Recent PDF analysis and its applications

Pavel Nadolsky
in collaboration with
Q.-H. Cao, J. Huston, H.-L. Lai, J. Pumplin,
D. Stump, W.-K. Tung, C.-P. Yuan

Michigan State University

May 28, 2008

Outline

- An overview of CTEQ6.6 family of PDF's
- Constraints on the gluon PDF from $t\bar{t}$ and jet production
- PDF reweighting in Monte-Carlo integration

Global analysis at Michigan State/Taiwan/Washington

- a part of the Coordinated Theoretical Experimental study of QCD (CTEQ) in U.S.A.
- development of general-purpose PDF's

  (Wu-Ki Tung and collaborators)
- new CTEQ6.6M standard set and 44 extreme eigenvector sets *(arXiv:0802.0007)*
  - improved treatment of s, c, b PDF's
  - correlation analysis of collider observables
  - available in the LHAPDF-5.4 library and at www.cteq.org

CTEQ6.5 and CTEQ6.6: advanced treatment of heavy quarks

1. full implementation of the **general-mass "SACOT-χ" scheme**
   - differences in predictions for c, b scattering ($F_2^{c,b}(x, Q^2)$, etc.), EW precision cross sections, as compared to the zero-mass CTEQ6.1

*Tung et al., JHEP 0702, 053 (2007); **CTEQ6.5***

2. exploration of **free strange PDF's** and/or asymmetric strange sea

$$s_+(x) \neq r\left(\bar{u}(x) + \bar{d}(x)\right), \quad s_-(x) \neq 0,$$
$$\text{where } s_\pm(x) \equiv s(x) \pm \bar{s}(x)$$

*Lai et al., JHEP 0704, 089 (2007); **CTEQ6.5S***

3. PDF's with **nonperturbative charm**
   - $c(x, \mu_0 = m_c) \neq 0$ due to low-energy charm excitations (as opposed to $g \to c\bar{c}$ radiative production)

*Pumplin et al., PRD 75, 054029 (2007); **CTEQ6.5C***

CTEQ6.6 study consolidates these developments

PDF family	Number of PDF sets	$s_+(x)$	$s_-(x)$	$\alpha_s(M_Z)$	Nonpert. charm PDF
6.6	45	free	0	0.118	No

- CTEQ6.6M + **44** extreme eigenvector sets
  - $s_+(x)$ is **independent** of $\bar{u}(x) + \bar{d}(x)$
- $s_-(x) = 0$, in agreement with the data at 90% c.l.
  - the preference for $s_-(x) \neq 0$ remains marginal: $\Delta\chi^2 = -15$ for $\int_0^1 x s_-(x, \mu_0) dx = 0.0018$, $\sqrt{2N_{NuTeV}} = 22$

CTEQ6.6 study consolidates these developments

PDF family	Number of PDF sets	$s_+(x)$	$s_-(x)$	$\alpha_s(M_Z)$	Nonpert. charm PDF
6.6	45	free	0	0.118	No
6.6C	4	free	0	0.118	Yes

- CTEQ6.6C: updated PDF's with intrinsic charm

CTEQ6.6 study consolidates these developments

PDF family	Number of PDF sets	$s_+(x)$	$s_-(x)$	$\alpha_s(M_Z)$	Nonpert. charm PDF
6.6	45	free	0	0.118	No
6.6C	4	free	0	0.118	Yes
6.6A	4	free	0	0.112-0.125	No

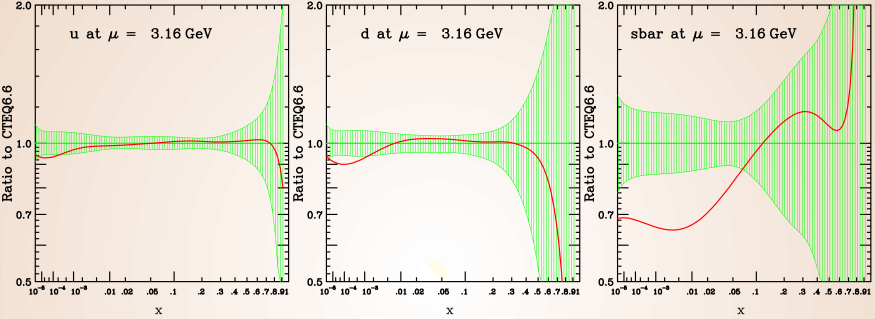
- CTEQ6.6C: updated PDF's with intrinsic charm
- CTEQ6.6A: PDF's for $\alpha_s(M_Z) = 0.112 - 0.125$

CTEQ6.6 study consolidates these developments

PDF family	Number of PDF sets	$s_+(x)$	$s_-(x)$	$\alpha_s(M_Z)$	Nonpert. charm PDF
6.6	45	free	0	0.118	No
6.6C	4	free	0	0.118	Yes
6.6A	4	free	0	0.112-0.125	No

