Recent progress in global PDF analysis

Graeme Watt

University College London

DIS 2008, London
7th April 2008

In collaboration with A. D. Martin, W. J. Stirling and R. S. Thorne

Introduction

- **20 years** since publication of first MRS analysis:

  A. D. Martin, R. G. Roberts and W. J. Stirling, "Structure-function analysis and ψ, jet, W, and Z production: Determining the gluon distribution," Phys. Rev. D **37** *(1988) 1161.*

- **Personnel changes:** R. S. Thorne (1998–), R. G. Roberts (–2004), G.W. (2006–). Hence MRS → MRST → MSTW.
- Focus on a few recent developments in MSTW 2008 analysis:
  1. Dynamic determination of **tolerance**.
  2. Tevatron Run II inclusive **jet** data.
  3. Tevatron Run II **W and Z** boson production.
  4. **Strangeness** in global PDF analysis.

Data sets fitted in MSTW 2008 (prel.) analysis

Data set	$N_{\text{pts.}}$
H1 MB 99 e^+p NC	8
H1 MB 97 e^+p NC	64
H1 low Q^2 96–97 e^+p NC	80
H1 high Q^2 98–99 e^-p NC	126
H1 high Q^2 99–00 e^+p NC	147
ZEUS SVX 95 e^+p NC	30
ZEUS 96–97 e^+p NC	144
ZEUS 98–99 e^-p NC	92
ZEUS 99–00 e^+p NC	90
H1 99–00 e^+p CC	28
ZEUS 99–00 e^+p CC	30
H1/ZEUS $e^{\pm}p$ F_2^{charm}	83
H1 99–00 e^+p incl. jets	24
ZEUS 96–97 e^+p incl. jets	30
ZEUS 98–00 $e^{\pm}p$ incl. jets	30
DØ II $p\bar{p}$ incl. jets	110
CDF II $p\bar{p}$ incl. jets	76
CDF II $W \to l\nu$ asym.	22
DØ II $W \to l\nu$ asym.	10
DØ II Z rap.	28
CDF II Z rap.	29

Data set	$N_{\text{pts.}}$
BCDMS μp F_2	163
BCDMS μd F_2	151
NMC μp F_2	123
NMC μd F_2	123
NMC $\mu n/\mu p$	148
E665 μp F_2	53
E665 μd F_2	53
SLAC ep F_2	37
SLAC ed F_2	38
NMC/BCDMS/SLAC F_L	31
E866/NuSea pp DY	184
E866/NuSea pd/pp DY	15
NuTeV νN F_2	53
CHORUS νN F_2	42
NuTeV νN xF_3	45
CHORUS νN xF_3	33
CCFR $\nu N \to \mu\mu X$	86
NuTeV $\nu N \to \mu\mu X$	84
All data sets	**2743**

- Red = New w.r.t. MRST 2006 fit.

Input parameterisation in MSTW 2008 NLO (prel.) fit

At input scale $Q_0^2 = 1\ \text{GeV}^2$:

$$
\begin{aligned}
xu_v &= A_u\, x^{\eta_1}(1-x)^{\eta_2}(1+\epsilon_u\sqrt{x}+\gamma_u\, x)\\
xd_v &= A_d\, x^{\eta_3}(1-x)^{\eta_4}(1+\epsilon_d\sqrt{x}+\gamma_d\, x)\\
xS &= A_S\, x^{\delta_S}(1-x)^{\eta_S}(1+\epsilon_S\sqrt{x}+\gamma_S\, x)\\
x\bar{d}-x\bar{u} &= A_\Delta\, x^{\eta_\Delta}(1-x)^{\eta_S+2}(1+\gamma_\Delta\, x+\delta_\Delta\, x^2)\\
xg &= A_g\, x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g\sqrt{x}+\gamma_g\, x)+A_{g'}\, x^{\delta_{g'}}(1-x)^{\eta_{g'}}\\
xs+x\bar{s} &= A_+\, x^{\delta_S}\,(1-x)^{\eta_+}(1+\epsilon_S\sqrt{x}+\gamma_S\, x)\\
xs-x\bar{s} &= A_-\, x^{\delta_-}(1-x)^{\eta_-}(1-x/x_0)
\end{aligned}
$$

- A_u, A_d, A_g and x_0 are determined from sum rules.
- 20 parameters allowed to go free for eigenvector PDF sets, cf. 15 for MRST eigenvector PDF sets.

