Parton distributions for the LHC

Graeme Watt

University College London

Cavendish HEP seminar
University of Cambridge
3rd June 2008

Pictorial representation of a typical LHC event

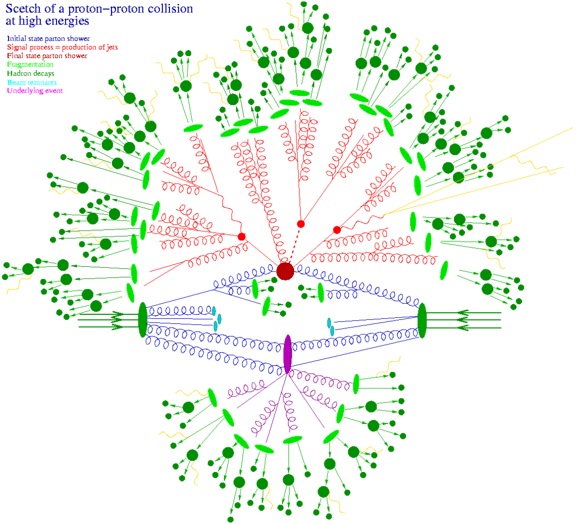
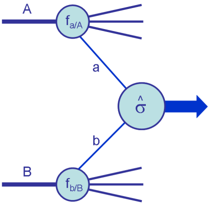


Figure from `http://www.sherpa-mc.de`

- Parton Distribution Functions (PDFs), $f_{a/p}(x, Q^2)$, give the **density** of partons of flavour $a = q, g$ in the proton with momentum fraction x at a scale Q^2.

Fixed-order collinear factorisation at hadron colliders

- The "standard" perturbative QCD formalism: work at Leading-Order (LO), Next-to-Leading Order (NLO), etc.
- **Hadronic** cross sections given by convolution of **partonic** cross sections with PDFs:



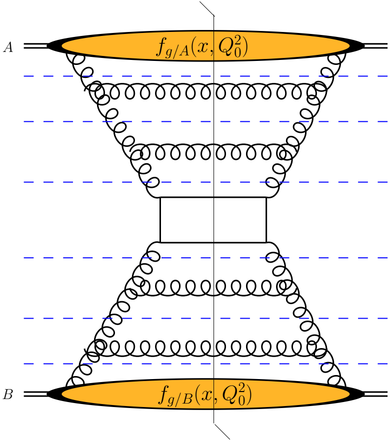
$$\sigma_{AB} = \sum_{a,b=q,g} \left[\hat{\sigma}_{ab}^{\rm LO} + \alpha_S(Q^2)\hat{\sigma}_{ab}^{\rm NLO} + \ldots\right] \otimes f_{a/A}(x_a, Q^2) \otimes f_{b/B}(x_b, Q^2)$$

PDF evolution: $$\frac{\partial f_a}{\partial \ln Q^2} = \sum_{a'=q,g} \left[P_{aa'}^{\rm LO} + \alpha_S P_{aa'}^{\rm NLO} + \ldots\right] \otimes f_{a'}.$$

α_S evolution: $$\frac{\partial \alpha_S}{\partial \ln Q^2} = -\beta^{\rm LO}\alpha_S^2 - \beta^{\rm NLO}\alpha_S^3 - \ldots.$$

Need to extract input values $f_a(x, Q_0^2)$ and $\alpha_S(M_Z^2)$ from data.

Diagrammatic interpretation of collinear factorisation



- **Diagram** represents cross section for e.g. dijet production.
- **Virtualities** of *t*-channel partons are strongly-ordered and increase towards hard subprocess.
- **Factorisation** between each "cell" separated by horizontal dashed lines.
- **Higher-order corrections** applied separately both to the splitting function cells and the hard subprocess cell.

Paradigm for PDF determination by "global analysis"

1. **Parameterise** the x dependence for each flavour $a = q, g$ at the input scale $Q_0^2 \sim 1\ \text{GeV}^2$ in some flexible form, e.g.

