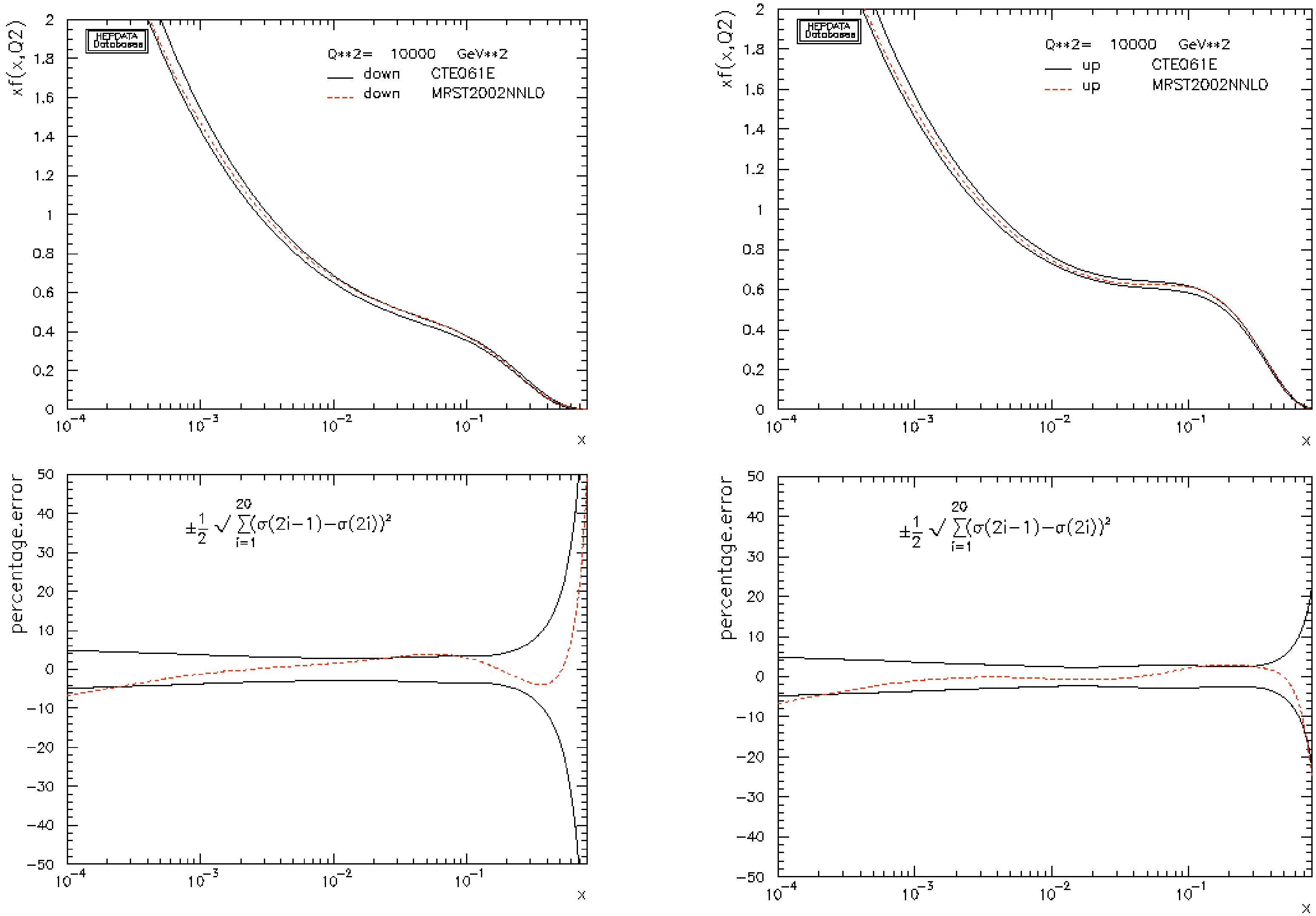
- Eigenvector 15 also causes largest error for measurement of top mass
  - more gg production
  - more initial state radiation



Compare MRST2002 to CTEQ6 error band



Compare MRST NNLO to CTEQ6 error band



Theoretical uncertainties for W/Z □

Reasonable stability order by order for (quark-dominated) W and Z cross-sections.