- CTEQ6.6C: updated PDF's with intrinsic charm
- CTEQ6.6A: PDF's for $\alpha_s(M_Z) = 0.112 - 0.125$
- All CTEQ6.6 sets are provided for $10^{-8} \leq x \leq 1$

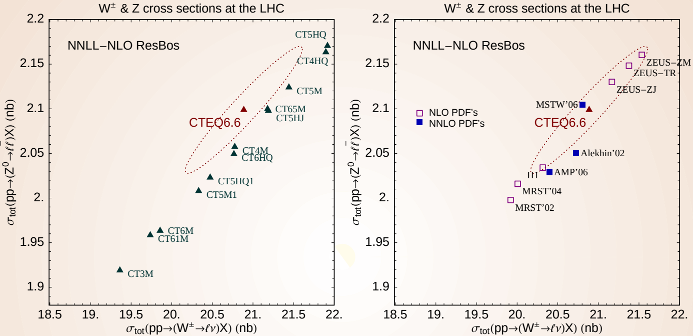
CTEQ6.6 PDF's



dashes: CTEQ6.1M (zero-mass scheme)

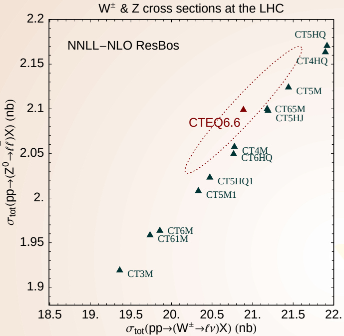
- CTEQ6.6 u, d are above CTEQ6.1 at $x \lesssim 10^{-2}$
  - The result of suppressed charm contribution to $F_2(x, Q)$ at HERA in the GM-VFN scheme
- very different strange PDF's

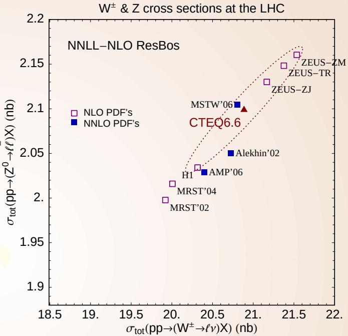
W and Z cross sections at the LHC



- At the LHC, $\sigma_{W,Z}$(CTEQ6.6M)$\approx 1.06\sigma_{W,Z}$(CTEQ6.1M)
  - reflects a 6% increase in light quark luminosities $\mathcal{L}_{q_i\bar{q}_j}(x_1,x_2,Q) = q_i(x_1,Q)\bar{q}_j(x_2,Q)$ at relevant x and Q

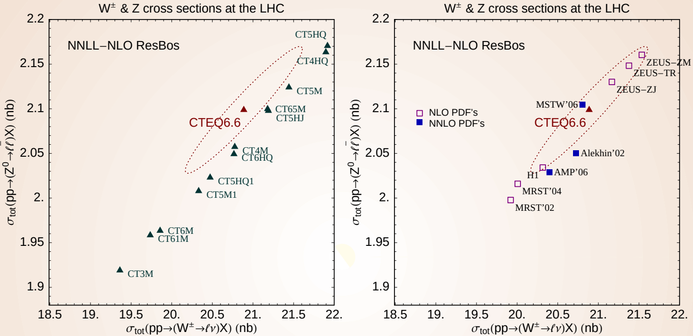
W and Z cross sections at the LHC





- Such changes in $\sigma_{Z,W}$ exceed NNLO corrections or anticipated experimental error of $\sim 1\%$
- Two latest MSTW predictions are compatible with the CTEQ6.6 result

W and Z cross sections at the LHC



W, Z production: which parton flavors drive the "experimental" PDF uncertainties?

Ideally, we would like to relate $\delta_{PDF}\sigma$ to (a few) specific $f_a(x,\mu)$ and explain (anti-)correlations between observables

Correlation analysis for collider observables

(J. Pumplin et al., PRD 65, 014013 (2002); P.N. and Z. Sullivan, hep-ph/0110378)

A technique based on the Hessian method

For $2N$ PDF eigensets and two cross sections X and Y:

$$\Delta X = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}$$

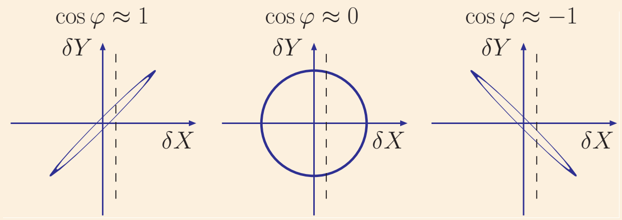
$$\cos\varphi = \frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$

$X_i^{(\pm)}$ are maximal (minimal) values of X_i tolerated along the i-th PDF eigenvector direction; $N = 22$ for the CTEQ6.6 set

Correlation angle φ

Determines the parametric form of the $X - Y$ correlation ellipse

$$
\begin{aligned}
X &= X_0 + \Delta X \cos\theta \\
Y &= Y_0 + \Delta Y \cos(\theta + \varphi)
\end{aligned}
$$