Use of eigenvector PDF sets (pioneered by CTEQ)

- Convenient to diagonalise covariance matrix $C \equiv H^{-1}$:

$$\sum_j C_{ij} v_{jk} = \lambda_k v_{ik},$$

  where λ_k is the kth eigenvalue and v_{ik} is the ith component of the kth orthonormal eigenvector $(i, j, k = 1, \ldots, N_{\rm parameters})$.
- Fitting groups produce eigenvector PDF sets $S_k^\pm$ with parameters a_i shifted from the global minimum:

$$a_i(S_k^\pm) = a_i^0 \pm t \sqrt{\lambda_k} v_{ik},$$

  with t adjusted to give the desired tolerance $T = \sqrt{\Delta\chi^2_{\rm global}}$.
- Then users can calculate uncertainties on a quantity F with

$$\Delta F = \frac{1}{2}\sqrt{\sum_k \left[F(S_k^+) - F(S_k^-)\right]^2},$$

  or using a formula to account for asymmetric errors.

Criteria for choice of tolerance $T = \sqrt{\Delta\chi^2_{\rm global}}$

Parameter-fitting criterion

- $T^2 = 1$ for 68% (1-σ) C.L., $T^2 = 2.71$ for 90% C.L.
- Appropriate if fitting consistent data sets with ideal Gaussian errors to a well-defined theory.
- **In practice:** minor inconsistencies between fitted data sets, and unknown experimental and theoretical uncertainties, so not appropriate for global PDF analysis.

Hypothesis-testing criterion (proposed by CTEQ)

- Much weaker than the parameter-fitting criterion: treat eigenvector PDF sets as alternative hypotheses.
- Determine T^2 from the criterion that each data set should be described within its 90% C.L. limit.

Criteria for choice of tolerance $T = \sqrt{\Delta\chi^2_{\rm global}}$

Parameter-fitting criterion

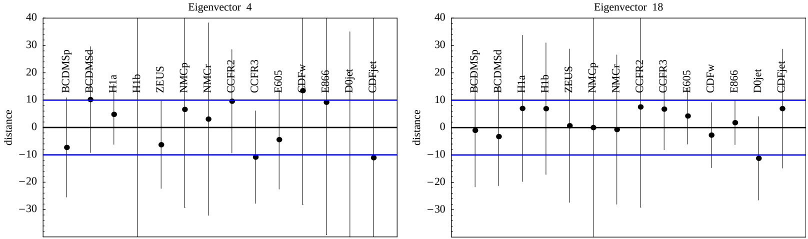
- $T^2 = 1$ for 68% (1-σ) C.L., $T^2 = 2.71$ for 90% C.L.
- Appropriate if fitting consistent data sets with ideal Gaussian errors to a well-defined theory.
- **In practice:** minor inconsistencies between fitted data sets, and unknown experimental and theoretical uncertainties, so not appropriate for global PDF analysis.

Hypothesis-testing criterion (proposed by CTEQ)

- Much weaker than the parameter-fitting criterion: treat eigenvector PDF sets as alternative hypotheses.
- Determine T^2 from the criterion that each data set should be described within its 90% C.L. limit.

Choice of tolerance by CTEQ [hep-ph/0201195]

- For each eigenvector, plot location of the minimum w.r.t. each data set and the 90% C.L. limits as the distance from the global minimum in units of $\sqrt{\Delta\chi^2_{\mathrm{global}}}$:



- A rough "average" over all eigenvectors gives $T = 10$ . . .
- . . . But $T = 10$ exceeds the 90% C.L. limits of some data sets.

Choice of tolerance by MRST [hep-ph/0211080]

"We estimate $\Delta\chi^2 = 50$ *to be a conservative uncertainty (perhaps of the order of a* 90% *confidence level or a little less than* 2σ*) due to the observation that* **an increase of 50 in the global** $\boldsymbol{\chi^2}$, *which has a value* $\chi^2 = 2328$ *for* 2097 *data points,* **usually signifies that the fit to one or more data sets is becoming unacceptably poor***. We find that* **an increase** $\Delta\chi^2$ **of 100 normally means that some data sets are very badly described** *by the theory."*

- Fairly qualitative statements.
- ⇒ Study more quantitatively in new MSTW analysis using same procedure applied in original CTEQ6 analysis.