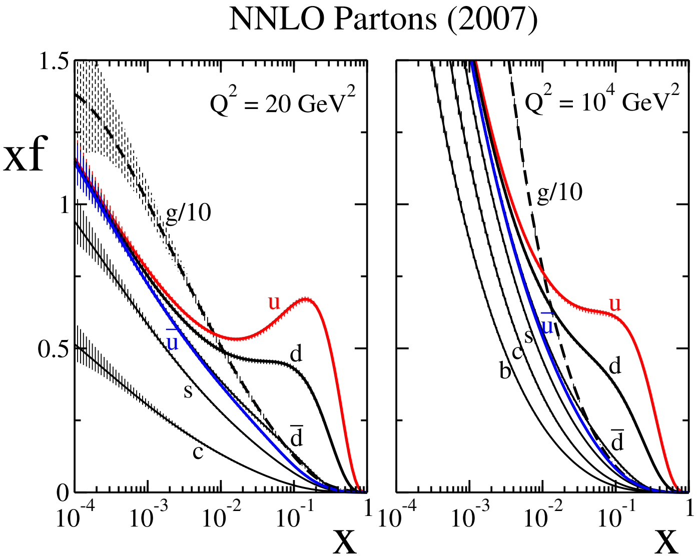
$$xf_a(x, Q_0^2) = A_a x^{\Delta_a} (1-x)^{\eta_a} (1 + \epsilon_a \sqrt{x} + \gamma_a x),$$

subject to number- and momentum-sum rule constraints.

2. **Evolve** the PDFs to higher scales $Q^2 > Q_0^2$ using the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution equations.
3. **Convolute** the evolved PDFs with the partonic cross sections to calculate hadronic cross sections corresponding to a wide variety of deep-inelastic scattering (and related) data.
4. **Vary** the input parameters $\{A_a, \Delta_a, \eta_a, \epsilon_a, \gamma_a, \ldots\}$ to minimise

$$\chi^2 = \sum_{i=1}^{N_{\text{pts.}}} \left(\frac{\text{Data}_i - \text{Theory}_i}{\text{Error}_i} \right)^2 .$$

Example of PDFs obtained from global analysis



Determination of parton distributions by global analysis

- An "industry" for more than 20 years. Regular updates as new data and theory become available.
- Major group: "**MRST**" = Martin+Roberts+Stirling+Thorne. Recently, "**MSTW**" = MRST−Roberts+G.W.
- Other major group: Coordinated Theoretical–Experimental Project on QCD (CTEQ). Recent CTEQ6.6 analysis by groups at Michigan State/Taiwan/Washington (MSTW).
- PDF sets publically available through LHAPDF interface: `http://projects.hepforge.org/lhapdf`
- Why bother?
  1. **A Fundamental Measurement**: a challenge to understand using methods of nonperturbative QCD.
  2. **A Necessary Evil**: essential input to perturbative calculations of signal and background at hadron colliders.

[J. Pumplin, talk at DIS 2005, hep-ph/0507093]

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.

Uncertainties in global PDF analysis

Theoretical errors

- *Examples:* input parameterisation form, neglected higher-order and higher-twist QCD corrections, electroweak corrections, choice of cuts, nuclear corrections, heavy flavour treatment.
- Difficult to quantify *a priori*. Often correction only known after an improved treatment/calculation is available.

Experimental errors: generally use Hessian method

- Assume $\chi^2_{\rm global}$ is quadratic about the global minimum $\{a_i^0\}$:

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

where the Hessian matrix $H_{ij} = \frac{1}{2}\frac{\partial^2\,\chi^2_{\rm global}}{\partial a_i \partial a_j}\Big|_{\rm minimum}$

Uncertainties in global PDF analysis

Theoretical errors

- *Examples:* input parameterisation form, neglected higher-order and higher-twist QCD corrections, electroweak corrections, choice of cuts, nuclear corrections, heavy flavour treatment.
- Difficult to quantify *a priori*. Often correction only known after an improved treatment/calculation is available.

Experimental errors: generally use Hessian method

- Assume $\chi^2_{\rm global}$ is quadratic about the global minimum $\{a_i^0\}$:

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

where the Hessian matrix $H_{ij} = \frac{1}{2}\frac{\partial^2\,\chi^2_{\rm global}}{\partial a_i \partial a_j}\Big|_{\rm minimum}$

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:

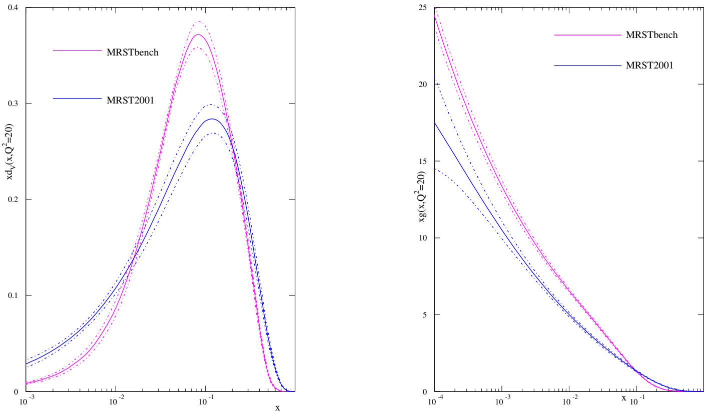
$$a_i(S_k^{\pm}) = a_i^0 \pm t\,\sqrt{\lambda_k} v_{ik},$$

  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.