However, changes of order 4%. Much bigger than uncertainty due to experimental errors.

This fairly good convergence is largely guaranteed because the quarks are fit directly to data.

From pdf uncertainties:

CTEQ obtain for $\alpha_S = 0.118$

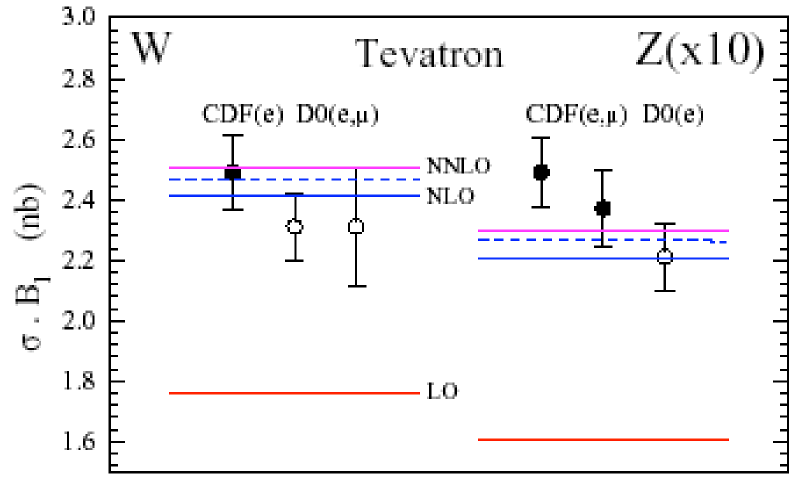
$$\Delta\sigma_W(\mathrm{LHC}) \approx \pm 4\% \quad \Delta\sigma_W(\mathrm{Tev}) \approx \pm 4$$

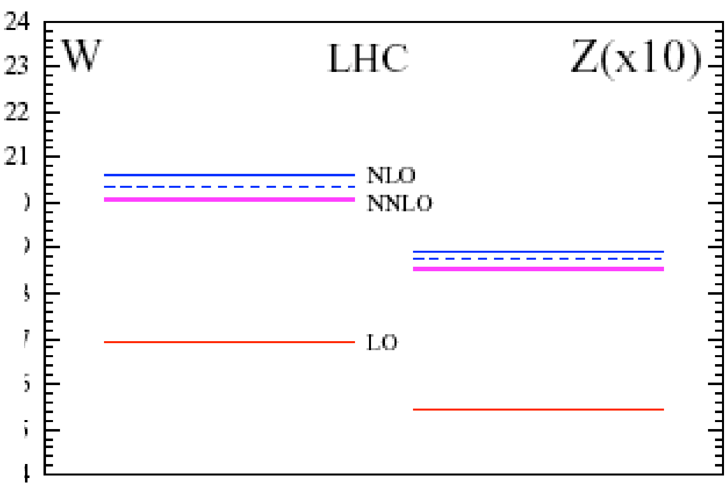
$$\Delta\sigma_H(\mathrm{LHC}) \approx \pm 5\%.$$

MRST use a wider range of data, and if $\Delta\chi^2 \sim 50$ find for $\alpha_S = 0.119$

$$\Delta\sigma_W(\mathrm{Tev}) \approx \pm 1.2\% \quad \Delta\sigma_W(\mathrm{LHC}) \approx \pm 2\%$$

