Parton Distribution Functions: their generation and use (especially at the LHC)

J. Huston

Michigan State University

and IPPP-Durham

2008 CTEQ-MCNET Summer School on QCD Phenomenology and Monte Carlo Generators

CTEQ

Some references

INSTITUTE OF PHYSICS PUBLISHING REPORTS ON PROGRESS IN PHYSICS

Rep. Prog. Phys. **70** (2007) 89–193 doi:10.1088/0034-4885/70/1/R02

Hard interactions of quarks and gluons: a primer for LHC physics

J M Campbell[1], J W Huston[2] and W J Stirling[3] CHS

[1] Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
[2] Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48840, USA
[3] Institute for Particle Physics Phenomenology, University of Durham, Durham DH1 3LE, UK

E-mail: j.campbell@physics.gla.ac.uk, huston@msu.edu and w.j.stirling@durham.ac.uk

Received 14 July 2006, in final form 6 November 2006
Published 19 December 2006
Online at stacks.iop.org/RoPP/70/89

Abstract

In this paper, we will develop the perturbative framework for the calculation of hard-scattering processes. We will undertake to provide both a reasonably rigorous development of the formalism of hard-scattering of quarks and gluons as well as an intuitive understanding of the physics behind the scattering. We will emphasize the role of logarithmic corrections as well as power counting in α_S in order to understand the behaviour of hard-scattering processes. We will include 'rules of thumb' as well as 'official recommendations', and where possible will seek to dispel some myths. We will also discuss the impact of soft processes on the measurements of hard-scattering processes. Experiences that have been gained at the Fermilab Tevatron will be recounted and, where appropriate, extrapolated to the LHC.

(Some figures in this article are in colour only in the electronic version)

- Also refer to lecture notes from previous CTEQ summer schools at http://www.phys.psu.edu/~cteq/
  - in particular lectures on pdf's from 2007 (Jeff Owens) from whom I've taken much material…but from whom there is still much to give

CTEQ

Some references

- I've also downloaded an ATLAS note I wrote almost 10 years ago
- A bit outdated but it still has some useful pedagogical information

April 16, 1999

LHC Guide to Parton Distribution Functions and Cross Sections

J. Huston

Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA

ATL-PHYS-99-008
28 Apr 1999

This ATLAS note is intended to serve as a pedagogical guide on the determination of, the proper use of, and the uncertainties of parton distribution functions and their impact on physics cross sections at the LHC. Portions of this note will be placed in the physics TDR.

I. INTRODUCTION

The calculation of the production cross sections at the LHC for both interesting physics processes and their backgrounds relies upon a knowledge of the distribution of the momentum fraction x of the partons in a proton in the relevant kinematic range. These parton distribution functions (pdf's) are determined by global fits to data from deep inelastic scattering (DIS), Drell-Yan (DY), and jet and direct photon production at current energy ranges. Two major groups, CTEQ and MRS, provide semi-regular updates to the parton distributions when new data and/or theoretical developments become available. The newest pdf's, in most cases, provide the most accurate description of the world's data, and should be utlilized in preference to older pdf sets. The newest sets from the two groups are CTEQ5 [1] and MRST [2]. As will be discussed in Section VII, the primary difference between the two pdf's lies in the size the gluon distribution at large x.

This note is intended to serve as a pedagogical summary; the author is a member of CTEQ and apologizes in advance for any bias in that direction.

C T E Q

MCnet

Understanding cross sections at the LHC

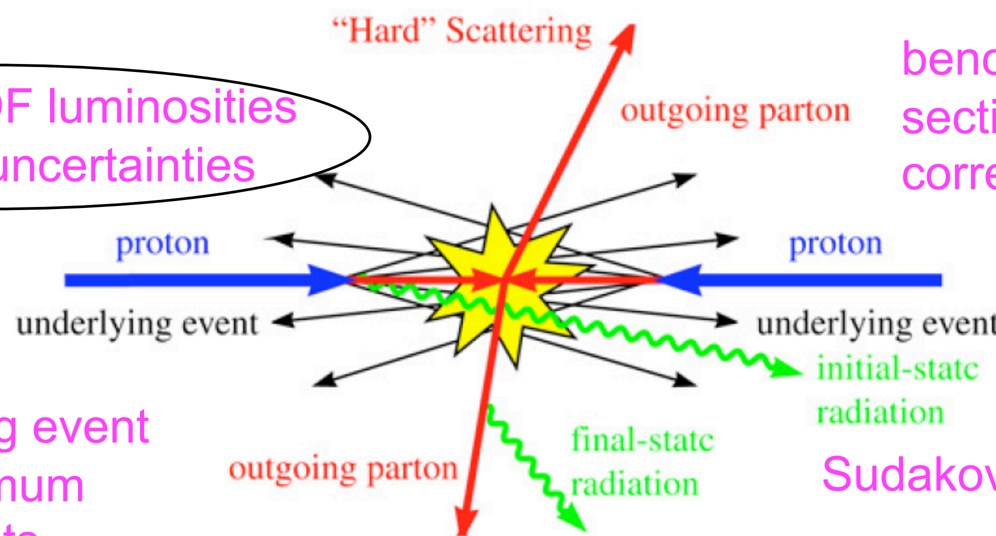
We're covering most of these topics at this summer school.

LO, NLO and NNLO calculations

K-factors

PDF's, PDF luminosities and PDF uncertainties

benchmark cross sections and pdf correlations



underlying event and minimum bias events

Sudakov form factors

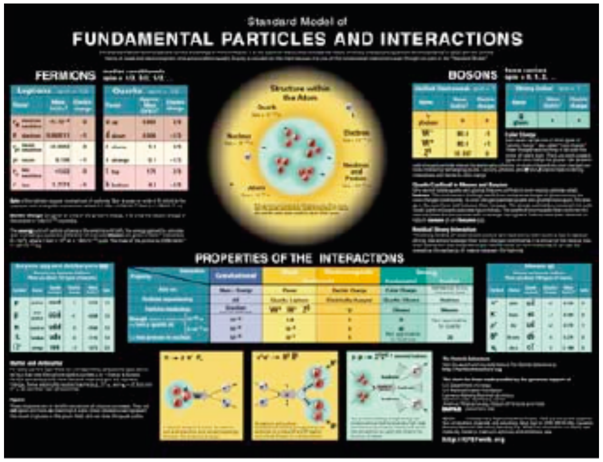
jet algorithms and jet reconstruction

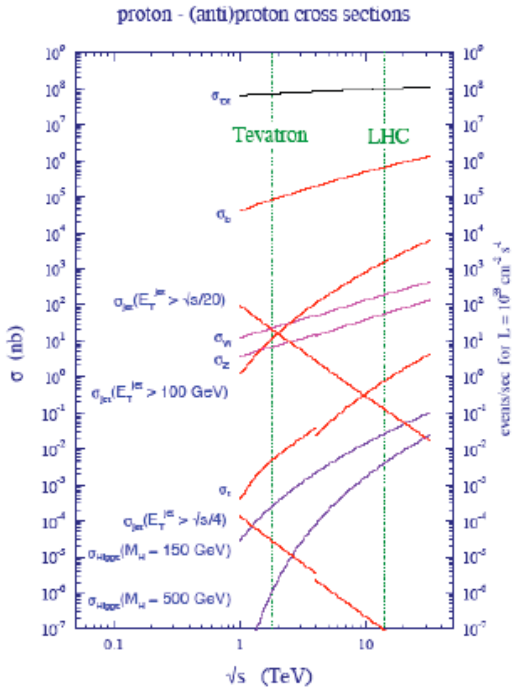
CTEQ

MCnet

Understanding cross sections at the LHC

- We're all looking for BSM physics at the LHC
- Before we publish BSM discoveries from the early running of the LHC, we want to make sure that we measure/understand SM cross sections
  - detector and reconstruction algorithms operating properly
  - SM physics understood properly
  - SM backgrounds to BSM physics correctly taken into account
  - and in particular (for these lectures at least) that pdf's and pdf uncertainties are understood properly





Parton distribution functions and global fits

- Calculation of production cross sections at the LHC relies upon knowledge of pdf's in the relevant kinematic region
- Pdf's 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 ->MRST2003->MRST2004 ->MSTW2008
  - CTEQ->CTEQ5->CTEQ6 ->CTEQ6.1->CTEQ6.5 ->CTEQ6.6

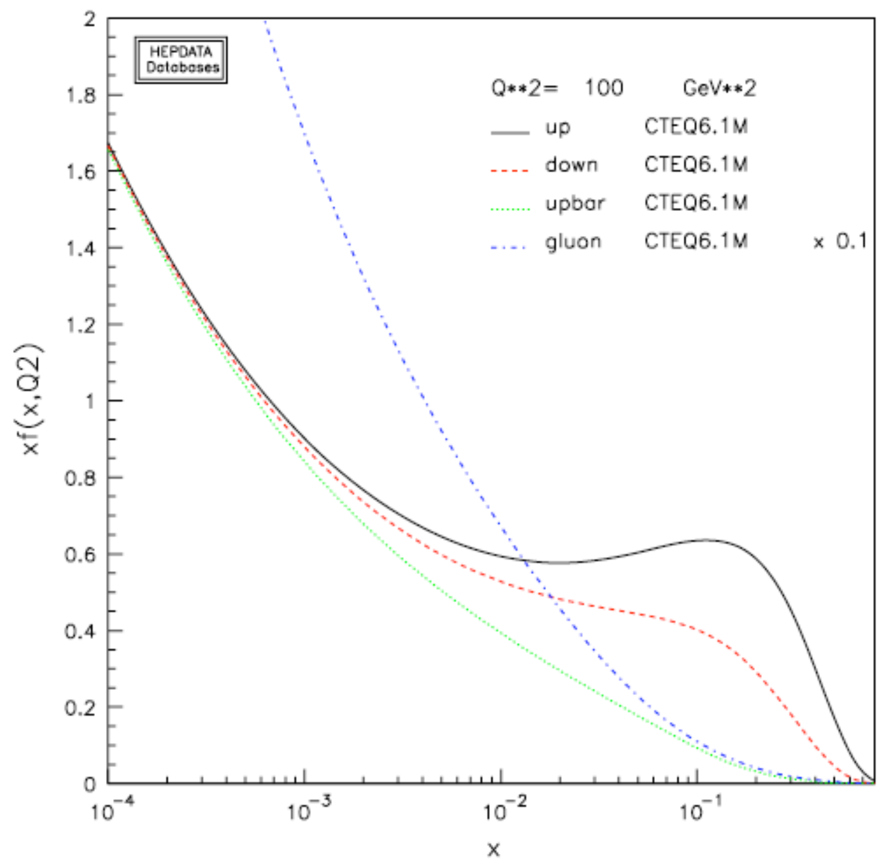


Figure 27. The CTEQ6.1 parton distribution functions evaluated at a Q of 10 GeV.

Review some basics: Drell Yan

- Consider Drell-Yan production
  - write cross section as

$$\sigma_{AB} = \int dx_a dx_b \; f_{a/A}(x_a) f_{b/B}(x_b) \; \hat{\sigma}_{ab\to X}$$

  - where $X=l^+l^-$
  - and $f_{a/A}(x_a)$ is the probability for parton a from hadron A to have a momentum fraction x_a, i.e. a parton distribution function, or pdf (and similar for $f_{b/B}$)
- Specifically the LO or Born term is shown on the lower right
  - note that the hard-scattering subprocess does not depend on any scale (at this point)

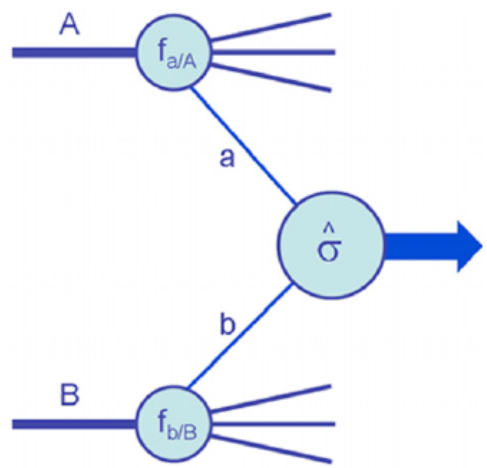
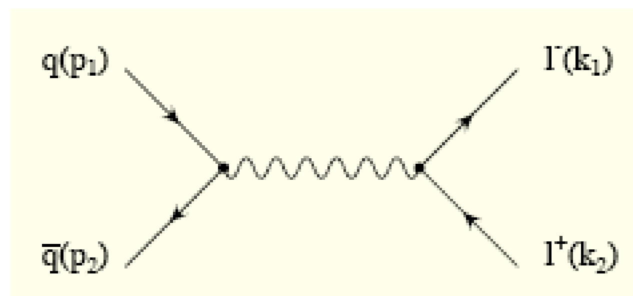


Figure 1. Diagrammatic structure of a generic hard-scattering process.