Benchmark PDFs for HERA–LHC workshop [R. Thorne]



- MRSTbench: $T^2 = 1$, restricted fit, $Q^2 \geq 9$ GeV^2.
- MRST2001: $T^2 = 50$, full global fit, $Q^2 \geq 2$ GeV^2.
- Non-overlapping uncertainty bands $\Rightarrow$ inconsistent data sets are included in full global fit.

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 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.
- $\Rightarrow$ 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

$$\chi^2_n < \left(\frac{\chi^2_{n,0}}{\xi_{50}}\right)\,\xi_{90}$$

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

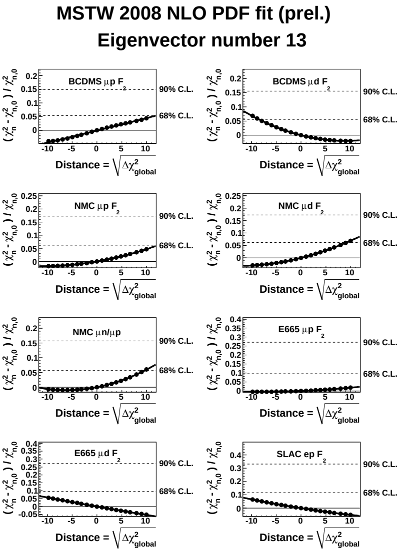
$$\int_0^{\xi_{90}} \mathrm{d}\chi^2\, f(\chi^2;N_n) = 0.90,$$

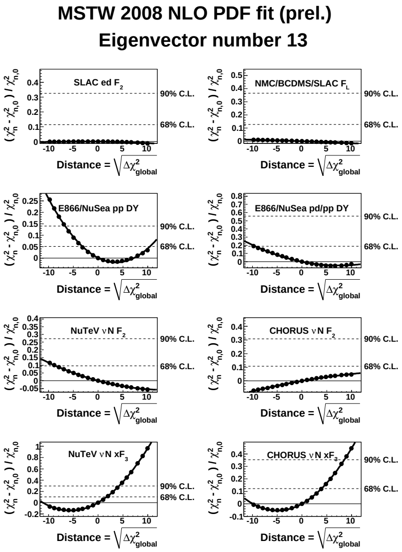
  where the probability density function is

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

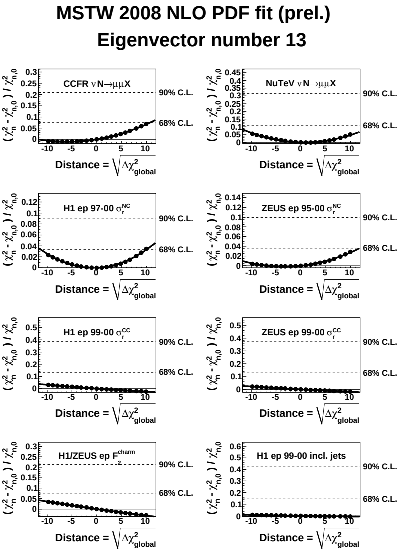
- $\xi_{50} \simeq N_n$ is the most probable value of the χ^2-distribution.
- $\chi^2_{n,0}$ for data set n is evaluated at the **global** minimum.
- **Rescale** by a factor $\chi^2_{n,0}/\xi_{50}$ since this often deviates from 1.
- Similarly for the 68% C.L. region.

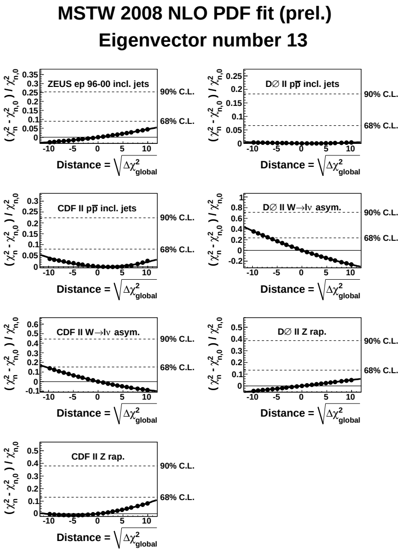
Fractional change in χ^2 along eigenvector number 13