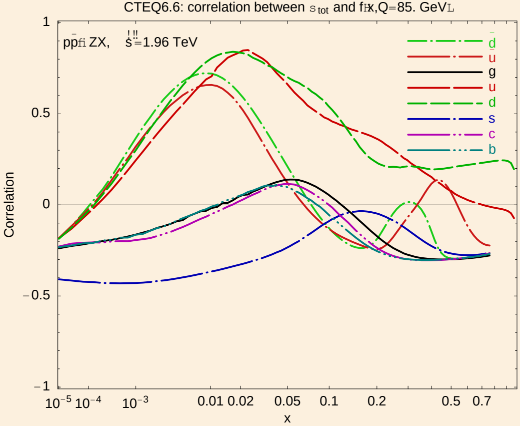
X_0, Y_0: best-fit values

$\Delta X, \Delta Y$: PDF errors

$\cos\varphi \approx \pm 1$:	Measurement of X imposes	tight	constraints on Y
$\cos\varphi \approx 0$:		loose	

Correlations $\cos\varphi$ between W, Z cross sections and PDF's

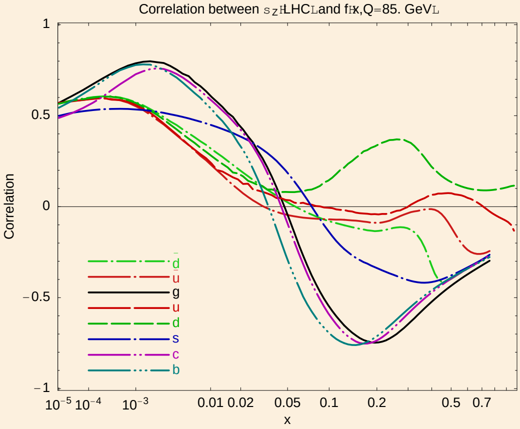
Tevatron Run-2



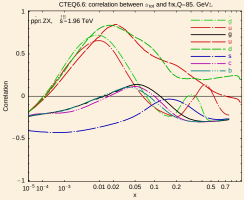
Similar correlations for W production

Correlations $\cos\varphi$ between W, Z cross sections and PDF's

LHC



Tevatron Run-2

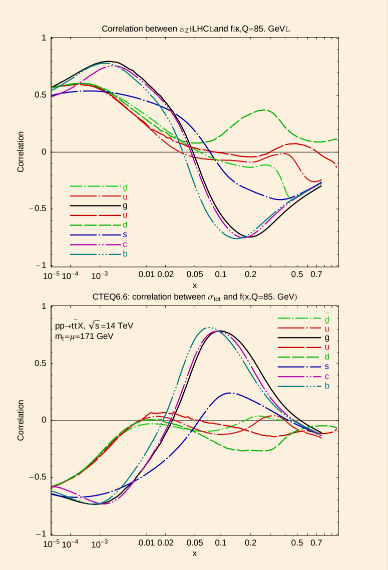


Similar correlations for W production

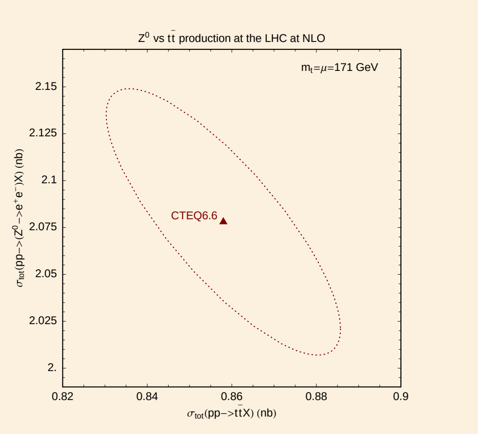
Correlations of Z and $t\bar{t}$ cross sections with PDF's

LHC Z, W cross sections are strongly correlated with $g(x)$, $c(x)$, $b(x)$ at $x \sim 0.005$

$\therefore$ they are strongly anticorrelated with processes sensitive to $g(x)$ at $x \sim 0.1$ ($t\bar{t}$, $gg \to H$ for $M_H > 300$ GeV) as a consequence of momentum sum rule

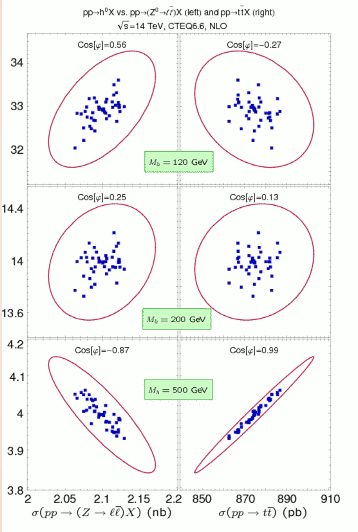


$t\bar{t}$ vs Z cross sections at the LHC



Measurements of $\sigma_{t\bar{t}}$ and σ_Z probe the same (gluon) PDF degrees of freedom at different x values

Correlations between $\sigma(gg \to H^0)$, σ_Z, $\sigma_{t\bar{t}}$



As M_H increases:

- $\cos\varphi(\sigma_H, \sigma_Z)$ decreases
- $\cos\varphi(\sigma_H, \sigma_{t\bar{t}})$ increases

$t\bar{t}$ production as a standard candle process

- Measurements of $\sigma_{t\bar{t}}$ with accuracy $\sim 5\%$ may be within reach
- would provide additional constraints on the large-x gluon PDF
- will be useful for monitoring of $\mathcal{L}_{LHC}$ luminosity in the first years and normalization of LHC event rates

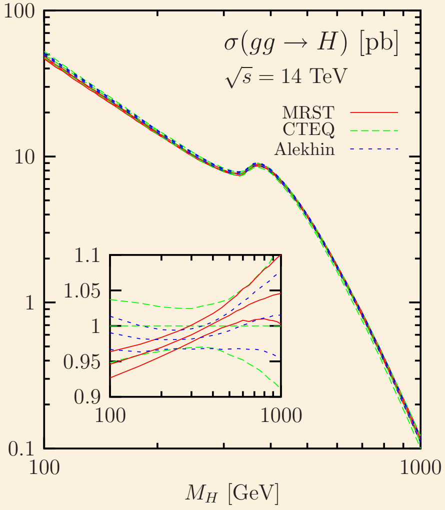
See also the talk by M. Csakon; Moch, Uwer, arXiv:0804.1476; Cacciari et al., arXiv:0804.2800; Kidonakis, Vogt, arXiv:0805.3844

$t\bar{t}$ production as a standard candle process

Uncertainties in $\sigma_{t\bar{t}}$ for $m_t = 171$ GeV

Type	Current	Projected	Assumptions
Scale dependence	11% (NLO)	$\sim 3-5\%$? (NNLO+resum.)	$m_t/2 \le \mu \le 2m_t$
PDF dependence	2%	1%?	1σ c.l.
m_t dependence	5% $\delta m_t = 2$ GeV	$< 3\%$ $\delta m_t = 1$ GeV	
Total (theory)	12%	$\sim 5\%$	
Experiment	8% (CDF)	5%?	

Gluon PDF's and LHC predictions

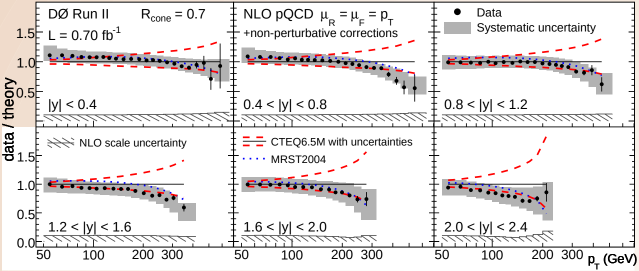


Djouadi, Ferrag, 2003

In gluon scattering processes ($gg \to t\bar{t}X$, $gg \to HX$, etc.), differences due to the choice of the PDF set may exceed Hessian PDF uncertainties, partly reflecting

- choice of the data sets constraining $g(x)$ (such as the jet data)
- treatment of systematic errors in these data

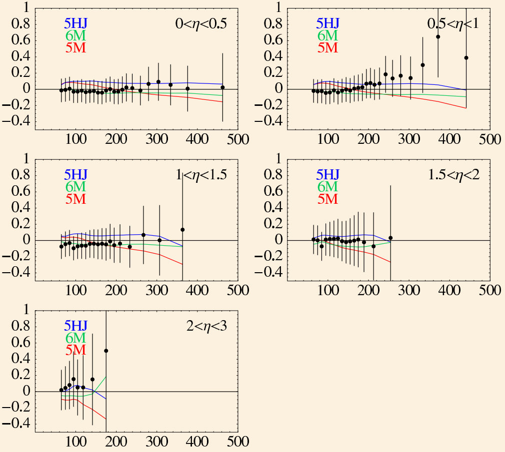
Inclusive jet production in the Tevatron Run-2



D0 Coll., arXiv:0802:2400 (700 pb^{-1}); similar tendency in CDF results (1.13 fb^{-1})

- (Almost) negligible statistical error
- sizable correlated systematic uncertainties
- Midpoint/k_T algorithm samples, corrected to parton level
- tend to lie below NLO theory based on CTEQ6.X/NLOJET++ at high E_T

Inclusive jet production in the Tevatron Run-1



- CDF, 1993: excess in high-E_T jet data over NLO theory based on contemporary "central-fit" PDFs *(CTEQ3M, 4M, 5M)*
- 2001: reaffirmed by D0 data in separate pseudorapidity regions (cf. figure)

- Accommodated by assuming a larger gluon PDF $g(x)$ at $x \gtrsim 0.1$ *(CTEQ4HJ, 5HJ, 6xM)*