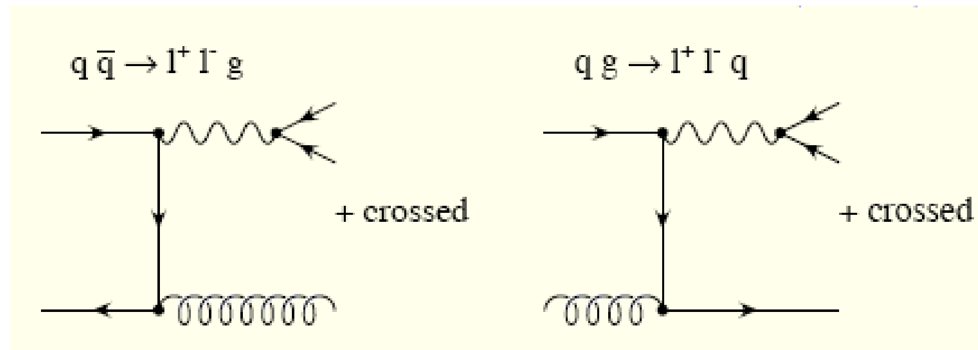


CTEQ

MCnet

Review some basics: Drell Yan

- At NLO, the lepton pair recoils against a quark or gluon
- Integrating over the transverse momentum of the recoiling parton generates logarithmic corrections originating from soft and collinear divergences
  - soft divergences cancel against contributions coming from virtual corrections
  - collinear divergences create logarithms that are the same as those in structure function calculations and thus can be absorbed, via DGLAP equations, in definition of parton distributions, giving rise to logarithmic violations of scaling
  - the pdf's (and σ-hat) now depend on the hardness Q of the process



$$\sigma_{AB} = \int dx_a dx_b \; f_{a/A}(x_a, Q^2)\, f_{b/B}(x_b, Q^2)\, \hat{\sigma}_{ab\to X}$$

...but

- Key point is that all logarithms appearing in Drell-Yan corrections can be factored into renormalized (universal) parton distributions
  - factorization
- But finite corrections left behind after the logarithms are not universal and have to be calculated separately for each process, giving rise to order α_s^n perturbative corrections
- So now we can write the cross section as

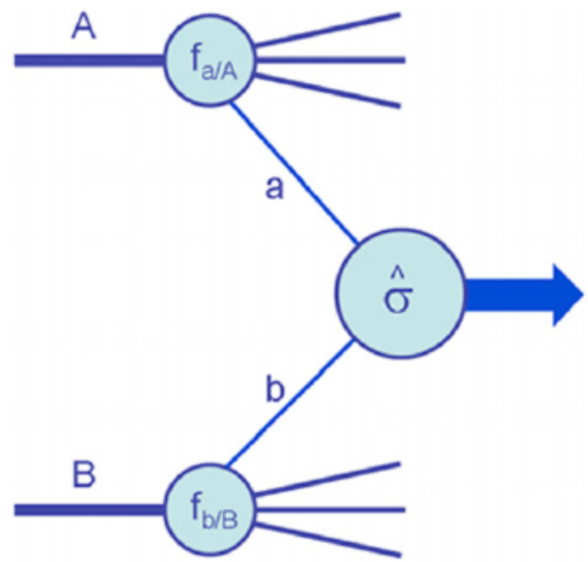


Figure 1. Diagrammatic structure of a generic hard-scattering process.

$$\sigma_{AB} = \int \mathrm{d}x_a \mathrm{d}x_b \; f_{a/A}(x_a, \mu_F^2) \; f_{b/B}(x_b, \mu_F^2) \; \times \; [\, \hat{\sigma}_0 \; + \; \alpha_S(\mu_R^2)\, \hat{\sigma}_1 \; + \; \cdots \,]_{ab \to X}.$$

- where μ_F is the factorization scale (separates long and short-distance physics) and μ_R is the renormalization scale for α_s
- nominally, they can be different but are usually chosen to be the same

An all-orders cross section has no dependence on μ_F and μ_R; a residual dependence remains (to order α_s^{n+1}) for a finite order (α_s^n) calculation

CTEQ

MCnet

DGLAP equations

- Parton distributions used in hard-scattering calculations are solutions of DGLAP equations (or in Italy the AP equations)
  - the DGLAP equations determine the scale dependence of the pdf's

DGLAP equations sum leading powers of $[\alpha_s \log\mu^2]^n$ generated by multiple gluon emission in a region of phase space where the gluons are strongly ordered in transverse momentum ($\log \mu >> \log (1/x)$)

$$\frac{\partial q_i(x,\mu^2)}{\partial \log \mu^2} = \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}\left\{P_{q_iq_j}(z,\alpha_S)q_j(\frac{x}{z},\mu^2) + P_{q_ig}(z,\alpha_S)\ g(\frac{x}{z},\mu^2)\right\},$$

$$\frac{\partial g(x,\mu^2)}{\partial \log \mu^2} = \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}\left\{P_{gq_j}(z,\alpha_S)q_j(\frac{x}{z},\mu^2) + P_{gg}(z,\alpha_S)g(\frac{x}{z},\mu^2)\right\},$$

  - the splitting functions have the perturbative expansions

$$P_{ab}(x,\alpha_S) = P_{ab}^{(0)}(x) + \frac{\alpha_S}{2\pi}P_{ab}^{(1)}(x) + \cdots$$

Thus, a LO calculation will contain σ_o and $P_{ab}^{(0)}$ (with a 1-loop α_s), a NLO calculation will contain, in addition, σ_1 and $P_{ab}^{(1)}$ (with a 2-loop a_s),a NNLO calculation will contain,in addition, σ_o and $P_{ab}^{(2)}$(with a 3-loop α_s)…

CTEQ

MCnet

Back to global fits

- With the DGLAP equations, we know how to evolve pdf's from a starting scale Q_0 to any higher scale
- …but we can't calculate what the pdf's are ab initio
  - one of the goals of lattice QCD
- We have to determine them from a global fit to data
  - factorization theorem tells us that pdf's determined for one process are applicable to another

- So what do we need
  - a value of Q_o (1.3 GeV for CTEQ, 1 GeV for MSTW) lower than the data used in the fit (or any prediction)
  - a parametrization for the pdf's
  - a scheme for the pdf's
  - hard-scattering calculations at the order being considered in the fit
  - pdf evolution at the order being considered in the fit
  - a world average value for α_s
  - a lot of data
    - with appropriate kinematic cuts
  - a treatment of the errors for the experimental data
  - MINUIT

CTEQ

MCnet

Back to global fits

- Parametrization: initial form
  - $f(x) \sim x^{\alpha}(1-x)^{\beta}$
  - estimate β from quark counting rules
    - $\beta = 2n_s - 1$ with n_s being the minimum number of spectator quarks
    - so for valence quarks in a proton (qqq), $n_s=2, \beta=3$
    - for gluon in a proton (qqqg), $n_s=3$, $\beta=5$
    - for anti-quarks in a proton (qqqqqbar), $n_s=4$, $\beta=7$
  - estimate α from Regge arguments
    - gluons and anti-quarks have $\alpha \sim -1$ while valence quarks have $\alpha \sim = 1/2$
  - but at what Q value are these arguments valid?

- What do we know?

1. we know that the sum of the momentum of all partons in the proton is 1 (but see later for modified LO fits)
2. we know the sum of valence quarks is 3
   - and 2 of them are up quarks and 1 of them is a down quark
   - we know that the net number of anti-quarks is 0, but what about dbar=ubar
3. we know that the net number of strange quarks (charm quarks/ bottom quarks) in the proton is 0
   - but we don't know if s=sbar locally

This already puts a lot of restrictions on the pdf's

Orders and Schemes

- Fits are available at
  - LO
    - CTEQ6L or CTEQ6L1
      - 1 loop or 2 loop α_s
    - in common use with parton shower Monte Carlos
    - poor fit to data due to deficiencies of LO ME's
  - LO*
    - better for parton shower Monte Carlos (see later)
  - NLO
    - CTEQ6.1 or CTEQ6.6
    - precision level: error pdf's defined at this order
  - NNLO
    - more accurate but not all processes known

- At NLO and NNLO, one needs to specify a scheme or convention for subtracting the divergent terms
- Basically the scheme specifies how much of the finite corrections to subtract along with the divergent pieces
  - most widely used is the modified minimal subtraction scheme (or MSbar)
  - used with dimensional regularization: subtract the pole terms and accompanying log 4π and Euler constant terms
  - also may find pdf's in DIS scheme, ,where full order α_s correction for F_2 in DIS absorbed into quark pdf's

CTEQ

MCnet

Scales and Masses

- Processes used in global fits are characterized by a single large scale
  - DIS-Q^2
  - lepton pair production-M^2
  - vector boson production-M_V^2
  - jet production-p_T^{jet}
- By choosing the factorization and renormalization scales to be of the same order as the characteristic scale
  - can avoid some large logarithms in the hard scattering cross section
  - some large logarithms in running coupling and pdf's are resummed
- Different treatment of quark masses and thresholds
  - zero mass variable flavor number scheme (ZM-VFNS)
  - fixed flavor number scheme (FFNS)
  - variable flavor number scheme (VFNS)

Zero mass variable flavor number scheme (ZM-VFNS)

- Start pdf evolution at charm threshold ($Q=m_c$=1.3 GeV)
  - set c and b distributions to zero at this scale (although can allow for possibility of intrinsic charm/bottom)
    - start b evolution at $Q=m_b$
  - all heavy quarks treated as massless
  - c and b pairs created by gluon splitting
  - adjust running coupling α_s as each flavor threshold is crossed since QCD β function depends on # of active flavors
  - in this approach, only mass effects are due to flavor thresholds and changing of β function
- Most commonly used CTEQ NLO pdf's prior to CTEQ6.5 (such as CTEQ6M, CTEQ6.1) are of this type
- Advantages
  - easy to implement
  - sums large logs of Q^2/m_Q^2 via DLGAP equation
  - asymptotically correct when $Q^2 >> m_Q^2$
- Disadvantages
  - does not treat heavy quark threshold correctly

CTEQ

MCnet

Fixed flavor number scheme

- Calculate heavy quark production from relevant subprocesses such as γ^*g->QQbar keeping only light quarks in DGLAP equations
- Only light quarks have pdf's
  - no charm or bottom quark pdf's
- Advantage
  - gets threshold behavior correct
- Disadvantage
  - does not resum potentially large logs of Q^2/m_Q^2

CTEQ

Variable flavor number scheme (VFNS)

MCnet

- This is the “just right” scheme
- It combines the ZM-VFNS and FFNS by interpolating between the FFNS (correct near threshold) and the ZM-FFNS (resums large logs)
- But it’s technically more complicated than the ZM-VFNS since there must be subtraction terms in order to avoid large logarithms
- All current (CTEQ6.6) and future NLO CTEQ pdf’s will be of this type
- Its use has an impact on predictions for the LHC

Data sets used in global fits (CTEQ6.6)

1. BCDMS F_2^{proton} (339 data points)
2. BCDMS $F_2^{deuteron}$ (251 data points)
3. NMC F_2 (201 data points)
4. NMC F_2^d/F_2^p (123 data points)
5. F_2(CDHSW) (85 data points)
6. F_3(CDHSW) (96 data points)
7. CCFR F_2 (69 data points)
8. CCFR F_3 (86 data points)
9. H1 NC e^-p (126 data points; 1998-98 reduced cross section)
10. H1 NC e^-p (13 data points; high y analysis)
11. H1 NC e^+p (115 data points; reduced cross section 1996-97)
12. H1 NC e^+p (147 data points; reduced cross section; 1999-00)
13. ZEUS NC e^-p (92 data points; 1998-99)
14. ZEUS NC e^+p (227 data points; 1996-97)
15. ZEUS NC e^+p (90 data points; 1999-00)
16. H1 F_2^c e^+p (8 data points;1996-97)
17. H1 $R\sigma^c$ for ccbar e^+p (10 data points;1996-97)
18. H1 R_σ^b for bbbar e^+p (10 data points; 1999-00)
19. ZEUS F_2^c e^+p (18 data points; 1996/97)
20. ZEUS F_2^C e+p (27 data points; 1998/00)
21. H1 CC e^-p (28 data points; 1998-99)
22. H1 CC e^+p (25 data points; 1994-97)
23. H1 CC e^+p (28 data points; 1999-00)
24. ZEUS CC e^-p (26 data points; 1998-99)
25. ZEUS CC e^+p (29 data points; 1994-97)
26. ZEUS CC e^+p (30 data points; 1999-00)
27. NuTev neutrino dimuon cross section (38 data points)
28. NuTev anti-neutrino dimuon cross section (33 data points)
29. CCFR neutrino dimuon cross section (40 data points)
30. CCFR anti-neutrino cross section (38 data points)
31. E605 dimuon (199 data points)
32. E866 dimuon (13 data points)
33. Lepton asymmetry from CDF (11 data points)
34. CDF Run 1B jet cross section (33 data points)
35. D0 Run 1B jet cross section (90 data points)

- 2794 data points from DIS, DY, jet production
- All with (correlated) systematic errors that must be treated correctly in the fit
- Note that DIS is the 800 pound gorilla of the global fit with many data points and small statistical and systematic errors
  - and fixed target DIS data still have a significant impact on the global fitting, even with an abundance of HERA data
- To avoid non-perturbative effects, kinematic cuts on placed on the DIS data
  - $Q^2>5$ GeV^2
  - $W^2(=m^2+Q^2(1-x)/x)>12.25$ GeV^2

Influence of data in global fit

- Charged lepton DIS (see Mandy's lectures)