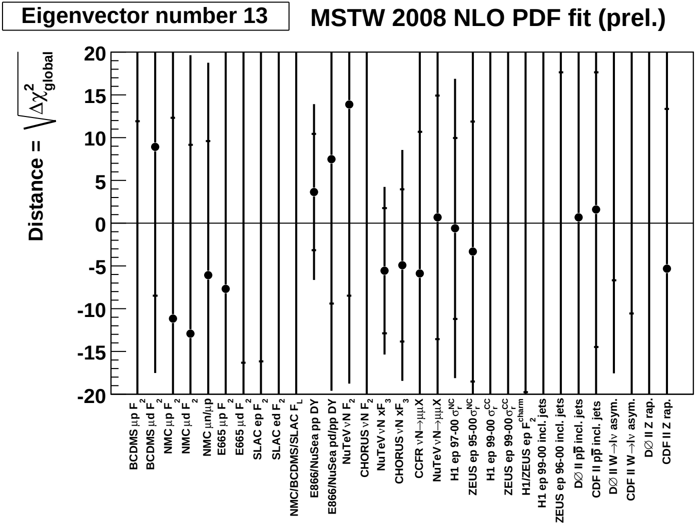


Fractional change in χ^2 along eigenvector number 13

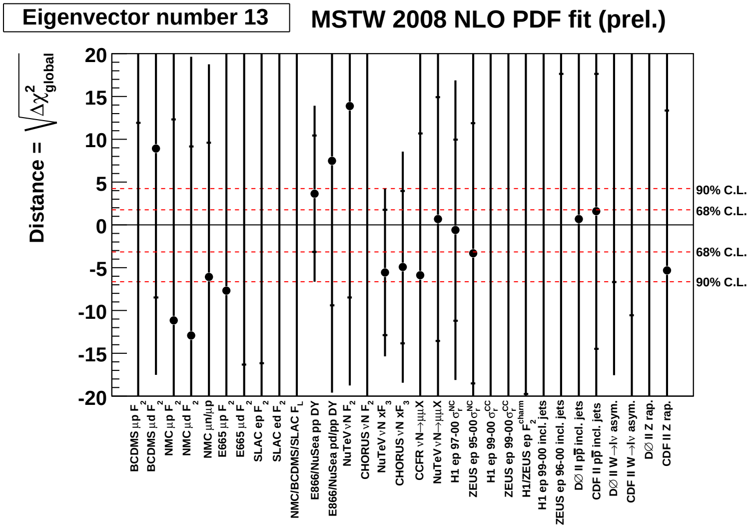




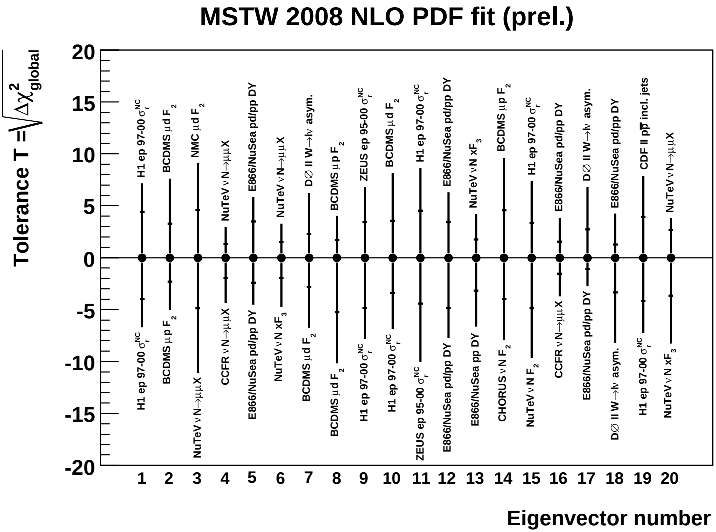
Determination of tolerance for eigenvector number 13



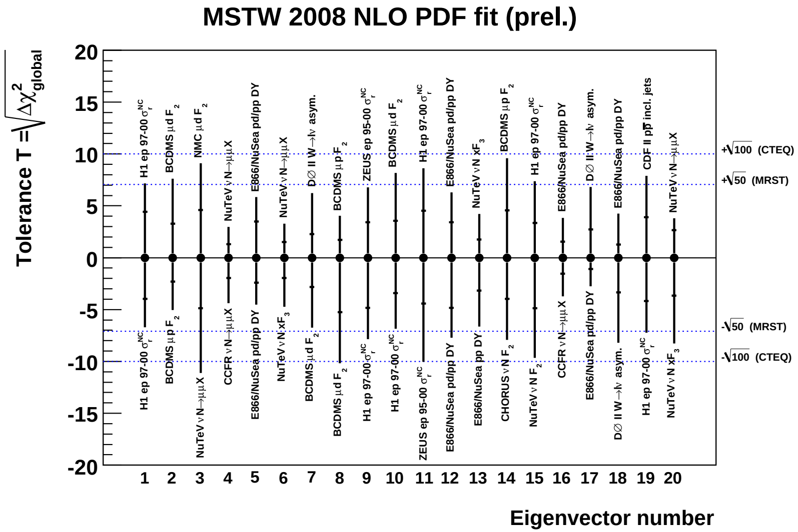
Determination of tolerance for eigenvector number 13