Impact of Run-2 jet data on global fits

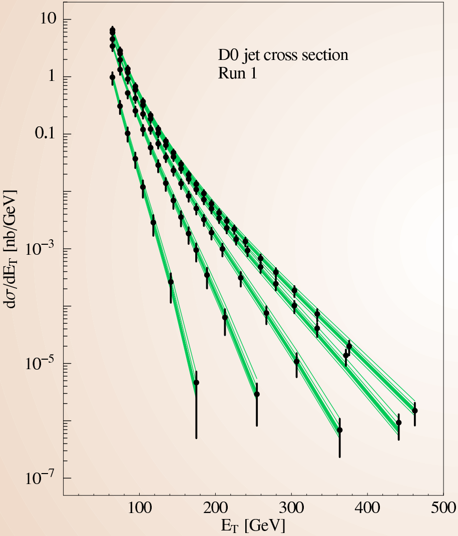
Huston, Lai, P.N., Pumplin, Stump, Tung, Yuan

Several issues affect the ability of the Run-2 data to constrain the PDF's

- reliability of theoretical predictions?
  - agreement between different NLO calculations at the level of a few percent
  - dependence on the choice of scale, R_{sep}, threshold corrections...
- compatibility of the Run-1 and Run-2 measurements?

Comparison of NLO calculations for $p\bar{p} \to j X$

Nagy, Olness, Soper, Pumplin, Yuan, in progress



Rapid falloff of $d\sigma/dp_T^{jet}$ affects numerical accuracy, especially at large η and/or large p_T^j

- ongoing pairwise comparisons of the existing NLO codes *(Ellis-Kunszt-Soper, NLOJET++, and FastNLO)*
- discrepancies were found, and some were resolved
- NLO theoretical errors can substantially exceed 2% *(D. Soper)*

Systematic uncertainties in CTEQ analysis

Pumplin, Stump, Tung, et al.: hep-ph/0101051; hep-ph/0201195

Minimization of a function

$$\chi^2(\{a\},\{r\}) = \sum_{expt}\left[\chi_D^2 + \chi_r^2\right]$$

with respect to PDF parameters $\{a\}$ (done numerically) and K systematic correlated parameters $\{r\}$ (done analytically)

$$\chi_D^2 = \sum_{i=1}^{N_e} \frac{1}{\alpha_i^2}\left(D_i - T_i(\{a\}) - \sum_{k=1}^{K} r_k \beta_{ki}\right)^2; \ \chi_r^2 = \sum_{k=1}^{K} r_k^2;$$

- N_e : number of data points
- D_i, T_i: data and theory
- $\alpha_i^2 = \sigma_i^2 + u_i^2$: total (stat.+syst.) uncorrelated errors

Systematic uncertainties in CTEQ analysis

Pumplin, Stump, Tung, et al.: hep-ph/0101051; hep-ph/0201195

Minimization of a function

$$\chi^2(\{a\},\{r\}) = \sum_{expt}\left[\chi_D^2 + \chi_r^2\right]$$

with respect to PDF parameters $\{a\}$ (done numerically) and K systematic correlated parameters $\{r\}$ (done analytically)

$$\chi_D^2 = \sum_{i=1}^{N_e}\frac{1}{\alpha_i^2}\left(D_i - T_i(\{a\}) - \sum_{k=1}^{K} r_k\beta_{ki}\right)^2;\ \chi_r^2 = \sum_{k=1}^{K} r_k^2;$$

- $\chi_D^2, \chi_r^2, \chi_{D+r}^2 \equiv \chi_D^2 + \chi_r^2$ satisfy χ^2 distributions with $N_e - K$, K, and N_e degrees of freedom
- A measure of the goodness of the fit: probabilities $P_a \equiv$ Prob $\left[\chi_a^2 > \chi_{a0}^2\right]$ for χ_a^2 to exceed the best-fit values χ_{a0}^2 ($a = D,\ r,\ D+r$)
- $P_a = 50\%$ corresponds to $\chi_a^2/d.o.f. = 1$

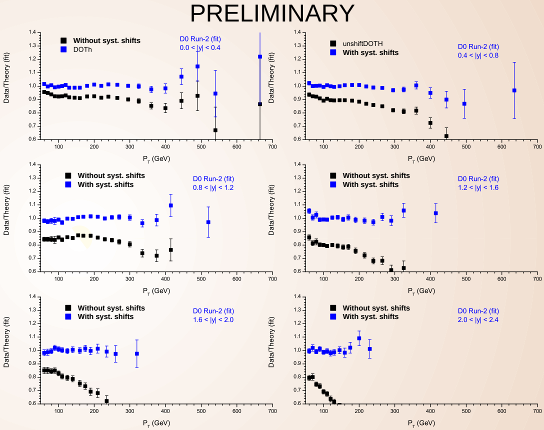
Comparing Run-1 and Run-2 data

Compare two fits:

(I) only to Run-1 data (CTEQ6.6)

(II) only to Run-2 data

Analytic minimization with respect to $\{r\}$ leads to effective shifts in the experimental data



These shifts are essential for getting a good fit to any jet data set

χ^2_r, χ^2_{D+r} (upper line) and P_r, P_{D+r} (lower line, in %)