Determination of 90% C.L. region following CTEQ6

- Define 90% C.L. region for each data set n (with N_n data points) as

$$\boxed{\chi_n^2 < \left(\frac{\chi_{n,0}^2}{\xi_{50}}\right)\,\xi_{90}}$$

- ξ_{90} is the 90th percentile of the χ^2-distribution with N_n d.o.f., i.e.

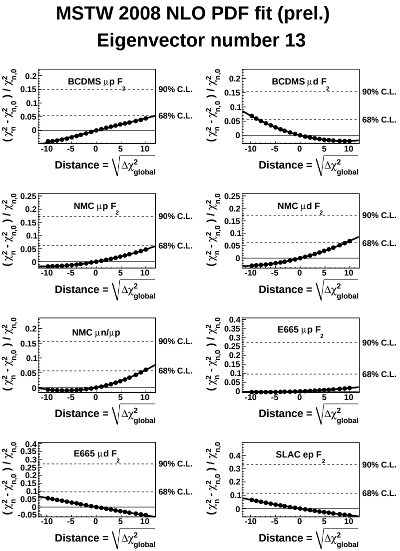
$$\int_0^{\xi_{90}} \mathrm{d}\chi^2 \; f(\chi^2; N_n) = 0.90,$$

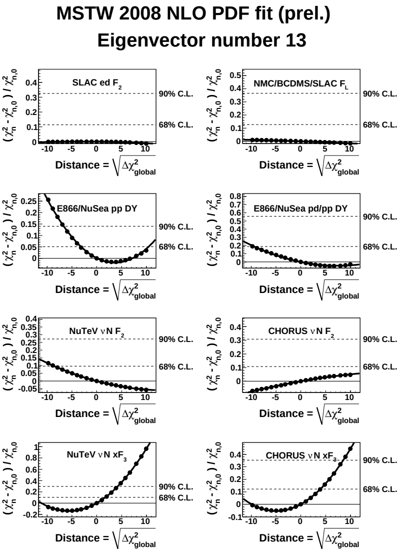
  where the probability density function is

$$f(z; N) = \frac{z^{N/2-1}\,\mathrm{e}^{-z/2}}{2^{N/2}\,\Gamma(N/2)}.$$

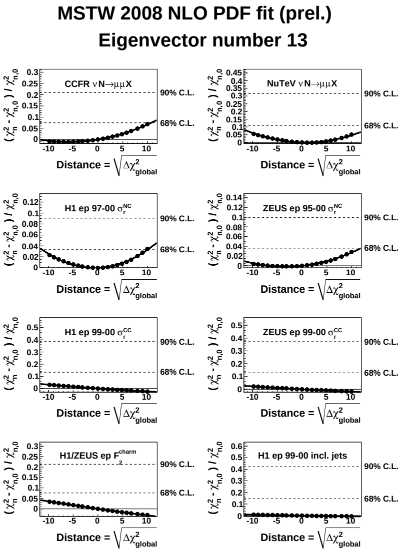
- $\xi_{50} \simeq N_n$ is the most probable value of the χ^2-distribution.
- $\chi_{n,0}^2$ for data set n is evaluated at the **global** minimum.
- **Rescale** by a factor $\chi_{n,0}^2/\xi_{50}$ since this often deviates from 1.
- Similarly for the 68% C.L. region.