Tolerance vs. eigenvector number

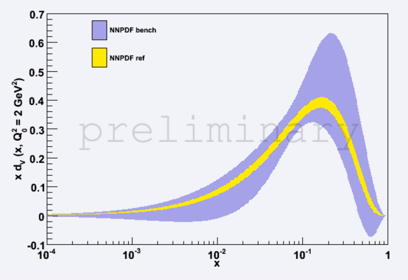


Tolerance vs. eigenvector number

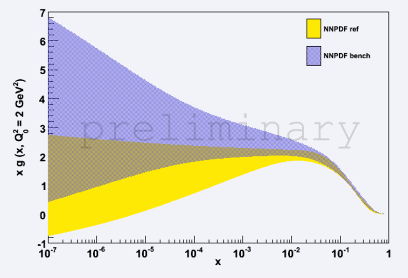


Alternative: neural network determination of PDFs

NNPDF Collaboration: use combination of Monte Carlo techniques and neural networks as unbiased interpolating functions.



d-valence



gluon

- Statistically correct with no parameterisation bias.
- Non-singlet fit published [hep-ph/0701127], with extension to full DIS and global fits in progress.

Inclusion of jet data in PDF fits

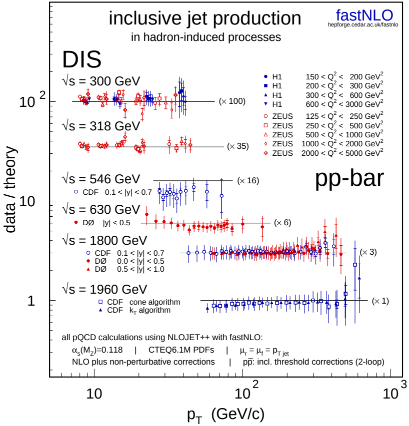
- **Tevatron jet data** are important in constraining the high-x gluon distribution. At NLO,

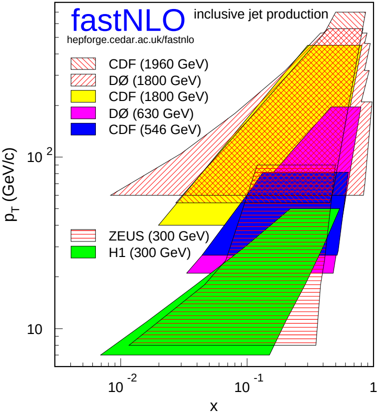
$$\sigma_{p\bar{p}} = \alpha_S^2(Q^2) \sum_{a,b=q,g} \left[\hat{\sigma}_{ab}^{\rm LO} + \alpha_S(Q^2)\hat{\sigma}_{ab}^{\rm NLO}\right] \otimes f_{a/p}(x_a, Q^2) \otimes f_{b/\bar{p}}(x_b, Q^2).$$

- Need multidimensional (Monte Carlo) phase space integration with cuts on E_T and η etc. $\Rightarrow$ takes hours/days of CPU time, so impractical to include in a PDF fit.
- **Old (approximate) solution:** calculate "K-factors" $(\sigma_{p\bar{p}}^{\rm NLO}/\sigma_{p\bar{p}}^{\rm LO})$ for a given set of PDFs. Then infer pseudogluon and $\Lambda_{\rm QCD}$ "data" points from $\sigma_{p\bar{p}}^{\rm LO}$.
- **New (rigorous) solution:** interpolate PDFs and α_S so that they can be factorised from $\hat{\sigma}_{ab}$. Replaces convolution with multiplication. Phase space integration only needs to be done once, with result stored in a grid for later use during a PDF fit.

"fastNLO": Fast pQCD Calculations for PDF Fits

Grids corresponding to Tevatron and HERA kinematic cuts provided by "**fastNLO**" project.[1]

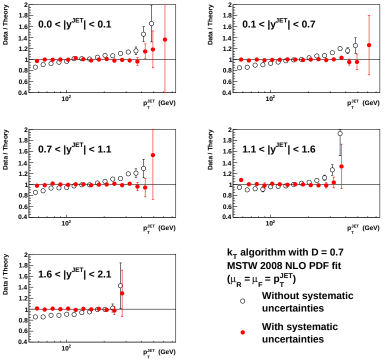




[1] T. Kluge, K. Rabbertz, M. Wobisch, hep-ph/0609285.