PRELIMINARY

Experiment	N_e	K	Fit I	Fit II
CDF Run-1	33	7	6.4, 53.2 49, 1.4	4.2, 51.1 76, 2.3
D0 Run-1	90	N/A	-, 88.9 -, 51	-, 114 -, 4.5
CDF Run-2 (Midpoint)	72	16	19.9, 111 23, 0.2	13.7, 99.5 62, 2
D0 Run-2 (Midpoint)	110	23	23.2, 123.9 45, 17	18.9, 122 71, 20

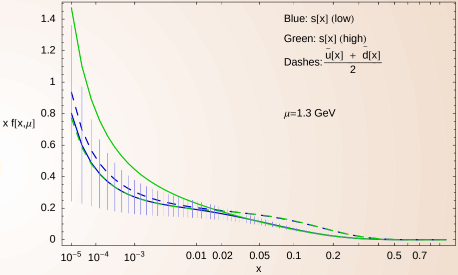
CDF: χ^2_D seems to be increased by two outlier points in Run-1, excessive scatter of data in Run-2 (no uncorr. syst. error provided)

Improved fit to D0 Run-1 implies worse fit to Run-2 data, and vice versa $\Rightarrow$ a hint of tension

Quark flavor (a)symmetry at small x

- For 3 active flavors at $Q_0 = m_c$, the exact $SU(3)_F$ symmetry would imply $\bar{u}(x) = \bar{d}(x) = \bar{s}(x)$ at $x \to 0$

- The actual data allows large violations of $SU(3)_F$ at $x < 10^{-2}$, including solutions with $\bar{s}(x) \approx 1.8\,(\bar{u}(x) + \bar{d}(x))$ at $x \to 0$



- flat direction in the PDF parameter space!

- can be constrained by measuring charged particle multiplicities $d\sigma(ep \to ehX)/(dQ^2 dx dz)$ (with separate samples of $h = \pi, K, p$) in SIDIS at HERA

PDF reweighting in Monte-Carlo integration

If $X_i^{(\pm)}$ and $\Delta X^2 = \sum_{i=1}^{N} \left(X_i^{(+)} - X_i^{(-)} \right)^2/4$ are computed in $2N = 44$ independent Monte-Carlo runs with $\bar{N}$ events each, their resulting estimates are given by

$$\overline{X}_i^{(\pm)} = X_i^{(\pm)} + \overline{\delta}_i^{(\pm)} \sim X_i^{(\pm)} + \frac{c}{\overline{N}^{1/2}} \text{ and}$$

$$\overline{\Delta X}^2 = \frac{1}{4} \sum_{i=1}^{N} \left(\overline{X}_i^{(+)} - \overline{X}_i^{(-)} \right)^2 \sim \Delta X^2 + \frac{c' N}{\overline{N}^{1/2}}$$

$\overline{\delta}_i^{(\pm)}$ is a **random** MC error dependent on the input PDF, arising, e.g., from importance sampling

As a result of the PDF dependence of $\overline{\delta}_i^{(\pm)}$, the error $\overline{\Delta X}^2 - \Delta X^2$ is increased by a factor $N \sim 22$

PDF reweighting in Monte-Carlo integration

- PDF reweighting generates the same sequence of events to compute each of $2N$ cross sections
  - all $\overline{\delta}_i^{(\pm)}$ are the same
  - $\overline{\Delta X}^2 = \Delta X^2$
- In multi-loop calculations, PDF reweighting saves CPU time drastically by reducing slow computations of hard-scattering matrix elements

FROOT: a theorist-friendly interface for Monte-Carlo reweighting

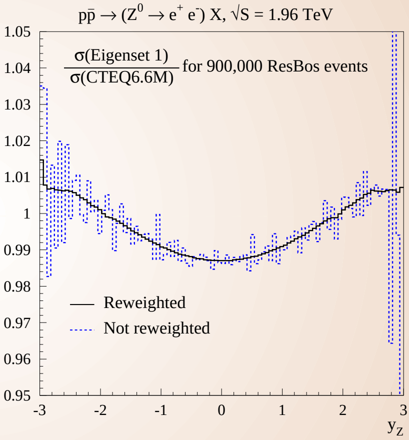
- Written in C, can be linked to standalone FORTRAN/C/C++ programs
- Simple – 170 lines of the code
- Writes the output directly into a ROOT ntuple; no need in intermediate PAW ntuples
- Flexible; new columns (branches) with PDF weights or events can be added into an existing ntuple
- Kinematical cuts, selection conditions can be imposed a posteriori in interactive or batch ROOT sessions
- implemented in ResBos (Q_T resummation for W, Z, H production)

FROOT: a theorist-friendly interface for Monte-Carlo reweighting

```
// These are the C functions accessible from Fortran.


extern "C" {
  //Initialization of the ROOT file
  void initrootnt_(const char *title, const char *access, int ltitle, int laccess);
  void reinitrootnt_(const char *access, int laccess);
  void addntbranch_(float *element, const char *chtag, int ltag);
  void fillntbranch_(const char *chtag, int ltag);
  int getnumbranches_();
  void rootntoutp_();
  void printnt_();
  void teststr_(const char *str, int lstr);

}//extern "C"
```