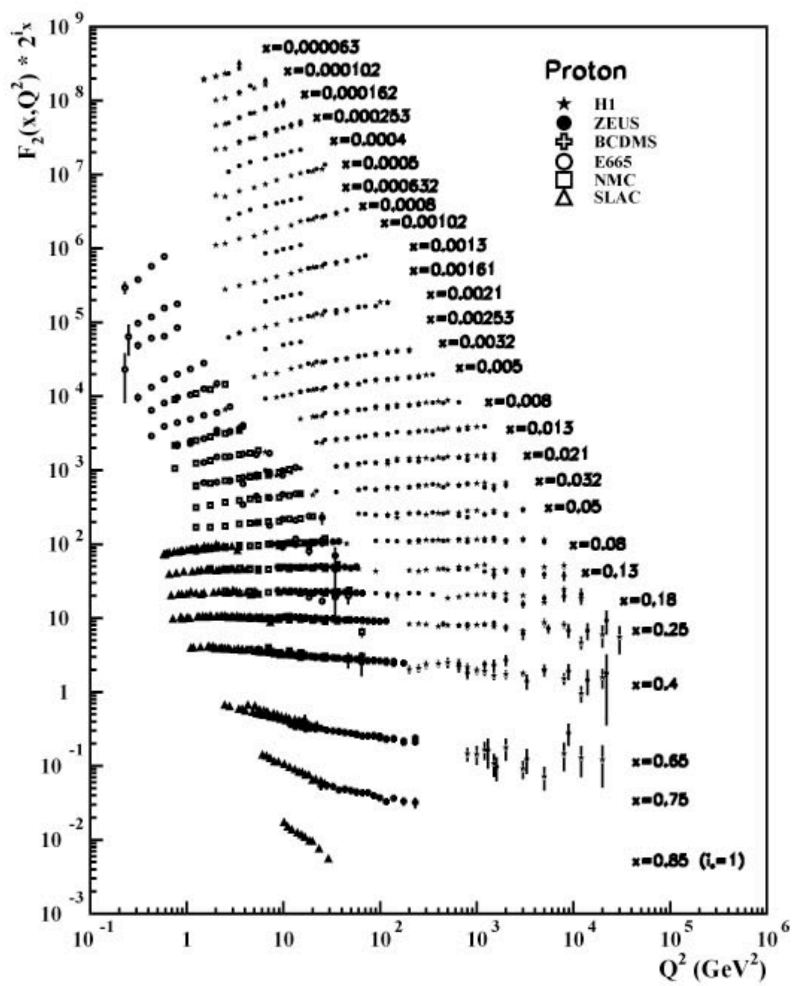
$$F_2(x,Q^2) = x\sum_i e_i^{\,2}[q_i(x,Q^2) + \bar{q}_i(x,Q^2)]$$

  - each flavor weighted by its squared charge
  - quarks and anti-quarks enter together
  - gluon doesn't enter, in lowest order, but does enter into the structure functions at NLO
  - also enters through mixing in evolution equations so gluon contributes to the change of the structure functions as Q^2 increases
  - at low values of x

$$Q^2\frac{dF_2}{dQ^2} \approx \frac{\alpha_s}{2\pi}\sum e_i^{\,2}\int_x^1 \frac{dy}{y}P_{qg}(y)G(\frac{x}{y},Q^2)$$

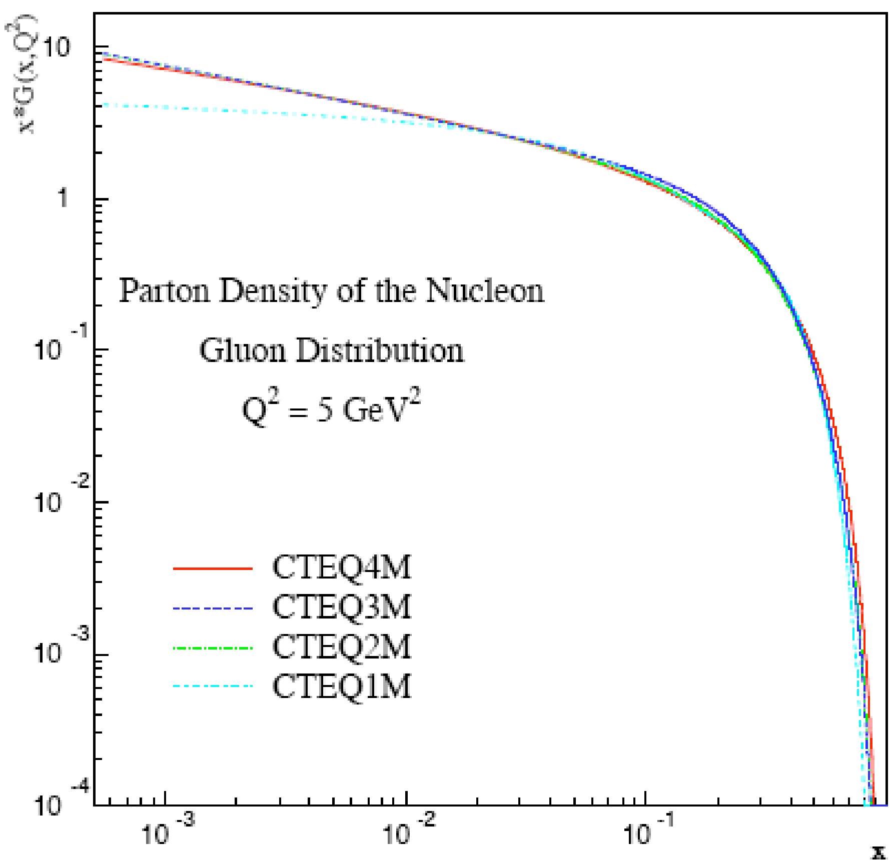
  - Q^2 dependence at small x is driven directly by gluon pdf

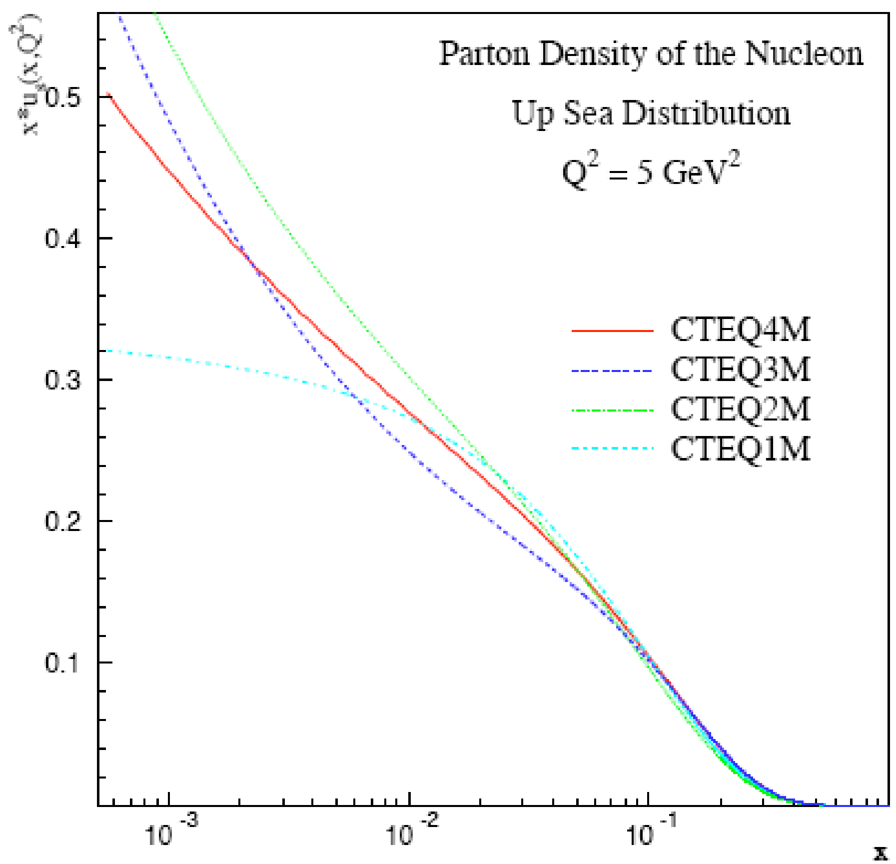
- At low x, structure functions increase with Q^2; at high x decrease



Some history

What caused the changes from CTEQ1 to CTEQ2 (and up)?
Infusion of low x HERA data. Before was just an extrapolation and *guess* was wrong.





CTEQ

MCnet

Neutrino DIS

$$F_2(x,Q^2) = x\sum_i [q_i(x,Q^2) + \overline{q}_i(x,Q^2)]$$

$$xF_3(x,Q^2) = x\sum_i [q_i(x,Q^2) - \overline{q}_i(x,Q^2)]$$

- additional structure function allows the separation of quarks and anti-quarks but not a complete flavor separation
- caveat: neutrino observables usually obtained using nuclear targets so there is added question of nuclear corrections

CTEQ

MCnet

Some observations from DIS

- DIS data provide strong constraints on the u and d distributions over the full range of x covered by the data
- The combination 4*ubar + dbar is well-constrained at small x
- The gluon is constrained at low values of x by the slope of the Q^2 dependence of F_2
  - momentum sum rule connects low x and high x behavior, but loosely

CTEQ

MCnet

dbar/ubar and Gottfried sum rule

$$S_G = \int_0^1 \frac{dx}{x}\left[F_2^{\,p}(x) - F_2^{\,n}(x)\right]$$

$$= \frac{1}{3} + \frac{2}{3}\int_0^1 dx\left(\bar{u}(x) - \bar{d}(x)\right)$$

$$= 0.235 \pm 0.026$$

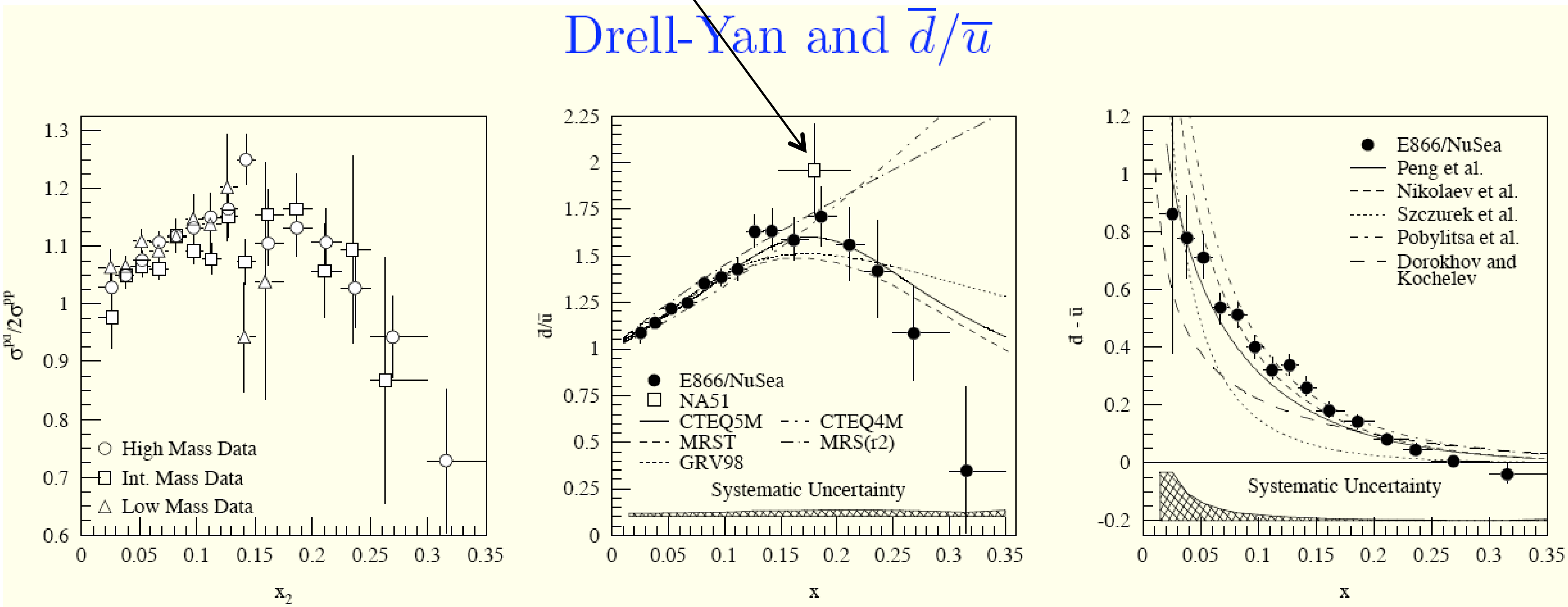
$$\bar{d} \neq \bar{u}$$

- Doesn’t tell us the x dependence though

- Why is $\bar{u} \neq \bar{d}$?
- Pion cloud argument
  - proton can fluctuate into a neutron and a positive pion
  - p->nπ^+->p
  - …or uud->(udd)($u\bar{d}$)
  - …or $\bar{d}>\bar{u}$
  - …so SU(2) symmetry of sea quarks is broken

Information from Drell-Yan

Note NA-51: only one data point but provided crucial information on dbar/ubar before E866



What about s and $\bar{s}$?