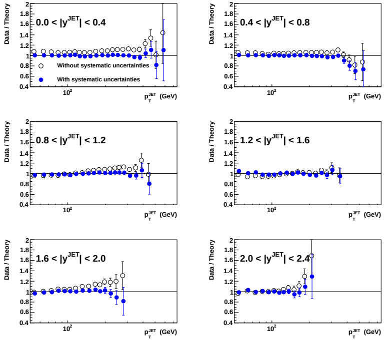
Description of Tevatron Run II inclusive jet data

CDF Run II inclusive jet data, χ^2 = 55 for 76 pts.



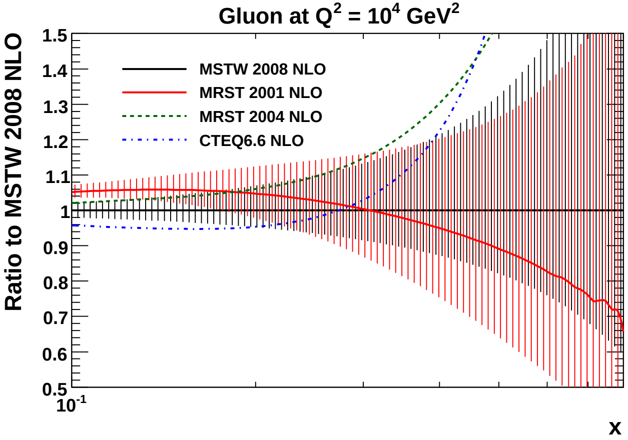
[hep-ex/0701051]

DØ 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



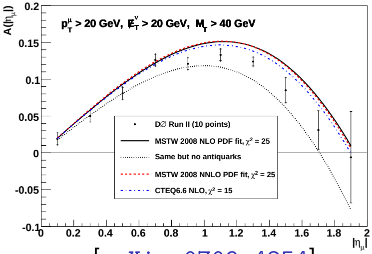
- Run II jet data prefer smaller gluon distribution at high x.

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

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

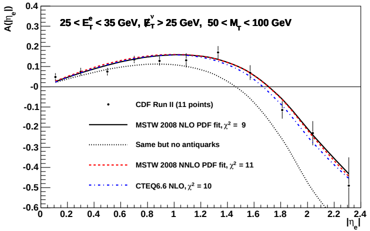
- Mainly constrains down quark.
- Calculations using FEWZ code [Melnikov, Petriello, hep-ph/0609070]

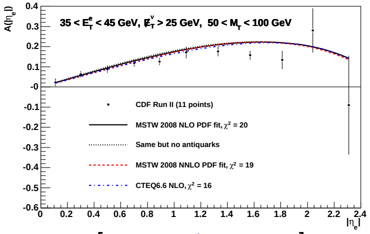
DØ data on lepton charge asymmetry from W→μν decays



[arXiv:0709.4254]

CDF data on lepton charge asymmetry from W→eν decays

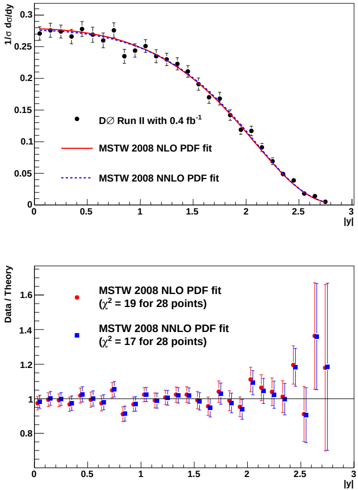




[hep-ex/0501023]

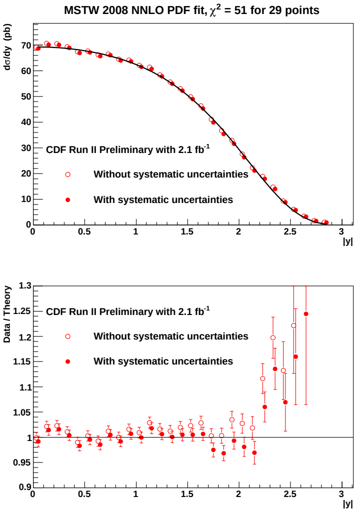
Z/γ^* rapidity distributions from Tevatron Run II