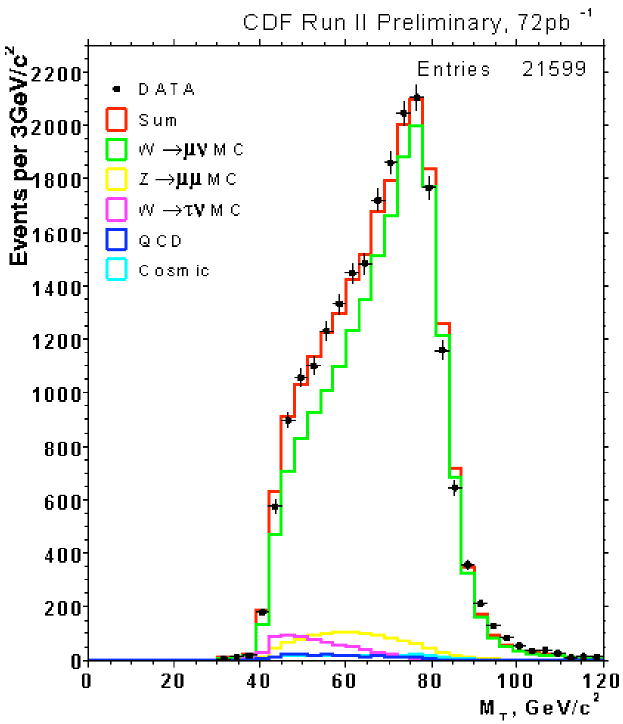
$$\Delta\sigma_H(\mathrm{Tev}) \approx \pm 4\% \quad \Delta\sigma_H(\mathrm{LHC}) \approx \pm 2\%.$$



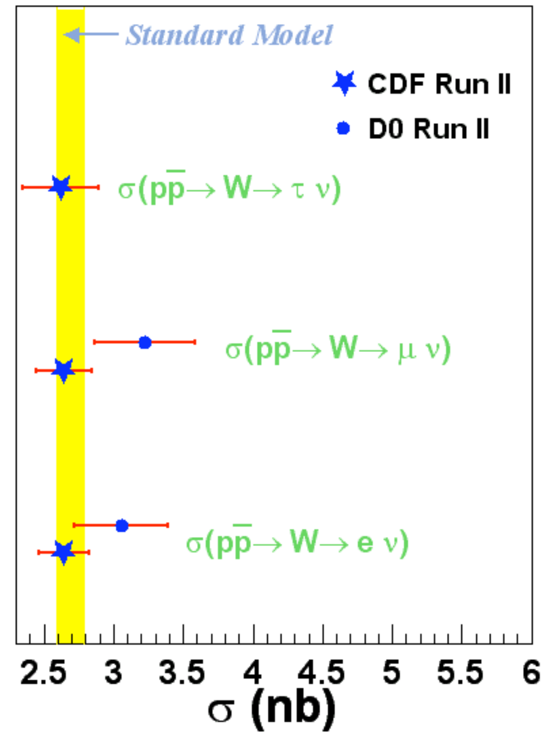


Note that CTEQ pdf error band is larger than MRST error band AND CTEQ □'s are ~2% higher than MRST

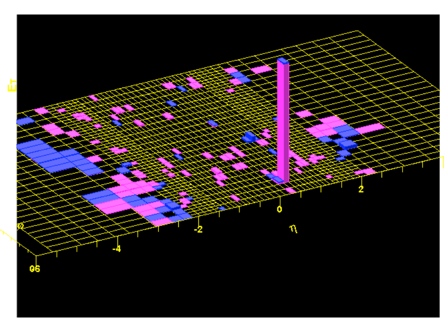
σ•B(W-->lν$_l$)



- Clean signature
  - isolated lepton
  - missing E_T
- High σ and S/B
- Ideal for calibration and precision measurements
  - standard candle for high p_T