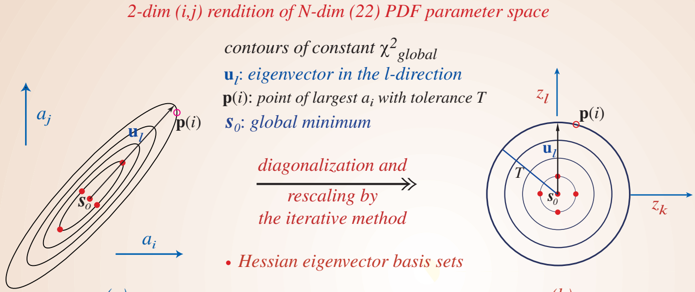


Concluding remarks

- CTEQ6.6 PDF's in general-mass scheme:
  - important differences from ZM-VFNS and some CTEQ6.5 predictions
  - must be used as the standard CTEQ set from now on
- Analysis of correlations in PDF parameter space is a powerful technique to explore connections between physics observables through shared PDF degrees of freedom
- Need to re-examine uncertainties in $SU(3)$ flavor composition at small x, large-x gluon PDF; new measurements of particle multiplicities in SIDIS at HERA, in $t\bar{t}$ production at the LHC would be helpful
- FROOT: a simple module for PDF reweighting of MC events and direct output into ROOT ntuples *(available by request)*

Backup slides

Tolerance hypersphere in the PDF space



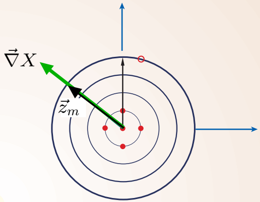
(a) Original parameter basis

(b) Orthonormal eigenvector basis

A hyperellipse $\Delta\chi^2 \leq T^2$ in space of N physical PDF parameters $\{a_i\}$ is mapped onto a hypersphere of radius T in space of N orthonormal PDF parameters $\{z_i\}$

Tolerance hypersphere in the PDF space

2-dim (i,j) rendition of N-dim (22) PDF parameter space

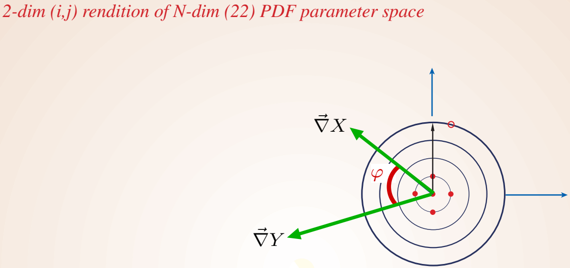


(b)
Orthonormal eigenvector basis

PDF error for a physical observable X is given by

$$\Delta X = \vec{\nabla} X \cdot \vec{z}_m = \left|\vec{\nabla} X\right| = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}$$

Tolerance hypersphere in the PDF space

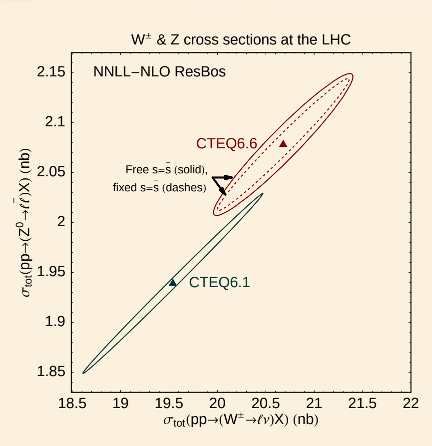


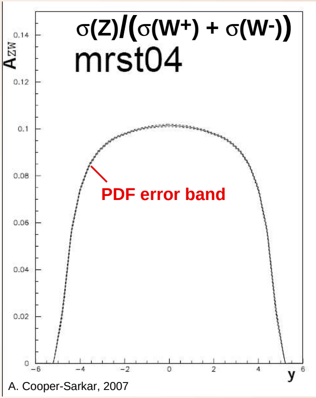
(b)
Orthonormal eigenvector basis

Correlation cosine for observables X and Y:

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

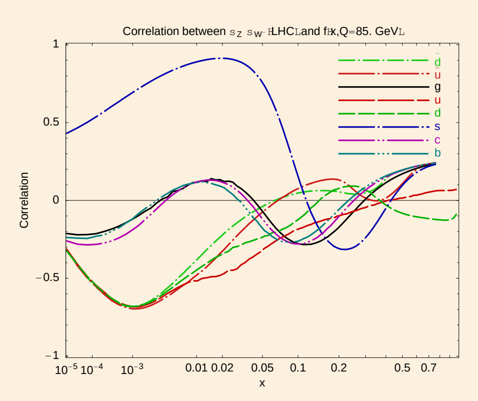
Correlations and ratio of W and Z cross sections