- Can get information from

$$W^+ s \rightarrow c$$

$$W^- \bar{s} \rightarrow \bar{c}$$

- Look from muon pairs in final state due to charm hadrons decaying semi-leptonically

$$c \rightarrow s\mu^+\nu$$

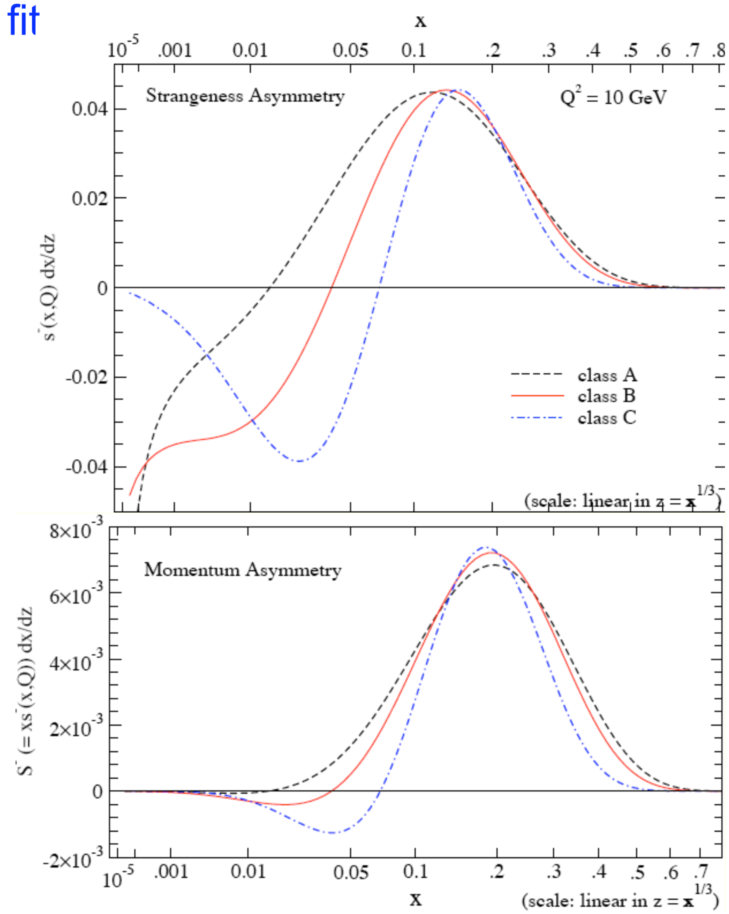
$$\bar{c} \rightarrow \bar{s}\mu^-\bar{\nu}$$

- Information from dimuon production in neutrino interactions

$$\nu N \rightarrow \mu^- c + X' \rightarrow \mu^-\mu^+ + X$$

$$\bar{\nu} N \rightarrow \mu^+ \bar{c} + X' \rightarrow \mu^+\mu^- + X$$

- So s carries somewhat more momentum than $\bar{s}$
- In previous fits, assumption was that s=$\bar{s}$; in CTEQ6.6 fit remove that assumption->2 new free parameters the fit



Inclusive jets and global fits

- We don't have many handles on the high x gluon distribution in the global pdf fits
- Best handle is provided by the inclusive jet cross section from the Tevatron

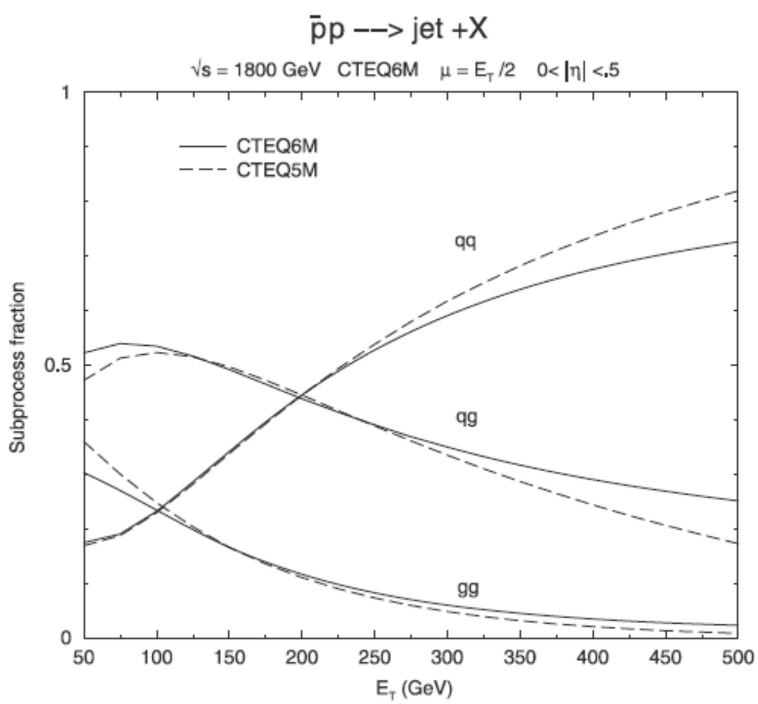


Figure 56. The subprocess contributions to inclusive jet production at the Tevatron for the CTEQ5M and CTEQ6M pdfs. The impact of the larger larger gluon at high x for CTEQ6 is evident.

- At high E_T (high x), gq is subdominant, but there's a great deal of freedom/uncertainty on the high x gluon distribution
  - about 42% of the proton's momentum is carried by gluons, and most of that momentum is at low x

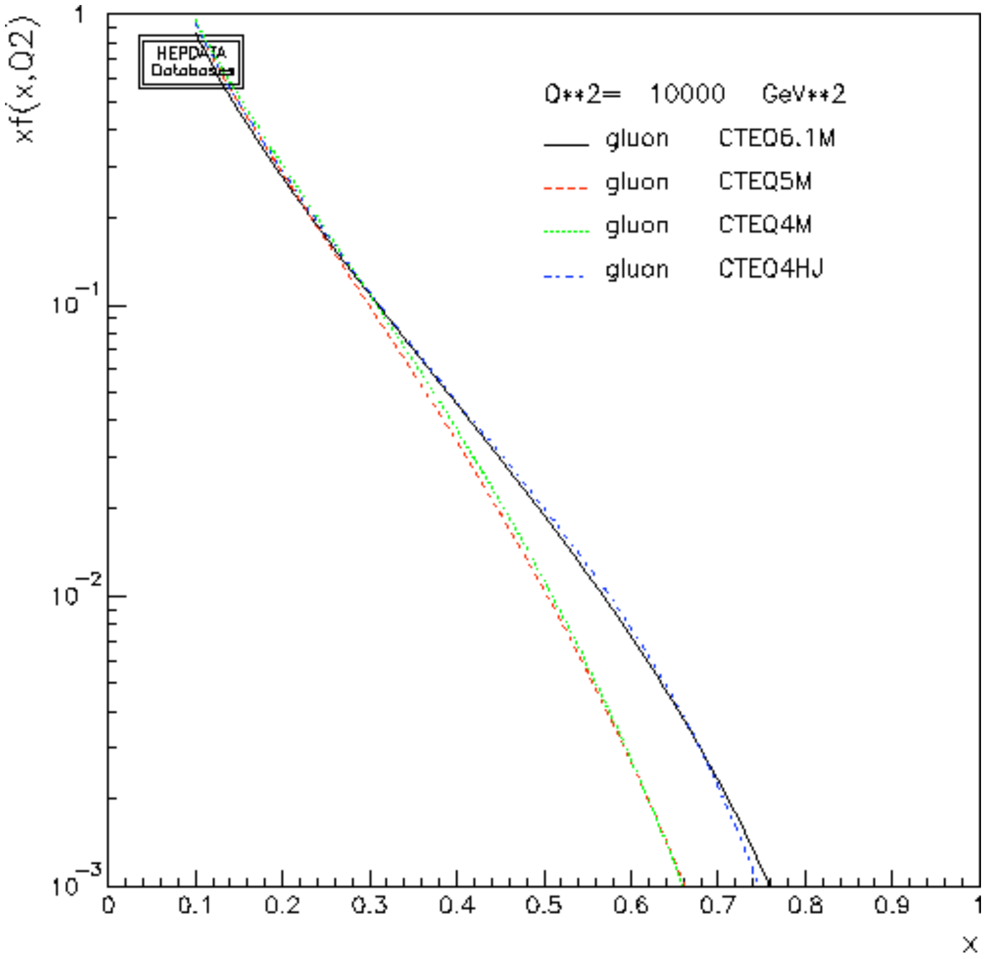
X Bin	Momentum fraction
10^{-4} to 10^{-3}	0.6%
10^{-3} to 0.01	3%
0.01 to 0.1	16%
0.1 to 0.2	10%
0.2 to 0.3	6%
0.3 to 0.5	5%
0.5 to 1.0	1%

TABLE I. The momentum fraction carried by gluons in a a given x bin at a Q value of 5 GeV.

- The inclusion of the CDF/D0 inclusive jet cross sections from Run 1 boosted the high x gluon distribution and thus the predictions for the high E_T jet cross sections

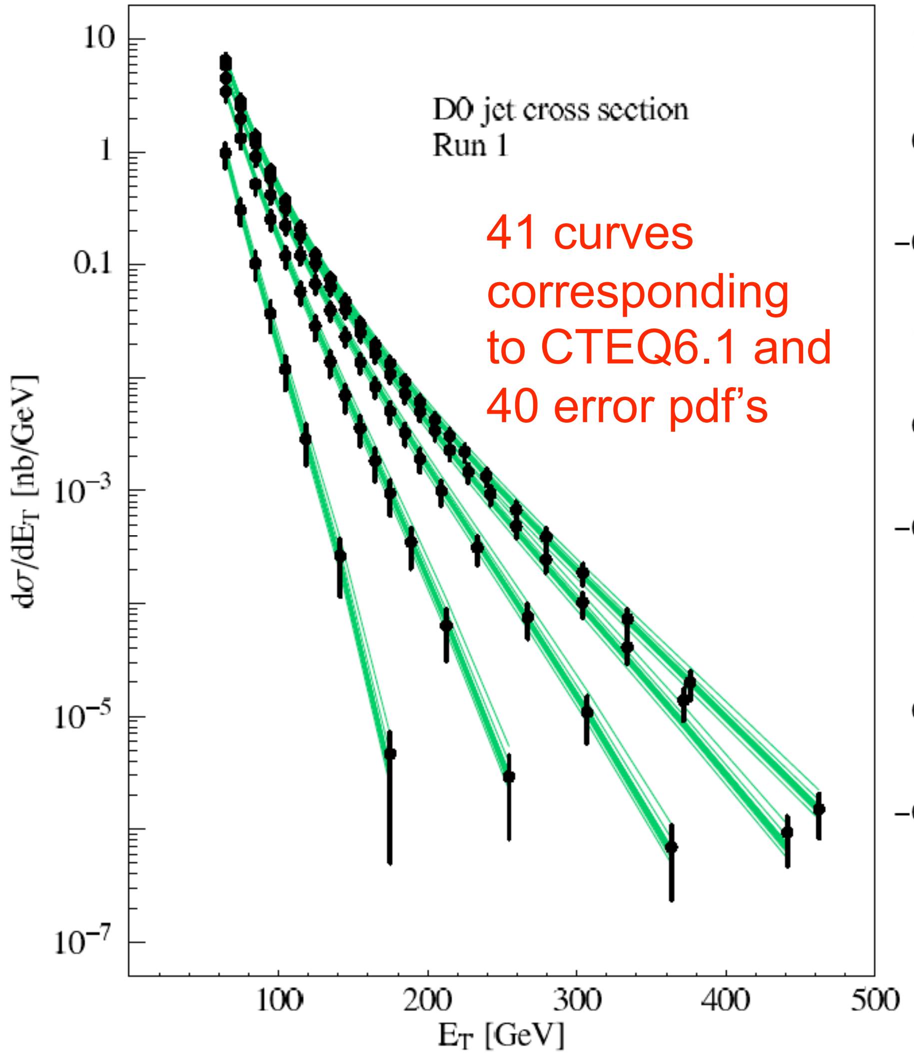
Some more history

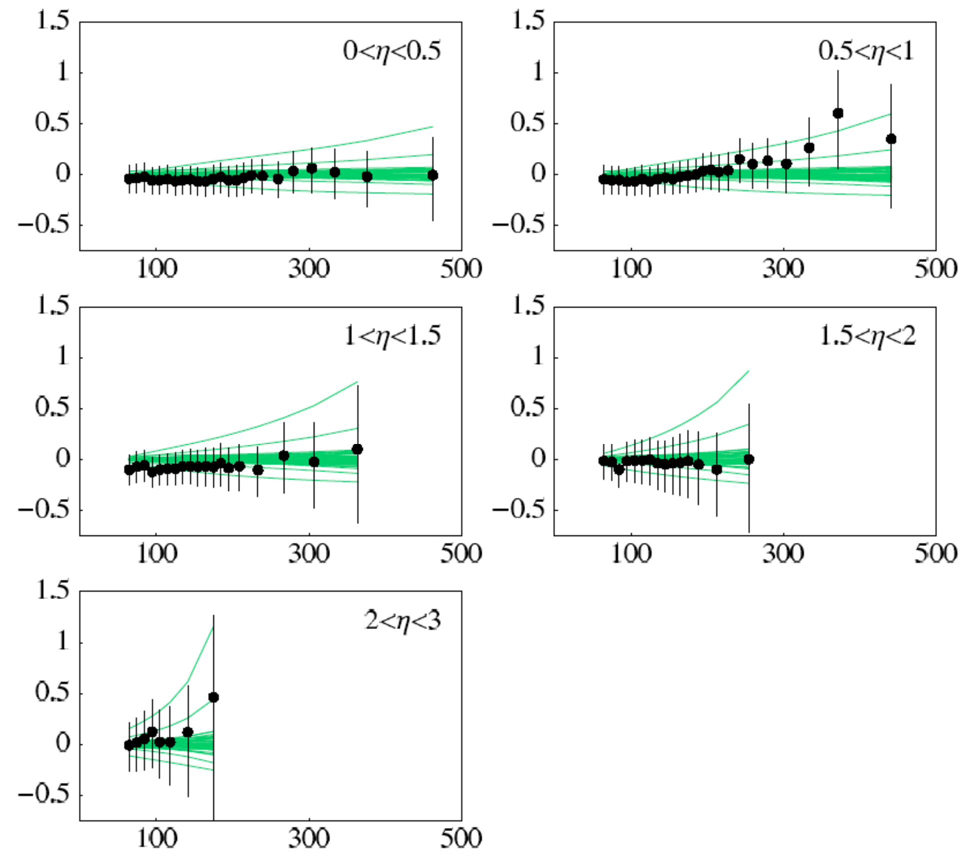
- Note that the high x gluon for CTEQ6.1 is much larger than that for either CTEQ4M or CTEQ5M. Why?
- Full inclusion of Run 1 jet data (especially from D0) which preferred to have a larger gluon
  - similar to the hypothesis of CTEQ4HJ
- Caveat: high x gluon is decreasing somewhat with the inclusion of the Run 2 jet data, which don't prefer as large of a high x gluon



CTEQ

Inclusive jets at the Tevatron





predictions using CTEQ6.1

CTEQ

MCnet

Global fitting: best fit

- Using our 2794 data points, we do our global fit by performing a χ^2 minimization
  - where D_i are the data points and T_i are the theoretical predictions; we allow for a normalization shift f_N for each experimental data set
    - but we provide a quadratic penalty for any normalization shift
  - where there are k systematic errors β for each data point in a particular data set
    - and where we allow the data points to be shifted by the systematic errors with the shifts given by the s_j parameters
    - but we give a quadratic penalty for non-zero values of the shifts s_j
  - where σ_i is the statistical error for data point i