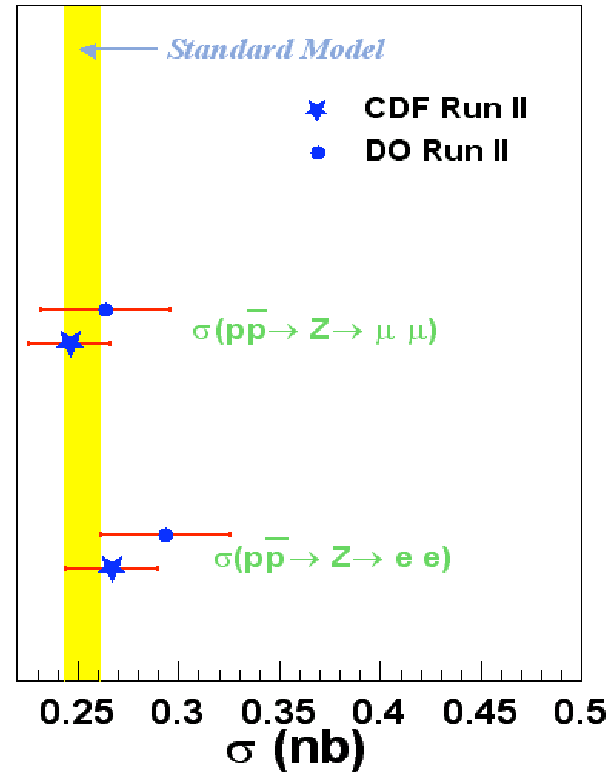


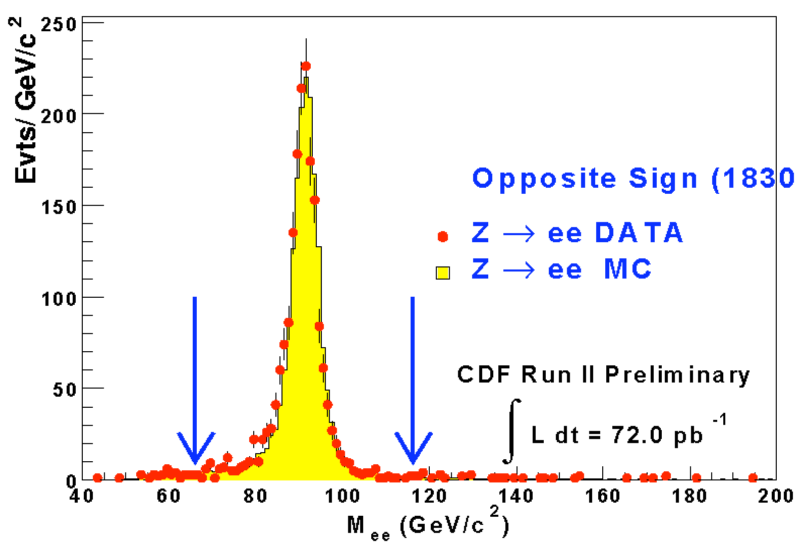
	Sample	Back.	σ•B(W-->lν$_l$) (nb)
e	38625	6%	**2.64**$\pm0.01_{stat}\pm0.09_{sys}\pm0.16_{lum}$
μ	21599	11%	**2.64**$\pm0.02_{stat}\pm0.12_{sys}\pm0.16_{lum}$
τ	2346	26%	**2.62**$\pm0.07_{stat}\pm0.21_{sys}\pm0.16_{lum}$



□•B(Z-->ll)