Fractional change in χ^2 along eigenvector number 13

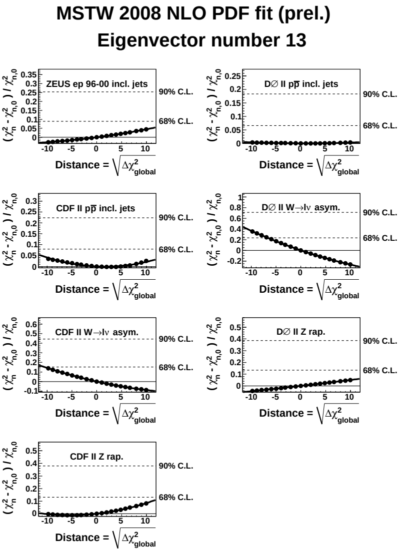




Fractional change in χ^2 along eigenvector number 13

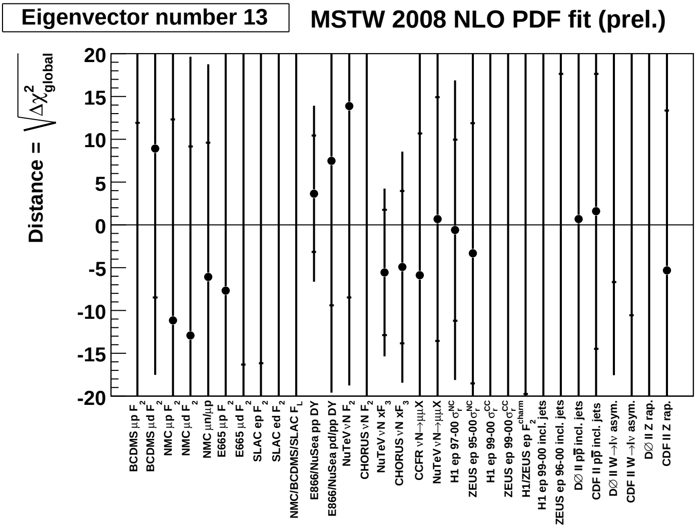


MSTW 2008 NLO PDF fit (prel.)
Eigenvector number 13

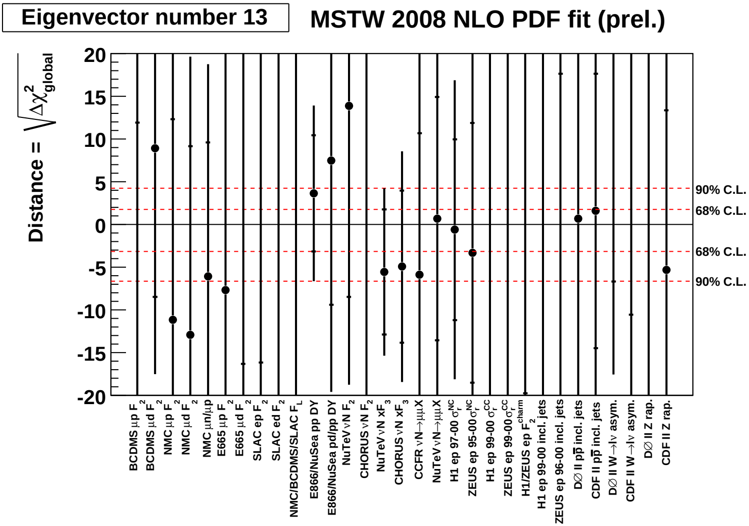


MSTW 2008 NLO PDF fit (prel.)
Eigenvector number 13

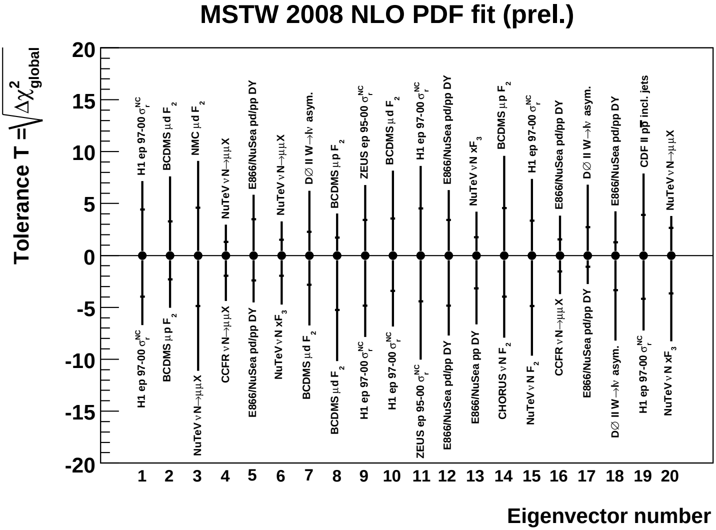
Determination of tolerance for eigenvector number 13