- For each data set, we calculate

$$\chi^2 = \sum_i \frac{\left[\left(f_N D_i - \sum_{j=1}^{k} \beta_{ij} s_j\right) - T_i\right]^2}{\sigma_i^2} + \sum_{j=1}^{k} s_j^2$$

- For a set of theory parameters it is possible to analytically solve for the shifts s_j,and therefore, continually update them as the fit proceeds
- To make matters more complicated, we may give additional weights to some experiments due to the utility of the data in those experiments (i.e. NA-51), so we adjust the χ^2 to be

$$\chi^2 = \sum_k w_k \chi_k^2 + \sum_k w_{N,k} \left[\frac{1-f_N}{\sigma_N^{norm}}\right]^2$$

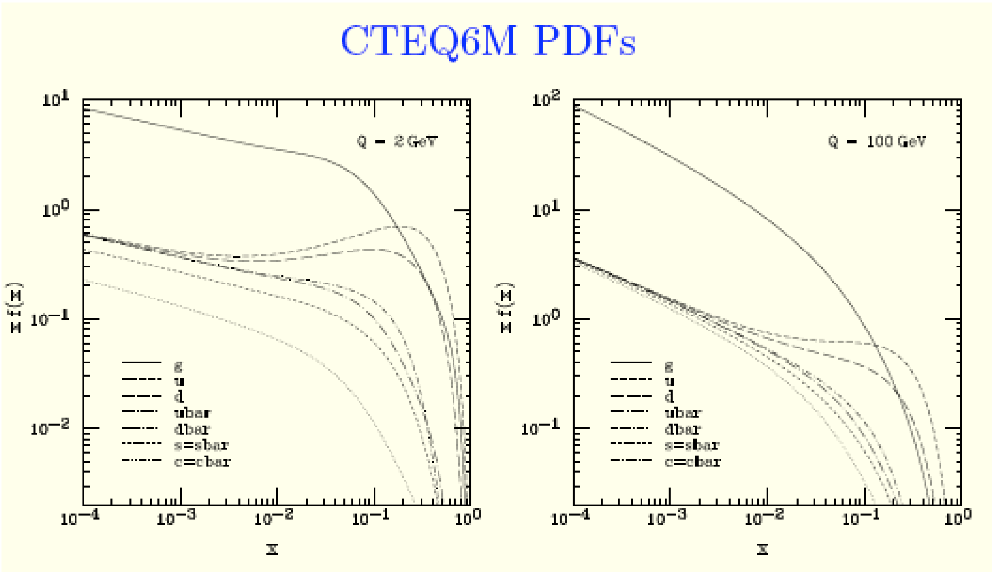
- where w_k is a weight given to the experimental data and $w_{N,k}$ is a weight given to the normalization

Minimization and errors

- Free parameters in the fit are parameters for quark and gluon distributions

$$f(x) = x^{(a_1-1)}(1-x)^{a_2} e^{a_3 x}[1+e^{a_4} x]^{a_5}$$

- Too many parameters to allow all to remain free
  - some are fixed at reasonable values or determined by sum rules
- 20 free parameters for CTEQ6.1, 22 for CTEQ6.6
  - 2 additional parameters for strange quark distributions
- Result is a global χ^2/dof on the order of 1
  - for a NLO fit
  - worse for a LO fit, since the LO pdf's can not make up for the deficiencies in the LO matrix elements

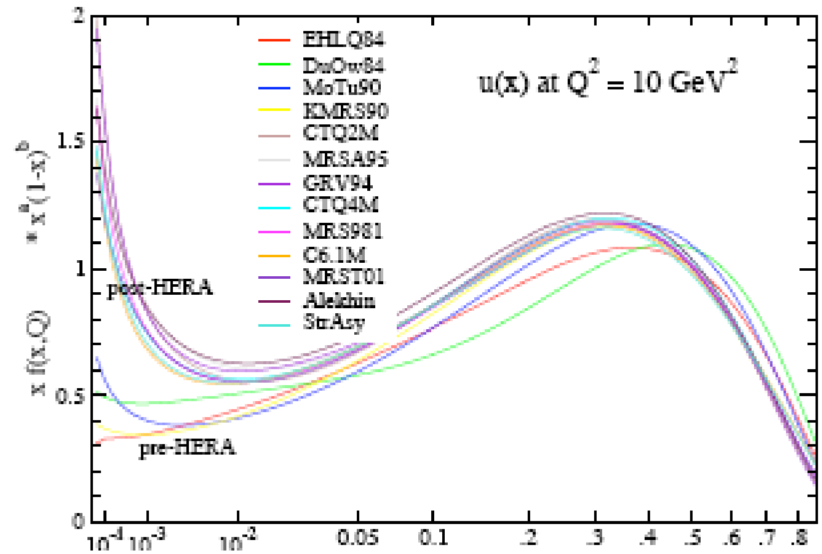


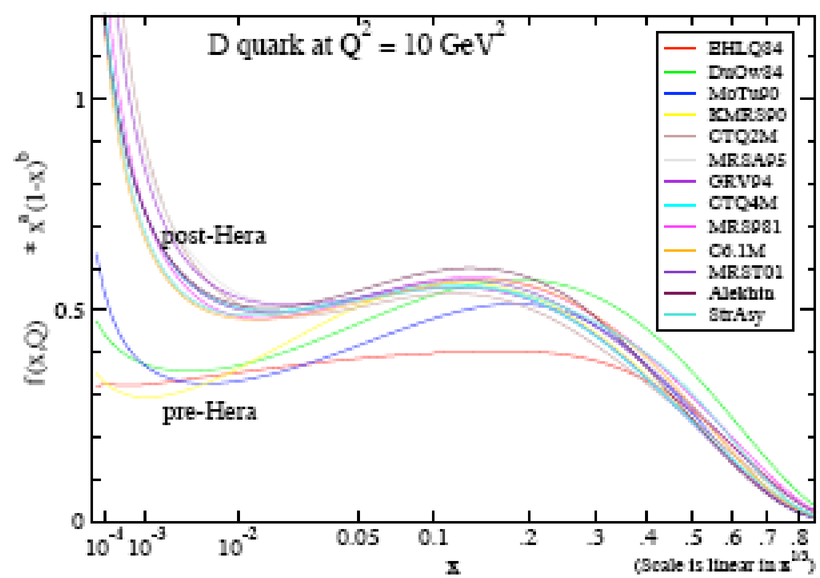
CTEQ

MCnet

PDF Errors: old way

- Make plots of lots of pdf's (no matter how old) and take spread as a measure of the error
- Can either underestimate or overestimate the error
- Review sources of uncertainty on pdf's
  - data set choice
  - kinematic cuts
  - parametrization choices
  - treatment of heavy quarks
  - order of perturbation theory
  - errors on the data
- There are now more sophisticated techniques to deal with at least the errors due to the experimental data uncertainties



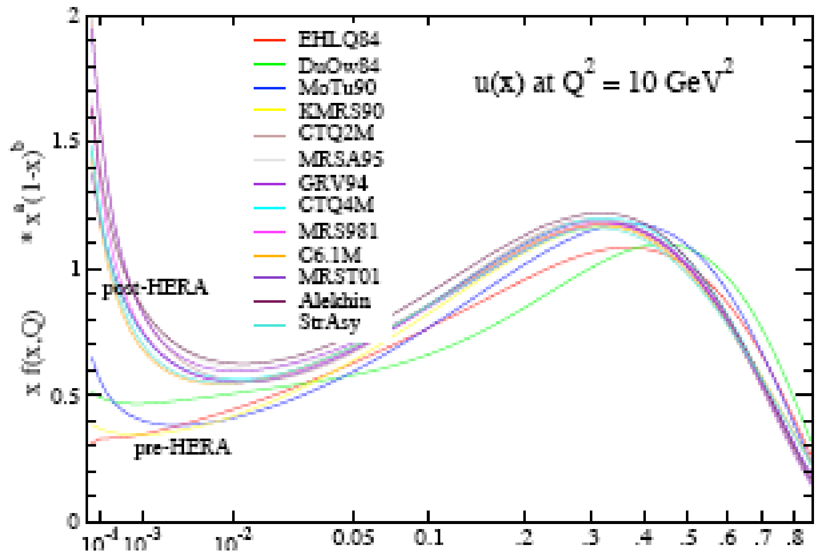


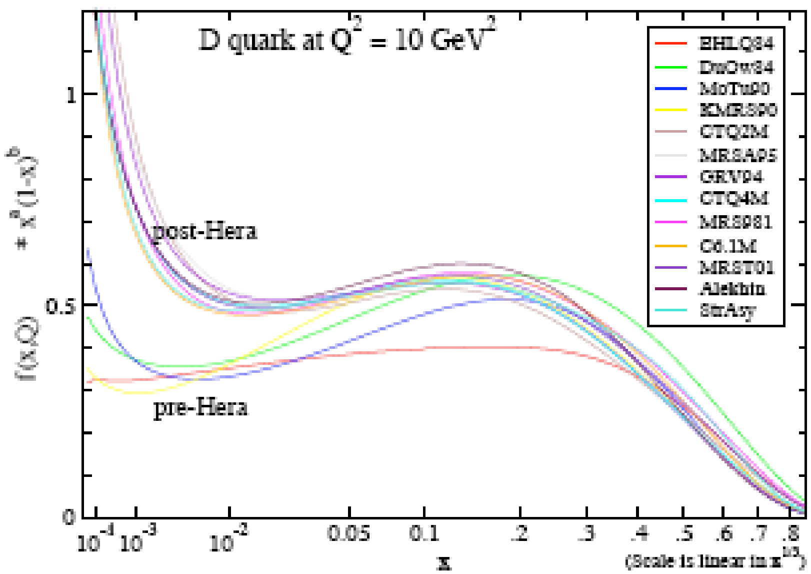
CTEQ

MCnet

PDF Errors: old way

Unlike fine wines, vintage pdf's are to be avoided.





PDF Errors: new way

- So we have optimal values (minimum χ^2) for the d=20 (22) free pdf parameters in the global fit
  - $\{a_\mu\}, \mu=1,\ldots d$
- Varying any of the free parameters from its optimal value will increase the χ^2
- It's much easier to work in an orthonormal eigenvector space determined by diagonalizing the Hessian matrix, determined in the fitting process

$$H_{uv} = \frac{1}{2}\frac{\partial\chi^2}{\partial a_\mu \partial a_\nu}$$

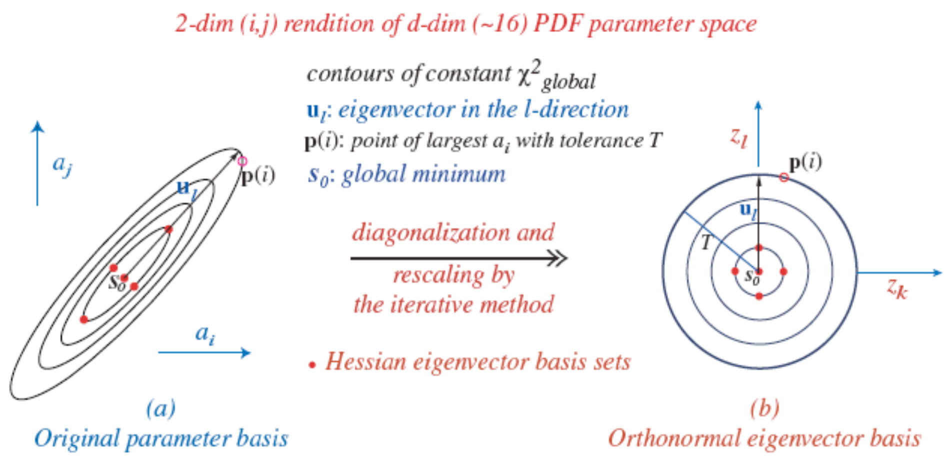


Figure 28. A schematic representation of the transformation from the pdf parameter basis to the orthonormal eigenvector basis.

To estimate the error on an observable X(a), due to the experimental uncertainties of the data used in the fit, we use the *Master Formula*

$$(\Delta X)^2 = \Delta\chi^2 \sum_{\mu,\nu} \frac{\partial X}{\partial a_\mu} \left(H^{-1}\right)_{\mu\nu} \frac{\partial X}{\partial a_\nu}$$

PDF Errors: new way

- Recap: 20 (22) eigenvectors with the eigenvalues having a range of >1E6
- Largest eigenvalues (low number eigenvectors) correspond to best determined directions; smallest eigenvalues (high number eigenvectors) correspond to worst determined directions
- Easiest to use Master Formula in eigenvector basis

$$\Delta X^{+}_{\max} = \sqrt{\sum_{i=1}^{N} [\max(X_i^{+} - X_0,\, X_i^{-} - X_0,\, 0)]^2},$$

$$\Delta X^{-}_{\max} = \sqrt{\sum_{i=1}^{N} [\max(X_0 - X_i^{+},\, X_0 - X_i^{-},\, 0)]^2}.$$

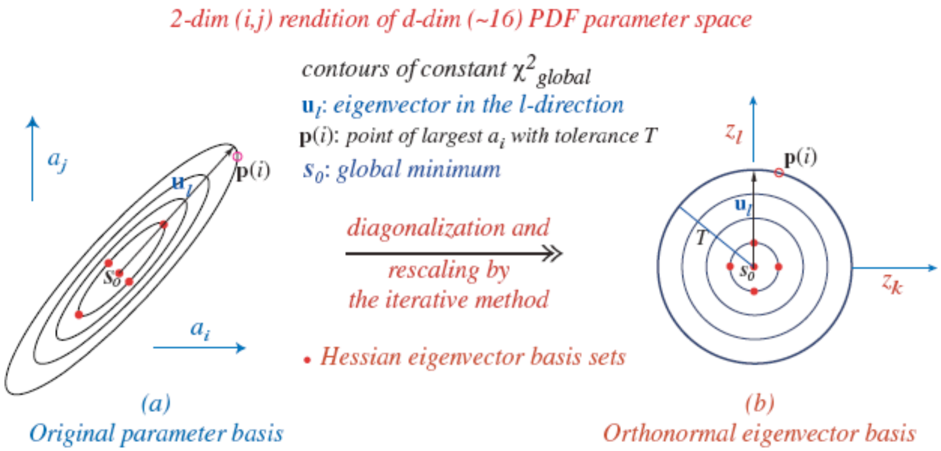


Figure 28. A schematic representation of the transformation from the pdf parameter basis to the orthonormal eigenvector basis.