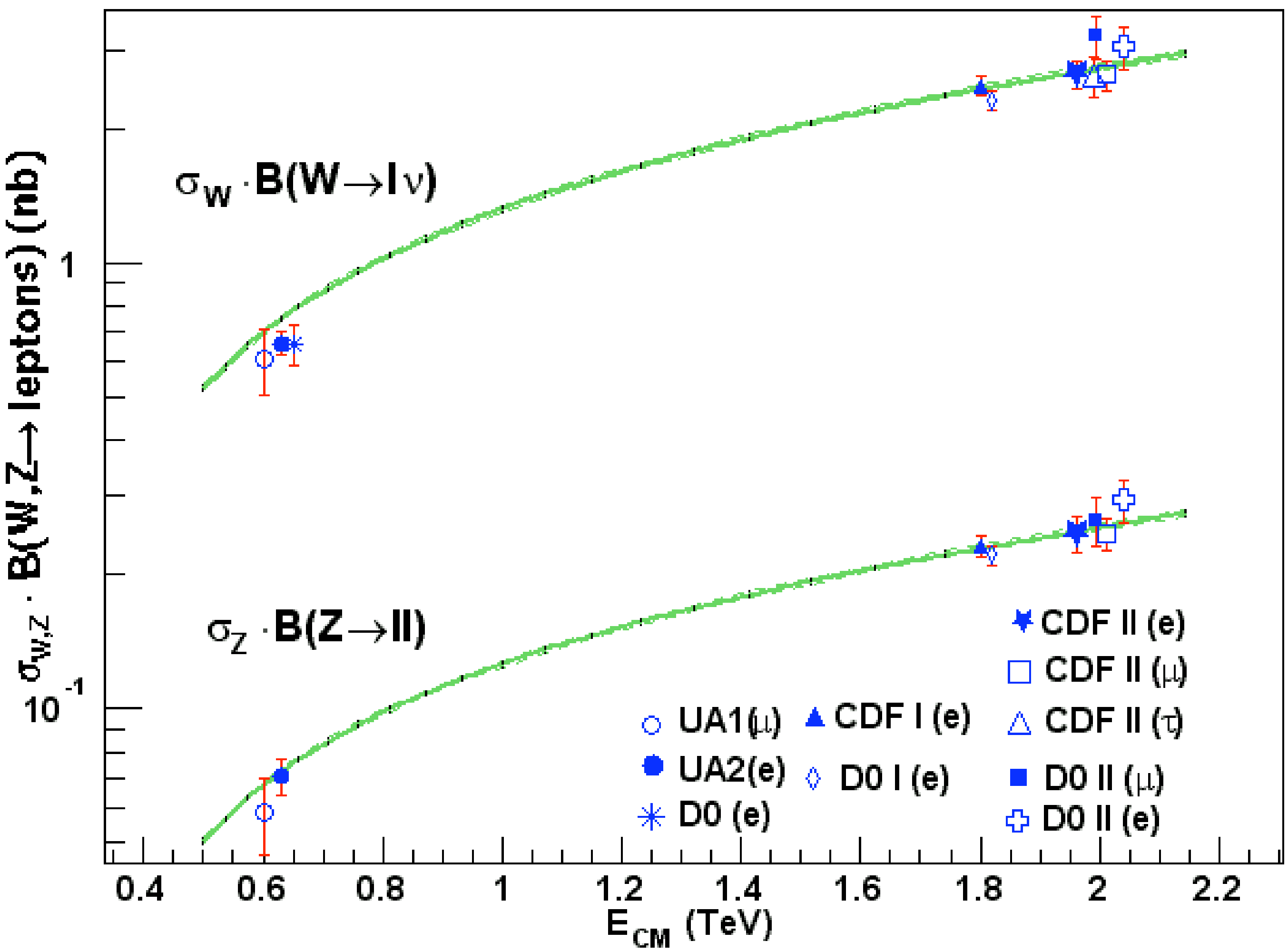
- Require two isolated leptons
  - negligible backgrounds
- Essential for detector calibrations
  - energy scale and resolution
  - ID efficiency





	Sample	Back.	□•B(Z-->ll) (pb)
e	1830	0.6%	$267 \pm 6_{stat} \pm 15_{sys} \pm 16_{lum}$
□	1631	0.9%	$246 \pm 6_{stat} \pm 12_{sys} \pm 15_{lum}$

W and Z cross sections



Luminosity determination

- With modern PDF's, QCD predictions can have smaller errors than normalization to the total inelastic cross section
  - measurement of luminosity can be a tricky business
  - …and in addition, in Run 1 CDF and D0 used values for inelastic cross section different by ~6% (so luminosities were different by 6%)
  - working on deriving a common value for the inelastic cross section for Run 2
- But, ratios of high p_T cross sections to W or Z will have smaller errors than absolute cross sections.
  - …so cross sections should be quoted that way as well

W mass

from mlm's talk at VLHC workshop

Example: accuracy in the extraction of the W cross-section

- NNLO total X-sections known, residual theory uncertainty ~few%.
- MC necessary to evaluate acceptance, and therefore total σ, to be compared with inclusive calculation.
- Effects other than NNLO seem to be have an effect on acceptance more important than the NLO-NNLO difference.

σ(W)	MRST 2000	MRST 2001
NLO		
Fnal	2.39	2.41
LHC	20.5	20.6
NNLO		
Fnal	2.51	2.50
LHC	19.9	20.0

Acceptance for lepton with p_T>20 GeV and |η|<2.5, using different parameters or approximations: easily over-10% differences

LO	LO, Γ_W=0	LO, no spin corr's	LO, PDF-CTEQ6.19	W+jet, Et>5GeV	Herwig LO (spin corr's)	Herwig MC@NLO (no spin c's)
0.4890(2)	0.4971(2)	0.5259(2)	0.5245(2)	0.5324(2)	0.5063	0.5575