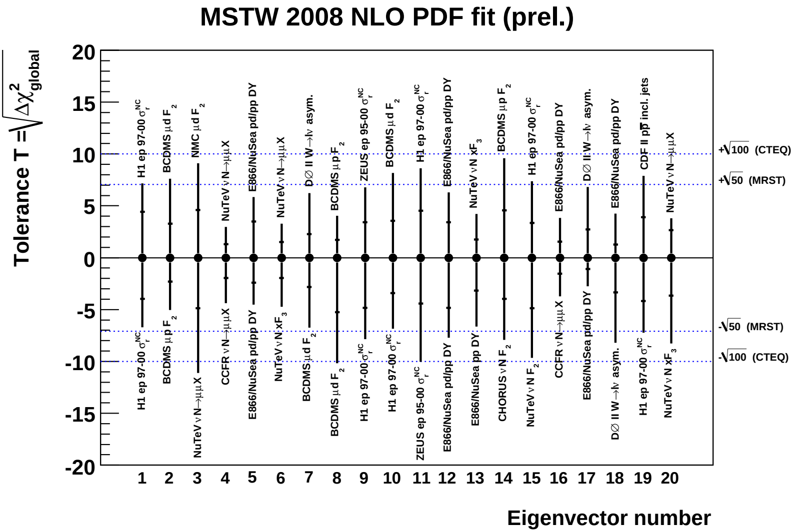
Determination of tolerance for eigenvector number 13



Tolerance vs. eigenvector number



Tolerance vs. eigenvector number



Treatment of jet data in MRST/MSTW analyses

MRST 2001–2006

- Fit six "pseudogluon" points at $Q^2 = 2000$ GeV^2 inferred from Tevatron Run I inclusive jet data.
- Comparison to actual jet data, calculated at LO with a K-factor, only made **after** the fit.

MSTW 2008

- Fit to Tevatron Run II and HERA DIS inclusive jet data.
- Complete treatment of correlated systematic errors.
- Use `fastNLO` code[a] to calculate NLO cross sections.
- At NNLO, include 2-loop threshold corrections[b] for Tevatron jet data and exclude HERA DIS jet data.

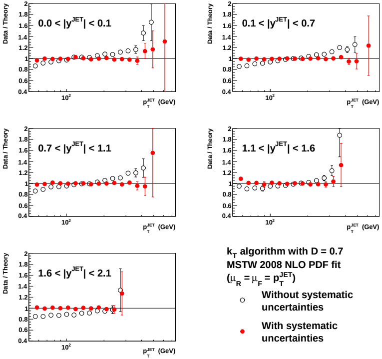
---

[a] Kluge, Rabbertz, Wobisch, hep-ph/0609285

[b] Kidonakis, Owens, hep-ph/0007268

Description of Tevatron Run II inclusive jet data

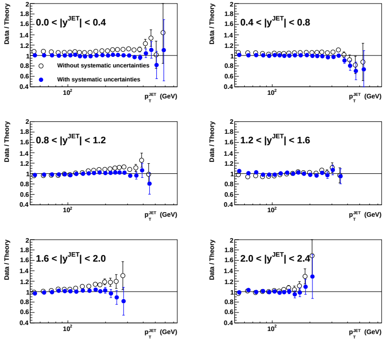
CDF Run II inclusive jet data, χ^2 = 61 for 76 pts.



[hep-ex/0701051]

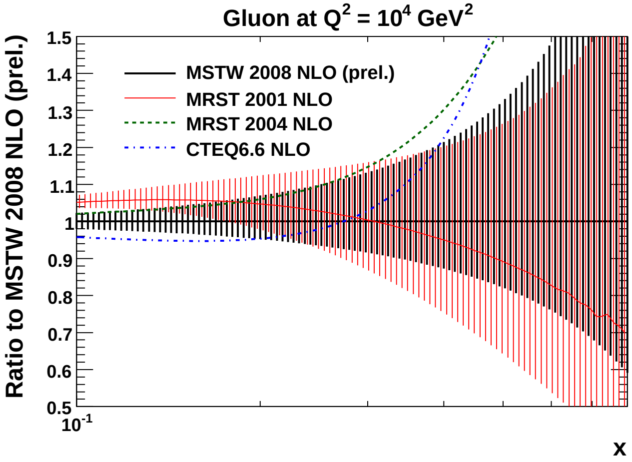
D∅ Run II inclusive jet data (cone, R = 0.7)

MSTW 2008 NLO PDF fit ($\mu_R = \mu_F = p_T^{JET}$), χ^2 = 115 for 110 pts.



[arXiv:0802.2400]

Impact of Run II jet data on high-x gluon distribution



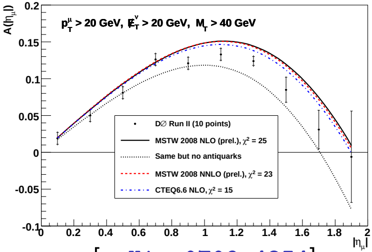
- Run II jet data prefer smaller gluon distribution at high x.