To estimate the error on an observable X(a), from the experimental errors, we use the *Master Formula*

$$(\Delta X)^2 = \Delta\chi^2 \sum_{\mu,\nu} \frac{\partial X}{\partial a_\mu} \left(H^{-1}\right)_{\mu\nu} \frac{\partial X}{\partial a_\nu}$$

where X_i^+ and X_i^- are the values for the observable X when traversing a distance corresponding to the tolerance T(=sqrt($\Delta\chi^2$)) along the i^{th} direction

C T E Q

MCnet

PDF Errors: new way

- What is the tolerance T?
- This is one of the most controversial questions in global pdf fitting?
- We have 2794 data points in the CTEQ6.6 data set (on order of 2000 for CTEQ6.1)
- Technically speaking, a 1-sigma error corresponds to a tolerance T(=sqrt($\Delta\chi^2$))=1
- This results in far too small an uncertainty from the global fit
  - with data from a variety of processes from a variety of experiments from a variety of accelerators
- For CTQE6.1, we chose a $\Delta\chi^2$ of 100 to correspond to a 90% CL limit
  - with an appropriate scaling for the larger data set for CTEQ6.6
- MSTW has chosen a $\Delta\chi^2$ of 50 for the same limit so CTEQ errors will be larger than MSTW errors

$$\Delta X^{+}_{\max} = \sqrt{\sum_{i=1}^{N} [\max(X_i^+ - X_0, X_i^- - X_0, 0)]^2},$$

$$\Delta X^{-}_{\max} = \sqrt{\sum_{i=1}^{N} [\max(X_0 - X_i^+, X_0 - X_i^-, 0)]^2}.$$

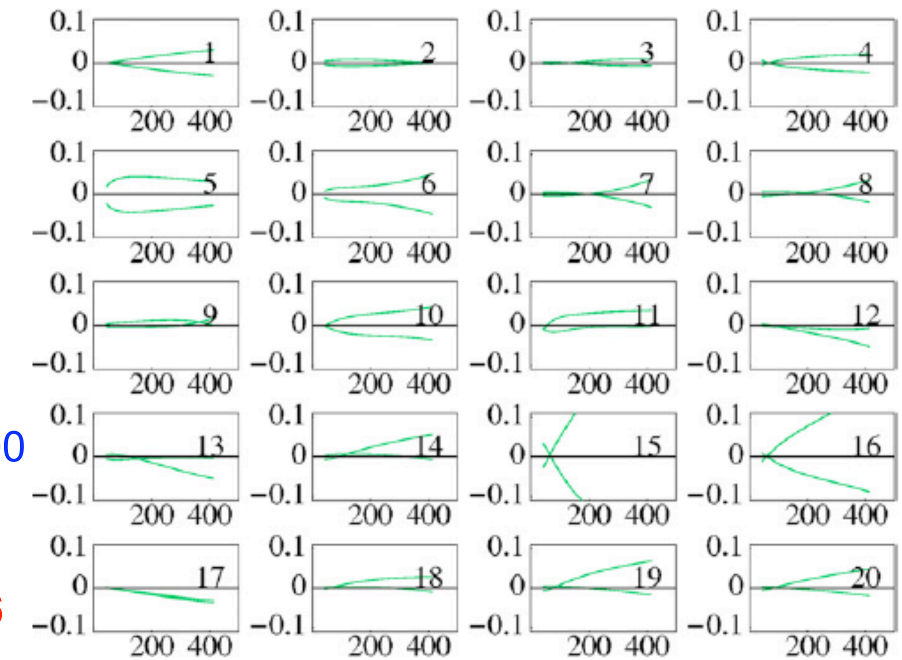


Figure 29. The pdf errors for the CDF inclusive jet cross section in Run 1 for the 20 different eigenvector directions. The vertical axes show the fractional deviation from the central prediction and the horizontal axes the jet transverse momentum in GeV.

CTEQ

MCnet

What do the eigenvectors *mean*?

- Each eigenvector corresponds to a linear combination of all 20 (22) pdf parameters, so in general each eigenvector doesn't mean anything?
- However, with 20 (22) dimensions, often eigenvectors will have a large component from a particular direction
- Take eigenvector 1 (for CTEQ6.1); error pdf's 1 and 2
- It has a large component sensitive to the small x behavior of the u quark valence distribution
- Not surprising since this is one of the best determined directions
  - the W mass pdf uncertainty at the Tevatron is due mainly to eigenvectors 1 and 2

Sets	Shape Parameter	Component
1, 2	BP(2, 1)	0.057911
1, 2	BP(2, 2)	-0.022688
1, 2	BP(2, 3)	0.015496
1, 2	BP(2, 4)	0.035277
1, 2	BP(2, 5)	frozen
1, 2	BP(1, 1)	0.888833
1, 2	BP(1, 2)	-0.161942
1, 2	BP(1, 3)	0.118204
1, 2	BP(1, 4)	0.268405
1, 2	BP(1, 5)	0.276392
1, 2	BP(0, 1)	0.038555
1, 2	BP(0, 2)	-0.006610
1, 2	BP(0, 3)	frozen
1, 2	BP(0, 4)	-0.017717
1, 2	BP(0, 5)	frozen
1, 2	BP(-1, 1)	-0.007668
1, 2	BP(-1, 2)	0.012745
1, 2	BP(-1, 3)	0.001851
1, 2	BP(-1, 4)	frozen
1, 2	BP(-1, 5)	0.001004
1, 2	BP(-2, 1)	0.117517
1, 2	BP(-2, 2)	-0.008357
1, 2	BP(-2, 3)	0.006504
1, 2	BP(-2, 4)	frozen
1, 2	BP(-2, 5)	frozen

Eigenvector 5 (pdf's 9 and 10)

- Low x behavior of the gluon
- Affects W rapidity distribution at the LHC and W mass at the LHC

Sets	Shape Parameter	Component
9, 10	BP(2, 1)	-0.073740
9, 10	BP(2, 2)	-0.001261
9, 10	BP(2, 3)	-0.000242
9, 10	BP(2, 4)	-0.006397
9, 10	BP(2, 5)	frozen
9, 10	BP(1, 1)	0.081017
9, 10	BP(1, 2)	0.143655
9, 10	BP(1, 3)	-0.090463
9, 10	BP(1, 4)	-0.087878
9, 10	BP(1, 5)	-0.152997
9, 10	BP(0, 1)	0.745961
9, 10	BP(0, 2)	-0.085072
9, 10	BP(0, 3)	frozen
9, 10	BP(0, 4)	-0.445408
9, 10	BP(0, 5)	frozen
9, 10	BP(-1, 1)	-0.099854
9, 10	BP(-1, 2)	0.189535
9, 10	BP(-1, 3)	0.017860
9, 10	BP(-1, 4)	frozen
9, 10	BP(-1, 5)	0.021630
9, 10	BP(-2, 1)	-0.345311
9, 10	BP(-2, 2)	0.001261
9, 10	BP(-2, 3)	-0.000159
9, 10	BP(-2, 4)	frozen
9, 10	BP(-2, 5)	frozen

C T E Q

What do the eigenvectors *mean*?

MCnet

- Take eigenvector 8 (for CTEQ6.1); error pdf's 15 and 16
- No particular direction stands out

Sets	Shape Parameter	Component
15, 16	BP(2, 1)	0.196388
15, 16	BP(2, 2)	0.387704
15, 16	BP(2, 3)	-0.226202
15, 16	BP(2, 4)	-0.411440
15, 16	BP(2, 5)	frozen
15, 16	BP(1, 1)	-0.193195
15, 16	BP(1, 2)	0.356604
15, 16	BP(1, 3)	0.018064
15, 16	BP(1, 4)	0.468888
15, 16	BP(1, 5)	0.376180
15, 16	BP(0, 1)	0.016734
15, 16	BP(0, 2)	-0.026136
15, 16	BP(0, 3)	frozen
15, 16	BP(0, 4)	-0.016537
15, 16	BP(0, 5)	frozen
15, 16	BP(-1, 1)	-0.176169
15, 16	BP(-1, 2)	0.136337
15, 16	BP(-1, 3)	0.074431
15, 16	BP(-1, 4)	frozen
15, 16	BP(-1, 5)	-0.030040
15, 16	BP(-2, 1)	-0.014533
15, 16	BP(-2, 2)	-0.067391
15, 16	BP(-2, 3)	0.049273
15, 16	BP(-2, 4)	frozen
15, 16	BP(-2, 5)	frozen

C T E Q

MCnet

What do the eigenvectors *mean*?

- Take eigenvector 15 (for CTEQ6.1); error pdf's 29 and 30
- Probes high x gluon distribution

creates largest uncertainty for high p_T jet cross sections at both the Tevatron and LHC

29, 30	BP(2, 1)	0.012701
29, 30	BP(2, 2)	-0.162018
29, 30	BP(2, 3)	0.018666
29, 30	BP(2, 4)	-0.111238
29, 30	BP(2, 5)	frozen
29, 30	BP(1, 1)	-0.003049
29, 30	BP(1, 2)	-0.001074
29, 30	BP(1, 3)	-0.034151
29, 30	BP(1, 4)	-0.005735
29, 30	BP(1, 5)	0.032812
29, 30	BP(0, 1)	-0.045923
29, 30	BP(0, 2)	0.873418
29, 30	BP(0, 3)	frozen
29, 30	BP(0, 4)	-0.241822
29, 30	BP(0, 5)	frozen
29, 30	BP(-1, 1)	-0.071419
29, 30	BP(-1, 2)	-0.067488
29, 30	BP(-1, 3)	0.100283
29, 30	BP(-1, 4)	frozen
29, 30	BP(-1, 5)	0.179551
29, 30	BP(-2, 1)	-0.009441
29, 30	BP(-2, 2)	-0.196100
29, 30	BP(-2, 3)	0.211281
29, 30	BP(-2, 4)	frozen
29, 30	BP(-2, 5)	frozen

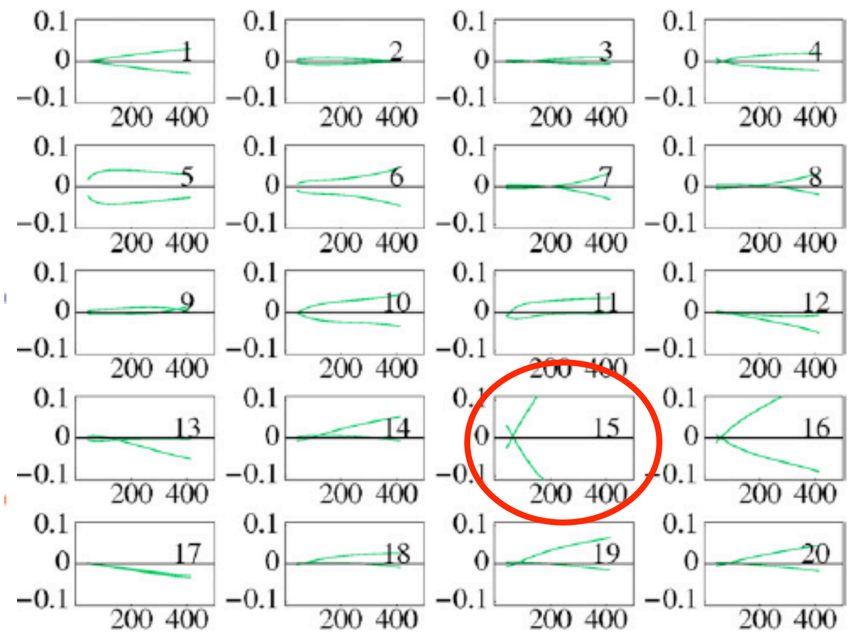


Figure 29. The pdf errors for the CDF inclusive jet cross section in Run 1 for the 20 different eigenvector directions. The vertical axes show the fractional deviation from the central prediction and the horizontal axes the jet transverse momentum in GeV.

CTEQ

MCnet

Aside: PDF re-weighting

- Any physical cross section at a hadron-hadron collider depends on the product of the two pdf's for the partons participating in the collision convoluted with the hard partonic cross section
- Nominally, if one wants to evaluate the pdf uncertainty for a cross section, this convolution should be carried out 41 times (for CTEQ6.1); once for the central pdf and 40 times for the error pdf's
- However, the partonic cross section is not changing, only the product of the pdf's
- So one can evaluate the full cross section for one pdf (the central pdf) and then evaluate the pdf uncertainty for a particular cross section by taking the ratio of the product of the pdf's (the pdf luminosity) for each of the error pdf's compared to the central pdf's

$$\sigma_{AB} = \int \mathrm{d}x_a \mathrm{d}x_b \; f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \; \hat{\sigma}_{ab \to X}$$

f^i is the error pdf and f^0 the central pdf

$$\frac{f^i{}_{a/A}(x_a, Q^2) f^i{}_{b/B}(x_b, Q^2)}{f^0{}_{a/A}(x_a, Q^2) f^0{}_{b/B}(x_b, Q^2)}$$

This works exactly for fixed order calculations and works well enough (see later) for parton shower Monte carlo calculations.