Effective use of pdf uncertainties

- PDF uncertainties are important both for precision measurements (W/Z cross sections) as well as for studies of potential new physics (a la jet cross sections at high E_T)
- Most Monte Carlo/matrix element programs have "central" pdf's built in, or can easily interface to PDFLIB
- Determining the pdf uncertainty for a particular cross section/distribution might require the use of many pdf's
  - CTEQ Hessian pdf errors require using 40 pdf's
  - GKK on the order of 100
  - **New: MRST2002->30 pdfs**
- Too clumsy to attempt to includes grids for calculation of all of these pdf's with the MC programs
- **->Les Houches accord #2 (also from 2001 workshop)**
  - each pdf can be specified by a few lines of information, if MC programs can perform the evolution
  - fast evolution routine will be included in new releases to construct grids for each pdf

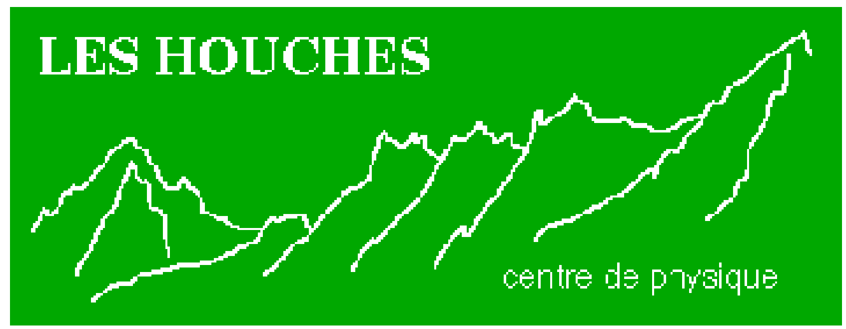
Les Houches accord #2 (LHAPDF)

- Using the interface is as easy as using PDFLIB (and much easier to update)
- First version has CTEQ6M, CTEQ6L, all of CTEQ6 error pdfs and MRST2001 pdfs
- See pdf.fnal.gov
- LHAPDF has been handed off to Durham who will provide support in perpetuity

- call InitiPDFset(*name*)
  - called once at the beginning of the code; *name* is the file name of external PDF file that defines PDF set
- call InitPDF(*mem*)
  - *mem* specifies individual member of pdf set
- call evolvePDF(*x*,*Q*,*f*)
  - returns pdf momentum densities for flavor *f* at momentum fraction *x* and scale *Q*

Les Houches update

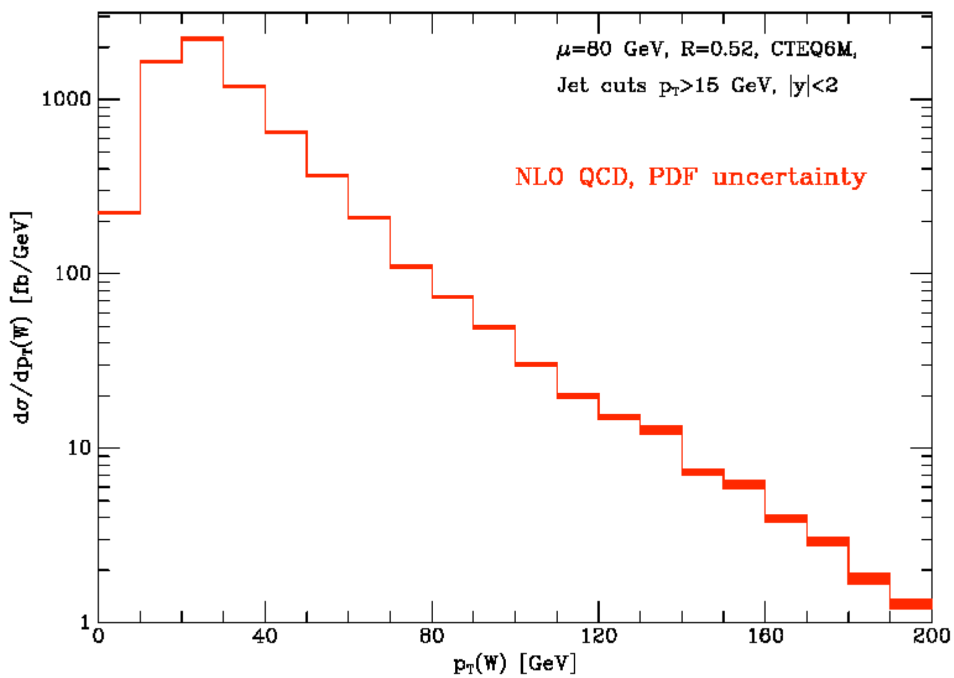
- Reminder: the big idea:
  - The Les Houches accords will be implemented in all ME/MC programs that experimentalists use
  - They will make it easy to generate the multi-parton final states crucial to much of the Run 2 and LHC physics programs and to compare the results from different programs
  - experimentalists/theorists can all share common MC *data sets*
  - They will make it possible to generate the pdf uncertainties for any cross sections measureable at the Tevatron/LHC