$W \rightarrow l\nu$ charge asymmetry from Tevatron Run II

$$A(\eta_l) = \frac{\mathrm{d}\sigma(l^+)/\mathrm{d}\eta_l - \mathrm{d}\sigma(l^-)/\mathrm{d}\eta_l}{\mathrm{d}\sigma(l^+)/\mathrm{d}\eta_l + \mathrm{d}\sigma(l^-)/\mathrm{d}\eta_l}$$

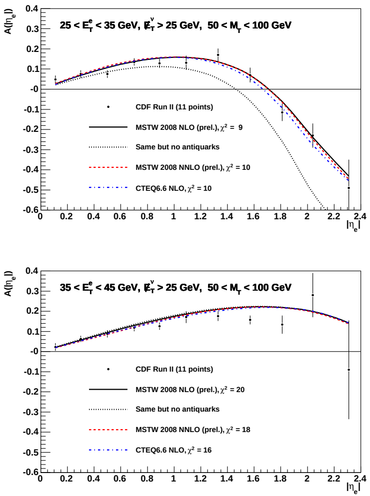
- Mainly constrains down quark.
- Calculations using `FEWZ` code [Melnikov, Petriello, `hep-ph/0609070`]

DØ data on lepton charge asymmetry from W→μν decays



[`arXiv:0709.4254`]

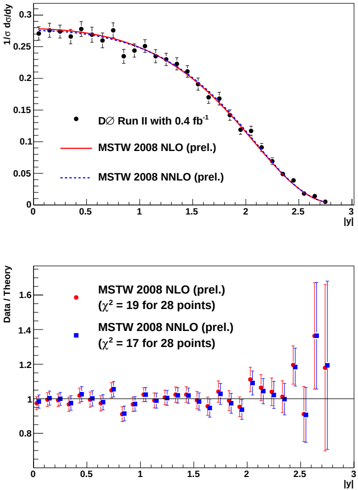
CDF data on lepton charge asymmetry from W→eν decays



[`hep-ex/0501023`]

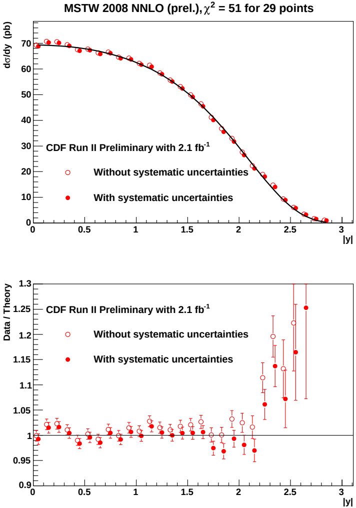
Z/γ^* rapidity distributions from Tevatron Run II

Z/γ* rapidity shape distribution from DØ



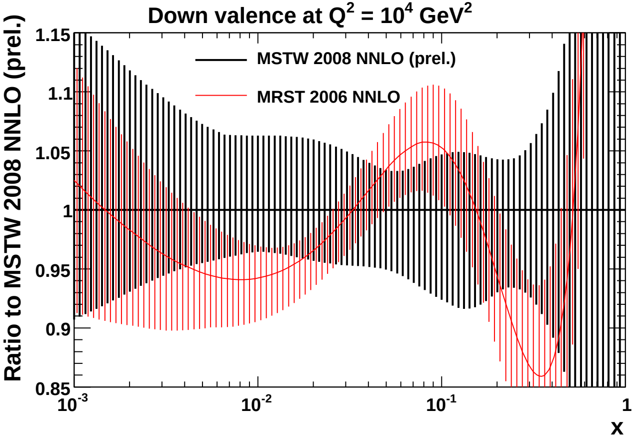
[hep-ex/0702025]

Z/γ* rapidity distribution from CDF



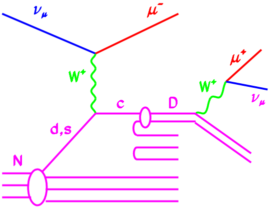
[CDF Preliminary, February 2008]

Impact on down valence quark distribution



- Different shape compared to MRST 2006 NNLO.
- (Better choice of parameters for uncertainty determination.)