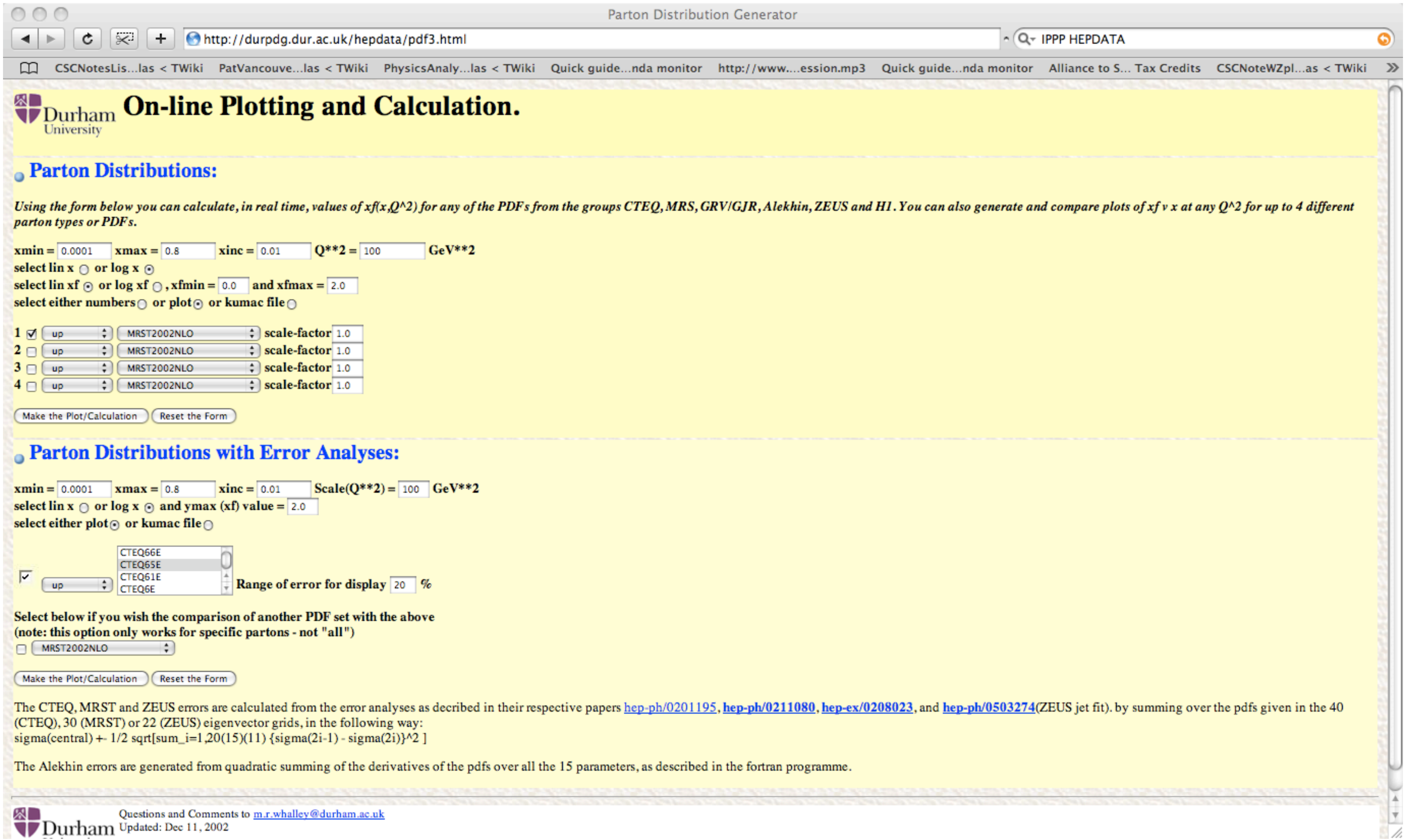
Most experiments now have code to easily do this…
and many programs will do it for you (MCFM)

CTEQ

MCnet

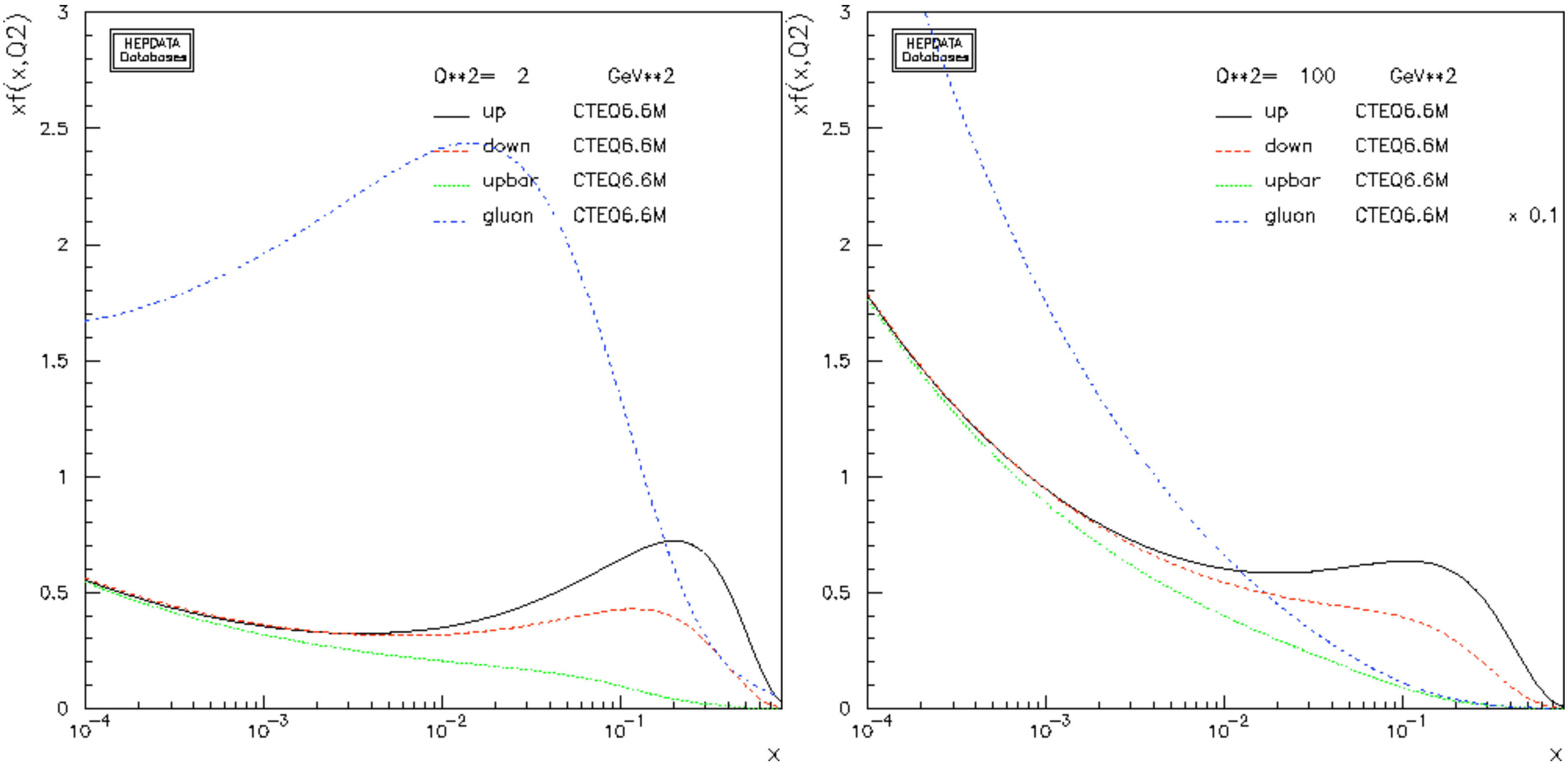
A very useful tool

Allows easy calculation and comparison of pdf's



Let's try it out

Up and down quarks dominate at high x, gluon at low x.
As Q^2 increases, note the growth of the gluon distribution, and to a lesser extent the sea quark distributions.

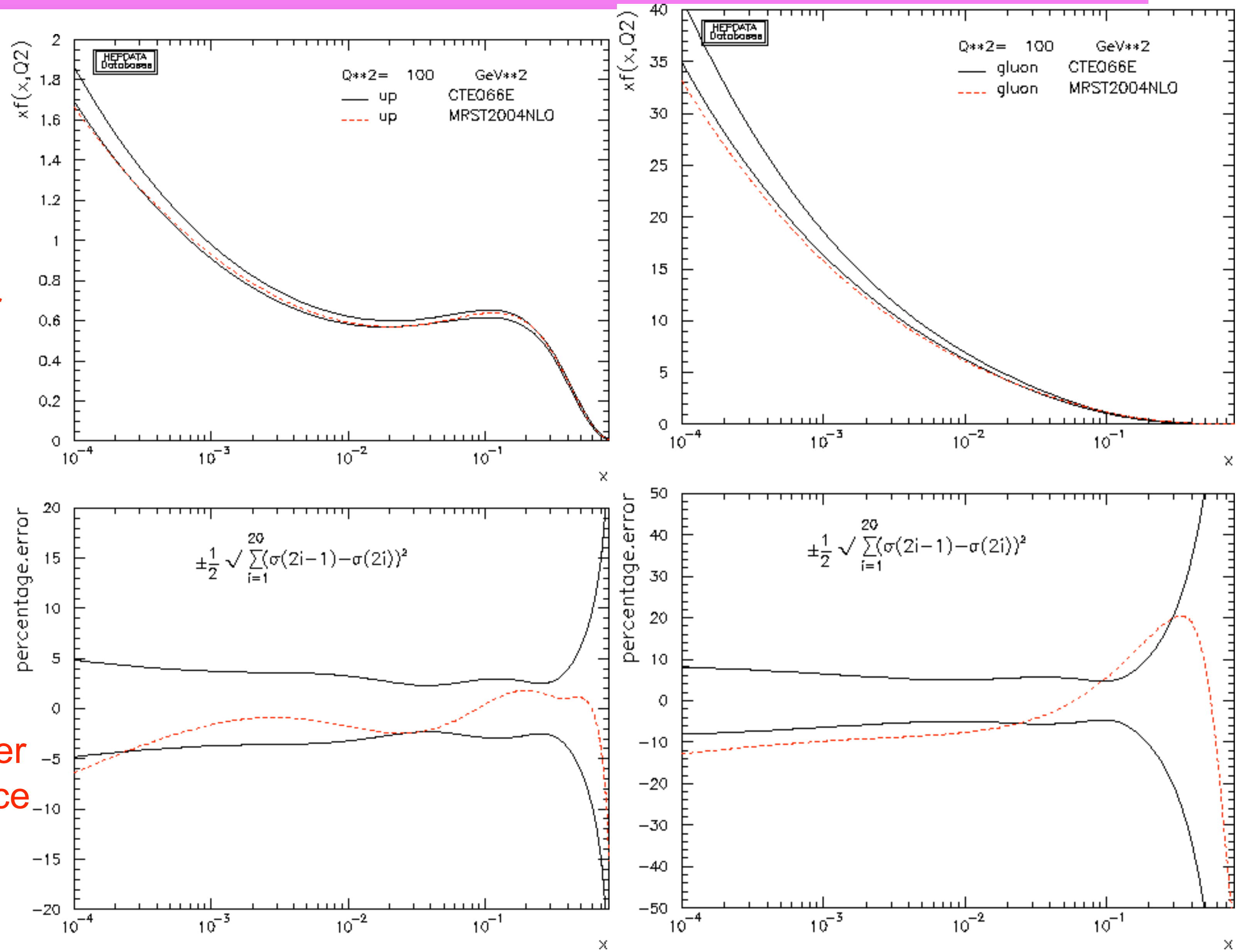


Uncertainties

uncertainties get large at high x

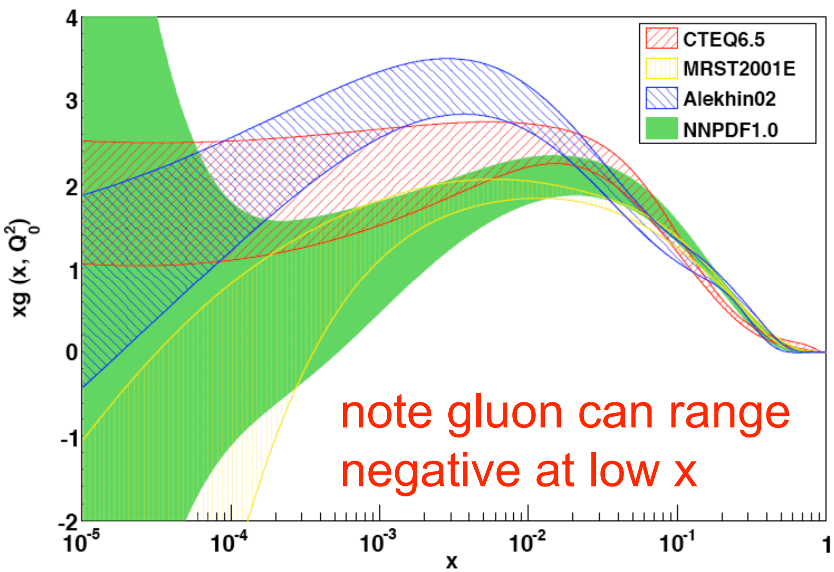
uncertainty for gluon larger than that for quarks