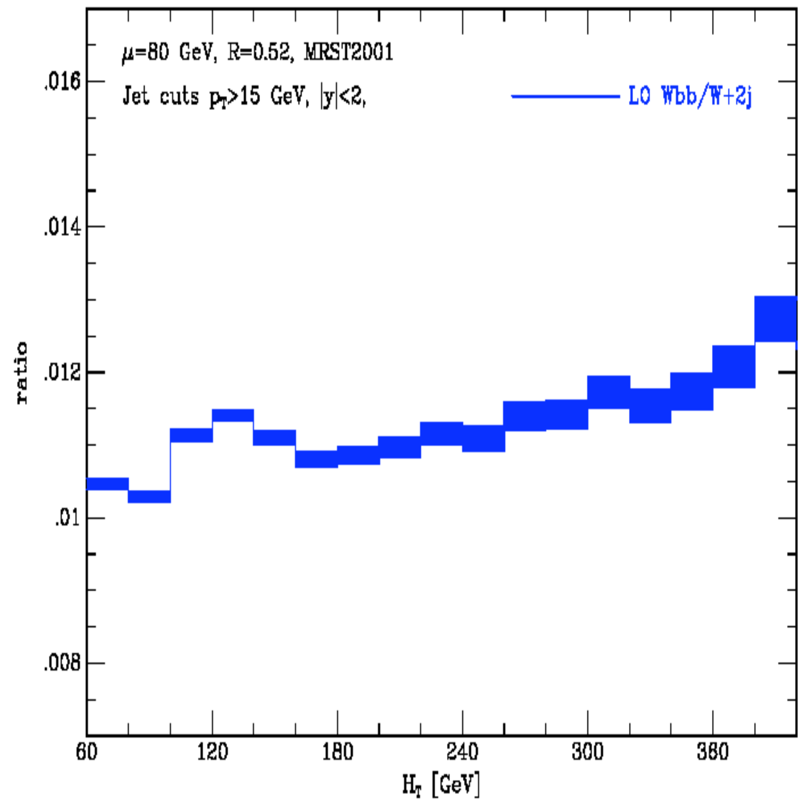


- One project started at Les Houches 2003 (and to be continued at Durham)
  - modify LHAPDF to keep central and all error pdf's in memory at same time

→ generate events with central pdf but keep track of pdf*pdf weight for error pdf's, for each event
beta version ready

Example: testing version





Another workshop in a nice place

- Workshop on Collider Physics at the Kavli Institute for Theoretical Physics at Santa Barbara
  - Jan 12 - April 2, 2004
  - emphases on:
    - pdf's
    - matrix element and Monte Carlo programs
    - NLO MC
    - new physics/EW physics
    - NLO and NNLO calculations
    - impact of data from the Tevatron
  - long-term stays (2-3 weeks) already subscribed but shorter term stays possible
  - opening conference on collider physics Jan. 12-16
    - speakers include Altarelli, Arkani-Hamed, Bethke, Carena, Campbell, Dixon, Frixione, Gehrmann, Huston, Hinchliffe, Kilgore, Kosower, Lykken, Peskin, Sterman, Stirling, Witherell, ,Weerts, Wolf, Yao
    - http://www.kitp.ucsb.edu/activities/collider_c04/register