Z/γ* rapidity shape distribution from D∅



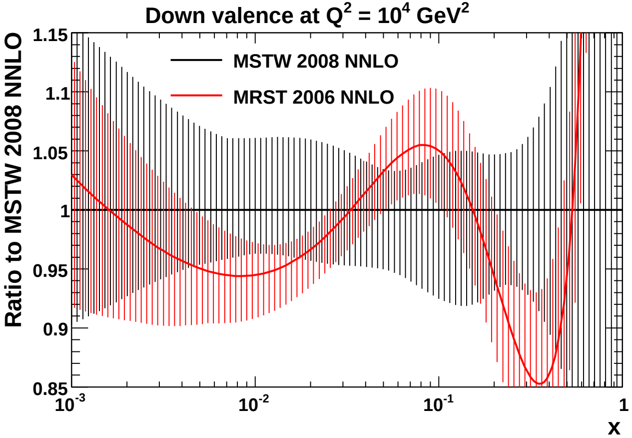
[hep-ex/0702025]

Z/γ* rapidity distribution from CDF



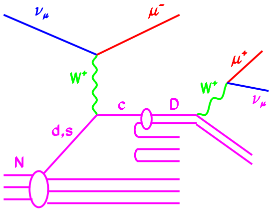
[CDF Preliminary, February 2008]

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

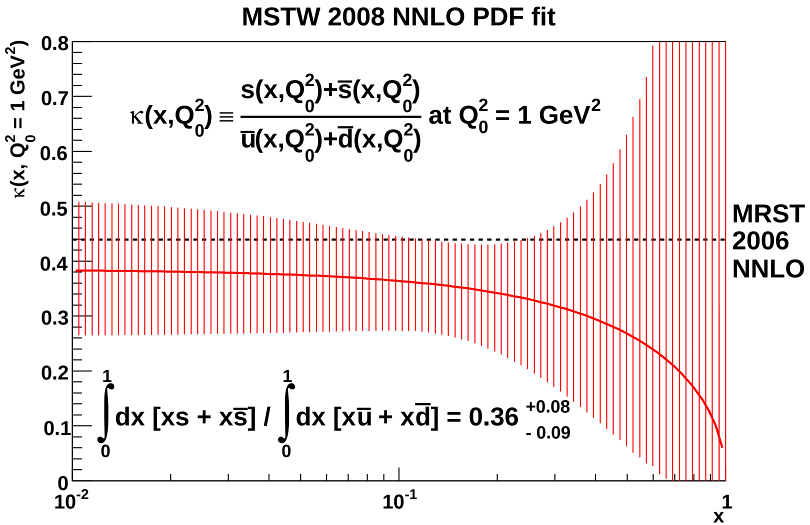
$$s(x,Q_0^2) = \bar{s}(x,Q_0^2) = \frac{\kappa}{2}\left[\bar{u}(x,Q_0^2) + \bar{d}(x,Q_0^2)\right], \text{ with } \kappa \approx 0.5.$$

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

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

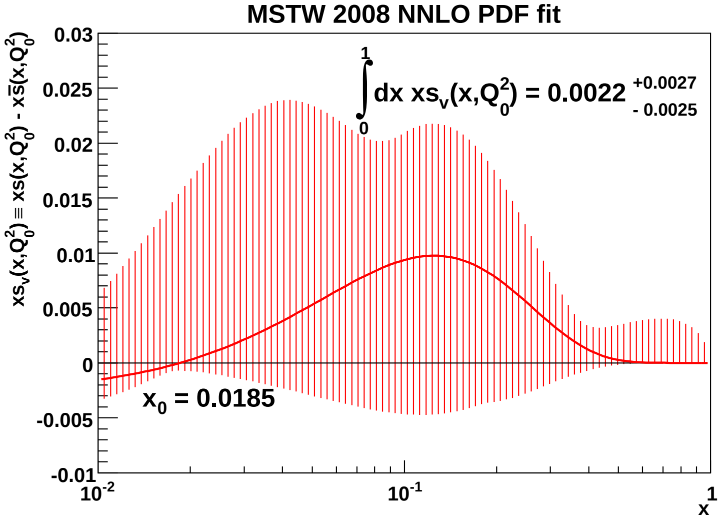
- x_0 fixed by zero strangeness: $\int_0^1 \mathrm{d}x\,\left[s(x,Q_0^2) - \bar{s}(x,Q_0^2)\right] = 0$.

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



- Suppression of strange sea at large x.