NuTeV/CCFR dimuon cross sections and strangeness



$$\frac{\mathrm{d}\sigma}{\mathrm{d}x\mathrm{d}y}(\nu_\mu N \to \mu^+\mu^- X) \propto \frac{\mathrm{d}\sigma}{\mathrm{d}x\mathrm{d}y}(\nu_\mu N \to \mu^- c\, X)$$

- ν_μ and $\bar{\nu}_\mu$ cross sections constrain s and $\bar{s}$.

- Can **relax assumption** made in previous MRST fits that

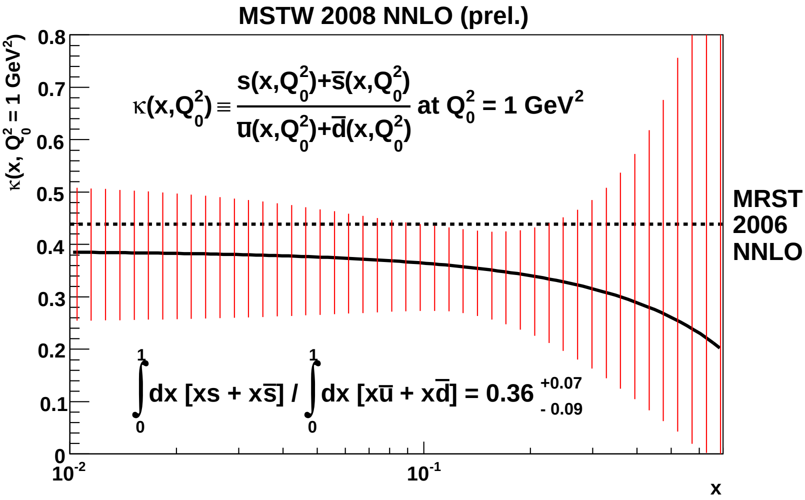
$$s(x,Q_0^2) = \bar{s}(x,Q_0^2) = \frac{\kappa}{2}\left[\bar{u}(x,Q_0^2) + \bar{d}(x,Q_0^2)\right], \text{ with } \kappa \approx 0.5.$$

- **Parameterise** at input scale of $Q_0^2 = 1$ GeV^2 in the form:

$$xs(x,Q_0^2) + x\bar{s}(x,Q_0^2) = A_+\,(1-x)^{\eta_+}\,xS(x,Q_0^2),$$
$$xs(x,Q_0^2) - x\bar{s}(x,Q_0^2) = A_-\,x^{0.2}(1-x)^{\eta_-}(1-x/x_0).$$

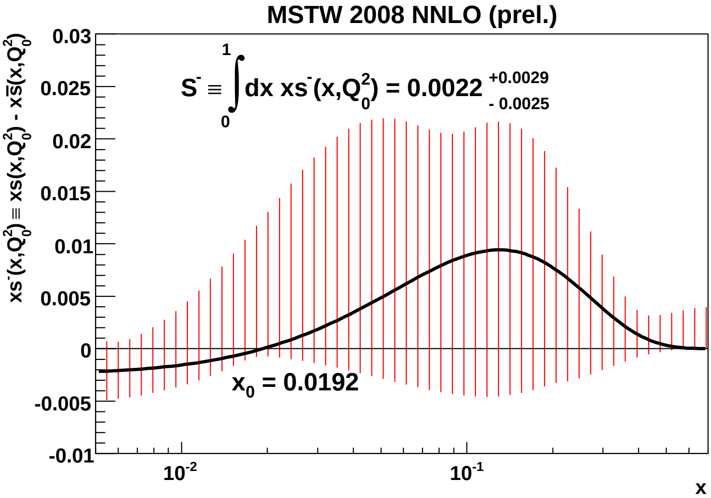
- General-mass variable-flavour-number scheme used for CC charm production up to NNLO (→ talk by R. Thorne).

Ratio of strange sea to non-strange sea: $(s+\bar{s})/(\bar{u}+\bar{d})$



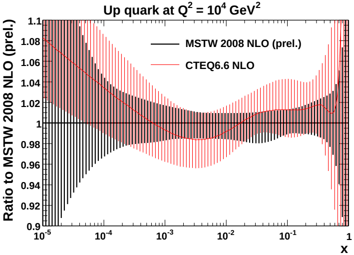
- Suppression of strange sea at large x.