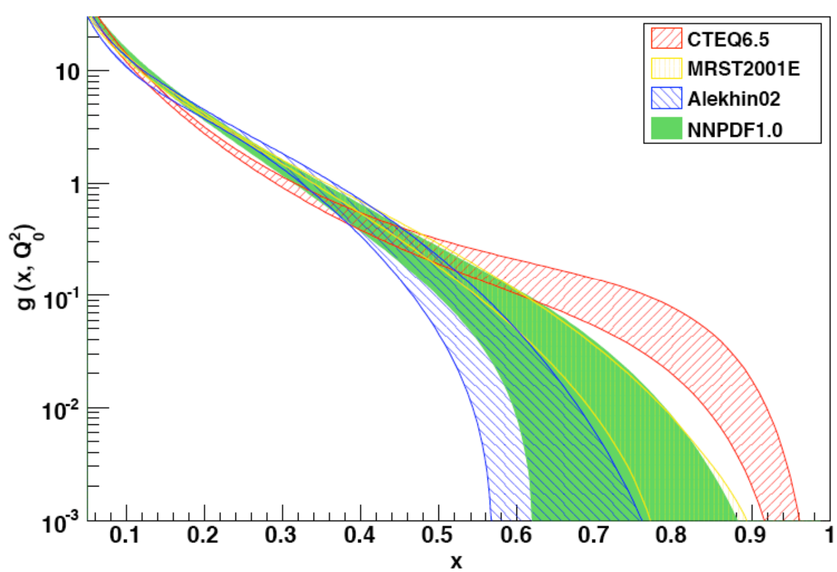
pdf's from one group don't necessarily fall into uncertainty band of another …would be nice if they did



Uncertainties and parametrizations

- Beware of extrapolations to x values smaller than data available in the fits, especially at low Q^2
- Parameterization may artificially reduce the apparent size of the uncertainties
- Compare for example uncertainty for the gluon at low x from the recent neural net global fit to global fits using a parametrization





Q^2=2 GeV^2

CTEQ

MCnet

Correlations

- Consider a cross section X(a)
- i^{th} component of gradient of X is

$$\frac{\partial X}{\partial a_i} \equiv \partial_i X = \frac{1}{2}(X_i^{(+)} - X_i^{(-)})$$

- Now take 2 cross sections X and Y
  - or one or both can be pdf's
- Consider the projection of gradients of X and Y onto a circle of radius 1 in the plane of the gradients in the parton parameter space
- The circle maps onto an ellipse in the XY plane
- The angle ϕ between the gradients of X and Y is given by

$$\cos\varphi = \frac{\vec{\nabla}X \cdot \vec{\nabla}Y}{\Delta X \Delta Y} = \frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$

- The ellipse itself is given by

$$\left(\frac{\delta X}{\Delta X}\right)^2 + \left(\frac{\delta Y}{\Delta Y}\right)^2 - 2\left(\frac{\delta X}{\Delta X}\right)\left(\frac{\delta Y}{\Delta Y}\right)\cos\varphi = \sin^2\varphi$$

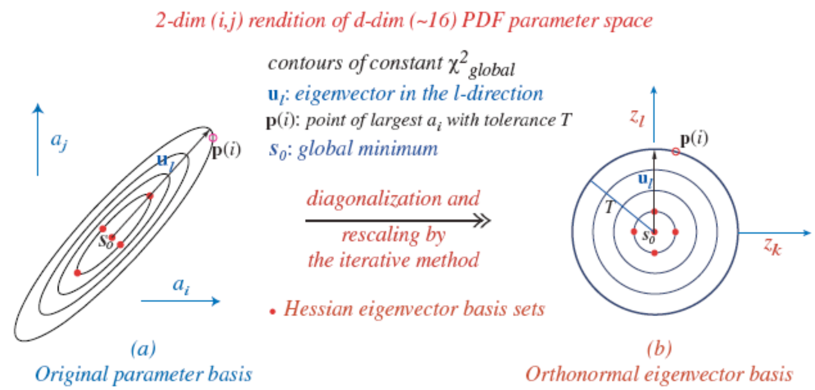


Figure 28. A schematic representation of the transformation from the pdf parameter basis to the orthonormal eigenvector basis.

- If two cross sections/pdf's are very correlated, then cosϕ~1
- …uncorrelated, then cosϕ~0
- …anti-correlated, then cosϕ~-1

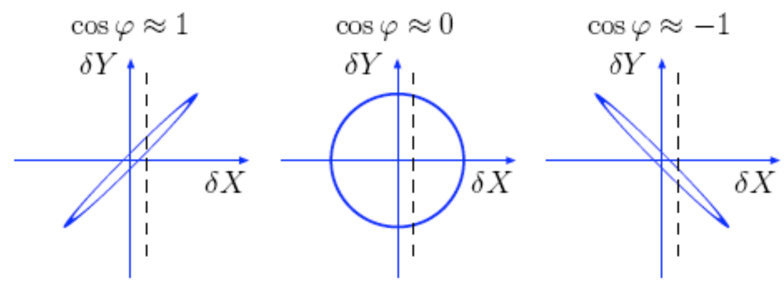


Figure 1: Dependence on the correlation ellipse formed in the $\Delta X - \Delta Y$ plane on the value of the correlation cosine $\cos\varphi$.

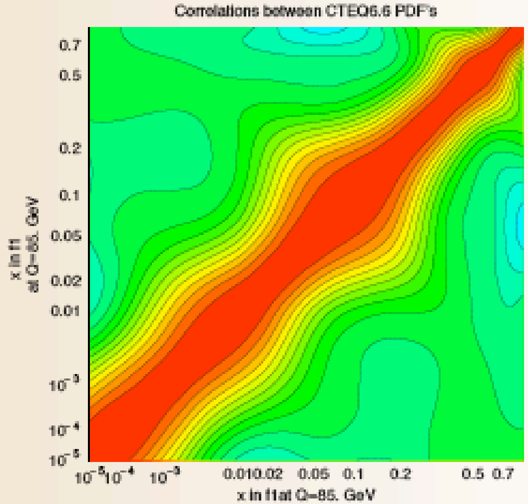
CTEQ

MCnet

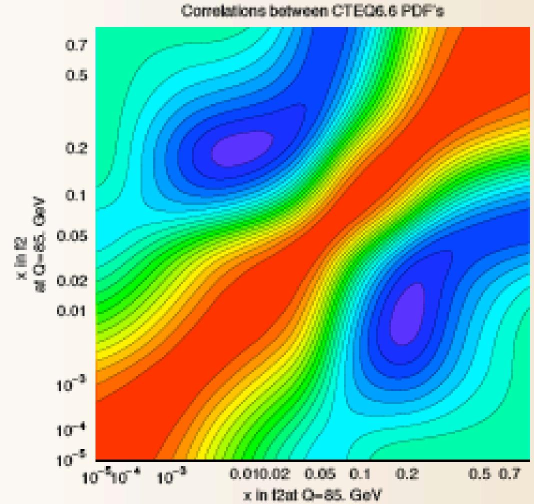
Correlations between pdf's

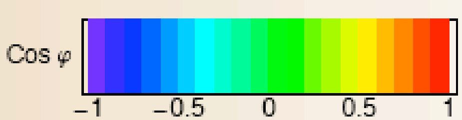
Correlations between $f(x_1, Q)$ and $f(x_2, Q)$ at $Q = 85$ GeV

$f_1(x_1, Q)$ vs. $f_1(x_2, Q)$



$f_2(x_1, Q)$ vs. $f_2(x_2, Q)$





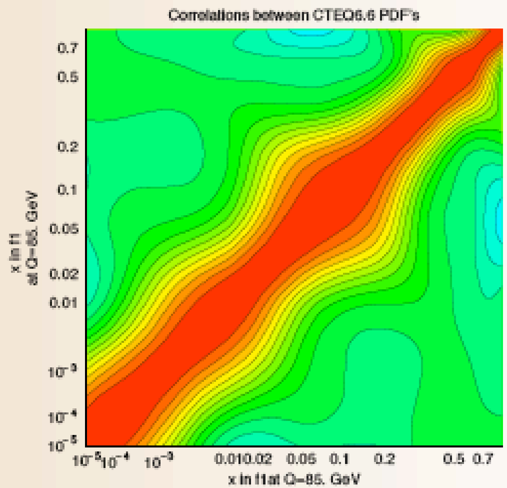
Pavel Nadolsky

Can you guess which PDF's these are?

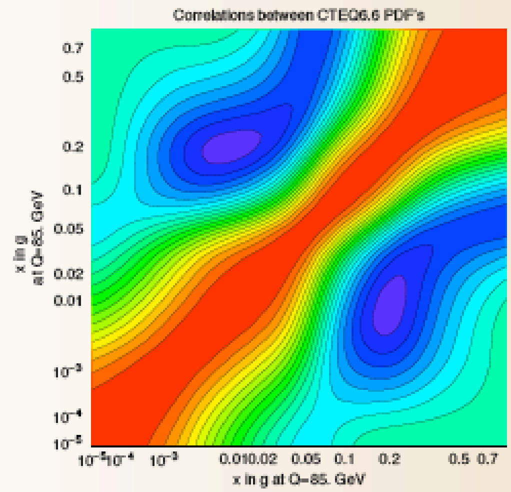
In-class answers

Correlations between $f(x_1, Q)$ and $f(x_2, Q)$ at $Q = 85$ GeV

$u(x_1, Q)$ vs. $u(x_2, Q)$

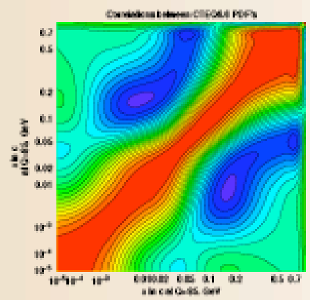


$g(x_1, Q)$ vs. $g(x_2, Q)$

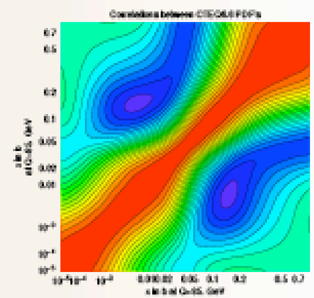


Correlation patterns look similar for g, c, b PDF's (no intrinsic charm here!)

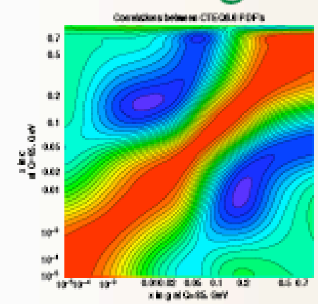
c vs. c



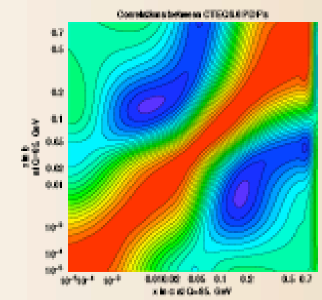
b vs. b



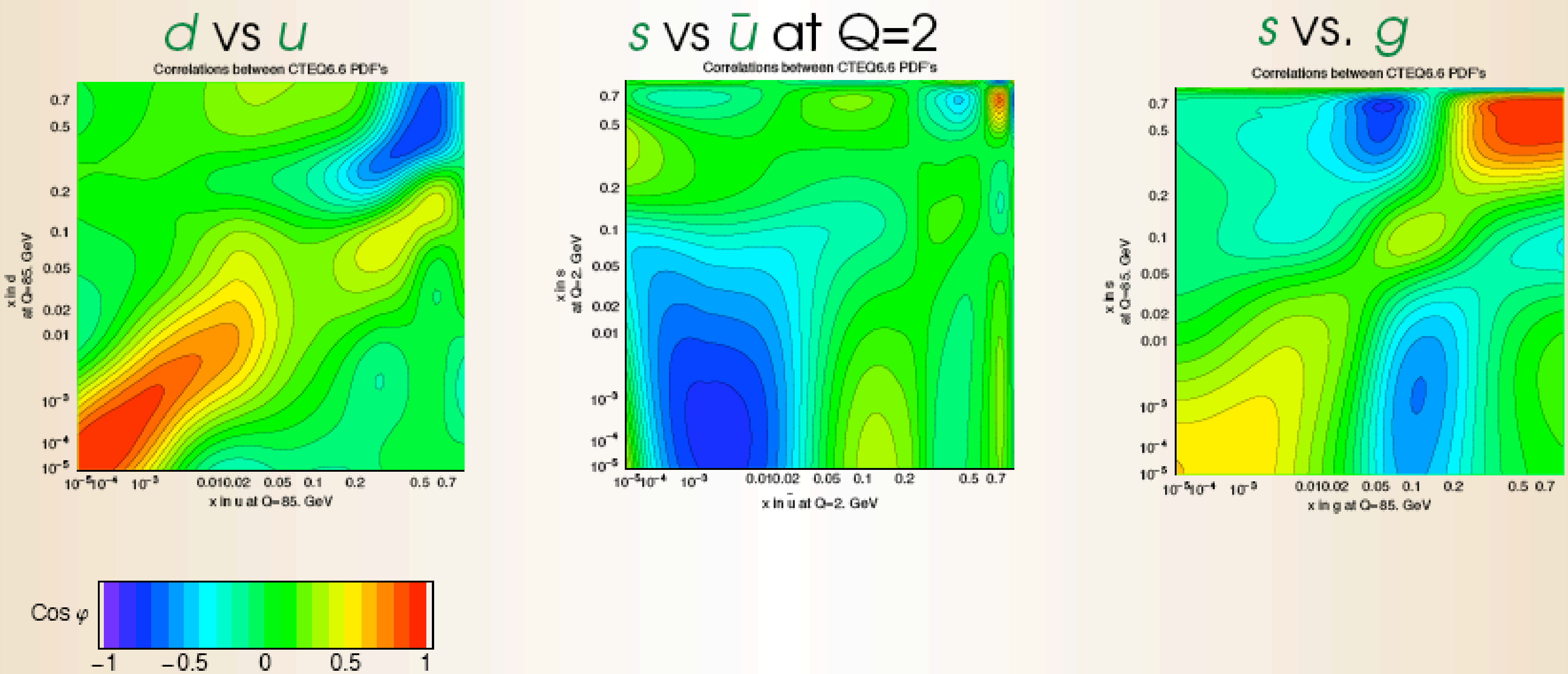
c vs. g



b vs. c



Correlations between $f_1(x_1, Q)$ and $f_2(x_2, Q)$ at $Q = 85$ GeV



Sometimes there is a clear physics reason behind the correlation (e.g., sum rules or assumed Regge-like behavior); sometimes not

C T E Q

MCnet

Try it yourself