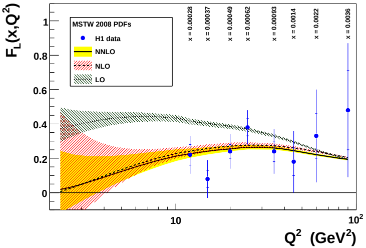
Strange sea asymmetry: $xs - x\bar{s}$



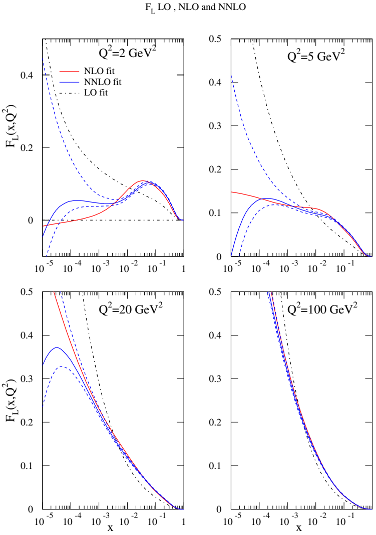
- Consistent with zero within 90% C.L. limit.

Longitudinal structure function $F_L(x, Q^2)$

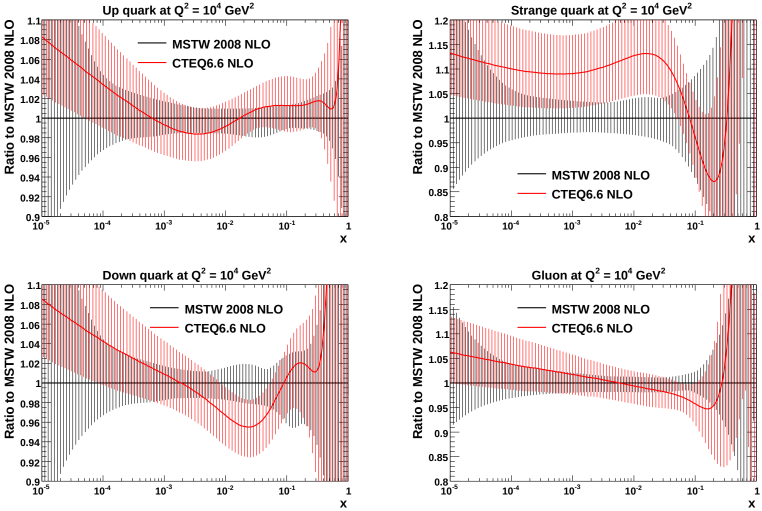
- HERA operated at **reduced proton beam energies** in last few months of running, primarily to measure F_L.



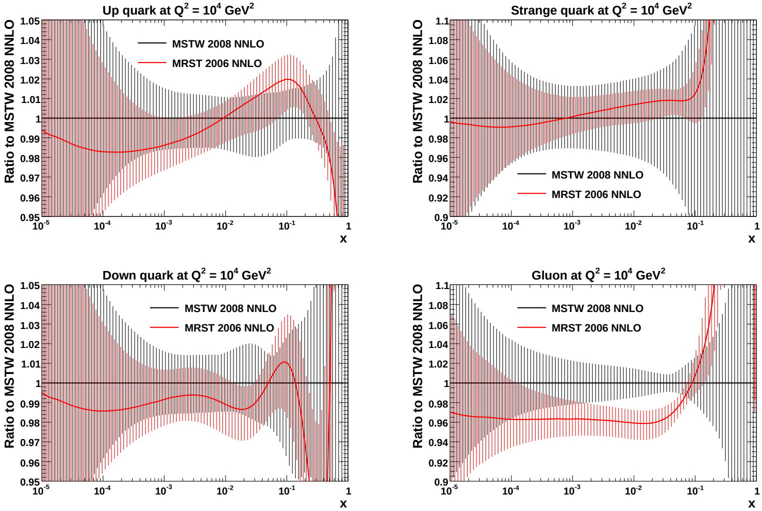
- H1 F_L data from arXiv:0805.2809.



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



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.
- Predictions stable in going from MRST 2006 to MSTW 2008.

For more information . . .

HERA–LHC workshop (2004–2008)

- Last meeting May 26–30 at CERN.
- http://www.desy.de/∼heralhc

PDF4LHC workshop (2008–)

- First meeting February 22–23 at CERN.
- http://indico.cern.ch/conferenceDisplay.py?confId=27439

Summary

- **Parton distributions** will be crucial for the interpretation of data and the identification of new physics at the LHC.
- **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.
- **Outlook:** Will provide LO, NLO, NNLO PDFs, each with 40 additional eigenvector PDF sets, for first LHC running.