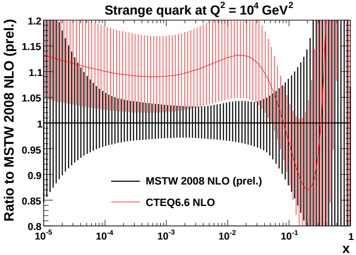
Strange sea asymmetry: $xs - x\bar{s}$

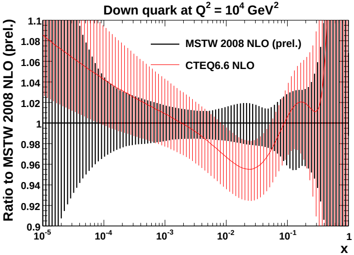


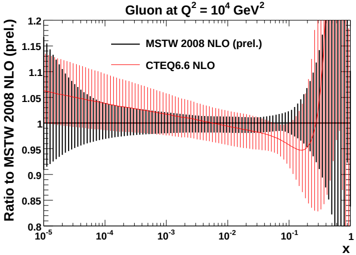
- Consistent with zero within 90% C.L. limit.

MSTW 2008 NLO (prel.) compared to CTEQ6.6

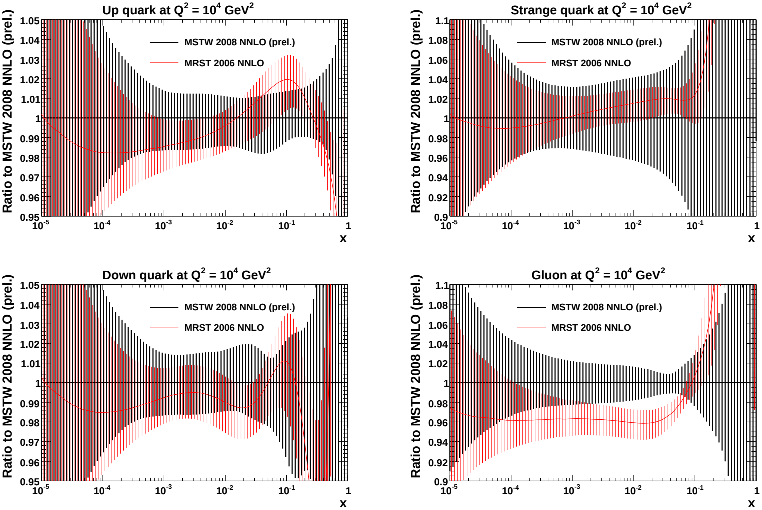

Up quark at $Q^2 = 10^4$ GeV^2


Strange quark at $Q^2 = 10^4$ GeV^2


Down quark at $Q^2 = 10^4$ GeV^2


Gluon at $Q^2 = 10^4$ GeV^2

MSTW 2008 NNLO (prel.) compared to MRST 2006



W and Z total cross sections at LHC (Tevatron)

	$B_{l\nu} \cdot \sigma_W$ (nb)	$B_{l^+l^-} \cdot \sigma_Z$ (nb)
MSTW 2008 NLO (prel.)	20.45 (2.650)	1.965 (0.2425)
MSTW 2008 NNLO (prel.)	21.44 (2.739)	2.043 (0.2512)

Ratio to MSTW 2008 (prel.)	σ_W	σ_Z
MRST 2006 NLO (unpublished)	1.002 (0.995)	1.009 (1.001)
MRST 2006 NNLO	0.995 (1.004)	1.001 (1.010)
MRST 2004 NLO	0.974 (0.990)	0.982 (1.000)
MRST 2004 NNLO	0.936 (0.991)	0.940 (1.003)
CTEQ6.6 NLO	1.019 (0.978)	1.022 (0.987)

- Increase from MRST 2004 to MRST 2006 due to change in heavy flavour prescription (→ talk by R. Thorne).
- Predictions stable in going from MRST 2006 to MSTW 2008.

Summary

- **Dynamic determination of tolerance:** different tolerance for each of the 40 eigenvector PDF sets ensuring that each data set is described within its 90% C.L. limit.
- **Tevatron Run II inclusive jet data** now included in global fit: smaller gluon distribution at high x than with Run I data.
- **Tevatron Run II W and Z data** also now included: some influence on down quark distribution.
- **Strange quark and antiquark** distributions are now constrained by NuTeV/CCFR dimuon data.
- **Predictions for W and Z total cross sections** at Tevatron and LHC are stable to addition of new data.