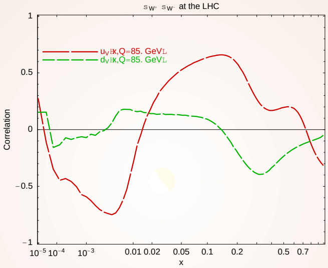
Radiative contributions, PDF dependence have similar structure in W, Z, and alike cross sections; cancel well in Xsection ratios

σ_Z/σ_W at the LHC



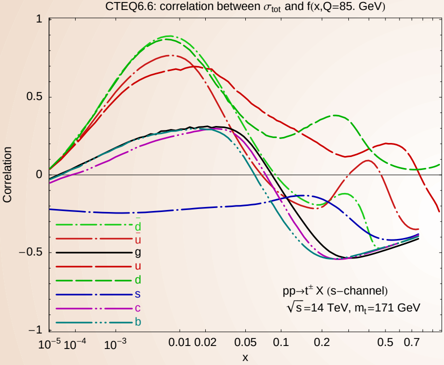
The remaining PDF uncertainty in σ_Z/σ_W is mostly driven by $s(x)$; increases by a factor of 3 compared to CTEQ6.1 as a result of free strangeness in CTEQ6.6

$\sigma(W^+)/\sigma(W^-)$

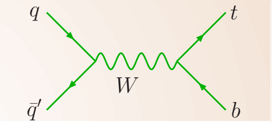


$\sigma(W^+)/\sigma(W^-) = 1.36 + 0.016$ (CTEQ6.6), 1.36 (MSTW'06NNLO), 1.35 (MRST'04NLO)

An example of a small correlation with the gluon



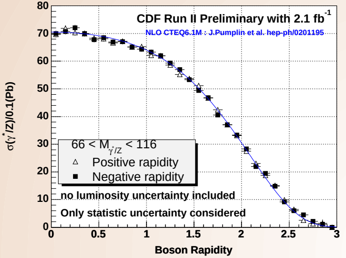
Single-top production (NLO)

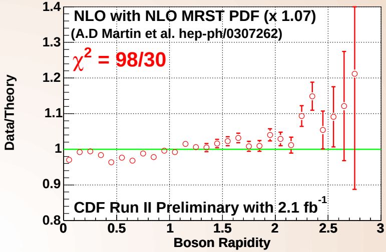


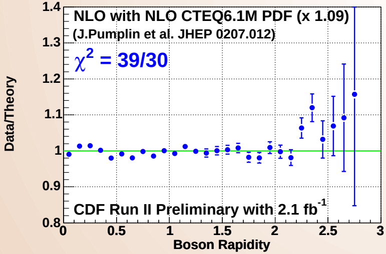
- typical $x \sim 0.01$
- mostly correlated with u, d PDF's

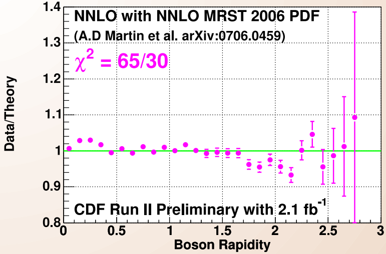
PDF uncertainties in W, Z *total* cross sections are irrelevant for some quark scattering processes (single-top, Z', ...)

$d\sigma/dy$ in $p\bar{p} \to \ell^+\ell^- X$ from CDF (2.1 fb^{-1})



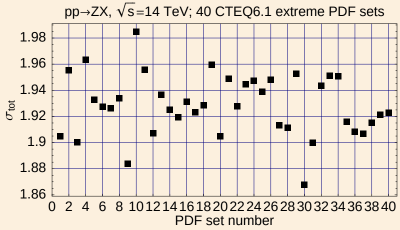






An inefficient application of the error analysis

☺ Compute σ_Z for 40 (44 in CTEQ6.6) extreme PDF eigensets

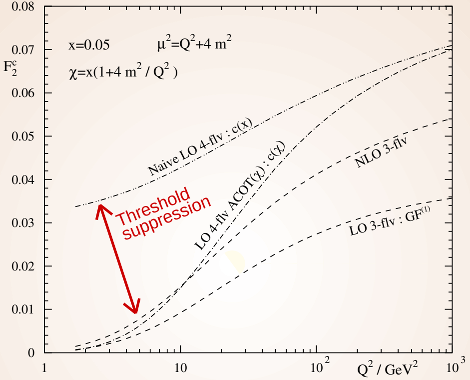


☺ Find eigenparameter(s) producing largest variation(s), such as #9, 10, 30

☹ It is not obvious how to relate abstract eigenparameters to physical PDF's $u(x)$, $d(x)$, etc.

A better insight is provided by correlation angles between σ and $f_a(x,Q)$

General-mass (ACOT-χ) factorization scheme



- Charm Wilson coefficient function is suppressed at $Q \to m_c$
- To keep agreement with DIS F_2 data, $u, d, \bar{u}, \bar{d}$ PDF's are enhanced at small x, as compared to the zero-mass (ZM-VFN) scheme